Engineered Trypsin Specificity

by

Walter Scott Willett Jr.

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

Date **University Librarian**

Degree Conferred: . . . . . .

For my family:

To my parents Walter and Janet,

and to my sisters Barbara and Patricia.

Thank you for your constant love and support.

You are all so very important to me.

And especially for Cynthia:

Your eternal faith and love are my inspiration.

Now we can begin our forever together.

Preface

I came to UCSF with a very simple idea in mind: to learn enough molecular biology to allow me to do interesting experiments with proteins. I didn't really care which project I worked on, but I did know that I wanted to be in Charly Craik's lab. Maybe it was his smiling picture in the UCSF Pharmaceutical Chemistry bulletin, or that someone at work had said, "Yeah, Charly's a pretty good guy". At any rate, I will never forget the day I met him in his office to discuss doing a rotation in his lab. That was October 17, 1989, the day of the big quake, and the middle of the World Series of Baseball. Charly and I were sitting in his "phone booth" office discussing something or other about some protease when the building started dancing. At first we didn't do anything, but then Charly said, "I think this is a big one, we'd better get out of here." We had just stepped out of his office when his bookshelves collapsed, raining a small refrigerator and several hundred pounds of books down upon the place we were just sitting. That's not what I would call an overly optimistic beginning, but certainly exciting enough. He ended up inviting me to do a rotation with Dr. Jeff Higaki in his lab, and I just never left.

Six years and a marriage later, it's time to leave. Already I've got the warm fuzzys about my experience here, but more importantly, I have a tremendous sense of gratification. I've always maintained that you don't have to be intelligent to get your doctorate, you just need to be very determined. I guess this document proves it. Of course, a place like UCSF makes doing science easy, and I owe a debt of gratitude to my collegues for teaching me so much, and to Charly for understanding and allowing me to

work in my own way. While there were times when I might have wanted more guidance, I know now that learning how to pick a scientific direction and follow through on my own has been one of the most important skills I've gained here in Craik world. Thank you Charly. So here I am, all filled with confidence and enthusiasm for scientific research. I don't think I could have asked for more from graduate school.

On a more personal note, I feel incredibly blessed by the people who've been part of my life for the last six years. I'll take with me to Boulder enough fond memories to last until my memory fails. I've been able to enjoy many productive and enlightening collaborations with a marvelous variety of people. I'd especially like to thank Dr. James P. Quigley for his great humor and compassion. I also need to acknowledge my housemates, the Key Brothers Jeff Parrott, Don Simonetti, Richie Most, Morgan Wright, and Marci Manfre (along with Blue and Christie). Our home has been a sanctuary for me and a great source of stability and comfort. I would also like to thank my family and let them know how much I love and miss them. I am forever grateful to all of you for the life we've shared.

The idea of leaving California for good gives me all sorts of strong feelings. I am leaving for the best of all possible worlds: to be with my wonderful wife, Cindy, so that we may begin our life together. I can't imagine a happier occasion. On the other hand, I've been here in California for nine years, and have experienced so much life and have grown so much (not just around the middle), and have built some truly meaningful friendships with some amazing people. It will be sad for me to leave them. But then, that's the way life is, eh? In the end, I am nearly overwhelmed with joy at the

blessings God has given me in my life. So I will leave California satisfied and filled with hope. What a marvelous way to feel.

Abstract

Engineered Trypsin Specificity

Walter Scott Willett Jr.

Structure-based protein design was used to engineer histidine specificity into the serine protease trypsin. Two methods were used: 1) substrate-assisted catalysis; and 2) metal-dependent substrate recognition. Substrate-assisted catalysis uses a histidine from bound substrate to reconstitute the catalytic triad of the protease. The concept was pioneered with subtilisin BPN', creating a protease specific for H-X or X-H bonds, where X is any large or hydrophobic amino acid (Carter & Wells (1987) *Science* **237**, 394-9). Removal of the catalytic histidine in trypsin (Tn H57A) produces an enzyme that recongnizes only HR, HK, RH, or KH diads as substrates, with catalytic efficiency up to 2% of trypsin. Trypsin H57A possesses specificity distinct from, and more stringent than that of subtilisin H64A. Histidine specificity has also been engineered into trypsin by creating a transition metal binding site that bridges the substrate and enzyme with metal ion, creating an intermolecular metal binding site. Application of simple geometric criteria to a structure-based enzyme-substrate model suggested that construction of trypsin N143H/E151H would enable the enzyme to bind a transition metal ion in a tetrahedal geometry with a P_2' His from substrate as the third ligand. Trypsin N143H/E151H is a metal-dependent protease able to cleave the sequence Y-AH only in the presence of nickel or zinc. As trypsin normally cleaves only after arginine or lysine, this demonstrates that a metal co-substrate can be used to modulate specificity in a designed fashion. A 100-fold

improvement in catalysis was achieved by removal of the negative charge in the S_1 subsite of trypsin H143/151, demonstrating that engineering multiple subsites can have additive effects on specificity. X-ray structural determination of a crystal complex of trypsin H143/151 with ecotin A86H provides the structural context of the observed activity. Four structures were solved by molecular replacement: one without metal, and one each with nickel, copper, or zinc present. Each metal ion binds specifically to the designed site with distinct geometry, suggesting that proteolytic activity correlates with the structure of this engineered substrate binding site.

Table of Contents

List of Tables

List of Figures

Chapter 1

Trypsin as a Model System for Protein Engineering

Introduction

Our understanding of the relationships between protein structure and function today is analogous to our knowledge of organic chemistry 130 years ago. In the 1860's the structural principles of aromatic chemical reactions that are well known and exploited today were first being discovered. At the time Kekulé proposed his structure of benzene in 1865, the atomic theory of matter was considered a good hypothesis yet was not unequivocally accepted as it is today. Kekulé was one of the first chemists to include a 3-dimensional view of atoms and molecules as an integral part of his theory of chemical reactivity, and so inextricably linked the structure of molecules with their observed chemistries (Hein, 1966, ch. 1) . As the new theory gained popularity, 3-dimensional models of molecules became common, and the field of aromatic chemistry was opened wide for exploration. This transformed a struggling chemical industry from alchemical empiricism to the world-wide industry we see today, starting with the explosion of growth in the chemical dye industry (Hein, 1966). In a similar fashion, we know that the 3-dimensional structures of macromolecules determine their functions, yet we don't know enough of the underlying principles that govern that relationship to be able to predict and control those interactions with the level of sophistication that an organic chemist possesses with small molecules.

Formation of unique structures, specific molecular recognition events, and discriminating reactivities are the hallmark of macromolecular interactions. Proteins are of paramount importance to every biological

system, and so exquisite is the relationship between their structure and function that they warrant the current research effort to discover and understand the ways in which they work. Furthermore, the vast numbers of structures and functions that proteins possess make them attractive targets for study because of the learning potential they present to the researcher. If we can understand the principles that govern macromolecular structure and function, we can understand biological systems at a level of complexity not possible before.

The advent of recombinant DNA technology in the 1970's has given us a powerful tool to study proteins at the atomic level by allowing us to clone a DNA sequence corresponding to a protein, manufacture that protein heterologously, and make deliberate mutations in that protein (Watson, et al., 1992). Thus, we are beginning to understand the principles of protein structure and function enough to be able to manipulate macromolecular interactions for specific purposes much the way organic chemists first learned to control reactions among small molecules. Progress at this stage seems mixed. We have learned how to do so much in so little time (it is now only 19 years since the first successful expression of human somatostatin and insulin), and yet are only able to utilize our new knowledge to a limited extent by testing principles and theories on small, simple, isolated systems. Hopefully, the knowledge we are gaining from this reductionist approach can be integrated into a larger understanding of biological systems. Protein engineering provides a path towards increasing our understanding of macromolecular interactions.

Examples of protein engineering applications are numerous, and range from the mundane to the arcane. While there are too many examples to review here, a few cases will illustrate the variety of applications available to the protein engineer. Proteins have been engineered to possess increased stability for industrial use in laundry detergents (Braxton and Wells, 1992, Wells and Estell, 1988) and organic solvents (Arnold, 1993), for improved therapeutic use (Paoni, et al., 1993), and for probes of particular structure-function relationships such as the role of conformational changes in activity (Trigo-Gonzalez, et al., 1992), the nature of allostery (Browner, et al., 1994), or the determinants of hormones binding to their receptors (Lowman, et al., 1991). As an example of progress in this field, the enzymes of a biosynthetic pathway have now been engineered in bacteria to create a strain which can synthesize indigo, the most widely used dye in the world, previously produced only by organic chemical means (Wick, 1995). As seen from this small subset of examples, proteins can be engineered for practical use as commercial products, or for the most academic of pursuits. Both goals can lead to an increased understanding of the fundamental rules governing protein structure-function relationships.

In particular, enzyme-substrate interactions are a useful arena in which to study the complex mixture of forces that determine specificity. These include hydrogen bonds, electrostatic, steric, and hydrophobic effects. For instance, hydrogen bond contributions to specificity have been analyzed by site-directed mutagenesis of tyrosyl t-RNA synthetase (Fersht, et al., 1985). Steric and hydrophobic interactions have been manipulated in the alcohol dehydrogenase substrate binding pocket to make the volume of the pocket larger, resulting in an enzyme capable of catalyzing reactions with much

larger substrates (Murali and Creaser, 1986) A charge reversal in the substrate binding pocket of aspartate aminotransferase created an enzyme selective for Arg and Lys, instead of the usual Asp amino acid (Cronin, et al., 1987). The Arg/Lys selectivity of the serine protease, trypsin, was modified by selective mutagenesis in the primary binding pocket of the enzyme (Craik, et al., 1985). The substrate specificity of subtilisin has been extensively engineered by altering charge-charge interactions (Wells, et al., 1987), as well as steric and hydrophobic interactions (Wells, et al., 1987). Subtilisin has also been engineered to efficiently form various peptide bonds rather than hydrolyze them, creating a peptide ligase useful for synthesizing proteins from peptides (Abrahmsen, et al., 1991). Proteinaceous protease inhibitors have been engineered to specifically inhibit target proteases. For example, the serine protease inhibitor, ecotin, has been engineered through phage-display to specifically inhibit u-PA, which has been implicated in cancer metastasis (Wang, et al., 1995). Also, a serpin has been engineered to be specific for furin and shown to prevent production of infectious measles virus particles *in vivo* (Watanabe, et al., 1995). These examples illustrate that both sides of the enzyme-substrate/inhibitor interaction can be manipulated. In most cases, specificity has been engineered by exploiting the simple concepts of charge-charge, steric, and hydrophobic interactions, then appropriately changing amino acids known to be at the interface between the enzyme and the substrate/inhibitor to effect the desired function.

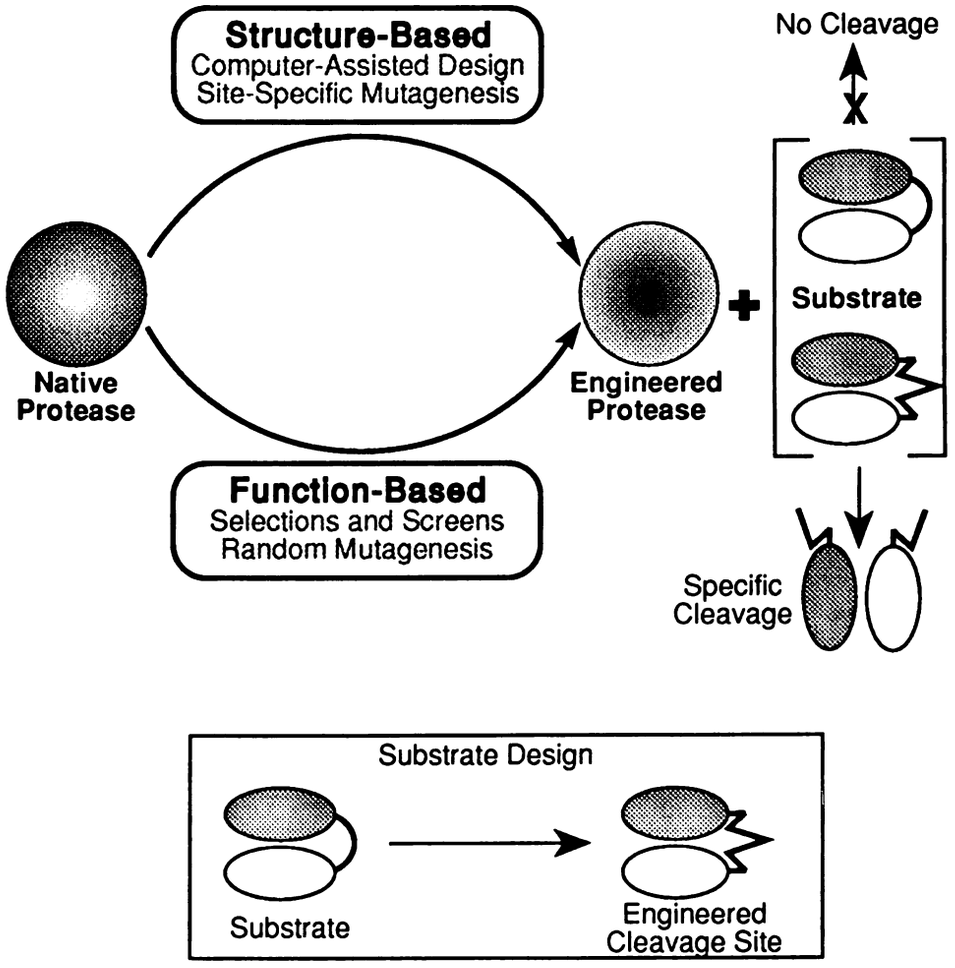
Protease-substrate interactions in particular present a useful platform upon which to test protein engineering hypotheses. Engineering protease activity and specificity is a worthwhile goal because proteases are key elements to every biological system, as well as extremely useful tools for both

analytical and industrial use. Consequently, if we can learn to control protease activity and specificity, we can gain insight into their natural roles and create reagents of practical use. The database of information created by these experiments can then be used to guide future protein engineering endeavors.

Protein engineering experiments can be split into two major types: structure-based, and function-based. In the structure-based approach, knowledge of the protein structure is required, allowing directed mutations to be made to effect the desired function or structure. The function-based approach requires no structural knowledge, rather it relies on genetic selections or functional screens to identify mutants possessing the desired trait. Both engineering methods have their advantages. The structure-based approach allows us to test our understanding of specific macromolecular interactions with targeted mutations, while the function-based approach allows us to screen large libraries of random mutations to achieve the objective, leaving the engineering details to a biological system. The advantages of each method are also a component of their disadvantages: the structure-based approach *requires* a high resolution structure and detailed functional information to succeed, while the function-based approach *requires* a selection or screen designed for a specific function. Figure 1.1 illustrates the two approaches applied to engineering protease specificity. Because our current understanding is limited, an iterative process is required where cleavage by an engineered protease is tested with an appropriate substrate, and design improvements are made with the aid of structural and functional information. In some cases the substrate/inhibitor can also be

engineered to match the specificity of a protease as diagrammed at the bottom of the figure.

Figure 1.1: Engineering Protease Specificity



From an operative point of view, it is helpful to work on a well defined system to perform protein engineering experiments. When an appropriate model system is chosen, there is an existing framework of information to use as a reference point in the interpretation of experimental results. This can be very useful as the results of these experiments tend to create questions as well

as answer them, so in this way we can build upon a foundation of data to advance the field of protein engineering. The serine protease trypsin serves as a good model system because it has been studied so extensively that there are many high-resolution structures of the enzyme in different states, and its mechanism of action is well known. As a result, it lends itself well to both the function-based and structure-based approaches of protein engineering. For example, a function-based genetic selection in bacteria was developed to find functional trypsins in a library of mutations in the primary binding pocket of trypsin that would create altered P_1 specificities (Evnin et al., 1990). This selection relied on detailed knowledge of the mechanism of trypsin in order to relieve an auxotrophy in bacteria thus allowing growth of cells harboring a functional mutant of trypsin. Phage display of trypsin has been accomplished as well in order to screen for altered primary specificities (Corey et al., 1993). Structure-based designs have also targeted the primary binding pocket in order to modulate Lys/Arg specificity (Craik et al., 1985). The catalytic triad of trypsin has mutated to elucidate the roles of specific constituents of the triad (Sprang et al., 1987; Higaki et al., 1989), or to find the minimal requirements for catalysis (Corey & Craik, 1992) and viable alternate geometries for the active-site Asp, His and Ser (Corey et al., 1992). The structure-based approach was also used to design a metal binding site capable of inactivating the enzyme in the presence of copper (Higaki et al., 1990). In all these trypsin engineering projects, knowledge of the structure and/or function of the enzyme was exploited to achieve the desired goal. In this thesis project, we have chosen to use exclusively the structure-based approach in order to engineer alternate specificities into trypsin.

Mechanism and Specificity of Serine Proteases

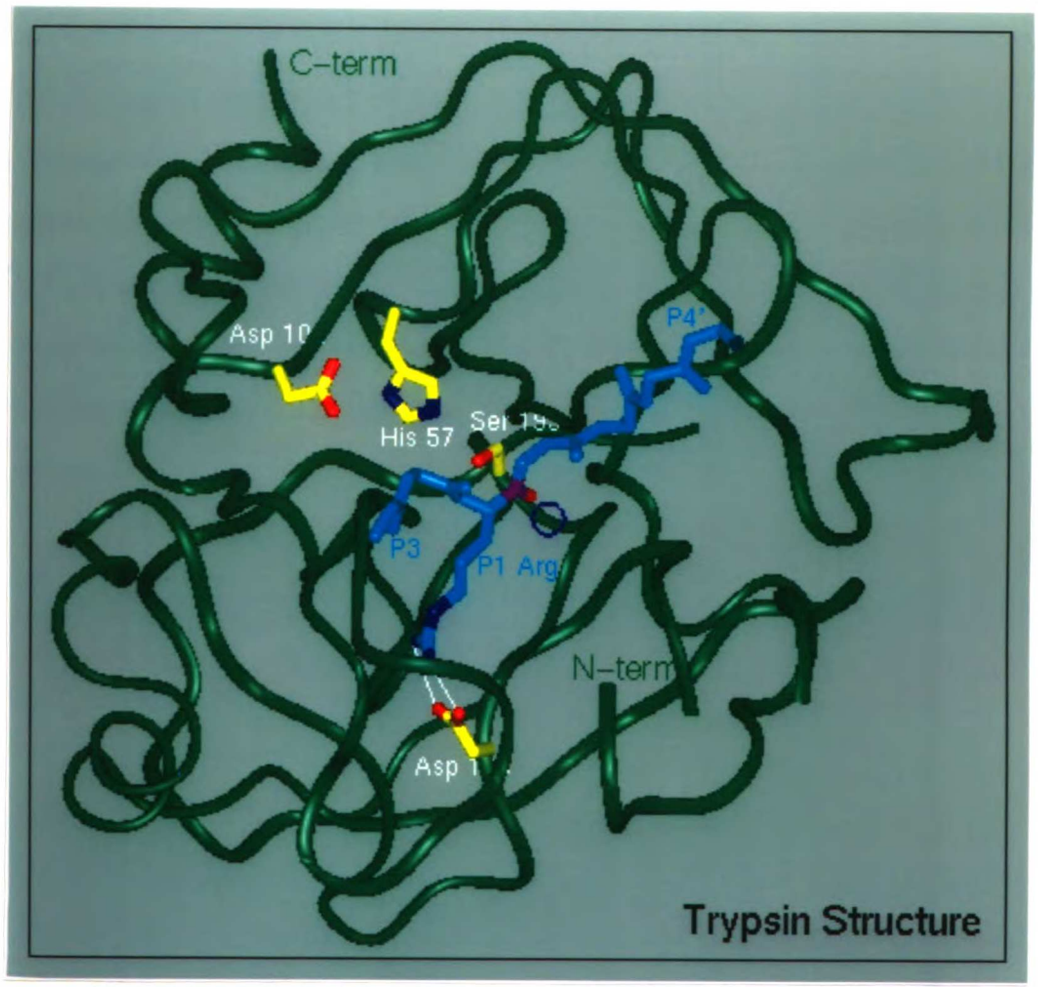
The foundation of these protein engineering experiments rests on our detailed understanding of the chemical mechanism and substrate specificity of serine proteases, in particular the digestive enzyme trypsin. The serine proteases all possess a catalytic triad comprised of the active-site aspartic acid, histidine, and serine, yet they fall into distinct structural classes, making them a classic example of convergent evolution (Liao, et al., 1992, Matthews, 1977). The results of these experiments are interpreted in the context of the activity of trypsin, a paradigm for serine proteases in the chymotrypsin structural family which possess specificity for the basic side chains arginine and lysine.

Structural families of serine proteases. Serine proteases fall into one of three structural groups. Classically, there were two structural motifs, chymotrypsin-like and subtilisin-like. Recently, however, a third structural framework has been revealed in the crystal structure or wheat serine carboxypeptidase II (Liao, et al., 1992, Liao and Remington, 1990). The chymotrypsin family share a double β–barrel motif with the catalytic triad positioned in the active-site cleft between the two domains. The subtilisin family are a single-domain, mixed α-β class, and wheat serine carboxypeptidase II is also a mixed α-β protein yet has a fold distinct from the subtilisin family [see Fig. 1 in (Perona & Craik, 1995)]. Although the three serine protease classes differ in their secondary and tertiary structural elements, they have all evolved a common enzymatic mechanism. The catalytic triads of the substilisin and chymotrypsin families are superimposable to within 0.5Å, and the Ser195 and His57 dyads are superimposble in all three classes, yet the position of Asp102 in wheat serine carboxypeptidase is not conserved relative to the other structural classes. This

illustrates the fact that there are alternate geometries possible for the catalytic triad of serine proteases, and that the three catalytic residues can be considered to be two dyads: Asp-His, and His-Ser (Perona& Craik, 1995, Fig. 1)

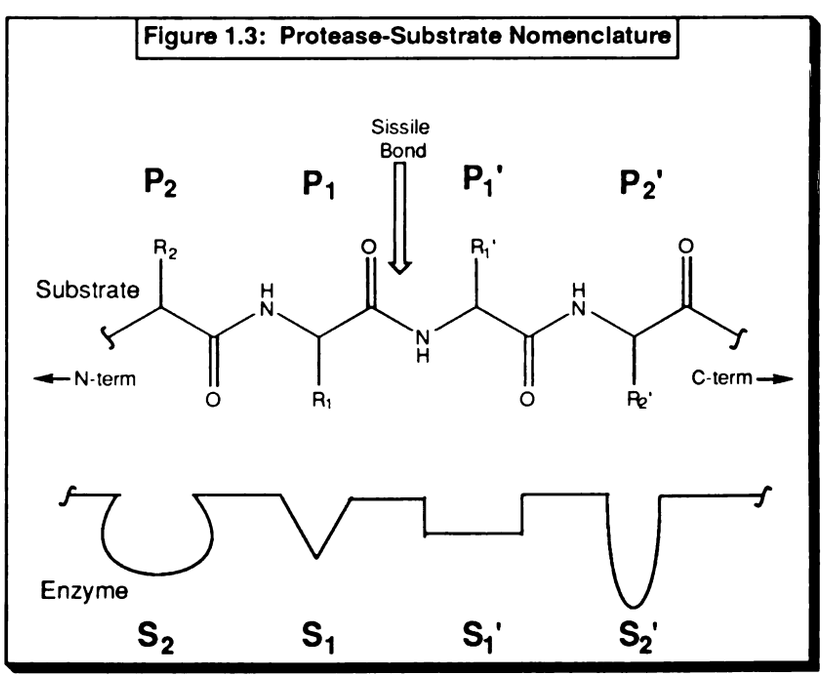
Figure 1.2 illustrates the overall structural fold of trypsin shown modelled with bound substrate (Willett, unpublished structure of rat trypsin H143/151 complexed with residies 82-88 of ecotin A86H) A P_1-Arg has been modelled in place of the P_1-Met84 in ecotin. Residues known to be important for substrate binding and catalysis are shown in yellow, and the N- and C-termini of mature trypsin are indicated in green. The members of the catalytic triad are behind the bound substrate (light blue), with the serine 195 hydroxyl group poised to attack the sissile bond of the P_1 arginine residue. The carbonyl group of the P_1 arginine (shown in purple and red) is pointing towards the oxyanion hole formed by the backbone amide nitrogens of Gly193 and Ser195 (indicated by the blue circle), and the guanidino group of the P_1-Arg forms a salt bridge with the carboxylate of Asp189. In the lower right of the figure there are two breaks in the backbone. These result from residues 113-117 being undefined in the electron density because of autolysis during crytallization (see Chapter 5, p 118).

Figure 1.2: Structure of Trypsin



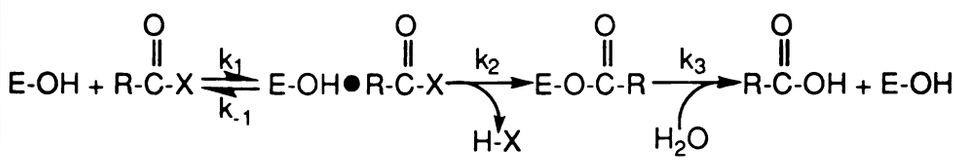
The nomenclature of protease-substrate interactions was defined by Schechter and Berger (Schechter and Berger, 1968), and is illustrated in figure 1.3. The amino acids of the substrate are numbered away from the sissile bond (shown with arrow). Residues on the N-terminal side of the sissile bond are numbered P_1, P_2, and so on (P for Peptide). Residues on the C-terminal side of the sissle bond are numbered P_1', P_2'. . . . The regions on the surface of the enzyme that bind the corresponding amino acids are called S_1, S_2, S_1', S_2', etc (S for Subsite). Note that there is sterospecificity in the interaction of substrate and protease because the bound substrate has N- to C-terminal polarity (i.e. P_n and P_n' are not equivalent). Also note that while the peptide binds to the enzyme as a contiguous section of primary sequence (usually an extended β-strand), the subsites are constructed from secondary and tertiary structural elements of the enzyme and need not be contiguous in sequence.

Figure 1.3: Protease-Substrate Nomenclature



Catalytic mechanism of serine proteases. The kinetic mechanism of amide and ester bond hydrolysis by serine proteases has been studied extensively (Bender and Killheffer, 1973, Neurath, 1985) The kinetic scheme involves binding of the substrate to form a Michaelis complex, an acylation reaction where a covalent acyl-enzyme is formed with concomitant first product release (an alcohol or amine), and a deacylation reaction followed by second product release. The kinetic sheme is illustrated in figure 1.4.

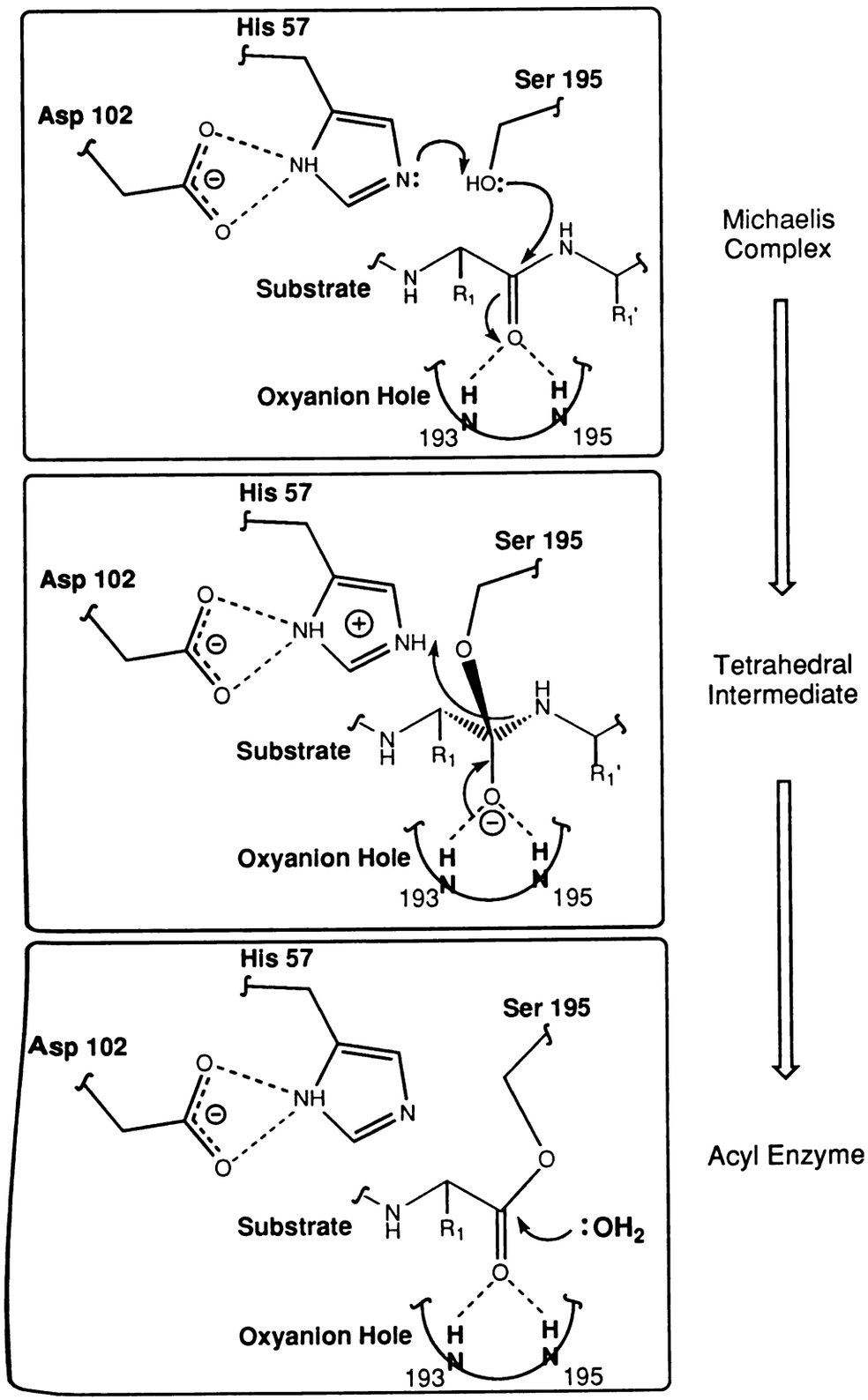
Figure 1.4: Serine Protease Catalytic Scheme



	Acylation-limited	Deacylation-limited
$K_S = \frac{k_{-1}}{k_1}$		
$K_M = K_S\left[\frac{k_3}{k_2 + k_3}\right]$	$K_M = K_S$	$K_M = K_S\left[\frac{k_3}{k_2 + k_3}\right]$
$k_{cat} = \left[\frac{k_2 k_3}{k_2 + k_3}\right]$	$k_{cat} = k_2$	$k_{cat} = k_3$

Serine proteases hydrolyze amide and ester bonds by an acyl transfer mechanism. Upon formation of a non-covalent Michaelis complex, the carbonyl carbon atom of the sissile bond is attacked by the eponymous serine of the catalytic triad to form a tetrahedral intermediate which collapses to the covalent acyl-enzyme (Figure 1.5). Histidine functions first as a general base to enhance the nucleophilicity of the serine, then as a general acid to donate a proton to the nascent leaving group.

Figure 1.5: Serine Protease Mechanism



The aspartic acid of the catalytic triad stabilizes the correct tautomer and conformer of His57 and is thought to contribute to catalysis by stabilizing the transient positive charge formed on the histidine. The deacylation reaction is hydrolysis of the acyl-enzyme ester, a water molecule acting as nucleophile to attack the acyl-enzyme. It is not known whether His57 is the catalytic general base in the deacylation reaction, or if it is specific-base catalyzed. Recent evidence with substrate-assisted catalysis mutants of subtilisin (Matthews & Wells, 1993) and trypsin (Corey et al., 1995) has shown that the catalytic histidine is not required at all for hydrolysis of the acyl-enzyme. P_1'-His containing peptides enhance the reaction rate, and since a P_1'-His is part of the leaving group during the acylation reaction, there is no histidine present during the deacylation reaction. These data suggest that the deacylation reaction is specific-base catalyzed.

Determinants of substrate specificity in serine proteases

Chymotrypsin-like proteases. The only well defined substrate binding pocket for the chymotrypsin family of enzymes is at S_1 resulting in the S_1 site being referred to as the primary binding pocket in these proteases (Perona and Craik, 1995). The S_1 site of chymotrypsin is large and hydrophobic in nature, while the S_1 site in elastase is still hydrophobic but sterically constrained at the mouth. The trypsin S_1 site is the same shape as the chymotrypsin pocket, but it contains a negatively charged Asp at the bottom of the pocket while the *Staphylococcus aureus* V8 protease pocket contains an important histidine (Drapeau, 1978). Hence chymotrypsin is specific for the large hydrophobic amino acids Tyr, Phe, and Trp at P_1, while elastase is specific for the small hydrophobic amino acids Ala and Val at P_1. Trypsin is specific for the basic amino acids Lys and Arg at P_1, and V8 protease is specific for Asp and Glu at

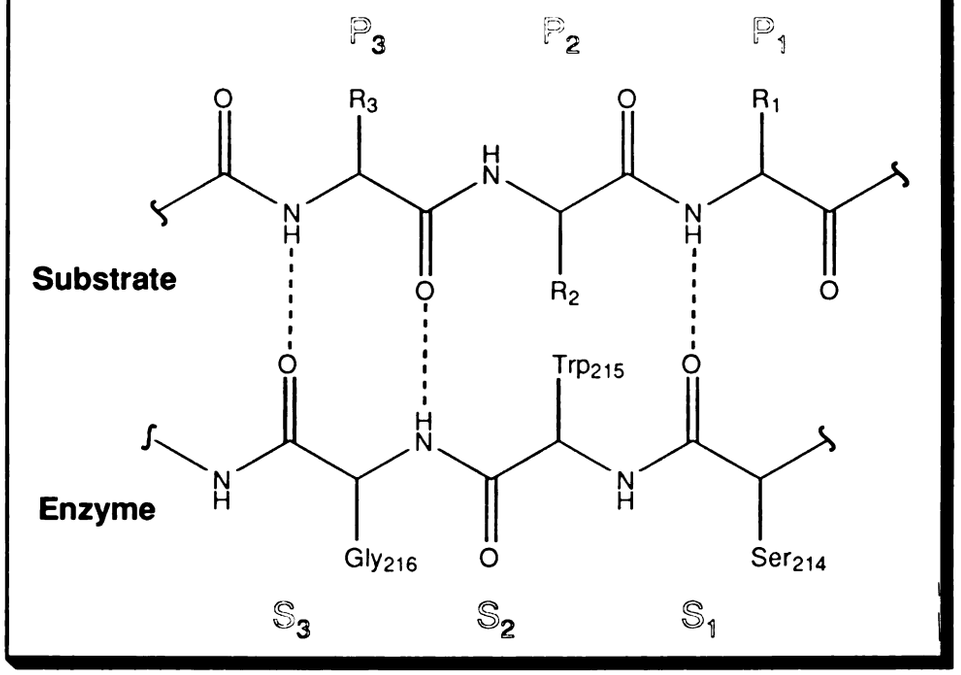
P_1. These enzymes share a common feature: their specificities are restricted to a narrow set of naturally occurring amino acids (Perona and Craik, 1995). An exception to this observation is found in fiddler crab collagenase (Grant and Eisen, 1980, Tsu, et al., 1994). Although fiddler crab collagenase belongs to the chymotrypsin structural class, it possesses high catalytic activity against chymotrypsin-like, elastase-like, and trypsin-like substrates. The S_1 site of collagenase demonstrates that the chymotrypsin scaffold is capable of maintaining high catalytic efficiency against a broad range of substrates in one enzyme. This is interesting from a protein engineering point of view because it may be easy to modulate the specificity of crab collagenase given its existing high level of activity against a structurally-varied set of substrates. Previous experience has shown that it is easier to remove functions from proteins than to add them. The bacterial enzyme α-lytic protease shares the double β-barrel structural fold of the chymotrypsin family, but is structurally divergent in the S_1 site and several surface loops from the other enzymes in this class (Fujinaga, et al., 1985). Alpha-lytic protease has the relative specificity of Ala > Met, Val, Gly > Nle > Leu > Phe (Bone, et al., 1991). Specificity engineering experiments in the P_1 pocket of α-lytic protease have yielded an understanding of the observed broad specificity in which the flexibility of the enzyme allows it to deform to accommodate varied substrates. This is in marked contrast to the structural rigidity found in the S_1 sites of other members of this family.

Subtilisin-like proteases. In contrast to the chymotrypsin-like enzymes, the hydrophobic binding site in subtilisin is not a well-defined, deep pocket, but rather a shallow groove resulting in a less stringent specificity for this protease. Subtilisin will cleave a variety of P_1 peptide bonds in the order Tyr,

Phe > Leu, Met, Lys > His, Ala, Gln, Ser >> Glu, Gly (Estell, et al., 1986, Wells, et al., 1987). A subfamily of subtilisin-like proteases, prohormone convertases, are specific for basic diads (K-K, K-R, R-R) at the P_2-P_1 sites. Examples of this class are kexin (Fuller, et al., 1989), and furin (van den Ouweland, et al., 1990). Although no 3-dimensional structure is currently available for this class of protease, homology and mutational studies have identified a series of negatively-charged amino acids that contribute to the observed specificity, presumably through electrostatic interactions (Creemers, et al., 1993, Siezen, et al., 1991).

Extended binding-site interactions. The primary specificities (P_1) of serine proteases have been modulated by changing residues that directly contact the substrate (Carter and Wells, 1987), emphasizing the importance of the P_1-S_1 interactions in catalytic activity. These experiments have suggested that efficient catalysis is dependent upon binding of the substrate in proper "catalytic register", meaning precise alignment of the sissile bond with the catalytic machinery of the protease (Craik et al., 1985). In addition to the S_1-P_1 interactions that are most commonly studied, the other substite interactions play an important role in the activity of serine proteases. X-ray crystallographic studies of several enzymes in the chymotrypsin class have revealed a set of specific hydrogen bond interactions between the P_1-P_3 residues of the substrate with the S_1-S_3 subsites on the enzyme (Huber & Bode, 1978). Peptide substrates lay across the subsites on the surface of the enzyme in an extended β-strand conformation, making an anti-parallel β-sheet with the 214-216 backbone region of the enzyme (Figure 1.6).

Figure 1.6: P_1-P_3 Interactions



The subtilisin family of proteases also exhibit specificity at the P_2-P_4 postitions, with the most marked effect at P_4 (Gron, et al., 1992). The S_4 subsite in subtilisin is a large hydrophobic pocket, and accordingly the observed order of specificities are Phe > Leu, Ile, Val > Ala at the P_4 position of the substrate. Structural analyses have shown that surface loops flanking the active site in the chymotrypsin family of enzymes play an important role in the specificity (Perona & Craik, 1995, Fig13). A striking example of this is enteropeptidase, which specifically activates trypsinogen to trypsin by recognizing DDDDK in the P_5-P_1 positions (Bode, et al., 1978). Enteropeptidase cleaves DDDDK in trypsinogen with 10^4 greater efficiency than does trypsin (Maroux, et al., 1971). While the origin of this relative rate enhancement is not well understood, sequence mapping has shown that 4

basic residues in a surface loop of the enzyme may make advantageous electrostatic interactions with the substrate (LaVallie, et al., 1993).

Extended subsite interactions (subsites other than S_1) can be exploited for engineering protease specificity. Therefore, it is useful to be able to measure effects of binding at the extended sites on catalysis, especially on the prime side of the scissile bond. An assay recently developed to measure specificity on the C-terminal side of the sissile bond (the prime side), has confirmed a predominantly hydrophobic preference by trypsin relative to chymotrypsin at P_1', whereas chymotrypsin hydrolyzes Arg and Lys residues at P_1' 10-fold better than trypsin (Schellenberger, et al., 1993,1994). The same assay shows that there is no stringent specificity on the prime side of the sissile bond for either protease. In addition, substrate phage display has allowed especially labile substrates to be found for subtilisin H64A (Matthews & Wells, 1993) These technologies may be used to map specificities of engineered proteases and aid in iterative improvement of protease designs.

Project scope. The scope of this project is limited to the structure-based approach for engineering the substrate specificity of trypsin. Simple concepts such as charge and shape complementarity, geometrical constraints, and matched polarity are used to guide the design process in order to predictably alter the specificity of the protease. For the purpose of engineering specificity into the extended binding sites of trypsin, it is helpful that the surface loops, and not core secondary structural elements, are responsible for the observed specificities. The prospect of engineering surface loops successfully is greater than for engineering core structures at this time.

Two independent methods are used to augment the normal Lys/Arg specificity of trypsin by creating histidine specificity at different subsites on the enzyme. The first is referred to as substrate-assisted catalysis (Carter and Wells, 1987), wherein the active site histidine of the protease is removed by mutagenesis. This in turn makes activity of the protease dependent on a histidine supplied by the substrate to the enzyme-substrate complex. The second method, referred to as metal-assisted specificity, utilizes an engineered transition metal binding site on the enzyme (Willett, et al., 1995) . This method relies on the natural affinity of histidine residues for the transition metal ions, copper, nickel, and zinc, creating a metal binding site that bridges the enzyme and substrate.

Chapter 2

Tools for Engineering Protease Specificity

Introduction

This chapter introduces the two methods used to engineer specificity into trypsin: substrate-assisted catalysis (chapter 3), and engineered metal binding sites referred to as metal-assisted catalysis (chapters 4 & 5). Both methods create histidine specificity in the subsites of the enzyme in addition to or in place of the requisite Arg/Lys specificity exhibited by trypsin at the P_1 position. Substrate-assisted catalysis adds histidine specificity to the P_2 or P_1' positions while maintaining the requirement for a basic residue at P_1. In contrast, the engineered metal binding site adds histidine specificity to the P_2' position of substrate and enables the enzyme to hydrolyze peptide bonds with tyrosine at the P_1 postion, in effect obviating the basic residue at that position. Both methods have in common the addition of histidine specificity to the enzyme through engineering of the subite interactions with the substrate.

Substrate-assisted catalysis

Substrate-assisted catalysis is a direct method for enhancing serine protease specificity to include histidine in addition to the cognate P_1 amino acid (see Fig. 1.3). The key to the method is based upon a detailed knowledge of the catalytic mechanism of serine proteases . In the normal function of the catalytic triad of serine proteases, the active-site histidine functions both as a general base and as a general acid during catalysis (see Figs. 1.4 & 1.5). The importance of the active-site histidine is underscored by the observation that its removal lowers the activity of subtilisin H64A by 10^6 in kcat/Km (Carter and Wells, 1987), and trypsin H57A by 10^4 (Corey, et al., 1992). Modeling of the

substrate-protease interactions of subtilisin have shown that a histidine supplied from the P_2 position of substrate could occupy an equivalent position in the H64A mutant that His64 occupied in the native enzyme, thereby reconstituting the catalytic triad in an intermolecular form (Carter and Wells, 1987). Subtilisin H64A was shown to have a 170-fold increase in kcat/Km for a tetrapeptide substrate (suc-FAXF-*p*-nitroanalide) containing a histidine at the P_2 position relative to a substrate with an alanine at that position. Although the specificity of subtilisin H64A is enhanced by requiring a His at P_2, the increase in specificity comes at a cost of decreased catalytic efficiency: subtilisin H64A is 5000-fold lower in activity on suc-FAHF-pNA than the wild type enzyme. This is most likely due to a compromised conformation of the P_2 histidine in the intermolecular catalytic triad. Subsequent engineering of subtilisin H64A resulted in a 155-fold increase in activity by introducing 5 additional mutations and using an optimized substrate, suc-AAHY-pNA (Carter, et al., 1991). The heptamutant subtilisin has been used to specifically cleave fusion proteins, demonstrating a practical use for the engineered protease. In addition, phage display technology was used to identify especially labile peptide substrates for subtilisin resulting in the observation that a histidine at the P_1' position of substrate is also able to participate in substrate-assisted catalysis (Matthews and Wells, 1993).

A success in protein engineering can be further established by testing the generality of the result. Techniques that can be applied to more than one system can be used as tools for protein engineering. If substrate-assisted catalysis can be shown to be a trait available to many serine proteases, the tool can prove to be powerful by allowing us to use the wide range of specificities found in serine proteases to create more site-specific enzymes. In this light,

the active-site His57 of trypsin was changed to Ala (trypsin H57A, chymotrypsin numbering) to test if substrate-assisted catalysis was applicable to the chymotrypsin family of serine proteases. Initial results using tetrapeptide substrates (suc-AAXR-pNA, where X=A, or H) were disappointing, showing that trypsin did not exhibit substrate-assisted catalysis on these substrates. However, when tested against peptide substrates that span the N-terminal as well as the C-terminal side of the sissile bond, trypsin H57A activity was increased 67-fold and 73-fold on a HR (P_2-His) and RH (P_1'-His) substrates, respectively, relative to substrates without histidines at those positions. This result demonstrates the importance of filling the subsites on trypsin for substrate-assisted catalysis to occur, and underscores the observation that the substrate sequence can be optimized as part of the engineering process.

An advantage to using trypsin for substrate-assisted catalysis is its pre-existing narrow specificity for the basic amino acids, Arg and Lys, at the P_1 position. By further requiring a histidine at position P_2 or P_1' of the substrate, trypsin H57A is quite site-specific (for models, see Fig 3.1). Histidine has a natural abundance of 2.1% in proteins, arginine 4.7%, and lysine 7.0% (Klapper, 1977). The occurrence of H-R/K or R/K-H diads is statistically 0.1% (HR or RH) and 0.15% (HK or KH), narrowing the potential cleavage sites of trypsin H57A relative to trypsin by a factor of 50. Coupled with the fact that not all of those recognition sites will be on the solvent accessible surface of the enzyme, the recognition sequence for trypsin H57A is indeed rare, and can complement the specificity of subtilisin H64A nicely. In contrast, wild type subtilisin has a broad specificity at the P_1 position, in order of decreasing activity Y > F > M > K, A, Q >> E, and therefore requires extensive engineering

to make its specificity more stringent, so that while subtilisin H64A is enhanced in specificity relative to the subtilisin, it is still less specific than trypsin H57A. This is a good reason to apply the concept of substrate-assisted catalysis to other members of the serine protease family. Trypsin H57A has been shown to cleave peptides specifically at R/K-H and H-R/K with the best kcat/Km value only 60-fold lower than trypsin for a peptide with histidines at both the P_2 and P_1' positions. Trypsin H57A has also been shown to specifically cleave a globular protein, ornithine decarboxylase, at a naturally ocurring trypsin H57A recoginition site (Corey, et al., 1995) (see Fig. 3.2). It has also been used to specifically cleave two fusion proteins engineered to contain the recognition site, one a plant hormone receptor with a His-tag for purification, and the other a trypsinogen mutant with an altered propeptide-enzyme junction (see Figs 3.3 &3.4). Still, while trypsin H57A may possess stringent specificity, the challenge we are now faced with is how to increase the hydrolysis rate of a target sequence in order to improve the enzyme.

Transition metal binding sites as engineering tools

Metal ions serve a variety of roles in proteins, the most important being to enhance structural stability of the protein, to participate in catalysis, and to mediate macromolecular recognition (Glusker, 1991, pp.1-5). As a result approximately one third of all known proteins require a metal ion for their structure or function (Ibers and Holm, 1980). Different metal ions serve different roles in proteins, with each case determined by the chemical and structural properties of the metal ions and the functional groups on the protein that comprise the metal binding site.

Alkaline earth metals such as calcium and magnesium serve structural as well as functional roles and are found mainly bound to 6, 7, or 8 oxygen atoms from protein carboxylate groups or carbonyls in an octahedral geometry. Calcium binding proteins can be split into two main groups: 1) those that are stabilized by Ca^{+2}, such as in trypsinogen (Bode, et al., 1978; Huber and Bode, 1978), and 2) those that bind Ca^{+2} reversibly, performing regulatory roles by controlling the concentration of free Ca^{+2} as observed with EF hand structures of proteins involved in cellular messaging systems (Herzberg and James, 1985; Szebenyi and Moffat, 1986). Magnesium is found mostly in proteins that utilize ATP and is important both for substrate recognition and phosphate group transfer (Mildvan, 1987).

Iron is found functioning as an electron carrier in electron transfer proteins such as rubredoxin where it is coordinated by 4 cysteines in a distorted tetrahedron and cytochrome c, or in the form of iron-sulfur clusters as found in ferredoxin (Adman, 1979). Porphyrin-bound iron catalyzes redox chemistry in cytochrome c peroxidase (Edwards, et al., 1987) and cytochrome P_{450}.(Poulos, et al., 1985). Iron is most commonly known to store and transport molecular oxygen by the heme proteins myoglobin (Kendrew, et al., 1960) and hemoglobin (Perutz and Lehmann, 1968) where it is liganded in a square bipyramidal gometry by 4 nitrogen atoms from the porphyrin ring, from a histidine in the protein, and by molecular oxygen or CO_2.

The transition metals nickel, copper, and zinc, serve a variety of structural and catalytic roles in proteins and their binding sites vary accordingly. Copper functions as a redox center in the type I, or blue copper proteins azurin and plastocyanin (Adman, et al., 1978; Norris, et al., 1983) where it is bound by two

histidines, a cysteine, and a methionine in a distorted tetraheral geometry. Copper ions are used as oxygen carriers in haemocyanin, found in some marine animals, where it is bound by nitrogen atoms in a fashion analogous to iron in heme proteins (Gaykema, et al., 1984). Nickel is less abundant in biological systems than either copper or zinc (Glusker, 1991, Ch.1). However, nickel is found urease, where it functions as a bi-nickel center, one ion bound by 3 ligands and the other by 5 (Jabri, et al., 1995). Zinc is ubiquitous in biological systems and functions in structural and catalytic roles (Christianson, 1991). For example, zinc is important for mediating protein/nucleic acid recognition through structures known as zinc fingers where the zinc ion is usually coordinated by 2 histidines and two cysteines in a tetrahedral geometry (Berg, 1986). Zinc is a nucleophilic catalyst in carbonic anhydrase where it is liganded by three histidines and one water in a tetrahedral geometry (Eriksson et al., 1988). Zinc also catalyzes peptide bond hydrolysis by carboxypeptidase A it is liganded by two histidines, a glutamate, and a water molecule in a tetrahedral geometry (Quiocho and Lipscomb, 1971; Rees, et al., 1983), and also in thermolysin, where it is also bound in a tetrahedral geometry (Holmes and Matthews, 1982).

We have learned a great deal about the structure of metal binding sites through numerous X-ray crystallographic, and spectroscopic studies of metal ions bound to small molecules and proteins. These studies have revealed that metal binding sites represent one of the most tractable protein design problems because an engineered site requires a minimal number of mutations (two ligands will suffice in most cases), and application simple geometric considerations. With the aid of recombinant DNA techniques, we can now use what we know to design metal binding sites in proteins to effect

desired changes in structure and function. Engineered metal binding sites thus make excellent tools for engineering proteins, allowing us to test our understanding of protein structure/function relationships, and build a database of engineering results.

Applications of engineered metal binding sites. Engineered metal binding sites are useful tools for protein engineering because they are essentially simple in design, can exist as discrete structural units, and have a wide variety of applications as evidenced by the various functions of metals in proteins. Metal binding sites have been engineered with natural and unnatural residues into peptides to enhance α–helicity (Ghadiri and Fernholz, 1990; Ruan, et al., 1990) (Ghadiri and Choi, 1990), or to stabilize structures of de novo designed proteins (Handel and Degrado, 1990), (Pessi, et al., 1993) (Regan and Clarke, 1990) (Handel, et al., 1993) (Klemba, et al., 1995). Bipyridyl groups have been synthesized onto the end of peptides to promote double- (Koert, et al., 1990) or triple-helical structures (Ghadiri, et al., 1992; Lieberman and Sasaki, 1991) Engineered metal binding sites have been used to regulate catalysis in *Staphylococcal* nuclease (Corey and Schultz, 1989) and trypsin (Higaki, et al., 1992; Higaki, et al., 1990; McGrath, et al., 1993), or to create potential catalytic sites in antibodies (Roberts, et al., 1990). A metal binding site has been used to effect conformational changes as a probe for allosteric interactions in glycogen phosphorylase (Browner, et al., 1994). Purification tags have been made with His-X_3-His or His_6- metal binding sites (Arnold & Haymore, 1991; Suh, et al., 1991; Todd, et al., 1991). Metal binding sites can be used to enhance the thermal stability of proteins (Braxton & Wells, 1992; Kuroki, et al., 1989) to affect substrate specificity (Willett, et al., 1995), to create laboratory reagents for specifically crosslinking (K. Brown, personal

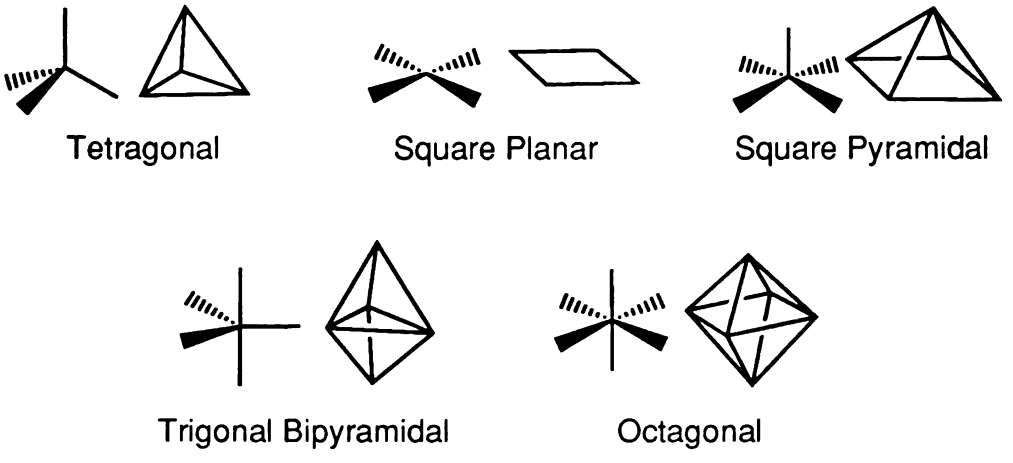
communication) or cleaving proteins (Mack, et al., 1988), and to study the nature of metal binding sites in proteins (Hellinga, et al., 1991; Hellinga & Richards, 1991; Ippolito & Christianson, 1994). These expamples illustrate the great variety of potential uses of designed metal binding sites in proteins that establish the technology as a general method for the study of protein structure-function relationships, and the creation of useful reagents.

Principles of metal binding. Two important characteristics of metal binding sites are the coordination number of the metal ion, and the stereochemistry of the site. The coordination number is the number of ligands binding the metal ion, and the stereochemistry is the geometry of the ligands about the metal ion which define a coordination polyhedron. Ligands are the atoms or groups of atoms that surround a metal ion in a protein. Table 2.1 combines small molecule and protein data to show the most common geometries associated with a given coordination number (Brown, 1988; Glusker, 1991). Figure 2.1 illustrates those coordination polyhedra.

Table 2.1: Metal Binding Site Polyhedra

Coordination Number	Observed Geometries
4	tetragonal, square planar
5	square pyramidal, trigonal bipyramidal
6	octahedral

Figure 2.1: Coordination Polyhedra



Major metal binding ligands in proteins are carboxylate groups from Asp and Glu residues, imidazole rings from histidines, thiol groups from cysteines, thioether groups from methionines, and from the main chain, carbonyl oxygens, and amide nitrogens (Voet and Voet, 1990). It is helpful to consider ligand-metal ion interactions in terms of Lewis acid-base theory (Lewis, 1923). Metal ions are electron deficient, and act as Lewis acids by accepting electron pairs from the electron-rich ligand acting as a Lewis base.

Metal ion (Lewis acid) + Ligand (Lewis base) ——> Metal-ligand Complex

An important concept that helps explain metal-ligand interactions is polarizability (Williams, 1959) and its contribution to Hard Soft Acid Base (HSAB) theory as proposed by Pearson (Pearson, 1966; Pearson, 1968a; Pearson, 1968b; Pearson, 1986) Polarizability is a measure of how easily an electron

cloud can be induced to have a dipole in an electric field. A more polarizable atom has less tightly held electrons, and therefore its electrons are more deformable. The reverse is true of a less polarizable atom. In terms of metal ions, a hard atom generally has a small ionic radius, and a high positive charge, meaning a high charge density. A soft metal ion has a large radius, and a small positive charge, or a low charge density. This gives rise to HSAB theory which essentially states that hard acids preferentially react with hard bases, and soft acids preferentially react with soft bases. This theory is borne out when observing metals bound to proteins and small molecules. Table 2 summarizes the biologically important metal ions and ligands in the context of HSAB theory (Glusker, 1991, p. 5)

Table 2.2: Hard/Soft Classification

Classification	Cation	Ligand
Hard	H^+, Li^+, Na^+, K^+, Mg^{+2}, Ca^{+2}, Mn^{+2}, Cr^{+3}, Co^{+3}	H_2O, OH^-, ROH, RO^-, NH_3, CO_3^{-2}, RCO_2^- (mainly O ligands)
Soft	Cu^+, Ag^+, Au^+, Tl^+, Pd^{+2}, Pt^{+2}, Cd^{+2}, Hg^+, Hg^{+2}	RSH, RS^-, R_2S, CN^-, I^- (mainly S ligands)
Borderline	Cu^{+2}, Ni^{+2}, Zn^{+2}, Fe^{+2}, Fe^{+3}, Co^{+2}, Sn^{+2}, Pb^{+2}, Rh^{+3}, Ir^{+3}, Ru^{+3}	Imidazole, RNH_2 (mainly N ligands)

The polarizability of a metal ion correlates with its ability to form Π bonds in complexes. Hard acids and bases form mostly ionic bonds because the filled d-orbitals of the metal ion that usually form the highest occupied molecular

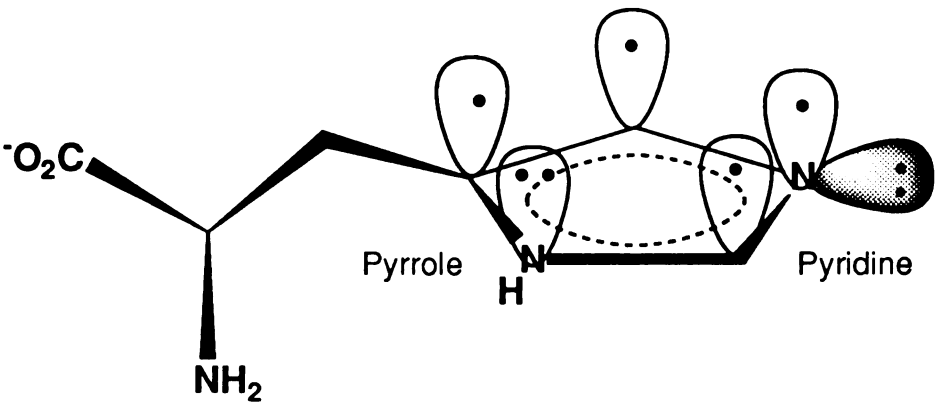
orbital (HOMO) of the complex, and the vacant ligand orbitals of proper symmetry that contribute to the lowest unoccupied molecular orbital (LUMO) of the complex are very different in energy. In contrast, the HOMO and LUMO energies in soft acids and bases are closer in energy, resulting in d-electrons from the metal ion being transferred to the ligand in order to delocalize negative charge built up on the metal ion by σ-bond donation of electrons from the ligands. This Π bond formation yields a partial covalent nature to the metal complexes of soft and borderline acids and bases. In biological systems, Na^{+}, and K^{+}, bind O weakly, Mg^{+2} and Ca^{+2} bind O moderately, but the transition metals Zn^{+2}, Cu^{+2} bind N and S strongly making them good for engineering tight-binding complexes (Williams, 1959). It is interesting to note that most biological metal-ligand interactions are between hard metal ions and hard ligands and are ionic in nature. Soft metal ions (e.g. Hg^{+2}) and soft ligands (e.g. CO, HS, and CN^{-}) are poisons because they bind strongly to their soft counterparts in proteins and inhibit them. Another advantage to using transition metals as engineering tools is that they usually have a coordination number of 4, where Ca^{+2}, Mg^{+2}, and Na^{+} have coordinations numbers of 6, 7, or 8. From a protein engineering viewpoint, it is easier to design 4 liganding goups than 6, 7, or 8. It is because of the strong complex formation and the lower coordination number of transition metals that they are the choice for engineering metal binding sites.

Histidine as engineered ligand. Once the choice is made to design transition metal binding sites as opposed to other metals, the choice of ligand atoms available to the protein engineer is preferably to N and S. Histidine, methionine, and cysteine are the most common ligands observed in proteins that coordinate through N or S atoms. Of those three, histidine is the

preferred choice because of the oxidation problems associated with the sulfur atom found in methionine and cysteine. Not all proteins are stable or functional in the presence of the reducing agents required to keep the sulfur-containing amino acids in their metal binding states, so the situation is best avoided in the design stages. On a practical level, inroducing cysteines onto the surface of a protein can create purification and aggregation problems for the protein chemist, adding difficulty to an already complex problem. Methionines can also be oxidized, and furthermore the long side chain has many conformations available to it (4χ angles) relative to histidine (2χ angles), making the task of placing the thioether moiety in the desired location for metal binding more difficult. While oxygen ligands are also found for zinc metalloproteins (Christianson, 1991), transition metal binding to imidazoles is stronger, and so carbolylates are not used here as engineered ligands. As a result, histidines have been used exclusively to design transition metal binding sites for this series of experiments.

The chemical and structural aspects of the histidine imidazole ring as metal ligand are worth considering because they explain the metal binding properties of this side chain. This five-membered heterocycle is structurally similar to pyridine and pyrrole, and can be considered as a planar aromatic ring system with (4n + 2) Π-electrons as shown in figure 2.2 (only the tops of the p orbitals are shown) (Sundberg and Martin, 1974).

Figure 2.2: Aromatic Nature of Histidine



The Π system of electrons in the imidazole ring consists of one electron from the unhybridized p-orbital of each trigonal ring carbon and the pyridine nitrogen, and two electrons from the unhybridized p orbital of the pyrrole nitrogen, making an aromatic sextet. Metal (or proton) complexation is only through the sp^2-hybridized lone pair on the pyridine nitrogen (shown shaded) because the lone pair on the pyrrole nitrogen is part of the aromatic sextet and complex formation there would result in loss of aromaticity. This explains why generally only one metal ion is bound to the imidazole ring at a time. Resonance stabilization energy for imidazole has been calculated to be 15.4 kcal/mol by the method of Dewar (Dewar, et al., 1969). This is comparable to energies of 20.0, 20.9, and 8.5 kcal/mol calculated for benzene, pyridine, and pyrrole, respectively.

Imidazole possesses two properties important in metal binding: basicity, and Π electron acceptor capability. When compared to pyridine and NH_3, imidazole is intermediate in both properties with the order of increasing

basicity (pKa shown in parentheses) being opposite of the Π electron acceptor capability.

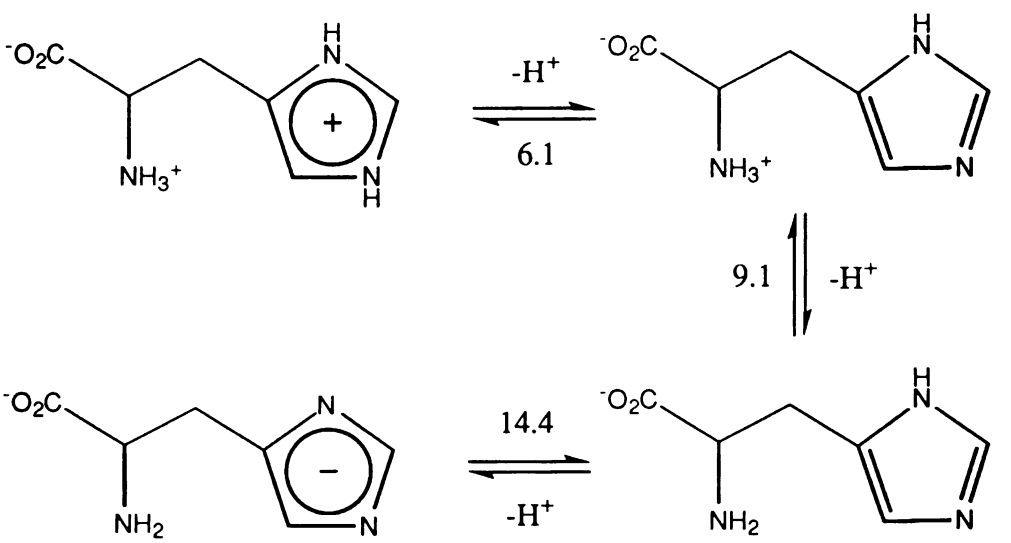
increasing basicity---------->

pyridine (5.2) < histidine imidazole (6.1) < NH_3 (9.3)

<---------- increasing Π acceptor capability

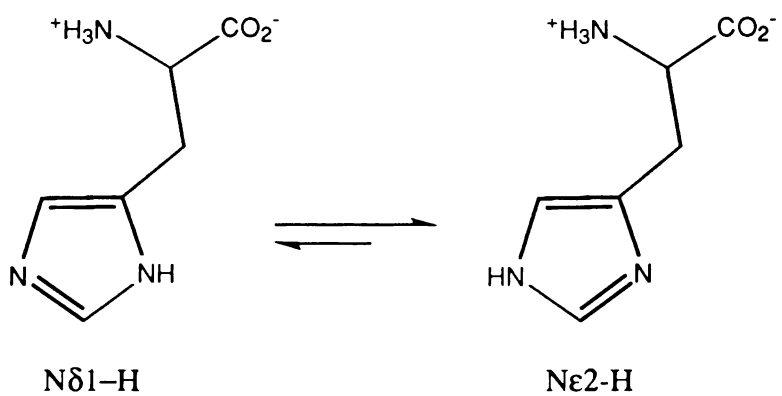
Increasing basicity correlates with increased ability to form strong sigma bonds with protons and metal ions, while increased Π acceptor ablility correlates with stronger Π bonding capabilities. Histidine has a balance between the two properties, which makes it an optimal metal binding residue in proteins. The acid-base equilibrium constants of histidine are shown in figure 2.3 below.

Figure 2.3: Histidine Acid-Base Equilibria



In the context of a protein, the relevant pKa is that of the imidazole ring. A value of 6.1 is close to physiological pH, making histidine a general acid and general base catalyst. Histidine is capable of two tautomeric forms, as illustrated in figure 2.4.

Figure 2.4: Histidine Tautomers

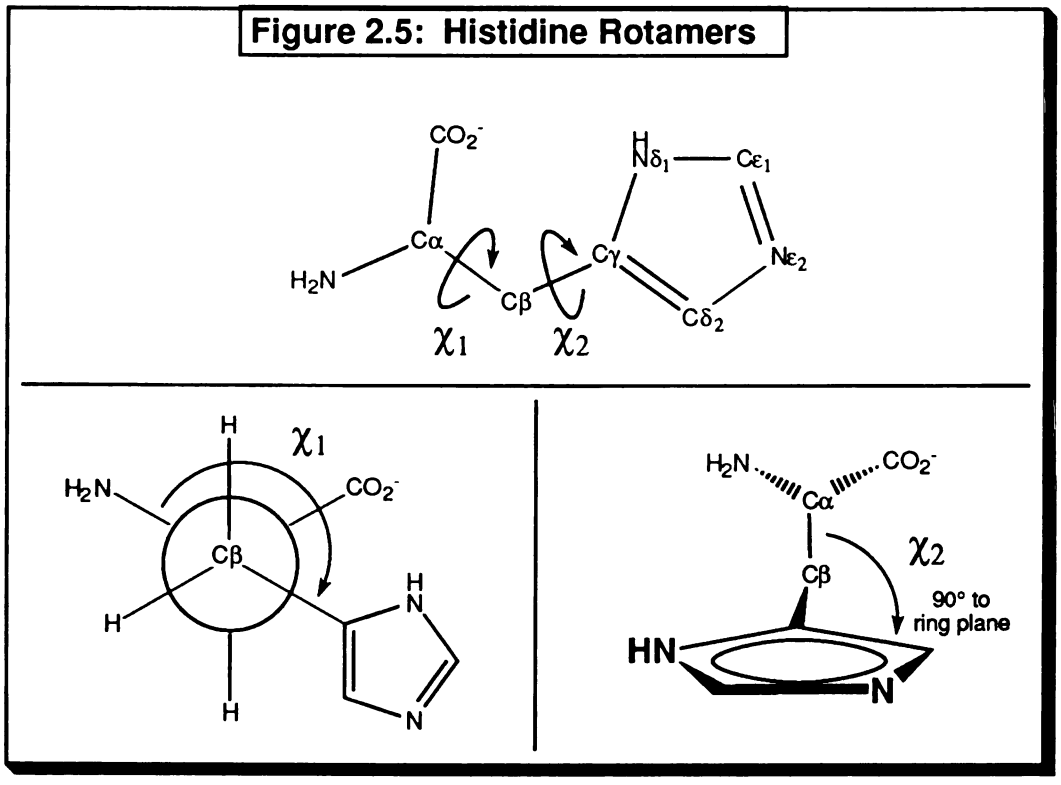


Nδ1–H Nε2-H

In solution, the Nε2-H tautomer of free histidine is favored 4-fold, so that 80% is found as this tautomer (Reynolds, et al., 1973). However, the energy levels of the two forms are close enough that, in proteins, the local structure determines the tautomeric form. This is important for X-ray crystallographic studies where protons are not visible, and fitting a histidine side chain into electron density in the proper orientation can be ambiguous. In such cases, a crystallographer or modeler can look for possible hydrogen bond interactions to resolve the ambiguity, although a tautomeric assignment will not always be possible. Fortunately, a metal ion bound to a histidine removes the ambiguity because the nitrogen not bound to metal is most likely to be protonated under physiological conditions.

Analysis of three-dimensional protein structures has revealed that statistically, histidine side chains adopt certain conformations as described by the χ_1 and χ_2 angles (Ponder and Richards, 1987). The most preferred χ_1 angles are ±60°, and 180° (lower left of fig. 2.4), while χ_2 is usually ±90° (lower right of fig. 2.4). χ_1 refers to rotation about the Cα-Cβ bond, and is defined as the dihedral angle formed by the backbone amide nitrogen, Cα, Cβ, and Cγ. χ_2 refers to the rotation of the imidazole ring about the Cβ-Cγ bond, and is defined as the dihedral angle formed by Cα, Cβ, Cγ, and Cδ2. Figure 2.5 illustrates the atom labeling and the two χ angles for histidine.

Figure 2.5: Histidine Rotamers



Structural aspects of histidine-metal complexes in proteins. A survey of zinc metalloprotein structures has revealed that histidines bind zinc atoms mostly in the direction of the sp^2-hybridized lone pair of the pyridine nitrogen which is in the plane of the imidazole and which bisects the angle

made by the liganding nitrogen, and the two adjacent carbon atoms in the imidazole ring (Chakrabarti, 1990). Interestingly, the majority of zinc-bound histidines in proteins are in the Nδ1-H tautomer. In addition, the bond lengths observed for most transition metal complexes with histidine range from 1.9-2.3Å. The stability constants of various metals binding to histidine have been measured and are shown below in increasing order:

$$Ca^{+2} < Mg^{+2} < Mn^{+2} < Fe^{+2} < Co^{+2} < Ni^{+2} < Cu^{+2} > Zn^{+2}$$

The transition metals nickel, copper and zinc bind with the strongest affinity (Irving and Williams, 1953) making them the logical target for engineering a metal binding site with histidines. These observations are important for engineering transition metal binding sites into proteins, and must be considered at the outset of the design process. Implicit in this design philosophy is the idea that at our current level of understanding, we can best succeed by imitating what is observed in nature.

Transition metal chemistry. A basic knowledge of the chemistry of transition metals is important for understanding the properties of metals complexed to proteins. The coordination number, geometries and chemical properties of transition metal complexes can be qualitatively understood by considering the d-electron orbitals of the metal ions in combination with ligand field and molecular orbital theories. The scope of this work is limited to experiments with the first row transition elements, nickel, copper, and zinc. These are the last three elements in the first transition series that fill the 3-d orbitals. Table 2.3 shows the electronic configurations, and the most commonly observed coordination numbers (CN) and geometries of small molecule and protein complexes with these three metals.

Table 2.3: Transition Metal Parameters

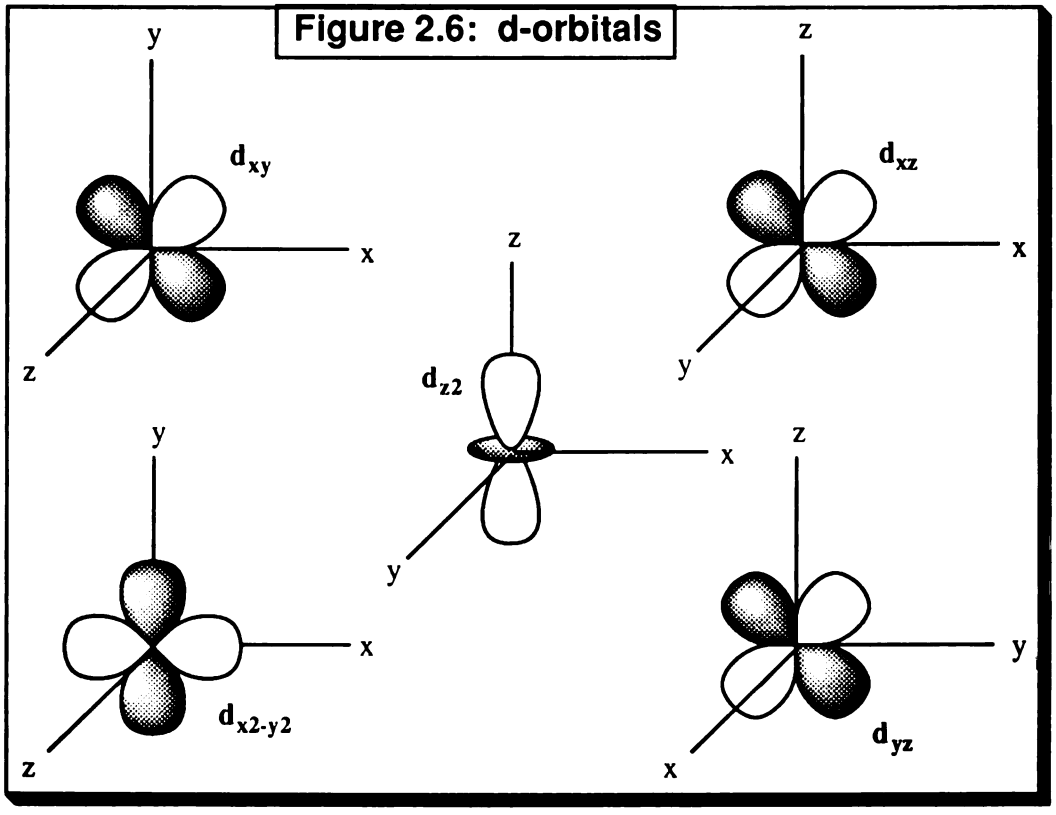
Ion	Configuration	CN	Geometry
Ni^{+2}	d^8	4	square planar sometimes tetrahedral
Cu^{+2}	d^9	4	square planar
Zn^{+2}	d^{10}	4	tetrahedral

The coordination number can be related to the number of d-electrons on the metal ion by the 18-electron rule, and this can then help predict a geometry of the complex (Collman, 1980). The coordination number is the number of σ-bonds in a given complex, with the electrons in the σ-bonds usually donated by the ligand atomic orbitals. The basis of the 18-electron rule derives from the stability of a completely filled third shell, meaning the electronic configuration $3s^2 3p^6 3d^{10}$, and the bonding molecular orbitals that result from the matching symmetry of the overlapping atomic orbitals of ligand and metal ion. The rule states:

$$n + 2(CN_{max}) = 18$$

Where n is the number of d-electrons on the metal ion, and CN_{max} is the maximum coordination number possible for the complex. It makes sense that the CN decreases with increasing d-electrons because those electrons usually fill non-bonding molecular orbitals that are situated between the ligands. Using this rule, CN_{max} can be calculated for Ni^{+2}, Cu^{+2}, and Zn^{+2} to be 4 or 5, with 4 being commonly observed in protein structures. Consideration of the spatial arrangement of d-orbitals with molecular orbital

and ligand field theory leads to an explanation of the observed geometries and stabilities of metal complexes (Collman, 1980). Figure 2.6 illustrates the angular dependence of the d-orbitals.



Ligand field theory is a post-facto method of describing the splitting of the metal d-orbital energies based on an assumed geometry. A central idea of the theory is that filled, non-bonding d-orbitals on the metal ion are more stable when they are directed away from the ligands. Further, d-orbitals that are oriented towards the ligands are destabilized by those interactions. A molecular orbital approach to the problem results in the idea that greater splitting of the molecular orbitals occurs when there is more overlap of ligand and metal atomic orbitals that are close in energy, giving rise to more stable bonding and less stable anti-bonding molecular orbitals. A factor which

contributes to the stability of a metal complex is the ability for Π bonding. Π bonding occurs when low lying empty ligand orbitals of correct symmetry accept non-bonding metal d-electrons, thereby reducing negative charge density on the metal ion and stabilizing the complex. Nickel, copper, and zinc all have the ability to Π-bond in their preferred geometries. On steric grounds, a tetrahedral arrangement of ligands is preferred over square planar, but square planar geometry results with d^8 and d^9 complexes such as Ni^{+2} and Cu^{+2} because of the more stable electronic character in this geometry (Collman, 1980).

In the context of a protein, the observed geometry of a metal binding site will result from a balance of energetic contributions. One one side, the metal ion will prefer geometries that stabilize its d-electrons, leading to a lower engergy complex. On the other side, the conformational constraints placed on the ligands by the local structure of the protein may cause deviation of the geometry from an ideal case, or not permit binding at all. The geometric requirements of the metal and the flexibility of the protein will therefore combine in an engineered metal binding site to yield a particular structure and binding affinity of the site.

Metal binding site design considerations. When designing a metal binding site in a protein, there are several important factors to consider (Hellinga & Richards, 1991). As a first-level criterion, proper metal coordination and geometry must be satisfied so that there are no steric or charge-charge conflicts resulting from the introduced ligands. The bond lengths and angles should agree with those found naturally for the metal ion to be used. For the transition metals, nickel, copper, and zinc, a nominal

bond length of 2.0Å was assumed for these experiments, alowing for small (<0.5Å) movements of the protein. The side chain rotamers of the ligands should be the "preferred" rotamers (Ponder and Richards, 1987), although observation of histidine rotamers in metalloproteins reveals that relatively few possess exactly preferred χ angles (W. S. Willett, personal observation). The observed range of rotamers offers some tolerance in modeling the site, although sites modelled with rotamers that are far from "preferred" may suffer in affinity for metal due to the energetic price of adopting non-preferred rotamers. Two secondary structural motifs have been recognized in protein metal binding sites comprising two protein ligands (Regan, 1995). Sites may be formed from an His-X_3-His sequence on α-helices, or an His-X-His sequence on a β-sheet. These sequences in their proper structural context place the histidines in a possible metal-binding orientation, offering a simple method of metal binding site design. His-X_3-His sites have been introduced on surface helices for purification with immobilized metal affinity chromatography (Arnold & Haymore, 1991). Interior close packing in the hydrophobic core of the protein must be taken into account if the site is to be in the interior of the protein. In the interior of the protein, all hydrogen bonds are usually satisfied. A survey of naturally occurring metal binding sites has shown that they usually have a first shell of hydrophilic liganding residues surrounded by a hydrophobic shell yielding a high hydrophobicity contrast in a radial direction from the metal ion (Yamashita, et al., 1990) The secondary shell of H-bond interactions that surround a metal binding site and stabilize the conformations of the ligands may be important in pre-organizing the site, thereby reducing the entropic cost of metal binding (Christianson and Alexander, 1989). As a practical consideration, it is easier to design sites on

the exterior, or solvent-exposed surfaces of a protein where the flexibility is generally higher and interior packing as well as hydrophobicity contrast is not a consideration. In addition, the metal binding sites in this work are designed to mediate macromolecular recognition, an event that occurs at solvent-accessible surfaces of proteins. Given our current understanding of protein structure-function relationships, these design considerations strive to imitate what is observed in nature.

Chapter 3

Trypsin Specificity Increased Through Substrate-Assisted Catalysis

Abstract

Histidine 57 of the catalytic triad of trypsin was replaced with alanine to determine if the resulting variant would be capable of substrate-assisted catalysis (Carter, P. & Wells, J. A. (1987) *Science* 237, 394-9). A 2.5-fold increase in kcat/Km was observed on tri-or tetra-peptide substrates containing *p*-nitroanalide leaving groups and histidine at P2. In contrast, hydrolysis of peptide substrates extending from P6 to P6' is improved by 70-300 fold by histidine in the P2 or P1' position. This preference creates new protease specificities for sequences HR↓, R↓H, HK↓, and K↓H. The ability of histidine from either the P2 or the P1' position of substrate to participate in catalysis emphasizes the considerable variability of proteolytically active orientations which can be assumed by the catalytic triad. Trypsin H57A is able to hydrolyze fully folded ornithine decarboxylase with complete specificity at a site containing the sequence HRH. Trypsin H57A was compared to enteropeptidase in its ability to cleave a propeptide from trypsinogen. Trypsin H57A cleaved the propeptide of a variant trypsinogen containing an introduced FPVDDDHR cleavage site only 100-fold slower than enteropeptidase cleaved trypsinogen. The selective cleavage of folded proteins suggests that trypsin H57A can be used for specific peptide and protein cleavage. The extension of substrate-assisted catalysis to the chymotrypsin family of proteolytic enzymes indicates that it may be possible to apply this strategy to a wide range of serine proteases and thereby develop

various unique specificities for peptide and protein hydrolysis. Most of this chapter is being published (Corey et al., (1995) *Biochemistry* , in press).

Introduction

Proteases are valuable tools for the manipulation of proteins and peptides. However, in comparison to the stringent specificities of restriction endonucleases, most proteases lack selectivity for unique primary sequences. Highly selective proteases do exist which possess the ability to cleave proteins at unique sites, including Factor Xa (Nagai & Thogersen, 1987), enteropeptidase (Kitamoto et al., 1994, Light et al., 1980), tobacco etch virus (TEV) protease (Parks et al., 1994), and picornavirus 3C protease (Walker et al., 1994). Yet their collective specificities are few, and the discrimination against closely related sequences is poor. As a result, there are applications for which no adequate protease is available. While the range of protease specificities will continually expand as naturally occurring proteases are identified, the current lack of selective proteases has made the development of new protease specificities a target for protein design.

Various strategies have been developed to confer novel substrate specificities on serine proteases. Plasminogen activators have been coupled to anti-fibrin antibodies to selectively target blood clots (Haber et al., 1989, Yang et al., 1994), and chymotrypsin has been biotinylated to localize proteolysis to avidin-bound substrates (Bayer et al., 1990). Proteases have been altered in the primary specificity pocket by rational design or by random mutagenesis (for a review, see Perona & Craik, 1995), engineered in the extended binding subsites so that a metal ion bridges the enzyme to the substrate (see chapters 4 & 5) (Willett et al., 1995), and modified by the removal of a catalytic residue

that is subsequently provided by the substrate (Carter & Wells, 1987). The last of these strategies was applied to subtilisin by removing the active site histidine, making the enzyme specific for sequences containing histidine adjacent to small amino acids.

Subtilisin BPN' was altered by substituting histidine 64 with alanine to make optimal catalysis dependent on the presence of histidine at either the P2 or P1' position (Schechter & Berger, 1968) of the substrate (Matthews & Wells, 1993). The resulting subtilisin H64A was specific for hydrolysis at HX↓ or X↓H peptide linkages where X is a small hydrophobic amino acid. This variant has been shown to cleave a recombinant parathyroid hormone fusion protein more efficiently than thrombin or enteropeptidase (Forsberg et al., 1991). Additional site-directed mutagenesis of subtilisin H64A was used to increase the hydrolysis rate of the enzyme for particular substrates (Carter et al., 1991). Subsequently, phage display was used to identify especially labile substrate sequences which are cleaved more efficiently by the enzyme (Matthews & Wells, 1993).

We applied the substrate-assisted catalysis strategy to the chymotrypsin family of serine proteases. Although the three-dimensional folds of the chymotrypsin and subtilisin families are unrelated, the catalytic triads of the enzymes superimpose with a root-mean-square difference of less than 0.2Å (Corey & Craik, 1992 and references therein). The enzymes of the chymotrypsin family range from the digestive enzymes trypsin, chymotrypsin, and elastase whose specificity is determined primarily by their S1 subsites (for review, see Perona & Craik, 1995), to the highly selective proteases of the blood coagulation pathway such as thrombin (Vu et al., 1991),

tissue plasminogen activator (Madison et al., 1995), and complement Factor D (Narayana et al., 1994), which interact more extensively with their substrates. If substrate-assisted catalysis were to prove general for serine proteases, their diverse substrate preferences would allow the development of various stringent specificities.

Histidine 57 of the catalytic triad of rat anionic trypsin was replaced with alanine, and the catalytic properties of the variant enzyme were characterized on both synthetic and natural substrates. Trypsin was chosen as an archetypal member of the structurally-related chymotrypsin family of serine proteases. Trypsin is specific for the cleavage of Arg and Lys at P_1, so that introduction of histidine specificity at P_2 by trypsin H57A would yield H-R/K or R/K-H specificities not possessed by subtilisin H64A or any other known protease. The catalytic properties of trypsin H57A and trypsin were compared on extended peptide substrates which closely mimic target substrate sequences within fusion proteins. Trypsin H57A was also analyzed for its ability to efficiently hydrolyze folded proteins at unique sites. In spite of the dissimilarity in tertiary structure between trypsin and subtilisin, substrate-assisted catalysis by trypsin H57A occurs with both peptide and protein substrates. This result demonstrates the potential to generate new protease specificities by adapting the chymotrypsin family of serine proteases for substrate-assisted catalysis.

Materials and Methods

Materials. Peptides were synthesized by solid phase synthesis on an Applied Biosystems (Foster City, CA) model 430A or a Symphony Multiplex (Rainin, Woburn MA) peptide synthesizer using Fmoc (N-(fluorenyl)

methoxycarbonyl) chemistry. Each peptide was desalted using a Whatman C-4 cartridge (Fisher). The molecular weights of the peptides and their proteolyzed fragments were confirmed by Mass Spectral analysis using a VG (Altrincham, England) 30-250 Quadrapole mass spectrometer. Wild type rat anionic trypsin and trypsin H57A were obtained as described (Corey et al., 1992) from a bacterial expression system, or from a yeast expression system (Hedstrom et al., 1992). Proteases obtained from either expression system were indistinguishable. Purified recombinant ornithine decarboxylase from *Leishmania donovani* (LODC) (Osterman et al., 1995) was obtained from Dr. Margaret Phillips (University of Texas Southwestern Medical Center at Dallas). Automated Edman degradation of trypsin H57A cleavage products of ornithine decarboxylase was performed on an Applied Biosystems model 477A amino acid sequencer with an online model 120A pTH amino acid analyzer using standard conditions. The peptidyl *p*-nitroanilide (pNA) substrate, succinyl-Ala-His-Arg-pNA (sucAHR-pNA), was obtained from Dr. Paul Carter (Genentech, South San Francisco, CA) and succinyl-Ala-Ala-Ala-Arg-pNA (suc-AAAR-pNA) was obtained from Bachem.

Computer modeling methods. The Insight II program (Biosym Technologies) running on a Silicon Graphics Indigo XS24 was used to display the X-ray crystal structure of rat anionic trypsin D102N complexed to the bacterial serine protease inhibitor ecotin (McGrath et al., 1994). The trypsin/ecotin interface at the active site is used as a model of the enzyme substrate interaction with ecotin representing bound substrate. To model the potential of histidine from a substrate to participate in catalysis by trypsin H57A, the amino acids at either the P2 or the P1' positions of ecotin (Thr83 or

Met85) were replaced with histidine, and the P1 amino acid (Met84) was replaced with a lysine. Histidine 57 on trypsin was replaced with alanine to create a model of trypsin H57A. The histidine side chains on the substrate model were rotated manually to find a position that formed suitable hydrogen bonds with Ser195 of trypsin. Criteria for good hydrogen bonds were bond distances between 2.5 and 3.5 Å, and a nearly coplanar arrangement of the imidazole-ring nitrogens of the histidine from the substrate, and the hydroxyl proton from Ser195 of trypsin.

Kinetic assays. The cleavage of the peptides by trypsin or trypsin H57A was monitored at 220 nm by reverse phase HPLC (Rainin) using a Microsorb 5 μm 300 Å reversed-phase column (4.6 mm x 25 cm) (Rainin) and a 10-50% gradient of 0.1% trifluoroacetic acid in doubly-distilled water (buffer A) and 0.08% trifluoroacetic acid in 95% acetonitrile, 5% doubly-distilled water (buffer B) (Corey & Craik, 1992). Cleavage assays were performed at 37°C in 100 mM NaCl/20 mM $CaCl_2$/50 mM Tris-HCl, pH 8.0 buffer. Assays with trypsin required 2-20 minutes and were terminated through the addition of 0.5 volumes of 1% aqueous trifluoroacetic acid. Assays with trypsin H57A required 2 hours for substrates containing histidine at P2, or P1' and 48 hours for substrates lacking histidine at these positions. Assays with trypsin H57A were terminated by freezing on dry ice. The integrated areas of the cleavage products and starting material after HPLC analysis were used to determine the rate of cleavage at varied substrate concentrations. Eadie Hofstee analysis of the initial rate of peptide cleavage was used to determine k_{cat} and K_m values. The location of cleavage within the peptides was confirmed by mass-spectral analysis of the peptide fragments recovered after enzymatic digestion and

HPLC purification. Assays for the hydrolysis of peptides containing D-histidine were performed as described above and were analyzed after a 48 hour incubation at 37°C.

Cleavage of LODC by trypsin H57A was performed at 37°C in a buffer consisting of 20 mM Tris-HCl, pH 7.5, 2 μM ß-mercaptoethanol, 0.02% Brij-35, 0.5 mM EDTA, and 20 μM pyridoxal phosphate. The cleavage reactions were stopped periodically over 1-48 hours by freezing on dry ice and the appearance of hydrolyzed products was monitored periodically over 72 hours by SDS-PAGE. Kinetic constants for the cleavage of LODC by trypsin H57A were determined using LODC concentrations of 0.5, 2, 3, 4, 5, and 6 μM and 0.2 μM trypsin H57A at 37°C in a buffer containing 25 mM Tris-HCl, pH 7.5, 0.02 Brij-35, 2.5 mM DTT, and 20 μM pyridoxal phosphate, for times ranging from 3.5 to 6.5 hours. These concentrations result in rates in the linear range of the reaction and involved less than 20% conversion of substrate. Reactions were stopped by freezing at -70°C and cleavage products were separated by SDS-PAGE using 12% polyacrylamide. Gels were stained with Coomassie brilliant blue R-250 for one hour and then destained overnight in 10% acetic acid, 5% methanol, 85% water. Intensities of substrate and product bands were quantitated on a model 300A scanning densitometer from Molecular Dynamics which was operated with Image Quant 3.0 software. The percentage of LODC cleaved at a given LODC concentration was determined by calculating P/S+P where P was the intensity of the product band and S was the intensity of the substrate band. These ratios were used to calculate initial velocities of cleavage. The kinetic parameters k_{cat} and K_m were derived from Eadie-Hofstee analysis, and were done in duplicate. Cleavage of substrates

sucAHR-pNA and sucAAPR-pNA were monitored by UV spectrophotometry as described (Corey & Craik, 1992).

Production of recombinant trypsinogen or trypsin. Trypsinogen and trypsinogen variants were constructed and expressed in yeast as described previously (Willett et al., 1995) using site-directed mutagenesis and the following synthetic oligonucleotides to mutate the bluescript-based mutagenesis plasmid, pST: GGAGGCAAGCACAGCTGCCAGGG, D189H oligo; GGTGTCTGCGGCCGCATGCTATAA, H57A oligo; GTGGATGATGA TCATAGAATCGTTGGAGG, D14H/K15R oligo. The mismatched basepairs are underlined. The DNA encoding the recombinant proteins was sequenced in its entirety in pST to confirm the desired mutations and ensure that unwanted mutations did not occur elsewhere in the gene. Mutant forms of the protein were expressed in bacteria for a rapid determination of kinetic parameters as described previously (Corey et al., 1992). To obtain larger amounts (>10mg) of the recombinant proteins, the yeast expression system was used. Concentrations of trypsin or trypsinogen were calculated from absorbance at 280 nm using an extinction coefficient of 34,300 M^{-1}.

Cleavage of Trypsinogen. Cleavage of trypsinogen by trypsin H57A was performed in 100 mM Tris, 1 mM $CaCl_2$, pH 8.0 at 37°C. A typical reaction contained 20μM trypsinogen and 1μM trypsin H57A in a 50μl volume in a 0.5ml siliconized eppendorf tube. Cleavage of trypsinogen by enteropeptidase was accomplished using the same reaction conditions except that the enteropeptidase concentration was 50nM in the reaction. Ten-microliter samples were taken from the reaction mix at various times and analyzed by Coomassie-stained SDS-PAGE containing 12% acrylamide. Cleavage of

trypsinogen yields trypsin, which upon SDS-PAGE analysis is indistinguishable from the trypsin H57A used to cleave the propeptide. As a result, the extent of cleavage by trypsin H57A was monitored by the disappearance of trypsinogen. Trypsinogen was well-resolved from mature trypsin, allowing indirect observation of its loss during conversion to trypsin. The intensity of the Coomassie-stained trypsinogen band was quantitated on a model 300A scanning densitometer from Molecular Dynamics which was operated with Image Quant 3.0 software and was used to evaluate cleavage by trypsin H57A.

Results

Computer modeling. Replacement of amino acids at P2 and P1' with histidines from a potential substrate was modeled from the structure of trypsin complexed with ecotin (McGrath et al., 1994). The modeling suggests a molecular basis for substrate-assisted catalysis by trypsin (Fig. 3.1). The histidine introduced at P2 has a χ_1=-139.5° and a χ_2=-71.3°. The histidine introduced at P1' has a χ_1=-19.3° and a χ_2=–51.6°. With the exception of the -19.3° χ_1 angle observed in the P1' model, these χ angles fall in the observed range of histidine rotamers (Ponder & Richards, 1987). The modeled histidines make no short contacts less than 3.0Å other than hydrogen bonds. Implicit in the modeling is the hypothesis that the histidine brought to the active site by the substrate can participate in the reaction by abstracting the proton from Ser195 to allow nucleophilic attack of the sissile bond in the substrate. The modeling suggests that the trypsin scaffold can allow a histidine in either the P2 or the P1' position of the substrate to bind in a

conformation placing one of the histidine ring nitrogens in a proper position to abstract the Ser195 proton.

Numerous structural and functional studies of trypsin show that His57 has its $N\delta_1$ proton hydrogen bonded to the Asp102 carboxylate at a distance of 3.0 to 3.5 Å, and its $N\varepsilon_2$ unprotonated, poised to accept the Ser195 hydroxyl proton from a distance of 2.8 Å (Fig. 3.1a). Substituting a histidine at P_2 on the substrate now allows $N\varepsilon_2$ of the imidazole ring to hydrogen bond to Asp102 at distances of 3.6 and 2.5Å to the carboxylate oxygens, and Nδ1 to accept the Ser195 proton from a distance of 2.9Å (Fig. 3.1b). Yet if the P_2 histidine were to donate the proton back to the leaving group amine as His 57 in trypsin does during the acylation reaction, it would have to rotate about the Cα-Cß bond to move nearly 2 Å. Subsequently, to participate in the deacylation reaction, the histidine would have to move back closer to its original modeled position to accept a proton from water. This would be a large directed movement for a histidine side chain which lacks the structural constraints of His 57 in trypsin.

The $N\delta_1$ from histidine substituted at P_1' can form a hydrogen bond with Ser195 at a distance of 2.6Å, but the Nε2 is 6.7 and 6.4Å from the oxygens of the Asp 102 carboxylate group, too far away to benefit from a potential hydrogen bond (Fig. 3.1c). A histidine at P_1' on the substrate can only participate in the acylation reaction because it is on the leaving group side of the scissile bond. It is not known whether the histidine donates a proton to its own incipient N-terminus or if the proton comes from the solvent. The model shows the P1' His Nδ1 2.7Å from its backbone nitrogen and 65° out of the plane of the imidizole ring, suggesting that it may be able to donate the necessary proton. Presumably, hydroxide ion from water attacks the acyl-

enzyme intermediate to catalyze the deacylation reaction. Modeling of D-histidine at either P2 or P1' resulted in sub-van der Waals contacts with several adjacent residues and distances from the imidazole nitrogens to the Ser195 hydroxyl group that did not allow formation of hydrogen bonds.

Figure 3.1. Modeling substrate-assisted catalysis in trypsin. a) Model of the trypsin active site displaying the catalytic triad and the S1 subsite. The direction of the substrate backbone is indicated with N and C, the oxyanion hole is indicated with the blue backbone nitrogens of 193 and 195 and the red carbonyl oxygen of the P_1 Lys b) Modeled interactions of substrate-assisted catalysis in trypsin H57A showing bound substrate contributing the catalytic histidine from the P2 position. c) Substrate contributing the catalytic histidine from the P1' position.

Figure 3.1: Substrate-assisted Catalysis Models
(a) Trypsin with Bound Substrate

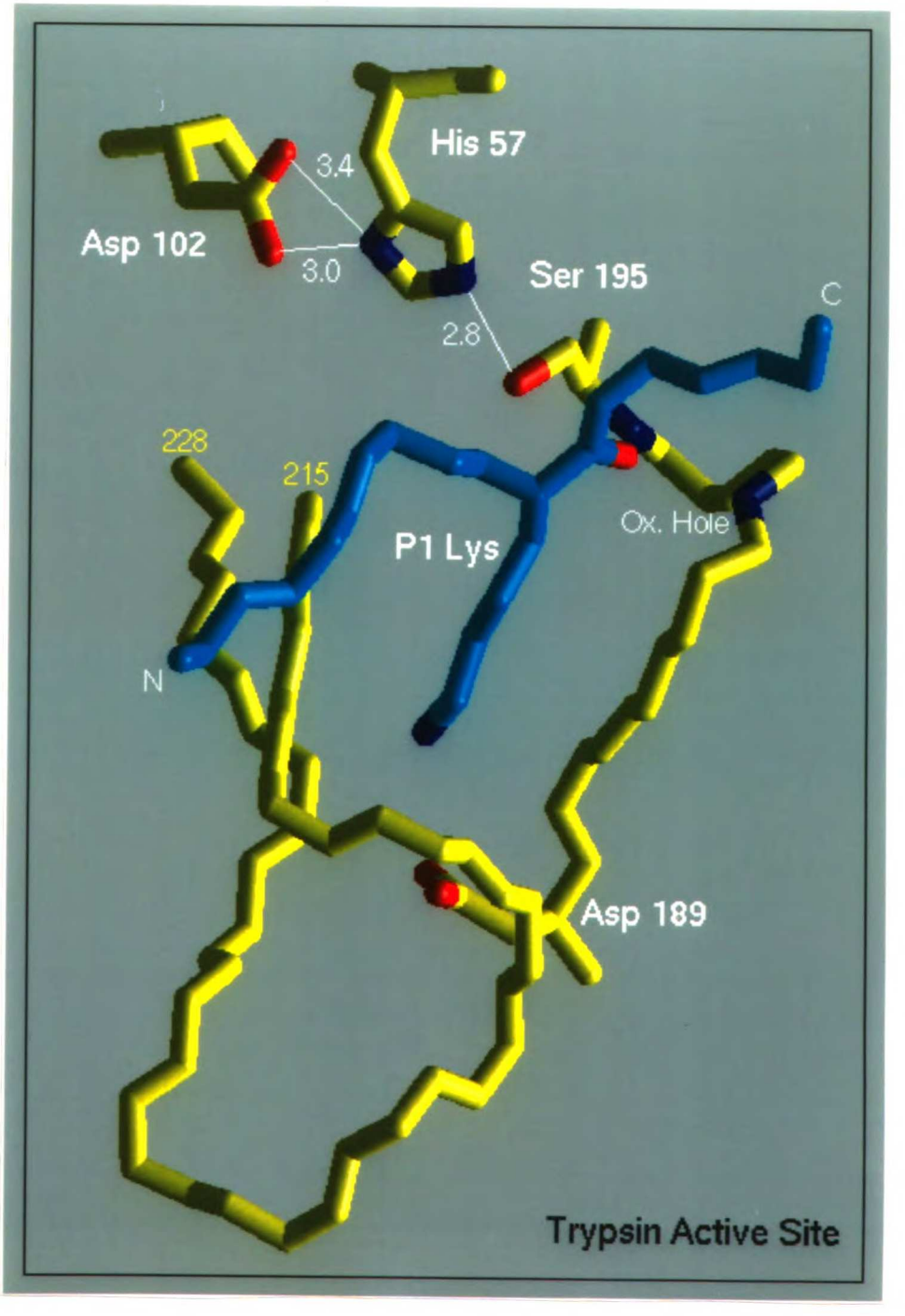


Figure 3.1: Substrate-assisted Catalysis Models
(b) Trypsin H57A with P_2-His substrate

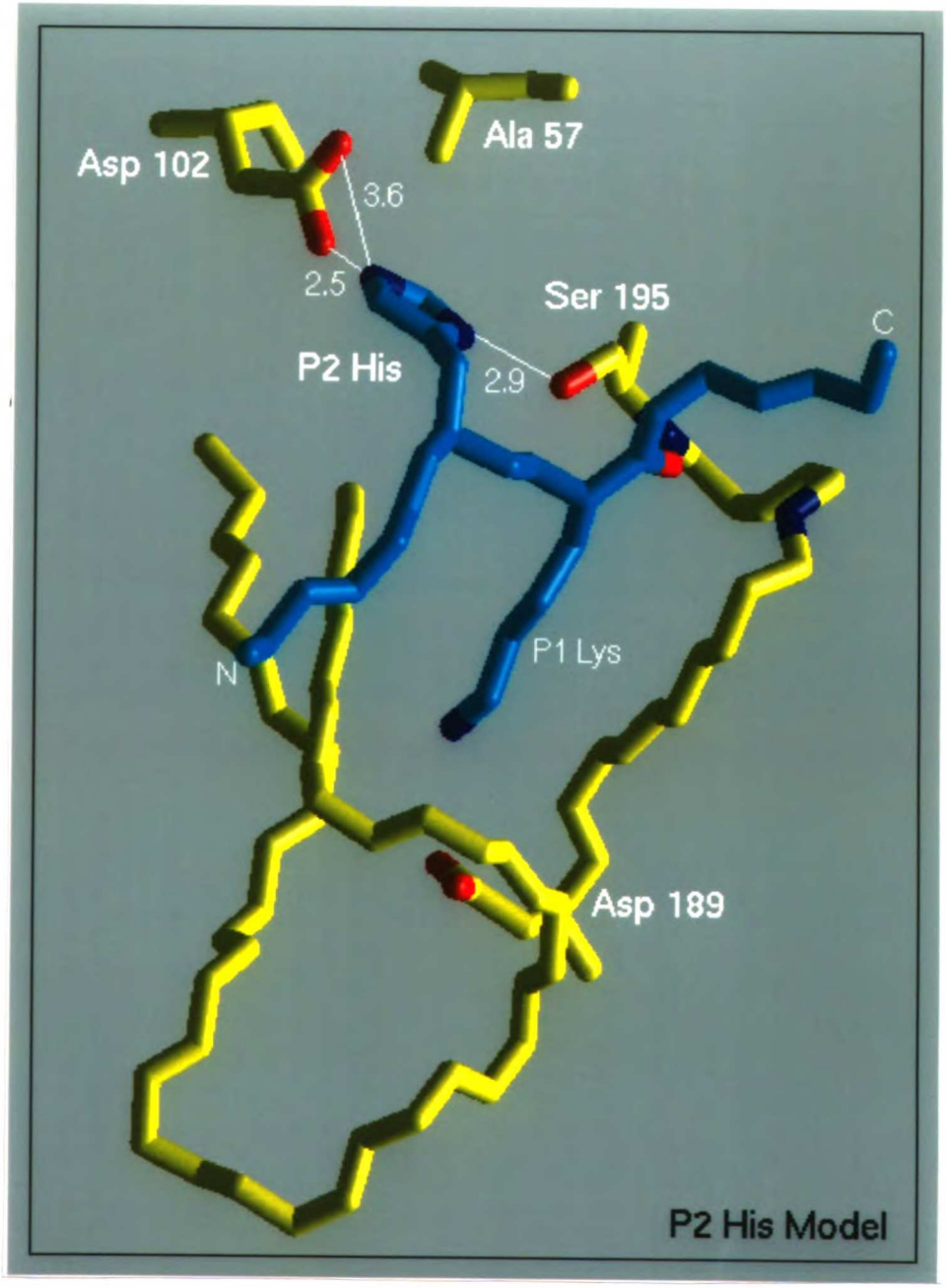
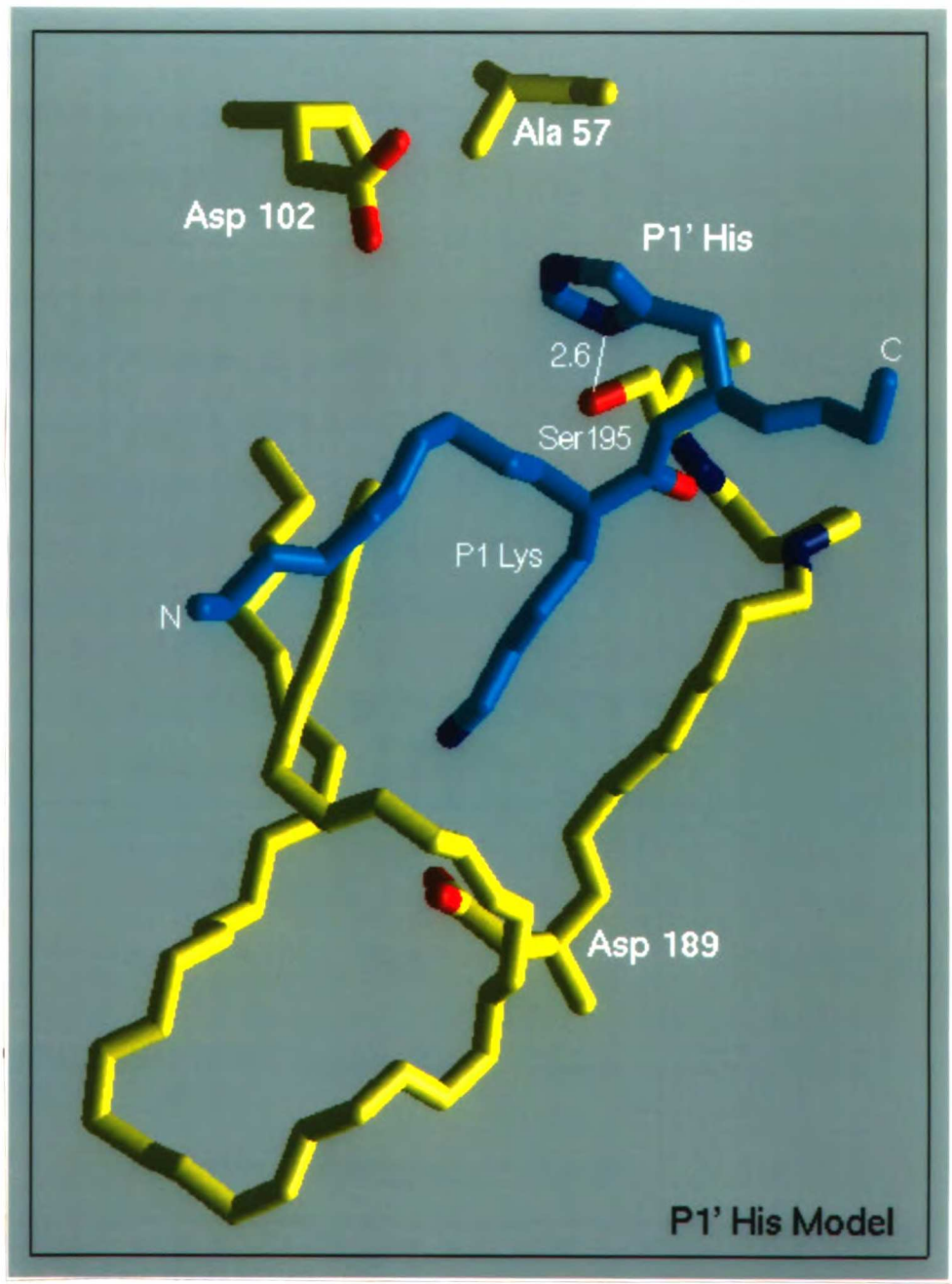


Figure 3.1: Substrate-assisted Catalysis Models
(c) Trypsin H57A with P_1'-His Substrate



Activity of trypsin H57A towards labile amide substrates. The efficiency of substrate-assisted catalysis by trypsin H57A was initially assayed using tetrapeptide (P_4 to P_1) or tripeptide (P_3 to P_1) substrates containing *p*-nitroanilide (pNA) leaving groups (Table 3.1). The substrates were sucAHR-pNA, which contains a histidine at P_2 that could potentially participate in substrate-assisted catalysis, and sucAAPR-pNA, which lacks histidine and would not be expected to participate in catalysis. Trypsin H57A catalyzed the hydrolysis of both substrates much more slowly than did trypsin (Table 3.1), by a factor of 29,500 for sucAAPR-pNA, and 5,300 for sucAHR-pNA. Hydrolysis by trypsin H57A of sucAHR-pNA was accelerated by 2.5-fold relative to the hydrolysis by trypsin H57A of sucAAPR-pNA. This slight acceleration compares to a 2-fold decrease in the activity of trypsin for sucAHR-pNA relative to sucAAPR-pNA.

Table 3.1: Synthetic Substrate Kinetics[a]

Enzyme	Substrate	kcat (s^{-1})	Km (μM)	kcat/Km ($M^{-1}s^{-1}$)	Relative kcat/Km
Trypsin	sucAAPR-pNA	43±1	22±1	$1.9x10^6$	29500
	sucAHR-pNA	36±1	41±1	$0.9x10^6$	13250
Trypsin H57A	sucAAHR-pNA	0.002± .0002	31±2	67	1
	sucAHR-pNA	0.007±0.001	40±3	167	2.5

[a]Substrates were sucAAPR-pNA and sucAHR-pNA. Assays were carried out in 1ml of 100mMTris/100mM NaCl/1mM $CaCl_2$, pH 8.0 at 25°C. [trypsin]=2nM, [trypsin H57A]=500nM. Activity was measured by increase in absorbance at 410nM, and kinetic constants were calculated by fitting the data to the Michaelis-Menten equation.

Activity of trypsin H57A towards peptide substrates extending from P_6 to P_6'. We had shown previously that the catalysis of labile amide linkages by variant trypsins with altered catalytic triads did not always accurately predict the level of catalysis towards extended peptides with normal peptide bonds (Corey & Craik, 1992). Therefore, to further our evaluation of trypsin H57A, a series of extended peptides were synthesized to probe: 1) the extent of catalysis by trypsin H57A towards peptides; 2) the relative importance of Lys or Arg at P_1, and; 3) the possibility that histidine at P_1' of the substrate might also be able to participate in catalysis. Direct comparison of cleavage by trypsin H57A of (I) (Table 3.2), which contained a histidine at P_2, and (II) which contained a proline at P_2, revealed that k_{cat}/K_m for the cleavage of extended peptide substrates was increased 70-fold by the presence of a histidine at P_2. This increase contrasts with a 2-fold decrease in the activity of trypsin upon the alteration of the P_2 proline of (II) to a P_2 histidine in (I). The observed rate enhancement was more dramatic when lysine was located at P_1. Trypsin H57A cleaved the P_2-histidine, P_1-lysine-containing substrate (III) as rapidly as its arginine-containing counterpart (I), while trypsin H57A failed to yield measurable cleavage of peptide (IV) which contained a lysine at P_1 but which lacked histidine.

Table 3.2: Peptide Kinetics of Trypsin H57A[a]

	Trypsin			Trypsin H57A		
Substrate	kcat (s^{-1})	Km (μM)	kcat/Km ($M^{-1}s^{-1}$)	kcat (s^{-1})	Km (μM)	kcat/Km ($M^{-1}s^{-1}$)
(I) YLVG**HR**GFFYDA	16	15	$1.1x10^6$	$4.7x10^{-3}$	47	100
(II) YLVG**PR**GFFYDA	31	12	$2.6x10^6$	$0.05x10^{-3}$	34	1.5
(III) YLVG**HK**GFFYDA	1.2	12	$0.1x10^6$	$5.2x10^{-3}$	46	110
(IV) YLVG**PK**GFFYDA	1.8	63	$0.03x10^6$		nd	
(V) YLVG**HRH**FFYDA	2.2	78	$0.03x10^6$	$21x10^{-3}$	47	466
(VI) YLVG**PRH**FFYDA	2.3	11	$0.22x10^6$	$2.8x10^{-3}$	25	110
(VII) GGSGPF**GRH**ALVPE	7.5	3.9	$1.9x10^6$	$13x10^{-3}$	38	350
(VIII) YLVG**PRGH**FYDA	5.0	33	$0.15x10^6$	$0.12x10^{-3}$	130	0.9
(IX) YLVG**(D)HR**GFFYDA		na			na	
(X) YLVG**PR(D)H**FFYDA		na			na	

[a]The error in these determinations was ±20% for kcat and Km. nd-not determined, activity observed was too low over a 24 hr period to calculate accurate kinetic constants. na-not active, no hydrolytic activity observed over 24 hr period.

It had previously been shown that subtilisin H64A exhibited an enhanced cleavage of peptides containing histidine at P_1' (Matthews & Wells, 1993). To assay whether trypsin H57A would also possess this activity, peptides (V) and (VI) were designed to contain histidine at either P_1' and P_2 or at P_1' alone. Assays towards peptide (VI), which contains a histidine at P_1', revealed an 80-fold rate enhancement relative to peptide (II) which contains a glycine a P_1', an acceleration which was similar to that observed for hydrolysis by trypsin H57A of substrate (I) which contained a P_2 histidine. Substrate (V) contained histidine at both P_1' and P_2 and was a 4-5 fold better substrate in terms of k_{cat}/K_m than either of the single-histidine containing substrates (I) or (VI). The positions from which substrate imidazole rings may substitute for the histidine of the catalytic triad is limited, however, as substrates (IX) and (X) containing D-histidine at either the P_2 or the P_1' position were not hydrolyzed by trypsin H57A. This is consistent with the modeling results showing that the D-histidine residues are unable to adopt a catalytically active conformation. Similarly, no rate enhancement was observed for the hydrolysis by trypsin H57A of (VIII), a substrate containing L-histidine at P_2'. Efficient hydrolysis could also be achieved using substrates containing optimized subsite interactions. Peptide (VII), which contained histidine at P_2 and whose P_4, P_3, and P_1'-P_4' residues had been identified within a particularly labile trypsin substrate (Ding et al, 1995), was processed by trypsin H57A with a kcat/Km which was 3-4 fold greater than for the other single histidine-containing substrates (I) or (IV).

Cleavage of protein substrates. To examine whether trypsin H57A could cleave a folded protein with complete specificity, trypsin H57A-catalyzed

hydrolysis of ornithine decarboxylase from *Leishmania donovani* (LODC) was analyzed. LODC is a 708 amino acid protein containing thirty-seven arginines and eleven lysines (Hanson et al., 1992). There is an HRH site at amino acids 246-248 and an RH site at amino acids 325-326. Treatment of LODC with trypsin H57A yielded two fragments of 247 and 461 amino acids in length (Fig. 3.2, lanes 4-7). No other products were evident after incubation for 42 hours (lane 7). N-terminal amino acid sequence analysis confirmed that cleavage was occurring at the HRH site between R247 and H248. Trypsin H57A did not cleave at the secondary RH site at residues 325-326, emphasizing that the structure of the protein substrate may obscure particular cleavage sequences. The HRH site occurs at a junction between the C-terminal region common to all parasite decarboxylases and an N-terminal extension which has only been observed in LODC (Hanson et al., 1992) and is presumably more solvent-exposed than the R325-H326 site. The initial rate of cleavage was examined as a function of substrate concentration to determine the kinetic constants for the specific hydrolysis of LODC by trypsin H57A. The turnover number, $k_{cat,}$ was 2.0×10^{-4} sec^{-1}, the binding constant, $K_{m,}$ was 2.4 μM, and k_{cat}/K_m was 83 $M^{-1}s^{-1}$. Limited proteolysis of LODC by trypsin also afforded the 461 amino acid product, but there were many other proteolysis products as well, and the 247 amino acid product was not present. Further proteolysis by trypsin resulted in the total degradation of LODC, whereas extended digestions with trypsin H57A yielded complete digestion at the target site with no appearance of secondary hydrolysis.

Figure 3.2: Cleavage of LODC by Trypsin H57A

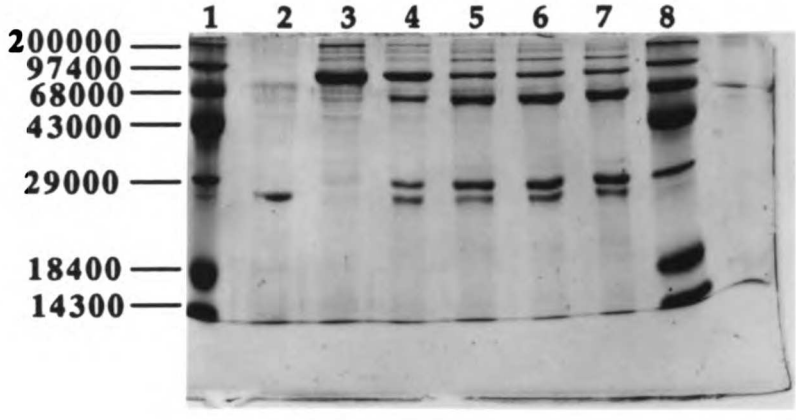


Figure 3.2. Cleavage of LODC by trypsin H57A. Cleavage reactions contained trypsin H57A (1.7μM) and LODC (3.2 μM). Reactions were performed as described in materials and methods. Lane 1, molecular weight markers; Lane 2, trypsin H57A. Lane 3, LODC; Lane 4-7 incubation of LODC with trypsin H57A for 4, 17, 27, and 42 hours.

Trypsin H57A was also used to cleave the pro-peptide of a variant trypsinogen. Three mutations were introduced into trypsinogen so that it could be used as a substrate for trypsin H57A (Fig. 3.3). The D189H mutation was introduced into the primary specificity pocket so that, upon activation, the resultant trypsin would be inactive, thereby removing the possibility of autoactivation and subsequent autolysis products. We have previously shown that trypsin D189H is 250-fold reduced in k_{cat}/K_m on synthetic lysine-containing ester substrates, with no detectable amidase activity (Willett et al., 1995). Two other mutations, D15H and K16R, were introduced at the propeptide-enzyme junction to create a substrate sequence for trypsin H57A (DDDHR-IVGG vs. DDDDK-IVGG in trypsinogen). Trypsin H57A cleaved the variant pro-peptide with complete specificity with a kcat/Km of 1.9 $M^{-1}s^{-1}$. This was ~100-fold slower than enteropeptidase cleaved trypsinogen as judged from Coomassie-stained gels of reaction products. Although there are several Lys and Arg residues located on surface loops of trypsin and one other KH sequence at position 90-91, there was no non-selective cleavage observed by trypsin H57A of either the wild-type or the variant trypsinogen (Fig. 3.3). In contrast to the stringent specificity of trypsin H57A, the more relaxed specificity of enteropeptidase permitted cleavage of the variant trypsinogen propeptide (D_3HR) as well as the wild-type propeptide (D_4K) with comparable efficiency. Trypsin also cleaved the mutant trypsinogen, but yielded a ladder of unwanted proteolytic by-products (data not shown).

Figure 3.3: Cleavage of HR-trypsinogen by Trypsin H57A

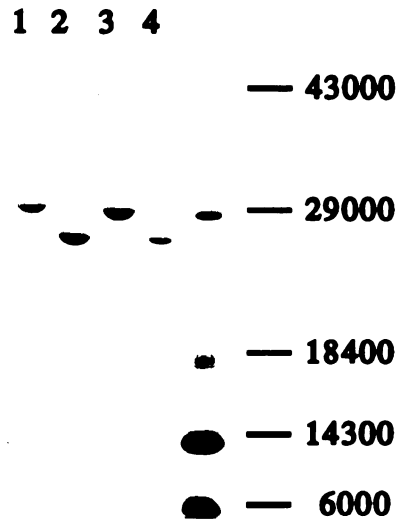


Figure 3.3. Coomassie-stained SDS-PAGE analysis of trypsin H57A cleavage of DDDHR trypsinogen. Lane 1, trypsinogen D189H; lane 2, trypsinogen D189H cleaved by enteropeptidase; lane 3, trypsinogen D14H/K15R/D189H; lane 4, trypsinogen D14H/K15R/D189H cleaved by trypsin H57A; right lane, low molecular weight markers.

Cleavage of purification-tag fusion proteins. A practical use of trypsin H57A is to specifically cleave purification tags from their fusion partners, thereby liberating the mature protein of interest for activity and/or structural studies. Another test case for trypsin H57A specificity and activity came as a fusion protein of a plant ethylene receptor response region (ETR-RR). A His_{10} affinity purification leader was fused to the C-terminal third (~15kDa) of the cytoplasmic domain of this transmembrane protein and expressed using the T7 promoter system supplied in a kit from Invitrogen (A. Grantz & S. H. Kim, personal communication). This experiment proved to be a case of serendipity as well as a control for trypsin H57A activity. There is an HK dyad at amino acid 50 out of 136 that separates two phosphorylation sites in this fragment. The goal of the experiment was to cleave the protein at this HK dyad in order to study the two phosphorylation sites independently. Cleavage of the protein at this site would have yielded a 10kDa and a 5kDa product (Fig. 3.4), yet upon exhaustive incubation of the protein with trypsin H57A, only one product was visible on a coomassie-stained SDS gel at about 14kDa. Closer inspection of a cloning artifact explained the unexpected cleavage. The expression vector contained a sequence encoding a His_{10} peptide fused to a factor Xa cleavage site, which was in turn fused to the N-terminus of the ETR-RR fragment:

H_2N-Met-$(His)_{10}$-Ile-Glu-Gly-Arg—His-Met-ETR-RR. . .

Figure 3.4: ETR-RR Cleavage by Trypsin H57A

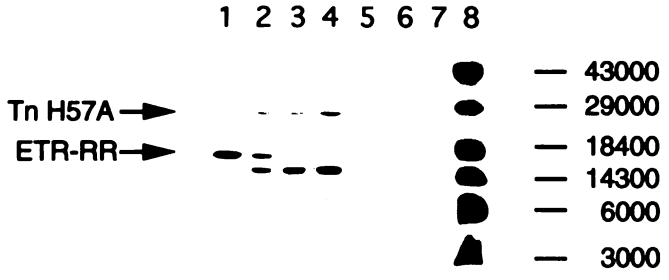


Figure 3.4: Cleavage of ETR-RR by trypsin H57A. Reaction conditions were: 100mM tris, pH8, [Tn H57A]= 500nM, or [trypsin]=50nM, [ETR-RR]=5μM, T=37°C. Lane 1: uncut ETR-RR, lane2: 30min incubation with Tn H57A, lane 3: 4hrs with Tn H57A, lane 4: 8hrs with Tn H57A, lanes 5-7: same time points, except wild-type trypsin was used.

The cloning artifact resulted from use of an Nde I site (CATATG) to insert the ETR-RR fragment into the vector. This fortuitously placed a histidine directly C-terminal to the arginine of the factor Xa site, thus creating a trypsin H57A cleavage site. The observed cleavage is consistent with this explanation. The intended cleavage at the internal HK site was not seen, even in 4M urea with 10mM DTT present, indicating that this site is not completely solvent-accessible. However, cleavage at the fusion protein junction was entirely specific and complete after two hours incubation. Cleavage with trypsin yielded almost complete degradation of substrate within 30 minutes, demonstrating the extreme specificity of trypsin H57A. Another test case was created using a $(His)_7$ tag fused to the N-terminus of HIV integrase. Specific cleavage was observed (data not shown), although at a much (>10-fold) slower rate than with ETR-RR as substrate. These experiments demonstrate the practicality of trypsin H57A as a processing enzyme, along with the idea that optimization of substrate is a worthwhile goal when engineering site-specific proteases.

Discussion

A goal of protein engineering is the creation of variant proteins which afford a useful perspective on the structure and function of the wild-type protein and on related enzymes. A test of whether this perspective has been obtained is the ability to alter function in a prescribed fashion. If successful, these alterations might also yield catalysts that possess pactical value as experimental tools or therapeutic agents. To be of practical value, modified enzymes must combine a turnover rate that is sufficient for a given experimental purpose with a high degree of specificity. Engineered proteases

which possess these properties would have a number of applications. Novel proteases with altered specificities could supplement existing proteases in the mapping of proteins much as restriction endonucleases with eight base recognition sequences complement those which recognize six base sites for the mapping of nucleic acids. Modified proteases could also be used to cleave folded proteins at naturally occurring cleavage sites to generate desired truncated products or to isolate protein domains. Alternatively, recombinant fusion proteins can be cleaved at engineered sites to liberate the desired native protein. Eventually these principles will allow us to create therapeutic proteases capable of cleaving target sites *in vivo*.

Site-directed mutagenesis was used to replace the catalytic histidine of trypsin with alanine to produce trypsin H57A. The ability of this variant to participate in substrate-assisted catalysis was examined using synthetic amide, extended peptide, and protein substrates. Catalysis by trypsin H57A towards synthetic substrates containing labile amide bonds showed a 2.5-fold increase in k_{cat}/K_m when histidine was present at P2. This differential reactivity was much less than had been observed earlier for subtilisin H64A, where the introduction of histidine at P2 resulted in a 200-fold increase in k_{cat}/K_m. Activity of trypsin H57A was then examined using a series of extended peptide substrates containing histidine at various subsites relative to lysine or arginine at P_1. These assays revealed a greater than 70-fold increase in k_{cat}/K_m for the hydrolysis of peptides containing histidine at P2 or P1' by trypsin H57A. The increase in catalysis caused by the presence of a P_2 or P_1' histidine in the substrate resulted primarily from an increased k_{cat},

supporting the hypothesis that histidine enhances hydrolysis through direct participation in the catalytic mechanism.

The efficiency of substrate-assisted catalysis could be increased either by the simultaneous inclusion of P_2 and P_1' histidines in the substrate or by using a substrate whose subsite occupancy is known to yield efficient catalysis by trypsin. The k_{cat}/K_m of trypsin H57A towards a peptide which contained histidine at P_2 and P_1' was only 55-fold lower than that of trypsin on the same peptide, an impressive result for a variant trypsin which lacks a member of the catalytic triad. This shows that substrate specificity can be varied without a dramatic reduction in catalytic efficiency. The improved rate enhancement observed with this peptide substrate may have been due to the ability of the two histidines in the substrate to increase the effective concentration of the general base required for catalysis. The restoration of activity to both trypsin H57A and subtilisin H64A through the presence of histidine at either the P_2 or the P_1' position of substrate is clear evidence for the diversity of catalytically competent orientations that are possible for histidine to assume. The highest level of catalysis observed by trypsin H57A on a peptide with one histidine was achieved with GGSPGPFGRHALVPE. This peptide contains a histidine at P_1' and amino acids that form favorable subsite interactions yielding a low K_m and high k_{cat}/K_m (Ding et al., 1995). This shows that it is possible to optimize the rate of substrate cleavage through the selection of labile substrate sequences.

A possible reason for the low level of activity by trypsin H57A on synthetic amide substrates relative to peptide substrates is the absence of extended subsite interactions C-terminal to the sissile bond. The peptide substrate

occupies these subsites allowing the binding energy to be translated into increased catalysis. For pNA substrates, the S_1' subsite of the protease is not filled by an amino acid but by a chromogenic or fluorogenic leaving group. The other prime sites, S_2', S_3', S_4' etc., are not filled at all, so that the binding energy contributed by the prime-side residues is not available for catalysis. Hence, exclusive use of synthetic substrates may mask increased activity of a variant trypsin towards peptide bonds (Corey & Craik, 1992). The binding energy contributed by the P' residues may induce a conformational change in the enzyme necessary for efficient catalysis. For example, trypsin D102S/S214D, which relocates the negative charge of the catalytic triad, exhibits only a 10-fold increase in k_{cat} relative to the D102S variant for the hydrolysis of labile synthetic substrates, but exhibits a 7200-fold increase in k_{cat} for the hydrolysis of peptide linkages relative to trypsin D102S (Corey & Craik, 1992, Corey & Craik, 1993). This effect is consistent with the increase in serine protease activity observed towards peptide substrates as additional P' amino acid residues are present (Bauer et al., 1981, Bizzozero & Dutler, 1987) or as they are varied (Schellenberger et al., 1993).

Just as the hydrolysis of labile synthetic amide linkages by proteases does not always reflect their ability to hydrolyze the unactivated amide linkages of peptides, the hydrolysis of peptides may not always reflect the capacity of a protease to cleave folded proteins. Folded secondary and tertiary structure can obstruct potential cleavage sites and can prevent protein hydrolysis. Trypsin H57A hydrolyzed an internal amide linkage in ornithine decarboxylase from *Leishmania donovoni* (LODC) and a variant propeptide from recombinant trypsinogen. Measurement of the kinetic constants for the cleavage of LODC

by trypsin H57A revealed that both K_m and k_{cat} were lower than had been observed for the cleavage of peptide substrates. The measured k_{cat}/K_m was 82 $M^{-1}s^{-1}$, a rate which is quite slow relative to cleavage of peptide substrates by wild-type trypsin. Cleavage by trypsin H57A of the propeptide from the recombinant trypsinogen occurred with a k_{cat}/K_m of 1.9 $M^{-1}s^{-1}$, at least two orders of magnitude down from enteropeptidase cleavage of the propeptide. Whereas enteropeptidase has evolved to cleave the D_4K sequence, trypsin is a poor enzyme against this substrate, probably due to the negative charges on the P_2-P_5 residues.

However, even this low rate of catalysis permits practical applications of the variant enzyme given a sufficient supply of protease. Indeed, the more critical consideration in utilizing a selective protease is its discrimination between targeted and untargeted catalysis during the desired period required to process substrate. After extended incubation, the only observed hydrolysis of LODC by trypsin H57A is at the target HRH at position 246-248. A similar level of discrimination was also observed for cleavage of the propeptide of trypsinogen. Potential trypsin H57A cleavage sites HR↓, HK↓, R↓H, and K↓H are not uncommon in proteins. However, many of these sites will occur within regions which are inaccessible to proteolysis when the protein is folded, so that sequences which are susceptible to trypsin H57A will be relatively rare. As has been done with factor Xa (Sahin-Tóth et al., 1995), trypsin H57A can be used to map sites that are proteolytically labile, and therefore, solvent accessible. Trypsin H57A has the advantage over factor Xa in that it requires less significant alteration in the target protein to create a cleavage site.

Substrate-assisted catalysis by serine proteases has now been shown for members of both the chymotrypsin and subtilisin families in spite of their structural divergence, and may prove to be a general property accessible to many serine proteases. Further structural and kinetic studies are being performed to characterize the substrate/enzyme interactions involved so that activity may be increased through additional mutagenesis. Subsequent modifications may be used not only to optimize the efficiency, but also to refine the selectivity of this enzyme. It should be possible to extend substrate-assisted catalysis to serine proteases with chymotrypsin or elastase specificities for large hydrophobic or small hydrophobic amino acid sidechains at P_1. It may also be possible to narrow the existing specificities of selective proteases such as tissue plasminogen activator or thrombin, to create unique selectivities for engineered or native protein substrates. Such control of specificity through the application of protein engineering principles represents an important step towards the rational design of proteases capable of recognizing specific target proteins.

Chapter 4

Engineered Metal-Assisted Specificity in Trypsin

Abstract

Histidine substrate specificity has been engineered into trypsin by creating metal binding sites for Ni^{+2} and Zn^{+2} ions. The sites bridge the substrate and enzyme on the leaving-group side of the scissile bond. Application of simple steric and geometric criteria to a crystallographically-derived enzyme-substrate model suggested that histidine specificity at the P2' position might be acheived by a tridentate site involving amino acid residues 143 and 151 of trypsin. Trypsin N143H/E151H hydrolyzes a P2'-His-containing peptide (AGPYAHSS) exclusively in the presence of nickel or zinc with a high level of catalytic efficiency. Since cleavage following the tyrosine residue is normally highly disfavored by trypsin, this result demonstrates that a metal cofactor can be used to modulate specificity in a designed fashion. The same geometric criteria applied in the primary S1 binding pocket suggested that the single-site mutation D189H might effect metal-dependent His specificity in trypsin. However, kinetic and crystallographic analysis of this variant showed that the design was unsuccessful because His189 rotates away from substrate causing a large perturbation in adjacent surface loops. This observation suggests that the reason specificity modification at the trypsin S1 site requires extensive mutagenesis is because the pocket cannot deform locally to accommodate alternate P1 side chains. By taking advantage of the extended subsites, an alternate substrate specificity has been engineered into trypsin. This chapter has been published previously [Willett et al., (1995), *Biochemistry* **34,** 2172-80].

Introduction

Metal binding sites in proteins serve a variety of functions, including structural stability, catalysis, folding, and molecular recognition. Nearly one third of all known proteins require metal ions for their structure or function (Ibers and Holm, 1980). This indicates that a thorough understanding of the chemistry of protein-metal interactions will yield increased insight into many physiological processes as well as provide a basis for protein engineering efforts. The wealth of information about metalloproteins has allowed the design of simple metal binding sites for specified functions (Higaki, et al., 1992). These sites have been incorporated into recombinant proteins to aid in purification, to enhance structural stability, to facilitate structure determination by X-ray crystallography, and to regulate activity (Tainer, et al., 1991). For example, metal binding sites have been incorporated successfully into *Staphylococcal* nuclease (Corey and Schultz, 1989) and trypsin (Higaki, et al., 1990) to reversibly inhibit these enzymes. A calcium binding site was introduced into subtilisin to stabilize the enzyme against thermal denaturation (Braxton and Wells, 1992), and a nickel binding site incorporated into glycogen phosphorylase allowed insight into the mechanism of allosteric regulation (Browner, et al., 1994). These reports indicate that enzymes can be engineered to bind metal ions at designed sites, providing a powerful new tool to address structure-function relationships and for redesigning enzymatic activity.

Trypsin functions well as a model system to test the design of metal binding sites because it has an established expression/purification system, well-defined substrates for kinetic analysis by continuous and discontinuous assays, and reproducible crystallization conditions for X-ray structure analysis of the modified proteins. Trypsin specificity at the P_1 substrate position is exclusively

toward Arg and Lys side chains for amide hydrolysis (Evnin, et al., 1990), while there is little preference at position P_2' (Schellenberger, et al., 1994). Attempts to modify the P_1 substrate preference with a small number of amino acid substitutions has met with little success (Graf, et al., 1987, Sprang, et al., 1988); however, a specificity conversion to large hydrophobic residues was effected by extensive exchange of amino acids for those of chymotrypsin (Hedstrom, et al., 1994, Hedstrom, et al., 1992). As an alternative way to alter substrate specificity of trypsin, the use of transition metal binding sites has been explored. Using this approach, subsites on the enzyme were engineered to bind a metal ion which bridges to a metal-binding amino acid within the substrate. The concept is illustrated in Fig. 4.1 using histidines introduced into trypsin to create metal-dependent P_2'-His specificity. This chapter explores the structural environments of both the S_1 and S_2' sites of trypsin as potentially suitable targets for the introduction of metal-dependent specificity. We describe the successful design, production, and kinetic analysis of a trypsin variant exhibiting metal-dependent histidine specificity.

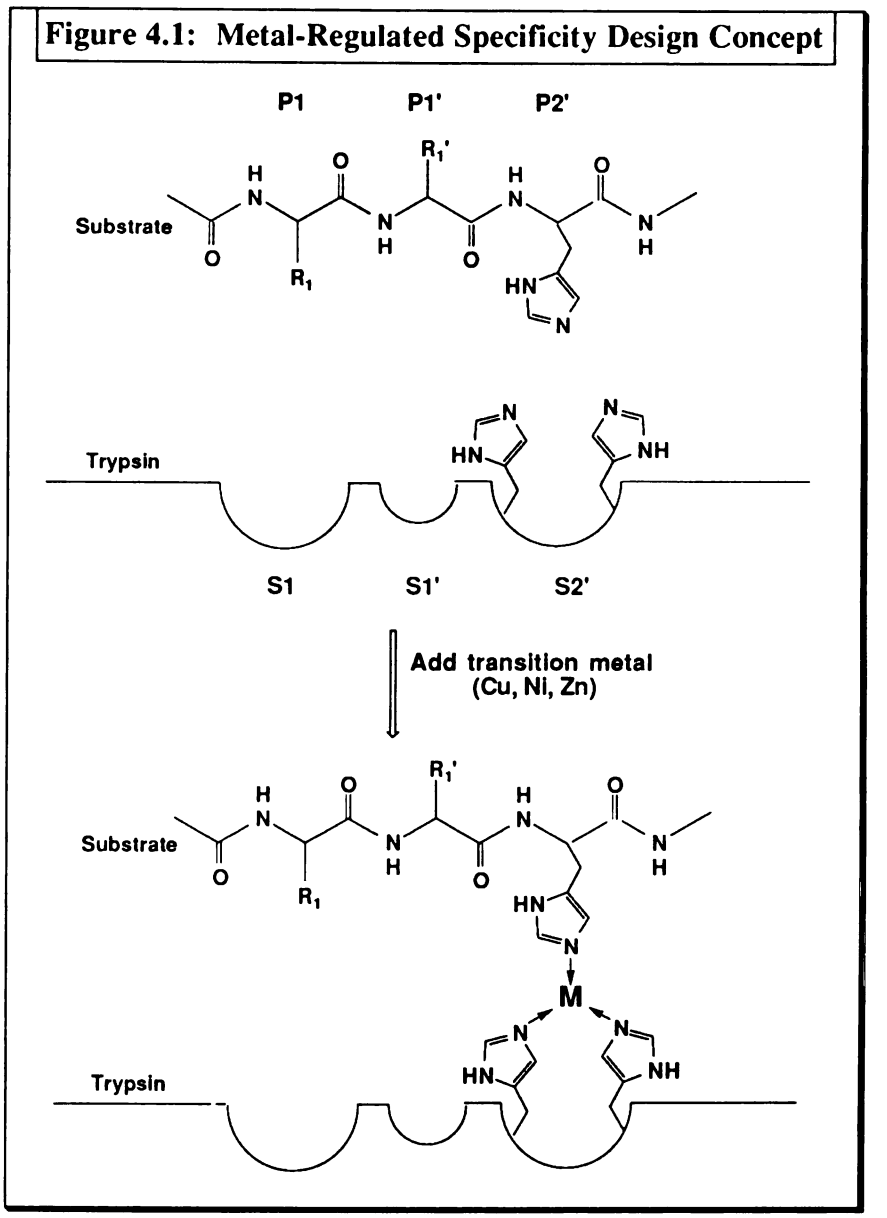


Figure 4.1: Schematic representation at position P_2' of the trypsin-substrate complex showing enzyme subsites (S_1-S_2') and substrate residues (P_1-P_2') according to the nomenclature of Schechter and Berger (Schechter and Berger, 1968). The sissile bond is between residues P_1 and P_1', and the primary binding determinant is the P_2' residue This concept involves an engineered metal binding site on the enzyme allowing a metal ion to bridge enzyme and substrate so that substrate binds and is cleaved only in the presence of metal ion.

Materials and Methods

Materials *Saccharomyces cerevisiae* strain DLM101α [*Mat a, leu 2-3,-112 his2 3-11,-15 can1, ura 3Δ, pep4Δ [cir°]DM23*] was obtained from Dr. L. Hedstrom. *Escherichia coli* strain X90 (F' *lac I^q, lac ZY, pro AB/ Δ-(lac-pro), ara, mal A, argEam, thi, rif^r*)was obtained from Dr. A Vershon. All restriction enzymes, T_4 DNA ligase, and T_4 DNA polymerase were purchased from New England Biolabs, Inc., and used according to manufacturer's recommendations. Enteropeptidase (EK2) was purchased from Biozyme, Inc. SP Sepharose Fast Flow resin was purchased from Pharmacia. Benzamidine agarose was purchased from Pierce, and Affigel 10 was purchased from Bio Rad. BPTI was purchased from Merck (Trasylol) and coupled to the Affigel beads according to the manufacturer's protocol. Nα-Benzyloxycabonyl-L-glycylprolylarginine 7-amino-4-methylcoumarin (Z-GPR-AMC) was purchased from Bachem Biosciences, Inc. The peptide substrates AGPYAHSS and AGPYAASS were purchased from BioServ Labs (San Jose, CA). Oligonucleotides were synthesized on an Applied Biosystems 380B DNA synthesizer (Foster City, CA).

Computer Modeling. Modeling was performed using the InsightII program (BioSym Technologies, Inc) displaying the crystal structure of rat anionic trypsin mutant D189S complexed to BPTI on a Silicon Graphics workstation (Perona, et al., 1994a). Potential metal binding sites utilizing histidines were modeled according to the following criteria: (1) The side chains were solvent-accessible for exogenous metal binding; (2) the distances between the α-carbons of the unsubstituted side chains were less than 13Å to allow imidazole side chains to achieve optimal distances of about 2Å for metal chelation; (3) the His side chains possessed one of five favorable rotamers (Ponder and Richards, 1987); (4) all atoms of the substituted His side chains lay further than 3.0Å from other atoms

of the protein unless a hydrogen bond could be formed. Subroutines from the program InsightII were used to perform the alignments and torsional angle adjustments of the side chains (Higaki, et al., 1990).

Modeling in S_1 was performed interactively using InsightII. Asp189 and the P_1 lysine of BPTI were replaced by histidines, and moved until a suitable coordination geometry was found. Second generation variants in S_1 were designed using the FRODO (Jones, 1978) and XPLOR (Brunger, et al., 1987) computer packages.

Mutagenesis and Expression . Mutagenesis was performed by the method of Kunkel (Kunkel, 1985) on a bluescript-based plasmid (pST) containing the gene for rat anionic trypsinogen fused to the yeast α-factor leader and ADH/GAPDH promoter sequences (Hedstrom, et al., 1992). The BamHI/SalI fragment of the resulting trypsin variant was subcloned into the corresponding vector fragment of the yeast expression vector (pYT). Yeast was transformed by electroporation in 2mm cells using a BioRad Gene Pulser set at 600V, 200Ω, and 25μF. Transformed cells were selected on uracil deficient plates, and grown at 30°C for two days. Colonies were restreaked onto leucine-deficient plates and grown for 2 days at 30°C to increase plasmid copy number (Erhart and Hollenberg, 1983). Colonies from the Leu- plates were grown in Leu- minimal media containing 8% glucose for two days (Gardell, et al., 1988). Trypsinogen was expressed and secreted into the media by inoculating a one liter flask containing 1% yeast extract/2% bactopeptone/2% glucose, and allowing growth at 30°C for 48-72 hours with shaking (250 rpm in New Brunswick shaker). For a detailed protocol on trypsinogen expression, see Appendix 1.

Purification . Secreted trypsinogen was isolated by centrifugation (10 min @5000xG) to remove cells, and the resulting supernatant was adjusted to pH 3.0 by addition of glycine-HCl to 50mM. This solution was recentrifuged (10 min @5000xG), loaded onto an SP-sepharose fast flow column, and trypsinogen was eluted with a 0-1.25M NaCl gradient. Fractions containing trypsinogen as determined by SDS-PAGE were pooled and dialyzed against 10mM MES/1mM $CaCl_2$, pH6 at 4°C. Enteropeptidase (EK2) dissolved in 10mM MES, pH6 at 1mg/ml was added to the dialyzed trypsinogen pool at a ratio of ~1:100 EK2 to trypsinogen. Activation was monitored by SDS-PAGE. This activation mixture was incubated at 37°C for 2-4 hours to cleave the pro-peptide and liberate mature trypsin, then loaded onto an appropriate affinity column equilibrated in 10mM MES/1mM $CaCl_2$, pH6. Variants containing an intact P_1 pocket were purified over benzamidine agarose, whereas variants containing the D189H mutation were purified via BPTI-coupled agarose to utilize the extended binding contacts necessary for purification of trypsins containing an altered P_1 pocket. The affinity column (either benzamidine or BPTI) was washed with 20 column-volumes of equilibration buffer containing 0.5M NaCl to remove unactivated trypsinogen and trypsin degradation products. Pure trypsin was step-eluted with 0.1M HCO_2H/0.5M NaCl/1mM $CaCl_2$. Fractions containing trypsin were diafiltered using a Centricon 10 concentrator (Amicon) into 1mM HCl/1mM $CaCl_2$ at pH3 for storage. All purified trypsins were stored under these conditions at 4°C. For a detailed trypsin purification protocol, see Appendix 2.

Kinetic parameter determination. Kinetic assays of all trypsin variants were performed in 5mM PIPPS/1mM Tris/100mM NaCl/ 1mM $CaCl_2$, pH8. This buffer was chosen for low metal binding capacity, with Tris functioning as a carrier for metal ions (Bai and Martell, 1969). The fluorogenic substrate Z-GPR-

AMC was used to test activity against a typical amide trypsin substrate, and Z-Lys-thiobenzyl ester was used to detect low levels of activity.

S_2' variants. An octapeptide substrate, AGPYAHSS (Bioserv, Inc., San Jose, CA), was used to observe specificity at P_2'. This allowed us to measure small kinetic effects at P_2' because of the very slow turnover of P_1-Tyr by trypsin. A discontinuous peptide cleavage assay was performed as follows: 2.5µl of 40µM enzyme in 1mM HCl/1mM$CaCl_2$ was added to 87.5µl of assay buffer (5mM PIPPS/1mM Tris/100mM NaCl/1mM $CaCl_2$, pH8) containing either 228µM $CuCl_2$, 228µM $NiCl_2$, 228µM $ZnCl_2$, or 0.5mM EDTA as a control. Ten microliters of peptide substrate dissolved in assay buffer were added to the enzyme mixture, and the reaction was allowed to proceed at 37°C for 2 h. The final enzyme concentration was 1µM, metal concentration was 200µM, and substrate concentration ranged from 200µM to 2mM. The reaction was stopped by diluting 10µl of reaction mix into 140µl of 0.1%TFA containing 0.001% dimethylformamide as an internal standard. Cleavage products were analyzed by RP-HPLC on a C_{18} guard column (Perkin Elmer) with a 0-30% acetonitrile/0.1%TFA gradient over 2.5 minutes. Absorbance was measured at 220nM. Quantitation of product peak areas was performed by the Rainin HPLC Dynamax program. Initial rates were calculated and plotted vs. substrate concentration, and the curve was fitted to the Michaelis-Menton equation using Kaleidograph on a Macintosh computer.

S_1 variants. Variants designed to exhibit metal-dependent specificity at P_1 were tested against 200µM suc-AAPH-pNA, suc-AAPD-pNA, and suc-AAPE-pNA in the presence and absence of 100µM $CuCl_2$, $NiCl_2$, or $ZnCl_2$. Enzyme

concentration was 100nM. These enzymes were also tested against Z-Lys-thiobenzyl ester at an enzyme concentration of 2μM.

Crystallography. A 1:1 stoichiometric ratio of Tn D189H:bovine pancreatic trypsin inhibitor (BPTI) crystallized after 4-7 days at 4°C in hanging drops (McPherson, 1989) at a final concentration of 10mg/mL. The well solutions contained 19% PEG 4K, 0.2M ammonium acetate, 0.1M sodium citrate (pH 6.0-6.25); equal volumes (1-3 μl) of the protein and well solutions were mixed. The two crystals used for data collection were both approximately 0.8x0.25x0.25 mm in size with space group $P3_221$ and cell dimensions a=92.8Å, c=62.2Å. Diffraction data were recorded to 1.6Å on an R-axis image plate system (Sato, et al., 1992). Ninety six frames of data were collected with a scan angle of 1.3-1.4° per frame. The recording time was 30 minutes per frame. The data were reduced using the R-axis software package (Sato, et al., 1992). Phases were determined using the structure of trypsin D189S complexed with BPTI (Perona, et al., 1994a) as a starting model from which the side chains of Lys15 of BPTI, Ser189, Ile183, Gly158, Val199, Gly183, Tyr228, Ser190, and all waters in the P_1 pocket were deleted. The initial R factor prior to refinement was 36%. Positional and B-factor refinement was accomplished with XPLOR (Brunger, et al., 1987) and was alternated with model building into electron density maps computed from coefficients ($2F_O$-F_C) and (F_O-F_C) using FRODO (Jones, 1978) running on an Evans and Sutherland PS390 graphics workstation. This completed model has a cumulative R_{cryst} of 18.4% for all data with $I/\sigma(I) > 1$ in the resolution range 6.0-1.8Å (Table 4.3). An omit map was also calculated following deletion of residues 189 and 217-226 and subsequent refinement. This map confirmed that this region of the structure had been correctly rebuilt (Fig. 4.7c).

Figure 4.2: Model of Metal Binding Site at S_2'

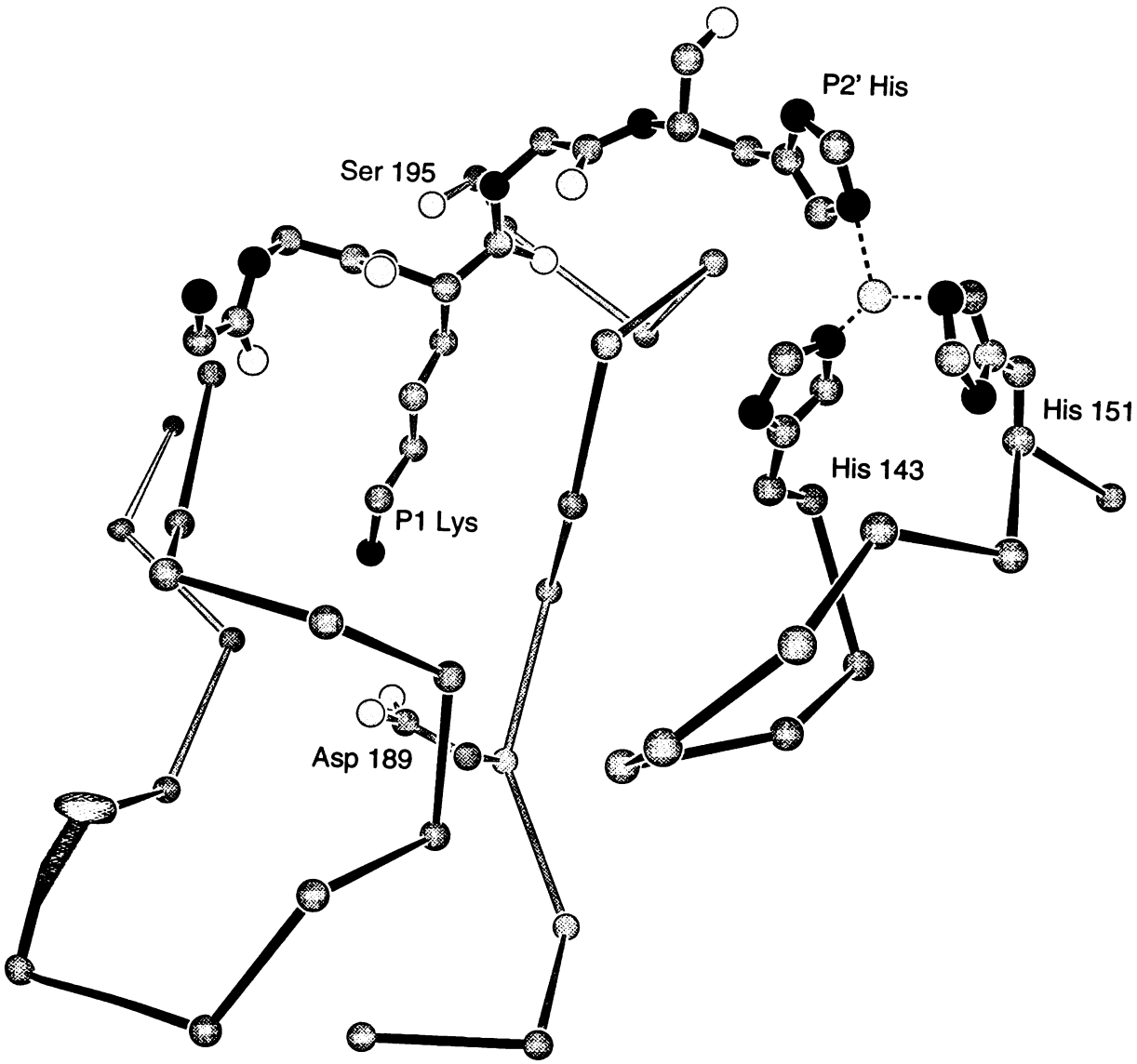


Figure 4.2: Computer model of Tn H143/151 interacting with bound substrate. The P1 lysine of BPTI is shown interacting with Asp189 at the bottom of the P1 pocket of trypsin. The designed metal binding site at S2' shows histidines at 143 and 151 forming a tridentate site with the P2' histidine of the substrate and bridging metal ion. Histidine-metal bond lengths are 1.5-2.0Å, and bond angles are 102-133°. Coordination geometry is roughly tetrahedral.

Results

Histidines were used to create metal binding sites in trypsin because they are known to coordinate transition metals. The variant enzyme designs were based on the X-ray crystal structure of rat anionic trypsin D189S complexed to BPTI (Perona, et al., 1994a). The gene encoding rat anionic trypsinogen was altered by site-specific mutagenesis, and variant and wild-type trypsins were expressed in yeast and purified to homogeneity by affinity chromatography. Variant enzymes were designed to contain metal binding sites in the S_2' or S_1 regions of the enzyme to effect metal-dependent substrate specificity towards substrates containing histidine at either P_2' or P_1.

Metal-dependent P_2' Specificity. Modeling of the trypsin/BPTI structure shows that residues 143 and 151 of the enzyme are candidates for a metal binding site that bridges to a substrate containing a histidine at P_2', forming an intermolecular tridentate site (Fig. 4.2). The modeling suggests histidine-metal bond lengths of 1.5–2.0Å and bond angles of 102-133°, showing roughly tetrahedral geometry. Based on observed structures of metal binding sites in proteins, transition metals preferring tetrahedral geometry (Zn^{+2} or maybe Ni^{+2}) should bind at this site. Alternatively, each of the single mutants, N143H, or E151H might also bind substrate through a metal ion in a bidentate fashion, although presumably with lower affinity than a tridentate site. Although three liganding atoms should bind a metal ion with higher affinity than two, optimal geometry of the site may be the

dominant factor to metal binding, leading to a bidentate site with higher affinity. The model prediction of short metal-ligand bond lengths (<2Å) implies that there would have to be some local structural changes in the protein to allow for tight binding of metal. Trypsins N143H, E151H and N143H/E151H were constructed, expressed, and purified to homogeneity. Each variant trypsin in this series was tested against Z-GPR-AMC to demonstrate structural integrity. These three variants possess wild-type levels of activity toward this P_1-Arg substrate which occupies the P_1-P_3 sites on the enzyme (Table 4.1).

Table 4.1: Amide Kinetics of S_2' Variants[a]

Enzyme	k_{cat} $(min)^{-1}$	$K_m(\mu M)$	k_{cat}/K_m
trypsin	3500	14.1	248
Tn N143H	2670	17.7	151
Tn E151H	4160	18.2	228
Tn H143/151	2970	20.4	145
Tn E151Q	2955	16.2	182

[a]Fluorogenic substrate Z-GPR-AMC was hydrolyzed in 10mM PIPPS/1mM tris/100mM NaCl/1mM $CaCl_2$, pH 8, at 25°C. [E]=2.5nM. These variants all possess wild-type levels of activity against this substrate.

A discontinuous RP-HPLC peptide cleavage assay was used to observe specificity at P_2' in the presence and absence of transition metal ions. The substrate octapeptide AGPYAHSS contains a Tyr at P_1 to maximize the effect of turnover due to binding at P_2' because very fast hydrolysis rates of a substrate containing Arg or Lys at P_1 could mask metal-dependent specificity at P_2'. Wild-

type trypsin fails to cleave AGPYAHSS at a measurable rate (Fig. 4.3). None of the enzymes cleave this peptide substrate in the absence of metal ion, and neither Tn N143H nor Tn E151H shows activity in the presence of copper, nickel, or zinc. However, the double variant N143H/E151H displays activity in the presence of either nickel or zinc, but not copper (Fig. 4.3). The peptide AGPYAASS was used as a control to demonstrate the requirement for histidine at P_2'. This substrate is not cleaved by any of the enzymes either in the presence or absence of metal ions. The AGPYAHSS substrate was also incubated with 500μM nickel and zinc in the absence of enzyme to test for metal-mediated cleavage. No cleavage was observed after 15 hours incubation at 37°C, demonstrating that both enzyme and added metal are required to obtain substrate turnover.

Mass spectral analysis was used to determine the cleavage site in the peptide substrate. The cleavage products were collected from the HPLC effluent and dried by vacuum centrifugation. Figure 4.4 illustrates the RP-HPLC assay of the reaction products. The 789.4 Dalton substrate AGPYAHSS is cleaved to generate the 407.2 Dalton AGPY peptide and the 401.2 Dalton AHSS peptide. This result agrees with the masses and retention times predicted by the program MacPro Mass, and confirms cleavage following the tyrosine residue, thereby placing the histidine at P_2'.

Figure 4.3: Kinetic parameters of AGPYAHSS hydrolysis. Turnover by Tn H143/151 with Zn was only 250-fold lower than chymotrypsin. There was no observable activity towards the peptide substrate by any trypsin molecule in the absence of metal or with copper. The control peptide AGPYAASS, lacking a P2' histidine was not hydrolyzed with or without metal. Substrate incubated alone with metal was not hydrolyzed. The data shown for chymotrypsin were obtained without metal.

Figure 4.3: Kinetics of AGPYAHSS Hydrolysis

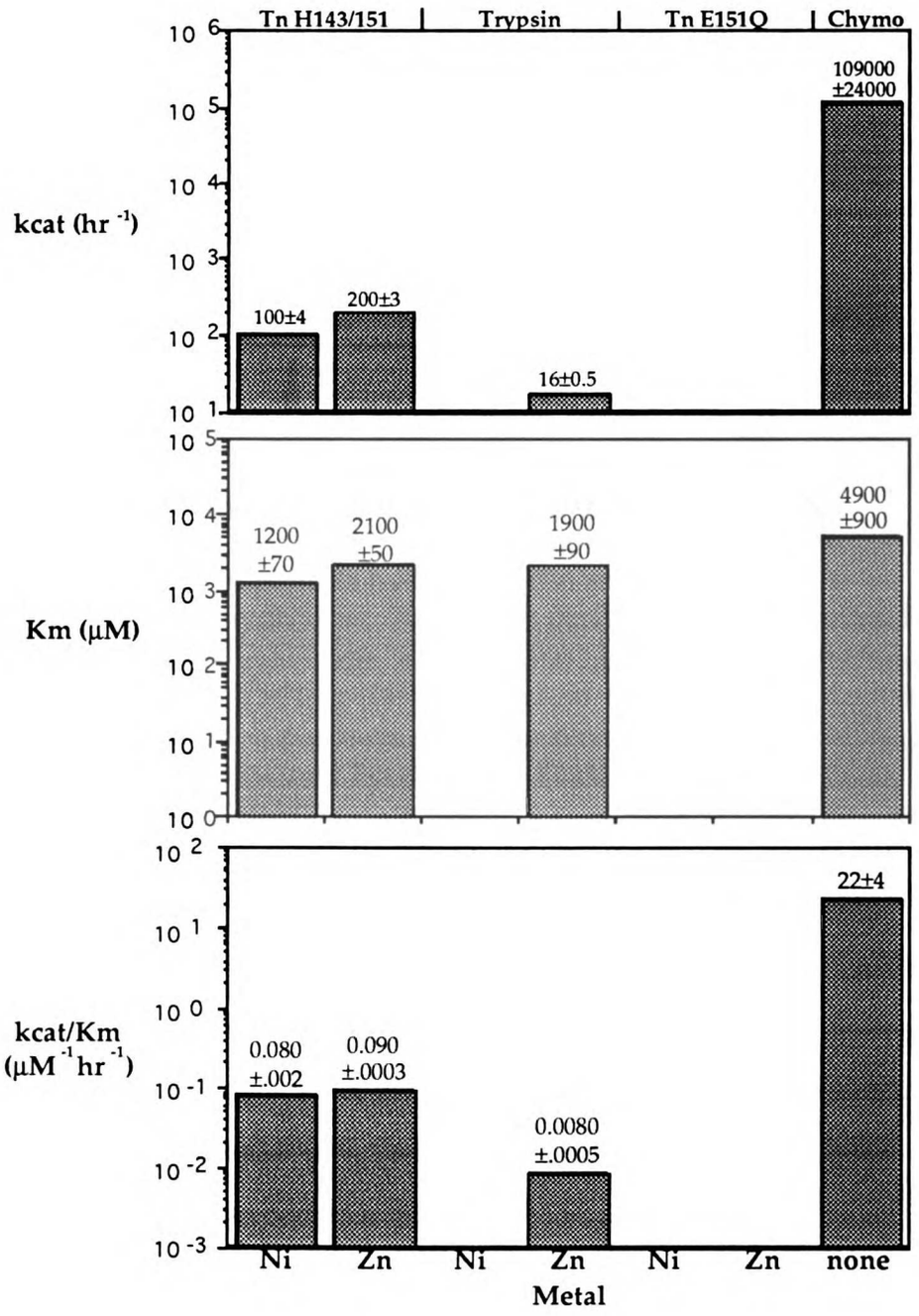


Figure 4.4: Peptide Cleavage HPLC Assay

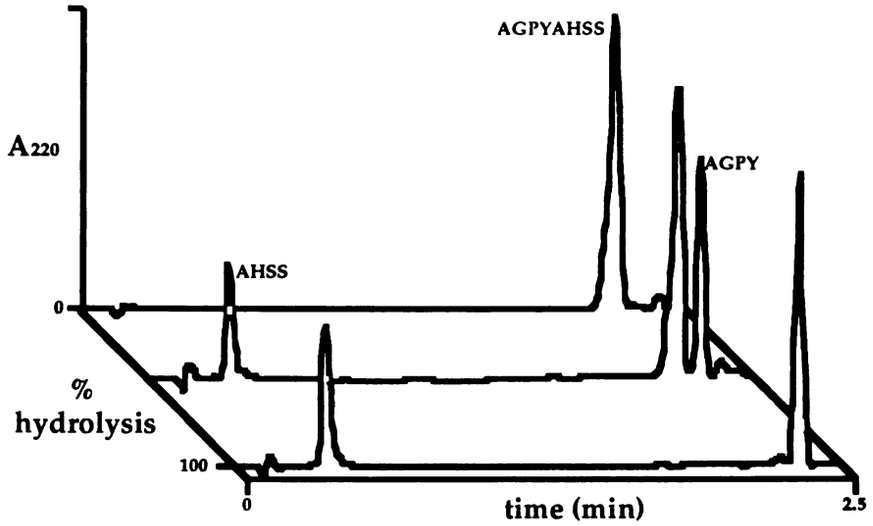


Figure 4.4: HPLC assay of peptide hydrolysis, illustrating progress of reaction from 0 to 100% hydrolysis. Assay was performed by incubation for 2 hours at 37°C of 1μM enzyme with 200μM metal ion in 10mM PIPPS/1mM Tris/100mM NaCl/1mM $CaCl_2$, pH8. The reaction was stopped by addition of TFA to 0.1%, and reaction products were separated over a C_{18} cartridge in a 0-30% acetonitrile gradient over 2.5 minutes. Peak identities were confirmed by mass spectrometry.

Kinetic analysis shows that TnN143H/E151H displays metal-dependent histidine specificity at P_2', with kcat/Km = (82±1.7)x10^{-3} and (93±0.3)x10^{-3} $\mu M^{-1}hr^{-1}$ in the presence of nickel and zinc, respectively. Wild-type trypsin does not hydrolyze AGPYAHSS in the presence of nickel, but does in the presence of zinc with kcat/Km = (8.3±.5)x10^{-3} $\mu M^{-1}hr^{-1}$, 10-fold lower than Tn N143H/E151H (Fig. 4.3). To address this unexpected activity of wild-type trypsin, Tn E151Q was made to remove possible metal-binding capability of the carboxylate of Glu 151 in the S_2' region. The glutamine side chain has a much lower affinity for metal ions than glutamate, and therefore would not be expected to bind the histidine substrate through a bridging metal ion. Trypsin E151Q has activity

indistinguishable from wild-type toward the Z-GPR-AMC substrate (Table 4.1), but shows no activity toward the AGPYAHSS substrate under any conditions. This indicates that the carboxylate of Glu151 is responsible for the observed activity of wild-type trypsin in the presence of zinc. Chymotrypsin was used as a positive control to establish a baseline for cleavage following the P_1-Tyr; hydrolysis of AGPYAHSS proceeds with kcat/Km = 22.5±3.8 $\mu M^{-1}hr^{-1}$. This rate is 275-fold greater than trypsin N143H/E151H in the presence of nickel or zinc.

The activity of trypsin N143H/E151H is dependent on the concentration of added Ni^{+2} or Zn^{+2} ion. The initial velocity of peptide hydrolysis increases with increasing zinc or nickel concentration until the metal and substrate concentrations are equal. Titration of enzyme activity by metal concentrations below substrate concentration displays Michaelis-Menten kinetics, demonstrating a bireactant system with metal ion acting as an obligate cofactor (Segel, 1975). When the concentration of metal exceeds that of substrate, activity decreases (data not shown). This trend is seen at substrate concentrations ranging from 200μM to 1mM. Figure 4.5 shows nickel and zinc titration of activity at 630μM substrate yielding an apparent zinc binding constant $K_{Zn_{app}}$=54μM, and an apparent nickel binding constant $K_{Ni_{app}}$=38μM.

Figure 4.5: Metal Titration of Activity

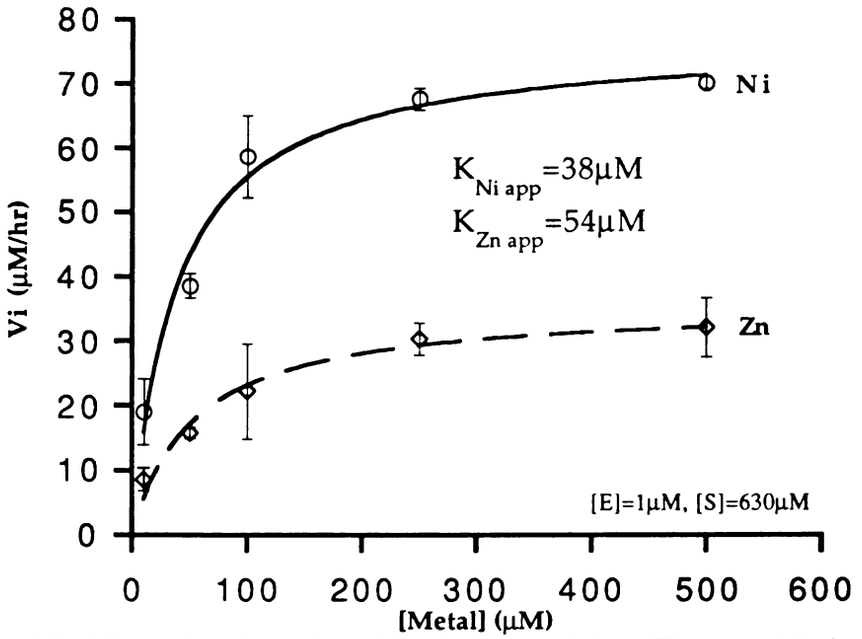


Figure 4.5: Metal titration of Tn H143/151 activity. The peptide hydrolysis assay was performed with increasing concentrations of nickel and zinc. These curves show the saturation of activity by metal ion, yielding apparent metal binding constants of 38 and 54μM, respectively, when the data were fit to the Michaelis-Menten equation.

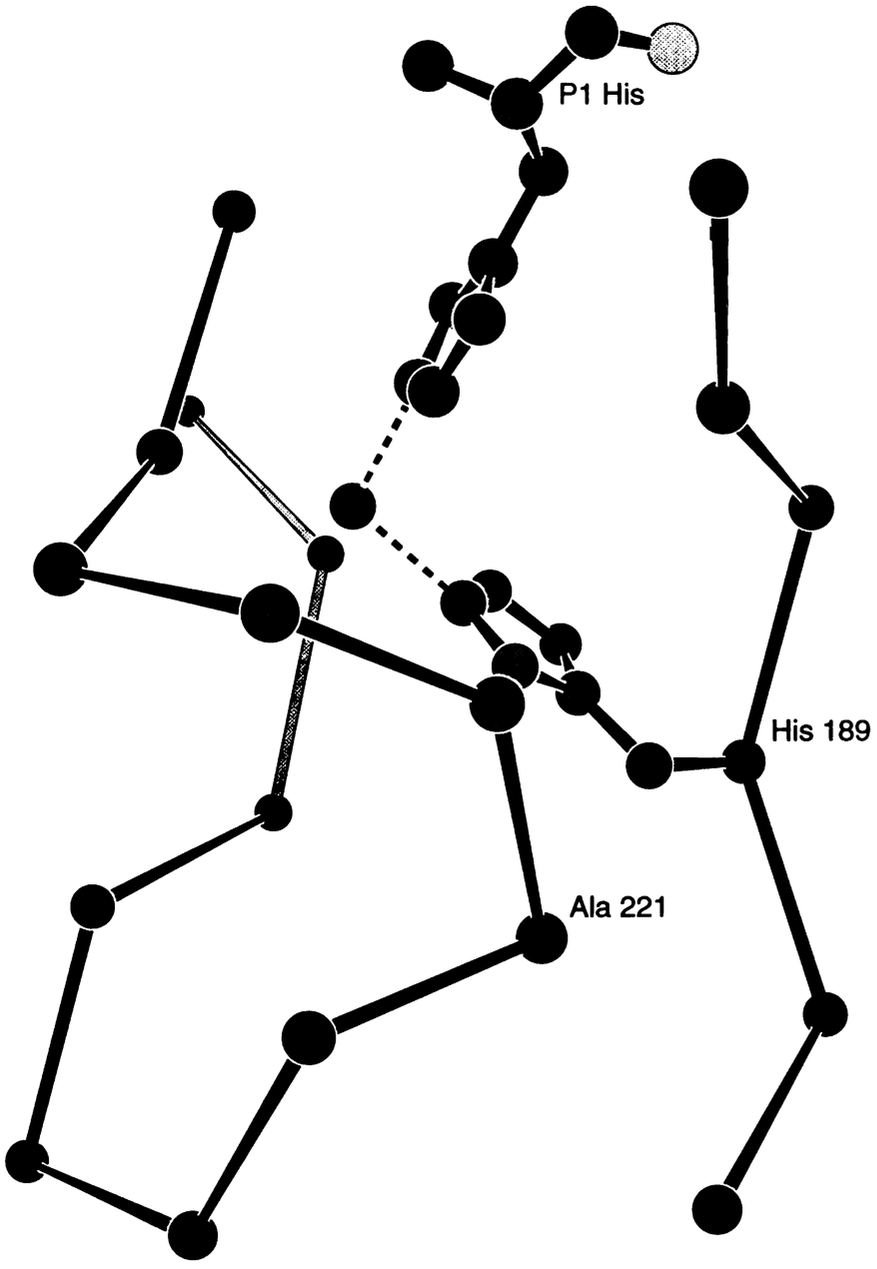
These apparent binding constants are comparable to those measured for bidentate sites, suggesting that only two histidines are bound to the metal ion or that the geometry of the three histidines is poor, resulting in lower affinity.

P_1 Specificity. Asp189 was mutated to His to engineer metal-regulated histidine specificity into the trypsin S_1 binding pocket. As at position P_2', the design principle relies on the affinity of transition metals for histidine side chains, such that a metal ion bridging the enzyme and substrate results in cleavage C-terminal to histidine. Modeling of the primary binding pocket in trypsin showed His189 able to assume the conformation necessary to create a

metal binding site in concert with a substrate P_1-His (Fig. 4.6). The two histidines coordinate a transition metal ion with bond distances of approximately 2Å and a N-metal-N angle of 100°. These bond lengths and angles are typically observed in metalloproteins binding transition metals, and both the enzyme and substrate histidines possess allowed rotamers (Ponder and Richards, 1987). In this conformation, however, the His189 side chain makes short steric contacts of 2.5-3.0Å with main-chain atoms of Ser190. Energy minimization carried out with the program XPLOR suggested that a small (~0.5Å) repositioning outwards of Ser190 could result in suitable van der Waals interactions with the imidazole ring without introducing other structural changes. It was therefore reasoned that His189 would be accommodated in the modeled conformation because such small rearrangements are common in response to single-site mutations in many proteins.

Figure 4.6: Computer model showing designed metal binding site in the trypsin S1 pocket involving His189 of trypsin and the P1 His of substrate. The imidazole-Cu bond lengths are 2Å, and the Nε2-Cu-Nε2 bond angle is 100°.

Figure 4.6: Tryspin D189H Model



Trypsin D189H was expressed and purified to homogeneity with a final yield of 12mg/l of culture. The enzyme possesses no observable activity against the amide substrate Succ-AAPH-pNA either in the presence or absence of 100μM $CuCl_2$, $NiCl_2$, or $ZnCl_2$. Likewise, there was no observable activity toward Succ-AAPD-pNA or Succ-AAPE-pNA, or toward the amide substrate Z-GPR-AMC. However, the enzyme does hydrolyze the more reactive Z-Lys-thiobenzyl ester substrate with a kcat/Km = 0.35±0.01 $min^{-1}\mu M^{-1}$ (corrected for background hydrolysis), demonstrating a stably folded protein and integrity of the catalytic triad. Wild type trypsin hydrolyzes this substrate with a kcat/Km = 78±1 $min^{-1}\mu M^{-1}$, 200-fold faster than D189H (Table 4.2).

Table 4.2: Ester Kinetics of S_1 Variants[a]

Enzyme	k_{cat} $(min)^{-1}$	K_m (μM)	k_{cat}/K_m
trypsin	6700±150	86±3	78±5
Tn D189H	740±40	2100±170	0.35±0.01
Tn H189/D226	760±5	1200±25	0.63±0.01

[a]Kinetic parameters of S_1 variants against Z-Lys-SBzl ester. Assays were performed in 10mM PIPPS/1mM tris 100mM NaCl/1mM$CaCl_2$, pH 8, at 25°C with [E]=2μM.

A second generation of variants was designed in the P_1 site. These additional mutations were introduced in an attempt to stabilize His189 in a suitable orientation for metal binding. The variants D189H/S190G and D189H/G226D were constructed to remove the serine hydroxyl bulk at position 190, and to introduce a hydrogen bonding partner for His189, respectively. The modeled position of Asp226 relative to the His189 and metal ion mimicks the carboxylate-

histidine-zinc interaction which forms a stable structure in the active site of metalloproteins (Christianson and Alexander, 1989). In each case modeling followed by energy minimization showed similar metal ion coordination geometry to that of D189H. Trypsins D189H/S190G, D189H/G226D, and D189H/S190G/G226D were expressed in yeast and purified to homogeneity by BPTI affinity chromatography. Kinetic analysis showed that those variants containing Gly190 exhibited no activity, in the presence or absence of metal cofactors, towards either amide or ester substrates. Trypsin D189H/G226D likewise showed no amidase activity towards Succ-AAPH-pNA under any conditions, but did hydrolyze Z-Lys-SBzl ester with kcat/Km = 0.63±0.01 $min^{-1}\mu M^{-1}$; a two-fold increase in turnover relative to D189H, or 100-fold lower than wild-type trypsin.

Crystal Structure of Trypsin D189H. Trypsin D189H was crystallized in the presence of BPTI and the structure determined at 1.8 Å resolution to understand the structural basis for the failure of the enzyme to exhibit the desired metal-regulated histidine specificity. Crystallographic results are summarized in Table 4.3. Main chain atoms of the active site residues His57, Asp102, and Ser195 are superimposable with rms deviations of 0.05 to 0.2 Å when compared to other structures of rat trypsin variants (McGrath, et al., 1993, Perona, et al., 1994a, Perona, et al., 1993a,b). The average pairwise rms deviation among the various structures is 0.33 to 0.40 Å for all backbone Cα atoms. Thus, the D189H mutation does not compromise the overall folding of the enzyme.

Table 4.3: Trypsin D189H Crystallographic Data

Space Group	$P3_221$
Cell Dimensions	**a**= **b**= 92.83Å, **c**=62.18Å
	$\alpha=\beta=90°$, $\gamma=120°$
Resolution	1.8Å
%Complete	95%
R_{merge}[a]	6.03%
R_{crys}[b]	18.4%
rms bond differences	0.011Å
rms angle differences	2.655°

[a] Rmerge=(ΣhΣi | <Fh>-Fhi |)/(ΣhFh) where Fh is the mean structure factor magnitude of i observations of symmetry-related reflections with Bragg index h. [b]Rcrys=(ΣhΣi | Fobs | - | Fcalc |)/Σ | Fobs |) where Fobs and Fcalc are the observed and calculated structure factor magnitudes.

The His 189 side chain is oriented toward the back of the pocket away from substrate. Its position contrasts strikingly to that of Asp189 in wild type trypsin, and precludes interaction with a metal ion to form a P_1-His binding site (Figs. 4.7a,b,c). The imidazole ring forms two hydrogen bonds: one with poor geometry between His189 $N\varepsilon_2$ and the carbonyl oxygen of Gly187 (N--H--O angle 131°, 2.65Å), and a second from His189 $N\delta_1$ to water molecule 554 (N--H--O angle 156°, 2.64 Å; Fig. 4.7b). This water bridges to the main-chain carbonyl group of Pro225. Sharp peaks in electron density maps calculated from coefficients (Fo-Fc) during refinement indicated that the peptide groups of both Pro 225 and Gly 187 have been rotated approximately 180° in ψ relative to wild-type trypsin. These backbone adjustments apparently occur to provide the hydrogen-bonding interactions with His189. A further consequence of the main-chain reorientations is a significant perturbation in the overall conformation of the two surface loops bridging the three β-strands of the P_1 pocket (Loops 1 and 2; Fig. 4.7). While the deviation in position of equivalent α-carbons within Loop

1 ranges from 0.2-1.0 Å, in Loop 2 the deviation is as great as 4 Å relative to wild-type trypsin. An inter-loop hydrogen bond bridging residues Gly188 and Ala221 is disrupted in this new conformation, although the main chain hydrogen bond between residues Asp223 and Leu185 is preserved. The water molecules nearest the entrance of the S1 pocket, which interact with the P_1-Lys of BPTI, occupy identical positions in this structure relative to wild-type trypsin. Other waters at the base of the pocket are found in novel locations resulting from the presence of His189 and the new conformations of Loops 1 and 2.

Figure 4.7a: Crystal structure of Tn D189H complexed to BPTI showing detail of P_1 pocket. Histidine 189 is pointed down and away from the top of the pocket and is unable to participate in metal or substrate binding. Relevant interatomic distances in Angstroms are: His189Nε_2 to carbonyl oxygen of Gly187, 2.65; His189Nδ_1 to water 554, 2.64; water 554 to water 556, 2.59; water 554 to water 555, 2.89; water 554 to carbonyl oxygen of Pro225, 2.89; and water 554 to carbonyl oxygen of Asn224, 2.71. Loop 1 is residues 185-189, and Loop 2 is residues 217-225. The disulfide bond between Cys191 and Cys220 is shown striped.

Figure 4.7b: Trypsin D189H structure superimposed upon wild-type trypsin. Wild-type trypsin is shown in black, while trypsin D189H is shown in grey. The His189 side chain is pointing down in the pocket as opposed to the Asp189 of wild-type trypsin. Also visible is the 219-225 loop (Loop 2) movement of 1-4Å. The disulfide bond between C191 and C220 is shown with the dashed line.

Figure 4.7c: Stereo $2F_O$-F_C map contoured at 1σ of trypsin D189H showing the area around His 189 and including the part of the 220's loop which deviated most from wild-type trypsin (Loop 2 in Fig. 4.7b). The electron density was displayed using the program CHAIN (Sack, 1988).

Figure 4.7a: Crystal Structure of Trypsin D189H

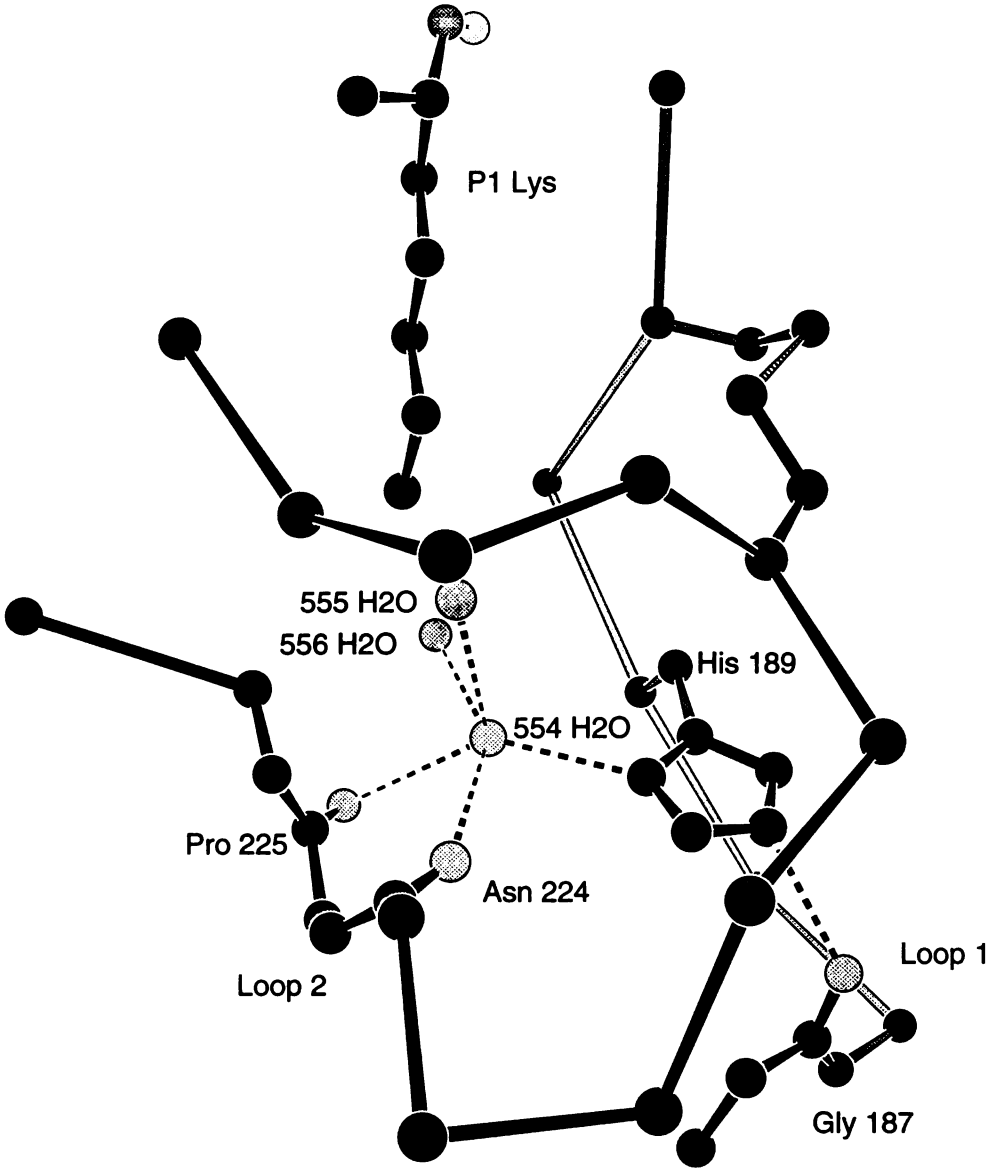


Figure 4.7b: Superposition of Trypsin with Tn D189H

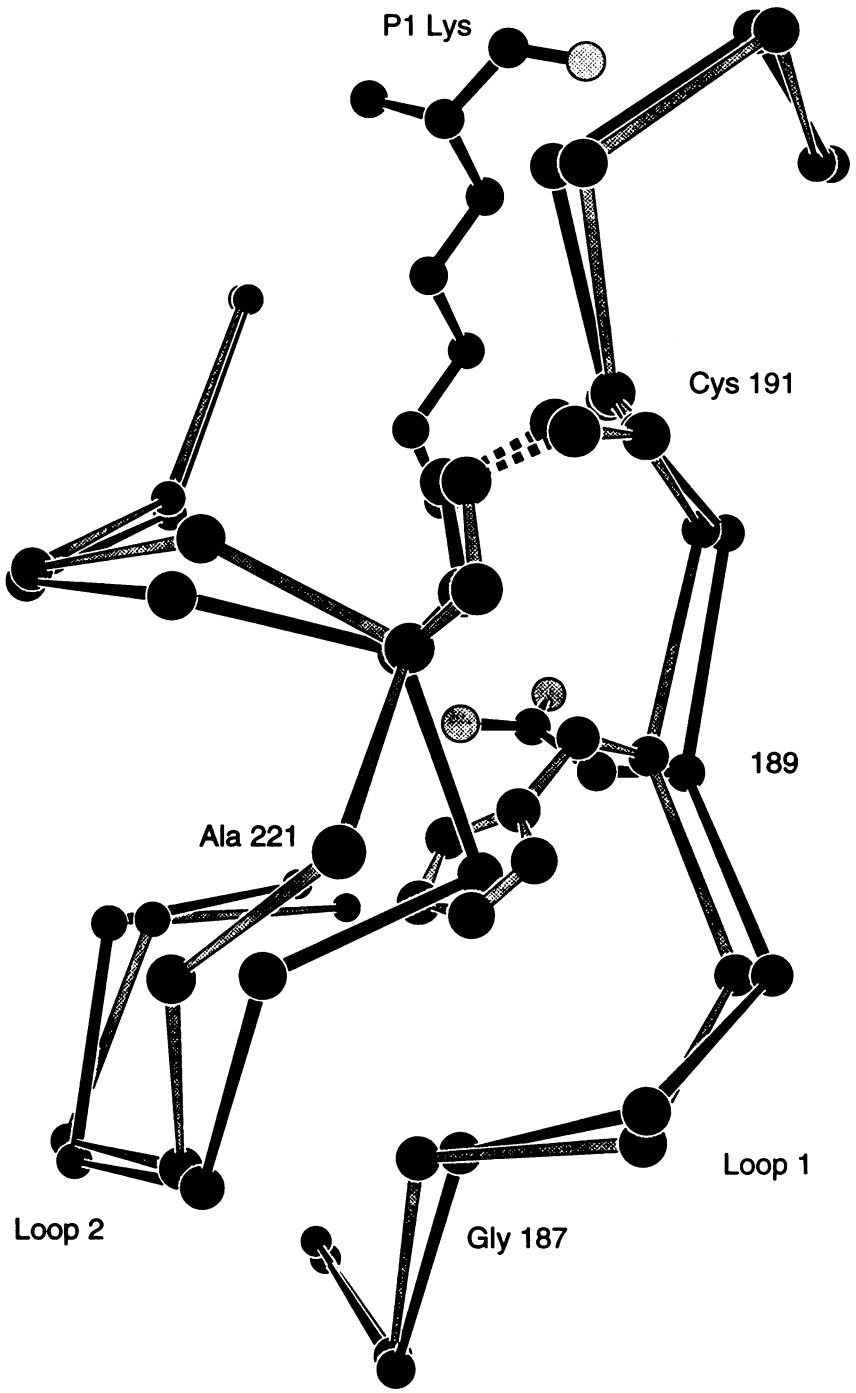
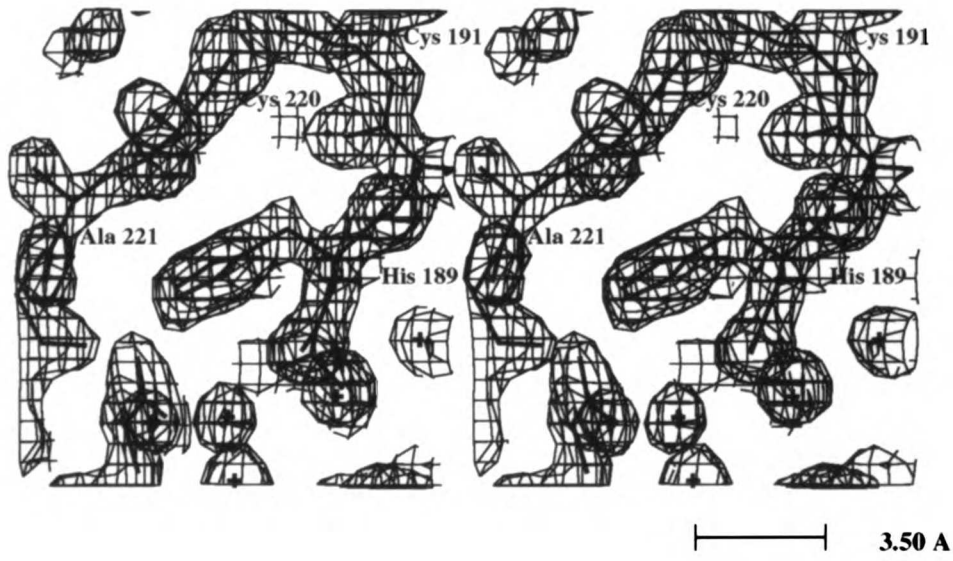


Figure 4.7c: Stereo $2F_o$-F_c Map of His189 Region



Discussion

P_2' Designs. Molecular modeling was successful in predicting a trypsin design which introduces metal-dependent histidine specificity at P_2'. Trypsin N143H/E151H cleaves a peptide substrate containing a histidine at P_2' in the presence of either nickel or zinc, but not in the presence of copper. The preferred ligand geometries of these three metal ions suggests the observed specificity: distorted octahedral and tetrahedral coordination geometries are observed with nickel and zinc, whereas copper as well as nickel are observed in a square planar geometry. Coordination of metal ions in proteins probably reflects balance of energetic contributions between protein flexibility and metal ion geometry. The S_2' site is formed by the juxtaposition of several surface loops well exposed to solvent. Possibly, the greater potential flexibility of these surface loops, relative to core β-strands critical to the tertiary fold, might be a key to the accomodation of metal binding without a loss of structural integrity. By contrast, the S_1 site is a deep cavity which is only partially solvent-accessible, and is formed by β-strands which are core secondary structure elements.

The activity of the enzyme towards the peptide substrate was strictly dependent on nickel or zinc concentration, showing saturation kinetics when metal concentrations were lower than substrate concentrations. These data obey Michaelis-Menten kinetics (Fig. 4.5) in a manner consistent with a two-reactant system where the metal ion acts as an obligate cofactor (Segel, 1975). At a substrate concentration of 630μM, zinc and nickel bind with apparent binding constants $K_{Zn_{app}}$= 54μM, and $K_{Ni_{app}}$=34μM. These binding constants are expected to be dependent on substrate concentration as the histidine-containing substrate itself has some intrinsic binding affinity for the metal ions. This is

manifested as a reduction in V_{max} when metal ion concentrations exceeds substrate concentrations. The decrease in activity at high metal ion concentrations is consistent with either an interaction between the metal and the substrate, rendering the substrate unavailable to the enzyme, or between metal and enzyme, causing direct enzyme inhibition.

Metal ions are used in naturally occurring proteins for a variety of purposes. In proteases they are used in catalysis and structural stability, but they are yet to be observed in proteases-substrate binding interactions. Other types of enzymes are known to use metal ions to bind substrates. Data for the restriction enzyme EcoRV show that Mg^{+2} is required for proper substrate specificity, and show relaxed substrate discrimination when Mg^{+2} is replaced by Mn^{+2} (Vermote and Halford, 1992). Additionally, studies of xylose isomerase have shown that the bound Mg^{+2} ions are important for proper substrate binding and subsequent catalysis (Jenkins, et al., 1992). These results indicate that metal ions are widely used for substrate recognition in enzymes other than proteases. Following this observation is the question of whether proteases also use this mechanism of substrate binding even though it has not been observed. Possibly trypsin N143H/E151H functions in a manner similar to undiscovered naturally occuring proteases employing metal-dependent molecular recognition.

The observation that zinc activates wild-type trypsin, but nickel or copper does not, raises interesting questions regarding the nature of metal ion specificity for coordination to His-Glu pairs. The selectivity may be geometrically based, requiring proper spatial orientation of ligands. Also, it is notable that copper ion does not activate any of these enzymes at any of concentrations used in this study. As copper is expected to bind tighter to histidines than either nickel or

zinc (Martell and Smith, 1974), one explanation for the failure of copper to activate the enzyme is that it binds too tightly to the enzyme/substrate complex to allow turnover. In this context, the designed metal binding site may function to inhibit the enzyme from hydrolyzing P_2' histidine-containing substrates in the presence of copper. It has been observed that Cu^{+2} inhibits wild-type trypsin with a Ki=225μM (W. S. W., unpublished result), indicating that copper interacts with the enzyme in a deleterious fashion. The copper ion may bind between the catalytic His57 and Asp102 causing the observed inhibition as silver ion was shown to do (Chambers, et al., 1974). Alternatively, the square planar geometry preferred by copper may not be possible within the structural constraints imposed at the S_2' site. Crystal structures of trypsin N143H/E151H complexed with a P_2'-His-containing inhibitor in the presence of different metal ions should provide further insight into this question.

Trypsin D189H Crystal Structure. The crystal structure of trypsin D189H reveals conformational changes in two surface loops (Loop1 and Loop2; Fig. 4.7b) adjacent to the S_1 pocket, resulting in a new network of intramolecular interactions. The structural changes result from the interaction of His189 with backbone amides of both loops, which leads to the ψ rotation of 180° of two peptide bonds and consequent adoption of the new conformation. It is of interest that the interaction of a His imidazole with adjacent main-chain groups is not uncommon in proteins (Baker and Hubbard, 1984). For example, in trypsin His40 makes a similar interaction with the carbonyl oxygen atom of Gly193, and His53 in the serine protease inhibitor ecotin (McGrath, et al., 1994) interacts with the carbonyl oxygen atom of Trp51 in a similar manner. The geometry in the latter case is very similar to that observed in the His189-Gly187 interaction. In both instances the hydrogen bond-accepting backbone carbonyl group lies in an

adjacent loop. Possibly, a favorable interaction of the His189 imidazole ring with the adjacent Gly187 main-chain provides part of the stabilization energy of the new loop conformations. These observations might profitably be used to reorient surface loops in protein engineering designs.

It is, however, difficult to estimate the subtle forces that favor a large rearrangement of the pocket to the new conformation rather than the modest relaxation of the Ser190 backbone apparently required to permit orientation of His189 toward substrate. The additional substitutions introduced into D189H were ineffective in stabilizing the modelled orientation of the imidazole ring. Since the Gly190 variants exhibit no activity at all it is possible that these enzymes are misfolded and do not adopt stable tertiary structures. Trypsin D189H/G226D manifests a weak esterase activity very similar to that of D189H. Thus, it is seems likely that His189 in this enzyme adopts a D189H-like orientation. None of the changes therefore appear to be sufficient to overcome the combination of forces stabilizing the new conformations of Loops 1 and 2.

The rearrangement of the S_1 pocket in trypsin D189H provides insight into the requirements for substrate specificity modification at this position. In contrast to the serine proteases subtilisin and α-lytic protease (Bone, et al., 1989, Estell, et al., 1986), mutagenesis of trypsin S_1-site residues directly in contact with substrate fails to improve catalytic efficiencies toward alternate P_1-amino acids (Graf, et al., 1987, Graf, et al., 1988; J. J. P. & C. S. C., unpublished observations). Instead, exchange of the surface loops Loop 1 and Loop 2, together with replacement of four S_1-site amino acids, will endow trypsin with a chymotrypsin-like substrate preference (Hedstrom, et al., 1994, Hedstrom, et al., 1992, Perona, et al., 1994b).

In this context it is of interest that the introduction of His189 causes a large structural change in Loops 1 and 2 rather than a small shifting of the main-chain directly in the P_1 side-chain binding site at Ser190. In α-lytic protease, greatly broadened substrate specificity profiles appear to arise from structural plasticity of the S_1 site, which includes small movements (0.5-1.0 Å) of the main chain in an area which directly contacts substrate (Bone, et al., 1991). Thus, the enzyme can deform to accommodate alternate P_1-substrate side chains in a manner which does not compromise positioning of the scissile bond relative to the catalytic machinery. Structural analysis of a number of S_1-site trypsin variants, however, shows some adjustment in second-shell residues but none in amino acids which directly contact substrate (Perona, et al., 1993a,b, Wilke, et al., 1991). Similarly, we observe here that in trypsin D189H an apparently small deformation of Ser190 to accommodate His189 does not occur. Thus, the His189 imidazole may be oriented away from substrate because there is no capacity for local deformability. The absence of even limited flexibility to permit orientation of His189 toward substrate supports the hypothesis that the requirement for extensive mutagenesis to obtain specificity modification in trypsin is due to a high degree of rigidity of the S_1 site (Perona, et al., 1994b).

Loops 1 and 2 also form part of the "activation domains" of trypsin and chymotrypsin, which become ordered only upon transition from the zymogen forms to the mature enzymes (Huber and Bode, 1978). The segments involved in this conformational transition include the N-terminal residues 16-19, the adjacent loop from residues 142 to 152, and Loops 1 and 2 together with adjacent residues at positions 216 and 189-193. Kinetic and structural analyses of trypsin mutants possessing partial chymotrypsin-like specificities suggests that the stability of the S_1 site in each native enzyme is highly dependent upon the integrity of the

adjacent structure (Hedstrom, et al., 1994). Both the conformations of Loops 1 and 2 as well as their interactions with surrounding portions of the structure are quite different in trypsin and chymotrypsin. Alternate strategies for stabilizing the S_1 site in trypsin and chymotrypsin thus appear to have evolved together with the divergence in substrate specificity.

In this context it is noteworthy that in trypsin D189H the entire "activation domain" (Huber and Bode, 1978), including Loops 1 and 2, is very well-ordered. The structure shows that the side chain of His189 occupies a central position between Loops 1 and 2, where it disrupts existing wild-type interactions and solvent structure in addition to producing the large conformational change in Loop 2. Therefore, the networks of intramolecular interactions present within the activation domains of trypsin and chymotrypsin do not represent exclusive solutions to the problem of producing a stable structure. Trypsin D189H is produced in the zymogen form and activated by enteropeptidase. This shows that the mutation does not block the normal processing pathway, although it may destabilize the partially folded zymogen state. The observation that D189H possesses an alternate stable conformation in the activation domain, rather than destabilizing the structure as observed in D189S (Hedstrom, et al., 1994), may be of significance to understanding the mechanism of zymogen activation in these enzymes.

The crystal structure of Tn D189H provides the structural basis for the observed kinetic result. Kinetic data can be used in conjunction with X-ray structural data as the basis for iterative design, making trypsin a good model system to test protein engineering hypotheses. These experiments showed however, that the primary specificity pocket is not easily reconfigured. The

obvious "improvements" suggested by the structure do not lead to the desired kinetic result of metal-dependent specificity. This reflects the current partial understanding of protein structure/function relationships, yet allows a significant addition to the database of protein engineering results.

This results in this chapter demonstrate that an engineered metal binding site may be used to effect histidine specificity in trypsin. Kinetic analysis was used to assess the success of the design, and X-ray crystallographic analysis was used to elucidate structural details of the mutant enzyme. Iterative use of kinetic and structural analysis may improve current levels of activity so that these design principles can be used to create highly specific, metal-dependent proteases.

Chapter 5

X-Ray Crystal Structures and Designed Improvement of a Metal-Activated Trypsin

Abstract

Analysis of the S_1 subsite interactions of trypsin with the P_1 residue of bound inhibitor suggested the addition of the D189S mutation to the metal-assisted trypsin H143/151 to improve the catalytic efficiency towards the tyrosine-containing peptide, AGP**Y**-**AH**SS. The triple-mutant trypsin N143H/E151H/D189S exhibits a 25-fold increase in activity with nickel, and 150-fold increase in activity with zinc relative to trypsin H143/151 on this peptide. In addition, activity of the trypsin H143/151/S189 towards an arginine-containing peptide, YLVGP**R**-**GH**FYDA, is titratable by copper, nickel, and zinc, yet in this case copper yields a 30-fold, nickel a 70-fold, and zinc a 350-fold increase, respectively, in activity over background hydrolysis without metal. These results demonstrate that engineering of multiple substrate binding subsites can be used to iteratively improve engineered protease specificity. This study suggests that delocalizing substrate binding interactions from a primary binding pocket to multiple subsites equalizes relative contributions from engineered subsites. The crystal structure of the metal-activated protease trypsin N143H/E151H complexed to the serine protease inhibitor ecotin A86H has been solved with and without the presence of copper, nickel, and zinc. These results elucidate the structural aspects of the observed metal-activated proteolysis. In the absence of metal, the histidines of the designed site are not bound to any metal ion and the 140-151 loop in trypsin that contains them is disordered. Crystals which were

soaked with copper, nickel, or zinc chloride contain metal bound by the histidines at the designed site and in the zinc structure, the 140-151 loop becomes ordered. Copper is bound with bond distances of 1.9-3.0Å and distorted square planar geometry. Nickel is bound with 2.0-2.6Å bond distances and square pyramidal geometry. Zinc is bound with 1.9-2.3Å distances and distorted tetrahedral geometry. In each case the fourth ligand is presumed to be water, however this is not visible in the zinc structure, while it is visible in both the copper and nickel structures.

Introduction

One of the goals of protein engineering is to add a desired function to an existing enzyme. Although our current understanding of macromolecular interactions allows only simple design experiments to be carried out successfully, there have been some promising results which demonstrate our ability to manipulate protein structure and function in a predetermined fashion. Because proteases have uses ranging from analytical laboratory tools, through therapeutic targets, to industrial processing enzymes, they are an important focus of protein engineering studies. For example, subtilisin has been engineered to possess altered specificity (Wells & Estell, 1988), peptide ligase activity (Abrahmsen, et al., 1991), enhanced thermal stability (Braxton and Wells, 1992) and increased stability in organic solvents (Arnold, 1993). Also, tissue-type plasminogen activator (t-PA) has been engineered to be more fibrin-specific in an effort to enhance its value as a therapeutic (Paoni, et al., 1993).

Using structure-based design, we have previously engineered histidine specificity into trypsin for the P_2' position of the substrate. This was done by

creating a metal binding site bridging enzyme and substrate so that the enzyme hydrolyzes a tyrosine peptide bond, AGP**Y**A**H**SS, only in the presence of nickel or zinc ions (Willett et al., 1995). Peptide substrates allow us to observe kinetic effects of enzyme-metal-substrate interactions at the P_2' position, while X-ray crystallography is instrumental in elucidating the structural component of the designed protease activity.

The pan-specific serine protease inhibitor, ecotin (McGrath, et al., 1994), serves as a model for bound substrate in these studies. The results are presented here of the X-ray crystallographic analysis of the designed metal binding site in trypsin N143H/E151H complexed to ectoin A86H without metal, and in the presence of copper, nickel, and zinc. In addition, further kinetic analysis and an iterative improvement of metal-activated hydrolysis by trypsin H143/151 and H143/151/S189 is presented.

Materials and Methods

Mutagenesis, expression and purification of trypsin and ecotin. Materials and methods involving trypsin have been described previously (Willett, et al., 1995), and are the same for the experiments reported here. Trypsin N143H/E151H/D189S was constructed by making single-stranded DNA from pST plasmid containing the Tn H143/151 mutant and performing site-directed mutatenesis (Kunkel, 1985) using the oligo GGGAGG**A**AAG**AGTTC**CTGCCAGGG (bold type denotes mismatch). The ecotin gene was mutated by site-directed mutagenesis to create ecotin A86H using the non-coding oligo GGGGCCA**C**AG**C**TGCTACTAC**GTG**ACGGGCC and expressed in E. *coli* strain X90 (F' *lac* I^q, *lac ZY, pro AB/ Δ-(lac-pro), ara, mal A, argEam, thi,* rif^r) as previously described (Wang, et al., 1995). Ecotin

A86H was purified as as follows. The periplasmic fraction from a 1 liter culture containing ecotin was passed through a 5ml DEAE Sepharose FF (Pharmacia) column equilibrated with 10mM Tris, pH 8.0. The unbound fraction was collected and made 20mM in Glycine from a 2M stock solution, then adjusted to pH 3.0 for cation exchange chromatography on a 15ml SP Sehpharose FF column (Pharmacia). The ecotin was eluted from the resin with a 75ml gradient of 0-1.5M NaCl in 20mM Glycine, pH 3.0. Fractions containing ecotin were dialyzed into water and concentrated to ~500μM ecotin (ε_{280}=21,860 M^{-1}) with a centriprep 10 concentrator (Amicon). This material was then loaded in 3ml lots onto a Vydac C4 prep column (2cm x 20cm) equilibrated with 0.1% TFA, and ecotin was eluted with a 0-40% acetonitrile/0.1% TFA gradient over 15 minutes. Fractions containing pure ecotin as judged by Coomassie-stained gel were lyophilized, resuspended in water, concentrated again with a centricon 10 (Amicon) to a concentration ≥1mM, and stored at 4°C.

Kinetic parameter determination. Kinetic parameter determination of peptide and synthetic substrate hydrolysis was described previously (see Chapter 4). For the case of trypsin H143/151/S189, 100nM enzyme was used in the peptide assay. All peptide assays were carried out in 5mM PIPPS/1mM Tris/100mM NaCl/1mM $CaCl_2$, pH 8.0 at 37°C.

Crystallization. Trypsin H143/151 which had been concentrated to ~1mM in 1mM HCl was mixed with ecotin A86H in equimolar amounts in a 0.5ml siliconized eppendorf tube on ice. To this mixture, 3M NaOAc and 1M Tris, pH 8.0 were added so that the final concentrations were as follows. Trypsin and ecotin were 375μM, NaOAc was 300mM, and Tris was 100mM. The

resulting pH of this solution was 8.0. The trypsin/ecotin complex was crytallized using the hanging drop method (McPherson, 1989) by mixing 2µl of protein solution with 2µl of well buffer and incubating the tray at 18°C. The well buffer contained 18-22% PEG 4K (Alltech), 300mM NaOAc/100mM Tris, pH 8.0/10mM $CaCl_2$. Crystals designated for soaking in metal solutions were stabilized in an artificial mother liquor of 25% PEG 4K/300mM NaOAc/10mM Tris, pH 8.0, and then soaked in increasing amounts of either copper, nickel, or zinc in artificial mother liquor. Typical metal soaks sequences were 1µM, 10µM, 100µM, and finally 1mM metal in artificial mother liquor for 30 minutes to 4hrs. The crystals were back-soaked into metal-deficient artificial mother liquor immediately prior to mounting in order to remove free and loosely bound metal ions. Additional details on the metal buffer preparation, crystallization, and metal soaks are provided in Appendix 4.

X-ray data collection and reduction. All X-ray diffraction data were collected at room-temperature on an R-axis IIc imaging plate with a Rigaku RU-200 rotating anode generator operating at 15 kW (50 mA and 300 kV) with a graphite monochromated CuKα radiation (λ=1.5418 Å). Exposure times were constant throughout each individual data set and were either 15 or 20 min./degree, depending on the intensity of the diffraction of the crystal in question. All diffraction data were indexed, integrated, scaled and merged with the **HKL** software package (Otwinowski, Z., Yale University, New Haven, CT, 1990).

Structure solution and data refinement All structures described were solved *via* the molecular replacement method, and were solved and refined

by Linda S. Brinen in collaboration with the laboratory of Robert J. Fletterick.. The apo structure was solved with the conventional rotation and translation function of the X-PLOR package, using one monomer of the original trypsin-ecotin complex structure as a search model (McGrath et al, 1994). It is interesting to note that the original trypsin-ecotin complex crystallized in a different monoclinic space group, $P2_1$. Subsequently, the three metal-containing structures were solved with either X-PLOR (Brunger et al, 1992) or AMoRE (Navaza, 1994), using the apo structure as a search model. Conventional refinement and fitting strategies were employed with X-PLOR and CHAIN, respectively.

Results

Trypsin-ecotin crystallization and metal soaks. The structural basis for the observed metal-activated hydrolysis was examined by X-ray crystallography. The bacterial serine protease inhibitor, ecotin, was chosen as a model for bound substrate because a convenient recombinant expression system allowed us to purify milligram quantities of a mutant possessing a histidine at the P2' position. Trypsin H143/151 was co-crystallized with ecotin A86H, each at 375μM, by the hanging drop method (McPherson, 1989) using PEG 4K as a precipitant and 100mM Tris pH 8.0/300mM NaOAc/10mM $CaCl_2$ as a buffer. Crystals sometimes appeared after 48hrs incubation, but more generally after 2 weeks to a month, with typical dimensions of 250x250x500μm. The crystals are monoclinic and belong to the space group C2, with one trypsin-ecotin complex in the asymmetric unit.

In addition to collecting data on a complex that had no metal present (apo-complex), the transition metal ions Cu^{+2}, Ni^{+2}, and Zn^{+2} were soaked into the

crystals to determine if the designed metal binding site bound each metal, and what the geometry and distances of the site were. The soaking protocol ramped the metal concentration up by 10-fold each time, and the crystals generally started to show fractures at metal concentrations of 100μM and higher. Due to the stress of the metal soaks, 20-50% of the crystals did not survive, and those that did were not always able to diffract the X-ray beam well enough to obtain a high resolution structure. As a result, several sets of data were collected in order to determine the X-ray stucture of each of the four cases: apo-, Cu-, Ni-, and Zn-complexes. A summary of the data collection and processing statistics is shown in Table 5.1. All four structures described belong to the monoclinic space group C_2 and have one trypsin-ecotin complex in the asymmetric unit.

Table 5.1: Data Collection and Reduction Statistics

Structure:	Apo	Copper	Nickel	Zinc
Unit Cell Parameters				
(Å) **a**=	85.15	86.35	86.38	85.98
(Å) **b**=	56.18	56.59	56.23	56.28
(Å) **c**=	80.68	81.40	81.23	81.86
β=	91.75°	92.99°	92.81°	92.18°
Max. resolution	1.8	2.3	2.0	2.2
Total Observations	86638	58971	56521	46301
Unique Observations	19067 (2σ)	16190 (1.5σ)	17281 (1.5σ)	10021 (1.5σ)
Completeness (%)	98.4	96.1	96.7	90.9
R_{merge} (%)	4.6	6.4	5.4	6.5

Refinement statistics are given in Table 5.2. These parameter values demonstrate the good quality of these structures.

Table 5.2: Refinement Statisitics

Structure:	Apo	Copper	Nickel	Zinc
Crystallographic R (%)	19.3	19.5	18.1	17.1
Free R (%)	28.5	27.7	25.6	26.7
<rms> bonds	0.005	0.006	0.006	0.005
<rms> angles	1.168	1.191	1.650	1.074

Copper, nickel, and zinc structures. Figures 5.1, 5.2, and 5.3 illustrate the local structure of copper, nickel, and zinc bound to the trypsin H143/151-ecotin H86 complex. The most obvious result is that each metal is bound to the designed site, and there is no evidence of non-specific binding of metal to other locations on the enzyme-inhibitor complex. While the histidines from the three structures are roughly superimposable, and the observed rotamers are nominal (Fig 5.4), there are some important differences in the local structure. Most notably, both the copper and nickel sites contain water as the fourth ligand to metal, albeit with long bonds of 3.0 Å and 2.6 Å, respectively. The zinc structure does not contain a water ligand. Also, the 143-151 loop in the zinc structure is well-defined, while in the apo, copper, and nickel structures, this loop is disordered from residue 146 to 149.

Figure 5.1: Copper Binding Site Structure

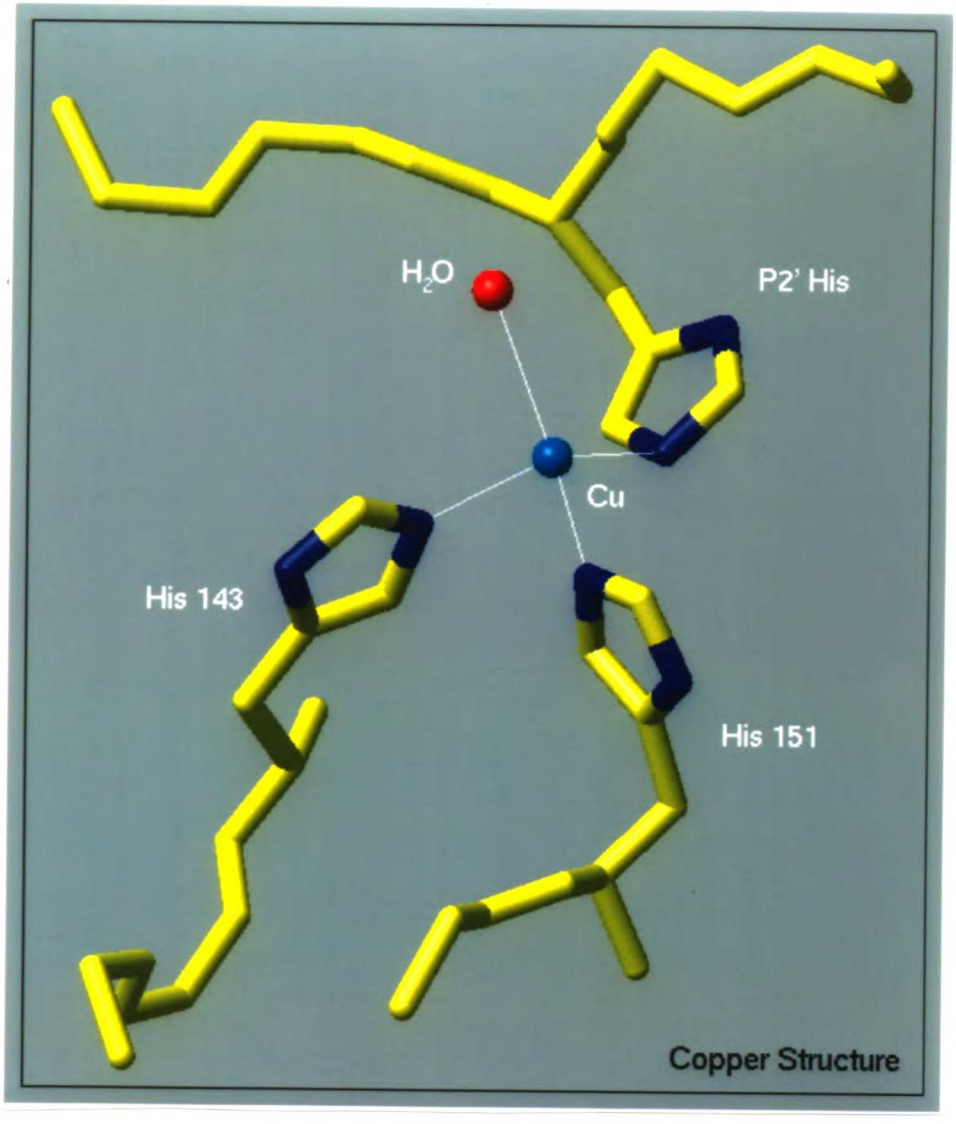


Figure 5.2: Nickel Binding Site Structure

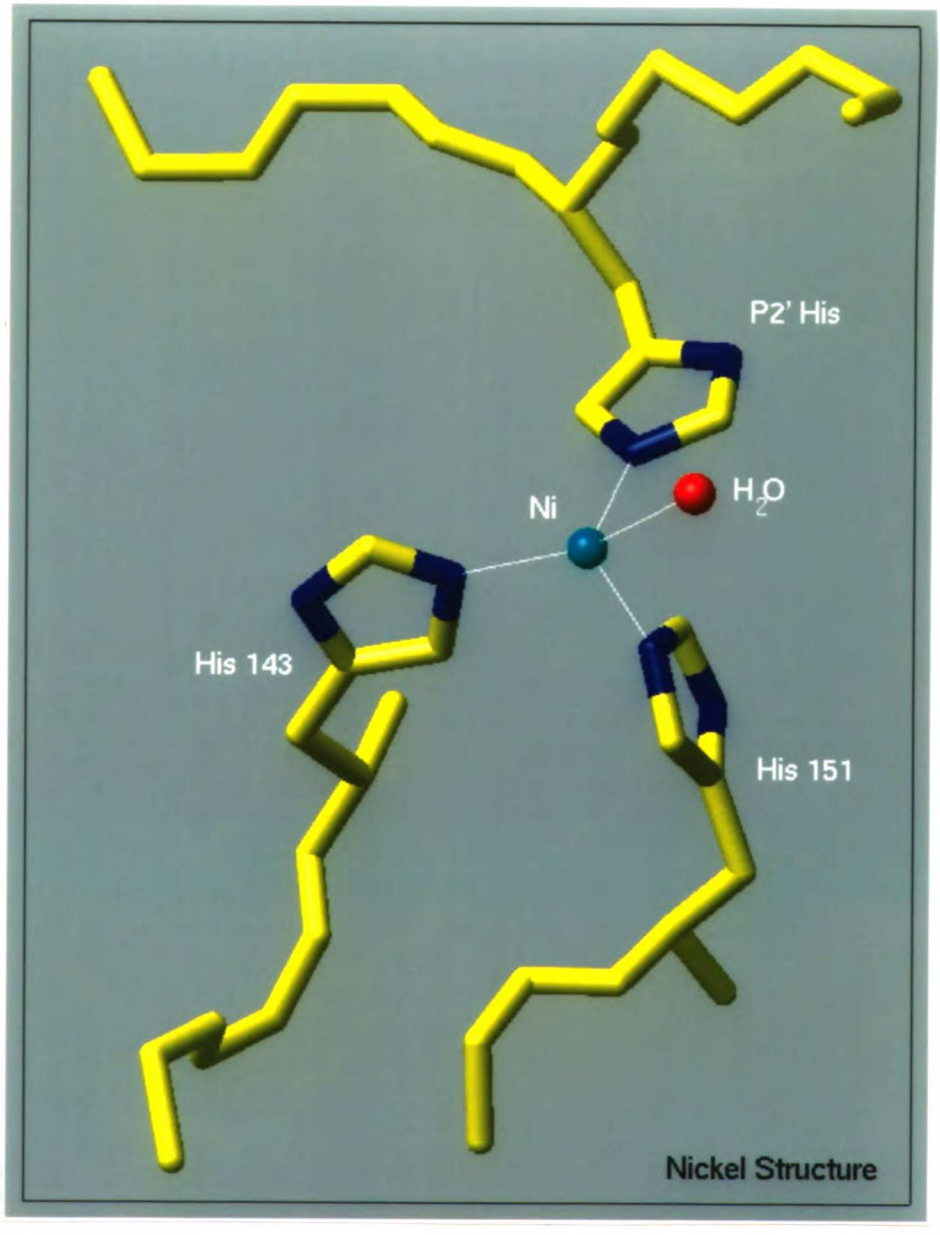


Figure 5.3: Zinc Binding Site Structure

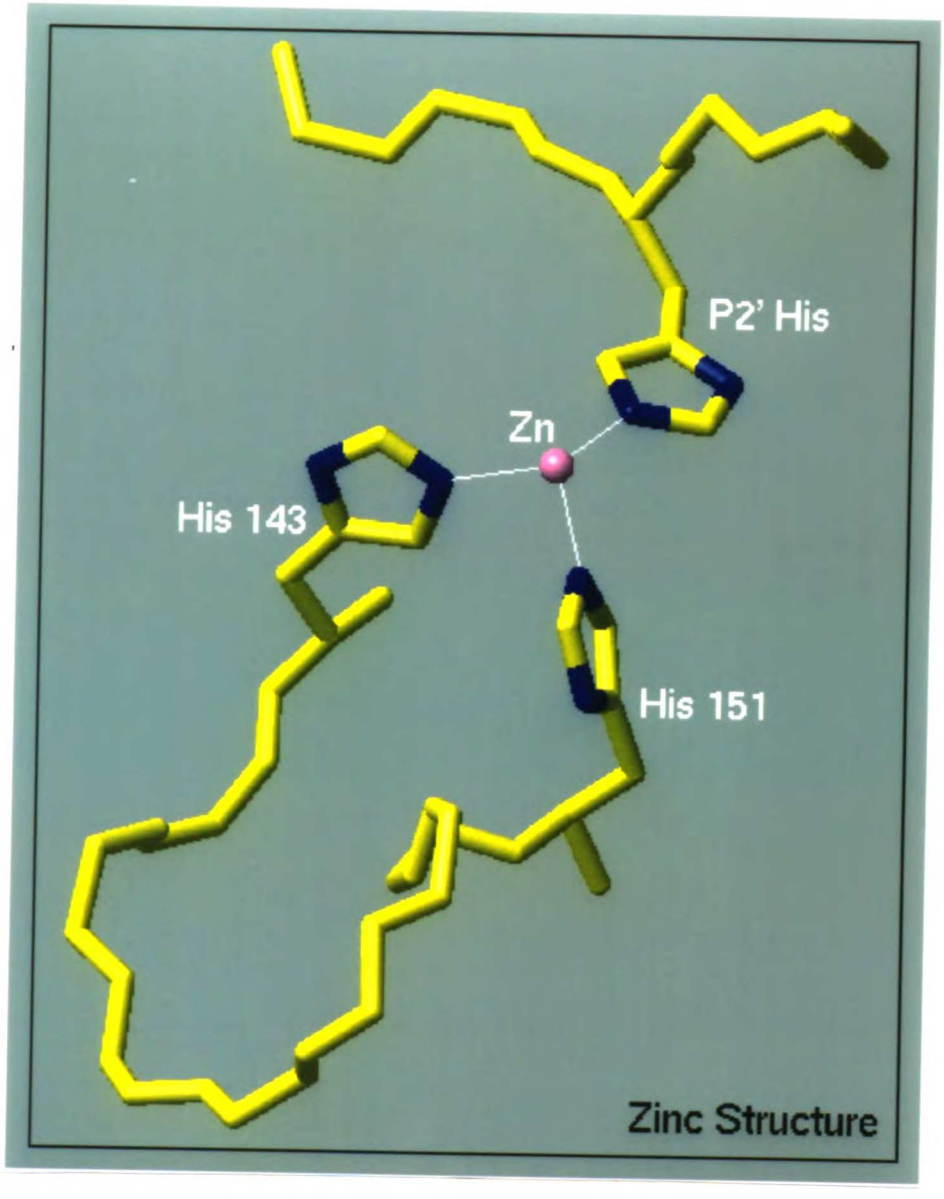
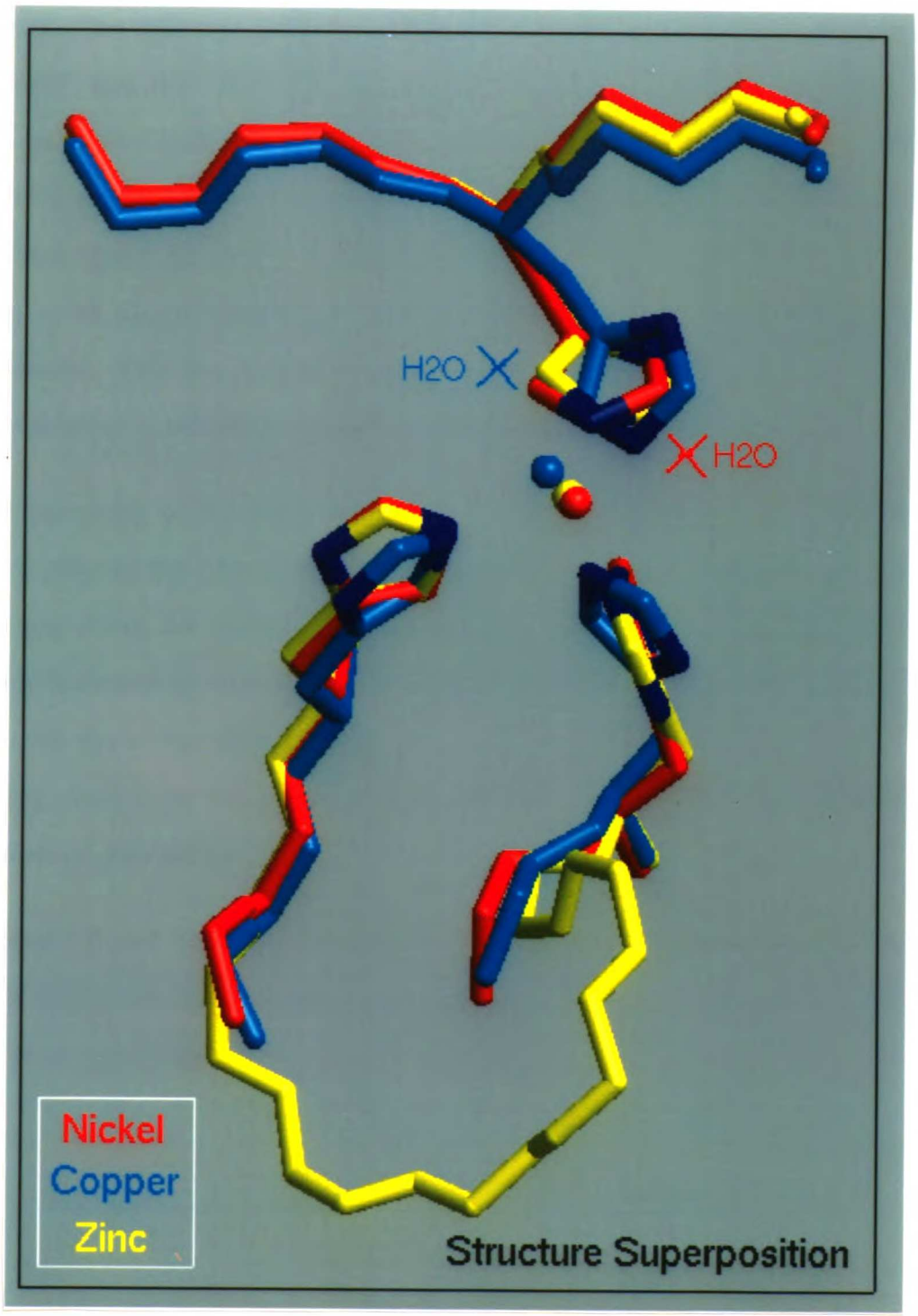


Figure 5.4: Superposition of Copper, Nickel, and Zinc Structures



Although none of the sites are exactly one geometry, viewing each structure in three dimensions with insight reveals the most closely related geometry. The zinc site is closest of the three to tetrahedral with bond angles of 93°, 106°, and 100° (Tab. 5.3). The copper structure is a distorted square planar geometry with the P_2' His Nε_2 on the opposite corner from the His143 Nε_2 (147°), and the water on the opposite corner from the H151 Nε_2 (131°). The nickel ligand geometry is neither tetrahedral, nor square planar, but appears to be square pyramidal, with the nickel ion at the center of the base of the pyramid. This is a geometry found in penta-coordinate complexes, but is also a subset of octahedral geometry with six ligands.

The planarity of the metal ions with the imidazole rings is an indication of the quality of the coordinate bond, as metal ions tend to bind in the plane of the ring along the direction of the nitrogen lone pair (Chakrabarti, 1991). The zinc is closest to coplanar with its three imidazole ligands (177°, 173°, and 165°), with the nickel close behind (180°, 170°, and 155°), and the copper deviating most from this planar preference (172°, 157°, 129°) (see table 5.3 for a definition of planarity).

Figure 5.5 and Table 5.3 summarize the quantitative differences in the metal binding site bond distances and angles, and in histidine rotamers found in the four structures.

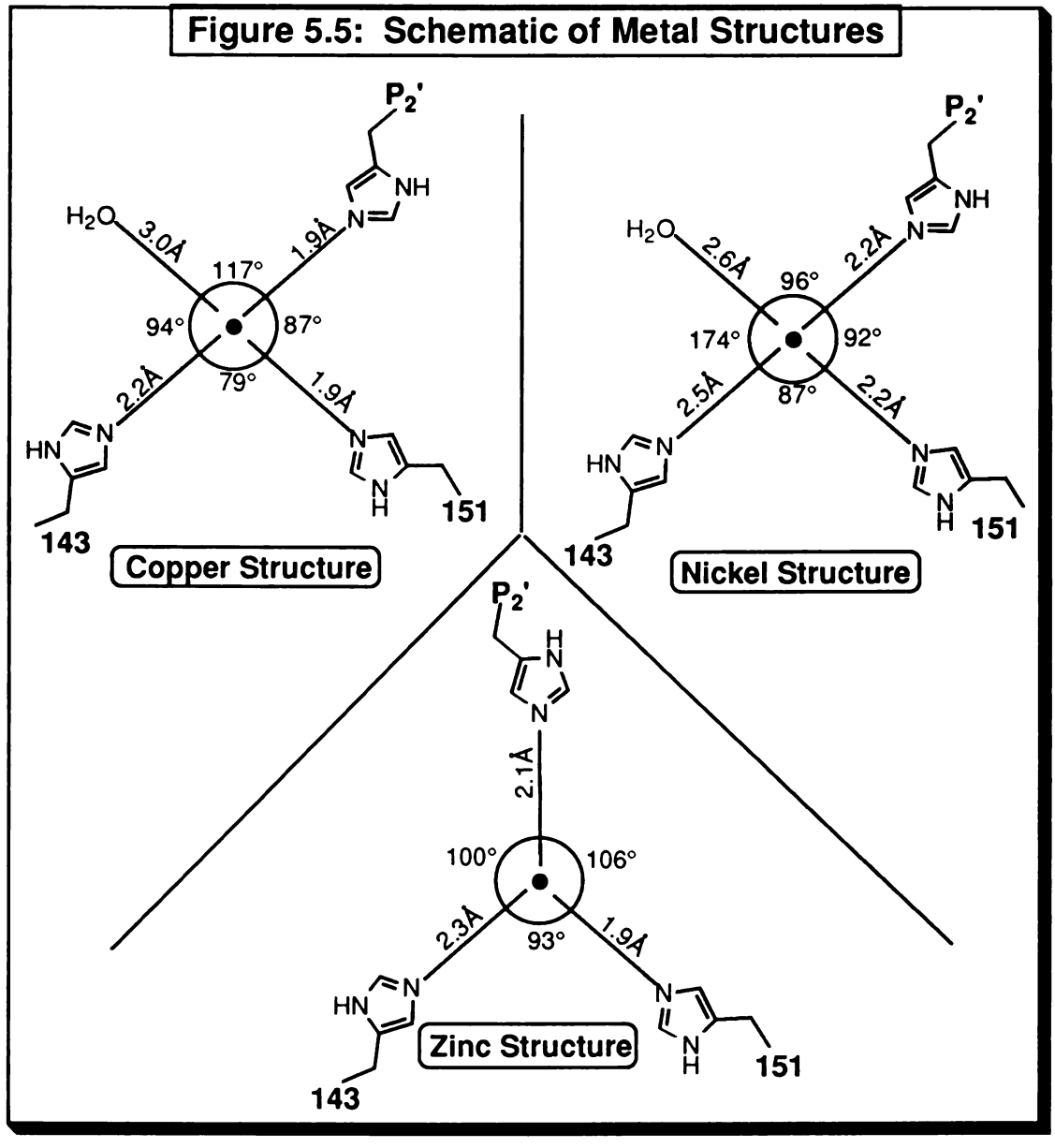


Figure 5.5: Schematic diagram of the three metal-bound structures of trypsin H143/151 complexed to ecotin A86H. The copper structure has distorted sqare planar geometry. The nickel structure has 4 out of 5 ligands of a square pyramidal geometry. The zinc structure has 3 out of 4 ligands of a distorted tetrahedral geometry. Note that in the copper and nickel structures the metal-water coordinate bond is the longest bond.

Table 5.3: Metal Binding Site Parameters

Parameter		Model	Apo	Copper	Nickel	Zinc
Distances (Å)[a]						
P2'His:Nε2-M^{+2}		1.9	-	1.9	2.2	2.1
H143:Nε2-M^{+2}		2.0	-	2.2	2.5	2.3
H151:Nε2-M^{+2}		1.6	-	1.9	2.2	1.9
H_2O:O-M^{+2}		-	-	3.0	2.6	-
Angles[b]						
P2'-Metal-H_2O		-	-	117°	96°	-
H_2O-Metal-143		-	-	94°	174°	-
143-Metal-151		119°	-	79°	87°	93°
151-Metal-P2'		123°	-	87°	92°	106°
P2'-Metal-143		105°	-	147°	78°	100°
H_2O-Metal-151		-	-	131°	91°	-
Metal-Imid. planarity[c]						
P2'His		160°	-	129°	-180°	-173°
His143		140°	-	172°	155°	165°
His151		-115°	-	157°	170°	-177°
His rotamers[d]						
P2'His:	χ_1	-175°	135°	-157°	-159°	-155°
	χ_2	-49°	-49°	-84°	-104°	-135°
His143:	χ_1	-70°	-68°	-61°	-62°	-64°
	χ_2	-175°	-123°	-170°	163°	172°
His151:	χ_1	-35°	172°	-54°	-94°	-71°
	χ_2	17°	149°	106°	-80°	91°

[a]Distances are measured between the Nε2 atom of the indicated histidine residue and the metal ion (M^{+2}). [b]Angles are measured between the Nε2 atoms of the indicated histidine residues and the metal ion. [c]The planarity of the metal ion with the imidazole ring of the histidine is measured with the dihedral angle formed by the Cγ, Cδ2, Nε2, and metal atoms. A dihedral angle of ±180° represents a metal ion that is coplanar

with the imidazole ring. $^{d}\chi_1$ is defined as the dihedral angle formed by the N, Cα, Cβ, and Cγ atoms of histidine, and χ_2 is formed by the Cα, Cβ, Cγ, and Nδ1 atoms. n/a=not applicable.

Comparison of model and zinc structures. The model for the design was presented previously (Willett, et al., 1995) and is shown again in Figure 5.5 with backbone atoms superimposed on the zinc structure. The model, shown in red, is made from a structure of BPTI bound to rat trypsin D189S while the zinc stucture is ecotin A86H bound to trypsin H143/151. At the top of the figure is the P1-P4' fragment of the bound inhibitor with a P2' histidine, and below the metal ion are His143, His151, and the surface loop that connects them. Two results are readily apparent. First, the modeled metal binding site and the resulting zinc structure are very similar, with an RMS deviation of 0.53Å for the Cα and Nε2 atoms of the 3 histidines. With this alignment, the positions of the metal atoms differ by 0.63Å. The second observation is that the conformation of the 143-151 loops differ radically; by 8-10Å at the widest separation. This is not suprising, given that the model was constructred from the BPTI-trypsin complex, and the crystallography was performed on ecotin-trypsin complexes. The histidine:Nε2-Zn bond distances in the zinc structure range from 1.9-2.3Å, and the geometry of the zinc binding site is distorted tetrahedral with the three observable angles ranging from 93-106° (Fig. 5.2). The model predicted 1.6-2.0Å bond distances from histidine:Nε2 to metal and distorted terahedral geometry as well, with bond angles being 105°, 119°, and 123°. In the zinc structure, the zinc is 7° ,15°, and 3° out of the plane of an imidazole ring for P2'His, His143, and His151, respectively. In contrast, the model predicted a less-optimal geometry, with the metal atom being 20°, 40°, and 65° out of the plane of the imidazole rings of the equivalent histidines.

Figure 5.6: Comparison of Model and Zinc Structures

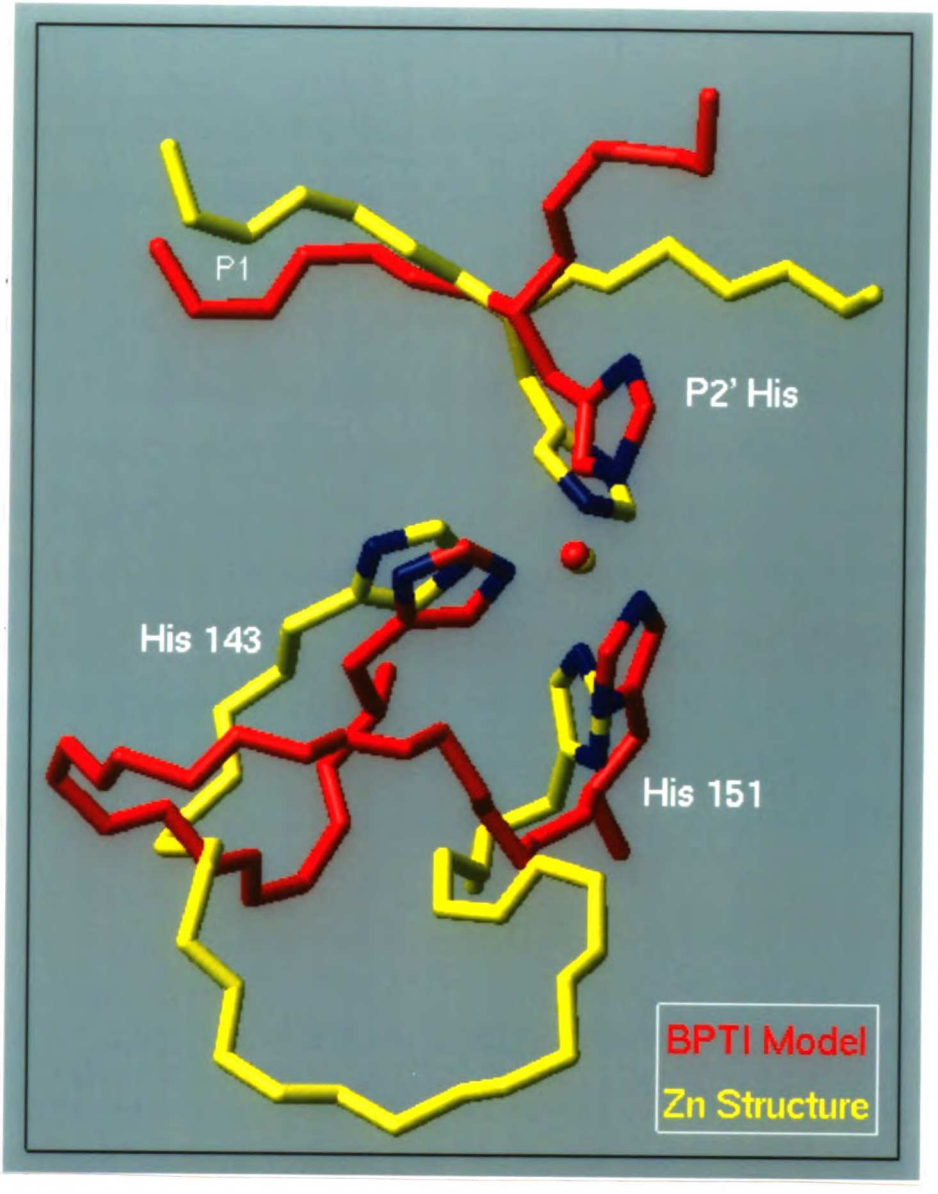
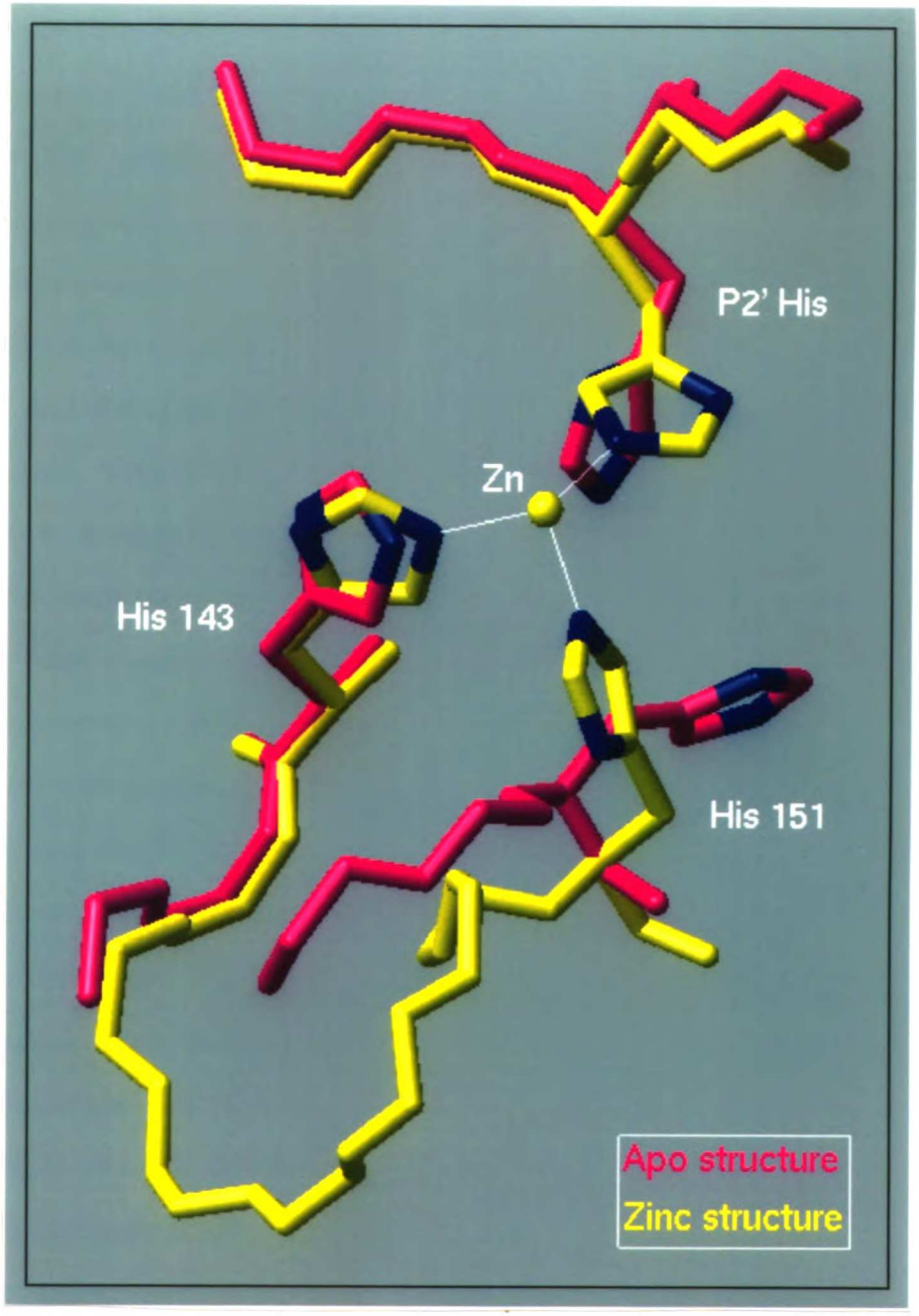


Figure 5.6. Superposition of BPTI model and trypsin H143151:ecotin H86 zinc structure. A striking similarity can be seen between the initial design and the actual crystallization result. The model is shown in red, and the zinc structure in yellow. The modeled and observed metal ions differ by only 0.63Å in position and the Cα and Nε2 atoms of the three histidines superimpose with a 0.53Å rms deviation. The most notable difference is observed between the 143-151 loops of the model and zinc structure, with an average displacement of 8-10Å.

Comparison of the apo and zinc structures. The zinc structure shows several notable differences from the apo structure, the most obvious of which is the coordination of the three histidines to the zinc atom (Fig. 5.6). Second, the electron density of the loop between 143 and 151 in the zinc structure is completely traceable, indicating a defined structure for this loop in the presence of zinc. In contrast, the electron density in the apo structure was not traceable from residues 146-148, indicating disordered local structure of this loop. Third, there is a substantial change in the χ_1 and χ_2 angles of His151 and the χ_2 angle of P2'-His compared to the apo structure (Table 5.3), resulting in a conformation able to bind zinc. And last, the backbone atoms from 149-151 have moved outward in the zinc structure by an average of 2Å, placing His151 in a position appropriate for binding zinc.

Figure 5.7. Superposition of apo and zinc structures highlighting structural changes of the complex upon binding of zinc. The apo structure is shown in pink and the zinc structure is shown in yellow, with the white lines representing the coordination of the zinc by the Nε2 of the histdines. Most notably, His151 rotates to participate in the zinc binding site, and the 143-151 loop moves and becomes ordered where it was not visible in the apo structure.

Figure 5.7: Comparison of Apo and Zinc Structures



There are also changes in the hydrogen bonding pattern of the three histidines in the zinc structure relative to the apo structure. In both the apo and zinc structures, the imidazole ring nitrogens of histidines 143 and 151 in trypsin and the P2'-His in ecotin make several possible hydrogen bond interactions with other atoms. In the apo structure, Nε2 of His143 is 3.0Å from the Oε1 of Gln192 and 3.3Å from the backbone carbonyl oxygen of Gln192, suggesting that the Nδ1 is protonated because the carbonyl oxygen can only participate as an H-bond acceptor. The Nε2 atom of His151 donates a hydrogen bond to two trypsin backbone carbonyl oxygens; one at 3.1Å to Trp141 and the other at 2.9Å to Pro152, suggesting that the Nδ1 of His151 is protonated. The P2'-His from ecotin of the apo structure is involved in three possible hydrogen bonds. Two are with a structured water at 2.9Å from Nδ1 and 3.7Å from Nε2, and the third is with the backbone carbonyl oxygen of Trp141 from trypsin at 3.7Å from Nε2. Given the long distances (>3.5Å) and poor geometry of these interactions, it is difficult to acertain the tautomeric form of His86 in the apo structure. In the zinc structure, Nε2 of His143, His151, and the P2'-His are all coordinated with zinc. Histidine143:Nε2 donates only one hydrogen bond to Gln192:Oε1 at 3.5Å where it made two in the apo structure. Histidine 151 has apparently flipped tautomers and now Nδ1 makes two hydrogen bonds: one to the backbone carbonyl oxygen of Val149 in trypsin at 2.7Å, and one to its own backbone N at 3.4Å. The P2'-His:Nδ1 in the zinc structure donates one hydrogen bond to the backbone carbonyl oxygen of His40 in trypsin.

Trypsin autolysis in crystals. When the first density maps of the trypsin-ecotin apo-complex were calculated, there was a region of trypsin from

residue 113-117 for which no electron density existed, suggesting that autolysis had occurred. Cleavage of the enzyme at positions 113 and 117 would generate three polypeptide fragments of 4, 97, and 143 amino acids in length. To confirm that suspicion, spare crystals were rinsed in artificial mother liquor and subjected to PAGE analysis. Although there was no detectable degradation product in the purified stock of enzyme, there was an autolysis product that must have been generated during the crystallization process which migrated at approximately 12kD on a reducing 12.5% acrylamide gel (data not presented). When the electrophoresis was repeated under non-reducing conditions, the 12kD band was not present, indicating that this trypsin fragment was bound to the rest of the molecule by one or more disulfide bonds, allowing the enzyme to maintain its stucture during crystallization. The structure of rat trypsin reveals 6 disulfide linkages connecting cysteines 22 to 157, 42 to 58, 128 to 232, 136 to 201, 168 to 182, and 191 to 220. One disulfide bond (22 to 157) links the Ile16-Lys133 fragment to the rest of the molecule. This 12kD fragment was blotted onto PVDF membrane and submitted for N-terminal sequencing. The N-terminal sequence data showed a major sequence of VATVA, confirming that autolysis was occurring at Arg117. There is a lysine at position 113 which could also have been cleaved. This autolysis presumably accounts for the lack of electron density for the segment 113-117.

Kinetic characterization of metal-assisted specificity. Metal-assisted histidine specificity was previously reported (Willett, et al., 1995) demonstrating that the engineered trypsin N143H/E151H (Tn H143/151) will hydrolyze tyrosine peptide bonds in the presence of nickel or zinc as long as a histidine is present at the P_2' position of the substrate. Because trypsin

normally cleaves Lys or Arg peptide bonds, the question remained as to whether a similar increase in activity would be observed for an Arg-X-His containing peptide. The peptide YLVGP**R-GH**FYDA, with an Arg at P_1 and a His at P_2', was purchased to address this question. HPLC analysis of the cleavage products (data not presented) show that 500μM copper, 500μM nickel, and 500μM zinc inhibit cleavage by 90%, 80%, and 20%±10%, respectively, as determined by percent hydrolysis of the peptide over 30 minutes. Trypsin E151Q, had been shown previously to be a non-metal-activated protease (Willett, et al., 1995), and was tested with the same substrate as a control experiment. Trypsin E151Q was inhibited by 500μM copper, 500μM nickel, and 500μM zinc by 90%, 20%, and 20%±10%, respectively. Similar experiments were performed using the synthetic fluorogenic substrate z-GPR-AMC which contains no amino acid residues C-terminal to the sissile bond. The results of these experiments show that copper inhibits trypsin H143/151 with this substrate by 80%, nickel by 30%, and zinc by 40%. The control protease, E151Q, is inhibited 50% by copper, 10% by nickel, and 25%±10% by zinc. These data show two main points: 1) nickel and zinc do not activate cleavage of an arginine peptide bond, rather, they seem to inhibit catalysis to a limited extent, and 2) copper non-specifically inhibits both proteases with both substrates 2-5 times more than do nickel or zinc.

Design improvement for metal-assisted specificity. Using structural information available to us, a trypsin was designed to exhibit increased turnover of a peptide containing the YAH and RGH sequences in the presence of metal. Previous studies in this lab (Perona et al., 1994b) have shown that removing the negative charge, Asp 189, in the S_1 subsite of trypsin and replacing it with the serine found in that position in

chymotrypsin (trypsin D189S) reduces the activity of the enzyme by 10^4 on synthetic tripeptide substrates. The mutation D189S was added to the N143H/E151H mutations to create trypsin N143H/E151H/D189S (H143/151/S189) with hopes of increasing the observed metal-activated hydrolysis of peptides. The design rationale was based on two main hypotheses: 1) removal of the negative charge at position 189 would increase metal-activated hydrolysis of a P_1-tyrosyl peptide bond in an extended peptide substrate containing a P_2' histidine; and 2) the activity towards an arginine substrate may be reduced enough so metal-activation of peptide hydrolysis containing a P_1 Arg and a P_2' His will be observed.

The results of the kinetic analysis of peptide hydrolysis are shown in figures 5.8 and 5.9. Figure 5.8 shows the metal titration of activity of trypsin H143/151/S189 against 800μM AGP**YAH**SS from 0 to 500μM metal as measured by the initial rate of hydrolysis, Vi (μM/min). Baseline hydrolysis without any metal averages 0.05μM/hr. Copper increases the activity 3-fold above baseline hydrolysis, and nickel by 10-fold. A dramatic 150-fold increase in activity was observed with 500μM zinc.

Figure 5.8
Tn H143/151/S189 Metal Titration of Activity

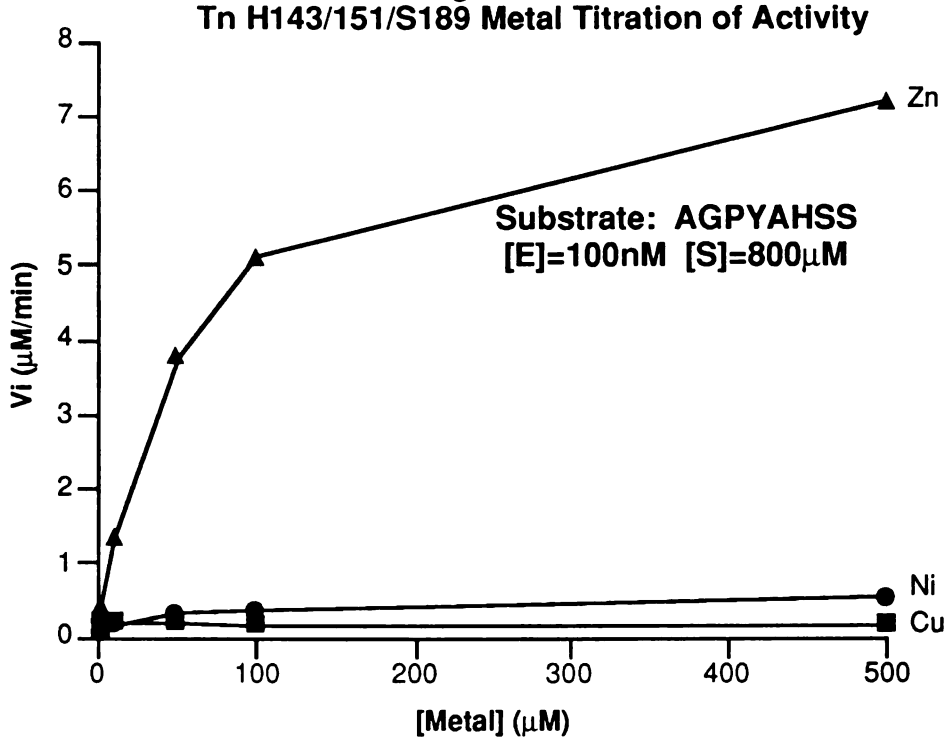


Figure 5.8: Copper, nickel and zinc titration of activity of trypsin H143/151/S189 against the peptide substrate AGPYAHSS. Activity is measured as the initial rate of hydrolysis, Vi (μM/min) at [E]=100nM, and [S]=800μM.

Figure 5.9 shows the relative activity for hydrolysis of AGP**Y-AH**SS and YLVGP**R-GH**FYDA by trypsin H143/151/S189 and a control protease, trypsin D189S, with and without 200 μMCu^{+2}, Ni^{+2}, or Zn^{+2}. Each enzyme exhibited a small amount of baseline activity, 0.001± 0.0005 μM/min, for hydrolysis of these peptides without metal. Figure 5.8a shows the relative kcat/Km for each of the enzymes against AGP**Y-AH**SS. Copper fails to increase the activity of either protease by a value greater than the error of the measurements. Nickel and zinc increase the activity of trypsin H143/151/S189 by 25-fold and 300-fold, respectively. The control protease, trypsin D189S, is unactivated by copper or nickel and only 2.5-fold by zinc. In addition, a control substrate

lacking the P_2' histidine, AGPY-**AA**SS, is only hydrolyzed at or below the baseline rate in all cases. Trypsin H143/151/S189 is activated by zinc 150-fold more relative to trypsin D189S. The kcat for hydrolysis of AGPY-**AH**SS by 200μM Ni^{+2} is 20±4.5 min^{-1} and the Km =1100±400 μM, yielding a kcat/Km of 0.020±0.004 $\mu M^{-1}min^{-1}$. With 200μM zinc, kcat=275±60 min^{-1}, and Km=1300±590μM, yielding a kcat/Km of 0.24±0.08 $\mu M^{-1}min^{-1}$. The 10-fold increase in activity with zinc relative to nickel is reflected in kcat, not Km. The zinc-activated kcat/Km of trypsin H143/151/S189 for this substrate is 65% of chymotrypsin activity on the same peptide, 0.37±.07$\mu M^{-1}min^{-1}$ (Willett, et al., 1995). This triple mutant represents a 150-fold improvement of metal-activated catalysis over the previous design, Tn H143/151.

Figure 5.9a:

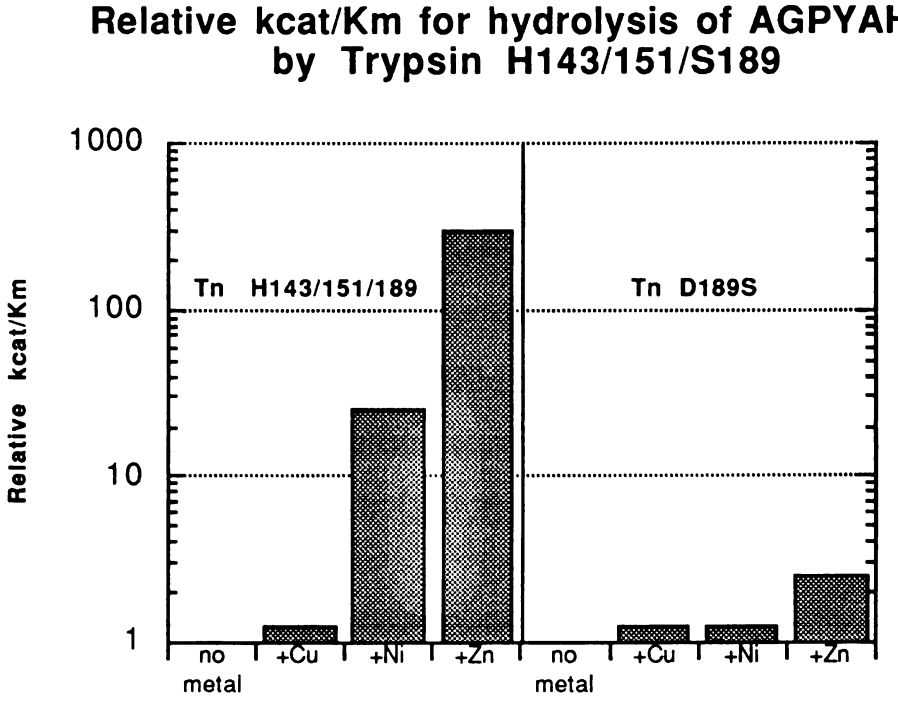
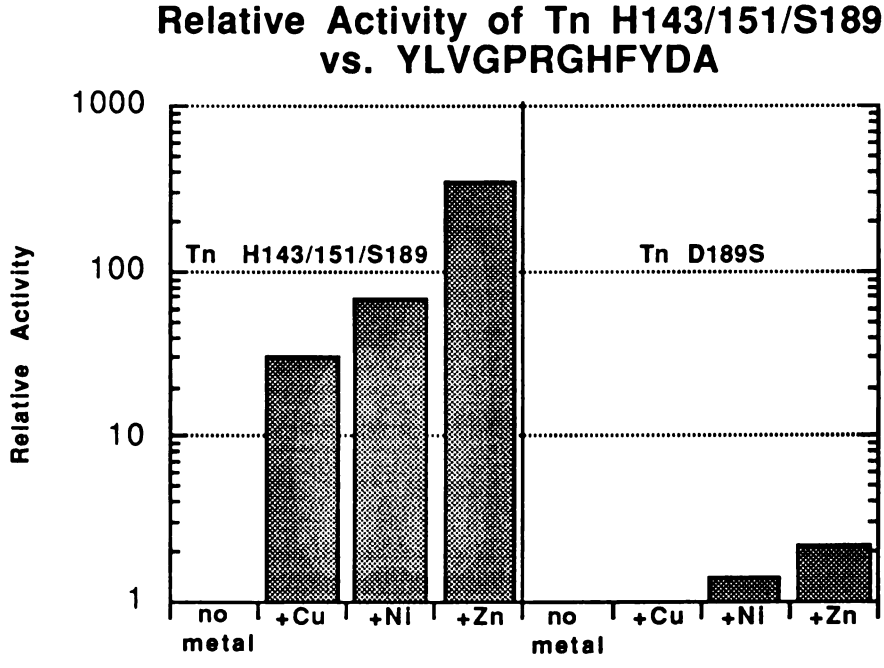


Figure 5.9. Relative activities of trypsin H143/151/S189 and a control protease, trypsin D189S, against AGP**YAH**SS (5.9a), and YLVGP**RGH**FYDA (5.9b). Values reported are ±25%. (a) Nickel activates hydrolysis of AGP**YAH**SS by trypsin H143/151/S189 by 25-fold, and zinc activates hydrolysis by 300-fold relative to baseline activity with copper, without metal, or by trypsin D189S. Only baseline activity was observed for a control substrate, AGPYAASS, lacking the P2' histidine. (b) Activity of trypsin H143/151/S189 against YLVGP**RGH**FYDA is sensitive to 200μM copper, nickel, and zinc showing 30-, 70-, and 350-fold increases in activity, respectively. The control protease, trypsin D189S, exhibited only baseline activity against YLVGP**RGH**FYDA (or slightly above) for each case.

Figure 5.9b illustrates the results of metal-activated hydrolysis of YLVGP**RGH**FYDA. Insufficient amounts of the peptide were available to determine kcat and Km values, so the relative activities are measured as the initial rate of hydrolysis (V_i) of the peptide under the given conditions. However, comparison of hydrolysis rates of YLVGP**RGH**FYDA and AGP**YAH**SS in the presence of 200μM zinc by Tn H143/151/S189 show that they are within 2-fold of each other. Furthermore, hydrolysis of YLVGP**RGH**FYDA by Tn H143/151/S189 plus 200μM zinc is within 10-fold of the activity observed by Tn H143/151 which possesses a wild-type S1 subsite. In contrast to AGP**YAH**SS, YLVGP**RGH**FYDA hydrolysis is activated by copper as well as nickel and zinc, whereas copper did not activate tyrosine hydrolysis. Copper activates activity 30-fold above background, nickel by 70-fold, and zinc by 350-fold. The control experiment using Tn D189S shows that none of the metals activate hydrolysis appreciably.

Figure 5.9b:



In the past, it has been shown that metal binding sites can be designed into proteins to stabilize structure, to regulate enzymatic activity (Corey and Schultz, 1989; Higaki, et al., 1990), and to modulate substrate specificity of a protease (Willett, et al., 1995). X-ray crystallography has been an effective tool for analyzing the resulting metal binding site (McGrath, et al., 1993). Crystallography has been used here to help explain the structural nature of an observed metal-dependent protease specificity of trypsin H143/151, and to suggest next-generation improvements of the design.

X-ray crystallographic results. In the absence of metal ion in the trypsin H143/151:ecotin A86H complex, the histidines comprising the engineered metal binding site are not in comformations that would bind a metal ion.

However, in the presence of copper, nickel or zinc, all three histidines form a metal binding site that closely resembles the model (Figs. 5.1-5.5). The geometry of the metal binding site is dependent on which metal is bound and seems to be dictated more by preferred geometry of the metal ion than the structural constraints of the protein. This result is consistent with the fact that the site was designed at the ends of a flexible surface loop in trypsin where it is conceivable for movements to occur with out much of an energetic barrier. Engineering on the solvent-accessible surface of the trypsin:ecotin (e.g. substrate) interface has proven successful in this case. In contrast, it is more difficult to design a metal binding site on the interior of a protein (Hellinga and Richards, 1991) where packing forces and extensive van der Waals contacts play a considerable role in the energetics of the site.

The Nε2-Zn bond lengths and geometry of the zinc binding site are closer to naturally occurring zinc binding sites than the model, with the most notable improvement being that the zinc is bound closer to the plane of the imidazole rings of the three histidines as observed for naturally occurring zinc binding sites (Christianson, 1991). In the model, the zinc averages 40° out of the plane of each imidazole ring, while in the zinc structure the average deviation from planarity is only 8° (Tab. 5.3). Although the model did not predict as optimal a site as is observed in the zinc structure, the design allowed for local movement of the metal binding site to accommodate a particular metal ion. This illustrates the advantage of designing a metal binding site in a solvent-accessible area on the protein (see Chapter 2: Metal Binding Site Design Considerations).

There is no well-defined electron density in the zinc structure for the water molecule that was modelled to be the fourth ligand to the metal ion. This is a suprising result because zinc is generally believed to be tetra-coordinated in aqueous solutions with tetrahedral geometry. An explanation is that there is a loosely bound water molecule, but that it is too disordered to observe at room temperature. Perhaps it would be visible at low temperature.

Design improvement for metal-assisted specificity. One of the advantages of using trypsin as a model system is the availability of high resolution structural as well as detailed kinetic information. We have used this information to make iterative design improvements, with the goal of increasing the metal-activated hydrolysis of the target substrate. Removal of the negative charge in the bottom of the pocket to create trypsin H143/151/S189 has resulted in a 150-fold improvement in hydrolysis rate of a YAH sequence. Trypsin H143/151/S189 is capable of cleaving a tyrosine substrate at 65% of the chymotrypsin-catalyzed rate. A 350-fold rate enhancement was observed in the presence of zinc for cleavage of a RGH sequence by trypsin H143/151/S189 relative to the rate observed with no metal present. This rate of hydrolysis is within 10-fold of the rate catalyzed by trypsin H143/151 which possesses a wild-type S_1 subsite. These results demonstrate that by utilizing both the S_1 and S_2' subsites on the enzyme as engineering targets, the substrate specificity of trypsin can be enhanced to require histidine at P_2', and to require zinc with minimal loss of catalytic efficiency. By crippling the S_1 subsite through removal of the negative charge at 189, and by adding a binding interatction at the P_2' subsite through a metal binding site, the design has in effect delocalized the primary specificity from the P_1-S_1 interaction to both the P_1-S_1 **and** P_2'-S_2' interactions. This may

allow for relatively similar binding energy contributions from both subsites instead of the majority of the binding energy coming from the primary binding pocket interaction. This may prove to be a general method for engineering protease specificity.

The order of activating potential of the three metal ions used, Zn>Ni>Cu can be explained partially by a combination of two factors: preferred geometry (See Tab. 2.1 and Fig. 2.1) (Glusker, 1991), and intrinsic affinity of metal ion for histidine (see Chapter 2) (Martell & Smith, 1974). Zinc prefers the tetrahedral arrangement of ligands that was designed into the site. Nickel prefers a square planar or sometimes tetrahedral geometry for tetra-coordinate complexes. Copper prefers a square planar geometry. The affinity for histidine of these metal ions is Cu>Ni>Zn. Catalytic efficiency directly correlates with the preference of the metal for tetrahedral geometry, and is inversely proportional to the affinity of the metal ion for histidine. These two factors taken together suggest an hypothesis for the observed kinetic effect. Given that the metal binding site is partially made by the leaving group of the acylation reaction in the course of the catalytic scheme (P_2'-His), it may be that the tighter binding metal creates a product-dissociation limiting case, while the metal ion that prefers tetrahedral geometry allows for the best alignment of the substrate, causing increased catalysis. Of course, the picture is more complicated because there are also metal-substrate, and metal-buffer equilibria, non-specific interactions of metal and enzyme, and protein dynamics to consider, but this simple explanation is interesting in light of the results. Analysis (*e.g.* spectroscopic) which could yield binding affinities of these metal ions to this designed site would help elucidate the determinants

of the observed kinetics, and lead towards an understanding of the mechanism of this metal-dependent catalysis.

Next-generation improvements of the design will include stabilizing His151 in the proper tautomer and conformation, so that an energetic price of side chain rotation does not have to be paid to bind metal ion. This may result in tighter binding of metal to the enzyme, leading to a more stable S_2' subsite for substrate P_2' histidine binding. Reorienting His151 may be accomplished by providing hydrogen bonding partners to His151:Nδ1 such as a carboxylate, or by redesigning the loop on which His151 resides to stabilize the desired conformation of His151.

The kinetics reported here, and the structures of this engineered metalloprotein demonstrate that simple geometric criteria can be used to design a metal binding site to effect a desired function. Still, these data only hint at a mechanism by which this metal binding site may function. In particular, the unexpected square-pyramidal geometry of the nickel site emphasizes the complex balance of forces between electronic stabilization of the metal ion (geometry), and energetic constraints of the local protein structure. The interplay of these two forces contributes to the selectivity of a given binding site for a particular metal ion. As such, this protease-inhibitor system may be useful for studying the nature of metal ion binding and selectivity in proteins.

Chapter 6

Low Hanging Fruit, Future Directions

Introduction

At this stage of the research, it is important to ask ourselves what we are learning, and if the goal of creating site-specific proteases is worthwhile. Continued experimentation in these areas is predicated upon the idea that we can increase our knowledge of fundamental aspects of protease-substrate recognition and catalysis, and/or can use these engineered proteases for practical applications. The results contained in this body of work provide a positive answer to the question and suggest experiments which can further advance the field of protein engineering. The future direction of this project can be broken down into several components: improvement of existing designs, combination of the two engineering methods (substrate-assisted catalysis and metal binding sites) in one protease, and application of these principles to other protease scaffolds. Undoubtedly, other methods for engineering specificity should and will be tried, but the work here points to several lines of experiments which can be pursued directly with the system that has been developed for trypsin. This "low-hanging fruit" will make a nice project(s) for students or post-docs entering the lab that wish to generate results quickly. In addition, these experiments can help build on our foundation of knowledge of protein structure function relationships adn provide the framework for more in-depth studies.

Substrate-assisted catalysis. In the case of substrate-assisted catalysis by trypsin H57A, there are several experiments already in progress that can be picked up and completed. First, crystals of trypsin H57A complexed with one

of two ecotin mutants have been obtained: ecotin M85H (mutant made by Dr. Ingrid Mehlhorn), and ecotin X83H/M84R. These two ecotin mutants were made to place a histidine in the P_1' and P_2 positions, respectively in order to act as structural models of subtrate-assisted catalysis. In retrospect, it would have been better to place an Arg in the P_1 position for both mutants, but that's for someone else to do. In addition, it may be useful to make an ecotin mutant with histidines at both P_2 and P_1' and an arginine at P_1, because that peptide yielded the highest turnover by trypsin H57A. At any rate, the crystal structures of these two trypsin-ecotin complexes may provide insight into the structural context of substrate-assisted catalysis by trypsin H57A. Specifically, the model predicts that $N\delta_1$ of the substrate His acts as the general base abstracting the proton from Ser195 and initiating catalysis. It will be interesting to see if this hypothesis is borne out in the crystal structure, or if the P_2 or P_1' His from substrate is ordered at all. Answers to these questions may provide direction for experiments aimed at improving the hydrolysis rate.

Kinetic studies of trypsin H57A with a series of synthetic substrates and peptides have shown that substrate-assisted catalysis will work in trypsin as long as the extended subsites are filled, and that a histidine can be donated from either the P_2 or P_1' position to yield the same catalytic efficiency. In all cases, the drop in catalytic efficiency is reflected in a reduced k_{cat}, not in a greatly increased K_m (Tab. 3.2). Studies on substrate-assisted catalysis by subtilisin BPN' H64A have provided insight into the mechanism of this protease and have suggested improvements for an increased rate of hydrolysis (Carter, et al., 1991). These improvements are centered on reconstituting the contribution of the catalytic carboxylate in the new

intermolecular triad. The trypsin H57A model shows a similarly compromised catalytic triad, suggesting experiments that would improve the catalytic efficiency of trypsin H57A should focus on optimizing the spatial orientation of the active-site residues relative to each other. The computer model with a P_1'-His substrate suggests that there is no stabilization of the histidine conformation or tautomer as is performed by Asp102 in the active site of trypsin (Fig. 3.1c). To that end, the obvious experiment to try is to imitate the role of Asp102 of trypsin with an amino acid either on the enzyme or on the substrate that could stabilize the correct tautomer and conformer of the P_1' catalytic histidine supplied by substrate. A computer search of the surface of trypsin in the region surrounding the substrate P_1' position may reveal a location for a carboxylate that would stabilize the catalytic substrate histidine. Perhaps a glutamate at position 102 might reach the substrate histidine at P_1'.

Similarly, the spatial arrangement of Asp 102 from trypsin H57A with a P_2-His is different from that found in the wild-type protease (Fig. 3.1a,b). This may be a cause for the observed decrease in k_{cat}. There is an opportunity for computer modeling, and for subsequent mutagenesis on trypsin or substrate to implement that idea. It could be as easy as adding the mutation D102E to H57A to allow the carboxylate to reach the imidazole ring of the P_1' His. The question is whether to wait for X-ray crystal structures of the trypsin-ecotin complexes, or to go ahead with the modeling and mutagenesis. Waiting for the crystal structure would adhere to the iterative approach we have used so far on this project, yet some simple modeling could get someone used to working with Insight II (Biosym, Inc.), and provide enough reason to go ahead with mutagenesis. This is a good example of the usefulness of

computer modeling. A model can be used to effectively rule out certain experiments due to obvious conflicts like steric overlap, etc., yet a model cannot reliably predict what will absolutely work in a protein engineering experiment. As a result, I have used computer modeling as a screen for reducing a list of possible mutations down to a set that is not *impossible* based on an X-ray structure.

Beyond a simple structure-based approach of modeling and mutating, a function-based approach may be more apporpriate for improving the rate of hydrolysis by trypsin H57A. Here, it should be noted that this approach can be used to engineer the substrate as well in order to achieve an optimized rate of hydrolysis (Matthews & Wells, 1993). Phage display technology could be used to engineer the protease as well as select for a substrate that is most labile to the enzyme (Corey et al., 1993). This strategy has already been used successfully to engineer macro molecular interactions of ecotin with urokinase (Wang et al., 1995). The design of a peptide library to make "substrate phage" should be relatively straight forward, but deciding which residues to mutate in trypsin may prove to be more difficult. Since catalytic efficiency seems to be related not just to binding of the substrate, but to binding in proper catalytic register (*i.e.* optimal alignment of the sissile bond with the catalytic machinery), mutated proteases will have to be screened against a series of substrates to find the best matches. This would be especially useful if one were designing a fusion protein junction meant for cleavage by trypsin H57A. In that manner, the optimal substrate could be engineered to match the engineered protease.

Another experiment which would enhance the usefulness of trypsin H57A would be showing that it is still active when immobilized to a solid support. In terms of downstream processing in the biotechnology industry, a required proteolytic processing step that can be performed with immobilized and reusable protease is an advantage. This method is preferred for its lower contamination of a process stream as well as the potential cost savings of a reusable protease "reactor". There are several commercially-available immobilization chemistries from which to choose, and the experiment would be easy to perform. Along these lines, we have at our disposal the techniques to add specific sequences to the protease that could aid in its directed covalent attachment to a solid support, such as adding lysines to the "back side" of the enzyme to increase the amount of protease coupled productively to a primary amine coupling reagent, such that the enzyme active site is available to the bulk fluid for catalysis. Demonstration of the ability to immobilize trypsin H57A either via a covalent attachment or via a His_6-tag on the C-terminus while retaining its activity would enhance its usefulness as an industrial and research enzyme.

One way to prove the generality of substrate-assisted catalysis as a protease engineering tool is to try it on various serine protease scaffolds. Fiddler crab collagenase (Tsu, et al, 1994) provides an excellent opportunity to test the hypothesis. In addition, because collagenase possesses a broad specificity, it may make a good candidate for engineering P_1 specificity. An engineered collagenase in combination with the H57A mutation may yield a useful protease with unique specificity. Only the experiment will be able to tell us if this will work.

Engineered metal binding sites. There are several opportunities for experimentation on this project. The state of results is still at the proof-of-principle stage where we have seen that a transition metal binding site can indeed be used to modify protease specificity. Now the challenge is to determine if this is a useful technology. The only substrates that have been cleaved with trypsin H143/151 or trypsin H143/151/S189 have been synthesized peptides designed to provide recognition and control sequences to the protease. These proteases have not been tested against large, globular proteins as substrates. It is expected that trypsin H143/151 will behave similarly to the wild-type enzyme as it still contains an intact S_1 site. However, trypsin H143/151/S189 may exhibit some useful cleavage patterns that are regulated by both metal ion identity (Ni^{+2}, Cu^{+2}, or Zn^{+2}) and concentration, because its S_1 site has been compromised by the D189S mutation. The obvious experiment to try is to hydrolyze a battery of globular proteins whose sequences are known in the presence or absence of various amounts of copper, nickel, and zinc, then to analyze the cleavage patterns under the various metal ion conditions. If it can be shown that hydrolysis is regulated by metal ion type and concentration, then an engineered metal binding site may prove to be a method for creating useful proteases whose acitivities are easily controlled. These proteases could be used for peptide or domain mapping.

Another use which has been successfully demonstrated for subtilisin is peptide ligation (Abrahmsen, et al., 1991) Since trypsin H143/151 has a metal-regulated specificity at P_2', it may also make a useful peptidyl ligase specific for histidines at the P_2' position. Thus a variant trypsin H143/151 could be part of a protein chemist's repertoire of enzymes used for building proteins from

peptides. Much more work would have to be performed to make this a practical use for an engineered metal binding site, but proof of principle should not be that difficult, and a simple HPLC assay could be used to demonstrate this alternative function of the enzyme.

A goal of this project has been to improve the engineered enzymes in an iterative fashion, using structural as well as functional information gathered in the process. Now that we have seen one round of iteration in design (Tn H143/151 —> Tn H143/151/S189) and a structural characterization of the enzyme with three bound metal ions, a logical next step is to improve the metal binding capability of the designed site on the enzyme. This can most likely be accomplished best by engineering the loop which contains H143 and H151 so that these two residues are "pre ordered" into the postitions appropriate for metal binding. Presumably, the entropic advantage gained by not having to order the histidine side chains upon metal binding will increase metal ion affinity of the site. There are two basic approaches to realize this goal: 1) mutate the loop so that the histidines are more likely to adopt the required metal-binding conformation, and 2) provide interactions that will stabilize the histidines in the positions observed in the bound metal state. Of these two approaches, the first is more nebulous as it is not clear how to design a loop to attain a desired conformation of any of its constituants. One strategy may be to introduce glycine residues surrounding the histidines to give more local flexibility, thus allowing the histidines to adopt their metal-binding conformations with less energetic strain on the enzyme. This is most appropriate for His151 which moves the most upon metal binding compared to the movement of His143 (Fig. 5.2). Although this

approach is naive, it may be viable in the absence of proven methods of loop modeling.

Phage display may be a method well suited to optimizing the metal binding capability of this loop by making a random library of mutations on the loop adjacent to the two histidines at positions 143 and 151 and screening for mutants that bind tightest to a copper-, nickel- or zinc-charged chelating sepharose column. The underlying hypothesis which suggests this experiment is that a better-designed bidentate site on the enzyme will bind metal tighter, meaning a slower off-rate, and the enzyme which binds metal tighter or has metal bound a greater percentage of time is more likely to bind a substrate with a P_2' His. In essence the idea is that if the enzyme is "pre-bound" with metal, the three-body problem requiring metal, enzyme, and substrate to come together is reduced to a two-body problem where the metalloenzyme binds substrate. In reality, this may not be possible because a bidentate site will have some maximum theoretical binding affinity for a given metal ion, and once that maximum is attained, a three-body problem may still be the case. Given the structural context of this particular design, the fourth ligand comes from solvent and not from the enzyme. Designing in the fourth ligand from the enzyme to create a tridentate binding site comprising enzyme ligands would require a substantial modeling effort which has not been the goal or scope of this work. However, this may be an attractive study for someone so inclined.

The phage display approach to modeling the metal-binding loop also has implications for a project aimed at studying the selectivity of a metal binding site for a particular metal ion. The nature of selectivity of metal binding sites

is only roughly understood (Glusker, 1991). It may be possible to screen for loop mutants which selectively bind one of the three transition metal ions used in this study. If a library of mutants were constructued which allowed the other metal-binding ligands (Asp, Glu, Cys, or Met) to occupy postitions 143 and/or 151 (or possibly another, better position on the loop), ligands other than histidine may be found which improve metal binding properties of the loop. An alternative approach analogous to the calcium binding loop placed into subtilisin BPN' (Braxton & Wells, 1992) would be to graft loops known to bind particular metal ions into the S_2' subsite of trypsin to see if their metal ion selectivity has been retained or not, or if they are able to function as substrate-binding sites. Subsequent kinetic and structural analyses of these mutants will provide insight into metal binding sites in proteins, in particular the determinants of an engineered metal binding site which enable it to productively bind substrate.

Another line of experiments which have been left open is spectroscopic characterization of the metal binding site. Electron Paramagnetic Resonace (EPR) (with Cu^{+2}) or NMR (with $^{113}Cd^{+2}$) experiments could yield interesting information concerning the dynamic nature of the metal binding site by allowing determination of metal-protein binding constants both in the presence and absence of bound ecotin A86H or a peptide substrate containing a P_2' His. A complex of trypsin H143/151/S189 with ecotin A86H would be suitable for such a study because of the low activity of that trypsin mutant. Even UV-vis spectroscopy can be used as a quick screen for bound ions such as copper. Experiments of this nature would complement the existing kinetic and structural studies nicely, and would give further insight into the metal binding properties of these proteins.

Combination of the two engineering methods. One of the goals of this project is to create proteases that are "site-specific", in a fashion analogous to restriction enzymes which exhibit extremely stringent specificity. One way to attempt this is to combine the two methods described in this dissertation: juxtapose substrate-assisted catalysis to the engineered metal binding site. Trypsin A57/H143/H151 or trypsin A57/H143/H151/S189 could easily be constructed, expressed, purified and tested against the appropriate peptide substrates. In theory, this protease should be specific for sequences that contain a histidine at either P_2 or P_1' *and* P_2', with either a Lys, Arg, or Tyr at P_1, depending on the particular S_1 subsite used (D189 or S189). Furthermore, the protease should be sensitive to transition metal ions. At first glance, this experiment seems like an "everything-but-the-kitchen-sink" type, but it would be interesting to see if the engineered specificities are additive or if each may detract from the other, resulting in a severly compromised enzyme. The most obvious problem is the complication of multiple histidines on a peptide in the presence of metal ions in solution. Such a system may exacerbate the problem due to the complex equilibria set up between hydrated metal ions and metal ions bound to enzyme, substrate, buffer, or combinations of all metal binding components. Still, it would be interesting to see if even more stringent specificity could be achieved by combining the two engineering methods.

Another approach is to utilize other subsites for engineering metal binding sites, perhaps on the N-terminal side of the sissile bond. Preliminary modeling of the P_3-S_3, and P_4-S_4 interactions suggests that an intermolecular bidentate site could be formed by changing residues Asn83, and/or Ala84 to histidines. This could feasibly generate metal-dependent P_3-, and P_4-His

specificity. This should also be considered with enzymes having a substantial extended binding subsite interactions such as crab collagenase.

Because the trypsin H143/151 and trypsin H143/151/S189 mutants are already purified and in the laboratory, more information may be gleaned by more detailed kinetic analysis of these proteases. Specifically, stopped-flow spectrometry experiments will be able to determine substrate on and off rates and possible support the hypothesis that the enzyme becomes product inhibited in the presence of metal ions due to tight binding of the leaving group. This set of experiments would add a depth to the project that would complement the structural work already undertaken.

Of course, all of the above experiments can be attempted in other proteases, or in other forms on the same protease. Although the trypsin-ecotin system is an excellent model system in which to test these hypotheses, the most likely candidate for another scaffold is crab collagenase because that clone is already being expressed recombinantly in the Craik lab, and its specificity and structure are becoming well characterized. Certainly there is a need for creativity; some of the most gratifying times in the laboratory come from seeing a casual idea start in the form of "what if. . .?" and take shape to become a body of work. Expanding these experiments to collagenase, or to other proteases that become the focus of the laboratory such as cruzain (Eakin et al., 1993) provides an excellent opportunity for training students and post doctoral fellows in the techniques of molecular biology and protein engineering. It also provides a powerful method for expanding our knowledge of the relationships between protein structure and function.

Abrahmsen, L., Tom, J., Burnier, J., Butcher, K. A., Kossiakoff, A. and Wells, J. A. (1991) Biochemistry 30, 4151-9.

Adman, E. T. (1979) *Biochim Biophys Acta* **549**, 107-44.

Adman, E. T., Stenkamp, R. E., Sieker, L. C. and Jensen, L. H. (1978) *J Mol Biol* **123**, 35-47.

Arnold, F. H. (1993) *Faseb J* **7**, 744-9.

Arnold, F. H. and Haymore, B. L. (1991) *Science* **252**, 1796-7.

Baker, E. N. and Hubbard, R. E. (1984) *Prog Biophys Mol Biol* **44**, 97-179.

Bass, S. H., Mulkerrin, M. G. and Wells, J. A. (1991) *Proc Natl Acad Sci U S A* **88**, 4498-502.

Bauer, C. A., Brayer, G. D., Sielecki, A. R. and James, M. N. (1981) *Eur J Biochem* **120**, 289-94.

Bayer, E. A., Grootjans, J. J., Alon, R. and Wilchek, M. (1990) *Biochemistry* **29**, 11274-9.

Bender, M. L. and Killheffer, J. V. (1973) *Crc Crit Rev Biochem* **1**, 149-99.

Berg, J. M. (1986) *Science* **232**, 485-7.

Bianchi, E., Sollazzo, M., Tramontano, A. and Pessi, A. (1993) *Int J Pept Protein Res* **41**, 385-93.

Bizzozero, S. A. and Dutler, H. (1987) *Arch Biochem Biophys* **256**, 662-76.

Bode, W. and Huber, R. (1976) *Febs Lett* **68**, 231-6.

Bode, W., Schwager, P. and Huber, R. (1978) *J Mol Biol* **118**, 99-112.

Bone, R. and Agard, D. A. (1991a) *Methods Enzymol* **202**, 643-71.

Bone, R. and Agard, D. A. (1991b) *Methods Enzymol* **202**, 643-71.

Bone, R., Fujishige, A., Kettner, C. A. and Agard, D. A. (1991a) *Biochemistry* **30**, 10388-98.

Bone, R., Fujishige, A., Kettner, C. A. and Agard, D. A. (1991b) *Biochemistry* **30**, 10388-98.

Bone, R., Silen, J. L. and Agard, D. A. (1989) *Nature* **339**, 191-5.

Braxton, S. and Wells, J. A. (1991) *J Biol Chem* **266**, 11797-800.

Braxton, S. and Wells, J. A. (1992) *Biochemistry* **31**, 7796-801.

Brown, I. D. (1988) *Acta Crystallographica Section B* **44**, 545-53.

Brown, S. (1992) *Proc Natl Acad Sci U S A* **89**, 8651-5.

Browner, M. F., Hackos, D. and Fletterick, R. J. (1994) *Nature Structural Biology* **1**, 327-333.

Brunger, A. T., Huber, R. and Karplus, M. (1987) *Biochemistry* **26**, 5153-62.

Carter, P., Abrahmsen, L. and Wells, J. A. (1991) *Biochemistry* **30**, 6142-8.

Carter, P., Nilsson, B., Burnier, J. P., Burdick, D. and Wells, J. A. (1989) *Proteins* **6**, 240-8.

Carter, P. and Wells, J. A. (1987) *Science* **237**, 394-9.

Chakrabarti, P. (1990a) *Protein Eng* **4**, 57-63.

Chakrabarti, P. (1990b) *Protein Eng* **4**, 49-56.

Chambers, J. L., Christoph, G. G., Krieger, M., Kay, L. and Stroud, R. M. (1974) *Biochemical adn Biophysical Research Communications* **59**, 70-74.

Chang, K. H., Liang, P. H., Beck, W., Scholten, J. D. and Dunaway-Mariano, D. (1992) *Biochemistry* **31**, 5605-10.

Chen, K. and Arnold, F. H. (1993) *Proc Natl Acad Sci U S A* **90**, 5618-22.

Chicz, R. M. and Regnier, F. E. (1989) *Anal Chem* **61**, 1742-9.

Christianson, D. W. (1991) *Adv Protein Chem* **42**, 281-355.

Christianson, D. W. and Alexander, R. S. (1989) *Journal of the American Chemical Society* **111**, 6412-19.

Clark, I. D., MacManus, J. P., Banville, D. and Szabo, A. G. (1993) *Anal Biochem* **210**, 1-6.

Collman, J. P. and Hegedus, L. S. (1980), "Principles and applications of organotransition metal chemistry" University Science Books, Mill Valley, California

Corey, D. R. and Craik, C. S. (1992) *J. Am. Chem. Soc.* **114**, 1784-90.

Corey, D. R. and Craik, C. S. (1993) *in* "Proteases and their Inhibitors; Fundamental and Applied Aspects" (Aviles, F. X., eds) Vol. pp. 425-44, Walter DeGruyer,

Corey, D. R., McGrath, M. E., Vasquez, J. R., Fletterick, R. J. and Craik, C. S. (1992) *Journal of the American Chemical Society* **114**, 4905-7.

Corey, D. R. and Schultz, P. G. (1989) *J Biol Chem* **264**, 3666-9.

Corey, D. R., Shiau, A. K., Yang, Q., Janowski, B. A. and Craik, C. S. (1993) *Gene* **128**, 129-34.

Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R. and Rutter, W. J. (1985) *Science* **228**, 291-7.

Craik, C. S., Roczniak, S., Sprang, S., Fletterick, R. and Rutter, W. (1987) *J Cell Biochem* **33**, 199-211.

Creemers, J. W., Siezen, R. J., Roebroek, A. J., Ayoubi, T. A., Huylebroeck, D., Van, de, Ven and Wj (1993) *J Biol Chem* **268**, 21826-34.

Cronin, C. N., Malcolm, B. A. and Kirsh, J. F. (1987) *J Am Chem Soc* **109**, 2222.

Cunningham, B. C., Mulkerrin, M. G. and Wells, J. A. (1991) *Science* **253**, 545-8.

Cunningham, B. C. and Wells, J. A. (1987) *Protein Eng* **1**, 319-25.

Cunningham, B. C. and Wells, J. A. (1991) *Proc Natl Acad Sci U S A* **88**, 3407-11.

Derewenda, Z. S., Derewenda, U. and Kobos, P. M. (1994) *J Mol Biol* **241**, 83-93.

Dewar, M. J. S., Harget, A. J. and Trinagstic, N. (1969) *Journal of the American Chemical Society* **91**, 6321.

Drapeau, G. R. (1978) *Can J Biochem* **56**, 534-44.

Eakin, A. E., McGrath, M. E., McKerrow, J. H., Fletterick, R. J. and Craik, C. S. (1993) *J Biol Chem* **268**, 6115-8.

Edwards, S. L., Nguyen, H. X., Hamlin, R. C. and Kraut, J. (1987) *Biochemistry* **26**, 1503-11.

Erhart, E. and Hollenberg, C. P. (1983) *J Bacteriol* **156**, 625-35.

Erpel, T., Hwang, P., Craik, C. S., Fletterick, R. J. and McGrath, M. E. (1992) *J Bacteriol* **174**, 1704.

Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G. and Wells, J. A. (1986) *Science* **233**, 659-63.

Evans, D. B., Tarpley, W. G. and Sharma, S. K. (1991) *Protein Expr Purif* **2**, 205-13.

Evnin, L. B., Vasquez, J. R. and Craik, C. S. (1990) *Proc Natl Acad Sci U S A* **87**, 6659-63.

Falke, J. J., Snyder, E. E., Thatcher, K. C. and Voertler, C. S. (1991) *Biochemistry* **30**, 8690-7.

Fehlhammer, H., Bode, W. and Huber, R. (1977) *J Mol Biol* **111**, 415-38.

Fersht, A. (1985), "Enzyme Structure and Mechanism" W. H. Freeman and Co., New York

Fersht, A. R., et al. (1985) *Nature* **314**, 235-8.

Fisher, C. L., Hallewell, R. A., Roberts, V. A., Tainer, J. A. and Getzoff, E. D. (1991) *Free Radic Res Commun* **1**, 287-96.

Ford, C. F., Suominen, I. and Glatz, C. E. (1991) *Protein Expr Purif* **2**, 95-107.

Forsberg, G., Baastrup, B., Rondahl, H., Holmgren, E., Pohl, G., Hartmanis, M. and Lake, M. (1992) *J Protein Chem* **11**, 201-11.

Forsberg, G., Brobjer, M., Holmgren, E., Bergdahl, K., Persson, P., Gautvik, K. M. and Hartmanis, M. (1991) *J Protein Chem* **10**, 517-26.

Fujinaga, M., Delbaere, L. T., Brayer, G. D. and James, M. N. (1985) *J Mol Biol* **184**, 479-502.

Fuller, R. S., Brake, A. and Thorner, J. (1989) *Proc Natl Acad Sci U S A* **86**, 1434-8.

Garavito, R. M., Rossmann, M. G., Argos, P. and Eventoff, W. (1977) *Biochemistry* **16**, 5065-71.

Gardell, S. J., Craik, C. S., Clauser, E., Goldsmith, E. J., Stewart, C. B., Graf, M. and Rutter, W. J. (1988) *J Biol Chem* **263**, 17828-36.

Gaykema, W. P. J., Jol, W. G. J., Vereijken, J. M., Soeter, N. M., Bak, J. J. and Beintema, J. J. (1984) *Nature* **309**, 23-9.

Ghadiri, M. R. and Choi, C. (1990) *J. A. C. S.* **112**, 1630-2.

Ghadiri, M. R. and Fernholz, A. K. (1990) *J Am. Chem. Soc.* **112**, 9633-9635.

Ghadiri, M. R., Soares, C. and Choi, C. (1992) *J. A. C. S.* **114**, 825-31.

Glusker, J. P. (1991) *Adv Protein Chem* **42**, 1-76.

Graf, L., Craik, C. S., Patthy, A., Roczniak, S., Fletterick, R. J. and Rutter, W. J. (1987) *Biochemistry* **26**, 2616-23.

Graf, L., Hegyi, G., Liko, I., Hepp, J., Medzihradszky, K., Craik, C. S. and Rutter, W. J. (1988a) *Int J Pept Protein Res* **32**, 512-8.

Graf, L., Jancso, A., Szilagyi, L., Hegyi, G., Pinter, K., Naray-Szabo, G., Hepp, J., Medzihradszky, K. and Rutter, W. J. (1988b) *Proc Natl Acad Sci U S A* **85**, 4961-5.

Grant, G. A. and Eisen, A. Z. (1980) *Biochemistry* **19**, 6089-95.

Greer, J. (1990) *Proteins* **7**, 317-34.

Gron, H., Meldal, M. and Breddam, K. (1992) *Biochemistry* **31**, 6011-8.

Haber, E., Quertermous, T., Matsueda, G. R. and Runge, M. S. (1989) *Science* **243**, 51-6.

Handel, T. and Degrado, W. F. (1990) *J. Am. Chem. Soc.* **112**, 6710-11.

Handel, T. M., Williams, S. A. and DeGrado, W. F. (1993) *Science* **261**, 879-85.

Hanson, S., Adelman, J. and Ullman, B. (1992) *J Biol Chem* **267**, 2350-9.

Hedstrom, L., Farr, J. S., Kettner, C. A. and Rutter, W. J. (1994a) *Biochemistry* **33**, 8764-9.

Hedstrom, L., Perona, J. J. and Rutter, W. J. (1994b) *Biochemistry* **33**, 8757-63.

Hedstrom, L., Szilagyi, L. and Rutter, W. J. (1992) *Science* **255**, 1249-53.

Hein, G. E. (1966) *in* "Kekulé Centennial: Symposium of the American Chemical Society" (Gould, R. F., eds) Vol. pp. chapter 1, Am. Chem. Soc., Washington D. C.

Hellinga, H. W., Caradonna, J. P. and Richards, F. M. (1991) *J Mol Biol* **222**, 787-803.

Hellinga, H. W. and Richards, F. M. (1991) *J Mol Biol* **222**, 763-85.

Herzberg, O. and James, M. N. (1985a) *Biochemistry* **24**, 5298-302.

Herzberg, O. and James, M. N. (1985b) *Nature* **313**, 653-9.

Higaki, J. N., Evnin, L. B. and Craik, C. S. (1989) *Biochemistry* **28**, 9256-63.

Higaki, J. N., Fletterick, R. J. and Craik, C. S. (1992) *Trends Biochem Sci* **17**, 100-4.

Higaki, J. N., Haymore, B. L., Chen, S., Fletterick, R. J. and Craik, C. S. (1990) *Biochemistry* **29**, 8582-6.

Holmes, M. A. and Matthews, B. W. (1982) *J Mol Biol* **160**, 623-39.

Huber, R. and Bode, W. (1978) *Acc Chem Res* **11**, 114-22.

Ibers, J. A. and Holm, R. H. (1980) *Science* **209**, 223-35.

Ippolito, J. A. and Christianson, D. W. (1994) *Biochemistry* **33**, 15241-9.

Irving, H. and Williams, R. J. P. (1953) *Journal of the Chemical Society* 3192.

Iverson, B. L., Iverson, S. A., Roberts, V. A., Getzoff, E. D., Tainer, J. A., Benkovic, S. J. and Lerner, R. A. (1990) *Science* **249**, 659-62.

Iverson, B. L. and Lerner, R. A. (1989) *Science* **243**, 1184-8.

Jabri, E., Carr, M. B., Hausinger, R. P. and Karplus, P. A. (1995) *Science* **268**, 998-1004.

Jameson, G. W., Roberts, D. V., Adams, R. W., Kyle, W. S. and Elmore, D. T. (1973) *Biochem J* **131**, 107-17.

Janda, K. D., Schloeder, D., Benkovic, S. J. and Lerner, R. A. (1988) *Science* **241**, 1188-91.

Jenkins, J., et al. (1992) *Biochemistry* **31**, 5449-58.

Jones, T. A. (1978) *Journal of Applied Crystallography* **11**, 268-272.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G. and Davies, D. R. (1960) *Nature* **185**, 422-7.

Kiefer, L. L., Krebs, J. F., Paterno, S. A. and Fierke, C. A. (1993) *Biochemistry* **32**, 9896-900.

Kitamoto, Y., Yuan, X., Wu, Q., McCourt, D. W. and Sadler, J. E. (1994) *Proc Natl Acad Sci U S A* **91**, 7588-92.

Klapper, M. H. (1977) *Biochem Biophys Res Commun* **78**, 1018-24.

Klemba, M., Gardner, K. H., Marino, S., Clarke, N. D. and Regan, L. (1995) *Structural Biology* **2**, 368-73.

Koert, U., Harding, M. M. and Lehn, J. M. (1990) *Nature* **346**, 339-42.

Kossiakoff, A. A., Chambers, J. L., Kay, L. M. and Stroud, R. M. (1977) *Biochemistry* **16**, 654-64.

Kunkel, T. A. (1985) *Proc Natl Acad Sci U S A* **82**, 488-92.

Kuroki, R., Taniyama, Y., Seko, C., Nakamura, H., Kikuchi, M. and Ikehara, M. (1989) *Proc Natl Acad Sci U S A* **86**, 6903-7.

LaVallie, E. R., Rehemtulla, A., Racie, L. A., DiBlasio, E. A., Ferenz, C., Grant, K. L., Light, A. and McCoy, J. M. (1993) *J Biol Chem* **268**, 23311-7.

Lee, H. R., et al. (1991) *Febs Lett* **287**, 53-6.

Lewis, G. N. (1923), "Valence and the Structure of Atoms and Molecules" Chemical Catalog Co., New York

Liang, P. H., Yang, G. and Dunaway-Mariano, D. (1993) *Biochemistry* **32**, 12245-50.

Liao, D. I., Breddam, K., Sweet, R. M., Bullock, T. and Remington, S. J. (1992) *Biochemistry* **31**, 9796-812.

Liao, D. I. and Remington, S. J. (1990) *J Biol Chem* **265**, 6528-31.

Lieberman, M. and Sasaki, T. (1991) *J. Am. Chem. Soc.* **113**, 1470-1.

Light, A., Savithri, H. S. and Liepnieks, J. J. (1980) *Anal Biochem* **106**, 199-206.

Liu, L. W., Vu, T. K., Esmon, C. T. and Coughlin, S. R. (1991) *J Biol Chem* **266**, 16977-80.

Lowman, H. B., Bass, S. H., Simpson, N. and Wells, J. A. (1991a) *Biochemistry* **30**, 10832-8.

Lowman, H. B., Cunningham, B. C. and Wells, J. A. (1991b) *J Biol Chem* **266**, 10982-8.

Lu, Y., Casimiro, D. R., Bren, K. L., Richards, J. H. and Gray, H. B. (1993) *Proc Natl Acad Sci U S A* **90**, 11456-9.

Mack, D. P., Iverson, B. L. and Dervan, P. B. (1988) *J Am Chem Soc* **110**, 7572-4.

Madison, E. L., Coombs, G. S. and Corey, D. R. (1995) *Journal of Biological Chemistry* **270**, 7558-62.

Maroux, S., Baratti, J. and Desnuelle, P. (1971) *J Biol Chem* **246**, 5031-9.

Martell, A. D. and Smith, R. M. (1974), "Critical Stability Constants" Plenum Press, New York

Matthews, B. W. (1977) *in* "Proteins" (Neurath, H. and Hill, R. L., eds) Vol. 3, pp. 404-590, Academic Press, New York

Matthews, D. J. and Wells, J. A. (1993) *Science* **260**, 1113-7.

McGrath, M. E., Erpel, T., Browner, M. F. and Fletterick, R. J. (1991) *J Mol Biol* **222**, 139-42.

McGrath, M. E., Erpel, T., Bystroff, C. and Fletterick, R. J. (1994) *The EMBO Journal* **13**, 1502-1507.

McGrath, M. E., Haymore, B. L., Summers, N. L., Craik, C. S. and Fletterick, R. J. (1993) *Biochemistry* **32**, 1914-9.

McGrath, M. E., Hines, W. M., Sakanari, J. A., Fletterick, R. J. and Craik, C. S. (1991) *J Biol Chem* **266**, 6620-5.

McPherson, A. (1989), "Preparation and Analysis of Protein Crystals" Robert E Kreiger Pub. Co., Malabar, FL

Merkle, D. L. and Berg, J. M. (1991) *Methods Enzymol* **208**, 46-54.

Mildvan, A. S. (1987) *Magnesium* **6**, 28-33.

Muchmore, D. C., McIntosh, L. P., Russel, C. B., Anderson, D. E. and Dahlquist, F. W. (1989) *Methods In Enzymology* **177B**, 44-73.

Murali, C. and Creaser, E. H. (1986) *Protein Eng* **1**, 55-7.

Nagai, K. and Thogersen, H. C. (1987) *Methods Enzymol* **153**, 461-81.

Narayana, S. V., Carson, M., el, K. O., Kilpatrick, J. M., Moore, D., Chen, X., Bugg, C. E., Volanakis, J. E. and DeLucas, L. J. (1994) *J Mol Biol* **235**, 695-708.

Navaza, J. (1994) *Acta Cryst. A* **50**, 157-63.

Neurath, H. (1985) *Fed Proc* **44**, 2907-13.

Norris, G. E., Anderson, B. F. and Baker, E. N. (1983) *J Mol Biol* **165**, 501-21.

Pal, G., Sprengel, G., Patthy, A. and Graf, L. (1994) *Febs Lett* **342**, 57-60.

Paoni, N. F., Chow, A. M., Pena, L. C., Keyt, B. A., Zoller, M. J. and Bennett, W. F. (1993) *Protein Eng* **6**, 529-34.

Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. and Dougherty, W. G. (1994) *Anal Biochem* **216**, 413-7.

Pearson, R. G. (1966) *Science* **151**, 172-7.

Pearson, R. G. (1968a) *J Chem Ed* **45**, 581-7.

Pearson, R. G. (1968b) *J Chem Ed* **45**, 643-8.

Pearson, R. G. (1986) *PNAS USA* **83**, 8440-1.

Perona, J. J. and Craik, C. S. (1995) *Protein Science* **4**, 337-60.

Perona, J. J., Craik, C. S. and Fletterick, R. J. (1993a) *Science* **261**, 620-2.

Perona, J. J., Evnin, L. B. and Craik, C. S. (1993b) *Gene* **137**, 121-6.

Perona, J. J., Evnin, L. B. and Craik, C. S. (1994a) *Gene* **137**, 121-6.

Perona, J. J., Hedstrom, L., Wagner, R. L., Rutter, W. J., Craik, C. S. and Fletterick, R. J. (1994b) *Biochemistry* **33**, 3252-9.

Perona, J. J., Rould, M. A. and Steitz, T. A. (1993a) *Biochemistry* **32**, 8758-71.

Perona, J. J., Tsu, C. A., Craik, C. S. and Fletterick, R. J. (1993b) *J Mol Biol* **230**, 919-33.

Perona, J. J., Tsu, C. A., McGrath, M. E., Craik, C. S. and Fletterick, R. J. (1993c) *J Mol Biol* **230**, 934-49.

Perutz, M. F. and Lehmann, H. (1968) *Nature* **219**, 902-9.

Pessi, A., Bianchi, E., Crameri, A., Venturini, S., Tramontano, A. and Sollazzo, M. (1993) *Nature* **362**, 367-9.

Ponder, J. W. and Richards, F. M. (1987) *J Mol Biol* **193**, 775-91.

Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C. and Kraut, J. (1985) *J Biol Chem* **260**, 16122-30.

Quiocho, F. A. and Lipscomb, W. N. (1971) *Adv Protein Chem* **25**, 1-78.

Rawlings, N. D. and Barrett, A. J. (1993) *Biochem J* 205-18.

Rees, D. C., Lewis, M. and Lipscomb, W. N. (1983) *J Mol Biol* **168**, 367-87.

Regan, L. (1995) *Trends Biochem Sci* **20**, 280-5.

Regan, L. and Clarke, N. D. (1990) *Biochemistry* **29**, 10878-83.

Reynolds, W. F., Peat, I. R., Freedman, M. H. and Lyerla, J. J. (1973) *J Am Chem Soc* **95**, 328-31.

Roberts, V. A., Iverson, B. L., Iverson, S. A., Benkovic, S. J., Lerner, R. A., Getzoff, E. D. and Tainer, J. A. (1990) *Proc Natl Acad Sci U S A* **87**, 6654-8.

Ruan, F., Chen, Y. and Hopkins, P. B. (1990) *J. A. C. S.* **112**, 9403-4.

Rypniewski, W. R., Perrakis, A., Vorgias, C. E. and Wilson, K. S. (1994) *Protein Eng* **7**, 57-64.

Sack, J. S. (1988) *J. Molecular Graphics* **6**, 224-5.

Sahin-Tóth, M., Dunten, R. L. and Kaback, H. R. (1995) *Biochemistry* **34**, 1107-12.

Sato, M., Yamamoto, M., Imada, K. and Katsuba, Y. (1992) *Journal of Applied Crystallography* **25**, 348-357.

Schechter, I. and Berger, A. (1968) *Biochem. Biophys. Res. Commun.* **27**, 157.

Schellenberger, V., Siegel, R. A. and Rutter, W. J. (1993a) *Biochemistry* **32**, 4344-8.

Schellenberger, V., Turck, C. W., Hedstrom, L. and Rutter, W. J. (1993b) *Biochemistry* **32**, 4349-53.

Schellenberger, V., Turck, C. W. and Rutter, W. J. (1994) *Biochemistry* **33**, 4251-7.

Segel, I. H. (1975), "Enzyme Kinetics" John Wiley & Sons, Inc., New York

Seymour, J. L., Lindquist, R. N., Dennis, M. S., Moffat, B., Yansura, D., Reilly, D., Wessinger, M. E. and Lazarus, R. A. (1994) *Biochemistry* **33**, 3949-58.

Sharma, S. K., Evans, D. B., Vosters, A. F., McQuade, T. J. and Tarpley, W. G. (1991) *Biotechnol Appl Biochem* **14**, 69-81.

Shin, D. H., Hwang, K. Y., Kim, K. K., Lee, H. R., Lee, C. S., Chung, C. H. and Suh, S. W. (1993) *J Mol Biol* **229**, 1157-8.

Siezen, R. J., de, V. W., Leunissen, J. A. and Dijkstra, B. W. (1991) *Protein Eng* **4**, 719-37.

Sprang, S., Standing, T., Fletterick, R. J., Stroud, R. M., Finer, M. J., Xuong, N. H., Hamlin, R., Rutter, W. J. and Craik, C. S. (1987) *Science* **237**, 905-9.

Sprang, S. R., Fletterick, R. J., Graf, L., Rutter, W. J. and Craik, C. S. (1988) *Crit Rev Biotechnol* **8**, 225-36.

Stroud, R. M., Kossiakoff, A. A. and Chambers, J. L. (1977) *Annu Rev Biophys Bioeng* **6**, 177-93.

Suh, S. S., Haymore, B. L. and Arnold, F. H. (1991) *Protein Eng* **4**, 301-5.

Sundberg, R. J. and Martin, R. B. (1974) *Chemical Reviews* **74**, 471-517.

Szebenyi, D. M. and Moffat, K. (1986) *J Biol Chem* **261**, 8761-77.

Tainer, J. A. and Roberts, V. A. (1990) *Nature* **348**, 589.

Tainer, J. A., Roberts, V. A. and Getzoff, E. D. (1991) *Current Opinion in Biotechnology* **2**, 582-591.

Takeuchi, Y., Noguchi, S., Satow, Y., Kojima, S., Kumagai, I., Miura, K., Nakamura, K. T. and Mitsui, Y. (1991) *Protein Eng* **4**, 501-8.

Thompson, R. C. and Blout, E. R. (1973) *Biochemistry* **12**, 57-65.

Todd, R. J., Van Dam, M. E., Casimiro, D., Haymore, B. L. and Arnold, F. H. (1991) *Proteins* **10**, 156-61.

Trigo-Gonzalez, G., Racher, K., Burtnick, L. and Borgford, T. (1992) *Biochemistry* **31**, 7009-15.

Tsu, C. A., Perona, J. J., Schellenberger, V., Turck, C. W. and Craik, C. S. (1994) *J Biol Chem* **269**, 19565-72.

van den Ouweland, A., van Duijnhoven, H., Keizer, G. D., Dorssers, L. C. and van de Ven, W. (1990) *Nucleic Acids Res* **18**, 664.

van Tilbeurgh, H., Jenkins, J., Chiadmi, M., Janin, J., Wodak, S. J., Mrabet, N. T. and Lambeir, A. M. (1992) *Biochemistry* **31**, 5467-71.

Varsani, L., Cui, T., Rangarajan, M., Hartley, B. S., Goldberg, J., Collyer, C. and Blow, D. M. (1993) *Biochem J* **291**, 575-83.

Vasquez, J. R., Evnin, L. B., Higaki, J. N. and Craik, C. S. (1989) *J Cell Biochem* **39**, 265-76.

Vermote, C. L. and Halford, S. E. (1992) *Biochemistry* **31**, 6082-9.

Vermote, C. L., Vipond, I. B. and Halford, S. E. (1992) *Biochemistry* **31**, 6089-97.

Voet, D. and Voet, J. G. (1990), "Biochemistry" Wiley, New York

Vosters, A. F., Evans, D. B., Tarpley, W. G. and Sharma, S. K. (1992) *Protein Expr Purif* **3**, 18-26.

Vu, T. K., Wheaton, V. I., Hung, D. T., Charo, I. and Coughlin, S. R. (1991) *Nature* **353**, 674-7.

Wang, C. I., Yang, Q. and Craik, C. S. (1995) *J Biol Chem* **270**, 12250-6.

Watanabe, M., Hirano, A., Stenglein, S., Nelson, J., Thomas, G. and Wong, T. C. (1995) *J Virol* **69**, 3206-10.

Watson, J. D., Gilman, M., Witkowski, J. and Zoller, M. (1992), "Recombinant DNA" Scientific American Books, New York

Wells, J. A. (1991) *Methods Enzymol* **202**, 390-411.

Wells, J. A., Cunningham, B. C., Graycar, T. P. and Estell, D. A. (1987a) *Proc Natl Acad Sci U S A* **84**, 5167-71.

Wells, J. A., Cunningham, B. C., Graycar, T. P., Estell, D. A. and Carter, P. (1987b) *Cold Spring Harb Symp Quant Biol* **52**, 647-52.

Wells, J. A. and Estell, D. A. (1988) *Trends Biochem Sci* **13**, 291-7.

Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P. and Estell, D. A. (1987) *Proc Natl Acad Sci U S A* **84**, 1219-23.

Wick, C. B. (1995) *Genetic Engineering News* **15**, 1.

Wilke, M. E., Higaki, J. N., Craik, C. S. and Fletterick, R. J. (1991) *J Mol Biol* **219**, 525-32.

Willett, W. S., Gillmor, S. A., Perona, J. J., Fletterick, R. J. and Craik, C. S. (1995) *Biochemistry* **34**, 2172-80.

Williams, R. J. P. (1959) *Enzymes* **1**, 391-441.

Wollmer, A., Rannefeld, B., Stahl, J. and Melberg, S. G. (1989) *Biol Chem Hoppe Seyler* **370**, 1045-53.

Yamashita, M. M., Wesson, L., Eisenman, G. and Eisenberg, D. (1990) *Proc Natl Acad Sci U S A* **87**, 5648-52.

Yang, W. P., Goldstein, J., Procyk, R., Matsueda, G. R. and Shaw, S. Y. (1994) *Biochemistry* **33**, p606-12.

Appendix 1: Trypsinogen Expression System

This appendix presents the important features of the trypsin expression system, including a plasmid map of the mutagenesis vector, pST. You will find here cloning strategies, media recipes, and tranformation and growth protocols. This appendix is meant as a practical guide to help you use the system. If you wish a more academic view of the material contained herein, consult the references found in Chapter 4 in the materials and methods section.

Summary

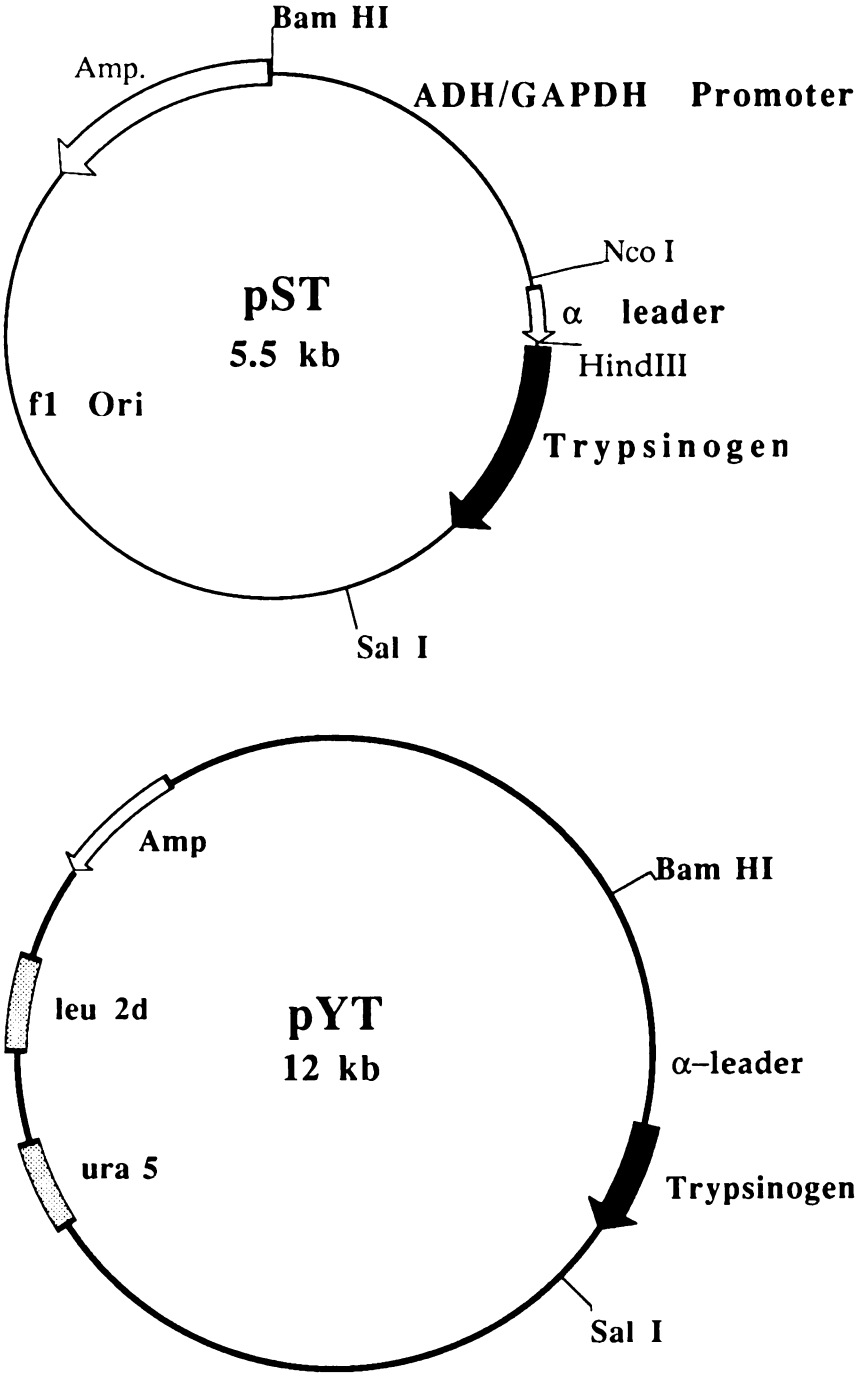
In short, trypsinogen is secreted as a soluble zymogen from *Saccharomyces cerevisiae* into the media as directed by the α-factor signal peptide. The trypsinogen gene is under control of the ADH/GAPDH hybrid promoter which is sensitive to glucose concentration: when glucose is depleted, the promoter is induced and the trypsinogen gene fused to the α-factor leader sequence is expressed. The protein is directed through the secretory pathway and released into the extracellular media. There are two proteolytic processing events, one by the Kex2 protease, and one by the ste13 gene product which separate the signal sequence from the N-terminus of trypsinogen upon secretion of the zymogen.

The secreted trypsinogen is separated from the cells by centrifugation, and purified over a cation exchange column, and optionally, a phenyl sepharose column. The purified trypsinogen is then activated by enteropeptidase to remove the propeptide and liberate mature trypsin, and the resulting trypsin is purified to homogeneity by affinity chromatography. The final purified

trypsin is dialyzed or diafiltered into 1mM HCl, pH 3.0 for storage at 4°C. This material can be stored for at least 6 months with full activity. If the trypsin is meant for crystallization trials, it is concentrated to ~1mM in the storage buffer prior to storage, and this material is used as soon as possible.

Plasmids. The expression of trypsinogen in *Saccharomyces cerevisiae* requires a two plasmid system: one for mutagenesis (pST), and one for expression (pYT). After mutagenesis and sequence confirmation have been performed in pST, the BamHI/SalI fragment containing the ADH/GAPDH promoter sequence, the α-factor leader sequence, and the trypsinogen gene is subcloned into the 12kB 2μm-based yeast expression plasmid, pYT. A schematic representation of the two plasmids is shown in figure A.1 below.

Figure A.1: Two plasmid Yeast Expression System



pST. The 5.5kB pST plasmid is bluescript-based (KS+) allowing efficient generation of uracil-laden single-stranded DNA for performing oligo-directed mutagenesis. The M13 ori is directed so that the non-coding strand of trypsinogen is rescued which means that mutagenic oligos should be coding (*i.e.*, Met=ATG). I use the program "snipper" found on the computer Socrates to design oligos, and I always try to introduce a new restriction site with the mutagenesis to simplify identification of the desired mutant by restriction sreening. This plasmid is nice to work with because it gives high yield in mini-prep procedures and behaves well during restriction screening and subcloning procedures. The sequence and restriction map of the pST BamHI/SalI fragment containing the trypsinogen gene is presented at the end of this appendix. For most manipulations with plasmid DNA, I prefer to use CsCl purified DNA (see appendix 5 for a protocol) in order to avoid the possibility that the cloning experiment did not work due to impure DNA. I realize this may seem a bit dogmatic, but because these molecular biology experiments are so "black box" in nature, I like the idea of avoiding uncertainty whenever I can. Besides, a CsCl prep of DNA yields enough to sequence *ad nauseum*, and to make a master stock for the -80°C freezer, and it's a good idea to give your collegues and collaborators DNA that they can trust.

For the subcloning steps, I usually digest 5-10μg of DNA with 10-20 units of BamHI and SalI (each) for one hour at 37°C in NEB buffer 3. The digest results in a 3kB fragment containing the desired Bam/Sal insert, and a 3.5kB fragment containing unwanted bluescript sequence. This, unfortunately, is a little close in size for easy cutting of the desired fragment, so I usually let the gel run long, until the dye is off the end of the gel for 10-15 minutes. I've

found that using fresh ethidium bromide stain, and only staining the gel for 30 seconds to 1 minute *maximum*, results in the most successful ligation experiments. The minimized staining time only allows penetration of the dye to the outer region of the gel, so that exposure to UV in order to visualize the band will only badly damage a portion of the DNA. As a general rule, I keep exposure of any DNA to UV light to a minimum, but especially DNA meant for ligations. A useful project for some interested party would be to remove the non-coding sequence from between the stop codon of trypsinogen and the SalI site. This would produce a smaller Bam/Sal insert fragment upon digestion, simplify the gel purification step, and cut down on background during ligation. I generally use geneclean (glassmilk) with the standard protocol to purify the 3kB fragment, and I elute the purified fragment in two steps of 15µl water.

pYT. The pYT plasmid is ~12kB in size and contains the necessary yeast sequences for replication and selection on uracil-minus and leucine-minus agar plates. This large plasmid gives a low yield in normal miniprep procedures (compared to bluescript), and can be fragile due to its large size. I have found through experience that using CsCl purified pYT and using Elutip (Schleicher and Schuell) to purify the Bam/Sal vector fragment gives the best results for ligating in the Bam/Sal insert fragment from pST. For the Elutip procedure, I cut the gel slice out, mince it with a clean, EtOH-sterilized razor blade, and elute the ~9kB fragment overnight in 5 mls of the low salt buffer in a 15ml falcon tube covered with foil (*a la* J. Vásquez).

Ligation and transformation into E. coli. Although it is good to try ratios of vector and insert in ligations, I have found that using 1 or 2µl of fragments

prepared as outlined above gives positive results. I do ligations in a total volume of 20µl at room temperature for 2-3 hrs., transform 5-10µl of the ligation reaction into $CaCl_2$-competent X90's, and put the rest of the ligation mix at 16°C for an overnight backup ligation reaction. For transformation into *E.* coli, I have found (conveniently) that if the ligation was successful, a 10 minute incubation of 10µl of ligation mix with 100µl of cells on ice, followed by a 1 minute "heat shock" at 37° or 42°C (dealer's choice), followed by addition of 1ml of LB and 10-20 minutes of outgrowth at 37°C in a 1.5ml eppendorf tube in the shaker will give plenty of colonies for subsequent mini preps and restriction screening. This butchered transformation procedure works wonderfully for intact plasmids, leaving you more time to think, or allowing you to go home earlier.

Electroporation into yeast. For this project, the *Saccharomyces cerevisiae* strain DLM101α has been used exclusively, although other strains may improve expression levels dramatically (Sergio Pichuantes, personal communication). I have also limited myself to transforming the yeast by electroporation because of the ease of the protocol. One you have obtained the correct pYT plasmid (verified by sequencing or restriction digest), use the following protocol to transform the plasmid into yeast.

Electroporation into Yeast.

1. Make a fresh streak of DLM101α from a glycerol stock onto YP+2% glucose plate and allow the colonies to grow at 30°C for 2 or 3 days or until they are 1-2mm in size. **All yeast growth protocols specify 30°C.**

2. Start a 50ml overnight culture of yeast in YP+2% glucose media by picking one colony from the YP plate, and inoculating the 50mls of media. The next morning, the cells should be between 0.6 and 1.0 OD_{600}. Filter 100mls of deionized (DI) water and place on ice or at 4°C.

3. When the proper OD's have been reached, put the culture on ice and harvest cells by centrifugation in a 50ml falcon tube in the SA600 rotor at 4k rpm for 5 minutes.

4. Wash the cells 3 times by resuspension in 15mls of the cold DI water, followed by centrifugtion as above. Keep the cells on ice. Finally resuspend the cells in ~1-2 mls of the cold water and keep them on ice. The resulting suspension should be a thick white slurry.

5. In sterile eppendorf tubes, place 1-2μg of pYT DNA and add 40μl of washed yeast cells. Mix by gentle pipetting, and place the mix on ice for 10 minutes.

6. When the DNA has incubated with the cells10 minutes, add the mix to hte bottom of a 2mm electroporation cuvette (Bio Rad). Tap the cell on the bench to ensure that the mix entirely bridges the electrodes at the bottom of the cuvette (no bubbles), and electroporate using the following settings: 200Ω, 600V, 25μF. You can monitor the time constant (τ) of the discharge; it should be ~4.8. Higher conductivity (more salt) leads to lower time constants.

7. Immediately add 1ml of room temp. SOS solution to the cuvette, mix, and pipette the entire contents into a 3ml culture tube. Outgrow this culture for 90 minutes @30°C in a shaking incubator.

8. Plate 200μl of the transformed mix onto a SD -uracil plate and incubate at 30°C for 2-3 days, or until colonies appear against the background of cell debris.

9. When colonies have appeared on the SD -ura plate, restreak two or three onto an SD -leucine plate, and incubate for another 2 days @30°C. Colonies resulting from this second selection are ready for growth in liquid media and subsequent expression trials. Note: you may skip the second selection on SD -leu plates if you wish to save time, however, I've found generally better expression levels when this step is included.

Expression of trypsinogen. At this point, there is another selection in SD -leucine liquid media, and then growth in rich YP+2% glucose media for expression of the gene.

To test for expression:

1. Inoculate 1ml of SD -leucine media +8% glucose with one colony picked from an SD -leucine plate, and grow overnight @30°C with shaking.

2. Inoculate 2ml of YP media +1% glucose with 100µl of the SD -lue culture, and let grow overnight to 72 hours @30°C with shaking.

3. Expression can be checked by spinning out the cells from 500µl of culture, and precipitating the trypsinogen by adding 1ml of cold 100% ethanol to 500µl of culture supernatant, incubating at -20°C for 30 min., centrifuging at maximum speed for 10 minutes, decanting the supernatant, resuspending the pellet in sample buffer, and running on an SDS gel. Running a time course is recommended. You should start to see expression after 40-48 hrs.

For Large-Scale expression:

1. Inoculate a 1ml culture as before for expression check, but use this culture to inoculate 50mls of SD -leucine media +8% glucose, and grow for 2 days @30°C with shaking.

2. At this point it is useful to make a glycerol stock of the culture by mixing 800µl of the culture with 200µl of 80% glycerol, followed by freezing at -80°C. This way you can avoid having to retransform and select colonies for future expression of this mutant.

3. Inoculate 1 liter of YP media +1 or 2% glucose and allow to grow for 48-72hrs @30°C with shaking.

4. Remove cells from trypsinogen in the media by centrifugation.

Recipes

YPG media/plates:

	200mls	1 liter
1% Bacto-yeast extract	2 g	10 g
2% Bacto-peptone	4 g	20 g
2% glucose (from 50% stk)	8 ml	40 ml
[2% agar, for plates only]	4 g	20 g

Autoclave 20 min @121°C. Be sure to add glucose only after autoclaving!

SOS media:
10. ml of 2M sorbital, sterile filtered
6.7 ml YPG media, sterilized
130 µl 1M $CaCl_2$, sterile filtered
3.17 ml sterile DI water
20 ml

SD +8% glucose minimal media/plates

	200 ml
0.67% bacto-yeast nitrogen base without amino acids	1.3 g
DI water	168 ml
amino acid mix (either Ura^- or Leu^-)	260 mg
[2% agar, for plates only]	4 g

Mix ingredients, ensure that pH is 5.5-6.0 - you may have to add some NaOH. Autoclave 20 min@121°C. Cool slightly, then add

8% glucose (from 50% sterile-filtered stock)	32 ml

Pour plates of 20 mls each.

Amino acid mixes

The amino acid mixes are powdered mixes for complementing the minimal SD media/plates. The ones I've made will lower the pH of a solution when added, so beware to check your pH, because agar will not set properly at pH<5. Add mix to the media prior to autoclaving at 1.3 g/liter.

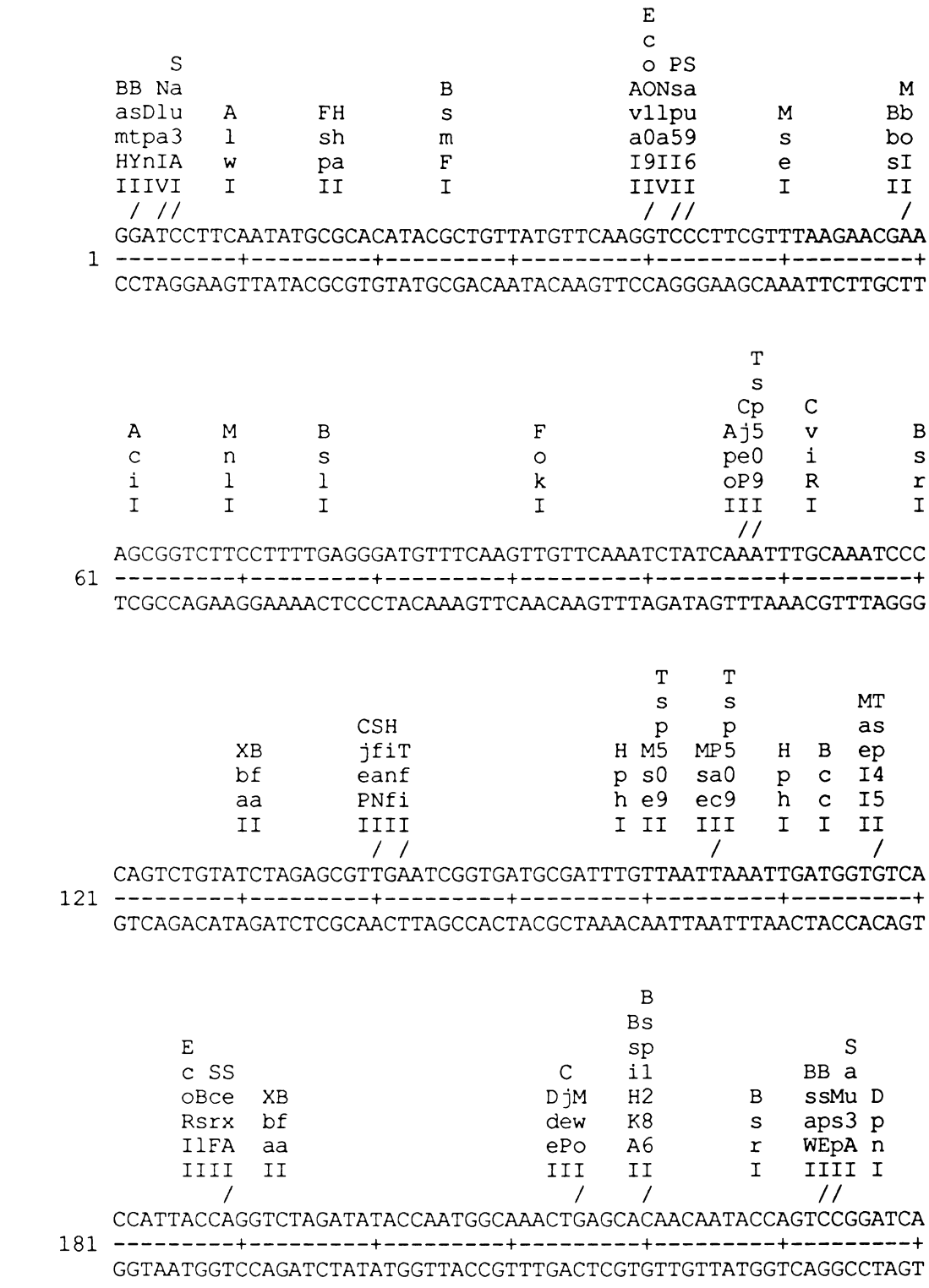
Ingredient	Amount	[Final] in media
Adenine sulfate	1.0 g	20 mg/l
L-tryptophan	1.0	20
L-histidine	1.0	20
L- arginine	1.0	20
L-methionine	1.0	20
L-tyrosine	1.5	30
L-isoleucine	1.5	30
L-lysine	1.5	30
L- phenylalanine	2.5	50
L-glutamate	5.0	100
L-aspartate	5.0	100
L-valine	7.5	150
L-threonine	10.0	200
L-serine	18.75	375
[uracil]	1.0	20
[L-leucine]	1.5	30
	total 59.25 g	

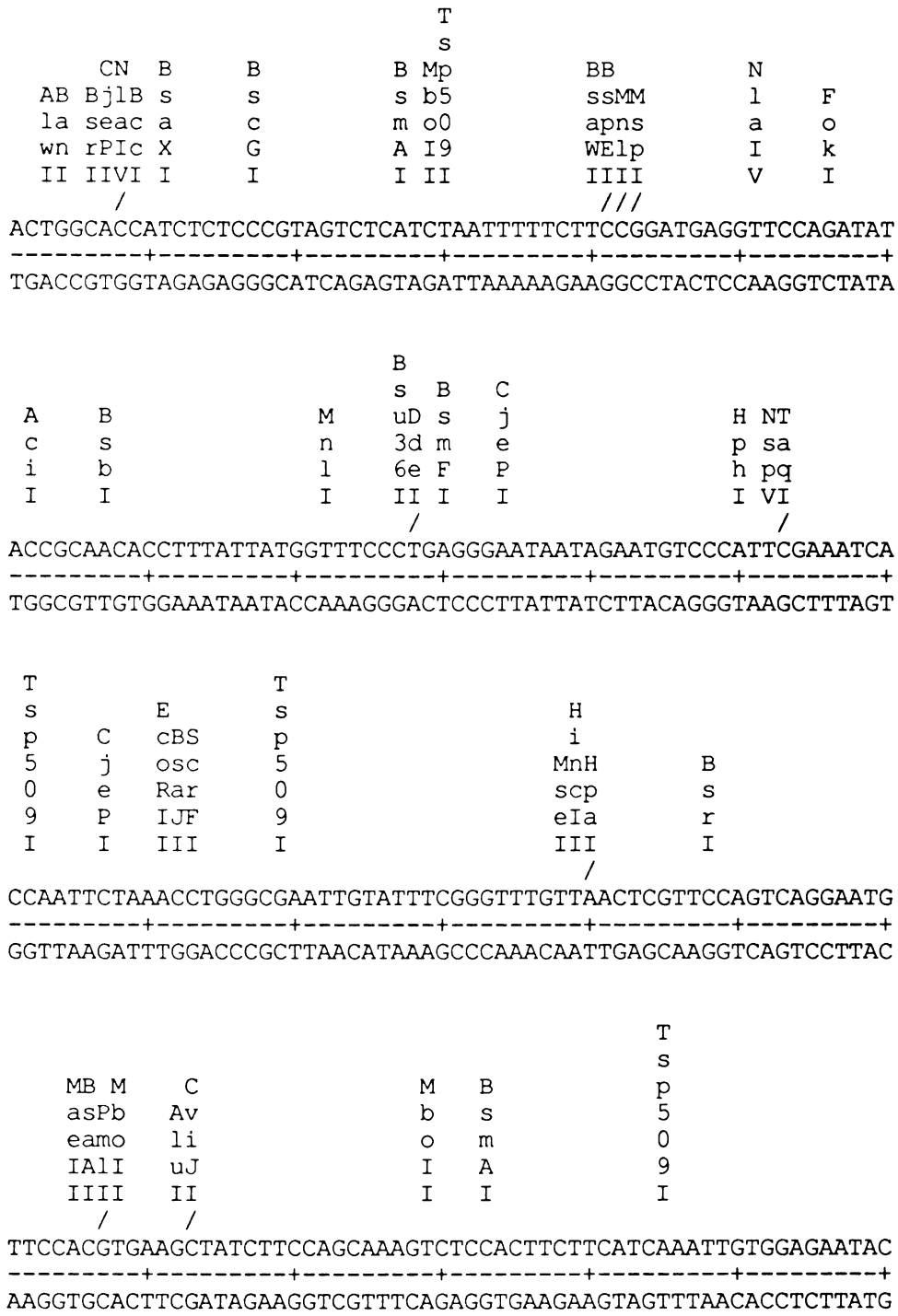
You can either leave the uracil and leucine out of a master stock, and add 20 or 30 mgs/liter, respectively, to media/plates, or you can make two amino acid mixes with either ingredient deleted and use appropriately. Store this mix in a cool dry place.

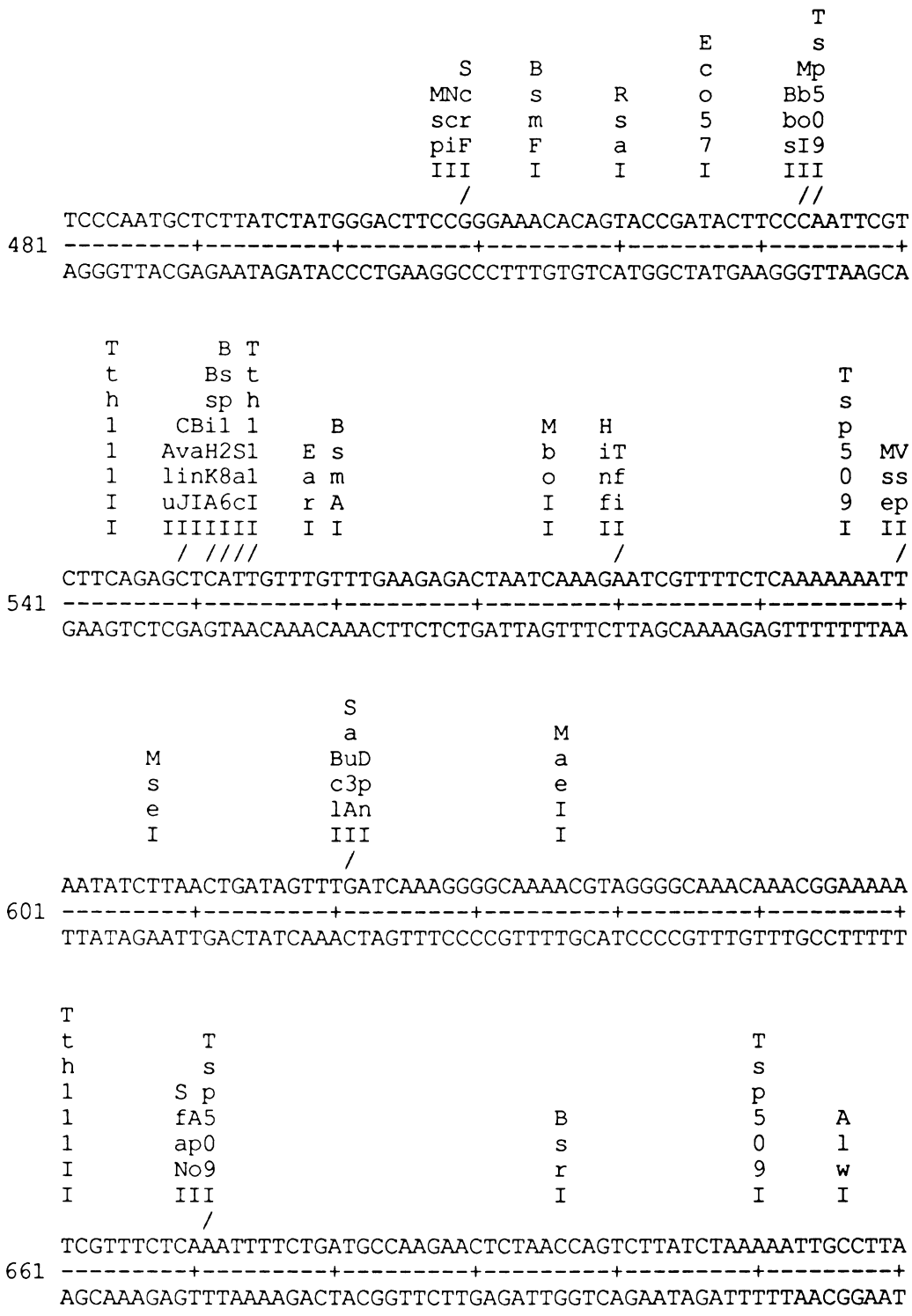
Restriction map of BamHI/SalI insert:

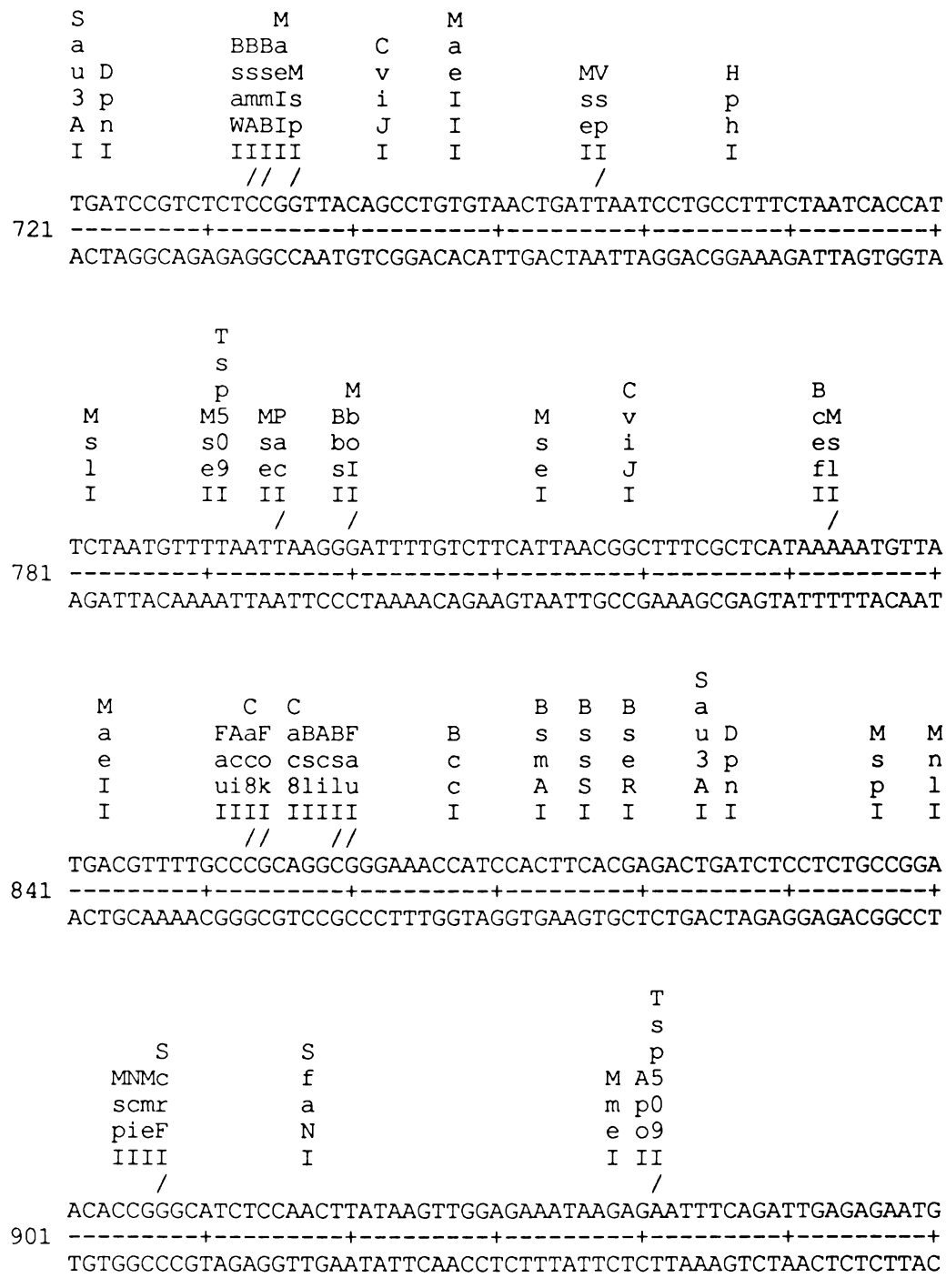
For the sake of brevity, only the restriction map of the BamHI/SalI insert fragment of pST is shown here in figure A.2. The rest of the sequence of pST is Bluescript KS+. The beginning of the α-factor leader sequence, the propeptide of trypsinogen, the start and end of trypsin are shown. The open reading frame is translated with the active-site residues of trypsin marked and indicated in bold.

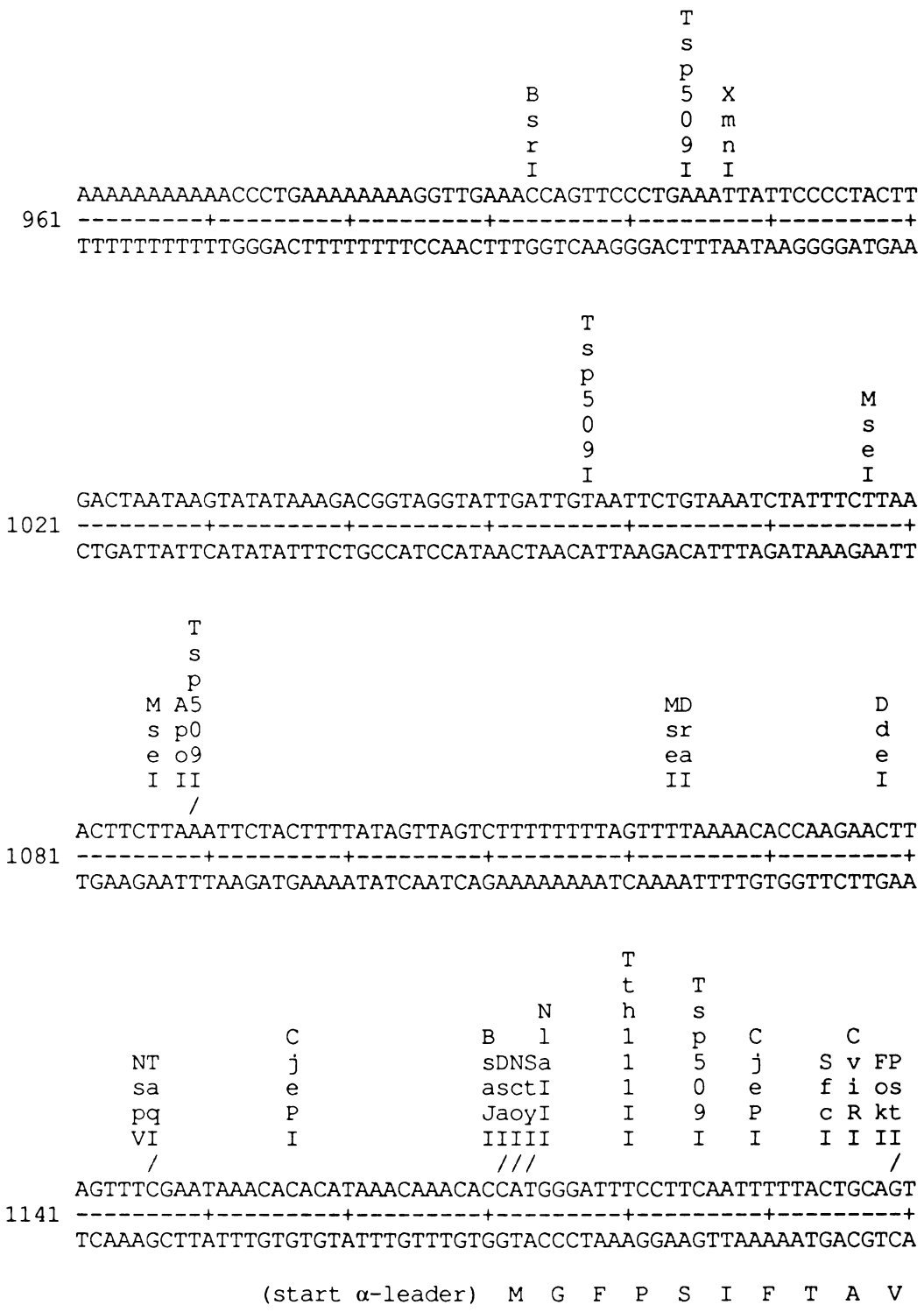
Restriction MAP of: pST BamHI/SalI insert from: 1 to: 2505bp

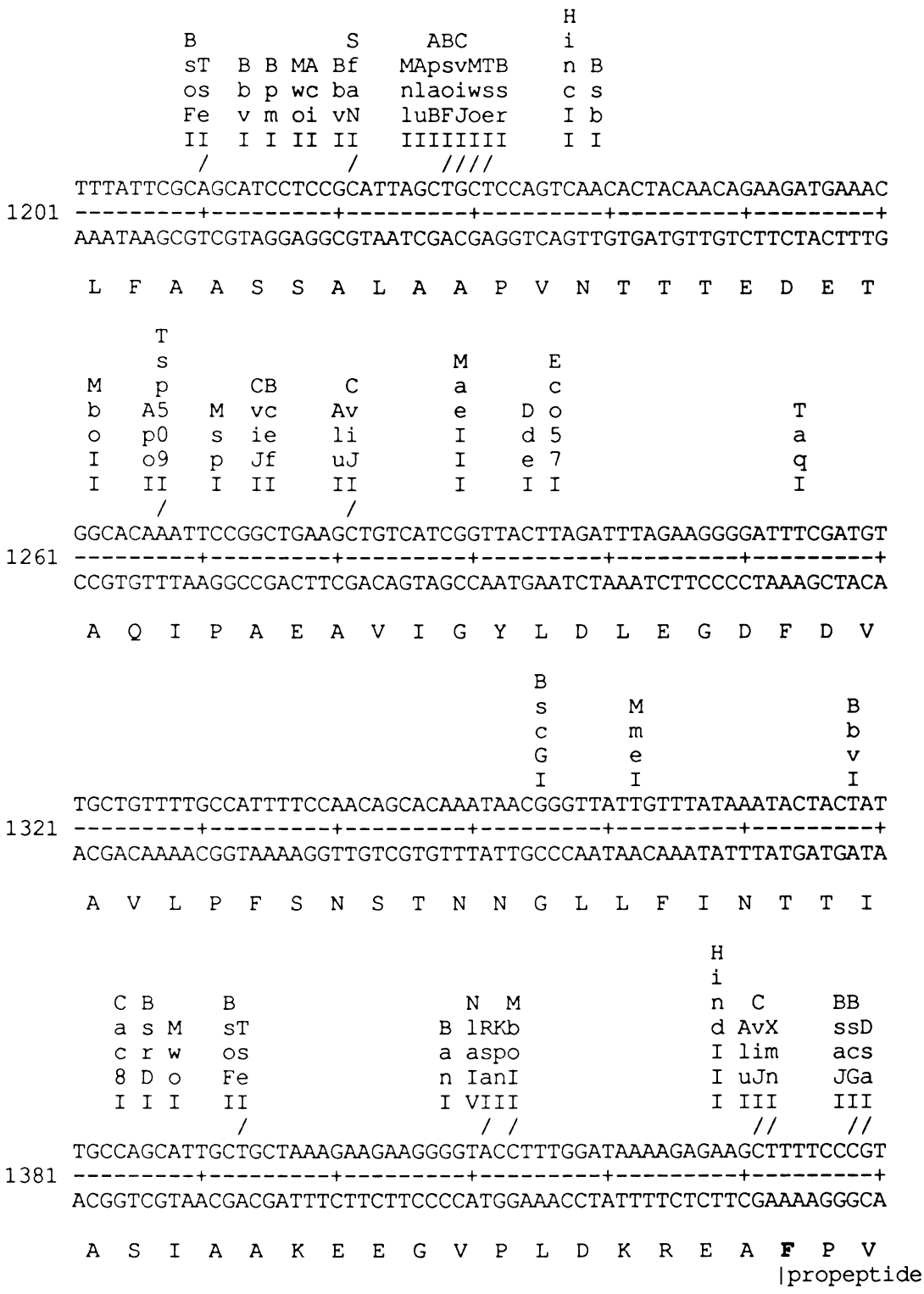


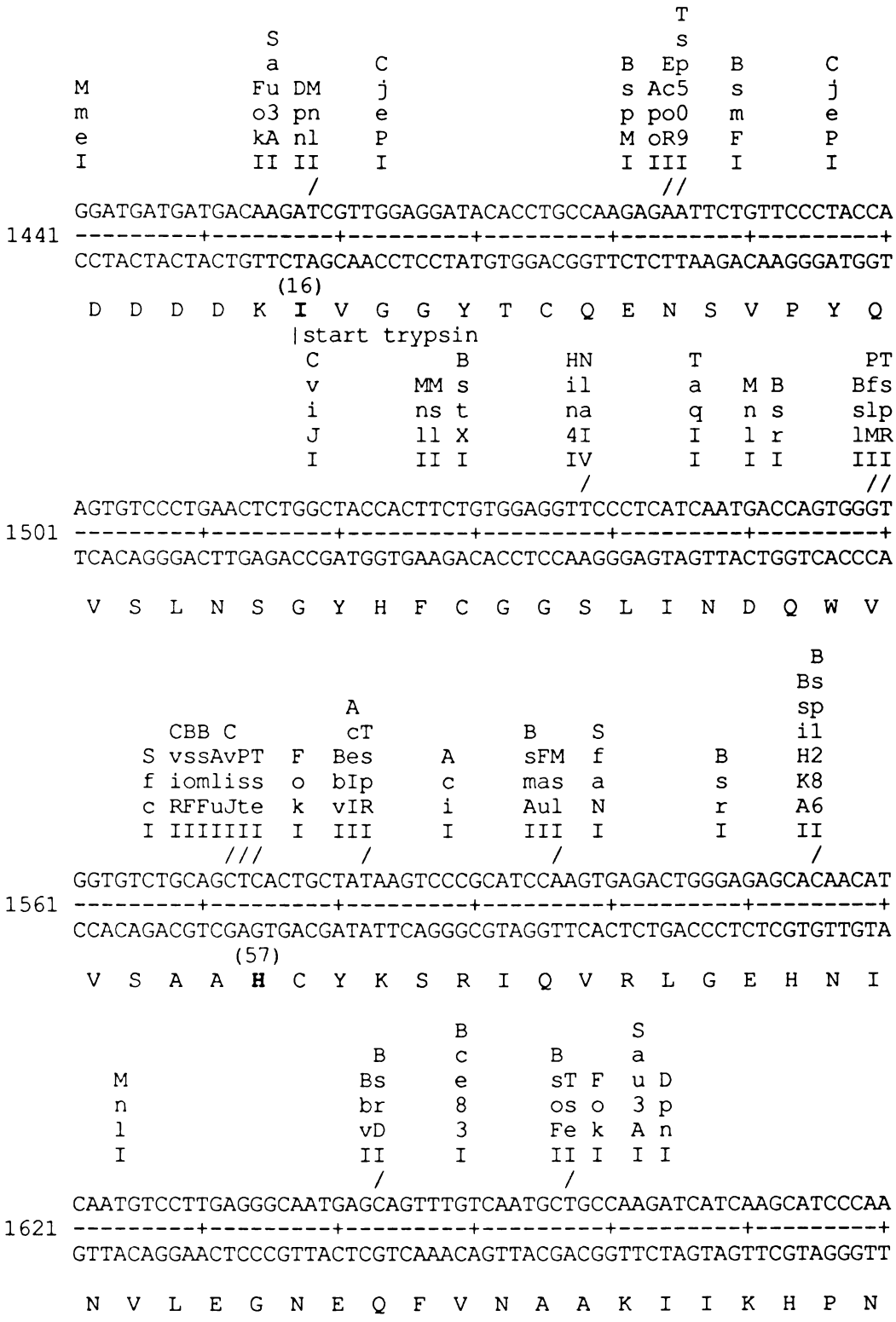

T
s
CN B B B Mp BB N
AB BjlB s s s b5 ssMM l F
la seac a c m o0 apns a o
wn rPIc X G A I9 WElp I k
II IIVI I I I II IIII V I
ACTGGCACCATCTCTCCCGTAGTCTCATCTAATTTTTCTTCCGGATGAGGTTCCAGATAT
241
TGACCGTGGTAGAGAGGGCATCAGAGTAGATTAAAAAGAAGGCCTACTCCAAGGTCTATA
B
s B C
A B M uD s j H NT
c s n 3d m e p sa
i b l 6e F P h pq
I I I II I I I VI
ACCGCAACACCTTTATTATGGTTTCCCTGAGGGAATAATAGAATGTCCCATTCGAAATCA
301
TGGCGTTGTGGAAATAATACCAAAGGGACTCCCTTATTATCTTACAGGGTAAGCTTTAGT
T T
s E s H
p C cBS p i
5 j osc 5 MnH B
0 e Rar 0 scp s
9 P IJF 9 eIa r
I I III I III I
CCAATTCTAAACCTGGGCGAATTGTATTTCGGGTTTGTTAACTCGTTCCAGTCAGGAATG
361
GGTTAAGATTTGGACCCGCTTAACATAAAGCCCAAACAATTGAGCAAGGTCAGTCCTTAC
T
s
MB M C M B p
asPb Av b s 5
eamo li o m 0
IAlI uJ I A 9
IIII II I I I
TTCCACGTGAAGCTATCTTCCAGCAAAGTCTCCACTTCTTCATCAAATTGTGGAGAATAC
421
AAGGTGCACTTCGATAGAAGGTCGTTTCAGAGGTGAAGAAGTAGTTTAACACCTCTTATG

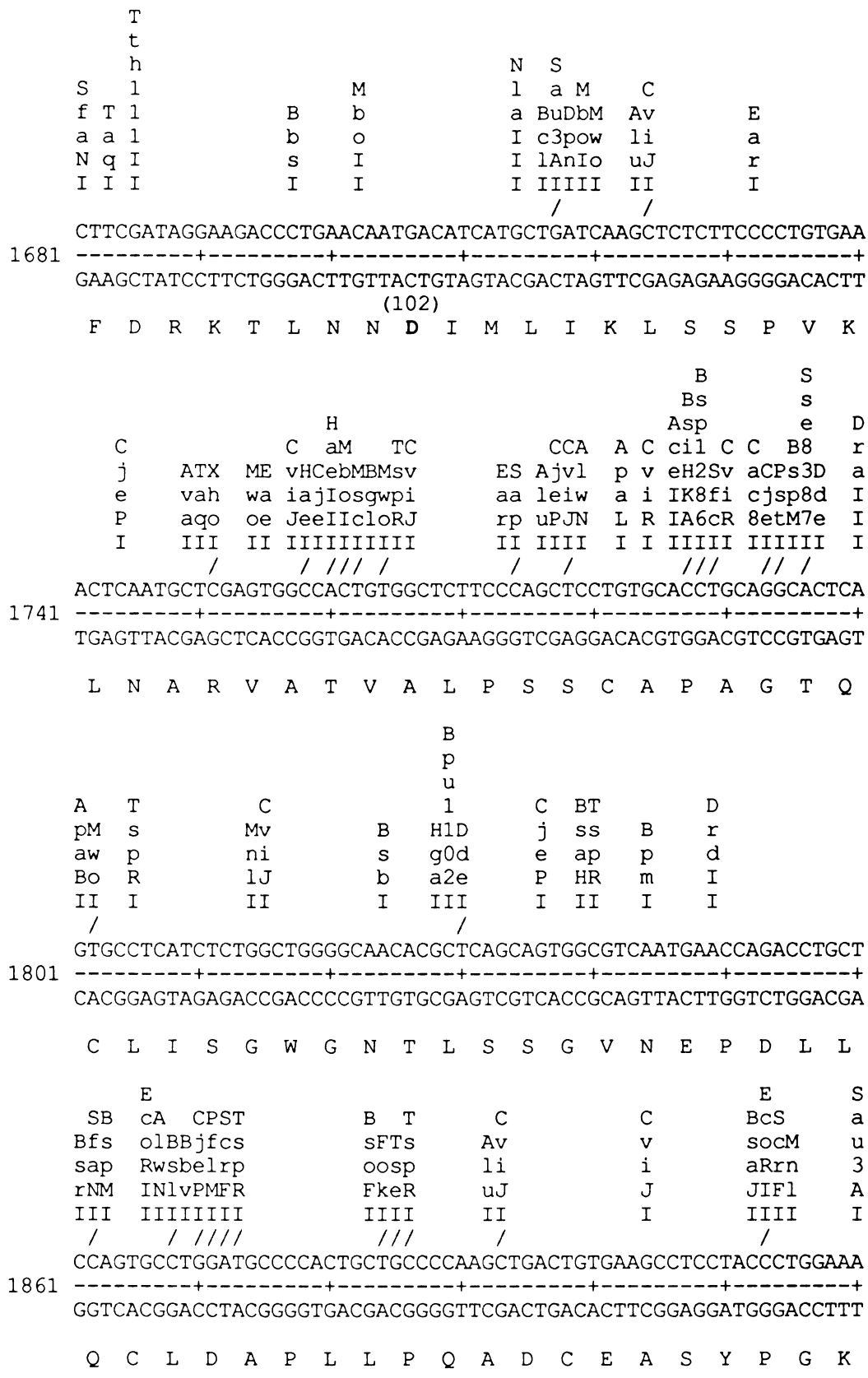

TCCCAATGCTCTTATCTATGGGACTTCCGGGAAACACAGTACCGATACTTCCCAATTCGT
481
AGGGTTACGAGAATAGATACCCTGAAGGCCCTTTGTGTCATGGCTATGAAGGGTTAAGCA
CTTCAGAGCTCATTGTTTGTTTGAAGAGACTAATCAAAGAATCGTTTTCTCAAAAAAATT
541
GAAGTCTCGAGTAACAAACAAACTTCTCTGATTAGTTTCTTAGCAAAAGAGTTTTTTTAA
AATATCTTAACTGATAGTTTGATCAAAGGGGCAAAACGTAGGGGCAAACAAACGGAAAAA
601
TTATAGAATTGACTATCAAACTAGTTTCCCCGTTTTGCATCCCCGTTTGTTTGCCTTTTT
TCGTTTCTCAAATTTTCTGATGCCAAGAACTCTAACCAGTCTTATCTAAAAATTGCCTTA
661
AGCAAAGAGTTTAAAAGACTACGGTTCTTGAGATTGGTCAGAATAGATTTTTAACGGAAT

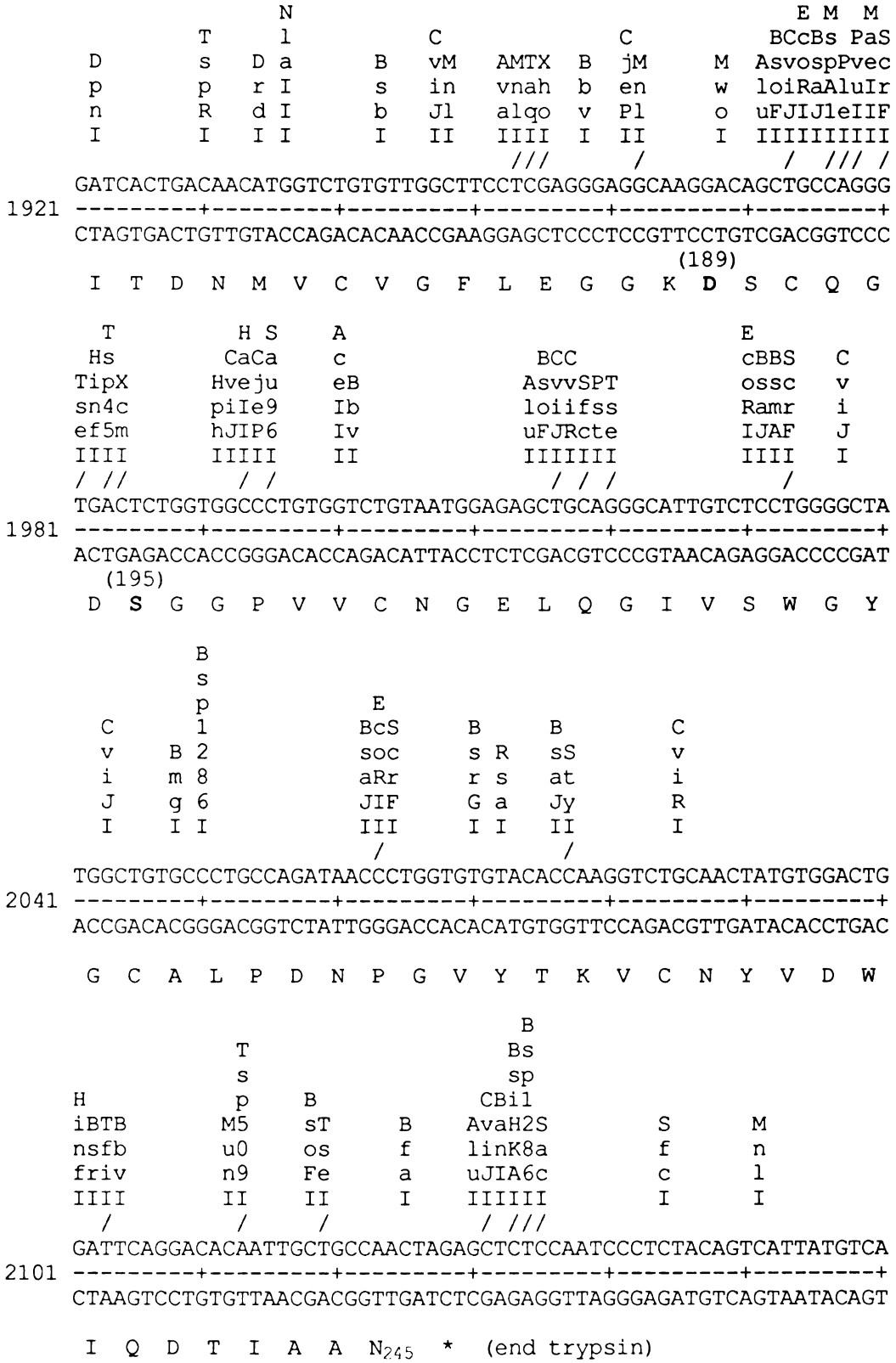

TGATCCGTCTCTCCGGTTACAGCCTGTGTAACTGATTAATCCTGCCTTTCTAATCACCAT
721
ACTAGGCAGAGAGGCCAATGTCGGACACATTGACTAATTAGGACGGAAAGATTAGTGGTA
TCTAATGTTTTAATTAAGGGATTTTGTCTTCATTAACGGCTTTCGCTCATAAAAATGTTA
781
AGATTACAAAATTAATTCCCTAAAACAGAAGTAATTGCCGAAAGCGAGTATTTTTACAAT
TGACGTTTTGCCCGCAGGCGGGAAACCATCCACTTCACGAGACTGATCTCCTCTGCCGGA
841
ACTGCAAAACGGGCGTCCGCCCTTTGGTAGGTGAAGTGCTCTGACTAGAGGAGACGGCCT
ACACCGGGCATCTCCAACTTATAAGTTGGAGAAATAAGAGAATTTCAGATTGAGAGAATG
901
TGTGGCCCGTAGAGGTTGAATATTCAACCTCTTTATTCTCTTAAAGTCTAACTCTCTTAC

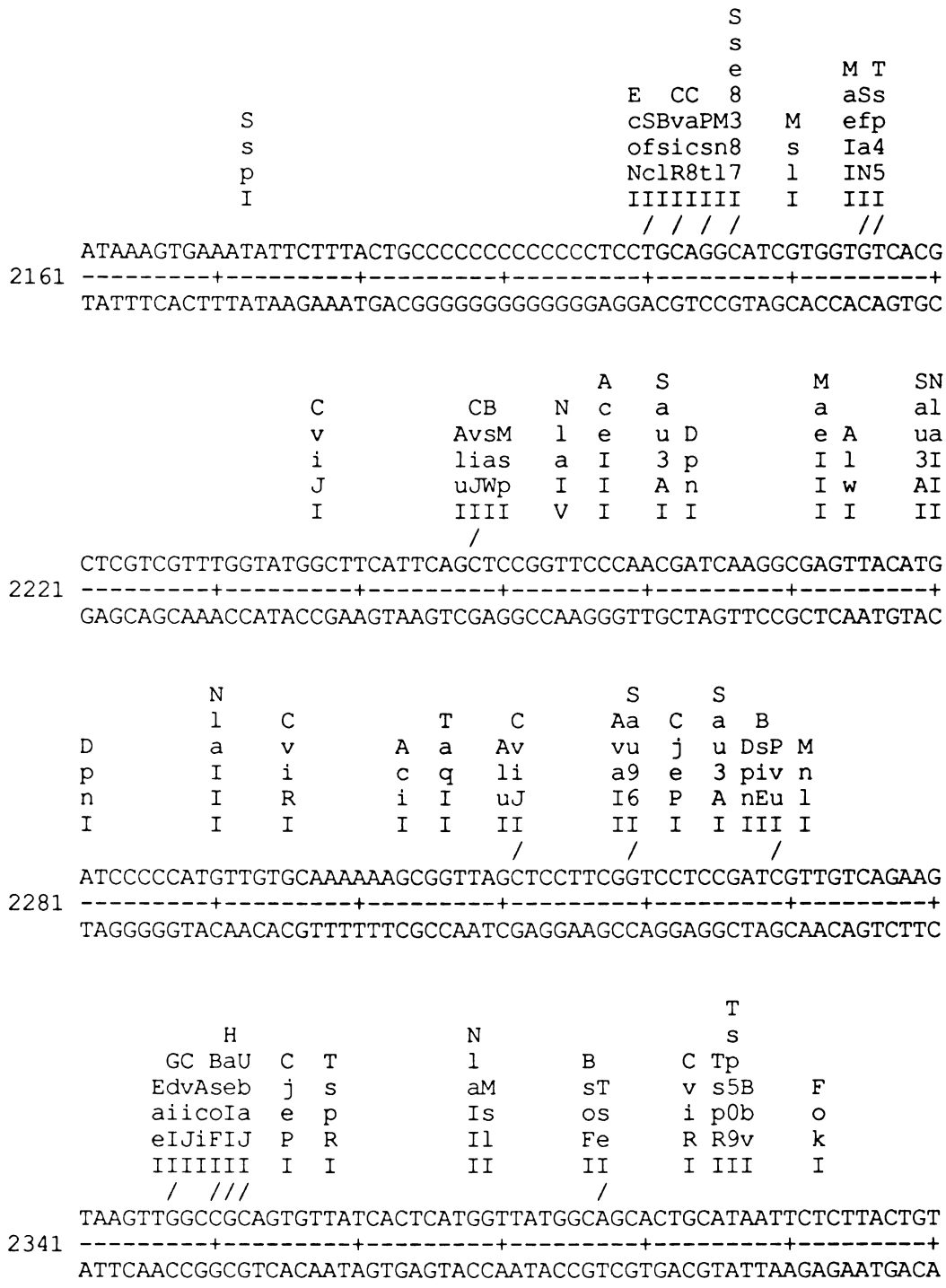

T
s
p
B
s
r
I
5
0
9
I
X
m
n
I
AAAAAAAAAAACCCTGAAAAAAAAGGTTGAAACCAGTTCCCTGAAATTATTCCCCTACTT
961
TTTTTTTTTTTGGGACTTTTTTTTCCAACTTTGGTCAAGGGACTTTAATAAGGGGATGAA
T
s
p
5
0
9
I
M
s
e
I
GACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAAATCTATTTCTTAA
1021
CTGATTATTCATATATTTCTGCCATCCATAACTAACATTAAGACATTTAGATAAAGAATT
T
s
p
M A5
s p0
e o9
I II
/
MD
sr
ea
II
D
d
e
I
ACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACACCAAGAACTT
1081
TGAAGAATTTAAGATGAAAATATCAATCAGAAAAAAAATCAAAATTTTGTGGTTCTTGAA
T
t
h
1
1
I
T
s
p
5
0
9
I
N
l
B
sDNSa
asctI
JaoyI
IIIII
///
C
j
e
P
I
C
j
e
P
I
NT
sa
pq
VI
/
S v FP
f i os
c R kt
I I II
/
AGTTTCGAATAAACACACATAAACAAACACCATGGGATTTCCTTCAATTTTTACTGCAGT
1141
TCAAAGCTTATTTGTGTGTATTTGTTTGTGGTACCCTAAAGGAAGTTAAAAATGACGTCA
(start α-leader) M G F P S I F T A V


TTTATTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAAC
1201
AAATAAGCGTCGTAGGAGGCGTAATCGACGAGGTCAGTTGTGATGTTGTCTTCTACTTTG
L F A A S S A L A A P V N T T T E D E T
GGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTTAGATTTAGAAGGGGATTTCGATGT
1261
CCGTGTTTAAGGCCGACTTCGACAGTAGCCAATGAATCTAAATCTTCCCCTAAAGCTACA
A Q I P A E A V I G Y L D L E G D F D V
TGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTATAAATACTACTAT
1321
ACGACAAAACGGTAAAAGGTTGTCGTGTTTATTGCCCAATAACAAATATTTATGATGATA
A V L P F S N S T N N G L L F I N T T I
TGCCAGCATTGCTGCTAAAGAAGAAGGGGTACCTTTGGATAAAAGAGAAGCTTTTCCCGT
1381
ACGGTCGTAACGACGATTTCTTCTTCCCCATGGAAACCTATTTTCTCTTCGAAAAGGGCA
A S I A A K E E G V P L D K R E A F P V
|propeptide


GGATGATGATGACAAGATCGTTGGAGGATACACCTGCCAAGAGAATTCTGTTCCCTACCA
1441
CCTACTACTACTGTTCTAGCAACCTCCTATGTGGACGGTTCTCTTAAGACAAGGGATGGT
(16)
D D D D K I V G G Y T C Q E N S V P Y Q
|start trypsin
AGTGTCCCTGAACTCTGGCTACCACTTCTGTGGAGGTTCCCTCATCAATGACCAGTGGGT
1501
TCACAGGGACTTGAGACCGATGGTGAAGACACCTCCAAGGGAGTAGTTACTGGTCACCCA
V S L N S G Y H F C G G S L I N D Q W V
GGTGTCTGCAGCTCACTGCTATAAGTCCCGCATCCAAGTGAGACTGGGAGAGCACAACAT
1561
CCACAGACGTCGAGTGACGATATTCAGGGCGTAGGTTCACTCTGACCCTCTCGTGTTGTA
(57)
V S A A H C Y K S R I Q V R L G E H N I
CAATGTCCTTGAGGGCAATGAGCAGTTTGTCAATGCTGCCAAGATCATCAAGCATCCCAA
1621
GTTACAGGAACTCCCGTTACTCGTCAAACAGTTACGACGGTTCTAGTAGTTCGTAGGGTT
N V L E G N E Q F V N A A K I I K H P N


CTTCGATAGGAAGACCCTGAACAATGACATCATGCTGATCAAGCTCTCTTCCCCTGTGAA
1681
GAAGCTATCCTTCTGGGACTTGTTACTGTAGTACGACTAGTTCGAGAGAAGGGGACACTT
(102)
F D R K T L N N D I M L I K L S S P V K
ACTCAATGCTCGAGTGGCCACTGTGGCTCTTCCCAGCTCCTGTGCACCTGCAGGCACTCA
1741
TGAGTTACGAGCTCACCGGTGACACCGAGAAGGGTCGAGGACACGTGGACGTCCGTGAGT
L N A R V A T V A L P S S C A P A G T Q
GTGCCTCATCTCTGGCTGGGGCAACACGCTCAGCAGTGGCGTCAATGAACCAGACCTGCT
1801
CACGGAGTAGAGACCGACCCCGTTGTGCGAGTCGTCACCGCAGTTACTTGGTCTGGACGA
C L I S G W G N T L S S G V N E P D L L
CCAGTGCCTGGATGCCCCACTGCTGCCCCAAGCTGACTGTGAAGCCTCCTACCCTGGAAA
1861
GGTCACGGACCTACGGGGTGACGACGGGGTTCGACTGACACTTCGGAGGATGGGACCTTT
Q C L D A P L L P Q A D C E A S Y P G K


GATCACTGACAACATGGTCTGTGTTGGCTTCCTCGAGGGAGGCAAGGACAGCTGCCAGGG
1921
CTAGTGACTGTTGTACCAGACACAACCGAAGGAGCTCCCTCCGTTCCTGTCGACGGTCCC
I T D N M V C V G F L E G G K D S C Q G
(189)
TGACTCTGGTGGCCCTGTGGTCTGTAATGGAGAGCTGCAGGGCATTGTCTCCTGGGGCTA
1981
ACTGAGACCACCGGGACACCAGACATTACCTCTCGACGTCCCGTAACAGAGGACCCCGAT
D S G G P V V C N G E L Q G I V S W G Y
(195)
TGGCTGTGCCCTGCCAGATAACCCTGGTGTGTACACCAAGGTCTGCAACTATGTGGACTG
2041
ACCGACACGGGACGGTCTATTGGGACCACACATGTGGTTCCAGACGTTGATACACCTGAC
G C A L P D N P G V Y T K V C N Y V D W
GATTCAGGACACAATTGCTGCCAACTAGAGCTCTCCAATCCCTCTACAGTCATTATGTCA
2101
CTAAGTCCTGTGTTAACGACGGTTGATCTCGAGAGGTTAGGGAGATGTCAGTAATACAGT
I Q D T I A A N_{245} * (end trypsin)



```
        N                    MT
        lS                   as
        afB                  ep        B   RS    H          D
        Iac                  I4        s   sc    p          d
        INc                  I5        r   aa    h          e
        III                  II        I   II    I          I
                              /             /
     CATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGA
2401 ---------+---------+---------+---------+---------+---------+
     GTACGGTAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCT
```

```
                                             H
              BU    B           T  SB      i
       B     Asb    s      H    aMNcs BAnST
       c     coa    i      g    qscra cccaa
       g     iFJ    E      a    IpiFH gcIlq
       I     III    I      I    IIIII IIIII
              /                  //    //
     ATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCGACGCT
2461 ---------+---------+---------+---------+----- 2505
     TATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGCTGCGA
```

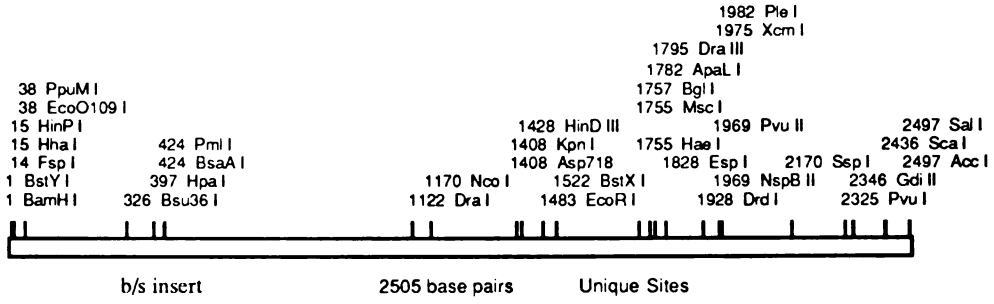
Enzymes that do cut:

AccI	AceIII	AciI	AluI	AlwI	AlwNI	ApaBI	ApaLI
ApoI	AvaI	AvaII	BamHI	BanI	BanII	BbsI	BbvI
BccI	Bce83I	BcefI	BcgI	BclI	BfaI	BglI	BmgI
BpmI	Bpu1102I	BsaAI	BsaHI	BsaJI	BsaWI	BsaXI	BsbI
BscGI	BseRI	BsiEI	BsiHKAI	BslI	BsmAI	BsmBI	BsmFI
BsoFI	Bsp1286I	BspEI	BspMI	BsrI	BsrDI	BsrGI	BssSI
BstXI	BstYI	Bsu36I	Cac8I	CjeI	CjePI	CviJI	CviRI
DdeI	DpnI	DraI	DraIII	DrdI	DrdII	DsaI	EaeI
EarI	Eco57I	EcoNI	EcoO109I	EcoRI	EcoRII	FauI	FokI
FspI	GdiII	HaeI	HaeIII	HgaI	HhaI	Hin4I	HincII
HindIII	HinfI	HpaI	HphI	KpnI	MaeII	MaeIII	MboII
MmeI	MnlI	MscI	MseI	MslI	MspI	MspA1I	MunI
MwoI	NciI	NcoI	NlaIII	NlaIV	NspV	PacI	PflMI
PleI	PmlI	Psp5II	PstI	PvuI	PvuII	RsaI	SacI
SalI	SapI	Sau96I	Sau3AI	ScaI	ScrFI	SexAI	SfaNI
SfcI	Sse8387I	SspI	StyI	TaqI	TaqII	TfiI	TseI
Tsp45I	Tsp509I	TspRI	Tth111II	UbaJI	VspI	XbaI	XcmI
XhoI	XmnI						

Enzymes that do not cut:

AatII	AflII	AflIII	AhdI	ApaI	AscI	AvrII	BaeI
BglII	Bpu10I	BsaI	BsaBI	BsgI	BsmI	Bsp24I	Bsp24I
BspGI	BspLU11I	BsrBI	BsrFI	BssHII	Bst1107I	BstEII	ClaI
EagI	EciI	Eco47III	EcoRV	FseI	HaeII	HgiEII	MluI
NarI	NdeI	NgoAIV	NheI	NotI	NruI	NsiI	NspI
Pfl1108I	PinAI	PmeI	PshAI	Psp1406I	RcaI	RleAI	RsrII
SacII	SanDI	SfiI	SgfI	SgrAI	SmaI	SnaBI	SpeI
SphI	SrfI	Sse8647I	StuI	SunI	SwaI	ThaI	Tth111I

Unique sites in Bam/Sal insert:



Appendix 2: Trypsin and Ecotin Purification

This appendix describes some of the details of the purification schemes I have developed for purifying trypsin from yeast, and ecotin from *E. coli*.

Trypsin Purification

1. Harvest media supernatant by centrifugation in 500ml bottles in GS3 rotor @7k rpm for 10 min.

2. Add 2M glycine, pH3.0 to culture supernatant to a concentration of 20mM. Adjust pH down to 3.0 with conc. HCl while stirring. Solution will probably turn slightly cloudy.

3. Filter solution through 0.45µm vacuum filter (500 ml capacity) using the supplied filter aids found in the box the filters come in. This is your SP column load. Take a 0.5ml sample of this and ethanol precipitate it for gel analysis.

4. Equilibrate a 25-50ml column (2.5cm ID)of SP sepharose FF (Pharmacia) with 20mM glycine, pH3.0. Pack this column as tightly as you can, and use a pump and flow adaptor for the column - this will give you the most theoretical plates, and increase the reproducibility of this chromatography.

5. Load the column at a linear flow rate of 120cm/hr or less. ($Q_{ml/min}=V_{cm/min} \times A_{cm}^2$), where Q is the volumetric flow rate, V is the linear flow rate, and A is the cross-sectional area of the column. Collect the flowthrough. The flowthrough should not contain any trypsinogen. Collect a 0.5ml sample of the flowthrough for EtOH precipitation and gel analysis.

6. After loading the column, wash it with 5 column volumes of equilibration buffer.

7. Elute the column with a 5-10 column volume gradient from 0 to 1.25M NaCl in equilibration buffer. I like to use only a 5 C.V. gradient of 150mls to minimize the volume handled in subsequent purification steps. Collect 4-5ml fractions. The trypsinogen should be eluted from the column at about 2/3 the way through the gradient, and usually lags just behind the most intense pigment peak.

8. Use 20µl of the fractions for samples on SDS-PAGE gels. Coomassie staining should let you see all you need to see if your expression levels are adequate.

9. Pool fractions containing trypsinogen and dialyze them against 10mM MES/1mM $CaCl_2$, pH6.0 @4°C overnight with at least one change. It is important to get rid of as much NaCl as you can in order to maximize the efficiency of the enteropeptidase cleavage.

   (Alternatively, you can pool the fractions and load them directly onto a phenyl sepharose column equilibrated with 20mM glycine/1M NaCl, pH3.0. The trypsinogen binds very well under these conditions. The column can be eluted by decreasing the NaCl concentration to 0 with a gradient, then running a 0-40% isopropanol gradient to elute the trypsinogen. This is a new step I was trying to improve the purification for crystallization trials. It needs optimization because the trypsinogen doesn't elute in a nice peak. Perhaps using a higher pH for this chromatography would help. If you include this step, dialyze the fractions containing trypsinogen to pH6.0 as above.)

10. Add 1/50 to 1/100 (w/w) of enteropeptidase to the pooled trypsinogen and incubate at 37°C for 1-2 hours to activate it to mature trypsin. This is the most problematic step in the purification. The main problem is this: active mutants of trypsin will start to degrade the product as soon as the enteropeptidase starts making it. The result is a tradeoff of yield for purity. Dead mutants of trypsin (*e.g.* H57A) do not suffer from this problem. Each case must be monitored either by following increase in trypsin activity or by SDS gels. This step need optimization or a solution to the problem.

11. After the activation step, load the activation solution onto an affinity resin, either BPTI agarose (affigel) or benzamidine agarose (Pierce) that has been equilibrated in 10mM MES/1mM $CaCl_2$, pH6.0. I've found that the BPTI agarose is better at removing yeast pigments from the prep than benzamidine agarose, but that yield may suffer as a result. Each tryspin mutant is different in this respect.

12. Wash the affinity column with 5-10 column volumes of 0.5M NaCl in equilibration buffer. Collect the wash for analysis.

13. Elute the trypsin with 100mM formic acid/0.5M NaCl. Collect 1ml fractions in siliconized eppendorf tubes and place them on ice. Take 20µl samples of the fractions for SDS gel analysis.

14. Pool the fractions containing the purest trypsin, and dialyze them exhaustively against 1mM HCl, pH3.0 .

15. Concentrate the trypsin using a centricon (Amicon) concentrator for final storage @4°C. Measure an A_{280} and calculate the trypsin yield using ε=34300 M^{-1}. Also check active-site titration. These two values should agree to within 10%. For trypsin destined for crystallization trials, concentrate to ~1mM (25mg/ml).

Ecotin Expression/Purification

Ecotin expression has been discussed elsewhere (Wang, et al, 1995) so it will not be described here to any detail. In short, ecotin is expressed in the periplasm of the *E. coli* strain 25A6 (W3110, *ton A, lon Δ, gal E, htp*P^{ts}) from the pTacTac plasmid (Muchmore, 1986). Cells are harvested by centrifugation, and the periplasmic fraction is isolated by treatment with lysozyme in 25% sucrose followed by another centrifugation to remove spheroblasts. At this point, I have modified the usual ecotin purification protocol to remove the boiling step, as isoelectric focusing get analysis of the resulting ecotins has shown that boiling produces unwanted derivatives of ecotin (data not shown here - it's in my notebook #5, p.1). The purification protocol I use is as follows.

1. Flow the periplasmic fraction through a small DEAE FF column that has been equilibrated with 10mM tris, pH8.0. Dialyze the flow-through fraction @4°C overnight against 1mM HCl to remove sucrose and precipitate some *E. coli* proteins. Change dialysis at least once. Alternatively, if you're in a hurry, you can skip this dialysis step, and proceed to step 2.

2. Add 2M glycine pH3.0 to the dialyzed periplasmic fraction and adjust the pH to 3.0. Remove precipitate by centrifugation and filter the supernatant through a 0.45μm vacuum filter.

3. Load this solution onto a 15ml SP sepharose FF column, wash the column with 5 column volumes of equilibration buffer, and elute the ecotin with a 5 column volume gradient of 0-1.5M NaCl in equilibration buffer.

4. Pool fractions containing ecotin, and dialyze them into 1mM HCl overnight @4°C.

5. Concentrate the pool down to <10ml for loading on to a C_4 reversed phase prep HPLC column equilibrated in 24% buffer B (95% acetonitrile/5% water/0.1%TFA). Load 3-5ml aliquots of the ecotin solution onto the C_4 prep column. The ecotin can be purified effectively by eluting the column with a 24-37% gradient of buffer B. This dynamax HPLC method is stored as C4.ec.prep or C4.ec.prep.B (this method saves solvent) on the prep HPLC computer.

6. Collect 5ml fractions from the HPLC, freeze them in crushed dry ice, and lyophilize them overnight. Resuspend the fractions in filtered DI water and analyze them on Coomassie-stained SDS gels.

7. Pool the fractions containing the purest ecotin, and concentrate them to ~1mM with a centricon, or centriprep concentrator. ε_{280}=21860 M^{-1}. Store the purified ecotin at 4°C.

These protocols have been used to purify several trypsin and ecotin mutants that have resulted in crystals of X-ray diffraction quality. They make a good starting point for someone trying to purify these two molecules for structural or functional work, but are open for optimization at several steps. Efforts to optimize the purification are predicated on the idea that pure protein will result in more reproducible crystallization. While this may not be true, at least you may be able to remove a variable in the "artful" process of protein crystallization.

Appendix 3: Crystallization and Metal Soaks

This appendix describes the protocol for preparing the trypsin and ecotin mutants for crystallization, and the subsequent transition metal ion soaks prior to mounting the crystals for data collection. Both the trypsin and ecotin stocks are first dialyzed into 1mM HCl, pH3.0 (or sometimes water for ecotin), concentrated to ~1mM in a centricon concentrator, and stored at 4°C until needed. The best results have been obtained when using proteins that are freshly purified. The complex formation and drop set up are as follows:

1. The molar concentration of each protein in the drops ranged from 300-375μM. The appropriate amount of purified trypsin was mixed with an equimolar amount of ecotin in a 0.5ml siliconized eppendorf tube on ice. This mixture was allowed to stand on ice for 30 min. to 1 hr.
2. 3 M NaOAc was added so that a final concentration of 300mM would result upon completion of protein complex set up.
3. 1 M Tris, pH 8.0 was added so that a final concentration of 100mM would result upon completion of protein complex set up.
4. Filtered DI water was added to bring the volume up to that required for the desired final reagent concentrations. The protein complex was kept on ice the entire time to minimize activity of trypsin at pH 8.0.
5. Two μl of protein mix were added to the surface of a siliconized cover slip, and 2μl of well buffer were added to that. Sometimes the two were mixed with the pipette, sometimes they were not. You be the judge.
6. All the crystallization trials were carried out at 18°C in the 8th floor cold room incubator. Crystals generally appeared after 3 days to a week if the tray was a good one. However, there were instances of crystal appearing only after a month or so.

Metal Soaks

In order to soak copper, nickel, or zinc into the crystals, the Tris concentration needed to be lowered from 100mM to 1mM. Therefore, the crystals were first transferred to 3ml of artificial mother liquor: 25% PEG 4K/300mM NaOAc/1mM Tris, pH 8.0/1mM $CaCl_2$, and soaked for at least 1hr @RT. Metal soaks were performed in steps in an effort to minimize shock to the crystals. Typical soaking steps were 30 min. to 2 hrs. with concentrations of1μM, 10μM, 100μM, and finally 1mM metal ion ($CuCl_2$, $NiCl_2$, or $ZnCl_2$). Most of the soaks were performed at room temperature, with some of the longer ones (some were overnight) at 18°C. As the metal concentration was increased, damage to the crystals became visible in the microscope as cracks or fissures in the crystals, with some crystals disintegrating completely. One advantage to this event was that it became easier to separate clusters of crystals when soaked in metal. Copper seemed to be the worst offender, damaging the crystals at lower metal concentration than nickel or zinc.

Dr. Linda Brinen had a good idea to back soak crystals quickly during the mounting process by floating the metal-soaked crystal through a plug of artificial mother liquor devoid of metal. We call this the "sinking backsoak." This turned out to work quite well, resulting in mounted crystals with the least amount of damage.

Preparation of metal-containing buffers. The kinetic results and the crystallography were dependent upon reproducible preparation of metal containing buffers. The problem with transition metal buffers is that metal ions are hydrated in aqueous solution. At alkaline pH where trypsin is most active, most of the metal is found in the metal-hydroxide form, which is

insoluble. Zinc and copper hydroxide seem to be less soluble than nickel hydroxide. Another complicating factor is that most common biological buffers bind metal ions to some extent. Therefore, the strategy we have adopted is to use as little buffer as is required (usually around 5mM), and to provide a "carrier" for the metal ions to keep them in solution at alkaline pH (McGrath et al., 1993). A "carrier" is a molecule which binds the metal ion preferentially over water, keeping it from forming the insoluble metal hydroxide species. We use Tris as a carrier, usually at 1mM concentration. I have found that as long as the ratio of Tris to metal is greater than 2, the metal will be soluble for long periods (more than 1 month) at pH 8.0. It is best to make up metal-containing buffers immediately prior to use. This is becuase slowly over time (more than a week or so) the metal hydroxide precipitate will form if it is going to. I keep a 1 M stock solution of $CuCl_2$, $NiCl_2$, and $ZnCl_2$, which have been sterile filtered at room temperature on the bench. These have lasted at least 5+ years (for the duration of my work here) and should be fine into perpetuity.

When adding metal solutions to buffers, keep in mind that they are good Lewis acids (electron deficient) and so will lower the pH of a solution in a concentration dependent manner (try checking the pH of a 1 M solution of $CuCl_2$). Because enzyme acitivities are highly dependent upon the pH of the solution, you must ensure that the resulting pH of your metal containing buffer is what you wish it to be.

Appendix 4: Fermentor Operation

I. Requirements

Be sure that there are no scheduled **air** or **water** interruptions. Read the manual. A little time spent reading can save major headaches later! Please pay attention to what you are doing, this is a potentially dangerous and expensive piece of equipment. Leave the fermentation area cleaner than when you found it.

II. Set up

A. Sterilize Attachments

Autoclave the following parts separately while filling and sterilizing the fermentor vessel:

a. a funnel wrapped in foil to simplify later additions,

b. air inlet and attached tubing,

c. acid/base/antifoam inlet and attached tubing,

d. dosing pump apparatus (pump head, tubing, and inlet),

e. a bottle for your sterile filtered DOSING additions,

f. any flasks which will be necessary to culture your inoculum.

Each of these inlets has a screw cap to cover the pointed end of the assembly. **Do not screw the caps on too tight**. Wrap foil around the end of each piece of tubing; *i.e.*, wrap the ends of the acid, base, and antifoam tubing separately.

B. Sterilize the Fermentor Vessel

1. Be sure the unit is clean.

2. Power up the unit by turning the key.

3. Turn on the air and cooling water valves beneath the sink.

4. Close the sampling valve, the black ball handle on the angle iron at the bottom of the fermentation chamber - CLOSED is DOWN.

5. Calibrate the pH electrode. Be sure the rubber cap is removed.

a. Turn the electrode pressure to 10 psi.

b. Check that the fluid levels of the electrode are at least as high as red and purple.

c. Rinse the electrode tip with distilled water.

d. Turn on the pH meter and the digital display. Turn the digital display selector to "pH".

e. Set the pH meter temperature to the temperature of the standardizing buffers.

f. Set the pH meter switch to **man.**

g. Immerse the pH probe in the first standardizing buffer.

h. Adjust the "ΔpH" knob to set the digital display to the pH of the buffer.

i. Rinse the tip and immerse it in the second standardizing buffer.

j. Adjust the "mv/pH" knob to set the digital display to the pH of the buffer.

k. Replace the pH probe in the fitting on the fermentation chamber.

l. Put enough water in the fermentation chamber to immerse the pH electrode.

m. Set the pH meter switch to **auto.**

6. Prepare the tank for assembly.

a. Orient the air inlet tube.

b. Prepare the culture media in the tank.

c. Add 2-3mls antifoam.

d. Stir until all media components are dissolved

e. Adjust the volume to the desired amount. Remember to take into account the volume of your inoculum and the volume of any added dosing.

f. Clean the top rim of the fermentation chamber with a little water and a kimwipe.

g. Be sure the O-ring is clean and seated properly. Remember Chernobyl, the Challenger, and the Walrus.

7 Put the top on top, making sure to match the air inlet with the drop tube (no air, no grow!).

a. Use the tightening tool to tighten the six (6) screws which hold the top on the tank. Be sure to use a rotating torque pattern (like on yer car tires).

b. Install the pressure gauge. Make sure you can read it while you are reading the control panel.

c. Install the pressure relief valve. Make sure the large black plastic ring spins freely by screwing it towards the round black knob. Consult the manual!

d. Install the foam sensor (the ceramic rod) near the center of the top away from other addition tubes, and plug it into the control panel.

f. Install the air exhaust condenser, and leave the coolant lines open to atmosphere .

g. Attach the exhaust hose to the exhaust condenser.

h. Plug all other ports with rubber septa, fix them into place with septum collars, and plug the septum collars with port plugs. Tighten the septum collars 1/8 turn past finger tight and only tighten the port plugs finger tight.

8. Set the stirring speed to 100 rpm.

9. Turn on the thermostat, and depress the "Fill Thermostat" switch until there's a steady flow from the waste water tube

10. **SET THE ELECTRODE PRESSURE TO 17 PSI!!!** The electrode pressure must ***always*** be greater than the tank pressure.

11. Double check, visually and manually, that the plugs are on top of the fermentation chamber and are tight.

12. Attach the blast shield.

15. Set Sterilization Time (30'). This is the amount of time that the fermentation chamber will be at sterilizing temperature and pressure.

16. Depress the "Vessel Size" switch to 5/10L.

17. Set sterilization switch to "Sterilization". Check the run periodically. The sterilization time indicator will flash until the fermentation chamber reaches sterilization temp. When sterilization begins, the sterilization time indicator will read out the amount of time left to sterilize, so you can estimate when the cycle will be done.

C. After Sterilization

1. When sterilization is complete, remove the blast shield. Attach inlet and outlet hoses of exhaust cooler and turn it on high. This will minimize loss of media due to evaporation. Expect to lose about 500 ml anyway. The fermentor will actively cool the fermentation chamber to the set fermentation temperature.

2. All attachments should be made as aseptically as possible. To help, add a small amount of 190 proof ethanol from the squirt bottle to the septum before making the attachment.

3. When the tank has reached fermentation temperature, attach the air inlet filter and hose. Remove the sheath from the attachment and poke it through the center of the septum of the air inlet port on the top of the tank. Attach the hose to the controller where it says aeration.

4. Turn the exhaust cooler valve to be open only $^1/_4$ to $^1/_2$ turn.

5. Attach the acid/base/antifoam inlets. Load the tubing into the respective peristaltic pumps and fill the acid, base, and antifoam tubes to the top of the inlet on top of the fermentor by depressing the manual buttons for each.

6. Attach the dosing inlet and pump head.

7. Equilibrate the fermentor to fermentation conditions.

   a. Set stirring rate to 750rpm.

   b. Activate pH control. Set the pH to the desired fermentation pH. Use the manual switch to hasten the approach to fermentation pH. (pH 7.4 is normative for *E. coli*)

8. Calibrate the pO_2 electrode. Turn on the pO_2 electrode, but not the control circuit.

    a. Purge the tank with argon. Argon is expensive; 5 minutes should be sufficient. Ensure that aeration input flow meters are all the way open.

    b. Set the pO_2 meter to zero using the zero knob.

    c. Saturate the media with air for ~5 min.

    d. Set the pO_2 meter to 100% by adjusting the slope knob.

    e. Set the pO_2 set point on the controller circuit to the desired level. (30% is normative for *E. coli*)

    f. Now turn on the pO_2 control circuit. You should hear the valve shut down the flow of air.

9. Set the electrode pressure to 10 psi again.

III. Culturing

1. Sterile filter a solution of desired additives into the ferment through one of the ports.

2. Inoculate.

IV. Harvesting

1. **Turn off the thermostat!!!!** Be certain that the pH, pO_2, and thermostat control circuits are active only as long as necessary and as long as the liquid in the bottom of the fermentation vessel covers the detectors for each system. Be sure to shut off the aeration valve before turning off the pO_2 control. It may be necessary to switch off the pO_2 control to allow air into the chamber to displace drained culture.

2. Turn off the pH, antifoam, exhaust cooler, and pO_2 control, and shut the valves to the aeration flowmeters.

V. Clean Up and Shutdown

1. Wash the fermentation chamber *i.e.*, it should be **clean** when you are done.

a. Fill the fermentation chamber with 10L of water.

b. Add a squirt of concentrated soap.

c. Stir the soapy water at 750 rpm ~ 5 minutes.

d. Stop the stirring, drain the soapy water, and rinse the fermentation chamber.

2. Remove the attachments from the top, rinse them clean with soap and water, dry them, and store them safely and securely in a place that's easily accessible.

   a. DO NOT LEAVE THE ATTACHMENTS IN A PILE WITH THE TUBING IN A GORDIAN KNOT.

   b. The foam sensor is brittle and should not be left where it can be crushed by some of the heavier attachments.

   c. The pressure gauge has a thin metal membrane that should be left in a safe place also.

   d. The air inlet attachment is heavy but has a thin metal tube which should not be bashed about into things. Also, the mass of this piece makes it a peril to other more delicate and fragile components.

3. Disconnect the aeration inlet from the top, unscrew the fastening bolts, and take the top off.

4. Clean the fermentation chamber top inside and out with soap and water. Rinse clean with water.

6. Scrub the inside of the fermentation chamber including the heating coils and stirring paddles with soap and water. Then rinse clean with water.

7. Replace the Viscolyt filled cap on the pH electrode.

8. Switch off the various control and metering systems on the control panel.

9. Turn the Air and Water valves off at the wall, and turn off the power to the fermentor.

Miscellaneous notes:

In the past phage-contaminated days, I've had to dissassemble the fermenter down to its nuts and bolts, and sterilize everything in site, including wiping down the entire fermentation room with iodine solution. While this is fun for me, I don't recommend this for anyone out to enjoy themselves. One thing to keep in mind when working with equipment like this, is that you should not have to force anything to go where it should go (this is a German-built machine, after all). The manual will explain just about anything you need to know: read it! A little water lubrication on O-rings goes a long way. Keep 0-rings squeeky clean, and always inspect them for signs of wear: remember, these have to hold an atmosphere of pressure during sterilization. It's a common practice in industry to perform a sterility-hold test on the tank after it's been sterilized with media in it. This should be at least 24 hours, but the tank should hold sterility out to 72 hours. The key to keeping this machine running well is to keep it scrupulously clean. This also ensures that you lower your risk of phage contamination in the room.

Appendix 5: Large-scale CsCl Plasmid Prep

A) **Solutions**

1. Solution I: 50mM glucose/25mM Tris/10mM EDTA, pH 8.0
2. Solution II: 200mM NaOH/1% SDS made fresh at RT
3. Solution III: 3M K/5M OAc (100g KOAc,60ml glacial AcOH,QS to 340ml)
4. 7.5M NH_4OAc
5. Isopropanol
6. TE
7. CsCl:HOH 1:1 (wt/wt)
8. 10 mg/ml ethidium bromide in water
9. CsCl saturated isopropanol/water mix

B) **Procedure** (per 1 liter of culture)

1. Spin down cells in GS3 rotor (7K rpm, 5 min). Discard sup.

2. Resuspend cells in 20ml Solution I in 250ml centrifuge bottle.

3. Add 40ml solution II. Invert gently to mix. Put on ice 10'.

4. Add 15ml solution III. Shake to mix until solution is not viscous. Leave on ice, 15'.

5. Spin in GSA in 250ml bottles for 20 min @10k rpm. Decant and save supernatant in clean 250ml centrifuge bottle. Volume should be ~70mls.

6. Add 35-40mls 7.5M NH_4OAc (1/2 vol), mix well, let sit on ice for 10 minutes, and centrifuge in GSA rotor for 20 min. @10k rpm. Decant supernatant into fresh 250ml centrifuge bottle.

7. Add 70 mls of isopropanol (~0.6 vol), incubate on ice for 30 minutes or over lunch, and pellet nucleic acid by centrifuging for 20 min. @10k rpm in the GSA rotor.

8. Decant supernatant and discard. Aspirate pellet until no residual liquid remains.

9. Resuspend pellet in 5 mls of TE and keep on ice. Mass out 4.6 g of CsCl into a 15 ml falcon tube and add resuspended pellet to it. Mix until all

the CsCl is dissolved, then add 0.5 ml ethidium bromide solution and place tube in the dark @RT to avoid UV damage.

10. Prepare large (5/8"x3") quick-seal tube: Carefully pipette 5 ml of 1:1 CsCl:H_2O into bottom of tube.

11. Centrifuge DNA-ethidium mix to remove precipitate by spinning @6k rpm in SA600 rotor for 10 minutes.

12. Pipette clarified DNA solution onto top of CsCl:H_2O in quick-seal tube with glass Pasteur pipette.

13. Fill remainder of tube with mineral oil, being careful to not introduce air bubble into the tube. This may take some practice.

14. Heat seal the tube, make sure that you have a balance tube that is ±50 mgs, and centrifuge overnight in the Beckman ultracenrifuge at 65k rpm in the Ti 80 rotor at 25°C. 12 hours should be sufficient.

15. Carefully remove the tubes from the centrifuge, and remove the DNA band from the tube with a 22 1/2 guage needle. Be sure to vent the top of the tube with another needle to avoid implosion. Place removed DNA band in 15 ml falcon tube.

16. Extract ethidium bromide from DNA by extraction with equal volume of saturated isopropanol:CsCl:water mix. Add saturated isopropanol to DNA, vortex, and let phases separate. Ethidium will turn bright pink in isopropanol phase. Remove this phase with P1000, using the blue tip as a mini separatory funnel. Repeat extractions until no color is observed in isopropanol phase.

17. Dilute DNA at least 3-fold with water to reduce CsCl concentration (typical volume of DNA after extraction is 1-1.5ml), and precipitate with cold 100% ethanol (~10mls)

18. Wash DNA with cold 70% ethanol, dry, and resuspend in 1ml TE. Quantitate yield with A_{260}.

Index