University of Alberta

Structural studies of protein:protein interactions

by

Katherine Sophie Bateman

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
Fall, 1999

University of Alberta

Library Release Form

Name of Author: Katherine Sophie Bateman

Title of Thesis: Structural Studies of Protein: Protein Interactions

Degree: Doctor of Philosophy

Year this Degree Granted: 1999

Katherine Bateman

#205-11104-84 Ave.
Edmonton, Alberta,
T6G 2R4

June 3 1999

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Structural studies of protein:protein interactions submitted by Katherine Sophie Bateman in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dr. M.N.G. James

Dr. B.D. Sykes

Dr. R.C. Bleackley

Dr. D.R. Bundle

Dr. I.A. Wilson

June 3, 1999

Dedicated to
Meleagris gallopavo,
the turkeys

Abstract

Turkey ovomucoid third domain (OMTKY3) is a potent inhibitor of serine proteinases. Coworkers of Michael Laskowski Jr. have synthesized numerous variants of OMTKY3 and have measured association constants for the mutants in complex with six serine proteinases. Wuyuan Lu has extended this data set to include two backbone variants of OMTKY3. In order to rationalize the trends observed in this large data set, high resolution crystal structures have been determined for the OMTKY3 P1 variants in complex with the bacterial serine proteinase *Streptomyces griseus* proteinase B (SGPB). The P1 position is the major specificity determinant for the canonical inhibitors.

Thirteen SGPB:OMTKY3 crystal structures have been determined previously, in the lab of Michael James and nine more are reported in this thesis. These structures include the β-branched, aromatic and positively-charged P1 variant complexes. Structures of backbone variants of OMTKY3 are also reported, one in complex with SGPB and another as a free monomer.

Structures of the β-branched P1 variant complexes reveal that the shape of these P1 residues are not complementary to the S1 binding pocket of SGPB, in agreement with the low association constants for this group. The aromatic P1 variant complexes have relatively high association constants. However, the large rings of the aromatic side chains require a slightly unfavourable χ_2 angle to fit into the S1 pocket. Positively-charged P1 residues are stabilized in the hydrophobic pocket of SGPB by hydrogen bonds to water molecules. An additional structure of the lysine variant complex was determined at a pH above the measured pK_a of the

P1 side chain. The deprotonated side chain no longer forms hydrogen bonds with water molecules but instead forms one with Asn^{36I} of the inhibitor. In this way, the hydrophobic S1 pocket only makes contact with the hydrophobic methylene atoms of the P1 lysine.

An ester bond variant of OMTKY3 was synthesized by Wuyuan Lu in order to examine the energetic contribution of one hydrogen bond between enzyme and inhibitor. A structure of this OMTKY3-COO-Leu^{18I} variant with SGPB confirms that only one hydrogen bond was removed and no other distortions are introduced with the mutation. A value of 1.5 kcal/mol can thus be attributed as the cost of losing one hydrogen bond. A reduced peptide bond variant, OMTKY3-CH_2-Asp^{19I}, was unable to inhibit chymotrypsin. The structure of this molecule did not reveal any gross structural distortions to the inhibitor and the lack of inhibition was attributed to the loss of an interaction between enzyme and substrate, that of the P1 carbonyl oxygen with the oxyanion hole, and, to an unfavourable interaction between the positively-charged amine group and the oxyanion hole.

Lastly, a structure of human pepsinogen A has been determined. The overall structure largely resembles that of other aspartic proteinase zymogens. An N-terminal prosegment prevents access to the preformed substrate binding site and residues from the prosegment stabilize the catalytic aspartic acid residues. An unfavourable χ_2 angle is imposed on Phe111 in the structures of human and porcine pepsinogen but structures of pepsin indicate that Phe111 adopts a favourable angle in the mature enzyme. This interaction has been implicated in the activation process of the pepsinogens.

Acknowledgements

Firstly, I would like to thank my supervisor, Michael James. I admire his expert knowledge of protein structure and his ability to share this knowledge with enthusiasm and I thank him for having the patience to allow and, especially, encourage independent learning.

The structures presented in this thesis could not have been determined without our collaborators, Nadia Tarasova, Michael Laskowski, Jr., and Wuyan Lu and I thank them for their interest and inspiration.

I thank the members of my supervisory committee, Randy Read, Doug Scraba, Chris Bleakley, Brian Sykes and the members of my examining committee, David Bundle and Ian Wilson for taking the time to follow my progress and read my thesis.

The James Lab has been a supportive and helpful group and I appreciate the numerous times that they have showed me how to use programs and equipment, listened to practice talks, collected data, grew crystals, told funny stories and many, many other things. Thank you Amir, Brian, Craig, Ernst, Jens, Jo, Ken, Kui, Maia, Marie, Masao, Nina, Stan and Steve.

Of course, I also found friends outside of the James lab in the other crystallography groups, neighbouring labs, soccer teams, fencing club, soft-ball team, fellow students, grocery store, etc., etc. Thank you very much.

I will *always* be indebted to my irreplaceable family and friends, and not just for the past 7 years. You are all funny and beautiful. To my parents, Sandra, Ian, Grammie, Cousins, Aunts, Uncles, Alex, Nina, Nathan, Alex, Melissa, Lynn, Claire, Brendan, Christine, Orly, Allan, Wendy, Wolf, Mr. Pink, the Sticks, and all the other pets that I have been fortunate enough to meet: Thank you, Thank you, Thank you!

Table of Contents

List of Tables

List of Figures

LIST OF SYMBOLS AND ABBREVIATIONS

Å	Ångstrom (10^{-10} meters)
Abu	α-amino butyric acid
Ahp	α-amino heptanoic acid
Ahx	α-amino hexanoic acid (norleucine)
Ape	α-amino pentanoic acid (norvaline)
B	Thermal motion parameter
BPTI	Bovine pancreatic trypsin inhibitor (Kunitz)
CHYM	α-chymotrypsin
$\|Fo\|$	Observed structure factor amplitude
$\|Fc\|$	Calculated structure factor amplitude
HLE	Human leukocyte elastase
K_a	Association equilibrium constant
K_a^{obs}	Observed association equilibrium constant
k_{cat}	First-order catalytic rate constant
K_m	Michaelis constant
$Lys^{18I\circ}$	Lys^{18I} with a neutral side chain ($-NH_2$)
Lys^{18I+}	Lys^{18I} with a positively charged side chain ($-NH_3^+$)
NMR	Nuclear magnetic resonance
P1	The residue contributing the carbonyl group to the reactive-site peptide bond (Schecter and Berger, 1967)
OMJPQ3	Japanese quail ovomucoid third domain
OMTKY3	Turkey ovomucoid third domain
OMSVP3	Silver pheasant ovomucoid third domain
PEG	Polyethylene glycol
PPE	Porcine pancreatic elastase
R	Standard crystallographic residual
rms	root-mean-square
S1	The preformed cavity in the enzyme to which the P1 residue binds (Schecter and Berger, 1967)
SGPA	*Streptomyces griseus* proteinase A
SGPB	*Streptomyces griseus* proteinase B
SGPE	*Streptomyces griseus* proteinase E
α_c	Calculated structure factor phase

Chapter 1:
Introduction

1.1 Thesis objectives

Twelve crystal structures, featuring various aspects of protein:protein interactions, are presented in this thesis. Chapter 2 describes the structure of human pepsinogen A, an aspartic proteinase zymogen. Chapters 3 to 6 present structures of *Streptomyces griseus* proteinase B (SGPB) in complex with 9 P1 variants (one was done at pH 7.4 and at pH of 10.7) of the ovomucoid third domain from turkey (OMTKY3). In addition, the structure of an OMTKY3 peptide bond variant that does not function as an inhibitor is presented in Chapter 6.

Several structures of aspartic proteinases zymogens, mature enzymes and activation intermediates have previously been determined (James and Sielecki,1986; Sielecki *et al.*, 1991; Fujinaga *et al.*, 1995; Moore *et al.*, 1995; Khan *et al.*, 1997; Hartsuck *et al.*, 1992; Abad-Zapatero *et al.*, 1990; Cooper *et al.*, 1990). Electrostatic interactions play an important role in the conversion from precursor to mature enzyme but many details of this process are not fully understood. Additional structures from different species or of different isoforms can help to explain the variations observed in the conversion from zymogen to mature proteinase (Foltmann and Jensen, 1982; Athauda *et al.*, 1989; Kagayama and Takahashi, 1989). Although human pepsinogen A is a single polypeptide chain, structural information from this zymogen can contribute towards understanding temporary protein:protein interactions as the association of the prosegment domain with the enzyme is very dependent on its environment.

Structures of SGPB in complex with some of the hydrophobic P1 variants of OMTKY3 are presented in Chapters 3 and 4. Chapter 3 includes the β-branched P1 variant complexes and Chapter 4 addresses the aromatic P1 variant complexes. These two groups have very different association equilibrium constants (Lu *et al.*, 1995). The aromatic P1 variants are

among the tightest-binding OMTKY3 variants with SGPB and the β-branched P1 variants are among the lowest. Reasons for these trends are examined in chapters 3 and 4. The differences in binding are probably more closely related to the shape of the P1 residues rather than the hydrophobicity of the side chains.

Chapter 5 examines what happens when positively-charged residues are introduced into a hydrophobic pocket. Structures of SGPB in complex with arginine and lysine P1 variants of OMTKY3 are presented. Comparisons are made with structures of SGPB in complex with negatively-charged and neutral P1 variants of OMTKY3, in addition to structures of complexes with other serine proteinases, having different P1 preferences, with positively charged P1 inhibitors.

An OMTKY3 variant having an ester bond replacing a peptide bond in order to remove a specific hydrogen bond common to complexes of canonical inhibitors with serine proteinases was synthesized by Wuyuan Lu (Lu *et al.*, 1997). The structure of OMTKY3:Thr^{17I}-COO-Leu^{18I} in complex with SGPB was determined to ensure that the mutation did not introduce any other distortions to the complex other than the absence of the hydrogen bond. This structure allows thermodynamic measurements (Lu *et al.*, 1997) to be directly related to one hydrogen bond.

Finally, Chapter 6 presents the structure of a backbone variant of OMTKY3, OMTKY3-CH_2-Asp^{19I}, that does not function as an inhibitor. The types of favourable protein:protein interactions that the mutation might have disrupted are examined.

1.2 Protein:protein interactions

Many biochemical processes rely on the specific but temporary association of protein molecules. Representative examples include interactions between hormones and receptors, enzymes and inhibitors, and enzymes and protein substrates. Most studies of protein:protein interactions have been performed on dimers, although the principles that govern the

association of two proteins are believed to extend to oligomeric complexes involving a larger number of molecules. The temporary protein interactions differ from the more permanent associations by the size of the buried surface area (Janin, 1995; Jones and Thornton, 1996) and by the type of residues that are found at an interface.

In general, the size of the interface of a permanent dimeric interaction increases with the molecular weight of the monomers (Janin, 1995; Jones and Thornton, 1996). Alternatively, antibodies and antigens, proteinases and inhibitors, and redox complexes all have interface areas of approximately 1600 $Å^2$ (Janin, 1995).

The type of residues found at the interface of transient protein:protein interactions are less hydrophilic than residues found on a protein's surface and less hydrophobic than residues found either in the interior of a protein or at a permanent dimer interface. The driving force for permanent protein associations and for protein folding is dominated by hydrophobic interactions. In contrast, electrostatic interactions, hydrogen bonds, hydrophobic interactions and van der Waals interactions all contribute towards the stabilization of the temporary protein complex.

Electrostatic interactions or salt bridges involve charged atoms. Opposite charges attract and like charges repel each other. Coulomb's Law calculates the energy of charged interactions as an inverse function of the distance between the charges and the dielectric constant of the medium between the charges.

Coulomb's Law

$$Eelec_{ij} = \frac{kq_iq_j}{\varepsilon r_{ij}}$$

k is a proportionality constant
q_i and q_j are the charges
ε is the dielectric constant of the solvent (= 1.0 for a vacuum)
r_{ij} is the distance between the charges q_i and q_j

Charged interactions distinguish protein binding from protein folding (Tanford, 1957). The desolvation of charged interactions during protein folding is quite destabilizing because the hydrophobic protein interior has a low dielectric constant and the ion pair cannot make interactions with the hydrophobic atoms that would compensate for lost interactions with the solvent. However, a protein interface is more hydrophilic than the protein core which means that polar atoms can interact with the charged residues and, therefore, can replace the solvent atoms that interact with the charges on an exposed surface (Xu *et al.*, 1997). Specific protein:protein interactions are encouraged by specific hydrophilic environments such that interactions between like charges are avoided (Xu *et al.*, 1997).

Hydrogen bonds are electrostatic interactions formed between a proton donor that has a partially-positive character, typically NH or OH and a proton acceptor that has a partially-negative character, typically =O or -O-. Some features of a hydrogen bond are distances in the range of 2.4 to 3.6 Å, from the donor atom to the acceptor atom, an angle greater than 100° for atoms D-H...A, and an angle between 130° and 170° for atoms D...A-C (D=donor atom, A=acceptor atom and C=carbon atom) (McDonald and Thornton, 1994). The distance between donor and acceptor atoms of a hydrogen bond is usually shorter or equal to a van der Waals contact.

Hydrogen bonds are responsible for stabilizing the secondary structures of α-helices and β-sheets. Hydrogen bonds between side-chain atoms stabilize side-chain conformations and can also contribute to the overall stability of a protein fold. This type of stabilization is indeed the case for residues of the serine proteinase inhibitor OMTKY3, and is discussed in chapter 6. A survey of protein structures from the PDB determined that 92% of potential hydrogen bonding atoms do participate in hydrogen bonds (McDonald and Thornton, 1994) and it has been estimated that a single hydrogen bond contributes approximately 1.5 kcal/mol towards conformational stability (Lu *et al.*, 1997b; Fersht *et al.*, 1985; Shirley *et al.*, 1992; Fersht, 1987; Groeger *et al.*, 1994). In Chapter 6, this value is verified

for the contribution of a single hydrogen bond towards the binding energy of the SGPB:OMTKY3 complex.

The polar interactions at an interface are balanced by hydrophobic interactions. Perhaps the hydrophobic interactions are more accurately described as dehydration reactions because it is the unfavourable entropy of ordering water molecules around a hydrophobic side chain (or lipid) and favourable entropy of releasing these water molecules that causes the hydrophobes to cluster together. The hydrophobicity of temporary protein interfaces are intermediate to the hydrophobicity of protein surfaces not involved in dimerization and protein interiors (Jones and Thornton, 1995 and 1996). Hydrophobic surface patches are also more prominent for permanent oligomeric interfaces than for temporary ones.

Van der Waals interactions are weak forces that occur between all atoms, regardless of charge. Attraction of atoms occurs at defined distances, known as the sum of the van der Waals radii. However, if the atoms get too close, they experience repulsion. The energy of these forces are calculated with the Lennard-Jones potential function.

Lennard-Jones Potential:

$$Evdw_{ij} = \frac{-A}{r_{ij}^{6}} + \frac{B}{r_{ij}^{12}}$$

in which r_{ij} is the distance between atoms i and j, and
A and B are proportionality constants related to the van der Waal radii (atom type)

The attractive force varies with r^{-6} and the repulsive force is much more sensitive and varies with r^{-12}. In order for van der Waals interactions to contribute towards the binding energy, the shape of the interface surfaces must be complementary. Gaps or cavities do not contribute towards stability and, in fact, have been shown to be deleterious (Eriksson *et al.* 1992).

1.3 Aspartic proteinase zymogens

The aspartic proteinases from mammals are all synthesized as inactive precursors. Crystal structures of the mammalian, gastric zymogens, porcine pepsinogen (James and Sielecki 1986; Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992) and human progastricsin (Moore *et al.*, 1995), have been determined previously and both proteins have the same overall conformation. The molecules consist of two large N- and C-terminal lobe domains with similar β-sheet topology and an active site cleft is formed at the intersection of the two domains. The lobes have probably arisen from gene duplication early in the evolution of aspartic proteinases (Tang *et al.*, 1978; Hobart *et al.*, 1984; Holm *et al.*, 1984). However, the features described thus far are shared by mature enzymes, including the locations of the catalytic aspartic acid residues, Asp32 and Asp215 (pepsin numbering) (Abad-Zapatero *et al.*, 1990; Cooper *et al.*, 1990; Sielecki *et al.*, 1990; Fujinaga *et al.*, 1995). An N-terminal extension called the prosegment distinguishes zymogen from enzyme. The prosegment fills the active site cleft and prevents substrate from binding. A highly-conserved lysine residue on the prosegment, Lys37P, forms salt-bridges with both aspartic acids and the catalytic residues are further stabilized by hydrogen bonds to two conserved tyrosine residues. Conversion from precursor to mature enzyme involves a conformational change (Perlmann *et al.*, 1963; McPhie, 1972). The prosegment residues must move from their position in the zymogen and allow access to the substrate binding site. The aspartic acid residues, Asp32 and Asp215 must also become catalytically active. The catalytic aspartic acid residues are responsible for activating a water molecule (Davies, 1990). The water molecule acts as the nucleophile and attacks the carbonyl carbon of the substrate's scissile bond. The newly-formed tetrahedral intermediate does not involve a covalent bond with the enzyme.

A recently determined structure of proplasmepsin II from *Plasmodium falciparum* demonstrates a novel inhibitory mode for an aspartic proteinase zymogen (Khazanovich-Bernstein *et al.*, 1999). In this

structure, the prosegment interacts with the C-terminal lobe and prevents the N- and C-terminal lobes from interacting as they would in a mature enzyme.

1.3.1 *Conversion of the gastric aspartic proteinase zymogens*

Formation of Intermediate I

When the zymogen is exposed to acidic conditions (pH < 3.5), aspartate and glutamate residues on the enzyme become protonated and lose their negative charge (Foltmann, 1988; Glick *et al.*, 1989). The loss of charge disrupts salt-bridge interactions with the positively-charged residues from the prosegment. In particular, the salt bridge from the prosegment lysine residue to the catalytic aspartic acid residues will be disturbed and the prosegment residues that block the active site cleft will be free to unfold. This unraveled, but intact, zymogen has been named 'intermediate I' (Foltmann and Jensen, 1982). At this point, a rapid increase in pH can restore the zymogen (McPhie, 1972).

Formation of intermediate II

Now that the substrate binding site has been exposed, intermediate I is capable of catalysis. Cleavage occurs at two sites, the pro-mature junction and within a helical region of the prosegment. These cleavages are autocatalytic (Herriott, 1939). A β-strand from the N-terminus of the prosegment still participates in a six-stranded β-sheet with the enzyme. This complex is known as intermediate II (Foltmann and Jensen, 1982). Intermediate II from progastricsin can be trapped and isolated by raising the pH. A structure of the trapped gastricsin intermediate II has been determined, revealing a partially-obstructed S1 pocket but an otherwise accessible active site (Khan *et al.*, 1997). In addition, a water molecule was observed in the same position as in mature pepsin; it forms hydrogen bonds with the catalytic aspartic acid residues, and is ready to act as a nucleophile.

Formation of the mature proteinase

Finally, the remaining prosegment residues dissociate from the enzyme and the β-strand is replaced by residues from the mature N-terminus. These residues ensure that the prosegment can not reassociate with the proteinase and act as an inhibitor.

1.4 Human pepsinogen A

Human pepsinogen A is the precursor of human pepsin and one of several digestive enzymes found in the stomach. The precursor is synthesized in the neutral environment of the chief cells (Waalewijen *et al.*, 1991) in the gastric mucosa and, after ingestion of food, it is then secreted into the acidic environment of the gastric lumen where it becomes converted into pepsin. An acidic environment is the only requirement for activation (Herriott, 1939). At low pH, the pepsinogens can activate each other (Al-Janabi *et al.*, 1972; Kageyama and Takahashi, 1987), but they are also capable of intramolecular activation (Bustin and Conway-Jacobs, 1971; Al-Janabi *et al.*, 1972; McPhie, 1972; Christensen *et al.*, 1977). Activation occurs either with one cleavage reaction and removal of the entire prosegment (Kageyama and Takahashi, 1987) (residues 1P-47P) or it can occur via an intermediate. The human pepsinogen A intermediate is initially cleaved between residues Leu23P and Lys24P. The prosegment residues 24P to 47P are then removed by a subsequent cleavage at the pro-mature junction (Foltmann, 1988b; Athauda *et al.*, 1989).

Five isoforms of human pepsinogen A have been identified and these were originally classified according to their electrophoretic mobility (Samloff, 1969). Further separation and classification of the isoforms have been performed with other techniques (Athauda *et al.*, 1989; Roberts *et al.*, 1995), but the actual number of isoforms is unclear. The isoforms separated by electrophoretic mobility have been numbered by decreasing anodal mobility, with isoform 5 being the most basic, and the most abundant (Kageyama *et al.*, 1989). Isoform 3 is the next abundant; it is the isoform whose structure is reported in this thesis. The remaining

isoforms are found in trace amounts, particularly isoform 1. Isoform 1 is also the only isoform believed to be associated with significant amounts of carbohydrate (Peek *et al.*, 1989; Pearson *et al.*, 1992) and it has an increased ability to degrade the mucoproteins which protect the stomach tissues (Pearson *et al.*, 1986; Pearson *et al.*, 1990; Pearson *et al.*, 1992). For these reasons, isoform 1 has been associated with the formation of ulcers (Walker *et al.*, 1980).

1.5 Serine proteinases

The serine proteinases comprise six clans and approximately thirty evolutionary families (Barrett and Rawlings, 1995). They provide an excellent example of convergent evolution. The three-dimensional arrangement of the catalytic serine and histidine residues is the same for enzymes from the subtilisin, chymotrypsin, and serine carboxypeptidase clans, although they all have different overall folds. Proteinases from the subtilisin clan have a single domain consisting of nine α-helices packed against a seven stranded, parallel β-sheet (Wright *et al.*, 1969; McPhalen and James, 1988). An active site cleft is formed at the junction of two β-barrel domains in the chymotrypsin clan (Matthews *et al.*, 1967; Steitz and Shulman, 1982) and the α/β hydrolase fold is found in members belonging to the serine carboxypeptidase clan (Liao *et al.*, 1992). An oxyanion hole also has been formed for each of these clans but the residues which make up the carbonyl oxygen atom binding site are different. In addition, an aspartic acid residue is always present, but the position of this residue with respect to the serine and histidine has not been strictly maintained. It is plausible that other folds may arise which support the same catalytic machinery. An antibody was generated that was capable of serine proteinase activity (Guo *et al.*, 1994). The crystal structure of the immunoglobulin revealed a serine-histidine dyad (Zhou *et al.*, 1994) but mutational analysis later showed that the serine residue was not neccessary for the catalytic activity of the antibody (Baca *et al.*, 1997).

Many structures have been determined of serine proteinases and the chymotrypsin clan has been particularly well studied. They all share two β-barrel domains, as previously discussed, but they display a range of substrate specificity, mostly as a result of differences in the S1 pocket. The chymotrypsin S1 pocket is hydrophobic and prefers hydrophobic residues at the P1 position, as do the bacterial proteinases SGPB and SGPA. Trypsin prefers postively-charged P1 side chains that can form a salt bridge with an aspartate residue at the bottom of the S1 pocket. Elastase has a small S1 pocket and, therefore, prefers small aliphatic residues. SGPE has a triad of histidines near the bottom of its S1 pocket and binds glutamate at P1 .

1.5.1 *Catalytic mechanism of the serine proteinases*

Michaelis-Menten complex

The first step of peptide bond cleavage by the serine proteinases is the formation of a Michaelis-Menten complex (Figure 1.1). The enzyme:substrate complex is stabilized by backbone interactions that form short segments of antiparallel β-sheet, in addition to interactions between the enzyme specificity pockets and the substrate side chains. The carbonyl oxygen of the bond to be cleaved accepts two hydrogen bonds from the oxyanion hole. A result of all of these interactions is that the carbonyl carbon of the scissile peptide bond is placed in a good position for nucleophilic attack by the Oγ of the catalytic serine residue. An angle of approximately 107° for the nucleophilic atom-carbonyl carbon-carbonyl oxygen is optimal for nucleophilic attack on a carbonyl group (Burgi *et al.*, 1973).

Nucleophilic attack

Nucleophilic attack by Ser195 Oγ is assisted by several factors including a hydrogen bond between His57 and Ser195 Oγ, a peptide bond between residues P2 and P1, and the oxyanion hole. The hydrogen bond between His57 Nε2 and Ser195 Oγ increases the nucleophilicity of Ser195 but a strong interaction between these residues is only formed in the presence of substrate. A peptide dipole from the peptide bond between P2 and P1

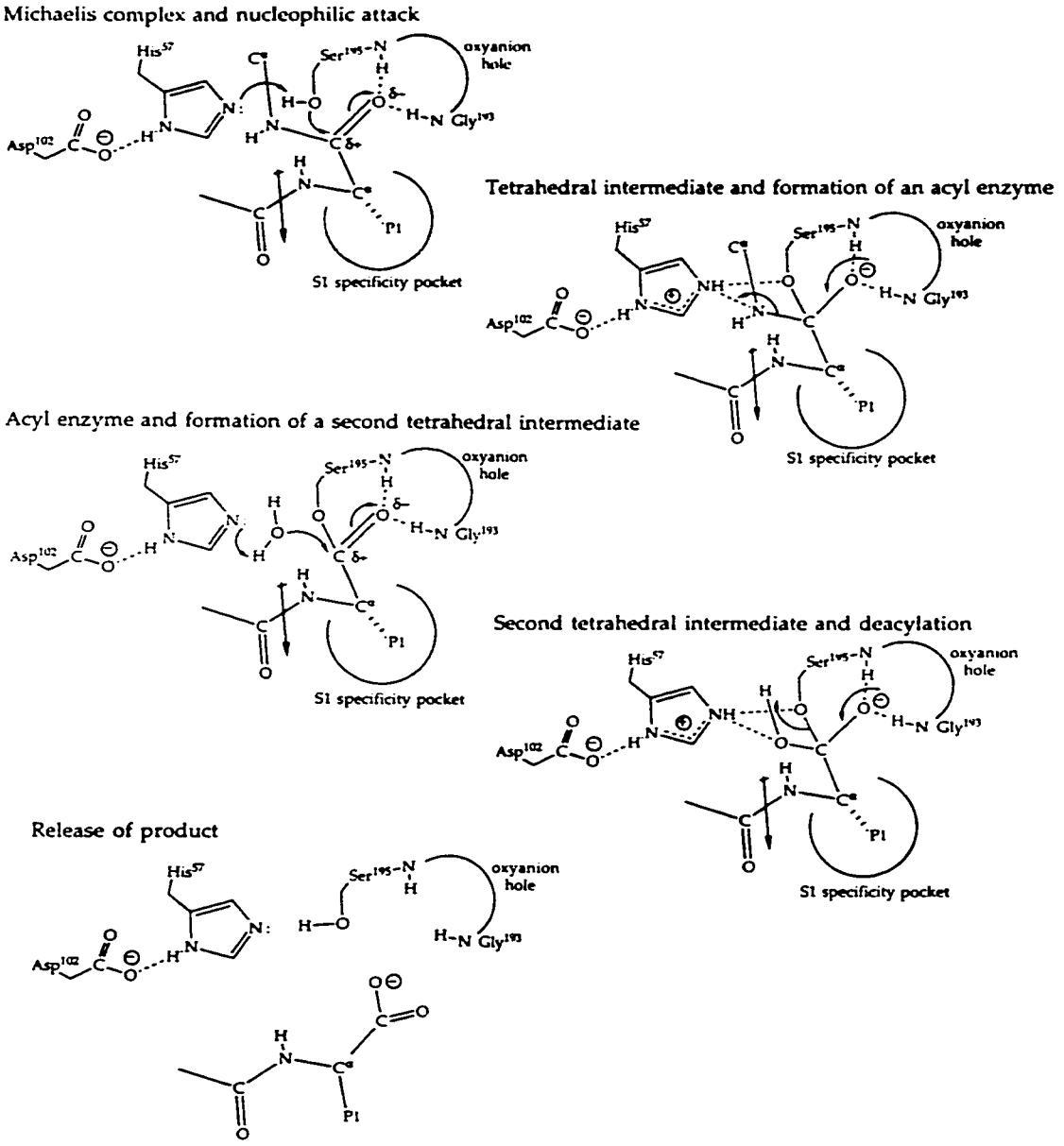

Figure 1.1 Catalytic mechanism of the serine proteinases.

points towards Ser195 Oγ and helps to provide an ideal environment for the generation of the nucleophile. These factors are especially important since serine normally has a pKa of ~15. Finally, a partial-positive charge on the carbonyl carbon of the scissile bond is enhanced by hydrogen bonding interactions of the carbonyl oxygen with the oxyanion hole. After nucleophilic attack, a covalent bond is formed between Ser195 Oγ and the carbonyl carbon resulting in a tetrahedral intermediate. The proton from Ser195 is transferred to His57 Nε2 and the resulting positive charge on the histidine is stabilized by a salt bridge with Asp102. A negative charge develops on the carbonyl oxygen of the substrate as the C=O double bond becomes a single bond and the interactions at the oxyanion hole are strengthened. Mutational studies of the three catalytic residues, Ser195, His57 and Asp102 have demonstrated their importance (Carter and Wells, 1987; Corey and Craik, 1992; Craik *et al.*, 1987). Mutation of these residues always results in a decrease of kcat. Any of the residual catalytic activity is believed to be associated with the oxyanion hole through stabilization of the transition state (Carter and Wells, 1987).

Acyl-enzyme intermediate

The next step involves formation of an acyl-enzyme intermediate. The proton from His57 is passed to the leaving group nitrogen and the C=O double bond is reformed as the scissile bond is cleaved. The newly-formed amino-terminus and residues on the primed side of the scissile bond are then released from the enzyme as products. The carbonyl-carbon atom of the acyl-enzyme intermediate still carries a partial-positive charge in preparation for the second nucleophilic attack. Acyl-enzyme adducts have been observed in structures of chymotrypsin and SGPA (Dixon *et al.*, 1991; Harel *et al.*, 1991; Blanchard and James, 1994).

Deacylation

The second tetrahedral intermediate results from the general-base assisted nucleophilic attack by a water molecule. The water enters the active site from the side of the leaving group (Perona *et al.*, 1993; Dixon *et al.*, 1991) and initially forms a hydrogen bond to His57 Nε2. A proton is donated to the histidine and the negative OH attacks the partial-positive

charge on the carbonyl carbon to form the second tetrahedral intermediate. Again, the negative charge on the carbonyl oxygen is stabilized by the oxyanion hole. His57 Nε2 transfers its proton to Ser195 Oγ and the final carboxylate product is released leaving the proteinase ready to catalyse another cleavage reaction.

1.6 Canonical inhibitors of serine proteinases

At least sixteen families of protein inhibitors are targeted towards serine proteinases (Laskowski and Kato, 1980), and can be grouped according to sequence, topology and method of binding. Table 1.1 lists some of the known families of canonical inhibitors. Structures are available for members from most of these families and common features of the canonical inhibitors are a compact shape and an exposed binding loop. Stability is often, but not always, achieved with numerous disulphide bonds and the overall conformations of the inhibitors can be quite different (Bode and Huber, 1992).

The canonical inhibitors are so named because they follow a standard inhibitory mechanism of tight binding accompanied by slow hydrolysis. Tight binding and formation of a Michaelis complex is a result of the exposed binding loop being highly complementary to the active site cleft of the proteinase. β-sheet hydrogen bonding interactions are formed with backbone atoms and side chains interact with specificity pockets. Interactions with the P1 residue are particularly important for inhibitors directed towards proteinases from the chymotrypsin clan, whereas interactions with the P4 residue are more dominant with the subtilisins. Generally, the contribution of individual side chains from the inhibitor binding loop can be considered to be additive (Lu, 1994). The exposed and extended conformation of the loop allows for minimal interaction between the specificity determinants. However, the canonical inhibitors are substrates, albeit poor ones. The specificity constants, k_{cat}/K_m for the inhibitors are comparable to those expected for good substrates, except that the individual values of k_{cat} and K_m are both very small. One

Table 1.1: Canonical inhibitors of serine proteinases

Source	Family
Microorganisms	Streptomyces subtilisin inhibitor (SSI) family*
	Ecotin family*
	Marinostatin family
Plants	Bowman-Birk family*
	Soybean trypsin inhibitor (STI) or Kunitz family*
	Squash family*
	Potato I family*
	Potato II family*
	Barley family
Animals	Bovine pancreatic trypsin inhibitor (BPTI) or Kunitz family*
	Pancreatic secretory trypsin inhibitor (PSTI) or Kazal family*
	Ascaris family*
	Chelonianin family*
	Antistasin family*

*Members from these families have structures determined (Bode and Huber, 1992 and references therein; Huang, 1995; Shin *et al.*, 1996; Lapatto *et al.*, 1997).

This table was adapted from Huang, 1995 and Lu, 1994.

difference between the canonical inhibitors and a good substrate is that the substrate lacks the close packed and extensive interactions of the proteinase:inhibitor complex. This complementarity contributes towards the enthalpy of forming a Michaelis complex. The entropic cost of binding is also much greater for the substrate because the segment to be cleaved is not restricted to a small number of conformations. Canonical inhibitors typically have some flexibility in the exposed binding loop which allows them to adapt to different proteinases (Fujinaga *et al.*, 1987) but the variation is limited by the relatively rigid scaffolding of the rest of the protein.

The proteinase:inhibitor Michaelis complex is so complementary that the activation energy towards formation of a tetrahedral transition state is increased. The canonical loop is very rigid within the complex through interactions with both inhibitor and enzyme. In order to reach the geometry of the transition state, hydrogen bonds would have to be broken and the inhibitor scaffold would have to undergo some strain.

Once cleavage has taken place, a second problem arises. The leaving group is not entirely free to leave. Many interactions keep the P1' residues near the active site, such as hydrogen bonding interactions with the enzyme and disulphide bridges to the inhibitor. This property has even been manipulated to form covalent, intact inhibitors from two peptide fragments in the enzymatic semi-synthesis of OMTKY3 variants (Ardelt and Laskowski, 1983; Bigler *et al.*, 1993).

1.7 *Streptomyces griseus* proteinase B

Streptomyces griseus proteinase B (SGPB) is one of the many proteinases found in the extracellular filtrate of *Streptomyces griseus*. The extracellular filtrate, commercially known as Pronase, also contains the serine proteinases SGPA, and SG trypsin. The amino acid sequence for SGPB was determined in Dr. Smillie's lab at the University of Alberta (Jurasek *et al.*, 1974).

SGPB belongs to the SA clan and shares a similar substrate specificity with chymotrypsin, the enzyme traditionally used to represent the clan. Medium- to large-sized hydrophobic residues, such as leucine, methionine, phenylalanine and tyrosine are preferred at the P1 position for these enzymes. However, the sequence identity between SGPB and chymotrypsin is low, at 20%. SGPB has a higher sequence identity with members of its own family, the S2 family, such as α-lytic protease (43%) and SGPA (61%). All members of the S2 family are synthesized as pre-pro-mature proteins. The 'pre' signal sequence directs the protein to the cell surface and the 'pro' sequence is believed to assist in the folding process of the proteinase (Baker *et al.*, 1992; Baardsnes *et al.* 1998). The pro sequences are more diverse among the family (Sidhu *et al.*, 1995). For example, α-lytic protease has a pro region of 166 amino acids (Baker *et al.*, 1992) and SGPB has only a 76 amino-acid pro region (Baardsnes *et al.*, 1998).

The James lab has determined several structures of SGPB, alone and in complex with protein inhibitors. These structures are free SGPB determined in an orthorhombic space group at pH 4.5 (Codding *et al.*, 1974; Delbaere *et al.*, 1979), free SGPB determined in a cubic space group at pH 7.0 (Blanchard and James, in preparation), SGPB in complex with a chymotrypsin inhibitor from potato tubers (Greenblatt *et al.*, 1989) and SGPB in complex with wild type OMTKY3 (Fujinaga *et al.*, 1982; Read *et al.*, 1983). Ten structures of SGPB in complex with several recombinant OMTKY3 variants have been determined by Kui Huang, also in the James lab (Huang, 1995; Huang *et al.*, 1995). This thesis extends the subset of structures to include a total of nineteen SGPB:OMTKY3 complexes.

1.8 Turkey ovomucoid inhibitor third domain

The third domain of turkey ovomucoid (OMTKY3) is a serine proteinase inhibitor belonging to the Kazal family of canonical protein inhibitors. OMTKY3 has a leucine at the P1 position and inhibits the enzymes chymotrypsin, subtilisin, porcine pancreatic elastase, human leucocyte elastase, SGPB and SGPB.

Ovomucoids are the major glycoprotein found in avian egg whites and are made up of three tandem Kazal domains of approximately sixty residues each. The third domain can be obtained by limited proteolysis and purification. Sequences have been determined for ovomucoid third domains from 101 avian species (Laskowski, Jr., *et al.*, 1987; Kato *et al.*, 1987). The greatest variation between ovomucoids from different species is within the region that contacts the proteinase. Thus, the ovomucoids, as a group, are able to display variation in their activity and specificity towards a wide range of serine proteinases. This observation distinguishes the evolution of the ovomucoids from that of other proteins because it is the functional region that exhibits the greatest sequence variation among ovomucoids from different species (Laskowski Jr. *et al.*, 1987). However, the structurally-important residues are highly conserved. Three disulphide bridges are present for all of the ovomucoid third domains. Other residues that participate in hydrogen bonding interactions and hydrophobic interactions in the small protein core have also been maintained (Laskowski, Jr. *et al.*, 1987).

OMTKY3 is being utilised to establish a sequence-reactivity algorithm. The ovomucoids display natural variation in enzyme specificity and, therefore, are an excellent choice for this endevour. In addition, the individual interactions of the specificity determinants do not interfere with each other. Dr. S. Anderson's group at Rutgers University has cloned a gene for OMTKY3 residues 6I to 56I. Each of the 20 coded amino acids have been substituted at the P1 position. In addition, 5 non-coded amino acids, Abu, Ape, Ahx, Ahp and homoserine, have been introduced at the P1 position by enzymatic semi-synthesis (Wieczorek *et al.*, 1987; Bigler *et al.*, 1993). Other positions have also been substituted, but this thesis will focus on the P1 variants. Dr. M. Laskowski, Jr.'s laboratory has measured the association equilibrium constant for each of the OMTKY3 variants in complex with six cognate enzymes as outlined in Table 1.2. Association equilibrium constants for OMTKY3 P1 variants with SGPB are in Figure 1.2. The thermodynamic association constants, Ka, are determined as follows (Empie and Laskowski, 1982; Ardelt and Laskowski, 1985):

Known concentrations of enzyme and inhibitor are mixed and equilibrated. Canonical inhibitors interact with the serine proteinases according to:

$$E + I \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} EI \underset{k^*_{on}}{\overset{k^*_{off}}{\rightleftharpoons}} E + I^*$$

E is the enzyme
I is the intact inhibitor
I* is the modified inhibitor (cleaved at the scissile bond)

Aliquots were measured for proteolytic activity using fluorogenic stubstrates to determine the concentration of free enzyme. K_a is calculated by measuring the free enzyme concentration at various concentrations of inhibitor. K_a is defined as:

$$K_a = \frac{[EI]}{[E] \times [I]} = \frac{[E_o] - [E]}{[E] \times ([I_o] - [E_o] + [E])}$$

$[E_o]$ is the total enzyme concentration
$[I_o]$ is the total inhibitor concentration
[E] is the concentration of free enzyme at equilibrium

The observed association constant is defined as:

$$K_a^{obs} = \frac{[EI]}{[E] \times ([I] + [I^*])} = \frac{K_a}{1 + [I^*] / [I]}$$

At neutral pH, hydrolysis of the inhibitor is very slow, therefore modified inhibitor is virtually undetectable and $K_a^{obs} = K_a$. The proper interpretation of this wealth of information on association equilibrium constants for the interacting pairs of proteins can only be done by determining high resolution crystal structures of the complexes; several are presented in this thesis.

Table 1.2 Association equilibrium constants, Ka* (M^{-1}), for 25 OMTKY3 P1 variants with 6 serine proteinases.

Enzymes:	CHYM	PPE	CARL	SGPA	SGPB	HLE
P_1:						
Gly	6.6×10^{6}	9.0×10^{8}	6.4×10^{8}	4.5×10^{7}	1.2×10^{7}	2.2×10^{7}
Ala	6.1×10^{7}	4.2×10^{10}	2.0×10^{10}	1.9×10^{9}	3.6×10^{8}	1.0×10^{9}
Abu	1.1×10^{9}	$\underline{3.3\times10^{11}}$	1.0×10^{11}	1.0×10^{10}	2.3×10^{9}	1.4×10^{10}
Ser	4.2×10^{7}	8.6×10^{8}	2.4×10^{9}	1.6×10^{8}	5.0×10^{7}	3.3×10^{7}
Cys	2.4×10^{9}	2.5×10^{10}	$\underline{3.3\times10^{11}}$	3.3×10^{11}	$\underline{5.7\times10^{10}}$	7.1×10^{9}
Ape	2.2×10^{10}	2.4×10^{11}	1.4×10^{11}	8.3×10^{10}	1.2×10^{10}	1.0×10^{10}
Val	1.4×10^{8}	8.8×10^{9}	8.9×10^{8}	2.1×10^{9}	3.3×10^{8}	1.4×10^{10}
Hse	2.5×10^{9}	7.5×10^{9}	5.6×10^{10}	6.5×10^{9}	1.7×10^{9}	2.6×10^{8}
Thr	9.8×10^{7}	2.9×10^{10}	2.6×10^{10}	2.0×10^{9}	2.5×10^{8}	1.3×10^{9}
Pro	$\underline{6.7\times10^{4}}$	5.9×10^{5}	$\underline{7.9\times10^{4}}$	$\underline{4.6\times10^{4}}$	$\underline{3.6\times10^{4}}$	1.9×10^{5}
Ahx	8.0×10^{10}	1.9×10^{11}	6.7×10^{10}	2.3×10^{11}	2.1×10^{10}	9.9×10^{8}
Leu	1.9×10^{11}	4.2×10^{10}	3.4×10^{10}	3.0×10^{11}	$\underline{5.6\times10^{10}}$	6.1×10^{9}
Ile	9.9×10^{7}	5.9×10^{9}	1.4×10^{8}	1.5×10^{8}	2.9×10^{7}	$\underline{2.1\times10^{10}}$
Met	1.0×10^{11}	1.3×10^{10}	8.1×10^{10}	2.6×10^{11}	2.7×10^{10}	7.3×10^{8}
Asn	6.6×10^{8}	6.9×10^{7}	2.7×10^{8}	2.6×10^{8}	1.8×10^{8}	9.0×10^{5}
Asp	1.0×10^{9}	7.3×10^{4}	1.6×10^{7}	5.1×10^{5}	3.9×10^{5}	$\underline{1.6\times10^{4}}$
Ahp	3.3×10^{11}	2.4×10^{8}	8.9×10^{10}	$\underline{4.7\times10^{11}}$	2.8×10^{10}	1.3×10^{8}
Lys	1.0×10^{8}	4.6×10^{4}	1.2×10^{8}	5.6×10^{7}	2.6×10^{8}	3.8×10^{5}
Gln	1.3×10^{9}	4.3×10^{7}	1.3×10^{10}	1.2×10^{9}	6.9×10^{8}	2.3×10^{7}
Glu	2.5×10^{6}	8.4×10^{4}	8.6×10^{8}	7.9×10^{6}	2.3×10^{6}	4.1×10^{4}
His	1.2×10^{9}	3.7×10^{4}	1.3×10^{10}	2.6×10^{9}	3.1×10^{9}	1.1×10^{5}
Phe	2.4×10^{12}	4.9×10^{4}	1.2×10^{10}	1.8×10^{11}	5.5×10^{9}	9.9×10^{5}
Arg	2.2×10^{8}	$\underline{4.8\times10^{3}}$	2.4×10^{8}	8.2×10^{7}	1.9×10^{8}	3.3×10^{4}
Tyr	$\underline{8.1\times10^{12}}$	7.5×10^{3}	1.6×10^{10}	1.0×10^{11}	3.3×10^{9}	8.3×10^{4}
Trp	3.3×10^{12}	2.5×10^{4}	2.6×10^{9}	8.9×10^{6}	2.4×10^{6}	1.8×10^{4}

*Experiments were conducted at pH 8.3 and at 21°C. The accuracy of the Ka values was 20%. The largest Ka in each column is doubly underlined and the smallest is singly underlined. CHYM: α-chymotrypsin, PPE: porcine pancreatic elastase, CARL: subtilisin Carlsberg, SGPA: *Streptomyces griseus* proteinase A, SGPB: *Streptomyces griseus* proteinase B, HLE: human leukocyte elastase.

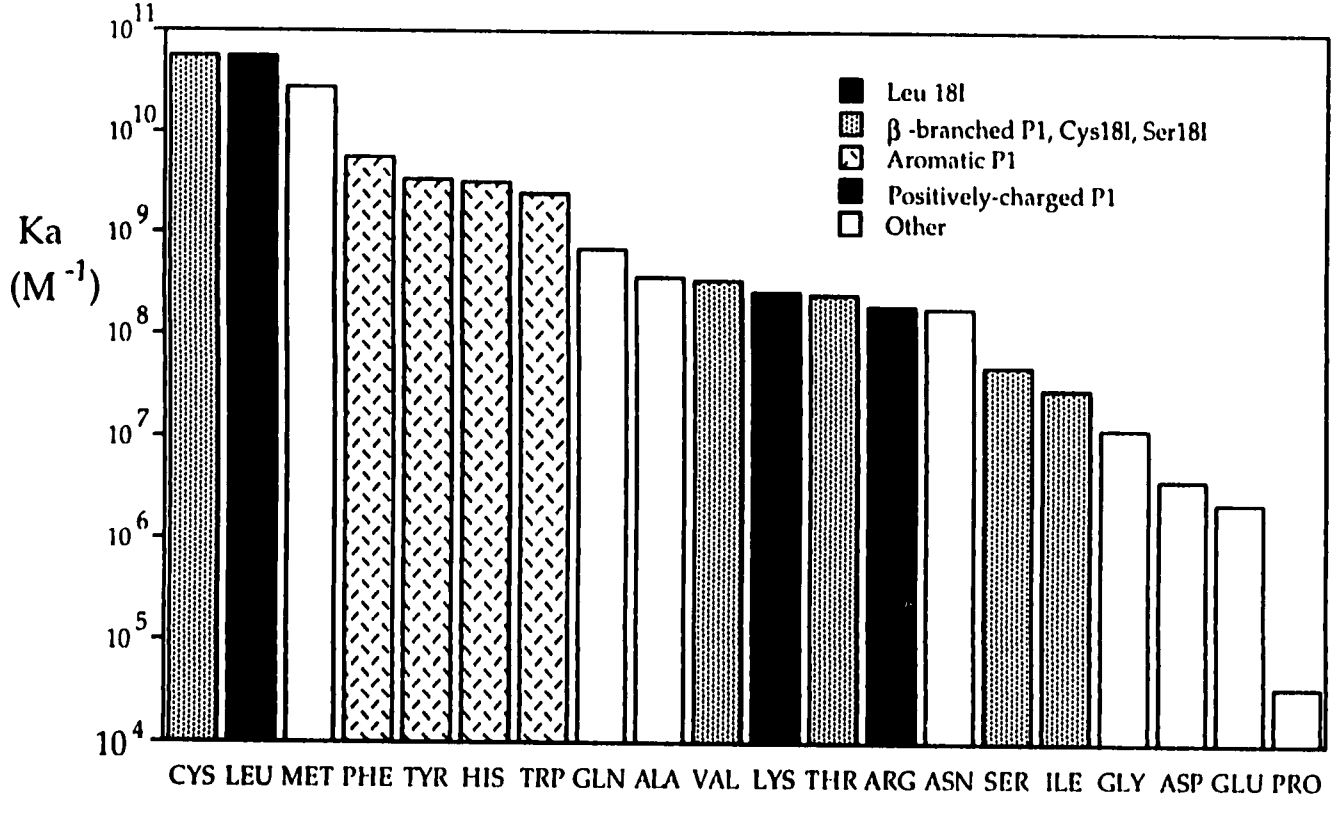

Figure 1.2 Association equilibrium constants (Ka) for OMTKY3 P1 variants with SGPB. The bar headings refer to the SGPB:OMTKY3 complex with that residue at the P1 position.

1.9 Structures of OMTKY3 P1 variants in complex with SGPB

Kui Huang's Ph.D. thesis presented the crystal structures of ten SGPB:OMTKY3 complexes (Huang, 1995). A structure of recombinant OMTKY3-Leu^{18I} (residues 6I-56I) with SGPB was determined to ensure that the absence of the first five residues did not have an impact on the structure in the region of the reactive site loop. Kinetic studies suggested that this was so (Wieczorek *et al.*, 1987), as did the previously determined structure of wild type OMTKY3 in complex with SGPB (Read *et al.*, 1982; Fujinaga *et al.*, 1983). Residues 1I-6I were not included in this model because of a lack of electron density.

Structures were determined for the complexes SGPB:OMTKY3-Gly^{18I} and SGPB:OMTKY3-Ala^{18I}. In order to determine the contribution of hydrophobic interactions to the free energy of binding, models of the straight-chain, aliphatic residues Abu, Ape, Ahx and Ahp were built at the P1 position into the S1 pocket of SGPB. A linear correlation was discovered between the change in free energy of binding (with respect to SGPB:OMTKY3-Gly^{18I} as the reference model) and the buried hydrophobic surface area of the SGPB:OMTKY3-X^{18I} models. The slope of this line gave an estimate of 34.1 cal/mol/Å^2 for the hydrophobic effect.

Structures were also presented of SGPB:OMTKY3-Asn^{18I}, SGPB:OMTKY3-Gln^{18I}, SGPB:OMTKY3-Asp^{18I}, and SGPB:OMTKY3-Glu^{18I}. P1 residues Asp^{18I} and Glu^{18I} displayed elevated pK_as within the S1 pocket of SGPB (Qasim *et al.*, 1995) and structures of SGPB:OMTKY3-Asp^{18I}, SGPB:OMTKY3-Glu^{18I} and SGPB:OMTKY3-Gln^{18I} were additionally determined at pH 10.7, significantly above the pK_a of Asp^{18I} and Glu^{18I} in the complex. SGPB:OMTKY3-Gln^{18I} was determined at two pH levels (6.8 and 10.7) to ensure that any changes in the S1 pocket of the ionizable residues were only due to the de-protonation of the P1 side chains. The two SGPB:OMTKY3-Gln^{18I} structures were identical as expected. In both the Asp^{18I} and Glu^{18I} complexes at high pH, the S1 pocket recruited a potassium ion to balance the negative charge on the P1 side chain.

Chapter 2:
X-ray crystallographic structure of human pepsinogen A

2.1 Introduction

Human pepsinogen A is the inactive protein precursor of pepsin, an aspartic proteinase found in the stomach. A 47 amino-acid, N-terminal prosegment is removed by autolytic cleavage to form the mature enzyme after pepsinogen has been exposed to acidic pH. Aspartate and glutamate residues become protonated with a drop in pH. Protonation results in the loss of salt bridges that are present between prosegment and pepsin residues. A conformational rearrangement of the prosegment allows access to the substrate binding site and the enzyme becomes active. However, the events of prosegment rearrangement are not fully understood and a comparison of several aspartic proteinase, zymogen and intermediate structures will facilitate an understanding of this process. Structures of human (Fujinaga *et al.*, 1995) and porcine pepsin (Sielecki *et al.*, 1990), porcine pepsinogen (Sielecki *et al.*, 1991), human progastricsin (Moore *et al.*, 1995) and gastricsin intermediate II (Khan *et al.*, 1997) have been determined previously. The structure of human pepsinogen A will be presented in this chapter.

2.2 Materials and Methods

Human pepsinogen A was generously provided by Dr. Nadya Tarasova. Crystals of the protein were grown by Dr. Maia Cherney from 1.7M K_2HPO_4/NaH_2PO_4 buffer at pH 7.3. Diffraction data from these crystals were not strong enough to collect locally. However, suitable data were collected to high resolution at the Photon Factory synchrotron in Tsukuba, Japan. Data were collected to 1.7Å, but very few reflections were collected in the shell from 1.8 to 1.7Å. Therefore, this shell was not included in the refinement. Data were indexed with the program Weiss (Higashi, 1989) and the data from all of the image plates were merged with an R-factor of 7.36% (Table 2.1).

Table 2.1: Crystallographic Data

Crystal system and space group	Orthorhombic $P2_12_12$
Unit cell dimensions (Å)	a=91.6, b=105.2, c=40.2
Resolution range (Å)	20.0-1.80
Total number of reflections	143166
No. of unique reflections	33491
Average redundancy	4.3
Completeness of Data (%)	
overall	80.2
highest resolution shell (range (Å))	58.3 (1.83-1.80)
Rmerge[a]	
overall	0.074
highest resolution shell (range (Å))	0.351 (1.83-1.80)
$\langle I/\sigma(I)\rangle$	
overall	8.01
highest resolution shell (range (Å))	1.80 (1.83-1.80)
Rcryst[b] (R-free (for 10% of the reflections))	0.207 (0.258)
Total number of protein atoms[c]	2852
Total number of water molecules	276
Average B-factors ($Å^2$)	
main chain atoms	23
side chain atoms	24
solvent atoms	46
Rms deviation from ideal stereochemistry	
Bond lengths (Å)	0.019
Bond angles (°)	2.050
Deviation of planar groups (°)	0.018
Rms coordinate error (Å)	
Sigmaa estimation (Read, 1986)	0.26
Luzzati estimation (Luzzati, 1952)	0.25

[a]$Rmerge=\Sigma_{hkl}\,[(\Sigma_i\,|I_i-\langle I\rangle|)/\Sigma_i\,I_i]$

[b]$Rcryst=\Sigma_{hkl}\,||F_o|-|F_c||/\Sigma_{hkl}\,|F_o|$, Rcryst values were calculated with the data in the resolution range indicated and without a $\sigma(I)$ cutoff.

[c]The number of protein atoms includes those atoms from residues having alternate conformations.

Molecular replacement methods were used to solve the structure. Rotation and translation functions (for data from 10-3.8Å) were performed with the CCP4 (CCP4., 1994) program, AMoRe (Navaza, 1994) using porcine pepsinogen (Sielecki *et al.*, 1991) as a search model. Human pepsinogen has 81% sequence identity with porcine pepsinogen. Waters were removed but no other alterations to the search model were made. Rigid body refinement was carried out in AMoRe (Navaza, 1994) and once the position of the molecule in the unit cell was determined, amino acids of porcine pepsinogen were substituted with those of human pepsinogen A. Side chains were adjusted to fit the calculated electron density maps.

Very little continuous electron density (contoured at 1σ) was present near the junction between the prosegment and the enzyme. Human pepsinogen A has a two-residue insertion relative to the search model near this location. Therefore, residues 40P[1] to 2 were left out of the human pepsinogen A model until further refinement improved the electron density map and it was possible to fit residues in this region with more confidence. Omit maps ($2|F_o|-|F_c|$) were calculated after simulated annealing with X-PLOR (Brünger, 1992b) in an attempt to find the best model for those residues which were difficult to locate. Refinement was initially carried out with X-PLOR (Brünger, 1992b) and then completed with TNT (Tronrud, 1992). Both programs were used with Maximum Likelihood targets, as implemented by Pannu and Read, 1996. In between the rounds of refinement, the model was inspected and fitted to the electron density using the programs FRODO (Jones, 1985) and O (Jones *et al.*, 1991). Three residues, Arg15P, Arg36P and Gln227, appeared to have density suggesting alternate conformations for their side chains. Refinement cycles were performed on protein models that included only one of the conformations, for each of the alternate side chain positions before concluding that the alternate conformations were,

[1]P denotes residues belonging to the prosegment. Numbering restarts at the N-terminus of the mature enzyme.

indeed, valid. Omit maps were also calculated and inspected to help fit the side chains of these residues. Geometry root mean squared (rms) deviations and Ramachandran plots were calculated with the CCP4 (CCP4., 1994) program Procheck (Laskowski *et al.*, 1993).

Superimposition rms deviation calculations between human pepsinogen A and other structures were performed with InsightII (BiosymTechnologies, 1993). The numbers of Cα atoms that were used for the calculations are indicated in Table 2.2.

2.3 Results

The crystal structure of Human pepsinogen A (Figure 2.1) was determined in space group $P2_12_12$ with one molecule per asymmetric unit. Unit cell dimensions were a=91.6Å, b=105.2Å, c=40.2Å, $\alpha=\beta=\gamma=90°$. The data collected from two crystals were 80.2% complete to 1.8 Å. Porcine pepsinogen (Sielecki *et al.*, 1991) was a successful search model for the molecular replacement. The best solution after the initial rotation function in AMoRe(Navaza, 1994) had an rmsd of 7σ above the mean, the next highest being only 4σ above the mean. After the translation function of the top solution and rigid body refinement, the correlation coefficient was 51.5% and Rcryst was 40.1%. Figure 2.2 shows an electron density map calculated from this model. After porcine pepsinogen residues were substituted with those of human pepsinogen A, the electron density maps and molecular model were improved by cycles of refinement and model building. The final model includes 2852 non-hydrogen protein atoms and 276 water molecules. Three residues, Arg15P, Arg36P and Gln227 were found to exist in alternate conformations. An example of the electron density from this final model can be seen in Figure 2.3. The final Rcryst for data from 20-1.8Å is 20.7% with an R-free (Brünger, 1992a) of 25.8%, calculated from 10% of the data. An estimation of rms coordinate error, as calculated by program SIGMAA (Read, 1986; CCP4., 1994), is 0.26±0.01Å. Additional refinement statistics can be found in Table 2.1.

Table 2.2: Rms differences of human pepsinogen A to several aspartic proteinases and zymogens.

Protein	Rms deviation (Å)	Number of Cα atoms used in calculation
Human progastricsin	2.52	336
Porcine pepsinogen	0.91	362
Human Pepsin	1.74	314
C-term onto N-term	2.85	62
Porcine pepsin onto human pepsin with pepstatin	0.72	326

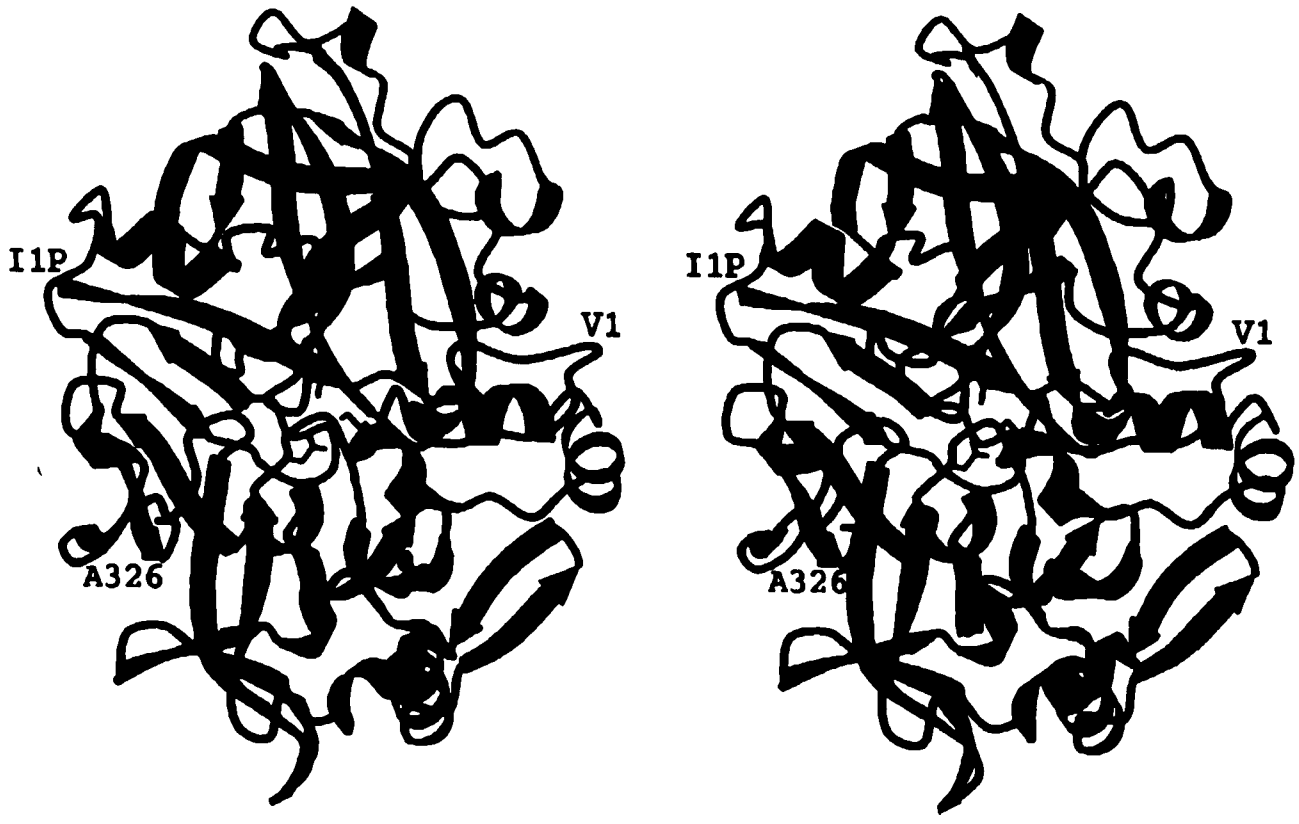

Figure 2.1 Human pepsinogen A. The prosegment has been drawn in blue, the first thirteen pepsin residues have been drawn in purple and the remainder of the molecule has been drawn in green. Side chains for catalytic aspartic acid residues Asp32 and Asp215 (red) and residue Lys37P (blue) have also been included.

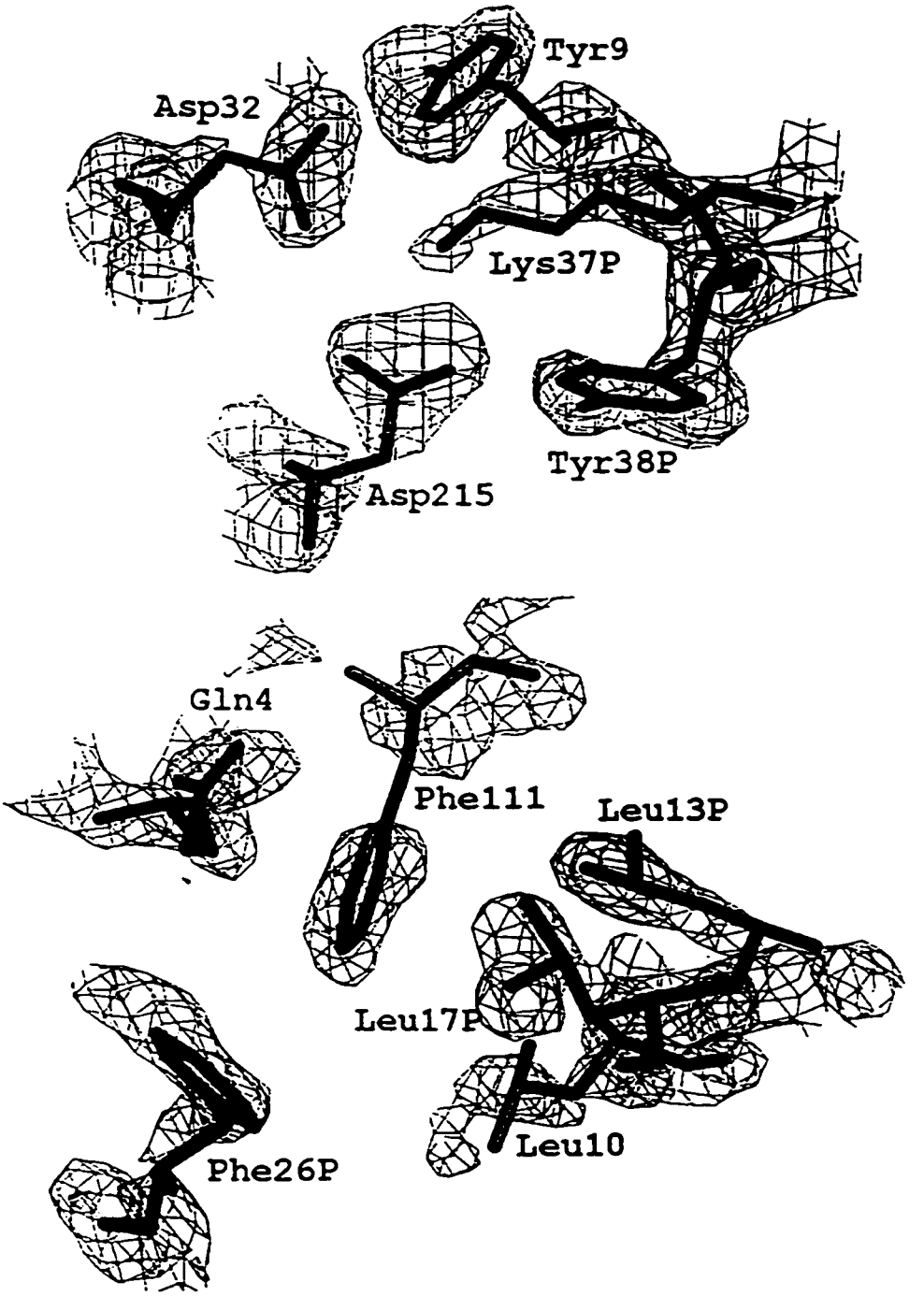

Figure 2.2 Electron density maps (2Fo-Fc) at 1.5σ drawn at the active site (top) and around residue Phe111 (bottom). The map was calculated after rotation, translation and rigid body refinement of the search model. The atomic models are from the final model.

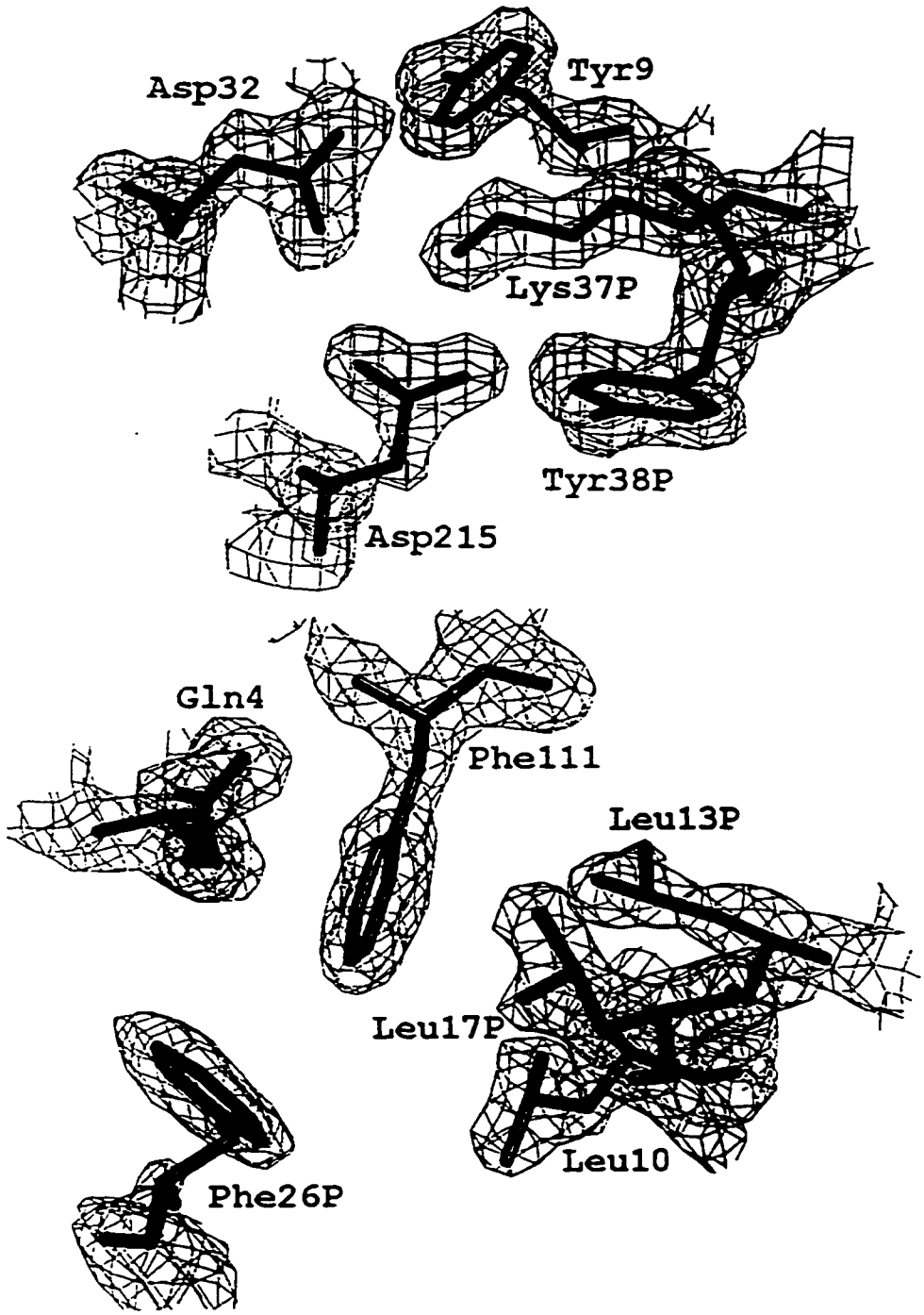

Figure 2.3 Final electron density maps (2Fo-Fc) at 1.5σ drawn at the active site (top) and around residue Phe111 (bottom) with the final models.

The geometry statistics for human pepsinogen A are also good (Table 2.1). 90% of the residues can be found in the most favoured regions of the Ramachandran plot Residues found in the generously allowed regions are Asp159, Asn200, Glu208, and Asp242. Each of these residues is located in a region of relative disorder; such regions are limited to a few surface loops; 43P-1, 157-161,199-209 and 240-243 (Figure 2.4). Electron density maps calculated for these regions were unclear and it was difficult to locate the positions of the atoms with confidence. The relative disorder of these residues is also illustrated by the isotropic B-values which are higher in comparison than those for the remainder of the molecule (Figures 2.4 and 2.5). No residues, other than glycines, are found in the disallowed regions of the Ramachandran plot.

Description of human pepsinogen A. The structure of human pepsinogen A can be divided into four domains; an N- and C-terminal lobe, a central, six-stranded β-sheet, and an N-terminal extension referred to as the prosegment (Figure 2.1). The N-terminal lobe (13-175) consists mostly of β-sheet with a few small α-helical segments. One catalytic residue, Asp32, extends from β-strand 25-32. Similarly, catalytic residue Asp215, sits on the edge of a β-strand from the C-terminal lobe in close proximity to Asp32 to form the active site. A substrate binding cleft is formed where the two lobes meet. The C-terminal lobe topology generally resembles that of the N-terminal lobe. The resemblance can be seen in Figure 2.6, from a superimposition of the two domains. Generally, the core β-sheets of the lobes superimpose well but the outer, remaining, structural elements show more divergence. An rmsd calculation between the lobes can be found in Table 2.2.

The N- and C-terminal lobes sit against the central, six-stranded β-sheet (Figure 2.1). Prosegment residues 1P to 7P provide the first strand of the β-sheet. From there, the prosegment residues adopt a helix-turn-helix conformation (Figure 2.7) followed by a short 3_{10} helix that is located near the active site. A random coil region follows the 3_{10} helix and contains the junction between prosegment and the first pepsin residues. The

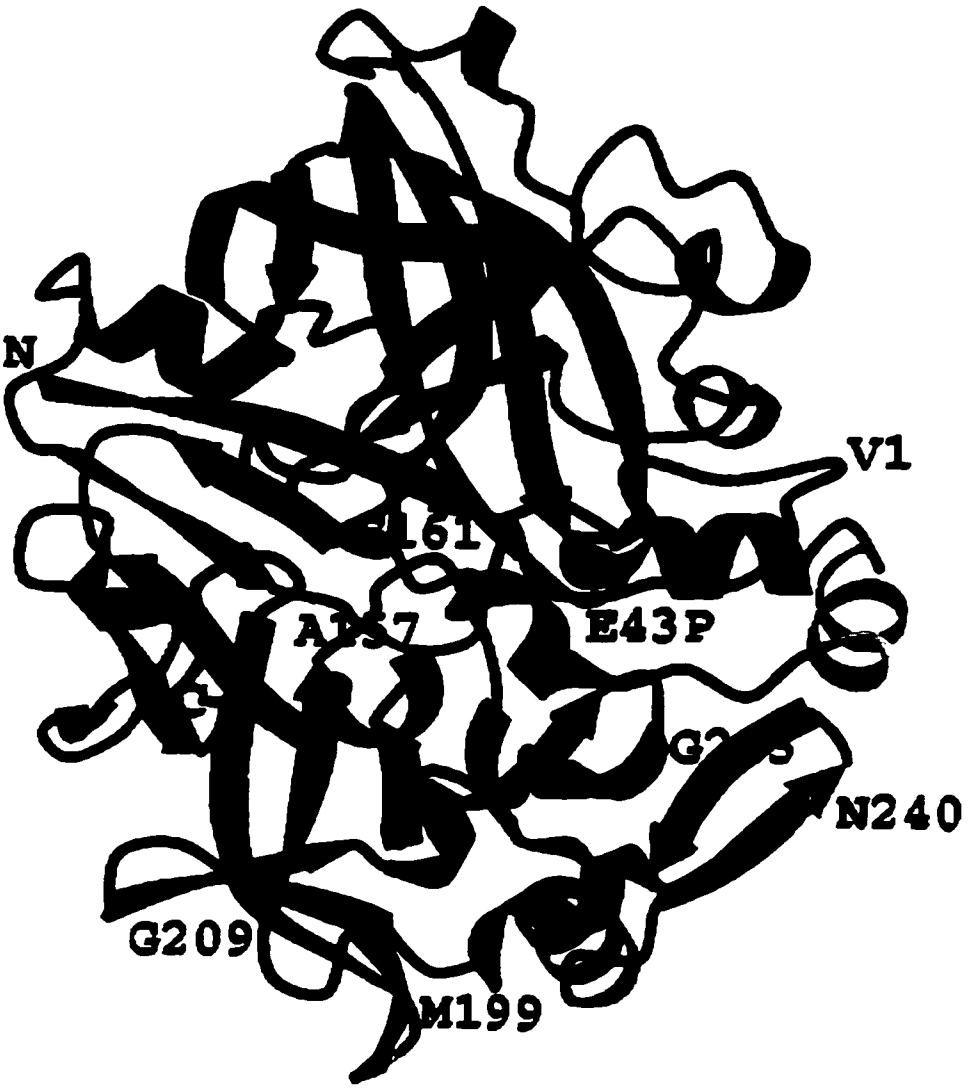

Figure 2.4 Human pepsinogen A has been coloured according to B-factor. Red represents the highest B-factors and blue represents low B-factors. This figure is in the same orientation as Figure 2.1.

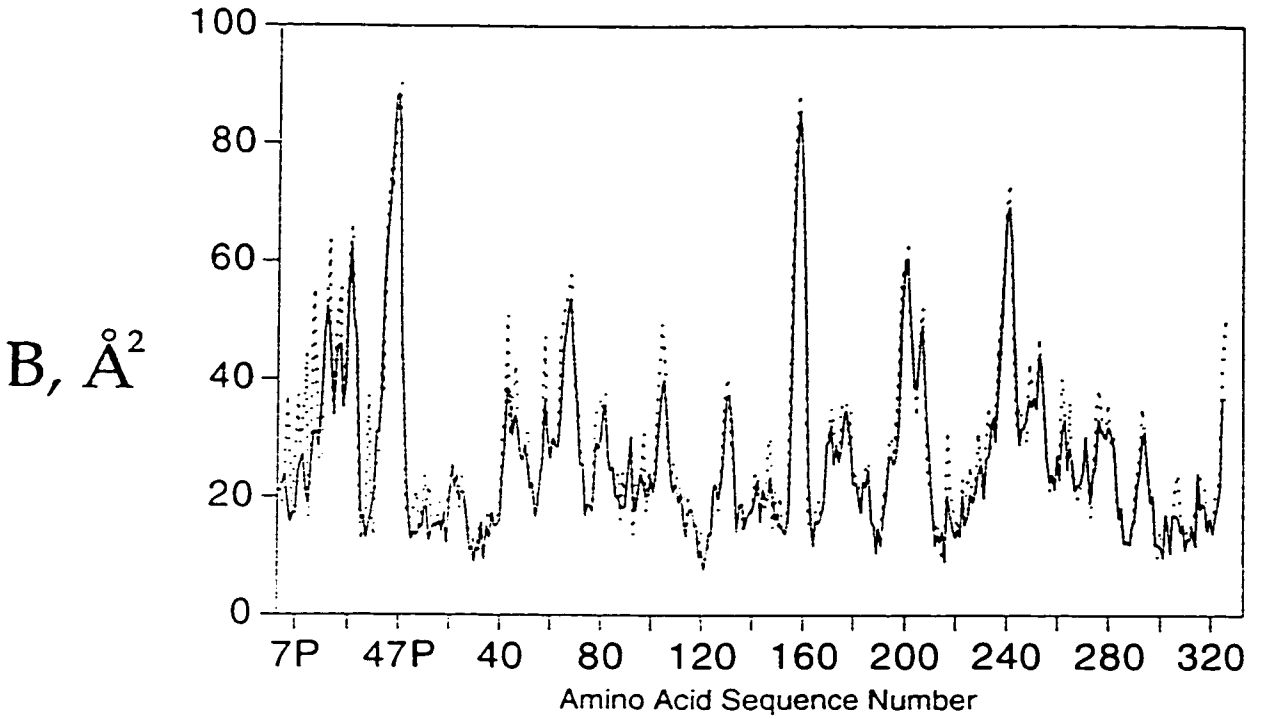

Figure 2.5 Average temperature factors for human pepsinogen A. Solid lines represent B-factors averaged over main chain atoms. Broken lines represent B-factors averaged over side chain atoms.

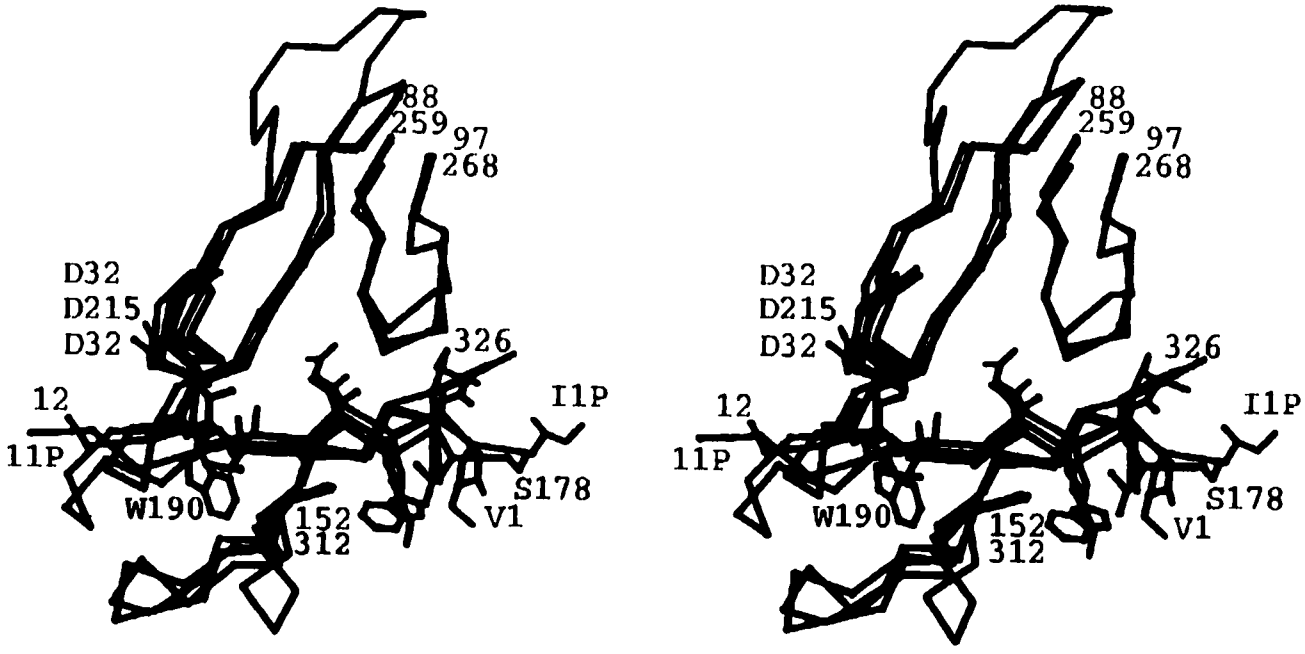

Figure 2.6 Superimposition of human pepsin (red) and the C-terminal domain of human pepsinogen A (green) onto the N-terminal domain of human pepsinogen A (blue) in the vicinity of the first β-strand in the central, 6-stranded β-sheet. The side chains of residues Asp32, Asp215, Trp190 and those belonging to the first β-strand have been drawn in ball-and-stick. The remaining residues have been depicted as a Cα trace.

random coil region is also an area with weak electron density and probably has considerable flexibility (Figures 2.4 and 2.5).

The overall conformation of human pepsinogen A resembles that of the previously determined zymogen structures, porcine pepsinogen (James and Sielecki, 1986; Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992) and human progastricsin (Moore *et al.*, 1995). A thorough description of aspartic proteinase zymogen structure can be found in the papers presenting these proteins. The coil region following the 3_{10} helix has the largest differences among the zymogens. The rms deviation calculations (Table 2.2) indicate that human pepsinogen A more closely resembles porcine pepsinogen than does shuman progastricsin. Human pepsinogen A has three additional residues in the prosegment with respect to porcine pepsinogen (Figure 2.11) resulting in a longer coil region following the 3_{10}-helix (Figure 2.7). Human pepsinogen A has four additional prosegment residues relative to human progastricsin. However, the pepsinogen prosegment coil region follows an entirely different path than that of progastricsin (Figure 2.8). These two zymogen structures diverge at residue 39P (human pepsinogen A numbering) with pepsinogen residues 41P to 4, going around 'the flap' (residues 68-86). In contrast, the progastricsin prosegment passes behind the flap. In order to accommodate the progastricsin prosegment in this arrangement, another loop from residues 125-135 (Figure 2.9) must turn away with respect to the equivalent loop in human pepsinogen A. The paths of the polypeptide chains are reunited at residue 6 of the mature enzymes.

The overall secondary structure of pepsinogen residues 13-326 is maintained from zymogen to mature pepsin (Fujinaga *et al.*, 1995). There is a small opening of the N- and C- terminal lobes of human pepsinogen with respect to human pepsin that results in the elevated rms deviation between these two molecules (Table 2.2). Pepsin residues 1-13 occupy very different positions in the zymogen (Figure 2.1) than in mature pepsin (Fujinaga *et al.*, 1995) (Figure 2.6). Residues 1-6 must travel from the active site cleft to provide the first β-strand in the six-stranded β-sheet. These residues actually occupy the positions of the

prosegment residues 2P-7P that are no longer present in mature pepsin (Figure 2.6).

The prosegment strand only shares two residues in common with the first pepsin residues and these residues occupy the same position in the β-sheet: a proline and a leucine residue at positions 6P and 7P in the prosegment and at positions 5 and 6 in the N-terminus of pepsin. The superimposition of N- and C-lobes aligns residues 179-164 with this first β-strand and again the proline is conserved but leucine has been replaced by a valine (Figure 2.6). Inspection of the structure in this area suggests that a leucine residue would bump into Trp^{190} (Figures 2.6 and 2.9). The rest of this strand, although the secondary structure is similar in all three cases, contains very different amino acids. In fact, strand 178-185 is mostly hydrophobic (S L N W V P V T); N-terminal pepsin residues 1-7 include three acidic amino acids (V D E Q P L E) and prosegment residues 1P-8P contain a positively-charged lysine residue (I M Y K V P L I).

The prosegment contains an abundance of basic residues in general. Table 2.3 catalogues the charged residues of pepsinogen. Thirteen of the positively-charged residues belong to the prosegment and will be lost when the prosegment is cleaved, leaving only four positively-charged amino acids out of 326 remaining residues in the mature pepsin. The potentially negatively-charged residues, aspartate and glutamate are distributed more evenly throughout the pepsinogen molecule.

A list of salt bridge interactions between positively and negatively-charged residues of human pepsinogen A can be found in Table 2.4. Some of these are conserved (Figure 2.11), including the very important salt bridges between prosegment residue Lys37P and the catalytic aspartic acid residues 32 and 215. Asp32 and Asp215 also accept stabilizing hydrogen bonds from tyrosine side chains 38P and 9, respectively. Salt bridges Lys4P to Asp171, Arg14P to Asp11 and His30P to Glu7, also occur between prosegment and pepsin residues. Salt bridge, Lys4P to Asp171, occurs between the first two strands of the central, six-stranded β-sheet. The next salt-bridges, Arg14P to Asp11 and His30P to Glu7, are on either

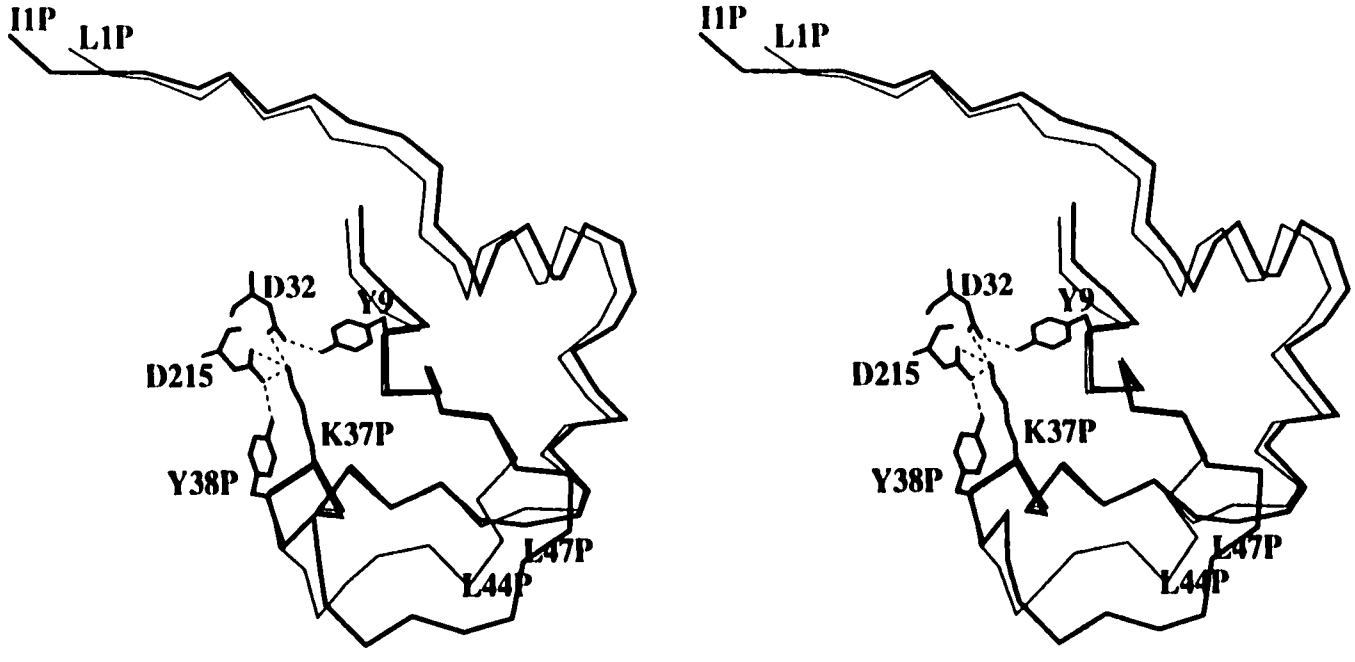

Figure 2.7 Superimposition of the porcine pepsinogen (thin lines) and human pepsinogen A (thick lines) prosegments depicted as a Cα trace. Side chains have been included for human pepsinogen residues Lys37P, Tyr38P, and also for the catalytic aspartic acid residues Asp32 and Asp215.

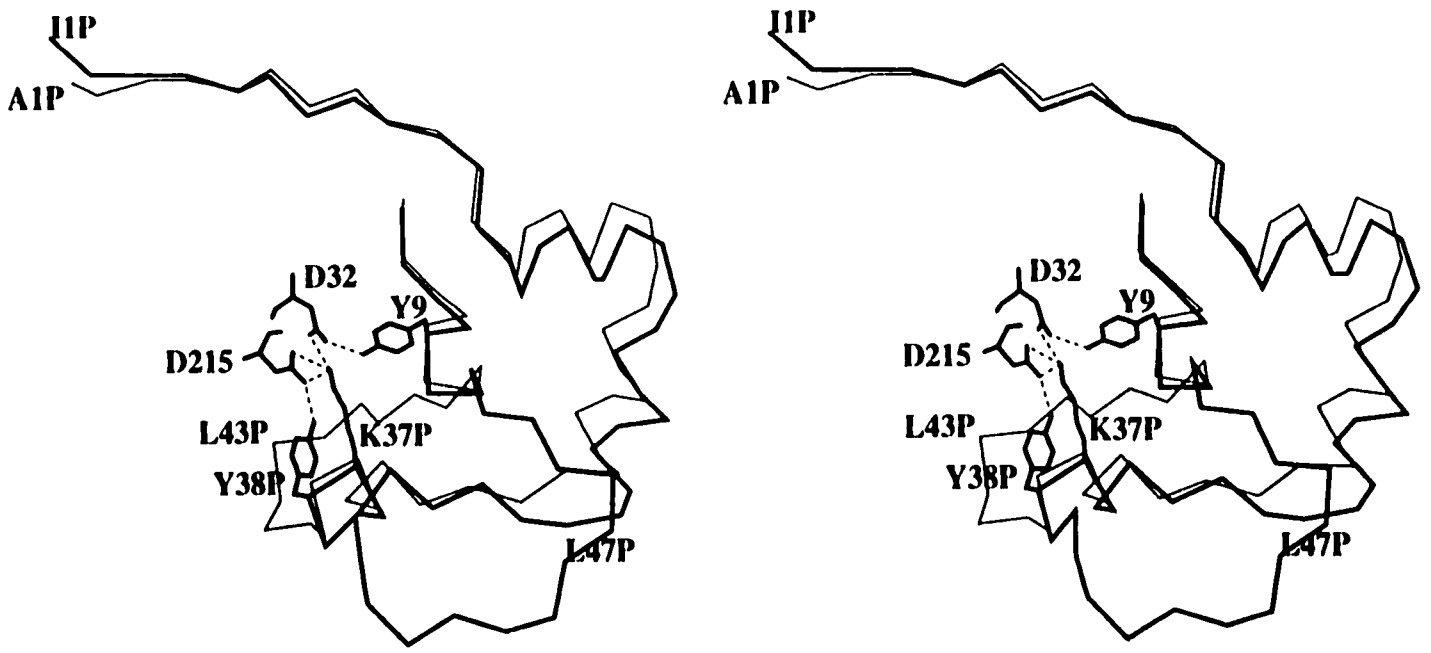

Figure 2.8 Superimposition of the human progastricsin (thin lines) and human pepsinogen A (thick lines) prosegments depicted as a Cα trace. Side chains have been included for human pepsinogen residues Lys37P, Tyr38P, and also for the catalytic aspartic acid residues Asp32 and Asp215.

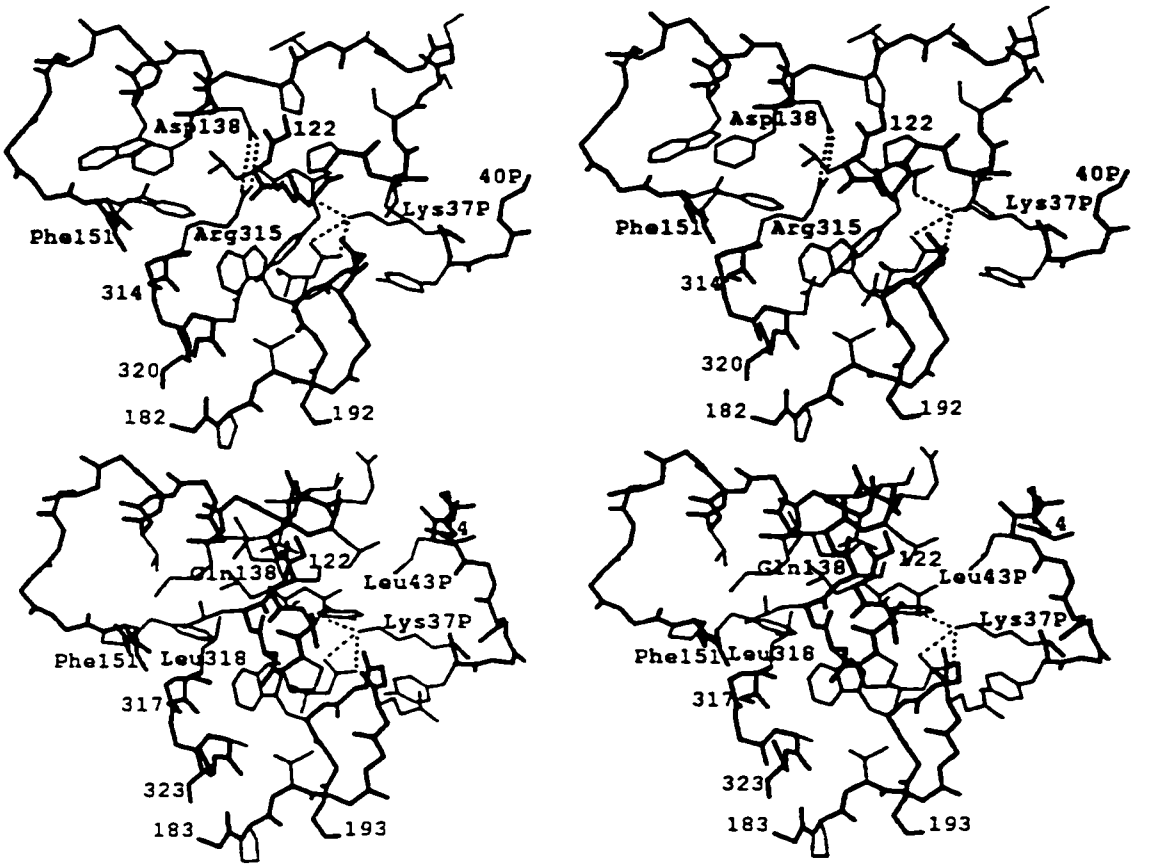

Figure 2.9 Stereo view of human pepsinogen A (top panel) in the vicinity of the salt bridge between Asp138 and Arg315. Side chains for residues near the salt bridge have been included as have those for active site residues. Only backbone atoms have been drawn for remaining residues. Salt bridged hydrogen bonds have been depicted as dotted lines. A stereo view of human progastricsin (bottom panel) has been drawn in the same region and same orientation as human pepsinogen A, although residues 125 to 135 and 40P to 4 have very different arrangements.

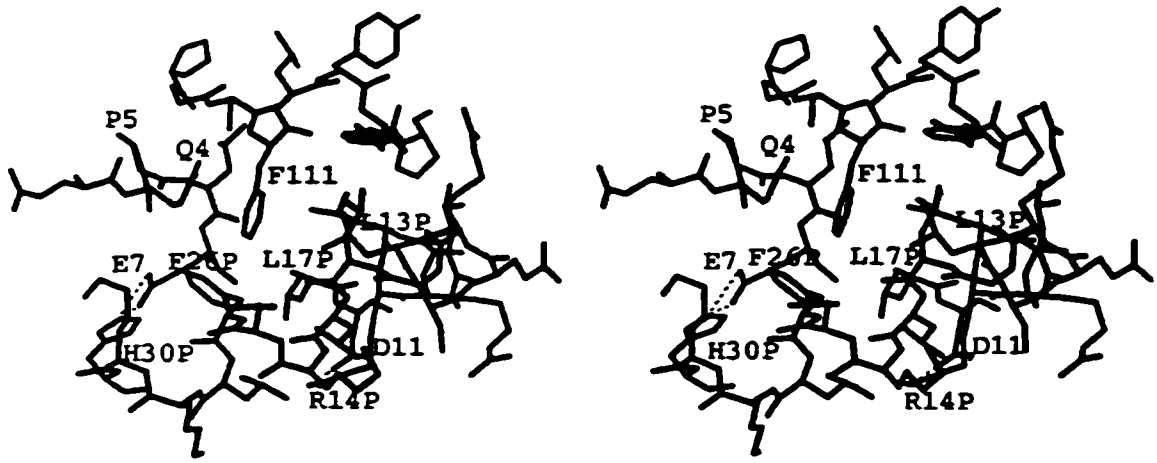

Figure 2.10 Stereo view of Phenylalanine 111 and its surrounding residues. Salt bridges between Arg14P and Asp11 and between His30 and Glu7 are drawn as dotted lines.

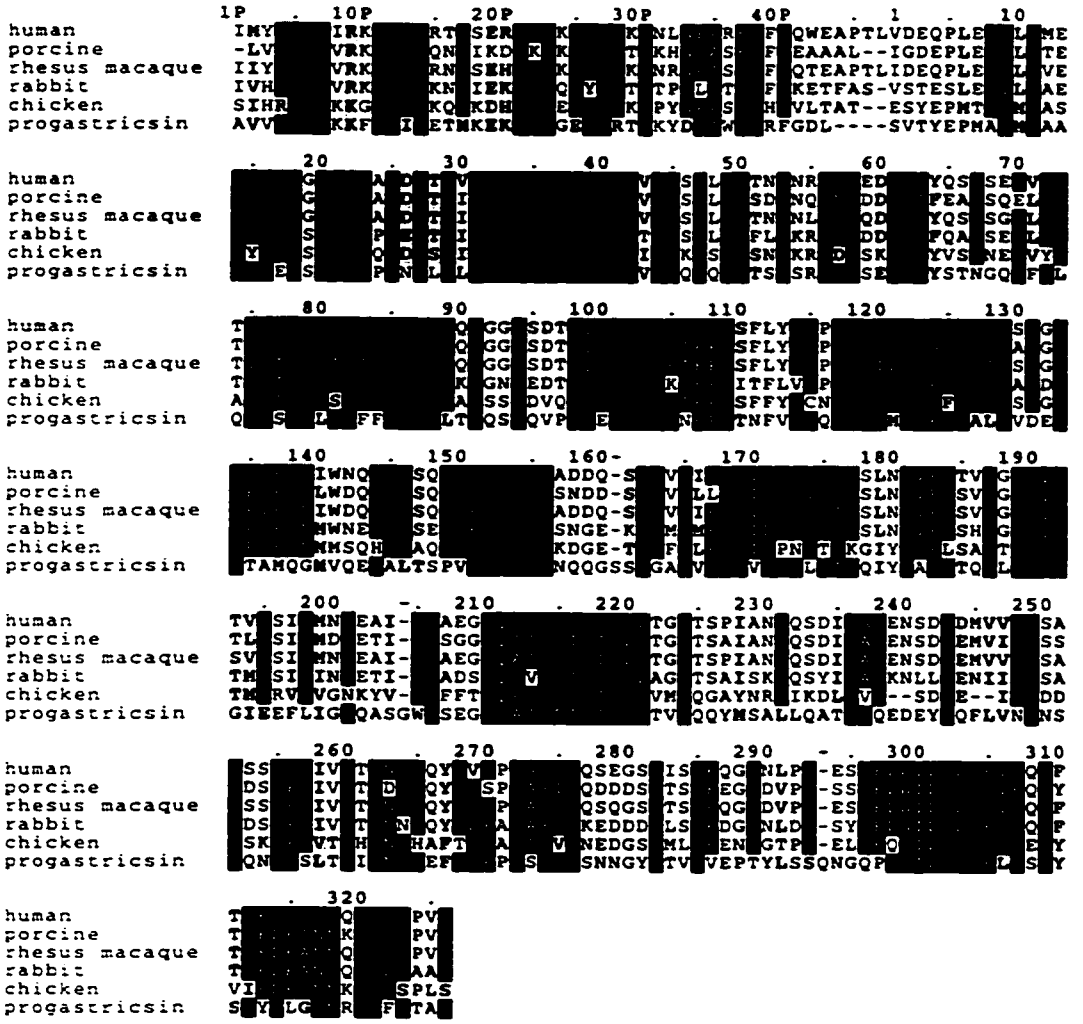

Figure 2.11 Sequence alignment of five pepsinogens: human (Sogawa, 1983; Evers, 1989), porcine (Foltmann, 1988c), macaque (Evers, 1988), rabbit (Kageyama, 1990) and chicken (Baudys, 1983)) and human progastricsin (Hayano, 1988; Taggart, 1989). Identical residues are boxed in black and conservative substitutions are boxed in grey. Numbering is that of human pepsinogen A. Dashes indicate deletions with respect to other sequences.

side of the helix-turn-helix (Ser12P to His30P) segment and link this part of the prosegment to the pepsin portion of the molecule (Figure 2.10). Salt bridges His53-Asp118 and Asp138-Arg315 (Figure 2.9) are contained within the pepsin portion of the zymogen molecule. Both appear to have been conserved among the pepsinogens from various sources (Figure 2.11), but Asp138 and Arg315 are absent from the progastricsins.

Hydrophobic amino acid Phe111 is notable because it has a χ_2 angle of 0° (Figures 2.1 and 2.10). Zero degrees is a very unfavourable χ_2 angle for the aromatic amino acids (Schrauber *et al.*, 1993; Dunbrack *et al.*, 1994). Phe111 is held in such a conformation by packing interactions with surrounding residues Leu17P, Phe26P, Gln4, two of which belong to the helix-turn-helix portion of the prosegment. In the structure of mature, human pepsin, with the prosegment removed, Phe111 is able to adopt a more favourable χ_2 angle of 70° (Fujinaga *et al.*, 1995) . Phe111 also distinguishes the pepsinogens from the progastricsins. Although many features of the helix-turn-helix environment appear to be conserved throughout the aspartic proteinase zymogens, Phe111 is only conserved among the pepsinogens (Figure 2.11).

Amino acid (Kageyama and Takahashi, 1980; Athauda *et al.*, 1989; Roberts *et al.*, 1995) and nucleotide sequencing (Sogawa *et al.*, 1983; Evers *et al.*, 1987; Evers *et al.*, 1989) of human pepsinogen A isoforms have identified five sequence locations with amino-acid substitutions (Table 2.5). Three of these residues, Glu43P, Gln160, and Ala203 are located on the surface of the zymogen where side chains have little interaction with the rest of the molecule. Furthermore, residues Glu43P and Gln160 are found in regions that have insufficient electron density to unequivocally determine their identity from the electron density map. Glu43P and Gln160 are also the only two locations where amino-acid substitution has an effect on charge. As a lysine, residue 43P can potentially form a salt bridge with Glu294, an acidic residue nearby (Figure 2.12).

Residue Leu291 is found on a loop, close to the prosegment (Figure 2.12). The structure of mature human (Fujinaga *et al.*, 1995) pepsin indicates

Table 2.3: Distribution of charged residues in human pepsinogen A

Residue	Human pepsin (326 a.a.)	Human pepsinogen A prosegment (47 a.a.)	Total
Asp	21	1	22
Glu	14	2	16
Arg	3	5	8
Lys	0	7	7
His	1	1	2

Table 2.4: Salt bridges of human pepsinogen A

Donor	Acceptor
Lys4p	Asp171
Lys11p	Glu19p
Arg14p	Asp11
His30p	Glu7
Lys37p	Asp32
Lys37p	Asp215
His53	Asp118
Arg315	Asp138

that this loop moves towards the active site once the prosegment has been removed and the environment of Leu291 changes slightly from pepsinogen to pepsin. The potential substitution at position 291 is a conservative one, from leucine to valine. Porcine pepsinogen has a valine at position 291. Most of the contacts that residue 291 makes with surrounding residues are the same for both porcine (Sielecki *et al.*, 1991) and human pepsinogen. A superimposition of porcine pepsin (Sielecki *et al.*, 1990) onto human pepsin with pepstatin bound to the active site (Fujinaga *et al.*, 1995) (not shown) reveals that, in the pepsin environment, Leu291 contacts residue P_3' of the inhibitor. It appears from the superimposition that a valine at this position would not.

Val30 is also located in the substrate binding site. In the structure of mature human pepsin in complex with pepstatin (Fujinaga *et al.*, 1995), Val30 makes van der Waal contacts with the P_1 residue of the inhibitor (Figure 2.12). This position also has a conservative substitution between valine and leucine for the five isoforms (Table 2.5). A superimposition of human progastricsin, which has a leucine at position 30 onto human pepsin with pepstatin (not shown) places the leucine such that it makes close contacts with the P_1 residue.

2.4 Discussion

The structures of aspartic proteinase zymogens (James and Sielecki, 1986; Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992; Moore *et al.*, 1995) have presented a simple explanation for their inactivity. A prosegment fills and blocks access to a preformed active site cleft. Catalytic aspartic acid residues are stabilized in the same configurations that they adopt in the mature enzyme, by interactions with the prosegment. These interactions do not mimic that of a substrate and, therefore, the peptide bonds of the prsegment are not subject to catalytic hydrolysis. For example, tyrosine residues 9 and 38P, sit on opposite sides of the catalytic aspartic acid residues and donate hydrogen bonds to the side chains but there are no peptide bonds in the right conformation to be cleaved. In fact, the tyrosines are 18 residues apart. The conserved prosegment residue

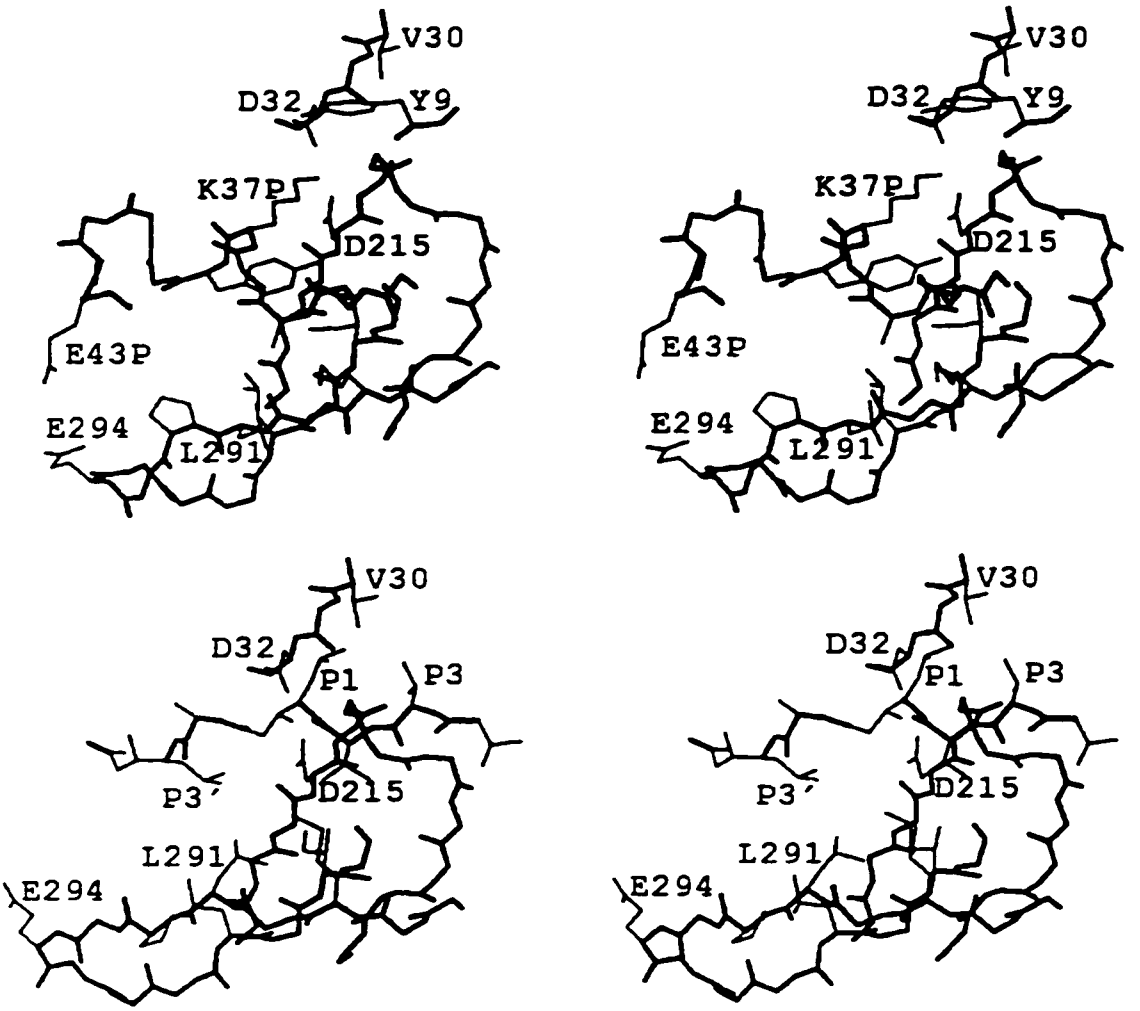

Figure 2.12 Stereo view of the active site cleft of human pepsinogen A (top panel) and human pepsin bound with pepstatin (bottom panel). Both structures are shown in the same orientation. Side chains have been shown for selected residues; L291 and its neighbours, 35P, 289,292,298,300; V30; D32; D215; K37P; Y38P; Y9; 300; 43P; 294; and pepstatin residues, P_3' Sta, P_2'Ala, P_1Sta, P_2Val, P_3Val, P_4Iva)

Table 2.5: Isoform substitutions of human pepsinogen A

Isoform	Conversion speed	43p	30	160	203	291
1	no data	---	---	---	---	---
2	slow	E	V	---	---	---
3	fast	E	V	**Q**	*A*	*L*
4	slow	K/**E**	L	**K**	**T**	**V**
5	slow	K	L	**Q**	**A**	**L**

Amino acids in bold font were determined by nucleotide sequencing, those in italics were determined by X-ray crystallography and the rest were determined by amino-acid sequencing.

Lys37P further stabilizes Asp32 and Asp215 through the formation of hydrogen-bonded salt bridges with these aspartic acids.

Elucidation of the pathway from zymogen to mature enzyme has not been so simple, and many of the details remain speculative. Conformational rearrangement of the prosegment is an intriguing part of the pathway. Even before rearrangement has begun, the prosegment appears to have more flexibility than the rest of the molecule. The junction between prosegment and the first pepsin residues is largely exposed to solvent and residues 43P to 1 have high B-factors and very weak electron density (Figure 2.4). The 47P-1 peptide bond is well exposed to hydrolysis by enzymes that are already active. Other sections of the prosegment are clearly traced in electron density maps and have more contact with the pepsin part of the zymogen (Figure 2.10) but the B-factors are still higher than those for the rest of the molecule (Figure 2.4). B-factors that are higher than average are the result of several possibilities. One is that each prosegment in the crystal has a slightly different conformation. Another is that the region in question has movement and of course, elevated B-factors can reflect both possibilities. All of these possibilities imply that the prosegment has flexibility and will readily undergo a conformational change.

What triggers a conformational change of the prosegment? The positively-charged, 47 amino-acid prosegment is tenuously held against the pepsin portion of the molecule by a few hydrophobic interactions (Figure 2.10), one β-strand (Figure 2.1), and by five salt bridges (Table 2.4), all of which are sufficient in the neutral environment of the chief cells. However, in the acidic environment of the stomach, exposed glutamate and aspartate residues will lose their negative charge, due to protonation, leaving the positively-charged residues without counter-ion stabilization. The prosegment structure, which already appears flexible, will be further disrupted from packing against the pepsin part of the zymogen. Although the aspartic proteinase zymogens share similar distributions of charged amino acids and therefore, all zymogens have been proposed to react to acidic conditions in a similar manner; the conversion rates vary

among homologues (Foltmann and Jensen, 1982) and among isoforms (Foltmann, 1988b; Athauda *et al.*, 1989; Kageyama and Takahashi, 1989) (Table 2.5). The conversion rate might be affected by the number of salt bridges between prosegment and enzyme. The difference in the number of salt bridges between isoforms 3 and 5 might explain the different conversion rates between the two isofoms (Table 2.5). Residue Glu43P, in isoform 3, is substituted by a lysine (Figure 2.12). As a lysine, 43P is potentially close enough to Glu294 to form an additional salt bridge and perhaps provide more stability to the isoform 5 zymogen. However, not all salt bridges contribute towards stability and the context of a salt bridge is very important (Kohn *et al.*, 1998).

Turning towards the different conversion rates between progastricsin, which is very fast (Foltmann and Jensen, 1982), and the pepsinogens (Foltmann, 1988b; Athauda *et al.*, 1989; Kageyama and Takahashi, 1989), it is tempting to implicate the different arrangement of the prosegments (Figure 2.8). A superimposition of human progastricsin onto human pepsinogen A indicates that the peptide bonds between the last prosegment residue and the first pepsin or gastricsin residue start at very different locations but must eventually arrive at the active sites that are well superimposed. The prosegment differences also appear to influence the arrangement of surrounding loops (Figure 2.9). In particular, the progastricsin loop from 125-135 turns toward the location of a salt bridge, present in the pepsinogens, between residues Asp138 and Arg315. This salt bridge may prevent the pepsinogen loop 125-135 from adopting the progastricsin conformation. However, any suggestions as to how these structural differences at neutral pH might direct conversion at acidic pH would require much more experimental evidence.

An additional difference between pepsinogens and progastricsins is the presence of Phe111 (Figure 2.11). In the pepsinogen structures, Phe111 is located in a hydrophobic pocket formed by prosegment and pepsin residues (Figure 2.3 and 2.10). The hydrophobic face of the helix-turn-helix segment packs against the phenylalanine such that it is required to adopt an unfavourable χ_2 angle of 0° (Schrauber *et al.*, 1993; Dunbrack *et*

al., 1994). Once the salt bridges on either side of the helix-turn-helix have been disrupted by a decrease in pH and cleavage has taken place, it is plausible that the phenylalanine has enough freedom to find a more favourable conformation, thereby releasing the prosegment from packing against the enzyme.

The progastricsins do not have such a phenylalanine in the hydrophobic pocket formed by the helix-turn-helix. Perhaps this is why researchers have been able to isolate a progastricsin intermediate that has been cleaved, but still retains the prosegment through non-covalent interactions (Foltmann *et al.*, 1982; Khan *et al.*, 1997). This intermediate was isolated by rapidly raising the pH after it had initially been lowered to stimulate conversion. There are two explanations for these intermediates. Interactions that were initially disrupted by lowering the pH, were allowed to reassociate after raising the pH, or the initial disruptions were limited to the active site region. If the former is true, it is possible to suggest a reason why isolation of a pepsinogen intermediate by a similar manner has not been successful. The hydrophobic association between prosegment and enzyme may be difficult to re-establish once Phe111 has been allowed to adopt a more favourable χ_2 conformation.

The first thirteen pepsin residues contribute towards pepsinogen inactivity. They undergo a large conformational change with enzyme conversion but are not considered prosegment residues because they remain with the enzyme after the prosegment has been removed (Figure 2.1). In fact, the first six of these thirteen residues replace the β-strand originally formed by prosegment residues 2P-7P (Figure 2.6). This β-strand exchange maintains the central 6-stranded β-sheet while subtly shifting the nature of the first strand from being positively charged, a characteristic which would be disruptive in an acidic environment, to polar. The aspartate and glutamate residues at positions 2 and 3 would presumably be protonated and without a negative charge at the low pH of the stomach. The conserved proline and leucine residues in both the prosegment and enzyme strands probably act something like a 'hook' to anchor the strands. The proline residue would reduce flexibility and the

leucine residue packs perfectly into the hydrophobic pocket as the 'hook'. Interestingly, the sixth strand of the central β-sheet, Leu179 to Val184, which superimposes onto the first strand in a superimposition of N- and C-domains, also contains a proline at the identical position (Figure 2.6). Proline at position 183 is followed by a valine, which like the leucine in the first β-strand, fits perfectly into its own hydrophobic pocket. Both of these residues are well conserved but this strand is not subject to exchange and does not need a 'hook' to target the strand.

Variation in sequence among the isoforms of human pepsinogen A extends to variation in the enzyme conversion rate and possibly to substrate specificity. The isoforms are known to differ in their rates of conversion (Foltmann, 1988b; Athauda *et al.*, 1989; Kageyama and Takahashi, 1989). Perhaps the different rates can be related to additional salt bridges, as discussed previously. Other factors that might have some influence on the conversion rate include residue Leu291, which changes its environment from zymogen to enzyme (Figure 2.12). Differences between a valine and leucine at this position may be able to affect the ease with which the conformational change takes place. Valine, for example, is β-branched which places more constraints on ϕ and φ angles. However, a more specialized role for Leu291 is participation in substrate specificity, as proposed by Fujinaga *et al.*, 1995. In the structure of mature pepsin with pepstatin, Leu291 makes van der Waals contacts with the Sta residue at P_3'. A valine at position 291 would presumably allow a larger side chain at the P_3' position but this result should be noted with caution as both residues Val291 and StaP_3' are disordered in the pepsin structure. Residue 30 would also be expected to affect substrate specificity. Val30 makes a contact with the P_1 Sta residue of pepstatin (Figure 2.12). A leucine at position 30 would likely be more restrictive.

A purpose for substitutions at the other positions which differ among the isoforms is unclear. The change at position 160 has an affect on charge and a lysine at this position may be able to form a salt bridge with the C-terminal carboxy group. However, this residue is very poorly defined in the crystal structure and it is difficult to speculate on whether or not

residue 160 affects function. Residue 203 also sits on the surface of the molecule, far from the active site and does not appear to have any influence on either conversion rate or specificity but residues far from the active site of other enzymes have been shown to affect activity (Imoto *et al.*, 1994; Heyduk *et al.*, 1991; Perona *et al.*, 1995) and residue 203 may be able to influence the conversion rate or catalytic activity of human pepsinogen A.

Chapter 3:
Deleterious effects of β-branched residues in the S1 specificity pocket of SGPB: crystal structures of OMTKY3 variants Ile^{18I}, Val^{18I}, Thr^{18I} and Ser^{18I} in complex with SGPB

3.1 Introduction

The turkey ovomucoid third domain (OMTKY3) belongs to the Kazal family of proteinase inhibitors. These inhibitors bind tightly to their cognate enzymes because the reactive site loop of the inhibitor is geometrically complementary to the active site of the enzymes that they inhibit. The inhibitors are then hydrolysed very slowly. A primary specificity determinant of the inhibitors is the P1 residue of the reactive site loop that fits into the S1 pocket of the proteinase. Equilibrium binding constants have been measured for six serine proteinases in complex with P1 mutants of OMTKY3 (Table 1.2) (Bigler *et al.*, 1993; Lu *et al.*, 1997). The K_a's for the β-branched P1 variants are relatively low for most of the serine proteinases studied with the exception of the elastases. X-ray structures have been determined for OMTKY3-Ile^{18I}, OMTKY3-Val^{18I} and OMTKY3-Thr^{18I} in complex with SGPB in order to determine how the β-branched side chains are disruptive at the P1 position. In addition, the structure of SGPB:OMTKY3-Ser^{18I} has been determined and a model of SGPB:OMTKY3-Cys^{18I} will be presented, based on the the structures of SGPB:OMTKY3-Ser^{18I} and SGPB:OMTKY3-Leu^{18I} (Huang *et al.*, 1995).

3.2 Materials and Methods

Dr. L.B. Smillie kindly provided SGPB that was purified from Pronase according to the procedure of Jurasek *et al.*, 1979. OMTKY3 variants were generously provided by Dr. M. Laskowski, Jr.. The cloning, overexpression and purification of the OMTKY3 variants have been described (Lu, 1994; Lu *et al.*, 1997).

The OMTKY3 P1 variants were dissolved with SGPB in distilled water in a 1.5 to 1 molar ratio, respectively. Crystals of the four complexes were grown by the hanging drop method with protein concentrations of 9mg/ml in each drop. The precipitant was 4-10% PEG 4000 for the initial crystals. Streak seeding (Stura and Wilson, 1992) was used to improve crystal size and typical precipitant concentrations for these experiments were 1-4% PEG 4000. The pH was buffered by 50mM $NaKHPO_4$ in the range of 6.8-7.4.

All four data sets were collected at room temperature with a DIP 2030H image plate detector (Mac Science Co., Ltd.) mounted on a Rigaku rotating anode X-ray generator operating at 45kV and 75mA. Double mirror focusing optics were used to align the beam. Diffraction data were processed and reduced with the programs Denzo and Scalepack (Otwinowski and Minor, 1996).

Initial phases were generated with CCP4 (CCP4., 1994) using the native SGPB:OMTKY3 complex structure (Read *et al.*, 1983; Fujinaga *et al.*, 1982). Leu^{18I} was retained as the P1 residue for calculation of the first electron density map. Waters were not included and Val^{6I} was also absent from the model (electron density was sparse for the first six amino acids of the native OMTKY3 inhibitor.) To ensure that the correct rotamer was chosen for Val and Thr P1 residues, models were also refined with χ_1 rotated by 180° and the resulting maps were inspected. Graphics programs FRODO (Jones, 1985), O (Jones *et al.*, 1991) and Xtalview (McRee, 1993) were all used to examine models and maps. After the correct P1 residue was fit into the electron density and Val^{6I} was added to the N-terminus of OMTKY3, the model was subjected to simulated annealing in X-PLOR (Brünger, 1992b). Water positions from the native SGPB:OMTKY3 model were inspected, along with density calculated from the X-PLOR refined model. If electron density was present at a level of 1σ in the $2|F_o|-|F_c|$ maps and 2.5σ in the $|F_o|-|F_c|$ map, the water was included in the model. Additional waters were added based on electron density and hydrogen bonding potential. Side chains were adjusted to fit the maps and all further refinement was carried out with TNT (Tronrud,

1992). Several cycles of model adjustment, water picking and refinement ensued.

Simulated annealing was carried out on three models of the SGPB:OMTKY3-Ser^{18I} complex; the original 'major' conformation, a second alternate conformation and Ser^{18I} substituted for alanine. An alternate conformation was subsequently included in all further refinement. Occupancies were not refined against the data themselves, but three sets of occupancies (0.6, 0.4; 0.5, 0.5; 0.7, 0.3) were each subjected to 10 cycles of TNT refinement in order to determine the best occupancy value.

Superimposition rms deviation calculations for comparisons of final coordinate sets were performed in Xtalview (McRee, 1993). Models were assessed with programs Procheck (Laskowski *et al.*, 1993) and Whatcheck (Rodriguez *et al.*, 1998) at several stages during refinement and after completion. Cavity volumes in the S1 pocket were calculated both with voidoo (Kleywegt and Jones, 1994) and with Grasp (Nicholls *et al.*, 1991). Waters were removed from the pocket but since the pocket is solvent accessible, waters 15 and 67 were retained as dummy atoms to define a closed cavity. The radius of the probe was 1.4 Å in both voidoo and Grasp.

3.3 Results

Diffraction quality crystals of the four SGPB:OMTKY3-X18 complexes grew overnight after seeding. The space group was $P2_1$ and the unit cell dimensions can be found in Table 3.1. The asymmetric unit contained one SGPB:OMTKY3 complex. It was more difficult to obtain good crystals of Ser^{18I} and Ile^{18I} than the Val^{18I} and Thr^{18I} P1 variants. The crystal quality is reflected in some of the statistics (Table 3.1), particularly the R-merge. However, the data were of more than adequate quality and resolution to produce clearly interpretable electron density maps.

Initial molecular replacement maps for all four structures are shown in Figures 3.1, 3.2, 3.3, and 3.4. Even though the map phases were calculated with a leucine at the P1 position, the identity and orientation of the true P1 residue was obvious. Val^{18I} and Thr^{18I} residues were flipped 180° about χ_1, relative to the initial fit and then refined. The resulting $|F_o|-|F_c|$ and $2|F_o|-|F_c|$ maps indicated that the original conformation was correct and distinct.

During refinement, a positive peak (at 2.5σ $|F_o|-|F_c|$) appeared near Ser^{18I} Cβ (Figure 3.5). The $2|F_o|-|F_c|$ density (at 1σ) also had a blip in the same region and the possibility of an alternate conformation of the side chain was explored. Simulated annealing runs were performed with an alanine at the P1 position. The $|F_o|-|F_c|$ omit map (at 2.5σ) showed a large positive peak at the original Oγ position and only a very small peak at the potential second position. A model of the complex with Ser^{18I} Oγ in the second position was also subjected to simulated annealing. The Oγ moved back to its major position but a positive difference peak remained at the second Oγ site. An alternate Ser^{18I} conformation was fit into the positive electron density peak and refined with TNT (Tronrud, 1992). The resulting map indicated that an alternate conformation was present and both Oγ atoms (χ_1A=-46°, χ_1B=40°) were included in subsequent models.

Final models and electron density maps for the four complex structures can be seen in Figures 3.6, 3.7, 3.8, and 3.9. Error estimates, R-factors and other statistics for the completed structures can be found in Table 3.2. 90% of the residues from SGPB:OMTKY3-Val^{18I}, 86% of the residues from SGPB:OMTKY3-Ile^{18I}, 88% of the residues from SGPB:OMTKY3-Thr^{18I} and 90% of the residues from SGPB:OMTKY3-Ser^{18I} are found in the most favoured regions of the Ramachandran plot. Only one residue, Asn100 from SGPB, is found in the disallowed region for all of the SGPB:OMTKY3 complex structures. Asn100 is located in a sharp turn following a cis-proline, Pro99A, and the electron density for this residue has been consistantly well defined. These statistics indicate that all four structures are of good quality.

Table 3.1 Data Collection Statistics

	Ile^{18Ia}	Val^{18I}	Thr^{18I}	Ser^{18I}
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit Cell a(Å)	45.49	45.47	45.66	45.52
b(Å)	54.74	54.61	54.53	54.59
c(Å)	45.59	45.57	45.59	45.58
β(°)	119.21	119.16	119.24	119.16
Maximum resolution (Å)	1.90	1.60	1.65	1.80
Total no. of observations	34403	93994	68327	57863
No. of unique reflections	12477	24537	21917	18888
Average redundancy	2.8	3.8	3.1	3.1
Rmerge[b] :				
overall	0.144	0.076	0.068	0.122
highest resolution shell	0.456	0.235	0.326	0.455
Resolution range of last shell (Å)	1.94-1.90	1.63-1.60	1.69-1.66	1.83-1.79
$\langle I/\sigma(I)\rangle$:				
overall	6.05	12.80	13.24	7.71
highest resolution shell	1.37	2.97	2.17	1.10
Resolution range of last shell (Å)	1.91-1.90	1.61-1.60	1.66-1.65	1.81-1.79
Completeness of Data:				
overall (%)	80.5	94.8	91.5	96.7
highest resolution shell (%)	42.5	75.0	59.9	72.0
Resolution range of last shell (Å)	1.93-1.90	1.63-1.60	1.68-1.65	1.83-1.80

[a]The column headings refer to the SGPB:OMTKY3 complex with that amino acid at the P1 position.

[b]$Rmerge=\Sigma_{hkl}\,[(\Sigma_i\,|I_i-\langle I\rangle|)/\Sigma_i\,I_i]$

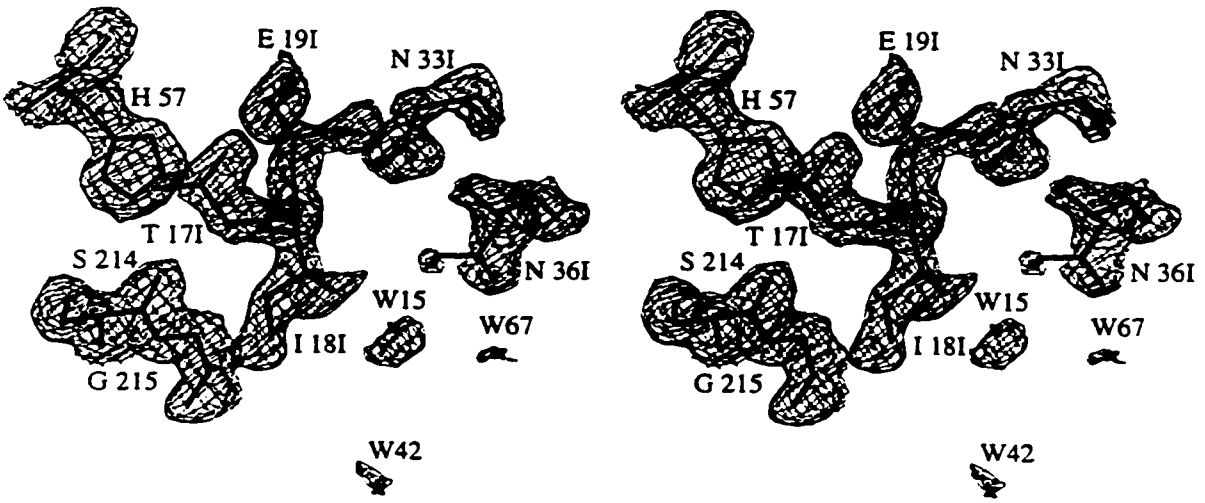

Figure 3.1 Initial electron density map of SGPB:OMTKY3-Ile^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

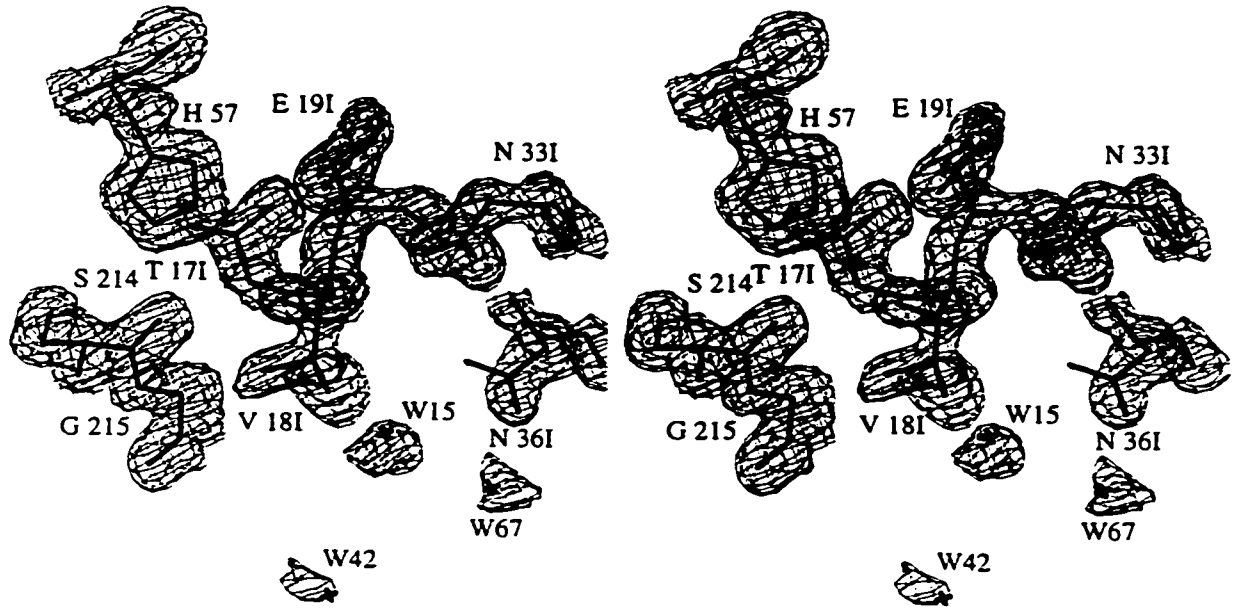

Figure 3.2 Initial electron density map of SGPB:OMTKY3-Val[18I] in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

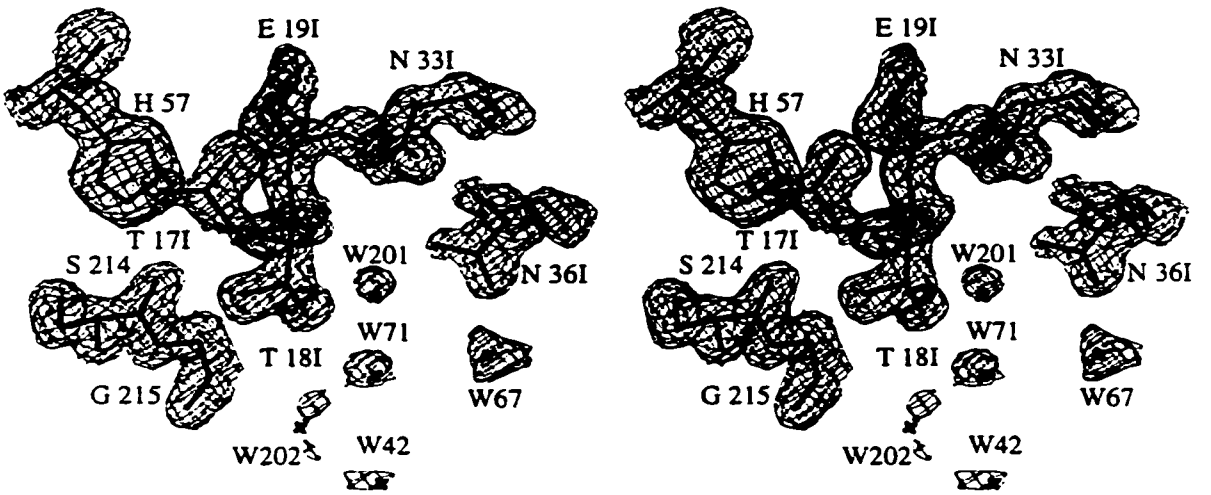

Figure 3.3 Initial electron density map of SGPB:OMTKY3-Thr^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

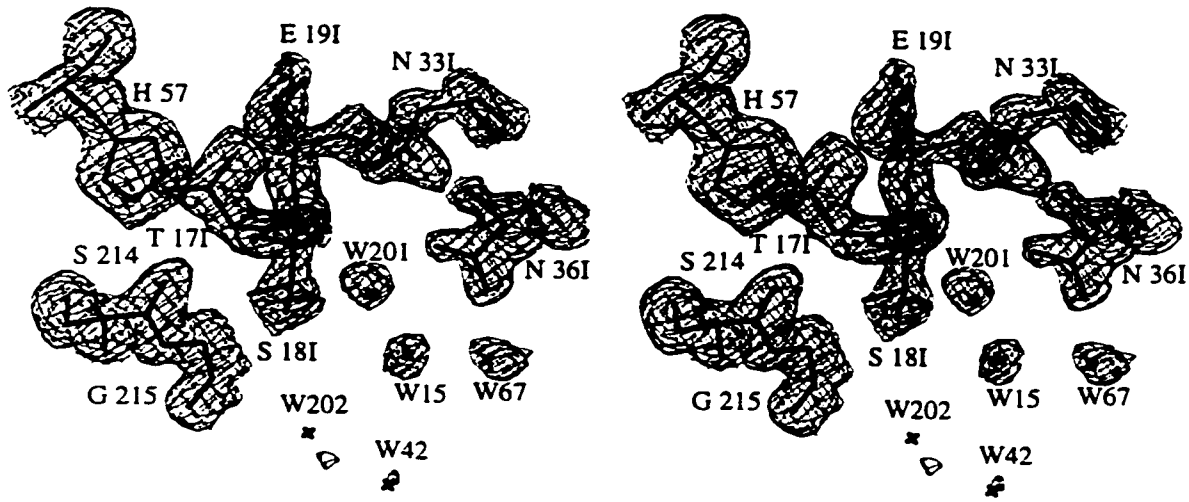

Figure 3.4 Initial electron density map of SGPB:OMTKY3-Ser^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

Figures 3.10 and 3.11 illustrate why χ_1 angles of approximately 40° (SGPB:OMTKY3-Ile^{18I}, 33°; SGPB:OMTKY3-Val^{18I}, 47°; SGPB:OMTKY3-Thr^{18I}, 39°) are preferred by the β-branched P1 residues over χ_1 angles 180° away, which would direct the γ atoms towards Pro^{192B}, Glu^{193} and Asp^{194} of the S1 pocket. These SGPB residues form the 'right' wall at the top of the specificity pocket and there is not enough room to fit the β-branched γ atoms without either adjusting the SGPB residues or adjusting the backbone of the inhibitor.

All of the structures superimpose well (Table 3.3), whether the superimposition was calculated over SGPB, OMTKY3 or both molecules. Additional superimpositions were calculated with the structure of SGPB:OMTKY3-Leu^{18I}; the results of these are also given in Table 3.3. Figures 3.12 to 3.16 show the S1 pocket results for several sets of superimpositions calculated using the 740 backbone atoms from the enzyme. All of the superpositions demonstrate the strong agreement among the backbone atoms of inhibitor residues P3-P2' in the substrate binding pocket. The hydrogen bonding distances for common interactions in the SGPB:OMTKY3 complexes are, perhaps, more revealing, particularly those for SGPB:OMTKY3-Ile^{18I} (Tables 3.4, 3.5 and 3.6). These distances are generally greater for this complex in comparison to the others, second only to SGPB:OMTKY3-Pro^{18I}. The proline side chain introduces more drastic changes than any other P1 substituent and will be discussed more thoroughly in Chapter 6.

The largest difference between SGPB:OMTKY3-Ile^{18I} and SGPB:OMTKY3-Leu^{18I} are: the absence of water 71 in the SGPB:OMTKY3-Ile^{18I} structure and the movement (≅1 Å) of the main-chain atoms from Gly^{215} to Ser^{217} towards the P1 amino acid in the SGPB:OMTKY3-Leu^{18I} structure (Figure 3.12). SGPB:OMTKY3-Val^{18I} and SGPB:OMTKY3-Ile^{18I} share the same number and position (Figures 3.13, 3.17, and 3.18) of waters in the S1 pocket but Val^{18I} sits slightly deeper in the pocket than Ile^{18I}. The comparison between SGPB:OMTKY3-Val^{18I} and SGPB:OMTKY3-Leu^{18I} is similar to that between SGPB:OMTKY3-Ile^{18I} and SGPB:OMTKY3-Leu^{18I}, as water 71 is not present in the SGPB:OMTKY3-Val^{18I} structure and also

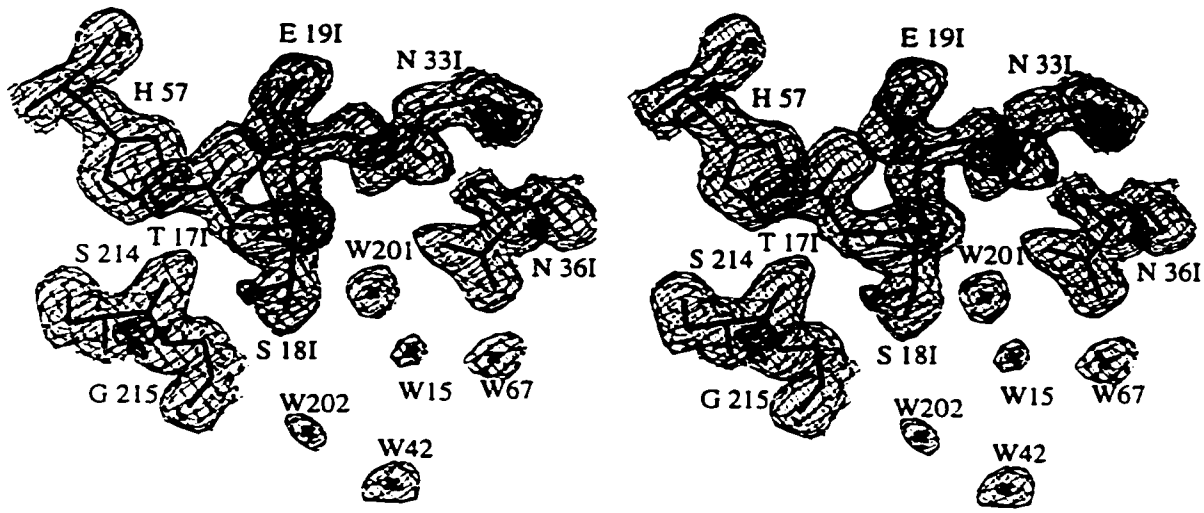

Figure 3.5 The final model of SGPB:OMTKY3-Ser^{18I}, superimposed on electron density maps calculated with an intermediate model that did not include the alternate conformation of Ser^{18I}. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ (thin lines) and $|F_o|-|F_c|$, contoured at 2.5σ (thick lines).

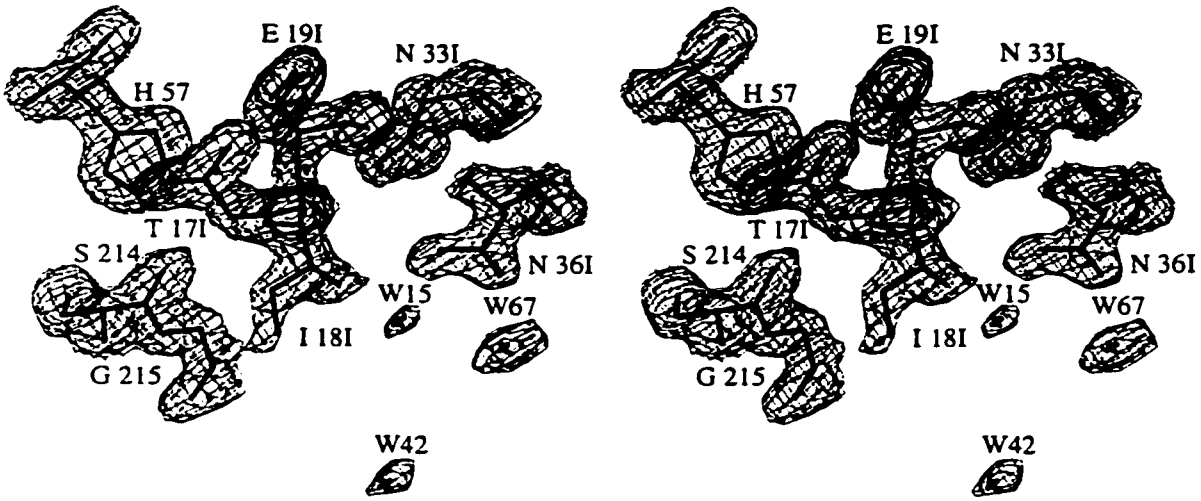

Figure 3.6 Final electron density map of SGPB:OMTKY3-Ile^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

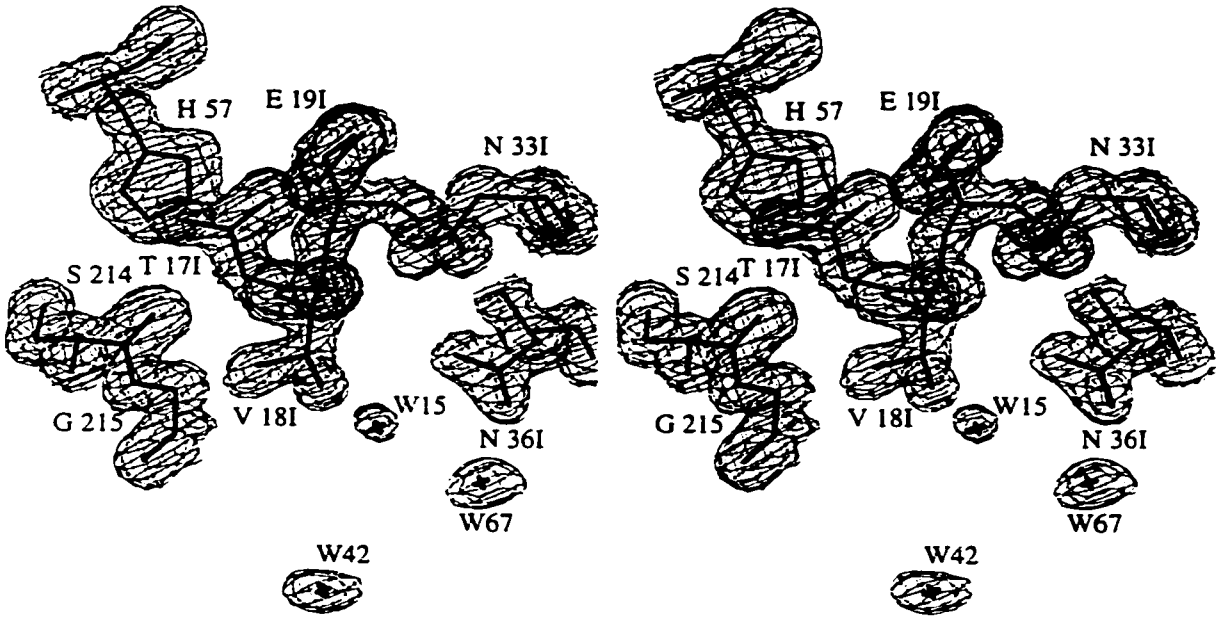

Figure 3.7 Final electron density map of SGPB:OMTKY3-Val^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

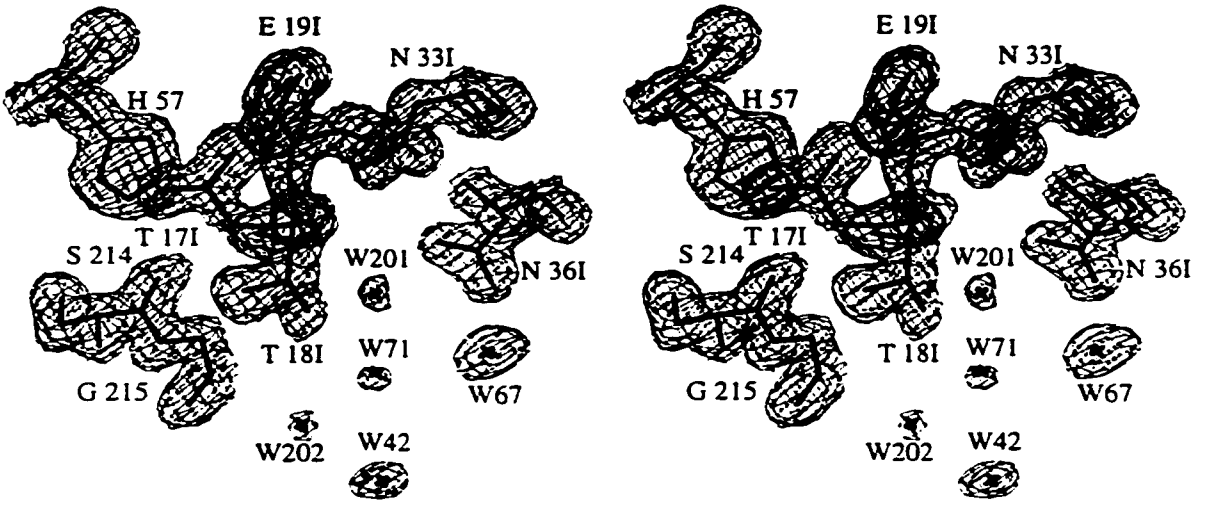

Figure 3.8 Final electron density map of SGPB:OMTKY3-Thr^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

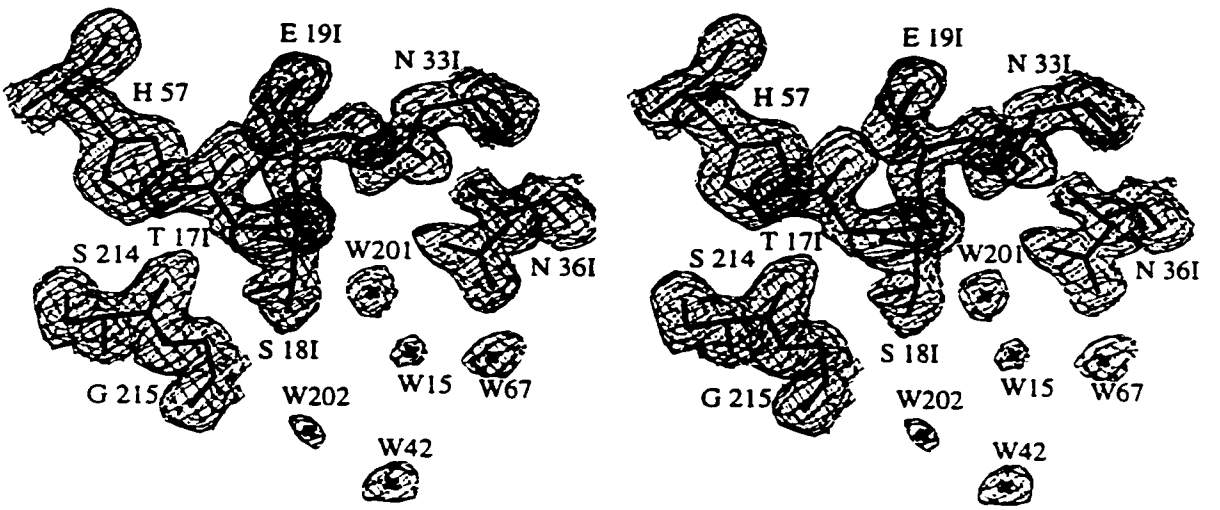

Figure 3.9 Final electron density map of SGPB:OMTKY3-Ser^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

Table 3.2 Refinement Statistics

	Ile^{18I} [a]	Val^{18I}	Thr^{18I}	Ser^{18I}
No. of reflections used	12477	24537	21917	18888
Resolution Range (Å)	20-1.90	20-1.60	20-1.65	20-1.80
Rcryst[b]	0.199	0.174	0.169	0.169
No. of protein atoms[c]	1703	1712	1710	1703
No. of solvent atoms	109	158	164	152
rms deviation from ideal stereochemistry				
bond distance (Å)	0.014	0.009	0.008	0.012
bond angle (°)	1.076	0.906	0.954	1.051
planar groups (Å)	0.013	0.011	0.013	0.013
Average B-values ($Å^2$)				
main-chain atoms	20	15	17	15
side-chain atoms	23	18	21	19
solvent atoms	38	37	39	37
Error Estimates (Å)				
Luzzati (Luzzati, 1952)	0.22	0.16	0.17	0.18
SIGMAA (Read, 1986)	0.28	0.16	0.19	0.21

[a]The column headings refer to the SGPB:OMTKY3 complex with that amino acid at the P1 position.

[b]$Rcryst=\Sigma_{hkl} \, | \, |F_o| - |F_c| \, | \, / \Sigma_{hkl} \, |F_o|$
R-factors were calculated with all of the data in the resolution range indicated and without a $\sigma(I)$ cutoff.

[c]The number of protein atoms includes those atoms from alternate conformations.

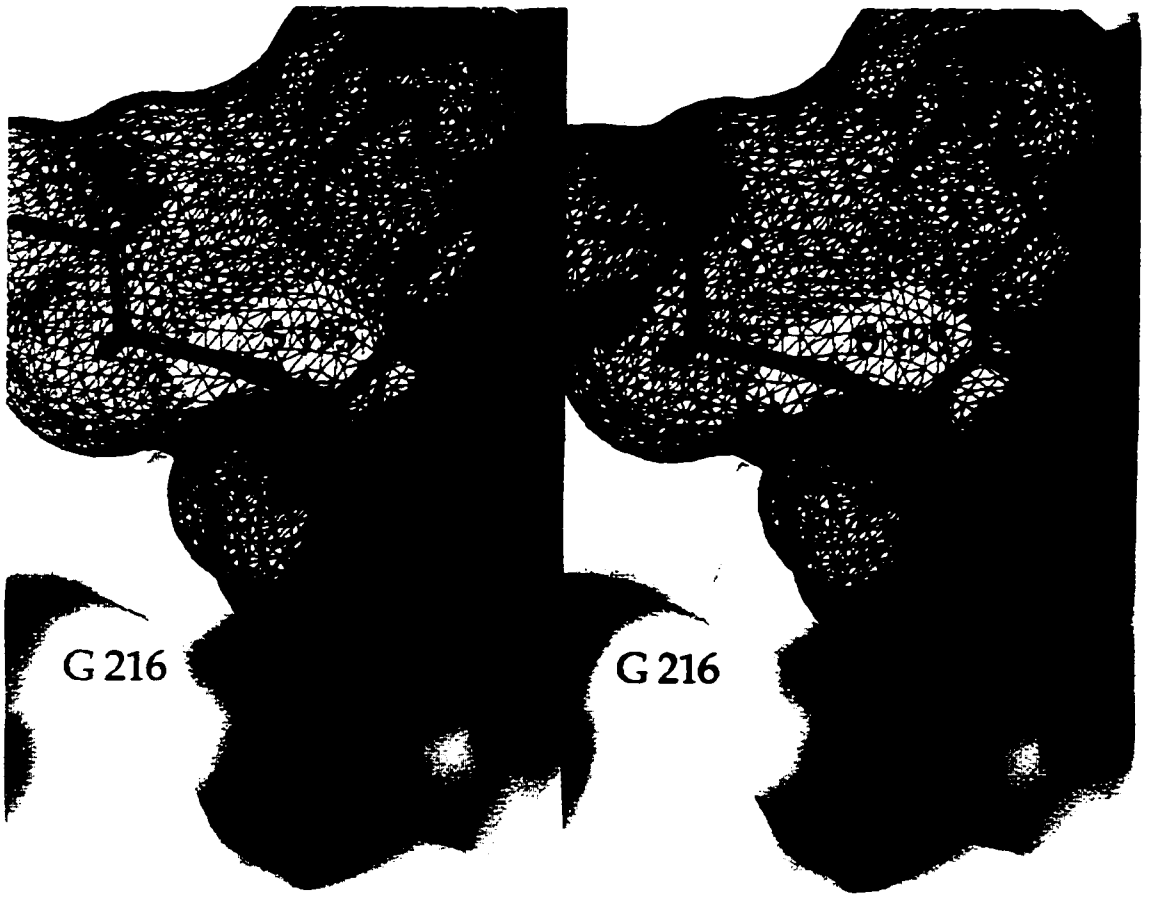

Figure 3.10 Molecular surface of SGPB, white, in complex with OMTKY3-Val18I in the region of the S1 specificity pocket. The molecular surface of residues from the inhibitor has been depicted as a mesh around the atoms (oxygens are red, nitrogens are blue and carbon atoms are gold). The close approach of Ser195 Oγ to the carbonyl carbon of the P1 residue can be seen as an intersection between the molecular surfaces of enzyme and inhibitor. Surface areas were calculated from van der Waals radii with the program GRASP.

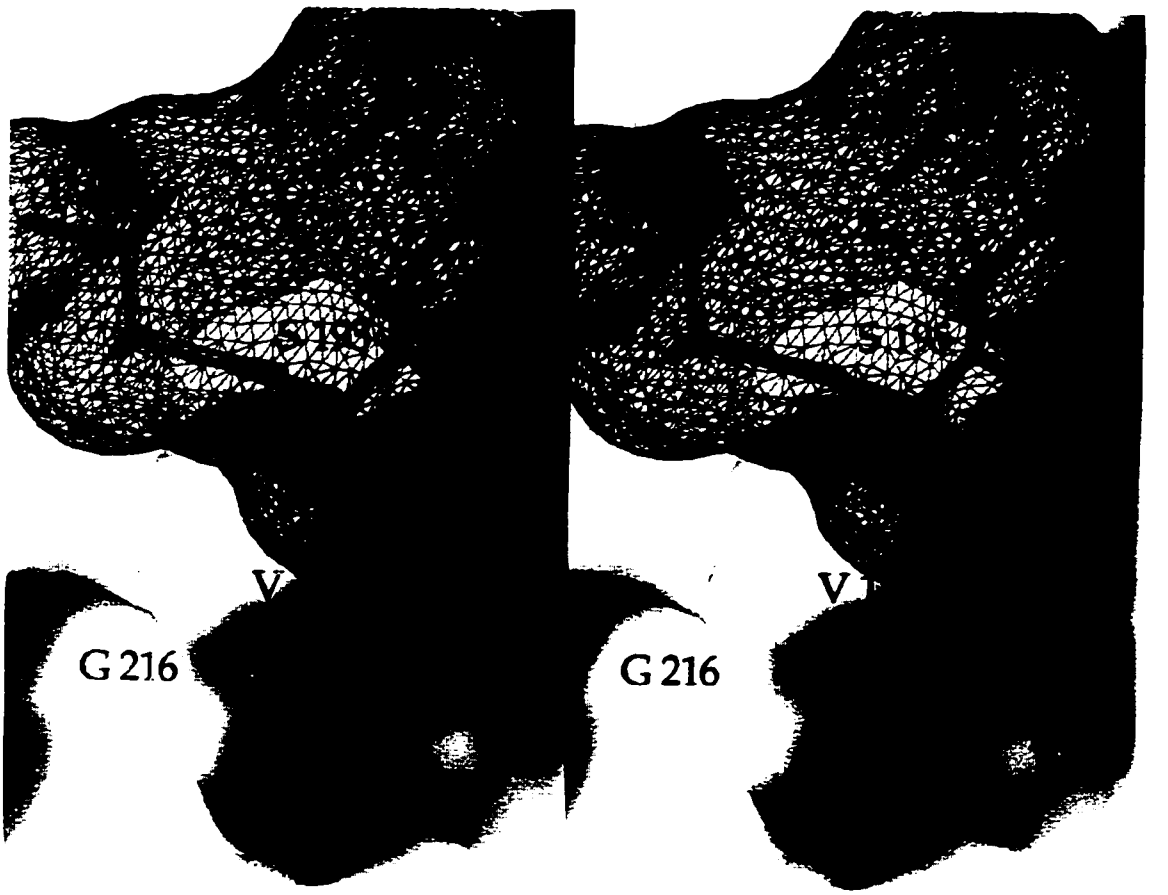

Figure 3.11 Molecular surface of SGPB, white, in complex with OMTKY3-Val^{18I} in the region of the S1 specificity pocket. The P1 residue has been rotated 180° about χ_1 from its actual orientation.

Table 3.3 Rms differences (Å) of main-chain atoms among the structures of the various SGPB:OMTKY3 complexes.

	Ile^{18I}	Val^{18I}	Thr^{18I}	Ser^{18I}
Leu^{18I}	0.15[1]	0.13	0.13	0.13
	0.16[2]	0.14	0.13	0.15
	0.17[3]	0.15	0.16	0.15
Ile^{18I}		0.11	0.12	0.12
		0.16	0.16	0.17
		0.13	0.15	0.14
Val^{18I}			0.05	0.06
			0.06	0.08
			0.06	0.06
Thr^{18I}				0.06
				0.08
				0.07

[1]The top number in each column refers to calculations among the SGPB molecules of the SGPB:OMTKY3 complexes. 740 main chain atoms (N, Cα, C, O) from 185 residues were used in each calculation.
[2]The middle number in each column refers to calculations among the OMTKY3 variant molecules of the SGPB:OMTKY3 complexes. 204 main chain atoms from 51 residues were used in each calculation.
[3]The bottom number in each column refers to calculations among the SGPB:OMTKY3 complexes. 944 main chain atoms (740 from SGPB and 204 from OMTKY3) from 236 residues were used in each calculation.

lacks the movement of the main chain from Gly^{215} to Ser^{217} towards P1. There was a small positive peak in the electon density ($|F_o| - |F_c|$, 2.5σ) at the position of water 71 in the SGPB:OMTKY3-Val^{18I} maps but a water molecule refined poorly at this position and had a very high B-factor. SGPB:OMTKY3-Val^{18I} and SGPB:OMTKY3-Thr^{18I} have the same number of atoms in the P1 residue but have significant differences in the solvent distributions of their S1 pockets (Figure 3.15). Thr^{18I} is slightly more buried in the S1 pocket than is Val^{18I} and Thr Oγ makes a hydrogen bond with Ser195 (Figure 3.19). Water 15 from SGPB:OMTKY3-Val^{18I} is 1.7 Å away from water 201 from SGPB:OMTKY3-Thr^{18I} and therefore these were deemed to be separate water sites. Both structures share waters 42 and 67. SGPB:OMTKY3-Thr^{18I} additionally has waters 71 and 202.

The alternate conformation of Ser^{18I} Oγ (Ser^{18I} OγB) superimposed onto Thr^{18I} Oγ extremely well after superimposition of the enzymes (Figure 3.16). The Ser Cβ-SerOγA bond is 28° from the Thr Cβ-Cγ bond. The S1 pockets of both complex structures share four waters, 42, 67, 201 and 202. In addition SGPB:OMTKY3-Ser^{18I} water 15 is 2.8 Å away from SGPB:OMTKY3-Thr^{18I} water 71. Hydrogen bonding patterns for these two complexes (Figures 3.19 and 3.20) are slightly different even though most of the waters are in the same position. Hydrogen bonding analysis of the SGPB:OMTKY3-Ser^{18I} S1 pocket has been done with trepidation due to the fact that two alternate conformations of Ser^{18I} Oγ are present and it is not possible to tell which water positions accompany which alternate conformation.

The volume of the S1 cavity with Ser^{18I} in the A conformation was 76 $Å^3$, calculated by voidoo (Kleywegt and Jones, 1994) and 82 $Å^3$, calculated by Grasp (Nicholls *et al.*, 1991) (Figure 3.21). The cavity size with Ser^{18I} in the B conformation was 12 $Å^3$ larger as calculated by voidoo, and 19 $Å^3$ larger as calculated by Grasp (Figure 3.22). The volume of the S1 cavity for SGPB:OMTKY3-Leu^{18I} is 18 $Å^3$. When the P1 leucine residue is substituted by a cysteine residue with a χ_2 angle of -60°, the cavity size is identical, 18 $Å^3$.

Table 3.4 Hydrogen bond distances (Å) for common SGPB:OMTKY3 interactions

	Atoms participating in a hydrogen bond							
SGPB OMTKY3	G216 O P3 NH	G216 NH P3 O	S214 O P1 NH	S195 Oγ P1 NH	G193 NH P1 O	S195 NH P1 O	R41 O P2' NH	R41 NH P2' O
Variant								
Leu^{18I}	2.94	2.96	3.67	2.80	2.49	3.07	2.98	3.11
Ile^{18I}	3.05	3.01	3.60	2.84	2.73	3.10	2.88	3.18
Val^{18I}	2.93	2.94	3.53	2.90	2.65	3.00	2.96	3.10
Thr^{18I}	2.93	2.96	3.48	2.91	2.61	3.05	2.97	3.09
Ser^{18I}	2.92	2.95	3.40	2.83	2.58	3.02	2.96	3.05
Trp^{18I}	2.97	2.87	3.46	2.80	2.59	3.08	2.95	3.06
His^{18I}	2.99	2.93	3.53	2.86	2.60	3.08	2.91	3.08
Phe^{18I}	2.93	2.98	3.59	2.88	2.55	3.04	2.97	3.03
Tyr^{18I}	2.95	2.97	3.67	2.71	2.59	2.88	2.92	3.12
Arg^{18I}	2.86	2.90	3.57	2.82	2.64	2.99	2.96	3.13
Lys^{18I+}	2.90	2.95	3.60	2.79	2.56	2.96	2.93	3.06
$Lys^{18I\circ}$	2.88	2.90	3.63	2.82	2.59	3.05	2.93	3.14
$COO\text{-}Leu^{18I}$	2.98	2.94	4.00	2.78	2.55	3.15	2.94	3.13
Pro^{18I}	2.83	3.34	4.08	3.35	2.40	3.55	3.37	3.34

The row headings refer to the SGPB:OMTKY3 complex with that residue at the P1 position.

Table 3.5: Hydrogen bond distances (Å) for common interactions between the catalytic residues

	Atoms participating in a hydrogen bond		
	S195 Oγ	H57 Nδ1	D102 Oδ1
	H57 Nε2	D102 Oδ1	S214 Oγ
Variant			
Leu^{18I}	2.60	2.84	2.63
Ile^{18I}	2.71	2.94	2.86
Val^{18I}	2.62	2.94	2.58
Thr^{18I}	2.58	2.90	2.59
Ser^{18I}	2.53	2.86	2.52
Trp^{18I}	2.55	2.84	2.60
His^{18I}	2.59	2.84	2.60
Phe^{18I}	2.62	2.88	2.56
Tyr^{18I}	2.66	2.79	2.53
Arg^{18I}	2.57	2.81	2.62
Lys^{18I+}	2.55	2.86	2.52
Lyn$^{18I\circ}$	2.52	2.79	2.58
COO-Leu^{18I}	2.61	2.87	2.67
Pro^{18I}	2.75	2.97	2.37

The row headings refer to the SGPB:OMTKY3 complex with that residue at the P1 position.

Table 3.6 Hydrogen bond distances (Å) for common interactions between OMTKY3 residues

	Atoms participating in a hydrogen bond			
	P1' Oε1	P1' Oε1	P2 O	P1' O
	P2 Oγ1	P1' NH	P15' Nδ2	P15' Nδ2
Variant				
Leu^{18I}	2.72	2.60	2.97	2.94
Ile^{18I}	2.87	2.52	3.12	3.09
Val^{18I}	2.66	2.62	3.01	3.06
Thr^{18I}	2.60	2.63	3.03	3.03
Ser^{18I}	2.68	2.67	2.94	3.04
Trp^{18I}	2.68	2.64	3.08	2.94
His^{18I}	2.66	2.66	3.01	3.00
Phe^{18I}	2.56	2.70	2.96	2.97
Tyr^{18I}	2.65	2.70	2.95	3.05
Arg^{18I}	2.63	2.60	2.99	3.00
Lys^{18I+}	2.59	2.70	3.05	3.00
Lys$^{18I\circ}$	2.67	2.67	2.93	2.97
COO-Leu^{18I}	2.75	2.55	3.03	3.03
Pro^{18I}	3.17	3.08	3.08	2.70

The row headings refer to the SGPB:OMTKY3 complex with that residue at the P1 position.

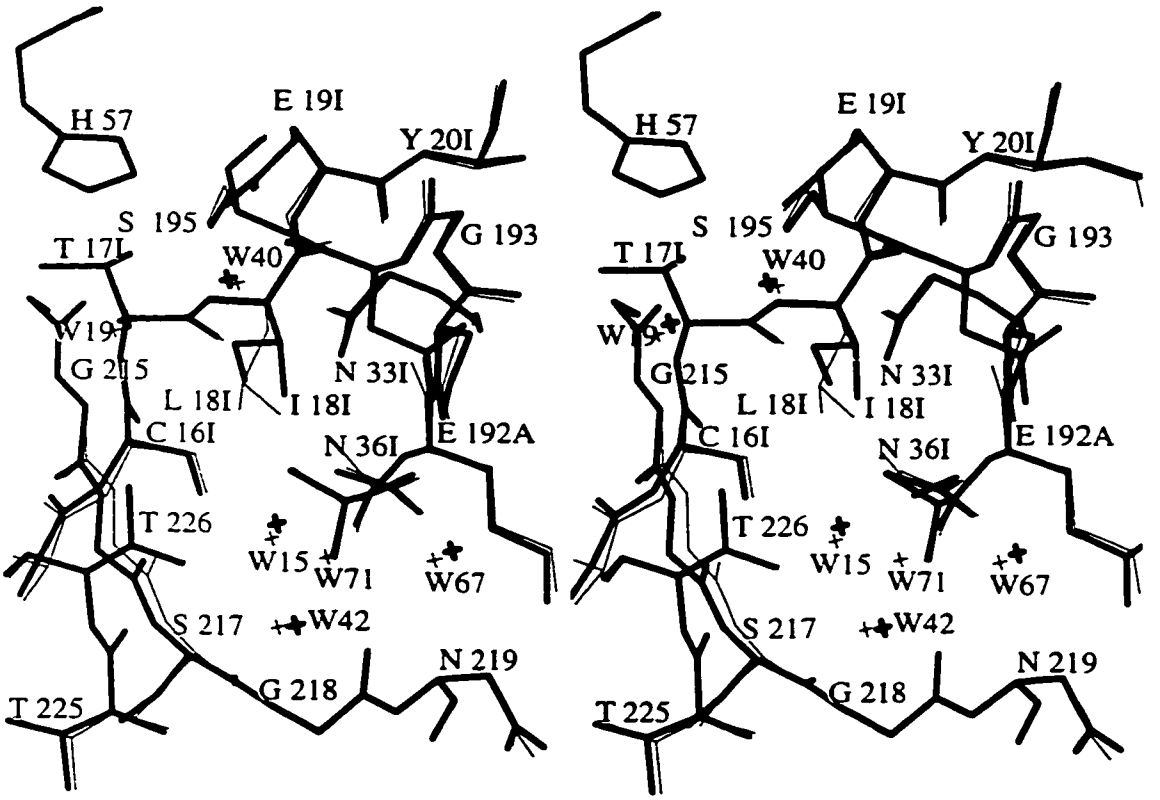

Figure 3.12 Superimposition of SGPB:OMTKY3-Ile^{18I} (thick lines) onto SGPB:OMTKY3-Leu^{18I} (thin lines) in the region of the active site and S1 specificity pocket. Water molecules have been drawn as crosses.

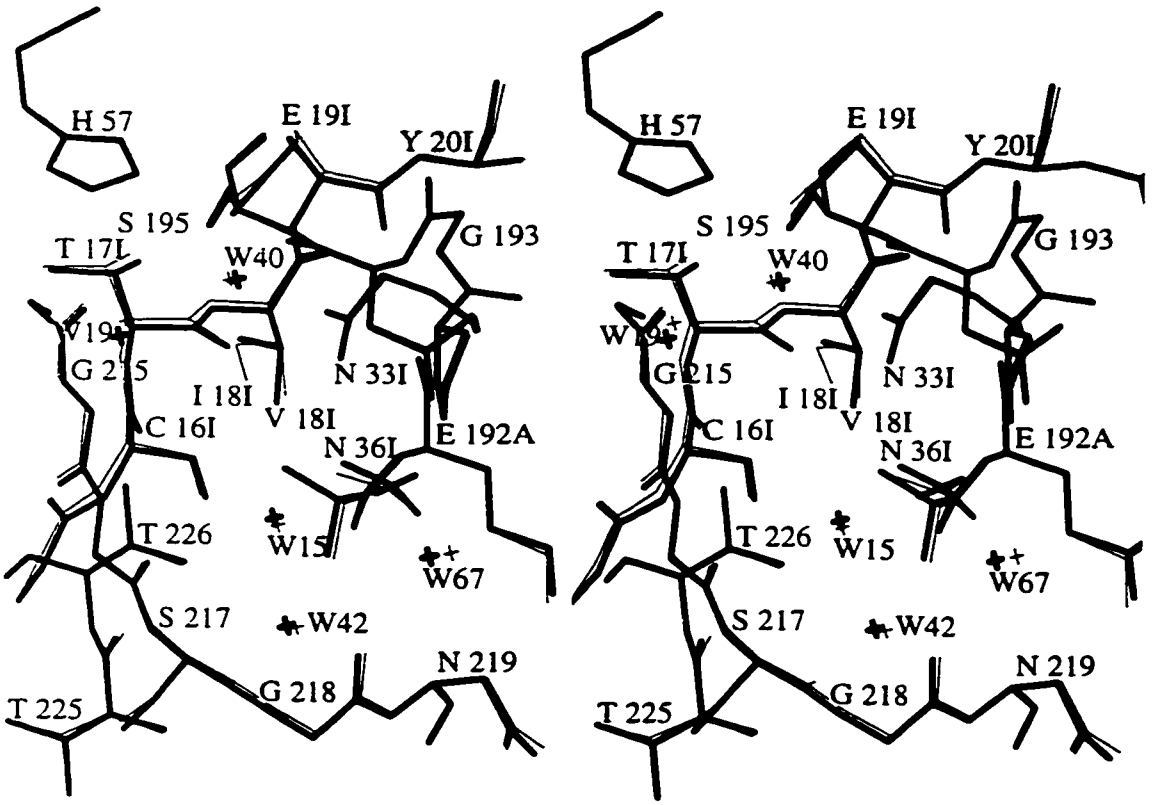

Figure 3.13 Superimposition of SGPB:OMTKY3-Val^{18I} (thick lines) onto SGPB:OMTKY3-Ile^{18I} (thin lines) in the region of the active site and S1 specificity pocket. Water molecules have been drawn as crosses.

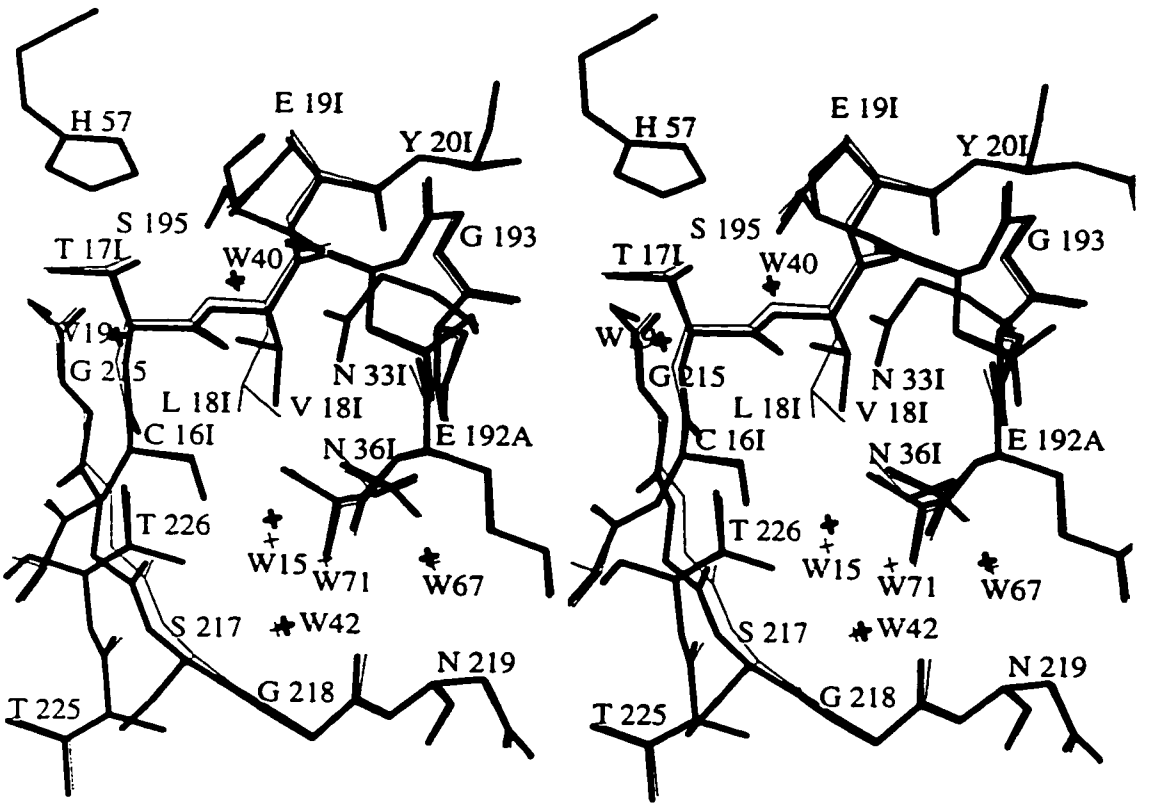

Figure 3.14 Superimposition of SGPB:OMTKY3-Val^{18I} (thick lines) onto SGPB:OMTKY3-Leu^{18I} (thin lines) in the region of the active site and S1 specificity pocket. Water molecules have been drawn as crosses.

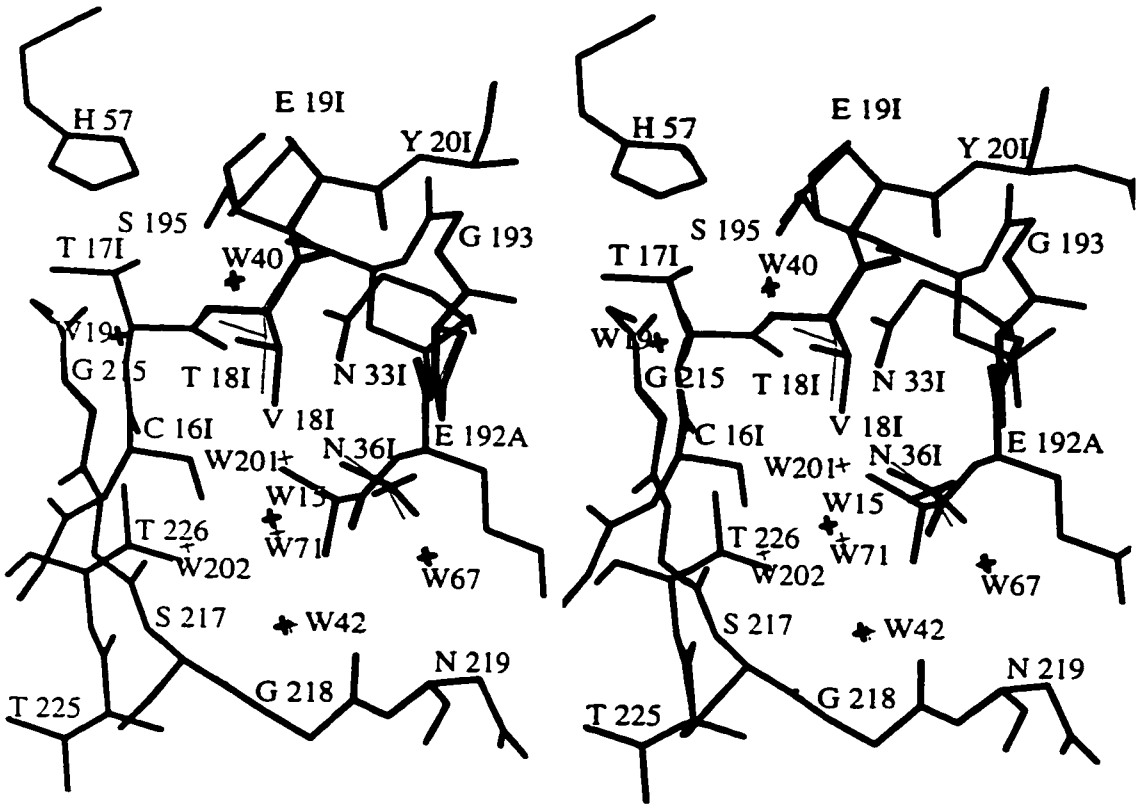

Figure 3.15 Superimposition of SGPB:OMTKY3-Val^{18I} (thick lines) onto SGPB:OMTKY3-Thr^{18I} (thin lines) in the region of the active site and S1 specificity pocket. Water molecules have been drawn as crosses.

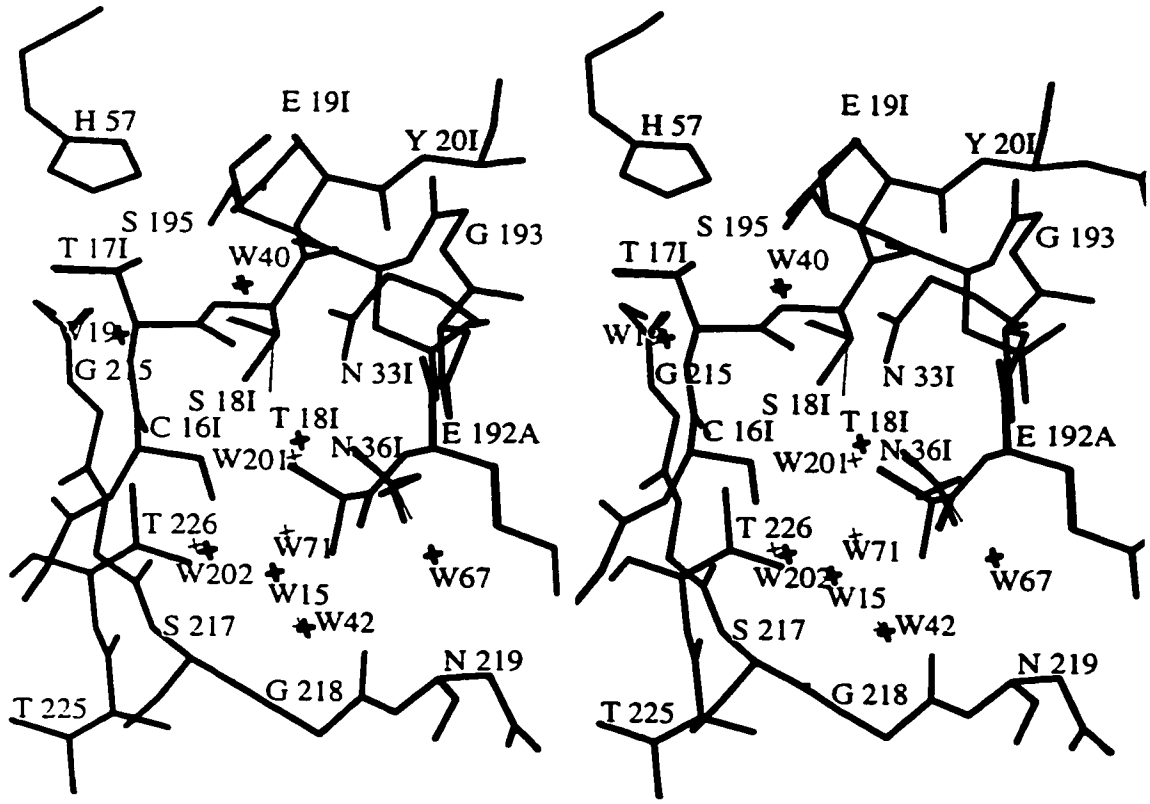

Figure 3.16 Superimposition of SGPB:OMTKY3-Ser^{18I} (thick lines) onto SGPB:OMTKY3-Thr^{18I} (thin lines) in the region of the active site and S1 specificity pocket. Water molecules have been drawn as crosses.

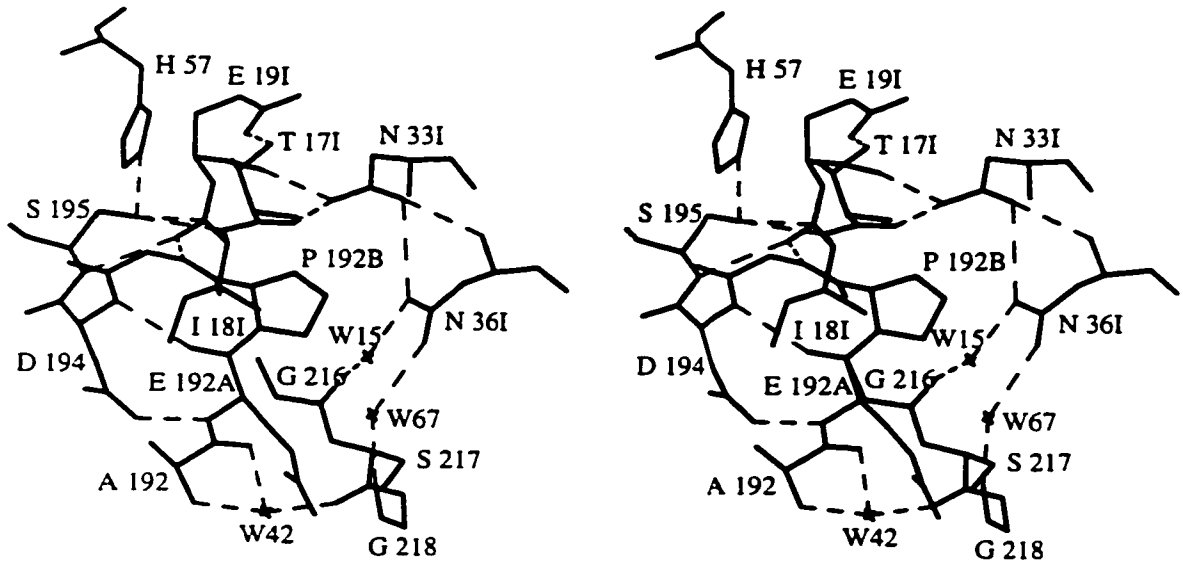

Figure 3.17 Stereoscopic view of the active site and S1 pocket from the SGPB:OMTKY3-Ile^{18I} complex. Water molecules have been depicted as crosses. Hydrogen bonds have been indicated by dashed lines.

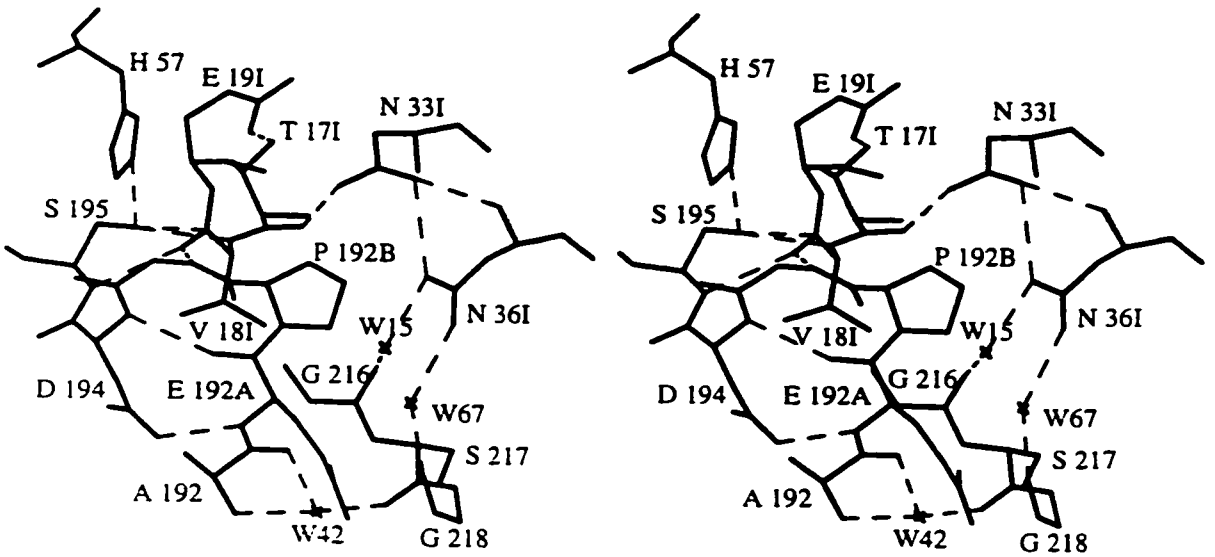

Figure 3.18 Stereoscopic view of the active site and S1 pocket from the SGPB:OMTKY3-Val^{18I} complex. Water molecules have been depicted as crosses. Hydrogen bonds have been indicated by dashed lines.

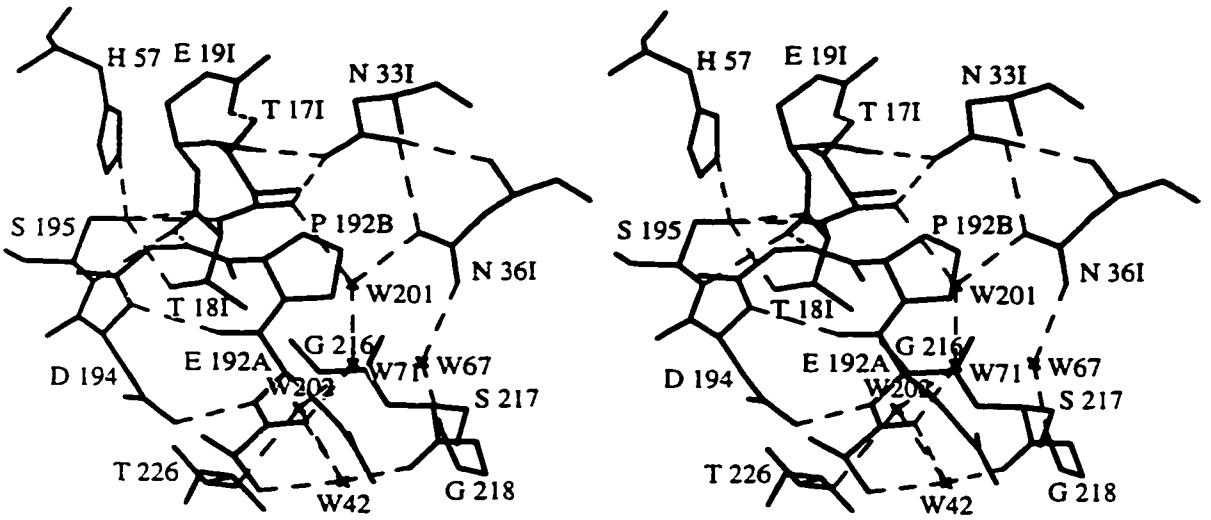

Figure 3.19 Stereoscopic view of the active site and S1 pocket from the SGPB:OMTKY3-Thr^{18I} complex. Water molecules have been depicted as crosses. Hydrogen bonds have been indicated by dashed lines.

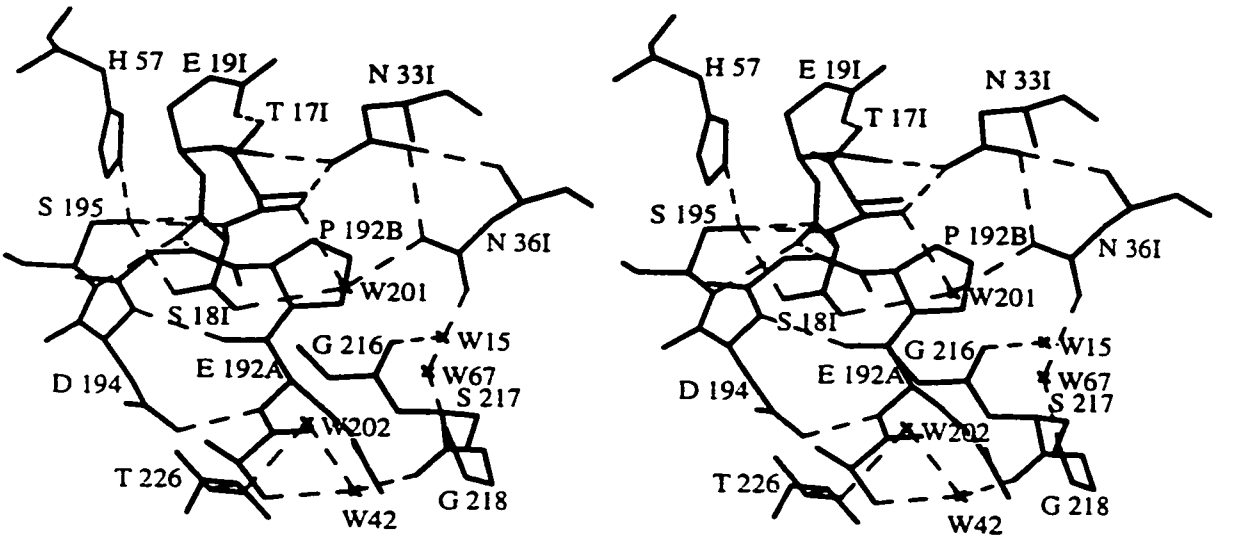

Figure 3.20 Stereoscopic view of the active site and S1 pocket from the SGPB:OMTKY3-Ser^{18I} complex. Water molecules have been depicted as crosses. Hydrogen bonds have been indicated by dashed lines.

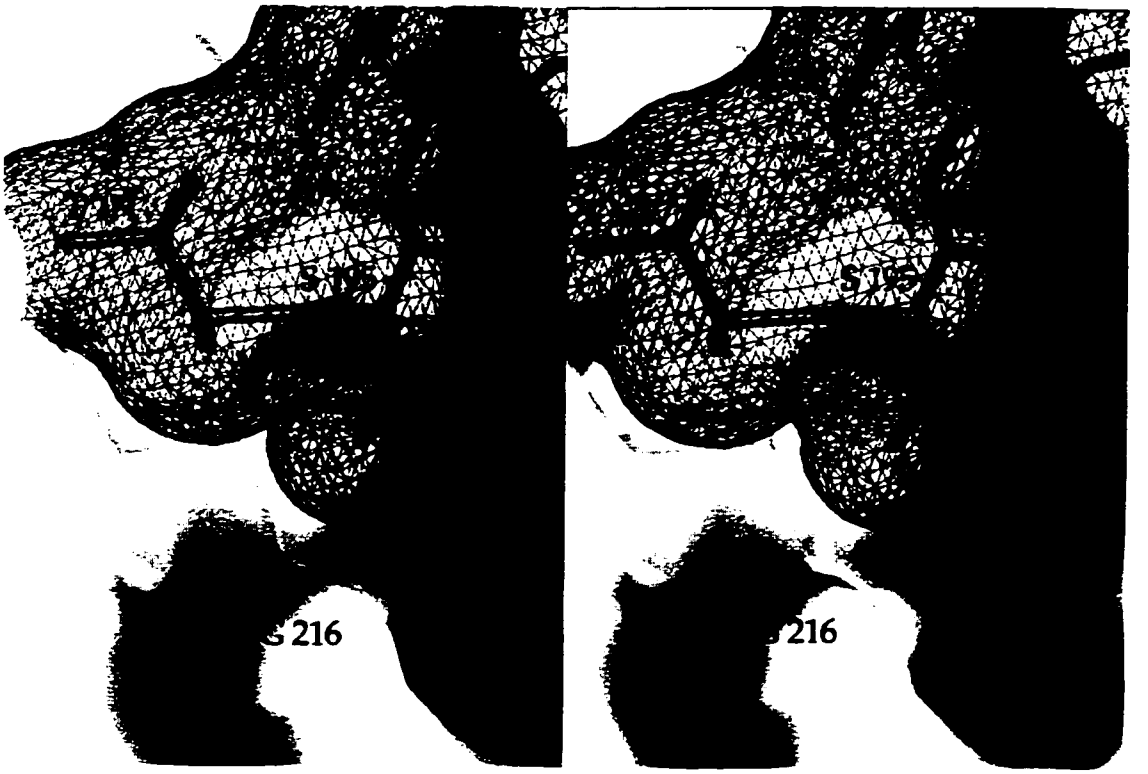

Figure 3.21 Molecular surface of SGPB, white, in complex with OMTKY3-Ser[18I] in the region of the S1 specificity pocket. Only conformation A of the P1 residue has been included. The molecular surface of residues from the inhibitor has been depicted as a mesh around the atoms (oxygens are red, nitrogens are blue and carbon atoms are gold). Surface areas were calculated from van der Waals radii with the program GRASP.

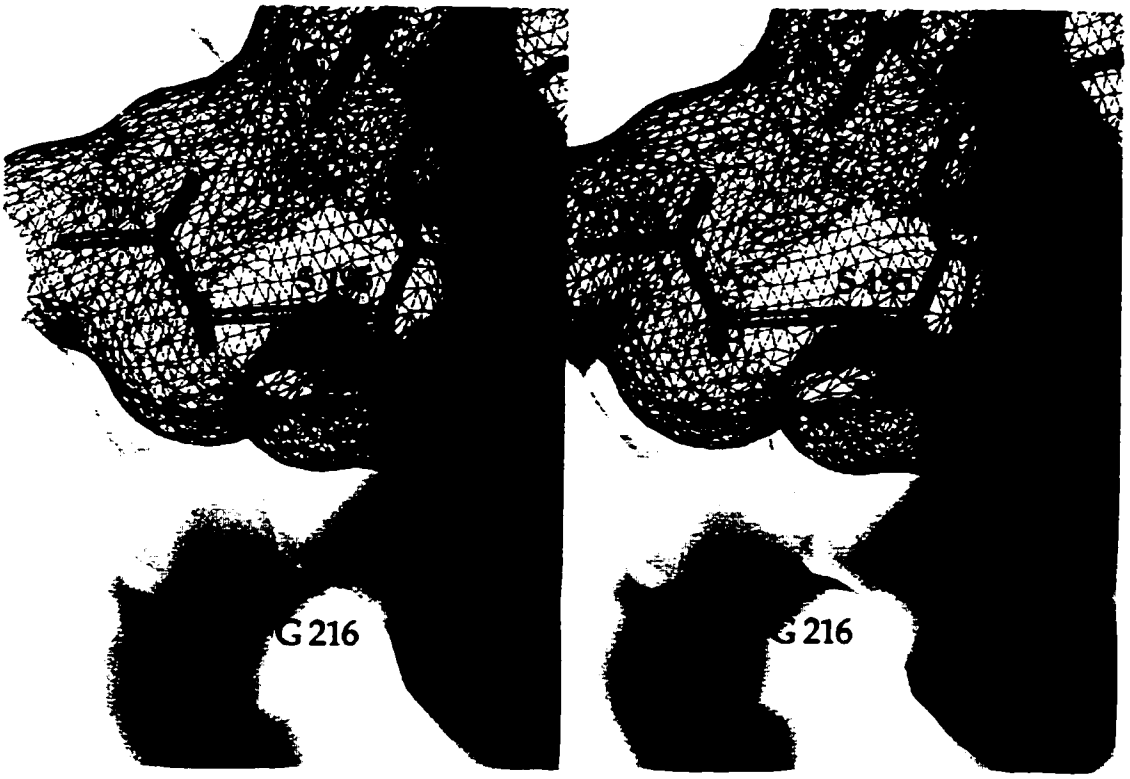

Figure 3.22 Molecular surface of SGPB, white, in complex with OMTKY3-Ser[18I] in the region of the S1 specificity pocket. Only conformation B of the P1 residue has been included. The molecular surface of residues from the inhibitor has been depicted as a mesh around the atoms (oxygens are red, nitrogens are blue and carbon atoms are gold). Surface areas were calculated from van der Waals radii with the program GRASP.

3.4 Discussion

A β-branched residue at the P1 position of OMTKY3 is generally deleterious for most of the serine proteinases with the exception of the elastases (Lu *et al.*, 1997; Bigler *et al.*, 1993). This is particularly well illustrated by the K_a's for enzymes chymotrypsin, subtilisin, SGPA and SGPB (Table 1.2). SGPB, like chymotrypsin prefers hydrophobic P1 side chains that are not branched at the β-carbon. The association constants for SGPB:OMTKY3-Cys^{18I} and SGPB:OMTKY3-Leu^{18I} are the highest of the natural P1 amino acids, followed closely by SGPB:OMTKY3-Met^{18I}. Of these top three P1 variants, only the structure of SGPB:OMTKY3-Leu^{18I} has been determined (Read *et al.*, 1983; Fujinaga *et al.*, 1982; Huang *et al.*, 1995). The Leu at P1 fits ideally into the hydrophobic S1 pocket of SGPB which is narrow at the top (closest to the catalytic residue Ser^{195}) and broadens gradually towards the bottom (Figures 3.12 and 3.23). The β-branched P1 residues do not complement this shape, leaving space at the bottom of the pocket, where it is widest. Cavities are rare in protein:protein recognition sites (Janin and Chothia, 1990) and complementarity involves close packing between proteins. Eriksson *et al.*, 1992, have suggested that the destabilization of a protein in response to cavity creating mutations is proportional to the cavity size. In contrast to the empty bottom, the narrow top of the S1 pocket has to accommodate all of the side chain atoms (Figure 3.10). This 'tight fit' also demands that the γ-atoms of the β-branches point away from Pro^{192B}-Asp^{194} to avoid steric clash with the sides of the pocket (3.11). The observed χ angles for these residues are neither unfavourable nor unusual (McGregor *et al.*, 1987; Swindells *et al.*, 1995).

Of the β-branched P1 variants, SGPB:OMTKY3-Val^{18I} has the highest K_a, 3.3×10^8 (M^{-1}), followed by SGPB:OMTKY3-Thr18I, 2.5×10^8 (M^{-1}) and SGPB:OMTKY3-Ile^{18I}, 2.9×10^7 (M^{-1}) (Lu, 1995; Lu *et al.*, 1997). Although serine is not a β-branched amino acid, the SGPB:OMTKY3-Ser^{18I} complex has been included in this paper because of the similarities between serine and threonine with respect to size and polarity. SGPB:OMTKY3-Ser^{18I} is also an interesting addition to the three β-branched P1 side chains because

Figure 3.23 Molecular surface of SGPB, white, in complex with OMTKY3-Leu^{18I} in the region of the S1 specificity pocket. The molecular surface of residues from the inhibitor has been depicted as a mesh around the atoms (oxygens are red, nitrogens are blue and carbon atoms are gold). Surface areas were calculated from van der Waals radii with the program GRASP.

of the two alternate conformations for Ser^{18I} Oγ. As a result, the P1 residue looks like a β-branched amino acid in the electron density (Figure 3.9). The final reason for including SGPB:OMTKY3-Ser^{18I} is that the K_a is similarly low along with the β-branched P1s, at 5.0×10^7 (M^{-1}).

The P1 valine side chain does not extend into the SGPB P1 pocket, as does an ideal P1 residue like leucine (Figures 3.10, 3.14 and 3.23). The van der Waals interactions between Val^{18I} and the S1 substrate pocket are limited to residues forming the pocket's top surface (Figure 3.10). P1 leucine has additional contact with residues Gly^{215} and Ala^{192} at the sides of the specificity pocket (Figure 3.23). The absence of water 71 from the SGPB:OMTKY3-Val^{18I} structure is possibly because the S1 pocket is emptier compared with the SGPB:OMTKY3-Leu^{18I} structure. Therefore, the interactions required to stabilize a water such that it appears in electron density are not strong enough. In terms of entropy, it would be more favourable to have fewer waters ordered within the pocket. However, the detrimental effects of the β-branched side chain at P1 and a large cavity (Eriksson *et al.*, 1992) appear to be too unfavourable to overcome this entropic advantage.

The SGPB:OMTKY3-Thr^{18I} association constant is slightly worse than that for SGPB:OMTKY3-Val^{18I} (Lu *et al.*, 1997). Oγ of Thr^{18I} replaces the slightly larger methyl group of Val^{18I} (Figure 3.15) and gives the side chain its polar nature. The Oγ makes a hydrogen bond with Ser^{195} Oγ (Figure 3.19). This favourable interaction allows Thr^{18I} to sit further into the S1 pocket compared to Val^{18I} (Figure 3.15). There are also more ordered waters in the S1 pocket of SGPB:OMTKY3-Thr^{18I}. An increase in ordered waters reduces the entropy of the complex and perhaps explains why SGPB:OMTKY3-Thr^{18I} has a slightly lower K_a.

SGPB:OMTKY3-Ile^{18I} displays the weakest association constant of the β-branched P1s with $Ka=2.9 \times 10^7$ (M^{-1}) (Lu *et al.*, 1997). This value is also 2×10^3 fold lower than that of its P1 isoform OMTKY3-Leu^{18I} ($Ka=5.6 \times 10^{10}$ (M^{-1})). Simply stated, the isoleucine side chain is not complementary to the S1 specificity pocket of SGPB. Tables 3.4, 3.5 and 3.6 indicate that the

association between SGPB and OMTKY3-Ile^{18I} is not as snug as the others (with the exception of SGPB:OMTKY3-Pro^{18I}), even among the β-branched P1 residues. The β-branching prevents Ile^{18I} from adopting a χ1 angle of -75°, which would allow more room for Ile^{18I} Cδ1. Instead, at a χ1 angle of 33.2°, Cγ1 and Cδ1 sit back against the top of the pocket while the left side of the pocket, particularly the position occupied by P1 Leu^{18I} Cγ is left empty (Figure 3.12). A comparison between SGPB:OMTKY3-Val^{18I} and SGPB:OMTKY3-Ile^{18I} is interesting because several studies have indicated that a methyl/methylene group contributes 1.0-1.5 kcal/mol towards the stabilization of a protein hydrophobic core, (Kellis *et al.*, 1989), or the stabilization of the hydrophobic interface of an enzyme inhibitor complex (Molina *et al.*, 1994; Fersht and Serrano, 1993). Therefore, one would expect enhanced binding of approximately 1.5 kcal/mol for SGPB:OMTKY3-Ile^{18I} over SGPB:OMTKY3-Val^{18I} because of the extra methylene group. However, the opposite appears to be true. SGPB:OMTKY3-Ile^{18I} is destabilized by 1.4 kcal/mol, (as calculated from the difference in K_a), over SGPB:OMTKY3-Val^{18I}. Except for the addition of the Cδ1 methyl group of Ile^{18I}, the structures are basically identical (Figure 3.13). Therefore, the strain on SGPB:OMTKY3-Ile^{18I} must be steric. This is not the case for a comparison between SGPB:OMTKY3-Val^{18I} and SGPB:OMTKY3-Leu^{18I}. SGPB:OMTKY3-Leu^{18I} also has an additional methylene group with respect to SGPB:OMTKY3-Val^{18I}, but has 3.0 kcal/mol enhanced binding over SGPB:OMTKY3-Val^{18I}. Unlike the SGPB:OMTKY3-Val^{18I}/SGPB:OMTKY3-Ile^{18I} comparison, the S1 pockets of SGPB:OMTKY3-Val^{18I} and SGPB:OMTKY3-Leu^{18I} are not identical (Figure 3.14) and several factors, including the methylene group, must contribute to the enhanced binding of SGPB:OMTKY3-Leu^{18I}. All of these results indicate that the shape of the S1 pocket plays a very important role in dictating the specificity and there is not neccessarily a linear dependence upon hydrophobicity (Dorovska *et al.*, 1972).

The association constant for SGPB:OMTKY3-Ser^{18I} is similarly low (5.0×10^7 (M^{-1})) to that of SGPB:OMTKY3-Ile^{18I}. There are two alternate conformations for the primary specificity determinant, Ser^{18I} in the structure of SGPB:OMTKY3-Ser^{18I} (Figure 3.20). Two conformations and a

poor Ka suggest that the side chain might be somewhat dissatisfied in both positions. However, analysis of two alternate conformations is difficult because one cannot say whether the two positions are sampled by each Ser^{18I} in the crystal or whether some of the serines at the P1 position are in one conformation and the others are in the alternate conformation. The data could also be a reflection of both possibilities.

Ser^{18I} B, the conformation with the lower occupancy, has Oγ siting tightly against the top wall of the S1 pocket, leaving a large cavity between enzyme and inhibitor (Figure 3.22). This position, however, has the advantage of making a stabilizing hydrogen bond to Ser^{195} Oγ (Figure 3.20). In addition, serine favours χ_1 angles near +60° (when its ϕ and φ angles fall into a coil region of the Ramachandran plot) (Schrauber *et al.*, 1993; Swindells *et al.*, 1995; McGregor *et al.*, 1987; Stapley and Doig, 1997). Ser^{18I} B has a χ_1 angle of 40.4°, much closer to the preferred angle than Ser^{18I} A, (χ_1=-45.9°). Ser^{18I} A, the position of higher occupancy, has Oγ siting in the middle of the pocket and more closely following the path of P1 leucine, an ideal P1 residue for SGPB (Figure 3.21). The size of the cavity in the S1 pocket (82 $Å^3$ defined as the volume accessible to a probe of radius 1.4 Å) is smaller than that of Ser^{18I} B (101 $Å^3$). As already mentioned, structural destabilization by a cavity is thought to be proportional to the cavity size (Eriksson *et al.*, 1972). Perhaps the smaller P1 cavity left by Ser^{18I} A contributes towards the stabilization of this orientation. Thus, the serine variant appears to be subjected to competing forces, the negative influence of cavity formation versus the preferred χ_1 conformation and hydrogen bonding to the enzyme.

An intriguing comparison is that between P1 serine and P1 cysteine. The difference in K_a between these two variants is 1.14×10^3 (M^{-1}) , but one would expect their conformations in the specificity pocket to be nearly identical. Unfortunately, no structure has been determined for SGPB:OMTKY3-Cys^{18I}. Considering a model of SGPB:OMTKY3-Cys^{18I} where Sγ of P1 cysteine superimposes onto that of Ser OγA, one can imagine a few reasons why P1 cysteine would be preferred by SGPB. The SH group is more hydrophobic than OH and in addition, is slightly larger.

Therefore, the Cys Sγ would fill up more of the hydrophobic specificity pocket. This aspect would also make it more difficult for the cysteine to move within the pocket and sample alternate conformations. If the SH group were to sample the conformation occupied by Ser^{18I} B, the hydrogen bond would not be as strong as that between OH of Ser^{18I} and Oγ of Ser^{195} and would probably not compete with the entropy lost in creating the large S1 cavity. In fact, it has been observed that protein sulphur atoms are rarely involved in hydrogen bonds (McDonald *et al.*, 1994). Lastly, cysteine favours $\chi 1$ angles of -60° (Schrauber *et al.*, 1993; McGregor *et al.*, 1987; Stapley and Doig, 1997). As previously mentioned, serine prefers a χ_1 angle of +60°. Therefore, there would be less strain on a cysteine in the A conformation in comparison to Ser^{18I}A at P1. Another example of a protein discriminating between a serine and a cysteine is cysteinyl-tRNA synthetase from *E.coli* (Fersht and Dingwall, 1979). It was calculated from an amino acid dependent ATP/pyrophosphate exchange reaction that the error rate for the mis-incorporation of serine for cysteine was less than 10^{-9}. The van der Waals interactions between either -S- or -O- with a $-CH_2-$ group and the small geometry differences between the residues were suggested as the properties that the enzyme would use to distinguish the isosteric amino acids.

Since the Ka for SGPB:OMTKY3-Cys^{18I} is identical to that of SGPB:OMTKY3-Leu^{18I}, it is tempting to suggest that the SGPB residues, Gly215 to Ser217, follow the same path in the SGPB:OMTKY3-Cys^{18I} complex as they do in the SGPB:OMTKY3-Leu^{18I} structure (Huang et al., 1995). Although this path is only slightly, (within 1 Å), different from the path found in other complex structures (Figures 3.12 and 3.14), it is consistantly found in SGPB:OMTKY3 structures with a leucine at the P1 position (Read *et al.*, 1983; Fujinaga *et al.*, 1982; Huang *et al.*, 1995; SGPB:OMTKY3-COO-Leu^{18I}, Chapter 6) and may be the result of an ideal fit of Leu^{18I} into SGPB's S1 pocket. A result of the SGPB:OMTKY3-Cys^{18I} model, with SGPB coordinates taken from SGPB:OMTKY3-Leu^{18I}, is that the cavity in the S1 pocket is the same size as the cavity found in the

SGPB:OMTKY3-Leu^{18I} structure; as it should be if Cys^{18I} and Leu^{18I} are isofunctional P1 substituents.

Chapter 4:
Crystal structures of SGPB in complex with OMTKY3 aromatic P1 variants Trp^{18I} and His^{18I}

4.1 Introduction

OMTKY3 inhibits serine proteinases by binding tightly to the enzyme and undergoing a slow hydrolysis of its reactive site peptide bond (Laskowski and Kato, 1980). The P1 position on the reactive site loop of OMTKY3 plays a predominant role in the interaction between proteinase and inhibitor. Leu^{18I} is the P1 residue in wild-type OMTKY3. This residue has been substituted by the nineteen other, coded, amino-acid residues and the Ka's for all 20 variants in complex with six serine proteinases have been measured (Table 1.2) (Bigler *et al.*, 1993; Lu *et al.*, 1997). The Ka's for the aromatic P1 variants of OMTKY3 in complex with SGPB are among the highest for the SGPB:OMTKY3 complexes. The X-ray structures of SGPB:OMTKY3-Trp^{18I} and SGPB:OMTKY3-His^{18I} are presented in this chapter and are compared to the structures of the other aromatic P1 variants, SGPB:OMTKY3-Tyr^{18I} and SGPB:OMTKY3-Phe^{18I} determined by Kui Huang. A comparison will also be made with the complex that has the highest Ka among the SGPB:OMTKY3 structures determined, SGPB:OMTKY3-Leu^{18I}.

4.2 Materials and Methods

Dr. L. B. Smillie kindly provided SGPB, purified from Pronase according to the procedure of Jurasek *et al.*, 1979. OMTKY3 P1 variants were generously provided by Dr. M. Laskowski, Jr.. The cloning, overexpression and purification of the OMTKY3 variants have been described (Lu, 1994; Lu *et al.*, 1997a). Aromatic OMTKY3 P1 variants were mixed with SGPB in a 1.5:1.0, OMTKY3/SGPB molar ratio. Crystals were grown from this protein solution by the hanging drop method. 4 to 10% PEG4000 was used as a precipitant and the buffer was $NaKHPO_4$ in a pH range of 6.8 to 7.6. The measured pKa for His^{18I} in the complex is 4.3 (Qasim *et al.*, 1995). Therefore, the side chain will be uncharged in these

crystallization conditions. Streak seeding (Stura and Wilson, 1992) was used in crystallization experiments with lower concentrations of PEG4000(2-4%) in order to improve crystal quality. Suitable crystals of the two SGPB:OMTKY3 complexes were mounted in glass capillaries and exposed to X-rays. Diffraction data from the SGPB:OMTKY3-Trp^{18I} crystal were collected on a SDMS area detector (Hamlin, 1985; Xuong *et al.*, 1985). Latticepatch (Klinger and Kretsinger, 1989) was used to choose the series of ϕ, χ and ω settings for the goniometer neccessary to cover the required volume of the unique part of the reciprocal lattice. The SGPB:OMTKY3-His^{18I} data set was collected with a DIP image plate detector (Mac Science Co., Ltd.). Graphite monochromated X-rays were generated by a Rigaku rotating anode operating at 40 kV and 150 mA for the SGPB:OMTKY3-Trp^{18I} data set. The Rigaku generator was operated at 45 kV and 75 mA and the beam was aligned with double mirror focusing optics for the SGPB:OMTKY3-His^{18I} data collection. Data collection typically lasted for three days.

Intensities from the SDMS data set were processed with the San Diego software (Howard *et al.*, 1985). Denzo and Scalepack (Otwinowski and Minor, 1996) were used for the DIP data.

The structure of native OMTKY3 in complex with SGPB (Read *et al.*, 1983; Fujinaga *et al.*, 1982), determined in the same space group and unit cell as the two complex crystals described in this paper, was used as the model for the initial phase calculations. Solvent molecules from this model were not included but leucine was retained as the P1 residue. $2|F_o|-|F_c|$ electron density maps were calculated with CCP4 SIGMAA (CCP4, 1994; Read, 1986) and inspected either with O (Jones *et al.*, 1991) (SGPB:OMTKY3-Trp^{18I}) or Xtalview (McRee, 1993) (SGPB:OMTKY3-His18I). The N-terminal residue of the OMTKY3 variants, Val^{6I}, was added to the search model. The first six OMTKY3 amino acids, including Val^{6I}, were highly mobile in the SGPB:OMTKY3 structure. For this reason, the OMTKY3 variants were synthesized without the first five amino acids and began with Val^{6I}. The P1 leucine was substituted with the appropriate residue and the model was subjected to simulated

annealing and energy minimization in X-PLOR (Brünger, 1992b). Electron density maps with coefficients $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ were calculated and inspected along with the new models. Waters from the native OMTKY3:SGPB complex structure were superimposed onto the density and those that fit $|F_o|-|F_c|$ density contoured at 2.5σ and $2|F_o|-|F_c|$ density contoured at 1σ were included in subsequent models.

All further refinement was carried out with the program TNT (Tronrud, 1992). The same criteria, described above, was used for the addition of new water molecules. Waters were also inspected to ensure that appropriate hydrogen bond donors or acceptors were present. Special attention was paid to waters with B-values above 50 $Å^2$ and those waters with B-values above 70 $Å^2$ were removed.

The imidazole ring of His^{18I} in the SGPB:OMTKY3-His^{18I} structure had two possibile orientations based on the calculated electron density maps; χ_2 of -50.4°, or χ_2 180° away of -129.6°. The structure was refined with His^{18I} in each of the two possible rotamers with TNT and additionally with X-PLOR. Since the resulting B-factors for the side chains did not give a strong indication of the true orientation, the side chain was modeled based on the potential hydrogen bonding patterns.

Electron density maps and molecular models were inspected with programs O (Jones *et al.*, 1992), FRODO (Jones, 1985), and Xtalview (McRee, 1993). Models were subjected to analysis by the Procheck (Laskowski *et al.*, 1993) and Whatcheck (Rodriguez *et al.*, 1998) programs at several stages of refinement and after completion of the refinement process. Error analysis was performed with Biomol and CCP4 SIGMAA (Collaborative Computational Project Number 4., 1994). Figures were made with Xtalview. Pairwise calculations of rms deviations between models were made with Xtalview.

4.3 Results and Discussion

Crystals of the aromatic P1 variants of OMTKY3 in complex with SGPB grew in space group $P2_1$, and in the same unit cell as the complex of native OMTKY3 and SGPB (Read *et al.*, 1983) (Table 4.1). High resolution data sets were collected to 1.95 Å (SGPB:OMTKY3-Trp^{18I}) and 1.80 Å (SGPB:OMTKY3-His^{18I}). Additional data collection statistics can be found in Table 4.1.

Initial electron density maps superimposed onto the final models are in Figures 4.1 and 4.2. SGPB in complex with native OMTKY3 was used to calculate the map parameters $|F_c|$ and α_c. The phasing model maintained leucine as the P1 residue, but the actual P1 side chain was evident from the initial map for each of the complex structures (Figures 4.1 and 4.2). However, the orientation of the histidine imidazole ring at the P1 position was not obvious. Refinement of the two possible χ_2 rotamers and comparison of the results did not clarify the orientation of the histidine. The model presented in this paper has a χ_2 angle of -50.4° and was chosen based on hydrogen bonding analysis. In the present conformation, Nε2 donates a hydrogen bond to water 202 and Nδ1 hydrogen bonds to water 201. In a 'flipped' orientation, with a χ_2 angle of 130°, the imidazole ring nitrogen would have no hydrogen bonding partners. It is also possible that His^{18I} binds to the S1 pocket of SGPB in either conformation but the model with hydrogen bonds to the imidazole ring nitrogens appears to be the most likely conformation.

Final electron density maps superimposed onto the final models are in Figures 4.3 and 4.4. The Rcryst values are 0.145 and 0.163 for SGPB:OMTKY3-Trp^{18I} and SGPB:OMTKY3-His^{18I}, respectively. Geometry statistics for the completed models are also good and can be found in Table 4.2 with the refinement statistics. 88.3% of the residues from SGPB:OMTKY3-Trp^{18I} and 89.4% of the residues from SGPB:OMTKY3-His^{18I} are found in the most favoured regions of the Ramachandran plot. Only one residue, Asn100 from SGPB, is found in the disallowed region for all of the SGPB:OMTKY3 complex structures. Asn100 is located in a

Table 4.1 Data Collection Statistics

	Trp^{18Ia}	His^{18I}
Space Group	$P2_1$	$P2_1$
Unit Cell a(Å)	45.50	45.44
b(Å)	54.82	54.72
c(Å)	45.63	45.60
β(°)	119.06	119.01
Maximum resolution (Å)	1.95	1.80
Total number of observations	37625	55846
No. of unique reflections	13629	18437
Average redundancy	2.8	3.0
Rmerge[b] :		
overall	0.066	0.060
highest resolution shell	0.228	0.171
Resolution range of last shell (Å)	2.07-1.92	1.85-1.80
$<I/\sigma(I)>$:		
overall	12.44	15.88
highest resolution shell	3.22	3.78
Resolution range of last shell (Å)	1.94-1.92	1.81-1.79
Completeness of Data:		
overall (%)	94.4	89.8
highest resolution shell (%)	70.4	64.7
Resolution range of last shell (Å)	1.98-1.95	1.83-1.80

[a]The column headings refer to the SGPB:OMTKY3 complex with that amino acid at the P1 position.

[b]$Rmerge=\Sigma_{hkl} [(\Sigma_i |I_i-<I>|)/\Sigma_i I_i]$

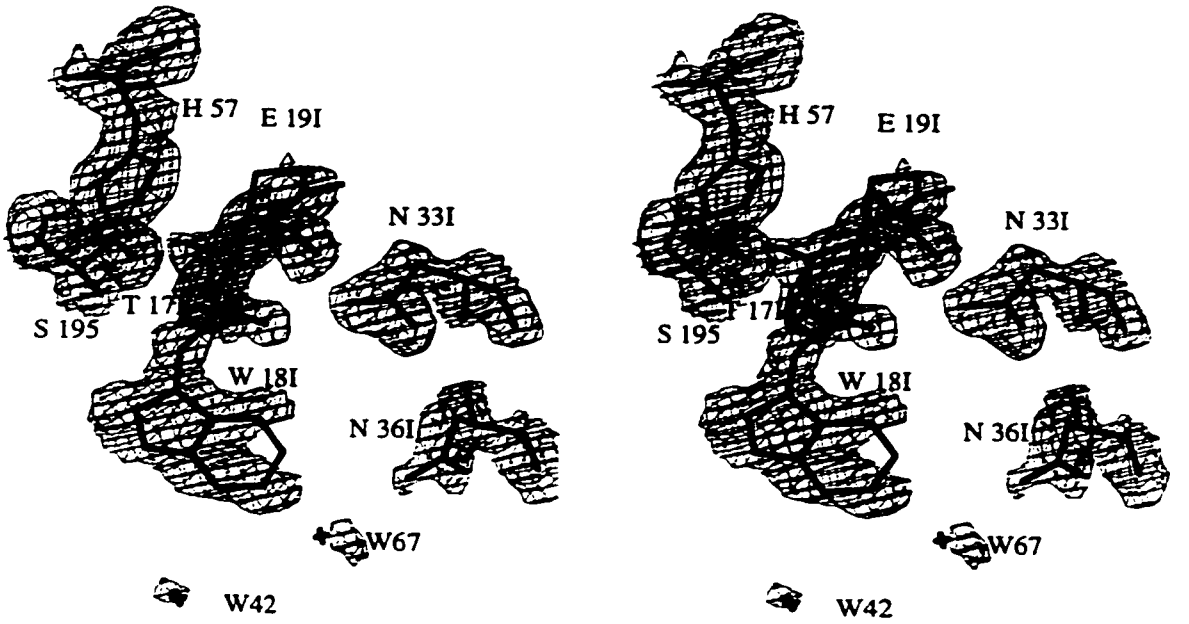

Figure 4.1 Initial electron density map of SGPB:OMTKY3-Trp^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

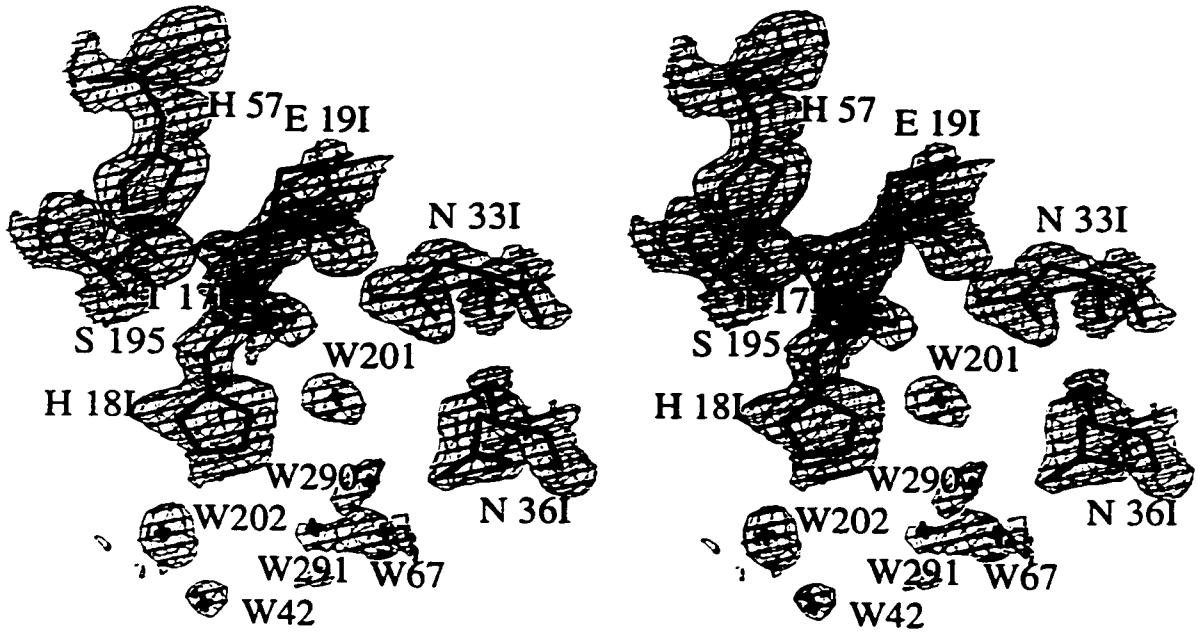

Figure 4.2 Initial electron density map of SGPB:OMTKY3-His[18I] in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

sharp turn following a cis-proline, Pro99A, and the electron density for this residue has been consistantly well defined.

The OMTKY3 variants with aromatic P1 residues have very similar association equilibrium constants in complex with SGPB (Table 1.2 and Figure 1.2) (Lu *et al.*, 1997). These Ka's are also amongst the highest for those SGPB:OMTKY3 variants studied. All of the tightest-binding variants (from SGPB:OMTKY3-Cys^{18I} to SGPB:OMTKY3-Abu^{18I}, Table 1.2) have medium-to-large sized, hydrophobic amino-acids at the P1 position. Of course, P1 residues Tyr^{18I}, Trp^{18I}, and His^{18I} are not entirely hydrophobic. Tyr^{18I} OH makes two hydrogen bonding interactions with water molecules inside the S1 pocket. The imidazole nitrogens of His^{18I} also form hydrogen bonds with water molecules (Figure 4.5). These solvent moleucules appear to be responsible for altering the environment of the S1 pocket in order to accommodate the different polar characteristics of the aromatic rings but also without having a large effect on the Ka's. Water molecules have often been responsible for modulating specificity. A water molecule was shown to replace the CH_2OH group of D-galactose in the interaction of L-arabinose binding protein with L-arabinose. Similarly tight binding was observed for both sugars (Quiocho *et al.*, 1989). The oligopeptide-binding protein is an extreme example of promiscuous ligand binding (Tame *et al.*, , 1994). The lack of side-chain specificity has been accomplished by hydrogen bonds to the main chain of the ligand and by water-mediated interactions with the side chains (Tame *et al.*, 1994). However, even though solvent molecules assist in modulating specificity there should be an entropy cost for ordering additional water molecules in the S1 pocket. Perhaps the extra water molecules that are found in the structures of SGPB:OMTKY3-His^{18I} and SGPB:OMTKY3-Tyr^{18I} explain the lower Ka's for these complexes among the tightest-binding variants (Table 1.2 and Figure 1.2).

In contrast to SGPB:OMTKY3-His^{18I} and SGPB:OMTKY3-Tyr^{18I}, the side-chain nitrogen of Trp^{18I} in the SGPB:OMTKY3-Trp^{18I} complex is without a hydrogen bonding acceptor (Figure 4.6). This unsatisfied hydrogen bond

Table 4.2 Refinement Statistics

	Trp^{18Ia}	His^{18I}
No. of reflections used	13629	18437
Resolution Range (Å)	20-1.95	20-1.80
Rcryst[b]	0.145	0.163
No. of protein atoms[c]	1707	1713
No. of solvent atoms	139	154
rms deviation from ideal stereochemistry		
bond distance (Å)	0.010	0.011
bond angle (°)	1.155	1.013
planar groups (Å)	0.011	0.013
Average B-values ($Å^2$)		
main-chain atoms	13	15
side-chain atoms	17	18
solvent atoms	34	36
Error Estimates (Å)		
Luzzati (Luzzati, 1952)	0.17	0.17
SIGMAA (Read, 1986)	0.14	0.18

[a]The column headings refer to the SGPB:OMTKY3 complex with that amino acid at the P1 position.

[b]$Rcryst=\Sigma_{hkl} \, ||Fo|-|Fc|| / \Sigma_{hkl} \, |Fo|$
R values were calculated with all of the data in the resolution range indicated and without a $\sigma(I)$ cutoff.

[c]The number of protein atoms includes those atoms from alternate conformations.

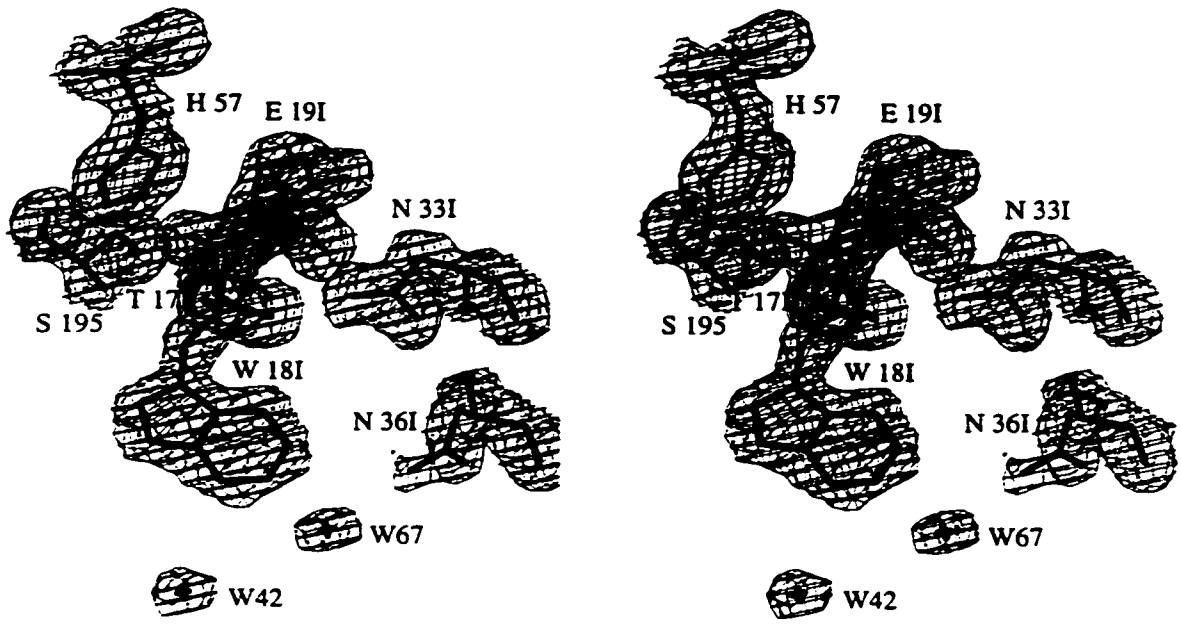

Figure 4.3 Final electron density map of SGPB:OMTKY3-Trp^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

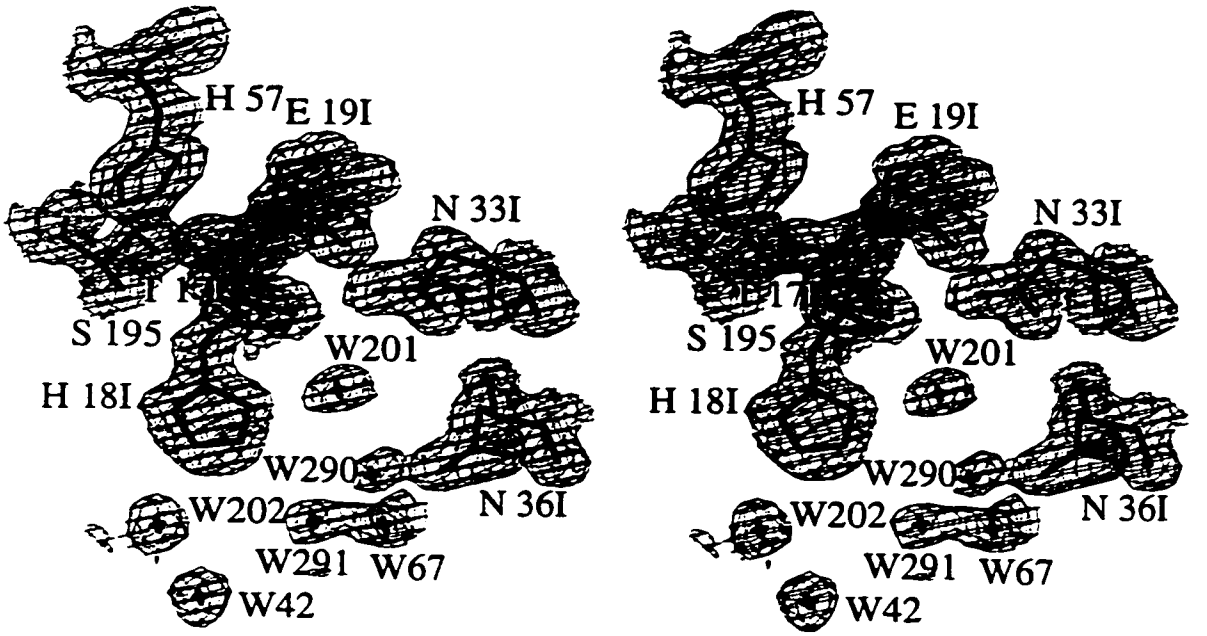

Figure 4.4 Final electron density map of SGPB:OMTKY3-His^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

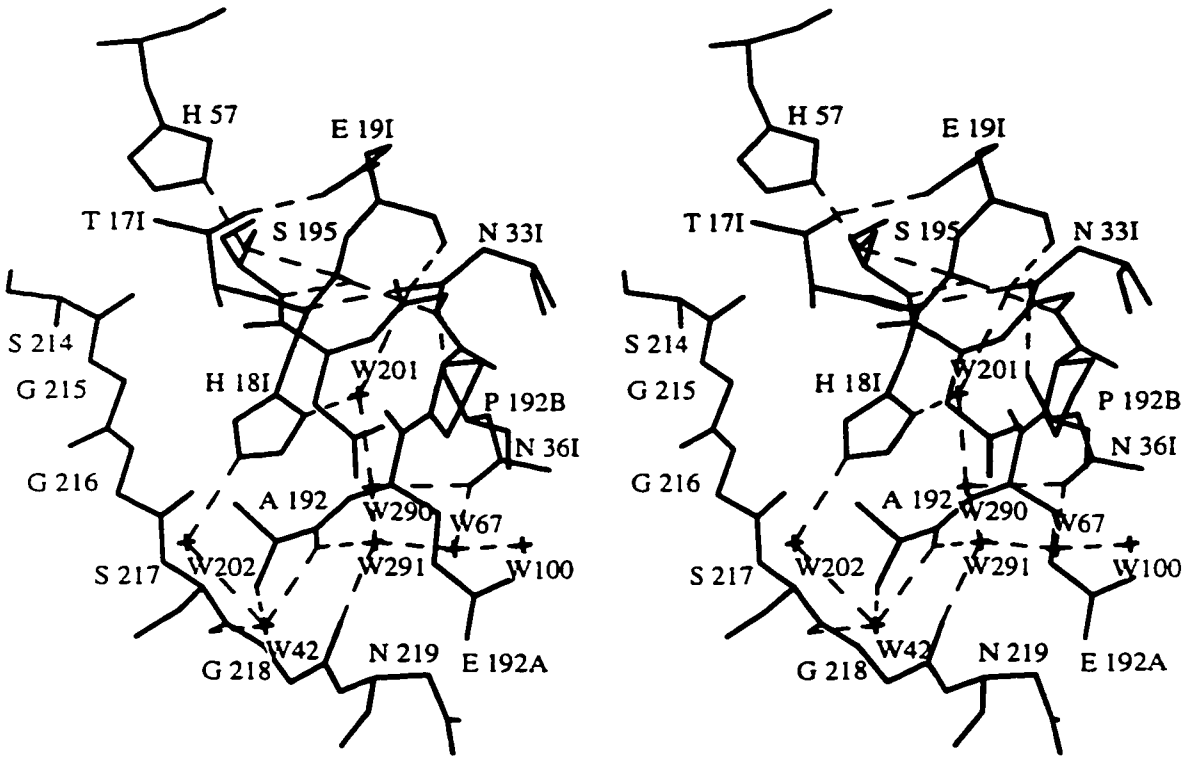

Figure 4.5 Stereoscopic view of the active site and S1 pocket from the SGPB:OMTKY3-His^{18I} complex. Water molecules have been depicted as crosses. Hydrogen bonds have been indicated by dashed lines.

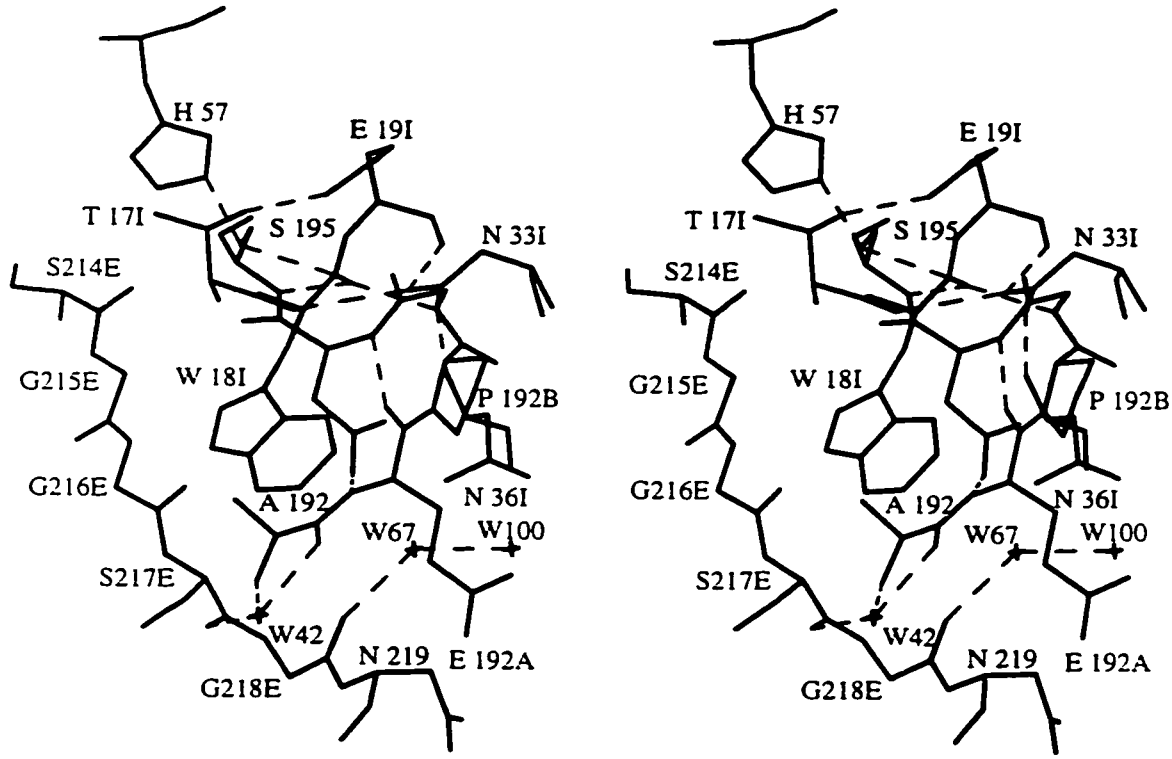

Figure 4.6 Stereoscopic view of the active site and S1 pocket from the SGPB:OMTKY3-Trp^{18I} complex. Water molecules have been depicted as crosses. Hydrogen bonds have been indicated by dashed lines.

may partially account for SGPB:OMTKY3-Trp^{18I} having the lowest Ka among the aromatic P1 complexes.

The aromatic residues have the largest P1 side chains in the OMTKY3 variant population thus far. In the structures of SGPB:OMTKY3-Trp^{18I}, SGPB:OMTKY3-His^{18I}, and SGPB:OMTKY3-Tyr^{18I}, inhibitor residue Asn^{36I} adopts an unusual conformation in comparison to its conformation in other SGPB:OMTKY3 complexes, (Table 4.4), and it turns away from the P1 residue (Figures 4.7 and 4.8). In this way, the side chain bulkiness is accommodated by the inhibitor and not by SGPB's S1 pocket. Atypical conformations of Asn^{36I} have also been observed in the SGPB:OMTKY3-Arg^{18I} and SGPB:OMTKY3-Lys$^{18I\circ}$ complex structures (Table 4.4 and Chapter 5). In these structures, Asn^{36I} helps to accommodate the charge, as well as the large size of the P1 side chains.

It was mentioned in Chapter 3 that the backbone atoms of SGPB residues, Gly215 to Ser217, move slightly towards the P1 residue in the structure of SGPB:OMTKY3-Leu^{18I} (Huang *et al.*, 1995) with respect to most of the other SGPB:OMTKY3 structures (Figures 4.8 and 4.9). It was also suggested that this feature was shared by the P1 residues that best fit the S1 pocket of SGPB. The aromatic P1 variant structures do not display this conformation, (Figures 4.7 and 4.8), even though their association equilibrium constants are relatively high (Table 1.2 and Figure 1.2). Why are the aromatic P1 variants less than ideal? The large size of these P1 residues should be an attribute. SGPB:OMTKY3-Trp^{18I} displaces two water molecules with respect to SGPB:OMTKY3-Leu^{18I} (Figure 4.8) and therefore, it should have lower entropy costs. A larger side chain in the S1 pocket also leaves a smaller cavity in the S1 pocket and, indeed, the cavity volumes for the aromatic SGPB:OMTKY3 complexes are negligable. However, unlike the typical preference for χ_2 angles close to either 60° (g-), 180° (t), or -60° (g+) (Schrauber *et al.*, 1993 ; McGregor *et al.*, 1987 ; Dunbrack and Karplus, 1994 ; Janin *et al.*, 1978), aromatic amino-acids favour χ_2 angles close to 90° and -90°. Figure 4.7 shows a superposition of the SGPB:OMTKY3 complexes that have an aromatic P1 residue. The rings of all four residues lie in the same plane, with a χ_2 of

Table 4.3 RMS differences (Å) of main chain atoms among the structures of the various SGPB:OMTKY3 complexes.

	His^{18I}	Phe^{18I}	Tyr^{18I}
Trp^{18I}	0.10[1]	0.09	0.11
	0.11[2]	0.12	0.13
	0.11[3]	0.10	0.12
His^{18I}		0.11	0.12
		0.11	0.12
		0.12	0.13
Phe^{18I}			0.09
			0.10
			0.10

[1]The top number in each column refers to calculations among the SGPB molecules of the SGPB:OMTKY3 complexes. 740 main chain atoms (N, Cα, C, O) from 185 residues were used in each calculation.
[2]The middle number in each column refers to calculations among the OMTKY3 variant molecules of the SGPB:OMTKY3 complexes. 204 main chain atoms from 51 residues were used in each calculation.
[3]The bottom number in each column refers to calculations among the SGPB:OMTKY3 complexes. 944 main chain atoms (740 from SGPB and 204 from OMTKY3) from 236 residues were used in each calculation.

Table 4.4: χ angles (°) of the aromatic P1 residues and Asn^{36I}.

Residue	Variant	χ_1	χ_2
P1 I18	Trp^{18Ia}	-61.0	-54.9
	His^{18I}	-66.1	-50.4
	Phe^{18I}	-65.0	-48.4
	Tyr^{18I}	-68.8	-45.8
Asn^{36I}	Leu^{18I}	-90.3	-129.2
	Ile^{18I}	-94.7	-125.0
	Val^{18I}	-91.8	-119.8
	Thr^{18I}	-85.6	-109.9
	Ser^{18I}	-87.4	-100.3
	Trp^{18I}	-154.8	-119.4
	His^{18I}	-166.4	80.9
	Phe^{18I}	-94.8	-116.4
	Tyr^{18I}	-147.6	-109.0
	Arg^{18I}	-93.8	29.2
	Lys^{18I+}	-87.2	-130.8
	$Lys^{18I\circ}$	-86.6	73.3
	COO	-93.3	-133.8

[a]The variant headings refer to the SGPB:OMTKY3 complex with that residue at the P1 position. COO refers to SGPB:OMTKY3-COO-Leu^{18I}.

roughly -50° (Table 4.4). In fact, most of the P1 residues are restricted to this same plane (Figure 4.9). This conformation is imposed on the large, aromatic, side chains by the walls of the S1 pocket (residues Ser214-Gly216 and Gly192A-Pro192B) Chymotrypsin also imposes one orientation on the aromatic P1 side chains but since the pocket is deeper than in SGPB, the six-membered ring of the indole group is deepest into the pocket rather than towards the inhibitor as in SGPB:OMTKY3-Trp^{18I} (Steitz *et al.*, 1969). In contrast to the aromatic P1 residues, the large side chains of Ahp^{18I}, Ahx^{18I} and Lys^{18I}, have additional χ angles that can be rotated to assist in accommodation. A linear relationship was observed between the buried hydrophobic surface area and ΔG for the straight-chain, aliphatic P1 variant complexes (Huang *et al.*, 1995) (reproduced in Figure 4.10). This relationship does not hold for the aromatic P1 variant structures. The lower K_as for the aromatic P1s among the top SGPB:OMTKY3 complexes is perhaps due to the strain of the less favourable χ_2 angle that is required for the residues to fit into the pocket.

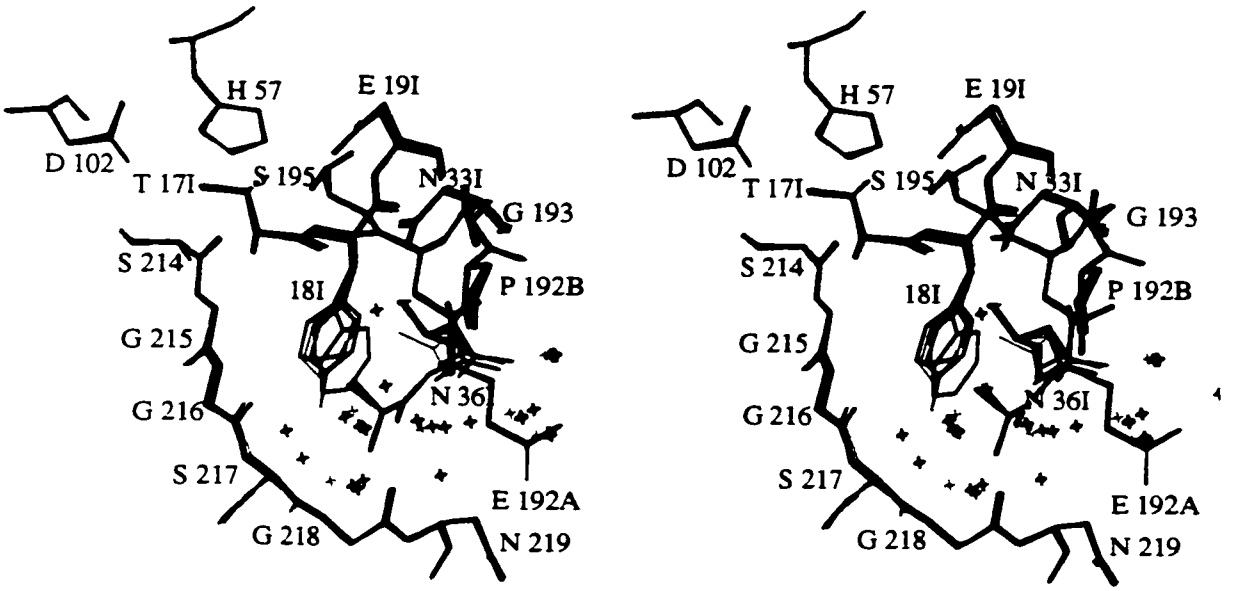

Figure 4.7 Superposition of SGPB in complex with the aromatic P1 variants of OMTKY3, Trp^{18I} (green), His^{18I} (blue), Phe^{18I} (yellow) and Tyr^{18I} (red). Water molecules have been depicted as crosses.

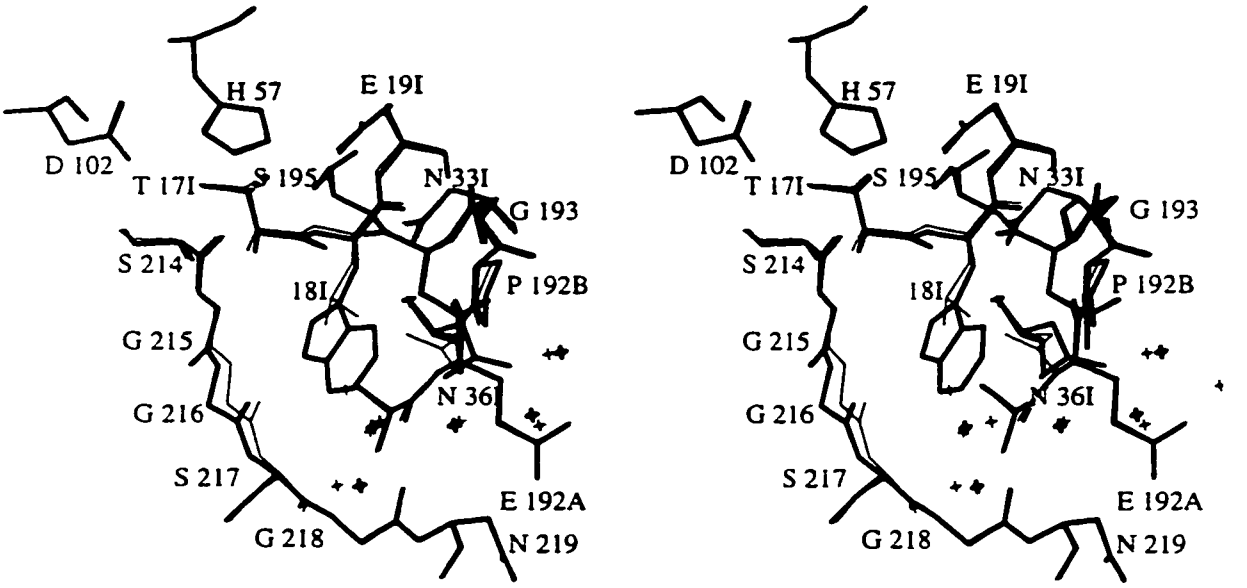

Figure 4.8 Superposition of SGPB:OMTKY3-Trp^{18I} (thick lines) onto SGPB:OMTKY3-Leu^{18I} (thin lines) in the region of the active site and S1 specificity pocket. Water molecules have been drawn as crosses.

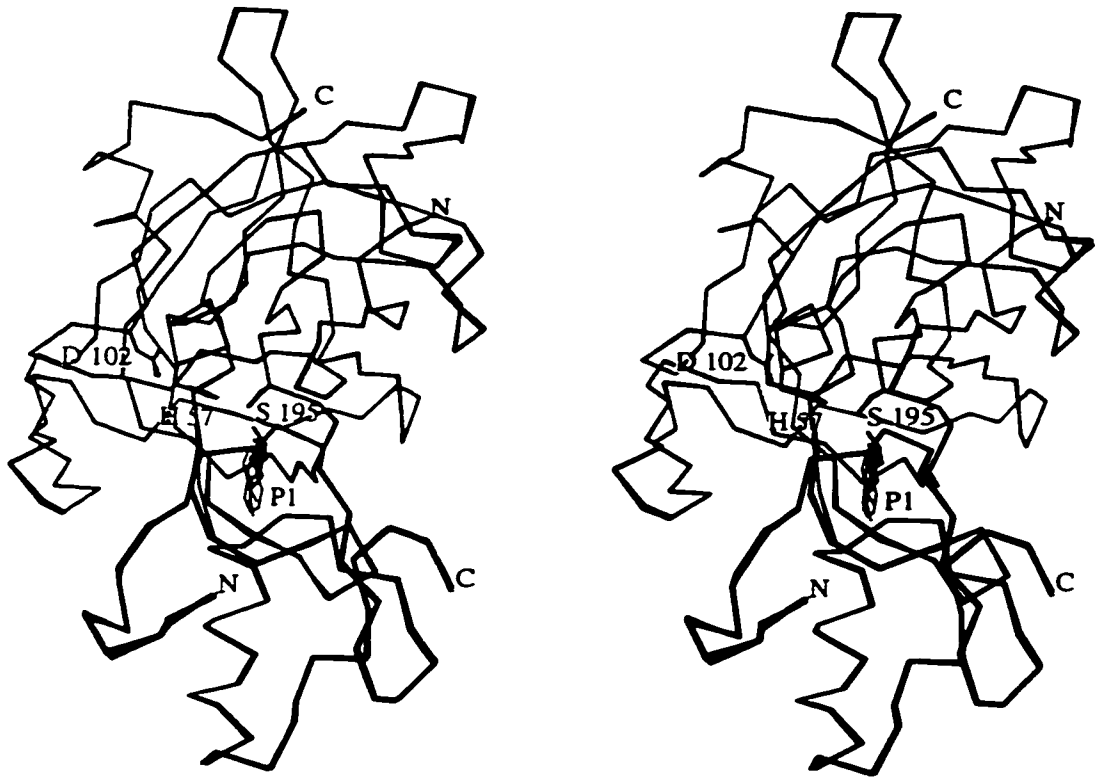

Figure 4.9 Cα trace of 18 OMTKY3 P1 variants in complex with SGPB (Leu^{18I}, Gly^{18I}, Ala^{18I}, Asp^{18I}, Glu^{18I}, Asn^{18I}, Gln^{18I}, Phe^{18I}, Tyr^{18I}, Ile^{18I}, Val^{18I}, Thr^{18I}, Ser^{18I}, Arg^{18I}, Lys^{18I}, Trp^{18I}, His^{18I}, COO-Leu^{18I}). The superimposition was calculated over the backbone atoms of the SGPB molecules. As a result, the SGPB molecules appear to be drawn in thin lines and the OMTKY3 molecules in thick lines but this is an illusion and simply indicates a higher deviation among the OMTKY3 molecules after superimposition of the enzymes. Side-chain atoms for the catalytic residues, His57, Asp102, and Ser195, and for each of the P1 residues have been included. The N- and C-terminus of both enzyme and inhibitor have also been indicated.

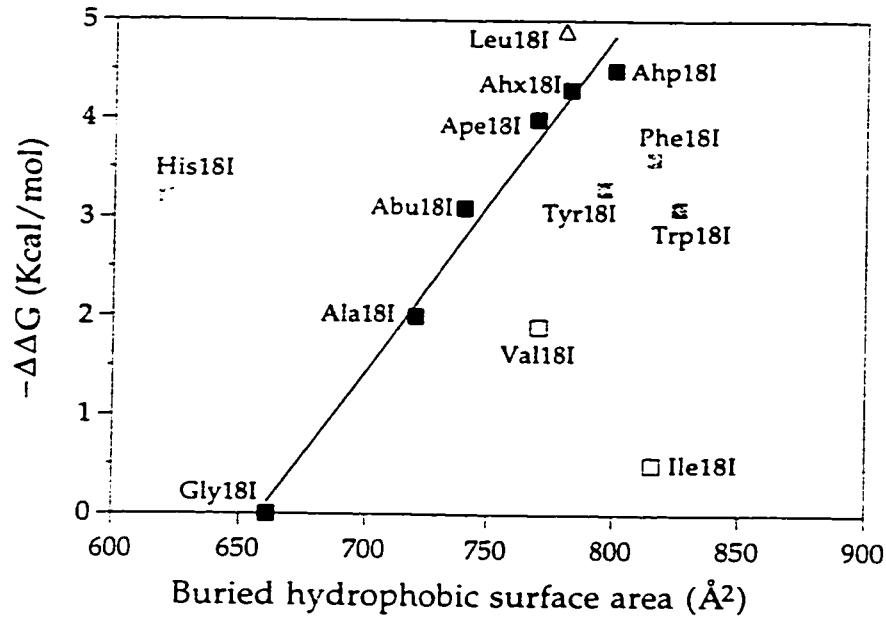

Figure 4.10 Buried hydrophobic surface area versus -ΔΔG. Data for the complexes with straight-chain, aliphatic P1 residues are indicated with black squares. Aromatic P1 variant complexes are indicated with grey squares, β-branched P1 variant complexes are indicated with white squares and SGPB:OMTKY3-Leu^{18I} is indicated with a triangle. ΔΔGs were calculated from the association equilibrium constants relative to that of SGPB:OMTKY3-Gly^{18I} (Huang *et al.*, 1995). The straight line was least squares fit to the data points indicated by black squares. Models of SGPB:OMTKY3 complexes with P1 residues Abu^{18I}, Ape^{18I}, Ahx^{18I} and Ahp^{18I} were constructed using the coordinates from SGPB:OMTKY3-Ala^{18I} (Huang *et al.*, 1995). Surface areas were calculated with ms (Connelly, 1983) and buried hydrophobic surface areas (BHSA) were calculated with the following equation.

BHSA = (HSA of SGPB) + (HSA of OMTKY3-X^{18I})
- (HSA of SGPB:OMTKY3-X^{18I})

This figure was adapted from Huang *et al.*, 1995.

Chapter 5:
Accommodation of positively-charged residues in a hydrophobic specificity pocket: crystal structures of SGPB in complex with OMTKY3 variants Lys^{18I} and Arg^{18I}

5.1 Introduction

The equilibrium binding constants (Ka) have been measured for six serine proteinases in complex with P1 variants of OMTKY3 (Table 1.2) (Bigler *et al.*, 1993; Lu *et al.*, 1997). SGPB, a serine proteinase from bacteria, shows a preference for medium-sized, hydrophobic amino acids at the P1 position. Positively-charged, negatively-charged and β-branched residues at this position are tolerated, but the Ka's for these variant complexes are much lower than those of the other variants. Additional experiments have been performed in order to calculate the pKa's for some of the ionizable P1 residues in complex with SGPB and with chymotrypsin (Qasim *et al.*, 1995; Qasim *et al.*, 1999) K_a's were also calculated for both the protonated and unprotonated P1 side chains of ionizable residues. Thorough analysis and proper interpretation of these data require structures of the different complexes. Structures of SGPB:OMTKY3-Glu^{18I} and SGPB:OMTKY3-Asp^{18I} have previously been determined at pH values above and below the pKa of the P1 side chains (Huang, 1995). In that study it was shown that a potassium ion was recruited into the S1 pocket of the negatively-charged variant complexes in order to balance the charge on the P1 side chain. A structure of OMTKY3-Lys^{18I} in complex with chymotrypsin has also been determined (Ding *et al.*, 1999) and in this chapter, X-ray crystal structures of the positively-charged OMTKY3 P1 variants, Arg^{18I} and Lys^{18I} complexed to SGPB will be presented. A structure of SGPB:OMTKY3-Lys^{18I} has additionally been determined at pH 10.7, ~ 2 pH units above the measured pKa for $Lys^{18I\circ}$ in the S1 pocket of SGPB. This variant with an uncharged lysyl side chain is refered to as Lys18I, the charged lysyl variant is Lys^{18I+}.

5.2 Materials and Methods

Crystallization. Dr. L. B. Smillie kindly provided SGPB, purified from Pronase according to the procedure of Jurasek *et al.*, 1979. OMTKY3 variants were generously provided by Dr. M. Laskowski, Jr.. The cloning, overexpression, and purification of the OMTKY3 P1 variants have been described (Lu, 1994; Lu *et al.*, 1997).

Crystals of the two SGPB:OMTKY3 P1 variant complexes were grown by the hanging drop method. Enzyme and inhibitor were mixed in a molar ratio of 1:1.5 and the protein concentrations of these solutions were approximately 30mg/ml. 1.5µl aliquots of protein solution were mixed with 3.5µl of a well solution that contained from 6-10% PEG4000 and 50mM $NaKHPO_4$ buffer, with a pH from 7.1 to 7.4. The final protein concentrations in the drops were 9mg/ml. Crystals that appeared from these conditions were used for streak seeding (Stura and Wilson, 1992). The seeding well conditions and drop preparation were similar to those described above except that PEG4000 concentrations were much lower, typically 1 to 4%.

Crystals of SGPB:OMTKY3-$Lys^{18I\circ}$ with a neutral P1 residue were prepared as just described. Those crystals suitable for data collection were quickly transferred to solutions containing 4% PEG4000, and 100mM $NaKHPO_4$ at pH 10.7. The crystals were soaked for 24 hours before being mounted in a capillary and exposed to X-rays.

Data Collection. Crystals of the molecular complexes were mounted in glass capillaries and X-ray diffraction data were collected either on a San Diego multiwire proportional counter (Hamlin, 1985; Xuong *et al.*, 1985) (SGPB:OMTKY3-Lys^{18I+}) or on a DIP 2030H image plate detector (Mac Science Co., Ltd.) using double mirror focusing optics (SGPB:OMTKY3-Arg^{18I}, SGPB:OMTKY3-$Lys^{18I\circ}$). X-rays were generated with a Rigaku RU200 rotating anode generator either operating at 40kV and 150mA for the SGPB:OMTKY3-Lys^{18I+} data collection or at 45kV and 75mA for the SGPB:OMTKY3-Arg18I and SGPB:OMTKY3-$Lys^{18I\circ}$ data sets. Each data

collection lasted approximately three days and was carried out at room temperature. SDMS data were scaled and merged with the SDMS software (Howard *et al.*, 1985). Denzo and Scalepack (Otwinowski and Minor, 1996) were used for reduction and scaling of the DIP data.

Structure Determination. The initial electron density maps of the three complexes were calculated with phases from the structure of SGPB in complex with the native OMTKY3 (Read *et al.*, 1983, Fujinaga *et al.*, 1982). Water molecules were removed from the model but leucine was retained as the P1 residue. The initial $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ maps were calculated with CCP4, SIGMAA (CCP4., 1994) and displayed with programs O (Jones *et al.*, 1991) (SGPB:OMTKY3-Lys^{18I+}) and Xtalview (McRee, 1993) (SGPB:OMTKY3-$Lys^{18I\circ}$; SGPB:OMTKY3-Arg^{18I}). The P1 residue was mutated to the appropriate amino acid and adjusted to fit the density. The N-terminal residue of the OMTKY3 variants, Val^{6I}, was also added to the inhibitor coordinates, as this residue was not present in the structure determination of native OMTKY3 in complex with SGPB (Read *et al.*, 1982; Fujinaga *et al.*, 1983). The electron density for Val^{6I} was clear in the initial $2|F_o|-|F_c|$ maps.

Refinement. The starting models, with the correct P1 residue, were subjected to simulated annealing and energy minimization in X-PLOR (Brünger, 1992b). Electron density maps were calculated from the resulting models and inspected with O and Xtalview. Water molecules from the SGPB:OMTKY3-Leu^{18I} structure were superimposed onto the electron density maps and those that fit peaks of positive density on $|F_o|-|F_c|$ maps of height at least 2.5σ and $2|F_o|-|F_c|$ maps of height at least 1σ were included in the model. All further refinement was carried out with TNT (Tronrud, 1992). Additional water molecules were assigned throughout the refinement process. As before, electron density at 2.5σ for $|F_o|-|F_c|$ maps and 1σ for $2|F_o|-|F_c|$ maps were required before water molecules were included in the parameter sets. Water molecule environments were also examined to ensure that meaningful hydrogen bonding partners were present. Side-chain adjustments were made to fit the observed electron density and the models were subjected to

stereochemical analysis by Procheck (Laskowski *et al.*, 1993) and Whatcheck (Rodriguez *et al.*, 1998) at various stages throughout the refinement. Special attention was paid to water molecules with B-values higher than 55Å^2 and those waters with B-values above 70Å^2 were deleted.

Structure comparison and figures. Superimpositions of enzymes, inhibitors and complex structures were performed with Xtalview (McRee, 1993). Superimpositions of chymotrypsin, trypsin and SGPA with SGPB were performed with the lsq commands in O (Jones *et al.*, 1991). Figures were made with many programs; Xtalview, Raster3d (Bacon and Anderson, 1988) and Grasp (Nicholls *et al.*, 1991).

5.3 Results

Crystals of SGPB in complex with OMTKY3 variants Arg^{18I} and Lys^{18I} were grown in space group P2$_1$, having the same unit cell dimensions as SGPB:OMTKY3 (Read *et al.*, 1983; Fujinaga *et al.*, 1982). The unit cell parameters and additional data collection statistics can be found in Table 5.1. Initial electron density maps, calculated with phases from the structure of wild type OMTKY3 in complex with SGPB (Read *et al.*, 1983; Fujinaga *et al.*, 1982), are in Figures 5.1, 5.2, and 5.3. The actual P1 side chain was obvious on these initial maps.

The SGPB:OMTKY3-Lys^{18I+} structure was in the final stages of refinement when the structure of SGPB:OMTKY3-Lys$^{18I\circ}$ was being determined. The deprotonation of Lys$^{18I\circ}$ was thought to be successful from the initial map (Figure 5.3) by comparison to the model of SGPB:OMTKY3-Lys^{18I+}. Waters 15 and 202 were not visible in the density and Nζ appeared to be in a different position. Further refinement confirmed these differences and they were attributed to the loss of a proton on Nζ of Lys$^{18I\circ}$.

Final electron density maps and models are given in Figures 5.4, 5.5, and 5.6. The R-factors for these models are 0.167 (SGPB:OMTKY3-Arg^{18I}),

0.148 (SGPB:OMTKY3-Lys^{18I+}), and 0.170 (SGPB:OMTKY3-Lys$^{18I\circ}$). Geometry statistics and error analysis can be found in Table 5.2. 89.4% of the residues from SGPB:OMTKY3-Arg^{18I}, 87.8% of the residues from SGPB:OMTKY3-Lys^{18I+} and 89.9% of the residues from SGPB:OMTKY3-Lys$^{18I\circ}$ are found in the most favoured regions of the Ramachandran plot. Only one residue, Asn100 from SGPB, is in the disallowed region for all of the SGPB:OMTKY3 complex structures. Asn100 is located in a sharp turn following a cis-proline, Pro99A, and the electron density for this residue has been consistantly well defined. None of the solvent atoms near the positively-charged P1 residues displayed the characteristics of negatively-charged ions with respect to electron density peak height or coordination. Therefore, they were treated as water molecules.

Figure 5.7 shows the S1 pocket of the two SGPB:OMTKY3-Lys^{18I} complex structures after superimposition of SGPB in the two complexes. When the P1 lysine in the complex is deprotonated by raising the pH above its pKa, two water molecules are released and Nζ moves out of the S1 pocket relative to the charged lysine at P1.

In all of the other SGPB:OMTKY3 complex structures, Asn^{36I} has been modeled with a χ_2 of roughly -120° (Table 4.4, Chapter 4), such that Nδ2 donates a hydrogen bond to Asn^{33I} Oδ1 (Figure 5.7). However, in the structure of SGPB:OMTKY3-Lys$^{18I\circ}$, Asn^{36I} was modeled with a χ_2 of 73° so that Oδ1 could accept a hydrogen bond from Lys$^{18I\circ}$ Nζ (Figure 5.7). Unfortunately, it was not possible to determine the actual orientation of Asn^{36I} from the atomic peak heights in the electron density map.

Waters 15 and 202, present at pH 7.2, both accept hydrogen bonds from Lys^{18I+} Nζ. Lys^{18I+} Nζ donates a third hydrogen bond to the carbonyl O of Ala^{192}. Water 202 also accepts a hydrogen bond from Thr226 Oγ1 and donates hydrogen bonds to Thr225 O and to water 42.

SGPA has higher association equilibrium constants than SGPB in complex with the OMTKY3 variants, except for those that are positively-charged (Lu *et al.*, 1997). In order to speculate why these enzymes display

Table 5.1 Data Collection Statistics

	Arg^{18Ia}	Lys^{18I+}	$Lys^{18I\circ}$
Space Group	$P2_1$	$P2_1$	$P2_1$
Unit Cell a(Å)	45.31	45.50	45.45
b(Å)	54.80	54.85	54.69
c(Å)	45.49	45.60	45.63
β(°)	119.02	119.06	119.33
Maximum resolution (Å)	1.65	1.80	1.75
Total no. of observations	78238	47037	68100
No. of unique reflections	22171	17246	19056
Average redundancy	3.5	2.7	3.6
Rmerge[b]:			
overall	0.118	0.056	0.173
highest resolution shell	0.288	0.187	0.359
Resolution range of last shell (Å)	1.68-1.65	1.94-1.80	1.78-1.75
$\langle I/\sigma(I)\rangle$:			
overall	8.09	18.27	7.44
highest resolution shell	2.86	2.26	1.72
Resolution range of last shell (Å)	1.81-1.79	1.82-1.80	1.76-1.75
Completeness of Data:			
overall (%)	94.6	94.8	95.9
highest resolution shell (%)	55.1	40.8	70.4
Resolution range of last shell (Å)	1.68-1.65	1.83-1.80	1.78-1.75

[a]The column headings refer to the SGPB:OMTKY3 complex with that amino acid at the P1 position.

[b]$Rmerge=\Sigma_{hkl}\,[(\Sigma_i\ |I_i-\langle I\rangle|)/\Sigma_i\ I_i]$

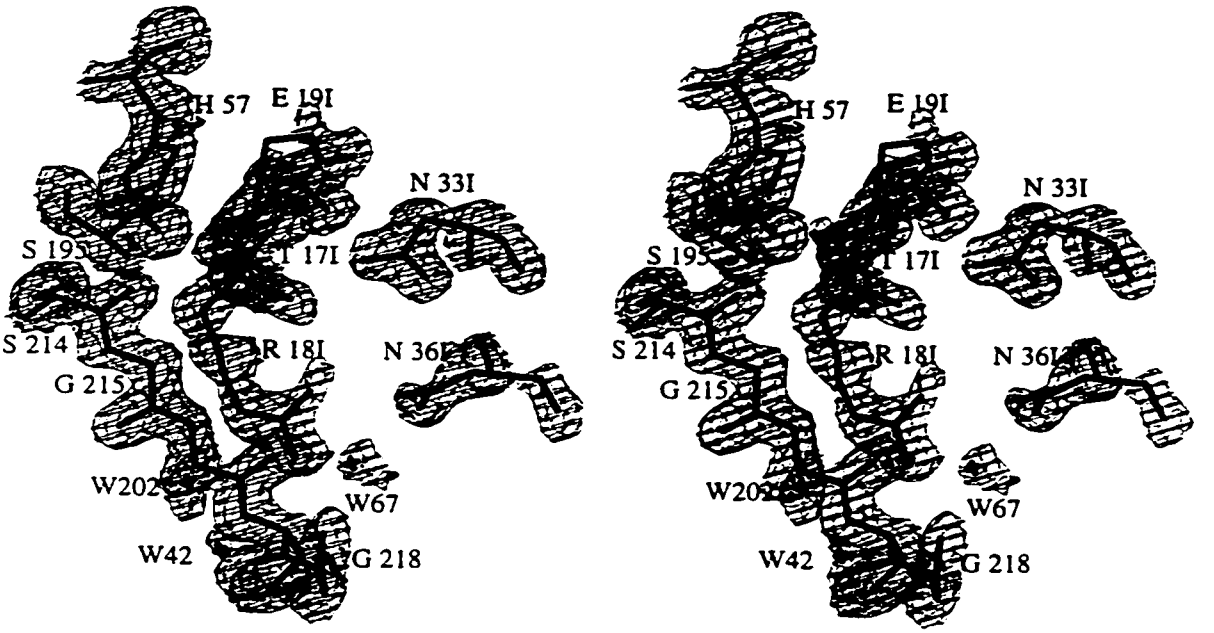

Figure 5.1 Initial electron density map of SGPB:OMTKY3-Arg^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

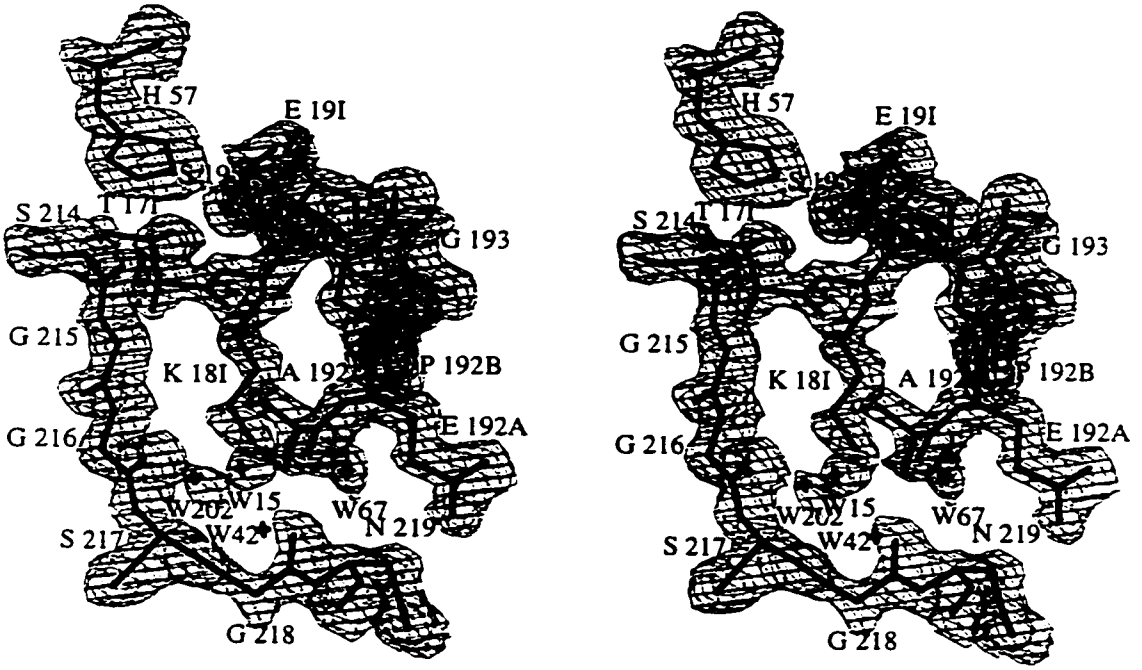

Figure 5.2 Initial electron density map of SGPB:OMTKY3-Lys^{18I+} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

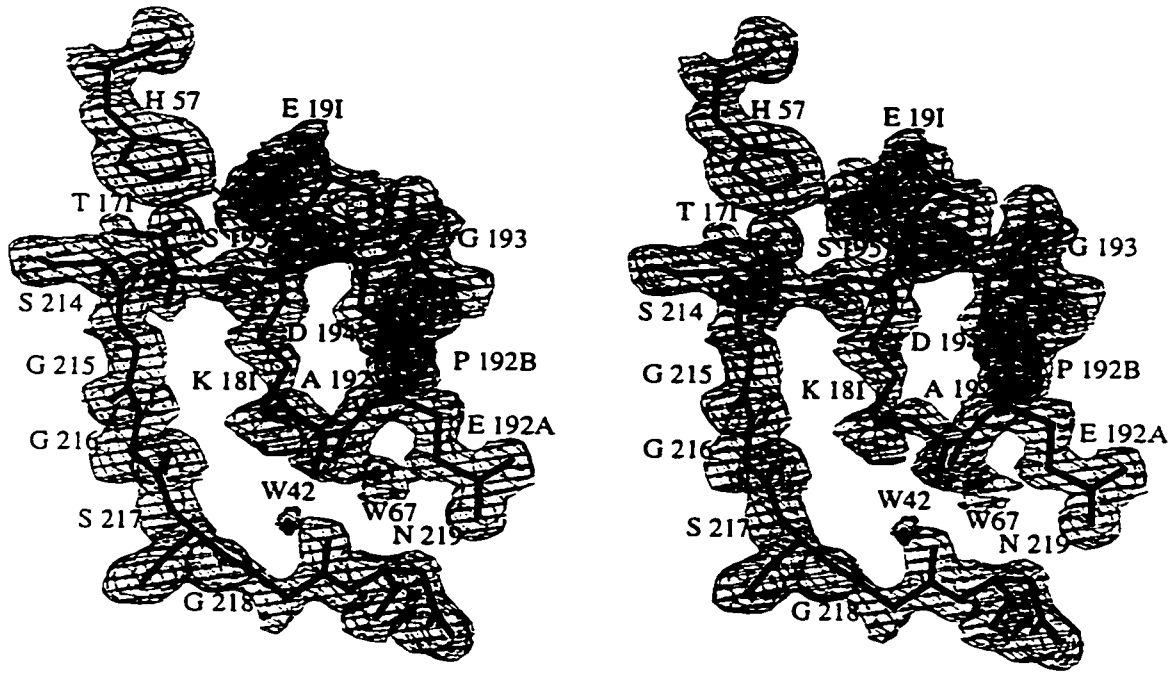

Figure 5.3 Initial electron density map of SGPB:OMTKY3-Lys^{181o} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

Table 5.2 Refinement Statistics

	Arg^{18Ia}	Lys^{18I}	Lyn^{18I}
No. of reflections used	22171	17246	19056
Resolution Range (Å)	20-1.65	20-1.80	20-1.75
$Rcryst^b$	0.167	0.148	0.170
No. of protein atoms[c]	1723	1700	1718
No. of solvent atoms	173	151	174
rms deviation from ideal stereochemistry			
bond distance (Å)	0.012	0.011	0.011
bond angle (°)	1.070	1.075	1.016
planar groups (Å)	0.014	0.014	0.014
Average B-values ($Å^2$)			
main-chain atoms	12	14	12
side-chain atoms	16	18	16
solvent atoms	34	36	34
Error Estimates (Å)			
Luzzati (Luzzati, 1952)	0.18	0.16	0.17
SIGMAA (Read, 1986)	0.21	0.12	0.22

[a]The column headings refer to the SGPB:OMTKY3 complex with that amino acid at the P1 position.

[b]$Rcryst=\Sigma_{hkl} \, | \, |F_o| - |F_c| \, | \, / \Sigma_{hkl} \, |F_o|$
R values were calculated with all of the data in the resolution range indicated and without a $\sigma(I)$ cutoff.

[c]The number of protein atoms includes those atoms from alternate conformations.

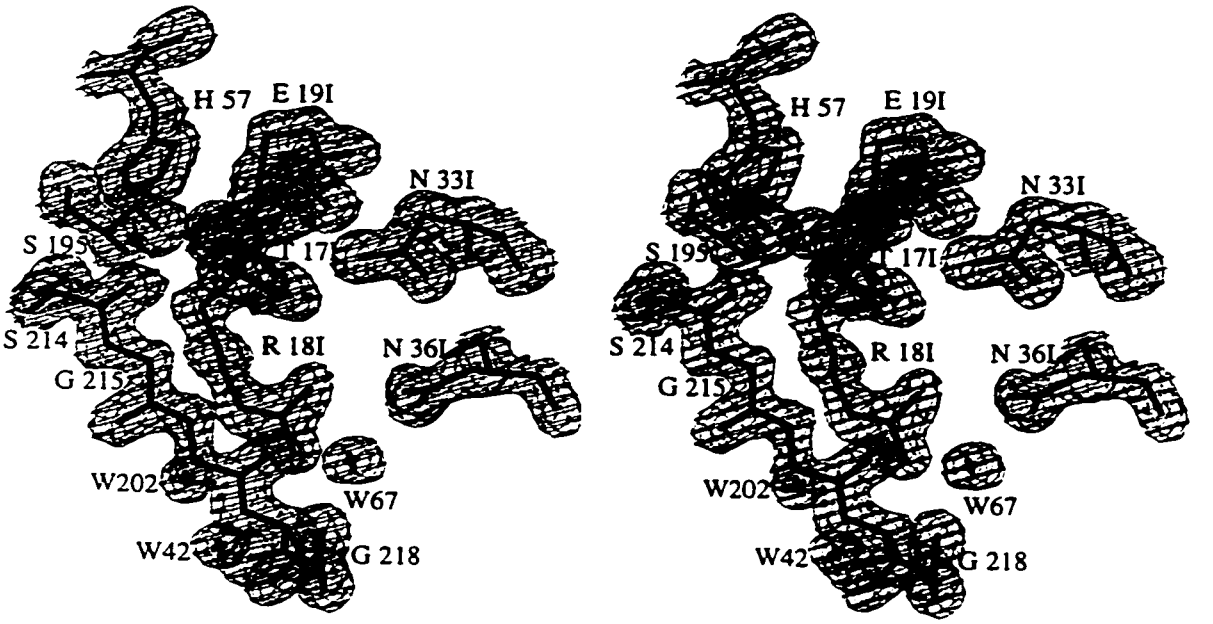

Figure 5.4 Final electron density map of SGPB:OMTKY3-Arg^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

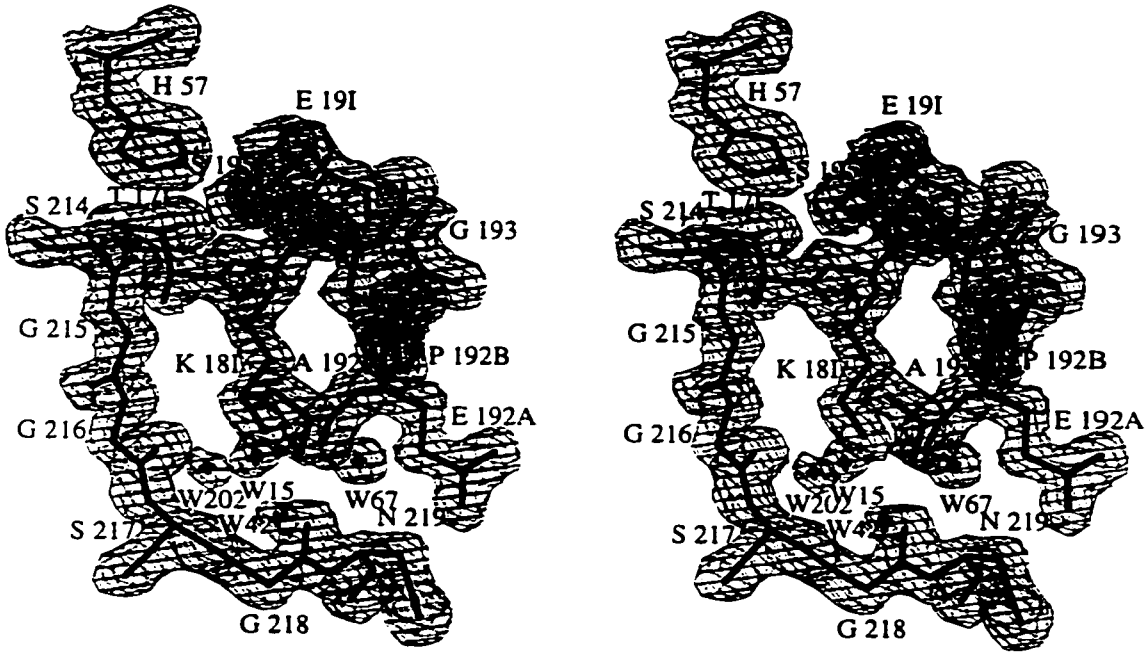

Figure 5.5 Final electron density map of SGPB:OMTKY3-Lys^{18I+} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

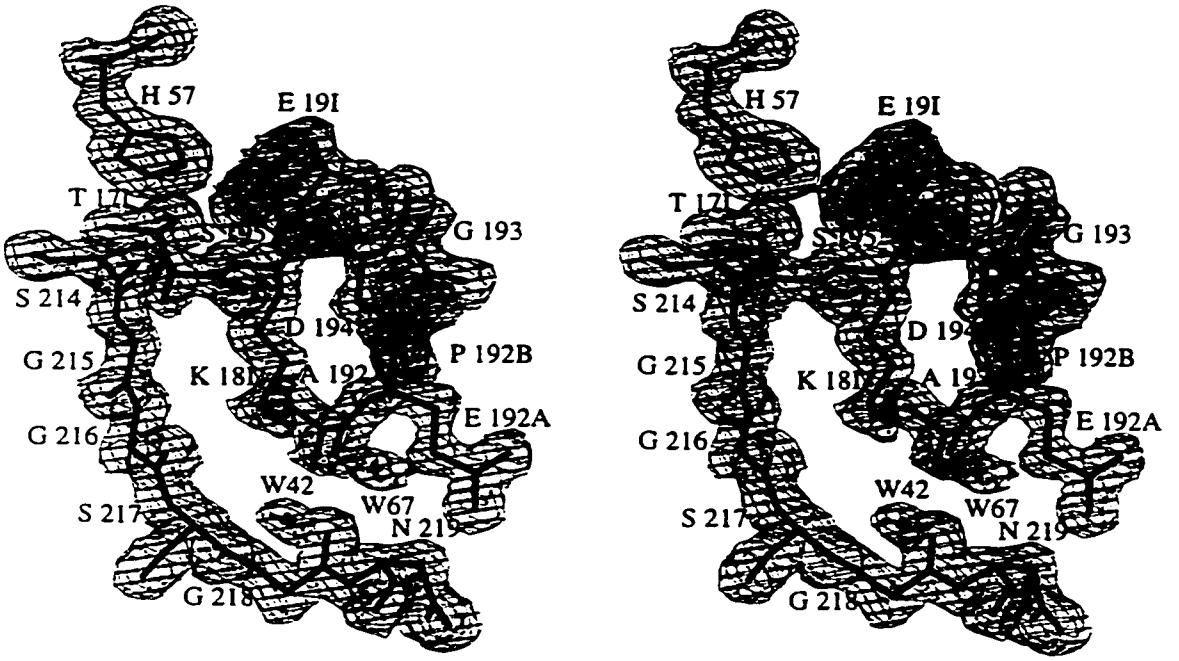

Figure 5.6 Final electron density map of SGPB:OMTKY3-Lys$^{18I°}$ in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

different patterns in their Kas, a comparison has been made between the structures of SGPB:OMTKY3-Lys^{18I+} and SGPA in complex with a tetrapeptide (James *et al.*, 1980). Figure 5.8 highlights a potential difference between the S1 pockets of the two enzymes. Most of the residues lining the S1 pocket are identical in the two enzymes but Thr226 has a different conformation in each enzyme. In SGPA the side chain Oγ1 makes a hydrogen bond with Tyr228 OH (phenylalanine 228 in SGPB). On the other hand, Thr226 Oγ1 in SGPB:OMTKY3-Lys^{18I+} is free to coordinate water 202. This interaction would help to stabilize the solvent molecule in the hydrophobic S1 pocket of SGPB when it is present. Water 202 additionally accepts a hydrogen bond from Lys^{18I+} Nζ. This water molecule is also present in the structure of SGPB:OMTKY3-Arg^{18I} and similarly accepts a hydrogen bond from Thr226 and the P1 arginine (Figure 5.4).

A comparison has also been made between OMTKY3-Lys^{18I+} in complex with SGPB and with chymotrypsin (Ding *et al.*, in preparation) (Figure 5.9). The S1 pockets have vastly different shapes in the two serine proteinases. For example, the chymotrypsin pocket is deeper. Amino acid 226, at the bottom of the pocket is a glycine rather than a threonine as in SGPB. In addition, the polypeptide chains of chymotrypsin and SGPB follow different paths at residue Ser217. As a result, the chymotrypsin S1 pocket appears "cave-like" whereas the SGPB S1 pocket is shaped more like a cleft than a cave (Figures 5.10 and 5.11).

The trypsin S1 pocket also appears deeper and more "cave-like" than the S1 pocket of SGPB (Huang *et al.*, 1993) (Figures 5.10, 5.12, 5.23 and 5.14). Residue 226 is a glycine, as in chymotrypsin but the most well known difference is the presence of Asp189 at the bottom of the S1 pocket. This acidic residue confers specificity for positively-charged residues at the P1 position. Residue Gln192 and the path of the trypsin segment from Ser214 to Cys220 create the more cave like appearance of the trypsin S1 pocket in comparison to SGPB.

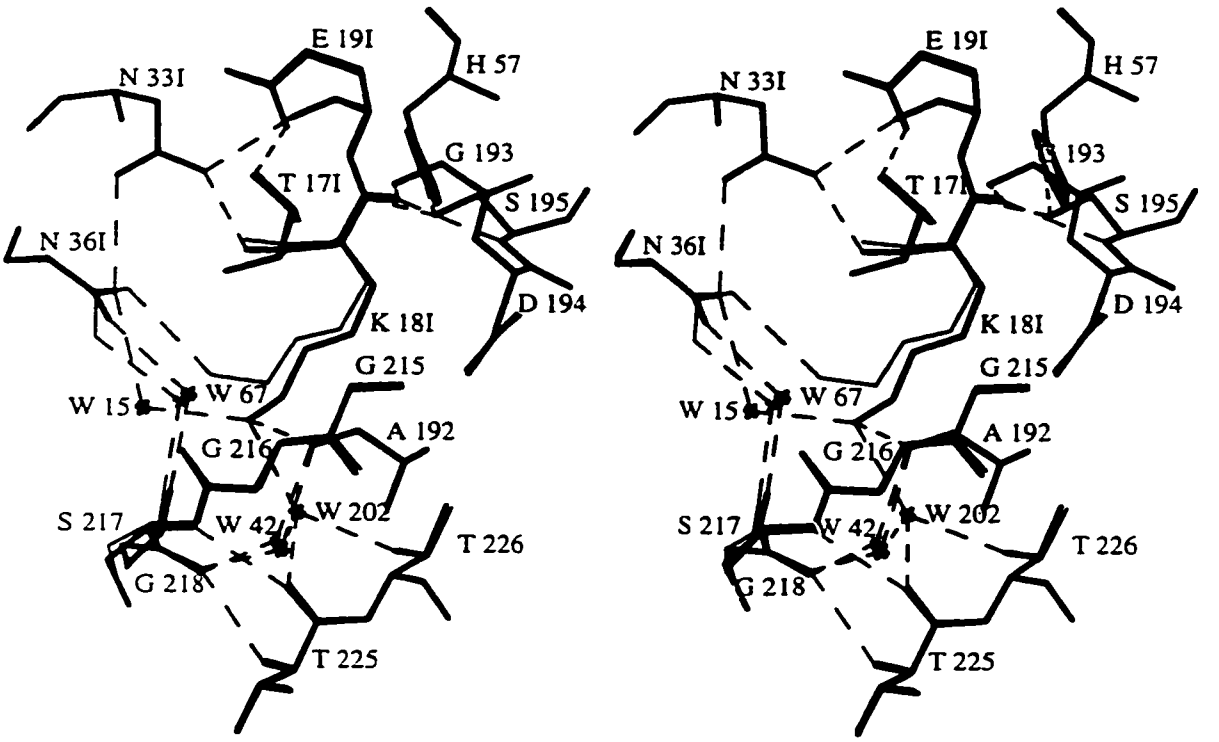

Figure 5.7 Superposition of SGPB in complex with OMTKY3-Lys^{18I+} (thick lines) and OMTKY3-Lys$^{18I\circ}$ (thin lines) in the region of the P1 lysine. Water molecules have been drawn as crosses. Hydrogen bonds are indicated by dashed lines.

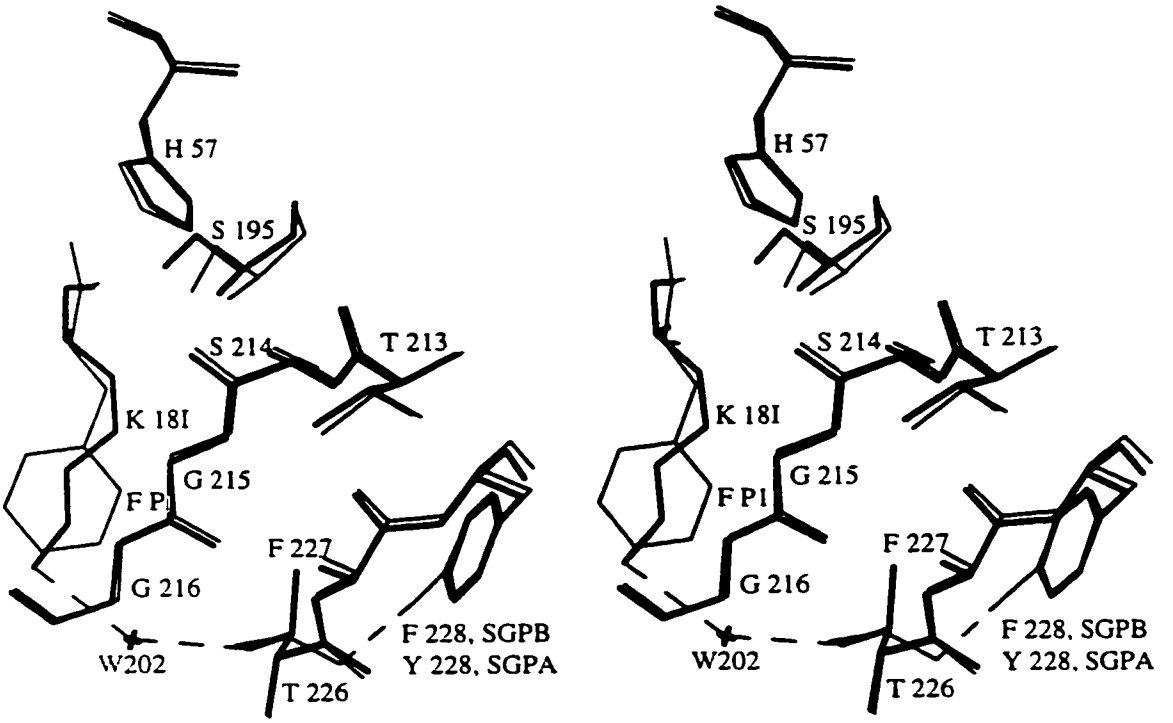

Figure 5.8 Superimpostition of SGPB:OMTKY3-Lys^{18I+} (thick lines) onto SGPA in complex with a tetrapeptide (thin lines), in the region of the active site and S1 specificity pocket. The positions of water molecules are represented by crosses. Hydrogen bonds highlighting the interactions of residue Thr226 from both SGPB and SGPA have been indicated by dashed lines.

Rms deviation values for the SGPB:OMTKY3 complexes are in Table 5.3. All of these values are low. Rms deviation values for the SGPB complexes superimposed with SGPA, chymotrypsin and trypsin complexes are in Table 5.4. The superimposition of SGPB onto SGPA has the lowest rms deviation for this group of calculations.

5.4 Discussion

The association equilibrium constant for SGPB with OMTKY3-Lys^{18I} is 2.6×10^8 (M^{-1}) when measured at pH 8.3 (Table 1.2 and Figure 1.2). However, this pH is very close to the pK_a for Lys^{18I} (8.72), while the inhibitor is in complex with SGPB (Dr. M. Laskowski, Jr., personal communication). Lysine residues with lowered pKas have been found in other proteins such as acetoacetate decarboxylase (6.0) (Kokesh and Westheimer, 1971), ribonuclease A (9.0) (Jentoft *et al.*, 1979) and apolipoprotein A-I (8.3) (Sparks *et al.*, 1992). Dr. Laskowski, Jr.'s lab calculated equilibrium association constants (from experimental data) for SGPB:OMTKY3-Lys^{18I} where both the free and the complexed inhibitor have a neutral P1 residue ($K_a=4.2 \times 10^9$ (M^{-1})) and where both the free and the complexed inhibitor have a positively charged lysine residue at P1 ($K_a=1.8 \times 10^8$ (M^{-1})). The S1 pocket of SGPB is hydrophobic and prefers medium-sized aliphatic residues, such as leucine and methionine. Therefore, the lysine side chain becomes more complementary to the pocket when the positive charge is removed. SGPB's preference for uncharged side chains at P1 is also reflected in the relatively high association constant for SGPB in complex with OMTKY3-Ahp^{18I} ($K_a=2.8 \times 10^{10}$ (M^{-1})). (Ahp is a non-coded amino acid. The side chain is isosteric to lysine with a methyl group replacing $N\zeta$.) Structures of the SGPB:OMTKY3-Lys^{18I} complex have been determined at two pH values, above and below the pK_a for Lys^{18I} in the S1 pocket, in order to rationalize the differences in K_a more specifically.

It was initially surprising that the neutral side chain of P1 $Lys^{18I\circ}$ sat further out of the S1 pocket than the charged Lys^{18I+} at P1 (Figure 5.7). However, the position of $N\zeta$ in relation to the S1 pocket is not an

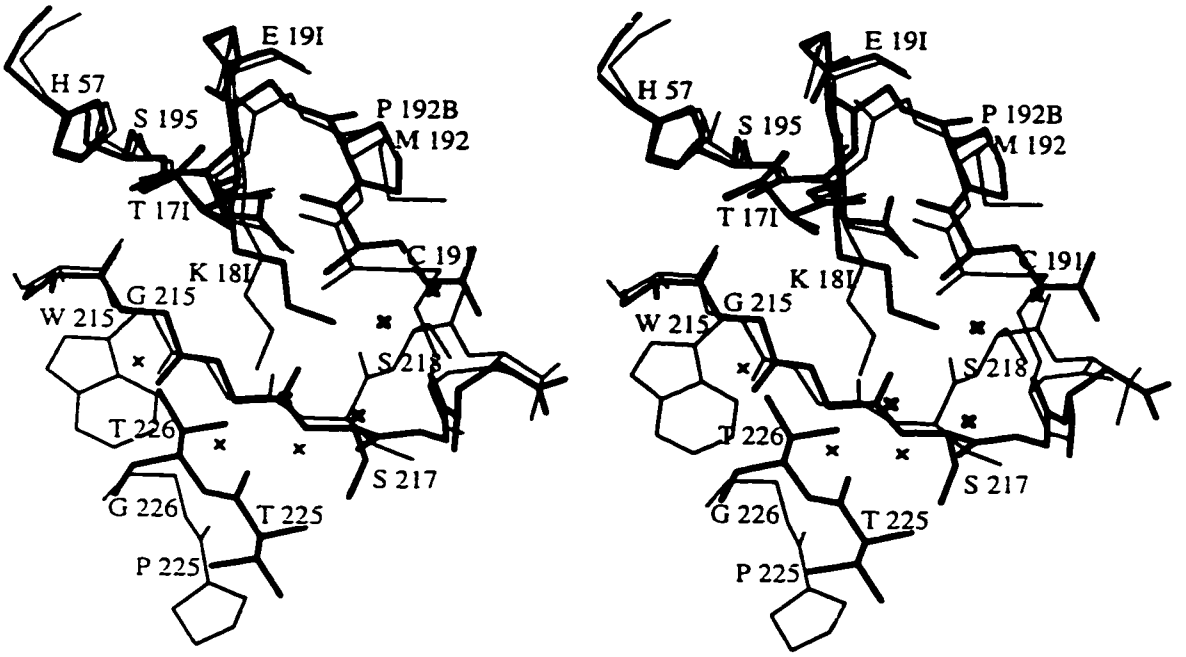

Figure 5.9 Superimposition of SGPB:OMTKY3-Lys^{18I+} (thick lines) onto CHYM:OMTKY3-Lys^{18I} (thin lines) in the region of the active site and S1 specificity pocket.

Figure 5.10 Molecular surface of SGPB, white, in complex with OMTKY3-Lys^{18I+} in the region of the S1 specifity pocket. Carbon atoms of OMTKY3-Lys^{18I+} have been coloured green, oxygen atoms red, nitrogen atoms blue and sulfur atoms yellow. The location of several SGPB residues are indicated with labels. The molecular surfaces for this figure and for figures 5.11 and 5.14 were calculated with GRASP (Nicholls *et al.*, 1991).

Figure 5.11 Molecular surface of chymotrypsin, white, in complex with OMTKY3-Lys^{18I} in the region of the S1 specifity pocket. Carbon atoms of OMTKY3-Lys^{18I} have been coloured green, oxygen atoms red, nitrogen atoms blue and sulfur atoms yellow. The location of several chymotrypsin residues are indicated with labels.

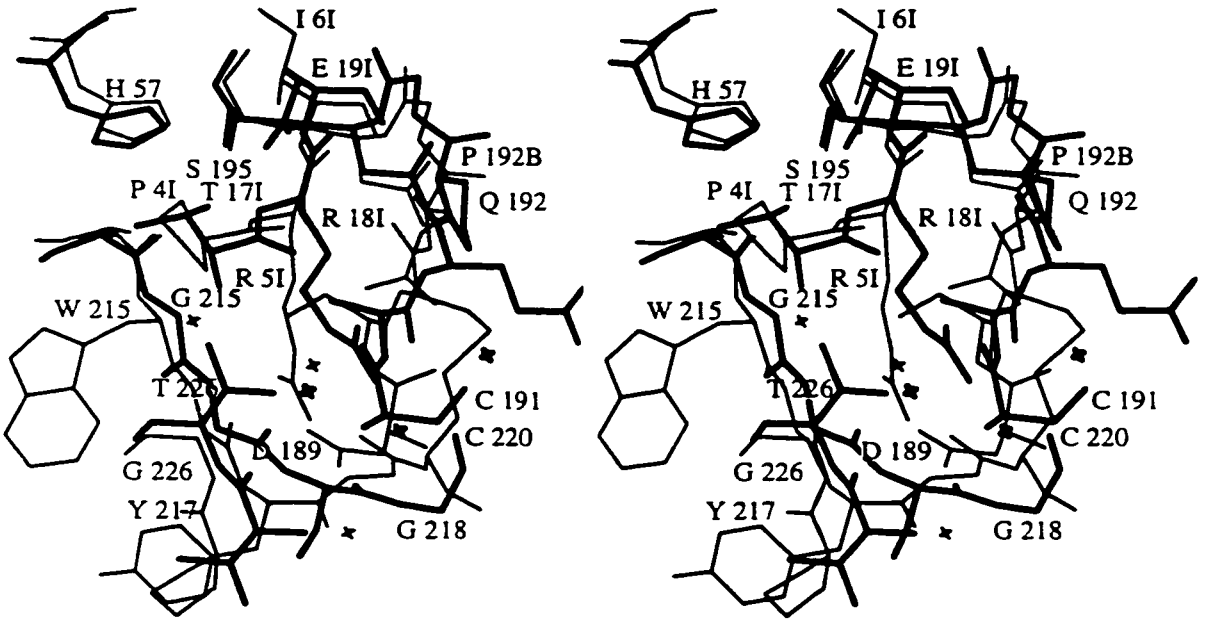

Figure 5.12 Superimposition of SGPB:OMTKY3-Arg^{18I} (thick lines) onto Trypsin in complex with the inhibitor from bitter gourd (thin lines) in the region of the S1 pocket. Water molecules are represented by crosses.

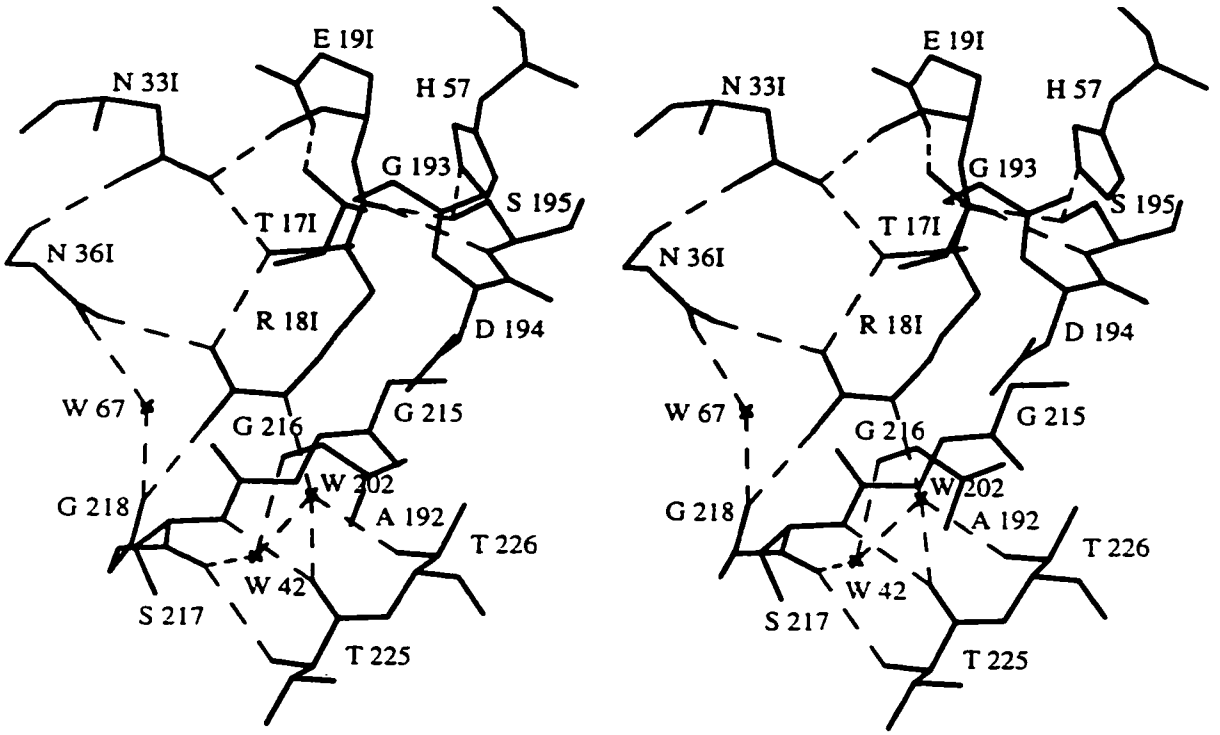

Figure 5.13 Hydrogen bonding interactions in the S1 pocket of SGPB:OMTKY3-Arg^{18I}. Water molecules are represented by crosses. Hydrogen bonds have been drawn as dashed lines.

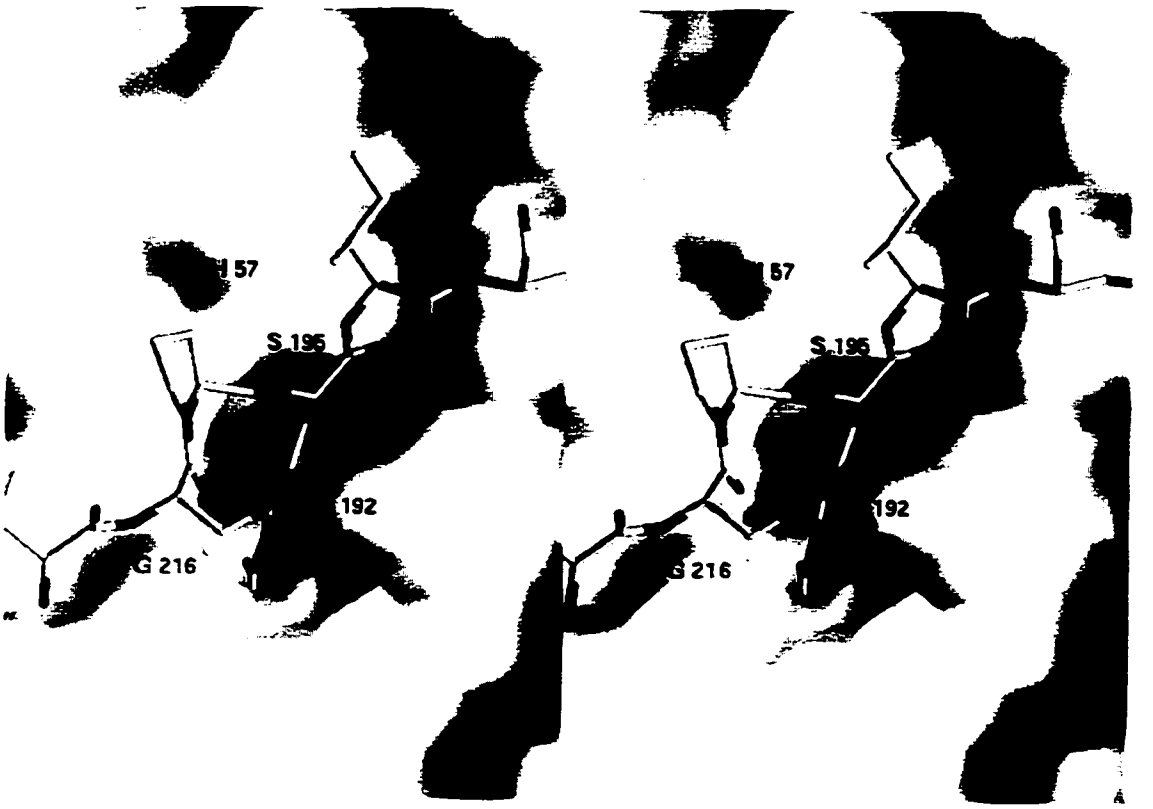

Figure 5.14 Molecular surface of trypsin, white, in complex with an inhibitor from bitter gourd in the region of the S1 specifity pocket. Carbon atoms of the inhibitor have been coloured green, oxygen atoms red, nitrogen atoms blue and sulfur atoms yellow. The location of several trypsin residues are indicated with labels.

Table 5.3 Rms differences (Å) of main-chain atoms among the structures of the various SGPB:OMTKY3 complexes.

	$Lys^{18I\circ}$	Arg^{18I}
Lys^{18I+}	0.12[1]	0.09
	0.11[2]	0.10
	0.13[3]	0.11
$Lys^{18I\circ}$		0.09
		0.13
		0.11

[1]The top number in each column refers to calculations among the SGPB molecules of the SGPB:OMTKY3 complexes. 740 main-chain atoms (N, Cα, C, O) from 185 residues were used in each calculation.
[2]The middle number in each column refers to calculations among the OMTKY3 variant molecules of the SGPB:OMTKY3 complexes. 204 main-chain atoms from 51 residues were used in each calculation.
[3]The bottom number in each column refers to calculations among the SGPB:OMTKY3 complexes. 944 main-chain atoms (740 from SGPB and 204 from OMTKY3) from 236 residues were used in each calculation.

Table 5.4: Rms differences of Cα atoms for three sets of superimpositons.

Complex structures	rms deviation (Å)	number of Cα atoms from the enzymes used in the calculation
SGPA:tetrapeptide SGPB:OMTKY3-Lys^{18I+}	0.69	172
CHYM:OMTKY3-Lys^{18I+} SGPB:OMTKY3-Lys^{18I+}	1.83	122
Trypsin:bitter gourd inhibitor SGPB:OMTKY3-Arg^{18I}	1.78	122

indication of the degree of repulsion between side chain and pocket. On the contrary, a positively-charged lysine side-chain requires two additional water molecules compared to SGPB:OMTKY3-Lys$^{18I\circ}$, in order to dissipate and neutralize the charge in a hydrophobic environment (Figure 5.7). Therefore, the positively-charged Nζ needs to sit further in the pocket so that the waters are accommodated. In contrast, the neutral lysine side chain donates an intramolecular hydrogen bond to Asn^{36I} Oδ1. In this way, the S1 hydrophobic pocket is only presented with lysine's four methylene groups and has minimal interaction with the polar end of the side chain (Figure 5.7). The absence of ordered waters 15 and 202 in the structure of SGPB:OMTKY3-Lys$^{18I\circ}$, also means that this complex has lower entropy costs than SGPB:OMTKY3-Lys^{18I+}. The sum of these energy contributions results in tighter binding for SGPB:OMTKY3-Lys$^{18I\circ}$ over SGPB:OMTKY3-Lys^{18I+} by ~1.8kcal/mol.

In comparison to other SGPB:OMTKY3 structures, Asn^{36I} adopts a somewhat abnormal conformation in the structure of SGPB:OMTKY3-Lys$^{18I\circ}$ (Table 4.4, Chapter 4). It must first be pointed out, that the positions of Asn^{36I} atoms Nδ2 and Oδ1 were chosen based on hydrogen bonding analysis and not on the experimental X-ray diffraction data and it is possible that the side chain adopts the typical conformation found in most of the SGPB:OMTKY3 variant complexes. However, this conformation would leave the P1 lysine without a hydrogen bonding partner. Asn^{33I}, which typically accepts two hydrogen bonds from Asn^{36I}, one from the main chain NH and one from the side chain Nδ2, is an important residue in the stabilization of the reactive site loop. Asn^{33I} is important because Nδ2 donates two hydrogen bonds, one each to the carbonyl oxygen atoms of Thr^{17I} and Glu^{19I} on either side of the scissile peptide bond. Stabilization of the reactive site loop in this way disfavours cleavage and, therefore, OMTKY3 acts as an inhibitor rather than substrate. Asn^{33I} also donates the same two hydrogen bonds to the carbonyl oxygen atoms of Thr^{17I} and Glu^{19I} in the SGPB:OMTKY3-Lys$^{18I\circ}$ structure; as well, Asn^{33I} Oδ1 accepts one hydrogen bond from the main chain NH of Asn^{36I}. The hydrogen bond from the side chain Nδ2 of Asn^{36I} to Asn^{33I} Oδ1 in most of the other SGPB:OMTKY3 complexes is

probably not crucial for stabilization of the reactive site loop, especially since Asn^{36I} contributes towards stabilization of the lysine side chain in the S1 specificity pocket. Asn^{36I} also displays natural variation among ovomucoids from different sources (Laskowski, Jr., *et al.*, 1987). This natural variation further suggests that Asn^{36I} is not absolutely required for inhibitor stability.

pK_a's for the P1 residue from OMTKY3-Lys^{18I} have also been determined in complex with chymotrypsin (Qasim *et al.*, 1999). The pK_a is 7.35, lower than that of the ovomucoid P1 lysine when in complex with SGPB. In addition, deprotonation of the lysine residue has a more dramatic effect on the K_a for binding with chymotrypsin than for SGPB. The K_a for CHT:OMTKY3-Lys^{18I+}, with a positive P1 is 4.2×10^7 (M^{-1}) and with a neutral P1, the K_a is 4.6×10^9 (M^{-1}), a 380-fold increase. Only a 23 fold increase was observed for SGPB:OMTKY3-Lys^{18I}. These differences are not surprising since the P1 lysine is more buried in the chymotrypsin S1 pocket than the S1 pocket of SGPB (Ding *et al.*, in preparation)(Figure 5.9). However, structures of chymotrypsin in complex with BPTI, a Kunitz inhibitor, (Capasso *et al.*, 1997; Scheidig *et al.*, 1997) reveal an alternate conformation for a P1 lysine in the chymotrypsin S1 pocket. In these structures, the lysine does not extend deeply into the pocket but turns out, and in this way resembles more closely the position of Lys^{18I} of SGPB:OMTKY3-Lys^{18I}. The pK_a for BPTI-Lys^{15I} in complex with chymotrypsin is 8.8 and reflects that the side chain has a less hydrophobic environment than it would deep inside the pocket. Changes in pH also have a less dramatic effect on the K_a of Lys^{15I} in BPTI. For a positive lysine, the K_a equals 4×10^7 (M^{-1}) and for a neutral lysine, the K_a equals 8×10^8 (M^{-1}).

The association constant for the OMTKY3-Arg^{18I} variant with SGPB ($K_a = 1.9 \times 10^8$) is only slightly lower than that for OMTKY3-Lys^{18I} (2.6×10^8 (M^{-1})) (Table 1.2 and Figure 1.2). This difference was not surprising since arginine has a larger side chain and would bury a larger surface area on forming the complex. Asn^{36I} also adopts an unusual conformation so that $O\delta1$ can accept a hydrogen bond from Arg^{18I} $N\eta1$ (Figure 5.13, Table

4.4). Again, the Asn^{36I} side-chain orientation was modeled based on hydrogen bonding analysis and not on electron density.

Positively-charged potassium ions, identified by peak height and coordination, were observed in the structures of SGPB:OMTKY3-Glu^{18I} and SGPB:OMTKY3-Asp^{18I} which were determined above the pK_a for the carboxylate groups of the buried P1 side chains at pH 10.7 (Huang, 1995). The ions formed salt bridges with the negatively-charged P1 residues and the cost of burying a negative charge in the hydrophobic S1 pocket was estimated to be 5.6 kcal/mol. In contrast, there is no evidence of negatively-charged ions in the S1 pockets of SGPB:OMTKY3-Lys^{18I+} or of SGPB:OMTKY3-Arg^{18I}. Lysine and arginine residues have longer side chains than glutamic and aspartic acids and the charged nitrogen atoms on the ends of the residues are closer to the bulk solvent than are the carboxyl groups of Glu^{18I} and Asp^{18I}. In other words, the negative charges are in a slightly different environment than the positive charges and perhaps this influences the recruitment of ions to the complex at high pH.

SGPB vs. SGPA The OMTKY3 variants with positively-charged P1 residues are the only two variants for which SGPB has a higher K_a than SGPA (Lu *et al.*, 1997). The S1 binding pockets are very similar in the two enzymes (Figure 5.8). Glu192B in SGPB is a glutamine in SGPA but only the main chain atoms of this residue line the S1 pocket. Residue 226, although it is a threonine in both proteinases, has a different conformation in SGPB than it has in SGPA (James *et al.*, 1980). In SGPA, Thr226 has a χ_1 angle of 60°; Oγ1 makes a hydrogen bond with Tyr228 OH and Cγ2 sits at the back of the S1 pocket. Since residue 228 is a phenylalanine in SGPB, there is no possibility of a hydrogen bond between Phe228 and Thr226. Therefore in SGPB, Thr226 has a χ_1 angle of -55°. In this conformation, Oγ1 superimposes onto Thr226 Cγ2 from SGPA (Figure 5.8). In this conformation, it is able to hydrogen bond to water 202 when that water is present in a complex structure. Water 202 has been found in several SGPB:OMTKY3 complex structures thus far, including the two positively-charged P1 variants, SGPB:OMTKY3-Lys^{18I+}

and SGPB:OMTKY3-Arg^{18I}. In these two complexes, water 202 forms hydrogen bonds with the basic side chains and with Thr226 Oγ1 (Figures 5.4 and 5.5). Water 202 helps to stabilize the positive charge on the P1 lysine or arginine side chain and the Thr226 conformation observed in SGPA does not allow for a favourable environment for stabilization of water 202 and the positive charge. On the other hand, Thr226 Cγ2 of SGPA creates a more favourable environment for the hydrophobic P1 residues. This observation is in agreement with the hydrophobic P1 variants of OMTKY3 having higher K_as when in complex with SGPA rather than SGPB.

Thr226 also distinguishes SGPB from the serine proteinases trypsin and chymotrypsin. A glycine is found at this position in the other two enzymes resulting in a deeper S1 pocket. This deeper pocket allows long P1 residues like lysine and arginine to adopt a fully extended conformation within the S1 pockets of the pancreatic enzymes. Trypsin, which prefers positively-charged P1 side chains, has an aspartic acid (Asp189) at the bottom of the pocket, whereas the others do not. Conformations of P1 lysine and arginine in the SGPB:OMTKY3 complexes would not be appropriate for trypsin or chymotrypsin because the segments forming the front edge of the pockets (217-220) follow different paths than SGPB (Figures 5.9 and 5.12) (Ding *et al.*, 1999; Huang *et al.*, 1993). All of these differences result in the SGPB P1 pocket having more of a cleft shape than the deep pockets of the serine proteinases trypsin and chymotrypsin (Figures 5.10, 5.11, 5.13 and 5.14).

Chapter 6:
Contribution of peptide bonds towards binding: crystal structures of OMTKY3-COO-Leu^{18I} in complex with SGPB and OMTKY3-CH2-Asp^{19I}

6.1 Introduction

The interactions between serine proteinases and the canonical protein inhibitors involve several intermolecular main-chain hydrogen bonds as do many other protein:protein interactions. In an effort to evaluate the energetic contribution of one such hydrogen bond, Wuyuan Lu introduced an ester bond between the P2 and P1 residues of OMTKY3, a serine proteinase inhibitor from the Kazal family (Lu *et al.*, 1997). The amide NH of the peptide bond, normally at this position, forms a birfurcated hydrogen bond with Oγ of the catalytic serine residue, Ser195, and the carbonyl O of Ser214 in complexes of OMTKY3 with serine proteinases (Fujinaga *et al.*, 1987; Huang, 1995; Bode *et al.*, 1986). The mutation to an ester bond would remove this hydrogen bond from a complex with a serine proteinase. Association equilibrium constants have been determined for wild type OMTKY3 and the ester variant in complex with six cognate enzymes and it was determined that the mutation to an ester, weakened the binding of OMTKY3 to the enzymes by 1.5 kcal/mol (Lu *et al.*, 1997). A structure of the mutant OMTKY3 variant in complex with a serine proteinase is absolutely required to ensure the protein:protein interactions are identical in both complexes except for the loss of the single hydrogen bond. This structure will be presented in this chapter.

Another backbone mutant of OMTKY3 was synthesized with a reduced peptide bond between P1 and P1' (W. Lu, personal communication). No inhibition towards chymotrypsin was observed for this variant. A structure of the reduced peptide bond variant was determined in order to speculate why this variant was an unsuccessful inhibitor.

6.2 Materials and Methods

Dr. L.B. Smillie kindly provided SGPB, purified from Pronase according to the procedure of Jurasek *et al.*, 1979. OMTKY3-COO-Leu^{18I} and OMTKY3-CH2-Asp^{18I} were generously provided by Dr. W. Lu. The variants were obtained by total synthesis as described by Lu *et al.*, 1997b.

Enzyme:inhibitor complex crystals of SGPB:OMTKY3-COO-Leu^{18I} were grown by the hanging drop method from protein solutions containing SGPB and OMTKY3-COO-Leu^{18I} in a 1:1.5 ratio. 4% PEG 4000 was used as the precipitant for SGPB:OMTKY3-COO-Leu^{18I} with 50mM $NaKHPO_4$ as the buffer (pH 7.1). Streak seeding (Stura and Wilson, 1992) was used to obtain diffraction quality crystals. The protein concentration in the drops was approximately 10mg/ml. Crystallization conditions for OMTKY3-CH2-Asp^{19I} were determined with the Hampton crystal screen (Hampton Research). Protein concentrations of approximately 10 mg/ml were used in the hanging drops. Diffraction quality crystals were grown in condition 32, 2.0M ammonium sulphate. No buffer was included in the well.

Suitable crystals were mounted in a capillary and data were collected on a DIP image plate detector (Mac Science Co., Ltd.). X-rays were generated with a Rigaku rotating anode generator, RU-200 BH, operating at 45kV and 75mA. Data collection for the complex crystals lasted three days. The small unit cell for the OMTKY3-CH2-Asp^{19I} crystals allowed a complete data set to be collected in 30 hours. The diffraction data for both data sets were processed with Denzo (Otwinowski and Minor, 1996).

The molecular replacement solution for the complex structure was straightforward. The SGPB:OMTKY3 (wild type) (Read *et al.*, 1983; Fujinaga *et al.*, 1982) structure, which was determined in the same space group and unit cell as the complex structure reported in this paper, was used as the molecular replacement model. Electron density maps were inspected with Xtalview (McRee, 1993). The backbone nitrogen atom, NH, of the P1 residue was replaced by an oxygen in SGPB:OMTKY3-COO-

Leu^{18I}. No other alterations were necessary. The new model was subjected to simulated annealing and energy minimization in X-plor (Brünger, 1992b). Further refinement was carried out with TNT (Tronrud, 1992). Parameter files for both X-plor and TNT were edited to include values for an ester bond. Ester bond parameters were obtained from previously determined parameter files (Kleywegt, 1995; Fraser *et al.*, 1992; Schmid *et al.*, 1983). Values included in these files were chosen using data from the Cambridge Structural Database (Fraser *et al.*, 1992).

The OMTKY3-CH2 structure was also determined by molecular replacement. The OMTKY3 coordinates from the SGPB:OMTKY3 (wild type) (Read *et al.*, 1983; Fujinaga *et al.*, 1982) structure were used as the search model. 5% of the reflections were removed before the structure determination and refinement in order that an R-free value could be calculated to monitor refinement (Brünger, 1992a). Rotation and translation functions were calculated by AMORE (Navaza, 1994) with data from 10 to 3 Å. AMORE was also used for rigid body fitting before the solution was subjected to simulated annealing in X-plor. The first X-plor refinement did not include the reduced peptide bond and the P1' residue was left as Glu^{19I}. The peptide bond was subsequently reduced, the Glu^{19I} side chain was replaced by an Asp^{19I} side chain and the model underwent an additional refinement with X-plor. In a manner similar to the SGPB:OMTKY3-COO-Leu18I complex, parameter files were edited to include values for a reduced peptide bond. These parameters were chosen using data from the Cambridge Structural Database. All further refinement was carried out in TNT, also using adjusted parameter files.

Electron density maps and models were examined and adjusted with the graphic program Xtalview (McRee, 1993). Model analysis was carried out with the programs Procheck (Laskowski *et al.*, 1993) and Whatcheck (Rodriguez *et al.*, 1998) at various stages of refinement and for the final model. Xtalview was used to calculate rms deviations among the structures in a pairwise fashion. Figures were made with a variety of programs: Xtalview, GRASP (Nicholls *et al.*, 1991), and Raster3D (Bacon and Anderson, 1988).

6.3 Results

Crystals of the enzyme:inhibitor complex were successfully grown in the traditional SGPB:OMTKY3 space group of $P2_1$ and unit cell of a=45.41 Å, b=54.65 Å, c=45.53 Å, β=119.13° with one complex per asymmetric unit. The crystallization conditions for the OMTKY3-reduced peptide variant, were found by screening conditions with the Hampton kit (Hampton Research). The best crystals appeared from condition 32, 2.0M ammonium sulphate, although crystals grew from nearly every condition in the screen. The space group and unit cell for this crystal was $P2_1$, a=23.10 Å, b=36.07 Å, c=26.11 Å and β=97.58°.

Diffraction data were collected to 1.7Å and 1.65Å resolution for the SGPB:OMTKY3-COO-Leu^{18I} crystals and the OMTKY3-CH2-Asp^{19I} crystals respectively. The completeness of the data, R-merge and other statistics pertaining to data collection can be found in Table 6.1.

AMORE (Navaza, 1994) was used to calculate rotation and translation functions for the OMTKY3 coordinates from the SGPB:OMTKY3 complex structure (Read *et al.*, 1983; Fujinaga *et al.*, 1982) The correlation coefficients for the top three rotation function solutions for the OMTKY3-CH2-Asp^{19I} data were 40.7 (5.7σ above the mean density), 32.0 (4.5σ) and 27.8 (3.9σ). Translation improved the correlation coefficients of the first solution to 59.8 with an Rcryst of 40.3%. The correlation coefficients and Rcryst for the next two solutions were 32 and 50.9% and 14.6 and 57%. After rigid body fitting, the top two solutions had the same correlation coefficient (72.8) and Rcryst (34.8 %). These solutions were related by the crystallographic 2_1 screw axis about y. AMORE results using P2 as the space group confirmed that $P2_1$ was the correct choice.

The initial map for all of the SGPB:OMTKY3 variant complex structures calculated with the SGPB:OMTKY3 (wild type) (Read *et al.*, 1983; Fujinaga *et al.*, 1982) model indicated the position and orientation of the P1 residue. In contrast, the switch from a peptide bond to an ester bond was not noticeable in the electron density map for SGPB:OMTKY3-COO-Leu^{18I}

calculated with the original peptide bond SGPB:OMTKY3-Leu^{18I} model (Figure 6.1).

The initial map for OMTKY3-CH2-Asp^{19I} is in Figure 6.2. Adjustments of the search model were made to fit the calculated density as well as possible before simulated annealing in X-plor (Brünger, 1992b). After this refinement, negative $|F_o|-|F_c|$ density appeared around the carbonyl oxygen of Leu^{18I} and near the carboxyl oxygens of Glu^{19I} (Figure 6.3).

Electron density maps calculated with the final models are in Figures 6.4 and 6.5. The final Rcryst values are 0.183 and 0.196 for SGPB:OMTKY3-COO-Leu^{18I} and OMTKY3-CH2-Asp^{19I}, respectively. Table 6.2 includes geometry statistics for the final models as well as the final number of atoms and waters. 87.0% of the residues from SGPB:OMTKY3-COO-Leu^{18I} and 87.2% of the residues from OMTKY3-CH2-Asp^{19I} are found in the most favoured regions of the Ramachandran plot.

Pair-wise superimpositions have been calculated for the structures described in this paper (Table 6.3) and for other complex structures of SGPB and OMTKY3 variants (Tables 6.4 and 6.5). The molecules from the SGPB:OMTKY3-Pro^{18I} structure show the poorest agreement with the other complex structures. The poor agreement was true for superimpositions calculated with enzyme atoms, inhibitor atoms or both. Figures 6.6 and 6.7, showing superimpositions of SGPB:OMTKY3-Pro^{18I} and SGPB:OMTKY3-Leu^{18I}, illustrate that most of the differences are located near the active site, and involve the reactive site loop and residues of the active site of SGPB. Tables 3.4, 3.5, and 3.6 from Chapter 3 are a collection of hydrogen-bonding distances near the active site. The distances for the SGPB:OMTKY3-Pro^{18I} complex show the largest deviations from the other structures.

The rms deviations between SGPB:OMTKY3-COO-Leu^{18I} molecules and those from other complex structures are in contrast to the SGPB:OMTKY3-Pro^{18I} results and are generally quite small (Table 6.3). Figures 6.8 and 6.9 nicely illustrate the agreement. Only a very small

Table 6.1 Data Collection Statistics

	SGPB:OMTKY3-COO-Leu[18I]	OMTKY3-CH2-Asp[19I]
Space Group	$P2_1$	$P2_1$
Unit Cell a(Å)	45.41	23.10
b(Å)	54.65	36.07
c(Å)	45.53	26.11
β(°)	119.13	97.58
Maximum resolution (Å)	1.70	1.65
Total no. of observations	61078	16933
No. of unique reflections	18894	4800
Average redundancy	3.2	3.5
Rmerge[a]:		
overall	0.120	0.083
highest resolution shell	0.288	0.245
Resolution range of last shell (Å)	1.74-1.71	1.68-1.65
$\langle I/\sigma(I)\rangle$:		
overall	7.86	10.64
highest resolution shell	2.14	2.51
Resolution range of last shell (Å)	1.71-1.70	1.68-1.65
Completeness of Data:		
overall (%)	87.8	92.9
highest resolution shell (%)	47.1	47.6
Resolution range of last shell (Å)	1.73-1.70	1.68-1.65

[a]$Rmerge=\Sigma_{hkl}\,[(\Sigma_i\,|I_i-\langle I\rangle|)/\Sigma_i\,I_i]$

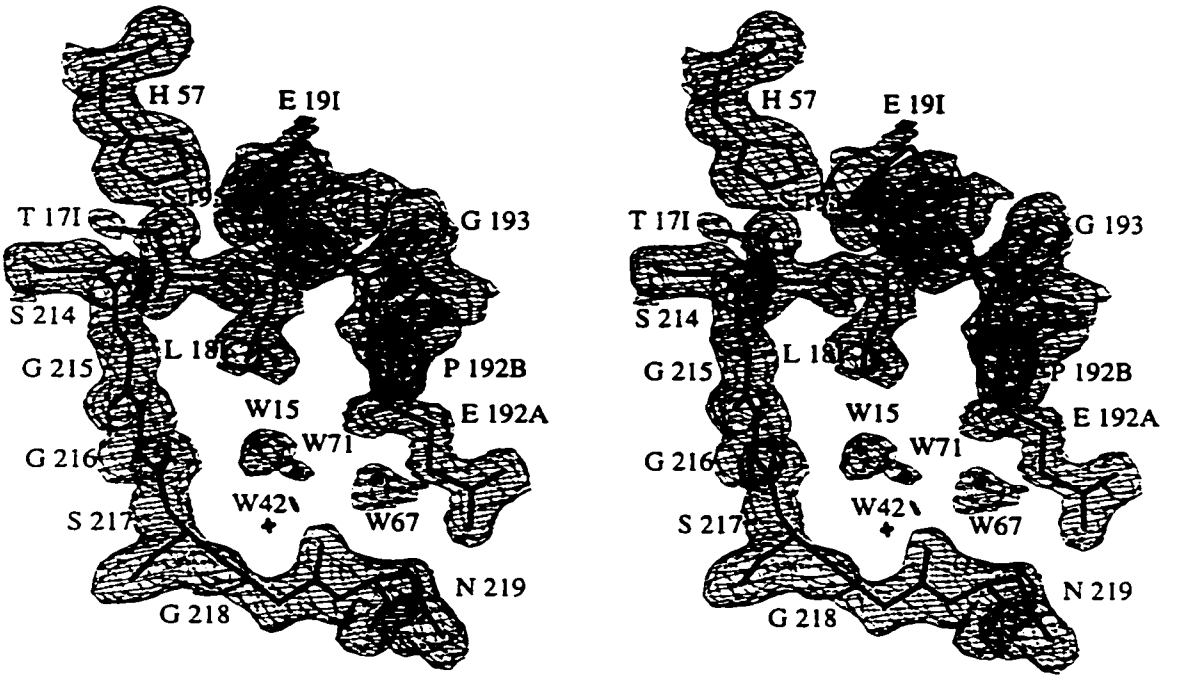

Figure 6.1 Initial electron density map of SGPB:OMTKY3-COO-Leu^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

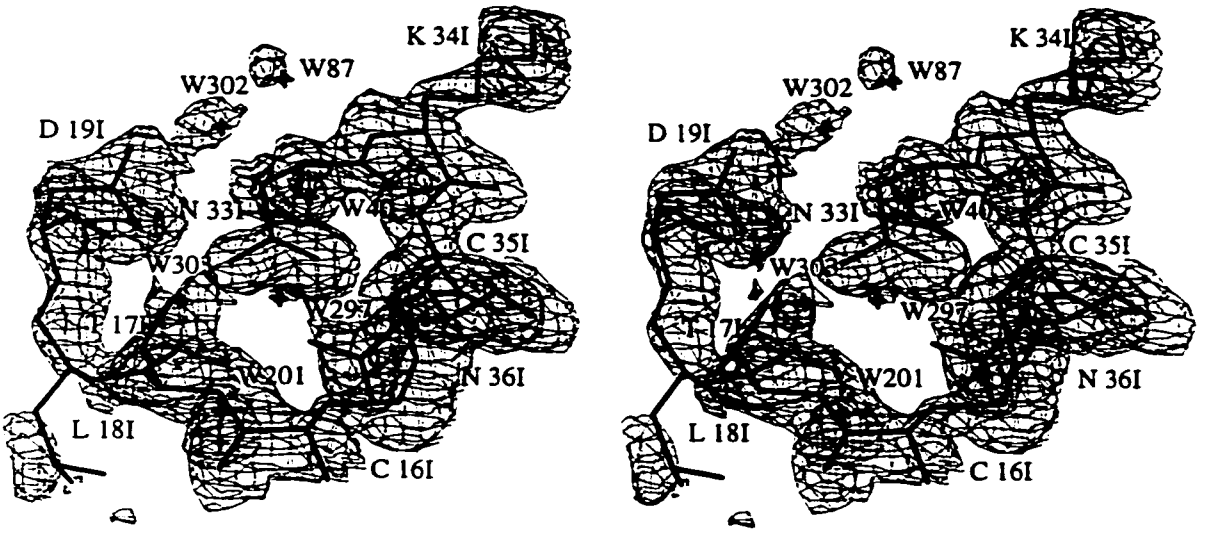

Figure 6.2 Initial electron density map of OMTKY3-CH2-Asp^{19I} superimposed onto the final model in the vicinity of the reduced peptide bond. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

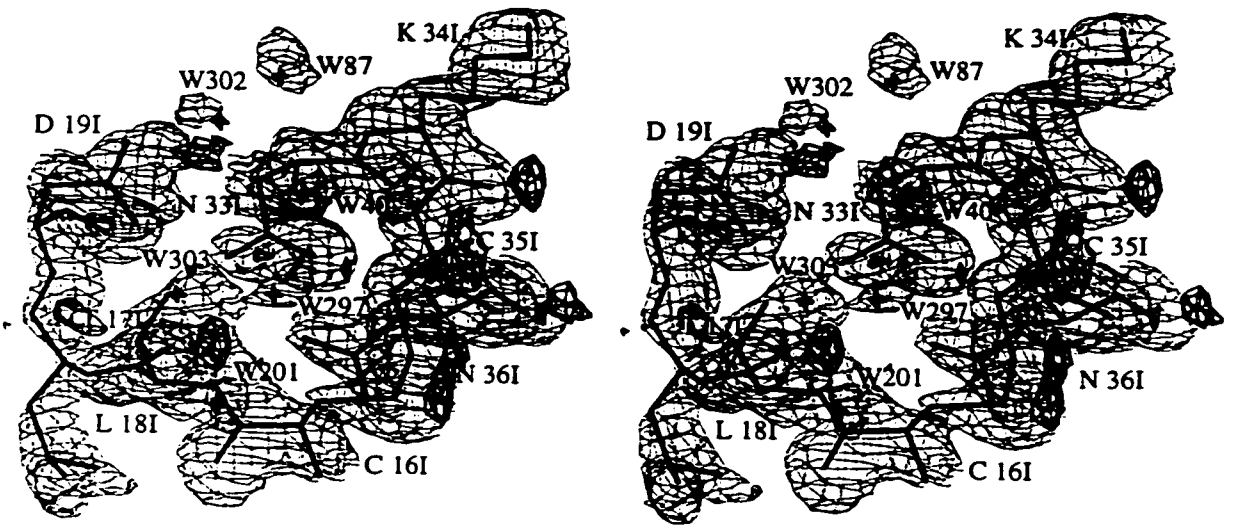

Figure 6.3 The final atomic model of OMTKY3-CH2-Asp^{19I} superimposed onto electron density maps that were calculated after the initial simulated annealing and energy minimization with X-plor. The model for these maps did not yet contain a reduced peptide bond. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ (light lines) and $|F_o|-|F_c|$, contoured at -2.5σ (dark lines).

Table 6.2 Refinement Statistics

	SGPB:OMTKY3-COO-Leu[18I]	OMTKY3-CH2-Asp[19I]
No. of reflections used	18894	4800
Resolution Range (Å)	20-1.70	20-1.65
Rcryst[a] (R-free)	0.183	0.196 (0.230)
No. of protein atoms[b]	1713	385
No. of solvent atoms	140	41
rms deviation from ideal stereochemistry		
bond distance (Å)	0.011	0.012
bond angle (°)	0.985	1.115
planar groups (Å)	0.012	0.011
Average B-values (Å^2)		
main-chain atoms	19	16
side-chain atoms	23	21
solvent atoms	39	36
Error Estimates (Å)		
Luzzati (Luzzati, 1952)	0.19	0.23
SIGMAA (Read, 1986)	0.27	0.26

[a]$Rcryst=\Sigma_{hkl} \, | \, |F_o| - |F_c| \, | \, / \Sigma_{hkl} \, |F_o|$
R values were calculated with all of the data in the resolution range indicated and without a $\sigma(I)$ cutoff.

[b]The number of protein atoms includes those atoms from alternate conformations.

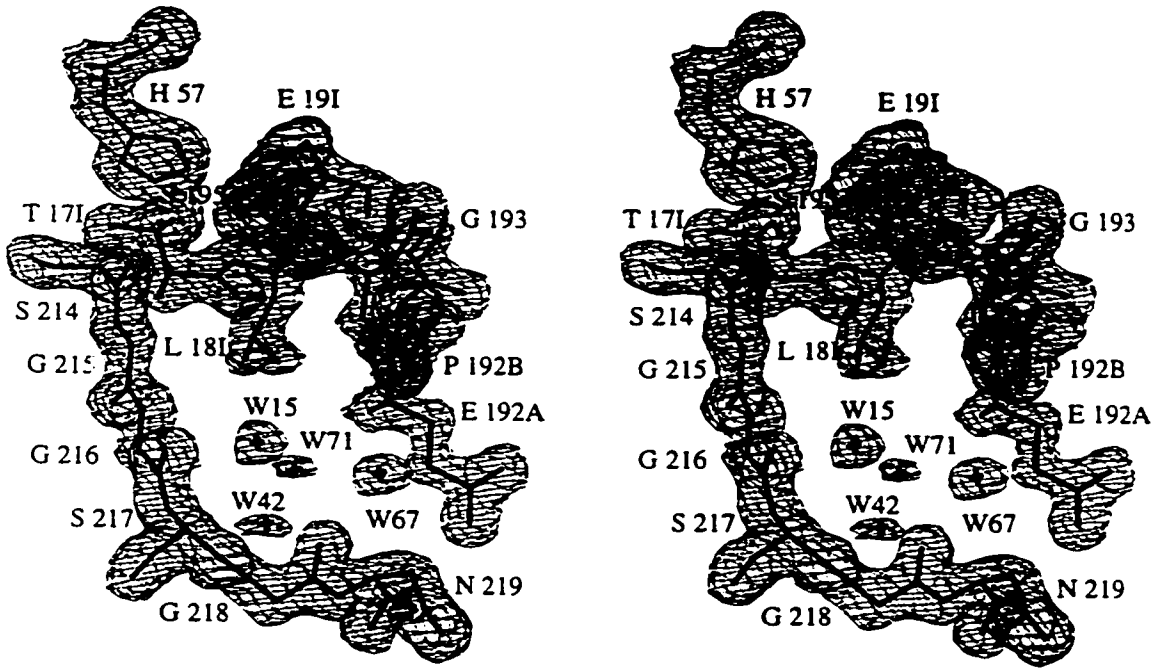

Figure 6.4 Final electron density map of SGPB:OMTKY3-COO-Leu^{18I} in the region of the S1 substrate pocket and active site, superimposed onto the final model. Map coefficients are $2|F_o|-|F_c|$, contoured at 1σ.

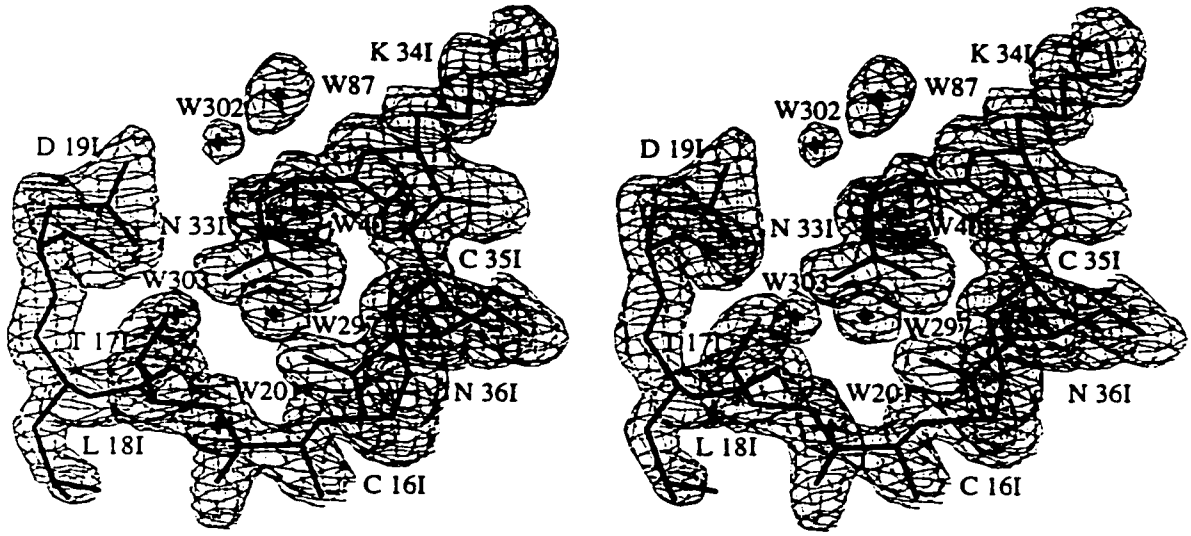

Figure 6.5 Final electron density map of OMTKY3-CH2-Asp^{19I} in the vicinity of the reduced peptide bond, superimposed onto the final model. Map coefficients are 2|Fo|-|Fc|, contoured at 1σ.

deviation of the ester bond between Thr^{17I} and Leu^{18I} with respect to the peptide bond of the SGPB:OMTKY3-Leu^{18I} structure (Huang *et al.*, 1995) is apparent.

Rms deviation calculations for ovomucoids from several sources are found in Table 6.4 and Table 6.5. The results for OMTKY3 variants taken from SGPB:OMTKY3 complex structures are relatively small (Table 6.5). The results for OMTKY3-CH2-Asp^{19I} superimpositions onto OMTKY3 variants taken from within a SGPB:OMTKY3 complex structure (Table 6.5) are larger than those between the complexed OMTKY3 variants. However, these rms deviations do not differ to any greater or lesser extent than the results for superimpositions with other ovomucoids for which structures have been determined on their own, (not bound to enzyme), or bound to an enzyme other than SGPB (Table 6.4). Figure 6.10 shows the superimposition of OMTKY3-CH2-Asp^{19I} onto OMTKY3-Leu^{18I} from SGPB:OMTKY3-Leu^{18I}. Most of the deviation is found in the reactive site loop. Figure 6.11 is the superimposition of seven ovomucoids, OMTKY3-CH2-Asp^{19I}, OMTKY3 from SGPB:OMTKY3 (Read et al., 1983), OMTKY3 from HLE:OMTKY3 (Bode *et al.*, 1986), OMTKY3 from CHYM:OMTKY3 (Fujinaga et al., 1987), OMTKY3 determined by NMR (Hoogstraten *et al.*, 1995), OMJPQ3 (Papmokos *et al.*, 1982), and OMSVP3 (Bode *et al.*, 1985). This figure illustrates the rms deviations among ovomucoids from different environments. The deviations for these superimpositions are not limited to the reactive site loop and generally can be found in the surface loops. The rms deviations for the NMR structure with the other ovomucoids are the largest (Table 6.4) but superimpositions of the core region of OMTKY3 (29I-45I) from this structure onto the others results in lower rms deviations of around 0.53 Å. Most of the differences between the NMR structure and the other ovomucoids are also located in the surface loops. Residue Asn^{33I}, from the ovomucoids superimposes very well, except for the NMR structure (Hoogstraten *et al.*, 1995). On the other hand, the side chain of Asn^{36I} shows very different conformations in several of the structures.

Table 6.3 RMS differences (Å) of main-chain atoms among the structures of the various SGPB:OMTKY3 complexes.

	Pro^{18I}	$-COO-Leu^{18I}$
Leu^{18I}	0.25[1]	0.10
	0.23[2]	0.11
	0.27[3]	0.11
Pro^{18I}		0.24
		0.20
		0.25

[1]The top number in each column refers to calculations among the SGPB molecules of the SGPB:OMTKY3 complexes. 740 main-chain atoms (N, Cα, C, O) from 185 residues were used in each calculation.
[2]The middle number in each column refers to calculations among the OMTKY3 variant molecules of the SGPB:OMTKY3 complexes. 204 main-chain atoms from 51 residues were used in each calculation.
[3]The bottom number in each column refers to calculations among the SGPB:OMTKY3 complexes. 944 main-chain atoms (740 from SGPB and 204 from OMTKY3) from 236 residues were used in each calculation.

Table 6.4 RMS differences (Å) of main-chain[a] atoms among the structures of the ovomucoids from various sources.

	1ovo	2ovo	1omt	1ppf	3sgb	1cho
CH2[b]	0.50	0.38	1.06	0.41	0.41	0.63
1ovo		0.60	1.08	0.61	0.72	0.75
2ovo			1.11	0.43	0.42	0.75
1omt				0.92	1.05	0.96
1ppf					0.39	0.52
3sgb						0.64

[a]200 main-chain atoms (N, Cα, C, O) from 50 residues (7I-56I) were used in each calculation. (199 main-chain atoms were used in calculations with CH2 as Leu18I O was absent.)

[b]CH2 OMTKY3-CH2-Asp19I
1ovo OMJPQ3 (Papamokos *et al.*, 1982)
2ovo OMSVP3 (Bode *et al.*, 1986)
1omt OMTKY3 NMR structure (Hoogstraten *et al.*, 1995)
1ppf OMTKY3 from the complex HLE:OMTKY3 (Bode *et al.*, 1986)
3sgb OMTKY3 from the complex SGPB:OMTKY3 (Read *et al.*, 1983)
1cho OMTKY3 from the complex CHT:OMTKY3 (Fujinaga *et al.*, 1987)

Table 6.5 RMS differences (Å) of main chain atoms among the structures of the various SGPB:OMTKY3 complexes.

	THR	SER	HIS	TRP	LYS+	LYS	ARG	ILE	COO	LEU	PHE	TYR	ALA	GLY	PRO	CH2
VAL	0.05	0.06	0.07	0.09	0.11	0.09	0.09	0.11	0.09	0.13	0.11	0.12	0.10	0.12	0.24	
	0.06	0.08	0.11	0.12	0.09	0.10	0.12	0.16	0.10	0.14	0.13	0.14	0.12	0.12	0.22	0.45
	0.06	0.06	0.11	0.12	0.12	0.11	0.11	0.13	0.11	0.15	0.13	0.15	0.11	0.13	0.25	
THR		0.06	0.07	0.10	0.11	0.09	0.10	0.12	0.09	0.13	0.11	0.12	0.10	0.12	0.24	
		0.08	0.11	0.13	0.10	0.12	0.14	0.16	0.10	0.13	0.13	0.14	0.12	0.11	0.22	0.43
		0.07	0.12	0.14	0.13	0.14	0.13	0.15	0.13	0.16	0.12	0.16	0.13	0.14	0.27	
SER			0.08	0.09	0.09	0.09	0.08	0.12	0.10	0.13	0.10	0.11	0.09	0.11	0.23	
			0.13	0.12	0.09	0.11	0.12	0.17	0.12	0.15	0.14	0.15	0.13	0.12	0.23	0.46
			0.11	0.11	0.11	0.11	0.10	0.14	0.12	0.15	0.12	0.14	0.11	0.12	0.25	
HIS				0.10	0.11	0.10	0.09	0.12	0.10	0.14	0.11	0.12	0.10	0.12	0.24	
				0.11	0.12	0.14	0.15	0.16	0.11	0.15	0.11	0.12	0.11	0.12	0.21	0.44
				0.11	0.12	0.13	0.12	0.16	0.14	0.17	0.12	0.13	0.11	0.13	0.25	
TRP					0.09	0.10	0.09	0.13	0.10	0.12	0.09	0.11	0.09	0.10	0.23	
					0.11	0.14	0.15	0.16	0.12	0.16	0.12	0.13	0.12	0.12	0.20	0.47
					0.10	0.13	0.12	0.15	0.13	0.15	0.10	0.12	0.10	0.11	0.25	
LYS+						0.12	0.09	0.15	0.13	0.15	0.12	0.12	0.09	0.10	0.23	
						0.11	0.10	0.18	0.13	0.15	0.13	0.15	0.11	0.11	0.23	0.47
						0.13	0.11	0.17	0.15	0.17	0.13	0.14	0.10	0.11	0.25	
LYS							0.09	0.13	0.10	0.13	0.11	0.13	0.11	0.13	0.25	
							0.13	0.19	0.15	0.15	0.16	0.17	0.15	0.14	0.23	0.48
							0.11	0.15	0.12	0.15	0.14	0.16	0.13	0.14	0.25	
ARG								0.12	0.11	0.14	0.12	0.14	0.10	0.12	0.24	
								0.22	0.18	0.21	0.18	0.18	0.16	0.15	0.26	0.49
								0.16	0.14	0.18	0.15	0.17	0.12	0.13	0.26	
ILE									0.11	0.15	0.14	0.15	0.14	0.16	0.26	
									0.12	0.16	0.16	0.16	0.17	0.18	0.23	0.44
									0.11	0.17	0.16	0.18	0.16	0.17	0.27	
COO										0.10	0.10	0.13	0.11	0.13	0.24	
										0.11	0.10	0.13	0.12	0.12	0.20	0.42
										0.11	0.12	0.15	0.13	0.14	0.25	
LEU											0.08	0.12	0.11	0.12	0.25	
											0.11	0.15	0.14	0.13	0.23	0.43
											0.11	0.15	0.13	0.14	0.27	
PHE												0.09	0.08	0.10	0.23	
												0.10	0.10	0.11	0.21	0.44
												0.10	0.09	0.10	0.25	
TYR													0.11	0.11	0.23	
													0.13	0.13	0.22	0.45
													0.12	0.12	0.25	
ALA														0.09	0.23	
														0.11	0.21	0.44
														0.10	0.24	
GLY															0.21	
															0.23	0.44
															0.24	
PRO																0.47

The column and row headings refer to the SGPB:OMTKY3 complex structure with that residue at the P1 position.
The top number in each column refers to calculations among the SGPB molecules of the SGPB:OMTKY3 complexes. 740 main chain atoms (N, Cα, C, O) from 185 residues were used in each calculation.
The middle number in each column refers to calculations among the OMTKY3 variant molecules of the SGPB:OMTKY3 complexes. 204 main chain atoms from 51 residues were used in each calculation. Calculations in the last column are between OMTKY3 variant molecules from SGPB:OMTKY3 complexes with OMTKY3-CH2-Asp19I. 199 main chain atoms from 50 residues were used for these calculations.
The bottom number in each column refers to calculations among the SGPB:OMTKY3 complexes. 944 main chain atoms (740 from SGPB and 204 from OMTKY3) from 236 residues were used in each calculation.

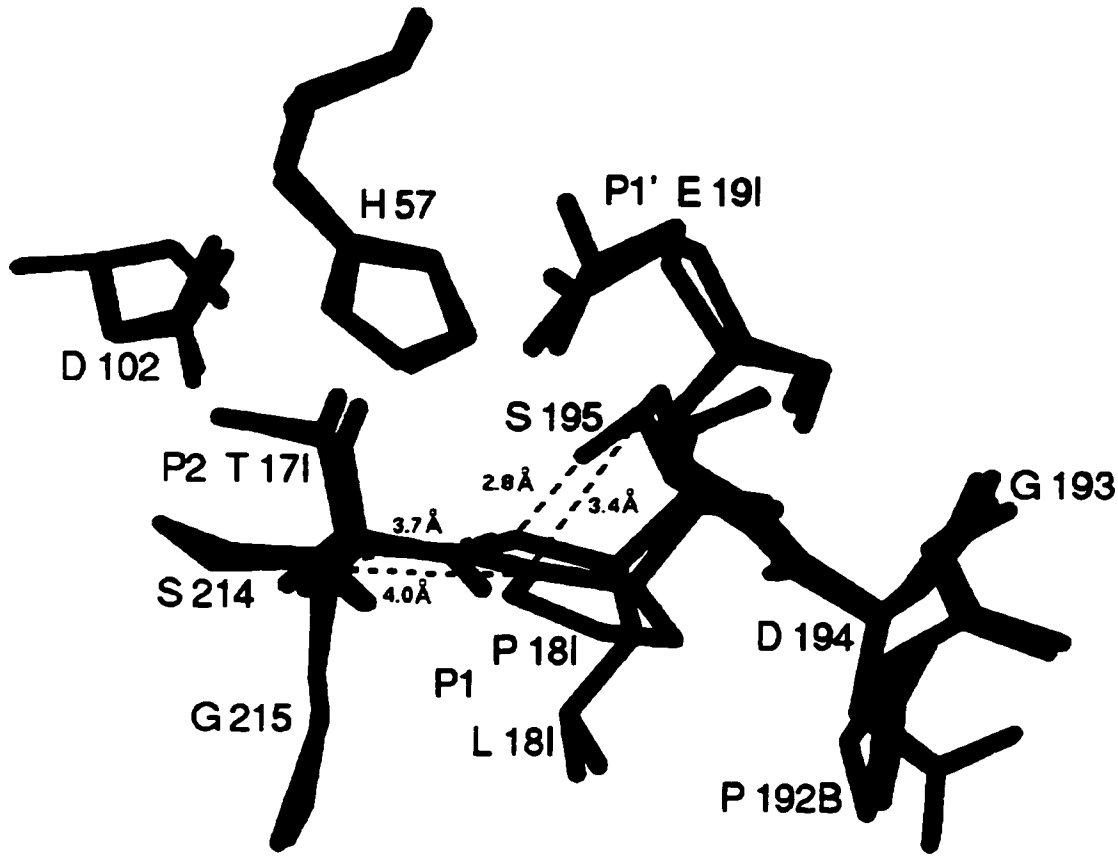

Figure 6.6 Close up of the active site after superimposition of the SGPB molecules from SGPB:OMTKY3-Pro^{18I} and SGPB:OMTKY3-Leu^{18I}. Oxygen atoms are red and nitrogen atoms are blue. Carbon atoms from SGPB:OMTKY3-Pro^{18I} have been coloured orange and those from SGPB:OMTKY3-Leu^{18I} have been coloured green.

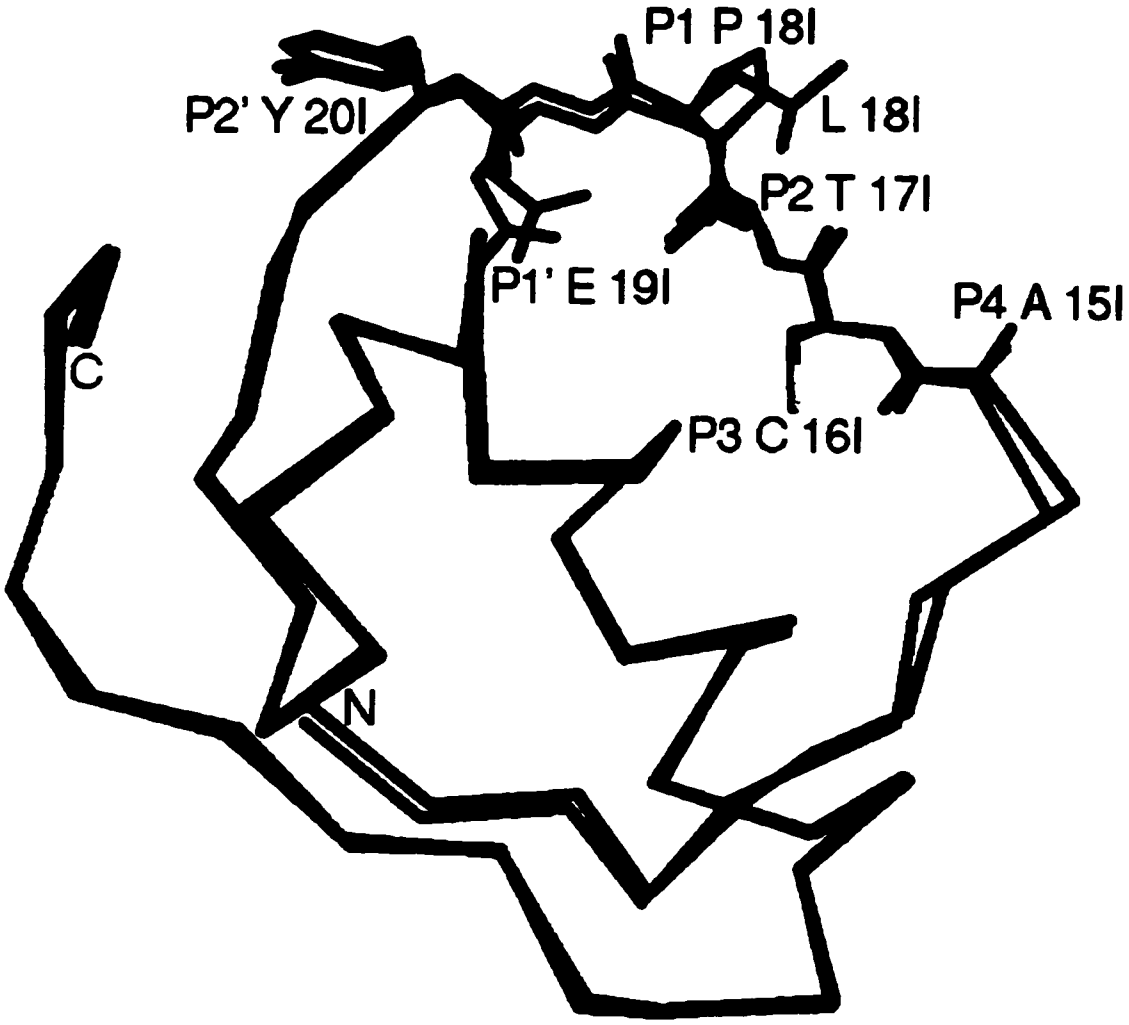

Figure 6.7 Superimposition of the OMTKY3 variants from SGPB:OMTKY3-Leu^{18I} (green) and SGPB:OMTKY3-Pro^{18I} (orange). All atoms have been depicted for the segment from P4 to P2' (oxygens are red, nitrogens are blue and sulphurs are yellow). The remaining residues are depicted as a Cα trace.

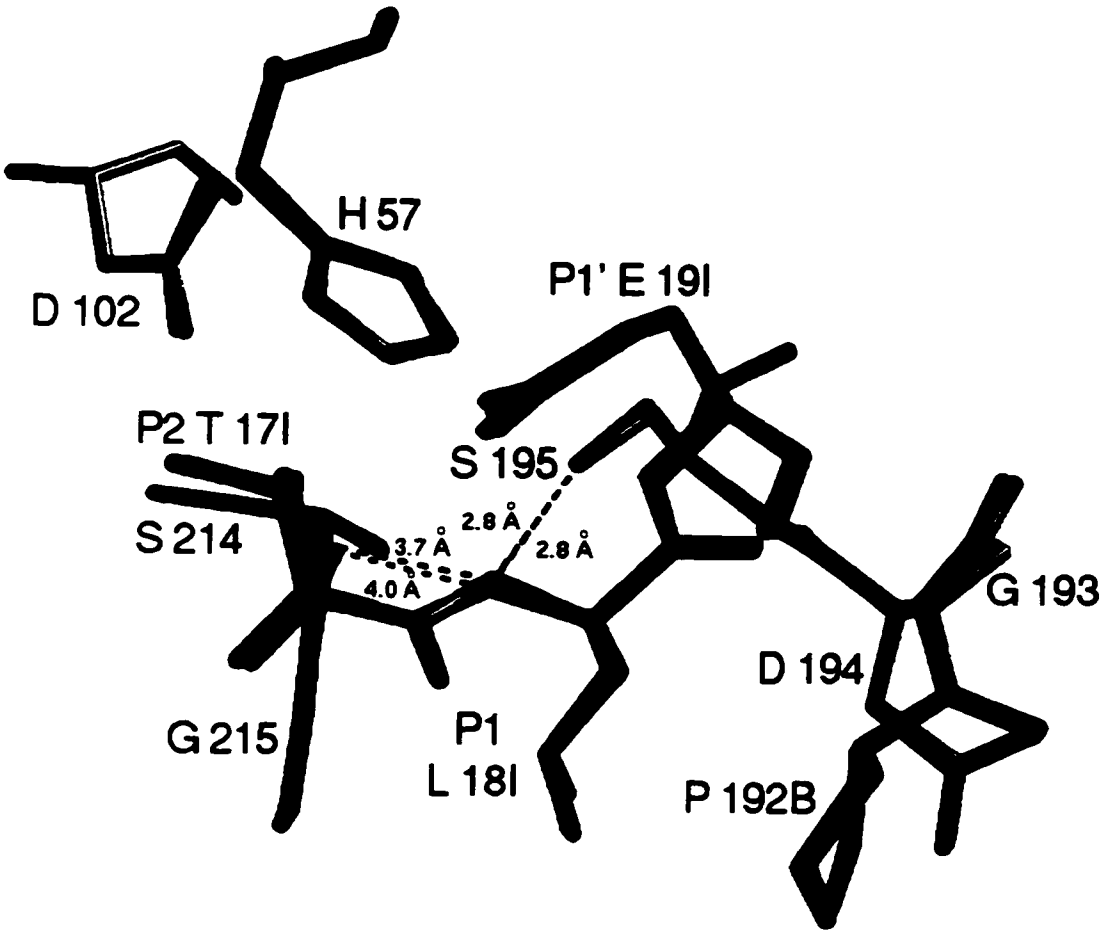

Figure 6.8 Close up of the active site after superimposition of the SGPB molecules from SGPB:OMTKY3-COO-Leu^{18I} and SGPB:OMTKY3-Leu^{18I}. Oxygen atoms are red and nitrogen atoms are blue. Carbon atoms from SGPB:OMTKY3-COO-Leu^{18I} have been coloured pink and those from SGPB:OMTKY3-Leu^{18I} have been coloured green.

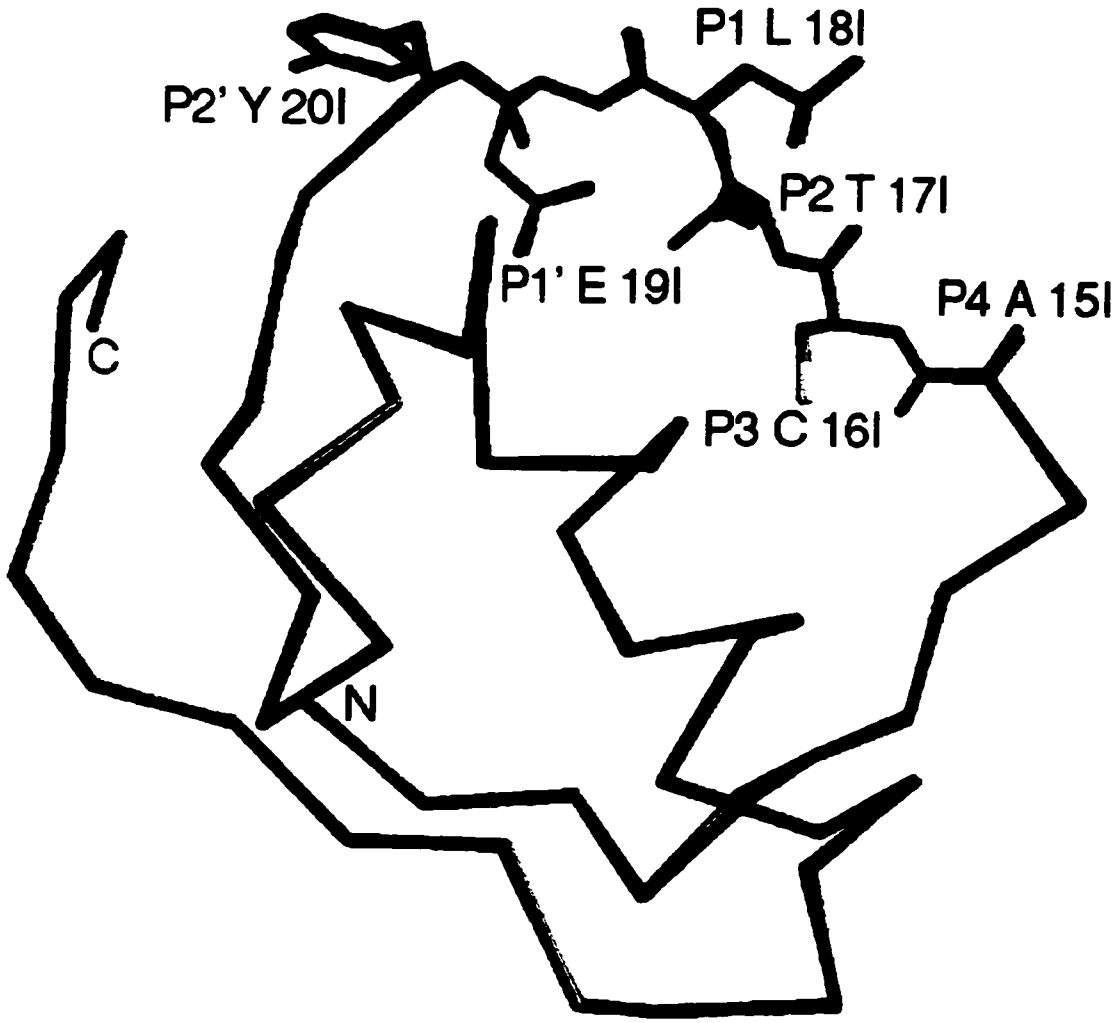

Figure 6.9 Superimposition of the OMTKY3 variants from SGPB:OMTKY3-Leu^{18I} (green) and SGPB:OMTKY3-COO-Leu^{18I} (pink). All atoms have been depicted for the segment P4 to P2' (oxygens are red, nitrogens are blue and sulphurs are yellow). The remaining residues are depicted as a Cα trace.

6.4 Discussion

The OMTKY3 variant with an ester bond connecting Thr^{17I} to Leu^{18I} has an equilibrium association constant of 1.1×10^9 (M^{-1}) with SGPB (Lu *et al.*, 1997b). Compared to the OMTKY3-Leu^{18I} variant with a peptide bond ($Ka=2.0 \times 10^{10}$ (M^{-1})), this difference in Ka corresponds to a $\Delta\Delta G°$ of 1.7 kcal/mol. Kas have been measured for several enzymes in complex with OMTKY3-COO-Leu^{18I} and in complex with the wild type variant (Lu *et al.*, 1997b). The average change in $\Delta G°$ for the enzymes was 1.5 kcal/mol and this value was estimated to be the cost of losing a bifurcated hydrogen bond donated by Leu^{18I} NH to Ser^{195} $O\gamma$ and Ser^{214} O. This estimate is in agreement with values of 0.5-1.8 kcal/mol from other studies where hydrogen bonds between side-chain atoms have been removed (Fersht *et al.*, 1985; Shirley *et al.*, 1992; Fersht, 1987). Similar measurements with a backbone mutant of BPTI in association with trypsin resulted in an estimate of 1.9 kcal/mol (Groeger *et al.*, 1994). Structures have been determined for thermolysin in complex with a peptide inhibitor and with its ester analog. The difference in intrinsic binding energy for these two molecules with thermolysin, 4.1 kcal/mol,was attributed to the loss of a specific hydrogen bond (Tronrud *et al.*, 1987; Bash *et al.*, 1987; Bartlett and Marlowe, 1987). However, this hydrogen bond was later re-examined by Grobelny *et al.*, 1989, and a value of 1.5 kcal/mol was determined. A structural comparison of the structures SGPB:OMTKY3-Leu^{18I} (Huang *et al.*, 1995) and SGPB:OMTKY3-COO-Leu^{18I} indicates that much of the change in $\Delta G°$ between these two complexes can also be attributed to one hydrogen bond. In general, the structures are identical, except for the absence of the amide hydrogen from the SGPB:OMTKY3-COO-Leu^{18I} complex. The rms deviation for all backbone atoms is 0.11 (Table 6.3). Figures 6.8 and 6.9 highlight the lack of major differences. Even in the vicinity of the ester bond, differences between the two complexes are minimal. Ester bonds, like peptide bonds are planar and the C-O-C angle (113°, Simmoneta and Carra, 1969) is only slightly smaller than the C-N-C angle (122°, Engh and Huber, 1991). After superimposition of the enzyme molecules, the distance between SGPB:OMTKY3-COO-Leu^{18I} ester O and SGPB:OMTKY3-Leu^{18I} NH is 0.23 Å. Another minor, but notable,

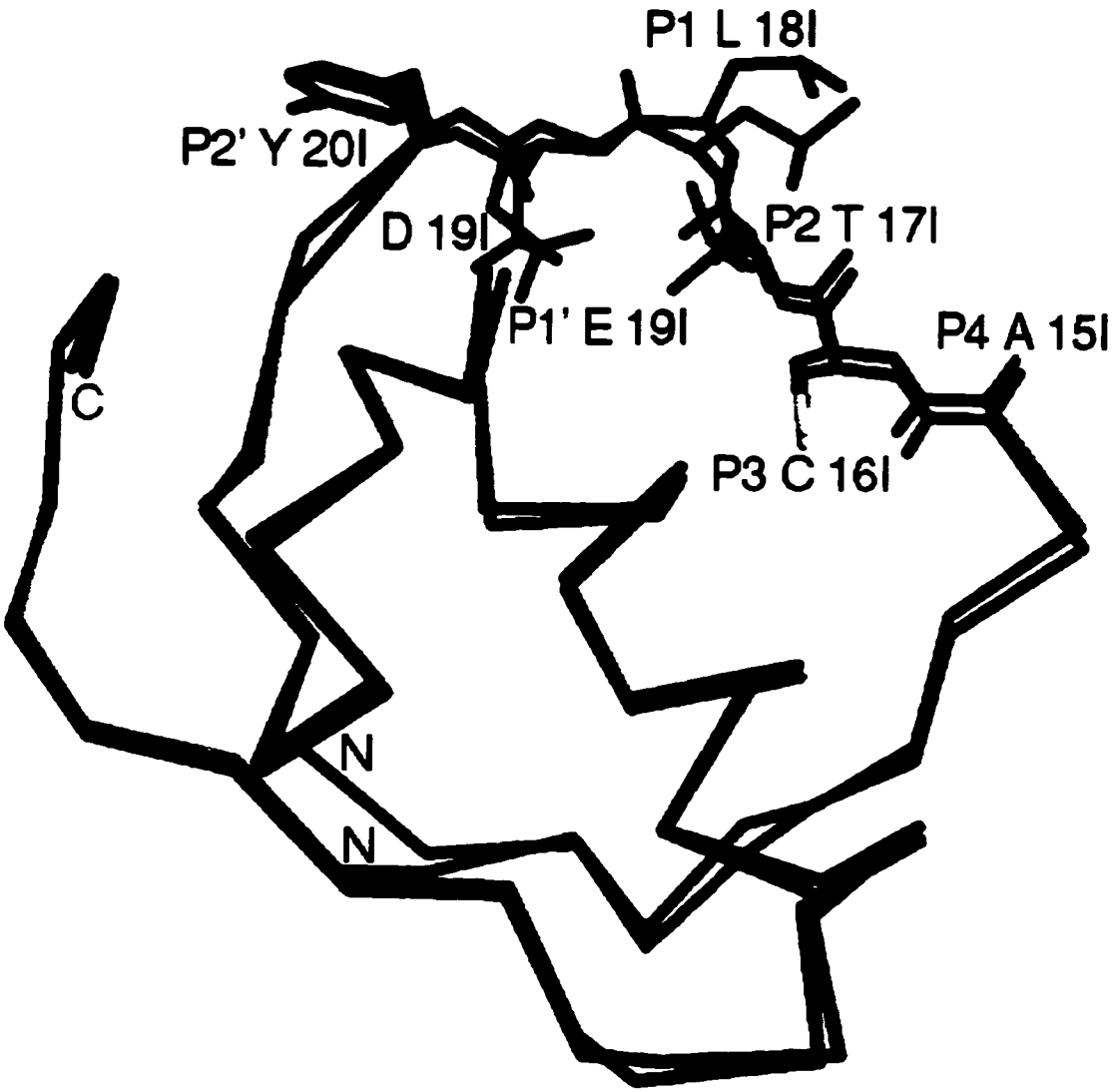

Figure 6.10 Superimposition of OMTKY3-CH2-Asp19I (violet) onto OMTKY3-Leu18I (green) (from the complex structure of SGPB:OMTKY3-Leu18I). Molecules have been depicted as a Cα trace except for residues P4 to P2', for which all atoms have been depicted (oxygens are red, nitrogens are blue and sulphurs are yellow).

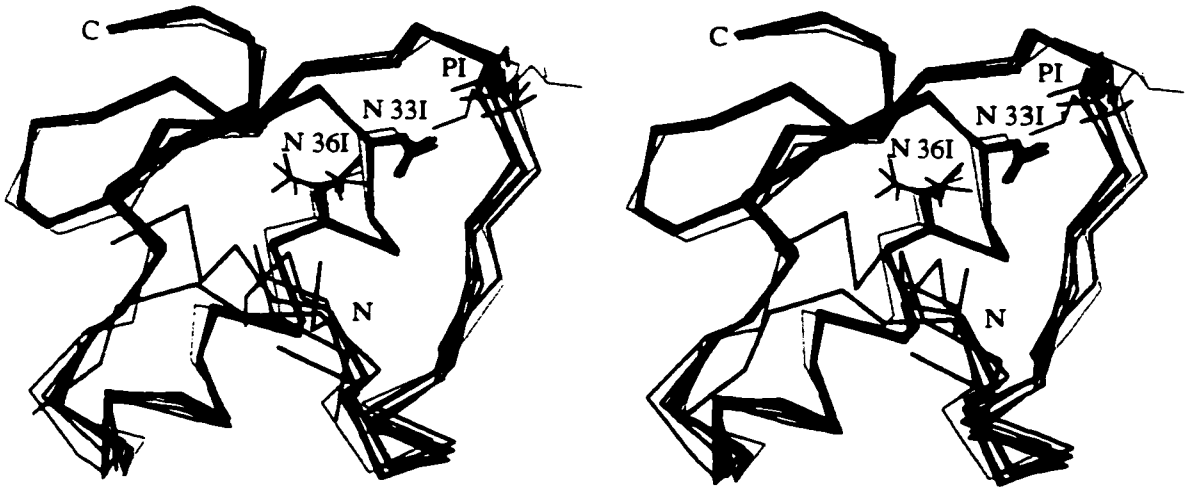

Figure 6.11 Superimposition of OMTKY3-CH2-Asp^{19I} (dark blue), OMTKY3-Leu^{18I} from 3sgb (pink), OMTKY3 NMR solution (yellow), OMTKY3 from 1cho (purple), OMTKY3 from 1ppf (red), OMJPQ3 (light blue), and OMSVP3 (green). Side chain atoms have been included for the residues P1, Asn^{33I} and Asn^{36I}. The remaining residues have been drawn as a Cα trace.

difference is between the carbonyl O of Ser214 from the two complex structures of 0.25 Å. Such variation is not significant at the resolution of these complex structures but the hydrogen bonding interaction between the backbone NH and the carbonyl O of Ser214 has been lost and therefore, it is not surprising to see a slightly longer distance between the ester O and the carbonyl O of Ser214. The peptide dipole from the peptide bond between the P2 and P1 residues has also been reduced with an ester bond introduced at this position. The P2-P1 peptide dipole helps to lower the pKa of Ser195 Oγ in order to facilitate the creation of a good nucleophile. After formation of the tetrahedral intermediate, the backbone NH loses its hydrogn bond to Ser195 Oγ and points more towards the carbonyl O of Ser214, thus stabilizing the tetrahedral intermediate. None of these interactions are possible with an ester bond between residues P2 and P1.

A proline at position 18I, also means the loss of the same amide hydrogen and potential hydrogen bond but in addition, a proline introduces major distortions in the reactive site loop as well (Figures 6.6 and 6.7; Tables 3.4, 3.5, and 3.6). The strain of accommodating this variant is reflected in the very low value of K_a, 3.6×10^4 (M^{-1}) (Lu *et al.*, 1997). The difference in ΔG between this variant and SGPB:OMTKY3-Leu^{18I} is 8.3 kcal/mol. The structural differences between these two complexes are distributed throughout several residues from the binding loop and from the enzyme, particularly the catalytic residue Ser^{195}, whose γ oxygen is 0.97 Å from Ser^{195} Oγ in SGPB:OMTKY3-Leu^{18I} (Figure 6.6) after superimposition of the enzymes. Ser195 is not in an ideal position for nucleophilic attack on the carbonyl carbon with an angle of 85° for the Ser195 Oγ, carbonyl C and carbonyl O atoms (Burgi *et al.*, 1973). Even if a more appropriate structural comparison is made between SGPB:OMTKY3-Pro^{18I} and SGPB:OMTKY3-Gly^{18I} (ΔΔG=3.4kcal/mol), too many factors contribute towards the change in ΔG to equate binding energy with structural features.

OMTKY3-CH2-Asp^{19I} is another backbone mutant of OMTKY3 but this variant was unable to inhibit chymotrypsin. Reduction of the peptide

bond at this location has three consequences that probably contribute to the lack of inhibition. Firstly, unlike an ester bond, the reduced peptide bond does not have to remain planar. Therefore, it would be more costly, in terms of entropy, to restrict the conformation of Leu^{18I}-Asp^{19I} to the conformation observed within the other SGPB:OMTKY3 complexes, than if a peptide bond (or perhaps ester bond) were present. Secondly, a fundamental interaction between enzyme and inhibitor would be missing in a CHYM:OMTKY3-CH2 complex. That interaction is the P1 carbonyl oxygen atom accepting two hydrogen bonds from the oxyanion hole (Ser195 NH and Gly193 NH). This oxyanion hole is well conserved in active serine proteinases. Although these hydrogen bonds are generally discussed in terms of their contribution towards the catalytic mechanism (stabilization of the negative charge developing on the carbonyl oxygen of the scissile bond), they must also play a role in substrate and canonical inhibitor recognition. Lastly, amines are more basic than amides and the backbone N of the reduced peptide bond is almost certainly protonated. It would be very unfavourable to have a positive charge close to an oxyanion hole that is designed to stabilize a negative charge. In contrast, inhibitors with a reduced peptide at the P1-P1' position have been complexed with aspartic proteinases (Foundling *et al.*, 1987; Sugana *et al.*, 1987; Miller *et al.*, 1989). Unlike the serine proteinases, the aspartic proteinases do not have an easily identifiable oxyanion hole and the presence of two catalytic aspartic acid residues creates a favourable environment for a positive charge.

No other structural anomalies are apparent in the crystal structure of OMTKY3-CH2-Asp^{19I} that would indicate the lack of inhibitory action. The rms deviation between OMTKY3-CH2-Asp^{19I} and OMTKY3 molecules bound to serine proteinases are no larger than the rms deviations among the ovomucoid structures determined independently from the enzymes they inhibit (Tables 6.4 and 6.5 and Figures 6.10 and 6.11).

Chapter 7: Conclusions

Twelve refined crystal structures have been reported in this thesis: a structure of the zymogen, human pepsinogen A, ten structures of nine OMTKY3 variants in complex with SGPB and a structure of an OMTKY3 variant having a reduced peptide bond.

Human pepsinogen A

Human pepsinogen A is made up of a central, six-stranded, anti-parallel β-sheet and N- and C-terminal lobes, also consisting mostly of strands of anti-parallel β-sheet. The first strand of the central β-sheet is provided by the N-terminal portion of the prosegment (Tyr3P to Leu7P). Following this strand, the prosegment is made up of two α-helices, a 3_{10} helix, and a coil region. The junction between the prosegment and the beginning of the mature enzyme is located in the coil region. Pepsinogen is inactive because the prosegment fills the active site cleft thereby precluding substrate binding. In addition, catalytic residues Asp32 and Asp215 are electrostatically stabilized by residues Lys37P, Tyr38P and Tyr9. The structures of porcine pepsinogen, human progastricsin, and human pepsin share similar overall conformations. Notable features of human pepsinogen A in comparison with human pepsin are three β-strands (Val1-Glu7 of pepsin, Met2P-Ile8P and Leu179-Thr185 of pepsinogen) that have similar secondary structure but generally different side chains and a phenylalanine at position 111 that has a χ_2 angle of 0° in human pepsinogen A, and 70° in human pepsin. These features will be discussed in subsequent paragraphs.

OMTKY3 P1 variants in complex with SGPB

In an effort to design a sequence-to-reactivity algorithm, people in the lab of Dr. Michael Laskowski, Jr. have synthesized a huge number of OMTKY3 variants and has measured the association equilibrium constants for all of these variants with six cognate enzymes. Dr. Michael

James' group has formed a collaboration with Dr. Laskowski in order to study the relationship between the three dimensional structures and the thermodynamic measurements of the SGPB:OMTKY3 complex series. Towards this end, Kui Huang, a former graduate student of Dr. James has determined the structures of SGPB:OMTKY3-Gly^{18I}, SGPB:OMTKY3-Ala^{18I}, SGPB:OMTKY3-Pro^{18I}, SGPB:OMTKY3-Gln^{18I}, SGPB:OMTKY3-Asn^{18I}, SGPB:OMTKY3-Glu^{18I}, SGPB:OMTKY3-Asp^{18I}, SGPB:OMTKY3-Phe^{18I}, SGPB:OMTKY3-Tyr^{18I}, and SGPB:OMTKY3-Leu^{18I}. The SGPB:OMTKY3-Gln^{18I}, SGPB:OMTKY3-Asp^{18I}, and SGPB:OMTKY3-Glu^{18I} structures were determined at two different pH values, 6.5 and 10.7.

The structures presented in this thesis were the β-branched P1 residues, SGPB:OMTKY3-Ile^{18I}, SGPB:OMTKY3-Val^{18I}, and SGPB:OMTKY3-Thr^{18I}; SGPB:OMTKY3-Ser^{18I}; the positively charged P1 residues, SGPB:OMTKY3-Arg^{18I} and SGPB:OMTKY3-Lys^{18I}; the aromatic P1 residues, SGPB:OMTKY3-Trp^{18I} and SGPB:OMTKY3-His^{18I}; and the backbone variant, SGPB:OMTKY3-COO-Leu^{18I}. SGPB:OMTKY3-Lys^{18I} was also determined at high pH, such that Nζ of the P1 lysine was unprotonated and neutral (-NH_2). A model of SGPB:OMTKY3-Cys^{18I} was suggested based on the structure of SGPB:OMTKY3-Ser^{18I}. In addition, a backbone mutant of OMTKY3 with a reduced peptide bond between the P1 and P1' residues was synthesized by Wuyuan Lu. This variant did not inhibit chymotrypsin (Dr. W. Lu, personal communication). A crystal structure of this OMTKY3 variant has also been determined in order to elucidate the possible reasons for its inactivity.

SGPB's S1 pocket is able to tolerate all of the naturally occurring amino acids at the P1 position, as presented within the framework of OMTKY3, including P1 proline. The β-branched residues are among the poorest P1 residues in this series, as indicated by the Ka's. The structures reveal a somewhat empty cavity near the bottom of the pocket and an excessively tight fit for the β-branched atoms near the top. Ile^{18I} enhances this discourse because Cδ also sits tightly against the back of the pocket. The β-branching of Ile^{18I} prevents an alternate χ_1 conformation which might allow more room for Cδ1.

Ser^{18I} has two alternate conformations. The conformation of highest occupancy points away from the back of the pocket. The alternate conformation, Ser^{18I} B, adopts a more commonly observed χ_1 angle, (for serines with ϕ and ψ in a coil region), and donates a hydrogen bond to the Oγ atom of the catalytic serine, Ser^{195}. However, this alternate conformation leaves a larger cavity in the S1 pocket than conformation A. The structural de-stabilization introduced by large internal cavities was suggested as a reason for conformation A having a higher occupancy. The resultant relatively weak Ka for SGPB:OMTKY3-Ser^{18I} was attributed to the fact that a P1 serine was being subjected to opposing forces; a favourable hydrogen bond and prefered $\chi 1$ angle versus the favourable term of a smaller cavity.

The positively-charged P1 variants of OMTKY3 have Kas that are similar to the β-branched P1 residues for the association with SGPB. It follows that the cost of accommodating the positive charge and large size of arginine and lysine is similar to the cost of accommodating the β-branched valine, threonine and isoleucine. However, the de-stabilizing properties of branching and charge are quite different. In order to stabilize the positive charge, the P1 side chain makes several hydrogen bonds to water molecules within SGPB's S1 pocket. Arg^{18I} additionally forms a hydrogen bond to Asn^{36I}. This residue from the inhibitor also assists in accommodating the large size of the side chain by moving from its more commonly observed orientation in other OMTKY3 variants.

Association constants for SGPB:OMTKY3-Lys^{18I} indicate that SGPB prefers a neutral lysine side chain at the P1 position and it was determined that the pK_a of Lys^{18I} in the environment of the S1 pocket was lowered to 8.72 from 10.05 (Dr. M. Laskowski, personal communication). A structure of the SGPB:OMTKY3-Lys^{18I} complex at pH 10.7 reveals that Lys^{18I} no longer forms hydrogen bonds with water molecules within the pocket but instead forms one intramolecular hydrogen bond with the side chain of Asn^{36I}. The SGPB S1 pocket is thus confronted with only the hydrophobic methylene carbon atoms of the side chain of Lys^{18I}.

In contrast to the complex structures discussed thus far, the aromatic P1 variant complexes have Kas that are among the highest in the series (Table 1.2 and Figure 1.2). The aromatic rings of all of the P1 side chains lie in the same plane. The slightly unfavourable χ_2 angle adopted by the side chains is imposed by the sides of the pocket. Asn^{36I} changes conformation to accommodate the large size of Trp^{18I}, Phe^{18I} and His^{18I}. The large size of the association equilibrium constants are probably due to the hydrophobic nature of the aromatic amino-acids, complementing the hydrophobic S1 pocket of SGPB. The nitrogens of the His^{18I}'s imidazole ring form hydrogen bonds with water molecules, but Trp^{18I}, also the largest of the aromatic P1s, has an unsatisfied hydrogen bond donor, Nε. These reasons can contribute to an explanation for SGPB:OMTKY3-Trp^{18I} having the lowest K_a of the aromatic P1 variants with SGPB.

The amide nitrogen of OMTKY3's P1 residue donates a hydrogen bond to the catalytic serine residue of SGPB in all of the SGPB:OMTKY3 variant complexes except for SGPB:OMTKY3-Pro^{18I}. SGPB:OMTKY3-Pro^{18I} is not suitable for studying the loss of this hydrogen bond because a proline at P1 introduces distortions in the inhibitor backbone as well as in several residues of the enzyme. Therefore, the contribution from a single hydrogen bond cannot be isolated from the other effects. To remove the one hydrogen bond without disrupting the other interactions between enzyme and inhibitor, the peptide bond between Thr^{17I} and Leu^{18I} was replaced by an ester bond (Lu *et al.*, 1997). A crystal structure of SGPB:OMTKY3-COO-Leu^{18I} was determined and the structure verified that no disruptions were introduced with this mutation. The K_a for SGPB:OMTKY3-COO-Leu^{18I} is 1.1×10^9 (M^{-1}) (Lu *et al.*, 1997). This corresponds to $\Delta\Delta G°$ of 1.7 kcal/mol weaker binding for the ester relative to the peptide ($K_a = 2 \times 10^{10}$ (M^{-1}) for SGPB:OMTKY3-Leu^{18I}). Since the only difference between SGPB:OMTKY3-COO-Leu^{18I} (Huang *et al.*, 1995) and SGPB:OMTKY3-Leu^{18I} is the loss of one hydrogen bond, 1.7 kcal/mol is the contribution of that hydrogen bond to the free energy of binding.

Another backbone mutant of OMTKY3 was synthesized, this time with a reduced peptide bond between Leu^{18I} and Asp^{19I}. The resulting OMTKY3

variant was unable to function as an inhibitor. The structure of OMTKY3-CH2-Asp^{19I} did not reveal any gross alterations relative to wild type OMTKY3. However, unlike a peptide bond or an ester bond, a reduced peptide bond is not restrained to planarity. Entropically, it would be more costly for the OMTKY3-CH2-Asp^{19I} variant to adopt the conformation seen in the typical SGPB:OMTKY3 complex than an OMTKY3 variant with a planar peptide bond. An important element in the catalytic mechanism of the serine proteinases is the interaction between the carbonyl oxygen of the P1 residue and the oxyanion hole. This interaction, which would be missing from CHYM:OMTKY3-CH2-Asp^{19I}, is also important for the recognition of a substrate or canonical inhibitor by the enzyme. Not only is the carbonyl oxygen missing from OMTKY3-CH2-Asp^{19I}, but the amine nitrogen will be protonated and a positive charge near the oxyanion hole would be very unfavourable.

χ angles

Side chains have preferred χ angles, particularly for χ_1 and χ_2 such that the steric repulsion of atoms within the residue is minimized. This thesis provides several examples of χ angles that are not in an ideal conformation. Phe111 of human pepsinogen A is in an outright unfavourable conformation with a χ_2 angle of 0°. The side chain is kept in this conformation by hydrophobic interactions to surrounding residues. Phe111, only being well conserved among the pepsinogens, also suggests a reason why a pepsinogen intermediate cannot be trapped in the same manner as the progastricsin intermediate. Once the activation process has been started and cleavage has taken place, Phe111 is able to adopt a more favourable χ_2 angle of 70°, close to 90°, as it has in the structures of the mature human and porcine pepsins. The prosegment could no longer re-associate with the enzyme in the same way as in the zymogen. Thus, the activation process cannot be reversed. In contrast, progastricsin must proceed to the strand exchange between prosegment and enzyme before the activation cannot be halted.

The aromatic P1 residues in the ovomucoid variants have χ_2 angles of -50° in the complex structures of SGPB:OMTKY3. Such an angle is less favourable than the expected angle of -90° but is not entirely unfavourable. This minor strain might contribute towards the hydrophobic, aromatic SGPB:OMTKY3 complexes having lower Kas than the hydrophobic but straight-chained or γ-branched P1 variants.

SGPB:OMTKY3-Cys^{18I} has a K_a of 5.7×10^{10} (M^{-1}), whereas SGPB:OMTKY3-Ser^{18I} has a K_a of 5×10^7 (M^{-1}). Although the χ_1 preference for the two P1 side chains cannot not account for such a large difference in K_a, it probably contributes towards the discrepancy.

Complementarity

The strength of a protein:protein interaction depends on how well the associating surfaces are complementary to each other, both structurally and electrostatically. This thesis has presented examples of interactions that are highly complementary, interactions that are accommodated depending on the changing environment, and interactions that are merely tolerated because the surrounding, adjacent interactions are complementary. Finally, this thesis presents an example of a protein:protein interaction that has been destroyed by a minor chemical change to the interface.

The SGPB substrate binding site is pre-formed and rigid. The canonical inhibitor proteins, like OMTKY3, are effective proteinase inhibitors because they are complementary to the pre-formed sites and therefore, can bind very tightly. The primary specificity determinant for the serine proteinases with the canonical inhibitors is the interaction between the P1 residue and the S1 pocket but many other interactions also provide stability to the complex. For example, β-sheet interactions are present between the backbone of the P3 and P2' residues and the enzyme.

This thesis has focused on interactions at the S1 binding pocket in particular. The pocket is pre-formed and superimposes very well for all

of the SGPB:OMTKY3 complex structures determined thus far, with the exception of SGPB:OMTKY3-Pro^{18I}. Ideal P1 residues for SGPB, are hydrophobic, like the pocket, and medium sized. Variants with the highest K_as are SGPB:OMTKY3-Cys^{18I}, SGPB:OMTKY3-Leu^{18I}, SGPB:OMTKY3-Met^{18I}, SGPB:OMTKY3-Lys$^{18I\circ}$, SGPB:OMTKY3-Glu$^{18I\circ}$ and the aromatic P1 variants of OMTKY3 with SGPB. The other interactions between SGPB and OMTKY3 are strong enough and the S1 pocket is large enough that all of the naturally occurring amino acid P1 residues have been tolerated. However, those P1 OMTKY3 variants that poorly complement the S1 pocket, also have low K_as. The pocket itself does not adjust to accommodate the P1 side chains, with the exception of SGPB:OMTKY3-Pro^{18I} (In order to accommodate the P1 proline residue in a fashion that is close to the canonical binding, the residues lining the active site and S1 pocket are perturbed from their normal positions). Water molecules must re-arrange in the S1 pockets in order to form hydrogen bonds with polar or charged side chains. Bulky side chains are accommodated by a movement of a residue from the inhibitor, Asn^{36I}.

Backbone interactions near the P1 residue were also studied in this thesis. The loss of a hydrogen bond between the P1 amide with Ser^{195} Oγ cost 1.7 kcal/mol but was tolerated. Conversely, the reduction of the peptide bond from P1 to P1' was not tolerated and inhibition was abolished.

The structures of human pepsinogen A and human pepsin provide examples of three β-strands (Val1-Glu7 of pepsin, Met2P-Ile8P and Leu179-Thr185 of pepsinogen) which have similar structural environments in terms of secondary structure but differ markedly in the nature of their side chains. In other words, maintenance of some complementary structural interactions allows accommodation of different side chains at other positions. A proline and leucine (valine for the C-terminus) (Pro6P, Pro5, Pro183, Leu7P, Leu6 and Val184) provide stability for all strands. At neutral pH, positively charged side chains from the prosegment (Ile1P-Ile8P) associate with negative charges on the enzyme. A drop in pH protonates the carboxylate groups so that a β-strand with polar side chains (Val1-Glu7) can be accommodated at the

same location. A similar strand on the C-terminal lobe of human pepsinogen A and human pepsin (Leu179-Thr185) accommodates hydrophobic side chains at both neutral and low pH.

The twelve structures presented in this thesis of protein:protein interactions between an enzyme and its prosegment, an inhibitor residue and enzyme specificity pocket, and an OMTKY3 variant incapable of inhibition enhance the thermodynamic measurements performed in Dr. Laskowski's laboratory and provide a resource for future work in the understanding and manipulation of protein:protein interactions.

References

Abad-Zapatero, C., Tydel, T.J., & Erickson, J. (1990) Revised 2.3 Å structure of porcine pepsin: Evidence for a flexible subdomain. *Proteins* **8**: 62-81.

Al-Janabi, J., Hartsuck, J.A., & Tang, J. (1972) Kinetics and mechanism of pepsinogen activation. *J. Biol. Chem.* **247**: 4628-4632.

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.

Athauda, S. B. P., Tanji, M., Kageyama, T., & Takahashi, K. (1989) A comparative study on the NH_2-terminal amino acid sequences and some other properties of six isozymic forms of human pepsinogens and pepsins. *J. Biochem.* **106**: 920-927.

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

Baca, M., Scanlan, T.S., Stephenson, R.C., & Wells, J.A. (1997) Phage display of a catalytic antibody to optimize affinity for transition-state analog binding. *Proc. Natl. Acad. Sci USA* **94**: 10063-10068.

Bacon, D. J., & Anderson, W.F. (1988) A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics* **6**: 219-220.

Baker, B. M., & Murphy, K.P. (1997) Dissecting the energetics of a protein-protein interaction: The binding of ovomucoid third domain to elastase. *J. Mol. Biol.* **268**: 557-569.

Baker, E. N., & Hubbard, R.E. (1984) Hydrogen Bonding in Globular Proteins. *Prog. Biophys. Molec. Biol.* **44**: 97-179.

Baker, D., Sohl, J.L., & Agard, D.A. (1992) A protein-folding reaction under kinetic control. *Nature* **356**: 263-264.

Barrett, A.J. & Rawlings, N.D. (1995) Perspectives in biochemistry and biophysics. Families and clans of serine peptidases. *Arch. Biochem. Biophys.* **318**: 247-250.

Bartlett, P., & Marlowe, C.K. (1987) Evalutaion of intrinsic binding energy from a hydrogen bonding group in an enzyme inhibitor. *Science* **235**: 569-571.

Bash, P. A., Singh, U.C., Brown, F.K., Langridge, R., Kollman, P.A. (1987) Calculation of the relative change in binding free energy of a protein-inhibitor complex. *Science* **235**: 574-576.

Baudys, M., & Kostka, V. (1983) Covalent structure of chicken pepsinogen. *Eur. J. Biochem.* **136**: 89-99.

Bernstein, F. C., Koetzle, T.F., Williams, G.J.B., Meyer,Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**: 535-542.

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.

BiosymTechnologies (1993). InsightII, version 2.3.0. San Diego.

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.

Blanchard, H. and James, M.N.G. (in preparation) X-ray crystallographic structure of *Streptomyces griseus* proteinase B at its enzymatically optimal pH.

Bode, W., Epp, W., 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.

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.

Brünger, A. T. (1992a) The Free-R value: a novel statistical quality for assessing the accuracy of crystal structures. *Nature* **355**: 472-474.

Brünger, A. T. (1992b). X-PLOR: A system for X-ray crystallography and NMR,version 3.1. Yale University Press, New Haven, CT.

Bürgi, H.B., Dunitz, J.D., & Shefter, E. (1973) Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J. Amer. Chem. Soc.* **95**: 5065-5067.

Bustin, M., & Conway-Jacobs, A. (1971) Intramolecular activation of porcine pepsinogen. *J. Biol. Chem.* **246**: 615-620.

Capasso, C., Rizzi, M., Menegatti, E., Ascenzi, P., & Bolognesi, M. (1997) Crystal structure of the bovine alpha-chymogrypsin:Kunitz inhibitor complex. An example of multiple protein:protein recognition sites. *J. Mol. Recognit.* **10**:26-35.

Carter, P. & Wells, J.A. (1987) Dissecting the catalytic triad of a serine protease. *Nature* **332**: 564-568.

Christensen, K. A., Pedersen, V.B., & Foltmann, B. (1977) Identification of an enzymatically active intermediate in the activation of porcine pepsinogen. *FEBS Lett.* **76**: 214-218.

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

Connolly, M.L. (1983) Analytical molecular surface calculation. *J. Appl. Cryst.* **16**: 548-558.

Cooper, J. B., Khan, G., Taylor, G., Tickle, I.J., & Blundell, T.L. (1990) X-ray analyses of aspartic proteinases. II. Three-dimensional structure of the hexagaonal crystal form of porcine pepsin at 2.3 Å resolution. *J. Mol. Biol.* **214**: 199-222.

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.

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

Davies, D.R. (1990) The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* **19**: 189-215.

Ding, J., Qasim, M.A., Laskowski, M., Jr., & James, M.N.G. (to be published) Crystal structure of the Lys18 variant of turkey ovomucoid inhibitor third domain complexed with α-chymotrypsin at 2.3 Å.

Dixon, M.M., Brennan, R.G., & Matthews, B.W. (1991) Structure of γ-chymotrypsin in the range pH 2.0 to pH 10.5 suggests that γ-chymotrypsin has a covalent acyl-enzyme adduct at low pH. *Int. J. Biol. Macromol.* **13**: 89-96.

Dorovska, V. N., Varfolomeyev, S.D., Kazanskaya, N.F., Klyosov, A.A., & Martinek, K. (1972) The influence of the geometric properties of the active centre on the specificity of α-chymotrypsin catalysis. *FEBS Lett.* **23**: 122-124.

Dunbrack, J., R., & Karplus, M. (1994) Conformational analysis of the backbone-dependent rotamer preferences of protein side chains. *Nature Struct. Biol.* **1**: 334-340.

Dykes, C. W., & Kay, J. (1976) Conversion of pepsinogen into pepsin is not a one-step process. *Biochem. J.* **153**: 141-144.

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

Engh, R.A. & Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* **A47**:392-400.

Eriksson, A. E., Baase, W.A., Zhang, X.-J., Heinz, D.W., Blaber, M., Baldwin, E.P., & Matthews, B.W. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**: 178-183.

Evers, M. P. J., Zelle, B., Peeper, D.S., Mager, W.H., Planta, R.J., Eriksson, A.W., & Frants, R.R. (1987) Molecular cloning of a pair of human pepsinogen A genes which differ by a Glu-Lys mutation in the activation peptide. *Hum. Genet.* **77**: 182-187.

Evers, M. P. J., Zelle, B., Bebelman, J.P., Pronk, J.C., Mager, W.H., Planta, R.J., Eriksson, A.W., & Frants, R.R. (1988) Cloning and sequencing of rhesus monkey pepsinogen A cDNA. *Gene* **65**: 179-185.

Evers, M. P. J., Zelle, B., Bebelman, J.P., van Beusechem, V., Kraakman, L, Hoffer, M.J., Pronk, J.C., Nager, W.H., Planta, R.J., Eriksson, A.W., & Frants, R.R. (1989) Nucleotide sequence comparison of five human pepsinogen A (PGA) genes: Evolution of the PGA multigene family. *Genomics* **4**: 232-239.

Fersht, A. R., & Dingwall, C. (1979) Cysteinyl-tRNA synthetase from *Escherichia coli* does not need an editing mechanism to reject serine and alanine. High binding energy of small groups in specific molecular interactions. *Biochemistry* **18**: 1245-1249.

Fersht, A. R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wildinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., & Winter, G. (1985) Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**: 235-238.

Fersht, A. R. (1987) The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* **12**: 301-304.

Fersht, A. R., & Serrano, L. (1993) Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**: 75-83.

Foltmann, B., & Jensen, A.L. (1982) Human progastricsin. Analysis of intermediates during activation into gastricsin and determination of the amino acid sequence of the propart. *Eur. J. Biochem.* **128**: 63-70.

Foltmann, B. (1988a) Structure and function of proparts in zymogens for aspartic proteinases. *Biol. Chem. Hoppe-Seyler* **369**: 311-314.

Foltmann, B. (1988b) Activation of human pepsinogens. *FEBS Lett.* **241**: 69-72.

Foltmann, B. (1988c). Aspartic proteinases: alignment of amino acid sequences. Proceedings of the 18th Linderstrom-Lang Conference, 4-8 July

1988. (B. Foltmann, ed.). Elsinore, Denmark, University of Copenhagen, Copenhagen: pp7-20.

Foundling, S.I., Cooper, J., Watson, F.E., Cleasby, A., Pearl, L.H., Sibanda, B.L., Hemmings, A., Wood, S.P., B.undell, T.L., Valler, M.J., Norey, C.G., Kay, J., Boger, J., Dunn, B.M., Leckie, B.J., Jones, D.M., Atrash, B., Hallett, A., & Szelke, M. (1987) High resolution X-ray analyses of renin inhibitor-aspartic proteinase complexes. *Nature* **327**: 249-352.

Fraser, M.E., Strynadka, N.C.J., Bartlett, P.A., Hanson, J.E., & James, M.N.G. (1992) Crystallographic analysis of transition-state mimics bound to penicillopepsin: Phosphorus-containing peptide analogues. *Biochemistry* **31**: 5201-5213.

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. USA* **79**: 4868-4872.

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

Fujinaga, M., Chernaia, M.M., Tarasova, N.I., Mosimann, S.C., & James, M.N.G. (1995) Crystal structure of human pepsin and its complex with pepstatin. *Protein Science* **4**: 960-972.

Glick, D. M., Shalitin, Y., & Hilt, C.R. (1989) Studies on the irreversible step of pepsinogen activation. *Biochemistry* **28**: 2626-2630.

Grobelny, D., Goli, U.B., & Galardy, R.E. (1989) Binding energetics of phosphorus-containing inhibitors of thermolysin. *Biochemistry* **28**: 4948-4951.

Groeger, C., Wenzel, H.R., & Tschesche, H. (1994) BPTI backbone variants and implications for inhibitory activity. *Int. J. Peptide Protein Res.* **44**: 166-172.

Guo, J., Huang, W., & Scanlan, T.W. (1994) Kinetic and mechanistic characterization of a hightly active hydrolytic antibody: Evidence for the formation of an acyl intermediate. *J. Am. Chem. Soc.* **116**: 6062-6069.

Hamlin, R. (1985) Multiwire area X-ray diffractometers. *Methods Enzymol.* **114**: 416-452.

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

Hartsuck, J. A., Koelsch, G., & Remington, S.J. (1992) The high-resolution crystal structure of porcine pepsinogen. *Proteins* **13**: 1-25.

Hayano, T., Sogawa, K., Ichihara, Y., Fuji-Kuriyama, Y, & Takahashi, K. (1988) Primary structure of human pepsinogen C gene. *J. Biol. Chem.* **263**: 1382-1385.

Herriott, R. M. (1939) Kinetics of the formation of pepsin from swine pepsinogen and identification of an intermediate compound. *J. Gen. Physiol.* **22**: 65-78.

Heyduk, T., Michalczyk, R., & Kochman, M. (1991) Long-range effects and conformational flexibility of aldolase. *J. Biol. Chem.* **266**: 15650-15655.

Higashi, T. (1989) The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J. Appl. Cryst.* **22**: 9-18.

Hobart, P. M., Fugliano, M., O'Connor, B.A., Schaefer, I.M., & Chirgwin, J.M. (1984) Genomic DNA sequence analysis of human renin. *Proc. Nat. Acad. Sci., U.S.A.* **81**: 5026-5030.

Holm, I., Ollo, R., Panthier, J.-J., & Rougeon, F. (1984.) Evolution of aspartyl proteases by gene duplication: the mouse renin gene is organised in two homologous clusters of four exons. *EMBO J.* **3**: 557-562.

Honig, B., & Nicholls, A. (1995) Classical Electrostatics in Biology and Chemistry. *Science* **268**: 1144-1149.

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 NOSEY data. *Protein Sci.* **4**:2289-2299.

Howard, A. J., Nielsen, C., & Xuong, N.H. (1985) Software for a diffractometer with multiwire area detector. *Methods Enzymol.* **114**: 452-472.

Huang, K. (1995) Structural studies of the interactions between serine proteinases and protein inhibitors. Ph.D. Thesis, University of Alberta, Edmonton, Alberta.

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, Gly18 variants of turkey ovomucoid inhibitor third domain with *Streptomyces griseus* proteinase B. *Protein Sci.* **4**: 1985-1997.

Huang, Q., Liu, S., & Tang, Y. (1993) The refined 1.6 Å resolution crystal structure of the complex formed between porcine β-trypsin and MCTI-A, a trypsin inhibitor of the squash family. *J. Mol. Biol.* **229**:1022-1036.

Imoto, T., Ueda, T., Tamura, T., Isakari, Y., Abe, Y., Inoue, M., Miki, T., Kawano, K., & Yamada, H. (1994) Lysozyme requires fluctuation of the active site for the manifestation of activity. *Protein Eng.* **7**: 743-748.

James, M. N. G., & Sielecki, A.R. (1986) Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* **319**: 33-38.

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

Janin, J. (1995) Principles of protein-protein recognition from structure to thermodynamics. *Biochimie* **77**: 497-505.

Janin, J. (1997) The kinetics of protein-protein recognition. *Proteins* **28**: 153-161.

Janin, J., Wodak, S., Levitt, M., & Maigret, B. (1978) Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* **125**: 357-386.

Janin, J., & Chothia, C. (1990) The structure of protein-protein recognition sites. *J. Biol. Chem.* **265**: 16027-16030.

Jentoft, J.E., Jentoft, N., Gerken, T.A., & Dearborn, D.G. (1979) 13c NMR studies of ribonuclease A methylated with [13c] formaldehyde. *J. Biol. Chem.* **254**: 4366-4370.

Jones, S., & Thornton, J.M. (1995) Protein-protein interactions: A review of protein dimer structures. *Prog. Biophys. Molec. Biol.* **63**: 31-65.

Jones, S., & Thornton, J.M. (1996) Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**: 13-20.

Jones, T. A. (1985) Diffraction methods for biological macromolecules. Interactive computer graphics: FRODO. *Meth. Enzymol.* **115**: 157-171.

Jones, T. A., Zou, J.Y., Cowan, S.W., & Kjeldgaard, D. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**: 110-119.

Kageyama, T., & Takahashi, K. (1980) Isolation of an activation intermediate and determination of the amino acid sequence of the activation segment of human pepsinogen A. *J. Biochem.* **88**: 571-582.

Kageyama, T., & Takahashi, K. (1987) Activation mechanism of monkey and porcine pepsinogens A. *Eur. J. Biochem.* **165**: 483-490.

Kageyama, T., Ichinose, M., Miki, K., Athauda, S.B., Tanji, M., & Takahashi, K. (1989) Difference of activation processes and structure of activation peptides in human pepsinogens A and progastricsin. *J. Biochem.* **105**: 15-22.

Kageyama, T., Tanabe, K., & Koiwai, O. (1990) Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequences of cDNAs, molecular evolution and gene expression during development. *J. Biol. Chem.* **265**: 17031-17038.

Kellis, J.T., Nyberg, K., & Fersht, A.R. (1989) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* **28**: 4914-4922.

Khan, A.R., Cherney, M.M., Tarasova, N.I., & James, M.N.G. (1997) Structural characterization of activation intermediate 2 on the pathway to human gastricsin. *Nature Struct. Biol.* **4**: 1010-1015.

Khan, A.R. & James, M.N.G. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Science* **7**: 815-836.

Khazanovich-Bernstein, N., Cherney, M.M., Loetscher, H., Ridley, R.G., & James, M.N.G. (1999) Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. *Nature Struct. Biol.* **6**: 15-21.

Kleywegt, G.J. (1995) Dictionaries for Heteros. CCP4/ESF-EACBM Newsletter on Protein Crystallography. pp. 45-50.

Kleywegt, G.J. & Jones, T.A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Cryst.* **D50**: 178-185.

Klinger, A. L., & Kretsinger, R.H. (1989) Latticepatch-an interactive graphics program to design data measurement strategies for area detectors. *J. Appl. Cryst.* **22**: 287-293.

Kohn, W.D., Kay, C.M., & Hodges, R.S. (1998) Orientation, positional, additivity and oligomerization-state effects of interhelical ion pairs in α-helical coiled-coils. *J. Mol. Biol.* **283**: 993-1012.

Kokesh, F.C. & Westheimer, F.H. (1971) A reporter group at the active site of acetoacetate decarboxylase. II. Ionization constant of the amino group. *J. Am. Chem. Soc.* **93**: 7270-7274.

Kraulis, P. (1991) Molscript: a program to produce both detailed and schematic plot of protein structures. *J. Appl. Cryst.* **24**: 946-950.

Lapatto, R., Krengel, U., Schreuder, H.A., Arkema, A., de Boer, B., Kalk, K.H., Wol, W.G., Grootenhuis, P.D., Mulders, J.W., Dijkema, R., Theunissen, H.J., & Dijkstra, B.W. (1997) *EMBO J.* **16**: 5151-5161.

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

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.

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

Liao, D., Breddam, K., Sweet, R.M., Bullock, T., & Remington, S.J. (1992) Refined atomic model of wheat serine carboxypeptidase II at 2.2Å resolution. *Biochemistry* **31**: 9796-9812.

Lins, L., & Brasseur, R. (1995) The hydrophobic effect in protein folding. *FASEB J.* **9**: 535-540.

Lu, W. (1994) Energetics of the interactions of ovomucoid third domain variants with different serine proteinases. Ph.D. Thesis, Purdue University, West Lafayette, Indiana.

Lu, W., Apostol, I., Qasim, M.., Warne, N., Wynn, R., Zhang, W.L., Anderson, S., Chiang, Y.W., Ogin, E., Rothberg, I., Ryan, K., & Laskowski M., Jr. (1997a) Binding of amino acid side-chains to S1 cavities of serine proteinases. *J. Mol. Biol.* **226**: 441-461.

Lu, W., Qasim, M.A., Laskowski, M., Jr., & Kent, S.B.H. (1997b) Probing intermolecular main chain hydrogen bonding in serine proteinase-protein inhibitor complexes: Chemical synthesis of backbone-engineered turkey ovomucoid third domain. *Biochemistry* **36**: 673-679.

Luzzati, V. (1952) Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Cryst.* **5**: 802-810.

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

McCammon, J.A. (1998) Theory of biomolecular recognition. *Curr. Opin. Struct. Biol.* **8**: 245-249.

McDonald, I. D., & Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**: 777-793.

McGregor, M. J., Islam, S.A., & Sternberg, M.J.E. (1987) Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**: 295-310.

McPhalen, C.A., & James, M.N.G. (1988) Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin C-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry* **27**: 6582-6598.

McPhie, P. (1972) A spectrophotometric investigation of the pepsinogen-pepsin conversion. *J. Biol. Chem.* **247**: 4277-4281.

McRee, D. E. (1993). Appendix B: Xtalview users' guide. A software system for protein crystallography. (D. E. McRee, ed.) San Diego, CA, Academic Press.

Merritt, E. A., & Murphy, M.E.P. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Cryst.* **D50**: 869-873.

Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H., & Wlodawer, A. (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* **246**:1149-1152.

Molina, M. A., Marino, C., Oliva, B., Aviles, F.X., & Quero, E. (1994) C-tail valine is a key residue for stabilization of complex between potato inhibitor and carboxypeptidase A. *J. Biol. Chem.* **269**(34): 21467-21472.

Moore, S. A., Sielecki, A.R., Chernaia, M.M., Tarasova, N.I., and James, M.N.G. (1995) Crystal and molecular structures of human progastricsin at 1.62 Å resolution. *J. Mol. Biol.* **247**: 466-485.

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

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

Nicholls, A., Sharp, K., & Honig, B. (1991) Protein folding and association: Insights from the interfacial and thermodynamic properities of hydrocarbons. *Proteins* **11**: 281-296.

Otwinowski, Z., & Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. (C. W. Carter, Jr., Sweet, R.M., eds.) pp. 307-326, New York, London, Academic Press.

Pannu, N. S., & Read, R.J. (1996) Improved structure refinement through maximum likelihood. *Acta Cryst.* **A52**: 659-668.

Papamokos, E., Weber, W., Bode, W., Huber, R., Empie, M.E., 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.

Pearson, J. P., Ward, R., Allen, A., Roberts, N.B., & Taylor, W.H. (1986) Mucus degradation by pepsin: comparison of mucolytic activity of human pepsin 1 and pepsin 3: implications in peptic ulceration. *Gut* **27**: 243-248.

Pearson, J. P., Blackburn, A., Allen, A., & Venables, C. (1990) Pepsin 5 (gastricsin): atypical pH profile of mucolytic activity. *Biochem. Soc. Trans.* **18**: 125s.

Pearson, J. P., Blackburn, A., Lynn, S., Allen, A., & Roberts, N.B. (1992) Human pepsin 1, a complex of proteoglycan and protein. *Biochem. Soc.Trans.* **20**: 356s.

Peek, K., Roberts, N.B., & Taylor, W.H. (1989) Heterogeneity of human pepsin 1, as shown by high performance ion-exchange chromatography. *J. Chromatography* **476**: 491-498.

Perlmann, G. E. (1963) The optical rotatory properties of pepsinogen. *J. Mol. Biol.* **6**: 452-464.

Perona, J.J., Craik, C.S., & Fletterick, R.J. (1993) Locating the catalytic water molecule in serine proteinases. *Science* **261**: 620-621.

Perona, J.J., Hedstrom, L., Rutter, W.J., & Fletterick, R.J. (1995) Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry* **34**: 1489-1499.

Qasim, M.A., Ding, J., James, M.N.G., Anderson, S., Ganz, P.J., Saunders, C., & Laskowski, M., Jr. (in preparation) The Lys+ side chain of succinyl Ala Ala Pro Lys p-nitroanilide is deeply inserted into the specificiy pocket of chymotrypsin in the transition state complex.

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.

Quiocho, F. A., Wilson, D.K., & Vyas, N.K. (1989) Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **340**: 404-407.

Read, F. J., Fujinage, 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.

Read, R. J. (1986a) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**: 140-149.

Read, R. J., & James, M.N.G. (1986b). Introduction to the protein inhibitors: X-ray crystallography. Proteinase Inhibitors. (A. J. S. Barrett, G., ed.). Amsterdam, Elsevier Science Publishers B.V.: 301-336.

Roberts, N. B., & Taylor, W.H. (1979) Action of human pepsins 1, 2, 3 and 5 on the oxidized B-chain of insulin. *Biochem J.* **179**: 183-190.

Roberts, N. B., Peek, K., Keen, J.N., & Taylor, W.H. (1995) Five human gastric aspartic proteinases: N-terminal amino acid sequences and amino acid composition. *Int. J. Biochem. Cell Biol.* **27**: 133-137.

Rodriguez, R., Chinea, G., Lopez, N., Pons, T., and Vriend, G. (1998) Homology modeling, model software evaluation: three related resources. *Bioinformatics* **14**: 523-528.

Samloff, I. M. (1969) Slow moving protease and the seven pepsinogens. Electrophoretic demonstration of the existence of eight proteolytic fractions in human gastric mucosa. *Gastroenterology* **57**: 659-669.

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

Scheidig, A.J., Hynes, T.R., Pelletier, L.A., Wells, J.A., & Kossiakoff, A.A. (1997) Crystal structure of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI) : engineering of inhibitors with altered specificities. *Protein Sci.* **6**: 1806-1824.

Schmid, M.F., Tronrud, D.E., & Matthews, B.W. (1983) Structural studies of a bacteriochlorophyll-containing protein. *Chem. Scr.* **21**:69-73.

Schrauber, H., Eisenhaber, F., & Argos, P. (1993) Rotamers: To be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J. Mol. Biol.* **230**: 592-612.

Shin, D.H., Song, H.K., Seong, I.S., Lee, C.S., Chung, C.H., & Suh, S.W. (1996) Crystal structure analyses of uncomplexed ecotin in two crystal forms: implications for its function and stability. *Protein Sci.* **11**: 2236-2247.

Shirley, B. A., Stanssens, P., Hahn, U., & Pace, C.N. (1992) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31**: 725-732.

Sidhu S.S., Kalmar, G.B., Willis, L., & Borgford, T.J. (1995) Protease evolution in *Streptomyces griseus*: discovery of a dimeric serine protease. *J. Biol Chem.* **270**: 7594-7600.

Sielecki, A. R., Fedorov, A.A., Boodhoo, A., Andreeva, N.S., & James, M.N.G. (1990) Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J. Mol. Biol.* **214**: 143-170.

Sielecki, A. R., Fujinaga, M., Read, R.J., & James, M.N.G. (1991) Refined structure of porcine pepsinogen at 1.8 Å resolution. *J. Mol. Biol.* **219**: 671-692.

Simmoneta, M. & Carra, S. (1969) in *The Chemistry of Carboxylic Acids and Esters*, Patai, S., Ed., Interscience, New York, pp. 1-52.

Sogawa, K., Fuji-Kuriyama, Y., Mizukami, Y., Ichihara, Y., & Takahashi, K. (1983) Primary structure of human pepsinogen gene. *J. Biol. Chem.* **258**: 5306-5311.

Sparks, D.L., Phillips, M.C., & Lund-Katz, S. (1992) The conformation of apolipoprotein A-I in discoidal and spherical lipoprotein particles. 13c NMR studies of lysine ionization behaviour. *J. Biol Chem.* **267**: 25830-25838.

Stapley, B. J., & Doig, A.J. (1997) Free energies of amino acid side-chain rotamers in α-helices, β-sheets and α-helix N-caps. *J. Mol. Biol.* **272**: 456-464.

Steitz, T.A., Henderson, R., & Blow, D.M. (1969) Structure of crystalline α-chymotrypsin. *J. Mol. Biol.* **46**: 337-348.

Steitz, T.A. & Shulman, R.G. (1982) Crystallographic and NMR studies of the serine proteases. *Ann. Rev. Biophys. Bioeng.* **11**: 419-444.

Stura, E.A. & Wilson, I.A. (1992) Seeding techniques. In *Crystallization of nucleic acids and proteins. A practical approach* (Ducruix, A. & Giege, R. eds). pp. 99-126, IRL Press, Oxford.

Suguna, K., Padlan, A., Smith, C.W., Carlson, W.D., & Davies, D.R. (1987) Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: Implications for a mechanism of action. *Proc. Natl. Acad. Sci. USA* **84**:7009-7013.

Swindells, M. B., MacArthur, M.W., & Thornton, J.M. (1995) Intrinsic φ, ψ propensities of amino acids, derived from the coil regions of known structures. *Nature Struct. Biol.* **2**: 596-603.

Taggart, R. T., Cass,L.G., Mohandas, T.K., Derby, P., Barr, P.J., Pals, G., & Bell, G.I. (1989) Human pepsinogen C (progastricsin). Isolation of cDNA clone, localization to chromosome 6, and sequence homology with pepsinogen A. *J. Biol. Chem.* **264**: 375-379.

Taggart, R. T. (1992) Genetic variation of human aspartic proteinases. *Scand. J. Clin.Lab Invest.* **52**(suppl.): 111-119.

Tame, J.R.H., Murshudov, G.N., Dodson, E.J., Neil, T.K., Dodson, G.G., Higgins, C.F., & Wilkinson, A.J. (1994) The structural basis of sequence-independent peptide binding by OppA protein. *Science* **264**: 1578-1581.

Tanford, C. (1957) The configuration of globular proteins in aqueous solution and its dependence on pH. In: Neuberger, A., ed. *Symposium on protein structure.* London: Metheun & Co., pp35-65.

Tang, J., James, M.N.G., Hsu, I.N., Jenkins, J.A., & Blundell, T.L. (1978) Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* **271**: 618-621.

Thornton, J. M., Singh, J., Campbell, S., & Blundell, T.L. (1988) Protein-protein recognition via side-chain interactions. *Bioch. Soc. Trans.* **16**: 927-930.

Tronrud, D. E., Holden, H.M., & Matthews, B.W. (1987) Structures of two thermolysin-inhibitor complexes that differ by a single hydrogen bond. *Science* **235**: 571-573.

Tronrud, D. E. (1992) Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Cryst.* **A48**: 912-916.

Waalewijen, R. A., Meuwissen, S.G.M., Pals, G., & Hoetsmit, E.C.M. (1991) Location of pepsinogens (A and C) and cellular differentiation of pepsinogen-synthesizing cells in human gastric mucosa. *Eur. J. Cell Biol.* **54**: 55-66.

Walker, V., & Taylor, W.H. (1980) Pepsin 1 secretion in chronic peptic ulceration. *Gut* **21**: 766-771.

Wendt, H., Leder, L., Harma, H., Jelesarov, I., Baici, A., & Bosshard, H.R. (1997) Very rapid, ionic strength-dependent association and folding of a heterdimeric leucine zipper. *Biochemistry* **36**: 204-213.

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

Wright, C.S., Alden, R.A., & Kraut, J. (1969) Structure of subtilisin BPN' at 2.5 Å resolution. *Nature* **221**: 235-242.

Xu, D., Lin, S.L., & Nussinov, R. (1997) Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *J. Mol. Biol.* **265**: 68-84.

Xuong, N. H., Sullivan, D., Neilsen, C., & Hamlin, R. (1985) Use of the multiwire area detector diffractometer as a national resource for protein crystallography. *Acta Crystallogr. B* **41**: 367-369.

Zhou, G.W., Guo, J., Huang, W., Fletterick, R.J., & Scanlan, T.S. (1994) The three-dimensional structure of a catalytic antibody with active site similarity to serine proteases. *Science* **265**: 1059-1064.