Title
Structural studies of alpha-lytic protease at sub-Angstrom resolution reveal insights into the mechanisms of serine protease catalysis and kinetic stability

Author
Fuhrmann, Cynthia Nichole

Publication Date
2005

Peer reviewed|Thesis/dissertation

Structural studies of alpha-lytic protease at sub-Ångstrom resolution reveal insights into the mechanisms of serine protease catalysis and kinetic stability

by

Cynthia Nichole Fuhrmann

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

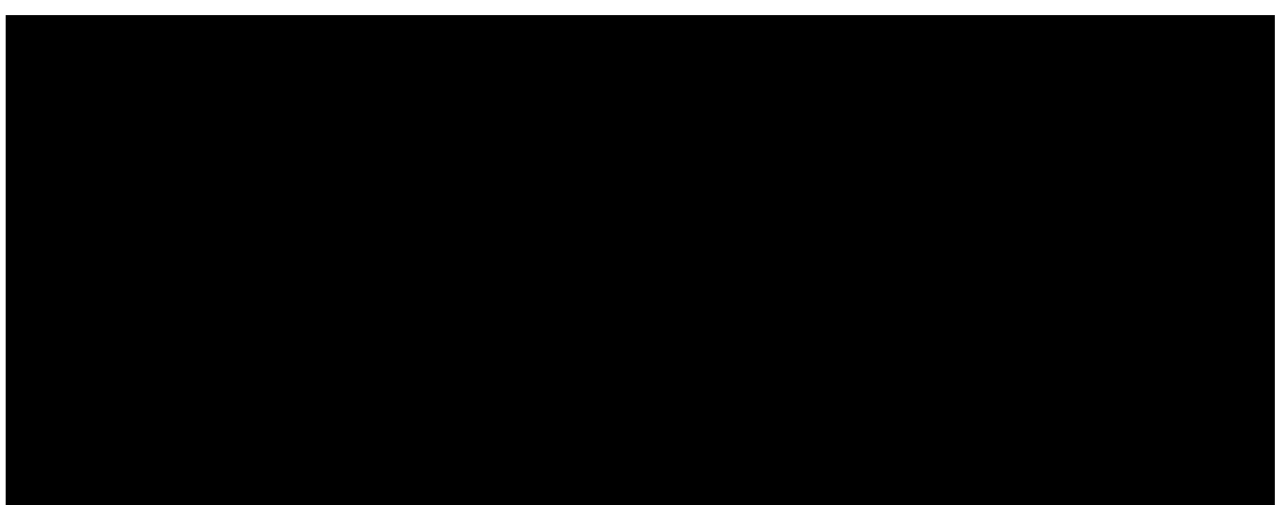
Biochemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO



Date University Librarian

Degree Conferred:.....

I dedicate this thesis to my parents

Patricia and Eckehard Fuhrmann

for their continual love and support,

whatever my endeavors may be.

Foreward

Completing a Ph.D. Thesis requires determination, skill, and luck luck luck (times 10), and the critical component: support of friends and family. And so, here I'd like to thank the people who made my eight years here at UCSF the best they could be.

Thank you to,

Dr. David Agard, my advisor, mentor, teacher, and friend, who is likely the smartest person I'll ever meet. He gave me a home in his lab and a computational project that could bring me back to my Chemistry roots. Under the light of David's brilliant ideas and excellency in science, I have strived to be the best I can be. David is a rare PI who recognizes the value of developing skills and pursuing interests that are non-scientific but parallel in importance to future careers, and in doing so he has greatly enriched my graduate career.

To my undergraduate advisor, Dr. Leonard Hjelmeland, who gave me the thoughtful guidance: "The most important skill you could learn in graduate school is how ask questions." He is the strongest-willed person that I know, and has overcome so many seemingly-insurmountable odds. Yet he always sees the light in life, as I strive to as well.

To the Agard Lab, and specifically the αLP subgroup and crystallographers (the numbers of which have grown!), for great guidance and comradeship. The Agard Lab continues to be a very happy place to be.

To my Tetrad classmates and Biophysics "honorary" classmates.

To the many people who I have had the joy of working with on *non-science* projects, including students and staff in the Graduate Students' Association, Career Center, and Student Activity Center; as well as faculty in the Program in Biological Sciences, Department of Biochemistry, School of Medicine, Graduate Division, and other professional schools, who have welcomed me into the academic administrative world not as a student, but as a colleague with a student perspective. And a huge note of appreciation to the amazingly talented graduate students and postdocs who joined me in creating the Preparing Future Faculty program at UCSF. It has been so great to work with such exciting and visionary people!

To my parents and sister, for more than I can express.

And to my husband Brian, who brings such light into my life simply with his smile. We have been great collaborators in science, and even better collaborators in life. He has been so supportive throughout the many stresses of graduate school. Marrying him was by far my greatest success in graduate school.

Finally, a special thank you to Brian (and his ex-appendix) for making the days leading up to the completion of this thesis so eventful.

Structural studies of α-lytic protease at sub-Ångstrom resolution reveal insights into the mechanisms of serine protease catalysis and kinetic stability

Cynthia N. Fuhrmann

Laboratory of David A. Agard

For many decades, α-lytic protease (αLP), an extracellular bacterial protease secreted by *Lysobacter enzymogenes,* has served as a model in both mechanistic and structural studies for chymotrypsin-like serine proteases.

In order to address questions regarding the catalytic mechanism of serine proteases, I determined three structures of αLP at ultra-high resolution: (1) the free enzyme at its active pH ($\alpha LP_{pH\,8}$; 0.83Å resolution), (2) αLP at pH 5 ($\alpha LP_{pH\,5}$; 0.82Å resolution), and (3) αLP bound to a peptidyl boronic acid inhibitor, MeOSuc-Ala-Ala-Pro-boroVal (αLP+boroVal(gol); 0.90Å resolution). The latter two structures provided analogs to the transition states of the catalytic reaction. The unexpected covalent binding of a glycerol molecule to the tetrahedral boronate in αLP+boroVal(gol) provided the most accurate and highly-resolved model of the tetrahedral intermediate for acylation (TI_1) determined to date. Structure-specific radiation damage was apparent in electron density for these structures, and special data collection techniques were used to minimize radiation damage observed at the Ser195-boronic acid adduct.

A comparison of all three structures elucidated the structural changes that occur during the first step of the catalytic reaction (ES → TI_1). We propose that upon deprotonation of His57, Ser195 undergoes a conformational change that is responsible for

preventing the TI_1→ ES back-reaction. Based on precise atomic positions obtained at sub-Ångstrom resolution, we have concluded that the His57...Asp102 interaction is a standard ionic hydrogen bond in both the free enzyme and the acylation transition state, and not a low-barrier hydrogen bond as had been previously debated. Instead, I propose that transition state stabilization by chymotrypsin-like serine proteases is achieved through a network of optimized hydrogen bonds that position the catalytic triad and stabilize the Ser195-substrate tetrahedral adduct. In particular, I identified a short, ionic hydrogen bond between His57 and the amide of the substrate leaving group that may play the primary role in catalyzing the second step of the acylation reaction.

A surprising outcome from these studies was the discovery of a site of conformational strain that appears to be evolutionarily linked to αLP's kinetic stability. Distortion of Phe228, a conserved residue in the protein core, is estimated to account for ~4 kcal/mol in conformational energy. The rearrangement and subsequent distortion of Phe228 by co-evolved residues surrounding it supports the hypothesis that tight packing in αLP is a key component of αLP's folding energy landscape. Subsequent mutational studies support this hypothesis, confirming our discovery of the first known functionally-relevant distortion.

Table of Contents

List of Tables

List of Figures

Introduction

α-Lytic protease (αLP) was first discovered in 1964 by Canadian microbiologists studying the soil bacterium, *Lysobacter enzymogenes* (Katznelson et al., 1964). Secreted into the soil, αLP's functional role is to proteolyze microorganisms for nourishment by its host bacterium (Whitaker et al., 1965). The first-known bacterial homolog of the pancreatic serine proteases, αLP has proved to be an excellent model for studying mechanisms of proteolytic catalysis and substrate specificity (reviewed in Fuhrmann et al., 2004b). Furthermore, recent studies of αLP's folding energy landscape have revealed that the protein is kinetically trapped in its thermodynamically-unstable native state (Sohl et al., 1998), creating a new paradigm for the protein folding field (reviewed in Cunningham et al., 1999).

αLP as a model for understanding serine protease catalysis

Serine proteases comprise one of the largest and most-studied families of enzymes. Together, proteases account for ~2% of the genes found in most organisms, with serine proteases being the largest family within this group (Hedstrom, 2002a). They are involved in a wide range of biological processes, including digestion, apoptosis, reproduction, signal transduction, hormone processing, viral protein processing, blood clotting, and the immune response (Hedstrom, 2002b). Consequently, serine proteases are often studied in great detail due to their potential as drug targets.

The chymotrypsin family of serine proteases, of which αLP is a member, have a characteristic double β-barrel fold, with the catalytic center located at the interface of the two domains (Figure 0.1) (Blow et al., 1969). The active site is comprised of the conserved "catalytic triad": Asp102, His57, and Ser195,[1] which together are necessary for efficient catalysis by the protease.

Figure 0.1. The structure of αLP.
The structure of αLP is shown (Brayer et al., 1979; Fuhrmann et al., 2004a; Rader & Agard, 1997), with the N-terminal domain shown in dark grey and the C-terminal domain in light grey. The catalytic triad (Asp102, His57, Ser195) is highlighted in green.

[1] Throughout this manuscript, residues are numbered based on homology to chymotrypsin (Fujinaga et al.,1985). A key relating this numbering scheme to αLP's sequential numbering is provided in Appendix A.

Crystallographic and kinetic studies have established the general mechanism by which serine proteases act (described here for the chymotrypsin family; Figure 0.2).

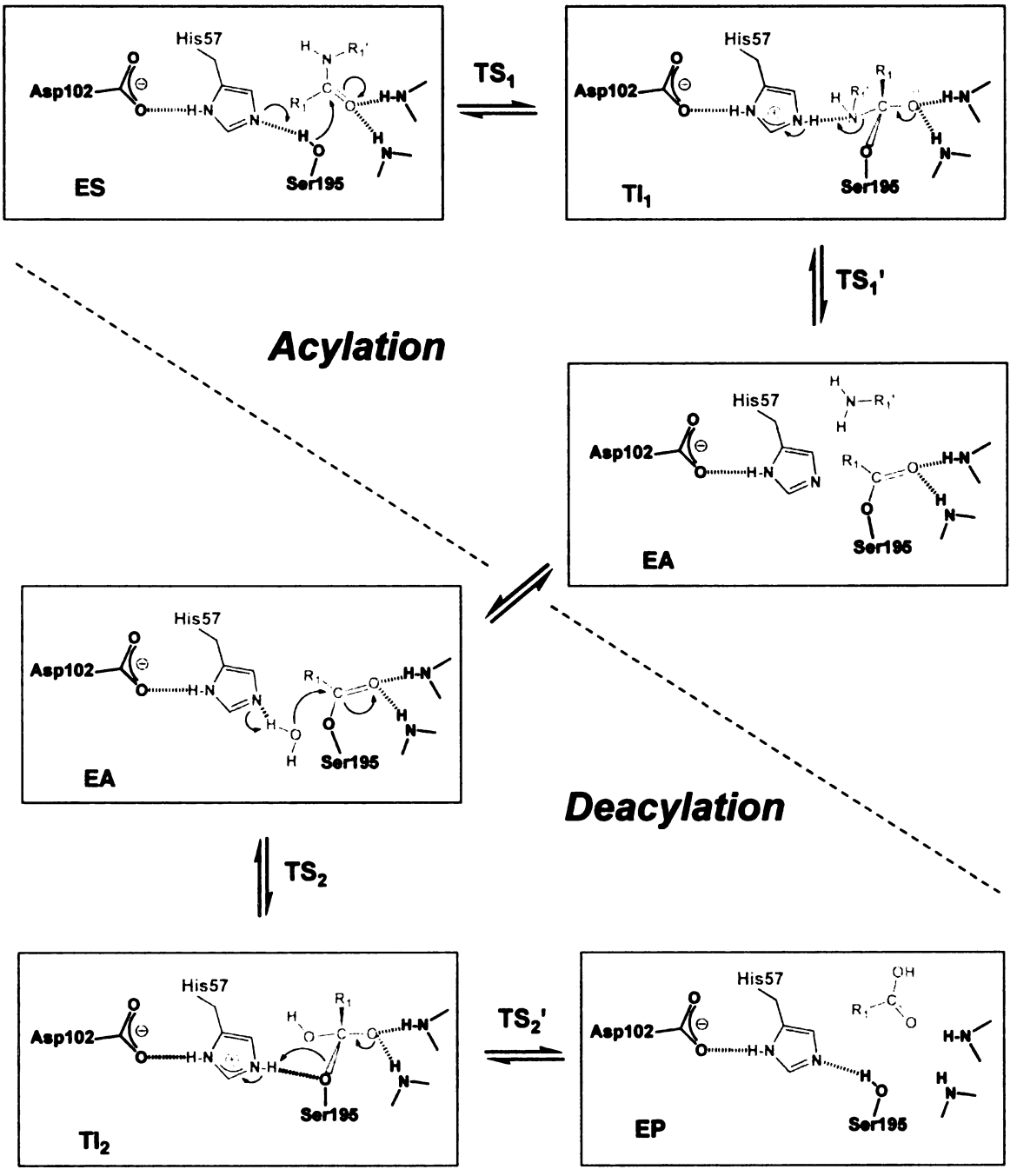


Figure 0.2. The general mechanism for catalysis by chymotrypsin-like serine proteases.

In the first stage of the catalytic mechanism (acylation), Ser195 attacks the carbonyl of the substrate's scissile bond, forming a metastable tetrahedral intermediate (TI_1). In what is proposed to be a concurrent step (Ishida & Kato, 2003), His57 deprotonates Ser195 Oγ, increasing the nucleophilicity of Ser195 and thereby promoting this first step of the reaction. Next, His57 donates a proton to the amide of the scissile bond, allowing release of the C-terminal portion of the cleaved substrate and leading to formation of the acyl enzyme (EA). The acylation reaction thereby occurs in two steps, passing through two transition states (TS_1 and TS_1') and a metastable tetrahedral intermediate (TI_1). The second stage of the reaction (deacylation) follows a similar mechanism, with a water molecule becoming the active nucleophile, passing through a second metastable tetrahedral intermediate (TI_2) prior to release of the product. According to this general mechanism, His57 is held in position via hydrogen-bonding with Asp102, the third member of the catalytic triad. Negatively-charged Asp102 is also believed to enhance the basicity of His57 by stabilizing its positively-charged state in the TS and TI states. In each step of the reaction, the backbone amides of Gly193 and Ser195 (termed the "oxyanion hole") stabilize the oxyanion of the tetrahedral adduct via hydrogen-bonding and electrostatic interactions.

Although the general mechanism of catalysis by serine proteases is understood, questions remain surrounding the precise mechanism by which the enzyme achieves transition state stabilization. Two prominent controversies have focused on the role of the His57-Asp102 hydrogen bond and, specifically, the position of the hydrogen atom between these two residues.[2] In 1969, Blow *et al.* proposed that protonation of His57 may be coupled with protonation of Asp102 in what has been called the "charge-relay", or "two-proton

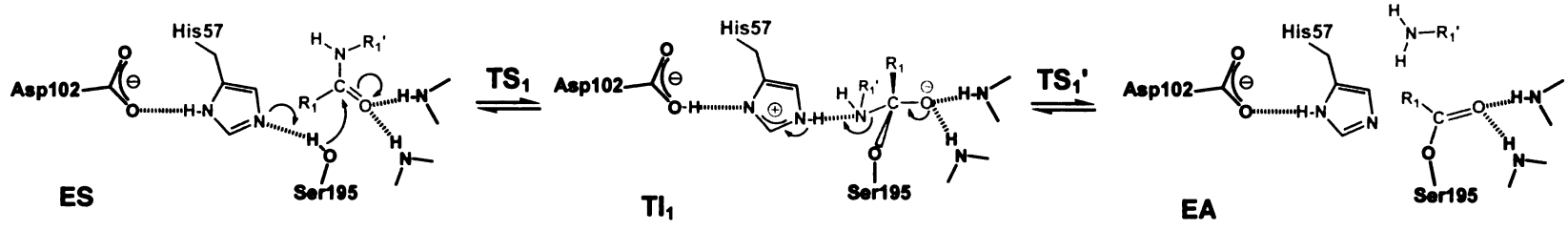
[2] This hydrogen atom has been dubbed the "most expensive hydrogen atom in NIH history." (David A. Agard, personal communication)

transfer," hypothesis (Figure 0.3a) (Blow et al., 1969; Hunkapiller et al., 1973). This proposal was controversial, as it required a large change in pK_a for both His57 and Asp102. ^{15}N NMR experiments published by multiple groups in the late 1970s and early 1980s provided evidence against the two-proton transfer hypothesis (reviewed in (Bachovchin, 1986)). In 1981, the 2.2Å resolution neutron diffraction study of trypsin bound to a transition state-mimetic inhibitor showed the hydrogen atom located on His57, and not on Asp102 (Kossiakoff & Spencer, 1980).

More recently, Frey *et al.* proposed an alternative mechanism involving the His57-Asp102 hydrogen-bond (Frey et al., 1994). Dubbed the "low-barrier hydrogen bond" hypothesis, it proposed that a low-barrier hydrogen bond (LBHB) is formed between His57 and Asp102 in the first step of the acylation reaction (Figure 0.3b). Accounting for up to ~20 kcal/mol in energy stabilization, formation of this LBHB would stabilize the transition state and increase the basicity of His57, catalyzing the acylation reaction. In the same issue of Science, a similar catalytic role by LBHB's was proposed for other enzymes, including ketosteroid isomerase, triose-phosphate isomerase, citrate synthase, orotidylate decarboxylase, mandelate racemase, aconitase, thermolysin, and carboxypeptidase (Cleland & Kreevoy, 1994).

Although the transition states for the proposed catalytic mechanism of serine proteases cannot be isolated, studies can be done on the tetrahedral intermediates (or mimics of these tetrahedral intermediates). Because TI_1 and TI_2 are close in energy to their corresponding transition states, by Hammond's postulate we can assume that their structures are also similar to those of the transition states (Hammond, 1955). However, the metastable

The Charge Relay / Two-Proton Transfer Hypothesis



The Low Barrier Hydrogen Bond Hypothesis

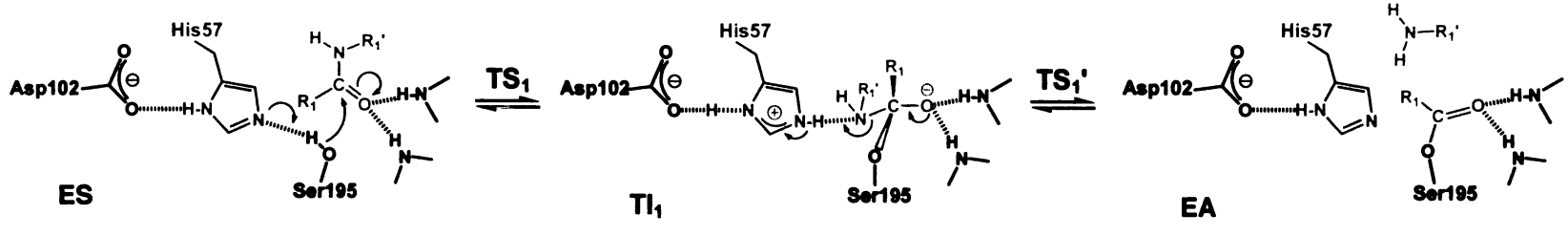


Figure 0.3. Controversies surrounding the role of the His57-Asp102 hydrogen bond.
See text for references.

nature of the tetrahedral intermediates makes their isolation difficult. In a rare crystallographic study, Wilmouth *et al.* pH-triggered deacylation of a stable acyl-enzyme complex and captured diffraction data showing tetrahedral electron density in the active site, providing the first experimental observation of the TI_2 state (Wilmouth et al., 2001). Despite these promising results, the metastable nature of TI_1 and TI_2 make it nearly impossible to study these states in any structural detail.

Most of our knowledge of the structures and chemical characteristics of the tetrahedral intermediates and transition states have been derived through studies of transition state analogs. Various natural and synthetic inhibitors have been identified that covalently bind at Ser195, forming stable adducts that mimic the tetrahedral geometry of these states. These include trifluoromethyl ketones, aldehydes, and boronic acids (which mimic deacylation transition states) and phosphonates (which mimic the acylation transition states) (Figure 0.4). While these analogs do mimic the TI and TS states to some extent, the chemical properties that stabilize their tetrahedral form also detract from being a perfect mimic. For example, in boronic acid and phosphonate inhibitors, the carbonyl carbon of the substrate is substituted with a boron and phosphorous atom, respectively. Although aldehydes and trifluoromethyl ketones do feature a carbon-centered tetrahedral adduct, the substituents on this tetrahedron have very different properties from those found in natural substrates.

Despite these caveats, transition state analogs do mimic functional characteristics of substrates, justifying their use as mimics of the TI and TS states. For example, the specificity profile for inhibition of α-lytic protease by peptide boronic acids is highly correlated with that of αLP's catalytic activity towards equivalent substrates (Bone et al., 1991a).

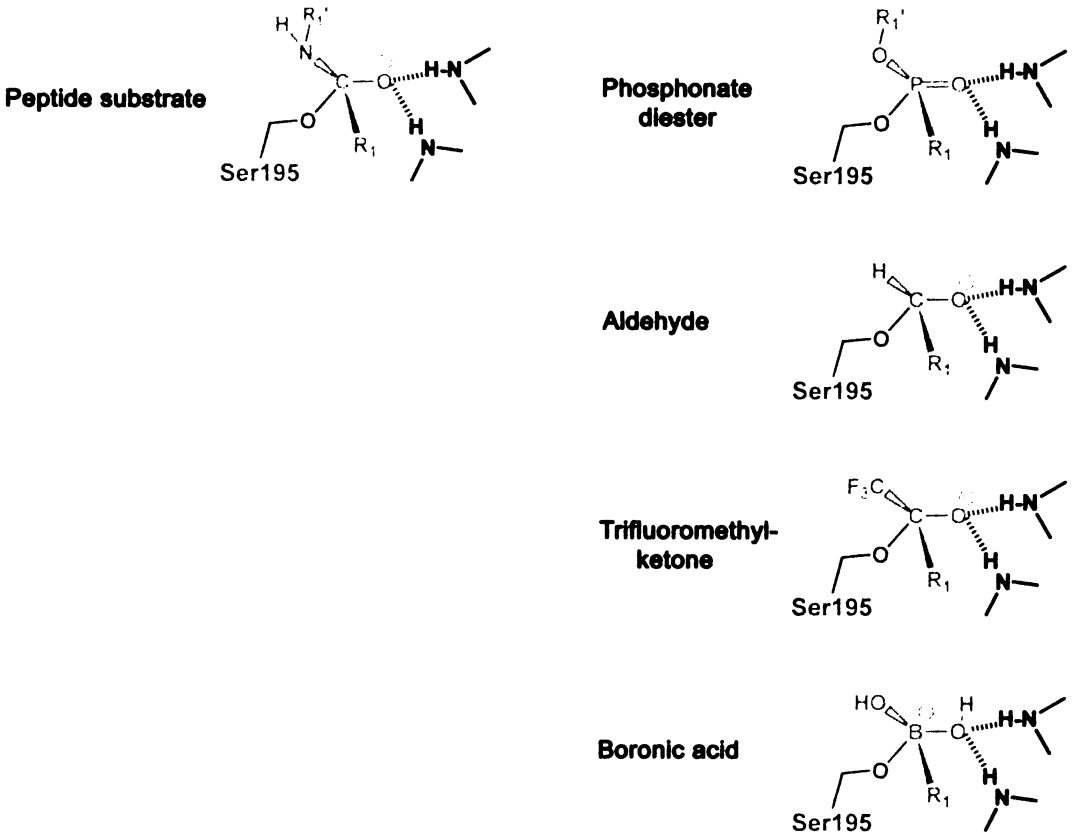


Figure 0.4. Common transition state analogs.
Peptidyl transition state analogs are shown in their predicted enzyme complex state. A standard substrate-enzyme complex is shown for reference.

Transition state analogs, therefore, provide the best opportunity to study the TI and TS states in structural detail, but interpretation of such results should include consideration of their inherent differences from natural substrates.

For decades, αLP has served as a model for studying serine protease catalysis. Initially, αLP was valued for its single histidine residue (His57), which facilitated spectroscopic studies of this important catalytic residue. As a result, there is a multitude of

structural and kinetic data available for this protease. Furthermore, as discovered in preliminary work for this thesis, αLP forms crystals that have a very high resolution limit of diffraction, allowing detailed analyses of αLP structure. This thesis describes studies of the mechanism for serine protease catalysis, utilizing ultra-high resolution diffraction data to address long-standing questions in the field (including those described above). In doing so, I solved the crystals structures of αLP in its active free-enzyme state and of αLP transition state analogs (αLP at pH 5, and αLP bound to a peptidyl boronic acid inhibitor).

αLP as a model for understanding mechanisms by which extracellular proteases achieve kinetic stability

αLP has also been studied extensively as a model for protein folding and stability. Kinetic studies of the folding and unfolding barriers for αLP revealed a surprising result: unlike its mammalian homologs, αLP's native state is thermodynamically less stable than the unfolded or intermediate states (Figure 0.5) (Sohl et al., 1998). Instead of thermodynamic stability, αLP achieves stability through a large, highly cooperative barrier to unfolding ($t_{1/2}$ =1.2 years). The price paid for this high unfolding energy barrier is an even larger energy barrier for folding ($t_{1/2}$ ~ 1000 years). As a solution to this folding dilemma, αLP is expressed with an N-terminal pro region that acts as a folding catalyst (Silen & Agard, 1989; Silen et al., 1989; Sohl et al., 1998). The pro region performs two tasks that are essential for proper folding of αLP. First, it stabilizes the folding transition state by 18.2 kcal/mol, thereby accelerating the folding process by 10^9 (Sohl et al., 1998). Second, the pro region binds

tightly to αLP (Baker et al., 1992; Peters et al., 1998), stabilizing the native conformation and thus shifting thermodynamic equilibrium towards the folded state. Digestion of the pro region by αLP releases the folded protease, which remains kinetically trapped in its native conformation (Cunningham & Agard, 2004; Sohl et al., 1998).

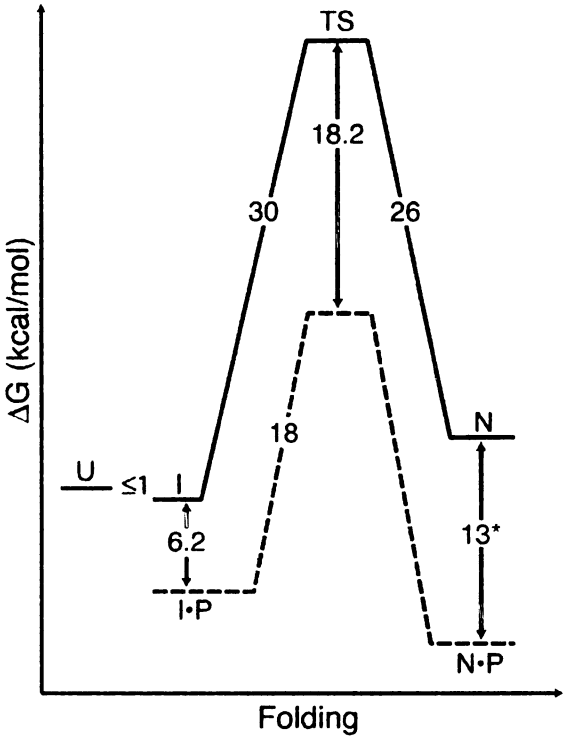


Figure 0.5. The free energy landscape for folding of αLP.
Free energy landscape for folding of αLP with (dashed line) and without (solid line) its pro region. Folding of αLP from its unfolded (U) or intermediately-folded (I) states requires co-expression with pro region (P), which catalyzes folding by stabilizing the transition state (TS). Pro region tightly binds αLP, thermodynamically stabilizing the native state (NP) relative to the intermediate state (IP). Values are in kcal/mol, and are based on experiments done at 4°C (* indicates experiments were performed at 25°C). Figure modified from Sohl, *et al.* (1998).

Kinetic analyses of homologous bacterial pro-proteases have revealed that kinetic stability may be a general property of these proteins (Truhlar et al., 2004). Like αLP, these proteases are expressed with an N-terminal pro region, varying in size from ~80 to ~170 amino acids (B. Kelch, personal communication). Comparison of αLP to a short-pro region-containing homolog, *S. griseus* protease B (SGPB), indicated that the size of the pro region correlates with the height of the energetic barriers to folding and unfolding, the stability of the native state relative to the unfolded forms, and the degree of transition state stabilization achieved by Pro (Truhlar et al., 2004).

Why evolve such a mechanism for stability? It has been hypothesized that kinetic stability is optimal for a protease such as αLP, which must survive at high concentrations in a harsh highly-proteolytic environment (Jaswal et al., 2002). Thermodynamically-stable proteins tend to have lower barriers to folding and unfolding, allowing occasional sampling of either partially or fully unfolded states. A large and highly cooperative barrier to unfolding would prevent dynamic fluctuations such as breathing motions, which would otherwise make αLP susceptible to proteolysis by neighboring proteases (Jaswal et al., 2002).

αLP does indeed display minimal structural flexibility. In crystal structures, αLP has unusually low atomic B-factors (12.7Å^2 at room temperature (Fujinaga et al., 1985); 5.4Å^2 at 120 K (Rader & Agard, 1997)). Furthermore, αLP has the highest protection factors measured by hydrogen exchange NMR to date, with ~30 amide protons displaying protection factors greater than 10^9 (Jaswal et al., 2002). Additionally, while most proteins exhibit limited flexibility primarily for a small core of residues (Li & Woodward, 1999), protection factors of greater than 10^4 were observed for more than half of αLP's amide protons. What structural characteristics of extracellular proteases provide such rigidity? Structural rigidity

is hypothesized to be the result of tight packing within the protein. αLP contains an unusually high concentration of glycines (16%, compared to an average of 9% for proteases in the chymotrypsin family), which are hypothesized to help create a rigid core by mediating tight turns in the fold of the protein (Baker et al., 1998).

The hypothesis that kinetic stability can enhance protease survival has been tested using "gladiator" assays (Jaswal et al., 2002). Briefly, samples of αLP, trypsin, and chymotrypsin were mixed together and allowed to proteolyze each other. Throughout time, samples were taken and assayed for protease activity specific to each protease, thus determining the amount of folded, active protease remaining. In this gladiator assay, αLP was >100-fold more resistant to proteoloysis than its thermodynamically stable homologs (Jaswal et al., 2002). A gladiator assay comparing survival of αLP, SGPB, and trypsin revealed similar results, with SGBP and αLP out-lasting trypsin by a factor of 8 and 20, respectively (Truhlar et al., 2004). Although trypsin has an unfolding barrier height similar to that of αLP, it was degraded much faster than αLP due to its greater native state fluctuations (Truhlar et al., 2004). This finding highlighted the importance of native state fluctuations in determining protease longevity.

While the energetics of folding are well characterized, the structural features of αLP that contribute to this unusual folding energy landscape are still unknown. Chapter 1 of this thesis describes a surprising result that came out of the ultra-high resolution structure of αLP that provided a key structural clue as to how αLP derives its kinetic stability.

The questions that guided this research were:

1. Can ultra-high resolution x-ray diffraction data be obtained for α-lytic protease?
2. What are the structural effects of x-ray radiation on αLP during collection of ultra-high resolution data?
3. How is the pK_a of His57 affected by the crystallographic environment?
4. How does the chemical structure of a protease-peptidyl boronic acid complex compare with that of a true tetrahedral intermediate?
5. Is there a low-barrier hydrogen bond between His57 and Asp102 in the acylation tetrahedral intermediate and transition states?
6. How does αLP catalyze proteolysis?
7. What are the structural features of αLP that derive its kinetic stability?

I addressed these questions by analyzing ultra-high resolution crystal structures of free αLP (solved at pH 8 and at pH 5) and αLP bound to peptidyl boronic acid transition state analogs.

Chapter 1: The 0.83Å resolution crystal structure of α-lytic protease reveals the detailed structure of the active site and identifies a source of conformational strain

Preface

I originally pursued the ultra-high resolution crystal structure of αLP with the intention of using this data to develop a method of refining multiple conformers of the protein, to be used to compare conformational sampling by the wild-type protein and a specificity mutant, Met190→ Ala, in the free and inhibitor-bound states. As I learned more about the serine protease field, I became interested in mechanisms of serine protease catalysis and, specifically, the low-barrier hydrogen bond hypothesis. These interests led to the work described in Chapters 2 and 3.

Here, I present interesting and unexpected results that came out of the $\alpha LP_{pH\,8}$ structure. The discovery of the Phe228 distortion, and its location in the core of the C-terminal domain (giving the first suggestion that this distortion may contribute to αLP's kinetic stability), catapulted me into the protein-folding subgroup of the Agard Lab. Coincident with my discovery of the Phe228 distortion, Brian Kelch performed and analyzed an alignment of bacterial proteases with short and long pro-regions. Brian discovered that a ringed residue was conserved at position 228, but that the surrounding residues co-varied with pro-region size. I performed a structural alignment of protease structures representing both families, leading to the realization that although Phe228 was conserved in sequence, its conformation in space co-varied with pro-region size. Nobuyuki Ota calculated the energy cost associated with the observed distortion of Phe228.

The work presented here was published in *Journal of Molecular Biology* (2004), volume 338, pages 999-1013. I am first author on this paper, and performed all experimental

work relating to solving this structure. Brian Kelch is the second author on this paper, having contributed to analysis of the structural results (as described above), and to the thoughtful considerations for how this distortion may affect protein stability and folding.

I dedicated this paper to Dr. Peter Kollman, a member of my Thesis Committee until his passing away in May 2001. Its inscription read, "to Dr. Peter A. Kollman, whose dynamic enthusiasm for science and life will continue to inspire all who knew him."

Summary

The crystal structure of the extracellular bacterial serine protease α-lytic protease (αLP) has been solved at 0.83Å resolution at pH 8 ("$\alpha LP_{pH\,8}$"). This ultra-high resolution structure allows accurate analysis of structural elements not possible in previous structures. Hydrogen atoms are visible, and confirm active site hydrogen bonding interactions expected for the apo enzyme. In particular, His57 Nδ1 participates in a normal hydrogen bond with Asp102 in the catalytic triad, with a hydrogen atom visible 0.83 ± 0.06 Å from the His Nδ1. The catalytic Ser195 occupies two conformations, one corresponding to a population of His57 that is doubly-protonated, the other to the singly-protonated His57. Based on the occupancy of these conformations, the pK_a of His57 is calculated to be ~8.8 when a sulfate ion occupies the active site. This 0.83Å structure has also allowed critical analysis of geometric distortions within the structure. Interestingly, Phe228 is significantly distorted from planarity. The distortion of Phe228, buried in the core of the C-terminal domain, occurs at an estimated energetic cost of 4.1 kcal/mol. The conformational space for Phe228 is severely limited by the presence of Trp199, which prevents Phe228 from adopting the rotamer observed in many other chymotrypsin family members. In αLP, the only allowed rotamer leads to the deformation of Phe228 due to steric interactions with Thr181. We hypothesize that tight packing of co-evolved residues in this region, and the subsequent deformation of Phe228, contributes to the high cooperativity and large energetic barriers for folding and unfolding of αLP. The kinetic stability imparted by the large, cooperative

unfolding barrier plays a critical role in extending the lifetime of the protease in its harsh environment.

Introduction

Secreted by the soil bacterium *Lysobacter enzymogenes*, α-lytic protease's (αLP) primary role is to lyse soil microorganisms and proteolyze their contents for nourishment of the host bacterium (Whitaker et al., 1965). αLP's unusual structural stability, evolved to maximize its lifetime in such a harsh environment, has made it a very useful model for studying how protein folding mechanisms can impact native-state properties. In addition, αLP's single histidine residue has made it an ideal system for spectroscopic experiments exploring the catalytic mechanism of serine proteases.

Numerous kinetic and structural studies have focused on αLP as a model system for understanding the catalytic mechanism of serine proteases (Fuhrmann et al., 2004b). Like other serine proteases, αLP's catalytic mechanism proceeds through two tetrahedral transition states with His57 acting as the general base and Ser195 as the primary nucleophile. These two residues, along with Asp102, form a hydrogen-bonding network called the "catalytic triad," which is conserved among members of the chymotrypsin family and has also been identified in other enzyme families. The particular nature of hydrogen bonds formed within the catalytic triad, and their energetic contributions toward catalysis, have been the cause of debate among enzymologists over several decades (see Introduction of thesis).

Extracellular bacterial serine proteases are unique to their mammalian counterparts in that they have evolved to maximize functional longevity in their harsh extracellular environments. Extensive studies of αLP have provided insight into the remarkable mechanisms of folding and stability evolved by this family of proteases. Quite surprisingly, one feature of αLP folding energetics is that, unlike its mammalian homologues, αLP's native state is thermodynamically less stable than either the fully unfolded protein or a molten globule-like folding intermediate (Figure 1.1) (Sohl et al., 1998). Instead of

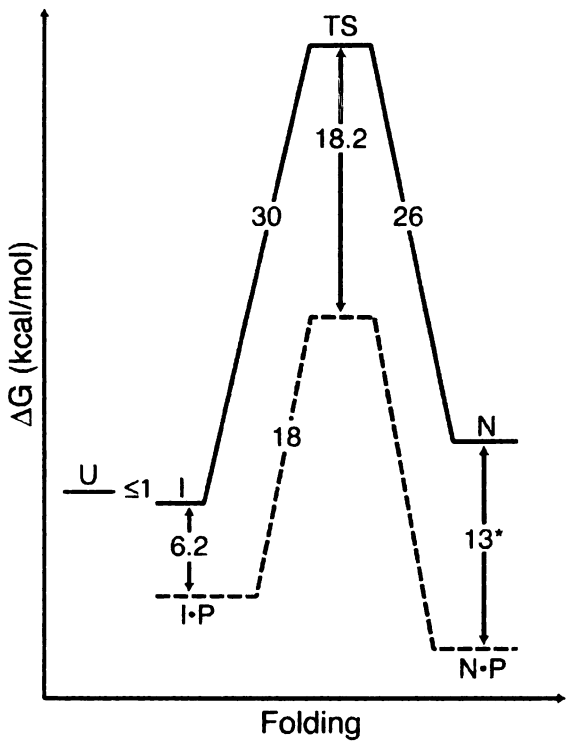


Figure 1.1. Free energy landscape for folding of αLP.
Free energy landscape for folding of αLP with (dashed line) and without (solid line) its pro region. Folding of αLP from its unfolded (U) or intermediately-folded (I) states requires co-expression with pro region (P), which catalyzes folding by stabilizing the transition state (TS). Pro region tightly binds αLP, thermodynamically stabilizing the native state (NP) relative to the intermediate state (IP). Values are in kcal/mol, and are based on experiments done at 4°C (* indicates experiments were performed at 25°C). Figure modified from Sohl, *et al.* (1998).

thermodynamic stability, αLP is stabilized kinetically, being trapped in the metastable native state by a large, highly cooperative barrier to unfolding. It has been proposed that kinetic stability provides an optimal mechanism for achieving resistance to proteolysis, and hence the longest possible lifetime under harshly proteolytic conditions (Cunningham et al., 1999). Indeed, experiments demonstrate that αLP is far more resistant to proteolysis than its mammalian homologs, trypsin and chymotrypsin (Jaswal et al., 2002). A highly cooperative barrier to unfolding and exceptional native-state rigidity prevent dynamic fluctuations such as breathing motions, which would otherwise make αLP susceptible to proteolysis. In support of this, αLP displays unusually low B-factors in crystal structures (12.7Å^2 at room temperature (Fujinaga et al., 1985); 5.4Å^2 at 120 K (Rader & Agard, 1997)), and has the highest protection factors measured by H-X NMR to date (Jaswal et al., 2002). αLP contains an unusually high concentration of glycine (16%, compared to an average of 9% for proteases in the chymotrypsin family), which is suspected of helping create a rigid core by mediating tight turns in the fold of the protein.

The negative consequence of the large barrier to unfolding is the even larger barrier to folding ($t_{1/2}$ = 1800 yr.) (Sohl et al., 1998). As a solution, αLP is expressed with a covalently-attached 166 residue N-terminal pro region that acts as a folding catalyst, accelerating folding by nearly a factor of 10^{10} (Sohl et al., 1998). In addition to stabilizing the rate-limiting folding transition state, the pro region also binds tightly to αLP (K_i = 0.32 nM), stabilizing the native conformation and thereby shifting the thermodynamic equilibrium in favor of the native state (Baker et al., 1992; Peters et al., 1998). Digestion of the pro region releases

folded, active αLP, which remains kinetically trapped in its metastable native conformation by a large barrier to unfolding ($t_{1/2}$ = 1.2 yr.) (Sohl et al., 1998).

Kinetic stability and pro region-catalyzed folding appear to be common characteristics of extracellular bacterial proteases (Truhlar et al., 2004) and are probably critical features of eukaryotic vacuolar and lysosomal proteases as well (Shinde & Inouye, 1995). A general trend has emerged that links protease resistance, folding energetics, and pro region size among chymotrypsin-like proteases. Although all αLP subfamily members contain homologous pro regions they vary in size from ~76 to ~170 amino acids (Cunningham et al., 1999). Truhlar *et al.* have characterized the folding energy landscape of a short pro region-containing protease, *S. griseus* protease B (SGPB) (Truhlar et al., 2004). Comparison of the folding energetics of SGPB and αLP suggested that both the height of the folding barrier and the thermodynamic stability of the native state, relative to the unfolded states, correlate with pro region size. Additionally, resistance to proteolysis—mediated by a large, highly cooperative barrier to unfolding for both proteases—also correlates with pro region size. However, little is known about the particular structural determinants for these proteases' unusual folding energetic landscapes.

We have solved the crystal structure of αLP at its active pH (pH 8) at 0.83Å resolution ("$\alpha LP_{pH\,8}$"). At this sub-angstrom resolution, the position of all atoms, including hydrogen atoms, can be determined accurately and precisely. This has allowed an accurate structural description of the position and hydrogen bonding of catalytic residues in the free enzyme. Additionally, we have discovered a distortion in the core of the C-terminal domain

(in the sidechain of Phe228) that we propose to be a general mechanism utilized by the large pro region-containing proteases to optimize protease resistance and protein longevity.

Results and Discussion

Data collection and refinement

The crystal structure of $\alpha LP_{pH\,8}$ was solved to 0.83 Å resolution. Data for this structure were collected at the Stanford Synchrotron Radiation Laboratory's Beamline 9-1 at 100 K using an x-ray wavelength of 0.785 Å. Crystals were grown under the same conditions as for previous αLP experiments (1.3 M Li_2SO_4, 20 mM Tris at pH 8.0) (Bone et al., 1987; Bone et al., 1989b; Brayer et al., 1979). A large crystal of approximate dimensions 1.3 x 0.45 x 0.45mm was used, allowing us to translate the beam along the crystal periodically during data collection to minimize radiation damage. Perhaps because of αLP's unusual rigidity as indicated by hydrogen exchange experiments (Jaswal et al., 2002), such high-resolution diffracting crystals are routinely obtained. In order to optimize resolution, data were collected in two passes on a single crystal: (1) a "high resolution" pass designed to capture the weaker outer reflections and (2) a 1.12Å "low resolution" pass to record data on the high-intensity lower resolution reflections (see Table 1.1). Reflections were observed at the edge of the Mar345 detector (corresponding to a resolution of 0.78 Å). Data from the two passes were scaled and merged (R_{symm}(average) = 5.9%), and resolution truncated so as to ensure a

Table 1.1. Statistics for data collection and structure refinement of $\alpha LP_{pH\ 8}$

Data Set	"High" Resolution	"Low" Resolution
Resolution Range (Å)	1.7 - 0.78	∞ - 1.12
Detector Distance (mm)	99	200
Mar plate diameter (mm)	345	345
Φ (°) Collected / ΔΦ (°)	60 / 0.5	60 / 1.0
Exposure Time (sec.)	65 - 90	5
Data Statistics		
Space Group	$P3_221$	
Unit Cell (a, c, Å)	65.9, 79.7	
Mosaicity	~ 0.4	
Limiting Resolution (Å)	0.83	
Total # Reflections	810,461	
# Unique Reflections	187,431	
I/σ(I)	28.5 (3.0)*	
% Completeness	99.6 (98.8)*	
R_{merge}**(Å)	5.9 (40.7)*	
Structure Refinement		
Resolution range (Å)	20 – 0.83	
R (# reflections in working set)	8.57% (168,627)	
R_{free} (# reflections in test set)	9.91% (18,758)	
R.M.S. bond lengths† (Å)	0.025	
R.M.S. angle distances† (Å)	0.046	
Protein residues (# residues with alternate conformations)	198 (23)	
SO_4^{2-} / glycerol molecules	4 / 2	
Water (whole/partial occupancy)	189 / 354	
Average isotropic B-factors ($Å^2$)		
Protein atoms	6.9	
Solvent atoms	22.0	

* Highest resolution bin = 0.84 - 0.83 Å.

** R_{merge} (as calculated by Scalepack (Fox & Holmes, 1966; Otwinowski & Minor, 1997)).

† As calculated by SHELXL-97 (Sheldrick & Schneider, 1997).

good signal-to-noise ratio in the highest resolution bin (Table 1.1). The data were very complete overall (99.6%) as well as in the highest resolution bin (98.8%).

αLP crystals are of space group $P3_221$, with one molecule per asymmetric unit. As calculated using the method of Matthews (Matthews, 1968), these crystals have a solvent content of 51% (v/v). Although this value is typical for most protein crystals (average 50%, v/v), it is significantly higher than that reported for other ultra-high resolution crystal structures (average 35%, v/v) (Schmidt & Lamzin, 2002).

The model was refined using the 1.5Å structure of αLP (1TAL.pdb (Rader & Agard, 1997)) as a starting structure, with 10% of the data reserved for calculation of R_{free}. Anisotropic B-factors were refined for every heavy atom, and hydrogen atoms included in the model as "riding hydrogens" (see Methods). In the latter stages of refinement, geometric restraints were released for residues having a single conformation. Improved electron density maps allowed for modeling of partially-occupied water molecules, yielding a final model with R/R_{free} of 8.57% / 9.91%. Complete refinement statistics are given in Table 1.1.

The 0.83Å structure of $\alpha LP_{pH\,8}$ is very similar to the corresponding structure at 1.5 Å resolution (1TAL.pdb (Rader & Agard, 1997)), with an RMSD of 0.071Å for main chain atoms. Higher resolution information has allowed the modeling of twice as many discrete alternate conformations, more accurate modeling of Asn and Gln side chains, and more detailed modeling of solvent molecules (as described below).

Modeling alternate conformations

αLP is a very rigid protein, with low isotropic B-factors (average 6.9 ± 3.5 $Å^2$ for all protein atoms). Alternate conformations were added only when difference maps clearly indicated their presence at contour levels above 2σ. Twenty-three residues were assigned discrete alternate conformations in $αLP_{pH\,8}$, indicating the higher level of detail visible in electron density maps at sub-angstrom resolution compared to 1.5Å resolution (where 10 residues were modeled with discrete alternate conformations) (Rader & Agard, 1997). At ultra-high resolution, secondary conformations present at less than 15% occupancy were readily visible (for example, ~13% for Cys189 and Cys220A). Additionally, some residues displayed three distinct conformations (Met190 and Arg192).

Two pairs of cysteine residues (Cys42/Cys58 and Cys189/Cys220A) occupy conformations in both the oxidized and reduced form. These pairs form disulfide bonds in their major conformation, but difference maps contoured to 3σ contained both positive and negative peaks indicative of a second conformation of these residues corresponding to their reduced form (~16% reduced for Cys42/Cys58, and ~13% reduced for Cys189/Cys220A). Both pairs of cysteines are solvent-exposed, and Cys42/Cys58 are located adjacent to the active site. Reduction of disulfide bonds—a common observation in ultra-high resolution crystal structures—has been discussed previously as a consequence of long exposures to synchrotron radiation (Burmeister, 2000; Ravelli & McSweeney, 2000; Schroder Leiros et al., 2001; Weik et al., 2000). It has been shown that radiation damage occurs predominantly to residues in the vicinity of water molecules or enzymatic active sites, as seen in our structures. However, our structure shows no evidence of other forms of radiation damage,

such as decarboxylation of Asp or Glu residues or the loss of methylthio groups from Met residues.

Modeling solvent

In the final stages of refinement, the large majority of difference electron density peaks (contoured at 3σ) appeared in the solvent region of the map. Several partially-occupied waters were modeled into these peaks, placed manually or using the peak search algorithms found in SHELX-Pro (Sheldrick & Schneider, 1997) and QUANTA . The final model contains 189 fully- and 354 partially occupied water molecules. Although a detailed analysis of water structure has yet to be pursued, it is evident that pentagonal and hexagonal networks of waters—in some cases overlapping multiple networks—exist in αLP crystals, as described previously for crambin (Teeter et al., 2001).

In addition to waters, four sulfate ions are present in the structure. One of these sulfate ions (Sul202) lies in the vicinity of the active site (see below). Glycerol molecules are also present in the structure. In preparation for mounting under the cryostream, the crystal was looped through cryosolvent containing precipitant solution and 15% glycerol. In the 2-10 second interval when the crystal was coated in cryosolvent prior to flash-freezing, glycerol molecules penetrated the lattice in high occupancy (up to ~80%). The final model of $\alpha LP_{pH\,8}$ contains two glycerol molecules per asymmetric unit. One of these glycerol molecules (Gol210) occupies the S2 pocket, mediating crystal contacts to a neighboring protein molecule. This high occupancy glycerol was observed in previous αLP structures

solved at cryogenic temperatures, including the 1.5Å αLP structure, where it was mistaken as a Tris molecule (Derman & Agard, 2000; Rader & Agard, 1997).

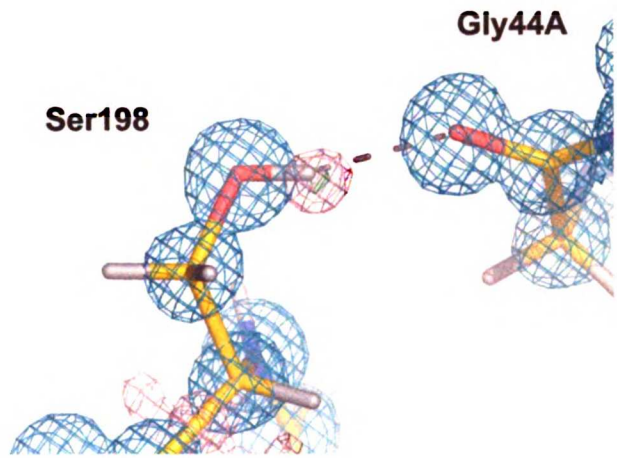
Structural Features at Ultra-High Resolution

Several unique features of the electron density are apparent in this structure at ultra-high resolution. Contoured at 1.5σ, σ_A-weighted $2F_o$-F_c electron density appears bulbous; at higher contour levels atoms are discreet, appearing like beads on a string (Figure 1.2). As recognized by Kuhn, *et al.* for the 0.78Å resolution crystal structure of subtilisin (Kuhn et al., 1998), carbon, nitrogen, and oxygen atoms can be easily distinguished by their electron density in σ_A-weighted $2F_o$-F_c maps contoured to 4σ. The appropriate orientation of Asn and Gln residues is unambiguous at these resolutions. High B-factors and difference electron density maps led us to flip Asn101, Asn166, and Gln182 relative to their positions in the 1.5Å resolution structure of αLP. Double bonds can be readily differentiated from single bonds by their characteristically shorter bond lengths when refined without geometrical restraints. In several cases it is possible to see asymmetries in the $2F_o$-F_c electron density that extend between atoms in π-bonding systems.

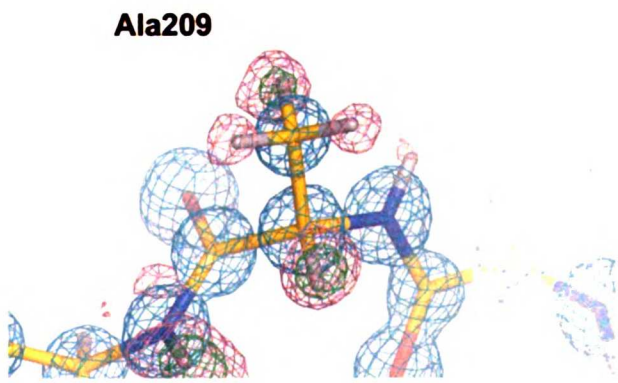
As seen in several other structures at atomic resolution (Schmidt & Lamzin, 2002), hydrogen atoms were clearly visible in difference electron density maps prior to their addition to the model. Hydrogen atoms were added to the model fairly early in the refinement because of their prominence in the electron density above the noise level. For instance, significant σ_A-weighted F_o-F_c peaks were visible above 2σ for hydrogens on ~85% of the Cα atoms and ~65% of the backbone amides, with the majority of these signals also

Figure 1.2. Structural features at ultra-high resolution.
At 0.83Å resolution, atom types can be distinguished. Electron density is more continuous along π-bonds than for σ-bonds. Hydrogen atoms are visible in difference electron density maps. The position of hydrogen atoms is unambiguous for uniquely-ordered (a) hydroxyl groups, (b) methyl groups, and (c) waters 449 and 450. σ_A-weighted $2F_o$-F_c electron density is contoured at 2σ (cyan), and σ_A-weighted F_o-F_c density at 2σ (pink) and 3σ (green). The difference electron density maps displayed are chosen from the appropriate stages of refinement: (b) shows σ_A-weighted F_o-F_c maps calculated for αLP prior to modeling any hydrogen atoms, (a) shows σ_A-weighted F_o-F_c maps calculated prior to addition of hydroxyl hydrogens to the model, and (c) shows σ_A-weighted F_o-F_c maps calculated from the final model. In all figures, the final model (including hydrogen atoms) is represented by sticks.

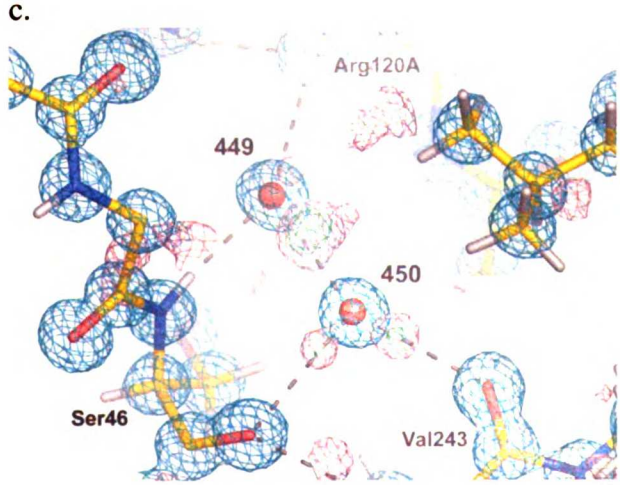
a.



b.



c.



visible at 3σ. Peaks of appropriate geometry were also visible on side chains, including polar hydrogens on several Ser, Thr, and Tyr residues. These polar hydrogens are unambiguously held in position by hydrogen bonding interactions (Figure 1.2a). Several methyl hydrogens were also visible as distinct triplicate peaks (Figure 1.2b). Hydrogen atoms were visible on Gol210 in the S2 pocket, and were included in the model for this solvent molecule. Hydrogen atoms were also observed on water molecules uniquely positioned by hydrogen bonding interactions (Figure 1.2c). Unique ordering of methyl groups and water molecules has also been observed in other structures solved at atomic resolution (Katona et al., 2002; Kuhn et al., 1998; Minasov et al., 2002).

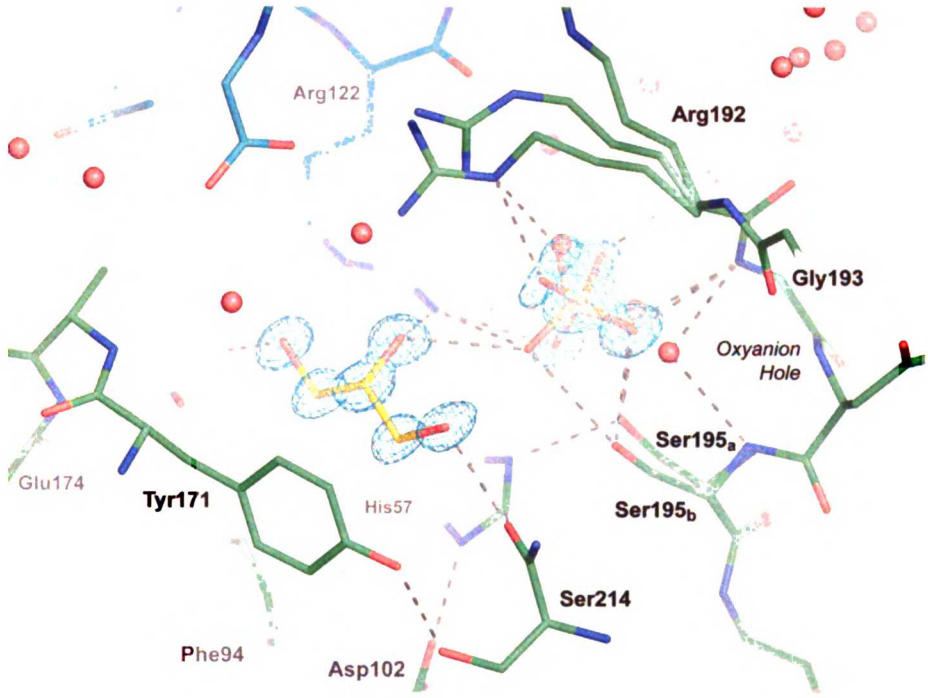
The Active Site: The protonation state of His57

As seen in previous structures, a sulfate ion is bound near the active site (Brayer et al., 1979; Fujinaga et al., 1985; Rader & Agard, 1997). In the 0.83Å structure of $\alpha LP_{pH\,8}$, this sulfate (Sul202) is present in two conformations (Figure 1.3). Using a novel R_{free} refinement approach, we reliably determined the total occupancy of Sul202 to be 77 ± 2% (see Methods and Figure 1.4). This corresponds to a $K_{d,\,\alpha LP\bullet Sul}$ of 0.38 ± 0.04 M for our crystal (grown in 1.3M Li_2SO_4). Both conformations of Sul202 hydrogen-bond to Ser195 Oγ and to Gly193 N in the oxyanion hole. Sul202 is further stabilized by ionic interactions with Arg122 (through a crystal contact) and Arg192. In the absence of Sul202, three water molecules maintain hydrogen-bonding interactions with Ser195, Gly193, and Arg192. The close proximity of Sul202 to the active site may be a direct cause of the pK_a perturbation observed in αLP crystals (see below) (Smith et al., 1989).

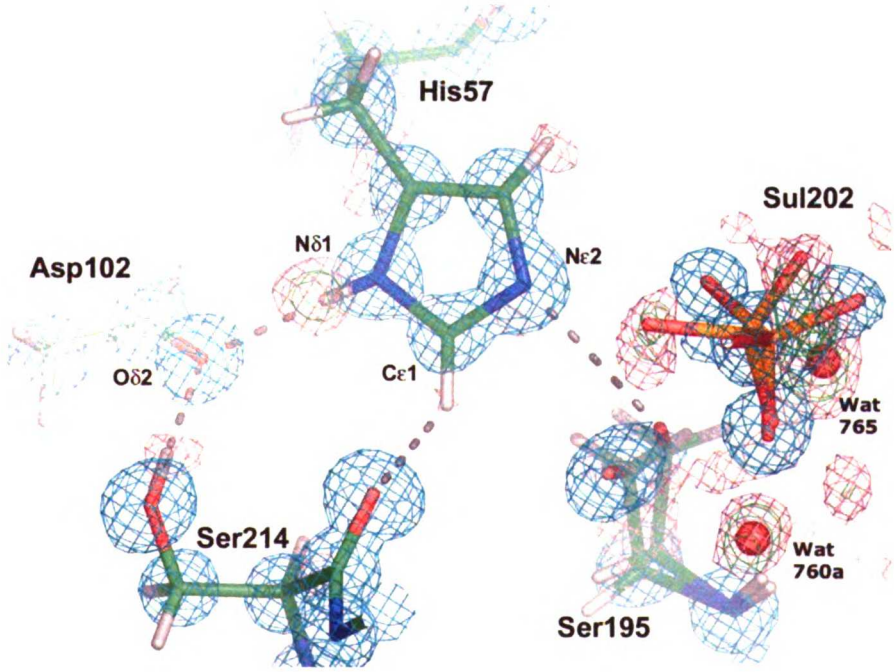
Figure 1.3. The active site of $\alpha LP_{pH\ 8}$.
(a) Two large solvent molecules, a sulfate ion (Sul202) and a glycerol molecule (Gol210), are present near the active site. Sul202 hydrogen-bonds with Ser195, Gly193 in the oxyanion hole, and is electrostatically stabilized by nearby Arg122 and Arg192. Gol210 occupies the S2 binding pocket of αLP, which is bordered by Tyr171, Phe94, and His57. For simplicity, $2F_o$-F_c electron density (contoured at 2σ) is shown only for Sul202 and Gol210. (b) Ser195 occupies two conformations, with only the minor conformation ($Ser195_a$ Oγ) in ideal geometry for hydrogen bonding with His57. σ_A-weighted F_o-F_c electron density peaks are visible at 2σ (pink) for all His57 imidazole hydrogens except Nε2 (which is only partially protonated at pH 8 in the apo-enzyme). His57 Nδ1 and Asp102 Oδ2 interact via a standard hydrogen bond. The final model is represented by sticks and water molecules displayed as red spheres. The hydrogen atoms on the His57 imidazole ring and on Ser214 Oγ were refined with no restraints. σ_A-weighted $2F_o$-F_c electron density maps are drawn at 2σ (cyan). σ_A-weighted F_o-F_c maps, calculated prior to the addition of His57 and polar hydrogen atoms to the model, are contoured at 3σ (green) and 2σ (pink).

(Figure 1.3, see caption on previous page)

a.



b.



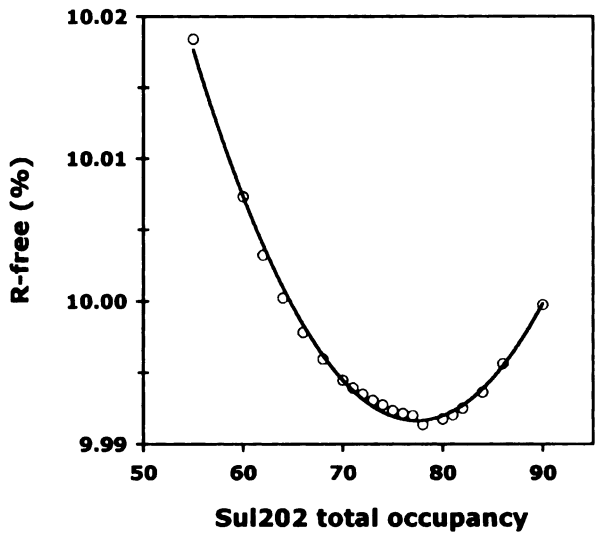


Figure 1.4. Using R_{free} to determine the total occupancy for Sul202 in $\alpha LP_{pH\ 8}$. R_{free} plotted as a function of the total occupancy assigned for Sul202 conformations. The input file for each refinement trial was identical, except that the occupancies of $Sul202_a$ and $Sul202_b$ were fixed to values whose sum equaled the "total occupancy" shown on the x-axis (with the ratio of $Sul202_a$:$Sul202_b$ kept the same for each trial). These data were fit to a poynomial, and its minimum value ($77.4 \pm 2.1\%$) taken to be the total occupancy of Sul202 used in subsequent refinement cycles. (See Materials and Methods for a more thorough description.)

In the first step of the catalytic mechanism, His57 acts as a general base, abstracting a proton from the catalytic Ser195. His57 occupies a single, well-ordered conformation in this free enzyme structure. However, Ser195 adopts two conformations (Figure 1.3). One conformation ($Ser195_a$; ~30% occupancy) forms the canonical hydrogen bond to His57 Nε2. In the second conformation ($Ser195_b$; ~70% occupancy), Oγ is displaced by ~1Å, no longer forming this hydrogen bond due to bad angular geometry (Table 1.2). Interestingly, this conformation is essentially identical to that observed in structures of αLP complexed with transition state mimetic inhibitors (Bone et al., 1987; Bone et al., 1989a; Bone et al., 1989b;

Bone et al., 1991a; Mace & Agard, 1995), suggesting that it represents the position of Ser195 when His57 is doubly-protonated. This agrees with difference electron density peaks corresponding to hydrogen atoms around the His57 ring. As expected, distinct σ-weighted F_o-F_c peaks are visible at contour level 2σ off atoms Nδ1, Cε1, and Cδ2 of the histidine ring (see Figure 1.3b); in addition, an electron density peak corresponding to Nε2-H becomes visible at contour levels at or below 1.5σ (data not shown), providing supporting evidence that a population of His57 is protonated at this site. However, the 70%/30% ratio of doubly-protonated His57 is a higher proportion than would be expected for an unperturbed histidine (pK_a ~ 7) at pH 8. Therefore, the His57 pK_a must be higher in our crystal structure. As seen in other crystal structures of αLP, a sulfate ion (Sul202) is present in the enzyme active site. The negative charge of Sul202 should stabilize the positive charge of doubly-protonated His57, raising its pK_a. At ultra-high resolution, we can quantify this effect (see Methods). Indeed, the overall pK_a for His57 in our crystals (which were grown in 1.3M Li_2SO_4) is ~8.4. This agrees with the value determined by Smith *et al.* (pK_a ~ 7.9) by solid-state NMR on αLP crystals grown in 0.6M Li_2SO_4 (see below) (Smith et al., 1989). A similar shift in pK_a was also observed for αLP in a solution of Li_2SO_4, suggesting that it is the presence of Li_2SO_4, and not an artifact of the crystalline state, that perturbs the pK_a (Farr-Jones, 1989).

Using the relative occupancies of Sul202 and the His57 protonation states determined in our crystal structure, we calculated that the subset of αLP molecules with Sul202 bound in the active site have a pK_a of 8.8 ± 0.1 (see Methods and Scheme 1). This corresponds to a calculated overall pK_a shift to 8.1 ± 0.8 for crystals if soaked in 0.6M Li_2SO_4, which agrees (within error) with the value determined experimentally by solid-state NMR on αLP crystals grown under these conditions (Smith et al., 1989). Our calculations suggest that it is the

presence of SO_4^{2-} ions, and not an artifact of the crystalline state of the protein, that causes the shift in pK_a observed for αLP crystals.

Hydrogen Bonding in the Active Site

Perhaps the most discussed hydrogen bond in the serine protease field is that between His57 and Asp102 of the catalytic triad. It has been proposed that this hydrogen bond becomes shorter and stronger in the transition state of the reaction, providing a significant portion of the energy required for catalysis. In this structure of the apo enzyme, His57 forms a standard hydrogen bond with Asp102 (2.770 ± 0.005 Å distance between His57 Nδ1 and Asp102 Oδ2; see Table 1.2). Prior to the addition of hydrogens to the model, a peak in the σ_A-weighted F_o-F_c difference electron density map off His57 Nδ1 was especially strong (Figure 1.3b). A hydrogen atom was added to the model, and its position refined against the data. The Nδ1—H bond length refined to 0.83 ± 0.06Å, which is close to the canonical bond length for atoms participating in a standard hydrogen bond. These data confirm that a standard hydrogen bond (and not a low-barrier hydrogen bond, as has been proposed for the transition state of the reaction (Frey et al., 1994)) exists between His57 and Asp102. This is corroborated by NMR data for the apo enzyme (Ash et al., 1997; Bachovchin, 1985).

Table 1.2. Hydrogen-bonding geometries for $\alpha LP_{pH\,8}$.

Interaction (D...A)	*d*(D...A) (Å)	*d*(D-H) (Å)	D-H...A (°)
Gly193 N ... Sul202 $O1_a$	2.820(9)		162(5)
Gly193 N ... Sul202 $O3_b$	2.85(3)		154(5)
Gly193 N ... $Wat760_a$	2.90(3)	1.37(6)	142(4)
Gly193 N ... Wat765	2.78(4)		115(4)
Ser195 N ... $Wat760_a$	2.96(3)	*n/d*	141
His57 Cε1 ... Ser214 O	3.033(6)	1.00(6)	126(3)
Ser214 Oγ ... Asp102 Oδ2	2.698(4)	0.98(7)	174(6)
His57 Nδ1 ... Asp102 Oδ2	2.770(5)	0.83(6)	170(2)
His57 Nε2 ... Ser195 $O\gamma_a$	2.826(15)	*n/d*	139
His57 Nε2 ... Ser195 $O\gamma_b$	2.962(5)		120

D-H bond lengths are stated in cases where the position of the hydrogen atom was refined without restraints. *n/d* indicates that the position of the hydrogen atom was not refined against the data. The value provided for D-H...A angle geometry in these cases is based on a hydrogen atom whose position was fixed at optimal geometry by SHELXL-97.

In a 1994 comparative study of serine protease structures solved to date, Derewenda *et al.* noted that in all structures, the carbonyl of Ser214 lay in close proximity to His57 Cε1, fulfilling the requirements of an unusual C-H...O hydrogen bond (Derewenda et al., 1994). These authors postulated that a weak hydrogen bond between His57 Cε1 and Ser214 O helps to delocalize the positive charge of the histidine ring, thereby weakening the Nε2-H bond and promoting breakdown of the tetrahedral intermediate to the acyl enzyme complex. This C-H...O bond is also present in our ultra-high resolution structure of $\alpha LP_{pH\,8}$. In $\alpha LP_{pH\,8}$, the donor-acceptor distance (His57 Cε1 ... Ser214 O) is 3.033 ± 0.006 Å (Table 1.2); the hydrogen—acceptor distance (His57 Cε1-H ... Ser214 O) is 2.33 ± 0.05Å, which agrees with the average value (2.3 ± 0.14Å) observed by Derewenda *et al.* in the 22 serine protease structures they analyzed. A similar hydrogen bond was also observed in the 0.95Å resolution

structure of an acyl enzyme complex, porcine pancreatic elastase bound to a peptide carboxylate (Katona et al., 2002).

In the tetrahedral transition states, the backbone amides of Gly193 and Ser195 participate in hydrogen bonding interactions with the substrate, stabilizing the transient oxyanion. In our free enzyme structure, Gly193 N hydrogen-bonds with the sulfate ion located in the active site, Sul202, with a standard hydrogen-bond distance of 2.83Å (Table 1.2). A distinct peak was present off Gly193 N in the σ_A-weighted F_o-F_c map contoured at 3σ. When the Gly193 N hydrogen was added to the model at the fixed canonical position as a "riding hydrogen," some of this F_o-F_c density remained, located at a position just beyond the hydrogen atom. Refinement of Gly193 N-H without restraints resulted in the loss of this F_o-F_c density, and yielded an N-H bond distance of 1.37 ± 0.06 Å. Although such extreme delocalization of a hydrogen atom is suggestive of a short strong hydrogen bond, the donor-acceptor distance (2.83Å) is too long to allow such a bond. Ser195 N-H, which only hydrogen-bonds with one partially-occupied water molecule ($Wat760_a$), was visible in difference maps only when contoured at or below 2σ.

Geometric Distortions: A Role in Protein Stability

At 0.83Å resolution, the data is presumably a better indication of atomic positions and bond geometries than are empirical force fields. Therefore, in order to determine the most accurate structure possible, all geometric restraints were released for residues with a single conformation during the final stages of refinement. Refinement was well behaved in the absence of geometrical restraints, with atomic positions remaining essentially the same

after 10 cycles of refinement (0.008Å RMSD). During subsequent rounds of refinement without restraints, main chain bond lengths and angles acquired a broader sampling than in earlier, restrained models. In a comparison of the last fully-restrained model (R/R_{free} = 10.77% / 11.86%) and the final model (R/R_{free} = 8.57% / 9.91%), RMS deviations from ideal values increased from 0.015Å to 0.025Å for bond lengths and from 0.031Å to 0.046Å for angle (1, 3)-distances (as defined by SHELXL-97 (Sheldrick & Schneider, 1997)).

The final model was checked for unusual geometries using PROCHECK (Laskowski et al., 1993). All residues lay well within the allowed regions of the Ramachandran plot, with only one amino acid, Ala39, in the generously allowed region. This distortion is most likely a crystallographic artifact (Ala39 participates in hydrogen bonding both in a crystal contact and in a β-hairpin). While the vast majority of residues maintained standard geometries even in the absence of restraints, one particular amino acid--Phe228--was clearly distorted from planarity. To our knowledge, such an observation has not previously been discussed for high-resolution structures of macromolecules.

The distortion of Phe228 is distributed throughout the sidechain, with a Cα-Cβ-Cγ angle deformed from tetrahedral (117.4 ± 0.3°), and further distortion in the phenylalanine ring itself. The degree of deformation from planarity increases through the ring from Cγ to Cζ. The angle measured between the Cβ-Cγ and the Cγ-Cζ vectors is 5.8° (Figure 1.5b). Phe228 is very well-ordered in the crystal: electron density is clean (Figure 1.5b), atomic ESD's are low (average 0.0094 ± 0.0009 Å for Phe228 sidechain heavy atoms, compared to 0.025 ± 0.034 Å average ESD for all sidechains in the structure), and atomic B-factors are low (average is 5.3 ± 0.6 $Å^2$ for sidechain carbon atoms, compared to 7.9 ± 4.2 $Å^2$ for all

$\alpha LP_{pH\,8}$ sidechain atoms). Furthermore, the existence of this distortion even prior to the release of geometrical restraints supports the validity of the observed conformation.

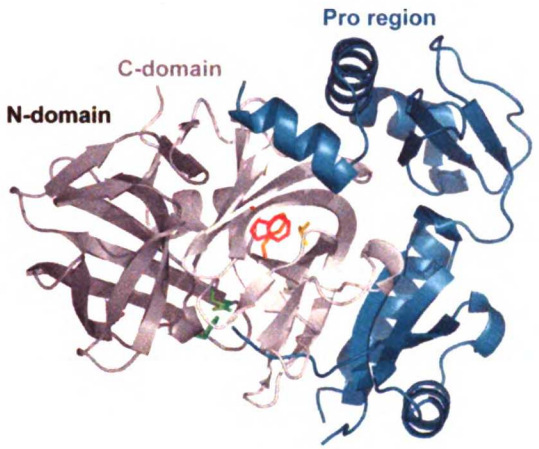
Phe228 is located in the core of the C-terminal domain (Figure 1.5a), wedged between the sidechains of Thr181 and Trp199. A surface representation of residues surrounding Phe228 suggests that the primary cause of Phe228's distortion is a steric interaction with Thr181 (Figure 1.5c), resulting in the deformation of the Phe228 bond angle at Cβ, and the progressive increase in deformation from Cγ to the tip of the Phe228 ring (Cζ). Quantum mechanical calculations performed on the isolated side chain of Phe228 *in vacuo* indicate that the observed deformation from ideality comes at a significant energetic cost of ~4.1 kcal/mol. Interestingly, numerous lines of evidence have indicated that the C-terminal domain of αLP plays a central role in the origin of αLP's unusual folding energetic landscape, suggesting that the Phe228 distortion is of functional significance.

Figure 1.5. Tight packing in the C-terminal domain distorts Phe228.
Phe228 (orange) is distorted from planarity by nearby Trp199 (red) and Thr181 (yellow). (a) The Pro-αLP complex (4PRO.pdb) (Sauter et al., 1998). αLP is a double-beta barrel structure consisting of N-terminal (dark grey) and C-terminal (light grey) domains. The Pro region (teal) catalyzes folding of αLP, with most Pro-αLP interactions occurring with αLP's C-terminal domain. Interestingly, the distorted Phe228 is located in the core of the C-terminal domain. For reference, the catalytic triad (Asp102, His57, Ser195) is displayed in green. (b) illustrates the distortion of Phe228. The σ_A-weighted $2F_o$-F_c electron density map is displayed for Phe228 at contour level 2σ (cyan). The angle between vectors Cβ→Cγ and Cγ→ Cζ is indicated. (c) van der Waals surfaces for a selection of residues within 5Å of Phe228 show the tight packing in this region.

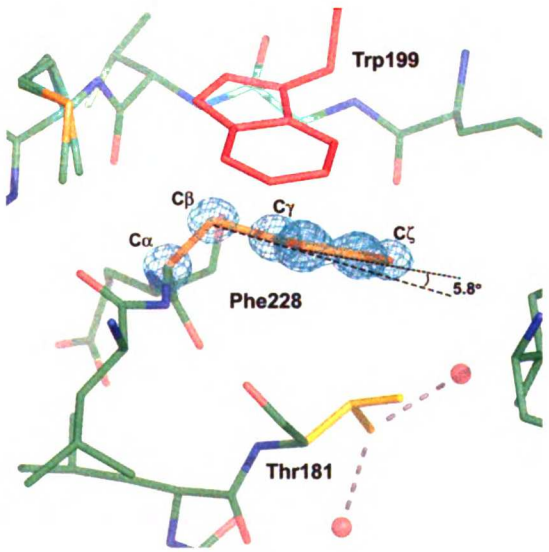
(Figure 1.5 shown on next page.)

(**Figure 1.5**, caption on previous page)

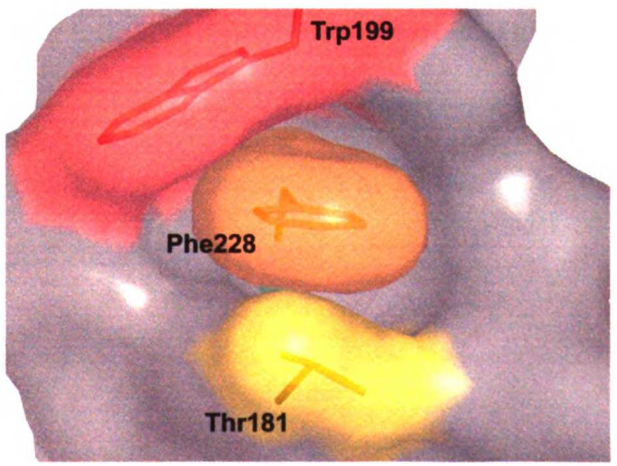
a.



b.



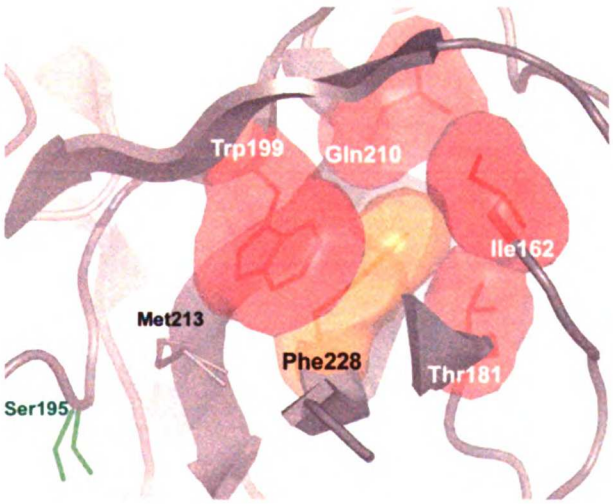
c.



In order to gain a better understanding of the particular residues contributing to the folding and unfolding energetics of bacterial serine proteases, we aligned the primary sequences of several proteases, and compared their sequences based on pro region size. Although position 228 is conserved among chymotrypsin family members to be Phe, Tyr, or His, this amino acid is surrounded by a cluster of residues— Ile162, Thr181, Trp199, and Gln210 in αLP—whose identities correlate strongly with pro region size (Figure 1.6). Interestingly, structural alignments indicate that, although conserved, residue 228 occupies one of two conformations in these proteases depending on the type of amino acid present at position 199 (Figure 1.6c). Bacterial proteases requiring only a short pro region for folding, and mammalian proteases (whose folding is pro region-independent), contain non-aromatic hydrophobic residues at position 199 (Figure 1.6b). Atomic-resolution structures of pancreatic porcine elastase (1GVK.pdb (Katona et al., 2002) and 1QNJ.pdb (Wurtele et al., 2000)) and bovine trypsin (1HJ9.pdb (Schroder Leiros et al., 2001)) show that in these proteases, Phe228 is undistorted. However, in long pro region-containing proteases such as αLP, large cyclic residues at position 199 severely limit the rotamer choices allowed for Phe228, restricting it to an alternate position where bad steric interactions with Thr181 cause deformation of the Phe228 sidechain (Figure 1.6c). Interestingly, the 1.5Å resolution crystal structure of *S. griseus protease E* (SGPE; 1HPG.pdb (Nienaber et al., 1993)), a protease expressed with a mid-sized pro region, shows that in this protease His199 similarly positions His228 against Val181. Although not commented on in the literature, SGPE His228 also demonstrates a significant distortion from planarity.

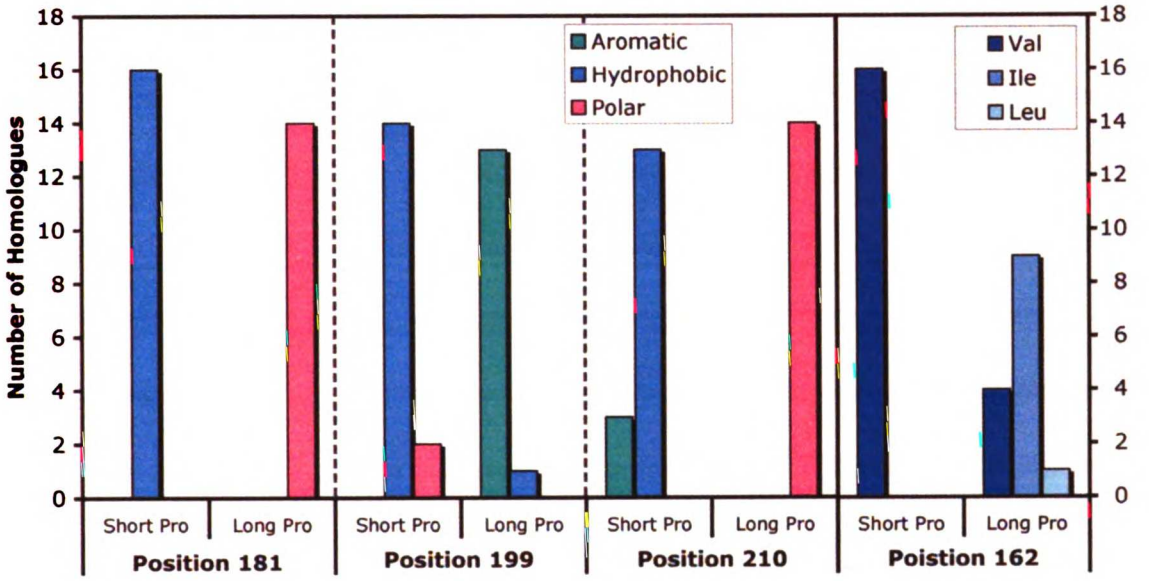
Figure 1.6. A cluster of residues surrounding Phe228 co-vary with pro region size. Sequence alignment of thirty serine proteases, containing pro regions ranging in length from zero to 175 amino acids, identified a cluster of 4 residues adjacent to residue 228 that co-vary with pro region size. (a) These residues in $\alpha LP_{pH\,8}$ (Ile162, Thr181, Trp199, and Gln210, in red; Phe228 is orange) are displayed using surface representations. Met 213 (grey) and Ser195 (green), which do not exemplify co-variance with pro region size, are displayed for reference. (b) describes the distribution of amino acid types for these residues in short (< 140 amino acids) and long (155-175 amino acids) pro region-containing proteases. (c) Structural alignment of representative chymotrypsin family members containing different-length pro regions: elastase (no pro region, blue (1HAX.pdb) (Wilmouth et al., 2001)), *S. griseus* protease B (short pro region, magenta (3SGB.pdb) (Read et al., 1983)), *S. griseus* protease E (mid-sized pro region, green (1HPG.pdb) (Nienaber et al., 1993)), and $\alpha LP_{pH\,8}$ (long pro region, orange). In proteases containing mid-sized and long pro regions, cyclic residues at position 199 sterically occlude the conformation of 228 observed in short pro region-containing proteases, restricting residue 228 to a conformation in which it is tightly packed against residue 181. The structural alignments were performed in O (Jones et al., 1990).

a.

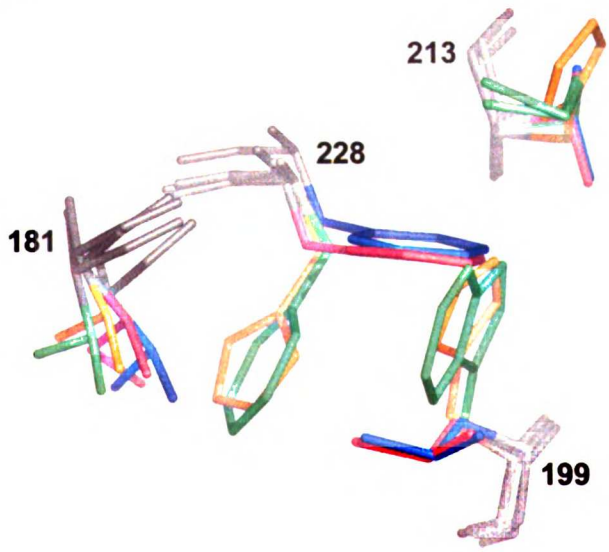


(**Figure 1.6**, continued from previous page)

b.



c.



αLP has achieved functional longevity by evolving a large, almost perfectly cooperative energy barrier to unfolding, protecting αLP from proteolysis by effectively trapping the protease in its native state. It has been hypothesized that this energy landscape largely derives from a tightly packed structure in which a high percentage of glycines mediate tight turns throughout the protein structure, essentially eliminating breathing motions (Jaswal et al., 2002). In addition, we believe that the strain introduced by the bent Phe228 sidechain acts like the spring in a spring-loaded latch or mousetrap, increasing the cooperativity of unfolding and thereby further reducing the frequency of sampling unfolded states. However, the significant energetic cost to distort Phe228 during the folding process acts both to increase the kinetic barrier to folding, and to further destabilize the native state relative to the intermediate or unfolded states (Figure 1.7). Evolution of a large, tightly-binding pro region overcomes these challenges, by catalytically reducing the barrier to folding, and by transiently shifting the equilibrium in favor of the native state during the folding process.

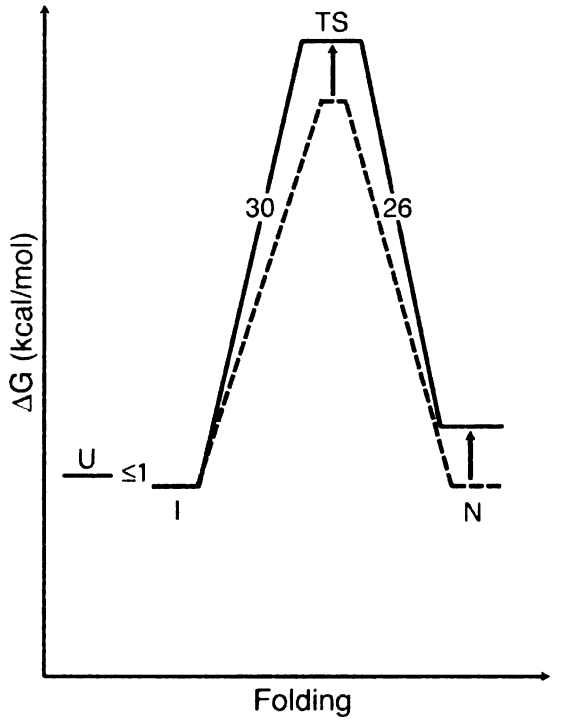


Figure 1.7. Proposed effect of the Phe228 distortion on αLP folding energetics
Folding energetics for αLP in the absence of Pro region, with (solid lines (Sohl et al., 1998)) and without (dashed lines) the distortion in Phe228. We hypothesize that introducing the Phe228 distortion increases the energy of the transition state (TS) and destabilizes the native state (N). The higher folding barrier and resulting thermodynamic instability of the native state relative to the intermediate state (I) requires a more effective Pro region. αLP benefits in that the strain introduced by the bent Phe228 sidechain increases the cooperativity of unfolding, reducing αLP's susceptibility to proteolysis.

Conclusion

A multitude of structural data is available for serine proteases in general, and for αLP in particular, including over 40 crystal structures at ≥1.2Å resolution. However, data at sub-angstrom resolution has allowed accurately precise analysis of structural details previously not possible. Based on occupancies of active site residue conformations, we have calculated the pK_a of His57 in the crystal to be ~8.4 overall, and specifically ~8.8 when a sulfate ion is bound in the active site. Electron density is visible for hydrogen atoms, and their position can be refined with remarkable accuracy. This structure confirms that the His57-Asp102 hydrogen bond in the apo enzyme is a conventional, and not a low-barrier, hydrogen bond. By revealing a site of conformational distortion that appears to be evolutionarily linked to pro region size, this ultra-high resolution crystal structure of $\alpha LP_{pH\,8}$ has provided critical new insights into the origins of folding and unfolding barriers and their relationship to kinetic stability. The rearrangement and subsequent distortion of Phe228 by conserved residues in the core of αLP supports the hypothesis that tight packing in αLP is a key component of the energetic landscape of protein folding and unfolding. It is the reliability of data at ultra-high resolution that makes us confident in the deformation of Phe228, and allows the design of future mutagenic experiments on αLP based on this observation. It will be interesting to look for geometric distortions in other ultra-high resolution crystal structures, and to use this new level of structural analysis to gain unique insights into the role of geometric strain in protein function.

Materials and Methods

Crystal Preparation

Protein was expressed in *Escherichia coli* (Mace et al., 1995), and purified as described (Derman & Agard, 2000; Mace & Agard, 1995). (See Appendices B and C for an overview of expression and purification protocols.) αLP crystals were grown by the hanging drop vapor diffusion method in 10 μl drops containing 5 μl of concentrated protein (15-20 mg/ml) and 5 μl of precipitant solution (1.3 M lithium sulfate and 20 mM Tris-sulfate, pH 8.0). Hanging drops were equilibrated over 1 mL of precipitant solution for 24-72 hours prior to seeding. A microseed suspension was prepared, using the Hampton Seed Bead Kit™, from one small crystal in 50 μl of precipitant solution. One to three crystals grew in drops seeded with 1 μl of a 1:4,000 or 1:16,000 dilution of the microseed suspension. Crystals grew to their maximum size following incubation for ~5 weeks at room temperature.

Data Collection

Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (beam line 9-1) on a Mar345 detector using x-rays of wavelength 0.785 Å. The crystal was whisked through a droplet of precipitant solution containing 15% glycerol, then flash-frozen under the N_2 gas cryo-stream (100 K). Data collection was done in two passes on a single

crystal of approximate dimensions 1.3 x 0.45 x 0.45 mm (Table 1.1). In the first ("high resolution") pass, a long exposure time and small phi-rotation were used to maximize the intensity of the weak ultra-high resolution reflections while minimizing their overlap. A minimal crystal-to-detector distance coupled with a short x-ray wavelength (0.785Å) reduced the spacing between reflections, allowing collection of data at ultra-high resolution. Spots were seen at the edge of the detector, corresponding to a resolution of 0.78Å. The stronger, lower resolution reflections that had been blocked by a beam stop in the "high resolution" pass were collected in a second "low resolution" pass. To minimize the damaging effects of long x-ray exposure, the crystal was translated across the x-ray beam periodically during data collection.

Data were processed in Denzo (Otwinowski & Minor, 1997) and scaled and merged in Scalepack (Fox & Holmes, 1966). Resolution was limited to satisfy the criteria that 50% of the data in the highest resolution bin have an $I/\sigma(I)$ greater than 2.0. The resulting resolution limit was 0.83Å. Complete data statistics are provided in Table 1.1.

Model Refinement

The 1.5 Å structure of αLP (1TAL.pdb (Rader & Agard, 1997)), crystallized under the same conditions and with similar unit cell dimensions, was used as a starting model for the refinement of the ultra-high resolution structure. Prior to any refinement, this model—including alternate conformations and waters—agreed with the data with an R-factor of 21.2% (Table 1.3). The initial stages of refinement were carried out in CNS against a maximum likelihood target using data to 1.2Å (Brünger et al., 1998; Pannu & Read, 1996).

Ten percent of the data were reserved as a test set for calculation of R_{free} (Brünger, 1992b). Based on visual inspection of electron density maps from the first round of refinement, inappropriate waters were removed from the model. Alternate cycles of minimization, B-factor, and occupancy refinement were performed, resulting in R / R_{free} of 18.0% / 18.8%. Waters and alternate conformations were added conservatively where difference electron density and hydrogen bonding interactions clearly indicated their presence. The final model from refinement in CNS contained 11 alternate conformations, 3 sulfate ions, and one glycerol molecule (R / R_{free} = 17.4% / 18.3%).

Refinement was continued in SHELXL-97, a program used for refining small molecule structures and macromolecular structures at atomic resolution (Sheldrick & Schneider, 1997). Once again, 10% of all data were reserved for calculation of R_{free}. Data to 0.90 Å was added to the refinement working set, and anisotropic B-factors modeled for each atom. This resulted in the largest decrease in R and R_{free}, to 13.4% and 15.0%. An improvement in electron density map quality allowed further modeling of alternate conformations and waters. The remainder of the data (to 0.83Å) was included in the structure factor calculations, yielding R/R_{free} of 12.2% / 13.5% after minor adjustments and refinement to convergence.

At this point, the majority of +3.0σ peaks in σ_A-weighted F_o-F_c difference maps in the core of the protein lay in positions expected for hydrogen atoms. As is standard practice for macromolecular crystallography at atomic resolution, hydrogen atoms were included in the model by SHELXL-97 as "riding hydrogens." Hydrogen atoms were rigidly placed on heavy atoms using the HFIX command in SHELXL-97, and their geometry constrained with the AFIX command (hence no parameters were added to the model). Adding "riding" hydrogens

to the model resulted in a drop in R-factors by about 1%, to R/R_{free} of 11.1% / 12.3%. In order to determine a model unbiased by geometrical restraints, we released all restraints (DFIX, DANG, FLAT, CHIV as defined in SHELXL), including non-bonding distances (BUMP), for 171 of the 198 amino acids (excluding those displaying multiple conformations). Refinement remained stable through later cycles of minimization in the absence of these restraints. Although there was a minimal effect on R-factors (Table 1.3), slight changes in geometry did occur (see Results). At this point, hydrogens were added to the His57 ring (Nδ1-H, Cε1-H, Cδ2-H), and their positions allowed to refine (with no AFIX restraint) for the remainder of the refinement. Because σ_A-weighted F_o-F_c difference electron density maps contoured at 2σ indicated unique positions for several methyl and polar hydrogens, these hydrogens were placed by structure factor calculations about the parent atom rather than by pre-defined geometry. In subsequent cycles of refinement, torsion angles were adjusted to keep the hydrogen atoms in positions of maximum electron density (using AFIX 137 and AFIX 147 for $-CH_3$ and –OH hydrogens, respectively). Finally, several cycles of refinement were devoted to modeling partial waters in the solvent region. Waters were placed manually and with the help of SHELXL-97 (Sheldrick & Schneider, 1997) and QUANTA , and their occupancies fixed to 100%. The occupancies of waters with negative F_o-F_c electron density (contoured at -3σ), high B-factors, or a distance of less than 2.45Å from other atoms were reduced to a fixed value of 50% or linked to those of nearby alternate conformations. Some waters in the model may represent atoms of larger molecules present at low occupancy (such as SO_4^{-2}, Tris, glycerol, and short peptides), but assignment of such molecules was unclear based on the electron density.

Table 1.3. Summary of the refinement process for $\alpha LP_{pH\,8}$

Refinement stage	Data[a]	Parameters	Protein atoms	Solvent atoms[b] (W / PW / NW)	R_{work} (%)	R_{free} (%)[c]
A. Refinement in CNS						
Initial model (1.2Å)[d]	56,452	-	1,474	104 / 189 / 15	21.17	21.21
Final CNS Model	56,452	*n/a*	1,475	43 / 201 / 21	17.40	18.30
B. Refinement in SHELXL-97						
Refinement of CNS model (1.2Å)	56,581	6,831	1,475	244 / 0 / 21	16.13	17.76
Minor model adjustments	56,581	6,811	1,467	244 / 0 / 21	16.14	17.76
Data to 0.90Å	132,811	6,811	1,467	244 / 0 / 21	16.84	17.80
General model adjustments	132,811	6,719	1,473	217 / 2 / 21	17.46	18.69
Anisotropic B	132,811	15,115	1,473	217 / 2 / 21	13.35	14.98
General model adjustments	132,811	16,058	1,507	244 / 42 / 38	11.83	13.45
Data to 0.83Å	168,627	16,058	1,507	244 / 42 / 38	12.31	13.79
General model adjustments	168,627	16,207	1,527	230 / 54 / 38	12.15	13.49
Add H (including $-CH_3$)	168,627	16,207	2,970	230 / 54 / 38	11.14	12.30
General model adjustments	168,627	16,645	2,992	246 / 76 / 38	10.77	11.86
Restraints released	168,627	16,645	2,992	246 / 76 / 38	10.75	11.87
Add H (His57)	168,627	16,654	2,995	246 / 76 / 38	10.74	11.86
Waters added	168,627	17,174	2,995	302 / 121 / 38	9.88	11.26
Add H (-OH)	168,627	17,207	3,028	302 / 121 / 38	9.86	11.25
General model adjustments, Final model	168,627	19,125	3,125	189 / 354 / 53	8.57	9.91

[a] Number of reflections in the working set.

[b] W=number of fully-occupied water molecules; PW=number of partially-occupied water molecules; NW=number of atoms from other solvent molecules (glycerol and SO_4^{-2}), some of them partially-occupied.

[c] The test set contained ~10% of the data (selected randomly), or 18,758 reflections for data 20Å – 0.83Å resolution.

[d] Initial model includes all atoms from 1TAL.pdb (except Tris); R-factors are prior to rigid body refinement.

The final model, containing 23 residues with alternate conformations, 4 sulfate ions, 2 glycerol molecules, 189 whole waters, and 354 partially-occupied waters converged to a working R-factor of 8.57% and R_{free} of 9.91%. Blocked least-squares refinement performed on this model provided an estimation of the errors for atomic positions. The final R-factor was 8.70% for all data, and 7.61% for data with $F_o > 4\sigma(F_o)$. The model was analyzed for geometric distortions using PROCHECK (Laskowski et al., 1993). The model coordinates are available from the Protein Data Bank (1SSX.pdb) or from our laboratory website, http://www.msg.ucsf.edu/agard.

αLP structures were aligned by Cα atoms using LSQMAN (Kleywegt & Jones, 1994). RMSD's are reported for main chain atoms (N, Cα, C) unless otherwise indicated. All figures were drawn using PyMOL (Delano, 2002) unless otherwise noted.

Determining the Occupancy of Sulfate 202

One sulfate ion (Sul202) was of particular interest due to its proximity to the enzyme's active site. Knowledge of the total occupancy of Sul202 allows the calculation of two important equilibrium constants in the context of $\alpha LP_{pH\,8}$ crystals: (1) K_d for SO_4^{2-} in the αLP active site, and (2) the pK_a of His57 (see below). Perhaps because of a variety of alternate conformations of amino acids and solvent molecules in this region (including Sul202), the total occupancy of Sul202 varied during the refinement process (average 69 ± 4 %). In order to determine an accurate value free of model bias, we ran a series of tests in parallel and determined the effect of Sul202 occupancy on R_{free}. Each trial consisted of 30 cycles of minimization, with the occupancies of $Sul202_a$ and $Sul202_b$ fixed such that each test

exhibited a unique total occupancy for Sul202. The ratio of occupancies ($Sul202_a$: $Sul202_b$) was fixed as 5.7 (as determined during the standard structural refinement) for all trials. The resulting R_{free} values were plotted against total occupancy for Sul202 (Figure 1.4) and these data fitted with a 2nd-order polynomial. The minimum of this polynomial (77.4 ± 2.1%) was taken to be the optimal total occupancy of Sul202. The use of this method was justified by the excellent fit of R_{free} values to the polynomial (R_{fit} = 0.9981). The occupancy of waters 760_a, 764, and 765, whose short distances from Sul202 indicate that these waters are present only its absence, was determined by repeating these tests with their occupancies fixed at 10%, 15%, and 20% (with 15% occupancy yielding the polynomial with the lowest minimum R_{free}; data not shown). A similar method was attempted (using 73% as the total occupancy of Sul202) in order to determine the most accurate ratio of $Sul202_a$ to $Sul202_b$, but the scatter in resulting R_{free} values demonstrated that the data was being over-determined (data not shown). Therefore, in the final cycles of refinement of this structure, the occupancies of $Sul202_a$ and $Sul202_b$ were allowed to refine, but restrained to sum to 77.4%. The occupancies of waters 760_a, 764, and 765 were fixed to 15%.

Calculation of the pK_a of His57

Because a sulfate ion (Sul202) is present in only a fraction of the molecules in the crystal lattice, we suggest that two equilibria are required to describe the protonation state of His57. As shown in Scheme 1, His57 can assume one of two pK_a values in the crystal lattice: pK_a 7 in the absence of Sul202 ($pK_{a,\,dLP}$), and a higher pK_a in the presence of Sul202 ($pK_{a,\,dLP\bullet Sul}$). Because crystallography allows the determination of the occupancy of Sul202 (77 ±

2% total occupancy for our crystal grown in 1.3M Li_2SO_4, corresponding to a $K_{d,\alpha LP\bullet Sul}$ of 0.38 ± 0.04 M), we can calculate $pK_{a,\alpha LP\bullet Sul}$. Using the equilibrium represented in Scheme 1, we calculated that in the presence of sulfate in the active site, the $pK_{a,\alpha LP\bullet Sul}$ of His57 is 8.8 ± 0.1. Using our values for $K_{d,\alpha LP\bullet Sul}$ and $pK_{a,\alpha LP\bullet Sul}$, we calculated that the overall pK_a for αLP crystals at 0.6M Li_2SO_4 would be 8.1 ± 0.8.

Ab initio Deformation Energy Calculations for Phe228

In order to estimate the conformational energy cost of deformation for Phe228, we calculated the difference in energy between the (1) intact residue conformation from the x-ray structure and (2) an optimized residue conformation. The residue of interest was isolated from the refined crystal structure (which had been refined in SHELXL with no geometrical restraints). The sidechain was truncated at Cα, which was modeled as a methyl group by adding hydrogens at optimized orientations using X-PLOR (Brünger & Karplus, 1988; Brünger, 1992a). Restricted Hartee-Fock calculations using the GAUSSIAN/98 program with the 6-31G* basis set were carried out on the isolated sidechain without geometry optimizations (Frisch et al., 1998). Energy of the corresponding optimized conformation was calculated using the same method, this time with full geometry optimizations. The relative energy difference between the distorted and optimized conformations was provided as the deformation energy in Hartree units, and then converted to kcal/mol.

Pro region Covariation Analysis

Thirty bacterial, pro region-containing proteases were identified by using αLP as a query in a BLAST search. The pro region N-termini were deduced by using the SignalP server to predict the signal sequence cleavage sites and the pro region C-termini were deduced based on homology to αLP. The 14 proteases that had pro regions of comparable size to that of alpha-lytic protease (~155-175 residues) were assigned to the "long pro region" bin. The 16 proteases with pro regions that were significantly shorter ($\leq$ 140 residues) were assigned to the "short pro region" bin. The multiple sequence alignment of these 30 proteases was made using the ClustalW tool in MacVector 7.2 (Accelrys, Inc.) using the Blosum30 weight matrix (Thompson et al., 1994). Residues at certain positions in the multiple sequence alignment were then classified according to amino acid type and totaled in each bin. A selection of these residues are displayed in Figure 1.6b.

Acknowledgments

Special thanks to Dr. Christopher Waddling for his helpful discussions and significant assistance during the process of refinement. We also thank Drs. N. Sauter and S. Harris for assistance in data collection, and Drs. L. Rice, S. Harris, and A. Shiau for insightful discussions. We greatly appreciate Dr. E.L. Cunningham and S.M.E. Truhlar for critical reading of the manuscript. C.N.F. was a National Science Foundation Predoctoral Scholar and B.A.K. supported by a Howard Hughes Medical Institute Predoctoral Fellowship. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. This work was funded through the Howard Hughes Medical Institute.

Postscript

Following this work, Brian Kelch has pursued kinetic and structural analyses of αLP mutants to test the hypothesis that conformational strain introduced at Phe228 contributes to αLP's kinetic stability (Kelch et al., manuscript in preparation). Brian's approach follows three strategies:

(1) to reduce the conformational strain of Phe228 by removing the cause of this strain (Thr181→ Gly; called "αLP_{TG}"),

(2) to remove the strain directly by mutating Phe228 to an alanine ("αLP_{FA}"), and

(3) to mutate residues surrounding Phe228 to mimic those of small pro region-containing proteases (e.g. SGPB). This includes mutants:

   a. Thr181→ Ile, Trp199→ Leu, called "αLP_{TIWL}"

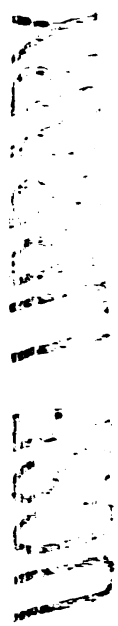
   b. Thr181→ Ile, Trp199→ Leu, Gln210→ Ile, called "αLP_{TIWLQI}."

In particular, the SGPB-mimic mutants, αLP_{TIWL} and αLP_{TIWLQI}, give striking results. Atomic resolution structures of both mutants show that mutation of the residues surrounding Phe228 allows Phe228 to adopt the conformation observed for SGPB (Read et al., 1983), as predicted by our structural alignments (Figure 1.6). Both mutants exhibited a stabilized transition state for folding as compared to that for αLP. For αLP_{TIWLQI}, which better mimics the packing interactions within the core of SGPB, both the transition state and the native state were stabilized relative to those states in αLP. Interestingly, the stabilization effect of the mutations is greater for the transition state (2.5 kcal/mol) than for the native state (0.5 kcal/mol) (Kelch et al., manuscript in preparation). One interpretation of these results is that

in αLP, Phe228 is more distorted in the transition state than in the native state (see Conclusions section of this thesis).

Brian has also been studying the kinetic and structural characteristics of two extremophilic homologs of αLP, *Thermobifida fusca* protease A (TFPA) and *Nocardiopsis alba* protease (NAPase). Preliminary data reveal that both TFPA and NAPase have folding energy profiles similar to that of αLP, with unfolding energy barriers even greater than that measured for αLP (Kelch & Agard, manuscript in preparation; Kelch et al., manuscript in preparation). Interestingly, crystal structures of these proteases show that the distortion of Phe228 is conserved among these three proteases, further supporting our hypothesis that this distortion is an evolved characteristic designed to enhance the kinetic stability of large-pro region-containing proteases.

Chapter 2: Structural changes induced by x-ray radiation damage in αLP-peptidyl boronic acid complexes



Preface

Several ultra-high resolution crystal structures of α-lytic protease (αLP) bound to peptidyl boronic acid inhibitors indicated $F_o - F_c$ difference electron density in the active site that could not be modeled using the assumption that the inhibitor-protease adduct was of tetrahedral geometry. As described here, examination of F_o-F_o difference electron density maps comparing complete low-resolution data sets collected before and after the high-resolution data collection indicated that this mysterious electron density was a result of radiation damage. With concern that radiation damage would complicate analyses discussed in Chapter 3, I explored two strategies to obtain data with minimal damage. The first strategy was to alter the data collection protocol to distribute dose over a larger volume of the crystal, resulting in the successful collection of data with no detectable radiation damage; the resulting structure (αLP+boroVal(gol)) is described in Chapter 3. In the second strategy, we applied a two-state approximation to the damage process, allowing a "zero-dose" set of structure factors to be extrapolated from the time dependent change in reflection intensities due to damage (Diederichs et al., 2003). This portion of the work was done in collaboration with Kay Diederichs at Universität Konstanz, who processed and scaled the data, applying the zero-dose extrapolation. Preliminary results confirmed the success of this method in reducing structural changes caused by radiation damage. However, minimization of this second structure converged prematurely in the refinement process despite indications of additional solvent molecules needing to be modeled. Additionally, several hydrogen atoms in the active site were not visible in electron density maps. It is unclear whether the lack of hydrogen atoms and premature refinement convergence are a reflection of the lower resolution (1.04Å) of this data as compared to other αLP data sets, or if these are a result of

noise or errors being introduced by the extrapolation method. In either case, the success of the extrapolation method in reducing radiation damage verifies this as a method worth pursuing. Further analysis of the extrapolated data is required, and perhaps different strategies applied when collecting data for this method (see Postscript).

With the exception of the zero-dose extrapolation (as noted above), I performed all experiments and analyses described in this chapter.

Abstract

Peptidyl boronic acid inhibitors bind to the catalytic serine of serine proteases, forming a boron-centered tetrahedral adduct that mimics the tetrahedral intermediates and transition states of the proteolytic mechanism. Here we describe collection of ultra-high resolution crystallographic data for α-lytic protease bound to MeOSuc-Ala-Ala-Pro-boroVal, and the detection of structural changes within the active site that occur as a result of radiation damage. Although radiation damage was not evident in global properties of the data, such as R_{merge} or a reduction in resolution over time, site-specific damage was evident in electron density maps calculated to compare data collected before and after the high-resolution data set. Structural changes observed include reduction of the three disulfide bonds in αLP and, interestingly, conformational changes at the Ser195-boronic acid adduct. We propose that these conformational changes may be a direct result of x-ray irradiation, stimulating breakage of the boron-carbon bond within the inhibitor and resulting in a trigonal-planar boronic acid-Ser195 adduct. An improved data collection strategy was designed to minimize radiation damage during high-resolution data collection. Furthermore, for one data set where damage had been detected, a two-state model of radiation damage was exploited to extrapolate the data to a level of zero-radiation dose.

Introduction

Advances in x-ray crystallographic technology, including the prevalent use of synchrotron radiation, have pushed the resolution limit for macromolecular crystal structure determination. The first utlra-high resolution (< 1.2Å) crystal structure of a protein was solved in 1984. Since then, hundreds of protein structures have been solved at this resolution, including 79 structures (of proteins larger than 100 amino acids) at sub-angstrom resolution. Sub-angstrom resolution data allows more detailed analysis of structures, with the ability to release geometric restraints during refinement, atomic positions determined with high precision, and hydrogen atoms becoming visible in electron density maps. At resolutions > 1.2Å, it is standard practice to describe the spread of electron density about each atom by a single parameter: an isotropic B-factor. With more data points collected at ultra-high resolution (and atomic detail distinguished in electron density), the spread of electron density about individual atoms can be described anisotropically. Atomic resolution data therefore provide a wealth of information about the dynamic as well as static properties of each atom.

The long exposure times, use of x-ray wavelengths < 1Å, and the consequent high radiation doses necessary for obtaining such high resolution, however, have introduced new challenges to the field. Ultra-high resolution crystal structures frequently suffer from structural effects of radiation damage, including reduction of disulfide bonds, demethylation of methionine residues, and decarboxylation of acidic residues (Burmeister, 2000; Ravelli & McSweeney, 2000; Weik et al., 2000). Experiments have shown that specific radiation damage is dose-dependent, with these particular residues being most susceptible to damage.

It has also been observed that enzyme active sites may be susceptible to radiation damage (Weik et al., 2000).

At the atomic level, radiation damage occurs via two mechanisms. In some cases, absorption of a high-energy photon by a protein atom causes ejection of an electron from its inner shell, resulting in "primary" radiation damage. Primary ionization events can occur in the solvent region or within the protein itself. Damage can also be mediated by free radicals that were themselves created by radiation damage. Such "secondary" damage is frequently caused by highly reactive solvent species such as hydroxyl or hydrogen radicals, or by hydrated electrons. For example, decarboxylation is likely caused by electron ejection-induced oxidation of carboxylic acid groups (O'Neill et al., 2002). Furthermore, the high electrophilicity of sulfur can lead directly to electron-trapping of sulfurs in disulfide bonds, leading to bond breakage (Carugo & Carugo, 2005; Weik et al., 2002). Disulfide bonds are also believed to be particularly susceptible to primary radiation damage because sulfur atoms have the highest absorbance cross-section among protein atoms at commonly used wavelengths (Weik et al., 2000). Other forms of radiation damage that have been observed include de-hydroxylation of tyrosine residues, loss of the methylthiol group from methionines, and loss of sulfur from cysteine sidechains (for a review of these observations and their corresponding mechanisms, see (Burmeister, 2000)). Most mechanisms of radiation damage occur via secondary events. Collecting data at cryogenic temperature (100K) both stabilizes the crystallographic lattice and limits the diffusion of free radicals within the crystal, minimizing radiation damage by secondary events.

A member of the chymotrypsin family of serine proteases, αLP has been used as a model to study serine protease mechanism for decades. In the general mechanism, the

enzyme's catalytic serine (Ser195) attacks the scissile carbonyl of the substrate, forming a metastable tetrahedral intermediate structure (TI_1). In this reaction, neighboring His57 acts as the general base, abstracting the proton from Ser195 Oγ, increasing its nucleophilicity. His57 then donates the proton to the amide of the substrate, catalyzing release of the C-terminal portion of the substrate and resulting in formation of the "acyl enzyme" complex. In the second step of the reaction (deacylation), His57 catalyzes attack by a water molecule on the acyl enzyme adduct, creating a second metastable tetrahedral intermediate (TI_2). Finally, His57 donates a proton to Ser195, facilitating release of the cleaved peptide product. Transition state analogs allow structural analysis of the metastable tetrahedral intermediate structures in atomic detail. In particular, peptidyl boronic acids, in which the C-terminal carbon is substituted with a boron, are excellent mimics of TI_2 (Figure 2.1) (Kettner et al., 1988; Bone et al., 1987). In the presence of glycerol as a cryoprotectant, glycerol molecules rapidly diffuse into the crystal lattice, and covalently bind to the boronic acid to provide a mimic of the TI_1 state (Figure 2.1c; Chapter 3).

Here we describe radiation damage detected at the active site of α-lytic protease (αLP) in structures of αLP bound to MeOSuc-Ala-Ala-Pro-boroVal solved at ultra-high resolution. Specifically, we detect conformational changes at Ser195 and boroValP1 that result from radiation damage that occurs during the long exposure to x-ray radiation required to obtain the ultra-high resolution data. These conformational changes may have been induced by radiation damage directly via cleavage of the B-Cα bond within the inhibitor. Two strategies were explored to obtain data without the influence of radiation damage. In the first, high-resolution data collection was optimized to distribute the damage across a large number of crystal positions, with low-resolution data sets collected before and after each

partial high-resolution set. The resultant structure, "αLP+boroVal(gol)," provided a mimic of TI_1 without radiation damage. In the second case, "αLP+boroVal," radiation damage was allowed to accrue throughout the collection of very high-resolution data. Using a two-state approximation, the method of Diederichs *et al.* was used to computationally extrapolate to a zero-damage data set (Diederichs et al., 2003).

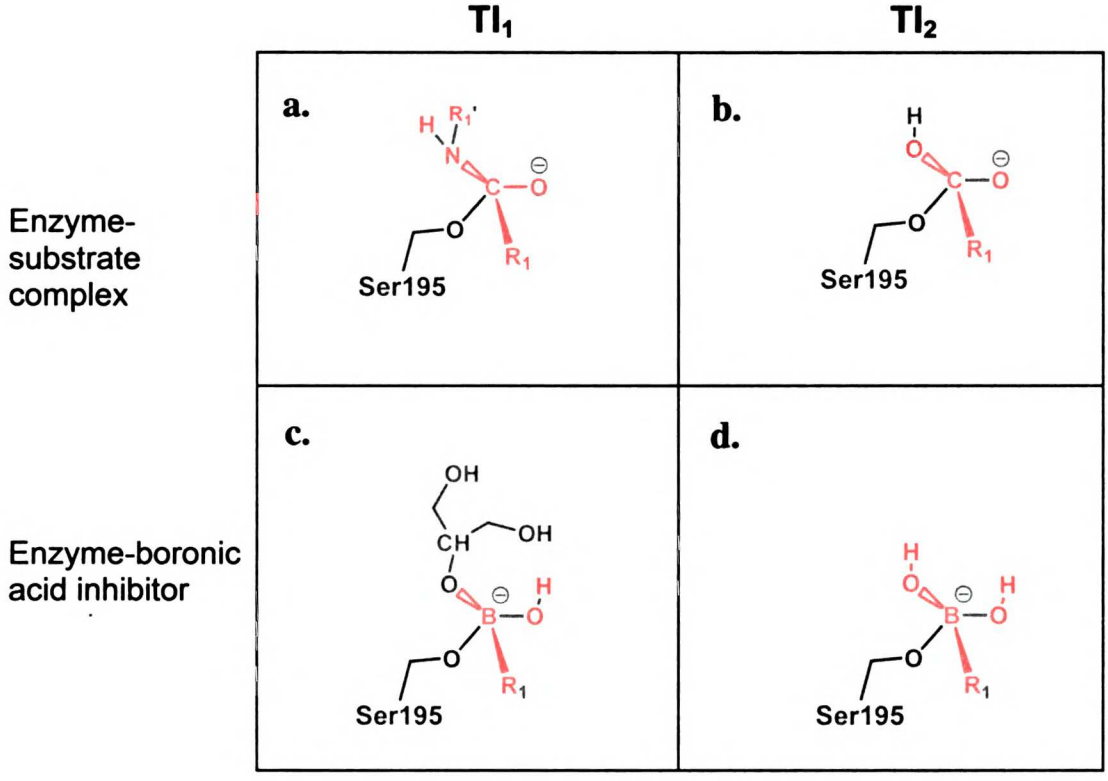


Figure 2.1. Boronic acid inhibitor as a mimic of the acylation and deacylation tetrahedral intermediates.
The schematics above illustrate the predicted chemical structures of TI_1 and TI_2 for the enzyme-substrate (a, b) and enzyme-inhibitor (c, d) complexes. The catalytic serine is colored in blue, substrate/inhibitor in red, reactive water in purple, and glycerol in green. Note that when data was collected using glycerol as a cryoprotectant, a glycerol molecule covalently bound to O_2 of the inhibitor, providing a mimic of the acylation tetrahedral intermediate.

Results

Isolation of a mimic of TI_2

In order to study the catalytic mechanism of αLP at ultra-high resolution, we set out to solve the crystal structures of αLP bound to a variety of peptidyl boronic acid inhibitors. Such complexes should mimic the tetrahedral intermediate state for the deacylation step of the catalytic mechanism (TI_2). Surprisingly, when data were analyzed for crystals that had been only very briefly soaked in cryoprotectant (15% glycerol), it was discovered that in most cases a glycerol molecule covalently attached to the boronic acid and occupied the S1' pocket[3], providing a mimic of the first tetrahedral intermediate (TI_1; see Chapter 3). Even in structures for which initial difference electron density maps failed to indicate that glycerol was bound, further refinement often revealed a glycerol bound to the boronic acid at low occupancy (results not shown). In order to obtain a proper mimic of TI_2, PEG-300 was used as a cryoprotectant in place of glycerol. PEG-300, a molecule larger than glycerol, should be unable to bind in the S1' pocket. Data was collected on αLP bound to MeOSuc-Ala-Ala-Pro-boroVal in 20% PEG-300 ("αLP+boroVal*"[4] and "αLP+boroVal"). As expected, two water molecules were observed in the S1' pocket where glycerol had previously been modeled (data not shown).

---

[3] Substrate residues are labeled according to the nomenclature of Schechter and Berger Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27, 157-62., where the residue N-terminal to the scissile bond is denoted P1, and is bound in the enzyme's S1 specificity pocket. Residues extending N-terminal from the cleaved bond are P1, P2, etc.; those extending C-terminally are P1', P2', etc.

[4] Structures demonstrating radiation damage are indicated with an asterix.

Detection of radiation damage in αLP crystals

*Indications of radiation damage in αLP+boroVal**

Collection of ultra-high resolution data requires extensive x-ray radiation exposure. In order to test for radiation damage at specific structural sites within the protein molecules, a complete low-resolution (~1.2Å) data set was collected before and after the high-resolution (~0.9Å) data collection. Because total exposure in each low-resolution set was ~5% of the exposure required for the high-resolution data, their effect on the structure was negligible. This allowed calculation of $F_{o,\,after} - F_{o,\,before}$ difference maps, which were used to identify conformational changes induced by radiation damage. Phases for these maps were provided by the model of αLP solved at pH 8 (1SSX.pdb (Fuhrmann et al., 2004a)). As shown in Figure 2.2, four regions of the protein demonstrated significant structural changes due to radiation damage. Three of these regions correspond to reduced disulfide bonds. In particular, Cys42/Cys58 (Figure 2.2c) and Cys189/Cys220A (Figure 2.2d) both demonstrate broken disulfide bonds at significant occupancy (~31% and ~34%, respectively). Reduction of these disulfide bonds was described previously for $\alpha LP_{pH\,8}$ (Chapter 1 (Fuhrmann et al., 2004a)), and for other structures of αLP solved at ultra-high resolution by this author (see Table 2.1). In αLP+boroVal*, Cys137/Cys159 is also reduced, although with less frequency than other disulfide bonds (Figure 2.2e). In F_o-F_c difference maps calculated for $\alpha LP+boroVal*_{11}$, negative density was observed on Cys159 stretching along the disulfide bond, but no positive density was visible (despite positive density indicating new

Table 2.1. Effects of radiation damage in ultra-high resolution crystal structures of αLP.

Reference (Chapter / Appendix)	Structure	Cys42-Cys58	Cys189-Cys220A	Cys137-Cys159	Resolution (Å)	R / R_{free}
Ch 1	WT αLP (pH 8)	16.0	13.0	-	0.83	8.6 / 9.9
Ch 3	WT αLP (pH 5.1)	25.2	20.2	11	0.82	9.9 / 11.0
Ch 2, Ap. E	WT αLP + MeOSuc-AAP-boroVal (in PEG)	31.4*	34.4*	-	0.89	12.5 / 14.8*
Ch 3	WT αLP + MeOSuc-AAP-boroVal... glycerol (no damage)	0	0	0	0.90	9.2 / 10.3
Ap. E	MA190 (pH 8)	17.3*	19.9*	-	0.86	12.9 / 14.6*
Ap. E	MA190 + MeOSuc-AAP-boroAla ... glycerol (C.N. Fuhrmann and C.Waddling)	14.3*	12.0*	-	0.90	9.8 / 12.0*
Ap. E	WT αLP + MeOSuc-AAP-boroVal ... glycerol	16.0*	18.3*	-	0.86	11.3 / 13.1*
Ap. E	WT αLP + MeOSuc-AAP-boroAla ... glycerol	18.3*	17.8*	-	0.89	9.5 / 11.5*
Ap. E	WT αLP + Ac-P-boroVal (in glycerol)	20.2*	-	-	0.88	12.6 / 14.2*
Ap. E	WT αLP + MeOSuc-AAP-boroVal (in glycerol)	26.2*	22.9*	-	0.86	11.6 / 13.0*

* Indicates structures were not fully refined (see reference); therefore, results are preliminary.

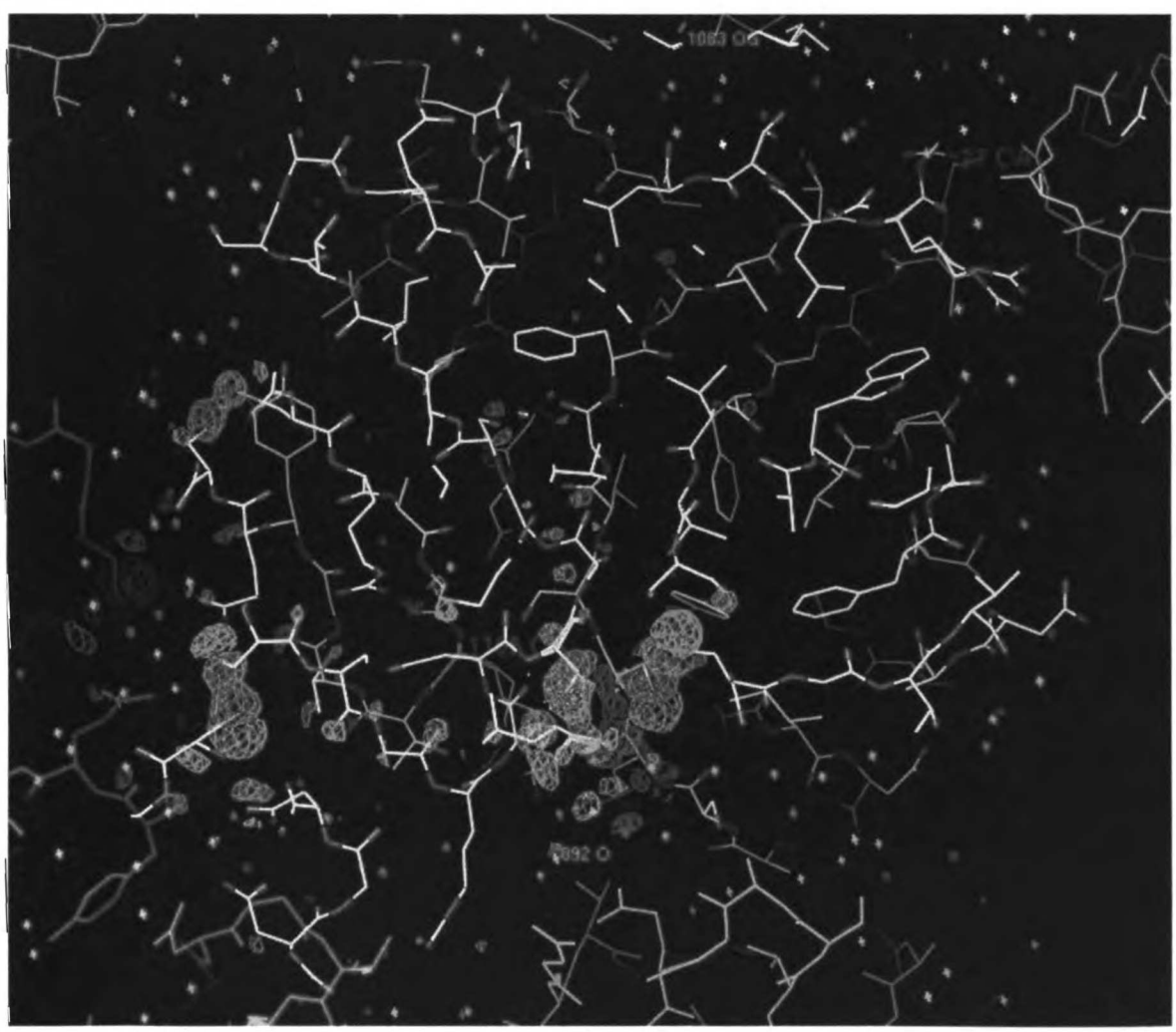
conformations for Cys137 and Cys159 in $F_{o,\,after} - F_{o,\,before}$ maps). An alternate conformation of Cys137 and Cys159 has only been observed in $\alpha LP_{pH\,5}$ solved at 0.82Å resolution (Chapter 3). All three disulfide bonds are solvent-exposed, allowing radiation damage by primary or secondary events.

A cluster of structural changes was evident in the vicinity of Cys42/Cys58. Backbone atoms of neighboring residues 40-43 exhibited conformational change, with particularly strong $F_{o,\,after}$-$F_{o,\,before}$ difference density centered on the Leu41-Cys42 peptide bond. This difference density corresponds to a smearing of $2F_o$-F_c density and relatively high B-factors observed in other αLP structures (see Chapter 3), including structures without radiation damage. The second conformation of Leu41 C=O observed in $F_{o,\,after}$-$F_{o,\,before}$ difference density is likely a result of steric clash between the backbone carbonyl group of Leu41 and the new conformation of Cys42 Sδ created upon disulfide reduction. Second conformations of Wat911, which hydrogen-bonds to Leu41 O, and other nearby waters are also observed.

$F_{o,\,after} - F_{o,\,before}$ difference density was also observed near Sδ of Met190 and Met213. The sidechains of both methionine residues are located in the S1 binding pocket, positioned between Cys189 and boroValP1 (see below). Because positive as well as negative difference density is present on these sidechains, and because similar density is observed for neighboring backbone atoms, we conclude that these observations are due to general conformational changes of the Met sidechains, rather than loss or modification of methylthiol groups (as is often observed under conditions of radiation damage).

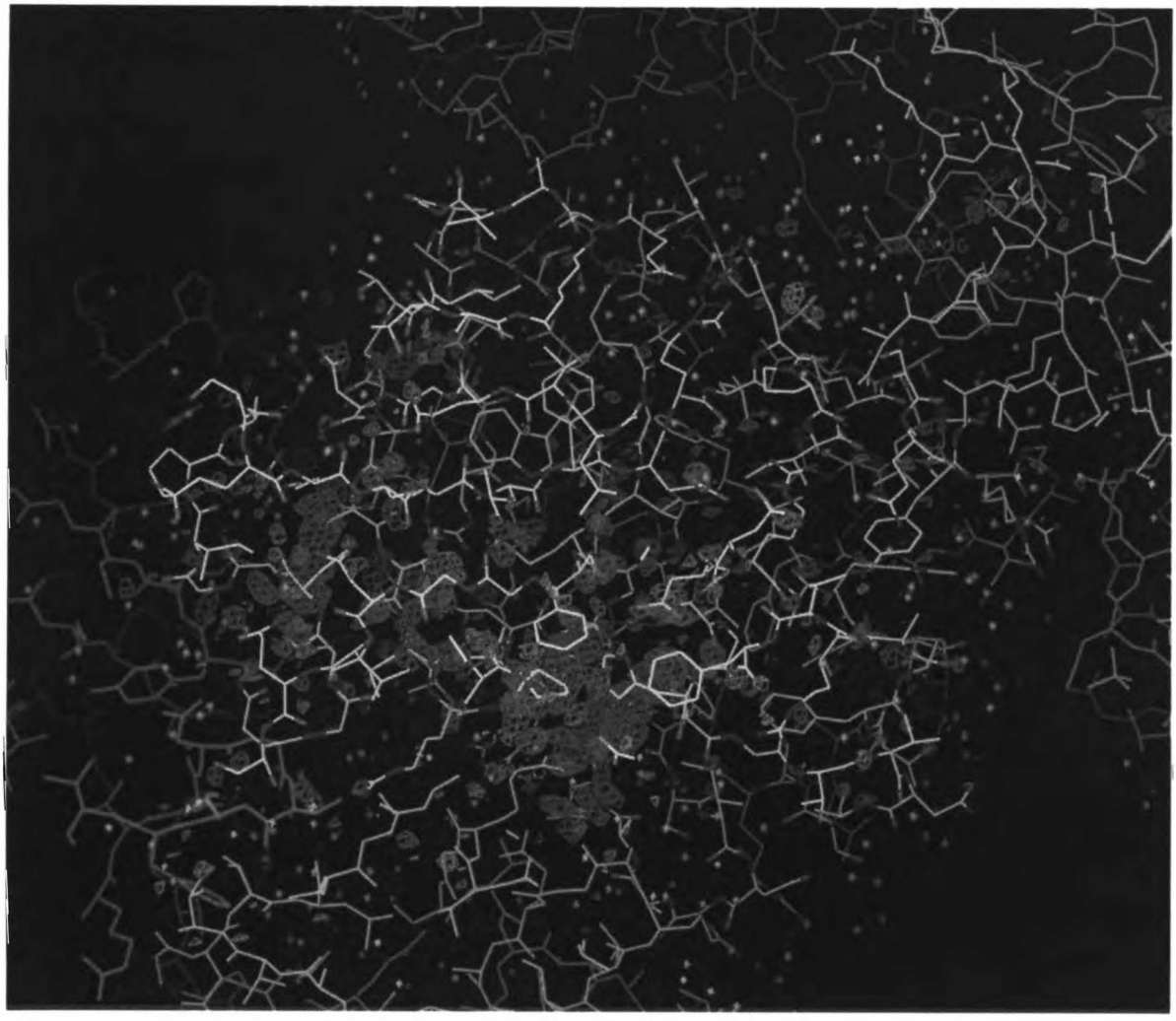
Figure 2.2. $F_{o, after}$ - $F_{o, before}$ comparing data collected before and after high-resolution data collection for αLP+boroVal*.
(a) Contoured at ±4σ, damage is primarily evident at disulfide bonds and at the Ser195-inhibitor adduct, shown here in a cross-section of the protein. (b) At lower contour levels (±3σ), a broader swath of conformational changes is apparent in the protein, covering a distance of ~25Å. (c-e) show radiation damage for these regions. Positive density (+4σ), indicative of more electron density after x-ray irradiation, is colored in green. Negative density (-4σ), indicative of less electron density following irradiation, is colored in magenta. αLP is shown in a sticks representation, colored by atom, while neighboring proteins are displayed as a solid color.

a.



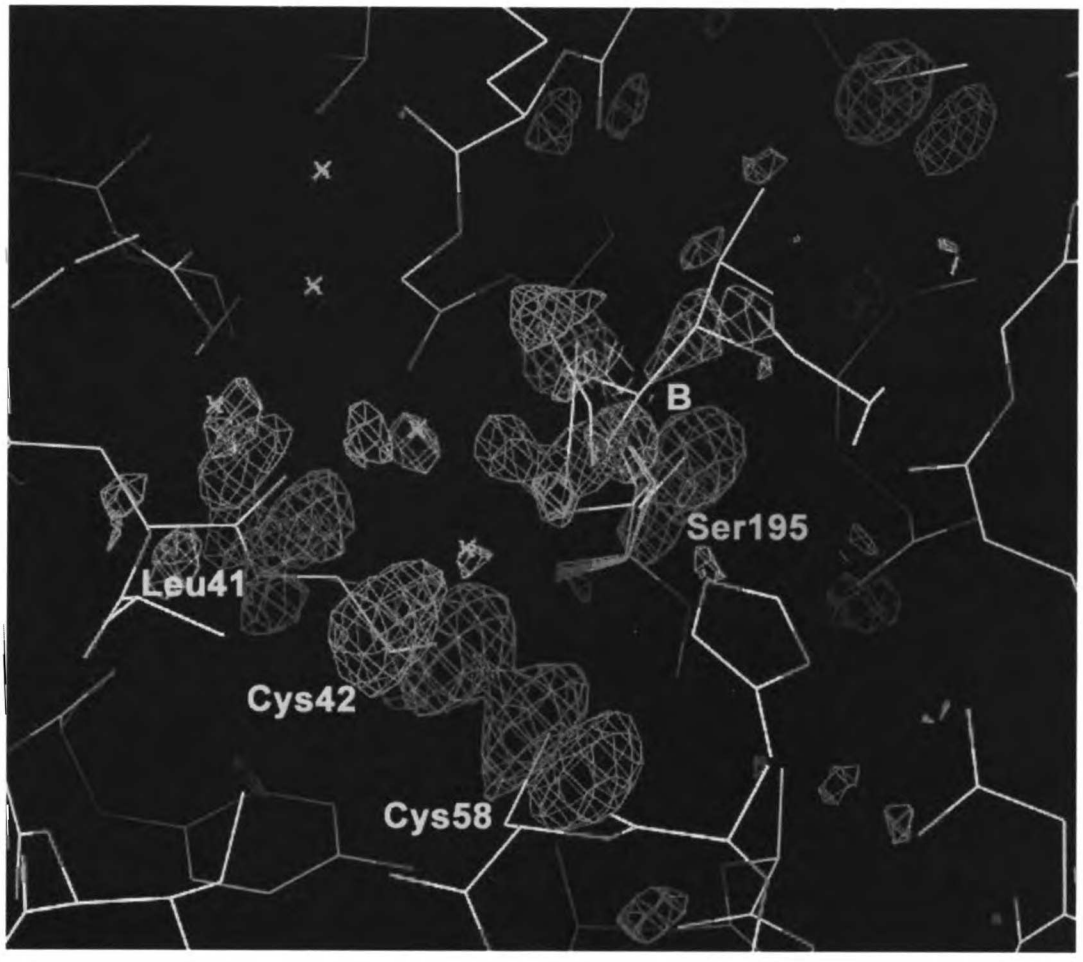
(Figure 2.2, continued)

b.

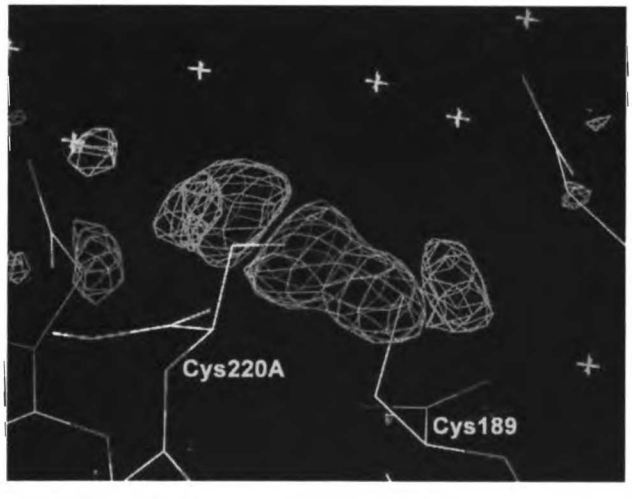


(Figure 2.2, continued)

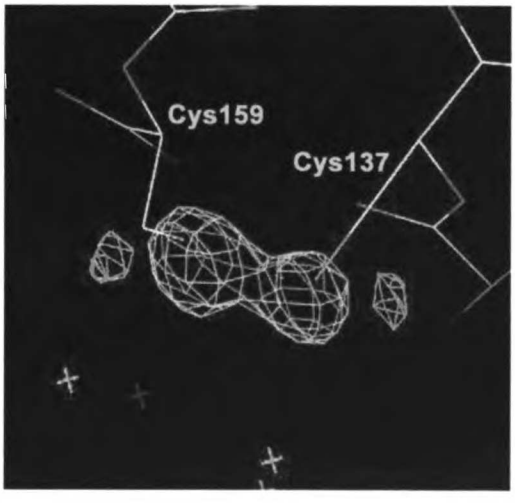
c.



d.



e.



Structural changes observed near the active site

In several crystal structures of αLP bound to a boronic acid inhibitor, negative and positive F_o-F_c electron density was observed in the immediate vicinity of the boron-Ser195 adduct. Figure 2.3 shows an example of this density in αLP+boroVal*, although similar disorder was observed for all other αLP+boronic acid complexes solved at ultra-high resolution. These difference maps were suggestive of a second and perhaps third conformation of the sidechain for Ser195 and of the boronic acid moiety. Analysis of $F_{o,\,after}$ – $F_{o,\,before}$ for αLP+boroVal* indicated that this conformational disorder was a result of radiation damage.

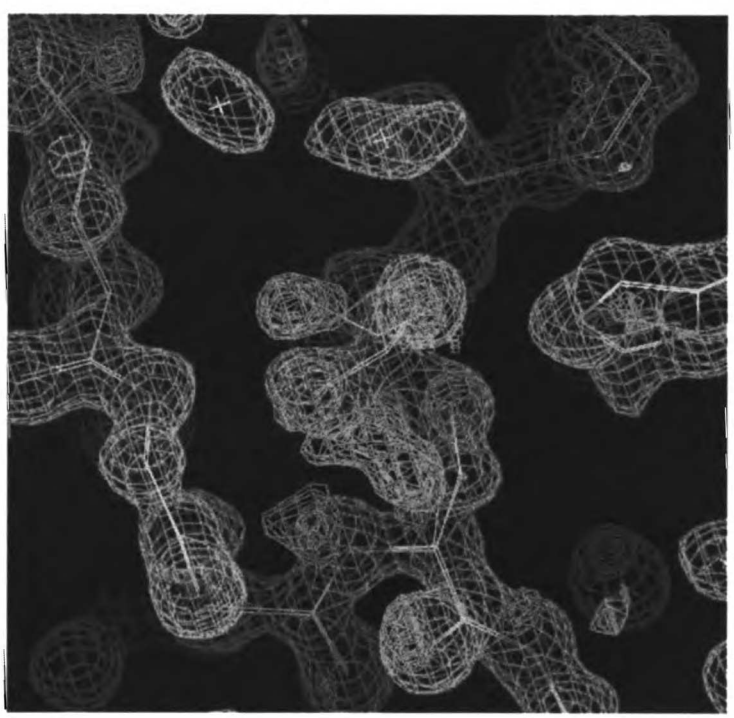


Figure 2.3. F_o-F_c difference electron density at the Ser195-boroValP1 adduct. Difference electron density maps are shown for the preliminary model of αLP+boroVal*. The yellow model illustrates the tetrahedral alternate conformations that were added to the model to accommodate this difference density. $2F_o$-F_c electron density (1.5σ) is colored purple, positive F_o-F_c density colored in green (+3σ) and cyan (+4σ), and negative F_o-F_c density (-3σ) colored in red.

αLP's active site lies adjacent to the Cys42/Cys58 cysteine pair, with Ser195 Cβ ~4Å from the disulfide bond. $F_{o,after}$-$F_{o,before}$ difference maps illustrate that the primary conformation of Ser195 Cβ-Oγ is less occupied following the collection of ultra-high resolution data. BoroValP1 and the carbonyl of ProP2 are also less occupied, with strongest $F_{o,after}$-$F_{o,before}$ difference density on boroValP1 O1 and O2, and stretching along the Cα–B bond. Correspondingly, positive $F_{o,after}$-$F_{o,before}$ difference density was observed adjacent to boroValP1 O1, and between O1 and Ser195 Cβ. The $F_{o,after}$-$F_{o,before}$ difference density agrees remarkably with the F_o-F_c difference density observed during refinement of other structures of αLP bound to boronic acid inhibitors (see Figure 2.3), suggesting that these conformational changes are a direct or indirect result of radiation damage accrued during collection of high-resolution data.

Within the catalytic triad are weak indications of conformational change. When $F_{o,after} - F_{o,before}$ maps are contoured at lower levels to reveal lower-occupancy or more subtle structural changes, negative density appears stretching between Oδ1 and Oδ2 of Asp102. Although this density is difficult to interpret, it may be an indication of a small percentage of Asp102 residues becoming decarboxylated. Positive and negative $F_{o,after} - F_{o,before}$ difference density is also observed at His57 Nε2, which may indicate a slight wobble movement of this residue in αLP+boroVal*$_{11}$. Indeed, B-factors for atoms in the His57 ring increase by 13% on average. Radiation damage has been observed previously in the active sites of enzymes (Weik et al., 2000), although it is unclear whether or not active sites are generally more susceptible to radiation damage than other protein regions (Leiros et al., 2001).

The number of residues in the vicinity of the active site that demonstrate conformational changes as a result of radiation damage is extensive. Looking at $F_{o,after} - F_{o,}$

$_{before}$ difference electron density maps contoured at an intermediate level of ±3σ, conformational changes cover a region 25Å across in two dimensions. The most intense region of difference density extends between Cys42/Cys58 and Cys189/Cys220A, sandwiching Ser195/boroValP1, Met190, Met213, and disordered backbone atoms of residues 189-192. Structural changes also extend away from this region along the length of the residue 41-47 β-strand and in residues 196-198 of the adjacent β-strand. The close vicinity of known primary targets of radiation damage (Cys42/Cys58 and Cys189/Cys220A) to the Ser195/boroValP1 adduct make it difficult to determine if disorder at the latter residues is caused by radiation damage directly, or simply reflects conformational (and not chemical) changes due to nearby radiation damage.

Using B-factors as an indicator of radiation damage

Results from $F_{o, after} - F_{o, before}$ difference maps agreed with another indicator of radiation damage: increased atomic B-factors. A preliminary model of αLP (including only one conformation for every residue) was refined against each low-resolution data set (αLP+boroVal*$_3$ and αLP+boroVal*$_{11}$). Although atomic B-factors were generally lower in αLP+boroVal*$_{11}$ compared to αLP+boroVal*$_3$, particular atoms did show a large increase in B-factor. Figure 2.4 shows the largest percent increase in isotropic B-factor for atoms in each residue of the structure. Residues with the largest percent increase included Cys42/Cys58, Ser195, and Cys189/Cys220A. Although regions showing highest-occupancy and most significant structural changes due to radiation damage (as evident in $F_{o, after} - F_{o, before}$ maps) did demonstrate an increase in B-factor, the inverse was not necessarily true. Several

residues showing an increase in B-factor of more than 20% did not show significant conformational changes in $F_{o,\,after} - F_{o,\,before}$ electron density maps. Nevertheless, this method provides an additional useful tool for identifying potential regions of radiation damage.

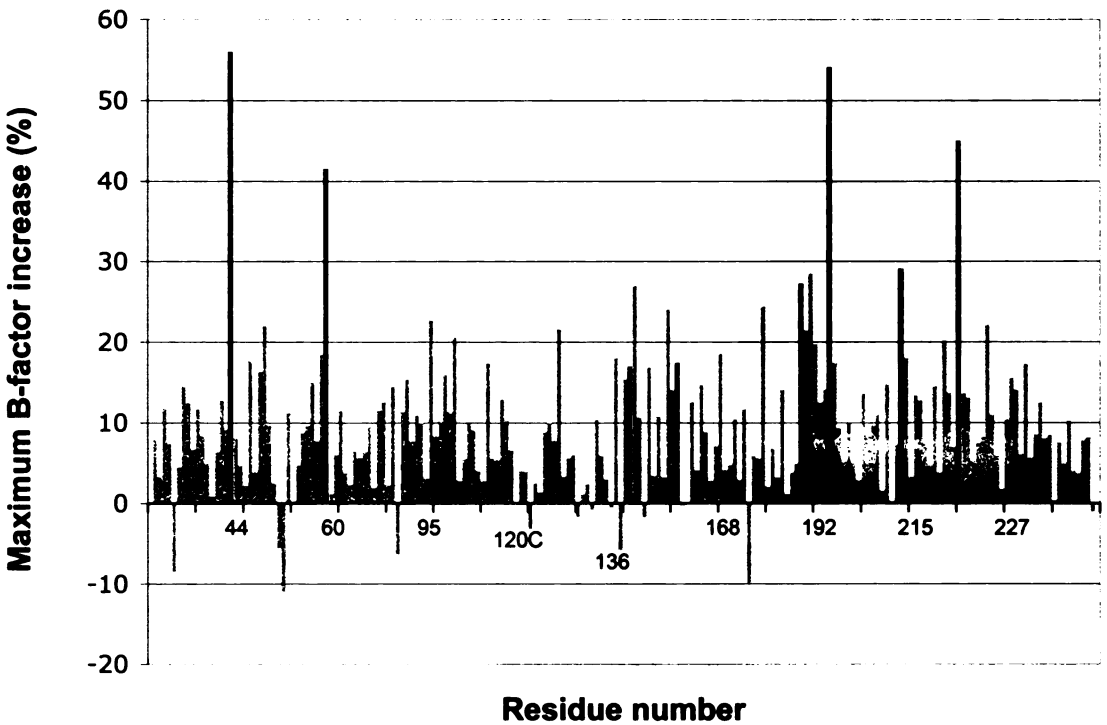


Figure 2.4. Increase in B-factors as an indicator of radiation damage.
For each residue in αLP+boroVal*, the largest percent change in B-factor for any atom in that residue is shown. Residues are numbered using homology to chymotrypsin. Cysteine residues are colored in red, methionine residues in green, and Ser195 in blue. Percent change in B-factor was calculated as follows:

$$\% \Delta B = 100 \times (B_{after} - B_{before}) / B_{before}$$

Radiation damage is caused by the long exposures of high-resolution data collection

Because collection of a complete low-resolution data set requires approximately the same amount of x-ray exposure as one frame of high-resolution data, we hypothesized that minimal radiation damage occurred during collection of any given low-resolution data set. As a test of this hypothesis, we refined the structure of αLP+boroVal* against the first low-resolution data set collected (αLP+boroVal*$_3$), and compared electron density maps to those obtained from refinement of αLP+boroVal* against the last low-resolution data set collected (αLP+boroVal*$_{11}$). As expected, there was no indication of structure-specific radiation damage in F_o-F_c difference electron density maps calculated from low-resolution data collected prior to high-resolution data collection. In contrast, strong positive and negative F_o-F_c difference electron density indicated damage had occurred in αLP+boroVal*$_{11}$. This electron density corresponded well with that observed in F_o-F_o maps.

Modeling structural changes at the Ser195/boroValP1 adduct

In order to learn more about the novel conformations of Ser195 and boroValP1 that were formed as an effect of radiation damage, a model of αLP+boroVal was refined against the last low-resolution data collected (data set #11). This data should exhibit the highest occupancy of radiation-induced structural changes, aiding in modeling efforts. Two approaches were used to model the Ser195/boroValP1 alternate conformations. First, an alternate conformation of the Ser195-boroValP1 adduct was added, in which tetrahedral geometry was maintained about the boron, but Ser195 Oγ and boroValP1 O1 were shifted to

accommodate positive difference density. This approach was based on the assumption that radiation damage occurring in nearby residues (i.e., Cys42/Cys58) caused conformational changes in the active site. Although this approach did accommodate positive density at O1, strong positive density remained stretching between Ser195 Cβ and boroValP1 O1a.

Another possibility is that radiation damage occurred directly in the active site, causing cleavage of a covalent bond in the boron adduct and allowing sampling of multiple conformations. Because B-O bonds are fairly stable (S. Kahl, personal communication), such bond cleavage was most likely to occur at the B-Cα bond. This would result in a trigonal planar boron adduct at Ser195. Presumably the N-terminal portion of the peptide inhibitor would remain bound in the binding pocket, stabilized by hydrogen-bonding interactions. In an attempt to model this trigonal planar product of radiation damage, second conformations of Ser195 Cβ, Ser195 Oγ, and boroValP1 O1, B, O2 were added to the model, and restraints assigned to maintain trigonal planar geometry. Even after adding two trigonal planar models, positive density remained off Ser195 Cβ.

Neither strategy fully accounted for the difference density observed in F_o-F_c or $F_{o,\,after}$ – $F_{o,\,before}$ maps. For both models, residual F_o-F_c density adjacent to Ser195 Cβ remained. One possibility is that the B-Oγ bond did cleave, resulting in a second conformation of Ser195. For the trigonal-planar trials described above, adding a third conformation of Ser195 in which the sidechain was rotated about the chi-1 torsion angle did accommodate the residual positive density. This conformation may be stabilized by van der Waals interactions with the reduced conformation of Cys42. Such a rotation of Ser195 would only be allowed if the boroValP1 B...Ser195 Oγ bond were cleaved. This would result in reformation of a trigonal planar peptidyl boronic acid and the free Ser195. Further studies are required to

definitively assign conformations to the structural changes observed in radiation-damaged αLP-boronic acid complex crystals.

Data collection strategies for obtaining αLP structures with minimal radiation damage

The scientific goal for solving structures of αLP bound to boronic acid inhibitors was to study the active site and catalytic triad in very precise structural detail. However, it was evident that radiation damage was a significant contributor to conformational disorder observed in the active site and thus not likely to be of functional significance. Such radiation damage likely also caused systematic errors in the data, which could adversely influence our interpretation of this data. Therefore, we set out to alter data collection strategies in an effort to obtain ultra-high resolution crystal structures with minimal radiation damage. These structures are summarized in Table 2.2.

Table 2.2. Summary of data collection parameters and statistics for αLP+boroVal structures.

	Structure		
A. Data Collection	αLP+boroVal*	αLP + boroVal(gol)	αLP + boroVal
Damaged?	Yes	No	No
Mimic	TI_2	TI_1	TI_2
Cryoprotectant	20% PEG-300	15% glycerol	20% PEG-300
Crystal dimensions (mm)	~0.3 x 0.3 x 0.7	~0.3 x 0.3 x 0.5	~0.3 x ~0.15 x 0.5
λ (Å)	0.785	0.750	0.785
Collimator (μm)	100	100	200
Maximum total exposure at one position			
Time (min)	~70	~20	~47
Flux (estimate) (photons/sec)	*Not available*	*Not available*	0.85×10^{15}
Energy (estimate) (J)	-	-	6,000
B. Data Statistics			
Space Group	$P3_221$	$P3_221$	$P3_221$
Unit Cell (a, c, Å)	65.6, 79.8	65.8, 79.5	65.6, 79.7
Mosaicity	~0.2	~0.17	~0.25
Limiting Resolution (Å)	0.89	0.90	1.04
# Unique Reflections	151,635	147,100	93,513
$I/\sigma(I)$	36.4 (3.3)	27.9 (4.8)**	33.9 (6.4)
% Completeness	100.0 (99.8)	99.9 (99.5)**	97.8 (95.0)
R_{merge}‡(Å)	7.7 (46.1)	8.5 (44.8)**	4.9(54.1)

‡R_{merge} (as calculated by Scalepack (Fox & Holmes, 1966))

Energy of radiation affects the absorbed dose

The ultimate goal when minimizing radiation damage is to minimize the number of photons absorbed by molecules within the crystal. The characteristic absorbance by a protein crystal is determined by the atomic composition of the crystal and the x-ray energy (wavelength) used (Drenth, 1999). We measured the absorbance of radiation at two different beam energies (8 keV and 15.5 keV) by a single crystal containing αLP complexed with MeOSuc-Ala-Ala-Pro-boroVal cryoprotected with 15% glycerol. For both tests, the beam was positioned on a horizontal edge of the crystal. The crystal was then rotated about phi (with the axis of rotation along the length of the crystal). During rotation, the beam alternately passed through thin and thick regions of the crystal, the thickest region having a path length of 0.3mm. The flux of the beam after passing through the crystal was measured by a photodiode. The resulting energy was plotted as a function of phi-angle (see Figure 2.5). At 8 keV (corresponding to a wavelength of ~1.5Å), the energy of the beam decreased by ~27% due to the change in absorbance between the thinnest and thickest widths of the crystal. This agrees well with published absorbance by a protein crystal of this size under these conditions (Drenth, 1999). In contrast, at 15.5 keV (wavelength ~0.8Å) there was at most a 6% difference in absorbance between the thickest and thinnest widths of the crystal. All ultra-high resolution data for αLP have been collected using x-ray wavelengths of 0.75-0.8Å since (by Bragg's Law) a short wavelength provides higher resolution data for a given diffraction angle, allowing capture of high-resolution reflections on the detector. As is suggested by this experiment, a secondary benefit of using these high-energy wavelengths is reduction of the amount of radiation absorbed, and thus a decrease in the damage done to these crystals during irradiation. While these are intriguing results, they should be

considered with caution. Higher-energy wavelengths provide stronger diffraction power, thereby decreasing the amount of exposure necessary to acquire reflections of a given intensity. Unfortunately, we did not pair this experiment with collection of diffraction data (which would allow calculation of absorption at each wavelength relative to a given diffraction intensity). We therefore cannot conclusively determine the advantage of using short wavelengths.

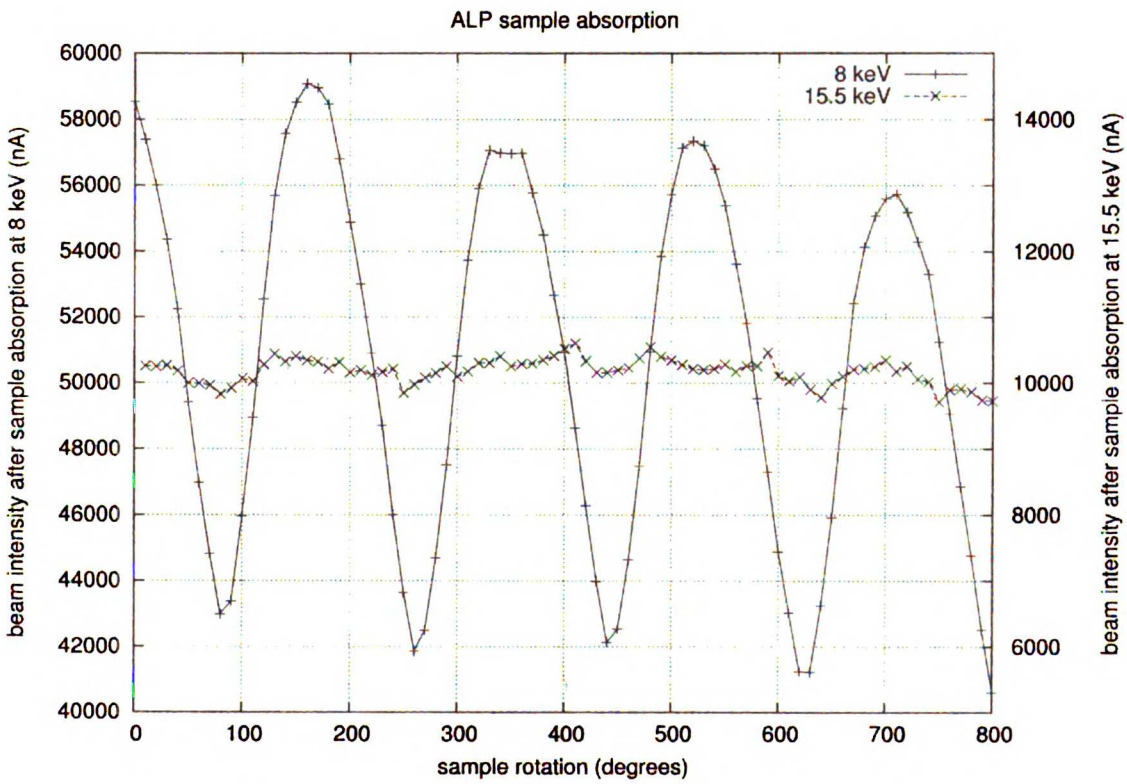


Figure 2.5. High-energy light is absorbed less efficiently than low-energy light. The beam was positioned on the edge of the crystal, and the crystal rotated about phi such that the beam routinely passed first through the edge of the crystal and then through its full width. The beam intensity was measured by a photodiode placed in the path of the beam on the opposite side of the crystal. Two beam energies were tested: 8 keV and 15.5 keV, corresponding to wavelengths of ~1.5 and 0.8Å, respectively. The latter high-energy wavelength is typically used in collection of high-resolution data for αLP.

Strategy for minimizing radiation damage during data collection: Isolation of a TI_1 mimic

In order to isolate a mimic of the tetrahedral intermediate for acylation (TI_1), we collected data on a crystal containing αLP complexed with MeOSuc-Ala-Ala-Pro-boroVal, soaked in a 15% glycerol solution. To minimize radiation damage, we sought to distribute collection of high-resolution data across as many positions of the crystal as possible. Data was collected on a single crystal of dimensions ~0.3 x 0.3 x 0.5 mm. A 100μm collimator was used, reducing the size of the beam relative to the crystal and thereby allowing for more translations along the crystal.

During data collection the crystal was translated nine times along both the length and width of the crystal (Figure 2.6a), with only 15-17 frames of high-resolution data collected at each position. In order to verify that no radiation damage had occurred, a "low-resolution" data set (to ~1.6Å resolution) was collected prior to and after each high-resolution data set. Therefore, at any given position on the crystal, three data sets were collected: (1) a complete low-resolution data set ("pre-high"), (2) ~12° of high-resolution data (20% of a complete data set), and (3) a complete low-resolution data set ("post-high"). Each high-resolution data set overlapped by at least 1 frame of data (0.7°) with the previous high-resolution data set, allowing accurate merging of the data sets. The total exposure time for each low resolution set was ~140 seconds, compared to ~1200 seconds for the zone of high-resolution data. $F_{o\,(post\text{-}high)} - F_{o\,(pre\text{-}high)}$ maps were calculated for each position of the crystal to test for detectable damage, such as reduction of disulfide bonds, which may have occurred during the high-resolution data collection. In each difference map there was no indication of any specific

a.

b.

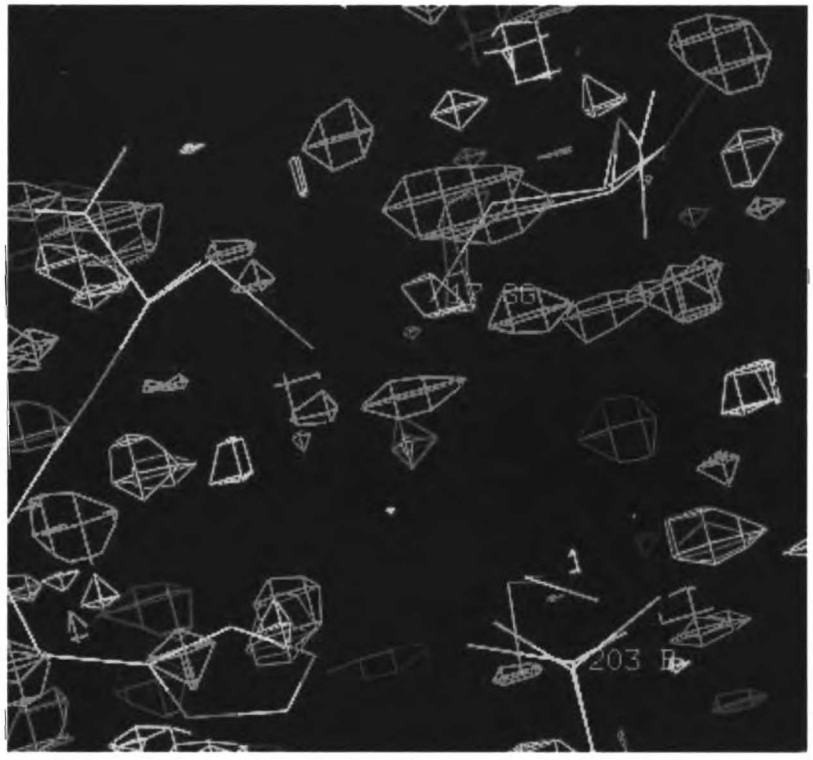


Figure 2.6. Collection of data with minimal radiation damage: αLP+boroVal(gol). (a) The photo shown here is of similar dimensions as the crystal on which data was collected. At position 1 (green circle), low-resolution data was collected. High-resolution data was then divided into nine other positions on the crystal (yellow circles). At each position, a complete low-resolution data set was collected before and after the high-resolution data collection. This allowed calculation of $F_{o,\,after} - F_{o,\,before}$ differences maps, which were used to detect radiation damage (b). $F_{o,\,after} - F_{o,\,before}$ electron density is colored in green ($+2\sigma$) and magenta (-2σ); no radiation damage was detected above noise level in these maps.

changes to the structure, indicating that little or no radiation damage had occurred during the high-resolution data collection (Figure 2.6). As a secondary test, refinement was pursued using one of the low-resolution data sets collected after the corresponding high-resolution exposure, and F_o-F_c difference maps calculated. These electron density maps confirmed the previous results by showing no reduction of disulfide bonds or other indicators of radiation damage (Figure 2.7).

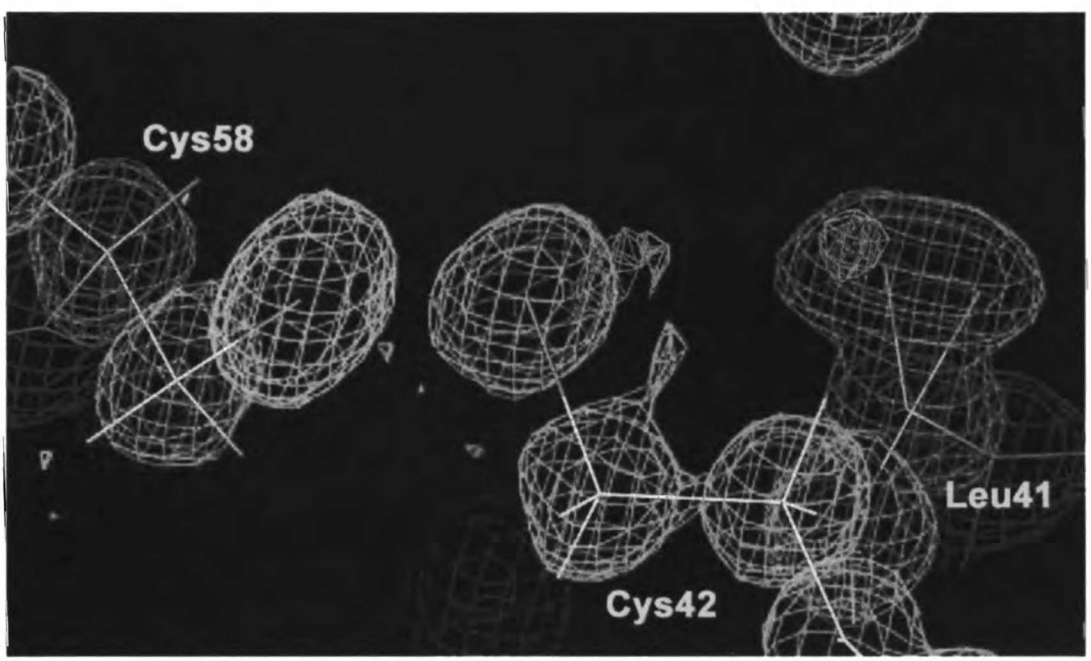


Figure 2.7. F_o-F_c electron density maps indicate no radiation damage in αLP+boroVal(gol).
$2F_o$-F_c electron density (1.5σ) is colored purple, positive F_o-F_c density colored in green (+3σ) and cyan (+4σ), and negative F_o-F_c density (-3σ) colored in red. The final model for αLP+boroVal(gol) is shown as a stick model.

Although this strategy was successful with αLP+boroVal(gol), later attempts to repeat the experiment with other boronic acid inhibitors or with PEG-300 as a cryoprotectant were unsuccessful. Although the reason for this disparity is unclear, we expect that the likely cause is variation in environmental and/or experimental conditions at the time of data collection, rather than the chemical variation introduced by using other inhibitors or a different cryoprotectant. Care was taken to use similar experimental parameters in later attempts. However, beamline conditions are intrinsically variable due to continually-changing ring current and updates made to hardware and software at the beamline. To ensure that a similar doseage is applied, future attempts should focus on matching the intensities and signal-to-noise ratios for a given resolution range in the successful αLP+boroVal(gol) data sets. For these tests to be meaningful, crystals of the same width of αLP+boroVal(gol) should be used. In addition to total dose, dose *rate* is difficult to control for. Previous studies have provided conflicting results regarding the importance of dose rate on the prevalence of damage (Leiros et al., 2001; Sliz et al., 2003). Future efforts should focus on tests of various experimental conditions, including x-ray energy (i.e., wavelength), flux/time of exposure (affected by slit width and wavelength), and area of exposure (i.e., collimator size). Other variables may include size and condition of the crystal used. Collecting data at lower cryogenic temperatures may also help to reduce radiation damage due to secondary ionization effects (Teng & Moffat, 2002). If possible preliminary tests should be performed consecutively in one day at the synchrotron, and F_o-F_o maps calculated as data is collected.

Extrapolating data to zero-dose: Isolation of a TI_2 mimic

Since subsequent attempts to collect ultra-high resolution data for αLP+boroVal (a mimic of TI_2) without radiation damage were unsuccessful, a second strategy was utilized: to allow damage to occur in a controlled fashion while collecting very highly-redundant data, and to then use computational methods to extrapolate the intensities of each reflection back to zero-dose. This method, published by Diederichs, *et al.* (2003) (Diederichs et al., 2003), relies on the assumption that radiation damage is two-state and proceeds linearly over time (see Methods).

Intrinsic to the method is the ability to calculate a good estimate for the decay rate (β) of each reflection. This requires (1) a large and broad sample set of any given unique reflection (achieved by collecting highly-redundant data spread over a large accumulated dose) and (2) that every unit cell is exposed uniformly throughout the experiment (by using a crystal that is smaller than the size of the beam). For αLP+boroVal, the former was achieved by collecting 275° of data (344 frames), yielding an average redundancy of 16.7. The second requirement was more challenging to fulfill. Previous experiments had suggested that fairly large crystals (typically wider than 0.2 mm) are required to obtain ultra-high resolution data for αLP (data not shown). Faced with a limited supply of crystals, we collected data on an αLP crystal with a width of ~0.3 mm. In order to increase the area of exposure on the crystal, the beam was focused at its greatest width using a 200μm collimator, and centered across the width of the crystal. As a conservative first attempt of this method, the exposure time was estimated with the intent to obtain ~1Å resolution data.

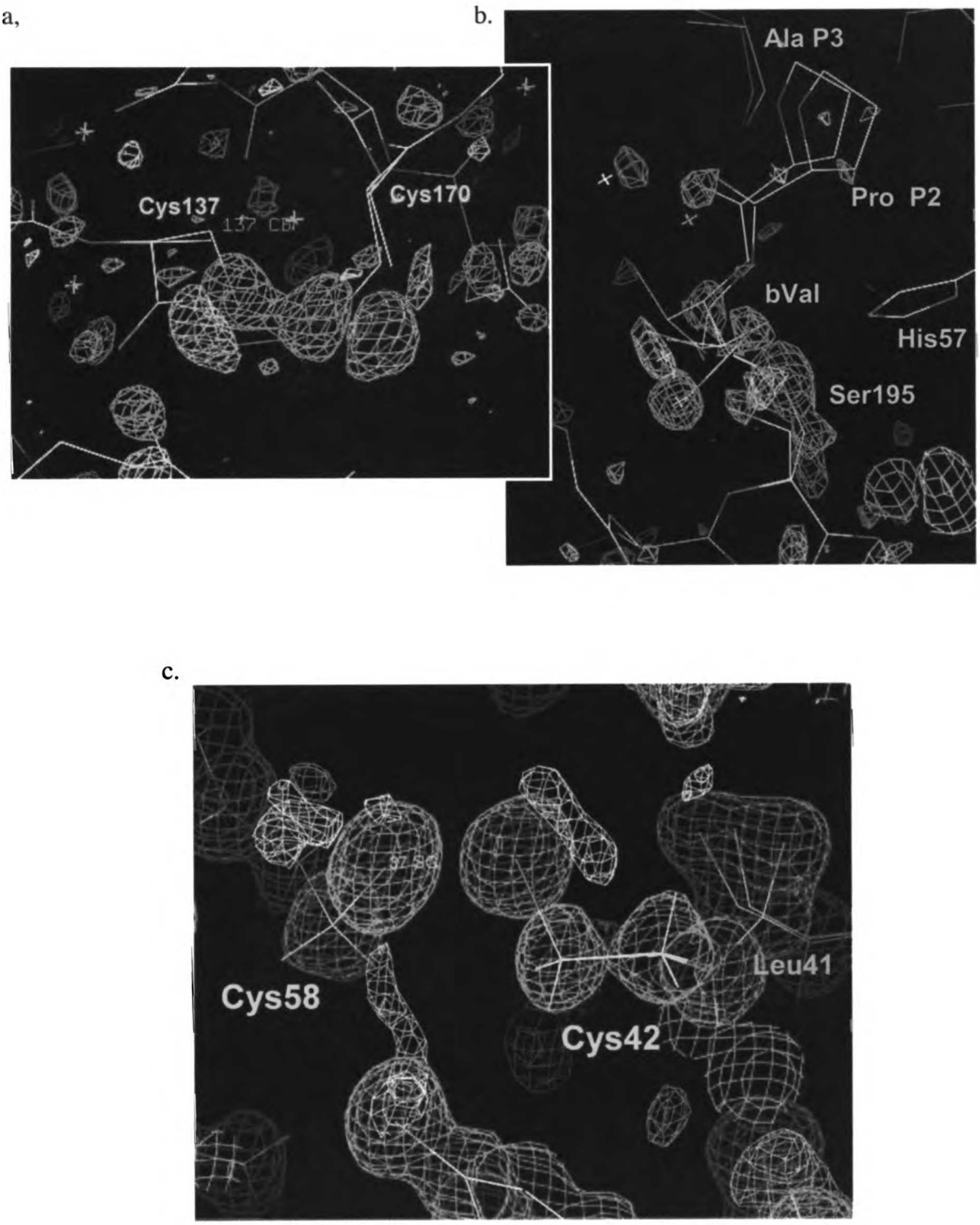
There was no indication of radiation damage in the overall data statistics. If radiation damage had occurred, this should be manifest as time-dependent changes in intensity for particular reflections. This can be detected globally by merging data collected early (low dose) with data collected at increasingly later stages, and calculating R_{merge} for these data combinations. Surprisingly, there was not an increasing trend for R_{merge}, giving no global indication of radiation damage (K. Diederichs, personal communication). There was also no indication of decreasing resolution during data collection as a function of dose. Extrapolation to zero-dose did result in a decrease in R_{merge}, although slight. These data would suggest that extrapolation was unnecessary.

The true test of the effectiveness of the extrapolation, however, lay in the electron density maps. $F_{o, corr} - F_{o, un}$ difference electron density maps were calculated to compare the extrapolated data ($F_{o, corr}$) to the uncorrected data ($F_{o, un}$), using phases from αLP+boroVal* (Figure 2.8). These maps were nearly identical to the $F_{o, after} - F_{o, before}$ difference maps calculated for αLP+boroVal* (Figure 2.2). The zero-dose extrapolation clearly reduced the occupancy of alternate conformations that had resulted from radiation damage. Furthermore, upon refinement of a preliminary model, $F_o - F_c$ difference electron density maps gave no indication of radiation damage (Figure 2.8c), suggesting that the extrapolation calculation had successfully eliminated the effects of radiation damage.

Figure 2.8. Zero-dose extrapolation was successful in reducing radiation damage. (a, b) $F_{o, corr} - F_{o, un}$ difference maps comparing data pre- and post-extrapolation. Positive density (+3s; green) indicates regions of more electron density following zero-dose extrapolation. Negative density (-3s) is colored magenta. (c) Refinement of αLP+boroVal against the extrapolated data indicates no radiation damage in F_o-F_c difference maps. $2F_o$-F_c electron density (1.5σ) is colored purple, positive F_o-F_c density colored in pink (+2σ) and green (+3σ), and negative F_o-F_c density (-3σ) colored in red.

(Figure 2.8 shown on next page.)

Figure 2.8 (caption on previous page)



Conclusion

Here we show that the long x-ray exposure time required to obtain ultra-high resolution data results in clearly observable radiation damage in αLP crystals. Although not affecting global data statistics such as R_{merge} or resolution, comparison of low-resolution data collected before and after the long-exposure high-resolution data revealed specific structural changes in $F_{o,\,after}$-$F_{o,\,before}$ maps. Effects of radiation damage included reduction of all three disulfide bonds, with Cys42/Cys58 and Cys189/Cys220A exhibiting reduction at the highest occupancy (up to ~34%). A large region of the protein extending between Cys42/Cys58 and Cys189/Cys220A demonstrated multiple conformations, with particularly prominent changes occurring at the Ser195-boroValP1 adduct in the active site. One possible explanation for this conformational disorder is that primary or secondary ionization events caused breakage of the boroValP1 Cα--B bond, resulting in a trigonal planar boronic acid structure at Ser195. While this is likely to occur, additional studies are required to definitively assign structural models to all of the conformations of Ser195 and boroValP1 observed in the damaged enzyme, and to determine whether these structural changes are a direct or indirect result of radiation damage. Collectively, these studies provide a strong reminder that high-resolution data collection should be sandwiched between short-exposure low-resolution data sets (before and after high-resolution data collection) whenever possible, to allow accurate assessment of radiation damage via $F_{o\,after} - F_{o\,before}$ difference electron density maps and refinement.

In order to obtain high quality structural information, two different data collection strategies were explored to minimize effects of radiation damage. In the first method, high-resolution data collection was spread over several positions on the crystal. In one case, this resulted in a complete data set with essentially no visible radiation damage. This structure (αLP+boroVal(gol)) was further analyzed in Chapter 3. The second method used highly-redundant data and a linear extrapolation method to estimate the intensity of each reflection at zero radiation dose. This method was successful in eliminating all effects of radiation damage, as evident in F_o-F_o and F_o-F_c maps, and should be pursued in future studies.

Methods

Where possible, methodology was described in detail within the Results section of this Chapter. The section below provides additional detailed information, where appropriate.

αLP was expressed, purified, and crystallized as described in Appendix B-D. The inhibitor, MeOSuc-Ala-Ala-Pro-(D,L)-boroVal-OH, was dissolved in 20mM Tris·sulfate at pH 7.5 to a final concentration of 0.15M, and was added to αLP crystals as described below. Unless otherwise noted, data were indexed in Denzo (Otwinowski & Minor, 1997), scaled and merged in Scalepack (Fox & Holmes, 1966), and models refined using SHELXL-97 (Sheldrick & Schneider, 1997). Figures were made using ChemDraw Pro and O (Jones et al., 1990).

Data collection and refinement for "damaged" data

Data was collected for αLP+boroVal* (αLP bound to MeOSuc-Ala-Ala-Pro-boroVal, showing radiation damage) following typical procedures followed by this author when collecting ultra-high resolution data for αLP (summarized in Table 2.3). Data was collected at Beamline 8.2.2 at the Advanced Light Source, using a crystal of approximate dimensions 0.3 x 0.3 x 0.7 mm and wavelength of 0.785Å. Prior to data collection, the αLP crystal was incubated with 1 μL of 0.15M MeOSuc-Ala-Ala-Pro-(D,L)-boroVal overnight (15.5 hours). The crystal was looped through 20% PEG-300 cyroprotectant, then mounted directly on the cryo-stream (temperature 98 K).

A complete low-resolution data set was collected before and after collection of the high-resolution data. In the initial low-resolution data set (1.22Å resolution), 70 frames of data were collected, each with an exposure time of 3 seconds over 1° oscillation angle. This data set was denoted "#3", and the model of αLP+boroVal refined against it called "αLP+boroVal$_3$". The beam was then moved to a new position on the crystal, and high-resolution data collected (30 second exposures over 36 frames). The oscillation angle used in this first high resolution data set was two wide (with too many overlapped reflections), so the data collection was stopped and a narrower oscillation angle (0.7°) used to collect the final 36° of data. The first wedge of high-resolution data was then re-collected, and spread over this and another position on the crystal.

Following collection of this high-resolution data, two low-resolution data sets were collected (to 1.4Å resolution, with exposure times of 1-2 seconds per frame), and finally a medium-resolution data set was collected to 1.1Å resolution (set "# 11"; 4 seconds of radiation for each of the 70 frames). These three short-exposure data sets were collected at the position on the crystal where most of the high-resolution data had been collected, allowing analysis of radiation damage that had occurred.

In order to test for radiation damage, $F_{o,\,after} - F_{o,\,before}$ difference maps were calculated (where $F_{o,\,before}$ were represented by data set #3 and $F_{o,\,after}$ were represented by data set #11). F_o-F_o maps were calculated using CNS (Brünger et al., 1998). Data sets were scaled together, and maps calculated using phases derived from the 0.83Å resolution model of the free enzyme (1SSX.pdb (Fuhrmann et al., 2004a)).

Table 2.3. Data collection parameters for αLP+boroVal*

Data Set	Low	Ultra-high				Low	Lower	Medium
	3	4	5	6	7	8	10	11
Beam position on crystal	1	2	2	3	1 or 2?	Same as set 7	Same as set 7	Same as set 7
No. of Frames	70	36	52	24	28	46	70	59
Resolution (Å)	20-1.22	2-0.89	2-0.88			30 - 1.4	30 - 1.4	30 - 1.1
Time (s)	3	30.0 - 31.2	31.5 - 33.2	33.2 - 34.2	35.0 - 36.0	2-1.7 (6.6s for frame #46)	1.0	4.0 - 4.1
Distance (mm)	200	106.9	106.9	106.9	106.9	230	230	150
$\Delta\phi$ (°)	1	1	0.7	0.7	0.7	1.5	1	1
ϕ (°)	0 – 70	0 – 36	34 – 70.4	0 – 16.8	16.8 – 36.4	0 – 69	0 – 70	0 – 70
λ (Å)	0.785							

In order to test for radiation damage occurring during low-resolution data collection, a model of αLP+boroVal was refined against the low resolution data set #3. This model is denoted "αLP+boroVal$_3$". The 0.90Å structure of αLP complexed with MeOSuc-Ala-Ala-Pro-boroAla (with glycerol bound) was used as a starting structure (see Appendix E). In order to minimize model bias, alternate conformations, the inhibitor, glycerol molecules, and partially-occupied waters were removed, as well as sulfate ions with occupancy < 50% and water molecules with B-factors > 50. Refinement was performed in SHELXL-97 (Sheldrick & Schneider, 1997). Ten percent of the reflections were reserved for calculation of R_{free}. Rigid body refinement was performed initially against data to 1.5Å, resulting in an R / R_{free} of 23.4% / 23.8%. Water molecules with negative difference density were removed, and all atoms in the model were minimized by conjugate-gradient least squares refinement against all data (R / R_{free} = 22.3% / 25.6%). Subsequent rounds of refinement included modeling waters and the inhibitor and adding alternate conformations (R / R_{free} = 20.4% / 23.6%). Anisotropic B-factors were added to the model, reducing R-factors by > 4%. Since only a preliminary model was required for the purposes of this study, refinement was ceased at R / R_{free} of 14.2% / 18.5%. This model contained 7 residues with alternate conformations, 246 fully- and 21 partially-occupied waters, and 2 sulfate ions.

In an attempt to model radiation damage occurring during high-resolution data collection, the model was refined against all data. This model is denoted "αLP+boroVal*". All data collected (except the first portion of high-resolution data) were scaled and merged, with a limiting resolution of 0.89Å. Five percent of the reflections were reserved for calculation of R_{free}; care was taken to ensure that no reflections that had been used in the

working set during refinement of αLP+boroVal$_3$ were included. The starting model used for refinement was the final model of αLP+boroVal$_3$. Twenty-five cycles of minimization against data to 1.2Å resulted in an R / R_{free} of 13.7% / 17.8%. Data was added to 1.0Å, and solvent and alternate conformations added to the model (R / R_{free} = 12.7% / 15.2%). Data was added to 0.89Å and riding hydrogen atoms added to the model (R / R_{free} = 12.6% / 15.0%). Hydrogen atoms were not added to $-CH_3$, -OH, $-NH_3$, or to key atoms in the structure, including His57, Gly193 N, Ser195 N, AlaP3 N, Ala P4 N. Discouraged by unsuccessful attempts to model the Ser195-boroValP1 adduct, refinement of this structure was not completed. The "final" model has an R / R_{free} of 12.5% / 14.8%, and contains 24 residues with alternate conformations, 1,111 hydrogen atoms, 236 fully- and 95 partially-occupied waters, and 4 sulfate ions. In addition to some protein conformations needing to be modeled, R-factors are high largely because of the relatively low number of waters included in the model.

Another strategy for modeling radiation damage was to refine the model against data containing the most amount of damage (data set #11). This model is labeled "αLP+boroVal*$_{11}$". Ten percent of the data were reserved for R_{free}; reflections of resolution > 1.22Å that had been flagged in αLP+boroVal$_3$ were once again flagged here, in order to minimize model bias. αLP+boroVal$_3$ was used as a starting model. From this model, various conformations of Ser195 and boroValP1 were tested as candidates to account for the F_o-F_c difference density observed.

Collecting data with minimal damage: αLP+boroVal(gol)

In order to form the αLP+boroVal complex, one microliter of a 0.14M solution of MeOSuc-Ala-Ala-Pro-(D,L)-boroVal-OH was added to a drop containing αLP crystal, and allowed to incubate for 17.5 hours. Prior to mounting on the goniometer, the crystal was equilibrated in a cryoprotectant containing 15% glycerol for 45 seconds. The crystal was then frozen directly in the stream of gaseous nitrogen, which was maintained at a temperature of 100 K throughout data collection. Data was collected at the Advanced Light Source (HHMI beamline 8.2.2), using x-rays of wavelength 0.75Å. Table 2.4 provides an example of parameters for data collection at a single position on the crystal (shown here for position 5).

Table 2.4. Characteristic data collection parameters for αLP+boroVal(gol)

	Low_{5-1}	$High_{5-2}$	Low_{5-3}
Frames (#)	35	17	35
Resolution (Å)	1.6	0.9	1.6
Time (s)	4	70	4.2
Distance (mm)	270	120	270
$\Delta\phi$ (°)	2	0.7	2
ϕ (°)	220–290	252.9–264.8	220–290
λ (Å)	0.750	0.750	0.750

Parameters are listed for data collection at position 5 on the crystal (see Figure 2.6a).

Data sets were indexed using HKL2000 (Otwinowski & Minor, 1997). High-resolution data sets were scaled and merged with two representative low-resolution data sets using Scalepack (Fox & Holmes, 1966). The resulting data set was limited to a resolution of 0.90Å (see Table 2.2).

Extrapolating data to zero-dose

A detailed description of this method if provided in Diederichs, *et al.* (2003). Briefly, in any given data set, each unique reflection (*hkl*) will have an intensity of y_{ij} that is measured on multiple frames (denoted by index *j*), and in each frame measured multiple times (index *i*). When radiation damage occurs, such damage is reflected as a change in y_{ij} that is linear under conditions of constant radiation dose. The linear relationship describing the change in intensities (y_{ij}) of a given *hkl* as a result of accumulated dose (x_{ij}) has a slope of β. A linear extrapolation can then be used to calculate the predicted intensity for *hkl* at x=0:

$$y_{ij}^{zero} = y_{ij} - \beta x_{ij}$$

with an associated error of:

$$\sigma_{ij}^{zero} = |\sigma_{ij}^2 + (\sigma_\beta x_{ij})^2|^{1/2}$$

Data collection for zero-dose extrapolation

This experiment was performed on an αLP crystal that had equilibrated for 66 days with one microliter of 0.15M MeOSuc-Ala-Ala-Pro-(D,L)-boroVal-OH. In our experience, the αLP-inhibitor complex is stable in the crystalline context for long periods of time. The crystal was allowed to equilibrate in 20% PEG-300 for 20 seconds prior to mounting on the goniometer. Data were collected at ALS beamline 8.2.2 using a wavelength of 0.785Å.

Data was collected in dose mode (automatically adjusting exposure time to account for beam decay), and the collection ran uninterrupted for the first 344 frames of data. The flux of the beam was measured before and after the high-resolution data collection, so that total dose could be calculated. Then, two more sets of high-resolution data were collected. Because the high-resolution data sets provided adequate low-resolution data, additional collection of low-resolution data was unnecessary. Data collection parameters are listed in Table 2.5.

Only the first 344 frames of data were used in this study. Prior to the extrapolation, all data were indexed and scaled (but not merged) using XSCALE from the XDS suite (Kabsch, 1993). The zero-dose correction was then performed on each reflection using the O-dose option built into XSCALE (Table 2.6) (Diederichs et al., 2003). This corrected data was once again scaled and then merged to provide a final complete data set.

Table 2.5. Data collection parameters for the zero-dose extraplation method (αLP+boroVal)

	High-resolution data set
Frames (#)	344
Resolution (Å)	1.04
Time (s)	8 – 8.6
Distance (mm)	150
$\Delta\phi$ (°)	0.8
ϕ (°)	0 – 275.2
λ (Å)	0.785
Collimator	200 μm

Table 2.6. Statistics from the zero-dose extrapolation (as output by XDS/XSCALE).

Resolution Limit	N(unique)	N(corr)	N(corr) / N(error)	σ(I) (corr) / σ(I) (uncorr)	N(free)
2.71	5666	3764	22.9	0.836	54674
2.15	5502	3449	20.9	0.867	50125
1.88	5365	3201	18.4	0.882	46457
1.59	10545	4956	15.5	0.914	73152
1.42	10627	4553	13.7	0.925	67223
1.36	5084	1856	13.4	0.933	27381
1.3	6063	2030	13.1	0.938	29927
1.26	4628	1512	13.3	0.939	22270
1.22	5345	1649	13.1	0.94	24451
1.19	4428	1336	12.8	0.942	19781
1.15	6617	1892	12.6	0.946	28045
1.13	3692	989	12.8	0.947	14402
1.1	5989	1583	12.7	0.952	23433
1.08	4386	1086	13.5	0.952	15968
1.06	4668	1106	13.3	0.957	16254
1.04	5078	1144	14.4	0.959	16898
TOTAL	93683	36106	15.1	0.913	530441

N(unique) = Number of unique reflections with enough symmetry-related observations to determine a decay factor b(h)

N(corr) = Number of 0-dose extrapolated unique reflections

N(error) = Number of unique extrapolated reflections expected to be overfitted. A large ratio of N0-DOSE/NERROR justifies the data correction as carried out here.

σ(I) (corr) = mean value of Sigma(I) for 0-dose extrapolated data

σ(I) (uncorr) = mean value of Sigma(I) for the same data but without 0-dose extrapolation.

N(free) = degrees of freedom for calculating S_corr

Calculation of radiation exposure during data collection

An estimate of the incident energy (E, in Joules) on the surface of a crystal can be calculated using the following equation:

$$E = nte\lambda$$

where n is the flux of the beam (number of photos per second), t is the time of exposure (seconds), e = 1.6x10^{-19} J/eV, and λ is the energy of the beam (dependent on the wavelength used: 15,800 eV for λ = 0.785Å). Because the flux of the beam is dependent on the slit settings, synchrotron ring current, and wavelength used, the flux had to be recorded at a time reasonably close to the time of data collection. For αLP+boroVal, intensity of the beam was measured by a photodiode and flux calculated using the script "flux.com," written by Dr. James Holton for ALS Beamlines 8.2.1, 8.2.2, and 8.3.1. It should be noted that in this case "flux.com" measured the flux of the x-ray beam after passing through the crystal (0.85 x 10^{15} photons/second). Therefore, the energy calculated here is an underestimation of the true energy of incidence. However, because the vast majority of photons pass through the crystal (with much lower incidence of absorbance at high energy, see Figure 2.5), this estimation should be a good approximation. It should also be noted that the "flux.com" calculation is also an estimation, based on estimates of the beam intensity distribution, beam aperture size, etc (Nave & Garman, 2005). Unfortunately, the flux was not measured for other data collections, so we cannot make a direct comparison of energy dose absorbed by other crystals in the experiments described in this Chapter.

Postscript

The structure of αLP+boroVal(gol) was refined to completion; results of its analysis are described in Chapter 3.

Unfortunately, the data for αLP+boroVal (which was extrapolated to zero dose) was not sufficient for answering the detailed catalytic questions of my thesis. Hydrogen atoms in the active site, particularly those on His57, which were of most interest, were visible at best as weak peaks in the F_o-F_c density. This may be a result of the lower-resolution data (1.04Å) acquired during the experiment. Another concern was the increase in noise in the electron density maps, as compared to ultra-high resolution data sets collected. In addition, refinement of this structure converged prematurely (despite difference maps indicating additional modeling was needed in the solvent regions) and with a large (3%) difference in R and R_{free}. It is possible that error was introduced during the extrapolation, likely due to an inaccurate determination of the decay factor, β. Factors that may have contributed to the inaccuracy of β include (1) changes to *y* that were too small to detect relative to their error in measurement, and (2) using a crystal of greater width than the beam diameter. This method should certainly be tried again in the future, this time using a smaller crystal, and with longer exposure per frame (which would provide higher resolution data as well as a stronger measurement of radiation damage).

Chapter 3: Sub-Ångstrom resolution crystal structures of a-lytic protease transition state analogs elucidate the role of short hydrogen bonds in serine protease catalysis

Preface

For both structures presented here, I collected the data, refined the structures, and analyzed the results. I expressed/purified the protein and grew the crystals for the αLP+boroVal(gol) structure. Rotation student Caleb Bashor expressed and purified the αLP protein used for the $\alpha LP_{pH\,5}$ structure. The protein was concentrated and crystals at pH 5.1 grown by rotation student Matthew Daugherty. The original purpose of these low-pH crystals was to aid in isolation of the Michaelis-Menten (ES) complex. Although initial attempts towards that goal were unsuccessful, the low-pH crystals were put to good use as a transition state analog in the study that I pursued here.

The results in this chapter will be submitted for publication in Fall 2005.

Summary

In order to address questions regarding the specific mechanism of catalysis by serine proteases, we have solved the x-ray crystal structures of two transition state analogs of α-lytic protease (αLP) at ultra-high resolution. In the first structure, αLP at pH 5 ($\alpha LP_{pH\,5}$; 0.82Å) the catalytic histidine is doubly-protonated, as it is in the transition states for the proteolytic reaction. We also solved the structure of αLP bound to the potent peptidyl boronic acid inhibitor, MeOSuc-Ala-Ala-Pro-boroVal ("αLP+boroVal(gol)"; 0.90Å). In αLP+boroVal(gol), the boron of the inhibitor is covalently bound to the catalytic serine with tetrahedral geometry. Unexpectedly, a glycerol molecule formed a covalent adduct with the boronate oxygen, mimicking the substrate leaving group and providing the first ultra-high resolution model of the tetrahedral intermediate for the acylation reaction (TI_1). Both models show that the His57-Asp102 hydrogen bond in the catalytic triad is a normal ionic hydrogen bond, and not a low-barrier hydrogen bond (LBHB) as previously hypothesized. In αLP+boroVal(gol), there is no evidence of, or requirement for, histidine-flipping during the acylation step of the reaction. Rather, we propose that the enzyme has evolved a network of relatively short hydrogen bonds in the vicinity of the active site that collectively stabilize the transition states. In particular, we have identified a short ionic hydrogen bond (SIHB) between His57 Nε2 and the mimic of the substrate's leaving group in αLP+boroVal(gol) that likely plays a key role in promoting forward progression of the TI_1 → acyl-enzyme reaction. Identification of this SIHB provides experimental evidence that a short donor-acceptor distance and downfield ^1H chemical shift alone cannot be used as sole indicators of a LBHB.

Introduction

The chymotrypsin family of serine proteases, the largest class of proteases in nature, utilizes the conserved catalytic triad (Asp102, His57, and Ser195) to catalyze hydrolysis of peptide bonds. Although the general mechanism for serine protease catalysis (Figure 0.2) is well established (as described in the Introduction to this thesis), some critical questions remain. Perhaps the most heated debate has focused on what factors are most critical for transition state stabilization in the acylation reaction. In 1994, Frey *et al.* proposed that in the transition states and tetrahedral intermediate for acylation, His57 and Asp102 form an especially strong hydrogen bond, called a low-barrier hydrogen bond (LBHB; Figure 3.1a) (Frey et al., 1994; Frey, 2004). Because a LBHB can store 10-20 kcal/mol of energy (Cleland & Kreevoy, 1994), Frey hypothesized that formation of such a hydrogen bond would promote formation of the transition state by stabilizing the Asp-His diad and increasing the basicity of His57, enhancing catalysis of the first step of the acylation reaction. Others argue against a special role for a LBHB between His57 and Asp102, hypothesizing that pre-organization of the active site environment, utilizing standard electrostatic and van der Waals interactions, is primarily involved in transition state stabilization (Ash et al., 1997; Kollman et al., 2001; Warshel, 1998).

In a standard hydrogen bond, the probability of a hydrogen atom being located on the acceptor atom or the donor atom is determined by the depth of the energy well at each of these positions, with an energy barrier restricting hydrogen atom exchange (for illustration of this, see GarciaViloca et al., 1997a). By contrast, in a LBHB, the energy barrier of hydrogen

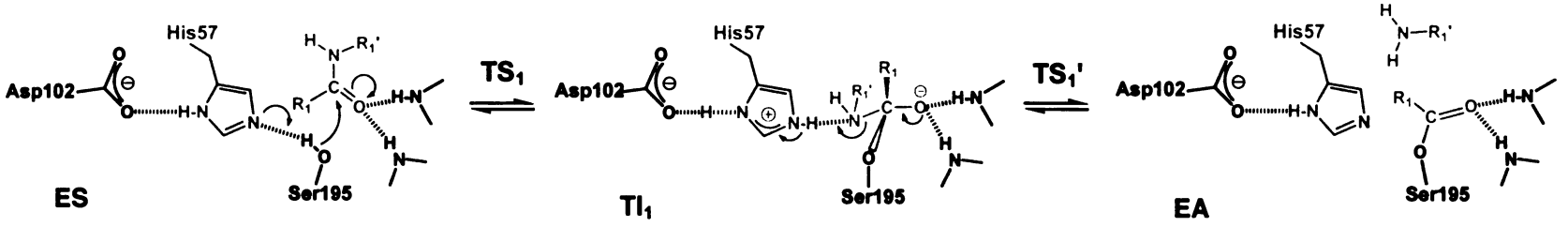


Figure 3.1. The low barrier hydrogen bond hypothesis.
This simplified schematic illustrates the mechanism for the first stage of proteolytic catalysis, acylation. The general base in the reaction, His57, is highlighted in blue; the substrate is shown in red. The reaction passes through two tetrahedral transition states (TS_1 and TS'_1). The low barrier hydrogen bond hypothesis predicts that a LBHB is formed between His57 and Asp102 in the transition states and TI_1. If this were true, the hydrogen atom between these two residues would be shared equally by donor and acceptor (illustrated here in TI_1).

transfer is lower than the first vibrational energy level for the hydrogen. For this special case, quantum theory calculates that, in contrast to a standard hydrogen-bond where the hydrogen is located on the donor atom, in a LBHB the hydrogen atom has the highest probability of being located equidistant between its two parent heavy atoms (GarciaViloca et al., 1997a). Formation of a LBHB requires a donor-to-acceptor distance of less than 2.65Å for a nitrogen-oxygen pair such as His57 and Asp102 (Frey et al., 1994). This short distance between heavy atoms, paired with an equidistant position of the hydrogen atom, constitutes a widely used diagnostic for a LBHB. Because ultra-high resolution x-ray crystallography can both (1) precisely determine the position of heavy atoms within 0.01Å and (2) visualize hydrogen atoms in difference electron density maps, this is an excellent technique to use to identify LBHB's.

Theories such as the LBHB hypothesis are inherently difficult to address experimentally due to the transient nature of transition states. In order to gain some structural understanding of these states, analogs mimicking certain aspects of the transition states (or the metastable tetrahedral intermediates) are studied (Hammond, 1955). In 1998, Kuhn, *et al.* published the 0.78Å resolution structure of subtilisin at pH 5.9 (Kuhn et al., 1998). At this pH, the catalytic histidine was presumed doubly-protonated and positively charged, providing a partial-mimic of the transition states. This structure clearly showed a short distance separating His Nδ1 and Asp Oδ2 (2.62Å), and electron density for a hydrogen atom 1.2Å from Nδ1 (1.5Å from Oδ2). Although dubbed a "catalytic hydrogen bond" by the authors, this work has been sited as a key piece of evidence for the existence of a LBHB in serine protease transition states (Frey, 2004).

Although low pH does indeed mimic certain aspects of the transition states of the serine protease mechanism, peptide inhibitors that covalently bind Ser195 and mimic the tetrahedral intermediate states (TI) provide a more chemically-accurate description of the transition states (TS). Until now, sub-Ångstrom resolution crystallographic data has not been available for tetrahedral intermediate structures of serine proteases. However, lower (1.4Å, 1.5Å) resolution crystal structures of chymotrypsin bound to trifluoromethyl ketone inhibitors Ac-Phe-CF_3 or Ac-Leu-Phe-CF_3 have shown a distance of 2.6Å between His57 and Asp102, with an estimated coordinate error of 0.15Å (Neidhart et al., 2001). Combined with ^{1}H NMR chemical shift, deuterium isotope effect, and D/H fractionation factor data, these structures supported the LBHB hypothesis. However, although these transition state analogs do offer a good mimic of certain aspects of the TI_2 state (including the tetrahedral adduct and he oxyanion), -CF_3 is bulkier than the corresponding substrate's terminal oxygen in TI_2 or eaving group in TI_1, creating a steric clash with His57 and likely altering the conformation f His57 relative to a true TI.

Peptide boronic acids offer an excellent alternative analog of the deacylation trahedral transition state (Figure 2.1) (Bone et al., 1987; Kettner & Shenvi, 1984). The ecificity profile for inhibition by peptide boronic acids is highly correlated with that of alytic activity towards substrates (Bone et al., 1991a). Unlike the corresponding strates, these inhibitors form a stable boron-centered tetrahedral structure (K_i = 6.4 nM for OSuc-Ala-Ala-Pro-boroVal-OH bound to αLP (Kettner et al., 1988)), making these tides ideal for structural studies. Furthermore, the boronic acid moiety displays a terminal oxyl group in the S1' pocket, providing a more accurate mimic of this group than offered ifluoromethyl ketone complexes.

Decades ago, α-lytic protease's single histidine residue (a member of the catalytic triad) marked it as an ideal protease for optical spectroscopic and NMR studies. As a result, numerous kinetic, spectroscopic, x-ray crystallographic, and theoretical studies have focused on this enzyme. The wealth of knowledge available for this protease, coupled with our ability to isolate αLP crystals that diffract to sub-Ångstrom resolution, make it an ideal system in which to address these questions. In Chapter 1, we described the structure of αLP at pH 8 at 0.83Å resolution. Here we present two transition-state analog structures of αLP at ultra-high resolution: first, as a comparison to subtilisin, αLP at pH 5 ($\alpha LP_{pH\,5}$); second, αLP bound to the transition state mimetic inhibitor, MeOSuc-Ala-Ala-Pro-boroVal-OH ("αLP+boroVal(gol)"). Interestingly, the covalent binding of a glycerol molecule to the boronate in the latter structure has provided us with, to our knowledge, the best available mimic of the acylation tetrahedral intermediate. Although ultra-high resolution crystal structures have been described for chymotrypsin-like serine proteases in the apo-enyzme (Fuhrmann et al., 2004a; Wurtele et al., 2000), acyl enzyme (Katona et al., 2002; Wilmouth et al., 2001), and peptide-associated (Rypniewski et al., 2001; Schmidt et al., 2003) states, this is the first ultra-high resolution crystal structure of a tetrahedral intermediate/transition state mimic for this family. This structure has allowed us to address the existence of a LBHB in the catalytic triad and other questions regarding the precise mechanism of acylation during proteolytic hydrolysis by chymotrypsin-like serine proteases.

Results

Data for the structure of αLP at pH 5.1 ($\alpha LP_{pH\,5}$) was obtained to 0.82Å resolution, providing the highest resolution structure of αLP to date. The structure of αLP bound to MeOSuc-Ala-Ala-Pro-boroVal-OH (αLP+boroVal(gol)) was solved to 0.90Å resolution, the highest resolution structure of any serine protease bound to a transition state mimetic inhibitor. Statistics for data collection and model refinement are provided in Table 3.1. The final models were determined to high accuracy, as indicated by very low R-factors (R/R_{free} of 8.10 / 9.34% and 8.14 / 9.20%, respectively). Both models exhibit a low level of thermal disorder, with an average B-factor for all protein atoms of 5.8 and 3.5Å^2 for $\alpha LP_{pH\,5}$ and αLP+boroVal(gol), respectively. As described previously, distinct hydrogen atoms were visible on the polar atoms of several Ser, Thr, and Tyr residues, on most aliphatic and aromatic carbons (including methyl groups), and on some distinctly-ordered waters (Fuhrmann et al., 2004a). Additionally, in both structures the sidechain of Phe228 is significantly distorted. This deformation, first observed in the 0.83Å structure of apo-αLP at pH 8, is thought to be an evolved structural element important for deriving the protein's unusual kinetic stability (Fuhrmann et al., 2004a).

Data Collection Strategy for aLP+boroVal(gol)

Earlier crystallographic studies of αLP bound to boronic acid inhibitors at ultra-high resolution revealed that the Ser195-boron adduct is susceptible to radiation damage at the

high x-ray dose necessary for such resolution (see Chapter 2). Taking advantage of the large size of the αLP crystal (~0.3 x 0.3 x 0.5 mm), it was possible to spread the ultra-high resolution data collection over nine positions on the crystal, minimizing x-ray exposure acquired at any one position. A complete "low"-resolution data set (~1.2Å), having a total exposure equal to one to two frames of ultra-high resolution data, was collected at each position before and after the ultra-high resolution data collection. Several tests confirmed that no damage had occurred during the ultra-high resolution data collection (see Methods). High-resolution data sets from each position were scaled and merged with "low"-resolution data, resulting in a complete data set to 0.90Å resolution (Table 3.1).

Table 3.1. Data collection statistics for $\alpha LP_{pH\,5}$ and αLP+boroVal(gol)

	Data Set	
A. Data Statistics	αLP (pH 5.1)	αLP + boroVal
Space Group	$P\,3_2\,2\,1$	$P\,3_2\,2\,1$
Unit Cell (a, c, Å)	65.7, 79.5	65.8, 79.5
Mosaicity	~ 0.2	~0.17
Limiting Resolution (Å)	0.82	0.90
Total # Reflections	1,405,443	1,279,938
# Unique Reflections	189,525	147,100
$I/\sigma(I)$	38.7 (3.0)*	27.9 (4.8)**
% Completeness	97.9 (84.5)*	99.9 (99.5)**
R_{merge}***(Å)	4.6 (37.5)*	8.5 (44.8)**
B. Structure Refinement		
Resolution range (Å)	20 – 0.82	20 - 0.90
R (# reflections in working set)	8.10 (179,637)	8.14 (139,610)
R_{free} (# reflections in test set)	9.34 (9,467)	9.20 (7,352)
R.M.S. bond lengths† (Å)	0.026	0.016
R.M.S. angle distances† (Å)	0.046	0.045
Protein residues (# residues with alternate conformations)	198 / 34	202 / 21
SO_4^{2-} / glycerol molecules	11 / 2	8 / 5
Water (whole/partial occupancy)	148 / 435	160 / 389
Average isotropic B-factors		
Protein atoms (Å^2)	5.8	6.1
Solvent atoms (Å^2)	19.4	20.5

* Highest resolution bin for αLP (apo enzyme, pH 5.1) = 0.83 - 0.82 Å.
** Highest resolution bin for αLP + boroVal...gol = 0.91 - 0.90 Å.
*** R_{merge} (as calculated by Scalepack[74])

General Structural Features of αLP at pH 5

The structure of αLP at pH 5 ($\alpha LP_{pH\,5}$) was determined to 0.82Å resolution. At ultra-high resolution precise features of the structure can be modeled, including hydrogen atoms and low-occupancy alternate conformations of protein atoms. Thirty-four residues were modeled with alternate conformations. Four areas of the protein displayed discrete disorder continuing along the backbone atoms of two or more residues: 59-59A, 66-82 (4 residues), 201A-202, and 219A-219B. The glycine loop formed by residues 66-82 had also been observed previously in $\alpha LP_{pH\,8}$ as a candidate for multiple conformations, with highly anisotropic B-factors and visibly smeared $2F_o$-F_c electron density. Smeared $2F_o$-F_c electron density and F_o-F_c difference density (contoured at 3σ) also clearly indicated the presence of multiple conformers for the carbonyl oxygen of Leu41 in $\alpha LP_{pH\,5}$, which hydrogen bonds to a water in the S1' pocket of the enzyme. This anisotropy was modeled with a second conformer of Leu41 C and O, which refined to an occupancy of ~24%. Interestingly, the major conformation of Leu41 O is sterically occluded by the sidechain of $Cys42_b$ (with a distance of 2.35Å separating $Leu41_a$ O from $Cys42_b$ Sγ). Furthermore, difference density indicative of multiple conformations of Leu41 C=O was visible in F_o-F_o maps, suggesting that conformational disorder was caused by radiation damage during the collection of ultra-high resolution data. However, because similar anisotropy is also observed for Leu41 in αLP+boroVal(gol), radiation damage is not required for this conformational heterogeneity. And, because similar smearing of $2F_o$-F_c electron density was observed for Leu41 in $\alpha LP_{pH\,8}$,

we are also unable to draw a correlation between dynamic motion of Leu41 and protease activity.

The crystal's extended exposure to x-rays resulted in visible radiation damage, as was previously observed for $\alpha LP_{pH\,8}$ (Fuhrmann et al., 2004a). Two disulfide bonds, Cys42/Cys58 and Cys189/Cys220A, are partially reduced (25% and 20%, respectively). In $\alpha LP_{ph\,5}$, two conformations of Cys42/Cys58 are observed in the reduced form, with occupancies of 20% and 5%. $\alpha LP_{pH\,5}$'s third disulfide bond (Cys137/Cys159) was also reduced, but at low occupancy (~11%).

Comparing $\alpha LP_{pH\,5}$ and $\alpha LP_{pH\,8}$

The general structure of $\alpha LP_{pH\,5}$ is nearly identical to that of $\alpha LP_{pH\,8}$ (rmsd for main chain atoms 0.066Å; rmsd for all non-hydrogen atoms 0.259Å) (Fuhrmann et al., 2004a). As in $\alpha LP_{pH\,8}$, $\alpha LP_{pH\,5}$ contains two solvent molecules in its binding pockets: a sulfate ion (Sul202) in the active site, and a glycerol (Gol210) in the S2 pocket (see Figure 3.2a). Sul202 is present in two conformations ($Sul202_a$, $Sul202_b$), with a total occupancy of 89%. This sulfate ion forms a hydrogen bonding network in the active site, including hydrogen bonds to His57, Ser195 Oγ, Gly141 N in the oxyanion hole, Gol210, Arg192, and Arg122 (in a crystal contact). Atoms O2 and O4 of $Sul202_a$ (occupancy 76%) form electrostatic interactions with positively-charged residues His57 and Arg192/Arg122, respectively.

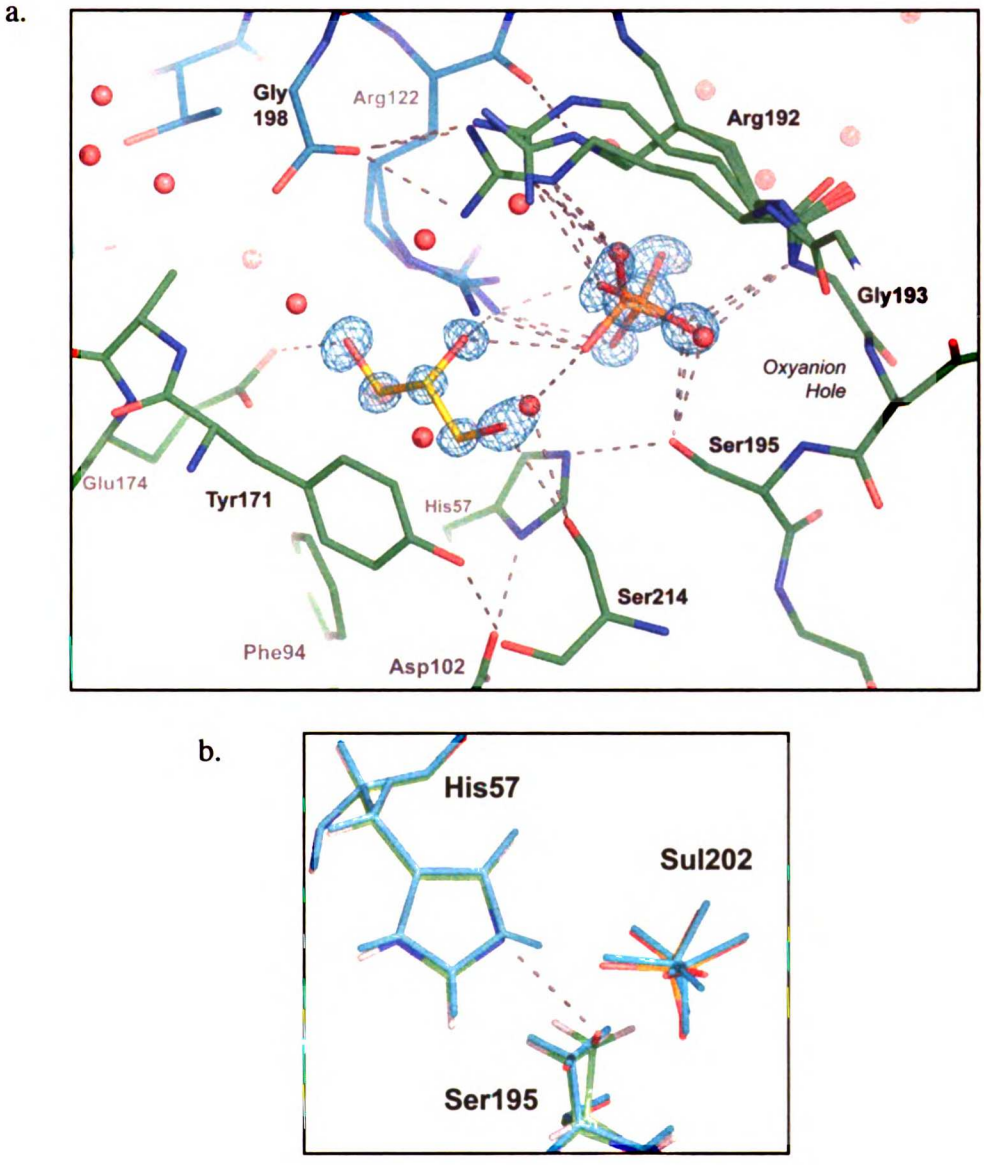


Figure 3.2. The active site of $aLP_{pH\,5}$

(a) $2F_o$-F_c electron density (contoured at 2σ) shows the position of two solvent molecules in the substrate binding pockets of αLP. Sulfate 202 occupies the region of the active site, and is stabilized by hydrogen bonding interactions with Ser195, His57, the oxyanion hole, and nearby arginine residues. A glycerol molecule occupies the S2 binding pocket. A neighboring molecule in the crystal lattice is represented in cyan, and water molecules by red spheres. (b) An overlay of $aLP_{pH\,8}$ (atoms colored by element) and $\alpha LP_{pH\,5}$ (cyan); the $\alpha LP_{pH\,5}$ structure has been offset slightly to allow visualization of both structures. In $\alpha LP_{pH\,8}$, we previously hypothesized that Ser195 occupies one of two different positions depending on the protonation state of His57 (Fuhrmann et al., 2004a). The structure of αLP at pH 5, in which all αLP molecules are protonated at His57 Nε2, confirms this hypothesis. Ser195 occupies one conformation, which corresponds to the conformation previously predicted in $\alpha LP_{pH\,8}$ to be that associated with $His57^{+}$.

Therefore, it would be expected that the two negative charges of $Sul202_a$ reside primarily on these two oxygens. Indeed, the S-O2 and S-O4 bond lengths are longer than those of S-O1 and S-O3 (1.48 and 1.49Å, compared to 1.45 and 1.43Å), despite restraints applied during the refinement to promote a bond length of 1.42Å for all sulfate bonds. We have therefore assigned partial negative charges to $Sul202_a$ O2 and O4 (Figure 3.4a). In the absence of Sul202, two water molecules maintain hydrogen-bonding interactions with Arg192 and Arg122, and with Ser195 and the oxyanion hole, respectively (Figure 3.2a).

Although $\alpha LP_{pH\,5}$ and $\alpha LP_{pH\,8}$ are nearly identical in their overall structures, notable differences do exist in the vicinity of the active site. At pH 5, His57 should be protonated at Nε2 in all αLP molecules of the crystal. Indeed, in $\alpha LP_{pH\,5}$ a peak corresponding to Nε2-H is visible in σ-weighted F_o-F_c difference maps (Figure 3.4a). Several structural changes occur in the vicinity of the active site in response to the protonation of His57 Nε2. One primary effect is that Ser195 occupies only one conformation at pH 5 (see Figure 3.2b and Discussion). In $\alpha LP_{pH\,5}$, Sul202 is present at higher occupancy than in $\alpha LP_{pH\,8}$; correspondingly, Arg192 and Arg122 both exhibit a higher occupancy for conformations that hydrogen bond to Sul202. These primary structural changes cause subtle differences in backbone positions of surrounding residues. These structural observations are confirmed in $F_{o,\,\alpha LP\,pH\,5}$ - $F_{o,\,\alpha LP\,pH\,8}$ maps (data not shown).

General Structural Features of αLP+boroVal(gol)

The boronic acid inhibitor MeOSuc-Ala-Ala-boroVal-OH inhibits αLP proteolysis by forming a covalent attachment to Ser195, providing a structural mimic of the proteolytic

transition state (Figure 3.3). During initial refinement of αLP+boroVal(gol), both positive electron density and σ_A-weighted $2F_o$-F_c density (extending from boroValP1[5] to the first carbonyl of MeOSucP5) indicated successful binding of the inhibitor in the binding pocket. Covalent linkage of the boronate to Ser195 was evident by a short distance between Ser 195 Oγ and the boron of the inhibitor, boroValP1 B (1.46Å), as well as continuous electron density extending from Ser195 Oγ to boroValP1 B (Figure 3.3b). In this structure, the sidechain of Ser195 occupies a conformation similar to that observed in $\alpha LP_{pH\,5}$. The inhibitor, bound at complete occupancy, was initially modeled as a single conformation, but difference maps indicated a second conformation for AlaP3-AlaP4 (occupancy ~21%). Three hydrogen bonds (Ser214 O... boroValP1 N, Gly216 N... AlaP3 O, and AlaP3 N...Gly216 O) stabilize the inhibitor-enzyme complex, forming a 3-stranded β-sheet (see Figure 3.3b, Table 3.2). Electron density corresponding to the N-terminal methoxysuccinyl tail (MeOSucP5) was observed extending from AlaP4 in both conformations and included in the model as minor conformations.

The short atom-atom distance and difference electron density maps clearly indicated the covalent addition of a glycerol molecule to the boronate. This glycerol (GolP1'), introduced to the crystal as a component of the cryosolvent, is present in three conformations with a total occupancy of 65%. In any one of the three conformations, the glycerol is bound via a secondary ester linkage to boroValP1 O2 (Figure 3.5). The terminal oxygen O1 of two of the conformations for GolP1' forms hydrogen bonds with Leu41 O and boroValP1 O1 in

[5] Substrate residues are labeled according to the nomenclature of Schechter and Berger (1967) where the residue N-terminal to the scissile bond is denoted P1, and is bound in the enzyme's S1 specificity pocket. Residues extending N-terminal from the cleaved bond are P1, P2, etc.; those extending C-terminally are P1', P2', etc.

a.

b.

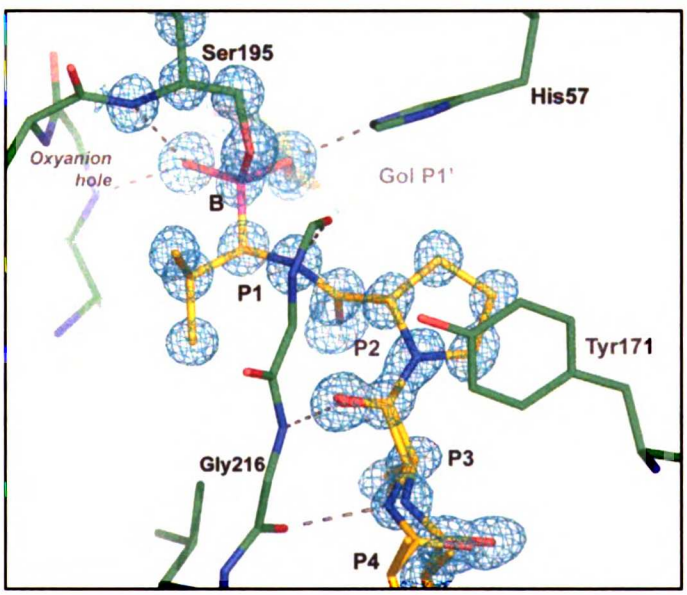


Figure 3.3. The boronic acid inhibitor, MeOSuc-Ala-Ala-Pro-boroVal-OH, binds Ser195 in the active site of αLP, mimicking the tetrahedral intermediate.
(a) Ribbon diagram of αLP, showing the position of the inhibitor (yellow) in the enzyme's binding pocket. Sidechains for the catalytic triad (Asp102, His57, Ser195) are shown in green. For simplicity, glycerol and water molecules are not shown. (b) σ_A-weighted $2F_o$-F_c electron density maps (cyan; contoured at 2σ) illustrate the covalent attachment of the boronic acid inhibitor to Ser195. αLP interacts with the boronic ester directly through three hydrogen bonds, and with the length of the inhibitor through three additional backbone hydrogen bonds. A glycerol molecule (GolP1') covalently attached to the boronate mimics the leaving group, providing a model of TI_1.

the oxyanion hole (Table 3.2). In the absence of this glycerol molecule, a water maintains both of these hydrogen bonds. A similar water forms this contact in the apo-enzyme structures. Although GolP1' does not satisfy the typical hydrogen bonds expected of a natural substrate (Bone et al., 1991b), it does lie in the S1' binding pocket of the enzyme and is chemically bonded to the tetrahedral adduct (with an O1-C bond distance of 1.45Å), providing a mimic of the substrate's leaving group. One can therefore consider αLP+boroVal(gol) as a model for the acylation tetrahedral intermediate (TI_1).

Table 3.2. Hydrogen-bonding geometries.
The geometry for key hydrogen bonds in $\alpha LP_{pH\,8}$ (Fuhrmann et al., 2004a), $\alpha LP_{pH\,5}$, and αLP+boroVal(gol) are given. D-H bond lengths are stated in cases where the position of the hydrogen atom was refined without restraints. *n/d* indicates that the position of the hydrogen atom was not refined against the data. The value provided for D-H...A angle geometry in these cases is based on a hydrogen atom whose position was fixed at optimal geometry by SHELXL-97 (Sheldrick & Schneider, 1997). Hydrogen bonds are presumed between $Ser195_b$ and Sul202 in $aLP_{pH\,8}$ based on the assumption that Oγ-H has the same position as that in $\alpha LP_{pH\,5}$. Hydrogen bonds marked with an asterix (*) are those indicated in Figure 3.7.

Table 3.2. Hydrogen-bonding geometries. (caption on previous page)

Interaction (D...A)	apo-αLP, pH 8 (Chapter 1)			apo-αLP, pH 5			αLP + boroVal		
	d(D...A) (Å)	*d*(D-H) (Å)	D-H...A (°)	*d*(D...A) (Å)	*d*(D-H) (Å)	D-H...A (°)	*d*(D...A) (Å)	*d*(D-H) (Å)	D-H...A (°)
Gly193 N ... Sul202$_a$ O1	2.820(9)	1.37(6)	162(5)	2.813(6)	1.05(4)	162(4)	-	-	-
Gly193 N ... Sul202$_a$ O3	-	-	-	3.280(8)	1.05(4)	133(3)	-	-	-
Gly193 N ... Sul202$_b$ O3	2.85(3)	1.37(6)	154(5)	2.880(18)	1.05(4)	168(4)	-	-	-
Gly193 N ... Wat760$_a$	2.90(3)	1.37(6)	142(4)	2.736(4)	1.05(4)	148(4)	-	-	-
Gly193 N ... Wat765	2.78(4)	1.37(6)	115(4)	-	-	-	-	-	-
Ser195 N ... Wat760$_a$	2.96(3)	*n/d*	141	-	-	-	-	-	-
Gly193 N ... boroValP1 O1	-	-	-	-	-	-	2.733(5)*	0.85(5)	146(5)
Ser195 N ... boroValP1 O1	-	-	-	-	-	-	2.848(5)*	0.76(5)	158(4)
Ser195$_b$ Oγ ... Sul202$_a$O1	2.64	*n/d*	*n/d*	2.640(6)	0.98(7)	164(6)	-	-	-
Ser195$_b$ Oγ ... Sul202$_b$O3	2.77	*n/d*	*n/d*	2.85(2)	0.98(7)	175(6)	-	-	-
His57 N ... Asp102 Oδ1	2.820	*n/d*	148	2.804(5)	*n/d*	149	2.819*	*n/d*	149
Ser214 Oγ ... Asp102 Oδ2	2.698(4)	0.98(7)	174(6)	2.705(4)	0.98(5)	169(5)	2.688(5)*	0.99(6)	168(5)
His57 Nδ1 ... Asp102 Oδ2	**2.770(5)**	**0.83(6)**	**170(2)**	**2.755(5)**	**0.88(5)**	**165(4)**	**2.734(5)***	**1.03(5)**	**155(4)**
His57 Cε1 ... Ser214 O	3.033(6)	1.00(6)	126(3)	3.015(5)	0.86(5)	129(4)	3.091(6)	0.91(5)	138(4)
His57 Nε2 ... Ser195$_a$ Oγ	2.826(15)	*n/d*	139	-	-	-	-	-	-
His57 Nε2 ... Ser195$_b$ Oγ	2.962(5)	*n/d*	120	3.001(4)	0.97(5)	131(4)	3.011(5)	0.90(5)	128(4)
His57 Nε2 ... boroValP1 O2	**-**	**-**	**-**	**-**	**-**	**-**	**2.645(4)***	**0.90(5)**	**164(5)**
His57 Nε2 ... Sul202$_a$ O2	2.80	*n/d*	*n/d*	2.760(5)	0.97(5)	132(4)	-	-	-
Glycerol$_a$ O1 ... Leu41$_a$ O	-	-	-	-	-	-	2.633(11)	*n/d*	*n/d*
Glycerol$_a$ O1 ... Leu41$_b$ O	-	-	-	-	-	-	2.96	*n/d*	*n/d*
Wat523$_a$ ... Leu41$_a$ O	-	-	-	2.891(8)	*n/d*	*n/d*	2.381(6)	*n/d*	*n/d*
Wat523$_b$ ... Leu41$_b$ O	-	-	-	2.506(11)	*n/d*	*n/d*	2.76(2)	*n/d*	*n/d*
boroValP1 O1... glycerol$_a$ O1	-	-	-	-	-	-	2.744(12)	0.98	144
boroValP1 N ... Ser214 O	-	-	-	-	-	-	3.110(5)*	*n/d*	164
Gly216 N ... AlaP3$_a$ O	-	-	-	-	-	-	2.934(11)*	*n/d*	160
AlaP3$_a$ N ... Gly216 O	-	-	-	-	-	-	3.10*	*n/d*	164

The αLP+boroVal(gol) model is nearly identical to the 0.83Å structure of the apo enzyme ($\alpha LP_{pH\,8}$), with an RMSD of 0.177Å for main chain atoms, and 0.530Å for all non-hydrogen protein atoms. Except for residues in symmetry-related molecules involved in crystal contacts with the binding pocket, significant structural differences are located primarily in the vicinity of the binding pocket. Binding of the inhibitor displaces the sulfate ion (Sul202) and glycerol molecule (GolP2) occupying the binding pocket in apo-αLP, causing shifts in local water networks. Main chain atoms of adjacent residues 171 and 173, forming a β-hairpin adjacent to the S2 binding pocket, move by as much as ~0.8Å upon binding the inhibitor. This is likely due to the loss of a hydrogen bond network present in the apo structure, which links GolP2 to the backbone nitrogen of Ala173 through hydrogen bonds with a water molecule.

Interestingly, upon binding inhibitor the beta-strand corresponding to residues 214-217 shifts ~0.8Å toward the inhibitor. Motion of this strand has been characterized previously for the apo- and bound forms of the enzyme, and its dynamics hypothesized to regulate substrate specificity (Bone et al., 1989b; Bone et al., 1991a; Mace et al., 1995; Miller & Agard, 1999; Ota & Agard, 2001; Rader & Agard, 1997). This collapse of the binding pocket is likely driven by formation of three hydrogen bonds between the backbone of this strand and the inhibitor (Figure 3.3b). Two of these hydrogen bonds replace interactions in the apo enzyme to a glycerol (GolP2) and partially-occupied water, yielding a gain of one hydrogen bond in the peptide-bound state. Displacement of this enzyme strand significantly reduces at Ser214 to a Cα-to-Cα distance of ~0.2Å. Nearby residues in the catalytic triad reflect a correspondingly slight displacement of 0.15-0.3Å; therefore, local

hydrogen bonds within the catalytic triad have similar geometries between the apo and TI_1 states (Table 3.2). Finally, Gly193 (forming the oxyanion hole) shifts away from the active site by ~0.4Å relative to $\alpha LP_{pH\,8}$, likely due to steric interactions with boroValP1 O1.

The Ser195-Boronate Adduct

At ultra-high resolution, it is possible to decisively describe the geometry of the boroValP1-Ser195 adduct. To our knowledge, this is the first high-occupancy structure of a tetrahedral transition state analogue to allow this analysis. In αLP+boroVal(gol), the adduct is of essentially tetrahedral geometry. Angles about the boron range from 105° to 112°(Table 3.3). This slight distortion from a perfect tetrahedron is similar to that observed in small molecule crystallography of other boron-centered tetrahedral structures (Allen, 2002; Bruno et al., 2002), and may be due to the highly heterogeneous environment of the boronate

A. Bond Length (Å)	
B-Oγ	1.466(6)
B-O1	1.485(6)
B-O2	1.519(7)
B-Cα	1.607(8)
B. Angle (°)	
Oγ-B-O1	112.3(4)
Oγ-B-O2	104.8(4)
Oγ-B- Cα	108.7(4)
O1-B-O2	107.5(4)
O1-B- Cα	111.6(4)
O2-B- Cα	111.8(4)

Table 3.3. Boronate geometries.
Bond lengths and angles are provided for Ser195 Oγ, boroVal P1 B, boroVal P1 O1 (which hydrogen bonds to Gly193 N and Ser195 N in the oxyanion hole), boroVal P1 O2 (which hydrogen bonds with His57 Nε2), and boroVal P1 Cα.

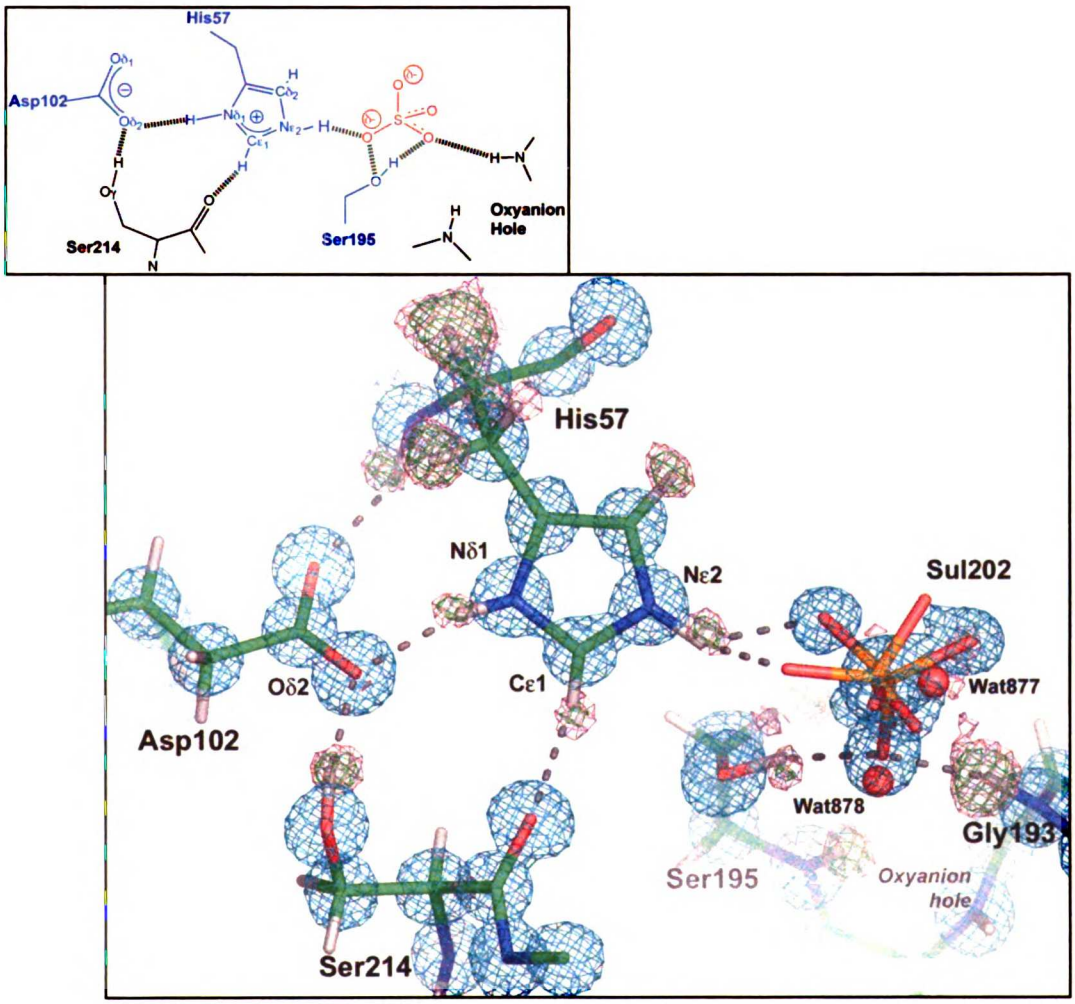
oxygens, including hydrogen bonding interactions with His57 and residues of the oxyanion hole. As expected, all three B-O bonds are approximately the same length (average 1.49±0.03Å) and are shorter than the less polar B-Cα bond (1.61Å; Table 3.3). These bond lengths are nearly identical to those seen in structures of tetrahedral boron-centered compounds solved by small molecule crystallography (Allen, 2002; Bruno et al., 2002).

Active Site Hydrogen Atoms in the Transition State Analogs

As expected for these highly-resolved structures, hydrogen atoms were visible in σ_A-weighted F_o-F_c density contoured at 2σ on the great majority of backbone amides in $\alpha LP_{pH\,5}$ and αLP+boroVal(gol). Because of the opportunity for bias to be introduced when calculating omit maps, hydrogen atoms were added to the models late in refinement, allowing visualization of hydrogen atoms as peaks in σ_A-weighted F_o-F_c difference electron density maps. In the active sites, most mechanistic protein hydrogens were visible at contour levels at or greater than 3σ before their addition to the model (see Figures 3.4 and 3.5). The positions of active site hydrogen atoms involved in hydrogen-bonding were refined without restraints; the resulting donor-to-hydrogen distances are listed in Table 3.2.

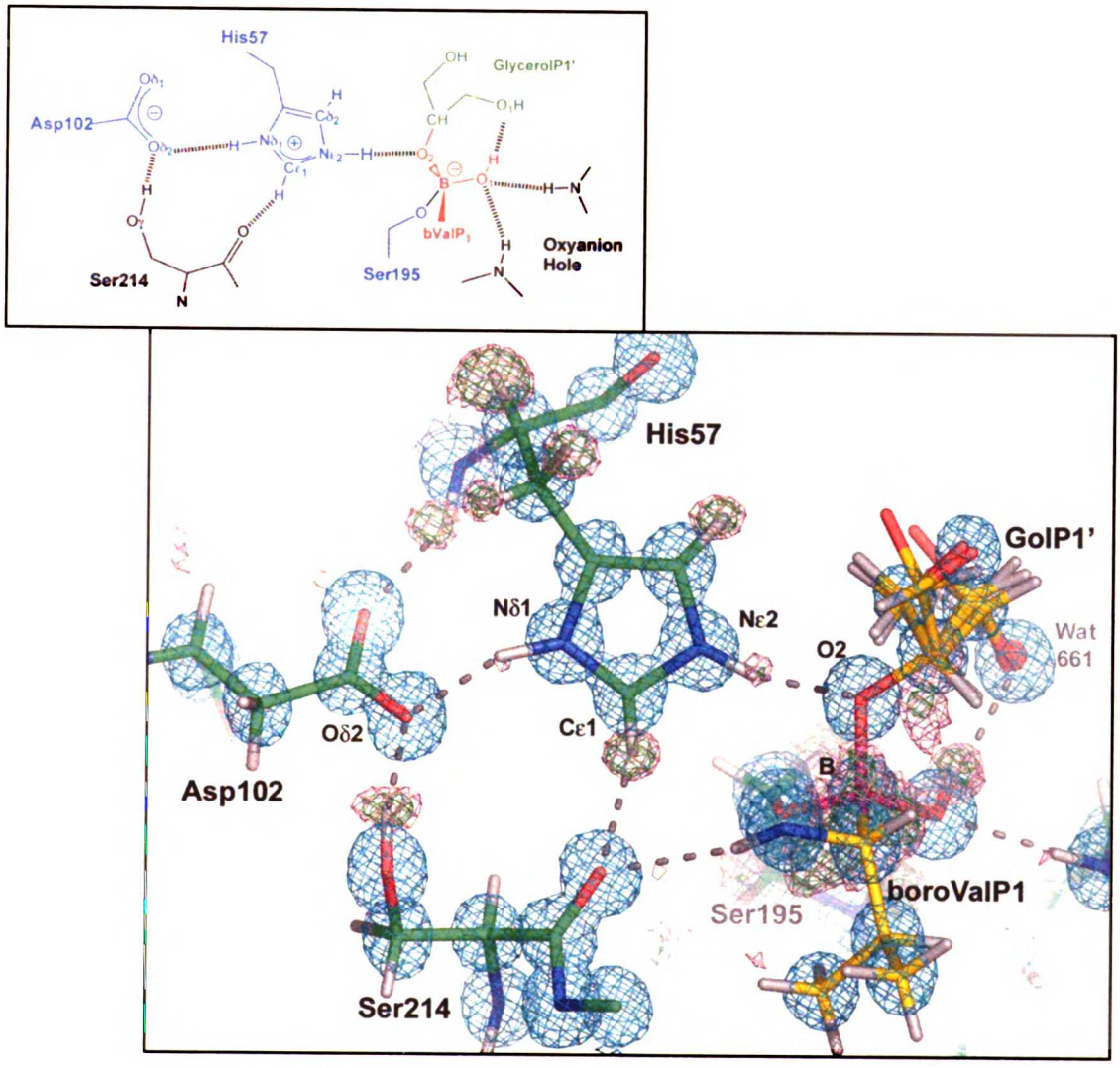
Figure 3.4. Experimentally-observed hydrogen atoms in the active site region. An illustrative schematic along with the refined model and electron density maps are shown for the active sites of (a) $\alpha LP_{pH\,5}$ and (b) αLP+boroVal(gol). For both models, σ_A-weighted $2F_o$-F_c electron density maps are drawn at 2σ (cyan), and σ_A-weighted F_o-F_c maps at 3σ (green) and 2.5σ (pink). F_o-F_c difference electron density maps were calculated prior to addition of active site hydrogens to the model (see Methods). Sticks represent the final refined models, with the positions of key hydrogen atoms refined (see Table 3.2). Dashed lines illustrate hydrogen-bonding interactions.

a.



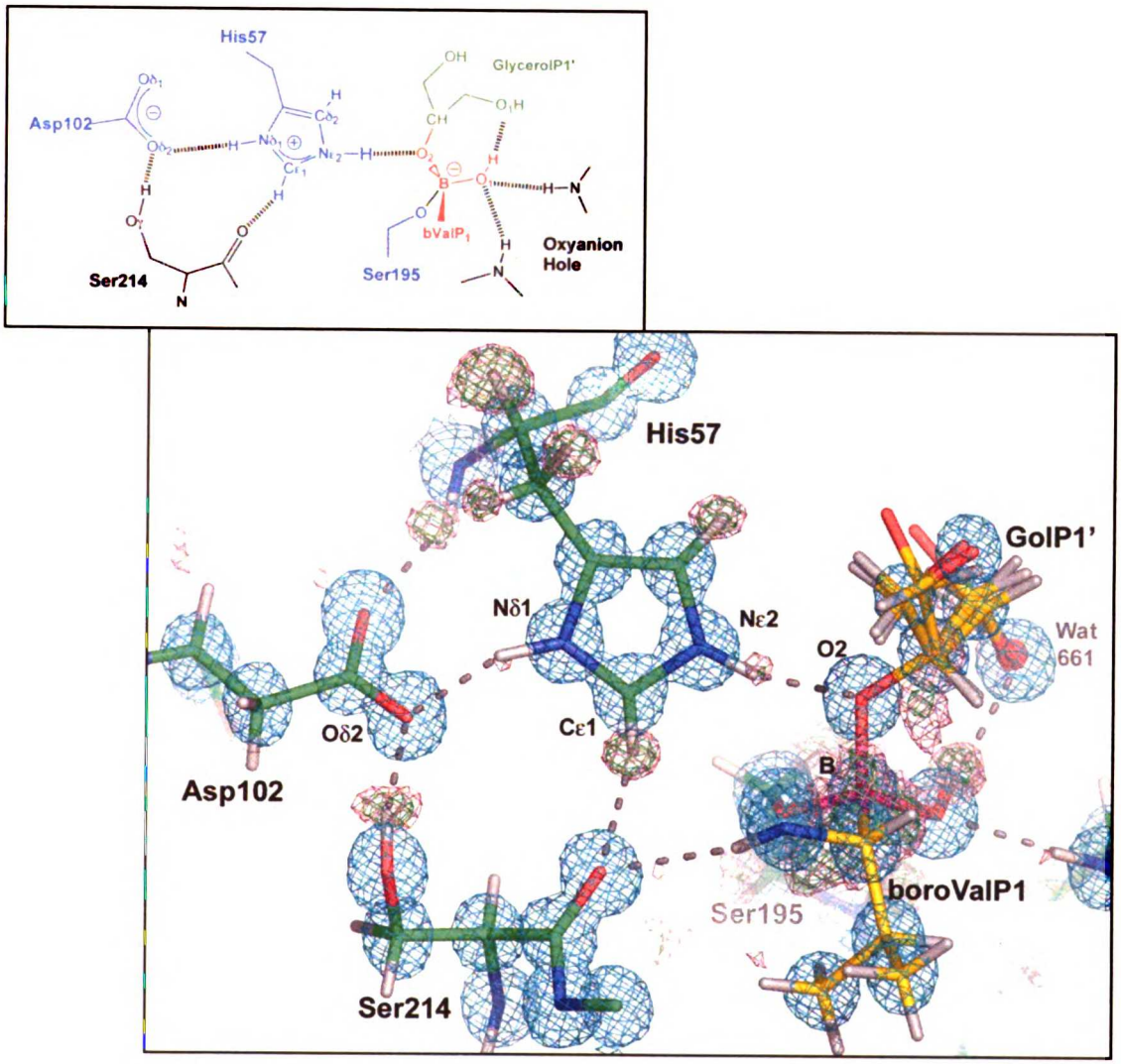
(Figure 3.4, continued)

b.



(Figure 3.4, continued)

b.



Ser195 in $\alpha LP_{pH\,5}$

In $\alpha LP_{pH\,5}$, Ser195 occupies a single conformation: that expected for Ser195 in the presence of a doubly-protonated His57 (Fuhrmann et al., 2004a). Interestingly, a hydrogen atom was visible on Ser195 Oγ in σ_A-weighted difference electron density maps (Figure 3.4a). This hydrogen atom had not been visible on the corresponding conformation of Ser195 in $\alpha LP_{pH\,8}$ ($Ser195_b$) because it overlapped in space with electron density for $Ser195_a$ Oγ in that structure. Not surprisingly, Ser195 Oγ-H is positioned in a plane ideal to hydrogen-bond with Sul202. When allowed to refine, this hydrogen refined to a position 0.98Å from its parent atom. This bond is longer than that predicted from canonical values as defined by SHELXL-97 (0.84Å) (Sheldrick & Schneider, 1997), but is identical to the bond length that is observed for the refined hydrogen atom on Ser214 Oγ, which hydrogen-bonds to Asp102 Oδ2.

The Oxyanion Hole

One of the prominent structural features of the serine protease mechanism is a conserved pocket of the protein designed to stabilize the oxyanion formed in the tetrahedral transition states. In αLP, this "oxyanion hole" is composed of the backbone amides of Gly193 and Ser195, which participate in hydrogen bonding interactions with the substrate's oxyanion. In both transition state analogs studied here, hydrogen atoms were visualized as distinct peaks on both amides in σ_A-weighted F_o-F_c maps calculated prior to the addition of

hydrogen atoms to the model. To our knowledge, this is the first time oxyanion hole hydrogen atoms have been visible by x-ray crystallography in a transition state analog.

In $\alpha LP_{pH\,5}$, Gly193 N (but not Ser195 N) is involved in hydrogen bonding interactions, primarily with Sul202 (Figure 3.2, Table 3.2). Both oxyanion hydrogen atoms were visible in σ_A-weighted F_o-F_c maps contoured at 4σ (Figure 3.5a). Gly193 N-H refined to a position 1.06Å from its parent atom. This is in contrast to the previously-solved structure of $\alpha LP_{pH\,8}$, where the position of 193 N-H refined to be 1.37(6)Å from its parent. Surprisingly, Ser195 N-H, which does not interact in hydrogen bonds in $\alpha LP_{pH\,5}$, is a very short bond (0.65Å). The refined position of this hydrogen agrees with the position of F_o-F_c density observed off Ser195 N.

In αLP+boroVal(gol), O1 of the tetrahedral boronate lies in the oxyanion hole. Interestingly, σ_A-weighted F_o-F_c difference electron density (contoured at 3σ) was visible off boroValP1 O1 (Figure 3.5). Addition of a restrained hydroxyl-hydrogen atom to O1 resulted in the disappearance of this density. Upon release of restraints, this hydrogen refined to a position 0.98Å from O1. The O1-H bond is aligned to facilitate hydrogen bonding to the terminal hydroxyl group of glycerol P1', with a donor-acceptor distance of 2.73Å. Additionally, boroValP1 O1 participates in hydrogen bonding with Gly193 N-H and Ser195 N-H of the oxyanion hole, with N...O1 distances of 2.73 and 2.85Å, respectively (Table 3.2). In early σ_A-weighted F_o-F_c maps calculated prior to the addition of hydrogen atoms to the model, a strong peak was visible off Ser195 N (at contour level 4σ; see Figure 4) and a weaker peak off Gly193 N (at contour level 2σ). The positions of Gly193 N-H and Ser195 N-H refined to be 0.85Å and 0.75Å from their donor atoms, respectively.

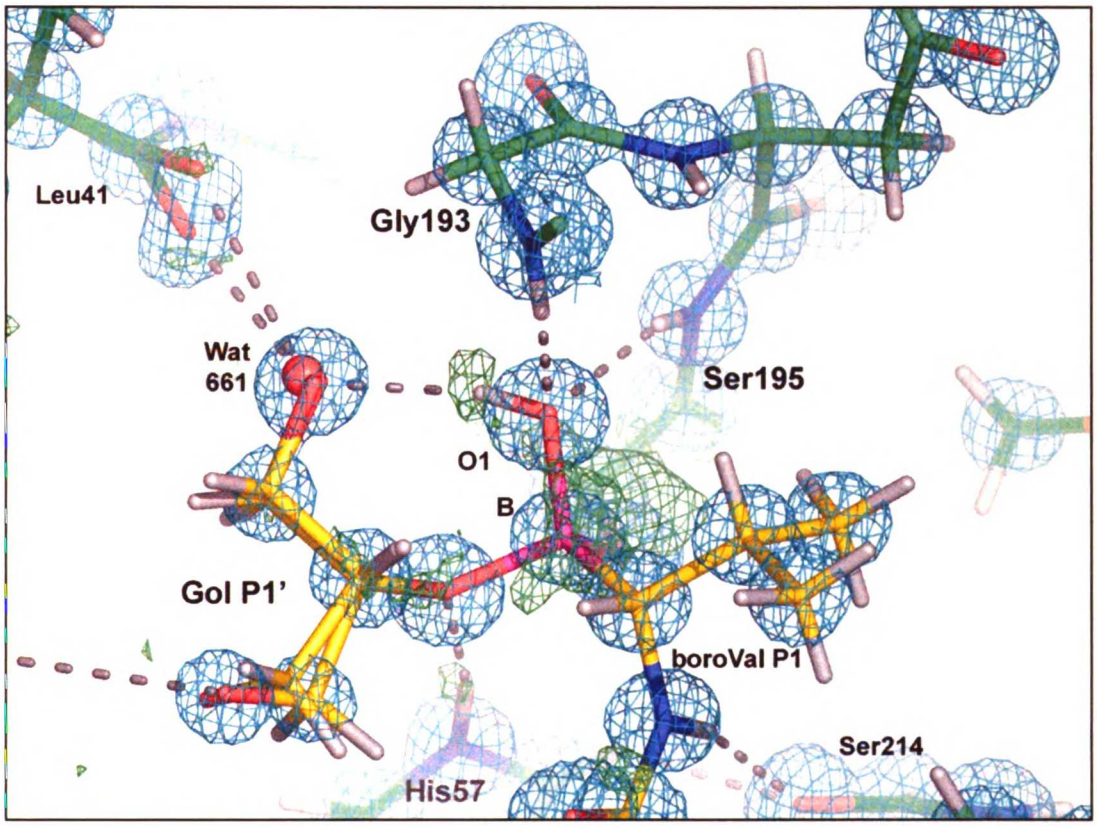


Figure 3.5. The boronic acid is protonated in the oxyanion hole.
$2F_o$-F_c electron density is contoured at 2σ (cyan), and σ_A-weighted F_o-F_c density contoured at 3σ. Even at this high contour level, a clear difference peak is visible off boroValP1 O1. O1 forms hydrogen bonding interactions with Gly193 and Ser195 in the oxyanion hole, and with the terminal oxygen of Glycerol P1'. Positive F_o-F_c density on the boron is consistent with the negative charge of the complex located on the boron, as is expected for a protonated boronic ester. σ_A-weighted F_o-F_c maps were calculated prior to addition of the boroValP1 O1 or His57 Nε2 hydrogen atoms to the model. The stick model shows the final structure of αLP+boroVal(gol), in which the position of O1-H has been refined.

As discussed below, protonation of boroValP1 O1 confirms that the negative charge of the tetrahedral adduct in boronic acid transition state analogs lies on the boron, rather than directly within the oxyanion hole. Indeed, in the final stages of refinement, positive σ_A-weighted F_o-F_c difference electron density on boroValP1 B indicated that the five electrons defined for a boron at this position was not sufficient (Figure 3.5). To address this, the occupancy of boron was allowed to refine. The occupancy refined to 88% of a carbon atom, indicating that the true density of B in boroValP1 is 5.3 electrons. Whether this value represents a partial negative charge or is an underestimate of the six electrons expected for a full formal charge is unknown.

Ser214 Positions the Catalytic Triad

Ser214, a conserved residue among serine proteases, has in the past been referred to as the fourth member of the catalytic triad due to the conserved hydrogen bonds that it forms to the sidechains of Asp102, His57, and the backbone amide of the substrate's P1 residue (Krem & Di Cera, 2001). The Ser214 Oγ ...Asp102 hydrogen bond maintains essentially identical geometry in $\alpha LP_{pH\,8}$, $\alpha LP_{pH\,5}$, and αLP+boroVal(gol), with an average donor-acceptor distance of 2.70±0.01Å in the three structures (Table 3.2). In all three structures, a hydrogen atom was visible in positive σ_A-weighted F_o-F_c electron density maps on Oγ, positioned in line with this hydrogen bond (see Figure 3.4) (Fuhrmann et al., 2004a). This hydrogen atom refined to identical positions in all three structures, with an Oγ-H distance of 0.98±0.05 Å in $\alpha LP_{pH\,5}$ and 0.99±0.06 Å in αLP+boroVal(gol).

The backbone carbonyl of Ser214 acts as acceptor for an unusual C-H...O hydrogen bond with His57 Cε1-H. This conserved hydrogen bond was previously recognized and proposed to play a key role in the mechanism of serine proteases (Derewenda et al., 1994). In αLP+boroVal(gol), the donor-acceptor distance for this hydrogen bond (3.09Å) is slightly longer than that observed in the apo enzyme at pH 5 (3.02Å) or pH 8 (3.03Å) (Fuhrmann et al., 2004a). Positive σ_A-weighted F_o-F_c electron density is clearly visible in a distinct peak off Cε1 in all three structures (Figure 3.4), with a refined Cε1-H bond length similar to the expected value of 0.95Å (see Table 3.2) (Sheldrick & Schneider, 1997).

His57

According to the general mechanism for serine protease catalysis, His57 is expected to be protonated at both Nδ1 and Nε2 in all of the tetrahedral intermediate and transition states. Solved at a pH below the pK_a of His57, the structure $\alpha LP_{pH\,5}$ provides a mimic of the transition state because His57 is doubly protonated. Indeed, in $aLP_{pH\,5}$ all hydrogen atoms on His57 were visible in σ_A-weighted F_o-F_c electron density maps, including Nε2-H, confirming the protonation state of His57 (Figure 3.4a). Availability of the Nε2-H hydrogen atom allows hydrogen bonding to both conformations of Sul202 and to Ser195 Oγ (Table 3.2).

Two of the most famous controversies in enzymology have focused on the precise character of the His57 Nδ1...Asp102 Oδ2 hydrogen bond in the transition state (Blow et al., 1969; Frey et al., 1994). In $\alpha LP_{pH\,5}$, the Nδ1...Oδ2 hydrogen bond is 2.76Å, which is only slightly shorter than the distance observed in $\alpha LP_{pH\,8}$ (2.77Å) (Fuhrmann et al., 2004a). This distance is significantly longer than the corresponding hydrogen bond observed in the low-

pH structure of subtilisin (2.62Å) (Kuhn et al., 1998). In αLP+boroVal(gol), which is a better mimic of the transition state, this hydrogen bond distance is 2.74Å (Table 3.2). Although shorter than the Nδ1-Oδ2 distance observed in the apo-enzyme, this hydrogen bond is still too long to be a short strong or low-barrier hydrogen bond (see Discussion). In $\alpha LP_{pH\,5}$, a clean peak of difference electron density contoured at 3σ indicated the position of a hydrogen atom on Nδ1 (Figure 3.4a). This hydrogen refined to a distance 0.88Å away from its parent atom (Nδ1). In αLP+boroVal(gol), a strong difference electron density peak (3σ) was observed in early stages of refinement at a position expected for a hydrogen atom on Nδ1. Surprisingly, the height of this peak decreased in the later stages of refinement (Figure 3.4b), perhaps suggesting that this density is noise. However, the prior shape of the peak, and the lack of thermal or conformational disorder in surrounding residues (average B-factor 3.9±0.4 for His57 and Asp102) indicate that this is not the case here. The position of this peak is 1.03±0.05Å from Nδ1 (1.76Å from Asp102 Oδ2). Therefore, both transition state analog structures ($\alpha LP_{pH\,5}$ and αLP+boroVal(gol)) indicate that this hydrogen is clearly located on His57, and not equally between His57 and Asp102, as would be predicted by the LBHB hypothesis.

During the acylation reaction, His57 donates its Nε2-H hydrogen atom to the nitrogen of the substrate's leaving group, allowing cleavage of the substrate and production of the acyl enzyme. In the tetrahedral intermediate mimetic structure αLP+boroVal(gol), His57 forms a short hydrogen bond with boroValP1 O2, the boronate mimic of the leaving group nitrogen. His57 Nε2 lies at a favorable angle from O2 for hydrogen bonding (164°), at a distance of 2.645±0.004Å (Table 3.2). A peak is visible in σ_A-weighted F_o-F_c difference electron density maps contoured at 3σ (Figure 5b). This hydrogen refines to a position 0.90±0.05Å from its

parent atom, Nε2, and 1.75Å from the acceptor atom, O2. Short hydrogen bonds are rare in the structure (see Discussion), and the overall distribution of hydrogen bond distances in αLP+boroVal(gol) agrees with those published for other ultra-high resolution structures (need to check this). Indeed, this is the shortest N...O hydrogen bond in the major conformation of αLP+boroVal(gol) (the second shortest hydrogen bond has a *d*(D...A) of 2.73Å). In contrast to the Nε2...O2 interaction, His57 forms only a weak hydrogen bond with Ser195, with a donor-acceptor distance of 3.01Å, and a D-H...A angle of 128°.

Discussion

We solved the structures of two transition state analogs, alpha-lytic protease at pH 5.1 ($\alpha LP_{pH\,5}$) and αLP bound to a transition state mimetic inhibitor, MeOSuc-Ala-Ala-Pro-Val boronic acid (αLP+boroVal(gol)) at sub-angstrom resolution (0.82Å and 0.90Å, respectively). Both structures were determined to very high accuracy, with the lowest R-factors reported to date for a protein larger than 100 residues (R/R_{free} of 8.10 / 9.34% for $\alpha LP_{pH\,5}$ and 8.14 / 9.20% for αLP+boroVal(gol); Table 3.1). Although the refined structures are nearly identical to the corresponding structures solved at lower resolution according to RMSD (see Results), information regarding the precise positioning of atoms have emerged at ultra-high resolution that allow us to address long-standing questions regarding the mechanism of proteolytic catalysis.

Structural effects of the protonation of His57 in the apo enzyme

In the previously-solved 0.83Å resolution crystal structure of αLP at pH 8 ($\alpha LP_{pH\,8}$) it had been observed that Ser195 was present in two conformations at pH 8 (Fuhrmann et al., 2004a). One of these conformations ($Ser195_a$) formed a hydrogen bond with His57 Nε2; in the other conformation, $Ser195_b$ Oγ was shifted by ~1Å, increasing its distance to His57 Nε2 and bringing it to an angle that was less favorable for hydrogen bonding (see Figure 3.2b). It was hypothesized that this latter conformation of Ser195 corresponded with a population of αLP molecules within the crystal for which His57 was protonated at Nε2, implying that the pK_a of His57 had shifted from 7 to 8.8 in αLP crystals (likely due to the close proximity of a sulfate ion in the active site). Protonation at Nε2 would cause steric clash with the canonical conformation of Ser195, causing it to shift to the second conformation.

The structure of αLP at pH 5 confirms this proposition. According to this hypothesis, at pH 5 100% of αLP molecules would be protonated at His57 Nε2, and therefore only the second conformation of Ser195 would be present in the structure. As illustrated in Figure 3.2b, this is indeed the case. In $\alpha LP_{pH\,5}$, Ser195 occupies one conformation, which is essentially identical to that of $Ser195_b$ in $\alpha LP_{pH\,8}$. $\alpha LP_{pH\,5}$ thereby confirms the hypothesis presented in Fuhrmann, *et al.* (2003), and justifies the pK_a calculation performed in that study.

Modeling the Acylation Transition State Intermediate

The αLP+boroVal(gol) structure is the best mimic of the tetrahedral intermediate (and therefore the TS and TS' states) for the acylation reaction determined at sub-angstrom resolution. As is known, the MeOSuc-Ala-Ala-Pro-boroVal-OH inhibitor forms a stable tetrahedral adduct with Ser195, mimicking the transition state intermediate of the deacylation reaction (Bone et al., 1987). Surprisingly, we found a glycerol molecule covalently bound to the boronic acid at ~65% occupancy. The glycerol (GolP1') occupies the S1' pocket, mimicking the leaving group of the substrate and therefore providing us with a model of the tetrahedral intermediate for the acylation reaction (TI_1, and by extraolation TS and TS'). A component of the cryosolvent, glycerol molecules entered the crystal lattice during the 45-second soak of the crystal prior to mounting under the cryostream. Attack of the boronate by a glycerol molecule has been observed previously in αLP crystals in cryosolvent soaks as short as 10 seconds (C. Fuhrmann and D. Agard, unpublished results), and reflects the reactivity of B-O to hydroxyl groups.

This 0.90Å resolution structure allows precise analysis of the geometry of the Ser195-inhibitor adduct. The geometry about the boron deviates only slightly from tetrahedral, with angles ranging from 105.7° to 114.7° (Table 3.3). Similar angles have been observed in structures of compounds containing a central tetrahedral boron (Allen, 2002; Bruno et al., 2002). Therefore, we conclude that the ±5° deviation from perfect tetrahedral geometry in αLP+boroVal(gol) is typical for a boronic acid in a heterogeneous environment such as the enzyme active site.

Peptide boronic acids have been used for decades to mimic αLP's transition state in structural studies (Fuhrmann et al., 2004b). Although scientists have suspected that the boronate oxygen in the oxyanion hole is protonated, this ultra-high resolution crystal structure provided the first opportunity to observe this hydrogen atom, and confirms that the negative charge in boronic acid transition state analogs lies on the central boron. Although an excellent mimic of the tetrahedral intermediate, a boronic acid-bound protease differs from a true TI or TS state primarily in its smaller number of protons at the tetrahedron's central atom, the presence of negative charge on this central atom (rather than in the oxyanion hole), and the presence of a hydrogen atom on the oxyanion mimic.

The TI_1 → EA transition: No evidence for a His57 ring flipping mechanism

It has been proposed that a "reaction-driven" ring flip of His57 is required for His57 to act both as a general base (in the formation of the first tetrahedral intermediate) and as an acid (donating a proton to the substrate leaving group to promote formation of the acyl enzyme) (Ash et al., 2000). Because ultra-high resolution crystallographic data provides information regarding both the precise positioning of atoms and the direction of motion about these atoms, the transition state analog structures presented here allow us to address this proposal.

In the first step of the mechanism (as modeled by $\alpha LP_{pH\,8}$), neutral His57 is positioned to abstract the proton from Ser195 in order to facilitate nucleophilic attack on the scissile peptide carbon (Fuhrmann et al., 2004a). In the second step of the proteolytic

reaction—breakdown of the tetrahedral intermediate into the acyl enzyme complex—it was postulated that His57 ring flipping is necessary to promote protonation of the leaving group over back-protonation of Ser195 (Ash et al., 2000). However, we see that this is not the case. In our tetrahedral acylation transition state analog structure, αLP+boroVal(gol), His57 is optimally positioned to hydrogen bond (through Nε2) to the mimic of the leaving group nitrogen, boroValP1 O2 (see later discussion). Additionally, Nε2 hydrogen bonding with Ser195 Oγ is disfavored (Table 3.2). Therefore, in our structure, flipping of His57 is unnecessary, as His57 is pre-positioned to favor formation of the acyl enzyme complex, and to prevent back-protonation of Ser195.

Our experimental observation of His57 Nε2 hydrogen bonding to the leaving group in αLP+boroVal(gol) (a mimic of TI_1) was predicted by QM/MM simulations performed by Ishida, *et al.* that calculated the energy profile for the acylation reaction (Ishida & Kato, 2003). These theoretical studies predicted that upon becoming tetrahedral, the scissile bond would rotate to a position where the –NH leaving group hydrogen-bonded to His57 Nε2. The calculations showed that rotation about the scissile bond would occur prior to formation of TI_1, leaving the substrate in a position favorable for formation of the EA complex.

Our data suggests that movement of Ser195 also occurs in the ES→ TI_1 transition. In both $\alpha LP_{pH\,5}$ and αLP+boroVal(gol), the sidechain of Ser195 has moved such that Oγ is located ~1Å away from its position in $\alpha LP_{pH\,8}$ where His57 is neutral. A comparison of $\alpha LP_{pH\,8}$ and $\alpha LP_{pH\,5}$ (Figure 3.2b) show that this repositioning of Ser195 occurs upon protonation of His57 at Nε2. Based on this data, we hypothesize that abstraction of the Oγ proton and movement of the Ser195 sidechain occur in concert during the EA→ TI_1 transition.

If His57 ring flipping were a mechanism required for transfer of the proton from Nε2 to the leaving group, this phenomenon would have to occur either in the TI_1 state, or in the transition state for the TI_1 -> EA reaction. To date, only indirect experimental evidence have been used to promote the His57 ring-flipping model (Ash et al., 2000; Haddad et al., 2005). Most recently, NMR studies of αLP published by Haddad *et al.* suggested that His57 is more mobile in apo-αLP at low pH than at high pH (Haddad et al., 2005). The authors stated that such movement could encompass motions as small as a wobble or as large as a ring-flip, making their observation consistent with the ring-flipping hypothesis. Our ultra-high resolution crystal structures of αLP at low pH ($\alpha LP_{pH\,5}$) and as an analog of TI_1 (αLP+boroVal(gol)) allow us to characterize the motion of His57 in the doubly-protonated state, and to test for ring-flipping or wobbling. At sub-angstrom resolution, atoms around the His ring can be readily identified in σ_A-weighted $2F_o$-F_c maps at high contour levels (data not shown), where nitrogen atoms appear distinctly larger than carbon atoms. More precisely, refined occupancies and B-factors directly report on electron content. If two sets of superimposed conformations of His57 were present in the crystals (one flipped, one not), this would be reflected in deviations of atomic occupancies and B-factors, which are not observed in either $\alpha LP_{pH\,5}$ or αLP+boroVal(gol).

Haddad *et al.* argue that His57 may not be observed in the flipped position in crystal structures, since only a small fraction of molecules are required to be "flipped" in order to make the mechanism effective (Haddad et al., 2005). Even in this case, if flipping is to occur one would expect to observe atomic motions indicative of wobble about the Cβ-Cγ bond in the TS and TI_1 states. If wobbling occurred, each atom of the ring should display anisotropy perpendicular to the ring, and Cβ and Cγ should have less anisotropic character of motion.

However, refined atomic anisotropic B-factors for His57 in $\alpha LP_{pH\,5}$ and in αLP+boroVal(gol) suggest a rigid body motion of the entire residue diagonal to the plane of the ring (Figure 3.6). This rigid body motion of His57, rather than histidine flipping, may have been the motion previously observed by NMR (Haddad et al., 2005).

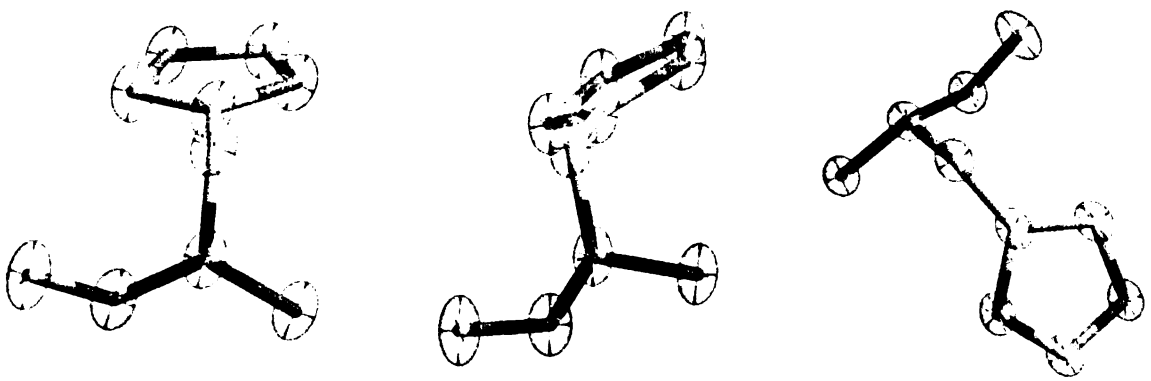


Figure 3.6. Anisotropic motion of His57 in αLP+boroVal(gol).
His57 is shown in three orientations, with atomic anisotropy represented by ellipsoids drawn at the 50% probability level. Anisotropic motion for all atoms in His57 are of comparable intensity and direction, with no indication of imidazole ring-flipping. This figure was made using Raster3D/rastep (Merritt & Bacon, 1997) and Molscript (Kraulis, 1991).

In summary, based on the αLP+boroVal(gol) complex structure, which is the best structure of the acylation tetrahedral intermediate mimic to date, there is no indication that His57 flipping is necessary or occurring in αLP. As is evident in both apo-αLP and αLP+boroVal(gol), when His57 Nε2 is protonated Ser195 occupies a different conformation, forming at best a weak hydrogen bond to His57. Furthermore, as evidenced by αLP+boroVal(gol), the movement of Ser195 occurs prior to formation of TI_1. His57 is then

already in an optimal position in TI_1 to donate the proton to the substrate leaving group, making a histidine-flipping mechanism unnecessary and even detrimental to progression of the reaction. These results are consistent with other ultra-high resolution crystal structures of serine proteases at low pH (Katona et al., 2002), predictions by QM/MM modeling of the acylation reaction (Ishida & Kato, 2003), and NMR data previously used to promote the ring-flip hypothesis (Ash et al., 2000; Haddad et al., 2005).

His57 Nδ1 – Asp102 Oδ2 : Debating the low-barrier hydrogen bond

It has been proposed that His57 and Asp102 participate in a low-barrier hydrogen bond (LBHB) when His57 is doubly-protonated, increasing the basicity of His57 and thereby facilitating the first step of the acylation reaction (Cleland & Kreevoy, 1994; Cleland et al., 1998; Frey et al., 1994; Frey, 2004). By definition, a LBHB is one in which the classic energy barrier to proton transfer is lower in energy than the first vibrational energy level of that atom (Figure 3.1b). The formation of a LBHB requires two structural features. First, the distance between the donor and acceptor atoms must be less than the van der Waals radii of these atoms (<2.65Å for N ...O (Frey et al., 1994)). Second, quantum mechanical calculations show that in a LBHB the hydrogen atom will be located equidistant between donor and acceptor (GarciaViloca et al., 1997a). These two characteristics of a LBHB should cause a significant downfield chemical shift in the proton NMR spectra for Nδ1, as has been observed for several serine proteases at low pH or bound to inhibitors that mimic the

tetrahedral intermediate (Cleland et al., 1998; Frey et al., 1994), including αLP bound to MeOSuc-Ala-Ala-Pro-boroVal-OH (16.0 ppm) (Bachovchin et al., 1988).

According to the LBHB hypothesis for serine protease catalysis, a LBHB forms between His57 Nδ1 and Asp102 Oδ2 in states TS, TI, and TS' during the acylation reaction (Figure 3.1). This LBHB would stabilize the His-Asp pair, facilitating deprotonation of Ser195 Oγ by His57 and thereby catalyzing the first step of the acylation reaction. Rather, in αLP+boroVal(gol), which mimics the acylation tetrahedral intermediate, we observe that Nδ1 and Oδ2 are separated by 2.734±0.005Å, a distance favorable for a normal H-bond but not compatible with a LBHB. Additionally, the Nδ1-H hydrogen atom refined to a position 1.03±0.05Å from Nδ1 and is not positioned midway between His57 and Asp102 as required by the LBHB hypothesis (Table 3.2 and Figure 3.4). Although this signal is weak compared to that observed in the apo-αLP structure, the peak height is similar to those observed for several amide hydrogens in the structure. Furthermore, the peak we observe is not indicative of a second conformation of His57. If this were the case, anisotropic B-factors should suggest movement in the direction of the putative hydrogen peak. Instead, the shortest axis of anisotropic motion lies in this direction (Figure 3.6).

In addition to inhibitor-bound transition state analogues, samples of the apo enzyme buffered at a pH below His57's pK_a are also frequently studied as mimics of the transition state due to the positively charged nature of His57. In our structure of αLP at pH 5 presented here, the His57...Asp102 hydrogen bond has geometry similar to that of αLP+boroVal(gol). The donor-acceptor distance (2.755 ± 0.005Å) and donor-hydrogen bond length (0.88 ± 0.05Å) definitively show that the His57...Asp102 hydrogen bond is not a LBHB.

Our observation that Nδ1-H is located on His57 and is not involved in a LBHB agrees with Kossiakoff and Spencer's 1980 neutron diffraction study of deuterated bovine trypsin bound to a transition state analog (monoisopropylphosphoryl group), in which they showed that the hydrogen atom is located on His57 (and not on Asp102) (Kossiakoff & Spencer, 1980). The geometry observed for Nδ1-H in $\alpha LP_{pH\,5}$ and αLP+boroVal(gol) is also consistent with structures of serine proteases previously solved at low pH to mimic the transition state (Betzel et al., 2001; Katona et al., 2002). These data collectively argue against the hypothesis that the His57-Asp102 hydrogen bond is a LBHB. In contrast to these data are two model systems that have frequently been used as the primary examples of observed LBHBs in serine proteases: chymotrypsin bound to peptide trifluoromethylketone and subtilisin at low pH.

The LBHB hypothesis has been most thoroughly studied in complexes of chymotrypsin with peptidyl trifluoromethyl ketone inhibitors (Cht-TFK). Although atomic-resolution crystallographic data is not available for a Cht-TFK complex, lower-resolution structures of Cht bound to Acetyl-Phe-CF_3 (1.4Å resolution; 1GG6.pdb) and to Acetyl-Leu-Phe-CF_3 (1.5Å resolution) have each assigned a distance of 2.6Å to the His57...Asp102 pair (Neidhart et al., 2001). Additional structural characterization of these complexes, including downfield chemical shifts (Lin et al., 1998a), low H/D fractionation factors (Lin et al., 1998b), activation enthalpies for hydrogen exchange (Lin et al., 1998b), and negative tritium isotope effects (Westler et al., 2002), are also consistent with the His57...Asp102 hydrogen-bond being a LBHB. However,, like all transition state analogs, Cht-TFK complexes are imperfect mimics of the transition states or tetrahedral intermediates. In particular, the $-CF_3$ group located in the position of the scissile amide sterically precludes formation of a

hydrogen bond to His57 Nε2, a key interaction in the acylation reaction. One result of this might be the repositioning of His57 away from $-CF_3$ and towards Asp102, falsely creating a structural environment indicative of a LBHB.

A second system that demonstrates strong evidence for a LBHB in the catalytic triad is *Bacillus lentus* subtilisin. In the 0.78Å resolution structure of this serine protease at pH 5.9, the His Nδ1...Asp Oδ2 distance was 2.62Å; furthermore, a peak for the hydrogen atom was observed nearly equidistant between Nδ1 and Oδ2, being highly suggestive of a LBHB (Kuhn et al., 1998). While there may be several explanations for this observation, perhaps the most likely is that a LBHB may occur in subtilisin family members, at least under the experimental conditions of the Kuhn *et al.* structure (at low pH and in the presence of peptide proteolysis products). However, our data and that of others suggest that such a His-Asp LBHB is absent in the ground state and acylation tetrahedral intermediate in chymotrypsin family members, and is therefore likely not a general requirement for catalysis by serine proteases.

Kirsch and colleagues tested the functional importance of the putative His-Asp LBHB in subtilisin BPN' by mutating the catalytic Asp to a Cys (Stratton et al., 2001). As evidenced by the loss of the downfield NMR signal assigned to His Nδ1-H, this mutant lacked the putative LBHB. Based on the observed 50-fold decrease in k_{cat}, the authors concluded that if it had existed, the LBHB contributed much less to transition state stabilization (2.2 kcal/mol) than originally predicted (5-10 kcal/mol) (Stratton et al., 2001). Indeed, arguments against the LHBH hypothesis state that the covalent nature of a LBHB would in fact be less effective in stabilizing the transition state compared to the electrostatic effects of a normal hydrogen

bond (GarciaViloca et al., 1997b; Kim et al., 1999; Shan & Herschlag, 1996; Shan et al., 1996; Warshel & Papazyan, 1996).

If the His57-Asp102 interaction is indeed not a LBHB in boronic acid-bound αLP, then why does this complex produce a chemical shift characteristic of that found for hydrogen atoms involved in LBHBs (16.0 ppm (Bachovchin et al., 1988))? Multiple theoretical studies have predicted that while a LBHB should produce an NMR signal in the characteristic range of 16-20 ppm, the detection of such a signal could also be due to an ionic hydrogen bond (GarciaViloca et al., 1997a; Guthrie, 1996; Shokhen & Albeck, 2004; Warshel & Papazyan, 1996). Our observations confirm this prediction. In conclusion, our observations in the αLP+boroVal(gol) and $\alpha LP_{pH\,5}$ structures, and the Asp→ Cys mutation experiments in subtilisin (Stratton et al., 2001), show that a LBHB is not necessary for efficient catalysis, and that the strong downfield shift observed by NMR does not necessarily imply the presence of a LBHB.

His57 Nε2 forms a short ionic hydrogen bond *to the leaving group mimic*

Unlike commonly-studied peptidyl trifluoromethyl ketone-protease complexes (which contain bulky $-CF_3$ at the position of the leaving group), the protease-boronic acid adduct is capable of hydrogen bonding directly to His57, providing a better mimic of the His57-leaving group interaction in the transition state. As described above, a glycerol molecule bound to the boronic acid moiety in the S1' pocket of our αLP+boroVal(gol) complex, allowing analysis of this structure as a mimic of the first tetrahedral intermediate (TI_1). In this structure, His57 is positioned to form a short hydrogen bond to boroValP1 O2 (distance =

2.65Å) with ideal geometry (Table 3.2). The donor-acceptor distance observed for this hydrogen bond is suggestive of a low-barrier hydrogen bond. This agrees with NMR data for αLP in the presence of MeOSuc-Ala-Ala-Pro-boroVal-OH showing a downfield chemical shift of 16.5 ppm for Nε2-H (Bachovchin et al., 1988). However, unlike a LBHB (where the hydrogen atom is by definition equidistant between donor and acceptor), Hε2 in αLP+boroVal(gol) is clearly located on His57 Nε2 (Figure 3.4b). This observation demonstrates that the relationship between donor-acceptor distance and delocalization of the hydrogen atom in a hydrogen bond is not strictly correlated. Clearly it is possible to have a short heavy atom distance and an asymmetric hydrogen bond. This point was previously predicted by Schutz and Warshel (Schutz & Warshel, 2004). Using valence bond theory to calculate energy profiles for the theoretical LBHB and ionic hydrogen bond, Schutz and Warshel showed that a short hydrogen bond can be either ionic or covalent in character (the latter being a LBHB). The shape of the energy profile for proton transfer defines whether a short hydrogen bond is ionic or a LBHB, with a symmetric energy profile (and $\Delta pK_a = 0$) being required for a so-called LBHB (Figure 3.1b). As the distance between donor and acceptor decreases, the pK_a's of each group approach each other. However, the critical distance at which ΔpK_a equals zero is influenced by the environment surrounding the hydrogen bond; this allows for ionic hydrogen bonds to occur at short donor-acceptor distances. The observation that Nε2-H is located on His57 in αLP+boroVal(gol), despite the short donor-acceptor distance of 2.65Å, demonstrates experimentally that this is indeed a short ionic hydrogen bond (SIHB).

Interestingly, the short distance between donor (Nε2) and acceptor (O2) atoms in the His57-leaving group mimic of αLP+boroVal(gol) (coupled with a D-H...A angle

approaching linearity) suggests that this hydrogen has a reduced energetic barrier to transfer relative to a hydrogen bond of standard length, promoting protonation and release of the leaving group. Indeed, the unusual character of this hydrogen bond (based on NMR data for serine proteases bound to various boronic acid inhibitors) has been alluded to in papers scattered throughout the history of serine proteases (Ash et al., 1997; Bachovchin et al., 1988; Robillard & Shulman, 1974; Steitz & Shulman, 1982). Further studies should be devoted to the His57 Nε2—leaving group hydrogen bond to discern its true energetic contribution to proteolytic catalysis.

αLP+boroVal(gol) allows prediction of the energetic landscape for acylation

We have described two key observations based on the ultra-high resolution crystal structures of αLP. First, by the TI_1 state Ser195 has already undergone a conformational change that prevents the back-reaction to reformation of reactants. Secondly, in αLP+boroVal(gol) His57 Nε2 forms a short hydrogen bond to the leaving group mimic (boroValP1 O2). These observations suggest that the acylation tetrahedral intermediate structure we have isolated is biased toward the acyl enzyme complex, and significantly away from the Michaelis complex, along the reaction coordinate. Interestingly, two quantum mechanical studies of the acylation reaction have independently predicted that the lowest-energy tetrahedral intermediate conformation is that which has already undergone a conformational change to disrupt the His57 Nε2—Ser195 Oγ hydrogen bond and promote hydrogen bonding to the leaving group (Daggett et al., 1991; Ishida & Kato, 2003). All of these data suggest that TI_1 more closely resembles TS_1' than TS_1 structurally. By the

Hammond postulate (Hammond, 1955), we extrapolate that TI_1 is also closer in energy to TS_1' than to TS_1. The short hydrogen bond between His57 Nε2 and the leaving group may further reduce the $TI_1 \rightarrow TS_1'$ energy barrier.

A network of short hydrogen bonds stabilize the catalytic triad

Interestingly, three of the four shortest N...O hydrogen bonds in the αLP+boroVal(gol) structure form interactions directly in the active site (His57 Nε2-H and Gly193 N-H hydrogen bonding to the boronic acid) or within the catalytic triad (His57 Nδ1-H... Asp102 Oδ2). Figure 3.7 shows the distribution of hydrogen bond distances within αLP+boroVal(gol), and the relative positions of active site, catalytic triad, and enzyme-inhibitor hydrogen bonds within this distribution. All hydrogen bonds formed between enzyme residues within or to the catalytic triad (blue squares), or within the active site to the inhibitor (red squares), are among the shortest 25% of hydrogen bonds found within the structure. The only exception is the His57 Nε2...Ser195 Oγ hydrogen bond, which we have argued is weaker in TI_1 to promote forward progression of the reaction (see discussion above). Adding to this hydrogen bond network is the conserved C-H...O hydrogen bond connecting the backbone of Ser214 to His57. In contrast to hydrogen bonds found at our near the active site, the three hydrogen bonds that form between the enzyme and the inhibitor backbone are weaker hydrogen bonds (d(D..A) = 2.93 – 3.11Å). We propose that the enzyme has evolved to stabilize the tetrahedral intermediate and transition states by creating a preorganized active site that is optimized for transition state stabilization. Residues are optimally positioned by a collection of relatively strong ionic hydrogen bonds in the vicinity

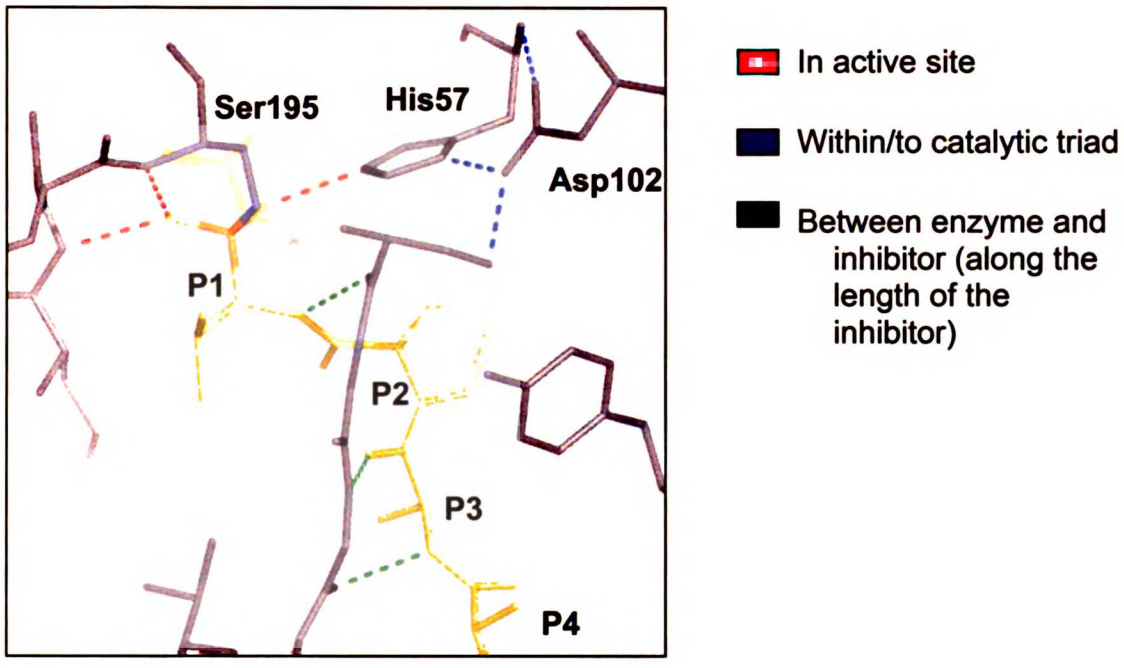
of the active site. Furthermore, the fact that hydrogen bonds involved in positioning the substrate in the enzyme binding pocket are of average donor-acceptor distances suggests that the enzyme's evolution favored maximizing k_{cat} rather than decreasing K_m. This phenomenon has been inferred from kinetic data demonstrating the "k_{cat} effect": that shortening substrates from three to one amino acids long decreases the k_{cat} 2500-fold, but K_m is only increased 3-fold (R. Bone and D. Agard, unpublished results; similar effects observed for other proteases (Case & Stein, 2003; Thompson & Blout, 1973)). Optimizing catalysis over substrate binding would benefit the enzyme in the final steps of proteolysis by facilitating product release. This may be a general strategy used by serine proteases (and enzymes in general) to optimize both catalysis and enzyme turnover.

Figure 3.7. Distribution of hydrogen bond lengths in αLP+boroVal(gol).
The histogram (b) shows the distribution of donor-acceptor distances for all hydrogen bonds in αLP+boroVal(gol) between nitrogen and oxygen atoms (N...O) and two oxygen atoms (O...O), with the average D...A distance denoted by a dashed line. Hydrogen bonds formed between the enzyme and inhibitor in the active site (i.e., to the boronic acid moiety directly) are colored red, and along the length of the inhibitor backbone are colored green. Enzyme intramolecular hydrogen bond interactions found within or to the catalytic triad are colored in blue. The position of these hydrogen bonds within the structure are illustrated in (a), where the inhibitor is represented by yellow sticks, and the enzyme in grey. Hydrogen bonds identified in each of these three categories are indicated by (*) in Table 3.2. Hydrogen bonds involving minor alternate conformations are not included.

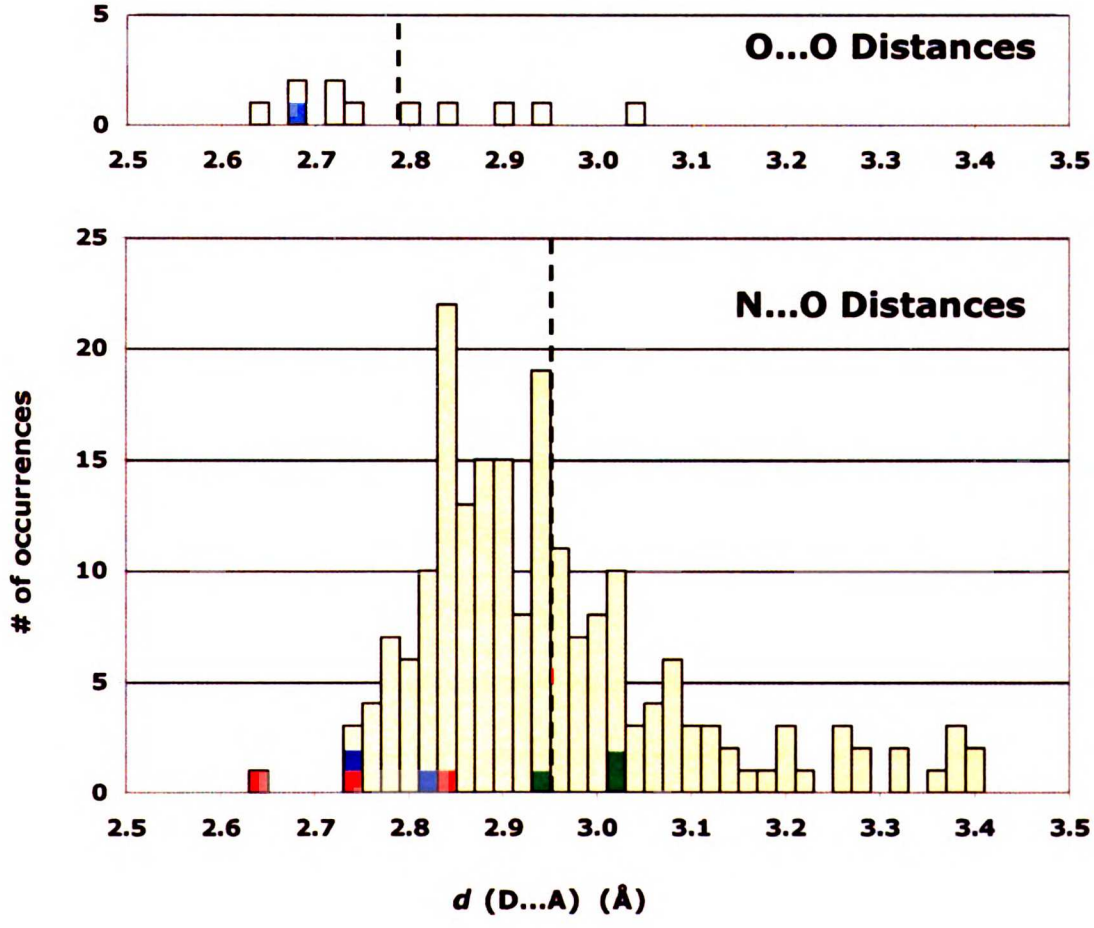
(Figure shown on next page.)

(Figure 3.7, caption on previous page)

a.



b.



Conclusion

We have solved the ultra-high resolution crystal structures of two transition state analogs of alpha-lytic protease, αLP at pH 5 and αLP bound to a peptidyl boronic acid (αLP+boroVal(gol)). The unexpected covalent binding of a glycerol molecule to the tetrahedral boronate in αLP+boroVal(gol) provided the best current model of the tetrahedral intermediate for acylation, allowing us to address questions of catalysis that have existed in the serine protease field for decades. Our crystallographic structures show that histidine flipping is not required for forward progression through the reaction mechanism and suggest, rather, that movement of Ser195 prior to the formation of TI_1 destabilizes hydrogen bonding to His57 and thereby prevents the TI_1 -> ES back-reaction. Based on precise atomic positions obtained at sub-angstrom resolution, we have concluded that a LBHB between Asp102 and His57 does not exist in the acylation transition state or intermediate. Although a His-Asp LBHB may have been observed in subtilisin (Kuhn et al., 1998), it is not required as a general mechanism for serine proteases. Instead, transition state stabilization by chymotrypsin-like serine proteases is likely achieved through a network of optimized hydrogen bonds that position the catalytic triad and stabilize the Ser195-substrate tetrahedral adduct. In particular, we identified a short ionic hydrogen bond (SIHB) between His57 and the amide of the leaving group that may play the primary role in catalyzing the second step of the acylation reaction. Combined, our observations contradict the common use of a downfield ^1H chemical shift and short donor-to-acceptor distance as litmus tests for a LBHB, and provide experimental evidence for a preorganized collection of relatively strong ionic hydrogen bonds in the vicinity of the active site that likely provide the driving force for proteolytic catalysis.

Materials and Methods

Crystal Preparation

αLP protein was expressed and purified from its native bacterium, *Lysobacter enzymogenes* (see Appendix B). As for previous experiments published by these authors, αLP was never lyophilized during the progression of protein purification. The crystal used in the determination of αLP+boroVal(gol) was grown by methods described previously, using precipitant solution composed of 1.3 M Li_2SO_4 and 20 mM Tris, buffered to pH 8 (Fuhrmann et al., 2004a). Low pH crystals were grown under similar conditions, but equilibrated with precipitant solution at pH 4.25. These crystals grew much more slowly than those equilibrated at pH 8, but retained the ability to diffract to sub-angstrom resolution.

The boronic acid inhibitor (MeOSuc-Ala-Ala-Pro-boroVal-OH) was prepared by dissolving MeOSuc-Ala-Ala-Pro-(D,L)-boroVal-OH in 20 mM Tris-sulfate (pH 7.5) to a final concentration of 0.14 M. One microliter of this solution was added to a drop containing αLP crystals. The drop equilibrated for 17.5 hours prior to data collection, allowing formation of the enzyme-inhibitor complex. Although a racemic mixture of the D and L forms of the inhibitor was used, only the L form was visible in electron density (likely due to the preferential binding of this form by at least 100-fold (Kettner et al., 1988)).

Data collection

αLP at pH 5.1

Data was collected at the Advance Light Source (HHMI beamline 8.2.1) on a single crystal of approximate dimensions 0.2 x 0.2 x 0.4 mm. Prior to mounting on the cryostream, the crystal was whisked through a cryoprotectant solution composed of precipitant solution (1.3 M Li_2SO_4 and 20 mM Tris) and 15% glycerol. At the time of data collection, the cryoprotectant was measured to be pH 5.1.

As described previously, ultra-high resolution data collection requires a short detector-to-crystal distance coupled with a short wavelength (to ensure that the outer high-resolution reflections will be captured on the detector) and long exposure times (in order to detect these weakly-diffracted reflections). Because low-resolution reflections tend to become oversaturated during these long exposure times, additional data sets must be collected to optimally capture these "low-resolution" reflections. Here, we collected high-resolution data at two positions on the crystal, and later merged this data to obtain a complete "high-resolution" data set (to 0.86Å resolution). We then collected two complete data sets using shorter exposure times, providing a "medium resolution" set (to 1.15Å resolution) and a "low resolution" set (to 1.24Å resolution). Because the crystal continued to display strong diffraction potential, we collected an additional, "ultra-high resolution," data set with the detector offset along 2-theta, providing data to 0.78Å resolution. All four data sets were integrated using Denzo (Otwinowski & Minor, 1997) and scaled and merged in Scalepack (Fox & Holmes, 1966), providing a single data set that was 100% complete and highly

redundant (see Table 3.1). The final resolution was determined to be 0.82Å to ensure a median signal-to-noise ratio greater than two and an R_{merge} less than 50% for the highest-resolution data.

αLP bound to mAAP-boroVal...glycerol

Data was collected at the Advanced Light Source (HHMI beamline 8.2.2), using x-rays of wavelength 0.75Å, on a single hexagonal crystal of dimensions ~0.3 x 0.3 x 0.5 mm. Prior to mounting on the goniometer, the crystal was equilibrated in a cryoprotectant containing 15% glycerol for 45 seconds. The crystal was then frozen directly in the stream of gaseous nitrogen, which was maintained at a temperature of 100 K throughout data collection.

In previous studies of αLP bound to peptidyl boronic acid inhibitors, we observed boron-specific radiation damage caused by the long x-ray exposures necessary to obtain ultra-high resolution data (see Chapter 2). Therefore, in this experiment, we took special care to minimize the dose of radiation exposed to any one position of the crystal during collection of high-resolution data. During data collection, the crystal was translated nine times along the length or width of the crystal, with only 15-17 of the total ~100 frames of high-resolution data collected at any one position. At each position, a low-resolution data set was collected before and after the collection of high resolution data. These low-resolution data sets allowed calculation of $F_{o\,(post\text{-}high)} - F_{o\,(pre\text{-}high)}$, which were used to confirm that no specific structural damage had occurred during the long-exposure high resolution data collection.

As a further test for radiation damage, an αLP model was refined using a 1.6Å resolution data set collected "post-high" at one position (see Chapter 2). In this model, σ_a-weighted F_o-F_c maps also showed no indication of x-ray radiation damage. In these tests, reduction of disulfide bonds Cys42/Cys58 and Cys189/Cys220A was used as an indicator for x-ray radiation damage. Even when such damage is present in as few as 10% of the molecules, this is clearly visible in F_o-F_c difference maps at high contour levels (3σ), providing an easy test for x-ray radiation damage at the molecular level. Collection and analysis of this data is described in more detail in Chapter 2.

Data sets were indexed using HKL2000 (Otwinowski & Minor, 1997). High-resolution data sets were scaled and merged with two representative low-resolution data sets using Scalepack (Fox & Holmes, 1966). The resulting data set was limited to a resolution of 0.90Å (see Table 3.1).

Model refinement

Both $\alpha LP_{pH\,5}$ and αLP+boroVal(gol) were refined in SHELXL-97 (Sheldrick & Schneider, 1997) using strategies similar to those used to solve the 0.83Å structure of $\alpha LP_{pH\,8}$ (Fuhrmann et al., 2004a). Throughout the refinement of each model, 5% of the data were set aside as a test set for calculation of R_{free} as validation of the model (Brunger, 1992).

αLP at pH 5.1

Refinement was begun using the previously-determined structure of αLP at pH 8 (1SSX.pdb, 0.83Å resolution) (Fuhrmann et al., 2004a). In order to minimize model bias, the starting model was stripped of disordered waters (i.e., if partially-occupied or with an isotropic temperature factor greater or equal to 50) and of minor alternate conformations of amino acids (with occupancies of major conformations returned to 100%). Anisotropic B-factors and hydrogen atoms were not included in the starting model, and sulfate and glycerol molecules were removed. Following rigid body refinement and minimization, the starting model had an R / R_{free} of 19.6% / 21.6%. During the early stages of refinement, major changes were made to the model while minimizing against data to 1.2Å resolution, including editing of waters, the addition of three sulfate ions and one glycerol molecule, and inclusion of nine amino acid alternate conformations (R / R_{free} = 16.7% / 18.2%). At this point, data was added to 0.9Å resolution, and all atoms modeled with anisotropic temperature factors, resulting in a reduction of R / R_{free} to 12.9% / 14.3%. Addition of 32 whole and 85 partially-occupied waters, one glycerol molecule, and six amino acid alternate conformations reduced R / R_{free} to 11.2% / 12.6%. Hydrogen atoms at this stage were included as "riding hydrogens," with fixed geometry as defined by SHELXL-97 (Sheldrick & Schneider, 1997), and the model minimized against data to 0.82Å resolution (R / R_{free} = 10.9% / 12.1%). In order to ensure no model bias for analysis of difference electron density maps, hydrogen atoms were not added to key atoms in the catalytic triad, Ser214, and oxyanion hole until the final stages of model building. Because positive difference electron density maps showed unique positions for hydrogen atoms on methyl, primary amine, and hydroxyl groups in

residues with a single conformation, these hydrogens were added to the model using the AFIX 137 and AFIX 147 instructions, which refine torsion angles to position hydrogen atoms in maximum electron density. After most of the model was complete (R / R_{free} = 8.35% / 9.64%), hydrogens were added to His57, Ser195 Oγ and N, Ser214 Oγ, and Gly193 N as riding hydrogens. After five cycles of minimization, restraints were removed from these hydrogen atoms and their positions allowed to refine throughout the rest of the refinement process. At this stage, restraints were also removed from all non-hydrogen atoms in residues of a single conformation. This resulted in a decrease in R and R_{free} of 0.016% and 0.002%, respectively.

The final model (R / R_{free} = 8.10%/ 9.34%) contained 36 residues with alternate conformations (including four residues with three or more conformations), 148 full-occupancy waters, 435 partially-occupied waters, 11 sulfate ions, and 2 glycerol molecules. Blocked least-squares refinement was performed to calculate an estimate of the errors for atomic positions. The size of the model required that it be broken into three blocks corresponding to (1) residues 1 through 100, (2) residues 98 through 212, and (3) all solvent atoms. The final working R-factor for this model was 8.17% for all data and 7.28% for data with $F_o > 4\sigma(F_o)$.

$F_{o,pH\,5} - F_{o,\,pH\,8}$ difference maps were calculated to compare $\alpha LP_{pH\,5}$ to $\alpha LP_{pH\,8}$ (1SSX.pdb (Fuhrmann et al., 2004a)). Using CNS (Brünger et al., 1998), data from each structure were scaled and then maps calculated using phases from 1SSX.pdb.

αLP bound to mAAP-boroVal...glycerol

The previously-solved structure of αLP bound to methoxysuccinyl-Ala-Ala-Pro-boroAla (with glycerol bound; unpublished data) was used as a starting model for refinement. In order to minimize model bias, all hydrogen atoms, 114 water molecules (partially occupied or having temperature factor greater than 50), one partially-occupied glycerol molecule, and one partially-occupied sulfate ion were removed prior to refinement. Twenty-two alternate conformations of side chains were also removed, and the occupancy of major conformations reset to unity. The inhibitor was not included in this initial model.

Refinement of the model was performed entirely in ShelxL-97, following strategies described previously for the 0.83Å resolution structure of αLP (pH 8) (Fuhrmann et al., 2004a). Rigid body refinement and subsequent minimization of the initial model showed an excellent fit to the data, with R/R_{free} of 18.4% / 20.4%. Early in the refinement, data was restricted to 1.2Å resolution to allow efficient modeling of waters and alternate conformations. Temperature factors were described using an isotropic model initially. Both σ_A-weighted $2F_o$-F_c and difference electron density maps clearly indicated that the inhibitor was covalently bound to the protein at Ser195 Oγ. Electron density was visible as punctate atoms along the length of the inhibitor from boroVal P1 to the methoxysuccinyl-Ala peptide bond. Therefore, the inhibitor was modeled as Ala-Ala-Pro-boroVal, with a resulting R / R_{free} of 16.4% / 18.2%. Next, anisotropic B-factors were added to the model for all atoms, and minimized against data to 1.0Å resolution (R / R_{free} = 12.4% / 14.1%). This was followed by additional modeling of waters and sidechain alternate conformations (R / R_{free} = 11.7% / 13.3%). Difference electron density corresponding to hydrogen atoms was visible for the

majority of hydrogens in the structure. Hydrogen atoms were added to the model as "riding hydrogens," meaning that their position was defined by geometry and not allowed to refine. Hydrogen atoms were not added to methyl, primary amine, or hydroxyl groups, or to key atoms in the active site (including all atoms of the inhibitor and His57, atoms in the oxyanion hole (Gly193 N and Ser 195 N), Ser214 Oγ, and Gly216 N). At this stage, R / R_{free} were 11.2% / 12.5%. Positive difference electron density (contoured to 3.0σ) suggested the presence of a glycerol molecule bound to the boronic acid of the inhibitor. This glycerol was modeled as a secondary adduct to boroVal P1 O2. At later stages of refinement, it became evident that this glycerol molecule occupied two other conformations, both secondary adducts to boroVal P1 O2, with a total occupancy of 71%. Electron density extending off the N-terminal end of the inhibitor was modeled with the full-length methoxysuccinyl group (R / R_{free} = 9.6% / 10.7%). This group refined to an occupancy of 43%, suggesting that other conformations of the methoxysuccinate are disordered. At this time, riding hydrogens were added to methyl, primary amine, and hydroxyl groups (using the AFIX 137 and 147 commands), and to atoms in the inhibitor. In later stages of refinement, it became apparent from positive difference maps that residues MeOSuc P5 through Ala P3 were present in a second conformation, which refined to an occupancy of 22%. Finally, all geometric restraints were removed from residues with single conformations, allowing atomic positions to refine without bias from canonical geometries. "Unrestrained" minimization resulted in a decrease in R by 0.037%, and of R_{free} by 0.049%.

In the last step of refinement, hydrogen atoms were added to key atoms in the active site, and their positions allowed to refine freely. In order to obtain estimates for the errors of atomic coordinates, blocked least squares refinement was performed, breaking the structure

into three blocks: residues 1-100, 98-211, and 209-1000. The final model contained 21 resides with alternate conformations, 160 fully- and 389 partially-occupied waters, 8 sulfate ions, and 5 glycerol molecules. The model is among the most accurate solved to date, with an R of 8.14% and R_{free} of 9.20%.

$F_{o,\ \alpha LP+bVal} - F_{o,\ 1SSX.pdb}$ electron density maps were calculated using methods similar to those for $\alpha LP_{pH\ 5}$ to directly observe differences between the structures of αLP+boroVal(gol) and the apo enzyme at pH 8 (Fuhrmann et al., 2004a). All models were aligned using LSQMAN (Kleywegt & Jones, 1994), and reported RMSD's were calculated for main chain atoms (Cα, C, N) except where otherwise noted. Structural figures were made using Pymol (Delano, 2002) unless otherwise indicated; chemical schematics were drawn in ChemDraw.

Acknowledgments

Thanks are given to Dr. C. Kettner for providing the boronic acid inhibitor and Dr. R.C. Rizzo for access to the Cambridge Structural Database. C.N.F. wishes to thank Dr. M. Garcia-Viloca for originally spurring her interest in the LBHB debate, and Drs. A. Mildvan, J. Kirsch, D. Truhlar, and J. Gao for helpful discussions of hydrogen-bond theory. Portions of this research were carried out at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. This work was funded through the Howard Hughes Medical Institute.

Conclusions

Can ultra-high resolution x-ray diffraction data be obtained for α-lytic protease?

Prior to this work, the highest-resolution structure of αLP was solved at 1.5Å resolution (Rader & Agard, 1997). During the course of my thesis work, I have collected data and begun refinement for twelve distinct structures of αLP, all at resolution of 0.90Å or better (see Appendix E). Some of these structures were not fully pursued due to complications that I later discovered were caused by effects of radiation damage at the Ser195-boronic acid adduct (Chapter 2); others may be pursued by myself or other members of the lab in the future. Three of these structures were solved to completion, and are discussed in detail in this thesis. I assisted Brian Kelch in ultra-high resolution data collection for a thirteenth structure (the double mutant of αLP, Thr181→ Ile / Trp199→ Leu), and he has pursued refinement of that structure for publication (Kelch et al., manuscript in preparation).

What was done differently here to obtain such high resolution data? To my knowledge, I made no changes to the prescribed conditions for growing αLP crystals. The great enhancement in resolution, however, was likely the result of improved techniques and hardware available during data collection. The advent of bright, tunable synchrotron sources allowed efficient data collection at wavelengths where radiation absorption (and therefore damage to crystals) could be minimized (Chapter 2). The freezing of the crystals under a liquid-N_2 cryostream, now a standard practice for data collection, greatly reduces damage caused by heating the crystal and secondary ionization events. The development of large, highly-sensitive detectors with a broader dynamic range aided in recording weak ultra-high

resolution reflections. These advancements have increased the resolution of many protein structures available. At the start of this thesis work, there was only one structure of a protein larger than 100 amino acids deposited in the Protein Data Bank at a resolution at or higher than 0.9Å (0.015% of all structures of similarly-sized protein deposited). In the past seven years, 29 more such structures have been deposited (including my 0.83Å structure of αLP at pH 8), representing 0.16% of the structures deposited during this time period—a ten-fold increase in prevalence of such high resolution.

While pursuing this work, I had ample opportunities to test various techniques for obtaining ultra-high resolution data for αLP. Glycerol and PEG-300 have both been used successfully as cryoprotectants (Chapter 2). Freezing αLP crystals can be done directly in the cryo-stream, and no advantage was observed when using a block-freeze method (where the cryo-stream is blocked while the pin holding the looped crystal is being mounted on the goniometer). If anything, data collected on these crystals was not as high-quality as data collected on crystals mounted in the running stream. I also tried flash-annealing, a method in which the stream is blocked and the crystal is allowed to warm up for a few seconds, then re-freeze. This method has been used to improve resolution of some types of crystals (Harp et al., 1999; Yeh & Hol, 1998), but if anything had a deleterious effect on resolution of αLP crystals. Importantly, soaking peptidyl boronic acid inhibitors into αLP crystals did not noticeably reduce their diffractive power.

Collecting data on large crystals had two benefits. First, crystals with a wide cross-section allowed diffraction by more unit cells, improving the diffractive power of the crystal. Because high-energy wavelengths were used in these experiments (0.75 – 0.80Å), absorption of x-rays by the crystal was not significant (Figure 2.5). Secondly, long crystals provided the

opportunity to spread high-resolution data collection over multiple positions on the crystals, minimizing radiation damage (Chapter 2). In order to obtain large crystals, I made an effort to sample a range of seeding solution concentrations so as to isolate one or two crystals per drop, allowing maximum growth for each individual crystal. In my experience, αLP crystals of width ≥ 0.3mm have been required for obtaining complete data sets at resolution ≤ 0.9Å. Crystals as large as 0.5mm in width have been used, although such large crystals frequently exhibit snowman-shaped reflections suggestive of lattice shifts within the crystal.

When collecting ultra-high resolution data, it is especially important to collect short-exposure, low-resolution data sets before and after high-resolution data collection, (with an "after-high" low-resolution (~1.3Å) data set collected at *every* position of the crystal where high-resolution was collected). Prior to refining any structure, $F_{o,\,after} - F_{o,\,before}$ difference electron density maps should be calculated to compare these low-resolution data sets. Analysis of such maps will reveal areas of structure-specific radiation damage, which may otherwise be misinterpreted as being functionally relevant alternate conformations. When high- and low-resolution data sets are merged, the strong features of radiation damage will likely only be observed if the "after" (and not "before") low-resolution data set is used. In the latter case, the strong intensities of low-resolution data can supersede any effects of radiation damage present in the high-resolution data, and it can go un-noticed in F_o-F_c maps. Extensive radiation damage, however, can introduce systematic errors into the data, making analysis of precise features in the electron density difficult. Methods for minimizing radiation damage in ultra-high resolution data collection are discussed in Chapter 2.

Relatively high-resolution data (to ~1.4Å) can easily be attained from αLP crystals using short exposure times (0.5 – 2 seconds) at today's synchrotron sources, even when

tuning the beam to short wavelengths (which have very low flux). Data to ~1.2Å typically required ~2-fold increase in exposure time compared to ~1.4Å data, and ultra-high resolution data (0.82-0.90Å) required an additional ~20- to 60-fold jump in exposure time. I have performed long-exposure tests on αLP crystals, and the maximum resolution reflections that I have observed were to ~0.76Å. Further studies should be done in the future, however, to determine the current resolution limit for αLP diffraction.

A question frequently asked is, "why does αLP diffract to such high resolution compared to other protein crystals?" It has been proposed that low solvent content (~35%, v/v) contributes to the ability of certain protein crystals to diffract to ultra-high resolution (Schmidt & Lamzin, 2002). This, however, is not the case for αLP, which has a solvent content of ~51% (v/v), close to the average value for all protein crystals (Schmidt & Lamzin, 2002). The excellent diffraction power of αLP may instead lie in its inherent rigidity. As described in Chapter 1, unlike other proteins, αLP undergoes essentially no "breathing motions" in its native state. αLP has extremely low atomic B-factors (12.7Å^2 at room temperature (Fujinaga et al., 1985); 5.4Å^2 at 120 K (Rader & Agard, 1997)), and hydrogen-deuterium exchange experiments have shown that αLP has the highest protection factors known for any protein (Jaswal et al., 2002). Such rigidity likely contributes to the low mosaicity of αLP crystals (typically ~0.2-0.3). The combination of low crystal mosaicity and lack of conformational disorder within the unit cell may be important factors contributing to the excellent diffracting power of these crystals.

What are the structural effects of x-ray radiation on αLP during collection of ultra-high resolution data?

The long exposure times required to obtain ultra-high resolution data frequently result in radiation damage to protein crystals. Effects commonly observed include reduction of disulfide bonds, decarboxylation of Asp and Glu residues, loss of methylthiol groups from Met residues, and dehydroxylation of Tyr residues. Radiation damage has also been observed in enzyme active sites (Weik et al., 2000).

In order to understand the structural changes that occur in αLP upon radiation damage, I studied the effects of radiation damage in αLP complexed with MeOSuc-Ala-Ala-Pro-boroVal-OH. As described above, complete low-resolution data sets were collected before and after the high-resolution data collection. These data sets were then compared (by calculation of $F_{o,\,after} - F_{o,\,before}$ electron density maps) to detect structure-specific effects of radiation damage. The most prominent structural changes were centered at four sites. Three of these corresponded to reduction of the αLP's disulfide bonds, with the frequency of disulfide breakage being Cys42/Cys58 ~ Cys189/Cys220A > Cys137/Cys159 (Figure 2.2). Similar results were attained for other αLP structures (Table 2.1). $F_{o,\,after} - F_{o,\,before}$ difference maps revealed subtle structural changes extending along β-sheets adjacent to these disulfide bonds, with effects occurring as far as 25Å away from disulfide cleavage (Figure 2.2b).

Surprisingly, prominent structural changes due to radiation damage also occurred at the Ser195-boronic acid adduct (Figure 2.2c). The peaks observed in $F_{o,\,after}$-$F_{o,\,before}$ maps agreed with those previously noted in F_o-F_c difference density for various complexes of αLP with peptidyl boronic acids (e.g., Figure 2.3). This difference density proved challenging to

model, and it remains unclear what particular structural changes these peaks correspond to. Interestingly, Ser195-boroValP1 is located adjacent to Cys42/Cys58. Therefore, it is unclear whether the structural changes observed in these electron density maps are a direct effect of radiation damage at this site, or are simply conformational changes caused by radiation damage at Cys42/Cys58. The latter possibility is unlikely, since neither Cys42 or Cys58, in the covalent or reduced forms, are within van der Waals distance from the Ser195-boroValP1 adduct. If the active site is a target of direct radiation damage, the most likely outcome is cleavage of the B-Cα bond within boroValP1, which is expected to be weaker than the B-O bonds. This would result in a trigonal-planar Ser195-$B(OH)_2$ adduct. Modeling the adduct in this way did account for most of the difference electron density observed. Positive density did remain, however, and so the true set of conformational changes occurring in these boronic acid complexes is still unclear.

How is the pK_a of His57 affected by the crystallographic environment?

The general mechanism proposed for proteolytic catalysis by serine proteases predicts that Ser195 hydrogen-bonds to His57 in the free-state of the enzyme (see Introduction and Figure 0.2). Surprisingly, several of the crystal structures of serine proteases in the un-bound state showed Ser195 to be at a distance too far from His57 for this hydrogen bond to form (James et al., 1980; Matthews et al., 1977). Interestingly, the initial crystal structures of αLP raised this conundrum as well (Brayer et al., 1979; Fujinaga et al., 1985). Other crystal structures, however, did support formation of a hydrogen-bond between His57 and Ser195 in the resting state of the enzyme (Bode et al., 1983; Tsukada & Blow, 1985).

The ultra-high resolution crystal structure of αLP solved at pH 8 (Chapter 1) gave clues as to the reason for these confusing results. Here, two conformations of Ser195 were visible. The major conformation (~70% occupancy) mimicked that observed in previous crystal structures of αLP. In the minor conformation of Ser195 (~30% occupancy), however, Oγ was positioned ~1Å away, in a position favorable for hydrogen-bond formation with His57 (Table 1.2, Figure 1.3). Interestingly, comparison of this structure with transition state analog structures of αLP showed that the major conformation of Ser195 was the same as that observed when Ser195 Oγ participates in formation of the tetrahedral intermediate structure. I therefore hypothesized that His57 was doubly-protonated in a sub-population of molecules within the crystal, and that this population displayed the alternate conformation of Ser195. This hypothesis was confirmed upon solving the crystal structure of αLP at pH 5, where His57 should be doubly-protonated in 100% of the protein molecules (Chapter 3). In $\alpha LP_{pH\,5}$ only one conformation of Ser195 was present, and (as expected) this conformation corresponded to the major conformation observed in $\alpha LP_{pH\,8}$ (Figure 3.2).

Because αLP's His57 has a pK_a of ~7, we were surprised that 70% of the molecules in our αLP crystal (equilibrated at pH 8) were doubly-protonated. This result suggested that some characteristic of αLP crystals influenced the pK_a of His57. Utilizing the occupancies of the Ser195 conformations, I calculated that the average pK_a of His57 in our αLP crystal was 8.4 (see Chapter 1), 1.4 units above the published value.

What characteristics of αLP crystals caused such a shift in pK_a? The likely explanation was that the negative charge of a nearby sulfate ion (Sul202; Figure 1.3) was stabilizing the positively-charged state of His57, increasing its pK_a. To what degree does this sulfate have such an effect? In my crystal, Sul202 was bound in the active site in ~77% of the

protein molecules. Assuming that the presence of Sul202 was the only influencing factor, I calculated that in the presence of Sul202 the pK_a for His57 rises to 8.8 (Chapter 1).

NMR chemical shift experiments have previously been used to determine the average pK_a of αLP's His57 in the presence of sulfate ion, both in solution and in the crystalline form (Farr-Jones, 1989; Smith et al., 1989). However, NMR experiments afford only the opportunity to calculate the average pK_a. In contrast, ultra-high resolution crystallography allowed calculation of the microscopic pK_a of His57 in the presence of sulfate. Upon taking the concentration of sulfate ions into account, the value we calculated was in good agreement with the average values determined experimentally by NMR (Chapter 1). Therefore, we conclude that the major factor influencing the pK_a of His57 in αLP crystals is the presence of a nearby sulfate ion.

This phenomenon raises the question: can αLP crystals be grown without Sul202 bound in the active site, allowing observation of His57 and Ser195 in their native states? In crystals of the free enzyme, Sul202 interacts electrostatically with Arg192 and with Arg122 in a neighboring symmetry-related molecule. Its role in forming this crystal contact suggests that Sul202 (or another negatively-charged ion) is required for crystal formation. However, once the crystals are formed, Sul202 is not required for maintenance of stable crystals. For example, crystals remain structurally stable during soaks with peptide boronic acid inhibitors. These inhibitors displace Sul202. As a result, Arg122 and Arg192 occupy different conformations in which they satisfy electrostatic stabilization through a crystal contact with Glu174 and a sulfate ion, respectively. This observation suggests that Sul202 may not be necessary for crystal growth. Future crystallographic studies could address this question by

trying other crystallographic conditions for crystal formation, or by trying to soak in other molecules that might displace Sul202 from the active site.

How does the enzyme promote forward progression through the TI_1 → EA transition?

In the early 1970s, enzymologists raised the question: if His57 is hydrogen-bonding to Ser195 in the catalytic triad, what prevents this interaction in the tetrahedral intermediate state? How is protonation of the substrate amide favored over protonation of Ser195 in the TI_1 → EA transition (Polgar, 1972; Satterthwait & Jencks, 1974; Wang, 1970)?

In 2000, the Bachovchin group proposed a mechanism by which this conundrum could be overcome: a flipping of the His57 ring (via a 180° rotation about the Cβ-Cγ bond) occurs in the transition states (Ash et al., 2000; Haddad et al., 2005). In this conformation, they argued, His57 would be better positioned to hydrogen-bond to Ser195, promoting formation of the acyl enzyme. This proposal was based on the observation of a C-H...O hydrogen bond formed between His57 Cε1 and a nearby backbone carbonyl, which could be substituted in the flipped state with a hydrogen bond through His57 Nε2 (for a more thorough discussion, see Chapter 3). Because of the energetic cost required to flip His57 (and thereby disrupt three hydrogen-bonds formed in its canonical state), this mechanism has not been given much credence by other enzymologists (Hedstrom, 2002b). However, the question remains: how does the enzyme promote forward progression through the TI_1→ EA transition?

The structure of αLP+boroVal(gol) provides an excellent mimic of the TI_1 state (Chapter 3).At sub-Ångstrom resolution, αLP+boroVal(gol) definitively shows that His57 is

in the canonical, not flipped, conformation. Furthermore, anisotropic B-factors show no evidence for wobbling or flipping motions of the His57 ring (Figure 3.6). In fact, the structure suggests that such a mechanism is not even required for forward progression of the reaction. In αLP+boroVal(gol), His57 is already aligned to form an ideal hydrogen bond to the substrate amide leaving group, and forms at most a weak hydrogen bond with Ser195 (Table 3.2). Therefore, upon formation of TI_1, the catalytic triad is already aligned to promote forward progression of the reaction. This suggests that some conformational change within the catalytic triad occurred in the E→ ES or ES→ TI_1 transitions. What is this conformational change?

Analysis of the $\alpha LP_{pH\,8}$ and $\alpha LP_{pH\,5}$ crystal structures provides a clue as to how this is achieved. As discussed above, two conformations of Ser195 were observed in $\alpha LP_{pH\,8}$, which corresponded to the protonated and unprotonated forms of His57 (Chapters 1, 3). We concluded that upon protonation of His57 at Nε2, the sidechain of Ser195 moves to a second conformation, which is the conformation of Ser195 observed in the TI_1 state (Chapter 3). Based on this conclusion, we propose that during the ES → TI_1 transition, protonation of His57 by Ser195 causes a similar conformational change of Ser195. In this way, deprotonation of Ser195 by His57 catalyzes the first step of the acylation reaction in two ways: (1) by increasing the nucleophilicity of Ser195, and (2) by altering the position of Ser195 to one that disfavors hydrogen-bonding to His57 Nε2, preventing the back-reaction. A secondary benefit of (2) is that the catalytic triad becomes positioned to favor hydrogen-bonding between His57 and the substrate leaving group, promoting the next step of the acylation reaction (TI_1 → EA). The conformational changes described in this hypothesis

would contribute towards the high energy of the TS_1 state, formation of which constitutes the rate-limiting step of the proteolytic reaction.

How does the chemical structure of a protease-peptidyl boronic acid complex compare with that of a true tetrahedral intermediate?

By definition, no stable transition state analog structure is a perfect mimic of its corresponding metastable tetrahedral intermediate or transition states. Therefore, understanding the limitations of particular transition state analogs is necessary for appropriate interpretation of experimental data.

Here, we have used peptidyl boronic acids to mimic the tetrahedral intermediate and transition states. These inhibitors are designed to bind the catalytic Ser195 as a stable tetrahedral adduct (via the boron), mimicking the TI_2 state (Kettner & Shenvi, 1984; Kettner et al., 1988). Our experiments using glycerol as a cryoprotectant resulted in crystal structures in which a glycerol molecule bound covalently to the boronic acid, filling the S1' pocket and providing a mimic of TI_1 (αLP+boroVal(gol), Chapters 2 and 3). In this structure, hydrogen bonding interactions observed in previous mimics of the TI_1 state (Bone et al., 1991b) were not mimicked by the glycerol. However, many other characteristics of the αLP+boroVal(gol) structure did illustrate its advantages as a TI_1 mimic.

As expected, in the presence of the inhibitor αLP's catalytic triad maintains the geometry for the TI_1 state, without any significant conformational changes (as has been seen upon binding other inhibitors). Importantly, a hydrogen atom is observed on His57 Nε2 (Figure 3.5), confirming that in this analog (as is expected for a true transition state or

intermediate) His57 is doubly protonated. Also as expected, geometry about the boron is tetrahedral (Table 3.3). Unlike other transition state analogs (such as trifluoromethyl ketones and aldehydes), peptidyl boronic acids display a hydrogen bond acceptor (typically a hydroxyl group) in the active site, allowing formation of a hydrogen bond with His57 Nε2. Formation of this hydrogen bond mimics a key step in the acylation reaction (see above). In αLP+boroVal(gol), this acceptor is an ether oxygen (mimicking the amide of the scissile bond), and is in excellent geometry to accept the proton from His57 Nε2 (Table 3.2). As discussed below, the short donor-acceptor distance for this hydrogen bond (as represented in αLP+boroVal(gol)) would be expected to facilitate transfer of the His57 Nε2 hydrogen atom to the amide leaving group, promoting cleavage of the peptide bond and formation of the acyl enzyme. The structural interpretation of αLP+boroVal(gol)—that is, that the TI_1 state is more similar in structure and energy to TS_1' than to TS_1—agrees with recent predictions of the energy profile for acylation of a true substrate as calculated using QM/MM methods (Ishida & Kato, 2003).

Although αLP+boroVal(gol) is an excellent mimic of the hydrogen bonding interactions expected for the catalytic triad in the TI_1 and TS_1' states, some caveats do need to be considered when analyzing boronic acid structures. First, a true tetrahedral intermediate complex is expected to contain an oxyanion, which results from the nucleophilic attack of Ser195 on the substrate carbonyl. This oxyanion is stabilized through hydrogen bonding interactions with the enzyme's backbone amide groups in the oxyanion hole. In αLP+boroVal(gol), a clear signal was observed in positive F_o-F_c difference maps adjacent to boroValP1 O1, which resides in the oxyanion hole (Figure 3.4). This signal was strongly indicative of a hydrogen atom, and refinement of a hydrogen atom in that position resulted in

disappearance of this difference density. Although many scientists in the field have expected that peptidyl boronic acids are protonated in the oxyanion hole, this was the first decisive experimental evidence of this.

If the boronic acid is protonated in the oxyanion hole, then the negative charge on the complex likely resides primarily on the central boron. This was confirmed by observation of positive F_o-F_c difference density on the boron (indicating the presence of more electrons than present in a typical boron atom; Figure 3.4). When the occupancy of the boron was refined, its electron content was determined to be 5.3 electrons. Whether this value represents a partial negative charge or is an underestimate of the six electrons expected for a full formal charge is unknown. Therefore, the charge distribution within the active site is different in complexes with peptidyl boronic acids as compared to true substrates. In a tetrahedral intermediate complex, three charged residues interact via hydrogen-bonding: Asp102, His57, and the tetrahedral intermediate. For a true substrate, the negative charge residues 4-5Å away from His57 Nε2, in the oxyanion hole. In a peptidyl boronic acid complex the negative charge residues mainly on the boron, which is ~1.2Å closer to His57. This could have the effect of attracting His57 away from Asp102, lengthening the His57-Asp102 hydrogen bond and shortening the His57-boroValP1 hydrogen bond. Therefore, it is important to test our interpretations of this structure by comparing our results to other transition state analog structures at ultra-high resolution.

Is there a low-barrier hydrogen bond between His57 and Asp102 in the acylation tetrahedral intermediate and/or transition states?

One of the major goals for my thesis was to use ultra-high resolution crystallography of αLP transition state analogs to address the LBHB hypothesis. According to this hypothesis, proposed in 1994 by Frey *et al.*, a LBHB is formed between His57 and Asp102 in the first transition state, enhancing the basicity of His57 and thereby catalyzing the acylation reaction (Frey et al., 1994; Frey, 2004). A LBHB would be particularly effective at stabilizing the transition state because of the high energy stored in this special, covalent hydrogen bond. Structurally, a LBHB is characterized by a short distance of < 2.65Å between donor and acceptor atoms, and by the location of the hydrogen atom equidistant between donor and acceptor (Frey et al., 1994). Most evidence for a LBHB existing in the TS_1 state is indirect, coming from ^1H NMR chemical shifts, D/H fractionation factors, and deuterium isotope effects detected for transition state analog structures. In 1998, a 0.78Å resolution structure of subtilisin at pH 5.9 provided the first and only direct evidence for a LBHB between His and Asp in a catalytic triad (Kuhn et al., 1998). In this structure, the distance between His Nδ1 and Asp Oδ2 was 2.62Å, and density attributed to a hydrogen atom was observed nearly equidistant between the two atoms. However, as a low-pH mimic of the transition state, the subtilisin structure was a non-ideal transition state analog. In addition to lacking a tetrahedral adduct and corresponding negative charge of the oxyanion, the structure was described to have disordered peptide proteolysis products present in the binding pocket; both of these characteristics could have an effect of pushing His towards Asp, inducing a LBHB. αLP's ability to diffract to sub-Ångstrom resolution afforded the

excellent opportunity for collecting data on a more appropriate mimic of the acylation transition state.

To address this question, I solved two transition state analog structures of αLP: αLP at pH 5.1 (to mimic the doubly-protonated state of His57 in the TS and TI states) and αLP bound to MeOSuc-Ala-Ala-Pro-boroVal... glycerol ("αLP+boroVal(gol)", as a mimic of TI_1) (Chapter 3). In both of these structures, the distance between His57 Nδ1 and Asp102 Oδ2 was too long to be a LBHB (2.76Å and 2.73Å, respectively; see Table 3.2). Furthermore, F_o-F_c difference density corresponding to a hydrogen atom was visible adjacent to Nδ1 (and not equidistant between Nδ1 and Oδ2, as is required for a LBHB) in both structures (Figure 3.5). Therefore, these two chemically-distinct transition state analogs showed that a LBHB is likely not present in the transition states for serine protease catalysis by αLP.

Although we observed electron density adjacent to Nδ1 corresponding to a hydrogen atom, this electron density peak was much weaker than peaks corresponding to other atoms in the His57 ring (Figure 3.5b). The weak intensity of this Nδ1-H peak may indicate a shift in electron density through the His57 ring in the TI_1 state. Quantum mechanical studies should be pursued to calculate the distribution of electron density within His57 in this structure.

In the previous section, I pointed out the possibility that our observation of a "lengthened" hydrogen bond between His57 and Asp102 in αLP+boroVal may be a result of the proximity of the negatively-charged boron to the positively-charged His57. A similar argument could be made for $αLP_{pH\ 5}$, which has a sulfate ion bound in the active site (Figure 3.2). In spite of the possible repercussions of these effects, the conclusion that

His57...Asp102 is not a LBHB remains a significant result. Importantly, our data points out the caveats of other evidence used to defend the LBHB hypothesis. As discussed in Chapter 3, the evidence presented most frequently in support of the LBHB proposal is ^{1}H NMR data showing that the His57 Nδ1-H is shifted far downfield, which is interpreted as being involved in a LBHB. Rather, our data for αLP+boroVal(gol) and $\alpha LP_{pH\,5}$ show that a strong downfield chemical shift, as has been observed for His57 under both of these conditions (Bachovchin, 1986; Bachovchin et al., 1988), is not necessarily indicative of a LBHB. Rather, we provide experimental evidence that such chemical shift data can be due to a standard ionic hydrogen bond, as predicted previously by theoretical methods (GarciaViloca et al., 1997a; Guthrie, 1996; Shokhen & Albeck, 2004; Warshel & Papazyan, 1996).

The most extensive studies supporting a LBHB between His57-Asp102 have been performed on chymotrypsin complexed with trifluoromethyl ketone inhibitors. As discussed in Chapter 3, the observation of a short hydrogen bond in these complexes could be an indirect result of the chemical characteristics of this mimic. The trifluoromethyl group, located where the scissile amide would be in a standard substrate, may sterically clash with His57, pushing this residue towards Asp102 and inducing a short ionic or low-barrier hydrogen bond. Although experimental evidence collected on these inhibitors is consistent with formation of a LBHB between His57 and Asp102, there has been no direct observation of the hydrogen atom. Ultra-high resolution crystallography or neutron diffraction should be pursued for chymotrypsin-trifluoromethyl ketone complexes in order to distinguish the His-Asp hydrogen bond as a short ionic hydrogen bond or a LBHB.

How does αLP catalyze proteolysis?

Although it is doubtful that a LBHB between His57 and Asp102 stabilizes the transition states in serine protease catalysis, the αLP+boroVal(gol) did suggest another mechanism by which hydrogen bonding interactions aid in catalysis.

In αLP+boroVal(gol) we observe a short ionic hydrogen bond between His57 Nε2 and boroValP1 O2, the mimic of the substrate leaving group (Table 3.2). The short distance between donor and acceptor implies a low energy barrier for proton transfer, which would promote the $TI_1 \rightarrow EA$ step of the reaction (Chapter 3).

The His57...boroValP1 hydrogen bond is one of three hydrogen bonds contacting the boronic acid directly (Figure 3.7a). Interestingly, these three hydrogen bonds are among the shortest hydrogen bonds in the structure (Figure 3.7b, Table 3.2). Similarly, three hydrogen bonds that act to position His57 and Asp102 in the catalytic triad are also very short. Collectively, these six hydrogen bonds are expected to be those most important to stabilization of the transition state. It is therefore advantageous to the protease for these hydrogen bonds to be short and thus stronger in interaction energy.

In contrast to the network of short hydrogen bonds identified amongst the catalytic residues, the three hydrogen bonds by which the enzyme binds the length of the substrate (along the backbone of residues P1 and P2) have a donor-acceptor distance close to or longer than the average for N...O hydrogen bonds (Figure 3.7). This suggests that the enzyme has evolved to preferentially maximize transition state stabilization over substrate binding.

This type of evolutionary pressure has been observed previously in studies of serine protease catalysis. Kinetic studies have shown that as substrate length decreases from three to one amino acid, k_{cat} / K_m decreases 7500-fold (Table 4.1, Bone, R., Kettner, C., Agard, D., unpublished results). Surprisingly, this increase in catalytic efficiency is not due to a decrease in K_m, but rather to a large increase in k_{cat}. This effect is seen amongst many different proteases, and has been called the "k_{cat} effect." It seems that the enzyme has evolved two mechanisms by which to stabilize the transition state, one short-range (strong hydrogen bonds interacting within the catalytic triad and active site) and one longer-range (via binding interactions along the length of the inhibitor). Although the benefit of short hydrogen bonds is intuitive, the question of how lengthening the substrate reduces the k_{cat} remains a mystery. How is binding energy converted into transition state stabilization?

Table 4.1. Kinetic constants for substrate cleavage by αLP (the k_{cat} effect)

Substrate P_3 P_2 $P_1 \Downarrow P_1'$	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m
suc-Ala-p-Na	0.013 ± 0.001	44 ± 4	0.28 ± 0.003
suc-Ala-Ala-p-Na	0.39 ± 0.02	35 ± 2	11.3 ± 0.1
suc-Ala-Ala-Ala-p-Na	32 ± 1.5	15 ± 0.7	2100 ± 32
Inhibitor P_4 P_3 P_2 P_1	K_i (nM)		
Acetyl-boroVal	722000		
Acetyl-Pro-boroVal	3300		
Acetyl-Ala-Pro-boroVal	24		
m-Ala-Ala-Pro-boroVal	6.4		

* Bone, R; Kettner, C.; Agard, D.A., unpublished results.

One possibility is that binding events influence dynamics at the active site, with optimal dynamics (in direction and/or amplitude) achieved when all binding pockets are bound by substrate. If dynamics do play a role in catalysis, then the short hydrogen bonds observed in the active site and among the catalytic triad may play a role in optimizing (or perhaps minimizing) these dynamic motions. Because inhibition of αLP by peptidyl boronic acids of different lengths follows the same trend as the k_{cat} effect (Table 4.1), crystallographic studies of these inhibitor complexes could be used to test this hypothesis. The excellent ratio of data to parameters afforded by ultra-high resolution crystallography allows calculation of anisotropic B-factors and even the modeling of multiple conformations for every atom, providing dynamical information about the protein. Do the direction or amplitude of the principle axes of motion change within the enzyme active site or along the inhibitor depending on inhibitor length? Results from these studies could be compared to results from computational approaches, such normal mode analysis.

What are the structural features of αLP that contribute towards its high folding and unfolding energetic barriers?

The energetic characteristics of αLP's folding and unfolding barriers are well-understood. However, a major unanswered question is: how has αLP achieved such a large and cooperative unfolding energy barrier? What structural features of αLP have created such a high-energy transition state for folding? The discovery of a distorted phenylalanine residue (Phe228) in the core of the protein may provide some of these answers. This distortion, described for $\alpha LP_{pH\,8}$ in Chapter 1, extends along the length of the Phe228 sidechain,

including a distortion of ~5° for Cα-Cβ-Cγ and of ~6° along the benzene ring (calculated between Cγ and Cζ) (Figure 1.5). First identified in $\alpha LP_{pH\,8}$, this distortion has been observed in all other ultra-high resolution crystal structures of αLP solved to date (data not shown), and for extremophilic, large-pro region-containing homologs of αLP (Kelch & Agard, manuscript in preparation; Kelch et al., manuscript in preparation). However, this distortion is not observed in proteases requiring a short pro region or no pro region for folding. What causes this structural difference, and is the distortion functionally significant?

Interestingly, sequence alignments of pro region-containing extracellular proteases showed that every residue surrounding Phe228 co-varied with pro region size (Figure 1.6), while in contrast Phe or Phe-like residues were strongly conserved at position 228. Structural alignments of proteases representing the small- and large-pro region-containing protease subfamilies revealed that choice of rotamer, and hence positioning of Phe228 in three-dimensional space, was determined by the residues around it (Figure 1.6). This finding challenges the paradigm that a conserved residue must adopt the same conformation in homologous proteins, and has implications for fields such as bioinformatics, structure prediction, and molecular modeling.

Studies on αLP homologs have revealed that pro region size correlates with the height of the folding and unfolding barriers for these proteases (Truhlar et al., 2004). Therefore, our observation that Phe228's position and degree of distortion co-varies with pro region size suggests that the conformational strain of Phe228 was evolved as a mechanism to increase the folding and unfolding barriers of kinetically stable proteases.

This hypothesis is supported by recent mutagenic studies of αLP's C-terminal core. Kinetic analyses of mutants directly or indirectly eliminating the Phe228 distortion in all

cases showed an increase in both folding and unfolding rates (Kelch et al., manuscript in preparation). The fact that folding rates increased for all αLP mutants, despite loss of packing interactions in the core of the protein, points to the significance of the Phe228 distortion. For example, a mutant in which Phe228 is mutated to Ala folds ~6-fold faster than wild-type αLP, despite a major disruption of the hydrophobic core (Kelch et al., manuscript in preparation). For all other known experiments in which a core Phe residue was mutated to Ala, decreased folding rates were observed (B. Kelch, personal communication). Together this data suggests that the Phe228 distortion does indeed increase the energetic barrier for folding of αLP.

Interestingly, Brian's results also suggest that Phe228 is more strained in the unfolding transition state than in the native state. All mutations that relieve the strain preferentially stabilize the transition state over the native state (Kelch et al., manuscript in preparation). For example, kinetic studies of a triple mutant of αLP designed to mimic the core of SGPB (Thr181→ Ile, Trp199→ Leu, Gln210→ Ile) show that the transition state was stabilized by ~2 kcal/mol more than was the native state relative to wild-type αLP. This mutant is particularly interesting, as it induces the SGPB-like conformation of Phe228 while maintaining a relatively tightly-packed core. Based on these results, we predict that Phe228 is even more distorted in the transition state than it is in the native state, and thereby increases the height of the unfolding barrier. We conclude that strain plays a functional role by increasing the unfolding barrier, enhancing the kinetic stability of large-pro region-containing proteases and thus increasing protease longevity.

Another way to test this hypothesis is to reverse this approach, mutating the core of SGPB to be more like αLP; that is, create a triple mutant of SGPB, Ile181 → Thr, Leu199→

Trp, Ile210→ Gln. The expectation is that the resulting arrangement of residues surrounding Phe228, in particular Trp199, would force Phe228 into the conformation observed in αLP, inducing a similar distortion. This would be expected to destabilize both the transition and native states, with a stronger destabilizing effect on the transition state.

To our knowledge, this is the first geometric distortion that has been shown to be functionally relevant. A survey of 35 ultra-high resolution structures deposited in the PDB (all containing at least 151 amino acids) revealed 25 structures with distorted Phe, Tyr, Trp, or His rings (9 structures with distortions as significant as that observed in αLP) (data not shown). Although such distortions might be considered artifacts in lower resolution structures, at ultra-high resolution electron density for well-ordered residues (with low atomic B-factors and high occupancy) these represent true observations. Some of these distortions may be due to strain introduced upon crystal formation. However, our discovery of a functionally-relevant geometric distortion in αLP suggests that other observed geometric distortions may also be relevant, either to protein stability or to other functions of the protein. For example, structural strain could be utilized to modulate the energetics of conformational changes associated with allostery and catalysis. As the number of sub-Ångstrom resolution crystal structures increases, similar surveys of geometric distortions should be pursued, and in appropriate cases mutagenic studies done to test for the relevance of such distortions.

Appendix A. Key to the chymotrypsin convention of labeling αLP residues

In order to facilitate comparison of αLP to members of the larger chymotrypsin family of serine proteases, residues are labeled in the text of this thesis using chymotrypsin homology numbering (Fujinaga et al., 1985). Appendix A provides the key to compare this homology numbering scheme with the sequential numbering for αLP.

The catalytic residues are highlighted with bold squares, and residues involved in distortion of Phe228 (Chapter 1) are highlighted with ovals.

(Appendix A, caption on previous page)

aLP Sequential	1	2	3	4	5	6	7	8	9	10
Cht Homology	**15A**	**15B**	**16**	**17**	**18**	**19**	**31**	**32**	**33**	**34**
	ALA	ASN	ILE	VAL	GLY	GLY	ILE	GLU	TYR	SER
	11	12	13	14	15	16	17	18	19	20
	35	**36**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**44A**
	ILE	ASN	ASN	ALA	SER	LEU	CYS	SER	VAL	GLY
	21	22	23	24	25	26	27	28	29	30
	45	**46**	**47**	**48**	**48A**	**48B**	**48C**	**49**	**50**	**51**
	PHE	SER	VAL	THR	ARG	GLY	ALA	THR	LYS	GLY
	31	32	33	34	35	36	37	38	39	40
	52	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**59A**	**59B**
	PHE	VAL	THR	ALA	GLY	HIS	CYS	GLY	THR	VAL
	41	42	43	44	45	46	47	48	49	50
	60	**61**	**62**	**64**	**65**	**66**	**67**	**81**	**82**	**83**
	ASN	ALA	THR	ALA	ARG	ILE	GLY	GLY	ALA	VAL
	51	52	53	54	55	56	57	58	59	60
	84	**85**	**87**	**88**	**88A**	**89**	**90**	**91**	**94**	**95**
	VAL	GLY	THR	PHE	ALA	ALA	ARG	VAL	PHE	PRO
	61	62	63	64	65	66	67	68	69	70
	100	**101**	**102**	**103**	**104**	**105**	**106**	**107**	**108**	**109**
	GLY	ASN	ASP	ARG	ALA	TRP	VAL	SER	LEU	THR
	71	72	73	74	75	76	77	78	79	80
	110	**111**	**112**	**113**	**114**	**119**	**120**	**120A**	**120B**	**120C**
	SER	ALA	GLN	THR	LEU	LEU	PRO	ARG	VAL	ALA
	N-terminal domain			*C-terminal domain*						
	81	82	83	84	85	86	87	88	89	90
	120D	**120E**	**120G**	**120H**	**120I**	**120J**	**120K**	**121**	**122**	**123**
	ASN	GLY	SER	SER	PHE	VAL	THR	VAL	ARG	GLY

(Appendix A, continued)

aLP Sequential	91	92	93	94	95	96	97	98	99	100
Cht Homology	**124**	**125**	**129**	**130**	**131**	**132**	**133**	**134**	**135**	**136**
	SER	THR	GLU	ALA	ALA	VAL	GLY	ALA	ALA	VAL
	101	102	103	104	105	106	107	108	109	110
	137	**138**	**139**	**140**	**141**	**142**	**143**	**156**	**157**	**158**
	CYS	ARG	SER	GLY	ARG	THR	THR	GLY	TYR	GLN
	111	112	113	114	115	116	117	118	119	120
	159	**160**	**161**	**162**	**163**	**164**	**165**	**166**	**167**	**168**
	CYS	GLY	THR	ILE	THR	ALA	LYS	ASN	VAL	THR
	121	122	123	124	125	126	127	128	129	130
	169	**170**	**171**	**173**	**174**	**175**	**176**	**177**	**178**	**179**
	ALA	ASN	TYR	ALA	GLU	GLY	ALA	VAL	ARG	GLY
	131	132	133	134	135	136	137	138	139	140
	180	**181**	**182**	**183**	**184**	**185**	**189**	**190**	**191**	**192**
	LEU	THR	GLN	GLY	ASN	ALA	CYS	MET	GLY	ARG
	141	142	143	144	145	146	147	148	149	150
	193	**194**	**195**	**196**	**197**	**198**	**199**	**200**	**201**	**201A**
	GLY	ASP	SER	GLY	GLY	SER	TRP	ILE	THR	SER
	151	152	153	154	155	156	157	158	159	160
	202	**207**	**208**	**209**	**210**	**211**	**212**	**213**	**214**	**215**
	ALA	GLY	GLN	ALA	GLN	GLY	VAL	MET	SER	GLY
	161	162	163	164	165	166	167	168	169	170
	216	**217**	**218**	**219**	**219A**	**219B**	**219C**	**219D**	**220**	**220A**
	GLY	ASN	VAL	GLN	SER	ASN	GLY	ASN	ASN	CYS
	171	172	173	174	175	176	177	178	179	180
	221	**222**	**222A**	**222B**	**222C**	**223**	**224**	**225**	**226**	**227**
	GLY	ILE	PRO	ALA	SER	GLN	ARG	SER	SER	LEU
	181	182	183	184	185	186	187	188	189	190
	228	**229**	**230**	**231**	**232**	**233**	**234**	**235**	**236**	**237**
	PHE	GLU	ARG	LEU	GLN	PRO	ILE	LEU	SER	GLN
	191	192	193	194	195	196	197	198		
	238	**239**	**240**	**241**	**242**	**243**	**244**	**245**		
	TYR	GLY	LEU	SER	LEU	VAL	THR	GLY		

Appendix B: Protocol for αLP expression in *Lysobacter enzymogenes*

(I modified the protocol, using as a starting point Jonathan Davis's Thesis, Chapter 2, 1996)

Bachovchin's Media (Recipe for 1L media)

Casamino acids		2 g
MSG	(187.14g/mol)	12 g
		(or glutamic acid + NaOH to pH 7.0)
$K_2HPO_4 \cdot H_2O$	(174.18g/mol)	1.2 g
NaCl	(58.44 g/mol)	1.2 g
$MgSO_4 \cdot 7H_2O$	(246.48 g/mol)	0.6 g
$ZnSO_4 \cdot 7H_2O$	(287.54 g/mol)	0.004 g
$MnSO_4 \cdot H_2O$	(169.01g/mol)	0.003 g
$Fe_2(SO_4)_3$	(399.9 g/mol)	0.015 g
		(add $Fe_2(SO_4)_2$ to each flask individually)

Adjust to pH 7.0 (with HCl).

Autoclave.

Add:	Sucrose (sterile filtered)	60 g **(see note, next page)
	MEM vitamins (100x)	1:100 (increases yield by 10%)

Can make stocks of metals, sterilize, and add to 1x media.

Fe_2SO_4 remains solid even in 1x media, but other ingredients should dissolve.

Expression Protocol:

1. Grow up L-plates from freezer stock (see box marked lysobacter) → 2 nights at 28° C

2. Pick 1 colony in 100 mL Bachovchin's media (in 500 mL flask) → grow over 1-2 nights at 74 rpm (28°C) *(I had good success at 100 rpm, June2001)*

3. Inoculate 300 mL Bachovchin's media (in 2800 mL triple-dimpled Fernbach) from the overnight, 1:50 dilution (*I inoculated with 10mL*)→ grow at 28°C @ 72-74 rpm (*again, I used 100rpm; but watch to prevent frothing!!*)

4. Maintain pH < 9.0 (pH tends to increase over time in Bachovchin's Media)

5. Assay activity until it plateaus (monitor cleavage of substrate, suc-APA-paranitroanilide, at 410 nm)

6. Growth probably for 2-3 days. *(expected yield 100-400mg αLP)*

** *Sucrose appears to be very important! 10g/L sucrose gives yield of only about 30mg/L αLP. 80g/L gives about 200 mg/L αLP; 60g/L gave most αLP (about 400mg/L). (June 2001)*

Appendix C. Overview of αLP purification

1. Cation exchange 1 (performed at 4 °C)
    a. Batch-bind supernatant to S-sepharose beads
    b. Wash with 10 mM glycine, pH 8.4
    c. Bump off column with 200 mM NaCl (in 10 mM glycine, pH 9.6)
2. Cation exchange 2
    a. HPLC 40mL column (by Vydac)
    b. Wash with 10 mM NaH_2PO_4 (pH 7.2)
    c. Elute by salt gradient, αLP elutes at about 0.12 M NaOAc (in 10 mM NaH_2PO_4, pH 7.2)
3. Monitor yield by activity assays-- monitor cleavage of substrate (suc-APA-pNa) at 410 nm in 0.1 M Tris-Cl, pH 8.0)
4. Check for purity by MALDI mass spec
    a. αLP should give a peak at ~19850
    b. If a second band appears at 19280, this likely corresponds to an autolysis product (for which the six C-terminal residues have been cleaved)
5. Remove salt by dialysis in ddH_2O
6. Store at -80°C (although is fairly stable at 4°C, or even 25°C)
7. Concentrate to 15-20 mg/mL for crystallization

Appendix D. Protocol for growing αLP crystals

Well solution:	(0.5-1 mL sln. per well, filter solution to remove dust before using)	
	1.3 M Li_2SO_4	(128.0 g/mol → 83.2 g for 500 mL)
	20 mM Tris	(121.14 g/mol → 1.2114 g for 500 mL)
	pH _.00 w/ H_2SO_4	(~30-40 mL of 0.1M H_2SO_4 for pH 8.00)
		(note: concentrated H_2SO_4 = 36 Normal)
Drop composition:	5 uL protein (15-20 mg/mL))	
	5 uL well solution	
	(spin protein @43,000 rpm for 20 min. 2x to remove dust; use new special dust-free centrifuge tubes each time)	

Day 1: Set up trays

1. Transfer "dust-free" protein and well solution to new centrifuge eppindorf tube (1:1 ratio = protein:well sln.). Then, centrifuge one more time @43,000 rpm for 20 min to be sure dust is at bottom of tube.
2. Label tray on one side; label tray lid on adjacent side. LABEL: Tray #, protein, your initials, date, well solution (include pH!)

3. For first row of tray:
    a. Grease rims
    b. Add 1 mL well solution
    c. Cover slip w/protein drop
        i. Dust cover slip with compressed air
        ii. Apply 10 ul drop (protein + well solution) to slip. I've found that the best technique is to pipet some of the solution out, THEN touch that solution drop to the cover slip. The pipet tip should never touch the glass. (The best tips for this purpose are the Fisherbrand 30 uL filtered tips, cat #02-707-54—the solution doesn't tend to ride up on these tips as it does on others)
        iii. Set slide upside-down on grease
    d. Check all coverslips/grease seal in row to be sure they are air-tight
4. Repeat for all rows

Day 2: Set up trays (see above)

Day 4: Seed crystals

1. Prepare seed crystals (see Hampton Seed Bead Kit)
    a. Prepare seed bead tube with 50 uL well solution at edge of tube (keep tube laying open, on its side

b. Loop 1 small-medium crystal, and place in drop of solution in seed bead tube. Check loop to be sure that crystal dropped into solution

c. Carefully slide solution to bottom of seed bead tube, being sure that crystal comes with it (no small drops stay behind on side of tube)

d. Vortex tube 90 seconds, stopping every 20 seconds

2. Make serial dilutions of seed bead mixture: "1/1,000" (add 950 uL to seed bead tube), "1/4,000", "1/16,000", "1/32,000"

3. Label lid of tray according to how dilutions will be applied (I usually do columns = 1/1000, 1/4000, 1/4000, 1/16000, 1/16000, 1/32000), and indicate on tray lid the date of seeding.

4. Add 1 uL of appropriate seed solution to each drop. Use the P2 pipetman. Make sure that the grease-slide seal is air-tight (double-check this when the entire tray is done).

Alternative method (using 2 seed solutions)

Day 1: Set up all trays

Day 3: Seed crystals on 1/2 trays

Day 4: Seed crystals on rest of trays

Appendix E. Description of αLP structures solved

These tables are meant to document some of the αLP structures that may not have been completed during the tenure of this thesis. The tables summarize the purpose for solving the structures, collection strategy, resolution, final stage of refinement, and preliminary results. Complete information is available in laboratory notebooks and data backup DVDs.

All work was done by CNF (protein purification, crystal growth, data collection, data processing, and structure refinement) unless otherwise noted.

Structure	**WT αLP at pH 8 (PDB code = 1SSX)**
Date of collection	January 7, 2000
Goal for Experiment	To collect ultra-high resolution data for an αLP crystal (up to this point, the highest resolution obtained was collected by me at ALS in 1999, to 1.03Å resolution). Ultra-high resolution data would allow us to pursue multiple-conformer refinement.
Final State of Structure	Published *JMB* **338**: 999-1013 (2004)
Crystal Preparation	
Protein purified	November 10, 1999
Date of tray set-up	November 22, 1999
Precipitant solution	1.3M Li_2SO_4, 20mM Tris at pH 8
Final crystal size (mm)	~0.47 x 0.47 x 1.3
Data Collection	
Beamline (detector)	SSRL 9-1 (MAR 345)
Cryoprotectant	15% glycerol (in precipitant solution)
Notes	Data collected in 2 passes, first pass a high-resolution set (1.7 – 0.78Å) and second pass low-resolution data (∞ - 1.12Å).
Data Statistics / Refinement	
Limiting resolution	0.83
R / R_{free}	8.57 / 9.91
Refinement Program	CNS / SHELXL-97
Directory	~cynthia/ULTRA_HIGH/WT_analysis/shelx
Conclusions	See Chapter 1.

Structure	**WT αLP at pH 5.1**
Date of data collection	December 5, 2003
Goal for Experiment	Determine protonation state of His57 at low pH, which mimics the transition state. This will provide a comparison to the 0.78Å subtilisin structure published by Kuhn, *et al.*, which showed that the His57-Asp102 formed a short strong hydrogen bond at pH 5.9.
Final State of Structure	*To be published – see Chapter 3.*
Crystal Preparation	*Crystals grown by rotation student Matthew Daugherty.*
Protein purified	Winter Quarter 2003 *(Expressed and purified by Caleb Bashor)*
Date of tray set-up	April 22, 2003 *(Tray set up by Matt Daugherty)*
Precipitant solution	1.3M Li_2SO_4, 20mM Tris at pH 4.25
Final crystal size (mm)	~0.2 x 0.2 x 0.4
Data Collection	
Beamline (detector)	ALS HHMI 8.2.1 (CCD ADSC q210)
Cryoprotectant	15% glycerol (in precipitant solution), pH 5.1
Data Collection Notes	Data was collected in four passes: high resolution (to 0.86Å, collected over 2 positions on the crystal), medium resolution (to 1.15Å), low resolution (to 1.24Å), and finally ultra-high resolution (to 0.78Å, with the detector off-set along 2-theta). Data collection recorded in brown notebook entitled "αLP inhibitors of different lengths."
Data Statistics / Refinement	
Limiting resolution	0.82 Å
R / R_{free} (%)	8.1 / 9.3
Refinement Program	SHELXL-97
Conclusions	See Chapter 3.

*On Dec. 5, 2003, data was also collected on another low-pH crystal. This data is labeled with prefix "αLP_lowpH1__.img".

Structure	**αLP, mutation Met190(138) → Ala ("MA190")**
Date of collection	May 17, 2002
Goal for Experiment	Compare dynamics in WT αLP, MA190, inhibitor-bound WT αLP and inhibitor-bound MA190. Inspired by work by J. Mace and S. Rader.
Final State of Structure	*Unpublished.*
Crystal Preparation	
Protein purified	November 11, 1999
Date of tray set-up	December 13, 1999 (Drop from tray #2-3 (Nov. 20, 1999) was re-equilibrated over 50% precipitant/50% dH_2O for 3 days prior to transfer to tray #23-2 for re-seeding.)
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Final crystal size (mm)	~0.3 x 0.3 x 0.5
Data Collection	
Beamline (detector)	SSRL 9-1 (MAR 345)
Cryoprotectant	15% glycerol (in precipitant solution), pH 8
Notes	Collected data in two passes. First, high resolution data (120 second exposures in dose mode, with beam at 60 mA; 120 frames. Moved crystal after frame 67). Low resolution data collected at position on crystal at midpoint between both high-res data sets. 9 sec exposures for 60 frames in dose mode. Collection notes in blue "Crystallography Data Collection" notebook.
Data Statistics / Refinement	
Limiting resolution	0.86 Å
R / R_{free} (%)	12.9 / 14.6
Refinement Program	SHELXL-97
Location of files	~cynthia/ULTRA_HIGH/MA190_analysis2/shelx
Conclusions	Structure is well worth pursuing!

* Data for MA190 also previously collected on January 8, 2000 by CNF. Resolution limit 0.79 Å. Data collected on crystal #1-3(D6), of approximate dimensions 0.33 x 0.33 x 1.0 mm. Protein purified by HPLC on November 11, 1999 and crystal trays set up on November 21, 1999.

Structure	**αLP, mutation Met190(138) → Ala inhibited by MeOSuc-Ala-Ala-Pro-boroAla (with glycerol bound) ("MA190+boroAla...glycerol")**
Date of collection	January 9, 2000
Goal for Experiment	Compare dynamics in WT αLP, MA190, inhibitor-bound WT αLP and inhibitor-bound MA190. Inspired by work by J. Mace and S. Rader.
Final State of Structure	*Not published.*
Crystal Preparation	
Protein purified	November 11, 1999
Date of tray set-up	
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Final crystal size (mm)	
Data Collection	
Beamline (detector)	SSRL 9-1 (MAR 345)
Cryoprotectant	15% glycerol (in precipitant solution), pH 8
Notes	Data collected in two passes: a high-resolution pass, then a low-resolution pass. *Data processed by CNF and Krista Shipley.*
Data Statistics / Refinement	*Early stages of refinement by Dr. C. Waddling.*
Limiting resolution	0.90 Å
R / R_{free} (%)	9.8 / 12.0
Refinement Program	CNS / SHELXL-97
Final directory	~cynthia/ULTRA_HIGH/ma190_ala_excerpts/ Revisited_refinement2
Conclusions	Structure solved to near-completion. However, not included in publication due to difficulty of modeling the glycerol in the P1' pocket. Nε2-H not visible. His57-Asp102 not a short hydrogen-bond (2.74Å and H located on Nε2).

Structure	**αLP inhibited by MeOSuc-Ala-Ala-Pro-boroAla (with glycerol bound) ("WT+boroAla...glycerol")**
Date of collection	January 9, 2000
Goal for Experiment	Compare dynamics in WT αLP, MA190, inhibitor-bound WT αLP and inhibitor-bound MA190. Inspired by work by J. Mace and S. Rader.
Final State of Structure	*Not published.*
Crystal Preparation	
Protein purified	November 10, 1999
Date of tray set-up	December 12, 1999 (Drop from tray #3-1 (Nov. 21, 1999) was re-equilibrated over 50% precipitant/50% dH_2O for 2 days prior to transfer to tray #22-1 for re-seeding.)
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Final crystal size (mm)	?
Data Collection	
Beamline (detector)	SSRL 9-1 (MAR 345)
Cryoprotectant	15% glycerol (in precipitant solution), pH 8
Notes	Data collected in two passes: one high-resolution pass, (2.8 to 0.88Å) and one low-resolution pass (to 1.7Å).
Data Statistics / Refinement	
Limiting resolution	0.89
R / R_{free}	9.5 / 11.5
Refinement Program	CNS / SHELXL-97
Conclusions	Inhibitor has a major second conformation along its length (80/20% occupancy). Difference maps suggest that this conformation may also extend to the Ser-BO adduct, and to the bound glycerol. Regions of multiple conformers through backbone include 161-164 (separation of 1.2Å). Hydrogen not visible off Nε2. His57-Asp102 not a short hydrogen-bond (2.76Å and H located ~1.02Å from Nε2).

Structure	αLP + inhibited by MeOSuc-Ala-Ala-Pro-boroGly
Date of collection	February 18, 2002
Goal for Experiment	Explore structural basis of αLP's low catalytic activity (due to low k_{cat}) for substrates with Gly at the P1 position.
Final State of Structure	*Not published.*
Crystal Preparation	
Protein purified	
Date of tray set-up	
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Final crystal size (mm)	
Data Collection	
Beamline (detector)	SSRL 9-1
Cryoprotectant	glycerol
Notes	
Data Statistics / Refinement	Data processing begun by Matthew Daugherty
Limiting resolution	0.85 or 0.86 Å
R / R_{free} (%)	-
Refinement Program	-
Location of files	~cynthia/DATA_COLLECTION/wt_inh_ssrl/wt_gly ~mdaugher/WT_GLY
Conclusions	No maps made yet, but data is strong. Should be continued! (Likely will have radiation damage; see Chapter 2.)

Structure	**αLP + inhibited by MeOSuc-Ala-Ala-Pro-boroVal (in PEG cryosolvent)**
Date of collection	March 5, 2004
Goal for Experiment	Obtain a structural mimic of TI_2 (with no glycerol in the S1' pocket) Hoped that using PEG-300 as a cryoprotectant may also prevent multiple conformers observed at boroValP1-Ser195... but this was not the case.
Final State of Structure	See Chapter 2
Crystal Preparation	
Protein purified	December 3, 2003 (expression & first step of purification in June 2001; protein stored at 4°C at ~pH 5)
Date of tray set-up	December 10, 2003
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Final crystal size (mm)	0.3 x 0.3 x 0.7 (estimate)
Data Collection	
Beamline (detector)	ALS HHMI 8.2.2 (CCD ADSC q315)
Cryoprotectant	20% PEG-300
Notes	Low-resolution data collected before and after high-resoltuion data sets. High resolution data collected at 2 or 3 positions on the crystal. The last data set collected was of medium resolution (1.1Å).
Data Statistics / Refinement	
Limiting resolution	0.89 Å
R / R_{free} (%)	12.5 / 14.8 (preliminary refinement only)
Refinement Program	SHELXL-97
Location of files	~cynthia/INH-LENGTH/wt_inh4_peg/shelx
Conclusions	Initial refinement was performed on the first low-resolution data set collected (shelx_low3). This would provide a model without influence from radiation damage. Refinement was then continued against all data (high resolution merged with low resolution data). Refinement stopped due to multiple conformations at Ser195-boroValP1 (likely due to radiation damage, see Chapter 2). The last data set collected (1.1Å resolution) was used to refine a "post-damage" model, "αLP+boroVal*$_{11}$". See Chapter 2.

Structure	**αLP inhibited by MeOSuc-Ala-Ala-Pro-boroVal (in PEG cryosolvent)**
Date of collection	December 5, 2004
Goal for Experiment	Structure of TI_2 (no glycerol bound in S1' pocket), at high resolution for an accurate description of catalytic triad geometries and hydrogen atom location. Structure without radiation damage: Obtain data that could be extrapolated to zero-dose radiation using the method of Diederichs (2003).
Final State of Structure	See Chapter 2
Crystal Preparation	
Protein purified	December 3, 2003
Date of tray set-up	December 10, 2003
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Inhibitor soak	1μL of inhibitor added on September 30, 2004
Final crystal size (mm)	~ 0.3 x 0.5
Data Collection	
Beamline (detector)	ALS HHMI 8.2.2 (CCD ADSC q315)
Cryoprotectant	20% PEG (soaked for 20 seconds)
Notes	Collected data to resolution ~1Å, allowing minimal dose yet resolution high enough that H's could be visible. Collected 275° of data (in 344 frames) at one position on the crystal. (High redundancy is required for extrapolation method.) Beam centered in middle of crystal width, so most unit cells were continually exposed. Collected additional data, but this was not included in analysis.
Data Statistics / Refinement	Extrapolated to zero-dose by Kay Diederichs, Jan. 25 2005 Refinement performed by CNF
Limiting resolution	1.04Å
R / R_{free} (%)	11.0 / 13.8 (structure not completed; refinement converged at this stage; had not followed up to completion at time of Thesis publication)
Refinement Program	SHELXL-97
Location of files	~cynthia/INH-LENGTH/inh4peg_redundant (refinement in shelx-KD-zero)
Conclusions	F_o-F_o maps indicate less damage in extrapolated data than in the original data set. F_o-F_c difference maps in refined

	model continue to show no indication of radiation damage on Cys residues. Structure is indeed a mimic of TI_2: boronate hydrogen-bonds to three waters in the S1' pocket. Hydrogen atoms are weak, and difficult to interpret. F_o-F_c density at similar contour level near His57 may suggest motion of the residue? Unclear: further analysis needed. Did not notice significant differences between this TI_2 mimic and aLP+boroVal TI_1 structure (see Chapter 2), but further analysis should be performed.

Structure	**αLP + inhibited by Ac-Pro-boroVal**
Date of collection	February 18, 2002?
Goal for Experiment	Compare dynamics in the active sites of a series of structures of αLP bound to boronic acid peptides of different lengths. (k_{cat} effect)
Final State of Structure	*Unpublished.*
Crystal Preparation	
Protein purified	
Date of tray set-up	
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Final crystal size (mm)	
Data Collection	
Beamline (detector)	SSRL 9-1?
Cryoprotectant	glycerol
Notes	Crystal was whisked through glycerol, but no glycerol molecule visible in the S1' pocket in initial electron density maps.
Data Statistics / Refinement	
Limiting resolution	0.88 Å
R / R_{free} (%)	12.6 / 14.2 (not complete!)
Refinement Program	SHELXL-97
Location of files	~cynthia/ULTRA_HIGH/wt_inh2_analysis/shelx
Conclusions	Refinement stopped due to multiple conformations at Ser195-boroValP1 (likely due to radiation damage, see Chapter 2).

Structure	**αLP + inhibited by Ac-Ala-Pro-boroVal**
Date of collection	February 18, 2002?
Goal for Experiment	Compare dynamics in the active sites of a series of structures of αLP bound to boronic acid peptides of different lengths. (k_{cat} effect)
Final State of Structure	*Unpublished.*
Crystal Preparation	
Protein purified	
Date of tray set-up	
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Final crystal size (mm)	
Data Collection	
Beamline (detector)	SSRL 9-1?
Cryoprotectant	glycerol
Notes	Crystal was whisked through glycerol, but no glycerol molecule visible in the S1' pocket in initial electron density maps.
Data Statistics / Refinement	
Limiting resolution	0.88 Å
R / R_{free} (%)	15.0 / 17.0 (preliminary refinement)
Refinement Program	SHELXL-97
Location of files	~cynthia/ULTRA_HIGH/wt_inh3_analysis/shelx
Conclusions	Refinement stopped due to multiple conformations at Ser195-boroValP1 (likely due to radiation damage, see Chapter 2).

Structure	**αLP + inhibited by MeOSuc-Ala-Ala-Pro-boroVal**
Date of collection	February 18, 2002?
Goal for Experiment	Compare dynamics in the active sites of a series of structures of αLP bound to boronic acid peptides of different lengths. (k_{cat} effect)
Final State of Structure	*Unpublished.*
Crystal Preparation	
Protein purified	
Date of tray set-up	
Precipitant solution	1.3 M Li_2SO_4, 20 mM Tris at pH 8
Final crystal size (mm)	
Data Collection	
Beamline (detector)	SSRL 9-1?
Cryoprotectant	glycerol
Notes	Crystal was whisked through glycerol, but no glycerol molecule visible in the S1' pocket in initial electron density maps.
Data Statistics / Refinement	
Limiting resolution	0.86 Å
R / R_{free} (%)	11.6 / 13.0 (not complete!)
Refinement Program	SHELXL-97
Location of files	~cynthia/ULTRA_HIGH/wt_inh4_analysis/shelx
Conclusions	Refinement stopped due to multiple conformations at Ser195-boroValP1 (likely due to radiation damage, see Chapter 2).

References

QUANTA 2000. Accelrys Inc.

Allen, F. H. (2002). The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr B* 58, 380-8. CSD accession numbers JIQHUM, JIQJEY, VUSTUY, HEMBOQ.

Ash, E. L., Sudmeier, J. L., De Fabo, E. C. & Bachovchin, W. W. (1997). A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment. *Science* 278, 1128-32.

Ash, E. L., Sudmeier, J. L., Day, R. M., Vincent, M., Torchilin, E. V., Haddad, K. C., Bradshaw, E. M., Sanford, D. G. & Bachovchin, W. W. (2000). Unusual 1H NMR chemical shifts support (His) C-epsilon 1...O==C H-bond: Proposal for reaction-driven ring flip mechanism in serine protease catalysis. *Proc Natl Acad Sci U S A* 97, 10371-6.

Bachovchin, W. W. (1985). Confirmation of the assignment of the low-field proton resonance of serine proteases by using specifically nitrogen-15 labeled enzyme. *Proc Natl Acad Sci U S A* 82, 7948-51.

Bachovchin, W. W. (1986). 15N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: evidence for a moving histidine mechanism. *Biochemistry* 25, 7751-9.

Bachovchin, W. W., Wong, W. Y., Farr-Jones, S., Shenvi, A. B. & Kettner, C. A. (1988). Nitrogen-15 NMR spectroscopy of the catalytic-triad histidine of a serine protease in peptide boronic acid inhibitor complexes. *Biochemistry* 27, 7689-97.

Baker, D., Silen, J. L. & Agard, D. A. (1992). Protease pro region required for folding is a potent inhibitor of the mature enzyme. *Proteins* 12, 339-44.

Baker, P. J., Sawa, Y., Shibata, H., Sedelnikova, S. E. & Rice, D. W. (1998). Analysis of the structure and substrate binding of Phormidium lapideum alanine dehydrogenase. *Nat Struct Biol* 5, 561-7.

Betzel, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S. & Singh, T. P. (2001). Structure of a serine protease proteinase K from Tritirachium album limber at 0.98 A resolution. *Biochemistry* 40, 3080-8.

Blow, D. M., Birktoft, J. J. & Hartley, B. S. (1969). Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* 221, 337-40.

3ode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G. & Bartunik, H. (1983). Refined 2 A X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase. Crystallization, structure determination, crystallographic refinement, structure and its comparison with bovine trypsin. *J Mol Biol* 164, 237-82.

Bone, R., Shenvi, A. B., Kettner, C. A. & Agard, D. A. (1987). Serine protease mechanism: structure of an inhibitory complex of alpha- lytic protease and a tightly bound peptide boronic acid. *Biochemistry* 26, 7609-14.

Bone, R., Frank, D., Kettner, C. A. & Agard, D. A. (1989a). Structural analysis of specificity: alpha-lytic protease complexes with analogues of reaction intermediates. *Biochemistry* 28, 7600-9.

Bone, R., Silen, J. L. & Agard, D. A. (1989b). Structural plasticity broadens the specificity of an engineered protease. *Nature* 339, 191-5.

Bone, R., Fujishige, A., Kettner, C. A. & Agard, D. A. (1991a). Structural basis for broad specificity in alpha-lytic protease mutants. *Biochemistry* 30, 10388-98.

Bone, R., Sampson, N. S., Bartlett, P. A. & Agard, D. A. (1991b). Crystal structures of alpha-lytic protease complexes with irreversibly bound phosphonate esters. *Biochemistry* 30, 2263-72.

Brayer, G. D., Delbaere, L. T. & James, M. N. (1979). Molecular structure of the alpha-lytic protease from Myxobacter 495 at 2.8 Angstroms resolution. *J Mol Biol* 131, 743-75.

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

Brünger, A. T. & Karplus, M. (1988). Polar hydrogen positions in proteins: Empirical energy function placement and neutron diffraction comparison. *Proteins* 4, 148-156.

Brünger, A. T. (1992a). *X-PLOR Version 3.1. A system for x-ray crystallography and NMR*, Yale University Press, New Haven and London.

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

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* D54 (Pt 5), 905-21.

no, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr B* 58, 389-97.

meister, W. P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr D Biol Crystallogr* 56 (Pt 3), 328-41.

ugo, O. & Carugo, K. D. (2005). When X-rays modify the protein structure: radiation damage at work. *Trends Biochem Sci* 30, 213-9.

e, A. & Stein, R. L. (2003). Mechanistic origins of the substrate selectivity of serine proteases. *Biochemistry* 42, 3335-48.

land, W. W. & Kreevoy, M. M. (1994). Low-barrier hydrogen bonds and enzymic catalysis. *Science* 264, 1887-90.

land, W. W., Frey, P. A. & Gerlt, J. A. (1998). The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* 273, 25529-32.

ningham, E. L., Jaswal, S. S., Sohl, J. L. & Agard, D. A. (1999). Kinetic stability as a mechanism for protease longevity. *Proc Natl Acad Sci U S A* 96, 11008-14.

ningham, E. L. & Agard, D. A. (2004). Disabling the folding catalyst is the last critical step in alpha-lytic protease folding. *Protein Sci* 13, 325-31.

gett, V., Schroder, S. & Kollman, P. A. (1991). Catalytic pathway of serine proteases: Classical and quantum mechanical calculations. *J Am Chem Soc* 113, 8926-8935.

ano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.

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

man, A. I. & Agard, D. A. (2000). Two energetically disparate folding pathways of alpha-lytic protease share a single transition state. *Nat Struct Biol* 7, 394-7.

derichs, K., McSweeney, S. & Ravelli, R. B. (2003). Zero-dose extrapolation as part of macromolecular synchrotron data reduction. *Acta Crystallogr D Biol Crystallogr* 59, 903-9.

nth, J. (1999). *Principles of protein x-ray crystallography*. 2nd edition edit, Springer-Verlag, Groningen.

r-Jones, S. (1989). Serine protease mechanism studied by 15N-NMR of the active-site histidine of alpha-lytic protease. Graduate Thesis, Tufts University.

Fox, G. C. & Holmes, K. C. (1966). An alternative method of solving the layer scaling equation of Hamilton, Rollet and Sparks. *Acta Cryst* 20, 886-891.

Frey, P. A., Whitt, S. A. & Tobin, J. B. (1994). A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* 264, 1927-30.

Frey, P. A. (2004). Strong hydrogen bonding in chymotrypsin and other serine proteases. *J Phys Org Chem* 17, 511-520.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B. G., Chen, W., Wong, M. W., Andres, J. L., Head-Gordon, M., Replogle, E. S. & Pople, J. A. (1998). Gaussian 98 Revision A.7 edit. Gaussian, Inc., Pittsburgh, PA.

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

Fuhrmann, C. N., Ota, N., Rader, S. D. & Agard, D. A. (2004b). Alpha-lytic protease. In *Handbook of Proteolytic Enzymes* 2 edit. (Barrett, A., Rawlings, N. D. & Woessner, J. F., eds.), Vol. 2. 2 vols. Academic Press, Cambridge.

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

GarciaViloca, M., Gelabert, R., GonzalezLafont, A., Moreno, M. & Lluch, J. M. (1997a). Is an extremely low-field proton signal in the NMR spectrum conclusive evidence for a low-barrier hydrogen bond? *J Phys Chem A* 101, 8727-8733.

GarciaViloca, M., GonzalezLafont, A. & Lluch, J. M. (1997b). On pK(a) matching as a requirement to form a low-barrier hydrogen bond. A theoretical study in gas phase. *J Phys Chem A* 101, 3880-3886.

Guthrie, J. P. (1996). Short strong hydrogen bonds: can they explain enzymic catalysis? *Chem Biol* 3, 163-70.

Haddad, K. C., Sudmeier, J. L., Bachovchin, D. A. & Bachovchin, W. W. (2005). alpha-lytic protease can exist in two separately stable conformations with different His57 mobilities and catalytic activities. *Proc Natl Acad Sci U S A* 102, 1006-11.

Hammond, G. S. (1955). A correlation of reaction rates. *J Am Chem Soc* 77, 334-8.

Harp, J. M., Hanson, B. L., Timm, D. E. & Bunick, G. J. (1999). Macromolecular crystal annealing: evaluation of techniques and variables. *Acta Crystallogr D Biol Crystallogr* D55, 1329-1334.

Hedstrom, L. (2002a). Introduction: proteases. *Chem Rev* 102, 4429-30.

Hedstrom, L. (2002b). Serine protease mechanism and specificity. *Chem Rev* 102, 4501-24.

Hunkapiller, M. W., Smallcombe, S. H., Whitaker, D. R. & Richards, J. H. (1973). Carbon nuclear magnetic resonance studies of the histidine residue in alpha-lytic protease. Implications for the catalytic mechanism of serine proteases. *Biochemistry* 12, 4732-43.

Ishida, T. & Kato, S. (2003). Theoretical perspectives on the reaction mechanism of serine proteases: the reaction free energy profiles of the acylation process. *J Am Chem Soc* 125, 12035-48.

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

Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. (2002). Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* 415, 343-6.

Jones, T. A., Bergdoll, M. & Kjeldgaard, M. (1990). O: A macromolecular modeling environment. In *Crystallographic and Modeling Methods in Molecular Design* (Bugg, C. & Ealick, S. E., eds.), pp. 189-195. Springer-Verlag Press.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 26, 795-800.

Katona, G., Wilmouth, R. C., Wright, P. A., Berglund, G. I., Hajdu, J., Neutze, R. & Schofield, C. J. (2002). X-ray structure of a serine protease acyl-enzyme complex at 0.95-A resolution. *J Biol Chem* 277, 21962-70.

Katznelson, H., Gillespie, D. C. & Cook, F. D. (1964). Studies on the Relationships between Nematodes and Other Soil Microorganisms. 3. Lytic Action of Soil Myxobacters on Certain Species of Nematodes. *Can J Microbiol* 10, 699-704.

elch, B. A. & Agard, D. A. Structure and folding of a kinetically stable, thermophilic protease. *Manuscript in preparation.*

elch, B. A., Fuhrmann, C. N., Ho, J. & Agard, D. A. Sidechain distortion modulates the energetics of a protein folding landscape. *Manuscript in preparation.*

elch, B. A., Humphries, E., Thomason, A., Mitsuiki, S. & Agard, D. A. Extreme acid resistance in a kinetically stable protease: A folding and structural study. *Manuscript in preparation.*

ettner, C. A. & Shenvi, A. B. (1984). Inhibition of the serine proteases leukocyte elastase, pancreatic elastase, cathepsin G, and chymotrypsin by peptide boronic acids. *J Biol Chem* 259, 15106-14.

ettner, C. A., Bone, R., Agard, D. A. & Bachovchin, W. W. (1988). Kinetic properties of the binding of alpha-lytic protease to peptide boronic acids. *Biochemistry* 27, 7682-8.

im, Y., Lim, S. & Kim, Y. (1999). The role of a short and strong hydrogen bond on the double proton transfer in the formamidine-formic acid complex: Theoretical studies in the gas phase and in solution. *J Phys Chem A* 103, 6632-6637.

leywegt, G. J. & Jones, T. A. (1994). A super position. *CCP4/ESF-EACBM Newsletter on Protein Crystallography* 31, 9-14.

ollman, P. A., Kuhn, B., Donini, O., Perakyla, M., Stanton, R. & Bakowies, D. (2001). Elucidating the nature of enzyme catalysis utilizing a new twist on an old methodology: quantum mechanical-free energy calculations on chemical reactions in enzymes and in aqueous solution. *Acc Chem Res* 34, 72-79.

ossiakoff, A. A. & Spencer, S. A. (1980). Neutron diffraction identifies His 57 as the catalytic base in trypsin. *Nature* 288, 414-6.

Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24, 946-50.

Krem, M. M. & Di Cera, E. (2001). Molecular markers of serine protease evolution. *Embo J* 20, 3036-45.

Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). The 0.78 A structure of a serine protease: Bacillus lentus subtilisin. *Biochemistry* 37, 13446-52.

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

Leiros, H. K., McSweeney, S. M. & Smalas, A. O. (2001). Atomic resolution structures of trypsin provide insight into structural radiation damage. *Acta Crystallogr D Biol Crystallogr* 57, 488-97.

Li, R. & Woodward, C. (1999). The hydrogen exchange core and protein folding. *Protein Sci* 8, 1571-90.

Lin, J., Cassidy, C. S. & Frey, P. A. (1998a). Correlations of the basicity of His 57 with transition state analogue binding, substrate reactivity, and the strength of the low-barrier hydrogen bond in chymotrypsin. *Biochemistry* 37, 11940-8.

Lin, J., Westler, W. M., Cleland, W. W., Markley, J. L. & Frey, P. A. (1998b). Fractionation factors and activation energies for exchange of the low barrier hydrogen bonding proton in peptidyl trifluoromethyl ketone complexes of chymotrypsin. *Proc Natl Acad Sci U S A* 95, 14664-8.

Mace, J. E. & Agard, D. A. (1995). Kinetic and structural characterization of mutations of glycine 216 in alpha-lytic protease: a new target for engineering substrate specificity. *J Mol Biol* 254, 720-36.

Mace, J. E., Wilk, B. J. & Agard, D. A. (1995). Functional linkage between the active site of alpha-lytic protease and distant regions of structure: scanning alanine mutagenesis of a surface loop affects activity and substrate specificity. *J Mol Biol* 251, 116-34.

Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol* 33, 491-7.

Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, T. & Kraut, J. (1977). Re-examination of the charge relay system in subtilisin comparison with other serine proteases. *J Biol Chem* 252, 8875-83.

Merritt, E. A. & Bacon, D. J. (1997). Raster3D photorealistic molecular graphics. *Methods Enzymol* 277, 505-24.

Miller, D. W. & Agard, D. A. (1999). Enzyme specificity under dynamic control: a normal mode analysis of alpha-lytic protease. *J Mol Biol* 286, 267-78.

Minasov, G., Wang, X. & Shoichet, B. K. (2002). An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. *J Am Chem Soc* 124, 5333-40.

Nave, C. & Garman, E. F. (2005). Towards an understanding of radiation damage in cryocooled macromolecular crystals. *J Synchrotron Radiat* 12, 257-60.

Neidhart, D., Wei, Y., Cassidy, C., Lin, J., Cleland, W. W. & Frey, P. A. (2001). Correlation of low-barrier hydrogen bonding and oxyanion binding in transition state analogue complexes of chymotrypsin. *Biochemistry* 40, 2439-47.

Nienaber, V. L., Breddam, K. & Birktoft, J. J. (1993). A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry* 32, 11469-75.

O'Neill, P., Stevens, D. L. & Garman, E. F. (2002). Physical and chemical considerations of damage induced in protein crystals by synchrotron radiation: a radiation chemical perspective. *J Synchrotron Radiat* 9, 329-32.

Ota, N. & Agard, D. A. (2001). Enzyme specificity under dynamic control II: Principal component analysis of alpha-lytic protease using global and local solvent boundary conditions. *Protein Sci* 10, 1403-14.

Otwinowski, Z. & Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. In *Methods Enzymol*, Vol. 276, pp. 307-326. Academic Press, Boston.

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

Peters, R. J., Shiau, A. K., Sohl, J. L., Anderson, D. E., Tang, G., Silen, J. L. & Agard, D. A. (1998). Pro region C-terminus: protease active site interactions are critical in catalyzing the folding of alpha-lytic protease. *Biochemistry* 37, 12058-67.

Polgar, L. (1972). On the role of hydrogen-bonding system in the catalysis by serine proteases. *Acta Biochim Biophys Acad Sci Hung* 7, 29-34.

Rader, S. D. & Agard, D. A. (1997). Conformational substates in enzyme mechanism: the 120 K structure of alpha-lytic protease at 1.5 A resolution. *Protein Sci* 6, 1375-86.

Ravelli, R. B. & McSweeney, S. M. (2000). The 'fingerprint' that X-rays can leave on structures. *Structure Fold Des* 8, 315-28.

Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. (1983). Structure of the complex of Streptomyces griseus protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-A resolution. *Biochemistry* 22, 4420-33.

Robillard, G. & Shulman, R. G. (1974). High resolution nuclear magnetic resonance studies of the active site of chymotrypsin. II. Polarization of histidine 57 by substrate analogues and competitive inhibitors. *J Mol Biol* 86, 541-58.

Rypniewski, W. R., Ostergaard, P. R., Norregaard-Madsen, M., Dauter, M. & Wilson, K. S. (2001). Fusarium oxysporum trypsin at atomic resolution at 100 and 283 K: a study of ligand binding. *Acta Crystallogr D Biol Crystallogr* 57, 8-19.

Satterthwait, A. C. & Jencks, W. P. (1974). The mechanism of the aminolysis of acetate esters. *J Am Chem Soc* 96, 7018-31.

Sauter, N. K., Mau, T., Rader, S. D. & Agard, D. A. (1998). Structure of alpha-lytic protease complexed with its pro region. *Nat Struct Biol* 5, 945-50.

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

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

Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). Trypsin revisited: crystallography AT (SUB) atomic resolution and quantum chemistry revealing details of catalysis. *J Biol Chem* 278, 43357-62.

Schroder Leiros, H. K., McSweeney, S. M. & Smalas, A. O. (2001). Atomic resolution structures of trypsin provide insight into structural radiation damage. *Acta Crystallogr D Biol Crystallogr* 57, 488-97.

Schutz, C. N. & Warshel, A. (2004). The low barrier hydrogen bond (LBHB) proposal revisited: the case of the Asp... His pair in serine proteases. *Proteins* 55, 711-23.

Shan, S. O. & Herschlag, D. (1996). The change in hydrogen bond strength accompanying charge rearrangement: Implications for enzymatic catalysis. *Proc Natl Acad Sci U S A* 93, 14474-9.

Shan, S. O., Loh, S. & Herschlag, D. (1996). The energetics of hydrogen bonds in model systems: Implications for enzymatic catalysis. *Science* 272, 97-101.

Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: High-resolution refinement. In *Methods Enzymol* (C. W. Carter, J. & Sweet, R. M., eds.), Vol. 277, pp. 319-343. Academic Press, Boston.

Shinde, U. & Inouye, M. (1995). *Intramolecular chaperones and protein folding*. Molecular Biology Intelligence Unit, R.G. Landes Company, Austin.

Shokhen, M. & Albeck, A. (2004). Is there a weak H-bond --> LBHB transition on tetrahedral complex formation in serine proteases? *Proteins* 54, 468-77.

Silen, J. L. & Agard, D. A. (1989). The alpha-lytic protease pro-region does not require a physical linkage to activate the protease domain in vivo. *Nature* 341, 462-4.

Silen, J. L., Frank, D., Fujishige, A., Bone, R. & Agard, D. A. (1989). Analysis of prepro-alpha-lytic protease expression in Escherichia coli reveals that the pro region is required for activity. *J Bacteriol* 171, 1320-5.

Sliz, P., Harrison, S. C. & Rosenbaum, G. (2003). How does radiation damage in protein crystals depend on X-ray dose? *Structure (Camb)* 11, 13-9.

Smith, S. O., Farr-Jones, S., Griffin, R. G. & Bachovchin, W. W. (1989). Crystal versus solution structures of enzymes: NMR spectroscopy of a crystalline serine protease. *Science* 244, 961-4.

Sohl, J. L., Jaswal, S. S. & Agard, D. A. (1998). Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* 395, 817-9.

Steitz, T. A. & Shulman, R. G. (1982). Crystallographic and NMR studies of the serine proteases. *Annu Rev Biophys Bioeng* 11, 419-44.

Stratton, J. R., Pelton, J. G. & Kirsch, J. F. (2001). A novel engineered subtilisin BPN' lacking a low-barrier hydrogen bond in the catalytic triad. *Biochemistry* 40, 10411-6.

Teeter, M. M., Yamano, A., Stec, B. & Mohanty, U. (2001). On the nature of a glassy state of matter in a hydrated protein: Relation to protein function. *Proc Natl Acad Sci U S A* 98, 11242-7.

Teng, T. Y. & Moffat, K. (2002). Radiation damage of protein crystals at cryogenic temperatures between 40 K and 150 K. *J Synchrotron Radiat* 9, 198-201.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-80.

Thompson, R. C. & Blout, E. R. (1973). Dependence of the kinetic parameters for elastase-catalyzed amide hydrolysis on the length of peptide substrates. *Biochemistry* 12, 57-65.

Truhlar, S. M., Cunningham, E. L. & Agard, D. A. (2004). The folding landscape of Streptomyces griseus protease B reveals the energetic costs and benefits associated with evolving kinetic stability. *Protein Sci* 13, 381-90.

Tsukada, H. & Blow, D. M. (1985). Structure of alpha-chymotrypsin refined at 1.68 A resolution. *J Mol Biol* 184, 703-11.

Wang, J. H. (1970). Directional character of proton transfer in enzyme catalysis. *Proc Natl Acad Sci U S A* 66, 874-81.

Warshel, A. & Papazyan, A. (1996). Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds. *Proc Natl Acad Sci U S A* 93, 13665-70.

arshel, A. (1998). Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem* 273, 27035-8.

'eik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc Natl Acad Sci U S A* 97, 623-8.

/eik, M., Berges, J., Raves, M. L., Gros, P., McSweeney, S., Silman, I., Sussman, J. L., Houee-Levin, C. & Ravelli, R. B. (2002). Evidence for the formation of disulfide radicals in protein crystals upon X-ray irradiation. *J Synchrotron Radiat* 9, 342-6.

Vestler, W. M., Frey, P. A., Lin, J., Wemmer, D. E., Morimoto, H., Williams, P. G. & Markley, J. L. (2002). Evidence for a strong hydrogen bond in the catalytic dyad of transition-state analogue inhibitor complexes of chymotrypsin from proton-triton NMR isotope shifts. *J Am Chem Soc* 124, 4196-7.

Vhitaker, D. R., Roy, C., Tsai, C. S. & Jurasek, L. (1965). Lytic enzymes of Sorangium sp. A comparison of the proteolytic properties of the alpha- and beta-lytic proteases. *Can J Biochem* 43, 1961-70.

Wilmouth, R. C., Edman, K., Neutze, R., Wright, P. A., Clifton, I. J., Schneider, T. R., Schofield, C. J. & Hajdu, J. (2001). X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. *Nat Struct Biol* 8, 689-94.

Wurtele, M., Hahn, M., Hilpert, K. & Hohne, W. (2000). Atomic resolution structure of native porcine pancreatic elastase at 1.1 A. *Acta Crystallogr D Biol Crystallogr* 56 (Pt 4), 520-3.

Yeh, J. I. & Hol, W. G. (1998). A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. *Acta Crystallogr D Biol Crystallogr* 54 (Pt 3), 479-80.