Effects of inflammatory conditions on the structure of fibrin clots: modeling a flexible protein dimer with molecular dynamics simulations

Eric Pederson

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Gianluca Interlandi, Chair

Wendy Thomas

Lutz Maibaum

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

Effects of inflammatory conditions on the structure of fibrin clots:
modeling a flexible protein dimer with molecular dynamics simulations

Eric Pederson

Chair of the Supervisory Committee:
Research Assistant Professor Gianluca Interlandi
Department of Bioengineering

Upon vascular injury, fibrin is converted from fibrinogen and polymerizes to form long protofibrils. Protofibrils laterally aggregate via the αC domain to form fibrin fibers, which bind to platelets and form a clot. Inflammation recruits pro-inflammatory neutrophils to the site of injury, which produce HOCl via myeloperoxidase. Inflammation is associated with abnormal clot morphologies, which can lead to thrombosis and bleeding. Fibrin Met^{476}, located in the αC domain, oxidizes upon exposure to myeloperoxidase-derived HOCl, which subsequently disrupts lateral aggregation and produces abnormal clots. The mechanism by which oxidation affects αC domain structure, dynamics and self-association was investigated by enhanced-sampling molecular dynamics techniques and free energy calculations. Replica exchange molecular dynamics simulations with an implicit solvent were used to efficiently sample αC-domain dimer conformations. Low-energy conformers were identified from the free energy landscape minima and were used for free energy calculations after equilibration in explicit solvent. Using this method, I propose the hydrophobic core model of αC polymerization, which states that Met^{476} is a docking spot for αC polymerization.

Contents

List of Figures

List of Tables

Glossary

2-D: 2-dimensional

3-D: 3-dimensional

α2-AP: α2-antiplasmin

αC: C-terminus of Aα chain

AFB: Adaptive force bias

AFM: Atomic force microscopy

ALA: Alanine

aMD: Accelerated molecular dynamics

Å: Ångstrom (0.1 nm)

APR: Acute phase response

ASP: Aspartic acid

BAR: Bennett's acceptance ratio

BLAST: Basic local alignment search tool

Ca^{2+}: Calcium ion

CHARMM: Chemistry at Harvard macromolecular mechanics

CYS: Cysteine

CV: Collective variable

DAMP: Damage-associated molecular pattern

DHA: Docosahexaenoic acid

DMD: Discrete molecular dynamics

DNA: Deoxyribonucleic acid

DNase: Deoxyribonuclease

DoF: Degree(s) of freedom

DSC: Differential scanning calorimetry

DSF: Differential scanning fluorimetry

DVT: Deep vein thrombosis

EPA: Eicosapentaenoic acid

FCS: Fluorescence correlation spectroscopy

FEL: Free energy landscape

FpA: Fibrinopeptide A

FpB: Fibrinopeptide B

FEP: Free energy perturbation

FF: Force field

FFT: Fast Fourier transform

FnE: Fibrin fragment E

FS: Femtosecond

FIX(a): Factor IX(a)

FV(a): Factor V(a)

FVII(a): Factor VII(a)

FVIII(a): Factor VIII(a)

FX(a): Factor X(a)

FXI(a): Factor XI(a)

FXII(a): Factor XII(a)

FXIII(a): Factor XIII(a)

GBSA: Generalized Born with surface area

GLN: Glutamine

GAS: Group A *Streptococci*

GBS: Group B *Streptococci*

HC: Hydrophobic core

HDX: Hydrogen-deuterium exchange

HIS: Histidine

HMWK: High molecular weight kininogen

HOCL: Hypochlorous acid

HPC: High performance computing

IDP: Intrinsically disordered protein

IDR: Intrinsically disordered region

IFN: Interferon

IL: Interleukin

kDa: Kilodalton

LPS: Lipopolysaccharide

LYS: Lysine

mAb: Monoclonal antibody

MD: Molecular dynamics

MET: Methionine

MET(O): Methionine sulfoxide

MSRA: Methionine sulfoxide reductase A

MSRB: Methionine sulfoxide reductase B

MPO: Myeloperoxidase

NADPH: Nicotinamide adenine dinucleotide phosphate

NAMD: Nanoscale molecular dynamics

nm: Nanometer

NET: Neutrophil extracellular trap

NETosis: NET formation

NMR: Nuclear magnetic resonance

NOE: Nuclear Overhauser effect

PAI: Plasminogen activator inhibitor

PAMP: Pathogen-associated molecular pattern

PAR: Protease-activated receptor

PBC: Periodic boundary conditions

PRN: Pseudorandom number generator

PDB: Protein Database

PL: Phospholipid

PPI: Protein-protein interaction

PTase: Prothrombinase

PTM: Post-translational modification

PRO: Proline

RA: Rheumatoid arthritis

R_g: Radius of gyration

RBD: Rigid-body docking

REMD: Replica exchange molecular dynamics

RMSD: Root mean square deviation

RMSF: Root mean square fluctuations

ROS: Reactive oxygen species

SASA: Solvent accessible surface area

SDMD: Steered discrete molecular dynamics

SER: Serine

SLE: Systemic lupus erythematosus

SMD: Steered molecular dynamics

TF: Tissue factor

TI: Thermodynamic integration

TLR: Toll-like receptor

TNF: Tissue necrosis factor

tPA: Tissue plasminogen activator

TIP3P: 3-pointed water molecule model

T-REMD: Temperature replica exchange molecular dynamics

uPA: Urokinase

VAL: Valine

VMD: Visual molecular dynamics

vdW: van der Waals

WHAM: Weighted histogram analysis method

Acknowledgments

I would like to thank my advisor, Gianluca Interlandi for being a great mentor during my graduate school experience. His wealth of knowledge and patience encouraged me to grow as a scientist. A very special thank you to Wendy Thomas for being an incredible secondary mentor. Current and former Thomas lab members, especially Laura Carlucci, Casey Kiyohara, Amy Stegmann and Molly Mollica deserve recognition for their camaraderie and helpfulness throughout my time in the Interlandi lab. My loving parents, Nels Pederson and Magdalena James-Pederson, always supported my decision to pursue science from a young age. I'm grateful for both of my brothers, Anders and Christian for being great siblings, friends and roommates. Brian Feeley, a close friend from childhood, helped keep me level-headed throughout my time in Washington. He was able to see my thesis defense, which was an absolute treat. Between exploring western Washington with me and finding good food in Seattle, Chris Woods helped keep graduate school interesting and lots of fun. Finally, I would like to thank the Biological Physics, Structure and Design (BPSD) program and the Department of Bioengineering for supporting me from my rotations through the rest of my graduate experience.

Dedication

To my loving and supportive family and friends.

Chapter 1

Introduction.

Normally, fibrin meshes stabilize platelets to form a clot in case of vascular injury. Such an injury triggers local inflammation, which involves the recruitment and activation of neutrophils. Activated neutrophils can lead to fibrin oxidation, which prevents proper clot formation and can lead to thrombosis and bleeding [1]. However, the mechanism by which fibrin oxidation leads to poor health outcomes is not fully understood. In this chapter, I will discuss the complex relationships between fibrinogen, the αC domain, coagulation, inflammation and neutrophil activity from Section 1.1 to Section 1.3.2.3 in order to exemplify the complexity of pathways involving fibrinogen and its oxidation. Current protein-protein docking algorithms are not designed to dock flexible peptides or intrinsically disordered proteins and have failed to identify low-energy dimer candidates of the αC domain. To meet this demand for a flexible-peptide docking algorithm, I explore a method to model the dimer of a flexible protein by using T-REMD with an implicit solvation model and then refining identified minima of the FEL with explicit-solvent simulations. Free energy calculations were performed on the low-energy dimer candidates to to estimate free energy changes of polymerization upon oxidation. The theoretical basis for these simulations is described later in this chapter. Based upon MD simulations presented in this dissertation, I propose the

hydrophobic core (HC) model that Met^{476} mediates αC polymerization by nucleating the HC of the αC polymer. Oxidation of Met^{476} disrupts the HC that forms between αC domains, which prevents lateral aggregation of protofibrils and leads to thrombosis and coagulopathy.

1.1 Fibrinogen structure and function.

Fibrinogen is a soluble, 340 kDa clotting factor (Factor I) that is secreted by hepatocytes and consists of $(A\alpha,B\beta,\gamma)_2$, chains in a symmetric fashion that are linked by disulfides at the central E domain. The E domain, which contains the N-termini of Aα, Bβ and γ chains, is flanked by two coiled-coil regions and globular D domains (Figure 1.1). The C-terminal Bβ and γ regions form the D domains, while each C-terminal Aα chain forms the disordered αC region (Figure 1.1). Upon activation by thrombin into fibrin, fibrinopeptides, FpA and FpB, are released from the E domain, which expose the knobs on the E domain. Holes 'a' and 'b' are located on the D domains and facilitate 'knob-hole' interactions with knob 'A' and 'B' on the E domain. 'A:a' knob-hole interactions are considered the primary driver of protofibril formation. The resulting protofibril is half-staggered with a periodicity of 22.5 nm [2]. Prior to activation, the αC domains occlude the holes and prevent protofibril formation. Upon activation, the αC domains transition from intramolecular to intermolecular contacts to maximize entropy. Activated fibrin rapidly forms protofibrils, which laterally aggregate via the αC domains [2,3]. Fibrin polymers form a mesh of thick fibers that stabilize platelets at the site of injury and prevent bleeding.

The human Aα chain is expressed in two isoforms *in vivo*. Isoform 2 (Fibrin-340) is the most common variant of the Aα chain and has a largely intrinsically disordered C-terminus with the exception of the folded, compact αC domain [4,5]. Isoform 1 (Fibrin-420) is formed by alternative splicing and includes an extended C-terminal globular (α_EC) domain that shares approximately 40% homology to the C-terminal Bβ and γ chains [6]. Although being considered the canonical isoform in the UniProt database [7], isoform 1 accounts for less than 3% of all fibrinogen molecules in the blood. Therefore, "αC domain" in this dissertation will

from now on refer to the Aα N-terminal sub-domain of isoform 2 [8]. Circulating fibrin(ogen) is fairly heterogeneous due to Bβ and γ splice variants as well as proteolytic degradation that removes portions of the αC region [9].

1.1.1 The αC domain.

More than 120 years after the discovery of fibrinogen in 1859, the compact αC domain was identified by differential scanning calorimetry (DSC) [10,11]. Due to the high flexibility of the human αC region, structural analysis by X-ray crystallography or NMR have been limited (Figure 1.1). Hydrogen-deuterium exchange (HDX) experiments on the human Aα chain indicate that the αC region fully exchanges (>90%) with solvent after 10 seconds [12]. Due to these limitations, there has been debate if the αC region is a fully intrinsically disordered region or if compact conformations are sampled at relevant frequency [13]. DSC and NMR experiments indicate a compact, folded domain between residues 425-503 that interacts with copies of itself in a cooperative manner [4,11,14]. The αC domain consists of a true β-hairpin and a pseudohairpin, which adopts a hairpin-like conformation, but lacks the backbone hydrogen bonding patterns for a β-hairpin. This is due to the strand orientation which features $2_{\uparrow}$ $1_{\downarrow}$ $3_{\downarrow}$ $4_{\uparrow}$ strand pattern, where strands 1 and 4 contain the N- and C-terminus of the αC domain, respectively. The parallel strand orientation between strands 1 and 3 is not conducive to hydrogen bonding compared to the antiparallel interactions between strands 2 and 1 [15]. This limits the stability of the monomer species, however, the αC polymer forms cooperatively with a free energy change of -6.7 kcal/mol per additional monomer [4]. Although the human αC domain aggregates at low concentrations, which make structure prediction from NMR data strikingly difficult, a homology model was produced based on the bovine structure and the human sequence (Figure 1.2) [4]. Atomic force microscopy (AFM) has been used to visualize fibrin molecules and fibrin-fibrin interactions. The AFM images detail how fibrin monomers interact with one another to form protofibrils. The protofibrils can be found laterally aggregating together via αC domains. αC domains feature an oblong shape tethered to the triple helical and globular regions by the αC region [16].

1.1.1.1 Previous attempts to model αC polymers.

Upon the discovery of the human αC domain by DSC [11], many efforts were made in vain to understand the tertiary structure of the αC domain. The solution state NMR structure of the homologous bovine αC domain N-terminal β-hairpin was elucidated [17] before both hairpins were resolved [14, 15]. The high β-hairpin content prompted the hypothesis that the αC polymer could form amyloid-like fibrils. Different models were proposed, which relied heavily on backbone amide interactions [15]. During this study, disulfide bonds were introduced to "lock" the hairpins together, in an effort to see if polymerization was affected by β-hairpin swapping. The disulfide bonds slowed down and decreased the degree of αC polymerization [15]. This suggests that rigid-body motion and flexibility of the hairpins are important for polymerization.

1.2 Haemostasis.

Haemostasis is a series of complex processes to keep blood cells inside damaged vessels in the event of a vascular injury. These processes can be separated into three distinct, yet overlapping phases: primary, secondary and tertiary haemostasis. Primary haemostasis is characterized by vasoconstriction, which helps stop blood flow, and platelet plug formation. During primary haemostasis, bivalent fibrin binds to $\alpha_{\mathrm{IIb}}\beta_3$ integrin, the major platelet surface receptor and aggregates platelets at the site of injury [2]. Secondary haemostasis is characterized by activation of the coagulation cascade, including the activation of fibrinogen into fibrin by thrombin, and the deposition of insoluble fibrin fibers around platelets. Finally, during tertiary haemostasis, proteases degrade fibrin clots in a process called fibrinolysis [2, 18]. There is a delicate balance between fibrin activation and fibrinolysis. If too much fibrin is activated, thrombosis may occur, while an excess of fibrinolysis may lead to depletion of fibrinogen in the blood, causing bleeding [2].

1.2.1 Coagulation pathways.

Under the cascade model, coagulation can be separated into three pathways: the extrinsic, the intrinsic and the common pathways (Figure 1.3). Activation of the cascade triggers enzymes that activate proenzymes in order to amplify the signal to form a clot. There are complementary inhibitory mechanisms to prevent thrombosis. The extrinsic pathway is triggered by vascular injury, which exposes Tissue Factor (TF). TF binds to FVII, producing FVIIa, which initiates the extrinsic pathway (Figure 1.3) [19]. The intrinsic pathway is activated when FXII binds to a negatively charged surface, such as collagen, polyphosphates, high molecular weight kininogen (HMWK), kallikrein, or lipopolysaccharide (LPS) from bacterial cell walls and is further upregulated by thrombin activity [20–22]. FXIIa activates FXI, which activates FIX. Thrombin activates FVIII, which forms the tenase complex (enzyme complex that activates FX) with FIXa in the presence of phospholipids (PL) and Ca^{2+} ions. Both of these pathways result in tenase complexes, which are responsible for activating FX. FXa then binds to thrombin-activated FVa in the presence of PL and Ca^{2+} ions to form the prothrombinase (PTase) complex, which activates thrombin [18, 23]. Thrombin then cleaves N-terminal fibrinopeptides from fibrinogen to release the αC domains from the central E domain, producing fibrin. This step allows for knob-hole interactions to form protofibrils, followed by the lateral aggregation of protofibrils into fibers mediated by αC domain homodimers, Since the development of the cascade model, a cell-based approach has been proposed, which accounts for the roles immune and endothelial cells play in coagulation, which better approximates *in vivo* conditions [18, 24].

The cell-based model of coagulation describes the interplay between active, inactive platelets, and TF-bearing cells, such as subendothelial and fibroblast cells, which are typically shielded from blood flow. The cell-based model occurs with four overlapping phases: initiation, amplification, propagation and stabilization. Initiation is characterized by the formation of tenase and PTase activity on TF-bearing cells. Low levels of FIXa, FXa and thrombin are produced on TF-bearing cells. When these protein levels rise above a threshold,

coagulation is initiated. Next, during amplification, platelets are activated by adhering to the site of injury and the presence of thrombin. Thrombin then activates FV, FVIII (which is a cofactor of PTase), and FXI before the platelet surface is remodeled by the localization of clotting factors. Then, during propagation, activated platelets form tenase and PTase complexes by co-localization, which results in a much higher production of thrombin than during initiation, which is enough to activate sufficiently large amounts of fibrinogen [24,25]. During stabilization, thrombin activates FXIII, which is a transglutaminase capable of cross-linking Gln and Lys residues along the αC region. This cross-linking has been shown to improve mechanical and chemical stability of the clot [2,25]. If lateral aggregation is inhibited, FXIIIa is unable to cross-link the αC regions, which impairs secondary and tertiary haemostasis [2].

1.2.2 Anticoagulation pathways.

Endogenous anticoagulants target different parts of the coagulation pathways (Figure 1.3). Antithrombin (AT) is a serine protease inhibitor, which targets thrombin, FIXa, FXa, FXIa, and FXIIa and has more enzymatic activity in the presence of heparin localized on endothelial cells. Thrombin is mostly inhibited by AT, although α2 macroglobin and α1-antitrypsin target thrombin as well. Tissue factor pathway inhibitor (TFPI) inhibits formation of FXa in the extrinsic pathway with Protein S as a cofactor. Protein C is activated by thrombin and inhibits FVa and FVIIIa, which are required for tenase and PTase activity in the intrinsic and common pathways, respectively [25]. The anticoagulation pathway shown in Figure 1.3 ensures that activation of the coagulation pathway occurs only above a certain threshold.

1.2.3 Fibrinolytic pathway.

During tertiary haemostasis, plasminogen can be cleaved by either tissue plasminogen activator (tPA) or urokinase (uPA) into plasmin, which proteolytically cleaves fibrin (Figure 1.4) [26]. tPA is secreted by endothelial cells and is the primary mode of fibrin degradation. uPA is localized near the cell surface, which act on cell-bound plasminogen [27]. Plasmin recognizes cleavage sites all along the Aα, Bβ and γ chains due to their sequence

homology. The C-terminus of the Aα chain is the most susceptible to cleavage. An early plasmin product, Fragment X, is characterized by cleavage of the αC regions. Early studies of the αC region involves treating fibrin with plasmin and isolating the αC fragments [28]. Later fibrin degradation products lack C-termini of multiple chains [28]. Plasminogen activation inhibitors (PAI-1/PAI-2) and α2-antiplasmin (α2-AP) are capable of slowing fibrinolysis by degrading the plasminogen activators (tPA and uPA) and plasmin, respectively [26]. PAI-1 is more ubiquitous than PAI-2, which is expressed during pregnancy and pathological conditions. During the stabilization phase, thrombin-activateable fibrinolysis inhibitor (TAFI) plays an important role in downregulation of fibrinolysis by cleaving C-terminal Lys residues of fibrin, which decreases the number of available tPA and uPA binding sites and therefore limiting the production of plasmin [29].

1.3 Inflammation.

1.3.1 Role of neutrophils in inflammation.

Each day, about 10^{11} neutrophils are produced in the bone marrow via granulopoiesis, destroyed via apoptosis or necrosis throughout the body and cleared by macrophages via efferocytosis [30, 31]. Neutrophils can become activated upon exposure to stimuli such as damage-associated molecular patterns (DAMPs) or pathogen-associated molecular patterns (PAMPs). Activated neutrophils extravasate into the tissue, where they mediate acute and chronic inflammation by clearing pathogens and recruiting immune cells [32]. Activated neutrophils employ different antimicrobial tactics, including recruitment of dendritic cells, macrophages, natural killer and T cells via chemokine production; degranulation; reactive oxygen species (ROS) production; autophagy; neutrophil extracellular traps (NET) by NETosis; protease activity (protease-3 and cathepsin G); and microparticle release [30, 33]. Inflammatory neutrophils multi-task and may perform many of these processes at once, while continually recruiting other neutrophils [30, 32]. NETosis is characterized by the breakdown of nuclear structures, the mixing of nuclear and cytoplasmic material and its extrusion into

the extracellular space to physically encapsulate pathogens and kill them via oxidative enzymes such as NADPH oxidase and MPO by autophagy [33]. Some pathogens have evolved to express DNases to defend against NETs [33]. NETosis typically involves neutrophil death, however certain cytokines and LPS have been shown to promote neutrophil survival, which may promote chronic inflammation [31]. MPO is highly expressed in the neutrophil, where it accounts for approximately 5% of the neutrophil mass and is required for NET formation [34,35]. Neutrophils can release MPO in several mechanisms: neutrophil degranulation, apoptosis or necrosis and by NETosis. MPO deficiency has an incidence of 1 in 2,000-4,000 in European populations and 1 in 55,000 in Japan, [35] which confers a heightened sensitivity to certain bacterial and fungal infections but may also confer some resistance to cardiovascular disease and chronic inflammation [31, 34, 36]. While neutrophils have potent inflammatory effects, neutrophils may also downregulate inflammation by converting to resolving neutrophils. The resolution response is largely mediated by these resolving neutrophils. Neutrophil microparticles may carry pro-inflammatory cargo in inflammatory neutrophils, and pro-resolving cargo in resolving neutrophils [30, 32].

1.3.1.1 Acute inflammation.

Upon vascular injury, inflammatory neutrophils are the first immune cells recruited to the site of stimulus [33]. NETs are structures containing depolymerized chromatin, oxidative enzymes such as nicotinamide adenine dinucleotide phosphate (NADPH) oxidase and MPO to produce ROS. NADPH oxidase produces H_2O_2, and MPO utilizes H_2O_2 to catalyze the following reaction:

$$H_2O_2 + X^- \xrightarrow{MPO} H_2O + HOX$$

X^- is a halide ion, which is most often Cl^- under physiological conditions, and HOX would then correspond to HOCl. HOCl is a strong oxidizing agent, which converts exposed Met residues into Met(O) with a rate constant on the order of 10^7 $M^{-1} \cdot s^{-1}$, as a way of neutralizing pathogen machinery. The *Escherichia coli* proteins are highly susceptible to oxidation via

HOCl treatment; up to 50% of all *E. coli* Met residues were oxidized following MPO treatment [37]. Interestingly, neutrophil cytoplasm contain high concentrations of methionine sulfoxide reductase B (MsrB) [38] and may help attenuate oxidative stress inside the neutrophil. MsrB is localized inside the neutrophil (including during degranulation) and the rate constant for reduction of free methionine sulfoxide is approximately 37 $M^{-1} \cdot s^{-1}$ [39]. This suggests extracellular proteins containing methionine sulfoxide are metabolized into fragments faster than the reduction reaction.

Inflammation resolution is important for haemostasis and tissue homeostasis [33,40]. The resolution phase is mediated, in part, by ω-3 fatty acids docosahexaenoic (DHA) and eicosapentaenoic acid (EPA). DHA and EPA are converted into D series resolvins, protectins and E series resolvins, respectively, which compete with pro-inflammatory lipids, such as arachidonic acid [32]. D and E series resolvins halt neutrophil migration and E resolvins can induce efferocytosis (clearing of dead/dying neutrophils by macrophages) [41]. Inflammatory neutrophils may transition into resolving neutrophils upon exposure to prostaglandin E2, where they downregulate inflammatory pathways. Resolving neutrophils cease pro-inflammatory cytokine production, produce pro-resolving lipid mediators and secrete pro-resolving microparticles, which express annexin A1 and phosphatidylserine (PS) on the surface [32]. Annexin A1 triggers neutrophil apoptosis, and triggers efferocytosis. Surface PS is a well-known 'eat-me' signal, which further facilitates efferocytosis. Inflammatory neutrophils rely on efferocytosis in order to transition into resolving neutrophils before destruction by apoptosis and clearance by phagocytosis. Failure to clear the dead neutrophils and/or failure to clear the stimulus or pathogen may induce chronic inflammation [30]

1.3.1.2 Chronic inflammation.

If a stimulus, such as LPS, is not cleared and the host is unable to initiate the resolution response, the host develops chronic inflammation. Autoimmune disorders are a unique form of chronic inflammation, where the adaptive immune system targets an autoantigen and is perpetuated by the inability to clear the stimulus. For patients with autoimmune disor-

ders, such as psoriasis, rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), activated neutrophils permeate tissues such as skin, synovial tissue and fluid and kidney glomeruli, respectively and perpetuate the inflammatory response [31]. One notable feature of chronic inflammation is the damage of tissue and subsequent fibrosis: the build up of collagen and connective tissue at the affected site [40]. Chronic inflammatory diseases are often associated with cardiovascular disorders due to the sensitivity of haemostasis to inflammation. For example, RA patients have a 50% higher cardiovascular risk and mortality than the general population. Autopsies of patients with SLE showed a presence of coronary atherosclerosis at 40%, compared to 2% of the control group [42]. Chronic inflammation is complex and targets many tissues, imposing long-term damage.

1.3.2 Fibrin and inflammation.

Inflammation and coagulation are intrinsically linked processes. Inflammation triggers the expression of TF in endothelial cells and leukocytes, which initiates the coagulation cascade, in part, by activating thrombin [43]. Fibrinogen is also a critical part of the acute phase response (APR).

1.3.2.1 Acute phase response.

The APR is characterized by a decrease in normal blood proteins called "negative" acute phase proteins (APP), and an increase in "positive" APPs, primarily from altered hepatic expression, including C-reactive protein, serum amyloid A, haptoglobin and cytokines. These cytokines either affect cell growth (positively or negatively), pro-inflammatory cytokines and anti-inflammatory cytokines . The pro-inflammatory group (TNF-$\alpha/\beta, IL-1\alpha/\beta$, IL-6, IL-8 and IFN-$\alpha$) induce fever via inflammation and promote induction of other cytokines via positive feedback regulation [44, 45]. IL-6 is the main driver of the APR and since all three fibrinogen promoter regions are IL-6-responsive, fibrinogen synthesis is a major consequence of IL-6 signalling [44–46]. Upon exposure to PAMPs or DAMPs, IL-6 is synthesized at the site of injury and travels through the bloodstream, where it triggers platelet production

in the bone marrow, fibrinogen and C-reactive protein production in the liver as well as pro-inflammatory T-cell modulation [9, 45, 46].

1.3.2.2 Regulation of cell aggregation and inflammation via fibrinogen-integrin interactions.

Fibrin mediates inflammation and immune cell aggregation by recognizing integrins with β_2 and β_3 subunits: $\alpha_{\mathrm{IIb}}\beta_3$, $\alpha_{\mathrm{V}}\beta_3$, $\alpha_{\mathrm{M}}\beta_2$, and $\alpha_{\mathrm{x}}\beta_2$ (Table 1.2) [43]. $\alpha_{\mathrm{IIb}}\beta_3$ is platelet-specific and is recognized by RGD motifs on the Aα chain and a dodecapeptide in the γ chain via the β_3 domain. Platelet aggregation is partially mediated by $\alpha_{\mathrm{IIb}}\beta_3$-fibrin interactions [47]. $\alpha_{\mathrm{V}}\beta_3$ is known as the vitronectin receptor, which recognizes the same motifs as $\alpha_{\mathrm{IIb}}\beta_3$ and helps recruit macrophages [43, 47–49]. Chronic activation of $\alpha_{\mathrm{V}}\beta_3$ is associated with atherosclerosis, a chronic inflammatory disease associated with thrombosis and myocardial infarction [50]. Chronic inflammation is, in part, mediated by $\alpha_{\mathrm{V}}\beta_3$ activation, which prevent macrophages from performing efferocytosis and inhibits resolution [50]. $\alpha_{\mathrm{M}}\beta_2$ (also known as CD11b/CD18) is expressed on neutrophils and monocytes, which strongly enhances inflammation by leukocytes [43] and triggers fibrinolysis via uPA secretion by neutrophils [29]. $\alpha_{\mathrm{x}}\beta_2$ (also known as CD11c/CD18) is expressed on monocytes, macrophages and dendritic cells. $\alpha_{\mathrm{x}}\beta_2$ recognizes the GPRP motif on fibrin and competes with $\alpha_{\mathrm{M}}\beta_2$ for binding [48, 49]. Fibrin activation of leukocyte integrins ($\alpha_{\mathrm{V}}\beta_3$,$\alpha_{\mathrm{M}}\beta_2$ and $\alpha_{\mathrm{x}}\beta_2$) triggers TF production and upregulates NF-κB signaling, which prevents initiation of resolution and impairs tertiary haemostasis [51]. NF-κB is a transcription factor that activates expression of pro-inflammatory genes including integrins and is activated by TNF-α or LPS via the PI3K/AKT pathway as well as IL-6 [46, 49]. Fibrin degradation products, such as FpA, FpB, D fragments, FnE, and Bβ15-42 have been shown to possess pro-inflammatory properties and are markers for organ dysfunction, although no dose-dependent relationship exists between disease severity and fibrin fragment levels. FpA and FpB serve as chemoattractants for neutrophils, macrophages and monocytes, establishing an early connection between fibrin activation and inflammation [43]. Fibrinogen can initiate and mediate inflammation via a variety of mechanisms [43].

1.3.2.3 Met^{476} and innate immunity.

MPO-produced HOCl is capable of oxidizing three fibrinogen methionine residues: αMet^{476} (will be referred to as Met^{476}), βMet^{367} and γMet^{78}. Oxidized βMet^{367} and γMet^{78} residues are not expected to drastically affect fibrin function. γMet^{78} is found in the triple helical region of fibrinogen and is not expected to disrupt binding interactions. βMet^{367} lies in hole 'b' in the D domain, which forms intermolecular interactions with knob 'B' in the E domain. Based on MD simulations, βMet^{367} oxidation is not expected to disrupt the binding site [39]. The exact role of the 'B:b' interactions in polymerization are cryptic as the BβD432A mutation does not impact protofibril formation or lateral aggregation. Although 'B:b' interactions can form when the 'A:a' interactions are impaired due to hereditary dysfibrinogenemia [3]. In contrast, Met^{476} lies in the αC domain, which mediates lateral aggregation between fibrin protofibrils into thick fibers. Disruption of αC domain interactions by HOCl treatment has been shown by mass spectrometry to produce abnormally dense blood clots [1]. Secondary haemostasis is impaired upon Met^{476} oxidation due to the failure to form fibers and failure to cross-link αC regions via FXIIIa. Tertiary homeostasis is also impaired due to the impermeability of the resulting fibrin clot which prevents proteolytic degradation. Swapping the human αC domain with the analogous chicken αC domain prevents protofibril lateral aggregation and have very similar physical properties as HOCl-treated fibin clots [52].

Fibrin clots are capable of trapping pathogens and inhibiting infection, as part of the innate immune system. Thrombin-activated plasma is capable of killing Group A *Streptococci* (GAS), Group B *Streptococci* (GBS), and *Staphylococcus aureus*, which is, in part, mediated by fibrin-binding proteins on the pathogen surface [53]. However, certain pathogens, such as *Yersinia pestis*, express fibrinolytic enzymes that allow the pathogen to evade the immune system [54]. As discussed in Section 1.3.1.1, MPO activity is highly associated with antimicrobial activity due to the inactivation of pathogen proteins containing exposed methionines [37]. However, HOCl produced by MPO has off-target effects and can affect nearly any protein with an exposed methionine. Met^{476} oxidation inhibits fibrinolysis by altering

clot morphology and has demonstrated antimicrobial activity against GAS [1, 54, 55]. To determine if this trait is common among other closely-related mammals, a BLAST search using the αC domain (residues 441-496) was performed [56]. The top non-redundant 16 results, as well as *Rattus norvegicus* and *Bos taurus* were aligned using the ClustalOmega algorithm [57] (Figure 1.5). Like the human sequence, gorillas, rats and two new world monkey species have a Met at this position. Most mammals listed, however, have an Ile residue at this position, which has a similar side chain SASA and hydrophobicity as Met, but is insensitive to oxidation (Figure 1.5) [58]. For instance, an Ile to Met mutation is tolerated in egg lysozyme due to their similar chemical properties [59]. This suggests the inflammation-associated antimicrobial properties of the αC domain, as part of the innate immune system, may be unique to some species, including humans. While this antimicrobial function provides some evolutionary advantage, oxidized fibrin clots have an increased the risk of thrombosis and may lead to bleeding by fibrin depletion [1, 55, 60].

1.4 Molecular dynamics simulations.

MD is a computational tool that uses Newton's three laws of motion to numerically calculate atomic trajectories given an initial state, a force field (FF), periodic boundary conditions (PBCs) and user-defined simulation parameters. With the development of high-performance computing (HPC) clusters and parallel computing algorithms, MD simulations have been able to address a growing number of biological, chemical and physical problems. For protein systems, MD simulations can be used, among other applications, to study protein-ligand and protein-protein interactions, minimize low-resolution electron microscope structures or homology models, and calculate FELs.

1.4.1 General principles and system configuration.

In its simplest form, MD simulations take a series of atomic positions, and produces a time-series of updated atomic positions. One major component of MD simulation configuration

is the FF, which describes the interactions between atoms in space. The CHARMM36 FF uses a modified Lennard-Jones potential to approximate van der Waals (vdW) forces along with the Coulomb potential to approximate electrostatic interactions . The parameter and topology files complement one another to encode the CHARMM36 FF for use by MD simulation software such as CHARMM or NAMD. The parameter file lists the numerical constants necessary to evaluate the forces and energies simulated by MD. The topology file describes the connectivity between atoms in order to generate the necessary structure files for MD [61,62]. Most parameter and topology files describe the canonical amino acids and water; new files are developed for post-translationally modified residues such as phosphoserine, methionine sulfoxide or N-linked-glycosylated-asparagine [63].

PBCs are often used to avoid finite size effects and allow for the implementation of particle mesh Ewald (PME), which is a very fast method to calculate pairwise interactions with a complexity of $\mathcal{O}(N \log N)$ where N is the number of atoms in the system, compared to $\mathcal{O}(N^2)$ for the conventional approach. PME calculates potentials based on the Fourier transform of the system within the PBC, which can be implemented using a fast Fourier transform (FFT) algorithm [64]. In order to improve performance time, the vdW interactions are switched on and off between a relevant window, typically between 6 (switch on) and 12 (non-bonded cutoff) Å. Electrostatic interactions are computed classically within this range and PME is used to calculate long-range electrostatics beyond the non-bonded cutoff. The list of non-bonded atoms is assembled using a pair-list cutoff that is 1–2 Å above the non-bonded cutoff. This ensures that atoms that move outside of the non-bonded cutoff between timesteps don't cause an unphysical jump in vdW or electrostatic potentials [65].

Hydrogen covalent bonds fluctuate over femtosecond timescales. As a result, a much shorter time step, 0.5–1 fs, is required in order to maintain accuracy. However, an algorithm such as SHAKE, which restrains hydrogen covalent bonds to a fixed length, reduces the number of DoF and allows for the use of a 2 fs time step (2–4-fold improvement in efficiency) [66]. The longer time step makes longer simulation times more feasible.

NAMD assigns velocities and accelerations to atoms throughout a simulation based on

the temperature and a thermostat is used to hold temperature constant. In this dissertation, classical MD is performed using the Langevin thermostat with a piston to hold pressure constant. The Langevin thermostat is used for simulations described in this dissertation due to the incorporated stochasticity and ease of use within NAMD [67, 68]. .

Static atom configurations need to be slowly heated until the target temperature is reached before production MD simulations can be performed. Once solvated and salt ions added, energy minimization of the hydrogen atoms is performed while the heavy atoms are restrained, followed by another minimization step with no restraints. To equilibrate the system, harmonic restraints are applied to all heavy protein atoms, while simulated in solvent for less than a nanosecond. This allows for the solvent to equilibrate around the solute and expand. Equilibration continues during the first 10 ns of the production run and is typically excluded from analysis.

1.4.2 Implicit solvation.

Explicit solvent simulations typically have large numbers of atoms, which scale on the order of r^3 with the size of the periodic box length, r. Implicit solvation, in contrast, reduces the number of atoms in the system, which can drastically decrease simulation run time. The free energy of solvation, ΔG_{sol}, is comprised of a non-electrostatic and electrostatic terms:

$$\Delta G_{sol} = \Delta G_{non\text{-}pol} + \Delta G_{pol}, \quad (1.1)$$

ΔG_{sol} is defined as the reversible work required to move a solute from a vacuum to solution and the goal of implicit solvation methods is to approximate this value for each solute [69]. $\Delta G_{non\text{-}pol}$ is the amount of energy required to solvate a solute with no partial charges and can be approximated using the the solvent accessible surface area (SASA) of the atom. ΔG_{pol} accounts for the energy associated between solvent-solute electrostatic interactions and is typically approximated using the Poisson-Boltzmann (PB) equation, however this

computationally intensive and is infeasible for calculating MD trajectories. An alternative method, the generalized Born (GB) model, approximates the ΔG_{pol} term with the Born free energy. Using the GB model, each charged atom is assigned a calculated Born radius, α_i, based on a parameterized data set [70, 71]. The Born radius defines the distance from the center of the atom where charge is screened; this can be visualized as the molecular surface [72]. Higher salt concentrations screen charges more effectively, which increases the effective distance between the two atoms by shrinking the effective Born radii. The effective Born free energy, $\Delta G_{pol,i}$, is related to the effective Born radius term by:

$$\Delta G_{pol,i} = -\frac{1}{2}(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w})\frac{q^2_i}{\alpha_i}, \quad (1.2)$$

where ϵ_p is the dielectric constant of the protein, ϵ_w is the dielectric constant of the water, q^2_i is the partial charge of atom i and α_i is the Born radius of atom i upon solvation [70]. To approximate vdW interactions and solvent cavitation — the energy required to create a "solute sized cavity" [73] — SASA is calculated for each atom in the following manner:

$$\Delta G_{surf} = \sum_{i=1}^{N} \gamma A_i, \quad (1.3)$$

where N is the number of atoms in the system, γ is the surface tension term in kcal mol^{-1} Å^{-2}, and A_i is the SASA of the atom in Å^2 [70, 71]. This implementation is known as the Onufriev, Bashford, Case generalized Born with surface area (GBSA) model [70, 71, 73, 74]. For calculating potentials between pairwise Born radii, six effective regimes of distances are used to apply descreening effects within a cutoff (greater than 4× the effective Born radius of the atom). These regimes from furthest distance to closest to sphere$_i$ are as follows: sphere$_j$ beyond cutoff, sphere$_j$ partially within cutoff, a smoothing regime, sphere$_j$ not overlapping, sphere$_j$ overlapping and sphere$_j$ inside sphere$_i$ [71, 75].

GBSA improves sampling due to the lack of friction of water present in explicit solvent simulations. Implicit solvation methods may introduce artefacts into the system, which may impact the hypothesis being tested. For example, GB implicit solvation may overstabilize salt bridges, while surface area based models, like GBSA, overstabilize pair-wise non-polar interactions [69]. Using an implicit solvation model requires thorough testing to minimize the possibility of artefact introduction.

1.4.3 Enhanced sampling techniques.

Classical MD simulations sample a conformational space at a single temperature and may be trapped inside a deep local energy minimum. In this case, the energy well will be well-sampled. Enhanced sampling techniques address the 'local minimum problem' by exploiting statistical mechanics to ask thermodynamic questions about the system. Accelerated molecular dynamics (aMD) uses a "boost potential" when energy drops below a threshold, which signifies sampling deep wells. This boost potential diminishes the depth of energy wells, which allows for the traversing of energy barriers without changing the system temperature. Following an aMD simulation, the energy landscape is reconstructed by taking into account each potential boost in a process called reweighting [76, 77]. Metadynamics is an enhanced sampling technique that uses a collective variable (CV), which is a function of atomic coordinates such as RMSD, distance, or R_g, to bias the simulation during runtime. Metadynamics enhance sampling by disfavoring visiting previously-sampled conformational space by adding Gaussian potentials to the force-field. The algorithm encourages sampling of new low-energy conformations in CV space. Metadynamics simulations that converge produce a flat free energy landscape (FEL), which allows for the reconstruction of the original FEL in CV space. Variations on metadynamics include, but are not limited to, free energy perturbation (FEP), thermodynamic integration (TI), umbrella sampling, weighted histogram analysis method (WHAM), adaptive force bias, and steered MD (SMD) [78]. Temperature replica exchange MD (T-REMD) is another enhanced sampling technique used to sample large regions of conformational space with reasonable efficiency by exploiting parallel com-

puting algorithms. T-REMD ensembles can be analyzed using CVs, however the CVs do not influence the simulation parameters, in contrast to metadynamics. In and , T-REMD and FEP are used, respectively, to probe the effects of oxidation on fibrinogen αC domain dimerization.

1.4.3.1 Replica exchange molecular dynamics.

T-REMD uses parallel simulations, called replicas, running at multiple temperatures in order to simultaneously sample wells thoroughly and traverse energy barriers. Each replica corresponds to the canonical ensemble at their respective temperatures and satisfies the detailed balance, which allows for the calculation of ensemble averages of each replica. T-REMD is different from classical MD in that T-REMD doesn't calculate trajectories. Instead, T-REMD is a tool to sample canonical ensembles, which utilizes "unphysical" replica swapping [79].

Overlap between replica energy distributions indicates the likelihood of swapping between replicas, according to the Metropolis criterion described in Equation 1.4 (Figure 1.6). The probability of accepting a swap, P_{target}, between adjacent replicas, i and j, is given by:

$$P_{target} = \min\{1, e^{(\beta_\mathrm{i}-\beta_\mathrm{j})(U(r_i^N)-U(r_j^N))}\}, \tag{1.4}$$

where $\beta_i = (1/k_B T_i)$ is the reciprocal temperature of the replica, U is the potential energy of the configuration and r_i^N specifies the coordinates of N atoms in system i. There are different schools of thought when it comes to $\mathrm{P_{target}}$ selection. If $\mathrm{P_{target}}$ is too low, the efficiency of the algorithm drops because it would take longer for each replica to converge. The upper limit for $\mathrm{P_{target}}$ depends heavily on the system in question, although it seems exchanging more frequently is beneficial [80]. The minimum temperature should represent the relevant context of the protein, such as room temperature (300 K) or physiological temperature (310 K), for example. The maximum temperature needs to be sufficient to traverse energy barriers. The choice of $\mathrm{P_{target}}$ can be limited by computing resources. For example, a high $\mathrm{P_{target}}$ value requires more replicas, which requires more computer nodes and therefore more computational time [80]. Using the online web server available at `http://folding.bmc.uu.`

`se/remd/`, temperatures can be estimated by providing P_{target}, the minimnum and maximum temperatures and the number of DoF in the system (solvent and protein) to estimate the heat capacity of the system. The algorithm provides a list of temperatures and the number of required replicas to achieve the P_{target} [81]. In the NAMD replica exchange package, the provided temperatures are approximated using:

$$T_i = T_{min} \cdot e^{(ln\frac{T_{max}}{T_{min}})(\frac{i}{N_{replicas}-1})}, \tag{1.5}$$

where T_i is the temperature of replica i, T_{min} and T_{max} are the minimum and maximum temperatures, respectively and $N_{replicas}$ is the number of replicas.

Each replica energy distribution depends on the heat capacity of the system, which can change throughout the simulation [81]. In order to preserve the average kinetic energy of the system upon replica swapping, $\frac{3}{2}Nk_BT$, momenta must be reassigned so that the new momenta are scaled appropriately according to:

$$\text{p'}_i = \sqrt{\frac{T_{new}}{T_{old}}}p_i, \tag{1.6}$$

where p_i is the old momenta for replica i, p_i' is the new momenta for replica i and T_{old} and T_{new} are the old and new temperatures, respectively [79]. Certain behavior can affect the energy distributions during the simulation, however. For example, if T_{max} were too high and inconsistently melted one or more of the replicas, the energy distribution would narrow. One of these affected replicas could become trapped in the high-temperature replica with no sufficient energy overlap with its neighbor to perform a swap. In this case, the P_{target} would diminish over simulation time, making the simulation less efficient and risking accuracy of the canonical ensembles. To address this issue, one could choose a more optimal T_{max} or increase P_{target}.

M-dimensional FELs can be constructed *post hoc* by creating an M-dimensional histogram with M CVs for $M+1$ axes. For example, a 2-dimensional histogram is converted

to a 3-dimensional FEL by:

$$G_{\text{bin}} = -k_B T_{replica} \ln(N_{\text{frames}}), \tag{1.7}$$

where G_{bin} is the Gibbs free energy of the bin (z-axis), k_B is the Boltzmann constant, $T_{replica}$ is the temperature of the replica, and N_{frames} is the number of frames in the bin. There are other methods of approximating FELs such as a disconnectivity graph, which utilizes hierarchical clustering to identify the most relevant conformers. This method does not report energies of states [82]. Multidimensional (>2 CVs) FELs may be useful for certain applications, but are difficult to visualize. CVs can be thought of as reaction coordinates in chemistry, in that both are low-dimensional representations of a FEL that contains many degrees of freedom (DoF). A CV should be able to describe large or important changes in structure. Choosing an atomic distance that doesn't exceed 3 Å as a CV may not be as useful as the R_g, which describes the global distribution of atoms in the polymer, unless those 3 Å are significant for the protein dynamics or catalysis. Choosing an optimal CV pair is non-trivial and often requires trial and error.

1.4.3.2 Method to model binding between flexible peptides or proteins

Docking suites such as Rosetta and ZDOCK use rigid-body transformations to optimize an energy function for shape complementarity, among other physical features [83–86]. RBD approaches, including methods with flexible side chains such as RosettaDock, are designed for structures with interacting domains. RosettaDock allows for flexibility of linker regions by introducing perturbations of the ϕ and ψ backbone angles. Docking two peptides with RosettaDock, is inefficient because the method calculates the interaction energy based upon the definitions of a linker and domains [84]. FlexPep is a docking protocol available through Rosetta that docks flexible peptides to receptor proteins, but the receptor must have a pre-identified binding pocket [87]. ClusPro uses a FFT approach, followed by a clustering step and MD refinement in CHARMM to clear clashes [88]. RBD approaches are limited by their ability to sample conformations. There is a need for novel approaches to flexible peptide-

peptide docking. Enhanced sampling methods, as described in Section 1.4.3, overcome this challenge by sampling as many conformations as possible despite a large number of DoF. I propose using T-REMD to describe the conformational landscape of αC-domain dimerization followed by the identification of low-energy dimers using clustering. Weak restraints on the centers of mass may be used to improve sampling of dimer structures.

1.4.3.3 Free energy perturbation.

FEP uses MD trajectories to estimate the $\Delta\Delta$G of structural changes, although in this case we're referring to oxygen, although it could extend to any post-translational modification (PTM). In contrast to T-REMD, FEP uses a CV, λ, to scale the potential energy of an atom to effectively turn it 'on' or 'off' throughout the course of the simulation. This process is called an alchemical transformation and FEP uses these transformations to estimate the free energy changes between states in the thermodynamic cycle shown in Figure A.1 [89,90]. Alchemical transformations are interpreted using λ, which varies from 0 to 1 to switch atoms on or off using:

$$G(\lambda)_{\text{oxidation}} = \lambda G_{\text{ox}} + (1-\lambda)G_{\text{unox}}, \tag{1.8}$$

where the Gibbs energies, G_{unox} and G_{ox} [90]. The structure where λ=0 is known as the reference, and λ=1 refers to the target structure, which will be denoted by 'unox' and 'ox', respectively [91].

During small windows of simulation time (dλ), the energy of the system is calculated such that the total free energy change is estimated by the summation:

$$\Delta G_{oxidation} = G_{ox} - G_{unox} = \sum_{\lambda=0}^{1} -RT \ln \langle e^{-\Delta E/RT} \rangle_{\lambda}, \tag{1.9}$$

where $\Delta E = E_{\lambda + d\lambda} - E_{\lambda}$, R is the real gas constant, and $\langle\rangle_{\lambda}$ is the canonical ensemble average over λ [90]. The length of these windows can vary based upon the size and complexity of the system in question and each window includes an equilibration period prior to data collection. Both the dimer (reference) and monomer, modeled as a tripeptide (target), structures are

used to evaluate how oxidation affects dimerization. The forward and backward ($0 \rightarrow 1$ and $0 \leftarrow 1$) simulations are used to estimate the free energy change using the Bennett's Acceptance Ratio (BAR). The derivation of ΔG^{BAR} is shown in equation 1.10:

$$e^{\beta(\Delta G^{BAR}-C)} = \frac{\langle f(\beta(C-\Delta U))\rangle_{ox}}{\langle f(\beta(\Delta U-C))\rangle_{unox}}, \tag{1.10}$$

where $f(x) = \frac{1}{1+e^x}$, $\langle\rangle$ is the canonical ensemble average over the unoxidized states, and C is an iteratively calculated and self-consistent constant. BAR loses accuracy when applied to large energy gaps and therefore, FEP requires an overlap between sampled states. A dual topology approach is used to maintain overlap, which encodes both the non-interacting reference and target alchemical groups in the simulation (Figure 1.7) [90, 92]. The $\Delta\Delta G_{ox}$ is estimated by comparing the ΔG values from the FEP simulations by:

$$\Delta\Delta G_{ox}^{FEP} = \Delta G_{ox}^{BAR} - \Delta G_{unox}^{BAR} \tag{1.11}$$

The sign of $\Delta\Delta G_{ox}^{FEP}$ indicates if dimerization upon oxidation is unfavorable (+) or favorable (-). In Appendix A, I describe how FEP was performed on representative frames (equilibrated in explicit solvent) to

1.4.4 Nuclear Overhauser effect restraints.

Nuclear Overhauser effect (NOE) restraints are calculated based on dipole-dipole interaction data from NOE spectroscopy (NOESY) 2-D NMR experiments and are often included in NMR solution structure entries in the Protein Database (PDB) [93]. NOE signals decay on the order of the sixth power of the interproton distance. In order to properly account for time averaging, an inverse sixth-power term must be used in the following manner,

$$r_{NOE} = \langle r(t)^6\rangle^{-1/6}, \tag{1.12}$$

where $r(t)$ is the collection of sampled distances from the NOESY experiment and t is time [94]. NOEs can be used to understand the congruence between a structural model and experimental data. To do this, interproton distance ranges collected from NOESY

experiments are converted into *post hoc* passive restraints. To calculate the number of NOE violations based on an MD trajectory, r_{NOE} is calculated for all NOE restraints. r_{NOE} is then compared to the NOE violation distances and distances above the upper bound are considered violated [95]. NOE restraints can be mutually exclusive, ambiguous, or both, since NMR captures ensembles of structures that can differ significantly between states. This heterogeneity limits the number of total NOEs that can be used for analysis [94]. The existence of mutually exclusive NOE restraints suggests that most structures will have an ambient number of NOE violations.

In contrast, restraints can be defined within a simulation, where a force is applied between atoms and across dihedral angles to maintain the interproton distance within the NOE restraint. Active NOE restraints are often used for NMR structure refinement and can be used as CVs for metadynamics. [93,94]. In , I used NOE restraints *post hoc* to probe whether implicit solvation introduced artefacts compared to explicit solvation.

Cytokine	**Cell types**	**Functions**
TNF-α	Macrophages	Phagocyte (neutrophil, monocyte, macrophage, dendritic cell and mast cell) activation [44, 45, 49]
TNF-β	Monocytes T-cells	Triggers phagocytosis, stimulates cytokine production [44, 45]
IL-1	Macrophages, B-cells dendritic cells	Stimulates neutrophil cell proliferation from bone marrow [44, 45]
IL-6	T_H-cells, macrophages, fibroblasts	Fibrinogen synthesis and IgG production [44–46]
IL-8	Macrophages	Neutrophil chemotaxis [45]
IFN-γ	T-cells	Macrophage activation, increases neutrophil, and monocyte function [44, 45]

Table 1.1:: Pro-inflammatory cytokine functions

1.5 Tables.

Integrin	**Cell type**	**Additional functions**
$\alpha_{\mathrm{M}}\beta_2$	Neutrophils, monocytes	Enhances fibrinolysis, TF production [49]
$\alpha_{\mathrm{x}}\beta_2$	Monocytes, macrophages, dendritic cells	Enhances TF production [49]
$\alpha_{\mathrm{IIb}}\beta_3$	Platelets	Platelet aggregation [47]
$\alpha_{\mathrm{V}}\beta_3$	Platelets	Recruits macrophages and inhibits efferocytosis [47]

Table 1.2:: Fibrinogen-binding integrins and their functions

1.6 Figures.

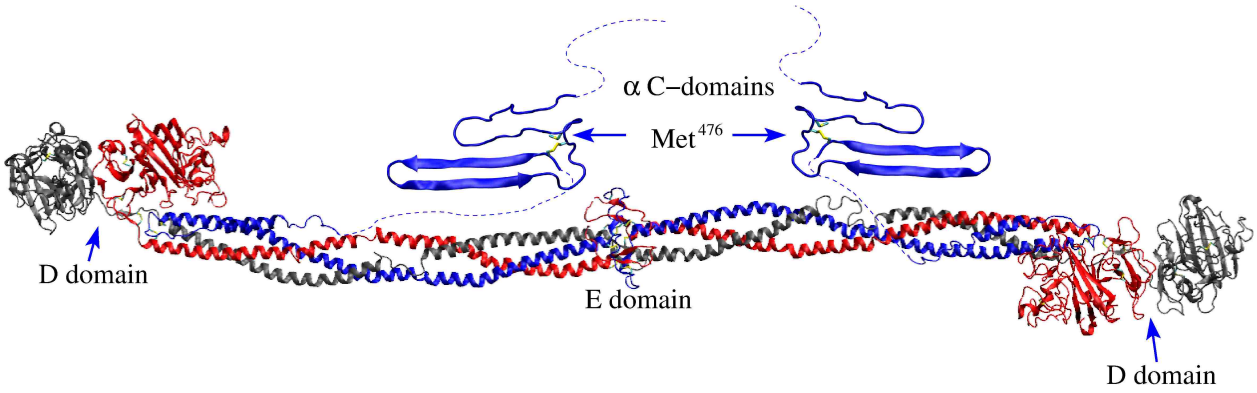

Figure 1.1:: Structural organization of fibrinogen

The X-ray crystal structure of fibrinogen (PDB ID 3GHG) and the human homology model of the αC domain. The Aα chain is shown in blue, the Bβ chain is shown in red and the γ chain is shown in gray. The side chain of Met^{476} in the αC domain is shown in the stick and ball representation and labeled. Dotted lines correspond to unresolved electron density in the αC region. Adapted from [96].

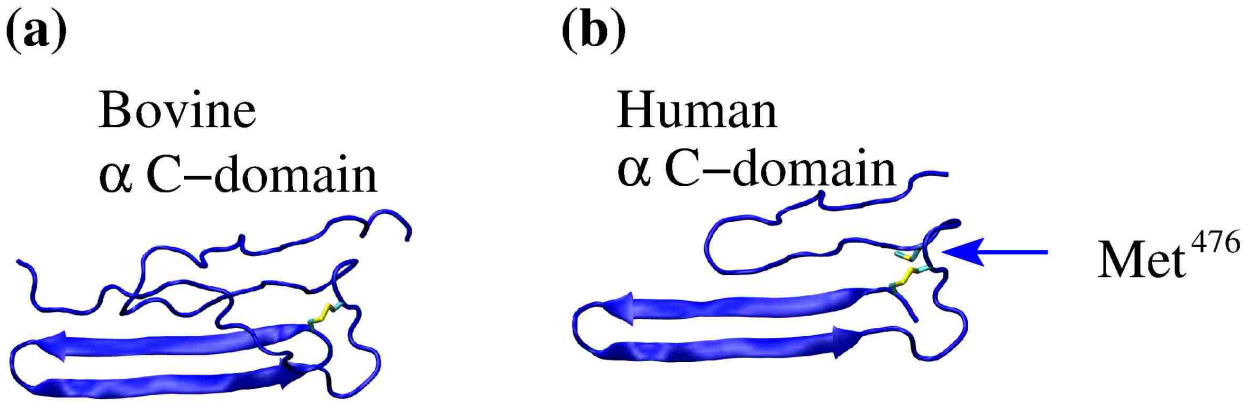

Figure 1.2:: The αC domain structure

(a) The NMR solution structure of the bovine fibrinogen αC domain (PDB ID 2JOR) and **(b)** the human homology model of the αC domain. β-hairpin and pseudohairpin regions are labeled.

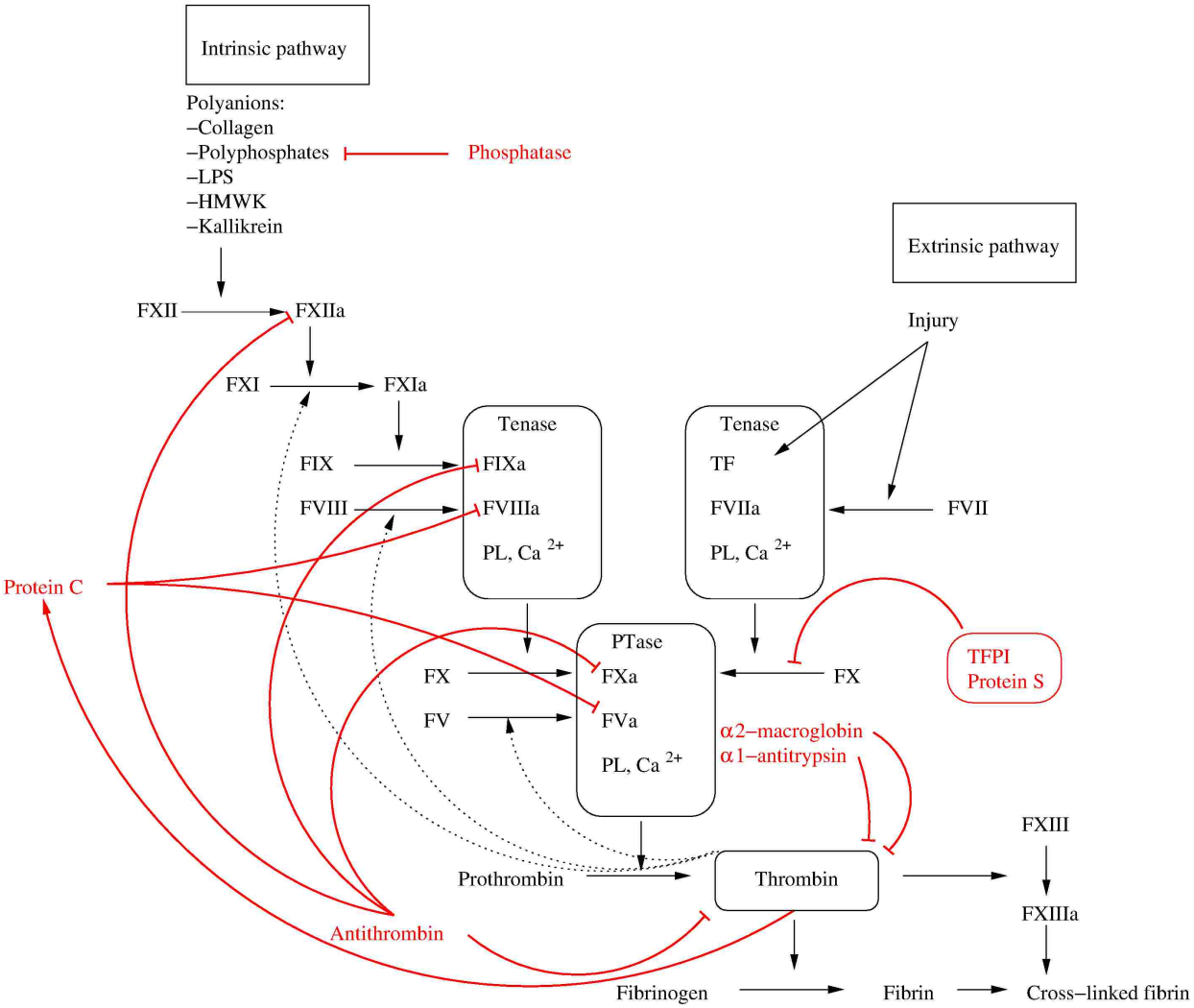

Figure 1.3:: Coagulation cascade and anticoagulation pathway

The intrinsic and extrinsic pathways result in Tenase and PTase formation (black). PTase activates thrombin (which converts fibrinogen into fibrin), the intrinsic pathway via positive feedback loops (dotted) and Protein C. Antithrombin is the main actor in the anticoagulation pathway (red), which restricts the activation of the intrinsic pathway and common pathways. TFPI and Protein S inhibit the extrinsic pathway via FX.

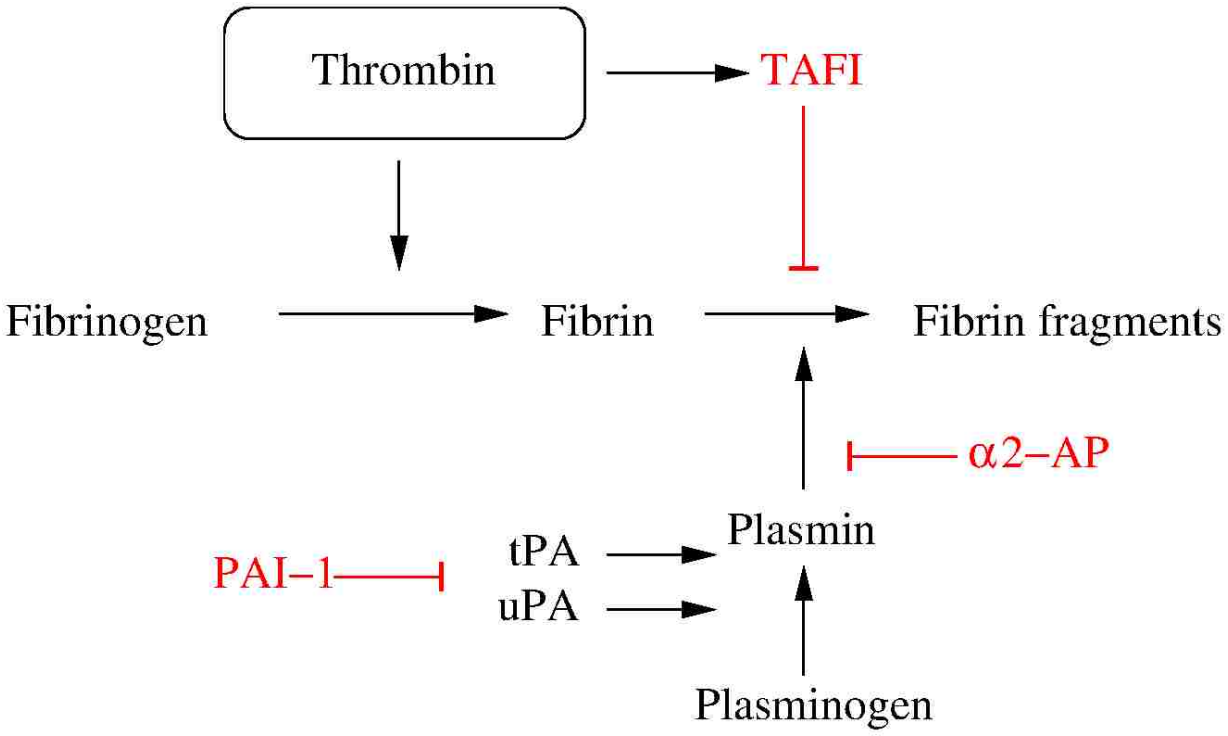

Figure 1.4:: The fibrinolytic pathway

Plasminogen is activated by tissue plasminogen activator (tPA) and urokinase (uPA) into plasmin, which proteolytically cleaves fibrin. Plasminogen activator inhibitor (PAI-1) inhibits tPA and uPA. α2-antiplasmin (α2-AP) is a serine protease inhibitor that targets plasmin. Thrombin-activatable fibrinolysis inhibitor (TAFI) cleaves C-terminal plasmin binding sites on fibrin, inhibiting fibrinolysis.

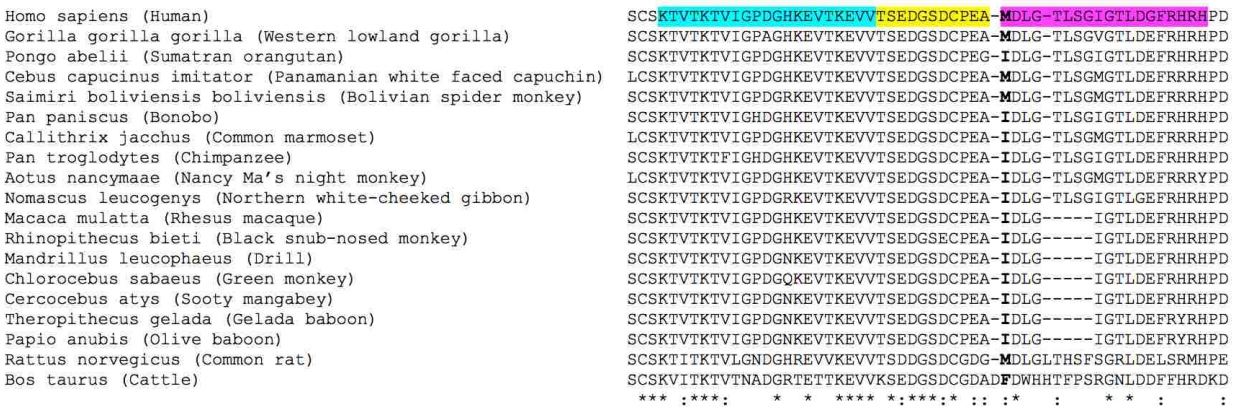

Figure 1.5:: Multiple sequence alignment of putative αC domains

Sequence identity is denoted by "*" and similarity is denoted by ":" ($<$ 3 differences between sequences and must have similar hydrophobicity, charge, polarity, or size) The human β-hairpin is highlighted in cyan, the linker region is highlighted in yellow and the pseudohairpin is highlighted in magenta. Residues aligned with human Met^{476} are in boldface.

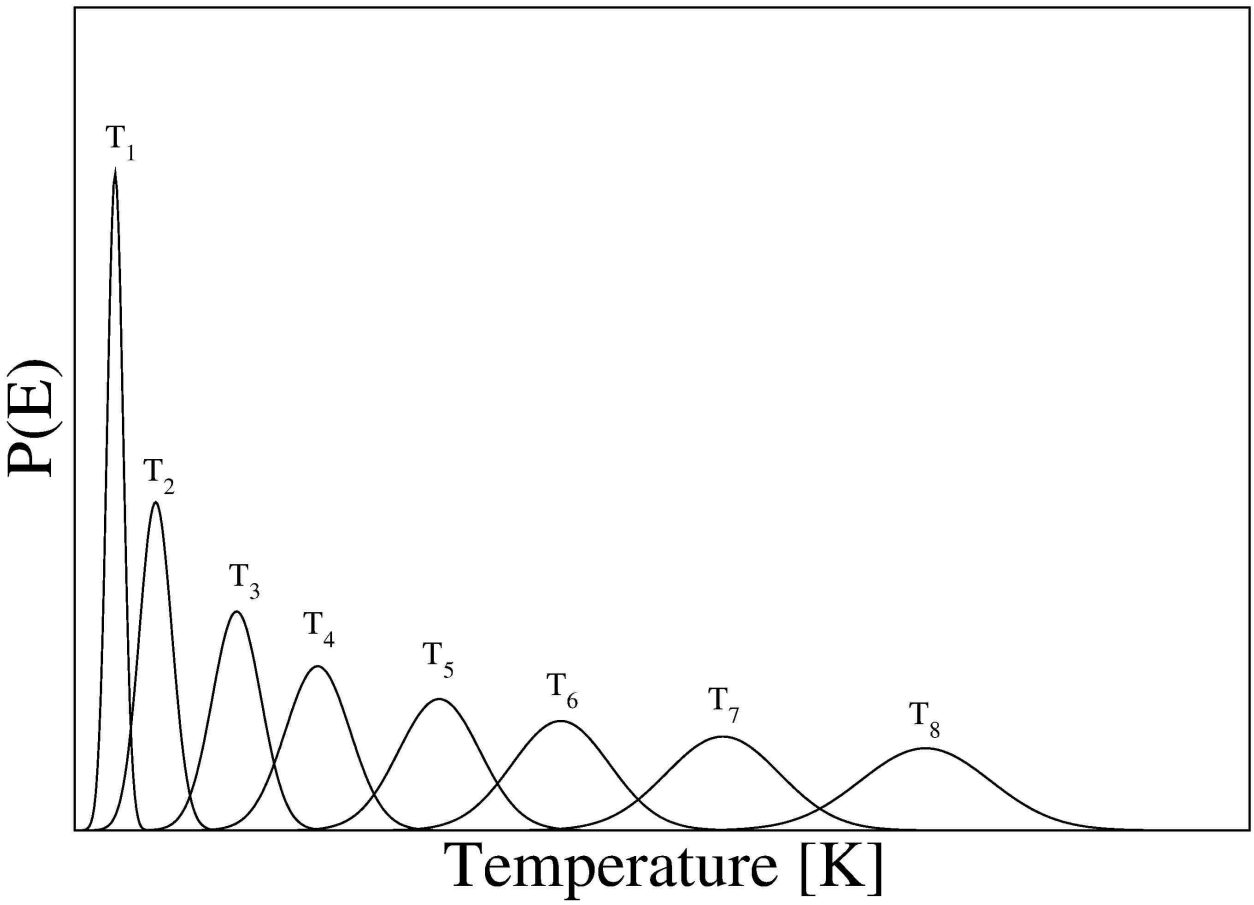

Figure 1.6:: Energy distributions of replicas in T-REMD

Gaussian energy distributions, P(E), at a given mean temperature, T, of each replica. To maintain overlap between replicas and avoid overlapping with non-adjacent replicas, spacing between temperatures increases (according to Equation 1.5) and the standard deviation of each distribution increases. P_{target} is proportional to the energy overlaps between adjacent replicas.

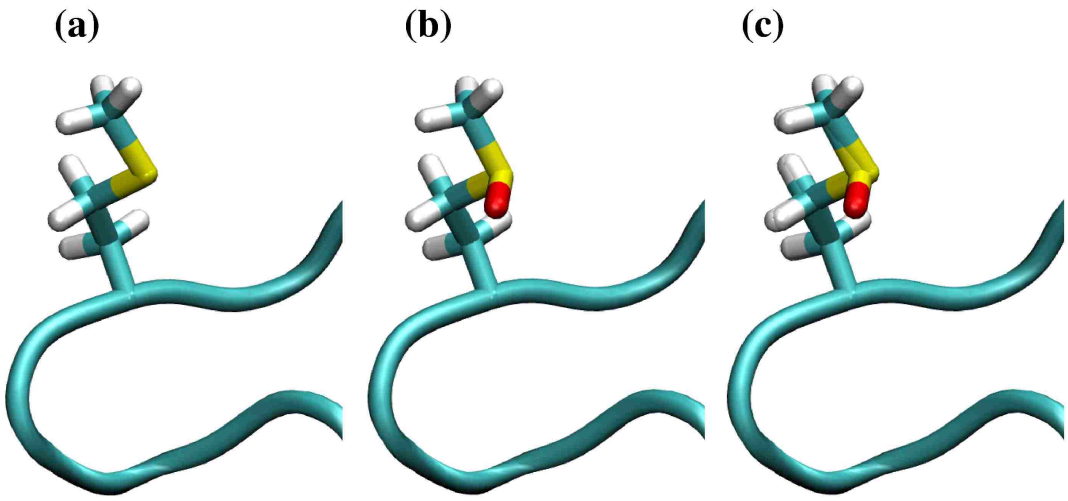

Figure 1.7:: Dual topology for FEP calculations allows for sampling overlap

(a) Met^{476} when $\lambda = 0$ (reference), **(b)** Met(O)^{476} when $\lambda = 1$ (target) and **(c)** Superposition of Met^{476} and Met(O)^{476} side chains, which do not interact with one another under the dual topology approach. All other atoms in the system are treated classically, which allows for the required overlap between sampled states, **(a)** and **(b)**.

Chapter 2

Modeling a flexible protein dimer

2.1 Introduction

Protein-protein interactions (PPIs) are incredibly important for biological functions, so there is a demand for accurate models of protein binding in basic science and industrial settings. Many proteins contain highly flexible regions that make studying interactions involving these regions difficult. Intrinsically disordered proteins (IDPs) account for 11% of proteins catalogued by Swiss-Prot and it is estimated that 35-51% of human proteins contain flexible loop regions [97]. IDP interactions are important for signal transduction, scaffolding, transcription (DNA-protein interactions), cell cycle regulation and chaperoning [98]. IDPs can adopt transient secondary structures during binding and these "fuzzy complexes" can still exhibit disordered behavior even in bound states [98]. Other examples where the binding between flexible proteins needs to be modeled in detail include disease-related misfolding and aggregation. For example, the prion (PrP^{c}) hydrophobic domain undergoes a large unfolding transition before rearranging disulfide bonds, followed by a relaxation to a misfolded multimer state (PrP^{sc}) [99, 100]. Finally, it is often also necessary to model a flexible part of a protein that is missing from the experimentally determined structure. For example, the blood protein von Willebrand factor (VWF) contains a 34-residue long, highly flexible region between the D'D3 domains and the A1 domain that is known to bind to the core

of the A1 domain and to regulate binding to blood platelets [101]. However, there is currently no resolved structure of the A1 domain with the long linker, and yet it would improve understanding how this regulatory region functions.

Protein docking software aim to understand the nature of PPIs. Most docking software, including Rosetta, utilize some form of rigid body docking (RBD). For example, Rosetta-based docking is a powerful technique that works for many proteins but struggles with targets that undergo conformational changes upon binding [102]. PPIs between flexible proteins, which includes IDPs and IDRs, are difficult to study with these approaches due to the large numbers of DoF and structural heterogeneity.

To address the need for a new method to model PPIs, I propose an MD-based approach that can capture the structural heterogeneity of flexible proteins despite the large number of DoF. The method uses an initial guess for the docked configuration, but because it uses enhanced-sampling techniques, there is no dependence on the initial conformation. Here, I describe the use of temperature replica exchange MD (T-REMD), an enhanced-sampling technique, to produce free energy landscapes, which can be represented as 2-D histograms. Next, the deepest minima are probed for the representative frames, which are used for analysis. A similar process, using T-REMD, was performed on the amelogenin monomer, which is an intrinsically disordered protein, although there are some important differences [103]. I will refer to this previous method as the clustering method. The proposed method constructs a FEL using the frames from the T-REMD simulation, whereas the clustering method organizes frames into substates. By constructing a FEL, thermodynamic information is preserved and the global minimum can be identified, which is an advantage over the clustering method. Both methods utilize similar hierarchical clustering strategies to identify representative frames, however the clustering method uses all frame, rather than the subset of structures identified from a local minimum. The described MD-based methods are far slower than RBD approaches, but MD can be used on disordered and flexible proteins, regions or peptides. In this chapter, I detail the individual steps to model a flexible protein dimer (the αC domain).

The αC domain is a largely flexible region of the fibrinogen Aα chain. NMR experiments on the human αC domain have been informative but have failed to resolve a structure, due to the high aggregation propensity [4]. The bovine αC domain, which shares 57% sequence identity to the bovine αC domain, was used to construct the homology model used throughout this dissertation [4, 14]. Two copies of the homology model were arranged in a planar fashion such that the pseudohairpin of one monomer was adjacent to the β-hairpin of the other monomer. An MD simulation was performed on this initial structure with restraints on the CoM between each monomer. A frame was taken from the end of the simulation to initialize the T-REMD and sample dimer structures. The sampling of dimer conformations by T-REMD is sufficient that the initial frame is irrelevant during analysis.

2.2 Methods & Materials

2.2.1 Explicit solvent simulations

Explicit solvent simulations approximate *in vivo/in vitro* conditions by including water and salt atoms in the simulation. Prior to performing such simulations, the system must be equilibrated. Harmonic constraints were applied to the heavy atoms of the protein and performed 0.2 ns of simulation time at 300 K. 300 and 310 K are common temperature options for MD simulations, as they correspond to room and physiological temperature, respectively, although we have chosen to use 300 K for explicit solvent simulations. In simulations with Met(O) residues, harmonic constraints were kept on all heavy atoms, excluding those atoms in the Met(O) residue and sequentially neighboring residues, and equilibrated for 2 ns. After these steps, harmonic constraints were released. For 300-K simulations, the first 10 ns are considered equilibration and are not used for analysis. During equilibration and production runs, the temperature is held constant by using the Langevin thermostat [67] in NAMD with a damping coefficient of 1 ps^{-1}. In explicit solvent simulations, the pressure was held constant at 1 atm by applying a pressure piston to mimic physiological conditions [68]. Explicit solvent simulations were initiated from the NMR structure of bovine αC domain

or a human homology model of the αC domain. Each simulation was started with different initial random velocities to ensure that each trajectory is not identical to other trajectories that used the same initial structure.

2.2.1.1 Nuclear Overhauser effect and restraints

NOE distance restraints of the bovine structure [14] were used to assess how similar a set of simulation conditions are to native conditions. A NOE distance restraint was considered violated *in silico* if the equation $\langle r(t)^{-6} \rangle^{-\frac{1}{6}} \leq {r_{\max}}^{\exp} + 0.5$ Å was not fulfilled, where $r(t)$ is the interproton distance, at simulation time t, ${r_{\max}}^{\exp}$ is the experimentally determined upper distance limit, and $\langle\,\rangle$ represents a time average. Here, a 0.5 Å buffer term is included to take edge cases into account. The buffer term is not necessary, but can be used to include restraints that fall within reasonable instrumental or experimental error, without introducing undue uncertainty.

NOE violations are reported as percentages by the equation $\mathrm{NOE}_{\%}=\frac{N_{\mathrm{violations}}}{N_{\mathrm{r}^{\exp}}} \times 100\%$ where $\mathrm{N}_{\mathrm{violations}}$ is the number of NOE violations and $\mathrm{N}_{\mathrm{r}^{\exp}}$ is the number of experimentally determined distance ranges. A large number of NOE violations may indicate non-native sampling, however, NOE distance restraints can be mutually exclusive and therefore, most MD trajectories will have a background level of NOE violations.

2.2.2 Implicit solvation model selection

There are many options when choosing an implicit solvent for production MD runs. The Generalized Born (GB) model approximates the Poisson-Boltzmann (PB) equations as described in Section 1.4.2 and is very efficient compared to PB-based approaches and explicit solvent [70]. The GB model does not account for hydrophobic interactions, however. To rectify this, other implicit solvent models such as GBSA and FACTS build on the GB model, but these implementations may introduce artefacts, such as introducing secondary structure biases that are not reflective of the true secondary structure. The ideal implicit solvent model would be efficient for use with T-REMD and introduces the fewest number of artefacts.

2.2.2.1 Generalized Born implicit solvent

One set of implicit solvent simulations were run with the GBSA model [70,71]. The generalized Born model estimates the electrostatic contribution of atoms to estimate the solvation energy. SASA is also considered to approximate non-polar contribution of the solvation energy. A user-supplied surface tension scalar term, $\gamma = 0.0054$ kcal/mol/Å^2, is used to adjust the contribution of the SASA term [70,71]. The electrostatic interactions cutoff was set to 12 Å and van der Waals interactions were treated with the use of a switch function starting at 10 Å and turning off at 12 Å. A cutoff of 14 Å was used for the list of non-bonded pairs. Dynamics were integrated with a time step of 2 fs. The covalent bonds involving hydrogens were rigidly constrained by means of the SHAKE algorithm with a tolerance of 10^{-8}. Snapshots were saved every 10 ps for trajectory analysis.

2.2.2.1.1 FACTS implicit solvent

Another set of implicit solvent simulations were run with the FACTS model. FACTS implicit solvation is a parameterized extension of the GB model that takes into account the geometry of the solute and uses the SASA to estimate the non-polar contributions of the solvation free energy [104]. A preset surface tension scalar term, $\gamma = 0.0075$ kcal/mol/Å^2, is used to the contribution of the SASA term [104]. The electrostatic interactions cutoff was set to 12 Å and van der Waals interactions were treated with the use of a switch function starting at 10 Å and turning off at 12 Å. A cutoff of 14 Å was used for the list of non-bonded pairs. Dynamics were integrated with a time step of 2 fs. The covalent bonds involving hydrogens were rigidly constrained by means of the SHAKE algorithm with a tolerance of 10^{-8}. Snapshots were saved every 10 ps for trajectory analysis.

2.2.2.1.2 Comparing GBSA and FACTS implicit solvent with the bovine and human αC domains

Implicit solvents are important for this method because they reduce the number of atoms in the system, which means fewer replicas are required. There are different implicit solvent models that calculate the free energy of solvation, ΔG_{solv}. The two discussed here, GBSA and FACTS, were used in αC domain monomer simulations to compare trajectories.

The bovine αC domain can be used to assess the utilities of GBSA and FACTS since NOE violations calculated from the trajectories can be compared to explicit solvent simulations. The human αC domain simulations are then compared to the bovine simulations to ensure structure stability and minimize artefact introduction.

To ensure artefacts are not introduced, 50 ns simulations were performed in explicit solvent, (Figure 2.1), GBSA implicit solvent (Figure 2.2) and FACTS implicit solvent (Figure 2.3) and the secondary structure elements of each frame were calculated using the STRIDE algorithm in VMD [105, 106]. In explicit solvent, the β-hairpin features β-strands separated by a turn, which is consistent with the initial frame indicating explicit solvent simulations maintain stability of the αC domain. 3_{10}-helices and α-helices are sampled but are transient and located in the pseudohairpin (Figure 2.1). GBSA implicit solvent shows similar secondary structure elements although there is more α-helical structure compared to explicit solvent (Figure 2.1 and Figure 2.2). In contrast, FACTS introduced the π-helix artefact, which is also transient, but π-helices are not found in either the explicit solvent or GBSA simulations.

NOE violations of trajectories of bovine αC domains indicate that FACTS shows similar number of violations to GBSA. The ideal implicit solvent will minimize NOE violations compared to explicit solvent. NOE analysis reveals that GBSA has a clear advantage over FACTS when it comes to the pseudohairpin. For example, both implicit solvents show an increase in NOE violations in the inter-pseudohairpin category, but GBSA shows statistically significantly fewer violations for the total pseudohairpin and intra-pseudohairpin regions ($p < 0.05$) (Figure 2.4). This is noteworthy as the Met^{476} residue lies in the pseudohairpin region, which indicates that the GBSA implicit solvent is an attractive option.

Figure 2.5 shows the number of β-hairpin hydrogen bonds in each solvent condition. Interestingly, none of the implicit solvents recapitulated the hydrogen bond stabilities compared to the NMR structure. However, human GBSA approximates the number of hydrogen bonds present in the human explicit solvent simulations. For example, the first 4 hydrogen bonds listed show decreased stability compared to the NMR structure by both human GBSA

and explicit solvent. For the bovine simulations, the last 4 hydrogen bonds are destabilized by FACTS and GBSA. The agreement between GBSA and explicit solvent suggests that it should be used to preserve native structure.

In order to assess stability of the bovine and human αC domains in different solvents, the mean C_α RMSD of each segment (β-hairpin, linker and pseudohairpin) is shown in Figure 2.6a. The differences between human and bovine αC domains in explicit solvent shows a statistically significant difference by Students T test ($p < 0.05$), but the difference is negligible in terms of distance in Å. There is no significant difference between the human explicit and GBSA simulations, which suggests GBSA implicit solvent may be a reliable option. In addition, the RMSF of each residue is plotted in Figure 2.6b, which describes the local flexibility. There are no statistically significant differences between bovine αC domain in explicit solvent and human αC domain in explicit solvent or GBSA.

I have demonstrated that the bovine and human αC domains in explicit solvent behave similarly to one another. The human GBSA simulations agree with the corresponding explicit solvent in terms of secondary structure, NOE violations, hydrogen bonding, structural deviation (RMSD) and flexibility (RMSF). Furthermore, FACTS is only available on multicore machines with a license, which is prohibitively expensive for our purposes. As a result, GBSA was chosen as the implicit solvent for production runs of the human αC domain.

2.2.3 Enhanced sampling methods

The enhanced sampling simulations, T-REMD, were performed with NAMD using the CHARMM36 force field [62,65]. The force field parameters for methionine sulfoxide were obtained from the SwissSidechain website [63] adapted per analogy for the CHARMM36 force field. Accelerated MD and metadynamics simulations can also be utilized, as long as there is enough data to construct a free energy landscape. In this study, T-REMD was used to traverse barriers in the FELs of αC-domain dimers. T-REMD simulations were performed in GBSA implicit solvent to reduce friction, atom count and the number of required replicas. Without friction, αC domains are free to sample more conformational space, resulting in an improvement in

sampling efficiency.

T-REMD uses temperature to traverse energy barriers by performing simultaneous simulations, replicas, of the same system at a nearby temperature. Each replica is assigned a temperature value, which corresponds to a Gaussian distribution of energy [79]. If two replicas sample the overlap between these distributions, it is theoretically impossible to distinguish which replicas corresponds to which temperatures. During this instance, the replicas swap temperatures according to the Metropolis-Hastings criterion. Multiple replicas allow a simulations to sample other wells that were previously computationally intensive using classical MD. Temperature selection is critical for successful T-REMD simulations. First, the lowest temperature (T_{min}) is chosen, which typically has some sort of significance, such as 300 K (room temperature) or 310 K (physiological temperature). After the simulations are complete, all frames sampled at T_{min} are typically used for analysis. The T_{max} replica should not exceed the melting temperature of the system. If the protein melts in the T_{max} replica, the heat capacity of the system is lowered (a more dispersed energy distribution), which affects the overlap with its neighbor. The exchange probability between other replicas would be so low that the efficiency of the simulations decrease over time. A proper temperature range will allow for the protein to undergo large conformational changes without melting the solute. Next, the number of replicas was chosen using this online tool: `http://folding.bmc.uu.se/remd/`, which estimates the heat capacity of the system by calculating the numbers of DoF in the system. The algorithm calculates a series of temperatures based on the desired temperature range, the exchange attempt frequency (EAF) and the desired exchange probability, P_{acc} [79]. P_{acc} is a function of the EAF, the temperature spacing between replicas and the relaxation rate of potential energy between adjacent replicas [107]. The EAF should be high, but the autocorrelation time of sampling represents the upper limit of P_{acc} [80]. If the EAF is too fast, the atoms do not have time for relaxation, and relevant information is lost [80, 107, 108].

In this study, the exchange attempts occurred every 10 picoseconds with a desired P_{acc} = 0.45. After running a short, 1 nanosecond simulation of the bovine αC domain in GBSA

implicit solvent at 330 K and 360 K. The 330 K and 360 K simulations melt the β-hairpin within 400 ps and 100 ps, respectively Figure 2.7. With an estimated upper limit of 350 K, the online temperature calculator suggested the following temperatures: 300.00 K, 306.35 K, 312.83 K, 319.45 K, 326.19 K, 333.05 K, 340.06 K and 347.19 K. The replica exchange module available in NAMD approximates these temperatures using the formula,

$$T_i = T_{min} \cdot e^{(ln\frac{T_{max}}{T_{min}})(\frac{i}{N_{replicas}-1})}, \tag{2.1}$$

where T_i is the temperature of replica i, $T_{min} = 300$ K and $T_{max} = 347$ K are the minimum and maximum temperatures, respectively and $N_{replicas}$ is the number of replicas. The resulting list of temperatures is as follows: 300.0 K, 306.30 K, 312.74 K, 319.31 K, 326.02 K, 332.87 K, 339.86 K, and 347.00 K. P_{acc} of the unoxidized and oxidized αC-domain dimers were 0.43 $\pm$ 0.04 and 0.40 $\pm$ 0.01, respectively. Restraints on the β-hairpin were applied to improve sampling and restraints on the centers of mass were applied to ensure that only dimer structures were being sampled. It is sometimes desirable to simulate the monomer states as well in order to understand the unbound to bound transition. Here, we are only interested in the dimer conformations and without a restraint, we would needlessly sample monomer states. In explicit solvent simulations of the αC domain monomer, the β-hairpin structure is largely unperturbed Figures 2.1 and 2.5. Restraining the β-hairpin is used to recapitulate this behavior in implicit solvent and is therefore unlikely to introduce bias.

2.2.4 Constructing free energy landscapes

Two collective variables must be chosen to construct a three-dimensional energy landscape. In this study, the radii of gyration were used to represent the FEL of the αC-domain monomers from the enhanced sampling runs. Collective variables should be chosen appropriately to ensure that they can represent large structural changes. The FEL is constructed from a two-dimensional histogram. A text file containing two columns representing the radii of gyration of each frame is converted into a two-dimensional histogram using the "hist-counts2" function in MATLAB. Each bin in the histogram can be converted to an energy

value using a Boltzmann distribution $G_{\text{Bin}} = -RT\ \ln(N_{\text{FramesInBin}})$ where G_{Bin} is the Gibbs free energy of the bin, R is the ideal gas constant in (kcal g^{-1} mol^{-1} K^{-1}), T = 300 K and $N_{\text{FramesInBin}}$ is the number of frames in a particular bin.

2.2.4.1 Binning and minima selection

Bin widths can dramatically affect minima selection if they are too large or too small. If the bin widths are too large, wells are difficult to define and each well will be excessively heterogeneous. If bin widths are too small, however, there are lots of minima that are too similar to one another, introducing noise. Bin size choice will affect the number of minima, and is very important for minima selection. When comparing FELs across different states (such as unoxidized and oxidized states) bin widths should be consistent to ensure that the integral of both surfaces are also consistent. Small deviations may be tolerated to ensure that bins are properly spaced,

Minima are defined by the "imregionalmin" function in MATLAB, which finds bins with a single deep edge that is surrounded by higher values. When deciding on bin size, the minima must not be near one another to avoid redundancy. To ensure that only relevant minima are subject to analysis, minima are selected based on their depth. Using a modified Boltzmann equation, $f_{\text{B}} = e^{(\text{E}_0\text{-E}_\text{i})/\text{RT}}$ where E_0 is the energy from the deepest well, E_i is the energy of the ith well, R is the ideal gas constant, T = 300 K and $0 \leq f_{\text{B}} \leq 1$. f_{B} is the theoretical transition probability between the most populated state and the ith state. A cutoff for f_{B} must be defined *a priori* for minima selection and it must be consistent throughout analysis. In this study, a cutoff of f_{B}=0.2 was used although this value will depend on the individual system and the depths of the wells. This method utilizes statistical mechanics to estimate the number of relevant minima, although there could be other strategies with different systems.

2.2.5 Identifying dimer conformers

In order to identify a conformer from a single bin in the FEL, average linkage hierarchical clustering was performed on all frames in the aforementioned bin, as is common when

depositing NMR ensembles to the PDB [109], which is available as part of the Chimera software package [110]. First, the distance matrix is constructed based on the RMSD values between all frames. Average linkage hierarchical clustering, measures the distance between two clusters as the average distance between all states each cluster. This distance is then used to calculate the average spread of each cluster. Alternatively, hierarchical clustering can be performed with single or maximum linkage distances. This decision is context specific, although the average linkage approach is suitable for our needs. The penalty value, P, is calculated based on the average spread and the number of clusters, N. N is chosen based on the value that minimizes P. Finally, the frames that approximate the centroid of the cluster are considered the representative frame of that minimum [109]. These representative frames are then subject to 50 ns explicit water simulations to assess stability and the introduction of artefacts by implicit solvation. Furthermore, the representative frames are considered dimer conformers.

2.3 Discussion

I present a novel method to identify low-energy protein dimer conformations using enhanced sampling techniques, implicit solvent and clustering. This method is extensible to a considerable degree as it can be used to study monomers, dimers, or higher order structures as well even though the work discussed here pertains to dimers. As previously discussed, many docking algorithms use rigid body transformations to optimize shape complementarity. However, these methods fail to capture the natural flexibility of many protein-protein interactions. In an effort to utilize the flexibility of proteins in dimer conformations, enhanced sampling techniques, such as T-REMD, are used to sample the conformational space of the state in question (monomer, dimer, post-translationally modified residues, etc.). T-REMD is only one of many possible enhanced sampling techniques that could be employed, although T-REMD is fairly straightforward and can be performed without much information about the system. T-REMD can be implemented in very efficient ways to minimize the amount of run

time required. In this study, simulating 1650 atoms for 1 μs took over 30 days of simulation time, including time waiting in the queue. For each new system, implicit solvent methods should be compared to explicit solvent in order to ensure that artefacts are not introduced and the solvent preserves the native structure. Helical proteins, for example, would likely utilize a different implicit solvent model to avoid overstabilization of non-helical secondary structure. Once the simulations are complete, *post hoc* analysis of the trajectories are used to construct 2-D histograms and FELs in an effort to preserve thermodynamic information. The global and local minima are labeled and clustering is used to identify the representative frames of each minimum. This method fulfills a gap in the current literature about modeling dimers of flexible or intrinsically disordered proteins.

2.4 Figures

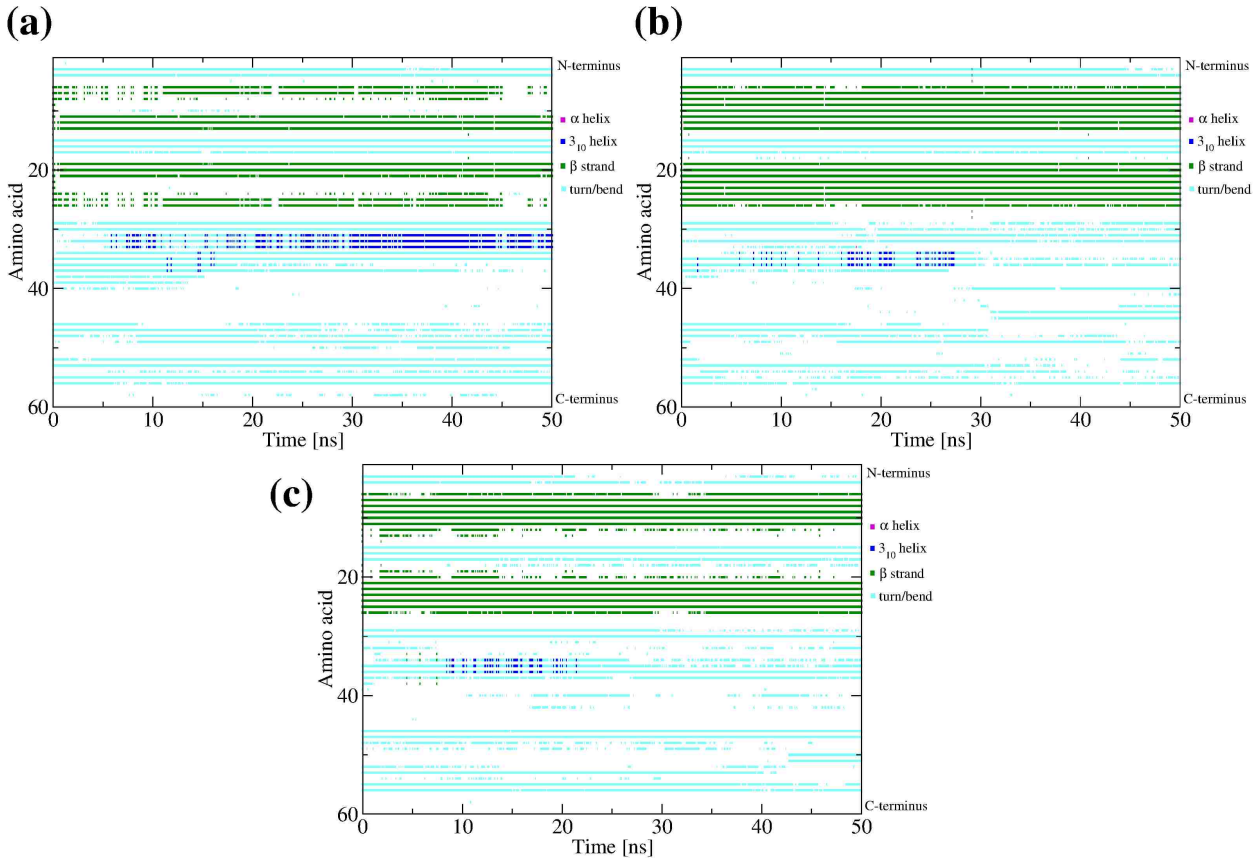

Figure 2.1:: Secondary structure elements of bovine αC domain in explicit solvent.

Time series analysis of secondary structure elements for each residue in the αC domain based on the STRIDE algorithm. Cyan corresponds to turns/bends, green corresponds to β-strands, pink corresponds to α-helices and blue corresponds to 3_{10}-helices.

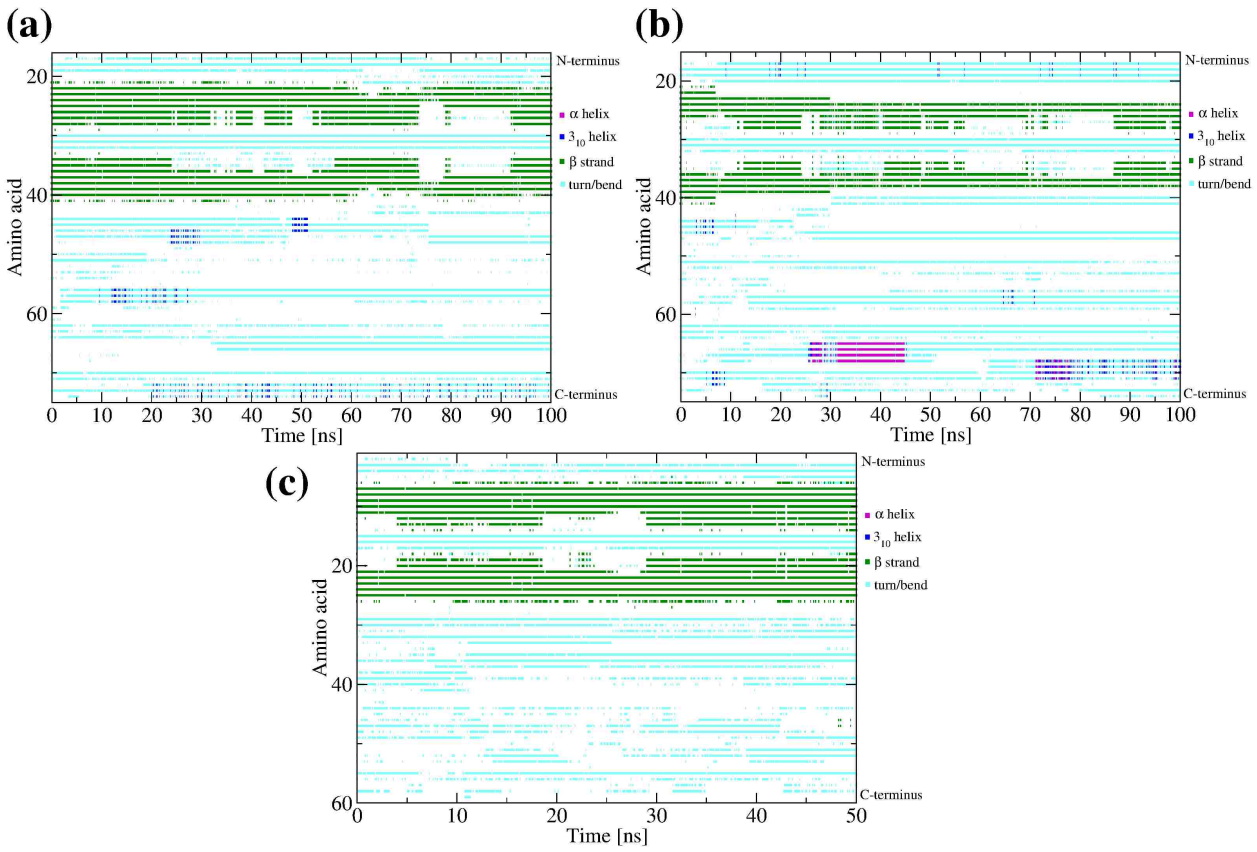

Figure 2.2:: Secondary structure elements of bovine αC domain in GBSA implicit solvent.

Time series analysis of secondary structure elements for each residue in the αC domain based on the STRIDE algorithm. Cyan corresponds to turns/bends, green corresponds to β-strands, pink corresponds to α-helices and blue corresponds to 3_{10}-helices.

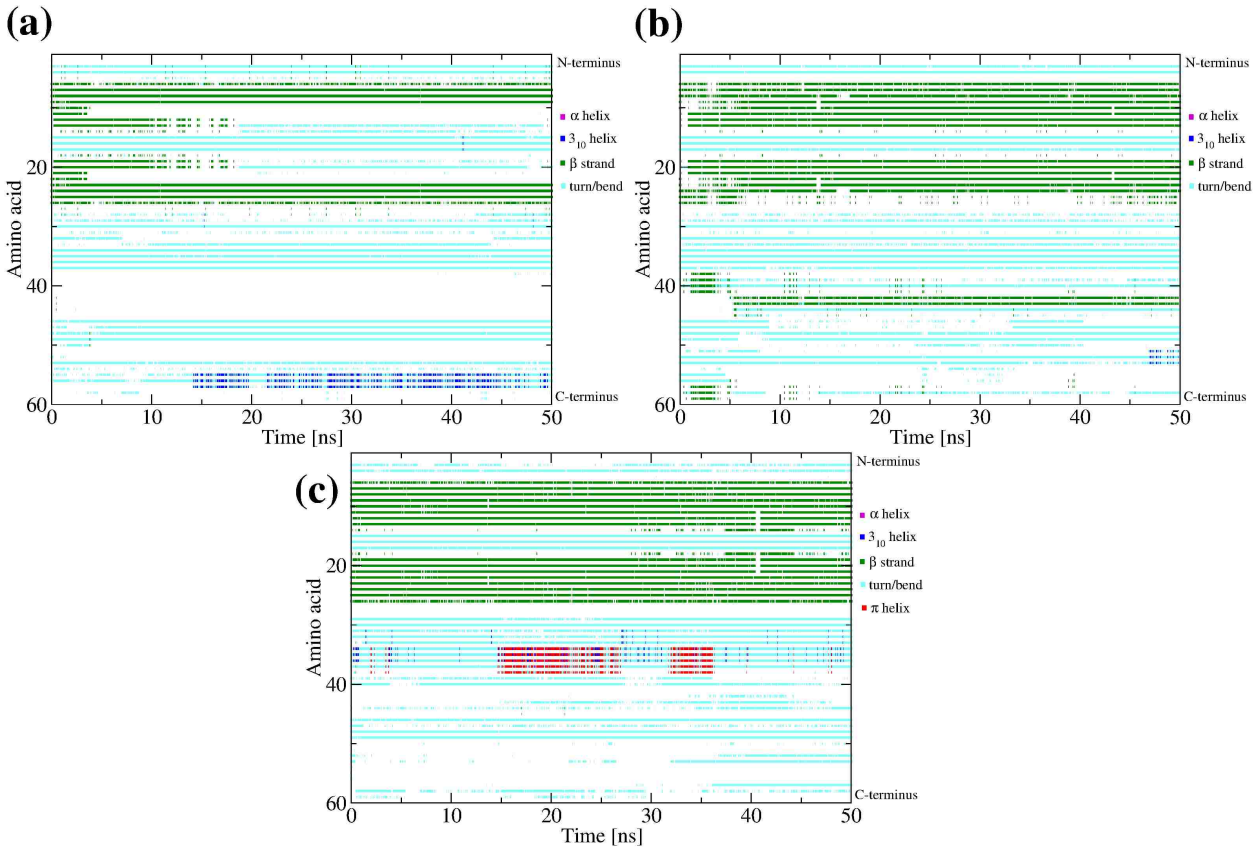

Figure 2.3:: Secondary structure elements of bovine αC domain in FACTS implicit solvent.

Time series analysis of secondary structure elements for each residue in the αC domain based on the STRIDE algorithm. Cyan corresponds to turns/bends, green corresponds to β-strands, pink corresponds to α-helices, blue corresponds to 3_{10}-helices and red corresponds to π-helices.

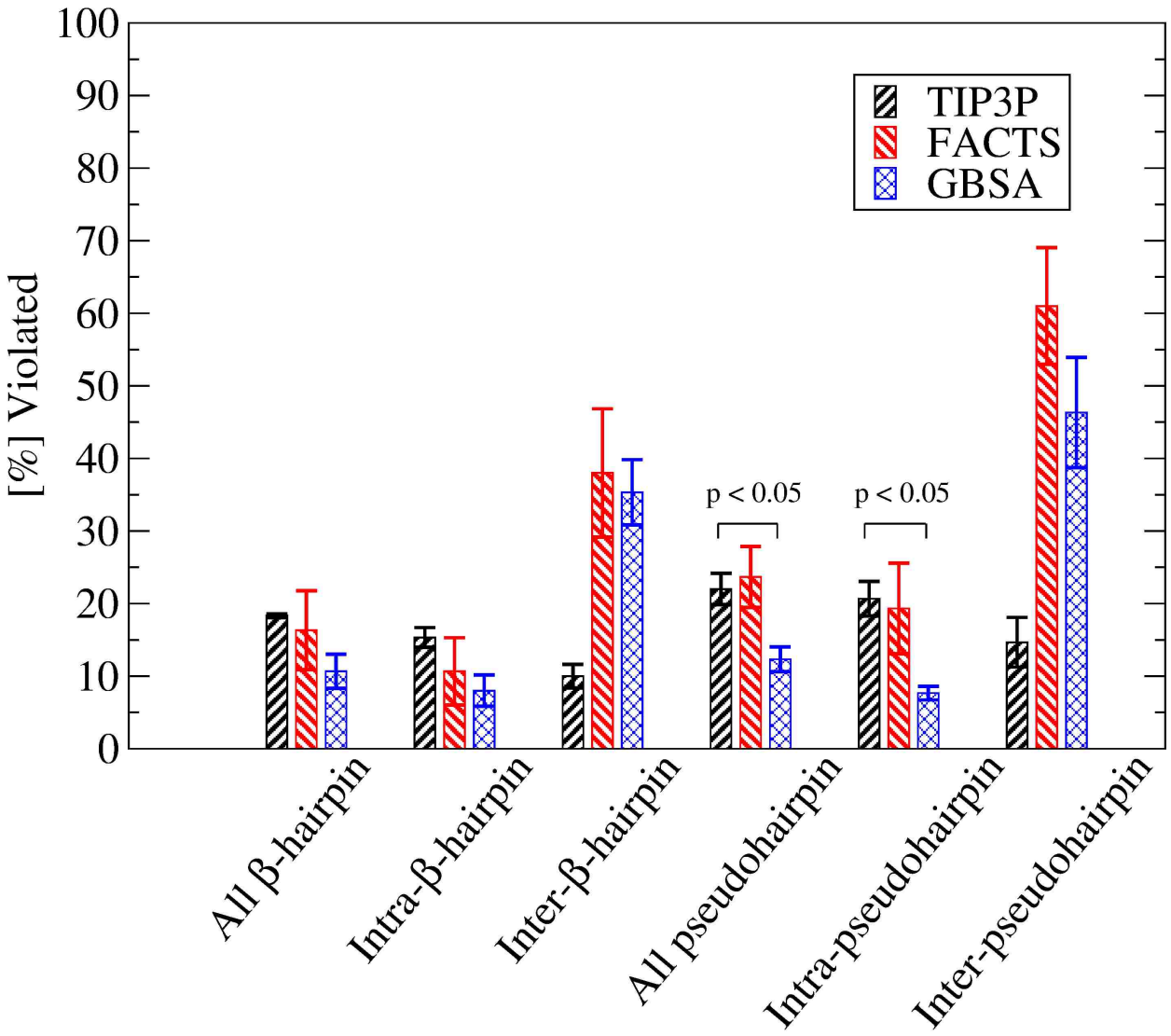

Figure 2.4:: NOE violations of Bovine αC domain monomer in explicit solvent, FACTS and GBSA implicit solvent.

Percent NOE violations for the β-hairpin, linker and pseudohairpin regions for the bovine αC domain in TIP3P explicit solvent, FACTS and GBSA implicit solvent in black, red and blue bars, respectively. The values were averaged over three 300-K simulations. Error bars represent the standard error of the mean.

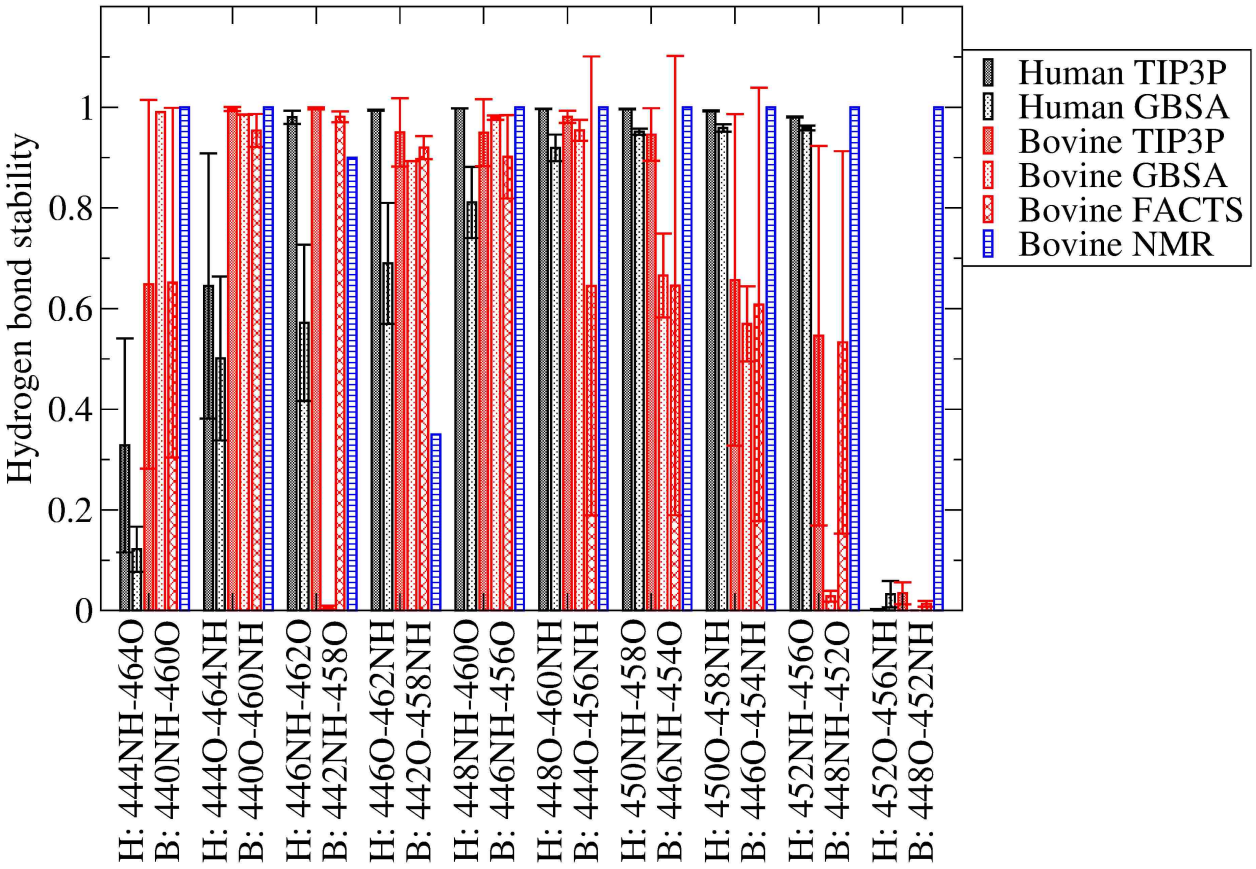

Figure 2.5:: Hydrogen bond stabilities for human and bovine αC domain simulations

Average hydrogen bond stabilities were calculated for ten hydrogen bonds along the human β-hairpin region. The values were averaged over three 300-K simulations. Error bars represent the standard deviation.

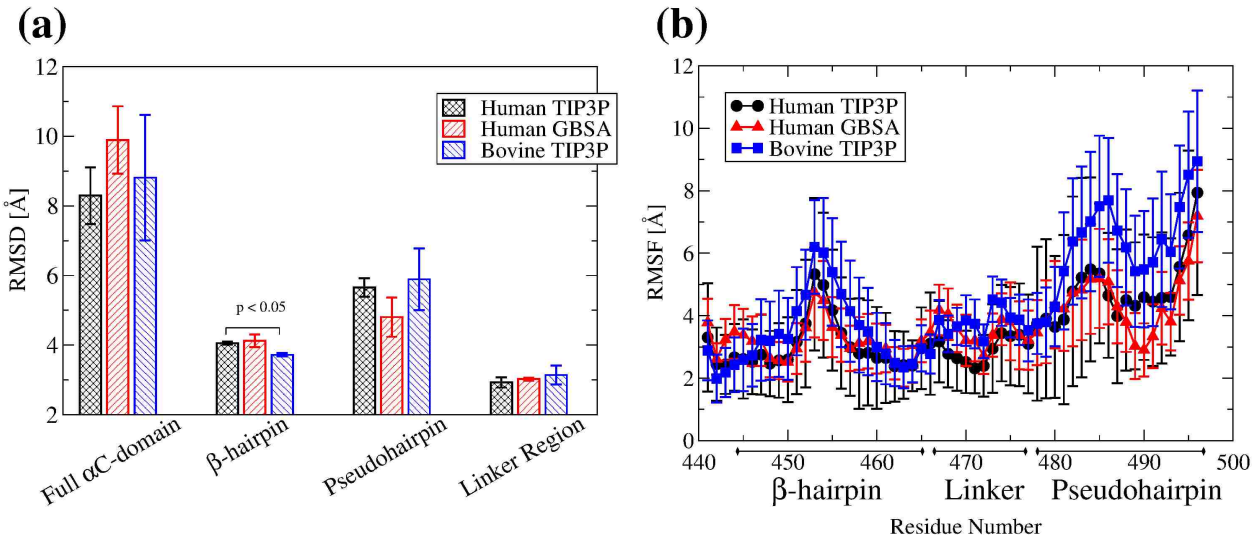

Figure 2.6:: Comparison of RMSD and RMSF between human and bovine αC domain in explicit solvent and GBSA.

(a) Mean RMSD of each region of the human and bovine αC domains in TIP3P explicit solvent and the human αC domain in GBSA implicit solvent. The values were averaged over three 300-K simulations. Error bars represent the standard error of the mean. **(b)** Mean RMSF of each residue in the human and bovine αC domains in TIP3P explicit solvent and the human αC domain in GBSA implicit solvent. The values were averaged over three 300-K simulations. Error bars represent the standard error of the mean.

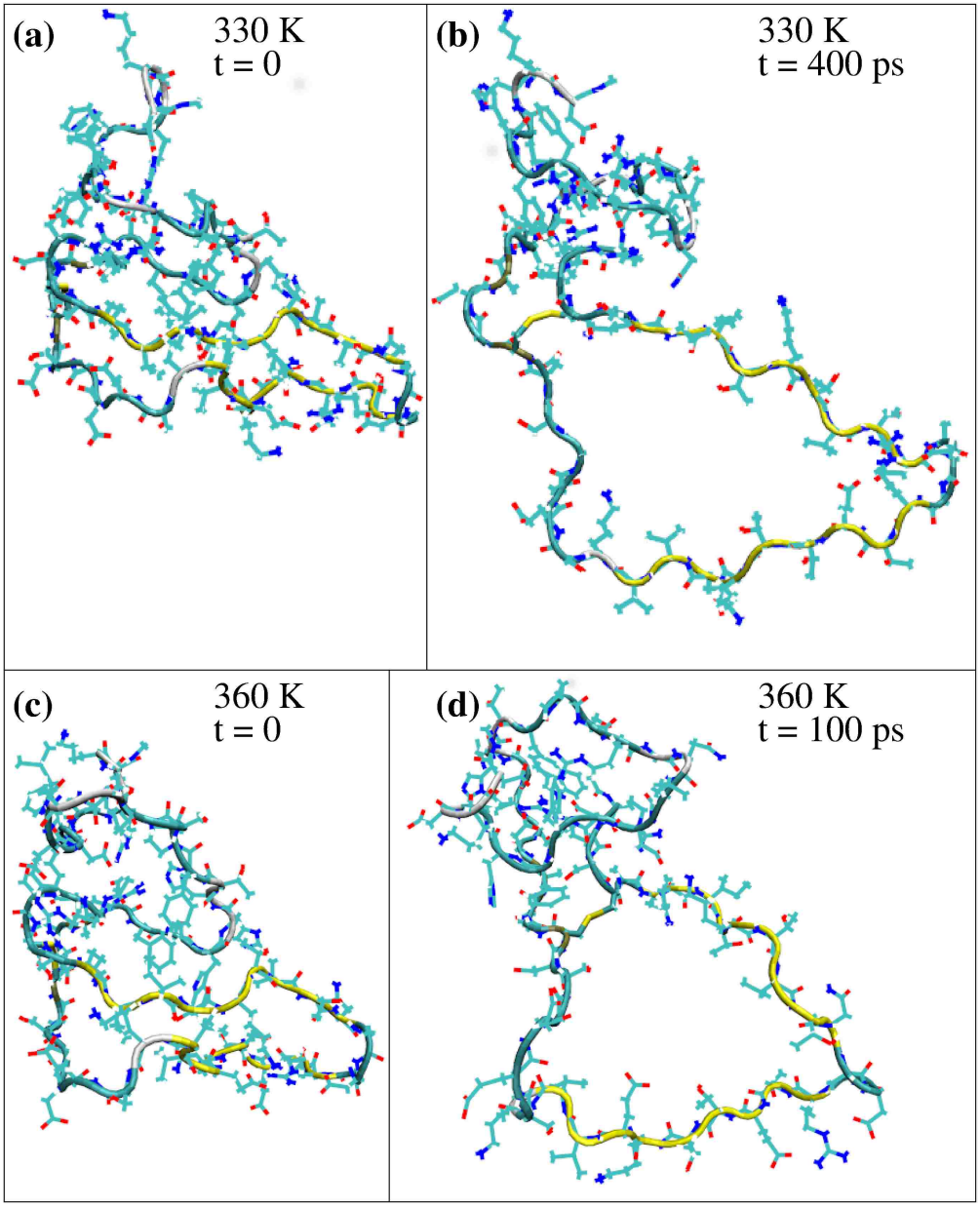

Figure 2.7:: High temperature simulations of the bovine αC domain in GBSA implicit solvent.

(a) Folded bovine αC domain after equilibration at 330 K. **(b)**Unfolded bovine αC domain after 400 ps at 330 K. **(c)** Folded bovine αC domain after equilibration at 360 K. **(d)** Unfolded bovine αC domain after 100 ps at 300 K.

Chapter 3

Oxidation-induced destabilization of the fibrinogen αC-domain dimer investigated by molecular dynamics simulations.

The work described in this chapter comes from a manuscript published in Proteins [96]

Abstract.

Upon activation, fibrinogen is converted to insoluble fibrin, which assembles into long strings called protofibrils. These aggregate laterally to form a fibrin matrix that stabilizes a blood clot. Lateral aggregation of protofibrils is mediated by the αC domain, a partially structured fragment located in a disordered region of fibrinogen. Polymerization of αC domains links multiple fibrin molecules with each other enabling the formation of thick fibrin fibers and a fibrin matrix that is stable but can also be digested by enzymes. However, oxidizing

agents produced during the inflammatory response have been shown to cause thinner fibrin fibers resulting in denser clots, which are harder to proteolyze and pose the risk of deep vein thrombosis and lung embolism. Oxidation of Met^{476} located within the αC domain is thought to hinder its ability to polymerize disrupting the lateral aggregation of protofibrils and leading to the observed thinner fibers. How αC domains assemble into polymers is still unclear and yet this knowledge would shed light on the mechanism through which oxidation weakens the lateral aggregation of protofibrils. This study used temperature replica exchange molecular dynamics simulations to investigate the αC-domain dimer and how this is affected by oxidation of Met^{476}. Analysis of the trajectories revealed that multiple stable binding modes were sampled between two αC domains while oxidation decreased the likelihood of dimer formation. Furthermore, the side chain of Met^{476} was observed to act as a docking spot for the binding and this function was impaired by its conversion to methionine sulfoxide.

3.1 Introduction.

Fibrinogen is a plasma protein that is essential for the blood coagulation process. In its inactive form, fibrinogen is normally soluble. However, once activated through the action of thrombin it is called fibrin and it becomes insoluble, forming a matrix that stabilizes a blood clot. The structure of fibrinogen consists of six chains (two Aα, two Bβ and two γ chains, respectively) arranged in an elongated manner (Figure 3.1a) [13]. The N-termini are located in a central nodule called the E-domain while the C-termini of Bβ and two γ chains form distally located globular folds called D-domains (Figure 3.1a) [13]. On the other hand, each of the Aα chains contains an intrinsically disordered C-terminal region that is not resolved in the X-ray crystallographic structure of fibrinogen [13]. Located within this disordered region is the so called αC domain whose N-terminal sub-domain presents an ordered structure in NMR experiments with the bovine sequence (Figure 3.1b; for simplicity, "αC domain" in this manuscript will from now on refer to the N-terminal sub-domain). The NMR conformers reveal that the αC domain consists of a β-hairpin and a pseudohairpin linked together also by a disulfide bond (Figure 3.1b) [14,17,111]. Thrombin activates fibrinogen and turns it into fibrin by cleaving N-terminal peptides of the Aα and Bβ chains (so called fibrinopeptides). The exposed "knobs" in the remaining chains interact with so called "holes" in the D-domains of another fibrin molecule in a half-staggered arrangement that leads to protofibrils [3]. The αC domains are thought to interact with the E nodule in fibrinogen, but upon cleavage of the fibrinopeptides they are liberated and interact intermolecularly mediating the lateral aggregation of protofibrils into thick fibrin fibers [3, 16]. There is experimental evidence that during such intermolecular interactions the αC domains polymerize into so called αC polymers [15]. However, there are currently neither experimental nor computational studies that shed light on the structure of αC polymers at atomic level of detail.

Yet, understanding the structural details of the polymerization process of αC domains is crucial to explain structural modifications in fibrin fibers observed in the presence of oxidants produced during inflammation [1]. The inflammatory response is a defense mechanism,

which evolved in higher organisms as a protection against infection by external pathogens. Under inflammatory conditions, neutrophils secrete the enzyme myeloperoxidase, which converts hydrogen peroxide into the oxidizing agent hypochlorous acid (HOCl). A study using mass spectrometry revealed that HOCl converts specific methionine residues in fibrinogen to methionine sulfoxide causing the formation of a dense network of thin fibers [1]. In particular, Met^{476} in the αC domain was found to be oxidized and this is postulated to disrupt the lateral aggregation of protofibrils [1,55]. This is supported by the observation that the fibrin clots obtained under oxidizing conditions exhibit similar characteristics as fibrin gels where human αC domain was replaced with the chicken sequence [52]. The resulting denser fibrin clots under oxidizing conditions are more resistant to fibrinolysis and were also found to be weaker [1]. This mechanism likely evolved as a protection against pathogens by trapping them inside a clot [37,54]. However, oxidized clots can lead to excessive bleeding right after traumatic injury while later on they can detach from the site of injury, travel in the vascular system and lead to a pulmonary embolism [1,5,60,112,113]. Furthermore, chronic inflammation may also increase the risk of thrombosis associated with fibrinogen oxidation [37,114]. For these reasons, it is essential to study the structural mechanism by which oxidation disrupts the normal function of the αC domain. To date, structural studies of the αC domain have been very limited. In fact, currently only a homology model of the human αC domain exists based on the NMR structure of its bovine sequence [4], highlighting the difficulty of using experimental structural determination techniques with this fibrinogen fragment.

In this study, we used molecular dynamics (MD) simulations with the enhanced sampling method, temperature replica exchange MD (T-REMD), to investigate the dimer of αC domains and how this is altered by oxidation of Met^{476}. The goal was to understand how Met^{476} contributes to the formation of the dimer and how conversion to methionine sulfoxide impairs its function. In order to improve sampling efficiency, the T-REMD simulations were performed using an implicit solvent model [82,103]. The obtained models of the αC-domain dimers were tested and further analyzed through simulations in explicit solvent.

3.2 Materials and Methods.

3.2.1 Initial conformations.

The coordinates for the NMR solution structure of bovine αC domain were used to start simulations of the bovine sequence [14]. Simulations with the human αC domain were started from a homology model based on the bovine sequence [4]. The initial conformation for the simulations of the oxidized structure was obtained by replacing Met^{476} with methionine sulfoxide per analogy and subsequently performing 100 steps of steepest descent minimization with the program CHARMM [61] *in vacuo* while holding all atoms fixed except the mutated residue. While it is possible that methionine oxidation could result with methionine sulfone, with two oxygen atoms bonded to the sulfur, this is unlikely given that mass spectrometry data suggests that only the sulfoxide form is present at detectable levels [1].

3.2.2 General setup of the systems.

The MD simulations were performed with the program NAMD [65] using the CHARMM36 force field [62]. The force field parameters for methionine sulfoxide were obtained from the SwissSidechain website [63] and adapted per analogy for the CHARMM36 force field. Simulations were performed in either explicit or implicit solvent and are summarized in Table 3.1.

3.2.2.1 Explicit solvent simulations.

The TIP3P model of water was used in the explicit solvent simulations. The monomeric αC domain was inserted into a cubic water box with side lengths of 75 Å, while the side lengths of 100 Å and 150 Å were used for the unoxidized and oxidized dimeric αC-domain, respectively. Chloride and sodium ions were added to neutralize the system and approximate a salt concentration of 150 mM. The water molecules overlapping with the protein or the ions were removed if the distance between the water oxygen and any atom of the protein or any ion was smaller than 3.1 Å. The resulting systems contained a total of ca. 39,000, 94,000 and 324,000

atoms for monomeric, unoxidized dimeric and oxidized dimeric αC domain, respectively. To avoid finite size effects, periodic boundary conditions were applied. After solvation, each system underwent 500 steps of minimization while the coordinates of the heavy atoms of the protein were held fixed, which was followed by 500 steps with no restraints. Electrostatic interactions were calculated within a cutoff of 10 Å while long-range electrostatic effects were taken into account by the particle mesh Ewald summation method [64]. Van der Waals interactions were treated with the use of a switch function starting at 8 Å and turning off at 10 Å. A cutoff of 12 Å was used to generate the list of non-bonded atom pairs, which was updated every 20 ps [115].

3.2.2.2 Implicit solvent simulations.

In the implicit solvent simulations, solvation effects were taken into account through the generalized Born model with the solvent accessible surface area term (GBSA) [70, 71]. The electrostatic interactions cutoff was 12 Å and van der Waals interactions were treated with the use of a switch function starting at 10 Å and turning off at 12 Å. A cutoff of 14 Å was used for the list of non-bonded atom pairs. In both, explicit and implicit solvent simulations, the dynamics were integrated with a time step of 2 fs. The covalent bonds involving hydrogens were rigidly constrained by means of the SHAKE algorithm with a tolerance of 10^{-8}. Snapshots were saved every 10 ps for trajectory analysis.

3.2.3 Equilibration and room-temperature simulations.

Before production runs, harmonic constraints were applied to the positions of all heavy atoms of the protein to equilibrate the system at 300 K for a total duration of 0.2 ns. To equilibrate the position of atoms around a methionine sulfoxide side chain, harmonic constraints were kept on all heavy atoms except those of the methionine sulfoxide residue and the neighboring amino acids, and equilibration was continued for another 2 ns. After these equilibration steps, the harmonic constraints were released. In each room-temperature simulation, the first 10 ns of unconstrained simulation were also considered part of the equilibration and were thus

not used for analysis. During, both equilibration and production, the temperature was kept constant at 300 K by using the Langevin thermostat [67] with a damping coefficient of 1 ps^{-1}. In explicit solvent simulations, the pressure was held constant at 1 atm by applying a pressure piston [68]. Explicit solvent simulations were started from either the bovine or the human αC domain, respectively (Table 3.1). Implicit solvent simulations were run with the human αC domain to compare solvation models and to justify the use of GBSA in the enhanced sampling runs described below (Table 3.1). Each simulation was started with different initial random velocities to ensure that different trajectories were sampled despite the same initial conformation was used in multiple runs. Explicit solvent simulations at 300 K were also started from representative conformations of free energy minima identified in enhanced sampling runs described below (Table 3.1).

3.2.4 Analysis of hydrogen bonds and side chain contacts.

The room-temperature trajectories were screened for persistent hydrogen bonds and side chain contacts. A hydrogen bond was said to be formed if the H$\cdots$A distance was not larger than 2.7 Å and the D-H$\cdots$A angle was at least 120, where a donor D and an acceptor A could both be either an oxygen or a nitrogen. A native side chain contact was defined to occur when the distance between the centers of mass of the two side chains was within 6 Å. Hydrogen bonds and side chain contacts present in at least 60% of the frames of a simulation were considered to be persistent in that particular simulation.

3.2.5 Nuclear Overhauser effect distance restraints.

The MD trajectories with the bovine αC domain were compared with distance restraints derived from Nuclear Overhauser effect (NOE) experiments [14]. A NOE distance restraint was considered violated in the simulations if the equation $\langle r(t)^{-6} \rangle^{-\frac{1}{6}} \leq {r_{\max}}^{\exp} + 0.5$ Å was not fulfilled, where $r(t)$ is the interproton distance, at simulation time t, ${r_{\max}}^{\exp}$ is the experimentally determined upper distance limit, and $\langle\,\rangle$ represents a time average. A 0.5 Å buffer term is added to account for edge cases while introducing a minimal amount of

uncertainty. NOE violations are reported as percentages, which are defined by the equation $\mathrm{NOE}_{\%}=\frac{N_{\mathrm{violations}}}{N_{\mathrm{r}^{\mathrm{exp}}}} \times 100\%$ where $\mathrm{N}_{\mathrm{violations}}$ is the number of NOE violations and $\mathrm{N}_{\mathrm{r}^{\mathrm{exp}}}$ is the number of experimentally determined distance ranges. NOE violations were calculated also along the 20 NMR conformers for comparison.

3.2.6 Enhanced sampling simulations of the dimer.

3.2.6.1 Temperature replica exchange molecular dynamics simulations.

Temperature replica exchange molecular dynamics (T-REMD) utilizes temperature to overcome free energy barriers [79]. Here, we used T-REMD to explore the free energy landscape (FEL) of the αC-domain dimer in the unoxidized state and compare it to the FEL in the oxidized state. The GBSA implicit solvation model [73, 82, 103] was used to decrease the number of atoms, thus requiring fewer replicas. The simulations consisted of eight replicas with the following temperatures: 300.0 K, 306.30 K, 312.74 K, 319.31 K, 326.02 K, 332.87 K, 339.86 K, 347.00 K. This set of temperatures was obtained using the temperature generator for REMD simulations available online through http://folding.bmc.uu.se/remd/ [81] indicating a desired swapping frequency of 0.45. After every 10 picoseconds, temperatures between adjacent replicas were swapped according to a Metropolis criterion [79, 81]. After running the simulations, the achieved swapping frequencies were 0.43 $\pm$ 0.04 for the unoxidized simulations and 0.40 $\pm$ 0.01 for the oxidized simulations. To enhance sampling, a 2 Å restraint on the RMSD of each β-hairpin was added. This is justified because we do not expect the β-hairpins to significantly change conformation upon dimerization. The distance between the centers of mass of the monomer was also restrained to avoid the single monomers diffusing away from each other. Each of the eight replicas was run for 125 ns of simulation time for a total of 1 μs of sampling for both unoxidized and oxidized configurations, respectively.

3.2.6.2 Representation of the free energy landscape.

The radius of gyration of each monomer was used to represent the FEL of the αC-domain dimer sampled in the T-REMD simulations. The resulting FEL is a two-dimensional histogram, and is represented as a matrix where the x- and y-axes correspond to the radius of gyration of each monomer and the z-axis is the energy of the bin derived using the Boltzmann distribution $G_{\mathrm{Bin}} = -RT\ \ln(N_{\mathrm{FramesInBin}})$ where G_{Bin} is the Gibbs free energy of the bin, R is the ideal gas constant in (kcal $\mathrm{g^{-1}}$ $\mathrm{mol^{-1}}$ $\mathrm{K^{-1}}$), $T = 300$ K and $N_{\mathrm{FramesInBin}}$ is the number of frames in a particular bin. The bin width was chosen as 1/3 Å and 1/4 Å for the simulations with the unoxidized and oxidized dimer, respectively.

3.2.6.3 Selection of minima.

The following criteria were used to define the minima in the obtained FELs. First, the bin with the lowest free energy was defined as the deepest minimum. Then, those bins whose Boltzmann factor relative to the deepest minimum was ≥ 0.2 were considered local minima. The Boltzmann factor was defined as $f_{\mathrm{B}} = e^{(\mathrm{E_0\text{-}E_i})/\mathrm{RT}}$ where E_0 is the energy from the deepest well, E_{i} is the energy of the ith well, R is the ideal gas constant, $T = 300$ K and $0 \leq f_{\mathrm{B}} \leq 1$. f_{B} is the theoretical transition probability between the most populated state and the ith state. If two local minima were separated by a single bin and their energies differed by no more than 1 kcal/mol, the two bins were merged into one single minimum. The so identified deepest minimum and local minima were used to suggest models of the αC-domain dimer. For this purpose, hierarchical clustering was performed on all frames in each minimum. The centroid of the largest cluster was calculated and the frame nearest the centroid was chosen as the representative frame of a particular minimum [109]. The so obtained models of the αC-domain dimer, in both the unoxidized and oxidized state, were further analyzed through explicit solvent simulations.

3.3 Results

3.3.1 Flexibility of monomers: bovine vs. human.

Before modeling the dimer, it is necessary to study the conformational plasticity of a single monomer. While experimental coordinates exist for the αC domain sequence, only a homology model of the human αC domain is available [4]. For this reason, explicit solvent simulations were run with both, human and bovine αC domain (Table 3.1), and the results were compared (Figure 3.2). Both human and bovine αC domain present large C_α root mean square fluctuations (RMSF) for the pseudohairpin (Figure 3.2a), while the C_α root mean square deviation (RMSD) from the initial conformation for both hairpins was generally below 4 Å (Figure 3.2b). Thus, most of the flexibility is likely due to the rigid-body motion of the pseudohairpin, which is facilitated by the rather flexible 11-residue linker between the β-hairpin and the pseudohairpin regions (Figure 3.2b). On the other hand, the β-hairpin is rather stable presenting persistent backbone hydrogen bonds in both bovine and human αC domain simulations (Figure 3.3). The similarity between bovine and human αC domain suggests that the homology model of the human sequence is an acceptable approximation.

To exclude that the observed flexibility could be an artefact of the simulation setup, the trajectories with the bovine sequence were compared to NOE-derived distance restraints [14]. The percentage of violations for the entire αC domain and within the two main regions (β-hairpin and pseudohairpin) were generally below 20% and were comparable to the percentage of violations within the 20 NMR conformers themselves (Figure 3.3). Interestingly, the simulations presented fewer violations between each region and the rest of the protein than the NMR conformers themselves (Figure 3.4). This indicates that the flexibility observed in the simulations is consistent with experimental data and the rigid-body motions of the β-hairpin and the pseudohairpin with respect to each other is likely a realistic characteristic of the αC domain.

3.3.2 Models of the αC-domain dimer

3.3.2.1 Justification for the use of GBSA.

The T-REMD method was used here to overcome energy barriers and sample the FEL of the αC-domain dimer. A previous T-REMD study with just the αC domain monomer was performed in explicit solvent and the size of the system required a large number of replicas [39]. A similar study with the dimer would be computationally prohibitive because an even larger number of replicas would be required for temperature swaps to occur. In this study, the GBSA implicit solvation model was used to reduce the number of atoms and replicas [73, 82, 103]. To test whether GBSA is a suitable approximation, 300-K simulations were performed with human αC domain using GBSA and were compared to the explicit solvent simulations. The αC domain presented a similar flexibility in the implicit solvent as in the explicit solvent simulations (Figure 3.2b). Generally, the β-hairpin has a similar pattern of hydrogen bond formation between explicit and implicit solvent simulations (Figure 3.3a). The exception are the two hydrogen bonds between residues 446 and 462, which are slightly less stable in implicit solvent (Figure 3.3a). Interestingly, in the NMR conformers of the bovine sequence [14], the backbone of residue 458 (corresponding to residue 462 in the human sequence) presents a relatively large flexibility as indicated by the relatively weak 458NH$\cdots$O442 hydrogen bond and the fact that only one NOE distance was measured involving residue 458 (Figure 3.3a). Thus, the discrepancy between implicit and explicit solvent in this region could be due to the former sampling a larger conformational space, which is likely in part due to the absence of the friction normally provided by water molecules.

3.3.3 Comparison between the unoxidized and the oxidized αC-domain dimer through REMD simulations.

3.3.3.1 Models of the αC-domain dimer in the unoxidized and oxidized states.

In order to overcome energy barriers, REMD simulations were performed with the αC-domain dimer. One set of 8 replicas was run with the unoxidized αC domain and another

set after replacing Met^{476} with $Met(O)^{476}$ in both monomers. For each set, the frames were projected onto the radius of gyration of each monomer (Figure 3.5). For the unoxidized state, seven distinct minima were identified (Figure 3.5a), while for the oxidized state three distinct minima were observed (Figure 3.5b). Representative structures from the minima highlight the variety in binding modes for both the unoxidized (Figure 3.6a) and the oxidized (Figure 3.7a) dimers. However, the oxidized αC domain can access a relatively smaller number of binding modes than in the unoxidized state (Figure 3.5). If we assume that all conformations sampled in all minima of either the unoxidized or oxidized FEL define a congregate dimeric state, then we can compare the thermodynamic stability of the unoxidized dimer with that of the oxidized dimer and estimate the change of the ΔG of dimerization upon oxidation, i.e., $\Delta\Delta G^{\text{oxidation}}_{\text{dimer}}$. One can write $\Delta\Delta G^{\text{oxidation}}_{\text{dimer}} = (G^{\text{ox}}_{\text{dimer}} - G^{\text{ox}}_{\text{monomer}}) - (G^{\text{unox}}_{\text{dimer}} - G^{\text{unox}}_{\text{monomer}})$. If we substitute $G^{\text{ox}}_{\text{dimer}} = -RT\ \ln(\sum_{i=1}^{3} N_i)$ and $G^{\text{unox}}_{\text{dimer}} = -RT\ \ln(\sum_{i=1}^{7} N_i)$, and assume that oxidation does not significantly alter the thermodynamic stability of single monomers (i.e., $G^{\text{unox}}_{\text{monomer}} \approx G^{\text{ox}}_{\text{monomer}}$), we obtain $\Delta\Delta G^{\text{oxidation}}_{\text{dimer}} = 6.23$ kcal/mol. N_i is the number of frames in the ith minimum.

3.3.3.2 Stability of the dimer models and role of Met^{476} in dimer formation.

Molecular dynamics simulations in explicit solvent were started from the representative structures in each minimum (Figures 3.6 and 3.7) to test the stability and further investigate differences between unoxidized and oxidized dimer models (Table 3.1). The radius of gyration for the total dimer was calculated along the explicit solvent simulations and compared to its value in the initial structure. In five out of seven simulations with the unoxidized dimer, the average radius of gyration (calculated over the last 40 ns of the in total 50-ns long simulations) was within 2 standard deviations from the value calculated for the initial structure, while in one run it was slightly smaller and in another slightly larger (Figure 3.6b). In two out of three simulations with the oxidized dimer, the average radius of gyration was also within two standard deviations from the initial value while in one run it was slightly smaller (Figure 3.7b). Calculated for each single monomer, the radius of gyration was generally also

within two standard deviations from the value of the initial conformation (supplementary Figure S1). These observations illustrate that the dimer models obtained from the T-REMD simulations are stable along the explicit solvent simulations. The explicit solvent simulations were also used to analyze differences between unoxidized and oxidized dimer at the site of Met^{476}. Analysis of the solvent accessible surface area (SASA) revealed that in all simulations with the unoxidized dimer at least one of the two Met^{476} was on average partially buried (see Figure 3.8a and its caption for the definition of "partially buried"). In comparison, in one out of three simulations with the oxidized dimer, none of the methionine residues was partially solvent exposed (Figure 3.8a). In general, Met^{476} had a smaller SASA then $Met(O)^{476}$ (Figure 3.8b, averages calculated using only the methionine with the smaller SASA for each run). Analysis of persistent side chain contacts revealed that Met^{476} was shielded from the solvent in the simulations with the unoxidized dimer in part by interdomain (Section 3.7a-e) and in part by intradomain side chain contacts (Section 3.7f). However, in the simulations with the oxidized dimer, $Met(O)^{476}$ presented only intradomain but no interdomain side chain contacts (Section 3.7f). This is probably due to the increase in hydrophilicity upon conversion of Met^{476} to methionine sulfoxide, which makes this side chain more likely to be solvent exposed. These observations support the idea that Met^{476} could serve as a docking spot for the αC-domain dimerization process but oxidation impairs this function.

3.4 Discussion.

The oxidizing agent HOCl produced during inflammation is known to alter the mechanical characteristics of fibrin clots rendering them weaker but also denser and thus more difficult to proteolyze. Although this is part of a defense mechanism meant to trap and incapacitate a pathogen, it can lead to a clot to detach from a site of injury and cause a lung embolism. The altered clot properties under inflammatory conditions have been linked to the oxidation of Met^{476} in the αC domain [1], a region of fibrinogen that is thought to polymerize intermolecularly linking protofibrils to thick fibrin fibers [15, 16]. However, because of its flexibility

experimental studies of human αC domain and its polymerization have been limited. Here, MD simulations were used to propose models of the αC-domain dimer and study the effect of Met^{476} oxidation. The following four conclusions emerge from this study.

First, the β hairpin and the pseudohairpin of the αC domain undergo rigid-body motions with respect to each other. This flexibility is key for the dimerization process. In fact, an experimental study showed that engineering a disulfide bond that locks the motion of the two hairpins with respect to each other prevents dimerization [15]. The flexibility of the hairpins likely allows multiple binding modes between two or more αC domains.

Second, multiple distinct binding modes were observed in enhanced sampling simulations. This information could be useful to guide future NMR studies of αC-domain polymerization. The availability of multiple binding modes is likely to increase the rate of binding between αC domains facilitating the lateral aggregation of protofibrils into thick fibrin fibers [116]. On the other hand, oxidation of Met^{476} was observed in the enhanced sampling simulations to decrease the thermodynamic stability of the αC-domain dimer. The change in the free energy of dimerization was estimated to be 6.23 kcal/mol upon oxidation. Interestingly, the binding free energy between αC domains has been determined by sedimentation to be -6.7 kcal/mol [4] suggesting that oxidation may drastically weaken the dimer.

Third, in each dimer model of the unoxidized state Met^{476} of at least one of the monomers is partially buried and in most cases it is shielded from the solvent by interdomain side chain contacts. This supports a model for the αC-domain dimerization process where the side chain of Met^{476} serves as a docking spot. Since most dimer models presented one Met^{476} partially buried and one completely solvent exposed, it can be speculated that the solvent exposed methionine residue serves as a docking spot for a third αC domain, and so on, enabling polymerization. However, in the oxidized state, $Met(O)^{476}$ did not engage in interdomain side chain contacts thus weakening the polymerization process.

Fourth, this study presents an example of how the binding between two flexible peptides can be modeled through computational simulations. This would be challenging for conventional docking software like AutoDock [117] or the docking mode of Rosetta [83],

since the first assumes the involved molecules to be rigid and the second would need to be trained with a large set of similar examples. Thus, MD simulations are a powerful tool to study the binding function of flexible peptides and miniproteins and how this is affected by post-translational modifications such as oxidation.

In conclusion, the present study provides an explanation how oxidation of a methionine residue in the fibrinogen αC domain impairs the lateral aggregation of protofibrils leading to thinner fibrin fibers and the observation of a denser fibrin clot [1]. In general, the destabilizing effect of methionine oxidation is likely an important link between inflammation and the thrombotic function of blood proteins. In fact, methionine oxidation has been found to activate the coagulatory protein von Willebrand factor by destabilizing one of its domains [118]. The dimer models presented here could be used to suggest the design of molecules that stabilize the interaction between two αC-domain dimers in order to maintain a normal clot morphology also under oxidizing conditions. For this purpose, the interface of the dimer models could be searched for pockets that can be used as docking spots for designed molecules administered to patients at risk of lung embolism such as after traumatic injury or surgery.

3.5 Acknowledgements.

We would like to thank Dr. Wendy Thomas and Dr. Jim Pfaendtner for helpful and interesting discussions. The simulations were performed in part on the Comet supercomputer at the San Diego Supercomputing Center thanks to a XSEDE allocation [119] with grant number TG-MCB140143, which is made available through NSF support, and in part on the Hyak supercomputer at the University of Washington. This research was financially supported by a NIH Career Development Award K25HL118137, a NIH Grant RO1AI50940 and a University of Washington Royalty Research Fund Award A132953 to GI. We have no conflict of interest to declare.

3.6 Tables.

Name	**Solvent**	**Starting Structure**	**4° Structure**	**Type**	**Duration [ns]**
Bovine Expl.	TIP3P	2JOR	Mono	300 K	3×50
Human Expl.	TIP3P	Homology model	Mono	300 K	3×100
Human Impl.	GBSA	Homology model	Mono	300 K	3×100
Human REMD	GBSA	Homology model	Dimer	REMD	8×125
Human Minima	TIP3P	Homology model minima 1-7	Dimer	300 K	7×50
Human Ox REMD	GBSA	Homology model $Met^{476} \rightarrow Met(O)^{476}$	Dimer	REMD	8×125
Human Ox Minima	TIP3P	Homology model minima 1-3	Dimer	300 K	3×50

Table 3.1:: Simulation systems.

	Minimum						
Side Chain Contact	1	2	3	4	5	6	7
Ser^{441}-Met^{476}	0.75						
Cys^{442}-Met^{476}			0.94				
Lys^{444}-Met^{476}			0.95				0.90
Val^{464}-Met^{476}							0.69
Ser^{466}-Met^{476}							0.96
Cys^{472}-Met^{476}						0.93	
Pro^{473}-Met^{476}					0.67	0.87	
His^{492}-Met^{476}					0.61		
Pro^{495}-Met^{476}					0.73		

Table 3.2:: Persistent interdomain side chain contacts involving Met^{476}.

3.7 Figures

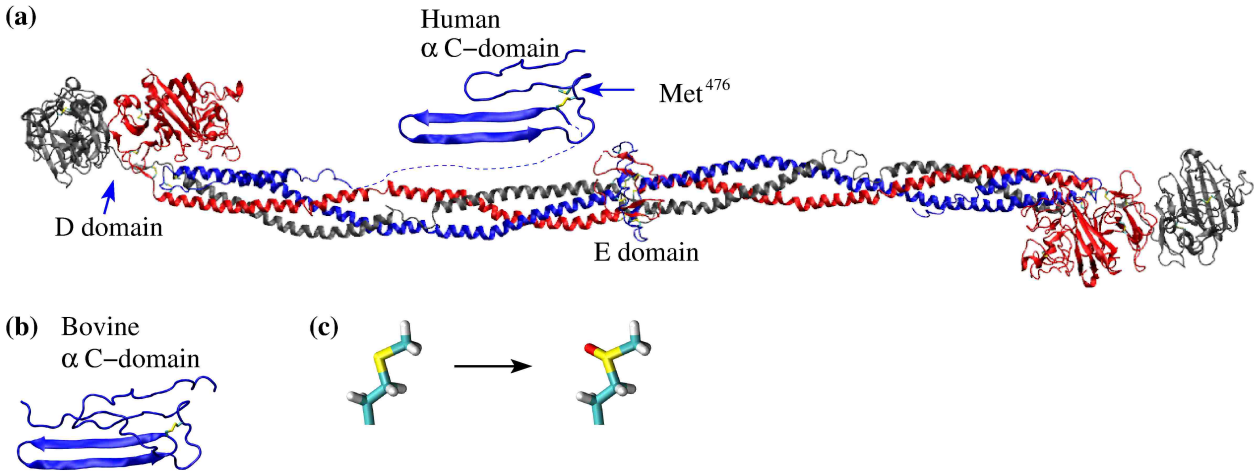

Figure 3.1:: Available structures of fibrinogen and the αC domain and illustration of methionine oxidation.

(a) The X-ray crystal structure of fibrinogen (PDB ID 3GHG) and the human homology model of the αC domain. The side chain of Met^{476} in the αC domain is shown in the stick and ball representation and labeled. **(b)** The bovine αC domain NMR solution structure (PDB ID 2JOR). **(c)** Conversion pathway of methionine to methionine sulfoxide.

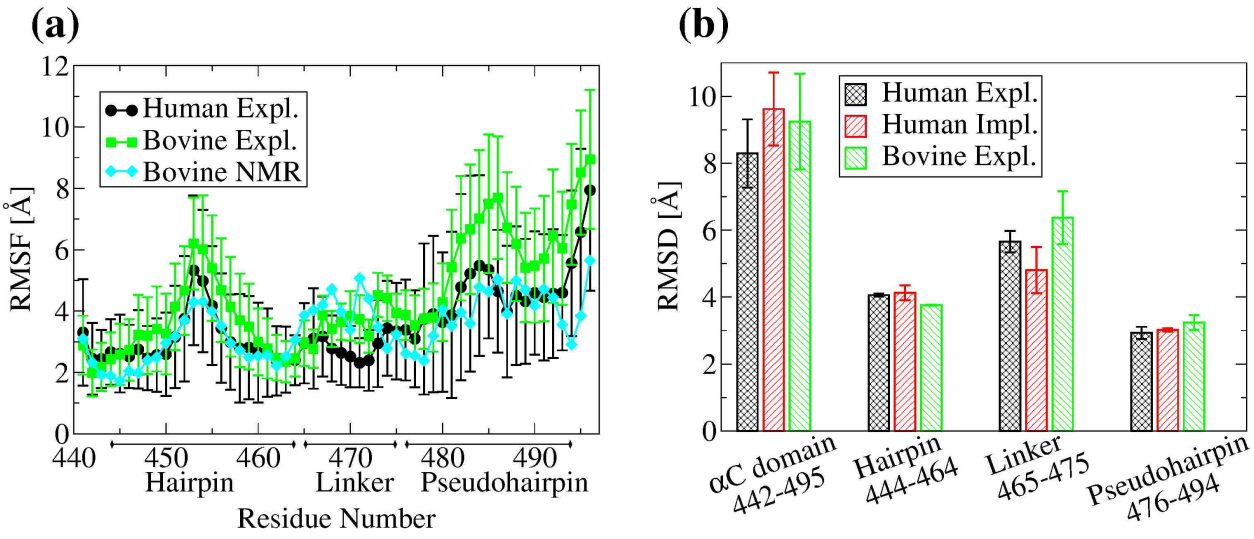

Figure 3.2:: Backbone flexibility at room temperature.

(a) Mean C_α RMSF calculated from the last 40 ns of the 50 ns of bovine simulations and 90 ns of the last 100 ns of human simulations in explicit solvent. Averages were computed over three 300-K simulations. Error bars represent the standard error of the mean. **(b)** The C_α RMSD from the initial conformation averaged over three 300-K simulations for the last 40 ns of the 50 ns bovine simuations and the last 90 of the 100 ns of the human simulations. The RMSD is calculated for the αC domain without the N- and C-terminal (residues 442-494), the β-hairpin region (residues 444-464), the linker region (residues 465-475) and the pseudohairpin region (residues 476-494). Error bars represent the standard error of the mean over three simulations.

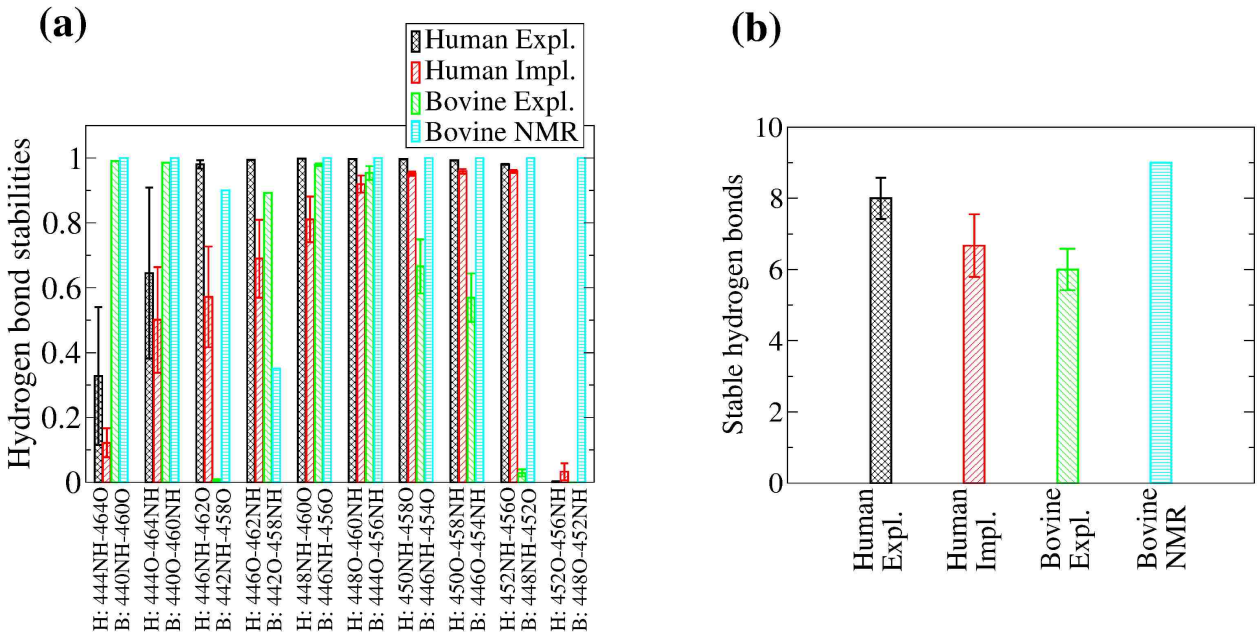

Figure 3.3:: Formation of backbone hydrogen bonds in the β-hairpin.

(a) Average hydrogen bond stabilities were calculated for ten hydrogen bonds along the human β-hairpin region. The values were averaged over three 300-K simulations. Error bars represent the standard deviation. **(b)** The number of persistent hydrogen bonds per simulation were averaged over three 300-K runs. Error bars represent the standard error of the mean. The definition of a persistent hydrogen bond is given in "Materials and Methods".

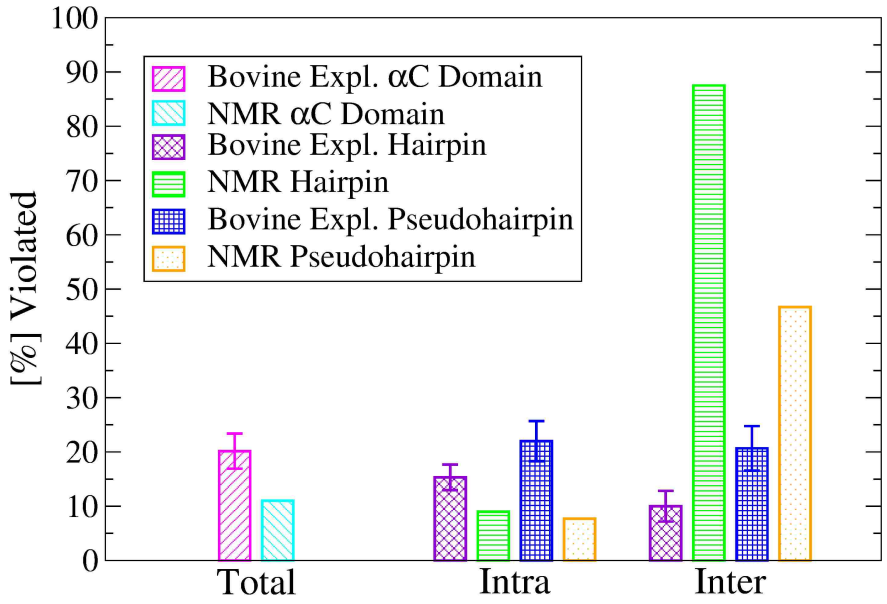

Figure 3.4:: Percent NOE violations along the simulations with the bovine sequence. The number of violations were computed using the last 40 ns of the in total 50-ns long simulations at 300 K. Values were averaged over three runs and the standard error of the mean was calculated and shown as error bars. NOE violations of PDB ID 2JOR were computed across all NMR conformers. Violations between either the β-hairpin or the pseudohairpin and the rest of the protein are reported as "inter". Violations occurring within the β-hairpin or pseudohairpin regions are reported as "intra".

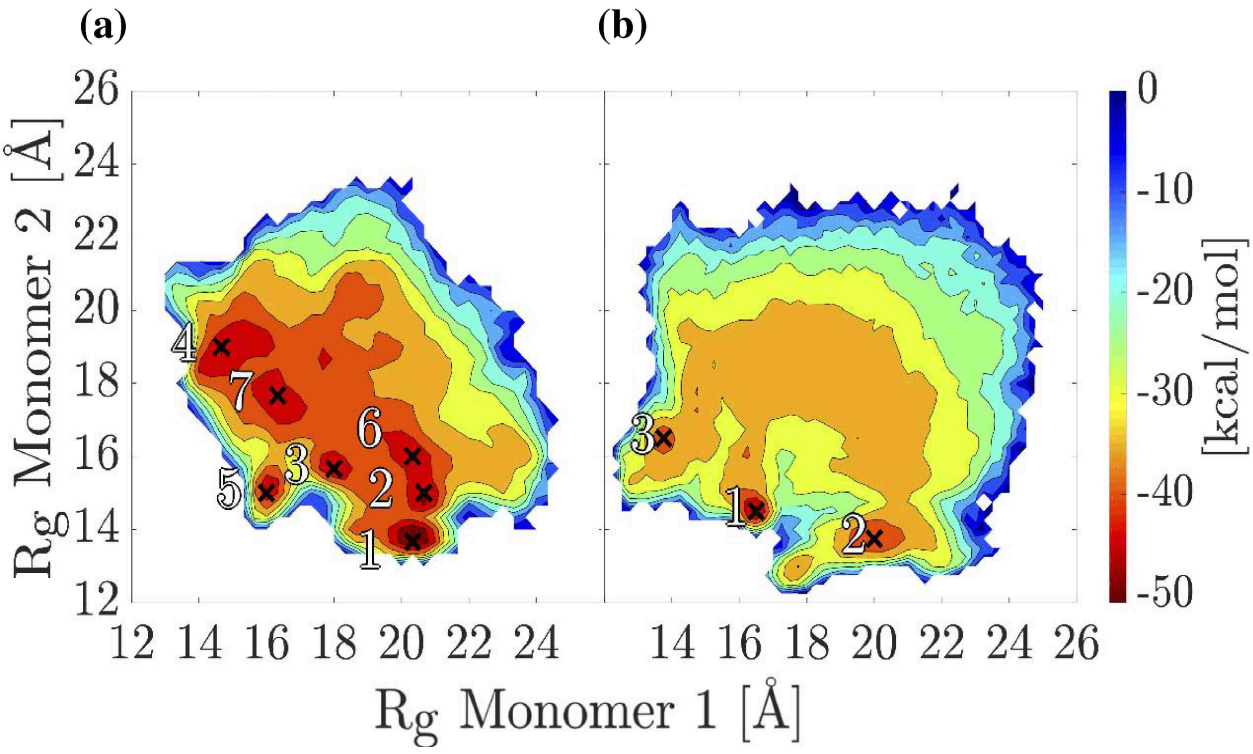

Figure 3.5:: Free energy landscapes of the αC-domain dimer projected onto the radii of gyration,$R_{g\,1}$ and $R_{g\,2}$.

(a) T-REMD simulations of the unoxidized αC-domain dimer. The seven deepest minima are labeled. **(b)** T-REMD simulations of the oxidized αC-domain dimer. The three deepest minima are labeled. Minima are labeled by crosses.

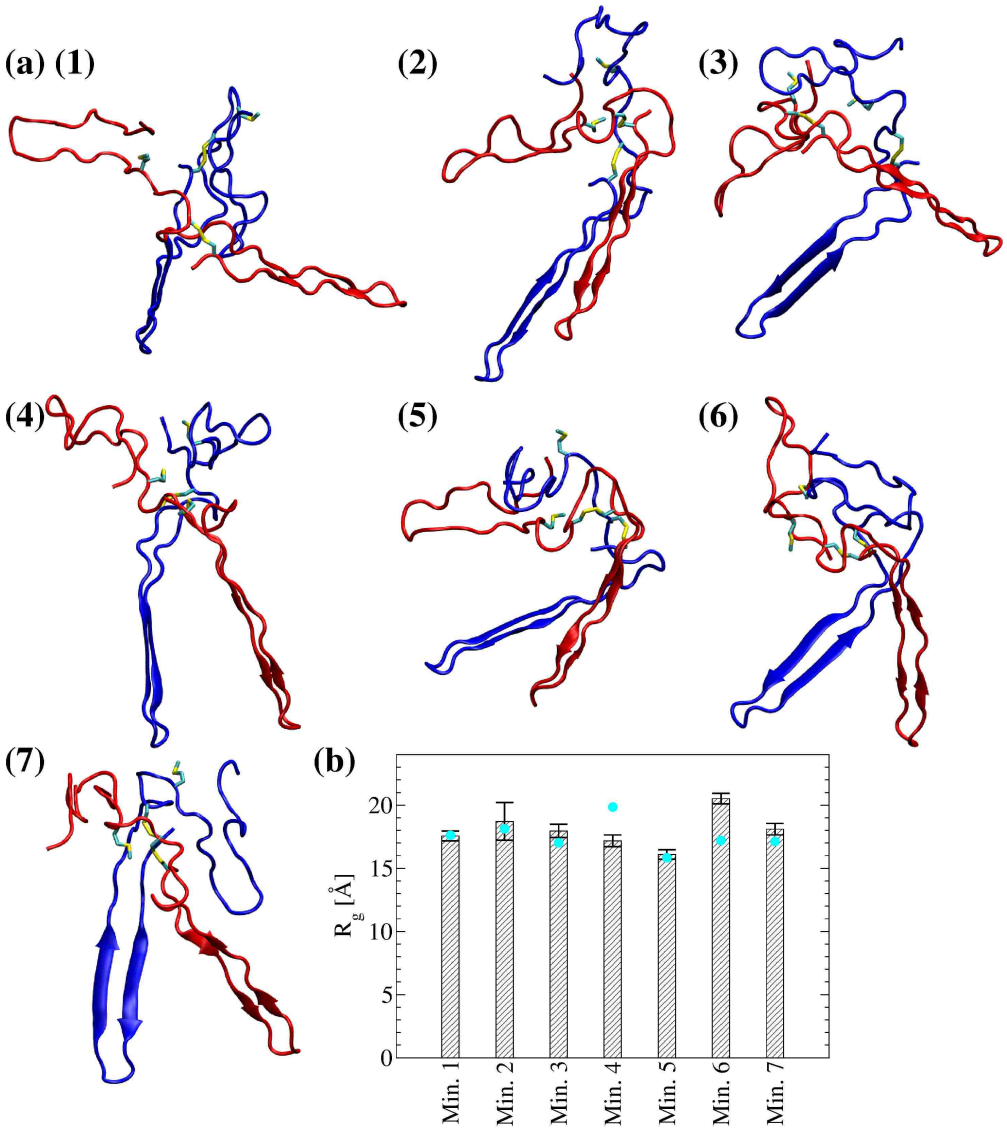

Figure 3.6:: Representative dimer conformations from the local minima of the FEL shown in Figure 3.5a.

(a-g) Met^{476} and the disulfide bond Cys^{442}-Cys^{472} are shown in the stick and ball representation. Monomer 1 is colored in blue and monomer 2 is colored in red. **(h)** The radius of gyration of the dimer computed from the last 40 ns of the in total 50-ns simulations. Cyan circles correspond to the radius of gyration of the dimer of the representative frame in each bin. Error bars represent the standard deviation.

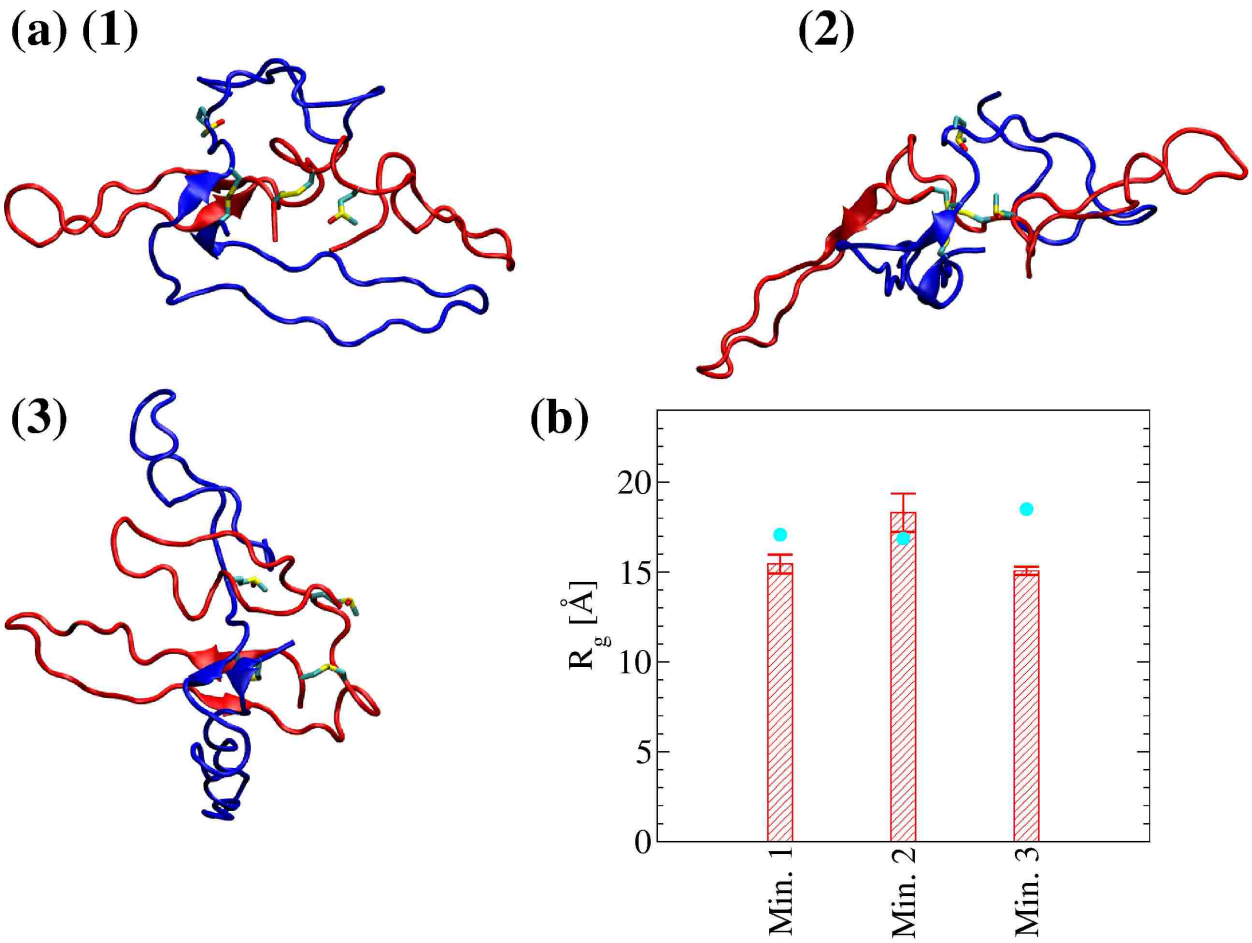

Figure 3.7:: Representative dimer conformations from the local minima of the FEL shown in Figure 3.5b.

(a-c) Met(O)^{476} and the disulfide bond Cys^{442}-Cys^{472} are shown in stick and ball representation. Monomer 1 is colored in blue and monomer 2 is colored in red. **(d)** The radius of gyration of the dimer computed from the last 40 ns of the in total 50-ns simulations. Cyan circles correspond to the radius of gyration of the dimer of the representative frame in each bin. Error bars represent the standard deviation.

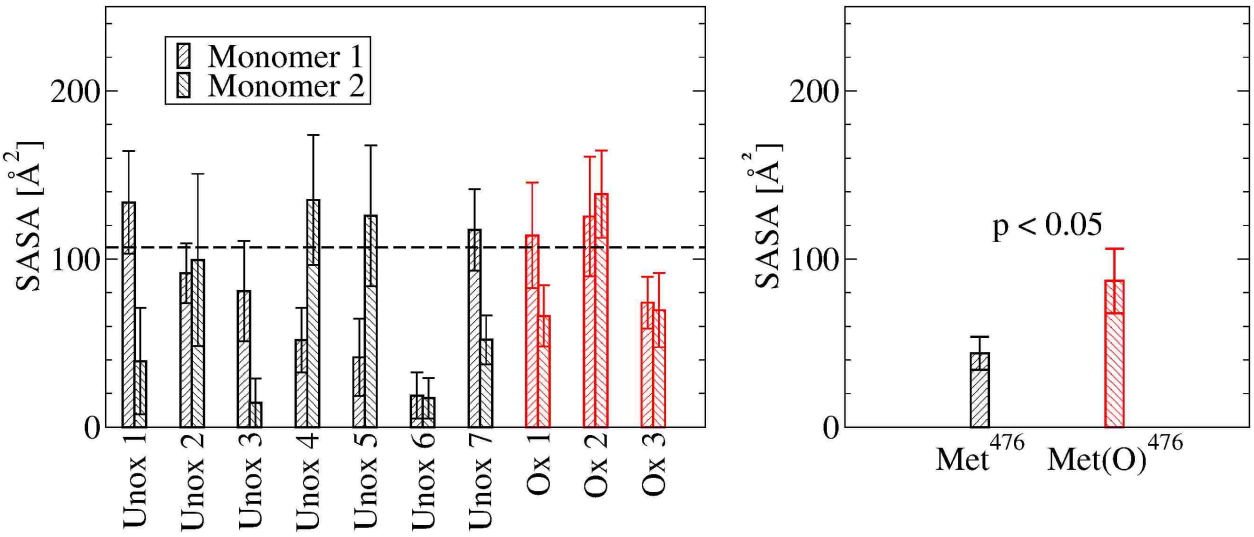

Figure 3.8:: Solvent accessible surface area (SASA) of each Met^{476} residues in the explicit solvent simulations started from the energy minima.

(a) SASA of each individual Met^{476}. The values were computed from the last 40 ns of the in total 50-ns simulations. The horizontal dashed line indicates the cutoff value of 106.88 Å^2 below which a side chain was considered partially buried. This value was calculated as 60% of the SASA of Met^{476} in the unfolded state of the monomeric αC domain. This was obtained by averaging the SASA of Met^{476} along the last 40 ns of an in total 50-ns long explicit-solvent simulation with a tripeptide containing Met^{476} flanked by Ala^{475} and Asp^{477}. **(b)** Average SASA of the most buried of the two Met^{476} in each explicit solvent simulation. Error bars in **(a)** represent the standard deviation and in **(b)** the standard error of the mean, respectively. A difference was considered statistically significant if the p-value (calculated from a one-tailed Student's t-test) was smaller than 0.05 (indicated in the figure).

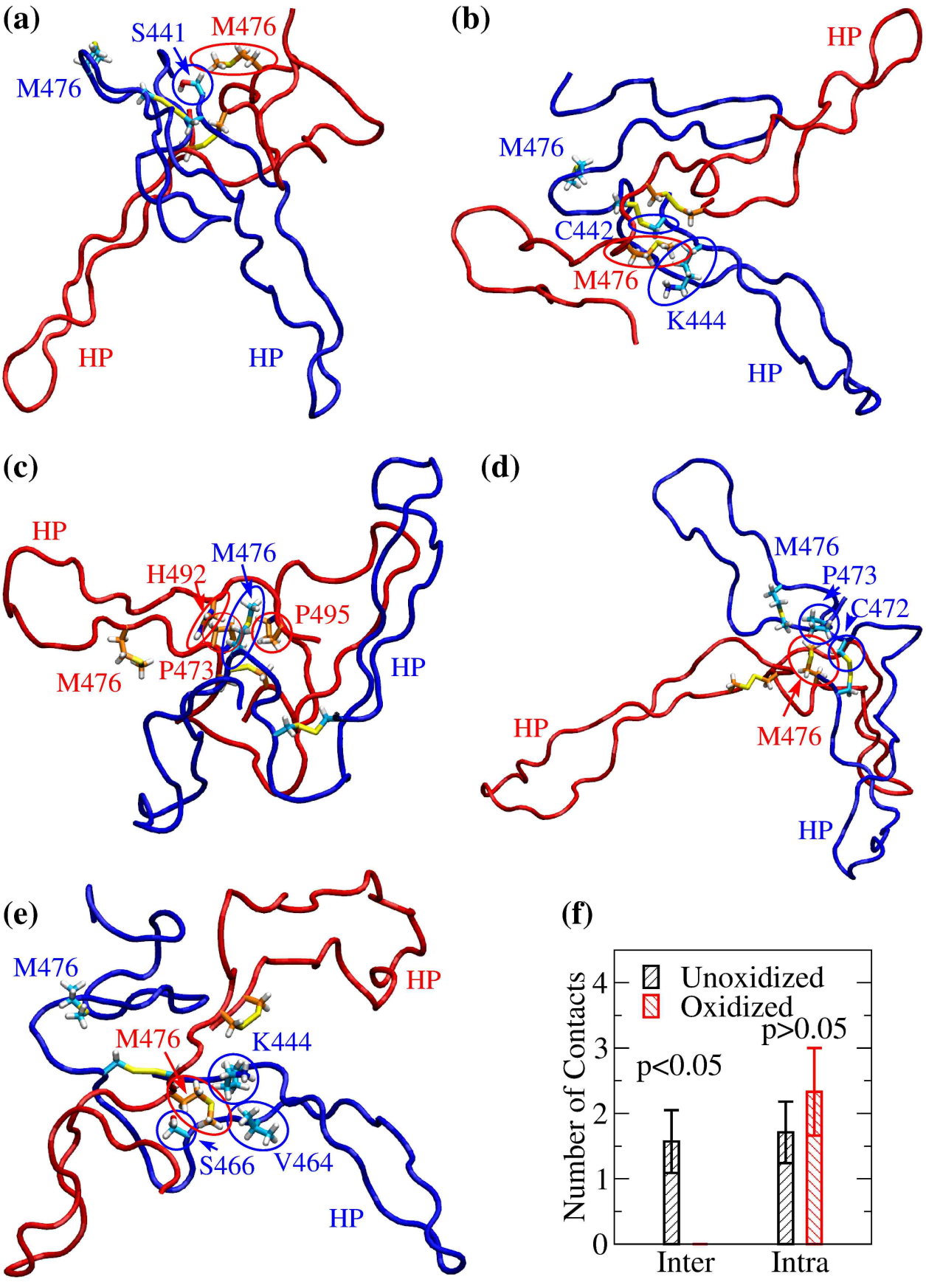
(a)
S441
M476
M476
HP
HP
(b)
HP
M476
C442
M476
K444
HP
(c)
HP
M476
H492
P495
M476
P473
HP
(d)
M476
P473
C472
HP
M476
HP
(e)
M476
HP
M476
K444
S466
V464
HP
(f)
Unoxidized
Oxidized
p>0.05
p<0.05
Number of Contacts
0
1
2
3
4
Inter
Intra

Figure 3.9:: Visualization of interdomain side chain contacts involving Met^{476}. Side chains involved in persistent interdomain contacts with Met^{476} in the explicit solvent simulation started from **(a)** minimum 1, **(b)** minimum 3, **(c)** minimum 5, **(d)** minimum 6, and **(e)** minimum 7, respectively (Table 3.2). The conformations shown correspond to snapshots sampled after 10 ns in each respective simulation, except for the snapshot in **(c)**, which was sampled after 29 ns in the simulation started from minimum 5. For distinction, monomer 1 is colored in blue and monomer 2 in red. Side chains are shown in the stick and ball representation and labeled. The respective carbon atoms are colored either in cyan (if belonging to monomer 1) or in orange (if belonging to monomer 2). Side chains from either monomer 1 or monomer 2 involved in interdomain contacts are highlighted by blue or red circles, respectively. The Met^{476} side chain not involved in interdomain contacts is also indicated and labeled but not circled. The location of the β-hairpin is indicated by the label "HP" colored accordingly. **(f)** Average number of inter- and intradomain contacts involving Met^{476} in the simulations started from the minima of the FEL of unoxidized and oxidized dimer, respectively. Error bars indicate standard errors of the mean. A difference was considered statistically significant if the p-value (calculated from a one-tailed Student's t-test) was smaller than 0.05 (indicated in the figure).

3.8 Supplemental information

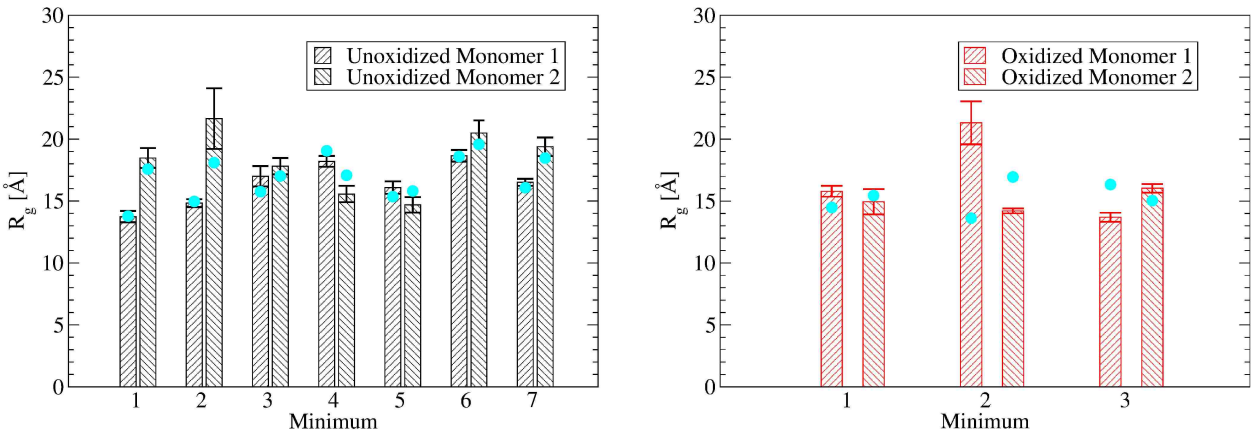

Figure S3.1:: The radius of gyration of each monomer.

(a) The radius of gyration of each monomer in the unoxidized dimer computed from the last 40 ns of the 50-ns simulations. **(b)** The radius of gyration of each monomer in the oxidized dimer computed from the last 40 ns of the 50-ns simulations. Cyan circles correspond to the radius of gyration of the dimer of the representative frame in each bin. Error bars represent the standard deviation.

Chapter 4

Free energy calculations.

4.1 Introduction

MD simulations can be used to calculate free energy differences between states, which is important for understanding chemical reactions, binding affinities, and protein dynamics. Two schools of thought can be used to perform free energy calculations. Work described in Chapter 3 exemplifies the "density-of-states" approach, in which a large conformational space is sampled and degenerate states indicate low-energy conformers. In contrast, FEP uses a "work" approach, which uses short trajectories to sample the paths between states and estimate $\Delta\Delta G^{ox}$, which describes the change in free energy of dimerization upon oxidation [120]. In this chapter, I detail how FEP is implemented and how its results bolster the argument that Met^{476} oxidation inhibits αC-domain dimerization.

4.2 Materials and Methods.

The change in the free energy of dimerization due to the oxidation of a methionine residue was estimated using alchemical transformations [89] in combination with the thermodynamic cycle [90] represented in Figure 4.1 in a similar manner to the FEP calculations performed in Interlandi, 2018 [118]. The alchemical transformations were performed through FEP calculations [121]. The seven representative frames — the centroids of the largest cluster

— from each minimum in the FEL were simulated for 50 ns in explicit solvent to assess stability of the conformations in the unoxidized state. The 40 ns, 45 ns and 50 ns frames were extracted from each of the seven trajectories and subject to alchemical transformations in the forward and backward direction in explicit solvent. In the forward transformation, Met side chain is slowly converted to Met(O) which contains an oxygen atom doubly covalently bonded to the sulfur atom. The conformation achieved in the forward transformation is then used to initialize the backward transformation where Met(O) is converted back to Met. The amount of work required for each transformation is calculated regularly during the simulations. The forward and backward calculations were then combined and a value for the ΔG_{alch} of the oxidation reaction was obtained using the Bennett's acceptance ratio method [91] implemented in the ParseFEP plugin for VMD [105]. Each forward and backward transformation was performed for 1 ns during the parameter, λ, was incrementally varied from 0 (Met) to 1 (Met(O)) and from 1 to 0, respectively in time intervals of 0.025 ns for a total of 40 intermediate states. The first half of the each time window involved equilibration and the second half data collection. A soft core term was introduced to the van der Waals potential to avoid singularities [122]. Similar ΔG values were obtained when the alchemical transformation was performed for only 1 ns and with time intervals of 0.025 ns, indicating that the side chain has likely thoroughly sampled its local environment in the nanosecond time scale. Alchemical transformations were performed in the dimer and monomeric states. The latter was approximated by the situation where a methionine is fully solvent exposed, which was accomplished by using an Ala-Met-Asp tripeptide, mimicking the Met^{476} local environment. By considering the thermodynamic cycle Figure 4.1, the difference in the free energy of dimerization upon methionine oxidation can be approximated by the difference between the ΔG values calculated from the alchemical simulations in the dimer and monomer states.

4.3 Results.

Representative frames from each minima in the unoxidized simulations were subject to explicit solvent simulations to assess stability Table 3.1. Unbiased selection of frames from those simulations were then used to estimate the free energy change of dimerization upon oxidation by FEP, denoted by $\Delta\Delta G^{ox}_{FEP}$. ΔG^{1}_{alch} values are shown in Figure 4.2 by black and red bars. ΔG^{2}_{alch} is also reported in Figure 4.2 in blue. ΔG^{1}_{alch} values that are above the blue dashed line correspond to $\Delta\Delta G^{ox}_{FEP} > 0$ since $\Delta\Delta G^{ox}_{FEP} = \Delta G^{2}_{alch} - \Delta G^{1}_{alch}$. The mean free energy change due to oxidation, $\Delta\Delta G^{ox}_{FEP}$, is 1.74 ± 1.24 kcal/mol excluding those without buried Met/Met(O) residues. A residue is considered buried if the mean SASA over 40 of the 50-ns explicit solvent simulations is less than 106.88 Å^2 (Figure 3.8). This result indicates that oxidation makes dimerization less favorable.

4.4 Discussion.

Met^{476} oxidation is predicted to disrupt αC polymerization and I performed FEP with BAR to assess $\Delta\Delta G_{ox}$. Work described in Chapter 2 highlights the importance of a hydrophobic core to nucleate αC polymerization. Using FEP, I demonstrate that the hydrophobic core can be disrupted by the introduction of Met(O)^{476}. The energy associated with hydrogen bonds is 1–3 kcal/mol [123], which is comparable to $\Delta\Delta G^{ox}_{FEP}$. A rough estimation of $\Delta\Delta G^{ox}_{FEL}$ from the FELs calculated in Chapter 2 was approximately 6.2 kcal/mol. Assuming that the monomeric unoxidized and oxidized states have approximately the same energy when solvated, $\Delta\Delta G^{ox}_{FEL}$ and $\Delta\Delta G^{ox}_{FEP}$ can be reduced to $\Delta G^{ox}_{FEL} = 6.2$ kcal/mol and $\Delta G^{ox}_{FEP} = 1.74$ kcal/mol, respectively. This energy approximates the solvation energies of Met and desolvation of Met(O) residues, and provides insight into the mechanism of impaired αC domain polymerization.

In this system, FEP samples the two endpoints of the oxidation reaction in the dimer and monomer cases but does not provide insight about the encounter complex. While the magnitude of $\Delta\Delta G^{ox}_{FEP}$ may be on the order of a single hydrogen bond, there are possible kinetic

effects that cannot be described using FEP. Figure 4.3 shows the conversion of FELs from unoxidized to oxidized αC domain. $\Delta\Delta G^{ox}_{polym}$, shown in Figure 4.3 is approximated by $\Delta\Delta G^{ox}_{FEL}$ and $\Delta\Delta G^{ox}_{FEP}$ using FELs and FEP, respectively. Ultimately, oxidation stabilizes both the monomeric and dimeric states, although the monomeric state is more favorable. The slight disagreement in magnitude between FEP and FEL results is likely negligible as both values are of the same order of magnitude. Most importantly, $\Delta\Delta G^{ox}_{FEP}$ and $\Delta\Delta G^{ox}_{FEL}$ are both positive, indicating that oxidation renders dimerization unfavorable. $\Delta\Delta G^{ox}$ correlates with the solvent exposure of Met^{476}/$Met(O)^{476}$, which is consistent with previous FEP studies of Met oxidation in VWF A2 domain [118]. Therefore, we can conclude that Met oxidation is unfavorable when Met is buried.

4.5 Figures.

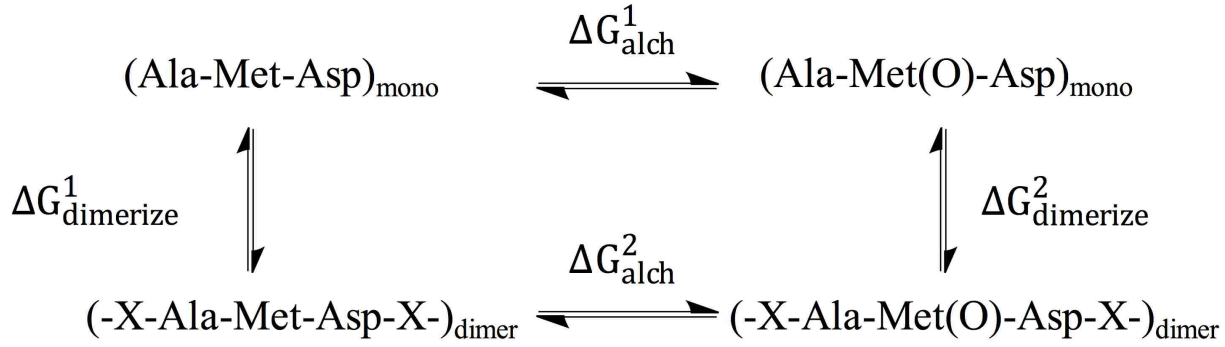

Figure 4.1:: Thermodynamic cycle used to estimate the change in free energy of dimerization upon oxidation of Met^{476}.

The horizontal reactions correspond to the conversion of Met^{476} to Met(O)^{476} in the dimerized state, respectively. The vertical reactions describe the dimerization in the unoxidized and oxidized state, respectively. $\Delta G^{1,2}_{alch}$ is calculated as described in the text. The change in free energy of dimerization can then be derived as follows: $\Delta\Delta G^{ox}_{\mathrm{FEP}} = \Delta G^2_{dimerize} - \Delta G^1_{dimerize} = \Delta G^2_{alch} - \Delta G^1_{alch}$

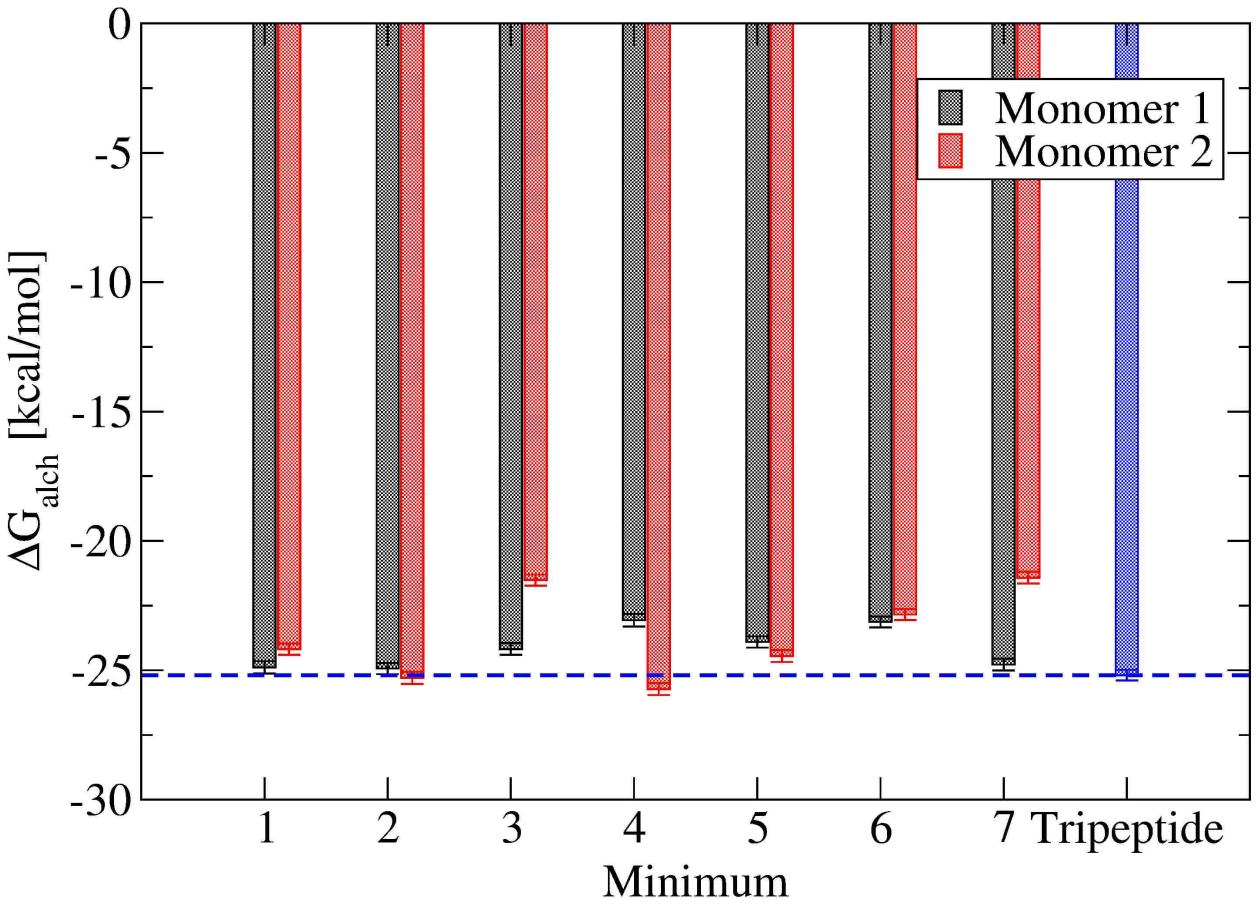

Figure 4.2:: Mean ΔG_{alch} for seven minima and a tripeptide

Black and red bars correspond to the mean ΔG^2_{alch} of each Met residue on monomers 1 and 2, respectively across three different configurations. The blue bar and blue dashed line correspond to the mean ΔG^1_{alch} across three different configurations. Error bars represent the variance [105].

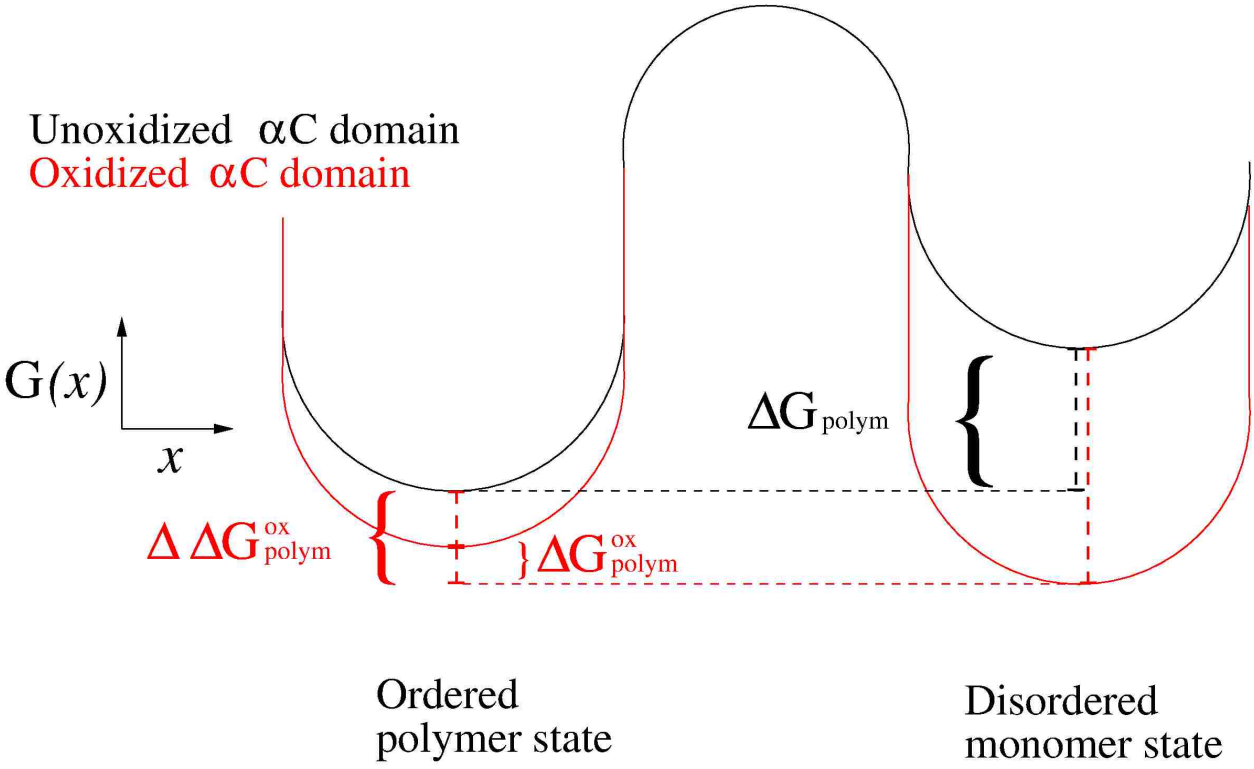

Figure 4.3:: Free energy diagram of αC polymerization

αC domain polymerization reaction. The reaction coordinate of polymerization, x, and the free energy at coordinate x, G(x), represent the x- and y-axes, respectively. The FEL of unoxidized αC domain (black) depicts ΔG_{polym}, which is the free energy of polymerization. Likewise, ΔG^{ox}_{polym} depicts the free energy change between monomeric and dimeric states of the oxidized αC domain. The FEL of the oxidized αC domain (red) shows stabilized monomeric and dimeric states. $\Delta\Delta G^{ox}_{polym}$ describes the energy difference between the unoxidized dimeric state and the oxidized monomeric state. Unoxidized αC domain polymerization is cooperative and ΔG_{polym} decreases by -6.7 kcal/mol per monomer [4], which means the energy well of the polymer state decreases as the polymer grows. The FEL of the unoxidized αC domain is dynamic, although this is not explicitly depicted. In contrast, oxidized αC domain dimerization is not cooperative.

Chapter 5

Conclusions.

5.1 Identification of dominant subpopulations of peptide conformers using enhanced sampling methods.

In Chapter 2, I describe, in detail, a process of identifying low-energy dimer conformers of flexible peptides. A T-REMD simulation was used to sample dimer conformations and GBSA implicit solvation was used to improve sampling, due to the lack of friction of water, and reduce the number of required replicas. The R_g of each monomer were used as the CVs to construct a 2-D histogram, which was converted into a FEL using the equations described in Section 2.2.6.2. The structures that comprised the lowest bins of the deepest wells were hierarchically clustered and the representative frames of the largest clusters were identified [4]. The representative frames were subject to explicit solvent simulations to assess dimer stability and to probe for artefact introduction by implicit solvation. This process can be implemented with metadynamics or aMD simulations as long as they converge and can be used to produce a FEL.

5.1.1 Comparison to rigid-body docking

Enhanced-sampling-based methods are computationally intensive by comparison to FFT-based methods like ZDOCK, but they are designed to sample large swaths of conformational

space. These methods may provide a theoretical advantage to RBD when dealing with intrinsically disordered regions (IDR), proteins with unknown flexibility or proteins that undergo large conformational changes upon binding.

5.2 The hydrophobic core model of αC polymerization: a new understanding of the αC domain.

The work described in Chapter 2 proposes a new model of the αC polymer that predicts amorphous binding modes between monomers (the hydrophobic core (HC) model). The HC model predicts that Met^{476} acts a docking spot by nucleating a hydrophobic patch in aqueous solution, which allows for the binding of two αC monomers. In the majority of the dimer models, we observe that only one methionine residue is buried while the other is exposed. The exposed Met^{476} in the dimer likely serves as a docking spot to recruit an additional αC-domain monomer. Each iteration would further stabilize and grow the αC polymer [4, 15].

Some intrinsically disordered proteins can undergo a transition from a disordered state to an ordered state upon binding to another protein. Proteins that only differ between ordered and disordered states by 5–15 kcal/mol, like the unoxidized αC region, may exhibit a "fly-casting" mechanism which allows them to fold and subsequently interact with another disordered partner to form a more stable complex [124]. The αC domain lies in an IDR, and becomes more flexible upon oxidation [1]. αC domains likely exhibit this disordered-to-ordered transition upon binding [4, 15], however, if the oxidized αC domain has a $\Delta G_{polym} < 5$ kcal/mol, simultaneous folding and polymerization is unlikely [124]. Free energy calculations described in Chapter 4 indicate a substantial loss in ΔG_{polym}, which is due to a small stabilization of the oxidized dimer state due to the introduction of hydrogen bonding. However, oxidation stabilizes the unbound state even more (Figure 4.3). The addition of the oxygen reduces the number of available hydrophobic residues capable of forming a HC, which stabilizes the energy well of the monomeric oxidized state, driving the equilibrium towards the monomeric

state (Figure 5.1) . The equilibration time for oxidized αC polymerization would likely far slower than the degradation rate of fibrin *in vivo* and would therefore never be sampled under physiological conditions. The HC model is consistent with results from simulations and previous literature on Met^{476} oxidation and fibrin clot morphology [1, 3, 39, 60].

5.3 Summary

Fibrinogen is one of the principle components of the coagulation cascade and regulates inflammation. Too much inflammation can induce oxidation of fibrin molecules, which leads to coagulopathy, thrombosis and bleeding. Despite the lack of structural knowledge of the αC domain, I performed MD simulations and free energy calculations to understand how oxidation affects αC domain dynamics. The work described in this dissertation aimed to resolve the energetic and structural consequences of αC domain oxidation and to further the current understanding of flexible regions in proteins. By using enhanced sampling MD techniques, I developed a new process of identifying low-energy conformers from constructed FELs that is more equipped to handle protein flexibility than currently-available RBD programs. Using this technique and free energy calculations, I propose a thermodynamic model, the HC model, by which oxidation disrupts the HC of the αC polymer, favoring the monomer state. Prior to this work, the common hypothesis was that the αC polymer is driven by amyloid-like interactions. I demonstrate that the αC-domain dimer interface is highly amorphous and, like many protein-protein interactions, is driven by the hydrophobic effect rather than hydrogen bonds from β-sheet stacking.

5.4 Future Directions

The *in silico* work presented in this dissertation requires further testing and experimental validation. The computational process to identify flexible conformers should be applied to other peptides with well-characterized dynamics. The effects of selecting different CVs for analysis on peptide identification and reproducibility should also be further examined.

There is significant value in probing the accuracy of the HC model using biophysical and biochemical experiments. Fluorescence correlation spectroscopy (FCS), as one example, measures the diffusion times of proteins, which is correlated to their size (and number of binding partners). FCS could be used to study the relative subpopulations of the monomer and polymer states before and after oxidation in a way to probe the HC model. Oxidomimetic mutations such as $Met^{476}Gln$ could be used in lieu of HOCl treatment, which is non-specific, due to the similar side chain chemistry (Figure 5.2).

5.5 Figures.

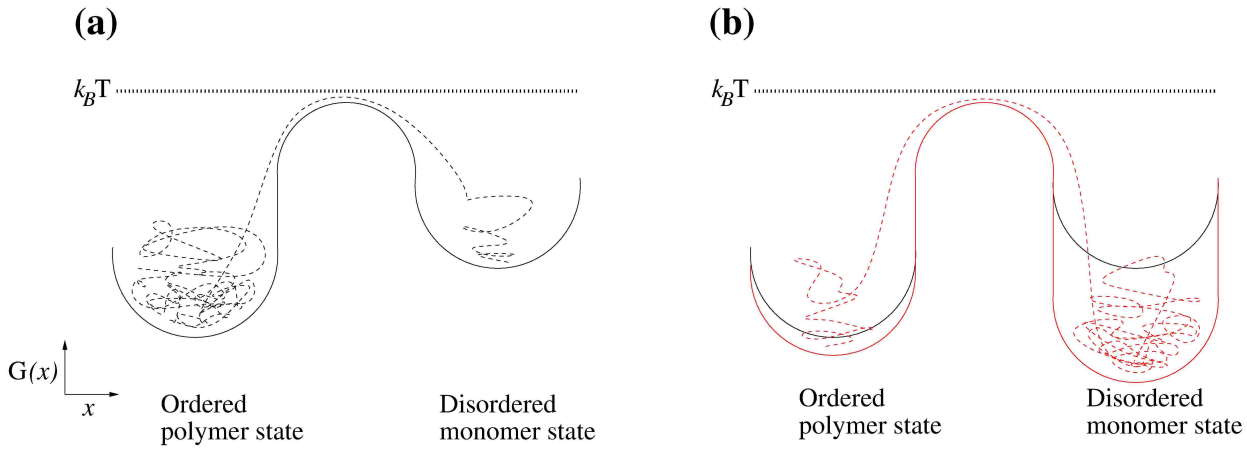

Figure 5.1:: Free energy diagram with trajectories of αC domain polymerization

The reaction coordinate of polymerization, x, and the free energy at coordinate x, G(x), represent the x- and y-axes, respectively. **(a)** Free energy diagram of unoxidized αC polymerization (solid black lines). Black dotted lines represent a theoretical trajectory of an αC domain as it samples the polymer and monomer states and crosses the energy barrier. **(b)** Free energy diagram of oxidized αC polymerization (solid red lines). Red dotted lines represent a theoretical trajectory of an αC domain that thermodynamically favors the monomeric state as oxidation slightly stabilizes the dimeric state but stabilizes the monomeric state even more.

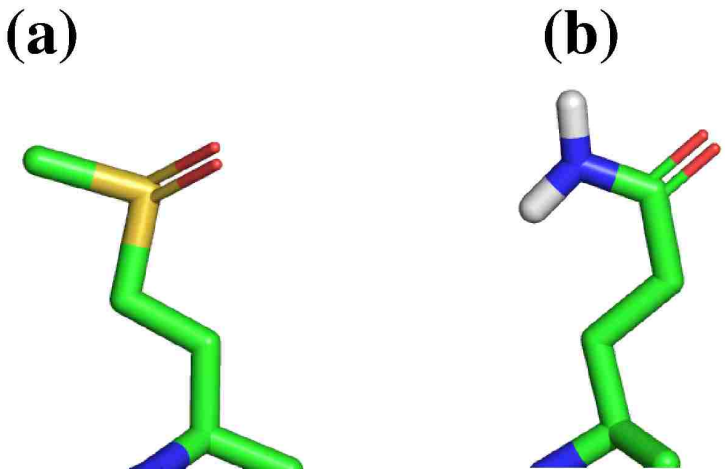

Figure 5.2:: Structural similarity between Met(O) and Gln residues

(a) The methionine sulfoxide side chain. **(b)** The glutamine side chain.

Bibliography

[1] K. M. Weigandt, N. White, D. Chung, E. Ellingson, Y. Wang, X. Fu, and D. C. Pozzo, "Fibrin Clot Structure and Mechanics Associated with Specific Oxidation of Methionine Residues in Fibrinogen," *Biophysical Journal*, vol. 103, pp. 2399–2407, 2012.

[2] J. W. Weisel and R. I. Litivinov, "Fibrin formation, structure and properties," *Subcell Biochem.*, vol. 82, pp. 405–456, 2017.

[3] J. W. Weisel and R. I. Litivinov, "Mechanisms of fibrin polymerization and clinical implications," *Blood*, vol. 121, no. 10, pp. 1712–1719, 2013.

[4] G. Tsurupa, R. R. Hantgan, R. A. Burton, I. Pechik, N. Tjandra, and L. Medved, "Structure, Stability, and Interaction of the Fibrin(ogen) αC-Domains," *Biochemistry*, vol. 48, no. 51, pp. 12191–12201, 2009.

[5] J. W. Weisel and L. Medved, "The Structure and Function of the αC Domains of Fibrinogen," *Annals New York Academy of Sciences*, vol. 936, pp. 312–327, 2001.

[6] Y. Fu, L. Weissbach, P. W. Plant, C. Oddoux, Y. Cao, T. J. Liang, S. N. Roy, C. M. Redman, and G. Grienenger, "Carboxy-terminal-extended variant of the human fibrinogen α subunit: a novel exon conferring marked homology to β and γ subunits," *Biochemistry*, vol. 31, no. 48, pp. 11968–11972, 1992.

[7] "UniProtKB - P02671 - FGA, `www.uniprot.org/uniprot/P02671`," 2019.

[8] D. Applegate, L. S. Steben, K. M. Hertzberg, and G. Grienenger, "The α_{E}C domain of human fibrinogen-420 is a stable and early plasmin cleavage product," *Blood*, vol. 95, pp. 2297–2303, 2000.

[9] M. P. M. de Maat and M. Verschuur, "Fibrinogen heterogeneity: inherited and noninherited," *Current Opinion in Hematology*, vol. 12, no. 5, 2005.

[10] B. Blombäck, "Fibrinogen: Evolution of the Structure-Function Concept," Fibrinogen 2000 Congress, Ann NY Aca Of Sci., 2000.

[11] P. L. Privalov and L. V. Medved, "Domains in the fibrinogen molecule," *Journal of Molecular Biology*, vol. 159, no. 4, pp. 665–683, 1982.

[12] J. J. Marsh, H. S. Guan, S. Li, P. G. Chiles, D. Tran, and T. A. Morris, "Structural Insights into Fibrinogen Dynamics Using Amide Hydrogen/Deuterium Exchange Mass Spectrometry.," *Biochemistry*, vol. 52, no. 32, pp. 5491–5502, 2013.

[13] J. M. Kollman, L. Pandi, M. R. Sawaya, M. Riley, and R. F. Doolittle, "Crystal Structure of Human Fibrinogen," *Biochemistry*, vol. 48, no. 18, pp. 3877–3886, 2009.

[14] R. A. Burton, G. Tsurupa, R. R. Hantgan, N. Tjandra, and L. Medved, "NMR Solution Structure, Stability and Interaction of the Recombinant Bovine Fibrinogen alphaC-Domain Fragment.," *Biochemistry*, vol. 46, no. 29, pp. 8550–8560, 2007.

[15] G. Tsurupa, I. Pechik, R. I. Litvinov, R. R. Hantgan, N. Tjandra, J. W. Weisel, and L. Medved, "On the Mechanism of αC Polymer Formation in Fibrin," *Biochemistry*, vol. 51, no. 12, pp. 2526–2538, 2015.

[16] A. Zhmurov, A. D. Protopopova, R. I. Litvinov, P. Zhukov, J. W. Weisel, and V. Barsegov, "Atomic Structural Models of Fibrin Oligomers," *Structure*, vol. 26, pp. 857–868, Jun 2018.

[17] R. A. Burton, G. Tsurupa, L. Medved, and N. Tjandra, "Identification of an Ordered Compact Structure within the Recombinant Bovine Fibrinogen αC-Domain Fragment by NMR," *Biochemistry*, vol. 45, no. 7, pp. 2257–2266, 2006.

[18] M. H. Periayah, A. S. Halim, and A. Z. Mat Saad, "Mechanism Action of Platelets and Crucial Blood Coagulation Pathways in Hemostasis," *International journal of hematology-oncology and stem cell research*, vol. 11, no. 4, pp. 319–327, 2017.

[19] Y. Nemerson and D. Repke, "Tissue factor accelerates the activation of coagulation factor VII: The role of a bifunctional coagulation cofactor," *Thrombosis Research*, vol. 40, no. 3, pp. 351–358, 1985.

[20] Y. Kawamoto and M. Kaibara, "Procoagulant activity of collagen. Effect of difference in type and structure of collagen," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1035, no. 3, pp. 361–368, 1990.

[21] S. A. Smith, N. J. Mutch, D. Baskar, P. Rohloff, R. Docampo, and J. H. Morrissey, "Polyphosphate modulates blood coagulation and fibrinolysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 903–908, 01 2006.

[22] F. Müller, N. J. Mutch, W. A. Schenk, S. A. Smith, L. Esterl, H. M. Spronk, S. Schmidbauer, W. A. Gahl, J. H. Morrissey, and T. Renné, "Platelet polyphosphates are proinflammatory and procoagulant mediators *in vivo*," *Cell*, vol. 139, pp. 1143–1156, 12 2009.

[23] J. W. Suttie and C. M. Jackson, "Prothrombin structure, activation and biosynthesis," *Physiological Reviews*, vol. 57, no. 1, 1997.

[24] M. Hoffman and D. M. I. Monroe, "A Cell-based Model of Hemostasis," *Thrombosis and Haemostasis*, vol. 85, no. 6, pp. 958–965, 2001.

[25] S. Palta, R. Saroa, and A. Palta, "Overview of the coagulation system," *Indian journal of anaesthesia*, vol. 58, no. 5, pp. 515–523, 2014.

[26] W. P. Fay, N. Garg, and M. Sunkar, "Vascular functions of the plasminogen activation system," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 27, no. 6, pp. 1231–1237, 2007.

[27] F. Blasi, "Urokinase and urokinase receptor: A paracrine/autocrine system regulating cell migration and invasiveness," *BioEssays*, vol. 15, pp. 105–111, 2019/03/19 1993.

[28] P. J. Gaffney, "Breakdown products of fibrin and fibrinogen: molecular mechanisms and clinical implications," *Journal of Clinical Pathology*, vol. 33, no. 14, pp. 10–17, 1980.

[29] G. Cesarman-Maus and K. A. Hajjar, "Molecular mechanisms of fibrinolysis," *British Journal of Haematology*, vol. 129, pp. 307–321, 2019/03/19 2005.

[30] O. Soehnlein, S. Steffens, A. Hidalgo, and C. Weber, "Neutrophils as protagonists and targets in chronic inflammation," *Nature Reviews Immunology*, vol. 17, p. 248, 2017.

[31] S. Caielli, J. Banchereau, and V. Pascual, "Neutrophils come of age in chronic inflammation," *Current opinion in immunology*, vol. 24, pp. 671–677, 12 2012.

[32] H. R. Jones, C. T. Robb, M. Perretti, and A. G. Rossi, "The role of neutrophils in inflammation resolution," *Seminars in Immunology*, vol. 28, no. 2, pp. 137–145, 2016. SI: Neutrophils in innate and adaptive immunity.

[33] P. Kruger, M. Saffarzadeh, A. N. R. Weber, N. Rieber, M. Radsak, H. von Bernuth, C. Benarafa, D. Roos, J. Skokowa, and D. Hartl, "Neutrophils: Between Host Defence, Immune Modulation, and Tissue Injury," *PLoS Pathogens*, vol. 11, pp. 1–22, 03 2015.

[34] D. Odobasic, A. R. Kitching, and S. R. Holdsworth, "Neutrophil-Mediated Regulation of Innate and Adaptive Immunity: The Role of Myeloperoxidase," *Journal of immunology research*, vol. 2016, p. 2349817, 2016.

[35] K. D. Metzler, T. A. Fuchs, W. M. Nauseef, D. Reumaux, J. Roesler, I. Schulze, V. Wahn, V. Papayannopoulos, and A. Zychlinsky, "Myeloperoxidase is required for neutrophil extracellular trap formation: implications for innate immunity," *Blood*, vol. 117, no. 3, pp. 953–959, 2011.

[36] D. Kutter, P. Devaquet, G. Vanderstocken, J. M. Paulus, V. Marchal, and A. Gothot, "Consequences of total and subtotal myeloperoxidase deficiency: Risk or benefit ?," *Acta Haematologica*, vol. 104, no. 1, pp. 10–15, 2000.

[37] H. Rosen, S. J. Klebanoff, Y. Wang, N. Brot, J. W. Heinecke, and X. Fu, "Methionine oxidation contributes to bacterial killing by the myeloperoxidase system of neutrophils.," *Proc Natl Acad Sci USA.*, vol. 106, no. 44, pp. 18686–18691, 2009.

[38] C. Achilli, A. Ciana, A. Rossi, C. Balduini, and G. Minetti, "Neutrophil granulocytes uniquely express, among human blood cells, high levels of Methionine-sulfoxide-reductase enzymes," *Journal of Leukocyte Biology*, vol. 83, pp. 181–189, 2019/04/03 2008.

[39] P. R. Burney, N. White, and J. Pfaendtner, "Structural Effects of Methionine Oxidation on Isolated Subdomains of Human Fibrin D and αC Regions," *PLoS ONE*, vol. 9, no. 1, p. e86981, 2014.

[40] C. N. Serhan, "Pro-resolving lipid mediators are leads for resolution physiology," *Nature*, vol. 510, pp. 92–101, 06 2014.

[41] C. N. Serhan, N. Chiang, and T. E. Van Dyke, "Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators," *Nature reviews. Immunology*, vol. 8, pp. 349–361, 05 2008.

[42] A. Durante and S. Bronzato, "The increased cardiovascular risk in patients affected by autoimmune diseases: review of the various manifestations," *Journal of clinical medicine research*, vol. 7, pp. 379–384, 06 2015.

[43] C. Jennewein, N. Tran, P. Paulus, P. Ellinghaus, J. A. Eble, and K. Zacharowski, "Novel aspects of fibrin(ogen) fragments during inflammation," *Molecular medicine (Cambridge, Mass.)*, vol. 17, pp. 568–573, May-Jun 2011.

[44] E. Gruys, M. J. M. Toussaint, T. A. Niewold, and S. J. Koopmans, "Acute phase reaction and acute phase proteins," *Journal of Zhejiang University. Science. B*, vol. 6, pp. 1045–1056, 11 2005.

[45] M. D. Turner, B. Nedjai, T. Hurst, and D. J. Pennington, "Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1843, no. 11, pp. 2563–2582, 2014.

[46] T. Tanaka, M. Narazaki, and T. Kishimoto, "IL-6 in inflammation, immunity, and disease," *Cold Spring Harbor perspectives in biology*, vol. 6, no. 10, p. a016295.

[47] A. Salsmann, E. Schaffner-Reckinger, F. Kabile, S. Plançon, and N. Kieffer, "A New Functional Role of the Fibrinogen RGD Motif as the Molecular Switch That Selectively Triggers Integrin αIIbβ3-dependent RhoA Activation during Cell Spreading," *Journal of Biological Chemistry*, vol. 280, pp. 33610—33619, June 2005.

[48] R. O. Hynes, "Integrins: Versatility, modulation, and signaling in cell adhesion," *Cell*, vol. 69, no. 1, pp. 11–25, 1992.

[49] N. Sándor, S. Lukácsi, R. Ungai-Salánki, N. Orgován, B. Szabó, R. Horváth, A. Erdei, and Z. Bajtay, "CD11c/CD18 Dominates Adhesion of Human Monocytes, Macrophages and Dendritic Cells over CD11b/CD18," *PLoS ONE*, vol. 11, p. e0163120, 09 2016.

[50] A. S. Antonov, G. N. Antonova, D. H. Munn, N. Mivechi, R. Lucas, J. D. Catravas, and A. D. Verin, "$\alpha_V\beta 3$ integrin regulates macrophage inflammatory responses via PI3 kinase/Akt-dependent NF-κB activation," *Journal of cellular physiology*, vol. 226, pp. 469–476, 02 2011.

[51] M. M. Aleman, B. L. Walton, J. R. Byrnes, and A. S. Wolberg, "Fibrinogen and red blood cells in venous thrombosis," *Thrombosis research*, vol. 133 Suppl 1, pp. S38–S40, 05 2014.

[52] L. Ping, L. Huang, B. Cardinali, A. Profumo, O. V. Gorkun, and S. T. Lord, "Substitution of the human αC-region with the analogous chicken domain generates a fibrinogen with severely impaired lateral aggregation: fibrin monomers assemble into protofibrils but protofibrils do not assemble into fibers," *Biochemistry*, vol. 50, no. 42, pp. 9066–9075, 2011.

[53] L. I. Påhlman, M. Mörgelin, G. Kasetty, A. I. Olin, A. Schmidtchen, and H. Herwald, "Antimicrobial activity of fibrinogen and fibrinogen-derived peptides –a novel link between coagulation and innate immunity," *Thrombosis and Haemostasis*, vol. 109, no. 05, pp. 930–939, 2013.

[54] C. T. Esmon, J. Xu, and F. Lupu, "Innate immunity and coagulation," *Journal of Thrombosis and Haemostasis*, vol. 9, pp. 182–188, 2011.

[55] J.-P. Collet, J. L. Moen, Y. I. Veklich, O. V. Gorkun, S. T. Lord, G. Montalescot, and J. W. Weisel, "The αC domains of fibrinogen affect the structure of the fibrin clot, its physical properties, and its susceptibility to fibrinolysis," *Blood*, vol. 106, no. 12, pp. 3824–3830, 2005.

[56] T. Madden, *The NCBI Handbook*, ch. 16: The BLAST Sequence Analysis Tool. Bethesda, MD: National Center for Biotechnology Information, 2003.

[57] H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, and R. Lopez, "Analysis Tool Web Services from the EMBL-EBI," *Nucleic acids research*, vol. 41, pp. W597–600, July 2013.

[58] S. Miller, J. Janin, A. M. Lesk, and C. Chothia, "Interior and surface of monomeric proteins," *Journal of Molecular Biology*, vol. 196, no. 3, pp. 641–656, 1987.

[59] T. Ohmura, T. Ueda, Y. Hashimoto, and T. Imoto, "Tolerance of point substitution of methionine for isoleucine in hen egg white lysozyme," *Protein Engineering, Design and Selection*, vol. 14, pp. 421–425, 06 2001.

[60] N. J. White, Y. Wang, X. Fu, J. C. Cardenas, E. J. Martin, D. F. Brophy, C. E. Wade, X. Wang, A. E. St. John, E. B. Lim, S. A. Stern, K. R. Ward, J. A. López, and D. Chung, "Post-translational oxidative modification of fibrinogen is associated with coagulopathy after traumatic injury," *Free Radicals Biology and Medicine*, vol. 96, pp. 181–189, 2016.

[61] B. Brooks, C. I. Brooks, A. J. Mackerell, L. Nilsson, R. Petreela, B. Roux, *et al.*, "CHARMM: The Biomolecular Simulation Program," *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.

[62] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, Jr., "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting

Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles," *Journal Chemical Theory and Computation*, vol. 8, pp. 3257–3273, Sep 2012.

[63] D. Gfeller, O. Michielin, and V. Zoete, "SwissSidechain: a molecular and structural database of non-natural sidechains," *Nucleic Acids Research*, vol. 41, no. D1, pp. E327–E332, 2013.

[64] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald - an N.log(N) method for Ewald sums in large systems," *Journal of Chemical Physics*, vol. 98, pp. 10089–10092, 1993.

[65] L. Kalé, R. Skeel, M. Bhandakar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten, "NAMD2: Greater Scalability for parallel molecular dynamics," *Journal of Computational Chemistry*, vol. 151, pp. 283–312, 1999.

[66] R. Elber, A. P. Ruymgaart, and B. Hess, "SHAKE parallelization," *The European physical journal. Special topics*, vol. 200, pp. 211–223, 11 2011.

[67] T. Schneider and E. Stoll, "Molecular Dynamics study of a 3-dimensional one-component model for distortive phase transitions," *Physical Review B*, vol. 17, no. 3, pp. 1302–1322, 1978.

[68] S. E. Feller, Y. H. Zhang, R. W. Pastor, and B. R. Brooks, "Constant-pressure molecular-dynamics simulations - the Langevin piston method," *Journal of Chemical Physics*, vol. 103, no. 11, pp. 4613–4621, 1995.

[69] J. Chen, r. Brooks, Charles L, and J. Khandogin, "Recent advances in implicit solvent-based methods for biomolecular simulations," *Current opinion in structural biology*, vol. 18, pp. 140–148, 04 2008.

[70] A. Onufriev, D. Bashford, and D. A. Case, "Modification of the Generalized Born Model Suitable for Macromolecules," *The Journal of Physical Chemistry B*, vol. 104, pp. 3712–3720, 04 2000.

[71] A. Onufriev, D. Bashford, and D. A. Case, "Exploring protein native states and large-scale conformational changes with a modified generalized born model," *Proteins: Structure, Function, and Bioinformatics*, vol. 55, pp. 383–394, 2019/04/10 2004.

[72] I. Samish, J. Gu, and M. L. Klein, *Structural Bioinformatics*, ch. 37 PROTEIN MOTION: SIMULATION, p. 922. John Wiley & Sons, Ltd, 2nd ed., 2011.

[73] J. Zhu, Y. Shi, and H. Liu, "Parametrization of a Generalized Born/Solvent-Accessible Surface Area Model and Applications to the Simulation of Protein Dynamics," *Journal of Physical Chemistry B*, vol. 106, no. 18, pp. 4844–4853, 2002.

[74] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics," *Journal of the American Chemical Society*, vol. 112, pp. 6127–6129, 08 1990.

[75] M. Schaefer and C. Froemmel, "A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution," *Journal of Molecular Biology*, vol. 216, no. 4, pp. 1045–1066, 1990.

[76] D. Hamelberg, J. Mongan, and J. A. McCammon, "Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules," *The Journal of Chemical Physics*, vol. 120, pp. 11919–11929, 2019/04/08 2004.

[77] Y. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker, and J. A. McCammon, "Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation," *Journal of Chemical Theory and Computation*, vol. 10, pp. 2677–2689, 07 2014.

[78] L. Sutto, S. Marsili, and F. L. Gervasio, "New advances in metadynamics," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, pp. 771–779, 2019/04/08 2012.

[79] D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications and new perspectives," *Physical Chemistry Chemical Physics*, vol. 7, pp. 3910–3916, 2005.

[80] D. J. Sindhikara, D. J. Emerson, and A. E. Roitberg, "Exchange Often and Properly in Replica Exchange Molecular Dynamics," *Journal of Chemical Theory and Computation*, vol. 6, pp. 2804–2808, 09 2010.

[81] A. Patriksson and D. van der Spoel, "A temperature predictor for parallel tempering simulations," *Physical Chemistry Chemical Physics*, vol. 10, pp. 2073–2077, 2008.

[82] Y. Chebaro, A. J. Ballard, D. Chakraborty, and D. J. Wales, "Intrinsically Disordered Energy Landscapes," *Scientific Reports*, vol. 5, May 22 2015.

[83] A. Sircar, S. Chaudhury, K. P. Kilambi, M. Berrondo, and J. J. Gray, "A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19," *Proteins*, vol. 78, pp. 3115–3123, 11 2010.

[84] A. M. Wollacott, A. Zanghellini, P. Murphy, and D. Baker, "Prediction of structures of multidomain proteins from structures of the individual domains," *Protein science : a publication of the Protein Society*, vol. 16, pp. 165–175, 02 2007.

[85] R. Chen, L. Li, and Z. Weng, "ZDOCK: An initial-stage protein-docking algorithm," *Proteins: Structure, Function, and Bioinformatics*, vol. 52, pp. 80–87, 2019/04/18 2003.

[86] B. G. Pierce, K. Wiehe, H. Hwang, B.-H. Kim, T. Vreven, and Z. Weng, "ZDOCK

server: interactive docking prediction of protein-protein complexes and symmetric multimers," *Bioinformatics (Oxford, England)*, vol. 30, pp. 1771–1773, 06 2014.

[87] B. Raveh, N. London, L. Zimmerman, and O. Schueler-Furman, "Rosetta FlexPepDock *ab-initio*: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors," *PLOS ONE*, vol. 6, pp. e18934–, 04 2011.

[88] D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov, and S. Vajda, "The ClusPro web server for protein-protein docking," *Nature Protocols*, vol. 12, pp. 255–278, 02 2017.

[89] C. Chipot and D. Pearlman, "Free energy calculations. the long and winding gilded road," *Mol Sim*, vol. 28, no. 1–2, pp. 1–12, 2002.

[90] P. Kollman, "Free-energy calculations - Applications to chemical and biochemical phenomena," *Chemical Reviews*, vol. 93, no. 7, pp. 2395–2417, 1993.

[91] N. Lu, D. Kofke, and T. Woolf, "Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods," *Journal of Computational Chemistry*, vol. 25, no. 1, pp. 28–39, 2004.

[92] V. B. Luzhkov, "On relation between the free-energy perturbation and Bennett's acceptance ratio methods: Tracing the influence of the energy gap," *The Journal of Chemical Physics*, vol. 132, p. 194104, 2019/04/16 2010.

[93] L. Wang and B. R. Donald, "An efficient and accurate algorithm for assigning nuclear overhauser effect restraints using a rotamer library ensemble and residual dipolar couplings," in *2005 IEEE Computational Systems Bioinformatics Conference (CSB'05)*, pp. 189–202, 2005.

[94] D. A. Pearlman and P. A. Kollman, "Are time-averaged restraints necessary for nuclear magnetic resonance refinement?: A model study for DNA," *Journal of Molecular Biology*, vol. 220, no. 2, pp. 457–479, 1991.

[95] A. E. Torda, R. M. Scheek, and W. F. van Gunsteren, "Time-averaged nuclear overhauser effect distance restraints applied to tendamistat," *Journal of Molecular Biology*, vol. 214, no. 1, pp. 223–235, 1990.

[96] E. N. Pederson and G. Interlandi, "Oxidation-induced destabilization of the fibrinogen αC-domain dimer investigated by molecular dynamics simulations," *Proteins*, May 2019.

[97] P. Tompa, "Intrinsically unstructured proteins," *Trends in Biochemical Sciences*, vol. 27, no. 10, pp. 527–533, 2002.

[98] L. Mollica, L. M. Bessa, X. Hanoulle, M. R. Jensen, M. Blackledge, and R. Schneider, "Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment," *Frontiers in molecular biosciences*, vol. 3, pp. 52–52, 09 2016.

[99] P. Tompa, G. E. Tusnády, P. Friedrich, and I. Simon, "The role of dimerization in prion replication," *Biophysical journal*, vol. 82, pp. 1711–1718, 04 2002.

[100] B. M. Coleman, C. F. Harrison, B. Guo, C. L. Masters, K. J. Barnham, V. A. Lawson, and A. F. Hill, "Pathogenic mutations within the hydrophobic domain of the prion protein lead to the formation of protease-sensitive prion species with increased lethality," *Journal of virology*, vol. 88, pp. 2690–2703, 03 2014.

[101] G. Interlandi, O. Yakovenko, A.-Y. Tu, J. Harris, J. Le, J. Chen, J. López, and W. E. Thomas, "Specific electrostatic interactions between charged amino acid residues regulate binding of von Willebrand factor to blood platelets," *Journal of Biological Chemistry*, 09 2017.

[102] J. J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker, "Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations," *Journal of Molecular Biology*, vol. 331, no. 1, pp. 281–299, 2003.

[103] A. Apicella, M. Marascio, V. Colangelo, M. Soncini, A. Gautieri, and C. J. Plummer, "Molecular dynamics simulations of the intrinsically disordered protein amelogenin," *Journal of Biomolecular Structure and Dynamics*, vol. 35, no. 8, pp. 1813–1823, 2016.

[104] U. Haberthür and A. Caflisch, "FACTS: Fast analytical continuum treatment of solvation," *Journal of Computational Chemistry*, vol. 29, pp. 701–715, 2019/05/22 2008.

[105] W. Humphrey, A. Dalke, and K. Shulten, "VMD: visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[106] M. Heinig and D. Frishman, "STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins," *Nucleic acids research*, vol. 32, pp. W500–W502, 07 2004.

[107] E. Rosta and G. Hummer, "Error and efficiency of replica exchange molecular dynamics simulations," *The Journal of chemical physics*, vol. 131, pp. 165102–165102, 10 2009.

[108] D. A. Kofke, "On the acceptance probability of replica-exchange Monte Carlo trials," *The Journal of Chemical Physics*, vol. 117, pp. 6911–6914, 2019/04/15 2002.

[109] L. A. Kelley, S. P. Gardner, and M. J. Sutcliffe, "An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies," *Protein Engineering Design and Selection*, vol. 9, no. 11, pp. 1063–1065, 1996.

[110] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera—A visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, pp. 1605–1612, 2019/06/03 2004.

[111] L. V. Medved, O. V. Gorkun, and P. L. Privalov, "Structural organization of C-terminal parts of fibrinogen Aα-chains," *FEBS Lett.*, vol. 160, pp. 291–295, 1983.

[112] M. Martinez, J. W. Weisel, and H. Ischiropoulos, "Functional impact of oxidative posttranslational modifications on fibrinogen and fibrin clots," *Free Radical Biology and Medicine*, vol. 65, pp. 411–418, 2013.

[113] R. Davenport, J. Manson, H. De'Ath, S. Platton, A. Coates, S. Allard, D. Hart, R. Pearse, K. J. Pasi, P. MacCallum, S. Stanworth, and K. Brohi, "Functional definition and characterization of acute traumatic coagulopathy," *Critical Care Medicine*, vol. 39, no. 12, pp. 2652–2658, 2011.

[114] B. Lipinski, "Pathophysiology of oxidative stress in diabetes mellitus," *Journal of Diabetes Complications*, vol. 15, no. 4, pp. 203–210, 2001.

[115] T. N. Heinz and P. H. Hünenberger, "A Fast Pairlist-Construction Algorithm for Molecular Simulations under Periodic Boundary Conditions," *Journal of Computational Chemistry*, vol. 25, pp. 1474–1486, 2004.

[116] M. Rocco, M. Molteni, M. Ponassi, G. Giachi, M. Frediani, A. Koutsioubas, A. Profumo, D. Trevarin, B. Cardinali, P. Vachette, F. Ferri, and J. Pérez, "A Comprehensive Mechanism of Fibrin Network Formation Involving Early Branching and Delayed Single- to Double-Strand Transition from Coupled Time-Resolved X-ray/Light-Scattering Detection," *Journal of the American Chemical Society*, vol. 136, pp. 5376–5384, 04 2014.

[117] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *Journal of Computational Chemistry*, vol. 30, pp. 2785–2791, 12 2009.

[118] G. Interlandi, "Destabilization of the von Willebrand factor A2 domain under oxidizing conditions investigated by molecular dynamics simulations," *PLoS ONE*, vol. 13, no. 9, pp. 1–17, 2018.

[119] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr, "XSEDE: Accelerating Scientific Discovery," *Computing in Science & Engineering*, vol. 16, no. 5, p. 62074, 2014.

[120] D. Wu and D. A. Kofke, "Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation," *The Journal of Chemical Physics*, vol. 123, p. 054103, 2019/06/03 2005.

[121] R. W. Zwanzig, "High-temperature equation of state by a perturbation method 1. nonpolar gases," *Journal of Chemical Physics*, vol. 22, no. 8, pp. 1420–, 1954.

[122] M. Zacharias, T. Straatsma, and J. McCammon, "Separation-shifted scaling, a new scaling method for lennard-jones interactions in thermodynamic integration," *Journal of Chemical Physics*, vol. 100, no. 12, pp. 9025–9031, 1994.

[123] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*, ch. 1, p. 43. New York: W.H. Freeman, 2002.

[124] A. Caflisch, "Folding for binding or binding for folding?," *Trends in Biotechnology*, vol. 21, pp. 423–425, 2019/04/17 2003.

Appendix A

Free energy landscapes.

The FELs of the unoxidized αC-domain dimers were calculated at 306.30 K, 312.74 K, 319.31 K, 326.02 K, 332.87 K, 339.86 K and 347.00 K (Figures A.1 to A.8). The FELs of the oxidized αC-domain dimers were calculated at using the same temperature series (Figures A.9 to A.16).

A.1 Figures.

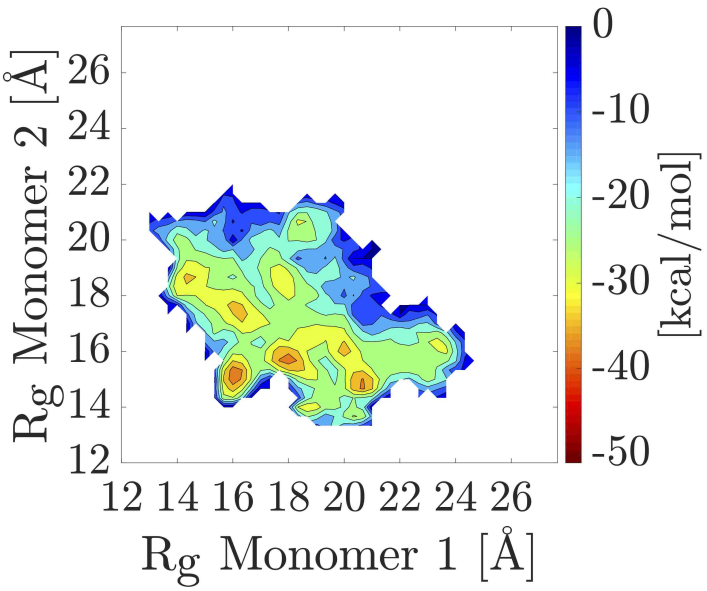

Figure A.1:: Free energy landscapes of the unoxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 300.0 K.

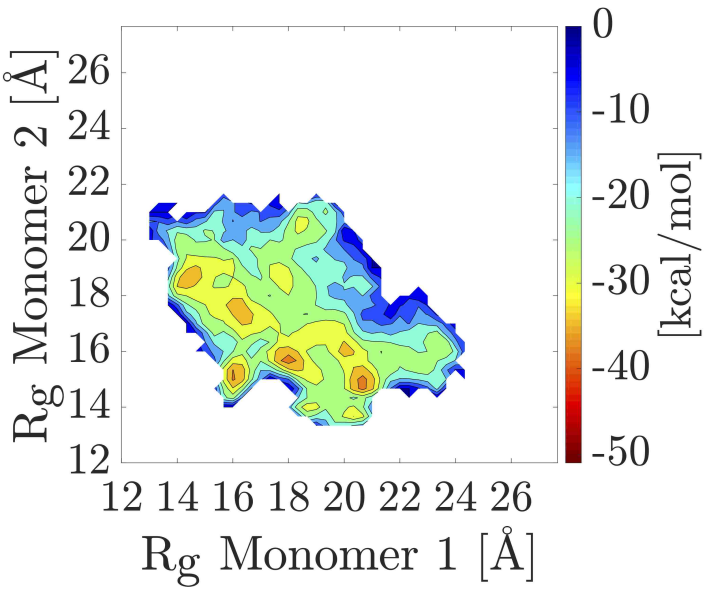

Figure A.2:: Free energy landscapes of the unoxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 306.30 K.

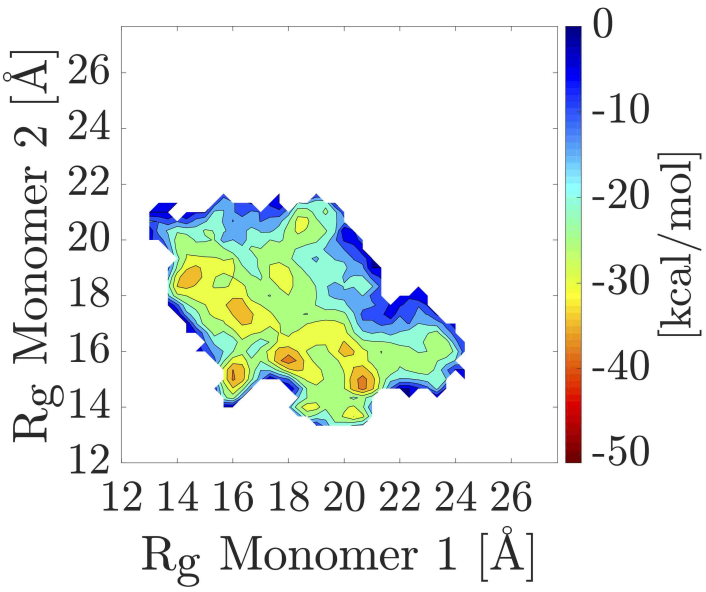

Figure A.3:: Free energy landscapes of the unoxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 312.74 K.

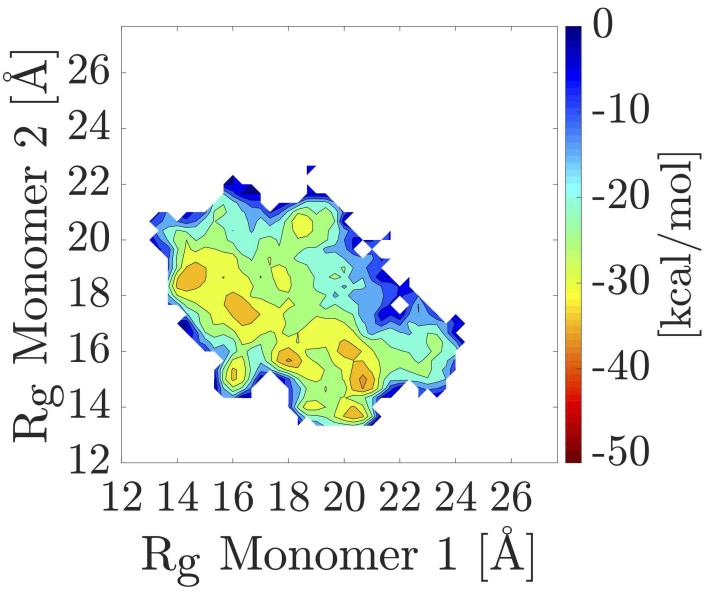

Figure A.4:: Free energy landscapes of the unoxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 319.31 K.

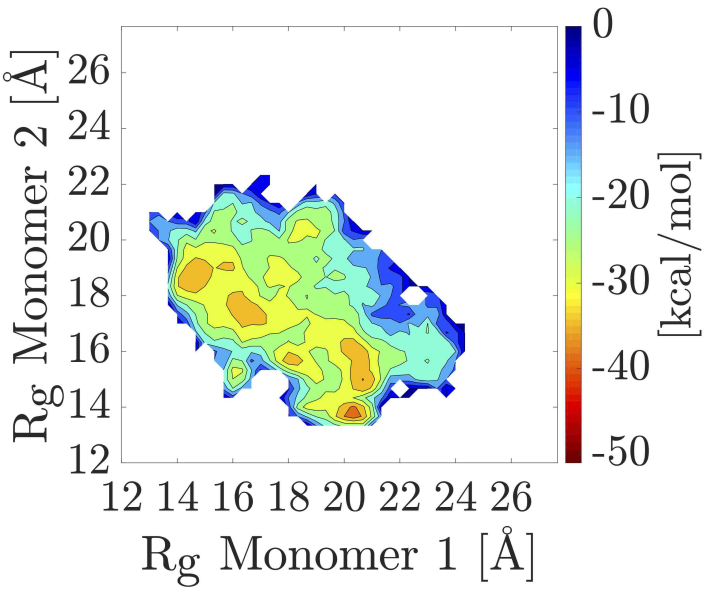

Figure A.5:: Free energy landscapes of the unoxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 326.02 K.

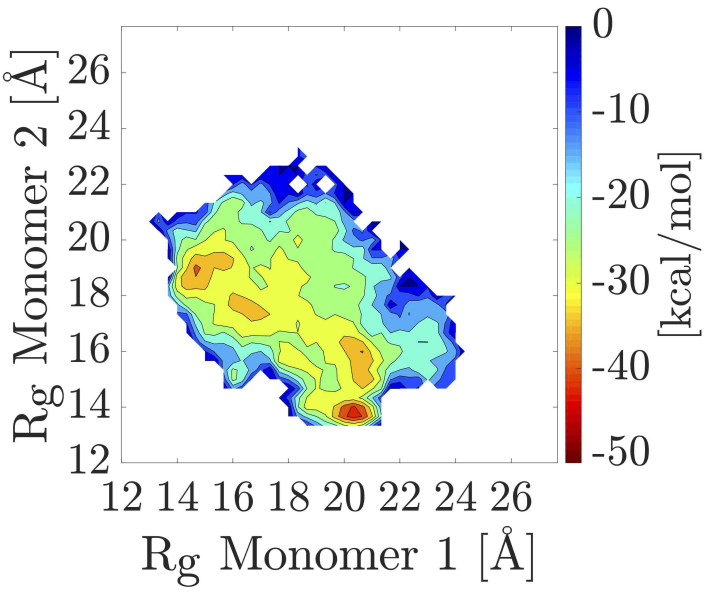

Figure A.6:: Free energy landscapes of the unoxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 332.87 K.

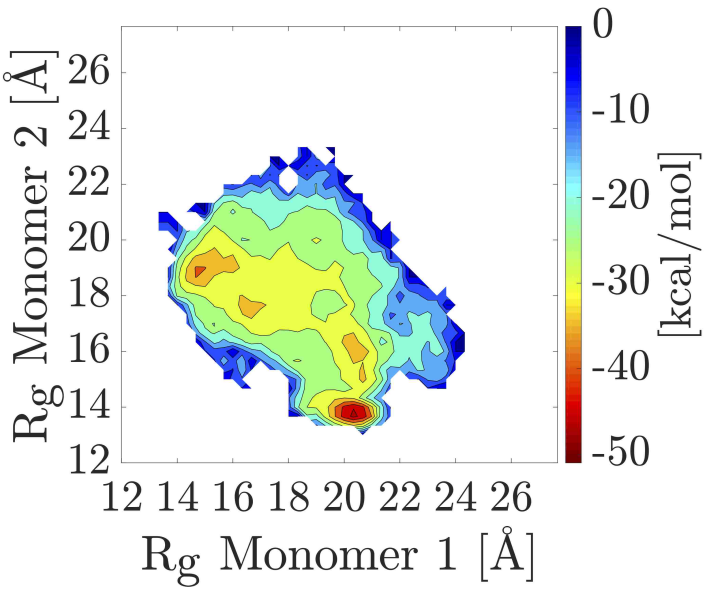

Figure A.7:: Free energy landscapes of the unoxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 339.86 K.

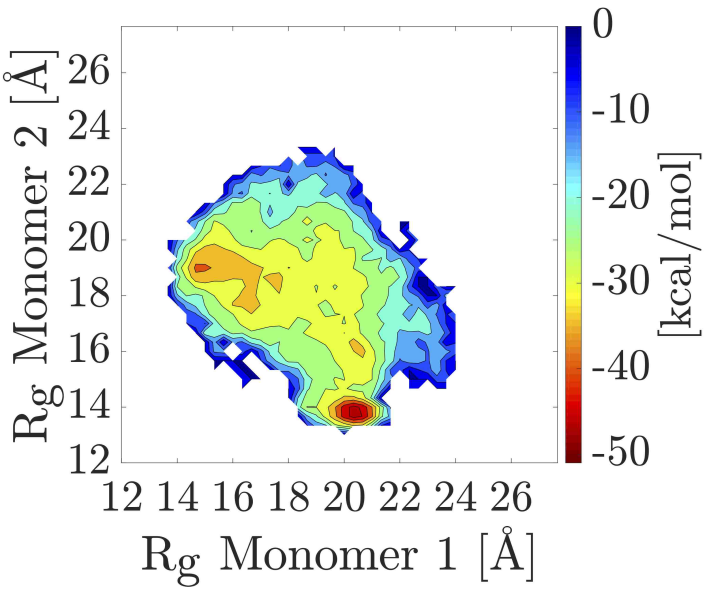

Figure A.8:: Free energy landscapes of the unoxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 347.00 K.

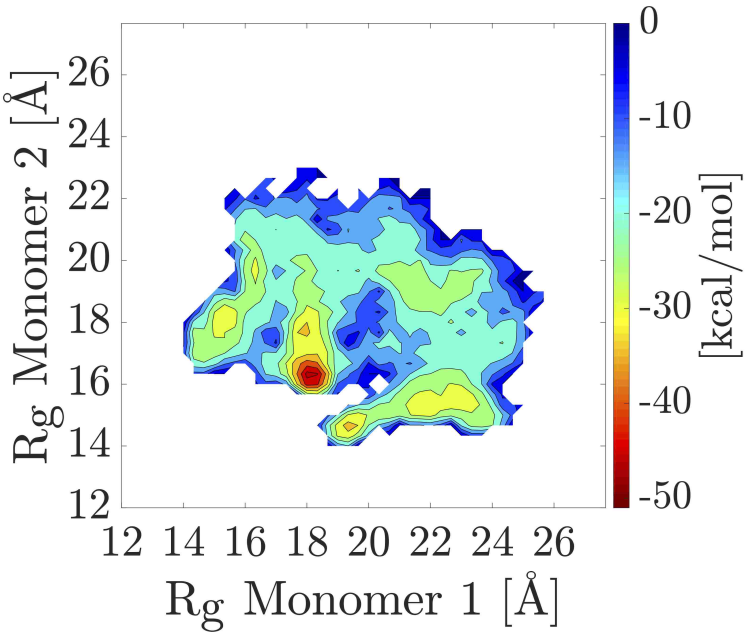

Figure A.9:: Free energy landscapes of the oxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 300.0 K.

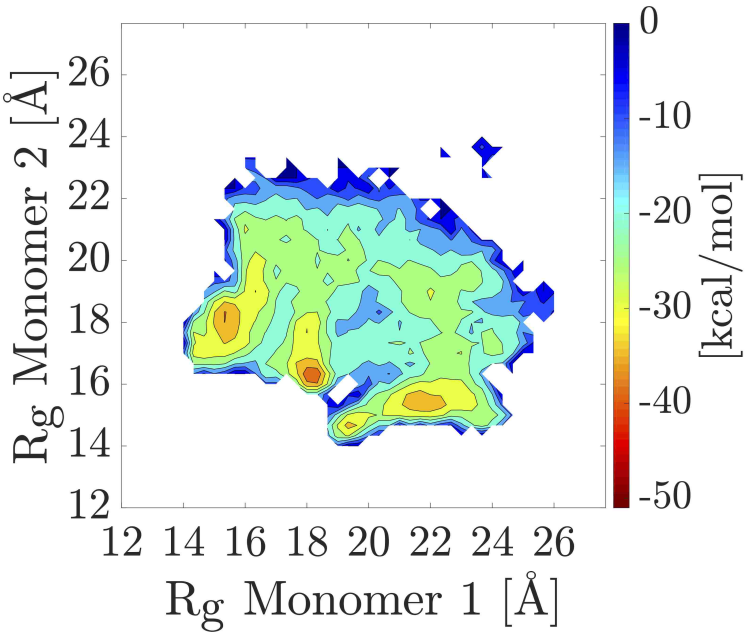

Figure A.10:: Free energy landscapes of the oxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 306.30 K.

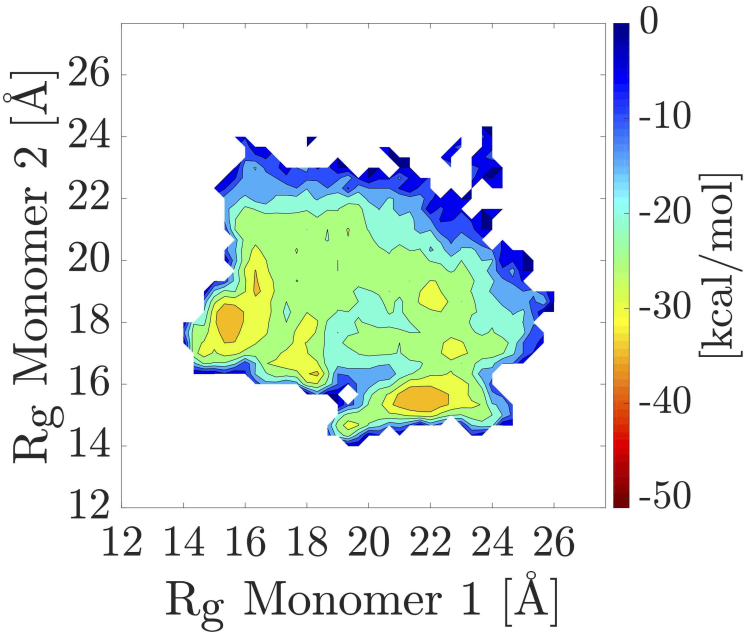

Figure A.11:: Free energy landscapes of the oxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 312.74 K.

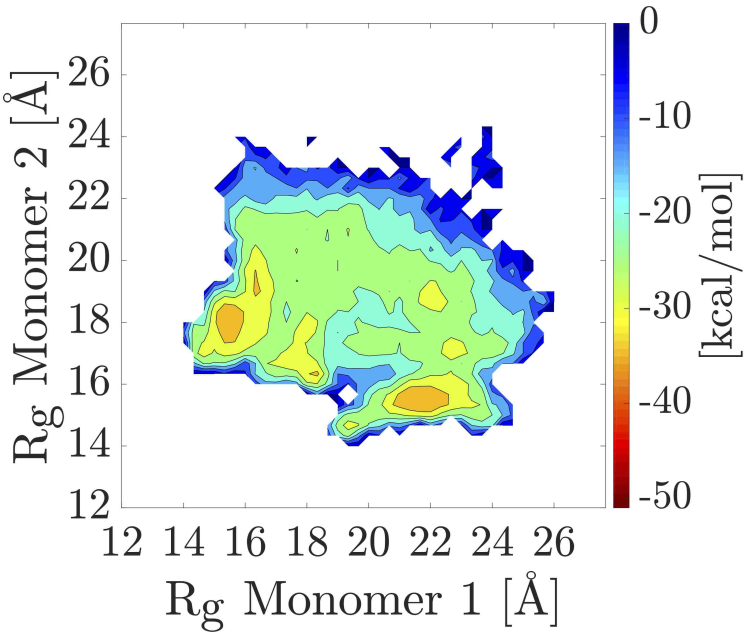

Figure A.12:: Free energy landscapes of the oxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 319.31 K.

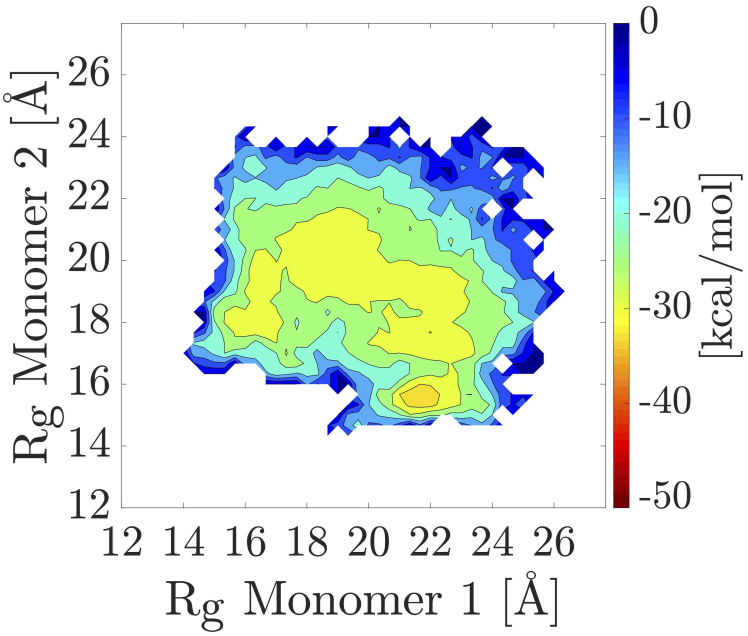

Figure A.13:: Free energy landscapes of the oxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 326.02 K.

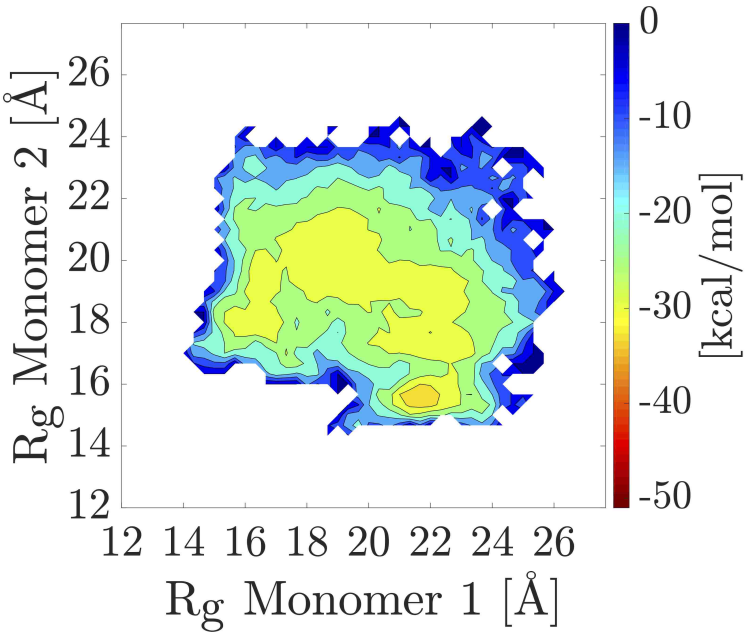

Figure A.14:: Free energy landscapes of the oxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 332.87 K.

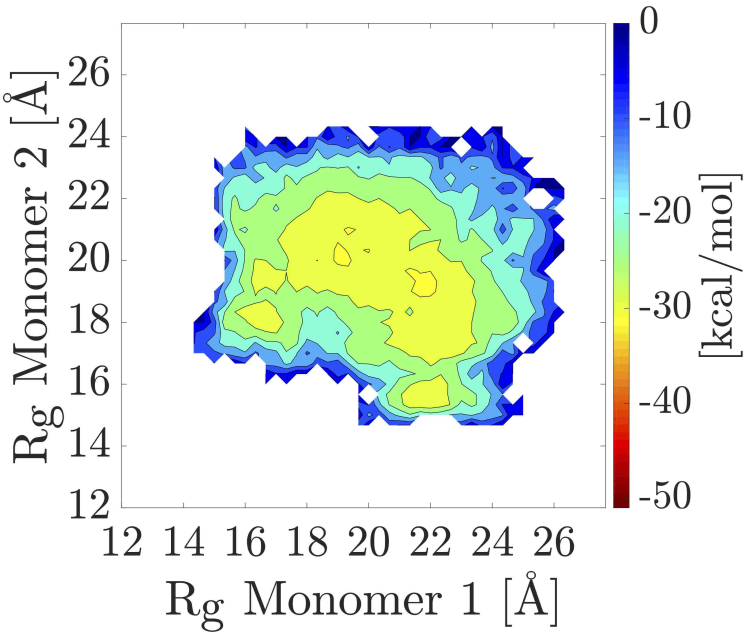

Figure A.15:: Free energy landscapes of the oxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 339.86 K.

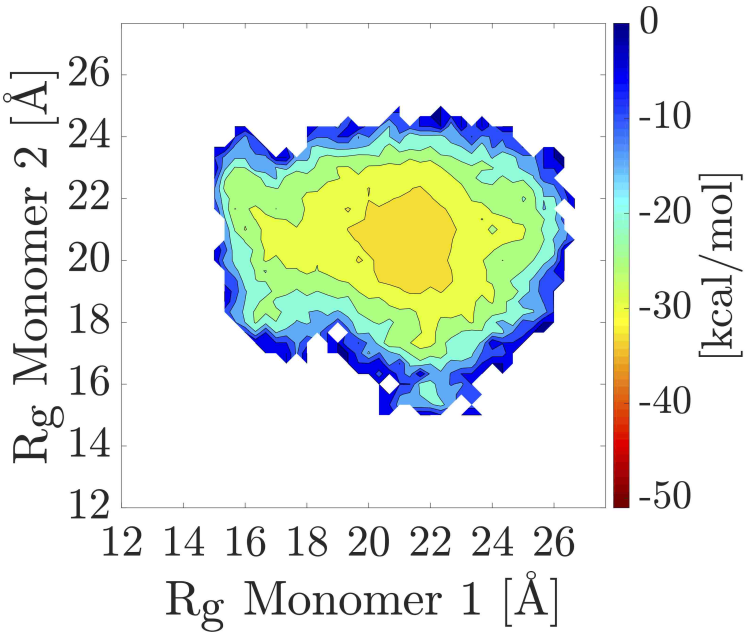

Figure A.16:: Free energy landscapes of the oxidized αC-domain dimer projected onto the radii of gyration $R_{g\,1}$ and $R_{g\,2}$ with frames sampled at 347.00 K.