The Mechanism of Collagenolysis: A Substrate-centric View

by

Paul S. Nerenberg

B.S. Physics
Johns Hopkins University, 2004

Submitted to the Department of Physics
In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2010

Signature of Author . . . .
Department of Physics
December 16, 2009

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

Certified by . . . .
Leonid A. Mirny
Associate Professor of Health Sciences and Technology and Physics
Co-supervisor

Accepted by . . . .
Krishna Rajagopal
Professor of Physics
Associate Department Head for Education

The Mechanism of Collagenolysis: A Substrate-centric View

by

Paul S. Nerenberg

Submitted to the Department of Physics
on December 16, 2009, in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Physics

Abstract

Collagenolysis (collagen degradation) is a physiological process involved in normal tissue maintenance, but excessive collagenolysis has been associated with the progression of cancer metastasis, atherosclerosis, and other diseases. Despite considerable efforts to understand the steps involved, the exact mechanism of collagenolysis remains unknown. One proposed mechanism suggests that the enzymes that degrade collagen, collagenases, physically unwind the triple-helical structure of collagen to gain access to the peptide bond that is cleaved. This unwinding mechanism would in principle have large energetic requirements, but neither ATP nor other energy-rich molecules are necessary for collagenolysis. An alternative mechanism is that collagen exists in multiple states, some featuring structures that are unfolded in the vicinity of the collagenase cleavage site, and that collagenases preferentially bind to and stabilize these partially unfolded structures before degradation occurs.

The focus of this work is to investigate this alternative mechanism, particularly as it pertains to the conformational ensemble of collagen, using both experimental and computational methods. In particular, this work concentrates on: (1) devising a reaction scheme for this mechanism and verifying that it can explain existing experimental observations; (2) generating a structural model of the type I collagen collagenase cleavage site using molecular dynamics simulations; (3) validating this structural model by performing degradation experiments and analyzing them in light of the reaction scheme described in (1); and (4) validating a similar structural model for type III collagen by designing and characterizing a self-assembling collagen-like model peptide that includes the sequence of the collagenase cleavage site. Together, these data present a detailed and comprehensive analysis of the conformational ensemble of collagen near the collagenase cleavage site and its role in the molecular mechanism of collagen degradation.

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

Co-supervisor: Leonid A. Mirny
Title: Associate Professor of Health Sciences and Technology and Physics

Acknowledgements

"Art is I, science is we." – Claude Bernard

Much like it takes a village to raise a child, I believe that for most of us it takes the help and support of many people – friends, colleagues, mentors, and family – to get a PhD. Because I only get to write one of these, please forgive me in advance for acknowledging the many people who contributed to my success in one way or another.

I would first like to thank my thesis committee: Prof. Collin Stultz (my thesis advisor), Prof. Leonid Mirny (my co-advisor), Prof. Alexander van Oudenaarden, and Prof. Mehran Kardar. Not only have they helped shape my work through our discussions, but they also have served in many other capacities throughout my graduate career including academic advisor, oral exam committee members, and perhaps most importantly, my teachers.

As a graduate teaching assistant, I have had the good fortune to work with many members of the MIT faculty. I must first thank Dr. Michael Bove of the MIT Media Lab. Mike graciously offered me a position as the TA for the Media Lab Freshman Program after I had left my previous lab, thus endowing me with the long-term funding that ultimately enabled me to find a new home in Collin's lab. Equally, if not more importantly, it was through this position that I came to realize just how much I enjoyed teaching undergraduate students. With that one small decision, Mike changed my life greatly for the better, and I am forever grateful to him for it. Next, I would like to thank Prof. Alan Guth and Prof. Rob Simcoe, with whom I TAed 8.01 and 8.02 for the first time. I would also like to acknowledge Andy Neely, who spent countless hours of his time helping myself and the rest of the teaching staff prepare and run our classes. His knowledge and insights were truly invaluable, and his enthusiasm for the students was contagious. Finally, I must thank Dr. Peter Dourmashkin and Prof. Eric Hudson, with whom I TAed 8.01 and 8.02 for the second time. In addition to being truly dedicated teachers, Peter and Eric have selflessly served as my mentors and supporters throughout the entire time I have known them and for that I feel both extremely fortunate and deeply thankful.

At this point, I must dedicate some verbiage to my thesis advisor, Prof. Collin Stultz. I didn't get off to the best start with Collin. When I first e-mailed him about joining the group, his reply automatically went to my spam folder (and was subsequently deleted without me ever seeing it). Luckily, he replied to me again when I sent him a second e-mail…and thankfully his message escaped the abyss of the spam folder! I must credit Collin for taking a rather large leap of faith in taking me on as a student. When I first joined the lab, I still had yet to pass my second general exam, I was TAing full-time, and I had only the barest of backgrounds in biophysics. Despite all that, Collin saw the potential in me and was confident in my abilities (even when I was not). The rest is, as they say, history.

As a mentor, Collin has continuously challenged me and allowed me to explore many topics in my research (probably to his occasional consternation). With his guidance, I have learned how to critically evaluate both my own work and the work of others, as well as ask the "right" scientific questions and propose ways to answer those questions. In addition, he has given me

the opportunity to gain valuable experience being a mentor myself and in general has advocated the development of skills that will be critical in becoming an independent scientist after graduate school.

Beyond his talents as a mentor, Collin also has a knack for assembling an exceptional group of people around him. Being a part of Collin's group has made going to work a real pleasure every day and undoubtedly helped me remain intensely passionate about science. My fellow members of the Stultz lab have become like a second family to me, and I will miss them dearly. It is my sincere hope that we never lose touch, regardless of where we may end up.

I must first and especially thank Ramon Salsas-Escat, with whom I have worked side-by-side for more than three years on our favorite subject, collagen, with only occasional digressions into our second favorite subject, pork products. I think I would probably have driven most people insane in that amount of time, but Ramon stoically endured the ranting and raving of a crazy kid from California, and together we not only produced meaningful scientific insights, but also became good friends.

Next, I would like to thank Dr. Austin Huang, who sat either behind or next to me for the better part of my time here (until he graduated) and taught me nearly everything that I know about CHARMM and MATLAB. I know for a fact that Austin is a Zen master because if he weren't, my unending questions during the first year I was in the lab would surely have driven him crazy. Austin and I also shared many conversations about our two loves – science and music – over the course of hundreds of lunches, and I very much enjoyed our time together.

I must thank Christine Phillips, who exposed me to an entirely different field of research and with whom I hope to collaborate long into the future. Christine poked and prodded me with many questions in our work together, forcing me to become more knowledgeable in many techniques along the way, and ultimately made me a better scientist. On a more personal note, we have shared in the many ups and downs of grad school, particularly in having our significant others on the opposite side of the country as we finished up, and I am especially thankful for her friendship during this time.

I would like to thank Thomas Gurry, a visiting master's student who spent this past summer with our group working on a project that I had only vague sketches of in my mind and nevertheless managed to generate valuable insights through his quick learning and ingenuity. I am also grateful to Dr. Chris Schubert, Elaine Gee, Charles Fisher, Orly Ullman, and Joyatee Sarker for our many discussions (both serious and not-so-serious) and fun times together. And although she is a recent arrival, I would be remiss in not thanking Dr. Sophie Walker, who has already become a good friend and source of invaluable life coaching during my thesis writing process.

Finally, I must thank the undergraduate students with whom I interacted and mentored in my time in the Stultz lab: Veena Venkatachalam, Gadini Delisca, William Wyatt, Madhavi Gavini, Steven Pennybaker, and Anjali Muralidhar. It is through working with all of you that I discovered that mentoring students was another one of my life's passions.

Outside of the lab I have been extremely fortunate to have a fantastic group of friends. I would like to acknowledge the many friends I made as a part of the PGSC, the GSC, the Muddy, and MIT Cycling. As integral parts of this last organization, I must thank Nick Loomis, AJ Schrauth, Zach LaBry, Jason Sears, Ilana Brito, and Eric Edlund for their tough love, good spirits, and being troopers as we shared many long car rides up and down the East Coast over the years. There is no question in my mind that cycling was an indispensible form of therapy throughout my tenure at MIT, and I am very grateful to my friends for getting me involved in both the sport and our team. Somewhat closer to home, I would like to thank a few of my dear friends in the physics department: Alan Hoffman, Adam Kocoloski (and his wife Hillary), Andrew Puckett, Arturo Dominguez (and his wife Carolina), Robyn Sanderson, and Andrew Grier. We have suffered together over countless problem sets and general exams, cheered each other in our successes, and supported each other in our failures. In particular, I have hung out with Arturo, Robyn, and Grier a large chunk of the last five and a half years and am amazed that they still talk to me after all this time.

Next, I must thank my family. Even though we aren't actually related, I consider my old friends Tristan, Adam, and Roni to be a part of my family and would like to thank them for cheering me on and providing a connection to the outside world over the past few years. As for people who I am actually related to, I must thank my parents – Mom, Dad, Dotty, and Joel – for their unending support. Graduate school is in many ways like a marathon, and it was only with their love that I got through the rough patches. They have served as my counselors on countless occasions throughout my time at MIT, even though they were many miles away. They also have been shameless in embarrassing me to stay how proud they are of me, which contrary to what I might say, does actually provide a welcome (and necessary) boost. I must also thank my grandmother (or Nana, as I know her) for serving as an inspiration to me not only in graduate school, but throughout my life.

Last, but certainly not least, I must thank my fiancée and partner in life, Heather. Words will likely not do her justice, but I will try to express my gratitude nonetheless. As young scientists in graduate school, we have experienced many of the same joyous victories and many of the same frustrating defeats, and as partners, we have been each other's cheerleaders, confidants, and caretakers throughout it all. I feel incredibly fortunate to have found someone with whom I can share life's journeys and quite confident in saying that her love and support have been an unshakable foundation upon which my own successes have been built and made all the more meaningful.

To all those who have supported me in my scientific and personal endeavors, I dedicate this work to you.

Contents

List of Figures

Chapter 1

Introduction

The most abundant proteins in the human body are members of the collagen family (Gelse et al. 2003, Brinckmann 2005). Collagens are the primary constituents of the extracellular matrix (ECM), which confers mechanical and biochemical properties to nearly every type of tissue (Gelse et al. 2003). Collagens are synthesized within specific cells and then exported to the extracellular space where they link together with other collagen molecules to form the basis of interstitial matrix or basement membrane (Gelse et al. 2003, Shoulders and Raines 2009). (In basement membrane, it has been shown that collagen can also link with other molecules; e.g. laminin (Kalluri 2003).) As the degradation of the ECM, and more specifically collagen, plays a central role in regular tissue maintenance, as well as the progression of several diseases, such as atherosclerosis, cancer, and arthritis, developing a detailed understanding of the molecular mechanism of collagenolysis is of particular interest (Celentano and Frishman 1997, Barnes and Farndale 1999, Bode et al. 1999, McDonnell et al. 1999, Gelse et al. 2003, Nerenberg et al. 2007, Libby 2008).

1.1 Collagen

At present, there are 29 known distinct types of collagen (Shoulders and Raines 2009). While the sequences and domain organizations of these proteins vary widely, they share the common feature of containing at least one region in which three polypeptide chains, each in a left-handed polyproline II-like helix conformation, supercoil around one another to form a right-handed triple helix linked by interchain hydrogen bonds (Figure 1-1) (Gelse et al. 2003, Brodsky and Persikov 2005, Engel and Bächinger 2005, Shoulders and Raines 2009). In addition, each of the chains is staggered by one residue within the triple helix and all of the peptide bonds are in the *trans* conformation (Brodsky and Persikov 2005, Engel and Bächinger 2005, Shoulders and Raines 2009). Remarkably, the basic structure of collagen was proposed over 50 years ago on

the basis of relatively low resolution X-ray fiber diffraction data (Ramachandran and Kartha 1954, Ramachandran and Kartha 1955, Rich and Crick 1955, Rich and Crick 1961).

For the remainder of this work, we will focus on the fibrillar collagens (types I, II, III, V, XI, XXIV, and XXVII), and in particular on collagen types I and III. Both of these collagens feature chains of approximately 1000 residues in length and have greater than 95% of their sequences located within a single long triple-helical region (Gelse et al. 2003). Type I collagen is a heterotrimer consisting of two α1(I) chains and one α2(I) chain, while type III collagen is a homotrimer consisting of three α1(III) chains (Gelse et al. 2003, Brinckmann 2005).
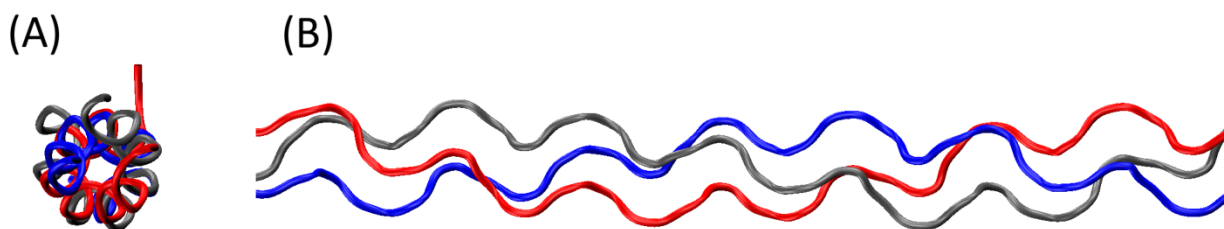
Figure 1-1: Characteristic structure of the collagen triple helix. Views of the backbone atoms of the triple helix along the (A) axial and (B) longitudinal directions. Each polypeptide chain is indicated with a different color. Coordinates taken from PDB entry 1BKV.

1.1.1 The relationship between the sequence and the structure of the collagen triple helix

The collagen triple helix is enabled by a repeating tripeptide sequence motif, Gly-X1-X2, where Gly is glycine and X1 and X2 can be any amino acid (Gelse et al. 2003, Brodsky and Persikov 2005, Engel and Bächinger 2005). This sequence is necessary because the sidechain position of glycine is directed toward the center of the triple helix and any larger sidechain (beyond the proton of glycine) would interfere with the packing of the three chains (Brodsky and Persikov 2005, Engel and Bächinger 2005, Shoulders and Raines 2009). Indeed, one well known class of genetic disorders, osteogenesis imperfecta, typically involves a glycine substitution that leads to a disruption of the triple helix (Myllyharju and Kivirikko 2001). As the sidechains of the X1 and X2 residues are directed outward, this means that the majority of collagen's sidechains are solvent-exposed, regardless of their hydrophobicity (Figure 1-2) (Brodsky and Persikov 2005, Engel and Bächinger 2005, Shoulders and Raines 2009). A sequence analysis of the triple-

helical regions of twelve types of collagen found that proline (Pro or P) occurs in the X1 position 28% of the time and hydroxyproline (Hyp or O), a post-translational modification of proline (see Section 1.1.2), occurs in the X2 position 38% of the time (Ramshaw et al. 1998). In addition, the most common triplet sequence is GPO, which occurs just over 10% of the time (Ramshaw et al. 1998). Nonetheless, the sequences of the collagens are quite diverse, and there are many short stretches in the sequence with little or no imino acid (i.e., proline or hydroxyproline) content (Figure 1-3). In section 1.1.3, we will return to these imino-poor regions and how they pertain to the molecular mechanism of collagenolysis.
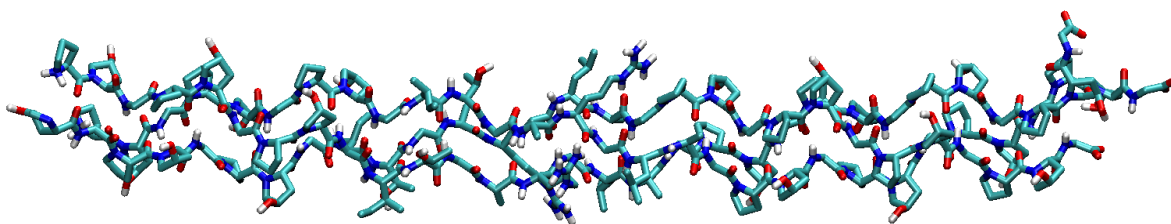
Figure 1-2: Structure of a triple-helical collagen-like model peptide demonstrating that the sidechains of the X1 and X2 position residues are directed outward from the triple helix and into solvent. Coordinates taken from PDB entry 1BKV.

As collagens have a minimum sequence length of several hundred residues and are relatively insoluble, it is worth noting that much of the work done to elucidate the basic biophysical properties of collagen, including the structure and thermodynamics of the triple helix, has been performed using synthetic collagen-like model peptides (Fields and Prockop 1996). In particular, the first high resolution structure of the triple helix was obtained just over 15 years ago with a homotrimeric collagen-like model peptide (Bella et al. 1994). It was through this structure that the interchain hydrogen bonding pattern of the triple helix, in which the amide proton of Gly acts as a donor and the amide oxygen of a residue in the X1 position of a neighboring chain acts as an acceptor, could finally be confirmed (Figure 1-4) (Bella et al. 1994). While collagen-like model peptides have been an invaluable tool in investigating the structure of collagen at the molecular level, one of the still unanswered and important questions in the field is how accurately these relatively short model peptides can represent the larger collagen molecule.

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

```
QLSYGYDEKSTGGISVPGPMGPSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKN
GDDGEAGKPGRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAP
GQMGPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGPRGSEGPQ
GVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQGPGGPPGPKGNSGEP
GAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPK
GPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGAR
GQAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQ
GLPGPAGPPGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDA
GAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK
GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPP
GPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVPGPPGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRP
GEVGPPGPPGPAGEKGSPGADGPAGAPGTPGPQGIAGQRGVVGLPGQRGERGFPGLPGPSGEPGKQGPS
GASGERGPPGPMGPPGLAGPPGESGREGAPGAEGSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGPV
GPAGKSGDRGETGPAGPAGPVGPVGARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSP
GEQGPSGASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPSAG
FDFSFLPQPPQEKAHDGGRYYRA
```

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

```
QYDGKGVGLGPGPMGLMGPRGPPGAAGAPGPQGFQGPAGEPGEPGQTGPAGARGPAGPPGKAGEDGHP
GKPGRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTPGQTGAR
GLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGAVGNAGPAGPAGPRGEV
GLPGLSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAAGATGARGLVGEPGPAGSK
GESGNKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPPGPPGLRGSPGSRGLPGADGRAGVMGPPGSR
GASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGEPGNI
GFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGPPGPQGVQGGKGEQGPAGPPGFQGLPGPS
GPAGEVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGSRGPSGPPGPDGNKGEPGVVGAV
GTAGPSGPSGLPGERGAAGIPGGKGEKGEPGLRGEIGNPGRDGARGAPGAVGAPGPAGATGDRGEAGAA
GPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGAKGPKGENGVVGPTGPVGAAGPAGPN
GPPGPAGSRGDGGPPGMTGFPGAAGRTGPPGPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRTGEVGAV
GPPGFAGEKGPSGEAGTAGPPGTPGPQGLLGAPGILGLPGSRGERGLPGVAGAVGEPGPLGIAGPPGAR
GPPGAVGSPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKGERGYPGNIGPVGAAGAPGPHGPVGPAGKH
GNRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKGEPGEKGPRGLPGLKGHNGLQGLPGIAGHHGDQGAP
GSVGPAGPRGPAGPSGPAGKDGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPGPPG
```

Type II collagen sequence (UniProtKB-SwissProt accession no. P02458)

```
QMAGGFDEKAGGAQLGVMQGPMGPMGPRGPPGPAGAPGPQGFQGNPGEPGEPGVSGPMGPRGPPGPP
GKPGDDGEAGKPGKAGERGPPGPQGARGFPGTPGLPGVKGHRGYPGLDGAKGEAGAPGVKGESGSPGEN
GSPGPMGPRGLPGERGRTGPAGAAGARGNDGQPGPAGPPGPVGPAGGPGFPGAPGAKGEAGPTGARGPE
GAQGPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPGPQGATGPLGPKGQT
GEPGIAGFKGEQGPKGEPGPAGPQGAPGPAGEEGKRGARGEPGGVGPIGPPGERGAPGNRGFPGQDGLA
GPKGAPGERGPSGLAGPKGANGDPGRPGEPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQ
GARGQPGVMGFPGPKGANGEPGKAGEKGLPGAPGLRGLPGKDGETGAAGPPGPAGPAGERGEQGAPGPS
GFQGLPGPPGPPGEGGKPGDQGVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGLPGTPGTDGPK
GASGPAGPPGAQGPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGGRGLTGPIGPPGPAGAN
GEKGEVGPPGPAGSAGARGAPGERGETGPPGPAGFAGPPGADGQPGAKGEQGEAGQKGDAGAPGPQGPS
GAPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPPGPSGKDGPKGARGDSGPP
GRAGEPGLQGPAGPPGEKGEPGDDGPSGAEGPPGPQGLAGQRGIVGLPGQRGERGFPGLPGPSGEPGKQ
GAPGASGDRGPPGPVGPPGLTGPAGEPGREGSPGADGPPGRDGAAGVKGDRGETGAVGAPGAPGPPGSP
GPAGPTGKQGDRGEAGAQGPMGPSGPAGARGIQGPQGPRGDKGEAGEPGERGLKGHRGFTGLQGLPGPP
GPSGDQGASGPAGPSGPRGPPGPVGPSGKDGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP
GPGIDMSAFAGLGPREKGPDPLQYMRA
```

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

```
QYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGPPGAI
GPSGPAGKDGESGRPGRPGERGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGLKGENGLP
GENGAPGPMGPRGAPGERGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLR
GGAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPN
GIPGEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPP
GPSGPRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPP
GPQGLQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGAPGERGPPGLAGAPGLRGGAGPPGPE
GGKGAAGPPGPPGAAGTPGLQGMPGERGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPA
GQPGDKGEGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVA
GPPGGSGPAGPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNT
GAPGSPGVSGPKGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGES
GKPGANGLSGERGPPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHP
GPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAP
GSPGPAGQQGAIGSPGPAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPP
GPPGAPGPCCGGVGAAAIAGIGGEKAGGFAPYYG
```

Figure 1-3: Amino acid sequences of the fibrillar collagens. The sequences for the α1 and α2 chains of type I collagen (a heterotrimer containing two α1 chains and one α2 chain) are shown on the previous page. The sequences for types II and III collagen (both homotrimers) are shown on this page. The collagenase cleavage site is highlighted in red, while other potential cleavage sites (pseudo-cleavage sites) are highlighted in cyan. Potential interchain salt bridge pairs are underlined. Because these sequences were obtained from genomic data and hydroxylation of proline residues occurs post-translationally, all hydroxyproline residues are shown as prolines.
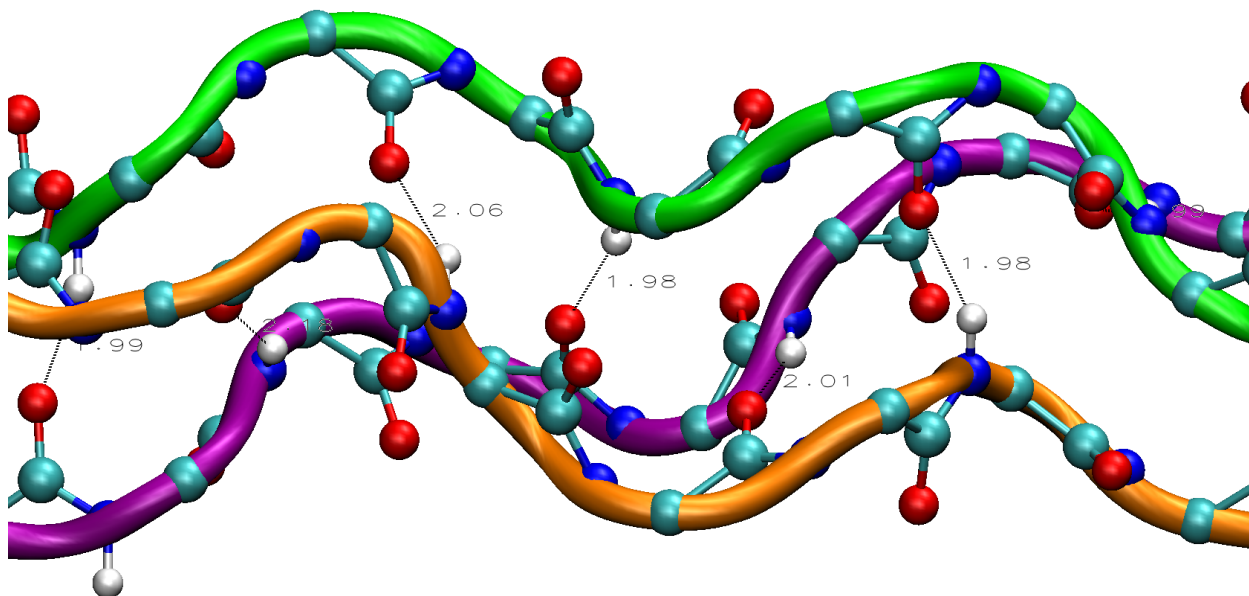
Figure 1-4: Interchain hydrogen bonding pattern of the collagen triple helix. The amide proton of a Gly residue on one chain serves as a donor, while the amide oxygen of an X1 residue on a neighboring chain serves as an acceptor. Backbone atoms (including those participating in the interchain hydrogen bonds) are shown in a CPK representation, while each of the polypeptide chains is traced in a different color. The leading strand is shown in orange, the middle strand in green, and the lagging strand in magenta. Interatomic distances are given in Ångstroms.

1.1.2 Determinants of triple-helical stability

In the previous section, we discussed the various sequence and structural characteristics that uniquely distinguish the collagen triple helix. While these basic aspects of collagen structure are fairly well understood, many of the physical determinants of triple-helical stability have only been made clear in recent years, while others have yet to be explained.

Given that one of main characteristics of the collagen triple helix is its interchain hydrogen bonding pattern and the fact that in fibrillar collagen there are approximately 1000 interchain hydrogen bonds, it has been suggested that these bonds are the primary source of stability for the collagen triple helix (Privalov 1982, Brodsky and Persikov 2005, Engel and Bächinger 2005, Shoulders and Raines 2009). Testing this hypothesis directly, however, is difficult, as methods that alter the hydrogen bonding pattern of collagen may also perturb other properties of the protein. Nonetheless, several recent studies using collagen-like model peptides with either ester or alkane isosteres of amino acids to eliminate either the donors or acceptors of the interchain

hydrogen bonds suggest that these hydrogen bonds contribute substantially to stability (Jenkins et al. 2005, Dai and Etzkorn 2009). This work has yielded estimates for the free energy of interchain hydrogen bond formation of approximately -2.0 kcal/mol in a model peptide context (Jenkins et al. 2005). While such estimates are high relative to estimates for hydrogen bond strength in globular proteins, it is worth noting that the interchain hydrogen bonds of collagen are in many ways distinct from the intrachain hydrogen bonds of the α-helix or the β-sheet and perhaps have more in common with base pair hydrogen bonds of double-stranded nucleic acids. Because of the chemical modifications made to these residues to disrupt the interchain hydrogen bonds, one cannot exclude the possibility that secondary effects are responsible for the observed destabilization of the protein. Nevertheless, these data represent the most straightforward probe to date of the contribution of interchain hydrogen bonds to triple-helical stability.

While it involves the amide proton of every third glycine residue, the interchain hydrogen bonding pattern of collagen does not constrain the possible amino acids that occupy the X1 position. From the aforementioned sequence data, however, it is clear that this position is often occupied by proline (Ramshaw et al. 1998). The abundance of prolines is likely to help individual collagen chains assume a polyproline II-type conformation, which in turn would reduce the entropic loss upon folding of the triple helix (Shoulders and Raines 2009). Conversely, due to their secondary amino group, prolines are more likely than other amino acids to sample *cis* peptide bond conformations, which are unfavorable for triple helix formation (Engel and Bächinger 2005, Shoulders and Raines 2009). Interestingly, model peptides in which the Gly-Pro peptide bond has been locked into a *trans* conformation have been shown to be less stable than the corresponding "regular" model peptides (i.e., ones in which the Gly-Pro peptide bond is unaltered and can fluctuate between *cis* and *trans* conformations), further complicating the issue (Dai et al. 2008). Host-guest studies, however, have suggested that proline is the amino acid with the highest triple-helical propensity when placed in the X1 position (Persikov et al. 2000).

After synthesis from the ribosome, collagen chains undergo extensive post-translational modification (Myllyharju 2005). The most prevalent of these modifications is the hydroxylation of proline residues in the X2 position by the enzyme prolyl 4-hydroxylase to yield hydroxyproline (Gelse et al. 2003, Myllyharju 2005). Nearly three decades ago it was observed that the thermal stability of various collagens depends strongly on their hydroxyproline content

(Burjanadze 1979, Privalov 1982). In the time since then, numerous studies with collagen-like model peptides have further confirmed this finding (Brodsky and Persikov 2005, Engel and Bächinger 2005, Shoulders and Raines 2009). Given these data and its relative prevalence in the X2 position, developing an understanding of the mechanism by which hydroxyproline stabilizes the collagen triple helix has been the focus of considerable research over the years.

One hypothesis was that Hyp formed hydrogen bonding networks through water, and analyses of the first crystal structure of a collagen-like model peptide suggested that water-mediated hydrogen bonds do in fact exist between Hyp residues and backbone atoms (Bella et al. 1994, Bella et al. 1995). However, similar hydrogen bonding networks were also seen in crystal structures of peptides lacking Hyp (Brodsky and Persikov 2005). A novel approach to further probe the mechanism of Hyp stabilization was the incorporation of fluoroproline (Flp) into the X2 position of model peptides (Holmgren et al. 1998, Holmgren et al. 1999). Flp is not capable of forming strong hydrogen bonds with water, but maintains the electronegative character of the hydroxyl group (Dunitz and Taylor 1997). The model peptides containing Flp were then shown to be hyperstable (i.e., more stable than those containing Hyp) (Holmgren et al. 1999). These data suggested that the origin of Hyp's stabilizing effect was not through water-bridged hydrogen bonds, but instead through a stereoelectronic effect (Holmgren et al. 1999, Bretscher et al. 2001). More specifically, it was suggested that the electron-withdrawing ability of Hyp (and Flp) stabilizes the pyrrolidine ring in a specific conformation (or pucker) which in turn pre-organizes the φ/ψ angles of that residue to be in a range appropriate for formation of a triple helix (Bretscher et al. 2001). In addition, it was found that the Cγ-*exo* ring pucker of Hyp also stabilizes the *trans* conformation of the peptide bond through an $n \rightarrow \pi^*$ interaction between the backbone oxygens of the Hyp residue and the preceding X1 residue (Bretscher et al. 2001). Thus, Hyp is thought to stabilize the collagen triple helix in two ways: by stabilizing both the appropriate φ/ψ conformation and the appropriate peptide bond conformation of residues in the X2 position.

Beyond the imino acids, proline and hydroxyproline, host-guest studies using model peptides have identified other amino acids and combinations of amino acids that yield similar thermal stabilities to the GPO triplet sequence (Persikov et al. 2000, Persikov et al. 2002). When placed in the X2 position, arginine is able to match the stabilizing ability of hydroxyproline (Yang et al. 1997). Moreover, X2 position arginines are prevalent in types I, II, and III collagen, existing in

12-13% of the triplets – more than all other amino acids except for hydroxyproline and lysine (see below) (Ramshaw et al. 1998). The exact mechanism of this stabilization is not known, but both a crystallographic structure and a molecular dynamics simulation of a collagen-like model peptide that contains X2 position arginines suggest that the arginine sidechain is able to fold "across" the triple helix and hydrogen bond with the backbone amide oxygen of a neighboring chain (Kramer et al. 1999, Kramer et al. 2001, Stultz 2002).

A more complex stabilizing interaction was found for pairs of lysine (Lys or K) and glutamate (Glu or E) or lysine and aspartate (Asp or D) residues. When added by themselves in either the X1 or the X2 position, these residues destabilize the triple helix relative to Pro or Hyp, respectively (Persikov et al. 2000). However, it was found that when Lys is in the X2 position and either Glu or Asp is in the X1 position of the next triplet (e.g., GPKGEO or GPKGDO) that these sequences were just as stable as an all-GPO sequence (Persikov et al. 2002, Persikov et al. 2005). In addition, a sequence analysis of fibrillar collagens revealed a higher occurence of GxKGEx and GxKGDx motifs than would be expected by chance alone (Persikov et al. 2005).

Melting experiments performed with model peptides at a low pH in which the acidic sidechains of Glu/Asp would be protonated (and therefore electrically neutral) demonstrated significantly lower thermal stabilities for the GPKGEO and GPKGDO sequences (Persikov et al. 2005). Therefore, it was suggested that the physical basis for the increased stability of the Lys-Glu and Lys-Asp pairs was the formation of interchain salt bridges between the sidechains (Persikov et al. 2005). Recent molecular dynamics simulations with these sequences supports these conclusions and have also shed light on the free energy differences for interchain salt bridge formation due to the one residue stagger of the triple helix (see Appendix A). Such interchain salt bridges are also likely to play a key role in the stabilization of bacterial and viral collagen-like proteins, as these organisms lack prolyl 4-hydroxylase and therefore their proteins contain no hydroxyproline (Xu et al. 2002, Rasmussen et al. 2003, Persikov et al. 2005).

1.1.3 Sequence characteristics of the collagenase cleavage site

The fibrillar collagens are degraded by enzymes known as collagenases (see Section 1.2) at a single site characterized by the tripeptide sequence G-[I/L]-[A/L] that is located approximately three-fourths from the N-terminus of the protein. One might naively guess that there are many such triplets within each collagen molecule, and indeed, there are three potential cleavage sites in

the α1 chain of type I collagen, five in the α2 chain of type I collagen, six in type II collagen, and five in type III collagen (Figure 1-3). One of the principle aims in investigating the molecular mechanism of collagenolysis has been to understand the specificity of the collagenase cleavage site, or in other words, why only one site is cleaved in native collagen when there are many potential cleavage sites.

One possible explanation for the specificity of the collagenase cleavage site is that the conformational ensemble of this site and its surrounding sequence differs substantially from the sequences surrounding the other potential cleavage sites (pseudo-cleavage sites). To this end, Fields extensively cataloged the sequence characteristics of both the real and pseudo-cleavage sites in types I, II, and III collagen (Fields 1991). According to this analysis, the collagenase cleavage sites are distinguished by having relatively imino-rich sequences up to four triplets N-terminal from the scissile bond, while the four triplets C-terminal of the scissile bond were imino-poor (0-2 proline/hydroxyproline residues in this stretch) and contained only a single X2 position arginine (Fields 1991). Moreover, a re-examination of these same eight triplet sub-sequences also reveals that there are no interchain salt bridge pairs in the vicinity of the collagenase cleavage site (Figure 1-3). Given a lack of sequence elements that are known to provide triple-helical stability, these data support the notion that the sequence of the collagenase cleavage site is less thermally stable than the pseudo-cleavage sites and therefore may sample a unique conformational ensemble that includes non-triple-helical conformations (Fields 1991, Stultz 2002).

1.2 Collagenases

The fibrillar collagens are degraded *in vivo* by a subfamily of Zn^{2+}- and Ca^{2+}-containing enzymes known as collagenases, which belong to the larger matrix metalloproteinase (MMP) family (Fields 1991, Borkakoti 1998, Lauer-Fields et al. 2002, Overall 2002). At present, there are 23 known members of the MMP family (Tallant et al. 2009). These enzymes share a common catalytic domain structure, as well as a consensus active site sequence (HExxHxxGxxH) in which the catalytic Zn^{2+} ion is coordinated by three histidines and a glutamate serves as the catalytic residue for peptide bond hydrolysis (Borkakoti 1998, Tallant et al. 2009). In addition, most members of the MMP family contain additional non-catalytic domains. For instance, the

collagenases (MMP-1, MMP-8, and MMP-13), stromelysins (MMP-3 and MMP-10), and metalloelastase (MMP-12) all contain a hemopexin-like domain that is C-terminal to the catalytic domain (Lauer-Fields et al. 2002). Other non-catalytic domains include the fibronectin type II modules of the gelatinases (MMP-2 and MMP-9) and the trans-membrane and cytoplasmic domains of the membrane-type MMPs (MMP-14 through MMP-17) (Lauer-Fields et al. 2002).
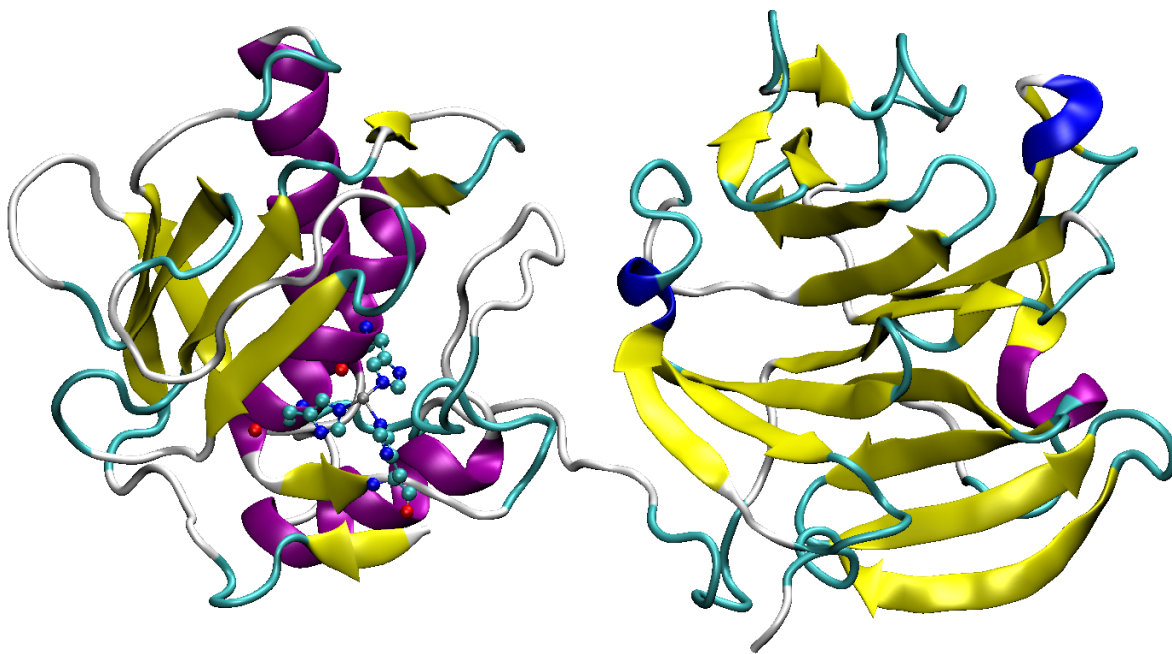
Figure 1-5: X-ray crystal structure of the collagenase MMP-1 shown in cartoon representation. The catalytic domain is on the left, while the hemopexin-like domain is on the right. The active site histidine and glutamate residues, as well as the catalytic zinc ion, are shown in CPK representation. Coordinates are taken from PDB entry 2CLT.

The MMP catalytic domain contains five major β-strands and three α-helices (Figure 1-5) (Borkakoti 1998, Bode et al. 1999, Tallant et al. 2009). Two of the active site histidines and the catalytic glutamate are contained in the second α-helix, while the third active site histidine is located in an adjacent loop region (Borkakoti 1998, Bode et al. 1999, Tallant et al. 2009). The conserved glycine between the second and third histidines of the active site enables a sharp turn in the polypeptide chain, as it is in a φ/ψ conformation that would be unfavorable for any other amino acid (Tallant et al. 2009). The mechanism of catalysis has been the subject of several

studies. It is generally thought that hydrolysis of the scissile bond occurs via a nucleophilic attack by a bound water molecule that is polarized by the catalytic glutamate and Zn^{2+} ion (Pelmenschikov and Siegbahn 2002, Gomis-Rüth 2008, Tallant et al. 2009).

The second major domain of collagenases is the C-terminal hemopexin-like domain. This domain is characterized by a four-bladed β-propeller structure in which each of the "blades" consists of four β-strands (Figure 1-5) (Borkakoti 1998, Bode et al. 1999). Each blade is connected to the next via a short connecting strand at the outer rim of the domain, and the C-terminus of the fourth blade is linked to the first strand with a disulfide bond, which is thought to stabilize the entire domain (Borkakoti 1998, Bode et al. 1999). Degradation experiments carried out with collagenase deletion mutants lacking the hemopexin-like domain, as well as various binding assays, suggest that the hemopexin-like domain contains important binding sites for collagen (Murphy et al. 1992, Chung et al. 2004, Tam et al. 2004, Lauer-Fields et al. 2009). The exact binding site(s) of the hemopexin-like domain are unknown, but modeling studies based on an x-ray fiber diffraction model of type I collagen and binding studies conducted with heterotrimeric model peptides containing the collagenase cleavage site suggest that this binding site is at least three triplets C-terminal to the cleavage site (Ottl et al. 2000, Perumal et al. 2008).

In collagenases, the catalytic domain is linked to the C-terminal hemopexin-like domain via a short linker region (~15 residues). Interestingly, chimeric enzymes in which the linker region from a non-collagenase MMP (e.g., MMP-3) is substituted into a collagenase (e.g., MMP-8) show substantially reduced ability to degrade type I collagen (Hirose et al. 1993). In addition, alanine scanning experiments focusing on the linker region of MMP-8 also revealed changes in enzyme activity (Knäuper et al. 1997). Together, these data suggest that the linker may play an important role in the overall function of the enzyme.

Despite the high sequence homology of the collagenases (51-60% sequence identity, 67-75% sequence similarity), each one displays a distinct preference for the various fibrillar collagens. MMP-1 degrades type III collagen most efficiently, while MMP-8 and MMP-13 prefer to degrade types I and II collagen, respectively (Overall 2002). These differences in degradation efficiency may arise from subtle differences in the binding contacts made by the catalytic and hemopexin-like domains (see Chapter 4), as well as differences in the rate of catalysis due to sequence differences near the active site.

1.3 The paradoxical mechanism of collagenolysis

While the collagenolytic ability of MMPs, and specifically the collagenases, has been well documented for several decades, the availability of X-ray crystal structures of collagenases and collagen-like model peptides in the last 15 years has raised a number of questions regarding the molecular mechanism of collagenolysis.

The most obvious discrepancy gleaned from these structures is the fact that the active site cleft of collagenases is only 5-8 Å wide, while the collagen triple helix is 15 Å in diameter (Lauer-Fields et al. 2002, Overall 2002, Chung et al. 2004). Therefore, barring any major conformational change of the active site, it appears impossible that the folded triple helix can fit into the collagenase active site. Even if the active site was able to undergo a major conformational change, however, crystallographic studies of model peptides have revealed that the scissile bond would be hidden within the folded triple helix and would therefore remain inaccessible to the active site of the enzyme (Figure 1-6) (Kramer et al. 1999, Stultz 2002).
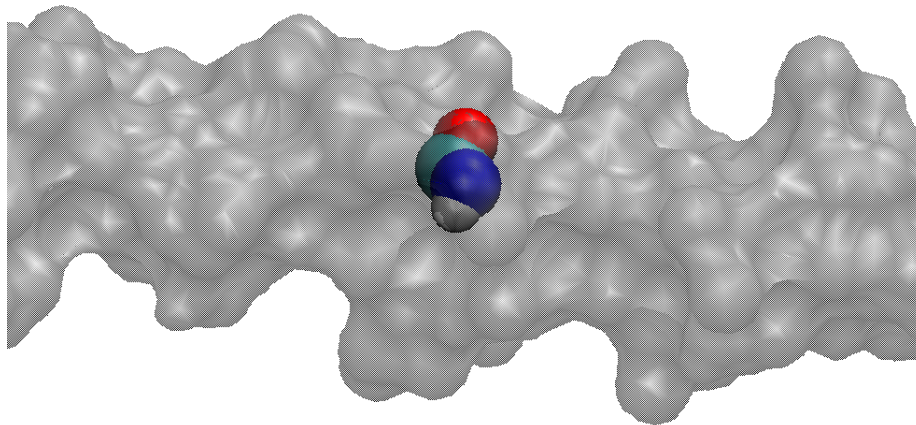
Figure 1-6: Limited solvent accessibility of the scissile bond within the collagen triple helix. The solvent-exposed surface of the collagen molecule (including backbone and sidechains) is shown in gray, while the atoms of the scissile bond are shown as colored van der Waals spheres.

As a result of the above observations, it has been thought that the collagen triple helix must in some way become unfolded such that individual chains would be able to fit into the

collagenase active site and the scissile bond would be solvent-exposed (Lauer-Fields et al. 2002, Overall 2002, Stultz 2002). There have been two hypothetical molecular mechanisms for collagenolysis put forward to explain how this unfolding occurs. One hypothesis is that collagen exists in a well-folded triple-helical state and that collagenases unwind the triple helix upon binding through the coordinated action of the catalytic and hemopexin-like domains. This active unwinding of the triple helix would not only separate the three chains, but also expose and subsequently enable the cleavage of the scissile bond. Broadly speaking, the active unwinding hypothesis corresponds to an induced fit mechanism in that the degradation-prone conformations of collagen exist *only in the presence of the enzyme*.

While active unwinding of helices has prior precedent in biology (e.g., DNA helicases), such processes typically require energy input in the form of ATP or other energy-rich molecules (Caruthers and McKay 2002). Collagenolysis, however, does not require energy input (Overall 2002, Chung et al. 2004). Therefore, an alternative hypothesis is that the conformational ensemble of collagen in the vicinity of the collagenase cleavage site includes relatively low energy structures that are partially unfolded with solvent-exposed scissile bonds even in the absence of collagenases (Stultz 2002). When collagenases are present, however, they could then bind and cleave these partially unfolded conformers. This hypothesis corresponds to a conformational selection mechanism in that the degradation-prone conformations of collagen exist *regardless of whether the enzyme is present* and that the enzyme "selects" the appropriate conformers for degradation.

Throughout this work, we investigate the molecular mechanism of collagenolysis from the standpoint of this alternative hypothesis. Specifically, this work aims to improve our understanding of the conformational ensemble of collagen at the collagenase cleavage site, using both computational (Chapters 2, 3, and 4) and experimental methods (Chapters 4 and 5). Ultimately these data suggest that collagenolysis occurs via a conformational selection mechanism in which collagenases bind and cleave partially unfolded conformers.

Chapter 2

A Conformational Selection Mechanism for Collagenolysis

This chapter is adapted from: Nerenberg PS, Salsas-Escat R, and Stultz CM. Do collagenase unwind triple-helical collagen before peptide bond hydrolysis? Reinterpreting experimental observations with mathematical models. Proteins 2008; 70:1154-1161.

Collagenolysis is an enigmatic process. Collagen-specific proteases have catalytic sites that are too narrow to accommodate the triple-helical structure of collagen, and the scissile bonds within the triple helix are 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 prior to peptide bond cleavage (Overall 2002, Chung et al. 2004). According to this theory, collagenases not only hydrolyze scissile bonds in collagen, but they also function as triple-helicases – 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). MMP-1, in particular, contains a catalytic domain and a hemopexin-like domain that are connected through a short linker region (Murphy and Knäuper 1997). While the catalytic domain alone can cleave denatured collagen, it cannot degrade collagen in its triple-helical form (Clark and Cawston 1989, Murphy and Knäuper 1997, Overall 2002). When type I collagen is incubated with an excess of MMP-1's catalytic domain, no significant collagen degradation products are formed – an observation which has been viewed as inconsistent with the notion that spontaneous unwinding of collagen occurs in the vicinity of the collagenase-cleavage site (Chung et al. 2004). In addition, while the catalytically inactive form of MMP-1 (E200A) cannot hydrolyze collagen, it can facilitate collagenolysis by enzymes that only cleave denatured collagen (Chung et al. 2004). These observations lend credence to the conjecture that collagenases locally unwind triple-helical collagen prior to collagenolysis.

Although the aforementioned experimental studies imply that partially unfolded states of collagen do not occur spontanesously, other studies suggest that collagen is conformationally labile in the vicinity of the collagenase cleavage site (Fan et al. 1993, Fiori et al. 2002, Leikina et al. 2002, Sacca et al. 2002, Stultz 2002, Stultz and Edelman 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 2002, Stultz and Edelman 2003).

2.1 Estimating the free energy barrier for active unwinding

Most theories that support a triple-helicase function for MMPs require: (i) collagen to at least bind both the catalytic and hemopexin-like domains; and (ii) coordinated domain motions that produce lateral tension, axial compression, or bending of the collagen triple helix (Overall 2002). These domain motions would, in principle, have large energetic requirements, however collagenolysis does not require energy input (Fields 1991, Murphy and Knäuper 1997, Overall 2002, Chung et al. 2004). Using a standard kinetic scheme coupled with transition state theory, we can estimate the energetic barrier associated with this enzyme-mediated reaction.

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:

$$N + E \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C_{NE} \tag{2.1}$$

$$C_{NE} \xrightarrow{k_{unwind}} C_{VE} \tag{2.2}$$

$$C_{VE} \xrightarrow{k_{cat}^{VE}} P + E \tag{2.3}$$

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 the parameters from the unwinding reaction scheme, the experimentally measured Michaelis-Menten parameters are then given by (Kuby 1990):

$$K_m^{exp} = \frac{k_{cat}^{VE}}{k_{unwind} + k_{cat}^{VE}} \frac{k_{unwind} + k_{off}}{k_{on}} \tag{2.4}$$

$$k_{cat}^{exp} = \frac{k_{unwind} k_{cat}^{VE}}{k_{unwind} + k_{cat}^{VE}} \tag{2.5}$$

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

$$k_{cat}^{exp} \approx \frac{k_{unwind} k_{cat}^{VE}}{k_{cat}^{VE}} = k_{unwind} \tag{2.6}$$

Thus with equation (2.6) we can estimate k_{unwind} as being equal to the experimentally measured catalytic rate k_{cat}^{exp}. We can then use transition state theory to calculate the free energy barrier for unwinding as follows (Voet and Voet 2004):

$$k_{unwind} \approx \frac{k_B T}{h} \exp\left(\frac{-\Delta G^\dagger}{RT}\right) = \upsilon \exp\left(\frac{-\Delta G^\dagger}{RT}\right) \tag{2.7}$$

$$\Delta G^\dagger = -RT \log\left(\upsilon^{-1} k_{unwind}\right) \tag{2.8}$$

At 298 K, υ is 6.25 ps^{-1}, and with $k_{cat}^{exp} \approx 0.015$ s^{-1}, we find $\Delta G^\dagger \approx 20$ kcal/mol. It is important to note that this barrier corresponds to the reaction *when the enzyme is present*. Moreover, barriers of this magnitude are characteristic of processes that typically do not occur on biological time scales (Dugave and Demange 2003, Hollowell et al. 2007).

2.2 A two state model of collagen degradation

Given that collagenolysis does not require energy input, we explored a reaction scheme that does not rely on enzyme-mediated unwinding (Figure 2-1). 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), with the equilibrium determined by the constant K_{eq}.

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 that the native state and the vulnerable state differ only 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 et al. 1993, Fiori et al. 2002, Sacca et al. 2002, Stultz 2002).

$$C_{NE} \underset{k_{on}^{N}}{\overset{k_{off}^{N}}{\rightleftarrows}} N \underset{k_2}{\overset{k_1}{\rightleftarrows}} V \underset{k_{off}^{V}}{\overset{k_{on}^{V}}{\rightleftarrows}} C_{VE} \xrightarrow{k_{cat}^{VE}} P + E$$
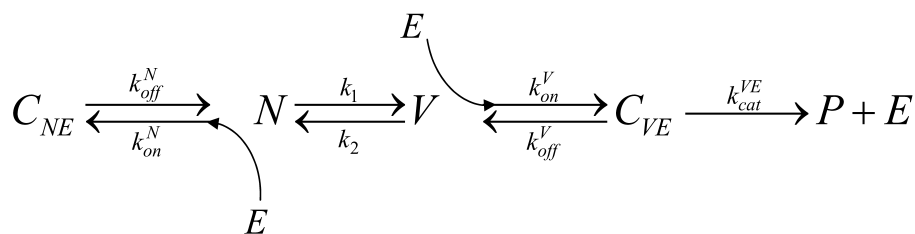
Figure 2-1: Reaction scheme used for numerical simulations. N denotes the native state of collagen (i.e., its triple-helical structure); E denotes full length MMP-1; C_{NE} is the enzyme-substrate complex; C_{VE} denotes the complex formed between the enzyme, E, and the partially unfolded state, V; and P denotes the products of collagenolysis.

In this formalism, collagenases can bind either the native state (with binding constant K_{bind}^{NE}) to form the C_{NE} complex or the vulnerable state (with binding constant K_{bind}^{VE}) to form the C_{VE} complex. The enzyme then cleaves the C_{VE} complex in which the scissile bonds are accessible with a catalytic rate k_{cat}^{VE}; the enzyme is unable to cleave the C_{NE} complex in which collagen is in its native triple-helical conformation. In addition, we explicitly exclude a step describing direct unfolding in the presence of the enzyme, $C_{NE} \rightleftharpoons C_{VE}$, 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 (e.g., enzyme-mediated unwinding).

Under steady state conditions, the experimentally observed rate of collagenolysis, k_{cat}^{exp}, may be a function of the various constants associated with our model (K_{eq}, K_{bind}^{NE}, K_{bind}^{VE}, and k_{cat}^{VE}). To understand this relationship, we again refer to the simple Michaelis-Menten model used for fitting these experiments (Kuby 1990). This model is described by the following reaction scheme:

$$N + E \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} C_{NE} \tag{2.9}$$

$$C_{NE} \xrightarrow{k_{cat}^{exp}} P + E \tag{2.10}$$

We can relate the rate of cleavage of the vulnerable state of collagen (or catalytic rate) in our proposed reaction scheme to the catalytic rate of this simple enzymatic scheme by noting that the rate of product creation $\left(\dfrac{d[P]}{dt} \right)$ must be the same for both:

$$k_{cat}^{VE}[C_{VE}]_t = k_{cat}^{exp}[C_{NE}]_t \;\rightarrow\; k_{cat}^{VE} = k_{cat}^{exp} \frac{[C_{NE}]_t}{[C_{VE}]_t} \tag{2.11}$$

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

$$\begin{aligned} 0 &= \frac{d[C_{VE}]}{dt} \\ &= k_{on}^{V}[V][E] - k_{off}^{V}[C_{VE}] - k_{cat}^{VE}[C_{VE}] \\ &= k_{on}^{V}[V][E] - \left(k_{off}^{V} + k_{cat}^{VE} \right)[C_{VE}] \end{aligned}$$

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

$$[C_{VE}] = \frac{k_{on}^{V}[V][E]}{\left(k_{off}^{V} + k_{cat}^{VE} \right)} \tag{2.12}$$

If we assume that the binding of native state collagen is similar between both schemes, we can apply the steady state approximation to the intermediate native state-enzyme complex:

$$0 = \frac{d\left[C_{NE}\right]}{dt}$$

$$= k_{on}^{N}[N][E] - k_{off}^{N}\left[C_{NE}\right]$$

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

$$\left[C_{NE}\right] = \frac{k_{on}^{N}[N][E]}{k_{off}^{N}} \tag{2.13}$$

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

$$k_{cat}^{VE} = k_{cat}^{exp}\frac{[C_{NE}]_{t}}{[C_{VE}]_{t}}$$

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

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

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

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

If we assume that $k_{off}^{V} \gg k_{cat}^{VE}$ (i.e., the rapid equilibrium hypothesis (Voet and Voet 2004)), we find:

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

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

$$= k_{cat}^{exp}\left(K_{eq}\right)^{-1}\left(K_{bind}^{NE}\right)\left(K_{bind}^{VE}\right)^{-1}$$

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

$$k_{cat}^{exp} \approx k_{cat}^{VE}K_{eq}\left(\frac{K_{bind}^{VE}}{K_{bind}^{NE}}\right) \tag{2.14}$$

Thus, 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. This steady state relationship demonstrates that small values for k_{cat}^{exp} 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 solved the ordinary differential equations (ODEs) arising from the reaction scheme shown in Figure 2-1 to obtain estimates of the extent of collagenolysis as a function of K_{eq}, K_{bind}^{VE}, and K_{bind}^{NE}. Unlike the derivation presented above, 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 1μM collagen with enzyme for 48 h (Chung et al. 2004).

2.3 Computational Methods

The reaction scheme of Figure 2-1 leads to a set of ordinary differential equations (ODEs) as listed below:

$$\frac{d[N]}{dt} = -k_1[N] + k_2[V] - k_{on}^N[N][E] + k_{off}^N\left[C_{NE}\right] \tag{2.15}$$

$$\frac{d[V]}{dt} = -k_2[V] + k_1[N] - k_{on}^V[V][E] + k_{off}^V\left[C_{VE}\right] \tag{2.16}$$

$$\frac{d\left[C_{NE}\right]}{dt} = -k_{off}^N\left[C_{NE}\right] + k_{on}^N[N][E] \tag{2.17}$$

$$\frac{d\left[C_{VE}\right]}{dt} = -k_{off}^V\left[C_{VE}\right] + k_{on}^V[V][E] - k_{cat}^{VE}\left[C_{VE}\right] \tag{2.18}$$

$$\frac{d[P]}{dt} = k_{cat}^{VE}\left[C_{VE}\right] \tag{2.19}$$

$$\frac{d[E]}{dt} = -k_{on}^N[N][E] + k_{off}^N\left[C_{NE}\right] - k_{on}^V[V][E] + k_{off}^V\left[C_{VE}\right] + k_{cat}^{VE}\left[C_{VE}\right] \tag{2.20}$$

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 (Welgus et al. 1982). Type I collagen contains two α1(I) chains and one α2(I) chain. The rate constant for denatured α1(I) chains is approximately 230 h^{-1} and the rate constant for denatured α2(I) chains is approximately 750 h^{-1}. Therefore to estimate the catalytic rate constant for type I collagen we use k_{cat}^{VE} = (1/3) * (2*230 h^{-1} + 750 h^{-1}) ≈ 400 h^{-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, and therefore scissile bonds in vulnerable conformations may not be as exposed as scissile bonds in denatured states. Using an upper bound 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 48 h after exposure of the catalytic domain to collagen using this value. Moreover, conducting numerical simulations with lower values of k_{cat}^{VE} produces the same result. Lastly we note that estimates for the remaining rate constants are determined by scaling the rate constants from predetermined initial values to values that match the desired the equilibrium constants $K_{eq} = k_1/k_2$, $K_{bind}^{VE} = k_{on}^V/k_{off}^V$, and $K_{bind}^{NE} = k_{on}^N/k_{off}^N$. In particular, the forward (or on) rates are held constant, while the off rates are scaled accordingly.

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 (48 h), 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}. While 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; i.e., 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 1 μM.

2.4 Predictions of the conformational selection model of collagenolysis

2.4.1 Results of the model at low enzyme concentrations

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 was made to be consistent with prior experiments (Chung et al. 2004). Figure 2-2 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:

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

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

III. When $K_{bind}^{VE} > 10^3$ M^{-1} and $K_{eq} > 10^{-1}$ (or ~10^1 for $K_{bind}^{NE} = 10^8$ M^{-1}) significant collagen degradation occurs.

35

Figure 2-2: Fraction of collagen degraded at 298 K after 48 h with $[E]/[\text{Total Collagen}] = 10^{-2}$ 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}, or (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 thereby demonstrating 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} at 298 K for full length enzyme (Chung et al. 2004). In addition, if $K_{bind}^{NE} \geq 10^8$ M^{-1} then little collagen degradation will be observed after 48 h, assuming $K_{eq} \approx 10^0$. Hence the fact that significant collagenolysis does occur at room temperature suggests that $K_{bind}^{NE} < 10^8$ M^{-1}. We note that accurate estimates for K_{bind}^{NE} at 298 K are difficult to obtain experimentally as our data suggests 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 10^6 M^{-1} – and therefore well below this upper bound (Welgus et al. 1980).

2.4.2 Results of the model at high enzyme concentrations

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 studied how K_{eq}, K_{bind}^{VE}, and K_{bind}^{NE} affect the rate of collagen-degradation when a high enzyme concentration is present (Figure 2-3). In Figure 2-3 we have explicitly labeled the region corresponding to incubating wild-type enzyme with collagen (i.e., $K_{bind}^{VE} > 10^3$ 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.

Prior experiments which incubated collagen with high concentrations of MMP-1 mutants lacking a hemopexin-like domain did not observe any collagenolysis after 48 h (Chung et al. 2004). While 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 MMP-1 for collagen. This is especially poignant given that MMP-1 hemopexin-like domains likely contain sites that bind collagen (Clark and Cawston 1989, Windsor et al. 1991, Murphy et al. 1992, Bigg et al. 1997). Figure 2-3 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}.
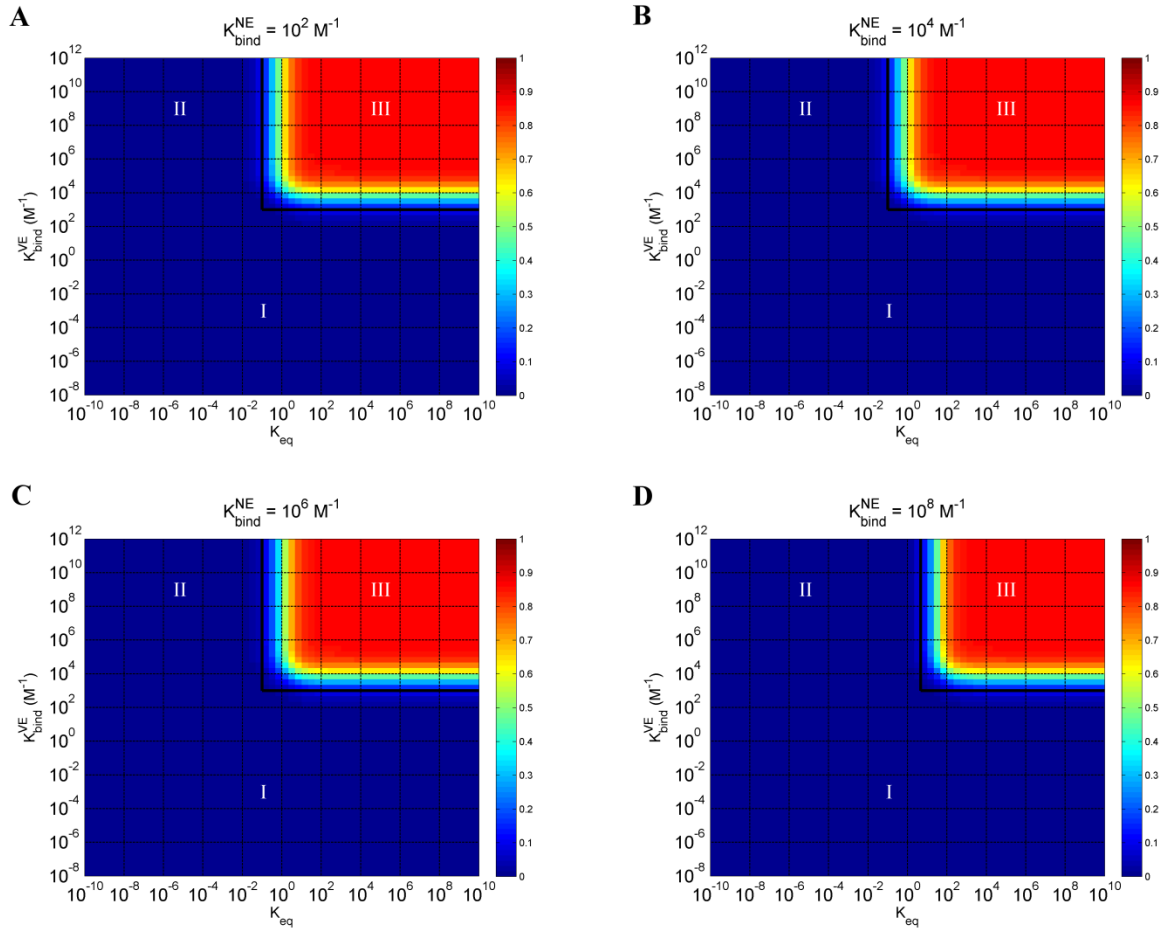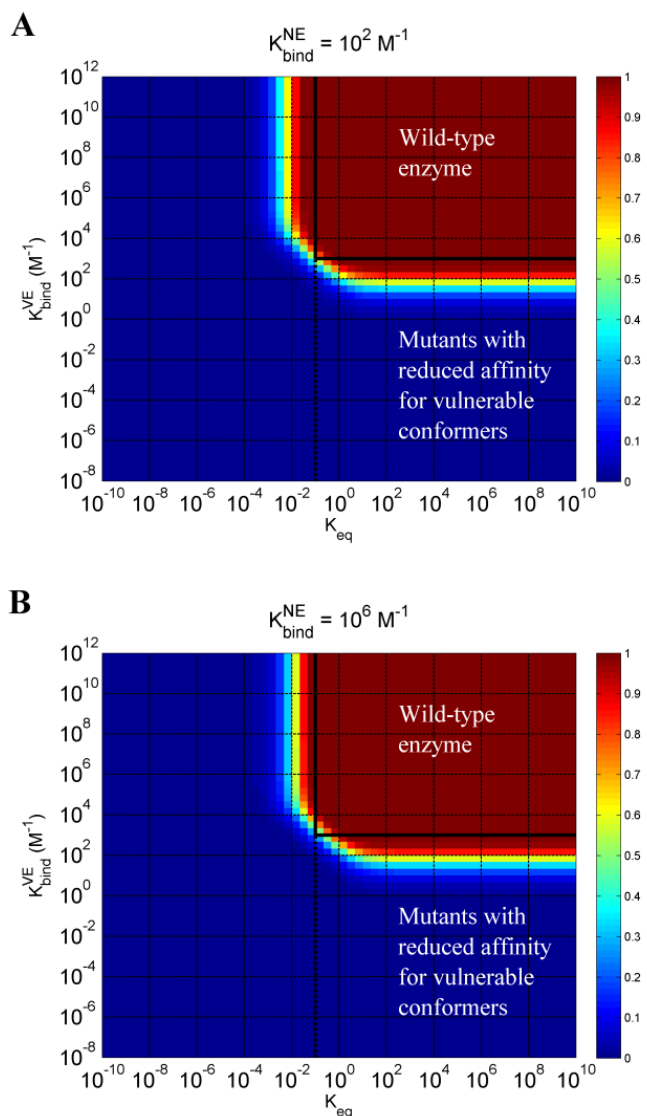

Figure 2-3: Fraction of collagen degraded at 298 K after 48 h with $[E]/[\text{Total Collagen}] = 1$ and (A) $K_{bind}^{NE} = 10^2$ M^{-1} or (B) $K_{bind}^{NE} = 10^6$ M^{-1}.

2.4.3 Calculation of the Destabilization Energy

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, then little collagen degradation will be observed after a 48 h 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 MMP-1, and E_m denote a mutant form of MMP-1 that binds vulnerable states with reduced affinity such that incubating E_m with collagen for 48 h results in degradation of less than 1% of the total collagen present. The two binding reactions are given by:

$$E + V \xrightarrow{\Delta G = -RT \ln \left(K_{bind}^{VE} \right)} EV \tag{2.21}$$

$$E_m + V \xrightarrow[\Delta G_m = -RT \ln \left(K_{\min}^{VE} \right)]{} E_m V \tag{2.22}$$
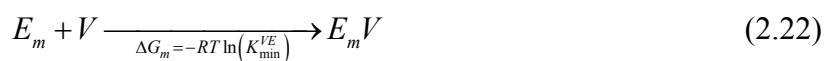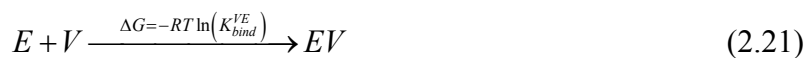
When $K_{bind}^{VE} > K_{\min}^{VE}$ the destabilization energy is given by:

$$\Delta \Delta G = \Delta G_m - \Delta G = RT \ln \left(\frac{K_{bind}^{VE}}{K_{\min}^{VE}} \right) \tag{2.23}$$

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 48 h. We note that $K_{\min}^{VE}$ is a function of K_{eq} and therefore that the destabilization energy is a function of both K_{bind}^{VE} and K_{eq}.

The main point here is that if MMP-1 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 ($K_{bind}^{NE} \approx 10^6$ M^{-1} and $K_{bind}^{VE} \approx 10^6$ M^{-1}) the destabilization energy is less than 7.5 kcal/mol. Moreover, when $K_{bind}^{VE} > 10^3$ M^{-1} 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 2-4). Therefore modest changes in the binding affinity – as expected with forms of MMP-1 that lack collagen binding sites in the hemopexin-like domain – will have important consequences on the extent of collagenolysis.
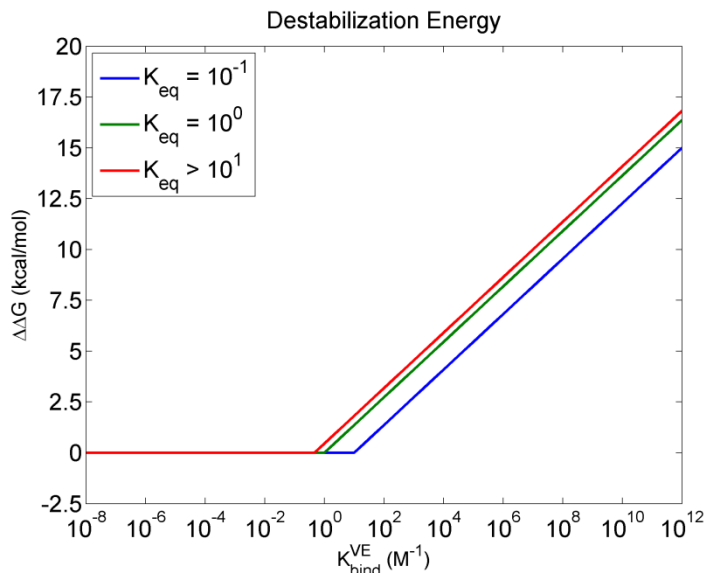
Figure 2-4: Destabilization energies at 298 K with $[E] = 1$ µM and $K_{bind}^{NE} = 10^6$ M^{-1}. Energies plotted for $K_{eq} = 10^{-1}$, 10^0, and $>10^1$. Because K_{min}^{VE} does not vary for $K_{eq} > 10^1$ (Figure 2-3B), the destabilization energy curve is the same for all $K_{eq} > 10^1$. Given 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.

2.5 Discussion

Since the first crystallographic structures of MMPs were published, the apparent mismatch between the width of the MMP active site pocket (~8 Å) and the width of the collagen triple helix (15 Å) 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; i.e., that 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 MMP-1 alone, MMP-1(ΔC), cannot cleave collagen (Clark and Cawston 1989, Chung et al. 2004) – a finding consistent with the notion that preformed partially unwound states of collagen do not occur spontaneously. However, our data demonstrates that the lack of collagen degradation by MMP-1(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. As the hemopexin-like domain of MMP-1 contains sites that bind collagen (Clark and Cawston 1989, Windsor et al. 1991, Murphy et al. 1992, Bigg et al. 1997), it is likely that MMP-1(ΔC) and full-length enzyme have different affinities for both native and vulnerable states. However, as is evident from Figures 2.2 and 2.3, 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 MMP-1 hemopexin-like domain (Chung 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 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 observed the CD spectrum.

3. It has been noted that collagen can be cleaved when incubated with the catalytic domain of MMP-1(ΔC) and severed MMP-1 hemopexin-like domains (Hpx$_{MMP-1}$) (Chung et al. 2004). The lack of a linker connecting MMP-1(ΔC) and Hpx$_{MMP-1}$ makes active triple-helical unwinding unlikely, as a helicase function would require coordinated domain movements. If we consider, however, 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.

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; e.g., full length MMP-3, catalytic domain of MMP-3 alone, and HLE (Chung et al. 2004). While these data have been interpreted to mean that E200A unwinds triple-helical collagen, they are explained by the notion that pre-existing 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 MMP-1(ΔC) may be further understood by considering how connected catalytic and hemopexin-like domains affect the collagenolytic ability of MMP-1. 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 MMP-1(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.

Chapter 3

A Structural Model for the Type I Collagen Collagenase Cleavage Site

In Chapter 2, we demonstrated that a conformational selection mechanism in which partially unfolded conformers of collagen can be bound and cleaved by MMPs can explain a wide variety of prior experimental observations that had previously been interpreted in the framework of an active unwinding (induced fit) mechanism for collagenolysis. In this work, we wish to generate a structural model for the collagenase cleavage site of type I collagen to determine if low energy partially unfolded states in fact exist, and if they do, we would like to characterize their conformation(s) and dynamics.

Approximately 90% of collagen found in the body is fibril-forming collagen – types I, II, III, V and XI – that crosslink to form fibrillar structures (Gelse et al. 2003). Of the fibril-forming (or fibrillar) collagens, type I collagen is the most common. Type I collagen is a major structural constituent of bone, tendon, skin, and a myriad of other tissues (Gelse et al. 2003). Therefore understanding the mechanism of degradation for the fibrillar collagens, specifically type I collagen, may lead to the design of novel therapies that slow the progression of many human diseases that involve abnormal collagen catabolism.

Fibrillar collagens are degraded by a family of Zn^{2+}-dependent proteases known as matrix metalloproteinases (MMPs); e.g., primarily the collagenases MMP-1, MMP-8, and MMP-13 (Bode et al. 1999, Jones et al. 2003). MMPs are multi-domain proteins that generally contain a catalytic domain connected via a linker region to a hemopexin-like domain (Bode et al. 1999). Some MMPs contain additional domains, e.g. the fibronectin type II-like collagen binding domains of the gelatinases, MMP-2 and MMP-9, or the cytoplasmic domains of the membrane-

type MMPs (Overall 2002). Despite these differences in the structure of various MMPs, a common paradox arises from analyses of different MMPs structures and the structure of collagen. Crystallographic studies of collagen-like model peptides suggest that the scissile bond that is cleaved by collagenases is not solvent accessible and that the triple-helical collagen structure cannot fit into the MMP active site (Fields et al. 1987, Kramer et al. 2001, Overall 2002). Consequently peptide bond hydrolysis of scissile bonds in collagen must involve some form of collagen unfolding that leads to the formation of a solvent-exposed scissile bond that can fit into the MMP active site (Overall 2002, Chung et al. 2004, Tam et al. 2004). Prior experimental studies have been interpreted to mean that collagen does not spontaneously unfold in solution and that unwinding of the triple helix is due to the simultaneous binding and subsequent coordinated motion of collagenase domains in a process referred to as molecular tectonics (Overall 2002). However, a growing body of experimental and theoretical results suggests that the collagen triple helix may be more flexible than originally thought, and that thermal fluctuations in the structure of collagen may lead to partial unfolding in the region of the cleavage site (Fan et al. 1993, Fiori et al. 2002, Leikina et al. 2002, Stultz 2002, Makareeva et al. 2008, Nerenberg et al. 2008). In particular, theoretical calculations on collagen-like model peptides suggest that regions near the collagenase cleavage site can adopt both well-folded (native) and partially unfolded (vulnerable) states (Stultz 2002). In this formalism, collagenolysis occurs when collagenases bind to and subsequently cleave pre-formed vulnerable states of collagen (Nerenberg et al. 2008).

A number of experimental and computational works have focused on models of type III collagen, a homotrimeric fibrillar collagen, in large part due to the ease of synthesizing self-assembling homotrimeric model peptides (Fields and Prockop 1996, Brodsky and Persikov 2005). For example, the model peptide T3-785, which contains a portion of the type III collagen sequence immediately C-terminal to the collagenase cleavage site, has been studied using x-ray crystallography, NMR, and detailed molecular dynamics simulations (Fan et al. 1993, Kramer et al. 2001, Stultz 2002, Stultz and Edelman 2003). Type I collagen, conversely, is a heterotrimer composed of two identical α1(I) chains and one α2(I) chain (Gelse et al. 2003). Synthesis of heterotrimeric model peptides is therefore more difficult and often requires the construction of covalent linkages to join chains with distinct sequences (Fields et al. 1996, Fields and Prockop 1996, Ottl et al. 1996, Fiori et al. 2002).

Experiments with full length type I collagen have been performed that provide some insight into the mechanism of collagen degradation. In particular, it has been argued that MMP-1 cleaves α1 and α2 chains in type I collagen with different efficiencies (Chung et al. 2004). Previous experiments incubated type I collagen with catalytically inactive MMP-1 at 25 °C followed by the introduction of other active proteases (e.g., catalytic domains of various MMPs and human leukocyte elastase). The result was that α1 chains were cleaved more rapidly than the α2 chain (Chung et al. 2004). From these data it has been argued that inactive MMP-1 preferentially bound to the α2 chain, leaving the corresponding sites on α1 chains amenable to cleavage by noncollagenolytic proteases. That is, since inactive MMP-1 was unable to cleave the scissile bond on α2 chains, it likely remained bound to the α2 chain for a relatively long period of time, thereby protecting the scissile bond from being cleaved by other proteases. Therefore, to explore the physical basis underlying any potential preference for MMP-1 association with either the α1 or α2 chains of type I collagen, we conducted unfolding simulations of each chain in type I collagen.

3.1 Computational Methods

3.1.1 The sequence of type I collagen inferred from sequence homologs

Sequences for human α1 and α2 chains were obtained from the UniProtKB/Swiss-Prot sequence database (accession numbers P02452 and P08123, respectively) (The UniProt Consortium 2008). Since these amino acid sequences are deduced from nucleotide sequences, and hydroxylation is a post-translational modification, it is impossible to determine which prolines are hydroxylated from these data alone. Fortunately, the entire amino acid sequence of the homologous bovine type I collagen and a large fraction of the amino acid sequence of the homologous chicken type I collagen have been determined using Edman degradation (Bornstein and Traub 1979, Dixit et al. 1979). As expected, when an imino acid appears in the X2 position of G-X1-X2 collagen triplets in these sequences, it is most often a hydroxyproline (Bornstein and Traub 1979, Gelse et al. 2003, Brodsky and Persikov 2005). In fact, in the entire sequence of both bovine and chicken α1 chains there are only four locations where hydroxylation at the X2 position does not occur (Highberger et al. 1982, Glanville et al. 1983). The most notable exception is a triplet that is N-terminal to the triplet containing the scissile bond (yellow residues in Figure 3-1). Given the

significant sequence homology between bovine/chicken and human type I collagen, particularly in the α1 chain, we modeled a proline in this triplet of human type I collagen instead of a hydroxyproline (Figure 3-1).

```
α1 chain
Bovine: GEKGAOGADGPAGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLO
Chick:  GEKGSOGADGPIGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLO
Human:  GEKGSOGADGPAGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLO

α2 chain
Bovine: GEKGPSGEQGTAGPOGTOGPQGLLGAOGFLGLOGSRGERGLOGVA
Chick:  GEKGPSGEAGAAGPOGTOGPQGILGAOGILGLOGSRGERGLOGIA
Human:  GEKGPSGEAGTAGPOGTOGPQGLLGAOGILGLOGSRGERGLOGVA
```

Figure 3-1: Alignments of sequences of type I collagen from bovine and chicken collagen in the region of the collagenase cleavage site. The sequence of human collagen is also shown where the hydroxylation pattern is deduced from the sequence alignment. The triplets containing the scissile bonds are colored in magenta. Residues that vary across species are underlined. The GTP triplet of the α1 chain and GTO triplet of the α2 chain are indicated in yellow (respectively).

3.1.2 Construction of the initial model

The sequences of the α1 and α2 chains of type I collagen used in the potential of mean force calculations included 45 residues in total – the triplet containing the scissile bond and the 7 triplets adjacent to that triplet in both the N and C terminal directions. The initial structure, which contained only heavy atoms, was first generated using the Triple Helix Builder (Rainey and Goh 2004). A polar hydrogen model of collagen was then constructed with CHARMM version 33b2 (Brooks et al. 2009) using the CHARMM19 extended-atom force field and parameters (Neria et al. 1996). To briefly describe it, the CHARMM19 force field has a potential energy function of the following form:

$$U = U_{bonded} + U_{non-bonded} = \left(U_{bonds} + U_{angles} + U_{dihedrals} + U_{U-B} + U_{impropers}\right) + \left(U_{vdW} + U_{elec}\right)$$

$$= \left(\sum_{i,j} k_{ij}(r_{ij} - r_{ij}^0)^2 + \sum_{i,j,k} k_{ijk}(\theta_{ijk} - \theta_{ijk}^0)^2 + \sum_{i,j,k,l} k_{ijkl}(1 + \cos(n_{ijkl}\phi_{ijkl} - \phi_{ijkl}^0)) \right.$$

$$\left. + \sum_{i,k} k_{ik}^{U-B}(r_{ik} - r_{ik}^0)^2 + \sum_{i,j,k,l} k_{ijkl}^{impr}(\phi_{ijkl} - \phi_{ijkl}^0)^2 \right) \qquad (3.1)$$

$$+ \left(\sum_{i,j} 4\eta_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right) + \frac{q_i q_j}{\varepsilon r_{ij}} \right)$$

The first three energy terms in equation 3.1 are the familiar bond stretching, bending (angle), and torsion (dihedral) terms for atoms connected by bonds. The fourth term is the Urey-Bradley term to constrain distances between atoms separated by two bonds, and the fifth term is the improper dihedral term for enforcing the planarity of a group of four bonded atoms. The last two terms are the energy terms for non-bonded atom pairs: the first of these is the van der Waals energy and the second is the electrostatic energy. Parameters for hydroxyproline were obtained from previous computational studies (Stultz 2002, Stultz and Edelman 2003).

This structure was then minimized with 100 steps of steepest descent (SD) minimization in vacuo to relieve bad contacts. Different models of collagen were constructed, each having a different arrangement of chain stagger (e.g., the α2 chain can be either chain A, B, or C in the triple helix). Initial simulations were run with each model, but only the model having α2 chain as chain C led to stable trajectories having a well-folded structure for all three chains.

3.1.3 Solvation and equilibration of model structure

The minimized structure was overlaid with a 30 Å sphere of equilibrated TIP3P water molecules (about the center of the protein) (Jorgensen et al. 1983). Water molecules overlapping with the protein structure were deleted; the remaining waters were then subjected to 5 ps of molecular dynamics at 300 K, with all atoms of the protein remaining fixed. This solvent overlay procedure was repeated two more times (always deleting any overlapping waters) to ensure adequate solvation and yielded a total of 3403 TIP3P water molecules in the solvent sphere. Taking into account the volume excluded by the collagen triple helix, this number of water molecules is consistent with a density of ~1000 kg m^{-3}.

The solvated system was then minimized using 200 steps of SD, followed by 200 steps of conjugate gradient minimization. The minimized system was heated linearly over 20 ps to 300 K and equilibrated for an additional 80 ps at 300 K with a Nosé-Hoover thermostat (Hoover 1985). To allow for side chain relaxation during minimization, heating, and equilibration without significant distortion of the triple helix, the backbone atoms of the protein were harmonically restrained with a force constant of 100 kcal mol^{-1} Å^{-2}. Using atom-based cutoffs, electrostatic interactions were switched to zero between 8 and 12 Å, and van der Waals interactions were shifted to zero at 12 Å. The nonbonded list was truncated at 13 Å. Bond lengths involving hydrogen atoms were fixed with SHAKE/Roll and RATTLE/Roll, and a time step of 1 fs was used with a velocity Verlet integrator (Lamoureux and Roux 2003).

3.1.4 Umbrella sampling and pmf calculations

The basic principle behind umbrella sampling is to run a simulation with a relatively strong biasing potential, $U_i^{bias}(\xi)$, which is some known function of the reaction coordinate of interest, ξ, that forces the system to extensively sample small regions of that reaction coordinate (Torrie and Valleau 1974). By conducting many such biased simulations, the full range of the reaction coordinate can thus be sampled in detail. Ultimately, however, we wish to obtain the unbiased probability distribution, $p(\xi)$, from these biased simulations so that we can construct a potential of mean force:

$$W(\xi) = -\frac{1}{\beta}\log\big(p(\xi)\big) + C \qquad (3.2)$$

where C is typically set such that the minimum value of $W(\xi)$ is zero (i.e., $C = \frac{1}{\beta}\log\big(\max(p(\xi))\big)$) (Roux 1995). The unbiased probability distribution, $p_i(\xi)$, is related to the biased probability distribution from a simulation run with a biasing potential $U_i^{bias}(\xi)$ as follows:

$$p_i(\xi) = p_i^*(\xi)e^{\beta U_i^{bias}(\xi)}\left\langle e^{-\beta U_i^{bias}(\xi)}\right\rangle_\xi$$
$$= p_i^*(\xi)e^{\beta U_i^{bias}(\xi)}e^{-\beta F_i} \qquad (3.3)$$

where $p_i^*(\xi)$ is the biased probability distribution (Roux 1995). A typical form for $U_i^{bias}(\xi)$ (and the form used in this work) is a harmonic potential:

$$U_i^{bias}(\xi) = \frac{1}{2}k_i\left(\xi - \xi_i^{cen}\right)^2 \tag{3.4}$$

where k_i is the force constant and ξ_i^{cen} is the "center" value of the reaction coordinate for biased simulation window i. From equation 3.3 (specifically, the first two terms $p_i^*(\xi)e^{\beta U_i^{bias}(\xi)}$), we can see that if a biased simulation frequently samples some value of the reaction coordinate ξ that is relatively far from ξ_i^{cen}, then this suggests that the unbiased probability is quite large at ξ.

The primary difficulty of umbrella sampling is in correctly combining the unbiased probability distributions obtained by the different simulations. This requires first solving for the F_i (see equation 3.3) and then appropriately weighting the contributions of the various probability distributions. In particular, solving for the F_i is difficult because to calculate it exactly requires knowledge of the unbiased probability distribution over the entire range of ξ – the very thing we are trying to obtain. The exact manner in which this is done is described later in this section.

The potential of mean force for each chain was thus constructed using umbrella sampling with the average interchain nitrogen-to-oxygen (NO) distance within the (solvated) reaction region serving as the reaction coordinate. This reaction region encompassed the seven central triplets of the sequence surrounded by the solvent sphere, with the scissile bond being contained in the middle triplet. Simulations for a given chain constrained NO distances between that chain and the other two, e.g. a simulation for chain A would constrain the distances between chains A and B as well as chains A and C, hence modeling a chain separating from its two partners.

Simulation windows were centered on 1.79-6.64 Å in 0.14 Å intervals using a harmonic biasing potential with a force constant of 196 kcal mol^{-1} Å^{-2}. In practice, total NO distance (the number of NO pairs multiplied by average NO distance) is used for the simulations, which yields an integer, rather than fractional, increment between windows. Before beginning the umbrella sampling protocol, the system underwent an additional 100 ps of equilibration at 300 K with the H-bond lengths constrained to the starting structure. The starting window for all three chains was 2.93 Å, as this was the average interchain NO distance after side chain equilibration. Each

window was simulated for 100 ps, including 40 ps of equilibration. The value of the reaction coordinate was saved every 50 fs, yielding 1200 data points per window. In some windows the equilibration time was observed to be longer, thus necessitating longer simulations. In these cases the simulation length was 200 ps, with 80 ps of equilibration, yielding 2400 data points per window. Parameters for these molecular dynamics simulations were the same as those described in the previous section.

An unbiased probability distribution (and thus potential of mean force) for each chain was constructed from the biased probability distribution of each window using the weighted histogram analysis method (WHAM) (Kumar et al. 1992, Souaille and Roux 2001). The WHAM equations are as follows:

$$p(\xi) = \sum_{i=1}^{N_W} \frac{n_i p_i^*(\xi)}{\sum_{j=1}^{N_W} n_j e^{-\beta\left(U_j^{bias}(\xi) - F_j\right)}} \tag{3.5}$$

$$e^{-\beta F_i} = \int d\xi \, e^{-\beta U_i^{bias}(\xi)} p(\xi) \tag{3.6}$$

where N_W is the total number of biased simulations and n_i is the number of data points for biased simulation i. By making an initial guess for the F_i in equation 3.5, one can then iteratively solve equations 3.5 and 3.6 until reaching some arbitrary convergence criterion. In this work, we iterated until the root-mean-square difference between successive sets of F_i was less than 10^{-6}. The resulting pmfs were then smoothed in MATLAB using a robust linear fit method with a data span of 5%.

3.1.5 Calculation of average NO distances and Cα RMS fluctuations

Average NO distances were obtained using the simulation window centered closest to the energy minimum of interest. Relevant NO distances were recorded every 50 fs using CHARMM, summed, and then divided by the number of data points in that window (a simple average). Cα root-mean-square (RMS) fluctuations were obtained using the same simulation windows. An average structure for each window and RMS fluctuations from that average were calculated using CHARMM.

3.2 Conformational thermodynamics and flexibility near the cleavage site

Non-hydroxylation of proline residues in the X2 position of G-X1-X2 triplets is curious not only because of its relative rarity, but also because the stability of collagen-like sequences is a function of their hydroxyproline content (Jenkins and Raines 2002, Brodsky and Persikov 2005, Engel and Bächinger 2005). Therefore, we postulated that non-hydroxylation of prolines in the vicinity of the collagenase cleavage site would affect the conformational thermodynamics of collagen in this region.

We computed conformational free energy profiles for unfolding all three chains (two α1 and one α2) in type I collagen using the sequence shown in Figure 3-1. As we were interested in the conformational thermodynamics of the region immediately surrounding the cleavage site, a stochastic boundary method similar to the one outlined in previous work was used (Stultz 2002). Simulations were conducted at 300K, a temperature comparable to previous degradation experiments that were conducted at 298 K (25 °C) (Chung et al. 2004).

Folded triple-helical collagen conformations (Figure 3-2A) have interchain hydrogen bonds (H-bonds) between the inward-facing amide hydrogen of glycine residues and the carbonyl oxygen of X1 residues in G-X1-X2 triplets (Figure 3-2B). As nitrogen (of Gly) and oxygen (of X1) are the heavy atoms involved in these H-bonds, unfolding simulations were conducted using a reaction coordinate corresponding to the average interchain backbone nitrogen-to-oxygen (NO) distance. NO distances below 3.6 Å are consistent with the formation of a hydrogen bond between the amide hydrogen and carbonyl oxygen; i.e., bonds present in the folded state. Consequently, this reaction coordinate enabled us to more fully explore differences in H-bond patterns between different conformers. A free energy profile (also known as potential of mean force or pmf) for unfolding each of the three chains in type I collagen was obtained by varying the average NO distance between one chain and the other two, thereby allowing us to directly monitor the separation of one chain from the other two.

(a)

(b)

```
Chain A (α1):  GEKGSOGADGPAGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLO
                 ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
Chain B (α1):   GEKGSOGADGPAGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLO
                 ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙
Chain C (α2):    GEKGPSGEAGTAGPOGTOGPQGLLGAOGILGLOGSRGERGLOGVA
                 ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙ ↙
Chain A (α1):  GEKGSOGADGPAGAOGTOGPQGIAGQRGVVGLOGQRGERGFOGLO
```
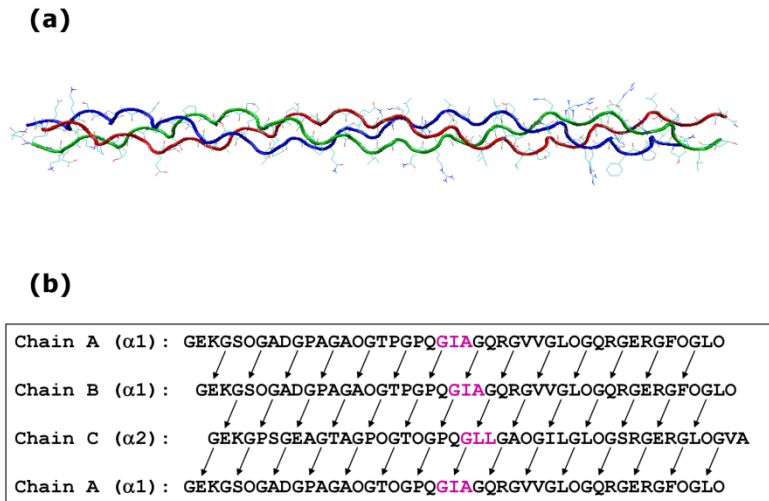
Figure 3-2: (A) The prototypical triple-helical structure of type I collagen in the region of the collagenase cleavage site (generated with the Triple Helix Builder (Rainey and Goh 2004)). Chain A is colored blue, Chain B is colored green, and Chain C is colored red. (B) Hydrogen bonding pattern in the triple helical conformation. The triplets containing the scissile bonds are colored in magenta. Arrows point from hydrogen bond donors to hydrogen bond acceptors.

3.2.1 Potentials of mean force for unfolding

Despite the fact that each α1 chain is in a distinct chemical environment (Figure 3-2), the pmfs for both α1 chains are very similar (Figures 3-3A and 3-3B). Both pmfs have a single deep energy minimum at an average NO distance of 3.1 Å and 3.2 Å for chains A and B, respectively (Figures 3-3A and 3-3B). Hence, on average the interchain H-bonds are preserved in structures corresponding to these minima. By contrast, the pmf for unfolding the α2 chain has two distinct local energy minima. In addition to the global energy minimum at an average NO distance of 3 Å, there is another minimum at 3.8 Å that has an energy within 1 kcal/mol of the lowest energy state (Figure 3-3C). As the average NO distance of the second minimum is slightly greater than 3.6 Å, this suggests that on average, interchain hydrogen bonds in structures corresponding to the second minimum are broken. Taken together, these data suggest that unfolding either α1 chain is energetically unfavorable, while α2 chain unfolding occurs more readily and leads to the formation of a state that has, on average, disrupted interchain hydrogen bonds.
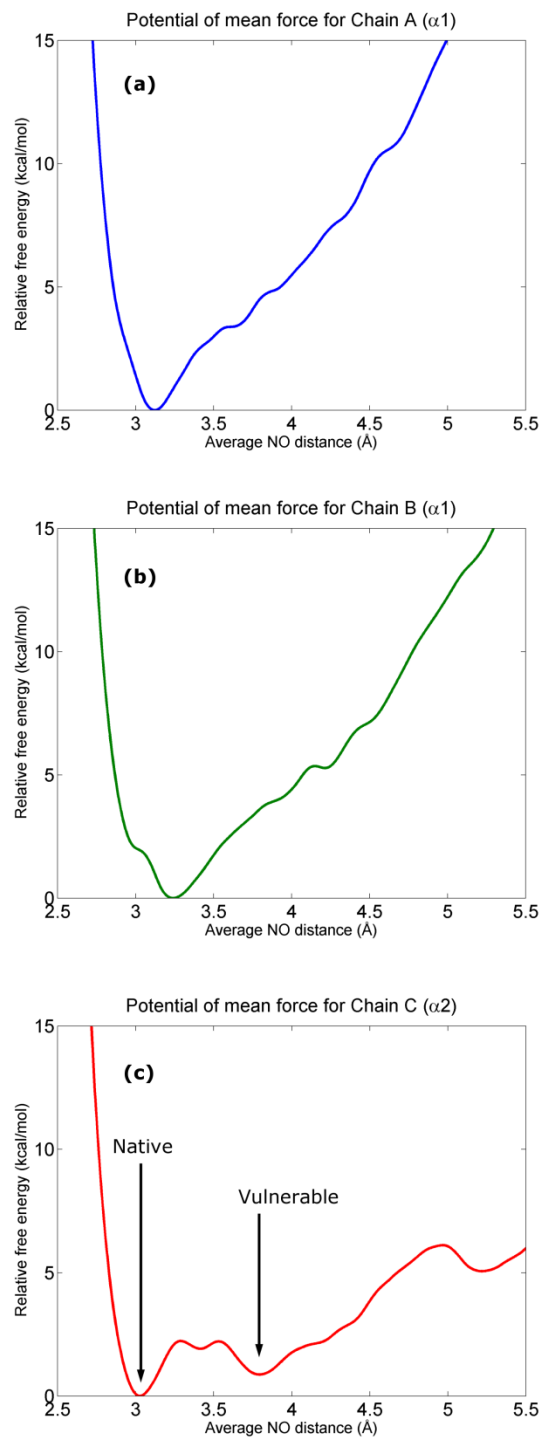
Figure 3-3: Potential of mean force for (A) chain A (α1), (B) chain B (α1), and (C) chain C (α2) of type I collagen.

3.2.2 Representative structures of the native and vulnerable states

Representative structures from the α2 chain global energy minimum (the "native" state), which occurs at an average NO distance of 3 Å, resemble the familiar triple-helical structure of collagen (Figure 3-4A). Although two NO distances in the vicinity of the triplet that contains the scissile bond are slightly above 3.6 Å, most hydrogen bonds in the vicinity of the scissile bond have NO distances below 3.3 Å (Figure 3-4B). Overall, these data argue that the triple-helical structure is preserved about the collagenase cleavage site.
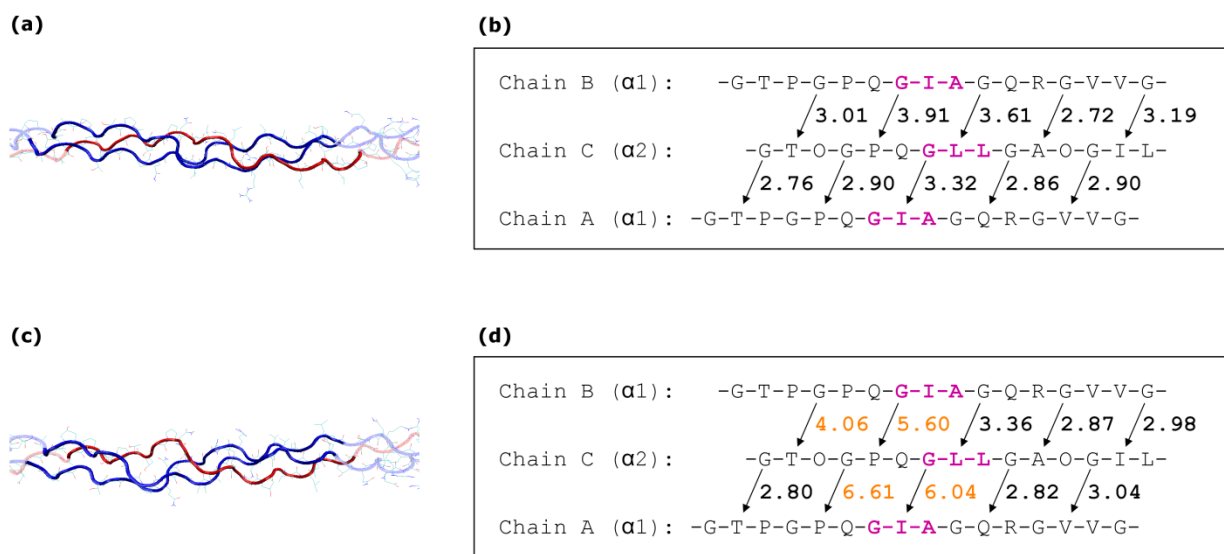
Figure 3-4: (A) Representative structure from the native state of type I collagen. The α2 chain is colored in red. (B) Interchain NO distances of the native state of type I collagen. The triplets containing the scissile bonds are colored in magenta. (C) Representative structure of the vulnerable state of type I collagen. (D) Interchain NO distances of the vulnerable state of type I collagen. NO distances greater than 4 Å are highlighted in orange.

Representative structures corresponding to the energy minimum at 3.8 Å, however, are partially unfolded in the region N-terminal to the collagenase cleavage site (Figure 3-4C). Moreover, the relatively large NO distances involving the scissile bond and triplets N-terminal to the cleavage site, suggest that H-bonds are broken in this region (Figure 3-4D). In particular, the GPQ triplet of the α2 chain that is immediately N-terminal to the central scissile bond-containing triplet

(GLL) neither donates nor accepts hydrogen bonds from other chains (Figure 3-4D). In addition, the GLL triplet, which contains the scissile bond, no longer donates a hydrogen bond to an α1 chain (Figure 3-4D). We therefore refer to structures corresponding to this minimum as belonging to the vulnerable state (Figure 3-3C). This classification is based on our previous formalism that defined vulnerable conformations as states have partially unfolded structures in the vicinity of the collagenase scissile bond (Stultz 2002, Stultz and Edelman 2003).
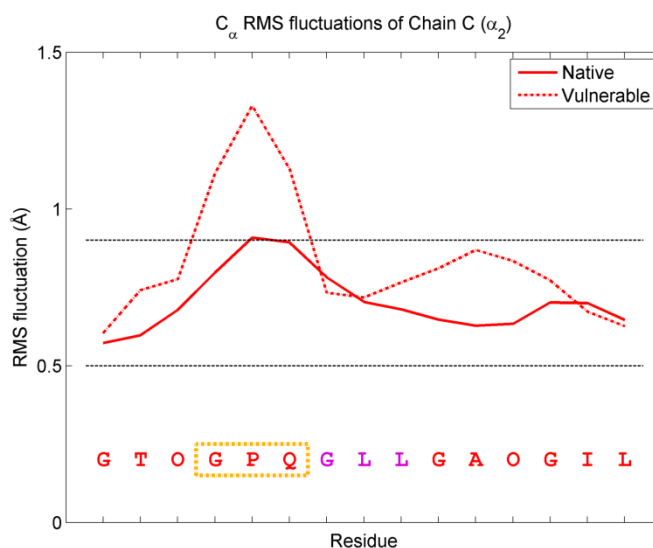
Figure 3-5: The RMS fluctuations of the α2 chain (chain C) backbone C_α atoms in the native and vulnerable states of type I collagen. The triplet containing the scissile bond is colored in magenta. In the vulnerable state, the triplet boxed in orange does not hydrogen bond to any residues in the adjacent α1 chains. Black dashed lines indicate the range of fluctuations within two standard deviations of mean fluctuation in the native state.

3.2.3 Dynamics of the native and vulnerable states

To characterize the effect that disruption in hydrogen bonding pattern has on the dynamics of the structure in the vicinity of the collagenase cleavage site, we calculated the root-mean-square (rms) fluctuations of backbone Cα atoms in both the native and the vulnerable states for the α2 chain (Figure 3-5). In the native state, the GPQ triplet, immediately N-terminal to the scissile bond triplet, has mildly elevated fluctuations – a finding consistent with the observation that one of the NO distances of this triplet is elevated (Figure 3-4B). However, rms fluctuations are

increased in the vulnerable state, relative to fluctuations in the native state. In the vulnerable state, the NO distances of the GPQ triplet are significantly elevated and consequently, the associated rms fluctuations are significantly larger (Figure 3-5). As the GPQ triplet does not hydrogen bond to any residues in adjacent α1 chains, it has the largest rms fluctuations and the greatest flexibility.

3.3 The effect of hydroxylation near the cleavage site

Our data suggest that unfolding of the α2 chain in the vicinity of the collagenase cleavage site is energetically favorable relative to unfolding of α1 chains in type I collagen. This is somewhat surprising given that the α2 chain contains more hydroxyproline residues than the α1 chain near the scissile bond triplet. In particular, we model the X2 position of a G-X1-X2 triplet, which is N-terminal to the scissile bond in α1 chains, as containing a proline instead of a hydroxyproline while the corresponding position in the α2 chain contains a hydroxyproline (yellow residues in Figure 3-1). Since collagen stability typically increases as the hydroxyproline content increases, this finding is at first counterintuitive (Jenkins and Raines 2002, Brodsky and Persikov 2005, Engel and Bächinger 2005). Therefore, to further explore the role that hydroxylation at this site has on the conformational thermodynamics of collagen, we conducted additional unfolding simulations with this proline in the α1 chain replaced with a hydroxyproline. We henceforth refer to this mutant sequence as the GTO mutant.

3.3.1 Potentials of mean force for unfolding

The resulting pmfs of all three chains in the GTO mutant have global energy minima with an average NO distance less than 3.6 Å (Figure 3-6). In particular, the vulnerable state that was previously observed in unfolding simulations of the α2 chain, no longer corresponds to a stable state in the α2 unfolding simulations with the GTO mutant (Figure 3-6C). Hence, α2 chain unfolding does not occur in the mutant structure. This is an interesting finding in light of the fact that the GTO mutation only occurs on α1 chains and not on α2 chains.
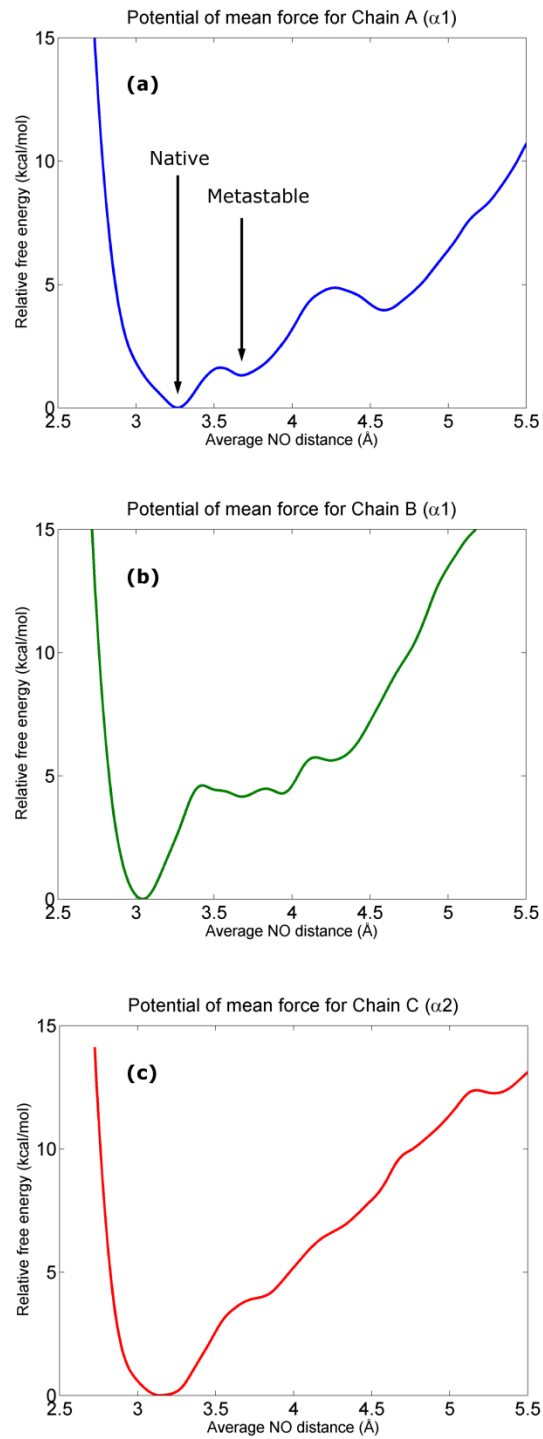
Figure 3-6: Potential of mean force for (a) chain A (α1), (b) chain B (α1), and (c) chain C (α2) of GTO mutant type I collagen.

While the pmf for α2 chain unfolding now has one distinct minimum, the pmf for the α1 chains, each have two minima. One α1 chain has a shelf-like broad minimum in the range of 3.4-4 Å (Figure 3-6B), which occurs at a relative free energy of 4.2 kcal/mol; i.e., approximately 7 RT at 300 K. Consequently, it is unlikely that this state is significantly sampled at room temperature. The other α1 chain has minima at 3.7 Å and 4.6 Å (Figure 3-6A). The minimum at 4.6 Å has a relative free energy of 4.0 kcal/mol, which again suggest that this state is rarely sampled at room temperature. The minimum at 3.7 Å has a relative free energy of 1.3 kcal/mol (~2 RT), suggesting that this metastable state may correspond to an accessible state at room temperature (Figure 3-6A). Therefore to explore the structure and dynamical properties of these states we analyzed representative structures from these minima.

(a)

```
Chain C (α2):        -G-T-O-G-P-Q-G-L-L-G-A-O-G-I-L-
                     /2.85 /3.08 /2.95 /3.01 /5.09
Chain A (α1): -G-T-O-G-P-Q-G-I-A-G-Q-R-G-V-V-G-
                     /2.86 /2.96 /2.91 /3.12 /2.92
Chain B (α1):    -G-T-O-G-P-Q-G-I-A-G-Q-R-G-V-V-G-
```

(b)

```
Chain C (α2):        -G-T-O-G-P-Q-G-L-L-G-A-O-G-I-L-
                     /5.31 /4.37 /2.94 /3.15 /6.56
Chain A (α1): -G-T-O-G-P-Q-G-I-A-G-Q-R-G-V-V-G-
                     /3.02 /2.98 /3.02 /2.90 /2.92
Chain B (α1):    -G-T-O-G-P-Q-G-I-A-G-Q-R-G-V-V-G-
```
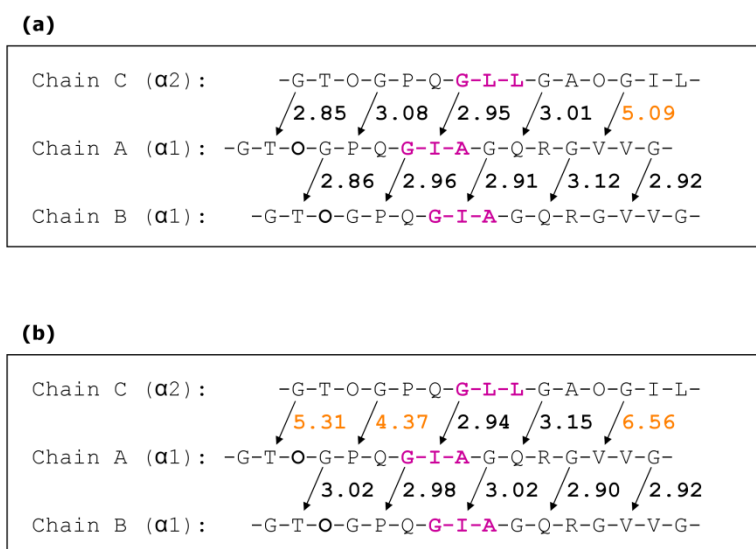
Figure 3-7: Interchain NO distances of the (A) native and (B) metastable states of GTO mutant type I collagen. The triplets containing the scissile bonds are colored in magenta. Broken hydrogen bonds (NO distances greater than 4 Å) are shown in orange. In both states, no triplets have broken donor and acceptor bonds, and the scissile bond triplet remains hydrogen bonded to both neighboring chains.

3.3.2 Representative structures of the native and metastable state

Structures corresponding to the lowest energy state with an average NO distance of 3.3 Å adopt a triple helical structure with mostly preserved interchain hydrogen bonds (Figure 3-7A). H-bonds are preserved near the scissile bond triplet, and triplets N-terminal to the scissile bond triplet all have NO distances below 3.1 Å (Figure 3-7A). One H-bond, which is C-terminal to the scissile bond triplet is broken in this structure, having an NO distance of ~5 Å (Figure 3-7A). Representative structures of the metastable state contain broken hydrogen bonds both N-terminal and C-terminal to the scissile bond (Figure 3-7B). In both the native and metastable states, the central triplet that contains the scissile bond maintains both hydrogen bonds. Additionally, although the metastable state has more broken hydrogen bonds relative to the native state, no triplet is dissociated from both of its neighboring chains. In contrast to the vulnerable state, associated with α2 chain unfolding (Figure 3-4D), none of the triplets in the α1 metastable state have both H-bonds broken (Figure 3-7B).
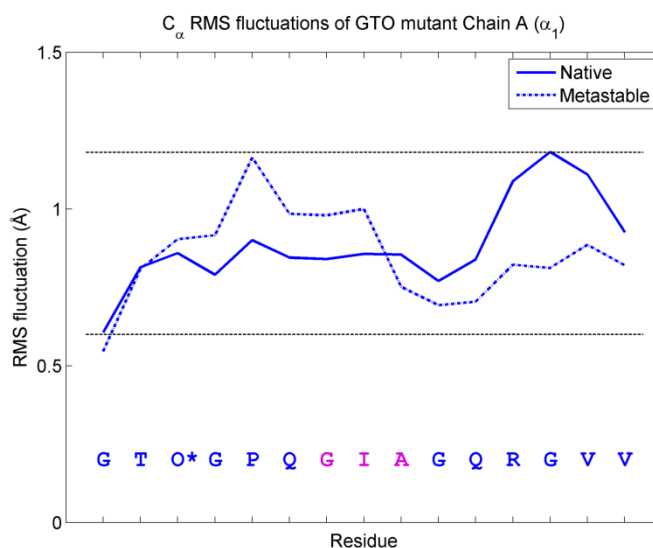
Figure 3-8: RMS fluctuations of the α1 chain (chain A) backbone C_α atoms in the native and metastable states of GTO mutant type I collagen. The hydroxyproline that has been substituted for proline is denoted O*. The triplet containing the scissile bond is colored in magenta. Black dashed lines indicate the range of fluctuations within two standard deviations of the native state mean fluctuation.

3.3.3 Dynamics of the native and metastable states

To examine the effect of this hydrogen bonding pattern on the dynamics of the chain, we computed rms fluctuations of the α1 chain backbone Cα atoms for the native and metastable states of the GTO mutant (Figure 3-6A). Both the native and metastable states have a similar range of rms fluctuations (Figure 8). In the native state, the rms fluctuations of the GVV triplet, which is near the C-terminus, has slightly increased fluctuations (Figure 3-8) – an expected finding given that a hydrogen bond is broken near the C-terminus (Figure 3-7B). This triplet, however, has reduced rms fluctuations in the metastable state (Figure 3-8) – a finding which is likely due to the fact that it has a smaller hydrogen bond distance with neighboring chain B. In the metastable state, most of the rms fluctuations are elevated in the N-terminal region. This is consistent with the observation that two H-bonds are broken N-terminal to the central scissile bond in the metastable state.

3.4 Discussion

The molecular mechanism of collagenolysis is an enigma. Studies aimed at deciphering the structural changes needed to hydrolyze scissile bonds in collagen help to elucidate the complex series of events that must occur to enable collagen degradation. To deduce structural changes in type I collagen that may facilitate collagenase recognition and cleavage of scissile bonds in collagen, we explored the conformational free energy landscape of collagen-like peptides that model regions near scissile bonds in collagen.

Although the DNA sequences of human α1 and α2 collagen chains are known, it is difficult to deduce the precise amino-acid sequence of collagen chains from these data because collagen undergoes extensive post-translational modifications (Gelse et al. 2003). Most notably, a *sine qua non* of collagen post-processing is the hydroxylation of a significant number of proline residues in the X2 position of G-X1-X2 triplets. Since deciphering the precise hydroxylation pattern is not possible from the DNA sequence alone, the exact amino-acid sequences of human α1 and α2 chains have yet to be experimentally determined. However, as the hydroxylation pattern of proline residues in bovine and chick collagen are known, we used the significant DNA sequence homology between bovine and chick sequences to deduce the hydroxylation pattern of the corresponding human sequences (Bornstein and Traub 1979, Dixit et al. 1979, Highberger et al. 1982, Glanville et al. 1983).

Interestingly, the alignment of sequences from the bovine and chick sequences suggest that a proline residue located in the X2 position of an upstream triplet in the α1 chain is not hydroxylated. Collagen unfolding simulations using an amino-acid sequence having a proline in this position suggest that unfolding of the α2 chain near the collagenase cleavage site is energetically favored relative to unfolding α1 chains. More precisely, we find that α2 chains can adopt a partially unfolded state having an energy similar to that of the native triple-helical conformation.

This partially unfolded, vulnerable, state has a number of interesting structural properties. As a rule, every G-X1-X2 triplet in the α2 chain of native type I collagen donates a hydrogen bond to one α1 chain and accepts a hydrogen bond from the other α1 chain (Brodsky and Persikov 2005, Engel and Bächinger 2005). However, in the vulnerable state, the α2 triplet that contains the scissile bond does not donate a hydrogen bond to a neighboring α1 chain and consequently only forms one hydrogen bond with another α1 chain. In addition, the triplet which is immediately N-terminal to the scissile bond does not form any hydrogen bonds with neighboring α1 chains; i.e., this triplet dissociates from the other two chains. The loss of both hydrogen bonds is associated with increased chain flexibility, as demonstrated by an increase in the rms fluctuations of C_α atoms in residues belonging to this triplet. An increase in backbone fluctuations suggests that the formation of interchain hydrogen bonds helps to restrict the motion of collagen chains, and therefore local disruption of the interchain hydrogen bonding may constitute a mechanism for imparting increased structurally lability to localized regions of the collagen structure. Overall, our data support the hypothesis that hydrogen bonding between adjacent chains has a direct effect on chain flexibility.

3.4.1 A critical non-hydroxylation near the cleavage site

Our results were obtained using an amino acid sequence where a proline residue in the X2 position of a triplet upstream from the cleavage site is not hydroxylated. To explore the effect that this assumption has on our results, we conducted additional unfolding simulations where this proline residue is replaced with a hydroxyproline. Interestingly, we find that the hydroxylation of this proline residue, which is found in α1 chains, leads to a dramatic change in the conformational free energy profile of the α2 chain. The vulnerable state of the α2 chain disappears when the prolines on the α1 chains are hydroxylated. These data therefore suggest

that hydroxylation of residues on the α1 chains can affect the conformational thermodynamics on the α2 chain, and more specifically, that the hydroxyproline residues can have stabilizing effects on adjacent chains.

It has been demonstrated that the stabilizing effect of hydroxyproline on collagen is mainly stereoelectronic in nature (Holmgren et al. 1999, Bretscher et al. 2001). Specifically, the hydroxyl group of hydroxyproline stabilizes the pyrrolidine ring in the Cγ-*exo* pucker, which in turn stabilizes both the φ/ψ and ω (peptide bond) angles in a conformation that is compatible with the triple-helical structure (Jenkins and Raines 2002, Brodsky and Persikov 2005, Engel and Bächinger 2005). Given this, hydroxyproline mediated stabilization of α1 chains allows other chains to form stabilizing contacts and hydrogen bonds to these restricted α1 chains. In fact, while our GTO mutant is able to adopt a metastable state, all of the triplets in this state form at least one H-bond to a neighboring chain. Additionally, it has been argued that hydroxyproline residues can form stabilizing dipole-dipole interactions with other hydroxyproline residues on adjacent chains (Improta et al. 2008). Hence hydroxyproline residues may directly interact with other chains, in a manner that does not involve the formation of hydrogen bonds, to increase the stability of adjacent chains. Overall, our data are consistent with these notions and suggest that non-hydroxylation of a proline residue may play a crucial role in determining the conformational flexibility and dynamics in the vicinity of the collagenase cleavage site.

3.4.2 Interpreting prior experimental results in light of the α2 chain vulnerable state

Incubating type I collagen with inactive MMPs protects the α2 chain from peptide bond hydrolysis by other proteases that hydrolyze the MMP-specific scissile bond, thereby suggesting that MMPs bind to α2 chains in the vicinity of the collagenase cleavage site (Chung et al. 2004). Localized unfolding and increased flexibility of the triplet upstream from the collagenase cleavage site may provide a mechanism that enables MMPs to recognize regions near the cleavage site. As such, our data helps to explain and clarify these experimental observations.

A more recent study utilized x-ray fiber diffraction of collagen fibrils at room temperature to deduce structural properties of type I collagen *in situ* (Orgel et al. 2006, Perumal et al. 2008). Using the electron density obtained from experiments on fibrillar collagen, a relaxed, energy-minimized, model of collagen was constructed and compared to an idealized triple-helical structure (Perumal et al. 2008). An analysis of these structures suggests that the structure of

fibrillar collagen significantly differs from the idealized triple-helical structure in a region of the α2 chain that is in the vicinity of the collagenase-scissile bond; i.e., the α2 chain is more dissociated from the center of the triple-helix relative to the other α1 chains (Perumal et al. 2008). Additionally, residues immediately upstream from the scissile bond in the α2 chain are more dissociated from the center of the triple-helix than residues immediately downstream from the scissile bond. These data are consistent with our results that were not obtained with a fibrillar model of collagen. As a result, our findings are consistent with the notion that preferential unfolding of the α2 chain in the vicinity of the collagenase cleavage site is an inherent property of the local sequence of collagen near the cleavage site and not a function of additional contacts that may be present in the collagen fibril.

Recent experiments performed at 37 °C using type I collagen and MMP-8 suggest similar catalytic rates for the α1 and α2 chains, implying that collagenases do not strongly prefer one sequence over the other (Gioia et al. 2007). However, as type I collagen is known to be thermally unstable at 37 °C and particularly flexible in the region of the cleavage site (Leikina et al. 2002, Makareeva et al. 2008), it is likely that all three chains exist in a partially unfolded state with disrupted hydrogen bonds at this higher temperature. Although CD measurements indicate that some fraction of triple-helical structure exists in type I collagen at 37 °C (Gioia et al. 2007), CD spectroscopy can rarely distinguish small localized changes in structure. Furthermore, data from NMR experiments and computational simulations of collagen-like model peptides suggest that regions near the collagenase cleavage site are relatively unstable and may be the first to unfold (Fan et al. 1993, Fiori et al. 2002, Stultz 2002, Stultz and Edelman 2003). If all three chains are equally unfolded at 37 °C, then the inability of MMP-8 to distinguish between the different chains in type I collagen is expected.

Similar experiments with type I collagen and MMP-2, again at 37 °C, imply that gelatinases may prefer binding to the α1 chains, therefore cleaving the scissile bond of the α2 chain with a higher catalytic rate (Gioia et al. 2007). When evaluating the results of experiments utilizing MMP-2, however, it is worth noting that MMP-2 contains a fibronectin type II-like collagen binding domain (CBD) in addition to the catalytic and hemopexin-like domains found in the fibrillar collagenases (Overall 2002). In addition, there are data experimental data which imply that the CBD, and not the hemopexin-like domain, is primarily responsible for MMP-2 binding to collagen (Tam et al. 2004). Hence the preference of MMP-2 for particular exposed collagen

chains will be dictated by inherent preferences that this domain may have for a given collagen chain.

Overall our data are consistent with the notion that incubating type I collagen with catalytically inactive MMP-1 leads to binding near the collagenase cleavage site. Prior work suggests that collagenases bind to partially unfolded states and stabilize vulnerable conformers in a manner that can facilitate collagenolysis (Nerenberg et al. 2008). Similarly, inactive MMP-1 may bind to the α2 chain and stabilize partially unfolded states having disrupted interchain hydrogen bonds, in turn promoting peptide bond hydrolysis of scissile bonds in the α1 chains by other proteases. Our data offer a detailed picture into the mechanism of collagen degradation and highlight the importance of subtle sequence variations in the vicinity of the collagenase cleavage site.

Chapter 4

Conformational Selection in Type I Collagen Degradation

This chapter is adapted from: Nerenberg PS*, Salsas-Escat R*, and Stultz CM. Cleavage site specificity and conformational selection in type I collagen degradation. Submitted. (*Both authors contributed equally to this work.)

In Chapter 3 we generated a structural model for the collagenase cleavage site of type I collagen using MD simulations. These data suggest that type I collagen exists in two distinct conformations: a well-folded triple-helical (native) state and a partially unfolded (vulnerable) state. To validate this structural model, we wish to conduct degradation experiments using deletion mutants of collagenases that contain only the catalytic domain, as these enzymes are thought to be a probe for unfolded states (see Chapter 2). In addition, we wish to test how well the resulting degradation time courses can be fit using the conformational selection model described in Chapter 2. Thus, degradation experiments provide any avenue not only to validate our structural model of type I collagen, but also to test the applicability of our conformational selection model.

Data obtained from circular dichroism and differential scanning calorimetry experiments suggest that there is considerable heterogeneity in the stability of the triple helix along the collagen chain (Makareeva et al. 2008). Moreover, NMR studies involving collagen-like model peptides in solution and molecular dynamics (MD) simulations suggest that at low temperatures (i.e., temperatures below collagen's melting point) regions of type I collagen near the collagenase cleavage site can adopt conformations that have relatively solvent-exposed scissile bonds (Fan et al. 1993, Fiori et al. 2002, Nerenberg and Stultz 2008).

A number of experiments have explored whether collagen can adopt conformations at low temperatures that could, in principle, be recognized and cleaved by proteases. General proteases like pepsin, chymotrypsin and trypsin have been used in assays for partially unfolded states of collagen (Bruckner and Prockop 1981). Early experiments incubated type I collagen with trypsin

and no collagen degradation was observed (Miller et al. 1976). More recent experiments incubated type I collagen with high concentrations of the catalytic domain of MMP-8 (CMMP-8) at room temperature and, again, no degradation was found (Schnierer et al. 1993). Similar results have been reported with the catalytic domain of MMP-1 (CMMP-1) (Chung et al. 2004). If the region near the collagenase cleavage site is able to spontaneously adopt partially unfolded states in solution that can bind 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 the aforementioned studies that suggest that type I collagen adopts partially unfolded states in the vicinity of the cleavage site in solution. These latter degradation experiments support a theory that dictates that 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 lead to active unwinding of the triple-helical structure (Overall 2002, Chung et al. 2004, Tam et al. 2004, Gioia 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 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 the unique cleavage site is recognized by the collagenase.

In this work we demonstrate for the first time that the degradation of type I collagen at temperatures 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 only contain the catalytic domain occurs at the unique collagenase cleavage and not at other potential cleavage sites. Thus both peptide bond hydrolysis and enzyme specificity is achieved with the catalytic domain alone. As full length enzyme is thought to be necessary for collagenase-mediated unwinding, our data unambiguously demonstrate that enzyme-mediated unwinding is not required for collagenolysis *in vitro*. We therefore analyze our data in light of a conformational selection where thermal fluctuations at the cleavage site cause collagen to adopt unfolded conformations that are complementary to the collagenase catalytic site. Overall, our

findings suggest that type I collagen can adopt locally unfolded states at room temperature and that collagenolysis occurs when collagenases cleave these locally unfolded states.

4.1 Experimental and computational methods

Acknowledgement: The degradation experiments and gel densitometry described in sections 4.1.1 and 4.1.2 were performed by my labmate Ramon Salsas-Escat.

4.1.1 Degradation experiments

All degradation reactions were performed at room temperature. The temperature was verified using a calibrated thermometer. All recorded temperatures were 24 °C and no variation in the temperature was observed. All experiments were carried out in TNC buffer, containing 100 mM Tris HCl (VWR International), 10 mM $CaCl_2$ (Sigma-Aldrich Co), pH 7.6. Bovine type I collagen, (BD Biosciences), was obtained at 3 mg/ml in 0.012 M HCl. MMP mutants that only contain the catalytic domain (CMMP-1 and CMMP-8) were purchased from Enzo Life Sciences. Full length MMP-1 (FMMP-1) was purchased from Anaspec and was provided in a preserving solution containing 1 mg/ml of BSA.

As purification of type I collagen often results in protein that is contaminated with type III collagen, purchased collagen samples were repurified using differential salt precipitation (Epstein 1974, Trelstad et al. 1976, Miller and Rhodes 1982). First type I collagen was diluted to a concentration of 0.5 mg/ml in 0.5 M AcOH. The solution was then dialyzed against low salt buffer (0.1 M NaCl, 50 mM Tris, pH 7.5, at 4 °C) followed by dialysis against a high salt buffer (1.8M NaCl, 50 mM Tris, pH 7.5, at 4 °C), in which type III collagen preferentially precipitates (Trelstad et al. 1976). At this point, the sample was centrifuged at 4 °C for 30 min at 16000g and the supernatant, containing purified type I collagen, was then dialyzed into TNC buffer. Purity of the final type I solution was confirmed by running degradation experiments with full length MMP-1 and no type III collagen degradation products were observed.

Degradation experiments employed purified type I collagen at a concentration of 150 µg/ml. All enzymes were incubated with 4-aminophenyl mercuric acetate (APMA, Sigma-Aldrich Co.) as previously described (Clark 2001) and mixed with type I collagen to a final concentration of 25 µg/ml for CMMP-8 and 40 µg/ml CMMP-1. Reactions were stopped by the addition of SDS-Laemmli buffer (BioRad Laboratories) with β-Mercapto Ethanol (Sigma-Aldrich Co.) and

boiled for 10 minutes. The degradation products were run in 4-12% gradient gels and stained with Coomassie colloidal blue. Gelatin was generated by boiling type I collagen in TNC buffer for 10 minutes. Gelatin at 150 µg/ml was then incubated with 13.8 µg/ml CMMP-8 and 16.6 µg/ml CMMP-1.

An unstained CMMP-8 gel was transferred to a PVDF membrane, which was stained with Coomassie blue. The ¼ α1 and ¼ α2 bands were cut from the gels corresponding to the CMMP-8 experiments and sent for sequencing to the Tufts University Core Facility, using an ABI 494 Protein Sequencer. The ¼ α1 and α2 bands from the CMMP-1 reaction could not be sequenced due to the low amount of collagen that is degraded by CMMP-1.

4.1.2 Densitometry analysis

Densitometry was performed with a Kodak Gel Logic 100 Imaging System. Bands were imaged using an automatic lane and band fitting method (Kodak Molecular Imaging Software v4.0.0). The percentage of type I collagen degradation by CMMP-8 or CMMP-1 was 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}, α1, α2, and the ¼ and ¾ bands). γ and β bands correspond to N-terminally crosslinked collagen molecules (Veis and Anesey 1965, French et al. 1987, Shigemura et al. 2004, Morimoto et al. 2009). The γ bands correspond to a trimer of chains (Shigemura et al. 2004, Morimoto et al. 2009). The β bands correspond to a dimer with two α1 chains (β11) or one α1 and one α2 chains (β12), with only one crosslink (French et al. 1987, Shigemura et al. 2004, Morimoto 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.

4.1.3 Numerical simulations using conformational selection model

As in Chapter 2, the reaction scheme shown in Figure 4-1 was converted into a set of ordinary differential equations (ODEs), with the equilibrium and binding constants expressed in terms of various rate constants: $K_{eq} = \dfrac{k_1}{k_2}$, $K_{bind}^{NC} = \dfrac{k_{on}^{NC}}{k_{off}^{NC}}$, and $K_{bind}^{VC} = \dfrac{k_{on}^{VC}}{k_{off}^{VC}}$.

We then constructed a set of ODEs that describe the time evolution of each species:

$$\frac{d[N]}{dt} = -k_1[N] + k_2[V] - k_{on}^{NC}[N][C] + k_{off}^{NC}[N \cdot C] \tag{4.1}$$

$$\frac{d[V]}{dt} = -k_2[V] + k_1[N] - k_{on}^{VC}[V][C] + k_{off}^{VC}[V \cdot C] \tag{4.2}$$

$$\frac{d[N \cdot C]}{dt} = -k_{off}^{NC}[N \cdot C] + k_{on}^{NC}[N][C] \tag{4.3}$$

$$\frac{d[V \cdot C]}{dt} = -k_{off}^{VC}[V \cdot C] + k_{on}^{VC}[V][C] - k_{cat}[V \cdot C] \tag{4.4}$$

$$\frac{d[P]}{dt} = k_{cat}[V \cdot C] \tag{4.5}$$

$$\frac{d[C]}{dt} = -k_{on}^{NC}[N][C] + k_{off}^{NC}[N \cdot C] - k_{on}^{VC}[V][C] + k_{off}^{VC}[V \cdot C] + k_{cat}[V \cdot C] \tag{4.6}$$
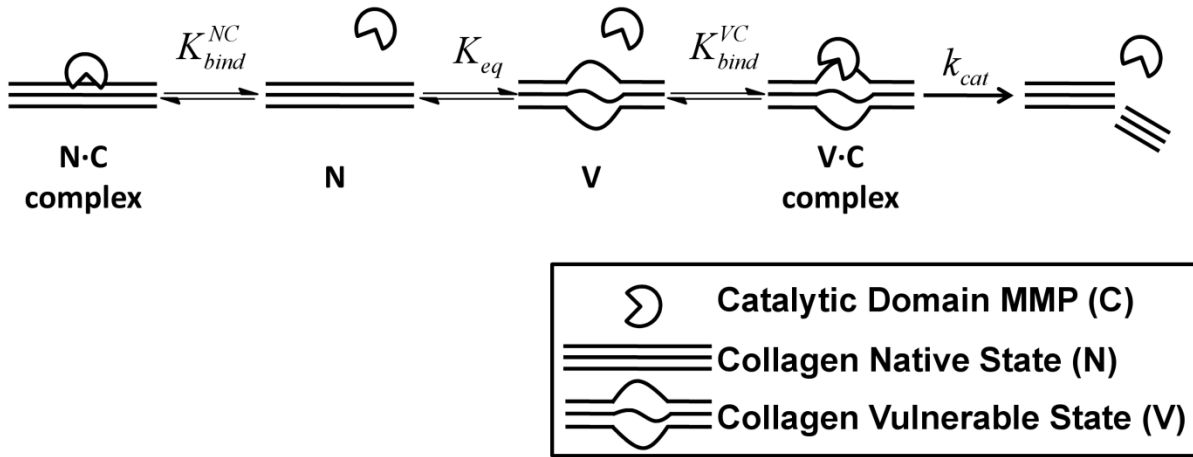
Figure 4-1: A conformational selection mechanism for collagenolysis with the MMP catalytic domains. Collagen exists in an equilibrium between native (N) and vulnerable (V) states with the equilibrium determined by K_{eq}. The catalytic domain of MMPs (C) interacts in a non-specific manner with the native state with binding constant K_{bind}^{NC}, yielding the $N\cdot C$ complex. C binds to the vulnerable state, V, state with binding constant K_{bind}^{VC}, forming the $V\cdot C$ complex. The $V\cdot C$ complex is then degraded with catalytic rate k_{cat}.

We estimated the native state binding rate constants, k_{on}^{NC} and k_{off}^{NC} (and therefore K_{bind}^{NC}), using previously determined constants for MMP-1 binding to a heterotrimeric collagen-like peptide that contains the type I collagen collagenase cleavage site (Trimer D) 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 (Ottl et al. 2000). Previous studies with a related reaction scheme, however, demonstrated that the model is not sensitive to the choice of K_{bind}^{NC} (Nerenberg et al. 2008).

The catalytic rate constant of the vulnerable state, k_{cat}, for CMMP-1 was estimated from the experimentally measured catalytic rates of MMP-1 degrading gelatin (Welgus et al. 1982, Nerenberg et al. 2008). Because similar data does not exist for CMMP-8, we set lower and upper bounds for the catalytic rate based on previously published data. The lower bound of k_{cat} for CMMP-8 was set equal to that of CMMP-1, as it is known that MMP-8 has a greater catalytic rate than MMP-1 for the same substrate and at the same temperature (Welgus et al. 1982, Gioia et al. 2002). Upper bounds for the catalytic rate of CMMP-8 were obtained from experimentally measured rate constants of FMMP-8 degrading linear collagen-like peptides (Netzel-Arnett et al. 1991). This approach yielded a range of k_{cat} for CMMP-8 spanning 0.11-11.1 s^{-1}.

To convert a given K_{eq} into the appropriate conformational transition rate constants, k_1 and k_2, we used initial rate constant values, k_1^{init} and k_2^{init}, equal to 10^6 s^{-1} (i.e., we assumed microsecond conformational transitions) and then multiplied them by a scale factor α according to:

$$\left. \begin{array}{l} k_1 = \alpha^{1/2} k_1^{init} \\ k_2 = \alpha^{-1/2} k_2^{init} \end{array} \right\} \quad K_{eq} = \frac{k_1}{k_2} = \alpha \frac{k_1^{init}}{k_2^{init}}$$

In this way, the appropriate rate constants could be determined for any given value of K_{eq}. We used a similar approach to determine the vulnerable state binding rate constants, k_{on}^{VC} and k_{off}^{VC}, from a given K_{bind}^{VC}. The initial rate constants for binding the vulnerable state, $k_{on}^{VC,init}$ and $k_{off}^{VC,init}$ were set equal to experimentally measured on and off rates for MMP-1 binding to a collagen-like model peptide that is largely unfolded at room temperature (Ottl et al. 2000).

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:

$$\text{RMSE} = \sqrt{n^{-1} \sum_{t=1}^{n} \left(F_{model}(t) - F_{exp}(t) \right)^2} \tag{4.7}$$

where n is the number of experimental time points.

4.2 Results of degradation experiments

In a prior work we introduced a conformational selection model in which 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 et al. 2008). MMPs can bind to either state, but collagenolysis occurs only when MMPs bind to vulnerable states. The model is based on the premise that collagens adopt 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 et al. 2008). To test this, we exposed type I collagen to both the catalytic domains of MMP-8 and MMP-1 (denoted as CMMP-8 and CMMP-1, respectively).

Bovine type I collagen was incubated with high concentrations of CMMP-8 at room temperature – a temperature well below the melting temperature of type I collagen (Privalov 1982, Leikina et al. 2002). After 48 h, type I collagen degradation was observed in solutions

containing CMMP-8 (Figure 4-2A, 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 MMP-1 (FMMP-1, Figure 4-2A, lane 9). To demonstrate that our results are not specific to MMP-8, we exposed type I collagen to a deletion mutant containing only the catalytic domain of MMP-1 (Figure 4-2B, lanes 1-5). The resulting cleavage pattern is the same as that observed with CMMP-8, suggesting that CMMP-1 recognizes and cleaves the same site, albeit higher concentrations of enzyme and longer incubation times were required to observe degradation with CMMP-1. N-terminal amino acid sequencing of the ¼ α1 and ¼ α2 bands confirms that the CMMP-8 deletion mutant cleaves at the unique cleavage site recognized by wild-type enzyme. The ¼ α1 and α2 bands from the CMMP-1 reaction could not be sequenced due to the low amount of collagen that is degraded by CMMP-1. The cleavage pattern by CMMP-1, as in the case of CMMP-8, corresponds to the same ¼ and ¾ fragments resulting from incubation with FMMP-1 (Figure 4-2B, lane 7). 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 (Bornstein and Traub 1979). If these potential sites were cleaved, the pattern of fragments observed would be significantly different from the FMMP-1 control. For this reason, we believe that CMMP-1 mediated cleavage, as is the case with CMMP-8, is also occurring at the true collagenase cleavage site.

To demonstrate that our results are not 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 enzyme. SDS-PAGE of the solutions after 6 days of incubation did not exhibit any degradation bands (data not shown). In addition, to confirm that our findings are not due to contamination of our collagen samples with unfolded type I collagen chains (gelatin), we incubated CMMP-8 and CMMP-1 with gelatin. Both CMMP-8 and CMMP-1 cleave gelatin at several sites yielding multiple degradation bands on SDS-PAGE (Figure 1A lane 8 and Figure 1B lane 6) (Welgus et al. 1982, Chung et al. 2004). Since these bands are not seen when CMMP-8 and CMMP-1 are incubated with collagen, contamination of our collagen sample with unfolded collagen chains does not explain our results.
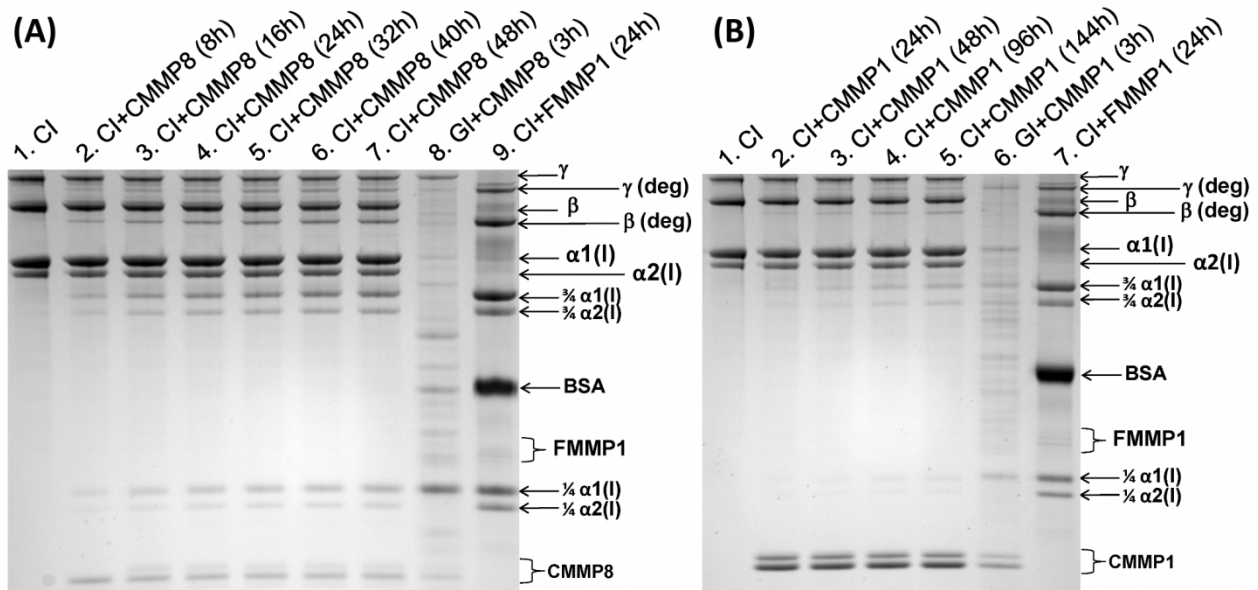
Figure 4-2: Degradation profiles of type I collagen at room temperature with CMMP-8 and CMMP-1. Intact type I collagen (CI) bands include monomeric α1(I) and α2(I) bands, dimeric β bands and trimeric γ bands. The β and γ aggregates correspond to N-terminally crosslinked collagen molecules (Veis and Anesey 1965, French et al. 1987, Shigemura et al. 2004, Morimoto et al. 2009). CI degradation bands include α1(I) and α2(I) ¾ and ¼ fragments, and degradation of crosslinked chains, β$_{deg}$ and γ$_{deg}$. (A) *Lane 1:* CI (150 µg/ml). *Lanes 2 to 7:* CI (150 µg/ml) incubated with the catalytic domain of MMP-8 (CMMP-8) (25 µg/ml) for 8, 16, 24, 32, 40 and 48 h, respectively. *Lane 8:* Type I collagen gelatin (GI) (150 µg/ml) incubated with CMMP-8 (13.8 µg/ml) for 4 h. *Lane 9*: Type I collagen (150 µg/ml) incubated with full length MMP-1 (FMMP-1) (1.3 µg/ml) for 24 h. This lane contains a bovine serum albumin (BSA) band since FMMP-1 is supplied in a buffer containing 1 mg/ml BSA. (B) CI incubated with CMMP-1 *Lane 1:* CI (150 µg/ml); *Lanes 2 to 5:* CI (150 µg/ml) incubated with the catalytic domain of MMP-1 (CMMP-1) (40 µg/ml) for 24, 48, 96 and 144 h, respectively. *Lane 6:* Type I collagen gelatin (GI) (150 µg/ml) incubated with CMMP-1 (16.6µg/ml) for 4 h. *Lane 7*: Type I collagen (150 µg/ml) incubated with full length MMP-1 (FMMP-1) (1.3 µg/ml) for 24 h. This lane contains a BSA band since FMMP-1 is supplied in a buffer containing 1 mg/ml BSA.

4.3 Interpreting type I collagen degradation experiments using the conformational selection model

Our data demonstrate that peptide bond hydrolysis at the unique collagenase cleavage site can occur *in vitro* with only the catalytic domain of MMPs. Since CMMP-8 and CMMP-1 cleave unfolded collagen at multiple sites but triple-helical collagen at only one site, we interpret our findings in light of a model where local unfolding of collagen at the unique cleavage site enables collagenases to gain access to their corresponding scissile bonds (Nerenberg et al. 2008). In this model, collagen can exist in both native triple-helical (N) and vulnerable states (V), and collagenolysis occurs only when the enzyme (consisting of only the catalytic domain, C) binds and cleaves the vulnerable state. The associated reaction scheme is shown in Figure 4-1. We use $N{\cdot}C$ and $V{\cdot}C$ 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.

This reaction scheme naturally leads to a set of ODEs that can be solved numerically, yielding the concentration of each species as a function of time. 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 and equilibrium constants associated with each of the various steps in the mechanism. In practice, however, if one is interested in the behavior of the system on long time scales, the model depends only on three equilibrium/binding constants (K_{eq}, K_{bind}^{NC}, and K_{bind}^{VC}) and one catalytic rate constant (k_{cat}) (Nerenberg et al. 2008). In the present case we estimate K_{bind}^{NC} using previously determined collagenase binding constants, which were obtained under conditions when the triple-helical state is expected to be most stable (Ottl et al. 2000). (We note again, however, that the model is not sensitive to the choice of equilibrium constant for binding to the native state, K_{bind}^{NC} (Nerenberg 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 et al. 1982, Netzel-Arnett 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). To determine estimates for the missing parameters, we fit the ODEs arising from the model shown in Figure 4-1 to experimental degradation data.

α1′: NH₂-(GPO)₃GPQGIAGQRGVVGLCGG-OH

α2:   NH₂-(GPO)₃GPQGLLGAOGILGCCGG-OH

α1:   NH₂-(GPO)₃GPQGIAGQRGVVGCGG-OH
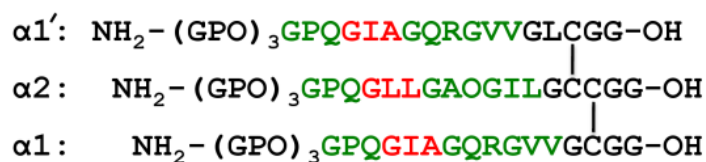
Figure 4-3: Sequence of the type I collagen-like model peptide Trimer A from Ottl *et al*. used in prior degradation experiments with CMMP-8 (Ottl et al. 2000). The type I collagen sequence surrounding the collagenase cleavage site is shown in green, while the triplets containing the scissile bonds are indicated in red. The disulfide linkages of the C-terminus cystine knot region are indicated with the black vertical lines.

4.3.1 Test of the conformational selection model on model peptide degradation data

We begin by focusing our analysis on previous degradation experiments that incubated CMMP-8 at room temperature with a heterotrimeric type I collagen-like model peptide that contains the collagenase cleavage site and its surrounding residues (Figure 4-3) (Ottl et al. 2000). The melting temperature of this peptide is 9 °C, and experiments were performed at 25 °C; 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 states 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 4-1 provides a test of the method.

Using the reaction scheme outlined in equations (4.1.1-4.1.4), we computed the amount of peptide that would be degraded as function as time for a range of K_{eq}, $\left(K_{bind}^{VC}\right)_{CMMP-8}$, and k_{cat} values and compared these degradation profiles to the corresponding experimental data. For each triplet, $\left(K_{eq}, \left(K_{bind}^{VC}\right)_{CMMP-8}, k_{cat}\right)$, we computed the root-mean-square error (RMSE) between the degradation time course obtained with the model and the experimental data. While a relatively wide range of values for K_{eq} and $\left(K_{bind}^{VC}\right)_{CMMP-8}$ were tried ($10^{-6} \leq K_{eq} \leq 10^{6}$, 10^{0} M^{-1} $\leq$ $\left(K_{bind}^{VC}\right)_{CMMP-8} \leq 10^{12}$ M^{-1}), the best fits are obtained when $K_{eq} > 30$ and $\left(K_{bind}^{VC}\right)_{CMMP-8} = 0.7$-$1.1$ x 10^{6} M^{-1}, regardless of the value of k_{cat} that we used (0.11 s$^{-1} \leq k_{cat} \leq 11.1$ s^{-1}) (Figures 4-4A to 4-4C). 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 4-4D).

(A)

RMS error in Trimer A deg. time course, $k_{cat}^{VC} = 1.1 \times 10^{-1}$ s^{-1}

(B)

RMS error in Trimer A deg. time course, $k_{cat}^{VC} = 1.1 \times 10^{0}$ s^{-1}

(C)

RMS error in Trimer A deg. time course, $k_{cat}^{VC} = 1.1 \times 10^{1}$ s^{-1}

(D)

Comparison of model and experimental time courses
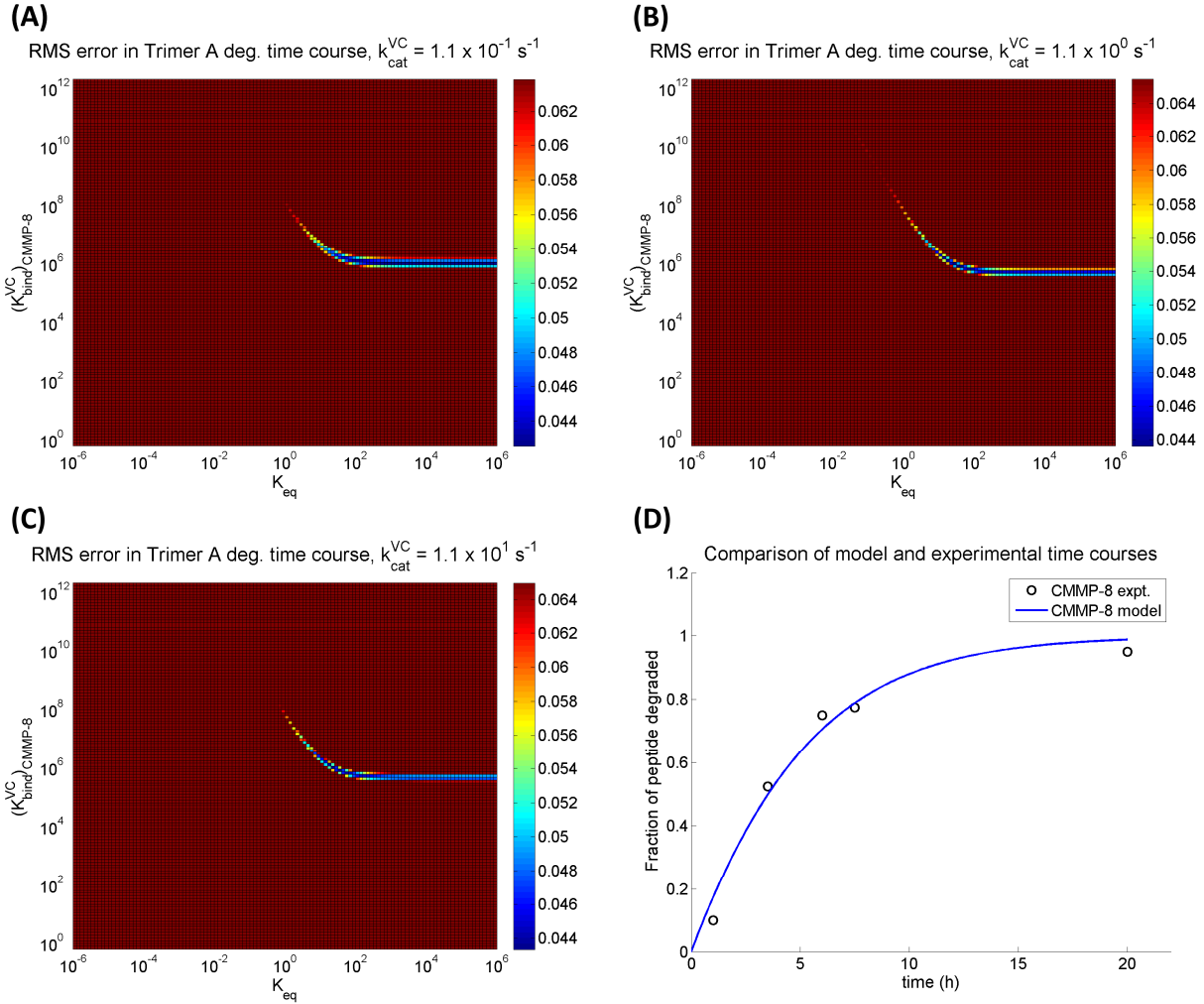
Figure 4-4: Conformational selection and degradation of Trimer A by CMMP-8. (A)-(C) Root mean square error (RMSE) of two state degradation model for Trimer A model peptide degradation time course. For each pair of values $\left(K_{eq}, K_{bind}^{VC}\right)$, the fraction of model peptide degraded by CMMP-8 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 $\left(K_{eq}, K_{bind}^{VC}\right)$ 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 $\left(K_{bind}^{VC}\right)_{CMMP-8}$ range from 0.7-1.1 x 10^6 M^{-1}. (D) Comparison of model and experimental degradation time courses over 20 h for best fit values K_{eq} = 60, $\left(K_{bind}^{VC}\right)_{CMMP-8}$ = 0.7 x 10^6 M^{-1}, and k_{cat} = 1.1 s^{-1}. Experimental data is indicated by the black circles; the model time course is indicated by the blue line.

4.3.2 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 4-2A. We then fit the reaction scheme shown in equations (4.1.1-4.1.4) 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, found that $\left(K_{bind}^{VC}\right)_{CMMP-8} = 0.7\text{-}1.1 \times 10^6 \text{ M}^{-1}$. As this peptide is a model for the collagenase cleavage site, we used this range of $\left(K_{bind}^{VC}\right)_{CMMP-8}$ in our numerical calculations of the model for type I collagen.

We again computed the RMSE between the degradation time courses obtained with the model using many different values of K_{eq} and compared these results to the experimentally determined degradation time course. Varying k_{cat} of CMMP-8 over two orders of magnitude, as we did for the aforementioned model peptide data, did not change the best fit values for K_{eq} and caused the minimum RMSE to again vary by only 2% (Figure 4-5). The best fits between the model and experiment are achieved when $K_{eq} = 1.7\text{-}2.1 \times 10^{-3}$ (Figure 4-5), and the corresponding degradation plot obtained from the model agrees well with experiment (Figure 4-6). The value for K_{eq} at room temperature suggests that the folded triple-helical native state is more favorable at room temperature.

4.3.3 Determining K_{bind}^{VC} for CMMP-1

We applied the same methodology to understand the basis for CMMP-1's markedly lower degradation efficiency as compared to CMMP-8 (Figure 4-2). Given that K_{eq} is an inherent property for type I collagen and the temperature of the experiments, and both CMMP-1 and CMMP-8 type I collagen degradation experiments were performed at room temperature, K_{eq} is the same for both reactions. Thus, we used our degradation model to determine the range of $\left(K_{bind}^{VC}\right)_{CMMP-1}$, estimating k_{cat} for CMMP-1 to be a weighted average of experimentally measured catalytic rates of MMP-1 degrading gelatin (Welgus et al. 1982, Nerenberg et al. 2008). With $K_{eq} = 1.7\text{-}2.1 \times 10^{-3}$, the best fits are found for $\left(K_{bind}^{VC}\right)_{CMMP-1} = 0.9\text{-}1.3 \times 10^4 \text{ M}^{-1}$ (Figure 4-7). These data suggest that CMMP-1's binding constant for the vulnerable state of collagen is nearly

two orders of magnitude lower than that of CMMP-8. Expressed as a difference of binding free energies, $\Delta\Delta G_{bind} = \left(\Delta G_{bind}\right)_{CMMP-1} - \left(\Delta G_{bind}\right)_{CMMP-8}$, the binding of CMMP-1 to type I collagen is approximately 2.5 kcal/mol less favorable than the binding of CMMP-8 to type I collagen.

(A)

RMS error in type I collagen + CMMP-8 degradation time course

$K_{bind}^{VC} = 0.7 \times 10^6\ M^{-1}$

- $k_{cat} = 0.11\ s^{-1}$
- $k_{cat} = 1.1\ s^{-1}$
- $k_{cat} = 11\ s^{-1}$

RMS error

K_{eq} (type I collagen)

(B)

RMS error in type I collagen + CMMP-8 degradation time course

$K_{bind}^{VC} = 1.1 \times 10^6\ M^{-1}$

- $k_{cat} = 0.11\ s^{-1}$
- $k_{cat} = 1.1\ s^{-1}$
- $k_{cat} = 11\ s^{-1}$

RMS error
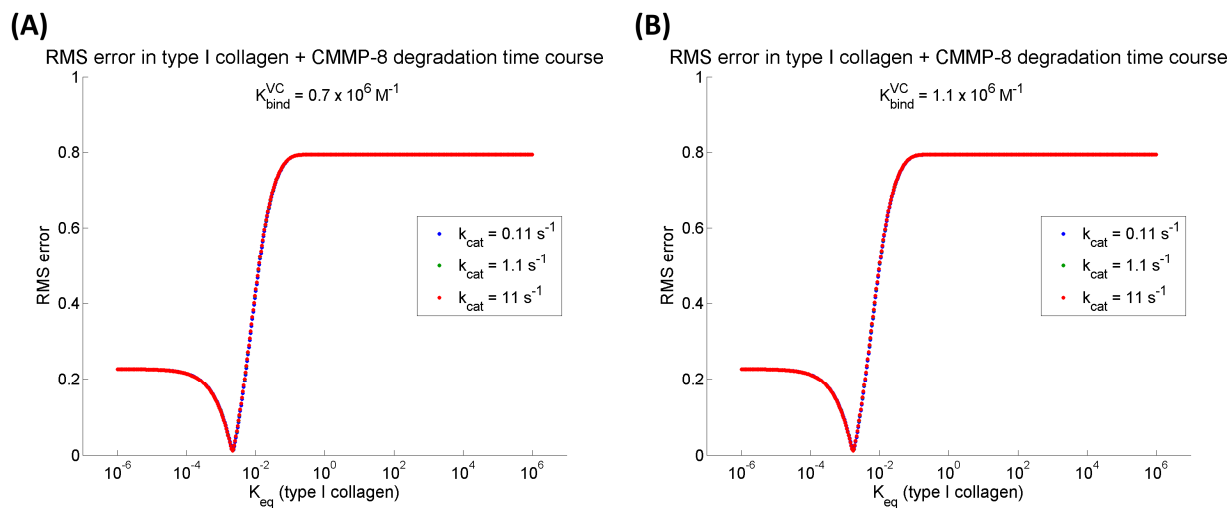
K_{eq} (type I collagen)

Figure 4-5: Root mean square error (RMSE) of two state degradation model for type I collagen degradation time course. For each value of K_{eq}, the fraction of type I collagen degraded by CMMP-8 was computed 8, 16, 24, 32, 40, and 48 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 indicate K_{eq} values that provide good fits to the experimental degradation data. Model fitting was 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 $\left(K_{bind}^{VC}\right)_{CMMP-8}$. As the RMSE curves for each value of k_{cat} reach a minimum at the same value of K_{eq}, this demonstrates that varying k_{cat} does not affect the best fit values for K_{eq}.
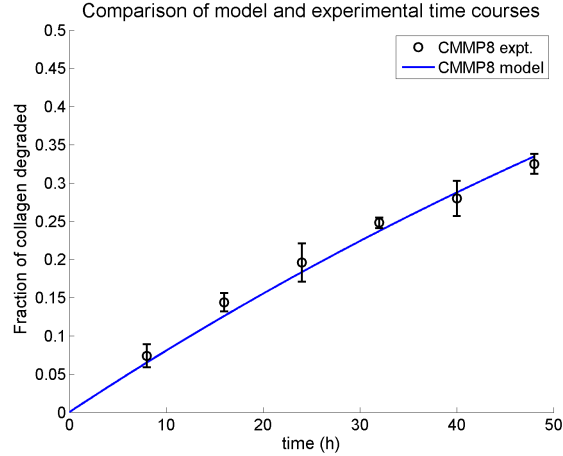
Figure 4-6: A comparison of model and experimental degradation time courses over 48 h for best fit values $K_{eq} = 2.1 \times 10^{-3}$, $\left(K_{bind}^{VC}\right)_{CMMP8} = 0.7 \times 10^6$ M^{-1}, and $k_{cat} = 1.1$ s^{-1}. Experimental data is indicated by the black circles with error bars; the model time course is indicated by the blue line.

(A)

(B)
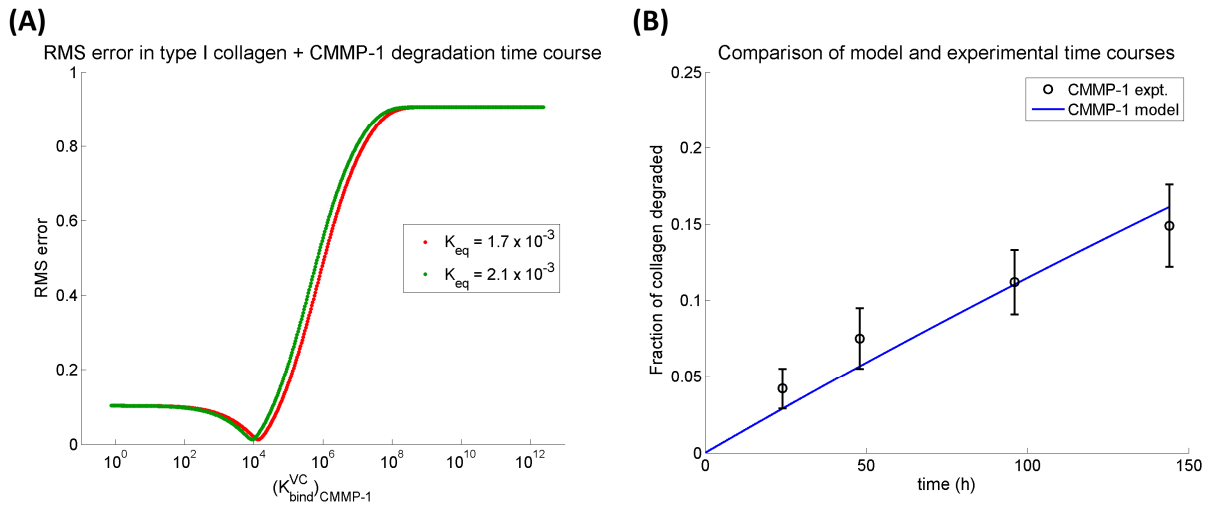
Figure 4-7: Conformational selection and degradation of type I collagen by CMMP-1. (A) RMSE of the conformational selection model over the time course for degradation experiments utilizing type I collagen and CMMP-1. For each value of $\left(K_{bind}^{VC}\right)_{CMMP-1}$, the fraction of degraded type I collagen was computed at 24, 48, 96, and 144 h. The root mean square difference was then calculated using these fractions and the experimentally measured fraction of degraded collagen. RMSE curves are computed using the lower and upper bounds of K_{eq}. (B) A comparison of model and experimental degradation time courses over 48 h for best fit values $K_{eq} = 2.1 \times 10^{-3}$, $\left(K_{bind}^{VC}\right)_{CMMP-1} = 0.9 \times 10^4$ M^{-1}, and $k_{cat} = 0.11$ s^{-1}. Experimental data is indicated by the black circles with error bars; the model time course is indicated by the blue line.

4.4 Discussion

Our results unambiguously demonstrate that collagenolysis can occur with MMP deletion mutants that contain only the catalytic domain and therefore indicate that the hemopexin-like domain is not required for peptide bond hydrolysis and enzyme specificity at room temperature. Moreover, while both CMMP-1 and CMMP-8 cleave completely unfolded type I collagen chains at multiple sites, both enzymes cleave type I collagen *in vitro* at only one site that corresponds to the unique cleavage site recognized by full length collagenases. We therefore interpret these findings 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 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 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 were 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 were performed, it correctly predicts that the degradation experiments were performed at a relatively high temperature for the heterotrimeric peptide and a relatively low temperature for collagen.

Our results offer an explanation for the different type I collagen efficiencies of CMMP-1 and CMMP-8. 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 degradation experiments. However, we find that the binding of CMMP-1 to type I collagen is 2.5 kcal/mol less favorable than the binding of CMMP-8. 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. These observations suggest that subtle changes in binding between similar homologous enzymes can lead to significant differences in the overall reaction kinetics.

Although several potential cleavage sites exist in collagen, only one unique site is cleaved by MMPs (Kuivaniemi et al. 1988, Tromp et al. 1988, Fields 1991). Recent observations suggest that cleavage site specificity is mediated by interactions of non-catalytic domains with distinct sites on collagen and that these interactions may play an important role in enzyme specificity in the fibrillar state (Perumal et al. 2008, Erat et al. 2009). Since we find that cleavage at the unique collagenase cleavage site is achieved in the absence of non-catalytic domains, the non-catalytic domains are not required for cleavage site specificity *in vitro* at room temperature. The relatively low local stability of the imino-poor cleavage site ensures that it unfolds at low temperatures and is preferentially recognized by MMP catalytic domains (Fields et al. 1987, Stultz 2002, Nerenberg and Stultz 2008, Salsas-Escat and Stultz 2009). As the temperature increases, more regions of the molecule unfold and additional sites may be recognized by MMP deletion mutants (Gioia et al. 2002, Taddese et al. 2009). At body temperature it is likely that there are a number of locally unfolded regions in the collagen chain, and in this scenario localization of the enzyme via non-catalytic domains will help to ensure that the correct cleavage site is recognized. In addition, the binding of the hemopexin-like domain at sites near the unique cleavage site in collagen effectively increases the local concentration of the enzyme in the vicinity of the cleavage site, thereby making enzyme-mediated degradation more efficient.

Overall, our data are consistent with the notion that, at room temperature, the conformational ensemble of type I collagen includes locally unfolded conformations that have relatively exposed scissile bonds. These locally unfolded conformations encode cleavage site specificity and are the basis of a conformational selection mechanism in which collagenases recognize and cleave these pre-existing locally unfolded states. This degradation mechanism presents a framework both for understanding the basic interaction of collagen with collagenases and for therapeutic strategies to modulate excessive collagenolysis associated with many diseases.

Chapter 5

Conformational Selection in Type III Collagen Degradation

This chapter is adapted from: Nerenberg PS, Salsas-Escat R, and Stultz CM. Collagenase cleavage site specificity in type III collagen. In preparation.

In Chapter 4, we validated the vulnerable state structural model of type I collagen by performing degradation experiments using deletion mutants of collagenases that contained only the catalytic domain. The fact that degradation occurred at the same location as with full length enzyme confirmed the existence of partially unfolded states in the vicinity of the collagenase cleavage site, but also suggested that the specificity of the cleavage site was encoded in its conformational ensemble, rather than through interactions with the hemopexin-like domains of collagenases. We wish to further investigate both the vulnerable state model and the issue of collagenase cleavage site specificity in the context of another fibrillar collagen, type III collagen. If type III collagen were shown to sample partially unfolded states that encode for the specificity of the collganase cleavage site, this would further support the hypothesis that collagenolysis occurs when collagenases recognize, bind, and cleave partially unfolded conformers of collagen.

It has long been known that native triple-helical collagen is cleaved at only one site by collagenases, whereas in denatured collagen (gelatin), collagenases are able to cleave at several sites (Welgus et al. 1982, Fields 1991). Human type III collagen contains four pseudo-cleavage sites in addition to the true, unique collagenase cleavage site (Fields 1991). These same sequence data indicate that there are significant differences in the compositions of the amino acid sequences surrounding the collagenase cleavage site and the pseudo-cleavage sites (Fields 1991), but on their own offer relatively little insight into any differences in the physical characteristics of these sites. Determining the physical basis for why collagenases cleave only a single site in native collagen, but multiple sites in denatured collagen, is an important step towards developing a more complete understanding of the molecular mechanism of collagen degradation.

Matrix metalloproteinases (MMPs) are Zn^{2+}-dependent enzymes that hydrolyze the peptide bond linking glycine and isoleucine or glycine and leucine residues within G-[I/L]-[A/L] triplets in collagen (Fields 1991, Lauer-Fields et al. 2002). The collagenases, a subset of the MMP family (MMP-1, -8, and -13), are characterized structurally by having two domains – a catalytic domain connected via a linker region to a (non-catalytic) hemopexin-like domain (Murphy and Knauper 1997, Borkakoti 1998, Lauer-Fields et al. 2002). Other MMPs have additional non-catalytic domains. The gelatinases (MMP-2 and -9), for example, contain a fibronectin type II domain. Although, as mentioned previously, there are several G-[I/L]-[A/L] triplets within the collagen molecule, the collagenases cleave native collagen at one unique site.

Many factors may contribute to the specificity of the collagenase cleavage site. One such factor may be the interactions between native collagen and collagenases, and specifically the binding contacts thought to be made by the non-catalytic domains with the triple helix (Murphy et al. 1992, Borkakoti 1998, Lauer-Fields et al. 2002, Nerenberg et al. 2008, Perumal et al. 2008, Erat et al. 2009, Lauer-Fields et al. 2009). Recent modeling studies, for instance, have suggested that non-catalytic domains of collagenases and gelatinases recognize and bind to an RGER motif in the α1 chains of type I collagen (Perumal et al. 2008, Erat et al. 2009). These data, along with theoretical studies and degradation experiments performed with deletion mutants of collagenases (i.e., mutant enzymes that lack non-catalytic domains) (Murphy et al. 1992, Chung et al. 2004, Nerenberg et al. 2008), suggest that there are important interactions between native collagen and the non-catalytic domains of collagenases, some of which may explain the specificity of the collagenase cleavage site.

Another factor that may contribute to the specificity of the cleavage site is the conformational ensemble of native collagen near the collagenase cleavage site. In particular, it has been suggested that the collagenase cleavage site in native collagen is less thermally stable than the pseudo-cleavage sites (Miller et al. 1976, Highberger et al. 1979, Ryhänen et al. 1983, Birkedal-Hansen et al. 1985, Fields 1991). Cleavage of native type III collagen in the immediate vicinity of the collagenase cleavage site at temperatures below its melting temperature has been observed in prior degradation experiments utilizing several different proteases (e.g., trypsin or elastase) (Miller et al. 1976, Wang et al. 1978, Mainardi et al. 1980, Birkedal-Hansen et al. 1985). These data have been interpreted to indicate that the region surrounding the collagenase cleavage site lacks normal triple helicity and therefore that non-collagenolytic proteases can freely access

peptide bonds in that region. A recent molecular dynamics study further probed this hypothesis by characterizing the conformational ensembles of the regions surrounding both the collagenase cleavage site and the pseudo-cleavage sites in type III collagen (Salsas-Escat and Stultz 2009). The results of this study suggested that the scissile bonds of the collagenase cleavage site are more solvent-exposed (and therefore vulnerable to enzymatic cleavage) than those of the pseudo-cleavage sites (Salsas-Escat and Stultz 2009). Moreover, the states sampled by the collagenase cleavage site were found to be more complementary to the collagenase active site than those sampled by the pseudo-cleavage sites (Salsas-Escat and Stultz 2009). Together with the previous observations that degradation of native type III collagen occurs with non-collagenolytic proteases near the collagenase cleavage site, these data suggest that the specificity of the collagenase cleavage site may be encoded in a local conformational ensemble that uniquely differs from the conformational ensembles of the pseudo-cleavage sites.

In this chapter we explore the physical basis of collagenase cleavage site specificity using a combination of biochemical experiments and modeling approaches. To probe whether binding contacts on the hemopexin-like domains of collagenases are responsible for the specificity of the collagenase cleavage site, we incubated native type III collagen with deletion mutants of collagenases that lack the hemopexin-like domains. To examine the thermal stability and conformational ensembles of these sequences, we designed a set of homotrimeric model peptides (C3S1-C3S5) containing the sequences surrounding the collagenase cleavage site and pseudo-cleavage sites of type III collagen. We then performed melting experiments to assess the global thermal stability of these sequences and heteronuclear 2D (HSQC) NMR experiments, fit with simple conformational models, to characterize the conformational distributions of both the collagenase cleavage site and the pseudo-cleavage site peptides.

5.1 Experimental methods

5.1.1 Type III collagen degradation experiments

Recombinant human type III collagen expressed in the yeast *Pichia pastoris* (AB73160, Abcam Inc.) is received at 3.3 mg/ml in 0.01 M HCl. Full length MMP-1 (FMMP-1, product no. SE-361) and deletion mutants that contain only the catalytic domains (CMMP-1 and CMMP-8, product no. SE-180 and SE-255) were purchased from Biomol International LP. Nu-Page 4-12% gradient gels (NP0335BOX) and Colloidal Blue Staining kit (SKLC6025) were obtained from Invitrogen.

Type III collagen was diluted in the final reaction vessel to a concentration of 200 μg/ml, while enzyme concentrations ranged from 17-33 μg/ml, as described in Figure 5-2. Both full length and the catalytic domains of MMP-1 and MMP-8 were activated using APMA (Sigma-Aldrich Co.) as previously described (Clark 2001). All degradation reactions were performed using a compact thermomixer (Eppendorf AG) at 25 °C, which can control the temperature from 4 °C to 99 °C, with an accuracy of ±1 °C. All reactions were carried out in TNC buffer, containing 100 mM Tris HCl (VWR International), 10 mM $CaCl_2$ (Sigma-Aldrich Co.), pH 7.6. Reactions were stopped by the addition of SDS-Laemmli buffer (BioRad Laboratories) with β-Mercapto Ethanol (Sigma-Aldrich Co.) and boiled for 10 minutes. The degradation products were run in 4-12% gradient gels and stained with Coomassie colloidal blue.

Unstained CMMP-1 and CMMP-8 gels were transferred to PVDF membranes, which were stained with Coomassie blue. The ¼α1 bands were cut and sent for sequencing to the Tufts University Core Facility, where they were sequenced with an ABI 494 Protein Sequencer.

5.1.2 Collagen-like model peptide synthesis

Self-assembling homotrimeric collagen-like model peptides were designed to contain six triplets of type III collagen sequence surrounding and including either the collagenase cleavage site or one of four pseudo-cleavage sites (Figure 1). The type III collagen sequence triplets were flanked by GPO triplets in both the N- and C-terminal directions (Figure 1B) to facilitate peptide folding and provide thermal stability (Sakakibara et al. 1973, Fields and Prockop 1996, Persikov et al. 2000, Engel and Bächinger 2005). The model peptides (C3S1-C3S5) were synthesized by the Tufts University Core Facility using an ABI 431 peptide synthesizer using Fmoc chemistry based on a 0.1 mmol scale. Each sequence was synthesized as two variants to optimize NMR assignments – one variant with ^{15}N-labels located at Gly7 and Ala18 and the other with [15]N-

labels located at Ala18 and Gly34. All peptides were purified using reversed-phase high pressure liquid chromatography on a C18 column, and their identities were confirmed using an ABI Voyager-DE Pro MALDI-TOF mass spectrometer.

(A)

```
S1 (936-962):  -GAQ-GPO-GAO-GPL-GIA-GIT-GAR-GLA-GPO-
S2 (945-971):  -GPL-GIA-GIT-GAR-GLA-GPO-GMO-GPR-GSO-
S3 (675-701):  -GDA-GAO-GER-GPO-GLA-GAO-GLR-GGA-GPO-
S4 (777-803):  -GDK-GEG-GAO-GLO-GIA-GPR-GSO-GER-GET-
S5 (990-1016): -GER-GPO-GPQ-GLO-GLA-GTA-GEO-GRD-GNO-
```

(B)

```
C3S1: (GPO)₃-GAO-GPL-GIA-GIT-GAR-GLA-(GPO)₆
C3S2: (GPO)₃-GIT-GAR-GLA-GPO-GMO-GPR-(GPO)₆
C3S3: (GPO)₃-GER-GPO-GLA-GAO-GLR-GGA-(GPO)₆
C3S4: (GPO)₃-GAO-GLO-GIA-GPR-GSO-GER-(GPO)₆
C3S5: (GPO)₃-GPQ-GLO-GLA-GTA-GEO-GRD-(GPO)₆
```

Figure 5-1: (A) The amino acid sequences of the collagenase cleavage site (S1) and four pseudo-cleavage sites (S2-S5) in human type III collagen, based on UniProtKB/Swiss-Prot accession number P02461. The scissile bond triplets are indicated in red. The S2 pseudo-cleavage site, which is present in the S1 sequence, and the S1 collagenase cleavage site, which is present in the S2 sequence, are shown in violet for sequences S1 and S2, respectively. (B) The amino acid sequences of homotrimeric collagen-like model peptides (C3S1-C3S5), which contain a portion of the amino acid sequences in (A). GPO repeats are indicated in blue, real type III collagen sequences are indicated in green, and the scissile bond triplets are indicated in red.

5.1.3 Determination of peptide melting temperatures

Peptide samples were dissolved in 50 mM sodium phosphate buffer, pH 6.4. They were then allowed to equilibrate for 72 hours at 4 °C to allow sufficient time for folding to reach equilibrium. Samples were then diluted 50x to a concentration of 0.08 mM (80 µM), appropriate for CD spectroscopy, and then equilibrated at 4 °C for an additional 24 hours.

CD spectra were recorded for all peptides at 4 °C on a Jasco J810 spectropolarimeter. Cuvettes of 1 mm path length were used, and the temperature of the cells was controlled using a

Jasco PTC-423S Peltier temperature controller. Wavelength scans from 190-240 nm were performed at a scan rate of 10 nm/min in 0.1 nm steps with an 8 s averaging time. Scans were repeated five times and averaged. For the melting temperature experiments, we monitored the ellipticity at 224 nm. The temperature was increased from 4 °C to 88 °C at a rate of 0.1 °C/min (6 °C/h). Although this heating rate did not allow for equilibrium to be reached, the goal of these experiments was to obtain relative, rather than absolute melting temperatures. Assuming a model in which there is a two state equilibrium between folded and monomer states, the folded fraction, $F(T)$, was calculated according to:

$$F(T) = \frac{\theta(T) - \theta_M(T)}{\theta_{TH}(T) - \theta_M(T)} \tag{5.1}$$

where $\theta(T)$ is the measured ellipticity and $\theta_{TH}(T)$ and $\theta_M(T)$ are the ellipticities of the triple-helical and monomer states, respectively, at a temperature T (Persikov et al. 2000, Persikov et al. 2004, Hyde et al. 2006). Both $\theta_{TH}(T)$ and $\theta_M(T)$ were approximated by using linear fits to the measured ellipticity in the appropriate temperature ranges. The melting temperature, T_m, for each peptide was calculated by finding the temperature at which the folded fraction of the peptide was 0.5 (i.e., $F(T_m) = 0.5$).

5.1.4 NMR spectroscopy of model peptides

Peptide samples were dissolved in 50 mM sodium phosphate buffer, pH 6.4, with 10% D_2O at a concentration of 4.0 mM. They were then allowed to equilibrate for 72 hours at 4 °C to allow sufficient time for folding to reach equilibrium.

NMR experiments were performed on a custom-built 591 MHz spectrometer with a Nalorac 5 mm indirect detection triple resonance probe at the Francis Bitter Magnet Laboratory. ^{1}H-^{15}N HSQC experiments were carried out at 10 °C and data were acquired with a spectral width of 8000 Hz in the ^{1}H dimension and 3000 Hz in the ^{15}N dimension (2048 x 512 complex points) in 16 scans. All data were phased and processed using NMRPipe (Delaglio et al. 1995). Peak assignments were made as described in section 5.2. Peak volumes were determined by fitting the peaks with Lorentzian line shapes, subtracting the baseline, and integrating the resulting volumes using Sparky (Goddard and Kneller).

5.2 Degradation of type III collagen by collagenase deletion mutants

Human type III collagen was incubated with CMMP-1 at 25 °C – a temperature well below the melting temperature of type III collagen (Birkedal-Hansen et al. 1985). After 3.5 h, type III collagen degradation was observed in solutions containing CMMP-1, and nearly complete degradation was seen after 48 h (Figure 5-2A). Degradation bands exhibit the familiar ¾ and ¼ fragments that are associated with cleavage at the unique collagenase cleavage site and can also be observed when type III collagen is degraded by FMMP-1 (Figure 5-2A). Amino acid sequencing of the ¼ bands confirms that the CMMP-1 deletion mutant cleaves at the same site as the corresponding wild-type full length enzymes.

To demonstrate that our results are not specific to MMP-1, we exposed type III collagen to a deletion mutant containing only the catalytic domain of MMP-8 (Figure 5-2B). The resulting cleavage pattern is the same as that observed with FMMP-1 (Figure 5-2A), and amino acid sequencing of the ¼ bands again demonstrates that CMMP-8 recognizes and cleaves the same site as FMMP-1. Although cleavage of type III collagen with the catalytic domain alone occurs at the same site that is recognized by full length enzyme, cleavage with FMMP-1 is more efficient than with the catalytic domain alone, as almost all collagen is degraded with FMMP-1 after 24 h (Figure 5-2A). Lastly we note that our data is not explained by contamination of the original collagen sample with full length MMPs, as solutions of type III collagen incubated without any added MMPs do not exhibit any degradation after incubation for 48 h (data not shown).

5.3 Thermal stability of collagen-like model peptides

We obtained collagen-like model peptides that contain regions corresponding to the true collagenase cleavage site and pseudo-cleavage sites (Figure 5-1). All five model peptides have CD spectra at 4 °C that contain a minimum near 196-198 nm and a peak at 224 nm; i.e., spectra that are characteristic of a folded triple-helical structure (Figure 5-3A). Interestingly, model peptide C3S1, which contains the collagenase cleavage site sequence, appears to have a lower overall triple helicity than model peptides C3S2-C3S5, which contain the pseudo-cleavage site sequences and have similar helicities (Figure 5-3A).
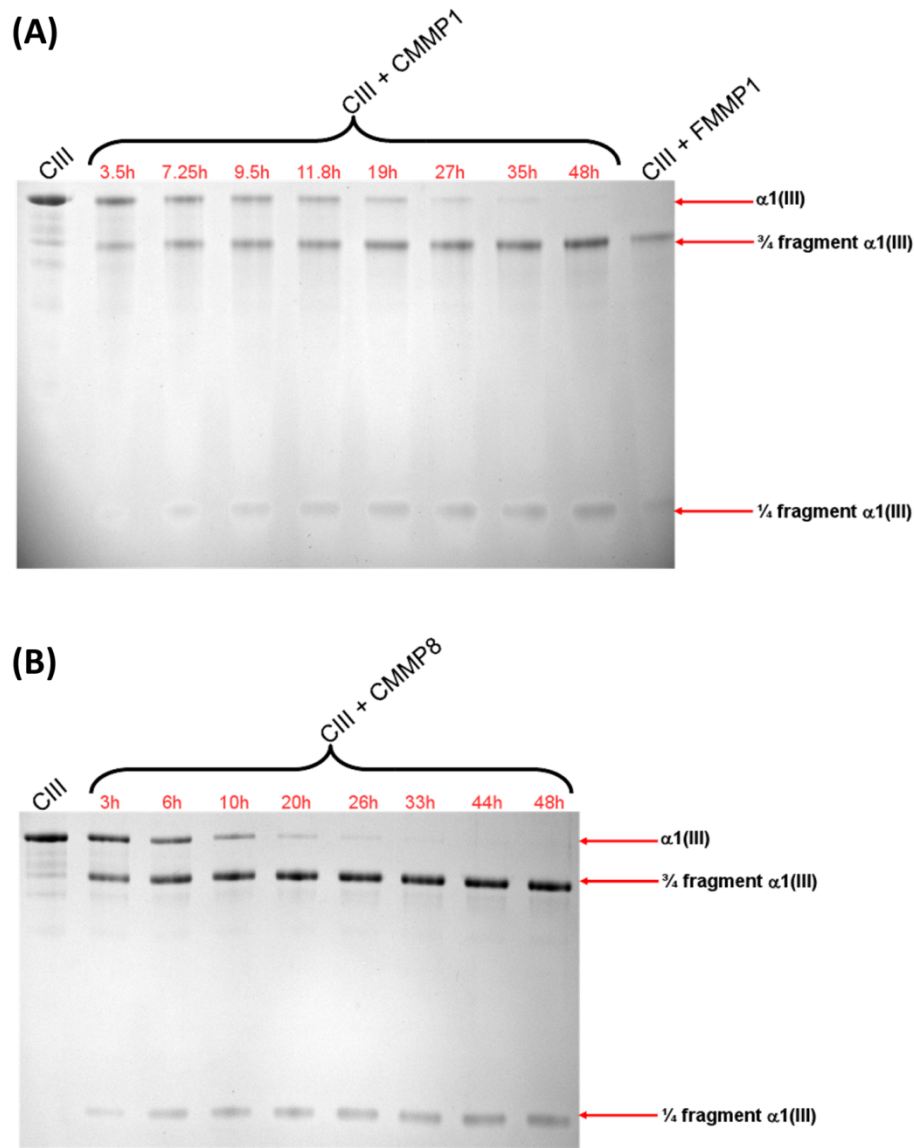
Figure 5-2: Degradation profiles of type III collagen (CIII) at 25 °C with MMP-1 and MMP-8 using both the full length enzymes and only the catalytic domains. (A) *Lane 1:* CIII (200 µg/ml). *Lanes 2 to 9:* CIII (200 µg/ml) incubated with the catalytic domain of MMP-1 (CMMP-1) (33 µg/ml) for 3.5, 7.25, 9.5, 11.8, 19, 27, 35, and 48 h, respectively. *Lane 10:* CIII (200 µg/ml) incubated with and full length MMP-1 (FMMP-1) (17 µg/ml) for 24 h. (B) *Lane 1:* CIII (200 µg/ml). *Lanes 2 to 9:* CIII (200 µg/ml) incubated with the catalytic domain of MMP-8 (CMMP-8) (33 µg/ml) for 3, 6, 10, 20, 26, 33, 44, and 48 h, respectively.
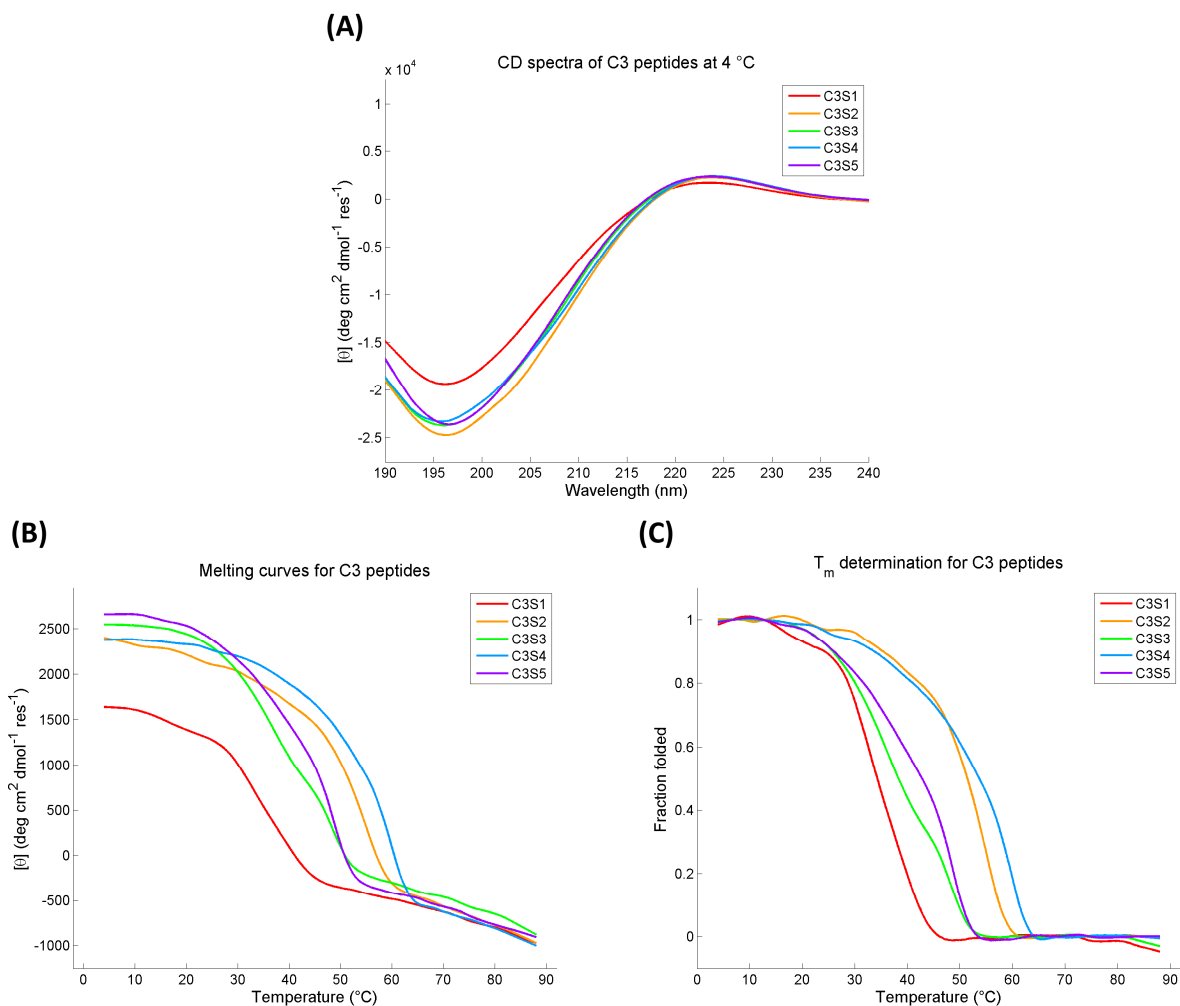
Figure 5-3: Circular dichroism for homotrimeric model peptides C3S1-C3S5. (A) The spectra of all five peptides at 4 °C display the canonical features of collagen triple helices – a peak at 224 nm and minimum at 196-198 nm. (B) Melting temperature experiments reveal that the characteristic peak at 224 nm is eliminated in a cooperative transition, indicative of the presence of a triple helix. (C) Determination of melting temperature, as described in section 5.1.3. The melting temperatures for all five peptides, in order from C3S1 to C3S5 are: 34.0 °C, 51.4 °C, 37.9 °C, 53.8 °C, and 42.6 °C. Notably, C3S1 is the peptide that contains the collagenase cleavage site sequence.

Although each peptide adopts a triple-helical structure at 4 °C, they each have very different thermal stabilities (Figures 5-3B and 5-3C). In particular, model peptide C3S1 has the lowest T_m (34.2 °C) (Figure 5-3C). The type III collagen sequence contained in this model peptide also has the lowest imino acid content – one proline and one hydroxyproline – of any of the model peptide sequences and only a single X2-position arginine (Figure 5-1B). By comparison, model peptide C3S4 has the highest melting temperature (53.6 °C) of all the model peptides (Figure 5-3C) and contains the most imino acids and two X2-position arginines of all five peptides (Figure 5-1B). Overall, the relative ordering of melting temperatures for the model peptides correlates well with the imino acid and X2-position arginine content of those sequences, as has been suggested in prior studies of triple-helical propensities (Persikov et al. 2000).

5.4 HSQC spectra of collagen-like peptides

While melting temperature experiments measure the global thermal stability of the model peptides, we also wished to quantify the local triple helicities of these peptides both near and far from the scissile bond. NMR spectroscopy offers the ability to monitor the local chemical environment around specific residues, thus providing a measure of how often that site samples either folded triple-helical or unfolded conformations. Given that model peptides C3S2-C3S5 have similar CD spectra at 4 °C, we elected to perform NMR experiments using the most thermally stable of these peptides, C3S4, along with the real collagenase cleavage site peptide, C3S1. To this end, model peptides C3S1 and C3S4 were synthesized in two variants – one with ^{15}N-labeled residues at Gly7 and Ala18 (the alanine residue immediately C-terminal of the scissile bond) and one with ^{15}N-labeled residues at Ala18 and Gly34 (Figures 5-4A and 5-5A). Heteronuclear 2D (HSQC) experiments were performed at 10 °C, well below the melting temperature of either C3S1 or C3S4. Additionally, these experiments were performed at pH 6.4 to ensure that the side chains of acidic residues in the model peptides would remain unprotonated. Prior studies have shown that these residues that may contribute to thermal stability via electrostatic interactions (Venugopal et al. 1994, Persikov et al. 2000, Persikov et al. 2002).

Well resolved HSQC spectra containing both folded (triple-helical) and unfolded peaks were obtained for peptides C3S1 and C3S4 (Figures 5-4B/C and 5-5B/C). Due to each peptide variant having only one ^{15}N-labeled glycine and one ^{15}N-labeled alanine, peaks were well separated into

(A)

```
Chain A: -G-P-O-G-P-O-G-A-O- ... -G-P-L-G-I-A-G-I-T- ... -G-P-O-G-P-O-G-P-O-

Chain B:  -G-P-O-G-P-O-G-A-O- ... -G-P-L-G-I-A-G-I-T- ... -G-P-O-G-P-O-G-P-O-

Chain C:   -G-P-O-G-P-O-G-A-O- ... -G-P-L-G-I-A-G-I-T- ... -G-P-O-G-P-O-G-P-O-

Chain A: -G-P-O-G-P-O-G-A-O- ... -G-P-L-G-I-A-G-I-T- ... -G-P-O-G-P-O-G-P-O-
```
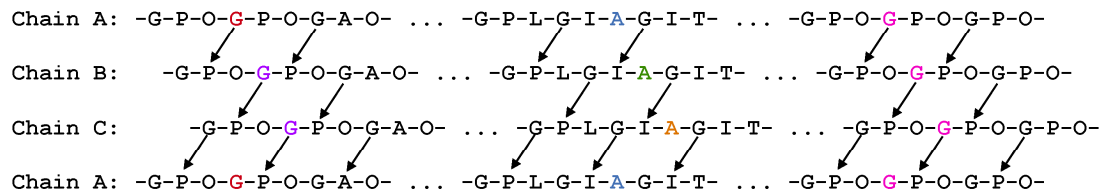
(B)
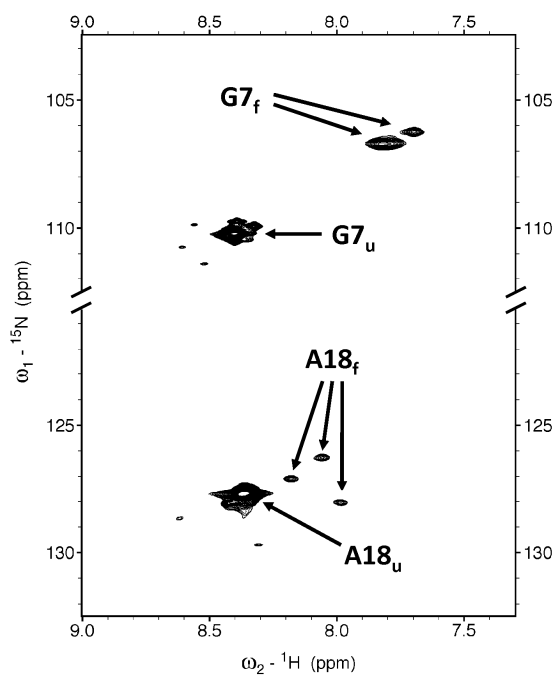
(C)
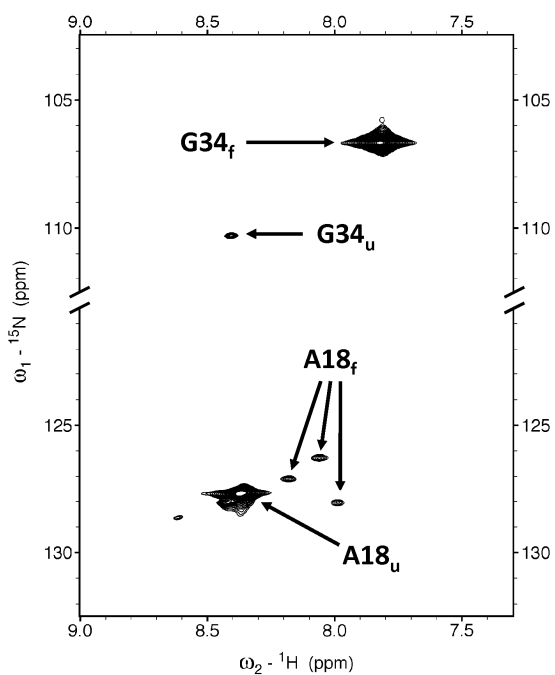
Figure 5-4: (A) Interchain hydrogen bonding pattern of model peptide C3S1, which contains the collagenase cleavage site, in the folded triple-helical state. This pattern shows the various hydrogen bonding partners to the ^{15}N-labeled Gly7 (red and purple), Ala18 (blue, green, and orange), and Gly34 (pink) sites. Different colors indicate different chemical environments. ^{1}H-^{15}N HSQC spectra obtained at 10 °C of (B) the C3S1 variant with ^{15}N labels at Gly7 and Ala18 and (C) the C3S1 variant with ^{15}N labels at Ala18 and Gly34. Peaks corresponding to folded triple-helical conformers are indicated with a subscript f, while peaks corresponding to unfolded conformers are indicated with a subscript u.
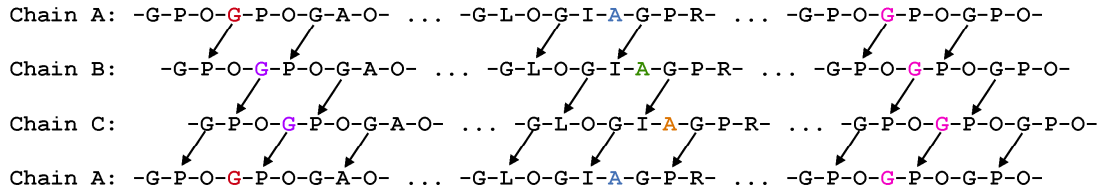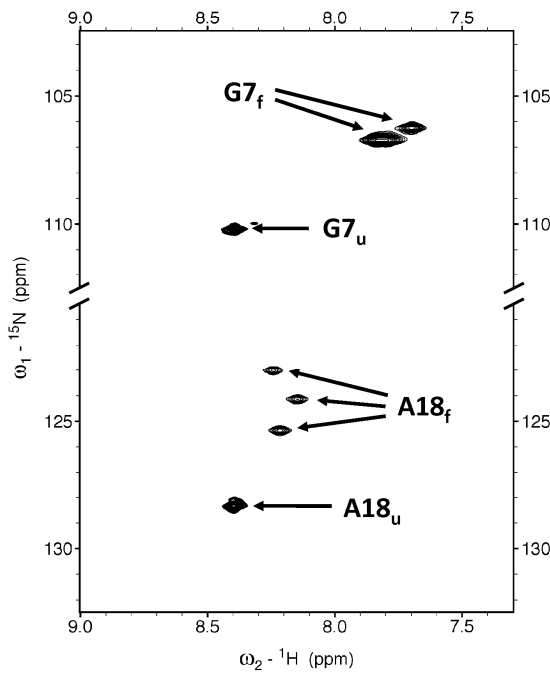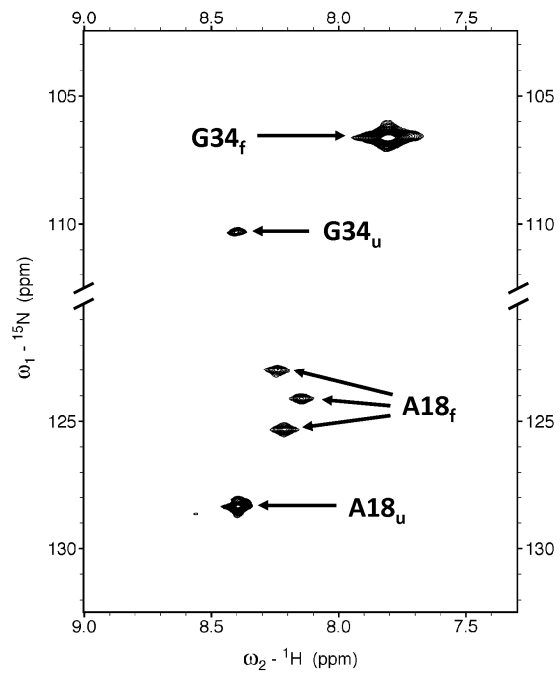
Figure 5-5: (A) Interchain hydrogen bonding pattern of model peptide C3S4, which contains the collagenase cleavage site, in the folded triple-helical state. This pattern shows the various hydrogen bonding partners to the ^{15}N-labeled Gly7 (red and purple), Ala18 (blue, green, and orange), and Gly34 (pink) sites. Different colors indicate different chemical environments. ^{1}H-^{15}N HSQC spectra obtained at 10 °C of (B) the C3S4 variant with ^{15}N labels at Gly7 and Ala18 and (C) the C3S4 variant with ^{15}N labels at Ala18 and Gly34. Peaks corresponding to folded triple-helical conformers are indicated with a subscript f, while peaks corresponding to unfolded conformers are indicated with a subscript u.

the characteristic chemical shift regions for those residues (Ulrich et al. 2008). Specific assignment of the peaks to folded and unfolded conformers was aided by previously published model peptide studies (Li et al. 1993, Buevich et al. 2000, Li et al. 2005, Li et al. 2007). In addition, we found that the number of folded peaks for each residue was consistent with the different triplet sequences that are interchain hydrogen bond partners with the triplets containing the ^{15}N-labeled residues (Figures 5-4A and 5-5A). For instance, the Gly7 label in both C3S1 and C3S4 yields two peaks – a doublet peak and a singlet peak that are close together in chemical shift space (Figures 5-4B and 5-5B). This observation makes sense in light of the fact that two of the labeled sites have one GPO triplet and one GAO triplet as hydrogen bonding partners, while one of the labeled sites has two GPO triplets as its partners (Figures 5-4A and 5-5A). Similar analyses can explain the appearance of three folded peaks for Ala18 and a single large folded peak for Gly34 (Figures 5-4B/C and 5-5B/C).

For peptides C3S1 and C3S4, the folded and unfolded peaks were fit with Lorentzians and their volumes integrated using Sparky (Goddard and Kneller). By comparing the folded and unfolded peak volumes, the fraction of folded conformers (or fractional triple helicity) at each labeled site can be calculated according to:

$$S_f = \frac{vol_f}{vol_f + vol_u} \tag{5.2}$$

where S is the labeled residue, and vol_f and vol_u are the volumes of the folded and unfolded peaks for that residue, respectively. For model peptide C3S1, which contains the collagenase cleavage site, the fractional triple helicities were as follows: $G7_f = 0.34$, $A18_f = 0.12$, and $G34_f = 0.95$. These data suggest that the Gly34 site, embedded in the C-terminal (GPO)$_6$ region, is almost entirely folded. The Gly7 site, located in the N-terminal (GPO)$_3$ region, has a substantially lower triple helicity. Finally, the Ala18 site, located one residue C-terminal to the scissile bond, is the least folded of all three sites. For model peptide C3S4, which contains the pseudo-cleavage site, the fractional triple helicities were: $G7_f = 0.75$, $A18_f = 0.67$, and $G34_f = 0.92$. These triple helicities suggest that the C-terminus regions of both peptides are similarly folded, but that the scissile bond and N-terminus regions of the collagenase cleavage site peptide (C3S1) are substantially less folded than the analogous regions in the more thermally stable pseudo-cleavage site peptide (C3S4).

5.5 Conformational models of collagen-like peptides

We wish to characterize the types of structures that are present in solution for both C3S1 and C3S4 under steady state conditions. Previous model peptide studies using residue-specific NMR diffusion measurements suggest that partially disordered folding intermediates may also be present in steady state conditions (Li et al. 2005). While these folding intermediates are kinetic intermediates and not equilibrium states, the lifetimes of these intermediates may be long relative to the NMR timescale, and therefore they may contribute to the measured folded and unfolded peaks (Buevich et al. 2000, Li et al. 2005). The types of structural states that may contribute to the HSQC peaks are shown in Figure 5-6. The fully unfolded state (X_1) and the fully folded state (X_4) represent the beginning and endpoints of the folding reaction. In addition, previous NMR studies on collagen-like model peptides demonstrate that folding typically proceeds from an imino-rich C-terminus to the N-terminus, hence a C-terminal intermediate is included (state X_2) (Liu et al. 1996, Buevich et al. 2000, Li et al. 2005). Recent experiments with model peptides have shown that folding can also proceed in the N-to-C direction, especially when the N-terminal region is relatively imino-rich, therefore we include an N-terminal folding intermediate (state X_3) (Frank et al. 2003).

Each of these states makes different contributions to the folded and unfolded peaks in the HSQC spectra. State X_1 contributes unfolded peaks at every labeled site, while state X_2 contributes folded peaks for the Gly34 label and unfolded peaks for the Gly7 and Ala18 labels. Similarly, state X_3 conformers would contribute folded peaks for the Gly7 label and unfolded peaks for the Ala18 and Gly34 labels, and state X_4 contributes to folded peaks for all labeled residues.

5.5.1 Four state model

We first explore a model containing all four of the aforementioned states X_1-X_4. In addition to the equations that relate the probabilities of the states (γ_1, γ_2, γ_3, γ_4) to the fractional triple helicities (G7$_f$, A18$_f$, G34$_f$), we also require that the probabilities of the states sum to one:

$$G7_u = 1 - G7_f = \gamma_1 + \gamma_2 \tag{5.3}$$

$$G7_f = \gamma_3 + \gamma_4 \tag{5.4}$$
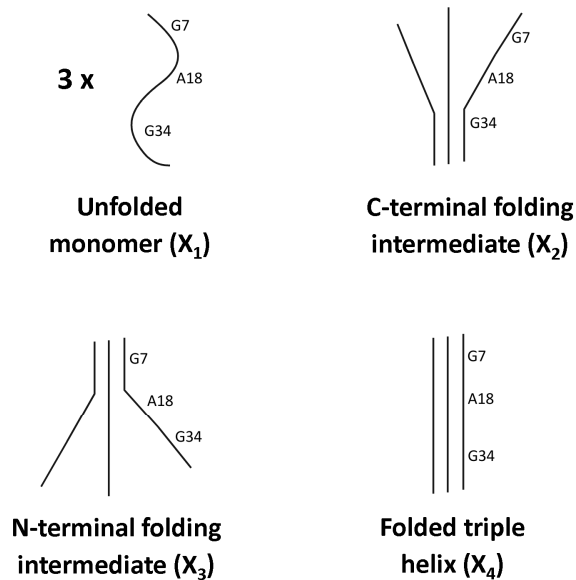
Figure 5-6: Four possible conformational states of collagen-like model peptides that may contribute to the folded and unfolded peaks observed in HSQC spectra. State X_1 is an unfolded monomer that yields unfolded peaks for all three labeled sites (Gly7, Ala18, and Gly34). State X_2 is a C-terminal folding intermediate with a triple-helical C-terminal GPO repeat region that yields folded peaks for Gly34, while the rest of the peptide is disordered and yields unfolded peaks for Gly7 and Gly34. State X_3 is an N-terminal folding intermediate with a triple-helical N-terminal GPO repeat region that yields folded peaks for Gly7, while the rest of the peptide is disordered and yields unfolded peaks for Ala18 and Gly34. State X_4 is a folded triple helix that yields folded peaks for all three labeled sites.

$$A18_u = 1 - A18_f = \gamma_1 + \gamma_2 + \gamma_3 \tag{5.5}$$

$$A18_f = \gamma_4 \tag{5.6}$$

$$G34_u = 1 - G34_f = \gamma_1 + \gamma_3 \tag{5.7}$$

$$G34_f = \gamma_2 + \gamma_4 \tag{5.8}$$

$$\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 1 \tag{5.9}$$

The solution for the probabilities of states in this model is:

$$\gamma_1 = 1 - G7_f + A18_f - G34_f,\ \gamma_2 = G34_f - A18_f,\ \gamma_3 = G7_f - A18_f,\ \gamma_4 = A18_f \tag{5.10}$$

In this solution, it is apparent that there is no requirement for the triple helicities of any of the labeled sites to be equal to each other, which is consistent with our experimental data. Plugging in the fractional triple helicities for C3S1 to this solution, however, we find:

$$\gamma_1 = -0.17,\ \gamma_2 = 0.83,\ \gamma_3 = 0.22,\ \gamma_4 = 0.12$$

Because the probability of state X_1 (γ_1) is less than zero, this solution is unphysical. When using the data for C3S4, we find that:

$$\gamma_1 = 0.00,\ \gamma_2 = 0.25,\ \gamma_3 = 0.08,\ \gamma_4 = 0.67$$

While this solution is physical in the sense that no probabilities are less than zero, it is unlikely that the probability of state X_1, the unfolded monomer, is exactly zero. This would imply that the free energy of folding for this peptide was negative infinity and that no monomeric species exists in solution. Given that we have obtained an unphysical solution for C3S1 and an implausible solution for C3S4, we therefore exclude the four state model. The states listed in Figure 5-6 are clearly insufficient to explain the HSQC data.

5.5.2 Five state model

Several studies suggest that collagen and collagen-like peptides can adopt partially unfolded states in solution at temperatures that are well below the melting temperature of these systems (Ryhänen et al. 1983, Birkedal-Hansen et al. 1985, Fields 1991, Fiori et al. 2002, Stultz 2002, Nerenberg and Stultz 2008, Ravikumar and Hwang 2008, Salsas-Escat and Stultz 2009). Consequently, we included a state X_5 that features a partially unfolded triple helix in which the

imino-rich N- and C-termini are folded and the imino-poor central region is unfolded (Figure 5-7A). Such a state would contribute folded peaks for the Gly7 and Gly34 labels and an unfolded peak for the Ala18 label.

We then consider a five state model containing the states X_1-X_5:

$$G7_u = \gamma_1 + \gamma_2 \tag{5.11}$$

$$G7_f = \gamma_3 + \gamma_4 + \gamma_5 \tag{5.12}$$

$$A18_u = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_5 \tag{5.13}$$

$$A18_f = \gamma_4 \tag{5.14}$$

$$G34_u = \gamma_1 + \gamma_3 \tag{5.15}$$

$$G34_f = \gamma_2 + \gamma_4 + \gamma_5 \tag{5.16}$$

$$\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 = 1 \tag{5.17}$$

The solution for this model is:

$$\gamma_1 = free,\ \gamma_2 = 1 - \gamma_1 - G7_f,\ \gamma_3 = 1 - \gamma_1 - G34_f,\ \gamma_4 = A18_f,$$
$$\gamma_5 = \gamma_1 + G7_f - A18_f + G34_f - 1 \tag{5.18}$$

For the five state model there are more variables (five) than constraints (three triple helicities plus the constraint on the sum of the probabilities). Therefore, one of the probabilities, γ_1, is a free parameter and there can be a range of physical solutions in which all constraints are satisfied and none of the probabilities is less than zero (Figures 5-7B and 5-7C). Interestingly, in all of the physical solutions to the five state model for C3S1, the collagenase cleavage site peptide, the partially unfolded state is preferred to the folded state (Figure 5-7B). Conversely, in all of the physical solutions to the five state model for C3S4, the pseudo-cleavage site peptide, the folded state is substantially preferred to the partially unfolded state (Figure 5-7C).

(A) **(B)** **(C)**

G7

A18

G34

**Partially unfolded
triple helix (X_5)**
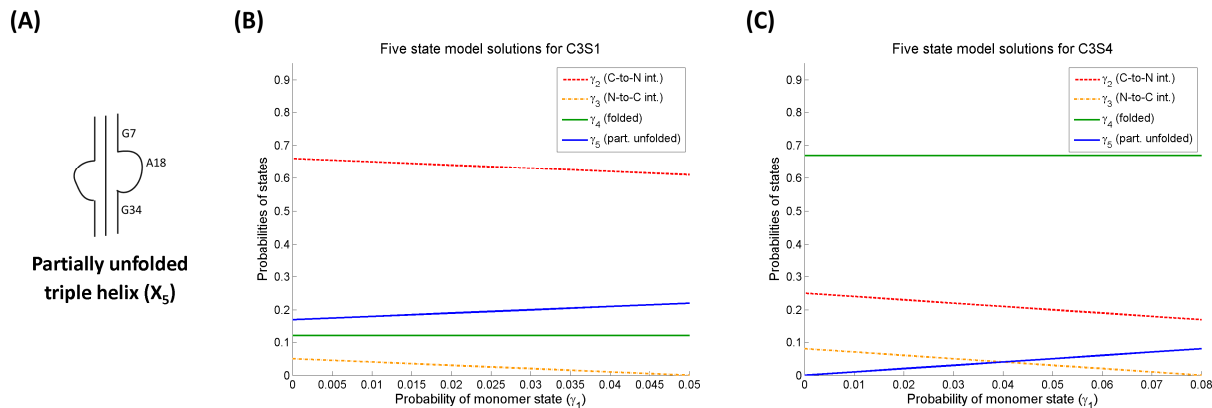
Figure 5-7: (A) Partially unfolded conformational state (X_5) of collagen-like model peptides that yields folded peaks for Gly7 and Gly34, but an unfolded peak for Ala18. Assuming a five state conformational ensemble, we calculate probabilities of the conformational states X_2-X_5 as a function of the probability of the unfolded monomer state X_1 for (B) the collagenase cleavage site peptide C3S1 and (C) the pseudo-cleavage site peptide C3S4. Because there are more variables than constraints in the model for a five state ensemble, the probability of state X_1 is a free variable. The conformational ensembles shown on the plots are consistent with the triple helicities at each labeled site, as determined by integrating the peaks from the HSQC spectra. For all physical solutions to the five state model, C3S1 prefers the partially unfolded state to the folded state, while C3S4 prefers the folded state to the partially unfolded state.

5.6 Discussion

Because of its importance in understanding the molecular mechanism of collagen degradation, the physical basis of collagenase cleavage site specificity has been the subject of considerable study during the past three decades. Prior experimental and theoretical studies have suggested that interactions between native collagen and the hemopexin-like domain of collagenases may be one factor that contributes to cleavage site specificity (Murphy et al. 1992, Borkakoti 1998, Lauer-Fields et al. 2002, Chung et al. 2004, Nerenberg et al. 2008, Perumal et al. 2008, Erat et al. 2009, Lauer-Fields et al. 2009). Other studies have postulated that the conformational ensemble of the collagenase cleavage site may play an important role in determining this specificity (Miller et al. 1976, Highberger et al. 1979, Ryhänen et al. 1983, Birkedal-Hansen et al. 1985, Fields 1991, Salsas-Escat and Stultz 2009). The goal of the present study was to investigate which, if any, of these factors might explain collagenase cleavage site specificity.

To understand the influence of binding contacts between native collagen and the non-catalytic hemopexin-like domains of collagenases, we conducted degradation experiments in

which we incubated native type III collagen with collagenase deletion mutants that lack hemopexin-like domains. These deletion mutants cleave type III collagen at the same location as full length collagenases, albeit with lower efficiencies (Figure 5-2). These results categorically demonstrate that enzyme specificity can be achieved with the catalytic domain alone, but suggest that the hemopexin-like domain still plays an important role in the degradation process (e.g., by localizing the catalytic domain to the cleavage site and increasing the overall efficiency of collagenolysis).

To assess the thermal stabilities and local triple helicities of the collagenase cleavage site and pseudo-cleavage site sequences in type III collagen, we designed and synthesized a set of homotrimeric collagen-like model peptides containing each potential cleavage site. We performed melting temperature experiments, monitoring the global triple helicities of the peptides with circular dichroism spectroscopy. We observed that the model peptide C3S1, which contains the collagenase cleavage site sequence, had the lowest overall triple helicity and lowest melting temperature of all five peptides (Figure 5-3). These data suggest that the collagenase cleavage site has markedly lower thermal stability than any of the pseudo-cleavage sites and may sample non-triple-helical conformations even at temperatures well below T_m.

Thermal stability, however, is a global measure of protein folding stability and does not directly assess the local folding characteristics of a sequence. To probe the local triple helicities of the model peptides at temperatures below T_m, we conducted ^{1}H-^{15}N HSQC NMR experiments at 10 °C, utilizing ^{15}N-labeled residues at Gly7, Ala18, and Gly34, and integrated the folded and unfolded peaks for these residues to obtain the fractional triple helicity for each site. For both C3S1 and C3S4, the lowest triple helicity is observed at Ala18, which is one residue C-terminal to the scissile bond. Additionally, comparing the data for both peptides, we find that the collagenase cleavage site peptide C3S1 is substantially less folded near the scissile bond than the pseudo-cleavage site peptide C3S4.

To better understand the conformational distributions of these peptides on the NMR timescale, we fit the NMR data to both four and five state conformational models. We found, however, that only the five state model was able to yield physical solutions to the NMR data. Notably, this model includes a partially unfolded state in which the Ala18 site was locally unfolded and solvent-exposed. The solutions to the five state model suggest that both model peptides may sample partially unfolded states in the immediate vicinity of the scissile bond at

temperatures well below the melting temperature of the triple helix. Moreover, these solutions suggest that the collagenase cleavage site peptide C3S1 preferentially samples these partially unfolded states, while the pseudo-cleavage site peptide C3S4 only rarely samples partially unfolded states. In addition, the solutions for both C3S1 and C3S4 suggest that these peptides sample more C-to-N folding intermediates than N-to-C intermediates; i.e., that these peptides fold predominantly along the C-to-N direction. These findings make sense in light of the fact that the peptides feature a longer imino-rich GPO repeat region (6 GPO triplets) at the C-terminus than at the N-terminus (3 GPO triplets) and are therefore more likely to nucleate and fold from the C-terminus.

Taken together, the melting temperature experiments, fractional triple helicities obtained from NMR experiments, and conformational models of these model peptides all support the notion that the collagenase cleavage site in type III collagen is thermally unstable and frequently samples partially unfolded states at body temperature. These data also suggest that the pseudo-cleavage sites are substantially more stable and correspondingly sample partially unfolded states less frequently. Together with recent computational studies, which suggest that the collagenase cleavage site samples partially unfolded states that more readily bind in the collagenase active site than partially unfolded states sampled by the pseudo-cleavage sites (Salsas-Escat and Stultz 2009), these data suggest that local unfolding in the vicinity of the scissile bond is the primary determinant of collagenase cleavage site specificity in type III collagen. Developing an understanding of the physical basis for collagenase cleavage site specificity represents a significant advancement in the understanding of the molecular mechanism of collagen degradation and may help direct future efforts to control collagenase activity/collagen degradation as it applies to both normal and disease-associated physiological processes.

Chapter 6

Conclusions and Future Directions

At the beginning of this work, we introduced the notion that collagen must adopt non-triple-helical conformations at the collagenase cleavage site in order to explain two paradoxical aspects of collagenolysis – that the collagenase active site cleft is too narrow and that the scissile bonds of the collagenase cleavage site are hidden within the collagen triple helix (Overall 2002, Stultz 2002, Chung et al. 2004). In principle these non-triple-helical conformations could be achieved in one of two ways: (1) an active unwinding of the triple helix by collagenases or (2) regular thermal fluctuations that result in collagen sampling non-triple-helical conformations even in the absence of enzyme. Given the fact that collagenolysis does not require energy input (Overall 2002, Chung et al. 2004) and that the estimated energetic requirements for active unwinding are large (see Chapter 2), we pursued the idea that conformational ensemble of collagen may include non-triple-helical conformations in which the scissile bonds are exposed and vulnerable to degradation (Fields 1991, Stultz 2002). Thus, the overall aim of this work has been to develop a detailed understanding of the conformational ensemble of collagen at the collagenase cleavage site and how it plays a role in the molecular mechanism of collagenolysis.

In Chapter 2, we demonstrated that a degradation mechanism in which collagen exists in two distinct conformational states, a well-folded triple-helical state (native) and a partially unfolded state (vulnerable), near the collagenase cleavage site can explain prior experimental observations. In this formalism, collagen is degraded when collagenases bind and cleave vulnerable conformers. To develop a structural model of collagen near the collagenase cleavage site, we conducted molecular dynamics (MD) simulations of type I collagen in Chapter 3. These simulations revealed that there were two distinct low energy conformations: a well-folded native state and a partially unfolded vulnerable state in which interchain hydrogen bonds to the triplet containing the cleavage site were broken. To validate this structural model for type I collagen, in Chapter 4 we conducted degradation experiments where we incubated type I collagen with deletion mutants of collagenases that contained only the catalytic domains at a temperature well

below its melting temperature. These experiments not only verified the existence of a partially unfolded vulnerable state, but also demonstrated that the specificity for the collagenase cleavage site does not require the presence of the hemopexin-like domain. Moreover, we were able to fit the resulting degradation time courses from these experiments using a slightly modified version of the conformational selection model for degradation introduced in Chapter 2. Finally, in Chapter 5 we validated the vulnerable state structural model for type III collagen using a combination of experimental and modeling techniques. We again carried out degradation experiments using collagenase catalytic domains and supplemented these experiments with circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy experiments on collagen-like model peptides. The latter experiments demonstrated that the collagenase cleavage site sequence is the least stable of any potential cleavage site sequences in type III collagen, and simple conformational models fit to the NMR data suggest that partially unfolded states are prevalent at temperatures below the melting temperature of the triple helix.

Taken together, these observations provide compelling evidence that the conformational ensembles of types I and III collagen include partially unfolded states in which the collagenase cleavage site is solvent-exposed and accessible to the active site of collagenases, even at temperatures that are low relative to the melting temperatures of these collagens. In addition, our degradation experiments demonstrate that the specificity of the collagenase cleavage site is maintained even in the absence of the hemopexin-like domain and consequently suggest that specificity is likely encoded in the unique conformational ensemble of the sequence surrounding and including the collagenase cleavage site.

6.1 The mechanism of collagenolysis with full length enzyme

In light of these observations we expand our previous reaction scheme for the catalytic domain alone to consider the effects of the hemopexin-like domain on collagenolysis (Figure 6-1). 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 et al. 1992, Overall 2002, Tam et al. 2004, Perumal et al. 2008, Lauer-Fields et al. 2009). When the native state is bound, an $N{\cdot}H$ 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 $N{\cdot}H$

complex, and the putative binding site for the hemopexin-like domain in type I collagen is removed from the cleavage site (Perumal et al. 2008, Erat 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.
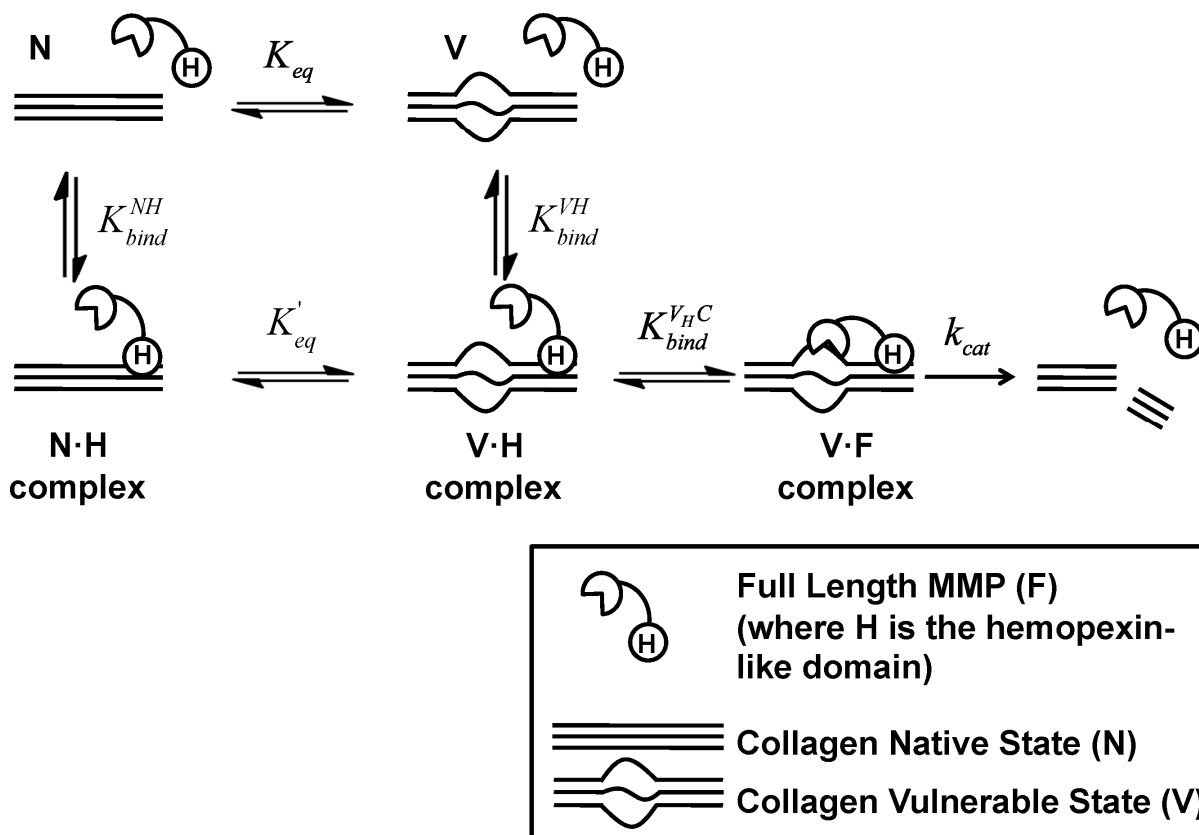
Figure 6-1: Conformational selection mechanism for collagenolysis with full length collagenases. Collagen exists in an equilibrium between native (*N*) and vulnerable (*V*) states determined by the equilibrium constant K_{eq}. Full length MMP (*F*) can bind to the native state of collagen via the hemopexin-like domain (*H*) with binding constant K_{bind}^{NH}, forming the *N·H* complex. The *N·H* complex can transition to the *V·H* complex via the equilibrium between native and vulnerable states defined by $K_{eq}^{'}$. The full length enzyme, *F*, can bind directly to the vulnerable state, *V*, via the hemopexin domain (*H*) forming the *V·H* complex, determined by the binding constant K_{bind}^{VH}. Once a *V·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·F* complex. The *V·F* complex is then degraded with catalytic rate k_{cat}.

6.1.1 The change in effective concentration of vulnerable states in the presence of enzyme

To develop a quantitative understanding of 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, $K_{eq}^{eff} = \frac{([V]+[V \cdot H]+[V \cdot F])}{([N]+[N \cdot H])}$. To simplify the calculation, we first assume that the hemopexin-like domain binds the native and vulnerable states with the same affinity ($K_{bind}^{NH} = K_{bind}^{VH} \equiv K_{bind}^{H}$). It should be noted that if the hemopexin-like domain does in fact preferentially bind and consequently stabilize the vulnerable state (i.e, $K_{bind}^{VH} > K_{bind}^{VH}$, as suggested in Chapter 2) then the preceding assumption will lead us to underestimate K_{eq}^{eff}. In this scenario, there is a closed form solution for K_{eq}^{eff} (see Appendix B):

$$K_{eq}^{eff} = K_{eq} \left(1 + \left(\frac{K_{bind}^{H}[E]}{1 + K_{bind}^{H}[E]} \right) K_{bind}^{V_H C} \right) \tag{6.1}$$

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 ($[E] = [E]_{tot} = 0$) 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 6-2. 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.
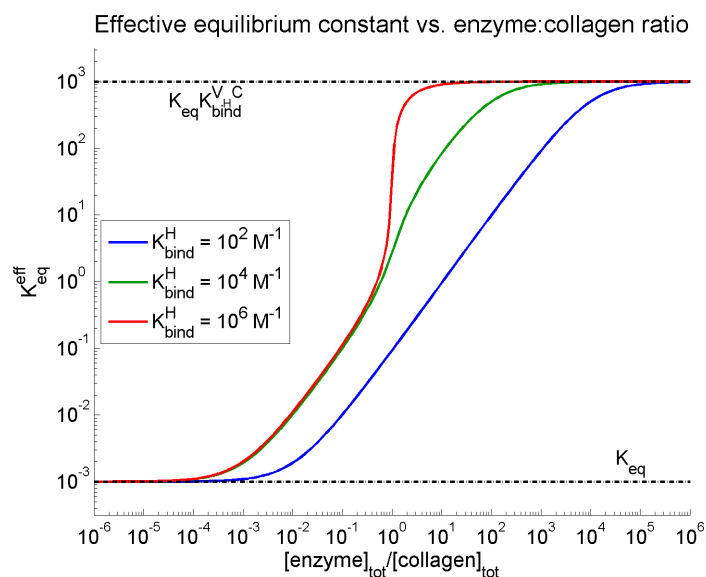
Figure 6-2: Change in the effective conformational equilibrium of collagen in the presence of full length collagenase. Computed values for the effective equilibrium constant, K_{eq}^{eff}, as a function of total enzyme-to-collagen ratio for three possible values of the hemopexin-like domain binding constant, K_{bind}^{H}, ranging from sub-millimolar to micromolar affinity. For these calculations, $[\text{collagen}]_{\text{tot}} = 10^{-6}$ 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 $K_{eq} K_{bind}^{V_H C}$.

6.1.2 Novel approaches to the inhibition of collagenolysis

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 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 6-1. Although great effort has been directed towards the design of such inhibitors, 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 et al. 2007, Georgiadis and Yiotakis 2008, Tallant et al. 2009). The reaction scheme in Figure 6-1 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 (e.g., decreasing K_{bind}^{NH} and K_{bind}^{VH}) may prove to be an effective mechanism to modulate collagenolysis (Folgueras et al. 2004, Nerenberg et al. 2007, Lauer-Fields 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. By utilizing these additional steps in the reaction mechanism, collagen degradation could potentially be inhibited in a specific and efficient manner.

6.2 The conformational ensemble of collagen *in vivo*

Throughout this work, we have conducted both experimental and computational studies with collagen (or collagen-like model peptides) in solution. In the body, however, types I and III collagen are generally not found free in solution, but rather are incorporated into various fibrillar structures as part of the extracellular matrix (Gelse et al. 2003). Thus, an implicit assumption in the conclusions of our work has been that the conformational ensemble of collagen in a fibrillar context does not differ substantially from the conformational ensemble of collagen in solution.

Previous experimental data suggest that both types I and III collagen are thermally stabilized in the fibrillar context relative to being in solution (Birkedal-Hansen et al. 1985). This in turn may result in a stabilization of the native state relative to the vulnerable state (i.e, a reduction in K_{eq}). In addition, modeling studies based on x-ray fiber diffraction data suggest that access to the collagenase cleavage site in type I collagen is limited by the telopeptides of neighboring collagen molecules in the fibril (Perumal et al. 2008), which may necessitate a modification of the basic mechanism presented in Figure 6-1. Interestingly, however, these same studies suggest that at the collagenase cleavage site the α2 chain is the most dissociated of the three polypeptide chains from the central axis of the triple helix (Perumal et al. 2008). This implies that the conformational ensemble of type I collagen in a fibrillar context may be similar to the solution state ensemble (and in particular to the conformational ensemble described in Chapter 3). In other words, the conformational ensemble of collagen at the collagenase cleavage site may be

unaffected by the additional intermolecular contacts of the fibril. Nonetheless, additional study is needed before we can be totally confident that data derived from solution state studies of collagen and collagenolysis is applicable to the *in vivo* degradation of fibrillar collagen.

One of the main tools of this work has been the use of MD simulations to elucidate the conformational ensemble of collagen at an atomistic level of detail. Currently MD simulations of fibrillar collagen have been limited to more coarse-grained methods (e.g., one "bead" per several residues (Buehler 2008)) than the ones used in Chapter 3, but the continuously increasing power of computing clusters suggests that MD studies of the conformational ensemble of collagen molecules contained in a microfibril conducted at an atomistic level of detail will become practical in the very near future. The collagen fibril presents a challenge for detailed biophysical experiments due to its sheer size and the current impossibility of being synthesized (which excludes, for example, the radiolabeling of specific residues for NMR experiments), but one possible avenue for improvement in the near future may be higher resolution structures obtained from x-ray fiber diffraction studies. The resolution of these fiber diffraction structures, however, will likely be limited to ~4 Å in the axial direction and ~10 Å in the equatorial direction (Joseph Orgel, personal communication) – a resolution that is significantly less than that typically obtained with x-ray crystallography of other peptides/proteins and insufficient to accurately model sidechain atoms (Wlodawer et al. 2008). Moreover, such structures would still present only a static picture of collagen in the fibrillar context, whereas our work suggests that collagen, like many other biopolymers, is inherently dynamic.

6.3 Other outstanding issues in collagen structure and thermodynamics

Beyond understanding the conformational ensemble of collagen, many of the basic biophysical questions regarding the structure and thermodynamics of the collagen triple helix remain unanswered and ripe for investigation using computational and experimental methods.

One such question is the mechanism by which X2 position arginines stabilize the triple helix. As mentioned in Chapter 1, arginine is quite prevalent in this position, occurring in 12-13% of all triplets in types I, II, and III collagen (Ramshaw et al. 1998). X-ray crystal structures and MD simulations of model peptides suggest that the arginine sidechain hydrogen bonding with the backbone amide oxygen of a neighboring chain may be the source of this stabilization (Kramer et al. 1999, Kramer et al. 2001, Stultz 2002). Prior experimental studies with model peptides

show, however, that this stabilization requires adequate separation of arginine residues – if the arginines occur in adjacent triplets in the amino acid sequence, they destabilize the triple helix (Yang et al. 1997). Thus, it is thought that unfavorable electrostatic interactions between these sidechains can lead to destabilization of the triple helix. These observations and hypotheses could be validated using MD simulation studies similar to those described in Appendix A, as well as other computational techniques (e.g., Poisson-Boltzmann calculations).

Another question relates to one of the common post-translational modifications in collagen – the hydroxylation of lysine residues. Like the hydroxylation of proline residues, lysine hydroxylation occurs before the folding of the triple helix (Gelse et al. 2003, Myllyharju 2005). Unlike proline hydroxylation, however, lysine hydroxylation occurs at relatively small number of sites (i.e., most lysines in the fibrillar collagens are not hydroxylated) (Myllyharju 2005). Additionally, these residues are typically involved in the intermolecular crosslinks that are formed when collagen molecules are assimilated into a fibril (Gelse et al. 2003, Myllyharju 2005). As such, there is a paucity of data on the effect of hydroxylysine on triple-helical stability. Intriguingly, increased hydroxylysine content has been correlated with osteogensis imperfecta, a class of genetic diseases in which glycine substitution mutations destabilize the triple helix and slow its folding (Tenni et al. 1993, Lehmann et al. 1995, Myllyharju and Kivirikko 2001). It is unknown whether this is simply a consequence of the collagen chains remaining in an unfolded state for a longer time (and therefore being overmodified) or if this is an alternative biological strategy for triple helix stabilization (noting that lysine by itself destabilizes the triple helix relative to many residues). A combined computational and experimental (host-guest) approach using model peptides could ascertain what, if any, effect hydroxylysine has on triple-helical stability.

In addition, hydroxylysine residues also serve as an attachment point for carbohydrates, in a process known as glycosylation (Gelse et al. 2003, Myllyharju 2005). In globular proteins, glycosylation is a ubiquitous form of post-translational modification that is thought to modify both folding stability and conformational preferences (Sola et al. 2007, Shental-Bechor and Levy 2009). While glycosylation is known to impact the packing of collagen fibrils, its effect on the triple-helical stability and the conformational preferences of collagen is unknown (Brinckmann et al. 1999, Notbohm et al. 1999, Myllyharju 2005). Here again, a combination of experiments

and molecular simulations may elucidate the molecular-level consequences of a post-translational modification.

Finally, for the sake of completeness, it would be worthwhile to conduct a comprehensive analyses of the conformational ensembles of all the potential cleavage sites of types I and II collagen, as has been done for type III collagen (Salsas-Escat and Stultz 2009). The data presented in this work support the notion that the collagenase cleavage sites in all three of these major fibrillar collagens readily sample partially unfolded states that are *uniquely* vulnerable to enzymatic cleavage. To test this hypothesis at the atomistic level of detail and unambiguously exclude other potential hypotheses, however, these data should be combined with comprehensive analyses of the conformational ensembles of all the potential cleavage sites (as in Salsas-Escat and Stultz 2009), as well as degradation experiments with type II collagen similar to those presented in Chapters 4 and 5. In addition, while this work has focused on the fibrillar collagens, it would be likewise interesting to explore the conformational ensembles of other collagens (e.g., type IV collagen, the primary constituent of the basement membrane) to see if partially unfolded states exist and may be responsible for the degradation of these proteins. Nature often has a way of recycling biological mechanisms that are robust and can work in many different contexts – it would no doubt be satisfying to be able to unify the degradation of the various collagens in the framework of a single molecular mechanism.

Bibliography

Abbenante G and Fairlie DP. Protease inhibitors in the clinic. Med Chem 2005; 1:71-104.

Barnes MJ and Farndale RW. Collagens and atherosclerosis. Exp Gerontol 1999; 34:513-525.

Bella J, Brodsky B, and Berman HM. Hydration structure of a collagen peptide. Structure 1995; 3:893-906.

Bella J, Eaton M, Brodsky B, and Berman HM. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. Science 1994; 266:75-81.

Bigg HF, Shi YE, Liu YE, Steffensen B, and Overall CM. 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 1997; 272:15496-15500.

Birkedal-Hansen H, Taylor RE, Bhown AS, Katz J, Lin HY, and Wells BR. Cleavage of bovine skin type III collagen by proteolytic enzymes. Relative resistance of the fibrillar form. J Biol Chem 1985; 260:16411-16417.

Bode MK, Mosorin M, Satta J, Risteli L, Juvonen T, and Risteli J. Complete processing of type III collagen in atherosclerotic plaques. Arterioscler Thromb Vasc Biol 1999; 19:1506-1511.

Bode W, Fernandez-Catalan C, Tschesche H, Grams F, Nagase H, and Maskos K. Structural properties of matrix metalloproteinases. Cell Mol Life Sci 1999; 55:639-652.

Borkakoti N. Matrix metalloproteases: variations on a theme. Prog Biophys Mol Biol 1998; 70:73-94.

Bornstein P and Traub W. The chemistry and biology of collagen. In: Neurath H, Hill RL, editors. The Proteins. 3rd ed. Volume 4. New York: Academic Press; 1979. p 411-632.

Bretscher LE, Jenkins CL, Taylor KM, DeRider ML, and Raines RT. Conformational stability of collagen relies on a stereoelectronic effect. J Am Chem Soc 2001; 123:777-778.

Brinckmann J. Collagens at a glance. Top Curr Chem 2005; 247:1-6.

Brinckmann J, Notbohm H, Tronnier M, Acil Y, Fietzek PP, Schmeller W, Müller PK, and Batge B. Overhydroxylation of lysyl residues is the initial step for altered collagen cross-links and fibril architecture in fibrotic skin. J Invest Dermatol 1999; 113:617-621.

Brodsky B and Persikov AV. Molecular structure of the collagen triple helix. Adv Protein Chem 2005; 70:301-339.

Brooks BR, Brooks III CL, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, and Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009; 30:1545-1614.

Bruckner P and Prockop DJ. Proteolytic enzymes as probes for the triple-helical conformation of procollagen. Anal Biochem 1981; 110:360-368.

Buehler MJ. Hierarchical Nanomechanics of Collagen Fibrils: Atomistic and Molecular Modeling. In: Fratzl P, editor. Collagen: Structure and Mechanics. New York: Springer; 2008. p 510.

Buevich AV, Dai QH, Liu X, Brodsky B, and Baum J. Site-specific NMR monitoring of cis-trans isomerization in the folding of the proline-rich collagen triple helix. Biochemistry 2000; 39:4299-4308.

Burjanadze TV. Hydroxyproline content and location in relation to collagen thermal stability. Biopolymers 1979; 18:931-938.

Caruthers JM and McKay DB. Helicase structure and mechanism. Curr Opin Struct Biol 2002; 12:123-133.

Celentano DC and Frishman WH. Matrix metalloproteinases and coronary artery disease: a novel therapeutic target. J Clin Pharmacol 1997; 37:991-1000.

Chung L, Dinakarpandian D, Yoshida N, Lauer-Fields JL, Fields GB, Visse R, and Nagase H. Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. EMBO J 2004; 23:3020-3030.

Clark IM. Matrix metalloproteinase protocols. Totowa, N.J.: Humana Press; 2001. xv, 545 p. p.

Clark IM and Cawston TE. Fragments of human fibroblast collagenase. Purification and characterization. Biochem J 1989; 263:201-206.

Conway JG, Wakefield JA, Brown RH, Marron BE, Sekut L, Stimpson SA, McElroy A, Menius JA, Jeffreys JJ, Clark RL, and et al. Inhibition of cartilage and bone destruction in adjuvant arthritis in the rat by a matrix metalloproteinase inhibitor. J Exp Med 1995; 182:449-457.

Cowan KN, Jones PL, and Rabinovitch M. Elastase and matrix metalloproteinase inhibitors induce regression, and tenascin-C antisense prevents progression, of vascular disease. J Clin Invest 2000; 105:21-34.

Dai N and Etzkorn FA. Cis-trans proline isomerization effects on collagen triple-helix stability are limited. J Am Chem Soc 2009; 131:13728-13732.

Dai N, Wang XJ, and Etzkorn FA. The effect of a trans-locked Gly-Pro alkene isostere on collagen triple helix stability. J Am Chem Soc 2008; 130:5396-5397.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, and Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995; 6:277-293.

Di Sebastiano P, di Mola FF, Artese L, Rossi C, Mascetta G, Pernthaler H, and Innocenti P. Beneficial effects of Batimastat (BB-94), a matrix metalloproteinase inhibitor, in rat experimental colitis. Digestion 2001; 63:234-239.

Dixit SN, Mainardi CL, Seyer JM, and Kang AH. Covalent structure of collagen: amino acid sequence of alpha 2-CB5 of chick skin collagen containing the animal collagenase cleavage site. Biochemistry 1979; 18:5416-5422.

Dugave C and Demange L. Cis-trans isomerization of organic molecules and biomolecules: implications and applications. Chem Rev 2003; 103:2475-2532.

Dunitz JD and Taylor R. Organic fluorine hardly ever accepts hydrogen bonds. Chem Eur J 1997; 3:89-98.

Engel J and Bächinger HP. Structure, stability and folding of the collagen triple helix. Top Curr Chem 2005; 247:7-33.

Epstein EH, Jr. (Alpha1(3))3 human skin collagen. Release by pepsin digestion and preponderance in fetal life. J Biol Chem 1974; 249:3225-3231.

Erat MC, Slatter DA, Lowe ED, Millard CJ, Farndale RW, Campbell ID, and Vakonakis I. Identification and structural analysis of type I collagen sites in complex with fibronectin fragments. Proc Natl Acad Sci U S A 2009; 106:4195-4200.

Fan P, Li MH, Brodsky B, and Baum J. Backbone dynamics of (Pro-Hyp-Gly)10 and a designed collagen-like triple-helical peptide by 15N NMR relaxation and hydrogen-exchange measurements. Biochemistry 1993; 32:13299-13309.

Fields CG, Grab B, Lauer JL, Miles AJ, Yu YC, and Fields GB. Solid-phase synthesis of triple-helical collagen-model peptides. Lett Peptide Sci 1996; 3:3-16.

Fields GB. A model for interstitial collagen catabolism by mammalian collagenases. J Theor Biol 1991; 153:585-602.

Fields GB and Prockop DJ. Perspectives on the synthesis and application of triple-helical, collagen-model peptides. Biopolymers 1996; 40:345-357.

Fields GB, Van Wart HE, and Birkedal-Hansen H. Sequence specificity of human skin fibroblast collagenase. Evidence for the role of collagen structure in determining the collagenase cleavage site. J Biol Chem 1987; 262:6221-6226.

Fiori S, Sacca B, and Moroder L. Structural properties of a collagenous heterotrimer that mimics the collagenase cleavage site of collagen type I. J Mol Biol 2002; 319:1235-1242.

Folgueras AR, Pendas AM, Sanchez LM, and Lopez-Otin C. Matrix metalloproteinases in cancer: from new functions to improved inhibition strategies. Int J Dev Biol 2004; 48:411-424.

Frank S, Boudko S, Mizuno K, Schulthess T, Engel J, and Bachinger HP. Collagen triple helix formation can be nucleated at either end. J Biol Chem 2003; 278:7747-7750.

French MF, Mookhtiar KA, and Van Wart HE. Limited proteolysis of type I collagen at hyperreactive sites by class I and II Clostridium histolyticum collagenases: complementary digestion patterns. Biochemistry 1987; 26:681-687.

Ganea E, Trifan M, Laslo AC, Putina G, and Cristescu C. Matrix metalloproteinases: useful and deleterious. Biochem Soc Trans 2007; 35:689-691.

Gelse K, Pöschl E, and Aigner T. Collagens--structure, function, and biosynthesis. Adv Drug Deliv Rev 2003; 55:1531-1546.

Georgiadis D and Yiotakis A. Specific targeting of metzincin family members with small-molecule inhibitors: progress toward a multifarious challenge. Bioorg Med Chem 2008; 16:8781-8794.

Gioia M, Fasciglione GF, Marini S, D'Alessio S, De Sanctis G, Diekmann O, Pieper M, Politi V, Tschesche H, and Coletta M. 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 2002; 277:23123-23130.

Gioia M, Monaco S, Fasciglione GF, Coletti A, Modesti A, Marini S, and Coletta M. 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 2007; 368:1101-1113.

Glanville RW, Breitkreutz D, Meitinger M, and Fietzek PP. Completion of the amino acid sequence of the alpha 1 chain from type I calf skin collagen. Amino acid sequence of alpha 1(I)B8. Biochem J 1983; 215:183-189.

Goddard TD and Kneller DG. SPARKY 3. University of California, San Francisco.

Gomis-Rüth FX. Structure and mechanism of metallocarboxypeptidases. Crit Rev Biochem Mol Biol 2008; 43:319-345.

Han S, Makareeva E, McBride DJ, Phillips CL, Schwarze U, Pace JM, Byers PH, Visse R, Nagase H, and Leikin S. Type I collagen homotrimers may alter tissue remodeling. American Society for Matrix Biology Biennial Meeting. Volume 27, Supp. 1; 2008. p 29.

Highberger JH, Corbett C, Dixit SN, Yu W, Seyer JM, Kang AH, and Gross J. Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the complete primary structure of the helical portion of the chick skin collagen alpha 1(I) chain. Biochemistry 1982; 21:2048-2055.

Highberger JH, Corbett C, and Gross J. Isolation and characterization of a peptide containing the site of cleavage of the chick skin collagen alpha 1[I] chain by animal collagenases. Biochem Biophys Res Commun 1979; 89:202-208.

Hirose T, Patterson C, Pourmotabbed T, Mainardi CL, and Hasty KA. 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 1993; 90:2569-2573.

Hohenester E and Engel J. Domain structure and organisation in extracellular matrix proteins. Matrix Biol 2002; 21:115-128.

Hollowell HN, Younvanich SS, McNevin SL, and Britt BM. Thermodynamic analysis of the low- to physiological-temperature nondenaturational conformational change of bovine carbonic anhydrase. J Biochem Mol Biol 2007; 40:205-211.

Holmgren SK, Bretscher LE, Taylor KM, and Raines RT. A hyperstable collagen mimic. Chem Biol 1999; 6:63-70.

Holmgren SK, Taylor KM, Bretscher LE, and Raines RT. Code for collagen's stability deciphered. Nature 1998; 392:666-667.

Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A 1985; 31:1695-1697.

Hyde TJ, Bryan MA, Brodsky B, and Baum J. Sequence dependence of renucleation after a Gly mutation in model collagen peptides. J Biol Chem 2006; 281:36937-36943.

Improta R, Berisio R, and Vitagliano L. Contribution of dipole-dipole interactions to the stability of the collagen triple helix. Protein Sci 2008; 17:955-961.

Jenkins CL and Raines RT. Insights on the conformational stability of collagen. Nat Prod Rep 2002; 19:49-59.

Jenkins CL, Vasbinder MM, Miller SJ, and Raines RT. Peptide bond isosteres: ester or (E)-alkene in the backbone of the collagen triple helix. Org Lett 2005; 7:2619-2622.

Jones CB, Sane DC, and Herrington DM. Matrix metalloproteinases: a review of their structure and role in acute coronary syndrome. Cardiovasc Res 2003; 59:812-823.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, and Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys 1983; 79:926-935.

Kalluri R. Basement membranes: structure, assembly and role in tumour angiogenesis. Nat Rev Cancer 2003; 3:422-433.

Knäuper V, Docherty AJ, Smith B, Tschesche H, and Murphy G. Analysis of the contribution of the hinge region of human neutrophil collagenase (HNC, MMP-8) to stability and collagenolytic activity by alanine scanning mutagenesis. FEBS Lett 1997; 405:60-64.

Kramer RZ, Bella J, Brodsky B, and Berman HM. The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. J Mol Biol 2001; 311:131-147.

Kramer RZ, Bella J, Mayville P, Brodsky B, and Berman HM. Sequence dependent conformational variations of collagen triple-helical structure. Nat Struct Biol 1999; 6:454-457.

Kuby SA. A Study of Enzymes: Enzyme Catalysis, Kinetics, and Substrate Binding. Boca Raton: CRC Press; 1990. 488 p.

Kuivaniemi H, Tromp G, Chu ML, and Prockop DJ. 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. Biochem J 1988; 252:633-640.

Kumar S, Bouzida D, Swendsen RH, Kollman PA, and Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. J Comput Chem 1992; 13:1011-1021.

Lamoureux G and Roux B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. J Chem Phys 2003; 119:3025-3039.

Lauer-Fields JL, Chalmers MJ, Busby SA, Minond D, Griffin PR, and Fields GB. Identification of specific hemopexin-like domain residues that facilitate matrix metalloproteinase collagenolytic activity. J Biol Chem 2009; 284:24017-24024.

Lauer-Fields JL, Juska D, and Fields GB. Matrix metalloproteinases and collagen catabolism. Biopolymers 2002; 66:19-32.

Lehmann HW, Rimek D, Bodo M, Brenner RE, Vetter U, Worsdorfer O, Karbowski A, and Muller PK. Hydroxylation of collagen type I: evidence that both lysyl and prolyl residues are overhydroxylated in osteogenesis imperfecta. Eur J Clin Invest 1995; 25:306-310.

Leikina E, Mertts MV, Kuznetsova N, and Leikin S. Type I collagen is thermally unstable at body temperature. Proc Natl Acad Sci U S A 2002; 99:1314-1318.

Li MH, Fan P, Brodsky B, and Baum J. Two-dimensional NMR assignments and conformation of (Pro-Hyp-Gly)10 and a designed collagen triple-helical peptide. Biochemistry 1993; 32:7377-7387.

Li Y, Brodsky B, and Baum J. NMR shows hydrophobic interactions replace glycine packing in the triple helix at a natural break in the (Gly-X-Y)n repeat. J Biol Chem 2007; 282:22699-22706.

Li Y, Kim S, Brodsky B, and Baum J. Identification of partially disordered peptide intermediates through residue-specific NMR diffusion measurements. J Am Chem Soc 2005; 127:10490-10491.

Libby P. The molecular mechanisms of the thrombotic complications of atherosclerosis. J Intern Med 2008; 263:517-527.

Liu X, Siegel DL, Fan P, Brodsky B, and Baum J. Direct NMR measurement of folding kinetics of a trimeric peptide. Biochemistry 1996; 35:4306-4313.

Mainardi CL, Hasty DL, Seyer JM, and Kang AH. Specific cleavage of human type III collagen by human polymorphonuclear leukocyte elastase. J Biol Chem 1980; 255:12006-12010.

Makareeva E, Mertz EL, Kuznetsova NV, Sutter MB, DeRidder AM, Cabral WA, Barnes AM, McBride DJ, Marini JC, and Leikin S. Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. J Biol Chem 2008; 283:4787-4798.

McDonnell S, Morgan M, and Lynch C. Role of matrix metalloproteinases in normal and disease processes. Biochem Soc Trans 1999; 27:734-740.

Miller EJ, Finch JE, Jr., Chung E, Butler WT, and Robertson PB. 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 1976; 173:631-637.

Miller EJ and Rhodes RK. Preparation and characterization of collagens. In: Cunningham LW, Fredericksen DW, editors. Methods in enzymology Vol 82, Structural and contractile proteins, Part A Extracellular Matrix. Volume 82. New York: Academic Press; 1982. p 41-44.

Morimoto K, Kawabata K, Kunii S, Hamano K, Saito T, and Tonomura B. Characterization of Type I Collagen Fibril Formation Using Thioflavin T Fluorescent Dye. Journal of Biochemistry 2009; 145:677-684.

Murphy G, Allan JA, Willenbrock F, Cockett MI, O'Connell JP, and Docherty AJ. The role of the C-terminal domain in collagenase and stromelysin specificity. J Biol Chem 1992; 267:9612-9618.

Murphy G and Knauper V. Relating matrix metalloproteinase structure to function: why the "hemopexin" domain? Matrix Biol 1997; 15:511-518.

Myllyharju J. Intracellular Post-Translational Modifications of Collagens. Top Curr Chem 2005; 247:115-147.

Myllyharju J and Kivirikko KI. Collagens and collagen-related diseases. Ann Med 2001; 33:7-21.

Nerenberg PS, Salsas-Escat R, and Stultz CM. Collagen--a necessary accomplice in the metastatic process. Cancer Genomics Proteomics 2007; 4:319-328.

Nerenberg PS, Salsas-Escat R, and Stultz CM. Do collagenases unwind triple-helical collagen before peptide bond hydrolysis? Reinterpreting experimental observations with mathematical models. Proteins 2008; 70:1154-1161.

Nerenberg PS and Stultz CM. Differential unfolding of alpha1 and alpha2 chains in type I collagen and collagenolysis. J Mol Biol 2008; 382:246-256.

Neria E, Fischer S, and Karplus M. Simulation of activation free energies in molecular systems. J Chem Phys 1996; 105:1902-1921.

Netzel-Arnett S, Fields GB, Birkedal-Hansen H, and Van Wart HE. Sequence specificities of human fibroblast and neutrophil collagenases. J Biol Chem 1991; 266:6747-6755.

Notbohm H, Nokelainen M, Myllyharju J, Fietzek PP, Müller PK, and Kivirikko KI. Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis in vitro. J Biol Chem 1999; 274:8988-8992.

Orgel JP, Irving TC, Miller A, and Wess TJ. Microfibrillar structure of type I collagen in situ. Proc Natl Acad Sci U S A 2006; 103:9001-9005.

Ottl J, Battistuta R, Pieper M, Tschesche H, Bode W, Kuhn K, and Moroder L. Design and synthesis of heterotrimeric collagen peptides with a built-in cystine-knot. Models for collagen catabolism by matrix-metalloproteases. FEBS Lett 1996; 398:31-36.

Ottl J, Gabriel D, Murphy G, Knauper V, Tominaga Y, Nagase H, Kroger M, Tschesche H, Bode W, and Moroder L. Recognition and catabolism of synthetic heterotrimeric collagen peptides by matrix metalloproteinases. Chem Biol 2000; 7:119-132.

Overall CM. Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. Mol Biotechnol 2002; 22:51-86.

Pelmenschikov V and Siegbahn PE. Catalytic mechanism of matrix metalloproteinases: two-layered ONIOM study. Inorg Chem 2002; 41:5659-5666.

Persikov AV, Ramshaw JA, Kirkpatrick A, and Brodsky B. Amino acid propensities for the collagen triple-helix. Biochemistry 2000; 39:14960-14967.

Persikov AV, Ramshaw JA, Kirkpatrick A, and Brodsky B. Peptide investigations of pairwise interactions in the collagen triple-helix. J Mol Biol 2002; 316:385-394.

Persikov AV, Ramshaw JA, Kirkpatrick A, and Brodsky B. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. Biochemistry 2005; 44:1414-1422.

Persikov AV, Xu Y, and Brodsky B. Equilibrium thermal transitions of collagen model peptides. Protein Sci 2004; 13:893-902.

Perumal S, Antipova O, and Orgel JP. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. Proc Natl Acad Sci U S A 2008; 105:2824-2829.

Peterson JT, Hallak H, Johnson L, Li H, O'Brien PM, Sliskovic DR, Bocan TM, Coker ML, Etoh T, and Spinale FG. Matrix metalloproteinase inhibition attenuates left ventricular remodeling and dysfunction in a rat model of progressive heart failure. Circulation 2001; 103:2303-2309.

Privalov PL. Stability of proteins. Proteins which do not present a single cooperative system. Adv Protein Chem 1982; 35:1-104.

Privalov PL. Stability of proteins. Proteins which do not present a single cooperative system. Adv Protein Chem 1982; 35:1-104.

Notbohm H, Nokelainen M, Myllyharju J, Fietzek PP, Müller PK, and Kivirikko KI. Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis in vitro. J Biol Chem 1999; 274:8988-8992.

Orgel JP, Irving TC, Miller A, and Wess TJ. Microfibrillar structure of type I collagen in situ. Proc Natl Acad Sci U S A 2006; 103:9001-9005.

Ottl J, Battistuta R, Pieper M, Tschesche H, Bode W, Kuhn K, and Moroder L. Design and synthesis of heterotrimeric collagen peptides with a built-in cystine-knot. Models for collagen catabolism by matrix-metalloproteases. FEBS Lett 1996; 398:31-36.

Ottl J, Gabriel D, Murphy G, Knauper V, Tominaga Y, Nagase H, Kroger M, Tschesche H, Bode W, and Moroder L. Recognition and catabolism of synthetic heterotrimeric collagen peptides by matrix metalloproteinases. Chem Biol 2000; 7:119-132.

Overall CM. Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. Mol Biotechnol 2002; 22:51-86.

Pelmenschikov V and Siegbahn PE. Catalytic mechanism of matrix metalloproteinases: two-layered ONIOM study. Inorg Chem 2002; 41:5659-5666.

Persikov AV, Ramshaw JA, Kirkpatrick A, and Brodsky B. Amino acid propensities for the collagen triple-helix. Biochemistry 2000; 39:14960-14967.

Persikov AV, Ramshaw JA, Kirkpatrick A, and Brodsky B. Peptide investigations of pairwise interactions in the collagen triple-helix. J Mol Biol 2002; 316:385-394.

Persikov AV, Ramshaw JA, Kirkpatrick A, and Brodsky B. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. Biochemistry 2005; 44:1414-1422.

Persikov AV, Xu Y, and Brodsky B. Equilibrium thermal transitions of collagen model peptides. Protein Sci 2004; 13:893-902.

Perumal S, Antipova O, and Orgel JP. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. Proc Natl Acad Sci U S A 2008; 105:2824-2829.

Peterson JT, Hallak H, Johnson L, Li H, O'Brien PM, Sliskovic DR, Bocan TM, Coker ML, Etoh T, and Spinale FG. Matrix metalloproteinase inhibition attenuates left ventricular remodeling and dysfunction in a rat model of progressive heart failure. Circulation 2001; 103:2303-2309.

Privalov PL. Stability of proteins. Proteins which do not present a single cooperative system. Adv Protein Chem 1982; 35:1-104.

Privalov PL. Stability of proteins. Proteins which do not present a single cooperative system. Adv Protein Chem 1982; 35:1-104.

Rainey JK and Goh MC. An interactive triple-helical collagen builder. Bioinformatics 2004; 20:2458-2459.

Ramachandran GN and Kartha G. Structure of collagen. Nature 1954; 174:269-270.

Ramachandran GN and Kartha G. Structure of collagen. Nature 1955; 176:593-595.

Ramshaw JA, Shah NK, and Brodsky B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. J Struct Biol 1998; 122:86-91.

Rasmussen M, Jacobsson M, and Bjorck L. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. J Biol Chem 2003; 278:32313-32316.

Ravikumar KM and Hwang W. Region-specific role of water in collagen unwinding and assembly. Proteins 2008; 72:1320-1332.

Rich A and Crick FH. The structure of collagen. Nature 1955; 176:915-916.

Rich A and Crick FH. The molecular structure of collagen. J Mol Biol 1961; 3:483-506.

Roux B. The calculation of the potential of mean force using computer simulations. Comp Phys Comm 1995; 91:275-282.

Ryhänen L, Zaragoza EJ, and Uitto J. Conformational stability of type I collagen triple helix: evidence for temporary and local relaxation of the protein conformation using a proteolytic probe. Arch Biochem Biophys 1983; 223:562-571.

Sacca B, Fiori S, and Moroder L. Studies of the local conformational properties of the cell-adhesion domain of collagen type IV in synthetic heterotrimeric peptides. Biochemistry 2003; 42:3429-3436.

Sacca B, Renner C, and Moroder L. The chain register in heterotrimeric collagen peptides affects triple helix stability and folding kinetics. J Mol Biol 2002; 324:309-318.

Sakakibara S, Inouye K, Shudo K, Kishida Y, Kobayashi Y, and Prockop DJ. Synthesis of (Pro-Hyp-Gly) n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxypyroline. Biochim Biophys Acta 1973; 303:198-202.

Salsas-Escat R and Stultz CM. Conformational selection and collagenolysis in Type III collagen. Proteins 2009.

Schnierer S, Kleine T, Gote T, Hillemann A, Knauper V, and Tschesche H. The recombinant catalytic domain of human neutrophil collagenase lacks type I collagen substrate specificity. Biochem Biophys Res Commun 1993; 191:319-326.

Shental-Bechor D and Levy Y. Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. Curr Opin Struct Biol 2009; 19:524-533.

Shigemura Y, Ando M, Harada K, and Tsukamasa Y. Possible degradation of type I collagen in relation to yellowtail muscle softening during chilled storage. Fisheries Science 2004; 70:703-709.

Shoulders MD and Raines RT. Collagen structure and stability. Annu Rev Biochem 2009; 78:929-958.

Sola RJ, Rodriguez-Martinez JA, and Griebenow K. Modulation of protein biophysical properties by chemical glycosylation: biochemical insights and biomedical implications. Cell Mol Life Sci 2007; 64:2133-2152.

Souaille M and Roux B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Comput Phys Commun 2001; 135:40-57.

Stultz CM. Localized unfolding of collagen explains collagenase cleavage near imino-poor sites. J Mol Biol 2002; 319:997-1003.

Stultz CM and Edelman ER. A structural model that explains the effects of hyperglycemia on collagenolysis. Biophys J 2003; 85:2198-2204.

Taddese S, Jung MC, Ihling C, Neubert AH, and Schmelzer CE. MMP-12 catalytic domain recognizes and cleaves at multiple sitesin human skin collagen type I and type III. Biochim Biophys Acta 2009.

Tallant C, Marrero A, and Gomis-Ruth FX. Matrix metalloproteinases: Fold and function of their catalytic domains. Biochim Biophys Acta 2009.

Tam EM, Moore TR, Butler GS, and Overall CM. 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 2004; 279:43336-43344.

Tenni R, Valli M, Rossi A, and Cetta G. Possible role of overglycosylation in the type I collagen triple helical domain in the molecular pathogenesis of osteogenesis imperfecta. Am J Med Genet 1993; 45:252-256.

Torrie GM and Valleau JP. Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. Chem Phys Lett 1974; 28:578-581.

Trelstad RL, Catanese VM, and Rubin DF. Collagen fractionation: separation of native types I, II and III by differential precipitation. Anal Biochem 1976; 71:114-118.

Tromp G, Kuivaniemi H, Stacey A, Shikata H, Baldwin CT, Jaenisch R, and Prockop DJ. Structure of a full-length cDNA clone for the prepro alpha 1(I) chain of human type I procollagen. Biochem J 1988; 253:919-922.

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, and Markley JL. BioMagResBank. Nucleic Acids Res 2008; 36:D402-408.

The UniProt Consortium. The universal protein resource (UniProt). Nucleic Acids Res 2008; 36:D190-195.

Veis A and Anesey J. Modes of Intermolecular Cross-Linking in Mature Insoluble Collagen. Journal of Biological Chemistry 1965; 240:3899-&.

Venugopal MG, Ramshaw JA, Braswell E, Zhu D, and Brodsky B. Electrostatic interactions in collagen-like triple-helical peptides. Biochemistry 1994; 33:7948-7956.

Voet D and Voet JG. Biochemistry. Hoboken: Wiley; 2004. 1616 p.

Wang HM, Chan J, Pettigrew DW, and Sodek J. Cleavage of native type III collagen in the collagenase susceptible region by thermolysin. Biochim Biophys Acta 1978; 533:270-277.

Welgus HG, Jeffrey JJ, and Eisen AZ. The collagen substrate specificity of human skin fibroblast collagenase. J Biol Chem 1981; 256:9511-9515.

Welgus HG, Jeffrey JJ, Stricklin GP, and Eisen AZ. The gelatinolytic activity of human skin fibroblast collagenase. J Biol Chem 1982; 257:11534-11539.

Welgus HG, Jeffrey JJ, Stricklin GP, Roswit WT, and Eisen AZ. Characteristics of the action of human skin fibroblast collagenase on fibrillar collagen. J Biol Chem 1980; 255:6806-6813.

Windsor LJ, Birkedal-Hansen H, Birkedal-Hansen B, and Engler JA. An internal cysteine plays a role in the maintenance of the latency of human fibroblast collagenase. Biochemistry 1991; 30:641-647.

Wlodawer A, Minor W, Dauter Z, and Jaskolski M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J 2008; 275:1-21.

Xu Y, Keene DR, Bujnicki JM, Hook M, and Lukomski S. Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. J Biol Chem 2002; 277:27312-27318.

Yang W, Chan VC, Kirkpatrick A, Ramshaw JA, and Brodsky B. Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides. J Biol Chem 1997; 272:28837-28840.

Appendix A

The Contribution of Interchain Salt Bridges to Triple Helical Stability in Collagen

This appendix is adapted from: Gurry T, Nerenberg PS, and Stultz CM. The Contribution of Interchain Salt Bridges to Triple Helical Stability in Collagen. Submitted.

Introduction

The collagens form a family of extracellular matrix proteins that play important roles in maintaining the structural integrity of a number of tissues including blood vessels, bone, ligaments and tendons (Vogel 1974, Myllyharju and Kivirikko 2004). Given the prevalence of collagen in a relatively large number of tissues, it is not surprising that collagen itself has been implicated in several common human diseases (Vogel 1974, Byers and Steiner 1992, Myllyharju and Kivirikko 2001, Myllyharju and Kivirikko 2004). Consequently, studies aimed at understanding the structure and metabolism of collagen are of particular importance.

Collagen's structure is characterized by three distinct amino acid chains that fold together to form a triple helix. This helical structure contains interchain backbone hydrogen bonds between the glycine amide protons and the carbonyl oxygens of residues in the X position of the corresponding GXY triplet on the adjacent chain (Brodsky and Persikov 2005). A high abundance of prolines in the X position and hydroxyprolines in the Y position has been noted in sequenced collagen peptides, accounting for roughly 20% of residues in these positions (Persikov et al. 2000). Proline residues induce a polyproline II-like helical conformation in individual strands, greatly reducing their entropy in the unfolded state and thereby the resulting loss of entropy upon triple helix formation (Bella et al. 1994, Shoulders and Raines 2009). The mechanism behind hydroxyproline-induced triple helical stabilization may be more complex. It was initially thought that the hydroxyl groups helped coordinate both intra- and intermolecular hydration networks, but other models suggest that the stereoelectronic effect of the hydroxyl group stabilizes a pucker of the pyrrolidine ring and backbone dihedral angles that are conducive

to the formation of triple-helical structure (Bella et al. 1994, Bretscher et al. 2001).

Biochemical experiments in which single residue mutants were created in homotrimeric model collagen peptides containing the sequence $(GPO)_3GXY(GPO)_4$, where either the X or the Y residue was mutated from proline or hydroxyproline, reveal insights into the contribution of different residues to triple-helical stability (Persikov et al. 2000). Compared to the highly stable control sequence of $(GPO)_3GPO(GPO)_4$, which had a melting temperature of 47.3˚C, a lysine in position Y lowered the melting point to a T_m of 41.5˚C, as did glutamate and aspartate residues in position X, which had T_m values of 42.9˚C and 40.1˚C, respectively (Persikov et al. 2000). However, when these mutations were combined in a model peptide with the sequence $(GPO)_3GXYGX'Y'(GPO)_3$, where the X' and Y residues were mutated simultaneously, striking stabilization of the triple helix was observed, with T_m values of 47.8˚C and 47.1˚C in the $(GPO)_3GPKGEO(GPO)_3$ and $(GPO)_3GPKGDO(GPO)_3$ mutants, respectively; i.e., stabilities comparable to a $(GPO)_8$ peptide (Persikov et al. 2005). Clearly, interactions which are not accounted for by the additive effect of individual residue stabilization were promoting triple helical stability in these peptides.

Additional studies suggest that interchain electrostatic interactions between lysine residues in the Y position and acidic residues in the X' position underlie this unexpected increase in stability (Persikov et al. 2005). In particular, experiments performed at a pH below the pKa of the glutamate side chain significantly reduces the melting temperature of the $(GPO)_3GPKGEO(GPO)_3$ peptide (Persikov et al. 2005). Moreover, energy minimization of triple-helical structures *in vacuo*, and with limited water molecules, suggest that *intra*-chain salt bridges are energetically favorable as long as the oppositely charged residues are separated by at most 2 residues (Katz and David 1990). Furthermore, χ^2 tests on the frequency of KGE and KGD residue pairs in fibrillar collagen sequences suggest that these residues occur more frequently together than would be expected by their individual frequencies in collagen alone, suggesting that these regions may play additional roles in modulating collagen stability (Persikov et al. 2005). Despite these results, the precise energetic contribution of *interchain* salt bridges to triple-helical stability has not been fully explored.

To clarify the energetic contribution that interchain salt bridges have on triple-helical stability, we calculated the relative contribution of salt bridge interactions to the folding free energy of the triple-helical structure. As each amino acid chain in collagen has a precise chain

126

stagger, three distinct salt bridge configurations are possible – all of which are examined in this work. In this manner, we present a comprehensive assessment of the relationship between salt bridge formation and triple-helix stability.

Methods

Construction of the initial models

Peptides with capped ends (acetylated on the N-terminal residues and amidated on the C-terminal residues) were built using the Triple-Helical Collagen Building Script (THe BuScr) 1.06 and CHARMM 35b2 with the CHARMM22/CMAP all-atom parameters (Mackerell et al. 2004, Rainey and Goh 2004, Brooks et al. 2009). The peptides were then solvated with 4946 (for the $(GPO)_3GPKGEO(GPO)_3$ peptide) or 4952 (for the $(GPO)_3GPKGDO(GPO)_3$ peptide) TIP3P water molecules using the MMTSB tool set (Feig et al. 2004). Water molecules were subjected to a cylindrical stochastic boundary potential of radius 22.5Å and length 100Å using the Miscellaneous Mean Field Potential (MMFP) with one cylindrical and two planar constraints. Prior to solvation, the model triple helices were brought to salt bridge configurations by restraining the distance between the lysine Nζ and the glutamate Cδ or aspartate Cγ, initially with a force constant of 20 kcal mol^{-1}Å^{-2} for 500 steps of steepest decent and 1000 steps of ABNR minimization, followed by 1000 more of ABNR with a force constant of 2 kcal mol^{-1}Å^{-2}.

Molecular dynamics simulations

Each system was equilibrated for 250ps followed by 10ns of production molecular dynamics simulations using CHARMM 35b2 (Brooks et al. 2009). The system was linearly heated to a temperature of 298.15K over 50ps during equilibration and then coupled to a Nosé-Hoover set at the same temperature (Evans and Holian 1985). Coordinates were saved every 0.1ps. Initial simulations suggested that protein would adopt slightly bent conformations where the ends of the molecule would wander outside of the simulation cylinder. Similar observations have been made in other simulations of triple-helical peptides (Ravikumar and Hwang 2008). To ensure that physiologically relevant (i.e., triple-helical) states were sampled and that the peptides remained within the solvent region for the duration of the simulations, all simulations employed a harmonic constraints on the first and last three residue backbone atoms with a force constant of 2 kcal mol^{-1}Å^{-2}.

Umbrella sampling

We calculated the free energy of intra-chain salt bridge formation using umbrella sampling (Torrie and Valleau 1974, Roux 1995). The approach outlined here is similar to what was described in our previous work (Nerenberg and Stultz 2008). The reaction coordinate was defined as the inter-atomic distances between the lysine $N\zeta$ and the glutamate $C\delta$ (in the case of $(GPO)_3GPKGEO(GPO)_3$), or between the lysine $N\zeta$ and the aspartate $C\gamma$ (in the case of $(GPO)_3GPKGDO(GPO)_3$). The carboxyl carbons were chosen as opposed to the carboxyl oxygens because measuring inter-atomic distances to the carbon atoms circumvents ambiguities that arise from rotations in the $C\gamma$-$C\delta$ bond in the glutamate or the $C\beta$-$C\gamma$ bond in aspartate (e.g., each carbonyl oxygen can accept a hydrogen bond from the lysine side chain). Simulations of the triple-helical peptide suggested that an inter-atomic distances of 3.6Å and 3.4Å correspond to a salt bridges in the $(GPO)_3GPKGEO(GPO)_3$ and $(GPO)_3GPKGDO(GPO)_3$ peptides, respectively.

Values of the reaction coordinate, ξ, were sampled starting from 2.4Å and up to 13.0Å in 0.2Å increments. In practice, the starting value for these runs was 3.4Å and sampling was done from 3.4→13.0Å and 3.4Å→2.4Å. A harmonic biasing potential (with a force constant of 80 kcal mol^{-1}Å^{-2}) was introduced at each window to bias the system towards a particular value of ξ. Each window was run for 250ps. Trajectories were saved every 0.1ps, resulting in 2500 frames per window. For each window, we plotted the running average of the system's internal energy and fit an exponential distribution to determine the relaxation time constant, τ. The equilibration period was then defined as 3τ. The production portion of the trajectory began at the end of this equilibration period. Data from the production runs was combined to form the unbiased probability distribution using the WHAM algorithm (Kumar et al. 1992, Souaille and Roux 2001). Similar to what was described above, umbrella sampling simulations employed additional restraints on the first and last three terminal residues (force constant of 2 kcal mol^{-1}Å^{-2} as noted in the previous section) to ensure that we simulated states in the vicinity of the triple-helical state, and that the peptides remained within the solvated stochastic boundary sphere for the duration of the simulations. To ensure that the precise choice of the harmonic restraint did not unduly influence our results, we computed the energy associated with these terminal residue harmonic constraints for all umbrella sampling windows. Windows where the contribution from these harmonic restraints was significant (>1.0 kcal mol^{-1}) were excluded from the pmf calculation. With this convention only windows centered about values of ξ greater than 12.0Å

for the C-A interaction in $(GPO)_3GPKGEO(GPO)_3$ were excluded.

Throughout this work, we adopt the notation ξ_{AB} to denote the inter-atomic distance between the lysine on chain A and acidic residue on chain B. Similar definitions apply for ξ_{BC}, and ξ_{CA}.

Results

In tropocollagen the different collagen chains are staggered with respect to each other resulting in asymmetric interactions between chains. Consequently, interchain salt bridges between the A-B, B-C and C-A pairs have distinct orientations (Figure 1). In light of this, we examined the effect of each type of salt bridge on triple-helical stability.

Dynamical Trajectories of Salt-Bridge Containing Peptides

We begin with an analysis of thermal fluctuations in interchain salt bridge containing structures. Ten nanosecond simulations of the equilibrated $(GPO)_3GPKGEO(GPO)_3$ (henceforth referred to as the GPKGEO peptide) and $(GPO)_3GPKGDO(GPO)_3$ (henceforth referred to as the GPKGDO peptide) systems were performed to observe the behavior of the interaction partners. To characterize the behavior of the salt bridge pair, we use the inter-atomic distance between the Nζ of lysine and the Cδ glutamate (in the case of the GPKGEO peptide), and the distance between the Nζ of lysine and the Cγ of aspartate (in the case of GPKGDO peptide).

Simulations of the GPKGEO triple-helical peptide highlight the difference between the interchain interactions (Figure 2). The inter-residue distance between the side chains of Lys and Glu in the A-B interaction mainly fluctuates between two states in GPKGEO – the first has an average distance of 3.6Å (which corresponds to a salt bridge between the two residues), and the second has an average distance of 5.9Å (where the salt bridge is broken). Overall, during the 10ns simulation the A-B salt bridge is formed approximately 42% of the time (Figure 2). The B-C salt bridge is stable about a mean value of 3.6Å for 74% of the 10ns. The C-A salt bridge forms a salt bridge 45% of the time with a mean distance of 3.6Å and what appears to be a secondary state at 5.2Å (Figure 2).

By contrast, in simulations of the GPKGDO peptide, the A-B salt bridge is stable during the 10ns (formed 96% of the time) about a mean value of 3.4Å. The B-C salt bridge is slightly less so (86% of the time), also about a mean value of 3.4Å, and the system appears to sample a secondary state with average distance 5.8Å, along with another state at higher values of ξ_{BC}.

129

The C-A salt bridge appears to be the least stable of the three, being formed 44% of the time. When not in a salt bridge, the system samples a secondary state with mean value 6.0Å (Figure 3).

The Effect of Salt Bridge Formation on Triple-helical Stability

To quantitatively assess the effect of interchain salt bridge formation on triple-helical stability, we use the thermodynamic cycle shown in Figure 4. In path 1, the three collagen chains fold to a state that contains a salt bridge ($\Delta G_{\text{fold}}^{\text{on}}$ is the associated free energy change), and in path 2 folding leads to a triple helical structure where the salt bridge is broken ($\Delta G_{\text{fold}}^{\text{off}}$ is the associated free energy change). The contribution of the salt bridge to triple-helical stability is given by the difference $\Delta\Delta G = \Delta G_{\text{fold}}^{\text{on}} - \Delta G_{\text{fold}}^{\text{off}}$. In principle, one could perform simulations to directly compute $\Delta G_{\text{fold}}^{\text{on}}$ and $\Delta G_{\text{fold}}^{\text{off}}$, however the direct calculation of the free energy associated with such folding reactions is computationally demanding. This is especially true for this system because simulating the unfolded state would require sampling all possible configurations of three dissociated chains. However, as the thermodynamic path shown in Figure 4 demonstrates, the effect of salt bridge formation on protein stability is equivalent to the free energy associated with salt bridge formation in the folded state; i.e., $\Delta\Delta G$ can be calculated from an analysis of the folded state alone.

The free energy difference associated with salt bridge formation, $\Delta\Delta G$, could be computed using a number of methods – each of which makes a different set of assumptions. For example, one could slowly turn off the electrostatic interactions between the two side chains that form the salt bridge and use thermodynamic integration to compute the resulting free energy change (Hendsch and Tidor 1993, Du et al. 2000). Strictly speaking, this method constitutes one way to calculate the electrostatic contribution to $\Delta\Delta G$, however, the overall electrostatic contribution is somewhat distinct from the contribution of a "salt bridge". In particular, two oppositely charged side chains that are far apart are not said to be in a salt bridge. Since electrostatic interactions are, in general, long range in character, these well separated side chains will still make some electrostatic contribution to the overall energy of the system. Hence the "broken salt-bridge" state does not correspond to a state where all electrostatic interactions between the side chains are turned off.

Here we use umbrella sampling to compute the contribution of the salt bridge to protein

stability. In this approach, the beginning and end states of the reaction are defined solely by the distance between the two side chains in question and long range electrostatic interactions are considered in both the state where the salt bridge is formed and the state where it is broken. Central to this method is a clear definition of the states corresponding to the conformation where the salt bridge is formed and the state corresponding to the salt bridge being broken. To identify these end states we use the dynamical trajectories shown in Figures 1 and 2 as a guide. In each trajectory, the system fluctuates between two distinct regions of conformational space, where the first region corresponds to a relatively low value of the reaction coordinate having representative structures that place the oppositely charged side chains within hydrogen bonding distance of one another. The second state corresponds to a larger value of the reaction coordinate where representative structures have side chains that are too far apart to allow direct hydrogen bonding to happen. We therefore begin our analysis with an investigation of the pmf for each interchain salt bridge in the folded state and interpret these findings in light of the dynamical trajectories discussed above. Insights from this analysis are then used to compute quantitative estimates for $\Delta\Delta G$.

Potentials of mean force for salt bridges in GPKGEO

The pmf for the A-B interchain interaction contains two well defined minima (Figure 5). The global energy minimum is located at ξ_{AB}=3.3Å, and corresponds to the state where an interchain salt bridge is formed between the lysine and glutamate side chains, consistent with the state with mean value 3.6Å in the dynamical trajectory in Figure 2. The second minimum is located at ξ_{AB}=5.5Å, a value similar to 5.9Å, the mean value of the secondary state in Figure 2. The relative free energy difference between the two states is quite modest – only 0.6 kcal/mol. This observation is in qualitative agreement with the data shown in Figure 2. In particular, unrestrained simulations that begin with the AB salt bridge, frequently sample states where the salt bridge is broken.

The pmf for the B-C interaction also contains two energy minima, where the global energy minimum corresponds to an interchain salt bridge at ξ_{BC}=3.7Å (Figure 6). The second minimum is more shallow and is located at ξ_{BC}=5.7Å and has a free energy that is 2.4 kcal/mol ($\sim 4k_BT$ at room temperature) higher than the ground state. Indeed, unrestrained simulations suggest that B-C salt-bridges are considerably more stable (Figure 2).

The pmf for the C-A interaction contains two energy minima (Figure 7). The global energy minimum is found at ξ_{CA}=3.7Å and corresponds to an interchain salt bridge. Again, the second minimum is shallow and is located at ξ_{CA}=4.7Å. The first of these two values is consistent with the mean values extracted from the trajectories in Figure 2 (3.6Å). The second state has a value ξ_{CA} that is somewhat lower than the average value in Figure 2, but representative structures in both cases correspond to states where the side chains are well beyond hydrogen bond distance. The second state has free energy 1.2 kcal/mol relative to the ground state.

Analysis of solvent molecules within hydrogen-bonding distance of the lysine and glutamate side chains suggests that in each case, the secondary local energy minimum forms when the side chains separate enough to allow water molecules to enter the space between the side chains. For example, for the A-B interaction, at a separation of 5.5Å a single water molecule can hydrogen bond to both side chains. An analysis of the window that is restrained to sample around ξ_{AB}=5.5Å, suggests that different water molecules take turns hydrogen bonding to both side chains. Overall, an intermediate water molecule is present 94% of the time in this state. Similarly, for metastable states in the B-C and C-A interactions, water molecules hydrogen bond to both side chains 77%, and 69% of the time, respectively.

Potentials of mean force for salt bridges in GPKGDO

The pmf for the A-B interaction in GPKGDO contains a prominent energy minimum located at ξ_{AB}=3.3Å (Figure 8), consistent with the value of 3.4Å observed in the dynamical trajectory in Figure 3. This state corresponds to a conformation where an interchain salt bridge is formed between the lysine and the aspartate residues. Trajectories of the triple-helical conformation infrequently sample a state having ξ_{AB}~5Å that corresponds to a state where the salt bridge is broken (Figure 3). This state corresponds to a shallow "shelf" on the conformational free energy surface which has a free energy that is 3.6 kcal/mol higher than the ground state (Figure 8).

The pmf for the B-C interaction in GPKGDO contains two energy minima (Figure 9). The global energy minimum is found at ξ_{BC}=3.3Å (also similar to the value of 3.4Å in Figure 3), and corresponds to an interchain salt bridge. The second energy minimum is found at ξ_{BC}=5.0Å, that has a free energy which is 2.9 kcal/mol relative to the ground state (approximately $5k_BT$ at room temperature). Consequently, this state is infrequently sampled in dynamical trajectories that begin with the B-C salt bridge formed (Figure 3).

Both metastable states for the A-B and B-C interaction occur at interatomic distances that allow individual water molecules to hydrogen bond to both states, in a manner similar to what was described for the GPKGEO trajectories above. For the shallow metastable state for the A-B interaction, approximately 92% of the time different water molecules again take turns hydrogen bonding to the two side chains. For the B-C metastable state, this occurs approximately 89% of the time.

The pmf for the C-A interaction in GPKGDO has two energy minima (Figure 10). The global minimum is located at ξ_{CA} =3.7Å, which corresponds to an interchain salt bridge. In addition, a secondary minimum can be found at ξ_{CA} =5.4Å with a relative free energy of 1.1 kcal/mol. Despite distances similar to the secondary states in previously described interaction pairs indicative of an intermediate water molecule, we only find individual water molecules 32% of the time. The secondary minimum also has the characteristic of spanning a greater range than that in other interaction pairs; values of ξ_{CA} between 4.5Å and 6.8Å are all below 2 kcal/mol. Such larger distances are therefore likely to correspond to states where several water molecules intervene, bridging the two side chains in a more complicated array. This is consistent with the dynamical timeseries in Figure 3, which appears to sample states in this vicinity when not in a salt bridge (defined as having values of ξ_{CA} >4.2Å).

Calculation of ΔΔ*G* values

We use the different pmfs to calculate ΔΔ*G* – the contribution of each salt bridge to triple-helical stability – as outlined in Figure 4. ΔΔ*G* is then calculated as the difference in energy between the salt bridge state and the local energy minima (or the relatively flat region of the conformational free energy surface a la A-B interaction in the GPKGDO peptide) corresponding to conformations where the salt bridges are broken. These values are shown in Table 1. In all cases, salt bridge formation stabilizes the triple-helical structure. However, for the GPKDEO peptide, B-C salt bridges are the most stabilizing and in the GPKGDO peptide, A-B salt bridge pairs are the most favorable.

Discussion

The effect of salt bridges on protein stability has long been a subject of great interest (Horovitz et al. 1990, Hendsch and Tidor 1993, Sindelar et al. 1998, Makhatadze et al. 2003). It has been

demonstrated that for globular proteins that the contribution of salt bridges to protein stability can vary greatly depending on the system studied. Hyperthermophilic proteins, for example, generally contain more salt bridges than their normophilic counterparts and it is believed that these salt bridges contribute, on average, to the increased thermal stability of these systems (Kumar and Nussinov 2001). The effects of salt bridges on normophilic globular proteins are still somewhat controversial. There are data to suggest that salt bridge formation in these systems is often not stabilizing and when it is, it likely only marginally contributes to overall stability (Horovitz et al. 1990, Hendsch and Tidor 1993, Sindelar et al. 1998).

The effect of salt bridge formation on fibrillar proteins, like collagen, has yet to be fully explored. Previous work has demonstrated that the introduction of single acidic or basic residues into $(GPO)_3GXY(GPO)_4$ host-guest peptides leads to a significant decrease in the peptide melting temperature (Persikov et al. 2000). However, double mutants $(GPO)_3GXYGX'Y'(GPO)_3$ that contain pairs of oppositely charged residues led to an unexpected increase in protein stability, suggesting that interchain salt bridges may be an important regulator of triple-helical stability (Persikov et al. 2005).

In the present study we explored the energetic contribution of salt bridges to the thermodynamics of triple-helix folding. First, we demonstrate that an analysis of the relative effect of salt-bridge formation on triple-helical folding can be determined from an analysis of the folded state alone. Calculated conformational free energy landscapes for the folded triple-helical structure were computed using a reaction coordinate that varied the distance between the side chains of the oppositely charged residues, thereby sampling states where the salt bridge is formed and states where it is broken. All of the resulting free energy profiles contain a prominent global energy minimum corresponding to the conformation that contains a salt bridge and a second metastable state that corresponds to a conformation where the side chains are too far apart. In each case, the global energy minimum and the metastable state correspond to states sampled in unrestrained dynamical trajectories of the salt bridge pairs.

Interestingly, we found that salt bridges can exhibit a range of energetic contributions to triple-helical stability. Some salt bridges have only a marginal effect on stability (A-B pair in GPKGEO), while others significantly stabilize the folded state (B-C in GPKGEO and A-B in GPKGDO). Overall, we found that salt bridges can stabilize collagen by 0.6-3.6 kcal/mol, which is similar to the folding stability (1-5 kcal/mol) contributed by both buried and surface salt

bridges in globular proteins (Anderson et al. 1990, Tissot et al. 1996, Strop and Mayo 2000, Makhatadze et al. 2003). The asymmetry between interchain interactions is a critical property of interchain electrostatic interactions in collagen. Both an unrestrained simulation (Figure 3) and the $\Delta\Delta G$ value calculated from the potential of mean force (Figure 8) suggest that the A-B salt bridge is the most stabilizing of the three possible salt bridge interactions in GPKGDO, presumably due to the favorable geometries of the lysine and aspartate residues in question for adopting this conformation. In contrast, the C-A salt bridge of the same peptide offers a mere 1.1 kcal/mol of stability compared to the metastable state. In this latter interaction pair, the two residues' Cα atoms are brought into close proximity. The length of the lysine side chain as compared to that of aspartate requires unfavourable stereochemical contortions to adopt a salt bridge conformation. This stereochemistry is relaxed at higher distances, and is likely to be responsible for the relative stability of the secondary state rather than intermediate water molecules.

Recently obtained experimental data in which a heterotrimeric model peptide comprised of $(POG)_{10}$, $(PKG)_{10}$ and $(DOG)_{10}$ chains was observed to form in a single triple-helical register in solution (Fallas et al. 2009). In these experiments, $(POG)_{10}$, $(PKG)_{10}$ and $(DOG)_{10}$ peptides chains were combined in a 1:1:1 ratio, heated to 85°C, then cooled to 25°C and incubated overnight at room temperature (Gauba and Hartgerink 2007). In principle, this mixture could yield three different homotrimers and six different heterotrimers (not including different possible chain registers). However, a single heterotrimer – one that included one copy of each peptide sequence (deemed the KDO peptide) – was the dominant product (Fallas et al. 2009). Moreover, this heterotrimer was found to exist in only one of six possible chain registers; multi-dimensional NMR experiments unambiguously identified chain A as the $(PKG)_{10}$ chain, chain B as the $(DOG)_{10}$ chain, and chain C as the $(POG)_{10}$ chain, as well as demonstrated the formation of interchain salt bridges, involving lysine and aspartate residues, between chains A and B (Fallas et al. 2009). Thus, the observed chain register for the heterotrimeric DKO peptide is consistent with our observation that the lysine-aspartate salt bridge formed between chains A and B is the most energetically favorable.

Our results shed light on the structural basis for the triple helical stabilization that results in peptides containing these residue pairs. Interchain electrostatic interactions may play an important role in local stabilization of vertebrate triple helical collagen in regions requiring

added stability. Furthermore, it has been noted that a high natural abundance of these residue pairs occurs in viral and bacterial collagen-like peptides (Xu. et al. 2002, Rasmussen et al. 2003). Prokaryotes lack the prolyl hydroxylase enzyme responsible for hydroxylation of the proline residues in the Y position (Mohs et al. 2007), and may therefore call on other mechanisms to stabilize the triple helical structure of such proteins. Also, viruses infecting hosts which lack this enzyme may require similar alternatives to enhance the stability of collagen-like genes. It may be possible to adopt similar approaches in the construction of synthetic collagens by introducing salt bridges in strategic locations.

Experimental melting temperatures have confirmed the ability of these residue pairs to recover the stability of a pure GPO-containing peptide (Persikov et al. 2005, Fallas et al. 2009). Stability of the triple helix in regions lacking hydroxyproline may rely strongly on the presence of these salt bridges, and their disruption could lead to abnormalities in the structure and metabolism of collagen. A detailed biophysical understanding of the contribution of salt bridges to triple helical stability could therefore form the basis of detailed sequence comparisons between diseased and healthy individuals in an up-and-coming age of high-throughput genomic sequencing.

Acknowledgements

References

Anderson DE, Becktel WJ, and Dahlquist FW. pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. Biochemistry 1990; 29:2403-2408.

Bella J, Eaton M, Brodsky B, and Berman HM. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. Science 1994; 266:75-81.

Bretscher LE, Jenkins CL, Taylor KM, DeRider ML, and Raines RT. Conformational stability of collagen relies on a stereoelectronic effect. J Am Chem Soc 2001; 123:777-778.

Brodsky B and Persikov AV. Molecular structure of the collagen triple helix. Adv Protein Chem 2005; 70:301-339.

Brooks BR, Brooks III CL, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, and Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009; 30:1545-1614.

Byers PH and Steiner RD. Osteogenesis imperfecta. Annu Rev Med 1992; 43:269-282.

Du Q, Beglov D, and Roux B. Solvation Free Energy of Polar and Nonpolar Molecules in Water: An Extended Interaction Site Integral Equation Theory in Three Dimensions. J Phys Chem B 2000; 104.

Evans DJ and Holian BL. The Nose-Hoover thermostat. J Chem Phys 1985; 83:4069-4074.

Fallas JA, Gauba V, and Hartgerink JD. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. J Biol Chem 2009; 284:26851-26859.

Feig M, Karanicolas J, and Brooks III CL. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Gr Mod 2004; 22:377-395.

Gauba V and Hartgerink JD. Surprisingly High Stability of Collagen ABC Heterotrimer: Evaluation of Side Chain Charge Pairs. J Am Chem Soc 2007; 129:15034-15041.

Hendsch ZS and Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Science 1993; 3:211-226.

Horovitz A, Serrano L, Avron B, Bycroft M, and Fersht AR. Strength and Cooperativity of Contributions of Surface Salt Bridges to Protein Stability. Journal of Molecular Biology 1990; 216:1031-1044.

Katz EP and David CW. Energetics of Intrachain Salt-Linkage Formation in Collagen. Biopolymers 1990; 29:791-798.

Kumar S, Bouzida D, Swendsen RH, Kollman PA, and Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. J Comput Chem 1992; 13:1011-1021.

Kumar S and Nussinov R. How do thermophilic proteins deal with heat? Cell Mol Life Sci 2001; 58:1216-1233.

Mackerell AD, Jr., Feig M, and Brooks III CL. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 2004; 25:1400-1415.

Makhatadze GI, Loladze VV, Ermolenko DN, Chen X, and Thomas ST. Contribution of Surface Salt Bridges to Protein Stability: Guidelines for Protein Engineering. J Mol Biol 2003; 327:1135-1148.

Mohs A, Silva T, Yoshida T, Amin R, Lukomski S, Inouye M, and Brodsky B. Mechanism of Stabilization of a Bacterial Collagen Triple Helix in the Absence of Hydroxyproline. J Biol Chem 2007; 282:29757-29765.

Myllyharju J and Kivirikko KI. Collagens and collagen-related diseases. Ann Med 2001; 33:7-21.

Myllyharju J and Kivirikko KI. Collagens, modifying enzymes and their mutations in humans, flies and worms. Tr Gen 2004; 20:33-43.

Nerenberg PS and Stultz CM. Differential unfolding of α1 and α2 chains in type I collagen and collagenolysis. J Mol Biol 2008; 382:246-256.

Persikov AV, Ramshaw JAM, Kirkpatrick A, and Brodsky B. Amino acid propensities for the collagen triple-helix. Biochemistry 2000; 39:14960-14967.

Persikov AV, Ramshaw JAM, Kirkpatrick A, and Brodsky B. Electrostatic Interactions Involving Lysine Make Major Contributions to Collagen Triple-Helix Stability. Biochemistry 2005; 44:1414-1422.

Rainey JK and Goh MC. An interactive triple-helical collagen builder. Bioinformatics 2004; 20:2458-2459.

Rasmussen M, Jacobsson M, and Bjorck L. Genome-based Identification and Analysis of Collagen-related Structural Motifs in Bacterial and Viral Proteins. J Biol Chem 2003; 278:32313-32316.

Ravikumar KM and Hwang W. Region-specific role of water in collagen unwinding and assembly. Proteins 2008; 72:1320-1332.

Roux B. The calculation of the potential of mean force using computer simulations. Comp Phys Commun 1995; 91:275.

Shoulders MD and Raines RT. Collagen structure and stability. Annu Rev Biochem 2009; 78:929-958.

Sindelar CV, Hendsch ZS, and Tidor B. Effects of salt bridges on protein structure and design. Protein Science 1998; 7:1898-1914.

Souaille M and Roux B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Comp Phys Commun 2001; 135:40-57.

Strop P and Mayo SL. Contribution of surface salt bridges to protein stability. Biochemistry 2000; 39:1251-1255.

Tissot AC, Vuilleumier S, and Fersht AR. Importance of two buried salt bridges in the stability and folding pathway of barnase. Biochemistry 1996; 35:6786-6794.

Torrie GM and Valleau JP. Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. Chem Phys Lett 1974; 28:578-581.

Vogel HG. Correlation Between Tensile Strength and Collagen Content in Rat Skin. Effect of Age and Cortisol Treatment. Conn Tis Res 1974; 2:177-182.

Xu. Y, Keene DR, Bujnicki JM, Hook M, and Lukomski S. Streptococcal Scl1 and Scl2 Proteins Form Collagen-like Triple Helices. J Biol Chem 2002; 277:27312-27318.

	GPKGEO $\Delta\Delta G$ (kcal/mol)	GPKGDO $\Delta\Delta G$ (kcal/mol)
A-B	-0.6	-3.6
B-C	-2.4	-2.9
C-A	-1.3	-1.1

Table 1: The free energy contribution of each salt bridge in the GPKGEO and GPKGDO peptides to triple-helical stability.

Chain A : -G-P-O-G-P-O-**G-X-Y-G-X′-Y′**-G-P-O-G-P-O-

Chain B : -G-P-O-G-P-O-**G-X-Y-G-X′-Y′**-G-P-O-G-P-O-

Chain C : -G-P-O-G-P-O-**G-X-Y-G-X′-Y′**-G-P-O-G-P-O-

Chain A : -G-P-O-G-P-O-**G-X-Y-G-X′-Y′**-G-P-O-G-P-O-
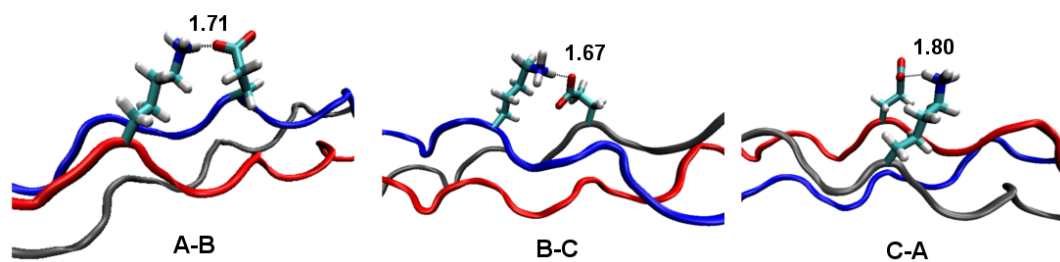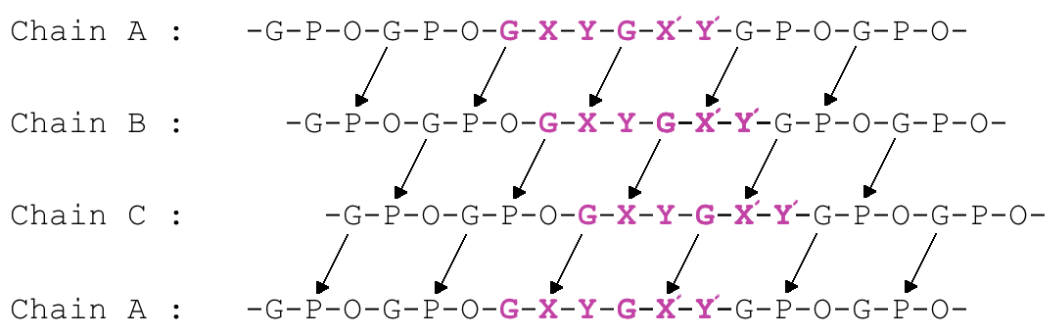
Figure 1: Backbone hydrogen bonding pattern of the a collagen-like sequence (above) and the three interchain interactions we consider – A-B, B-C and C-A (below).
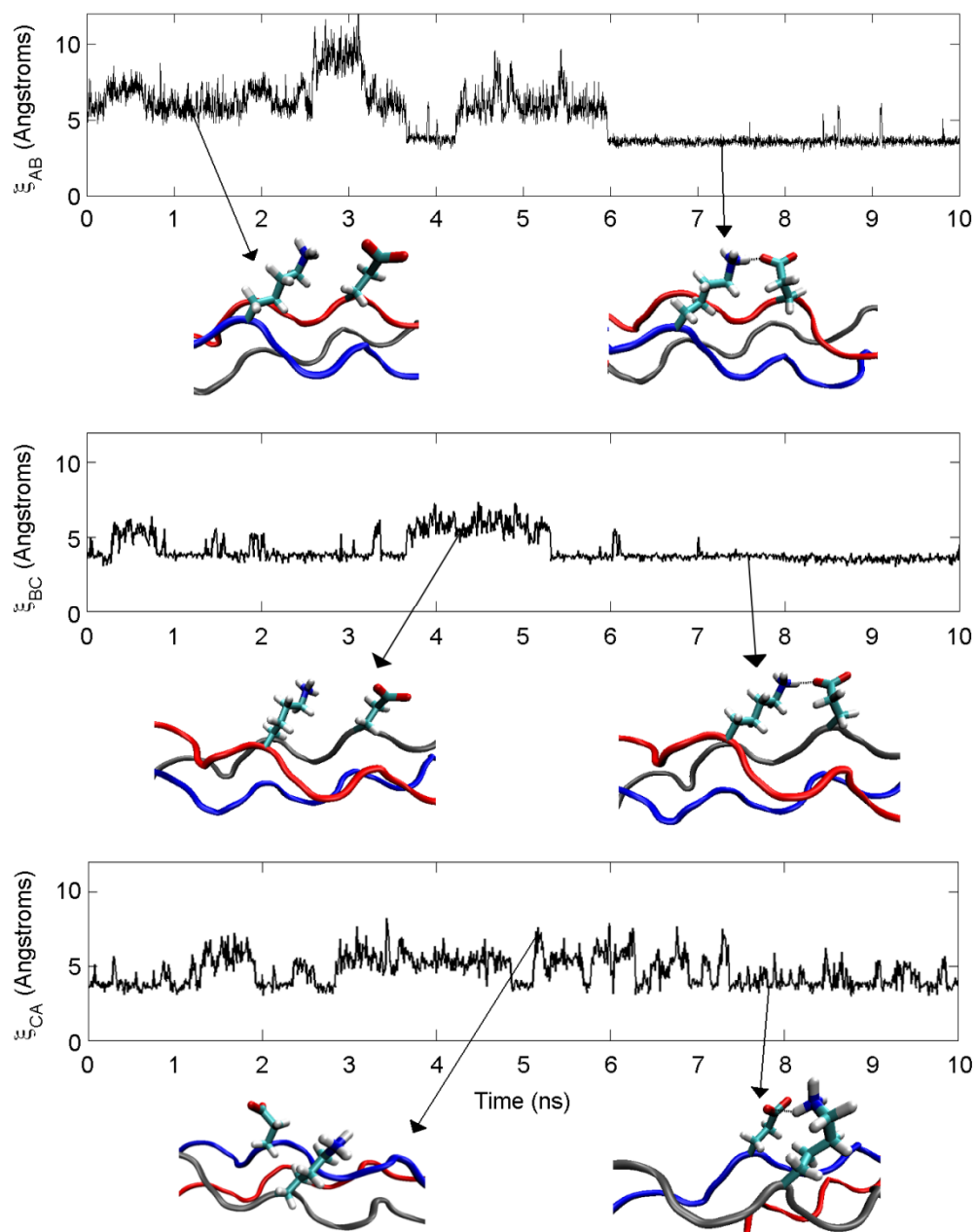
Figure 2: Dynamical timeseries data for the GPKGEO peptide. Snapshots from the various states are shown below the timeseries for each interchain interaction.
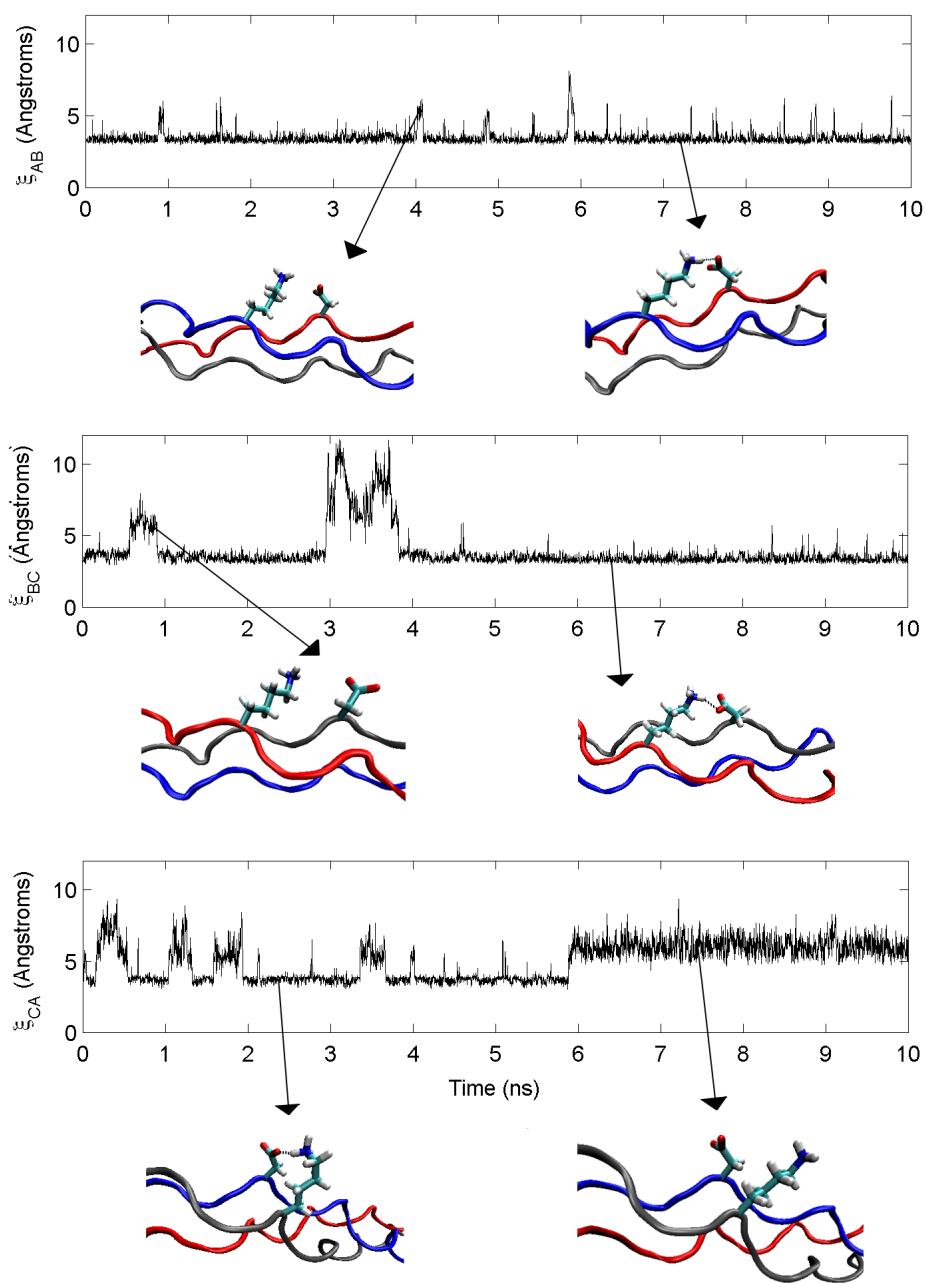
Figure 3: Dynamical timeseries data for the GPKGDO peptide. Snapshots from the various states are shown below the timeseries for each interchain interaction.
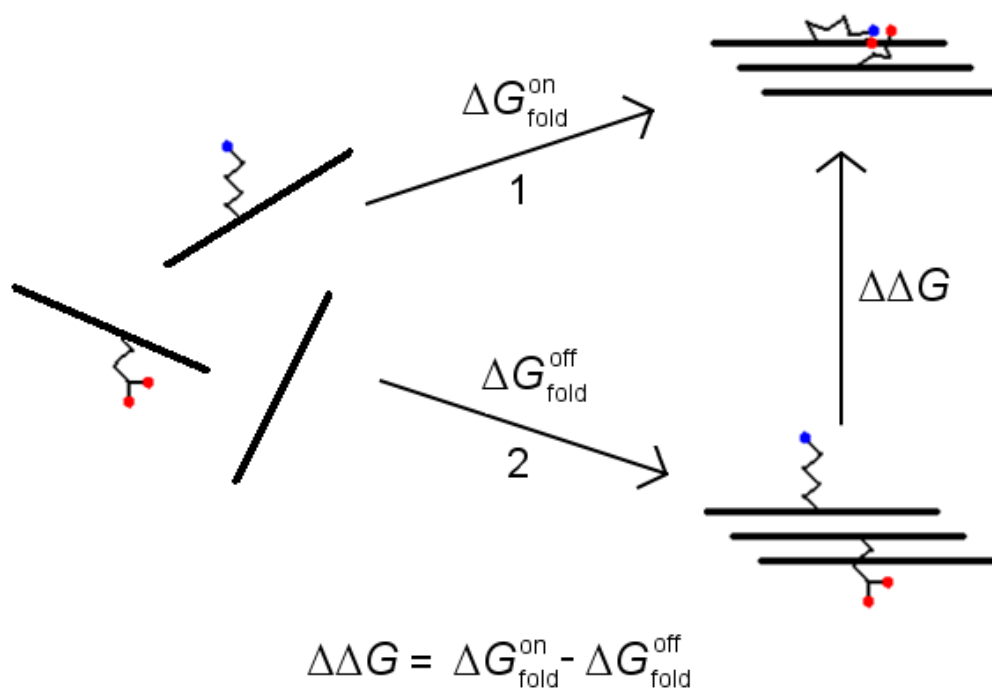
Figure 4: Thermodynamic cycle of collagen triple-helix formation involving the electrostatic interactions under study.
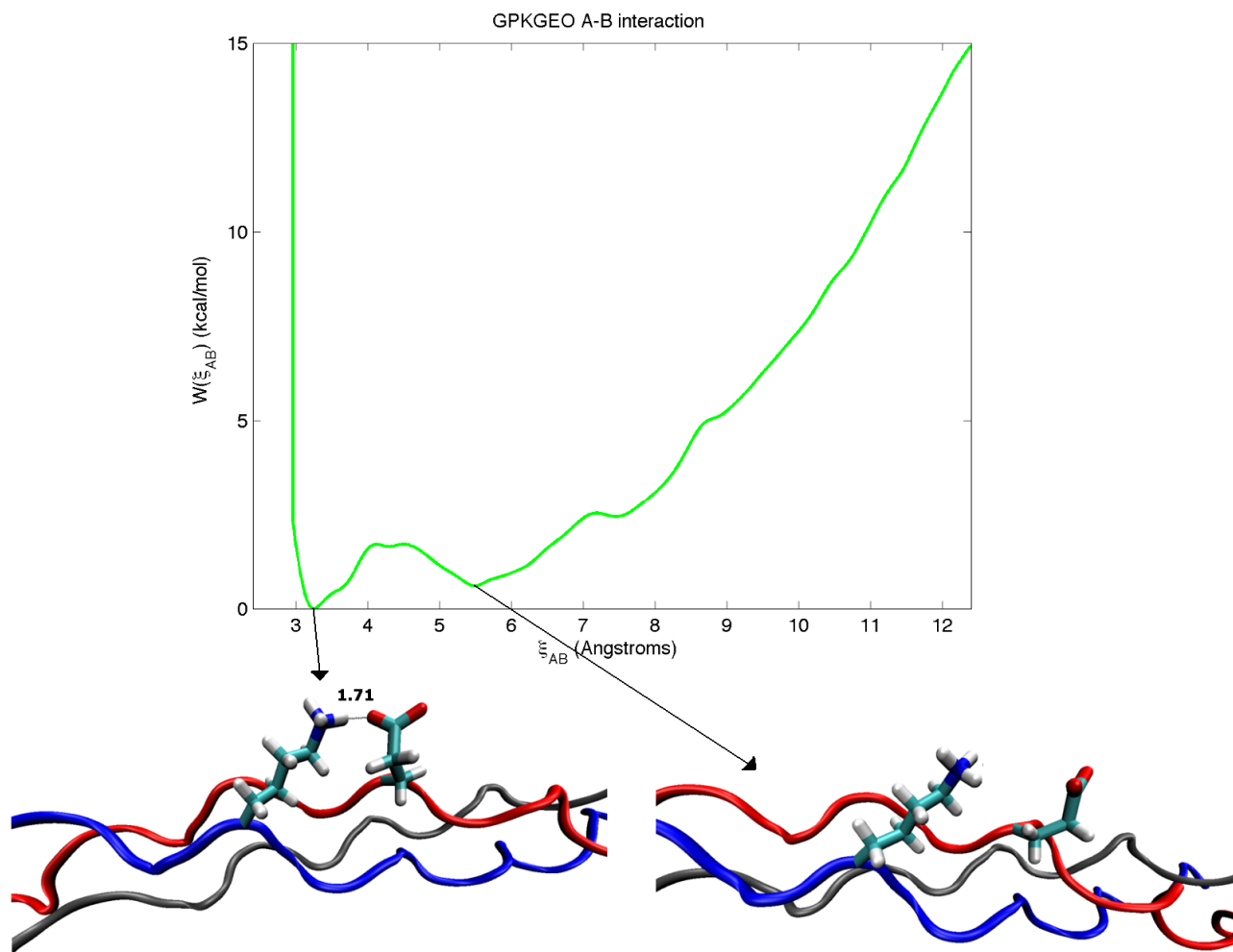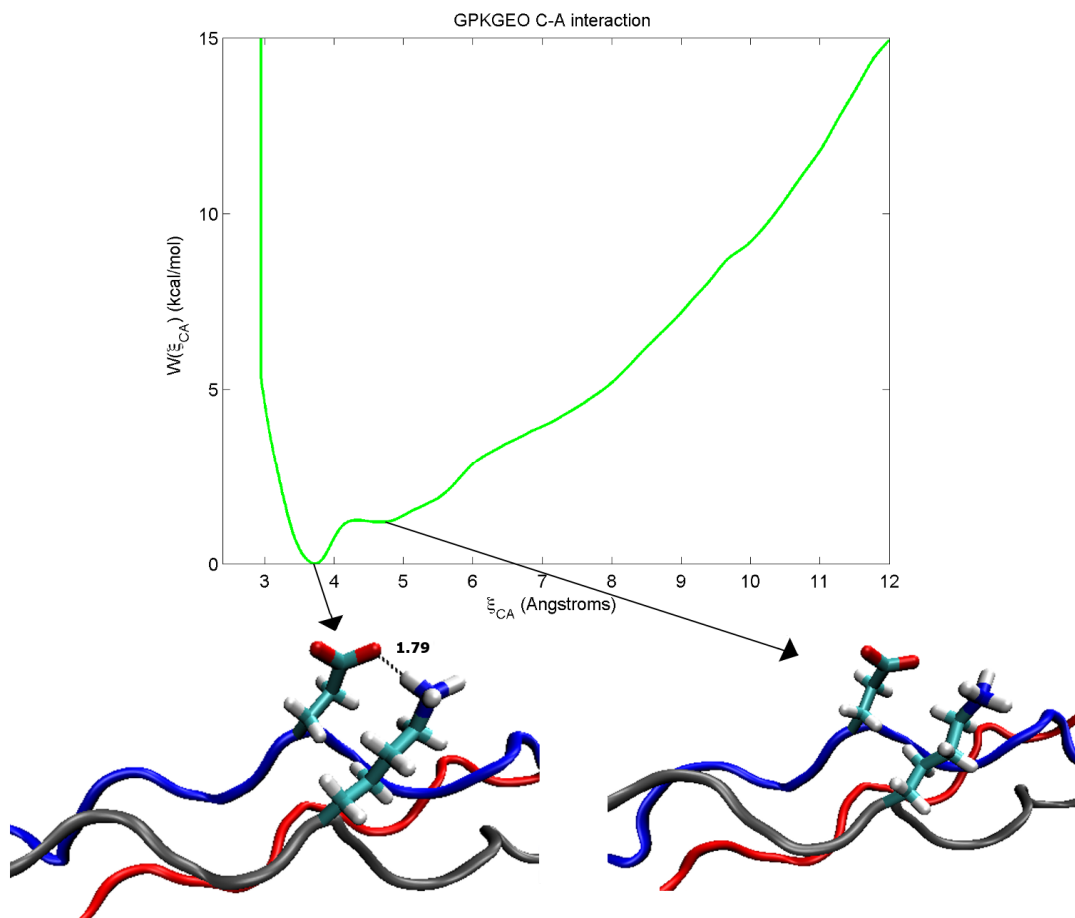
Figure 5: Potential of mean force constructed using the ξ_{AB} reaction coordinate for GPKGEO. Also shown are snapshots inside umbrella sampling simulation windows centered on $\xi_{AB}=3.4$Å and $\xi_{AB}=5.6$Å, respectively, corresponding to the two energy minima of the pmf.
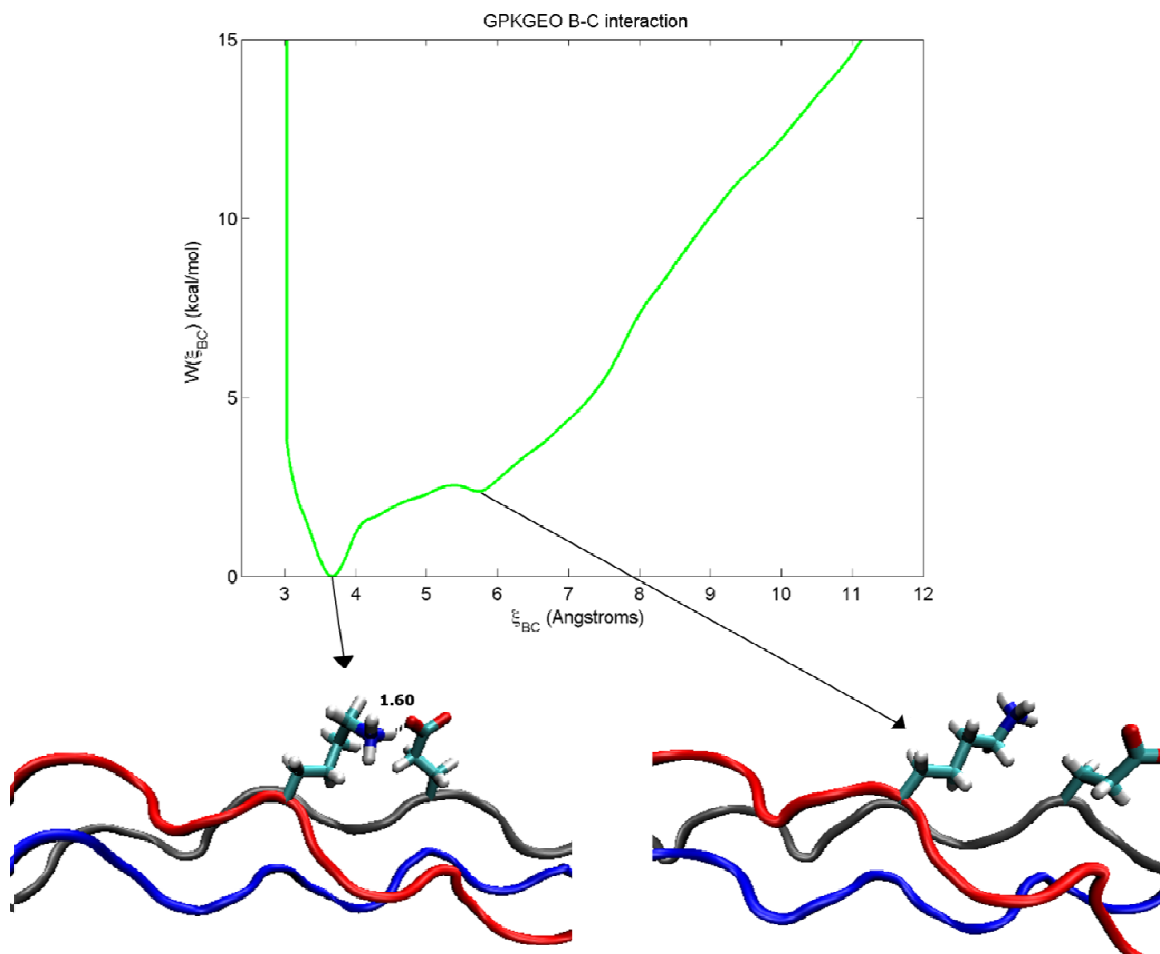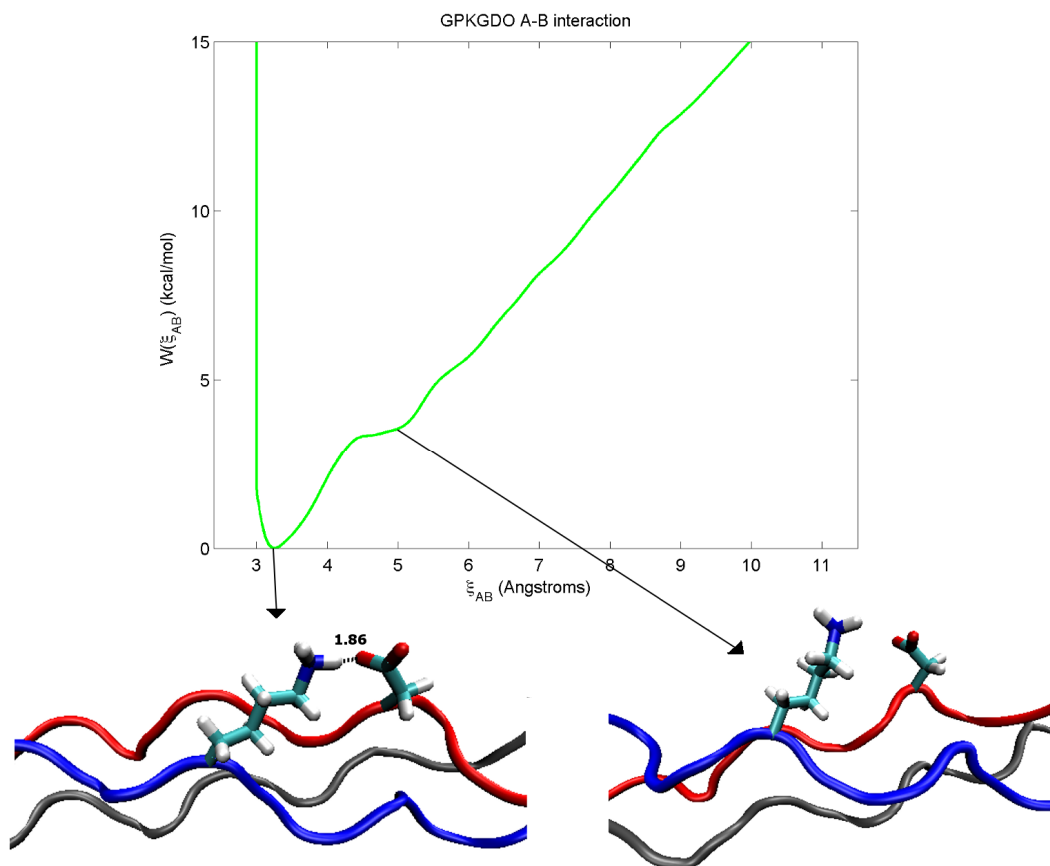
Figure 6: Potential of mean force constructed using the ξ_{BC} reaction coordinate for GPKGEO. Also shown are snapshots inside umbrella sampling simulation windows centered on $\xi_{BC}=3.8$Å and $\xi_{BC}=5.8$Å, respectively, corresponding to the two energy minima of the pmf.

Figure 7 - Potential of mean force constructed using the ξ_{CA} reaction coordinate for GPKGEO. Also shown are snapshots inside umbrella sampling simulation windows centered on ξ_{CA}=3.8Å and ξ_{CA}=4.8Å, respectively, corresponding to the two energy minima of the pmf.
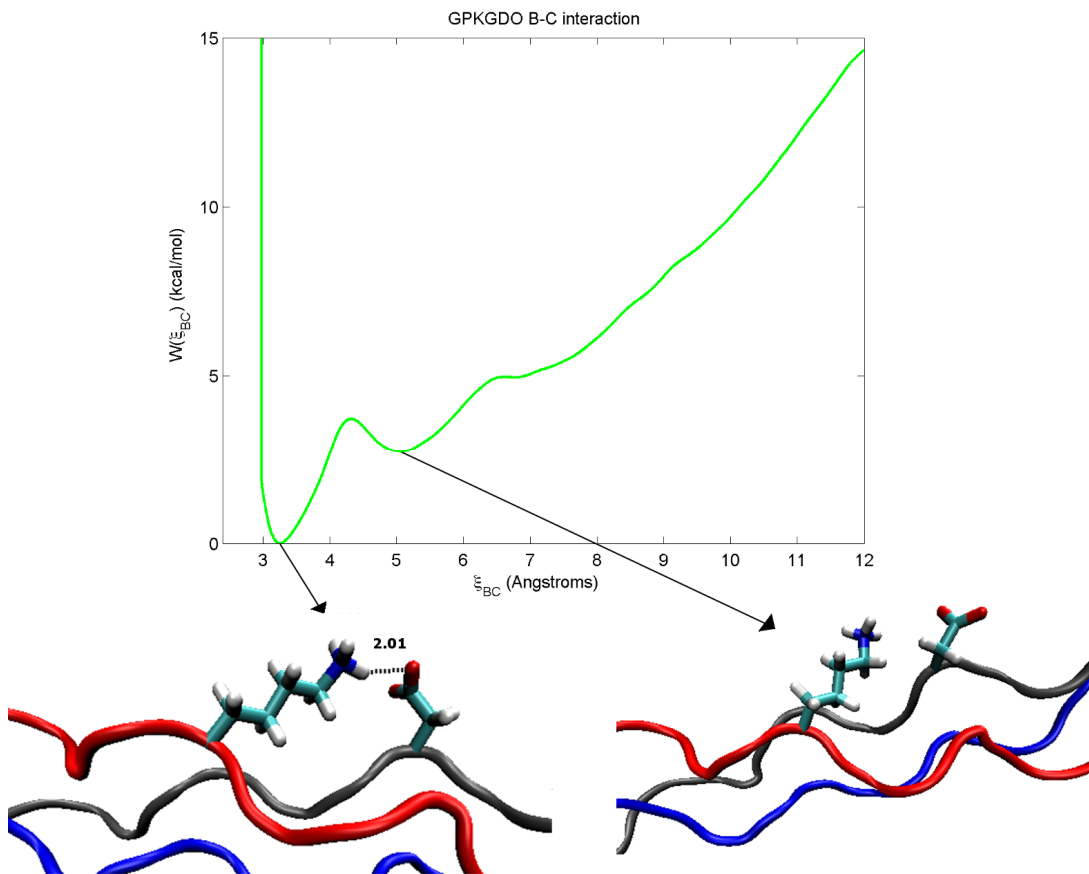
Figure 8: Potential of mean force constructed using the ξ_{AB} reaction coordinate for GPKGDO. Also shown are snapshots inside umbrella sampling simulation windows centered on ξ_{AB}=3.4Å and ξ_{AB}=5.0Å, respectively, corresponding to the two energy minima of the pmf.

Figure 9: Potential of mean force constructed using the ξ_{BC} reaction coordinate for GPKGDO. Also shown are snapshots inside umbrella sampling simulation windows centered on ξ_{BC}=3.4Å and ξ_{BC} =5.0Å, respectively, corresponding to the two energy minima of the pmf.
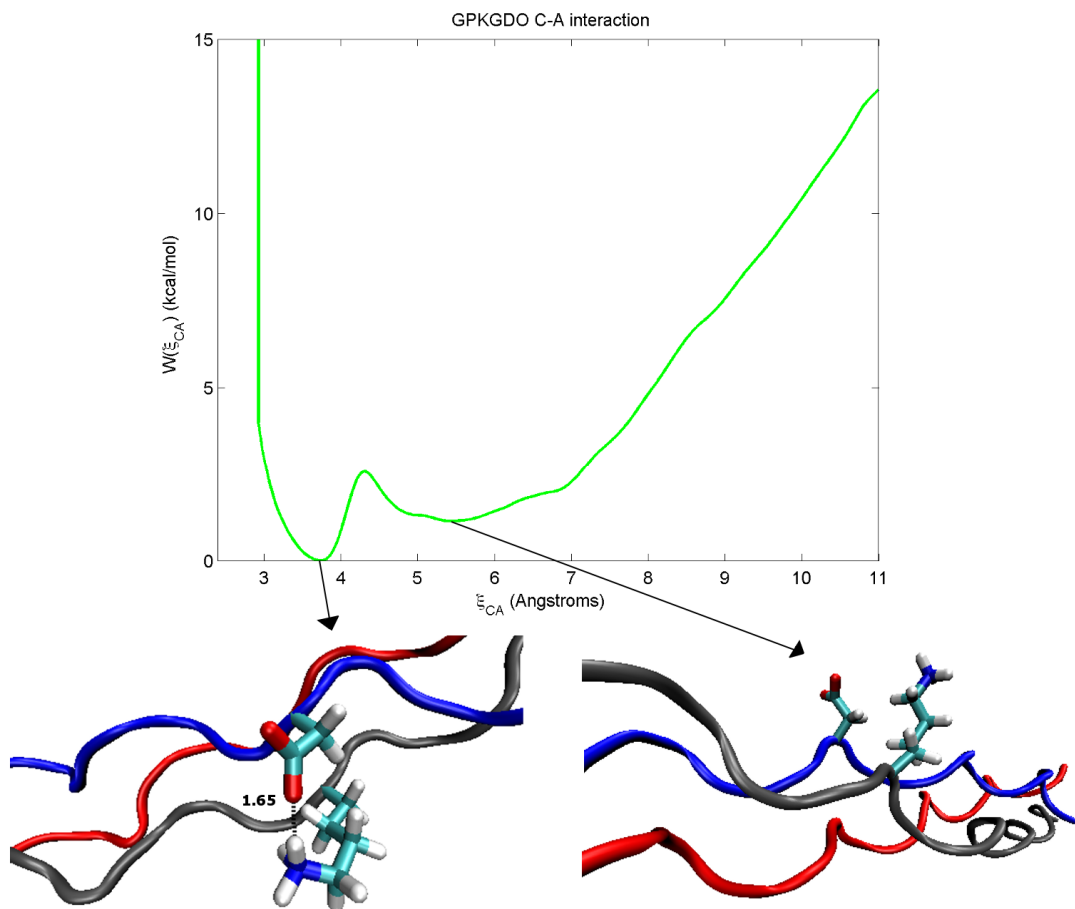
Figure 10: Potential of mean force constructed using the ξ_{CA} reaction coordinate for GPKGDO. Also shown are snapshots inside umbrella sampling simulation windows centered on ξ_{CA}=3.8Å and ξ_{CA}=5.4Å, respectively, corresponding to the two energy minima of the pmf.

Appendix B

Derivation of the effective equilibrium constant in the presence of full length enzyme

We first assume that the full length enzyme in Figure 6-1 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:

$$\frac{d[N]}{dt} = \frac{d[V]}{dt} = \frac{d[N \cdot H]}{dt} = \frac{d[V \cdot H]}{dt} = \frac{d[V \cdot F]}{dt} = \frac{d[E]}{dt} = 0 \tag{1}$$

where $[E]$ is the free enzyme concentration and the remaining variables are the species enumerated in Figure 6-1. We wish to then calculate the equilibrium ratio of all vulnerable states to all native states:

$$K_{eq}^{eff} = \frac{[V] + [V \cdot H] + [V \cdot F]}{[N] + [N \cdot H]} \tag{2}$$

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

$$K_{eq} = \frac{[V] + [V \cdot H]}{[N] + [N \cdot H]} \tag{3}$$

Proof:

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

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

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

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

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

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

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

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

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

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

$$= K_{eq}$$

Using equation (3) and the fact that $K_{bind}^{H} = \dfrac{[N \cdot H]}{[N][E]} = \dfrac{[V \cdot H]}{[V][E]}$ and $K_{bind}^{V_H C} = \dfrac{[V \cdot F]}{[V \cdot H]}$, we can calculate K_{eq}^{eff} from equation (2) as follows:

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

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

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

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

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

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

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

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

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

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

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

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

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

Therefore, we have:

$$K_{eq}^{eff} = K_{eq} \left(1 + \left(\frac{K_{bind}^{H} [E]}{1 + K_{bind}^{H} [E]} \right) K_{bind}^{V_H C} \right) \tag{4}$$