The Role of Unfolded States in Collagen Degradation

by

Ramon Salsas Escat

B.S. Chemistry
Institut Químic de Sarrià, Barcelona, Spain, 2003

Submitted to the Computational and Systems Biology Program
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Computational and Systems Biology

at the

Massachusetts Institute of Technology

June 2010

Signature of Author……………………………………………………………………………
Computational and Systems Biology Program
March 16, 2010

Certified by……………………………………………………………………………………
Collin M. Stultz
Associate Professor of Electrical Engineering and Computer Science
and Health Sciences and Technology
Thesis Supervisor

Accepted by…………………………………………………………………………………...
Christopher Burge
Associate Professor of Biology
Director, Computational and Systems Biology Graduate Program

The Role of Unfolded States in Collagen Degradation

by

Ramon Salsas Escat

Submitted to the Program of Computational and Systems Biology
on March 16, 2010 in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Computational and Systems Biology

ABSTRACT

Excessive collagen degradation (collagenolysis) has been implicated in a series of diseases such as tumor metastasis, atherosclerosis and arthritis. There are still several unresolved questions about the mechanism of collagenolysis. First, the prototypical structure of the collagen triple helix does not fit into the active site of collagenases, the enzymes responsible of cleaving collagen. Moreover, the scissile bond that is degraded during collagenolysis is hidden from solvent. Therefore it is widely agreed that collagen unfolding must occur in order for collagenolysis to proceed. Some proposed mechanisms suggest that collagenases actively unfold collagen in order to expose the cleavage site, but no direct evidence of such mechanisms has been provided. Second, while several potential cleavage sites exist in the sequence of collagen, only one is cleaved in triple helical collagen.

The hypothesis of this work is that locally unfolded states exist in collagen in the absence of collagenases. They occur as a result of the natural thermal fluctuations in the structure of collagen. Collagenolysis occurs when collagenases bind and cleave these unfolded states. In this work, a combination of computational and experimental methods is presented in order to test this hypothesis. Initially, computational results suggest that locally unfolded states are ubiquitous along the structure of collagen. However, it is shown that not all unfolded states are created equal, and that the precise sequence in the vicinity of the true collagenase cleavage site in type III collagen allows collagen to sample locally unfolded states that are complimentary to the collagenase active site. Therefore, it is hypothesized that cleavage site specificity is encoded in the nature of the unfolded states. Next, it is shown that types I and III collagen can be bound and cleaved at the actual cleavage site by just the catalytic domain of collagenases, a finding in apparent contradiction with previous work in this field. These results are interpreted in light of a novel conformational selection mechanism in which collagenases only cleave locally unfolded, vulnerable states. Finally, based on the new mechanism of collagenolysis presented here, new strategies to regulate collagenolysis are proposed.

Thesis Supervisor: Collin M. Stultz
Title: Associate Professor of Electrical Engineering and Computer Science
and Health Sciences and Technology

Acknowledgments

Dedicat als meu pares, Ramon i Núria, i a la meva dona, Melike.

To my parents, Ramon and Núria, and to my wife, Melike.

I would like to start this round of acknowledgments by thanking my thesis committee members. My advisor, Collin Stultz, the Chair of my committee, Amy Keating, and Thomas Schwartz and Michael Smith. Thanks to all of them for being part of this endeavor and their help during this process. I've had the pleasure to rotate in the Keating lab, as well as taking fantastic classed with Amy and Thomas, from which I learnt a great deal. They have been great teachers. I would also like to thank Michael Smith, from Boston University, who has jumped aboard this boat at the last moment to contribute with his expertise in order to evaluate this work.

Next, I would like to thank personally my advisor, Collin Stultz. Collin and I met a while ago, before I started my graduate school at MIT. For this reason, I would like to say that this thesis is the end result of almost seven years of my life at MIT, luckily only five of these in grad school. I arrived in September 2003 with a suitcase and a vague idea that I was going to learn "Biochemistry" in general at MIT. This was an exciting prospect after an organic chemistry undergraduate education. I was here to carry out a one year research project, part of my master's thesis, or equivalent, in Barcelona. Soon after joining the Edelman lab, from the Health Sciences and Technology Division, and surveying the different projects that were going on, I approached a *very talented postdoctoral fellow by the name* of Collin M. Stultz, since I was interested on his research. It took a while to get his attention, since he was busy in the process of applying for professor positions. Soon after, we decided to work together and he taught me from the basics of molecular dynamics simulations to American slang words that I had never learnt back in Spain. After taking a break for a summer internship in Europe, I returned to MIT in the fall of 2004 to join the newly formed Computational Biophysics Laboratory. I am very grateful to Collin for supporting me for a year, while continuing research in his laboratory and applying for PhD programs. After being accepted to the Computational and Systems Biology PhD program, I decided to continue my research in the Stultz lab, a path that led us to, amongst other things, this thesis. During my extended interaction with Collin, I have enjoyed his mentoring, both inside and outside the lab. We have got our share of heated scientific discussion, and I think that both of us have benefitted from it. Thanks to Collin I can consider myself a scientist. With his guidance, I have learnt how to approach scientific problems, and how to criticize my own work and that of others.

Next, I would like to thank Mercedes Balcells. Merche recruited me in Barcelona to join the MIT adventure in the first place. She also mentored me during my time in the Edelman lab,

and we have remained friends after that. She has been the anchor of the IQS community at MIT, half a mother and half a friend to all the students coming from Barcelona. Thanks also to Elazer Edelman, whose lab I joined my first year at MIT. He allowed me to have my first taste of real research. More importantly, he made me feel valued for my work and contribution to the lab, not an easy thing in a very well established lab where I was the most junior member.

My time in the Computational Biophysics Laboratory wouldn't have been the same without the people in it. A special mention for Dr. Paul Nerenberg, a name that appears quite often in this thesis. This is because we have worked together very closely in some of the collagen projects. I am amazed how a chemist and a physicist have managed to get along and speak the same language. But we have, and while I have roamed around the wet lab more than him, and he is a master of Matlab more than me, we can still comment on each others' work very easily. I am convinced that I have influenced him on developing a new appreciation of the world of Spanish pork products, the same way I learnt from him many aspects of the American culture unknown to me.

Let me continue by thanking all the other members, current and former, from the Stultz lab. This family of notables includes: Dr. Christian Schubert (the α-Helices Wizard), Elaine Gee (Lady Fibronectin), Dr. Austin Huang (Dr. Tau), Charles Fisher (the Man Who Sees the World in Equations), Orly Ullman (the α-Synuclein Woman, or lady triop), Joyatee Sarker (Ms. My Heart Goes Boom or Ms. Heart Beat) and Dr. Sophie Walker (Dr. Kiwi, or the discovery that there is whole new way to speak English that I didn't know about). It has been a fun ride to share with all of you. I would also like to thank the two undergraduate students that I mentored during these years, who contributed to the experimental part of my thesis, William M. Wyatt and Steven J. Pennybaker.

If there is an activity that has kept me sane during all these years, it is basketball. That is the time when the only oxygen left in my brain is used to decide the next play. There is no time to think about collagen! The perfect way to disconnect. I want to thank everyone I have met on the court during these seven years. The invention of "noon ball", which has been going on for over 30 years on Mondays, Wednesdays and Fridays, should be awarded some kind of MIT prize.

Next, I would like to thank the European Club at MIT. I have met many friends there, and had the opportunity to learn so many things that are not taught in classes or laboratories and have a lot of fun (wine and cheese parties, Eurovision parties, European Career Fair dinners…). I would like to make a special mention of my friends Dr. Mike Koeris, Vice President of the European Club and who we met on the basketball court and introduced me to the Club, Pedja Djuranovic, the President of the European Club and Dr. Wiep Klaas Smits, who I had the privilege to work very closely with when chairing the organization of the 13th European Career Fair.

6

I wanted to keep my last words for my family, who I have terribly missed during all these years. I would like to acknowledge my parents Ramon and Núria. They have always supported me with every one of my choices. They knew that my one year trip to MIT back in 2003 could lead to an extended stay. And when it did, they encouraged me, not without sadness not to see me return, since they reckoned it was the best for my future. Without their support it would have been impossible to come to MIT in the first place, since there was no financial support I could apply for. I would also like to include in my acknowledgements the closest members of my family, my fantastic grandmother Esperança, my aunt and uncle Margarita and Miquel and my cousins Roger, Enric, Marta and Pau. I know that they have all enjoyed my explanations in Catalan of the science that I am doing at MIT, showing them how collagen degradation occurs over a Christmas dinner. While in the distance, they have been by my side and they have welcomed me every time I have been back home. Moltes gràcies a tots i fins aviat!

If it was not enough with one family, I managed to pick up a second one along the way of my PhD years. It all started in 2005, when I met Melike Lakadamyalı in Boston. We got engaged in 2007, and got married in 2008. And with her came her great Turkish Cypriot family: my new parents Şener and Zehra, in-laws, aunts, uncles and cousins Fezile, Savaş, Huray, Çidem, Polen, Hüseyin, Harun… (to mention a few, since Cypriot families are so large that I would need a second thesis to mention them all). I would like to thank them all for making me feel like one more of the family and showing me a great time during the needed vacation periods during my thesis. Çok teşekkürler!

Finally, I would like to make a special mention of my wife Melike. She has been by my side since I started my PhD and supported me during all the good and bad times. A scientist herself, she knows how hard graduate student life can be at times. Somebody else would not understand the long hours, total dedication to the project and the "Sundays in the Lab" (in honor of Uri Alon). We have grown together, both personally and scientifically. We have both helped each other during this time, and will continue to do so in the future.

Contents

List of Figures

Introduction

I.1. Collagen

I.1.1. The collagen triple helix

Collagen is composed of three polypeptide chains that form a triple helical structure. Each of the chains can be quite long, about 1000 amino acids or more in length in some collagen types. A repeating sequence of three amino acids (G-X1-X2) forms this structure, in which every third amino acid is glycine (Ramachandran and Kartha 1955; Rich and Crick 1955; Rich and Crick 1961). The tight internal packing of the triple-helix only allows a residue without a side-chain to occupy every first position. Many of the remaining positions in the chain are occupied by proline (Pro or P) and a post-translational modification of proline, hydroxyproline (Hyp, or O) (Brodsky and Ramshaw 1997). The three polypeptide chains, each in an extended left-handed polyproline II-helix conformation, are supercoiled in a right-handed manner around a common axis, with a stagger of one residue between adjacent chains (Brodsky and Persikov 2005). A classic triple helix is shown in Figure I.1, corresponding to the PDB file 1BKV (Kramer, Bella et al. 2001).
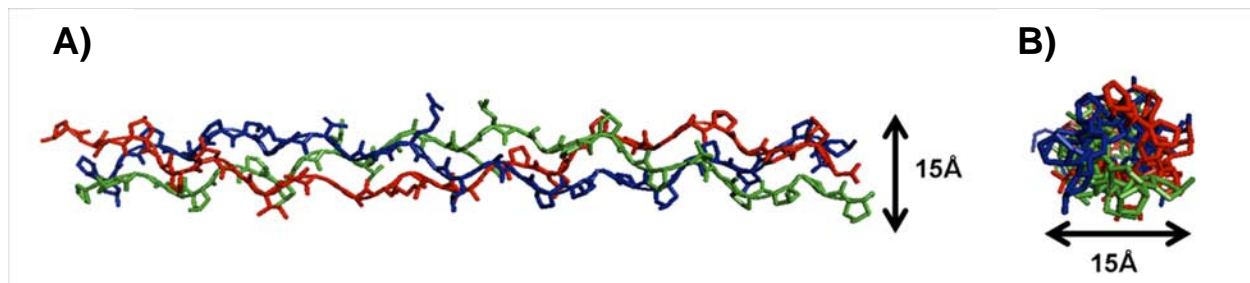
Figure I.1. Representation of the triple helix of collagen. The figure is based on the crystal structure of a type III collagen like peptide containing a biologically relevant sequence, PDB file 1BKV (Kramer, Bella et al. 2001). A) Lateral view of the collagen triple helix. Each chain is 30 amino acids in length. The diameter of the triple helix is about 15Å. The backbone is shown in cartoon representation, and the side chains in stick representation. B) View down the barrel of the collagen triple helix. Both representations were generated with Pymol (DeLano 2002).

Regions in the collagen sequence that are rich in GPO (glycine-proline-hydroxyproline) repeats (imino rich regions) are essential for the overall triple helical structure of collagen, but represent only 10% of the total type I collagen sequence (Shoulders and Raines 2009). Imino acids are responsible for triple helical stability, while other amino acids can stabilize the triple helix to a lesser degree or even destabilize it (Persikov, Ramshaw et al. 2000; Persikov, Ramshaw et al. 2005). Indeed, not only imino acids are found in the X1 and X2 positions of the G-X1-X2 triplets, but other amino acids can be present. There are several stretches within different collagen sequences that are relatively imino poor in that few or no imino acids are present. These imino poor regions, rather than playing a structural role, are thought to carry biological information, such as cell recognition or protein binding sites (Brodsky and Persikov 2005). Most importantly, the regions downstream of the cleavage sites in several types of collagen have a low content of imino acids and may play an important role in collagen degradation (Fields 1991).

I.1.2. Collagen types and their role in normal and disease states

About one third of all of proteins in the human body belong to the class of structural proteins known as collagens (Shoulders and Raines 2009). Up to 29 vertebrate collagen types have been discovered to this date (Shoulders and Raines 2009). They can be subdivided into several subfamilies. Fibrillar collagens, the most abundant, include type I, II, III, V and XI (Hulmes 2002; Gelse, Poschl et al. 2003). Other collagens fall under the network-forming, fibril associated with interrupted triple helices (FACIT), membrane associated with interrupted triple helices (MACIT) and multiple triple-helix domains and interruptions (MULTIPLEXINs) categories (Brinckmann 2005; Shoulders and Raines 2009). Interestingly, collagens are also present in invertebrates (Boot-Handford and Tuckwell 2003), and triple helical domains resembling collagen's structure have been discovered in bacteria (Xu, Keene et al. 2002; Mohs, Silva et al. 2007).

In humans, fibrillar collagens form fibers that add significant tensile strength to a variety of body structures. To mention a few, type I collagen is the most abundant collagen and is present in tendon, bone, ligament, type II collagen is the main component of cartilage, and type III collagen is present in skin and blood vessels. Type I collagen is a heterotrimer composed of two $\alpha1$(I) chains and one $\alpha2$(I) chain (Ricard-Blum, Ruggiero et al. 2005). Type II collagen is a homotrimer composed of three $\alpha1$(II) chains (Gelse, Poschl et al. 2003; Shoulders and Raines 2009). Type III collagen is a homotrimer composed of three $\alpha1$(III) chains (Gelse, Poschl et al. 2003). In all three types of collagen, each of the chains is about 1000 amino acids in length.

Collagen is involved is several disease states that affect the human body. The inefficient biosynthesis of collagen or the synthesis of structurally defective collagen, for example, leads to disorders such as the family of Ehlers-Danlos syndromes (Parapia and Jackson 2008). In type IV

Ehlers-Danlos syndrome, for example, mutations in the type III collagen gene lead to a deficiency of type III in the vessel walls, which, in turn, results into debilitated blood vessels that can rupture (Germain 2007). Another disorder is osteogenesis imperfecta (OI), also known as brittle bone disease, characterized by fragile bones susceptible to breaking (Gajko-Galicka 2002). Several mutations in the $\alpha1(I)$ and $\alpha2(I)$ chains of type I collagen are responsible for a variety of different phenotypes in osteogenesis imperfecta (Gajko-Galicka 2002). The previously discussed disorders have in common the fact that mutations in the collagen genes or in the enzymes required for collagen biosynthesis are responsible for the disease. This confines these disorders to a relatively small number of patients. Ehlers-Danlos syndrome, for example, affects an estimated 1 in 10000 to 25000 individuals, 5-10% of which correspond to type IV Ehlers-Danlos (Germain 2007), while OI occurs in 1 out of 10000 births in the US (Martin and Shapiro 2007).

Other collagen-related disorders that affect larger number of individuals have been associated with excessive collagen degradation. In these disorders, no mutations in collagen or in the enzymes required for its biosynthesis need to occur. Therefore individuals that correctly synthesize collagen are susceptible to these disorders. Amongst these disease states we find atherosclerosis and tumor metastasis. Atherosclerotic heart disease is the most prevalent of these conditions and affects about 80 million people in the US (Lloyd-Jones, Adams et al. 2009). Mature atherosclerotic plaques are lipid-laden protrusions from the inner lining of arteries (Celentano and Frishman 1997; Barnes and Farndale 1999; McDonnell, Morgan et al. 1999). These abnormal evaginations contain an overlying layer of collagen that separates the internal lipid-laden contents within the plaque from the blood. When excessive collagen degradation occurs at the plaque, the outer layer of collagen debilitates and can rupture. When this protective

lining of collagen is degraded, the contents of the atherosclerotic plaque are exposed to the blood, leading to the formation of a clot in the blood vessel itself; i.e., an event which prevents flow through the artery. This sequence of events is the major cause of catastrophic heart attacks (Barnes and Farndale 1999).

Another disorder which has been linked to abnormal collagen metabolism is tumor metastasis (McDonnell, Morgan et al. 1999; Nelson, Fingleton et al. 2000; Nerenberg, Salsas-Escat et al. 2007). For a further details about collagen and cancer metastasis, please refer to a review written by the Stultz group (Nerenberg, Salsas-Escat et al. 2007). In light of the clinical importance of collagen degradation, methods which provide a deeper understanding of the molecular mechanism of collagen degradation may shed light on the events underlying a number of important diseases.

I.1.3. Collagen biosynthesis

The biosynthesis of collagen is a complex process that requires the participation of many enzymes and coenzymes. This process can only be carried out in cells that contain the appropriate cellular machinery capable of introducing post-translational modifications (Gelse, Poschl et al. 2003; Koide and Nagata 2005; Salsas-Escat and Stultz 2009). The main steps of collagen biosynthesis are outlined in Figure I.2. Initially, procollagen α chains are synthesized in the endoplasmic reticulum, ER (I). Next, selected prolines and lysines are hydroxylated (II). This is followed by O-glycosilation of selected hydroxylysines (III). Then, the 3 procollagen α chains are ready for self assembly and the procollagen triple-helix forms (IV). This structure is then secreted (V and VI) and the propeptides are then cleaved (VII) to form collagen monomers. These monomers self-assemble into fibrils that present a banding pattern of periodicity of about 70nm, and can covalently crosslink to each other (VIII). Finally, the fibrils aggregate to form a collagen fiber (IX).
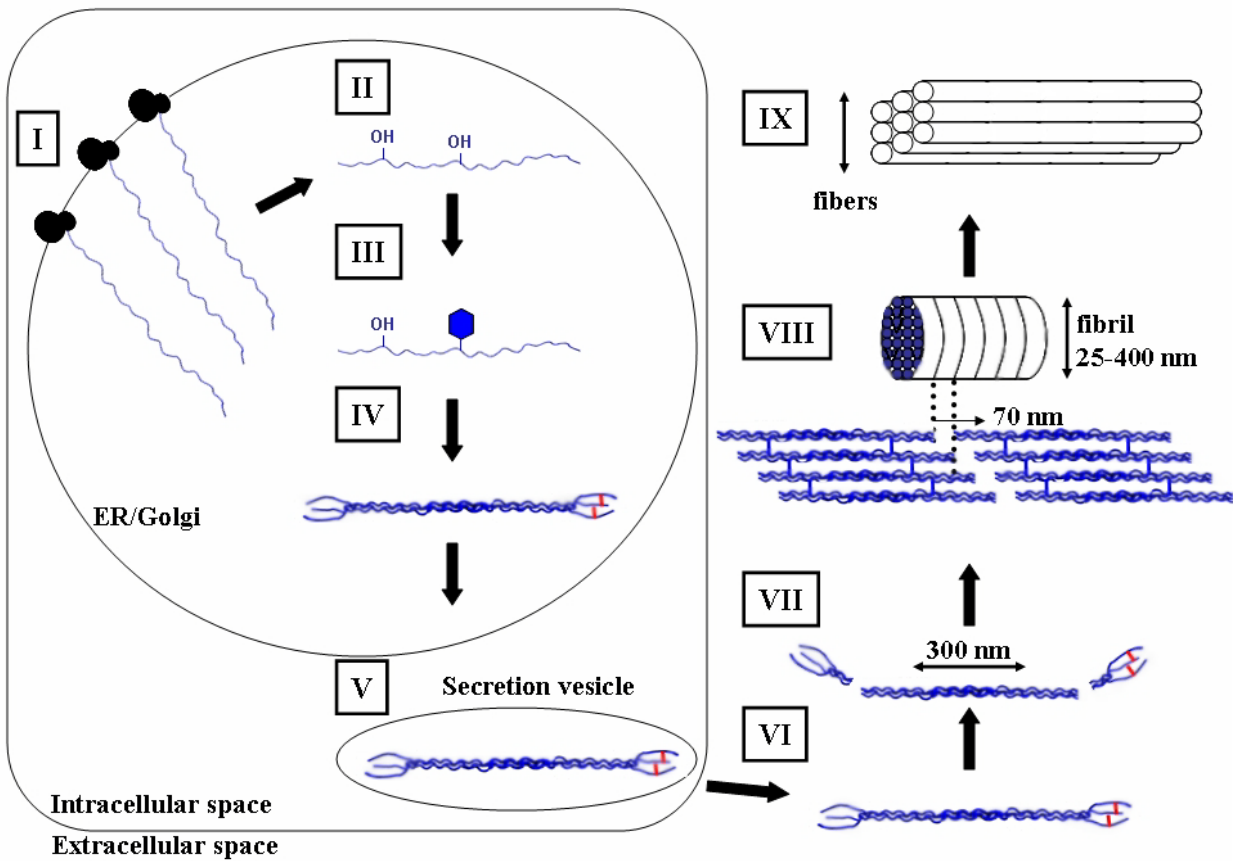
Figure I.2. Collagen biosynthesis. Schematic representation of collagen biosynthesis starting from the synthesis of the procollagen chains in the endoplasmic reticulum (ER) and finishing in the formation of collagen fibers (Hohenester and Engel 2002; Gelse, Poschl et al. 2003; Salsas-Escat and Stultz 2009).

I.2. Collagen degradation

I.2.1. Matrix metalloproteinases: general structural and sequence motifs

Matrix metalloproteinases are a family of Zn^{2+} and Ca^{2+} dependent endopeptidases that are capable of cleaving components of the extracellular matrix, including collagen, fibronectin and laminin (Chakraborti, Mandal et al. 2003). For example, a number of members of the MMP family such as fibroblast collagenase (MMP1), neutrophil collagenase (MMP8), collagenase-3 (MMP13), collagenase 4 (MMP18, found in frog but not in humans), membrane type I MMP (MMP14) and MMP2 (gelatinase A) are the only enzymes that can cleave native triple helical fibrillar collagens (Lauer-Fields, Juska et al. 2002; Chakraborti, Mandal et al. 2003; Song, Wisithphrom et al. 2006). There are at least 28 members of the MMP family. All but two (MMP7 and MMP26) contain at least the two domains shown in Figure I.3. This structure, corresponding to MMP1, consists of two domains - a catalytic domain of approximately 160-170 amino acids, and a hemopexin-like domain of about 200 amino acids – connected by a linker of variable length, between 15 and 65 residues (Tallant, Marrero et al. 2009). Some members of the MMP family contain additional domains, e.g. the transmembrane domain of MMP14 or the fibronectin-like domains of MMP2 (Lauer-Fields, Juska et al. 2002; Overall 2002; Bode and Maskos 2003; Tallant, Marrero et al. 2009).

MMPs are expressed as zymogens and can be activated intra or extracellularly (Pei and Weiss 1995; Chakraborti, Mandal et al. 2003). These zymogens include an approximately 80 residue long propeptide at the N-terminus of the catalytic domain that prevents access to the catalytic site (Overall 2002). The propeptide can be cleaved by membrane-type MMPs or other proteases, including activated MMPs (Jones, Sane et al. 2003). The catalytic domain contains a conserved sequence motif **HEXXHXXGXXH**, with glutamate being the catalytic residue and

the histidine residues responsible for coordinating the Zn^{2+} ion. In the absence of the hemopexin-like domain, some MMPs retain catalytic activity for cleaving gelatin (unfolded collagen), but lose efficiency to cleave triple helical collagen. The hemopexin is considered to be necessary for collagen binding and it has been hypothesized that it is involved in the active unfolding of the collagen triple helix (Overall 2002; Chung, Dinakarpandian et al. 2004). Several structures of different MMP domains or some full length MMPs have been previously reported (Maskos 2005).
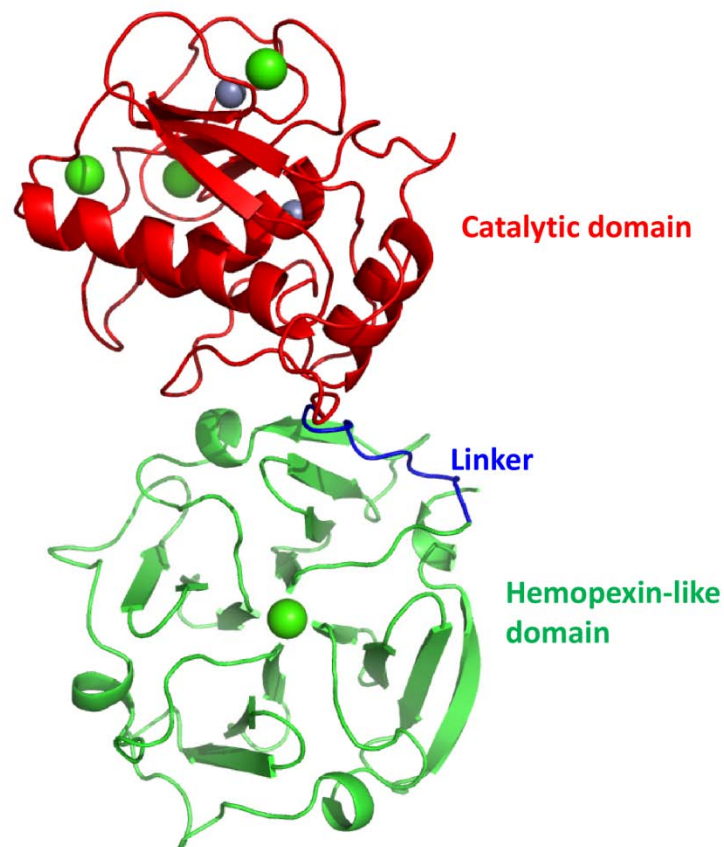
Figure I.3. Structure of matrix metalloproteinase-1. The crystal structure of MMP1 (PDB: 1FBL), which is the MMP archetypal structure (Li, Brick et al. 1995). The catalytic domain is shown in red, the linker in blue, the hemopexin-like domain in green, the calcium ions as green spheres and the zinc ions as grey sphere. This figure was created with molecular visualization package Pymol (DeLano 2002).

I.3. The mechanism of collagen degradation by MMPs: an unanswered question

The steps underlying the mechanism of collagen degradation are mysterious. Collagen degradation begins with the cleavage of collagen monomers by specific proteases called matrix metalloproteinases (MMPs), which cleave collagen at unique cleavage sites that are characterized by a covalent bond between a Glycine residue and a Leucine or Isoleucine residue, followed by an Alanine or Leucine, G~I/L-A/L, where ~ represents the bond cleaved (Fields 1991). What is curious about the degradation process itself, is that the collagen triple helix has a diameter of approximately 15Å (Figure I.1) while the catalytic domain of MMPs has a catalytic site that is only 5Å wide, suggesting that the folded structure of collagen cannot fit into the catalytic site, as seen in Figure I.4 (Fields 1991; Chung, Dinakarpandian et al. 2004; Lang, Braun et al. 2004).
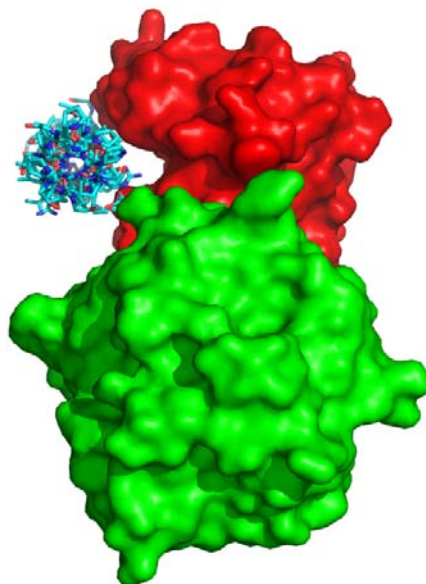
Figure I.4. Alignment of the collagen triple helix (15Å diameter) with the active site of MMP1 (5Å opening). The catalytic domain is shown in red and the hemopexin-like domain is shown in green. This figure was created with molecular visualization package Pymol (DeLano 2002).

Moreover, an analysis of the solvent accessible surface of the structure of collagen suggests that the scissile bond which is cleaved by collagenases is buried within the structure of collagen (Figure I.5), implying that even if the catalytic site could accommodate collagen, the bond that has to be cleaved is still inaccessible (Stultz 2002). As the catalytic site of MMPs cannot accommodate the folded structure of collagen, and the collagenase scissile bond is hidden within the triple-helical structure (Figure I.5), the precise mechanism of collagen cleavage by MMPs remains unclear.
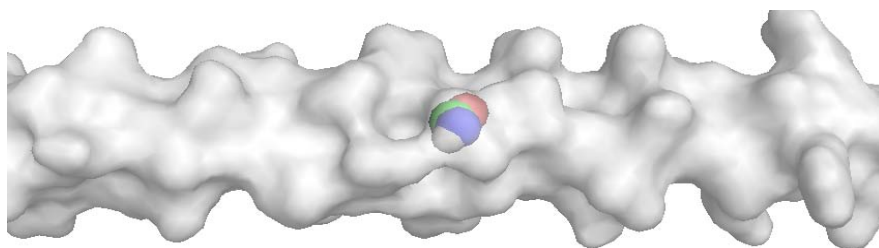
Figure I.5. Solvent accessibility of the scissile bond within the triple helix. Surface representation of the collagen triple helix, in transparent grey, with the four atoms forming the scissile bond shown in color. This figure was created with molecular visualization package VMD (Humphrey, Dalke et al. 1996).

There a number of hypotheses that explain the structural changes that must occur to facilitate collagen degradation. For example, it has been proposed that collagen is actively unwound by MMPs in a process known as "molecular tectonics" (Overall 2002; Bode and Maskos 2003). Figure I.6 shows a possible scenario in which a collagenase may actively unfold the collagen triple helix using a clamping mechanism. The problematic issue surrounding such mechanisms is that molecular tectonics requires the different MMP domains to undergo significant structural rearrangements that, in principle, would require large amounts of energy (Overall 2002). However, as collagen degradation does not require energy input *in vitro*, molecular tectonic theories have yet to be validated (Murphy and Knauper 1997; Overall 2002).
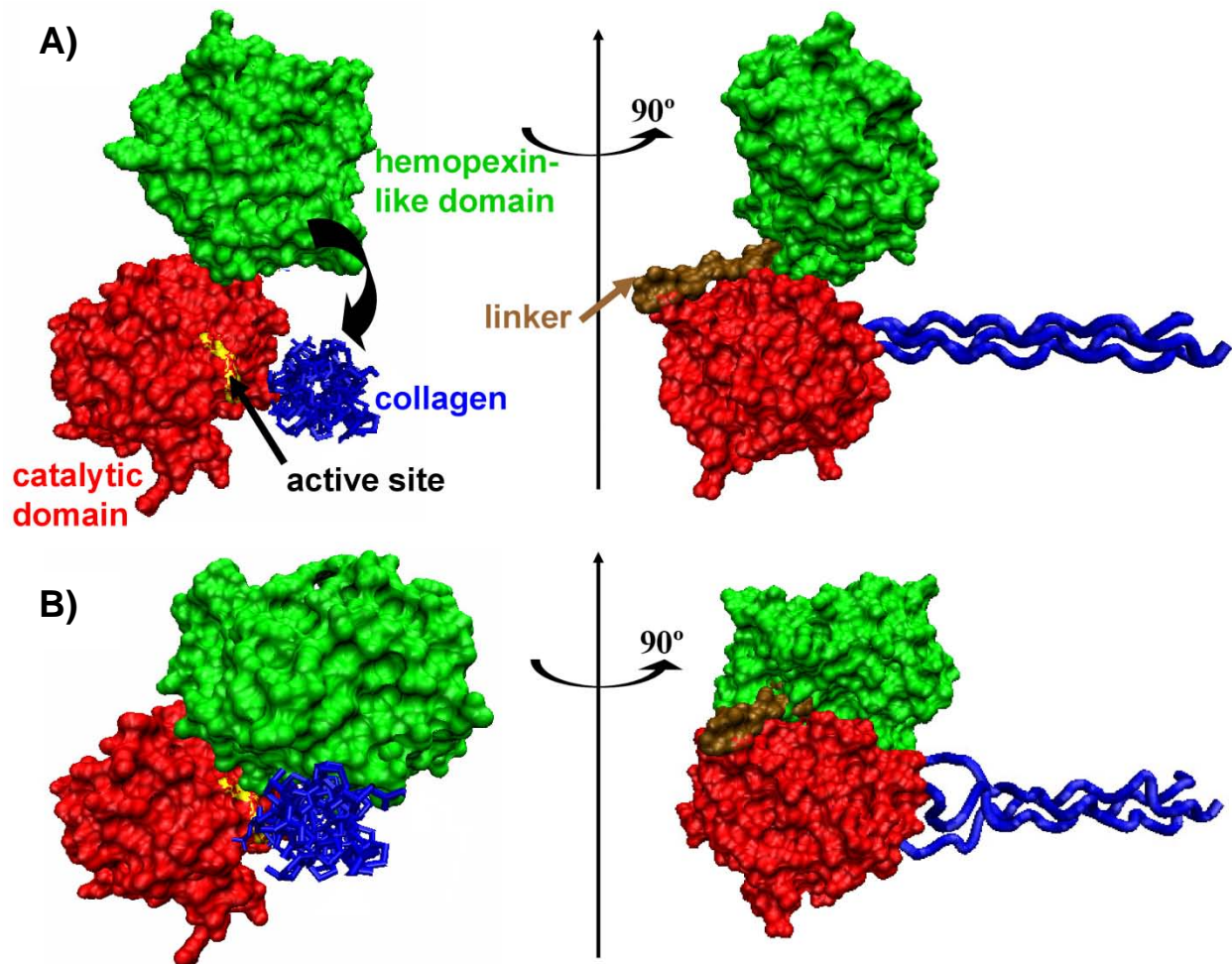
Figure I.6. Molecular tectonics - a clamping mechanism. Each figure shows a down the barrel (left) and longitudinal (right, 90° rotation) view with respect to a collagen monomer. In A), the fully folded triple helix (blue) sits near the catalytic domain (red) of MMP. It is apparent that the triple helix doesn't fit in the active site (yellow). In order to unfold collagen, the hemopexin-like domain (green) has to move towards the collagen triple helix and 'clamp' it; B) The triple helix of collagen is locally unfolded as a result of the concerted action of the catalytic and hemopexin-like domains (coordinated by the linker, in brown). The local unfolding then allows a single chain to fit into the active site. All figures were created with the molecular visualization package VMD (Humphrey, Dalke et al. 1996).

I.4.   Unfolded states and collagen degradation

One of the unifying themes of the research carried out in the Stultz lab is an interest in unfolded states (Huang and Stultz 2007; Huang and Stultz 2008; Yoon, Venkatachalam et al. 2009). When the unfolded state is the predominant state of a protein, we refer to them as natively unfolded proteins. These proteins take on highly flexible and conformationally heterogeneous states that determine function, although the function of many of these natively unfolded proteins is unknown. Natively unfolded proteins play a key role in diseases such as Alzheimer's disease (Aβ protein), Parkinson's disease (α-Synuclein) or tauopathy (tau protein) (Taylor, Hardy et al. 2002; Dyson and Wright 2005; Vendruscolo 2007). Computational methods are highly suitable to study these natively unfolded proteins (Vendruscolo 2007).

Natively folded proteins can also undergo motions and structural fluctuations that are important for enzyme catalysis, ligand binding and formation of protein complexes, amongst others (Vendruscolo 2007). This dynamic view of proteins is an important addition to the central dogma of modern protein science in which the sequence of a protein determines its structure, which in turn, determines its function. Indeed, the repetitive sequence of collagen and the presence of imino acids dictate collagen's triple helical structure (Brodsky and Persikov 2005). Consequently, its triple helical structure allows collagen to form fibrils and perform its structural function. In this thesis we study the structure of collagen regarded as a mobile, breathing entity. Therefore, the protein sequence determines the ensemble of structures that a protein adopts, and the dynamic fluctuations within the ensemble (Karplus and Kuriyan 2005). It is this combination of ensemble and dynamics that ultimately determines the function of a protein. Indeed, in the recent years it has become apparent that collagen itself may not be a rigid, inflexible, molecule.

That is, like most proteins, collagen can undergo small fluctuations in its structure at physiologic temperatures and unfold locally while maintaining a globally folded triple helix.

Experimentally, the existence of locally unfolded states has been suggested in studies of collagen and collagen like peptides, short peptides that fold into a triple helix. The collagen triple helix is typically thought to be resistant to proteases like trypsin or pepsin, a property which is used in the purification of acid soluble collagen at cold temperatures (Haralson and Hassell 1995). Nevertheless, it has been shown that, at temperatures below its melting temperature, type III collagen can be degraded by trypsin (Miller, Finch et al. 1976; Birkedal-Hansen, Taylor et al. 1985) or thermolysin (Wang, Chan et al. 1978) in the imino poor region downstream of the collagenase cleavage site. This suggests a disordered or interrupted triple helix.

Work with collagen-like peptides is also consistent with the existence of locally unfolded states in collagen. An NMR study showed that the central glycine residue in the imino poor region C-terminal to the type III collagenase cleavage site exchanges proton with solvent faster than a pure $(POG)_{10}$ peptide (a peptide containing imino acids in all X1 and X2 positions), suggesting a more labile triple helix at that position for the former peptide (Fan, Li et al. 1993). Another NMR study on a type I collagen-like peptide that contained an imino rich region and a relative imino poor region including the type I collagen cleavage site, revealed two independent structural domains. The imino rich region appeared to be well folded, while the imino poor region appeared to lack triple helical structure (Fiori, Sacca et al. 2002). Moreover, dynamical simulations on type I collagen and type III collagen-like peptides suggest that partial unfolding at room temperature does occur (Stultz 2002; Nerenberg and Stultz 2008; Salsas-Escat and Stultz 2009).

These data support the fact that local unfolding is an inherent property of the collagen triple helix and its sequence that occurs naturally without the action of collagenases. Taking this concept one step further, we formulate a conformational selection hypothesis for collagen, in which the existence of locally unfolded states, independent of collagenases, is necessary for collagen degradation. In other words, the interaction of preexisting unfolded states of collagen and collagenases is required for collagen degradation to occur. In fact, in a paper published during the course of this thesis, a two-state, conformational selection model was developed and was used to explain previous experimental observations in which deletion mutants of MMP1, containing only the catalytic domain, failed to cleave type I collagen at room temperature. These experimental results were interpreted to mean that locally unfolded states in collagen do not exist (Chung, Dinakarpandian et al. 2004; Nerenberg, Salsas-Escat et al. 2008). The two-state model is shown in Figure I.7. In short, collagen exists in an equilibrium between native (N) and vulnerable (V) states. While collagenases can bind to both states, yielding C_{NE} and C_{VE}, respectively, degradation can only occur when vulnerable states are bound. For details about this work, please refer to Appendix 1.

$$C_{NE} \underset{k_{on}^N}{\overset{k_{off}^N}{\rightleftarrows}} \quad N \underset{k_2}{\overset{k_1}{\rightleftarrows}} V \underset{k_{off}^V}{\overset{k_{on}^V}{\rightleftarrows}} C_{VE} \xrightarrow{k_{cat}^{VE}} P + E$$
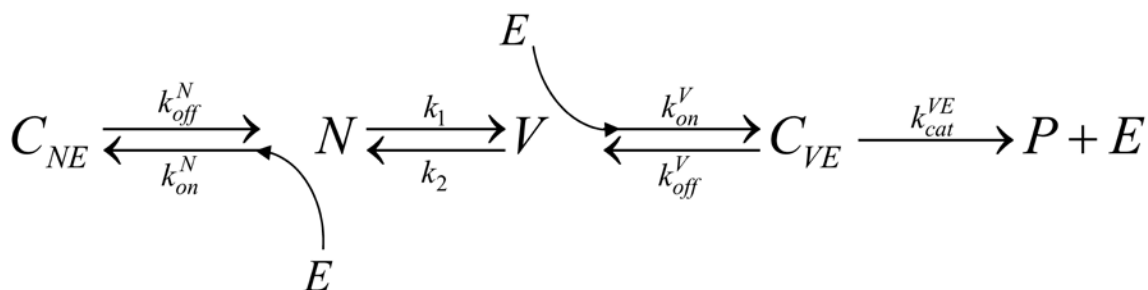
Figure I.7. Conformational selection mechanism for collagen degradation. N denotes the native state of collagen (i.e., its triple-helical structure); E denotes a collagenase; C_{NE} is the enzyme-substrate complex; C_{VE} denotes the complex formed between the enzyme E and the partially unfolded state V; and P denotes the products of collagenolysis. The model itself, as outlined here, makes no assumption about the form of collagen and therefore is applicable to both collagen in its monomeric form as well as in its fibrillar form. The individual rate constants for each transition are shown.

Given all the data suggesting the existence of unfolded states in collagen, it may be that the classic triple-helical structure of collagen found in many textbooks does not accurately represent its structure under physiologic conditions. Collagen degradation in the body occurs at 37°C and mainly in the fibrillar state. The detailed study of collagen structure and degradation under these conditions is experimentally and computationally complex. However, it has been shown that collagen can be cleaved by collagenases *in vitro* at room temperature (Chung, Dinakarpandian et al. 2004). For this reason, we restrict our computational and experimental *in vitro* studies to room temperature, which is well below collagen's melting temperature (Leikina, Mertts et al. 2002). Using computational tools, in Chapters 1 and 2 we intend to develop a deeper understanding of the normal thermal fluctuations that the collagen structure experiences. However, studying conformational fluctuations in proteins is often a difficult task as the thermal fluctuations of interest are often quite small. In this thesis we demonstrate how computational tools can be used to probe these small structural changes at an atomistic level of detail.

I.5. The problem of cleavage site specificity

In addition to the aforementioned unfolding question of collagen degradation, there is another aspect of collagen degradation that is of much debate. This is the issue of collagenase cleavage site specificity. Collagenases cleave triple helical collagen at sites defined by the sequence G~I/L-A/L. While several of these sites exist in the sequence of collagen (Figure I.8), only one in each chain is cleaved by collagenases in triple helical collagen (Fields 1991).

Certain proteases have, in general, well defined sequence specificities. For example, trypsin cleaves after positively charged residues (Olsen, Ong et al. 2004), pepsin preferentially after phenyalanine and leucine residues (Fruton and Bergmann 1938; Baker 1951) and chymotrypsin after aromatic residues (Hudaky, Kaslik et al. 1999). Therefore, from the analysis of the sequence of a target protein it can be predicted what fragments will result from the incubation with any of these proteases, given enough time for the protease to completely process its target. Limited proteolysis relies on the short exposure of general proteases to a target protein so that only the most accessible peptide bonds that satisfy the specificity requirements of the protease in question will be cleaved (Hubbard 1998). For example, limited exposure of trypsin to a protein will result in cuts at the most exposed positively charged residues, while arginine or lysine residues buried in the structure will not be readily cut. Only when the partially cut protein starts unfolding, these otherwise inaccessible residues will be accessed and cleaved by the protease. In light of this, it is apparent that accessibility of a peptide bond of interest to the active site of a protease is a key to the cleavage of such peptide bond.

An explanation of the issue of collagenase cleavage site specificity can be formulated using the same approach. Only collagen triplets of the sequence G~I/L-A/L that are accessible to the collagenase active site will be cleaved. The main difference with the description of limited

proteolysis is that, in principle, there are no G~I/L-A/L triplets totally buried in the collagen structure, since collagen is a triple helical protein, not a globular one. Due to its triple helical structure, all side chains in collagen are solvent exposed, which is not the case in a globular protein. Nevertheless, in Section I.3 it has been discussed that the scissile bond in the canonical triple helical structure is hidden from solvent, and that the collagen triple helix does not fit into the active site of collagenases. The existence of locally unfolded states, discussed in Section I.4, may play a key role in allowing certain scissile bonds to become accessible to collagenases. While all the potential cleavage sites in collagen share the G~I/L-A/L sequence, they are surrounded by different stretches of amino acids. These differences in the surrounding sequences may result in different local conformations at each site. It is possible that the locally unfolded states at each of the potential cleavage sites are different, and that this helps determine cleavage site specificity.

Type I collagen α1 chain sequence (UniProtKB-SwissProt accession no. P02452)

QLSYGYDEKSTGGISVPGPMGPSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKN
GDDGEAGKPGRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDAKGDAGPAGPKGEPGSPGENGAP
GQMGPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGPRGSEGPQ
GVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAP**GIA**GAPGFPGARGPSGPQGPGGPPGPKGNSGEP
GAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPK
GPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGAR
GQAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQ
GLPGPAGPPGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDA
GAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK
GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPP
GPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRP
GEVGPPGPPGPAGEKGSPGADGPAGAPGTPGPQ**GIA**GQRGVVGLPGQRGERGFPGLPGPSGEPGKQGPS
GASGERGPPGPMGPP**GLA**GPPGESGREGAPGAEGSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGPV
GPAGKSGDRGETGPAGPAGPVGPVGARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSP
GEQGPSGASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPSAG
FDFSFLPQPPQEKAHDGGRYYRA

Type I collagen α2 chain sequence (UniProtKB-SwissProt accession no. P08123)

QYDGKGVGLGPGPMGLMGPRGPPGAAGAPGPQGFQGPAGEPGEPGQTGPAGARGPAGPPGKAGEDGHP
GKPGRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTPGQTGAR
GLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGAVGNAGPAGPAGPRGEV
GLPGLSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGATGARGLVGEPGPAGSK
GESGNKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPPGPPGLRGSPGSRGLPGADGRAGVMGPPGSR
GASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGEPGNI
GFPGPKGPTGDPGKNGDKGHA**GLA**GARGAPGPDGNNGAQGPPGPQGVQGGKGEQGPAGPPGFQGLPGPS
GPAGEVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGSRGPSGPPGPDGNKGEPGVVGAV
GTAGPSGPSGLPGERGAAGIPGGKGEKGEPGLRGEIGNPGRDGARGAPGAVGAPGPAGATGDRGEAGAA
GPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGAKGPKGENGVVGPTGPVGAAGPAGPN
GPPGPAGSRGDGGPPGMTGFPGAAGRTGPPGPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRTGEVGAV
GPPGFAGEKGPSGEAGTAGPPGTPGPQ**GLL**GAP**GIL**GLPGSRGERGLPGVAGAVGEPGPL**GIA**GPPGAR
GPPGAVGSPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKGERGYPGNIGPVGAAGAPGPHGPVGPAGKH
GNRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKGEPGEKGPRGLPGLKGHNGLQGLP**GIA**GHHGDQGAP
GSVGPAGPRGPAGPSGPAGKDGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPGPPG

Type III collagen sequence (UniProtKB-SwissProt accession no. P02461)

QYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGPPGAI
GPSGPAGKDGESGRPGRPGERGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGLKGENGLP
GENGAPGPMGPRGAPGERGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLR
GGAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPN
GIPGEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPP
GPSGPRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPP
GPQGLQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGAPGERGPP**GLA**GAPGLRGGAGPPGPE
GGKGAAGPPGPPGAAGTPGLQGMPGERGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPA
GQPGDKGEGGAPGLP**GIA**GPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVA
GPPGGSGPAGPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNT
GAPGSPGVSGPKGDAGQPGEKGSPGAQGPPGAPGPL**GIA**GITGAR**GLA**GPPGMPGPRGSPGPQGVKGES
GKPGANGLSGERGPPGPQGLP**GLA**GTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHP
GPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAP
GSPGPAGQQGAIGSPGPAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPP
GPPGAPGPCCGGVGAAAIAGIGGEKAGGFAPYYG

Figure I.8. Amino acid sequences of human types I and III collagens. The actual collagenase cleavage sites are highlighted in green, while the potential cleavage sites are highlighted in red. These sites are not cleaved by collagenases in triple helical collagen. Since these sequences were obtained from genomic data, all hydroxyproline residues are being shown as proline.

I.6. Hypothesis and aims of this thesis

The main hypothesis of this thesis is that locally unfolded states in collagen exist in the absence of collagenases. Degradation occurs when collagenases cleave these locally unfolded states. Also, it is hypothesized that the nature of the locally unfolded states at each cleavage site is a determinant of cleavage site specificity.

In order to test these hypotheses, three main chapters are present in this work. The first chapter "Influence of the C-terminal Sequence to a Cleavage Site on Local Unfolding" aims to study the relationship between the sequence C-terminal to two potential cleavage sites and the existence of unfolded states, using a combination of computational and experimental methods. The second chapter "Characterizing the Conformational Ensemble of Collagen Cleavage Sites" takes the computational methods used in Chapter 1 one step further, expanding the study to multiple cleavage sites and the sequences upstream and downstream to these cleavage sites. In particular, it aims to study all the potential cleavage sites in type III collagen using computational methods, with hopes to understand the nature of the different locally unfolded states that occur at each cleavage site. The third chapter "Investigating the Influence of Unfolded States in Cleavage Site Specificity" aims to establish if the existence of unfolded states is a determinant of cleavage site specificity, by computational and experimental methods.

In addition, two appendices are present in this thesis. In Appendix 1, the early development of the two-state collagen degradation model and testing of its predictive power against earlier experimental data is demonstrated. Finally, Appendix 2 aims to test the two-state collagen degradation model developed in Appendix 1. The goal is to be able to confirm experimental predictions of the two-state degradation model and also gain insight into the thermodynamics of unfolded states and of the collagen-collagenase interaction.

Chapter 1:
Influence of the C-terminal Sequence to a Cleavage Site on Local Unfolding

Chapter 1 is adapted from: Salsas-Escat, R. and Stultz C.M. (2009). "The Molecular Mechanics of Collagen Degradation: Implications for Human Desease." <u>Experimental Mechanics</u> **49**(1): 65-77.

1.1 Introduction

Although several G~I/L-A/L triplets can be found in collagen, only one of these triplets in native collagen is cleaved by matrix metalloproteinases. That is, peptide bond hydrolysis occurs at one unique site despite the fact that several potential cleavage sites exist. A comparison of collagen sequences from several species suggests that the amino acid sequence surrounding the true collagenase cleavage site has unique sequence characteristics (Fields 1991). Namely, the true collagenase cleavage site has: 1) a relatively high imino acid content (4 or more imino acids in 12 residues) in the 4 triplets N-terminal (upstream) to the cleavage site; 2) a low imino acid content (less than 2 imino acids in 12 residues) in the 4 triplets C-terminal (downstream) to the cleavage site; and 3) an arginine residue downstream from the scissile bond (in position $P_5^{'}$ or $P_8^{'}$). While these observations provide insight into sequence properties that are correlated with the true cleavage site, it is not obvious why such sequence differences enable the actual cleavage site to be recognized and cleaved by MMPs. Since it has been established that the presence of imino acids stabilizes the triple helix (Persikov, Ramshaw et al. 2000), we focus on the second sequence rule presented. We hypothesize that a low content of imino acids in the 4 triplets C-terminal to the cleavage site is necessary to allow the existence of locally unfolded states. In Chapter 1 we investigate the influence of the C-terminal sequence to a cleavage site on the nature of the locally unfolded states existing next to such cleavage site.

Most biochemical experiments measure average properties from a large collection of molecules. Spectroscopic measurements, for example, are performed on solutions containing micromolar concentrations of protein, which have on the order of 10^{20} molecules. Single molecule experiments can be more informative but even so, they can be difficult to perform and rarely yield information at an atomistic level of detail. By contrast, molecular dynamics (MD)

simulations have the advantage of generating detailed pictures of molecular phenomena at an arbitrary level of detail (Schlick 2000). These characteristics make MD simulations well-suited to study the small conformational fluctuations that may occur in collagen. These fluctuations are of particular interest as they may facilitate the recognition of collagen by proteases. To gain insight into the conformational fluctuations of collagen, we could compute trajectories using a model of the collagen molecule. As collagen itself contains over 3000 amino acids and well over 30,000 atoms, detailed molecular simulations on a system of this size would require extensive computational resources. Consequently, we use a reduced representation that focuses on the amino acids near the true collagenase cleavage site. In order to achieve that, we simulate collagen-like peptides, short peptides that contain a biologically relevant sequence surrounded by proline-hydroxyproline-glycine (POG) repeats to ensure that the peptides fold in solution. This also allows us to perform experiments on the same peptides that are being simulated.

1.2 Results

The peptide T3-785 (IP1 in this work) models an imino-poor segment from type III collagen which is downstream from the unique collagenase cleavage site (Figure 1.1). T3-785 has been shown to adopt a triple helical structure at temperatures well below body temperature (Kramer, Bella et al. 1999). The other peptide studied in this work, IP2, models a region of collagen that is adjacent to a bond that is not cleaved in native collagen, but is cleaved in denatured (i.e., unfolded) collagen (Fields 1991) (Figure 1.1) - a finding which suggests that the conformation of this region is not partially unfolded in solution. We demonstrate how data obtained from MD simulations can be used to make useful inferences into the relative stability of these two peptides and that such observations ultimately lead to insights regarding the mechanism of collagenolysis.

A)

```
C3:  -A-O-G-P-L-G~I-A-G-I-T-G-A-R-G-L-A-G-P-O-G-M-O-G-P-R-G-S-O-G-
IP1: P-O-G-P-O-G*P-O-G-I-T-G-A-R-G-L-A-G-P-O-G-P-O-G-P-O-G-P-O-G
```

B)

```
C3:  -E-R-G-P-O-G^L-A-G-A-O-G-L-R-G-G-A-G-P-O-G-P-E-G-G-K-G-A-A-G-
IP2: P-O-G-P-O-G*P-O-G-A-O-G-L-R-G-G-A-G-P-O-G-P-O-G-P-O-G-P-O-G
```

Figure 1.1. Sequences of peptides IP1 and IP2. A) Alignment of residues 776-805 in type III collagen (C3) and IP1 (T3-785) peptide. The symbol O denotes the residue hydroxyproline. The underlined residues correspond to the imino-poor region. The covalent bond cleaved by collagenases in type III collagen is denoted by a ~. The corresponding position in IP1 is denoted by a *. B) Alignment of residues 515-544 in type III collagen and IP2 peptide. In this case, the potential cleavage site is denoted by a ^. This site is not cleaved in native collagen. The corresponding position in IP2 is denoted by a *.

MD simulations use a potential function to generate molecular trajectories. While the potential energy of the system is an important parameter in this regard, it is the free energy of the system that is most important for understanding physical properties. Potential energies tell us

about the energy of an individual conformation - a parameter that is inaccessible experimentally. By contrast, free energies tell us about the behavior of large collections of molecules. As typical laboratory experiments are performed on solutions containing molecules at nano-to-micromolar concentrations, methods that compute free energies allow us to directly correlate calculated quantities with experiment. In this regard, the potential of mean force plots (pmfs, also known as free-energy profiles) of IP1 and IP2 are of particular interest. From the free energy profiles, we can estimate the relative stability of different conformers, including partially unfolded states.

Figure 1.2 shows the calculated free energy profiles of both peptides. The x-axis displays the radius of gyration, γ, - a convenient measure which indicates the extent in which a protein is unfolded - and the y-axis corresponds to the free energy. The pmf for IP1 has two minima, one at $\gamma=18.5\text{Å}$, close to that of the crystal structure of T3-785, and another at $\gamma=18.9\text{Å}$. The pmf for IP2 only has one minimum at $\gamma=18.2\text{Å}$ - a value suggesting that the structure of this peptide in solution is more compact than the crystallographic structure. Qualitatively these data suggest that IP1 is conformationally labile relative to IP2. However, an added advantage of these calculations is that the pmf itself is a rigorous description of the conformational landscape of each peptide. Once the pmf is known, all properties that depend on the peptide's conformation can be calculated. To quantify the conformational flexibility of each sequence, we compute the conformational entropy of each pmf; i.e., a quantity that describes the number of accessible states the peptide can have (see Methods). We find that $T\Delta S =200$ cal/mol higher for IP1, confirming that the conformational landscape of IP1 has higher entropy than the conformational landscape of IP2. However, the overall value of this entropy difference is quite small. To put this value into perspective, we note that estimates of the entropy changes associated with complete unfolding of globular proteins is on the order of 1500 to 1800 cal/mol/residue at 300K (Privalov

and Dragan 2007). Therefore the entropic difference between the conformational landscape of IP1 and IP2 is less than the entropic change associated with the complete unfolding of a single residue. These findings suggest that while the IP1 can adopt two distinct states, neither of these states is completely unfolded and the structural differences between the two conformers are likely to be quite localized.
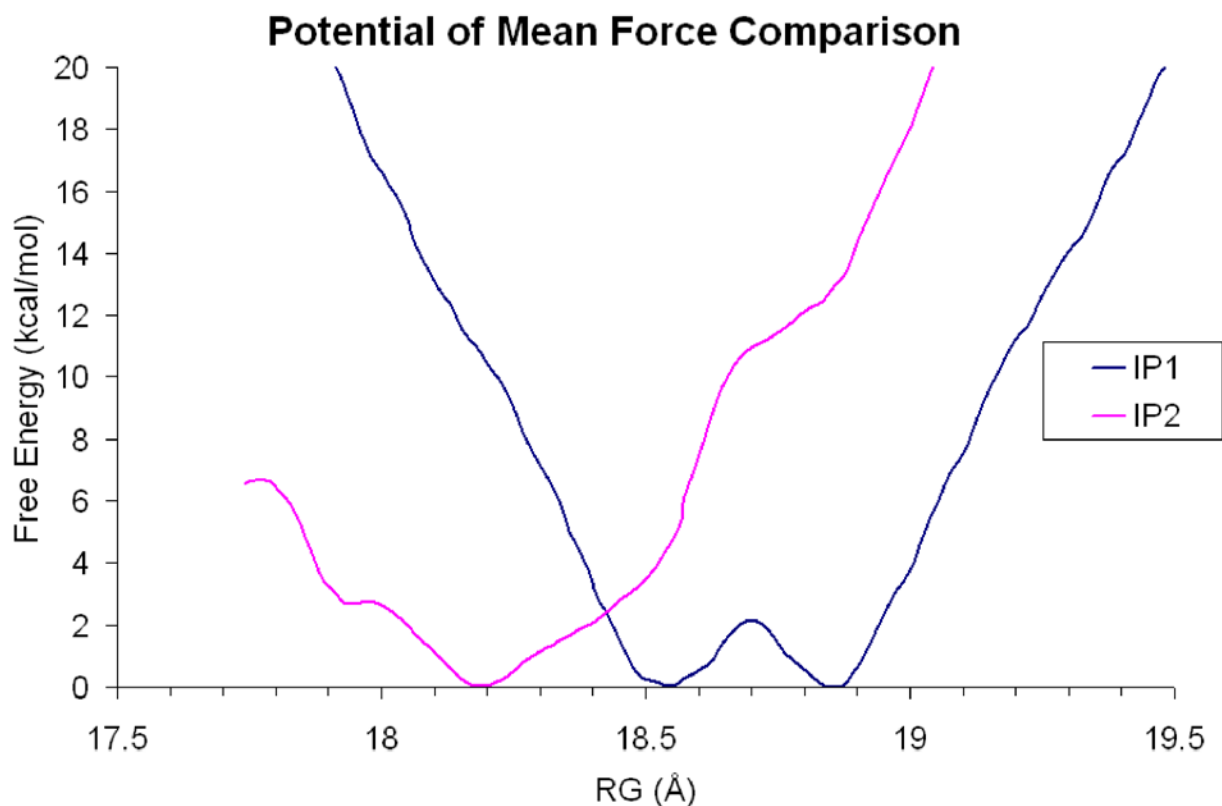
Figure 1.2. Comparison of potential of mean force for IP1 (blue) and IP2 (magenta). The pmf for IP1 is obtained from a previous work (Stultz 2002). IP1 presents 2 minima, one near the radius of gyration of the crystal structure (18.5Å) and a second near 18.9Å. IP2 presents only one minimum, near 18.2Å.

An analysis of representative structures from the simulations confirms this observation. While the structure of IP1 at the minimum corresponding to $\gamma=18.5\text{Å}$ is quite similar to the crystallographic structure, the structure corresponding to $\gamma=18.9\text{Å}$ has a relatively localized region that is unfolded (Figure 1.3 A and B) (Kramer, Bella et al. 1999; Stultz 2002). Interestingly, this region is adjacent to the collagenase cleavage site. Given this, we refer to conformers, like that shown in Figure 1.3 B, as vulnerable because these structures are, in principle, more vulnerable to cleavage by collagenases. To quantify structural differences between these conformers we compute the change in solvent accessibility surface (SAS) for backbone residues of the unfolded chain in each structure with respect to the well folded triple helix (Figure 1.3 D). To calculate the solvent accessibility, we use the Lee and Richards algorithm (Lee and Richards 1971), with a sphere radius of 1.4Å. Positive peaks indicate that the structure of a given residue in the conformation of interest is more exposed relative to the conformation of that residue in the crystal structure. These data corroborate the qualitative observations noted in Figure 1.3 A and B. That is, regions near the true collagenase cleavage site are relatively solvent exposed and accessible.

A) T3-785: RG=18.5 Å

B) IP1: RG=18.9 Å

C) IP2: RG=18.2 Å

D)
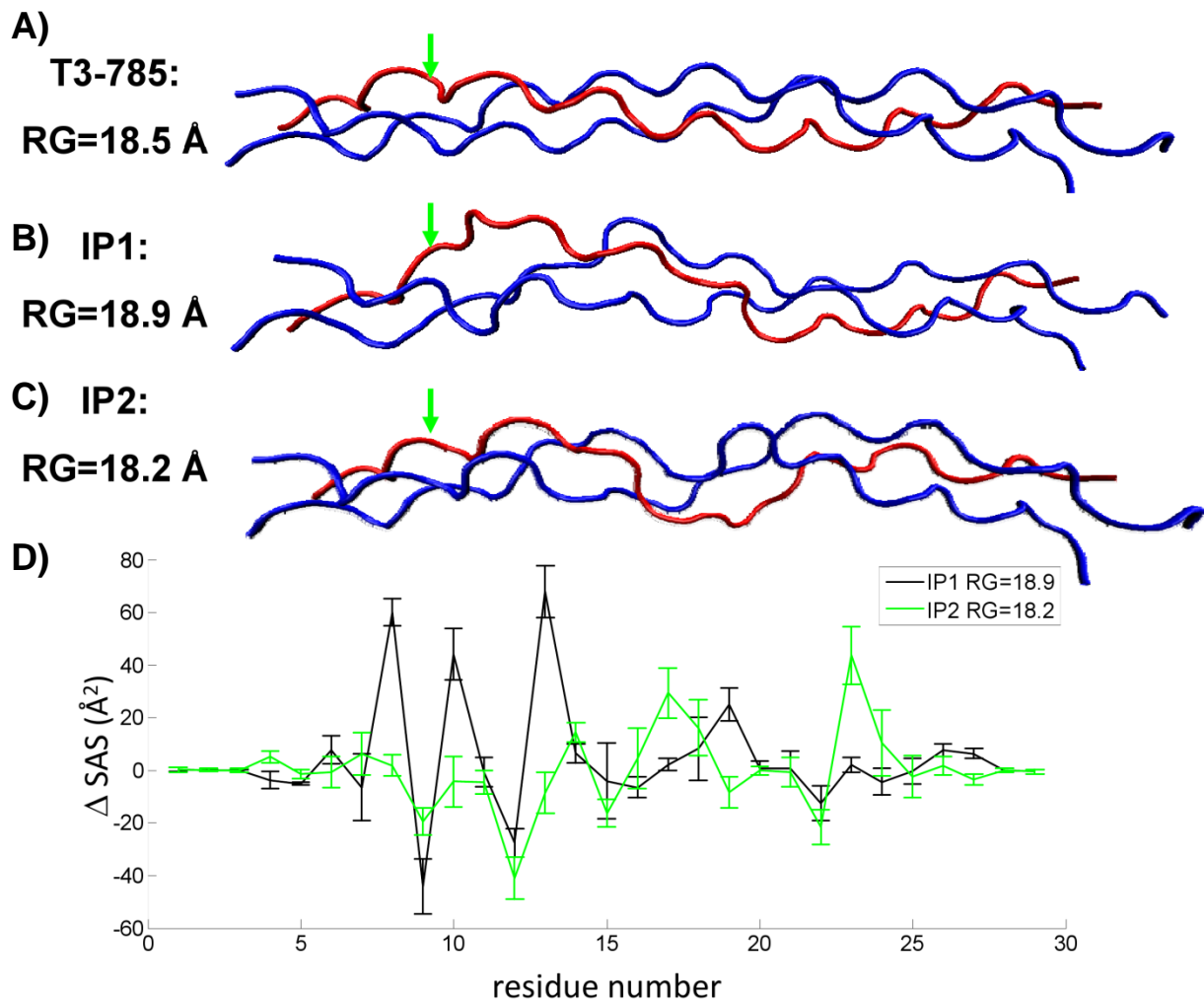
Figure 1.3. Comparison of the structural ensembles of each energy minima in Figure 1.2. A) Crystal structure of T3-785. The chain which is unfolded in the molecular simulations is shown in red. The green arrow points to the location where the scissile bond would be in type III collagen. B) Representative structure for IP1 at the minimum corresponding to radius of gyration 18.9Å. C) Representative structure for IP2 at the minimum corresponding to radius of gyration 18.2Å. D) Solvent accessibility surface (SAS) differences for the vulnerable state of IP1 (black) and for the structure lowest energy state of IP2 (green). Positive peaks indicate higher solvent accessibility than in the crystal structure. Figures A, B and C have been created with the molecular visualization package VMD (Humphrey, Dalke et al. 1996).

The simulations predict that peptide IP1 can sample compact as wells as unfolded conformations near room temperature while IP2 samples mainly compact confromations in solution. To determine whether these observations are consistent with the structure of these peptides *in vitro*, we obtain experimental estimates of the relative stability of both peptides. We begin by noting that the melting point of a solution containing a specific protein is the temperature in which half of the protein molecules in solution become unfolded. The more stable a given protein, the higher its melting temperature. Our data suggest that at temperatures near room temperature, IP1 starts to unfold, while IP2 retains much native like character. Consequently, we would expect IP2 to have a higher melting temperature than IP1.

Melting curves for IP1 and IP2 are obtained at variable heating rates (Figure 1.4 A and B). From each melting curve, one can estimate the melting temperature (Figure 1.4 C). These data demonstrate that the melting curve is highly dependent on the heating rate (Privalov 1982). Ideally, the quantity that is of most interest is the melting temperature obtained by heating the peptide at an infinitely slow rate. Consequently, we use the various melting curves to extrapolate a melting curve corresponding to a heating rate of 0ºC/h (Figure 1.4 C). The resulting melting temperatures for IP1 and IP2 (Table 1.1) demonstrate that IP1 does, in fact, have a slightly lower melting temperature, suggesting that it unfolds first as the temperature is raised.
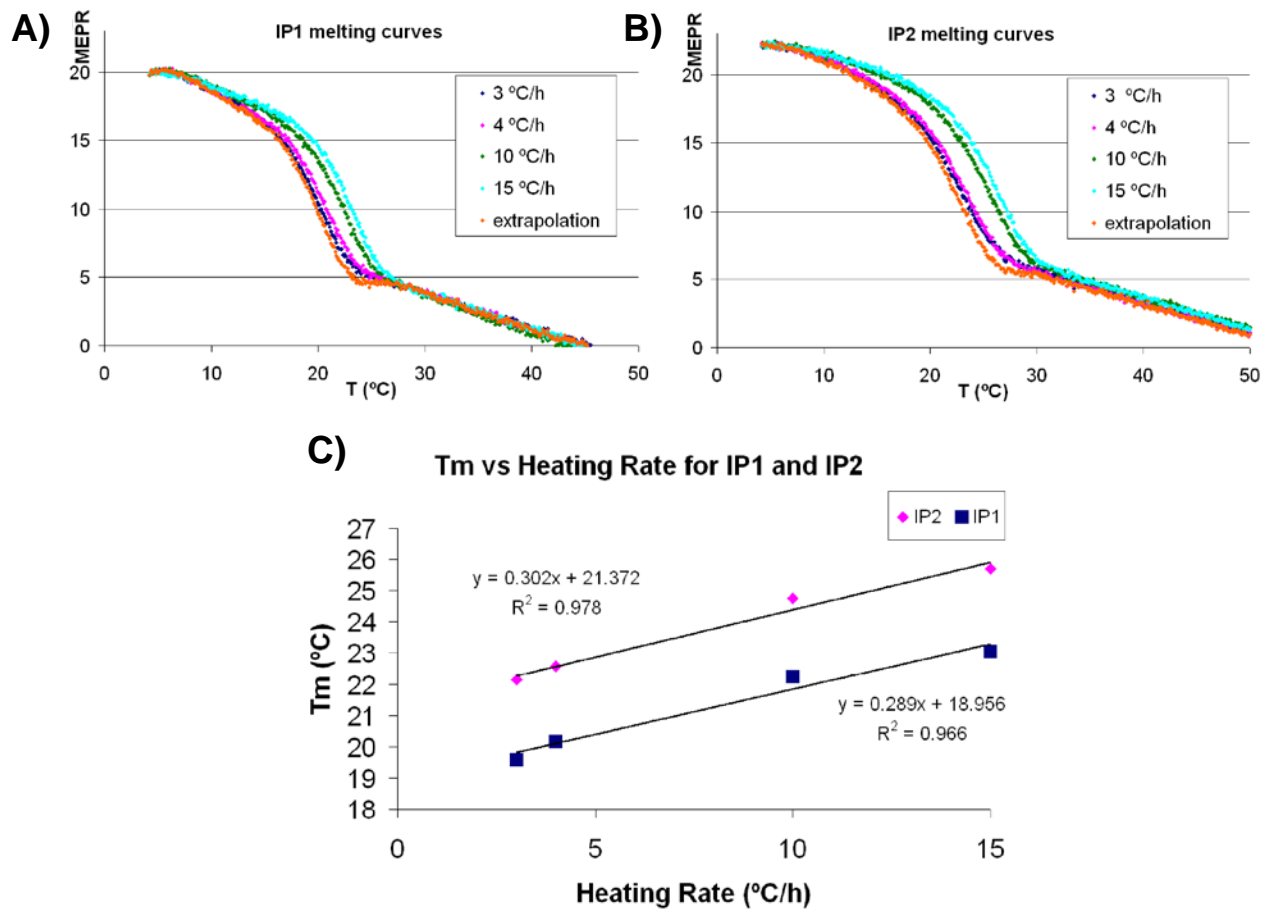
Figure 1.4. Melting curves at different heating rates for A) IP1 and B) IP2. C) Melting point extrapolation for IP1 and IP2.

Table 1.1 Melting temperature for peptides IP1 and IP2

	Melting temperature (°C)
IP1	18.9±0.2
IP2	21.4±0.1

1.3 Discussion

Proteins are truly fascinating. They are complex molecules that are responsible for a myriad of diverse functions. To perform their prescribed functions, they often need to exhibit some degree of flexibility during their biological lifetime (Bahar and Rader 2005). Such structural fluctuations range from small thermal vibrations, where protein atoms oscillate about some average position, to large conformational changes, like the structural alterations associated with the myosin 'molecular motor' (Sellers and Veigel 2006). In practice, it is often quite difficult to visualize and/or quantify the degree of structural flexibility that is available to any given protein. Protein structures are usually determined by well-established experimental techniques such as x-ray crystallography and NMR spectroscopy (Voet and Voet 2004; Nelson and Cox 2005). X-ray crystallography uses x-ray radiation to obtain a diffraction pattern of a crystal which contains a large number of protein structures that are oriented in an orderly and recurring pattern. As protein crystals may contain a relatively heterogeneous population of conformations, meaningful x-ray crystal structures are only obtained when the different structures within the crystal are all quite similar. Fortunately, this is the case for most proteins. That is, under physiologic conditions most proteins only make small deviations from their average structure. To quantify the degree of thermal fluctuations within the crystal, x-ray crystallographers often report a measure of the thermal motion of atoms within the structure; i.e., B factors. The average structure, together with these B-factors, contains a wealth of information as they encode the magnitude of the structural fluctuations that each atom in a protein undergoes in the environment of a protein crystal. Nevertheless, decoding their information content can be quite problematic. Artifacts caused by disorder within a protein crystal can lead to erroneous conclusions that are based on analyses of B-factors. NMR experiments, by contrast, can

sometimes give insights into the degree of structural flexibility that is inherent to specific regions of a protein. However, due to technical limitations it is not always feasible to perform these experiments on an entire protein. Consequently, it is not always possible to make detailed conclusions about conformational flexibility in proteins from experimental data alone (Brooks III, Karplus et al. 1988; Voet and Voet 2004; Nelson and Cox 2005).

Given these observations, it is likely that crystallographic structures of collagen-like peptides within the structural database do not fully capture the extent of conformational flexibility in the vicinity of the collagenases cleavage site. This is especially poignant given that crystal structures of collagen like-peptides are obtained under conditions that dramatically differ from physiologic states (Kramer, Bella et al. 1999). A number of experimental studies suggest that this may in fact be the case (Fiori, Sacca et al. 2002; Sacca, Fiori et al. 2003; Brodsky and Persikov 2005). While NMR and gross structural experiments have implied that collagen may be flexible in some areas, they provide only limited information on the detailed conformational changes that are involved.

In this chapter, detailed molecular simulations on collagen-like peptides that model two distinct regions of collagen reveal the necessary structural characteristics that a sequence must have to enable degradation to occur. In particular, our data suggest that the amino acid sequence near the true collagenase cleavage site is conformationally labile relative to another sequence next to a potential cleavage site that is not degraded. Experimentally, the melting temperature for IP1 is significantly lower than IP2, a finding which agrees with our observations; i.e., the pmfs imply that unfolded states of IP1 are relatively prevalent at temperatures near room temperature while IP2 retains a more compact structure at such temperatures. Analyzing the most important sequence differences between both peptides, both IP1 and IP2 are relatively imino-poor in

46

content but IP2 has a hydroxyproline residue in its 11th position (Figure 1.1). This likely confers extra stability, preventing unfolding near the collagenase cleavage site. These results suggest that small variations in sequence next to a cleavage site, especially in the imino acid context, can lead to significant differences in stability and conformational dynamics in the neighborhood of the cleavage site.

1.4　Methods

1.4.1　Molecular dynamics simulations

Dynamical simulations of molecules have a long history. Early atomistic simulations were used to calculate scattering trajectories associated with simple chemical reactions (Hirschfelder, Eyring et al. 1936). Further applications were focused on hard sphere liquids and near ideal gasses (Alder and Wainwright 1959; Rahman 1964). Simulations of protein systems first appeared in the early 1970s and demonstrated that previously unappreciated insights into protein dynamics could be obtained from detailed simulations of modest length (Levitt and Warshel 1975; McCammon, Gelin et al. 1977). Today, molecular dynamics (MD) simulations are broadly applied to different proteins to make observations that would otherwise not have been possible with standard biochemical experiments (Hansson, Oostenbrink et al. 2002; Karplus and McCammon 2002). Next, we briefly outline the main features of a MD simulation and the type of data that is needed to generate molecular trajectories.

1.4.1.1 The initial model

The first step in simulating a molecular system is to obtain an appropriate model of the system of interest. Many types of models exist to represent a biomolecular system, ranging from the very detailed, which model subatomic particles, to the very simplified which represent groups of atoms as single spheres. In the present case, we model atoms in the molecule as points in space. Each atom has an associated mass and an associated partial charge. As the computational cost of the simulation is directly related to the number of atoms that are explicitly modeled, we only model heavy atoms (e.g., C, N, O, and S) and atoms which are covalently bonded to the electronegative atoms N, O and S. In other words, hydrogen atoms that are bound

to carbon atoms (also known as non-polar hydrogens) are not explicitly represented. Instead, hydrogen-carbon clusters are represented by a single point. The mass and van der Waals radius of these points is adjusted to include both the mass and van der Waals radius of a carbon atom and the bonded hydrogen atoms. This representation is referred to as a polar hydrogen set. Its main advantage is that it reduces the number of atoms that need to be explicitly simulated, thereby reducing the computational cost. Nevertheless, the model explicitly represents hydrogen atoms that can engage in non-covalent hydrogen bonds.

In order to perform a molecular dynamics simulation, the three dimensional coordinates of the atoms of the system must be known. They are usually obtained from structures determined by x-ray crystallography, and less often by nuclear magnetic resonance (NMR), and can be found in the Protein Data Bank repository (www.pdb.org). As hydrogen atoms are typically not seen in x-ray crystallographic structures, the position of polar hydrogens must be inferred from the coordinates of the heavy atoms that they are bound to. Once the positions of all relevant atoms are known, the system can be solvated to model the surrounding environment of the protein. In this work, we use an explicit model of water where each water molecule contains an oxygen atom bonded to two hydrogen atoms.

1.4.1.2 The potential energy function

At the heart of all molecular dynamics simulations is the potential energy function. Once a potential energy function is defined for any given protein conformation, then trajectories can be calculated using the corresponding Euler-Lagrange equation.

The potential energy function can be divided into interactions between atoms that are connected through covalent bonds and interactions between atoms that are separated by 4 or

more covalent bonds. The latter set of interactions is often referred to as non-bonded interactions:

$$V = V_{BONDED} + V_{NON-BONDED} \tag{1.1}$$
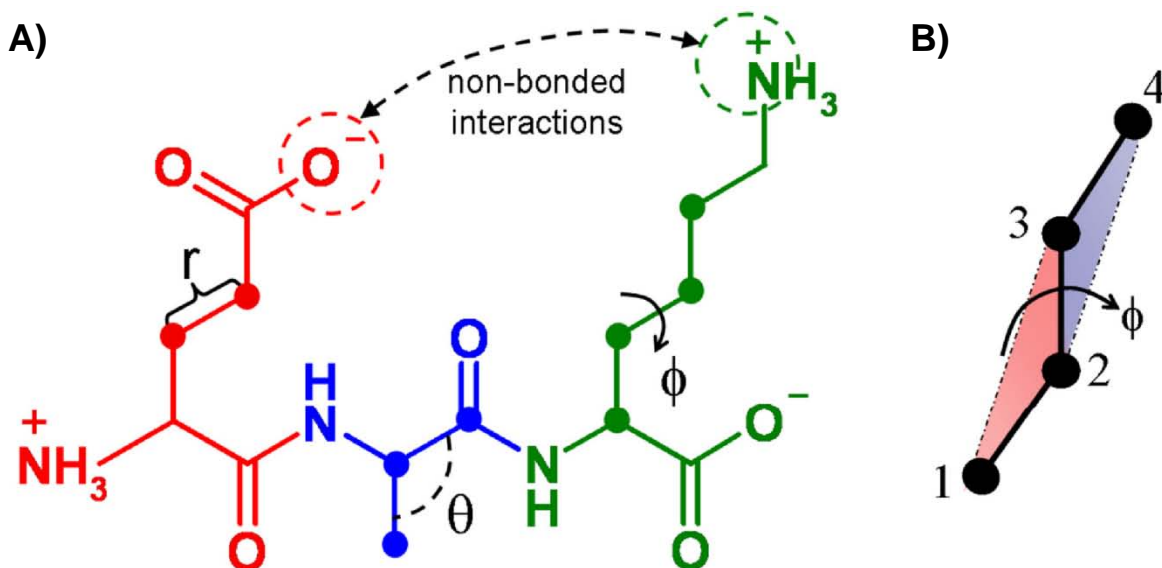
Figure 1.5. Potential energy interactions. A) Interactions between atoms connected by covalent bonds in a Glutamate-Alanine-Lysine tripeptide which are included in typical potential energy models. In glutamate (red), two atoms represented by red dots are connected by a bond. In alanine (blue), three atoms represented by blue dots form an angle. In lysine (green), four atoms represented by green dots define a dihedral angle. Also, non-bonded interactions between circled atoms, separated by more than 3 covalent bonds, are shown. B) Schematic representation of a dihedral angle. It is defined by the angle between the plane formed by atoms 1, 2 and 3 (red plane), and the plane formed by atoms 2, 3 and 4 (blue plane).

Bonded interactions

Bonded interactions are associated with atoms that are connected through covalent bonds (Figure 1.5). They can be subdivided as follows:

$$\begin{aligned} V_{BONDED}\,(r_1, &\ldots, r_{N-1}, \theta_1, \ldots, \theta_{N-2}, \varphi_1, \ldots, \varphi_{N-3}) \\ &= V_{BOND}\,(r_1, \ldots, r_{N-1}) + V_{ANGLE}\,(\theta_1, \ldots, \theta_{N-2}) \\ &+ V_{DIHEDRAL}\,(\phi_1, \ldots, \phi_{N-3}) \end{aligned} \tag{1.2}$$

where $r_1, \ldots, r_{N-1}$ denotes the distances corresponding to the N-1 bonds in the system (where N is the number of atoms in the protein), respectively. Similarly, $\theta_1, \ldots, \theta_{N-2}$ and $\phi_1, \ldots, \phi_{N-3}$ denote the N-2 bond angles and N-3 dihedral angles in the system, respectively. For simplicity we express the potential energy as a function of the internal coordinates of the system (e.g., atom distances, bond angle measurements, etc.) instead of the Cartesian coordinates. We recognize, however, that all of these equations could be written in terms of the more traditional Cartesian coordinates without loss of generality.

The bond stretching term models the energy associated with small vibrations about covalent bonds. Typically this term is modeled with a harmonic function of the form:

$$V_{BOND}(r_i) = \sum_i K_{b,i}\left(r_i - r_i^{eq}\right)^2 \qquad (1.3)$$

where r_i^{eq} is the equilibrium bond length of the ith bond and $K_{b,i}$ is the force constant for the ith bond. Force constants and equilibrium bond lengths are chosen to agree with data obtained from x-ray crystallography and infra-red spectroscopy.

The energetic contribution associated with bending the bond angle is also expressed as a simple harmonic function:

$$V_{ANGLE}(\theta_i) = \sum_{ANGLES} K_{\theta,i}\left(\theta_i - \theta_i^{eq}\right)^2 \qquad (1.4)$$

where θ_i^{eq} is the equilibrium angle of the ith angle and $K_{\theta,i}$ is the force constant for the ith angle.

Rotation about dihedral angles is the only bonded term that is not modeled by a harmonic function:

$$V_{DIHEDRAL}(\varphi_i) = \sum_{DIHEDRALS} K_{\varphi,i}\left(1 + cos(n\phi_i - \delta)\right) \qquad (1.5)$$

where $K_{\varphi,i}$ is force constant, n the multiplicity of that function, and δ is the phase shift. A sinusoidal function is appropriate here because the dihedral energy can have several local energy minima.

Non-bonded interactions

The energetic contribution for non-bonded interactions (Figure 1.5 A) can be subdivided into an electrostatic contribution and a van der Waals (vdW) contribution, Equation (1.6):

$$V_{NON-BONDED}(r_{ij}) = V_{vdW}(r_{ij}) + V_{ELEC}(r_{ij})$$

(1.6)

where r_{ij} is the distance between non-bonded atoms i and j.

The electrostatic contribution is the familiar Coulombic interaction term:

$$V_{ELEC}(r_{ij}) = \sum_{i,j} \frac{Q_i Q_j}{D r_{ij}}$$

(1.7)

where Q_i and Q_j are the charges on atoms i and j, respectively, and D is the dielectric constant of the surrounding medium (typically set to 1 when explicit solvent is used in the simulations). At the beginning of the simulations, each atom in the protein of interest is assigned a partial atomic charge based on detailed quantum mechanical calculations on model systems, which is used for all subsequent electrostatic calculations.

The vdW term models the interaction that arises from dispersion forces associated with all atomic interactions. The equation follows the so-called 6-12 potential, describing the Lenard-Jones potential energy:

$$V_{vdW}(r_{ij}) = \sum_{i,j} \varepsilon_{ij} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{A_{ij}}{r_{ij}} \right)^{6} \right]$$

(1.8)

where ε_{ij} determines the well depth and A_{ij} determines the "vdW radius"; i.e., the distance where the vdW energy starts to precipitously increase. In the absence of any other outside forces, two atoms generally do not get closer than their vdW radii.

1.4.1.3 Integrating the Equations of Motion

Once the potential energy of the system is specified, Newton's equations of motion are obtained by applying the Euler-Lagrange equation. The result is that the force on any atom is a function of the derivative of the potential energy:

$$-\frac{dV}{dq_i} = m_i\ddot{q}_i = F_i \tag{1.9}$$

where q_i denotes the Cartesian coordinates of the system, $\ddot{q}_i$ denotes the 3N accelerations, and F_i is the force associated with the ith degree of freedom. At any given time, the force can be calculated from the derivative of the potential. Once the force is known, the velocity, $\dot{q}_i$, of each atom can be obtained after some small time step:

$$\dot{q}_i(t + \Delta t) = \dot{q}_i(t) + \ddot{q}_i(t)\Delta t + O\Delta t^2 \tag{1.10}$$

Similarly, once the velocity is known, the positions can be calculated as follows:

$$q_i(t + \Delta t) = q_i(t) + \dot{q}_i(t)\Delta t + \frac{1}{2}\ddot{q}_i(t)\Delta t^2 + O\Delta t^3 \tag{1.11}$$

In order to use Equations (1.10) and (1.11), one must have an initial set of Cartesian coordinates for all atoms in the system and an initial set of atomic velocities. The initial Cartesian coordinates can be obtained from the x-ray crystal structure and the initial velocities are assigned to the protein to ensure that the system is at some pre-specified temperature.

1.4.2 Potential of mean force calculations

All molecular dynamics simulations are performed with potential energy function within the program CHARMM (Chemistry at HARvard Molecular Mechanics) (Brooks III, Bruccoleri et al. 1983). All peptides are constructed using the polar hydrogen set, as previously described. Molecular dynamics for IP1 are performed in a previous work (Stultz 2002) and in this work the data is just reanalyzed. Molecular dynamics simulations for IP2 are carried out using the same reported methodology (Stultz 2002), with a few modifications. As a crystallographically determined structure of IP2 is not available, we model the mutant structure by substituting key amino acids in the T3-785 structure with residues found in IP2. This approach is appropriate because crystallographic structures of collagen-like peptides with distinct sequences have been shown to adopt triple-helical structures. Subtle variations arise in the side-chain packing and the helical twist - variations that, in principle, can be captured using the molecular modeling techniques mentioned below. For both IP1 and IP2, the system is solvated using a sphere of 20Å of diameter of explicit water molecules. Simulations are performed at radius of gyration (γ) values ranging from 17.7 to 19.5Å, with increments of 0.1Å, starting at the value for the initial structure (18.5Å), and proceeding in the forward and reverse directions. The radius of gyration γ is calculated for the backbone atoms of chain A of each peptide within the reaction region, and corresponds to:

$$\gamma = \sqrt{\frac{\sum_{i=1}^{N} m_i R_i^2}{\sum_{i=1}^{N} m_i}} \qquad (1.12)$$

where m_i is the mass of a given atom and R_i is the distance of such atom to the center of mass.

Each simulation is 50ps in length, with a time step of 1 femtosecond, where the first 20ps are considered equilibration and the last 30ps are the production time. During the production time, 3000 values of the radius of gyration are collected in order to create a histogram of values of γ at that simulation window. From the histogram, the biased probability for a given window can be calculated, by dividing the number of occurrences of a certain bin of the histogram over the 3000 values collected. This local biased probability distribution can be unbiased following the procedure explained in the next section.

Simulations are performed using a stochastic boundary approach as previously described (Stultz 2002). Stochastic boundary methods limit the computational requirements of the system by only simulating a relatively small region within the protein of interest (Figure 1.6) (Brooks III and Karplus 1983; Brooks III, Brunger et al. 1985). Regions that fall outside of the simulated area are retained in the computation but do not undergo full molecular dynamics. Instead they are restrained to adopt conformations that are near the crystallographic coordinates.
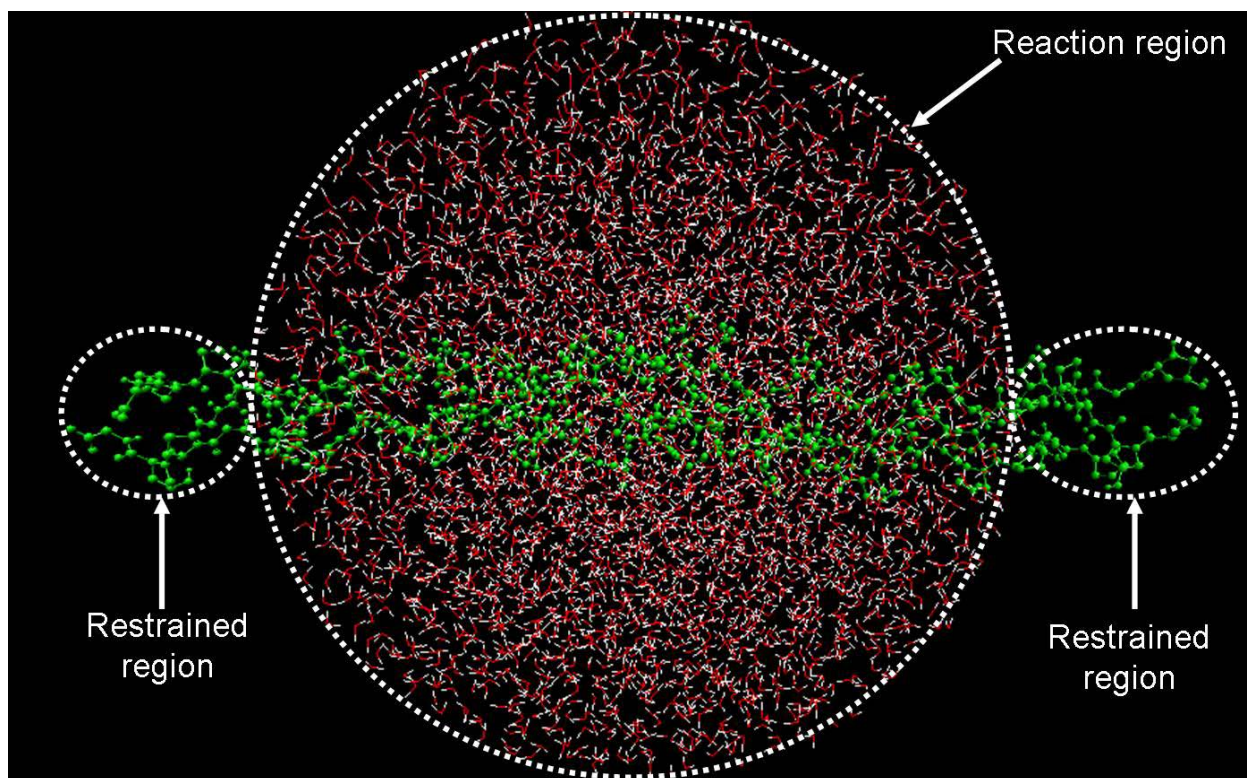
Figure 1.6. The stochastic boundary method (Brooks III and Karplus 1983; Brooks III, Brunger et al. 1985) applied to a collagen-like peptide. The reaction region undergoes full molecular dynamics. The portion of the protein that falls within this sphere is exposed to water molecules at a density near that found in physiologic systems. Waters are kept within the reaction region using a stochastic boundary potential. The two restrained regions do not undergo molecular dynamics and are not solvated. This figure has been created with the molecular visualization package VMD (Humphrey, Dalke et al. 1996).

1.4.3 Obtaining an unbiased potential of mean force from a biased simulation

We start by defining the probability distribution for a given window as a function of the reaction coordinate γ, in this case the radius of gyration of the chain A within the reaction region.

$$p_i(\gamma) = \frac{\int e^{-\beta U_i(\gamma,\xi)} d\xi}{z} \tag{1.13}$$

where $\beta = \frac{1}{k_B T}$ and k_B is the Boltzmann's constant, T is the temperature in degrees Kelvin, z is the partition function of the system and $U_i(\gamma, \xi)$ is the potential energy as a function of the radius of gyration and the Cartesian coordinates ξ.

The sampling of different values of radius of gyration is ensured by introducing a biasing energy term that forces the system to sample conformation around a given value of radius of gyration. Starting at the equilibrated value of radius of gyration, this is increased or decreased by increments of 0.1Å, using the following expression:

$$U_{bias,i}(\gamma) = A(\gamma - \gamma_i)^2 \tag{1.14}$$

where γ is the value of radius of gyration that the system takes at a given time of the simulation, γ_i is the reference value to which the a given window is biased and A is the force constant of the bias, in this case 500 kcal/Å^2. Therefore, the effective potential energy for that is used to simulate a window is given by the sum:

$$U_i(\gamma, \xi) + U_{bias,i}(\gamma) \tag{1.15}$$

From each window, we obtain the biased probability distribution as a function of γ.

$$p_i^*(\gamma) = \frac{\int e^{-\beta\left(U_i(\gamma,\xi)+U_{bias,i}(\gamma)\right)} d\xi}{z^*} \tag{1.16}$$

where z^* is the biased partition function. If we now multiply and divide by the partition function z (multiplying by 1) we obtain:

$$p_i^*(\gamma) = \frac{\int e^{-\beta\left(U_i(\gamma,\xi)+U_{bias,i}(\gamma)\right)}d\xi}{z^*} \cdot \frac{z}{z} \tag{1.17}$$

Rearranging terms we obtain:

$$
\begin{aligned}
p_i^*(\gamma) &= \frac{\int e^{-\beta\left(U_i(\gamma,\xi)+U_{bias,i}(\gamma)\right)}d\xi}{z^*} \cdot \frac{z}{z} \\[2mm]
&= \frac{\int e^{-\beta(U_i(\gamma,\xi))}d\xi \cdot e^{-\beta\left(U_{bias,i}(\gamma)\right)}}{z^*} \cdot \frac{z}{z} \\[2mm]
&= \frac{\int e^{-\beta(U_i(\gamma,\xi))}d\xi}{z} \cdot \frac{z}{z^*} \cdot e^{-\beta U_{bias,i}(\gamma)} \\[2mm]
&= p_i(\gamma) \cdot \frac{z}{z^*} \cdot e^{-\beta U_{bias,i}(\gamma)} = p_i^*(\gamma)
\end{aligned} \tag{1.18}
$$

Isolating the unbiased probability we obtain:

$$p_i(\gamma) = p_i^*(\gamma) \cdot \frac{z^*}{z} \cdot e^{\beta U_{bias,i}(\gamma)} \tag{1.19}$$

which relates the unbiased probability with the biased probability. The ratio of partition functions is just a number that can be adjusted to allow the data from different simulations to match, and can be expressed as $\frac{z^*}{z} = e^{-\beta F_i}$, where the F_i constants can be determined using the SPLICE method (Stultz 2002), or the weighted histogram analysis method in Chapter 2. If we take logarithms on both sides of the previous relation, we obtain the free energy W_i as a function of γ.

$$W_i(\gamma) = -k_B T ln\big(p_i(\gamma)\big)_i \tag{1.20}$$

$$
\begin{aligned}
W_i(\gamma) &= -k_B T ln\big(p_i^*(\gamma) \cdot e^{-\beta F_i} \cdot e^{\beta U_{bias,i}(\gamma)}\big) \\[2mm]
&= -k_B T ln\big(p_i^*(\gamma)\big) - U_{bias,i}(\gamma) + F_i
\end{aligned} \tag{1.21}
$$

In short, in the SPLICE method, the overlap between the unbiased probability distribution of two contiguous windows is found. Next, the slopes in the overlap region are compared, and

the two distributions are pieced together where the slopes match the best. In order to piece the two fragments of pmf together, one of the pieces is arbitrarily shifted in order to bring it close to the other piece. This is equivalent to adjusting the constant F_i. This method is repeated until all the windows have been matched together.

1.4.4 Flexibility calculations

The conformational flexibility of sequence in solution, is related to the configurational entropy of the peptide as follows (McQuarrie 2000):

$$S = -k_B T \int_{\gamma_{min}}^{\gamma_{max}} p(\gamma) log\big(p(\gamma)\big) d\gamma \qquad (1.22)$$

where k_B is the Boltzmann constant, T the temperature and p the probability density described in Equation (1.23).

$$p(\gamma) = \frac{e^{-\beta W(\gamma)}}{\int_{\gamma_{min}}^{\gamma_{max}} e^{-\beta W(\gamma')} d\gamma'} \qquad (1.23)$$

where β is $1/k_B T$, $W(\gamma)$ is the pmf for a given sequence and γ_{min} and γ_{max} denote values of the reaction coordinate where the overall probability falls below 10^{-6}.

1.4.5 Peptide preparation

Two collagen-like peptides are studied in this work: A) T3-785, which models the imino-poor segment next to the actual cleavage site in type III collagen, which we refer to as peptide IP1, and B) an imino-poor segment next to a potential cleavage site in type III collagen, with similar sequence to IP1. We refer to the latter as peptide IP2. The sequences for the two peptides are shown in Figure 1.1. Both peptides contain multiple POG repeats to ensure that they will fold into a triple helix.

All peptides are obtained to >95% purity from Genemed Synthesis. Peptides are received as a lyophilized powder and are then resuspended in phosphate buffered saline (PBS) to a final concentration of 10mg/ml and allowed to sit for at least 24 hours at 4ºC prior to any experiments. Samples are then stored at -80ºC until experiments could be performed.

1.4.6 Circular dichroism

Circular dichroism (CD) spectroscopy is based on the differential absorption of circularly polarized light by the peptide bonds in a protein. CD experiments have been successfully used to gain insights into the overall conformation of proteins (Kelly, Jess et al. 2005). As the CD spectrum of fully folded collagen has a unique shape, CD spectroscopy of collagen and collagen-like peptides can be used to screen proteins and peptides for the presence of triple-helical structure (Bhatnagar and Cough 1996). All CD spectroscopic measurements are obtained with a Jasco J-810 spectropolarimeter. Compressed nitrogen is used to minimize the amount of air in the sample chamber during the experiments, as air interferes with spectral readings at low wavelengths (Venyaminov and Yang 1996).

1.4.7 Peptide melting point and melting curve determination

Peptide samples are thawed by storing them at 4ºC for 24h. Melting curves and melting temperatures are obtained by heating the peptide of interest at distinct heating rates (3, 4, 10, and 15ºC/h). Heating begins with the sample at a temperature of 4ºC and ends at a temperature of 50ºC, with two repetitions in each case. The ellipticity value at the wavelength of 225nm is chosen to monitor the structural change in the peptide during the melting experiment. This value corresponds to the maximum in CD spectra of a triple helical structure (in the range 190-240nm (Bhatnagar and Cough 1996)). Measurements are made using a peptide concentration of 1mg/ml.

The raw output measures the ellipticity in mili-degrees at a given wavelength, which is then converted to the standard circular dichroism units of molar ellipticity per residue - a function of sample concentration, peptide molecular weight, cell cuvette length, peptide residue length, and the ellipticity in milli-degrees (Venyaminov and Yang 1996). To obtain the melting point as a function of the heating rate, we fit each melting curve to the sigmoidal function:

$$f(T) = A \left(\frac{1}{1 + e^{B(T-T_m)}} \right) \tag{1.24}$$

where T_m is the melting temperature and A, B and C are constants.

We would ideally like to obtain the melting temperature of the protein at a heating rate near 0°C/h; i.e., when the protein is heated infinitely slowly. Consequently, the different melting curves, obtained at distinct heating rates, are used to estimate the melting curve at a heating rate of 0°C/h. Data at each temperature interval (separated by 0.1°C increments) for the different melting curves (recorded at different heating rates) is extrapolated as well to a heating rate of 0°C/h.

Chapter 2:
Characterizing the Conformational Ensemble of Collagen Cleavage Sites

2.1 Introduction

In collagen degradation, in order for peptide bond hydrolysis to occur at one unique site, MMPs must: 1) localize to a region in the vicinity of the unique scissile bond; 2) bind residues that comprise the scissile bond at the catalytic site; and 3) hydrolyze the corresponding peptide bond. Recent studies suggest that for some collagen types, localization to the unique cleavage site is mediated by the binding of non-catalytic MMP domains (e.g., the hemopexin-like domain) to specific amino acid sequences near the actual cleavage site (Perumal, Antipova et al. 2008; Erat, Slatter et al. 2009). For example, in type I collagen, a short stretch of 4 residues (RGER) downstream from the actual cleavage site has been implicated as a true binding site for such non-catalytic domains (Perumal, Antipova et al. 2008; Erat, Slatter et al. 2009).

Once the enzyme is localized to a region near the scissile bond, the catalytic domain must then bind the residues that contain the peptide bond to be hydrolyzed. Even if the localization step was the sole responsible for cleavage specificity, there is still the problem that the triple helix does not fit in to the active site of the collagenase. Here, we focus on the second step in the degradation process. In particular, we wish to determine whether unfolded states exist at the collagenase cleavage site and if binding of the catalytic domain to these states plays any role in determining MMP cleavage site specificity. The precise question we address is: if one could localize an MMP near each potential cleavage site, could the catalytic domain then bind the corresponding scissile bond without the need of active unwinding of the collagen triple helix?

In this thesis we hypothesize that naturally occurring locally unfolded allow collagen degradation to occur. The nature of such locally unfolded states may be a function of the sequence surrounding each potential cleavage site, making some of them suitable for binding and degradation, while others not. In Chapter 1, we demonstrate that the sequence C-terminal to two

cleavage sites has an influence on the structural characteristics of the locally unfolded states occurring next to such cleavage sites. While these results are interesting, they present several limitations. Mainly, the studied sequences do not include the actual cleavage site. This is due to the fact that IP1 (T3-785) was designed without the cleavage site (Kramer, Bella et al. 2001). Second, they ignore the effect of the sequence N-terminal to the cleavage site. An third, only two sequences are compared. In favor of those sequences though, they can easily be studied experimentally since they are embedded between POG repeats that allow them to fold in solution. Adding the cleavage site itself and the sequence N-terminal to it may complicate possible experiments, i.e., the longer the peptide is the harder it will be to fold and more POG repeats will be needed at both ends of the peptide. However, from a computational point of view, adding more amino acids to the system only results in longer simulation times, which can be assumed by using more computer time. For this reason, a more complete computational study of the influence of sequence on the nature of locally unfolded states is presented here in Chapter 2. In particular, several potential cleavage sites are studied in peptides that include the cleavage sites themselves surrounded by their upstream and downstream amino acid sequences.

To assess the ability of different potential cleavage sites to adopt locally unfolded states that are amenable to binding by the MMP catalytic domain, we compute the conformational free energy surface of amino acid sequences near each potential cleavage site in type III collagen using detailed molecular simulations. The goal in Chapter 2 is to characterize the conformational states that the different potential cleavage sites sample, as a function of their surrounding sequence. The goal in Chapter 3 is to determine whether the binding of the catalytic domain of collagenases to these unfolded states is sufficient to explain cleavage site specificity.

2.2   Results

2.2.1 Free energy profiles of potential cleavage sites in type III collagen

We begin by mapping the free energy landscape of regions that contain potential cleavage sites in type III collagen. As type III collagen only contains a relatively small number (5) of potential cleavage sites (type I has 8 and type II has 6) this helps to ensure that we are able to fully map the free energy landscape of all potential cleavage sites in a reasonable amount of CPU time (Seyer and Kang 1981; Cheah, Stoker et al. 1985; Kuivaniemi, Tromp et al. 1988; Tromp, Kuivaniemi et al. 1988; Su, Lee et al. 1989; Peri, Navarro et al. 2003; Mishra, Suresh et al. 2006).

Each of the type III collagen potential cleavage sites, together with their N- and C-terminal flanking triplets, is shown in Figure 2.1. One sequence is centered about the true collagenase cleavage site (S1) while the other sequences (S2-S5) are centered about potential cleavage sites that are not degraded. We note that since the true collagenase cleavage site is near another potential cleavage site, peptides S1 and S2 contain both the true cleavage site in addition to a non-cleaved potential site (Figure 2.1).

```
S1(757-807):GDA-GQO-GEK-GSO-GAQ-GPO-GAO-GPL-G~IA-GIT-GAR-GLA-GPO-GMO-GPR-GSO-GPQ
S2(766-816):GSO-GAQ-GPO-GAO-GPL-GIA-GIT-GAR-G*LA-GPO-GMO-GPR-GSP-GPQ-GVK-GES-GKO
S3(496-546):GDA-GAO-GAO-GGK-GDA-GAO-GER-GPO-G*LA-GAO-GLR-GGA-GPO-GPE-GGK-GAA-GPO
S4(598-648):GPI-GPO-GPA-GQO-GDK-GEG-GAO-GLO-G*IA-GPR-GSO-GER-GET-GPO-GPA-GFO-GAO
S5(811-861):GES-GKO-GAN-GLS-GER-GPO-GPQ-GLO-G*LA-GTA-GEO-GRD-GNP-GSD-GLO-GRD-GSO
```

Figure 2.1. Sequences corresponding to the 5 potential cleavage sites in type III collagen. The true collagenase cleavage site (defined by the triplet GIA) is colored in green, while potential cleavage sites (three GLA triplets and one GIA triplet), that are not cleaved, are colored in red. The actual scissile bond is depicted with a ~ sign, while the scissile bonds corresponding to triplets that are not cleaved are depicted with a * sign. The numbering in parenthesis corresponds to the position of the first and last amino acids in the sequence of type III collagen.

We compute free energy profiles (also known as potentials of mean force or pmf) for unfolding homotrimeric triple-helical structures of peptides S1-S5 at 300K with explicit solvent using the radius of gyration as a reaction coordinate (Figure 2.2 A-E). Representative structures from local energy minima in each pmf are also shown. The pmf for peptide S1, the true cleavage site, contains two minima that have similar energies and representative structures from each minimum are partially unfolded in the region that contains the scissile bond (Figure 2.2 A). The pmf for peptide S2 is similar to the S1 pmf in that it also has two minima with comparable energies (Figure 2.2 B). This result is not surprising since peptides S1 and S2 share 50% of their sequence, in particular an imino poor stretch of sequence GIA-GIT-GAR-GLA that is expected to form a more labile and mobile triple helix due to the lack of imino acids. This increased mobility may allow collagen to sample more states, as reflected in the two energy minima of both pmf plots. The rest of potential cleavage sites S3-S5 have pmf plots with only one energy minimum (Figure 2.2 C-E). Interestingly, while there is some degree of unfolding in all the representative structures from each minimum, the scissile bond itself seems to be relatively hidden from solvent in sequences S2-S5 (Figure 2.2 B-E). Moreover, these results differ from the pmfs in Chapter 1 in that no well folded triple helices are present in the pmfs of sequences S1-S5. This may be due to the presence of POG repeats in the peptides from Chapter 1, while the peptides studied here only contain their wild type sequences, which contain a smaller number of imino acids.
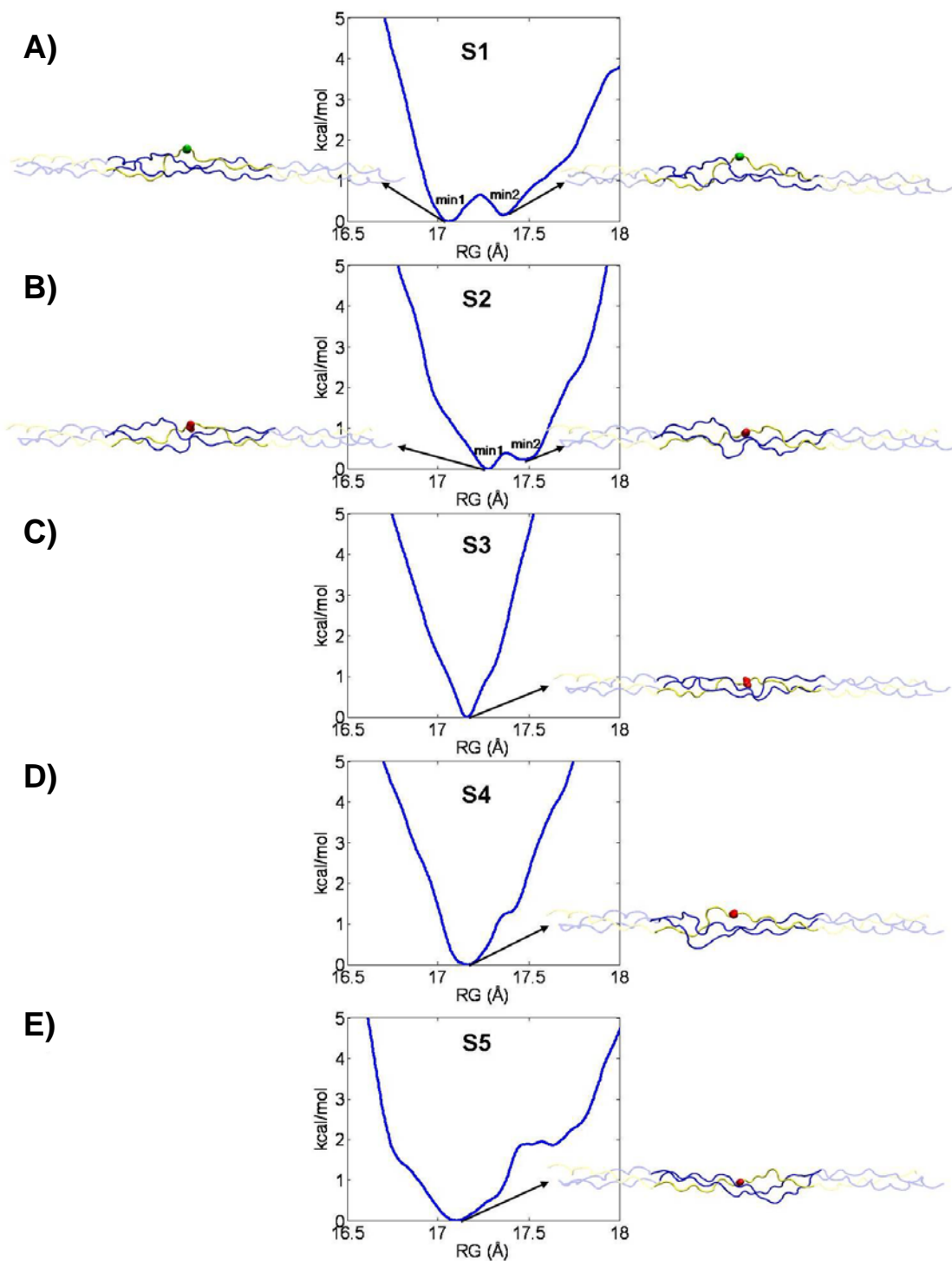
Figure 2.2. Potential of mean force (pmf) plots for the five potential cleavage sites in type III collagen (sequences listed in Figure 2.1). Representative structures from each energy minima are shown. The four atoms comprising the scissile bond (H, N, C and O) are shown in green for the true cleavage site and in red for the four potential cleavage sites.

68

While these data suggest that each of the potential cleavage sites is locally unfolded in the vicinity of the cleavage site, an analysis of representative structures from the low energy states of each pmf suggest that the different sequences correspond to distinct unfolded ensembles. As the goal of Chapter 2 is to compare and contrast the different structural ensembles for each of the potential cleavage sites, it is desirable to develop simple metrics that facilitate the comparison of unfolded ensembles at each potential cleavage site. In this regard, for each sequence we compute: 1) a *Vulnerability Score* that quantifies the average scissile bond exposure and 2) a *Flexibility Score* that quantifies the conformational entropy of each sequence. Large *Vulnerability Scores* suggest that the scissile bond is, on average, solvent exposed and consequently relatively accessible to collagenases. Similarly, large *Flexibility Scores* suggest that the chain can adopt a relatively large number of conformational states at 300K. It is important to note that flexibility and vulnerability are, in principle, distinct concepts. If a sequence adopts many different conformations during its biological lifetime, it is by definition flexible, however if none of these conformations have solvent exposed scissile bonds, then the sequence will have a low *Vulnerability Score*. Conversely, a sequence that adopts only one state in solution may be very vulnerable if this state has an exposed scissile bond. Such a sequence, however, is clearly not flexible. Together, the *Vulnerability Score* and the *Flexibility Score* constitute a rigorous set of metrics that we use to compare different free energy profiles.

Computed *Vulnerability Scores* for sequences S1-S5 confirm that the scissile bond site in the sequence corresponding to the true collagenase cleavage site has the highest vulnerability (Figure 2.3). By contrast, sequences S2-S5 have lower *Vulnerability Scores*. Taking into account that sequences S2-S5 are not cleaved in triple helical collagen, we suggest that vulnerability, i.e. the existence of vulnerable locally unfolded states, is a metric that describes whether a cleavage

site can be cleaved or not. Calculated *Flexibility Scores* are listed in Table 2.1. The listed values are relative to the *Flexibility Score* of sequence S1. Sequences S2-S5 all have negative *Flexibility Scores* and are therefore less flexible than sequence S1, which contains the true collagenase cleavage site. To put these values into context, we note that a previous estimate of the average entropy gain associated with complete unfolding of one amino acid to be 1500 to 1800 cal/mol/residue at 300K (D'Aquino, Gomez et al. 1996; Privalov and Dragan 2007). Therefore the conformational entropy changes listed in Table 2.1 are smaller than the entropy of complete unfolding of a single residue; i.e., differences in flexibility between sequences S1-S5 are all quite small. This is consistent with the representative structures from the different pmfs in Figure 2.2 in that the structural differences between sequences are localized to relatively small regions.
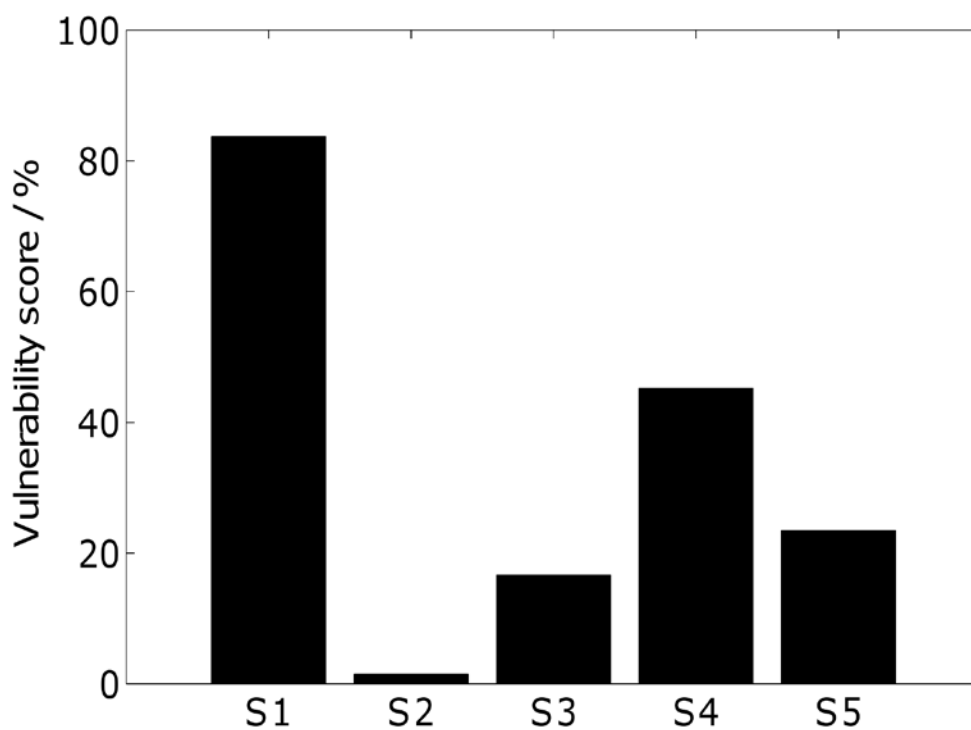
Figure 2.3. *Vulnerability Score* plots for the 5 cleavage sites in type III collagen. Vulnerability values are expressed as a percentage, with 100% being the vulnerability of the scissile bond of a G-I or G-L dipeptide free in solution.

Sequence (S)	$[F(S) - F(S1)]$ cal/mol at 300K
S1	0
S2	-114
S3	-573
S4	-363
S5	-78
S1M1	-153
S1M1e	-222
I785P	-132

Table 2.1. Entropy values for sequences S1-S5 and mutants S1M1, S1M1e and I785P

2.2.2 The effect of imino acids C-terminal to the S1 cleavage site on sequence vulnerability and flexibility

It has been previously hypothesized that one of the requirements for a collagenase cleavage site to be cleaved is that its C-terminal region must be mostly imino acid free (Fields 1991). This hypothesis was drawn from the comparison of the C-terminal sequences of true cleavage sites versus non degraded cleavage sites, amongst several collagen types and species. While these results yielded an interesting insight into the sequence requirements for a cleavage site to be amenable to collagenolysis, they didn't explain the mechanism by which the presence of imino acids C-terminal to a cleavage site would abrogate collagenolysis. The results from Chapter 1 show that the presence of a hydroxyproline residue in the imino poor sequence C-terminal to a potential cleavage site, corresponding to sequence S3, limits local unfolding. Since the cleavage site is not included in that sequence, it is not clear what the effect would the presence of an imino acid have.

Interestingly, it has been reported that a type III collagen mutant containing the mutation I785P is resistant to collagenolysis when compared to wild type collagen (Williams and Olsen 2009). This mutation substitutes an isoleucine residue in the adjacent triplet C-terminal to the true collagenase cleavage site in sequence S1 for a proline (Figure 2.4 A). When introducing a proline residue, type III collagen becomes resistant to cleavage by MMP1 at a 10:1 collagen:enzyme ratio (Figure 2.4 B). At lower values of collagen:enzyme ratio, like 5:1, some degradation of the I785P mutant occurs, while all wild type collagen is degraded (data not shown).

In order to investigate the cause for the marked resistance to cleavage of the I785P mutant, we also perform a potential of mean force simulation of the mutant and compute its

Vulnerability Score. The introduction of the proline residue results in a pmf with only one energy minimum, as opposed to the two energy minima for sequence S1 (Figure 2.4 C), suggesting that the I785P can sample less states than S1. In addition, the *Vulnerability Score* for the I785P mutant is much lower, about 5%, compared to the S1 score, which is above 80% (Figure 2.4 D). The mutant I785P is also less flexible than sequence S1 (Table 1). Therefore it is possible to explain, under the hypothesis of locally unfolded, vulnerable states, why the I785P mutant is cleaved with less efficiency by MMP1. The results suggest that the scissile bond is hindered from solvent in the I785P mutant, hence the low *Vulnerability Score*, which does not allow degradation to occur. Together with the previous result in which sequence S1 is the most vulnerable of all the wild type sequences, our results on the I785P mutant allow us to draw a connection between vulnerability and the possibility of a cleavage site to be degraded. Under the hypothesis that locally unfolded states are necessary for collagen degradation to occur, if a cleavage site is not vulnerable enough, binding of the catalytic domain will not occur and degradation will not proceed. This ultimately implies that vulnerability can be used as a marker of degradability at a cleavage site.

A)

```
S1(757-807): GDA-GQO-GEK-GSO-GAQ-GPO-GAO-GPL-GIA-GIT-GAR-GLA-GPO-GMO-GPR-GSO-GPQ
I785P      : GDA-GQO-GEK-GSO-GAQ-GPO-GAO-GPL-GIA-GPT-GAR-GLA-GPO-GMO-GPR-GSO-GPQ
```
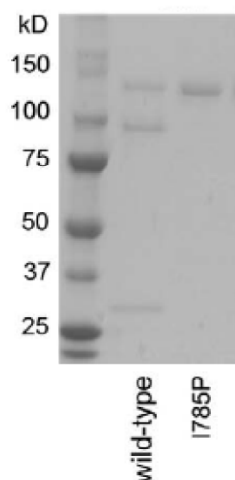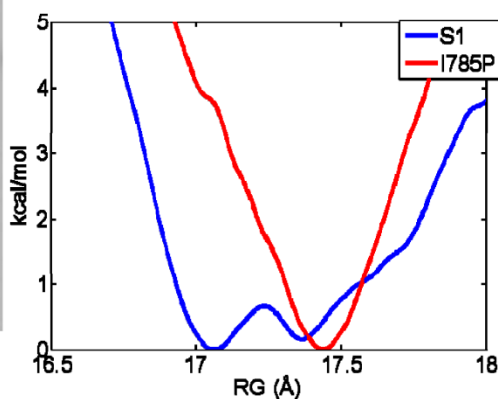
B)   **C)**   **D)**
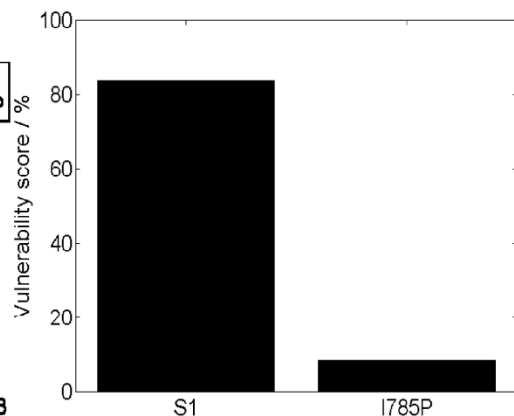
Figure 2.4. Effect of introducing the I785P mutation in type III collagen. A) Sequence corresponding to the S1 cleavage site and its I785P mutant in type III collagen. The true collagenase cleavage site (defined by the triplet GIA) is colored in green. The introduced proline residue is shown in magenta. B) Degradation of type III collagen, wild type, versus the degradation of the I785P mutant by MMP1, 10:1 collagen:enzyme ratio (Williams and Olsen 2009). In the wild type lane three bands can be seen. The intact α1(III) chain, top band, and the characteristic ¾ and ¼ fragments of the α1(III) chain, middle and bottom bands. In the I785P lane, only the intact α1(III) chain can be seen. C) Potential of mean force (pmf) plots for the S1 cleavage site (blue) and its I785P mutant (red). D) *Vulnerability Score* plots for the S1 cleavage site and its I785P mutant. Vulnerability values are expressed as a percentage, with 100% being the vulnerability of the scissile bond of a G-I dipeptide free in solution.

2.2.3 The effect of arginine on sequence vulnerability and flexibility

An early analysis of collagen sequences from different species identified sequence properties that are associated with the true collagenase cleavage site. Two important properties along these lines include a relatively paucity of imino acids downstream from the cleavage site and an arginine residue downstream from the cleavage site (usually in the P_5' or P_8' position) (Fields 1991). A number of previous studies have demonstrated that the presence of imino acids has an influence on the conformational stability of collagen (Li, Fan et al. 1993; Fiori, Sacca et al. 2002; Stultz 2002; Nerenberg and Stultz 2008) and in the previous section it has been shown to have an effect on the capability of a cleavage site to be degraded. On the other hand, while there are data to suggest that arginine residues also confer added stability to the triple-helical structure (Persikov, Ramshaw et al. 2000; Persikov, Ramshaw et al. 2000), the structural basis underlying any stabilizing effects of arginine has not been fully explored.

In crystallographic structures of collagen-like peptides, the downstream arginine side chain forms a hydrogen bond with a backbone carbonyl from a neighboring chain (Kramer, Bella et al. 2001). We hypothesize that this interaction may help to stabilize the downstream triple-helical structure. To test this, we design a mutant of sequence S1, called S1M1, which contains an arginine residue immediately N-terminal to the scissile bond containing triplet (Figure 2.5 A). The resulting pmf contains only one energy minimum (Figure 2.5 B). More importantly, the calculated *Vulnerability Score* of the resulting sequence is significantly less than the wild-type sequence (Figure 2.5 C), suggesting that arginine residues prevent significant unfolding at the cleavage site. Based on the previously proposed predictive power of the *Vulnerability Score*, we hypothesize that mutant S1M1 will be resistant to cleavage. Moreover, the flexibility of the mutant sequence is also significantly reduced (Table 2.1).

To test whether the formation of a hydrogen bond between this arginine side chain and other moieties on adjacent collagen chains is responsible for this change in vulnerability and flexibility, we rerun the unfolding simulations for S1M1 with the electrostatic interactions between the introduced arginine and the rest of the system turned off (sequence S1M1e, Figure 2.5 A). The goal is to understand whether electrostatic interactions between the arginine side chain and the rest of the protein play an important role in decreasing the vulnerability and flexibility of the sequence. While the resulting pmf still contains only one energy minimum (Figure 2.5 B), the *Vulnerability Score* is increased (almost to the value seen in the wild-type sequence) when these electrostatic interactions are excluded (Figure 2.5 C). The *Flexibility Score*, however, remains significantly below the corresponding wild-type value (Table 2.1).
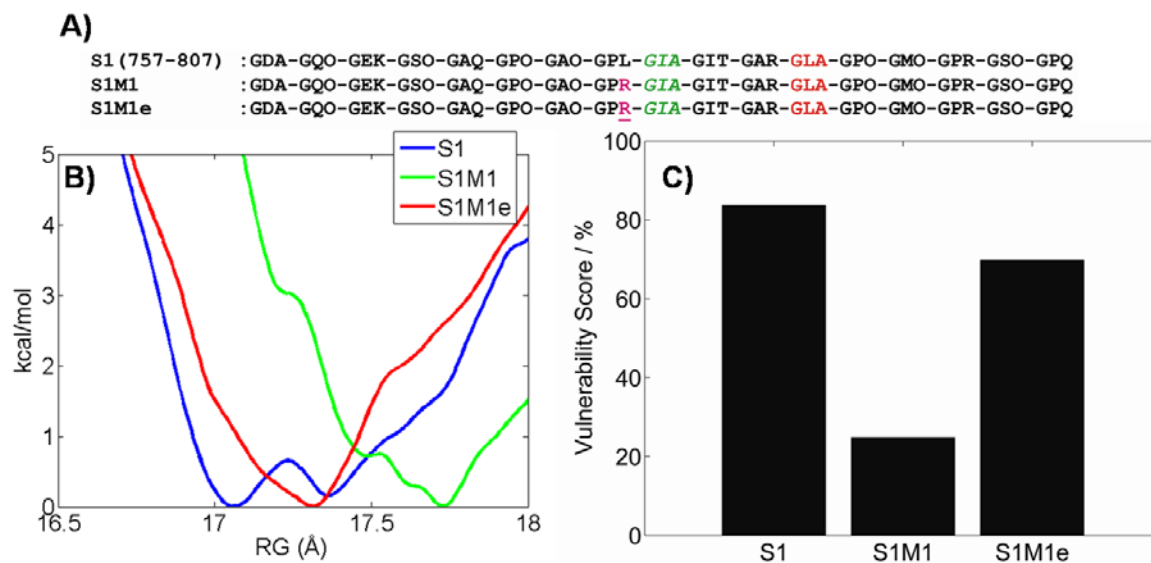
Figure 2.5. Effect of introducing an arginine residue N-terminal to the actual cleavage site in S1. A) Sequence alignment of the S1 and the S1M1/S1M1e corresponding mutants. The actual type III collagen cleavage site is shown in green. The nearby potential cleavage site is shown in red. Mutant residue positions are shown in magenta. S1M1 replaces a leucine with an arginine. S1M1e has the same mutation, but the electrostatic interactions between the side chain of the mutant arginine and other collagen chains are turned off. B) Comparison of the potential of mean force plots for S1 and its two mutant sequences. C) Vulnerability Score plot for S1 and its two mutant sequences.

2.3   Discussion

Although there are several potential cleavage sites in the sequence of type III collagen, only one is cleaved at temperatures below collagen's melting temperature. Prior attempts to understand this phenomenon were based on a comparison of sequences across different species and collagen types. While these studies yielded a number of important sequence rules that are correlated with the true collagenase cleavage site, they do not immediately make clear how these sequence variations translate into the experimentally observed specificity of MMPs (Fields 1991). The goals of Chapter 2 are to determine whether locally unfolded states exist in collagen, to characterize such unfolded states, and to establish whether the nature of these unfolded states contributes to MMP cleavage site specificity.

We begin with an analysis of the conformational free energy landscape of collagen-like sequences corresponding to the different potential cleavage sites in type III collagen. Low energy states from all five sequences exhibit some degree of partial unfolding. This is to be contrasted with prior simulations of imino rich homotrimeric triple-helical peptides that suggest that such sequences retain their folded triple-helical structure at 300K (Stultz 2002). Each of the sequences under consideration contains a stretch of 2-4 triplets that lack any imino acids. Consequently, the finding that these regions adopt structures that differ from the classic triple-helical fold is not surprising in light of the fact that each of the 19 imino poor naturally occurring amino acids contribute differently to triple-helical stability, where many amino acids can lead to destabilization of the helical structure when placed in positions X1 or X2 of G-X1-X2 triplets (Persikov, Ramshaw et al. 2000). Moreover, the fact that relatively imino poor regions of collagen adopt structures that differ from the prototypical collagen-like fold is consistent with other data that imply that there is considerably heterogeneity in triple helix structure and stability

77

along the collagen chain and that this heterogeneity may be important for the interaction of collagen with other proteins (Kramer, Bella et al. 1999; Persikov, Ramshaw et al. 2000; Kramer, Bella et al. 2001; Sacca, Fiori et al. 2003; Orgel, Irving et al. 2006; Makareeva, Mertz et al. 2008; Perumal, Antipova et al. 2008; Erat, Slatter et al. 2009). Our results show that, while unfolded conformations are present in low energy states of each potential cleavage site, stable states of the true cleavage site have the scissile bond in conformations that are the most vulnerable and the most flexible. Overall, our data suggest, at least for type III collagen, that differences in unfolded ensembles of potential cleavage sites may contribute to MMP specificity. In particular, we hypothesize that vulnerability is a marker of degradability at a cleavage site. Also, we propose an explanation based on the lack of vulnerable states to account for the experimental observation in which a type III mutant incorporating a I785P mutation next to the collagenase cleavage site is resistant to cleavage by MMP1 (Williams and Olsen 2009). Imino acids like Proline in the X1 position and Hydroxyproline in the X2 position of G-X1-X2 triplets have been shown to stabilize the triple helix the most (Persikov, Ramshaw et al. 2000). Since the proline residue in I785P is placed in the X1 position two residues C-terminal to the scissile bond, it is possible that it stabilizes the triple helix locally, impeding the local unfolding required for vulnerable states to occur.

A limitation of our results is that we focus on a small region of collagen in isolation, as opposed to in the fibrillar state. In a recent study, a relatively low resolution fiber diffraction structure of a type I collagen fibril was obtained at room temperature (Orgel, Wess et al. 2000; Orgel, Miller et al. 2001; Orgel, Irving et al. 2006; Perumal, Antipova et al. 2008). Full length porcine MMP1 was then manually docked at the cleavage site of type I collagen (Perumal, Antipova et al. 2008). However, chain exposure/dissociation in the fibrillar structure – which is

78

an average x-ray structure – was insufficient to allow the scissile bond to access the active site of MMP1 (Perumal, Antipova et al. 2008). This suggests that the region including the true scissile bond does not, on average, adopt states that are compatible with the catalytic site. In the same study it was also demonstrated that a number of potential cleavage sites in type I collagen, which are not cleaved, have similar degrees of chain dissociation from the helix center relative to the true cleavage site.

Unlike type III collagen, type I collagen (the focus of the fibrillar study) contains a hydroxyproline residue in the triplet immediately downstream from the scissile bond, which is known to stabilize the triple-helical structure (Persikov, Ramshaw et al. 2000; Vitagliano, Berisio et al. 2001; Mizuno, Hayashi et al. 2004; Nerenberg and Stultz 2008). Hence our observations regarding the accessible states of potential cleavage sites in type III collagen may not be directly applicable to other types of collagen. Moreover, the applicability of our results to collagen in the fibrillar state remains to be determined. Nevertheless, it is important to note that in order to determine whether a potential cleavage site is complementary to a given enzyme active site, one must consider all of the thermally accessible states of the system. This is particularly true for regions of the protein that are known to be partially unfolded, as such regions typically have a wide range of accessible states. For such sequences, an average structure may not provide an accurate representation of whether the protein can adopt complementary states for any given target (Lange, Lakomek et al. 2008). While our data suggest that all potential cleavage sites in type III collagen are similar in that they sample unfolded states, the exact nature of the corresponding unfolded ensembles significantly differs in that the true cleavage site samples conformations that are more vulnerable and flexible. These findings are in accord with other studies on intrinsically disordered proteins that suggest that unfolded states of proteins

have their own taxonomy, which is determined by their precise amino acid sequences (Tran, Wang et al. 2005; Dunker, Silman et al. 2008; Huang and Stultz 2008; Tran, Mao et al. 2008; Huang and Stultz 2009). In other words, not all unfolded states are created equal.

Lastly we note that the role of arginine residues in collagen degradation has recently been a subject of some interest (Perumal, Antipova et al. 2008; Erat, Slatter et al. 2009). Both $\alpha1$ and $\alpha2$ chains in type I collagen contain a RGER motif 14 residues downstream to the cleavage site. NMR and X-Ray crystallography studies have shown that this motif interacts with the 8 and 9 repeats of fibronectin I (Erat, Slatter et al. 2009). Moreover, docking studies involving MMP1 and fibrillar type I collagen suggest that this motif interacts with the MMP hemopexin-like domain (Perumal, Antipova et al. 2008). An analysis of the docked MMP1 structure points to the interaction between the RGER motif and a pocket between modules 3 and 4 of the hemopexin-like domain and it has been suggested that this interaction may lead to further destabilization of the triple helix (Perumal, Antipova et al. 2008). Our data suggest that electrostatic interactions between the arginine side chain and backbone moieties on adjacent collagen chains stabilize the triple-helical structure. Consequently, interactions between non-catalytic MMP domains and arginine residues that disrupt this interaction may lead to further destabilization of the triple-helical structure, thereby allowing nearby residues to sample non triple-helical states that are more compatible with the catalytic site (Perumal, Antipova et al. 2008). In addition, our data suggest that the placement of arginine residues themselves may influence stability at the cleavage site. The introduction of an arginine residue immediately upstream from the scissile bond containing triplet leads to a significant reduction in sequence vulnerability. Our calculations suggest that this is due to direct electrostatic interactions between the arginine side chain and backbone moieties on adjacent chains. The fact that the flexibility of the chain remains

unchanged when these electrostatic interactions are turned off argues that the intrinsic conformational preference of the arginine residue restricts collagen's flexibility.

In summary, in Chapter 2 we use molecular dynamics simulations to determine that locally unfolded states are present at all of the potential cleavage sites in type III collagen. These unfolded states are not created equal and differ in terms of vulnerability and flexibility, being the actual cleavage site the most vulnerable and flexible. We hypothesize that this combination is partly responsible for cleavage specificity at the actual cleavage site. Since it has been hypothesized that cleavage site specificity is mediated by binding sites in the hemopexin-like domain, Chapter 3 aims to test whether cleavage site specificity can be achieved by the selective interaction of the catalytic domain of a collagenase alone and the locally unfolded states described here in Chapter 2.

2.4   Methods

2.4.1 Model construction

The sequences S1 to S5 (Figure 2.1) are obtained from the Human Protein Reference Database (Peri, Navarro et al. 2003; Mishra, Suresh et al. 2006). As there are no existing structures for most regions of the type III collagen sequence, the triple helical structure of each sequence is generated using the THeBuScr (Triple Helical collagen Building Script) tool (Rainey and Goh 2004). This approach is appropriate because crystallographic structures of collagen-like peptides with distinct sequences have been shown to adopt similar triple-helical structures (Brodsky and Persikov 2005). Subtle variations arise in the side-chain orientation and the helical twist - variations that, in principle, can be captured using molecular dynamics. Each sequence is 17 triplets long, centered about the potential cleavage site.

2.4.1.1 Details of the molecular dynamics simulations

All molecular dynamics simulations are performed with CHARMM (Brooks III, Bruccoleri et al. 1983). The initial structures generated with THeBuScr are imported into CHARMM, and the positions of polar hydrogens are determined as previously described (Brunger and Karplus 1988). The resulting structure is then energy minimized with 100 steps of steepest descent minimization. Each structure is then solvated with a sphere of TIP3P water molecules of 30Å radius (for a total of 3245 water molecules). Simulations are performed at 300K using a stochastic boundary approach (Salsas-Escat and Stultz 2009). The 30Å radius contains 7 triplets, centered at the cleavage site. Stochastic boundary methods explicitly simulate a relatively small region within the protein of interest (Brooks III and Karplus 1983; Brooks III, Brunger et al. 1985). Regions that fall outside of the simulated region (reaction region) are

retained in the computation but do not undergo full molecular dynamics. Instead they are restrained to adopt conformations that are near their initial coordinates.
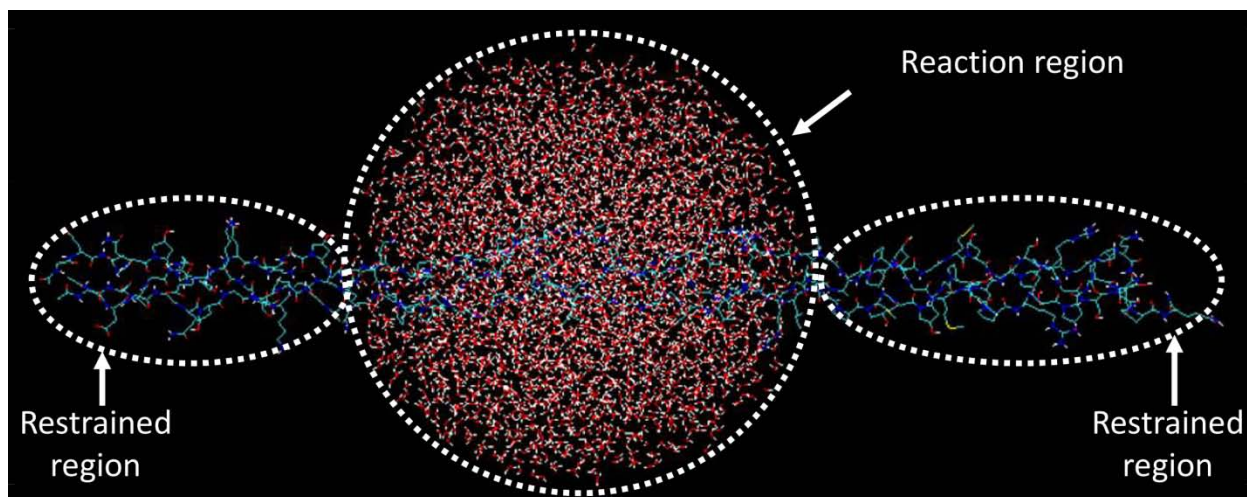
Figure 2.6. The stochastic boundary method (Brooks III and Karplus 1983; Brooks III, Brunger et al. 1985) applied to a collagen-like peptide from Chapter 2. The reaction region undergoes full molecular dynamics. The portion of the protein that falls within this sphere is exposed to water molecules at a density near that found in physiologic systems. Waters are kept within the reaction region using a stochastic boundary potential. The two restrained regions do not undergo molecular dynamics and are not solvated. This figure has been created with molecular visualization package VMD (Humphrey, Dalke et al. 1996).

Since we are interested in studying the conformational free energy landscape of regions nearby potential scissile bonds as they appear in the full sequence of type III collagen, a stochastic boundary method is appropriate. Simulations that do not employ a stochastic boundary would explore the conformational free energy landscape of these sequences in isolation. However, the isolated S1-S5 sequences likely do not adopt a triple helical fold in isolation at 300K. Using an online Collagen Peptide Tm Calculator (Persikov, Ramshaw et al. 2005), all of these sequences have a negative predicted melting temperature. Since these sequences are globally folded in the context of full triple helical collagen, the stochastic boundary method allows us to model the local structure of these sequences as they appear within full length collagen.

The solvated system is equilibrated for 150ps prior to the potential of mean force calculations. The potential of mean force is the free energy as a function of a reaction coordinate and in this work, the reaction coordinate, γ, corresponds to the radius of gyration of chain A of the tripeptide within the reaction region. This radius of gyration has been used as a reaction coordinate for unfolding for a number of systems, including collagen, and useful results have been obtained (Boczko and Brooks 1995; Brooks 1998; Stultz 2002; Salsas-Escat and Stultz 2009). The radius of gyration γ corresponds to:

$$\gamma = \sqrt{\frac{\sum_{i=1}^{N} m_i R_i^2}{\sum_{i=1}^{N} m_i}} \tag{2.1}$$

where m_i is the mass of a given atom and R_i is the distance of such atom to the center of mass.

The starting γ value for each sequence is determined from the average backbone radius of gyration of the non restrained central 7 triplets of each peptide from the last 100ps of equilibration. The reaction coordinate is sampled by increments and decrements of 0.1Å, for a total of 20 windows for each sequence. Each window is simulated for a total of 50ps long, with the last 30ps used for the production dynamics. During production dynamics, radius of gyration data is saved each 0.01ps, for a total of 3000 data points per window. Free energy profiles ae computed from these data using WHAM (Kumar, Rosenberg et al. 1992; Kumar, Rosenberg et al. 1995).

2.4.2 Essentials of the WHAM method

For a description on how to obtain an unbiased potential of mean force from a biased simulation, refer to section 1.4.3 in Chapter 1. From the simulations at each window we can calculate the biased $p_i^*(\gamma)$ probability and the bias potential $U_{bias,i}(\gamma)$ is known. The weighted histogram analysis method (WHAM) allows to determine the best set of values for F_i (Kumar, Rosenberg et al. 1992; Kumar, Rosenberg et al. 1995).

The two equations of the WHAM method are:

$$p(\gamma) = \sum_{i=1}^{N_W} \frac{n_i p_i^*(\gamma)}{\sum_{j=1}^{N_W} n_j e^{-\beta(U_{bias,j}(\gamma)-F_j)}} \tag{2.2}$$

$$e^{-\beta F_i} = \int e^{-\beta(U_{bias,j}(\gamma))} p(\gamma) \, d\gamma \tag{2.3}$$

where N_W is the number of biased simulations and n_i the number of data points for simulation i. The WHAM method starts by making a guess for all the F_i in Equation (2.2) and calculate a new set of F_i in Equation (2.3). This method is iterated until the root mean square difference between successive sets of F_i is less than 10^{-6}.

2.4.3 Vulnerability calculations

The *Vulnerability Score*, *V*, of a given sequence is defined as the average solvent accessible surface (SAS) of atoms (H, N, C and O) that form the potential scissile bond in question, Equation (2.4), normalized by the maximum value possible of SAS, Equation (2.5):

$$\langle SAS \rangle = \int_{\gamma_{min}}^{\gamma_{max}} SAS(\gamma) p(\gamma) d\gamma \tag{2.4}$$

$$V \equiv \frac{\langle SAS \rangle_\gamma}{SAS_{max}} \cdot 100 \tag{2.5}$$

where $SAS(\gamma)$ is the solvent accessible surface of the atoms that form the scissile bond in structures having a radius of gyration equal to γ, and $p(\gamma)$ is the probability density function over γ:

$$p(\gamma) = \frac{e^{-\beta W(\gamma)}}{\int_{\gamma_{min}}^{\gamma_{max}} e^{-\beta W(\gamma')} d\gamma'} \tag{2.6}$$

where β is $1/k_B T$, k_B is the Boltzmann constant, T is the temperature, $W(\gamma)$ is the pmf for a given sequence and γ$_{min}$ and γ$_{max}$ denote values of the reaction coordinate where the overall probability falls below 10^{-6}.

Calculating a *Vulnerability Score* entails first calculating the solvent accessible surface area for conformations with different radii of gyration. *Vulnerability Scores* are expressed as a percentage of the maximum SAS that a G-I or G-L dipeptide can have in solution. Hence to determine the maximum possible vulnerability, we simulate G-I and G-L dipeptides for 150ps at 300K (50ps equilibration, 100ps production) in a 20Å sphere of water (approximately 1000 TIP3P water molecules). The N-terminal nitrogen atom is fixed to avoid drifting of the dipeptide outside of the water sphere during the simulation. To calculate SAS_{max} for the reference dipeptides, the same algorithm is applied to individual structures resulting from the production

dynamics (1000 structures) and computing its average. We ensure that the calculated *Vulnerability Scores* are well converged in a given simulation window (Figure 2.7).
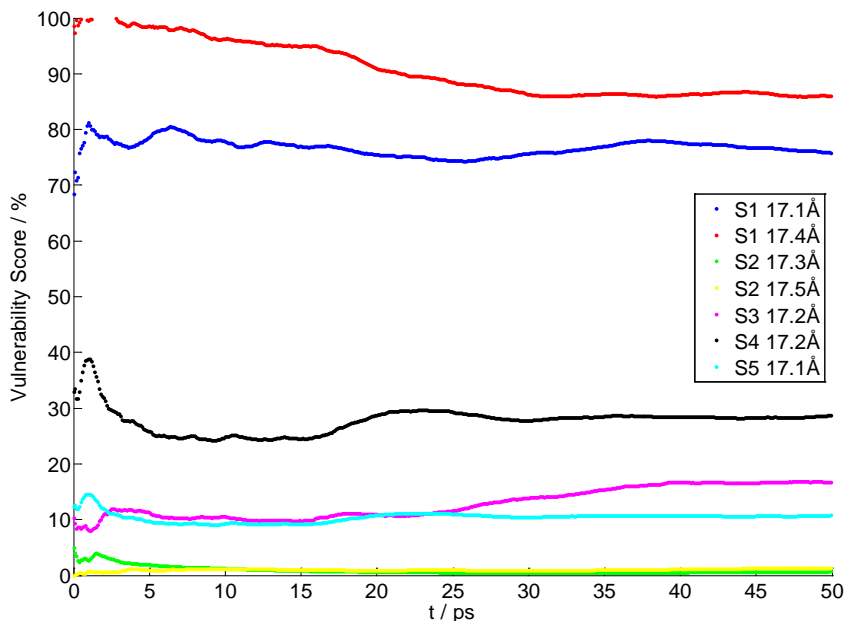
Figure 2.7. Average *Vulnerability Score* for the windows closest to the energy minima of each sequence.

2.4.4 Flexibility calculations

The *Flexibility Score, F,* which quantifies the number of conformational states that the sequence can adopt in solution, is related to the configurational entropy of the peptide as follows (McQuarrie 2000):

$$F \equiv -k_B T \int_{\gamma_{min}}^{\gamma_{max}} p(\gamma) log\big(p(\gamma)\big) d\gamma \tag{2.7}$$

where k_B is the Boltzmann constant, T the temperature and p the probability density described in Equation (2.6). The limits of integration are the same as those outlined in the *Vulnerability Score*.

Chapter 3:
Investigating the Influence of Unfolded States on Cleavage Site Specificity

Chapter 3 is adapted in part from:

and

3.1   Introduction

In Chapter 2 we determined that locally unfolded states are ubiquitous in type III collagen, and that the conformational ensemble at the actual collagenase cleavage site is the most vulnerable and flexible. In this section, we want to study whether the binding of the catalytic domain to these unfolded states is sufficient to determine cleavage site specificity. That is, even in the absence of the hemopexin-like domain, which is believed to encode cleavage site specificity (Perumal, Antipova et al. 2008; Erat, Slatter et al. 2009), can we still observe a single site cleavage at the actual cleavage site in triple helical collagen?

In order to investigate this, we first develop a novel rigid body docking method in which we dock 300 unfolded structures of each energy minima from Figure 2.2 into the catalytic domain of MMP8. The question we try to answer here is very simple. That is, since we have a heterogeneous population of unfolded states across the 5 potential cleavage sites, do any of them fit well into the active site of a collagenase? In short, we find that the locally unfolded states in sequence S1, the actual cleavage site, fit best into the active site of a collagenase. We then validate the docking results with the experimental degradation of type III collagen by the catalytic domain of MMP8 or MMP1 and observe that single cleavage at the true collagenase cleavage site can be obtained by the catalytic domain of MMP8 or MMP1 alone.

Furthermore, we look to generalize our experimental results for type III collagen to other collagen types. As stated in the Introduction chapter of this thesis, a number of experiments have explored whether collagen can adopt conformations at low temperatures that could, in principle, be recognized and cleaved by proteases. While type III collagen can be cleaved by general proteases at temperatures below its melting temperature (Miller, Finch et al. 1976; Wang, Chan et al. 1978; Birkedal-Hansen, Taylor et al. 1985), this is not the case for type I collagen. Early

experiments incubated type I collagen with trypsin at 25ºC and no collagen degradation was observed (Miller, Finch et al. 1976). More recent experiments incubated type I collagen with high concentrations of the catalytic domain of MMP8 (CMMP8) at room temperature and, again, no degradation was found (Schnierer, Kleine et al. 1993). Similar results have been reported with the catalytic domain of MMP1 (CMMP1) (Chung, Dinakarpandian et al. 2004). If the region near the collagenase cleavage site is able to spontaneously adopt partially unfolded states in solution that can be bound by the collagenase cleavage site, then one would expect incubation of the catalytic domain of MMPs with collagen to result in collagenolysis. Since this is not the case, it is difficult to reconcile these latter experimental observations with other studies (mentioned in the Introduction chapter) that suggest that type I collagen adopts partially unfolded states in the vicinity of the cleavage site in solution (Kadler, Hojima et al. 1988; Fiori, Sacca et al. 2002; Nerenberg and Stultz 2008). Consequently, these latter degradation experiments support a theory where the collagenase cleavage site does not spontaneously adopt partially unfolded states in solution (at temperatures below collagen's melting temperature), and that collagenolysis involves active unfolding of the collagen triple helix by MMPs. In this formalism the coordinated action of both the catalytic and hemopexin-like MMP domains leads to active unwinding of the triple-helical structure (Overall 2002; Chung, Dinakarpandian et al. 2004; Tam, Moore et al. 2004; Gioia, Monaco et al. 2007). In addition, recent data argues that the hemopexin-like domain may also play a role in determining cleavage site specificity by binding to specific secondary sites located near the unique collagenase cleavage site (Perumal, Antipova et al. 2008). Hence it has been argued that the hemopexin-like domain plays an essential role in both exposing the scissile bond and in ensuring that only one potential cleavage site is recognized by the collagenase.

In Chapter 3 we demonstrate for the first time that the degradation of type I collagen at room temperature - a temperature well below type I collagen's melting temperature - does not require the presence of the MMP hemopexin-like domain. Moreover, peptide bond hydrolysis with MMP mutants that contain only the catalytic domain occurs at the unique collagenase cleavage site and not at other potential cleavage sites. Thus both peptide bond hydrolysis and enzyme specificity are achieved with the catalytic domain alone at room temperature. As full length enzyme is thought to be necessary for collagenase-mediated unwinding, our data imply that enzyme-mediated unwinding is not required for collagenolysis *in vitro*. In Appendix 2, we analyze our data in light of a conformational selection model where thermal fluctuations at the cleavage site cause collagen to adopt unfolded conformations that are complementary to the collagenase active site. Overall, our findings suggest that both types I and III collagen can adopt locally unfolded states at room temperature and that collagenolysis occurs when collagenases cleave these locally unfolded states.

3.2 Results

3.2.1 Binding of potential cleavage sites to the MMP catalytic domain

To further explore whether the locally unfolded states sampled in the molecular dynamics simulations from Chapter 2 are consistent with the MMP catalytic site, we dock representative structures from each minimum arising from the various pmfs into the crystallographic structure of the MMP catalytic domain. Since there is considerable homology between the catalytic domains of different collagenases, we choose the catalytic domain from the crystallographic structure in the pdb that is solved to the highest resolution; i.e., MMP8 (PDB ID 2OY4) (Li, Brick et al. 1995; Berman, Westbrook et al. 2000; Bertini, Calderone et al. 2006).

Structures arising from simulations that restrain the triple-helical peptide to adopt a radius of gyration consistent with a local energy minimum are individually docked into the MMP8 binding site. Docking simulations are performed using a harmonic restraining potential that biases the rigid backbone of each partially unfolded structure to adopt the backbone position of a single stranded collagen-like peptide found in the binding site of a related MMP (Bertini, Calderone et al. 2006). Docked structures that have a low rms between the collagen-triple helix and the corresponding residues in a bound collagen-like peptide are most complementary to the active site (see Methods).

Figure 3.1 shows the results of the docking simulations. Structures belonging to minima S1min1 and S1min2 have the lowest rms differences, suggesting that these structures fit best within the MMP8 catalytic site. The complex with the lowest rms deviation arises from minimum S1min2 and is shown in Figure 3.2 A. The scissile bond involving the glycine and isoleucine residues from the partially unfolded chain fits within the catalytic site and is in close proximity to the catalytically essential Zn^{+2} metal of the active site (Figure 3.2 B); i.e., the

carbonyl oxygen from the scissile bond is 2.7Å from the catalytically important Zn^{+2} ion (Grams,

Reinemer et al. 1995). Moreover, the side chain of the isoleucine residue in position P_1' enters

the hydrophobic S_1' site (Bertini, Calderone et al. 2006), and the alanine residue sits in the

hydrophobic S_2' site (Brandstetter, Grams et al. 2001). Finally, the side chain of the leucine

residue in position P_2 is solvent exposed, as is the side chain of an asparagine residue in the

corresponding position, P_2, of the original single stranded peptide used as a reference for docking

(Bertini, Calderone et al. 2006). This suggests that the identity of the residue in position P_2 is not

critical for binding to CMMP8. These results show that the unfolded states that are most

vulnerable, corresponding to S1, also fit best into the active site of an MMP.
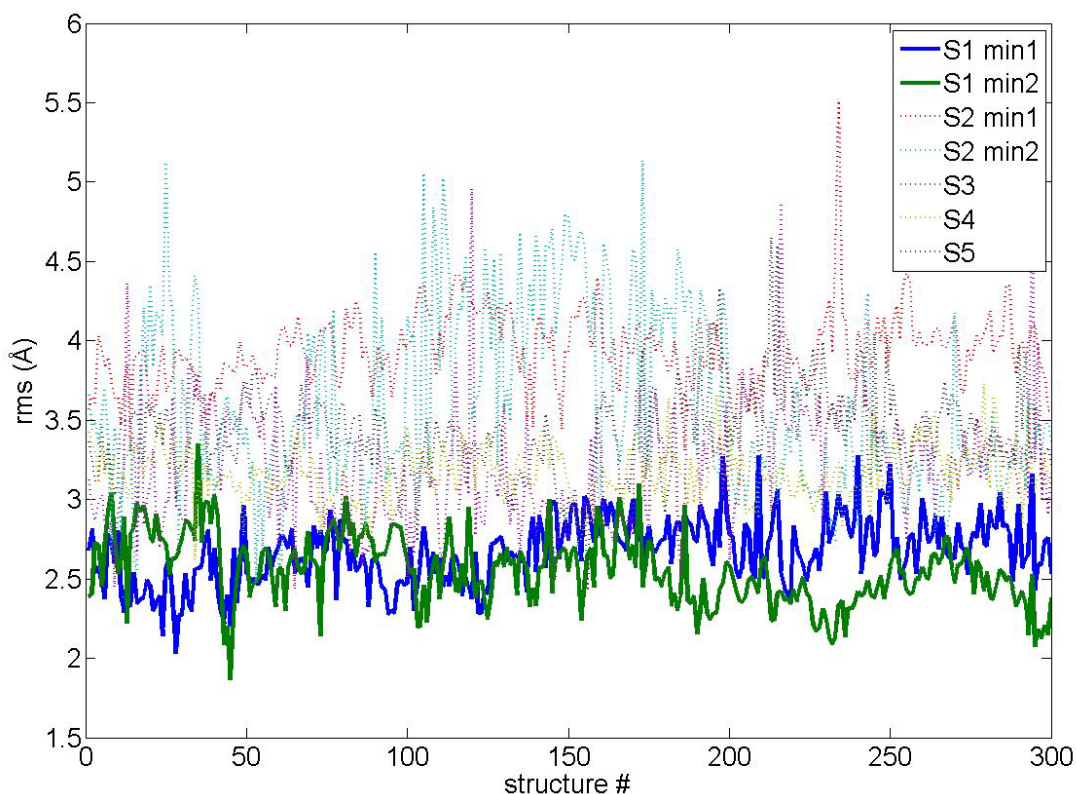
Figure 3.1. Results of docking studies. RMS values are computed between the backbone atoms of the collagen-like peptide in the co-crystal structure of MMP8 and the corresponding backbone atoms in the chain A of the triple helical peptide. All 300 structures from each energy minimum shown in Figure 2.2 are individually docked into CMMP8.
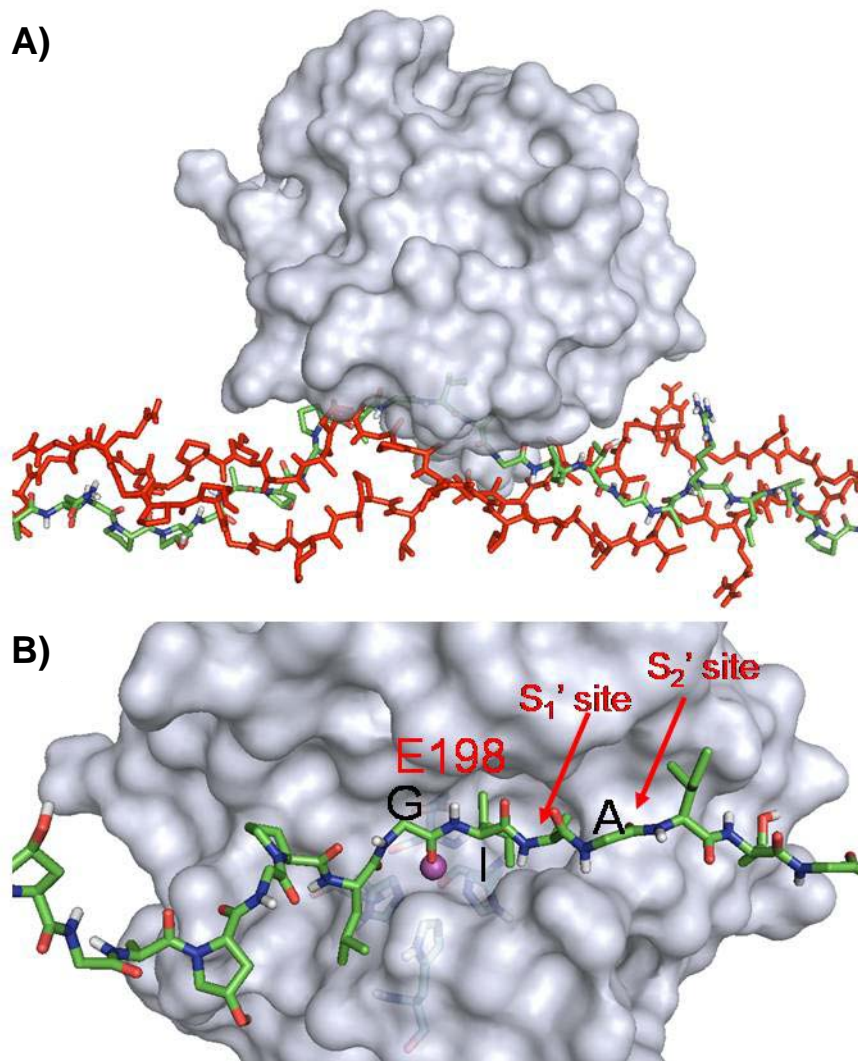
Figure 3.2. Molecular view of the best docking result. A) Overview of the best docked structure of S1 into the catalytic site of CMMP8. Chain A, which is unfolded in our simulations, is shown in stick representation and carbon atoms colored in green, while chains B and C are colored in red. The structure of CMMP8 is transparent. B) The catalytic site of the bound complex with chain A present (chains B and C deleted). The catalytically important E198 is shown, together with the 3 histidine residues that coordinate the Zn^{+2} ion. The isoleucine side chain sits in the S_1' site, while the alanine side chain sits in the S_2' site.

3.2.2 Collagenolysis without a hemopexin-like domain

In the previous section we show that the catalytic domain of a collagenase binds type III collagen at its actual cleavage site forming a complex that fulfills the putative interactions necessary for collagen degradation. While these computational results indicate that the local unfolding at the S1 site in type III collagen may help select for cleavage at that site and determine cleavage site specificity, it is not clear if such complex does occur experimentally. For this reason, we want to test, using collagen degradation experiments, whether the interaction between the catalytic domain of two collagenases (CMMP8 and CMMP1) and type III collagen results in cleavage at a single site like the docking results would suggest. Furthermore we want to generalize these results to other collagen types. Since locally unfolded states at the true collagenase cleavage site of type I collagen have been described (Nerenberg and Stultz 2008), the same type of experiments are carried out with type I collagen against CMMP8 and CMMP1.

Moreover, these experiments will help us test the predictive power of the two state collagen degradation model developed in Appendix 1. In this conformational selection model, collagen is able to adopt either a well-folded native triple-helical state or a vulnerable state where the region near the collagenase cleavage site is unfolded and solvent exposed (Nerenberg, Salsas-Escat et al. 2008). MMPs can bind to either state, but collagenolysis only occurs when MMPs bind to vulnerable states. The model is based on the premise that collagen adopts different conformations in solution and that collagenolysis occurs when the appropriate conformers are selected by the enzyme. A re-examination of collagen degradation experiments suggests that the failure to observe collagenolysis with MMP deletion mutants, which contain only the catalytic domain, is due to the fact that these mutant enzymes bind partially unfolded states of collagen with reduced affinity relative to the full length enzyme. A corollary of this result is that

collagenolysis could occur if collagen is exposed to relatively high concentrations of mutant enzymes and relatively long incubation times are used (Nerenberg, Salsas-Escat et al. 2008). To test this, we expose types I and III collagen to both the catalytic domains of MMP8 and MMP1 (denoted as CMMP8 and CMMP1, respectively) at high concentrations and for relative long incubation times.

3.2.2.1 Type III collagen degradation by CMMP8 and CMMP1

Human type III collagen expressed in *pichia Pastoris*, is incubated with high concentrations of CMMP1 at room temperature, a temperature below the melting temperature of type III collagen (Engel and Bachinger 2000). Figure 3.3 shows the degradation profile of type III collagen against the catalytic domain of MMP8 (CMMP8). Similar results are obtained with CMMP1 (Figure 3.4). In both figures, degradation of type III collagen by the catalytic domains of MMP1 and MMP8 result in the familiar ¾α1(III) and ¼α1(III) fragments that are obtained when type III collagen is degraded by a full length collagenase. This shows that type III collagen can be cleaved by the catalytic domain of a collagenase alone, validating the results of our docking simulation. Moreover, cleavage occurs at the cleavage site S1, suggesting that the hemopexin domain is not necessary for cleavage site specificity *in vitro* at room temperature.
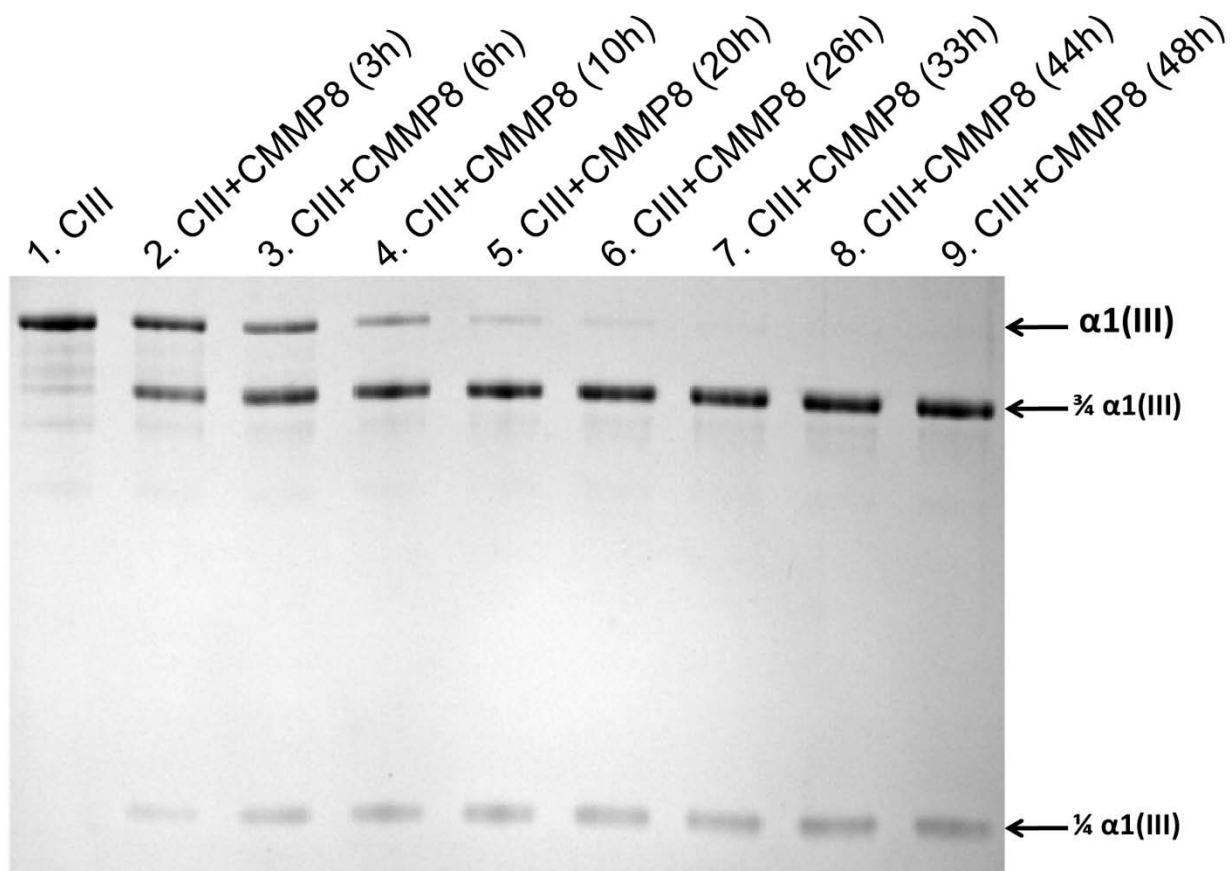
Figure 3.3. Degradation profiles of type III collagen at room temperature with CMMP8. Intact type III collagen bands $\alpha 1(III)$ and $\alpha 1(III)$ ¾ and ¼ degradation bands are shown. *Lane 1:* Type III collagen (200μg/ml, 0.66μM). *Lanes 2 to 9:* Type III collagen (200μg/ml, 0.66μM) incubated with the catalytic domain of MMP8 (CMMP8) (33μg/ml, 1.6μM) for 3, 6, 10, 20, 26, 33, 44 and 48h, respectively.
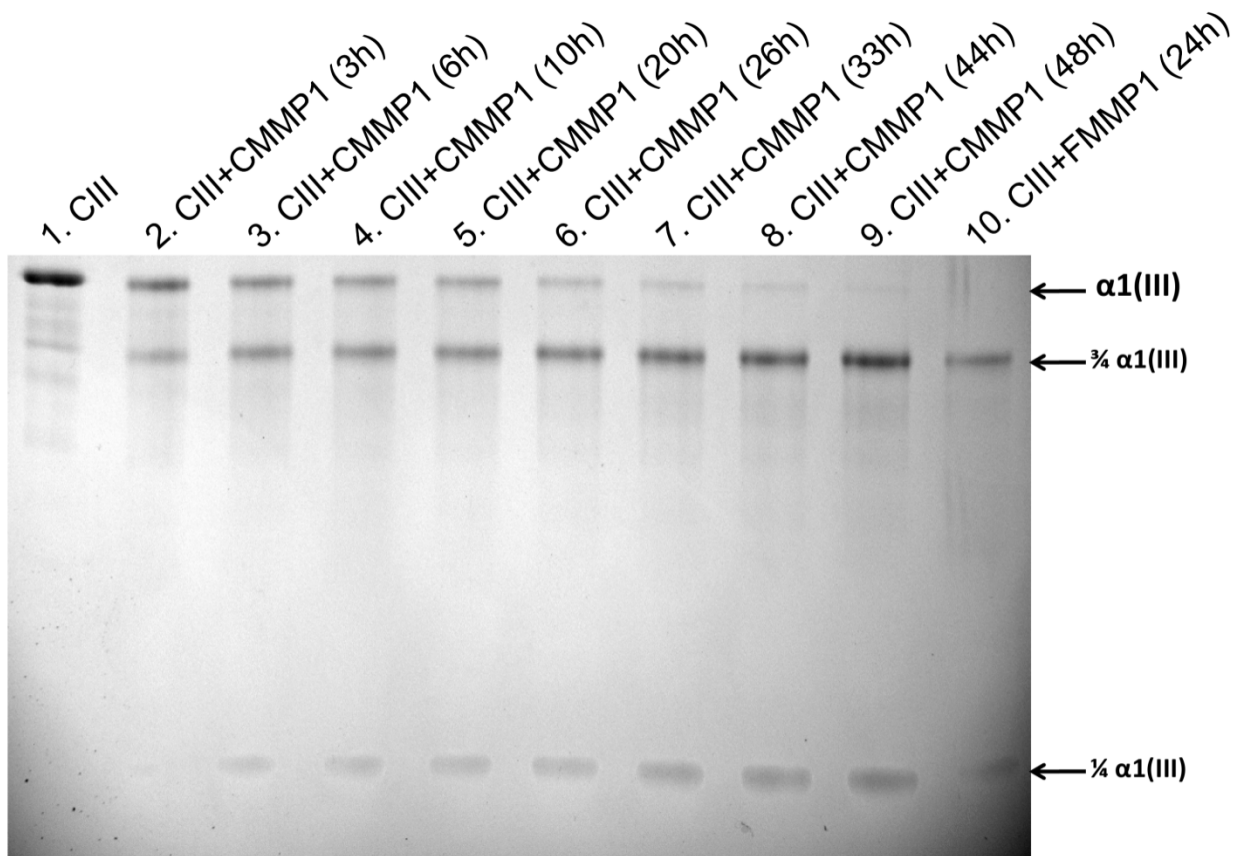
Figure 3.4. Degradation profiles of type III collagen at room temperature with CMMP1. Intact type III collagen bands α1(III) and α1(III) ¾ and ¼ degradation bands are shown. *Lane 1:* Type III collagen (200μg/ml, 0.66μM). *Lanes 2 to 9:* Type III collagen (200μg/ml, 0.66μM) incubated with the catalytic domain of MMP8 (CMMP1) (33.3μg/ml, 1.6μM) for 3, 6, 10, 20, 26, 33, 44 and 48h, respectively. *Lane 10:* Type III collagen (200μg/ml, 0.66μM) incubated with full length MMP1 (FMMP1) (1.7μg/ml, 30nM) for 24h.

3.2.2.2 Type I collagen degradation by CMMP8 and CMMP1

Bovine type I collagen is incubated with high concentrations of CMMP8 at room temperature - a temperature well below the melting temperature of type I collagen (Privalov 1982; Leikina, Mertts et al. 2002). After 48h type I collagen degradation is observed in solutions containing CMMP8 (Figure 3.5 A lanes 1-7). Degradation bands exhibit the familiar the ¾ and ¼ fragments that are associated with cleavage at the unique collagenase cleavage site by the wild

type collagenase MMP1 (FMMP1, Figure 3.5 A lane 9). N-terminal amino acid sequencing of the ¼α1 and¼α2 bands confirms that the CMMP8 deletion mutant cleaves at the unique cleavage site recognized by wild-type enzyme (data not shown). To demonstrate that our results are not specific to MMP8, we expose type I collagen to a deletion mutant containing only the catalytic domain of MMP1 (Figure 3.5 B lanes 1-5). Despite the fact that higher concentrations of enzyme and longer incubation times are used, type I collagen degradation with CMMP1 is significantly less efficient than that observed with CMMP8. The ¼ α1 and α2 bands from the CMMP1 reaction cannot be sequenced due to the low amount of collagen that is degraded by CMMP1. However, the resulting cleavage pattern is the same as that observed with FMMP1 and CMMP8. In bovine type I collagen, the closest potential cleavage sites of sequence G-[I/L]-[A/L] are 48 and 30 residues away from the true cleavage site in the α1 and α2 chains, respectively (Fietzek, Rexrodt et al. 1972; Fietzek, Wendt et al. 1972; Rauterberg, Fietzek et al. 1972; Wendt, von der Mark et al. 1972; Fietzek, Rexrodt et al. 1973; Fietzek and Kuhn 1975; Shirai, Hattori et al. 1998). If these potential sites were cleaved, the pattern of fragments observed would be significantly different from the FMMP1 control, suggesting that CMMP1 mediated cleavage, as is the case with CMMP8, occurs at the same site.

To confirm that our findings are not due to contamination of our collagen samples with unfolded type I collagen chains (gelatin), we incubated CMMP8 and CMMP1 with gelatin. Both CMMP8 and CMMP1 cleave gelatin at several sites yielding multiple degradation bands on SDS-PAGE (Figure 3.5 A lane 8, Figure 3.5 B lane 6) (Welgus, Jeffrey et al. 1982; Chung, Dinakarpandian et al. 2004). Since these bands are not seen when CMMP8 and CMMP1 are incubated with collagen, contamination of our collagen sample with unfolded collagen chains does not explain our results.

To further demonstrate that our results cannot be explained by contamination of the original collagen sample with low concentrations of full length MMPs, we incubated solutions of type I collagen containing 4-aminophenyl mercuric acetate (APMA), to activate any latent enzyme, but without any added collagenases. SDS-PAGE of the solutions after 6 days of incubation did not exhibit any degradation bands, thereby arguing that contamination with latent enzyme does not explain our findings (Figure 3.6 A). Similarly, to rule out the presence of other contaminating proteases in our CMMP1 and CMMP8 preparations, we incubated type I collagen and gelatin with CMMP1/8 in presence of 50mM EDTA (Figure 3.6 B and C). While EDTA is a known inhibitor of collagenases, it does not affect the activity of non-metal-dependent proteases, many of which can readily cleave gelatin (Courts 1955; Gendron, Grenier et al. 1999). Figure 3.6 B shows the result of incubating collagen and CMMP8 or CMMP1 in the presence of 50mM EDTA. In both cases, collagen degradation cannot be observed. Similarly, incubations of CMMP1 and CMMP8 with gelatin and EDTA did not reveal any evidence of degradation as shown by SDS-PAGE (Figure 3.6 C). Both results suggest that there are no proteases contaminating the CMMP1 and CMMP8 samples. For this reason, we can conclude that the degradation profiles obtained in Figure 3.5 are the result of collagen and gelatin degradation by CMMP1 and CMP8. Interestingly, Figure 3.6 C shows that CMMP8 completely cleaves gelatin after 3 hours but that CMMP1 preferentially cleaves the α2 chain. This is expected since, at 37C, the reported k_{cat} of α1 gelatin chains cleavage by MMP1 is $0.064s^{-1}$, while it is $0.208s^{-1}$ for α2 chains (Welgus, Jeffrey et al. 1982).
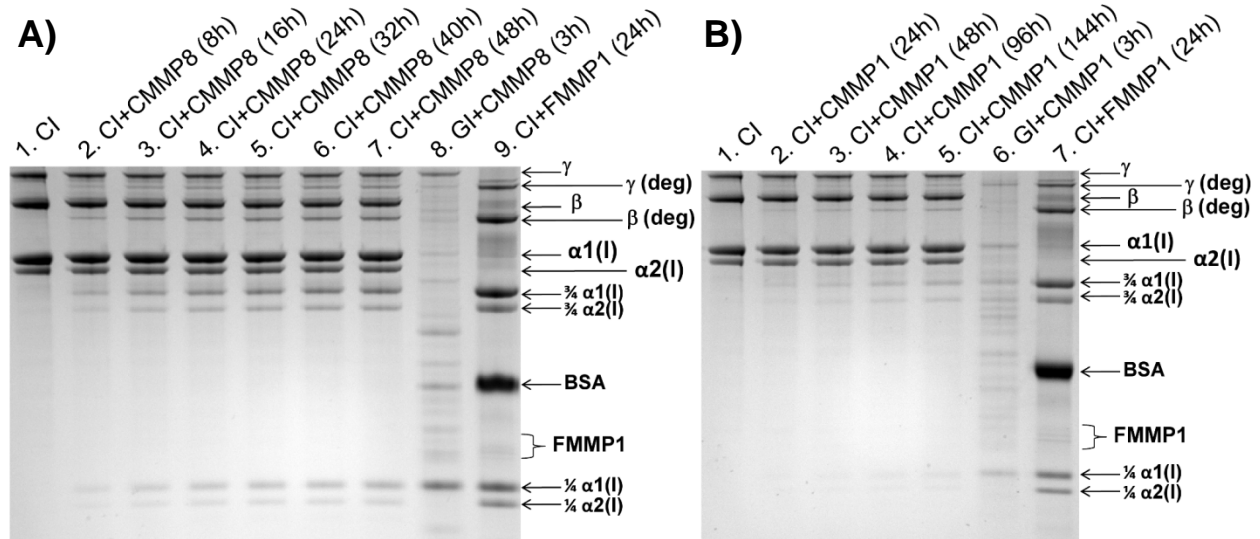
Figure 3.5. Degradation profiles of type I collagen at room temperature with CMMP8 and CMMP1. Intact type I collagen (CI) bands include monomeric α1(I) and α2(I) bands, dimeric β bands and trimeric γ bands. The β and γ aggregates correspond to N-terminally crosslinked collagen molecules (Veis and Anesey 1965; French, Mookhtiar et al. 1987; Shigemura, Ando et al. 2004; Morimoto, Kawabata et al. 2009). Type I collagen degradation bands include α1(I) and α2(I) ¾ and ¼ fragments, and degradation of crosslinked chains, β_{deg} and γ_{deg}. A) Type I collagen incubated with CMMP8. *Lane 1:* Type I collagen (150μg/ml, 0.5μM). *Lanes 2 to 7:* Type I collagen (150μg/ml, 0.5μM) incubated with the catalytic domain of MMP8 (CMMP8) (25μg/ml, 1.2μM) for 8, 16, 24, 32, 40 and 48h, respectively. *Lane 8:* Type I collagen gelatin (GI, 150μg/ml, 1.5μM) incubated with CMMP8 (13.8μg/ml, 0.67μM) for 3 h. *Lane 9*: Type I collagen (150μg/ml, 0.5μM) incubated with full length MMP1 (FMMP1) (1.7μg/ml, 30nM) for 24h. This lane contains a bovine serum albumin (BSA) band since FMMP1 is supplied in a buffer containing 1mg/ml BSA. B) Type I collagen incubated with CMMP1. *Lane 1:* Type I collagen (150μg/ml, 0.5μM); *Lanes 2 to 5:* Type I collagen (150μg/ml, 0.5μM) incubated with the catalytic domain of MMP1 (CMMP1) (40μg/ml, 2.0μM) for 24, 48, 96 and 144 h, respectively. *Lane 6:* Type I collagen gelatin (GI 150μg/ml, 1.5μM) incubated with CMMP1 (16.6μg/ml, 0.83μM) for 3 h. *Lane 7*: Type I collagen (150μg/ml, 0.5μM) incubated with full length MMP1 (FMMP1) (1.7μg/ml, 30nM) for 24h. As with lane 9 in A), this lane contains a BSA band since FMMP1 is supplied in a buffer containing 1mg/ml BSA.
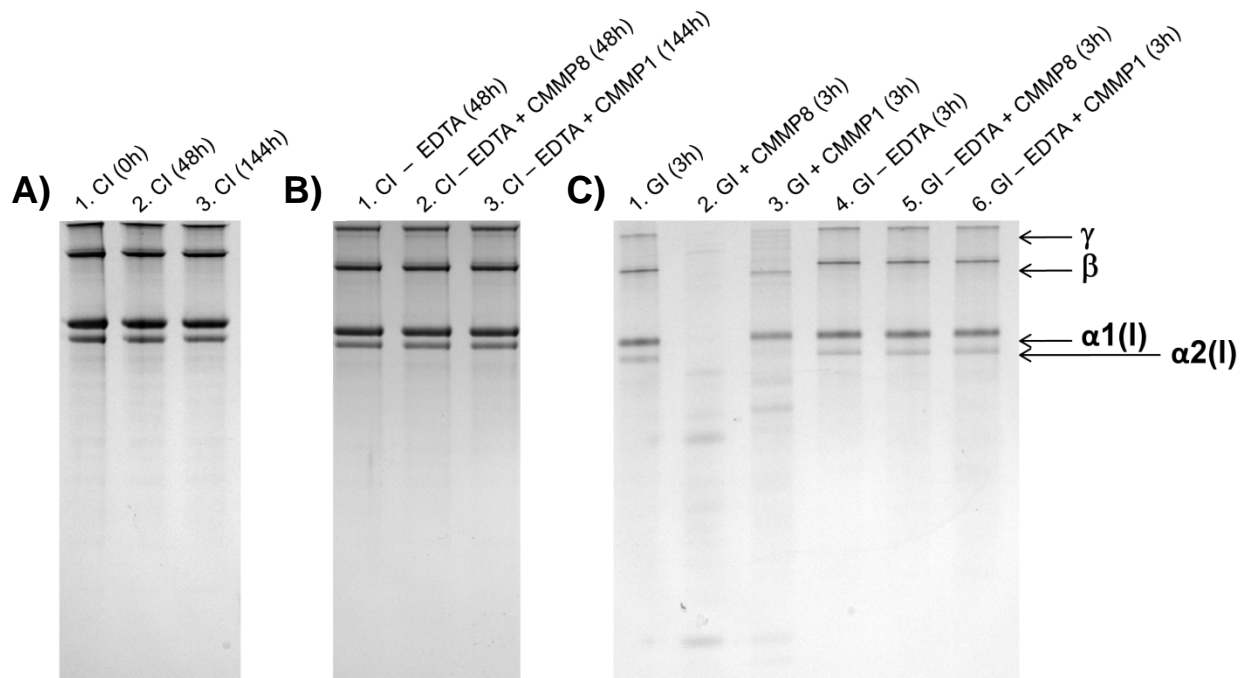
Figure 3.6. Testing for the presence of contaminating MMPs and other proteases. A) Type I collagen (150μg/ml, 0.5μM) incubated in the presence of APMA without the addition of any collagenases, for 0, 48 and 144h. No degradation bands can be observed, which argues that no contaminating MMPs are present. B) Testing for the presence of contaminating proteases in the CMMP1/8 solutions with collagen. *Lane 1*: Type I collagen (150μg/ml, 0.5μM) in reaction/EDTA buffer in the absence of collagenases incubated for 48h. *Lane 2*: Type I collagen (150μg/ml, 0.5μM) incubated against CMMP8 (25μg/ml, 1.2μM) in reaction/EDTA buffer. Degradation of gelatin cannot be observed. *Lane 3*: Type I collagen (150μg/ml, 0.5μM) incubated against CMMP1 (40μg/ml, 2.0μM) in reaction/EDTA buffer. Degradation of collagen cannot be observed. C) Testing for the presence of contaminating proteases in the CMMP1/8 solutions with gelatin. *Lane 1*: Type I collagen gelatin (GI 150μg/ml, 1.5μM) in reaction buffer in the absence of collagenases incubated for 3h. *Lane 2*: Type I collagen gelatin (GI 150μg/ml, 1.5μM) incubated against CMMP8 (1.4μg/ml, 67nM) in reaction buffer. Degradation of gelatin can be observed. *Lane 3*: Type I collagen gelatin (GI 150μg/ml, 1.5μM) incubated against CMMP1 (1.4μg/ml, 69nM) in reaction buffer. Degradation of gelatin can be observed. *Lane 4*: Type I collagen gelatin (GI 150μg/ml, 1.5μM) in reaction/EDTA buffer in the absence of collagenases incubated for 3h. *Lane 5*: Type I collagen gelatin (GI 150μg/ml, 1.5μM) incubated against CMMP8 (1.4μg/ml, 67nM) in reaction/EDTA buffer. Degradation of gelatin cannot be observed. *Lane 6*: Type I collagen gelatin (GI 150μg/ml, 1.5μM) incubated against CMMP1 (1.4μg/ml, 69nM) in reaction/EDTA buffer. Degradation of gelatin cannot be observed.

3.3 Discussion

In Chapter 2 we characterized the conformational ensemble of locally unfolded states in the different cleavage sites in type III collagen. Since the unfolded states at the actual cleavage site S1 are more vulnerable, we hypothesize that collagenases will preferentially bind and cleave these states. To test that, we develop a novel rigid docking method in which locally unfolded states in the context of triple helical collagen are bound by the catalytic domain of a collagenase. The docking studies suggest that the locally unfolded structures for sequence S1 (corresponding to the true collagenase cleavage site) are most complementary to the MMP catalytic site. The partially unfolded triple helical peptide fits within the catalytic site, placing hydrophobic residues into the known hydrophobic pockets in the enzyme (Grams, Reinemer et al. 1995). In addition it has been proposed that for catalysis to occur the carbonyl oxygen of the glycine residue, which forms part of the scissile bond, must be in close proximity to the Zn^{2+} ion; i.e., this metal is believed to be important in the binding of the substrate and to stabilize the transition state of the reaction (Browner, Smith et al. 1995; Pelmenschikov and Siegbahn 2002). Our complex places the carbonyl carbon of the glycine residue in the GIA triplet within 2.7Å of this ion and therefore is consistent with the notion that binding preformed, and complementary, unfolded states of sequence S1 would lead to peptide bond hydrolysis.

To test whether the interaction described by the docking simulations would actually result into collagen degradation, we incubate type III collagen with the catalytic domain of MMP1 and MMP8. The resulting degradation pattern suggests that CMMP1 and CMMP8 bind and cleave type III collagen at the actual cleavage site, even in the absence of the hemopexin-like domain. This suggests that only the locally unfolded states at the true collagenase cleavage site are being bound and cleaved by the catalytic domain of MMP1 and MMP8, which correlates with the

simulation results shown in Chapter 2. The same results are obtained with type I collagen. These results confirm the predictions of a two-state collagen degradation model in which collagenases bind to preexisting locally unfolded states, see Appendix 1 (Nerenberg, Salsas-Escat et al. 2008).

Previous data suggested that non-catalytic domains have binding sites for collagen and degradation experiments using only the catalytic domain of MMPs confirmed the importance of these domains in hydrolyzing the unique scissile bond (Clark and Cawston 1989; Windsor, Birkedal-Hansen et al. 1991; Hirose, Patterson et al. 1993; Nerenberg, Salsas-Escat et al. 2008). More recent studies proposed that non-catalytic domains can bind specific amino acid sequences near the collagenase cleavage site in some collagen types, and that this serves as a mechanism to localize the enzyme to a region near the true cleavage site (Perumal, Antipova et al. 2008; Erat, Slatter et al. 2009). However, our computational results suggest that unfolded states of collagen exist in the absence of collagenases. Moreover, we show computationally and experimentally that these states can be bound and cleaved by collagenases containing only the catalytic domain, achieving cleavage site specificity in the absence of the hemopexin-like domain. Therefore, we propose that the hemopexin-like domain is not required for either collagenolysis or cleavage site specificity *in vitro* at room temperature. This can be explained in light of the presence of locally unfolded states at the actual cleavage site of collagenases that are more vulnerable and complimentary to the catalytic domain of collagenases than other unfolded states at other potential cleavage sites. While our results seem to contradict previous data that suggest that the hemopexin-like domain is required for collagenolysis, we still believe that this domain plays a key role in collagen degradation. In both collagen degradation experiments presented in Chapter 3, very high concentrations of CMMP1 and CMMP8 are required in order to observe degradation. This suggests that, while degradation occurs, CMMP1 and CMMP8 are very

inefficient enzymes when cleaving triple helical collagen at room temperature. For this reason, we propose that the hemopexin-like domain is required to increase the efficiency of these enzymes. A more detailed discussion to this end is presented in the Conclusions section of this thesis.

Our data are explained by the fact that the unique scissile bond locally unfolds at low temperatures. However, as the temperature increases, microunfolding of collagen leads to the exposure of other potential cleavage sites that can be recognized by MMP deletion mutants (Taddese, Jung et al. 2009). Indeed, recent data suggest that collagen is thermally unstable at body temperature and it has been postulated that microunfolding of distinct collagen regions facilitates fibril formation and collagen catabolism (Leikina, Mertts et al. 2002). As the MMP hemopexin-like domain has binding sites near the collagenase cleavage site, it may help to localize the enzyme to a specific scissile bond, thereby ensuring that the correct site is cleaved at body temperature (Perumal, Antipova et al. 2008). Therefore while our data argue that the hemopexin-like domain is not needed for cleavage site specificity at low temperatures, it may play an important role in determining specificity at body temperature.

Conformational changes in proteins that are associated with binding reactions are typically grouped into two large categories: induced fit and conformational selection. Unlike induced fit, where folding and binding occur concomitantly, conformational selection dictates that a protein's binding partner preferentially recognizes pre-formed complementary structures (Shoemaker, Portman et al. 2000; Tsai, Ma et al. 2001; Sugase, Dyson et al. 2007; Dobbins, Lesk et al. 2008; Lange, Lakomek et al. 2008). One difference between conformational selection and induced fit is that conformational selection presupposes that the protein conformation, which is complementary to a given binding partner, exists in solution even when the binding partner is

absent (Sugase, Dyson et al. 2007; Yoon, Venkatachalam et al. 2009). Hence information as to what mechanism is at play for any given protein may be obtained from an analysis of the conformational thermodynamics of the protein of interest in the absence of its binding partner. The results presented in Chapter 2 and Chapter 3 suggest that at temperatures below the melting temperature of type III collagen, the unique cleavage site adopts conformations that are complementary to the catalytic site while sequences corresponding to other potential cleavage sites do not. Hence, in addition to localization by non-catalytic domains, the subsequent binding of the catalytic domain to residues that contain the scissile bond may also help to ensure that the true cleavage site in types III collagen is recognized and cleaved.

The degradation data of type I collagen by CMMP1 and CMMP8 obtained in Chapter 3 serve as the basis for testing the conformational selection model presented in Appendix 1. In Appendix 2, the model is used to fit the degradation data and obtain information about the equilibrium constant between native and vulnerable states, as well as the binding constants of CMMP1 and CMMP8 to the vulnerable states. In short, the results show that, at room temperature, the native states at the true cleavage site in type I collagen are more stable than the vulnerable states, and that CMMP8 binds collagen ~2.5kcal/mol more favorably than CMMP1.

3.4 Methods

3.4.1 Docking simulations

A co-crystal structure of a cleaved hexapeptide (electron density is seen for only 5 residues in the active site) corresponding to the consensus sequence of types I and III collagen in the active site of the catalytic domain of MMP12 (PDB ID 2OXZ (Berman, Westbrook et al. 2000; Bertini, Calderone et al. 2006) solved at 1.8Å resolution) is used as the starting structure for the docking simulations (Bertini, Calderone et al. 2006). Given the structural similarity between the catalytic domains of MMP12 and MMP8 (PDB ID 2OY2 (Berman, Westbrook et al. 2000; Bertini, Calderone et al. 2006) solved at 1.5Å resolution), with a backbone RMSD of only 0.67Å, the two structures are aligned and the two peptide fragments are positioned in the active site of MMP8. Selected residues are mutated in the peptide fragment *in silico* in order to obtain the corresponding amino acid sequence of the type III collagen cleavage site. The two fragments are covalently linked together and the final structure is energy minimized for 100 steps using the steepest descent algorithm. This structure, consisting of the catalytic domain of MMP8 and a docked 5 residue peptide containing the collagenase scissile bond, serves as our reference structure for docking simulations.

Each collagen structure corresponding to a given stable state from the pmf is docked into the active site of MMP8 using the bound 5mer as a reference. A total of 300 structures are docked for each energy minimum from the following pmf windows: S1min1, γ=17.3Å; S1min2, γ=17.5Å; S2min1, γ=17.3Å; S2min2, γ=17.5Å; S3, γ=17.2Å; S4, γ=17.2Å; S5, γ=17.1Å. We fix the internal coordinates of the triple helical peptides, and the position of the MMP8 and 5mer atoms. We then use a harmonic biasing potential that minimizes the rms between the backbone

(nitrogen, α-carbon and carbonyl carbon) atoms of the bound 5mer and those same atoms in chain A of the triple helical peptide. Molecular dynamics are run for 20ps in vacuum with the fixed CMMP8 and reference peptide, and a rigid triple helical peptide, with the biasing potential turned on with a force constant of $5kcal/mol/Å^2$. When the triple helical peptide collides with the fixed MMP8, due to the predominant van der Waals clashes, the rms stops decreasing and fluctuates around a given value. Figure 3.8 shows the evolution of the docking rms from a typical docking run. In the early stages, collagen is far from CMMP8 and the rms is therefore large. When collagen approaches CMMP8 in order to align with the reference peptide, the rms lowers, and the vdW contacts become relevant, preventing collagen from fully aligning with the reference peptide. The structure with the lowest rms of all docks, (1.86Å) corresponding to S1min2, is energy minimized using the EEF1 (Lazaridis and Karplus 1999) implicit solvent model for 200 steps of steepest descent followed by 200 steps of conjugate gradient, while fixing CMMP8 and chains B and C.
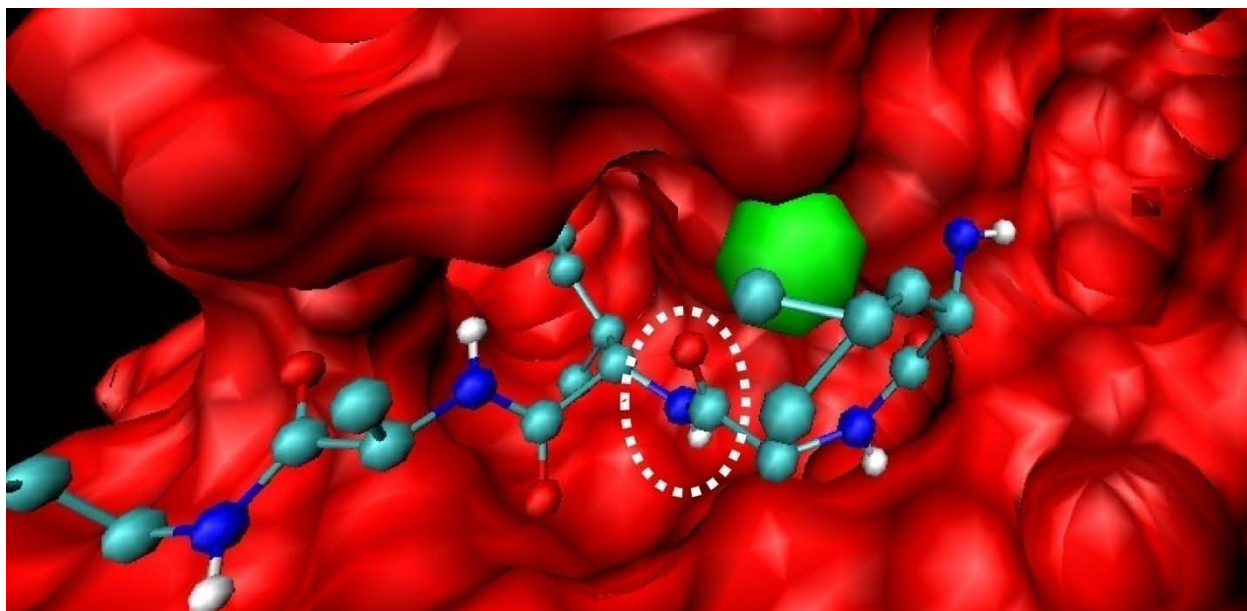
Figure 3.7. Detail of the catalytic domain of MMP8 with the 5 residue peptide bound in the catalytic site cleft. Inside the dotted white oval, the scissile bond between G~I is circled. The green sphere corresponds to the catalytically essential Zn^{2+} ion. This figure has been created with molecular visualization package VMD (Humphrey, Dalke et al. 1996).
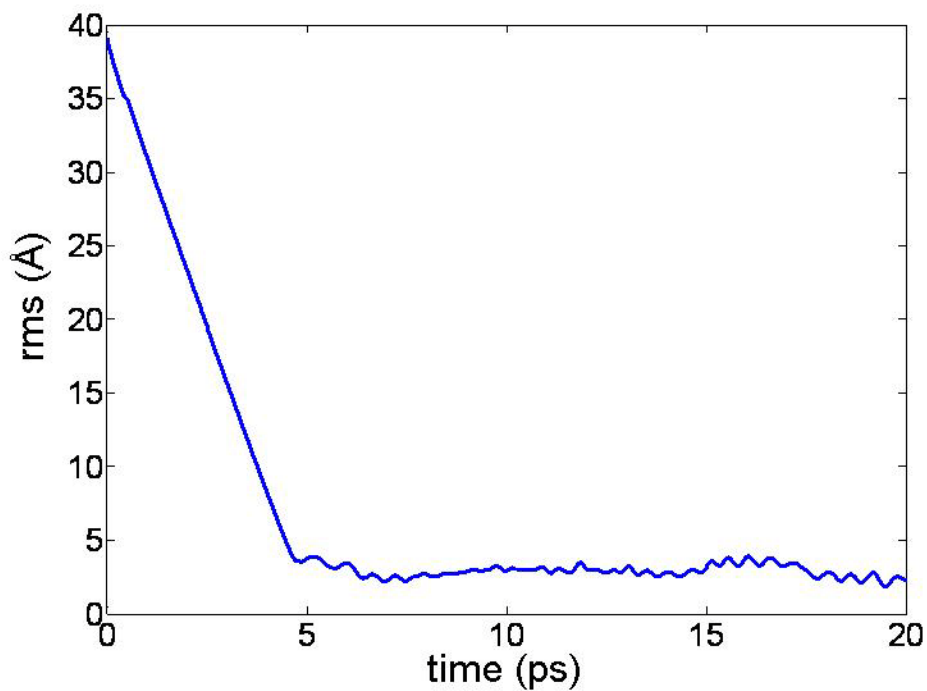
Figure 3.8. Evolution of rms during the docking trajectory for a typical docking run.

3.4.2 Collagen degradation experiments

3.4.2.1 Enzyme preparation and characterization

MMP mutants that only contain the catalytic domains of MMP1 (CMMP1) and MMP8 (CMMP8) are purchased from Enzo Life Sciences. Full length MMP1 (FMMP1) is purchased from Anaspec, and is provided in a preserving solution containing 1mg/ml of BSA. The specific activities of CMMP8 and CMMP1 are determined by the manufacturer at 37ºC to be 356U/μg and 1050U/μg, respectively. To verify that the mutant enzymes retained activity under our experimental conditions we measure the specific activities of CMMP1 and CMMP8 at 24ºC in our reaction buffer. Specific activities are determined using an assay based on the MMP1 Colorimetric Drug Discovery Kit from Enzo Lifesciences (BML-AK404-0001) (Weingarten and Feder 1985). In short, a 1mM solution of 5,5'-dithiobis(2-nitrobenzoic) acid Ellman's reagent, is prepared in reaction buffer. The enzyme (10nM CMMP1 or 13nM CMMP8, a 2500X dilution of the supplied enzyme) is then incubated in the latter buffer with 100μM of a colorimetric thiopeptolide of sequence Acetil-Proline-Leucine-Glycine-[2-mercapto-4-methyl-pentanoyl]-Leucine-Glycine-OEthyl (Weingarten, Martin et al. 1985) (Enzo Lifesciences BML-P125). This substrate is the same that is used by the manufacturer to measure the enzyme activity at 37ºC. The initial velocity is determined using the slope of the absorbance increase at 412nm after 3 minutes (Figure 3.9), yielding specific activities 289U/μg for CMMP8 and 766U/μg for CMMP1 (1U = 100pmol/min degradation of the thiopeptolide). The formula used to obtain the specific activity is based in the Beer-Lambert law, as follows:

$$A = l \cdot \varepsilon \cdot c \tag{3.1}$$

where A is the Absorbance, l is the path length in cm, ε is the molar absortivity ($13600M^{-1}cm^{-1}$ for 2-nitro-5-thiobenzoic acid, the product of the reaction) and c is the molar concentration. If we take derivatives with respect to time in both sides of the expression (3.1) and separate c as the ratio of moles over volume (V in L), we obtain:

$$\frac{dA}{dt} = \frac{l \cdot \varepsilon}{V} \cdot \frac{d(moles)}{dt} \tag{3.2}$$

since the activity is defined as the number of moles of substrate consumed in a period of time, isolating this quantity yields:

$$\frac{d(moles)}{dt} = \frac{V}{l \cdot \varepsilon} \cdot \frac{dA}{dt} \tag{3.3}$$

where $\frac{dA}{dt}$ (in units of absorbance per minute) is the slope in Figure 3.9. To obtain the specific activity, we only have to divide the activity value by the amount of enzyme used in the reaction, in µg.
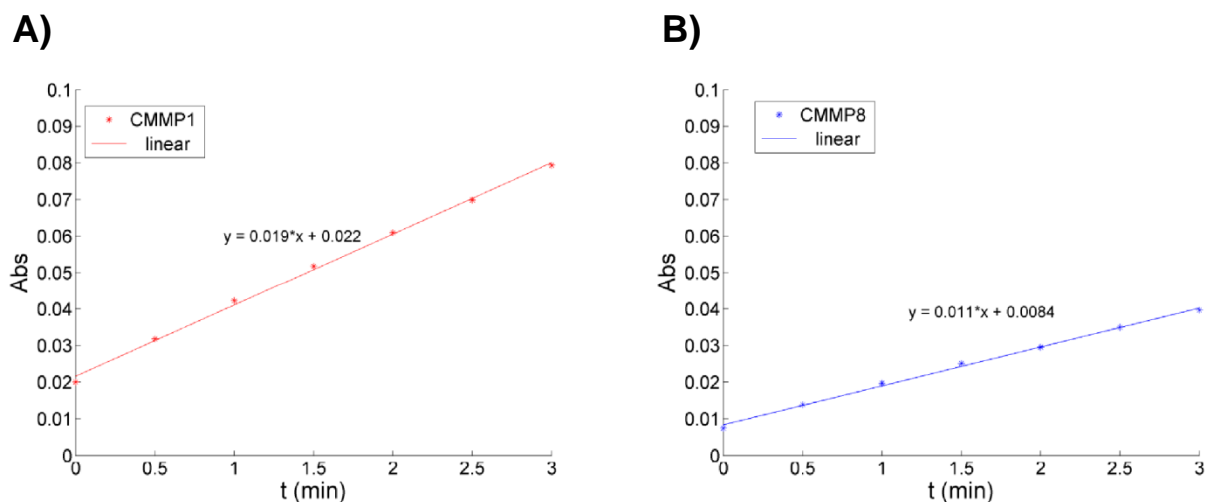
Figure 3.9. Determination of the specific activity of CMMP1 and CMMP8. Initial velocity measurement of the degradation of a thiopeptolide by A) CMMP1 and B) CMMP8.

112

3.4.2.2 Type III collagen degradation experiments

All experiments are carried out in a reaction buffer containing 100mM Tris HCl (VWR International), 10mM $CaCl_2$ (Sigma-Aldrich Co), pH 7.6. Human sequence type III collagen, expressed in *Pichia Pastoris* (Abcam), is obtained at 3.3mg/ml in 0.5M AcOH. Degradation experiments are performed at a collagen concentration of 200μg/ml (0.66μM) of type III collagen, by diluting the concentrated type III collagen into reaction buffer. Type III collagen expressed in yeast does not contain other collagen impurities. For this reason, collagen repurification is not needed.

3.4.2.3 Preparation of type I collagen

All experiments are carried out in a reaction buffer containing 100mM Tris HCl (VWR International), 10mM $CaCl_2$ (Sigma-Aldrich Co), pH 7.6. Bovine type I collagen, (BD Biosciences), is obtained at 3mg/ml in 0.012M HCl.

As purification of type I collagen often results in protein that is contaminated with type III collagen, purchased collagen samples are repurified using differential salt precipitation (Epstein 1974; Trelstad, Catanese et al. 1976; Miller and Rhodes 1982). First, type I collagen is diluted to a concentration of 0.5mg/ml (1.67μM) in 0.5M AcOH. The solution is then dialyzed against low salt buffer (0.1M NaCl, 50mM Tris, pH 7.5, at 4ºC) followed by dialysis against a high salt buffer (1.8M NaCl, 50mM Tris, pH 7.5, at 4ºC), in which type III collagen preferentially precipitates (Trelstad, Catanese et al. 1976). At this point the sample is centrifuged at 4ºC for 30min at 16000g and the supernatant, containing purified type I collagen, is then dialyzed against reaction buffer. Purity of the final type I solution is confirmed by running degradation experiments with full length MMP1 and no type III collagen degradation products are observed (Figure 3.10).
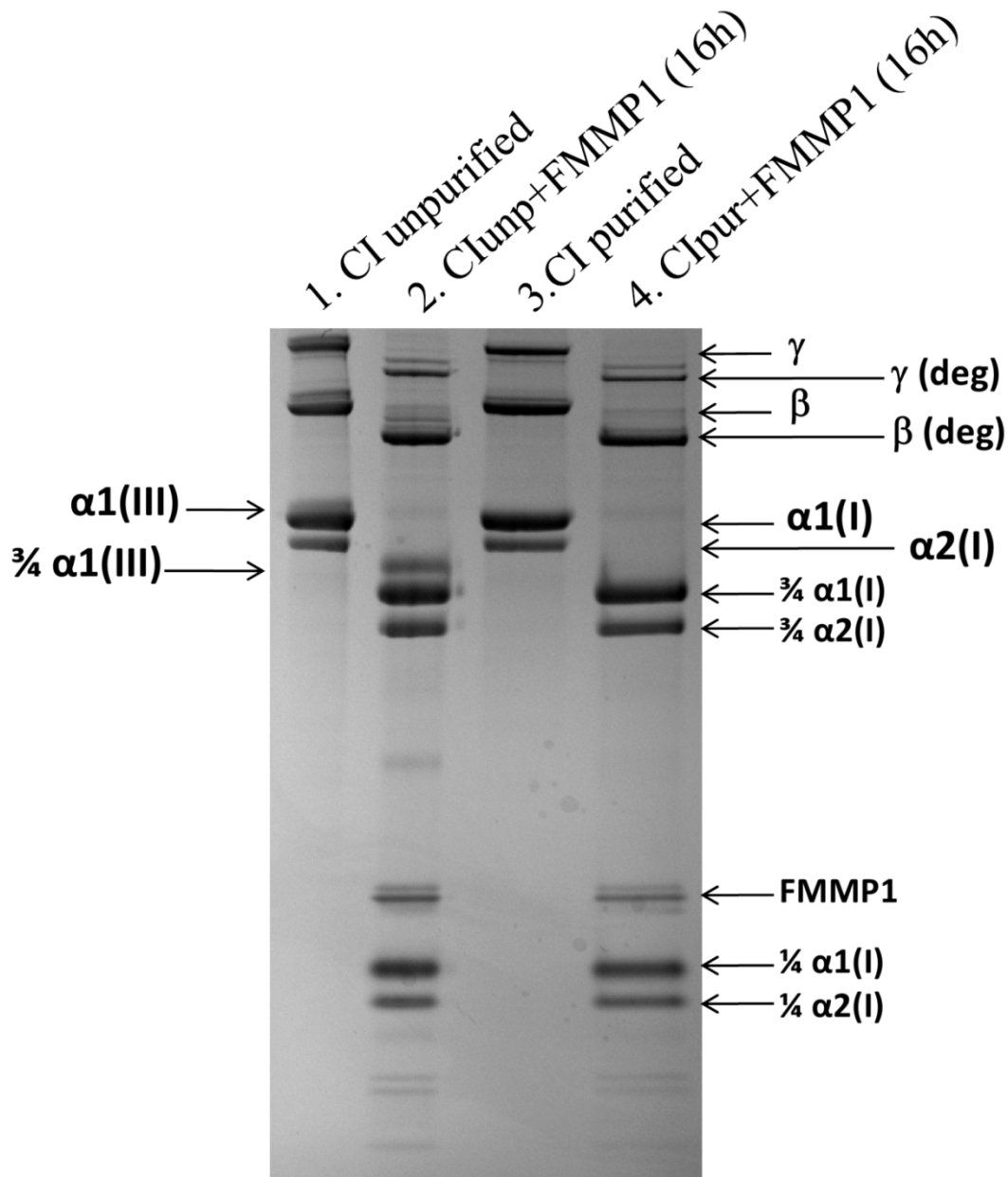
Figure 3.10. Purification profile of type I collagen. *Lane 1:* Unpurified type I collagen. α1(I), α2(I) and aggregation bands can be seen along an impurity of type III collagen, α1(III) band. *Lane 2:* Degradation of unpurified type I collagen. The degradation bands of type I collagen, ¾ and ¼ α1(I) and α2(I) bands can be seen. A small ¾ α1(III) band can also be observed. *Lane 3:* Purified type I collagen. Only α1(I), α2(I) and aggregation bands can be seen. *Lane 4:* Degradation of purified type I collagen. Only the degradation bands of type I collagen, ¾ and ¼ α1(I) and α2(I) bands can be seen.

3.4.2.4 Type I collagen degradation experiments

Degradation experiments are performed at a collagen concentration of 150µg/ml (0.5µM). Prior studies suggest that the critical concentration for fibril formation was determined to be around 4µg/ml at relatively low temperatures and under different experimental conditions (T=29ºC, buffer consisting of 20mM $NaHCO_3$, 117mM NaCl, 5.4mM KCl, 1.8mM $CaCl_2$, 0.81 mM $MgSO_4$, 1.03mM NaH_2PO_4, 0.01% NaN_3.) (Kadler, Hojima et al. 1987). To determine whether collagen at 150µg/ml (0.5µM) would remain soluble under our experimental conditions (T=24ºC, reaction buffer), we assess the extent of collagen aggregation as a function of time.

The aggregation of type I collagen at room temperature is followed by allowing samples to sit at 150µg/ml (0.5µM) for specified periods of time at room temperature, followed by centrifugation (at 16000g for 4min). The supernatant is then run in an SDS gel and stained using Coomassie colloidal blue. A similar procedure is followed in a prior work to assess the extent of collagen aggregation as a function of time (Kadler, Hojima et al. 1987). Figure 3.11 depicts the amount of collagen in the supernatant as a function of time. As the intensity of the collagen bands is constant, type I collagen remains soluble under our experimental conditions.

All degradation reactions are performed at room temperature. The temperature is verified using a calibrated thermometer and all recorded temperatures are 24ºC. Degradation experiments employed purified type I collagen at a concentration of 150µg/ml (0.5µM). All enzymes are incubated with 4-aminophenyl mercuric acetate, APMA, (Sigma-Aldrich Co) as previously described (Knäuper and Murphy 2001) and mixed with type I collagen to a final concentration of 25µg/ml (1.2µM) for CMMP8, 40µg/ml (2.0µM) for CMMP1, and 1.7µg/ml (30nM) for FMMP1. Reactions are stopped by the addition of SDS-Laemmli buffer (BioRad Laboratories) with β-Mercaptho Ethanol (Sigma-Aldrich Co) and boiled for 10 minutes. The degradation

products are run in 4-12% gradient gels and stained with Coomassie colloidal blue. Gelatin is generated by boiling type I collagen in reaction buffer for 10 minutes. Gelatin at 150μg/ml (3 chains X 0.5μM = 1.5μM) is then incubated with 13.8μg/ml (0.67μM) CMMP8 and 16.6μg/ml (0.83μM) CMMP1.

An unstained CMMP8 gel is transferred to a PVDF membrane, which is stained with Coomassie blue. The ¼ α1(I) and ¼ α2(I) bands are cut from the gels corresponding to the CMMP8 experiments and sent for sequencing to the Tufts University Core Facility, using an ABI 494 Protein Sequencer. The ¼ α1(I) and α2(I) bands from the CMMP1 reaction cannot be sequenced due to the low amount of collagen that is degraded by CMMP1.
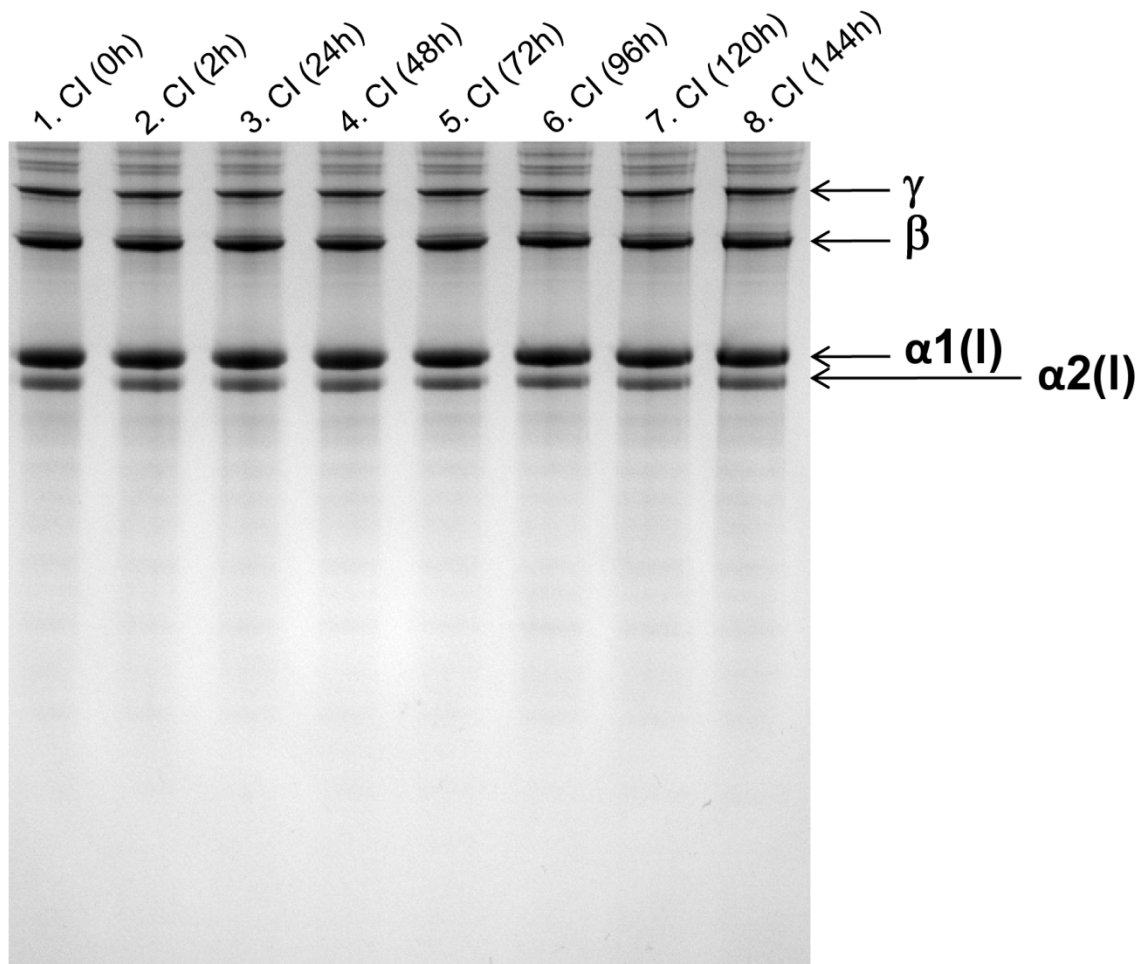
Figure 3.11. Control for the fibrillogenesis of collagen. Type I collagen (150μg/ml, 0.5μM) is allowed to sit at room temperature for up to 6 days, spun down and the supernatant run on SDS-PAGE. The intensity of the bands remains constant, suggesting that precipitation does not occur under the experimental conditions.

Conclusions

Collagen degradation is involved in a series of disorders that debilitate millions of individuals. In our quest to understand how this process works, we have encountered two scientifically challenging questions. First, how does collagen degradation occur given that the triple helix does not fit in the active site of a collagenase and the scissile bond in collagen is hidden from solvent? Second, how do collagenases cleave only at one site, leaving other potential cleavage sites untouched? We hypothesize that the answer to these questions can be found in the existence of locally unfolded states in collagen.

In order to test this hypothesis, we design a path that starts by characterizing the nature of locally unfolded states in collagen. In Chapter 1, we show that the nature of the unfolded states in type III collagen is dependent on the surrounding sequence of a given cleavage site. These results are interesting since they suggest that small sequence variations can lead to significant differences in the stability and structure of the conformational ensemble of the triple helix.

In Chapter 2, we extend the methodology from Chapter 1 to study all potential cleavage sites in type III collagen. Our results show that the locally unfolded states at the actual cleavage site are more vulnerable and flexible than at other potential cleavage sites. This suggests that active unfolding of the triple helix is not required for the existence of vulnerable states at the true cleavage site. Moreover, since not all the cleavage sites are as vulnerable, cleavage site specificity may be encoded in the conformation at each cleavage site and ultimately in the sequence surrounding each cleavage site. Type III collagen is not the only collagen type that contains more than one potential cleavage site. Future work in this direction would entail simulations of other collagen types to study if high vulnerability at the actual cleavage site is a conserved theme amongst collagen types. As a first step in this direction, other work from the Stultz lab has shown that locally unfolded states also exist at the true collagenase cleavage site in

120

type I collagen (Nerenberg and Stultz 2008). Also, it would be interesting to extend the methodology of the simulations presented here to the fibrillar state of collagen. Since this is the natural state of collagen in the body, we wish to study the influence of neighboring collagen molecules on the existence of vulnerable states. Structures of the fibrillar state of full length type I collagen have been made available (Perumal, Antipova et al. 2008). Although they are not of great resolution, they could be used as a scaffold on which to build a more detailed model of type I collagen to study the occurrence of vulnerable states in the fibrillar context. While fibrillar collagen is more stable than collagen in solution, we still expect that vulnerable states will exist in this scenario and that conformational selection will play a prominent role in the mechanism of collagenolysis.

In Chapter 3, we test whether the locally unfolded states described in Chapter 2 can be bound by the catalytic site of a collagenase. We also want to determine whether cleavage site specificity can be achieved in the absence of the hemopexin-like domain. To this end, we develop a novel rigid body docking method. Our results show that the locally unfolded states at the actual type III collagen cleavage site fit best inside of the catalytic domain of a collagenase. The docking results suggest that incubation of type III collagen with the catalytic domain of collagenases would result in degradation at a unique site. Indeed, our experimental results show that this is the case. Type III collagen in solution can be cleaved at the true cleavage site by using only the catalytic domain of MMP1 or MMP8 at room temperature. We also extend these experiments to type I collagen, the most abundant collagen type, to show that, against previous literature, locally unfolded states at the actual cleavage site can also be cleaved by the catalytic domain of a collagenase. Since the hemopexin-like domain is believed to be necessary to actively unfold the collagen triple helix, the results obtained in Chapter 3 strongly suggest that

unfolded states exist in the absence of collagenases. Moreover, cleavage site specificity is achieved in the absence of the hemopexin-like domain at room temperature. For this reason, we propose that cleavage site specificity at room temperature may be encoded in the nature of the unfolded states at each cleave site, as the results from Chapter 2 suggest. At higher temperatures, we expect vulnerable states to occur at other sites along the triple helix (Taddese, Jung et al. 2009). Therefore it is possible that the binding hemopexin-like domain in the vicinity of the true cleavage site contributes to specificity *in vivo*.

Future work related to Chapter 3 would include follow up on the structures obtained from the docking simulations. These docked structures of type III collagen are the only available snapshots of the interaction between collagen and a collagenase. These structures could serve as the starting point of more detailed molecular dynamics simulations to characterize the interaction between collagen and collagenases. Moreover, other simulations could be designed in which the hemopexin domain of MMPs could be added to the system in order to study the interaction of such domain with the triple helix.

In Appendix 1 we present a conformational selection model of collagenolysis in which collagen exists in an equilibrium between native and vulnerable states, and degradation occurs when collagenases bind and cleave the vulnerable states. In Appendix 2 we explain the degradation results of type I collagen by CMMP1 and CMMP8 from Chapter 3 using the two-state degradation model presented in Appendix 1. This type of analysis allows us to characterize the equilibrium between native (well folded) and vulnerable (locally unfolded) conformers in type I collagen in solution at room temperature. It also points to the binding affinity to the vulnerable state as the cause behind the marked difference in efficiency of CMMP1 and CMMP8 at cleaving type I collagen.

However, we recognize that collagen is typically degraded by full length collagenases in the body. For this reason, here we present a model for collagenolysis by full length MMP based in the existence of an equilibrium between native and vulnerable states in collagen. We expand the reaction scheme for the catalytic domain alone (Appendix 1 and 2) to consider the effects of the hemopexin-like domain on collagenolysis *in vitro* (Figure C.1). In this reaction scheme, collagen exists in an equilibrium between native (*N*) and vulnerable (*V*) states, and either state can be bound by collagenases via the hemopexin-like domain, which contains binding sites for collagen (Murphy, Allan et al. 1992; Overall 2002; Tam, Moore et al. 2004; Perumal, Antipova et al. 2008; Lauer-Fields, Chalmers et al. 2009). When the native state is bound, an *N·H* complex is formed that cannot be cleaved since the scissile bond in this structure is not accessible to the MMP active site. Since the catalytic domain of the enzyme is not bound to the scissile bond in the *N·H* complex, and the putative binding site for the hemopexin-like domain in type I collagen is removed from the cleavage site (Perumal, Antipova et al. 2008; Erat, Slatter et al. 2009), the *N·H* complex can transition to a vulnerable state (*V·H* complex) via a conformational change similar to the one experienced by unbound collagen. Once the *V·H* complex is formed, the catalytic domain can then bind the accessible scissile bond, yielding the *V·F* complex, which then goes on to form degraded protein.
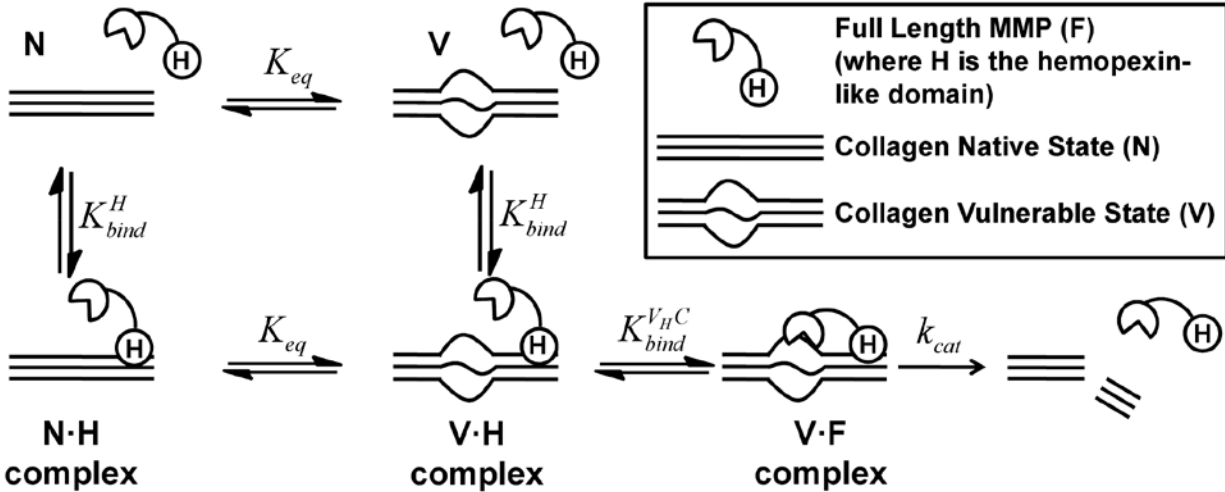
Figure C.1. A conformational selection mechanism for collagenolysis with full length collagenases. Collagen exists in an equilibrium between native (N) and vulnerable (V) states determined by the equilibrium constant K_{eq}. Full length MMP (F) can bind to the native state of collagen via the hemopexin-like domain (H) with binding constant K_{bind}^{H}, forming the $N{\cdot}H$ complex. The $N{\cdot}H$ complex can transition to the $V{\cdot}H$ complex via the equilibrium between native and vulnerable states, K_{eq}. The full length enzyme, F, can bind directly to the vulnerable state, V, via the hemopexin domain (H) forming the $V{\cdot}H$ complex, again determined by the binding constant K_{bind}^{H}. Once a $V{\cdot}H$ complex is formed, the catalytic domain of the full length MMP can bind to the vulnerable state with a binding constant $K_{bind}^{V_{H}C}$, forming the $V{\cdot}F$ complex. The $V{\cdot}F$ complex is then degraded with catalytic rate k_{cat}.

Chapter 3 and Appendix 2 analyze the degradation of type I collagen in solution by means of both experimental and numerical studies, but *in vivo*, collagen typically exists in a fibrillar state (Gelse, Poschl et al. 2003). Since previous studies suggest that type I collagen is thermally stabilized in a fibrillar context relative to the solution state, it is an open question as to whether the conformational ensemble of the collagenase cleavage site is the same in the fibril as it is in solution (Birkedal-Hansen, Taylor et al. 1985). Interestingly, however, recent fiber diffraction studies suggest that the α2(I) chain is relatively dissociated from the central axis of the helix at the collagenase cleavage site *in situ*; i.e. in the fibrillar state, the triple helix is disrupted near the cleavage site (Perumal, Antipova et al. 2008). While the fiber diffraction study is of insufficient resolution to provide reliable details on the atomic structure of the fibrillar state,

these data do suggest that the structure of the triple helix can vary considerably along its length and especially near the collagenase cleavage site. Thus, while *in vivo* collagen degradation clearly differs from the solution state degradation studies presented in this work, the mechanism outlined in Figure C.1 and the preceding discussion may still be a fair representation of the important aspects of *in vivo* collagen degradation, albeit the value of K_{eq} is likely much smaller in the fibrillar context.

To understand how the presence of enzyme might affect the relative concentration of vulnerable states, we consider an equilibrium in which the enzyme is catalytically inactive (i.e., the degradation step does not occur), and solve for the effective ratio of all vulnerable states to all native states, $K_{eq}^{eff} = ([V] + [V \cdot H] + [V \cdot F])/([N] + [N \cdot H])$. In this scenario, there is a closed form solution for K_{eq}^{eff} (see Methods in Appendix 2):

$$K_{eq}^{eff} = K_{eq} \left(1 + \left(\frac{K_{bind}^{H} [E]}{K_{bind}^{H} [E] + 1} \right) K_{bind}^{V_H C} \right)$$

where $[E]$ is the concentration of free enzyme, which itself is a function of the enzyme-to-substrate ratio, and the binding constants K_{bind}^{H} and $K_{bind}^{V_H C}$. We can better understand the behavior of K_{eq}^{eff} by evaluating it at two limiting conditions. In the first, no enzyme is present ($[E] = [E]_{\text{tot}}$ = 0) and/or the enzyme does not bind collagen ($K_{bind}^{H} = 0$); in either case $K_{eq}^{eff} = K_{eq}$. The second limiting condition is when the concentration of enzyme is in relative excess and K_{bind}^{H} is non-zero; in this case, $K_{eq}^{eff} \approx K_{eq} K_{bind}^{V_H C}$. Given that the catalytic domain is restricted to reside in the vicinity of the cleavage site when the hemopexin-like domain is bound, the entropic loss upon binding of the catalytic domain is relatively small when the enzyme is anchored to the protein by the hemopexin-like domain. Consequently, we expect that $K_{bind}^{V_H C} \gg 1$ and therefore

that $K_{eq} \ll K_{eq} K_{bind}^{V_H C}$. Accordingly, this suggests that as long as there is a non-zero total concentration of enzyme, we will have $K_{eq} < K_{eq}^{eff} < K_{eq} K_{bind}^{V_H C}$, as demonstrated in Figure C.2. In other words, the ratio of vulnerable states to native states, K_{eq}^{eff}, will always be greater than K_{eq} when enzyme is present simply due to the principle of mass action. In this way the presence of enzyme leads to an effective increase in the concentration of vulnerable, "unwound", conformers without requiring energy input or active enzyme-mediated unwinding.
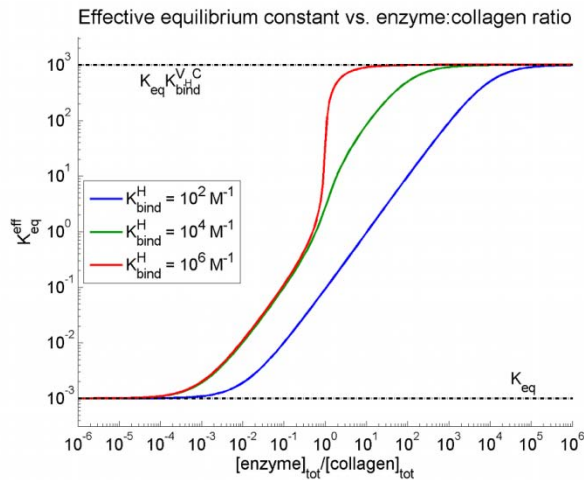
Figure C.2. Change in the effective conformational equilibrium of collagen in the presence of full length collagenase. Computed values for the effective equilibrium constant, K_{eq}^{eff}, as a function of total enzyme-to-collagen ratio for three possible values of the hemopexin-like domain binding constant, K_{bind}^H, ranging from sub-millimolar to micromolar affinity. For these calculations, $[\text{collagen}]_{tot} = 10^{-6}$ M, $K_{eq} = 10^{-3}$, and $K_{bind}^{V_H C} = 10^6$. At low enzyme-to-collagen ratios, K_{eq}^{eff} is approximately equal to K_{eq}, whereas at high enzyme-to-collagen ratios, K_{eq}^{eff} approaches $K_{eq} K_{bind}^{V_H C}$.

This analysis of the collagenolysis model by full length collagenases reconciles the concept of "active unfolding" with the natural occurrence of vulnerable states. Since the hemopexin-like domain preferentially binds in the vicinity of the cleavage site (Perumal, Antipova et al. 2008), the catalytic domain is ideally oriented to bind the rarely occurring vulnerable states. When such binding happens, the vulnerable states are stabilized in comparison

126

to when they occur freely in solution and no enzyme is around to bind them. Without the localization effect due to the binding of the hemopexin-like domain, the probability of productive encounter between a vulnerable state and a catalytic domain would be very low. Indeed, this is reflected in the very high concentrations of CMMP1 and CMMP8 that are required to observe type I collagen degradation, as opposed to the lower concentration of full length collagenases needed to observe the same result (Chapter 3 and Chung, Dinakarpandian et al. 2004). In summary, this model allows us to redefine the role of the hemopexin-like domain of collagenases. This domain was originally thought to be necessary for collagenolysis since it would play a key role in the active unfolding of the triple helix (Overall 2002; Chung, Dinakarpandian et al. 2004). However, the work presented in this thesis allows explaining collagenolysis without having to resort to active unwinding. Therefore we propose that the presence of the hemopexin-like domain makes collagenases much more efficient enzymes, as the probability of binding vulnerable states increases. Lastly, the hemopexin-like domain may be also involved in determining cleavage site specificity *in vivo*.

More importantly, the new degradation model based on the existence of an equilibrium between native and vulnerable states allows us to propose new avenues of regulating collagen degradation, which is the long term goal in the field. Historically, methods designed to prevent excessive collagen degradation in disorders of collagen metabolism have mainly focused on designing small molecule inhibitors of the MMP active site that would prevent binding and hydrolysis of the scissile bond (Folgueras, Pendas et al. 2004; Georgiadis and Yiotakis 2008). This corresponds to disrupting the reaction step involving the transition between the $V{\cdot}H$ and $V{\cdot}F$ complexes in Figure C.1. Although great effort has been directed towards the design of such inhibitors, poor clinical results have been obtained. Since the catalytic domains of many

collagenases are quite similar, selectivity issues and secondary side effects in clinical trials have been encountered and therefore only few inhibitors have been made available commercially (Abbenante and Fairlie 2005; Ganea, Trifan et al. 2007; Georgiadis and Yiotakis 2008; Tallant, Marrero et al. 2009). The reaction scheme in Figure C.1 highlights additional and potentially fruitful avenues for abrogating disease-associated collagenolysis. As others have noted, disruption of hemopexin-like domain binding to secondary sites on collagen could prove to be an effective mechanism to modulate collagenolysis (e.g., decreasing K_{bind}^H) (Folgueras, Pendas et al. 2004; Nerenberg, Salsas-Escat et al. 2007; Lauer-Fields, Chalmers et al. 2009). Our data suggest that mechanisms that lead to stabilization of the native state relative to the vulnerable state (i.e., decreasing K_{eq}) may have a similar desirable effect. One possible avenue of research is the study of small molecules that have the ability to bind to the native state of collagen at its actual cleavage site, preventing unfolding and reducing the number of vulnerable conformers. By utilizing these additional steps in the reaction mechanism, collagen degradation could potentially be inhibited in a specific and efficient manner.

To conclude, in this thesis we have used a combination of experimental and computational tools to study a biophysical problem of great complexity. By using only one of these approaches in isolation it would have probably been very difficult to achieve the level of understanding obtained in this thesis. The details provided by the molecular dynamics simulations would have been hard to observe using any other experimental technique alone. Since MD simulations rely on the accuracy of the underlying potential energy function, which is approximate, it is important to verify the observations arising from these simulations with experiment. Our collagen degradation studies confirmed the predictions from the MD simulations. Then, taking a step back to the computer world, numerical simulations of the

experimental data allow us to characterize the interaction between collagen and collagenases, as well as the conformational equilibrium at the collagen cleavage site. The synergy between simulation and experiment is one of the core values of the Computational and Systems Biology program. Here, applied to a biophysical problem, it proved a valuable approach to contribute to the current understanding of the mechanism of collagenolysis.

Bibliography

Abbenante, G. and D. P. Fairlie (2005). "Protease inhibitors in the clinic." <u>Med. Chem.</u> 1(1): 71-104.

Alder, B. and T. Wainwright (1959). "Studies in molecular dynamics. I. General Method." <u>J Chem Phys</u> 31(2): 459-466.

Bahar, I. and A. J. Rader (2005). "Coarse-grained normal mode analysis in structural biology." <u>Curr Opin Struct Biol</u> 15(5): 586-92.

Baker, L. E. (1951). "New synthetic substrates for pepsin." <u>J Biol Chem</u> 193(2): 809-19.

Barnes, M. J. and R. W. Farndale (1999). "Collagens and atherosclerosis." <u>Exp Gerontol</u> 34(4): 513-25.

Berman, H. M., J. Westbrook, et al. (2000). "The Protein Data Bank." <u>Nucleic Acids Res</u> 28(1): 235-42.

Bertini, I., V. Calderone, et al. (2006). "Snapshots of the reaction mechanism of matrix metalloproteinases." <u>Angew Chem Int Ed Engl</u> 45(47): 7952-5.

Bhatnagar, R. and C. Cough (1996). Circular dichroism of collagen and related peptides *Circular Dichroism and the Conformational Analysis of Biomolecules*. G. Fasman. New York, Plenum Press: pp.183-199.

Birkedal-Hansen, H., R. E. Taylor, et al. (1985). "Cleavage of bovine skin type III collagen by proteolytic enzymes. Relative resistance of the fibrillar form." <u>J Biol Chem</u> 260(30): 16411-7.

Boczko, E. M. and C. L. Brooks, 3rd (1995). "First-principles calculation of the folding free energy of a three-helix bundle protein." <u>Science</u> 269(5222): 393-6.

Bode, W. and K. Maskos (2003). "Structural basis of the matrix metalloproteinases and their physiological inhibitors, the tissue inhibitors of metalloproteinases." <u>Biol Chem</u> 384(6): 863-72.

Boot-Handford, R. P. and D. S. Tuckwell (2003). "Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest." <u>Bioessays</u> 25(2): 142-51.

Brandstetter, H., F. Grams, et al. (2001). "The 1.8-A crystal structure of a matrix metalloproteinase 8-barbiturate inhibitor complex reveals a previously unobserved mechanism for collagenase substrate recognition." <u>J Biol Chem</u> 276(20): 17405-12.

Brinckmann, J. (2005). "Collagens at a glance." <u>Top Curr Chem</u> 247: 1-6.

Brodsky, B. and A. V. Persikov (2005). "Molecular structure of the collagen triple helix." <u>Adv Protein Chem</u> 70: 301-39.

Brodsky, B. and J. A. Ramshaw (1997). "The collagen triple-helix structure." <u>Matrix Biol</u> 15(8-9): 545-54.

Brooks, C. L., 3rd (1998). "Simulations of protein folding and unfolding." <u>Curr Opin Struct Biol</u> 8(2): 222-6.

Brooks III, B., R. Bruccoleri, et al. (1983). "Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations." <u>J. Comput. Chem.</u> 4: 187–217.

Brooks III, C. L., A. Brunger, et al. (1985). "Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach." <u>Biopolymers</u> 24(5): 843-65.

Brooks III, C. L. and M. Karplus (1983). "Deformable stochastic boundaries in molecular dynamics." <u>J. Chem. Phys.</u> 79(12): 6312-6325.

Brooks III, C. L., M. Karplus, et al. (1988). <u>Proteins: a theoretical perspective of dynamics, structure, and thermodynamics</u>. New York, NY, J. Wiley.

Browner, M. F., W. W. Smith, et al. (1995). "Matrilysin-inhibitor complexes: common themes among metalloproteases." <u>Biochemistry</u> 34(20): 6602-10.

Brunger, A. T. and M. Karplus (1988). "Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison." <u>Proteins</u> 4(2): 148-56.

Celentano, D. C. and W. H. Frishman (1997). "Matrix metalloproteinases and coronary artery disease: a novel therapeutic target." <u>J Clin Pharmacol</u> 37(11): 991-1000.

Chakraborti, S., M. Mandal, et al. (2003). "Regulation of matrix metalloproteinases: an overview." <u>Mol Cell Biochem</u> 253(1-2): 269-85.

Cheah, K. S., N. G. Stoker, et al. (1985). "Identification and characterization of the human type II collagen gene (COL2A1)." <u>Proc Natl Acad Sci U S A</u> 82(9): 2555-9.

Chung, L., D. Dinakarpandian, et al. (2004). "Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis." <u>Embo J</u> 23(15): 3020-30.

Clark, I. M. and T. E. Cawston (1989). "Fragments of human fibroblast collagenase. Purification and characterization." <u>Biochem J</u> 263(1): 201-6.

Courts, A. (1955). "The N-terminal amino acid residues of gelatin. 3. Enzymic degradation." <u>Biochem. J.</u> 59(3): 382-6.

D'Aquino, J. A., J. Gomez, et al. (1996). "The magnitude of the backbone conformational entropy change in protein folding." <u>Proteins</u> 25(2): 143-56.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. Palo Alto, CA, DeLano Scientific.

Dobbins, S. E., V. I. Lesk, et al. (2008). "Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking." Proc Natl Acad Sci U S A 105(30): 10390-5.

Dunker, A. K., I. Silman, et al. (2008). "Function and structure of inherently disordered proteins." Curr Opin Struct Biol 18(6): 756-64.

Dyson, H. J. and P. E. Wright (2005). "Intrinsically unstructured proteins and their functions." Nat Rev Mol Cell Biol 6(3): 197-208.

Engel, J. and H. P. Bachinger (2000). "Cooperative equilibrium transitions coupled with a slow annealing step explain the sharpness and hysteresis of collagen folding." Matrix Biol 19(3): 235-44.

Epstein, E. H., Jr. (1974). "(Alpha1(3))3 human skin collagen. Release by pepsin digestion and preponderance in fetal life." J. Biol. Chem. 249(10): 3225-31.

Erat, M. C., D. A. Slatter, et al. (2009). "Identification and structural analysis of type I collagen sites in complex with fibronectin fragments." Proc Natl Acad Sci U S A 106(11): 4195-200.

Fan, P., M. H. Li, et al. (1993). "Backbone dynamics of (Pro-Hyp-Gly)10 and a designed collagen-like triple-helical peptide by 15N NMR relaxation and hydrogen-exchange measurements." Biochemistry 32(48): 13299-309.

Fields, G. B. (1991). "A model for interstitial collagen catabolism by mammalian collagenases." J Theor Biol 153(4): 585-602.

Fietzek, P. P. and K. Kuhn (1975). "The covalent structure of collagen: amino-acid sequence of the cyanogen-bromide peptides alpha1-CB2, alpha1-CB4 and alpha1-CB5 from calf-skin collagen." Eur J Biochem 52(1): 77-82.

Fietzek, P. P., F. W. Rexrodt, et al. (1973). "The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 from calf-skin collagen." Eur J Biochem 38(2): 396-400.

Fietzek, P. P., F. W. Rexrodt, et al. (1972). "The covalent structure of collagen. Amino-acid sequence of peptide 1-CB6-C2." Eur J Biochem 30(1): 163-8.

Fietzek, P. P., P. Wendt, et al. (1972). "The covalent structure of collagen: amino acid sequence of 1-CB3 from calf skin collagen." FEBS Lett 26(1): 74-6.

Fiori, S., B. Sacca, et al. (2002). "Structural properties of a collagenous heterotrimer that mimics the collagenase cleavage site of collagen type I." J Mol Biol 319(5): 1235-42.

Folgueras, A. R., A. M. Pendas, et al. (2004). "Matrix metalloproteinases in cancer: from new functions to improved inhibition strategies." Int. J. Dev. Biol. 48(5-6): 411-24.

French, M. F., K. A. Mookhtiar, et al. (1987). "Limited proteolysis of type I collagen at hyperreactive sites by class I and II Clostridium histolyticum collagenases: complementary digestion patterns." Biochemistry. 26(3): 681-7.

Fruton, J. S. and M. Bergmann (1938). "The Specificity of Pepsin Action." Science 87(2268): 557.

Gajko-Galicka, A. (2002). "Mutations in type I collagen genes resulting in osteogenesis imperfecta in humans." Acta Biochim Pol 49(2): 433-41.

Ganea, E., M. Trifan, et al. (2007). "Matrix metalloproteinases: useful and deleterious." Biochem. Soc. Trans. 35(Pt 4): 689-91.

Gelse, K., E. Poschl, et al. (2003). "Collagens--structure, function, and biosynthesis." Adv Drug Deliv Rev 55(12): 1531-46.

Gendron, R., D. Grenier, et al. (1999). "Inhibition of the activities of matrix metalloproteinases 2, 8, and 9 by chlorhexidine." Clin. Diagn. Lab. Immunol. 6(3): 437-9.

Georgiadis, D. and A. Yiotakis (2008). "Specific targeting of metzincin family members with small-molecule inhibitors: progress toward a multifarious challenge." Bioorg. Med. Chem. 16(19): 8781-94.

Germain, D. P. (2007). "Ehlers-Danlos syndrome type IV." Orphanet J Rare Dis 2: 32.

Gioia, M., S. Monaco, et al. (2007). "Characterization of the mechanisms by which gelatinase A, neutrophil collagenase, and membrane-type metalloproteinase MMP-14 recognize collagen I and enzymatically process the two alpha-chains." J. Mol. Biol. 368(4): 1101-13.

Grams, F., P. Reinemer, et al. (1995). "X-ray structures of human neutrophil collagenase complexed with peptide hydroxamate and peptide thiol inhibitors. Implications for substrate binding and rational drug design." Eur J Biochem 228(3): 830-41.

Hansson, T., C. Oostenbrink, et al. (2002). "Molecular dynamics simulations." Curr Opin Struct Biol 12(2): 190-6.

Haralson, M. A. and J. R. Hassell (1995). Extracellular matrix : a practical approach. Oxford New York, IRL Press ;
Oxford University Press.

Hirose, T., C. Patterson, et al. (1993). "Structure-function relationship of human neutrophil collagenase: identification of regions responsible for substrate specificity and general proteinase activity." Proc Natl Acad Sci U S A 90(7): 2569-73.

Hirschfelder, J., H. Eyring, et al. (1936). "Reactions involving hydrogen molecules and atoms." J Chem Phys 4: 170-177.

Hohenester, E. and J. Engel (2002). "Domain structure and organisation in extracellular matrix proteins." Matrix Biol 21(2): 115-28.

Huang, A. and C. M. Stultz (2007). "Conformational sampling with implicit solvent models: application to the PHF6 peptide in tau protein." Biophys J 92(1): 34-45.

Huang, A. and C. M. Stultz (2008). "The effect of a DeltaK280 mutation on the unfolded state of a microtubule-binding repeat in Tau." PLoS Comput Biol 4(8): e1000155.

Huang, A. and C. M. Stultz (2009). "Finding Order within Disorder: Elucidating the Structure of Proteins Associated with Neurodegenerative Disease." Future Medicinal Chemistry 1(3): 467-482.

Hubbard, S. J. (1998). "The structural aspects of limited proteolysis of native proteins." Biochim Biophys Acta 1382(2): 191-206.

Hudaky, P., G. Kaslik, et al. (1999). "The differential specificity of chymotrypsin A and B is determined by amino acid 226." Eur J Biochem 259(1-2): 528-33.

Hulmes, D. J. (2002). "Building collagen molecules, fibrils, and suprafibrillar structures." J Struct Biol 137(1-2): 2-10.

Humphrey, W., A. Dalke, et al. (1996). "VMD: visual molecular dynamics." J Mol Graph 14(1): 33-8, 27-8.

Jones, C. B., D. C. Sane, et al. (2003). "Matrix metalloproteinases: a review of their structure and role in acute coronary syndrome." Cardiovasc Res 59(4): 812-23.

Kadler, K. E., Y. Hojima, et al. (1987). "Assembly of collagen fibrils de novo by cleavage of the type I pC-collagen with procollagen C-proteinase. Assay of critical concentration demonstrates that collagen self-assembly is a classical example of an entropy-driven process." J. Biol. Chem. 262(32): 15696-701.

Kadler, K. E., Y. Hojima, et al. (1988). "Assembly of type I collagen fibrils de novo. Between 37 and 41 degrees C the process is limited by micro-unfolding of monomers." J Biol Chem 263(21): 10517-23.

Karplus, M. and J. Kuriyan (2005). "Molecular dynamics and protein function." Proc Natl Acad Sci U S A 102(19): 6679-85.

Karplus, M. and J. A. McCammon (2002). "Molecular dynamics simulations of biomolecules." Nat Struct Biol 9(9): 646-52.

Kelly, S. M., T. J. Jess, et al. (2005). "How to study proteins by circular dichroism." <u>Biochim Biophys Acta</u> 1751(2): 119-39.

Knäuper, V. and G. Murphy (2001). Methods for Studying Activation of Matrix Metalloproteinases *Matrix metalloproteinase protocols*. I. M. Clark. Totowa, N.J., Humana Press: pp.377-387.

Koide, T. and K. Nagata (2005). "Collagen biosynthesis." <u>Top Curr Chem</u> 247: 85-114.

Kramer, R. Z., J. Bella, et al. (2001). "The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence." <u>J Mol Biol</u> 311(1): 131-47.

Kramer, R. Z., J. Bella, et al. (1999). "Sequence dependent conformational variations of collagen triple-helical structure." <u>Nat Struct Biol</u> 6(5): 454-7.

Kuivaniemi, H., G. Tromp, et al. (1988). "Structure of a full-length cDNA clone for the prepro alpha 2(I) chain of human type I procollagen. Comparison with the chicken gene confirms unusual patterns of gene conservation." <u>Biochem J</u> 252(3): 633-40.

Kumar, S., J. Rosenberg, et al. (1992). "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method." <u>J Comp Chem</u> 13(8): 1011-1021.

Kumar, S., J. Rosenberg, et al. (1995). "Multidimensional free-energy calculations using the weighted histogram analysis method." <u>J Comp Chem</u> 16(11): 1339-1350.

Lang, R., M. Braun, et al. (2004). "Crystal structure of the catalytic domain of MMP-16/MT3-MMP: characterization of MT-MMP specific features." <u>J Mol Biol</u> 336(1): 213-25.

Lange, O. F., N. A. Lakomek, et al. (2008). "Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution." <u>Science</u> 320(5882): 1471-5.

Lauer-Fields, J. L., M. J. Chalmers, et al. (2009). "Identification of specific hemopexin-like domain residues that facilitate matrix metalloproteinase collagenolytic activity." <u>J. Biol. Chem.</u> 284(36): 24017-24.

Lauer-Fields, J. L., D. Juska, et al. (2002). "Matrix metalloproteinases and collagen catabolism." <u>Biopolymers</u> 66(1): 19-32.

Lazaridis, T. and M. Karplus (1999). "Effective energy function for proteins in solution." <u>Proteins</u> 35(2): 133-52.

Lee, B. and F. M. Richards (1971). "The interpretation of protein structures: estimation of static accessibility." <u>J Mol Biol</u> 55(3): 379-400.

Leikina, E., M. V. Mertts, et al. (2002). "Type I collagen is thermally unstable at body temperature." Proc Natl Acad Sci U S A 99(3): 1314-8.

Levitt, M. and A. Warshel (1975). "Computer simulation of protein folding." Nature 253(5494): 694-8.

Li, J., P. Brick, et al. (1995). "Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed beta-propeller." Structure 3(6): 541-9.

Li, M. H., P. Fan, et al. (1993). "Two-dimensional NMR assignments and conformation of (Pro-Hyp-Gly)10 and a designed collagen triple-helical peptide." Biochemistry 32(29): 7377-87.

Lloyd-Jones, D., R. J. Adams, et al. (2009). "Heart Disease and Stroke Statistics--2010 Update. A Report From the American Heart Association." Circulation.

Makareeva, E., E. L. Mertz, et al. (2008). "Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta." J Biol Chem 283(8): 4787-98.

Martin, E. and J. R. Shapiro (2007). "Osteogenesis imperfecta:epidemiology and pathophysiology." Curr Osteoporos Rep 5(3): 91-7.

Maskos, K. (2005). "Crystal structures of MMPs in complex with physiological and pharmacological inhibitors." Biochimie 87(3-4): 249-63.

McCammon, J. A., B. R. Gelin, et al. (1977). "Dynamics of folded proteins." Nature 267(5612): 585-90.

McDonnell, S., M. Morgan, et al. (1999). "Role of matrix metalloproteinases in normal and disease processes." Biochem Soc Trans 27(4): 734-40.

McQuarrie, D. A. (2000). Statistical mechanics. Sausalito, Calif., University Science Books.

Miller, E. J., J. E. Finch, Jr., et al. (1976). "Specific cleavage of the native type III collagen molecule with trypsin. Similarity of the cleavage products to collagenase-produced fragments and primary structure at the cleavage site." Arch Biochem Biophys 173(2): 631-7.

Miller, E. J. and R. K. Rhodes (1982). Preparation and characterization of collagens *Methods in enzymology Vol 82, Structural and contractile proteins, Part A Extracellular Matrix*. L. W. Cunningham and D. W. Fredericksen. New York, Academic Press. 82: pp.41-44.

Mishra, G. R., M. Suresh, et al. (2006). "Human protein reference database--2006 update." Nucleic Acids Res 34(Database issue): D411-4.

Mizuno, K., T. Hayashi, et al. (2004). "Hydroxylation-induced stabilization of the collagen triple helix. Acetyl-(glycyl-4(R)-hydroxyprolyl-4(R)-hydroxyprolyl)(10)-NH(2) forms a highly stable triple helix." J Biol Chem 279(36): 38072-8.

Mohs, A., T. Silva, et al. (2007). "Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline." <u>J Biol Chem</u> 282(41): 29757-65.

Morimoto, K., K. Kawabata, et al. (2009). "Characterization of type I collagen fibril formation using thioflavin T fluorescent dye." <u>J. Biochem.</u> 145(5): 677-684.

Murphy, G., J. A. Allan, et al. (1992). "The role of the C-terminal domain in collagenase and stromelysin specificity." <u>J. Biol. Chem.</u> 267(14): 9612-8.

Murphy, G. and V. Knauper (1997). "Relating matrix metalloproteinase structure to function: why the "hemopexin" domain?" <u>Matrix Biol</u> 15(8-9): 511-8.

Nelson, A. R., B. Fingleton, et al. (2000). "Matrix metalloproteinases: biologic activity and clinical implications." <u>J Clin Oncol</u> 18(5): 1135-49.

Nelson, D. and M. Cox (2005). <u>Lehninger's Principles of Biochemistry</u>. New York, NY, W.H. Freeman and Company.

Nerenberg, P. S., R. Salsas-Escat, et al. (2007). "Collagen--a necessary accomplice in the metastatic process." <u>Cancer Genomics Proteomics</u> 4(5): 319-28.

Nerenberg, P. S., R. Salsas-Escat, et al. (2008). "Do collagenases unwind triple-helical collagen before peptide bond hydrolysis? Reinterpreting experimental observations with mathematical models." <u>Proteins</u> 70(4): 1154-61.

Nerenberg, P. S. and C. M. Stultz (2008). "Differential unfolding of alpha1 and alpha2 chains in type I collagen and collagenolysis." <u>J Mol Biol</u> 382(1): 246-56.

Olsen, J. V., S. E. Ong, et al. (2004). "Trypsin cleaves exclusively C-terminal to arginine and lysine residues." <u>Mol Cell Proteomics</u> 3(6): 608-14.

Orgel, J. P., T. C. Irving, et al. (2006). "Microfibrillar structure of type I collagen in situ." <u>Proc Natl Acad Sci U S A</u> 103(24): 9001-5.

Orgel, J. P., A. Miller, et al. (2001). "The in situ supermolecular structure of type I collagen." <u>Structure</u> 9(11): 1061-9.

Orgel, J. P., T. J. Wess, et al. (2000). "The in situ conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen." <u>Structure</u> 8(2): 137-42.

Overall, C. M. (2002). "Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites." <u>Mol Biotechnol</u> 22(1): 51-86.

Parapia, L. A. and C. Jackson (2008). "Ehlers-Danlos syndrome--a historical review." Br J Haematol 141(1): 32-5.

Pei, D. and S. J. Weiss (1995). "Furin-dependent intracellular activation of the human stromelysin-3 zymogen." Nature 375(6528): 244-7.

Pelmenschikov, V. and P. E. Siegbahn (2002). "Catalytic mechanism of matrix metalloproteinases: two-layered ONIOM study." Inorg Chem 41(22): 5659-66.

Peri, S., J. D. Navarro, et al. (2003). "Development of human protein reference database as an initial platform for approaching systems biology in humans." Genome Res 13(10): 2363-71.

Persikov, A. V., J. A. Ramshaw, et al. (2000). "Collagen model peptides: Sequence dependence of triple-helix stability." Biopolymers 55(6): 436-50.

Persikov, A. V., J. A. Ramshaw, et al. (2005). "Prediction of collagen stability from amino acid sequence." J Biol Chem 280(19): 19343-9.

Persikov, A. V., J. A. Ramshaw, et al. (2000). "Amino acid propensities for the collagen triple-helix." Biochemistry 39(48): 14960-7.

Perumal, S., O. Antipova, et al. (2008). "Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis." Proc Natl Acad Sci U S A 105(8): 2824-9.

Privalov, P. L. (1982). "Stability of proteins. Proteins which do not present a single cooperative system." Adv Protein Chem 35: 1-104.

Privalov, P. L. and A. I. Dragan (2007). "Microcalorimetry of biological macromolecules." Biophys Chem 126(1-3): 16-24.

Rahman, A. (1964). "Correlation in the motion of atoms in liquid argon." Phys Rev 136(2A): A405-A411.

Rainey, J. K. and M. C. Goh (2004). "An interactive triple-helical collagen builder." Bioinformatics 20(15): 2458-9.

Ramachandran, G. N. and G. Kartha (1955). "Structure of collagen." Nature 176(4482): 593-5.

Rauterberg, J., P. Fietzek, et al. (1972). "The amino acid sequence of the carboxyterminal nonhelical cross link region of the alpha 1 chain of calf skin collagen." FEBS Lett 21(1): 75-79.

Ricard-Blum, S., F. Ruggiero, et al. (2005). "The collagen superfarmily." Top Curr Chem 247: 35-84.

Rich, A. and F. H. Crick (1955). "The structure of collagen." Nature 176(4489): 915-6.

Rich, A. and F. H. Crick (1961). "The molecular structure of collagen." J Mol Biol 3: 483-506.

Sacca, B., S. Fiori, et al. (2003). "Studies of the local conformational properties of the cell-adhesion domain of collagen type IV in synthetic heterotrimeric peptides." Biochemistry 42(12): 3429-36.

Salsas-Escat, R. and C. M. Stultz (2009). "The Molecular Mechanics of Collagen Degradation: Implications for Human Disease." Experimental Mechanics 49(1): 65-77.

Schlick, T. (2000). Molecular modeling and simulation. An interdisciplinary guide. New York, NY, Springer.

Schnierer, S., T. Kleine, et al. (1993). "The recombinant catalytic domain of human neutrophil collagenase lacks type I collagen substrate specificity." Biochem. Biophys. Res. Commun. 191(2): 319-26.

Sellers, J. R. and C. Veigel (2006). "Walking with myosin V." Curr Opin Cell Biol 18(1): 68-73.

Seyer, J. M. and A. H. Kang (1981). "Covalent structure of collagen: amino acid sequence of alpha 1(III)-CB9 from type III collagen of human liver." Biochemistry 20(9): 2621-7.

Shigemura, Y., M. Ando, et al. (2004). "Possible degradation of type I collagen in relation to yellowtail muscle softening during chilled storage." Fisheries Science. 70(4): 703-709.

Shirai, T., S. Hattori, et al. (1998). "The complete cDNA coding sequence for the bovine proalpha2(I) chain of type I procollagen." Matrix Biol 17(1): 85-8.

Shoemaker, B. A., J. J. Portman, et al. (2000). "Speeding molecular recognition by using the folding funnel: the fly-casting mechanism." Proc Natl Acad Sci U S A 97(16): 8868-73.

Shoulders, M. D. and R. T. Raines (2009). "Collagen structure and stability." Annu Rev Biochem 78: 929-58.

Song, F., K. Wisithphrom, et al. (2006). "Matrix metalloproteinase dependent and independent collagen degradation." Front Biosci 11: 3100-20.

Stultz, C. M. (2002). "Localized unfolding of collagen explains collagenase cleavage near imino-poor sites." J Mol Biol 319(5): 997-1003.

Su, M. W., B. Lee, et al. (1989). "Nucleotide sequence of the full length cDNA encoding for human type II procollagen." Nucleic Acids Res 17(22): 9473.

Sugase, K., H. J. Dyson, et al. (2007). "Mechanism of coupled folding and binding of an intrinsically disordered protein." Nature 447(7147): 1021-5.

Taddese, S., M. C. Jung, et al. (2009). "MMP-12 catalytic domain recognizes and cleaves at multiple sites in human skin collagen type I and type III." Biochim. Biophys. Acta. 1803: 20-28.

Tallant, C., A. Marrero, et al. (2009). "Matrix metalloproteinases: Fold and function of their catalytic domains." Biochim Biophys Acta.

Tam, E. M., T. R. Moore, et al. (2004). "Characterization of the distinct collagen binding, helicase and cleavage mechanisms of matrix metalloproteinase 2 and 14 (gelatinase A and MT1-MMP): the differential roles of the MMP hemopexin c domains and the MMP-2 fibronectin type II modules in collagen triple helicase activities." J. Biol. Chem. 279(41): 43336-44.

Taylor, J. P., J. Hardy, et al. (2002). "Biomedicine - Toxic proteins in neurodegenerative disease." Science 296(5575): 1991-1995.

Tran, H. T., A. Mao, et al. (2008). "Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins." J Am Chem Soc 130(23): 7380-92.

Tran, H. T., X. Wang, et al. (2005). "Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins." Biochemistry 44(34): 11369-80.

Trelstad, R. L., V. M. Catanese, et al. (1976). "Collagen fractionation: separation of native types I, II and III by differential precipitation." Anal. Biochem. 71(1): 114-8.

Tromp, G., H. Kuivaniemi, et al. (1988). "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain of human type I procollagen." Biochem J 253(3): 919-22.

Tsai, C. J., B. Ma, et al. (2001). "Structured disorder and conformational selection." Proteins 44(4): 418-27.

Veis, A. and J. Anesey (1965). "Modes of intermolecular cross-linking in mature insoluble collagen." J. Biol. Chem. 240(10): 3899-3908.

Vendruscolo, M. (2007). "Determination of conformationally heterogeneous states of proteins." Curr Opin Struct Biol 17(1): 15-20.

Venyaminov, S. and J. Yang (1996). Determination of Protein Secondary Structure *Circular Dichroism and the Conformational Analysis of Biomolecules*. G. Fasman. New York, Plenum Press: pp.69-107.

Vitagliano, L., R. Berisio, et al. (2001). "Structural bases of collagen stabilization induced by proline hydroxylation." Biopolymers 58(5): 459-64.

Voet, D. and J. G. Voet (2004). Biochemistry. New York, NY, J. Wiley & Sons.

Wang, H. M., J. Chan, et al. (1978). "Cleavage of native type III collagen in the collagenase susceptible region by thermolysin." <u>Biochim Biophys Acta</u> 533(1): 270-7.

Weingarten, H. and J. Feder (1985). "Spectrophotometric assay for vertebrate collagenase." <u>Anal. Biochem.</u> 147(2): 437-40.

Weingarten, H., R. Martin, et al. (1985). "Synthetic substrates of vertebrate collagenase." <u>Biochemistry.</u> 24(23): 6730-4.

Welgus, H. G., J. J. Jeffrey, et al. (1982). "The gelatinolytic activity of human skin fibroblast collagenase." <u>J. Biol. Chem.</u> 257(19): 11534-9.

Wendt, P., K. von der Mark, et al. (1972). "The covalent structure of collagen. The amino-acid sequence of the 112-residues. Amino-terminal part of peptide 1-CB6 from calf-skin collagen." <u>Eur J Biochem</u> 30(1): 169-83.

Williams, K. E. and D. R. Olsen (2009). "Matrix metalloproteinase-1 cleavage site recognition and binding in full-length human type III collagen." <u>Matrix Biol</u> 28(6): 373-9.

Windsor, L. J., H. Birkedal-Hansen, et al. (1991). "An internal cysteine plays a role in the maintenance of the latency of human fibroblast collagenase." <u>Biochemistry</u> 30(3): 641-7.

Xu, Y., D. R. Keene, et al. (2002). "Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices." <u>J Biol Chem</u> 277(30): 27312-8.

Yoon, M. K., V. Venkatachalam, et al. (2009). "Residual structure within the disordered C-terminal segment of p21(Waf1/Cip1/Sdi1) and its implications for molecular recognition." <u>Protein Sci</u> 18(2): 337-47.

Appendix 1:
A Conformational Selection Mechanism of Collagen Degradation

Appendix 1 is adapted from: Nerenberg, P.S., Salsas-Escat, R., and Stultz, C.M. (2008). "Do collagenases unwind triple-helical collagen before peptide bond hydrolysis? Reinterpreting experimental observations with mathematical models." Proteins **70**(4): 1154-61.

A1.1 Introduction

Collagenolysis is an enigmatic process. Collagenases have catalytic sites that are too narrow to accommodate the triple-helical structure of collagen, and the scissile bond within the triple helix is not solvent accessible (Fields 1991; Overall 2002; Stultz 2002). Consequently, the precise mechanism of collagenolysis is unclear. It has been suggested that collagenases gain access to their cleavage sites by actively unwinding triple-helical collagen before peptide bond cleavage (Overall 2002; Chung, Dinakarpandian et al. 2004). According to this theory, collagenases not only hydrolyze scissile bonds in collagen, but they also function as triple-helicases; that is, enzymes that locally unwind collagen's triple-helical structure. Matrix metalloproteases (MMPs) are zinc-dependent proteases that cleave extracellular matrix components at specific sites (Fields 1991; Overall 2002). MMP1, in particular, contains a catalytic domain and a hemopexin-like domain that are connected through a short linker region (Murphy and Knauper 1997). Although the catalytic domain alone can cleave denatured collagen, it cannot degrade collagen in its triple-helical form (Clark and Cawston 1989; Overall 2002; Chung, Dinakarpandian et al. 2004). When type I collagen is incubated with an excess of MMP1's catalytic domain, no significant collagen degradation products are formed - an observation that has been viewed as inconsistent with the notion that spontaneous unwinding of collagen occurs in the vicinity of the collagenase-cleavage site (Chung, Dinakarpandian et al. 2004). In addition, while the catalytically inactive form of MMP1 (E200A) cannot hydrolyze collagen, it can facilitate collagenolysis by enzymes that only cleave denatured collagen (Chung, Dinakarpandian et al. 2004). These observations lend credence to the conjecture that collagenases locally unwind triple helical collagen before collagenolysis. Most theories that support a triple-helicase function for MMPs require (i) collagen to at least bind both the catalytic

146

and hemopexin-like domains; and (ii) coordinated domain motions that produce lateral tension, axial compression, or bending of the collagen triple helix (Overall 2002).

In principle, these domain motions would have large energetic requirements; however, collagenolysis does not require energy input (Fields 1991; Murphy and Knauper 1997; Overall 2002; Chung, Dinakarpandian et al. 2004). Using a standard kinetic scheme coupled with transition state theory, we estimate the energetic barrier associated with this enzyme-mediated reaction to be ~20 kcal/mol (see Methods). This significant barrier exists when the enzyme is present. However, barriers of this magnitude are characteristic of processes that typically do not occur on biological time scales (Dugave and Demange 2003; Hollowell, Younvanich et al. 2007).

Although the aforementioned experimental studies imply that partially unfolded states of collagen do not occur spontaneously (Clark and Cawston 1989; Chung, Dinakarpandian et al. 2004), other studies suggest that collagen is conformationally labile in the vicinity of the collagenase cleavage site (Fan, Li et al. 1993; Fiori, Sacca et al. 2002; Leikina, Mertts et al. 2002; Stultz 2002; Sacca, Fiori et al. 2003). Accordingly, one alternate theory is that collagen exists as an equilibrium distribution of conformational states, where some states are partially unfolded in the vicinity of the collagenase cleavage site. Collagenolysis then occurs when collagenases bind to, stabilize, and cleave these partially unfolded conformers. Dynamical simulations of collagen-like model peptides suggest that collagen can adopt two distinct conformational states near its scissile bond (Stultz 2002). One state corresponds to the native triple-helical structure (N), and the other corresponds to a partially unfolded conformation that contains a relatively exposed scissile bond. We refer to this latter conformation as vulnerable (V), because collagen structures that adopt this conformation are relatively vulnerable to

collagenolysis (Stultz and Edelman 2003). In light of these observations, we reexamined prior experimental data on the interaction between MMP1 and type I collagen to determine whether these data are consistent with the aforementioned experimental observations on the kinetics of collagenolysis.

A1.2 Results

Collagenolysis is typically described by the standard kinetic scheme:

$$N + E \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C_{NE} \tag{A1.1}$$

$$C_{NE} \xrightarrow{k_{cat}^{NE}} P + E \tag{A1.2}$$

where N denotes the native state of collagen (i.e., its triple-helical structure), E denotes full length MMP1, C_{NE} is the enzyme-substrate complex, P denotes the products of collagenolysis, and the associated rate constants are explicitly shown. The rate at which collagen degradation products is formed is given by:

$$\frac{d[P]}{dt} = k_{cat}^{NE}[C_{NE}] \tag{A1.3}$$

where k_{cat}^{NE} is the experimentally observed turnover rate for type I collagen. In general, k_{cat}^{NE} is quite small, in the order of 0.015s⁻¹ for type I collagen (Welgus, Jeffrey et al. 1981). This slow turnover rate is rationalized using the argument that k_{cat}^{NE} measures both peptide bond hydrolysis and collagenase-mediated triple-helical unwinding, where the latter is an energetically costly reaction (Overall 2002; Chung, Dinakarpandian et al. 2004) (see Methods). Given that collagenolysis does not require energy input, we explored a reaction scheme that does not rely on enzyme-mediated unwinding. We begin by assuming that collagen exists as an equilibrium distribution of states as previously outlined; that is, N corresponds to the native triple-helical state and V corresponds to a partially unfolded conformation (Stultz 2002; Stultz and Edelman 2003). Our reaction scheme is based on the notion that collagenases bind and cleave vulnerable

conformers (Figure A1.1). Moreover, we explicitly exclude a step describing direct unfolding in the presence of the enzyme, $C_{NE} \rightleftarrows C_{VE}$, because we wish to determine whether previous experimental observations can be explained without appealing to a mechanism that invokes collagen unfolding while bound to the enzyme; for example, enzyme-mediated unwinding.

$$C_{NE} \underset{k_{on}^N}{\overset{k_{off}^N}{\rightleftarrows}} N \underset{k_2}{\overset{k_1}{\rightleftarrows}} V \overset{E}{\underset{k_{off}^V}{\overset{k_{on}^V}{\rightleftarrows}}} C_{VE} \xrightarrow{k_{cat}^{VE}} P+E$$

Figure A1.1. Reaction scheme used for numerical simulations. N denotes the native state of collagen (i.e., its triple-helical structure); E denotes full-length MMP1; C_{NE} is the enzyme-substrate complex; C_{VE} denotes the complex formed between the enzyme E and the partially unfolded state V; and P denotes the products of collagenolysis. The model itself, as outlined here, makes no assumption about the form of collagen and therefore is applicable to both collagen in its monomeric form as well as in its fibrillar form.

It is important to note that in this formalism the vulnerable state is quite distinct from the denatured state. We define a denatured conformation as a completely unfolded structure lacking triple helical structure. By contrast, vulnerable conformers have considerable native-like structure in which the native state and the vulnerable state differ in the local region about the scissile bond. Data from dynamical simulations and NMR experiments on collagen-like model peptides are consistent with the notion that such locally unfolded states of collagen exist (Fan, Li et al. 1993; Fiori, Sacca et al. 2002; Stultz 2002; Sacca, Fiori et al. 2003).

In this reaction scheme, hydrolysis of scissile bonds in collagen occurs only when collagenases recognize and cleave vulnerable conformers (Figure A1.1). Under steady-state conditions, the experimentally observed rate of collagenolysis [k_{cat}^{NE} in Eq. (A1.3)] is a function of $K_{eq} = [V]/[N] = k_1/k_2$ (the equilibrium constant dictating the relative concentrations of

vulnerable and native conformers), $K_{bind}^{VE} = [VE]/([V][E]) = k_{on}^V/k_{off}^V$ (the affinity of the enzyme for vulnerable states), and $K_{bind}^{NE} = [NE]/([N][E]) = k_{on}^N/k_{off}^N$ (the affinity of the enzyme for native states):

$$K_{cat}^{NE} \approx k_{cat}^{VE} K_{eq} K_{bind}^{VE} K_D^{NE} \tag{A1.4}$$

where $K_D^{NE} = (K_{bind}^{NE})^{-1}$ (see Methods for derivation). When K_{eq} is low, the native state is preferred and consequently collagen degradation will occur slowly. Likewise, mutations in either the enzyme or collagen that reduce the enzyme's affinity for vulnerable states will significantly lower K_{bind}^{VE}, and subsequently the rate at which collagen is degraded. Equation (A1.4) demonstrates that small values for K_{cat}^{NE} can be explained without invoking a mechanism that involves collagenase-mediated unwinding.

To explore whether this reaction scheme is consistent with prior experiments, we numerically solve the ODEs arising from the reaction scheme shown in Figure A1.1 to obtain estimates of the extent of collagenolysis as a function of K_{eq}, K_{bind}^{VE}, and K_{bind}^{NE}. Unlike the derivation of Eq. (A1.4), these calculations do not assume steady state conditions and therefore represent exact solutions to the chemical kinetics. Furthermore, to be consistent with prior experiments, we report the total fraction of collagen degraded after incubating 1 μM collagen with enzyme for 48h (Chung, Dinakarpandian et al. 2004).

We first consider the case where the enzyme concentration is significantly lower than the total collagen concentration, as is typically done in collagen-degradation experiments that employ full-length enzyme. In the present study, the precise choice of enzyme concentration is made to be consistent with prior experiments (Chung, Dinakarpandian et al. 2004). Figure A1.2 (A–D) depicts the fraction of collagen degraded as a function of K_{eq}, K_{bind}^{VE}, and K_{bind}^{NE}. The

collagen degradation profiles are quite similar over a large range of K_{bind}^{NE} values. In particular, for each value of K_{bind}^{NE} three distinct regimes are readily identified:

1. when binding to the vulnerable state is weak or disfavored ($K_{bind}^{VE} < 10^3 \text{M}^{-1}$), little or no collagen degradation occurs regardless of the value of K_{eq};

2. when binding is relatively strong ($K_{bind}^{VE} > 10^3 \text{M}^{-1}$) and vulnerable conformations are sampled infrequently, $K_{eq} < 10^{-1}$, little to no collagen degradation occurs even after a 48h incubation; and

3. when $K_{bind}^{VE} > 10^3 \text{M}^{-1}$ and $K_{eq} > 10^{-1}$ (or $\sim 10^1$ for $K_{bind}^{NE} = 10^8 \text{M}^{-1}$) significant collagen degradation occurs.
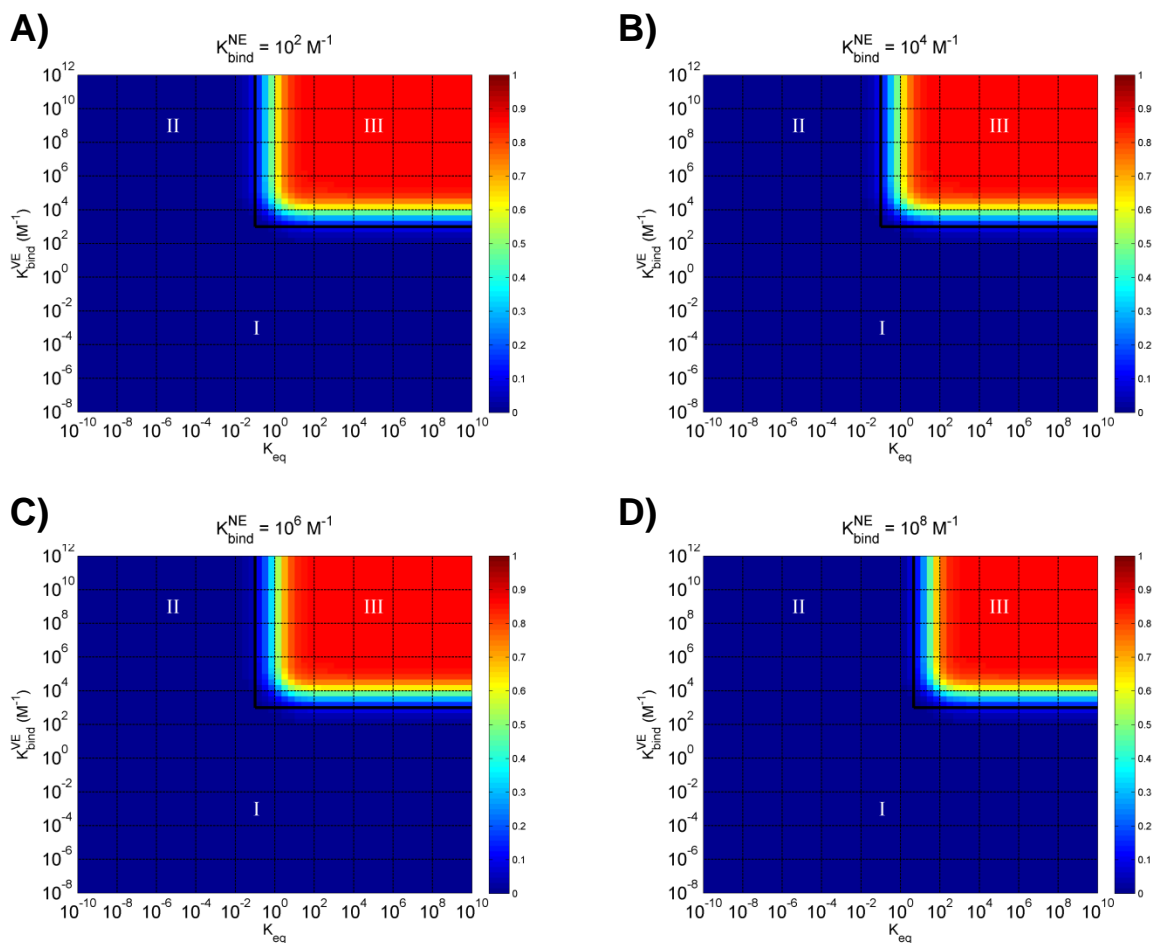
Figure A1.2. Fraction of collagen degraded at 298K after 48h with [E]/[total collagen] $= 10^{-2}$, and A) $K_{bind}^{NE} = 10^2 M^{-1}$, B) $K_{bind}^{NE} = 10^4 M^{-1}$, C) $K_{bind}^{NE} = 10^6 M^{-1}$, and D) $K_{bind}^{NE} = 10^8 M^{-1}$

These data demonstrate that collagenolysis requires vulnerable states to be sampled relatively frequently and relatively strong binding of collagenases to vulnerable conformers. Interestingly, unfolding simulations on a collagen-like peptide that models a region adjacent to the true collagenase cleavage site suggest that $K_{eq} \approx 10^0$ at 298 K, implying that the first criterion is met at room temperature (Stultz 2002). The fact that prior experiments have observed significant degradation when collagen is incubated with low enzyme concentrations at 298 K suggests that $K_{bind}^{VE} > 10^3 M^{-1}$ (Chung, Dinakarpandian et al. 2004). In addition, if $K_{bind}^{NE} = 10^8 M^{-1}$

then little collagen degradation will be observed after 48h, assuming $K_{eq} = 10^0$. Hence, the fact that significant collagenolysis does occur at room temperature suggests that $K_{bind}^{NE} < 10^8 \text{M}^{-1}$. We note that accurate estimates for K_{bind}^{NE} at 298K are difficult to obtain experimentally as our data suggest that a significant percentage of collagen molecules adopt a vulnerable conformation at this temperature. Nevertheless, binding constants for collagenase binding to fibrillar collagen at 298 K are near 10^6M^{-1}, and therefore well below this upper bound (Welgus, Jeffrey et al. 1980).

While full-length enzyme can cleave collagen at relatively low enzyme concentrations, experiments that employ mutant forms of enzyme may require much larger enzyme concentrations to observe degradation. Consequently, we study how K_{eq}, K_{bind}^{VE}, and K_{bind}^{NE} affect the rate of collagen-degradation when a high enzyme concentration is present (Figure A1.3 A and B). In Figure A1.3, we have explicitly labeled the region corresponding to incubating wild-type enzyme with collagen (i.e., $K_{bind}^{VE} > 10^3 \text{M}^{-1}$ and $K_{eq} > 10^{-1}$). One thing is apparent from these data. High enzyme concentrations do not necessarily guarantee that extensive collagenolysis will occur even when collagen is incubated with an equimolar concentration of enzyme. Collagenolysis is abrogated when the mutant of interest has reduced affinity for vulnerable states.

A)

B)

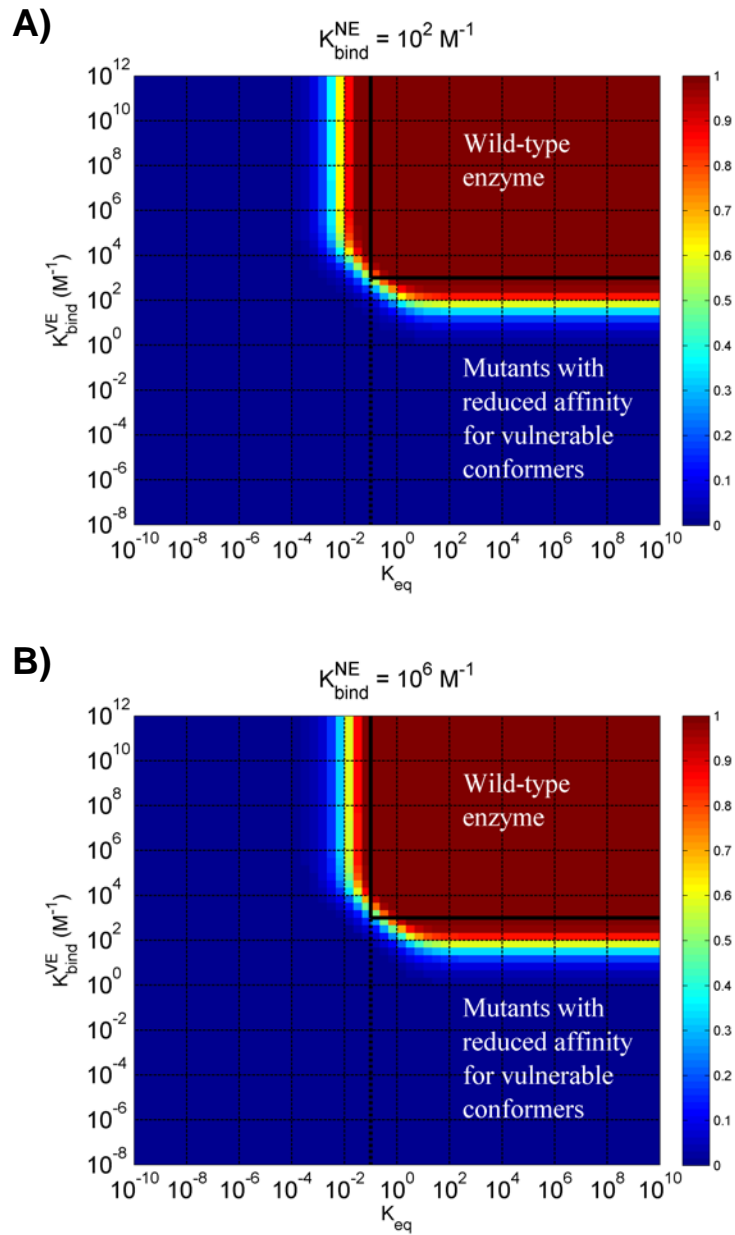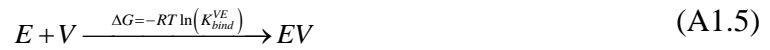Figure A1.3. Fraction of collagen degraded at 298K after 48h with [E]/[total collagen] = 1, and A) $K_{bind}^{NE} = 10^2 M^{-1}$, and B) $K_{bind}^{NE} = 10^6 M^{-1}$

Prior experiments that incubated collagen with high concentrations of MMP1 mutants lacking a hemopexin domain did not observe any collagenolysis after 48h (Chung, Dinakarpandian et al. 2004). Although these results have been interpreted to mean that partially

unfolded states do not occur spontaneously, they can also be explained by the notion that removal of the hemopexin-like domain reduces the affinity of MMP1 for collagen. This is especially poignant given that MMP1 hemopexin-like domains contain sites that bind collagen (Clark and Cawston 1989; Windsor, Birkedal-Hansen et al. 1991; Bigg, Shi et al. 1997). Figure A1.3 demonstrates that a substantial reduction in the binding affinity for the native state (K_{bind}^{NE}), which is expected for mutant enzymes lacking the hemopexin-like domain, does not noticeably affect the collagen degradation profile. Instead, as in the case of low enzyme concentrations, collagenolysis depends almost entirely on K_{eq} and K_{bind}^{VE}.

In general, mutations that destabilize the complex formed by enzyme and vulnerable conformations, such as the removal of collagen binding sites on the enzyme, will slow the rate of collagenolysis. If the destabilization is significant, little collagen degradation will be observed after a 48h incubation time. To quantify the amount of destabilization needed to reduce the fraction of collagen degraded to negligible values, we computed the destabilization energy as follows. Let E denote wild-type MMP1, and E_m denote a mutant form of MMP1 that binds vulnerable states with reduced affinity such that incubating E_m with collagen for 48h results in degradation of less than 1% of the total collagen present. The two binding reactions are given by:

$$E + V \xrightarrow{\Delta G = -RT \ln\left(K_{bind}^{VE}\right)} EV \tag{A1.5}$$

$$E_m + V \xrightarrow{\Delta G_m = -RT \ln\left(K_{min}^{VE}\right)} E_m V \tag{A1.6}$$

when $K_{bind}^{VE} > K_{min}^{VE}$ the destabilization energy is given by:

$$\Delta\Delta G = \Delta G_m - \Delta G = RT\ln\left(\frac{K_{bind}^{VE}}{K_{min}^{VE}}\right) \tag{A1.7}$$

where K_{min}^{VE} is the reduced binding constant associated with E_m. When K_{bind}^{VE} falls below K_{min}^{VE}, less than 1% of the total collagen will be degraded after 48h. We note that K_{min}^{VE} is a function of K_{eq} and therefore the destabilization energy is a function of both K_{bind}^{VE} and K_{eq}. The main point here is that if MMP1 binds the vulnerable state with a binding constant of K_{bind}^{VE}, then destabilizing the bound state by $\Delta\Delta G$ kcal/mol will reduce collagenolysis such that less than 1% of the total collagen present will be degraded. When the enzyme binds vulnerable and native states with dissociation constants in the micromolar range ($K_{bind}^{NE} = 10^6$ and $K_{bind}^{VE} = 10^6$) the destabilization energy is less than 7.5 kcal/mol (Figure A1.4). Moreover, when $K_{bind}^{VE} > 10^3$ and $K_{eq} > 10^{-1}$ (ranges suggested by low enzyme concentration simulations), the destabilization energy ranges from 2.5 to 12.5 kcal/mol (Figure A1.4). Therefore, modest changes in the binding affinity - as expected with forms of MMP1 that lack collagen binding sites in the hemopexin-like domain - will have important consequences on the extent of collagenolysis.
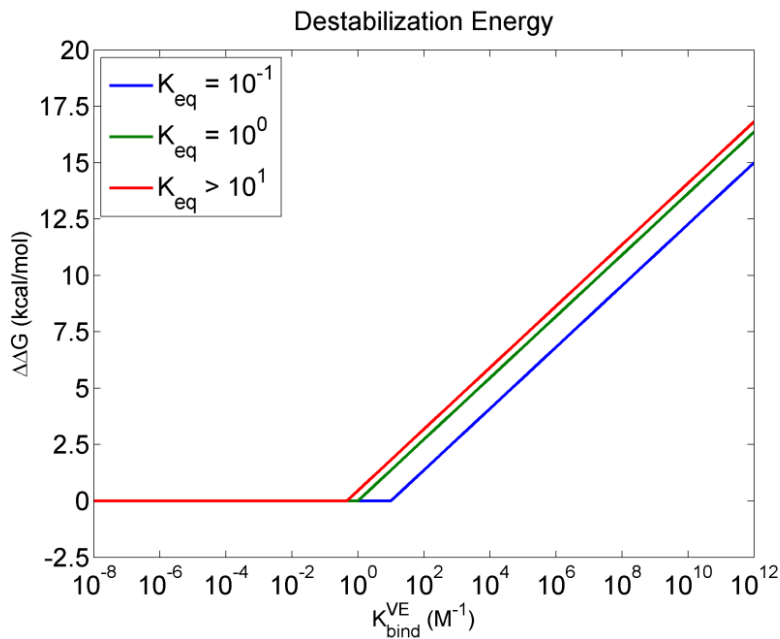
Figure A1.4. Destabilization energies at 298K with $[E] = 1$ μM and $K_{bind}^{NE} = 10^6 M^{-1}$. Energies plotted for $K_{eq} = 10^{-1}$, 10^0, and $> 10^1$. Because K_{min}^{VE} does not vary for $K_{eq} > 10^1$ (Figure A1.3 B), the destabilization energy curve is the same for all $K_{eq} > 10^1$. Given that the lower bound for degradation is $K_{eq} = 10^{-1}$, we observe that for a given K_{bind}^{VE}, the destabilization energy does not differ significantly over the range of K_{eq} suggested by our simulations.

A1.3 Discussion

Since the first crystallographic structures of MMPs were published, the apparent mismatch between the width of the MMP active site pocket (5Å) and the width of the collagen triple helix (15Å) has presented a formidable challenge in explaining the mechanism of collagen degradation (Fields 1991; Overall 2002; Stultz 2002). Previous explanations have focused on a hypothesis that relies on the active unwinding of the collagen triple-helix by MMPs; that is, the simultaneous binding of MMP domains and the coordinated motions of these domains could exert the necessary mechanical force to unravel the helix. Accordingly, a number of experimental observations have been interpreted within this paradigm. Our results, however, suggest that these experiments can be understood without appealing to a mechanism that requires MMPs to act as triple-helicases. In this regard, we note the following:

1. Previous experiments suggest that the catalytic domain of MMP1 alone, MMP1(ΔC), cannot cleave collagen (Clark and Cawston 1989; Chung, Dinakarpandian et al. 2004) - a finding consistent with the notion that preformed partially unwound states of collagen do not occur spontaneously. However, our data demonstrate that the lack of collagen degradation by MMP1(ΔC) is explained by the observation that removal of the hemopexin-like domain reduces the enzyme's affinity for (and thus stabilization of) the vulnerable state relative to full length enzyme (Clark and Cawston 1989; Murphy, Allan et al. 1992; Chung, Dinakarpandian et al. 2004). As the hemopexin-like domain of MMP1 contains sites that bind collagen (Clark and Cawston 1989; Windsor, Birkedal-Hansen et al. 1991; Bigg, Shi et al. 1997), it is likely that MMP1(ΔC) and full-length enzyme have different affinities for both native and vulnerable states. However, as is evident from Figure A1.2 and Figure A1.3, the extent of collagenolysis is more

dependent on the enzyme's affinity for vulnerable conformers. Moreover, the critical role of the hemopexin-like domain in stabilizing vulnerable states is further supported by the observation that human leukocyte elastase (HLE), an enzyme that normally does not cleave triple-helical collagen can hydrolyze collagen when it is combined with solutions containing the MMP1 hemopexin-like domain (Chung, Dinakarpandian et al. 2004).

2. The binding of linker and hemopexin-like domains obtained from MT1-MMP has been shown to significantly perturb the triple-helical structure of type I collagen as measured by CD spectroscopy (Tam, Moore et al. 2004). As this perturbation is similar to the observed change in CD spectra when collagen is thermally denatured, this observation has been interpreted to mean that the linker and hemopexin-like domain alone are functioning as triple-helicases. If, however, collagen exists as an equilibrium distribution of different conformational states, and binding to the vulnerable state by the enzyme stabilizes this conformation, then the principle of mass action dictates that the overall structure of collagen in solution will be shifted to vulnerable conformations, yielding a change in the observed CD spectrum.

3. It has been noted that collagen can be cleaved when incubated with the catalytic domain of MMP1(ΔC) and severed MMP1 hemopexin-like domains (Hpx$_{MMP1}$) (Chung, Dinakarpandian et al. 2004). The lack of a linker connecting MMP1(ΔC) and Hpx$_{MMP1}$ makes active triple-helical unwinding unlikely, as a helicase function would require coordinated domain movements. However, if we consider that binding of the hemopexin-like domain stabilizes the vulnerable state, then these experimental results can be understood in the context of preformed native and vulnerable states.

160

4. Collagenolysis occurs when collagen is placed in solutions containing catalytically inactive enzyme (E200A), and other enzymes that normally do not degrade tripe-helical collagen; for example, full-length MMP3, catalytic domain of MMP3 alone, and HLE (Chung, Dinakarpandian et al. 2004). While these data have been interpreted to mean that E200A unwinds triple-helical collagen, they are explained by the notion that preexisting vulnerable states of collagen are stabilized by E200A (as noted in point 2).

The observed lack of collagen degradation in the case where collagen is incubated with MMP1(ΔC) may be further understood by considering how connected catalytic and hemopexin-like domains affect the collagenolytic ability of MMP1. Without a connected hemopexin-like domain to (i) bind to the correct region of collagen, near the scissile bond; (ii) stabilize the vulnerable state; and (iii) ensure that the connected catalytic domain is in the vicinity of the scissile bond, the local concentration of catalytic domain near the collagenase cleavage site would be significantly smaller in experiments involving MMP1(ΔC) as opposed to full-length enzyme. Our results demonstrate that a straightforward reaction mechanism, which assumes that collagen is flexible in the vicinity of the collagenase cleavage site, can explain collagenolysis without implying that collagenases perform the energetically costly task of actively unwinding triple-helical collagen. Collagen, like all other biological heteropolymers, undergoes thermal fluctuations that cause it to sample distinct structures in the neighborhood of the native state, some of which feature an ''unwound'' or vulnerable portion of the triple helix near the collagenase cleavage site. Collagenolysis occurs when collagenases stabilize the appropriate vulnerable state and, by virtue of having connected catalytic and hemopexin-like domains,

greatly increase the concentration of catalytic domains in the immediate vicinity of the cleavage site.

A1.4 Methods

First, we construct a reaction scheme that describes collagenolysis without assuming that MMPs have any triple-helicase activity (see Figure A1.1). This reaction scheme leads to a set of ordinary differential equations (ODEs) as listed below:

$$\frac{d[N]}{dt} = k_2[V] + k_{off}^N[C_{NE}] - k_1[N] - k_{on}^N[N][E] \tag{A1.8}$$

$$\frac{d[V]}{dt} = k_1[N] + k_{off}^V[C_{VE}] - k_1[V] - k_{on}^V[V][E] \tag{A1.9}$$

$$\frac{d[C_{NE}]}{dt} = k_{on}^N[N][E] - k_{off}^N[C_{NE}] \tag{A1.10}$$

$$\frac{d[C_{VE}]}{dt} = k_{on}^N[V][E] - k_{off}^V[C_{VE}] - k_{cat}^{VE}[C_{VE}] \tag{A1.11}$$

$$\frac{d[P]}{dt} = k_{cat}^{VE}[C_{VE}] \tag{A1.12}$$

$$\frac{d[E]}{dt} = k_{off}^N[C_{NE}] + k_{off}^V[C_{VE}] + k_{cat}^V[C_{VE}] - k_{on}^N[N][E]$$
$$- k_{on}^V[V][E] \tag{A1.13}$$

To numerically solve these ODEs, we require estimates for each rate constant. As k_{cat}^{VE} corresponds to the turnover number when the scissile bond is exposed and readily accessible, we assume that an upper bound for K_{cat}^{VE} can be approximated by the catalytic rate constant of MMP-1 on denatured collagen chains (Fields, Van Wart et al. 1987). Type I collagen contains two $\alpha1(I)$ chains and one $\alpha2(I)$ chain. The rate constant for denatured $\alpha1(I)$ chains is ~230h^{-1} and the rate constant for denatured $\alpha2(I)$ chain is ~750h^{-1}. Therefore, to estimate the catalytic rate constant for type-I collagen, we use $k_{cat}^{VE} = (1/3) \times (2 \times 230h^{-1} + 750h^{-1}) \approx 400h^{-1}$. The

true value for the catalytic rate constant on k_{cat}^{VE} is likely significantly lower given that vulnerable states are not fully unfolded; that is, scissile bonds in vulnerable conformations may not be as exposed as scissile bonds in denatured states. Using this relatively high value for k_{cat}^{VE} in the numerical simulations provides a relatively stringent test of the model as we demonstrate that little-to-no collagenolysis occurs over 48h after exposure of the catalytic domain to collagen using this value. Moreover, numerical simulations with lower values of k_{cat}^{VE} produce the same result.

Estimates for the remaining rate constants are determined by the equilibrium constants $K_{eq} = k_1/k_2, K_{bind}^{VE} = k_{on}^V/k_{off}^V$, and $K_{bind}^{NE} = k_{on}^N/k_{off}^N$. The amount of collagen degraded over short time intervals is strongly dependent on the precise values of these rate constants. However, as we are interested in the fraction of collagen degraded after an extended period of time (48h), it is the equilibrium constants rather than the rate constants that determine the final amount of degradation. For example, k_1 determines the rate at which native states transition to vulnerable states - a process driven by thermal fluctuations. If one is interested in collagen degradation after a long period of time, then it is the ratio of the forward rate k_1 to the backward rate k_2 that is most important. Simulations conducted using a wide range (over 10 orders of magnitude) of values for each rate constant verify that the precise values do not influence the results so long as the equilibrium constants are fixed. For each set of numerical simulations, K_{bind}^{NE} is held at a fixed value, while K_{bind}^{VE} is allowed to vary from 10^{-8} to $10^{12}M^{-1}$, ensuring that a wide range of equilibrium binding constants is simulated. Similarly, K_{eq}, the equilibrium constant for native and vulnerable states of collagen is also allowed to vary from 10^{-10} to 10^{10}. Although we have not made any assumptions in our model as to whether collagen is in a soluble or fibrillar form, the range of equilibrium constants simulated likely includes regimes corresponding to both forms of

collagen. That is, it is likely that the cleavage site is not as accessible in fibrillar collagen relative to its state in collagen monomers.

This effect, however, can be captured in K_{eq}, the equilibrium constant describing the relative amounts of native and vulnerable states. If the accessibility of the collagenase cleavage site is reduced in the fibrillar state, then this is effectively modeled by a low K_{eq}. As we consider many different values for K_{eq} in our simulations, the current set of simulations likely covers both scenarios, that is, cleavage of isolated collagen monomers and fibrillar collagen. Numerical simulations were performed in MATLAB (© Mathworks) using the ODE solver ode15s with the total collagen concentration set to 1μM.

A1.4.1 Enzymatic cleavage with unwinding

In this scenario, the MMP first binds to collagen, then unwinds it, and finally cleaves it. We assume that unwinding is a slow, but irreversible reaction. The reaction scheme is given by:

$$N + E \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C_{NE} \qquad (A1.14)$$

$$C_{NE} \xrightarrow{k_{unwind}} C_{VE} \qquad (A1.15)$$

$$C_{VE} \xrightarrow{k_{cat}^{VE}} P + E \qquad (A1.16)$$

where k_{on} and k_{off} are the on and off rates for the enzyme binding native collagen, respectively. k_{unwind} is the rate constant for unwinding a native conformation to a vulnerable conformation, and k_{cat}^{VE} is the rate constant for degradation of the now-accessible scissile bond.

To estimate the rate constant for unwinding, k_{unwind}, we first note that degradation experiments are typically fit with a simple Michaelis-Menten kinetics model to yield the

parameters K_m^{exp} and k_{cat}^{exp}. Using parameters from the unwinding reaction scheme, the experimentally measured Michaelis-Menten parameters are then given by (Kuby 1990):

$$K_m^{exp} = \frac{k_{cat}^{VE}}{k_{unwind} + k_{cat}^{VE}} \cdot \frac{k_{unwind} + k_{off}}{k_{on}} \qquad (A1.17)$$

$$k_{cat}^{exp} = \frac{k_{unwind} \, k_{cat}^{VE}}{k_{unwind} + k_{cat}^{VE}} \qquad (A1.18)$$

If we assume the unwinding is the rate-limiting step (i.e., $k_{unwind} \ll k_{cat}^{VE}$), we find:

$$k_{cat}^{exp} \approx \frac{k_{unwind} \, k_{cat}^{VE}}{k_{cat}^{VE}} = k_{unwind} \qquad (A1.19)$$

Thus, with Equation (A1.18) we can estimate k_{unwind} as being equal to the experimentally measured catalytic rate k_{cat}^{exp}. We can then use the transition state theory to calculate the free energy barrier for unwinding as follows:

$$k_{unwind} \approx \frac{k_B T}{h} \exp\left(\frac{-\Delta G^+}{RT}\right) = v \exp\left(\frac{-\Delta G^+}{RT}\right) \qquad (A1.20)$$

Thus, the activation energy is given by:

$$\Delta G^+ = -RT log(v^{-1} k_{unwind}) \qquad (A1.21)$$

At 298 K, v is 6.25 ps⁻¹ (Voet, Voet et al. 2006), and with $k_{cat}^{exp} \approx k_{unwind} \approx$ 0.015 s⁻¹(Welgus, Jeffrey et al. 1981), we find $\Delta G^+ \approx$ 20 kcal/mol. It is important to note that this barrier corresponds to the reaction *when the enzyme is present*.

A1.4.2 Collagenolysis with preformed partially unfolded states in solution

In this scenario, collagen can exist in either a native (folded) or vulnerable (partially unfolded) state. MMP can bind to either state, but cleavage will only occur when it is bound to

the vulnerable state. The reaction scheme is given by Figure A1.1 and the resulting ODEs are listed in Equations (A1.8) to (A1.13).

We can relate the rate of cleavage of the vulnerable state of collagen (or catalytic rate) in this scheme to the catalytic rate, k_{cat}^{NE}, in Equation (A1.3) of the aforementioned simple enzymatic scheme (found in Results) by noting that the rate of product creation (dP/dt) must be the same for both:

$$k_{cat}^{VE}[C_{VE}]_t = k_{cat}^{NE}[C_{NE}]_t \rightarrow k_{cat}^{VE} = k_{cat}^{NE} \frac{[C_{NE}]_t}{[C_{VE}]_t} \qquad \text{(A1.22)}$$

We can again gain insight by examining the steady state approximation for the intermediate vulnerable state-enzyme complex:

$$0 = \frac{d[C_{VE}]}{dt} = k_{on}^V[V][E] - k_{off}^V[C_{VE}] - k_{cat}^{VE}[C_{VE}]$$
$$= k_{on}^V[V][E] - (k_{off}^V + k_{cat}^{VE})[C_{VE}] \qquad \text{(A1.23)}$$

Recombining the above equation to solve for $[C_{VE}]$ yields:

$$[C_{VE}] = \frac{k_{on}^V[V][E]}{(k_{off}^V + k_{cat}^{VE})} \qquad \text{(A1.24)}$$

Similarly applying the steady state approximation to the intermediate native state-enzyme complex yields:

$$0 = \frac{d[C_{NE}]}{dt} = k_{on}^N[N][E] - k_{off}^N[C_{NE}] \qquad \text{(A1.25)}$$

Recombining the above equation to solve for $[C_{NE}]$ yields:

$$[C_{NE}] = \frac{k_{on}^N[N][E]}{k_{off}^N} \qquad \text{(A1.26)}$$

We now can calculate the ratio $[C_{NE}]/[C_{VE}]$ and thus relate k_{cat}^{VE} and k_{cat}^{NE}:

$$k_{cat}^{VE} = k_{cat}^{NE} \frac{[C_{NE}]_t}{[C_{VE}]_t}$$

$$k_{cat}^{VE} = k_{cat}^{NE} \frac{k_{on}^N [N][E]}{k_{off}^N} \frac{(k_{off}^V + k_{cat}^{VE})}{k_{on}^V [V][E]} =$$

$$= k_{cat}^{NE} \frac{k_{on}^N [N]}{k_{off}^N} \frac{(k_{off}^V + k_{cat}^{VE})}{k_{on}^V [V]}$$

$$= k_{cat}^{NE} \frac{[N]}{[V]} \frac{k_{on}^N (k_{off}^V + k_{cat}^{VE})}{k_{off}^N k_{on}^V}$$

(A1.27)

$$= k_{cat}^{NE} \left(K_{eq}\right)^{-1} (K_{bind}^{NE}) \frac{(k_{off}^V + k_{cat}^{VE})}{k_{on}^V}$$

$$= k_{cat}^{NE} \left(K_{eq}\right)^{-1} (K_{bind}^{NE}) \left((K_{cat}^{VE})^{-1} + \frac{k_{cat}^{VE}}{k_{on}^V}\right)$$

If assume that $k_{on}^V \gg k_{cat}^{VE}$ (binding is much faster than catalysis - an assumption valid for many enzyme-substrate complexes (Stryer 1988)), we find:

$$k_{cat}^{VE} \approx k_{cat}^{NE} \left(K_{eq}\right)^{-1} (K_{bind}^{NE})(K_{cat}^{VE})^{-1} = k_{cat}^{NE} \left(K_{eq}\right)^{-1} \frac{K_{bind}^{NE}}{K_{cat}^{VE}} \qquad \text{(A1.28)}$$

Rearranging to obtain k_{cat}^{NE}, we find:

$$k_{cat}^{NE} \approx k_{cat}^{VE} \left(K_{eq}\right) \frac{K_{bind}^{VE}}{K_{cat}^{NE}} \qquad \text{(A1.29)}$$

A1.5 Bibliography Appendix 1

Bigg, H. F., Y. E. Shi, et al. (1997). "Specific, high affinity binding of tissue inhibitor of metalloproteinases-4 (TIMP-4) to the COOH-terminal hemopexin-like domain of human gelatinase A. TIMP-4 binds progelatinase A and the COOH-terminal domain in a similar manner to TIMP-2." J Biol Chem **272**(24): 15496-500.

Chung, L., D. Dinakarpandian, et al. (2004). "Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis." Embo J **23**(15): 3020-30.

Clark, I. M. and T. E. Cawston (1989). "Fragments of human fibroblast collagenase. Purification and characterization." Biochem J **263**(1): 201-6.

Dugave, C. and L. Demange (2003). "Cis-trans isomerization of organic molecules and biomolecules: implications and applications." Chem Rev **103**(7): 2475-532.

Fan, P., M. H. Li, et al. (1993). "Backbone dynamics of (Pro-Hyp-Gly)10 and a designed collagen-like triple-helical peptide by 15N NMR relaxation and hydrogen-exchange measurements." Biochemistry **32**(48): 13299-309.

Fields, G. B. (1991). "A model for interstitial collagen catabolism by mammalian collagenases." J Theor Biol **153**(4): 585-602.

Fields, G. B., H. E. Van Wart, et al. (1987). "Sequence specificity of human skin fibroblast collagenase. Evidence for the role of collagen structure in determining the collagenase cleavage site." J Biol Chem **262**(13): 6221-6.

Fiori, S., B. Sacca, et al. (2002). "Structural properties of a collagenous heterotrimer that mimics the collagenase cleavage site of collagen type I." J Mol Biol **319**(5): 1235-42.

Hollowell, H. N., S. S. Younvanich, et al. (2007). "Thermodynamic analysis of the low- to physiological-temperature nondenaturational conformational change of bovine carbonic anhydrase." J Biochem Mol Biol **40**(2): 205-11.

Kuby, S. A. (1990). Study of Enzymes: Enzyme Catalysis, Kinetics, and Substrate Binding. Boca Raton, CRC Press.

Leikina, E., M. V. Mertts, et al. (2002). "Type I collagen is thermally unstable at body temperature." Proc Natl Acad Sci U S A **99**(3): 1314-8.

Murphy, G., J. A. Allan, et al. (1992). "The role of the C-terminal domain in collagenase and stromelysin specificity." J Biol Chem **267**(14): 9612-8.

Murphy, G. and V. Knauper (1997). "Relating matrix metalloproteinase structure to function: why the "hemopexin" domain?" Matrix Biol **15**(8-9): 511-8.

Overall, C. M. (2002). "Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites." Mol Biotechnol **22**(1): 51-86.

Sacca, B., S. Fiori, et al. (2003). "Studies of the local conformational properties of the cell-adhesion domain of collagen type IV in synthetic heterotrimeric peptides." Biochemistry **42**(12): 3429-36.

Stryer, L. (1988). Biochemistry. New York, W.H. Freeman & Company.

Stultz, C. M. (2002). "Localized unfolding of collagen explains collagenase cleavage near imino-poor sites." J Mol Biol **319**(5): 997-1003.

Stultz, C. M. and E. R. Edelman (2003). "A structural model that explains the effects of hyperglycemia on collagenolysis." Biophys J **85**(4): 2198-204.

Tam, E. M., T. R. Moore, et al. (2004). "Characterization of the distinct collagen binding, helicase and cleavage mechanisms of matrix metalloproteinase 2 and 14 (gelatinase A and MT1-MMP): the differential roles of the MMP hemopexin c domains and the MMP-2 fibronectin type II modules in collagen triple helicase activities." J Biol Chem **279**(41): 43336-44.

Voet, D., J. G. Voet, et al. (2006). Fundamentals of biochemistry: life at molecular level. Hoboken, Wiley.

Welgus, H. G., J. J. Jeffrey, et al. (1981). "The collagen substrate specificity of human skin fibroblast collagenase." J Biol Chem **256**(18): 9511-5.

Welgus, H. G., J. J. Jeffrey, et al. (1980). "Characteristics of the action of human skin fibroblast collagenase on fibrillar collagen." J Biol Chem **255**(14): 6806-13.

Windsor, L. J., H. Birkedal-Hansen, et al. (1991). "An internal cysteine plays a role in the maintenance of the latency of human fibroblast collagenase." Biochemistry **30**(3): 641-7.

Appendix 2:
Testing a Conformational Selection Model of Collagen Degradation

A2.1 Introduction

In Chapters 1 and 2 we found that locally unfolded states exist in collagen, and hypothesized that they play a role in collagen degradation and collagenase cleavage site specificity. In Chapter 3 we have shown that these unfolded states can be cleaved by the catalytic domain of collagenases alone, suggesting that these unfolded states exist in the absence of collagenases. The results shown in Chapter 3 indicate that, while able to cleave triple helical collagen, CMMP1 and CMMP8 are very inefficient. This result verifies the prediction of the two state degradation model presented in the Introduction chapter, and explained in detail in Appendix 1. The failure to observe type I collagen degradation by CMMP1 and CMMP8 (Clark and Cawston 1989; Schnierer, Kleine et al. 1993; Chung, Dinakarpandian et al. 2004) is not due to the lack of locally unfolded states, but due to the low affinity of these enzymes towards these locally unfolded states compared to full length MMP1 and MMP8.

Since in Chapter 3 we unequivocally show that CMMP1 and CMMP8 are to cleave type I collagen, in Appendix 2 we want to model this reaction using the two state degradation model previously presented. In particular, we wish to characterize the equilibrium existing between folded and locally unfolded states, as well as the interaction between CMM1/8 and these locally unfolded states.

A2.2 Results

A2.2.1 A conformational selection model for collagenolysis

The data presented in Chapter 3 demonstrate that type I collagen degradation at the unique collagenase cleavage site can occur *in vitro* at room temperature with only the catalytic domain of MMPs. Since CMMP8 and CMMP1 cleave unfolded collagen at multiple sites (Figure 3.5 A lane 8, Figure 3.5 B lane 6), but triple-helical collagen at only one site (Figure 3.5 A lane 7, Figure 3.5 B lane 5), we interpret our findings in light of a model where unfolding of collagen at the unique cleavage site enables collagenases to gain access to their corresponding scissile bonds (Figure A2.1) (Nerenberg, Salsas-Escat et al. 2008). In this model, collagen can exist in both native triple-helical (*N*) and vulnerable states, or locally unfolded, (*V*), and collagenolysis only occurs when the enzyme (consisting of only the catalytic domain, *C*) binds and cleaves the vulnerable state. Vulnerable states correspond to the ensemble of conformers that have the collagenase cleavage site in an unfolded and solvent-exposed conformation. We use *N·C* and *V·C* to denote complexes of the native and vulnerable states, respectively, with the mutant enzyme that only contains the catalytic domain, and *P* denotes the degradation products released by the enzyme after cleavage.
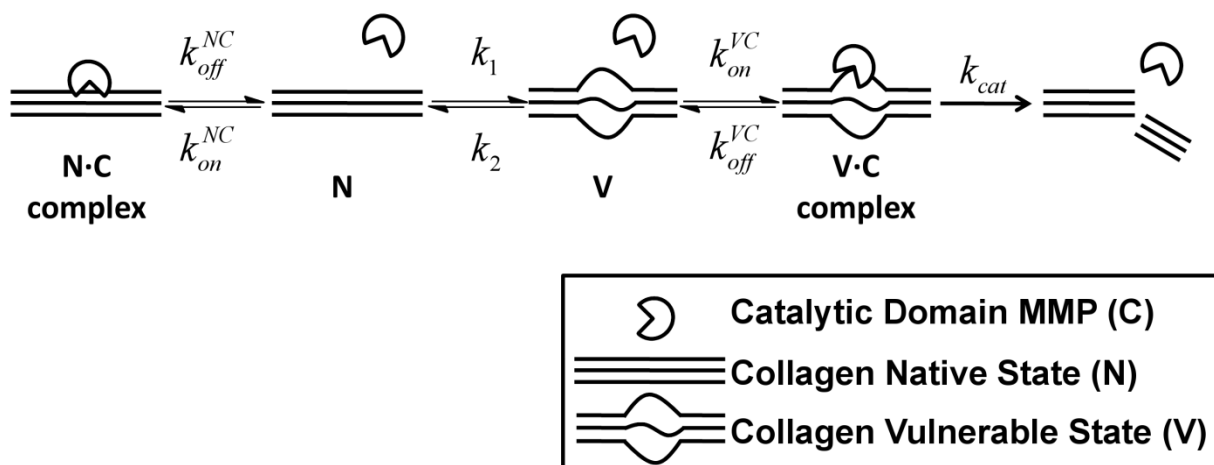
Figure A2.1. A conformational selection mechanism for collagenolysis with MMP catalytic domains. Collagen exists in an equilibrium between native (N) and vulnerable (V) states with the equilibrium determined by $K_{eq} = \frac{k_1}{k_2}$. The catalytic domain of MMPs (C) interacts in a non-specific manner with the native state with binding constant $K_{bind}^{NC} = \frac{k_{on}^{NC}}{k_{off}^{NC}}$, yielding the $N{\cdot}C$ complex. C binds to the vulnerable state, V, with binding constant $K_{bind}^{VC} = \frac{k_{on}^{VC}}{k_{off}^{VC}}$, forming the $V{\cdot}C$ complex. The $V{\cdot}C$ complex is then degraded with catalytic rate k_{cat}.

This reaction scheme naturally leads to a set of ordinary differential equations (ODEs) that can be solved numerically, yielding the concentration of each species as a function of time (see Methods). Numerical solutions of the ODEs are exact in that they do not make any assumptions about steady state behavior and they specifically account for the different rate constants associated with each of the different species in the reaction. Generating solutions for the model therefore requires inputs in the form of rate constants associated with each of the various steps in the mechanism.

Previous studies with a related reaction scheme demonstrated that the model is not sensitive to the specific choice of the individual rate constants when one is interested in the reactant concentrations over long times with respect to the individual rate constants themselves (Nerenberg, Salsas-Escat et al. 2008). In this scenario, the solutions to the ODEs depend more on

the ratio of the rate constants (Nerenberg, Salsas-Escat et al. 2008). Since our experiments occur over days and theoretical estimates for local unfolding/folding events within the collagen triple helix are on the order of inverse nanoseconds (Stultz 2002; Nerenberg and Stultz 2008), and rate constants for binding reactions in general are on the order of inverse milliseconds (Ottl, Gabriel et al. 2000), we explore the behavior of the system as a function of the associated equilibrium and binding constants, $K_{eq} = \frac{k_1}{k_2}$, $K_{bind}^{NC} = \frac{k_{on}^{NC}}{k_{off}^{NC}}$, and $K_{bind}^{VC} = \frac{k_{on}^{VC}}{k_{off}^{VC}}$. It is important to note that while we use these equilibrium/binding constants to describe the behavior of the model, the associated ODEs explicitly depend on the individual rate constants.
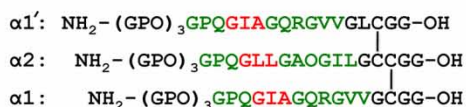
As described in the Methods section we estimate K_{bind}^{NC} using previously determined collagenase binding rate constants, which were obtained under conditions when the triple-helical state is expected to be most stable (Ottl, Gabriel et al. 2000). We note, however, that the model is not sensitive to the choice of binding constant for the native state, K_{bind}^{NC}, see Appendix 1 for more details (Nerenberg, Salsas-Escat et al. 2008). The catalytic rate constant, k_{cat}, corresponds to the rate of peptide bond hydrolysis after the enzyme has bound the unfolded region containing the cleavage site. Bounds for k_{cat} are therefore obtained from experimentally measured rate constants from MMP-mediated degradation of gelatin and unfolded collagen-like peptides (see Methods) (Welgus, Jeffrey et al. 1982; Netzel-Arnett, Fields et al. 1991). This leaves two undetermined parameters for the model: K_{eq} (the equilibrium constant describing the relative concentration of vulnerable and native conformers) and K_{bind}^{VC} (the binding constant of the catalytic domain for vulnerable states). The method by which K_{eq} and K_{bind}^{VC} are converted into forward/backward and on/off rate constants, respectively, for the numerical simulations is described in the Methods section.

To determine estimates for the missing parameters, we fit the ODEs arising from the model shown in Figure A2.1 to experimental degradation data. We begin by focusing on previous degradation experiments that incubated CMMP8 at room temperature with a heterotrimeric type I collagen-like model peptide that contains the collagenase cleavage site and its surrounding residues (Trimer A, Figure A2.2 A) (Ottl, Gabriel et al. 2000). The melting temperature of this peptide is 9ºC, and experiments were performed at 25ºC; hence, the peptide is largely unfolded at this temperature. In our model K_{eq} represents the equilibrium constant between native and vulnerable states, where the vulnerable ensemble includes all conformers that have the collagenase cleavage site in an unfolded and solvent-exposed conformation. Consequently, for this system we expect $K_{eq} > 1$, and therefore fitting these data to the model shown in Figure A2.1 provides a test of the method. Using the reaction scheme outlined in Figure A2.1, we computed the amount of peptide that would be degraded as a function of time for a range of K_{eq}, $\left(K_{bind}^{VC}\right)_{CMMP8}$, and k_{cat} values and compared these degradation profiles to the corresponding experimental data. For each triplet, K_{eq}, $\left(K_{bind}^{VC}\right)_{CMMP8}$, k_{cat}, we computed the root mean square error (RMSE) between the degradation time course obtained with the model and the experimental data (see Methods). While a relatively wide range of values for K_{eq} and $\left(K_{bind}^{VC}\right)_{CMMP8}$ are tried ($10^{-6} \leq K_{eq} \leq 10^{6}$, 10^{0} M$^{-1} \leq \left(K_{bind}^{VC}\right)_{CMMP8} \leq 10^{12}$ M^{-1}), the best fits are obtained when $K_{eq} > 30$ and $\left(K_{bind}^{VC}\right)_{CMMP8} = 0.7\text{-}1.1 \times 10^{6}$ M^{-1}, regardless of the value of k_{cat} that was used (0.11 s$^{-1} \leq k_{cat} \leq 11.1$ s^{-1}) (Figure A2.3). Moreover, varying k_{cat} by two orders of magnitude caused the minimum RMSE to vary by only 2%. As the best fits have $K_{eq} > 30$, these results suggest that the vulnerable state of the peptide dominates at room temperature - a finding in agreement with the experimental conditions, as discussed above. Moreover, the

predicted results from the model using these values show excellent agreement with experiment

(Figure A2.2 B).

A)

α1′: NH$_2$-(GPO)$_3$GPQ**GIA**GQRGVVGLCGG-OH

α2: NH$_2$-(GPO)$_3$GPQ**GLL**GAOGILGCCGG-OH
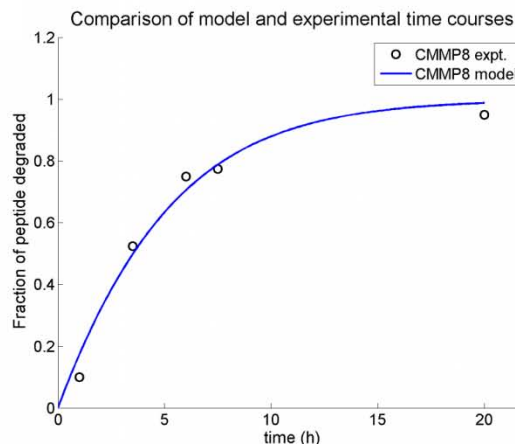
α1: NH$_2$-(GPO)$_3$GPQ**GIA**GQRGVVGCGG-OH

B)

Figure A2.2. Conformational selection and degradation of Trimer A by CMMP8. A) Sequence of the type I collagen-like model peptide Trimer A from Ottl *et al*. used in prior degradation experiments with CMMP8 (Ottl, Gabriel et al. 2000). The type I collagen sequence surrounding the collagenase cleavage site is shown in green, while the triplets containing the scissile bonds are indicated in red. The disulfide linkages of the C-terminus cystine knot region are indicated with the black vertical lines. B) Comparison of model and experimental degradation time courses over 20 h for best fit values $K_{eq} = 60$, $\left(K_{bind}^{VC}\right)_{CMMP8} = 0.7 \times 10^6 \text{ M}^{-1}$, and $k_{cat} = 1.1 \text{ s}^{-1}$. Experimental data are indicated by the black circles; the model time course is indicated by the blue line.
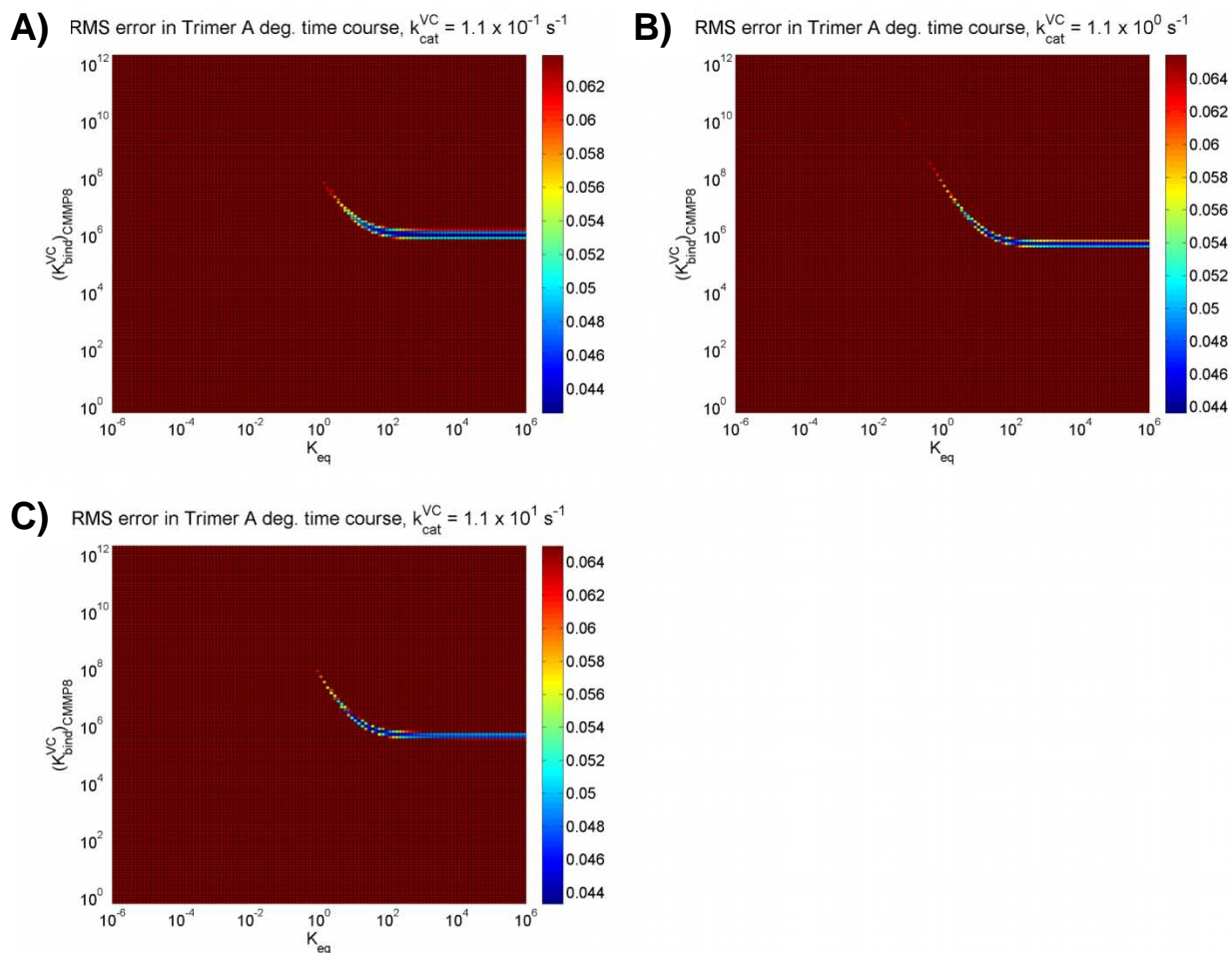
Figure A2.3. Root mean square error (RMSE) of the conformational selection model for Trimer A model peptide degradation time course. For each pair of values (K_{eq}, K_{bind}^{VC}), the fraction of model peptide degraded by CMMP8 was computed at 1, 3.5, 6, 7.5, and 20 h. The root mean square difference was then calculated using these fractions and the experimentally measured amount of degradation at the same time points. The lowest RMSEs (shown in dark blue) indicate (K_{eq}, K_{bind}^{VC}) pairs that provide good fits to the experimental degradation data. Model fitting was done for three values of k_{cat} spanning two orders of magnitude: A) 0.11 s^{-1}, B) 1.1 s^{-1}, and C) 11 s^{-1}. Over this range of k_{cat}, the best fit values of $(K_{bind}^{VC})_{CMMP8}$ range from 0.7-1.1 x 10^6 M^{-1}.

A2.2.2 Determining K_{eq} for type I collagen at room temperature

To obtain an estimate for K_{eq} at room temperature we quantified the extent of collagen degradation over time using the data shown in Figure 3.5 A. We then fit the reaction scheme shown in Figure A2.1 to these data to obtain an estimate for K_{eq}. We note that the previously discussed studies on a heterotrimeric type I collagen-like peptide, which contains the collagenase scissile bond, find that $\left(K_{bind}^{VC}\right)_{CMMP8} = 0.7\text{-}1.1 \times 10^6$ M^{-1}. As this peptide is a model for the collagenase cleavage site, we used this range of $\left(K_{bind}^{VC}\right)_{CMMP8}$ in our numerical calculations of the model for type I collagen.

We again compute the RMSE between the degradation time courses obtained with the model using many different values of K_{eq} and compare these results to the experimentally determined degradation time course. Varying k_{cat} of CMMP8 over two orders of magnitude did not change the best fit values for K_{eq} and causes the minimum RMSE to again vary by only 2% (Figure A2.5). The best fits between the model and experiment are achieved when $K_{eq} = 1.7\text{-}2.1 \times 10^{-3}$ (Figure A2.4 A), and the corresponding degradation plot obtained from the model agrees well with experiment (Figure A2.4 B). This value for K_{eq} suggests that the folded triple-helical native state is more favorable at room temperature.
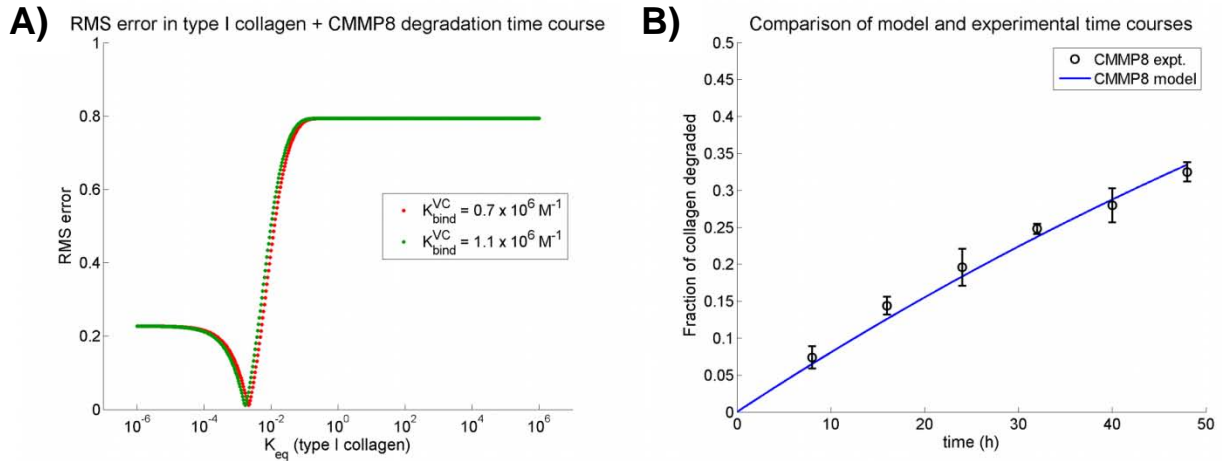
Figure A2.4. Conformational selection and degradation of type I collagen by CMMP8. A) Root mean square error (RMSE) of the conformational selection model over the time course for degradation experiments utilizing type I collagen and CMMP8. For each value of K_{eq}, the fraction of degraded type I collagen is computed at 8, 16, 24, 32, 40, and 48 h. The root mean square difference is then calculated using these fractions and the experimentally measured fraction of degraded collagen. RMSE curves are computed using the lower (red) and upper (green) bounds of $\left(K_{bind}^{VC}\right)_{CMMP8}$. B) A comparison of model and experimental degradation time courses over 48 h for best fit values $K_{eq} = 2.1$ x 10^{-3}, $\left(K_{bind}^{VC}\right)_{CMMP8} = 0.7$ x 10^{6} M^{-1}, and $k_{cat} = 1.1$ s^{-1}. Experimental data are indicated by the black circles with error bars; the model time course is indicated by the blue line.
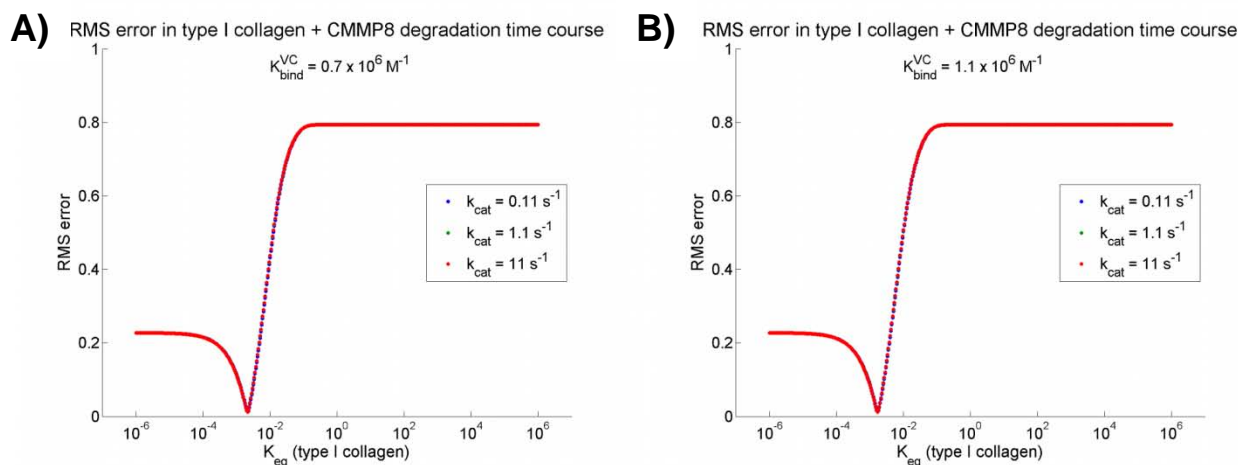
Figure A2.5. Root mean square error (RMSE) of conformational selection model for type I collagen degradation time course. For each value of K_{eq}, the fraction of type I collagen degraded by CMMP8 is computed 8, 16, 24, 32, 40, and 48 h. The root mean square difference is then calculated using these fractions and the experimentally measured amount of degradation at the same time points. The lowest RMSEs indicate K_{eq} values that provide good fits to the experimental degradation data. Model fitting is done for three values of k_{cat} spanning two orders of magnitude (0.11 - 11 s^{-1}) and for A) the lower bound (0.7 x 10^6 M^{-1}) and B) the upper bound (1.1 x 10^6 M^{-1}) of $\left(K_{bind}^{VC}\right)_{CMMP8}$. As the RMSE curves for each value of k_{cat} have the same minima, this demonstrates that varying k_{cat} does not affect the best fit values for K_{eq}.

A2.2.3 Determining K_{bind}^{VC} for CMMP1

Despite the fact that CMMP1 has a higher specific activity than CMMP8 under our experimental conditions (as measured against a thiopeptolide substrate, see Methods in Chapter 3), it degrades collagen with considerably lower efficiency than CMMP8 (Figure 3.5 A and B). To decipher the physical basis for this reduced activity, we apply the reaction scheme shown in Figure A2.1 to the CMMP1 degradation experiments.

Given that K_{eq} is an inherent property of type I collagen at a given temperature, and both CMMP1 and CMMP8 collagen degradation experiments were performed at room temperature, K_{eq} is the same for both reactions. Thus, we use our degradation model to determine the range of $\left(K_{bind}^{VC}\right)_{CMMP1}$. As k_{cat} corresponds to the catalytic rate once the protein has bound the vulnerable state (that is unfolded in the region of the scissile bond), it is estimated using previously determined k_{cat} values for MMP1 degrading gelatin from type I collagen (Welgus, Jeffrey et al. 1982).

With K_{eq} = 1.7-2.1 x 10^{-3}, the best fits are found for $\left(K_{bind}^{VC}\right)_{CMMP1}$ = 0.9-1.3 x 10^{4} M^{-1} (Figure A2.6 A and B). These data suggest that CMMP1's binding affinity for the vulnerable state of collagen is nearly two orders of magnitude lower than that of CMMP8. Expressed as a difference of binding free energies, $\Delta\Delta G_{BIND} = (\Delta G)_{CMMP1} - (\Delta G)_{CMMP8}$, the binding of CMMP1 to type I collagen is approximately 2.5 kcal/mol less favorable than the binding of CMMP8 to type I collagen.
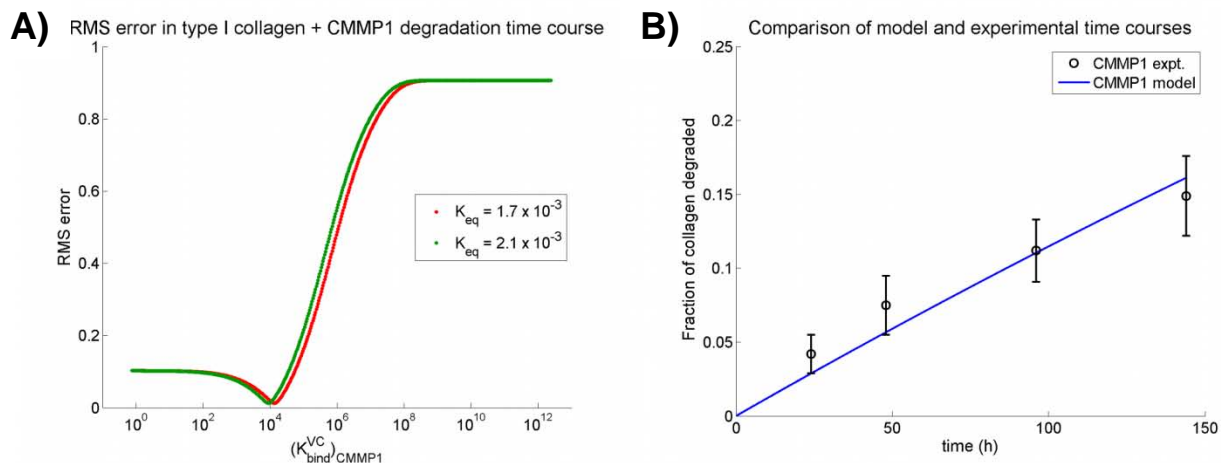
Figure A2.6. Conformational selection and degradation of type I collagen by CMMP1. A) RMSE of the conformational selection model over the time course for degradation experiments utilizing type I collagen and CMMP1. For each value of $\left(K_{bind}^{VC}\right)_{CMMP1}$, the fraction of degraded type I collagen is computed at 24, 48, 96, and 144 h. The root mean square difference is then calculated using these fractions and the experimentally measured fraction of degraded collagen. RMSE curves are computed using the lower (red) and upper (green) bounds of K_{eq}. B) A comparison of model and experimental degradation time courses over 48 h for best fit values K_{eq} = 2.1 x 10^{-3}, $\left(K_{bind}^{VC}\right)_{CMMP1}$ = 0.9 x 10^4 M^{-1}, and k_{cat} = 0.11 s^{-1}. Experimental data are indicated by the black circles with error bars; the model time course is indicated by the blue line.

A2.3 Discussion

Our results in Chapter 3 unambiguously demonstrate that collagenolysis *in vitro* can occur with MMP deletion mutants that contain only the catalytic domain, and that cleavage occurs at the unique cleavage site recognized by full length collagenases. This indicates that the hemopexin-like domain is not required for peptide bond hydrolysis and enzyme specificity *in vitro* at room temperature. In light of these observations we interpret our results using a conformational selection model where thermal fluctuations at the unique cleavage site cause the protein to sample partially unfolded vulnerable states that can then be recognized and cleaved by collagenases (Nerenberg, Salsas-Escat et al. 2008). This formalism allows us to estimate the relative amounts of native and vulnerable states at room temperature from an analysis of type I collagen degradation data. Estimates of K_{eq} for a small heterotrimeric peptide modeling the collagenase cleavage site in type I collagen suggest that the vulnerable state dominates at room temperature for this peptide - a finding in agreement with the measured melting temperature of this peptide (Ottl, Gabriel et al. 2000). For type I collagen at room temperature the calculated K_{eq} of 1.7-2.1 x 10^{-3} corresponds to a free energy difference of ~3.5 kcal/mol between the two states, where the folded triple-helical state is the most stable. This free energy difference corresponds to breaking 2-4 favorable hydrogen bonds, a finding in agreement with a previously proposed structure of the type I collagen vulnerable state (Nerenberg and Stultz 2008). Interestingly, this estimate for the relative amounts of vulnerable states for both the scissile bond-containing heterotrimeric peptide and type I collagen in solution are obtained from an analysis of the degradation data alone. That is, although the model itself does not explicitly contain information about the temperature at which the experiments are performed, it correctly

predicts that the degradation experiments are performed at a relatively high temperature for the heterotrimeric peptide and a relatively low temperature for collagen.

Previous experiments that incubated type I collagen with deletion mutants corresponding to the catalytic domain MMPs failed to discover any degradation products, thereby suggesting that collagen does not adopt partially unfolded conformations at room temperature (Schnierer, Kleine et al. 1993; Chung, Dinakarpandian et al. 2004). However, as our work demonstrates, at temperatures well below collagen's melting temperature, the relative concentration of partially unfolded conformers will be small, and consequently the rate of collagen degradation will be low (Nerenberg, Salsas-Escat et al. 2008). Under these conditions appreciable amounts of degradation will only be observed when relatively long incubation times and high enzyme concentration are used. More precisely, we find that two days of incubation time with CMMP8 and six days with CMMP1 are required to see evidence of collagen degradation.

Additionally, our results offer an explanation for the different type I collagen efficiencies of CMMP1 and CMMP8. While the catalytic rates for the two enzymes are likely different, our data suggest that variations in k_{cat} are insufficient to explain the differences in the overall degradation rates. On the other hand, we find that the binding of CMMP1 to the vulnerable state of type I collagen is 2.5 kcal/mol less favorable than the binding of CMMP8. This free energy difference corresponds to relatively small differences in the bound structures themselves; e.g., breaking 1-2 favorable hydrogen bonds could easily explain a difference of this magnitude. In addition, the notion that CMMP1 has a weaker binding affinity for vulnerable type I collagen chains than CMMP8 is consistent with our observation that CMMP1 degrades denatured type I collagen (gelatin) less efficiently than CMMP8 (Figure 3.6 C). Together these data suggest that

subtle changes in binding between similar homologous enzymes can lead to significant differences in the overall reaction kinetics.

A2.4 Methods

A2.4.1 Densitometry analysis

Densitometry was performed with a Kodak Gel Logic 100 Imaging System. Bands are imaged using an automatic lane and band fitting method (Kodak Molecular Imaging Software v4.0.0). The percentage of type I collagen degradation by CMMP8 or CMMP1 is measured by dividing the sum of the net intensities of the γ_{deg}, β_{deg}, ¼ and ¾ bands by the sum of the total net intensity of all the bands corresponding to total collagen in a given lane (γ, γ_{deg}, β, β_{deg}, $\alpha1$, $\alpha2$, and the ¼ and ¾ bands). γ and β bands correspond to N-terminally crosslinked collagen molecules (Veis and Anesey 1965; French, Mookhtiar et al. 1987; Shigemura, Ando et al. 2004; Morimoto, Kawabata et al. 2009). The γ bands correspond to a trimer of chains (Shigemura, Ando et al. 2004; Morimoto, Kawabata et al. 2009). The β bands correspond to a dimer with two $\alpha1$ chains ($\beta11$) or one $\alpha1$ and one $\alpha2$ chains ($\beta12$), with only one crosslink (French, Mookhtiar et al. 1987; Shigemura, Ando et al. 2004; Morimoto, Kawabata et al. 2009). These cannot be resolved using the 4-12% gradient gels, and are imaged together. Data are presented as the mean and standard deviation over three independent gels per experiment.

A2.4.2 Numerical simulations using a conformational selection model

We develop a reaction scheme based on a conformational selection model to determine whether our experimental observations can be explained using a formalism that assumes collagen can sample partially unfolded states in solution (Nerenberg, Salsas-Escat et al. 2008). In this scheme collagen exists in an equilibrium between native (*N*) and partially unfolded, or

187

vulnerable (V), states. The associated equilibrium constant is denoted by K_{eq}. The catalytic domain of MMPs (C) interacts in a non-specific manner with the native state with binding constant K_{bind}^{NC}, yielding the $N \cdot C$ complex and C can also bind to the vulnerable state, V, with binding constant K_{bind}^{VC}, forming the $V \cdot C$ complex. Binding of C to the native state, N, does not lead to collagenolysis while the $V \cdot C$ complex is degraded with catalytic rate k_{cat}.

The reaction scheme naturally leads to a set of ordinary differential equations (ODEs), with two rate constants corresponding to each equilibrium/binding constant (i.e., $K_{eq} = \frac{k_1}{k_2}$, $K_{bind}^{NC} = \frac{k_{on}^{NC}}{k_{off}^{NC}}$, and $K_{bind}^{VC} = \frac{k_{on}^{VC}}{k_{off}^{VC}}$). The set of ODEs that describes the time evolution of each species shown is given by:

$$\frac{d[N]}{dt} = -k_1[N] + k_2[V] - k_{on}^{NC}[N][C] + k_{off}^{NC}[N \cdot C] \tag{A2.1}$$

$$\frac{d[V]}{dt} = -k_2[V] + k_1[N] - k_{on}^{VC}[V][C] + k_{off}^{VC}[V \cdot C] \tag{A2.2}$$

$$\frac{d[N \cdot C]}{dt} = -k_{off}^{NC}[N \cdot C] + k_{on}^{NC}[N][C] \tag{A2.3}$$

$$\frac{d[V \cdot C]}{dt} = -k_{off}^{VC}[V \cdot C] + k_{on}^{VC}[V][C] - k_{cat}[V \cdot C] \tag{A2.4}$$

$$\frac{d[P]}{dt} = k_{cat}[V \cdot C] \tag{A2.5}$$

$$\frac{d[C]}{dt} = -k_{on}^{NC}[N][C] + k_{off}^{NC}[N \cdot C] - k_{on}^{VC}[V][C] + k_{on}^{VC}[V][C] \tag{A2.6}$$
$$+ k_{cat}[V \cdot C]$$

where P denotes the products of collagenolysis.

We use previously determined rate constants for MMP1 binding to a heterotrimeric collagen-like peptide that contains the type I collagen collagenase cleavage site at a temperature substantially less than its melting temperature (i.e., we expect this peptide to be in a well-folded triple-helical conformation at this temperature) to estimate the native state binding rate constants, k_{on}^{NC} and k_{off}^{NC} ($k_{on}^{NC} = 1.56 \ 10^3$ M^{-1} s^{-1}, $k_{off}^{NC} = 5.35 \ 10^{-3}$ s^{-1}) (Ottl, Gabriel et al. 2000). The catalytic rate constant of the vulnerable state, k_{cat}, for CMMP1 was estimated from the experimentally measured catalytic rates of MMP1 degrading gelatin (Welgus, Jeffrey et al. 1982). Because similar data does not exist for CMMP8, we set lower and upper bounds for the catalytic rate based on previously published data. The lower bound of k_{cat} for CMMP8 is set equal to that of CMMP1, as it is known that MMP8 has a greater catalytic rate than MMP1 for the same substrate and at the same temperature (Welgus, Jeffrey et al. 1982; Gioia, Fasciglione et al. 2002). Upper bounds for the catalytic rate of CMMP8 are obtained from experimentally measured rate constants of FMMP8 degrading linear collagen-like peptides (Netzel-Arnett, Fields et al. 1991). This approach yields a range of k_{cat} for CMMP8 spanning 0.11-11.1 s^{-1}.

Finding the concentration of each reactant species as function of time requires knowledge of each of the remaining rate constants $k_1, k_2, k_{on}^{VC}, and \ k_{off}^{VC}$. Previous studies with a related reaction scheme demonstrated that the model is not sensitive to the specific choice of the individual rate constants when one is interested in the reactant concentrations over long time periods (i.e., a time that is long with respect to the individual rate constants themselves, see Appendix 1) (Nerenberg, Salsas-Escat et al. 2008). In this case, the system behavior is dependent on the ratio of the rate constants, or equivalently the associated equilibrium binding constants K_{eq} and K_{bind}^{VC}. Since our experiments occur over days, which is very long with respect to the

folding/unfolding and binding processes of interest, we determine the behavior of the system as a function of K_{eq} and K_{bind}^{VC}.

To convert a given K_{eq} into the appropriate transition rate constants, k_1 and k_2, for the solution of the ODEs, we use initial rate constant values (k_1^{init} and k_2^{init}, set equal to 10^6 s^{-1}) that are consistent with the folding times for a number of small proteins; e.g., (Wittung-Stafshede, Lee et al. 1999; Kubelka, Chiu et al. 2006). These rate constants are then multiplied them by a scale factor α according to:

$$k_1 = \alpha^{\frac{1}{2}} k_1^{init} \ \ and \ \ k_2 = \alpha^{-\frac{1}{2}} k_2^{init} \tag{A2.7}$$

$$K_{eq} = \frac{k_1}{k_2} = \alpha \frac{k_1^{init}}{k_2^{init}} \tag{A2.8}$$

In this way, the appropriate rate constants can be determined for any value of K_{eq}. We use a similar approach to determine the vulnerable state binding rate constants, k_{on}^{VC} and k_{off}^{VC}, from a given K_{bind}^{VC}:

$$k_{on}^{VC} = \beta^{\frac{1}{2}} k_{on}^{VC,init} \ \ and \ \ k_{off}^{VC} = \beta^{-\frac{1}{2}} k_{off}^{VC,init} \tag{A2.9}$$

$$K_{bind}^{VC} = \frac{k_{on}^{VC}}{k_{off}^{VC}} = \beta \frac{k_{on}^{VC,init}}{k_{off}^{VC,init}} \tag{A2.10}$$

The initial rate constants for binding the vulnerable state, $k_{on}^{VC,init}$ and $k_{off}^{VC,init}$, are set equal to experimentally measured rates for MMP1 binding to a heterotrimeric collagen-like peptide that contains the type I collagen collagenase cleavage site at a temperature substantially greater than its melting temperature (i.e., we expect the triple helix to be mostly unfolded and

thus the scissile bonds to be solvent-exposed at this temperature) ($k_{on}^{VC,init}$ =4.5 10^2 M^{-1} s^{-1}, $k_{off}^{VC,init}$ =5.94 10^{-3} s^{-1}) (Ottl, Gabriel et al. 2000). With these rate constants, Equations (A2.1)-(A2.6) are then solved using the ode15s solver in MATLAB 2008a (© MathWorks) using the durations and initial concentrations of collagen and enzyme specific to the relevant degradation experiment(s).

Solutions to the model yield the concentration of collagen degradation products as a function of time, $P_{model}(t)$, while the degradation experiments yield the fraction of peptide/collagen degraded as a function of time, $F_{exp}(t)$. To relate the calculated concentrations to the measured degraded fractions, we normalize the concentration of degradation products to the initial substrate concentration: $F_{model}(t) = P_{model}(t)/S_{init}$, where S_{init} is the initial substrate concentration. We then calculate the root mean square error (RMSE) between experimental and model time courses according to:

$$RMSE = \sqrt{n^{-1} \sum_{t=1}^{n} \left(F_{model}(t) - F_{exp}(t) \right)^2} \qquad (A2.11)$$

where n is the number of experimental time points.

A2.4.3 Derivation of the effective equilibrium constant in the presence of full length enzyme

We first assume that the full length enzyme in Figure C.1 is catalytically inactive (i.e., degradation does not occur from the $V \cdot F$ complex). In this case, the system will reach an equilibrium defined by:

$$\frac{d[N]}{dt} = \frac{d[V]}{dt} = \frac{d[N \cdot H]}{dt} = \frac{d[V \cdot H]}{dt} = \frac{d[V \cdot F]}{dt} = \frac{d[E]}{dt} = 0 \qquad \text{(A2.12)}$$

where $[E]$ is the free enzyme concentration and the remaining variables are the species enumerated in Figure C.1. We wish to then calculate the equilibrium ratio of all vulnerable states to all native states:

$$K_{eq}^{eff} = \frac{[V] + [V \cdot H] + [V \cdot F]}{[N] + [N \cdot H]} \qquad \text{(A2.13)}$$

As an intermediate step to obtaining K_{eq}^{eff}, we first note that:

$$K_{eq} = \frac{[V] + [V \cdot H]}{[N] + [N \cdot H]} \qquad \text{(A2.14)}$$

Proof:

$$K_{eq} = \frac{[V] + [V \cdot H]}{[N] + [N \cdot H]} = \frac{[V]}{[N] + [N \cdot H]} + \frac{[V \cdot H]}{[N] + [N \cdot H]} = \frac{1}{\frac{[N] + [N \cdot H]}{[V]}} + \frac{1}{\frac{[N] + [N \cdot H]}{[V \cdot H]}}$$

$$= \frac{1}{\frac{[N]}{[V]} + \frac{[N \cdot H]}{[V]}} + \frac{1}{\frac{[N]}{[V \cdot H]} + \frac{[N \cdot H]}{[V \cdot H]}}$$

Using the fact that $K_{eq} = \frac{[V]}{[N]} = \frac{[V \cdot H]}{[N \cdot H]}$, we have:

$$\frac{[V] + [V \cdot H]}{[N] + [N \cdot H]} = \frac{1}{\frac{1}{K_{eq}} + \frac{[N \cdot H]}{[V]}} + \frac{1}{\frac{[N]}{[V \cdot H]} + \frac{1}{K_{eq}}} = \frac{K_{eq}}{1 + \frac{[N \cdot H]}{[V]} K_{eq}} + \frac{K_{eq}}{\frac{[N]}{[V \cdot H]} K_{eq} + 1}$$

$$= \frac{K_{eq}}{1 + \frac{[N \cdot H]}{[V]} \frac{[V]}{[N]}} + \frac{K_{eq}}{\frac{[N]}{[V \cdot H]} \frac{[V \cdot H]}{[N \cdot H]} + 1} = \frac{K_{eq}}{1 + \frac{[N \cdot H]}{[N]}} + \frac{K_{eq}}{\frac{[N]}{[N \cdot H]} + 1}$$

$$= \frac{K_{eq}[N]}{[N\} + [N \cdot H]} + \frac{K_{eq}[N \cdot H]}{[N] + [N \cdot H]} = K_{eq}\left(\frac{[N] + [N \cdot H]}{[N] + [N \cdot H]}\right) = K_{eq}$$

Using Equation (A2.14) and the fact that $K_{bind}^{H} = \frac{[N \cdot H]}{[N][E]} = \frac{[V \cdot H]}{[V][E]}$ and $K_{bind}^{V_H C} = \frac{[V \cdot F]}{[V \cdot H]}$ we can calculate K_{eq}^{eff} from Equation (A2.13) as follows:

$$K_{eq}^{eff} = \frac{[V] + [V \cdot H] + [V \cdot F]}{[N] + [N \cdot H]} = \frac{[V] + [V \cdot H]}{[N] + [N \cdot H]} + \frac{[V \cdot F]}{[N] + [N \cdot H]} = K_{eq} + \frac{[V \cdot F]}{[N] + [N \cdot H]}$$

$$= K_{eq} + \frac{[V \cdot H]K_{bind}^{V_H C}}{[N] + [N \cdot H]} = K_{eq} + \frac{K_{bind}^{V_H C}}{\frac{[N] + [N \cdot H]}{[V \cdot H]}} = K_{eq} + \frac{K_{bind}^{V_H C}}{\frac{[N]}{[V \cdot H]} + \frac{[N \cdot H]}{[V \cdot H]}}$$

$$= K_{eq} + \frac{K_{bind}^{V_H C}}{\frac{[N]}{[V \cdot H]} + \frac{1}{K_{eq}}} = K_{eq} + \frac{K_{bind}^{V_H C}K_{eq}}{\frac{[N]K_{eq}}{[V \cdot H]} + 1} = K_{eq} + \frac{K_{bind}^{V_H C}K_{eq}}{\frac{[N][V \cdot H]}{[V \cdot H][N \cdot H]} + 1}$$

$$= K_{eq} + \frac{K_{bind}^{V_H C}K_{eq}}{\frac{[N]}{[N \cdot H]} + 1} = K_{eq} + \frac{K_{bind}^{V_H C}K_{eq}}{\frac{1}{K_{bind}^{H}[E]} + 1} = K_{eq} + \frac{K_{bind}^{V_H C}K_{eq}K_{bind}^{H}[E]}{K_{bind}^{H}[E] + 1}$$

$$= K_{eq}\left(1 + \frac{K_{bind}^{V_H C}K_{bind}^{H}[E]}{K_{bind}^{H}[E] + 1}\right) = K_{eq}\left(1 + \left(\frac{K_{bind}^{H}[E]}{K_{bind}^{H}[E] + 1}\right)K_{bind}^{V_H C}\right)$$

A2.5 Bibliography Appendix 2

Chung, L., D. Dinakarpandian, et al. (2004). "Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis." Embo. J. **23**(15): 3020-30.

Clark, I. M. and T. E. Cawston (1989). "Fragments of human fibroblast collagenase. Purification and characterization." Biochem J **263**(1): 201-6.

French, M. F., K. A. Mookhtiar, et al. (1987). "Limited proteolysis of type I collagen at hyperreactive sites by class I and II Clostridium histolyticum collagenases: complementary digestion patterns." Biochemistry. **26**(3): 681-7.

Gioia, M., G. F. Fasciglione, et al. (2002). "Modulation of the catalytic activity of neutrophil collagenase MMP-8 on bovine collagen I. Role of the activation cleavage and of the hemopexin-like domain." J. Biol. Chem. **277**(26): 23123-30.

Kubelka, J., T. K. Chiu, et al. (2006). "Sub-microsecond protein folding." Journal of Molecular Biology **359**(3): 546-553.

Morimoto, K., K. Kawabata, et al. (2009). "Characterization of type I collagen fibril formation using thioflavin T fluorescent dye." J. Biochem. **145**(5): 677-684.

Nerenberg, P. S., R. Salsas-Escat, et al. (2008). "Do collagenases unwind triple-helical collagen before peptide bond hydrolysis? Reinterpreting experimental observations with mathematical models." Proteins. **70**(4): 1154-61.

Nerenberg, P. S. and C. M. Stultz (2008). "Differential unfolding of alpha1 and alpha2 chains in type I collagen and collagenolysis." J. Mol. Biol. **382**(1): 246-56.

Netzel-Arnett, S., G. B. Fields, et al. (1991). "Sequence specificities of human fibroblast and neutrophil collagenases." J. Biol. Chem. **266**(11): 6747-55.

Ottl, J., D. Gabriel, et al. (2000). "Recognition and catabolism of synthetic heterotrimeric collagen peptides by matrix metalloproteinases." Chem. Biol. **7**(2): 119-32.

Schnierer, S., T. Kleine, et al. (1993). "The recombinant catalytic domain of human neutrophil collagenase lacks type I collagen substrate specificity." Biochem Biophys Res Commun **191**(2): 319-26.

Schnierer, S., T. Kleine, et al. (1993). "The recombinant catalytic domain of human neutrophil collagenase lacks type I collagen substrate specificity." Biochem. Biophys. Res. Commun. **191**(2): 319-26.

Shigemura, Y., M. Ando, et al. (2004). "Possible degradation of type I collagen in relation to yellowtail muscle softening during chilled storage." Fisheries Science. **70**(4): 703-709.

Stultz, C. M. (2002). "Localized unfolding of collagen explains collagenase cleavage near imino-poor sites." <u>J. Mol. Biol.</u> **319**(5): 997-1003.

Veis, A. and J. Anesey (1965). "Modes of intermolecular cross-linking in mature insoluble collagen." <u>J. Biol. Chem.</u> **240**(10): 3899-3908.

Welgus, H. G., J. J. Jeffrey, et al. (1982). "The gelatinolytic activity of human skin fibroblast collagenase." <u>J. Biol. Chem.</u> **257**(19): 11534-9.

Wittung-Stafshede, P., J. C. Lee, et al. (1999). "Cytochrome b(562) folding triggered by electron transfer: Approaching the speed limit for formation of a four-helix-bundle protein." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **96**(12): 6587-6590.