Title
Visualizing enzyme-inhibitor interactions

Author
Wang, Stephanie Xuefang

Publication Date
2001

Peer reviewed|Thesis/dissertation

VISUALIZING ENZYME-INHIBITOR INTERACTIONS: AFFINITY AND SPECIFICITY

by

Stephanie Xuefang Wang

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

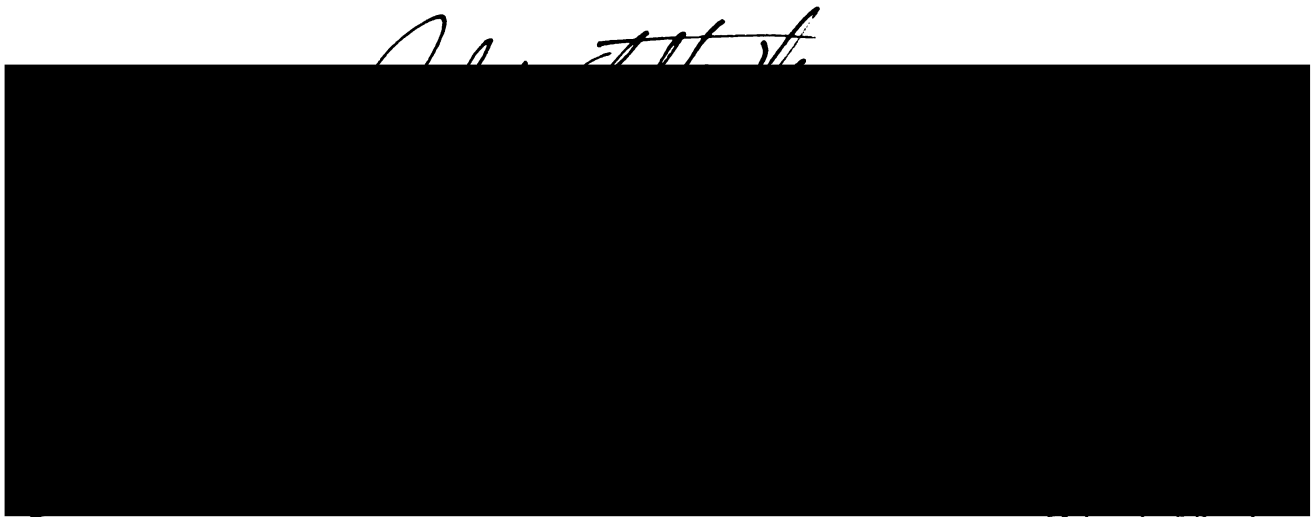
Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO



Date University Librarian

Degree Conferred: .....

This thesis is dedicated to my parents Xiao-Lan Zhu and Yuan-Fa Wang,

my sister Sophie Wang,

and my brother-in-law Hongtao Hou.

PREFACE

The moment has finally come for me to summarize my graduate school experience and to thank the people who helped me to shape my life in the past years. First and foremost, I would like to thank my research advisor Dr. Robert Fletterick. A survey once said that almost all graduate students never regretted the experience of going graduate school, however, a majority of them regretted not changing research lab. I was one of the lucky few who actually summoned the courage to switch laboratory midway through graduate school. Robert welcomed me into his lab and offered me every research opportunity in the lab. I thank Robert for his scientific acumen, for his encouragement, for allowing me to grow independently in research, and for creating a productive, friendly and supportive lab atmosphere that I enjoyed every day at work. Robert is truly an insightful teacher and a caring mentor.

Dr. Charly Craik was my outside-the-lab advisor. In the beginning of my crystallographic studies, I worked closely with the Craik lab. Charly allowed me to be part of his lab and treated me as one of his own students. Charly is among the professors who make the efforts to show their interest in students. He was always easy to talk to and his office was always open to students.

Dr. Tom Scanlan was the last of my thesis committee members, but I knew him from the very beginning of graduate school when I used to track him down at the beginning of every quarter, to ask him to sign my study list. I particularly want to thank

Tom for reading my thesis efficiently and fedexing back the signed title page in time for submission.

The Fletterick lab is wonderful home for me in the last three years of my graduate school. I was fortunate to have Sarah Gillmor as my supervisor, who taught me the basics of crystallography and helped me start my projects. Elena Sablin, Peter Hwang, and Manish Butte answered many of my technical questions. Elena, Jennifer Turner and Chuck Sindelar took the time to patiently proofread my manuscripts. I particularly want to thank Jennifer, who shared my interest in caramel frappuccino and Earl Gray tea, for her constant encouragement during the preparation of this thesis. I also want to thank Sabine Borngraber, Chuck, Maia Vinogradova, Carolyn Sousa, Richard Wagner, Russ Huber, Beatrice Darimont, Jenny Buchbinder, Linda Brinen, and recently Mary Jane Budny, Jennifer Ekstrom, Ben Sandler, Eugene Hur and Eric Slivka, who made my time in the lab full of fun. And there is always Debra Singer, who works quietly and efficiently to make our lives easier. I cannot possibly end this paragraph without thanking everyone for the warm and happy memories of ski trips to Tahoe. I shall remember those on-the-slope ski lessons from Robert, Peter, Sandy, Elena and Sergey and practice my skills so that I can feel comfortable going down any slopes.

My collaborators in the Craik lab have been terrific. Toshi Takeuchi and Chris Eggers have made my projects a lot easier and smoother by teaching me or performing the protein purification, and even setting up some crystallization trials. Jennifer Harris, though not an official collaborator, was a supportive friend and very generous to share her protocols and reagents.

The 10th floor msg community has been a wonderful place to learn and do science, the extremely open, friendly and helpful atmosphere was most impressive to me. People from our lab are free to wander into the Stroud lab or the Agard lab to borrow reagents or to seek scientific advice, and *vice versa*. I shall always remember the long hours in the graphic room or synchrotrons, sometimes well into the night, in the friendly company of Manish, Elena, Jennifer T., Sandy, Andy Shau, Kinkead Reiling...

I also would to thank my former advisor Dr. Dan Santi, who provided me the opportunity to work on a very challenging project in a fascinating system. The members of the Santi lab, Rachel Duncan, Lisa Huang, Lu Liu, Raghu Mirmira, Jung Ku, Ming Yu and Maria Diaz, have helped me grow into an experienced graduate student. Drs. C. C. Wang and Sue Miller, who welcomed me in their offices for scientific discussions or random chats, gave me valuable advice from which I have already benefited and will continue to benefit.

UCSF is an exciting place to learn and hear about high quality science, it is also a great place to meet interesting and intelligent people. Friends like Mark Kaplan, Keith Burdick, Herschel Wade, Karen Runk, Bridget Hanser shared fun time with me at parties, concerts, and other events. I also want to thank Jade Wang, Qi Wang, Walter Lau, Kuanhong Wang, and Jen Liou (the little *liu-wang* student club), for their support and friendship, and for making my life at UCSF enjoyable. I particularly want to thank Jade, who is my best friend all through graduate school and my roommate for the past three years, for being part of so many of my fond memories of San Francisco.

Finally, I want to gratefully thank my parents for their love and unwavering support for my education. They taught me my first lesson in science and have always encouraged me to pursue high learning. My sister has always been a role model for me. Her love, patience and tolerance have been most important to me. I want to express my deepest gratitude to my sister and her husband. Their support and encouragement have helped making everything possible.

Visualizing Enzyme-Inhibitor Interactions: Affinity and Specificity

Stephanie Xuefang Wang

Enzyme-inhibitor interactions help to reveal the mechanisms of many biological reactions and provide insights for designing therapeutic reagents. The activities of protease are regulated by natural or synthetic inhibitors, most of which bind to the active site of target proteases in a substrate-like fashion. The *E. coli* originated ecotin inhibits most chymotrypsin-family serine proteases by forming an ecotin$_2$protease$_2$ complex, regardless of the specificity of its targets.

To understand how the ecotin dimer interface contributes to its pan-specificity and the network like protein-protein interaction, I solved the structure of a monomeric variant of ecotin. The structure showed that the dimerization of ecotin was easily prevented through minor protein engineering, which indicates that the ecotin dimer may be evolved from monomers through a "domain swapping" type mechanism. The monomeric ecotin was used to simplify the complicated interactions in the ecotin-protease tetramer.

Using ecotin-trypsin as a model system, I studied the effects of mutations on protein-protein interactions. Three ecotin mutants, including an ecotin mutant with a C-terminal truncation, an ecotin with a primary site mutation and an ecotin double mutant with both the truncation and primary site mutations, were crystallized in the presence of

trypsin. The structures revealed that the C-terminal truncation resulted in minor conformational changes between the protein domains. The structures also indicated that the extensive interactions at the C terminal dimer interface might be responsible for the non-additivity of binding energies from the two protease-binding sites of ecotin.

I also determined the structure of bovine α-thrombin bound to a primary site mutant of ecotin, in order to reveal how thrombin evades the inhibition by wild type ecotin. One of the surface loops in thrombin moved drastically to permit the protease to accommodate the incoming ecotin, which indicates that the active site of thrombin may fluctuate between an open or a closed conformation. The extended interactions between thrombin and ecotin may serve as a template for predicting thrombin specificity and for providing insights for future inhibitor design.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIIGURES

Chapter 1

Introduction to the Thesis

1.1 Protein-Protein Interactions

Protein-protein interactions are found in almost all aspects of molecular and cell biology, from kinase-based signal transduction (Nishida & Gotoh, 1993; Ullrich & Schlessinger, 1990), receptor-based cell recognition and communication (Davis & Bjorkman, 1988; Gschwind et al., 2001; Lambie, 1996), antigen antibody-based immune response (Zamoyska, 1998), and protease-based digestion and regulation (Hollenberg, 1996). Protein-protein interactions differ from the interactions between proteins and small molecules in that they are more complicated, and often involve large and flexible surfaces (Antipova et al., 2001; Jones & Thornton, 1997). Through combined biological, chemical and structural studies, the principles and specificity of protein-protein interactions are being revealed despite of its complexity. For example, well-defined motifs such as the SH2 and SH3 domains have been identified (Pawson & Gish, 1992). The residues that contribute significantly to the binding affinity between two protein surfaces, *i.e.* the hot spots of interactions, have been defined by alanine-scanning technique (Bogan & Thorn, 1998; Clackson & Wells, 1995). In this thesis, I used the technique of x-ray crystallography to analyze the interactions between proteases and their macromolecular ligands.

Proteases are a class of enzymes, which by definition recognize physiological substrates that are only proteins. By cleaving specific peptide bonds in the substrates, proteases either activate them for their regulatory functions or degrade them for scavenging purpose. Protease-activated biological processes include digestion, blood

coagulation (Esmon, 2000), complement activation (Taylor et al., 1999), and programmed cell death (*i.e.* apoptosis) (Solomon et al., 1999). The regulation of proteases are achieved either by physiological inhibitors such as serpins (serine protease inhibitors) (Bode & Huber, 1992; Engh et al., 1995), or by small natural inhibitors such as hirudin (Rydel et al., 1991), bovine pancreatic trypsin inhibitor (BPTI) (Groeger et al., 1994; Ruhlmann et al., 1973) and their synthetic analogs (Lombardi et al., 1999).

Most inhibitors inhibit proteases by binding to the active site in a substrate-like fashion (Laskowski & Qasim, 2000). Therefore, studying the protease-inhibitor interactions is critical for understanding the mechanism of these enzymes and for designing potent therapeutic agents to regulate their activities. The substrate-like inhibitors often bind to the active site of target proteases via a long and extended surface loop called the reactive site loop (Figure 1-1). The residue that binds to the primary substrate pocket of the protease is called the reactive site residue or the P1 residue (Figure 1-1). It often provides the majority of the interaction energy, and thus is critical in determining the inhibition property of a protease inhibitor (Bode & Huber, 1992). Under most circumstances, the reactive site residue of a protease inhibitor matches the substrate specificity of the particular target protease it inhibits (Bode & Huber, 1992; Laskowski & Qasim, 2000). Details of the interactions between proteases and a macromolecular inhibitor are illustrated in Figure 1-1.

Figure 1-1. Interactions between a protease and a macromolecular substrate

The general interactions between a protease and its substrate or inhibitor is depicted. The amino acids in the reactive site loop are drawn out from N-terminus to C-terminal, and are labeled NH3...-P3-P2-P1-P1'-P2'-P3'...-COO- (Schechter & Berger, 1967). P1-P1' is the cleaved peptide bond, marked by a red arrow. Correspondingly, the pockets on enzyme are ...S3, S2, S1, S1, S2', S3'... so that P1 residue in substrate/inhibitor binds the S1 pocket of the protease, etc. P1 is also called the reactive site residue and S1 is the primary substrate-binding pocket.

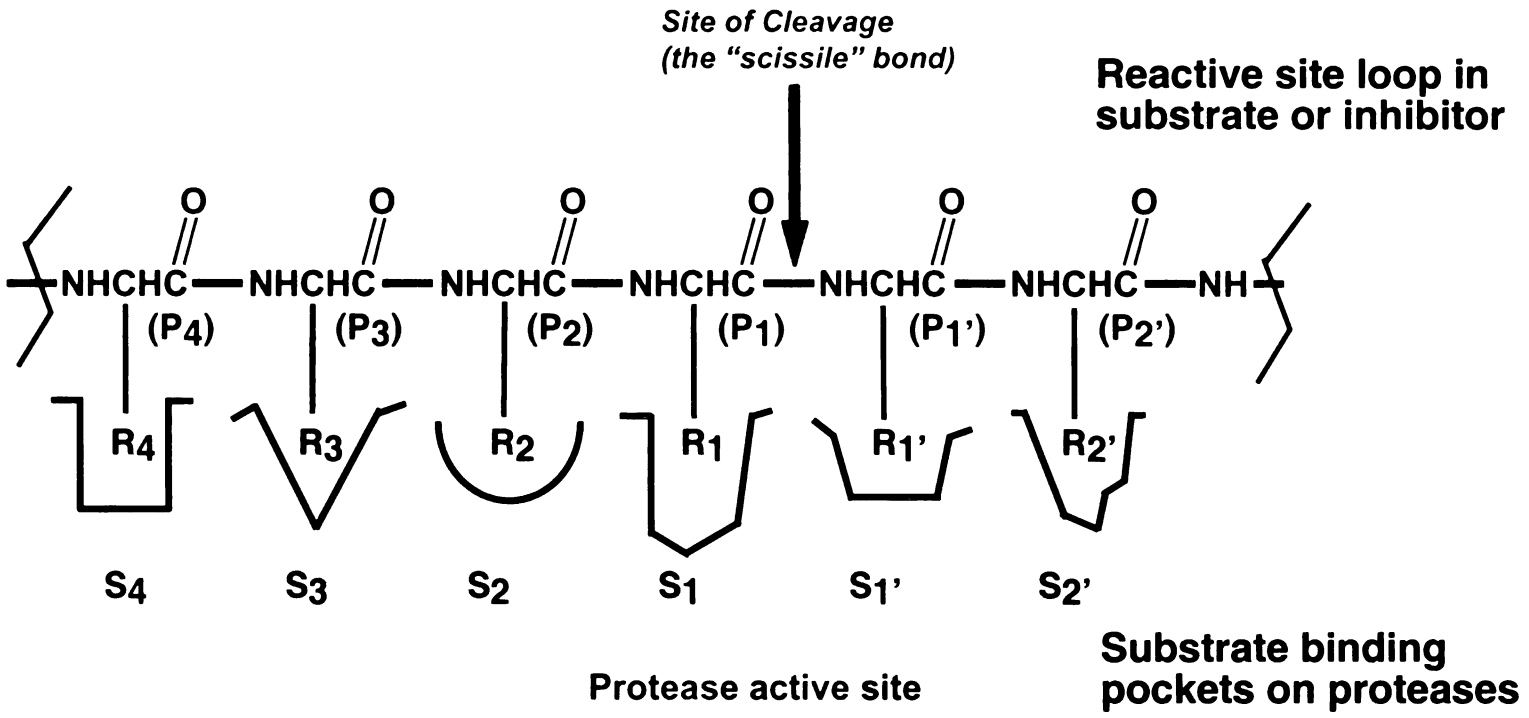


Figure 1-1. Interactions between a protease & another macromolecule

As depicted in Figure 1-1, the shapes of the side chains of the residues in the reactive-site loop from a substrate or inhibitor are often complementary to the substrate binding pockets on the proteases. Protease inhibitors, especially those not hydrolyzed by proteases during inhibition, can be used as probes for analyzing the protease-ligand interactions (Bode & Huber, 1992). The consecutive peptide sequence of a reactive site loop can be randomized by methods of phage display or chemical synthesis for optimized binding or enzyme activity (Harris et al., 1998; Wang et al., 1996). On the other hand, x-ray structures of protease-inhibitor complex allow researchers to visualize the shapes of the substrate binding pockets as well as the conformations of each residues in the reactive site loop, which gives insights into rational drug design (Morenweiser et al., 1997).

1.2 Ecotin and Protease Interactions

1.2.1 Ecotin

Ecotin, a small protein of 142 amino acids, was originally discovered from the periplasmic space of *Escherichia coli* for its ability to inhibit proteases such as trypsin (Chung et al., 1983). In fact, the name ecotin came from *Escherichia coli* trypsin inhibitor. Further analyses, however, revealed that ecotin could inhibit many other proteases in the chymotrypsin-fold serine protease family, which include chymotrypsin, elastase, blood coagulatory factor Xa, and factor XIIa and plasma kallikrein (Chung et al., 1983; Pal et al., 1994; Seymour et al., 1994; Ulmer et al., 1995). The reactive site of ecotin was determined to be a methionine at residue 84 (McGrath et al., 1994), which does not match the substrate specificity of most of its target proteases. In the ecotin-

trypsin structure, the side-chain of this methionine mimics the conformation of an optimized P1 residue such as the Lys residue in the trypsin-BPTI complex (McGrath et al., 1994; Ruhlmann et al., 1973). The side chain of a methionine is capable of forming hydrogen bonds (Groeger et al., 1994) as well as van del waals contacts, which suggest that it may serve as a compromised P1 residue to adapt to different substrate binding pockets.

Biochemical and crystallographic studies further reveals that ecotin likely exists as a dimer *in vivo*, regardless of the presence or absence of target proteases (Lee et al., 1999; McGrath et al., 1994; Perona et al., 1997; Shin et al., 1996). *Escherichia coli*, the natural host of ecotin, lives in the digestive tracks of host animals where various proteases reside in abundant quantity. Dimerization may be a mechanism the bacterium adopts to protect itself against its harsh living environment (McGrath et al., 1995). The core structure of ecotin is a seven-stranded β-barrel. Dimer formation allows the β -barrel from each ecotin molecule to pack tightly together, which possibly protects ecotin from being digested by surrounding proteases. The surface loops connecting the parallel β-strands in the ecotin core structure are exposed in solution (Shin et al., 1996). The secondary conformation of ecotin is quite similar to what has been observed for antibody structure (Chothia et al., 1989).

The unusual pan specificity sets ecotin apart from other protease inhibitors; its unique dimer conformation also sets ecotin apart from other protease inhibitors. For more than fifteen years, no homolog of ecotin was found from any other species.

Recently, Sarah Gillmor identified two ecotin-like genes in *Pseudomonas aeruginosa* and *Yersinia pestis* (Gillmor et al., 2000) by applying blast search in newly completed bacteria genomes. Sequence alignment showed the sequences of the binding loops in ecotin are also preserved in the two ecotin-like genes (Figure 1-2). One of the newly identified sequences, psatin (*i.e.* *Ps*eudomonas *a*eruginosa trypsin inhibitor), is being studied in the laboratory of Dr. Charles Craik at UCSF. I also found two more ecotin-like sequences from *Anopheles gambiae* and *Lycopersicon esculentum* (Figure 1-2).

Figure 1-2. Sequence alignment of ecotin and its homologs

The sequence alignment is modified from a figure from Sarah Gillmor's paper (Gillmor et al., 2000) where two ecotin-like sequences are found from *Pseudomonas aeruginosa* and *Yersinia pestis*. Two additional sequences are identified from *Anopheles gambiae* and *Lycopersicon esculentum*. The sequences corresponding to the binding loops from *E. coli* (ecotin), at both the primary binding site (50s, 80s) and the secondary binding site (60s, 110s), are highlighted by short orange bars.

Figure 1-2. Sequence alignment of ecotin & its homologs

Figure 1-3. Ecotin-protease interactions

The tetrameric organization of wild type ecotin-rat trypsin complex (McGrath et al., 1994). The primary, secondary binding site and the putative dimer interface are highlighted in the figure. Trypsin molecules are colored in blue. The two ecotin molecules are colored in dark red and gold, respectively.

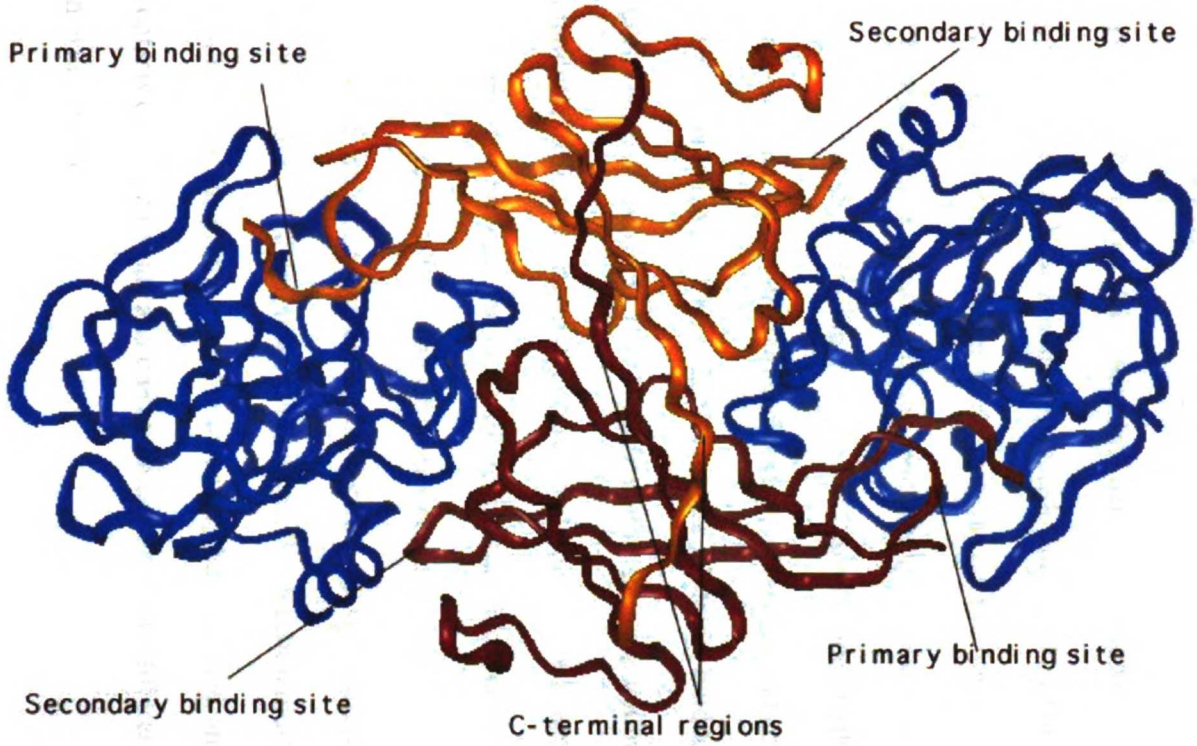


Figure 1-3. X-ray structure of wild type ecotin complexed with rat ainoic trypsin

1.2.2 Ecotin-Protease Interactions

1.2.2.1 The Ecotin-Protease Complex

Mary McGrath from our laboratory first determined the x-ray structure of wild type ecotin, in complex with rat anionic trypsin (McGrath et al., 1994). Ecotin-trypsin complex revealed a protease inhibition mode that was not previously discovered. Ecotin and trypsin form a complex of ecotin$_2$trysin$_2$, containing two ecotin and two trypsin molecules (McGrath et al., 1994)(Figure 1-3). This novel mode of inhibition was subsequently confirmed by the structures of ecotin-collagenase and ecotin-chymotrypsin (Lee et al., 1999; Perona et al., 1997). Most protease inhibitors bind to target protease in a 1:1 complex (Bode & Huber, 1992). Though another protease inhibitor *Streptomyces* subtilisin inhibitor (SSI) was shown to dimerize and bind two proteases, every SSI only interacts with only one protease through a single substrate-like contact at the protease active site (Takeuchi et al., 1991). The tetrameric ecotin-protease complex exists as an intricate network of complicated interactions (Figure 1-3), very different from the "beads on a string" model of the SSI-subtilisin interactions. Each ecotin molecule has two protease binding sites: the primary and the secondary binding sites, and every protease domain contacts one primary site and one secondary site from different ecotin molecules (McGrath et al., 1994).

The primary binding site consists of the 50's and the 80's loops of ecotin (Figure 1-3). The 80's loop is the reactive site loop, containing the P1 residue Methionine 84. It binds to trypsin active site in an extended conformation, similar to those defined for the

conical protease inhibitors (Ruhlmann et al., 1973). The P1 residue Met84 contributes to the pan-specificity of ecotin (McGrath et al., 1994). The 50's loop of ecotin is much shorter than the 80's loop. A disulfide bond connects residue 50 to residue 87, indicating that the function for the 50's loop is to stabilize and to support the 80's loop. The buried surface area at the primary binding site is about 1900 $Å^2$ (McGrath et al., 1994). The secondary binding site includes the 60's and the 110's loops (Figure 1-3); it contacts a protease surface that is distant from the active site. The residues contributing to protease binding are residues 67 to 70 from the 60's loop and residues 108 to 114 from the 110's loop (Gillmor et al., 2000; McGrath et al., 1994). Sequence analysis of ecotin and its analogs revealed the residues in the secondary binding loops might be less conserved than the reactive site binding 80's loop (Gillmor et al., 2000) (Figure 1-2). The same clusters of residues varied more than primary site residues in phage display analysis (Yang & Craik, 1998). The buried surface area at the secondary site is 950$Å^2$, only half of the size of the primary site (McGrath et al., 1994). The secondary site, though much smaller in size compared to the primary site, may be a site for engineering selectivity into ecotin for more specific inhibition against selected protease targets (Wang et al., 1996; Yang & Craik, 1998).

The primary binding site from one ecotin and the secondary binding site from the other ecotin molecule contact the same protease (Figure1-3). The special configuration is made possible by the presence of the dimer interface between the two ecotin molecules in the ecotin-protease tetramer (Figure 1-3). The ecotin dimer is formed by the C-termini (*i.e.* the arm regions, containing residues 125 to 142) from two ecotin molecules

(McGrath et al., 1994). During dimerization, the two arm regions from ecotin molecules extend out from the β-barrel core structure to grab its dimer mate, which may be a classic example of the "domain swapping" process (Bennett et al., 1994). The buried surface area at the dimer interface is around 3000Å^2 (McGrath et al., 1994), the gigantic size also suggests its significance.

1.2.2.2 Previous Biochemical and Structural Analyses of Ecotin-Protease Complex

The two protease-binding sites of ecotin, the primary and secondary sites, have been analyzed closely by mutational, biochemical and structural studies. It has been demonstrated that mutations at the primary site affected ecotin binding to protease targets (Gillmor et al., 2000; Yang & Craik, 1998; Yang et al., 1998). For example, P1 residue Met of ecotin can be mutated to protease specific residues such as Lys or Arg to improve its binding affinity for trypsin, factor XIa or activated protein C (Pal et al., 1994; Seymour et al., 1994; Ulmer et al., 1995). The unique secondary site indeed contributes to protease binding as mutating the protease contacting residues in the secondary site 60's loop to alanines (*i.e.* 67-70A$_4$ ecotin) dramatically decreased ecotin binding to rat trypsin by over 4500 folds (Yang et al., 1998). Furthermore, randomizing the protease contacting residues from the secondary site gave rise to consensus sequences with improved affinity, indicating that protease binding at the secondary site can also be manipulated (Yang & Craik, 1998).

Protease interactions at the two binding sites are not independent, even though they are separated by over 25 Å. For example, the 67-70A_4 ecotin mutant can be rescued by a simple optimization mutation at the P1 site (Met to Arg) (Yang et al., 1998). Structural analysis of the 67-70A_4/M84R double revealed that ecotin completely lost contact to the protease surface at the secondary site and rotated 14 degree towards the primary site (Gillmor et al., 2000). This observation suggests that binding at the primary site may be compromised in the original wild type ecotin-trypsin complex. When multiple alanine mutations severely weakened the binding interactions at the secondary site, protease is free to fully optimize its binding at the primary site to compensate the loss of binding at the secondary site. Indeed, the 67-70A_4/M84R double mutant binds to trypsin even slightly tighter than the M84R mutant (Yang et al., 1998).

1.3 Efforts to Analyze Details of Ecotin-Protease Interactions

1.3.1 A Monomeric Form of Ecotin

Ecotin dimer formation is critical for the formation of the final ecotin-protease complex. The dimer interface make it possible for two protease binding sites from different ecotin molecules to contact one target protease at the same time. Previous biochemical and structural studies (Gillmor et al., 2000; Yang & Craik, 1998; Yang et al., 1998) have revealed that the two binding sites are not independent: binding interactions at one site can be affected by the interactions at the other site and by the strength of interactions of the dimer interface. In order to simplify our analysis and to fully dissect the interactions in the complicated ecotin-protease complex, Christopher T. Eggers from

Dr. Charles Craik's laboratory at UCSF made an ecotin variant that exists as a monomer (mEcotin). I joined his efforts in characterizing the mEcotin protein (Eggers et al., 2001). The results of structural and some biochemical analyses of the mEcotin and its variants can be found in Chapter 2. Detailed protocol of some biochemical studies can be found in Appendix One.

1.3.2 The C-Terminal Truncation on Ecotin-Trypsin Interactions

The C-terminal arm region in ecotin, consisting of residues from 125 to 142 (McGrath et al., 1994), is believed to be critical for the formation of ecotin dimer as well as the subsequently formation of the ecotin-protease tetrameric complex. In Chapter 2, we tried to characterize the dimer interface by getting rid of it; in Chapter 3, we took another approach: decreasing the size of dimer interface by C-terminal truncation. Previously, deleting residues 133 to 142 rendered an ecotin mutant that exists as monomer by gel filtration (Pal et al., 1996). Similar truncation experiments were conducted by Yang and his colleagues from the laboratory of Dr. Charles Craik at UCSF (Yang et al., 1998). They also showed that deleting amino acids beyond residue 133 abolished ecotin expression, which indicated that these amino acids possibly actively participated in the proper folding of ecotin molecules. They also showed that the same truncation (*i.e.* deleting residues 133 to 142) weakened ecotin-trypsin interaction by over 200 fold (Yang et al., 1998). In both studies, however, it was established that the truncation of the last ten residues in ecotin does not disrupt the formation of a tetrameric complex in the presence of protease targets (Pal et al., 1996; Yang et al., 1998).

Using the interactions between ecotin and trypsin as a model system, I have tried to determine if and how much the C-terminal region of ecotin affects the tetrameric ecotin-protease complex. I have crystallized several mutants of ecotin in complex with rat anionic trypsin. These ecotin mutants include M84R (a single mutant with the P1 residue Methionine mutated to an Arginine), armless (a mutant containing a deletion of the C-terminal residue 133 to 142) and armless/M84R (a double mutant containing both the P1 Met to Arg mutation and the C-terminal truncation). The structures of all three ecotin-trypsin complexes are determined to high resolution (around or beyond 2 Å). Structural comparison analysis reveals that ecotin and trypsin in the tetrameric complex slightly rearrange their relative position upon the truncation at the C-terminal region. Such reorganization is slightly exaggerated in the presence of an optimized P1 residue. These new structures enhanced the present understanding on ecotin interactions with proteases. The tetrameric complex was quite rigid in the dimer interface: *i.e.* the hinge was not as flexible as what was suspected previously (McGrath et al., 1994). The rigidity cast by the C-terminal arm regions is responsible for the compromised binding at the primary and secondary binding site. Thus, truncating the C-terminus of ecotin seems to release the rigidity in the complex (Gillmor et al., 2000). These structures are presented in Chapter 3.

1.3.3 Interactions Between Ecotin M84R Mutant and Thrombin

Serine protease α-thrombin is one of the crucial enzymes regulating the blood coagulation process. It exhibits both pro- and anti coagulant activities, either through converting fibrinogen to clot-forming fibrin (Blomback et al., 1967; Hogg & Blomback,

1978) or through the thrombomodulin-mediated activation of protein C pathway (Esmon et al., 1982). The regulation of thrombin has been the focus of much research effort searching for clotting therapeutics. Thrombin is interesting to me also for another reason. As we know, ecotin inhibits most serine proteases of the chymotrypsin fold regardless of their substrate specificity (Chung et al., 1983). There are, however, in the chymotrypsin family, proteases that are not inhibited by ecotin. Thrombin is one such protease. Wild type ecotin failed to inhibit thrombin (Seymour et al., 1994; Ulmer et al., 1995). Modeling studies indicate that a surface loop in thrombin (the 60's loop) may be responsible for wild type ecotin's failure to inhibit thrombin. The 60's loop, which is located right above the active site of thrombin, clashes with residues 50 and 51 from ecotin. Surprisingly, a point mutation: M84R, which unlikely removes any of the structural barriers, turned ecotin into a moderate inhibitor (Seymour et al., 1994) against thrombin with an apparent inhibition constant of 1.5 nM, $2x10^4$ times better than wild type ecotin (data not shown). In order to elucidate how a single residue outside the steric-clashing area changes the interaction between ecotin and thrombin, I have determined the structure of bovine thrombin bound with the M84R ecotin mutant. The ecotin M84R-thrombin complex was isolated by gel filtration and subsequently crystallized. The final structure was refined to 2.5 Å. In chapter 4, the conformational changes that have enabled the formation of the ecotin M84R-thrombin complex are described and analyzed in details. The extended interactions between ecotin M84R mutant and the active site of thrombin and the possible insights such interactions provide on future thrombin inhibitor design are also discussed in chapter 4.

1.3.4 Ecotin and the Blood Coagulation Factor Xa

The blood clotting pathways are regulated by a series of serine proteases including factor IXa, factor Xa and factor VIIIa (figure 1-4, the blood coagulation pathways). The high similarity among these enzymes has made designing specific inhibitors a rather challenging task. In Chapter 4, I already described the interactions between ecotin M84R mutant and an important player of the blood coagulation pathway: thrombin. In Appendix Three, I included a brief summary of my study on the interaction between ecotin and the blood coagulation factor Xa. For the first time, factor Xa is co-crystallized with a macromolecular protease inhibitor. Also for the first time, Gla domain and both EGF domains from an intact light chain of factor Xa are visualized at the same time. It appears that all domain structures (*i.e.* Gla, EGF1 and EGF2) maintain their structural integrity. The two short regions connecting these structural motifs may act as flexible hinge-like switches, which may be regulated by the presence of Ca ion.

1.4 Summary

I have focused my graduate studies on the interaction between a general serine protease inhibitor ecotin and several serine proteases, including the simple digestive protease: trypsin, and the more complex regulatory proteases from the blood clotting pathways: thrombin and factor Xa. Ecotin is a unique protease inhibitor not only because of its pan specificity, but also because of its novel mode of inhibition. Large protein-protein interaction surfaces embedded in the ecotin-protease complex have helped our understanding of protease interactions. Previously, the structure of the ecotin-collagenase complex has allowed researchers to understand the interactions between collagenase and

its substrate: collagen (Perona et al., 1997). In my study, ecotin is crystallized with several different proteases with highly similar structures. I hope to define differences in the highly similar interaction surfaces between these proteases and ecotin variants. The ultimate goal is to be able to define more rigorous specificity into ecotin so it would recognize target protease more selectively.

1.5 A List of Ecotin Variants Discussed in Thesis

M84R: an ecotin mutant with its P1 residue Methionine 84 mutated to an Arginine;

67-70A_4: an ecotin mutant in which four protease-contacting residues in the secondary binding site, i.e. residue 67 though 70, are all mutated to alanines;

67-70A_4/M84R: an ecotin double mutant containing both the multiple alanine mutations in residues 67 to 70 and a point mutation at residue 84;

Y69F/D70P: an ecotin mutant selected from phage display study, a Tyr at residue 69 is mutated to a Phe and an Asp at residue 70 is replaced by a Pro;

Armless: an ecotin mutant in which the C-terminal residues 133 through 142 are deleted;

Armless/M84R: an ecotin double mutant containing both the c-terminal truncation and a point mutation of Met to Arg at residue 84;

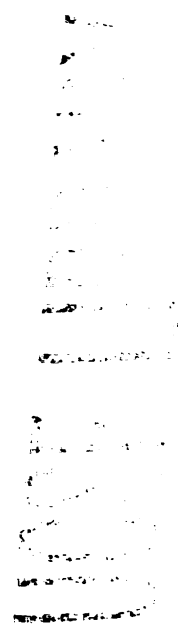
mEcotin: a monomeric form of ecotin which contains an inserted b-turn to fold back the extended C-terminus;

M84R mEcotin: a mEcotin variant that contains an M84R mutation;

67-70A_4/M84R mEcotin: a mEcotin variant that contains both an M84R mutation and an additional mutation of 67-70A4.

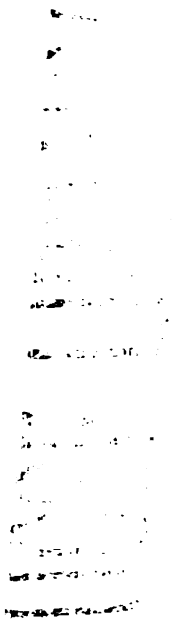
Chapter 2

Monomeric Ecotin



2.1 Introduction

In an ecotin-protease complex, there are five protein-protein interaction surfaces of three different types: two ecotin-protease interfaces between the primary binding loops and the active site of the protease; two ecotin-protease interfaces between the secondary binding loops and the C-terminus of the protease; and one ecotin-ecotin dimer interface at the C-terminal region of ecotin (McGrath et al., 1994). The first two types of interfaces are between ecotin and protease, while the last dimer interface is between two ecotin molecules. Although the dimer interface is the only interface that doesn't contact the protease target, it appears to be most essential. Without the dimer interface, it may be impossible for the intricate protein-protein interactions to occur (Figure 1-3). The presence of the dimer interface, however, also complicates the analysis of the interactions in the ecotin-protease complexes. Binding energy from the each protease-binding site can not be evaluated separately (Yang & Craik, 1998; Yang et al., 1998). The effects of the two protease-binding sites from ecotin will only be fully understood and explored if the dimer interface is disrupted.



Two ecotin molecules form an ecotin dimer by exchanging their C-terminal regions in a typical "domain-swap" type arrangement (Bennett et al., 1994). The dimer interface includes amino acid residues from 125 to 142 in the C-terminal region of ecotin (McGrath et al., 1994). The most obvious approach for making a monomeric form of ecotin may be to truncate the C-terminal arm region. In 1996, Pal *et al.* constructed an ecotin mutant without the last ten amino acids at its C-terminus. This mutant existed as a

monomer during gel filtration analysis, but it still formed an ecotin-protease tetramer in the presence of trypsin (Pal et al., 1996). In order to fully characterize the C-terminal arm region of ecotin, Stephen Q. Yang from the Craik laboratory at UCSF carried out sequential truncation experiments and subsequent kinetic analysis (Yang & Craik, 1998; Yang et al., 1998), accompanied by the structural analysis by Sarah A.Gillmor from our laboratory (Gillmor et al., 2000). Deletion beyond residue 132 failed to produce any ecotin protein, indicating that the residues in the C-terminal region may directly participate in the proper expression and folding of ecotin molecules (unpublished results). Truncating the last ten amino acids gave maximum deletion at the C-terminus while rendering normal ecotin expression, which agreed with the result from previous study. Armless, a mutant of ecotin containing a truncation from residue 133 to 142, was constructed (Yang & Craik, 1998). By fluorescence titration analysis, the dimerization constant (K_d) of the armless mutant was measured to be 540 nM, slightly higher than the K_d for the wild type ecotin at 220 nM. It was therefore concluded that the armless mutant could still dimerize, which disagreed with the result from the previous study (Pal et al., 1996). Both groups, however, concluded that an ecotin variant with ten amino acids deleted from its C-terminus could maintain the ability to form an ecotin-protease tetrameric complex. Truncation experiment does not appear to give rise to a monomeric form of ecotin that will bind to target protease in a 1:1 complex.

A second approach is taken to create a monomeric form of ecotin based on the domain-swap mechanism for dimerization. Sarah Gillmor from our laboratory formulated the original design. Instead of simple deletion, a sharp turn was inserted in

the arm region to retrieve the extended arm of ecotin from its dimer mate. The design of this ecotin variant and its characterization are discussed in this chapter. The significance of the arm region as evaluated by C-terminal truncation will be discussed in chapter 3.

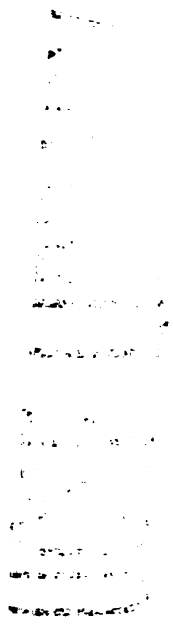
2.2 Material and Methods

2.2.1 Design of a Monomeric Form of Ecotin (mEcotin)

Based on the design by Sarah Gillmor, Christopher Eggers from Dr. Charles Craik's laboratory created an obligate ecotin monomer, denoted as mEcotin (Eggers et al., 2001). Detailed design for the mEcotin variant is elucidated in Figure 2-1. A three-residue insertion, Ala-Asp-Gly, is introduced after residue Trp130. The inserted residues and the Lys residue at 131 constitute the sequence of Ala-Asp-Gly-Lys, which was known to form a type I' β-turn in staph nuclease (Hynes & Fox, 1991; Loll & Lattman, 1989). Modeling studies showed that this type I' β-turn will allow the C-terminal region of an ecotin molecule to fold back on itself, instead of extending over to the other ecotin molecule to form the dimer. Modeling analysis also suggested that the retrieved arm should form similar interactions with the ecotin core structure to the original interactions formed by the C-terminus from the "borrowed" arm. This β-turn approach has been applied successfully to construct a monomeric variant of λ Cro, which also features an anti-parallel β-ribbon at the dimerization interface (Mossing & Sauer, 1990).

Figure 2-1 Design of mEcotin and scEcotin

Design of the monomeric mEcotin and a single chained scEcotin. The dimeric structure of the uncomplexed wild type ecotin (Shin et al., 1996) is depicted with only one subunit colored by secondary structure and the other subunit colored in gray. The dotted red lines represent the modeled mutations. The mEcotin mutant contains an Ala-Asp-Gly insertion after residue Trp130. The side chains are shown for the two Trp residues from each subunit, at position 67 and 130. The scEcotin mutant involves a Gly-Gly-Gly linker between the two subunits. Residues on one subunit is labeled prime and on the other labeled non-prime. The figure was prepared using Raster3D (Bacon & Anderson, 1988; Merritt & Bacon, 1997).



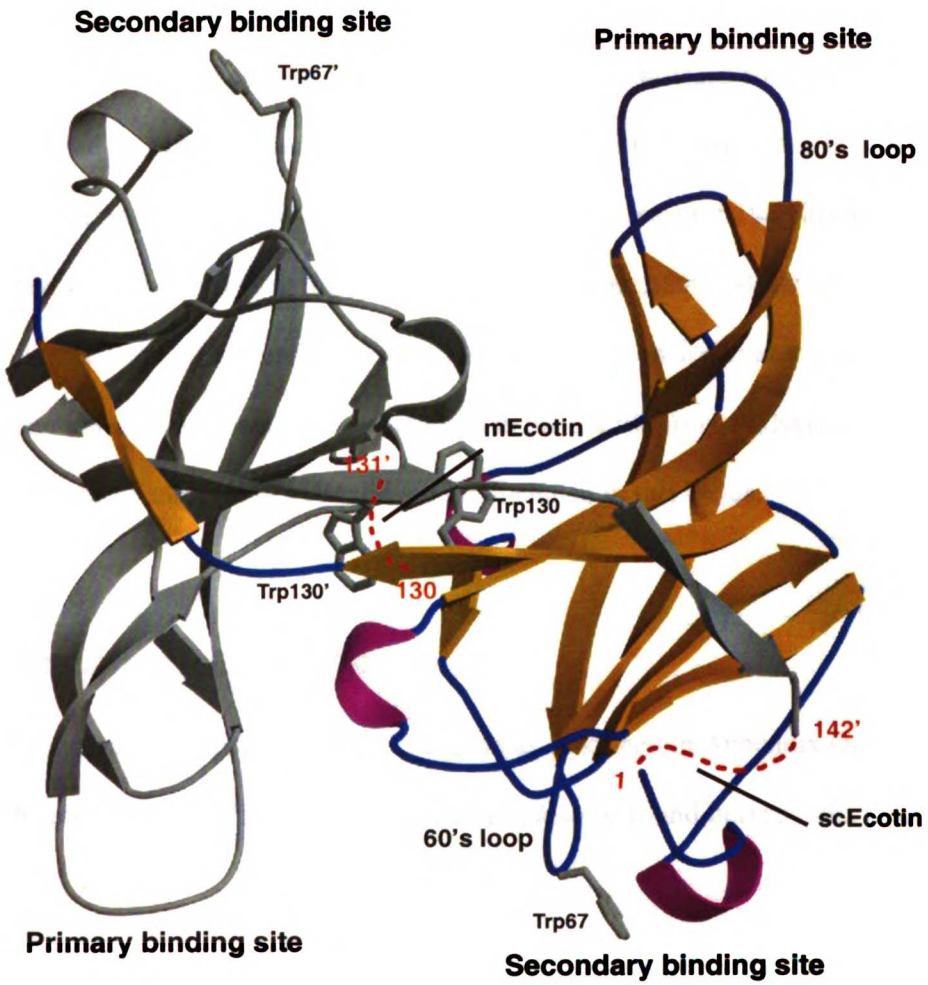
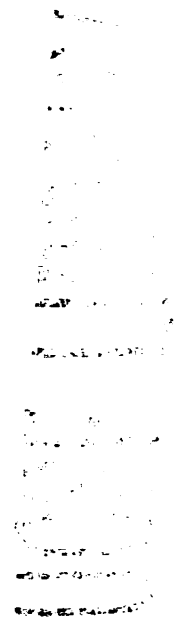


Figure 2-1. Design of mEcotin and scEcotin

Two additional mutations were then introduced into mEcotin to reduce the secondary site interactions or alter the primary binding site. Residues 67 to 70 in ecotin have been shown to be important for the ecotin secondary site interaction (Yang et al., 1998). In the 67-70A_4 mEcotin variant, four alanines replace the protease contacting residues 67 to 70, and an M84R mutation replaces the P1 Met with an Arg in the M84R mEcotin variant. An Arg at residue 84 is optimal against trypsin because a salt bridge can be established with the Asp189 at the base of the primary specificity pocket of trypsin. A single chained ecotin mutant (scEcotin) is also created as depicted in Figure 2-1. Except its function as a molecular weight standard in the gel filtration analysis, other results from biochemical analyses of scEcotin are not included in this thesis.



2.2.2 Purification of mEcotin Variant

The mEcotin variant is purified by C. T. Eggers as described in Appendix One. Details about the mutagenesis of the mEcotin variant can also be found in Appendix One.

2.2.3 Biochemical Analyses on mEcotin

The mEcotin variant is subject to various biochemical methods, including gel filtration, analytical centrifugation, fluorescence resonance energy transfer (FRET), fluorescence titration and kinetic parameter measurements (*i.e.* on and off rates K_{on}, K_{off}, apparent inhibition constants K_i and equilibrium constant K_d). Protocols and conditions for these analyses can be found in Appendix One. A summary of the biochemical analyses is also included in the result and discussion section of this chapter.

2.2.4 Crystallization of mEcotin

The purified and lipholized mEcotin protein was dissolved in ddH_2O at concentrations between 12-15 mg/ml and crystallized by hanging drop vapor diffusion method in the presence of 28-30% polyethylene glycol 4K, 0.1 M Tris at pH 8.0, 0.2 M NaAcetate and 15% glycerol. 1 ml aliquote of silicon oil and paraffin oil (at 1:1 ratio) was placed over 1ml of the well solution to control the vapor diffusion rate. A full data set was collected at Stanford Synchrotron Radiation Laboratory beamline 7-1 using a Mar345 imaging plate system. The data was evaluated and integrated using SCALEPACK/DENZO (Otwinowski & Minor, 1997). The data set was 97% complete to 1.8 Å. 200193 measurements of 15534 reflections where mean I/I_σ is 23.85 and R_{symm} to 1.8 Å is 9.6%. Processing statistics indicated that the crystal had a space group of $P4_32_12$ with unit cell parameters of a = b = 97.24 Å, c = 37.23 Å, $\alpha = \beta = \gamma = 90$ degree. Each asymmetric unit contained only one molecule of mEcotin. Solvent content in crystal is 55%.

2.2.5 Structure Determination and Refinement

A model of mEcotin was constructed based on the uncomplexed ecotin dimer structure (Shin et al., 1996), using only one ecotin molecule from the dimer structure. Five residues at the N-terminus and the last twelve amino acids at the C-terminus were omitted in our model for mEcotin. The mEcotin structure was determined by molecular replacement using programs from the CNS1.0 program suites (Brunger et al., 1998).

Data between 15.0 Å to 4.0 Å was used in the rotational search. Top 20 solutions were selected for translational search using data up to 2.0 Å. After the first round of rigid body refinement, the best solution from translational search resulted in an R-value of 45.2%. The mEcotin structure was refined reiteratively using the CNS1.0 program suites (Brunger et al., 1998). Between rounds of refinement, the structure was visualized and rebuilt using Quanta98 (Molecular Simulations Inc., San Diego CA). The structure has been refined to 2.0 Å. Final R-free is 26.64%, and R-working is 24.14% for data in the range of 6-2.0 Å with 2 sigma cut-off. The R.M.S.D. values for bond lengths and angles are 0.007 Å and 1.4 degree, respectively. Structure comparison study was performed using InsightII (Accelrys. San Diego, CA).

2.3 Results and Discussions

The goal for designing a monomeric ecotin was to create a stable variant that would be incapable of dimer formation even at high concentrations and would bind to proteases only at its primary site in the same orientation and conformation as a wild-type ecotin molecule does. In order to verify the design strategy and to determine if mEcotin maintains the structural integrity and protease-binding surface of wild type ecotin, the mutant is subject to various biochemical analyses, including gel filtration, analysis, fluorescence titration, analytical centrifugation. Finally, the X-ray crystal structure of mEcotin in its uncomplexed state was determined to high resolution.

The mEcotin variant was expressed and purified with relative ease from an *Escherichia coli* expression system. Various methods were adapted to assess its folding

properties as well as its ability to inhibit proteases. All the biochemical assays were performed by Christopher Eggers in Dr. Charles Craik's laboratory at UCSF. In order to provide a complete understanding of the protein, I have included some relevant results from the biochemical analyses. Additional experimental results can be found in Appendix One.

2.3.1 Protein Stability Analysis

The mEcotin protein and its variants were all stable in a purification protocol that included boiling, acidifying to pH 3.0, and reverse-phase HPLC, indicating that the design did not compromise the exceptional stability of ecotin (McGrath et al., 1991). The apparent melting temperature (K_m) of mEcotin is approximately 74°C, which is within error of the K_m of the wild type ecotin at 73°C (data not shown). The fluorescence spectrum of native mEcotin is red-shifted compared to the wild type ecotin by about 8 nm, possibly due to the more exposed Trp130. The mEcotin protein, however, is not destabilized by the exposure of its normally-buried hydrophobic surface, suggesting that the dominant energetic interactions at the dimer interface are between the C-terminal β-strand and the β-barrel core structure, which is present in both the wild type ecotin and mEcotin.

Table 2-1. Data and refinement statistics for the mEcotin crystal

Space group	$P4_32_12$
Cell (Å)	
a	97.24
b	97.24
c	37.23
$\alpha = \beta = \gamma$	90°
observations	200193
unique	15534
[a] R_{merge}	9.6% (three data sets merged together)
Resolution (Å)	1.80
completeness	97.5%
highest shell (1.86-1.80Å)	84.5%
$\langle I/I_\sigma \rangle$	23.85
Current R values (refined to 2.0Å):	
[b] $R_{working}$	24.14%
[c] R_{free}	26.64%
r.m.s.d. (angle)	1.4 degree
r.m.s.d. (bond length)	0.007Å

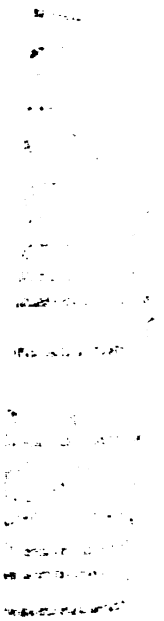
[a] $R_{merge} = \Sigma|(I - \langle I \rangle)| / \Sigma(I)$

[b] $R = \Sigma_{h,k,l}(|F_{obs}(h,k,l)| - k|F_{calc}(h,k,l)|)^2 / \Sigma_{h,k,l}|F_{obs}(h,k,l)|$

[c] free R: Cross-validation R calculated by omitting 10% of the reflections (Kleywegt & Brunger, 1996).

Figure 2-2. X-ray structure of mEcotin

The structure of mEcotin is depicted by its secondary rendering, showing β-sheets in yellow, helices in magenta and flexible loops and turns in blue. The introduced β-turn is highlighted in red. The 80's loop from the primary binding site and the 60's loop from the secondary site are labeled. The structure demonstrates the conservation of the flattened β-barrel fold and the folding back of the C-terminal arm onto the β-barrel. The figure is prepared using BOBSCRIPT & Raster3D (Bacon & Anderson, 1988; Merritt & Bacon, 1997).



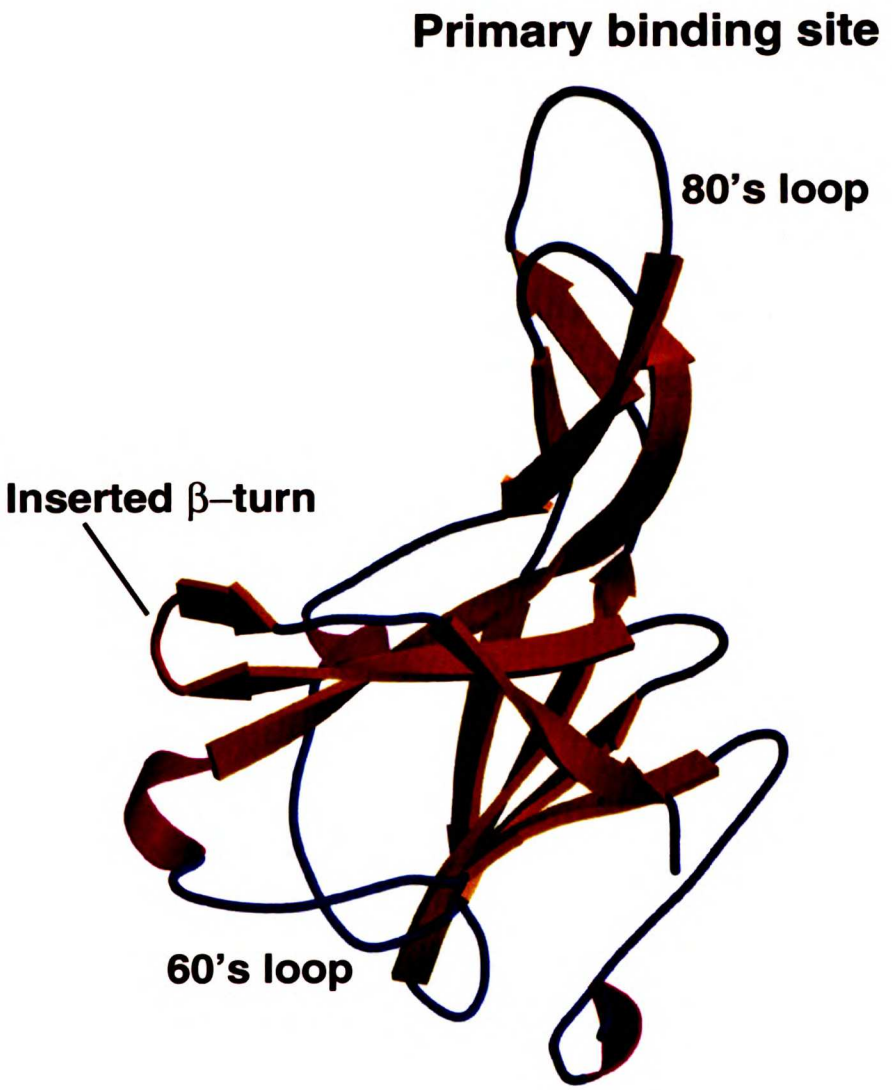


Figure 2-2. Structure of mEcotin

Figure 2-3. Electron density surrounding the inserted loop

The electron density around the engineered β-turn is calculated and shown here as a simulated annealing $3F_o$-$2F_c$ map contoured at 1σ. The three inserted amino acid residues following Trp130, Ala-Asp-Gly, are labeled as residues 130A, 130B, 130C, respectively. Electron density is missing for the side chain of the surface residue Lys131. The side chain is deleted from molecular structure. The difference map is calculated using CNS program suites (Brunger et al., 1998), and the figure is prepared by Raster3D (Merritt & Bacon, 1997).

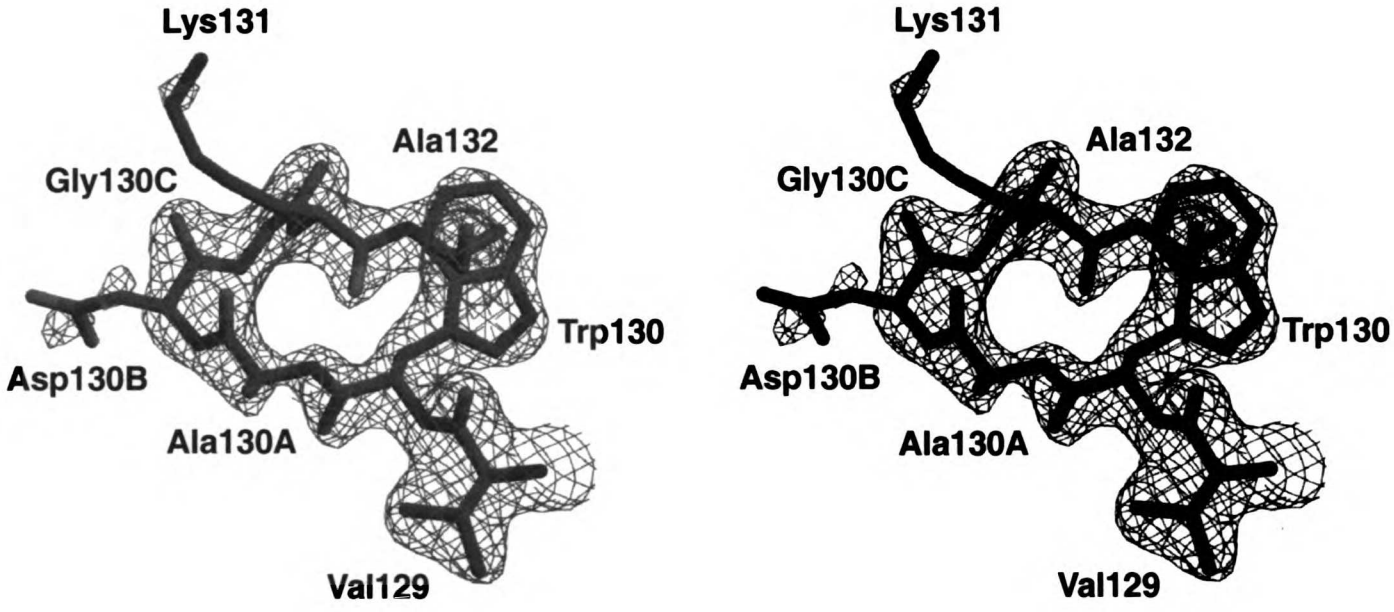


Figure 2-3. Election density map of the inserted loop

Table 2-2. Conformational changes in mEcotin

Different ecotin structures are compared to look for structural changes due to the inserted β-turn, r.m.s.d. Values in Å for C_α atoms from different regions are calculated between mEcotin and ecotin when complexed with trypsin D102N (McGrath et al., 1994), and mEcotin and ecotin in uncomplexed form (1 ECY.pdb).

	Ecotin-D102N (complexed)	Ecotin (uncomplexed)
Overall core residues (12-43, 59-63, 70-76, 103-130)	0.36	0.61
C-terminal residues (132-142)	0.36	0.52
Residues in the primary binding site (83-85)	0.04	0.14
Residues in the secondary binding site (108-112, 67-70)	0.27	0.98

2.3.2 X-ray Analysis of mEcotin Verifies Design

The mEcotin variant is crystallized in the space group of $P4_32_12$. Each asymmetric unit contains only one ecotin molecule. The final X-ray crystal structure of mEcotin is determined by molecular replacement. The structure has been refined to 2.0 Å. Data and structural refinement statistics are listed in Table 2-1. The structure clearly demonstrates that mEcotin exists as a monomeric molecule under our crystallization condition at high protein concentration of 15 mg/ml (Figure 2-2). The mEcotin structure supports our data from gel filtration and analytical centrifugation studies. This structure has been refined to 2.0 Å and clearly demonstrates that mEcotin exists as a monomer under the high protein concentrations of the crystallization conditions.

2.3.2.1 Detailed Structural Analysis of the mEcotin Molecule

The overall fold of mEcotin, depicted in Figure 2-2, is identical to that of the wild type ecotin, either in its uncomplexed form or bound to serine proteases (McGrath et al., 1994; Perona et al., 1997; Shin et al., 1996). The structure of mEcotin maintains a non-globular shape, made of a mostly anti-parallel seven-stranded β-barrel structure (Figure 2-2). Flexible loops connect the anti-parallel β-sheets and form the protease binding sites. The general fold lacks of helical structure, containing only three very short helices. The core residues of mEcotin (12-43, 59-63, 70-76, 103-130) show only a 0.36 Å r.m.s. deviation (Table 2-2) with respect to ecotin complexed with trypsin (McGrath et al., 1994) and 0.61 Å when compared to the uncomplexed ecotin (Shin et al., 1996). Both the primary and the secondary binding sites are essentially unaltered by the three-residue

insertion after Trp130. Two protruding loops, the 80's loop and 50's loop, still highlight the primary binding site of mEcotin. In the absence of a protease binding partner, the 80's loop, which binds to the active site of protease, is in a less extended conformation comparing to its bound form. The reactive site residues in the 80's loop, especially the side chains for residues Met84 and Met85, become more disorganized without the protease binding partner (Table 2-2). The B-factors for the residues in the reactive site loop are relatively higher when compared with those of the rest of the molecule. Similar results have been observed for the structure of wild type dimeric ecotin in its uncomplexed form (Shin et al., 1996). The disulfide bond between Cys50 and Cys87 is preserved in mEcotin, indicating that the 50's loop's main function remains as providing support to the active site binding 80's loop. The secondary protease binding site, which consists of the 60's and 110's loops, also reveals little conformational change in mEcotin when compared with the structure of an ecotin molecule in dimeric context (Table 2-2). The two loops remain short and form a relatively flat surface in comparison to the primary binding site.

At the C-terminus of the molecule, the structure verifies the design. The Ala-Asp-Gly insertion, together with Lys131, forms a type I' β-turn nearly identical to the modeled type I' turn from staph nuclease (residues 94 to 97 in 1A3V.pdb) (Wynn et al., 1997), with an r.m.s. deviation of 0.16 Å for the main-chain atoms (Figure 2-3). This turn successfully folds the last eleven residues back onto the β-barrel core, forming a β-strand that makes contacts similar to those formed by the C-terminal arm from the other ecotin subunit in the dimer structure. Seven of ten main-chain hydrogen bonds are

preserved in the mEcotin structure, with additional water-mediated hydrogen bonds to compensate the missing interactions.

2.3.2.2 More Stable or Less Stable: mEcotin *vs.* Ecotin in Dimer Form

The mEcotin structure seems to be more solvent-exposed than the dimeric ecotin structure. For example, residue Trp130, a residue critical for the packing of ecotin molecules (Yang & Craik, 1998), is usually deeply buried in the dimer interface and forms extensive hydrophobic contact with Trp130 from the other ecotin molecule in the existing structures of all dimeric ecotin molecules. In the structure of mEcotin, however, the electron density map around the β-turn revealed that the side chain of Trp130 in mEcotin is exposed in solvent and disordered (Figure 2-3). On the other hand, although crystallization of the mEcotin protein occurs rather easily, it was very difficult to obtain a crystal that would diffract well in all orientations: strong twinning always occur in certain orientations. The lack of perfect organization in the mEcotin crystals may be due to several factors: imperfect crystallization conditions, freezing-related disruption of the crystals, or instability of the mEcotin molecule itself. The crystal used in our structure analysis has a very high average B-factor, which also may be due to the instability of the mEcotin molecules. The above observations suggest that mEcotin is less stable than the dimeric form of ecotin.

It has been observed that ecotin-protease interactions stabilize the flexible binding loops in ecotin (Shin et al., 1996), which allow us to formulate the hypothesis that ecotin in complexed form is more stable than in uncomplexed form. Table 2-2 shows that the

difference between mEcotin and the dimeric ecotin in the ecotin-trypsin complex with ecotin is smaller than the difference between mEcotin and the uncomplexed dimeric ecotin. This observation is not only true for certain restricted regions of the ecotin molecule, but also true for the entire ecotin molecule, including the β-barrel core structure, the c-terminal arm and the protease binding loops in both the primary and the secondary sites (Table 2-2). The structure of mEcotin resembles more to the structure of ecotin bound to protease than to the structure of uncomplexed ecotin, which suggests that mEcotin is more stable than the uncomplexed dimeric ecotin. Conflicting conclusions suggest that further evidence is needed to conclude whether converting ecotin to mEcotin by inserting the abrupt β-turn has destabilized the protein.

2.3.3 mEcotin, Its Variants and Ecotin Dimerization

In the absence of any serine protease target, the mEcotin protein appeared monomeric (Eggers et al., 2001). In gel filtration analysis, the mEcotin protein had an observed molecular weight of 20,600 dalton, nearly half of the molecular weight of the wild type ecotin (39,000 dalton), at concentrations from 25 μM up to 1 mM. In analytical centrifugation analysis, the data for mEcotin fitted to a single species with a molecular weight of around 17,300 dalton, also nearly half of the molecular weight of the wild type ecotin (approximately 30,700 dalton). In the presence of serine protease target, the characteristics of the mEcotin protein and its variants appeared more complicated. The mEcotin protein initially formed a tetrameric complex with trypsin (Eggers et al., 2001). But when excess amount of mEcotin was added, an additional complex appeared with the molecular weight close to a heterodimer. Only when additional mutations were

introduced to the mEcotin variant to strengthen the primary binding site (M84R) and to severely weaken the secondary site (67-70A_4), a heterodimer became the only species other than the ecotin and protease components. The results from gel filtration analysis are shown in Figure 2-4. The detailed protocol is included in Appendix One. The monomeric nature of mEcotin is also confirmed by sedimentation equilibrium analysis, protocol and results of which can also be found in Appendix One.

The dimeric wild type ecotin can be easily converted to a monomeric mEcotin, which confirms that the dimeric nature of ecotin originated from a "domain swap" mechanism, similar to that seen in diptheria toxin and many other proteins (Bennett et al., 1994). Protein dimerization is frequently achieved by extending a terminal arm to embrace the opposite subunit (Richardson, 1981). This arm-exchange mechanism is observed in such proteins as methylamine dehydrogenase, tumor necrosis factor, and bleomycin-resistance protein, where a conserved proline at the hinge of the arm has been proposed to help keep the arm in a conformation suitable for oligomerization (Bergdoll et al., 1997). The hinge residue of ecotin appears to be Trp130, but it is unclear what prevents the C-terminus of the wild-type protein from folding back on itself before dimerization can occur.

Figure 2-4. Gel filtration analysis of ecotin variants

Gel filtration analysis of ecotin variants and trypsin D102N-ecotin complexes on a Superdex 200 column. The apparent molecular weight, top axis, was calculated from the elution volume, bottom axis, using protein standards. (a) The following ecotin variants were run separately: WT (solid), scEcotin (dotted), and mEcotin M84R (dashed). (b) WT ecotin was complexed with excess trypsin (solid). Trypsin was complexed with excess WT ecotin (dashed). Trypsin was complexed with excess mEcotin M84R 67-70A_4 (dotted). WT stands for wild type ecotin; scEcotin is a single chained formed ecotin variant; mEcotin, mEcotin M84R and mEcotin M84R 67-70A_4 are three monomeric ecotin variants.

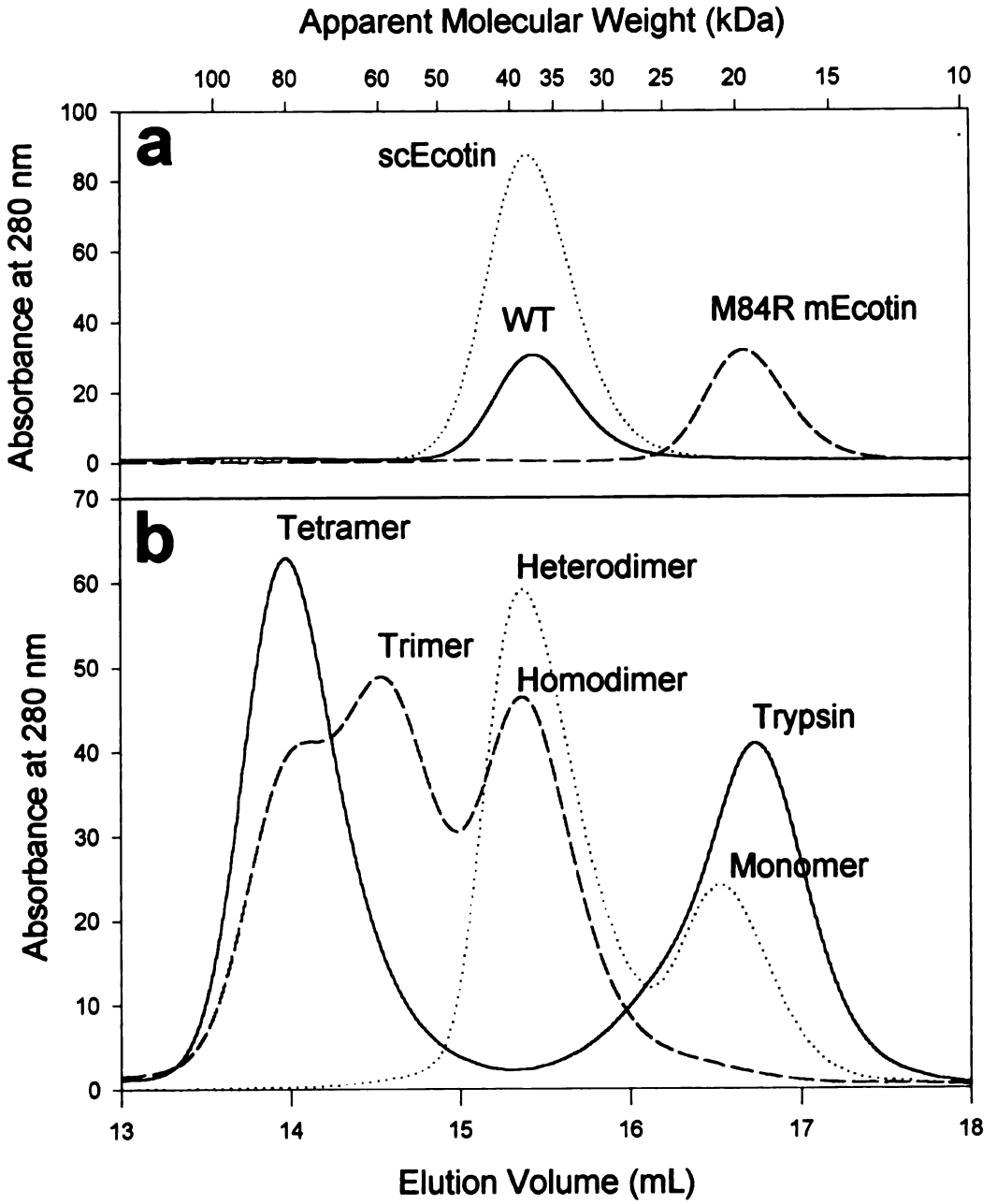


Figure 2-4. Gel filtration analysis of mEcotin & its variants

Table 2-3. Kinetic constants and equilibrium dissociation constants for ecotin variants.

This table is taken from the paper (Eggers et al., 2001). Detailed description of the experiments performed to obtained the kinetic constants can be found in Appendix One.

Table 2-3. Kinetic constants and equilibrium dissociation constants for ecotin variants.

A. Trypsin Ecotin Variant	k_{on} x 10^{-6} (M^{-1} sec^{-1})	k_{off} (sec^{-1})	K_d (k_{off}/k_{on})	K_i^*	$\Delta G_{binding}$ [a] (kcal/mol)
WT	0.86 ± 0.01 (2)[b]	1.16 ± 0.61 x 10^{-7} (3)	130 ± 70 fM	530 ± 80 fM (2)	-17.55 ± 0.31
scEcotin	0.90 ± 0.03 (2)	9.1 ± 1.4 x 10^{-8} (2)	100 ± 20 fM	650 ± 80 fM (2)	-17.72 ± 0.09
mEcotin	N.D. [c]	N.D.	N.D.	1.42 ± 0.03 μM (3)	-7.97 ± 0.01
$60A_4$ mEcotin	N.D.	N.D.	N.D.	1.40 ± 0.06 μM (3)	-7.98 ± 0.03
M84R	3.13 ± 0.11 (2)	4.9 ± 0.7 x 10^{-9} (2)	1.6 ± 0.2 fM	N.D.	-20.19 ± 0.09
M84R mEcotin	2.62 ± 0.11 (2)	1.05 ± 0.02 x 10^{-5} (1)	4.0 ± 0.2 pM	3.2 ± 1.0 pM (1)	-15.54 ± 0.03
M84R $60A_4$ mEcotin	2.56 ± 0.05 (2)	1.06 ± 0.11 x 10^{-5} (2)	4.2 ± 0.4 pM	2.1 ± 0.2 pM (1)	-15.52 ± 0.06
B. Chymotrypsin					
WT	1.93 ± 0.08 (2)	8.4 ± 2.8 x 10^{-9} (3)	4.4 ± 1.5 fM	N.D.	-19.58 ± 0.20
scEcotin	1.89 ± 0.05 (2)	5.0 ± 2.6 x 10^{-9} (2)	2.7 ± 1.4 fM	N.D.	-19.87 ± 0.30
mEcotin	0.89 ± 0.02 (3)	2.6 ± 0.8 x 10^{-6} [d]	N.D.	3.0 ± 0.9 pM (2)	-15.72 ± 0.18
$60A_4$ mEcotin	0.78 ± 0.07 (2)	2.5 ± 0.9 x 10^{-6} [d]	N.D.	3.2 ± 1.1 pM (2)	-15.67 ± 0.20
M84R	3.06 ± 0.06 (2)	1.05 ± 0.10 x 10^{-7} (2)	34 ± 3 fM	300 ± 30 fM (1)	-18.36 ± 0.06
M84R mEcotin	0.71 ± 0.03 (2)	4.9 ± 0.3 x 10^{-5} (2)	69 ± 5 pM	82 ± 4 pM (2)	-13.86 ± 0.05
M84R $60A_4$ mEcotin	0.63 ± 0.02 (2)	4.7 ± 0.5 x 10^{-5} (2)	74 ± 8 pM	108 ± 2 pM (2)	-13.81 ± 0.06

[a] Free energy of binding was calculated as RT ln K_i^* for mEcotin and $60A_4$ mEcotin and as RT ln K_d for all other variants.

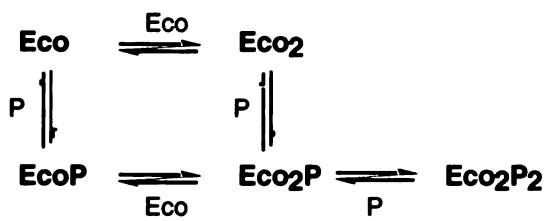
[b] Measured values represent the average of the number of independent determinations given in parentheses.

[c] N.D., not determined.

[d] Dissociation rate constant calculated as K_i^* x k_{on}.

2.3.4 mEcotin, Its Variant and Ecotin-Protease Interactions

The following scheme depicts the multiple equilibrium steps in the interactions between ecotin and proteases, where Eco denotes ecotin and P denotes protease. Intermediates including EcoP, Eco_2P have not been isolated in previous analyses (Pal et al., 1996; Yang et al., 1998).



Equilibrium between ecotin and protease

2.3.4.1 mEcotin and Its Variants Dissect the Ecotin-Protease Binding Interactions

The mEcotin mutant as well as several variants of mEcotin, such mEcotin M84R and mEcotin 67-70A_4/M84R, allowed us to separate the effects from the two protease-binding sites, which ultimately permits the isolation of all intermediates of the complex binding interactions (Eggers et al., 2001). Figure 2-3 clearly shows the formation of an $ecotin_2$ dimer, an ecotin-trypsin heterodimer, an $ecotin_2$-trypsin trimer and an $ecotin_2$-$trypsin_2$ tetramer.

The kinetic and equilibrium constants of both monomeric and dimeric ecotin variants were measured against rat trypsin and bovine chymotrypsin in order to determine the effect of the secondary binding site (Table 2-3). These enzymes display different

substrate specificity, but nevertheless both form high-affinity complexes with ecotin. Because many of these protease-inhibitor complexes are extremely stable, the dissociation constants were determined in most cases by taking the ratio of the dissociation and association rate constants. All kinetic studies were performed by Christopher Eggers, and the detailed description can be found in Appendix One.

2.3.4.2 mEcotin Reveals Contribution from Secondary Site

This mEcotin variant was shown to be monomeric at concentrations as high as 1 mM, by both analytical centrifugation and gel filtration, indicating that reversing the domain swap found in wild type ecotin raised the K_{dim} at least seven orders of magnitude. Additionally, mEcotin variants are capable of forming stable heterodimers with a protease. The mEcotin M84R 67-70A_4 variant, which has both a disrupted secondary site and dimer interface, forms heterodimers with trypsin on a gel filtration column (Figure 2-4). During kinetic studies, the binding energy at the secondary site ($\Delta\Delta G_{2ndary}$) is calculated as the difference between the binding energy of wild type ecotin and the binding energy of mEcotin variants (Table 2-3). The value of $\Delta\Delta G_{2ndary}$ changes upon mutating the P1 residue as well as upon changing protease target. For example, $\Delta\Delta G_{2ndary}$ is –9.58 kcal/mol out of the –17.55 kcal/mol total binding energy for wild type ecotin against trypsin, while $\Delta\Delta G_{2ndary}$ changes to only –3.86 kcal/mol out of the –19.58 kcal/mol binding energy for wild type ecotin against chymotrypsin.

The mEcotin M84R variant, however, appears to form tetramers, but not trimers, with trypsin at high concentrations (Figure 2-4). It would appear that the energy from

two secondary sites, but not one, is capable of driving dimerization with a disabled dimer interface. It is worth noticing, however, that disruption of the secondary site had no effect on the measured affinities of monomeric ecotin variants under our assay conditions, indicating that the heterodimer was the only relevant complex present when determining binding energetics (Table 2-3).

2.3.4.3 Non-additivity Reveals Compromised Binding at Both Sites

It was long observed in kinetic studies that binding effects from the two protease-binding sites are not additive (Yang & Craik, 1998; Yang et al., 1998). Such non-additivity may be due to compromised binding since an X-ray structure of ecotin 67-70A_4/M84R double mutant clearly shows that weakening the secondary site has given ecotin more freedom to binding at the primary site (Gillmor et al., 2000). The mEcotin and its variants are more selective: mEcotin favors chymotrypsin to trypsin by 10^6 fold, but M84R mEcotin binds to trypsin thirty times tighter than to chymotrypsin (Table 2-3). Kinetic data of mEcotin indicate that, depending on the protease target and the mutations present at its two binding sites, ecotin may achieve optimal binding at either the primary site or the secondary site, but not both. Figure 2-5 illustrates the concepts of optimal and sub-optimal binding at either site.

Figure 2-5. Model of compromised interactions at the two binding sites

The compromised binding at the protease binding sites of ecotin is illustrated in the following figures. 2-5A) depicts the optimized binding at the primary site, 2-5B) shows the optimized binding at the secondary site.

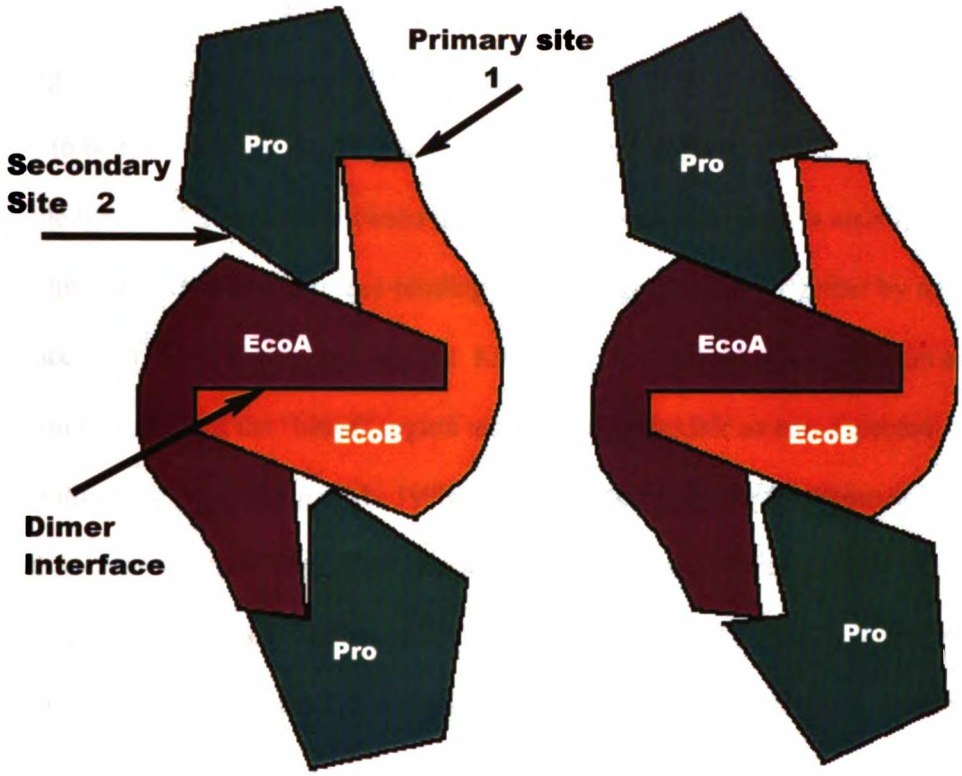


Figure 2-5. Compromised binding in ecotin-protease tetramer

Is it possible to achieve optimal binding at both sites? If it is possible, how can we improve binding at both sites simultaneously? We can only answer this question if we know the cause of the non-additivity (or dependence) between the two protease-binding sites. The bidentate-binding mode utilizing both the primary and secondary sites seems to be facilitated by the dimerization of the ecotin molecules. It is therefore highly possible that the dependence between the two protease-binding sites is also due to the dimer interface. The two protease-binding sites are tightly linked together by the dimer interface, also called the "hinge" region. Kinetic studies on both dimeric ecotin and mEcotin indicate that the "hinge" region may not be as flexible as one expected (McGrath et al., 1994; Yang et al., 1998). Thus, the formation of the tetrameric complex, though increases the binding affinity drastically, at the same time casts steric constrain on the four protein domains. We can safely assume that optimal binding between any two protein-domains only occurs in a certain orientation. In an ecotin-protease tetrameric complex, interactions occur at four ecotin-protease interfaces: Eco_{A1} & Pro_A, Eco_{A2} & Pro_B, Eco_{B1} & Pro_B, Eco_{B2} & Pro_A as well as one ecotin-ecotin dimer interface: Eco_A & Eco_B (Figure 2-5). It is almost impossible for all proteins to be at optimal orientations with respect to each other simultaneously in order for optimal binding to occur: compromise has to occur to permit the tetramer formation. Being in the center of the tetramer, the apparent rigidity of the dimer interface likely should account for some of the compromise occurring in the interactions, but the extent of its influence is not unknown.

2.4 Conclusions & Future Directions

Our crystal structure of mEcotin verifies this design strategy, revealing a monomeric protein with a β-turn of the same type as modeled (Figure 2-2). Furthermore, it shows that the basic structure of ecotin, including the primary site loops, remains essentially unchanged, making it likely that mEcotin binds to the active sites of proteases with the same affinity as a wild type ecotin monomer. The simplicity of the mutation needed to change ecotin from a monomer to a dimer or *vice versa* highlights the efficiency of "domain swapping" as a mechanism of evolution for oligomeric proteins. The complicated network of ecotin-protease interactions is simplified by the use of mEcotin mutant. In the following chapter, the dimer interface is severely reduced by a C-terminal deletion to test its effect on ecotin-protease tetramer formation, in the absence or presence of primary site mutation.

Chapter 3

Ecotin-Trypsin Interactions

3.1 Introduction

The *Escherichia coli* originated protease inhibitor ecotin inhibits most serine proteases of the chymotrypsin fold, regardless of their substrate specificity (Chung et al., 1983; Seymour et al., 1994; Ulmer et al., 1995). When inhibiting proteases, two ecotin molecules dimerize and bind two protease molecules to form a hetero-tetramer (Figure 1-3) (McGrath et al., 1994; Perona et al., 1997). Previous structural analyses reveal that an ecotin-protease complex contains a total of five protein-protein interfaces (Figure 1-3). These five interfaces can be further categorized into three types: two ecotin-protease interfaces between the primary binding site of ecotin and the active site of the protease; two ecotin-protease interfaces between ecotin secondary binding site and the C-terminus of the protease; and an interface between the C-termini of two ecotin molecules. The primary binding site from one ecotin and the secondary site from the other ecotin contact the same protease simultaneously, which is made possible by the "hinge"-like dimer interface. The unique tetrameric arrangement between the two ecotins and the two proteases is believed to account for ecotin's pan-specificity in protease inhibition (McGrath et al., 1994). The complexity and delicacy of the ecotin-protease tetramer have also made it a good choice for studying protein-protein interactions.

The characteristics of the protein-protein interfaces between ecotin and protease have been studied using methods of mutagenesis and x-ray crystallography. Both

protease-binding sites contribute to the strong affinity of ecotin for serine proteases, which can often be improved by mutations at either its primary protease-binding site or its secondary protease-binding site. For instance, ecotin P1 mutants M84K or M84R can strengthen the binding to the proteases that have a preference for basic residue at its primary substrate-binding pocket S1 such as trypsin, factor XIa, and thrombin (Seymour et al., 1994). The significance of the unique secondary-binding site is well established as abolishing the side chains of its protease-contacting residues severely reduced ecotin's ability to inhibit protease (Yang et al., 1998). Also, a Y69F/D70P double mutation at the secondary site increased ecotin's binding affinity to trypsin by 10 fold (Yang & Craik, 1998).

3.1.1 The Dimer Interface

Although the interface between the two ecotin is the largest protein-protein interface in ecotin-protease complexes, its function has remained less understood (Yang et al., 1998). In all existing ecotin-protease structures, the interface between the two ecotin molecules covers over 3000 A^2 in buried surface area (Gillmor et al., 2000; McGrath et al., 1994; Perona et al., 1997). This extremely large interface was immediately assumed to be responsible for the ecotin dimer formation, though no experimental evidence was obtained until 1996. Pal and colleagues showed that an ecotin variant with its C-terminus truncated from residue 133 to residue 142 was isolated as a monomer during gel filtration analysis (Pal et al., 1996). At the same time, Stephen Q. Yang, a graduate student in Dr.

Charles Craik's laboratory at UCSF, conducted a study on the characteristics of the C-terminus of ecotin. His early mutagenic studies showed that truncation beyond residue 133 rendered no ecotin expression (Qing and Sarah, unpublished results). A series of ecotin mutants were subsequently created, including an armless mutant which is identical to the mutant from the earlier study (Pal et al., 1996). This mutant exhibited an impaired affinity for trypsin: its apparent inhibition constant (K_i) was over 200 times weaker than the K_i of the wild type ecotin (Yang et al., 1998). Kinetic studies (Yang et al., 1998) and earlier gel filtration analysis (Pal et al., 1996), however, both suggested that the armless mutant could still form tetramers with target proteases such as trypsin (Pal et al., 1996; Yang et al., 1998). Modeling studies indicates that the 'dimer' interface in the armless mutant would be reduced to around 900 $Å^2$, only around 30% of its original size in wild type ecotin dimer. How ecotin can still form tetramer with protease targets after such a drastic decrease in surface area is unknown.

3.1.2 Non-additivity Between Two Binding Sites in Ecotin

The most interesting feature of the ecotin-protease interactions is the non-additivity of the energetic inputs from the two binding sites (Gillmor et al., 2000; Yang et al., 1998). For example, the apparent binding constant for ecotin M84R (an ecotin variant with its P1 residue Met mutated to Arg) is 0.38 nM, and the apparent binding constant for ecotin mutant 67-70A_4 (an ecotin variant with residues 67 to 70 all mutated to Ala) is 4420 nM. The apparent binding constant for the double mutant 67-70A_4/M84R is 0.18

nM instead of 1806 nM (Figure 3-1). The crystal structure of 67-70A_4/M84R-trypsin showed that the secondary site lost most of its contact with the trypsin molecule while the protease rotated 14° towards the primary site (Gillmor et al., 2000). It was concluded that protease binding at two sites in wild type ecotin-trypsin (WT-trypsin) might be compromised; hence, an M84R mutation combining with a severely weakened secondary site allowed near-optimal binding in the primary site. The compromised binding, thus, accounts for the non-additivity between primary and secondary sites. The source of the non-additivity, however, was not identified.

Figure 3-1. Non-additivity between two protease binding sites

The apparent binding constants of three ecotin variants are shown here, all against a common target protease: rat trypsin. For comparison purpose, a binding constant for 67-70A_4/M84R is calculated, provided that the mutations at the two protease-binding sites are additive. The mutants are:

M84R: an ecotin mutant with the P1 residue Met84 mutated to an Arg,

67-70A_4: an ecotin mutant with residues 67 to 70 all mutated to Alas,

67-70A_4/M84R: an ecotin double mutant that contains both the M84R mutation and the 67-70A_4 mutation.

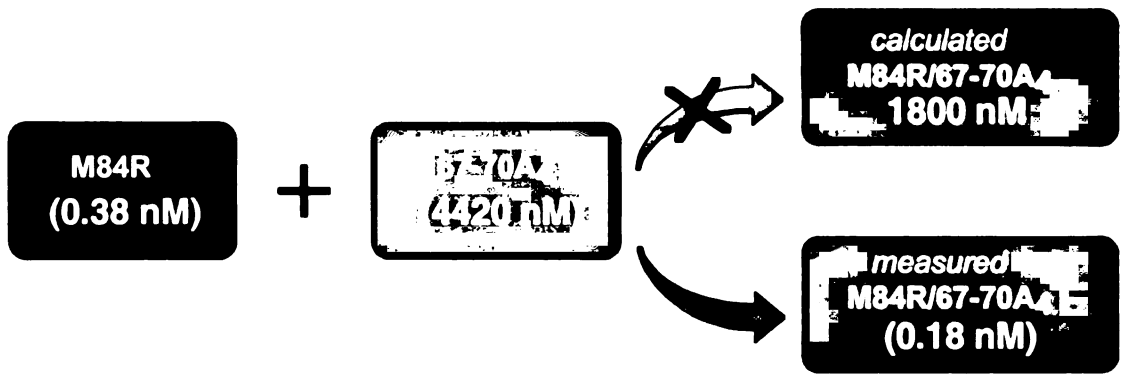


Figure 3-1. Non-additivity of mutations at primary and secondary binding sites in ecotin

Table 3-1. Inconsistency in binding energy from two protease binding sites

Binding energies for each protease-binding site are calculated from apparent binding constants measured using monomeric mEcotin and its variants. This table is based on a figure from a published paper (Eggers et al., 2001).

Ecotin primary site	Primary site (kcal/mol)	Secondary site (kcal/mol)
Rat trypsin		
Wild type	-8.0	-9.6
M84R	-15.5	-4.7
Bovine chymotrypsin		
Wild type	-15.7	-3.9
M84R	-13.9	-4.5

Compromised binding at the two protease-binding sites was further indicated in the studies of the mEcotin mutant (Eggers et al., 2001) (chapter 2). The binding energies of the primary and secondary sites were calculated after the interactions in the tetramer were dissected by using mEcotin and its variants (Table 3-1). As shown in Table 3-1, the binding energies contributed from both binding sites clearly varied with respect to specific mutations as well as protease partners, which indicates that the interactions between ecotin and protease at the two sites are not optimal at the same time. This variation is consistent with the data from previous mutagenic studies, which provides another piece of evidence for the non-additivity between the two binding sites.

Because the ecotin dimer interface lies in the center of a complicated protein-protein interaction network, I hypothesize that it may be responsible for the compromised binding at both sites. The dimer interface is usually regarded as the "hinge" region, which brings the two ecotin molecules together so that the two sites from different ecotin can contact protease simultaneously (McGrath et al., 1994). This "hinge" region spans across about 15 Å in width and over 60 Å in length. Consequently, the two binding sites are forced to be in a certain orientation, likely with very little rotational freedom for the optimization of the relative position of the two ecotins in binding. Therefore, the stiff "hinge-like" dimer interface may have caused the compromised binding at the two protease-binding sites in an ecotin-protease tetramer.

We can further hypothesize that the "hinge" can be made less rigid by reasonable truncation at the C-terminus, provided that the truncation itself will not significant change the tetrameric organization. With a more flexible "hinge", ecotin molecules may fully explore both protease-binding sites simultaneously.

In the study presented in this chapter, I have crystallized the armless mutant, which contains a deletion from residues 133 to 142, in complex with rat trypsin. By doing so, I intend to evaluate how the C-terminal deletion affects the structural organization of the tetramer. I also have crystallized two additional ecotin mutants in complex with trypsin. The two mutants are M84R (which contains a point mutation at residue 84) and armless/M84R (which contains both the C-terminal deletion and the point mutation). By determining the structure of these two additional ecotin-trypsin complexes, I attempt to test the hypothesis that the truncation at the C-terminus can give rise to a more flexible ecotin that forms more extended interactions with its target proteases.

Figure 3-2. Site of mutations in ecotin

The ecotin dimer (1ECZ.pdb) is rendered by secondary structure, with the two monomers colored in gray and light yellow, respectively. The locations of the mutations (M84R and C-terminal truncation from residues 133 to 142) are labeled and highlighted by CPK rendering of the corresponding residues.

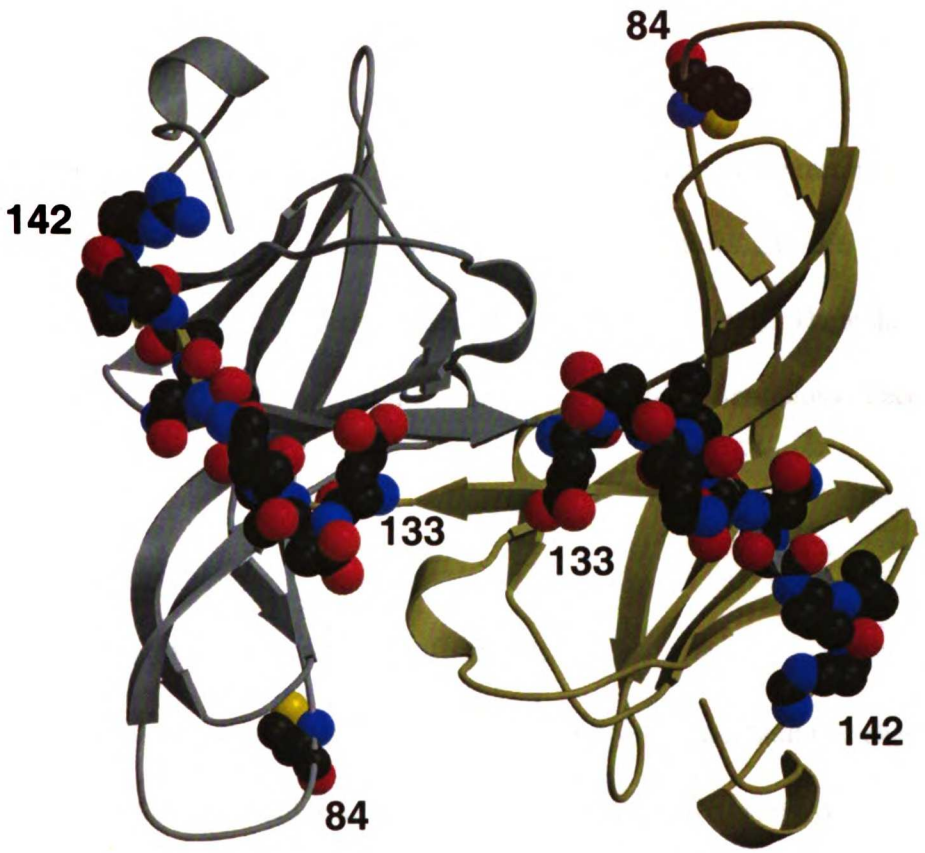


Figure 3-2. Site of Mutatioins in Ecotin

3.2 Material and Methods

3.2.1 Design of Armless, M84R and Armless/M84R

The armless mutant contains a deletion of residues 133 to 142 at the C-terminus. The M84R mutant contains a Met to Arg mutation at the P1 position in the reactive site loop of ecotin. The double mutant armless/M84R bears both mutations. All the mutagenesis was performed by Qing Yang as described (Yang et al., 1998). These three ecotin mutants complete a simple cycle of mutation. Locations of the mutations in ecotin molecule are depicted in Figure 3-2.

3.2.2 Protein Expression and Purification

Rat anionic trypsin mutant D102N, referred to as trypsin throughout this chapter, was used in all crystallization studies for its reduced proteolytic activity and yet remaining isostructural to wild type trypsin (Sprang et al., 1987). The gene for rat trypsinogen D102N was subcloned into the expression vector pPICZαA (Invitrogen, San Diego, CA) by Jennifer Harris. The vector containing trypsinogen D102N was linearized and transformed into *Pichia pastoris*. Protein expression and purification were performed as described (Gillmor et al., 2000; Willett et al., 1995) with minor modifications. Colonies were picked into 10 ml BMMY/Zeocin^{TM}, and grown overnight at 30°C with shaking. This culture was used to inoculate one liter growth media which contains autoclaved BMGY, 100 ml of 1 M Potassium Phosphate at pH 6.0, 100 ml 10x yeast nitrogen base

(134 g dissolved in 1000 ml ddH2O, filtered), and 2 ml of 500x biotin (20mg of biotin in 100 ml filtered ddH_2O). 5 mls of Methanol was added for induction of trypsin expression every 12 hours before the cultured was collected after 72 hours. Detailed recipes for all growth media and a copy of the purification procedure can be found in Appendix Two.

Secreted trypsinogen D102N was loaded on a Phenyl Sepharose column in 50 mM Tris, pH 8.0 and 2.5 M NaCl and eluted by decreasing NaCl gradient to 0 mM over 500 ml volume. The purified trypsinogen D102N was pooled and dialyzed against a buffer containing 50 mM Tris, pH 8.0 and 100 mM NaCl. Enteropeptidase (Biozyme) was added at 1/20 (w/w) and incubated overnight at 37°C to convert trypsinogen to trypsin in the presence of 10 mM $CaCl_2$. Activated trypsin bound to a Soybean Trypsin Inhibitor column and was eluted with 100 mM Acetic Acid, 100 mM NaCl, 10 mM $CaCl_2$ and 50 mM Tris at pH 3.0. All three ecotin variants (armless, M84R and armless/M84R) were constructed by S. Q. Yang and purified following a standard procedure (Yang et al., 1998).

Each purified ecotin mutant and trypsin were mixed at approximately 1: 1 stoichiometry by molar quantity. The mixture was incubated on ice for an hour before it was loaded on HPLC with a strong anion exchange column for further purification following a standard protocol as described (Gillmor et al., 2000; McGrath et al., 1991; McGrath et al., 1994). All purified complexes were in a solution containing 10 mM Tris

at pH 8.0 and 1 mM $CaCl_2$, and concentrated to 5 to 6 mg/ml for crystallization experiments.

3.2.3 Crystallization of Trypsin with Three Ecotin Variants

All purified trypsin-ecotin complexes were crystallized by hanging drop vapor diffusion method, under conditions similar to those for the crystals of the WT-trypsin complex (Gillmor et al., 2000; McGrath et al., 1994). The armless-trypsin complex was crystallized in 2 x 2 µl hanging drops with well solutions containing 10.5% to 14.5% polyethylene glycol 4K, 0.3 M Sodium Acetate, and 0.15 M Sodium Cacodylate (pH 6.2 and 6.5). The M84R-trypsin complex was crystallized in 2 x 2 µl hanging drops with well solutions containing of 13.5% to 16.5% polyethylene glycol 4K, 0.3 M Sodium Acetate, 0.15 M Sodium Cacodylate (pH 6.0 and 6.2). The armless/M84R-trypsin complex was crystallized under several conditions including well solutions containing 12.5% to 15.5% polyethylene glycol 3K or 4K with 0.3 M Sodium Acetate and 0.15 M Sodium Cacodylate at pH ranging from 6.0 to 6.5. Crystallization of the armless/M84R-trypsin complex was optimized with 15.5% polyethylene glycol 4K, 0.3 M Sodium Acetate and 0.2 M Sodium Hepes at pH 6.5, in the presence of 0.1 M Strontium Chloride.

All crystals are very stable; and can be found on trays that are over one year old. Moreover, the "old" crystals diffract to resolution beyond 2 Å, as well as fresh crystals do.

3.2.4 Structure Determination and Refinement

All diffraction data were collected at Stanford Synchrotron Radiation Laboratory (SSRL) on beam line 7-1 or 9-1, using MAR detector systems. For each ecotin-trypsin complex, diffraction data were collected on a single crystal. All data sets were integrated and scaled with DENZO/SCALEPACK (Otwinowski & Minor, 1997). The processing statistics indicated that all three complexes are crystallized in the space group of $P2_1$. All crystals diffract to high resolution. Detailed diffraction data statistics for all three complexes are listed in Table 3-2.

The search model for M84R-trypsin is simply the structure of the WT-trypsin complex (McGrath et al., 1994), while the search model for both armless-trypsin and armless/M84R-trypsin is the structure of the WT-trypsin with residues 133 to 142 deleted. Molecular replacement was used to determine the structures of all three ecotin-trypsin complexes. All structures are refined by the CNS program suites (Brunger et al., 1998), with a bulk solvent correction using data from 6 Å to the highest resolutions. Refinements include rigid body, minimization, simulated-annealing and B-factor refinement. The structures were visualized and rebuilt using Quanta98 (Molecular

Simulations Inc., San Diego CA). Detailed refinement statistics for all structures are also included in Table 3-2.

Buried surface areas in all complexes were measured using the program NACCESS (Hubbard & Thornton, 1993). The program GEM (Eric Fauman) was used to calculate the rotation/translation differences between the protein domains.

Table 3-2. Data and refinement statistics for all three ecotin-trypsin complexes

	M84R-trypsin	Armless-trypsin	M84R/Armless-trypsin
space group	$P2_1$	$P2_1$	$P2_1$
cell parameters (Å)			
a	62.05	50.87	51.06
b	79.35	89.02	88.69
c	80.01	81.34	81.57
β (degree)	97.3°	95.3	95.3°
total observations	407688	276468	309503
unique observations	80666	57817	57340
[a] R_{merge}	8.5%	5.5%	6.2%
highest shell	51.4% (1.76-1.73Å)	54.7% (1.92-1.89Å)	53.6% (1.93-1.90Å)
resolution	1.73 Å	1.89 Å	1.9 Å
completeness	93%	99.7%	98.5%
highest shell	92.7% (1.76-1.73Å)	99.9% (1.92-1.89Å)	92.3% (1.93-1.90Å)
$\langle I/I_\sigma \rangle$	10.1	20.5	18.3
Molecular Replacement:			
Corr. Coeff.	46.1%	49.4%	46.8%
[b] $R_{initial}$	44.0%	42.6%	43%
CNS Refinement:			
Resolution	6-1.8 Å	6-1.89 Å	6-1.9 Å
[b] $R_{crystal}$	20.9%	22.3%	20.6%
[c] R_{free}	22.9%	23.7%	23.0%
rms bond (Å)	0.005	0.01	0.007
rms angle (degree)	1.4 °	1.8 °	1.5 °

[a] $R_{merge} = \Sigma|(I - \langle I \rangle)| / \Sigma(I)$

[b] $R = \Sigma_{h,k,l}(|F_{obs}(h,k,l)| - k|F_{calc}(h,k,l)|)^2 / \Sigma_{h,k,l}|F_{obs}(h,k,l)|$

[c] free *R*: Cross-validation R calculated by omitting 10% of the reflections (Kleywegt & Brunger, 1996).

3.3 Results and Discussion

3.3.1 The Effects of Truncation at the C-terminus of Ecotin

The ecotin armless mutant contains a deletion from residue 132 to residue 142 at the C-terminus. Modeling studies indicated that the armless truncation would likely cause a 70% reduction from the 3200 $Å^2$ dimer interface of wild type ecotin. The K_i of the armless mutant was 200 times weaker than the K_i of the wild type ecotin (Yang et al., 1998). Additional experiments showed that armless mutant was isolated as a monomer by gel filtration method, indicating that truncating the last ten residues of ecotin was sufficient to abolish or severely weaken the interactions at the ecotin dimer interface (Pal et al., 1996). However, further gel filtration analysis also revealed that the armless mutant formed the an armless-protease tetramer in the presence of protease target (Pal et al., 1996; Yang et al., 1998). By determining the structure of armless mutant complexed with trypsin, I hope to evaluate to what extent the C-terminal arm region affects the interactions at the dimer interface, as well as the interactions in the subsequently formed tetramer.

Table 3-3. Tertiary and quaternary structural changes in ecotin-trypsin complexes

All structural comparison studies are done between an ecotin mutant-trypsin complex and WT-trypsin complex.

	M84R-trypsin (Å)	Armless-trypsin (Å)	M84R/Armless-trypsin (Å)
core residues[1] (ecotin monomer)	0.21	0.31	0.31
core residues (ecotin dimer)	0.30	0.55	0.63
primary site residues (83-85)	0.05	0.12	0.2
secondary site residues (67-70, 108-112)	0.23	0.16	0.17
rmsd of protease (protease monomer)	0.24	0.29	0.28
rotation (° degree)[3]	~2.5	~5	~6.5
translation (Å)[3]	~0.5	~2.4	~2.3
BSA[2] primary (Å^2)	1750	1540	1660
BSA secondary (Å^2)	980	930	930
Dimer interface (Å^2)	3220.4	921.9	855.3

[1]Core residues are defined as residues 12-43, 59-63, 70-76, 103-131 in ecotin.
[2]BSA: buried surface area.
[3]The rotational and translational movements of the protease domains are measured with respect to the core residues in each ecotin. Because the tetrameric structure does not possess a strict non-crystallographic symmetry, measurements of the movements are averaged to give the final value.

Figure 3-3. Structural comparison: WT-trypsin vs. Armless-trypsin

The structure of armless-trypsin is superimposed with the structure of the WT-trypsin complex by core residues from only ecotin (A) or from both ecotin molecules (B). The structure of armless-trypsin is colored in navy, and the structure of WT-trypsin is colored in gold. The core residues in ecotin are defined as residues 12 to 43, 59 to 63, 70 to 76, and 103 to 131. In both figures, the superimposed molecules are shown in both a front view and a side view. The figures are rendered by the program InsightII (Accelrys, San Diego, CA).

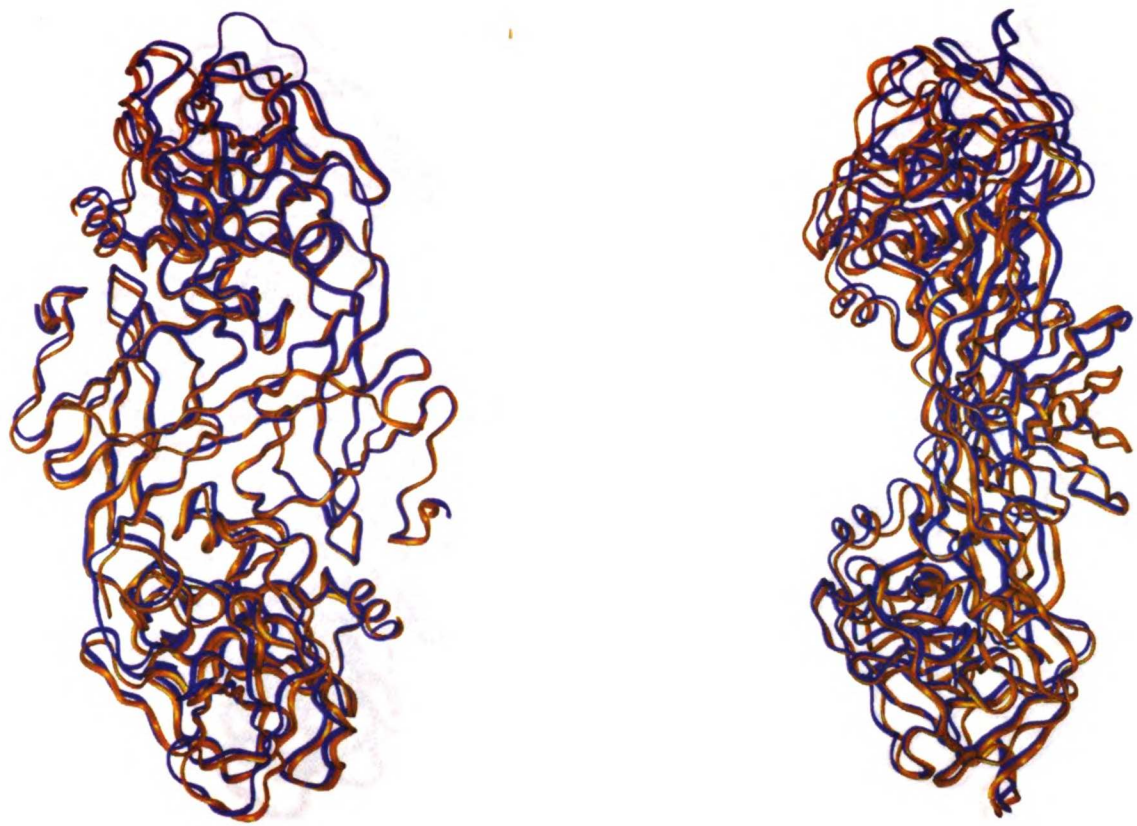


Figure 3-3A. Wild Type Ecotin-Trypsin *vs.* M84R-Trypsin

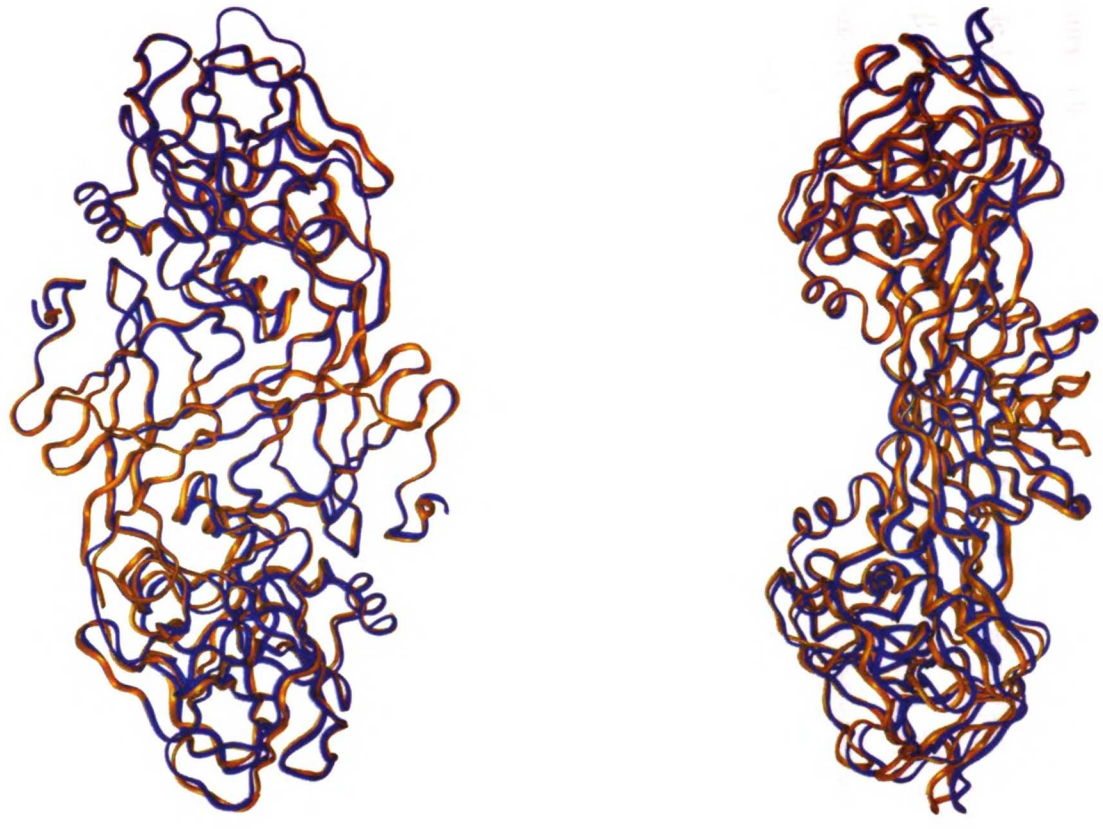


Figure 3-3B. Wild Type Ecotin-Trypsin *vs.* M84R-trypsin

3.3.1.1 Armless-Trypsin Maintains a Tetrameric Structure

The purified ecotin armless-trypsin complex was crystallized easily in the space group of $P2_1$. Every asymmetric unit in the unit cell contains only one $armless_2$-$trypsin_2$ tetramer. The unit cell dimensions of the armless-trypsin crystals, are 51 Å x 89 Å x 81Å (Table 3-2), differing slightly from those of the WT-trypsin crystal (62 Å x 83 Å x 81 Å). The β angle of the unit cell also changed slightly from 97.5° for WT-trypsin to 95.2° for armless-trypsin. The minor changes in crystal unit cell parameters seem to occur rather orderly and only in only certain directions. Crystal packing analyses show that the longest dimension of the tetramer, which does not seem to be affected by the truncation, aligns with the *c*-axis of the unit cell. The truncation-initiated slight rotation occurs along the *c*-axis, which leaves the cell dimension along the *c*-axis intact while causing minor changes in cell parameters along the *a*- and *b*- axes. Over all, the severe C-terminal truncation does not prevent ecotin molecules from forming a conventional tetramer with trypsin. The final x-ray structure of the armless-trypsin complex maintains the general overall features of the known ecotin-protease tetramer complexes.

3.3.1.1.1 Conformational Changes Between Protein Domains

When the structure of armless-trypsin is superimposed with the structure of WT-trypsin by the core residues from only one ecotin, the R.M.S.D. is 0.29 Å (Table 3-3, Figure 3-3A). When the two structures are superimposed based on core residues from

both ecotin molecules, the R.M.S.D. changes to 0.56 Å (Table 3-3, Figure 3-3B). The R.M.S.D. value nearly doubles when both ecotin molecules are used for the superimposition analysis, which is confirmed by the subtle but clearly visible structural differences (Figures 3-3A & 3-3B). The ecotin molecules in the tetrameric complex maintain their structural integrity; their relative orientation slightly changes as a result of the truncation. Deletion at the C-terminus of ecotin may have weakened the dimer interface, which allows the two ecotin-monomers to adjust their position with respect to each other. The movements between the ecotin domains, however, are not as great as those observed for the 67-70A_4/M84R double mutant (which evaluates a truncation of part of the secondary site) when bound to trypsin (Gillmor et al., 2000). The difference is less than 1 Å. Nevertheless, the minor movements support our hypothesis of a flexible "hinge" region.

The two trypsin domains in the armless-trypsin complex remain virtually unchanged with respect to the trypsin domains in the WT-trypsin structure. When each protease domain in armless-trypsin is superimposed with its counterparts in the WT-trypsin complex, the R.M.S.D. is smaller than 0.3 Å for over 230 equivalent C_α atoms (Table 3-3). The relative orientation between ecotin and trypsin changes slightly from what was observed in the WT-trypsin complex. For instance, when corresponding ecotin molecules are aligned by their core residues, the centers of mass (COMs) for the protease domain in the armless-trypsin complex have moved as far as 2.5 Å from its counterparts

in the WT-trypsin complex. Additional to the translational movements, there is a five-degree rotational difference between the corresponding protease domains.

3.3.1.1.2 Changes at Protein-Protein Interfaces

The rotational and translational movements of the protease domains with respect to the ecotin core only slightly alter the interactions between ecotin and trypsin (Table 3-3). The buried surface area at the primary binding site of ecotin is 1600 $Å^2$, slightly smaller than that at the primary site in the WT-trypsin complex. A hydrogen bond is observed between Leu52 from ecotin and Arg96 in trypsin. In all existing ecotin-trypsin structures, residues in the 50's loop do not contact any residues in the protease domain. Interactions between the 50's and protease domains have only been observed in the extended interactions between ecotin and regulatory proteases such as thrombin (Wang et al., 2001) and factor Xa (unpublished results). Our observation suggests that, even when against the simple digestive proteases such as trypsin, the 50's loop from ecotin functions other than as a simple support to the active site binding 80's loop. The buried surface at the secondary binding sites is only slightly larger than 900 $Å^2$, remaining very close to that of the WT-trypsin complex. The 60's and 110's loops of the secondary site of ecotin make the expected hydrogen bonds with the trypsin residues. Drastic reduction of the dimer interface is confirmed by the crystal structure of armless-trypsin. The dimer interface decreases to 920 $Å^2$, from more than 3000 $Å^2$ for the WT-trypsin complex. The protein-protein interactions in the dimer interface are also severely weakened: sixteen of the eighteen direct contacts are lost after the truncation.

3.3.1.1.3 C-Terminal Arm Region, Ecotin Dimerization and Tetramer Formation

Dimerization seems to be essential for the formation of ecotin-protease tetramers. Thus, it is puzzling that the armless mutant, after such severe reduction in its dimer interface, still is able to form a tetrameric complex with trypsin. The structural and functional data promote us to question the role of the arm region in ecotin dimer formation. Two pieces of experimental data suggest its critical contribution to dimerization. First, ecotin exists as a monomer after its dimer interface is severely reduced. For example, truncating the arm region indeed rendered ecotin monomeric as shown in the gel filtration studies (Pal et al., 1996). Second, ecotin exists as a monomer after its dimer interface is completely abolished. For example, when ecotin's extended arm region was retrieved back to its β-barrel core structure by protein engineering, as described in chapter 2, the modified ecotin (mEcotin) was shown to be monomeric by gel filtration and analytical centrifugation analyses (Eggers et al., 2001).

Another question likely to be raised is whether the dimer formation is necessary for the final tetramer formation. Initially, a positive answer seems obvious since dimerization of the ecotin molecules brings a primary site and a secondary site from two different ecotin molecules into a unique configuration so that both sites contact the same protease molecule simultaneously. Further examination indicates, however, that the presence of the dimer interface may be unnecessary for maintaining a stable ecotin-protease tetramer. For instance, if the interactions between ecotin and protease are strong

enough at both the primary and secondary sites, such as in the case of the M84R mEcotin mutant, the ecotin C-terminus is not needed for the formation of a tetrameric complex. As discussed in chapter 2, it is very difficult for all five protein-protein interfaces in the ecotin-protease complex to be at their optimal binding configurations at the same time, reducing or omitting the dimer interface should allow ecotin more freedom to optimize its binding interactions.

3.3.2 Non-additivity & Inconsistency of the M84R Mutation

In protein-protein interactions, if certain mutations do not alter the mechanism of interaction and if they occur over well-separated distance, their effects are often additive, such as those observed for subtilisin, tyrosyl-tRNA synthetase, trypsin and DHFR, and glutathione reductase (Wells, 1990). The two binding sites of ecotin are over 25 Å apart, however, mutational effects at these two sites are not additive (Figure 3-1) (Yang et al., 1998). Previous structural analysis demonstrated that this non-additivity came from the dependence between the two binding sites (Gillmor et al., 2000). It was shown that protease binding at one site was often affected by the mutation at the other site (Yang et al., 1998). The C-terminal arm region is not a part of either protease-binding site, however, non-additivity in binding energy is still found between mutations in the arm region and at the primary binding site. For instance, binding energy contributed from the dimer interface (ΔG_{dimer}) can be calculated by two different approaches. One is to measure the binding energy difference between the wild type ecotin and the armless

mutant, *i.e.* $\Delta G_{dimer} = \Delta G_{WT} - \Delta G_{armless} = -RT\ln K_{i(WT)} + RT\ln K_{i(armless)}$. The other approach is to measure the binding energy difference between two other ecotin variants: the M84R mutant and the armless/M84R double mutant, *i.e.* $\Delta G_{dimer} = \Delta G_{M84R} - \Delta G_{armless/M84R} = -RT\ln K_{i(M84R)} + RT\ln K_{i(armless/M84R)}$. The first equation gives a value of 3.1 kcal/mol, while the second equation give a value of –0.84 kcal/mol, both calculated based on data from Yang *et al.* (Yang et al., 1998). I hope to gain insights into the non-additivity from the crystal structures of the already determined armless-trypsin complex and two more ecotin-trypsin complexes: M84R-trypsin and armless/M84R-trypsin.

3.3.2.1 The X-ray Structure of Ecotin M84R-Trypsin

The primary substrate-binding pocket of trypsin includes residue Asp189, which contributes significantly to trypsin's preference for positively charged residues such as Lys and Arg at P1 site. In ecotin M84R mutant, the P1 residue Met84 is mutated to an Arg to increase its binding to trypsin. Previous analysis confirms that this mutant binds to rat trypsin about three times stronger than wild type ecotin (Yang et al., 1998). The purified ecotin M84R-trypsin complex readily crystallizes in the space group of $P2_1$ (Table 3-2). An $M84R_2$-$trypsin_2$ tetramer is found in every asymmetric unit. The cell parameters of the ecotin M84R-trypsin crystal are almost identical to those of the WT-trypsin complex. Thus, it is not surprising that the final structure of the ecotin M84R-trypsin is virtually identical to the structure of the WT-trypsin complex, with the differences between these two structures all within experimental errors (Table 3-3).

3.3.2.1.1 Structural Comparison M84R-Trypsin *vs.* WT-Trypsin

The structure of M84R-trypsin remains almost identical to the structure of WT-trypsin, when they are superimposed by core residues from either only one ecotin molecule or both ecotin molecules (Table 3-3). When the structures of ecotin M84R-trypsin and WT-trypsin are superimposed by the equivalent C_α atoms in the β-barrel core structure in one ecotin molecule, the R.M.S.D. is merely 0.21 Å (Table 3-3), which is comparable to the experimental errors. When the two structures are superimposed by the equivalent core C_α atoms from both ecotin molecules, the R.M.S.D. becomes 0.30 Å, which suggests that the relative position between the two ecotin molecules only changes slightly. The results suggest that the point mutation at P1 site of ecotin cause very little change in either the tertiary structure of the each ecotin or the quaternary organization between them.

When the structures of M84R-trypsin and WT-trypsin are superimposed by each protease domain, the R.M.S.D. for all equivalent C_α atoms are both around 0.24 Å (Table 3-3), which indicates that the protease domains maintain their typical conformations. When these two structures are superimposed based on core residues in each ecotin, the COMs of the corresponding protease domains in M84R-trypsin and WT-trypsin move around 0.5 Å. Though larger than the experimental error of 0.2 Å, this is considerably smaller than the COMs movement due to the armless mutation. Apparently, the point mutation at ecotin residue 84 has not caused significant changes in either the core

structure of trypsin (tertiary) or the relative position between ecotin and trypsin (quaternary).

3.3.2.1.2 The Protease Binding Sites and Dimer Interface

The buried surface area at the primary binding site of the M84R-trypsin complex is 1750 $Å^2$, slightly larger than that of the WT-trypsin complex. The interactions at the primary site, however, remain very similar to those of the WT-trypsin. Except for the hydrogen bonds formed by the side chain of the Arg84, the remaining hydrogen bond network at both binding sites in ecotin M84R-trypsin remains similar to that of the WT-trypsin complex. The buried surface area at the secondary binding site is 1000 $Å^2$, and the dimer interface is 3200 $Å^2$. Both sites remain very close to those of the WT-trypsin complex (Table 3-3).

The P1 Arg in the M84R-trypsin structure is superimposed with the P1 Met in WT-trypsin structure to reveal how these two residues adapt to the same S1 binding pocket in trypsin (Figure 3-4). There is little change in the size and shape of the S1 binding pocket. Their main chain atoms form identical interactions with the oxyanion hole and the catalytic Ser195 (ref, not shown in the figure), and the side chains for Met and Arg overlap fairly well. The Met side chain, however, fails to make any contact with Asp189, which is at the bottom of the S1 substrate-binding pocket. The Arg side chain in M84R-trypsin makes six direct contacts, including both hydrogen bonds and salt bridges,

with Asp189, Ser190, Gly219, and Ala221. As a result of the two salt bridges formed between Arg84 and Asp189, the side chain of Asp189 in M84R-trypsin moves slightly up towards the incoming P1 residue. The most prominent conformational change is found in Ser190. In both WT-trypsin and M84R-trypsin, Ser190 forms a hydrogen bond with Tyr228 (Figure 3-4). Closer examination reveals that the side chain of Ser190 in the M84R-trypsin structure points towards the incoming Arg and forms a hydrogen bond with one of the terminal nitrogen atoms. The side chain of Ser190 in WT-trypsin, however, rotates over 90 degree away the bound Met, in a conformation similar to that of the Ser190 in chymotrypsin structure (Scheidig et al., 1997; Yennawar et al., 1994).

In the structure of Squash Protease Inhibitor (SPI) bound to bovine trypsin (1PPE.pdb), the P1 Arg of SPI exhibits identical conformation to the Arg in our M84R-trypsin structure (Figure 3-4). Similar network of hydrogen bonds and salt bridges can be found between the P1 Arg and Asp189, Ser190, Gly219, Ala221, and Tyr228. The SPI-trypsin structure and M84R-trypsin structure clearly demonstrate that the S1 substrate-binding pocket is optimized for binding of charged residues. Residues aligning the substrate binding pocket of trypsin, however, are not entirely resistant to structural changes. For example, Ser190 can rotate either inward or outward, to accommodate the hydrophilic or hydrophobic nature of the incoming P1 residue.

Figure 3-4. Comparison of S1 pockets when bound to different substrates

This figures depicts subtle structural re-arrangement of the S1 substrate binding pockets when bound with different P1 residues. The structures of WT-trypsin and M84R-trypsin are compared to reveal differences in P1-S1 interactions up the point mutation. The structure of bovine trypsin (1PPE.pdb) bound with the squash protease inhibitor (SPI) is also included in the figure since SPI contains an Arg at its P1 site. All three structures are superimposed by residues that align the S1 binding pocket in trypsin *i.e.*, residues 189-195 and 219, 221, and 228. Nitrogen, oxygen and sulfur atoms are colored in blue, red, and yellow, respectively. The carbon atoms in residues of WT-trypsin, M84R-trypsin, and SPI-trypsin are colored in gold, green, and mauve, respectively, with main chain carbon atoms being slightly darker than the side chain atoms.

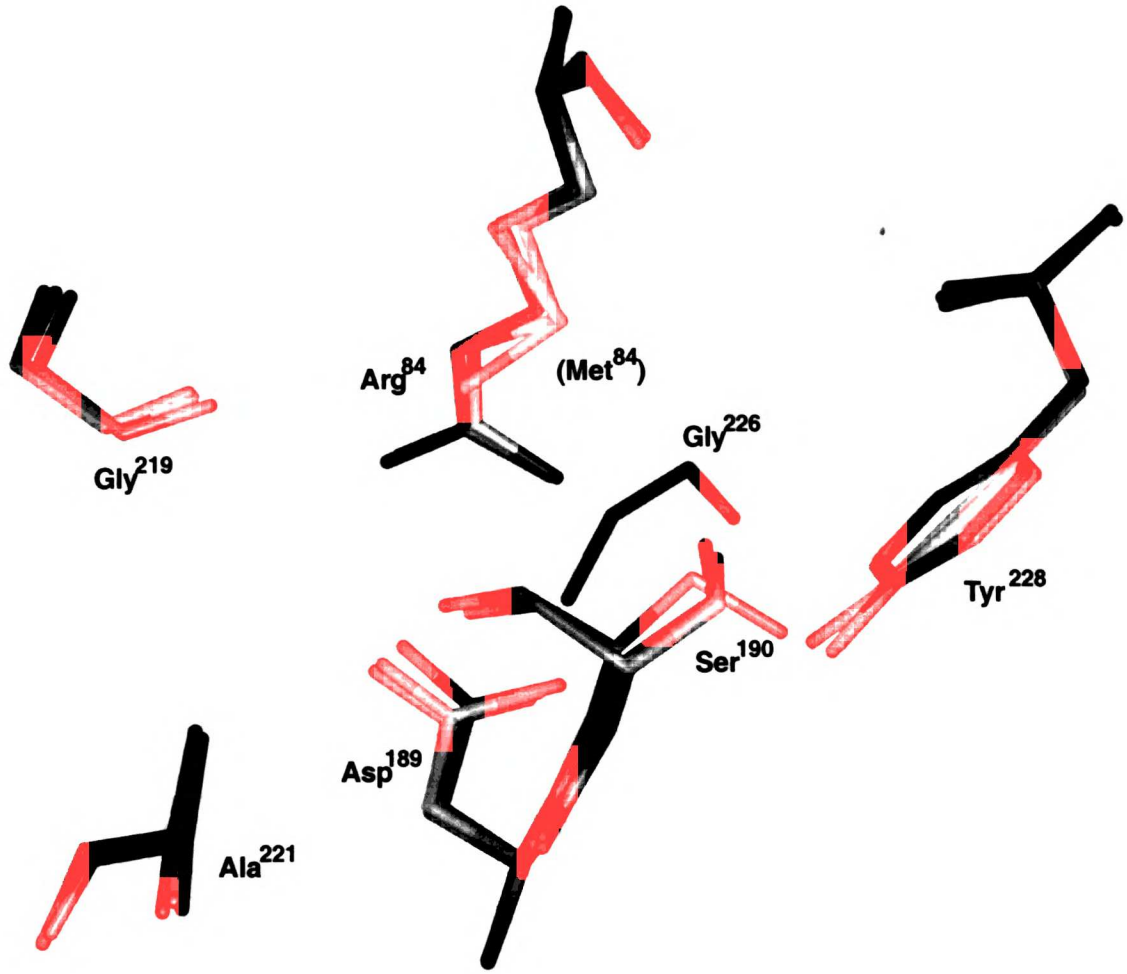


Figure 3-4. Comparison of S1 pocket of trypsin with different substrates

3.3.2.2 X-ray Structure of the Armless/M84R-Trypsin Complex

The ecotin armless/M84R mutant contains both the C-terminal deletion of residues 133 to 142 and the Met to Arg mutation at ecotin's P1 residue. The apparent K_i of this double mutant is ten times stronger than the K_i of wild type ecotin, indicating that the M84R mutation at the P1 residue not only completely rescued the loss of affinity from the C-terminal truncation but also enhanced ecotin's binding to trypsin. The ecotin armless/M84R-trypsin complex, like most ecotin-trypsin complexes, is also crystallized in the space group of $P2_1$. Every asymmetric unit of the unit cell contains only one tetramer. The cell parameters of the armless/M84R-trypsin crystals match those of the armless-trypsin complex (Table 3-2).

3.3.2.2.1 Conformational Changes Between Protein Domains in the Armless/M84R-Trypsin Complex

The structure of armless/M84R-trypsin more resembles the structure of armless-trypsin than the structure of either WT-trypsin or ecotin M84R-trypsin. When ecotin armless/M84R-trypsin is superimposed with the structure of WT-trypsin by the core residues in only one ecotin molecule, the R.M.S.D. is 0.31 Å. This value is similar to the R.M.S.D. between armless-trypsin and WT-trypsin (0.31 Å), but is slightly higher than the R.M.S.D. between M84R-trypsin and WT-trypsin (0.21 Å). When the structure of the armless/M84R-trypsin complex is superimposed with the structure of WT-trypsin

using core residues from both ecotin molecules, the R.M.S.D. increases to 0.63 Å, which is twice of the R.M.S.D. value from superimposing only one ecotin. The structure of each ecotin monomer remains unchanged, but the distance between the two monomers becomes even slightly greater than what has been observed for the armless-trypsin complex.

As expected, both protease domains retain their structural integrity. When compared with the WT-trypsin complex, the protease domains move in manner similar to those observed for the armless-trypsin complex. The COMs of the corresponding protease domains have moved as much as 2.3 Å after the two mutations are introduced in the double mutant. These protease domains also moved over 6 degree towards the ecotin primary site, corresponding to the translational movement (Table 3-3). The positional change propagated by the rotations is as far as 4 Å for residues in certain surface loops of the protease domains. While both armless-trypsin and armless/M84R-trypsin exhibit domain movements with respect to the structure of WT-trypsin, Figure 3-5 further illustrates that the structure of armless/M84R-trypsin diverges more from the structure of WT-trypsin than the structure of armless-trypsin does.

Figure 3-5. Domain movements after C-terminal truncation

This figure highlights the domain movements in the armless/M84R-trypsin structure. Relative conformational change in the ecotin domain is shown after the structures of WT-trypsin, armless-trypsin, and armless/M84R-trypsin are superimposed by all residues (16 to 74, 81 to 245) from one of the protease domain. In order to clearly visualize and compare the conformational change in ecotin, only half of the ecotin-trypsin tetramer is shown here. On the left (A) is a front view of the superimposed structures; on the right (B) is the side view of the same structures with the ecotin domains rotated 90° towards the readers. The structures of WT-trypsin, armless-trypsin, and armless/M84R are colored in gold, navy, and dark red, respectively.

Figure 3-5. Structural Changes In Wild Type, Armless & Armless/M84R

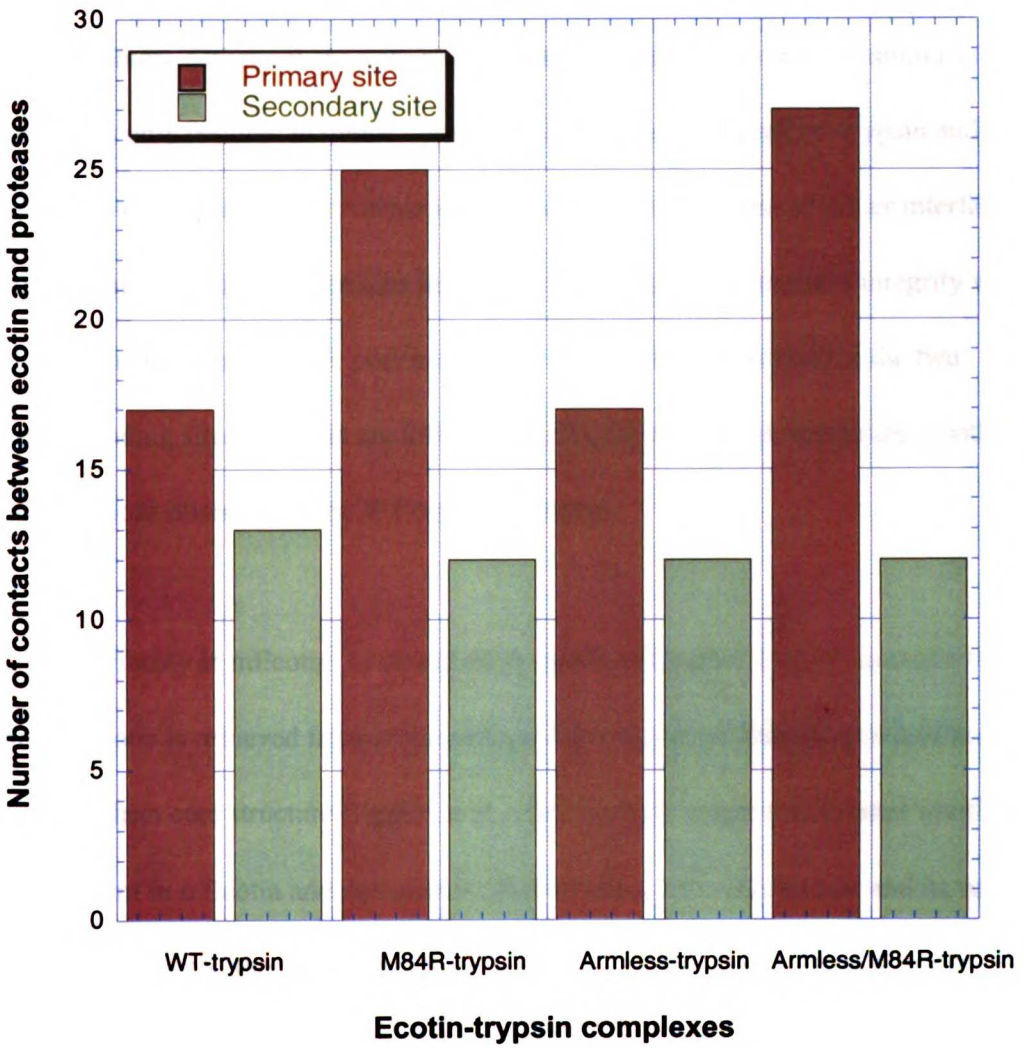
3.3.2.2.2 Detailed Analysis of Protein-Protein Interfaces

The buried surface areas at the two protease-binding sites in armless/M84R-trypsin remain close to those of the WT-trypsin complex (Table 3-3), but the specific interactions at these two sites seem strengthened. At the primary site, residues Ser79 and Arg54 from ecotin form two newly encountered hydrogen bonds with trypsin. At the secondary site, the 60's and 110's loops form up to six well defined hydrogen bonds. At the primary site, the total number of hydrogen bond formed between ecotin and trypsin in the armless/M84R-trypsin complex exceeds the number of hydrogen formed in any other ecotin-trypsin (Figure 3-6). Interactions at the secondary site, however, remain relatively unchanged.

The dimer interface in the armless/M84R-trypsin complex is 855 $Å^2$, even slightly smaller the dimer interface in the armless-trypsin complex. The enhanced primary site interactions from the M84R mutation may have twisted the "hinge" even further, and thus have decreased the dimer interface. Protein-protein interactions at the secondary binding site also seem strengthened in the armless/M84R-trypsin complex (Figure 3-6).

Figure 3-6. Changes of the number of ecotin-trypsin interactions

The numbers of ecotin-protease interactions at both binding sites in WT-trypsin, M84R-trypsin armless-trypsin, and armless/M84R-trypsin are compared in this figure. Primary site interaction is shown by a brick color bar, while the secondary site interaction is shown in light green.


Number of contacts between ecotin and proteases
30
25
20
15
10
5
0
Primary site
Secondary site
WT-trypsin
M84R-trypsin
Armless-trypsin
Armless/M84R-trypsin
Ecotin-trypsin complexes

3.3.3 Structural Changes in Ecotin-Trypsin Complexes and Insights into Ecotin-Trypsin Interactions

Although the C-terminal arm region may be essential for ecotin dimer formation (Eggers et al., 2001; Pal et al., 1996), it appears unimportant for maintaining an ecotin-trypsin tetramer. Structural analyses reveals that deleting the terminal ten amino acid residues has only resulted in minor changes in the structures of armless-trypsin and armless/M84R-trypsin. In both structures, after an up to 70% loss of dimer interface, the ecotin domains and trypsin domains mostly maintain their own structure integrity while slightly adjusting their relative position to each other. The interactions at the two protease-binding sites of ecotin are little affected by the truncation, remaining mostly similar to those observed in the WT-trypsin complex.

Previously in mEcotin, as described in details in Chapter 2, the C-terminal arm region in ecotin is retrieved from its conventional extending conformation to fold back onto its β-sheet core structure (Eggers et al., 2001), which ensure that a dimer interface does not form in mEcotin and its variants. Nevertheless, mEcotin mutant and its variants still form tetramer in the presence of protease targets, which was attributed to the strength of the binding interactions at the secondary site. Similarly, binding interactions at the secondary site of ecotin may again be the reason why an armless-trypsin tetramer does not collapse after the severe truncation.

Structural analyses of the armless-trypsin, M84R-trypsin and armless/M84R-trypsin complexes support our hypothesis that the large dimer interface may be the source for the compromised binding in ecotin-trypsin tetramers. As depicted in Figure 3-5, the truncation at ecotin C-terminus may in fact serve to release the structural rigidity in the ecotin-trypsin tetramer to ultimately allow improved binding at the two protease-binding sites. C-terminal truncation severely shortened the "arm" region, which ultimately results in a much smaller dimer interface, even slightly smaller than the dimer interface in armless-trypsin. Consequently, the ecotin monomers slightly adjust their position to each other, which results in slightly improved binding at the two binding sites.

Figure 3-7. Weakened dimer interface permits more freedom in binding?

This figure depicts our hypothesis of whether a weakened dimer interface in ecotin allows more freedom for protease binding. **A**, the deletion at the ecotin C-terminus leaves the interactions between two armless ecotin molecules a lot less stable than those between two wild type ecotin molecules. The instability may result in more flexibility in the ecotin molecules. **B**, the more "relaxed" ecotin dimer can adapt and bind to proteases with more flexibility. Therefore, more improved binding or even optimal binding at both protease-binding sites may be achieved.

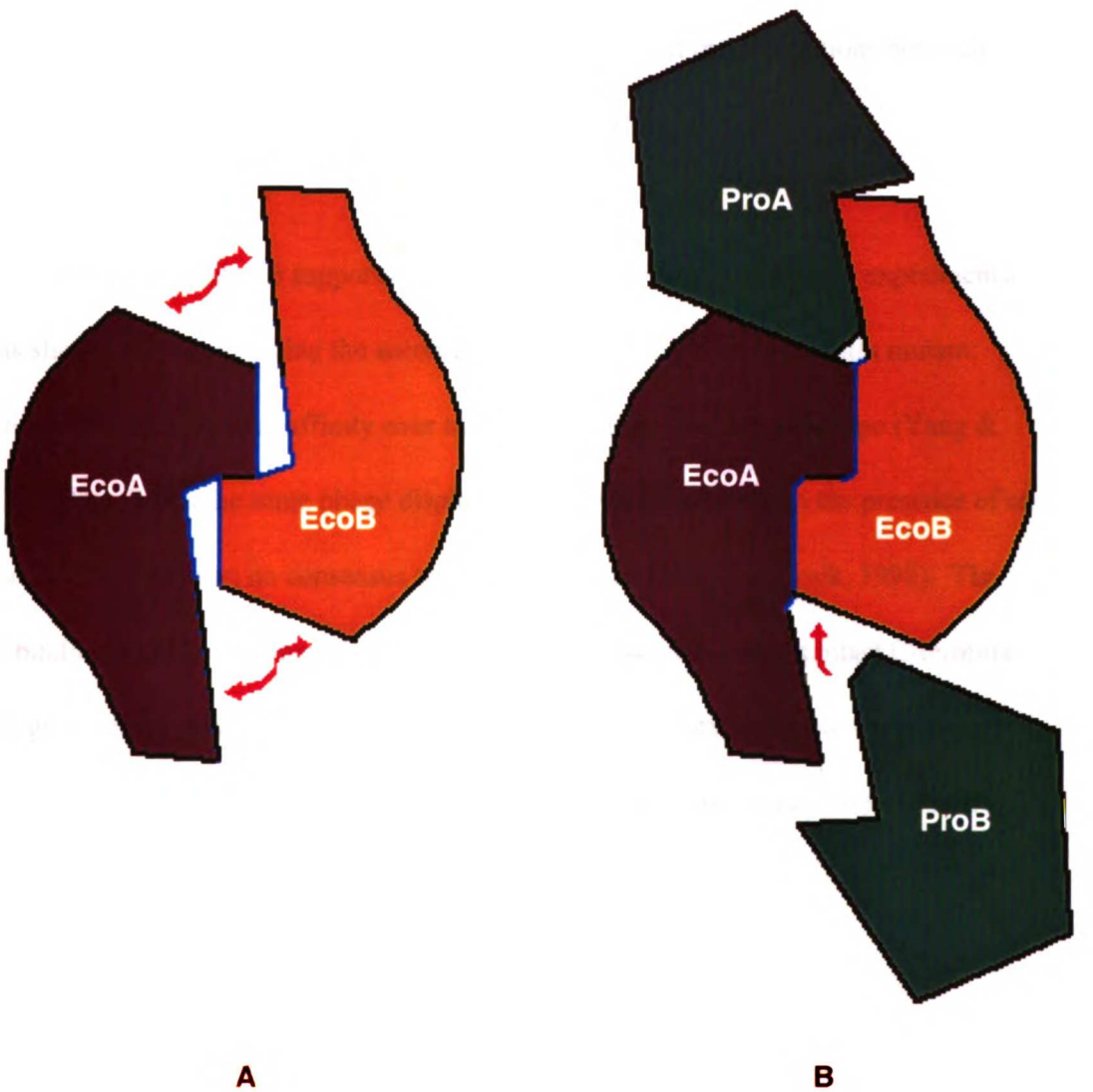


Figure 3-7. Weakened ecotin dimer interface allows better protease binding

In both armless-trypsin and armless/M84R-trypsin, interactions are found between the 50's loop of ecotin and trypsin, which has never been observed in other ecotin-trypsin complexes. It is possible that additional flexibility resulting from the C-terminal truncation may have contributed to the formation of the interactions between 50's loop and trypsin in both armless and armless/M84R.

Additional evidence supporting the model came from phage display experiments. It was shown that randomizing the secondary site sequence yielded an ecotin mutant: Y69F/D70P with a binding affinity over ten times stronger than the wild type (Yang & Craik, 1998). When the same phage display analysis was conducted in the presence of an Arg at P1 site, however, no consensus sequence was found (Yang & Craik, 1998). The tight-binding Arg84 strengthens binding at the primary site. Because an intact C-terminal arm region allows only limited mobility at the 'hinge', tight interactions at the primary site probably leaves less room for further optimization of the secondary site. Thus no consensus sequence could be reached for the secondary site.

3.4 Conclusions and Future Directions

Though maybe import for ecotin dimer formation, the C-terminal arm region appears unimportant for ecotin-trypsin tetramer formation. As suggested by our study, the severely shortened arm region in ecotin likely results in more freedom for protein-protein interactions at the two protease-binding sites. The more flexible armless mutation

may serve as a good starting point for further mutations at either primary or secondary site. If technical problem such as how to display the armless on phage can be worked out, it may even be used in phage display studies to see if consensus sequences with improved binding at both sites can be reached.

Chapter 4

Ecotin and Thrombin

4.1 Introduction

The serine protease α-thrombin is a key enzyme in the processes of thrombosis and haemostasis (Fenton, 1988). It exhibits both pro- and anti- coagulant activities, through converting fibrinogen to clot-forming fibrin (Blomback et al., 1967; Hogg & Blomback, 1978) or through the thrombomodulin-mediated activation of protein C (Esmon et al., 1982; Kisiel et al., 1977). Physiologically, the activities of thrombin are regulated by endogenous inhibitors including antithrombin III (Rosenberg & Damus, 1973), α2-macroglobulin (Stubbs & Bode, 1993), heparin cofactor II (Sheehan et al., 1994), protein C inhibitor (Rezaie et al., 1995) and protease nexins (Gronke et al., 1987). Alternatively, thrombin levels can be regulated by natural and synthetic protease inhibitors such as hirudin, triabin and rhodniin (Lombardi et al., 1999; Stubbs & Bode, 1993). Inhibitors, especially macromolecular inhibitors, have been essential in elucidating the mechanisms of actions of various proteases. The interactions between thrombin and different inhibitors have been studied extensively due to thrombin's biomedical significance.

Ecotin, a 142 amino acid protein discovered in the periplasmic space of *Escherichia coli* (Chung et al., 1983), has the unique property of inhibiting most serine proteases of the chymotrypsin fold, including trypsin, chymotrypsin and elastase (Chung et al., 1983; Seymour et al., 1994). The inhibitory profiles of ecotin against various serine

proteases share three common features (McGrath et al., 1994; Perona et al., 1997): 1) Ecotin dimerizes and inhibits cognate proteases at 1:1 stoichiometry. Crystal structures of ecotin-trypsin, ecotin-collagenase and ecotin-chymotrypsin all show the formation of a tetramer containing two ecotin and two protease molecules. 2) Ecotin inhibits proteases via binding at two different protease contact sites: the primary and secondary binding sites. Primary site loops of ecotin bind to the active site of target proteases in a substrate-like manner as observed in canonical protease inhibitors such as BPTI (Perona et al., 1993). The P1 residue in ecotin (Met84) mimics the interactions of a canonical P1 substrate residue. Unique secondary site loops of ecotin bind to a relatively flat surface of the protease that is distant from the active site. 3) The protein-protein interaction surfaces are extremely large in ecotin-protease complexes; the sum of the buried surface area is around 6000 $Å^2$ (Gillmor et al., 2000; Perona et al., 1997).

4.1.1 Previous Studies on Thrombin

Despite its apparent pan-specificity against chymotrypsin family serine proteases, ecotin fails to inhibit thrombin (Seymour et al., 1994). Thrombin contains nine extra amino acid residues inserted at residue 60 in comparison to the sequence of chymotrypsin (Blomback et al., 1967). This insertion loop contains bulky residues like Pro, Trp and Tyr, and appears relatively rigid forming a lip above the active site (Bode et al., 1989). Modeling analysis shows that Trp^{60D} (thrombin residue numbers are labeled as superscripts to distinguish them from ecotin residues) clashes with residues Cys50 and

Asn51 in the e50's binding loop of ecotin (residue numbers specifying the binding loops in ecotin follow the letter e to differ from thrombin). Previous studies showed that the insertion loop at thrombin residue 60 blocked inhibition by bovine pancreatic trypsin inhibitor (BPTI) (Le Bonniec et al., 1993). In that study, a thrombin desPPW mutant, created by removing three bulky rigid residues Pro^{60B}-Pro^{60C}-Trp^{60D}, permitted BPTI inhibition with a nanomolar Ki (Le Bonniec et al., 1993). While a single mutation of Met84 to Arg only improved ecotin binding to trypsin by three folds (Yang et al., 1998), the same mutation increased ecotin's inhibition against thrombin by $2x10^4$ times (Seymour et al., 1994). The inhibition constant (Ki) changed from approximately 30uM for wild type ecotin to 1.5nM for the M84R mutant (data not shown). Deleting the residues that cause steric hindrance in wild type thrombin explains BPTI's inhibition against the desPPW thrombin mutant (Le Bonniec et al., 1993). Even though thrombin clearly prefers an Arg at the P1 site, this crucial residue is only effective when bound to the active site (Laskowski & Qasim, 2000). Since ecotin access to the active site appears to be blocked by Trp^{60D}, it is difficult to explain why the Met84 to Arg mutation helps ecotin to inhibit thrombin.

4.1.2 Regulatory Proteases and Extended Interactions

Like other regulatory proteases, the substrate specificity of thrombin extends beyond the P1 residue: other amino acids flanking the cleaved scissile bond are also important for the substrate recognition of thrombin (Martin et al., 1996; Stubbs et al.,

1992). In comparison to digestive proteases, regulatory proteases often contain insertions of amino acids in the surface-located loops surrounding the active site. The characteristics of these loops are believed to be important for substrate recognition by the regulatory proteases (van de Locht et al., 1997). Details of the extended specificity of thrombin have been studied *in vitro* using synthetic substrates (Backes et al., 2000; DiBella & Scheraga, 1996; Le Bonniec et al., 1996; Rezaie & Olson, 1997). These data define the cleavage site to be L/I/V/F-X-P-**R**-non-polar/hydrophobic-hydrophobic/charged, from P4 to P2' (X stands for any amino acid, P1 residue Arg is bold). Fibrinogen peptides are cleaved at GGV**R**-GP and FSA**R**-GH. Interactions at P4-P2' sites can also be visualized by combining information from X-ray structures of many thrombin complexes with inhibitors or peptides (Bode et al., 1989; Martin et al., 1992; van de Locht et al., 1995; van de Locht et al., 1996). Among existing exogenous inhibitors, ecotin binds proteases over the largest interface, which may extend beyond the known interaction site for thrombin.

In order to elucidate how ecotin M84R and thrombin form a complex, we have determined the structure of the ecotin M84R-thrombin complex to 2.5 Å resolution.

4.2 Material and Methods

4.2.1 Gel Filtration and Gel Electrophoretic Studies

Purified bovine thrombin and ecotin M84R were prepared separately as described previously (McGrath et al., 1991; Owen et al., 1974). Thrombin-ecotin M84R was formed by incubating 1:1 molar ratio of each component and subsequently purified on a superdex-200 size exclusion column (Pharmacia, Germany) in 50mM Potassium Phosphate, pH 8.0 and 0.15M NaCl. Three peaks were isolated with calculated molecular weights of 33.4kDa, 64kDa and 99kDa, respectively. All three peaks were subject to both denaturing and native gel electrophoretic analyses. All peaks contain both bovine thrombin and ecotin M84R (gel picture) based SDS-PAGE gel analysis. Based on the molecular weight information, the first peak probably contains a mixture of bovine thrombin and dimeric ecotin; the second peak seems to contain a meta-stable complex of thrombin-M84R with a stoichiometry of 1:2; the third peak contains a complex of thrombin-M84R with stoichiometry of 2:2. The tetrameric complex from the third peak seems to be more stable than runs as a single band by native gel analysis.

4.2.2 Crystallization studies and Data Collection

The thrombin-ecotin complex isolated by gel filtration was used in the crystallization experiments. Orthorhombic crystals were grown at room temperature by hanging drops using the vapor diffusion technique, from 15% Polyethylene Glycol 6K,

0.1M Citric Acid pH 5.0 with 0.01M spermine-tetrahydrochloride. Three diffraction data sets were collected at Advanced Light Source (ALS) beam line 5-2 with a CCD camera and at Stanford Synchrotron Radiation Laboratory (SSRL) beam line 9-1 with a MAR image plate system. The data sets were evaluated and integrated using SCALEPACK/DENZO (Otwinowski & Minor, 1997). The crystal has a space group of $C222_1$. One asymmetric unit contains one thrombin and one ecotin.

4.2.3 Modeling and Structure Determination

A hetero-dimer model of bovine thrombin (1UVT.pdb) bound to ecotin at the protease active site was generated using InsightII (Biosym, Inc.), based on the trypsin-ecotin structure (McGrath et al., 1994). The 60's and 148's insertion loops were deleted from thrombin. The resulting structure was used as the search model to solve the structure of thrombin-ecotin M84R, using molecular replacement with rotational and translational functions from CNS 1.0 (Brunger et al., 1998). Data to 3.5 Å were used in the rotational search. Translational search and rigid body fitting for the top solutions using data to 2.5 Å gave a final solution with an initial R value of 43.5%. Refinement was performed using programs from CNS 1.0 (Brunger et al., 1998). Detailed data and refinement statistics are listed in Table 1 (PDB entry ID:1ID5). Visualization and rebuilding of the model was done with Quanta98 (Molecular Simulations Inc., San Diego CA).

4.3 Experimental Results

4.3.1 Analysis of Bovine Thrombin & Ecotin M84R by Gel Filtration

Following incubation on ice, a mixture of bovine thrombin and ecotin M84R at 1:1 stoichiometry gives rise to a complex which can be isolated by gel filtration method. By using gel filtration standard, the apparent molecular weight of the isolated complex is calibrated to be ~100 kDa. The calculated molecular weights of ecotin and bovine thrombin are16.1 kDa and 39.8 kDa, respectively. The apparent molecular weight of the isolated peak corresponds to the sum of the molecular weights of two ecotin molecules and two thrombin molecules, which is what we expect based on previous knowledge about the interactions between ecotin and target serine proteases.

Another gel filtration experiment is conducted for wild type ecotin and bovine thrombin. Under the same conditions for gel filtration experiment, only one peak is isolated with an apparent molecular weight of around ~35 kDa. This apparent molecular weight corresponds to that of bovine thrombin and as well as the molecular weight of ecotin dimer molecules since ecotin has been shown to exist as dimer under high concentration (Yang et al., 1998).

4.3.2 Analysis of Thrombin-M84R Complex by Native Gel

To fully determine the component of the isolated fractions from gel filtration experiment, the isolated peaks are further analyzed by gel electrophoretic experiments, under both denaturing and native conditions. SDS-PAGE gel analysis indicates that the complex with apparent molecular weight of ~100 kDa contain both thrombin and ecotin M84R. When run on a native gel, the same sample runs as a single band that is not found in either thrombin or ecotin M84R sample alone, indicating the formation of a stable complex between thrombin and ecotin M84R mutant. In the peak isolated from wild type ecotin and thrombin mixture, which has an apparent molecular weight of ~35 kDa, both ecotin and thrombin are observed by SDS-PAGE gel. This is because thrombin and dimeric ecotin have similar molecular weights that cannot be resolved by a superdex-200 gel filtration column, so both proteins elute out in the same peak. This conclusion is confirmed by native gel analysis because no new band is observed except the ones corresponding to thrombin and wild type ecotin. All evidences support the formation of a 2thrombin-2ecotin M84R tetrameric complex. No such complex is formed when mixing bovine thrombin and wild type ecotin under the same conditions.

4.3.3 Data Processing and Structure Determination

The data set was 97.7% complete to 2.5 Å and 88.7% complete in the highest shell (2.50 to 2.59 Å). There are a total of 239796 measurements of 20967 reflections where mean I/I_σ is 13.3 and R_{symm} to 2.5 Å is 10.8% (merging three data sets together).

Processing statistics indicated that the crystal has a space group of $C222_1$ with unit cell parameters of a = 88.5 Å, b = 165.4 Å, c = 83.3 Å, $\alpha = \beta = \gamma = 90$ degree. There is only half of the complex in an asymmetric unit, containing one thrombin and one ecotin molecule. Solvent content in crystal is 55%. Details of data statistics are listed in Table 4-1.

Table 4-1. Data and refinement statistics for ecotin M84R-thrombin complex

Structure	M84R-Thrombin
Space group	$C222_1$
Unit cell constants (Å)	
a	88.5
b	165.4
c	83.3
$\alpha = \beta = \gamma$	90°
Highest resolution (Å)	2.5
Total reflections	239796
Unique reflections	20967
Completeness	97.7%
Highest resolution shell	88.7% (2.59 Å to 2.50 Å)
I/I_σ	13.3
R_{merge} [a]	10.8% (three data sets)
Highest resolution shell	29.8% (2.59 Å to 2.50 Å)
Solvent content	55%
R [b] after molecular replacement	43.5%
Refinement resolution range	6–2.5 (Å)
R	19.5%
free R [c]	25.3%
water molecules	158
r.m.s.d. in bond length (Å)	0.006
r.m.s.d. in bond angle (°)	1.3

[a] $R_{merge} = \Sigma|(I - \langle I \rangle)| / \Sigma(I)$

[b] $R = \Sigma_{h,k,l}(|F_{obs}(h,k,l)| - k|F_{calc}(h,k,l)|)^2/\Sigma_{h,k,l}|F_{obs}(h,k,l)|$

[c] free R: Cross-validation R calculated by omitting 10% of the reflections (Kleywegt & Brunger, 1996).

Figure 4-1. X-ray structure of ecotin M84R-bovine thrombin

Structure of bovine thrombin complexed to ecotin M84R in secondary structures. Light chain (L) and heavy chain (H) of thrombin are shown in green, with catalytic triad in red. Ecotin is shown in purple; disulfide bonds and Calcium ion are highlighted in yellow and orange. The symmetry mate that helps to construct a full tetrameric complex is shown in gray. The figure was generated by Raster3D (Merritt & Bacon, 1997), with secondary structures rendered by Rasmol (Sayle & Milner-White, 1995).

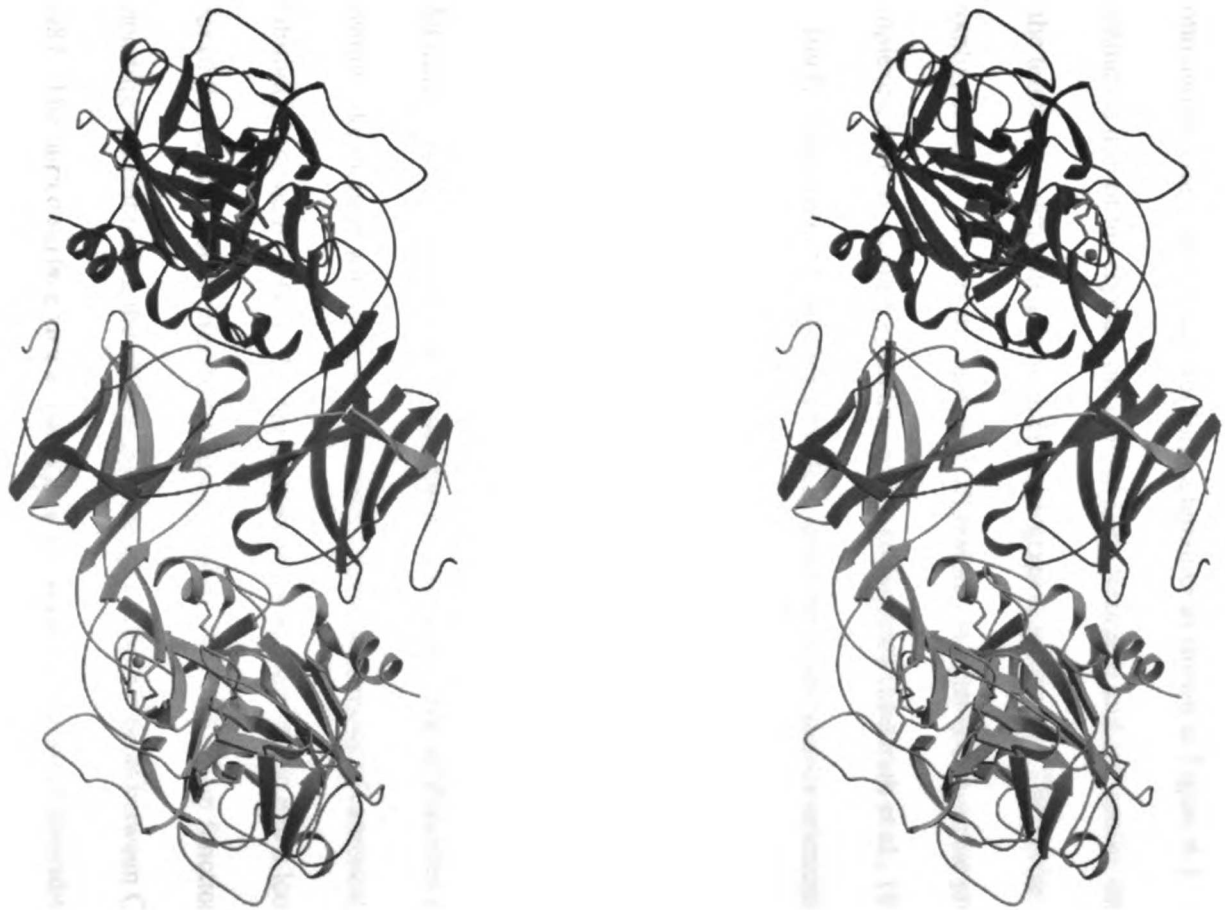


Figure 4-1. X-ray structure of thrombin-M84R

4.3.4 Structure of the Bovine Thrombin-Ecotin M84R Complex

Ecotin M84R-thrombin complex is crystallized in the $C222_1$ space group. Two ecotin and two thrombin molecules form a tetramer as shown in Figure 4-1. One asymmetric unit of the crystal contains only one thrombin and one ecotin; the other half of the tetramer is generated through crystallographic symmetry. The ecotin M84R-thrombin complex exhibits the same overall features as other ecotin-serine protease complexes, such as ecotin-trypsin and ecotin-collagenase (McGrath et al., 1994; Perona et al., 1997). The ecotin M84R dimer binds to thrombin in the similar orientation as wild type ecotin does in the ecotin-trypsin structure, contacting thrombin with two binding sites.

The primary binding site of ecotin M84R includes the e50's and e80's loops (McGrath et al., 1994). The e80's loop binds to the active site of thrombin in a substrate-like manner with characteristic main chain interactions of canonical protease inhibitors (Bode et al., 1992). A total of eight residues from ecotin e80's loop, on both sides of the scissile bond, directly contact thrombin. The e50's loop functions by supporting and stabilizing the e80's loop through a disulfide bond between Cys50 and Cys87. The buried surface area at the primary contact area between thrombin and ecotin is 2290 $Å^2$, which is larger than any other known protease-ecotin interaction site (McGrath et al., 1994; Perona et al., 1997). The secondary binding site, mainly consisting of the e60's and e110's loops, binds to the 99's loop and the C-terminal helix of thrombin

that are about 27 Å away from the active site. The total buried surface area at the secondary binding site is 1250 $Å^2$, also larger than any other secondary site in known ecotin-protease complexes (McGrath et al., 1994; Perona et al., 1997).

Compared with the structure of uncomplexed ecotin dimer (Shin et al., 1996), the two ecotin molecules adjust their relative position slightly upon binding to thrombin. However, they do not change their conformations on inhibiting thrombin as they do on inhibiting collagenase (Perona et al., 1997). The structure of ecotin M84R mutant exhibits few significant differences from the structure of wild type ecotin when bound to trypsin (McGrath et al., 1994) or in its unbound form (Shin et al., 1996), with r.m.s.d. of 0.39 Å and 0.48 Å for superimposing the conserved core structure that includes residues 12-43, 59-63, 70-76 and 103-131(Perona et al., 1997).

Figure 4-2. Electron density of the 60's insertion loop in thrombin

Electron density ($2F_{obs}$-F_{calc}) of the 60's insertion loop from the bovine thrombin-ecotin M84R complex contoured at 1σ. Density for the side chains of Trp^{60D} and Asn^{60G} is missing. The figure was produced using Raster3D (Merritt & Bacon, 1997).

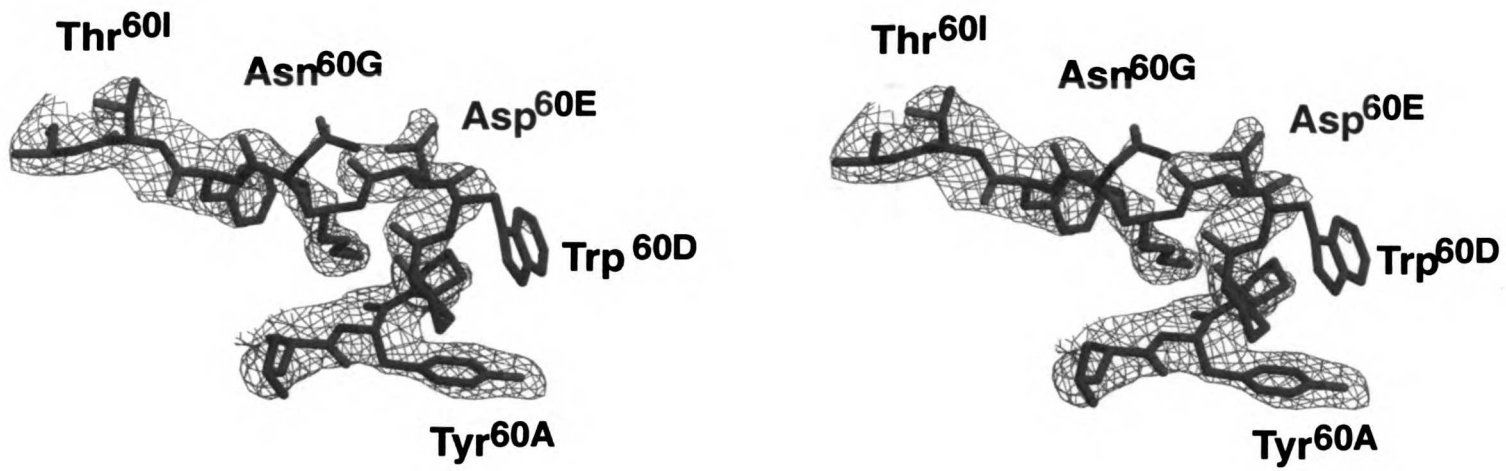


Figure 4-2. Electron density map of the 60's loop.

4.3.4.1 The thrombin structure

The overall thrombin structure in the complex maintains the classic morphology of serine proteases of the chymotrypsin fold (Bode et al., 1989). The globular structure of thrombin consists mainly of β strands, with only a few helical segments and various surface-located loops. The catalytic residues: His^{57}, Asp^{102} and Ser^{195} are located in the junction cleft formed by the two beta barrel domains of thrombin. The conformations of the catalytic triad and surrounding regions are well preserved (Bode et al., 1989). Our thrombin core structure, excluding the light chain and surface loops that include the 60's (Leu^{60} to Thr^{60I}), the 148's (Trp^{148} to Glu^{149E}), the 37's (Arg^{35} to Glu^{39}) and the 99's insertion loops (Arg^{93} to Leu^{99}), compares with published coordinates with r.m.s.d. smaller than 0.4 Å for over 160 equivalent C_α atoms. The four surface loops, however, have undergone significant movements to permit the binding to ecotin M84R.

The 60's insertion loop contains a fragment of nine amino acids with the sequence of Y^{60A}PPWDKNFT^{60I} (according to chymotrypsin nomenclature) (Stubbs & Bode, 1993). This loop is unique to thrombin with its amino acid sequence highly conserved across different species (Banfield & MacGillivray, 1992), indicating its significance in thrombin substrate recognition. Previous structural analyses have shown that the 60's insertion forms a rigid loop above the active site, and it exhibits little conformational change upon inhibitor binding (Bode et al., 1989; Martin et al., 1992). This loop's

clashing into the e50's loop from ecotin has been hypothesized to account for wild type ecotin's failure to bind thrombin. In our structure, the 60's loop moves significantly away from the substrate-binding cleft. The side chains of Trp^{60D} and Asn^{60G} become disordered upon ecotin binding (Figure 4-2). Moreover, the C_α of Trp^{60D} moves nearly 7 Å compared to the position of the same atom when thrombin is bound with small peptide inhibitors (Figure 4-3a). The apparent rigidity of the 60's loop suggests that there may be a significant energy cost to promote movements of this loop.

Figure 4-3. Conformations of surface loops in thrombin

Comparisons of conformations of **a**) the 60s, **b**) the 148's and **c**) the 99's **d**) the 37's loops bound with different inhibitors. Bovine thrombin structures, 1ETR.pdb (green, bound to small inhibitor 2MQPA), 1BBR.pdb (blue, bound to fibrinopeptide α2 7-16), 1TBR.pdb (yellow, bound to rhodniin), 1TOC.pdb (purple, bound to ornithodorin), 1UVT.pdb (gold, bound to small inhibitor MB14.1248) and the ecotin bound thrombin molecule (red), are superimposed based on core residues. Trp^{60D} in all structure in **a**) are shown in full stick model. The conformations of all loops are shown in ribbon diagram. All figures were generated with InsightII (Accelrys, San Diego, CA.).

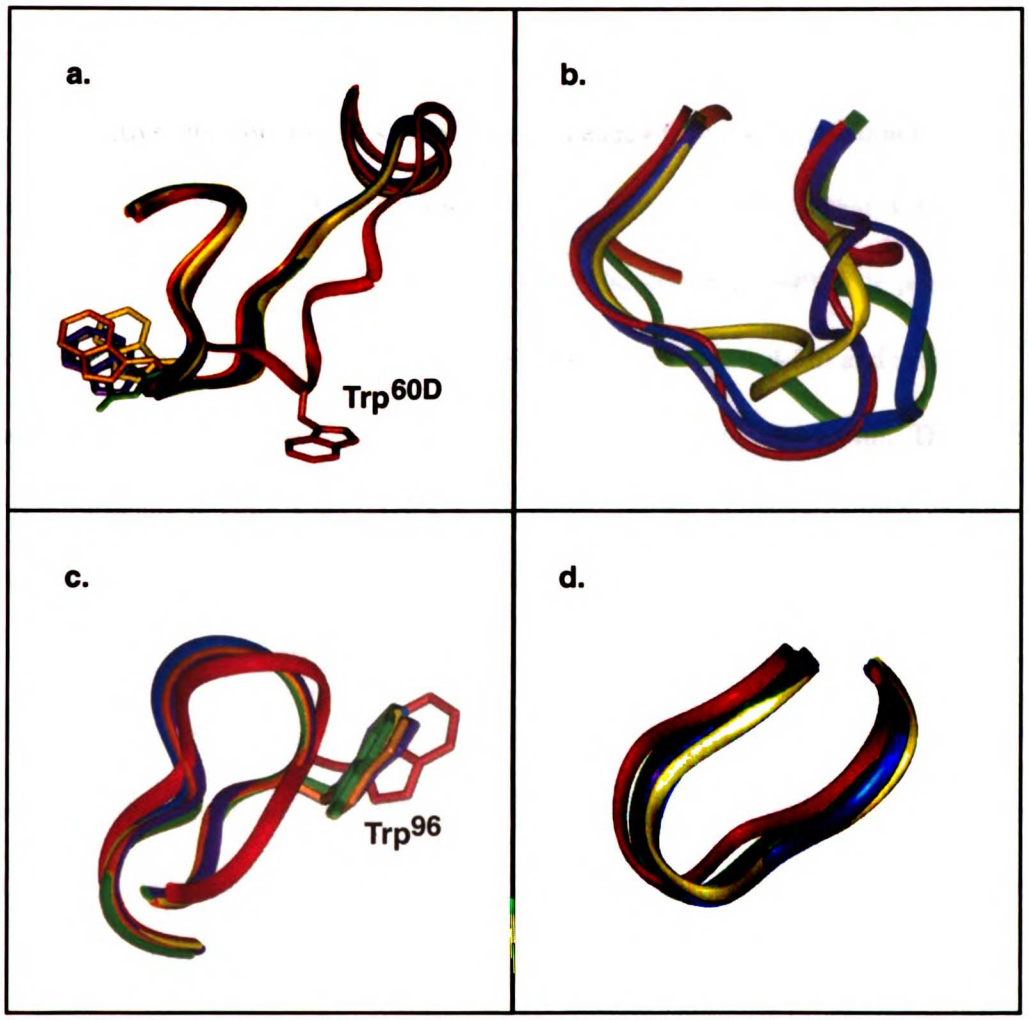


Figure 4-3. Comparison of thrombin surface loops when bound to different ligands

The 148's loop also moves significantly upon ecotin binding (Figure 4-3b), compared to the conformations when bound to small inhibitors. This loop contains a shorter insertion of $T^{149A}SVAE^{149E}$ (Stubbs & Bode, 1993), and is located on the opposite side of the active site from the 60's loop. The sequence of the 148's insertion is not as conserved as that of the 60's loop (Banfield & MacGillivray, 1992), neither is its conformation (Bode et al., 1989; Grutter et al., 1990; Martin et al., 1992). In our structure, the 148's loop adopts an extended conformation (Figure 4-3b) and is slightly disordered. No direct contacts are formed between the 148's loop and ecotin. This loop seems intrinsically flexible, suggesting a small energy cost for its conformational change.

The 37's and the 99's loops, two surface loops adjacent to the 60's insertion loop, also show minor conformational change upon ecotin binding (Figure 4-3a & 3b), when compared to the conformations of these two loops when bound to other inhibitors (Bode et al., 1989; Martin et al., 1996; Martin et al., 1992; Slon-Usakiewicz et al., 2000; van de Locht et al., 1997). In particular, Trp^{96}, a residue important for the catalytic activity of thrombin (DiBella & Scheraga, 1998), forms a hydrogen bond with Tyr100 from ecotin. Moreover, the side chain of this residue clearly differs from its conventional position in other structures (Figure 4-3a). Previous studies showed that the 37's and the 99's loops are involved in substrate binding at P2 and P1' sites (He et al., 1997; Le Bonniec et al., 1996). The small changes in the positions of these loops may result in size changes of the S2 and S1'substrate binding pockets, therefore allowing different amino acid recognition.

Figure 4-4. Extended interactions between thrombin and ecotin M84R

Details of the primary binding site interaction between bovine thrombin and ecotin M84R mutant are shown in **a)** a two-dimensional plot and **b)** a three-dimensional surface representation of thrombin surface. In **a)**, residues 79 through 86 in ecotin are laid out vertically as the P6 to P2' site. Residues from thrombin that make hydrogen bonds (linked by green dash line with distance labeled in red) or hydrophobic contacts (in brown with residues labeled in black) are drawn in each side of ecotin ligand. In **b)**, the same fragment from ecotin is shown in stick model horizontally from P6 to P2', with main chain atoms in navy and side chain atoms in blue. Thrombin is shown as a transparent surface with its C_α trace shown as green ribbon. Side chains of the residues directly contributing in ecotin binding are shown in stick mode in yellow. The catalytic triad is in red. **4a** and **4b** are generated using Ligplot (Wallace et al., 1995) and InsightII (Accelrys, San Diego, CA.), respectively. **4c** lists the cleavage sites of some physiological substrates of thrombin. Non-polar residues are in black, polar residues and Gly are in green, positively charged residues are in red and negatively charged residues are in blue.

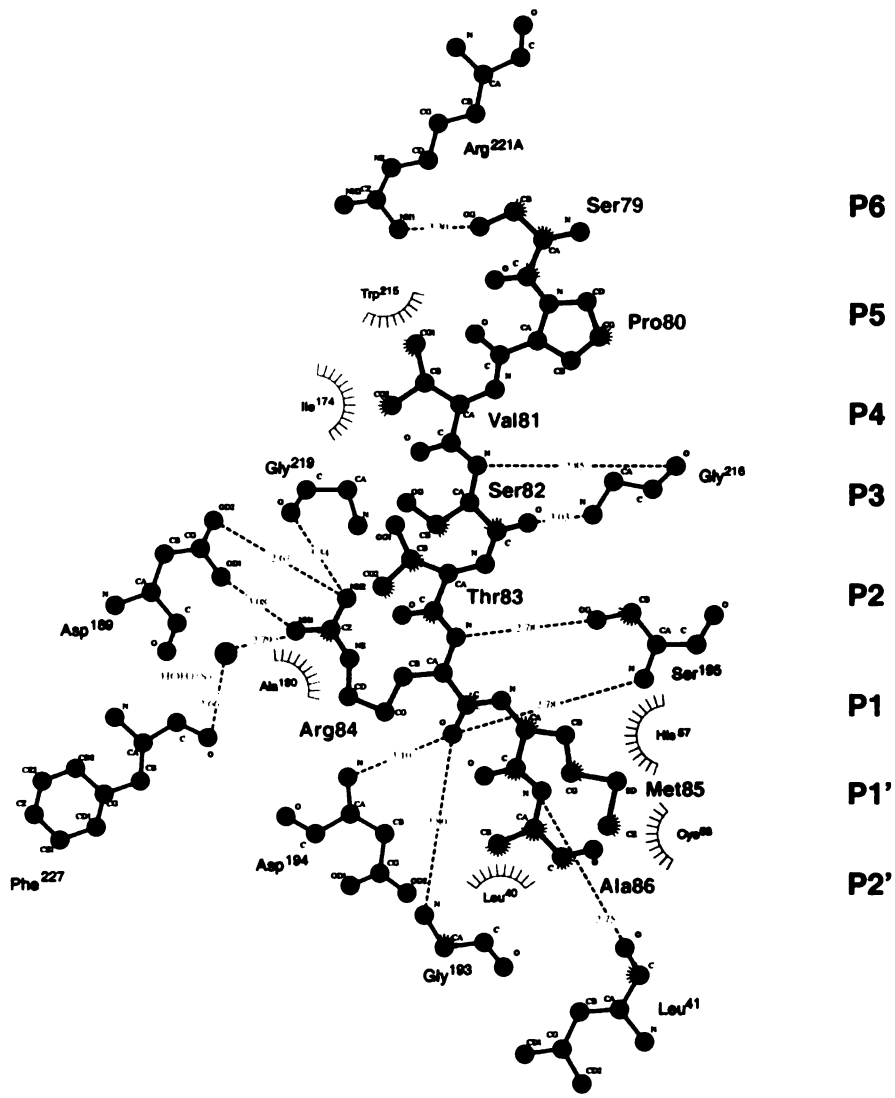


Figure 4-4. Extended interactions (a)

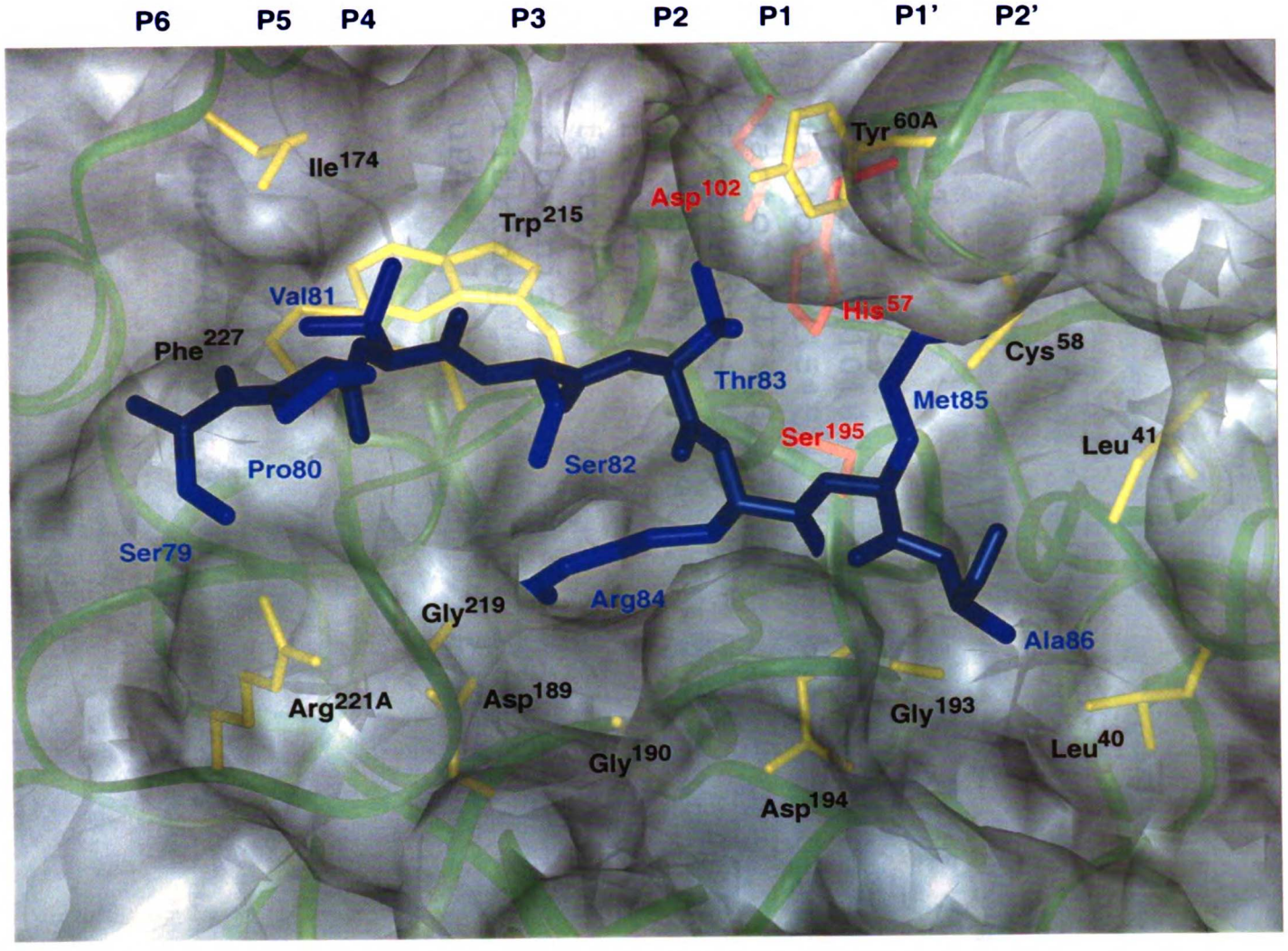


Figure 4-4. Extended interactions (b)

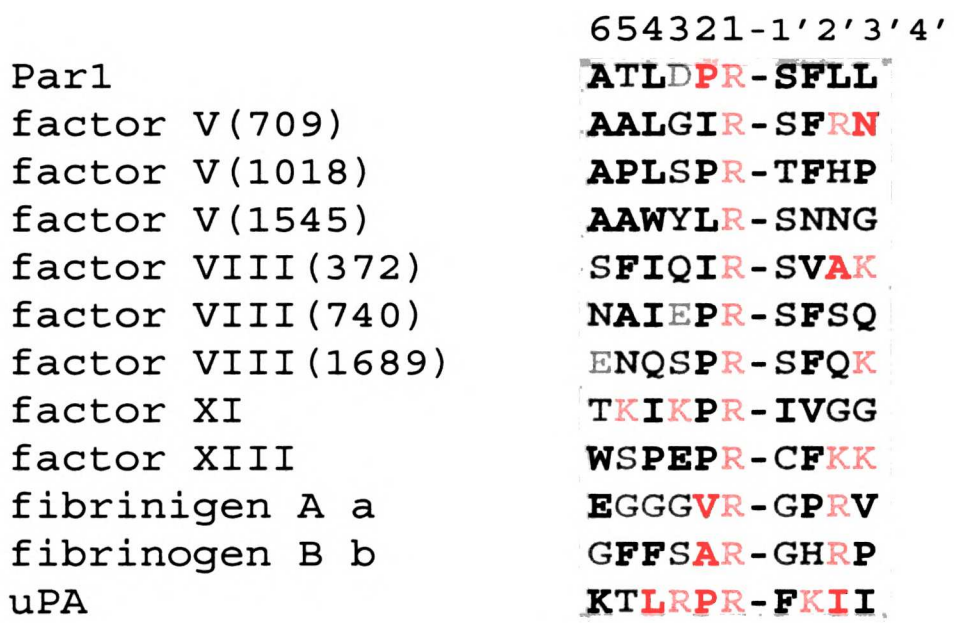


Figure 4-4. Extended interactions (c)

4.3.4.2 Extended Interactions between the Active Site of Bovine Thrombin and Ecotin M84R

The primary site e80's loop of ecotin M84R interacts extensively with the active site of thrombin. Though not optimized for recognition (except for the Met to Arg mutation), residues Ser79 (P6) to Ala86 (P2') from ecotin span across the active site cleft of thrombin, providing a platform for predicting amino acid preference at each substrate recognition site of thrombin.

Arg84, the only mutated residue in ecotin, fits exceptionally well into the S1 binding pocket of thrombin. Arg84 is the ecotin residue that forms most interactions with thrombin: it makes five direct hydrogen bonds and a salt bridge with five residues in thrombin in addition to two water-mediated hydrogen bonds and various hydrophobic interactions (Figure 4-4a). Gly^{193}, Asp^{194} and Ser^{195} coordinate with the main chain –N and the carboxyl –O of Arg84. Asp^{189}, which lies at the bottom of the S1 pocket, contributes to Arg side chain recognition by forming a salt-bridge. Gly^{219} makes an additional hydrogen bond with the terminal $-N_{\eta 2}$ of Arg84. Gly^{216} and Phe^{227} form two water-mediated hydrogen bonds with Arg84 side chain. Furthermore, Ala^{190} and Val^{213} coordinate the aliphatic side chain of Arg84 through hydrophobic interactions. Most interactions established between the side chain of Arg84 and thrombin could not be formed if this residue were a Met. Additional energy from these interactions may be essential for the change in inhibition profile.

It is not obvious whether Thr83 is the optimized residue at the P2 position for a thrombin inhibitor. The S2 pocket appears to have a predominately hydrophobic character. Leu^{99} and Tyr^{94} set the boundary in one direction; Trp^{215} and Tyr^{60A} further define this volume in the other direction. Trp^{96} forms a lid to close the binding pocket. As shown in Figure 4-4a and 4-4b, the S2 pocket seems small enough to exclude large residues such as Phe, Trp, Arg and Lys. *In vitro* studies with positional scanning libraries of fluorogenic peptides have shown that the S2 site of thrombin exhibits a strong preference for Pro (Backes et al., 2000). The preference for a Pro at the P2 position is well supported by the structure of human α-thrombin with D-Phe-Pro-Arg-chloromethylketone (PPACK) (Bode et al., 1989). In this structure, the 60's insertion loop is right above the active site and creates a small rigid pocket that seems to be selective for a Pro. Furthermore Pro is found as the P2 residue in many natural substrates or inhibitors of thrombin (Figure 4-4c). Although the thrombin-ecotin M84R structure indicates that the restriction by the 60's loop no longer exists upon the binding of certain macromolecules, Pro may still be favored as the P2 residue for having the additional role of providing rigidity to the substrate backbone to enhance extended binding.

Ser82 occupies the P3 position in ecotin. Two hydrogen bonds with Gly^{216} well coordinate the main chain –N and carboxyl –O in Ser82. Two water-mediated hydrogen bonds are formed between its side chain and main chain –N of Glu^{192} and carboxyl –O of Gly^{219}. There is no obvious preference for a Ser residue. As the S3 pocket is partially

solvent exposed, a variety of residues can be docked in the S3 pocket such as Asp, Lys and Arg. Conversely, the P4 position is much more selective with Val81 well recognized by thrombin. A local hydrophobic environment, which is limited in size, explains the preference for a Val at this position. As depicted in Figure 4-4b, Trp^{215} paves the base of the pocket and Ile^{174} makes significant hydrophobic contact with Val83. $Met^{180}$and Leu^{99}, not shown in the figure, also fortify a well-formed hydrophobic binding pocket suitable for a Val, Ile or Leu. Our observations at S3 and S4 substrate binding sites are consistent with *in vitro* studies with combinatorial substrate libraries (Backes et al., 2000) and with the sequences of some physiological substrates of thrombin (Figure 4-4c).

The role of Pro80 at the P5 position appears to be to create a rigid turn so that Ser79 at the P6 position can form a favorable hydrogen bond interaction with the side chain of Arg^{221A} from thrombin (Figure 4-4a & 4b). Lys^{224} is also in the vicinity of Ser79. Ser79 is the furthest residue N-terminal to the scissile bond that directly contacts thrombin; its side chain –OH is solvent exposed.

Our structure also reveals ecotin and protein interactions C-terminal to the P1 residue because ecotin stops the catalysis of serine protease prior to the cleavage of the scissile bond (McGrath et al., 1994). Figure 4b shows that the P1' residue Met85 takes an extended conformation in binding the substrate binding pocket, underneath the insertion loop at residue 60 of thrombin. The side chain of Met85 is sandwiched between two disulfide bonds: Cys^{42}-Cys^{58} from thrombin and Cys50-Cys87 from ecotin. Tyr^{60A}

defines the bottom of the S1' pocket. Pro^{60B}, Phe^{60H} and the aliphatic side chain of the buried Lys^{60F} are all within 6 Å to Met85. Leu^{41} and His^{57} further enhance the hydrophobicity of the S1' pocket. The P2' site is occupied by Ala86, the last ecotin residue in the e80's loop that directly contacts thrombin. Its main chain –N forms a hydrogen bond with the carboxyl –O in Leu^{41}. Its methyl side chain interacts extensively with the aliphatic side chain of Leu^{40}. A Val, Leu or Ile may also be well accommodated because of the local hydrophobicity at the S2' site. However, the S2' pocket is only partially hydrophobic. In the direction toward the exposing surface of the protease, Arg^{73}, Asn^{143}, Gln^{151} and Glu^{192} are within 8 Å of Ala86. Therefore, a charged residue with long side chain, such as an Arg or a Lys can possibly go through the local hydrophobic area and establish interactions with these polar residues.

The geometric vicinity of the S1' and S2' sites creates a combined specificity; i.e. residue bound in S1' affects the choice of residues at S2' site. When the S1' site is bound with a small non-charged residue, the S2' site can accommodate either a large hydrophobic residue such as a Phe to fill the large hydrophobic space, or a small hydrophobic or non-charged residue followed by a similar residue at the S3' site. In the presence of a large hydrophobic P1' residue such as a Phe, the S2' pocket will be limited to a large charged residue like a Lys, as found in uPA (Figure 4-4c). An Arg may also fit; other alternatives include small non-charged residues such as Cys, Ser, Ala or Gly.

4.4 Discussions and Conclusions

4.4.1 Thrombin Selects Ligand By Its Active Site & Surface Loops

A simple docking and fitting model does not explain the differences in interactions of thrombin with wild type or ecotin M84R mutant. The conformations of thrombin likely shift between a 'closed' state and an 'open' state; only in the 'open' state, the active-site-restricting surface loops move away to permit ligand binding to thrombin. It is possible that a transient complex between thrombin and ecotin M84R exists prior to the stable final complex we observe in the static X-ray structure. The major energetic contribution for the formation of the final ecotin M84R-thrombin complex comes from the numerous interactions the Arg side chain establishes with thrombin (figure 4b). A similar transient complex also likely exists between thrombin and wild type ecotin, but it lacks of the interaction affinity for the transition to a more stable complex that can then be isolated by gel filtration. Conformational changes, especially those in the surface loops surrounding thrombin active site, are required for the formation of the transient complexes. Such conformational changes often require high energetic costs, which suggests the meta-stable nature of the transient complexes. A survey of existing structures of thrombin-inhibitor complexes indicates that the main energy cost probably comes from the conformational change of the 60's loop because of its apparent rigidity in most thrombin structures. The 148's insertion loop seems to have a more flexible nature and likely a small energy cost for conformational changes.

Figure 4-5. Substrate-induced loop movement in thrombin

Thrombin from E192Q-BPTI complex (BPTI and the 60's and 148's loop shown in yellow) is superimposed with the ecotin M84R bound thrombin (the 60's, 148's loops and bound ecotin M84R in red). The conserved thrombin structures are in light gray. The 40's loop from BPTI and the e80's loop from ecotin are labeled. Figure was generated with Raster3D (Merritt & Bacon, 1997).

Figure 4-5. Ecotin bound thrombin vs. BPTI bound thrombin

Two approaches can be taken to permit stable thrombin-inhibitor complex formation, to reduce the cost in conformational change by deleting or mutating residues in the 60's loop such as the desPPW mutant (Le Bonniec et al., 1993), or to increase the energy of interactions as found in ecotin M84R-thrombin. Ecotin M84R-thrombin complex is not the only example where positive binding energy inputs overcomes the energy cost due to conformational changes, particularly in the 60's loop. The interactions between thrombin and ecotin M84R showed similar characteristics to those observed in the BPTI-thrombin E192Q complex (van de Locht et al., 1997). BPTI weakly inhibits wild type thrombin with a μM level Ki. A point mutation in thrombin: E192Q increased the inhibition of BPTI against thrombin by 500-fold (Guinto et al., 1994). The E192Q mutation in thrombin eliminates a negative repulsion between the terminal –O in Glu^{192} side chain and BPTI Cys14 and adds a hydrogen bond between Gln^{192}-$N_{\varepsilon 2}$ and Cys14-O of BPTI. The 'extra' interactions from both point mutations are critical for the final stabilization of the thrombin-M84R or E192Q-BPTI complex.

The structures of thrombin-ecotin M84R and thrombin E192Q-BPTI are the only two existing examples where 'rigid' insertion loop at thrombin residue 60 moves significantly upon inhibitor binding. Figure 4-5 shows that the 60's and the 148's loops in thrombin adopt different conformation when bound to different inhibitors since BPTI (58 amino acids) and ecotin (142 amino acids) differ much in their sizes and three-dimensional structures. Except for the conformation of the P1 residue, the reactive site

loop of BPTI differs considerably from that of ecotin. When the C_α atoms of seven residues from each reactive site loop are superimposed, the r.m.s.d. value is greater than 1.5 Å. The 40's loop in BPTI creates a bulkier surface than the corresponding surface in ecotin M84R. As a result, the 60's loop moves further away from the active site in the BPTI bound state than in the ecotin M84R bound state. On the opposite side, the residues at the end of the e80's loop in ecotin M84R protrude more than the corresponding side of BPTI. Consequently, the insertion loop at thrombin residue 148 in ecotin bound state moves further away from the active site than its BPTI bound state. It appears that the extent of the loop movement is determined by the conformations of the incoming inhibitors.

4.4.2 60's Loop and Its Possible Role in Substrate Recognition

Further analyses suggest that the 60's insertion loop in thrombin may be involved in functions more than accommodating an incoming ligand. Analyses of the packing of this loop by the Naccess program (Hubbard & Thornton, 1993) on the current structure and over 10 published thrombin coordinates show that little change has occurred in the side-chain solvent accessibility for almost all residues in the insertion loop at thrombin residue 60. The same analysis reveals variations in side solvent accessibility for residues in the insertion loop at thrombin residue 148. This observation and the highly conserved sequence suggest that the 60's insertion loop may also participate in substrate recognition. However, most residues in the 60's insertion loop of thrombin are too far

above the active site cleft to make contacts with bound substrates. Anchoring residues for the loop, Tyr^{60A} and Phe^{60H}, are exceptions that contribute to the hydrophobicity of the S2 and S1'sites. Distance difference matrix plots (data not shown) show that the movement of the 60's insertion loop is always accompanied by movements of the 37's and the 99's loops which are indeed involved in substrate recognition (Bode et al., 1989; He et al., 1997). Because the 60's insertion loop is sandwiched between the 37's and the 99's loops, it may influence substrate binding indirectly: by affecting the position of the adjacent 37's and 99's loops. On the other hand, ligand binding at the 37's or 99's loop may also cause conformational changes in the 60's insertion loop, though this possibility has not been seen beyond the observation reported.

4.4.3 Insights on Substrate Recognition by Thrombin

Designing selective inhibitors has been challenging since thrombin and other blood coagulation factors share considerable similarities in primary sequence and three-dimensional structures. Extended substrate specificity revealed here may be employed in the quest for more specific inhibitors. For example, placing a sharp turn followed by a negatively charged residue at P5, P6 and an extended hydrophobic group at P1' may produce a more potent inhibitor of thrombin. A Val at P4 position is also a good choice. Finally, we showed that a single point mutation, M84R in a surface interaction of 6000 $Å^2$, enables strong binding. Our work shows this occurs because the Arg84 appears to be a 'hot spot' for the interactions between ecotin and thrombin (Clackson & Wells, 1995).

It forms seven hydrogen bonds, one salt bridge as well as several hydrophobic interactions through its aliphatic side chain. Energy from these interactions is critical for stabilizing conformational changes required for binding and for the 10^4 times increase in affinity.

Appendix One:

Biochemical Analyses of the mEcotin Variants

In the biochemical analyses by Christopher Eggers, other ecotin variants were also constructed besides the mEcotin variant. These mutants include the following.

mEcotin: the monomeric form of ecotin with three residue insertion after Trp130;
mEcotin/M84R: an mEcotin variant with an additional M84R mutation at P1 site;
mEcotin/$60A_4$:an mEcotin variant with a weakened secondary site;
mEcotin/$60A_4$/M84R: an mEcotin variant that contains both the M84R mutation and the $60A_4$ mutation.

Protocols of various analyses are included here, they are taken directly from the paper (Eggers et al., 2001) without any modification. All kinetic data from the following studies are included in Table 2-3.

Purification of mEcotin

Ecotin variants were expressed in the *E. coli* ecotin gene deletion strain IMΔecoJ, which was derived from JM101 (Yang et al., 1998). All ecotin variants except scEcotin were expressed and purified essentially as described (Yang et al., 1998). WT and mEcotin were further purified for analytical centrifugation on a HiLoad 26/60 Superdex 75 preparative gel filtration column (Pharmacia), as were mEcotin and $60A_4$ mEcotin for kinetics against trypsin. The scEcotin variant was purified by periplasmic preparation, followed by acidification to pH 3.0, centrifugation, and dialysis into water. A 65% ammonium sulfate precipitation was performed, and the pellet was re-suspended and dialyzed into 10 mM Tris, pH 8.0. After concentration, the protein was loaded onto a

Mono Q HR 10/10 strong anion exchange column (Pharmacia). At a constant 10 mM Tris, pH 8.0, a slow gradient of NaCl was run up to 13 mM, where it was held until the ecotin eluted from the column. The molecular weights of ecotin variants were determined by matrix-assisted laser desorption time-of-flight (MALDI-TOF) mass spectrometry on a Biospectrometry Workstation (PerSeptive Biosystems, Inc., Framingham, MA). Approximately 1 μg of protein was analyzed on a matrix of sinapinic acid with a laser intensity of 2350. Apparent melting temperatures were determined by measuring the peak fluorescence wavelength and intrinsic fluorescence intensity upon excitation at 280 nm over a temperature range of 25-90° C.

Ecotin mutagenesis

Ecotin variants were constructed in the pTacTac vector (McGrath et al., 1991), based on the bacterial expression vector pHSe5 (Muchmore et al., 1989). Site-directed mutagenesis was performed by overlap extension PCR. The first half of the mEcotin gene was constructed off the template pTacTac: ecotin K131A–[DKG] Δ138-142 (mutation of Lys131 to Ala, insertion after residue 131 with DKG, and deletion of residues 138-142) using the reverse mutagenic primer 5' CTC CTC GGC CTT TCC ATC AGC CCA GAC GCG GTA CTT CAC 3' and the second half was constructed off the template pTacTac: ecotin A132-[DKG] Δ135-142 using the forward mutagenic primer 5' GAT GGA AAG GCC GAG GAG AAA ATT GAC AAC GCG GTA GTT CGC 3'. A second PCR using these two products and the outside primers was performed and the

product was digested and ligated into the pTacTac vector to produce the mEcotin construct, which inserted ADG after Trp130. The ecotin constructs M84R, $60A_4$, and M84R $60A_4$ were previously constructed (Yang et al., 1998). mEcotin M84R, mEcotin M84R $60A_4$, and mEcotin $60A_4$ were generated by digesting constructs with BamHI and BsmI and ligating the inserts into the pTacTac: mEcotin vector.

Gel filtration

Ecotin-protease complexes were analyzed on a Pharmacia Superdex 200 10/30 column run at 0.5 mL/min in 50 mM Tris, 150mM NaCl, 20mM $CaCl_2$, pH 8.0. Wild-type ecotin and mEcotin were analyzed on a Pharmacia Superdex 75 10/30 column run at 1.0 mL/min in 50 mM $Na_xH_yPO_4$, 150 mM NaCl, pH 7.5. Apparent molecular weights were calculated from a standard curve of the following proteins: ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa), BSA (67 kDa), ovalbumin (43 kDa), and myoglobin (17.6 kDa). Peaks were assigned by absorbance at 280 nm, except for WT ecotin on the Superdex 75, which was measured at 215 nm.

Analytical centrifugation

Sedimentation equilibrium experiments were run on a Beckman XL-A ultracentrifuge, using 6-position cells at 15000, 18500, 22000, 26000, and 30000 rpm. The buffer used was 50 mM $Na_xH_yP0_4$, 150 mM NaCl, pH 7.5. Sednterp 1.01 software (Hayes et al., 1997) was used to estimate the following values based on amino acid and

buffer composition: WT partial specific volume = 0.7435, mEcotin partial specific volume = 0.7422, buffer density = 1.00956. Global data analysis was performed by non-linear least-squares fitting using Origin 4.1 software (Microcal Software, Inc.). Both proteins were fit to single-particle models. WT ecotin data analyzed was at 1 μM total protein concentration at 215 nm at all five speeds. The mEcotin data analyzed was collected at 343 μM at 300 nm, and 172 and 86 μM at 295 nm. These three concentrations were monitored at 15000, 18500, 22000, and 26000 rpm.

Fluorescence Resonance Energy Transfer

0.29 mg of the succinimidyl ester of 5-(and 6-)-carboxyfluorescein or 1.0 mg of the succinimidyl ester of 5-(and 6-)-carboxytetramethylrhodamine in DMSO were added to 5 μg WT ecotin in a buffer of 0.1 M $NaHCO_3$, pH 8.3. The protein and dye were incubated at room temperature for 1 hour, and then hydroxylamine, pH 8.5, was added to 0.15 M and incubated 1 hour. The reacted protein was extensively dialyzed against 10 mM Tris, pH 8.0. Degree of labeling was determined by measuring the absorbance at 280 nm and at the absorbance maxima of the two dyes as approximately 1/2 for fluorescein-ecotin and 2/3 for rhodamine-ecotin.

To initiate the subunit exchange experiment, 10 μM rhodamine-ecotin was added to 0.5 μM of fluorescein-ecotin. The following day, 200 μM of unlabeled ecotin was added to the sample. The fluorescence at 520 nm, F, at time, t, was measured upon

excitation at 488 nm. The rate of subunit exchange, k_{ex}, was obtained by fitting the data to the following equation, where C_1 is the fluorescence at infinite time and C_1+C_2 represents the fluorescence at time zero:

$$F(t) = C_1 + C_2 e^{-k_{ex}t}$$

The fluorescence was normalized to the initial value, C_1+C_2, when rhodamine-ecotin was added and to the final value, C_1, when unlabeled-ecotin was added. A similar technique has been used to measure the subunit exchange of α-Crystallin (Bova et al., 1997).

Fluorescence Titration

Samples for fluorescence titration experiments were extensively dialyzed against 10 mM Tris, pH 8.0. The fluorescence intensity at 340 nm upon excitation at 280 nm was measured for WT, mEcotin, or scEcotin samples and for a sample of L-tryptophan of the same absorbance at 280 nm. Serial two-fold dilutions of each sample were made with dialysis buffer, and the ratio of ecotin fluorescence to tryptophan fluorescence was fit to the following equation:

$$F = F_M + (F_D - F_M)\left(\frac{-K_d + (K_d^2 + 8K_d[I_T])^{1/2}}{4[I_T]}\right)$$

where the relative fluorescence, F, is related to the monomeric fluorescence, F_M; the dimeric fluorescence, F_D; and the total inhibitor concentration, I_T.

Kinetics

All reactions were performed in a buffer of 50 mM Tris, pH8.0, 100 mM NaCl, and 20 mM $CaCl_2$. Ecotin variants were quantitated by titration with trypsin, except in the case of mEcotin and mEcotin $60A_4$, which were titrated against chymotrypsin. Equilibrium K_i values were determined at enzyme concentrations lower than 250 times the K_i to assure measurable curvature in the inhibition curve. Protease and inhibitor were allowed to incubate for a time much longer than that calculated to be necessary given the measured k_{on} and k_{off} values to reach within 1% of equilibrium, as long as 16 days. Protease activity was found to be stable over this period of as long as two weeks by leaving one of the assay tubes without inhibitor at 4°C for the duration of the incubation and by following the activity of a sample of protease at 25°C over time. For ecotin variants with subnanomolar K_i values, no substrate-induced enzyme dissociation was observed on the time-scale of the assay, so the fractional activity was equal to fractional free enzyme. Data was fit to the following equation for tight-binding inhibitors:

$$v_i / v_o = 1 - \frac{[E_o] + [I_o] + K_i - \sqrt{([E_o] + [I_o] + K_i)^2 - 4[E_o][I_o]}}{2[E_o]}$$

Variants with K_i values above 1_ M were analyzed at multiple substrate and inhibitor concentrations by traditional Michaelis-Menton kinetics:

$$v = \frac{V_{max}[S]}{[S] + K_m\left(1 + \frac{[I]}{K_i}\right)}$$

Dissociation rate constants were determined by forming protease-inhibitor complexes at low micromolar concentrations with a slight excess of inhibitor. Protease and inhibitor were incubated at room temperature for over an hour and then diluted to low nanomolar concentrations in 25°C reaction buffer with either human neutrophil elastase or trypsin S195A as an ecotin scavenger. Elastase was used at either 125 or 250 nM and trypsin S195A was used at concentrations of 0.5 to 2 μM. At selected time points, 500 μL aliquots were taken and substrate was added to 180 μM. The initial enzyme rate was determined by monitoring absorbance at 405 nm. A standard curve of enzyme activity was constructed so that the fractional free enzyme concentration could be determined. The enzyme rate, v, at time T was fit to the following equation,

$$v = v_o + v_{tot}\left(1 - e^{(-k_{off}T)}\right)$$

where v_o is the rate of substrate turnover at time zero, including scavenger activity, and v_{tot} is the expected rate if all of the enzyme were free, based on the standard curve. In the case of dimeric ecotin variants, a burst phase was observed, so the data was fit to the following biphasic equation,

$$v = v_o + v_{tot} F_{burst}\left(1 - e^{(-k_{burst} T)}\right) + v_{tot}(1 - F_{burst})\left(1 - e^{(-k_{off} T)}\right)$$

where F_{burst} is the fraction of enzyme undergoing the burst phase and k_{burst} is the rate at which enzyme dissociates during the burst.

Association rate constants were calculated by mixing known concentrations of enzyme and inhibitor at time zero and monitoring the loss of enzymatic rate as enzyme formed stable inhibitor complexes. The time was recorded when substrate was added to a concentration many times the K_M, effectively stopping inhibitor association. Because of the slow dissociation rates of proteases from ecotin variants, dissociation could be neglected on the time scale of association assays. Rates were fit to the following equation,

$$v = C\left([E_o] - \frac{[I_o][E_o]\left(e^{\left[([I_o]-[E_o])k_{on}T\right]} - 1\right)}{[I_o]e^{\left[([I_o]-[E_o])k_{on}T\right]} - [E_o]} \right)$$

where $[E_o]$ and $[I_o]$ are the initial concentrations of protease and inhibitor, respectively, and C is the conversion factor from enzyme concentration to activity. Uncertainty values given for all kinetics values in Table 1 are the sample standard deviations of independent determinations.

Appendix Two:

Purification of Rat Trypsin D102N mutant

The following is a detailed protocol for expressing D102N mutant of rat trypsin in *Pichia pastoris* from Jennifer Harris in the Craik laboratory, with only minor modification.

MEDIA RECIPES

BMGY: 10 g yeast extract, 20 g peptone and 10 ml glycerol dissolved in 790 ml ddH_2O, autoclave the solution before adding 100 ml 1M Potassium Phosphate at pH 6.0, 100 ml 10x yeast nitrogen base (134 yeast nitrogen base in 1000 ml ddH_2O, filtered 0.2 μm, stored at 4°C), and 2 ml 500x biotin (20 mg biotin in 100 ml ddH_2O, filtered and stored at 4°C).

BMMY: 10 g yeast extract, 20 g peptone dissolved in 795 ml ddH_2O, autoclave the solution and cool to room temperature before adding 100 ml 1M Potassium Phosphate at pH 6.0, 100 ml 10x yeast nitrogen base, 2 ml 500x biotin, and 5 ml methanol (HPLC grade).

EXPRESSION PROCEDURES

1) Streak glycerol stock onto RDB plate (Minimal media minus His), and let it grow at 30°C overnight up to 24 hours

2) Inoculate a 10ml culture in BMGY, and grow with constant shaking at 30°C over night.

Growth Step:

3). Inoculate the 10ml culture from step 2 into one liter of BMGY, let it grow w constant shaking at 250rpm at 30°C overnight, until OD_{600} of the culture reaches approximately 20.

4). Spin down the grown culture in autoclaved centrifuge bottles at 300rpm for 10 to 20min, keep the pellet.

Induction Step:

5). Re-suspend the pellet from one liter of growth media in 200 ml BMMY. Let it grow in **Baffled** flask at 30°C with constant shaking 250 rpm for 72 hours. To induce expression of D102N, add 500 µl of methanol per liter of cell culture for every 12 hours.

4 Spin down cell culture at 5000 rpm for 10min, keep the supernatant, then spin down again at 8-10,0000 rpm for 20 min to get rid of any residual stuff, keep the supernatant for purification step.

PURIFICATION PROCEDURES

7) Add NaCl to supernatant from previous step until concentration reaches 2.5 M (in cold room at 4°C over 6-10 hours). Stir the solution over night at 4°C.

8) Pack Phenyl-Sepharose fast flow column (25mL in gel volume), Pre-equilibrate the column with 50 mM Tris pH 8.0, 2.5 M NaCl, run this column at 4°C with active trypsin, D102N is inactive so it is okay to run the column step at room temperature.

9). Load the trypsin supernatant at 1ml/min onto column, wash column with 150ml 50 mM Tris pH 8, 2.5M NaCl

10). Elute the column with 600 ml gradient (300 ml + 300 ml) from 50 mM Tris pH8.0 2.5 M NaCl to 0 M NaCl 50 mM Tris pH8.0, pool trypsin-containing fractions and dialyze into 50 mM Tris pH 8, 100 mM NaCl.

11). Activate with Enteropeptidase (from Biozyme), use 1: 20 (enteropeptidase : trypsin), in the presence of 10 mM $CaCl_2$, run over Soybean Trypsin Inhib-Agarose column (Sigma), elute with 100 mM Acetic Acid, 100 mM NaCl, 10 mM $CaCl_2$, 1ml/min

12). Store the purified and acD102N trypsin after dialysis into 1mM HCl (using NaOH to adjust the pH to 3.0), and 10 mM $CaCl_2$ at 4°C.

Appendix Three:

Crystallization Studies on factor Xa with ecotin M84R

For the past a few months, I have been working with Eric Slivka and Eugene Hur on the crystallization studies of blood coagulation factor Xa (fXa) in the presence of ecotin mutant M84R, which is the strongest known inhibitor against fXa. We have successfully determined the structure. It is the first time that the intact fXa structure is co-crystallized with a macromolecular inhibitor. We hope to visualize the large interface between the two proteins to fully appreciate extended interactions. It will also be interesting to compare the structure of fXa-M84R with existing structure of thrombin-M84R to look for differences in specificity of the two highly homologous protease, to help us design stronger and more selective inhibitor. The structure is still under refinement. I have included in the following a short account of our study. Structures of present working model are included at the end.

INTRODUCTION

The blood coagulation factor Xa (fXa) catalyzes the conversion of prothrombin to the functional thrombin, which then converts fibrinogen to fibrin to initiate the blood clotting process (Furie & Furie, 1988). It achieves its function by binding to factor Va (fVa) and forming the prothrombinase (Krishnaswamy et al., 1987; Nesheim et al., 1979; Rudolph et al., 2001). fXa is upstream of thrombin in the blood coagulatory cascade (Davie et al., 1991; Furie & Furie, 1988). It may serve as a better drug target than thrombin as the coagulation effect is not yet amplified through the activation of

prothrombin molecules. Moreover, fXa-based thrombolic therapy is likely to have less side effects since it is not as multi-functional as thrombin (Bode et al., 1997).

fXa consists of two polypeptide chains, linked together by a disulfide bond (Padmanabhan et al., 1993). The heavy chain, with a molecular weight of 38 kilo-dalton, harbors its serine protease activity. Its sequence shares homology with other members of the chymotrypsin-fold serine proteases superfamily (Padmanabhan et al., 1993). The light chain of fXa consists of three non-catalytic domains: the N-terminal γ-carboxyglutamic acid (Gla) domain and two epidermal growth factor (EGF)-like repeats (Davie et al., 1991). Gla domain mediates the Calcium-dependent phospholipid membrane anchoring of the fXa protein (Skogen et al., 1984; Sunnerhagen et al., 1995). Because this domain causes complication for crystallization of fXa, it has been excluded during previous crystallization analysis (Brandstetter et al., 1996; Padmanabhan et al., 1993). The two EGF-like domains have similar three dimensional structures (Padmanabhan et al., 1993). The N-terminal EGF-like domain facilitates Calcium binding of the Gla domain (Persson et al., 1991). Its relative position to the Gla domain can be altered by the presence of Calcium (Sunnerhagen et al., 1996). Like their counter parts in factor IXa (fIXa), the interactions between the two EGF-like domain may contribute to the enzymatic activity of fXa (Celie et al., 2000).

It has been difficult to design specific inhibitor against fXa because it shares considerable homology with several other serine protease in the blood coagulation

pathway (Davie et al., 1991; Stubbs & Bode, 1994). The interactions between fXa and its inhibitors are not investigated as thoroughly as the interactions between thrombin and its inhibitors. Most of these small molecule inhibitors inhibit blood coagulation by inactivating the protease function of fXa. However, it was recently shown that a fXa-binding protein from snake venom inhibits blood coagulation by binding to the Gla domain of fXa (Mizuno et al., 2001). Thus, the Gla domain becomes a new target of anticoagulant drugs.

The *E. coli* originated protease inhibitor ecotin (Chung et al., 1983) inhibits most serine proteases of the chymotrypsin-fold, including trypsin, chymotrypsin and collagenase (Seymour et al., 1994; Ulmer et al., 1995). Previous structural analyses showed that ecotin achieves inhibition by forming a tetramer with its targets, using two distinct protease binding sites: the primary and secondary sites (McGrath et al., 1994; Perona et al., 1997; Wang et al., 2001). Wild type ecotin inhibits fXa with an inhibition constant (*Ki*) of 54 pM, which is the strongest among all known fXa inhibitors (Seymour et al., 1994). An M84R mutation at the P1 residue further increases the inhibition affinity by 5 fold (Seymour et al., 1994).

We have crystallized and determined the X-ray structure of human fXa in the presence of ecotin M84R mutant. Ecotin is the largest, and the first macromolecular inhibitor that has been co-crystallized with factor Xa. It is also the first time an intact fXa is co-crystallized with an inhibitor. The structure of fXa-M84R is likely to provide

insights into the interactions between fXa and the macromolecular ligand. The structure of this complex is compared with existing structure of thrombin-M84R to reveal how ecotin adapts to different protease to achieve high affinity binding (Wang et al., 2001).

MATERIAL AND METHODS

Purification of ecotin and factor Xa and ecotin-factor complex

Ecotin M84R mutant was purified by standard procedures as described (McGrath et al., 1991; Yang et al., 1998). fXa is purified from human plasma (Haematologic Technologies Inc.). Purified human fXa-ecotin M84R complex is purified by gel filtration in the presence of 150 mM NaCl, 2mM $MgCl_2$, and 20 mM Tris at pH 7.4.

Crystallization and structural determination of ecotin-factor Xa complex

Purified fXa-M84R is concentrated to 5.4 mg/ml and crystallized at room temperature by sitting drop method in 5x5 μl drops. Initial crystals were obtained by the PegIon screen (Hampton Research), in the presence of a well solution containing 20% PEG3350, and 0.2M Sodium Potassium Tartrate at pH 7.1. Final crystals were optimized by adding trace amount of glycerol (to a final concentration of 2%) and 1μl of MPD to 4μl of well solution to combine with 5μl of protein solution to form the 10 μl drop.

The orthorhombic crystals of factor Xa-ecotin M84R diffracted to 2.65 Å. A full data set was collected at the Advance Light Source ALS) beam line 5-2 with a CCD camera. The data sets were evaluated and integrated using SCALEPACK/DENZO (Otwinowski & Minor, 1997). The crystals of factor Xa-ecotin M84R complex belong to the space group *I*222, with a = 66.8Å, b = 108.0Å, c = 186.3Å and $\alpha = \beta = \gamma = 90°$. Data statistics indicates that one asymmetric unit contains only one factor Xa and one ecotin, which gives a solvent content of 57.6%.

Structure determination and refinement

The structure of factor Xa-M84R was solved by using molecular replacement with rotational and translational functions from CNS 1.0 (Brunger et al., 1998). Initially, a hetero-dimer model of ecotin M84R mutant bound to the active site of factor Xa (1FAX.pdb) was constructed using insight II (Accelrys, San Diego, CA), using the structure of trypsin-ecotin (McGrath et al., 1994). The light chain was omitted in the final model for molecular replacement. The resulting model, however, didn't give any rotational solutions that could be improved by further translational search. In a second approach, two separate search models were constructed for a two-step molecular replacement procedure. Data to 3.5 Å were used in both rotational searches. The first rotation search used a model that contained only the protease heaving chain from 1FAX.pdb. The top solutions from the first rotational search were fixed in a second rotational search using only the ecotin molecule as its search model. Translational search and rigid body fitting for the top solutions using data to 2.65 Å gave a final solution with

an initial R value of 43.5%. The final structure was refined to 2.65 Å with an R factor of 24.7% and an R_{free} of 26.1% Detailed data and refinement statistics are listed in Table 1.

Refinement was performed using programs from CNS 1.0 (Brunger et al., 1998). Visualization and building of the model was done with Quanta98 (Molecular Simulations Inc., San Diego CA). The Gla domain of factor Xa was generated by using the build module in Quanta98 (Molecular Simulations Inc., San Diego, CA). Buried surface areas between protein domains were calculated by the program NACCESS (Hubbard & Thornton, 1993).

Table A3-1. Structural and Refinement Statistics for fXa-M84R

Structure	M84R-Xa
Space group	I222
Unit cell constants (Å)	
a	66.8
b	108.0
c	186.3
$\alpha = \beta = \gamma$	90°
Highest resolution (Å)	2.64
Total reflections	437893
Unique reflections	20381
Completeness	91.1%
Highest resolution shell	79.2% (2.69 Å to 2.64 Å)
I/I_σ	11.7
R_{merge} [a]	9.2%
Highest resolution shell	63.1% (2.69 Å to 2.64 Å)
Solvent content	57.6%
R [b] after molecular replacement	43.5%
Refinement resolution range	20–2.65 (Å)
Starting R	41% (starting)
Starting free R [c]	39% (starting)
Current R	24.7%
Current free R [c]	26.1%
water molecules	

[a] $R_{merge} = \Sigma|(I - \langle I \rangle)| / \Sigma(I)$

[b] $R = \Sigma_{h,k,l}(|F_{obs}(h,k,l)| - k|F_{calc}(h,k,l)|)^2 / \Sigma_{h,k,l}|F_{obs}(h,k,l)|$

[c] free R: Cross-validation R calculated by omitting 10% of the reflections (Kleywegt & Brunger, 1996).

PRELIMINARY RESULTS

The ecotin M84R-factor Xa complex maintains the tetrameric over features as those in other ecotin-protease complexes, with half of the tetramer generated through crystallographic symmetry. During molecular replacement, a model based on the structure of ecotin-trypsin failed to give a solution for rotational search, which suggests that the relative orientation between the protease and ecotin have changed.

The analysis of fXa-M84R is not yet complete. In the following, I included structures of the working model of fXa-M84R complex.

Figure A3-1. Structures of the working model of the fXa-M84R complex

The complete fXa, containing both the light chain and heavy chain, is shown complex with ecotin M84R mutant. Only half of the tetramer is shown in a), the complete tetrameric complex, generated by crystallographic symmetry, is shown in b). The ecotin molecule is colored in purple. The catalytic heavy chain is colored in blue, the Gla (residues 1-45), the EGF1 (residues 46-83), and the EGF2 (residues (84-127) domains in the light chain are colored red, orange, and yellow, respectively. Part of the EGF domains (residues 60-86) is modeled in, and it is colored in a darker shade of orange to distinguish from the rest of the EGF1 domain. The fXa-M84R complex a) is rotated roughly 90° from its counterpart in the tetramer in figure b).

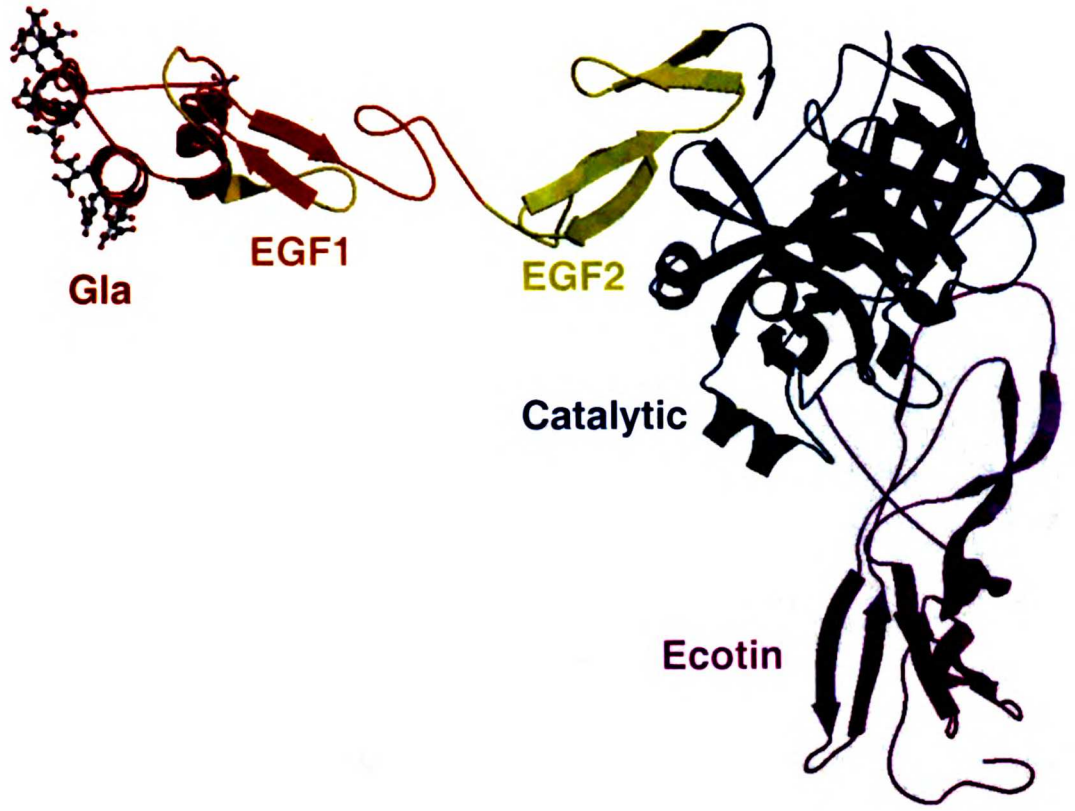


Figure 3A-1a. The structure of factor Xa-ecotin M84R

Figure A3-1b. The complete tetramer

REFERENCES

Antipova, A. S., Semenova, M. G., Belyakova, L. E. & Il'in, M. M. (2001). On relationships between molecular structure, interaction and surface behavior in mixture: small-molecule surfactant+protein. *Colloids Surf B Biointerfaces* **21**(1-3), 217-230.

Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S. & Ellman, J. A. (2000). Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin [see comments] [published erratum appears in Nat Biotechnol 2000 May;18(5):559]. *Nat Biotechnol* **18**(2), 187-93.

Bacon, D. J. & Anderson, W. F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics*(6), 2190220.

Banfield, D. K. & MacGillivray, R. T. (1992). Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species. *Proc Natl Acad Sci U S A* **89**(7), 2779-83.

Bennett, M. J., Choe, S. & Eisenberg, D. (1994). Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci U S A* **91**(8), 3127-31.

Bergdoll, M., Remy, M. H., Cagnon, C., Masson, J. M. & Dumas, P. (1997). Proline-dependent oligomerization with arm exchange. *Structure* **5**(3), 391-401.

Blomback, B., Blomback, M., Hessel, B. & Iwanaga, S. (1967). Structure of N-terminal fragments of fibrinogen and specificity of thrombin. *Nature* **215**(109), 1445-8.

Bode, W., Brandstetter, H., Mather, T. & Stubbs, M. T. (1997). Comparative analysis of haemostatic proteinases: structural aspects of thrombin, factor Xa, factor IXa and protein C. *Thromb Haemost* **78**(1), 501-11.

Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* **204**(2), 433-51.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). The refined 1.9 A crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *Embo J* **8**(11), 3467-75.

Bode, W., Turk, D. & Karshikov, A. (1992). The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci* **1**(4), 426-71.

Bogan, A. A. & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**(1), 1-9.

Bova, M. P., Ding, L. L., Horwitz, J. & Fung, B. K. (1997). Subunit exchange of alphaA-crystallin. *J Biol Chem* **272**(47), 29511-7.

Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K. & Engh, R. A. (1996). X-ray structure of active site-inhibited clotting factor Xa. Implications for drug design and substrate recognition. *J Biol Chem* **271**(47), 29988-92.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**(Pt 5), 905-21.

Celie, P. H., Lenting, P. J. & Mertens, K. (2000). Hydrophobic contact between the two epidermal growth factor-like domains of blood coagulation factor IX contributes to enzymatic activity. *J Biol Chem* **275**(1), 229-34.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R. & et al. (1989). Conformations of immunoglobulin hypervariable regions. *Nature* **342**(6252), 877-83.

Chung, C. H., Ives, H. E., Almeda, S. & Goldberg, A. L. (1983). Purification from Escherichia coli of a periplasmic protein that is a potent inhibitor of pancreatic proteases. *Journal of Biological Chemistry* **258**(18), 11032-8.

Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science* **267**(5196), 383-6.

Davie, E. W., Fujikawa, K. & Kisiel, W. (1991). The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* **30**(43), 10363-70.

Davis, M. M. & Bjorkman, P. J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature* **334**(6181), 395-402.

DiBella, E. E. & Scheraga, H. A. (1996). The role of the insertion loop around tryptophan 148 in tthe activity of thrombin. *Biochemistry* **35**(14), 4427-33.

DiBella, E. E. & Scheraga, H. A. (1998). Thrombin specificity: further evidence for the importance of the beta- insertion loop and Trp96. Implications of the hydrophobic interaction between Trp96 and Pro60B Pro60C for the activity of thrombin. *J Protein Chem* **17**(3), 197-208.

Eggers, C. T., Wang, S. X., Fletterick, R. J. & Craik, C. S. (2001). The Role of Ecotin Dimerization in Protease Inhibition. *J Mol Biol* **308**(5), 975-991.

Engh, R. A., Huber, R., Bode, W. & Schulze, A. J. (1995). Divining the serpin inhibition mechanism: a suicide substrate 'springe'? *Trends Biotechnol* **13**(12), 503-10.

Esmon, C. T. (2000). Regulation of blood coagulation. *Biochim Biophys Acta* **1477**(1-2), 349-60.

Esmon, N. L., Owen, W. G. & Esmon, C. T. (1982). Isolation of a membrane-bound cofactor for thrombin-catalyzed activation of protein C. *J Biol Chem* **257**(2), 859-64.

Fenton, J. W. d. (1988). Regulation of thrombin generation and functions. *Semin Thromb Hemost* **14**(3), 234-40.

Furie, B. & Furie, B. C. (1988). The molecular basis of blood coagulation. *Cell* **53**(4), 505-18.

Gillmor, S. A., Takeuchi, T., Yang, S. Q., Craik, C. S. & Fletterick, R. J. (2000). Compromise and accommodation in ecotin, a dimeric macromolecular inhibitor of serine proteases [In Process Citation]. *J Mol Biol* **299**(4), 993-1003.

Groeger, C., Wenzel, H. R. & Tschesche, H. (1994). BPTI backbone variants and implications for inhibitory activity. *Int J Pept Protein Res* **44**(2), 166-72.

Gronke, R. S., Bergman, B. L. & Baker, J. B. (1987). Thrombin interaction with platelets. Influence of a platelet protease nexin. *J Biol Chem* **262**(7), 3030-6.

Grutter, M. G., Priestle, J. P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J. & Stone, S. R. (1990). Crystal structure of the thrombin-hirudin complex: a novel mode of serine protease inhibition. *Embo J* **9**(8), 2361-5.

Gschwind, A., Zwick, E., Prenzel, N., Leserer, M. & Ullrich, A. (2001). Cell communication networks: epidermal growth factor receptor transactivation as the paradigm for interreceptor signal transmission. *Oncogene* **20**(13), 1594-600.

Guinto, E. R., Ye, J., Le Bonniec, B. F. & Esmon, C. T. (1994). Glu192-->Gln substitution in thrombin yields an enzyme that is effectively inhibited by bovine pancreatic trypsin inhibitor and tissue factor pathway inhibitor. *J Biol Chem* **269**(28), 18395-400.

Harris, J. L., Peterson, E. P., Hudig, D., Thornberry, N. A. & Craik, C. S. (1998). Definition and redesign of the extended substrate specificity of granzyme B. *J Biol Chem* **273**(42), 27364-73.

Hayes, D. B., Laue, T. & Philo, J. (1997). Sedimentation Interpretation Program version 1.01. *University of New Hampshire.*

He, X., Ye, J., Esmon, C. T. & Rezaie, A. R. (1997). Influence of Arginines 93, 97, and 101 of thrombin to its functional specificity. *Biochemistry* **36**(29), 8969-76.

Hogg, D. H. & Blomback, B. (1978). The mechanism of the fibrinogen-thrombin reaction. *Thromb Res* **12**(6), 953-64.

Hollenberg, M. D. (1996). Protease-mediated signalling: new paradigms for cell regulation and drug development. *Trends Pharmacol Sci* **17**(1), 3-6.

Hubbard, S. J. & Thornton, J. M. (1993). *'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University of London.*

Hynes, T. R. & Fox, R. O. (1991). The crystal structure of staphylococcal nuclease refined at 1.7 A resolution. *Proteins* **10**(2), 92-105.

Jones, S. & Thornton, J. M. (1997). Analysis of protein-protein interaction sites using surface patches. *J Mol Biol* **272**(1), 121-32.

Kisiel, W., Canfield, W. M., Ericsson, L. H. & Davie, E. W. (1977). Anticoagulant properties of bovine plasma protein C following activation by thrombin. *Biochemistry* **16**(26), 5824-31.

Kleywegt, G. J. & Brunger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure* **4**(8), 897-904.

Krishnaswamy, S., Church, W. R., Nesheim, M. E. & Mann, K. G. (1987). Activation of human prothrombin by human prothrombinase. Influence of factor Va on the reaction mechanism. *J Biol Chem* **262**(7), 3291-9.

Lambie, E. J. (1996). Receptor function at the junction. Cell-cell communication. *Curr Biol* **6**(9), 1089-91.

Laskowski, M. & Qasim, M. A. (2000). What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim Biophys Acta* **1477**(1-2), 324-37.

Le Bonniec, B. F., Guinto, E. R., MacGillivray, R. T., Stone, S. R. & Esmon, C. T. (1993). The role of thrombin's Tyr-Pro-Pro-Trp motif in the interaction with fibrinogen, thrombomodulin, protein C, antithrombin III, and the Kunitz inhibitors. *J Biol Chem* **268**(25), 19055-61.

Le Bonniec, B. F., Myles, T., Johnson, T., Knight, C. G., Tapparelli, C. & Stone, S. R. (1996). Characterization of the P2' and P3' specificities of thrombin using fluorescence-quenched substrates and mapping of the subsites by mutagenesis. *Biochemistry* **35**(22), 7114-22.

Lee, C. S., Seong, I. S., Song, H. K., Chung, C. H. & Suh, S. W. (1999). Crystallization and preliminary X-ray crystallographic analysis of the protease inhibitor ecotin in complex with chymotrypsin. *Acta Crystallogr D Biol Crystallogr* **55**(5), 1091-2.

Loll, P. J. & Lattman, E. E. (1989). The crystal structure of the ternary complex of staphylococcal nuclease, Ca2+, and the inhibitor pdTp, refined at 1.65 A. *Proteins* **5**(3), 183-201.

Lombardi, A., De Simone, G., Galdiero, S., Staiano, N., Nastri, F. & Pavone, V. (1999). From natural to synthetic multisite thrombin inhibitors. *Biopolymers* **51**(1), 19-39.

Martin, P. D., Malkowski, M. G., DiMaio, J., Konishi, Y., Ni, F. & Edwards, B. F. (1996). Bovine thrombin complexed with an uncleavable analog of residues 7-19 of fibrinogen A alpha: geometry of the catalytic triad and interactions of the P1', P2', and P3' substrate residues. *Biochemistry* **35**(40), 13030-9.

Martin, P. D., Robertson, W., Turk, D., Huber, R., Bode, W. & Edwards, B. F. (1992). The structure of residues 7-16 of the A alpha-chain of human fibrinogen bound to bovine thrombin at 2.3-A resolution. *J Biol Chem* **267**(11), 7911-20.

McGrath, M. E., Erpel, T., Browner, M. F. & Fletterick, R. J. (1991). Expression of the protease inhibitor ecotin and its co-crystallization with trypsin. *J Mol Biol* **222**(2), 139-42.

McGrath, M. E., Erpel, T., Bystroff, C. & Fletterick, R. J. (1994). Macromolecular chelation as an improved mechanism of protease inhibition: structure of the ecotin-trypsin complex. *Embo J* **13**(7), 1502-7.

McGrath, M. E., Gillmor, S. A. & Fletterick, R. J. (1995). Ecotin: lessons on survival in a protease-filled world. *Protein Sci* **4**(2), 141-8.

Merritt, E. A. & Bacon, D. J. (1997). *Raster3D: Photorealistic Molecular Graphics*. In *Methods in Enzymology*, Vol. 277, pp. 505-524. Academic Press.

Mizuno, H., Fujimoto, Z., Atoda, H. & Morita, T. (2001). Crystal structure of an anticoagulant protein in complex with the Gla domain of factor X. *Proc Natl Acad Sci U S A* **98**(13), 7230-4.

Morenweiser, R., Auerswald, E. A., van de Locht, A., Fritz, H., Sturzebecher, J. & Stubbs, M. T. (1997). Structure-based design of a potent chimeric thrombin inhibitor. *J Biol Chem* **272**(32), 19938-42.

Mossing, M. C. & Sauer, R. T. (1990). Stable, monomeric variants of lambda Cro obtained by insertion of a designed beta-hairpin sequence. *Science* **250**(4988), 1712-5.

Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. (1989). Expression and nitrogen-15 labeling of proteins for proton and nitrogen- 15 nuclear magnetic resonance. *Methods Enzymol* **177**, 44-73.

Nesheim, M. E., Taswell, J. B. & Mann, K. G. (1979). The contribution of bovine Factor V and Factor Va to the activity of prothrombinase. *J Biol Chem* **254**(21), 10952-62.

Nishida, E. & Gotoh, Y. (1993). The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends Biochem Sci* **18**(4), 128-31.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology* (Carter, C. W. J. & Sweet, R. M., eds.), Vol. 276: Macromolecular Crystallography, part A, pp. 307-26. Academic Press.

Owen, W. G., Esmon, C. T. & Jackson, C. M. (1974). The conversion of prothrombin to thrombin. I. Characterization of the reaction products formed during the activation of bovine prothrombin. *J Biol Chem* **249**(2), 594-605.

Padmanabhan, K., Padmanabhan, K. P., Tulinsky, A., Park, C. H., Bode, W., Huber, R., Blankenship, D. T., Cardin, A. D. & Kisiel, W. (1993). Structure of human des(1-45) factor Xa at 2.2 A resolution. *J Mol Biol* **232**(3), 947-66.

Pal, G., Sprengel, G., Patthy, A. & Graf, L. (1994). Alteration of the specificity of ecotin, an E. coli serine proteinase inhibitor, by site directed mutagenesis. *FEBS Lett* **342**(1), 57-60.

Pal, G., Szilagyi, L. & Graf, L. (1996). Stable monomeric form of an originally dimeric serine proteinase inhibitor, ecotin, was constructed via site directed mutagenesis. *FEBS Lett* **385**(3), 165-70.

Pawson, T. & Gish, G. D. (1992). SH2 and SH3 domains: from structure to function. *Cell* **71**(3), 359-62.

Perona, J. J., Tsu, C. A., Craik, C. S. & Fletterick, R. J. (1993). Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J Mol Biol* **230**(3), 919-33.

Perona, J. J., Tsu, C. A., Craik, C. S. & Fletterick, R. J. (1997). Crystal structure of an ecotin-collagenase complex suggests a model for recognition and cleavage of the collagen triple helix. *Biochemistry* **36**(18), 5381-92.

Persson, E., Bjork, I. & Stenflo, J. (1991). Protein structural requirements for Ca2+ binding to the light chain of factor X. Studies using isolated intact fragments containing the gamma- carboxyglutamic acid region and/or the epidermal growth factor-like domains. *J Biol Chem* **266**(4), 2444-52.

Rezaie, A. R., Cooper, S. T., Church, F. C. & Esmon, C. T. (1995). Protein C inhibitor is a potent inhibitor of the thrombin- thrombomodulin complex. *J Biol Chem* **270**(43), 25336-9.

Rezaie, A. R. & Olson, S. T. (1997). Contribution of lysine 60f to S1' specificity of thrombin. *Biochemistry* **36**(5), 1026-33.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Adv Protein Chem* **34**, 167-339.

Rosenberg, R. D. & Damus, P. S. (1973). The purification and mechanism of action of human antithrombin-heparin cofactor. *J Biol Chem* **248**(18), 6490-505.

Rudolph, A. E., Porche-Sorbet, R. & Miletich, J. P. (2001). Definition of a factor Va binding site in factor Xa. *J Biol Chem* **276**(7), 5123-8.

Ruhlmann, A., Kukla, D., Schwager, P., Bartels, K. & Huber, R. (1973). Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. Crystal

structure determination and stereochemistry of the contact region. *J Mol Biol* **77**(3), 417-36.

Rydel, T. J., Tulinsky, A., Bode, W. & Huber, R. (1991). Refined structure of the hirudin-thrombin complex. *J Mol Biol* **221**(2), 583-601.

Sayle, R. A. & Milner-White, E. J. (1995). RASMOL: biomolecular graphics for all. *Trends Biochem Sci* **20**(9), 374.

Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**(2), 157-62.

Scheidig, A. J., Hynes, T. R., Pelletier, L. A., Wells, J. A. & Kossiakoff, A. A. (1997). Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): engineering of inhibitors with altered specificities. *Protein Sci* **6**(9), 1806-24.

Seymour, J. L., Lindquist, R. N., Dennis, M. S., Moffat, B., Yansura, D., Reilly, D., Wessinger, M. E. & Lazarus, R. A. (1994). Ecotin is a potent anticoagulant and reversible tight-binding inhibitor of factor Xa. *Biochemistry* **33**(13), 3949-58.

Sheehan, J. P., Tollefsen, D. M. & Sadler, J. E. (1994). Heparin cofactor II is regulated allosterically and not primarily by template effects. Studies with mutant thrombins and glycosaminoglycans. *J Biol Chem* **269**(52), 32747-51.

Shin, D. H., Song, H. K., Seong, I. S., Lee, C. S., Chung, C. H. & Suh, S. W. (1996). Crystal structure analyses of uncomplexed ecotin in two crystal forms: implications for its function and stability. *Protein Sci* **5**(11), 2236-47.

Skogen, W. F., Esmon, C. T. & Cox, A. C. (1984). Comparison of coagulation factor Xa and des-(1-44)factor Xa in the assembly of prothrombinase. *J Biol Chem* **259**(4), 2306-10.

Slon-Usakiewicz, J. J., Sivaraman, J., Li, Y., Cygler, M. & Konishi, Y. (2000). Design of P1' and P3' residues of trivalent thrombin inhibitors and their crystal structures. *Biochemistry* **39**(9), 2384-91.

Solomon, M., Belenghi, B., Delledonne, M., Menachem, E. & Levine, A. (1999). The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. *Plant Cell* **11**(3), 431-44.

Sprang, S., Standing, T., Fletterick, R. J., Stroud, R. M., Finer-Moore, J., Xuong, N. H., Hamlin, R., Rutter, W. J. & Craik, C. S. (1987). The three-dimensional structure of

Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. *Science* **237**(4817), 905-9.

Stubbs, M. T. & Bode, W. (1993). A player of many parts: the spotlight falls on thrombin's structure. *Thromb Res* **69**(1), 1-58.

Stubbs, M. T. & Bode, W. (1994). Coagulation factors and their inhibitors. *Curr Opin Struct Biol* **4**(6), 823-32.

Stubbs, M. T., Oschkinat, H., Mayr, I., Huber, R., Angliker, H., Stone, S. R. & Bode, W. (1992). The interaction of thrombin with fibrinogen. A structural basis for its specificity. *Eur J Biochem* **206**(1), 187-95.

Sunnerhagen, M., Forsen, S., Hoffren, A. M., Drakenberg, T., Teleman, O. & Stenflo, J. (1995). Structure of the Ca(2+)-free Gla domain sheds light on membrane binding of blood coagulation proteins. *Nat Struct Biol* **2**(6), 504-9.

Sunnerhagen, M., Olah, G. A., Stenflo, J., Forsen, S., Drakenberg, T. & Trewhella, J. (1996). The relative orientation of Gla and EGF domains in coagulation factor X is altered by Ca2+ binding to the first EGF domain. A combined NMR- small angle X-ray scattering study. *Biochemistry* **35**(36), 11547-59.

Takeuchi, Y., Satow, Y., Nakamura, K. T. & Mitsui, Y. (1991). Refined crystal structure of the complex of subtilisin BPN' and Streptomyces subtilisin inhibitor at 1.8 A resolution. *J Mol Biol* **221**(1), 309-25.

Taylor, F. R., Bixler, S. A., Budman, J. I., Wen, D., Karpusas, M., Ryan, S. T., Jaworski, G. J., Safari-Fard, A., Pollard, S. & Whitty, A. (1999). Induced fit activation mechanism of the exceptionally specific serine protease, complement factor D. *Biochemistry* **38**(9), 2849-59.

Ullrich, A. & Schlessinger, J. (1990). Signal transduction by receptors with tyrosine kinase activity. *Cell* **61**(2), 203-12.

Ulmer, J. S., Lindquist, R. N., Dennis, M. S. & Lazarus, R. A. (1995). Ecotin is a potent inhibitor of the contact system proteases factor XIIa and plasma kallikrein. *FEBS Lett* **365**(2-3), 159-63.

van de Locht, A., Bode, W., Huber, R., Le Bonniec, B. F., Stone, S. R., Esmon, C. T. & Stubbs, M. T. (1997). The thrombin E192Q-BPTI complex reveals gross structural rearrangements: implications for the interaction with antithrombin and thrombomodulin. *Embo J* **16**(11), 2977-84.

van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., Hoffken, W. & Bode, W. (1995). Two heads are better than one: crystal structure of the insect

derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO J* **14**(21), 5149-57.

van de Locht, A., Stubbs, M. T., Bode, W., Friedrich, T., Bollschweiler, C., Hoffken, W. & Huber, R. (1996). The ornithodorin-thrombin crystal structure, a key to the TAP enigma? *Embo J* **15**(22), 6011-7.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**(2), 127-34.

Wang, C. I., Yang, Q. & Craik, C. S. (1996). Phage display of proteases and macromolecular inhibitors. *Methods Enzymol* **267**, 52-68.

Wang, S. X., Esmon, C. T. & Fletterick, R. J. (2001). Crystal Structure of Thrombin-Ecotin Reveals Conformational Changes and Extended Interactions. *Biochemistry* **40**(34), 10038-10046.

Wells, J. A. (1990). Additivity of mutational effects in proteins. *Biochemistry* **29**(37), 8509-17.

Willett, W. S., Gillmor, S. A., Perona, J. J., Fletterick, R. J. & Craik, C. S. (1995). Engineered metal regulation of trypsin specificity. *Biochemistry* **34**(7), 2172-80.

Wynn, R., Harkins, P. C., Richards, F. M. & Fox, R. O. (1997). Comparison of straight chain and cyclic unnatural amino acids embedded in the core of staphylococcal nuclease. *Protein Sci* **6**(8), 1621-6.

Yang, S. Q. & Craik, C. S. (1998). Engineering bidentate macromolecular inhibitors for trypsin and urokinase-type plasminogen activator. *J Mol Biol* **279**(4), 1001-11.

Yang, S. Q., Wang, C. I., Gillmor, S. A., Fletterick, R. J. & Craik, C. S. (1998). Ecotin: a serine protease inhibitor with two distinct and interacting binding sites. *J Mol Biol* **279**(4), 945-57.

Yennawar, N. H., Yennawar, H. P. & Farber, G. K. (1994). X-ray crystal structure of gamma-chymotrypsin in hexane. *Biochemistry* **33**(23), 7326-36.

Zamoyska, R. (1998). CD4 and CD8: modulators of T-cell receptor recognition of antigen and of immune responses? *Curr Opin Immunol* **10**(1), 82-7.