University of Alberta

Catalytic and Inhibitory Mechanisms of SARS Coronavirus Main Peptidase and Streptogrisin B - a Structural Perspective

by

Ting-Wai Lee

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

Edmonton, Alberta, Canada
Spring 2008

Your file Votre référence
ISBN: 978-0-494-45551-7
Our file Notre référence
ISBN: 978-0-494-45551-7

Abstract

Severe acute respiratory syndrome (SARS), a highly transmissible, infectious and often fatal disease, is caused by a coronavirus (CoV), whose main peptidase (M^{pro}) has been one of the most attractive targets for the development of anti-SARS drugs. Part One of this thesis reports six molecular structures of M^{pro} and its variant with an additional Ala at the N-terminus of each protomer ($M^{pro}_{+A(-1)}$). Structural comparisons suggest a dynamic equilibrium between the catalytically competent and the catalytically incompetent conformations of M^{pro}. Three of the molecular structures were determined in the presence of an aza-peptide epoxide (APE) inhibitor, revealing that the binding of APE to M^{pro} is in a substrate-like manner and follows an induced-fit model. Cys145 S^{γ} of M^{pro} nucleophilically attacks the epoxide atom C3 of APE, resulting in the formation of a covalent bond between these two atoms, and the opening of the conformationally strained epoxide ring of APE. These results refine our understanding of the dependence of the catalytic activity of M^{pro} on a variety of factors (e.g. pH), and provide clues for designing peptidomimetic inhibitors of M^{pro}.

Sequence-to-reactivity algorithms (SRAs) have potential applications in the design of drugs, vaccines and proteins. An additivity-based SRA has recently been developed to predict the equilibrium association constants for variants of a protein inhibitor, turkey ovomucoid third domain (OMTKY3), with a panel of six serine peptidases, one of which is streptogrisin B (SGPB). The evaluation of the performance of this SRA requires high-resolution molecular structures. Part Two of this thesis reports two independent high-resolution molecular structures of SGPB that unexpectedly represent mixtures of its second tetrahedral intermediates and enzyme-

product complexes. Structural comparisons show conformational changes in some active-site residues of SGPB favoring the progression of its catalytic mechanism. Also reported are the high-resolution molecular structures of the unbound wild-type OMTKY3 and its Ala32I variant (OMTKY3-Ala32I), and the SGPB:OMTKY3-Ala32I complex. These molecular structures help us to understand some of the non-additivity effects in the SRA, and provide the insight that the validity of the additivity approximation in the SRA requires protein-protein interactions in a rigid-body manner.

Dedicated to the memory of my mother

Acknowledgement

My first thanks certainly go to my mentor, Dr. Michael N. G. James. He has brought me to this very exciting field of structural biology. I am most impressed by his knowledge and experience in X-ray crystallography and structural biology, his enthusiasm in sharing his knowledge and experience, and his rigor in dealing with science. I appreciate very much the freedom and the independence he has allowed to me in doing research, giving me the chance to learn from hopeless trials and stupid mistakes. Michael is very friendly and open-minded, and cares about everyone's well being. I am very grateful to his understanding, patience, tolerance and encouragement in whatever aspects of my life over the last six years.

I thank Dr. J. N. Mark Glover and Dr. Charles F. B. Holmes very much for following up my academic progress, giving suggestions for my research, and providing advice and support for my development. I also thank Dr. John C. Vederas (Department of Chemistry, University of Alberta) and Dr. Robert J. Fletterick (Department of Biochemistry and Biophysics, University of California at San Francisco) very much for reviewing this thesis, and providing invaluable comments and suggestions for improvement.

It was great to collaborate with the groups of Dr. Lindsay D. Eltis (Department of Microbiology and Immunology, University of British Columbia) and Dr. James C. Powers (School of Chemistry and Biochemistry, Georgia Institute of Technology) on the study of SARS coronavirus main peptidase, and with the group of the late Dr. Michael Laskowski, Jr. (Department of Chemistry, Purdue University), particularly with Dr. M. A. Qasim (presently in the Department of Chemistry, Xavier

University of Louisiana), on the study of the sequence-to-reactivity algorithm for serine peptidases and their inhibitors. I recognize the importance of their help in providing good-quality peptidase and inhibitor samples, analyzing and interpreting the structural results, and reviewing my manuscripts.

Every member of the James' group, no matter whether he/she is now working here or elsewhere, is unforgettable. To name a few, Meitian Wang was very generous and patient in showing me how to use the computer programs at the beginning of my research. The efforts Kathy Bateman has made in the study of the sequence-to-reactivity algorithm, as evidenced by her PhD thesis, are amazing. Maia Cherney's talent in getting virtually every protein crystallized is very much appreciated. Ernst Bergmann, Joanne Lemieux and Jason Maynes are my role models. Bindu Pillai is our `PNAS Superwoman'. The professional spirit demonstrated by Sheraz Khan, our `X-ray Facility Manager', in managing all of the crystallography-related business is impressive. Jo Parrish's and Craig Garen's laughter is always cheering. Grace Garen cares about her sons very much, reminding me of my mother ... I have to stop here, otherwise Michael would complain.

I treasure very much the continuous support from my friends back in Hong Kong, including Ernst Lam, Douglas Lam, Damian Siu and Eric Wan. They always try their best in entertaining my long-distance calls, expressing their opinions and giving me advice on various issues, searching for journal articles that I cannot find here, shipping stuff to me, and so on. They are among the most influential people in many of my important decisions, including the one for my coming here six years ago.

I am also very grateful to my sister, Yee-Kay, for her care. She has been working very hard to clean up all the mess I have left home since I came to Canada six years ago. She does not know much about science, but she does try her best to understand the difficult situations I have to face as a scientist, to appreciate what I am doing, and to share her views as a layperson. I will not leave her alone, no matter where I am in the future.

My deepest gratitude of course goes to my mother. Without her in my mind, I would not be so determined to study abroad, to pursue biomedical research, and to obtain a PhD degree, and I would not be able to survive all the crises in the past 15 years. I have got lost for several times in my life, but I can always stand up and identify my goals again when I remember what she said to me. It is regrettable that she cannot be here to see how I am doing nowadays, but I wish she could share in my happiness from the other world. I wish she would not need to worry about me anymore. My whole career - not just this thesis; this thesis is just the beginning - will be dedicated to her.

The studies reported in this thesis were supported by research grants from the Canadian Institute of Health Research (CIHR) to Dr. Michael N. G. James. The study of SARS coronavirus main peptidase was also supported by a grant from the Protein Engineering Network of Centres of Excellence (PENCE), and a Strategic Grant from the Natural Sciences and Engineering Resesrch Council (NSERC) of Canada. I acknowledge the receipt of the 75th Anniversary Award from the Faculty of Medicine and Dentistry, University of Alberta in 2004. The X-ray diffraction data from all of the crystals in these studies were collected at the synchrotron Beamline 8.3.1 at the

Advanced Light Source (ALS) in the Lawrence Berkeley National Laboratory, under an agreement with the Alberta Synchrotron Institute (ASI). The ALS is operated by the US Department of Energy, and supported by the National Institutes of Health. Beamline 8.3.1 was funded by the National Science Foundation, the University of California and Henry Wheeler. The ASI synchrotron access program is supported by grants from the Alberta Science and Research Authority (ASRA), the Alberta Heritage Foundation for Medical Research (AHFMR) and the Western Economic Diversification (WED) of Canada.

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

αLP	α-lytic peptidase
AGln	Aza-glutamine
Å	Angström (10^{-10} m)
APE	Aza-peptide epoxide
CARL	Subtilisin Carlsberg
CHYM	Bovine chymotrypsin Aα
CMK	Chloromethyl ketone
CoV	Coronavirus
DPI	Diffraction-component precision index
HLE	Human leukocyte elastase
ID	Chain ID in the PDB file
K_{assoc}	Equilibrium association constant
M^{pro}	Main peptidase
$M^{pro}_{+A(-1)}$	Variant of main peptidase with an additional Ala at the N-terminus of each protomer
M_r	Molecular mass
OMTKY3	Recombinant turkey ovomucoid third domain
OMTKY3-Ala32I	Ala32I variant of recombinant turkey ovomucoid third domain
PPE	Porcine pancreatic elastase
rmsd	Root-mean-square difference
SARS	Severe acute respiratory syndrome

σ_r	Overall positional uncertainty
SGPA	Streptogrisin A
SGPB	Streptogrisin B
SRA	Sequence-to-reactivity algorithm
wtOMJPQ3	Wild-type Japanese quail ovomucoid third domain
wtOMSVP3	Wild-type silver pheasant ovomucoid third domain
wtOMTKY3	Wild-type turkey ovomucoid third domain

Chapter 1

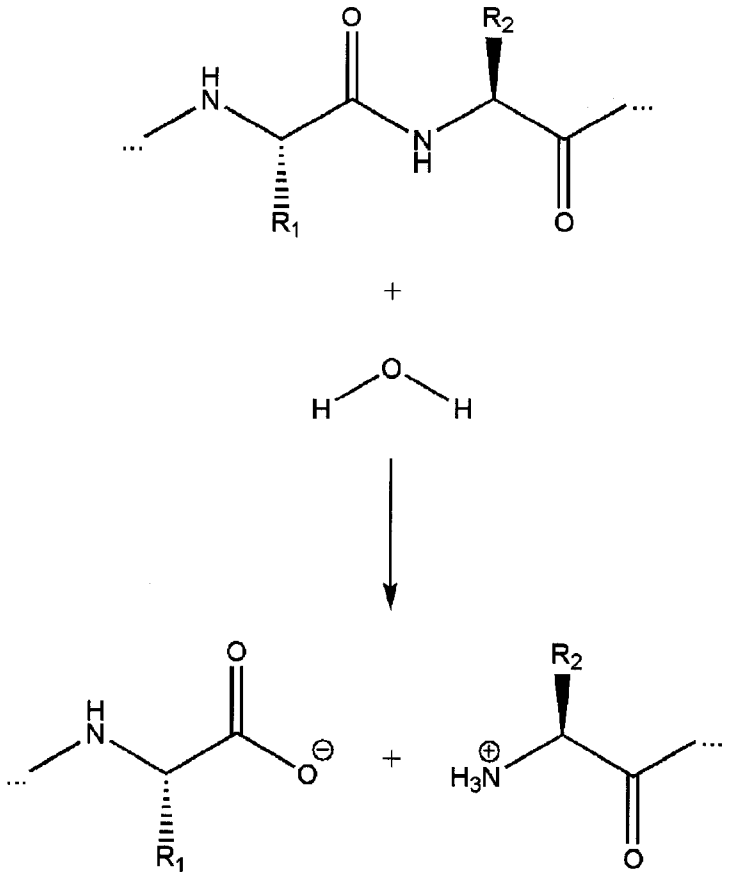
Introduction

1.1 Peptidases: why study them?

Proteins are linear polymers in which amino acids are connected in specified sequences by peptide bonds. A wide variety of biological processes, such as food digestion, blood coagulation, wound healing, tissue remodeling, infection, immunity, cell-cycle progression and apoptosis, require the hydrolysis of peptide bonds in proteins (Figure 1-1).[1] However, in physiological conditions, the rate of this reaction is too slow to support the survival of organisms. For example, at 37 °C and pH 7, the first-order rate constant for peptide-bond hydrolysis is in the order of 10^{-11} s^{-1}, corresponding to a half-life in the order of 10^2 years.[2] This necessitated the evolution of various peptidases - enzymes catalyzing peptide-bond hydrolysis. Of all genes in organisms whose genomes have been sequenced to date, approximately 2 % encode peptidases and their homologs.[3]

Each peptidase exhibits specificity in catalyzing peptide-bond hydrolysis. The specificity is based on the location of a peptide bond in the substrate, and on the identities of the amino-acid residues flanking a peptide bond. Differences in the location specificity have allowed the classification of peptidases into exopeptidases (peptidases catalyzing the hydrolysis of peptide bonds near the termini of the substrate polypeptide) and endopeptidases (peptidases catalyzing the hydrolysis of internal peptide bonds of the substrate). The two classes of peptidases have then been further classified according to the precise locations of their target peptide bonds in the substrates.[4,5] Alternatively, peptidases have been classified according to their catalytic

Figure 1-1 Hydrolysis of a peptide bond. R_1 and R_2 represent the side chains of the amino-acid residues forming the scissile peptide bond.



types (i.e. the identities of the key participants in their catalytic mechanisms), namely serine, cysteine, threonine, aspartic, glutamic and metallo.[6] Increased availabilities of amino-acid sequences and three-dimensional structures have motivated further classification of peptidases into clans and families. A clan consists of peptidases with similar three-dimensional structures and the same order of catalytic residues in their polypeptide chains; whereas a family is a subset of a clan, consisting of peptidases that show statistically significant relationships in the amino-acid sequences of their catalytically relevant parts.[4,7] Interestingly, some clans have later been combined into one that includes peptidases of different catalytic types (MEROPS - the peptidase database, http://merops.sanger.ac.uk/).[4]

Owing to their participation in a broad range of vital processes *in vivo*, the activities of peptidases require precise regulation. Regulatory mechanisms include the adjustments in the level of expression of peptidase genes, the trafficking and the compartmentalization of peptidases, the limited proteolysis for the activation of peptidase zymogens and for the structural modifications of peptidases, the degradation of peptidases, and the inhibition of peptidases by protein inhibitors (Section 1.7).[8,9] In humans, defects in peptidases and in their regulatory mechanisms have been found relevant to diseases including cancer and arthritis. Moreover, in many infections, peptidases encoded by the microorganisms act as the virulence factors.[1] A detailed understanding of the structure-function relationships in peptidases is essential, not only for appreciating how various diseases are caused by defects in peptidases and in their regulatory mechanisms, but also for discovering potent and selective inhibitors of undesirable or under-regulated peptidases.[10]

1.2 Serine peptidases

Each serine peptidase has a serine at its active site, whose side-chain hydroxyl group acts as a nucleophile in the catalytic mechanism for peptide-bond hydrolysis. Approximately one-third of all peptidases and peptidase homologs known so far are serine peptidases, grouped into 16 clans and 42 families (MEROPS - the peptidase database, http://merops.sanger.ac.uk/). Most serine peptidases adopt a fold similar to that of chymotrypsin, which consists of two β-barrels with the active site formed at their junction. At the active site of chymotrypsin, Ser195 (the nucleophilic serine) O^{γ} forms a hydrogen bond with His57 $N^{\varepsilon 2}$, and His57 $N^{\delta 1}$ forms a hydrogen bond with Asp102 $O^{\delta 2}$ (Figure 1-2a)[11] Collectively known as the catalytic triad, His57, Asp102 and Ser195 have been shown to be catalytically essential.[12,13,14]

The catalytic mechanism of chymotrypsin-like serine peptidases has been well established.[15] It proceeds through five steps (Figures 1-2a to 1-2e). For convenience, the residue numbering for all chymotrypsin-like serine peptidases is based on that of chymotrypsinogen Aα:[16]

(a) Formation of the Michaelis complex

The substrate enters the active site of the peptidase, leading to the formation of an enzyme-substrate complex (also known as the Michaelis complex). In this complex, Ser195 O^{γ} of the peptidase is positioned above the carbonyl C of the scissile peptide bond of the substrate at a distance of 2.6 Å to 2.9 Å, and the axis passing through these two atoms is roughly orthogonal to the plane defined by the atoms of the scissile peptide group. This geometry is prepared for the nucleophilic attack in step (b).[17,18] Residues Gly193 to Ser195 of the peptidase constitute an oxyanion hole,

Figure 1-2 Catalytic mechanism of chymotrypsin-like serine peptidases.[38] (a) The Michaelis complex. (b) The first tetrahedral intermediate. (c) The acyl-enzyme intermediate. (d) The second tetrahedral intermediate. (e) The unbound peptidase. Residues of the peptidase are shown in red. Substrate residues and the water molecule are shown in green. R_1 and R_2 represent the side chains of residues P1 and P2 of the substrate (See Figure 1-3), respectively. R and R' represent the polypeptide segments on the N-terminal and the C-terminal sides of the scissile peptide bond of the substrate, respectively. Hydrogen bonds are indicated by dashed lines. The dipole moment across the P2-P1 peptide bond of the substrate is indicated by a straight arrow with a cross at its positive end. Partial charges are indicated by δ+ and δ-. The curved arrows in each panel indicate the electron flow that drives the catalytic cycle to the following step.

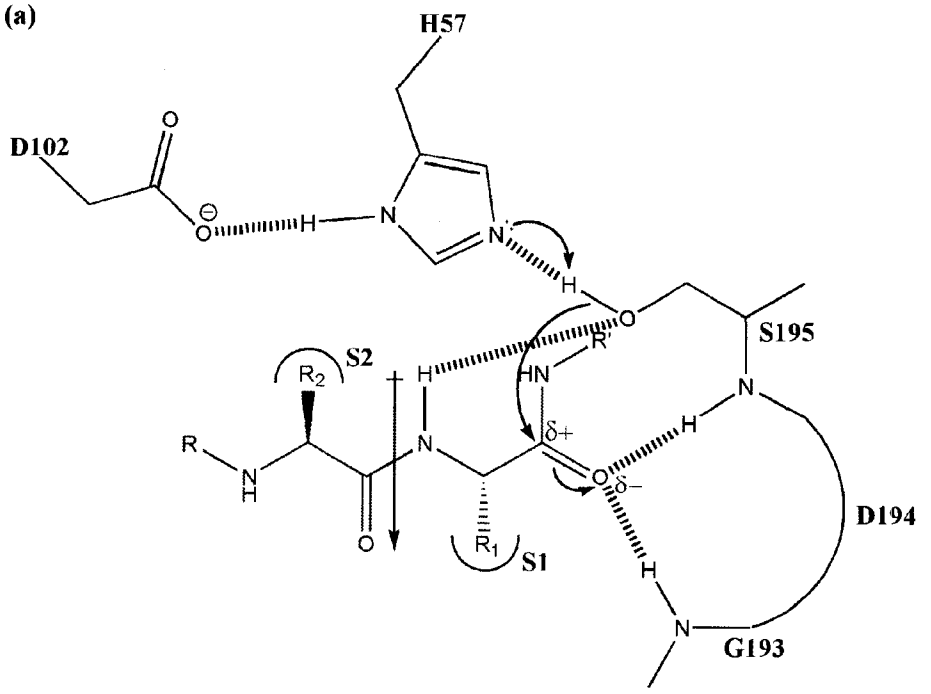
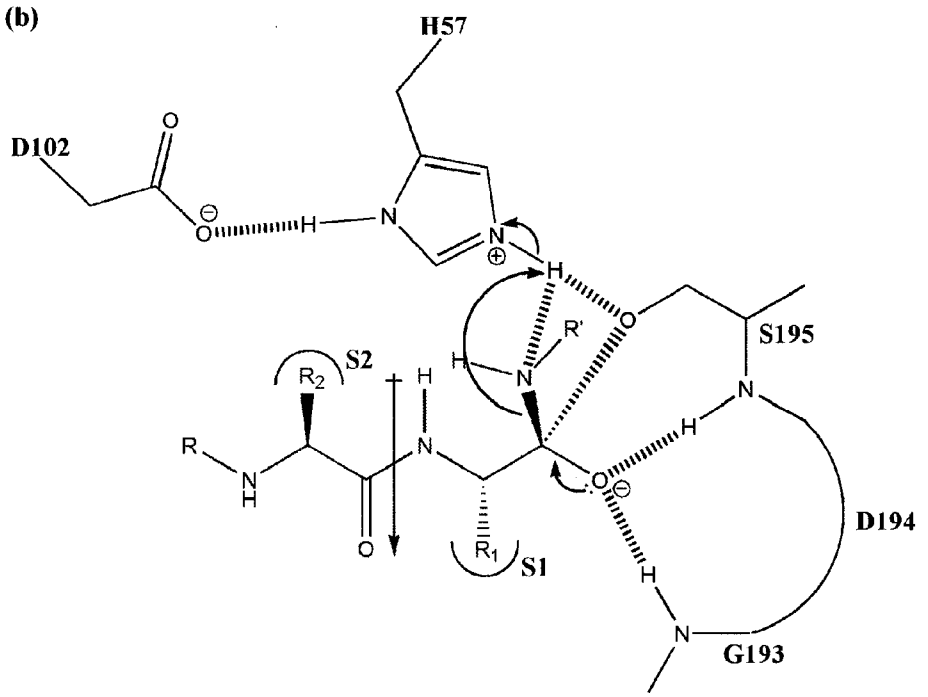


Figure 1-2 (continued)



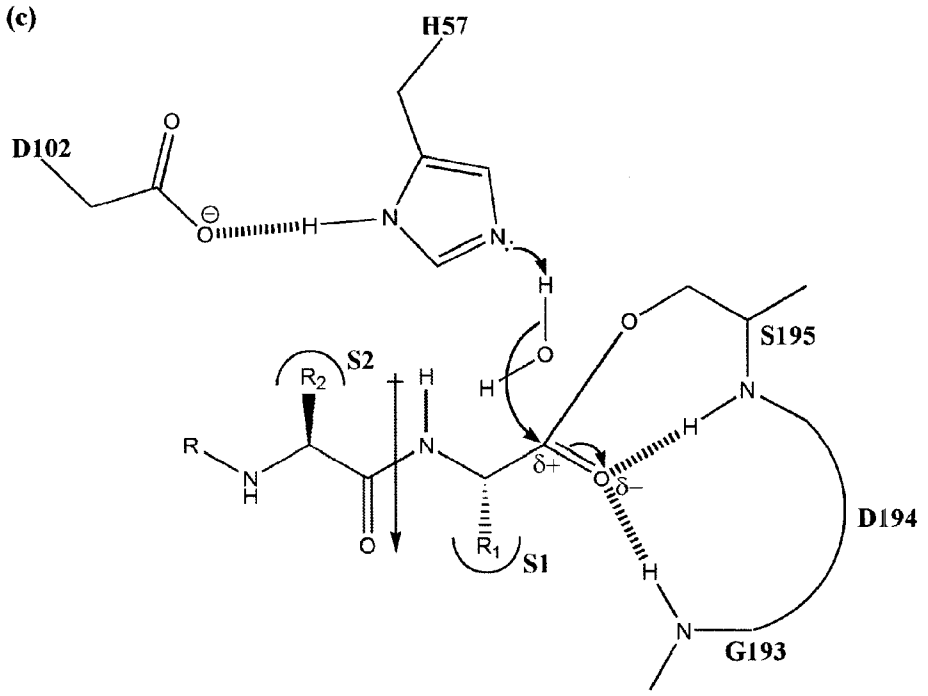
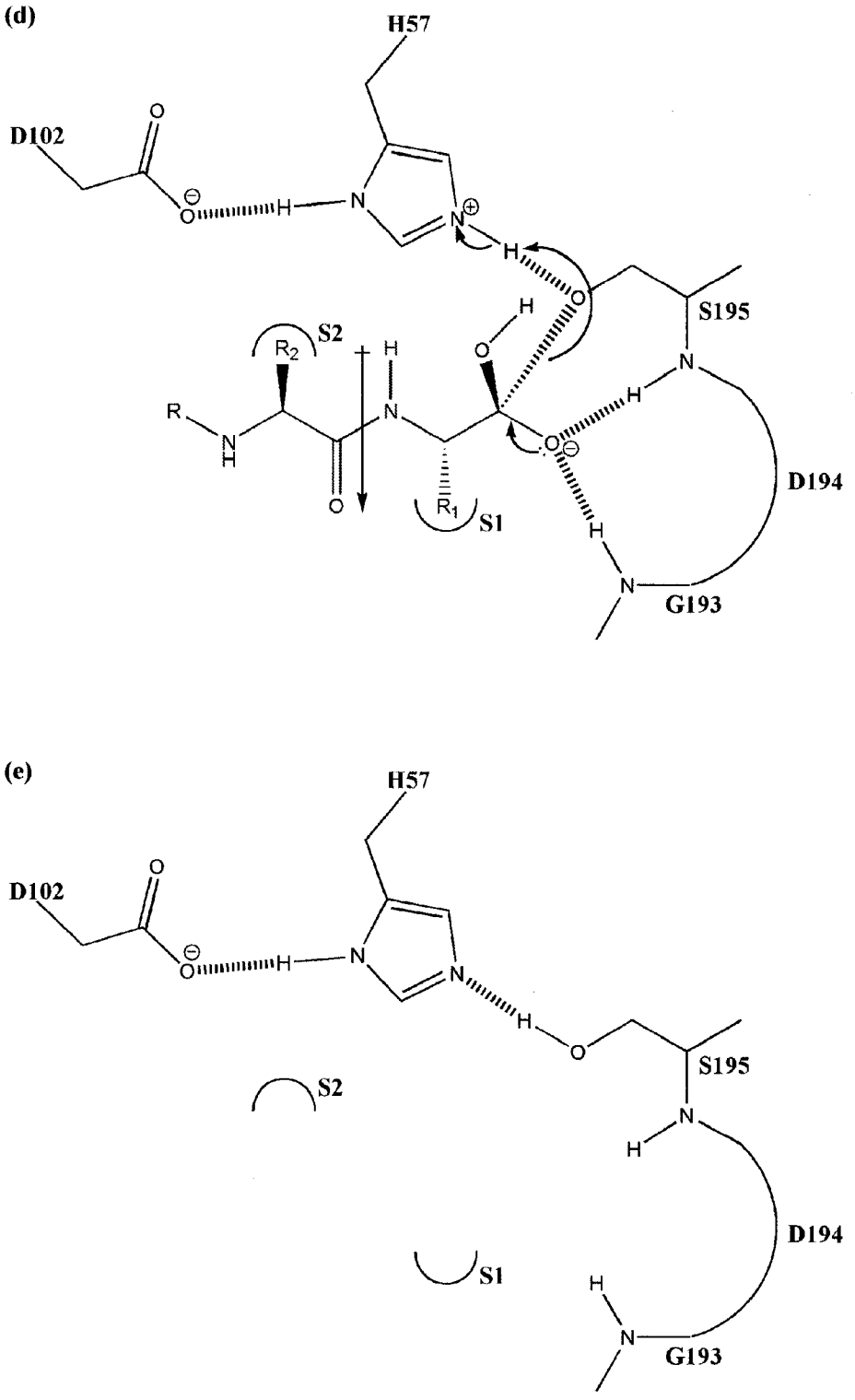


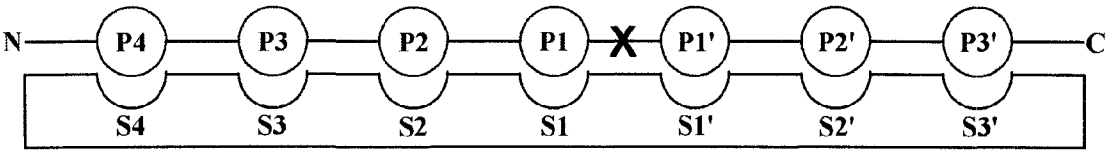
Figure 1-2 (continued)



accommodating the carbonyl O of the scissile peptide group of the substrate. This oxygen forms hydrogen bonds with the main-chain NH groups of Gly193 and Ser195 of the peptidase (Figure 1-2a).

Enzyme-substrate interactions are not limited to the local region of the active site. In fact, the substrate-binding region of the peptidase extends over seven residues of the substrate, with four of them preceding (P4 to P1, nomenclature based on that of Schechter and Berger[19]) and three of them following the scissile peptide bond (P1' to P3'; Figure 1-3).[20,21] The main chain of the substrate from P4 to P1 interacts with that of the peptidase from Ser214 to Gly216 in the fashion of an anti-parallel β-sheet, forming hydrogen bonds between the NH and the C=O groups from the two main chains.[10] The substrate-binding region of the peptidase can be divided into subsites S4 to S3', corresponding to sites where substrate residues P4 to P3' will bind (Figure 1-3). The local structure of each subsite determines which amino-acid residue is preferred at the corresponding substrate position. Wide structural and electrostatic variations have been observed in the S1 subsites, also known as the S1 specificity pockets because of their pocket-like shapes, of chymotrypsin-like serine peptidases, resulting in the broad range of substrate specificities they exhibit. For example, chymotrypsin prefers a substrate with a hydrophobic residue P1 (e.g. Phe, Leu), because its S1 specificity pocket is hydrophobic. Trypsin prefers a substrate with P1-Lys or P1-Arg, because the negatively-charged side chain of Asp189 at the bottom of its S1 specificity pocket can form an ionic interaction with the positively-charged side chain of Lys or Arg. Elastase prefers a substrate whose residue P1 has no (i.e. Gly) or a small aliphatic side chain (e.g. Ala, Val), because the size of its S1 specificity pocket

Figure 1-3 Schechter-Berger nomenclature for the substrate-binding subsites of a peptidase (sites S4 to S3'; red), and the residues of a substrate or a substrate-like inhibitor (residues P4 to P3'; green).[19] N and C indicate the N- and the C-termini, respectively, of the substrate or the substrate-like inhibitor. X indicates the scissile peptide bond of the substrate, or the reactive site of a protein inhibitor.



is limited by the side chains of Val216 and Thr226.[22] Streptogrisin E prefers a substrate with P1-Glu, because the histidine triad (His213, His199 and His228) at the bottom of its S1 specificity pocket can stabilize the negatively-charged side chain of Glu.[23]

(b) Formation of the first tetrahedral intermediate

The side-chain hydroxyl group of a free serine has a pKa of about 15. This indicates the lack of nucleophilicity of its O^{γ}. However, in the Michaelis complex, there are three features contributing to enhance the nucleophilicity of Ser195 O^{γ} of the peptidase (Figure 1-2a). Firstly, within the peptidase, Ser195 O^{γ} forms a strong hydrogen bond with His57 $N^{\varepsilon 2}$. Secondly, Ser195 O^{γ} of the peptidase forms a hydrogen bond with the main-chain NH of residue P1 of the substrate, and is positioned at the positive end of the dipole moment generated by the P2-P1 peptide group of the substrate. Thirdly, the two hydrogen bonds provided by the oxyanion hole of the peptidase polarize the C=O bond of the scissile peptide group of the substrate, increasing the electrophilicity of the carbonyl C of the scissile peptide bond. Thus, Ser195 O^{γ} of the peptidase donates its associated proton to His57 $N^{\varepsilon 2}$ of the peptidase, and performs a nucleophilic attack on the carbonyl C of the scissile peptide bond of the substrate. The protonated and positively-charged side chain of His57 is stabilized by its ionic interaction with the negatively-charged side chain of Asp102 within the peptidase. As a covalent bond forms between Ser195 O^{γ} of the peptidase and the carbonyl C of the scissile peptide bond of the substrate, the C=O bond of the scissile peptide group becomes a single bond and a full negative charge develops on its oxygen. This negatively-charged oxygen remains stabilized by the two hydrogen

bonds in the oxyanion hole of the peptidase. The carbonyl C of the scissile peptide bond of the substrate undergoes a change in hybridization from sp^2 to sp^3, and thus a change in bonding geometry from trigonal-planar to tetrahedral. As a result, the Michaelis complex has transformed into the first tetrahedral intermediate (Figure 1-2b).

(c) Formation of the acyl-enzyme intermediate

In the first tetrahedral intermediate, His57 $N^{\varepsilon 2}$ of the peptidase donates its associated proton through a strong hydrogen bond to the nitrogen of the scissile peptide bond of the substrate, thereby making the polypeptide segment C-terminal to the scissile peptide bond of the substrate a good leaving group. The scissile peptide bond of the substrate is then cleaved. This drives the restoration of the C=O double bond of the scissile peptide group and the trigonal-planar bonding geometry of its carbon. Consequently, the polypeptide segment C-terminal to the scissile peptide bond of the substrate is released, whereas the polypeptide segment on the N-terminal side remains associated with the peptidase in an ester bond with Ser195 O^{γ}. This complex is known as the acyl-enzyme intermediate (Figure 1-2c). Molecular structures have been reported for the acyl-enzyme intermediates of chymotrypsin[24,25,26] and streptogrisin A.[27]

Steps (a) to (c) are collectively known as the acylation stage, whereas steps (d) and (e) are collectively known as the deacylation stage:

(d) Formation of the second tetrahedral intermediate

A water molecule enters the active site of the peptidase from the side of the leaving group,[28] and forms a hydrogen bond with His57 $N^{\varepsilon 2}$ of the peptidase. This

water molecule is well positioned for a nucleophilic attack on the carbonyl C of residue P1 of the substrate.[18] Similar to that in step (b), the nucleophilic attack in this step is also favored by the increased electrophilicity of the carbonyl C of residue P1 of the substrate. The water molecule then donates one of its protons to His57 $N^{\varepsilon 2}$ of the peptidase, and the remaining nucleophilic hydroxide ion forms a covalent bond with the carbonyl C of residue P1 of the substrate. As in step (b), the C=O double bond of residue P1 of the substrate becomes a single bond, and the bonding geometry of its carbon becomes tetrahedral. The resulting complex is known as the second tetrahedral intermediate (Figure 1-2d). Molecular structures have been reported for the second tetrahedral intermediates of chymotrypsin[29] and trypsin.[30]

(e) Product release

Finally, Ser195 O^{γ} accepts the proton associated with His57 $N^{\varepsilon 2}$ within the peptidase and becomes a good leaving group. The covalent bond between Ser195 O^{γ} of the peptidase and the carbonyl C of residue P1 of the substrate then dissociates. As in step (c), this drives the restoration of the C=O double bond of residue P1 of the substrate and the trigonal-planar bonding geometry of its carbon. The release of the polypeptide segment N-terminal to the scissile peptide bond of the substrate completes the catalytic cycle of the peptidase (Figure 1-2e). Nonetheless, molecular structures have been reported for the enzyme-product complexes of streptogrisin A,[31] chymotrypsin[32] and trypsin.[30,33]

Interestingly, in addition to the catalytic triad, Ser214 is conserved among chymotrypsin-like serine peptidases. Its carbonyl O forms a hydrogen bond with His57 $C^{\varepsilon 1}$, and its O^{γ} forms a hydrogen bond with Asp102 $O^{\delta 2}$. His57 and Asp102 are

thus well positioned to carry out their functions in the catalytic mechanism.[34] The former hydrogen bond has also been proposed previously to affect the charge distribution of the protonated side chain of His57 in steps (b) and (d), promoting the function of His57 as a general base or a general acid at different stages of the catalytic mechanism.[35]

Chymotrypsin-like serine peptidases provide an excellent example of divergent evolution at the molecular level. The high similarities in amino-acid sequence and molecular structure of the chymotrypsin-like serine peptidases from mammalian pancreas (chymotrypsin, trypsin, elastase and thrombin) suggest that they have evolved from a common ancestral peptidase by gene duplication. Subsequent mutations in the gene duplicates have resulted in substitutions of residues lining the substrate-binding regions (in particular the S1 specificity pockets) of the evolved peptidases, and thus the difference in their substrate specificities.[36] Interestingly, some bacterial serine peptidases (α-lytic peptidase, streptogrisins A and B) are structurally very similar to the chymotrypsin-like serine peptidases from mammalian pancreas, despite the low amino-acid sequence identities (~20 %) among peptidases of the two different origins. This suggests that these peptidases also share a common ancestral peptidase existing before the divergence of prokaryotes and eukaryotes.[37] On the other hand, no significant relationships in amino-acid sequence or molecular structure are observed among the chymotrypsin-like, the subtilisin-like and the serine carboxypeptidase-like serine peptidases, but their active sites and substrate-binding regions can be accurately superimposed. This suggests that these three kinds of serine

peptidases have evolved independently and ended up with a common catalytic mechanism. This is an example of convergent evolution at the molecular level.[22,36,38]

1.3 Cysteine peptidases

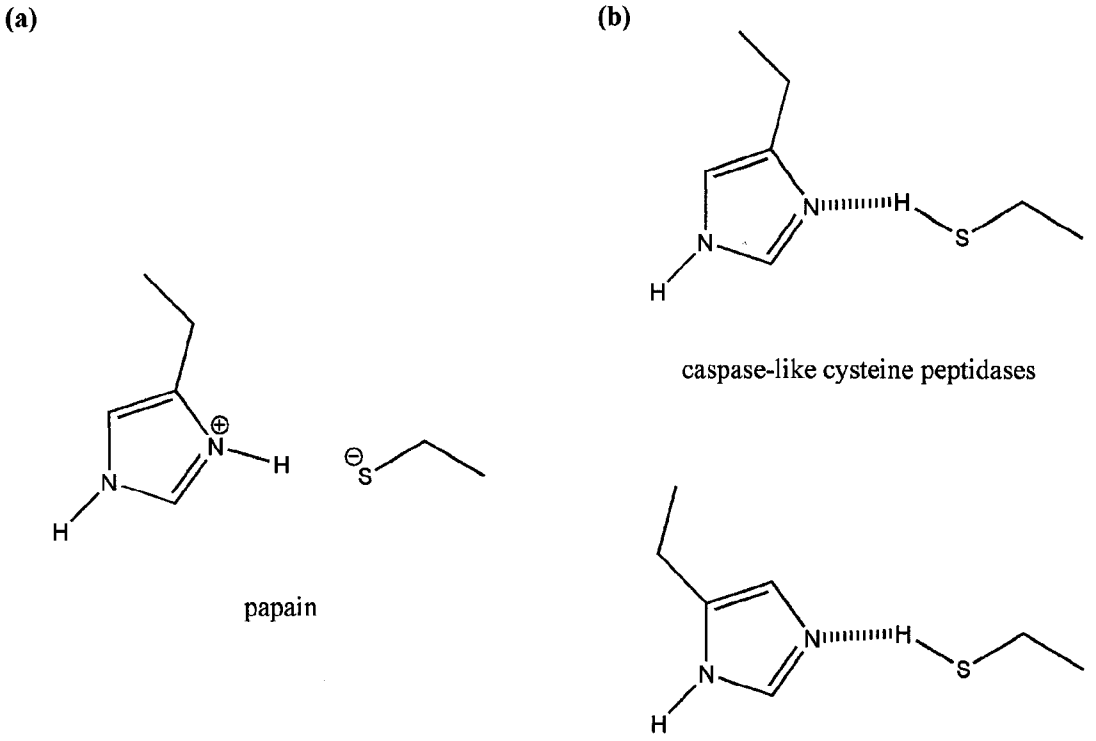
Each cysteine peptidase has a cysteine at its active site, whose side-chain sulfhydryl group acts as a nucleophile in the catalytic mechanism for peptide-bond hydrolysis. Approximately one-fifth of all peptidases and peptidase homologs known so far are cysteine peptidases, grouped into 12 clans and 60 families (MEROPS - the peptidase database, http://merops.sanger.ac.uk/). Folds similar to those of papain and caspases are adopted most frequently by cysteine peptidases.[10] Interestingly, the molecular structures of the 2A and the 3C peptidases from picornaviruses (all being cysteine peptidases) have been found to be chymotrypsin-like.[39,40,41,42] It has been suggested that the viruses have acquired some chymotrypsin-like serine peptidases from their hosts. Substitution of the catalytic Ser by Cys might have occurred in some of these serine peptidases during their evolution within the viruses. In order to include chymotrypsin-like peptidases of both the serine and the cysteine catalytic types, clan PA has been created (MEROPS - the peptidase database, http://merops.sanger.ac.uk/).[4,43]

Despite their structural diversity, cysteine peptidases exhibit similar features in their active sites (e.g. having a Cys and a His as the catalytically essential residues, having a similar spatial arrangement of catalytically essential residues and the presence of an oxyanion hole). This is the other example of convergent evolution at the molecular level. The catalytic mechanisms of cysteine peptidases are similar to that of chymotrypsin-like serine peptidases (Section 1.2). However, it has been shown

that the catalytic Cys and His form a thiolate-imidazolium ion pair for papain in its catalytically competent form (Figure 1-4a). The catalytic His cannot serve as a general base to enhance the nucleophilicity of the catalytic Cys in the formation of the first tetrahedral intermediate of papain.[43] In contrast, the catalytic Cys and His remain as a thiol-imidazole couple for the chymotrypsin-like and the caspase-like cysteine peptidases in their catalytically competent forms (Figure 1-4b).[44,45,46,47] Their catalytic mechanisms probably resemble that of chymotrypsin-like serine peptidases more closely.

Similar to the chymotrypsin-like serine peptidases, most cysteine peptidases have a catalytic triad containing an acidic residue (Asp or Glu) in addition to the catalytic Cys and His. However, the role of this acidic residue in some cysteine peptidases does not seem to correspond functionally to that of Asp102 in chymotrypsin-like serine peptidases (Section 1.2).[43] In the 3C peptidase from the hepatitis A virus, this acidic residue is an Asp. Previous structural studies have shown that its side chain does not form a hydrogen bond with the side chain of the catalytic His.[39,48,49] However, recent structural studies have shown that the $O^{\delta 2}$ of this Asp does form a hydrogen bond with the $N^{\delta 1}$ of the catalytic His, and thus likely plays a similar role as in the chymotrypsin-like serine peptidases.[50,51] The catalytic triad of papain contains an Asn, instead of an acidic residue, in addition to the catalytic Cys and His. The $O^{\delta 1}$ of this Asn is hydrogen-bonded to the $N^{\varepsilon 2}$ of the catalytic His. The contribution of this Asn to the catalytic activity of papain has been demonstrated.[52] Interestingly, caspases have a catalytic dyad containing the catalytic Cys and His only. The $N^{\varepsilon 2}$ of the catalytic His is hydrogen-bonded to the carbonyl O of a nearby

Figure 1-4 Catalytic Cys and His in cysteine peptidases. (a) The thiolate-imidazolium ion pair observed in papain. (b) The thiol-imidazole couple observed in the caspase-like and the chymotrypsin-like cysteine peptidases.



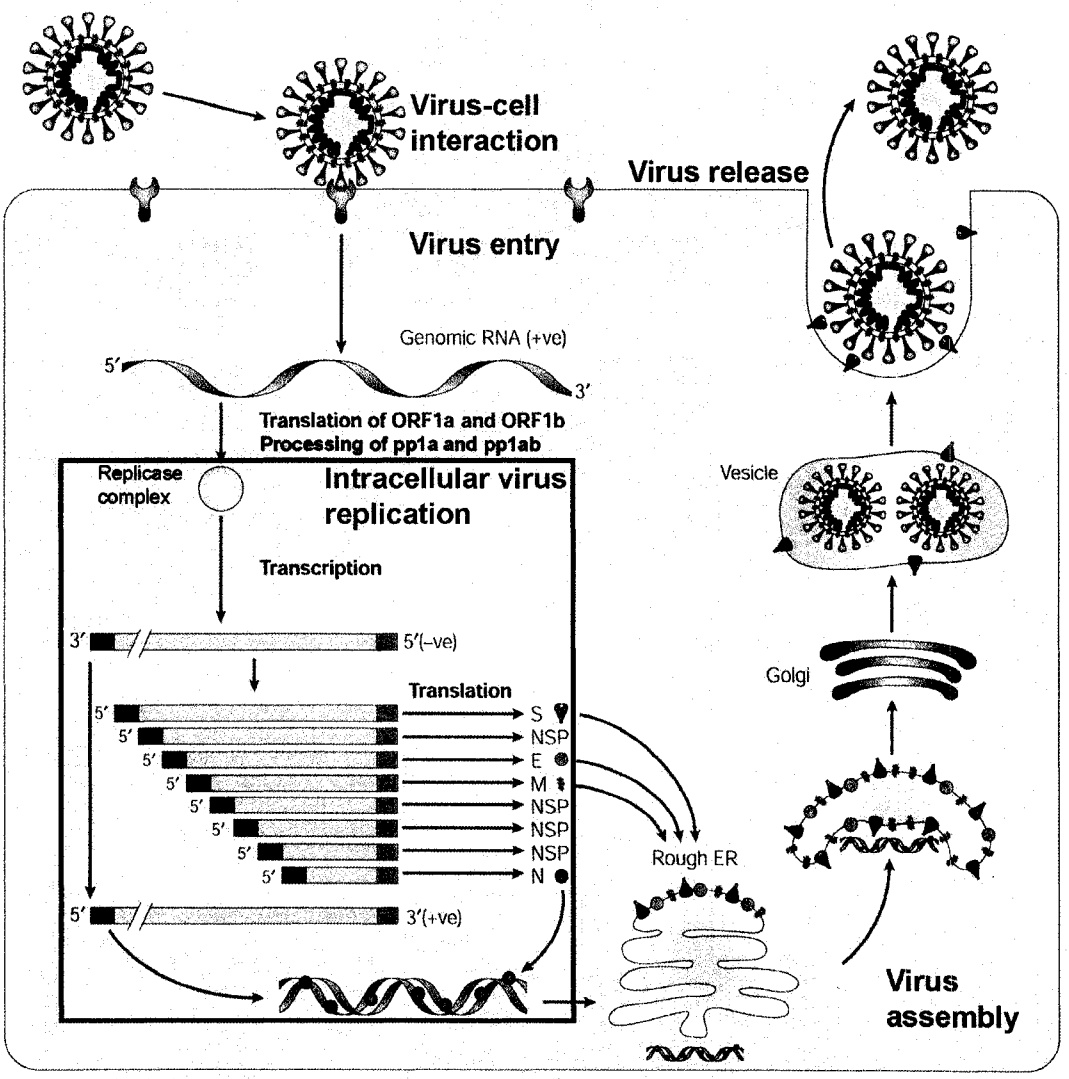
residue.[43] In contrast, gingipain (a caspase-like cysteine peptidase) has a Cys-His-Glu catalytic triad. The $O^{\varepsilon 1}$ and the $O^{\varepsilon 2}$ of the Glu are hydrogen-bonded to the $N^{\varepsilon 2}$ of the catalytic His.[53]

1.4 SARS coronavirus main peptidase

Severe acute respiratory syndrome (SARS) first emerged in China in November 2002. This highly transmissible, infectious and often fatal disease spread to 32 countries across five continents, causing close to 8,500 infections and over 900 deaths, until being contained by the summer of 2003. Several outbreaks of these infections in Asia were reported subsequently (World Health Organization – Severe acute respiratory syndrome, http://www.who.int/csr/sars/en/). Efforts in the development of anti-SARS vaccines and drugs have taken on paramount importance.

SARS is caused by a coronavirus (CoV).[54,55,56] It is an enveloped positive-sense single-stranded RNA virus infecting respiratory and gastrointestinal epithelial cells, macrophages, and other cell types, thereby causing systemic changes and damaging many vital organs such as lung, heart, liver, kidney and adrenal gland.[56,57] Anti-SARS therapeutics could target any one of the several major steps in the viral life cycle, such as the virus-cell interaction, the virus entry, the intracellular viral replication, the virus assembly and the virus release (Figure 1-5).[58,59] Extensive studies have been done on proteins involved in these steps.[60] The intracellular replication of SARS-CoV is mediated by a replicase complex derived from two virally coded polyprotein precursors, pp1a (486 kDa) and pp1ab (790 kDa).[61,62] The formation of this replicase complex requires the extensive processing of the two polyproteins by two cysteine peptidases encoded within them, namely the main

Figure 1-5 Life cycle of SARS-CoV.[58] Several major steps are highlighted in red. (http://mcb.web.psi.ch/teaching/Virology/Posters%2003/SARS_Wahl.htm; adopted with modifications)



peptidase (M^{pro}; also known as the 3C-like peptidase, $3CL^{pro}$, because of its similarity to the 3C peptidases from picornaviruses)[63] and the accessory papain-like peptidase 2 ($PL2^{pro}$).[62] $PL2^{pro}$ cleaves at three sites in the N-proximal regions of the two polyproteins. M^{pro} cleaves at 11 sites in the central and the C-terminal regions of the two polyproteins, releasing key proteins in the viral replication, such as an RNA-dependent RNA polymerase and a helicase (Figure 1-6).[62] As an essential protein, M^{pro} is an attractive molecular target for the development of anti-SARS drugs that act as its inhibitors.

Belonging to peptidase family C30 in clan PA (MEROPS - the peptidase database, http://merops.sanger.ac.uk/), M^{pro} has a molecular mass of 33.8 kDa per protomer. It exists in the catalytically active, homodimeric form over a wide range of concentrations in solution.[64,65,66,67,68,69] The crystal structures of M^{pro} in space group $P2_1$ showed that its two protomers are oriented almost perpendicular to one another. The N-terminal residues 1 to 7 of each protomer constitute the 'N-finger', of which Arg4 was shown to be mandatory for the dimerization and the catalytic activity of M^{pro}.[69] Beyond the N-finger, each protomer consists of three domains. Domains I (residues 8 to 101) and II (residues 102 to 184) comprise a chymotrypsin-like fold. Domain III (residues 201 to 300) has five α-helices and is connected to domain II by a long loop (residues 185 to 200). Each protomer has its own substrate-binding region situated in the cleft between domains I and II (Figure 1-7).[70] A mutagenesis study has confirmed that, similar to the main peptidases from human coronavirus strain 229E[71,72] and porcine transmissible gastroenteritis coronavirus,[73] M^{pro} is a cysteine peptidase with a Cys-His catalytic dyad at the active site.[74] As suggested by the

Figure 1-6 Processing of pp1a and pp1ab by M^{pro} (red arrows) and $PL2^{pro}$ (blue arrows). A helicase domain is found in pp1ab, but not in pp1a. (Reference 62; adopted with modifications)

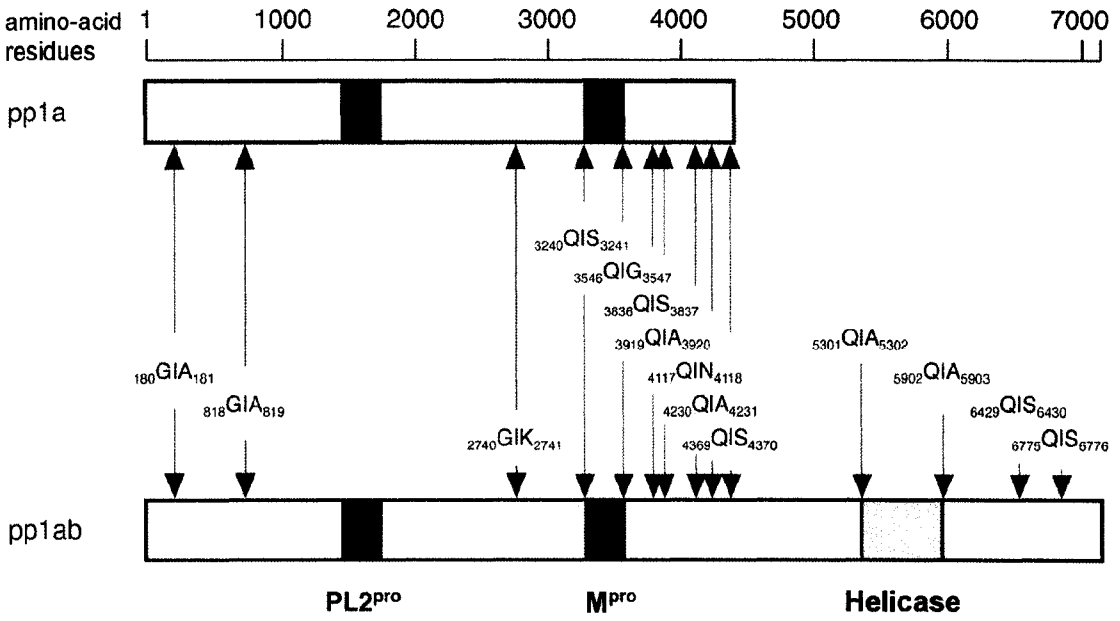
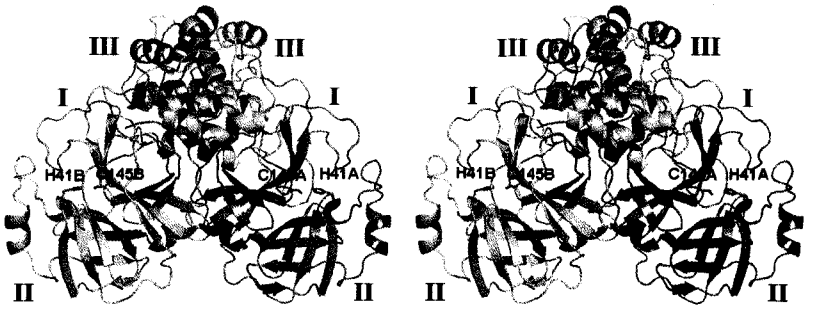


Figure 1-7 Overall structure of M^{pro}.[70] Protomers A (green) and B (cyan) have identical amino-acid sequences. The spatial arrangement of domains I, II and III is indicated for each protomer. The catalytic dyad (His41 and Cys145) is located at the junction of domains I and II of each protomer.



structure-based amino-acid sequence alignment of the main peptidases (including their flanking residues in the polyproteins) from SARS-CoV and other coronaviruses,[72] and confirmed by *in vitro* studies,[62,64] M^{pro} cleaves preferentially at a consensus sequence for residues P4 to P1' of substrates (the arrow indicates the cleavage site): (amino acid with a small side chain)-(any amino acid)-Leu-Gln↓(Ala, Ser, Gly, Asn).

The old drug cinanserin has been identified by virtual screening followed by experimental evaluations as a strong inhibitor of the replication of SARS-CoV, probably through targeting M^{pro}.[75] Also, a number of M^{pro} inhibitors have been proposed using structure-based discovery[76,77] and experimental screening.[78,79,80] Although the efficacies of many of these inhibitors were supported by assay results, the modes of action are unknown for most of them. Crystal structures have been determined for M^{pro} inhibited by only a few peptidomimetic compounds.[70,81,82]

1.5 Aza-peptide epoxide

Aza-peptide epoxides (APEs) were synthesized as a new class of inhibitors apparently specific for cysteine peptidases in peptidase clan CD,[83] including the legumains (family C13)[84] and the caspases (family C14).[85] Each APE has an aza-peptide component, with an epoxide moiety attached to the carbonyl group of residue P1. The side chain of residue P1 determines predominantly the target-peptidase specificity of an APE. The substituent on epoxide atom C2 also allows some tuning of both the inhibitory activity and specificity of APE towards a particular target peptidase. The aza-peptide component resembles a peptide, except that the C^{α} of residue P1 in the former is replaced by a nitrogen atom to form an aza-amino acid

residue. This introduces trigonal planar geometry to the α-atom of residue P1 and reduces the electrophilicity of the carbonyl C of the same residue, thereby making the carbonyl group of residue P1 resistant to nucleophilic attack.[86] It has been proposed that APEs inhibit their target peptidases irreversibly by a mechanism in which the S^{γ} of the catalytic Cys nucleophilically attacks one of the two epoxide carbon atoms (C2 or C3) of APE (Figure 1-8a).[83,84,85] This results in the opening of the conformationally strained epoxide ring, and the formation of a covalent bond between the S^{γ} of the catalytic Cys and the attacked APE atom.

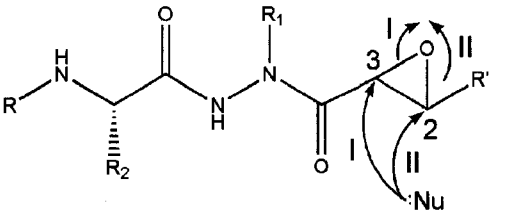
1.6 Streptogrisin B

Streptogrisin B,[87] also known as *Streptomyces griseus* peptidase B (SGPB), is one of the serine peptidases found in Pronase - a commercially available crude enzyme preparation obtained from the extracellular filtrate of *Streptomyces griseus*.[88] Using a synthetic oligonucleotide probe based on the amino-acid sequence of SGPB,[89] the gene encoding this peptidase, *sprB*, was identified in the genomic library of *Streptomyces griseus*.[90] Recent studies have shown that *sprB* is a member of the AdpA regulon in the A-factor regulatory cascade of *Streptomyces griseus*. This regulon is relevant to the secondary metabolism and the morphological differentiation of the bacterium.[91]

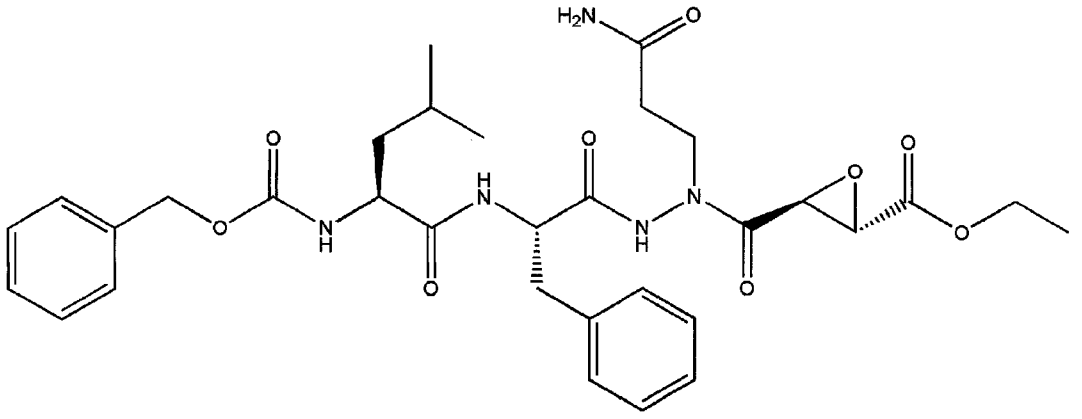
DNA sequence analysis showed that *sprB* encodes a polypeptide chain of 299 residues. However, SGPB obtained from Pronase has a molecular weight of 18.6 kDa, and consists of a single polypeptide chain of 185 residues that originates from the C-terminus of the *sprB*-encoded polypeptide chain.[90] The first 38 residues at the N-terminus of the *sprB*-encoded polypeptide chain constitute a signal peptide (the pre-

Figure 1-8 Aza-peptide epoxide (APE). (a) General design. The epoxide carbon atoms are numbered and their stereochemistries are omitted for simplicity. R and R' represent the groups on the N-terminus and the epoxide atom C2, respectively. R may represent additional amino acid residues to extend the aza-peptide component beyond residue P2. R_1 and R_2 represent the side chains of aza-amino-acid residue P1 and amino-acid residue P2, respectively. The proposed mechanism for the irreversible inhibition of a cysteine peptidase by APE is indicated by arrows (either route I or II). Nu represent a nucleophile (the S^{γ} of the catalytic Cys for a cysteine peptidase).[83,84,85] (b) Cbz-Leu-Phe-AGln-(2*S*,3*S*)EP-COOEt, the APE used in the studies described by Part One of this thesis. Cbz, the benzyloxycarbonyl group; AGln, aza-glutamine; EP, epoxide; COOEt, ethyl ester.

(a)



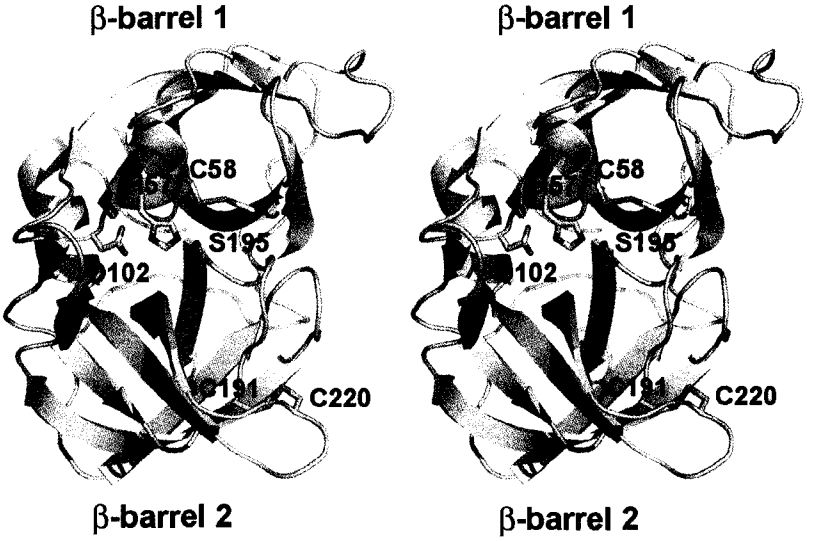
(b)



peptide) that directs the co-translational secretion of the polypeptide chain through the inner membrane of the bacterial cell. This peptide is removed by a membrane-associated signal peptidase.[92] The following 76 residues constitute the pro-peptide (Pro_{SGPB}), which is required for the secretion of SGPB into the extracellular medium.[93] Further studies have shown that Pro_{SGPB} catalyzes the folding of the remaining 185-residue segment at the C-terminus of the *sprB*-encoded polypeptide chain (the mature region), and stabilizes its native (properly folded) state.[94] Very likely, Pro_{SGPB} is proteolytically degraded to release the mature SGPB once the protein has been folded properly and secreted into the extracellular medium, as in the case of a close homolog of SGPB, α-lytic peptidase (αLP) from *Lysobacter enzymogenes*.[95]

SGPB belongs to peptidase family S1 (also known as the chymotrypsin family) in clan PA (MEROPS - the peptidase database, http://merops.sanger.ac.uk/). It has served as one of the models for studying the catalytic activities of serine peptidases. The native crystal structure of SGPB at pH 4.2 has been determined at resolutions up to 2.8 Å.[37,96,97] It adopts a chymotrypsin-like fold, composed of β-barrel 1 (residues Ile16 to Val119), β-barrel 2 (residues Ile124 to Gln229) and an α-helix (residues Val231 to Gly238). Each β-barrel contains one disulfide bond (β-barrel 1: Cys42-Cys58; β-barrel 2: Cys191-Cys220). The active site and the substrate-binding region form in the cleft between the two β-barrels, with the catalytic-triad residues His57 and Asp102 contributed by β-barrel 1, and Ser195 contributed by β-barrel 2 (Figure 1-9). The pH optimum of the catalytic activity of SGPB is reached in the pH range of 7.0 to 8.5.[98] Its S1 specificity for Phe, Tyr, Met and Leu has been demonstrated by

Figure 1-9 Overall structure of SGPB.[37,96,97] The locations of the catalytic triad (His57, Asp102 and Ser195) and the two disulfide bonds (Cys42-Cys58 and Cys191-Cys220) are indicated.



kinetic studies.[99,100] Substitution of Ser195 by Ala or Gly transformed SGPB into a catalyst for peptide ligation.[101]

SGPB and αLP[102,103,104,105,106] are structurally very similar. The glycine contents of both peptidases are high (SGPB: 17 %; αLP: 16 %; the average percentage for the chymotrypsin-like serine peptidases is 9 %). The conformational flexibility of Gly allows for the formation of tight turns in the molecular structures of both peptidases, thereby increasing their compactness and strengthening their intramolecular interactions. As a result, both peptidases exhibit high structural rigidity. The highly restricted dynamic properties of αLP have been demonstrated by X-ray crystallographic[104] and nuclear magnetic resonance (NMR) experiments.[107] The folding/unfolding energy landscapes of both SGPB and αLP show a large and highly cooperative energy barrier between the unfolded and the folded states.[94,108] The high energy costs associated with the sampling of conformations for SGPB probably explain its unusual stability and residual catalytic activity in acidic pH,[109,110] its resistance to guanidinium chloride as a denaturant,[94] and its resistance to proteolysis.[94] Meanwhile, the large and highly cooperative energy barrier to the folding of SGPB necessitates the co-evolution of Pro_{SGPB} as a folding catalyst.[94]

Extensive structural[111,112,113,114,115,116,117,118,119,120] (also Huang *et al.*, unpublished results; PDB accession codes: 2SGF and 1SGY) and thermodynamic studies[109,121,122] have been done on the interactions of turkey ovomucoid third domain (OMTKY3; Section 1.8) and its variants with a panel of six serine peptidases (including SGPB), for the development of a sequence-to-reactivity algorithm (SRA; Section 1.9).[123,124] It has been found that the S1 specificity pocket of SGPB can

accommodate the side chain of any of the 20 coded amino acids,[111,112,115,116,117,119,120] (also Huang *et al.*, unpublished results; PDB accession codes: 2SGF and 1SGY) resulting in a broad range of equilibrium association constants for SGPB with OMTKY3 and its variants (~10^{10} M^{-1} for Leu or Cys to 10^4 M^{-1} for Pro).[122] Computational analysis has shown that 79 % of this variation can be accounted for by the differences in the hydrophobic and the electrostatic interactions.[125]

1.7 Protein inhibitors

As mentioned in Section 1.1, protein inhibitors have evolved to inhibit peptidases exhibiting unwanted activities. Abnormalities in protein inhibitors have been found to underlie diseases such as cancer, emphysema and epilepsy. Some of these diseases may be treated by the administration of synthetic drugs or engineered proteins taking over the activities of protein inhibitors. The design of these therapeutic molecules requires a detailed understanding of peptidase-protein inhibitor interactions. Some abnormalities in protein inhibitors may also be corrected for by gene therapy. Similarly, in agriculture, crop plants may be genetically modified to express protein inhibitors to inhibit the peptidases of pests.[126]

As for peptidases, protein inhibitors have been classified into 33 clans and 53 families to date (MEROPS - the peptidase database, http://merops.sanger.ac.uk/). A clan consists of protein inhibitors with similar three-dimensional structures; whereas a family is a subset of a clan, consisting of protein inhibitors that show statistically significant relationships in amino-acid sequence. Interestingly, in contrast to single-domain protein inhibitors (also known as simple inhibitors), many protein inhibitors (also known as compound inhibitors) have evolved with multiple (2 to 15) domains in

a single polypeptide chain. These domains are not identical, but each of them acts as an inhibitor and has been classified independently. Most compound inhibitors are homotypic, with all domains coming from a single inhibitor family (e.g. turkey ovomucoid; Section 1.8); whereas few compound inhibitors are heterotypic, with domains coming from different inhibitor families (e.g. human eppin). Seven inhibitor families contain members inhibiting peptidases of multiple catalytic types, indicating that structurally dissimilar peptidases can be inhibited by the same mechanism, or that structurally similar protein inhibitors can act by different mechanisms.[126]

A variety of mechanisms have been observed for the inhibition of peptidases by protein inhibitors, including the standard mechanism (e.g. the inhibition of SGPB by OMTKY3), the induction of structural changes by the formation of an acyl-enzyme intermediate (e.g. the inhibition of trypsin by α-1-antitrypsin), the maintenance of the peptidase-product complex (e.g. the inhibition of carboxypeptidase A2 by leech carboxypeptidase inhibitor), the non-productive binding to the peptidase (e.g. the inhibition of caspase-7 by X-linked inhibitor of apoptosis), and the steric blockage of the active site and the substrate binding region of the peptidase (e.g. the inhibition of papain by stefin B).[9,126,127,128,129,130] This thesis concentrates on the standard mechanism, by which many protein inhibitors of serine peptidases act. A protein inhibitor of this type binds in a substrate-like manner to its target serine peptidase, forming a non-covalent peptidase-inhibitor complex analogous to the Michaelis complex (Section 1.2):

$$E + I \leftrightarrow E{:}I \qquad \text{(Equation 1-1)}$$

wherein E, I and E:I represent the peptidase, the inhibitor and the peptidase-inhibitor complex, respectively. An equilibrium association constant (K_{assoc}) can be defined for this peptidase-inhibitor system:

$$K_{assoc} = [E:I] / [E][I] \qquad \text{(Equation 1-2)}$$

Despite their structural diversity, each of the standard-mechanism protein inhibitors of serine peptidases has an exposed and extended loop that binds with an excellent shape complementarity to the substrate-binding region of the target serine peptidase upon the formation of the peptidase-inhibitor complex. A particular peptide bond in this loop, known as the reactive site, is presented to the active site of the peptidase, analogous to the scissile peptide bond of a true substrate. This loop is anchored to the rest of the inhibitor molecule (also known as the inhibitor scaffold) by disulfide bonds and intramolecular hydrogen bonds. This reduces the changes in the main-chain conformational angles and freedom of this loop upon the formation of the peptidase-inhibitor complex. In fact, the reactive-site loops of all standard-mechanism protein inhibitors of serine peptidases have quite consistent values of angles ϕ and ψ, or in other words, they adopt a canonical main-chain conformation, regardless of their states (unbound or peptidase-bound), the amino-acid sequences of their reactive-site loops, the structures of their scaffolds, and the identities of their target peptidases. The reactive-site loop of the inhibitor forms extensive intermolecular interactions with the peptidase in the peptidase-inhibitor complex, as does a true substrate with the peptidase in the Michaelis complex (Section 1.2). Additional interactions with the peptidase are contributed by some residues in the scaffold of the inhibitor.[9,20,21,127,128]

The values of k_{cat}/K_M for the hydrolysis of standard-mechanism canonical protein inhibitors catalyzed by serine peptidases are characteristic of very specific peptidase-substrate interactions. Nonetheless, both K_M and k_{cat} are very small, indicating very strong peptidase-inhibitor interactions and very slow hydrolysis, respectively.[131] The numerous intra- and inter-molecular interactions prevent the movement of the atoms of the reactive-site peptide bond of the inhibitor required for the progression of the catalytic mechanism of the peptidase (Section 1.2). The mechanism can hardly progress beyond the formation of the acyl-enzyme intermediate, because the C-terminal segment of the reactive-site loop of the inhibitor remains bound to the substrate-binding region of the peptidase. The α-amino group of residue P1' of the inhibitor is aimed ideally at the ester bond of the acyl-enzyme intermediate for transpeptidation, favoring the re-formation of the reactive-site peptide bond of the inhibitor. In fact, previous studies have shown that, with the aid of the peptidases, the modified inhibitors (whose reactive-site peptide bond has been hydrolyzed) convert predominantly into the virgin inhibitors (whose reactive-site peptide bond is intact) at low pH under kinetic control.[9,20,21]

Both the virgin and the modified inhibitors can form a complex with the peptidase. The resulting complexes are experimentally indistinguishable.[9,20,21] This can be better described by modifying Equation 1-1:

$$E + I^{\circ} \leftrightarrow (E{:}I^{\circ} \leftrightarrow E{:}I^{*}) \leftrightarrow E + I^{*} \qquad \text{(Equation 1-3)}$$

where I°, I*, E:I° and E:I* represent the virgin inhibitor, the modified inhibitor, the peptidase-virgin inhibitor complex and the peptidase-modified inhibitor complex,

respectively. With more precise definitions for [I] and [E:I], Equation 1-2 remains effective in describing the state of the peptidase-inhibitor system:

$$[I] = [I^\circ] + [I^*] \qquad \text{(Equation 1-4)}$$

$$[E:I] = [E:I^\circ] + [E:I^*] \qquad \text{(Equation 1-5)}$$

1.8 Turkey ovomucoid third domain

Ovomucoid is an abundant single-polypeptide glycoprotein present in avian egg whites. As a homotypic compound inhibitor, it is composed of three domains, all belonging to inhibitor family I1 (also known as the Kazal family) in clan IA (MEROPS - the peptidase database, http://merops.sanger.ac.uk/) and acting as a standard-mechanism canonical protein inhibitor of serine peptidases. Ovomucoids from closely related avian species vary strikingly in their inhibitory activities and specificities.[132] Measurement of the K_{assoc}'s of the ovomucoid third domains alone with a panel of six serine peptidases, including bovine chymotrypsin Aα (CHYM), human leukocyte elastase (HLE), porcine pancreatic elastase (PPE), streptogrisins A and B (SGPA and SGPB), and subtilisin Carlsberg (CARL) gave a huge range of inhibitory activities (K_{assoc}'s from 7.1×10^2 M^{-1} to 1.4×10^{12} M^{-1}) as well as a broad range of specificities.[133] (also Park *et al.*, unpublished results)

Examination of the crystal structures of the ovomucoid third domains from the Japanese quail (wtOMJPQ3)[134,135] and the silver pheasant (wtOMSVP3),[136] and analysis of the amino acid sequences of the ovomucoid third domains from 153 avian species[137,138,139] showed that the structurally important residues are strongly conserved, whereas the residues of the reactive-site loop are hypervariable. Examination of the crystal structures of the wild-type turkey ovomucoid third domain

(wtOMTKY3) bound to SGPB,[111,112] to HLE[113] and to CHYM[114] identified 10 residues of the inhibitor that are in contact with all of these peptidases. It was proposed that the wide spectrum of inhibitory activities and specificities exhibited by ovomucoid third domains is due to the hypervariability of these 10 consensus contact residues only. The structural effects of any changes in these consensus contact residues, and the contributions from the non-contact residues are considered negligible.

wtOMTKY3 consists of 56 residues, with its Leu18I-Glu19I peptide bond (the suffix I is used for residues of the inhibitor) acting as the reactive site in the inhibition of at least nine serine peptidases, including the six mentioned above.[113,140] Kinetic measurements showed that the truncation of the first five N-terminal residues of wtOMTKY3 has no effect on its inhibitory activity.[141] In the crystal structure of the SGPB:wtOMTKY3 complex, the first six N-terminal residues of the inhibitor could not be located in the electron density maps, indicating a high mobility for these residues.[111,112] The crystal structure of the SGPB:OMTKY3 complex, in which the first five N-terminal residues of wtOMTKY3 were truncated, confirmed that the structural effects of this truncation are minor.[115]

1.9 Sequence-to-reactivity algorithm for serine peptidases and their inhibitors

The recovery of the native structure and activity of the reduced and denatured bovine pancreatic ribonuclease A showed clearly that the sequence of a protein determines its structure.[142] The same structure and activity exhibited by the naturally-occurring and by the chemically-synthesized versions of the same enzyme supported

the tenet that the activity of a protein can also be determined by its sequence alone.[143] In view of the possibility of sequence-to-reactivity algorithms (SRAs) being widely applied in the design of drugs and vaccines, and in the engineering of proteins, considerable effort has been directed to the development of SRAs and it has become one of the main goals of certain fields of biochemistry. Most researchers attempt this development by dividing the SRAs into two sub-algorithms: sequence-to-fold and fold-to-activity. However, the challenges facing the development for the sequence-to-fold sub-algorithm present major impediments.[123,124]

OMTKY3 (wtOMTKY3 with its first five N-terminal residues truncated; Figure 1-10) was selected as the basic model for the development of an SRA that determines the inhibitory activity of an OMTKY3 variant from its sequence. Taking advantage of the structural similarity and rigidity of all the OMTKY3 variants, which justifies the use of the additivity approximation,[144,145] an additivity-based SRA was recently developed. The standard free energy change of association for an OMTKY3 variant with any of the six serine peptidases ($\Delta G°_{a\ other}$) is predicted by adding directly to the standard free energy change of association for OMTKY3 with that serine peptidase ($\Delta G°_{a\ OMTKY3}$) the increments due to the substitution of each hypervariable consensus contact residue ($\Sigma_i\ \Delta\Delta G°_{a\ Xi}$; Figure 1-11).[123,124]

In order to test this SRA, a total of 450 values for the standard free energy changes of association ($\Delta G°_a$) with the six serine peptidases were experimentally determined for some natural ovomucoid third domains (398 values), some other natural Kazal inhibitors (17 values) and the OMTKY3 variants previously engineered to inhibit human furin (35 values).[146] Of all these values, 64 % conformed to

Figure 1-10 Covalent structure of OMTKY3. Residues are named using both the conventional (i.e. Val6I to Cys56I, numbers in black) and the Schechter-Berger nomenclatures[19] (i.e. P13-Val to P38'-Cys, numbers in blue). The reactive site (the Leu18I-Glu19I peptide bond) is indicated by the red arrow. Residues Ala15I to Arg21I constitute the reactive-site loop. The ten residues in yellow are the hypervariable consensus contact residues. The two residues in green, although being consensus contact residues as well, are structurally important and invariable. Disulfide bonds are indicated by the black bars.

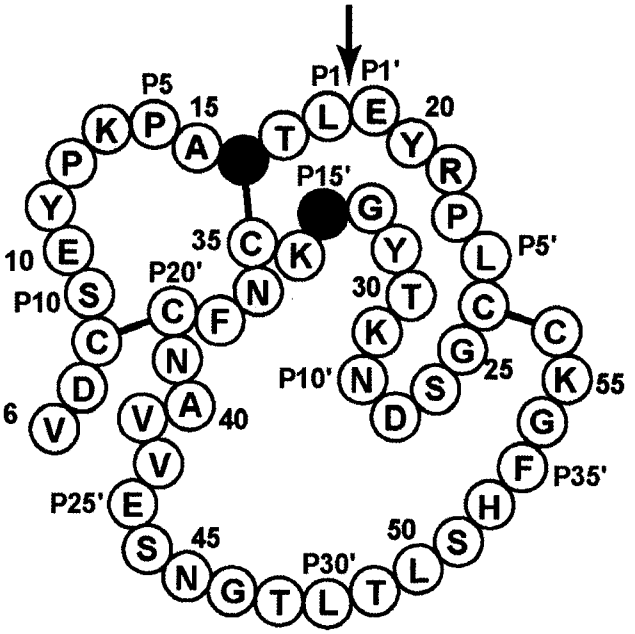
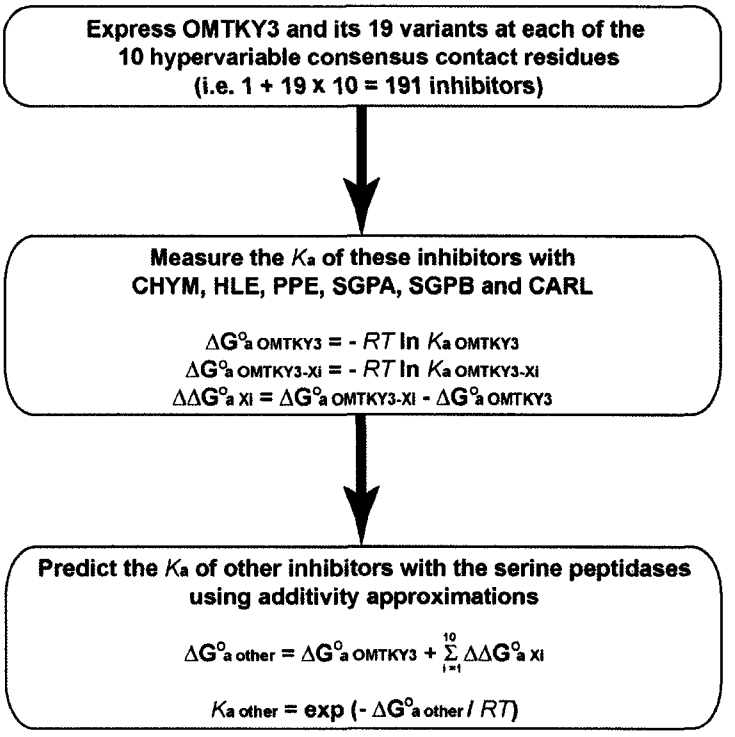


Figure 1-11 Development of the additivity-based SRA for the equilibrium association constants of six serine peptidases with variants of OMTKY3



additivity, 26 % to partial additivity and 9 % were non-additive. The distributions of these cases within the three classes of inhibitors are more or less consistent.[123,124,147]

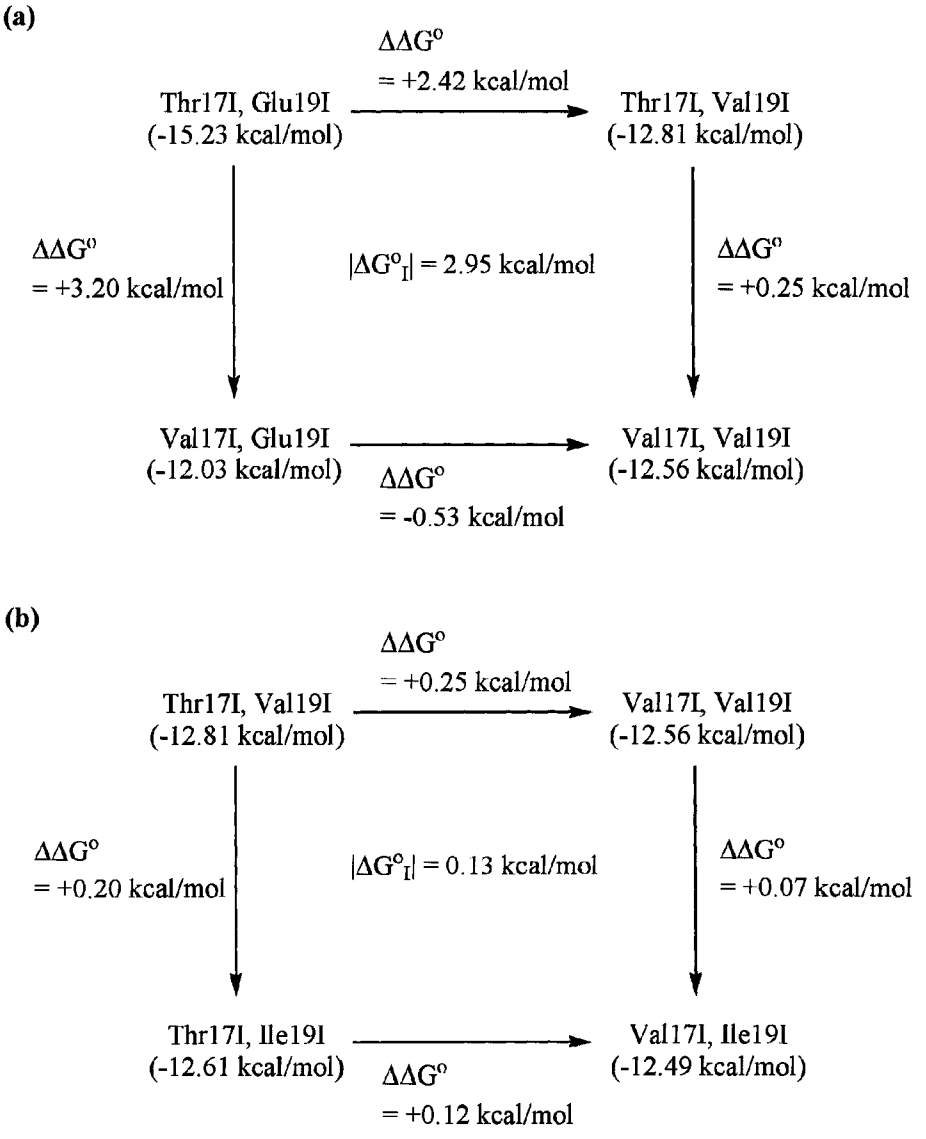
The validity of the additivity approximation requires that the interaction between any two consensus contact residues of an OMTKY3 variant remains unchanged upon the binding of the inhibitor to the six serine peptidases. Very commonly, avian ovomucoid third domains have Thr17I and Glu19I. The crystal structure of the unbound wtOMSVP3 shows a hydrogen bond (3.0 Å) between Thr17I $O^{\gamma 1}$ and Glu19I $O^{\varepsilon 1}$.[136] However, all of the crystal structures of the complexes of OMTKY3 (including the wild type and the residue-18I variants, except the Pro18I variant) with SGPB[111,112,115,116,119,120] and with CHYM[114] (also Ding *et al.*, unpublished results; PDB accession code: 1HJA) determined so far show a shortened hydrogen bond (mean distance: 2.6 Å, standard deviation: 0.1 Å) between the same pair of atoms. Double-mutant cycles illustrate the non-additivity effect in the SRA-based prediction of the $\Delta G°_a$ of CHYM with OMTKY3-Val17IVal19I as a result of the strengthening of this hydrogen bond upon the binding of OMTKY3 to CHYM. Hence, to avoid this non-additivity effect, the use of the SRA is restricted to predictions for only the OMTKY3 variants with either Thr17I or Glu19I (Figure 1-12).[124,147]

1.10 Objectives and organization of this thesis

This thesis is divided into Parts One and Two.

Part One covers the structural studies of SARS-CoV M^{pro}, its variant with an additional Ala at the N-terminus of each protomer ($M^{pro}{}_{+A(-1)}$), and the inhibition of these peptidases by an APE.[148,149] It has been determined that Cbz-Leu-Phe-AGln-

Figure 1-12 Double-mutant cycles.[124] (a) The non-additivity effect (indicated by $|\Delta G°_I| > 0.4$ kcal/mol) in the SRA-based prediction of the $\Delta G°_a$ of CHYM with OMTKY3-Val17IVal19I is due to the strengthening of the hydrogen bond between Thr17I and Glu19I of OMTKY3 upon the binding of OMTKY3 to CHYM. (b) This non-additivity effect is avoided (as indicated by $|\Delta G°_I| < 0.4$ kcal/mol) by restricting the use of the SRA to predictions for only the OMTKY3 variants with either Thr17I or Glu19I.



(2*S*,3*S*)EP-COOEt, an APE possessing an aza-glutamine (AGln) as residue P1 to mimic the S1 specificity of M^{pro} for Gln (Figure 1-8b), inhibited M^{pro} strongly with a k_{inact}/K_i = 1900 ± 400 $M^{-1}s^{-1}$.[148] Chapter 2 reports the molecular structures of the unbound M^{pro} and the M^{pro}:APE complex in space group C2, and the $M^{pro}_{+A(-1)}$:APE complex in space group $P2_12_12_1$. The mode of action for APE inhibiting M^{pro} and $M^{pro}_{+A(-1)}$, and the peptidase-inhibitor interactions are analyzed in details. The importance of the `N-fingers' in the dimerization of M^{pro} and $M^{pro}_{+A(-1)}$, and in the structural integrity of their S1 specificity pockets are examined.

Chapter 3 reports the molecular structures of the unbound $M^{pro}_{+A(-1)}$ and the $M^{pro}_{+A(-1)}$:APE complex in space group $P4_32_12$, and the unbound M^{pro} in space group $P2_1$. Considerable efforts have been directed from all over the world to the structural studies of M^{pro}, resulting in the availability of many molecular structures of M^{pro} and its variants over a pH range of 5.9 to 9.0 in a variety of space groups. Comparisons among these molecular structures suggest that the active sites and the S1 specificity pockets of M^{pro} are conformationally flexible, and that there is a dynamic equilibrium between the catalytically competent and the catalytically incompetent conformations of M^{pro}. Factors affecting the position of this equilibrium are discussed. In particular, the effects of APE binding suggest that the substrate binds to M^{pro} in an induced-fit manner.

Part Two covers the structural studies of SGPB, OMTKY3, the Ala32I variant of OMTKY3 (OMTKY3-Ala32I), and the SGPB:OMTKY3-Ala32I complex.[150,151] The molecular structure of the native SGPB has been determined so far at a relatively low resolution (2.8 Å) and at pH 4.2 only,[37,96,97] whereas those of the complexes of

SGPB with OMTKY3 and its variants have been determined at much higher resolution (higher than 2 Å).[111,112,115,116,117,119,120] (also Huang *et al.*, unpublished results; PDB accession codes: 2SGF and 1SGY) In order to place structural comparisons on an equal footing, the molecular structures of the native SGPB should be re-determined at comparable or higher resolutions. Chapter 4 reports the molecular structures of SGPB at the highest resolutions its crystals can afford (~1.2 Å) at both pH 4.2 and 7.3. Unexpectedly, additional electron density peaks were observed at the active site and in the substrate-binding region of SGPB in the computed maps at both pH values. Peak assignment followed by structure refinement with relaxed geometric restraints resulted in structures representing mixtures of the second tetrahedral intermediates and the enzyme-product complexes of SGPB existing in a pH-dependent equilibrium. Structural comparisons of SGPB in different conditions provide insights into the movement of atoms required for the progression of the catalytic mechanism of the chymotrypsin-like serine peptidases.

The molecular structures of both the SGPB:wtOMTKY3 and the SGPB:OMTKY3 complexes show that the main-chain regions of Leu18I and Gly32I, and the side chains of Tyr20I and Asn36I of the inhibitor interact with one another indirectly via Pro192BE of the peptidase.[111,112,115] In the unbound form, the interactions among these inhibitor residues would be different, and this difference would be the other origin of non-additivity effects in the SRA. In fact, non-additivity effects were found between residues 18I and 32I,[152] and between residues 20I and 32I[153] in the SRA-based predictions of the $\Delta G°_a$ of SGPB with OMTKY3 variants. To appreciate how residue 32I contributes to these non-additivity effects, Gly32I has

been changed to Ala for the addition of a small side chain (C^{β}) and the decrease in the main-chain conformational flexibility. This substitution retains essentially the tightness of the SGPB:OMTKY3 complex (K_{assoc} is 5.6×10^{10} M^{-1} for the SGPB:OMTKY3 complex, and 7.1×10^{9} M^{-1} for the SGPB:OMTKY3-Ala32I complex; both values were determined at 21 ± 2 C°, pH 8.3). This is advantageous to the crystallization of this complex. Chapter 5 reports the molecular structures of the unbound OMTKY3, the unbound OMTKY3-Ala32I and the SGPB:OMTKY3-Ala32I complex. Structural comparisons with the unbound inhibitors confirm our understanding of some previously addressed non-additivity effects. The Gly-to-Ala substitution does not change the main-chain conformational angles and freedom of residue 32I of OMTKY3, but it does change the stoichiometry and the conformation of the SGPB:OMTKY3 complex. The causes and the effects of these changes, particularly their effects on the validity of the SRA, are discussed.

1.11 References

1. Lopez-Otin, C. & Overall, C. M. (2002). Protease degradomics: a new challenge for proteomics. *Nat. Rev. Mol. Cell. Biol.* **3**, 509-519.

2. Smith, R. M. & Hansen, D. E. (1998). The pH-rate profile for the hydrolysis of a peptide bond. *J. Am. Chem. Soc.* **120**, 8910-8913.

3. Rawlings, N. D., Morton, F. R. & Barrett, A. J. (2006). MEROPS: the peptidase database. *Nucleic Acids Res.* **34**, D270-D272.

4. Gerhartz, B., Niestroj, A. J. & Demuth, H. (2002). Enzyme classes and mechanisms. In *Proteinase and peptidase inhibition: recent potential targets for drug development* (Smith, H. J., ed.), pp. 17-36. Taylor & Francis, Florence.

5. NC-IUBMB. (1992). *Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*, Academic Press, Orlando.

6. Hartley, B. S. (1960). Proteolytic enzymes. *Annu. Rev. Biochem.* **29**, 45-72.

7. Rawlings, N. D. & Barrett, A. J. (1993). Evolutionary families of peptidases. *Biochem. J.* **290**, 205-218.

8. Ellerbroek, S. M., Wu, Y. & Stack, M. S. (2002). Regulatory mechanisms for proteinase activity. In *Proteinase and peptidase inhibition: recent potential targets for drug development* (Smith, H. J., ed.), pp. 37-50. Taylor & Francis, Florence.

9. Otlewski, J., Krowarsch, D. & Apostoluk, W. (1999). Protein inhibitors of serine proteinases. *Acta Biochim. Pol.* **46**, 531-565.

10. Tyndall, J. D., Nall, T. & Fairlie, D. P. (2005). Proteases universally recognize beta strands in their active sites. *Chem. Rev.* **105**, 973-999.

11. Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967). Three-dimensional structure of tosyl-alpha-chymotrypsin. *Nature* **214**, 652-656.

12. Carter, P. & Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564-568.

13. Craik, C. S., Roczniak, S., Largman, C. & Rutter, W. J. (1987). The catalytic role of the active site aspartic acid in serine proteases. *Science* **237**, 909-913.

14. Corey, D. R. & Craik, C. S. (1992). An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *J. Am. Chem. Soc.* **114**, 1784-1790.

15. Polgar, L. (2004). Catalytic mechanisms of serine and threonine peptidases. In *Handbook of Proteolytic Enzymes* 2nd edit. (Barrett, A. J., Rawlings, N. D. & Woessner, J. F., eds.), Vol. 1, pp. 1440-1448. Elsevier, London.

16. James, M. N. G., Delbaere, L. T. & Brayer, G. D. (1978). Amino acid sequence alignment of bacterial and mammalian pancreatic serine proteases based on topological equivalences. *Can. J. Biochem.* **56**, 396-402.

17. Radisky, E. S. & Koshland, D. E., Jr. (2002). A clogged gutter mechanism for protease inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10316-10321.

18. Radisky, E. S., Lee, J. M., Lu, C. J. & Koshland, D. E., Jr. (2006). Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6835-6840.

19. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162.

20. Laskowski, M., Jr. & 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**, 324-337.

21. Laskowski, M. J., Qasim, M. A. & Lu, S. M. (2000). Interaction of standard mechanism, canonical protein inhibitors with serine proteinases. In *Protein-protein recognition* (Kleanthous, C., ed.), pp. 228-279. Oxford University Press, New York.

22. Branden, C. & Tooze, J. (1999). *Introduction to Protein Structure*. 2nd edit, Garland, New York.

23. Nienaber, V. L., Breddam, K. & Birktoft, J. J. (1993). A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry* **32**, 11469-11475.

24. Dixon, M. M. & Matthews, B. W. (1989). Is gamma-chymotrypsin a tetrapeptide acyl-enzyme adduct of alpha-chymotrypsin? *Biochemistry* **28**, 7033-7038.

25. Dixon, M. M., Brennan, R. G. & Matthews, B. W. (1991). Structure of gamma-chymotrypsin in the range pH 2.0 to pH 10.5 suggests that gamma-

chymotrypsin is a covalent acyl-enzyme adduct at low pH. *Int. J. Biol. Macromol.* **13**, 89-96.

26. Singh, N., Jabeen, T., Sharma, S., Roy, I., Gupta, M. N., Bilgrami, S., Somvanshi, R. K., Dey, S., Perbandt, M., Betzel, C., Srinivasan, A. & Singh, T. P. (2005). Detection of native peptides as potent inhibitors of enzymes. Crystal structure of the complex formed between treated bovine alpha-chymotrypsin and an autocatalytically produced fragment, Ile-Val-Asn-Gly-Glu-Glu-Ala-Val-Pro-Gly-Ser-Trp-Pro-Trp, at 2.2 Å resolution. *FEBS J.* **272**, 562-572.

27. Blanchard, H. & James, M. N. G. (1994). A crystallographic re-investigation into the structure of *Streptomyces griseus* proteinase A reveals an acyl-enzyme intermediate. *J. Mol. Biol.* **241**, 574-587.

28. Perona, J. J., Craik, C. S. & Fletterick, R. J. (1993). Locating the catalytic water molecule in serine proteases. *Science* **261**, 620-622.

29. Yennawar, N. H., Yennawar, H. P. & Farber, G. K. (1994). X-ray crystal structure of gamma-chymotrypsin in hexane. *Biochemistry* **33**, 7326-7336.

30. Rypniewski, W. R., Østergaard, P. R., Nørregaard-Madsen, M., Dauter, M. & Wilson, K. S. (2001). *Fusarium oxysporum* trypsin at atomic resolution at 100 and 283 K: a study of ligand binding. *Acta Cryst.* **D57**, 8-19.

31. James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. & Bauer, C. A. (1980). Structures of product and inhibitor complexes of Streptomyces griseus protease A at 1.8 A resolution. A model for serine protease catalysis. *J. Mol. Biol.* **144**, 43-88.

32. Harel, M., Su, C. T., Frolow, F., Silman, I. & Sussman, J. L. (1991). Gamma-chymotrypsin is a complex of alpha-chymotrypsin with its own autolysis products. *Biochemistry* **30**, 5217-5225.

33. Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). Trypsin revisited: crystallography at (sub)atomic resolution and quantum chemistry revealing details of catalysis. *J. Biol. Chem.* **278**, 43357-43362.

34. Krem, M. M. & Di Cera, E. (2001). Molecular markers of serine protease evolution. *EMBO J.* **20**, 3036-3045.

35. Derewenda, Z. S., Derewenda, U. & Kobos, P. M. (1994). (His) C(epsilon)-H...O=C < hydrogen bond in the active sites of serine hydrolases. *J. Mol. Biol.* **241**, 83-93.

36. Neurath, H. (1984). Evolution of proteolytic enzymes. *Science* **224**, 350-357.

37. Delbaere, L. T., Hutcheon, W. L., James, M. N. G. & Thiessen, W. E. (1975). Tertiary structural differences between microbial serine proteases and pancreatic serine enzymes. *Nature* **257**, 758-763.

38. Voet, D. & Voet, J. G. (1995). *Biochemistry*. 2nd edit, John Wiley & Sons, New York.

39. Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **369**, 72-76.

40. Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L.

& Worland, S. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* **77**, 761-771.

41. Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. G. (1997). Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* **273**, 1032-1047.

42. Petersen, J. F., Cherney, M. M., Liebig, H. D., Skern, T., Kuechler, E. & James, M. N. G. (1999). The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J.* **18**, 5463-5475.

43. Polgar, L. (2004). Catalytic mechanisms of cysteine peptidases. In *Handbook of Proteolytic Enzymes* 2nd edit. (Barrett, A. J., Rawlings, N. D. & Woessner, J. F., eds.), Vol. 1, pp. 1072-1079. Elsevier, London.

44. Sarkany, Z., Skern, T. & Polgar, L. (2000). Characterization of the active site thiol group of rhinovirus 2A proteinase. *FEBS Lett.* **481**, 289-292.

45. Sarkany, Z., Szeltner, Z. & Polgar, L. (2001). Thiolate-imidazolium ion pair is not an obligatory catalytic entity of cysteine peptidases: the active site of picornain 3C. *Biochemistry* **40**, 10601-10606.

46. Csoma, C. & Polgar, L. (1984). Proteinase from germinating bean cotyledons. Evidence for involvement of a thiol group in catalysis. *Biochem J.* **222**, 769-776.

47. Kembhavi, A. A., Buttle, D. J., Rauber, P. & Barrett, A. J. (1991). Clostripain: characterization of the active site. *FEBS Lett.* **283**, 277-280.

48. Bergmann, E. M., Mosimann, S. C., Chernaia, M. M., Malcolm, B. A. & James, M. N. G. (1997). The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *J. Virol.* **71**, 2436-2448.

49. Bergmann, E. M., Cherney, M. M., McKendrick, J., Frormann, S., Luo, C., Malcolm, B. A., Vederas, J. C. & James, M. N. G. (1999). Crystal structure of an inhibitor complex of the 3C proteinase from hepatitis A virus (HAV) and implications for the polyprotein processing in HAV. *Virology* **265**, 153-163.

50. Yin, J., Bergmann, E. M., Cherney, M. M., Lall, M. S., Jain, R. P., Vederas, J. C. & James, M. N. G. (2005). Dual modes of modification of hepatitis A virus 3C protease by a serine-derived beta-lactone: selective crystallization and formation of a functional catalytic triad in the active site. *J. Mol. Biol.* **354**, 854-871.

51. Yin, J., Cherney, M. M., Bergmann, E. M., Zhang, J., Huitema, C., Pettersson, H., Eltis, L. D., Vederas, J. C. & James, M. N. G. (2006). An episulfide cation (thiiranium ring) trapped in the active site of HAV 3C proteinase inactivated by peptide-based ketone inhibitors. *J. Mol. Biol.* **361**, 673-686.

52. Vernet, T., Tessier, D. C., Chatellier, J., Plouffe, C., Lee, T. S., Thomas, D. Y., Storer, A. C. & Menard, R. (1995). Structural and functional roles of asparagine 175 in the cysteine protease papain. *J. Biol. Chem.* **270**, 16645-16652.

53. Eichinger, A., Beisel, H. G., Jacob, U., Huber, R., Medrano, F. J., Banbula, A., Potempa, J., Travis, J. & Bode, W. (1999). Crystal structure of gingipain R: an

Arg-specific bacterial cysteine proteinase with a caspase-like fold. *EMBO J.* **18**, 5453-5462.

54. Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W., Cheung, M. T., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K. & Yuen, K. Y. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319-1325.

55. Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguiere, A. M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D., Schmitz, H. & Doerr, H. W. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967-1976.

56. Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang, J. Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J. & Anderson, L. J. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953-1966.

57. Ding, Y., Wang, H., Shen, H., Li, Z., Geng, J., Han, H., Cai, J., Li, X., Kang, W., Weng, D., Lu, Y., Wu, D., He, L. & Yao, K. (2003). The clinical pathology of severe acute respiratory syndrome (SARS): a report from China. *J. Pathol.* **200**, 282-289.

58. Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H. D. & Rappuoli, R. (2003). SARS--beginning to understand a new virus. *Nat. Rev. Microbiol.* **1**, 209-218.

59. Groneberg, D. A., Hilgenfeld, R. & Zabel, P. (2005). Molecular mechanisms of severe acute respiratory syndrome (SARS). *Respir. Res.* **6**, 8.

60. Bartlam, M., Yang, H. & Rao, Z. (2005). Structural insights into SARS coronavirus proteins. *Curr. Opin. Struct. Biol.* **15**, 664-672.

61. Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. (2001). Viral replicase gene products suffice for coronavirus discontinuous transcription. *J. Virol.* **75**, 6676-6681.

62. Thiel, V., Ivanov, K. A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E. & Ziebuhr, J. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305-2315.

63. Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**, 853-879.

64. Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J. & Lai, L. (2004). Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **279**, 1637-1642.

65. Shi, J., Wei, Z. & Song, J. (2004). Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra

domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. *J. Biol. Chem.* **279**, 24765-24773.

66. Kuo, C. J., Chi, Y. H., Hsu, J. T. & Liang, P. H. (2004). Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem. Biophys. Res. Commun.* **318**, 862-867.

67. Chou, C. Y., Chang, H. C., Hsu, W. C., Lin, T. Z., Lin, C. H. & Chang, G. G. (2004). Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry* **43**, 14958-14970.

68. Chen, S., Chen, L., Tan, J., Chen, J., Du, L., Sun, T., Shen, J., Chen, K., Jiang, H. & Shen, X. (2005). Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization: biochemical and thermodynamic investigation in conjunction with molecular dynamic simulations. *J. Biol. Chem.* **280**, 164-173.

69. Hsu, W. C., Chang, H. C., Chou, C. Y., Tsai, P. J., Lin, P. I. & Chang, G. G. (2005). Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *J. Biol. Chem.* **280**, 22741-22748.

70. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13190-13195.

71. Ziebuhr, J., Heusipp, G. & Siddell, S. G. (1997). Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J. Virol.* **71**, 3992-3997.

72. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763-1767.

73. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **21**, 3213-3224.

74. Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. (2004). 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* **43**, 4568-4574.

75. Chen, L., Gui, C., Luo, X., Yang, Q., Gunther, S., Scandella, E., Drosten, C., Bai, D., He, X., Ludewig, B., Chen, J., Luo, H., Yang, Y., Yang, Y., Zou, J., Thiel, V., Chen, K., Shen, J., Shen, X. & Jiang, H. (2005). Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro. *J. Virol.* **79**, 7095-7103.

76. Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A. & Freire, E. (2004). Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry* **43**, 4906-4912.

77. Jain, R. P., Pettersson, H. I., Zhang, J., Aull, K. D., Fortin, P. D., Huitema, C., Eltis, L. D., Parrish, J. C., James, M. N., Wishart, D. S. & Vederas, J. C. (2004). Synthesis and evaluation of keto-glutamine analogues as potent inhibitors of severe acute respiratory syndrome 3CLpro. *J. Med. Chem.* **47**, 6113-6116.

78. Wu, C. Y., Jan, J. T., Ma, S. H., Kuo, C. J., Juan, H. F., Cheng, Y. S., Hsu, H. H., Huang, H. C., Wu, D., Brik, A., Liang, F. S., Liu, R. S., Fang, J. M., Chen, S. T., Liang, P. H. & Wong, C. H. (2004). Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10012-10017.

79. Kao, R. Y., Tsui, W. H., Lee, T. S., Tanner, J. A., Watt, R. M., Huang, J. D., Hu, L., Chen, G., Chen, Z., Zhang, L., He, T., Chan, K. H., Tse, H., To, A. P., Ng, L. W., Wong, B. C., Tsoi, H. W., Yang, D., Ho, D. D. & Yuen, K. Y. (2004). Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetics. *Chem. Biol.* **11**, 1293-1299.

80. Blanchard, J. E., Elowe, N. H., Huitema, C., Fortin, P. D., Cechetto, J. D., Eltis, L. D. & Brown, E. D. (2004). High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. *Chem. Biol.* **11**, 1445-1453.

81. Ghosh, A. K., Xi, K., Ratia, K., Santarsiero, B. D., Fu, W., Harcourt, B. H., Rota, P. A., Baker, S. C., Johnson, M. E. & Mesecar, A. D. (2005). Design

and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. *J. Med. Chem.* **48**, 6767-6771.

82. Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Yuen, K. Y., Wong, L., Gao, G., Chen, S., Chen, Z., Ma, D., Bartlam, M. & Rao, Z. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **3**, e324.

83. Asgian, J. L., James, K. E., Li, Z. Z., Carter, W., Barrett, A. J., Mikolajczyk, J., Salvesen, G. S. & Powers, J. C. (2002). Aza-peptide epoxides: a new class of inhibitors selective for clan CD cysteine proteases. *J. Med. Chem.* **45**, 4958-4960.

84. James, K. E., Gotz, M. G., Caffrey, C. R., Hansell, E., Carter, W., Barrett, A. J., McKerrow, J. H. & Powers, J. C. (2003). Aza-peptide epoxides: potent and selective inhibitors of Schistosoma mansoni and pig kidney legumains (asparaginyl endopeptidases). *Biol. Chem.* **384**, 1613-1618.

85. James, K. E., Asgian, J. L., Li, Z. Z., Ekici, O. D., Rubin, J. R., Mikolajczyk, J., Salvesen, G. S. & Powers, J. C. (2004). Design, synthesis, and evaluation of aza-peptide epoxides as selective and potent inhibitors of caspases-1, -3, -6, and -8. *J. Med. Chem.* **47**, 1553-1574.

86. Powers, J. C., Asgian, J. L., Ekici, O. D. & James, K. E. (2002). Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* **102**, 4639-4750.

87. Qasim, M. A. (2004). Streptogrisin B. In *Handbook of proteolytic enzymes* 2nd edit. (Barrett, A. J., Rawlings, N. D. & Woessner, J. F., Jr., eds.), pp. 1458-1462. Elsevier, London.

88. Jurášek, J., Johnson, P., Olafson, R. W. & Smillie, L. B. (1971). An improved fractionation system for Pronase on CM-sephadex. *Can. J. Biochem.* **49**, 1195-1201.

89. Jurášek, L., Carpenter, M. R., Smillie, L. B., Gertler, A., Levy, S. & Ericsson, L. H. (1974). Amino acid sequence of *Streptomyces griseus* protease B, a major component of Pronase. *Biochem. Biophys. Res. Commun.* **61**, 1095-1100.

90. Henderson, G., Krygsman, P., Liu, C. J., Davey, C. C. & Malek, L. T. (1987). Characterization and structure of genes for proteases A and B from *Streptomyces griseus*. *J. Bacteriol.* **169**, 3778-3784.

91. Tomono, A., Tsai, Y., Ohnishi, Y. & Horinouchi, S. (2005). Three chymotrypsin genes are members of the AdpA regulon in the A-factor regulatory cascade in *Streptomyces griseus*. *J. Bacteriol.* **187**, 6341-6353.

92. von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683-4690.

93. Baardsnes, J., Sidhu, S., MacLeod, A., Elliott, J., Morden, D., Watson, J. & Borgford, T. (1998). *Streptomyces griseus* protease B: secretion correlates with the length of the propeptide. *J. Bacteriol.* **180**, 3241-3244.

94. Truhlar, S. M., Cunningham, E. L. & Agard, D. A. (2004). The folding landscape of *Streptomyces griseus* protease B reveals the energetic costs and benefits associated with evolving kinetic stability. *Protein Sci.* **13**, 381-390.

95. Cunningham, E. L. & Agard, D. A. (2004). Disabling the folding catalyst is the last critical step in alpha-lytic protease folding. *Protein Sci.* **13**, 325-331.

96. Codding, P. W., Delbaere, L. T., Hayakawa, K., Hutcheon, W. L., James, M. N. G. & Jurášek, L. (1974). The 4.5 Å resolution structure of a bacterial serine protease from *Streptomyces griseus*. *Can. J. Biochem.* **52**, 208-220.

97. Delbaere, L. T., Brayer, G. D. & James, M. N. G. (1979). The 2.8 Å resolution structure of *Streptomyces griseus* protease B and its homology with alpha-chymotrypsin and *Streptomyces griseus* protease A. *Can. J. Biochem.* **57**, 135-144.

98. Gertler, A. (1974). Inhibition of *Streptomyces griseus* protease B by peptide chloromethyl ketones: partial mapping of the binding site and identification of the reactive residue. *FEBS Lett.* **43**, 81-85.

99. Bauer, C. A. (1978). Active centers of *Streptomyces griseus* protease 1, *Streptomyces griseus* protease 3, and alpha-chymotrypsin: enzyme-substrate interactions. *Biochemistry* **17**, 375-380.

100. Sidhu, S. S. & Borgford, T. J. (1996). Selection of *Streptomyces griseus* protease B mutants with desired alterations in primary specificity using a library screening strategy. *J. Mol. Biol.* **257**, 233-245.

101. Elliott, R. J., Bennet, A. J., Braun, C. A., MacLeod, A. M. & Borgford, T. J. (2000). Active-site variants of *Streptomyces griseus* protease B with peptide-ligation activity. *Chem. Biol.* **7**, 163-171.

102. Brayer, G. D., Delbaere, L. T. & James, M. N. G. (1979). Molecular structure of the alpha-lytic protease from Myxobacter 495 at 2.8 Å resolution. *J. Mol. Biol.* **131**, 743-775.

103. Fujinaga, M., Delbaere, L. T., Brayer, G. D. & James, M. N. G. (1985). Refined structure of alpha-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. *J. Mol. Biol.* **184**, 479-502.

104. Rader, S. D. & Agard, D. A. (1997). Conformational substates in enzyme mechanism: the 120 K structure of alpha-lytic protease at 1.5 Å resolution. *Protein Sci.* **6**, 1375-1386.

105. Fuhrmann, C. N., Kelch, B. A., Ota, N. & Agard, D. A. (2004). The 0.83 Å resolution crystal structure of alpha-lytic protease reveals the detailed structure of the active site and identifies a source of conformational strain. *J. Mol. Biol.* **338**, 999-1013.

106. Fuhrmann, C. N., Daugherty, M. D. & Agard, D. A. (2006). Subangstrom crystallography reveals that short ionic hydrogen bonds, and not a His-Asp low-barrier hydrogen bond, stabilize the transition state in serine protease catalysis. *J. Am. Chem. Soc.* **128**, 9086-9102.

107. Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. (2002). Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* **415**, 343-346.

108. Sohl, J. L., Jaswal, S. S. & Agard, D. A. (1998). Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* **395**, 817-819.

109. Qasim, M. A., Ranjbar, M. R., Wynn, R., Anderson, S. & Laskowski, M., Jr. (1995). Ionizable P1 residues in serine proteinase inhibitors undergo large pK shifts on complex formation. *J. Biol. Chem.* **270**, 27419-27422.

110. Ardelt, W. & Laskowski, M., Jr. (1991). Effect of single amino acid replacements on the thermodynamics of the reactive site peptide bond hydrolysis in ovomucoid third domain. *J. Mol. Biol.* **220**, 1041-1053.

111. Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M., Jr. & James, M. N. G. (1982). Refined crystal structure of the molecular complex of *Streptomyces griseus* protease B, a serine protease, with the third domain of the ovomucoid inhibitor from turkey. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4868-4872.

112. Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. (1983). Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-Å resolution. *Biochemistry* **22**, 4420-4433.

113. Bode, W., Wei, A. Z., Huber, R., Meyer, E., Travis, J. & Neumann, S. (1986). X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J.* **5**, 2453-2458.

114. Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M., Jr. & James, M. N. G. (1987). Crystal and molecular structures of the complex of

alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J. Mol. Biol.* **195**, 397-418.

115. Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr. & James, M. N. G. (1995). Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci.* **4**, 1985-1997.

116. Bateman, K. S., Anderson, S., Lu, W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2000). Deleterious effects of beta-branched residues in the S1 specificity pocket of *Streptomyces griseus* proteinase B (SGPB): crystal structures of the turkey ovomucoid third domain variants Ile18I, Val18I, Thr18I, and Ser18I in complex with SGPB. *Protein Sci.* **9**, 83-94.

117. Bateman, K. S., Huang, K., Anderson, S., Lu, W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2001). Contribution of peptide bonds to inhibitor-protease binding: crystal structures of the turkey ovomucoid third domain backbone variants OMTKY3-Pro18I and OMTKY3-psi[COO]-Leu18I in complex with *Streptomyces griseus* proteinase B (SGPB) and the structure of the free inhibitor, OMTKY-3-psi[CH2NH2+]-Asp19I. *J. Mol. Biol.* **305**, 839-849.

118. Maynes, J. T., Cherney, M. M., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2005). Structure of the subtilisin Carlsberg-OMTKY3 complex reveals two different ovomucoid conformations. *Acta Cryst.* **D61**, 580-588.

119. Huang, K. (1995). *Structural studies of the interactions between serine proteinases and protein inhibitors*. PhD thesis, University of Alberta.

120. Bateman, K. S. (1999). *Structural studies of protein-protein interactions*. PhD thesis, University of Alberta.

121. Bigler, T. L., Lu, W., Park, S. J., Tashiro, M., Wieczorek, M., Wynn, R. & Laskowski, M., Jr. (1993). Binding of amino acid side chains to preformed cavities: interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P1 residues. *Protein Sci.* **2**, 786-799.

122. Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogin, E., Rothberg, I., Ryan, K. & Laskowski, M., Jr. (1997). Binding of amino acid side-chains to S1 cavities of serine proteinases. *J. Mol. Biol.* **266**, 441-461.

123. Lu, S. M., Lu, W., Qasim, M. A., Anderson, S., Apostol, I., Ardelt, W., Bigler, T., Chiang, Y. W., Cook, J., James, M. N. G., Kato, I., Kelly, C., Kohr, W., Komiyama, T., Lin, T. Y., Ogawa, M., Otlewski, J., Park, S. J., Qasim, S., Ranjbar, M., Tashiro, M., Warne, N., Whatley, H., Wieczorek, A., Wieczorek, M., Wilusz, T., Wynn, R., Zhang, W. & Laskowski, M., Jr. (2001). Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1410-1415.

124. Laskowski, M., Jr., Qasim, M. A. & Yi, Z. (2003). Additivity-based prediction of equilibrium constants for some protein-protein associations. *Curr. Opin. Struct. Biol.* **13**, 130-139.

125. Fujinaga, M., Huang, K., Bateman, K. S. & James, M. N. (1998). Computational analysis of the binding of P1 variants of domain 3 of turkey ovomucoid inhibitor to *Streptomyces griseus* protease B. *J. Mol. Biol.* **284**, 1683-1694.

126. Rawlings, N. D., Tolle, D. P. & Barrett, A. J. (2004). Evolutionary families of peptidase inhibitors. *Biochem. J.* **378**, 705-716.

127. Read, R. J. & James, M. N. G. (1986). Introduction to the protein inhibitors: X-ray crystallography. In *Proteinase Inhibitors* (Barrett, A. J. & Salvesen, G., eds.), pp. 301-336. Elsevier, New York.

128. Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433-451.

129. Bode, W. & Huber, R. (2000). Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim. Biophys. Acta.* **1477**, 241-252.

130. Otlewski, J., Jelen, F., Zakrzewska, M. & Oleksy, A. (2005). The many faces of protease-protein inhibitor interaction. *EMBO J.* **24**, 1303-1310.

131. Laskowski, M., Jr. & Kato, I. (1980). Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**, 593-626.

132. Rhodes, M. B., Bennett, N. & Feeney, R. E. (1960). The trypsin and chymotrypsin inhibitors from avian egg whites. *J. Biol. Chem.* **235(6)**, 1686-1693.

133. Empie, M. W. & Laskowski, M., Jr. (1982). Thermodynamics and kinetics of single residue replacements in avian ovomucoid third domains: effect on inhibitor interactions with serine proteinases. *Biochemistry* **21**, 2274-2284.

134. Weber, E., Papamokos, E., Bode, W., Huber, R., Kato, I. & Laskowski, M., Jr. (1981). Crystallization, crystal structure analysis and molecular model of the third domain of Japanese quail ovomucoid, a Kazal type inhibitor. *J. Mol. Biol.* **149**, 109-123.

135. Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I. & Laskowski, M., Jr. (1982). Crystallographic refinement of Japanese quail ovomucoid, a Kazal-type inhibitor, and model building studies of complexes with serine proteases. *J. Mol. Biol.* **158**, 515-537.

136. Bode, W., Epp, O., Huber, R., Laskowski, M., Jr. & Ardelt, W. (1985). The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). *Eur. J. Biochem.* **147**, 387-395.

137. Laskowski, M., Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M. W., Kohr, W. J., Park, S. J., Parks, K., Schatzley, B. L., Schoenberger, O. L., Tashiro, M., Vichot, G., Whatley, H. E., Wieczorek, A. & Wieczorek, M. (1987). Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. *Biochemistry* **26**, 202-221.

138. Laskowski, M., Jr., Apostol, I., Ardelt, W., Cook, J., Giletto, A., Kelly, C. A., Lu, W. Y., Park, S. J., Qasim, M. A., Whatley, H. E. & et al. (1990). Amino acid sequences of ovomucoid third domain from 25 additional species of birds. *J. Protein Chem.* **9**, 715-725.

139. Apostol, I., Giletto, A., Komiyama, T., Zhang, W. & Laskowski, M., Jr. (1993). Amino acid sequences of ovomucoid third domains from 27 additional species of birds. *J. Protein Chem.* **12**, 419-433.

140. Ardelt, W. & Laskowski, M., Jr. (1985). Turkey ovomucoid third domain inhibits eight different serine proteinases of varied specificity on the same ...Leu18-Glu19 ... reactive site. *Biochemistry* **24**, 5313-5320.

141. Wieczorek, M., Park, S. J. & Laskowski, M., Jr. (1987). Covalent hybrids of ovomucoid third domains made from one synthetic and one natural peptide chain. *Biochem. Biophys. Res. Commun.* **144**, 499-504.

142. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

143. Merrifield, B. (1986). Solid phase synthesis. *Science* **232**, 341-347.

144. Wells, J. A. (1990). Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509-8517.

145. Dill, K. A. (1997). Additivity principles in biochemistry. *J. Biol. Chem.* **272**, 701-704.

146. Lu, W., Zhang, W., Molloy, S. S., Thomas, G., Ryan, K., Chiang, Y., Anderson, S. & Laskowski, M., Jr. (1993). Arg15-Lys17-Arg18 turkey ovomucoid third domain inhibits human furin. *J. Biol. Chem.* **268**, 14583-14585.

147. Qasim, M. A., Lu, W., Lu, S. M., Ranjbar, M., Yi, Z., Chiang, Y. W., Ryan, K., Anderson, S., Zhang, W., Qasim, S. & Laskowski, M., Jr. (2003). Testing

of the additivity-based protein sequence to reactivity algorithm. *Biochemistry* **42**, 6460-6466.

148. Lee, T. W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. G. (2005). Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J. Mol. Biol.* **353**, 1137-1151.

149. Lee, T. W., Cherney, M. M., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. G. (2007). Crystal structures reveal an induced-fit binding of a substrate-like Aza-peptide epoxide to SARS coronavirus main peptidase. *J. Mol. Biol.* **366**, 916-932.

150. Lee, T.-W. & James, M. N. G. 1.2 Å-resolution crystal structures reveal the second tetrahedral intermediates of streptogrisin B (SGPB). *Biochim. Biophys. Acta* In press.

151. Lee, T.-W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2007). Structural insights into the non-additivity effects in the sequence-to-reactivity algorithm for serine peptidases and their inhibitors. *J. Mol. Biol.* **367**, 527-546.

152. Park, S. J. (1985). *Effect of amino acid replacements in ovomucoid third domains upon their association with serine proteinases*. PhD thesis, Purdue University.

153. Wynn, R. (1990). *Design of a specific human leukocyte elastase inhibitor based on ovomucoid third domains*. PhD thesis, Purdue University.

Part One SARS Coronavirus Main Peptidase

A version has been published for each chapter in Part One:

Chapter 2

Lee, T.-W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. G. (2005). *J. Mol. Biol.* **353**, 1137-1151.

Chapter 3

Lee, T.-W., Cherney, M. M., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. G. (2007). *J. Mol. Biol.* **366**, 916-932.

Chapter 2

Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide

2.1 Overview

This chapter reports the molecular structures of SARS-CoV M^{pro} at pH 6.5 in space group C2, both in the absence (resolution: 2.1 Å) and the presence of APE (resolution: 1.9 Å), and of an M^{pro} variant with an additional Ala at the N-terminus of each protomer ($M^{pro}_{+A(-1)}$) at pH 6.5 in space group $P2_12_12_1$, in the presence of APE (resolution: 2.3 Å). The mode of action for APE inhibiting M^{pro} and $M^{pro}_{+A(-1)}$, and the peptidase-inhibitor interactions are analyzed in details. Comparisons are made with the interactions between other cysteine peptidases and their inhibitors. Models have been built to explain why M^{pro} is inhibited by only one of the four diastereomers of APE. The importance of the `N-fingers' in the dimerization of M^{pro} and $M^{pro}_{+A(-1)}$, and in the structural integrity of their S1 specificity pockets are examined.

2.2 Materials and methods

Preparation of M^{pro}, $M^{pro}_{+A(-1)}$ and the APEs

The M^{pro} and the $M^{pro}_{+A(-1)}$ used in this study were prepared by the group of Lindsay D. Eltis in the Department of Microbiology and Immunology, University of British Columbia. $M^{pro}_{+A(-1)}$ was cloned, overexpressed and purified as described.[1] A clone expressing M^{pro} was generated using oligonucleotide-directed evolution to

delete the codon corresponding to the N-terminal Ala of $M^{pro}_{+A(-1)}$. Using this clone, M^{pro} was overexpressed and purified essentially as described for $M^{pro}_{+A(-1)}$.

The APEs used in this study, Cbz-Leu-Phe-AGln-(2*S*,3*S*)EP-COOEt and Cbz-Leu-Phe-AGln-(2*R*,3*R*)EP-COOEt, were prepared by the group of James C. Powers in the School of Chemistry and Biochemistry, Georgia Institute of Technology, using the methods established to synthesize other APEs[2,3,4] with minor modifications.

Crystallization, crystal soaking and cryo-protection

Crystallization, crystal soaking and cryo-protection were done by Maia M. Cherney in the group of Michael N. G. James. Before crystallization, both M^{pro} and $M^{pro}_{+A(-1)}$ were dialyzed against 20 mM NaCl and 20 mM Tris-HCl (pH 7.5), and concentrated to 10 mg/mL. All crystals were grown at ambient temperature by the hanging-drop vapor diffusion method. For the C2 crystals, the reservoir solution contained 50 mM ammonium acetate, 5 % (w/v) polyethylene glycol (M_r 10,000), 3 % ethylene glycol, 3 % dimethyl sulfoxide, 1 mM dithiothreitol and 0.1 mM MES (pH 6.5). The drop contained equal amounts of the M^{pro} solution and the reservoir solution. Block-shaped crystals grew in 2 to 3 days to a size of about 0.1 mm × 0.1 mm × 0.1 mm. For the $P2_12_12_1$ crystals, the reservoir solution had essentially the same composition as that for the C2 crystals, except the replacement of 5 % polyethylene glycol (M_r 10,000) and 3 % dimethyl sulfoxide by 6 % polyethylene glycol (M_r 8,000). The drop contained equal amounts of the $M^{pro}_{+A(-1)}$ solution and the reservoir solution. Needle-shaped crystals grew in 3 to 5 days to a size of about 0.05 mm × 0.05 mm × 0.5 mm. Crystals of good quality were selected and soaked overnight in drops with the same compositions as their reservoir solutions plus the

APEs chosen for this study at 3 mM. Cryo-protectants had essentially the same compositions as reservoir solutions, except for the inclusion of 25 % (v/v) glycerol in the case of the C2 crystals and the increase of ethylene glycol to 25 % in the case of the $P2_12_12_1$ crystals. Crystals were briefly soaked and then immediately flash-cooled in the liquid nitrogen for storage and shipment to the synchrotron beamline.

Data collection and processing, and structure solution and refinement

The X-ray diffraction data from all crystals were collected at the synchrotron Beamline 8.3.1 (equipped with an ADSC-Q210 CCD detector) at the Advanced Light Source in the Lawrence Berkeley National Laboratory. All data sets were indexed, scaled and merged using DENZO and SCALEPACK.[5] Structure solution and refinement were done in CCP4.[6,7] All structures were solved by the molecular replacement method,[8] using the molecular structure of the unbound M^{pro} at pH 8.0 in space group $P2_1$ (PDB accession code: 1UK2)[9] as the search model for the structure of the unbound M^{pro} in space group C2, and using the molecular structure of the unbound M^{pro} in space group C2 as the search model for the structure of the M^{pro}:APE complex in space group C2 and for the structure of the $M^{pro}_{+A(-1)}$:APE complex in space group $P2_12_12_1$. The structures of the unbound M^{pro} and the $M^{pro}_{+A(-1)}$:APE complex were solved using AMoRe,[10] and the structure of the $M^{pro}_{+A(-1)}$:APE complex was solved using MOLREP.[11] APE was located as outstanding electron densities in the substrate-binding region of the peptidase in both the $|F_o|-|F_c|,\alpha_c$ (contoured at 3σ and 4σ) and the $2|F_o|-|F_c|,\alpha_c$ (contoured at 1σ) maps for each structure. All structures were iteratively refined using REFMAC,[12] and manually

adjusted when needed using XtalView/Xfit.[13] The stereochemical qualities of the final structures were assessed using PROCHECK.[14]

Structure analysis

Graphical representations of the molecular structures were prepared using PyMOL (http://www.pymol.org/). Superpositions of molecular structures were done using ALIGN,[15,16] based on the main-chain atoms (amide N, C^{α}, and carbonyl C and O). The surface areas of molecular structures were calculated using NACCESS.[17] Peptidase-inhibitor interactions and dimer interactions were analyzed using LIGPLOT and DIMPLOT,[18] respectively.

Protein Data Bank accession codes

The atomic coordinates and the structure factors of all structures have been deposited in the Protein Data Bank. The accession code is 2A5A for the structure of the unbound M^{pro}, 2A5I for the structure of the M^{pro}:APE complex, and 2A5K for the structure of the $M^{pro}_{+A(-1)}$:APE complex.

2.3 Results

Structure determination

The parameters and statistics derived from data processing and structure refinement are summarized in Table 2-1. For the unbound M^{pro} and the M^{pro}:APE complex in space group C2, each asymmetric unit has only one protomer of the dimer. The two protomers of each dimer are related by the crystallographic two-fold symmetry (Figure 2-1). All residues of the protomer (residues 1 to 306) were identified in the electron density maps. In the Ramachandran plot for the molecular structure of the unbound M^{pro}, Asp33, Ala46 and Glu47 are in the generously allowed

Table 2-1 Parameters and statistics derived from X-ray diffraction data processing and structure refinement

	Unbound M^{pro}	M^{pro}:APE	$M^{pro}_{+A(-1)}$:APE
Data processing			
Wavelength (Å)	1.116	1.116	1.116
Resolution limit[a] (Å)	42.41 – 2.08	32.32 – 1.88	41.83 – 2.30
	(2.15 – 2.08)	(1.95 – 1.88)	(2.38 – 2.30)
Space group	C2	C2	$P2_12_12_1$
Unit-cell constants			
a (Å)	107.76	106.71	65.35
b (Å)	82.14	83.68	67.48
c (Å)	53.13	52.87	167.32
α(°)	90.00	90.00	90.00
β(°)	104.87	105.66	90.00
γ(°)	90.00	90.00	90.00
Mosaicity (°)	1.00	0.61	0.23
Total number of	50,960	74,092	134,161
reflections	(3,872)	(6,410)	(10,755)
Number of unique	26,196	35,576	33,305
reflections	(2,482)	(3,301)	(3,147)
Redundancy	1.9 (1.6)	2.0 (1.7)	4.0 (3.4)
Completeness (%)	95.8 (91.0)	96.3 (90.0)	99.2 (95.5)
R_{sym}[b] (%)	3.7 (22.5)	3.3 (31.5)	5.2 (51.8)
$I / \sigma(I)$	14.9 (3.4)	21.5 (2.1)	24.8 (2.0)

[a] Numbers in the parentheses refer to the highest resolution bins.

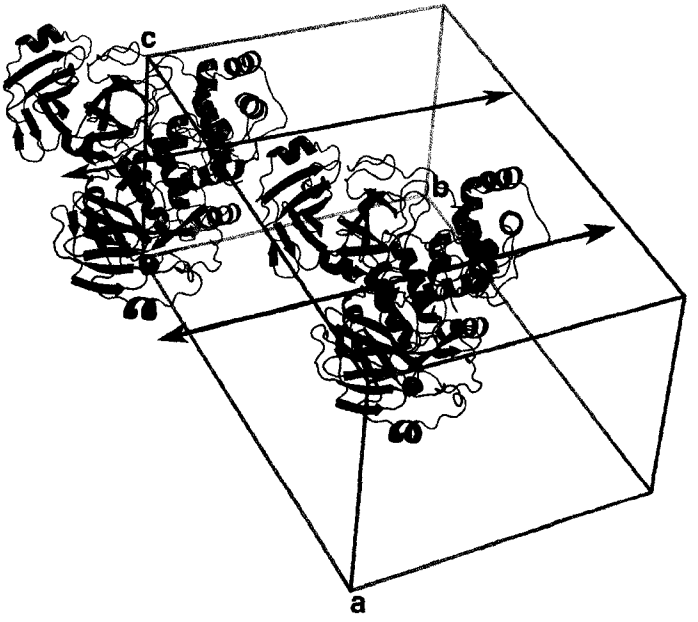
[b] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_{hkl,i} - \langle I_{hkl}\rangle| / \Sigma_{hkl}\Sigma_i I_{hkl,i}$, where $I_{hkl,i}$ and $\langle I_{hkl}\rangle$ are the i-th observed intensity and the average intensity of reflection hkl, respectively.

Table 2-1 (continued)

	Unbound M^{pro}	M^{pro}:APE	$M^{pro}_{+A(-1)}$:APE
Structure refinement			
Resolution range (Å)	32.25 – 2.08	32.32 – 1.88	35.92 – 2.30
R_{work}[c] (%)	19.2	19.8	17.5
R_{free}[c] (%)	24.6	24.2	24.9
Number of non-hydrogen atoms per asymmetric unit (average *B* factor, Å^2)			
Protein	2,371 (56.03)	2,371 (42.72)	4,690 (46.61)
APE	Not applicable	46 (45.28)	92 (65.29)
Solvent	150.5 (66.13)	264.5 (55.29)	279.5 (56.28)
rms deviation from ideal geometry			
Bond lengths (Å)	0.029	0.022	0.021
Bond angles (°)	2.801	2.342	1.944
Ramachandran plot			
Favored (%)	87.2	91.3	88.5
Allowed (%)	10.6	6.8	9.9
Generously allowed (%)	1.1	1.5	0.8
Disallowed (%)	1.1	0.4	0.8

[c] $R_{work} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and the calculated structure factor amplitudes of a particular reflection, respectively, and the summation is over 95 % of the reflections in the specified resolution range. The remaining 5 % of the reflections were randomly selected before the structure refinement and not included in the structure refinement. R_{free} was calculated over these reflections using the same equation as for R_{work}.[29]

Figure 2-1 M^{pro} in the C2 unit cell. The two protomers making up the dimeric peptidase are from two different asymmetric units (green and magenta) related by the crystallographic two-fold symmetry along the **b**-axis (shown as arrows).



regions, whereas Asn84, Tyr154 and Ile286 are in the disallowed regions. Asp33 $O^{\delta 2}$ forms a hydrogen bond with the phenolic OH of Tyr101 (2.9 Å). Asn84 $N^{\delta 2}$ forms a hydrogen bond with the carbonyl O of Glu178 (3.2 Å), and there may be hydrogen bonds and Van der Waals forces between the side chains of Asn84 and Lys180 as well. The side chains of Thr285 and Ile286 from opposite protomers make contact with one another at the dimer interface. The electron densities for the side chains of Ala46, Glu47 and Tyr154 are not well defined. Similarly, in the Ramachandran plot for the molecular structure of the M^{pro}:APE complex, Asp33, Asn84, Tyr154 and Asn277 are in the generously allowed region, whereas Ile286 is in the disallowed region. The electron densities for Tyr154 and Asn277 are not well defined.

For the $M^{pro}{}_{+A(-1)}$:APE complex in space group $P2_12_12_1$, there is a dimer in each asymmetric unit. Only residues 1A to 304A and 1B to 300B were identified in the electron density maps. In the Ramachandran plot for this structure, Asp33 and Asn84 of both protomers are in the generously allowed regions, whereas Tyr154 and Ile286 of both protomers are in the disallowed regions. The electron density for Tyr154 is not well defined. Superposition of protomers A and B yielded a root-mean-square difference (rmsd) of 0.28 Å for 1,070 out of 1,192 main-chain atoms. Differences in atomic positions (up to 4.18 Å) occur mainly in the N- and the C-terminal residues as well as those poorly defined residues on the flexible loops.

The three molecular structures reported here are in close agreement (Table 2-2), and are identical to the molecular structures of M^{pro} determined previously in space group $P2_1$, with regard to the protomer orientation and the protein fold.[9]

Table 2-2 Root-mean-square differences (rmsd; in Å) for superpositions of the molecular structures of M^{pro} and $M^{pro}_{+A(-1)}$ determined in different conditions[a]

	Unbound M^{pro} (C2, pH 6.5)[b]	**M^{pro}:APE (C2, pH 6.5)[b]**	**$M^{pro}_{+A(-1)}$:APE ($P2_12_12_1$, pH 6.5)[b]**
Unbound M^{pro}	-	PP: 0.24 (1,148)	PA: 0.35 (1,100)
(C2, pH 6.5)[b]	-	DD: 0.37 (2,292)	PB: 0.35 (1,116)
	-	-	DD: 0.65 (2,260)
M^{pro}:APE	-	-	PA: 0.39 (1,118)
(C2, pH 6.5)[b]	-	-	PB: 0.38 (1,133)
	-	-	DD: 0.74 (2,314)
Unbound M^{pro}	AP: 0.63 (1,142)	AP: 0.75 (1,157)	AA: 0.62 (1,131)
($P2_1$, pH 6.0)	BP: 0.32 (1,060)	BP: 0.37 (1,059)	BB: 0.39 (1,074)
	DD: 0.48 (1,881)	DD: 0.48 (1,840)	AB: 0.62 (1,148)
	-	-	BA: 0.34 (1,063)
	-	-	DD: 0.79 (2,270)
Unbound M^{pro}	AP: 0.92 (1,155)	AP: 0.99 (1,165)	AA: 0.76 (1,132)
($P2_1$, pH 7.6)	BP: 0.99 (1,161)	BP: 1.02 (1,150)	BB: 0.89 (1,162)
	DD: 1.03 (2,131)	DD: 1.04 (2,156)	AB: 0.82 (1,167)
	-	-	BA: 0.90 (1,158)
	-	-	DD: 0.83 (2,066)
Unbound M^{pro}	AP: 0.92 (1,152)	AP: 0.98 (1,148)	AA: 0.75 (1,136)
($P2_1$, pH 8.0)	BP: 0.96 (1,156)	BP: 1.01 (1,153)	BB: 0.85 (1,155)
	DD: 1.08 (2,199)	DD: 1.06 (2,176)	AB: 0.78 (1,136)
	-	-	BA: 0.84 (1,146)
	-	-	DD: 0.85 (2,132)

Table 2-2 (continued)

	Unbound M^{pro} (C2, pH 6.5)[b]	**M^{pro}:APE (C2, pH 6.5)[b]**	**$M^{pro}_{+A(-1)}$:APE ($P2_12_12_1$, pH 6.5)[b]**
M^{pro}:CMK	AP: 0.58 (1,143)	AP: 0.68 (1,145)	AA: 0.58 (1,155)
($P2_1$, pH 6.0)	BP: 0.33 (1,075)	BP: 0.37 (1,097)	BB: 0.41 (1,105)
	DD: 0.47 (1,859)	DD: 0.51 (1,922)	AB: 0.57 (1,142)
			BA: 0.40 (1,097)
			DD: 0.70 (2,295)

[a] Superpositions of molecular structures were done for pairs of protomers (PP, PA, AP, PB, BP, AA, BB and AB) and pairs of dimers (DD). The first letter corresponds to the part from the molecular structure listed along the left of the table and the second letter corresponds to the part from the molecular structure listed along the top of the table. P, protomer (applicable to the molecular structures in space group C2 only); A, protomer A; B, protomer B; D, dimer. All superpositions started with residues 3 to 300 of each protomer (total number of main-chain atoms: 1,192) included in the calculations. The number of main-chain atoms included in the final calculation of rmsd for each superposition is given in the parentheses.

[b] These are the molecular structures reported here.

Binding of APE to M^{pro} and $M^{pro}_{+A(-1)}$

Crystals of M^{pro} and $M^{pro}_{+A(-1)}$ were soaked in the solutions of Cbz-Leu-Phe-AGln-(2*S*,3*S*)EP-COOEt, Cbz-Leu-Phe-AGln-(2*R*,3*R*)EP-COOEt, and a racemic mixture of the 2*S*,3*S* and the 2*R*,3*R* diastereomers (*trans*). Only the 2*S*,3*S* diastereomer showed up in the electron density maps. APE binds in the substrate-binding regions of both M^{pro} and $M^{pro}_{+A(-1)}$ (Figures 2-2a and 2-2b). As visualized in all of the three molecular structures reported here, the residues forming the substrate-binding regions of both protomers of the peptidase are in the catalytically competent conformation, similar to their counterparts in the molecular structures of the main peptidases from other coronaviruses,[19,20] and to those in protomer A of the molecular structures of M^{pro} in space group $P2_1$.[9] In the molecular structure of the unbound M^{pro}, the catalytic dyad has a distance of 3.7 Å between His41 $N^{\varepsilon 2}$ and Cys145 S^{γ}, and Cys145 S^{γ} is coplanar with the atoms of the imidazole ring of His41. Superposition of the molecular structures of the unbound M^{pro} and the M^{pro}:APE complex shows that the binding of APE does not cause any major changes in the molecular structure of M^{pro} (Table 2-2). The C^{α}-C^{β} bond of Cys145 undergoes a 95 ° rotation (χ_1: from -64.0 ° to +30.7 °) accompanying the formation of a covalent bond with a distance of 2.01 Å between Cys145 S^{γ} of M^{pro} and the epoxide atom C3 of APE (Figures 2-3a and 2-3b). The distance of a C-S single bond is normally about 1.8 Å. However, with an estimated overall coordinate error (based on maximum likelihood) of 0.12 Å for the molecular structure of the M^{pro}:APE complex, the difference between the refined and the expected distances (0.2 Å) is not considered significant. This new covalent bond makes a torsion angle, O=C(P1-AGln)-C3(epoxide)-S^{γ}(Cys145), of 65.4 °

Figure 2-2 Binding of APE (orange) in the substrate-binding regions of M^{pro} and $M^{pro}_{+A(-1)}$. (a) Outstanding electron densities in the $|F_o|$-$|F_c|$,α_c maps for the structures of the M^{pro}:APE complex and protomer B of the $M^{pro}_{+A(-1)}$:APE complex. (b) Outstanding electron densities in the $|F_o|$-$|F_c|$,α_c maps for protomer A of the $M^{pro}_{+A(-1)}$:APE complex.

Figure 2-3 Interactions of APE with M^{pro} and $M^{pro}_{+A(-1)}$. (a) View from the 'floor' of the S1 specificity pocket in all of molecular structures reported here. (b) Schematic diagram for the peptidase-inhibitor interactions in the molecular structures of the M^{pro}:APE complex and protomer B of the $M^{pro}_{+A(-1)}$:APE complex. (c) Schematic diagram for the peptidase-inhibitor interactions in protomer A of the $M^{pro}_{+A(-1)}$:APE complex. Hydrogen bonds are shown as dashed lines, with their distance(s) (in Å) given in (b) and (c). In (b), the first value corresponds to the M^{pro}:APE complex, and the second value corresponds to protomer B of the $M^{pro}_{+A(-1)}$:APE complex. Residues of the peptidase in contact with APE are shown as arcs.

(a)

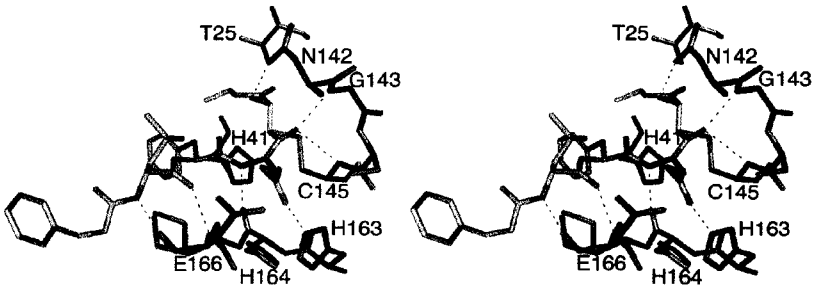
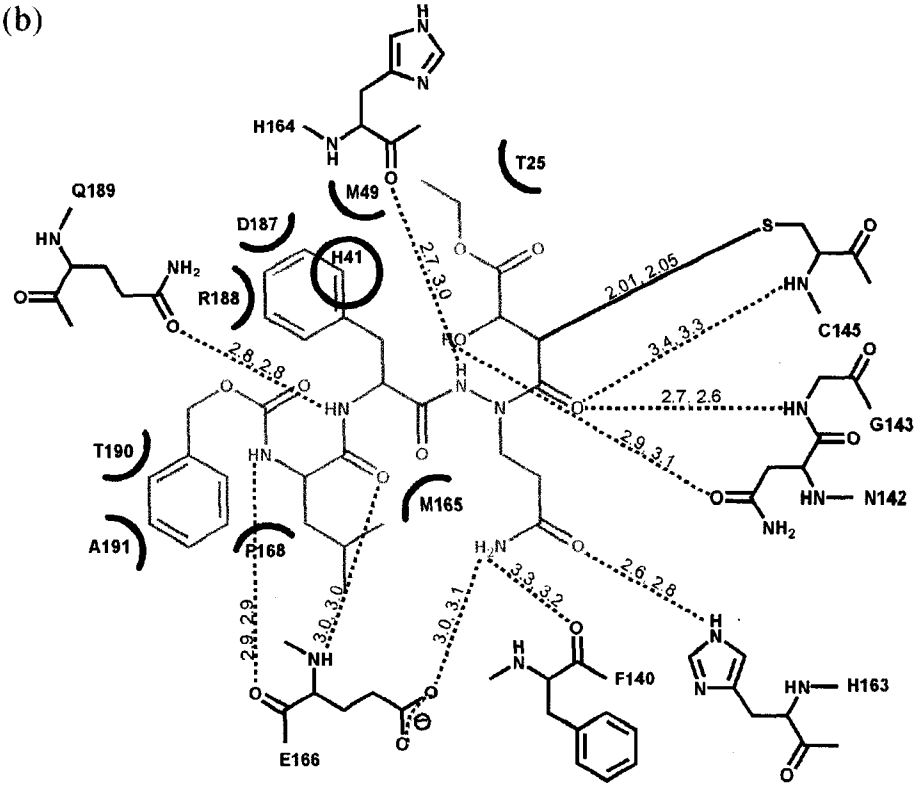
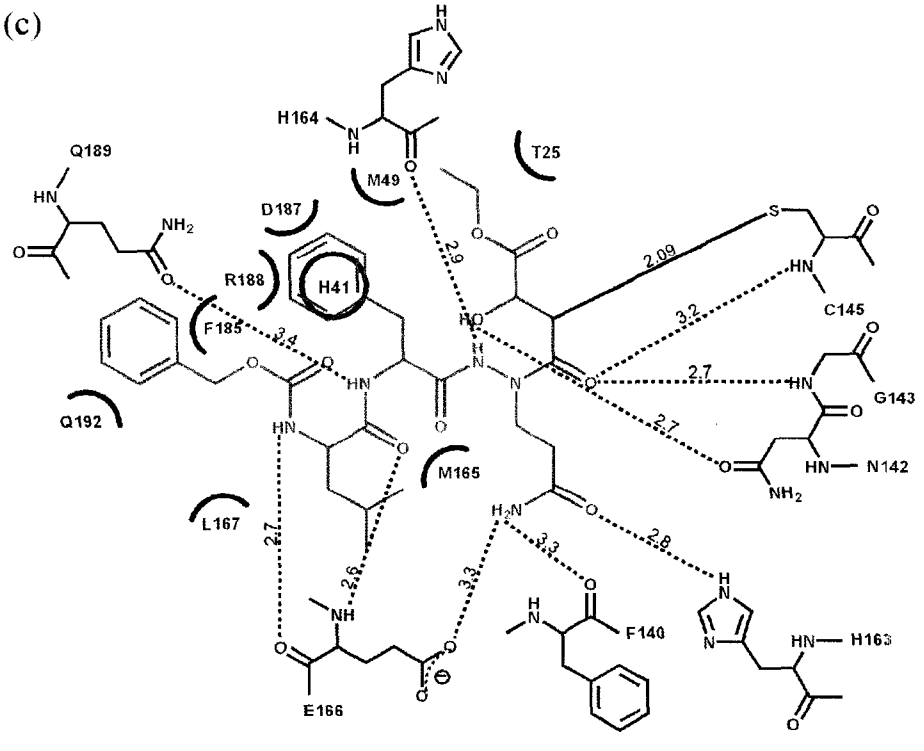


Figure 2-3 (continued)

(b)



(c)



(Figure 2-4c). The epoxide ring of APE opens, leaving a hydroxyl group on the atom C2. This hydroxyl group forms hydrogen bonds with Asn142 $O^{\delta 1}$ of M^{pro} (2.9 Å) and the carbonyl O of P2-Phe of APE (3.2 Å) (Figures 2-3a and 2-3b). The configurations of atoms C2 and C3 are inverted from 2*S*,3*S* to 2*R*,3*R*. The conversion at the atom C3 is due to the formation of this new covalent bond, whereas that at the atom C2 is a result of the change in the ranking order of substituents (Figure 2-4d). Asn142 and Cys145 of M^{pro} flank a short loop constituting the oxyanion hole, with the main-chain NH groups of Gly143 and Cys145 as hydrogen-bond donors to stabilize the carbonyl O of the scissile peptide group of the substrate. The conformation of this loop remains essentially unchanged upon the binding of APE. The carbonyl O of P1-AGln of APE is accommodated in the oxyanion hole of M^{pro}, accepting hydrogen bonds from the main-chain NH groups of Gly143 and Cys145 at distances of 2.7 Å and 3.4 Å, respectively (Figures 2-3a and 2-3b). The N^{α} of P1-AGln of APE remains sp^2-hybridized and adopts trigonal planar geometry. For P1-AGln, the equivalent to ϕ, C(P2-Phe)-N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln), is -84.8 ° (Figure 2-4a), and the equivalent to ψ, N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln)-C3(epoxide), is -7.3 ° (Figure 2-4b). With the errors in atomic coordinates considered, the trigonal plane centered at the N^{α} of P1-AGln is coplanar with that centered at the carbonyl C of P1-AGln, allowing the N^{α} to reduce the electrophilicity of the carbonyl C by π-electron delocalization. The equivalent to χ_1 of P1-AGln, N-N^{α}-C^{β}-C^{γ}, is -77.4 °.

Compared to that of the M^{pro}:APE complex, the molecular structure of the $M^{pro}{}_{+A(-1)}$:APE complex shows some differences in the geometry of binding. The rotation of the C_{α}-C_{β} bond of Cys145 of $M^{pro}{}_{+A(-1)}$ reaches a more positive value of

Figure 2-4 Newman projections. (a) The equivalent to ϕ of P1-AGln, C(P2-Phe)-N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln), (b) the equivalent to ψ of P1-AGln, N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln)-C3(epoxide), (c) the torsion angle O=C(P1-AGln)-C3(epoxide)-S^{γ}(Cys145), and (d) the torsion angle C(P1-AGln)-C3(epoxide)-C2(epoxide)-C(ethyl ester carbonyl) of APE in the M^{pro}:APE complex. (e) The equivalent to ϕ of P1-AGln, C(P2-Phe)-N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln), (f) the equivalent to ψ of P1-AGln, N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln)-C3(epoxide), (g) the torsion angles O=C(P1-AGln)-C3(epoxide)-S^{γ}(Cys145) and (h) the torsion angle C(P1-AGln)-C3(epoxide)-C2(epoxide)-C(ethyl ester carbonyl) of APE in protomers A and B of the $M^{pro}{}_{+A(-1)}$:APE complex.

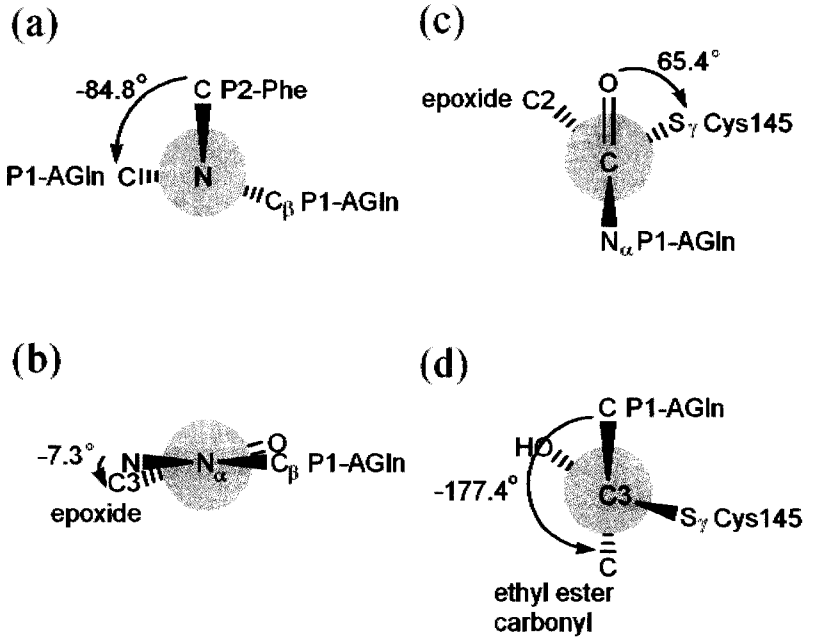
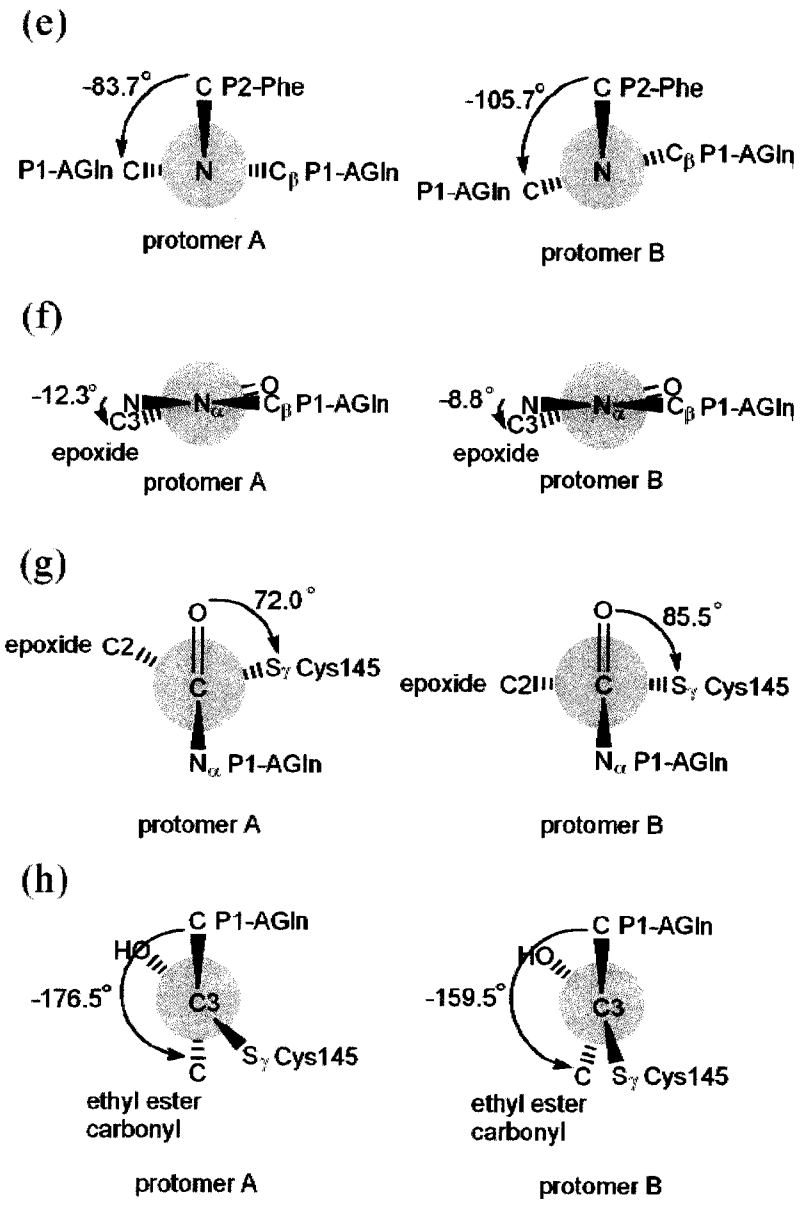


Figure 2-4 (continued)



χ_1 (protomer A: 47.5 °; protomer B: 46.5 °). The length of the covalent bond between Cys145 S^{γ} of $M^{pro}_{+A(-1)}$ and the epoxide atom C3 of APE is 2.09 Å in protomer A and 2.05 Å in protomer B (Figures 2-3a to 2-3c). Note that the estimated overall coordinate error (based on maximum likelihood) for the molecular structure of the $M^{pro}_{+A(-1)}$ complex is 0.17 Å. O=C(P1-AGln)-C3(epoxide)-S^{γ}(Cys145) makes a torsion angle of 72.0 ° in protomer A and 85.5 ° in protomer B (Figure 2-4g). In both protomers, the configurations of atoms C2 and C3 of APE are inverted from *2S,3S* to *2R,3R* (Figure 2-4h), and the N^{α} of P1-AGln of APE remains sp^2-hybridized and adopts trigonal planar geometry. For the P1-AGln of APE, the equivalent to ϕ, C(P2-Phe)-N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln), is -83.7 ° in protomer A and -105.7 ° in protomer B (Figure 2-4e), and the equivalent to ψ, N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln)-C3(epoxide), is -12.3 ° in protomer A and -8.8 ° in protomer B (Figure 2-4f). With the errors in atomic coordinates considered, the trigonal plane centered at the N^{α} of P1-AGln is roughly coplanar with that centered at the carbonyl C of P1-AGln in both protomers. The equivalent to χ_1 of P1-AGln, N-N^{α}-C^{β}-C^{γ}, is -82.7 ° in protomer A and -119.5 ° in protomer B.

The aza-peptide component of APE binds in the substrate-binding regions of both protomers of M^{pro} and $M^{pro}_{+A(-1)}$ in a substrate-like manner. Consistent with the results of the secondary-structure studies for the substrates of M^{pro},[21] the molecular structures of the M^{pro}:APE and the $M^{pro}_{+A(-1)}$:APE complexes show that the backbone of the aza-peptide component of APE interacts with that of residues His164 to Glu166 in both protomers of M^{pro} and $M^{pro}_{+A(-1)}$ through amide hydrogen-carbonyl

oxygen hydrogen bonding in the manner of an anti-parallel β sheet (Figures 2-3a to 2-3c).

Substrate-binding subsites

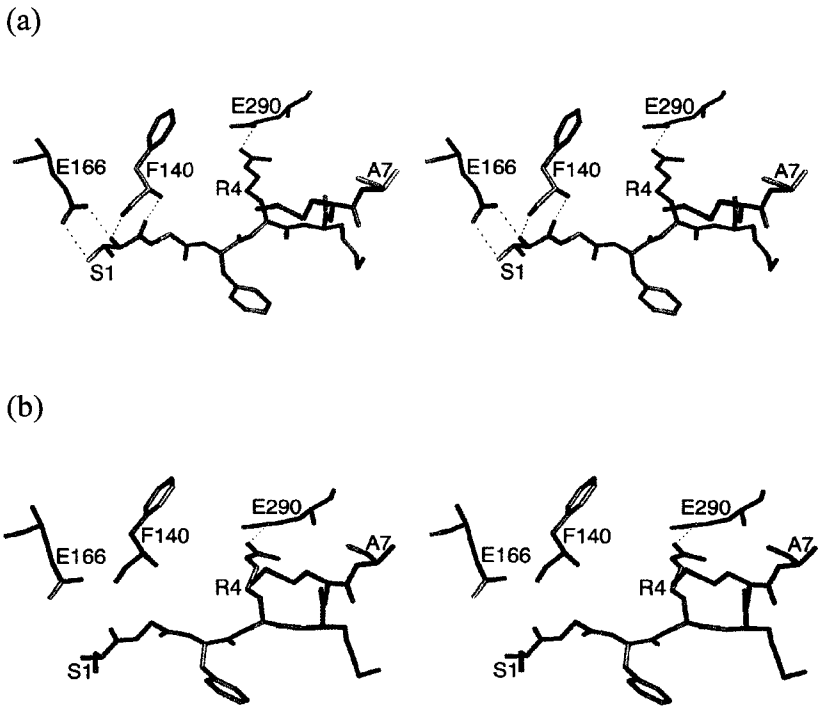
The predominant S1 specificity of M^{pro} for Gln is determined primarily by His163. In the molecular structure of the unbound M^{pro}, His163 $N^{\varepsilon 2}$ interacts with a chloride ion at a distance of 3.3 Å and in the plane of the imidazole ring of His163, whereas in the molecular structures of the M^{pro}:APE and the $M^{pro}{}_{+A(-1)}$:APE complexes, the chloride ion is displaced by the side-chain amide group of P1-AGln of APE with its $O^{\varepsilon 1}$ accepting a hydrogen bond from His163 $N^{\varepsilon 2}$ of M^{pro} and $M^{pro}{}_{+A(-1)}$ (2.6 to 2.8 Å). Additional hydrogen bonds may be donated, though not of ideal geometry, by the $N^{\varepsilon 2}$ of P1-AGln of APE to Glu166 $O^{\varepsilon 2}$ and the carbonyl O of Phe140 of M^{pro} and $M^{pro}{}_{+A(-1)}$ (Figures 2-3b and 2-3c). In both protomers of all of the three molecular structures, the phenyl ring of Phe140 interacts with the imidazole ring of His163 through π-stacking (distance between the geometric centers of the rings: 3.7 to 3.8 Å), and the latter is oriented for its $N^{\delta 1}$ to accept a hydrogen bond from the phenolic OH of Tyr161 (2.9 to 3.1 Å). This hydrogen bond maintains the neutral tautomeric state of the imidazole ring of His163 with its $N^{\varepsilon 2}$ protonated over a broad range of pH. This is crucial for the interaction of His163 of M^{pro} and $M^{pro}{}_{+A(-1)}$ with P1-Gln of the substrate.[9,19,20]

In M^{pro}, the integrity of the S1 specificity pocket in one protomer requires the protonated α-amino group of Ser1 from the other protomer. This residue is at the tip of the 'N-finger' (N-terminal residues 1 to 7) propagating between domain III of its parent protomer and domain II of the opposite protomer. In the unbound M^{pro} and the

M^{pro}:APE complex, the α-amino group of Ser1 of each protomer forms an ionic interaction with the side-chain carboxylate group of Glu166 of the opposite protomer (2.7 Å). Ser1 of each protomer also interacts with Phe140 of the opposite protomer through amide hydrogen-carbonyl oxygen hydrogen bonding (3.2 Å). These three residues are thus held together to form the `floor' of the S1 specificity pocket (Figure 2-5a). However, in both protomers of the $M^{pro}_{+A(-1)}$:APE complex, Ser1 is N-terminally blocked by the additional Ala, and the ionic interaction between the α-amino group of Ser1 and the side-chain carboxylate group of Glu166 is lost. The amide hydrogen-carbonyl oxygen hydrogen bonds between Ser1 and Phe140 are also lost, because Ser1 is no longer oriented properly. The additional Ala is disordered, leaving Ser1 unanchored and the `floor' of the S1 specificity pocket partly disrupted (Figure 2-5b). Nonetheless, the disruption of the S1 specificity pockets of $M^{pro}_{+A(-1)}$ by the presence of an additional Ala at the N-terminus of each protomer does not likely compromise the catalytic activity of $M^{pro}_{+A(-1)}$ significantly as compared with that of M^{pro}. In fact, the specific activity of the variant of M^{pro} with a ten-residue affinity tag at the N-terminus of each protomer is lower than that of the wild-type M^{pro} by less than an order of magnitude.[22]

M^{pro} has greatest preference for Leu and Ile as residue P2 of a substrate, followed by Phe, Val and Met in that order.[21,23] In the M^{pro}:APE and the $M^{pro}_{+A(-1)}$:APE complexes, the side chain of P2-Phe of APE fits snugly in the S2 specificity pocket of M^{pro} and $M^{pro}_{+A(-1)}$, where the interactions are mainly hydrophobic. The phenyl ring of P2-Phe of APE interacts with the imidazole ring of His41 of M^{pro} and $M^{pro}_{+A(-1)}$ through π-stacking (distance between the geometric centers of the two rings:

Figure 2-5 'Floor' of the S1 specificity pocket in M^{pro} and $M^{pro}_{+A(-1)}$. Residues Phe140, Glu166 and Glu290 of the first protomer (green) and the 'N-finger' (N-terminal residues 1 to 7) of the second protomer (magenta) are shown. (a) In the unbound M^{pro} and the M^{pro}:APE complex, Phe140 and Glu166 of the first protomer interacts with Ser1 of the second protomer to form the 'floor' of the S1 specificity pocket. (b) In the $M^{pro}_{+A(-1)}$:APE complex, the disordered N-terminal Ala of the second protomer (not shown) orients Ser1 poorly for any interactions with Phe140 and Glu166 of the first protomer. However, the interaction between Glu290 of the first protomer and Arg4 of the second protomer remains in both (a) and (b). Hydrogen bonds are shown as dashed lines.



4.3 to 4.6 Å). Superposition of the molecular structures of the unbound M^{pro} and the M^{pro}:APE complex shows that, upon the binding of APE, the side chain of Met49 of M^{pro} undergoes a large conformational change, thereby opening its S2 specificity pocket for the side chain of P2-Phe of APE. Also, the side chain of Gln189 of M^{pro} and $M^{pro}{}_{+A(-1)}$ is no longer disordered after its reorientation and formation of a hydrogen bond through its $O^{\varepsilon 1}$ with the main-chain NH of P2-Phe of APE (2.8 to 3.4 Å). This appears to secure P2-Phe of APE in the S2 specificity pocket of M^{pro} and $M^{pro}{}_{+A(-1)}$ (Figures 2-2a, 2-2b and 2-3a).

In the molecular structures of the M^{pro}:APE and the $M^{pro}{}_{+A(-1)}$:APE complexes, the side chain of P3-Leu of APE extends into the solvent and does not have any significant interactions with M^{pro} and $M^{pro}{}_{+A(-1)}$ (Figures 2-2a, 2-2b and 2-3a). This lack of interactions is consistent with the fact that no S3 specificity could be established for M^{pro}.[23]

The aza-peptide component of APE consists of only three residues, and its benzyloxycarbonyl (Cbz) group takes up part of the space for residue P4 of a substrate. M^{pro} has a shallow S4 subsite that accommodates small side chains (Ser, Thr, Val, Pro and Ala).[23] In the M^{pro}:APE complex and protomer B of the $M^{pro}{}_{+A(-1)}$:APE complex, the Cbz group of APE does not interact with the S4 subsite of M^{pro} and $M^{pro}{}_{+A(-1)}$. The position for the C^{α} of residue P4 of a substrate is occupied by the atom O2 of the Cbz group of APE. In this conformation, the benzyl group of APE makes contacts with Pro168 and with residues Thr190 to Gln192 of M^{pro} and $M^{pro}{}_{+A(-1)}$, and the Cbz group is exposed to the solvent (Figures 2-2a and 2-3b). In contrast, in protomer A of the $M^{pro}{}_{+A(-1)}$:APE complex, the benzyl group of APE squeezes into

and thereby widens the S4 subsite of $M^{pro}_{+A(-1)}$, so that it is snugly accommodated in this enlarged pocket now formed by residues Met165 to Pro168, Phe185, Gln192 and the main-chain atoms of Val186 (Figures 2-2b and 2-3c).

The S1' subsite of M^{pro} is also shallow, accommodating small side chains only (Ser, Ala, Gly, Asn and Cys).[21,23] In the design of APEs, several different epoxide derivatives were attached to the aza-peptide component to modulate the interactions of APE with the S1' subsite of a cysteine peptidase in clan CD.[2,3,4] However, in the M^{pro}:APE and the $M^{pro}_{+A(-1)}$:APE complexes, the S1' subsite of M^{pro} and $M^{pro}_{+A(-1)}$ does not apparently interact with APE. The epoxide atom C2 of APE sits close to the position for the C^{α} of residue P1' of a substrate, and the hydroxyl group on the atom C2 is exposed to the solvent. The ethyl ester group of APE lies against the `ceiling' of the active site lined by Leu27, Pro39, His41, and the peptide group between His41 and Val42.

Dimer interface

The molecular structures reported here show similar features at their dimer interfaces. The solvent-accessible surface area buried upon dimerization of M^{pro} and $M^{pro}_{+A(-1)}$ is from 1,250 $Å^2$ to 1,260 $Å^2$ per protomer. In the molecular structures of the unbound M^{pro} and the M^{pro}:APE complex, the crystallographic two-fold axis passes through the dimer interface and brings the pairs of interacting residues from opposite protomers into exact two-fold symmetry (Figure 2-1); whereas in the molecular structure of the $M^{pro}_{+A(-1)}$:APE complex, the dimer interface still exhibits approximate two-fold symmetry even in the absence of a crystallographic two-fold axis. In all of the three molecular structures, the dimer interface concentrates on one

face of each protomer: that containing residues in the 'N-finger', and domains II and III. The majority of the interactions involve residues in the 'N-finger' and domain II. Two ionic interactions, one between the α-amino group of Ser1 and the side-chain carboxylate group of Glu166 from opposite protomers, and one between the side chains of Arg4 and Glu290 from opposite protomers, are observed in the molecular structures of the unbound M^{pro} and the $M^{pro}_{+A(-1)}$:APE complex (Figure 2-5a). However, in the molecular structure of the $M^{pro}_{+A(-1)}$:APE complex, only the latter ion-pair is observed (Figure 2-5b). In contrast to the 'N-finger' immobilized at the dimer interface, the C-terminal loop (residues 301 to 306) of each protomer is highly mobile as it is exposed to the solvent and anchored by only 3 or 4 solvent-exposed hydrogen bonds to some residues along the rim of the dimer interface. Interestingly, in the unbound M^{pro} and the M^{pro}:APE complex, the C-terminal loop of each protomer extends towards the S1 specificity pocket in the opposite protomer, whereas in the $M^{pro}_{+A(-1)}$:APE complex, the C-terminal loop of protomer A (the only one visualized in this molecular structure) propagates away from the S1 specificity pocket in protomer B.

2.4 Discussion

The molecular structures reported here and the kinetic data[22] indicate that APEs have excellent potential as inhibitors of M^{pro} and are worthy of further evaluation in the development of lead compounds for anti-SARS drugs. The k_{inact}/K_i of Cbz-Leu-Phe-AGln-(2*S*,3*S*)EP-COOEt for M^{pro} is similar in magnitude to the corresponding values of the first-generation APEs produced to inhibit other cysteine

peptidases.[2,3,4] Optimization of the latter has yielded inhibitors of caspases with k_{inact}/K_i well over 10^6 $M^{-1}s^{-1}$.[4]

The excellent specificities of APEs for cysteine peptidases in clan CD[2,3,4] suggest that APEs have better potential as inhibitors of M^{pro} than do chloromethyl ketones (CMKs), the first class of potential inhibitors proposed on a structural basis. The molecular structure of the M^{pro}:CMK complex shows different and unexpected modes of binding for CMK to the two protomers of M^{pro}.[9] CMKs are highly active alkylating agents and react with good nucleophiles, such as hydroxyl and thiol groups. Therefore, they inhibit serine peptidases as well as cysteine peptidases.[24] A recent study has shown that CMKs inhibit efficiently some cysteine peptidases in clan CA, such as papain and the cathepsins.[25] This casts doubt on the utility of CMKs as specific inhibitors of M^{pro}. In contrast, the molecular structure of the M^{pro}:APE complex shows that the aza-peptide component of APE binds to M^{pro} in a substrate-like manner. The backbone of the aza-peptide component of APE forms amide hydrogen-carbonyl oxygen hydrogen bonds with that of residues His164 to Glu166 of M^{pro} in the manner of an anti-parallel β sheet. The side chains of residues P1 and P2 of APE occupy the S1 and the S2 specificity pockets of M^{pro}, respectively. Based on the definitions for the binding of epoxysuccinyl peptides to cysteine peptidases in clan CA,[24] this corresponds to the S and S' binding mode, with inclination to the S binding mode because the S subsites of M^{pro} makes the major contribution to the M^{pro}:APE interactions.

The molecular structures of the M^{pro}:APE and the $M^{pro}_{+A(-1)}$:APE complexes substantiate the mechanism by which APEs have been proposed to inhibit irreversibly

their target peptidases (Figure 1-8a). Nucleophilic attack on the epoxide atom C3, rather than the atom C2, of APE by Cys145 S^{γ} of M^{pro} and $M^{pro}_{+A(-1)}$ is consistent with the expected transition-state geometry for M^{pro}-catalyzed proteolysis. In caspase-3, the atom C3 is attacked;[26] whereas in caspase-1, the atom C2 is attacked.[4] In the case of epoxysuccinyl peptides inhibiting cysteine peptidases in clan CA, the position of attack depends on the orientation of the epoxysuccinyl peptide in the substrate-binding region of the peptidase.[24] E-64 binds to papain in the S binding mode, similar to the mode of APE binding to M^{pro}. However, in the papain:E-64 complex, the atom C2 is the one attacked.[27] Nucleophilic attack on the atom C3 is observed in the S' binding mode, as exemplified by CA-074 binding to cathepsin B.[28]

The kinetic data[22] and the results of the crystal-soaking experiments indicate that M^{pro} reacts with the 2*S*,3*S* diastereomer of APE only, and not with its 2*R*,3*R* diastereomer. Interestingly, the order of inhibitory activities for APEs towards most cysteine peptidases in clan CD is 2*S*,3*S* > 2*R*,3*R* > *trans* > *cis* (the racemic mixture of the 2*S*,3*R* and 2*R*,3*S* diastereomers).[2,3,4] In order to explain this, models have been built for each of the four diastereomers of APE in the substrate-binding regions of M^{pro} before the nucleophilic attack of its Cys145 S^{γ} (Figures 2-6a to 2-6d), based on the molecular structure of the M^{pro}:APE complex. These models suggest that, in order for APE to be accommodated in the substrate-binding regions of M^{pro}, the epoxide atom C3 of APE must be in the *S* configuration, otherwise the epoxide moiety would sterically clash with the `back-wall' of the active site of M^{pro} and with the aza-peptide component of APE itself (Figures 2-6b and 2-6c). In the *S* configuration, the atom C3 of APE is also in better geometry with respect to Cys145 S^{γ} of M^{pro} for nucleophilic

Figure 2-6 Models for each of the four diastereomers of APE in the substrate-binding region of M^{pro} before the nucleophilic attack by its Cys145 S^{γ}. (a) 2*S*, 3*S*, (b) 2*R*, 3*R*, (c) 2*S*, 3*R*, (d) 2*R*, 3*S*.

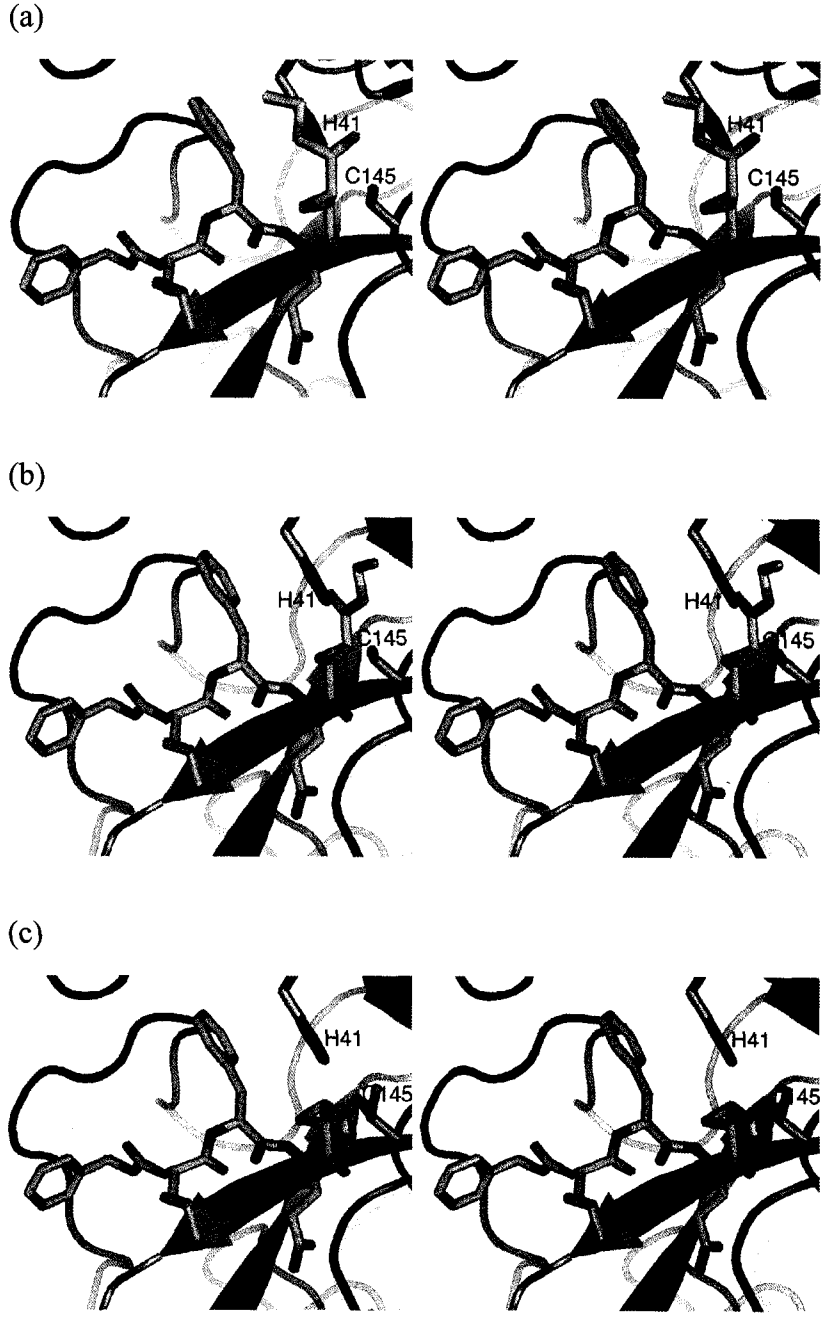
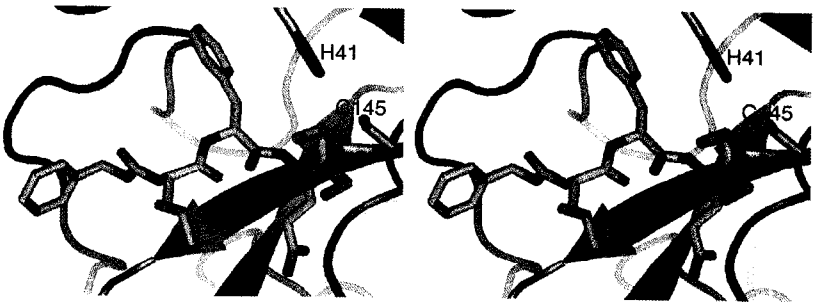


Figure 2-6 (continued)

(d)



attack. The atom C2 of APE must be in the *S* configuration as well, in order to allow for the interactions between the epoxide moiety and the active site of M^{pro}, otherwise the epoxide moiety would sterically clash with the loop constituting the oxyanion hole of M^{pro} (Figure 2-6d). The model for the 2*S*,3*S* diastereomer of APE also suggests that the distance between His41 $N^{\varepsilon 2}$ of M^{pro} and the epoxide O of APE is 4 to 5 Å, and that these two atoms are not well aligned for proton transfer (Figure 2-6a). The opening of the epoxide ring may involve two steps separated by a conformational rearrangement of APE: (1) the protonation of the epoxide O, and (2) the nucleophilic attack on the epoxide atom C3. From this model, it is not possible to determine the order in which these steps occur. It would be energetically more favorable for protonation to be the first step. M^{pro} may be sufficiently flexible in solution to allow for the alignment of His41 $N^{\varepsilon 2}$ of M^{pro} with the epoxide O of APE for proton transfer. In such a scenario, protonation and nucleophilic attack could occur in a concerted manner, enabling the epoxide ring of APE to open in a single step. Much of the inhibitory mechanism of APE towards M^{pro} remains to be elucidated. Rigorous treatment of this issue using methodologies in organic chemistry will be required.

All of the three molecular structures reported here successfully visualize the 'N-fingers' of both protomers of M^{pro}. As shown by the molecular structures of the unbound M^{pro} and the M^{pro}:APE complex, the N-finger of each protomer contributes two ionic interactions to the dimerization of M^{pro}: one between the protonated α-amino group of Ser1 and the side-chain carboxylate group of Glu166 from opposite protomers, and one between the side chains of Arg4 and Glu290 from opposite protomers. In the molecular structure of the $M^{pro}_{+A(-1)}$:APE complex, the former ion-

pair does not exist, but the substrate-binding regions of both protomers are still capable of accommodating APE in a similar manner as in the M^{pro}:APE complex. This suggests that the former ionic interaction is of less importance.

2.5 References

1. Blanchard, J. E., Elowe, N. H., Huitema, C., Fortin, P. D., Cechetto, J. D., Eltis, L. D. & Brown, E. D. (2004). High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. *Chem. Biol.* **11**, 1445-1453.

2. Asgian, J. L., James, K. E., Li, Z. Z., Carter, W., Barrett, A. J., Mikolajczyk, J., Salvesen, G. S. & Powers, J. C. (2002). Aza-peptide epoxides: a new class of inhibitors selective for clan CD cysteine proteases. *J. Med. Chem.* **45**, 4958-4960.

3. James, K. E., Gotz, M. G., Caffrey, C. R., Hansell, E., Carter, W., Barrett, A. J., McKerrow, J. H. & Powers, J. C. (2003). Aza-peptide epoxides: potent and selective inhibitors of Schistosoma mansoni and pig kidney legumains (asparaginyl endopeptidases). *Biol. Chem.* **384**, 1613-1618.

4. James, K. E., Asgian, J. L., Li, Z. Z., Ekici, O. D., Rubin, J. R., Mikolajczyk, J., Salvesen, G. S. & Powers, J. C. (2004). Design, synthesis, and evaluation of aza-peptide epoxides as selective and potent inhibitors of caspases-1, -3, -6, and -8. *J. Med. Chem.* **47**, 1553-1574.

5. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W., Jr. & Sweet, R. M., eds.), Vol. 276: Macromolecular Crystallography, Part A, pp. 307-326 Academic Press, New York, NY.

6. Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

7. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Cryst.* **D59**, 1131-1137.

8. Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-31.

9. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13190-13195.

10. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.

11. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.

12. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240-255.

13. McRee, D. E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.

14. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-91.

15. Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593-604.

16. Cohen, G. H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Cryst.* **30**, 1160-1161.

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

18. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.

19. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **21**, 3213-3224.

20. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763-1767.

21. Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J. & Lai, L. (2004). Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **279**, 1637-1642.

22. Lee, T. W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. G. (2005). Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J. Mol. Biol.* **353**, 1137-1151.

23. Thiel, V., Ivanov, K. A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E. & Ziebuhr, J. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305-2315.

24. Powers, J. C., Asgian, J. L., Ekici, O. D. & James, K. E. (2002). Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* **102**, 4639-4750.

25. Rozman-Pungercar, J., Kopitar-Jerala, N., Bogyo, M., Turk, D., Vasiljeva, O., Stefe, I., Vandenabeele, P., Bromme, D., Puizdar, V., Fonovic, M., Trstenjak-Prebanda, M., Dolenc, I., Turk, V. & Turk, B. (2003). Inhibition of papain-like cysteine proteases and legumain by caspase-specific inhibitors: when reaction mechanism is more important than specificity. *Cell Death Differ.* **10**, 881-888.

26. Ganesan, R., Jelakovic, S., Campbell, A. J., Li, Z. Z., Asgian, J. L., Powers, J. C. & Grutter, M. G. (2006). Exploring the S4 and S1 prime subsite specificities in caspase-3 with aza-peptide epoxide inhibitors. *Biochemistry* **45**, 9059-9067.

27. Varughese, K. I., Ahmed, F. R., Carey, P. R., Hasnain, S., Huber, C. P. & Storer, A. C. (1989). Crystal structure of a papain-E-64 complex. *Biochemistry* **28**, 1330-1332.

28. Yamamoto, A., Tomoo, K., Hara, T., Murata, M., Kitamura, K. & Ishida, T. (2000). Substrate specificity of bovine cathepsin B and its inhibition by CA074, based on crystal structure refinement of the complex. *J. Biochem. (Tokyo)* **127**, 635-643.

29. Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.

Chapter 3

Crystal structures reveal an induced-fit binding of a substrate-like aza-peptide epoxide to SARS coronavirus main peptidase

3.1 Overview

This chapter reports the molecular structures of the unbound SARS-CoV M^{pro} at pH 6.0 in space group $P2_1$ (resolution: 1.8 Å), and of $M^{pro}_{+A(-1)}$ at pH 6.5 in space group $P4_32_12$, both in the absence and the presence of APE (resolutions: 2.0 Å). Structural comparisons with M^{pro} (both the wild type and the variants) in other pH values, space groups and inhibitor-binding states show the conformational variability of the active sites and the S1 specificity pockets of the two protomers of M^{pro}. Detailed analysis and comparison of the catalytically competent and the catalytically incompetent conformations are presented. The results lead to the suggestion that these parts of M^{pro} are conformationally flexible, and establish a dynamic equilibrium between the two sets of conformations. How the position of this equilibrium could be affected by factors such as the pH, the integrity of the interactions among residues forming the `floors' of the S1 specificity pockets of M^{pro}, and the substrate binding is discussed. Interestingly, the effects of the substrate-like inhibitor binding suggest that the substrate binds to M^{pro} in an induced-fit manner, and this favors the progression of the catalytic mechanism of M^{pro}.

3.2 Materials and methods

Preparation of M^{pro}, $M^{pro}_{+A(-1)}$ and the APEs

The M^{pro} and the $M^{pro}_{+A(-1)}$ used in this study were prepared by the group of Lindsay D. Eltis in the Department of Microbiology and Immunology, University of British Columbia; whereas the APEs used in this study, Cbz-Leu-Phe-AGln-(2*S*,3*S*)EP-COOEt and Cbz-Leu-Phe-AGln-(2*R*,3*R*)EP-COOEt, were prepared by the group of James C. Powers in the School of Chemistry and Biochemistry, Georgia Institute of Technology. Experimental details are given in Section 2.2.

Crystallization, crystal soaking and cryo-protection

Crystallization, crystal soaking and cryo-protection were done by Maia M. Cherney in the group of Michael N. G. James. Before crystallization, both M^{pro} and $M^{pro}_{+A(-1)}$ were dialyzed against 20 mM NaCl and 20 mM Tris-HCl (pH 7.5), and concentrated to 10 mg/mL. All crystals were grown at ambient temperature by the hanging-drop vapor diffusion method. For the $P2_1$ crystals, the reservoir solution contained 2 % (w/v) polyethylene glycol (M_r 20,000), 3 % (v/v) dimethyl sulfoxide, 1 mM dithiothreitol and 0.1 M MES (pH 6.0). The final crystallization drop contained equal amounts of the M^{pro} solution and the reservoir solution. Crystals of a thick-plate habit grew in 3 to 5 days to a size of about 0.3 mm × 0.3 mm × 0.1 mm. For the $P4_32_12$ crystals, the growth conditions were the same as those for the $P2_12_12_1$ crystals (Section 2.2); the reservoir solution contained 50 mM ammonium acetate, 6 % (w/v) polyethylene glycol (M_r 8,000), 3 % (v/v) ethylene glycol, 1 mM dithiothreitol and 0.1 M MES (pH 6.5). The drop contained equal amounts of the $M^{pro}_{+A(-1)}$ solution and the reservoir solution. Bipyrimidal crystals grew in 3 to 5 days to a size of about 0.1

mm × 0.1 mm × 0.1 mm. Macroseeding was done to improve the quality of the crystals. Crystals of good quality were selected and soaked overnight in drops having the same compositions as their reservoir solutions plus the APE chosen for this study present at 3 mM. Cryo-protectants had essentially the same compositions as reservoir solutions, except for the inclusion of 25 % (v/v) ethylene glycol and the exclusion of dimethyl sulfoxide and dithiothreitol. Crystals were soaked for about 10 seconds and then immediately flash-cooled in the liquid nitrogen for storage and shipment to the synchrotron beamline.

Data collection and processing, and structure solution and refinement

The X-ray diffraction data from all crystals were collected at the synchrotron Beamline 8.3.1 (equipped with an ADSC-Q210 CCD detector) at the Advanced Light Source in the Lawrence Berkeley National Laboratory. All data sets were indexed, scaled and merged using DENZO and SCALEPACK.[1] Structure solution and refinement were done in CCP4.[2,3] All structures were solved by the molecular replacement method using MOLREP[4], with the molecular structure of the unbound M^{pro} in space group C2 (Chapter 2; PDB accession code: 2A5A) being used as the search model. In the structure of the $M^{pro}_{+A(-1)}$:APE complex, APE was located as outstanding electron densities in the substrate-binding region of $M^{pro}_{+A(-1)}$ in both the $|F_o|-|F_c|,\alpha_c$ (contoured at 3σ and 4σ) and the $2|F_o|-|F_c|,\alpha_c$ (contoured at 1σ) maps. A restraint was applied to the covalent bond between Cys145 S^{γ} of $M^{pro}_{+A(-1)}$ and the epoxide atom C3 of APE, on the basis that the length of a C-S single bond is normally about 1.8 Å. All structures were then iteratively refined using REFMAC,[5] and

manually adjusted when needed using XtalView/Xfit.[6] The stereochemical qualities of the final structures were assessed using PROCHECK.[7]

Structure analysis

Graphical representations of the molecular structures were prepared using PyMOL (http://www.pymol.org/). Superpositions of molecular structures were done using ALIGN,[8,9] based on the main-chain atoms (amide N, C^{α}, and carbonyl C and O). Peptidase-inhibitor interactions were analyzed using LIGPLOT.[10]

Protein Data Bank accession codes

The atomic coordinates and the structure factors of all the structures have been deposited in the Protein Data Bank. The accession code is 2GT7 for the structure of the unbound M^{pro}, 2GT8 for the structure of the unbound $M^{pro}_{+A(-1)}$, and 2GTB for the structure of the $M^{pro}_{+A(-1)}$:APE complex.

3.3 Results and discussion

Structure determination

The parameters and statistics derived from data processing and structure refinement are summarized in Table 1. M^{pro} crystallizes in space group $P2_1$ in conditions slightly different from those reported previously.[11] Each asymmetric unit contains both protomers (A and B) of the physiological dimer. In the electron density maps of the unbound M^{pro}, residues 1A to 44A, 50A to 305A and 1B to 302B could be identified. In the Ramachandran plot of this structure, Asp33A, Asn84A and Asn84B are in the generously allowed regions, and Asp33B, Glu47B, Tyr154A and Tyr154B are in the disallowed regions. In both independent protomers, the main-chain NH of Asp33 forms a hydrogen bond with Thr98 $O^{\gamma 1}$, and the carbonyl O of

Table 3-1 Parameters and statistics derived from X-ray diffraction data processing and structure refinement

	Unbound M^{pro}	Unbound $M^{pro}_{+A(-1)}$	$M^{pro}_{+A(-1)}$:APE
Data processing			
Wavelength (Å)	1.116	1.116	1.116
Resolution limit[a] (Å)	41.25 – 1.82	24.85 – 2.00	24.78 – 2.00
	(1.89 – 1.82)	(2.07 – 2.00)	(2.07 – 2.00)
Space group	$P2_1$	$P4_32_12$	$P4_32_12$
Unit-cell constants			
a (Å)	52.39	70.29	70.09
b (Å)	96.19	70.29	70.09
c (Å)	67.91	102.87	103.86
α(°)	90.00	90.00	90.00
β(°)	102.91	90.00	90.00
γ(°)	90.00	90.00	90.00
Mosaicity (°)	0.63	0.65	0.58
Number of unique reflections	56,644 (5,313)	17,964 (1,739)	18,098 (1,760)
Redundancy	2.0 (1.8)	6.8 (6.2)	7.1 (6.5)
Completeness (%)	96.4 (90.6)	99.3 (97.8)	99.8 (99.4)
R_{sym}[b] (%)	3.6 (20.1)	8.8 (48.8)	7.4 (44.4)
$< I / \sigma(I) >$	20.0 (3.4)	19.6 (4.2)	23.7 (4.9)

[a] Numbers in the parentheses refer to the highest resolution bins.

[b] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_{hkl,i} - <I_{hkl}>| / \Sigma_{hkl}\Sigma_i I_{hkl,i}$, where $I_{hkl,i}$ and $<I_{hkl}>$ are the i-th observed intensity and the average intensity of reflection hkl, respectively.

Table 3-1 (continued)

	Unbound M^{pro}	Unbound $M^{pro}_{+A(-1)}$	$M^{pro}_{+A(-1)}$:APE
Structure refinement			
Resolution range (Å)	33.12 – 1.82	24.85 – 2.00	24.78 – 2.00
R_{work}[c] (%)	17.0	17.5	19.6
R_{free}[c] (%)	21.7	24.1	27.0
Number of non-hydrogen atoms per asymmetric unit (average *B* factor, Å^2)			
Protein	4,753 (30.12)	2,332 (23.00)	2,332 (23.93)
APE	Not applicable	Not applicable	46 (22.69)
Solvent	591 (37.81)	162 (34.38)	127 (32.49)
rms deviation from ideal geometry			
Bond lengths (Å)	0.019	0.019	0.022
Bond angles (°)	1.778	1.765	2.064
Ramachandran plot			
Favored (%)	91.3	90.0	90.0
Allowed (%)	7.3	8.5	8.1
Generously allowed (%)	0.6	0.8	0.8
Disallowed (%)	0.8	0.8	1.2

[c] $R_{work} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and the calculated structure factor amplitudes of a particular reflection, respectively, and the summation is over 95 % of the reflections in the specified resolution range. The remaining 5 % of the reflections were randomly selected before the structure refinement and not included in the structure refinement. R_{free} was calculated over these reflections using the same equation as for R_{work}.[28]

Asp33 forms hydrogen bonds with Trp31 $N^{\varepsilon 1}$ and the main-chain NH of Asn95. Also, Asn84 $N^{\delta 2}$ forms a hydrogen bond with the carbonyl O of Glu178. The poorly defined electron densities of the side chains of Glu47B, Tyr154A and Tyr154B indicate dynamic disorder.

$M^{pro}_{+A(-1)}$ crystallized in space group $P2_12_12_1$ (Chapter 2) as well as $P4_32_12$. Each asymmetric unit of the latter contains only one protomer of the dimer. The two protomers of each dimer are related by the crystallographic two-fold symmetry axis parallel with the C-face diagonal of the unit cell. In the electron density maps of the unbound $M^{pro}_{+A(-1)}$, residues 3 to 300 of the protomer were clearly identified. In the Ramachandran plot of this structure, Asp33 and Ser139 are in the generously allowed regions, and Asn84 and Tyr154 are in the disallowed regions. The side chain of Tyr154 makes contact with those of Ile78 and probably Arg76 from a neighboring asymmetric unit. The electron density of Ser139 is not well defined. In the electron density maps of the $M^{pro}_{+A(-1)}$:APE complex, residues 2 to 300 of the protomer were identified. In the Ramachandran plot of this structure, Asp33 and Asn277 are in the generously allowed regions, and Asn84, Tyr154 and Ile286 are in the disallowed regions. The side chains of Thr285 and Ile286 from opposite protomers of the dimer contact each other. The electron density of Asn277 is not well defined.

Superpositions of the molecular structures reported here with all of those previously reported show no significant difference in the protomer orientation and the protein fold (Table 3-2).

Table 3-2 Root-mean-square differences (rmsd; in Å) for superpositions of the molecular structures of M^{pro} (wild type and variants) determined in different conditions[a]

	Unbound M^{pro} [b] ($P2_1$, pH 6.0)	**Unbound $M^{pro}_{+A(-1)}$ [b] ($P4_32_12$, pH 6.5)**	**$M^{pro}_{+A(-1)}$:APE[b] ($P4_32_12$, pH 6.5)**
Unbound M^{pro} [b] ($P2_1$, pH 6.0)	-	AP: 0.47 (1,075) BP: 0.55 (1,098) DD: 0.69 (2,082)	AP: 0.42 (1,077) BP: 0.43 (1,098) DD: 0.60 (2,142)
Unbound M^{pro} [15] ($P4_32_12$, pH 5.9)	PA: 0.50 (1,025) PB: 0.61 (1,076) DD: 0.70 (2,008)	PP: 0.30 (1,079) DD: 0.35 (2,158)	PP: 0.43 (1,080) DD: 0.45 (1,986)
Unbound M^{pro} [11] ($P2_1$, pH 6.0)	AA: 0.30 (1,125) BB: 0.22 (1,046) AB: 0.66 (1,166) BA: 0.60 (1,083) DD: 0.34 (2,169)	AP: 0.42 (1,094) BP: 0.46 (1,013) DD: 0.69 (2,130)	AP: 0.50 (1,092) BP: 0.43 (1,051) DD: 0.61 (2,162)
M^{pro}:CMK[11] ($P2_1$, pH 6.0)	AA: 0.33 (1,138) BB: 0.26 (1,093) AB: 0.62 (1,166) BA: 0.62 (1,104) DD: 0.32 (2,221)	AP: 0.49 (1,108) BP: 0.62 (1,143) DD: 0.72 (2,158)	AP: 0.49 (1,109) BP: 0.49 (1,102) DD: 0.66 (2,223)
Unbound M^{pro} [16] (C2, pH 6.5)	PA: 0.59 (1,121) PB: 0.29 (1,117) DD: 0.51 (2,053)	PP: 0.63 (1,109) DD: 0.63 (2,166)	PP: 0.48 (1,105) DD: 0.49 (2,142)
M^{pro}:APE[16] (C2, pH 6.5)	PA: 0.69 (1,138) PB: 0.30 (1,090) DD: 0.47 (1,816)	PP: 0.73 (1,115) DD: 0.69 (2,148)	PP: 0.52 (1,125) DD: 0.64 (2,208)
Unbound M^{pro} [18] ($P2_12_12$, pH 6.5)	PA: 0.85 (1,120) PB: 0.61 (1,119) DD: 1.02 (2,234)	PP: 0.85 (1,125) DD: 1.11 (2,196)	PP: 0.68 (1,093) DD: 0.91 (2,192)
Unbound M^{pro} [15] ($P2_12_12$, pH 6.6)	PA: 0.92 (1,144) PB: 0.64 (1,131) DD: 0.73 (2,025)	PP: 0.96 (1,177) DD: 1.06 (2,304)	PP: 0.75 (1,134) DD: 0.80 (2,222)

Table 3-2 (continued)

	Unbound M^{pro} [b] ($P2_1$, pH 6.0)	**Unbound $M^{pro}_{+A(-1)}$ [b] ($P4_32_12$, pH 6.5)**	**$M^{pro}_{+A(-1)}$:APE [b] ($P4_32_12$, pH 6.5)**
Unbound M^{pro} [23] ($P2_12_12$, pH 7.0)	PA: 1.11 (1,149) PB: 0.80 (1,135) DD: 0.91 (2,086)	PP: 1.12 (1,160) DD: 1.29 (2,308)	PP: 0.98 (1,170) DD: 1.08 (2,286)
Unbound M^{pro} [11] ($P2_1$, pH 7.6)	AA: 0.56 (1,091) BB: 0.94 (1,161) AB: 0.89 (1,168) BA: 0.88 (1,128) DD: 0.82 (2,030)	AP: 0.64 (1,108) BP: 0.80 (1,150) DD: 0.98 (2,258)	AP: 0.72 (1,138) BP: 0.86 (1,154) DD: 1.02 (2,174)
Unbound M^{pro} [11] ($P2_1$, pH 8.0)	AA: 0.60 (1,092) BB: 0.91 (1,156) AB: 0.89 (1,162) BA: 0.88 (1,136) DD: 0.84 (2,029)	AP: 0.68 (1,124) BP: 0.81 (1,155) DD: 0.95 (2,215)	AP: 0.72 (1,141) BP: 0.82 (1,138) DD: 0.97 (2,141)
Unbound $M^{pro}_{+A(-1)}$ [b] ($P4_32_12$, pH 6.5)	-	-	PP: 0.38 (1,092) DD: 0.49 (2,194)
$M^{pro}_{+A(-1)}$:APE [16] ($P2_12_12_1$, pH 6.5)	AA: 0.53 (1,114) BB: 0.38 (1,124) AB: 0.39 (1,116) BA: 0.53 (1,117) DD: 0.70 (2,282)	AP: 0.49 (1,015) BP: 0.54 (1,116) DD: 0.69 (2.150)	AP: 0.30 (1,017) BP: 0.35 (1,089) DD: 0.47 (2,110)
Unbound M^{pro}_{+SL} [c] ($P2_1$, pH 6.5)	AA: 0.57 (1,136) BB: 0.61 (1,126) AB: 0.20 (1,119) BA: 0.38 (1,118) DD: 0.99 (2,206)	AP: 0.53 (1,095) BP: 0.44 (1,066) DD: 0.63 (2,042)	AP: 0.42 (1,096) BP: 0.50 (1,068) DD: 0.60 (2,141)
M^{pro}-C145A [d]:product [23] (C2, pH 9.0)	AA: 0.97 (1,047) BB: 0.90 (1,152) AB: 0.83 (1,069) BA: 1.01 (1,133) DD: 0.84 (1,961)	AP: 0.85 (1,000) BP: 1.10 (1,142) DD: 1.02 (2,119)	AP: 0.94 (1,073) BP: 0.97 (1,161) DD: 0.91 (2,084)

Table 3-2 (continued)

	Unbound M^{pro} [b]	**Unbound $M^{pro}_{+A(-1)}$ [b]**	**$M^{pro}_{+A(-1)}$:APE [b]**
	($P2_1$, pH 6.0)	**($P4_32_12$, pH 6.5)**	**($P4_32_12$, pH 6.5)**
M^{pro}_{+GPLGS}[e]:I2[17]	AA: 0.55 (1,151)	AP: 0.81 (1,155)	AP: 0.75 (1,162)
($P2_1$, pH 6.0)	BB: 0.65 (1,174)	BP: 0.82 (1,135)	BP: 0.68 (1,139)
	AB: 0.86 (1,177)	DD: 1.05 (2,292)	DD: 1.00 (2,307)
	BA: 0.78 (1,142)		
	DD: 0.76 (2,326)		
M^{pro}_{+GPLGS}:N1[17]	AA: 0.31 (1,073)	AP: 0.65 (1,121)	AP: 0.56 (1,134)
($P2_1$, pH 6.0)	BB: 0.23 (1,097)	BP: 0.63 (1,111)	BP: 0.49 (1,134)
	AB: 0.68 (1,146)	DD: 0.78 (2,174)	DD: 0.75 (2,224)
	BA: 0.64 (1,118)		
	DD: 0.38 (2,187)		
M^{pro}_{+GPLGS}:N3[17]	AA: 0.30 (1,084)	AP: 0.62 (1,111)	AP: 0.53 (1,128)
($P2_1$, pH 6.0)	BB: 0.25 (1,116)	BP: 0.66 (1,132)	BP: 0.49 (1,133)
	AB: 0.66 (1,163)	DD: 0.80 (2,196)	DD: 0.71 (2,220)
	BA: 0.66 (1,136)		
	DD: 0.31 (2,166)		
M^{pro}_{+GPLGS}:N9[17]	AA: 0.35 (1,069)	AP: 0.65 (1,104)	AP: 0.58 (1,133)
($P2_1$, pH 6.0)	BB: 0.24 (1,094)	BP: 0.63 (1,111)	BP: 0.51 (1,150)
	AB: 0.72 (1,158)	DD: 0.84 (2,238)	DD: 0.74 (2,251)
	BA: 0.64 (1,126)		
	DD: 0.34 (2,162)		

Table 3-2 (continued)

[a] Superpositions of molecular structures were done for pairs of protomers (PP, PA, AP, PB, BP, AA, BB and AB) and pairs of dimers (DD). The first letter corresponds to the part from the molecular structure listed along the left of the table and the second letter corresponds to the part from the molecular structure listed along the top of the table. P, protomer (applicable to any structure whose asymmetric unit contains only one protomer); A, protomer A; B, protomer B; D, dimer. All superpositions started with residues 3 to 300 of each protomer (total number of main-chain atoms: 1,192) included in the calculations. The number of main-chain atoms included in the final calculation of rmsd for each superposition is given in the parentheses.

[b] These are the molecular structures reported here.

[c] M^{pro}_{+SL}: the M^{pro} variant with additional residues Ser-Leu at the N-terminus of each protomer (J. B. Bonanno *et al.*, unpublished results; PDB accession code: 1Q2W)

[d] M^{pro}-C145A: the C145A variant of M^{pro}

[e] M^{pro}_{+GPLGS}: the M^{pro} variant with additional residues Gly-Pro-Leu-Gly-Ser at the N-terminus of each protomer

Active sites and substrate-binding regions of the unbound M^{pro}

The previously reported molecular structure of M^{pro} in space group $P2_1$ at pH 6.0 showed the collapse of the active site and S1 specificity pocket in one of the protomers, whereas the molecular structures in the same space group at pH 7.6 and 8.0 showed all the active sites to be in the catalytically competent conformation (Table 3-3). Based on this trend, a pH-triggered switch for the catalytic activity of M^{pro} was proposed.[11] We have now grown crystals of M^{pro} in the same space group ($P2_1$ with the same unit-cell constants) at pH 6.0 under slightly different conditions. The active sites and the S1 specificity pockets of both protomers are in the catalytically competent conformation (Figures 3-1a, 3-1b, 3-2a, 3-2b, 3-3a and 3-3b). More specifically, superposition of protomers A and B of the resulting M^{pro} structure (rmsd: 0.50 Å for 1,070 out of 1,172 main-chain atoms) shows good agreement in most atomic positions. In both independent protomers, the catalytic dyad has a distance of 3.6 Å between His41 $N^{\varepsilon 2}$ and Cys145 S^{γ}, and Cys145 S^{γ} is coplanar with the atoms of the imidazole ring of His41. Residues Gly143 to Cys145 are in the proper conformation to form the oxyanion hole that accommodates the carbonyl O of the scissile peptide bond of the substrate. The position to be occupied by the carbonyl O is occupied by a water molecule. The oxygen of the water molecule forms hydrogen bonds at distances of 3.0 Å with the main-chain NH groups of Gly143 and Cys145 (Figures 3-2a and 3-2b).

Previous studies suggest strongly that the predominant S1 specificity of M^{pro} for Gln is determined primarily by the conserved residue His163.[12,13,14] In both independent protomers, the orientation of the imidazole ring of His163 is determined

Table 3-3 Molecular structures of M^{pro} (wild type and variants) determined in different conditions

Reference	Inhibitor	Space group	pH	In the catalytically competent conformation? Protomer A[a]	Protomer B[a]
	M^{pro}				
b	**-**	**$P2_1$**	**6.0**	**Yes**	**Yes**
15	-	$P4_32_12$	5.9	No	No
11	-	$P2_1$	6.0	Yes	No
11	CMK	$P2_1$	6.0	Yes	No
16	-	C2	6.5	Yes	Yes
16	APE	C2	6.5	Yes	Yes
18	-	$P2_12_12$	6.5	Yes	Yes
15	-	$P2_12_12$	6.6	No	No
23	-	$P2_12_12$	7.0	Yes	Yes
11	-	$P2_1$	7.6	Yes	Yes
11	-	$P2_1$	8.0	Yes	Yes
	$M^{pro}{}_{+A(-1)}$				
b	**-**	**$P4_32_12$**	**6.5**	**No**	**No**
b	**APE**	**$P4_32_12$**	**6.5**	**Yes**	**Yes**
16	APE	$P2_12_12_1$	6.5	Yes	Yes
	$M^{pro}{}_{+SL}$				
c	-	$P2_1$	6.5	Yes	Yes
	M^{pro}-C145A				
23	product	C2	9.0	Yes	Yes

Table 3-3 (continued)

Reference	Inhibitor	Space group	pH	In the catalytically competent conformation? Protomer A[a]	Protomer B[a]
	$M^{pro}{}_{+GPLGS}$				
17	I2	$P2_1$	6.0	Yes	Yes
17	N1	$P2_1$	6.0	Yes	Yes
17	N3	$P2_1$	6.0	Yes	Yes
17	N9	$P2_1$	6.0	Yes	Yes

[a] For any structure whose asymmetric unit contains only one protomer, that protomer represents both protomers A and B.

[b] These are the molecular structures reported here.

[c] J. B. Bonanno *et al.*, unpublished results; PDB accession code: 1Q2W

Figure 3-1 Electron densities in the $|F_o|-|F_c|,\alpha_c$ omit maps for residues Lys137 to Ser144 of M^{pro} and $M^{pro}_{+A(-1)}$. (a) Protomer A of the unbound M^{pro}. (b) Protomer B of the unbound M^{pro}. (c) Unbound $M^{pro}_{+A(-1)}$. (d) APE-bound $M^{pro}_{+A(-1)}$. The residues in (c) are in the catalytically incompetent conformation.

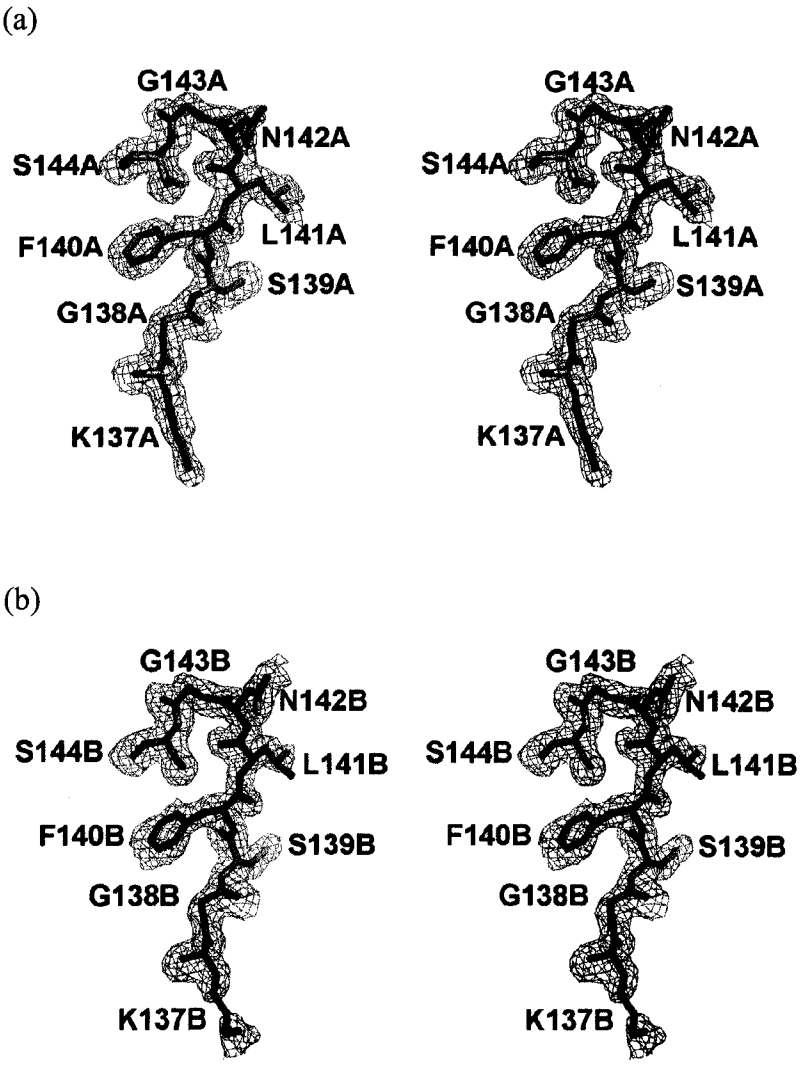
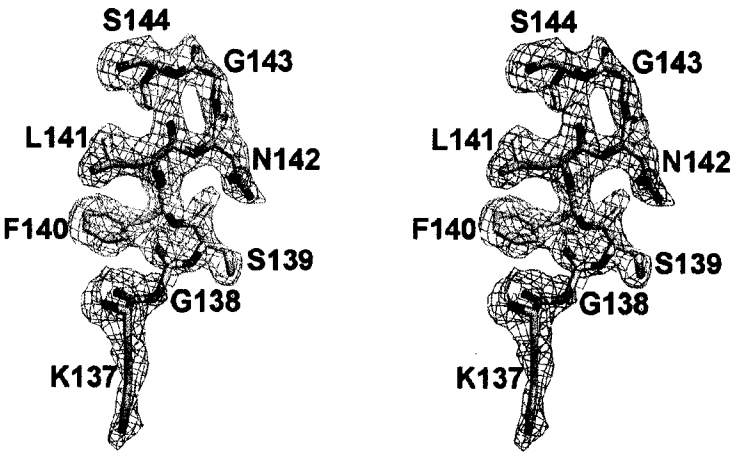


Figure 3-1 (continued)

(c)



(d)

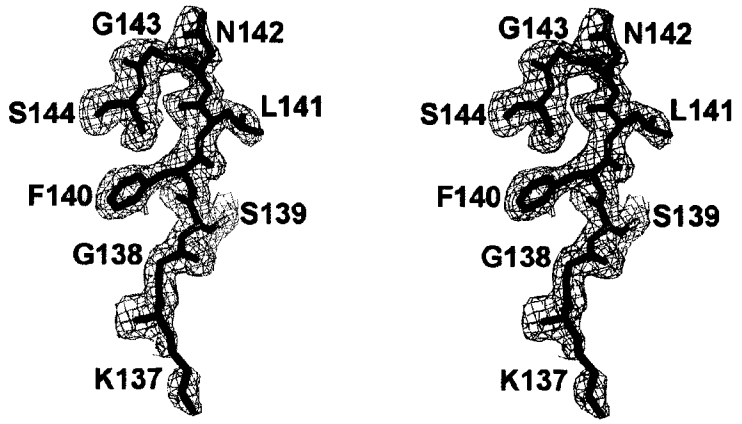
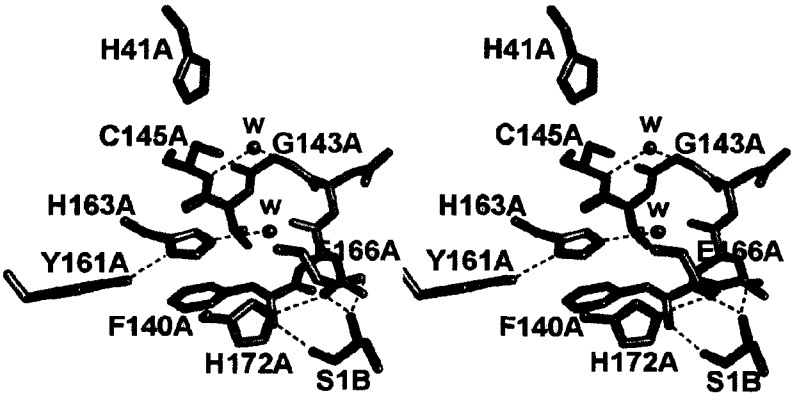


Figure 3-2 Active sites and S1 specificity pockets of M^{pro} and $M^{pro}_{+A(-1)}$, viewed from the S2 specificity pockets. (a) Protomer A of the unbound M^{pro}. (b) Protomer B of the unbound M^{pro}. (c) Unbound $M^{pro}_{+A(-1)}$. (d) APE-bound $M^{pro}_{+A(-1)}$ (For clarity, the APE is not shown.). Hydrogen bonds are indicated by dashed lines. Water molecules are labelled as w.

(a)



(b)

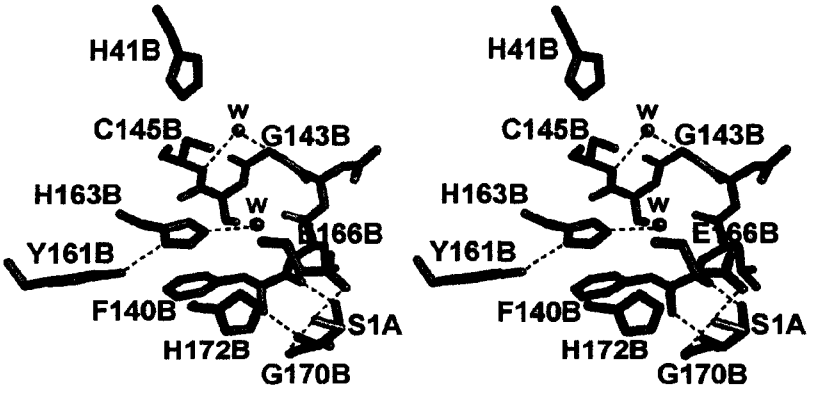
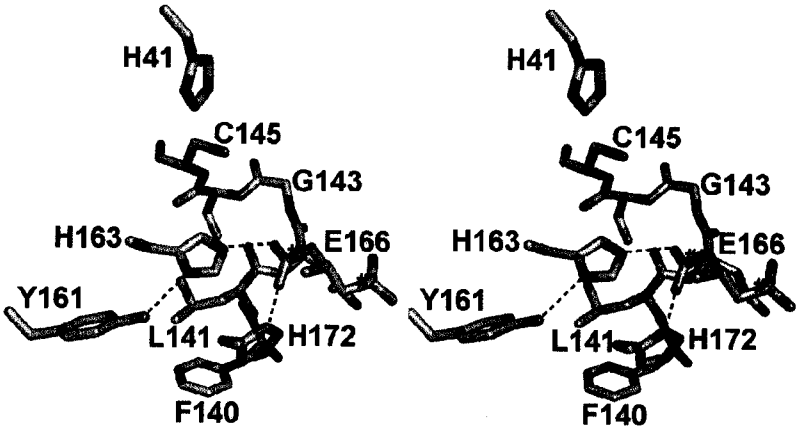


Figure 3-2 (continued)

(c)



(d)

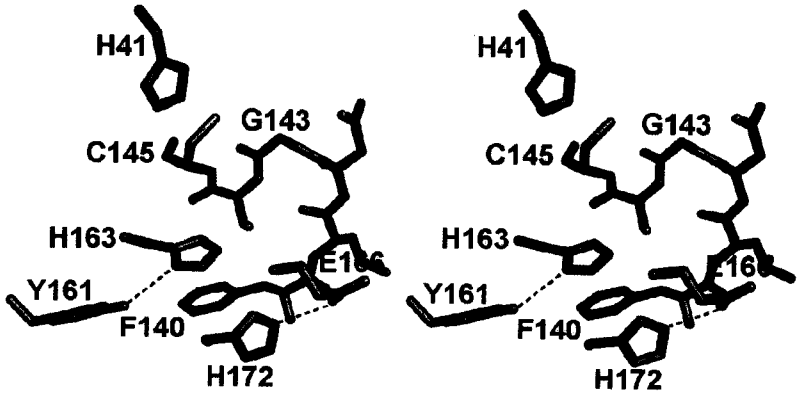
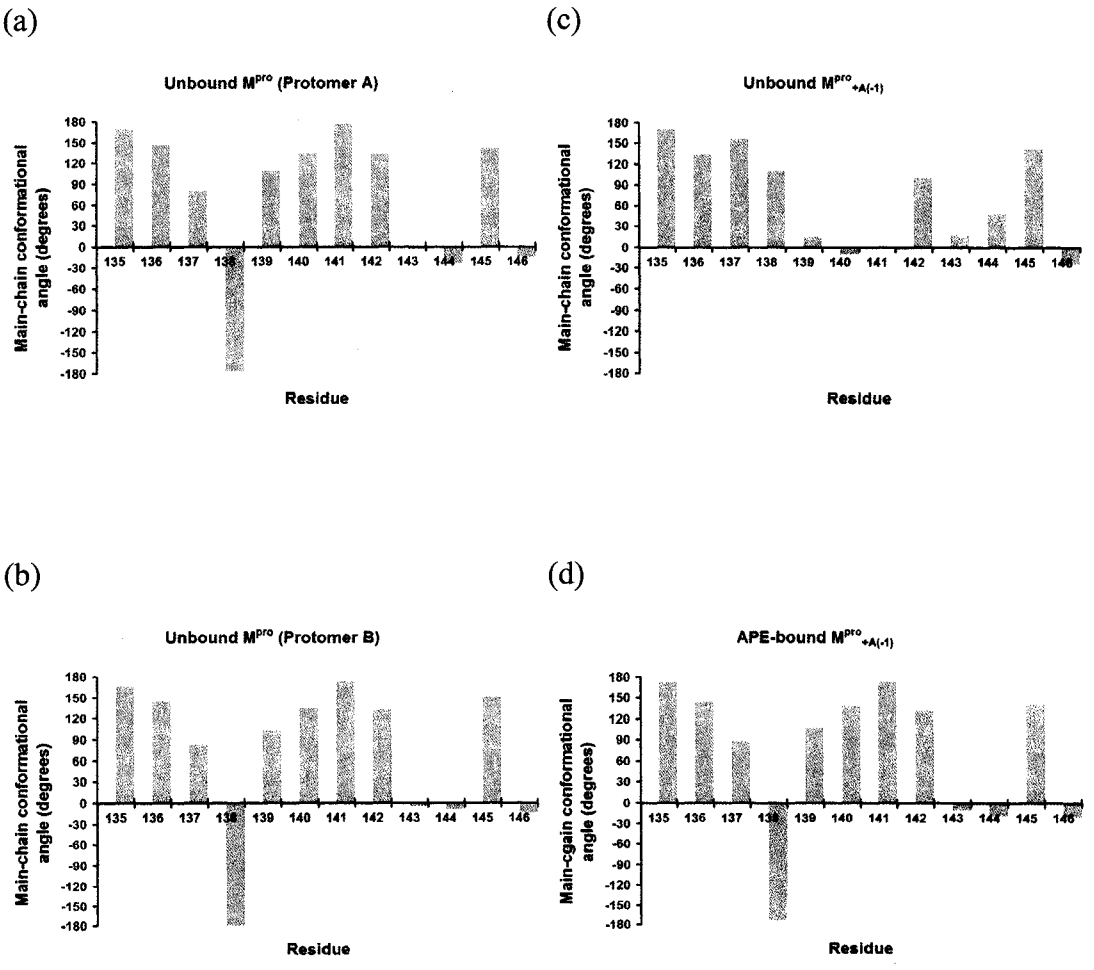


Figure 3-3 Main-chain conformational angles of residues Thr135 to Gly146 in the molecular structures of M^{pro} and $M^{pro}_{+A(-1)}$ reported here. The angles ϕ and ψ are represented by the yellow and the blue bars, respectively. (a) Protomer A of the unbound M^{pro} (overall positional uncertainty estimated based on maximum likelihood, σ_r: 0.15 Å). (b) Protomer B of the unbound M^{pro} (σ_r: 0.15 Å). (c) Unbound $M^{pro}_{+A(-1)}$ (σ_r: 0.23 Å). (d) APE-bound $M^{pro}_{+A(-1)}$ (σ_r: 0.28 Å).



by the hydrogen bonding between its $N^{\delta 1}$ and the phenolic OH of Tyr161 (3.0 Å) and by its π-stacking with the phenyl ring of Phe140 (distance between the geometric centers of the aromatic rings: 3.9Å). The position to be occupied by the side-chain carbonyl O of P1-Gln of the substrate is occupied by a water molecule, whose oxygen forms a hydrogen bond with His163 $N^{\varepsilon 2}$ (2.9 Å). This water molecule is coplanar with the atoms of the imidazole ring of His163. In each protomer, Phe140 and Glu166 interact with Ser1 of the opposite protomer to form the `floor' of the S1 specificity pocket (Figures 3-2a and 3-2b). In protomer A, Glu166A $O^{\varepsilon 1}$ forms a hydrogen bond (2.8 Å) with His172A $N^{\varepsilon 2}$, thereby constituting a `side wall' of the S1 specificity pocket similar to that in protomer A of previously reported molecular structures and the molecular-dynamic simulation model of M^{pro} at pH 6.0 (Figure 3-2a).[11,15] However, in protomer B of the molecular structure of M^{pro} reported here, the side chain of Glu166B rotates away, and His172B $N^{\varepsilon 2}$ forms a hydrogen bond (3.0 Å) with Ser1A O^{γ} of protomer A instead. Ser1A O^{γ} also forms a hydrogen bond (2.8 Å) with the carbonyl O of Gly170B (Figure 3-2b). The difference in the side-chain conformation of Glu166 in protomers A and B of the molecular structure of M^{pro} reported here indicates the weakness of the interaction between the side chains of Glu166 and His172, even though this interaction may acquire some ionic character by the protonation of His172 $N^{\delta 1}$. Contrary to those in protomer B of both the previously reported molecular structure and the molecular-dynamic simulation model of M^{pro} at pH 6.0,[11,15] the side chains of His163B and Glu166B in protomer B of the molecular structure of M^{pro} reported here do not interact with each other, probably because His163B is protonated only at its $N^{\varepsilon 2}$ atom and carries no charge.

As in the previously reported molecular structure of M^{pro} at pH 6.0,[11] protomers A and B in the current molecular structure show good agreement in the rest of the substrate-binding region, including the S2, the S4 and the S1' subsites. Superpositions with the molecular structures of M^{pro} in other conditions show that the rest of the substrate-binding regions in the molecular structure of M^{pro} reported here are in the catalytically competent conformation.

Active sites and substrate-binding regions of the unbound $M^{pro}_{+A(-1)}$

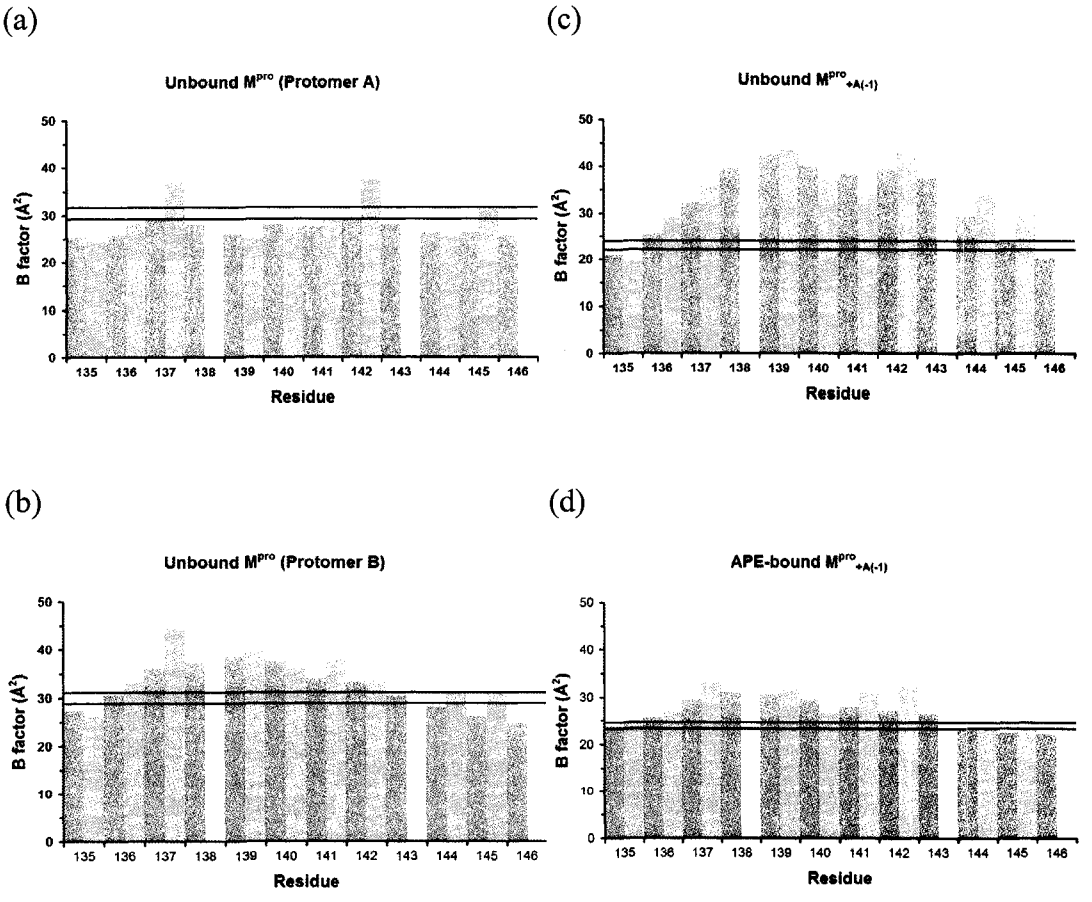
The asymmetric unit of the molecular structure of the $M^{pro}_{+A(-1)}$:APE complex in space group $P2_12_12_1$ contains both protomers of the $M^{pro}_{+A(-1)}$ dimer (Chapter 2). In that molecular structure, the active sites and the substrate-binding regions of both protomers are in the catalytically competent conformation (Table 3-3). In each protomer, the additional Ala at the N-terminus blocks Ser1 and disrupts its interactions with Phe140 and Glu166 of the opposite protomer. However, the `floor' of the S1 specificity pocket is only partly disrupted. More importantly, the presence of a ten-residue affinity tag at the N-terminus of each protomer reduces the specific activity of M^{pro} by less than an order of magnitude.[16] Similar observations are given by the molecular structure of the M^{pro} variant with additional residues Ser-Leu at the N-terminus of each protomer (M^{pro}_{+SL}; Table 3-3) (J. B. Bonanno *et al.*, unpublished results; PDB accession code: 1Q2W).

Attempts for the crystallization of $M^{pro}_{+A(-1)}$ in space group $P2_1$ have not been successful. Interestingly, $M^{pro}_{+A(-1)}$ crystallizes in space group $P4_32_12$ as well as in space group $P2_12_12_1$ under the same conditions. In space group $P4_32_12$, each asymmetric unit has one protomer of $M^{pro}_{+A(-1)}$. Superpositions of the resulting

molecular structure of the unbound $M^{pro}_{+A(-1)}$ with the molecular structures of M^{pro} in other conditions shows no differences in the protein fold or the protomer orientation (Table 3-2). Crystal contacts of the unbound $M^{pro}_{+A(-1)}$ in space group $P4_32_12$ do not involve any residues forming the active sites and the S1 specificity pockets of $M^{pro}_{+A(-1)}$. The molecular structure of the unbound $M^{pro}_{+A(-1)}$ in space group $P4_32_12$ shows good agreement in most atomic positions with that of the $M^{pro}_{+A(-1)}$:APE complex in space group $P2_12_12_1$ (Table 3-2). In the molecular structure of the unbound $M^{pro}_{+A(-1)}$ reported here, the catalytic dyad has a distance of 3.9 Å between His41 $N^{\varepsilon 2}$ and Cys145 S^{γ}, and Cys145 S^{γ} is coplanar with the atoms of the imidazole ring of His41. However, the oxyanion hole and the S1 specificity pocket are distorted (Figures 3-1c and 3-2c). The angles ϕ and ψ of residues Lys137 to Ser144 show dramatic differences as compared with those in the previously reported molecular structure of the $M^{pro}_{+A(-1)}$:APE complex (Figure 3-3c). The above-average B factors of these residues in the unbound $M^{pro}_{+A(-1)}$ indicate their high mobility relative to the rest of $M^{pro}_{+A(-1)}$ (Figure 3-4c). The oxyanion hole is not distorted as much. The main-chain NH groups of Gly143 and Cys145 are still oriented to donate hydrogen bonds that would stabilize the negatively-charged carbonyl O of the scissile peptide bond of the substrate, although no water molecule is found at the position to be occupied by the carbonyl O (Figure 3-2c).

The hydrogen bond between His163 $N^{\delta 1}$ and the phenolic OH of Tyr161 is preserved (3.1 Å). However, the imidazole ring of His163 is no longer π-stacked with the phenyl ring of Phe140. It makes contacts with the side chain of Leu141 instead (Figure 3-2c). Here the phenyl ring of Phe140 makes contacts with the side chains of

Figure 3-4 B factors of residues Thr135 to Gly146 in the molecular structures of M^{pro} and $M^{pro}_{+A(-1)}$ reported here. The main-chain and the side-chain B factors are represented by the red and the green bars, respectively. The average values of the main-chain and the side-chain B factors are indicated by the red and the green lines, respectively. There are no green bars for Gly138, Gly143 and Gly146, because these residues do not have side chains. (a) Protomer A of the unbound M^{pro}. (b) Protomer B of the unbound M^{pro}. (c) Unbound $M^{pro}_{+A(-1)}$. (d) APE-bound $M^{pro}_{+A(-1)}$.



Val114, Tyr126, Ile136 and His172, and the main-chain atoms of Lys137 and Gly138 from the parent protomer, and with the side chain of Arg4 from the opposite protomer. The carbonyl O of Phe140 also forms a long hydrogen bond with the phenolic OH of Tyr118 (3.4 Å). The electron density maps show two possible conformers of the side chain of Glu166. The occupancies of both conformers were fixed at 0.5 without refinement. In conformer #1, Glu166 $O^{\varepsilon 1}$ forms a hydrogen bond with His172 $N^{\delta 1}$ (2.3 Å), and Glu166 $O^{\varepsilon 2}$ forms a hydrogen bond with His163 $N^{\varepsilon 2}$ (2.8 Å); whereas in conformer #2, the side chain of Glu166 protrudes into the solvent (Figure 3-2c). Probably because of their high mobility, the additional Ala at the N-terminus, Ser1 and Gly2 of the protomer could not be identified in the electron density maps of the unbound $M^{pro}{}_{+A(-1)}$. Interactions similar to those forming the 'floors' of the S1 specificity pockets of M^{pro} are not observed in the unbound $M^{pro}{}_{+A(-1)}$.

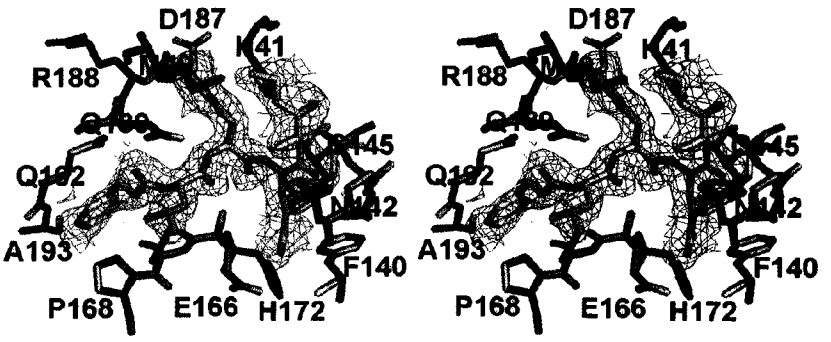
The rest of the substrate-binding region in the molecular structure of the unbound $M^{pro}{}_{+A(-1)}$ reported here is in the catalytically competent conformation, in good agreement with those in the previously reported molecular structures of the $M^{pro}{}_{+A(-1)}$:APE complex and of M^{pro}, except for some side chains whose conformational rearrangements are necessary in order to accommodate APE or the other inhibitors.

Binding of APE to $M^{pro}{}_{+A(-1)}$

Both the $P2_1$ crystals of M^{pro} and the $P4_32_12$ crystals of $M^{pro}{}_{+A(-1)}$ were soaked in solutions of the APE synthesized for this study in three stereochemical versions: the two diastereomers Cbz-Leu-Phe-AGln-(2*S*,3*S*)EP-COOEt and Cbz-Leu-Phe-AGln-(2*R*,3*R*)EP-COOEt, and the racemic mixture of these two diastereomers.

Figure 3-5 Interactions of APE (orange) with $M^{pro}_{+A(-1)}$ (green). (a) Outstanding electron densities in the $|F_o|-|F_c|,\alpha_c$ map of the $M^{pro}_{+A(-1)}$:APE complex. (b) View from the 'floors' of the S1 specificity pockets of $M^{pro}_{+A(-1)}$. (c) Schematic diagram for the peptidase-inhibitor interactions. Hydrogen bonds are shown as dashed lines, with their distances (in Å) given alongside. Residues of the peptidase in contact with APE are shown as arcs.

(a)



(b)

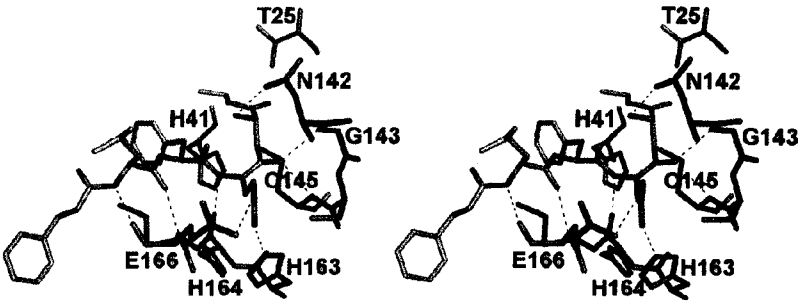
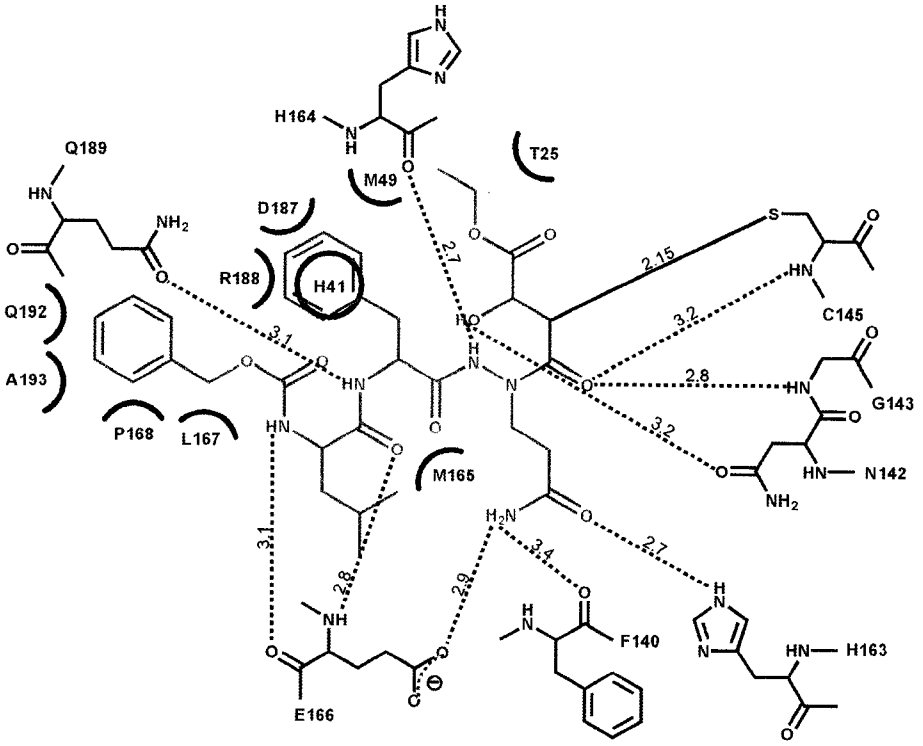


Figure 3-5 (continued)

(c)



Outstanding electron densities for APE were observed only in the electron density maps of $M^{pro}_{+A(-1)}$. They could be fitted by the 2*S*,3*S* diastereomer only (Figure 3-5a). The latter is consistent with the results of the studies on the inhibition of M^{pro} by APE in these versions. These results can be explained with the models for all of the four possible diastereomers of APE binding to M^{pro} (Chapter 2). Superpositions of the molecular structures of the unbound and the APE-bound $M^{pro}_{+A(-1)}$ reported here show that the binding of APE does not grossly affect the protein fold nor the protomer orientation of the $M^{pro}_{+A(-1)}$ dimer (Table 3-2). The same observations are found in the superpositions of the previously reported molecular structures of the unbound and the APE-bound M^{pro} in space group C2 (Chapter 2). The molecular structure of the $M^{pro}_{+A(-1)}$:APE complex in space group $P4_32_12$ also agrees well in most atomic positions with that in space group $P2_12_12_1$.

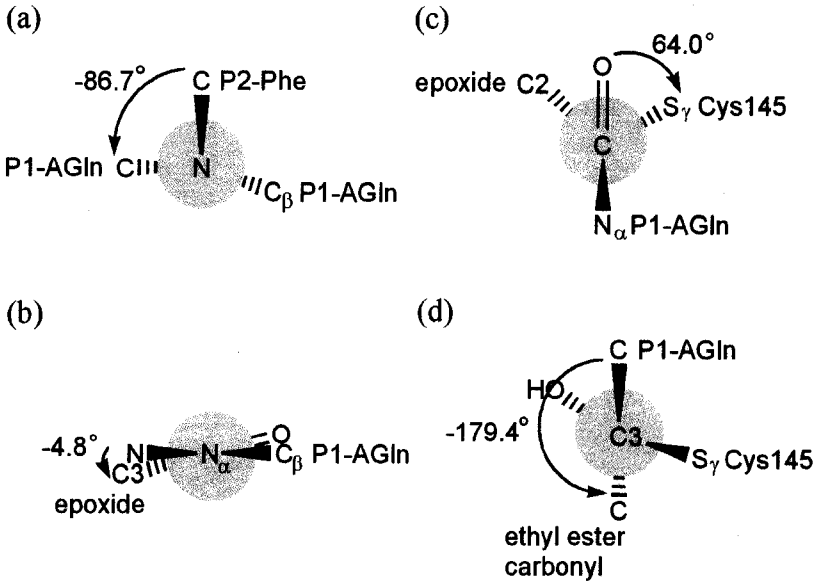
The structural consequences of the binding of APE to $M^{pro}_{+A(-1)}$ are very similar to those observed in the molecular structures of the unbound and the APE-bound M^{pro} in space group C2, and the APE-bound $M^{pro}_{+A(-1)}$ in space group $P2_12_12_1$ (Chapter 2). A covalent bond forms between Cys145 S^{γ} of $M^{pro}_{+A(-1)}$ and the epoxide atom C3 of APE (2.15 Å; Figures 3-5b and 3-5c). In the refinement of all of the molecular structures of the APE-bound M^{pro} and $M^{pro}_{+A(-1)}$, a restraint was applied to this C-S bond on the basis that the length of a C-S single bond is normally about 1.8 Å. Interestingly, the opened epoxide moiety of APE always tended to be tilted away from Cys145 of M^{pro} or $M^{pro}_{+A(-1)}$, thereby lengthening this C-S bond by 0.2 to 0.3 Å, even though the opened epoxide moiety of APE was manually moved back towards Cys145 of M^{pro} or $M^{pro}_{+A(-1)}$ regularly during the refinement process. Although such

lengthening is not considered significant as the overall positional uncertainties (based on maximum likelihood) of all of these molecular structures are in the range of 0.2 to 0.3 Å, it does suggest the possibility of this C-S bond being under strain and vulnerable to dissociation caused by a second nucleophilic attack at the epoxide atom C3 of APE (say, by an activated water molecule), leading to the speculation that APE could act as a reversible inhibitor of M^{pro}.

The conformation of the opened epoxide moiety and the main-chain conformation of P1-AGln of APE in the molecular structure of the $M^{pro}_{+A(-1)}$:APE complex reported here are essentially the same as those in the previously reported molecular structures of the M^{pro}:APE and the $M^{pro}_{+A(-1)}$:APE complexes (Figures 3-6a to 3-6d). Unlike the C^{α} of residue P1 of other inhibitors, the N^{α} of P1-AGln of APE is sp^2-hybridized and has a trigonal planar geometry. In order to be accommodated by the S1 specificity pocket of M^{pro} and $M^{pro}_{+A(-1)}$, the side chain of P1-AGln has to adopt a different conformation, in particular the equivalent to χ_1 ($N\text{-}N^{\alpha}\text{-}C^{\beta}\text{-}C^{\gamma}$). This angle is -123.6° in the molecular structure of the $M^{pro}_{+A(-1)}$:APE complex reported here. By contrast, in the molecular structures of the complexes of an M^{pro} variant with a series of peptidomimetic inhibitors, this angle is in the range of -65° to -80°.[17]

Interestingly, the binding of APE induces the recovery of the catalytically competent conformation of the oxyanion holes and the S1 specificity pockets of $M^{pro}_{+A(-1)}$ (Table 3-3; Figures 3-1d and 3-2d). The main-chain carbonyl O of P1-AGln of APE is accommodated in the oxyanion hole of $M^{pro}_{+A(-1)}$, forming hydrogen bonds with the main-chain NH groups of Gly143 (2.8 Å) and Cys145 (3.2 Å) (Figures 3-5b and 3-5c). The angles ϕ and ψ of residues Lys137 to Ser144 in the molecular

Figure 3-6 Newman projections of APE. (a) The equivalent to ϕ of P1-AGln, C(P2-Phe)-N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln). (b) The equivalent to ψ of P1-AGln, N(P1-AGln)-N^{α}(P1-AGln)-C(P1-AGln)-C3(epoxide). (c) The torsion angle O=C(P1-AGln)-C3(epoxide)-S^{γ}(Cys145). (d) The torsion angle C(P1-AGln)-C3(epoxide)-C2(epoxide)-C(ethyl ester carbonyl).



structure of the $M^{pro}_{+A(-1)}$:APE complex reported here are essentially the same as those in the previously reported molecular structure of the same complex (Figure 3-3d). In contrast to those in the molecular structure of the unbound $M^{pro}_{+A(-1)}$, the B factors of these residues in the molecular structures of the $M^{pro}_{+A(-1)}$:APE complex are close to the averages, indicating that the mobility of these residues is reduced upon the binding of APE (Figure 3-4d).

The hydrogen bond between His163 $N^{\delta 1}$ and the phenolic OH of Tyr161 is preserved (3.3 Å), and the π-stacking of the imidazole ring of His163 with the phenyl ring of Phe140 is recovered (distances between the geometric centers of the aromatic rings: 3.7 Å) (Figure 3-2d). His163 $N^{\varepsilon 2}$ no longer interacts with Glu166 $O^{\varepsilon 2}$, but forms a hydrogen bond (2.7 Å) with the side-chain carbonyl O of P1-AGln of APE instead (Figures 3-5b and 3-5c). The side chains of Glu166 and His172 interact with each other (3.0 Å; Figure 3-2d). The additional Ala at the N-terminus and Ser1 of the protomer could not be identified in the electron density maps of the $M^{pro}_{+A(-1)}$:APE complex. Interestingly, the side-chain amide NH_2 group of P1-AGln of APE forms hydrogen bonds, though not in ideal geometry, with the carbonyl O of Phe140 (3.4 Å) and atoms $O^{\varepsilon 1}$ and $O^{\varepsilon 2}$ of Glu166 (3.5 Å and 2.9 Å, respectively) (Figure 3-5b). Therefore, without the participation of Ser1 of the opposite protomer, Phe140 and Glu166 are tied together to form parts of the `floors' of the S1 specificity pockets of $M^{pro}_{+A(-1)}$.

Similar to APE in protomer A of the previously reported molecular structure of the $M^{pro}_{+A(-1)}$:APE complex, the benzyloxycarbonyl (Cbz) group of APE in the molecular structure of the same complex reported here squeezes into and slightly

widens the S4 subsites of $M^{pro}_{+A(-1)}$, as a result making contacts with Leu167, Pro168, Gln192 and Ala193 of $M^{pro}_{+A(-1)}$. Otherwise, the interactions of APE with the rest of the substrate-binding region observed in the molecular structure of the $M^{pro}_{+A(-1)}$:APE complex reported here are essentially the same as those observed in the previously reported molecular structures of the APE-bound $M^{pro}_{+A(-1)}$ and M^{pro} (Figures 3-5b and 3-5c).

Dynamic equilibrium for the conformation of the active sites and the S1 specificity pockets of M^{pro}

In contrast to the previously reported molecular structure of M^{pro} in space group $P2_1$ at pH 6.0,[11] the molecular structure of M^{pro} reported here (in the same space group and at the same pH) shows that the active sites and the S1 specificity pockets of both protomers are in the catalytically competent conformation (Table 3-3). In the determinations of both structures, the complete wild-type sequence (residues 1 to 306) of M^{pro} was overexpressed, purified and crystallized. Although different strategies might have been used in the X-ray diffraction data collection and processing, and in the structure solution and refinement, this could not result in the dramatic structural differences observed. The structural differences probably arise from the differences in the conditions for the preparation and the crystallization of M^{pro}. Similarly, the molecular structure of M^{pro} in space group $P2_12_12$ at pH 6.5 shows that both protomers are in the catalytically competent conformation,[18] whereas that in the same space group $P2_12_12$ and at a slightly higher pH (6.6) shows that both protomers of the latter are in the catalytically incompetent conformation.[15] The conformations of the collapsed active sites and S1 specificity pockets observed in the

catalytically incompetent protomers show some variability. This variability indicates that the active site and the S1 specificity pocket of each protomer of M^{pro} do not adopt a single conformation in solution at pH 5.9 to 6.6, but instead there is an ensemble of conformations. A particular conformation might be favored by a particular set of crystallization conditions.

However, the various conformations of the collapsed active sites and S1 specificity pockets share a common feature: the imidazole ring of His163 is not π-stacked with the phenyl ring of Phe140. π-stacking is an example of aromatic interactions.[19] In the molecular structures of the unbound M^{pro}, an aromatic interaction is observed between these two rings in an offset-stacked (i.e. π-stacking) or an edge-to-face fashion wherever the protomer is in the catalytically competent conformation (Figure 3-7). In both orientations, one or two hydrogen atoms (with partial positive charge) on the phenyl ring of Phe140 are positioned near the central region (with partial negative charge) of the imidazole ring of His163. According to the results of previous studies,[20,21] a single aromatic interaction as such is weak (the interaction energy may be in the range of 1 to 2 kcal/mol only), in contrast to the clusters of aromatic interactions commonly involved in the stabilization of protein structures.[22] Therefore, this interaction is susceptible to disruptions that could be caused by changes in a number of factors. The formation of this interaction can be viewed as a reversible process in dynamic equilibrium:

Aromatic interaction ↔ No aromatic interaction (Equation 3-1)

The position of the equilibrium could be determined in part by pH. At low pH (near or below the pK_{a1} of His163; Figure 3-8), His163 $N^{\delta 1}$ in a significant number of the

Figure 3-7 Aromatic interactions observed between Phe140 and His163 in the molecular structures of M^{pro} (wild type and variants). The dashed lines indicate the alignment of the partial positive charges (δ+) of the hydrogen atoms on the phenyl ring of Phe140 with the partial negative charges (δ-) in the central part of the imidazole ring of His163.

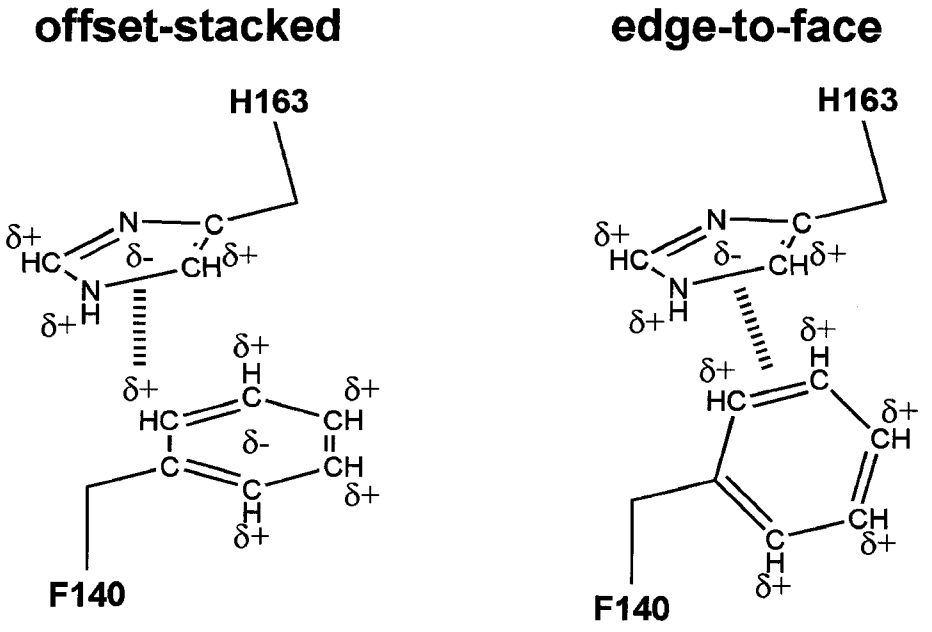
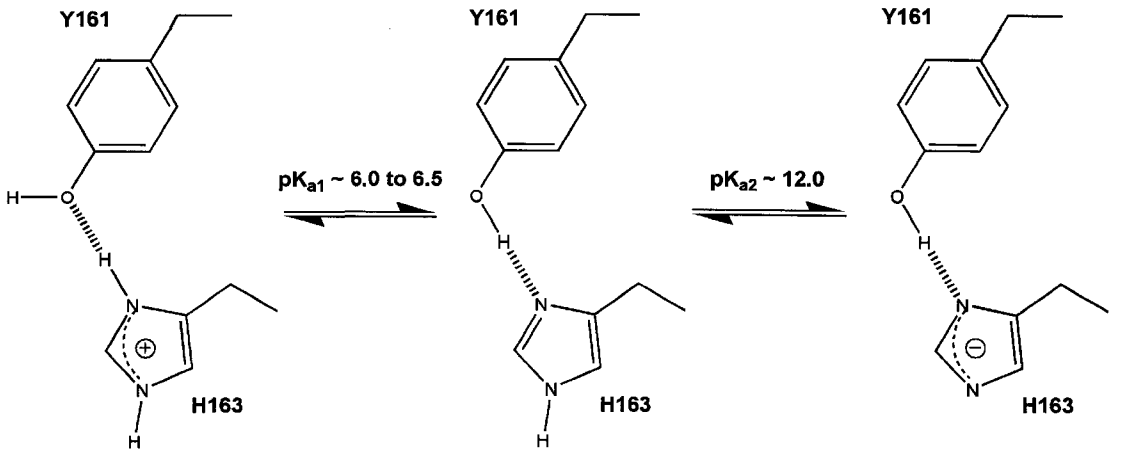


Figure 3-8 Protonation/deprotonation states of His163 hydrogen-bonded to Tyr161 in M^{pro} (wild type and variants). K_{a1} is the dissociation constant for the protonation of His163 $N^{\delta 1}$. K_{a2} is the dissociation constant for the deprotonation of His163 $N^{\varepsilon 2}$.



M^{pro} protomers is protonated, thereby introducing a positive charge on the imidazole ring of His163. This would disfavor its aromatic interaction with the phenyl ring of Phe140, and the position of the equilibrium would shift to the right. This is consistent with the trend exhibited by the molecular structures of the unbound M^{pro} at various pH values (Table 3-3). In the pH range of 7.0 to 9.0, most of the M^{pro} protomers have the aromatic interaction and are in the catalytically competent conformation (the left side of the equilibrium predominates); whereas in the pH range of 5.9 to 6.6, some of the M^{pro} protomers lose the aromatic interaction and are in the catalytically incompetent conformation (the position of the equilibrium shifts to the right).

A second factor in determining the position of the equilibrium could be the integrity of the interactions among Phe140 and Glu166 of each protomer, and Ser1 of the opposite protomer. In the molecular structure of $M^{pro}_{+A(-1)}$ reported here, the additional Ala at the N-terminus of the opposite protomer blocks Ser1 and disrupts its normal interactions with Phe140 and Glu166 of the parent protomer. This probably weakens the `conformational anchor' of Phe140, as indicated by its above-average B factors (Figure 3-4c), making the aromatic interaction of its phenyl ring with the imidazole ring of His163 vulnerable to disruption. In all of the molecular structures of the unbound M^{pro} (both the wild type and the variants) determined so far, wherever Phe140 and Glu166 of the parent protomer, and Ser1 of the opposite protomer do not interact normally, the B factors of residues Lys137 to Ser144 of the parent protomer are above the average, even if the parent protomer is in the catalytically competent conformation (e.g. the molecular structure of the unbound M^{pro} at pH 6.5 in space group $P2_12_12$[18]). In contrast, wherever the three residues interact normally, the B

factors of residues Lys137 to Ser144 of the parent protomer are close to the average (e.g. protomer A of the molecular structure of the unbound M^{pro} reported here; Figure 3-4a). These observations suggest that the interactions among the three residues can immobilize residues 137 to 144. Apparently, both the ionic interaction between the α-amino group of Ser1 and the side-chain carboxylate group of Glu166 from opposite protomers, and the amide hydrogen-carbonyl oxygen hydrogen bonds between Ser1 and Phe140 from opposite protomers contribute to the immobilization of residues Lys137 to Ser144. The molecular structure of the unbound M^{pro} at pH 7.0 in space group $P2_12_12$,[23] and protomer B of the molecular structure of the unbound M^{pro} reported here (Figure 3-4b) suggest that the absence of either interaction slightly compromises the immobilizing effect.

The increased vulnerability of the aromatic interaction to disruption would shift the position of the equilibrium to the right. This probably explains why the active sites and the S1 specificity pockets of both protomers are collapsed in the molecular structure of the unbound $M^{pro}_{+A(-1)}$ reported here. The M^{pro} variant with its three N-terminal residues truncated, $M^{pro}_{\Delta(1-3)}$, exists mainly as a dimer in solution but its catalytic activity is 24 % lower than that of M^{pro}.[24] Without the interactions among Phe140 and Glu166 of the parent protomer, and Ser1 of the opposite protomer, the position of the equilibrium would shift to the right, resulting in the decreased availability of the active sites and the S1 specificity pockets in the catalytically competent conformation. The binding of a substrate to $M^{pro}_{\Delta(1-3)}$ may induce the recovery of the catalytically competent conformation of the active sites and the S1

specificity pockets (to be discussed below), at an energetic cost accounted for by the reduced catalytic activity of $M^{pro}_{\Delta(1-3)}$.

Induced-fit binding of APE to M^{pro}

The molecular structures of the unbound and the APE-bound $M^{pro}_{+A(-1)}$ reported here show that the binding of APE to $M^{pro}_{+A(-1)}$ follows an induced-fit model, not the lock-and-key model. More particularly, the binding of APE induces the recovery of the catalytically competent conformation of the oxyanion holes and the S1 specificity pockets of both protomers of $M^{pro}_{+A(-1)}$. This induced fit is likely relevant to the binding of APE to M^{pro}, because of the conformational flexibility of the active sites and the S1 specificity pockets of M^{pro} discussed above. In all of the molecular structures of the inhibitor-bound M^{pro} (both the wild type and the variants) determined so far (Chapter 2),[11,17,23,25] except for protomer B in the complex of M^{pro} with a chloromethyl ketone (CMK),[11] the active sites and the S1 specificity pockets are in the catalytically competent conformation (Table 3-3). The induced-fit binding of APE to $M^{pro}_{+A(-1)}$ is probably driven by the formation of the covalent bond between Cys145 S^{γ} of $M^{pro}_{+A(-1)}$ and the epoxide atom C3 of APE. It would be energetically costly to break this C-S bond in order to repel APE and preserve the conformational flexibility of the active sites and the S1 specificity pockets of $M^{pro}_{+A(-1)}$. Therefore, the active sites and the S1 specificity pockets of $M^{pro}_{+A(-1)}$ have to adopt the catalytically competent conformation in order to accommodate a substrate mimic like APE. This explanation probably applies to the actual catalytic mechanism of M^{pro} as well: the acylation stage (i.e. formation of the covalent acyl-enzyme intermediate) may drive the induced-fit binding of a substrate to M^{pro}. After the deacylation stage (i.e.

breakdown of the covalent acyl-enzyme intermediate), the removal of the product may be favored by the recovery of the conformational flexibility of the active sites and the S1 specificity pockets of M^{pro}.

As the side chain of P1-AGln of APE occupies the S1 specificity pockets of $M^{pro}_{+A(-1)}$, the side-chain amide NH_2 group of P1-AGln of APE donates hydrogen bonds to the carbonyl O of Phe140 and to atoms $O^{\varepsilon 1}$ and $O^{\varepsilon 2}$ of Glu166 of $M^{pro}_{+A(-1)}$, in addition to the hydrogen bond between the side-chain carbonyl O of P1-AGln of APE and His163 $N^{\varepsilon 2}$ of $M^{pro}_{+A(-1)}$. Similarly, additional hydrogen bonds are observed in the molecular structures of the $M^{pro}_{+A(-1)}$:APE complex in space group $P2_12_12_1$, the M^{pro}:APE complex (Chapter 2), protomer A of the M^{pro}:CMK complex,[11] and the complex of M^{pro} with a peptidomimetic inhibitor.[25] In the molecular structures of the complexes of an M^{pro} variant with a series of peptidomimetic inhibitors, these hydrogen bonds seem to be water-mediated.[17] Consequently, even without the participation of Ser1 of the opposite protomer of $M^{pro}_{+A(-1)}$, Phe140 and Glu166 of the parent protomer are tied together by these hydrogen bonds to form parts of the `floors' of the S1 specificity pockets. These hydrogen bonds also conformationally anchor Phe140, as indicated by its close-to-average B factors (Figure 3-4d), and make the π-stacking of its phenyl ring with the imidazole ring of His163 resistant to disruption. As a result, the catalytically competent conformation of the active sites and the S1 specificity pockets of $M^{pro}_{+A(-1)}$ is rigidified. The position of the equilibrium described by Equation 3-1 would shift to the left. This probably serves as an additional explanation for the predominant S1 specificity of M^{pro} or $M^{pro}_{+A(-1)}$ for Gln, but not for Glu. Comparison of the molecular structure of the unbound M^{pro} with that of the

APE-bound M^{pro}, both in space group C2 at pH 6.5 (Chapter 2), does not show an obvious reduction in the B factors of residues Lys137 to Ser144 upon the binding of APE to M^{pro}, probably because Phe140 and Glu166 of the parent protomer, and Ser1 of the opposite protomer in the unbound M^{pro} interact normally and immobilize these residues.

In protomer B of the M^{pro}:CMK complex, the binding of CMK does not induce the recovery of the catalytically competent conformation of the active site and the S1 specificity pocket of M^{pro} (Table 3-3). The side chain of P1-Gln of CMK cannot be accommodated by the collapsed S1 specificity pocket so it protrudes into the solvent.[11] This could be attributed to the covalent bond between Cys145 S^{γ} of M^{pro} and the methylene C of CMK, whose rotation can orient the side chain of P1-Gln of CMK towards the solvent, thereby avoiding the steric hindrance due to the collapsed S1 specificity pocket. In protomer A of the M^{pro}:CMK complex, although the active site and the S1 specificity pocket of M^{pro} are in the catalytically competent conformation (Table 3-3), allowing the side chain of P1-Gln of CMK to occupy the S1 specificity pocket of M^{pro}, the unoccupied space in the rest of substrate-binding region of M^{pro} is large enough for the conformational rearrangements in the rest of CMK, resulting in an unexpected binding mode of CMK to protomer A of M^{pro} as well.[11] These unexpected binding modes are not observed in the molecular structures of the complexes of M^{pro} (both the wild type and the variants) with APE (Chapter 2), nor with the peptidomimetic inhibitors,[17,25] probably because their P' ester groups sterically hinder the rotation of the C-S covalent bond formed between the peptidases and the inhibitors. In the case of APE, the π-conjugation of the atom N^{α} and the main-

chain carbonyl group of P1-AGln restricts the orientation of the side chain of P1-AGln to point towards the S1 specificity pockets of M^{pro} or $M^{pro}_{+A(-1)}$. In the case of the peptidomimetic inhibitors, the carbon attacked by Cys145 S^{γ} corresponds to the carbonyl C of the scissile peptide bond of a substrate. Therefore, the resulting C-S bond pulls the inhibitor towards the `back wall' of the substrate-binding region. This reinforces the steric hindrance due to the P' ester group of the inhibitor against the rotation of the C-S bond.

Thus, a refined explanation could be provided for the pH dependence of the catalytic activity of M^{pro}.[11,15,26,27] At low pH, a significant number of the M^{pro} protomers are in the catalytically incompetent conformation. However, these protomers are conformationally flexible, and the recovery of their catalytically competent conformation can be induced by the binding of a substrate. The energetic cost associated with this induced-fit binding compromises the catalytic activity of M^{pro}. Nonetheless, this energetic cost is not necessarily the only cause of the apparent decrease in the catalytic activity of M^{pro}. The possibility of a second mechanism, such as the change in the protonation/deprotonation state of the catalytic dyad with pH, being involved in the pH dependence of the catalytic activity of M^{pro} should not be ruled out.

3.4 References

1. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W., Jr. & Sweet, R. M., eds.), Vol. 276: Macromolecular Crystallography, Part A, pp. 307-326 Academic Press, New York, NY.
2. Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.
3. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Cryst.* **D59**, 1131-1137.
4. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.
5. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240-255.
6. McRee, D. E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.
7. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-91.
8. Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593-604.

9. Cohen, G. H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Cryst.* **30**, 1160-1161.

10. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.

11. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13190-13195.

12. Ziebuhr, J., Heusipp, G. & Siddell, S. G. (1997). Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J. Virol.* **71**, 3992-3997.

13. Hegyi, A., Friebe, A., Gorbalenya, A. E. & Ziebuhr, J. (2002). Mutational analysis of the active centre of coronavirus 3C-like proteases. *J. Gen. Virol.* **83**, 581-593.

14. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763-1767.

15. Tan, J., Verschueren, K. H., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., Heisen, B., Mesters, J. R., Chen, K., Shen, X., Jiang, H. & Hilgenfeld, R. (2005). pH-dependent conformational flexibility of the SARS-

CoV main proteinase (M(pro)) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *J. Mol. Biol.* **354**, 25-40.

16. Lee, T. W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. G. (2005). Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J. Mol. Biol.* **353**, 1137-1151.

17. Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Yuen, K. Y., Wong, L., Gao, G., Chen, S., Chen, Z., Ma, D., Bartlam, M. & Rao, Z. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **3**, e324.

18. Xu, T., Ooi, A., Lee, H. C., Wilmouth, R., Liu, D. X. & Lescar, J. (2005). Structure of the SARS coronavirus main proteinase as an active C2 crystallographic dimer. *Acta Cryst.* **F61**, 964-966.

19. Hunter, C. A., Lawson, K. R., Perkins, J. & Urch, C. J. (2001). Aromatic interactions. *J. Chem. Soc., Perkin Trans. 2*, 651-669.

20. Burley, S. K. & Petsko, G. A. (1986). Dimerization energetics of benzene and aromatic amino acid side chains. *J. Am. Chem. Soc.* **108**, 7995-8001.

21. Serrano, L., Bycroft, M. & Fersht, A. R. (1991). Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. *J. Mol. Biol.* **218**, 465-475.

22. Burley, S. K. & Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**, 23-28.

23. Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P., Shr, H. L., Chang, G. G., Wang, A. H. & Liang, P. H. (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **280**, 31257-31266.

24. Hsu, W. C., Chang, H. C., Chou, C. Y., Tsai, P. J., Lin, P. I. & Chang, G. G. (2005). Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *J. Biol. Chem.* **280**, 22741-22748.

25. Ghosh, A. K., Xi, K., Ratia, K., Santarsiero, B. D., Fu, W., Harcourt, B. H., Rota, P. A., Baker, S. C., Johnson, M. E. & Mesecar, A. D. (2005). Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. *J. Med. Chem.* **48**, 6767-6771.

26. Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. (2004). 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* **43**, 4568-4574.

27. Chou, C. Y., Chang, H. C., Hsu, W. C., Lin, T. Z., Lin, C. H. & Chang, G. G. (2004). Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry* **43**, 14958-14970.

28. Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.

Part Two Streptogrisin B

A version of Chapter 4 has been accepted for publication:

Lee, T.-W. & James, M. N. G. *Biochim. Biophys. Acta* In press.

A version of Chapter 5 has been published:

Lee, T.-W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2007). *J. Mol. Biol.* **367**, 527-546.

Chapter 4

1.2 Å-resolution crystal structures reveal the second tetrahedral intermediates of streptogrisin B (SGPB)

4.1 Overview

This chapter reports the molecular structures of SGPB at pH 4.2 in space group $P2_12_12$, and at pH 7.3 in space group I23, both at the highest resolutions its crystals can afford (1.2 Å). Unexpectedly, additional electron density peaks were observed in the active site and the substrate-binding region of SGPB in the computed maps at both pH values. The approach taken to the assignment of these peaks is explained in detail. Structure refinement with relaxed geometric restraints resulted in structures representing mixtures of the second tetrahedral intermediates and the enzyme-product complexes of SGPB existing in a pH-dependent equilibrium. A detailed geometric analysis of the active-site residues of SGPB, the bound species and their interactions is presented. The molecular structures of SGPB in different conditions are also compared, in order to obtain insights into the movement of atoms required for the progression of the catalytic mechanism of serine peptidases.

4.2 Materials and methods

Preparation and crystallization of SGPB

SGPB, purified from Pronase as described previously,[1] was a generous gift from L. B. Smillie in the Department of Biochemistry, University of Alberta. Crystallization of SGPB in both space groups $P2_12_12$ and I23 was done as described previously.[2] An approximately 10 mg/mL solution of SGPB was prepared by

dissolving SGPB in a 10 mM solution of magnesium chloride. Diffraction-quality crystals in both space groups were grown overnight by the hanging-drop vapor diffusion method at ambient temperature. Each drop contained equal amounts of the SGPB solution and the reservoir solution. The reservoir solution for the block-shaped $P2_12_12$ crystals included 0.7 M potassium dihydrogen phosphate (pH 4.2), whereas that for the cubic-shaped I23 crystals included 1.0 M lithium sulfate and 0.1 M HEPES-NaOH (pH 7.3). For cryo-protection, all crystals were soaked for approximately 10 seconds in their respective reservoir solutions with 40 % (v/v) glycerol included, and then immediately flash-cooled in liquid nitrogen for storage and subsequent shipment to the synchrotron beamline.

Data collection and processing, and structure solution and refinement

The X-ray diffraction data from all crystals were collected at the synchrotron Beamline 8.3.1 (equipped with an ADSC-Q210 CCD detector) at the Advanced Light Source in the Lawrence Berkeley National Laboratory. The data sets from the $P2_12_12$ crystals were indexed, scaled and merged using HKL-2000,[3] whereas those from the I23 crystals were processed using DENZO and SCALEPACK.[3] Structure solution and refinement were done in CCP4.[4,5] Structures in both space groups were solved by the molecular replacement method, with the search model provided by the molecular structure of the SGPB:OMTKY3 complex (PDB accession code: 1SGR).[6] Both structures were solved using MOLREP,[7] iteratively refined using REFMAC,[8] and manually adjusted when needed using XtalView/Xfit.[9] At the end, both structures were refined anisotropically with riding hydrogen atoms included. The data-to-restraint weighting term used in the final refinement for the structure of SGPB in

space group $P2_12_12$ is 10.0, whereas that for the structure of SGPB in space group I23 is 30.0. The stereochemical qualities of the final structures were assessed using PROCHECK.[10]

Structure analysis

Graphical representations of the molecular structures were prepared using PyMOL (http://www.pymol.org/). Geometric calculations were done using GEOMCALC.[11] Superpositions of molecular structures were done using ALIGN,[12,13] based on the main-chain atoms (amide N, C^{α}, and carbonyl C and O). The diffraction-component precision index (DPI) was used as a measure of the overall positional uncertainty (σ_r) of a molecular structure.[14] The anisotropic thermal motions of atoms were analyzed using XtalView/Xfit.[9]

Protein Data Bank accession codes

The atomic coordinates and the structure factors of both structures have been deposited in the Protein Data Bank. The accession code is 2QA9 for the structure of SGPB in space group $P2_12_12$, and 2QAA for the structure of SGPB in space group I23.

4.3 Results and discussion

Structure determination

The parameters and statistics derived from X-ray diffraction data processing and structure refinement are summarized in Table 4-1. As reported previously, SGPB crystallized in the orthorhombic space group $P2_12_12$ at pH 4.2,[2,15,16] and in the cubic space group I23 at pH 7.3.[2] In space group $P2_12_12$, each asymmetric unit contains one SGPB molecule (molecule E). All residues of SGPB (residues 16E to 242E) were

Table 4-1 Parameters and statistics derived from X-ray diffraction data processing and structure refinement

	SGPB	
pH	4.2	7.3
Data processing		
Wavelength (Å)	0.954	1.127
Resolution limit[a] (Å)	19.42 – 1.18	38.01 – 1.23
	(1.22 – 1.18)	(1.27 - 1.23)
Space group	$P2_12_12$	I23
Unit-cell constants		
a (Å)	43.46	142.29
b (Å)	107.47	142.29
c (Å)	36.79	142.29
α(°)	90.00	90.00
β(°)	90.00	90.00
γ(°)	90.00	90.00
Mosaicity (°)	0.26	0.22
Number of unique reflections	56,646 (5,212)	136,369 (12,302)
Redundancy	3.6 (2.2)	8.5 (3.8)
Completeness (%)	98.3 (91.7)	99.0 (90.5)
R_{sym}[b] (%)	7.1 (29.1)	4.8 (40.4)
$< I / \sigma(I) >$	12.4 (3.1)	37.3 (3.0)

[a] Numbers in the parentheses refer to the highest resolution bins.

[b] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_{hkl,i} - <I_{hkl}>| / \Sigma_{hkl}\Sigma_i I_{hkl,i}$, where $I_{hkl,i}$ and $<I_{hkl}>$ are the i-th observed intensity and the average intensity of reflection hkl, respectively.

Table 4-1 (continued)

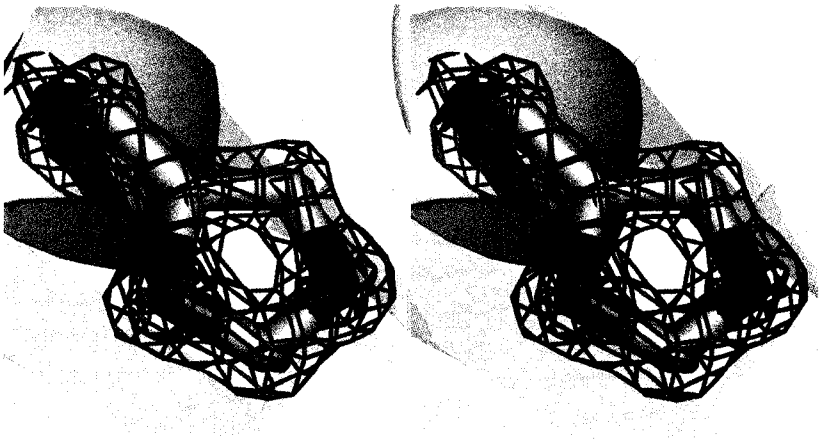
	SGPB	
pH	4.2	7.3
Structure refinement		
Resolution range (Å)	19.42 – 1.18	38.01 – 1.23
R_{work}[c] (%)	14.5	11.3
R_{free}[c] (%)	18.2	12.7
R^c *(%)*	14.7	11.3
Number of non-hydrogen atoms per asymmetric unit (average *B* factors, Å^2)		
Protein	Chain E: 1,350 (8.71)	Chain A: 1,353 (10.66)
	Chain I: 38 (4.55)	*Tyr and Leu: 18 (21.09)*
		Chain B: 1,342 (11.86)
		Tyr and Leu: 18 (18.47)
Solvent	332 (23.46)	658 (30.95)
rms deviation from ideal geometry		
Bond lengths (Å)	0.029	0.029
Bond angles (°)	2.467	2.414
Ramachandran plot		
Favored (%)	91.2	88.7
Allowed (%)	8.1	10.6
Generously allowed (%)	0.0	0.0
Disallowed (%)	0.7	0.7

[c] $R_{work} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and the calculated structure factor amplitudes of a particular reflection, respectively. The summation is over 95 % of the reflections in the specified resolution range. The remaining 5 % of the reflections were randomly selected before the structure refinement and not included in the structure refinement. R_{free} was calculated over these reflections using the same equation as for R_{work}.[52] R was calculated over all of the reflections in the specified resolution range.

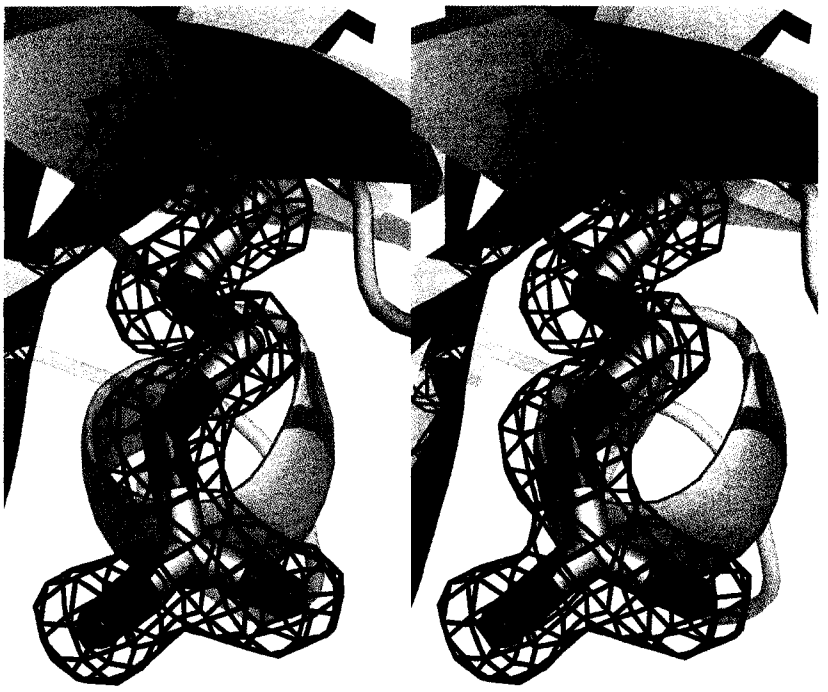
clearly defined in the electron density maps. Most atoms, including those of the active-site residues, could be located precisely in the $2|F_o|-|F_c|,\alpha_c$ map contoured at the 4.0σ to 5.0σ level (Figure 4-1a). In most cases, carbons, nitrogens and oxygens could be distinguished, thus the side-chain orientations of asparagines, glutamines and histidines could be determined unambiguously. Atoms of relatively solvent-accessible residues were located reliably in the same map contoured at the 1.0σ to 2.0σ level (Figure 4-1b). In order to eliminate the negative densities of the disulfide bond Cys42E-Cys58E in the $|F_o|-|F_c|,\alpha_c$ map (which suggests a damage of this disulfide bond probably by the high dose of X-rays used for collecting the high-resolution diffraction data), the occupancies of the S^{γ} of these two cysteines were decreased. Two conformers were identified for Met180E (occupancies: 0.75 for conformer A, 0.25 for conformer B). One-third of conformer A is regarded as methionine sulfoxide (Sme), whose atom O^{ε} satisfies a positive density centered less than 2 Å away from its atom S^{δ} in the $|F_o|-|F_c|,\alpha_c$ map contoured at the 3.0σ to 4.0σ level. The Ramachandran plot shows that 91.2 % of non-Gly and non-Pro residues are in the most favored regions, and 8.1 % are in the additionally allowed regions. The remaining 0.7 % is due to one residue, Asn100E, which is in the disallowed region (Table 4-1). Asn100E has well defined electron densities. Its main-chain conformation seems to be restricted sterically by its neighboring residues on the same loop and by residues Val176E to Tyr178E. In the crystal, the side chain of Asn100E from the asymmetric unit x, y, z makes contact with the side chain of Arg107E from an adjacent asymmetric unit at 0.5+x, 0.5-y, -z.

Figure 4-1 $2|F_o|-|F_c|,\alpha_c$ map of SGPB at pH 4.2. (a) The electron densities of the side chain of His57E, contoured at the 2.0σ (blue) and the 5.0σ (red) levels. It is clearly possible to distinguish carbons and nitrogens in this region. (b) The electron densities of the side chain of Arg182E, contoured at the 2.0σ level (blue).

(a)



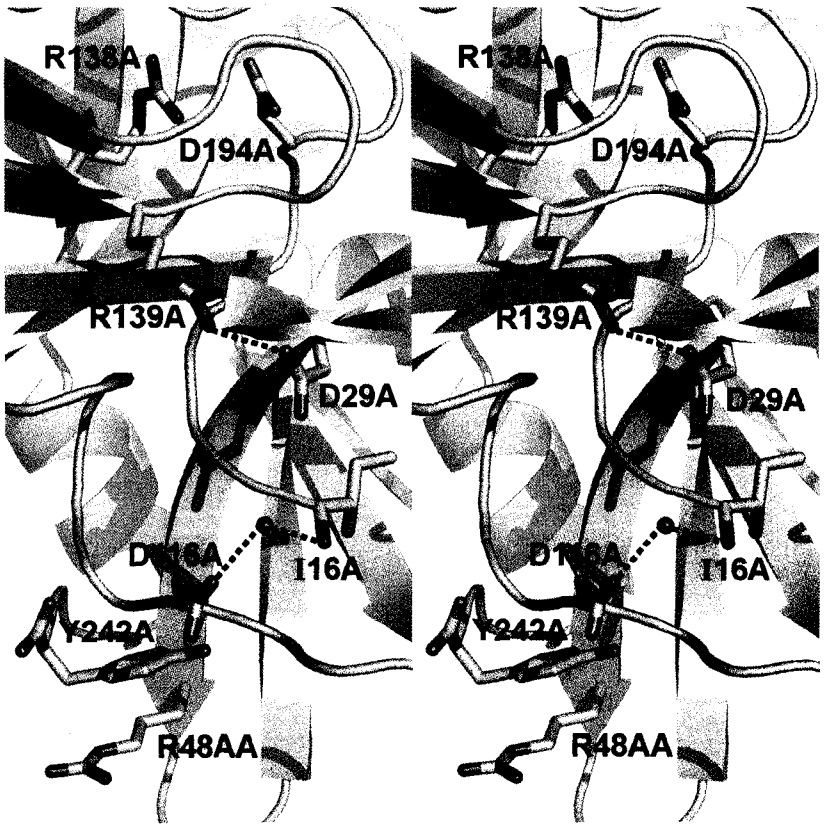
(b)



In space group I23, each asymmetric unit consists of two SGPB molecules (molecules A and B). All residues of SGPB (residues 16A to 242A, and residues 16B to 242B) were clearly defined in the electron density maps. The electron density maps in space group I23 are of equivalent quality to those in space group $P2_12_12$ (Figures 4-1a and 4-1b). The occupancies of the S^{γ} of Cys42, Cys58, Cys191 and Cys220 in both molecules A and B of SGPB were decreased to eliminate their negative densities in the $|F_o|$-$|F_c|$,α_c map. Two conformers could be identified for Cys42A, Cys220A and Cys42B. In each of them, only conformer A is oriented for disulfide bonding. As shown by the Ramachandran plot for molecules A and B of SGPB, 88.7 % of non-Gly and non-Pro residues are in the most favored regions, and 10.6 % are in the additionally allowed regions. As in space group $P2_12_12$, the remaining 0.7 % is due to Asn100A and Asn100B, which are in the disallowed region (Table 4-1). The electron densities of both Asn100A and Asn100B are well defined. Their steric environment is similar to that of Asn100E in space group $P2_12_12$. In the crystal, the side-chain amide NH_2 group of Asn100B from the asymmetric unit x, y, z forms hydrogen bonds with Ser201A O^{γ} from an adjacent asymmetric unit at -x, -y, z.

Overall, the molecular structures of SGPB reported here are identical to those determined previously.[2,15,16] Ionic interactions are observed in all of them, between the side chains of Asp29 and Arg139 in a hydrogen-bonding fashion (Figure 4-2; distance between Asp29 $O^{\delta 1}$ and Arg139 $N^{\eta 1}$: 2.9 Å), and between the side chains of Arg138 and Asp194 in a π-interacting fashion (Figure 4-2; distance between Arg138 C^{ζ} and Asp194 C^{γ}: 4.0 to 4.1 Å). In the structures of molecules A and B of SGPB at pH 7.3, the α-amino group of Ile16 (the N-terminus) and the side-chain carboxylate

Figure 4-2 Ionic interactions in molecule A of SGPB at pH 7.3. Hydrogen bonds are indicated by dashed lines. The ionic interaction between Asp29A and Arg139A (in a hydrogen-bonding manner), and that between Arg138A and Asp194A (in a π-interacting manner) are also observed in SGPB at pH 4.2 and in molecule B of SGPB at pH 7.3. The ionic interaction between Ile16A (the N-terminus) and Asp116A (in a water-mediated hydrogen-bonding manner, with the water molecule represented by a red sphere) is observed in molecules A and B of SGPB at pH 7.3 only. The charged groups of Ile16 and Asp29 are not oriented for ionic interaction in any of the molecular structures of SGPB reported here. Neither are those of Arg48A and Tyr242 (the C-terminus).



group of Asp116 also form an ionic interaction in a water-mediated hydrogen-bonding fashion (Figure 4-2; distance between the α-amino N of Ile16 and the oxygen of the water molecule: 2.8 Å; distance between the oxygen of the water molecule and Asp116 $O^{\delta 2}$: 2.6 to 2.7 Å). Importantly, Asp194 is positioned by its ionic interaction with Arg138 in such a way that the oxyanion hole of SGPB is in the catalytically competent conformation, with the backbone NH groups of Gly193 and Ser195 oriented for donating hydrogen bonds to the carbonyl O of the scissile peptide bond of a substrate. In the activation of chymotrypsinogen, a conformational rearrangement of its oxyanion hole is induced by the formation of a similar ionic interaction between the side-chain carboxylate group of Asp194 and the α-amino group of Ile16 released upon the hydrolysis of the Arg15-Ile16 peptide bond.[15,16,17]

The structural rigidity of SGPB is indicated by its low average B factors (8.71 to 11.86 $Å^2$) at both pH 4.2 and 7.3 (Table 4-1). The B factors of most residues are lower than the average values (Figures 4-3a, 4-3b and 4-3c). Superpositions of the molecular structure of SGPB at pH 4.2 with the structures of molecules A and B of SGPB at pH 7.3 show excellent agreement in most atomic positions (Table 4-2). Significant differences in atomic positions (greater than 2.7 times the rmsd) occur only at the C-terminus and in some loops near the molecular surface of SGPB, such as the loop from Asp60 to Ala63, and the segment of polypeptide chain from Gly117 to Ile124. Similar results were given by superpositions of the molecular structures of SGPB reported here with that from the complex of SGPB with polypeptide chymotrypsin inhibitor-1 (PCI-1) from Russet Burbank potato tubers,[18] and those from the complexes of SGPB with OMTKY3 and its variants determined so far

Figure 4-3 B factors (in $Å^2$) of the residues of SGPB in the molecular structures reported here. (a) SGPB at pH 4.2. (b) Molecule A and (c) molecule B of SGPB at pH 7.3. The main-chain and the side-chain B factors are represented by the orange and the purple vertical bars, respectively. The average values of the B factors of the main-chain and the side-chain atoms are indicated by the red and the blue lines, respectively. There are no purple bars for certain residues. These residues are glycines and do not have side chains.

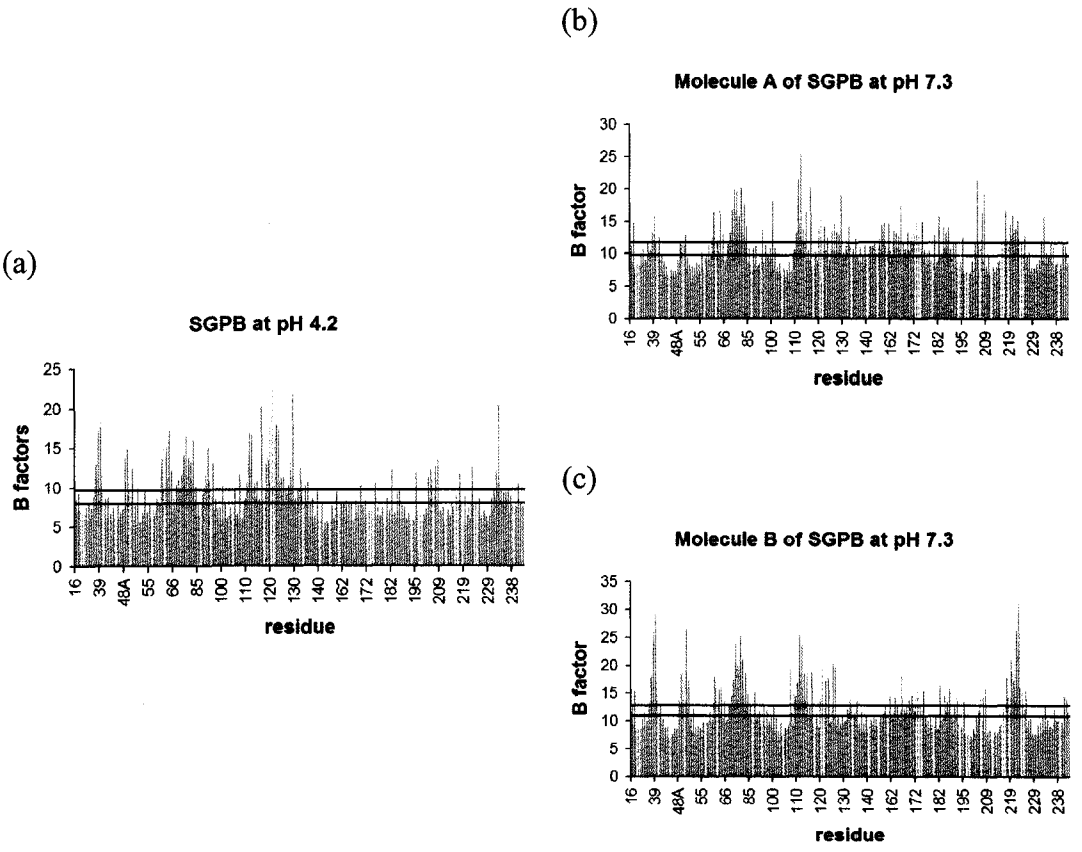


Table 4-2 Root-mean-square differences (rmsd; Å) for superpositions of the molecular structures of SGPB in different conditions[a]

	PDB	ID	**SGPB[b] (pH 4.2, $P2_12_12$)**	**SGPB[b] (pH 7.3, I23)**	
			E	A	B
SGPB[b] (pH 4.2, $P2_12_12$)	2QA9	E	-	0.34 (693)	0.34 (642)
SGPB[b] (pH 7.3, I23)	2QAA	A	-	-	0.26 (686)
SGPB:PCI-1[18]	4SGB	E	0.28 (659)	0.27 (701)	0.29 (689)
SGPB:wtOMTKY3[19,20]	3SGB	E	0.29 (645)	0.26 (699)	0.31 (713)
SGPB:OMTKY3[2]	1SGR	E	0.31 (650)	0.27 (709)	0.32 (723)
SGPB:OMTKY3-Gly18I[2]	1SGQ	E	0.33 (662)	0.27 (714)	0.27 (712)
SGPB:OMTKY3-Ala18I[2]	1SGP	E	0.31 (653)	0.27 (712)	0.24 (682)
SGPB:OMTKY3-Val18I[21]	1CT4	E	0.31 (652)	0.25 (700)	0.24 (678)
SGPB:OMTKY3-Ser18I[21]	1CT0	E	0.32 (661)	0.26 (709)	0.23 (682)
SGPB:OMTKY3-Thr18I[21]	1CT2	E	0.30 (649)	0.25 (702)	0.25 (690)
SGPB:OMTKY3-Ile18I[21]	1CSO	E	0.34 (665)	0.27 (708)	0.29 (707)
SGPB:OMTKY3-Trp18I[22]	2NU0	E	0.31 (651)	0.25 (702)	0.24 (688)
SGPB:OMTKY3-His18I[22]	2NU1	E	0.33 (662)	0.27 (711)	0.23 (681)
SGPB:OMTKY3-Tyr18I[c]	1SGY	E	0.33 (670)	0.26 (703)	0.26 (707)
SGPB:OMTKY3-Phe18I[c]	2SGF	E	0.29 (640)	0.26 (705)	0.28 (717)
SGPB:OMTKY3-Arg18I[22]	2NU2	E	0.34 (666)	0.26 (709)	0.24 (692)
SGPB:OMTKY3-Lys18I[22] (pH 7.1)	2NU3	E	0.34 (658)	0.28 (719)	0.24 (690)
SGPB:OMTKY3-Lys18I[22] (pH 10.7)	2NU4	E	0.31 (651)	0.26 (709)	0.27 (705)
SGPB:OMTKY3-Asp18I[23] (pH 6.5)	1SGD	E	0.33 (663)	0.28 (718)	0.28 (706)
SGPB:OMTKY3-Asp18I[23] (pH 10.7)	2SGD	E	0.36 (660)	0.31 (705)	0.32 (716)

Table 4-2 (continued)

	PDB	ID	SGPB[b] (pH 4.2, $P2_12_12$)	SGPB[b] (pH 7.3, I23)	
			E	A	B
SGPB:OMTKY3-Glu18I[23] (pH 6.5)	1SGE	E	0.32 (655)	0.27 (706)	0.29 (710)
SGPB:OMTKY3-Glu18I[23] (pH 10.7)	2SGE	E	0.36 (672)	0.30 (717)	0.31 (719)
SGPB:OMTKY3-Asn18I[23]	1SGN	E	0.35 (675)	0.27 (706)	0.29 (713)
SGPB:OMTKY3-Gln18I[23] (pH 6.5)	2SGQ	E	0.33 (666)	0.27 (713)	0.28 (715)
SGPB:OMTKY3-Gln18I[23] (pH 10.7)	3SGQ	E	0.33 (657)	0.27 (699)	0.28 (701)
SGPB:OMTKY3-Pro18I[24]	2SGP	E	0.37 (679)	0.30 (714)	0.30 (712)
SGPB:OMTKY3-COOLeu18I[24]	1DS2	E	0.30 (644)	0.25 (707)	0.26 (699)
SGPB:OMTKY3-Ala32I[25]	2GKV	E	0.34 (673)	0.30 (711)	0.37 (717)

[a] All superpositions started with residues 16 to 242 of SGPB (total number of main-chain atoms: 740) included in the calculations. The number of main-chain atoms included in the final calculation of rmsd for each superposition is given in the parentheses.

[b] These are the molecular structures reported here.

[c] Huang *et al.*, unpublished results

(Table 4-2).[6,19,20,21,22,23,24,25] (also Huang *et al.*, unpublished results; PDB accession codes: 2SGF and 1SGY). The less well-defined electron densities and the higher B factors of atoms in the loops (Figures 4-3a, 4-3b and 4-3c) indicate their higher conformational flexibilities and thus their higher susceptibilities to the effects of crystal packing. This is supported by the superposition of the structures of molecules A and B of SGPB at pH 7.3 (Table 4-2), which shows significant differences in atomic positions of some of these loops even though the two SGPB molecules are at the same pH.

Unexpected electron densities in the active site and the substrate-binding region

In the determination of the molecular structures (at resolutions higher than 2 Å) of trypsin, chymotrypsin and streptogrisin A (SGPA), the electron density maps showed unexpected positive densities in the active sites and the substrate-binding regions of these peptidases. These positive densities were finally assigned as oligopeptides. Depending on their electron-density connectivity with Ser195 O^{γ} in these peptidases, these oligopeptides were either bonded covalently to these peptidases in the form of second tetrahedral[26,27] or acyl-enzyme intermediates,[28,29,30,31] or they were bound non-covalently to these peptidases in the form of enzyme-product complexes.[27,32,33] In the electron density maps of SGPB at both pH 4.2 and 7.3, unexpected positive densities are also observed in its active site and substrate-binding region (Figures 4-4a, 4-4c and 4-4e). Assignment of these densities as molecular species from the reagents used in the preparation or the crystallization of SGPB was disproved by the subsequent refinement.

Figure 4-4 Unexpected electron densities in the active site and the substrate-binding region of SGPB. Panels (a), (c) and (e) show SGPB at pH 4.2, and molecules A and B of SGPB at pH 7.3, respectively, before map fitting and structure refinement. Panels (b), (d) and (f) show SGPB at pH 4.2, and molecules A and B of SGPB at pH 7.3, respectively, after map fitting and structure refinement. The $2|F_o|-|F_c|,\alpha_c$ maps are contoured at the 1.0σ level (dark blue), whereas the $|F_o|-|F_c|,\alpha_c$ maps are contoured at the 3.0σ level (magenta). Molecule I in (b), and the Tyr and the Leu molecules in (d) and (f) are colored green. Glycerol molecules in (d) and (f) are colored white and labelled 'GOL'. Water molecules in (b), (d) and (f) are represented by red spheres.

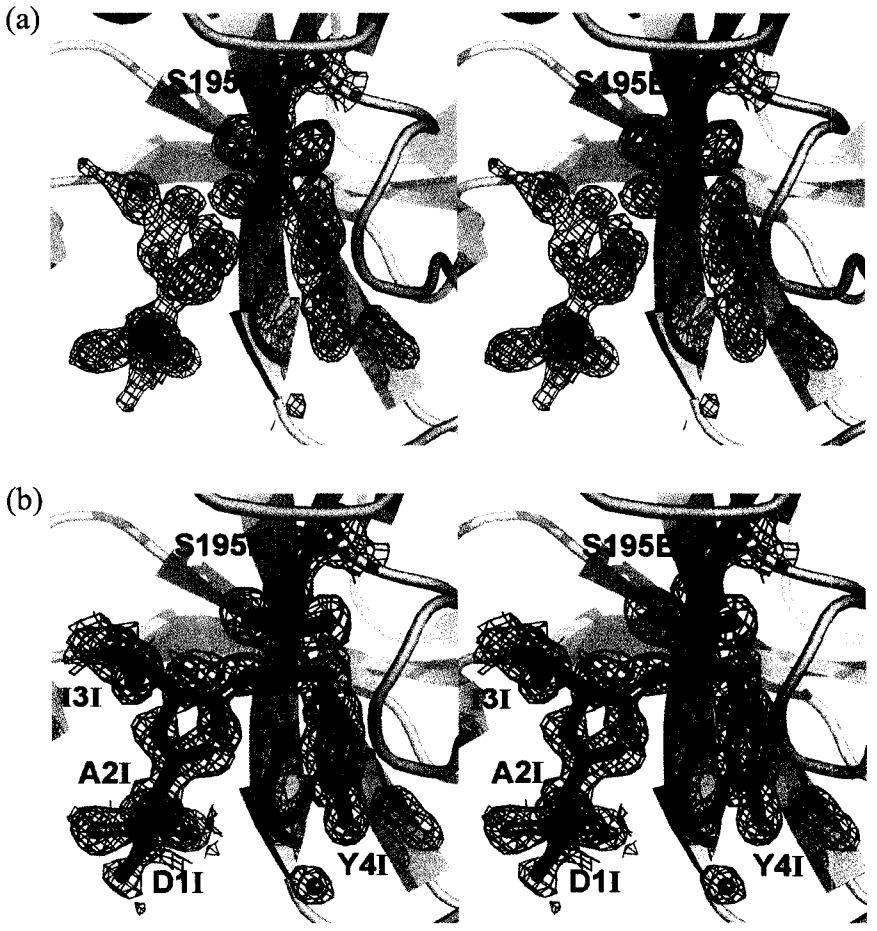
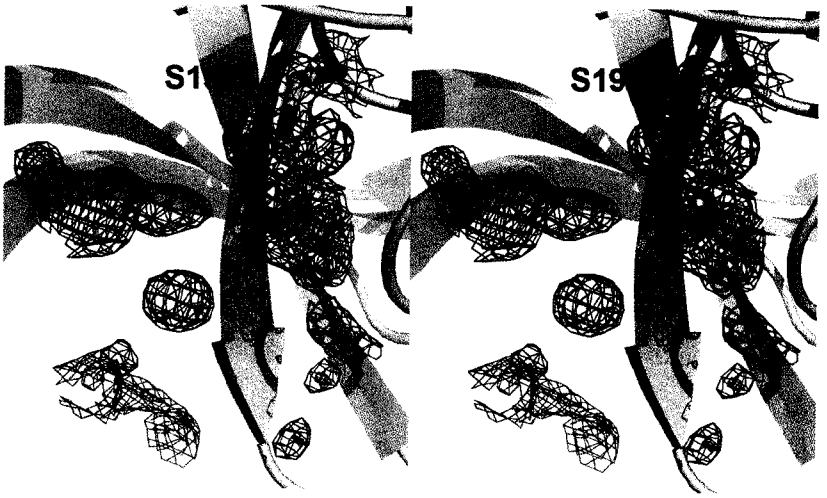


Figure 4-4 (continued)

(c)



(d)

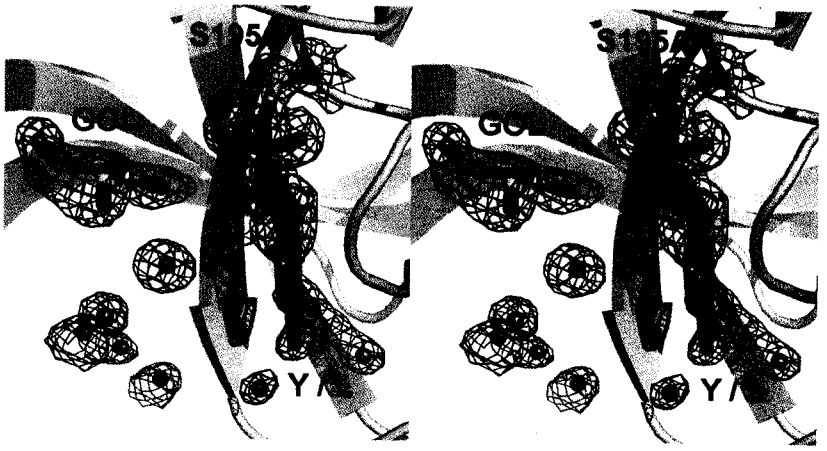
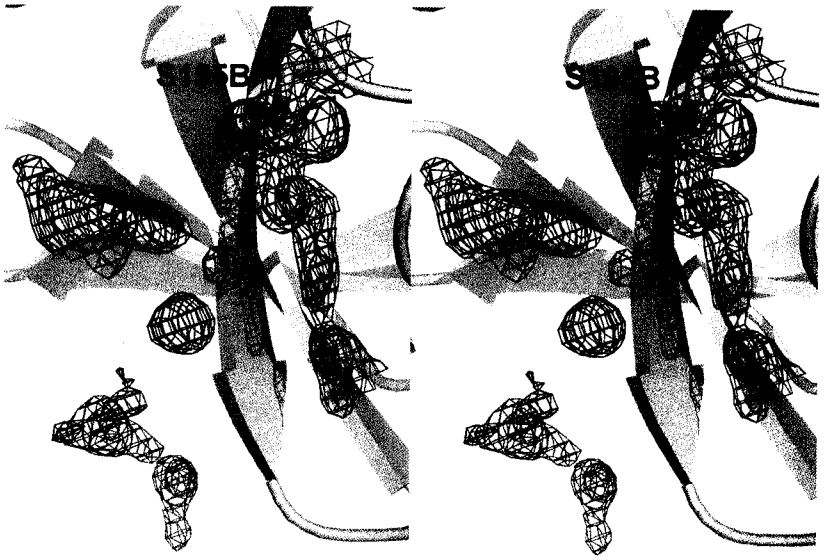
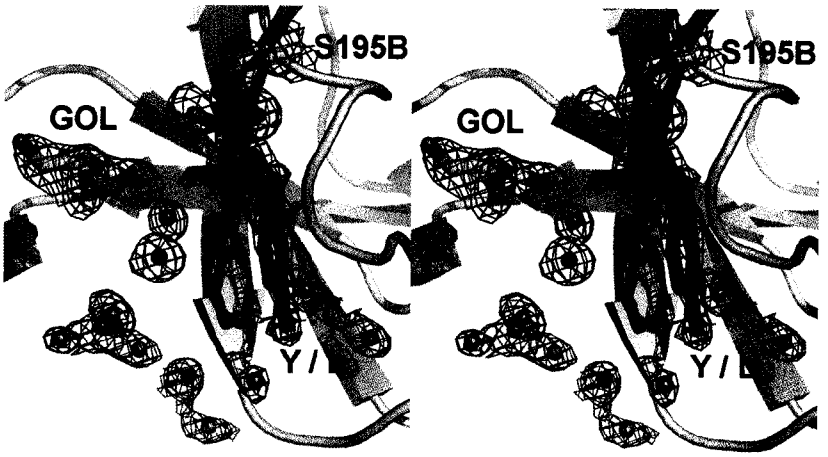


Figure 4-4 (continued)

(e)



(f)



In the electron density maps of the orthorhombic crystal at pH 4.2, the positive densities in the active site and the substrate-binding region of SGPB were continuous. They extended along residues Ser214E to Gly216E of SGPB (Figure 4-4a). Superposition of the molecular structure of the SGPB:OMTKY3 complex[6] onto that of SGPB in the electron density maps suggested the fitting of these positive densities with a tetrapeptide in a substrate-like manner. Similar to the oligopeptides trapped in the molecular structures of trypsin,[27] chymotrypsin[26,28,29,31,32] and SGPA,[30] this tetrapeptide probably arises from an autocatalytic proteolysis of SGPB during its preparation or crystallization, although its origin could not be identified definitively. The mass spectrum of the $P2_12_12$ crystal of SGPB showed a large number of closely spaced peaks below and above m/z = 18630.79, the calculated value for the monoisotopic molecular ion of SGPB (data not shown). The complexity of the spectrum also indicated that the SGPB molecules in the crystal have trapped not only a specific tetrapeptide, but in fact many oligopeptides of varying lengths and amino-acid sequences. This has been confirmed for chymotrypsin[28,32] and SGPA[30] by the high-performance liquid chromatography (HPLC) analysis of their crystal contents. Therefore, no single oligopeptide can actually satisfy the positive densities in the electron density maps of SGPB at pH 4.2 completely. Only the best-fitting tetrapeptide was selected from the amino-acid sequence of SGPB, based on its solvent accessibility in the molecular structure of SGPB, and its shape agreement with the positive densities. The best fit was achieved by the tetrapeptide Asp-Ala-Ile-Tyr (molecule I; corresponding to the sequence of residues Asp29 to Tyr32 of SGPB) at an occupancy of 0.70 (Figure 4-4b). Previous studies showed that the substrate-

binding region of SGPB extends over seven residues of a substrate (P4 to P3'), with four of them preceding the scissile peptide bond (P4 to P1).[34,35,36,37,38] The average B factor of molecule I (4.55 $Å^2$) is lower than that of molecule E (8.71 $Å^2$) of SGPB (Table 4-1), probably indicating an underestimation of the extent to which the active site and the substrate-binding region of SGPB are occupied by various molecular species (not limited to molecule I) in the $P2_12_12$ crystal. Furthermore, the mobility of molecule I may be restricted by crystal packing. In particular, Asp1I and Ala2I of molecule I make contacts with the side chains of Asp123E and Thr125E of molecule E of SGPB from the corresponding asymmetric unit in the neighboring unit cell. This probably explains why the average B factor of molecule I does not increase substantially at its N-terminus, unlike those of the oligopeptides trapped in trypsin,[27] chymotrypsin[26,28,29,31,32] and SGPA.[30]

In the electron density maps of the cubic crystal at pH 7.3, the positive densities in the active sites and the substrate-binding regions of both molecules A and B of SGPB exist as discontinuous patches spread along residues Ser214 to Gly216 of the SGPB molecules. The two largest patches occupy the S1 and the S2 specificity pockets (Figures 4-4c and 4-4e). The mass spectrum of the contents of the I23 crystal of SGPB is similar to that of the $P2_12_12$ crystal (data not shown). However, attempts to fit the positive densities in both molecules A and B of SGPB with the tetrapeptide Ala-Ala-Ala-Ala resulted in the development of extensive regions of negative densities surrounding the peptide bonds of this tetrapeptide. Independent fitting of the patches of positive densities was necessary. The patches in the S1 specificity pockets of both molecules A and B of SGPB were fitted with amino acids. This was

consistent with the fact that the catalytic activity of SGPB is maximal at pH 7.3,[34] so that the autocatalytic proteolysis of SGPB might progress further during crystallization at this pH than at pH 4.2. The identities of these amino acids were determined based on the shapes of the positive densities in the S1 specificity pockets. For molecule A of SGPB, the patch in the S1 specificity pocket was fitted with a tyrosine molecule and a leucine molecule, each at an occupancy of 0.50 (Figure 4-4d); whereas for molecule B of SGPB, it was fitted with the same two molecules, each at an occupancy of 0.40 (Figure 4-4f). The average B factors of the Tyr and the Leu molecules (21.09 $Å^2$ in molecule A, 18.47 $Å^2$ in molecule B) are higher than those of their parent SGPB molecules (10.66 $Å^2$ for molecule A, 11.86 $Å^2$ for molecule B; Table 4-1), probably indicating an overestimation of the occupancies of the Tyr and the Leu molecules in the active sites of both molecules A and B of SGPB in the I23 crystal. Moreover, unlike molecule I having extensive interactions with molecule E of SGPB at pH 4.2 (to be described below), the Tyr and the Leu molecules have few interactions with both molecules A and B of SGPB at pH 7.3, thereby acquiring greater mobility.

The best fit of the positive electron-density patches in the S2 specificity pockets of both molecules A and B of SGPB was achieved by full-occupancy glycerol molecules (from the cryoprotectant) positioned in the same orientation (Figures 4-4d and 4-4f). Interestingly, the atomic-resolution molecular structures of the unbound α-lytic peptidase (αLP) also showed the presence of a glycerol molecule in this orientation in its S2 specificity pocket.[39,40] The S2 specificity pockets of SGPB and αLP are essentially identical in structure, and probably favor the binding of

glycerol molecules in this orientation. The positive electron-density patches in the rest of the substrate-binding regions of molecules A and B of SGPB were assigned as water molecules (Figures 4-4d and 4-4f).

Active-site residues and their interactions with the bound species

Electron densities are observed for both of the α-carboxylate O and OXT in Tyr4I at pH 4.2, and in the bound Tyr and Leu molecules at pH 7.3 (Figures 4-4a to 4-4f). For each of these residues, the α-carboxylate O occupies the oxyanion hole of the parent SGPB molecule, forming hydrogen bonds with the backbone NH groups of its residues Gly193 and Ser195 (Table 4-3). Interestingly, in both molecules A and B of SGPB at pH 7.3, the electron densities of Ser195 O^{γ} of SGPB are connected to those of the α-carboxylate carbons of the bound Tyr and Leu molecules (Figures 4-4c and 4-4e). In the later stages of refinement, geometric restraints were relaxed, and the distances of these connections shortened, following the displacement of the α-carboxylate carbons of the bound Tyr and Leu molecules out of the planes defined by their respective C^{α}, and α-carboxylate O and OXT (Figures 4-4d and 4-4f). In the final structure, the out-of-plane displacement of the α-carboxylic carbons of the Tyr and the Leu molecules (Δ in Figure 4-5) bound in the active site of molecule A of SGPB is 0.3 Å, and the distance of these atoms from Ser195A O^{γ} of SGPB (d in Figure 4-5) is 1.55 Å (Table 4-4). For molecule B of SGPB, the final values of Δ and d are 0.3 Å and 1.69 Å, respectively (Table 4-4). In contrast, at pH 4.2, the electron densities of Ser195E O^{γ} of SGPB and the α-carboxylic C of Tyr4I became continuous only near the end of refinement (Figures 4-4a and 4-4b). The final values of Δ and d are 0.1 Å and 1.93 Å, respectively (Table 4-4). The various values of Δ and d

Table 4-3 Interactions among the active-site residues in the Michaelis complex and the second tetrahedral intermediates of SGPB

	Michaelis complex	**Second tetrahedral intermediates**		
pH	6.5	4.2	7.3	
σ_r (Å)[a]	0.190	0.064	0.045	
ID of SGPB	E	E	A	B
Bound species	OMTKY3	Asp-Ala-Ile-Tyr	Tyr, Leu	Tyr, Leu
Ser195 O^{γ} ... His57 $N^{\varepsilon 2}$ (Å)	2.6	3.1	3.3	3.1
Asp102 $O^{\delta 2}$ ... His57 $N^{\delta 1}$ (Å)	2.9	2.7	2.7	2.7
Asp102 $O^{\delta 2}$ ... Ser214 O^{γ} (Å)	2.6	2.6	2.7	2.7
His57 $C^{\varepsilon 1}$ ... Ser214 O (Å)	3.1	3.0	3.0	3.0
Gly193 N (SGPB) ... P1 O (bound species) (Å)	2.5[b]	2.8	2.8	2.7
Ser195 N (SGPB) ... P1 O (bound species) (Å)	3.1[b]	2.9	3.3	3.2
His57 $N^{\varepsilon 2}$ (SGPB) ... P1 OXT (bound species) (Å)	4.1[b]	2.7	2.7	2.7
Ser195 O^{γ} (SGPB) ... P1 N (bound species) (Å)	2.9	2.9	-	-

[a] σ_r, overall positional uncertainty.

[b] For the SGPB:OMTKY3 complex, the carbonyl O of Leu18I of OMTKY3 is considered equivalent to the α-carboxylate O of residue P1 of the bound species, whereas the amide N of Glu19I of OMTKY3 is considered equivalent to the α-carboxylate OXT of residue P1 of the bound species.

Figure 4-5 Geometric parameters describing the trajectory of the nucleophilic attack of Ser195 O^{γ} of SGPB on the α-carboxylate C of the bound species. Plane A (green) is defined by atoms C^{α} (green), OXT (green) and O (yellow) of residue P1 of the bound species, whereas plane B is defined by Ser195 O^{γ} of SGPB (red), and the α-carboxylic C (red) and O (yellow) of residue P1 of the bound species. d is the distance between Ser195 O^{γ} of SGPB and the α-carboxylic C of residue P1 of the bound species. Δ is the displacement of the α-carboxylic C of residue P1 of the bound species out of plane A. θ_1 is the angle between planes A and B. θ_2 is the angle O^{γ}(Ser195 of SGPB)-C-O(residue P1 of the bound species).

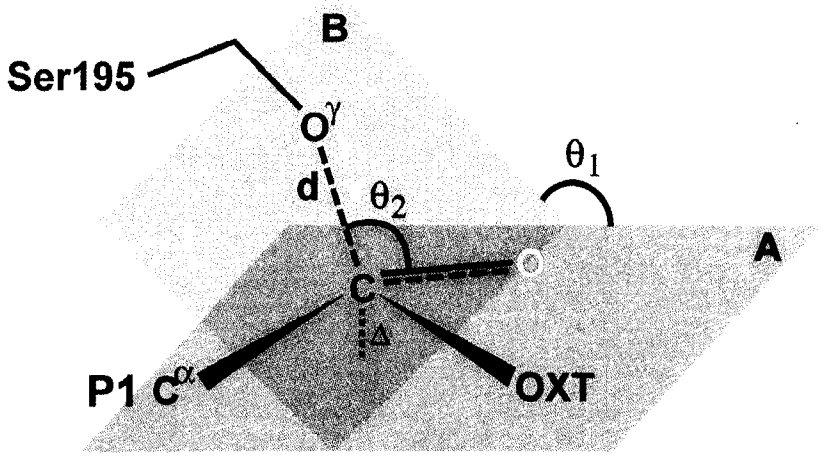


Table 4-4 Trajectory of the nucleophilic attack in the Michaelis complex and the second tetrahedral intermediates of SGPB

	Michaelis complex	**Second tetrahedral intermediates**		
pH	6.5	4.2	7.3	
σ_r (Å)[a]	0.190	0.064	0.045	
ID of SGPB	E	E	A	B
Bound species	OMTKY3	Asp-Ala-Ile-Tyr	Tyr, Leu	Tyr, Leu
Δ (Å)	0.0[b]	0.1	0.3	0.3
d (Å)	2.8[b]	1.93	1.55	1.69
θ_1 (°)	82.2[b]	88.8	86.3	82.8
θ_2 (°)	92.0[b]	97.3	111.9	105.7

[a] σ_r, overall positional uncertainty.

[b] For the SGPB:OMTKY3 complex, the carbonyl O of Leu18I of OMTKY3 is considered equivalent to the α-carboxylate O of residue P1 of the bound species, whereas the amide N of Glu19I of OMTKY3 is considered equivalent to the α-carboxylate OXT of residue P1 of the bound species.

exhibited by SGPB at pH 4.2 and 7.3 suggest that each of these continuous electron densities represents a mixture of the second tetrahedral intermediates and the enzyme-product complexes of SGPB existing in a pH-dependent equilibrium. Apparently, at pH 7.3, the second tetrahedral intermediates of SGPB are favored. At pH 4.2, the protonation of Ser195E O^{γ} of SGPB may be favored, leading to the dissociation of the covalent bond between this atom and the α-carboxylic C of Tyr4I. Hence, the enzyme-product complexes of SGPB dominate the mixture. The acyl-enzyme intermediates and the enzyme-product complexes of chymotrypsin also seem to maintain a pH-dependent equilibrium. However, its acyl-enzyme intermediates are favored at low pH, whereas its enzyme-product complexes are favored at high pH.[29]

In the molecular structures reported here, the angle θ_1 for the nucleophilic attack of Ser195 O^{γ} of SGPB on the α-carboxylic C of residue P1 of the bound species (Figure 4-5) is in the range of 82.8 ° to 88.8 ° (Table 4-4). In fact, geometric analyses of the Michaelis complex, the tetrahedral intermediate and the good acyl-enzyme intermediates of trypsin, and of various serine peptidase-protein inhibitor complexes showed that θ_1 is close to 90 ° in all cases.[25,41,42] In contrast, the angle of nucleophilic attack θ_2 (Figure 4-5) depends on the degree of tetrahedral distortion at the α-carboxylic C of residue P1 of the bound species. In both molecules A and B of SGPB at pH 7.3, θ_2 is close to 109.5 ° (Table 4-4), similar to that in the tetrahedral intermediate of trypsin.[42] At pH 4.2, θ_2 is close to 90 ° (Table 4-4), similar to those values in the Michaelis complex and the good acyl-enzyme intermediates of trypsin, and in various serine peptidase-protein inhibitor complexes.[25,41,42]

Although the molecular structures reported here represent the second tetrahedral intermediates of SGPB, they can also serve as models of the first tetrahedral intermediates of SGPB. The molecular structures of the complexes of SGPB with OMTKY3 and its variants[6,19,20,21,22,23,24,25] (also Huang *et al.*, unpublished results; PDB accession codes: 2SGF and 1SGY) serve as excellent models of the Michaelis complex of SGPB. Comparisons of these molecular structures show that, upon the formation of the tetrahedral intermediate, the carbonyl O of residue P1 of the bound species moves, and this is accompanied by the movement of residues Glu192A to Gly193 of SGPB (shifts in atomic positions: up to 1.0 Å) towards this atom (Figure 4-6). The purpose of this movement is probably to maintain the distance of the hydrogen bond between the atom O of residue P1 of the bound species and the backbone NH of Gly193 of SGPB (distances: 2.7 to 2.8 Å; Table 4-3), so that the oxyanion developed upon the formation of the tetrahedral intermediate can be stabilized effectively. Without this movement, the distance of this hydrogen bond would be 3.0 Å or greater in the tetrahedral intermediate. In contrast, the movement of residues Asp194 and Ser195 of SGPB upon the formation of the tetrahedral intermediate is small (shifts in atomic positions: 0.3 Å or smaller; Figure 4-6). This is probably because the O^{γ} atom of Ser195 of SGPB must be positioned for its nucleophilic attack on the carbonyl C of residue P1 of the substrate.

Moreover, adjustments occur in the side-chain conformational angles of His57 (by 10 ° to 15 ° in χ_1 and by 10 ° to 20 ° in χ_2) and Ser195 (by 20 ° to 25 ° in χ_1) of SGPB upon the formation of the tetrahedral intermediate (Table 4-5). As a result, the hydrogen bond between His57 $N^{\varepsilon 2}$ and Ser195 O^{γ} within SGPB is weakened, whereas

Figure 4-6 Movement of the oxyanion-hole residues of SGPB towards the carbonyl O of residue P1 of the bound species upon the formation of the tetrahedral intermediate. The molecular structure of SGPB at pH 4.2 reported here serves as a model of the tetrahedral intermediate of SGPB (colored yellow for SGPB and green for tetrapeptide I), whereas the molecular structure of the SGPB:OMTKY3 complex serves as a model of the Michaelis complex of SGPB (colored white for SGPB and blue for OMTKY3). For clarity, only residues P1 and P2 of the bound species are shown. Similar movements of the oxyanion-hole residues of SGPB are observed in both molecules A and B of SGPB at pH 7.3 reported here as well.

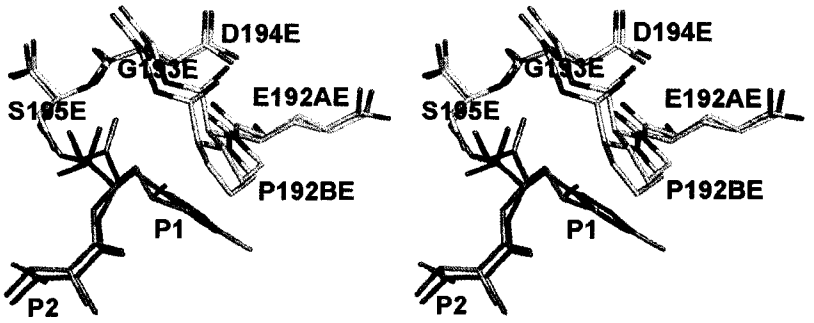


Table 4-5 Conformational angles of His57 and Ser195 in the Michaelis complex and the second tetrahedral intermediates of SGPB

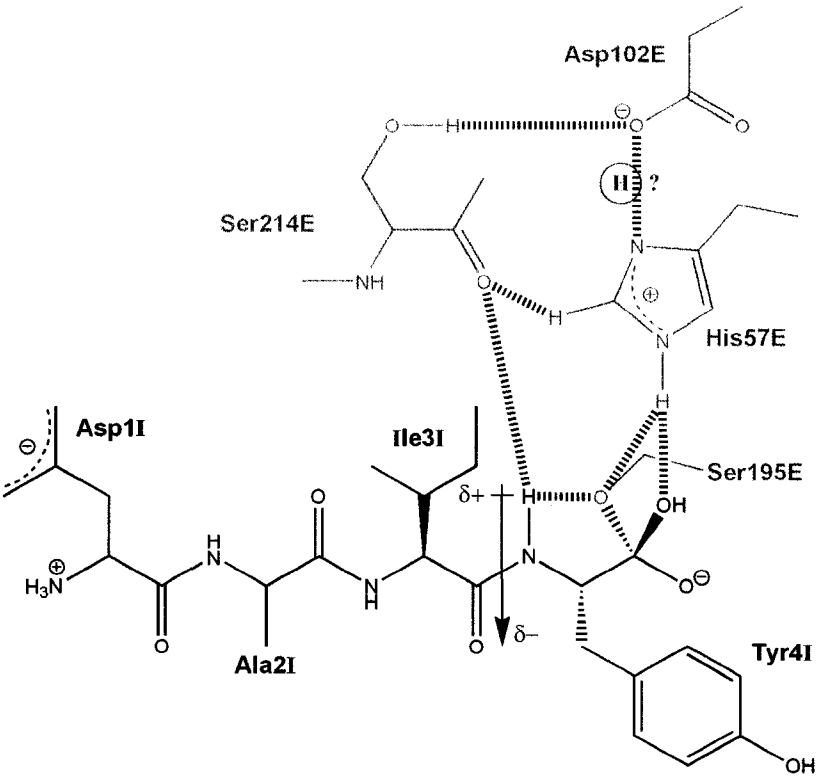
	Michaelis complex	**Second tetrahedral intermediates**		
pH	6.5	4.2	7.3	
σ_r (Å)[a]	0.190	0.064	0.045	
ID of SGPB	E	E	A	B
Bound species	OMTKY3	Asp-Ala-Ile-Tyr	Tyr, Leu	Tyr, Leu
His57				
ϕ (°)	-68.3	-75.2	-68.7	-73.1
ψ (°)	-16.1	-11.6	-16.6	-12.6
χ_1 (°)	63.7	74.3	76.8	75.6
χ_2 (°)	-88.6	-108.9	-101.9	-101.8
Ser195				
ϕ (°)	-46.3	-53.3	-54.1	-53.9
ψ (°)	136.0	140.7	140.6	141.2
χ_1 (°)	-85.8	-62.5	-62.5	-65.7

[a] σ_r, overall positional uncertainty.

that between His57 $N^{\varepsilon 2}$ of SGPB and the atom OXT of residue P1 of the bound species is strengthened (Figure 4-7; Table 4-3). In the SGPB-catalyzed proteolysis of true substrates, this would prevent the back-protonation of Ser195 O^{γ} by His57 $N^{\varepsilon 2}$ within SGPB, and would promote the protonation of the amide N of residue P1' of the substrate by His57 $N^{\varepsilon 2}$ of SGPB, thereby favoring the progression of the acylation stage. The previously proposed flipping of the imidazole ring of His57[43] does not seem to be required to drive the catalytic mechanism of SGPB. In fact, analysis of the anisotropic thermal motions of the atoms of His57 in both of the molecular structures of SGPB reported here shows that the wobbling of the imidazole ring of His57 about its C^{β}-C^{γ} bond is limited. The same conclusions were drawn for elastase,[44,45] trypsin[42] and αLP[40] as well.

Upon the formation of tetrahedral intermediates, His57 $N^{\varepsilon 2}$ of SGPB becomes protonated. The positively-charged imidazolium ring of His57 is stabilized by the negatively-charged side-chain carboxylate group of Asp102 within SGPB. In both of the molecular structures reported here, a hydrogen bond at a distance of 2.7 Å between His57 $N^{\delta 1}$ and Asp102 $O^{\delta 2}$ within SGPB is present (Figure 4-7; Table 4-3). Whether this is a low-barrier hydrogen bond (LBHB) cannot be determined in this study, mainly because the position of the proton in this hydrogen bond cannot be located in the 1.2 Å-resolution electron density maps here. Also, a distance of 2.7 Å is at the limit (2.65 Å) for the formation of an LBHB.[46] As discussed previously, the character of a hydrogen bond is influenced by its specific environment, and thus cannot be determined based on few simplistic definitions. Comprehensive

Figure 4-7 Schematic diagram of the catalytic triad (His57E, Asp102E and Ser195E) and Ser214E of SGPB (orange), and molecule I (green) in the molecular structure at pH 4.2. Hydrogen bonds are indicated by dashed lines. The dipole across the Ile3I-Tyr4I peptide bond is indicated by a straight arrow with a cross at its positive end. The proton in the hydrogen bond between His57E $N^{\delta 1}$ and Asp102 $O^{\delta 2}$ within SGPB could not be located in this study.



experimental results concerning various features of a hydrogen bond are required to establish reliably whether it is an LBHB.[40]

In addition to the catalytic triad (His57, Asp102 and Ser195), Ser214 is conserved among the chymotrypsin-like serine peptidases. Its carbonyl O is hydrogen-bonded to His57 $C^{\varepsilon 1}$, and its O^{γ} is hydrogen-bonded to Asp102 $O^{\delta 2}$ (Figure 4-7). His57 and Asp102 are thus positioned to carry out their functions in the catalytic mechanism.[47] The unusual oxygen-carbon hydrogen bond between the carbonyl O of Ser214 and His57 $C^{\varepsilon 1}$ has also been proposed previously to affect the charge distribution of the imidazolium ring of His57 in the tetrahedral-intermediate state, promoting the function of His57 as a general base or a general acid at different steps of the catalytic mechanism.[48] Both hydrogen bonds are observed in the molecular structures of SGPB reported here (Table 4-3).

In the molecular structure of SGPB at pH 4.2, the main chain of tetrapeptide I interacts with that of SGPB from Ser214E to Gly216E in the fashion of an anti-parallel β-sheet, forming hydrogen bonds between the NH and the carbonyl groups from the two main chains (2.9 to 3.3 Å). The angles ϕ and ψ of residues Asp1I to Tyr4I agree with those of residues P4 to P1 of the SGPB-bound OMTKY3[6] and the SGPB-bound PCI-1[18] (values not shown). The Ser195E O^{γ} of SGPB is positioned near the positive end of the dipole across the Ile3I-Tyr4I peptide bond, accepting a hydrogen bond from the main-chain NH of Tyr4I (Figure 4-7; Table 4-3). As discussed previously,[25] this indicates that serine peptidase catalysis is probably substrate-assisted.

S1 specificity pocket

As shown by its molecular structures in different conditions[2,6,15,16,18,19,20,21,22,23,24,25,36] (also Huang *et al.*, unpublished results; PDB accession codes: 2SGF and 1SGY), the S1 specificity pocket of SGPB is lined by residues Ala192 to Pro192B and residues Thr213 to Gly216. The side chains of Ala192 and Thr213 form the 'back wall' of this pocket. Interestingly, previous studies showed that simultaneous substitutions of Thr213 and some of these other lining residues could generate not only variants of SGPB with altered S1 specificities, but also those with altered thermodynamic stabilities.[49]

The S1 specificity pocket of molecule E of SGPB at pH 4.2 is occupied by the side chain of Tyr4I of molecule I (Figure 4-8a). In fact, the electron densities of Tyr4I $C^{\varepsilon 2}$ and C^{ζ} became continuous only in the later stages of refinement. A separate and weak electron density centered less than 2 Å away from Tyr4I $C^{\varepsilon 2}$ was assigned as a water molecule at an occupancy of 0.3 (occupancy of molecule I: 0.7). This water molecule forms a hydrogen bond with Thr226E $O^{\gamma 1}$ of SGPB (Figure 4-8a). In the final structure, the geometry at Tyr4I C^{ζ} is distorted from the normally observed trigonal planarity. The angles $C^{\varepsilon 1}$-C^{ζ}-O^{η} and $C^{\varepsilon 2}$-C^{ζ}-O^{η} are 112.8 ° and 128.1 °, respectively. Similar to those of Tyr18I in the SGPB:OMTKY3-Tyr18I complex (Huang *et al.*, unpublished results; PDB accession code: 1SGY), the angles χ_1 and χ_2 of Tyr4I are -65.9 ° and -34.0 °, respectively, and Tyr4I O^{η} is hydrogen-bonded to two water molecules (2.6 Å and 3.0 Å; Figure 4-8a).

The S1 specificity pockets of both molecules A and B of SGPB at pH 7.3 are occupied by the side chains of the Tyr and the Leu molecules at equal occupancies

Figure 4-8 S1 specificity pocket of SGPB. (a) SGPB at pH 4.2. (b) molecule A and (c) molecule B of SGPB at pH 7.3. The $2|F_o|-|F_c|,\alpha_c$ maps are contoured at the 1.0σ level (dark blue), whereas the $|F_o|-|F_c|,\alpha_c$ maps are contoured at the 3.0σ level (magenta). Water molecules are represented by red spheres; those observed in the molecular structure of the SGPB:OMTKY3-Tyr18I complex (Huang *et al.*, unpublished results; PDB accession code: 1SGY) are labeled `w'. Hydrogen bonds are indicated by dashed lines.

(a)

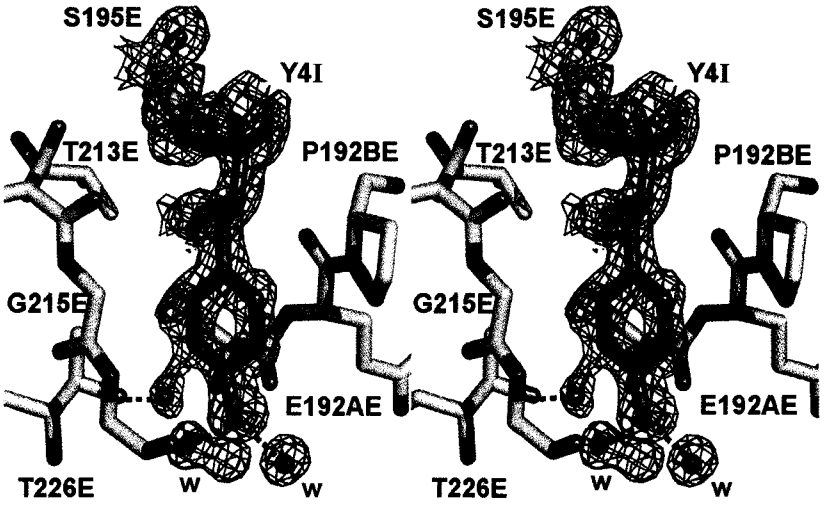
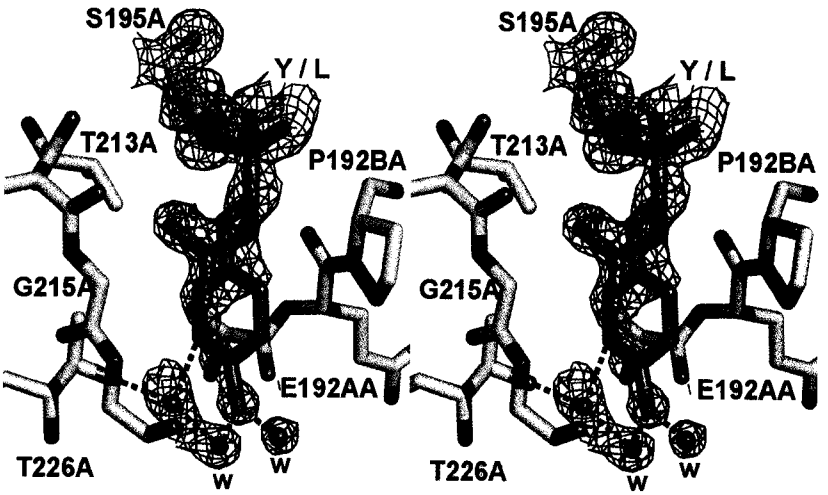
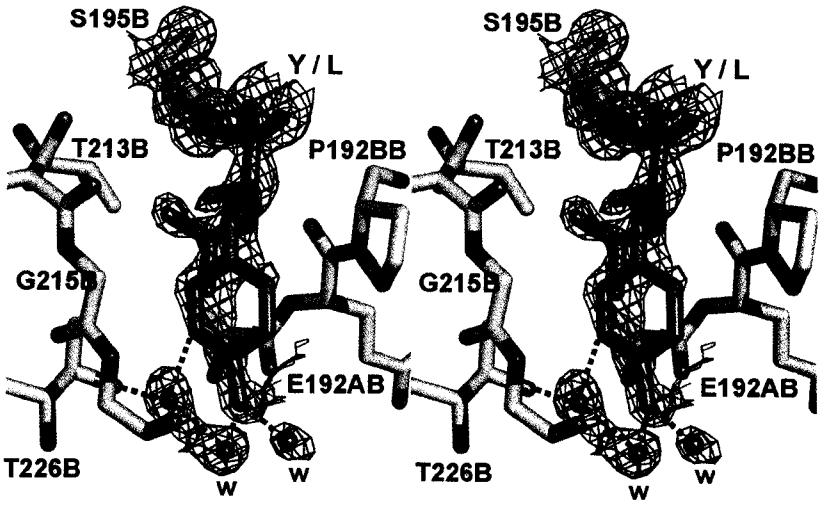


Figure 4-8 (continued)

(b)



(c)



(Figures 4-8b and 4-8c). For the Tyr molecules, the angles χ_1 and χ_2 are -67.0 ° and -22.5 °, respectively, in molecule A, and -70.1 ° and -45.3 °, respectively, in molecule B. The angles $C^{\varepsilon 1}$-C^{ζ}-O^{η} and $C^{\varepsilon 2}$-C^{ζ}-O^{η} are 115.7 ° and 130.0 °, respectively, in molecule A, and 112.9 ° and 131.5 °, respectively, in molecule B. The atom O^{η} of each Tyr molecule is hydrogen-bonded to two water molecules (2.3 Å and 3.3 Å in molecule A, 2.4 Å and 3.6 Å in molecule B). A full-occupancy water molecule forms hydrogen bonds with the atom $C^{\varepsilon 1}$ of each Tyr molecule (2.5 Å in molecule A, 2.4 Å in molecule B) and with Thr226 $O^{\gamma 1}$ of the parent SGPB molecule (2.7 Å in molecule A, 2.8 Å in molecule B; Figures 4-8b and 4-8c). For the Leu molecules, the angles χ_1 and χ_2 are 61.4 ° and -109.3 °, respectively, in molecule A, and 44.7 ° and -126.3 °, respectively, in molecule B. These contrast the angles χ_1 (-56.8 °) and χ_2 (173.3 °) of Leu18I of SGPB-bound OMTKY3,[6] probably indicating an alternative binding mode of the side chain of Leu in the S1 specificity pocket of SGPB. It is also possible that the side chain of each Leu molecule actually exhibits both the previously observed and the alternative conformers in the S1 specificity pocket of the parent SGPB molecule. However, the previously observed conformer occupies the same space as does the side chain of the Tyr molecule, and thus is not discernible in the electron density maps (Figures 4-8b and 4-8c).

Insights into the catalytic mechanism of serine peptidases

Extensive studies have been done to elucidate the catalytic mechanism of serine peptidases.[50] Importantly, the recently determined high-resolution molecular structures of serine peptidases, including the 1.2 Å-resolution molecular structures of SGPB reported here, allow for the elucidation of the mechanism in precise structural

details.[51] The molecular structures reported here represent the second tetrahedral intermediates of SGPB. In all of them, the angle θ_1 for the nucleophilic attack of Ser195 O^{γ} of SGPB on the α-carboxylate C of residue P1 of the bound species is close to 90 °. Comparisons among the molecular structures of the complexes of SGPB with OMTKY3 and its variants, which can serve as models of the Michaelis complex of SGPB, showed that residues Glu192A to Gly193 of SGPB move towards the carbonyl O of residue P1 of the bound species upon the formation of the tetrahedral intermediate, in order to maintain the effective stabilization of the oxyanion developed. Meanwhile, adjustments in the side-chain conformational angles of His57 and Ser195 of SGPB favor the progression of the catalytic mechanism of SGPB. No sign of ring flipping is observed for His57 of SGPB. Many of these observations are consistent with those given by the high-resolution molecular structures of other serine peptidases.

4.4 References

1. Jurášek, L., Carpenter, M. R., Smillie, L. B., Gertler, A., Levy, S. & Ericsson, L. H. (1974). Amino acid sequence of *Streptomyces griseus* protease B, a major component of Pronase. *Biochem. Biophys. Res. Commun.* **61**, 1095-1100.

2. Codding, P. W., Delbaere, L. T., Hayakawa, K., Hutcheon, W. L., James, M. N. G. & Jurášek, L. (1974). The 4.5 Å resolution structure of a bacterial serine protease from *Streptomyces griseus*. *Can. J. Biochem.* **52**, 208-220.

3. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W., Jr. & Sweet, R. M., eds.), Vol. 276: Macromolecular Crystallography, Part A, pp. 307-326 Academic Press, New York, NY.

4. Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

5. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Cryst.* **D59**, 1131-1137.

6. Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr. & James, M. N. G. (1995). Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci.* **4**, 1985-1997.

7. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.

8. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240-255.

9. McRee, D. E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.

10. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

11. Evans, P. & Fermi, J. (1995). GEOMCALC. Medical Research Council Laboratory of Molecular Biology, Cambridge.

12. Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593-604.

13. Cohen, G. H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Cryst.* **30**, 1160-1161.

14. Cruickshank, D. W. J. (1999). Remarks about protein structure precision. *Acta Cryst.* **D55**, 583-601.

15. Delbaere, L. T., Hutcheon, W. L., James, M. N. G. & Thiessen, W. E. (1975). Tertiary structural differences between microbial serine proteases and pancreatic serine enzymes. *Nature* **257**, 758-763.

16. Delbaere, L. T., Brayer, G. D. & James, M. N. G. (1979). The 2.8 Å resolution structure of *Streptomyces griseus* protease B and its homology with alpha-

chymotrypsin and *Streptomyces griseus* protease A. *Can. J. Biochem.* **57**, 135-144.

17. Stroud, R. M., Kossiakoff, A. A. & Chambers, J. L. (1977). Mechanisms of zymogen activation. *Annu. Rev. Biophys. Bioeng.* **6**, 177-193.

18. Greenblatt, H. M., Ryan, C. A. & James, M. N. G. (1989). Structure of the complex of *Streptomyces griseus* proteinase B and polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1 Å resolution. *J. Mol. Biol.* **205**, 201-228.

19. Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M., Jr. & James, M. N. G. (1982). Refined crystal structure of the molecular complex of *Streptomyces griseus* protease B, a serine protease, with the third domain of the ovomucoid inhibitor from turkey. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4868-4872.

20. Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. (1983). Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-Å resolution. *Biochemistry* **22**, 4420-4433.

21. Bateman, K. S., Anderson, S., Lu, W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2000). Deleterious effects of beta-branched residues in the S1 specificity pocket of *Streptomyces griseus* proteinase B (SGPB): crystal structures of the turkey ovomucoid third domain variants Ile18I, Val18I, Thr18I, and Ser18I in complex with SGPB. *Protein Sci.* **9**, 83-94.

22. Bateman, K. S. (1999). *Structural studies of protein-protein interactions*. PhD thesis, University of Alberta.

23. Huang, K. (1995). Structural studies of the interactions between serine proteinases and protein inhibitors. PhD thesis, University of Alberta.

24. Bateman, K. S., Huang, K., Anderson, S., Lu, W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2001). Contribution of peptide bonds to inhibitor-protease binding: crystal structures of the turkey ovomucoid third domain backbone variants OMTKY3-Pro18I and OMTKY3-psi[COO]-Leu18I in complex with *Streptomyces griseus* proteinase B (SGPB) and the structure of the free inhibitor, OMTKY-3-psi[CH2NH2+]-Asp19I. *J. Mol. Biol.* **305**, 839-849.

25. Lee, T.-W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2007). Structural insights into the non-additivity effects in the sequence-to-reactivity algorithm for serine peptidases and their inhibitors. *J. Mol. Biol.* **367**, 527-546.

26. Yennawar, N. H., Yennawar, H. P. & Farber, G. K. (1994). X-ray crystal structure of gamma-chymotrypsin in hexane. *Biochemistry* **33**, 7326-7336.

27. Rypniewski, W. R., Østergaard, P. R., Nørregaard-Madsen, M., Dauter, M. & Wilson, K. S. (2001). *Fusarium oxysporum* trypsin at atomic resolution at 100 and 283 K: a study of ligand binding. *Acta Cryst.* **D57**, 8-19.

28. Dixon, M. M. & Matthews, B. W. (1989). Is gamma-chymotrypsin a tetrapeptide acyl-enzyme adduct of alpha-chymotrypsin? *Biochemistry* **28**, 7033-7038.

29. Dixon, M. M., Brennan, R. G. & Matthews, B. W. (1991). Structure of gamma-chymotrypsin in the range pH 2.0 to pH 10.5 suggests that gamma-

chymotrypsin is a covalent acyl-enzyme adduct at low pH. *Int. J. Biol. Macromol.* **13**, 89-96.

30. Blanchard, H. & James, M. N. G. (1994). A crystallographic re-investigation into the structure of *Streptomyces griseus* proteinase A reveals an acyl-enzyme intermediate. *J. Mol. Biol.* **241**, 574-587.

31. Singh, N., Jabeen, T., Sharma, S., Roy, I., Gupta, M. N., Bilgrami, S., Somvanshi, R. K., Dey, S., Perbandt, M., Betzel, C., Srinivasan, A. & Singh, T. P. (2005). Detection of native peptides as potent inhibitors of enzymes. Crystal structure of the complex formed between treated bovine alpha-chymotrypsin and an autocatalytically produced fragment, Ile-Val-Asn-Gly-Glu-Glu-Ala-Val-Pro-Gly-Ser-Trp-Pro-Trp, at 2.2 Å resolution. *FEBS J.* **272**, 562-572.

32. Harel, M., Su, C. T., Frolow, F., Silman, I. & Sussman, J. L. (1991). Gamma-chymotrypsin is a complex of alpha-chymotrypsin with its own autolysis products. *Biochemistry* **30**, 5217-5225.

33. Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). Trypsin revisited: crystallography at (sub)atomic resolution and quantum chemistry revealing details of catalysis. *J. Biol. Chem.* **278**, 43357-43362.

34. Gertler, A. (1974). Inhibition of *Streptomyces griseus* protease B by peptide chloromethyl ketones: partial mapping of the binding site and identification of the reactive residue. *FEBS Lett.* **43**, 81-85.

35. Bauer, C. A. (1978). Active centers of *Streptomyces griseus* protease 1, *Streptomyces griseus* protease 3, and alpha-chymotrypsin: enzyme-substrate interactions. *Biochemistry* **17**, 375-380.

36. James, M. N. G., Brayer, G. D., Delbaere, L. T., Sielecki, A. R. & Gertler, A. (1980). Crystal structure studies and inhibition kinetics of tripeptide chloromethyl ketone inhibitors with *Streptomyces griseus* protease B. *J. Mol. Biol.* **139**, 423-438.

37. Laskowski, M., Jr. & 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**, 324-337.

38. Laskowski, M. J., Qasim, M. A. & Lu, S. M. (2000). Interaction of standard mechanism, canonical protein inhibitors with serine proteinases. In *Protein-protein recognition* (Kleanthous, C., ed.), pp. 228-279. Oxford University Press, New York.

39. Fuhrmann, C. N., Kelch, B. A., Ota, N. & Agard, D. A. (2004). The 0.83 Å resolution crystal structure of alpha-lytic protease reveals the detailed structure of the active site and identifies a source of conformational strain. *J. Mol. Biol.* **338**, 999-1013.

40. Fuhrmann, C. N., Daugherty, M. D. & Agard, D. A. (2006). Subangstrom crystallography reveals that short ionic hydrogen bonds, and not a His-Asp low-barrier hydrogen bond, stabilize the transition state in serine protease catalysis. *J. Am. Chem. Soc.* **128**, 9086-9102.

41. Radisky, E. S. & Koshland, D. E., Jr. (2002). A clogged gutter mechanism for protease inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10316-10321.

42. Radisky, E. S., Lee, J. M., Lu, C. J. & Koshland, D. E., Jr. (2006). Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6835-6840.

43. Ash, E. L., Sudmeier, J. L., Day, R. M., Vincent, M., Torchilin, E. V., Haddad, K. C., Bradshaw, E. M., Sanford, D. G. & Bachovchin, W. W. (2000). Unusual 1H NMR chemical shifts support (His) C(epsilon)1...O==C H-bond: proposal for reaction-driven ring flip mechanism in serine protease catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10371-10376.

44. Wilmouth, R. C., Edman, K., Neutze, R., Wright, P. A., Clifton, I. J., Schneider, T. R., Schofield, C. J. & Hajdu, J. (2001). X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. *Nat. Struct. Biol.* **8**, 689-694.

45. Liu, B., Schofield, C. J. & Wilmouth, R. C. (2006). Structural analyses on intermediates in serine protease catalysis. *J. Biol. Chem.* **281**, 24024-24035.

46. Frey, P. A., Whitt, S. A. & Tobin, J. B. (1994). A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* **264**, 1927-1930.

47. Krem, M. M. & Di Cera, E. (2001). Molecular markers of serine protease evolution. *EMBO J.* **20**, 3036-3045.

48. Derewenda, Z. S., Derewenda, U. & Kobos, P. M. (1994). (His) C(epsilon)-H...O=C < hydrogen bond in the active sites of serine hydrolases. *J. Mol. Biol.* **241**, 83-93.

49. Sidhu, S. S. & Borgford, T. J. (1996). Selection of *Streptomyces griseus* protease B mutants with desired alterations in primary specificity using a library screening strategy. *J. Mol. Biol.* **257**, 233-245.

50. Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501-4523.

51. Schmidt, A. & Lamzin, V. S. (2002). Veni, vidi, vici - atomic resolution unravelling the mysteries of protein function. *Curr. Opin. Struct. Biol.* **12**, 698-703.

52. A. T. Brünger, Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355 (1992) 472 - 475.

Chapter 5

Structural insights into the non-additivity effects in the sequence-to-reactivity algorithm for serine peptidases and their inhibitors

5.1 Overview

This chapter reports the molecular structures of the unbound OMTKY3 and its unbound Ala32I variant (OMTKY3-Ala32I), both in space group $P2_1$ and at the highest resolutions their crystals can afford (1.2 Å), and of the SGPB:OMTKY3-Ala32I complex in space group $P2_12_12_1$ at a resolution of 1.7 Å. Detailed comparisons are made for OMTKY3 in different conditions (different variants, unbound or bound to different serine peptidases), with respect to their overall structures, their structural rigidity, the main-chain conformations of their reactive-site loops and the interactions among the residues of their reactive-site loops. The results confirm our understanding of some previously addressed non-additivity effects in the SRA. The geometry of the interactions between the reactive site of OMTKY3 and the active sites of serine peptidases is also analyzed and compared with that typically observed in the catalytic intermediates of serine peptidases. Structural effects of the Gly-to-Ala substitution of residue 32I of OMTKY3 are examined in its unbound and SGPB-bound states. This substitution does not change the main-chain conformational angles and freedom of residue 32I of OMTKY3, but it does change the stoichiometry and the conformation of the SGPB:OMTKY3 complex. The detailed analysis for

identifying the causes of these changes is presented. The effects of these changes, particularly on the validity of the SRA, are discussed.

5.2 Materials and methods

Preparation and crystallization of the proteins

OMTKY3 and OMTKY3-Ala32I were cloned, overexpressed and purified by the group of the late Michael Laskowski, Jr. in the Department of Chemistry, Purdue University using the methods described previously.[1] Diffraction-quality crystals of both OMTKY3 and its Ala32I variant were grown by the hanging-drop vapor diffusion method at ambient temperature in 2 to 3 days, from a drop containing equal amounts of the protein solution (approximately 10 mg/mL in water) and the reservoir solution (Condition number 41 of the Hampton Research Crystal Screen: 0.1 M HEPES-NaOH (pH 7.5), 10 % (v/v) isopropanol and 20 % (w/v) polyethylene glycol 4000). SGPB, purified from Pronase as described previously,[2] was a generous gift from L. B. Smillie in the Department of Biochemistry, University of Alberta. An approximately 20 mg/mL solution of the SGPB:OMTKY3-Ala32I complex was prepared by mixing SGPB and OMTKY3-Ala32I in water in a 1:1.5 molar ratio. To remove the excess OMTKY3-Ala32I molecules, the complex solution was ultra-filtered at a molecular-mass cutoff of 10,000. Clusters of needle-shaped crystals of the complex were grown by the hanging-drop vapor diffusion method at ambient temperature within one week. The drop of the complex contained equal amounts of the complex solution and the reservoir solution (0.2 M sodium acetate trihydrate, 0.1 M Tris hydrochloride (pH 8.5) and 26% (w/v) polyethylene glycol 4000). A diffraction-quality fragment was selected from one of the crystal clusters. For cryo-

protection, all of the crystals were soaked for approximately 10 seconds in their respective reservoir solutions with 40 % (v/v) glycerol included, and then immediately flash-cooled in the liquid nitrogen for storage and subsequent shipment to the synchrotron beamline.

Data collection and processing, and structure solution and refinement

The X-ray diffraction data from all crystals were collected at the synchrotron Beamline 8.3.1 (equipped with an ADSC-Q210 CCD detector) at the Advanced Light Source in the Lawrence Berkeley National Laboratory. All data sets were indexed, scaled and merged using DENZO and SCALEPACK.[3] Structure solution and refinement were done in CCP4.[4,5] All structures were solved by the molecular replacement method, with search models provided by the molecular structure of the SGPB:OMTKY3 complex (PDB accession code: 1SGR).[6] All structures were solved using MOLREP,[7] iteratively refined using REFMAC,[8] and manually adjusted when needed using XtalView/Xfit.[9] At the end, the structures of the unbound OMTKY3 and its unbound Ala32I variant were refined anisotropically with riding hydrogen atoms included. The stereochemical qualities of the final structures were assessed using PROCHECK.[10]

Structure analysis

Graphical representations of the molecular structures were prepared using PyMOL (http://www.pymol.org/). Peptide planes and their tetrahedral distortions were calculated using GEOMCALC.[11] The surface areas of molecular structures were calculated using NACCESS.[12] The interactions and the shape complementarities of protein-protein interfaces were analyzed using LIGPLOT[13] and Sc,[14] respectively.

Superpositions of molecular structures were done using ALIGN,[15,16] based on the main-chain atoms (amide N, C^{α}, and carbonyl C and O). The representative of the solution-structure ensemble of wtOMTKY3 was chosen using OLDERADO.[17] The diffraction-component precision index (DPI) was used as a measure of the overall positional uncertainty (σ_r) of a molecular structure.[18]

Protein Data Bank Accession Codes

The atomic coordinates and the structure factors of all structures have been deposited in the Protein Data Bank. The accession code is 2GKR for the structure of the unbound OMTKY3, 2GKT for the structure of the unbound OMTKY3-Ala32I, and 2GKV for the structure of the SGPB:OMTKY3-Ala32I complex.

5.3 Results and discussion

Structure determination

The parameters and statistics derived from X-ray diffraction data processing and structure refinement are summarized in Table 5-1. Similar to the previously reported OMTKY3-$\psi[CH_2NH_2^+]$Asp19I,[19] both OMTKY3 and OMTKY3-Ala32I crystallized in space group $P2_1$ with only one OMTKY3 molecule in each asymmetric unit. All residues of the inhibitor (residues 6I to 56I) were clearly discernable in the electron density maps. Most atoms could be located precisely in the $2|F_o|-|F_c|,\alpha_c$ maps contoured at the 3.0σ to 4.0σ level (Figure 5-1a), whereas some atoms of the N-terminal residues, Leu18I, Glu19I and Arg21I of the reactive-site loop, the surface residues, and the C-terminal residues could only be reliably located in the same maps contoured at the 1.0σ to 2.0σ level (Figure 5-1b). In the Ramachandran plot for the molecular structure of the unbound OMTKY3, 92.9 % of non-Gly and non-Pro

Table 5-1 Parameters and statistics derived from X-ray diffraction data processing and structure refinement

	OMTKY3	OMTKY3-Ala32I	SGPB:OMTKY3-Ala32I
Data processing			
Wavelength (Å)	0.954	0.954	0.954
Resolution limit[a] (Å)	25.59 – 1.16 (1.20 – 1.16)	25.59 – 1.23 (1.27 – 1.23)	35.83 – 1.70 (1.76 – 1.70)
Space group	$P2_1$	$P2_1$	$P2_12_12_1$
Unit-cell constants			
a (Å)	22.90	23.05	48.60
b (Å)	34.43	34.69	53.03
c (Å)	25.79	25.80	88.63
α (°)	90.00	90.00	90.00
β (°)	97.25	97.30	90.00
γ (°)	90.00	90.00	90.00
Mosaicity (°)	0.51	0.32	0.72
Number of unique reflections	13,772 (1,286)	11,526 (1,067)	25,510 (2,457)
Redundancy	3.1 (2.0)	3.2 (2.3)	3.6 (3.5)
Completeness (%)	97.7 (91.3)	97.7 (91.7)	98.6 (96.6)
R_{sym}[b] (%)	3.9 (30.3)	3.6 (14.4)	4.1 (50.4)
$< I / \sigma(I) >$	21.7 (3.7)	23.8 (7.0)	24.7 (3.3)

[a] Numbers in the parentheses refer to the highest resolution bins.

[b] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_{hkl,i} - <I_{hkl}>| / \Sigma_{hkl}\Sigma_i I_{hkl,i}$, where $I_{hkl,i}$ and $<I_{hkl}>$ are the i-th observed intensity and the average intensity of reflection hkl, respectively.

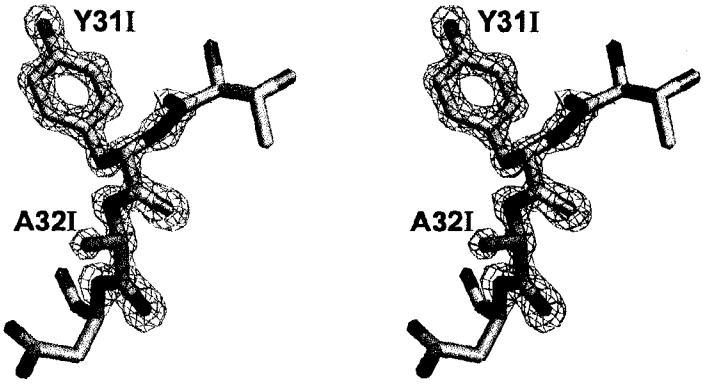
Table 5-1 (continued)

	OMTKY3	OMTKY3-Ala32I	SGPB:OMTKY3-Ala32I
Structure refinement			
Resolution range (Å)	25.59 – 1.16	25.59 – 1.23	35.83 – 1.70
R_{work}[c] (%)	14.7	13.0	21.4
R_{free}[c] (%)	17.7	15.1	24.8
R[c] (%)	15.0	13.2	Not applicable
Number of non-hydrogen atoms per asymmetric unit (average *B* factor, Å^2)			
Protein	Chain I: 392 (15.67)	Chain I: 393 (16.03)	Chain E: 1,311 (31.05) Chain A: 390 (31.59) Chain B: 388 (32.30)
Solvent	64 (32.63)	64 (35.07)	111 (43.66)
rms deviation from ideal geometry			
Bond lengths (Å)	0.018	0.021	0.018
Bond angles (°)	1.672	1.895	1.763
Ramachandran plot			
Favored (%)	92.9	95.3	88.4
Allowed (%)	7.1	4.7	11.2
Generously allowed (%)	0.0	0.0	0.0
Disallowed (%)	0.0	0.0	0.4

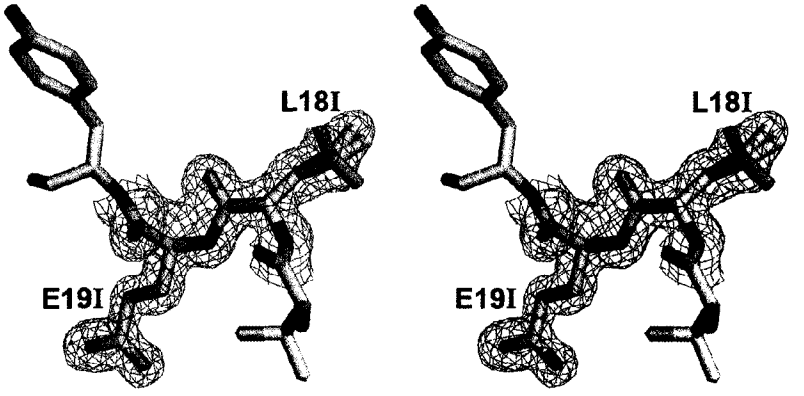
[c] $R_{work} = \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and the calculated structure factor amplitudes of a particular reflection, respectively. For the unbound OMTKY3 and the unbound OMTKY3-Ala32I, the summation is over 90 % of the reflections in the specified resolution range. The remaining 10 % of the reflections were randomly selected before the structure refinement and not included in the structure refinement. R_{free} was calculated over these reflections using the same equation as for R_{work}.[44] R was calculated over all the reflections in the specified resolution range. Similarly, for the SGPB:OMTKY3-Ala32I complex, 95 % of the reflections in the specified resolution range were used in the calculation of R_{work}, and the remaining 5 % of the reflections were used in the calculation of R_{free}.

Figure 5-1 Electron density map of the unbound OMTKY3-Ala32I. (a) $2|F_o|-|F_c|,\alpha_c$ map of Tyr31I and Ala32I contoured at 3.0σ. (b) $2|F_o|-|F_c|,\alpha_c$ map of Leu18I and Glu19I contoured at 1.0σ.

(a)



(b)



residues are in the most favored regions, and 7.1 % are in the additionally allowed regions. In the Ramachandran plot for the molecular structure of the unbound OMTKY3-Ala32I, 95.3 % of non-Gly and non-Pro residues are in the most favored regions, and 4.7 % are in the additionally allowed regions (Table 5-1).

Unexpectedly, the SGPB:OMTKY3-Ala32I complex crystallized in space group $P2_12_12_1$ with one SGPB molecule (molecule E) and two OMTKY3-Ala32I molecules (molecules A and B) in each asymmetric unit. All the residues of molecule E (residues 16E to 242E), molecule A (residues 6IA to 56IA) and molecule B (residues 6IB to 56IB) were identified in the electron density maps. Most atoms could be reliably located in the $2|F_o|-|F_c|,\alpha_c$ maps contoured at the 1.0σ to 2.0σ level. In the Ramachandran plot for the molecular structure of the SGPB:OMTKY3-Ala32I complex, 88.4 % of non-Gly and non-Pro residues are in the most favored regions, 11.2 % are in the additionally allowed regions, and 0.4 % (Asn100E) is in a disallowed region. With well-defined electron densities, Asn100E is on the loop propagating from strand *e* to strand *f* of β-barrel 1 of SGPB. In the crystal, the main-chain conformation of Asn100E seems to be sterically restricted by its neighboring residues on the *e-f* loop and by Val176E to Tyr178E on strand *i* of the same SGPB molecule (say, from the asymmetric unit x, y, z), and the side chain of Trp67E and the main chain of Arg81E of the SGPB molecule from the asymmetric unit 0.5 + x, 0.5 – y, -z. Asn100E $N^{\delta 2}$ of the SGPB molecule from the asymmetric unit x, y, z forms a hydrogen bond with the carbonyl O of Arg81E of the SGPB molecule from the asymmetric unit 0.5 + x, 0.5 – y, -z (3.5 Å).

Molecular structure of the unbound OMTKY3

The overall structure of the unbound OMTKY3 has the form of a wedge-shaped disc, comprising an extended loop (residues Ala15I to Arg21I) that contains the reactive site (the Leu18I–Glu19I peptide bond), a small triple-stranded antiparallel β-sheet (residues Leu23I to Ser26I, Asp27I to Tyr31I and Ser51I to His52I) with a type I β-turn (residues Gly25I to Asn28I) and an α-helix (residues Asn33I to Asn45I) positioned between the reactive-site loop and the β-sheet. OMTKY3 is structurally rigidified by three disulfide bonds (Cys8I–Cys38I, Cys16I–Cys35I and Cys24I–Cys56I), especially by the first two that anchor the reactive-site loop to the core of the inhibitor.

The molecular structure of the unbound ψ[$CH_2NH_2^+$]Asp19I variant of OMTKY3 has been determined previously in the same space group and unit-cell dimensions as those of the unbound OMTKY3 reported here.[19] In OMTKY3-ψ[$CH_2NH_2^+$]Asp19I, the reactive-site carbonyl group of OMTKY3 is reduced to a methylene group, and Glu19I of OMTKY3 is changed to Asp19I. Nonetheless, superposition of the molecular structures of OMTKY3 and OMTKY3-ψ[$CH_2NH_2^+$]Asp19I shows excellent agreement in atomic positions (Table 5-2). The molecular structures of the unbound wtOMJPQ3[20,21] and the unbound wtOMSVP3[22] have been determined previously in different space groups and unit-cell dimensions from those of the unbound OMTKY3 reported here. The amino-acid sequences of OMJPQ3 (wtOMJPQ3 without its first five N-terminal residues) and OMTKY3 differ at residues 17I, 18I, 19I, 23I, 32I and 51I, whereas those of OMSVP3 (wtOMSVP3 without its first five N-terminal residues) and OMTKY3 differ at residue 18I only.

Table 5-2 Root-mean-square differences (rmsd; in Å) for superpositions of the molecular structure of the unbound OMTKY3 with those of ovomucoid third domains in other conditions[a]

Reference	Enzyme	Inhibitor	ID	rmsd (Å)[b]
Unbound inhibitors				
19	-	OMTKY3-ψ[$CH_2NH_2^+$]Asp19I	I	0.27 (192)
20,21	-	OMJPQ3	A	0.39 (186)
			B	0.51 (194)
			C	0.44 (188)
			D	0.46 (187)
22	-	OMSVP3	-	0.41 (195)
23	-	wtOMTKY3 (solution structure)	-	1.46 (204)
c	**-**	**OMTKY3-Ala32I**	**I**	0.13 (194)
Inhibitors bound to serine peptidases				
24,25	SGPB	wtOMTKY3	I	0.46 (200)
6	SGPB	OMTKY3	I	0.46 (196)
32	HLE	wtOMTKY3	I	0.43 (198)
29	CHYM	wtOMTKY3	I	0.71 (202)
30	CARL	wtOMTKY3	I	0.59 (201)
31	CARL	OMTKY3	C	0.48 (186)
			D	0.85 (198)

[a] The molecular structure of the unbound OMTKY3 has also been compared with those of the SGPB-bound residue-18I variants of OMTKY3 determined so far (rmsd: 0.45 to 0.53 Å for 192 to 200 main-chain atoms) and with that of the CHYM-bound OMTKY3-Lys18I (rmsd: 0.61 Å for 203 main-chain atoms).

Table 5-2 (continued)

[b] All superpositions started with residues 6 to 56 of each inhibitor (total number of main-chain atoms: 204) included in the calculations. The number of main-chain atoms included in the final calculation of rmsd for each superposition is given in the parentheses.

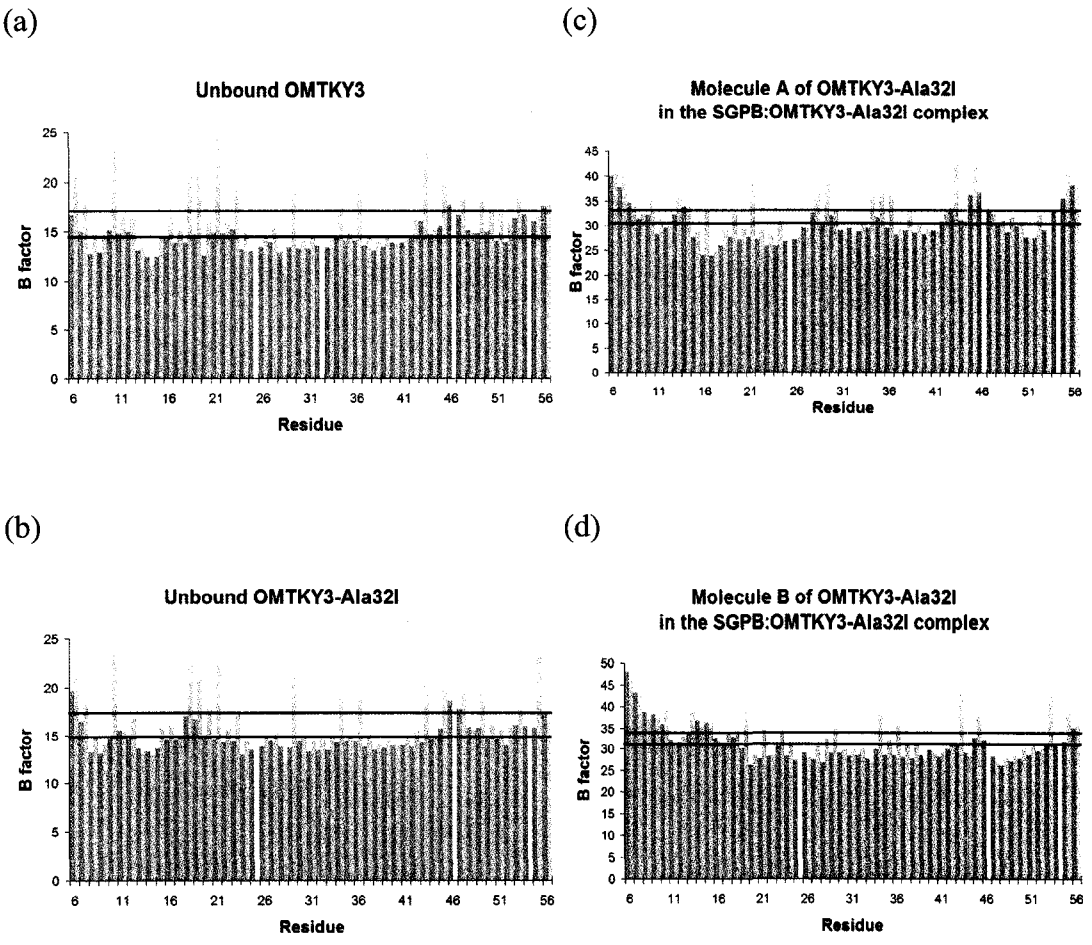
[c] This molecular structure is reported here.

Superpositions of the molecular structures of the unbound OMTKY3 with those of the unbound wtOMJPQ3 and the unbound wtOMSVP3 show slightly less good agreement in atomic positions (Table 5-2).

The molecular structure of the unbound OMTKY3 has also been compared with the representative solution structure of the unbound wtOMTKY3[23] (Table 5-2). Substantial differences in atomic positions occur in residues Val6I to Pro14I, the reactive-site loop, the type I β-turn, residues Asn45I to Thr47I, Leu50I, Lys55I and Cys56I. The extensive conformational flexibilities of these regions are shown by the comparisons of the 50 members of the solution-structure ensemble of the unbound wtOMTKY3. The backbone root-mean-square differences (rmsd) for these regions are greater than that for the whole structure (0.86 Å). Modeling of the representative solution structure of the unbound wtOMTKY3 in the unit cell defined by the unbound OMTKY3 (not shown) suggests that the conformational differences in most of these regions are caused by the crystal packing of the unbound OMTKY3. In the molecular structure of the unbound OMTKY3, the main-chain and the side-chain B factors of many of the residues in these regions are above the average values. Other residues in these regions have lower B factors, probably because of the dampening effects of crystal packing on their motions (Figure 5-2a).

The molecular structure of the unbound OMTKY3 reported here is considered the most relevant model of the unbound ovomucoid third domain for the SRA-related structural studies, because the SRA is based on OMTKY3 and most of the molecular structures previously determined for the development of the SRA contain an OMTKY3 variant. Previously, the molecular structure of the unbound wtOMSVP3 at

Figure 5-2 The B factors (in $Å^2$) of the residues of the inhibitor in the molecular structures reported here. (a) Unbound OMTKY3. (b) Unbound OMTKY3-Ala32I. (c) Molecule A and (d) molecule B of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex. The main-chain and the side-chain B factors are represented by the purple and the orange bars, respectively. The average values of the main-chain and the side-chain B factors are indicated by the blue and the red lines, respectively. There are no orange bars for Gly25I, Gly32I (in the unbound OMTKY3 only), Gly46I or Gly54I, because these residues do not have side chains.



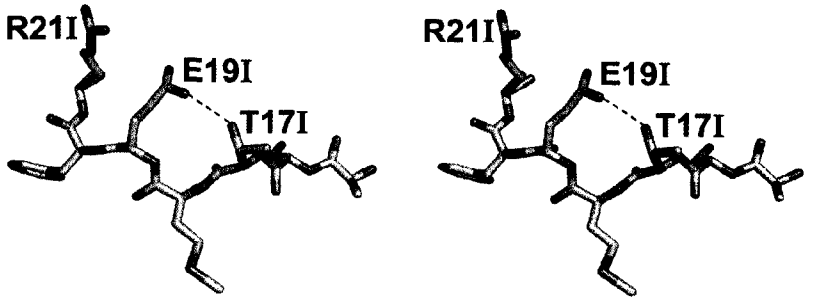
a lower resolution of 1.5 Å served as a model of the unbound ovomucoid third domain. OMTKY3 and OMSVP3 differ only in the identity of residue 18I (Leu in the former; Met in the latter). All of the molecular structures of the unbound avian ovomucoid third domains determined so far show that the side chain of residue 18I is exposed to solvent.

Interactions among Thr17I, Glu19I and Arg21I of OMTKY3

The molecular structure of the unbound wtOMSVP3 shows a direct hydrogen bond (3.0 Å) between Thr17I $O^{\gamma 1}$ and Glu19I $O^{\varepsilon 1}$ (Figure 5-3a).[22] In the molecular structure of the unbound OMTKY3, a hydrogen bond mediated by a water molecule with its oxygen 2.7 Å from Thr17I $O^{\gamma 1}$ and 2.8 Å from Glu19I $O^{\varepsilon 1}$ is observed (Figure 5-3b). Direct hydrogen bonds of shorter lengths (mean: 2.6 Å, standard deviation: 0.1 Å) between Thr17I $O^{\gamma 1}$ and Glu19I $O^{\varepsilon 1}$ are observed in the molecular structures of all the complexes of OMTKY3 (including the wild type and the residue-18I variants, except the Pro18I variant) with SGPB[6,24,25,26,27,28] and with CHYM[29] (also Ding *et al.*, unpublished results; PDB accession code: 1HJA) determined so far (Figure 5-3c). The decrease in the length of this hydrogen bond by approximately 0.4 Å indicates an increase in the strength of this hydrogen bond upon the binding of OMTKY3 to SGPB and CHYM. However, this hydrogen bond is not observed in the molecular structure of the CARL:wtOMTKY3 complex,[30] nor in that of the CARL:OMTKY3 complex.[31] Additionally, in the unbound OMTKY3, the main-chain NH of Glu19I forms a hydrogen bond (3.3 Å) with Thr17I $O^{\gamma 1}$. This hydrogen bond is observed in the CARL:wtOMTKY3 (3.6 Å) and the CARL:OMTKY3 (molecule C: 3.8 Å; molecule D: 3.6 Å) complexes as well. However, in the complexes of OMTKY3

Figure 5-3 Interactions among Thr17I, Glu19I and Arg21I of ovomucoid third domains in different conditions. (a) Unbound wtOMSVP3.[22] (b) Unbound OMTKY3. (c) SGPB-bound OMTKY3.[6] (d) Unbound OMTKY3-Ala32I. (e) Molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex. Hydrogen bonds and ionic interactions are indicated by dotted lines. Water molecules are indicated by w.

(a)



(b)

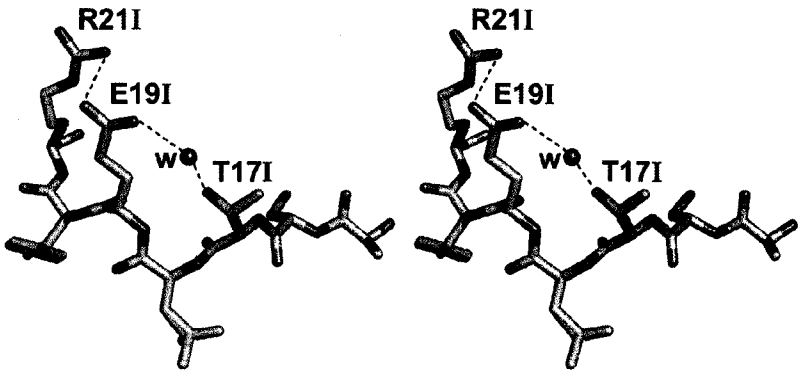
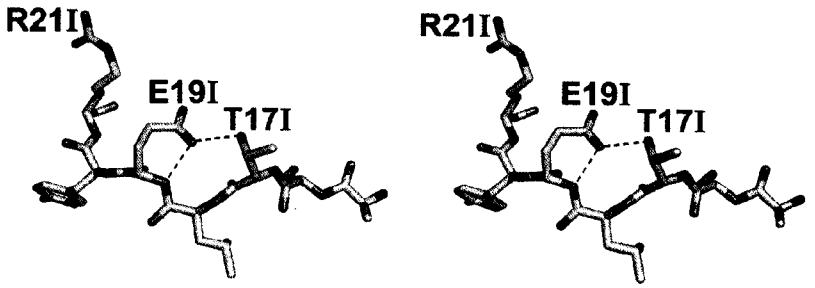
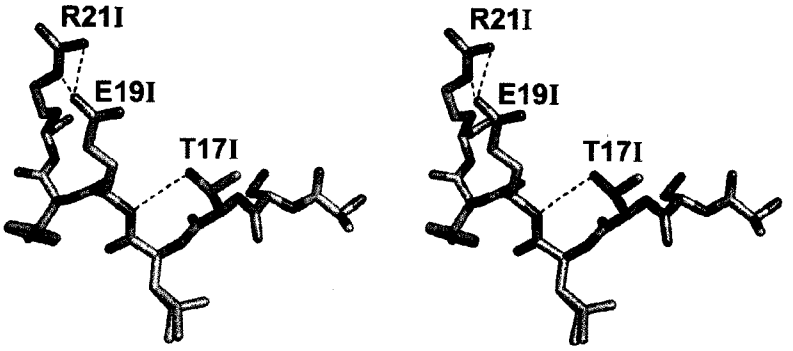


Figure 5-3 (continued)

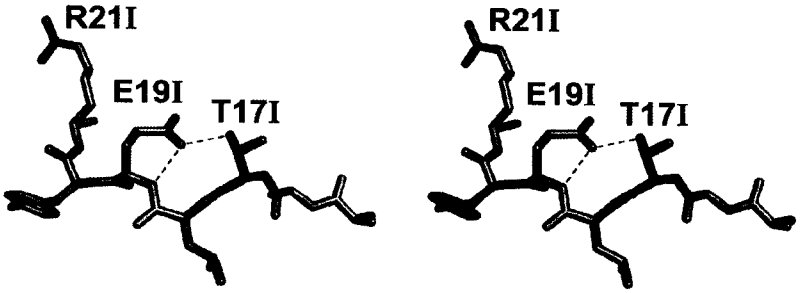
(c)



(d)



(e)



(including the wild type and the residue-18I variants) with SGPB, with CHYM and with HLE, the main-chain NH of Glu19I forms a hydrogen bond with Glu19I $O^{\varepsilon 1}$ (mean distance: 2.7 Å, standard deviation: 0.1 Å) instead. All of these differences in the interactions between Thr17I and Glu19I upon the binding of OMTKY3 to the various serine peptidases result in the non-additivity effects between residues 17I and 19I in the SRA-based predictions of the K_{assoc} values of OMTKY3 variants with these serine peptidases.

In the molecular structure of the unbound OMTKY3, ionic interactions occur between the side chains of Glu19I and Arg21I, with Glu19I $O^{\varepsilon 2}$ 3.3 Å and 3.2 Å away from Arg21I N^{ε} and $N^{\eta 1}$, respectively. These ionic interactions are not observed in the molecular structure of wtOMSVP3,[22] nor in any of the molecular structures of the complexes of OMTKY3 (including the wild type and the residue-18I variants) with SGPB,[6,19,24,25,26,27,28] with CHYM[29] (also Ding *et al.*, unpublished results; PDB accession code: 1HJA) and with HLE.[32] However, these ionic interactions are observed in the CARL:wtOMTKY3 complex,[30] and in molecule D of OMTKY3 in the CARL:OMTKY3 complex.[31] This raises the possibility of non-additivity effects between residues 19I and 21I in the SRA-based predictions of the K_{assoc} values of OMTKY3 variants with CARL.

Conformation of the reactive-site loop of OMTKY3

The rigidity of the overall structure of OMTKY3 suggests a lack of conformational change upon the binding of OMTKY3 to serine peptidases. In general, superpositions of the molecular structure of the unbound OMTKY3 with those of OMTKY3 (including the wild type and the residue-18I variants) bound to

SGPB, CHYM and HLE determined so far show excellent agreement in atomic positions (Table 5-2). The spread of the rmsd values for these superpositions (0.43 Å to 0.71 Å for 196 to 202 out of 204 main-chain atoms) indicates the difference in some local conformational changes required for the binding of OMTKY3 to different serine peptidases, and the difference in the space groups and the unit-cell dimensions of the complexes of OMTKY3 with different serine peptidases. The molecular structure of the CARL:OMTKY3 complex reveals OMTKY3 molecules in two different conformations (molecules C and D in the asymmetric unit of the CARL:OMTKY3 complex).[31] Superpositions show that the conformation of the unbound OMTKY3 resembles that of molecule C more closely than that of molecule D (Table 5-2).

Canonical protein inhibitors share a common main-chain conformation for the residues of their reactive-site loops. This common conformation is only affected in local regions of the reactive site loop whether the inhibitor is unbound or bound to its target serine peptidase.[33,34] As shown in Table 5-3, the angles ϕ and ψ of the residues of the reactive-site loops of the unbound OMTKY3, wtOMJPQ3[20,21] and wtOMSVP3[22] agree within the positional uncertainties of these molecular structures. However, upon the binding of OMTKY3 to SGPB,[6,24,25] CHYM,[29] HLE[32] and CARL,[30,31] adjustments in the angles ϕ and ψ of residues Ala15I to Glu19I of OMTKY3 are observed. The angles ϕ and ψ of Ala15I of OMTKY3 change significantly upon the binding of OMTKY3 to CARL. In fact, these two angles also change, though inconsistently, upon the binding of OMTKY3 to SGPB, CHYM and HLE. Some researchers have questioned the inclusion of Ala15I in the definition of

Table 5-3 Angles ϕ and ψ (in degrees) of the residues of the reactive-site loops of ovomucoid third domains in different conditions

Reference	Enzyme	Inhibitor	σ_r (Å)[a]	ID	Residue 15	
					ϕ	ψ
Unbound inhibitors (including molecule B of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)						
b	**-**	**OMTKY3**	**0.072**	**I**	**-141.7**	**151.4**
b	**-**	**OMTKY3-Ala32I**	**0.076**	**I**	**-140.7**	**150.6**
b	**SGPB**	**OMTKY3-Ala32I**	**0.241**	**B**	**-124.5**	**152.8**
20,21	-	OMJPQ3	0.317	A	-128.6	144.6
				B	-162.0	159.3
				C	-150.2	156.4
				D	-146.4	160.0
22	-	OMSVP3	0.219	-	-149.1	163.2
Mean					-142.9	154.8
(Standard deviation)					(11.3)	(5.7)
Inhibitors bound to serine peptidases (including molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)						
b	**SGPB**	**OMTKY3-Ala32I**	**0.241**	**A**	**-118.7**	**137.1**
24,25	SGPB	wtOMTKY3	0.485	I	-158.3	158.2
6	SGPB	OMTKY3	0.190	I	-157.6	153.1
32	HLE	wtOMTKY3	0.287	I	-127.4	148.3
29	CHYM	wtOMTKY3	0.281	I	-128.9	135.9
30	CARL	wtOMTKY3	0.057	I	-100.6	140.4
31	CARL	OMTKY3	0.150	C	-90.6	126.8
				D	-90.5	126.7
Mean					-121.6	140.8
(Standard deviation)					(25.3)	(10.9)
Difference of the mean for inhibitors bound to serine peptidases from that for unbound inhibitors					21.3	-14.0

Table 5-3 (continued)

Enzyme	Inhibitor	σ_r (Å)[a]	ID	Residue 16		Residue 17	
				ϕ	ψ	ϕ	ψ
Unbound inhibitors (including molecule B of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)							
-	**OMTKY3**	**0.072**	**I**	**-121.7**	**147.4**	**-79.7**	**173.4**
-	**OMTKY3-Ala32I**	**0.076**	**I**	**-117.5**	**151.5**	**-78.2**	**169.9**
SGPB	**OMTKY3-Ala32I**	**0.241**	**B**	**-123.1**	**143.5**	**-66.8**	**156.6**
-	OMJPQ3	0.317	A	-122.4	156.1	-65.4	155.9
			B	-116.8	154.1	-61.8	160.0
			C	-123.1	153.5	-61.4	162.6
			D	-123.4	153.8	-65.8	166.5
-	OMSVP3	0.219	-	-130.5	155.0	-86.9	173.8
Mean				-122.3	151.9	-70.8	164.8
(Standard deviation)				(3.9)	(4.0)	(8.9)	(6.7)
Inhibitors bound to serine peptidases (including molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)							
SGPB	**OMTKY3-Ala32I**	**0.241**	**A**	**-129.3**	**150.7**	**-70.8**	**158.7**
SGPB	wtOMTKY3	0.485	I	-128.5	146.6	-69.0	159.2
SGPB	OMTKY3	0.190	I	-124.7	144.2	-65.4	162.1
HLE	wtOMTKY3	0.287	I	-118.6	143.1	-73.1	155.7
CHYM	wtOMTKY3	0.281	I	-130.7	149.8	-67.8	159.9
CARL	wtOMTKY3	0.057	I	-132.3	156.3	-61.8	160.6
CARL	OMTKY3	0.150	C	-127.4	148.9	-56.8	156.4
			D	-127.2	152.7	-59.0	160.3
Mean				-127.3	149.0	-65.5	159.1
(Standard deviation)				(3.9)	(4.1)	(5.4)	(2.0)
Difference of the mean for inhibitors bound to serine peptidases from that for unbound inhibitors				-5.0	-2.9	5.3	-5.7

Table 5-3 (continued)

Enzyme	Inhibitor	σ_r (Å)[a]	ID	Residue 18		Residue 19	
				ϕ	ψ	ϕ	ψ
Unbound inhibitors (including molecule B of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)							
-	**OMTKY3**	**0.072**	I	**-84.9**	**6.4**	**-62.5**	**121.7**
-	**OMTKY3-Ala32I**	**0.076**	I	**-87.4**	**3.2**	**-60.7**	**117.1**
SGPB	**OMTKY3-Ala32I**	**0.241**	B	**-88.6**	**-1.8**	**-54.7**	**124.3**
-	OMJPQ3	0.317	A	-100.0	5.8	-54.1	128.4
			B	-90.8	-5.3	-43.4	125.2
			C	-104.6	19.6	-60.2	139.3
			D	-106.8	32.9	-56.5	145.9
-	OMSVP3	0.219	-	-96.0	9.3	-58.2	139.4
Mean				-94.9	8.8	-56.3	130.2
(Standard deviation)				(7.7)	(11.5)	(5.6)	(9.5)
Inhibitors bound to serine peptidases (including molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)							
SGPB	**OMTKY3-Ala32I**	**0.241**	**A**	**-110.3**	**33.4**	**-73.1**	**152.3**
SGPB	wtOMTKY3	0.485	I	-116.4	45.6	-83.0	153.0
SGPB	OMTKY3	0.190	I	-118.5	37.2	-78.9	156.8
HLE	wtOMTKY3	0.287	I	-101.3	31.4	-84.0	154.9
CHYM	wtOMTKY3	0.281	I	-107.4	31.6	-73.9	159.4
CARL	wtOMTKY3	0.057	I	-105.8	36.2	-77.1	133.6
CARL	OMTKY3	0.150	C	-110.9	37.5	-71.6	139.1
			D	-109.1	37.0	-75.4	130.3
Mean				-110.0	36.2	-77.1	147.4
(Standard deviation)				(5.2)	(4.2)	(4.3)	(10.6)
Difference of the mean for inhibitors bound to serine peptidases from that for unbound inhibitors				-15.1	27.4	-20.8	17.2

Table 5-3 (continued)

Enzyme	Inhibitor	σ_r (Å)[a]	ID	Residue 20		Residue 21	
				ϕ	ψ	ϕ	ψ
Unbound inhibitors (including molecule B of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)							
-	**OMTKY3**	**0.072**	**I**	**-99.7**	**104.7**	**-137.8**	**80.7**
-	**OMTKY3-Ala32I**	**0.076**	**I**	**-92.0**	**100.8**	**-133.9**	**75.1**
SGPB	**OMTKY3-Ala32I**	**0.241**	**B**	**-91.3**	**102.8**	**-135.1**	**74.5**
-	OMJPQ3	0.317	A	-98.7	101.5	-137.1	71.4
			B	-91.7	118.1	-167.7	71.4
			C	-107.0	93.1	-132.7	71.2
			D	-105.7	95.7	-135.4	72.5
-	OMSVP3	0.219	-	-99.0	92.8	-129.5	69.0
Mean				-98.1	101.2	-138.7	73.2
(Standard deviation)				(5.8)	(7.6)	(11.2)	(3.4)
Inhibitors bound to serine peptidases (including molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)							
SGPB	**OMTKY3-Ala32I**	**0.241**	**A**	**-101.7**	**107.3**	**-136.6**	**88.7**
SGPB	wtOMTKY3	0.485	I	-98.7	114.8	-148.8	93.6
SGPB	OMTKY3	0.190	I	-108.2	115.5	-145.7	88.4
HLE	wtOMTKY3	0.287	I	-102.0	106.7	-138.4	82.4
CHYM	wtOMTKY3	0.281	I	-113.4	106.7	-141.7	75.6
CARL	wtOMTKY3	0.057	I	-95.5	105.9	-141.7	74.4
CARL	OMTKY3	0.150	C	-94.9	107.8	-138.9	78.0
			D	-88.5	102.6	-132.8	83.2
Mean				-100.4	108.4	-140.6	83.0
(Standard deviation)				(7.4)	(4.2)	(4.7)	(6.4)
Difference of the mean for inhibitors bound to serine peptidases from that for unbound inhibitors				-2.3	7.2	-1.9	9.8

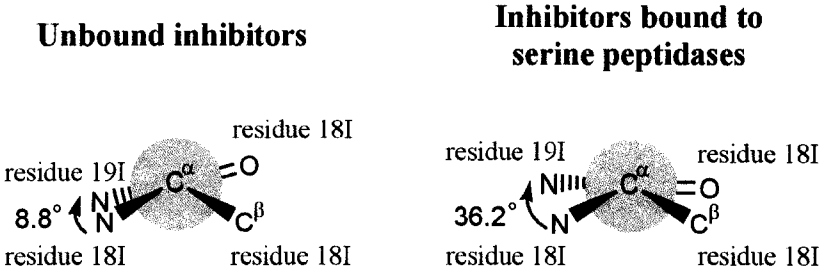
[a] σ_r, overall positional uncertainty.

[b] These are the molecular structures reported here.

the reactive-site loop of OMTKY3.[34] The angles ϕ and ψ (especially the latter) of Leu18I of OMTKY3 change quite consistently upon the binding of OMTKY3 to SGPB, CHYM, HLE and CARL (Table 5-3, Figure 5-4). These adjustments could be justified by the need for the insertion of the side chain of Leu18I of OMTKY3 into the S1 specificity pockets of these serine peptidases. They may be facilitated by the main-chain conformational adjustments in Thr17I and Glu19I of OMTKY3. The angle ψ of Thr17I and the angle ϕ of Glu19I of OMTKY3 change consistently upon the binding of OMTKY3 to SGPB, CHYM, HLE and CARL, whereas the angle ϕ of Thr17I and the angle ψ of Glu19I of OMTKY3 change consistently upon the binding of OMTKY3 to SGPB, CHYM and HLE only, but not to CARL (Table 5-3). This is probably relevant to the strengthening of the hydrogen bond between Thr17I $O^{\gamma 1}$ and Glu19I $O^{\varepsilon 1}$, and the switch in the hydrogen-bonding partner of the main-chain NH of Glu19I upon the binding of OMTKY3 to SGPB, CHYM and HLE, but not to CARL. The consistent changes in the angle ψ of Thr17I, the angles ϕ and ψ of Leu18I, and the angle ϕ of Glu19I upon the binding of OMTKY3 to the clan-PA serine peptidases SGPB, CHYM, and HLE, and to the clan-SB serine peptidase CARL indicate the convergent evolution of the active sites of serine peptidases in these two clans. The inconsistent changes in the angles ϕ and ψ further from the reactive site upon the binding of OMTKY3 to these serine peptidases indicate the structural differences in the rest of their substrate-binding regions, especially those between the clan-PA serine peptidases and the clan-SB ones (Table 5-3).

Interestingly, the molecular structures of the complexes of SGPB with all but one of the residue-18I variants of OMTKY3 determined so far[6,19,26,27,28] show that

Figure 5-4 Newman projections of the angle ψ of residue 18I of the unbound and the serine-peptidase bound ovomucoid third domains. The mean value of the angle is given along the arrow in each case.



substitutions of residue 18I do not affect the angles ϕ and ψ of Thr17I to Glu19I of the SGPB-bound OMTKY3 (values not shown). The only exception is the Leu-to-Pro substitution. The conformationally most restrictive Pro18I forces the angle ψ of Thr17I to 140.6 ° and the angle ϕ of Pro18I to -89.5 °. Probably because of the pull by the hydrogen bond between Thr17I $O^{\gamma 1}$ and Glu19I $O^{\varepsilon 1}$ (3.2 Å), the angle ψ of Glu19I comes to 141.6 °. The molecular structure of the CHYM:OMTKY3-Lys18I complex (Ding *et al.*, unpublished results; PDB accession code: 1HJA) also shows that the Leu-to-Lys substitution of residue 18I of OMTKY3 does not affect the angles ϕ and ψ of Thr17I to Glu19I of the CHYM-bound OMTKY3 (values not shown).

Interactions between the reactive site of OMTKY3 and the active sites of serine peptidases

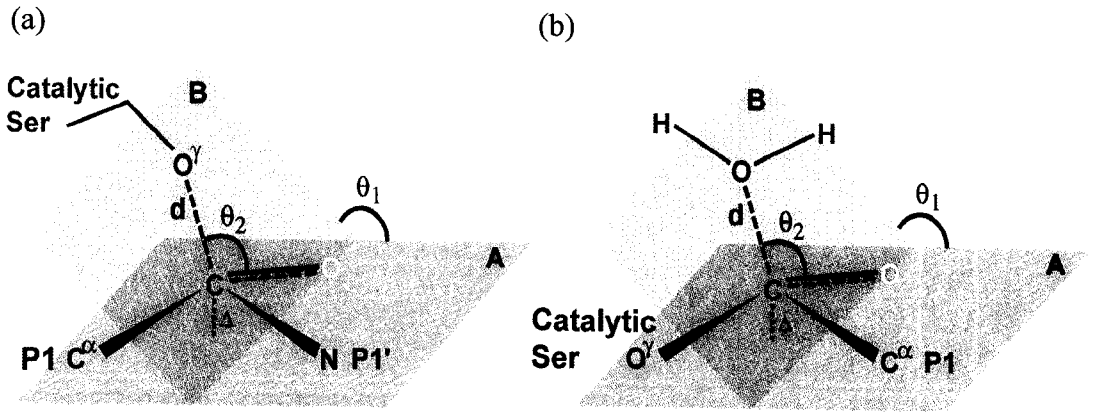
Previous studies showed that the dissociation of the complexes of both the virgin (i.e. the reactive site is intact) and the modified (i.e. the reactive site is hydrolyzed) wtOMTKY3 with eight serine peptidases, including CHYM, PPE, SGPA, SGPB and CARL, releases the virgin wtOMTKY3 predominantly (> 90 %).[35] This indicates that these serine peptidases catalyze both the hydrolysis and the transpeptidation of the reactive site of OMTKY3, even though the reactive site of OMTKY3 is intact in all of the molecular structures of OMTKY3 (including the wild type and the residue-18I variants) bound to SGPB,[6,19,24,25,26,27,28] CHYM,[29] (also Ding *et al.*, unpublished results; PDB accession code: 1HJA) HLE[32] and CARL[30,31] determined so far. Comparisons of the molecular structures of the modified wtOMJPQ3 and wtOMSVP3 with their virgin versions show that the hydrolysis of the reactive site increases the mobility of the reactive-site loop, but has little effect on

the interactions within the rest of the inhibitor.[36] The extensive interactions of OMTKY3 with SGPB, CHYM, HLE and CARL probably keep the reactive site of OMTKY3, no matter whether it is intact or hydrolyzed, oriented for the peptidase-catalyzed reactions in both directions.

Similar to previous studies,[37,38] geometric parameters are defined to describe the trajectory of the attack of the nucleophilic O (the catalytic Ser O^{γ} of the serine peptidase in the acylation stage, or the oxygen of the nucleophilic water molecule in the deacylation stage) on the reactive-site carbonyl C of OMTKY3 (Figures 5-5a and 5-5b). In all of the complexes of OMTKY3 and wtOMTKY3 with SGPB, CHYM, HLE and CARL, the catalytic triad of the serine peptidase is positioned for strong hydrogen bonding between the Ser O^{γ} and the His $N^{\varepsilon 2}$ (2.5 to 2.8 Å; Table 5-4), and between the His $N^{\delta 1}$ and the Asp $O^{\delta 2}$ (2.6 to 2.9 Å; Table 5-4) (Figure 5-5c). The parameter d is of the same order as the distance of a strong hydrogen bond (2.6 to 2.9 Å; Table 5-4), indicating some weak orbital interactions between the catalytic Ser O^{γ} of the peptidase and the reactive-site carbonyl C of the inhibitor. However, as indicated by the parameter Δ (maximum: 0.044 Å in the SGPB:wtOMTKY3 complex, with an overall positional uncertainty of 0.485 Å;[24,25] Table 5-4), a significant tetrahedral distortion of the reactive-site carbonyl C of OMTKY3, which could result from the change in its hybridization state from sp^2 to sp^3, is not observed. Both θ_1 (80.4 ° to 88.5 °; Table 5-4) and θ_2 (89.2 ° to 102.4 °; Table 5-4) are close to 90 °. All these distances and angles fall in the ranges of their corresponding values in other serine peptidase-protein inhibitor complexes,[37] and are also close to their corresponding values in the acylation of a Michaelis complex (with cucurbita pepo

Figure 5-5 Geometry of the reactive site of a protein inhibitor and the catalytic triad of a serine peptidase. Geometric parameters are defined to describe the trajectory of the attack on the reactive-site carbonyl C of the protein inhibitor. (a) The acylation stage, with the catalytic Ser O^{γ} of the serine peptidase as the nucleophile. Plane A (green) is defined by the atom C^{α} (green) and the carbonyl O (yellow) of residue P1, and the amide N of residue P1' of the inhibitor (green). Plane B (red) is defined by the catalytic Ser O^{γ} of the peptidase (red), and the carbonyl C (red) and O (yellow) of residue P1 of the inhibitor. d is the distance between the catalytic Ser O^{γ} of the peptidase and the carbonyl C of residue P1 of the inhibitor. Δ is the displacement of the carbonyl C of residue P1 of the inhibitor out of plane A. θ_1 is the angle between planes A and B. θ_2 is the angle O^{γ}(catalytic Ser of the peptidase)-C=O(residue P1 of the inhibitor). (b) The deacylation stage. Definitions are the same as in (a), except that the catalytic Ser O^{γ} of the peptidase is replaced by the oxygen of the nucleophilic water molecule (red), and that the amide N of residue P1' of the inhibitor is replaced by the catalytic Ser O^{γ} of the peptidase (green). (c) Catalytic triad of the serine peptidase (orange), and residues P3, P2 and P1 (with side chains R_3, R_2 and R_1, respectively) of the inhibitor (green). R and R' represent the N- and the C-terminal residues of the inhibitor, respectively. Hydrogen bonds are indicated by dashed lines. Partial charges are indicated by $\delta+$ and $\delta-$. The dipole across the P2-P1 peptide bond of the inhibitor is indicated by a straight arrow with a cross at its positive end. Electron flows in the acylation stage are indicated by curved arrows.

Figure 5-5 (continued)



(c)

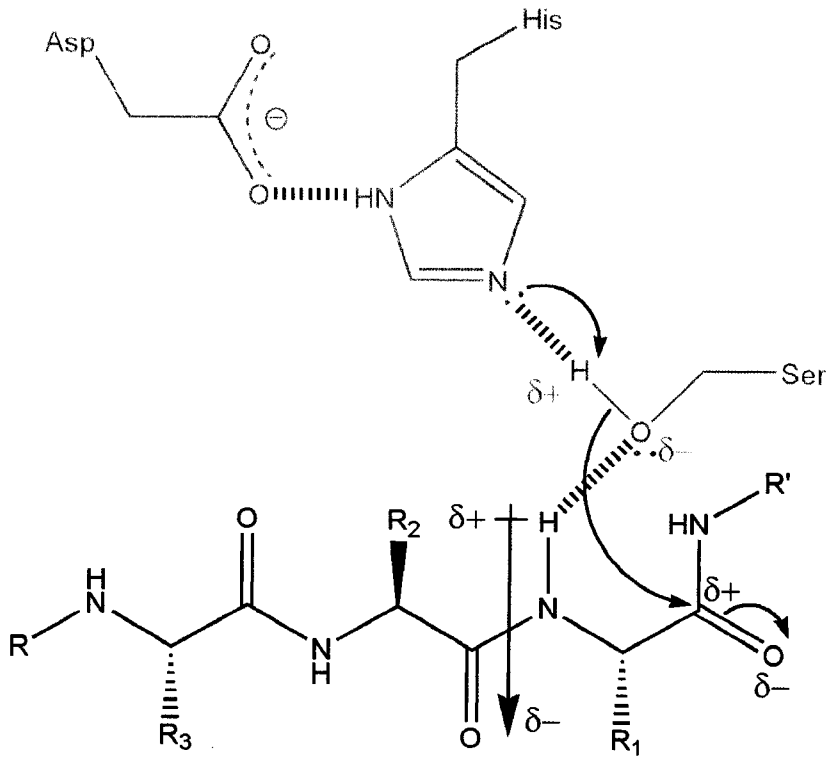


Table 5-4 Trajectory of the attack of the nucleophilic O of different origins on the reactive-site carbonyl C of ovomuvoid third domains in different conditions

Reference	Enzyme	Inhibitor / Substrate	σ_r (Å)[a]	ID	Δ (Å)[a,b]	d (Å)[a]	θ_1 (°)[a]	θ_2 (°)[a]
Unbound inhibitors (including molecule B of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)								
c	**-**	**OMTKY3**	**0.072**	**I**	**0.060**	**-**	**-**	**-**
c	**-**	**OMTKY3-Ala32I**	**0.076**	**I**	**0.025**	**-**	**-**	**-**
c	**SGPB**	**OMTKY3-Ala32I**	**0.241**	**B**	**0.041**	**-**	**-**	**-**
20,21	-	OMJPQ3	0.317	A	0.118	-	-	-
				B	0.054	-	-	-
				C	0.083	-	-	-
				D	0.059	-	-	-
22	-	OMSVP3	0.219	-	0.020	-	-	-
Inhibitors bound to serine peptidases (including molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)								
c	**SGPB**	**OMTKY3-Ala32I**	**0.241**	**A**	**-0.004**	**2.6**	**84.9**	**94.1**
24,25	SGPB	wtOMTKY3	0.485	I	+0.044	2.7	82.4	94.6
6	SGPB	OMTKY3	0.190	I	-0.018	2.8	82.2	92.0
32	HLE	wtOMTKY3	0.287	I	+0.037	2.8	82.9	99.1
29	CHYM	wtOMTKY3	0.281	I	+0.014	2.9	88.5	102.4
30	CARL	wtOMTKY3	0.057	I	-0.001	2.7	80.4	89.9
31	CARL	OMTKY3	0.150	C	-0.001	2.8	83.6	89.2
				D	+0.003	2.7	83.6	90.5
Michaelis complex, tetrahedral intermediate, acyl-enzyme intermediates of trypsin[d]								
39	trypsin	CPTI-II	0.095	I	+0.002	2.7	86.8	90.5
38	trypsin	leupeptin	0.040	A	+0.487	1.3	90.0	109.4
					(+0.405)	(1.4)	(89.8)	(100.3)
38	trypsin	AAPR	0.041	A	-	-	-	-
	(Water S25 as the nucleophile)				(+0.078)	(3.0)	(102.1)	(83.8)
38	trypsin	AAPK	0.061	A	-	-	-	-
	(Water S25 as the nucleophile)				(+0.183)	(2.5)	(94.2)	(96.1)
38	trypsin	GB	0.044	-	-	-	-	-
	(Water S25 as the nucleophile)				(+0.038)	(3.9)	(129.3)	(53.9)
	(Water S36 as the nucleophile)					(4.0)	(82.5)	(113.3)
	(Water S185 as the nucleophile)					(3.5)	(87.5)	(68.1)

Table 5-4 (continued)

Enzyme	Inhibitor / Substrate	σ_r (Å)[a]	ID	Nucleophilic O ... Catalytic His $N^{\varepsilon 2}$ (Å)[e]	Catalytic His $N^{\delta 1}$ ... Catalytic Asp $O^{\delta 2}$ (Å)	Catalytic Ser O^{γ} (peptidase) ... P1 amide N (inhibitor) (Å)
Inhibitors bound to serine peptidases (including molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex)						
SGPB	**OMTKY3-Ala32I**	**0.241**	**A**	**2.6**	**2.8**	**2.9**
SGPB	wtOMTKY3	0.485	I	2.5	2.8	2.8
SGPB	OMTKY3	0.190	I	2.6	2.9	2.9
HLE	wtOMTKY3	0.287	I	2.5	2.6	3.0
CHYM	wtOMTKY3	0.281	I	2.5	2.8	3.0
CARL	wtOMTKY3	0.057	I	2.7	2.6	2.9
CARL	OMTKY3	0.150	C	2.8	2.6	3.0
			D	2.7	2.7	2.9
Michaelis complex, tetrahedral intermediate, acyl-enzyme intermediates of trypsin[d]						
trypsin	CPTI-II	0.095	I	2.6	2.6	3.0
trypsin	leupeptin	0.040	A	3.0 (2.7)	2.8	2.9
trypsin (Water S25 as the nucleophile)	AAPR			3.0 (2.9)	2.8	2.8
trypsin (Water S25 as the nucleophile)	AAPK			3.0 (2.7)	2.8	2.8
trypsin (Water S25 as the nucleophile) (Water S36 as the nucleophile) (Water S187 as the nucleophile)	GB			3.8 (2.8) (3.6) (3.4)	2.8	---

[a] σ_r, overall positional uncertainty. For Δ, d, θ_1 and θ_2, values not in parentheses refer to the acylation stage (definitions in Figure 5-5a), whereas values in parentheses refer to the deacylation stage (definitions in Figure 5-5b).

Table 5-4 (continued)

[b] For unbound inhibitors and molecule B of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex, the absolute values of Δ are shown. For inhibitors bound to serine peptidases, a positive value of Δ indicates that both the reactive-site carbonyl C of the inhibitor and the nucleophilic O are on the same side of plane A (as defined in Figures 5-5a and 5-5b), whereas a negative value of Δ indicates that the reactive-site carbonyl C of the inhibitor and the nucleophilic O are on the opposite sides of plane A.

[c] These are the molecular structures reported here.

[d] The trypsin:CPTI-II complex represents a Michaelis complex of trypsin. The trypsin-leupeptin complex represents a tetrahedral intermediate of trypsin. The trypsin-AAPR and the trypsin-AAPK complexes represent good acyl-enzyme intermediates of trypsin. The guanidinobenzoyl (GB)-trypsin complex represents a bad acyl-enzyme intermediate of trypsin.

[e] For the hydrogen-bond distance between the catalytic His $N^{\varepsilon 2}$ of the peptidase and the nucleophilic O, a value not in parentheses refers to the catalytic Ser O^{γ} of the peptidase as the nucleophilic O in the acylation stage, whereas a value in parentheses refers to the oxygen of the nucleophilic water molecule in the deacylation stage.

trypsin inhibitor II, CPTI-II;[39] Table 5-4) and in the deacylation of two good acyl-enzyme intermediates (one with succinyl-Ala-Ala-Pro-Arg-*p*-nitroaniline, AAPR, and one with succinyl-Ala-Ala-Pro-Lys-*p*-nitroaniline, AAPK;[38] Table 5-4) of trypsin.

It is noteworthy that, in all of the complexes of OMTKY3 and wtOMTKY3 with SGPB, CHYM, HLE and CARL, and in the Michaelis complex, the tetrahedral intermediate (the trypsin-leupeptin complex)[38] and the good acyl-enzyme intermediates of trypsin, the catalytic Ser O^{γ} of the peptidase also forms a hydrogen bond (2.8 to 3.0 Å; Table 5-4) with the main-chain NH of residue P1 of the bound species. The catalytic Ser O^{γ} of the peptidase is positioned near the positive end of the dipole across the P2-P1 peptide bond of the bound species. Thus, the partial negative charge developed at the catalytic Ser O^{γ} upon its deprotonation by the catalytic His $N^{\varepsilon 2}$ within the peptidase is stabilized, and the nucleophilicity of the catalytic Ser O^{γ} for the reactive-site carbonyl C in the acylation stage is enhanced (Figure 5-5c). This serves as an example of substrate-assisted catalysis. Similar enhancement in nucleophilicity does not seem to happen to the oxygen of the nucleophilic water molecule in the deacylation stage.

Structural effects of the Gly-to-Ala substitution of residue 32I of the unbound OMTKY3

The overall structure of the unbound OMTKY3-Ala32I is virtually identical to that of the unbound OMTKY3. Superposition of the molecular structure of the unbound OMTKY3 with that of its unbound Ala32I variant (Table 5-2) shows excellent agreement in most atomic positions. The distributions of the main-chain and the side-chain B factors for all of the residues of the unbound OMTKY3 (in particular

the main-chain B factor of residue 32I) remain essentially unchanged upon the Gly-to-Ala substitution (Figures 5-2a and 5-2b). The unbound OMTKY3 and its unbound Ala32I variant agree well in the main-chain conformation of the reactive-site loop (Table 5-3) and of residue 32I (angles ϕ, ψ: -72.3 °, -26.8 ° for Gly, -68.4 °, -30.8 ° for Ala). The Gly-to-Ala substitution does not influence the spatial relationships among Leu18I, Tyr20I, residue 32I and Asn36I of the unbound OMTKY3. In the unbound OMTKY3-Ala32I, neither a direct nor a water-mediated hydrogen bond occurs between the side chains of Thr17I and Glu19I. Only a long hydrogen bond between Thr17I $O^{\gamma 1}$ and the main-chain NH of Glu19I (3.5 Å) is observed. Ionic interactions occur between Glu19I and Arg21I, with Arg21I N^{ε} and $N^{\eta 1}$ being 3.2 Å and 3.1 Å away from Glu19 $O^{\varepsilon 2}$, respectively (Figure 5-3d).

Molecular structure of the SGPB:OMTKY3-Ala32I complex

The asymmetric unit of the SGPB:OMTKY3-Ala32I complex includes one SGPB molecule (molecule E) and two OMTKY3-Ala32I molecules (molecules A and B) (Figure 5-6a), in contrast to those for the complexes of SGPB with other OMTKY3 variants determined so far that include one SGPB molecule and one OMTKY3 molecule.[6,19,24,25,26,27,28] The OMTKY3 molecules in the latter complexes correspond to molecule A of OMTKY3-Ala32I in the molecular structure of the SGPB:OMTKY3-Ala32I complex. As standard-mechanism protein inhibitors, OMTKY3 and its variants form a 1:1 complex with SGPB. So far there has been no evidence for the existence of a 1:2 complex of SGPB with any of the OMTKY3 variants in solution. The large value of the K_{assoc} of the SGPB:OMTKY3-Ala32I complex (~ 10^9 M^{-1}) indicates the possibility of using a highly homogeneous solution

Figure 5-6 The SGPB:OMTKY3-Ala32I complex. (a) Overall structure. Each asymmetric unit includes one SGPB molecule (molecule E, yellow) and two OMTKY3-Ala32I molecules (molecule A occupying the substrate binding region of SGPB, green; molecule B, red). (b) Occupation of the substrate-binding region of SGPB (yellow surface) by the reactive-site loop of molecule A of OMTKY3-Ala32I (green sticks). The positions of the catalytic triad of SGPB - Ser195E, His57E and Asp102E - are labeled as S, H and D, respectively. (c) The additional hydrogen bond (dotted line) in the SGPB:OMTKY3-Ala32I complex (yellow:green). The corresponding residues in the SGPB:OMTKY3 complex[6] (orange:white) are also shown.

(a)

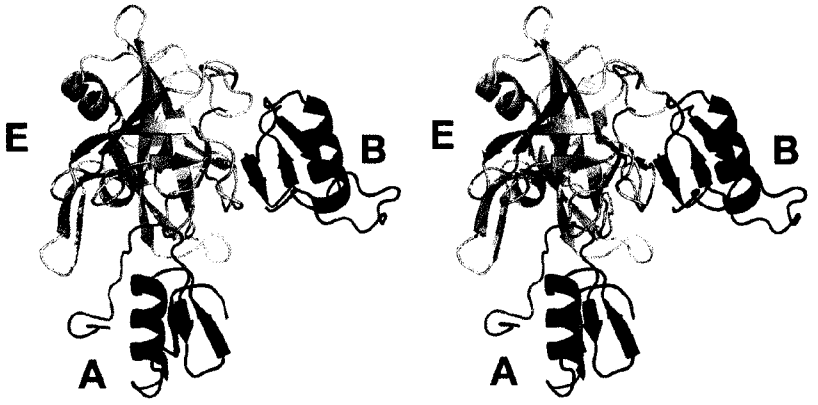
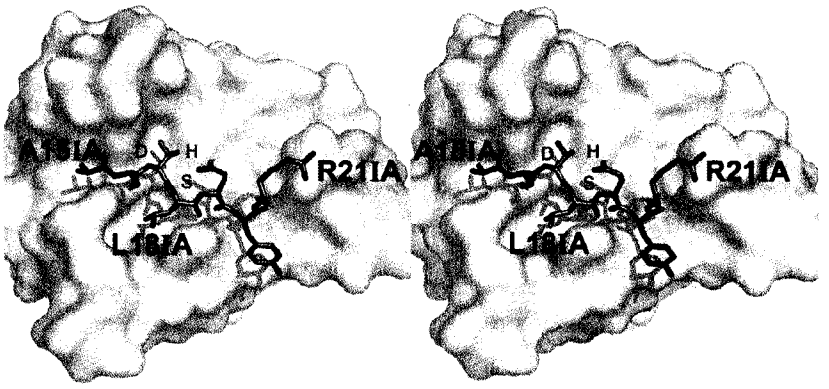
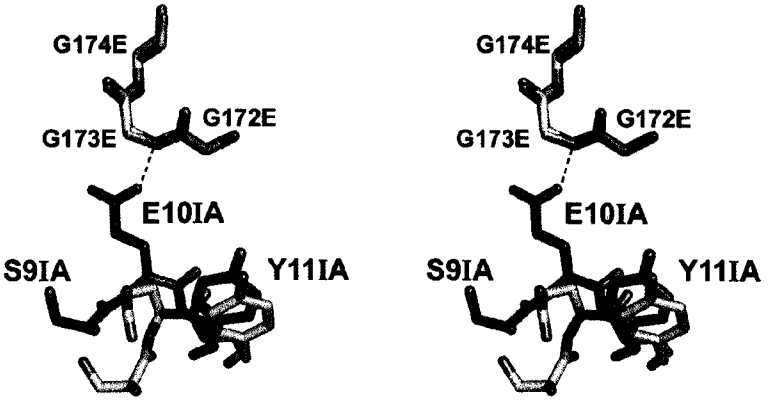


Figure 5-6 (continued)

(b)



(c)



of the complex (Section 5.2) in the attempts of growing crystals with the complex in a stoichiometry of 1:1, containing one SGPB molecule and molecule A of OMTKY3-Ala32I only, in space group $P2_1$. However, such attempts have not been successful.

The overall structures of all the three molecules are preserved in the complex. Molecule A of OMTKY3-Ala32I binds to SGPB in the expected manner, with the reactive-site loop of the former occupying the substrate-binding region of the latter (Figure 5-6b). In contrast to those of the unbound OMTKY3-Ala32I (Figure 5-2b), the main-chain and the side-chain B factors of the residues of the reactive-site loop of molecule A are generally near or below the average values (Figure 5-2c), probably as a result of the immobilization of the reactive-site loop of molecule A by the substrate-binding region of SGPB. A total of 1,272 $Å^2$ of solvent-accessible surface area is buried. The shape correlation statistic of this interface is 0.77, well within the range of values commonly observed at the peptidase-protein inhibitor interfaces (0.70 to 0.76).[14] Also, these two values are very close to those of the SGPB:OMTKY3 complex (1,209 $Å^2$ and 0.79, respectively).[6] The numerous and extensively distributed interactions between SGPB and molecule A in the SGPB:OMTKY3-Ala32I complex are in essentially the same pattern as those between SGPB and OMTKY3 in that complex, except that, strikingly, a hydrogen bond between Glu10IA $O^{\varepsilon 2}$ and the main-chain NH of Gly173E (2.5 Å) is observed in the SGPB:OMTKY3-Ala32I complex, but not in the SGPB:OMTKY3 complex (Figure 5-6c).

Superposition of the molecular structure of the unbound OMTKY3-Ala32I with that of molecule B of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex (rmsd: 0.43 Å for 195 out of 204 main-chain atoms) shows good agreement

in most atomic positions. The binding of molecule B to SGPB buries a total of 1,210 $Å^2$ of solvent-accessible surface area. The shape correlation statistic of this interface is 0.55, substantially lower than that of the interface between molecule A and SGPB. Three hydrogen bonds are observed at the interface, one between Ser51IB O^{γ} and Asp29E $O^{\delta 1}$ (2.7 Å), one between the carbonyl O of Leu50IB and the main-chain NH of Ala30E (3.2 Å), and one between the main-chain NH of Val6IB and Ser79E O^{γ} (2.7 Å). In addition, there are six contacts dispersed as small patches throughout the interface, involving six residues of SGPB (Arg41E, Ala68E, Ile113E, Pro114E, Ser141E and Gly156E) and four residues of molecule B (Thr47IB, Thr49IB, His52IB and Phe53IB). None of the residues of the reactive-site loop of molecule B are involved in these contacts. The reactive-site loop of molecule B is exposed to the solvent. Its main-chain conformation is in good agreement with that of the unbound OMTKY3-Ala32I (Table 5-3). In contrast to those of molecule A and the unbound OMTKY3-Ala32I (Figures 5-2b and 5-2c), the main-chain and the side-chain B factors of the N-terminal residues of the reactive-site loop of molecule B are near or above the average values (Figure 5-2d). This indicates the relatively high mobility of these residues, probably because of the lack of crystal contacts. In particular, the electron densities of Ala15IB and Cys16IB are not well defined. These features suggest that the interactions between SGPB and molecule B are of no biochemical or physiological significance.

Superposition of the SGPB molecule from the SGPB:OMTKY3-Ala32I complex with that from the SGPB:OMTKY3 complex[6] (rmsd: 0.27 Å for 695 out of 740 main-chain atoms) shows excellent agreement in most atomic positions,

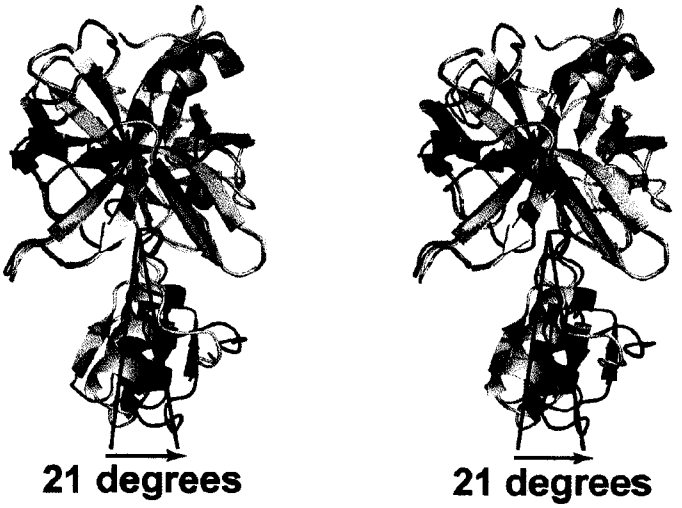
indicating that the structural effects on SGPB caused by the binding of molecule B of OMTKY3-Ala32I is minimal. Moreover, there are no interactions between molecules A and B in the same asymmetric unit of the SGPB:OMTKY3-Ala32I complex. Therefore, with regard to their structures and interactions with SGPB, molecules A and B in the same asymmetric unit probably have no influence on each other, either directly or indirectly via SGPB.

Structural effects of the Gly-to-Ala substitution of residue 32I of the SGPB-bound OMTKY3

Superposition of the SGPB:OMTKY3 complex with the SGPB:OMTKY3-Ala32I complex (with molecule B of OMTKY3-Ala32I excluded from the latter; rmsd: 0.27 Å for 712 out of 944 main-chain atoms) shows that the atomic positions of the peptidase molecules agree closely with each other, whereas the inhibitor molecules are well aligned only in their reactive-site loops. With the reactive-site loop as a hinge, the Gly-to-Ala substitution at residue 32I of OMTKY3 causes a rotation of the body of the inhibitor through an angle of 21 ° (Figure 5-7a). This rotation brings Glu10IA $O^{\varepsilon 2}$ and the main-chain NH of Gly173E into sufficiently close proximity to form of a short hydrogen bond (2.5 Å). Therefore, residue 10I should be added to the set of hypervariable contact residues of the SRA whenever the binding of an OMTKY3 variant with Ala32I is considered. This also raises the possibility of introducing additional hypervariable contact residues to the SRA as residue 32I of OMTKY3 is substituted with residues having even bulkier side chains, thereby possibly causing more dramatic conformational changes in the SGPB:OMTKY3 complex.

Figure 5-7 Hinged rigid-body rotation of the SGPB-bound OMTKY3 upon the Gly-to-Ala substitution of its residue 32I. (a) Superposition of the SGPB:OMTKY3 complex[6] (orange:white) with the SGPB:OMTKY3-Ala32I complex (yellow:green) excluding molecule B of OMTKY3-Ala32I. (b) Main-chain superposition of molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex (green) with the unbound OMTKY3-Ala32I (tint) and the SGPB-bound OMTKY3 (white). (c) Spatial relationships of Pro192BE of SGPB with the contact residues of OMTKY3: Leu18I, Tyr20I, Gly/Ala32I and Asn36I.

(a)



(b)

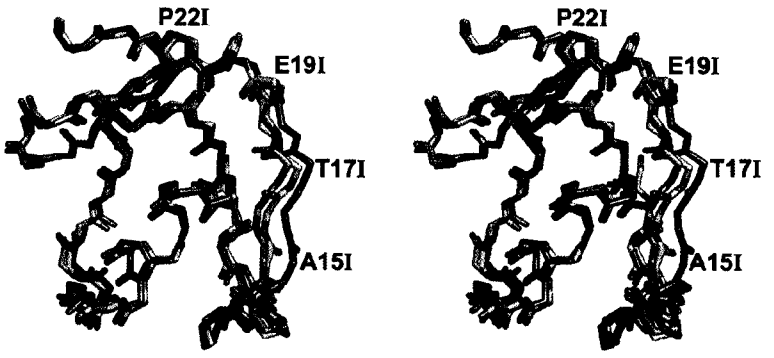
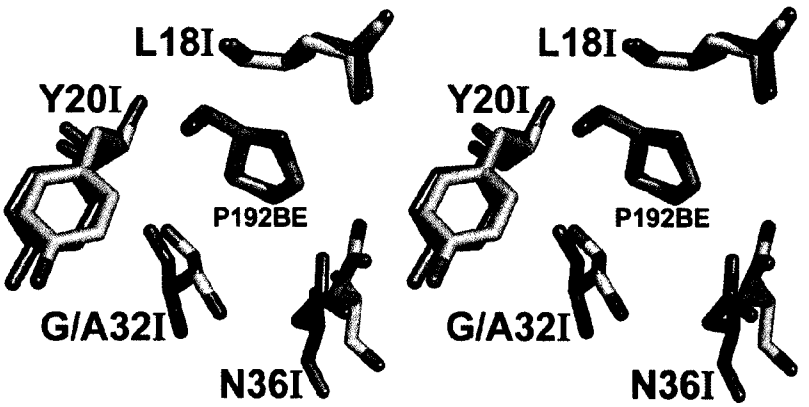


Figure 5-7 (continued)

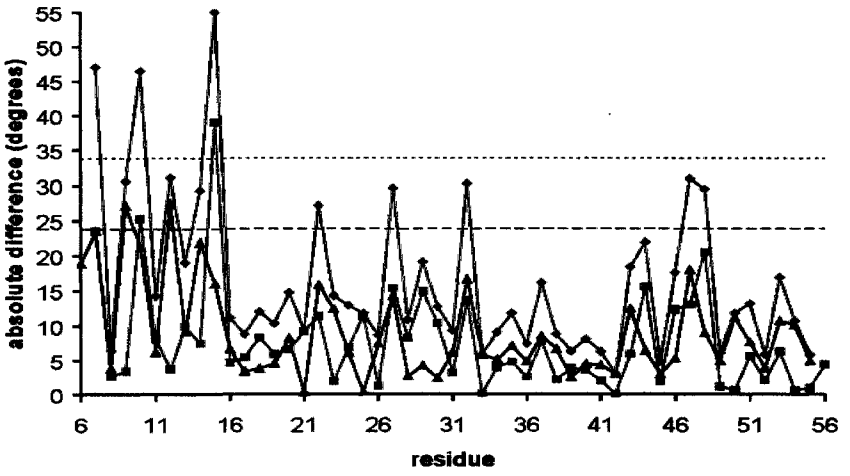
(c)



Superpositions of molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex with both the unbound OMTKY3-Ala32I (rmsd: 0.80 Å for 203 out of 204 main-chain atoms) and the SGPB-bound OMTKY3[6] (rmsd: 0.71 Å for 203 out of 204 main-chain atoms) show good alignment in all parts of the inhibitor molecules except for the residues of the reactive-site loops (Figure 5-7b). The main-chain conformational changes of the reactive-site loop, except those for Ala15I, of OMTKY3-Ala32I upon its binding to SGPB (as molecule A in the SGPB:OMTKY3-Ala32I complex) are consistent with those associated with the binding of OMTKY3 to SGPB (Table 5-3). These results show that the hinged rotation of molecule A in the SGPB:OMTKY3-Ala32I complex occurs in a rigid-body manner. In molecule A, Thr17IA $O^{\gamma 1}$ forms a direct hydrogen bond with Glu19IA $O^{\varepsilon 1}$ (2.6 Å). Glu19IA $O^{\varepsilon 1}$ also forms a hydrogen bond with the main-chain NH of Glu19IA (2.7 Å). However, no hydrogen bonds are observed between the side chains of Glu19IA and Arg21IA (Figure 5-3e). The geometry of the reactive site of molecule A and the catalytic triad of SGPB in the SGPB:OMTKY3-Ala32I complex shows virtually no difference from that observed in the SGPB:OMTKY3 complex (Table 5-4).

In order to locate the angles ϕ and ψ responsible for the hinged rigid-body rotation of molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex, the absolute differences in the angles ϕ and ψ between molecule A and the SGPB-bound OMTKY3, and the sum of these differences are plotted for all residues (Figure 5-8). Obviously, there is a significant peak for the angle ϕ of Ala15I, the N-terminus of the reactive-site loop. However, none of the angles on the C-terminal side of the reactive-site loop shows a significant peak. The peaks for the angles ϕ and ψ of

Figure 5-8 Plot of the absolute differences in the angles ϕ and ψ between the SGPB-bound OMTKY3[6] and molecule A in the SGPB:OMTKY3-Ala32I complex, and the sums of these differences. The absolute differences in the angles ϕ ($|\Delta\phi|$) and ψ ($|\Delta\psi|$) are plotted as red squares and green triangles, respectively, whereas the sums of these differences ($|\Delta\phi|+|\Delta\psi|$) are plotted as blue diamonds. The 2.7σ of $|\Delta\phi|$ and $|\Delta\psi|$ is indicated by the dashed line, and that of $|\Delta\phi|+|\Delta\psi|$ is indicated by the dotted line.



Pro22I, probably because of its conformational restrictions, are high but are not significant. Interestingly, similar trends are exhibited by the angles ϕ and ψ of OMTKY3 in the molecular structure of the CARL:OMTKY3 complex, resulting in the two different conformations of OMTKY3.[31] Nonetheless, in the SGPB:OMTKY3-Ala32I complex, due to the steric hindrance between Pro192BE and Ala32IA, it seems unlikely for molecule A to adopt two different conformations as were observed for the CARL:OMTKY3 complex.

Interactions among Leu18I, Tyr20I, residue 32I and Asn36I of SGPB-bound OMTKY3

In the SGPB:OMTKY3-Ala32I complex, Ala32IA C^{β} makes contacts with Pro192BE C^{β} and C^{γ}, and is located exactly at the position of Gly32I C^{α} of OMTKY3 in the SGPB:OMTKY3 complex (Figure 5-7c). Thus, it is highly likely that the hinged rigid-body rotation of molecule A of OMTKY3-Ala32I in the SGPB:OMTKY3-Ala32I complex is caused by the steric hindrance between the methyl group (the atom C^{β}) of Ala32IA and the atoms of Pro192BE. The results of this steric hindrance are indicated by the moderate peaks in the plots of $|\Delta\phi|$, $|\Delta\psi|$ and $|\Delta\phi|+|\Delta\psi|$ of residue 32I in Figure 5-8. In contrast, there is virtually no conformational change in the loop containing Pro192BE of SGPB in response to the Gly-to-Ala substitution of residue 32I of OMTKY3.

The contact between Leu18I of OMTKY3 and Pro192BE of SGPB is not affected by the Gly-to-Ala substitution of residue 32I of OMTKY3, as a result of the conservation of the main-chain conformation of the reactive-site loop of OMTKY3. However, the hinged rigid-body rotation of molecule A of OMTKY3-Ala32I in the

SGPB:OMTKY3-Ala32I complex moves the side chains of Tyr20IA and Asn36IA away from Pro192BE. This modulates the interactions (mediated by Pro192BE of SGPB) among Leu18I, Tyr20I, residue 32I and Asn36I of OMTKY3 (Figure 5-7c). This result shows that the non-additivity effects in the SRA due to changes in the interactions among these residues upon the binding of OMTKY3 to SGPB are complicated by substitutions of residue 32I of OMTKY3.

Relationships between the conformation and the crystal packing of the SGPB:OMTKY3 complex

Molecular structures of the complexes of SGPB with all the OMTKY3 variants determined so far show that residue 32I of OMTKY3 is not exposed to the solvent. It makes contacts with Pro192BE of SGPB. Therefore, the Gly-to-Ala substitution of residue 32I of OMTKY3 does not seem likely to modify the surface properties of the 1:1 SGPB:OMTKY3 complex, nor to change its preference for space group $P2_1$ in crystallization, unless the complex undergoes a conformational change. Assuming that the SGPB:OMTKY3-Ala32I complex had adopted a 1:1 stoichiometry (containing the SGPB molecule and molecule A of OMTKY3-Ala32I only), modeling studies (results not shown) suggest that the hinged rigid-body rotation of molecule A results in a conformational change of this complex that disfavors its packing in the $P2_1$ unit cell defined by the complexes of SGPB with other OMTKY3 variants determined so far. The $P2_12_12_1$ unit cell reported here is more spacious and better able to accommodate the SGPB:OMTKY3-Ala32I complex. It is even possible to fill up the remaining space by adding a second OMTKY3-Ala32I molecule (molecule B) to each unit of the complex. The Matthews coefficient of the $P2_12_12_1$

crystal is 1.9 $Å^3$/Da, within the range of values commonly observed for macromolecular crystals (1.7 $Å^3$/Da to 3.5 $Å^3$/Da).[40] Few crystal contacts involve molecule A, suggesting that the hinged rigid-body rotation of molecule A is not a consequence of the change in crystal packing.

Rigid-body protein-protein interactions - a requirement for the validity of the additivity approximation in the SRA

Using the additivity approximation, a computationally simple SRA that predicts the K_{assoc} of an OMTKY3 variant with six serine peptidases was developed.[41,42] A test of the SRA with 450 experimentally determined K_{assoc} values showed that 90 % of the predictions were accurate or partially accurate.[43] The validity of the additivity approximation requires that the interaction between any of the two hypervariable consensus contact residues of the inhibitor remains unchanged upon the binding of the inhibitor to the six serine peptidases. The overall structure of OMTKY3 is rigid, and remains unchanged upon the binding of OMTKY3 to the serine peptidases. This is one of the factors contributing to the high success rate of the SRA.[42] However, OMTKY3 is not an ideal rigid body. As shown by the molecular structures determined so far (including the three reported here), OMTKY3 has several conformationally flexible regions, one of which is its reactive-site loop. Important for the physiological function of OMTKY3, the conformational flexibility of the reactive-site loop allows for the adaptability of OMTKY3 in inhibiting a broad range of serine peptidases. Analysis of the molecular structures of the unbound OMTKY3 and OMTKY3 bound to different serine peptidases shows that adjustments do occur in the main-chain conformation of the reactive-site loop of OMTKY3 upon the

binding of the inhibitor to the peptidases (Table 5-3). These conformational adjustments are accompanied by changes in the interactions within some pairs of the hypervariable consensus contact residues of OMTKY3, resulting in the non-additivity effects in the SRA. For example, the changes in the angles ϕ and ψ of Thr17I to Glu19I of OMTKY3 upon the binding of OMTKY3 to SGPB, CHYM and HLE are accompanied by the strengthening of the hydrogen bond between Thr17I $O^{\gamma 1}$ and Glu19I $O^{\varepsilon 1}$ of OMTKY3, and the switch in the hydrogen-bonding partner of the main-chain NH of Glu19I of OMTKY3. This leads to the non-additivity effects for residues 17I and 19I of OMTKY3. The non-ideal rigid-body behaviors of OMTKY3 limit the performance of the SRA. In the test with 450 inhibitors, 10 % of the predictions were inaccurate.[43]

The problem in accounting for non-additivity effects remains somewhat unsettled. Its complexity is the primary reason. For example, the occurrence of non-additivity effects for a particular pair of residues of OMTKY3 depends on which peptidase the inhibitor binds to. Double substitutions of residues 17I and 19I of OMTKY3 have been studied, and some restrictions have been established for the use of the SRA. Initially, it was found that, for SGPB, CHYM and HLE being the target peptidases, SRA-based predictions are accurate only for the OMTKY3 variants with either Thr17I or Glu19I present (i.e. 38 of all possible variants). Later, it was suggested that additivity holds as long as there is no strong interaction, such as ionic or hydrogen-bonding interaction, between residues 17I and 19I of the OMTKY3 variant. This requirement could be fulfilled by 49 additional variants, as Gly, Ala, Val, Leu, Ile, Met and Phe normally do not interact strongly with each other.[42,43] The

results of recent studies further suggest that additivity also holds for either one of residues 17I and 19I being polar and the other one being non-polar. However, additivity does not hold if both residues 17I and 19I are charged or polar (Qasim, M. A., personal communications).

5.4 References

1. Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogin, E., Rothberg, I., Ryan, K. & Laskowski, M., Jr. (1997). Binding of amino acid side-chains to S1 cavities of serine proteinases. *J. Mol. Biol.* **266**, 441-461.

2. Jurášek, L., Carpenter, M. R., Smillie, L. B., Gertler, A., Levy, S. & Ericsson, L. H. (1974). Amino acid sequence of Streptomyces griseus protease B, A MAJOR COMPONENT OF Pronase. *Biochem. Biophys. Res. Commun.* **61**, 1095-1100.

3. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W., Jr. & Sweet, R. M., eds.), Vol. 276: Macromolecular Crystallography, Part A, pp. 307-326 Academic Press, New York, NY.

4. Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

5. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Cryst.* **D59**, 1131-1137.

6. Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr. & James, M. N. G. (1995). Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with Streptomyces griseus proteinase B. *Protein Sci.* **4**, 1985-1997.

7. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.

8. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240-255.

9. McRee, D. E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.

10. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

11. Evans, P. & Fermi, J. (1995). GEOMCALC. Medical Research Council Laboratory of Molecular Biology, Cambridge.

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

13. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.

14. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946-950.

15. Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593-604.

16. Cohen, G. H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Cryst.* **30**, 1160-1161.

17. Kelley, L. A. & Sutcliffe, M. J. (1997). OLDERADO: on-line database of ensemble representatives and domains. On Line Database of Ensemble Representatives And DOmains. *Protein Sci.* **6**, 2628-2630.

18. Cruickshank, D. W. J. (1999). Remarks about protein structure precision. *Acta Cryst.* **D55**, 583-601.

19. Bateman, K. S., Huang, K., Anderson, S., Lu, W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2001). Contribution of peptide bonds to inhibitor-protease binding: crystal structures of the turkey ovomucoid third domain backbone variants OMTKY3-Pro18I and OMTKY3-psi[COO]-Leu18I in complex with *Streptomyces griseus* proteinase B (SGPB) and the structure of the free inhibitor, OMTKY-3-psi[CH2NH2+]-Asp19I. *J. Mol. Biol.* **305**, 839-849.

20. Weber, E., Papamokos, E., Bode, W., Huber, R., Kato, I. & Laskowski, M., Jr. (1981). Crystallization, crystal structure analysis and molecular model of the third domain of Japanese quail ovomucoid, a Kazal type inhibitor. *J. Mol. Biol.* **149**, 109-123.

21. Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I. & Laskowski, M., Jr. (1982). Crystallographic refinement of Japanese quail ovomucoid, a Kazal-type inhibitor, and model building studies of complexes with serine proteases. *J. Mol. Biol.* **158**, 515-537.

22. Bode, W., Epp, O., Huber, R., Laskowski, M., Jr. & Ardelt, W. (1985). The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). *Eur. J. Biochem.* **147**, 387-395.

23. Hoogstraten, C. G., Choe, S., Westler, W. M. & Markley, J. L. (1995). Comparison of the accuracy of protein solution structures derived from conventional and network-edited NOESY data. *Protein Sci.* **4**, 2289-2299.

24. Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M., Jr. & James, M. N. G. (1982). Refined crystal structure of the molecular complex of *Streptomyces griseus* protease B, a serine protease, with the third domain of the ovomucoid inhibitor from turkey. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4868-4872.

25. Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. (1983). Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-Å resolution. *Biochemistry* **22**, 4420-4433.

26. Huang, K. (1995). *Structural studies of the interactions between serine proteinases and protein inhibitors*. PhD thesis, University of Alberta.

27. Bateman, K. S. (1999). *Structural studies of protein-protein interactions*. PhD thesis, University of Alberta.

28. Bateman, K. S., Anderson, S., Lu, W., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2000). Deleterious effects of beta-branched residues in the S1 specificity pocket of *Streptomyces griseus* proteinase B (SGPB): crystal structures of the turkey ovomucoid third domain variants Ile18I, Val18I, Thr18I, and Ser18I in complex with SGPB. *Protein Sci.* **9**, 83-94.

29. Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M., Jr. & James, M. N. G. (1987). Crystal and molecular structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J. Mol. Biol.* **195**, 397-418.

30. Horn, J. R., Ramaswamy, S. & Murphy, K. P. (2003). Structure and energetics of protein-protein interactions: the role of conformational heterogeneity in OMTKY3 binding to serine proteases. *J. Mol. Biol.* **331**, 497-508.

31. Maynes, J. T., Cherney, M. M., Qasim, M. A., Laskowski, M., Jr. & James, M. N. G. (2005). Structure of the subtilisin Carlsberg-OMTKY3 complex reveals two different ovomucoid conformations. *Acta Cryst.* **D61**, 580-588.

32. Bode, W., Wei, A. Z., Huber, R., Meyer, E., Travis, J. & Neumann, S. (1986). X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J.* **5**, 2453-2458.

33. Laskowski, M., Jr. & 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**, 324-337.

34. Laskowski, M. J., Qasim, M. A. & Lu, S. M. (2000). Interaction of standard mechanism, canonical protein inhibitors with serine proteinases. In *Protein-protein recognition* (Kleanthous, C., ed.), pp. 228-279. Oxford University Press, New York.

35. Ardelt, W. & Laskowski, M., Jr. (1985). Turkey ovomucoid third domain inhibits eight different serine proteinases of varied specificity on the same ...Leu18-Glu19 ... reactive site. *Biochemistry* **24**, 5313-5320.

36. Musil, D., Bode, W., Huber, R., Laskowski, M., Jr., Lin, T. Y. & Ardelt, W. (1991). Refined X-ray crystal structures of the reactive site modified ovomucoid inhibitor third domains from silver pheasant (OMSVP3*) and from Japanese quail (OMJPQ3*). *J. Mol. Biol.* **220**, 739-755.

37. Radisky, E. S. & Koshland, D. E., Jr. (2002). A clogged gutter mechanism for protease inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10316-10321.

38. Radisky, E. S., Lee, J. M., Lu, C. J. & Koshland, D. E., Jr. (2006). Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6835-6840.

39. Helland, R., Berglund, G. I., Otlewski, J., Apostoluk, W., Andersen, O. A., Willassen, N. P. & Smalas, A. O. (1999). High-resolution structures of three new trypsin-squash-inhibitor complexes: a detailed comparison with other trypsins and their complexes. *Acta Cryst.* **D55**, 139-148.

40. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

41. Lu, S. M., Lu, W., Qasim, M. A., Anderson, S., Apostol, I., Ardelt, W., Bigler, T., Chiang, Y. W., Cook, J., James, M. N. G., Kato, I., Kelly, C., Kohr, W., Komiyama, T., Lin, T. Y., Ogawa, M., Otlewski, J., Park, S. J., Qasim, S., Ranjbar, M., Tashiro, M., Warne, N., Whatley, H., Wieczorek, A., Wieczorek, M., Wilusz, T., Wynn, R., Zhang, W. & Laskowski, M., Jr. (2001). Predicting

the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1410-1415.

42. Laskowski, M., Jr., Qasim, M. A. & Yi, Z. (2003). Additivity-based prediction of equilibrium constants for some protein-protein associations. *Curr. Opin. Struct. Biol.* **13**, 130-139.

43. Qasim, M. A., Lu, W., Lu, S. M., Ranjbar, M., Yi, Z., Chiang, Y. W., Ryan, K., Anderson, S., Zhang, W., Qasim, S. & Laskowski, M., Jr. (2003). Testing of the additivity-based protein sequence to reactivity algorithm. *Biochemistry* **42**, 6460-6466.

44. Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.

Chapter 6

Conclusion

6.1 Part One - SARS coronavirus main peptidase

SARS-CoV M^{pro} plays an essential role in the life cycle of the virus. It is an attractive molecular target for the development of anti-SARS drugs that act as its inhibitors. Since the outbreak of SARS in 2002 and its rapid spread throughout early 2003, considerable efforts have been directed from all over the world to the studies of M^{pro}. This results in the availability of many molecular structures of M^{pro} and its variants over a pH range of 5.9 to 9.0 in a variety of space groups, among which the molecular structures of M^{pro} at pH 6.0 in space group $P2_1$ and at pH 6.5 in space group C2, and of $M^{pro}_{+A(-1)}$ at pH 6.5 in space group $P4_32_12$ are reported in Part One of this thesis. Comparisons among all of these molecular structures show that the protomer orientation and the protein fold of M^{pro} are preserved in various conditions, but the active site and the S1 specificity pocket of each protomer of M^{pro} adopt an ensemble of conformations, rather than a single conformation, in solution at pH 5.9 to 6.6. The catalytic competence of an M^{pro} protomer is apparently correlated with the integrity of the aromatic interaction between the imidazole ring of His163 and the phenyl ring of Phe140. This interaction is weak, and is susceptible to disruptions that could be caused by changes in a number of factors. The formation of this interaction can be viewed as a reversible process in dynamic equilibrium. Low pH (near or below pH 6.5) disfavors the formation of this interaction, because of the increased protonation and positive-charge development on the imidazole ring of His163. This leads to an increase in the number of M^{pro} protomers in the catalytically incompetent

conformation, thereby resulting in a decrease in the overall catalytic activity of M^{pro}. Introduction of additional residues to the N-terminus of an M^{pro} protomer (e.g. an additional Ala for $M^{pro}_{+A(-1)}$) disrupts the normal interactions of Ser1 of the parent protomer with Phe140 and Glu166 of the opposite protomer. This weakens the 'conformational anchor' of Phe140, resulting in the increased mobility of residues Lys137 to Ser144 of the opposite protomer, and thus the increased vulnerability of the aromatic interaction between the imidazole ring of His163 and the phenyl ring of Phe140 in the opposite protomer to disruptions. As at low pH, the formation of this aromatic interaction is disfavored, leading to an increase in the number of M^{pro} protomers in the catalytically incompetent conformation, thereby resulting in a decrease in the overall catalytic activity of M^{pro}.

Compared with the previously proposed pH-triggered switch for the catalytic activity of M^{pro},[1] the dynamic-equilibrium model proposed here provides refined explanations for the observed dependence of the catalytic activity of M^{pro} not only on the pH, but also on the integrity of the interactions among Phe140 and Glu166 of one protomer, and Ser1 of the opposite protomer. The effects of substrate binding can be explained using this model as well (to be discussed below). This model also raises the possibility of a second mechanism, such as the change with pH in the protonation/deprotonation state of the catalytic dyad, being involved in the pH dependence of the catalytic activity of M^{pro}.

Also reported in Part One of this thesis are the molecular structures of the M^{pro}:APE complex at pH 6.5 in space group C2, and of $M^{pro}_{+A(-1)}$ complex at pH 6.5 in space groups $P4_32_12$ and $P2_12_12_1$. APEs were synthesized as a new class of

inhibitors apparently specific for cysteine peptidases in peptidase clan CD, including the legumains (family C13) and the caspases (family C14). It has been determined that Cbz-Leu-Phe-AGln-(2*S*,3*S*)EP-COOEt, an APE with P1-AGln mimicking the S1 specificity of M^{pro} for Gln, inhibited M^{pro} strongly with a $k_{inact}/K_i = 1900 \pm 400\ M^{-1}s^{-1}$.[1,2] As shown by the molecular structures reported here, APE acts as a competitive inhibitor of M^{pro}, binding to the active sites and substrate-binding regions of both protomers. Cys145 S^{γ} of M^{pro} nucleophilically attacks the epoxide atom C3 of APE, leading to the formation of a covalent bond between these two atoms and the opening of the conformationally strained epoxide ring of APE. The rest of APE interacts with M^{pro} in a substrate-like manner. The main-chain carbonyl O of P1-AGln of APE occupies the oxyanion hole of M^{pro}. The main chain of the aza-peptide component of APE interacts with the main chain of residues His164 to Glu166 of M^{pro} through amide nitrogen-carbonyl oxygen hydrogen bonding in the fashion of an anti-parallel β-sheet. The side chains of P2-Phe and P1-AGln of APE occupy the S2 and the S1 specificity pockets of M^{pro}, respectively. In the S1 specificity pocket of M^{pro}, the side-chain carbonyl O of P1-AGln of APE forms a hydrogen bond with the atom $N^{\varepsilon 2}$ of the conserved residue His163 - the primary determinant of the predominant S1 specificity of M^{pro} for Gln. The side chain of P3-Leu of APE extends into the solvent. Taking up parts of the space for residue P4 of a substrate, the Cbz group of APE either squeezes into the S4 specificity pocket of M^{pro}, or is exposed to the solvent. The ethyl ester group of APE lies against the `ceiling' of the active site of M^{pro}. It has been found that M^{pro} can be inhibited by the 2*S*,3*S* diastereomer of APE only.[2] As shown by modeling studies on the basis of the molecular structures reported here, the

epoxide atom C3 of the 2*S*,3*S* diastereomer of APE is better positioned for the nucleophilic attack of Cys145 S^{γ} of M^{pro}. On the other hand, the binding of the other three diastereomers of APE to M^{pro} is disrupted by the steric clashes of their ethyl ester groups with the active site of M^{pro} or with the aza-peptide components of the APE molecules themselves.

Interestingly, comparison of the molecular structures of the unbound and the APE-bound $M^{pro}_{+A(-1)}$ shows that the binding of APE induces the recovery of the catalytically competent conformation of the active sites and the S1 specificity pockets of both protomers of $M^{pro}_{+A(-1)}$. This is probably driven by the formation of the covalent bond between Cys145 S^{γ} of $M^{pro}_{+A(-1)}$ and the epoxide atom C3 of APE, and favored by the following three features. Firstly, the side-chain amide NH_2 group of P1-AGln of APE donates hydrogen bonds to the carbonyl O of Phe140 and the side-chain carboxylate oxygens of Glu166 of $M^{pro}_{+A(-1)}$. Secondly, the ethyl ester group of APE interacts with the 'ceiling' of the active site of $M^{pro}_{+A(-1)}$. Thirdly, the main-chain conformation of P1-AGln of APE is restricted by the π-conjugation of its atom N^{α} and carbonyl group. Phe140 of $M^{pro}_{+A(-1)}$ is conformationally anchored by the hydrogen bond between its carbonyl O and the side-chain amide NH_2 group of P1-AGln of APE. This increases the resistance of the aromatic interaction between the phenyl ring of Phe140 and the imidazole ring of His163 to disruptions. Thus, the formation of this aromatic interaction is favored. The similarities of $M^{pro}_{+A(-1)}$ to M^{pro} and of APE to a true substrate, and the molecular structures of the inhibitor-bound M^{pro} (both the wild-type and the variants) reported elsewhere strongly suggest that the

binding of a true substrate to M^{pro} follows an induced-fit model as well, and this favors the progression of the catalytic mechanism of M^{pro}.

Supported by the biochemical data[2] and the molecular structures reported here, APE is an excellent lead compound for the development of peptidomimetic inhibitors of M^{pro}. Possible ways to improve the efficacy of APE as an M^{pro} inhibitor include the substitution of P3-Leu of APE with an amino acid having a hydrophilic side chain (e.g. Asn), the addition of an amino acid having a small side chain (e.g. Ser, Thr, Val, Pro, Ala) as residue P4 of APE, the replacement of the Cbz group of APE with a hydrophilic group (e.g. an acetyl group), and the attachment of some other substituents (rather than the ethyl ester group) to the epoxide atom C2 of APE for maximal interactions with the S1' subsite of M^{pro}. The molecular structures reported here do reveal some structural features of APE that contribute to its success as an M^{pro} inhibitor; for example, the presence of a conformationally strained epoxide ring, the presence of a hydrogen-bond acceptor (such as the side-chain carbonyl group of P1-AGln) and a hydrogen-bond donor (such as the side-chain NH_2 group of P1-AGln) at the tip of the side chain of residue P1, the presence of a substituent on the epoxide atom C2, and the presence of the atom N^{α} (rather than C^{α}) in residue P1. These features may be useful in the development of new peptidomimetic inhibitors of M^{pro}.

Two ionic interactions have been observed at the dimer interface of M^{pro}. Both of them involve residues of the 'N-finger' (N-terminal residues 1 to 7) of each protomer. Their contribution to the dimerization and the catalytic activity of M^{pro} is an issue of interest. The ionic interaction between the side chains of Arg4 and Glu290

from opposite protomers is observed in all of the molecular structures of M^{pro} and its variants determined so far, whereas that between the α-amino group of Ser1 and the side-chain carboxylate group of Glu166 from opposite protomers is observed only in some of the molecular structures. This suggests that the latter is of less importance in maintaining the dimerization and determining the protomer orientation of M^{pro}. In most of the molecular structures of M^{pro} and its variants determined so far, the loss of the ionic interaction between the α-amino group of Ser1 and the side-chain carboxylate group of Glu166 from opposite protomers is accompanied by the loss of the amide nitrogen-carbonyl oxygen hydrogen bonds between Ser1 and Phe140 from opposite protomers. As discussed above, this disfavors the formation of the aromatic interaction between the phenyl ring of Phe140 and the imidazole ring of His163, leading to an increase in the number of M^{pro} protomers in the catalytically incompetent conformation, thereby resulting in a decrease in the overall catalytic activity of M^{pro}.

6.2 Part Two - Streptogrisin B

The structural studies reported in Part Two of this thesis are mainly to provide the basis for the development of the SRA that predicts the K_{assoc} values of OMTKY3 variants with six serine peptidases, one of which is SGPB. Extensive studies have been done on the interactions of OMTKY3 and its variants with SGPB. The molecular structures of these complexes have been determined at resolutions higher than 2 Å. However, the initial molecular structure of the native SGPB was determined at a relatively low resolution (2.8 Å) and only at the acidic pH of 4.2. With the molecular structures of SGPB reported in Chapter 4 of this thesis (at the highest

resolutions its crystals can afford - 1.2 Å, and at both pH 4.2 and 7.3), structural comparisons can be placed on an equal footing. Unexpectedly, additional electron density peaks were observed in the active site and the substrate-binding region of SGPB in the computed maps at both pH values. The electron density peaks observed at pH 4.2 were assigned as a tetrapeptide, Asp-Ala-Ile-Tyr, occupying the active site and the substrate-binding region of SGPB in a substrate-like manner; whereas those observed at pH 7.3 were assigned as a tyrosine molecule and a leucine molecule existing at equal occupancies, with their side chains occupying the S1 specificity pockets of both SGPB molecules in the asymmetric unit. Refinement with relaxed geometric restraints resulted in molecular structures representing mixtures of the second tetrahedral intermediates and the enzyme-product complexes of SGPB existing in a pH-dependent equilibrium, apparently with the former favored at high pH. The trajectory of the nucleophilic attack of Ser195 O^{γ} of SGPB on the α-carboxylic C of residue P1 of the bound species is consistent with that observed previously in the Michaelis complex, the tetrahedral intermediate and the good acyl-enzyme intermediates of trypsin, and in various serine peptidase-protein inhibitor complexes. Particularly, the angle θ_1 for the nucleophilic attack is close to 90 ° in all cases. The molecular structures reported here can also serve as models of the first tetrahedral intermediates of SGPB, whereas those of the complexes of SGPB with OMTKY3 and its variants serve as excellent models of the Michaelis complex of SGPB. Structural comparisons show that residues Glu192A to Gly193 of SGPB move towards the carbonyl O of residue P1 of the bound species upon the formation of the tetrahedral intermediate, in order to maintain the effective stabilization of the

oxyanion developed. Meanwhile, adjustments occur in the side-chain conformational angles of His57 and Ser195 of SGPB, favoring the progression of the catalytic mechanism of SGPB. There is no sign of the previously proposed flipping of the imidazole ring of His57[3] observed in these structural results for SGPB.

Extensive studies have been done by many groups to elucidate the catalytic mechanism of serine peptidases. Importantly, the molecular structures of many serine peptidases have been determined recently at high resolutions. This allows for the elucidation of the mechanism in precise structural details. SGPB has served as one of the models for studying the catalytic activities of serine peptidases. Many of the observations given by the 1.2 Å-resolution molecular structures of SGPB reported here are consistent with those given by the high-resolution molecular structures of other serine peptidases, thus adding weight to our understanding of the catalytic mechanism of serine peptidases.

In the SRA-related structural studies, the molecular structure of the unbound wtOMSVP3 at a resolution of 1.5 Å[4] used to serve as a model of the unbound ovomucoid third domain. However, the SRA is based on OMTKY3, and most of the previously determined molecular structures contain an OMTKY3 variant. Chapter 5 of this thesis reports the molecular structure of the unbound OMTKY3 at a resolution of 1.2 Å. This is now considered the best available model of the unbound ovomucoid third domain. The molecular structures of OMTKY3 and wtOMSVP3 are virtually identical. The two inhibitors differ only in the identity of residue 18I (Leu in the former; Met in the latter). All of the molecular structures of the unbound avian

ovomucoid third domains determined so far show that the side chain of residue 18I is exposed to the solvent.

The SRA has been tested by the group of the late Michael Laskowski, Jr. with 450 experimentally determined K_{assoc} values. 90 % of the SRA-based predictions were found to be accurate or partially accurate. The remaining 10 % were found to be inaccurate, indicating that the additivity approximation - the basis of the SRA - is not applicable in certain situations. The validity of the additivity approximation requires that the interaction between any of the two hypervariable consensus contact residues of an OMTKY3 variant remains unchanged upon the binding of the inhibitor to the six serine peptidases. Structural comparisons show that the hydrogen bond between Thr17I $O^{\gamma 1}$ and Glu19I $O^{\varepsilon 1}$ of OMTKY3 is strengthened upon the binding of OMTKY3 to SGPB and CHYM, and the main-chain NH of Glu19I of OMTKY3 switches its hydrogen-bonding partner from Thr17I $O^{\gamma 1}$ to Glu19I $O^{\varepsilon 1}$ upon the binding of OMTKY3 to SGPB, CHYM and HLE. These changes explain the previously addressed non-additivity effects between residues 17I and 19I in the SRA-based predictions of the K_{assoc} values of OMTKY3 variants with SGPB, CHYM and HLE. Also, the ionic interaction between the side chains of Glu19I and Arg21I of OMTKY3 is apparently strengthened upon the binding of OMTKY3 to CARL. This raises the possibility of non-additivity effects between residues 19I and 21I in the SRA-based predictions of the K_{assoc} values of OMTKY3 variants with CARL.

Non-additivity effects were also found between residues 18I and 32I, and between residues 20I and 32I in the SRA-based predictions of the K_{assoc} values of OMTKY3 variants with SGPB. With the molecular structures of the unbound

OMTKY3-Ala32I and the SGPB:OMTKY3-Ala32I complex reported in Chapter 5 of this thesis, how residue 32I contributes to these non-additivity effects can be better understood. The main-chain conformational angles and freedom of residue 32I of OMTKY3 do not seem to contribute to these non-additivity effects, because structural comparison of the unbound OMTKY3 and the unbound OMTKY3-Ala32I shows that these parameters remain unchanged upon the Gly-to-Ala substitution. However, this substitution does change the stoichiometry of the SGPB:OMTKY3 complex. The asymmetric unit of the SGPB:OMTKY3-Ala32I complex contains one SGPB molecule (molecule E) and two OMTKY3-Ala32I molecules (molecules A and B), instead of showing the expected 1:1 stoichiometry. Detailed analysis of the intermolecular interactions in the SGPB:OMTKY3-Ala32I complex strongly suggest that the binding of molecule B of OMTKY3-Ala32I to SGPB is an artifact. It is of no biochemical or physiological significance. Its influence on SGPB and molecule A of OMTKY3-Ala32I is minimal. In contrast, molecule A of OMTKY3-Ala32I binds to SGPB in an expected manner, with the reactive-site loop of the former occupying the substrate-binding region of the latter. The pattern of the interactions between molecule A and SGPB, the main-chain conformation of the reactive-site loop of molecule A, and the geometry of the interactions between the reactive site of molecule A and the active site of SGPB are virtually the same as those observed between OMTKY3 and SGPB in that complex. Nonetheless, structural comparison of the SGPB:OMTKY3 complex with the SGPB:OMTKY3-Ala32I complex (with molecule B of OMTKY3-Ala32I excluded from the latter) shows that the Gly-to-Ala substitution of residue 32I of OMTKY3 leads to a rigid-body rotation of OMTKY3

through an angle of 21 °, with its reactive-site loop as the hinge. This rotation is caused by the steric hindrance between the methyl group of Ala32IA and the atoms of Pro192BE, and is effected mainly by the change in the angle ϕ of Ala15IA, the N-terminus of the reactive-site loop. As a result of this rotation, the side chains of Tyr20IA and Asn36IA move away from Pro192BE. This modulates the interactions (mediated by Pro192BE of SGPB) among Leu18I, Tyr20I, residue 32I and Asn36I of OMTKY3. Thus, the non-additivity effects in the SRA due to changes in the interactions among these residues upon the binding of OMTKY3 to SGPB are complicated by substitutions of residue 32I of OMTKY3. Additionally, this rotation brings the main-chain NH of Gly173E and Glu10IA $O^{\varepsilon 2}$ into sufficiently close proximity to form a hydrogen bond. Therefore, residue 10I should be added to the set of hypervariable contact residues of the SRA whenever the binding of an OMTKY3 variant with Ala32I is considered. This also raises the possibility of introducing additional hypervariable contact residues to the SRA as residue 32I of OMTKY3 is substituted with residues having even bulkier side chains, thereby possibly causing more dramatic conformational changes in the SGPB:OMTKY3 complex.

Results of the structural studies reported in Part Two of this thesis indicate that the structural rigidity of SGPB and OMTKY3 is one of the factors contributing to the high success rate of the SRA. The overall structures of both SGPB and OMTKY3 are rigid, and remain essentially unchanged upon association. However, SGPB and OMTKY3 are not ideal rigid bodies. Each of them has several conformationally flexible regions. In OMTKY3, one of these regions is the reactive-site loop. Important for the physiological function of OMTKY3, the conformational flexibility

of the reactive-site loop allows for the adaptability of OMTKY3 in inhibiting a broad range of serine peptidases. Analysis of the molecular structures of the unbound OMTKY3 and OMTKY3 bound to different serine peptidases shows that adjustments do occur in the main-chain conformation of the reactive-site loop of OMTKY3 upon the binding of the inhibitor to the peptidases. These conformational adjustments are accompanied by changes in the interactions within some pairs of the hypervariable consensus contact residues of OMTKY3, resulting in the non-additivity effects in the SRA. For example, the changes in the angles ϕ and ψ of residues Thr17I to Glu19I of OMTKY3 upon the binding of OMTKY3 to SGPB, CHYM and HLE are accompanied by the strengthening of the hydrogen bond between Thr17I $O^{\gamma 1}$ and Glu19I $O^{\varepsilon 1}$ of OMTKY3, and the switch in the hydrogen-bonding partner of the main-chain NH of Glu19I of OMTKY3. This leads to the non-additivity effects between residues 17I and 19I of OMTKY3. The non-ideal rigid-body behaviors of OMTKY3 limit the performance of the SRA.

6.3 References

1. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13190-13195.

2. Lee, T. W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. G. (2005). Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J. Mol. Biol.* **353**, 1137-1151.

3. Ash, E. L., Sudmeier, J. L., Day, R. M., Vincent, M., Torchilin, E. V., Haddad, K. C., Bradshaw, E. M., Sanford, D. G. & Bachovchin, W. W. (2000). Unusual 1H NMR chemical shifts support (His) C(epsilon)1...O==C H-bond: proposal for reaction-driven ring flip mechanism in serine protease catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10371-10376.

4. Bode, W., Epp, O., Huber, R., Laskowski, M., Jr. & Ardelt, W. (1985). The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). *Eur. J. Biochem.* **147**, 387-395.