Substrate Specificity, Active-Site Residues, and Function of the Tsp Protease

by

Kenneth Charles Keiler

B.S., Chemistry and Biology
Stanford University
June, 1989

M.S., Biology
Stanford University
June, 1989

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the
Massachusetts Institute of Technology
November, 1995

Signature of Author Department of Biology

Certified by Robert T. Sauer, Thesis Supervisor

Accepted by Frank Solomon, Chairman, Biology Graduate Committee

Substrate Specificity, Active-Site Residues, and Function of the Tsp Protease

by

Kenneth Charles Keiler

Submitted to the Department of Biology
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

The work presented in this thesis examines the proteolytic activity and *in vivo* function of the Tsp protease. Tsp is a periplasmic protease from *Escherichia coli* that selectively degrades proteins based on the residues at their C termini. Studies of the substrate specificity, enzymatic properties, and active-site residues of Tsp uncover a novel proteolytic pathway involving co-translational tagging of substrates mediated by the *ssrA* RNA.

Chapter 1 presents an overview of substrate selection by proteases and cellular mechanisms used to increase proteolytic selectivity.

Chapter 2 describes the characterization of the substrate specificity and enzymatic properties of Tsp. Cleavage assays of protein and peptide substrates show that Tsp efficiently degrades substrates that have nonpolar residues and a free α-carboxylate at the C terminus. In addition, Tsp displays a preference for substrates that are not stably folded. Mapping of the cleavage sites of several substrates using mass spectrometry reveals that Tsp cleaves its substrates at a discrete number of sites, but with rather broad primary sequence specificity. These data are consistent with a model in which Tsp cleavage of a protein substrate involves binding to the C-terminal tail of the substrate, transient denaturation of the substrate, and then recognition and hydrolysis of specific peptide bonds. These experiments were published as "C-terminal specific protein degradation: Activity and substrate specificity of the Tsp protease" (Keiler, K. C., Silber, K. R., Downard, K. M., Papayannopoulos, I. A., Biemann, K., and Sauer, R. T. (1995) *Protein Science 4*, 1507-1515).

Chapter 3 describes the identification of the active-site residues of Tsp. Using site-directed mutagenesis, 20 positions of Tsp were individually substituted with alanine, and the resulting proteases were assayed for activity *in vitro* and *in vivo*. Alanine substitutions at three positions, Ser430, Asp441, and Lys455, result in inactive proteases that have structures and substrate-binding properties similar to wild type. These data are consistent with a serine-lysine dyad mechanism, similar to those proposed for the LexA-like proteases, the type I signal peptidases, and the class A β-lactamases. This work is in press for publication as "Identification of active-site residues of the Tsp protease" (Keiler, K. C., & Sauer, R. T. (1995) *Journal of Biological Chemistry*).

Chapter 4 describes the characterization of the substrate C-terminal sequences that are recognized by Tsp. Following randomization of the final three residue positions of cytochrome b_{562}-WVAAA, which is a good substrate for Tsp, 54 different mutants with single residue substitutions were recovered. The steady-state expression levels of cytochrome variants bearing these mutant tails and cleavage of the purified variants by Tsp *in vitro* indicate that the C-terminal sequence of cytochrome b_{562} variants is important in determining their proteolytic fate in the cell and show that this degradation is mediated predominantly by Tsp. Tsp prefers small, uncharged residues at the C terminus, and hydrophobic side chains at the second and third positions. These studies have been submitted for publication as "Substrate specificity of the Tsp protease" (Keiler, K. C., & Sauer, R. T. (1995) *Journal of Biological Chemistry*).

Chapter 5 describes a bacterial system in which proteins translated from mRNAs without stop codons are targetted for degradation by C-terminal addition of an 11 residue ssrA-encoded peptide tag. Periplasmic proteins containing the peptide tag are degraded by Tsp, and cytoplasmic proteins containing the peptide tag are degraded by a different protease. This work has been submitted for publication as "A peptide-tagging system mediates degradation of proteins synthesized from damaged mRNA" (Keiler, K. C., Waller, P. R. H., and Sauer, R. T. (1995) *Science*).

Thesis Supervisor: Robert T. Sauer

Title: Whitehead Professor of Biochemistry

for Sarah, who provided the backbone for my graduate career

Acknowledgements

First and foremost I thank Bob Sauer for allowing me to work in his lab, for providing the intellectual and technical support for this work, and for his guidance in my development as a scientist.

Thanks also to the members of my thesis committee, Peter Kim, Paul Schimmel, Frank Solomon, and Dan Finley, for their support and advice.

I am grateful to my co-authors for their contributions to the publications which comprise this thesis: Karen Silber, Patrick Waller, Kevin Downard, Ioannis Papayannopoulos, Klaus Biemann, and Bob Sauer.

This thesis would not have been possible without the work of Karen Silber, who pioneered the Tsp project and provided the framework for many of the studies presented here. I also thank Karen for her exceptional generosity with her knowledge and experience.

Thanks to other Sauer Lab members past and present, especially Dawn, John, Bronwen, Joel, Dennis, Harold, Cliff, and Tracy. Special thanks to Patrick for his level-headed help in the hectic final months of my tenure.

I am grateful to my previous mentors, Teresa Wang, Jesse Edwards, and Carolyn Lavallee, who provided me with the background of knowledge and interest to pursue my present research.

Thanks to my parents, Mary and Joel Keiler, and my sister Leslie, who have encouraged and supported me throughout my education.

Finally, I am forever indebted to Sarah Ades for sharing her insights and listening to my crazy ideas.

Table of Contents

List of Tables

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

CHAPTER 1

Mechanisms of Proteolytic Selectivity

INTRODUCTION

Proteolysis occurs in all organisms and is critical for diverse processes, such as cell cycle control, secretion, pattern formation, signal transduction, and hormone processing, and for removing proteins that are damaged, misfolded, or inactive. To carry out these processes, cells have evolved a wide array of different proteases, at least 36 in *E. coli* alone (Maurizi 1992). The prevalence of proteases and the number of different substrate targets pose a critical challenge for cells: how to keep proteases from inadvertently degrading the wrong proteins. Indeed, cells are extremely efficient at regulating proteolysis. In bacteria, only about 5% of cellular proteins have a half-life less than 2 hours, and another 10% have a half-life from 5-25 hours (Maurizi 1992). The remaining proteins appear to be stable indefinitely *in vivo*, suggesting that they are not recognized as substrates by any of the proteases in the cell.

How is proteolytic selectivity attained? Substrate binding entails recognition of some combination of the primary, secondary, tertiary, or quaternary structure of the target protein. In this sense, proteolytic recognition is similar to any other protein-protein interaction. However, because the cleavage of a peptide bond is essentially irreversible in the cell, it is of utmost importance to prevent unwanted interactions. Several strategies have been employed to prevent inadvertent proteolysis. First, some proteases are exquisitely specific: they cleave only a minute fraction of the available peptide bonds. Specificity is generated through increasing discrimination among residues at the binding site, increasing the length of the amino acid sequence that is recognized, or combining elements of primary, secondary, tertiary and quaternary structure to limit the number of proteins that are recognized as substrates. Second, some proteases are isolated from the bulk of cellular protein in lysosomes, membranes, or macromolecular structures. In

these cases, the substrates are recognized by other proteins and transported to the compartment where they are degraded. Third, in some cases the recognition event is separated from the proteolytic event by using molecular tags. In systems such as the ubiquitin-proteasome pathway and the *ssrA* C-tagging pathway, substrates are recognized by enzymes that target the substrate for proteolysis by attaching a molecular tag. The protease can then recognize the tagged protein, or a combination of the tag and other features of the protein to further increase selectivity. This process also provides the opportunity for editing by removing a tag before proteolysis has occurred.

STRUCTURAL DETERMINANTS OF PROTEOLYIC CLEAVAGE

Whether proteolytic selectivity is generated through a highly specific protease, or by using additional enzymes to select appropriate substrates, the problem of recognizing a substrate among all the proteins in the cell is the same. *A priori*, any structural feature might be used as a proteolytic determinant. Primary sequence determinants are the most common among well-characterized proteases, but secondary, tertiary, and quaternary structures also serve as recognition determinants. In many cases, the unfolded structures of proteins are the forms recognized by proteases. Damage to proteins from chemical modification results in abnormal residues or structures that could also be recognized by proteases. It is likely that in most cases proteases recognize their substrates using a combination of structural features to unambiguously identify them among all the cellular proteins. The following sections examine structural features that have been identified as determinants of proteolysis.

Primary Structure: The most direct way to increase the specificity of a protease is to limit the number of peptide bonds it can cleave. In many cases, determinants are adjacent to the scissile bond, in the P1 and P1' positions (nomenclature of Schechter & Berger 1967). Relatively non-specific proteases, such as trypsin and elastase, cleave the peptide bond adjacent to a certain class of residues, positively charged residues in the case of trypsin and small, nonpolar residues in the case of elastase (Beynon & Bond 1989). Since all proteins are likely to have such residues, these enzymes are not very selective. However, if the requirements are extended beyond the P1 and P1' positions, a high degree of specificity can be achieved. For example, the Gpr protease from *Bacillus megaterium* cleaves substrate proteins at a seven residue motif Thr-Glu-(Phe,Ile)-Ala-Ser-Glu-Phe (Setlow 1988), which would occur at random about once every 3×10^8 peptide bonds. Studies with the Lon protease from *E. coli* suggest that the cleavage sites fall within the consensus sequence Φ–X_{3-4}-Leu-Ser-(Leu,X)-X_5-Ser-X-Φ (where Φ represents a hydrophobic residue and Xaa represents any residue), a sequence expected by chance in only about 1% of proteins (Gottesman & Maurizi 1992).

Primary sequence specificity is not necessarily limited to the cleavage site. Some proteases recognize and bind sequences in one region of their substrates, and cleave at remote, relatively nonspecific sites. Several examples of sequences remote from the cleavage site affecting proteolysis have been documented, such as the "destruction box" found in eukaryotic cyclins (Glotzer et al. 1991) and N- (Varshavsky 1992) and C-terminal sequences (Bowie & Sauer 1989; Parsell et al. 1990), but most of these signals appear to involve targetting or tagging pathways that are discussed below. There is no reason why a protease could not recognize a substrate sequence distinct from the cleavage site in addition to residues near the scissile bond, and such a determinant has been suggested for both Lon and Clp (Gottesman & Maurizi 1992).

Secondary Structure: Because of the limited forms of secondary structure that are found in proteins, proteases could not be very selective based on secondary structure alone. However, recognition of secondary structure in conjunction with sequence determinants increases the specificity of several proteases. Magaininase, a skin protease of *Xenopus laevis*, cleaves substrate peptides at a lysine residue within an amphipathic α-helix (Resnick et al. 1991). If the amphipathic nature of the helix is disrupted by deletion or insertion of charged residues, or if the lysine residue is moved to the hydrophobic face of the helix, the peptide is not recognized as a substrate. The poliovirus protease 3C cleaves the viral polyproprotein at Gln-Gly-Pro bonds within a turn or bend (Ypma-Wong et al. 1988). If the secondary structure of model peptide substrates is disrupted by removing the proline or by adding glycine residues at other positions, the peptide is not cleaved (Weidner & Dunn 1991). The crystal structure of the related rhinovirus protease 3C provides an explanation for the required secondary structure. The active site of the protease is in a crooked "canyon", and a bend is required to place the Gln-Gly bond in the catalytic site (Matthews et al. 1994). There are also many proteases that recognize their primary sequence motifs in a random coil. In the case of Lon protease, it appears that the primary sequence motif is recognized in an extended conformation (Waxman & Goldberg 1986). Recognition of a random coil could occur within the context of the native structure of the substrate protein, in a transiently unfolded region, or in a random coil resulting from permanent loss of the native structure through damage or misfolding. The number of proteases that have a preference for unfolded substrates suggests that this is a common means of structural recognition. However, it is possible that coil recognition is a secondary activity of many proteases, used to eliminate damaged or misfolded proteins, in addition to their specific substrates.

Tertiary and Quaternary Structure: Tertiary and quaternary structure have been identified most frequently as negative determinants of proteolysis.

Sequence determinants that would be recognized by proteases can be inaccessible in the native fold or buried in the protein-protein interface. The stability of the native structure has been shown to be an important determinant of proteolytic susceptibility in both prokaryotes and eukaryotes (Goldberg & St. John 1976; Parsell & Sauer 1989). In *E. coli* both Lon (Gottesman 1989) and Clp (Maurizi 1992) have been implicated in the degradation of unfolded proteins, and in eukaryotes the proteasome appears to be the major protease for unfolded proteins (Hershko & Ciechanover 1992; Jentsch 1992). However, the features of unfolded proteins that are recognized by these proteases have not been described.

The incorporation of proteins in multimeric complexes is also a determinant of proteolytic susceptibility. As with tertiary structure, quaternary interactions seem to serve primarily to shield proteins from degradation, most likely by masking recognition sites. Some proteins are selectively degraded to regulate the activity of the complex, while others are degraded only when they are found in excess of other components of the complex, presumably for housekeeping reasons. Examples of housekeeping include degradation of over-expressed ribosomal subunits, which are rapidly removed when they are in abundance of other subunits (Petersen 1990). Quaternary signals that appear to have a regulatory function include RcsA and RcsB in *E. coli*. RcsA is rapidly degraded by Lon, but is protected from degradation when it is bound to RcsB (Gottesman and Maurizi 1992). The selective proteolysis of RcsA can therefore be controlled by expression of RcsB.

There are a few examples of quaternary interactions that act as positive determinants of proteolysis. The p53 tumor suppressor protein is degraded rapidly via the ubiquitin-proteasome pathway (discussed below) when bound to the E6 oncoprotein. The proteolytic determinant, although it has not been identified, appears to be in the E6 protein. Binding to E6 targets p53 in a *trans* fashion for ubiquitination and subsequent degradation. (Scheffner et al. 1990; 1993)

Modified Proteins: Proteolysis of modified proteins is an important mechanism for removing abnormal proteins from the cell. A wide array of modifications, including oxidation, alkylation, esterification, amino acylation, deamination, and glycosylation, lead to rapid protein turnover *in vivo* (Stadtman 1990). In many cases, damaged proteins may be recognized through changes in tertiary or quaternary structure, but there is no reason *a priori* that modification of residues on the surface of proteins would lead to large structural changes. Therefore, it would be advantageous to recognize the modified elements of proteins directly, and to selectively degrade the modified proteins. Few of the proteases responsible for degrading modified proteins have been identified. It appears that the Tsp protease is identical to an activity in *E. coli* that rapidly degrades oxidized glutamine synthetase (Silber et al. 1992; Roseman and Levine 1986), but the mode of recognition of this substrate is not known.

SPATIAL ISOLATION OF PROTEASES

Proteolytic selectivity can be attained or increased through spatial isolation of proteases from the bulk of cellular proteins. The proteases responsible for lysosomal degradation are relatively non-specific. For example cathepsin D has an extremely broad specificity (P1 non-specific, P1' not Ala, Val, or Ile), and cathepsin B is a non-specific dipeptidylpeptidase (Beynon & Bond 1989). While these proteases will degrade most proteins they encounter, the proteases are isolated to lysosomes, and the only proteins they encounter have already been selected and transported for degradation.

Using RNase A as a probe for selective lysosomal proteolysis, Dice and coworkers found that sequences related to Lys-Phe-Glu-Arg-Gln target cytosolic proteins for lysosomal degradation (Dice et al. 1986; reviewed Dice 1990). The recognition and transport pathways have not been characterized, but prp73, a member of the hsp 70 family, binds the Lys-Phe-Glu-Arg-Gln region and is required for transport into lysosomes *in vitro* (Chiang et al. 1989).

Another method for spatially isolating proteases is through incorporation in multimeric complexes. The hallmark example is the proteasome, which contains most of the proteolytic activity in the cytoplasm of eukaryotes (Rock 1994). The 26S proteasome is a multimeric complex comprised of a 20S catalytic core, containing 28 subunits arranged in a cylindrical structure with a channel through the center (Lowe et al. 1995), and two 19S "caps" that bind over the openings of the channel and have regulatory functions (Peters et al. 1994; Hoffman et al. 1992). The proteasome contains at least five distinct proteolytic activities that are localized in the central channel (Orlowski et al. 1993; Djaballah et al. 1992), so they are not exposed to the bulk of cellular proteins. Only proteins that pass though the channel are subject to degradation. The methods for selecting proteins for degradation and inserting them into the channel have not been well-characterized, but some involve the ubiquitination pathway discussed below.

The signal peptide peptidases are twice removed from the bulk of cellular protein, localized in mutimeric complexes that are inserted in the membrane. The signal peptide peptidases are responsible for cleaving the signal sequence from secreted proteins as they cross the membrane. The recognition determinants for the signal peptide peptidases are relatively non-specific, requiring a ten residue hydrophobic region in the signal peptide (reviewed by von Heijne 1988). Nevertheless, signal peptide peptidases are highly selective because they are isolated in a membrane complex with other proteins in the translocation apparatus. The peptidases only encounter

proteins that have been selected for secretion. The pathway for selecting these proteins involves recognition by the signal recognition particle, a ribonucleoprotein complex, and subsequent interaction with chaperones, membrane receptors, and a protein pore in the membrane (reviewed by Luirink & Dobberstein 1994).

TAGGING SYSTEMS FOR SELECTIVE PROTEOLYSIS

The selectivity of proteolysis can also be increased by tagging systems, which spatially and temporally isolate substrate selection from the protease. The general features of these systems are that a substrate is recognized by an enzyme other than the protease and marked for degradation by covalently attaching a molecular tag. The protease then recognizes the tag and degrades the substrate. The selectivity of tagging systems arises from the combination of two separate recognition events. The protease screens only a subset of the cellular proteins, and may recognize features of the protein in addition to the tag to increase selectivity. Moreover, tagging systems provide the opportunity to proof-read the identification of a substrate before degradation.

The most thoroughly characterized tagging system is the ubiquitin-proteasome degradation pathway (for reviews see Jentsch & Schlenker 1995; Hochstrasser 1995; Ciechanover 1995). In this pathway, proteolytically unstable proteins are tagged for degradation through covalent attachment of ubiquitin, a small protein that is highly abundant in eukaryotic cells, and degraded by the proteasome (discussed above). Substrates for this pathway are recognized by a ubiquitin conjugating enzyme (E2), which mediates covalent attachment of ubiquitin to a lysine in the substrate, in many cases through

interaction with a ubiquitin ligase (E3). The enzymes responsible for most of the selectivity in this system are the E2s and E3s. There are many members of the E2 and E3 families, each of which recognizes a subset of the ubiquitin-proteasome substrates. Several ubiquitination signals have been identified to date. These include the identity of the N-terminal residue (N-end rule) (Varshavsky 1992), the "destruction box" of cyclins (Glotzer et al. 1991), and the *trans*-acting signal of E6-p53 discussed above. Through the ubiqitin-proteasome pathway a variety of signals converge on the same protease complex. This allows most of the proteolytic activity in the cytoplasm to be sequestered and accessed only by marked substrates (although some non-ubiquitinated proteins may be degraded by the proteasome as well). In addition to isolating the proteolytic activities from the non-substrate proteins in the cell, the ubiquitin-proteasome pathway may increase proteolytic selectivity by editing inadvertently tagged proteins. Some proteins that are not short-lived become ubiquitinated, either through mistakes in recognition or as part of a separate pathway, but the ubiquitin is removed by isopeptidases before proteolysis can occur (Ciechanover 1994).

Recently a proteolytic tagging system has been discovered in bacteria (Chapter 5). The central factor in this system is*ssrA* RNA, a small, metabolically stable RNA found in Gram-negative and Gram-positive bacteria, which tags proteins translated from damaged mRNAs to target them for degradation. One portion of the *ssrA* RNA can fold into a structure resembling the acceptor stem of tRNA, and has been shown to be aminoacylated with alanine by alanyl-tRNA synthetase *in vitro* (Komine et al. 1994; Ushida et al. 1994) Another portion of the *ssrA* RNA has an open reading frame that codes for a peptide that has been found at the C terminus of some peptide fragments of recombinant interleukin-6 (Tu et al. 1995). In Chapter 5, I show that the *ssrA* RNA mediates the attachment of an eleven residue peptide to the C terminus of nascent polypeptides transcribed from truncated transcripts. This peptide targets the protein for proteolysis both in

the cytoplasm and periplasm. In the periplasm, the Tsp protease is primarily responsible for degradation of these tagged proteins.

THE TSP PROTEASE

The C-terminal sequences of proteins have a profound affect on their proteolytic suscptibility in bacteria (Bowie & Sauer 1989; Parsell et al. 1990). The Tsp protease was purified as an activity that specifically degrades a variant of the N-terminal domain of λ-repressor with nonpolar residues (Trp-Val-Ala-Ala-Ala) at the C terminus, and does not degrade a variant with polar residues (Arg-Ser-Glu-Tyr-Glu) at the C terminus (Silber et al. 1992). A deletion of the *tsp* gene does not affect the half-lives of λ-repressor variants *in vivo*, suggesting that there are additional tail-specific proteases in *E. coli* (Silber & Sauer 1994). Nevertheless, Tsp is responsible for a majority of the tail-specific proteolytic activity in the periplasm (Chapter 4). The work in this thesis examines the characteristics of proteins and peptides that make them substrates for Tsp *in vitro* and *in vivo* (Chapters 2 & 4). The basic enzymatic properties (Chapter 2) and catalytic residues of Tsp are determined (Chapter 3), and a model for the catalytic mechanism is presented (Chapter 3). The role of Tsp *in vivo* is discussed in the context of the *ssrA* RNA tagging pathway (Chapter 5).

REFERENCES

Beynon, R. J. and Bond, J. S. (1989) "Proteolytic enzymes: a practical approach." IRL Press: Oxford.

Bowie, J. U. and Sauer, R. T. (1989) "Identification of C-terminal extensions that protect proteins from intracellular proteolysis." *J Biol Chem* **264** 7596-7602.

Chiang, H.-L., Terlecky, S. R., Plant, C. P., and Dice, J. F. (1989) "A role for a 70-kilodalton heat shock protein in lysosomal degradation of intracellular proteins." *Science* **246**, 382-385.

Ciechanover, A. (1994) "The ubiquitin-proteasome proteolytic pathway." *Cell* **79**, 13-21.

Dice, J. F. (1990) "Peptide sequences that target cytosolic proteins for lysosomal proteolysis." *Trends Biochem. Sci.* **15**, 305-309.

Dice, J. F., Chiang, H.-L., Spencer, E., and Backer, J. M. (1986) "Regulation of catabolism of microinjected ribonuclease A requires the amino-terminal 20 amino acids." *J. Biol. Chem.* **261**, 6853-6859.

Djaballah, H., Harness, J. A., Savory, P. J., and Rivett, A. J. (1992) "Use of serine-protease inhibitors as probes for the different proteolytic activities of the rat liver multicatalytic proteinase complex." *Eur. J. Biochem.* **209**, 629-634.

Glotzer, M. A., Murray, A. W., and Kirschner, M. W. (1991) "Cyclin is degraded by the ubiquitin pathway." *Nature* **349**, 132-138.

Goldberg, A. L. and St. John, A. C. (1976) "Intracellular protein degradation in mammalian and bacterial cells." *Annu. Rev. Biochem.* **45**, 747-803.

Gottesman, S. and Maurizi, M. R. (1992) "Regulation by proteolysis: energy-dependent proteases and their targets." *Microbiol. Rev.* **56**, 592-621.

Gottesman, S. (1989) "Genetics of proteolysis in *Escherichia coli*." *Annu. Rev. Genet.* **23**, 163-198.

von Heijne, G. (1988) "Transcending the impenetrable: how proteins come to terms with membranes." *Biochim. Biophys. Acta* **947**, 307-333.

Hershko, A. and Ciechanover, A. (1992) "The ubiquitin system for protein degradation." *Annu. Rev. Biochem*, **61**, 761-807.

Hochstrasser, M. (1995) "Ubiquitin, proteasomes, and the regulation of intracellular protein degradation." *Curr. Opin. Cell Biol.* **7**, 215-223.

Hoffman, L., Pratt, G., and Rechsteiner, M. (1992) "Multiple forms of the 20S multicatalytic and the 26S ubiquitin/ATP-dependent proteases from rabbit reticulocyte lysate." *J. Biol. Chem.* **267**, 22362-22368.

Jentsch, S. and Schlenker, S. (1995) "Selective protein degradation: a journey's end within the proteasome." *Cell* **82**, 881-884.

Jentsch, S. (1992) "The ubiquitin-conjugation system." *Annu. Rev. Genet.* **26**, 179-207.

Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K., and Inokuchi, H. (1994) "A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*." *Proc. Natl. Acad. Sci. USA* **91**, 9223-9227.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., Huber, R. (1995) "Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* Å resolution." *Science* **268**, 533-582.

Luirink, J. and Dobberstein, B. (1994) "Mammalian and *Escherichia coli* signal recognition particles." *Mol. Microbiol.* **11**, 9-13.

Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L., and Worland, S. (1994) "Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein." *Cell* **77**, 761-771.

Maurizi, M. R. (1992) "Proteases and protein degradation in *Escherichia coli*," *Experimentia* **48**, 178-201.

Orlowski, M., Cardozo, C., and Michaud, C. (1993) "Evidence for the presence of five distinct proteolytic components in the pituitary multicatalytic proteinase complex." *Biochemistry* **32**, 1563-1572.

Parsell, D. A., Silber K. R., and Sauer, R. T. (1990) "Carboxy-terminal determinants of intracellular protein degradation." *Genes & Dev.* **4**, 277-286.

Parsell, D. A. and Sauer, R. T. (1989) "The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*." *J. Biol. Chem.* **264**, 7590-7595.

Peters, J. M., Franke, W. W., and Kleinschmidt, J. A., (1994) "Distinct 19S and 20S subcomplexes of the 26S proteasome and their distribution in the nucleus and the cytoplasm." *J. Biol. Chem.* **269**, 7709-7718.

Petersen, C. (1990) "*Escherichia coli* ribosomal protein L10 is rapidly degraded when synthesized in excess of ribosomal protein L7/L12." *J. Bacteriol.* **172**, 431-436.

Resnick, N. M., Maloy, W. L., Guy, H. R., and Zasloff, M. (1991) "A novel endopeptidase from *Xenopus* that recognizes α-helical secondary structure." *Cell* **66**, 541-554.

Rock, K. L., Gramm, C., Rothsrein, L., Clark, K., Stein, R. Dick, L., Hwang, D., and Goldberg, A. L. (1994) "Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules." *Cell* **78**, 761-771.

Roseman, J. E. and Levine, R. L. (1986) "Purification of a protease from *Escherichia coli* with specificity for oxidized glutamine synthetase." *J. Biol. Chem.* **262**, 2101-2110.

Schechter, I. and Berger, A. (1967) "On the active site of proteases." *Biochem. Biophys. Res. Commun.* **27**, 157-162.

Scheffner, M., Huibregtse, J. M., Vierstra R. D., and Howley, P. M. (1993) "The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53." *Cell* **75**, 495-505.

Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J., and Howley, P. M. (1990) "The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53." *Cell* **63**, 1129-1136.

Setlow, P. (1988) "Small acid-soluble spore proteins of *Bacillus* species: structure, synthesis, genetics, function, and degradation." *Annu. Rev. Microbiol.* **42**, 319-338.

Silber, K. R. and Sauer, R. T. (1994) "Deletion of the *prc* (*tsp*) gene provides evidence for additional tail-specific proteolytic activity in *Escherichia coli* K12." *Mol. Gen. Genet.* **242**, 237-240.

Silber, K. R., Keiler, K. C., and Sauer, R. T. (1992) "Tsp: a tail-specific protease that selectively degrades proteins with nonpolar C termini." *Proc. Natl. Acad. Sci.* **89**, 295-299.

Stadtman, E. R. (1990) "Covalent modification reactions are marking steps in protein turnover." *Biochemistry* **29**, 6323-6331.

Tu, G.-F., Reid, G. E., Zhang, J.-G., Moritz, R. L., and Simpson, R. J. (1995) "C-terminal extension of truncated recombinant proteins in *Escherichia coli* with a 10Sa RNA decapeptide." *J. Biol. Chem.* **270**, 9322-9326.

Ushida, C., Himeno, H., Watanabe, T., and Muto, A. (1994) "tRNA-like structures in 10Sa RNAs of *Mycoplasma capricolum* and *Bacillus subtilis*." *Nucleic Acids Res.* **22**, 3392-3396.

Varshavsky, A. (1992) "The N-end rule." *Cell* **69**, 725-735.

Waxman, L. and Goldberg, A. L. (1986) "Selectivity of intracellular proteolysis: protein substrates activate the ATP-dependent protease (La)." *Science* **232**, 500-503.

Weidner, J. R. and Dunn, B. M. (1991) "Development of synthetic peptide substrates for the poliovirus 3C proteinase." *Arch. Biochem. Biophys.* **286**, 402-408.

Ypma-Wong, M. F., Filman, D. J., Hogle, J. M., and Semler, B. L. (1988) "Structural domains of the poliovirus polyprotein are major determinants for proteolytic cleavage at Gln-Gly pairs." *J. Biol. Chem.* **263**, 17846-17856.

CHAPTER 2

C-terminal Specific Protein Degradation: Activity and Substrate Specificity of the Tsp Protease

INTRODUCTION

Proteases are ubiquitous in biological systems. The structures, properties, and mechanisms of a large number and variety of extracellular proteases are known (Beynon and Bond, 1989), but our knowledge of intracellular proteases is far less complete. For example, intracellular proteolysis is extremely selective (Goldberg, 1992; Gottesman and Maurizi, 1992) and yet the simple question of what determines the substrate specificity of an intracellular protease has in no case been clearly answered. In yeast and higher eukaryotes, some proteins are marked for degradation by modification with ubiquitin (Hershko and Ciechanover, 1992), but in bacteria there is no ubiquitin system. Factors that have been shown to be important for the proteolytic fate of specific bacterial proteins include thermal stability (Parsell and Sauer, 1989), the identity of N-terminal sequences (Tobias et al., 1991), and the identity of C-terminal sequences (Bowie and Sauer, 1989; Parsell et al., 1990). A better understanding of proteolytic selectivity requires purification of the relevant proteases and biochemical studies of their interactions with substrates.

Tsp is a periplasmic protease of *Escherischia coli* that was purified based upon its ability to selectively degrade a variant of the N-terminal domain of λ repressor with the apolar C-terminal sequence WVAAA, but not to degrade a variant with the polar C-terminal sequence RSEYE (Silber et al., 1992). The gene encoding Tsp has been called both *tsp* and *prc* (Silber et al., 1992; Hara et al., 1991), and Tsp may be identical with protease Re (Park et al., 1988) and

with a protease purified based upon its ability to cleave oxidized glutamine synthetase (Roseman and Levine, 1987). The λ repressor variant that is a substrate for Tsp *in vitro* is still degraded in cells deleted for the *tsp* gene, indicating that another protease must be responsible for degradation of this substrate in the deletion strain (Silber and Sauer, 1994). However, cells deleted for the *tsp* gene are defective in the C-terminal processing of two periplasmic proteins, penicillin-binding protein 3 (Nagasawa et al., 1989) and tonB (R. Larsen and K. Postle, personal communication), suggesting that these proteins may be natural substrates of the Tsp protease. Tsp has no sequence homology to any characterized protease, and is not inhibited by a wide array of small-molecule protease inhibitors (Silber et al., 1992). *ctpA*, a gene with homology to *tsp*, has recently been identified in the cyanobacterium *Synechocystis* sp. PCC 6803 and is believed to encode a protease that is responsible for C-terminal processing of the D1 protein (Anbudurai et al., 1994).

In this chapter, I probe the characteristics of peptides and proteins that make them substrates for Tsp *in vitro*, determine the location of cleavage sites in several protein and peptide substrates, and present a preliminary characterization of the kinetic properties of Tsp. These studies confirm the importance of C-terminal sequence in determining substrate preference but show that additional factors are also involved. One of these factors appears to be thermodynamic stability because experiments with a set of related substrates with differing stabilities indicate that the denatured protein is the only proteolytically susceptible form. Another factor may be the presence of appropriate cleavage sites because I find that Tsp makes endoproteolytic cuts in its substrates at a limited number of discrete bonds.

RESULTS

Comparison of Tsp and $TspH_6$—To simplify the purification of Tsp, I constructed a *tsp* gene encoding six additional C-terminal histidine residues and purified the resulting gene product, called $TspH_6$, using nickel chelate chromatography. The CD spectra of Tsp and $TspH_6$ are almost identical and consistent with approximately 35-40% α-helix (see Fig. 1). The molecular weight of $TspH_6$ was determined to be 78.5 ± 5.3 kDa by sedimentation equilibrium centrifugation. The polypeptide molecular weight calculated from the sequence is 75.0 kDa, indicating that $TspH_6$ is monomeric, as has previously been shown for Tsp (Silber et al., 1992).

As shown in Figure 2, Tsp and $TspH_6$ have very similar activities in cleaving several protein substrates. In addition, time courses of the cleavage of Arc repressor and the #105 variant of the N-terminal domain of λ repressor by $TspH_6$ and Tsp are indistinguishable (not shown). These results indicate that the histidine tag does not affect the specificity or kinetic properties of Tsp. Moreover, because Tsp and $TspH_6$ are purified by completely different methods and yet display the same enzymatic activity, it is extremely unlikely that the observed proteolytic activity is due to a contaminant. The experiments described in the following sections were performed using the $TspH_6$ enzyme.

Importance of C-terminal sequences in different protein and peptide substrates —Four sets of proteins or peptides were tested as potential substrates for Tsp (Table 1). Within each set, the amount of cleavage is sensitive to the identity of residues or functional groups at the C-terminus.

For example, the two λ repressor variants and four Arc variants shown in Table 1A differ only in their C-terminal sequences and yet show dramatically different sensitivities to Tsp cleavage. In single point cleavage assays, the variants with relatively apolar C-terminal tails (λ #105, Arc, and Arc-st7) are good substrates for Tsp, whereas those with charged or highly polar tails (λ N-terminal domain, Arc-st5, and Arc-st6) are poor substrates. In addition, even though each of the Arc variants has the apolar sequence GRIGA in its C-terminal tail, only those with this sequence at the C-terminus (Arc and Arc-st7) are good substrates.

The presence of a free α-carboxyl group is also important in determining whether closely related peptides with apolar C-terminal sequences are cleaved efficiently by Tsp. As shown in Table 1B, peptides BAS7 and BAS9 differ only in whether the C-terminus is a carboxyl or carboxyamide group; the same is true for peptides BAS8 and BAS10. In both cases, the peptide with the α-carboxyl group is cleaved readily by Tsp, whereas the variant with the α-carboxyamide group is cleaved poorly or not cleaved at all (Table 1B; Fig. 3).

Site of proteolytic cleavage—To determine the sites at which Tsp cleaves its substrates, the digestion products of Arc, Arc-st7, the #105 variant of the N-terminal domain of λ repressor, and peptide BAS10 were analyzed by MALDI-TOF mass spectrometry. The major products from the digestion of each substrate are listed in Table 2 and are shown schematically in Figure 4. Several conclusions emerge from these studies. First, Tsp is clearly an endoprotease. Although Tsp recognizes determinants at the C-terminus of substrates, it cleaves these substrates at a limited number of discrete sites

within each polypeptide chain (Table 2; Fig. 4). These sites occur at different distances in the sequence from the C-terminus and map to diverse locations in the native secondary and tertiary structures. The P1 position of the scissile peptide bond is Ala, Ser, or Val in 14 of the 20 cleavage sites; Ile, Leu, Arg, and Lys occupy the P1 position in the remaining 6 sites. Hence, Tsp appears to prefer P1 residues that are small and uncharged or that have aliphatic side chains extending through the $C\gamma$, $C\delta$, or $C\varepsilon$ positions. The same general preferences are evident at the P1' position of the scissile peptide bond except there is no discernible preference for a small residue and the repertoire of allowed residues includes Met, Tyr, and Trp.

Kinetic properties and the importance of substrate stability—To determine approximate kinetic parameters for Tsp cleavage of a protein and a peptide substrate, initial rates of cleavage were determined for different substrate concentrations of Arc and peptide BAS9 by using CD to monitor the loss of α-helicity that occurs upon degradation. Figure 5A and B shows the rate versus substrate concentration plots for Tsp cleavage of Arc and BAS9, respectively. Fitting of these data yields the apparent kinetic parameters shown in Table 3. The apparent K_m for Tsp cleavage of peptide BAS9 is comparable to the K_m observed for Tsp cleavage of Arc, but k_{cat} and k_{cat}/K_m are almost 30-fold higher for cleavage of the peptide. Thus, Tsp-catalyzed degradation of peptide BAS9 significantly more efficient than Tsp-catalyzed degradation of Arc. By contrast, Tsp and chymotrypsin have similar values of k_{cat}/K_m using Arc as a substrate (Table 3).

One factor that might influence the efficiency with which Tsp cleaves different proteins or peptides is the equilibrium between the native and

denatured forms of the substrates. This could occur, for example, if the denatured substrate were actually the form in which the peptide bond was cleaved (see Discussion). To test this idea, the susceptibility to cleavage of Arc repressor variants with decreased stability (Arc-NK29, lysine for asparagine at position 29) or increased stability (Arc-PL8, leucine for proline at position 8) were compared with that of wild-type Arc. Arc-NK29 and Arc-PL8 each differ from wild-type Arc at a single amino acid residue and have the same apolar C-teminal sequence as Arc. Moreover, neither of the sites of mutation is adjacent to a scissile bond. The melting temperatures of these three proteins under comparable conditions are approximately 37 °C (Arc-NK29), 50 °C (Arc) and 70 °C (Arc-PL8) (M. Milla, personal communication). As shown in Figure 6, the hyperstable PL8 Arc variant is cleaved significantly more slowly than Arc, whereas the unstable NK29 variant is cleaved significantly more rapidly than Arc. These data suggest that substrates are cleaved only in a denatured form.

DISCUSSION

In this work, I have investigated the substrate specificity of Tsp and probed some of its basic enzymatic properties. Previous studies had shown that Tsp cleaves a variant of λ repressor's N-terminal domain with an apolar, but not a polar, C-terminal tail sequence (Silber et al., 1992). Here I have shown that Tsp also selectively cleaves variants of Arc repressor depending upon their C-terminal sequences. Arc variants with relatively apolar C-terminal residues are good substrates for Tsp, whereas those with polar C-terminal residues are poor substrates. The studies reported here also show that a free α-carboxyl group at the C-terminus of an appropriate substrate is

important in determining whether it will be cleaved by Tsp. Peptides BAS9 and BAS10, which have apolar C-terminal sequences and free α-carboxyl groups, are cleaved readily by Tsp, whereas otherwise identical peptides with amidated C-termini are cleaved very poorly. The finding that C-terminal determinants are important in determining cleavage susceptibility in several unrelated sets of proteins and peptides strongly suggests that this is a general mechanism by which Tsp discriminates among potential substrates.

How does the chemical nature of a protein's or peptide's C-terminal region influence whether it will be a good substrate for Tsp? Because Tsp is an endoprotease with some cleavage sites far from the C-terminus, it is unlikely that the C-terminal region of a substrate directly affects the chemistry of catalysis. There are several plausable models by which the C-terminal sequence of a substrate could affect Tsp cleavage. First, C-terminal sequences that are not strongly polar or charged could act as a site to which Tsp can initially bind to the substrate, with the α-carboxyl group forming part of this binding site. Such binding might be important simply for tethering Tsp to the substrate but might also be required to activate the protease in some fashion. Activation does not occur *in trans* because neither the #105 variant of the N-terminal domain of λ repressor nor a peptide corresponding to the last nine residues of #105 stimulates cleavage of the wild-type N-terminal domain of λ repressor (data not shown). I note that the tail sequences of Arc and the N-terminal domain of λ repressor are partially disordered and accessible in the native structures of these proteins (Breg et al., 1990; Weiss et al., 1987), which would allow them to act as binding sites for Tsp even in the native proteins. Second, the tail sequences could act indirectly by promoting unfolding or other conformational changes in the substrate which result in increased

susceptibility of the substrate to Tsp cleavage. This model, however, is not consistent with experiments showing that Tsp-sensitive variants with hydrophobic tails have stabilities and conformations comparable to Tsp-insensitive proteins with hydrophilic tails in both the λ repressor and Arc proteins (Parsell and Sauer, 1989; Milla et al., 1993). Moreover, although Tsp can discriminate between variants of λ repressor or Arc differing only at their C-termini, nonspecific proteases cleave variants with hydrophobic or hydrophilic tails equally well. (Silber, 1992). The peptide pairs examined here (BAS7/BAS9 and BAS8/BAS10) have little stable structure as assayed by CD, but their spectra are identical, indicating that amidation of the terminal carboxylate in peptides BAS7 and BAS8 does not lead to significant conformational changes. Because there is no evidence that apolar C-terminal sequences alter substrate conformation or stability, I favor the idea that these sequences act as binding sites for Tsp.

Mapping of the sites of Tsp cleavage in the primary structures of several protein and peptide substrates reveals a limited number of discrete sites of proteolysis that occur throughout the substrate sequences. Tsp cleaves peptide bonds with Ala, Ser, or Val (and to a lesser extent Ile, Leu, Lys, or Arg) at the P1 position and with these same side chains plus Met, Tyr, or Trp at the P1' position. These specificity preferences appear to be sufficiently lax that most proteins and peptides would be expected to contain some cleavage sites. I note, however, that many of the substrates contain peptide bonds that fit these broad rules but are not sites at which cleavage is observed. For example, peptide BAS10 is cleaved at the Ala^{2}-Arg^{3} bond but not at the Ala^{7}-Arg^{8} or Ala^{12}-Arg^{13} bonds. These latter bonds may not be cleaved because of additional side-chain preferences at adjoining residues (P2, P2', etc.) or

because other factors are operative. For example, if Tsp contains a site for binding the C-terminus of the substrate that is distinct from the active site, then the distances between these sites may introduce a requirement for some minimal distance between the C-terminus of the substrate and a cleavage site.

I find that an Arc repressor variant which is less stable than wild type is cleaved more rapidly and a variant which is more stable than wild type is cleaved more slowly. These results are consistent with a model in which denaturation of substrate is required prior to cleavage. Indeed, the positions of the cleavage sites in Arc and the N-terminal domain of λ repressor map to many different locations in the native secondary and tertiary structure, including the hydrophobic core, surface loops, unstructured regions, and the middle of α-helices. All of these sites can not be recognized in the context of the native structures of Arc or the N-terminal domain of λ repressor, as many of the scissile peptide bonds would be inaccessible to the protease in the native proteins. If global rather than local denaturation of protein substrates is a prerequisite for cleavage, then the global unfolding rate constant would need to be greater than or equal to k_{cat} for Tsp cleavage (> 0.2 sec^{-1} for Arc at 37 °C). In fact, the unfolding rate constant for Arc is ~3 s^{-1} at 37 °C (C. Waldburger, personal communication) and thus global unfolding of Arc could precede cleavage.

Kinetic analysis of substrate cleavage by Tsp is complicated because there are often multiple cleavage sites, and the proteolytic products may themselves be substrates for further cleavage or even inhibitors. For these reasons, only apparent values of kinetic parameters can be measured. Cleavage of Arc by Tsp at 37 °C and pH 8 is relatively slow (apparent k_{cat}/K_m =

3.8×10^3 M^{-1} s^{-1}) but is similar in rate to cleavage of the same substrate by chymotrypsin (apparent $k_{cat}/K_m = 4.4 \times 10^3$ M^{-1} s^{-1}). The cleavage of peptide BAS9 ($k_{cat}/K_m = 1.1 \times 10^5$ M^{-1} s^{-1}) by Tsp is almost 30-fold more efficient. I presume that BAS9 is a better substrate for Tsp than Arc because it is largely unfolded at 37 °C (B. Schulman, personal communication), although other factors may also contribute. For comparison, cleavage of pentapeptide substrates by elastase have k_{cat}/K_m values of approximately 2×10^4 M^{-1} s^{-1} (Bauer et al., 1976; Thomson and Blout, 1973). Thus, Tsp does not appear to be any less efficient in cleaving susceptible protein and peptide substrates than well-characterized enzymes such as chymotrypsin and elastase, but it is far more specific in its choice of substrates.

A protease such as Tsp almost certainly comes in contact with a wide variety of different proteins in the periplasm of the bacterial cell and must be able to discriminate reliably between appropriate and inappropriate substrates. My studies suggest that the chemical nature and structural accessibility of the C-terminal sequence of potential substrates play a central role in determining whether such proteins are cleaved by Tsp. Proteins with polar C-terminal residues or with C-terminal sequences that are buried within the native structure should not be cleaved by Tsp. Similarly, even if a protein has an accessible and apolar C-terminus that allowed transient Tsp binding, it could probably still escape cleavage if its native structure were sufficiently stable. By using a combination of specificity determinants for substrate recognition, Tsp is apparently able to avoid degrading the wrong proteins in the cell.

MATERIALS AND METHODS

Proteins and peptides—Wild-type Tsp was purified from *E. coli* strain X90 containing the Tsp over-producing plasmid pKS6-1w (Silber et al., 1992). Cells were grown in 6-12 L of LB broth at 37 °C with gentle shaking to an absorbance of 0.8 at 600 nm and induced by addition of 1 mM IPTG. The culture was grown for an additional 3 h, harvested by centrifugation and resuspended in 10 volumes of a sphereoplasting buffer containing 100 mM Tris-HCl (pH 8), 5 mM EDTA, 500 mM sucrose. After resuspension the cells were collected again by centrifugation, resuspended in 10 volumes of cold water to burst the outer membrane and release the periplasmic proteins, and the cytoplasmic and membrane portions were removed by centrifugation. The soluble periplasmic fraction was chromatographed on a 20 mL DEAE Sephacel column equilibrated in 10 mM Tris-HCl (pH 8), 100 mM KCl. The flow-through fraction was collected, concentrated to 0.5 mL by ultrafiltration, and chromatographed on a 24 mL Superose 6 column in 10 mM Tris-HCl (pH 8), 100 mM KCl using a Pharmacia FPLC apparatus. Fractions containing Tsp at greater than 95% purity, as assayed by SDS polyacrylamide gel electrophoresis, were pooled. Tsp was stored in column buffer at 0 °C with no loss of activity over a period of several months.

A gene encoding a variant of Tsp ($TspH_6$) containing six additional histidine residues at the C-terminus was constructed by subcloning the EcoRI-HindIII fragment from pKS6-1w (which contains all of the *tsp* gene with the exception of 66 bases at the 3′ end of the gene) into EcoRI and XbaI cut pBluescript II KS(+) together with an HindIII-XbaI oligonucleotide cassette designed to restore the wild type bases at the 3′ end of the gene followed by the

His_6 tag and normal termination codon. The structure of the resulting plasmid (pKK101) was confirmed by restriction mapping and DNA sequencing. The $TspH_6$ protein was purified from *E. coli* strain KS1000, which contains a deletion of the chromosomal *tsp* gene (Silber and Sauer, 1994), transformed with pKK101. Cells were grown, induced, and harvested as described above. The cell pellet was resuspended in 5 volumes of 100 mM Tris-HCl (pH 8), 200 mM KCl, 1 mM EDTA, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 1.4 mM 2-mercaptoethanol, lysed by sonication, and the lysate was cleared by centrifugation. The supernatant was applied to a 20 mL Ni-NTA column (Qiagen) equilibrated in 100 mM NaH_2PO_4, 10 mM Tris-HCl (pH 8). Bound proteins were eluted with a 500 mL gradient formed by linear mixing of the column buffer at pH 8 and column buffer titrated to pH 4.5 with HCl. Fractions containing the $TspH_6$ protein, which eluted between pH 5.3 and pH 4.8, were pooled, concentrated to 0.5 mL by ultrafiltration, and chromatographed on a 24 mL Superdex 75 column equilibrated in 10 mM Tris-HCl (pH 8), 100 mM KCl. Fractions containing $TspH_6$ at greater than 95% purity were pooled.

The Arc-st7 protein differs from wild-type Arc in having 11 additional C-terminal residues (HHHHHHGRIGA). The *arc-st7* gene was constructed by synthesis of an appropriate HindIII-ClaI cassette which was inserted between the HindIII and ClaI sites of plasmid pSA600 (Milla et al., 1993). Arc repressor, the Arc-st5, Arc-st6, and Arc-st7 variants, the N-terminal domain of λ repressor (residues 1-102), and the #105 variant of the N-terminal domain were purified by ion exchange chromatography, affinity chromatography, and gel filtration as described (Milla et al., 1993; Lim and Sauer, 1991) . The NK29 and PL8 mutants of Arc were gifts of Marcos Milla. Peptides BAS7-BAS10

were the generous gift of Brenda Schulman. The expected molecular weights of all peptides were confirmed by mass spectrometry.

Characterization of Tsp—The circular dichroism (CD) spectra of Tsp and $TspH_6$ (50 μg/ml) were recorded at 5 °C in CD buffer (50 mM phosphate (pH 7.0), 100 mM KCl) using an Aviv Model 60DS Circular Dichroism Spectrometer. The α-helical content was estimated using the program Varselec (Manavalan and Johnson, 1987). Sedimentation equilibrium centrifugation was performed in 10 mM Tris-HCl (pH 8), 100 mM KCl using a Beckman XL-A analytical ultracentrifuge. Molecular weight values were determined for 0.5 μM, 1.7 μM, 3.4 μM, and 5 μM $TspH_6$ at rotor speeds of 15,000 rpm and 13,000 rpm by nonlinear least squares fitting of the data to a single species molecular weight function ($C(r) = C(a)\exp[\omega^2M(1-\nu\rho)(r^2-a^2)/2RT]$) (Laue et al., 1992) using the program Nonlin (Brenstein, 1989; Johnson and Frazier, 1985).

Assays for proteolysis—For assays of Tsp or $TspH_6$ cleavage of protein substrates by gel electrophoresis, 0.5 μg of protease and 2 μg of substrate in 30 μL of reaction buffer (10 mM Tris-HCl (pH 8.0), 200 mM KCl) were incubated at 37 °C for 4 h, and the reaction products were electrophoresed on 15% polyacrylamide SDS Tris-tricine gels (Schagger and von Jagow, 1987). Cleavage of peptides by $TspH_6$ was monitored by HPLC on a reverse-phase column (gradient from 0 to 100% acetonitrile in 0.1% trifluoroacetic acid (TFA)) following incubation of 500 μg peptide with 3.4 μg $TspH_6$ in 1 mL of reaction buffer for 5 h at 37 °C. The extent of cleavage was quantified by the loss of area in the peak corresponding to the intact peptide and by the appearance of additional peaks in the chromatograph. Reverse phase column

chromatography was also used to monitor the time course of Tsp cleavage of protein substrates such as Arc and the #105 variant of the N-terminal domain of λ repressor. For these assays, 10 μg of substrate was mixed with 2 μg $TspH_6$ in 300 μl of reaction buffer at 37 °C, and 50 μl aliquots were removed from the reaction at 20 min intervals, quenched with 0.1% TFA, and analyzed by reverse-phase chromatography as described above.

Kinetic constants for cleavage of Arc repressor and peptide BAS9 were determined by mixing 50 nM $TspH_6$ with 2.5-220 μM Arc repressor or 1.6-150 μM BAS9 in reaction buffer at 37 °C and monitoring changes in CD ellipticity at 222 nm. The rate of cleavage in mdeg/sec was determined from a linear fit of the ellipticity versus time data and converted to units of μM substrate cleaved/sec/μM enzyme assuming that cleaved substrate has no ellipticity at 222 nm. K_M and k_{cat} values for Arc and BAS9 were determined by fitting the rate versus substrate concentration data to the hyperbolic function $v = k_{cat}[E_t]([S]/(K_M+[S]))$ using the program Nonlin. The relative rates of cleavage of Arc repressor variants were determined by mixing 45 nM Tsp with Arc, Arc-PL8, or Arc-NK29 at concentrations of 8 μM, and monitoring cleavage using the CD assay described above.

Mass spectrometry—The peptides produced by Tsp cleavage of λ #105, Arc, Arc-st7, and the peptide BAS10 were analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Hillenkamp et al., 1991). Typically 50-150 nmoles of substrate were digested swith 100 pmoles of Tsp or $TspH_6$ for 1-6 h. In some cases the digestion products were partially purified by reverse phase chromatography or nickel-NTA affinity chromatography prior to mass spectrometric analysis. A sample

of each digest was diluted with a 10 mM solution of 2,5-dihydroxybenzoic acid or 3,5-dimethoxy-4-hydroxycinnamic acid to a concentration of approximately 10 μM. 1 μL of each matrix/analyte solution was deposited and dried on the sample probe of a VESTEC 2000 linear time-of-flight mass spectrometer (Perseptive Biosystems). The samples were irradiated at 337 nm with a nitrogen laser (Model VSL337ND, Laser Science Inc.). Measured *m/z* values for the $[M+H]^+$ ions detected were matched with segments of the protein using a computer program (COMPOST) which has been previously described (Papayannopoulos and Biemann, 1991). In cases where more than one peptide from a given substrate could correspond to the measured mass, sequence assignments were made based on the consistency of the peptide termini with those from other peptide digestion products. In this regard, peptides formed through a single cleavage of the intact protein were favored over those requiring two cleavages. In several cases, sequence assignments were confirmed independently from the collision-induced dissociation mass spectra of the peptide products (see Biemann, 1990), or by N-terminal sequencing using the Edman degradation.

ACKNOWLEDGMENTS: I thank Bronwen Brown, Brenda Schulman, and Marcos Milla for gifts of proteins and peptides. I also thank Ray Larsen and Kathleen Postle for communicating unpublished data. Supported by NIH grants AI-15706 and RR-00317.

REFERENCES

Anbudurai PR, Mor TS, Ohad I, Shestakov SV, and Pakrasi HB. 1994. The CtpA gene encodes the C-terminal processing protease for the D1 protein of the photosystem II reaction center complex. *Proc Natl Acad Sci USA 91*: 8082-8086.

Bauer CA, Thompson RC, and Blout ER. 1976. The active centers of *Streptomyces griseus* protease 3, alpha-chymotrypsin, and elastase: enzyme-substrate interactions close to the scissile bond. *Biochemistry 15*: 1296-1299.

Beynon RJ, and Bond JS, ed. 1989. *Proteolytic enzymes: a practical approach.* Oxford IRL Press.

Biemann K. 1990. Srquencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods in Enzymology,* J.A. McCloskey, Editor. San Diego: Academic Press: 455-479.

Bowie JU, and Sauer RT. 1989. Identification of C-terminal extensions that protect proteins from intracellular proteolysis. *J Biol Chem 264*: 7596-7602.

Breg JN, van Opheusden JHJ, Burgering MJM, Boelens R, and Kaptein R. 1990. Structure of Arc repressor in solution: evidence for a family of beta-sheet DNA-binding proteins. *Nature 346*: 586-589.

Brenstein RJ. 1989. *NONLIN for Macintosh.* Carbondale, IL.: Robelko Software.

Goldberg AL. 1992. The mechanisms and functions of ATP-dependent proteases in bacterial and animal cells. *Eur J Biochem 203*: 9-23.

Gottesman S, and Maurizi MR. 1992. Regulation by proteolysis: energy-dependent proteases and their targets. *Microbiol Rev 56*: 592-621.

Hara H, Yamamoto Y, Higashitani A, Suzuki H, and Nishimura Y. 1991. Cloning, Mapping, and Characterization of the *Escherichia coli prc* gene, which is involved in C-terminal processing of penicillin-binding protein 3. *J Bacteriol 173*: 4799-4813.

Hershko A, and Ciechanover A. 1992. The ubiquitin system for protein degradation. *Annu Rev Biochem 61*: 761-807.

Hillenkamp F, Karas M, Beavis RC, and Chait BT. 1991. Matrix-assisted desorption/ionization mass spectrometry of biopolymers. *Anal Chem 63*: 1193A-1203A.

Johnson ML, and Frazier S. 1985. Nonlinear least-squares analysis. *Methods Enzymol 117*: 301-342.

Laue TM, Shah BD, Ridgeway TM, and Pelletier SL. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. *Analytical Ultracentrifugation in Biochemistry and Polymer Science,* Harding S, Rowe A, and Horton J, eds. Cambridge: Royal Society of Chemistry. 90-125.

Lim WA, and Sauer RT. 1991. The role of internal packing interactions in determining the structure and stability of a protein. *J Mol Biol 219*: 359-376.

Manavalan P, and Johnson WC, Jr. 1987. CD secondary structure analysis using variable selection. *Anal Biochem 167*: 76-85.

Milla ME, and Sauer RT. 1994. P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry 33*: 1125-1133.

Milla ME, Brown BM, and Sauer RT. 1993. P22 Arc repressor: enhanced expression of unstable mutants by addition of polar C-terminal sequences.*Protein Science 2*: 2198-2205.

Nagasawa H, Sakagami Y, Suzuki A, Suzuki H, Hara H, and Hirota Y. 1989. Determination of the cleavage site involved in C-terminal processing of penicillin-binding protein 3 of *Escherichia coli. J Bacteriol 171*: 5890-5893.

Pabo CO, and Lewis M. 1982. The operator-binding domain of lambda repressor: structure and DNA recognition. *Nature 298*: 443-447.

Park JH, Lee YS, Chung CH, and Goldberg AL. 1988. purification and characterization of protease Re, a cytoplasmic endoprotease in *Escherichia coli. J Bacteriol 170*: 921-926.

Parsell DA, Silber KR, and Sauer RT. 1990. Carboxy-terminal determinants of intracellular protein degradation. *Genes Dev 4*: 277-286.

Parsell DA, and Sauer RT. 1989. The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli. J Biol Chem 264*: 7590-7595.

Papayannopoulos IA, and Biemann K. 1991. A computer program (COMPOST) for predicting mass spectrometric information from known amino acid sequences. *J Am Soc Mass Spectrom 2*: 174-177.

Roseman JE, and Levine RL. 1987. Purification of a protease from *Escherichia coli* with specificity for oxidized glutamine synthetase. *J Biol Chem 262*: 2101–2110.

Schagger H, and von Jagow G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem 166*: 368-379.

Silber KR, and Sauer RT. 1994. Deletion of the *prc* (*tsp*) gene provides evidence for additional tail-specific proteolytic activity in *Escherichia coli* K-12. *Mol Gen Genet 242*: 237-240.

Silber KR, Keiler KC, and Sauer RT. 1992. Tsp: A tail-specific protease that selectively degrades proteins with nonpolar C termini. *Proc Natl Acad Sci USA 89*: 295-299.

Silber KR. 1992. The role of carboxyl termini in selective protein degradation. Doctoral Thesis. Massachusetts Institute of Technology.

Thomson RC, and Blout ER. 1973. Dependence of the kinetic parameters for the elastase-catalyzed amide hydrolysis on the length of peptide substrates. *Biochemistry 12*: 57-65.

Tobias JW, Shrader TE, Rocap G, and Varshvsky A. 1991. The N-end rule in bacteria. *Science 254*: 1374-1377.

Weiss MA, Karplus M, and Sauer RT. 1987. 1H NMR aromatic spectrum of the operator binding domain of the lambda repressor: resonance alignment with application to structure and dynamics. *Biochemistry 26*: 890–897.

A

Protein	Protein background:tail sequence	% cleavage
λ N domain	λ 1-97:RSEYE	0
λ #105 N domain	λ 1-97:WVAAA	95
Arc repressor	Arc 1-48:GRIGA	90
Arc-st5	Arc 1-48:GRIGAKNQHE	0
Arc-st6	Arc 1-48:GRIGAHHHHHH	20
Arc-st7	Arc 1-48:GRIGAHHHHHHGRIGA	90

B

Peptide	sequence	% cleavage
BAS9	Ac WARAAARAAARBAAB*COOH*	50
BAS7	Ac WARAAARAAARBAAB*CONH2*	<5
BAS10	Ac WARAAARAAARBGGB*COOH*	45
BAS8	Ac WARAAARAAARBGGB*CONH2*	<5

Table 1. Cleavage of protein and peptide substrates by Tsp. A: Protein substrates were incubated with Tsp for 8 h and cleavage was quantified by laser densitometry of SDS gels. B: Sequences of peptides tested as substrates of Tsp. B represents the non-natural amino acid amino-butyric acid. Peptides with the N terminus blocked by acetylation (*Ac*) or the C terminus blocked by amidation ($CONH_2$) are indicated. Peptides were incubated with Tsp for 5 h, chromatographed on an HPLC reverse phase column, and the amount of cleavage was quantified by integrating the peak corresponding to intact peptide.

Substrate	Measured *m/z* value[1]	Assigned Fragment	Assigned Cleavage[2]	Difference from calculated *m/z* value[3]
λ #105	11177.15	1-102	S...A	-2.77
	11032.40	1-100	S...A†	-5.36
	10967.11	1-99	S...V†	+1.42
	10676.52	1-97	S...L↓W	-4.72
	10563.14	1-96	S...S↓L	-5.04
	10120.10	1-92	S...S↓M	-4.55
	10003.07	1-91	S...V↓S	-4.51
	7132.32	1-66	S...A↓K	+8.12
	6762.50	1-62	S...A↓A	+6.78
	6413.92	1-59	S...A↓Y	+6.56
	6113.65	1-56	S...A↓L	+4.63
	1577.36	67-80	A↓K...I↓A	+0.49
	1489.77	80-91	S↓I...V↓S	+2.06
	1205.00	82-90	A↓R...A↓V	+0.65
	1161.54	92-102	V↓S...A[4]	+0.17
	1076.18	93-102	S↓M...A	+1.89
	1020.90	92-100	V↓S...A†	+0.69
	948.29	92-99	V↓S...V†	+0.16
	932.52	93-100	S↓M...A†	+0.39
	861.19	93-99	S↓M...V†	+0.14
	734.06	60-66	A↓Y...A↓K[4]	-1.79*
	629.45	97-102	S↓L...A	-1.31
	487.78	97-100	S↓L...A†	-0.83
Arc	6229.22	1-53	M...A[5]	+1.95
	5566.04	7-53	K↓M...A	+0.63
	3570.47	23-53	V↓R...A[5]	+0.44
	2678.29	1-22	M...V↓R[5]	+0.99
	2137.44	23-40	V↓R...R↓V	+1.11
Arc-st7	7263.96	1-61	M...R↓I	+1.10
	4844.48	23-64	V↓R...A	+1.09
	1753.94	26-40	V↓A...R↓V	-0.41
BAS10	1340.61	3-17	A↓R...B	+0.08

Table 2. Products of Tsp cleavage of substrates determined by MALDI-TOF mass spectrometry. [1]Measured *m/z* values represent average (i.e. polyisotopic) values for the $[M+H]^+$ ion. [2]Arrows indicate cleavage between the residues shown. [3]The difference between measured and calculated *m/z* values for the assigned sequence is ≤ 0.18% (except for sequence marked by * which was confirmed independently; see footnote 4). [4]Sequence confirmed by an independent MS/MS experiment (Biemann, 1990). [5]N-terminal sequence

confirmed by Edman degradation. †In addition to 1-102, the starting material contained small amounts of 1-100 and 1-99. Thus, fragments which end at position 99 or 100 are assumed to result from these truncated substrates and not from cleavage by Tsp.

Protease	Substrate	apparent kinetic parameters
Tsp	Arc	K_m = 50 (± 36) μM k_{cat} = 0.19 (± 0.05) s^{-1} k_{cat}/K_m = 3.8 (± 2.9) x 10^3 M^{-1} s^{-1}
Chy	Arc	k_{cat}/K_m = 4.4 (± 0.5) x 10^3 M^{-1} s^{-1}
Tsp	BAS9	K_m = 35 (± 13) μM k_{cat} = 3.7 (± 0.5) s^{-1} k_{cat}/K_m = 1.1 (± 0.4) x 10^5 M^{-1} s^{-1}

Table 3. Apparent kinetic parameters for the cleavage of Arc and the peptide BAS9 by Tsp and chymotrypsin (Chy). K_m and k_{cat} values were determined by fitting the data in Figure 4 to the hyperbolic function $v = k_{cat}$ [E_t] ([S] / (K_m + [S])).

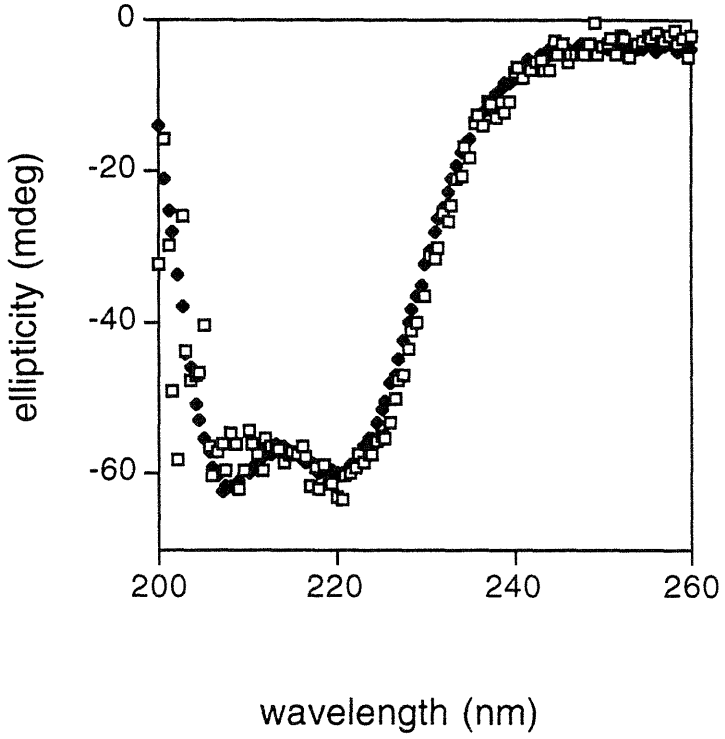


Fig. 1. CD spectra of the Tsp (filled diamonds) and $TspH_6$ (open squares) enzymes.

Fig. 2. Degradation of λ repressor and Arc substrates by Tsp and $TspH_6$. Reaction mixtures were incubated at 37 °C for 4 h, and cleavage was assayed by gel electrophoresis. At shorter incubation times the same cleavage products were observed, but at reduced levels. The arrow indicates the position of the Tsp and $TspH_6$ protein bands. (following page)

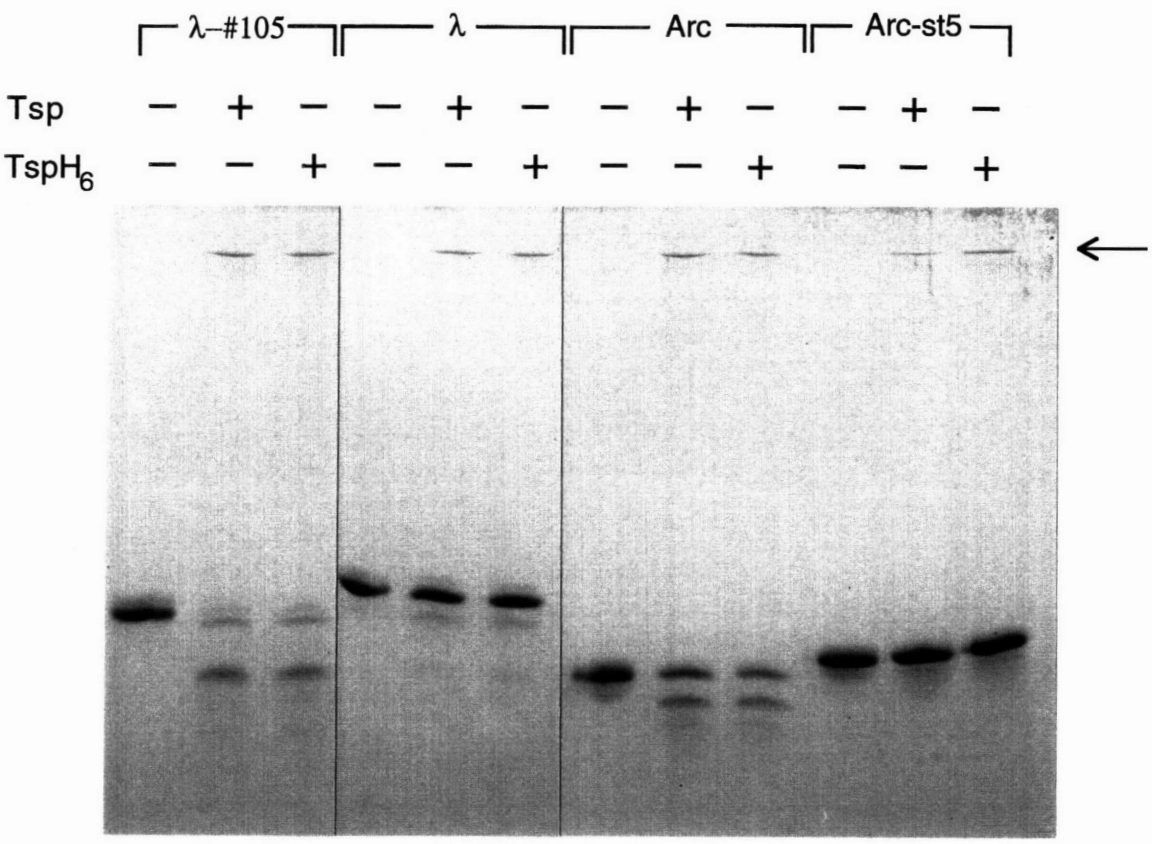

λ–#105
λ
Arc
Arc-st5
Tsp
TspH6

Fig. 3. Cleavage of peptides BAS10 (free α-carboxyl group) and BAS8 (α-carboxyamide group) by $TspH_6$. Reverse phase HPLC traces (A_{226} versus time) are shown after 0 and 2 h of incubation with $TspH_6$ (pH 8, 37 °C). (following page)

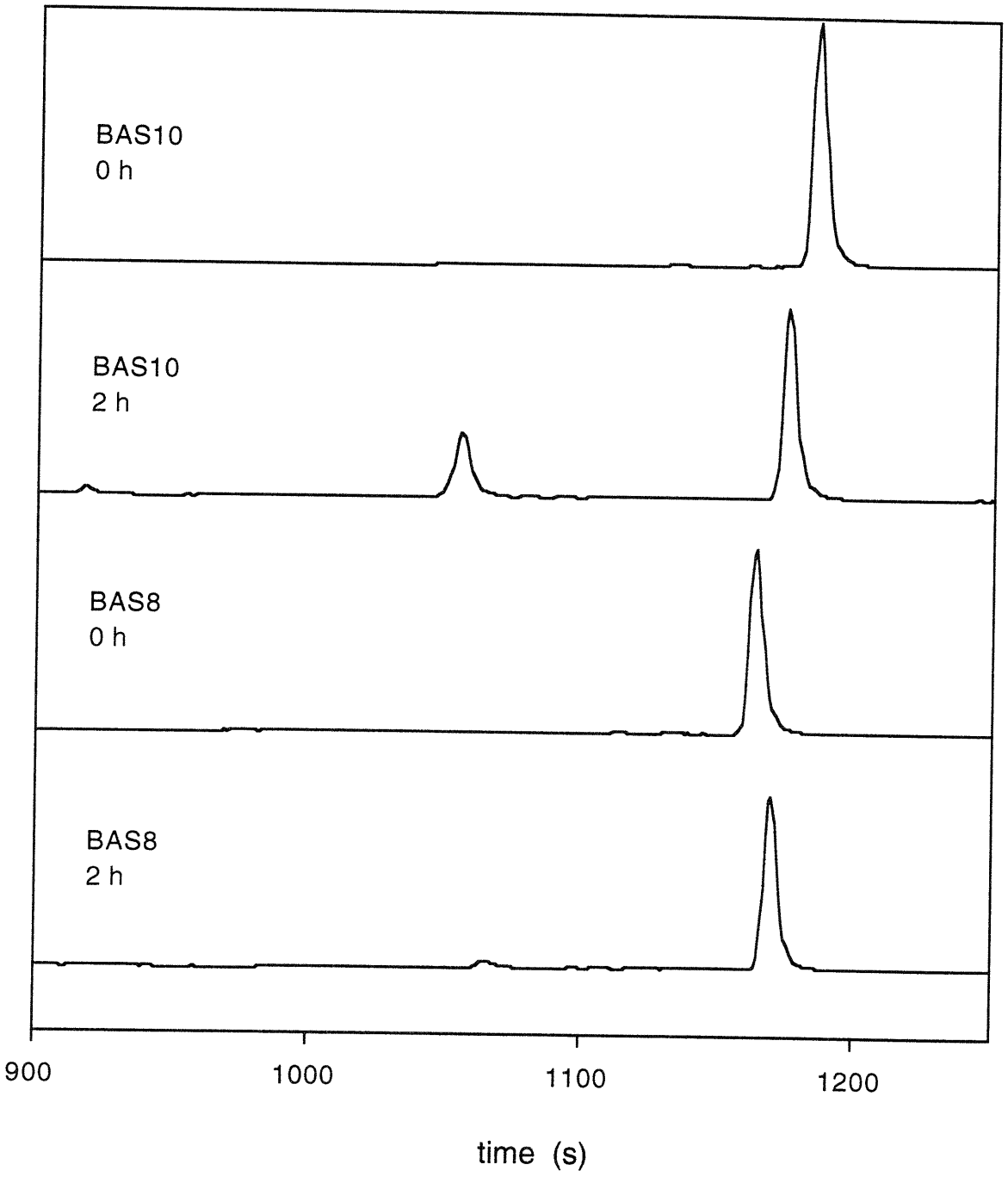
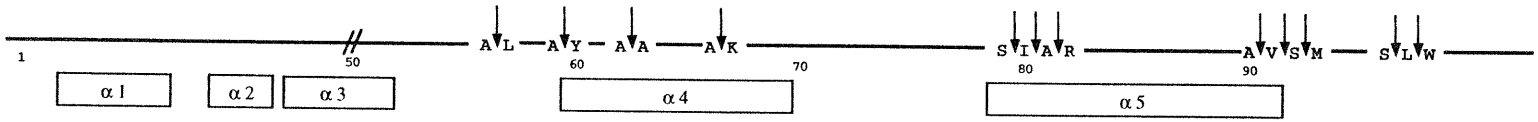
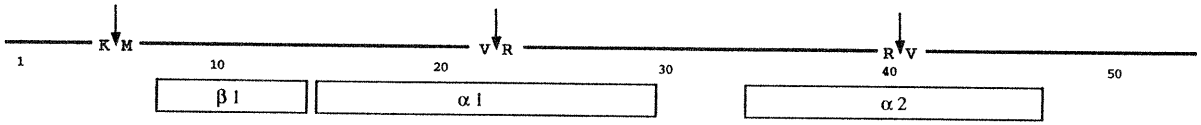

BAS10
0 h
BAS10
2 h
BAS8
0 h
BAS8
2 h
900
1000
1100
1200
time (s)

Fig. 4. Sites of Tsp cleavage are indicated by arrows in four protein and peptide substrates. Elements of secondary structure in λ repressor and Arc (Breg et al., 1990; Pabo and Lewis, 1982) are taken from the structures of the wild-type proteins. α-helices and β-strands are labeled. The first fifty residues of λ #105 are not to scale. (following page)

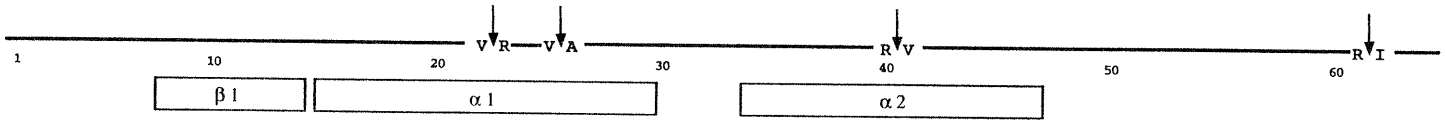
λ #105



Arc



Arc-st7



BAS10

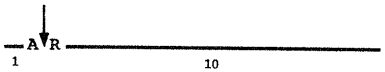
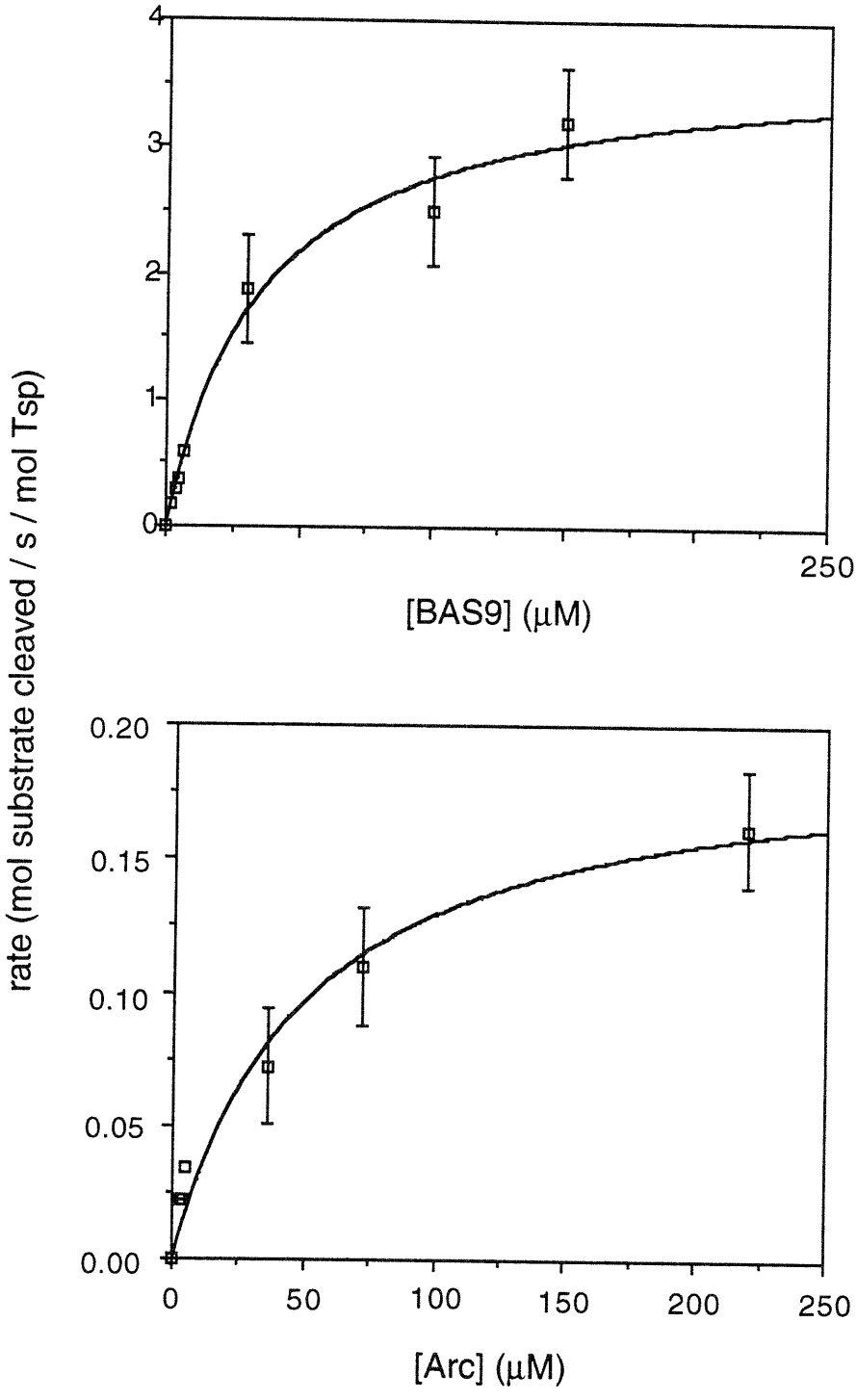


Fig. 5. Rate versus substrate concentration plots for Tsp cleavage of peptide BAS9 and Arc. Cleavage was assayed by changes in CD ellipticity. Error bars indicate the uncertainty in determining the rate in each assay. Points below 25 μM have errors of similar magnitude which have not been shown. (following page)


rate (mol substrate cleaved / s / mol Tsp)
4
3
2
1
0
250
[BAS9] (µM)
0.20
0.15
0.10
0.05
0.00
0
50
100
150
200
250
[Arc] (µM)

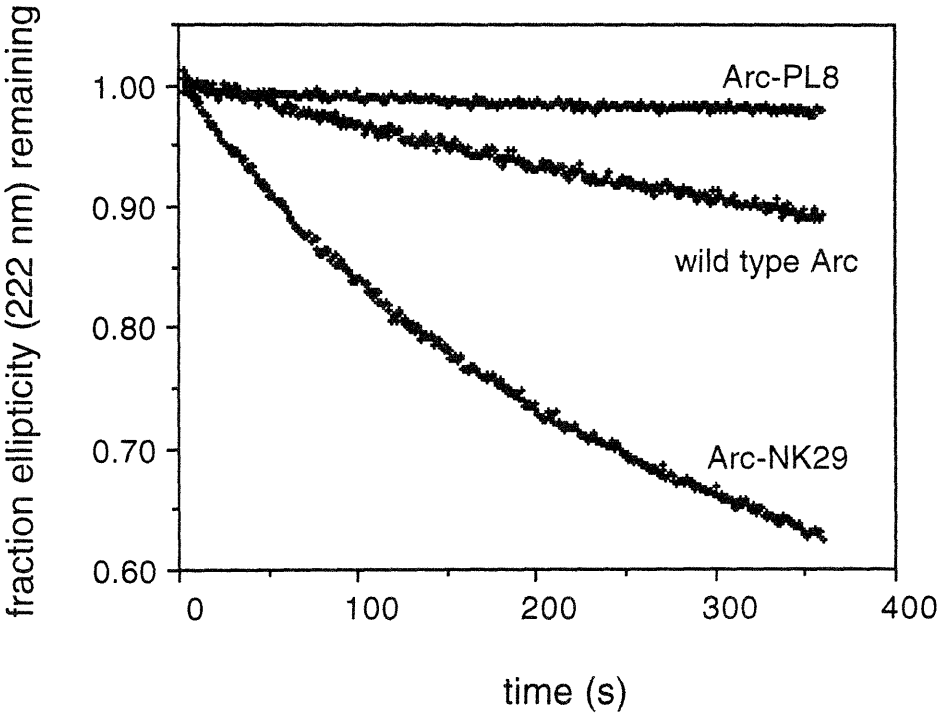


Fig. 6. Tsp cleavage of wild-type Arc and stability variants assayed by loss of CD ellipticity at 222 nm. The melting temperatures of PL8, Arc, and NK29 are approximately 70 °C, 50 °C, and 37 °C, respectively.

CHAPTER 3

Identification of Active Site Residues of the Tsp Protease

INTRODUCTION

Tsp is a periplasmic protease of *Escherichia coli* (1), which has been implicated in the processing of penicillin binding protein 3 (2) and TonB (Larsen, R., and Postle, K. personal communication) and appears to play some role in fatty acid transport mediated by the fadL protein (3). Deletion of the *tsp* gene (also known as *prc*) causes a temperature-dependent growth defect under conditions of osmotic stress (2). Biochemical studies *in vitro* suggest that Tsp selects substrate proteins based on the identity of side chains and functional groups at their C termini. For example, amidation of the α-carboxyl group of a good peptide substrate prevents cleavage by Tsp (4). Moreover, Tsp cleaves variants of Arc repressor and the N-terminal domain of λ-repressor that have apolar C-terminal residues but does not cleave variants of comparable stability that have polar C-terminal sequences (1, 4).

Tsp does not share sequence homology with well-characterized proteases or protease active-site sequences and cannot be classified as a serine, cysteine, aspartic, or metallo protease based on inhibitor studies (1). A portion of the Tsp sequence is homologous to a sequence motif repeated four times in the interphotoreceptor retinoid binding protein (IRBP) of mammals (1). It has been proposed that Tsp binds the apolar C termini of its substrates in the same fashion as the IRBP repeats bind retinoids. Recently, four bacterial genes encoding sequences with homology to Tsp have been determined (see Fig. 1). I assume that many of these genes encode proteases, although this is supported by genetic or biochemical evidence only in the case of CtpA (5).

In this chapter, I construct Tsp mutants by site-directed mutagenesis, test the mutant strains for growth defects *in vivo*, and assay the mutant proteins for protease activity *in vitro*. Inactive mutant proteins are assayed for gross structural perturbations using circular dichroism spectrometry. Finally, the binding of substrates to the inactive but structurally intact proteins is assayed. Using these approaches, I identify three Tsp residues (Ser430, Asp441, and Lys455) which have properties expected for active-site residues. The behavior of proteins altered at these positions suggests that Tsp may use a serine-lysine mechanism similar to those proposed for LexA, the type I signal peptidases, and the class A β-lactamases (6-9).

EXPERIMENTAL PROCEDURES

Plasmids and proteins —Wild-type Tsp with a six-histidine tag was purified from *E. coli* strain KS1000 (10) transformed with the over-producing plasmid pKK101 (4). In this plasmid, transcription of the *tsp* gene is controlled by an IPTG-inducible promoter. Cells were grown in 1 L of LB broth at 37 °C with gentle shaking to an absorbance of 0.8 at 600 nm and were induced by addition of 1 mM IPTG. The culture was grown for an additional 3 h, harvested by centrifugation and resuspended in a pH 8.0 buffer containing 6 M guanidine hydrochloride, 10 mM Tris-HCl, 100 mM sodium phosphate, and 20 mM imidazole. The cells were stirred for 1 h at 4 °C, insoluble material was removed by centrifugation at 12,000 rpm for 30 min, and the supernatant was loaded onto a 3 ml Ni-NTA column (Qiagen). The column was washed with 25 volumes of loading buffer, and bound protein was eluted with loading buffer plus 250 mM imidazole. This fraction was dialyzed against 10 mM Tris-HCl (pH 8.0), 100 mM KCl, and chromatographed

over a 2 ml Q-Sepharose column (Pharmacia) equilibrated in the same buffer. The flow-through fraction contained Tsp at greater than 95% purity.

Site-directed mutations were constructed in the *tsp* gene of pKK101 using the PCR mutagenesis technique of Higuchi (11), and confirmed by DNA sequencing. Variants were purified in the same manner as wild type, except that it was sometimes necessary to elute more acidic variants of Tsp from the Q-Sepharose column using 10 mM Tris-HCl (pH 8.0), 300 mM KCl. Wild-type Tsp and the G375A, G376A, E433A, T452A, and K455A variants were also purified without denaturation, to ensure that the properties of the purified proteins did not depend on the method of purification. In these cases, cells were grown, induced, and harvested as described above, and resuspended in 10 volumes of a sphereoplasting buffer containing 100 mM Tris-HCl (pH 8.0), 5 mM EDTA, 500 mM sucrose. After resuspension, the cells were collected by centrifugation, resuspended in 10 volumes of cold water to burst the outer membrane, and the cytoplasmic and membrane portions were removed by centrifugation. The soluble periplasmic fraction was brought to a final concentration of 10 mM Tris-HCl (pH 8.0), 100 mM sodium phosphate, 20 mM imidazole by dilution and addition of concentrated buffer, and chromatographed over a 2 ml Ni-NTA column. Tsp or its variants were eluted with loading buffer plus 250 mM imidazole, dialyzed against 10 mM Tris-HCl (pH 8.0), 100 mM KCl, and chromatographed over a Q-Sepharose column as described above. In each case where Tsp or its variants were purified under native conditions, their proteolytic activities *in vitro* and circular dichroism (CD) spectra were indistinguishable from those of the same proteins purfied under denaturing conditions.

The #105 variant of the N-terminal domain of λ-repressor, a good protein substrate for Tsp (1), was purified by ion-exchange chromatography, affinity chromatography, and gel filtration as described (12).

Activity and structural assays—To assay proteolytic cleavage activity, 0.5 μg of Tsp and 2 μg of the λ-repressor substrate were incubated at 37 °C for 20 h in 30 μl of cleavage buffer (10 mM Tris-HCl (pH 8.0) plus 20 mM NaCl). The reaction products were then electrophoresed on 15% polyacrylamide SDS Tris-tricine gels (13) and visualized by staining with Coomassie Blue. To quantify differences in cleasvage rates, 10 μM Arc repressor was incubated with 250 nM Tsp or a Tsp variant in cleavage buffer at 37 °C in an Aviv model 60DS circular dichroism spectrometer, and the change in ellipticity at 222 nm was monitored over time (4). The activity of Tsp variants *in vivo* was assayed by streaking *E. coli* strain KS1000 transformed with wild-type or mutant versions of pKK101 on low-salt agar plates containing 0.5% tryptone, 0.25% yeast extract, and 150 μg/ml ampicillin. Duplicate plates were incubated for two days at the permissive (37 °C) or the restrictive temperature (45 °C).

The CD spectra of Tsp or variants at concentrations of 50 μg/ml were recorded at 5 °C (to prevent auto-degradation), in buffer containing 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, using an Aviv Model 60DS Circular Dichroism Spectrometer.

Sequence Alignments—Sequences with homology to Tsp in GenBank release 89 were identified using the program BLAST (14), and multiple sequence alignments were performed using the PileUp program in the GCG package (15).

Substrate binding assay—The B chain of bovine insulin labeled with FITC at its amino terminus was purified from material purchased from Sigma using reverse phase chromatography under reducing conditions. Incubation of the FITC-labelled insulin B chain with Tsp in cleavage buffer (2 h, 37 °C, E:S::1:50) at 37 °C resulted in approximately 40% cleavage of the labelled protein as assayed by reverse-phase HPLC. Addition of 2 μg FITC-insulin to the λ-repressor cleavage reaction described above resulted in 40-60% inhibition of λ-repressor cleavage. For binding assays, equal amounts (667 nM) of Tsp and the FITC-labelled insulin B chain were incubated in cleavage buffer at 5 °C (at this temperature less than 1% of FITC-insulin was cleaved during these assays, as determined by reverse-phase HPLC analysis), allowed to equilibrate for 1 h, and the fluorescence emission spectrum was taken following excitation at 496 nm. In this assay, binding results in a 25-50% decrease in the intensity of the fluorescein fluorescence and an approximate 1.5 nm red shift in the fluorescence spectrum. Incubation of Tsp with fluorescein-iodoactamide results in only 5-10% changes in fluorescence intensity and no red shift, suggesting that binding to FITC-labelled insulin B chain is not mediated solely by the the fluorescein dye.

Chemical Inhibition Assays—The ability of diisopropylfluoro-phosphate (DFP) to inhibit Tsp was studied by incubating 1.5 μM Tsp with 1-50 mM DFP in cleavage buffer for 1 h, adding 2 μg of the λ-repressor substrate, and incubating for an additional 4 h. Degradation was assayed by gel electrophoresis, and the degree of inhibition was determined by comparing reactions incubated with DFP to control reactions containing all components except DFP. As a positive control, DFP-inhibition of bovine trypsin was assayed under identical conditions and found to be complete at each

concentration of DFP tested. Labelling with tritiated DFP (NEN-DuPont) was performed by incubating 1.5 μM Tsp or bovine trypsin with 12.5 μM tritiated DFP for 1 h, adding a second aliquot of tritiated DFP to a final concentration of 25 μM and incubating for an additional 1 h. Reaction products were separated on an SDS-polyacrylamide gel, stained with Coomassie Blue, soaked in Autofluor (National Diagnostics) for 1 h, dried, and exposed to film at -80 °C for 5 days. To quantify the extent of labelling, the stained bands were excised from the gel, suspended in scintillation fluid, and counted in a liquid scintillation counter. Inhibition of the S430C variant of Tsp by alkylating reagents was assayed by incubating 0.5 μg protein in cleavage buffer with 10 mM N-ethylmaleimide (NEM) or iodoacetamide at room temperature for 20 min, adding 2 μg of the λ-repressor substrate, incubating the mixture for 20 h at 37 °C, and assaying for substrate cleavage by gel electrophoresis.

RESULTS

Candidates for functionally important positions: The mature Tsp protein contains 660 amino acids (1). Positions for mutagenesis were chosen in several ways. First, each of the five histidine residues in Tsp was mutated to alanine because histidine is a common component of known proteolytic mechanisms (16). Second, Arg-371 and Arg-444 were mutated to alanine because these positions are conserved between Tsp and each of the known IRBP repeats, suggesting that these basic residues may play a role in binding substrate. Third, Ser-372, Gly-375, and Gly-376 were mutated because these residues (as well as Arg-371) fall in a small region which shows weak homology with an active-site region of chymotrypsin (RSNGGGAL in Tsp; MGDSGGPL in chymotrypsin). Finally, the alignment between Tsp and the

presumed protease sequences shown in Fig. 1 was used as a guide, since these proteins share 53% identity in the two regions shown but have little significant homology in other regions. Each conserved residue with a hydroxyl, carboxyl, or amino group within the homology region was mutated to alanine since these residues could potentially participate in proteolytic chemistry. In all, 20 positions in Tsp were selected, alanine-substitution mutations were constructed, and the mutant proteins were purified.

Activity of Tsp variants: Thirteen alanine-substituted Tsp variants showed significant proteolytic activity *in vitro* and seven had no detectable activity (Table 1, Fig. 2A). The inactive proteins include the G375A, G376A, S430A, E433A, D441A, T452A, and K455A variants. Activity of the mutants *in vivo* was assayed by rescue of the temperature-sensitive growth defect of a *tsp*-deletion strain on low salt medium. As shown in Table 1, growth at the restrictive temperature is rescued by plasmids encoding Tsp variants that are proteolytically active *in vitro* but is not rescued by plasmids encoding mutants that are inactive *in vitro*. This correlation suggests strongly that the conditional growth phenotype of a *tsp*-deletion strain is related to proteolytic activity and not to some other function of Tsp.

Structure and substrate binding: In principle, a Tsp mutant could be inactive because it folds improperly, because it cannot bind substrate, or because it cannot catalyze peptide-bond hydrolysis. To address the folding issue, the CD spectra of all inactive variants and several active variants were determined. Three inactive mutants (S430A, D441A, and K455A) had CD spectra that were similar to wild type (Fig. 3A). Although this assay cannot rule out small conformational changes that could cause inactivity, Ser-430,

Asp-441, and Lys-455 are the best candidates for residues that are involved directly in substrate binding or catalytic activity. The remaining four inactive mutants (G375A, G376A, E433A, and T452A) had spectra which differed markedly from that of wild-type Tsp (Fig. 3B), suggesting significant structural perturbations.

As an assay for substrate binding, wild-type Tsp and the S430A, D441A, and K455A mutants were incubated with FITC-labelled insulin B chain and the fluorescence spectra were recorded. As shown in Fig. 4, the spectrum of the FITC-labelled substrate shows decreased intensity and a small red shift in the presence of each of these proteins. No significant change in intensity or red shift is observed when FITC-labelled insulin B chain is incubated with the structurally perturbed G376A mutant or when the fluorescein dye is incubated with wild-type Tsp. These results suggest that the S430A, D441A, and K455A variants bind substrate in a fashion similar to wild-type Tsp, making it likely that the side chains altered by these mutations are involved directly in the catalytic mechanism.

DFP does not inhibit Tsp: The properties of the S430A mutant suggest that Ser-430 may be an active-site residue. For this reason, I tested whether DFP, an inhibitor of many serine proteases, could inactivate Tsp. After Tsp was incubated with 50 mM DFP for 1 h, it was still fully active in cleavage of the λ-repressor #105 substrate (data not shown). Moreover, incubation of Tsp with tritiated DFP resulted in less than 1% labelling compared to a bovine trypsin control.

Conservative mutations at positions 430, 441, and 455: In some proteases, cysteine can replace the active-site serine with retention of proteolytic activity (7, 8, 17, 18). To test this in Tsp, the S430C mutant was constructed and purified. The S430C variant is active *in vitro* (Fig. 2A), displaying about 5-10% wild-type activity (Fig. 2C), and can rescue the conditional growth defect of a *tsp*-deletion strain (Table 1). The activity of the S430C mutant is inhibited by the cysteine-modifying reagents NEM and iodoacetamide (Fig. 2, B and D). By contrast, these alkylating agents have no effect on the activity of wild-type Tsp which contains no cysteine residues (1). When the S430C mutant was stored for several days at 0 °C in the absence of reducing agents, activity was lost but could be restored by incubating with 50 mM DTT for 1 h at room temperature. This suggests that the activity of the S430C variant is sensitive to some type of reversible oxidation reaction.

To determine if the carboxyl group of Asp-441 is required for activity, the isosteric D441N variant was constructed and purified. This mutant retains approximately 10% of the wild-type activity *in vitro* and is active *in vivo* (Fig. 2, Table 1). This result indicates that the carboxylate of Asp441 is not essential for the proteolytic activity of Tsp. Since the D441A mutant is inactive, I presume that Asn-441 can partially substitute for Asp-441 by participating in similar hydrogen-bonding or steric interactions.

To probe the tolerance of position 455 to substitutions, the codon was randomized and 47 colonies were isolated, tested for activity *in vivo*, and sequenced. Although all residues except Met, Cys, and Trp were recovered, only genes encoding Lys at position 455 were active. The K455R and K455H variants were purified and found to be inactive when assayed for proteolytic

activity *in vitro* (Fig. 2, Table 1). These data indicate that other basic side chains cannot functionally substitute for Lys-455. The K455H variant has a wild-type CD spectrum, while the K455R variant has an altered CD spectrum. The structural disruption caused by the seemingly conservative K455R substitution may indicate that the Lys-455 side chain is partially buried in a way that does not allow larger side chains to be accommodated.

DISCUSSION

In this study, I have used site-directed mutagenesis to identify residues important for Tsp activity *in vitro* and *in vivo*. Ser-430, Asp-441, and Lys-455 have properties expected of active-site residues. Alanine-substitution mutations at these positions abolish activity, but do not appear to affect substrate binding or to perturb the Tsp structure. Another group of residues (Gly-375, Gly-376, Glu-433, and Thr-452) seem to be important for maintaining the structure of Tsp. Thirteen additional residues, including each of the histidine residues in Tsp and two highly-conserved arginines, do not appear to be functionally or structurally important.

The three active-site residues in Tsp appear similar in some respects to the residues of the catalytic triad of classical serine proteases. The traditional serine-protease mechanism involves a serine nucleophile, a histidine which acts as a general base to activate the hydroxyl group of the serine, and an aspartate which stabilizes the partial charge assumed by the histidine (19). Substitution of any of the catalytic triad residues with alanine in well-studied serine proteases such as subtilisin reduces activity by more than 1000-fold (20). The inactivity and partial activity, respectively, of the S430A and S430C

mutants of Tsp, are consistent with the possibility that Ser-430 acts as a nucleophile. Moreover, although Tsp is not sensitive to inhibitors of classical serine proteases, such as DFP and PMSF, the S430C variant is inhibited by modification of the thiol group. Lys-455 of Tsp could function as a general base, in place of histidine, to activate the serine. However, Asp-441 of Tsp must play a role somewhat different than the aspartate of the classical catalytic triad, since the D441N substitution in Tsp does not affect activity whereas the analogous substitution in trypsin reduces activity almost completely (21).

Tsp may be similar, in some respects, to the class A β-lactamases, which use a serine, lysine, and aspargine in hydrolysis of the lactam bond of penicillins and cephallosporins (9). Biochemical and crystallographic data suggest that the serine of β-lactamase is activated for nucleophillic attack by donating a hydrogen bond to the lysine, which in turn is stabilized by hydrogen bonds with an asparagine and another serine (9). There is also evidence for mechanisms involving a serine activated by a lysine for the bacterial proteases LexA, Lep (EC), and SipS (6-8). Like Tsp, these proteases have critical serine and lysine residues but are not readily reactive with PMSF or DFP. It is important to note that although Tsp resembles these proteases and β-lactamases in some ways, it does not appear to share sequence or structural homology with any of these enzymes.

The conservation of the critical serine, aspartate, and lysine residues of Tsp in CtpA and the other sequences shown in Fig. 1 suggests that the corresponding residues in these proteins will also be catalytically important. In the case of the IRBP repeats, two of the eight known sequences have a serine and aspartate at positions corresponding to Ser-430 and Asp-441 of Tsp.

However, none of the IRBP repeats contains a lysine or other basic residue at positions corresponding to Lys-455 of Tsp. This lack of conservation of key residues is consistent with the fact that proteolytic activity has not been observed for IRBP. Instead, it appears that IRBP retains the fold of the Tsp active-site region, without the catalytic residues.

ACKNOWLEDGMENTS: I thank Karen Silber and Bronwen Brown for gifts of protein, Sarah Ades for technical assistance, and Ray Larsen and Karen Postle for communicating unpublished data. Supported by NIH grants AI-15706 and AI-16892.

REFERENCES

1. Silber, K. R., Keiler, K. C., and Sauer, R. T. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 295-299

2. Hara, H., Yamamoto, Y., Higashitani, A., Suzuki, H., and Nishimura, Y. (1991) *J. Bacteriol.* **173**, 4799-4813

3. Azizan, A., and Black, P. (1994) *J. Bacteriol.* **176**, 6653-6662

4. Keiler, K. C., Silber, K. R., Downard, K. M., Papayannopoulos, I. A., Biemann, K., and Sauer, R. T. (1995) *Prot. Sci.* **4**, 1507-1515.

5. Shestakov, S.V., Anbudurai, P.R., Stanbekova, G.E., Gadzhiev, A., Lind,L.K. and Pakrasi,H.B. (1994) *J. Biol. Chem.* **269**, 19354-19359

6. Little, J. W. (1993) *J. Bacteriol.* **175**, 4943-4950

7. Tschantz, W. T., Sung, M., Delgado-Partin, V. M., and Dalbey, R. E. (1993) *J. Biol. Chem.* **268**, 27349-27354

8. van Dijl, J. M., de Jong, A., Venema, G., and Bron, S. (1995) *J. Biol. Chem.* **270**, 3611-3618

9. Strynadka, N. C. J., Adachi, H., Jensen, S. E., Betzel, C., Sutoh, K., and James, M. N. G. (1992) *Nature* **359**, 700-705

10. Silber, K. R., and Sauer, R. T. (1994) *Mol. Gen. Genet.* **242**, 237-240

11. Higuchi, R. (1990) in *PCR protocols: A Guide to Methods and Applications*: (Innis M.A., Gelfand, D. H., Sininsky, J. J. and White, T. J., eds) pp. 177-183, Academic Press, Inc., San Diego, CA

12. Lim, W. A., and Sauer, R. T. (1991) *J. Mol. Biol.* **219**, 359-376

13. Schagger, H. and von Jagow, G. (1987) *Anal. Biochem.* **166**, 368-379

14. Altschul S. F., Gish, W., Miller, W., Meyers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403-410

15. Genetics Computer Group, (1991), *Program Manual for the GCG Package, Version 7*, Madison, WI

16. Dunn, B. M. (1989) in *Proteolytic Enzymes: A Practical Approach* (Beynon, J. R., and Bond, J. S., eds) pp. 57-81, IRL Press, Oxford, UK

17. Nakatsuka, T., Sasaki, T., and Kaiser, E. T. (1987) *J. Am. Chem. Soc.* **109**, 3808-3810

18. Higaki, J. N., Evnin, L. B., and Craik, C. S. (1989) *Biochem.* **28**, 9256-9263

19. Antonov, V. K. (1993) *Chemistry of Proteolysis*, Springer-Verlag, Berlin

20. Carter, P., and Wells, J. (1988) *Nature* **332**, 564-568

21. Craik, C. S., Roczniak, S., Largman, C., and Rutter, W. J. (1987) *Science*, **237**, 909-913

variant	proteolysis of #105 λ-repressor substrate	growth on 1/2 LB no salt at 42 °C	wild-type CD spectrum
wild type	+	+	yes
H19A	+	+	
H34A	+	+	
H60A	+	+	
H203A	+	+	
D369A	+	+	yes
R371A	+	+	
S372A	+	+	
G375A	–	–	no
G376A	–	–	no
S428A	+	+	
S430A	–	–	yes
S430C	+	+	yes
S432A	+	+	yes
E433A	–	–	no
D441A	–	–	yes
D441N	+	+	yes
R444A	+	+	
E449A	+	+	yes
T452A	–	–	no
K455A	–	–	yes
K455H	–	–	yes
K455R	–	–	no
D505A	+	+	
H553A	+	+	

Table 1: Properties of wild-type Tsp and variants

Fig. 1. Sequence homology. (*top*) Schematic representation of the Tsp sequence. Regions of homology with CtpA, Aqu, BaoCtpA, and Ngu and with the IRBP repeats (1) are shown. (*bottom*) Sequence alignments. Numbers refer to the Tsp sequence, and residues which are identical in all sequences are shaded. Open circles indicate alanine substitutions which result in an active Tsp protease. Squares indicate substitutions which result in inactivity and structural perturbations. Black circles indicate substitutions which result in inactivity without structural perturbations or substrate binding defects. The accession numbers for the sequences are Tsp (P23865), CtpA (L25250), Aqu (S18125), BaoCtpA (L37094), and Ngu (U11547). (following page)

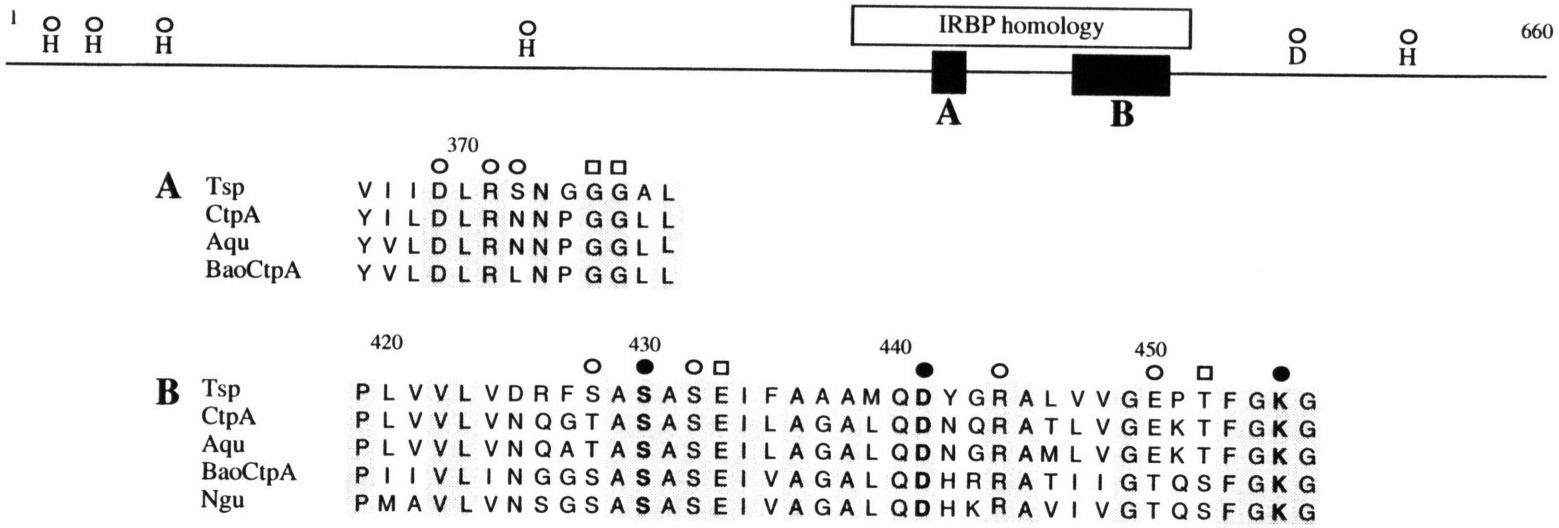
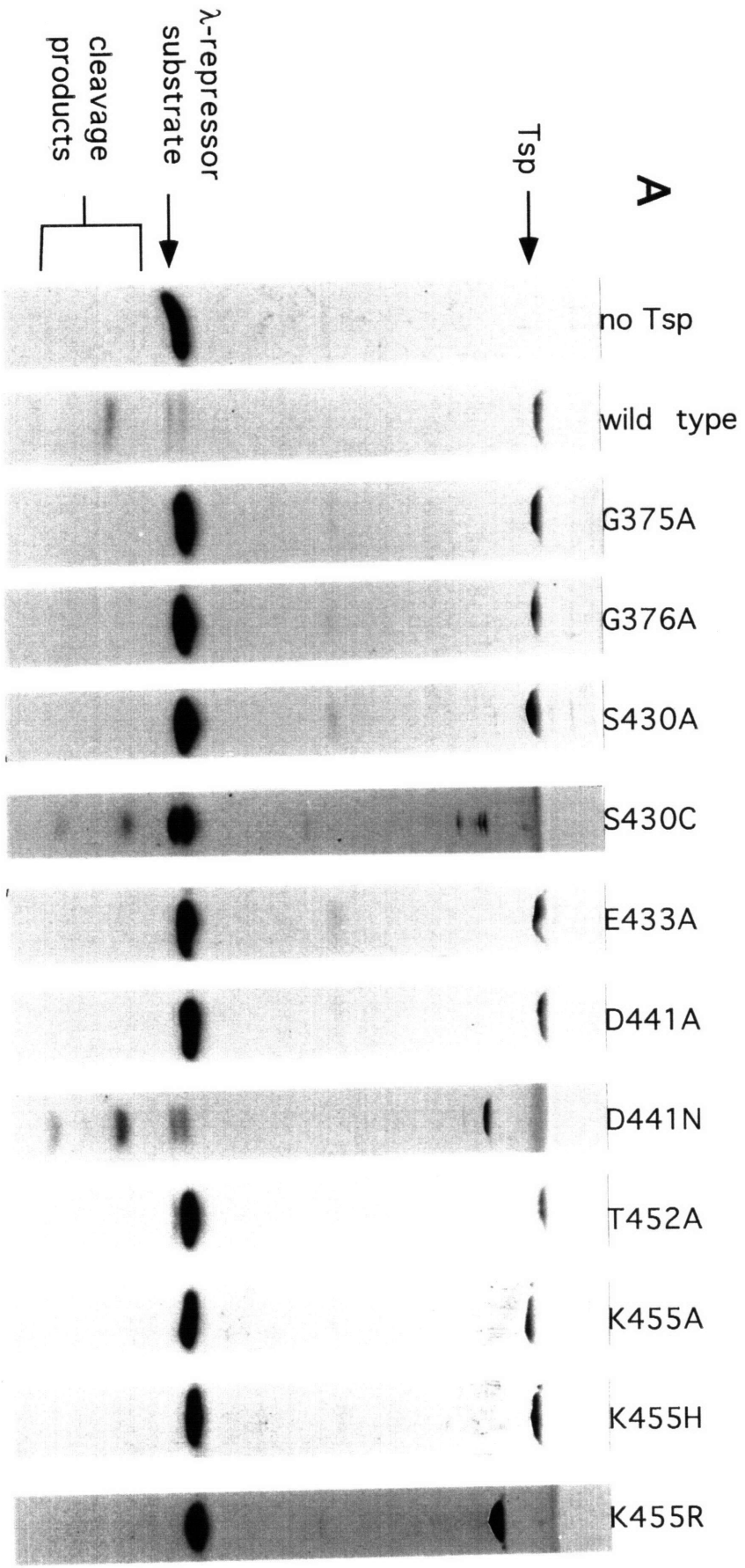
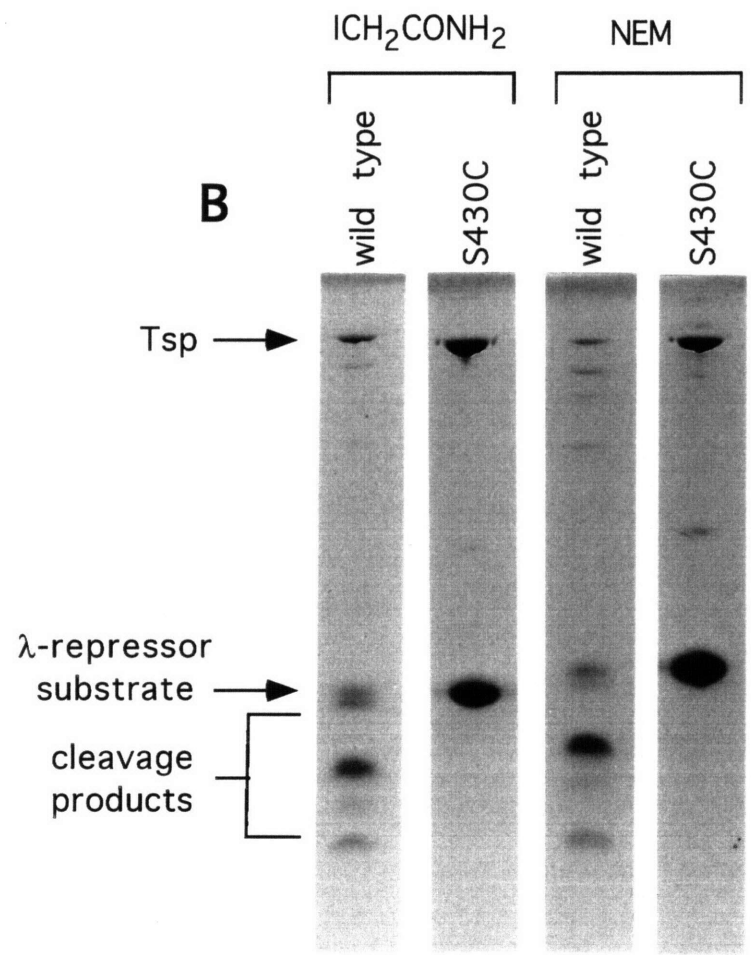
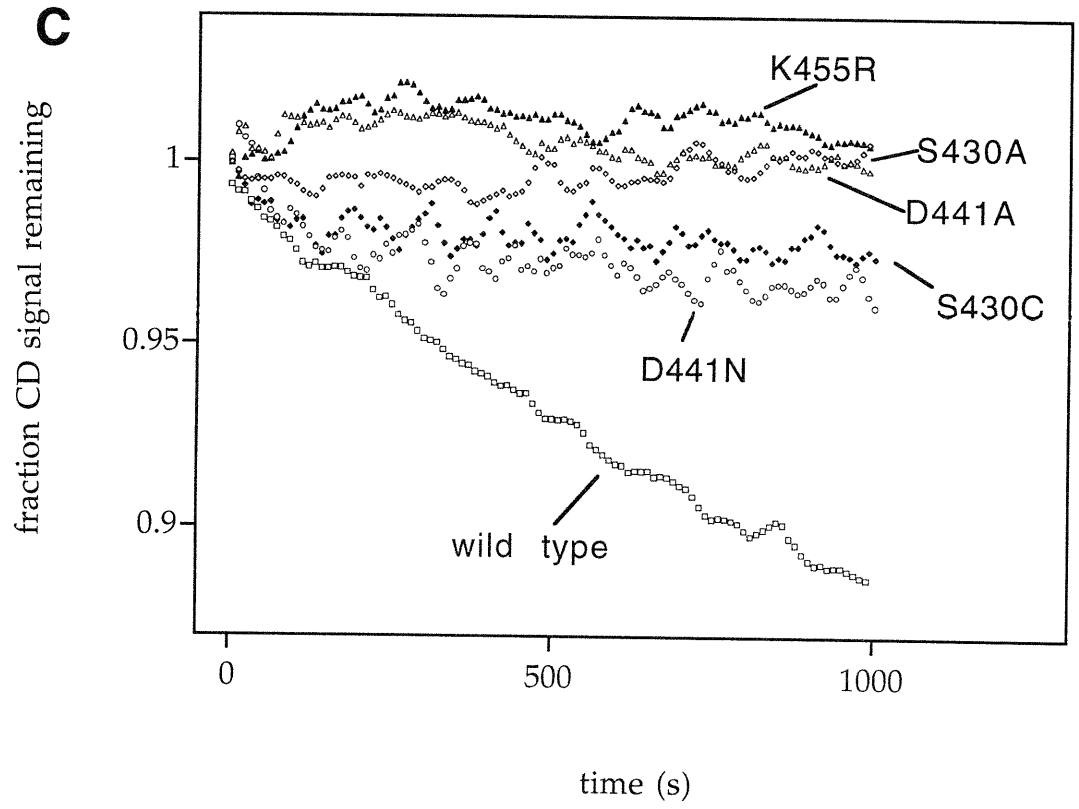
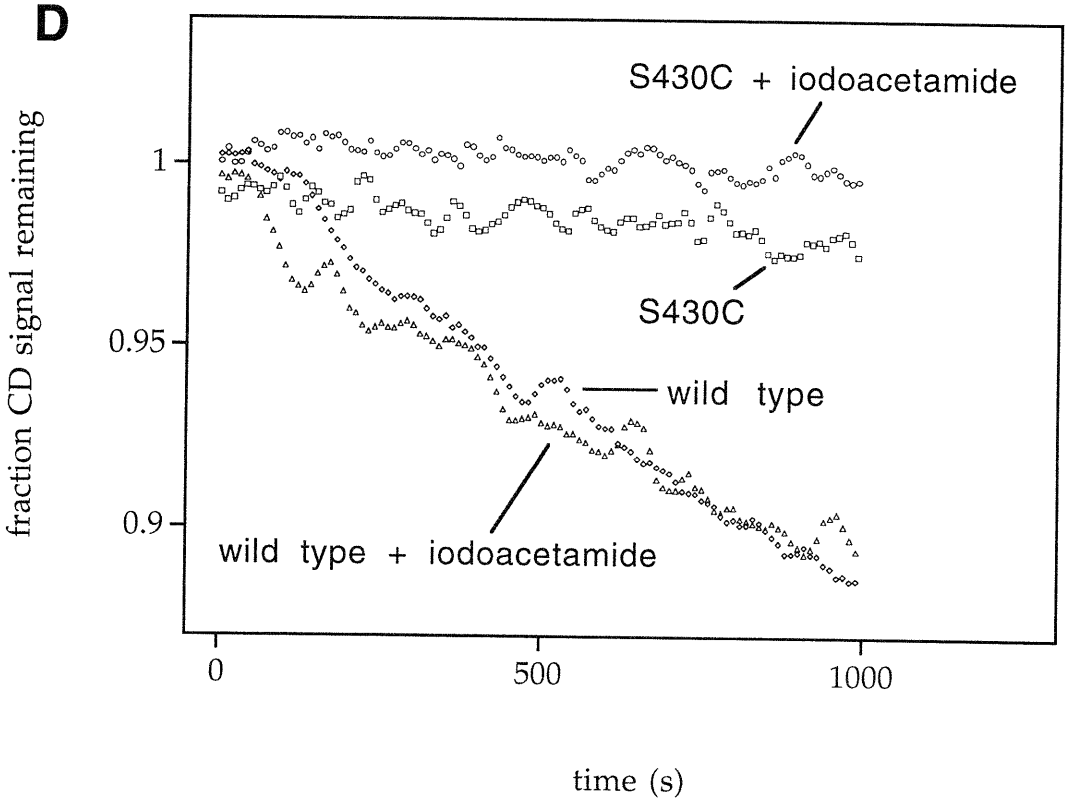

1
H
H
H
H
IRBP homology
D
H
660
A
B
370
A
Tsp
CtpA
Aqu
BaoCtpA
420
430
440
450
B
Tsp
CtpA
Aqu
BaoCtpA
Ngu

Fig. 2. Substrate cleavage by wild-type Tsp and selected variants. (*A*) Gel assays for degradation of the λ-repressor #105 substrate by wild-type Tsp and variants after a 20 h incubation at 37 °C. (*B*) Effects of incubation of wild-type Tsp or the S430C variant with 10 mM iodoacetamide or 10 mM NEM for 20 min at room temperature prior to the cleavage assay. Arrows indicate the positions of the intact Tsp and λ-repressor substrate bands. Approximately the same initial concentration of Tsp or variant was used in each assay. The final concentrations of enzyme are lower for wild-type Tsp and the S430C and D441N variants because of auto-digestion. Some Tsp variant bands appear to have different mobilities because they were electrophoresed on gels run for different periods of time. (*C*) Time courses of degradation of Arc repressor (10 μM) by wild-type Tsp and variants (250 nM) at 37 °C monitored by loss of CD signal at 222 nm. (*D*) Effects of incubation of wild-type Tsp or the S430C variant with 10 mM iodoacetamide and 18.5 mM ethanol for 20 min at room temperature prior to the Arc cleavage assay (conditions as in panel C). Control reactions in which the protease is incubated with 18.5 mM ethanol alone are shown. Incubation with ethanol alone results in a reduction in activity of about 5% for both wild-type Tsp and S430C. (following page)


A
Tsp
λ-repressor substrate
cleavage products
no Tsp
wild type
G375A
G376A
S430A
S430C
E433A
D441A
D441N
T452A
K455A
K455H
K455R


B
ICH2CONH2
NEM
wild type
S430C
wild type
S430C
Tsp
λ-repressor
substrate
cleavage
products


C
fraction CD signal remaining
1
0.95
0.9
0
500
1000
time (s)
K455R
S430A
D441A
S430C
D441N
wild type


D
fraction CD signal remaining
1
0.95
0.9
0
500
1000
time (s)
S430C + iodoacetamide
S430C
wild type
wild type + iodoacetamide

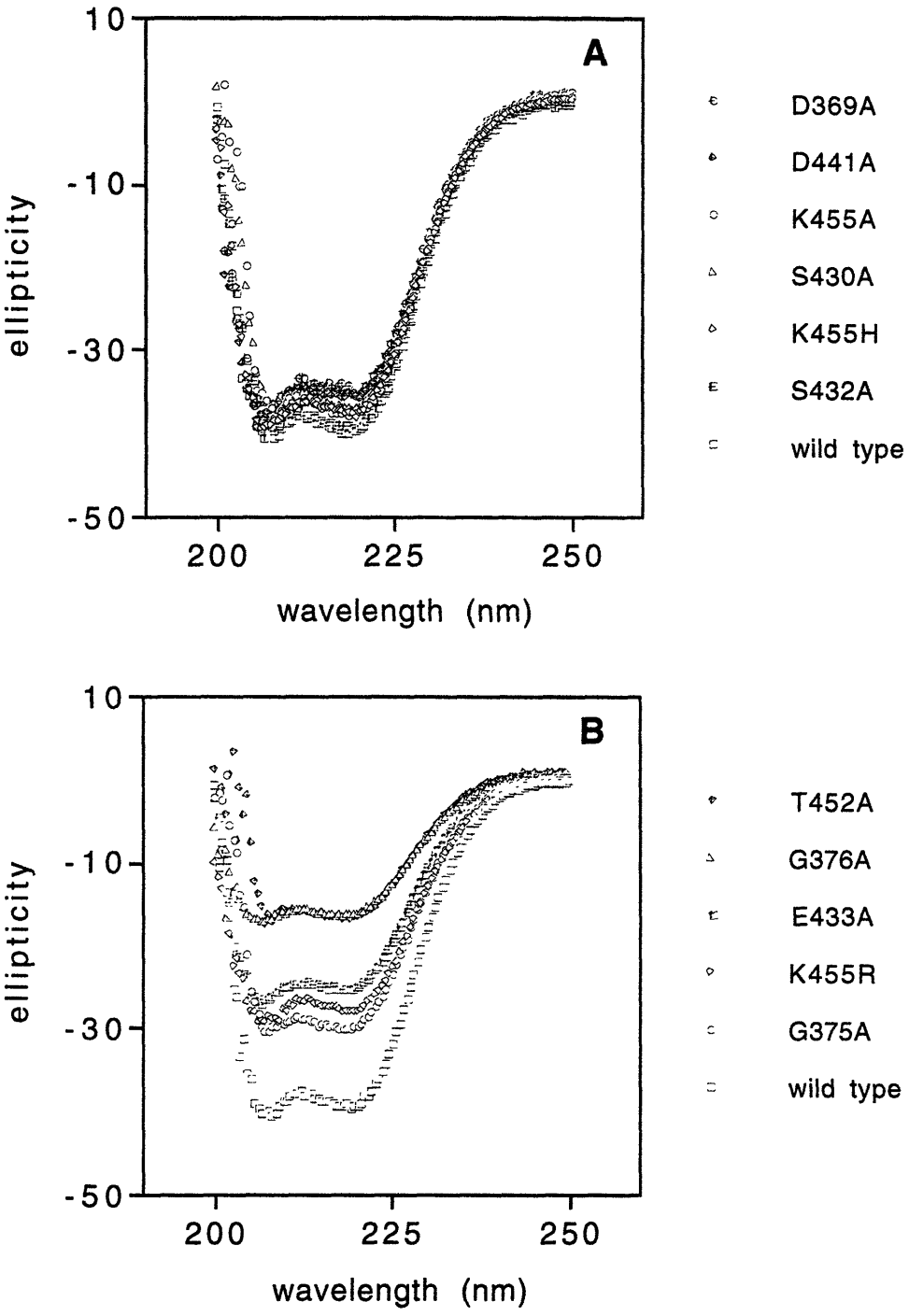


Fig. 3. Circular dichroism spectra of wild-type Tsp and variants. (*A*) Spectra of wild-type Tsp and structurally unperturbed variants. (*B*) Spectra of wild-type Tsp and structurally perturbed variants.

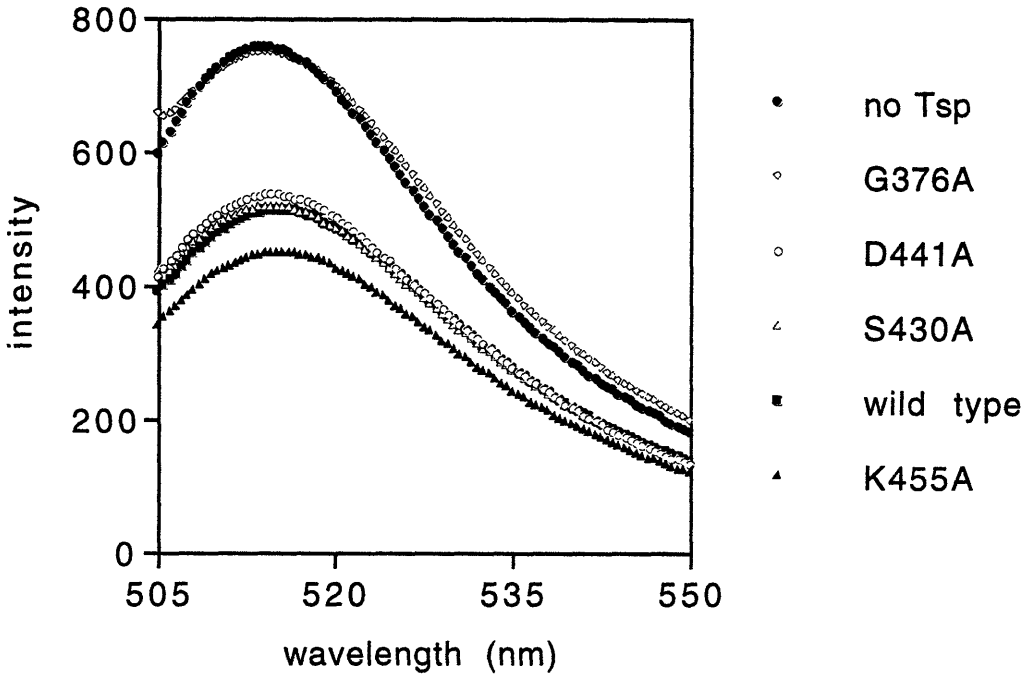


Fig. 4. Fluorescence emission spectra of FITC-labelled insulin B chain in the presence and absence of wild-type Tsp and variants. 667 nM of wild-type or variant Tsp was incubated with 667 nM FITC-labelled insulin B chain at 5 °C in buffer containing 10 mM Tris-HCl [pH 8.0] plus 20 mM NaCl, and the emission spectra were measured after excitation at 496 nm. Substrate binding is indicated by a reduction in intensity and a small red shift.

CHAPTER 4

Sequence Determinants of C-terminal Substrate Recognition by the Tsp Protease

INTRODUCTION

The observed selectivity of intracellular proteolysis implies the existence of mechanisms which allow proteases to discriminate between correct and incorrect protein substrates (1, 2). Some substrate recognition may be possible at the level of exposure of appropriate cleavage sites. For example, a protease might cleave almost any unfolded or partially folded protein in which peptide bonds flanked by the proper P1 and P1′ residues are accessible. This type of mechanism may serve to rid the cell of misfolded or unfolded proteins, but is unlikely to allow significant selectivity unless the local sequence determinants of cleavage site selection occur only rarely. Some protein substrates may be marked for degradation by covalent modification with molecular tags which then serve as recognition determinants in subsequent steps. The ubiquitin-proteosome system of eukaryotic cells is currently the best understood example of this type, although the determinants which cause particular proteins to be modified by ubiquitin are still not well understood (3). In bacteria, there are systems in which the identity of sequences at the C-terminal or N-terminal ends of proteins appear to serve as determinants of proteolytic degradation (4-6). In these cases, such sequences may serve as secondary binding sites which allow a protease to tether a substrate while waiting for rare unfolding events that expose the sites of primary cleavage.

Specific degradation of proteins with nonpolar C-terminal sequences was first reported for several cytoplasmic proteins in *Escherichia coli* (4, 5), but the protease that mediates this degradation has not been identified. Tsp (tail-specific protease) is a periplasmic protease of *E. coli* that was purified based on

its ability to differentially degrade two protein substrates that differed only in their C-terminal residues (7). The protein substrate cleaved by Tsp had a relatively apolar C-terminal sequence (WVAAA), while the protein resistant to Tsp cleavage had a relatively polar C-terminal sequence (RSEYE). Although this specificity *in vitro* is similar to that observed *in vivo* for cytoplasmic degradation (5), gene knockout experiments have shown that Tsp is not involved in cytoplasmic degradation (8). Experiments *in vitro* have established that Tsp is an endoprotease which cleaves substrates at discrete sites throughout the polypeptide chain in a reaction that depends upon the identity of the substrate's C-terminal sequence and requires the presence of a free α-carboxyl group (9). The precise determinants that allow some C-terminal sequences but not others to be recognized by Tsp are currently unknown. Moreover, it has not yet been established that C-terminal specific degradation of substrates by Tsp occurs in the cell. In this chapter, I address these issues by studying the susceptibility of cytochrome-b_{562} variants to Tsp-mediated proteolysis *in vivo* and *in vitro*. Cytochrome b_{562} is a periplasmic protein which can be readily detected in cell lysates because binding of the protein to heme gives rise to a characteristic red absorbance spectrum (10). I show that wild-type cytochrome b_{562} is resistant to Tsp-mediated cleavage but becomes a good substrate when a WVAAA C-terminal tail is added. To investigate sequence preferences at the C terminus, I constructed libraries of cytochrome b_{562}-WV<u>AAA</u> with each of the last three tail positions randomized individually, and assayed for steady-state levels of the modified variants in cells containing Tsp or deleted for Tsp. These results reveal different preferences at each of the three C-terminal positions and show that Tsp is the major periplasmic protease responsible for C-terminal-specific degradation of these substrates.

EXPERIMENTAL PROCEDURES

Strains, plasmids, and mutagenesis— *E. coli* strain X90 is *ara* Δ(*lac pro*) *gyrA argE*(Am) Rif$^-$ *thi-1*/F′ *lacI*^{q} *lac*$^+$ *pro*$^+$; *E. coli* strain KS1000 is X90 Δ*tsp(prc)3*::kan *eda-51*::Tn*10* (8). Plasmid pKK101 is a pBluescript-derived vector which encodes ampicillin resistance and the Tsp-His$_6$ protease under control of a *lac*-promoter (9). Plasmid pRW-1 (a gift from Michael Hecht) is a pEMBL-18 derived plasmid which encodes ampicillin resistance and the cytochrome-b_{562} gene under transcriptional control of a *lac*-promoter (11).

A plasmid (pCyb2) encoding a variant of cytochrome b_{562} with the C-terminal tail sequence WVAAA was constructed by ligating the *Pst*1-*BamH*1 backbone fragment from pRW-1 (the *Pst*1 site is near the 3′ end of the cytochrome b_{562} gene; the *BamH*1 site is roughly 150 bp downstream) with a double-stranded oligonucleotide cassette encoding the 3′ end of the gene, codons for the WVAAA sequence, and the wild-type stop codon and termination sequences. The structure of pCyb2 was confirmed by restriction mapping and DNA sequencing. To randomize the C-terminal codons of the cytochrome-b_{562}-WVAAA gene, the pCyb2 construction was repeated using an oligonucleotide cassette containing an equimolar mixture of G, A, T, and C at the appropriate codon. These libraries were transformed into X90 cells, single colonies were isolated, and the C-terminal sequences of genes from 60-75 independent candidates were determined by DNA sequencing.

Protein purification— Wild-type Tsp with a six histidine tag was purified from *E. coli* strain KS1000/pKK101 using nickel-chelate chromatography and ion-exchange chromatography as described (9).

Cytochrome b_{562} was purified from *E. coli* strain KS1000/pRW-1 using a protocol adapted from Ames *et al.* (12). Cells were grown at 37 °C with gentle shaking for 12 h in 1 L of 2X-YT broth supplemented with 150 μg/mL ampicillin and 1 mM IPTG and were harvested by centrifugation. The cell pellet was resuspended in 10 mL chloroform, incubated at room temperature for 15 min, and 100 mL of 10 mM Tris-HCl (pH 8.0) was added. This mixture was centrifuged for 20 min at 4,000 rpm, and the red aqueous supernatant containing cytochrome b_{562} and other periplasmic proteins was recovered. 10 mL of a 0.5 M potassium citrate (pH 4.0) buffer was added, and the pH was brought to 4.0 by addition of 1 M HCl. After stirring for 30 min at 4 °C, precipitated proteins were removed by centrifugation at 10,000 rpm for 30 min. The supernatant was loaded onto a Mono S column (Pharmacia) equilibrated in 25 mM potassium citrate (pH 4.0) and the column was developed with a linear gradient from 0 to 300 mM KCl in the same buffer. Fractions containing cytochrome b_{562}, which eluted between 75 and 100 mM KCl, were pooled, concentrated, and chromatographed on a Superose 12 column (Pharmacia) in buffer containing 50 mM Tris-HCl (pH 8.0), 0.1 mM EDTA. Fractions which contained cytochrome b_{562} at greater than 95% purity, as assayed by SDS-polyacrylamide gel electrophoresis, were pooled. Variant cytochrome-b_{562} proteins were purified in the same manner as the wild type.

Expression assays— The steady-state levels of cytochrome-b_{562} variants in *E. coli* strain X90 (which has a chromosomal copy of the *tsp* gene) or strain KS1000 (which is deleted for the *tsp* gene) were determined following 12 h growth at 37 °C after inoculating 50 μL of an overnight culture into 5 mL of 2X-YT broth supplemented with 150 μg/mL ampicillin, 1 mM IPTG, and 10 μg/mL $FeCl_2$. Cells were harvested by centrifugation, the cell pellets were

frozen in an ethanol/dry ice bath for 15 min, and thawed in a 4 °C water bath for 15 min. This freeze-thaw protocol releases cytochrome b_{562} (11). After three freeze-thaw cycles, the cell pellets were resuspended in 1 mL water, incubated on ice for 30 min, and the cells were removed by centrifugation at 4 °C. The supernatants were recovered and the absorbance spectra from 220 nm to 450 nm was recorded. Assays from three independent cell cultures were performed for each cytochrome-b_{562} variant, and the OD values at 426 nm (the absorbance maximum for cytochrome b_{562} with a reduced heme group) were averaged.

Pulse-chase assays— The half-lives of cytochrome-b_{562} variants *in vivo* were detrmined by growing cells to mid-log phase in M9 minimal medium containing no Cys or Met and inducing protein expression by addition of 1 mM IPTG. Twenty min after induction, a labelling pulse of 100 μCi of ^{35}S-methionine was added and 30 s later a chase of unlabelled L-methionine was added to a final concentration of 1.4 mg/mL. At different times, 0.5 mL aliquots were removed, immediately frozen in an ethanol-dry ice bath, and lysed by 3 cycles of thawing at 4 °C followed by refreezing. Samples were electrophoresed on SDS-polyacrylamide gels and the radiolabelled bands were visualized by phosphorimaging. Half-lives were determined by fitting the pulse-chase data to an exponential decay. For these calculations, the intensities of the induced bands were integrated using the phosphorimager ImageQuaNT software and were normalized to the intensities of a set of stable bands.

Cleavage assays— The cleavage of cytochrome-b_{562} variants by Tsp *in vitro* was monitored by the change in heme absorbance which accompanies

its dissociation from the cleaved protein. For this assay, 2.5 μM of the purified cytochrome b_{562} variant was incubated at 22 °C with 0.3 μM Tsp-His_6 in 50 mM Tris-HCl (pH 8.0), 20 mM NaCl, and the loss of absorbance at 418 nm was monitored (418 nm is the absorbance maximum for the protein with an oxidized heme group; oxidation of the heme in the purified protein occurs spontaneously but does not alter the structure or stability of cytochrome b_{562}). As a control, aliquots were removed from the cuvette at various time points and assayed for cleavage of cytochrome b_{562} by SDS gel electrophoresis. As expected, loss of absorbance at 418 nm was found to correlate with loss of the intact cytochrome-b_{562} band.

Stability of cytochrome-b_{562} variants— The circular dichroism spectra of cytochrome-b_{562} variants were measured using 25 μg/mL protein in buffer containing 50 mM Tris-HCl (pH 8.0), 0.1 mM EDTA. Thermal melts were performed using the same protein concentration and buffer composition by monitoring the ellipticity at 222 nm as a function of temperature. Melting curves were fit to a two-state transition between native and denatured protein by non-linear least squares fitting using the program NONLIN for Macintosh (13, 14).

RESULTS

Wild-type cytochrome b_{562} is not cleaved by Tsp *in vitro* (Fig. 1A) and is expressed at comparable levels in *tsp*$^+$ and *tsp*$^-$ cells as determined by heme absorbance (Fig. 1B). Neither finding is surprising. The Tsp protease is thought to prefer non-polar tails that are accessible in the folded protein (7, 9), whereas cytochrome b_{562} has a very polar C-terminal sequence (HQKYR)

which is relatively inaccessible in the crystal structure (15, 16). In an attempt to make cytochrome b_{562} a substrate for Tsp, I constructed a gene with the cytochrome coding sequence followed by codons for the C-terminal pentapeptide WVAAA. This sequence was chosen because purified Tsp is known to cleave variants of the N-terminal domain of λ repressor and of Arc repressor which have the WVAAA tail *in vitro* (7; unpublished data). As shown in Fig. 1A, the purified cytochrome-b_{562}-WVAAA protein is also cleaved by Tsp *in vitro*. The cytochrome-b_{562}-WVAAA protein is expressed at a much lower steady-state level in cells containing Tsp than in cells lacking Tsp (Fig. 1B), and pulse-chase experiments show that the reduced steady-state level is caused by increased intracellular degradation, which is Tsp-dependent (Table 1).

Steady-state expression levels depend on the identity of the C-terminal residues— Each of the three C-terminal residues of the cytochrome-b_{562}-WVAAA substrate was individually randomized and from 16-20 different amino acids were recovered at each position. The resulting variants were assayed for steady-state expression levels in *tsp*$^+$ and *tsp*$^-$ cells as shown in Fig. 2. In this figure, open bars represent expression levels in the absence of Tsp and closed bars represent expression levels in the presence of Tsp. In the cells containing Tsp, there are significant differences in expression levels depending on the chemical identity of the three C-terminal residues. These difference are discussed in greater detail below. In the cells lacking Tsp, all of the variants are expressed at similar levels. This suggests that Tsp is the major periplasmic protease responsible for C-terminal specific protein degradation of the cytochrome-b_{562} substrates.

Cytochrome variants with Ala, Cys, Val, Ser, and Thr at the C terminus (position 1) are expressed at the lowest steady-state levels in cells containing Tsp (Fig. 2), indicating that Tsp prefers substrates which have small, uncharged side chains at this position. Variants with polar side chains, large side chains, Pro, or Gly at the C-terminus are expressed at reasonably high levels even in the presence of Tsp. In many cases, the expression levels of these variants are as high as in cells deleted for Tsp (Fig. 2). This suggests that the identity of the C-terminal residue is extremely important in determining whether a protein will be a good or poor substrate for Tsp.

At the penultimate amino-acid residue (position 2), variants with Ala, Tyr, Ile, and Trp are expressed at the lowest levels in the presence of Tsp and variants with Arg, Lys and Gly are expressed at the highest levels. In general, hydrophobic residues at position 2 appear to be preferred by Tsp relative to hydrophilic residues. Moreover, only a few side chains at this position increase expression to levels comparable to those seen in *tsp*$^-$ cells. This suggests that position 2 is less important than position 1 in determining resistance to Tsp cleavage.

At the third position from the C terminus, Tsp prefers Ala, Leu, Val, and Ile. The least preferable side chains are Asn, Gln and Met. It is somewhat surprising that Leu and Met, which are often considered to be conservative substitutions for each other, have such different effects at this position. No side chains at position 3 increase steady-state expression to levels observed in the absence of Tsp, suggesting that this position is less important than either position 1 or 2 in determining resistance to Tsp cleavage.

Comparison of cleavage rates in vivo and in vitro— I purifed cytochrome-b_{562} variants with at least one stabilizing and one destabilizing residue at each of the three C-terminal positions, and assayed rates of cleavage by Tsp *in vitro* (Table 1). Tsp rapidly cleaved variants with the AAA, AAV, AYA, and LAA tails but did not cleave or only slowly cleaved variants with the AAK, AGA, and QAA tails. Pulse-chase experiments were used to determine half-lives for these same variants in the cell (Fig. 3). As expected, variants which had low steady-state expression levels (AAA, AYA, LAA) had the shortest half-lives, and variants which had high steady-state expression levels (AAK, AGA) had considerably longer half-lives (Fig. 2; Table 1). As shown in Table 1, the resistance of the cytochrome-b_{562} variants to cleavage by Tsp *in vitro* correlates reasonably well ($R^2 \approx 0.9$) with the half-lives of the variants in a *tsp*$^+$ strain *in vivo*. Thus, these data support the model that differential degradation of these proteins by Tsp *in vivo* is responsible for the observed differences in their half-lives and steady-state levels.

Effects of tail sequence on protein stability— In principle, the different C-terminal sequences of the cytochrome-b_{562} variants might affect susceptibility to Tsp by reducing the thermodynamic stability of the protein. To test for this possibility the stabilities of the purified cytochrome-b_{562} variants were measured by temperature denaturation monitored by CD spectroscopy (Fig. 4). The melting curve of each variant showed a T_m between 68 and 69 °C, indicating that the C-terminal mutations do not affect the global stability of cytochrome b_{562}.

DISCUSSION

The work presented here has established the importance of specific amino acids at each of the three C-terminal residues in determining whether a protein is efficiently cleaved by Tsp *in vitro* and *in vivo*. In previous studies based on screening of a small number of potential protein and peptide substrates, I had concluded that Tsp appeared to recognize substrates with non-polar or hydrophobic C-terminal residues and not to recognize substrates with polar C-terminal residues (7, 9). The data summarized in Fig. 2 reveal that this view is an over-simplification. While no good substrates have highly polar tails and most good substrates do have non-polar residues at the three C-terminal positions, there is considerable fine specificity. For example, small nonpolar residues are preferred relative to larger hydrophobic side chains at the C-terminal position. At the other two positions, there is no simple correlation between the size of nonpolar side chains and effects on Tsp cleavage. For example, alanine and tyrosine are the most destabilizing side chains at the penultimate residue, while valine has a significantly smaller effect at this position.

How do the C-terminal residues of a substrate affect its degradation by Tsp? Unlike systems in which sequence signals act to target substrates to subcellular compartments specialized for degradation (17), the C-terminal sequences that mediate degradation by Tsp appear to be recognized by the protease itself. This is shown most clearly by the strong correlation between the half-lives of the cytochrome-b_{562} variants *in vivo* and the resistance of these purified variants to cleavage by purified Tsp *in vitro*. Since other macromolecules are not required for Tsp-mediated degradation of substrates

with nonpolar tails, then Tsp must either recognize these C-terminal sequences directly or recognize indirect effects of these sequences on stability or structure. Several experiments suggest that Tsp recognizes the tail sequences directly. As shown here, C-terminal sequences that make cytochrome b_{562} a good substrate for Tsp do not alter the thermodynamic stability of the protein. The same is true of C-terminal sequences that make Arc repressor and the N-terminal domain of λ repressor good substrates for Tsp cleavage *in vitro* (7, 9). Moreover, it seems unlikely that destabilizing C-terminal tails act indirectly by allowing other sequence or structural determinants to be recognized because the same tail sequence (e. g., WVAAA) can make Arc repressor, λ repressor, and cytochrome b_{562} (proteins that differ in primary, secondary, tertiary, and quaternary structure) good substrates for Tsp cleavage. I believe that Tsp directly recognizes protein and peptide substrates by using a binding site which requires a free α-carboxyl group and side chains of the appropriate size, shape, and hydrophobicity at the last three residue positions of the substrate. Such a binding site would serve to tether the substrate to the enzyme, and binding of an appropriate C-terminal tail at this site might also function to activate the enzyme.

Our studies have shown that Tsp can efficiently degrade a periplasmic protein with a WVAAA C-terminal tail. An independent protease in the cytoplasm of *E. coli* is also capable of rapidly degrading proteins with WVAAA tails (5, 8). Why does *E. coli* use the Tsp system in the periplasm and an independent system in the cytoplasm to degrade proteins with certain C-terminal sequences? It seems unlikely that this is a general mechanism for removing unfolded or misfolded proteins from the cell. First, most damaged or misfolded proteins would not be expected to have the proper C-terminal

sequence to allow degradation by a C-terminal-specific pathway. Second, Tsp and its cytoplasmic counterpart are not limited to degrading unfolded proteins. Both the cytochrome-b_{562} variants studied here and the λ-repressor variants used to study cytoplasmic C-terminal-specific degradation are stably folded (5). Recent studies suggest that one function of tail-specific proteases is to work in conjunction with a peptide-tagging system that marks certain proteins in *E. coli* for degradation (18, 19). In this system, proteins translated from damaged mRNA are modified by C-terminal addition of a peptide with the sequence AANDENYALAA. The C-terminal residues of this peptide tag then render the tagged protein susceptible to degradation by Tsp or by its cytoplasmic counterpart.

ACKNOWLEDGMENTS: I thank Michael Hecht for the gift of pRW-1, and Cliff Robinson for helpful advice. Supported by NIH grants AI-15706 and AI-16892.

REFERENCES

1. Gottesman, S., and Maurizi, M. R. (1992) *Microbiol. Rev.* **56**, 592-621

2. Goldberg, A. L., and St. John, A. C. (1976) *Annu. Rev. Biochem.* **45**, 747-803

3. Ciechanover, A. (1994) *Cell* **879**, 13-21

4. Bowie, J. U., and Sauer, R. T. (1989) *J. Biol. Chem.* **264**, 7596-7602

5. Parsell, D. A., Silber, K. R., and Sauer, R. T. (1990) *Genes Dev.* **4**, 277-286

6. Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. (1991) *Science* **254**, 1374-1377

7. Silber, K. R., Keiler, K. C., and Sauer, R. T. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 295-299

8. Silber, K. R., and Sauer, R. T. (1994) *Mol. Gen. Genet.* **242**, 237-240

9. Keiler, K. C., Silber, K. R., Downard, K. M., Biemann, K., and Sauer, R. T. (1995) *Prot. Sci.* **4**, 1507-1515

10. Itagaki, E., and Hager, L. P. (1966) *J. Biol. Chem.* **241**, 3687-3695

11. Brunet, A. P., Huang, E. S., Huffine, M. E., Loeb, J. E., Weltman, R. J., and Hecht, M. H. (1993) *Nature* **364**, 355-358

12. Ames, G. F.-L., Prody, C., and Kustu, S. (1984) *J. Bact.* **160**, 1181-1183

13. Brenstein, R. J. (1989) *NONLIN for Macintosh.* Carbondale, IL.: Robelko Software

14. Johnson, M. L., and Frazier, S. (1985) *Methods Enzymol.* **117**, 301-342

15. Mathews, F. S., Bethge, P. H., and Czerwinski, E. W. (1979) *J. Biol. Chem.* **254**, 1699-1706

16. Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D., and Mathews, F. S. (1981) *J. Molec. Biol.* **148**, 427-448

17. Dice, J. F. (1990) *Trends Biochem. Sci.* **15**, 305-309

18. Keiler, K. C., Waller, P. R. H., and Sauer, R. T. submitted

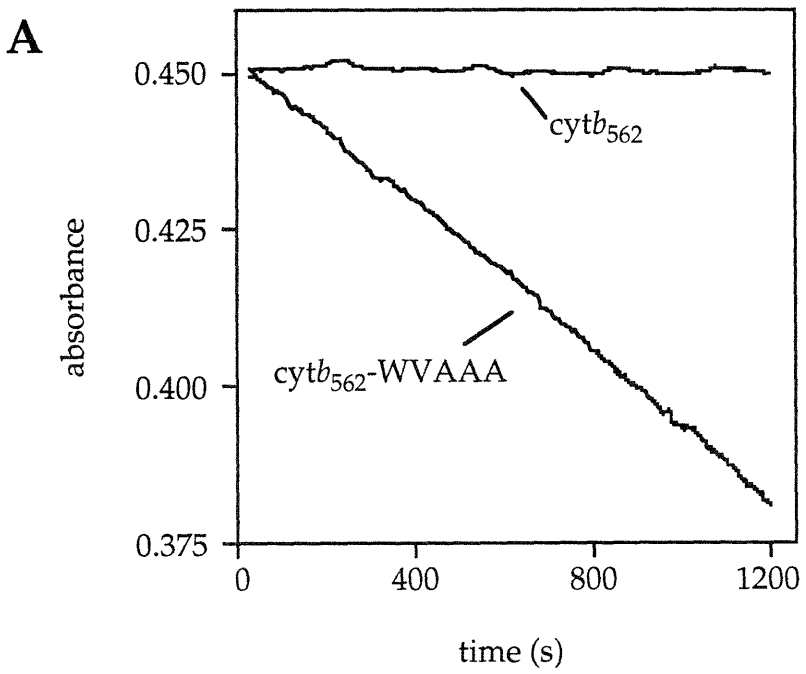
19. Tu, G.-F., Reid, G. E., Zhang, J.-G., Moritz, R. L., and Simpson, R. J. (1995) *J. Biol. Chem.* **270**, 9322-9326

Table 1: Degradation of cytochrome-b_{562} variants *in vivo* and *in vitro*.

	half-life *in vivo* (min)		relative resistance to cleavage
variant	*tsp*$^+$	*tsp*$^-$	by Tsp *in vitro* [1]
AAA	<2	>120	0.01
AYA	<2	>120	0.01
LAA	<2	>120	0.01
AAV	6	>120	0.02
AGA	36	>120	0.10
QAA	23	>120	0.13
AAK	>120	>120	0.30

[1] rate of cleavage of wild type / rate of cleavage of variant

Fig. 1. Sensitivity of cytochrome b_{562} and a variant bearing the WVAAA C-terminal tail to Tsp degradation *in vitro* and *in vivo*. (A) Cleavage of 2.5 μM of purified cytochrome b_{562} or cytochrome b_{562}-WVAAA by 0.3 μM Tsp *in vitro* monitored by loss of heme absorbance at 418 nm. (B) Steady-state levels of cytochrome b_{562} and cytochrome b_{562}-WVAAA in periplasmic fractions prepared from *tsp*$^+$ cells (X90) or *tsp*$^-$ cells (KS1000) determined by absorbance spectra. (following page)


A
0.450
0.425
0.400
0.375
absorbance
cytb562
cytb562-WVAAA
0
400
800
1200
time (s)

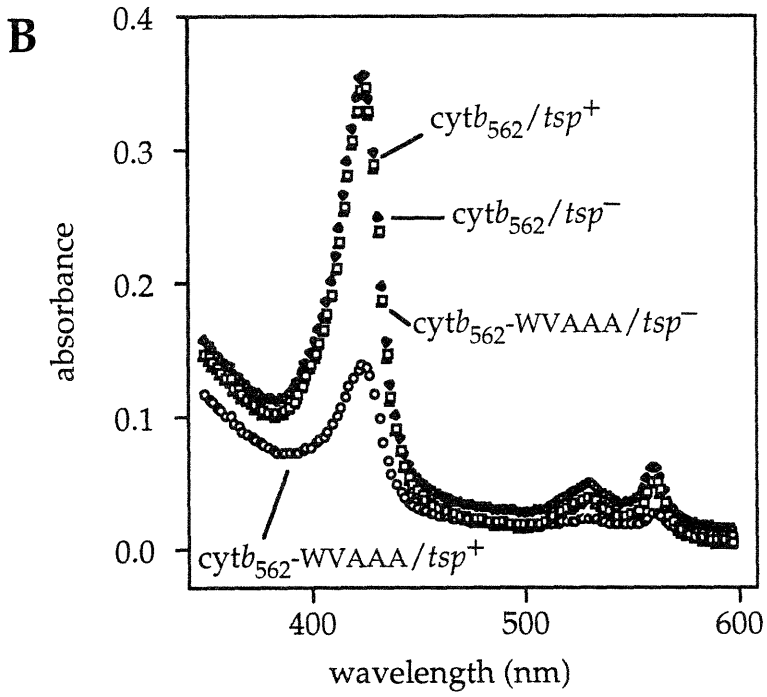
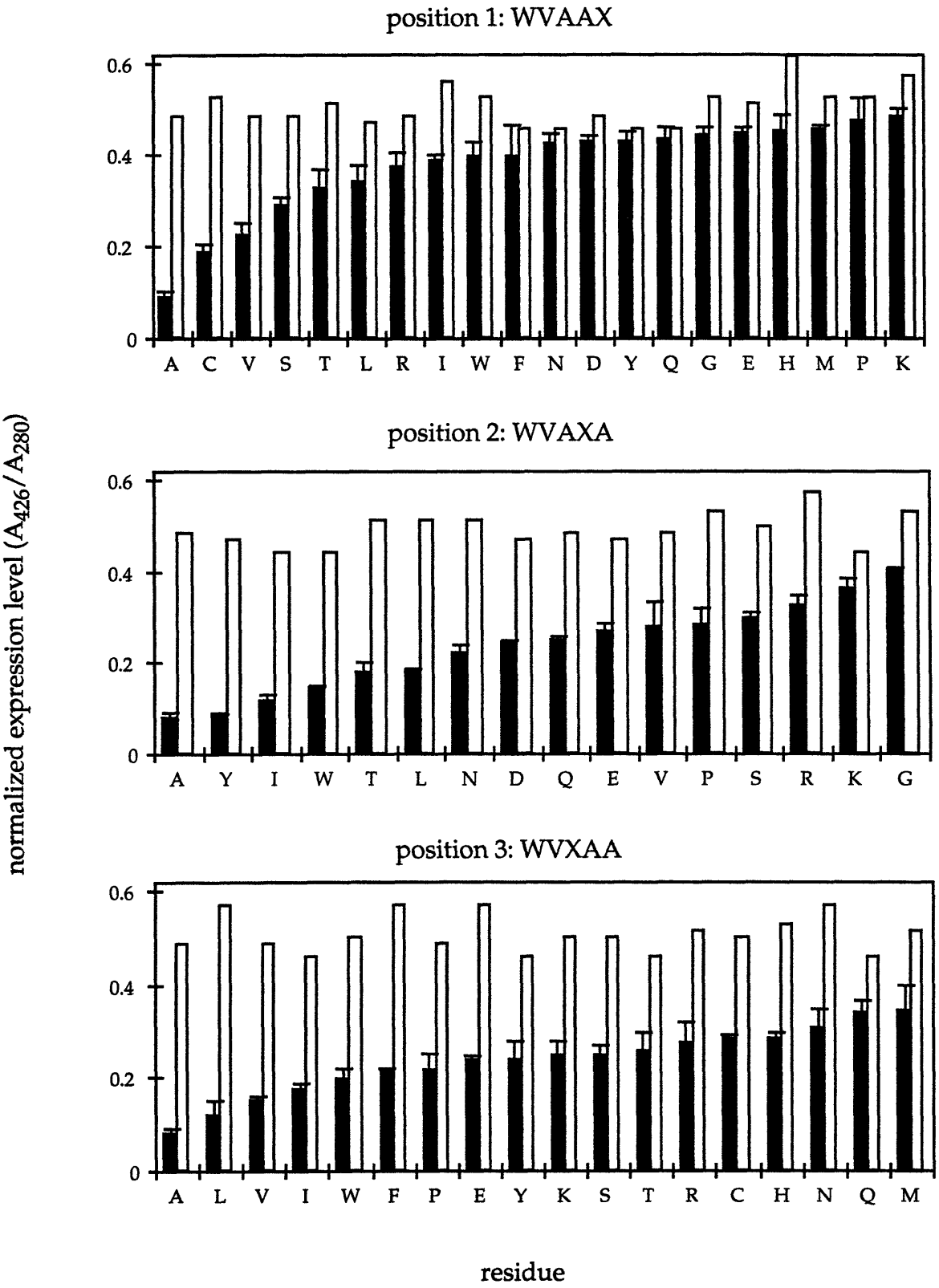

B
0.4
0.3
0.2
0.1
0.0
absorbance
cytb562/tsp+
cytb562/tsp−
cytb562-WVAAA/tsp−
cytb562-WVAAA/tsp+
400
500
600
wavelength (nm)

Fig. 2. Steady-state levels of cytochrome-b_{562} variants in the *tsp*$^+$ strain X90 (filled bars) and the otherwise isogenic *tsp*$^-$ strain KS1000 (open bars). The error bars indicate the standard deviation from the mean for three different X90 cultures. (following page)


position 1: WVAAX
0.6
0.4
0.2
0
A C V S T L R I W F N D Y Q G E H M P K
position 2: WVAXA
0.6
0.4
0.2
0
A Y I W T L N D Q E V P S R K G
position 3: WVXAA
0.6
0.4
0.2
0
A L V I W F P E Y K S T R C H N Q M
normalized expression level (A_{426}/A_{280})
residue

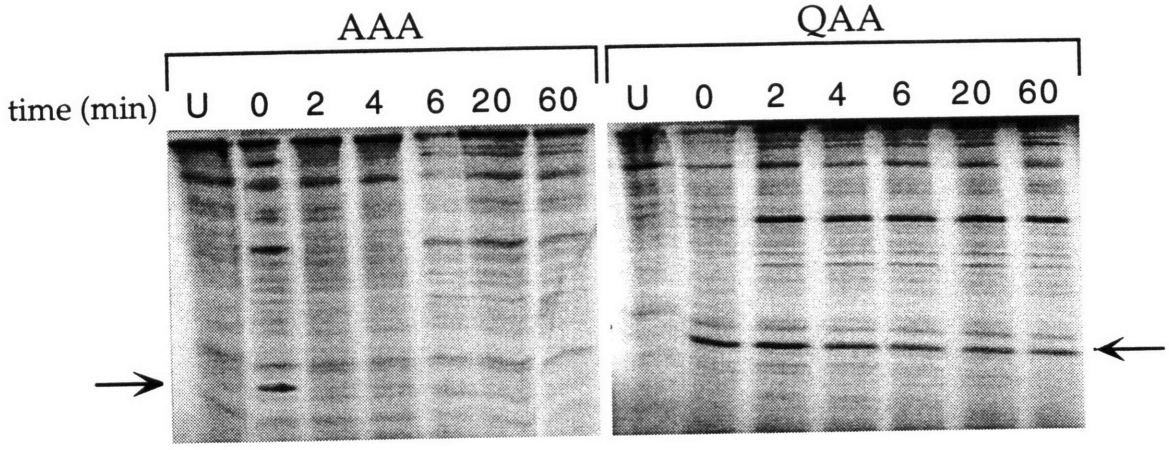


Fig. 3. Pulse-chase assays of the AAA and QAA cytochrome-b_{562} variants. Uninduced controls are shown in the lanes marked U. Arrow indicates the position of cytochrome-b_{562} variants.

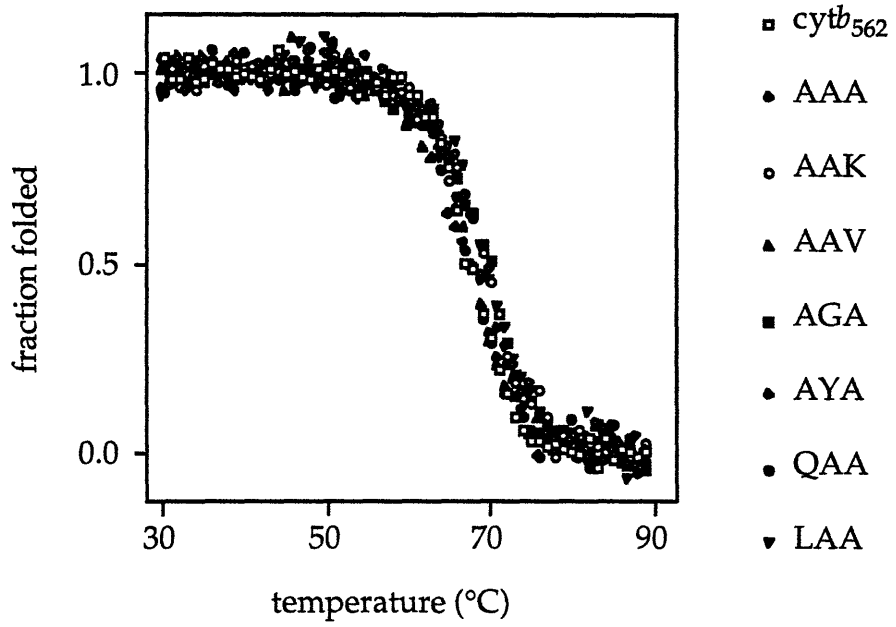


Fig. 4. Thermal denaturation of cytochrome-b_{562} variants. The stability of variants was determined by monitoring CD ellipticity as a function of temperature. The fraction of folded protein was determined by fitting the thermal denaturation curves to a two-state transition between native and denatured protein. Different variants are indicated by their three C-terminal residues.

CHAPTER 5

A Peptide-Tagging System Mediates Degradation of Proteins Synthesized from Damaged mRNA

Proteins translated from mRNAs without stop codons are modified by C-terminal addition of a peptide tag that is encoded by the *ssrA* gene of *E. coli*. These tagged proteins are subsequently degraded by C-terminal specific proteases. The tag appears to be added to the C-terminus of the nascent polypeptide chain by cotranslational switching of the ribosome from the damaged mRNA to *ssrA* RNA.

Cells contain an assortment of intracellular proteases but biological substrates and functions have been established for only a few of these enzymes (1). Even in these cases, it is generally not clear why some proteins but not others are degraded by a particular protease. In studies of recombinant interleukin-6 (IL-6) purified from inclusion bodies in *Escherichia coli*, Tu and colleagues made the surprising discovery that some molecules had been modified by C-terminal truncation at different positions in the sequence and addition of an 11 residue peptide tag (2). The last 10 residues of this peptide tag (AANDENYALAA) were found to be encoded by the *ssrA* gene of *E. coli* and tagging of IL-6 did not occur in *ssrA*$^-$ cells. The product of the *ssrA* gene is a stable 362 nucleotide RNA molecule, which has some tRNA-like properties and can be charged with alanine (3). I noticed that the C-terminal residues of the *ssrA* peptide tag (YALAA) were similar to a C-terminal tail sequence (WVAAA) recognized by Tsp (4), a periplasmic endoprotease, and by a cytoplasmic protease that also degrades proteins in a tail-dependent fashion (5). I reasoned, therefore, that tagging with the *ssrA* peptide might mark proteins for degradation by C-terminal specific proteases.

To test this hypothesis, I constructed genes encoding variants of the N-terminal domain of λ repressor (a cytoplasmic protein) and cytochrome b_{562} (a

periplasmic protein) in which the *ssrA* peptide-tag sequence (AANDENYALAA) was encoded at the DNA level as a C-terminal tail (6). Variants of both proteins were also constructed with a control tag (AANDENYALDD), because both Tsp and its cytoplasmic counterpart fail to cleave proteins with charged residues at the C-terminus (4, 5). In pulse-chase experiments (Fig. 1A, 1B), the λ-repressor and cytochrome b_{562} proteins with the *ssrA* peptide-tag were degraded with half-lives of less than 5 min, while proteins with the control-tag had half-lives greater than 1 h. When the peptide-tagged cytochrome b_{562} protein was expressed in a cell deleted for the Tsp protease, its half-life increased from a few minutes to more than 1 h (Fig. 1C). Tsp is a periplasmic protease and, as expected, degradation of the λ repressor-peptide tag protein in the cytoplasm was not affected in the *tsp*$^-$ strain (Table 1). As a result, it is clear that the presence of the peptide tag at the C-termini of these proteins leads to rapid intracellular degradation in both the cytoplasmic and periplasmic compartments of the cell. In the periplasm, this rapid degradation requires the presence of the Tsp protease.

The rapid degradation of the peptide-tagged cytochrome b_{562} protein in *tsp*$^+$ but not *tsp*$^-$ strains does not establish whether Tsp acts directly or indirectly. To provide evidence for a direct role, the *ssrA* peptide-tagged and control-tagged cytochrome b_{562} proteins were purified and incubated with purified Tsp *in vitro* (7). As shown in Fig. 1D, cleavage of the peptide-tagged cytochrome but not the control-tagged protein is observed, supporting a direct role for Tsp in the periplasmic degradation observed *in vivo*.

By what mechanism is the *ssrA*-encoded tag sequence be added to the C-terminus of a protein? How are certain proteins in the cell chosen for

modification by peptide tagging? Fig. 2 shows a model that addresses both questions and also provides a biological rationale for the system. I assume that certain mRNAs in the cell lack amber, ochre, or opal termination codons because they are truncated by premature termination of transcription or by nuclease cleavage. Translation of such damaged mRNAs presents two potential problems. First, the translated proteins will also be aberrant and may prove deleterious to the cell because of folding defects, uncoupling of enzyme and regulatory domains, etc. Second, the ribosomal factors that normally release the nascent polypeptide chain from the ribosome require the presence of termination codons (8). As a result, the ribosome might be expected to stall or idle upon reaching the 3′ end of the damaged mRNA, leaving the nascent polypeptide chain esterified to the tRNA for the last amino acid. In the model of Fig. 2, I propose: (*a*) alanine-charged *ssrA* RNA recognizes a ribosomal complex stalled at or near the end of an mRNA without a termination codon (Fig. 2, parts I & II); (*b*) the alanine from the *ssrA* RNA is added to the C-terminus of the nascent chain, creating a peptidyl-*ssrA* molecule (Fig. 2, parts II & III); (*c*) translation by the ribosome switches from the 3′ end of the damaged mRNA to the region of the *ssrA* RNA encoding the tag sequence (Fig. 2, parts III & IV; ref. 9); and (*d*) normal termination and release occur at the ochre termination codon following the peptide-tag region of *ssrA* RNA (Fig. 2, parts V & VI). By this model, any protein translated from an mRNA lacking a termination codon will be modified by peptide tagging, which then provides an opportunity for the tagged protein to be recognized and degraded by specialized intracellular proteases. In a sense, the presence of a termination codon provides a signal that the protein product represents a valid translation of an intact mRNA and thus can be released in an unmodified fashion from the ribosome.

To test this model, I cloned a strong, rho-independent transcriptional terminator from the *trp* operon (*trpAt*) before the translational termination codons in the 3′ regions of genes encoding the N-terminal domain of λ repressor and cytochrome b_{562} (10, 11). I reasoned that the transcripts from these genes (λ-repressor-1-93-M2-H_6-*trpAt* and cyt-b_{562}-*trpAt*, respectively) should terminate at the *trpA* terminator and thus, if the model is correct, that the protein product(s) of these prematurely terminated mRNAs would be modified by peptide tagging and then degraded. Fig. 3 shows pulse-chase experiments for the λ-repressor variant. In cells containing *ssrA* RNA, the protein is degraded with a half-life of a few minutes. In otherwise isogenic cells lacking *ssrA*, the protein is both smaller (as expected if peptide tagging does not occur) and is not rapidly degraded (12). The cyt-b_{562}-*trpAt* protein also appears to be degraded in the periplasm of *ssrA*$^+$ cells; in fact no protein is observed following a 30 s pulse even before initiation of a chase (Fig. 4). I believe that this represents extremely rapid degradation rather than a defect in synthesis because the cyt-b_{562}-*trpAt* protein was expressed and relatively long-lived in an otherwise isogenic *tsp*$^-$ strain (Fig. 4). Overall, the results of the experiments with the λ repressor and cytochrome constructs containing the *trpA* terminator confirm several predictions of the model. First, proteins translated from mRNAs lacking termination codons appear to be modified by *ssrA*-dependent peptide tagging. Second, these peptide-tagged proteins are rapidly degraded.

A second cytochrome b_{562} construct (cyt-b_{562}-M2-H_6-*trpAt*) in which the *trp* terminator was separated from the body of the protein by an M2 epitope and His_6 sequence was also constructed (11). In this case, comparison

of protein production in *tsp*$^+$ and *tsp*$^-$ strains suggests that the peptide-tagged protein is not completely degraded in cells containing Tsp (Fig. 4). Rather, the protein appears to be rapidly processed to a smaller form, which is then degraded with a half-life of roughly 15 min. The larger and smaller forms of the cyt-b_{562}-M2-H_6-*trpAt* proteins were purified from *tsp*$^+$ and *tsp*$^-$ strains and characterized by N-terminal sequencing and ion-spray mass spectrometry (13). The protein purified from the *tsp*$^-$ strain had an N-terminal sequence (NH_2–ADLED) that matches that of cytochrome b_{562} (14) but showed some heterogeneity by mass spectrometry. The largest and most abundant form had a mass of 15,637 daltons. This is within 0.06% of the mass (15,647 daltons) expected for a 140-residue protein that contains residues 1-129 from the cyt-b_{562}-M2-H_6-*trpAt* gene followed by the AANDENYALAA peptide tag (Fig. 5b). The protein purified from the *tsp*$^+$ strain had the expected N-terminal sequence and a mass of 13,919 daltons. This is within 0.05% of the mass (13,927 daltons) predicted if the protein consists of residues 1-123 of the gene sequence (Fig. 5b), and suggests that Tsp cleaves the mature protein at an A↓A sequence, thereby removing approximately 15-20 residues.

To confirm the structure proposed above for the peptide-tagged form of the cyt-b_{562}-M2-H_6-*trpAt* protein, I digested this protein and the one purified from the *tsp*$^+$ strain with trypsin and chromatographed the digests on a C18 column (Fig. 5a). Because the peptide tag is expected to contain a tyrosine residue, I monitored the column effluent by absorbance at 280 nm. Compared to the *tsp*$^+$ protein, the *tsp*$^-$ protein has two additional tryptic peptides that contain tyrosine. These difference peptides are labelled 1 and 2 in Fig. 5a. The amino acid sequence of peptide 1, determined by sequential Edman degradation, contains five residues encoded by the *trpA* terminator portion of

the cyt-b_{562}-M2-H_6-*trpAt* gene followed by the AANDENYALAA peptide tag (Fig. 5b). Peptide 2 was found to be a mixture of two sequences containing either three or four *trpA* terminator-encoded residues followed by the 11-residue peptide tag (Fig. 5b). Hence, in both cases, the sequences of the difference peptides correspond to those expected if translation of the cytochrome b_{562} variant ends within the *trpA* terminator and these nascent polypeptides are then modified by addition of alanine and the *ssrA*-encoded peptide-tag sequence (15).

In the model of Fig. 2, peptide tagging occurs cotranslationally. The strongest argument in favor of this mechanism is that it makes biological sense. I have shown that *ssrA*-dependent peptide-tagging occurs when an mRNA lacks a termination codon. Moreover, as required by the model, *ssrA* RNA can be charged with alanine and is associated with ribosomes (3). There are also observations that make it unlikely that peptide tagging occurs pre-translationally by some type of RNA splicing mechanism, or post-translationally by some type of peptide ligation. In their studies of peptide tagging of IL-6, Tu et al. found no evidence for mRNAs containing both IL-6 and *ssrA* peptide tag sequences (2). The presence of the first alanine of the peptide tag, which is neither mRNA nor *ssrA* RNA encoded, is also difficult to explain by pre-translational or post-translational mechanisms. Nevertheless, the mechanism suggested in Fig. 2 is still a working model which needs to be subjected to more rigorous tests, including isolation of the nascent peptidyl-*ssrA* molecule, which is predicted to be a covalent intermediate.

I have interpreted the results presented here as suggesting that proteins synthesized from mRNAs without termination codons are targets for peptide tagging. It is also possible that the secondary structure of the *trpA* terminator causes ribosome stalling and subsequent tagging, although this seems somewhat less likely because the majority of the terminator RNA, including the hairpin, is translated. It is straightforward, nevertheless, to imagine how certain RNA structures or chemical modifications might also lead to an interruption of normal translation, recruitment of *ssrA* RNA, and subsequent peptide tagging.

By themselves, the studies presented here do not show how peptide tagging leads to degradation by endoproteases like Tsp. However, previous studies have shown that the peptide tails, which are recognized by Tsp and its cytoplasmic counterpart, do not result in unfolding, decreased thermodynamic stability, or conformational changes of the protein to which they are attached (4, 5). In the free substrate proteins, these tails are accessible and apparently unstructured. The simplest model is that the tails are recognized and bound to a tethering site on the protease, with subsequent cleavage of the tethered substrate mediated by a separate active site (4, 16). Substrate cleavage by Tsp does not require ATP or other high-energy compounds, and thus any unfolding of the substrate, which is required prior to cleavage, must occur spontaneously or as a consequence of enzyme binding.

I believe that the peptide-tagging system and associated C-terminal specific proteases probably serve to rid cells of deleterious proteins synthesized from incomplete or otherwise damaged mRNAs. *ssrA*$^-$ strains

grow slowly, as might be expected if damaged proteins cannot be efficiently degraded (3, 17). Homologs of *ssrA* RNA, which can be charged with alanine and contain reasonable candidate sequences for peptide tags, have been identified in both gram-negative and gram-positive species (3). Proteases homologous to Tsp have also been found in widely varying bacterial species (16). Both findings suggest that systems involving peptide tagging and degradation by C-terminal specific proteases are likely to be widespread in the bacterial world.

There are obvious parallels between the peptide-tagging system described here and the ubiquitin system of eukaryotes (18). In both cases, substrates are modified by addition of peptide or protein tags which then mediate recognition by specialized proteases. There are also, however, significant differences. In the ubiquitin case, for example, modification occurs via addition to amino groups of an appropriate protein substrate by post-translational mechanisms mediated by specific ubiquitin-conjugating enzymes. It will be interesting to see whether eukaryotic cells also utilize independent peptide-tagging mechanisms, similar to those in bacteria, for eliminating proteins synthesized from damaged mRNAs or whether the spatial separation of transcription and translation in eukaryotic cells allows damaged mRNA to be recognized and degraded before it leaves the nucleus.

NOTES AND REFERENCES

1. S. Gottesman and M. R. Maurizi, *Microbiol. Rev.* **56**, 592 (1992); V. K. Antonov in *Chemistry of Proteolysis*, (Springer-Verlag, Berlin 1993).

2. G.-F. Tu, G.E. Reid, J.-G Zhang, R.L. Moritz and R.J. Simpson, *J. Biol. Chem.* **270**, 9322 (1995).

3. B.K. Ray and D. Apirion, *Molec. Gen. Genet.* **174**, 25 (1979); B.-K. Oh and D. Apirion, *Molec. Gen. Genet.* **229**, 52 (1991); Y. Komine, M. Kitabatake, T. Yokogawa, K. Nishikawa and H. Inokuchi, *Proc. Natl. Acad. Sci. USA* **91**, 9223 (1994); C. Ushida, H. Himeno, T. Watanabe and A. Muto, *Nucleic Acids Res.* **22**, 3392 (1994).

4. K.R. Silber, K.C. Keiler and R.T. Sauer, *Proc. Natl. Acad. Sci. USA* **89**, 295 (1992); K.C. Keiler, K.R. Silber, K.M. Downard, I.A. Papayannopoulos, K. Biemann and R.T. Sauer, *Protein Science* **4**, 1507 (1995).

5. D.A. Parsell, K.R. Silber, and R.T. Sauer, *Genes and Development* **4**, 277 (1990); K.R. Silber and R.T. Sauer, *Mol. Gen. Genet.* **242**, 237 (1994).

6. Plasmids encoding variants of cytochrome b_{562} with a C-terminal AANDENYALAA peptide-tag (pCytbSPT) or a AANDENYALDD control-tag (pCytbCPT) were constructed by ligating oligonucleotide cassettes encoding the C-terminal residues of cytochrome b_{562} followed by the appropriate peptide tail into pNS207, which directs overexpression of wild-type cytochrome b_{562} (14).

7. The *ssrA* peptide-tagged and control-tagged variants of cytochrome b_{562} were purified from IPTG-induced log-phase cultures of the *tsp*$^-$ *E. coli* strains KS1000/pCytbSPT and KS1000/pCytbCPT, respectively (5, 6). Following chloroform lysis, the lysate was acidified to pH 4, insoluble material was removed, and the proteins were purified by chromatography on CM-Acell, Superdex 75, and reverse phase C18 HPLC columns. The final material was > 95% pure as judged by SDS PAGE. Cleavage by Tsp *in vitro* was assayed by incubating 5 μM of the purified cytochrome b_{562} variant with 300 nM of purified Tsp (4) at 37 °C for 30 min and analyzing the reaction on SDS Tris-tricine gels.

8. J.W.B. Hershey in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, H.E. Umbarger and F.C. Neidhardt, Eds., (American Soc. Microbiology, Washington, DC, 1987) pp. 613-647.

9. Precedents exist for the skipping of as many as 50 nucleotides during translation of the T4 gene 60 mRNA. (W.M. Huang, S.Z. Ao, S. Casjens, R. Orlandi, R. Zeikus, R. Weiss, D. Winge and M. Fand, *Science* **239**, 1005 (1988)). This type of hopping from one region of an mRNA to another may be analogous to the ribosome switching proposed here between the 3′ end of the mRNA and the tag sequence of the *ssrA* RNA.

10. P.J. Farnham, T. Platt, *Cell* **20**, 739 (1980); R. Reynolds, R.M. Bermudez-Cruz, M.J. Chamberlin, *J. Mol. Biol.* **224**, 31 (1992).

11. A gene encoding residues 1-93 of λ repressor, followed by the M2 epitope (DYKDDDDK), a His_6 sequence (HHHHHH), and the *trpA* terminator (Fig. 5b) was constructed by PCR and cloned under control of a P_{trc} promoter in a pBR322-based plasmid. Genes containing the entire coding sequence of cytochrome b_{562} (including the signal sequence) followed by the *trpA* terminator or followed by the M2 epitope, His_6 sequence, and *trpA* terminator were constructed in a similar fashion.

12. The observation that significant amounts of stable protein are observed in *ssrA*$^-$ strains for the λ-repressor-1-93-M2-H_6-*trpAt* and cyt-b_{562}-*trpAt* constructs suggests that there must be some mechanism that allows *ssrA*-independent release of the nascent chain from mRNAs without termination codons. This mechanism may not operate in *ssrA*$^+$ strains or may occur at a rate slower than *ssrA*-mediated peptide tagging and release.

13. The cyt b_{562}-M2-H_6-*trpAt* proteins were purified from IPTG-induced cultures of *E. coli* strain X90 (*tsp*$^+$) or KS1000 (*tsp*$^-$) transformed with appropriate plasmids. Cells were lysed in 6 M guanidine-HCl by sonication, the insoluble fraction was removed by centrifugation, and the supernatant was applied to a 5 ml Ni-NTA agarose column. Bound proteins were eluted in buffer containing 250 mM imidazole, and chromatographed on a C18 reverse phase HPLC column. Fractions containing cyt b_{562}-M2-H_6-*trpAt* of greater than 95% purity by SDS PAGE were pooled, dried under vacuum, and resuspended in 10 mM Tris (pH 8.0), 20 mM KCl. Approximately 5 μg of purified protein was incubated with 0.1 μg TPCK-trypsin at 37 °C for 30 min, and the products were

separated on a C18 reverse phase HPLC column using a water /acetonitrile gradient in 0.1% TFA. N-terminal sequencing by sequential Edman degradation was performed at the MIT Biopolymers Facility, and masses were determined by electrospray mass spectrometry at the Harvard Microchemistry Facility.

14. H. Nikkila, R.B. Gennis and S.G. Sligar, *Eur. J. Biochem.* **202**, 309 (1991).

15. As shown in Fig. 5b, at least three different peptide-tagged forms of the cyt-b_{562}-M2-H_6-*trpAt* protein were purified from the *tsp*$^-$ strain. These differ in having three, four, or five *trpA* terminator-encoded residues. Likewise, numerous different peptide-tagged forms of IL-6 were identified by Tu et al. and size heterogeneity of the IL-6 mRNA was also observed (2). By our model, mRNA heterogeneity leads naturally to heterogeneity of the tagged protein products. This mRNA heterogeneity could arise from heterogeneity in transcriptional termination or by nuclease cleavage of full length mRNA (R. Reynolds, R. M. Bermudez-Cruz and M. J. Chamberlin, *J. Mol. Biol.* **224**, 31 (1992); E. Hajnsdorf, O. Steier, L. Coscoy, L. Teysset and P. Regnier, *EMBO J.* **13**, 3368 (1994); H. Causton, B. Py, R. S. McLaren and C. F. Higgins, *Mol. Microbiology* **14**, 731 (1994)).

16. K.C. Keiler and R.T. Sauer, *J. Biol. Chem.* (in press).

17. In addition to slow growth, *ssrA*$^-$ mutants have phenotypes that involve changes in intracellular proteolysis (21) and alterations in the lysis-lysogeny decisions of temperate phages (D. M. Retallack, L. J. Johnson

and D. I. Friedman, *J. Bacteriol.* **176**, 2082 (1994)). It is presently unclear whether these phenotypes arise either directly or indirectly from loss of the peptide-tagging activity of *ssrA* RNA with subsequent changes in targeted degradation, or are caused by other defects associated with loss of *ssrA* RNA.

18. A. Ciechanover, *Cell* **79**, 13 (1994); M. Hochstrasser, *Curr. Opin. Cell Biol.* **7**, 215 (1995); S. Jentsch and S. Schlenker, *Cell* **82**, 881 (1995).

19. For pulse-chase experiments, cells were grown to mid log-phase at 37 °C in M9 minimal medium (no Met or Cys). Protein expression was induced by addition of 1 mM IPTG and after 20 min, a labelling pulse of 100 μCi ^{35}S-methionine was added for 30 s (Figs. 1B, 1C, 3, 4) or 120 s (Fig. 1A). After this time, L-methionine was added to a final concentration of 1.4 mg/ml. At different times, 0.5 ml aliquots were removed and immediately lysed by boiling in SDS sample loading buffer (Fig. 1A, 3, 4), or alternatively were frozen in an ethanol-dry ice bath (Fig. 1B, 1C) and then lysed by 3 cycles of thawing at 4 °C followed by refreezing. Samples were electrophoresed on SDS polyacrylamide gels and the radiolabeled bands were visualized by autoradiography and/or by phosphorimaging. Half-lives were calculated by nonlinear least-squares fitting of data to an exponential decay. For these calculations, the intensities of the induced bands were integrated using the phosphorimager ImageQuant software and were normalized to a set of stable bands.

20. Strain X90 *ssrA1*::cat was constructed by P1 transduction of the *ssrA1*::cat allele from *E. coli* strain JK6257 (21) into *E. coli* strain X90.

21. J.E. Kirby, J.E. Trempy and S. Gottesman, *J. Bacteriol.* **176**, 2068, (1994).

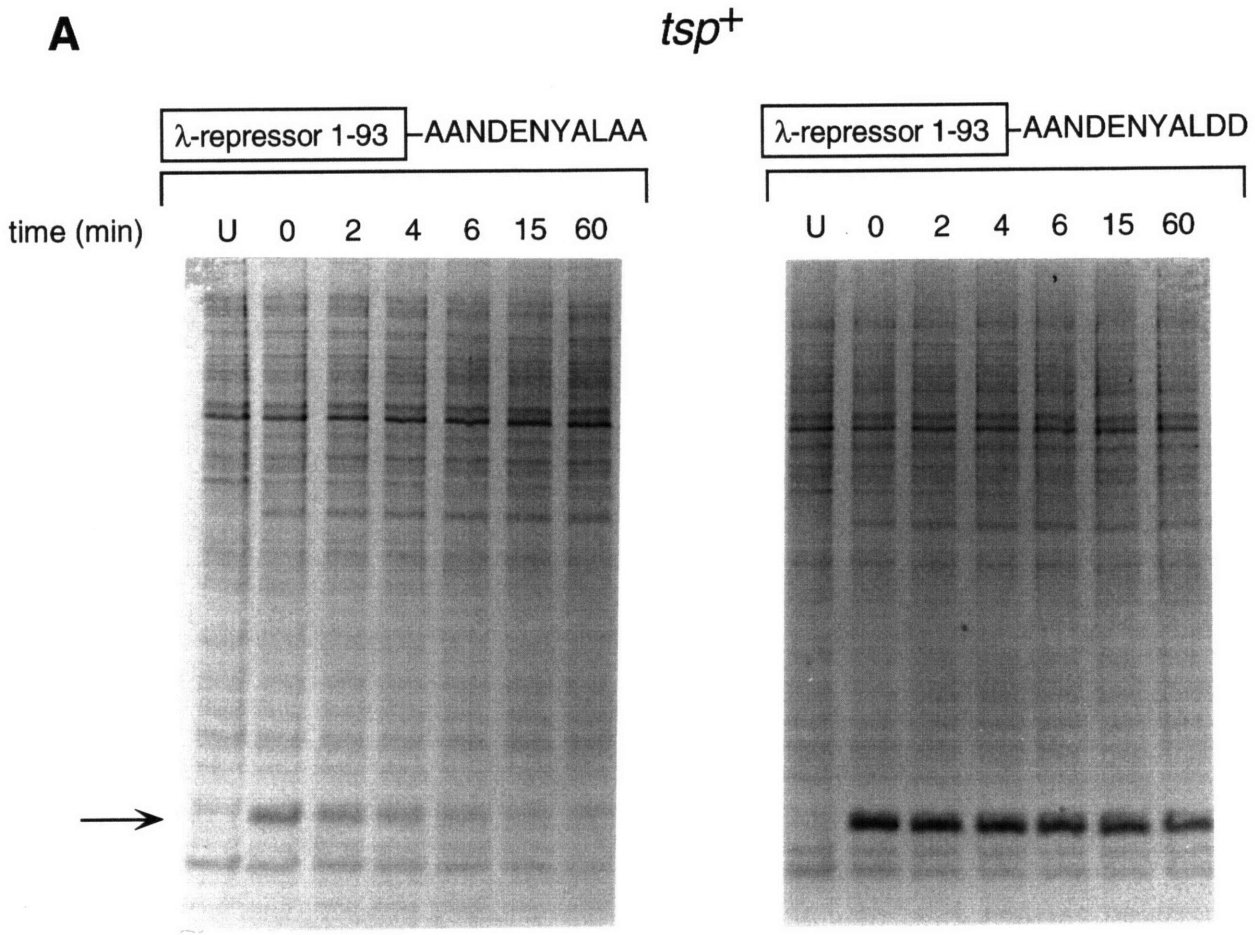
22. I thank Tania Baker, David Bartel, Alan Grossman, Chris Kaiser, and Tom RajBhandary for advice and comments; Susan Gottesman, Michael Hecht, and Steve Sligar for strains; Sarah Ades, Brenda Shulman, the MIT Biopolymers Lab, and the Harvard Microchemistry Facility for help. Supported by NIH grants AI-15706 and AI-16892.

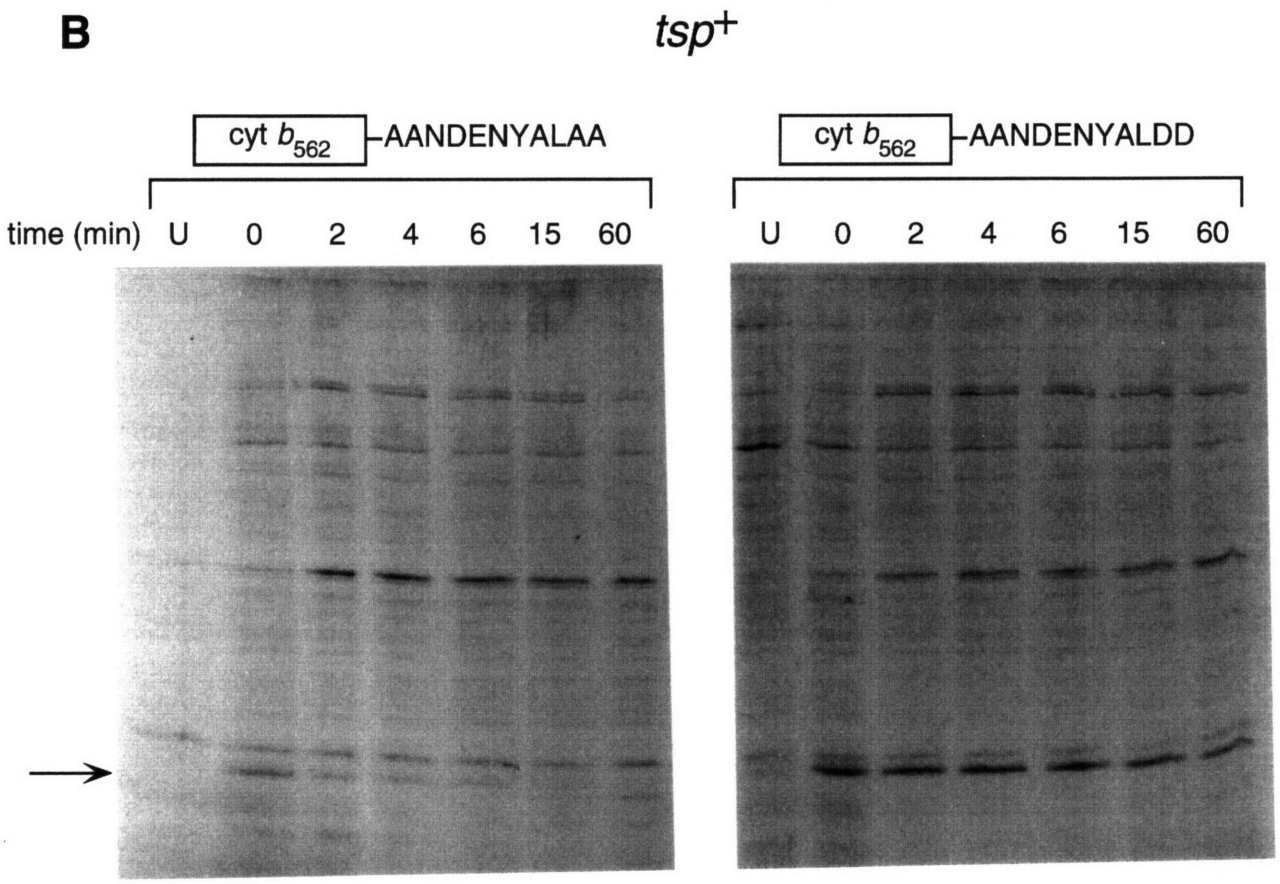
Table 1. Half-lives of the λ-repressor and cytochrome b_{562} constructs in *tsp*$^+$ *ssrA*$^+$ (X90), *tsp*$^-$ *ssrA*$^+$ (KS1000), and *tsp*$^+$ *ssrA*$^-$ (X90 *ssrA1*::*cat*) strains.

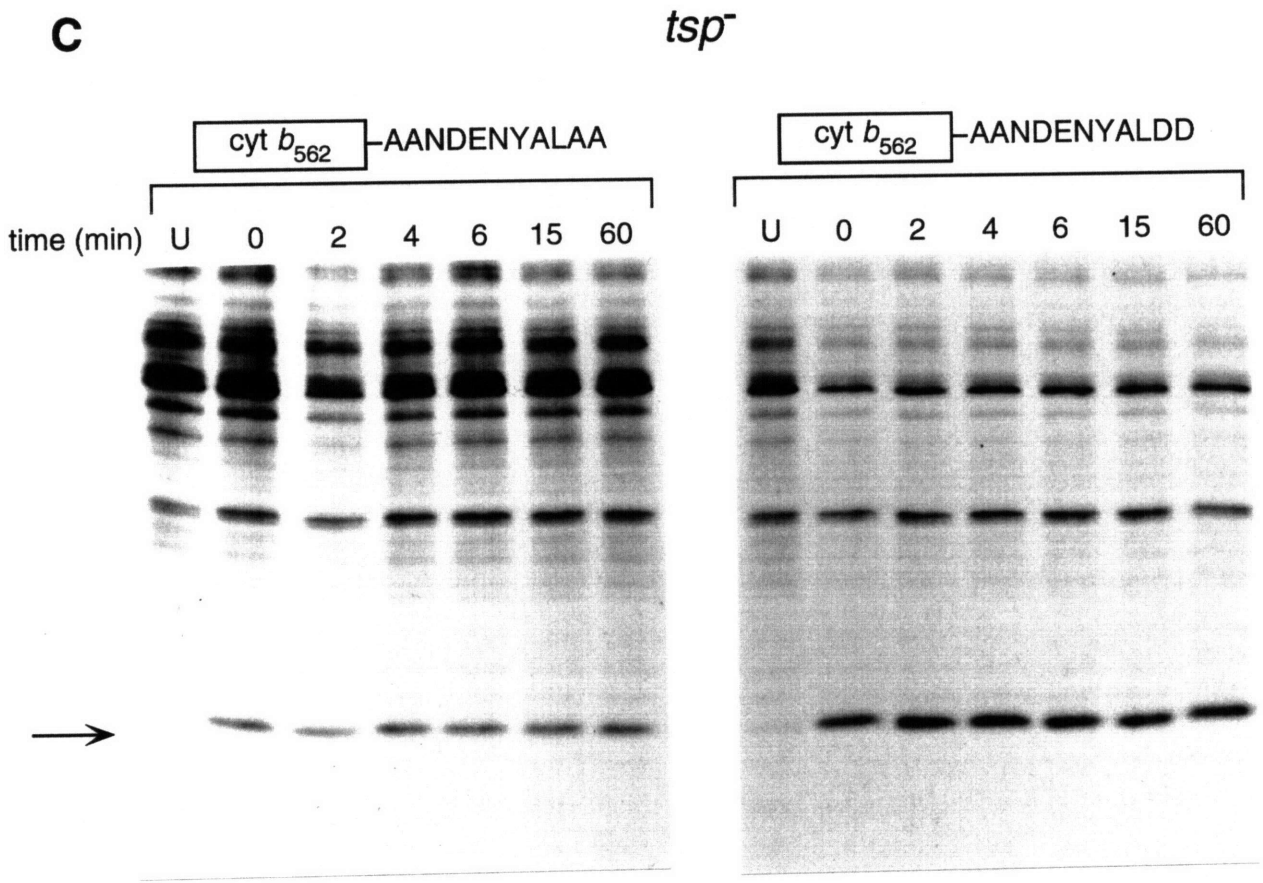
Protein constructs	Strain background		
	tsp $^+$ *ssrA* $^+$	*tsp* $^-$ *ssrA* $^+$	*tsp* $^+$ *ssrA* $^-$
λ-repressor 1-93 – AANDENYALAA	4 min	4 min	
λ-repressor 1-93 – AANDENYALDD	> 60 min	> 60 min	
λ-repressor 1-93 – M2 – H_6 – *trpAt*	2 min		> 60 min
cyt b_{562} – AANDENYALAA	4 min	> 60 min	
cyt b_{562} – AANDENYALDD	> 60 min	> 60 min	
cyt b_{562} – *trpAt*	< 30 s	~ 30 min	
cyt b_{562} – M2 – H_6 – *trpAt*	< 30 s *	~ 40 min	~ 60 min

* half-life shown refers to full length protein, a processed form of this protein has a half-life of ~ 15 min

Fig. 1 Degradation of variants of the N-terminal domain of λ-repressor or cytochrome b_{562} with the *ssrA* peptide tag (AANDENYALAA) or a control tag (AANDENYALDD). (**A**) Pulse-chase assays (19) for the λ-repressor variants in the tsp^+ *E. coli* strain X90 (5). Arrows indicate the induced protein. The lane marked U represents an uninduced control. (**B**) Pulse-chase assays of cytochrome b_{562} variants in X90. (**C**) Pulse-chase assays of the cytochrome b_{562} variants in the tsp^- strain KS1000 (5). (**D**) Degradation of the purified cytochrome b_{562}–AANDENYALAA variant but not the AANDENYALDD variants by Tsp *in vitro* (7). (following pages)


A
tsp+
λ-repressor 1-93
AANDENYALAA
λ-repressor 1-93
AANDENYALDD
time (min)
U 0 2 4 6 15 60
U 0 2 4 6 15 60


B
tsp+
cyt b562
-AANDENYALAA
cyt b562
-AANDENYALDD
time (min)
U
0
2
4
6
15
60
U
0
2
4
6
15
60


C
tsp-
cyt b562
-AANDENYALAA
cyt b562
-AANDENYALDD
time (min)
U
0
2
4
6
15
60
U
0
2
4
6
15
60

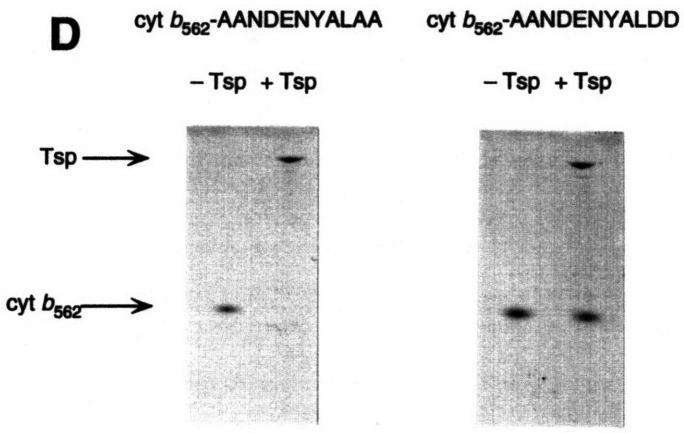

D
cyt b_{562}-AANDENYALAA
cyt b_{562}-AANDENYALDD
– Tsp + Tsp
– Tsp + Tsp
Tsp
cyt b_{562}

Fig. 2 Model for *ssrA* RNA-mediated peptide tagging of proteins synthesized from mRNA transcripts without termination codons. (**I**) The ribosome reaches the 3′ end of the mRNA, leaving the nascent polypeptide esterified to the tRNA in the A site. (**II**) The stalled complex is recognized by alanine-charged *ssrA* RNA, which binds in the P site. (**III**) Alanine from the *ssrA* RNA is added to the nascent polypeptide chain and the ribosome switches translation to the region of *ssrA* RNA encoding the peptide tag. (**IV**) The *ssrA* peptide tag is added cotranslationally to the growing nascent chain. (**V**) The *ssrA* peptide-tagged protein is released from the ribosome when the ochre termination codon of *ssrA* RNA is reached. (**VI**) The *ssrA* peptide-tagged protein can be recognized and degraded by specialized C-terminal specific intracellular proteases. (following page)

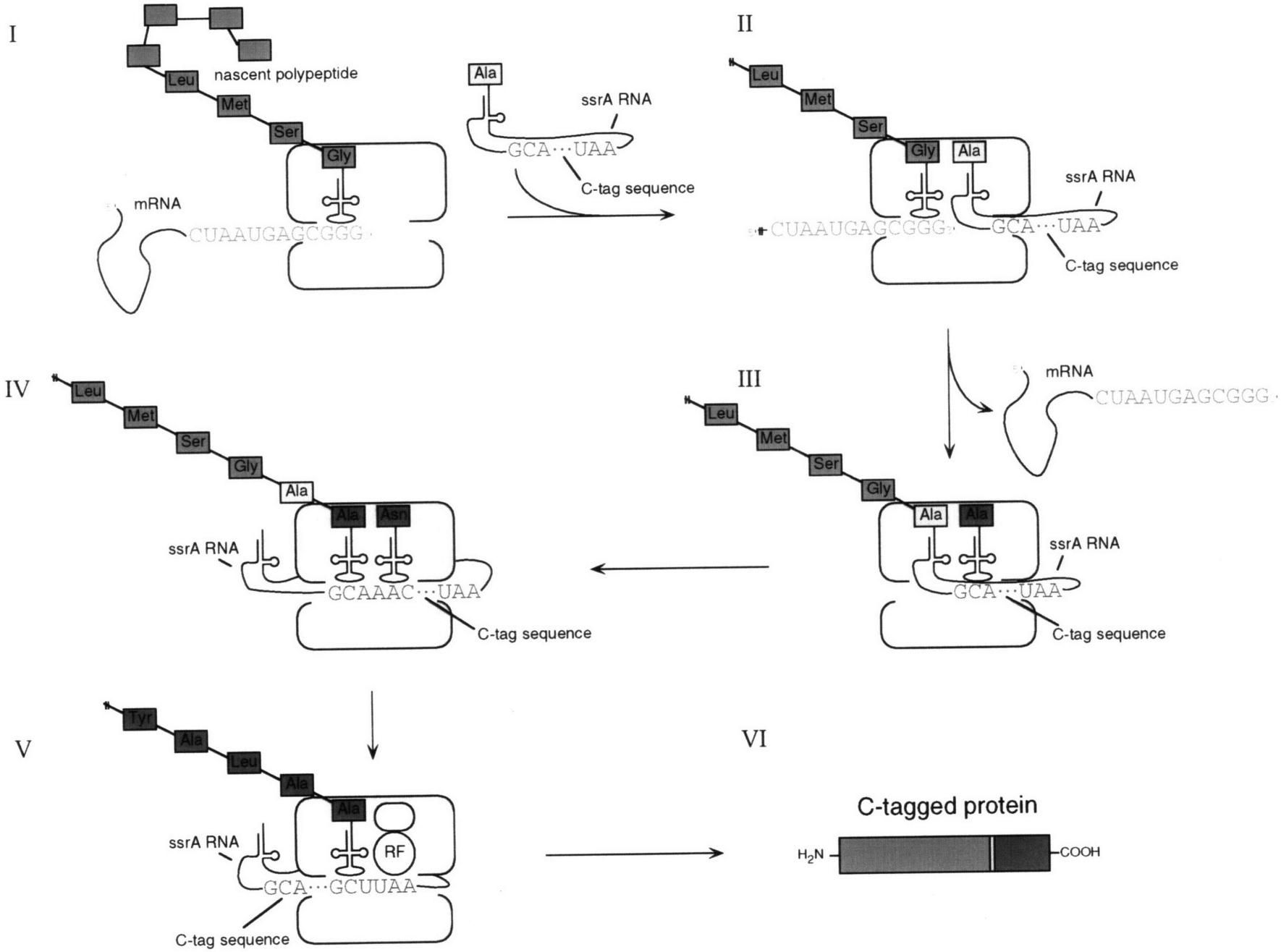

I
nascent polypeptide
Leu
Met
Ser
Gly
mRNA
CUAAUGAGCGGG
Ala
ssrA RNA
GCA···UAA
C-tag sequence
II
Leu
Met
Ser
Gly
Ala
ssrA RNA
CUAAUGAGCGGG
GCA···UAA
C-tag sequence
III
mRNA
CUAAUGAGCGGG
Leu
Met
Ser
Gly
Ala
Ala
ssrA RNA
GCA···UAA
C-tag sequence
IV
Leu
Met
Ser
Gly
Ala
Ala
Asn
ssrA RNA
GCAAAC···UAA
C-tag sequence
V
Tyr
Ala
Leu
Ala
Ala
RF
ssrA RNA
GCA···GCUUAA
C-tag sequence
VI
C-tagged protein
H2N
COOH

Fig. 3 Degradation of λ-repressor-1-93-M2-H_6-*trpAt* protein in *ssrA*$^+$ (X90) and *ssrA*$^-$ strains (X90 *ssrA1*::cat; ref. 20) assayed by pulse-chase experiments. The half-life of the protein in the *ssrA*$^+$ strain is 2 min. The half-life in the *ssrA*$^-$ strain is greater than 60 min. (following page)

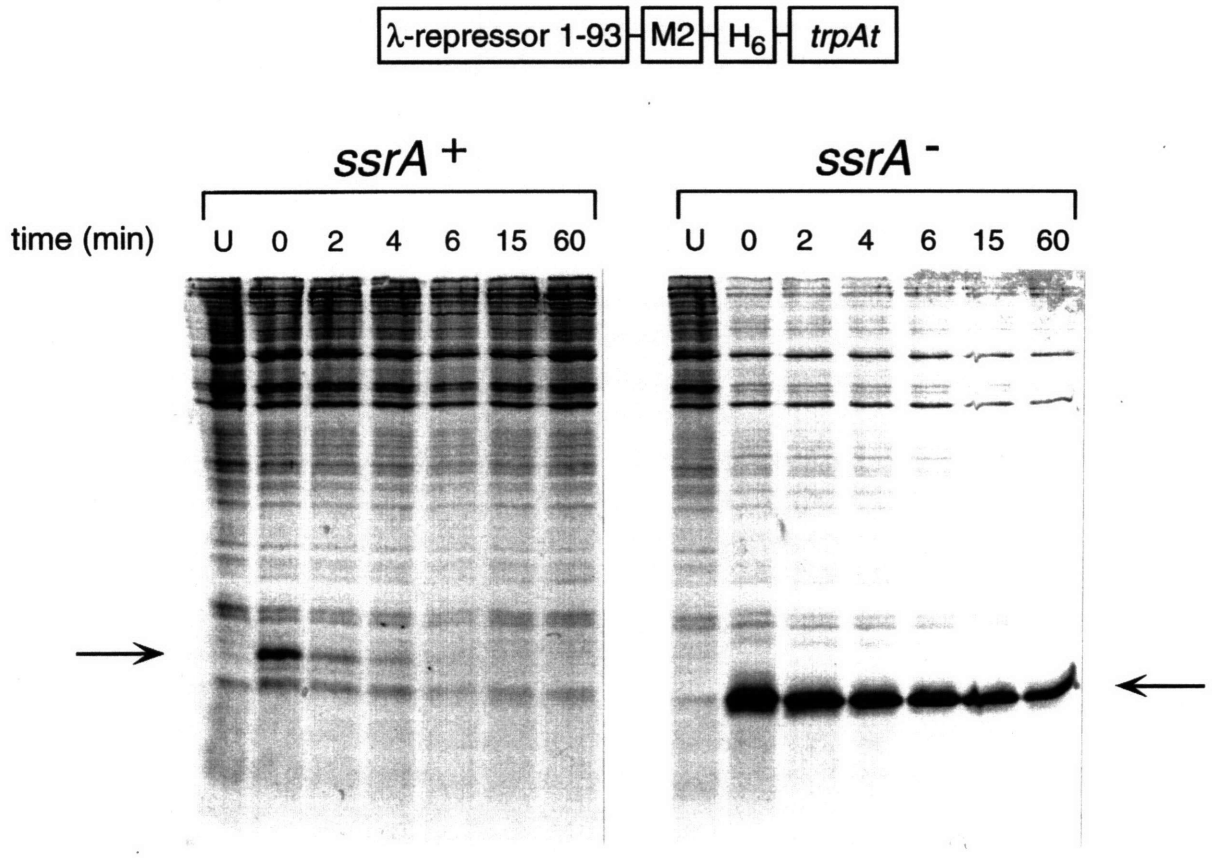

λ-repressor 1-93
M2
H6
trpAt
ssrA +
ssrA -
time (min)
U 0 2 4 6 15 60
U 0 2 4 6 15 60

Fig. 4 Expression of cyt-b_{562}-*trpAt* and cyt-b_{562}-M2-H_6-*trpAt* proteins in *tsp*$^+$ (X90) and *tsp*$^-$ (KS1000) strains following a 30 s ^{35}S-labelling pulse and 0 or 60 min of cold chase. Note that the degradation or processing appears to be sufficiently fast in the *tsp*$^+$ strain that no full length protein is observed at the 0 time point. (following page)

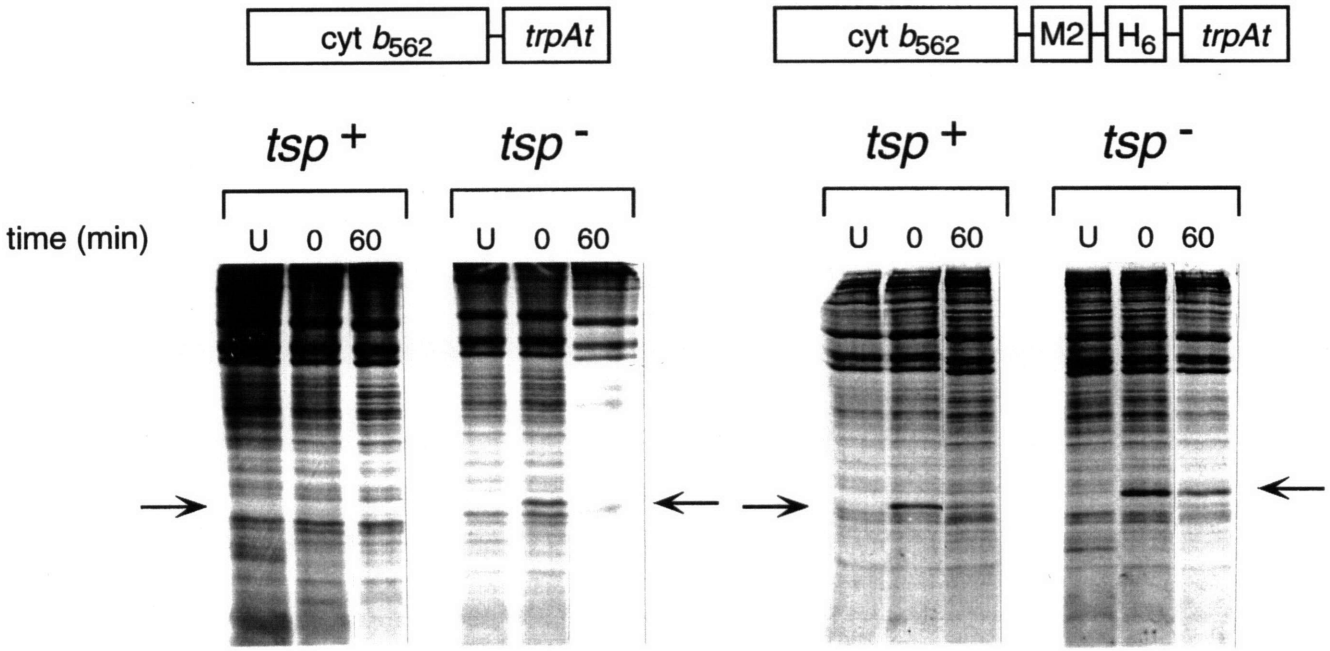
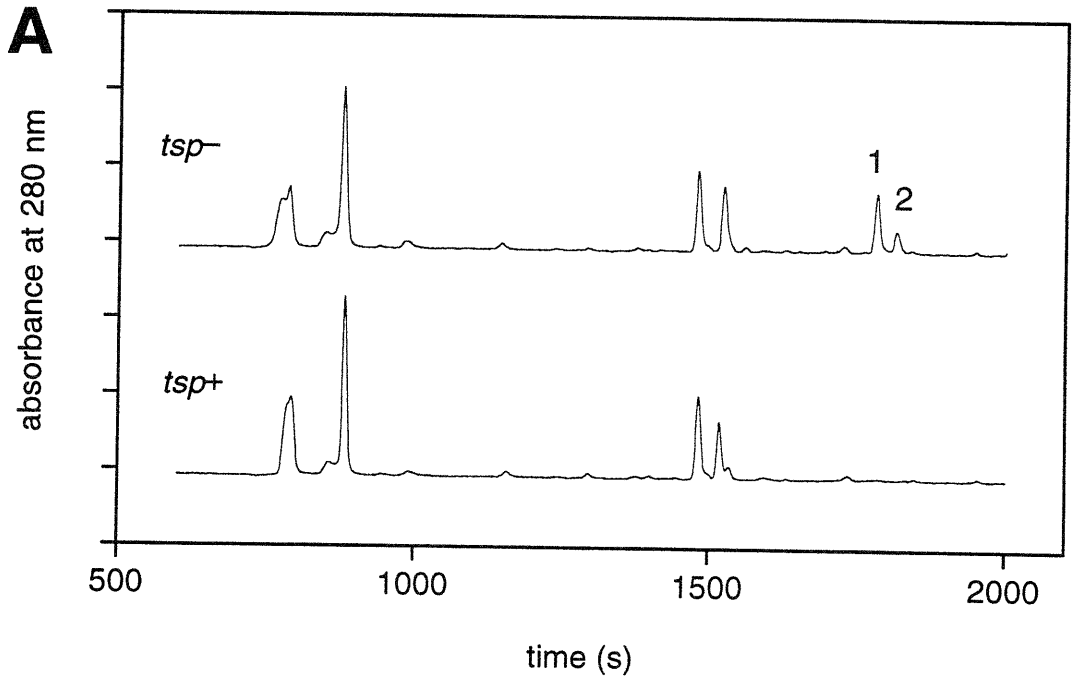
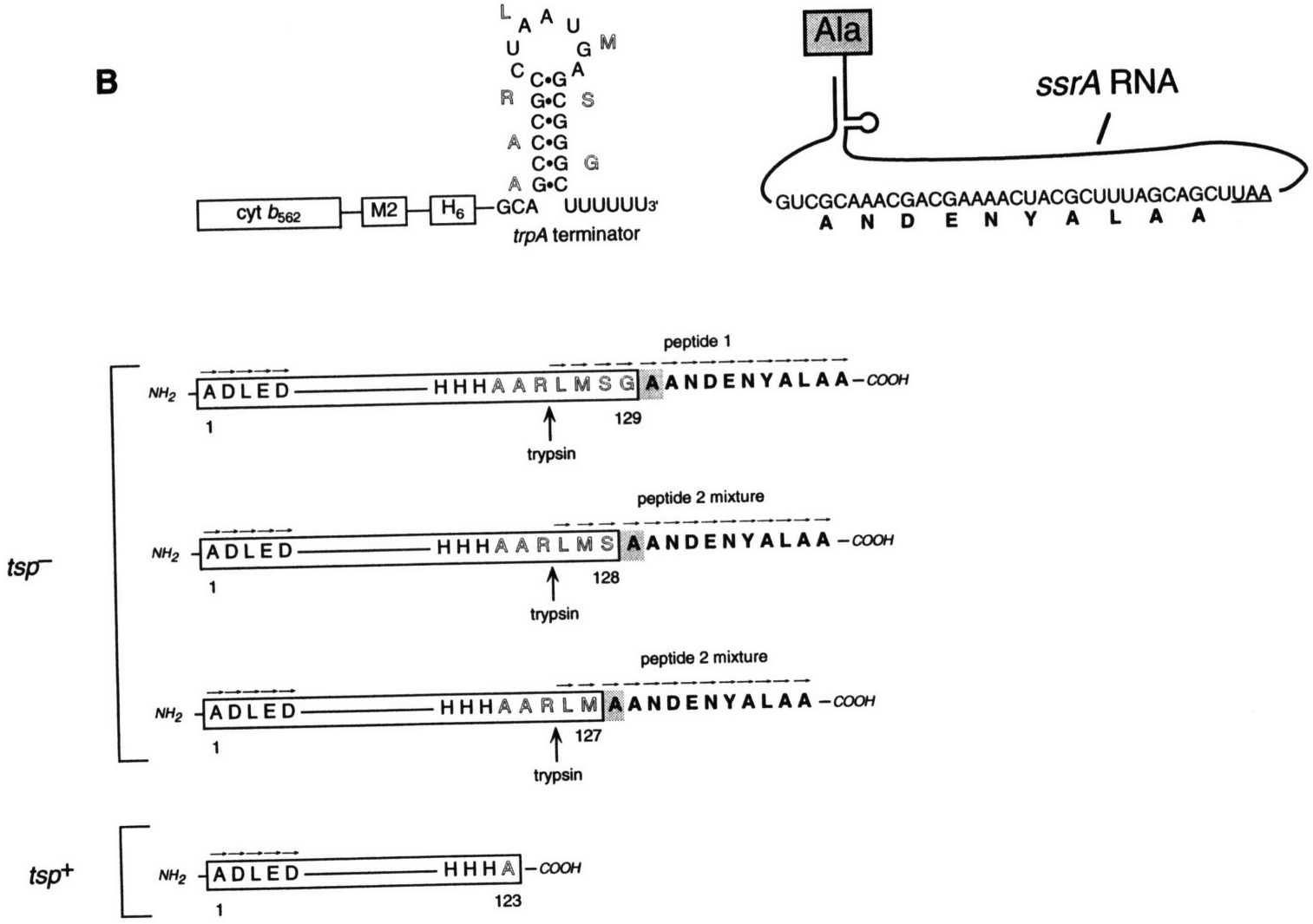

cyt b_{562}
trpAt
cyt b_{562}
M2
H_6
trpAt
tsp +
tsp -
tsp +
tsp -
time (min)
U 0 60
U 0 60
U 0 60
U 0 60

Fig. 5 Sequence characterization of the cyt b_{562}-M2-H_6-*trpAt* proteins purified from *tsp*$^+$ (X90) or *tsp*$^-$ (KS1000) cells. (**A**) Reverse-phase HPLC separation of tryptic digests monitored by absorbance at 280 nm. (**B**) *top* Diagram of the mRNA expected if transcription of the cyt-b_{562}-M2-H_6-*trpAt* gene terminates at the *trpA* terminator and amino acids encoded by the terminator (shown in outline). Diagram of alanine-charged *ssrA* RNA and the sequence of the encoded peptide tag (shown in bold). *bottom* Structures deduced for purified proteins based on sequencing of the intact proteins and the difference tryptic peptides, and the masses determined by ion spray mass spectrometry. Residues encoded by the cytochrome b_{562}, M2, His_6 or *trpAt* portions of the gene are boxed. Residues encoded by *trpAt* are shown in outline. The alanine esterified to the 3′ end of *ssrA* RNA is shaded. Residues from the peptide-tag coding region of *ssrA* RNA are shown in bold. Sequences determined by sequential Edman degradation are indicated by arrows. (following page)


A
absorbance at 280 nm
tsp–
tsp+
1
2
500
1000
1500
2000
time (s)


B
cyt b_{562}
M2
H_6
GCA UUUUUU3'
trpA terminator
Ala
ssrA RNA
GUCGCAAACGACGAAAACUACGCUUUAGCAGCUUAA
A N D E N Y A L A A
peptide 1
NH_2
ADLED
HHHAARLMSG
AANDENYALAA
COOH
1
129
trypsin
peptide 2 mixture
tsp⁻
NH_2
ADLED
HHHAARLMS
AANDENYALAA
COOH
1
128
trypsin
peptide 2 mixture
NH_2
ADLED
HHHAARLM
AANDENYALAA
COOH
1
127
trypsin
tsp+
NH_2
ADLED
HHHA
COOH
1
123