7)

DegP Related Proteases in *Escherichia coli* and Carboxyl-Terminal Tagging of Proteins for Degradation

by

Patrick Robert Henson Waller

B.A., Genetics
Cambridge University
June, 1988

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
February, 1996

Signature of Author Department of Biology

Certified by Robert T. Sauer, Thesis Supervisor

Accepted by Frank Solomon, Chairman, Biology Graduate Committee

DegP Related Proteases in *Escherichia coli* and Carboxyl-Terminal Tagging of Proteins for Degradation

by

Patrick Robert Henson Waller

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

Abstract

The work in the first part of this thesis presents the identification and characterization of *degQ* and *degS*, two *Escherichia coli* genes encoding homologs of the periplasmic protease DegP. The DegQ protein is shown to have the properties of a periplasmic serine endoprotease *in vitro*, and a plasmid expressing *degQ* rescues the temperature sensitive phenotype of a *degP* deletion strain. A plasmid expressing *degS* does not rescue this temperature sensitive phenotype. Deletions in the *degQ* gene cause no obvious growth defect, while those in the *degS* gene result in a small-colony phenotype which is only rescued by plasmids expressing *degS*. These results suggest that DegQ and DegS are functionally different, whereas DegQ and DegP are functionally similar. Biochemical characterization of the cleavage specificities of DegP and DegQ suggest that they both preferentially degrade transiently denatured or unfolded proteins, and cleave substrates after aliphatic, β-branched residues, which are typically buried in the hydrophobic cores of proteins.

The work in the second part of this thesis describes an *E. coli* 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.

Thesis Supervisor: Robert T. Sauer
Title: Whitehead Professor of Biochemistry

ACKNOWLEDGEMENTS

I thank all the members of the Sauer lab for their advice, support, and camaraderie. I especially thank Bob Sauer for encouraging me to pursue my interests in protein degradation. I also thank Harald Kolmar for his energetic pursuit of the biochemical properties of DegP during the collaborative work described in Chapter 2. I thank Ken Keiler for his insightful ideas that led to the collaborative work on peptide tagging described in Chapter 3.

I thank Alex Varshavsky and the members of his lab for their enthusiasm during my early years at M.I.T. They initiated me into the interesting and challenging world of protein degradation.

I thank the members of my thesis committee, Peter Kim, Tom Rajbhandary, Alan Grossman and Fred Goldberg for their interest in the various aspects of this work.

Finally, I thank Maria for her love and support throughout my graduate student years.

Table of Contents

List of Tables

List of Figures

Chapter 2

Chapter 3

Chapter 4

CHAPTER 1

Substrate Selectivity of Intracellular Protein Degradation

INTRODUCTION

Intracellular proteolysis is important in most organisms. The selective degradation or proteolytic cleavage of specific protein substrates is necessary for many physiological processes. Degradation of regulatory proteins allows for rapid changes in their steady state levels. The importance of such changes is illustrated by the requirement for cyclin degradation during eukaryotic cell cycle progression (Seufert et al., 1995). In *Escherichia coli*, an important growth regulator, the SulA protein, is also short-lived (Mizuawa and Gottesman, 1983). The degradation of key enzymes in metabolic pathways, such as mammalian HMG-CoA reductase (Chun et al., 1990), is essential to the regulation of intracellular metabolism. Some developmental proteins, such as the *ftz* protein of *Drosophila melanogaster*, are rapidly degraded in functionally important reactions (Kellerman et al., 1990). Proteolysis also modulates the levels of proteins involved in responses to environmental cues, such as the MATα2 repressor of *Saccharomyces cerevisiae*, the σ^{32} subunit of *E. coli* RNA polymerase, and the λ CII protein (Hochstrasser and Varshavsky, 1990; Tilly et al. 1989; Banuett et al., 1986).

In addition, cells must be able to eliminate proteins that are damaged. Damaged proteins are a waste of potentially useful metabolites and, in some cases, may be toxic due to altered activity or inappropriate interactions with other cellular components. Damage leading to severe structural and functional problems may result from errors during normal metabolic processes, such as misincorporation of rNTPs or amino acids during transcription or translation, premature termination of transcription or translation. Post-translational modifications, such as oxidation, deamidation, alkylation or glycosylation could

also be detrimental. Another potential source of protein damage is exposure to environmental stresses such as elevated temperatures, or extremes in pH.

A multitude of proteases carry out different degradative functions in the cell. In *E. coli*, for example, over 20 different proteases have been identified (Maurizi, 1992). Some of these enzymes, such as the *E. coli* Lon protease or the eukaryotic proteasome, are responsible for both the degradation of specific regulatory proteins and the selective degradation of damaged proteins (Gottesman and Maurizi, 1992; Ciechanover, 1994). The general features of both substrates and proteases, which are responsible for the selectivity of intracellular protein degradation, are the subject of the first part of this chapter. Two proteolytic sytems of *E. coli* are then discussed in greater detail and are the subject of the work presented in subsequent chapters.

SUBSTRATE SELECTIVITY

Cleavage site specificity. The primary amino acid sequence can be an important determinant of cleavage sites for certain proteases, because side chains that are immediately adjacent to the scissile bond often interact with binding sites in or near the protease active site. Chymotrypsin, for example, has a preference for aromatic residues at the P1 position (immediately N-terminal to the scissile bond), which are bound in a deep nonpolar cleft at the active site (. The P1 site preferences of proteases such as chymotrypsin and trypsin (strong preference for lysine and arginine at the P1 position) results in cleavage of specific peptide bonds, but are not generally useful for degradation of specific protein substrates, because almost all proteins will contain aromatic or positively charged residues.

However, significant substrate selectivity can be achieved if a protease requires several specific residues surrounding the scissile bond. The *Bacillus megaterium* Gpr protease, for example, cleaves substrates at a seven residue recognition sequence TE(F,I)ASEF (Setlow, 1988). This sequence would occur by chance less than once every 10^8 peptide bonds, ensuring that only those proteins containing the recognition motif are cleaved. The Lon protease of *E. coli* shows some preference for residues surrounding the scissile bond (Gotterman and Maurizi, 1992), where cleavage studies have identified the consensus sequence $\Phi X_{3-4}LS(L,X)X_5SX\Phi$ (Φ represents any hydrophobic residue). Such sequences would be expected in only about 1% of proteins. However, the poorly defined nature of this consensus sequence suggests that closely related sequences could also be cleaved to some degree. It is likely that the specificity of Lon for substrates such as SulA and RcsA does not reside uniquely in cleavage site recognition. Furthermore, the role of Lon in the selective degradation of misfolded proteins indicates that Lon is capable of cleaving peptide bonds found in a number of different proteins.

Secondary structure can also affect the cleavage of specific peptide bonds. Xenopus magaininase, for example, cleaves magainin peptides on the amino-terminal side of a specific lysine residue (Resnick et al., 1991. This cleavage specificity is largely due to the structural context of the Xaa-Lys bond. Cleavage occurs only when the bond is located on the hydrophilic face of an amphipathic α-helix. Deletions or insertions that disrupt the amphipathic nature of the helix abolish cleavage. In addition, the Xaa-Lys bond is not cleaved if it is moved to the hydrophobic face of the helix. Similarly, cleavage of the poliovirus polyprotein, by protease 3C, only occurs at Gln-Gly-Pro bonds located within a turn or a bend (Ypma-Wong et al., 1988). The basis for this specificity is revealed

by the structure of the rhinovirus 3C protease. The active site is located within a crooked cleft into which the Gln-Gly bond must fit for cleavage to occur (Matthews et al., 1994).

The combination of primary and secondary structural requirements for peptide bonds to bind the active site of a protease provides the basis for a wide range of substrate selectivity. Proteases such as chymotrypsin, which only require a hydrophobic residue, are not very selective. Enzymes such as protease 3C only cleave a Gln-Gly-Pro motif in a particular conformation and are quite selective. Other proteases have intermediate substrate selectivities based on their acitve-site specificities.

Degradation signals. Many substrates are targetted for proteolysis by specific degradation signals (particular sequence and structural motifs) that are often distant from the cleavage sites. Specific proteases, or regulatory proteins associated with these proteases, recognize these different degradation signals.

In *E. coli*, the identity of both the amino and carboxyl termini of a protein can be important determinants of its proteolytic susceptibility. The identity of the amino-terminal residue affects the degradation of test substrates such as β-galactosidase (Tobias et al., 1991). An amino-terminal Arg, Lys, Phe, Leu, Trp or Tyr leads to rapid degradation by the Clp protease. The ClpA subunit is required and may be involved in the recognition process. There is no strict correlation between the identity of a proteins carboxyl-terminal residue and its degradation rate. However, studies with the amino-terminal domain of λ repressor indicate that the properties of at least the five carboxyl-terminal residues of a protein do influence its proteolytic susceptibility (Parsell et al.,

1990). Proteins with non-polar residues at their carboxyl-terminus are rapidly degraded, whereas proteins with charged or polar residues are relatively stable. The Tsp protease is responible for periplasmic tail-specific proteolysis (Keiler and Sauer, 1996). *In vitro* experiments suggest that the non-polar carboxyl-terminal residues bind to a site in Tsp that is separate from the proteolytic active site (Keiler and Sauer, 1995). The cytoplasmic protease responsible for this tail-specific proteolysis has not yet been conclusively identified, but recent evidence suggests that the HflB gene product is involved (Herman et al., 1995; Herman, unpublished results), either directly or indirectly.

In eukaryotes the identity of the amino-terminal residue can also determine the proteolytic fate of specific proteins, and the number of destabilizing residues is larger than in *E. coli* (Varshavsky, 1992). The amino-terminal residue is one of several signals that target proteins for degradation via ubiquitination. Other known signals are the destruction boxes found in cyclins and the *S. cerevisiae* MATα2 protein (Glotzer et al., 1991; Hochstrasser and Varshavsky, 1990). These different signals are recognized by specific ubiquitin conjugating (E2) and ubiquitin ligase (E3) enzymes (reviewed by Jentsch and Schlenker, 1995; Hochstrasser, 1995; Ciechanover, 1995). There are many such enzymes which presumably all recognize different sequence and structural motifs . Recognition results in ubiquitination of the substrate. The first ubiquitin (ubiquitin is an abundant 76 residue protein) is attached through an isopeptide bond between its C-terminus and a lysine side chain in the substrate. Additional ubiquitin moieties can then be added sequentially by forming similar isopeptide bonds with Lys48 in the previous ubiquitin. The branched multiubiquitin chain targets the substrate for degradation by the proteasome. The ubiquitin moieties are not degraded but are removed by specific isopeptidases and recycled for use

in tagging other substrates. In this pathway ubiquitinated substrates are specifically degraded, but substrate selectivity occurs at the level of ubiquitination.

In eukaryotes rapidly degraded proteins often contain stretches rich in Pro, Glu, Ser and Thr residues. However, there are no conclusive experiments indicating whether such sequences are specific degradation signals.

Unfolded and modified proteins. Unfolded proteins may result from incorrect folding, exposure to high temperatures or extremes in pH, covalent modification, and truncation due to incomplete synthesis. Most proteases seem to preferentially degrade unfolded substrates, which may expose one or several of the cleavage motifs or degradation signals described above. In addition, unfolded polypeptide chains can more freely adopt the optimal conformation required for binding to a protease active site. Unfolded proteins also expose hydrophobic residues, which may selectively target unfolded proteins for degradation, either through direct interactions with a protease that binds hydrophobic regions, or via interactions with heat-shock proteins and chaperones.

Certain covalent modifications may also directly target a protein for degradation without causing unfolding. This may be the case for *E. coli* glutamine synthetase (Roseman and Levine, 1986). The oxidized form is rapidly degraded, but oxidation does not seem to have any major structural effect on the protein. It is proposed that the resulting increase in surface hydrophobicity renders the protein proteolytically susceptible. However, oxidation could also result in a small structural rearrangement exposing a cleavage motif or

degradation signal. Other covalent protein modifications, such as alkylation, deamination and glycosylation can also result in rapid degradation. These modifications may result in protein unfolding as discussed above, or may be specifically recognized by some proteases.

Protease localization. An additional source of substrate selectivity is protease localization. Within a multicellular organism, different proteases are expressed in different tissues, resulting in tissue-specific degradation of certain substrates. However, even within a single cell, proteases are compartmentalized, giving them access to a limited number of substrates. Mammalian lysosomal proteases, such as cathepsin B and D, are not very specific *in vitro* (Beynon and Bond, 1989). However, they only degrade proteins *in vivo* that are targetted to the lysosome, and this targetting process is specific. Studies with RNase A suggest that sequences related to KFERQ result in the selective transport of cytosolic proteins to the lysosome (Dice, 1990).

In *E. coli* different proteases are located in the cytoplasm, membranes (inner and outer), and periplasm. *E. coli* signal-peptide peptidases are located in the membrane, where they selectively degrade signal peptides. This selectivity is in part due to their recognition of substrates with hydrophobic sequences characteristic of signal peptides (von Heijne, 1988). However, membrane localization, within multimeric protein complexes responsible for protein secretion, ensures rapid signal-peptide processing. Nonetheless, membrane localization is not always a source of substrate selectivity. The *E. coli* HflB protease is membrane bound, but degrades soluble cytoplasmic proteins such as σ^{32} (Herman et al., 1995).

Many soluble cytoplasmic proteases form multimeric complexes, in which the active sites are sequestered in a central cavity. This allows the active sites to be relatively nonspecific, because selectivity is determined by factors which regulate access of substrates to the cavity. Moreover, once substrates enter the proteolytic cavity they are likely to be cleaved at multiple positions, by one or more active sites. The largest proteolytic complex is the eukaryotic 26S proteasome, responsible for much of the protein degradation in the cytoplasm (Rock, 1994). The proteasome is organised around a core of 28 subunits forming a cylinder with a channel through the middle (Lowe et al., 1995). The proteolytic active sites are located within this cylinder. These active sites do not have strong specificities (however, different cleavage specificities for the various 20S proteasomal subunits may be important to generate different peptides for antigen presentation) and much of the substrate selectivity occurs at the level of targetting proteins to the proteasome (reviewed in Hochstrasser, 1995). As discussed earlier, ubiquitination is one such targetting pathway. Other pathways are not well characterized, but probably involve the regulatory protein caps that cover the cylinder (Peters et al. 1994). In *E. coli*, the Lon and Clp proteases are responsible for the bulk of cytoplasmic ATP-dependent protein degradation (Goldberg, 1992). Both of these proteases also form multimeric complexes. The structural organisation of the Clp protease is similar to that of the eukaryotic proteasome. Negative staining electron microscopy suggests that a core cylinder of ClpP subunits is capped at either end by a ring of ClpA subunits (Kessel et al., 1995). The ClpP monomer contains the proteolytic active site, and the ClpA monomer is thought to be a regulatory, ATP-hydrolyzing, chaperonin-like subunit (Gottesman and Maurizi, 1992). The organisation of the Lon protease is much simpler, and consists of a homotetramer. However, each subunit contains

a proteolytic active site and a separate regulatory site (Gottesman and Maurizi, 1992).

DegP (Protease Do, HtrA)

Protease Do was initially purified from *E. coli* as a protein with strong hydrolytic activity against globin and casein (Goldberg et al., 1981). The *DegP* and *HtrA* genes were identified independently as being required for the degradation of specific periplasmic fusion proteins and growth at elevated temperatures, respectively (Strauch and Beckwith, 1988; Strauch et al., 1989; Lipinska et al., 1989). Subsequently these two genes were shown to be identical (Strauch et al., 1990) and to encode protease Do (Seol at al., 1991). These studies showed that DegP is a periplasmic serine endoprotease and is required for the degradation of unfolded or misfolded proteins. *In vitro* results suggest that DegP may specifically recognize some features characteristic of unfolded proteins. DegP rapidly degrades unfolded BSA, but not folded BSA (Swamy et al., 1983). However, DegP degrades unfolded and folded growth hormone equally slowly (Swamy et al., 1983). Gel filtration chromatography indicates that DegP is a multimer (Swamy et al., 1983). It is tempting to suggest that a cylindrical structure with a central core (analogous to the proteasome or the Clp protease) may only be accessible to unfolded proteins. Expression of the *degP* gene increases under conditions likely to result in the accumulation of unfolded or misfolded proteins in the periplasm. Elevated temperatures and overexpression of a range of outer membrane proteins result in increased transcription requiring the σ^E subunit of RNA polymerase (Erickson et al., 1987; Lipinska et al., 1988; Erickson and Gross, 1989; Mecsas et al., 1993; Hiratsu et al., 1995).

Overexpression of NlpE also involves the two component Cpx pathway (Danese et al., 1995; Snyder et al., 1995).

Several studies in *Salmonella typhimurium* and *Brucella abortus* implicate DegP homologs in pathogenic virulence (Johnson et al., 1991; Roop II et al., 1994; Baumler et al., 1994; Elzer et al., 1994). DegP may simply allow the bacteria to survive the stressful conditions, such as low pH and high temperature, it encounters during infection. Alternatively, DegP may be required to cleave specific host proteins, such as antibacterial peptides (Groisman et al., 1992).

The work presented in Chapter 2 describes the identification and characterization of DegQ and DegS, two DegP homologs in *E. coli*. Chapter 3 presents studies of cleavage specificity for DegP and DegQ. The results suggest that both proteases preferentially degrade unfolded substrates.

Tsp

Tsp was purified as a protease that selectively degrades substrates with non-polar C-termini (Silber et al., 1992). It was identified independently as being responsible for the C-terminal processing of penicillin-binding protein 3 (Hara et al., 1991). Tsp is a periplasmic protease, and is responsible for the selective degradation of periplasmic substrates with non-polar C-termini (Keiler and Sauer, 1996). A *tsp* gene deletion has no significant effect on the degradation of any cytoplasmic proteins tested (Silber and Sauer, 1994). Tsp recognizes the non-polar tails at a site distant from the proteolytic active site, does not exhibit strong cleavage site specificity, and degrades all substrates tested with an appropriate

C-terminus (Keiler et al., 1995). The strong preference for substrates with non-polar C-termini and the weak cleavage site specificity could result in the selective degradation of a specific protein, or in the degradation of all proteins tagged with an appropriate tail.

ssrA RNA

The *E. coli ssrA* RNA molecule was originally isolated as a small, relatively abundant, stable RNA (Lee et al., 1978). The mature 362 nucleotide molecule was shown to be derived from a 457 nucleotide precursor (Chauhan and Apirion, 1989; Subbarao and Apirion, 1989; Oh et al., 1990). Deletions in the *ssrA* gene result in slow growth, poor recovery from carbon starvation (Oh and Apirion, 1991), reduced motility on semisolid agar (Komine at al., 1994), activation of a Lon-independent pathway for degradation of SulA (Trempy and Gottesman, 1989; Trempy et al., 1994; Kirby et al., 1994), and improved repression by a variety of phage and bacterial repressors (Retallack and Friedman, 1995). These results show that *ssrA* RNA is important for *E. coli* growth, but give no strong clues as to its specific function. Analysis of the *ssrA* sequence suggests that the mature molecule may form a tRNA-like structure, with 7 nucleotides at the 5' end and 28 nucleotides at the 3' end arranged into acceptor and TFC stems (Komine et al., 1994). The sequence of the acceptor stem is characteristic of an alanine tRNA, and *ssrA* RNA can be charged with alanine *in vitro* (Komine et al., 1994). The growth defects of an *ssrA* deletion strain are not relieved by mutant *ssrA* genes with changes to the G-U pair in the acceptor stem. However, one of these mutant *ssrA* genes does restore normal growth if alanyl-tRNA synthetase is overexpressed (Komine et al., 1994). These results suggest that alanine charging of *ssrA* RNA is required for its function. Studies of recombinant murine

interleukin-6 (mIL-6) expressed in *E. coli* found that some of the proteins were C-terminally tagged, in an *ssrA* dependent fashion, with an alanine followed by a ten residue peptide encoded by the *ssrA* RNA molecule (Tu at al., 1995). The tagged proteins were all missing a variable number of the mIL-6 encoded C-terminal residues, and Northern analysis showed that a significant portion of the mIL-6 mRNA was truncated (Tu at al., 1995). These results are explained by the work presented in Chapter 4, which shows that proteins translated from an mRNA molecule with no translational termination codon are modified, in an *ssrA* dependent fashion, by C-terminal addition of an alanine followed by the *ssrA* encoded ten residue peptide. The model proposes that the tag is added cotranslationally, with the alanine being derived from the charged *ssrA* RNA. In addition, we show that the tag acts as a specific degradation signal recognised by Tsp in the periplasm and by an unidentified protease in the cytoplasm. This tagging system allows the cell to eliminate potentially toxic protein products translated from mRNA containing 3' truncations.

REFERENCES

Baird, L. and Georgopoulos, C. (1990) "Identification, cloning, and characterization of the *Escherichia coli sohA* gene, a suppressor of the *htrA* (*degP*) null phenotype." *J. Bacteriol.* **172**, 1587-1594.

Baird, L., Lipinska, B., Raina, S., and Georgopoulos, C. (1991) "Identification of the *Escherichia coli sohB* gene, a multicopy suppressor of the HtrA (DegP) null phenotype." *J. Bacteriol.* **173**, 5763-5770.

Banuett, F., Hoyt, M. A., McFarlane, L., Echols, H., and Herskowitz, I. (1986) "*hflB*, a new *Escherichia coli* locus regulating lysogeny and the level of bacteriophage lambda CII protein." *J. Mol. Biol.* **187**, 213-224.

Baumler, A. J., Kusters, J. G., Stojiljkovic, I., and Heffron, F. (1994) "*Salmonella typhimurium* loci involved in survival within macrophages." *Infect. Immun.* **62**, 1623-1630.

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

Brown, J. W., Hunt, D. A., and Pace, N. R. (1990) "Nucleotide sequence of the 10Sa RNA gene of the b-purple eubacterium *Alcaligenes eutrophus*." *Nucleic Acids Res.* **18**, 2820.

Cavard, D., Lazdunski, C., and Howard. S. P. (1989) "The acylated precursor form of the Colicin A lysis protein is a natural substrate of the DegP protease." *J. Bacteriol.* **171**, 6316-6322.

Chauhan, A. K. and Apirion, D. (1989) "The gene for a small stable RNA (10Sa RNA) of *Escheridchia coli*." *Mol. Microbiol.* **3**, 1481-1485.

Chun, R. T., Bar-nun, S., and Simoni, J. (1990) "The regulation of degradation of 3-hydroxy-3-methylglutaryl-CoA reductase requires a short-lived protein and occurs in the endoplasmic reticulum." *J. Biol. Chem.* **265**, 22004-22010.

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

Danese, P. N., Snyder, W. B., Cosma, C. L., Davis, L. J. B., and Silhavy, T. J. (1995) "The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP." *Genes and Dev.* **9**, 387-398.

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

Elzer, P.J., Phillips, R. W., Kovach, M. E., Peterson, K. M., and Roop II, R. M. (1994) "Characterization and genetic complementation of a *Brucella abortus* high-temperature-requirement A (*htrA*) deletion mutant." *Infect. Immun.* **62**, 4135-4139.

Erickson, J. W. and Gross, C. A. (1989) "Identification of the σ^E subunit of *Escherichia coli* RNA polymerase: a second alternate s factor involved in high-temperature gene expression." *Genes and Dev.* **3**, 1462-1471.

Erickson, J. W., Vaughn, V., Walter, W. A., Neidhardt, F. C., and Gross, C. A. (1987) "Regulation of the promoters and transcripts of *rpoH*, the *Escherichia coli* heat shock regulatory gene." *Genes and Dev.* **1**, 419-432.

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., Swamy, K. H. S., Chung C. H., and Larimore, F. S. (1981) "Proteases in *Escherichia coli*." *Methods Enzymol.* **80**, 680-702.

Goldberg, A. L., (1992) "The mechanism and functions of ATP-dependent proteases in bacterial and animal cells." *Eur. J. Biochem.* **203**, 9-23.

Groisman, E. A., Parra-Lopez, C., Salcedo, M., Lipps, C. J., and Heffron, F. (1992) "Resistance to host antimicrobial peptides is necessary for *Salmonella* virulence." *Proc. Natl. Acad. Sci. USA*. **89**, 11939-11943.

Gottesman, S. and Maurizi, M. R. (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.

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

Herman, C., Ogura, T., Tomoyasu, T., Hiraga, S., Akiyama, Y., Koreaki, I., Thomas, R., D'Ari, R., and Bouloc, P. (1993) "Cell growth and λ phage development controlled by the same essential *Escherichia coli* gene, *ftsH/hflB*." Proc. Natl. Acad. Sci. USA **90**, 10861-10865.

Herman, C., Thevenet, D., D'Ari, R., and Bouloc, P. (1995) "Degradation of σ^{32}, the heat shock regulator in *Escherichia coli*, is governed by Hf1B." *Proc. Natl. Acad. Sci. USA* **92**, 3516-3520.

Hiratsu, K., Amemura, M., Nashimoto, H., Shinagawa, H., and Makino K. (1995) "The *rpoE* gene of *Escherichia coli*, which encodes σ^{E}, is essential for bacterial growth at high temperature." *J. Bacteriol.* **177**, 2918-2922.

Hochstrasser, M., and Varshavsky, A. (1990) "*In vivo* degradation of a transcriptional regulator: the yeast α2 repressor." *Cell* **61**, 697-708.

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

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

Johnson, K., Charles, I., Dougan, G., Pickard, D., O'Gaora, P., Costa, G., Ali, T., Miller, I., and Hormaeche, C. (1991) "The role of a stress-response protein in *Salmonella typhimurium* virulence." *Mol. Microbiol.* **5**, 401-407.

Keiler, K. C. and Sauer, R. T. (1995) "Identification of active site residues of the Tsp protease." *J. Biol. Chem.* **270**, 28864-28868.

Keiler, K. C. and Sauer, R. T. (1996) "Substrate specificity of the Tsp protease." *J. Biol. Chem.* (in press).

Keiler, K. C., Silber, K. R., Downard, K. M., Papayannopoulos, I. A., Biemann, K., and Sauer, R. T. (1995) "C-terminal specific protein degradation: Activity and substrate specificity of the Tsp protease." *Protein Science* **4**, 1507-1515.

Kellerman, K. A., Mattson, D. M., and Duncan, I. (1990) "Mutations affecting the stability of the *fushi tarazu* protein of *Drosophila*." *Genes and Dev.* **4**, 1936-1950.

Kessel, M., Maurizi, M. R., Kim, B., Kocsis, E., Trus, B. L., Singh, S. K., and Steven, A. C. (1995) "Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26 S proteasome." *J. Mol. Biol.* **250**. 587-594.

Kirby, J. E., Trempy, J. E., and Gottesman, S. (1994) "Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli*." *J. Bacteriol.* **176**, 2068-2081.

Komine, Y., Kitabatake, M., Yokogawa, T., and Nishikawa, K. (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.

Lee, S. Y., Bailey, S. C., and Apirion, D. (1978) "Small stable RNAs from *Escherichia coli*: evidence for the existence of new molecules and for a new ribonucleoprotein particle containing 6S RNA." *J. Bacteriol.* **133**, 1015-1023.

Lipinska, B., Fayet, O., Baird, L., and Georgopoulos, C. (1989) "Identification, characterization, and mapping of the *Escherichia coli htrA* gene, whose product is essential for bacterial growth only at elevated temperatures." *J. Bacteriol.* **171**, 1574-1584.

Lipinska, B., Sharma, S., and Georgopoulos, C. (1988) "Sequence analysis and regulation of the *htrA* gene of *Escherichia coli*: a σ^{32}-independent mechanism of heat-inducible transcription." *Nucleic Acids Res.* **16**, 10053-10067.

Lipinska, B., Zylicz, M., and Georgopoulos, C. (1990) "The HtrA (DegP) protein, essential for *Escherichia coli* survival at high temperatures, is an endopeptidase." *J. Bacteriol.* **172,** 1791-1797.

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.

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.

Mecsas, J., Rouviere, P. E., Erickson, J. W., Donohue, T. J., and Gross, C. A. (1993) "The activity of σ^E, an *Escherichia coli* heat-inducible σ-factor, is modulated by expression of outer membrane proteins." *Genes and Dev.* **7,** 2618-2628.

Misuzawa, S., and Gottesman, S. (1983) "Protein degradation in *Escherichia coli*: the *lon* gene controls the stability of the SulA protein." *Proc. Natl. Acad. USA* **80,** 358-362.

Oh, B-K. and Apirion, D. (1991) "10Sa RNA, a small stable RNA of *Escherichia coli*, is functional." *Mol. Gen. Genet.* **229,** 52-56.

Oh, B-K., Chauhan, A. K., Isono, K., and Apirion, D. (1990) "Location of a gene (ssrA) for a small, stable RNA (10Sa RNA) in the *Escherichia coli* chromosome." *J. Bacteriol.* **172**, 4708-4709.

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

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.

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

Retallack, D. M. and Friedman, D. I. (1995) "A role for a small stable RNA in modulating the activity of DNA-binding proteins." *Cell* **83**, 227-235.

Retallack, D. M., Johnson, L. L., and Freidman, D. I. (1994) "Role for 10Sa RNA in the growth of λ-P22 hybrid phage." *J. Bacteriol.* **176**, 2082-2089.

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.

Roop II, R. M., Fletcher T. W., Sriranganathan, N. M., Boyle, S. M., and Schurig, G. G. (1994) "Identification of an immunoreactive *Brucella abortus* HtrA stress response protein homolog." *Infect. Immun.* **62**, 1000-1007.

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.

Scheffner, M., Nuber, U., and Hulbregtse, J. M. (1995) "Protein ubiquitination involving and E1-E2-E3 enzyme ubiquitin thioester cascade." *Nature* **373**, 81-83.

Seol, J. H., Woo, K. S., Jung, E. M., Yoo, S. J., Lee, C. S., Kim K., Tanaka, K., Ichihara, A., Ha, D. B., and Chung, C. H. (1991) "Protease Do is essential for survival of *Escherichia coli* at high temperatures: its identity with the *htrA* gene product." *Biochem. Biophys. Res. Commun.* **176**, 730-736.

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

Seufert, W., Futcher, B., and Jentsch, S. (1995) "Role of a ubiquitin-conjugating enzyme in degradation of S- and M-phase cyclins." *Nature* **373**, 78-81.

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. USA* **89**, 295-299.

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* K-12." *Mol. Gen. Genet.* **242**, 237-240.

Skorko-Glonek, J., Krzewski, K., Lipinska, B., Bertoli, E., and Tanfani F. (1995) "Comparison of the structure of wild-type HtrA heat shock protease and mutant HtrA proteins." *J. Biol. Chem.* **270**, 11140-11146.

Skorko-Glonek, J., Wawrzynow, A., Krzewski, K., Kurpierz, K., and Lipinska, B. (1995) "Site-directed mutagenesis of the HtrA (DegP) serine protease, whose proteolytic activity is indispensable for *Escherichia coli* survival at elevated temperatures." *Gene* **163**, 47-52.

Snyder, W. B., Davis, L. J. B., Danese, P. N., Cosma, C. L., and Silhavy, T. J. (1995) "Overproduction of N1pE, a new outer membrane lipoprotein, suppresses the toxicity of periplasmic LacZ by activation of the Cpx signal transduction pathway." *J. Bacteriol.* **177**, 4216-4223.

Strauch, K. L. and Beckwith, J. (1988) "An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins." *Genetics* **85**, 1576-1580.

Strauch, K. L., Johnson, K., and Beckwith, J. (1989) "Characterization of *degP*, a gene required for proteolysis in the cell envelope and essential for growth of *Escherichia coli* at high temperature." *J Bacteriol.* **171**, 2689-2696.

Strauch, M. A., Baumann, M., Friedman, D. I., and Baron, L. S. (1986) "Identification and characterization of mutations in *Escherichia coli* that selectively influence the growth of hybrid λ bacteriophages carrying the immunity region of bacteriophage P22." *J. Bacteriol.* **167**, 191-200.

Subbarao, M. N. and Apirion, D. (1989) "A precursor for a small stable RNA (10Sa RNA) of *Escherichia coli*." *Mol. Gen. Genet.* **217**, 499-504.

Swamy, K. H. S., Chung, C. H., and Goldberg, A. L. (1983) "Isolation and characterization of protease Do from *Escherichia coli,* a large serine protease containing multiple subunits." *Arch. Biochem. Biophys.* **224,** 543-554.

Terlecky, S., R. and Dice, J. F. (1993) "Polypeptide import and degradation by isolated lysosomes." *J. Biol. Chem.* **268,** 23490-23495.

Tilly, K., Spence, J., and Georgopoulos, C. (1989) "Modulation of stability of the *Escherichia coli* heat shock regulatory factor σ^{32}." *J. Bacteriol.* **171,** 1585-1589.

Trempy, J. E. and Gottesman, S. (1989) "Alp, a suppressor of Lon protease mutants in *Escherichia coli.*" *J. Bacteriol.* **171,** 3348-3353.

Trempy, J. E., Kirby, J. E., and Gottesman, S. (1994) "Alp suppression of Lon: dependence on the *slpA* gene." *J Bacteriol.* **176,** 2061-2067.

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.

Vogel, R. F., Entian, K-D., and Mecke, D. (1987) "Cloning and sequence of the *mdh* structural gene of *Escherichia coli* coding for malate dehydrogenase." *Arch. Microbiol.* **149,** 36-42.

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.

Zheng, C. and Friedman, D. I. (1994) "Reduced Rho-dependent transcription termination permits NusA-independent growth of *Escherichia coli*." *Proc. Natl. Acad. Sci. USA* **91**, 7543-7547.

CHAPTER 2

Characterization of *degQ* and *degS*, *Escherichia coli* Genes Encoding Homologs of the DegP Protease

Waller, P. R. H., and Sauer, R. T. 1996. Characterization of *degQ* and *degS*, *Escherichia coli* Genes Encoding Homologs of the DegP Protease. J. Bacteriol. **178:** 1146-1153.

INTRODUCTION

Proteolysis plays an important role in regulating the steady state levels of many intracellular proteins. In addition, proteolysis is required for the degradation of misfolded proteins, the correct processing of secreted proteins, and survival under physiologically stressful conditions such as exposure to high temperature or nutrient deprivation. These varied proteolytic functions are carried out by many enzymes. In *Escherichia coli*, for example, over twenty different intracellular proteases have been identified (18).

This chapter describes the identification and characterization of two *E. coli* genes, designated *degQ* and *degS*, which encode homologs of the DegP protease. DegP (HtrA) is a periplasmic serine protease that is required for *E. coli* growth at elevated temperatures (15, 17, 26), and is involved in the degradation of abnormal periplasmic proteins (25). The *degQ* gene is located immediately upstream of the *degS* gene, but each gene appears to be transcribed independently and neither is heat-inducible. The *degP* gene, by contrast, is heat inducible (16). Using biochemical and genetic analyses, we show that the DegQ protein is a periplasmic serine endoprotease, which can functionally substitute for DegP under some conditions. The DegS protein cannot substitute for DegP, but is required for normal bacterial growth. The *degQ* (*hhoA*) and *degS* (*hhoB*) genes have also recently been isolated as multicopy suppressors of the temperature/osmotic sensitive phenotype of an *E. coli* strain lacking another periplasmic protease (1).

MATERIALS AND METHODS

Bacterial strains and plasmids. The *E. coli* strains used in this work are listed in Table 1. Bacteria were grown in LB medium (10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl per liter). Low-salt medium contains 5 g of tryptone and 2.5 g of yeast extract per liter. High-salt medium is LB with an additional 15 g of NaCl per liter. Plates contained 1.5% agar. Ampicillin (100 μg/ml), chloramphenicol (20 μg/ml), and kanamycin (40 μg/ml) were included as needed. Standard techniques were used for plasmid constructions.

Subcloning and sequencing. The *degQ* and *degS* genes are located in a tandem array immediately downstream of the malate dehydrogenase (*mdh*) gene, and are oriented in the opposite direction from the *mdh* gene. A ~2.5 kb EcoRI-KpnI fragment and a ~3.0 kb HindIII fragment from Kohara phage 526 (14) were cloned into pUC19 and pBR322, respectively. The EcoRI-KpnI fragment contains the 3' end of the *mdh* gene and ~2.3 kb of downstream sequence. The HindIII fragment contains only sequence downstream from the *mdh* gene (Fig. 1A). The HindIII and EcoRI-KpnI fragments overlap by 813 bp. Regions from these fragments covering the sequence immediately downstream from the *mdh* gene were further subcloned into pUC19 using convenient restriction sites. Both DNA strands were sequenced using the double-stranded sequencing method described in the Sequenase 2.0 kit from United States Biochemical. The pUC19 forward and reverse primers were used for most of the sequencing. Primers were synthesized to sequence regions that could not be reached using the pUC19 primers.

Sequence analysis was carried out using the sequence analysis software package from the Genetics Computer Group, Inc.

RNA analysis. RNA was isolated from 10 ml cultures of *E. coli* strain MC1061, which were grown to an A_{600} of 0.8 in LB medium. Cells were chilled on ice, harvested by centrifugation, and resuspended in 0.5 ml of 30 mM sodium acetate (pH 5.2). These cell suspensions were mixed with 0.1 ml of 10% SDS and 0.5 ml of phenol (equilibrated with 30 mM sodium acetate (pH 5.2)). The mixture was incubated at 65 °C for 10 min and vortexed every 2 min. After microcentrifugation for 2 min, the aqueous supernatant was extracted once more with hot phenol, and then extracted once with chloroform. The RNA was precipitated with 1/4 volume of 3 M sodium acetate (pH 5.2) and 3 volumes of 100% ethanol. The pellet was resuspended in 50 µl water. The concentration of RNA was determined from the A_{260}.

For Northern-blot analysis, 20 µg of RNA was electrophoresed on a formaldehyde gel (6) and transferred to a nylon membrane using the Schleicher and Schuell Turboblot apparatus. The blot was prehybridized at 42 °C for 2 hr in a solution containing 5x SSPE, 7% SDS, and 40% formamide. Hybridization was carried out at 42 °C overnight in the same solution with a ^{32}P-labeled DNA probe, generated from a double-stranded DNA fragment using the random primed labeling kit from United States Biochemical. The blot was washed in 2x SSC, 0.1% SDS for 30 min at room temperature, and in 0.1x SSC, 0.1% SDS for 15 min at 42 °C.

For primer extension, 30 μg of RNA was incubated with 5 x 10^4 cpm of ^{32}P end-labeled primer in 15 μl of 150 mM KCl, 10 mM Tris-HCl (pH 8.0), and 1 mM EDTA at 65 °C for 90 min. After slowly cooling to room temperature, 30 μl of 30 mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 8 mM DTT, 0.2 mg/ml actinomycin D, 1 mM each of all four dNTPs, and 0.2 U/μl MoMULV reverse transcriptase (from New England Biolabs) were added to the annealed primer. The mixture was incubated at 42 °C for 1 hr, then 105 μl of 20 μg/ml RNase A in TE buffer (pH 7.5) plus 100 mM NaCl was added and incubation was continued at 37 °C for 15 min. The reaction was extracted with 150 μl phenol/chloroform/isoamyl alcohol. The DNA was precipitated with 15 μl of 3 M sodium acetate (pH 5.2), 2 μl of 1 mg/ml glycogen, and 300 μl of 100% ethanol. The pellet was washed with 70% ethanol and air dried. The pellet was resuspended in stop buffer (United States Biochemical Sequenase 2.0 kit) and electrophoresed on an 8% polyacrylamide gel containing 8.3 M urea. A sequence ladder, generated using the same primer in a sequencing reaction, was electrophoresed next to the primer extension product. The primers (5′ CCTGGAATCGACGCGACGGCCTGAAATGACGCCGA 3′ and 5′ GAGTGGAAAGCGGGTTAAGGCTGCGCAGCGAAGG 3′, for *degQ* and *degS*, respectively) were designed to anneal ~100 nucleotides downstream of the putative transcriptional start sites.

Construction of expression vectors. The *degP*, *degQ*, and *degS* genes were amplified by PCR from genomic DNA isolated from strain MC1061, and cloned into pAD100, a pDW239 derived plasmid with a pBR322 origin of replication (29). For all three genes, the primers used for amplification generated a new restriction site (NcoI for the *degQ* and *degP* genes, AflIII for the *degS* gene) that

includes the first ATG codon of the gene. In each case, the new site was used to clone the 5' end of the gene into the unique NcoI site of pAD100. In these constructs, each gene is under transcriptional control of the plasmid borne P_{trc} promoter, the ribosome binding site is plasmid encoded, and translation starts at the ATG codon which is part of the NcoI site. Primers A (5' AAATCCATGGCG AAAACCACATTAGCACTG 3') and B (5' CCGAATTCACAGATTGTAAGGAG AACC 3') were used to amplify the *degP* gene. The resulting fragment was digested with NcoI and EcoRI and cloned into the NcoI-EcoRI backbone of pAD100, generating pPW100. Primers C (5' GGAACCATGGCGAAACAAACCC AGCTGTTG 3') and D (5' TCTATCGGTACCCTGAGC 3') were used to amplify the tandem *degQ* and *degS* genes. The resulting fragment was digested with NcoI and KpnI and cloned into the NcoI-KpnI backbone of pAD100, generating pPW145. This plasmid was then digested with BamHI and BglII and religated, resulting in the removal of most of the *degS* coding sequence, generating pPW146. Primers E (5' GCCTCCAACATGTTTGTGAAGCTCTTAC 3') and D were used to amplify the *degS* gene. The resulting fragment was partially digested with AflIII and BamHI and cloned into the NcoI-BamHI backbone of pPW145, generating pPW147. For both DegP and DegQ, these constructions result in the replacement of Lys by Ala at position 2 in the signal sequence. This change does not appear to affect export, as both DegP and DegQ behaved like soluble periplasmic proteins upon subcellular fractionation of induced cultures using standard procedures described in the Qiagen (Qiaexpressionist) handbook. Primer F (5' TTCAGGTTAACCAGCGCACCACCGGCGTTAC CACGG 3') and primer A were used to amplify the amino-terminal half of the *degP* gene, and introduce a mutation changing the codon for the putative active site Ser^{236} to a

codon for Ala. The resulting fragment was digested with HpaI and NcoI, and cloned into pPW100 partially cut with HpaI and BamHI generating pPW105.

The expression levels of DegP, DegQ, DegS, and DegP SA236 upon induction of the P_{trc} promoter were similar as judged by electrophoresis of lysates from IPTG-induced cultures on SDS-Laemmli gels and staining with Coomassie Blue. The phenotype-rescue experiments described in the Results section were carried out with these plasmids in the absence of IPTG, because cell growth rates were significantly reduced in the presence of IPTG in each case. In the absence of IPTG, the *degP*, *degQ*, and *degS* genes are expressed at a sufficiently low level such that no visible protein is detected on a Coomassie stained SDS-Laemmli gel.

Purification of the DegQ protein. Strain PW151/pPW146 was grown in 2 liters of LB broth with ampicillin, at 30 °C, to an A_{600} of 0.6. The cells were induced with 0.4 mM IPTG and grown for a further 3 hr. The cells were harvested by centrifugation, resuspended in 30 ml of lysis buffer (50 mM Tris-Cl pH 8.0, 10 mM $MgCl_2$, 200 mM KCl) and lysed by sonication. The lysate was then diluted by addition of 120 ml lysis buffer and cleared by the addition of polyethylenimine to 1% and centrifugation. The supernatant was then precipitated with ammonium sulfate added to 90% saturation. The pellet was resuspended in 20 ml of 20 mM Tris-Cl (pH 8.0) and dialysed against the same buffer. This solution was then loaded onto a DEAE Sephacel column (1.5 x 5 cm). After extensive washing with the loading buffer, the bound proteins were eluted with a linear gradient of KCl from 0 to 150 mM. The DegQ protein eluted at ~60 mM KCl (in ~10 ml of pooled fractions). This solution was then dialysed against

20 mM potassium phosphate (pH 6.0), and passed through a CM Sepharose column (1.5 x 5 cm). The flow-through fraction contained the DegQ protein and was dialysed against 50 mM potassium phosphate (pH 7.4), concentrated, and loaded onto a Sephacryl S300 column (1 x 50 cm). The DegQ protein eluted from this column ahead of the remaining contaminants. Samples of this material showed no other visible bands on a Coomassie stained SDS-Laemmli gel.

The proteolytic activity of DegQ was assayed in 15 µl of 50 mM potassium phosphate (pH 7.4) with 5 µg of β-casein as a substrate. The reactions were incubated at 37 °C for 90 min, and electrophoresed on a 12% SDS-Laemmli gel. The inhibitory effect of DFP (diisopropylfluorophosphate), a serine protease inhibitor, was tested by adding different amounts of DFP (each dissolved in 0.5 µl of isopropanol) to the DegQ reaction mixtures 10 min before the addition of β-casein.

Construction of mutant strains. Deletions in the *degQ* and *degS* genes were constructed using a gene replacement protocol (10) in which a disrupted gene on a plasmid is exchanged for the corresponding chromosomal gene. Four deletion constructs, in which different regions of the *degQ* and *degS* genes were deleted or replaced were cloned into pMAK705 (a plasmid which confers resistance to chloramphenicol, and has a temperature-sensitive pSC101 origin of replication). Plasmid pPW136 contains the ~4.7 kb HindIII-KpnI fragment (Fig. 1A) with the ~0.4 kb EcoRV region deleted. Plasmid pPW138 contains the ~4.7 kb HindIII-KpnI fragment with the ~1.5 kb EcoRV-ScaI region deleted. Plasmid pPW140 contains the ~1.7 kb HindIII-KpnI fragment with the ~0.2 kb BamHI-ScaI region

deleted (the BamHI site was filled in using Klenow before ligation to the ScaI site). Plasmid pPW141 contains the ~0.8 kb EcoRI-HindIII fragment with the ~0.4 kb EcoRV region replaced by the kanamycin resistance gene from mini-Tn5 Km (7). For each construct, transformants of strain MC1061 were grown on LB chloramphenicol plates at 44 °C to select for integration of the plasmid into the chromosome via recombination between one of the regions flanking the deletion and the homologous sequence on the chromosome. Integrants were then grown overnight, at 30 °C, in 25 ml cultures of LB broth with chloramphenicol. During this growth period at the permissive temperature, the integrated constructs can resolve via homologous recombination to regenerate freely replicating plasmids. These cultures were plated on LB chloramphenicol, grown at 30 °C, and individual colonies were isolated and screened for plasmid content and structure by plasmid miniprep analysis. Plasmids that contain the complete gene segment (as opposed to the original deleted gene segment) should result from a resolution event involving the flanking region not used during integration, and their host cells should have the corresponding chromosomal deletion. Colonies containing plasmids with the complete gene segment were grown in culture in LB at 30 °C. Dilutions from each culture were plated, in duplicate, on LB plates. One of the plates was incubated at 30 °C, and the other at 37 °C. Growth at 37 °C, on non-selective medium, results in plasmid loss, because the temperature-sensitive pSC101 replicon is poorly functional at 37 °C. Similar numbers of colonies were obtained at 30 °C and 37 °C after gene replacement with plasmids pPW136 and pPW138, indicating that neither *degQ* nor *degS* is required for cell viability. The loss of the plasmid in candidate colonies was confirmed by the inability of the deletion strain to grow on LB chloramphenicol. The presence of the desired

deletion on the chromosome was confirmed by PCR analysis. The structures of the resulting chromosomal gene deletion/replacements are shown in Fig. 1B (strains PW147, PW148, PW149 and PW152 were generated using plasmids pPW136, pPW138, pPW140 and pPW141, respectively). An otherwise isogenic *degP* deletion strain (PW151) was constructed by transferring the *degP41* deletion from strain KS474 to strain MC1061 by P1 transduction. Transductants were selected for by kanamycin resistance, which is encoded by the insert in the *degP41* deletion construct.

Assay to test the plasmid rescue of the *degS* deletion phenotype. When a strain containing a *degS* deletion is grown in liquid medium, faster growing suppressor mutants take over the culture. This complicates direct assays for the ability of test plasmids to rescue the small colony phenotype, because competent cells prepared from *degS* deletion strains contain faster growing cells which form normal sized colonies irrespective of whether the plasmid compensates for the deletion. We therefore devised the following plasmid shuffle assay. Using the gene replacement protocol described above, the *degS* deletion construct on pPW140 was exchanged for the chromosomal copy of the gene. To maintain the resulting plasmid (pPW142, bearing the wild-type copy of the *degS* gene) the strain (PW149/pPW142) was grown at 30 °C in LB broth with chloramphenicol, and competent cells were made. These cells were transformed with test plasmids (containing a pBR322 origin of replication, which is compatible with the pSC101 origin of replication of pPW142), plated onto LB with ampicillin and chloramphenicol (which select for the test plasmids and pPW142 respectively), and incubated at 30 °C. Transformants were grown in LB broth with ampicillin

at 30 °C, plated on LB ampicillin plates, and incubated at 37 °C, which results in the loss of pPW142 containing the wild-type copy of the *degS* gene. If the test plasmid rescues the small colony *degS* deletion phenotype, the resulting colonies are of normal size. If the test plasmid does not rescue the phenotype, the resulting colonies are small.

Nucleotide sequence accession number. The GenBank accession number for the 2,714 bp nucleotide sequence reported in this work is U32495.

RESULTS

The *degQ* and *degS* gene sequences. At the beginning of these studies, a partial open reading frame encoding a sequence homologous to part of the DegP protein was known to be located downstream from the *E. coli* malate dehydrogenase (*mdh*) gene (3, 27). To examine potential structural and functional similarities of the protein(s) encoded by this region with the DegP protein, we subcloned and sequenced a 2.7 kb region from Kohara phage 526 (14), containing the 3' end and downstream region of the *E. coli mdh* gene. Two genes, designated *degQ* and *degS*, were identified. The open reading frames of the *degQ* and *degS* genes are oriented in the same direction, are separated by 92 bases, and appear to encode proteins of 455 and 355 residues, respectively. The gene organization, DNA sequence, and deduced amino acid sequences of the DegQ and DegS proteins are shown in Fig. 2. Identical DNA sequences for this region also have been reported by Bass (1) (GenBank accession number U15661) and Blattner (GenBank

accession number U18997). The *degQ* and *degS* genes are located at approximately 72-73 min on the *E. coli* chromosome.

Several sequence features suggest that *degQ* and *degS* may be transcribed and translated independently: (i) Sequences including potential -35 and -10 hexamers, characteristic of promoters recognized by σ^{70} containing RNA polymerase (9, 11, 19) are found upstream of both the *degQ* and *degS* genes (Fig. 2). (ii) Reasonable Shine-Dalgarno sequences (23) are present a few bases upstream of the putative first ATG codon of each gene (Fig. 2). (iii) Stem-loop structures followed by a run of T's, characteristic of rho-independent transcription terminators (4, 21), are present at the 3' ends of both genes (Fig. 2).

As shown in Fig. 3, the amino acid sequences deduced for the DegQ and DegS proteins share homology with each other and with the DegP sequence (note that the C-terminal sequence of DegP is different from that previously published (16), because of a frameshift error at the 3' end of the gene). DegQ and DegP are of similar size (455 and 474 residues, respectively) and are the most closely related, with strong homology (60% identity and 75% sequence similarity) over their entire length. One difference between DegQ and DegP is that the latter protein contains an insertion of approximately 20 residues (near position 80), which contains the only cysteines in either protein. This region is rich in glycine, serine, and proline suggesting that it may form a flexible loop in DegP. The DegS protein is 90 residues shorter than DegQ and DegP at its carboxy-terminus and also shows less sequence conservation (~37% identity and ~58% sequence similarity when compared with either DegP or DegQ), with the most significant

divergence occurring in the 70 amino-terminal residues. However, all three proteins show strong sequence conservation in the regions surrounding the His, Asp and Ser residues of the putative catalytic triad (3) (Fig. 3). This suggests that DegQ and DegS, like DegP, are serine proteases.

The 27 amino-terminal residues of DegQ have several characteristics of a signal peptide (12, 20), including positively charged residues at the amino-terminus (Lys at positions 2 and 3) followed by a stretch of hydrophobic residues and a potential signal peptide cleavage site between residues 27 and 28 (Ala-Val-Ala/Ser^{28}). Indeed, purified DegQ protein is cleaved at or near this site (see below). The signal peptides of DegQ and DegP (17) are very similar. The amino-terminal residues of the DegS protein also have some of the characteristics of a signal peptide, however an overproduced variant of DegS containing a C-terminal His_6 tag has the N-terminal sequence expected for the intact, unprocessed protein (E. Roche, personal communication).

RNA transcript analysis. The 5′ regions of *degQ* and *degS* do not have any sequences characteristic of heat-shock promoters suggesting that, unlike *degP*, these genes are not heat-inducible. To test this hypothesis, *E. coli* MC1061 cells were grown to an A_{600} of 0.8 in LB broth at 30 °C and then transferred to 44 °C. RNA was isolated from these cells at 0, 10, 20, and 40 min after the temperature shift. A Northern blot was prepared with duplicate RNA samples from each time point. The two halves of the blot were probed separately with ^{32}P-labeled fragments derived from the *degQ* and *degS* genes. In the resulting autoradiograms there is no notable increase in the transcript level from either

gene after the shift to 44 °C (Fig. 4A). As a control, one of these blots was then probed with a ^{32}P-labeled fragment derived from the *degP* gene. As expected (16) this gene shows a significant increase in transcript level after the shift to 44 °C (Fig. 4B).

The transcripts of ~1.5 kb and ~1.1 kb seen for the *degQ* and *degS* genes (Fig. 4A), respectively, are consistent with the transcript sizes expected from the promoters and terminators shown in Fig. 2. In addition, a band of ~2.6 kb is seen with both probes, consistent with partial transcriptional readthrough of the terminator at the end of the *degQ* gene.

To map the 5' ends of the RNAs encoding the *degQ* and *degS* genes, primer extensions were performed using oligonucleotides complementary to the 5' regions of the two genes. As shown in Fig. 5, both of these RNAs start at guanines located 10 bases downstream from the -10 hexamer of the putative promoter sequences. The features of these sequences that are characteristic of σ^{70} dependent promoters are indicated in Fig. 5C. Although we cannot rigorously rule out mRNA processing, the primer extension results and the sequence homology to the σ^{70} consensus promoter suggest that the smaller RNAs seen on the Northern blots arose from transcription from the *degQ* and *degS* promoters indicated in Fig. 2. Consistent with this model, genetic results involving insertion of a strong transcription termination signal in the *degQ* gene (see below) indicate that the *degS* gene can be transcribed from a promoter located at the 3' end of the *degQ* gene.

The purified DegQ protein is proteolytically active. To test the hypothesis that DegQ is a serine protease, it was purified to greater than 95% homogeneity from a *degP* deletion strain containing a plasmid expressing the *degQ* gene from an IPTG-inducible promoter. Incubation of DegQ with β-casein results in several cleavage fragments (Fig. 6), suggesting that DegQ is an endopeptidase. A series of fractions from the final Sephacryl S300 column used in the purification of DegQ were assayed for proteolytic activity. The amount of activity in each fraction correlated with the amount of DegQ protein, suggesting that the proteolysis observed is indeed due to DegQ and not to a contaminating protein. The degradation of β–casein was significantly inhibited by 5 mM DFP, a serine protease inhibitor, indicating that DegQ is a serine protease (Fig. 6).

To determine whether the predicted signal peptide is removed from DegQ, the amino-terminal sequence of the purified protein was determined (M.I.T. Biopolymers Laboratory). Two major sequences and one minor sequence were obtained. One of the major sequences (SIPGQ...) begins at Ser^{28} consistent with cleavage at the predicted site, the other (ASIPG...) begins at Ala^{27} indicating cleavage one residue upstream. The minor sequence (FQAVA...) begins at Phe^{23} indicating cleavage at a secondary signal peptide cleavage site (Ser-Ala-Ser/Phe^{23}).

A plasmid expressing the *degQ* gene rescues the phenotype of a *degP* deletion strain. The amino acid sequence of the DegQ protein and the results from the experiments *in vitro* described above suggest that DegQ may be functionally similar to DegP. We tested the ability of the *degQ* gene to rescue the temperature

sensitive growth defect due to a deletion in the *degP* gene. Strain PW151, which does not grow at 44 °C as a result of the *degP41* deletion, was transformed with plasmids pPW146 and pPW100 expressing the *degQ* and *degP* genes, respectively. Transformants with each plasmid grew well on LB ampicillin at 37 °C and 44 °C (Table 2). As a control, we transformed strain PW151 with plasmid pPW105 expressing a mutant DegP protein in which Ser^{236}, a putative active site residue (3, 24), is replaced by Ala. Transformants of PW151/pPW105 grew on LB ampicillin at 37 °C but not at 44 °C, suggesting that the temperature sensitive growth of the *degP41* deletion strain is due to a proteolytic defect, and is not an indirect effect of the physical absence of the DegP protein. In a similar experiment, we found that plasmid pPW147, expressing the DegS protein, was also unable to rescue the temperature sensitive phenotype of the *degP41* strain. None of these plasmids affected the ability of strain MC1061 to grow at 44 °C.

The *degS* gene is important for normal cell growth. To test whether *degQ* or *degS* play important roles in cell growth or viability, single and double *degQ* and *degS* gene deletions were constructed in the MC1061 strain background (Fig. 1). A strain in which the *degQ* gene bears a deletion (PW147) shows no obvious growth defect or notable colony morphology on LB agar plates. In contrast, a strain in which the *degS* gene contains a deletion (PW149) shows a small colony phenotype on LB plates (Fig. 7A). A strain bearing deletions in both the *degQ* and *degS* genes (PW148) shows the same small colony phenotype as the strain bearing only the deletion in the *degS* gene (PW149). A strain with a kanamycin resistance element inserted into the *degQ* gene (PW152) has no obvious phenotype, indicating that the strong transcriptional terminators associated with

this element (7) do not significantly affect expression of the downstream *degS* gene. None of the *degQ*, *degS* or *degQ degS* mutant strains showed any additional growth or colony morphology phenotypes on plates with different salt concentrations (LB, low salt, and high salt) or at different temperatures (room temperature, 30 °C, 37 °C, and 44 °C).

These single and double deletions in the *degQ* and *degS* genes were also constructed in the PW151 (MC1061 *degP41*) strain background. The resulting strains showed no obvious synthetic phenotype. All were temperature sensitive (the phenotype of the *degP41* mutation), and all strains carrying a deletion in the *degS* gene showed the small colony phenotype.

The small colony *degS* mutant phenotype is only rescued by plasmids expressing the *degS* gene. The ability of plasmids pPW146 and pPW100, expressing the *degQ* and *degP* genes, to restore normal growth to the *degS* gene deletion strain PW149 was tested (Table 2). A plasmid shuffle experiment described in the Materials and Methods section showed that neither plasmid rescues the small colony phenotype of PW149, whereas plasmid pPW147 expressing the *degS* gene from the same promoter restores normal colony size (Fig. 7B). None of these plasmids affected the colony size of strain MC1061.

DISCUSSION

In this work, we have identified and characterized two homologs of the *degP* gene. These genes, *degQ* and *degS*, are tandemly arranged at 72-73 min on the *E. coli* chromosome. The *degP* gene is located at 4 min.

The *degQ* gene is immediately upstream of the *degS* gene, and both genes are transcribed in the same direction. However, our results suggest that the two genes can be transcribed from independent promoters. A *degQ* transcript of ~ 1.5 kb and a *degS* transcript of ~ 1.1 kb are seen on Northern blots, and for each gene, primer extension identifies a putative transcription start site downstream from sequences characteristic of promoters recognized by a σ^{70} containing RNA polymerase. Furthermore, insertion of the kanamycin resistance element into the *degQ* gene does not result in the phenotype associated with a deletion in the *degS* gene. This indicates that the strong transcription termination signals flanking the kanamycin resistance gene do not exert a polar effect on the downstream *degS* gene, implying that *degS* transcription is not dependent on *degQ* transcription. Neither gene is heat-inducible, which is consistent with the presence of sequences characteristic of promoters recognized by a σ^{70} containing RNA polymerase, and the absence of any sequences characteristic of heat-inducible promoters. This is in contrast with the *degP* promoter which is heat-inducible and is recognized by σ^{E} (8, 16, 28). Clearly, our results do not rule out the possibility of other promoters which could regulate the expression of *degQ* and *degS* under conditions not tested here.

Several lines of evidence suggest that DegQ is a periplasmic serine endoprotease. Cleavage of β-casein by DegQ generates several large polypeptide fragments, in a reaction that can be inhibited by DFP. In addition, the DegQ signal peptide is removed post-translationally suggesting that DegQ is exported from the cytoplasm. Genetically, we have shown that DegQ can substitute for another periplasmic protease, DegP. In addition, Bass has demonstrated that DegQ (HhoA) can rescue the temperature/osmotic sensitive phenotype of a strain lacking the periplasmic Tsp (Prc) protease (1). The physiological basis for the conditional lethality of *prc* deletion strains is not understood, and there is no sequence homology between Tsp and DegP, DegQ or DegS. Nevertheless, the results of Bass indicate that these enzymes may be capable of cleaving some of the same substrates.

The ability of DegQ to rescue the temperature sensitive phenotype of the *degP41* strain indicates that DegQ is a periplasmic protease but also implies that DegQ shares some functional features with DegP (such as the ability to degrade the same substrates). This is not surprising, given the high degree of similarity between the two proteins. It is not clear whether higher intracellular levels of DegQ than DegP are required to rescue the *degP41* temperature sensitive defect. The amount of DegQ expressed from the chromosomal copy of the *degQ* gene is not sufficient to allow growth at elevated temperature in the absence of the *degP* gene. However, low levels of DegP may not support growth at high temperatures either. Under heat-shock conditions (when transcription of the *degP* gene is strongly induced), the intracellular level of DegP may be similar to the level of DegQ from an expression plasmid like pPW146.

The *degQ* gene is not important for normal *E. coli* growth. Furthermore, strains containing deletions in the *degQ* gene showed no obvious phenotype under a variety of growth conditions, including growth at 44 °C. This is in contrast to the *degP41* strain which does not grow at 44 °C. Several proteases, including DegP homologs in *Salmonella typhimurium* and *Brucella abortus*, have been implicated in the pathogenic virulence of these organisms (2, 13, 22). The DegQ protein might play a similar role in interactions between *E. coli* and a host organism.

The *degS* gene is important for normal *E. coli* growth. Strains containing a deletion in the *degS* gene have a small colony phenotype under all growth conditions tested. The DegS protein is smaller than either of the DegQ or DegP proteins and does not share the same overall degree of sequence homology, but the strong sequence conservation within the regions surrounding the putative catalytic triad residues suggests that DegS is also a serine protease. However, DegS must differ in some functional way from DegQ and DegP. A plasmid expressing the *degS* gene rescues the small colony phenotype of a *degS* mutant strain, but does not rescue the temperature sensitive *degP41* growth defect. The opposite is true of plasmids expressing the *degQ* or *degP* genes. They complement the *degP* mutant phenotype, but do not complement the *degS* mutant phenotype. The functional differences between DegS and DegQ/DegP could be caused by differences in protease activity or substrate specificity or by different subcellular localization. In contrast to DegP and DegQ which are predominantly in the soluble periplasmic fraction, the DegS protein is located almost entirely in the insoluble/membrane pellet in cellular fractionation experiments (E. Roche,

personal communication; unpublished results). In addition, DegP and DegQ, unlike DegS, are processed by the removal of amino-terminal signal sequences. The ability of DegS (HhoB) to substitute for Tsp (1), suggests that DegS is probably located on the periplasmic side of the membrane.

ACKNOWLEDGEMENTS

We thank Ken Rudd for providing us with Kohara phage 526, Steve Bass for discussing his results, and members of the Sauer lab for helpful advice. This work was supported by NIH grant AI-16892.

REFERENCES

1. **Bass, S. H., Q. Gu, and A. Christen.** 1996. Multicopy suppressors of Prc mutant *Escherichia coli* include two Htr (DegP) protease homologs (HhoAB), DksA, and RlpA. J. Bacteriol. **178:**1154-1161.

2. **Baumler, A. J., J. G. Kusters, I. Stojiljkovic, and F. Heffron.** 1994. *Salmonella typhimurium* loci involved in survival within macrophages. Infect. Immun. **62:**1623-1630.

3. **Bazan, J. F., and R. J. Fletterick.** 1990. Structural and catalytic models of trypsin-like viral proteases. Seminars in Virology **1:**311-322.

4. **Carafa, Y. d., E. Brody, and C. Thermes.** 1990. Prediction of Rho-independent *Escherichia coli* transcription terminators. J. Mol. Biol. **216:**835-858.

5. **Cassadaban, M. J., and S. N. Cohen.** 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli.* J. Mol. Biol. **138:**179-207.

6. **Chomczynski, P.** 1992. One-hour downward alkaline capillary transfer for blotting of DNA and RNA. Anal. Biochem. **201:**134-139.

7. **De Lorenzo, V., M. Herrero, U. Jakubzik, and K. N. Timmis.** 1990. Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. J. Bacteriol. **172:**6568-6572.

8. **Erickson, J. W., and C. A. Gross.** 1989. Identification of the σ^E subunit of *Escherichia coli* RNA polymerase: a second alternate σ factor involved in high-temperature gene expression. Genes Dev. **3:**1462-1471.

9. **Gross, C., M. Lonetto, and R. Losick.** 1992. Bacterial sigma factors, p. 129-176. *In* S. McKight and K. Yamamoto (ed.), Transcriptional Regulation, vol 1. Cold Spring Harbor Press, Cold Spring Harbor, N. Y.

10. **Hamilton, C. M., M. Aldea, B. K. Washburn, P. Babitzke, and S. R. Kushner.** 1989. New Method for Generating Deletions and Gene Replacements in *Escherichia coli.* J. Bacteriol. **171:**4617-4622.

11. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoters. Nucleic Acids Res. **15:**2343-2361.

12. **Jain, R. G., S. L. Rusch, and D. A. Kendall.** 1994. Signal peptide cleavage regions. J. Biol. Chem. **269:**16305-16310.

13. **Johnson, K., I. Charles, G. Dougan, D. Pickard, P. O'Gaora, G. Costa, T. Ali, I. Miller, and C. Hormaeche.** 1991. The role of a stress-response protein in *Salmonella typhimurium* virulence. Mol. Microbiol. **5:**401-407.

14. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495-508.

15. **Lipinska, B., O. Fayet, L. Baird, and C. Georgopoulos.** 1989. Identification, characterization, and mapping of the *Escherichia coli htrA* gene, whose product is essential for bacterial growth only at elevated temperatures. J. Bacteriol. **171:**1574-1584.

16. **Lipinska, B., S. Sharma, and C. Georgopoulos.** 1988. Sequence analysis and regulation of the *htrA* gene of *Escherichia coli*: a σ^{32}-independent mechanism of heat-inducible transcription. Nucleic Acids Res. **16:**10053-10067.

17. **Lipinska, B., M. Zylicz, and C. Georgopoulos.** 1990. The HtrA (DegP) protein, essential for *Escherichia coli* survival at high temperatures, is an endopeptidase. J. Bacteriol. **172:**1791-1797.

18. **Maurizi, M. R.** 1992. Proteases and protein degradation in *Escherichia coli.* Experimentia **48:**178-201.

19. **McClure, W. R.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. **54:**171-204.

20. **Oliver, D.** 1985. Protein secretion in *Escherichia coli.* Annu. Rev. Microbiol. **39:**615-648.

21. **Reynolds, R., R. M. Bermudez-Cruz, and M. J. Chamberlin.** 1992. Parameters affecting transcription termination by *Escherichia coli* RNA polymerase. J. Mol. Biol. **224:**31-51.

22. **Roop, R. M., T. W. Fletcher, N. M. Sriranganathan, S. M. Boyle, and G. G. Schurig.** 1994. Identification of an immunoreactive *Brucella abortus* HtrA stress response protein homolog. Infect. Immun. **62:**1000-1007.

23. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342-1346.

24. **Skorko-Glonek, J., K. Krzewski, B. Lipinska, E. Bertoli, and F. Tanfani.** 1995. Comparison of the structure of wild-type HtrA heat shock protease and mutant HtrA proteins. J. Biol. Chem. **270:**11140-11146.

25. **Strauch, K. L., and J. Beckwith.** 1988. An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins. Proc. Natl. Acad. Sci. USA **85:**1576-1580.

26. **Strauch, K. L., K. Johnson, and J. Beckwith.** 1989. Characterization of *degP*, a gene required for proteolysis in the cell envelope and essential for growth of *Escherichia coli* at high temperature. J. Bacteriol. **171:**2689-2696.

27. **Vogel, R. F., K. D. Entian, and D. Mecke.** 1987. Cloning and sequence of the *mdh* structural gene of *Escherichia coli* coding for malate dehydrogenase. Arch. Microbiol. **149:**36-42.

28. **Wang, Q. P., and J. M. Kaguni.** 1989. A novel sigma-factor is involved in expression of the *rpoH* gene of *Escherichia coli*. J. Bacteriol. **171:**4248-4253.

29. **Waugh, D. S., and R. T. Sauer.** 1993. Single amino acid substitutions uncouple the DNA binding and strand scission activities of FokI endonuclease. Proc. Natl. Acad. Sci. USA **90:**9596-9600.

Table 1. Bacterial strains.

Strain	Genotype	Source
KS474	KS272 *degP41*	(26)
MC1061	*hsdR2 araD139 galE15 galK16 rpsL*	(5)
PW147	MC1061 Δ*degQ1* (ΔEcoRV)	This study
PW148	MC1061 Δ*(degQ-degS)1* (ΔEcoRV-ScaI)	This study
PW149	MC1061 Δ*degS1* (ΔBamHI-ScaI)	This study
PW151	MC1061 *degP41*	This study
PW152	MC1061 Δ*degQ2*::*kan* (ΔEcoRV::*kan*)	This study

Table 2. The ability of plasmids expressing different *degP* homologs to rescue the temperature sensitive phenotype of PW151 (*degP41*), and the small colony phenotype of PW149 (Δ*degS1*).

Plasmid	Growth at 44 °C		Colony size	
	PW151 (*degP41*)	MC1061	PW149 (Δ*degS1*)	MC1061
Untransformed	No	Yes	Small	Normal
pPW100 (*degP*)	Yes	Yes	Small	Normal
pPW146 (*degQ*)	Yes	Yes	Small	Normal
pPW147 (*degS*)	No	Yes	Normal	Normal
pPW105 (*degP236*)	No	Yes	not tested	Normal

Figure 1. (**A**) Restriction map of the *degQ* and *degS* genes. Restriction enzymes: B, BamHI; E, EcoRI; EV, EcoRV; H, HindIII; K, KpnI; P, PstI; S, ScaI. (**B**) Chromosomal gene deletions (construction details in Materials and Methods). Strain PW147 has the 420 bp EcoRV fragment deleted (stippled region), removing the putative active site Ser of DegQ. The remaining 5' and 3' fragments of the *degQ* gene are in frame (solid regions). Strain PW148 has the 1,504 bp EcoRV-ScaI fragment deleted (stippled region), removing the putative active site Ser of DegQ, and the putative active site His and Asp of DegS. The remaining 3' fragment of the *degS* gene (hatched region) is out of frame with the 5' end of the *degQ* gene. Strain PW149 has the 222 bp BamHI-ScaI fragment deleted (stippled region), removing the putative active site His and Asp of DegS. The remaining 3' fragment (hatched region) is out of frame with the 5' end of the *degS* gene. Strain PW152 contains a kanamycin resistance gene (shaded region) in place of the 420 bp EcoRV fragment.

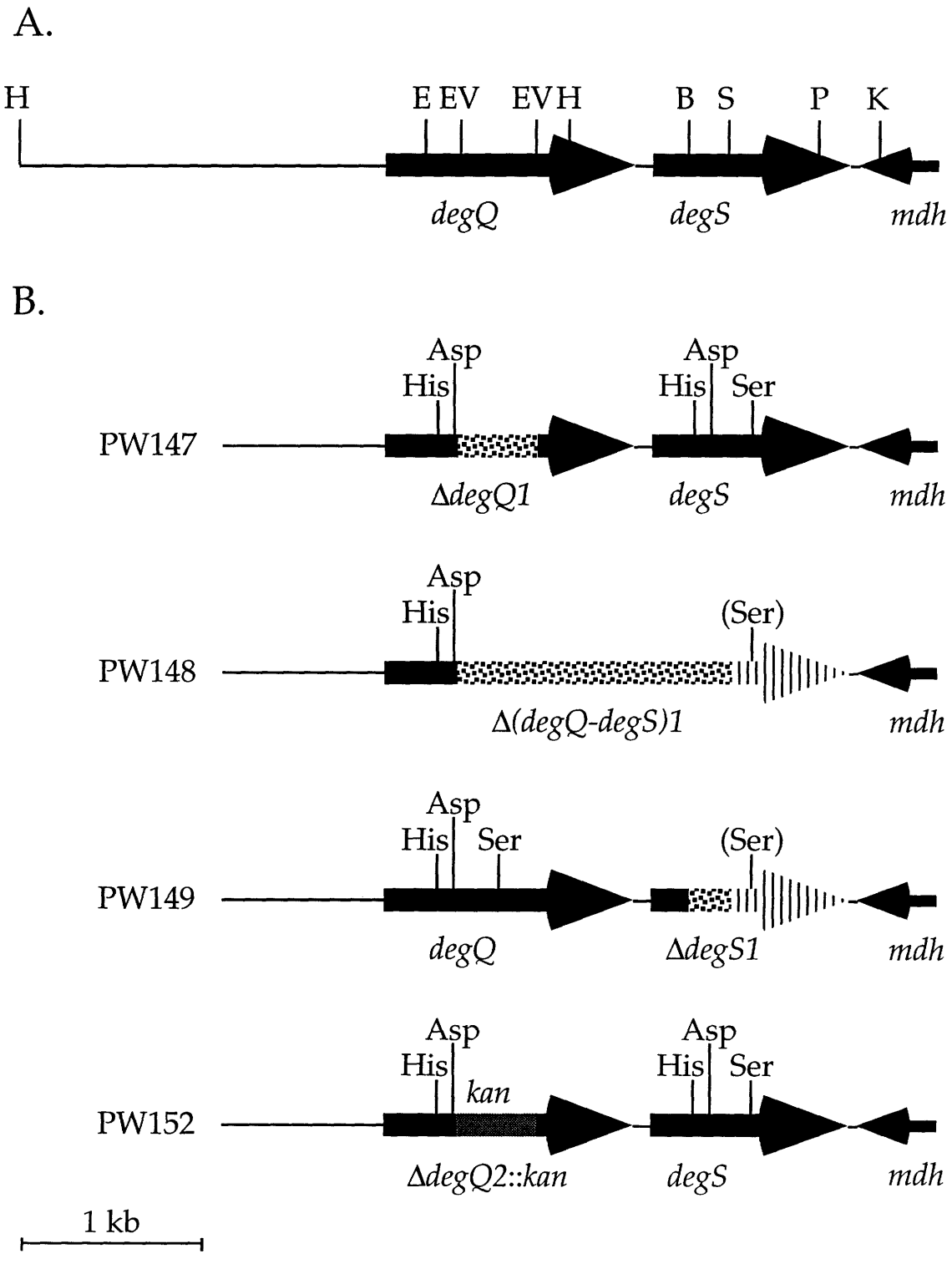

A.
H
E EV EVH
B S
P
K
degQ
degS
mdh
B.
PW147
Asp
His
Asp
His Ser
ΔdegQ1
degS
mdh
PW148
Asp
His
(Ser)
Δ(degQ-degS)1
mdh
PW149
Asp
His Ser
(Ser)
degQ
ΔdegS1
mdh
PW152
Asp
His
kan
Asp
His Ser
ΔdegQ2::kan
degS
mdh
1 kb

Figure 2. The DNA and deduced amino acid sequences of the *degQ* and *degS* genes. At the 5' end of both genes putative -35 and -10 promoter elements are indicated by bold, italicized type. Candidate Shine Dalgarno (S-D) sequences are underlined. At the 3' end of both genes the lowercase, underlined sequences are capable of forming stem-loop structures characteristic of rho-independent transcription terminators: two stem-loops followed by a series of T residues at the 3' end of *degQ*, and one stem-loop containing a series of T residues at the 3' end of *degS*. Note that the putative *degQ* transcription terminator overlaps with the putative *degS* promoter.

```
1     CCATTGCGCC CTCCTCTTCT CTCCCTCCCC GACTATCATT TAATCTGGTG TCTCATTGTT AGCCGTCTGA AAATTCAATA ACATCAAACT
                                                                  -35                         -10
                                                  M K   K Q T   Q L L S   A L A   L S V
91    GTTTTGAATC TCTTTTCTTA TCATTCAGGT ACGAGAGCAG GAATAATGAA AAAACAAACC CAGCTGTTGA GTGCATTAGC GTTAAGTGTC
                                           S-D       degQ
      G L T L   S A S   F Q A   V A S I   P G Q   V A D   Q A P L   P S L   A P M
181   GGGTTAACTC TCTCGGCGTC ATTTCAGGCC GTCGCGTCGA TTCCAGGCCA GGTTGCCGAT CAGGCCCCTC TCCCCAGTCT GGCTCCAATG
      L E K V   L P A   V V S   V R V E   G T A   S Q G   Q K I P   E E F   K K F
271   CTGGAAAAAG TGCTTCCGGC AGTGGTGAGC GTACGGGTGG AAGGAACGGC CAGTCAGGGA CAGAAAATCC CGGAAGAATT CAAAAAGTTT
      F G D D   L P D   Q P A   Q P F E   G L G   S G V   I I N A   S K G   Y V L
361   TTTGGTGATG ATTTACCGGA TCAACCTGCA CAACCCTTCG AAGGTTTAGG CTCCGGTGTC ATCATCAACG CCAGTAAAGG CTATGTGCTG
      T N N H   V I N   Q A Q   K I S I   Q L N   D G R   E F D A   K L I   G S D
451   ACCAACAACC ATGTGATTAA TCAGGCACAG AAAATCAGTA TTCAGCTCAA TGATGGGCGC GAGTTTGATG CAAAACTGAT TGGTAGCGAT
      D Q S D   I A L   L Q I   Q N P S   K L T   Q I A   I A D S   D K L   R V G
541   GACCAGAGCG ATATCGCCCT GTTACAAATT CAAAACCCGA GCAAATTAAC GCAAATCGCT ATTGCCGACT CCGATAAATT GCGCGTCGGT
      D F A V   A V G   N P F   G L G Q   T A T   S G I   V S A L   G R S   G L N
631   GATTTTGCCG TAGCGGTCGG TAACCCATTT GGCCTTGGGC AAACCGCCAC CTCTGGCATT GTTTCCGCAT TAGGCCGCAG CGGGTTGAAT
      L E G L   E N F   I Q T   D A S I   N R G   N S G   G A L L   N L N   G E L
721   CTTGAAGGTC TGGAAAACTT TATCCAGACA GATGCTTCCA TTAACCGCGG TAACTCCGGC GGTGCACTAT TAAACCTTAA CGGTGAGTTA
      I G I N   T A I   L A P   G G G S   V G I   G F A   I P S N   M A R   T L A
811   ATTGGCATCA ACACTGCAAT CCTTGCGCCT GGCGGCGGGA GCGTCGGGAT TGGATTTGCC ATCCCCAGTA ATATGGCGCG AACACTGGCG
      Q Q L I   D F G   E I K   R G L L   G I K   G T E   M S A D   I A K   A F N
901   CAGCAGCTTA TCGACTTTGG TGAAATCAAA CGCGGTTTGT TAGGCATCAA AGGCACCGAG ATGAGTGCCG ATATCGCCAA AGCCTTCAAC
      L D V Q   R G A   F V S   E V L P   G S G   S A K   A G V K   A G D   I I T
991   CTTGACGTGC AGCGTGGCGC GTTTGTCAGC GAAGTGTTGC CAGGTTCTGG CTCGGCAAAA GCGGGCGTCA AAGCGGGCGA TATTATTACC
      S L N G   K P L   N S F   A E L R   S R I   A T T   E P G T   K V K   L G L
1081  AGCCTCAACG GCAAACCGCT GAATAGCTTT GCTGAGTTGC GCTCTCGTAT CGCGACCACC GAGCCGGGCA CGAAAGTGAA GCTTGGCCTG
      L R N G   K P L   E V E   V T L D   T S T   S S S   A S A E   M I T   P A L
1171  CTGCGTAACG GCAAACCACT GGAAGTAGAA GTGACGCTCG ATACCAGCAC CTCTTCGTCG GCCAGCGCTG AAATGATCAC GCCAGCGCTG
      E G A T   L S D   G Q L   K D G G   K G I   K I D   E V V K   G S P . A A Q
1261  GAAGGTGCAA CGTTGAGCGA TGGTCAGCTA AAAGATGGCG GCAAAGGTAT TAAAATCGAT GAAGTTGTCA AAGGAAGCCC AGCTGCTCAG
      A G L Q   K D D   V I I   G V N R   D R V   N S I   A E M R   K V L   A A K
1351  GCTGGCTTGC AAAAAGACGA TGTGATCATT GGCGTCAACC GCGATCGGGT GAACTCGATT GCTGAAATGC GTAAAGTGCT GGCGGCAAAA
      P A I I   A L Q   I V R   G N E S   I Y L   L M R   *
1441  CCGGCCATCA TCGCCCTGCA AATTGTACGC GGCAATGAAA GCATCTATCT GCTGATGCGT TAATGTCgta aaccgggcat caggcTTACg
                                                                M F V   K L L   R S V A
1531  tgtgatgtcc ggttaacTCg tggtatgctg Ctgccgttcc cTTTTTTAAT GACGCCTCCA TCATGTTTGT GAAGCTCTTA CGTTCCGTTG
       -35                -10                             S-D       degS
       I G L   I V G   A I L L   V A M   P S L   R S L N   P L S   T P Q   F D S T
1621  CGATTGGATT AATTGTCGGC GCTATTCTGC TGGTTGCCAT GCCTTCGCTG CGCAGCCTTA ACCCGCTTTC CACTCCGCAA TTTGACAGTA
       D E T   P A S   Y N L A   V R R   A A P   A V V N   V Y N   R G L   N T N S
1711  CCGATGAGAC GCCTGCCAGC TATAATCTGG CGGTTCGCCG CGCCGCGCCA GCGGTGGTTA ACGTTTACAA CCGTGGTTTG AACACCAACT
       H N Q   L E I   R T L G   S G V   I M D   Q R G Y   I I T   N K H   V I N D
1801  CTCACAACCA GCTTGAGATC CGCACCCTGG GATCCGGTGT AATCATGGAT CAACGCGGTT ATATCATCAC CAATAAACAC GTCATCAACG
       A D Q   I I V   A L Q D   G R V   F E A   L L V G   S D S   L T D   L A V L
1891  ACGCCGATCA GATCATCGTC GCCTTACAGG ATGGACGTGT ATTTGAAGCA TTGCTGGTGG GATCTGACTC TCTAACCGAT CTGGCGGTAC
       K I N   A T G   G L P T   I P I   N A R   R V P H   I G D   V V L   A I G N
1981  TTAAAATTAA TGCCACTGGC GGTTTACCTA CCATTCCAAT TAATGCACGT CGCGTACCGC ACATTGGCGA CGTAGTACTG GCGATCGGTA
       P Y N   L G Q   T I T Q   G I I   S A T   G R I G   L N P   T G R   Q N F L
2071  ACCCGTACAA CCTCGGGCAG ACCATTACCC AGGGGATTAT TAGTGCCACG GGTCGAATCG GTCTGAACCC GACCGGGCGG CAAAACTTCC
       Q T D   A S I   N H G N   S G G   A L V   N S L G   E L M   G I N   T L S F
2161  TCCAAACCGA TGCTTCCATT AACCACGGTA ACTCTGGCGG CGCGCTGGTG AACTCGCTGG GCGAACTGAT GGGCATTAAT ACGCTGTCGT
       D K S   N D G   E T P E   G I G   F A I   P F Q L   A T K   I M D   K L I R
2251  TTGATAAGAG TAACGATGGC GAAACGCCGG AAGGTATCGG CTTTGCGATT CCTTTCCAGT TAGCAACCAA AATTATGGAT AAGCTGATCC
       D G R   V I R   G Y I G   I G G   R E I   A P L H   A Q G   G G I   D Q L Q
2341  GCGATGGTCG CGTGATCCGC GGCTACATTG GTATCGGCGG ACGTGAGATC GCACCACTGC ACGCGCAGGG CGGTGGTATA GATCAACTGC
       G I V   V N E   V S P D   G P A   A N A   G I Q V   N D L   I I S   V D N K
2431  AAGGGATCGT GGTTAATGAA GTGTCACCTG ACGGCCCGGC GGCGAATGCG GGTATTCAGG TCAACGATCT GATTATTTCG GTGGATAACA
       P A I   S A L   E T M D   Q V A   E I R   P G S V   I P V   V V M   R D D K
2521  AACCGGCCAT CTCTGCTCTG GAGACGATGG ATCAGGTGGC GGAAATTCGC CCTGGTTCGG TGATCCCTGT AGTAGTGATG CGTGATGATA
       Q L T   L Q V   T I Q E   Y P A   T N *
2611  AGCAGTTAAC GCTGCAGGTC ACCATTCAGG AATATCCGGC AACCAATTAA GTCGTGCGCT CAAaacaaaa aaccggagTC TGTGctccgg

2701  tttttattA TCCG
```

Figure 3. Alignment of the DegP, DegQ and DegS protein sequences. Positions that are identical in the different proteins are shaded. The putative catalytic triad residues (H, His; D, Asp; S, Ser) are boxed.

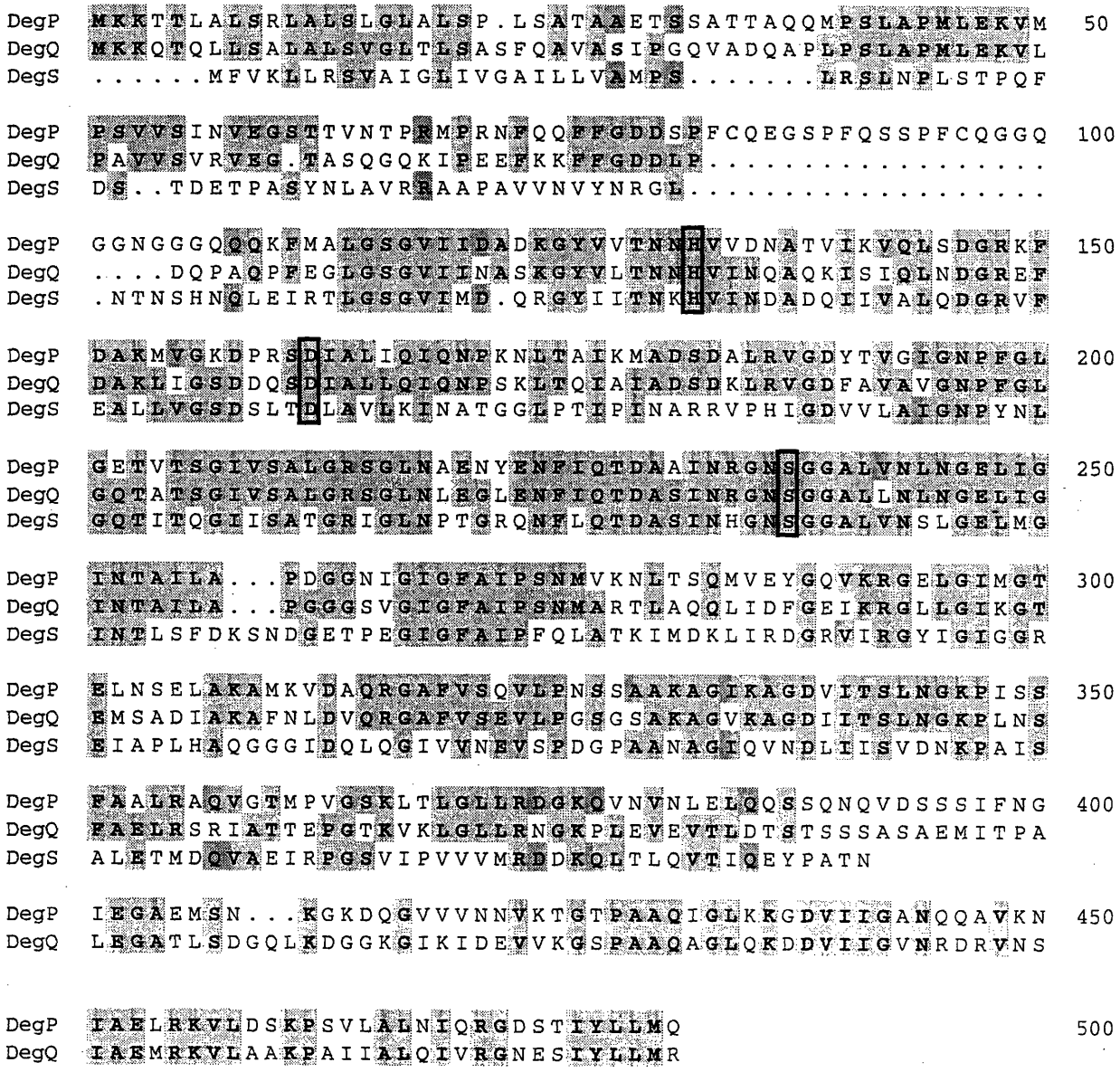
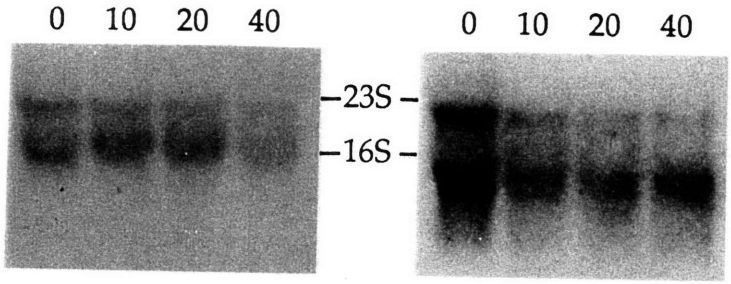

DegP MKKTTLALSRLALSLGLALSP.LSATAAETSSATTAQQMPSLAPMLEKVM 50
DegQ MKKQTQLLSALALSVGLTLSASFQAVASIPGQVADQAPLPSLAPMLEKVL
DegS ......MFVKLLRSVAIGLIVGAILLVAMPS.....LRSLNPLSTPQF
DegP PSVVSINVEGSTTVNTPRMPRNFQQFFGDDSPFCQEGSPFQSSPFCQGGQ 100
DegQ PAVVSVRVEG.TASQGQKIPEEFKKFFGDDLP.....
DegS DS..TDETPASYNLAVRRAAPAVVNVYNRGL.....
DegP GGNGGGQQQKFMALGSGVIIDADKGYVVTNNHVVDNATVIKVQLSDGRKF 150
DegQ ....DQPAQPFEGLGSGVIINASKGYVLTNNHVINQAQKISIQLNDGREF
DegS .NTNSHNQLEIRTLGSGVIMD.QRGYIITNKHVINDADQIIVALQDGRVF
DegP DAKMVGKDPRSDIALIQIQNPKNLTAIKMADSDALRVGDYTVGIGNPFGL 200
DegQ DAKLIGSDDQSDIALLQLQNPSKLTQLAIADSDKLRVGDFAVAVGNPFGL
DegS EALLVGSDSLTDLAVLKINATGGLPTIPINARRVPHIGDVVLAIGNPYNL
DegP GETVTSGIVSALGRSGLNAENYENFIQTDAAINRGNSGGALVNLNGELIG 250
DegQ GQTATSGIVSALGRSGLNLEGLENFIQTDASINRGNSGGALLNLNGELIG
DegS GQTITQGIISATGRIGLNPTGRQNFLQTDASINHGNSGGALVNSLGELMG
DegP INTAILA...PDGGNIGIGFAIPSNMVKNLTSQMVEYGQVKRGELGIMGT 300
DegQ INTAILA...PGGGSVGIGFAIPSNMARTLAQQLIDFGEIKRGLLGIKGT
DegS INTLSFDKSNDGETPEGIGFAIPFQLATKIMDKLIRDGRVIRGYIGIGGR
DegP ELNSELAKAMKVDAQRGAFVSQVLPNSSAAKAGIKAGDVITSLNGKPISS 350
DegQ EMSADIAKAFNLDVQRGAFVSEVLPGSGSAKAGVKAGDIITSLNGKPLNS
DegS EIAPLHAQGGGIDQLQGIVVNEVSPDGPAANAGIQVNDLIISVDNKPAIS
DegP FAALRAQVGTMPVGSKLTLGLLRDGKQVNVNLELQQSSQNQVDSSSIFNG 400
DegQ FAELRSRIATTEPGTKVKLGLLRNGKPLEVEVTLDTSTSSSASAEMITPA
DegS ALETMDQVAEIRPGSVIPVVVMRDDKQLTLQVTIQEYPATN
DegP IEGAEMSN...KGKDQGVVVNNVKTGTPAAQIGLKKGDVIIGANQQAVKN 450
DegQ LEGATLSDGQLKDGGKGIKIDEVVKGSPAAQAGLQKDDVIIGVNRDRVNS
DegP IAELRKVLDSKPSVLALNIQRGDSTIYLLMQ 500
DegQ IAEMRKVLAAKPAIIALQIVRGNESIYLLMR

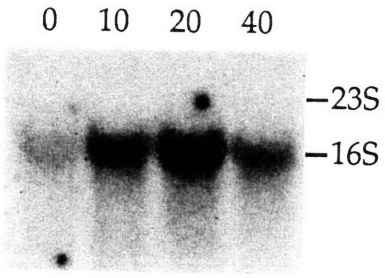
Figure 4. (**A**) Northern analysis of transcripts from the *degQ* and *degS* genes at 0, 10, 20 and 40 min after transferring cell cultures from 30 °C to 44 °C. The probes were ^{32}P-labeled EcoRI-HindIII and BamHI-PstI fragments from the *degQ* and *degS* genes, respectively. (**B**) The blot that was probed for *degQ* RNA was stripped (boiling 0.5x SSC, 0.5% SDS) and reprobed with a ^{32}P-labeled fragment derived from the *degP* gene. The migration of the 16S (1541 nucleotides) and 23S (2904 nucleotides) ribosomal RNAs is indicated.

A.



degQ *degS*

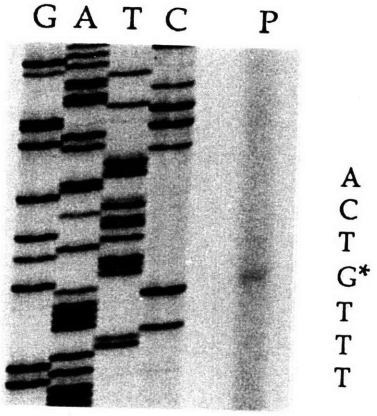
B.



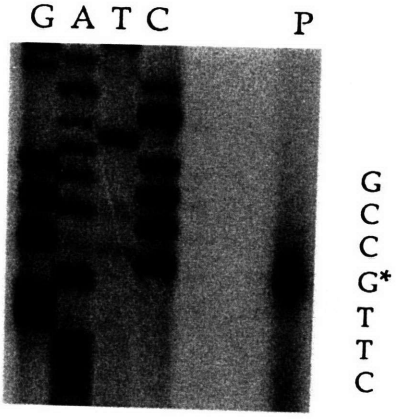
degP

Figure 5. Primer extension analysis of transcripts from (**A**) the *degQ* and (**B**) the *degS* genes. The primer extension reactions are run in the P lanes. The sequences of the template DNAs are run in the G, A, T, and C lanes. The sequences of the sense strands are shown for the regions surrounding the transcription start sites. For both genes, the nucleotide at the +1 position is indicated by an asterisk. (**C**) The sequences of the -35 and -10 regions identified by primer extension. The capitalized residues are identical to the consensus sequence for promoters recognized by σ^{70}.

A.



B.



C.

degQ 5' tcaTTGttAgccgtctgaaaattcaaTAacATcaaactGttt 3'
-35 -10 +1

degS 5' cgtgTGAtgtccggttaactcgtggTATgcTgctgccGttc 3'
-35 -10 +1

Figure 6. Proteolytic cleavage of β-casein by DegQ in the presence of different concentrations of DFP (0, 0.1, 1, 5, 10, and 20 mM). In lane C β-casein was incubated without DegQ. Reaction conditions are described in Materials and Methods. The molecular masses of the protein standards in lane M are 21.5, 30, and 46 kilodaltons.

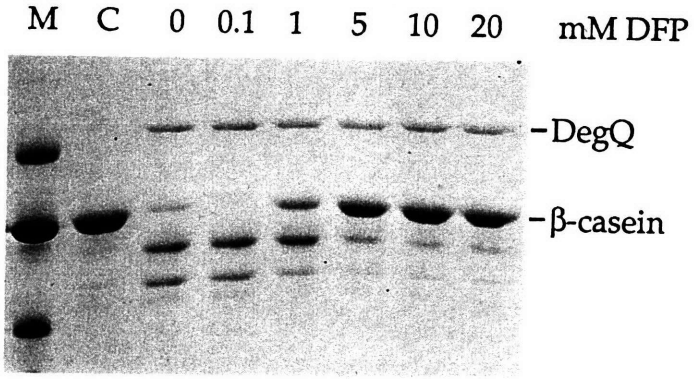
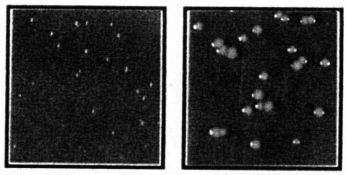

M C 0 0.1 1 5 10 20 mM DFP
DegQ
β-casein

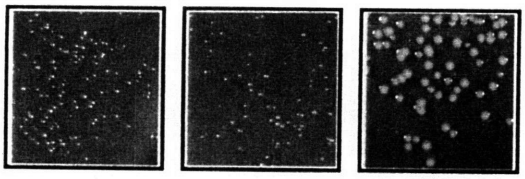
Figure 7. (**A**) The small colony phenotype of strain PW149 (Δ*degS1*). (**B**) Plasmids pPW100 and pPW146, expressing *degP* and *degQ*, respectively, do not restore normal colony size to PW149. Plasmid pPW147, expressing *degS* from the same promoter, restores normal colony size to PW149.

A.



PW149 MC1061

B.



PW149/ pPW100 (*degP*)

PW149/ pPW146 (*degQ*)

PW149/ pPW147 (*degS*)

CHAPTER 3

The DegP and DegQ Periplasmic Endoproteases of Escherichia coli: Specificity for Cleavage Sites and Substrate Conformation

Kolmar, H., Waller, P. R. H., and Sauer, R. T. 1996. The DegP and DegQ Periplasmic Endoproteases of *Escherichia coli*: Specificity of Cleavage Sites and Substrate Conformation. J. Bacteriol. *submitted*

INTRODUCTION

Intracellular proteases play important regulatory roles and also serve essential housekeeping functions by removing damaged or misfolded proteins (Maurizi, 1992). *Escherichia coli* contains at least two cytoplasmic proteases, Lon and Clp, which function to degrade abnormal proteins (Gottesman and Maurizi, 1992). The biochemical properties of these ATP-dependent proteases have been studied extensively (Goldberg, 1992). Relatively little, however, is known about the biochemical or structural properties of proteases which may be responsible for degradation of misfolded or abnormal proteins in the periplasmic compartment of bacteria. Such a function has been attributed to the periplasmic protease DegP, which is also known as HtrA or protease Do (Strauch and Beckwith, 1988; Lipinska et al., 1988; Strauch et al., 1989; Seol et al., 1991). DegP is required for survival of *E. coli* at elevated temperatures, and mutations in the *degP* gene result in decreased degradation of chimeric membrane and periplasmic proteins (Strauch and Beckwith, 1988; Strauch et al., 1989; Lipinska et al., 1989). The temperature sensitivity of *degP* mutants is reduced in strains which release periplasmic proteins into the medium because of outer-membrane defects, as expected if DegP is required for the removal of misfolded proteins which may accumulate at high temperatures (Strauch and Beckwith, 1988; Lipinska et al., 1990).

Several groups have recently identified another periplasmic protease of *E. coli*, DegQ (HhoA), which is homologous to DegP (Waller and Sauer, 1996; Bass et al., 1996; Blattner et al., unpublished). The DegQ and DegP proteins are

of similar size (455 and 474 residues, respectively) and display approximately 60% sequence identity. Overproduction of DegQ suppresses the temperature-sensitive defect of a *degP*$^-$ strain suggesting that the two enzymes must be capable of degrading similar substrates (Waller and Sauer, 1996). Based on sequence conservation, inhibition by diisopropyl fluorophosphate, and mutagenesis experiments, both enzymes seem to contain the Ser-His-Asp catalytic triad found in traditional serine proteases (Swamy et al., 1983; Skorko-Glonek et al., 1995; Waller and Sauer, 1996). Serum albumin, ovalbumin, growth hormone and a number of other proteins are not cleaved by DegP *in vitro*, nor are any of 20 synthetic peptides routinely used as protease substrates (Lipinska et al., 1990) but little else is known about the substrate preferences of either DegP or DegQ. In this work, we determine the cleavage-site specificities of DegP and DegQ for several model protein substrates *in vitro* and investigate the structural properties of these substrates that determine whether they will be cleaved by DegP and DegQ. We also show that both proteases can be crosslinked to dodecamers in solution.

MATERIALS AND METHODS

Bacterial strains and plasmids. The over-expressing plasmid vectors pPW100 and pPW146 contain the *degP* and *degQ* genes, respectively, under transcriptional control of a plasmid borne P_{trc} promoter (Waller and Sauer, 1996). *E. coli* strains PW147 (MC1061 Δ*degQ1*) and PW151 (MC1061 *degP41*) were used for DegP and DegQ production, respectively (Waller and Sauer, 1996). Bacteria were grown in LB medium (10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl per liter).

Protein purification, modification, and crosslinking. For DegP protein production, *E. coli* strain PW147/pPW100 was grown at 30 °C in 6 L of LB broth containing ampicillin (100 mg/L) to an A_{600} of 0.6. The cells were induced with 1 mM IPTG, grown for 3 h, harvested by centrifugation, resuspended in 80 ml lysis buffer (100 mM Tris-Cl pH 8.0, 200 mM KCl, 1 mM EDTA, 14 mM β-mercaptoethanol), and lysed by sonication. The lysate was diluted to 200 mL by addition of lysis buffer, polyethyleneimine was added to 1%, and after stirring for 45 min, the suspension was cleared by centrifugation. The supernatant was precipitated with ammonium sulfate added to 90% saturation, and the pellet was resuspended in 40 mL of 20 mM Tris-Cl (pH 8.0) and dialyzed against the same buffer. This solution was passed through a 50 mL DEAE-Sephacel column (Pharmacia) and the flow-through fraction was collected. This was dialyzed against 20 mM potassium phosphate (pH 6.4) and applied onto a 50 mL CM-Sepharose column (Pharmacia). After washing with loading buffer, bound proteins were eluted with successive steps increased each time by 40 mM KCl. DegP eluted at approximately 80 mM, and was dialyzed against 20 mM Tris-Cl (pH 7.4), 100 mM KCl. Following concentration by ultrafiltration (Centriprep 30), the material was loaded onto a 1 x 30 cm Superose 6 gel-filtration column (Pharmacia). DegP eluted from this column near the void volume, ahead of the remaining contaminants. After concentration of the protein solution by ultrafiltration, glycerol was added to 20% (v/v) and the material was stored in aliquots at -80 °C.

The DegQ protease was purified from *E. coli* strain PW151/pPW146 by a procedure similar to that described above involving polyethylenimine and ammonium sulfate precipitation, ion-exchange chromatography on DEAE-

Sephacel and CM-Sepharose, and gel filtration chromatography on Sephacryl S300 (Waller and Sauer, 1996). As judged by SDS-PAGE, both DegP and DegQ were greater than 95% pure following the gel-filtration column. In one DegP preparation, several minor contaminating bands of lower molecular weight were excised from the gel and identified as proteolytic fragments of DegP by N-terminal sequencing.

Variants of the phage P22 Arc repressor, including Arc-*st6* , Arc-*st11* and alanine-substitution mutants of these proteins were gifts of B. Brown and M. Milla (Brown et al., 1994; Milla et al., 1994). The *st6* tail encodes six C-terminal histidine residues, while the *st11* tail encodes the sequence $(His)_6$-Lys-Asn-Gln-His-Glu (Milla et al., 1993). The #105 variant of the N-terminal domain of λ repressor was kindly provided by K. Keiler. This protein consists of residues 1-97 of intact λ repressor followed by the sequence Trp-Val-Ala-Ala-Ala (Parsell et al., 1990). The β-lactamase/REI_v fusion protein was purified as described (Kolmar, 1992).

To reduce the disulfide bonds of the β-lactamase-REI fusion protein, a solution of 1.2 ml containing 0.25 mg of protein, 400 mM Tris-Cl (pH 8.6), 8 mM urea, 5 mM EDTA, and 0.11 M β-mercaptoethanol was incubated at 4 °C. After 2 h, 100 µL of 1.4 M iodoacetic acid (in 1 M NaOH) was added to alkylate the free cysteines and the solution was incubated at 4 °C for another h, followed by dialysis against 20 mM Tris-HCl (pH 7.4).

DSP [dithio*bis*(succinimidyl propionate)] (Pierce) was used as a crosslinking agent. Approximately 1.5 µg of DegP or DegQ protein in 24 µL of 20 mM Tris-Cl (pH 7.4), 100 mM KCl was mixed with 1 µL of a DSP solution (5

mg/mL in dimethylsulfoxide). After 15 min at room temperature, the crosslinking reaction was quenched by addition of 1/10 volume of 30 mM lysine, 250 mM Tris-Cl (pH 8). As a control, 1 μL of protein denaturation buffer (12.5 mM Tris, 100 mM glycine, 0.05% (w/v) SDS) was added to the protein solution prior to addition of DSP. The reaction mixture was electrophoresed through an acrylamide-agarose gel containing 2.2% (w/v) acrylamide, 0.2% (w/v) bisacrylamide, and 0.1% agarose (Baker and Mizuuchi, 1992).

Proteolysis assay. To assay DegP or DegQ cleavage of most protein substrates by gel electrophoresis, 50 μL of a solution containing 20 mM potassium phosphate (pH 7.4), 0.2 μM protease, and approximately 5 μM of the protein was incubated overnight at 37 °C and, 20 μL of the reaction products were electrophoresed through a 15% polyacrylamide-SDS-Tris-tricine gel (Schagger and von Jagow, 1987). Cleavage of the β-lacatamase/REI_v fusion protein was performed in 100 μl of 20 mM potassium phosphate (pH 7.4), 4 μM β-lacatamase/REI_v and 0.2 μM DegP or DegQ, respectively, with overnight incubation at 45 °C. To obtain relatively large quantities of cleavage fragments for protein chemistry, approximately 100 μM of the Arc-VA18-st6 variant or the λ-repressor #105 variant were dissolved in 150 μL of 20 mM potassium phosphate (pH 7.4), 1 μM DegP protease was added, and the mixture was incubated overnight at 37 °C.

HPLC analysis of the protease cleavage products was performed on a Beckman System Gold equipped with a diode-array detector. Solvents were A, 0.1% (v/v) TFA in water; and B, 0.1% (v/v) TFA in acetonitrile. All fractionations were performed using a C18 reverse phase column (3.9 x 150

mm, Waters Delta Pak) operating at a flow rate of 1 mL/min with a 3 min linear gradient from 0 to 15% solvent B beginning 2 min after injection, followed by a 30 min gradient from 15 to 60% solvent B, and a final 2 min gradient from 60 to 100% solvent B. Purified peptides were lyophilized and N-terminal sequences were determined by sequential Edman degradation by the MIT Biopolymers Laboratory.

RESULTS

We initially purified DegP from an *E. coli* strain containing chromosomal copies of the *degP* and *degQ* genes and found by N-terminal sequencing that the preparation was contaminated with a protein later shown to be DegQ. To avoid this problem in the studies reported here, DegP was purified from a strain containing the Δ*degQ1* deletion, and DegQ was purified from a strain containing the *degP41* deletion (Waller and Sauer, 1996). DegP and DegQ from these strains have similar but distinct properties during purification, and by this criteria would not be expected to copurify. As shown below, however, both proteins form large oligomers and thus in strains containing both proteases, contamination of one protein with the other may result because the two proteins form mixed oligomers.

DegP/DegQ protein substrates. Since natural biological substrates for DegP and DegQ are not known, we screened a collection of purified proteins to test if they were substrates for either protease. As noted previously, β-casein is a good substrate for both DegP and DegQ (Lipinska et al., 1990; Skorko-Glonek et al., 1995; Waller and Sauer, 1996). We also found that variants of Arc repressor and the N-terminal domain of λ-repressor were

cleaved by both proteases. By contrast, neither protease cleaved native bovine serum album, ovalbumin, carbonic anhydrase, or an N-terminal fusion of β-lactamase to the immunoglobulin domain REI_V (Kolmar, 1992). The last result was somewhat surprising as this chimeric protein is expressed at much higher levels in strains lacking functional DegP, suggesting that it is normally degraded by DegP *in vivo* (Kolmar, 1992). As we show later, this protein becomes susceptible to degradation by either DegP or DegQ if its disulfide bonds are broken.

As shown in Fig. 1, DegP and DegQ have very similar activities in cleaving Arc-*st6* and the #105 variant of the N-terminal domain of λ-repressor, producing cleavage products that are indistinguishable by HPLC. Time courses of cleavage were also similar for the two enzymes with both of these substrates (data not shown). This suggests that DegP and DegQ have very similar cleavage specificities, at least for these two protein substrates.

Sites of proteolytic cleavage. The #105 variant of λ-repressor and the VA18 mutant of Arc-*st6* were cleaved with DegP, the digestion products were purified by HPLC, and the N-terminal sequences of the major cleavage peptides were determined by Edman degradation. In all, seven different cleavage site sequences were identified (Table 1). Valine is the P1 residue in six of these sites and isoleucine, a closely related β-branched, aliphatic residue occupies the P1 position in the remaining site. For Arc, no peptide products resulting from cleavage after Val33 or Ile37 were recovered. Furthermore, the peptides resulting from cleavage after Val22 or Val25 contained Tyr38, indicating that the Val33-Asn34 and Ile37-Tyr38 peptide bonds are not cleaved by DegP. For λ repressor, no products resulting from cleavage after Ile21, 54,

69, 79 or Val73 were recovered. These results show that not all Val-Xaa and Ile-Xaa bonds are cleaved and suggest that there may be a preference for Val over Ile at the P1 position.

To directly address the question of a preference for Val over Ile, the IV37 and IV37/VI41 variants of Arc-*st11* (Milla and Sauer, 1995) were also digested by DegP, and the major cleavage products were purified and sequenced. The cleavage sites in both variants were the same, and identical to those seen in Arc-*st6* VA18. The result with the IV37 variant shows that replacing Ile with Val does not generate a new cleavage site at position 37. Moreover, the IV37/VI41 result shows that replacing Val with Ile does not prevent cleavage at position 41. Taken together, these experiments suggest that there is no strong preference for Val over Ile at the P1 position and that the lack of cleavage at specific Ile-Xaa may be caused by the sequence or structural contexts of these residues.

Are there preferences for specific residues or residue types at positions that flank the scissile peptide bonds in addition to those observed at P1? As shown in Table 1, a variety of chemically dissimilar side chains including alanine, serine, methionine, tyrosine, and arginine are found at the P1′ position of different cleavage sites. Moreover, there are also no clearly conserved positions at the other flanking positions (Table 1). To investigate the potential influence of side chains at the P3 to P3′ positions of a single cleavage site, we chose the Val41/Met42 cleavage site in Arc and analyzed the DegP-mediated cleavage of six variants in which residues 39 to 44 were individually replaced by alanine. As judged by SDS gel electrophoresis, cleavage following Val41 was not significantly altered by the nearby QA39,

RA40, MA42, EA43, or SA44 mutations, suggesting that the wild-type side chains at these positions are not crucial determinants of cleavage-site selection for DegP. As expected, the VA41 mutation did reduce cleavage at this position significantly although a small amount (< 5%) was detected by HPLC studies (data not shown). Although we have not detected significant side chain preferences at positions other than P1, some additional factors must influence cleavage since the Ile85-Tyr86 peptide bond in λ repressor is cleaved by DegP, whereas the chemically identical Ile37-Tyr38 bond in Arc is not cleaved.

Conformational constraints in protein degradation. The three-dimensional structures of both Arc and the N-terminal domain of λ repressor are known (Raumann et al., 1994; Pabo and Lewis, 1982), allowing inspection of the sites of DegP cleavage in the native structures of both proteins. In Arc, the Val22, Val25, and Val41 side chains are almost completely buried in the hydrophobic core of the native protein with fractional side chain accessibilities of 0.00, 0.01, and 0.00, respectively (Fig. 2). In the N-terminal domain dimer of λ repressor, the side chains of Val36, Val71, Ile84 and Val91 are also solvent inaccessible. The latter two residues, however, form part of the subunit interface in the metastable N-domain dimer (K_d=0.4 mM) and thus may be more exposed because the protein is expected to be largely monomeric at the substrate concentrations used for DegP digestion. Nevertheless, a majority of the P1 side chains recognized by DegP are inaccessible in the native, three-dimensional structure of the substrates. As a result, it seems probable that partial or complete unfolding of the substrates is required before cleavage by DegP can occur.

Importance of disulfide bridges in susceptability to DegP cleavage. As mentioned above, a fusion protein consisting of *E. coli* β-lactamase (Bla) and the variable domain of the Bence-Jones protein REI (REI_v) is a candidate to be a substrate for DegP *in vivo* (Kolmar et al., 1995; Kolmar, 1992). Fig. 3 shows that the melting temperature of this protein is approximately 45 °C. Thus, approximately half of the Bla/REI_v molecules are denatured at this temperature, each containing numerous Val-Xaa dipeptides which should be potential cleavage sites. Nevertheless, as shown in Fig. 4, the purified Bla/REI_v fusion protein is not significantly degraded by purified DegP or DegQ at 45 °C *in vitro*. The Bla/REI_v fusion protein contains two disulfide bonds, one in each domain. Reduction and S-carboxymethylation of these disulfide bonds results in a soluble form of the protein which no longer shows cooperative thermal unfolding (Fig. 3) and is presumably denatured. As shown in Fig. 4, this disulfide-free Bla/REI_v protein is readily cleaved both by DegP and DegQ, suggesting that the presence of intramolecular disulfide bonds in the denatured form of the oxidized protein must prevent degradation by these proteases.

The oligomeric state of DegP and DegQ. Although the subunit molecular weights of DegP and DegQ are roughly 50 kDa, both behave as much larger proteins in gel filtration experiments (Swamy et al., 1983; Waller and Sauer, 1996). Goldberg and coworkers reported that DegP (protease Do) could be isolated in different preparations in either of two forms which chromatographed at apparent molecular weights of approximately 300 kDa or 500 kDa in gel filtration experiments (Swamy et al., 1983). We observed similar behavior during the purification of DegP (but not DegQ). Occasionally, a mixture of both species was found during chromatography on

Superose 6, and rechromatography of the 300 kDa fraction produced the same two peaks again. This suggests that DegP can exist in two oligomeric forms which are interconvertible.

To characterize the oligomeric states of DegP and DegQ in greater detail, molecular crosslinking studies were performed with both proteins. Incubation with the bifunctional crosslinker DSP resulted in formation of a mixture of covalently linked oligomers (Fig. 5). The highest crosslinked oligomer observed was a dodecamer. The other main crosslinking products were hexamers and trimers. No crosslinked products were observed when DegP or DegQ were denatured with SDS prior to crosslinking (Fig. 5). Dodecamers of either protein would have molecular weights of approximately 600 kDa, within error of the rough estimates from gel filtration. Hexamers probably correspond to the 300 kDa species observed in gel filtration.

DISCUSSION

We have shown that the DegP and DegQ proteases display indistinguishable substrate specificities towards several model protein substrates *in vitro*. This result, the high level of homology between the two proteases, their common subcellular localization in the periplasm, and the finding that overproduction of DegQ rescues the temperature-sensitive growth defect of a *degP41* strain (Waller and Sauer, 1996) suggest that the two enzymes are essentially interchangable. Nevertheless, it is still the case that a *degP41 degQ*$^+$ strain cannot grow at high temperature. This might be explained if DegP and DegQ have slightly different substrate specificities or

activities *in vivo*. An alternative, and more likely explanation, involves the transcriptional regulation of the two proteins. Overexpression of DegP, but not DegQ, is induced by heat shock, and thus insufficient DegQ may be produced at high temperature to complement the *degP*$^-$ phenotype in an otherwise wild-type cell (Waller and Sauer, 1996).

DegP and DegQ are quite selective in their choice of cleavage sites. At the level of primary sequence, only Val-Xaa and Ile-Xaa peptide bonds appear to be efficiently cleaved, suggesting a preference for a β-branched side-chain at the P1 position. However, several Val-Xaa and Ile-Xaa bonds are not cleaved, indicating cleavage determinants in addition to the presence of a nonpolar β-branched residue. One reason for selective cleavage after specific Ile or Val residues may be the accessibility of these bonds to the protease active site. However, the cleaved and uncleaved bonds are all buried within the hydrophobic cores of the test substrates, Arc and the N-terminal domain of λ repressor. This suggests that local structural or sequence features may affect cleavage of specific peptide bonds.

The ability of DegP and DegQ to cleave bonds that are buried within the hydrophobic core of a substrate, and their preference for Val or Ile at the P1 position, suggest that these proteases recognize the non-native states of proteins. This may explain why DegP and DegQ efficiently degrade β-casein, which is largely unstructured in solution (Waxman and Goldberg, 1986; Ostoa-Saloma et al., 1990; Lipinska et al., 1990, Skorko-Glonek, 1995), but fail to cleave the native forms of proteins such as bovine serum albumin, ovalbumin, and carbonic andydrase.

Why are Arc and the N-terminal domain of λ repressor reasonably good substrates for DegP and DegQ, if their tertiary folds prevent access to the scissile peptide bonds along their polypeptide chains? In each case, the native structures of these small proteins are only modestly more stable than the denatured states and thus a small fraction of molecules are denatured at any given time. Moreover, the interconversion of the native and denatured forms of these proteins occurs very rapidly, with denaturation half-lives on the order of 10 s and millisecond refolding times (Milla and Sauer, 1994; Huang and Oas, 1995). Thus, if the denatured forms of Arc and the N-terminal domain of λ repressor are efficiently recognized and bound by DegP and DegQ, simple mass action could eventually lead to complete degradation. Because both repressor substrates are small (50-100 residues) and lack disulfides, their denatured states may also be relatively unstructured and thus more accessible to the active sites of DegP and DegQ. DegP and DegQ lack the ATP-dependent unfolding activities associated with cytoplasmic proteases like Lon and Clp. For this reason, it will be interesting to see if these periplasmic enzymes have evolved other mechanisms (e.g., tight binding to relatively large regions of extended regions of polypeptide chain) to bind substrates in their denatured forms. Such mechanisms might explain why DegP (and presumably DegQ) seem to be relatively inactive in cleaving small peptides (Lipinska et al., 1990). Indeed a peptide encompassing residues 37-53 of Arc is not cleaved by DegP *in vitro* (Kolmar, unpublished results). This peptide includes the Val-Met bond at position 41-42 which is efficiently cleaved in the context of the intact protein.

The ability of DegP and DegQ to cleave the β-lactamase-REIv protein only after its disulfide bonds are broken is similar to the 20S proteasome of

the archaebacterium *Thermoplasma acidophilum*, which only cleaves somatostatin after reduction of its disulfide bond (Wenzel and Baumeister, 1995). The crystal structure of this proteasome reveals 14 subunits in two seven-subunit rings with the active sites located in an inner channel of the complex (Lowe, 1995). Only unfolded proteins devoid of disulfide bonds appear capable of entering the channel, thereby preventing nonspecific proteolysis of folded proteins. The same structural organization is thought to be responsible for the selectivity of the 14-subunit Clp protease of *E. coli* (Kessel et al., 1995). Since DegP and DegQ are large oligomers as well, geometric restrictions in active site access may play some role in their substrate selectivity. In this regard, it is interesting that a monomeric periplasmic protease of *E. coli*, Tsp, appears to use a tethering site to bind specific proteins that have been targeted for degradation by addition of a C-terminal peptide tag (Keiler and Sauer, 1996). In this case, substrate selectivity is ensured not by the physical properties of the enzyme active site but rather by secondary recognition site. The 26S proteasome complex of eukaryotic cells may employ both strategies for selectivity since it has a 20S protease core similar to that of *T. acidophilum* but also recognizes proteins tagged for degradation by covalent attachment of ubiquitin (Hershko and Ciechanover, 1992; Goldberg, 1992; Wlodawer, 1995).

A preference for denatured substrates is consistent with the probable role of DegP in the degradation of unfolded proteins. Furthermore, the P1 site specificity for Val and Ile, residues present in the hydrophobic cores of most proteins, should allow DegP to cleave most unfolded proteins it would naturally encounter in the periplasm.

ACKNOWLEDGEMENTS. We thank Bronwen Brown, Ken Keiler, and Marcos Milla for proteins, and thank members of the Sauer lab for helpful advice. This work was supported by NIH grant AI-16892. H.K. was funded by a post-doctoral fellowship from the Deutsche Forschungsgemeinschaft.

REFERENCES

Baker T. A. and K. Mizuuchi. 1992. DNA promoted assembly of the active tetramer of the Mu transposase Genes Dev. **6**: 2221-2232.

Brown, B. M., M. E. Milla, T. L. Smith, and R. T. Sauer. 1994. Scanning mutagenesis of the Arc repressor as a functional probe of operator recognition. Nat. Struct. Biol. **1:** 164-168.

Cunningham, B. C., and J. A. Wells. 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine scanning mutagenesis. Science **244:** 1081-1085.

Danese, P.N., W. B. Snyder, C.L. Cosma, L.J.B. Davis, and T. J. Shilhavy. 1995. The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. Genes Dev. **9:** 387-398.

Goldberg, A. L. 1992. The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. Eur. J. Biochem. **203:** 9-23.

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

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

Hirota, Y., H. Suzuki, Y. Nishimura, and S. Yasuda. 1977. On the process of cellular division in *Escherichia coli*: A mutant of *E. coli* lacking a murein lipoprotein. Proc. Natl. Acad. Sci. USA. **74:**1417-1420.

Huang G. S., and T. G. Oas. 1995. Submillisecond folding of monomeric lambda repressor. Proc. Natl. Acad. Sci. USA. **92:**6878-6882.

Hubbard, S. J., J. M. Thornton, and S F. Campbell. 1992. Substrate recognition by proteinases. Farraday Discuss. **5:**13-23.

Keiler, K. C., and R. T. Sauer. 1996. Substrate specificity of the Tsp protease. J. Biol. Chem. (in press).

Kessel, M., M. R. Maurizi, B. Kim, E. Kocsis, B. L. Trus, S. K. Singh, and A. C. Steven. 1995. Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26 S proteasome. J. Mol. Biol. **250:**587-594.

Kolmar, H. 1991. Über die Faltungsstabilität einer variablen Immunglobulindomäne und deren Abhängigkeit von speziellen Mermalen der Aminsäuresequenz. Thesis, Universität Göttingen, F.R.G.

Kolmar, H., C. Frisch, K. Götze, and H.-J. Fritz. 1995. Immunoglobulin mutant library genetically screened for folding stability exploiting bacterial signal transduction. J. Mol. Biol. **251:**471-476.

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

Lipinska, B., S. Sharma, and C. Georgopoulos. 1988. Sequence analysis and regulation of the htrA gene of *Esherichia coli*:a σ_{32}-independent mechanism of heat-inducible transcription. Nucleic Acids Res. **16:**10053-10067.

Lipinska, B., O. Fayet, L. Baird, and C. Georgopoulos. 1989. Identification, characterization, and mapping of the gene, whose product is essential for bacterial growth only at elevated temperatures. J. Bacteriol. **171:**1574-1584.

Lipinska, B., M. Zylicz, and C. Georgopoulos. 1990. The HtrA (DegP) protein, essential for *Escherichia coli* survival at high temperatures. J. Bacteriol. **172:**1791-1797.

Löwe, J., D. Stock, B. Jap, P. Zwickl, W. Baumeister, and R. Huber. 1995. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. Science **268:**533-539.

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

Milla. M. E., B. M. Brown, and R. T. Sauer. 1993. P22 Arc repressor: enhanced expression of unstable mutants by addition of polar C-terminal sequences. Prot. Sci. **2:**2198-2205.

Milla, M., B. M. Brown and R. T. Sauer. 1994. Protein stability effects of a complete set of alanine substitutions in Arc repressor. Nat. Struct. Biol. **1:**518-523.

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

Milla, M., and R. T. Sauer. 1995. Critical side-chain interactions at a subunit interface in the Arc repressor dimer. Biochemistry **34:**3344-3351.

Ostoa-Saloma, P., J. Ramirez, and R. Perez-Montfort. 1990. Causes of the decrease in fluorescence due to proteolysis of α-casein. Biochem.Biophys. Acta. **1041:**146-152.

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

Parsell, D. A., and S. Lindquist. 1993. The function of heat-shock proteins in stress tolerance: Degradation and reactivation of damaged proteins. Annu. Rev. Genet. **27:** 437-496.

Raumann, B.E., M. A. Roult, C. O. Pabo, and R. T. Sauer. 1994. DNA recognition by beta-sheets in the Arc repressor-operator crystal structure. Nature **367:**754-757.

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

Seol, J.H., S. K. Woo, E. M. Jung, S. J. Yoo, C. S. Lee, K. Kim, K. Tanaka, A. Ichihara, D. B. Ha, and C. H. Chung. 1991. Protease Do is essential for survival of *Escherichia coli* at high temperatures: its identity with the *htrA* gene product. Biochem. Biophys. Res. Comm. **176:**730-736.

Skorko-Glonek, J., K. Krzewski, B. Lipinska, E. Bertoli, and F. Tanfani. 1995. Comparison of the structure of wild-type HtrA heat shock protease and mutant HtrA proteins. J. Biol. Chem. **270:**11140-11146.

Strauch, K. L., and J. Beckwith. 1988. An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins. Proc. Natl. Acad. Sci. USA **85:**1576-1580.

Strauch, K. L., K. Johnson, and J. Beckwith. 1989. Characterization of *degP*, a gene required for proteolysis in the cell envelope and essential for growth of *Escherichia coli* at high temperatures. J. Bacteriol. **171:**2689-2686.

Swamy, K. H. S., C. H. Chung, and A. L. Goldberg. 1983. Isolation and Characterization of protease Do from *Escherichia coli*, a large serine protease containing multiple subunits. Arch. Biochem. Biophys. **224:**543-554.

Waller, P. R. H., and R. T. Sauer. 1996. Characterization of *degQ* and *degS*, *Escherichia coli* genes encoding homologs of the DegP protease. J. Bacteriol. **178:**1146-1153.

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

Wenzel, T., and W. Baumeister. 1995. Conformational constraints in protein degradation by the 20S proteasome. Struct. Biol. **2:**199-204.

A. Wlodawer. 1995. Proteasome: a complex protease with a new fold and a distinct mechanism. Structure **3:** 417-420.

Table 1. DegP/DegQ cleavage sites in λ repressor and Arc. The numbers indicate the positions of the P1 residues in the two substrates. The site of the scissile bond is indicated by an arrow.

	P1 ↓ P1'	
λ #36	QESV	ADKM
λ #71	ILKV	SVEE
λ #84	AREI	YEMY
λ #91	YEAV	SMQP
Arc #22	LDLV	RKVA
Arc #25	VRKV	AEEN
Arc #41	YQRV	MESF

Figure 1. HPLC traces of the digestion products obtained from incubation of Arc-*st6* with DegP (Arc/DegP) or DegQ (Arc/DegQ), respectively. Uncleaved Arc is shown as a control. The sequence of the four N-terminal residues of each of the major Arc/DegP cleavage products was determined by Edman degradation and is shown above each of the corresponding peaks. The numbering denotes the position of the N-terminal residue of each fragment within the Arc protein sequence.

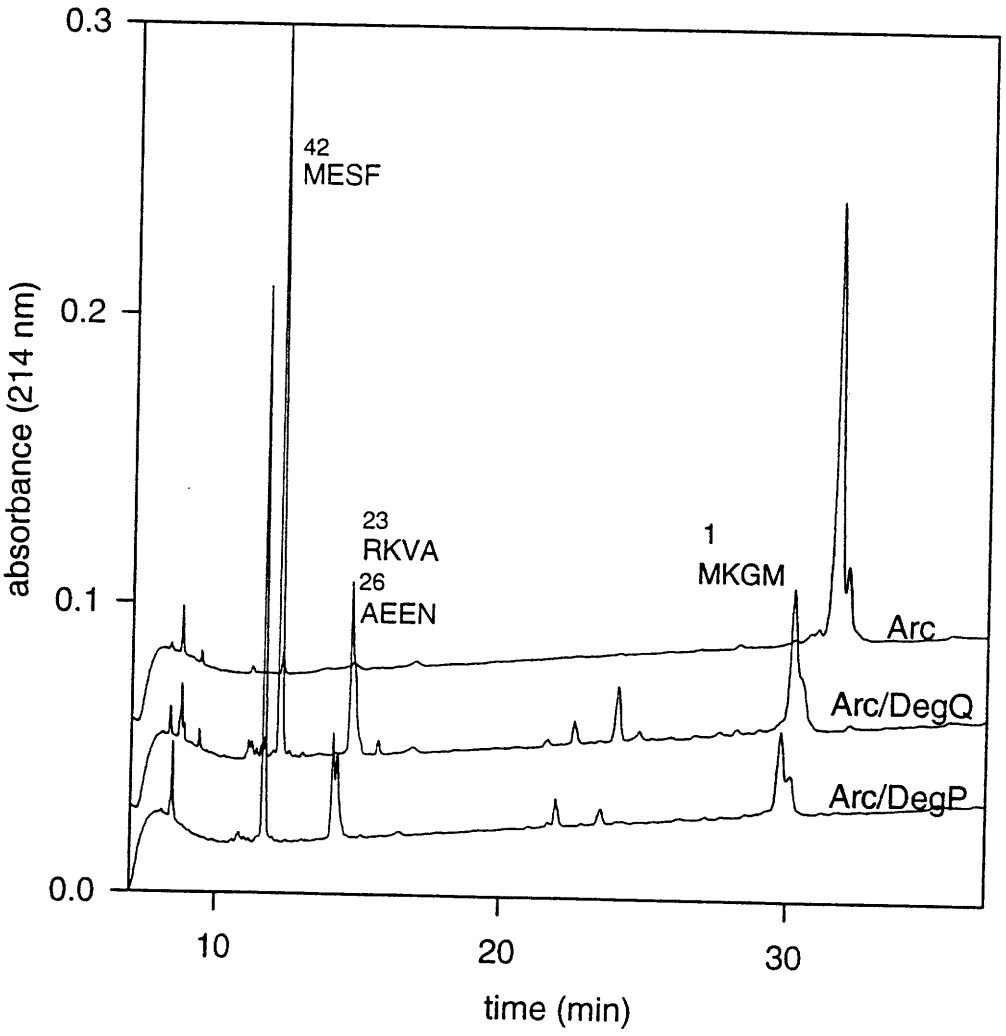

0.3
0.2
0.1
0.0
absorbance (214 nm)
42
MESF
23
RKVA
26
AEEN
1
MKGM
Arc
Arc/DegQ
Arc/DegP
10
20
30
time (min)

Figure 2. Structure of the Arc homodimer (Raumann, 1994), showing the positions of the side-chains of Val22, Val25, and Val41 on one of the two monomers.

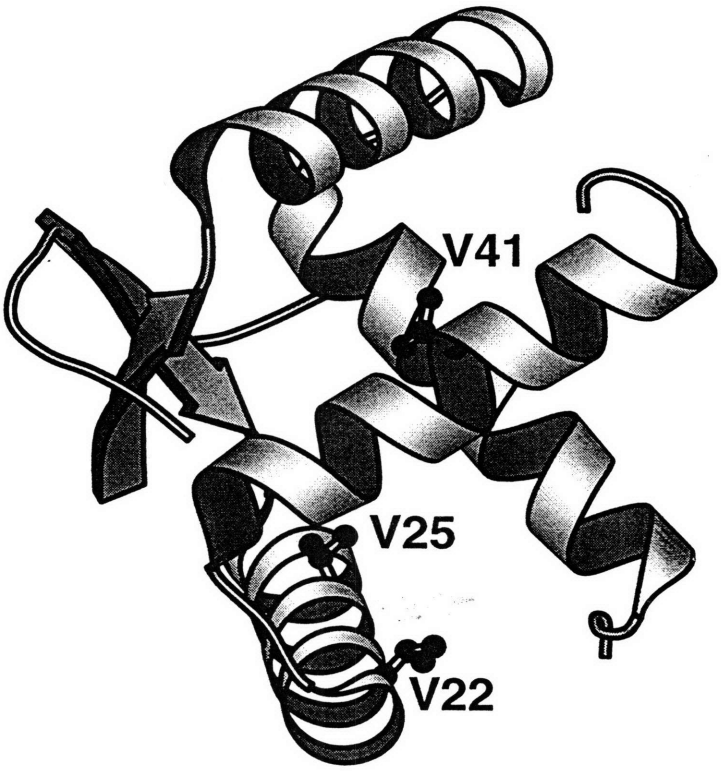

V41
V25
V22

Figure 3. Thermal denaturation of unmodified β-lactamase/REIv fusion protein (open circles) and reduced and carboxymethylated β-lactamase/REIv (open squares). Thermal unfolding was assayed by monitoring circular dichroism at 222 nm at a protein concentration of 1 mg/ml in 50 mM potassium phosphate, pH 7.4.

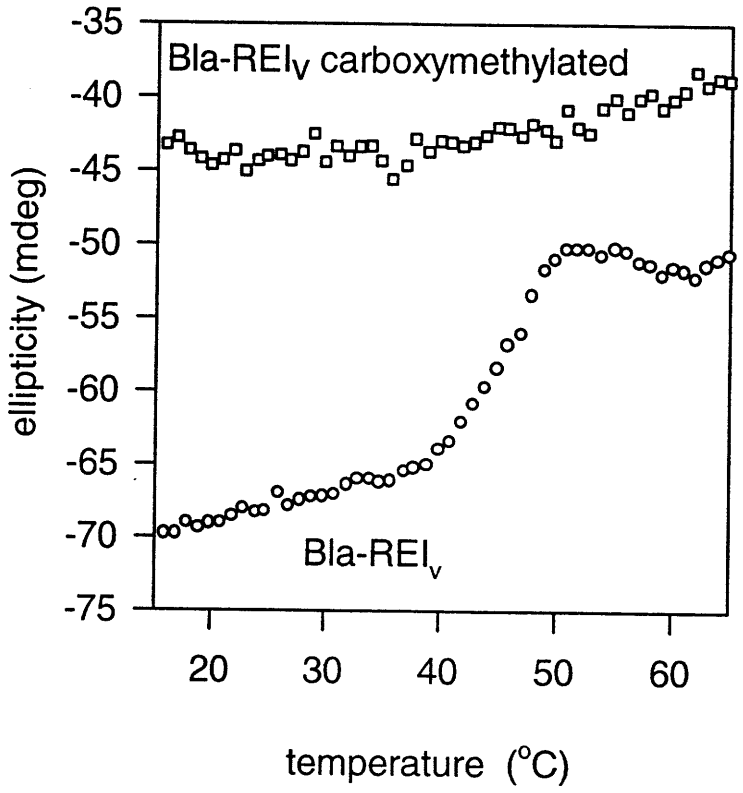

-35
-40
-45
-50
-55
-60
-65
-70
-75
ellipticity (mdeg)
Bla-REI$_V$ carboxymethylated
Bla-REI$_V$
20
30
40
50
60
temperature (°C)

Figure 4. HPLC analysis of β-lactamase/REIv following overnight incubation at 45 °C with DegP (**A**) or DegQ (**B**) using enzyme:substrate ratios of 1:20. In each panel, trace 3 is the Bla/REIv fusion protein alone; trace 2: is the Bla/REIv fusion protein following incubation with DegP or DegQ; and trace 1 is the reduced and carboxymethylated Bla/REIv fusion protein following incubation with DegP or DegQ.

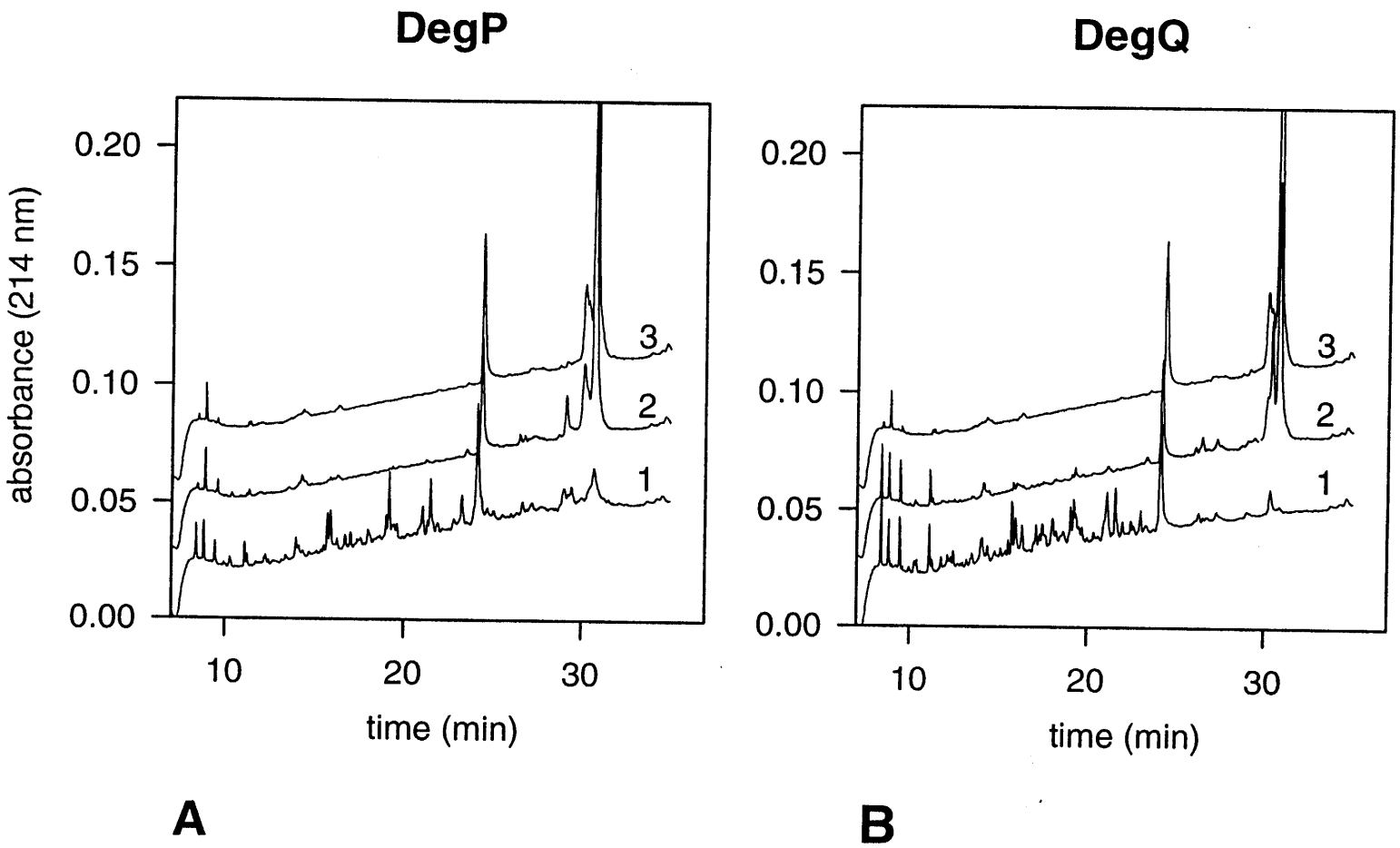
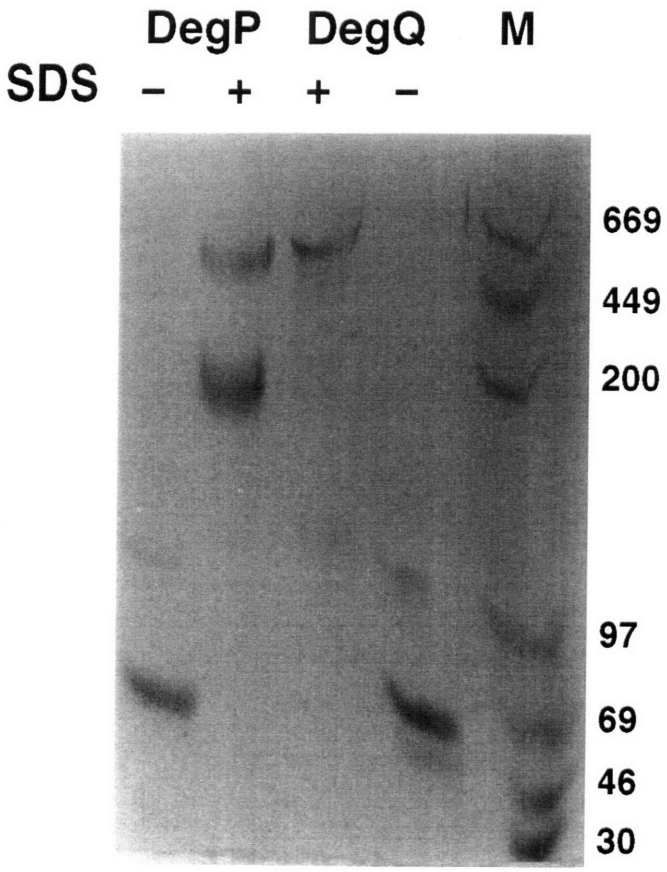

DegP
DegQ
absorbance (214 nm)
0.20
0.15
0.10
0.05
0.00
10
20
30
time (min)
1
2
3
A
B

Figure 5. Molecular crosslinking of DegP and DegQ with DSP. As a control, each protease was denatured by addition of SDS prior to incubation with DSP. Proteins were resolved on a 2.2% acrylamide-agarose gel (see Materials and Methods). M: Molecular weight marker with sizes indicated in kDa.


DegP
DegQ
M
SDS
– + + –
669
449
200
97
69
46
30

CHAPTER 4

Role of a Peptide Tagging System in the Degradation of Proteins Synthesized from Damaged mRNA

Keiler, K. C., Waller, P. R. H., and Sauer, R. T. 1996. Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA. Science **271:** 990-993.

Using variants of λ repressor and cytochrome b_{562}, we found that proteins translated from mRNAs without stop codons were modified by COOH-terminal addition of an *ssrA*-encoded peptide tag and subsequently degraded by COOH-terminal specific proteases in both the cytoplasm and periplasm of *E. coli*. The tag appears to be added to the COOH-terminus of the nascent polypeptide chain by cotranslational switching of the ribosome from the damaged mRNA to *ssrA* RNA.

Although many intracellular proteases have been identified (1), it is generally not known what makes certain proteins the targets of particular proteases. Studies of recombinant interleukin-6 (IL-6), purified from inclusion bodies in *Escherichia coli*, revealed that some IL-6 molecules are modified by COOH-terminal truncation at different positions in the sequence and addition of an 11-residue peptide tag at the truncation point (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 *ssrA* gene product is a stable 362-nucleotide RNA molecule, which has some tRNA-like properties and can be charged with alanine (3). The COOH-terminal residues of the *ssrA* peptide tag (YALAA) are similar to a COOH-terminal tail sequence (WVAAA) recognized by Tsp (4), a periplasmic endoprotease, and by a cytoplasmic protease which also degrades proteins in a tail-dependent fashion (5). Thus, we reasoned that tagging with the *ssrA* peptide might mark proteins for degradation by COOH-terminal specific proteases.

To test this hypothesis, we constructed genes encoding variants of the NH_2-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 COOH-terminal tail (6). Variants of each protein were also constructed with a control tag (AANDENYALDD), because both Tsp and its cytoplasmic counterpart fail to cleave proteins with charged residues at the COOH-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 strain 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 peptide-tagged λ-repressor protein in the cytoplasm was not affected in the *tsp*$^-$ strain (Table 1). Thus, the presence of the peptide tag at the COOH-termini of these proteins resulted in rapid intracellular degradation in both the cytoplasmic and periplasmic compartments of the cell. In the periplasm, this rapid degradation required the presence of the Tsp protease. When the variant cytochrome proteins were purified and incubated with Tsp in vitro (7), cleavage of the *ssrA* peptide-tagged variant but not the control-tagged variant, was observed (Fig. 1D). This supports a direct role for Tsp in the periplasmic degradation of peptide-tagged substrates in vivo (8).

By what mechanism could the *ssrA*-encoded tag sequence be added to the COOH-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. Certain mRNAs in the cell may lack stop codons because of premature termination or nuclease cleavage. Completing translation of such mRNAs is potentially

problematic, because the ribosomal factors that normally release the nascent polypeptide require the presence of termination codons (9). Thus, the ribosome might stall or idle upon reaching the 3′ end of the mRNA. We propose: (*a*) alanine-charged *ssrA* RNA recognizes a ribosome stalled at the end of an mRNA without a stop codon; (*b*) the alanine from the *ssrA* RNA is added to the COOH-terminus of the nascent chain, creating a peptidyl-*ssrA* molecule; (*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 (10); and (*d*) normal termination and release occur at the ochre termination codon following the peptide-tag region of *ssrA* RNA (Fig. 2). By this model, any protein translated from an mRNA lacking a termination codon will be modified by peptide tagging. As a result, aberrant polypeptides which might be deleterious to the cell can be recognized and degraded by specialized intracellular proteases.

To test this model, we cloned the *trpA* transcriptional terminator (*trpAt*) upstream of the translational-termination codons in the 3′ regions of genes encoding the NH_2-terminal domain of λ repressor and cytochrome b_{562} (11, 12). By our model, the transcripts (λ-repressor-1-93-M2-H_6-*trpAt* and cyt-b_{562}-*trpAt*, respectively) should terminate at *trpAt* and the resulting protein should be modified by peptide tagging and then degraded. In cells containing *ssrA* RNA, the λ-repressor variant was degraded with a half-life of a few minutes (Fig. 3, Table 1). In isogenic cells lacking *ssrA*, the protein was both smaller (as expected if peptide tagging did not occur) and degraded more slowly (Fig. 3, ref. 13). The cyt-b_{562}-*trpAt* protein also appeared to be rapidly degraded in *ssrA*$^+$ cells; in fact no protein was observed following a 30 s pulse (Fig. 4). The same protein was visibly expressed and longer lived (14) in an

otherwise isogenic *tsp*$^-$ strain (Fig. 4). Thus, as predicted by the model, proteins translated from mRNAs lacking termination codons appeared to be modified by *ssrA*-dependent peptide tagging. The half lives for these *ssrA* peptide-tagged proteins were very short (<2 min) in wild-type strains but were increased significantly (30-60 min) for peptide-tagged periplasmic variants in strains lacking the COOH-terminal specific protease Tsp. This again provides strong evidence that *ssrA*-mediated peptide tagging marks proteins for degradation by specialized proteases.

A second cytochrome-b_{562} protein (cyt-b_{562}-M2-H_6-*trpAt*) in which *trpAt* was separated from the body of the protein by an M2 epitope and His_6 sequence was also rapidly cleaved (half-life < 30 s) in cells containing Tsp, but in this case cleavage resulted in a metastable intermediate (Fig. 4). This proteolytic intermediate was then degraded with a half-life of approximately 15 min (15). In cells lacking the Tsp protease, a larger, uncleaved form of the protein was produced (Fig. 4). The intermediate and uncleaved forms of the cyt-b_{562}-M2-H_6-*trpAt* proteins were purified from *tsp*$^+$ and *tsp*$^-$ strains, respectively, and characterized by NH_2-terminal sequencing and ion-spray mass spectrometry (16). The protein purified from the *tsp*$^-$ strain had the NH_2-terminal sequence (ADLED) of cytochrome b_{562} (17) but was heterogeneous by mass spectrometry. The largest and most abundant form had a mass of 15,637 Da, within 0.06% of that expected for a protein containing residues 1-129 of 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 same NH_2-terminal sequence and a mass of 13,919 Da, within 0.05% of that expected for a protein consisting of residues 1-123 (Fig. 5B). This

suggests that Tsp initially cleaved the mature protein at an A↓A sequence, thereby removing approximately 15-20 residues.

To confirm the peptide-tagged structure proposed above, we digested the cyt-b_{562}-M2-H_6-*trpAt* proteins purified from *tsp*$^+$ and *tsp*$^-$ strains with trypsin and chromatographed the digests on a C18 column. Two difference peptides were isolated from the *tsp*$^-$ protein (labeled 1 and 2 in Fig. 5A) and were sequenced by Edman degradation. Peptide 1 contained four residues encoded by the *trpAt* portion of the cyt-b_{562}-M2-H_6-*trpAt* gene followed by the AANDENYALAA peptide tag (Fig. 5B). Peptide 2 was a mixture of sequences containing either two or three *trpAt*-encoded residues followed by the 11-residue tag (Fig. 5B). Thus, in both cases, the sequences of the difference peptides corresponded to those expected if translation of the cytochrome b_{562} variant ended within the *trpA* terminator and the nascent polypeptides were then modified by addition of alanine and the *ssrA*-encoded peptide-tag sequence (18).

Our results show that proteins synthesized from mRNAs lacking a translational termination codon are modified by addition of a AANDENYALAA peptide tag in a *ssrA*-dependent fashion (19). Tagged proteins are then degraded by specialized proteases in both the periplasm and cytoplasm. The connection between tagging and the absence of termination codons is readily explained by the cotranslational model shown in Fig. 2. The observations that*ssrA* RNA is associated with ribosomes, has tRNA-like properties, can be charged with alanine, and encodes all but the first alanine of the tag peptide (2, 3) also strongly support a cotranslational model. In addition, several observations make pre-translational RNA-splicing or post-

translational peptide-ligation mechanisms unlikely. In studies of IL-6 peptide tagging, mRNAs containing both IL-6 and *ssrA* peptide-tag sequences were not detected (2). Furthermore, 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- or post-translational mechanisms.

The *ssrA* peptide-tagging system and associated COOH-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, 20). Systems involving peptide tagging and degradation by COOH-terminal specific proteases are likely to be widespread in the bacterial world, because homologs of *ssrA* and Tsp have been found in widely varying gram-negative and gram-positive species (3, 21).

There are obvious parallels between the peptide-tagging system described here and the ubiquitin system of eukaryotes which also tags substrates and marks them for degradation (22). In addition, some substrates of the N-end rule pathway are modified by addition of specific amino-acids in a tRNA-dependent manner (23). However, both of these modifications are post-translational. It will be interesting to see whether eukaryotic cells also utilize independent cotranslational 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, 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, H. Inokuchi, *Proc. Natl. Acad. Sci. USA* **91**, 9223 (1994); C. Ushida, H. Himeno, T. Watanabe, A. Muto, *Nucleic Acids Res.* **22**, 3392 (1994).

4. K.R. Silber, K.C. Keiler, R.T. Sauer, *Proc. Natl. Acad. Sci. USA* **89**, 295 (1992); K.C. Keiler, et al., *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 COOH-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} (17).

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. Appropriate COOH-terminal peptide tails appear to be recognized and bound to a tethering site on Tsp or its cytoplasmic counterpart, with subsequent cleavage of the tethered substrate mediated by a separate protease active site (4, 21). 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. The tails themselves are unstructured and do not result in unfolding, decreased thermodynamic stability, or conformational changes in the attached protein (4, 5).

9. J.W.B. Hershey in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, F. C. Neidhardt, et al., Eds., (American Soc. Microbiology, Washington, DC, 1987) pp. 613-647.

10. Precedents exist for the skipping of as many as 50 nucleotides during translation of the T4 gene 60 mRNA. (W.M. Huang, et al., *Science* **239**, 1005 (1988)). This type of skipping 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.

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

12. 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.

13. 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}-M2-H_6-*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. The residual slow degradation of these λ and cytochrome variants produced in *ssrA*$^-$ strains could be due to any of a number of intracellular proteases.

14. The *ssrA* peptide-tagged forms of cyt-b_{562}-*trpAt* and cyt-b_{562}-M2-H_6-*trpAt* have half-lives of 30-40 min in *tsp*$^-$ strains and the cyt-b_{562}-AANDENYALAA protein has a half-life of slightly more than 1 h. Periplasmic proteases homologous to DegP can substitute for Tsp under some conditions (S. H. Bass, Q. Gu, A. Christen, *J. Bacteriol*, **178**, 1154), and may be responsible for the relatively slow, Tsp-independent degradation observed for tagged variants in the periplasm. The small differences in the half-lives of these variants in *tsp*$^-$ strains suggests that the position of the peptide tag or the presence of *trpAt* sequences may affect susceptibility to such proteases.

15. The processed form of the cyt-b_{562}-M2-H_6-*trpAt* protein may also be degraded by Tsp as its C-terminal alanine is the residue most preferred by Tsp (K. C. Keiler and R. T. Sauer, *J. Biol. Chem.* in press)

16. 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. NH_2-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.

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

18. 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 five, six, or seven *trpA* terminator-encoded residues. Likewise, numerous different peptide-tagged forms of IL-6 were identified by Tu et al (2). We assume that these tagged IL-6 proteins were not degraded because they form inclusion bodies. 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, M. J. Chamberlin, *J. Mol. Biol.* **224**, 31 (1992); E. Hajnsdorf, O. Steier, L. Coscoy, L. Teysset, P. Regnier, *EMBO J.* **13**, 3368 (1994); H. Causton, B. Py, R. S. McLaren, C. F. Higgins, *Mol. Microbiology* **14**, 731 (1994)).

19. The simplest model is that tagging occurs when an mRNA lacks a translational termination codon. However, the results do not exclude a possible role for the *trpAt* sequence.

20. In addition to slow growth, *ssrA*$^-$ mutants have phenotypes that involve changes in intracellular proteolysis (26) and alterations in the lysis-lysogeny decisions of temperate phages (D. M. Retallack, L. J. Johnson, 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 (see, for example, D.M. Retallack and D.I. Friedman, *Cell* **83**, 227 (1995)).

21. K.C. Keiler and R.T. Sauer, *J. Biol. Chem.* **270**, 28864 (1995).

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

23. D. K. Gonda, et al., *J. Biol. Chem.* **264**, 16700 (1989); E. Balzi, et al., *J. Biol. Chem.* **265**, 7464 (1990); T. E. Shrader, J. W. Tobias, A. Varshavsky, *J. Bacteriol.* **175**, 4364 (1993).

24. 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 labeling 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. All half-lives were calculated, using data from a minimum of 6 time points, 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.

25. Strain X90 *ssrA1*::cat was constructed by P1 transduction of the *ssrA1*::cat allele from *E. coli* strain JK6257 (26) into *E. coli* strain X90.

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

27. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

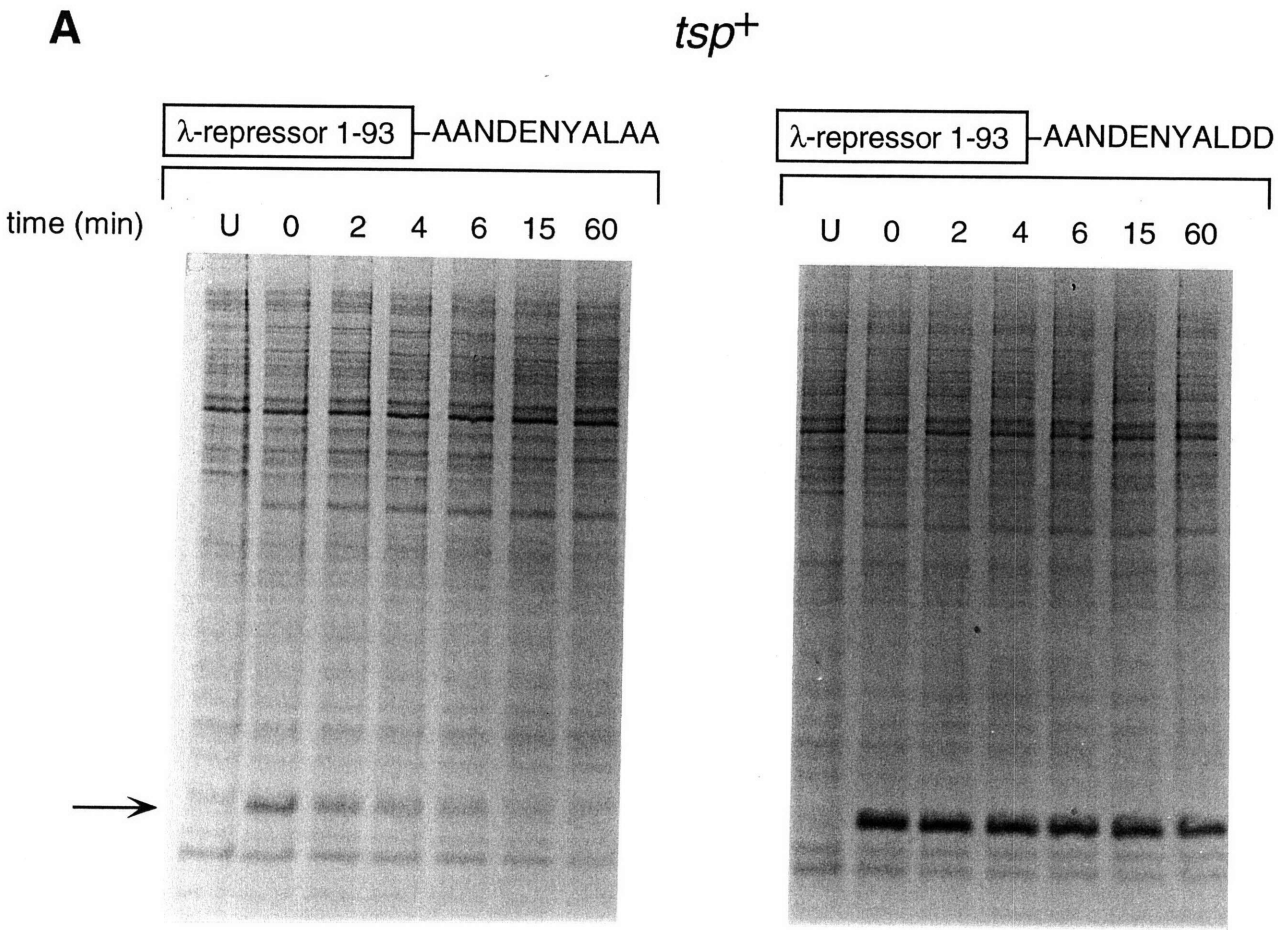
28. We thank T. Baker, D. Bartel, A. Grossman, C. Kaiser, and U. RajBhandary for advice and comments; S. Gottesman, M. Hecht, and S. Sligar for strains; S. Ades, B. Shulman, the MIT Biopolymers Lab, and the Harvard Microchemistry Facility for help. Supported by NIH grants AI-15706 and AI-16892.

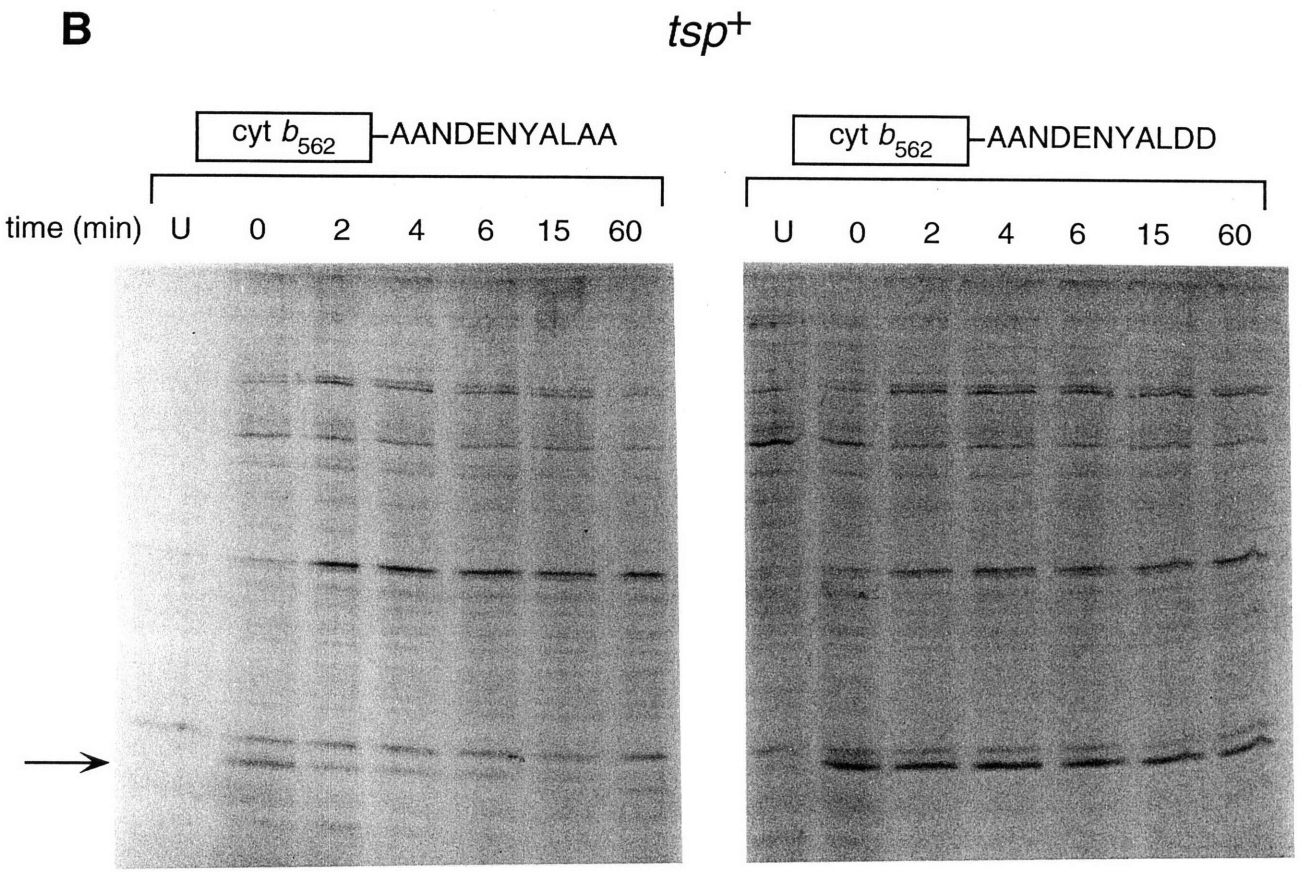
Table 1. Half-lives of the λ-repressor and cytochrome b_{562} constructs in *tsp*$^+$ *ssrA*$^+$ (X90), *tsp*$^-$ *ssrA*$^+$ (KS1000), and *tsp*$^+$ *ssrA*$^-$ (X90 *ssrA1*::*cat*) strains.

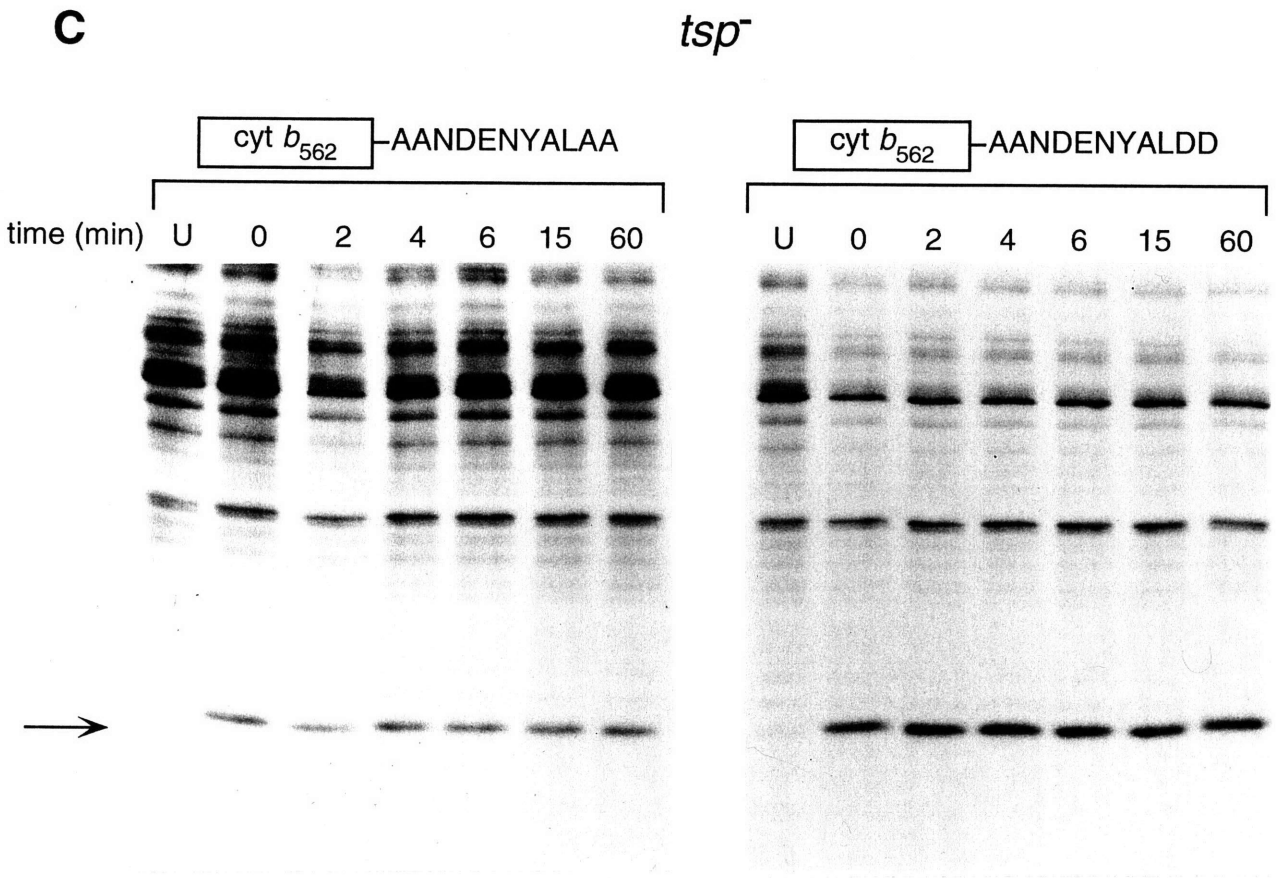
Protein constructs	Strain background		
	tsp $^+$ *ssrA* $^+$	*tsp* $^-$ *ssrA* $^+$	*tsp* $^+$ *ssrA* $^-$
λ-repressor 1-93 – AANDENYALAA	4 min	4 min	ND
λ-repressor 1-93 – AANDENYALDD	> 60 min	> 60 min	ND
λ-repressor 1-93 – M2 – H_6 – *trpAt*	2 min	ND	> 60 min
cyt b_{562} – AANDENYALAA	4 min	> 60 min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

* The half-life is that of the full length protein, a metastable proteolytic intermediate is seen and is degraded with a half-life of ~ 15 min. ND: not determined.

Figure 1. Degradation of variants of the N-terminal domain of λ-repressor or cytochrome b_{562} fused to the *ssrA* peptide tag (AANDENYALAA) or a control tag (AANDENYALDD). (**A**) Pulse-chase assays (24) 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 (24) of cytochrome-b_{562} variants in X90. (**C**) Pulse-chase assays (24) 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 variant by Tsp *in vitro* (7).


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

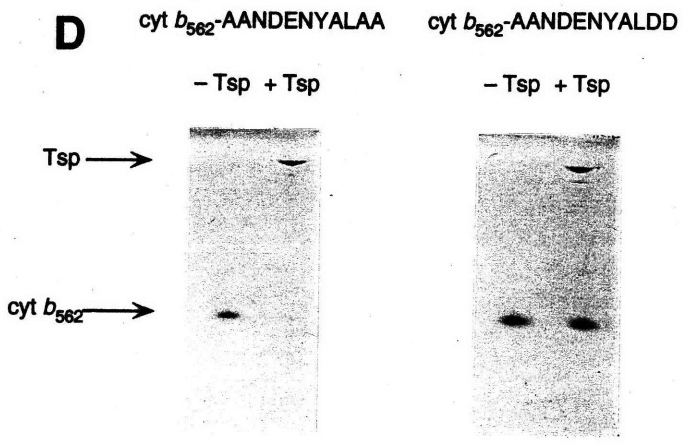

D
cyt b_{562}-AANDENYALAA
cyt b_{562}-AANDENYALDD
– Tsp + Tsp
– Tsp + Tsp
Tsp
cyt b_{562}

Figure 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.

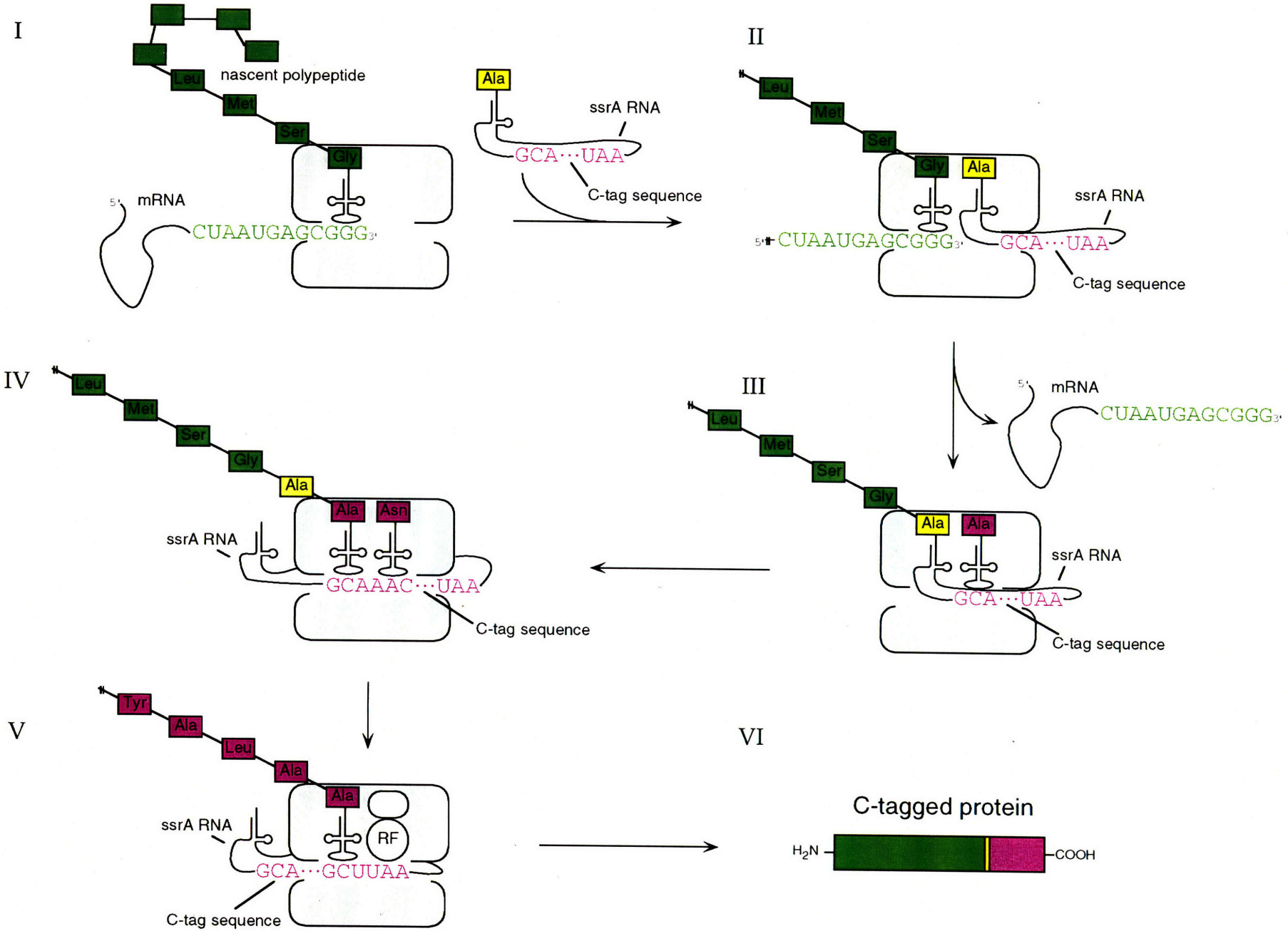

I
nascent polypeptide
Leu
Met
Ser
Gly
5'
mRNA
CUAAUGAGCGGG3'
Ala
ssrA RNA
GCA···UAA
C-tag sequence
II
Leu
Met
Ser
Gly
Ala
ssrA RNA
5'
CUAAUGAGCGGG3'
GCA···UAA
C-tag sequence
III
5'
mRNA
CUAAUGAGCGGG3'
Leu
Met
Ser
Gly
Ala
Ala
ssrA RNA
GCA···UAA
C-tag sequence
IV
Leu
Met
Ser
Gly
Ala
Ala
Asn
ssrA RNA
GCAAAC···UAA
C-tag sequence
V
Tyr
Ala
Leu
Ala
Ala
RF
ssrA RNA
GCA···GCUUAA
C-tag sequence
VI
C-tagged protein
H2N
COOH

Figure 3. Degradation of λ-repressor-1-93-M2-H_6-*trpAt* protein in *ssrA*$^+$ (X90) and *ssrA*$^-$ (X90 *ssrA1*::cat; ref. 25) strains assayed by pulse-chase experiments (24). 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.

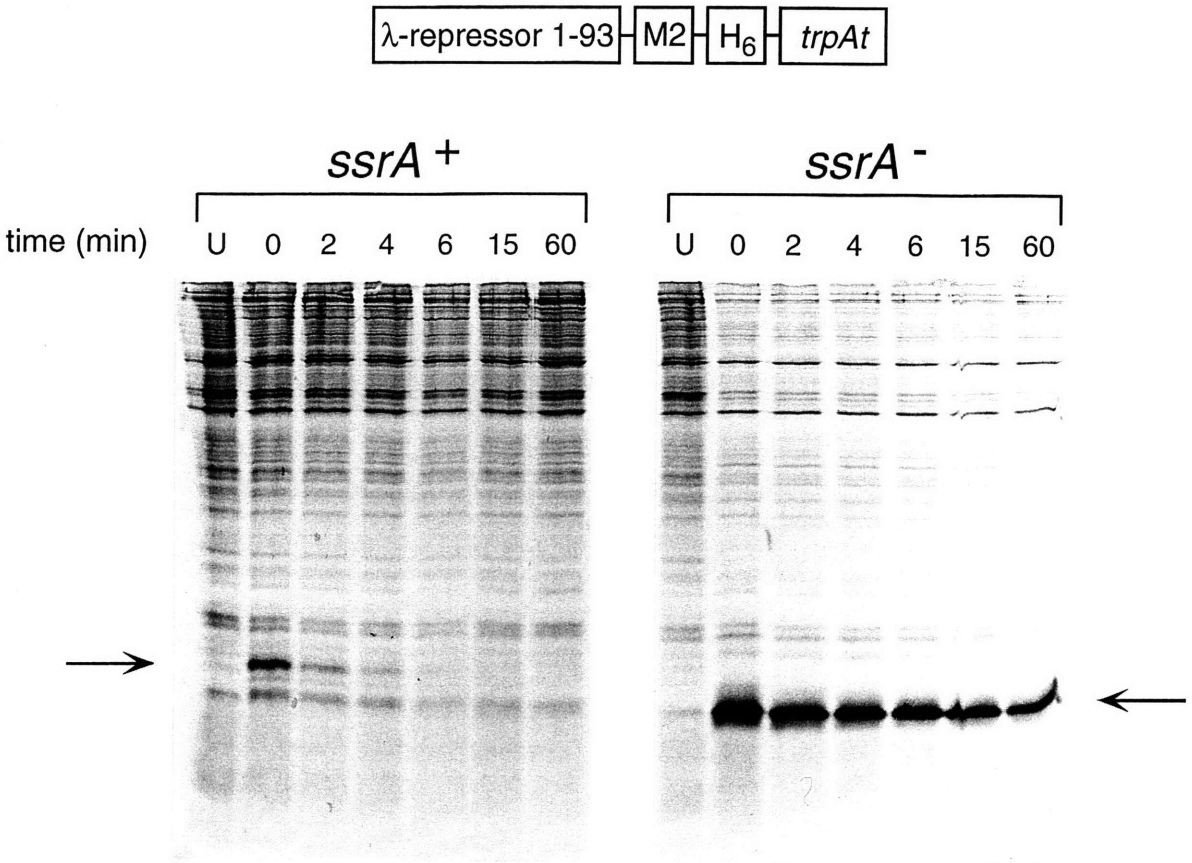

λ-repressor 1-93
M2
H6
trpAt
ssrA +
ssrA -
time (min)
U 0 2 4 6 15 60
U 0 2 4 6 15 60

Figure 4. Pulse-chase experiments (24) for cyt-b_{562}-*trpAt* and cyt-b_{562}-M2-H_6-*trpAt* 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.

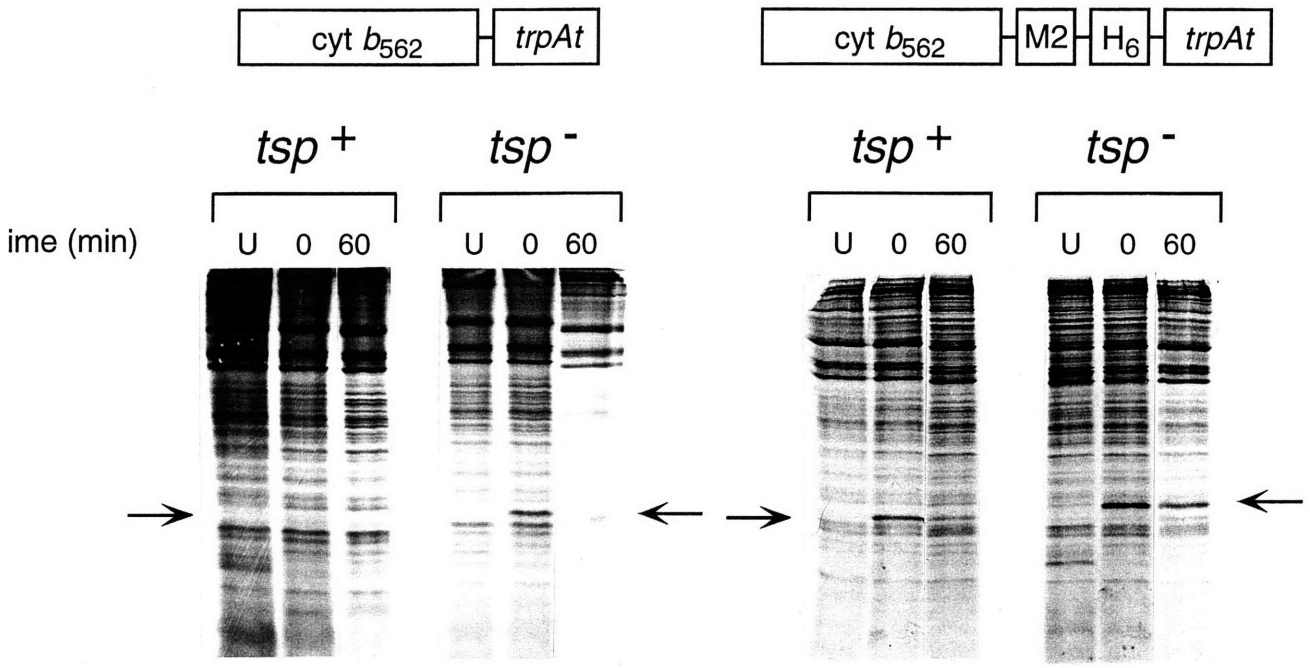
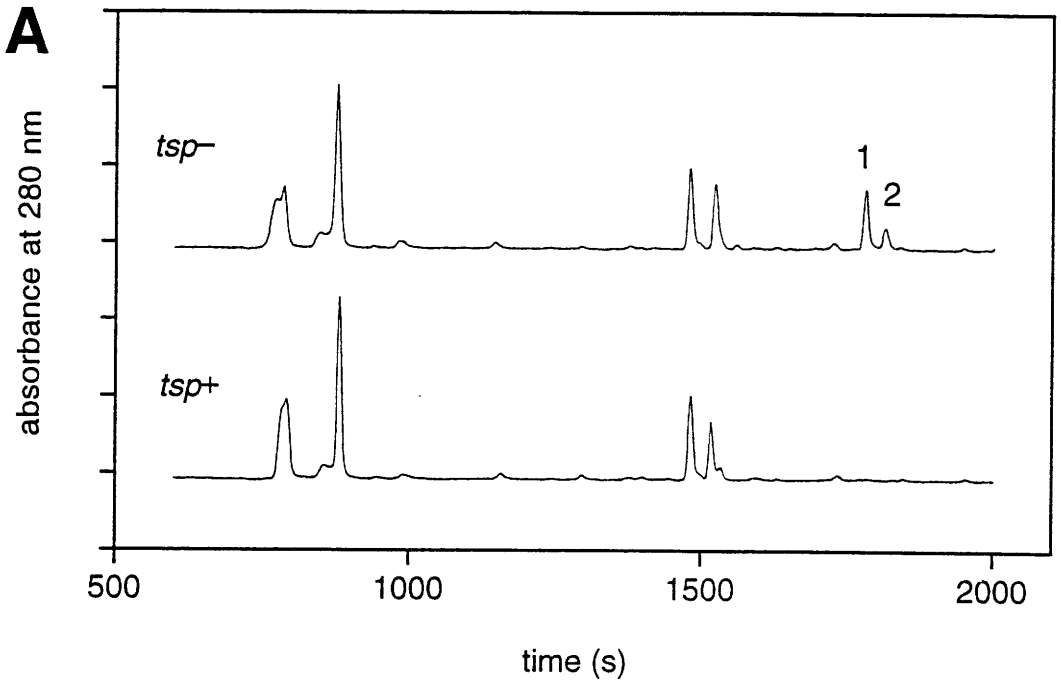
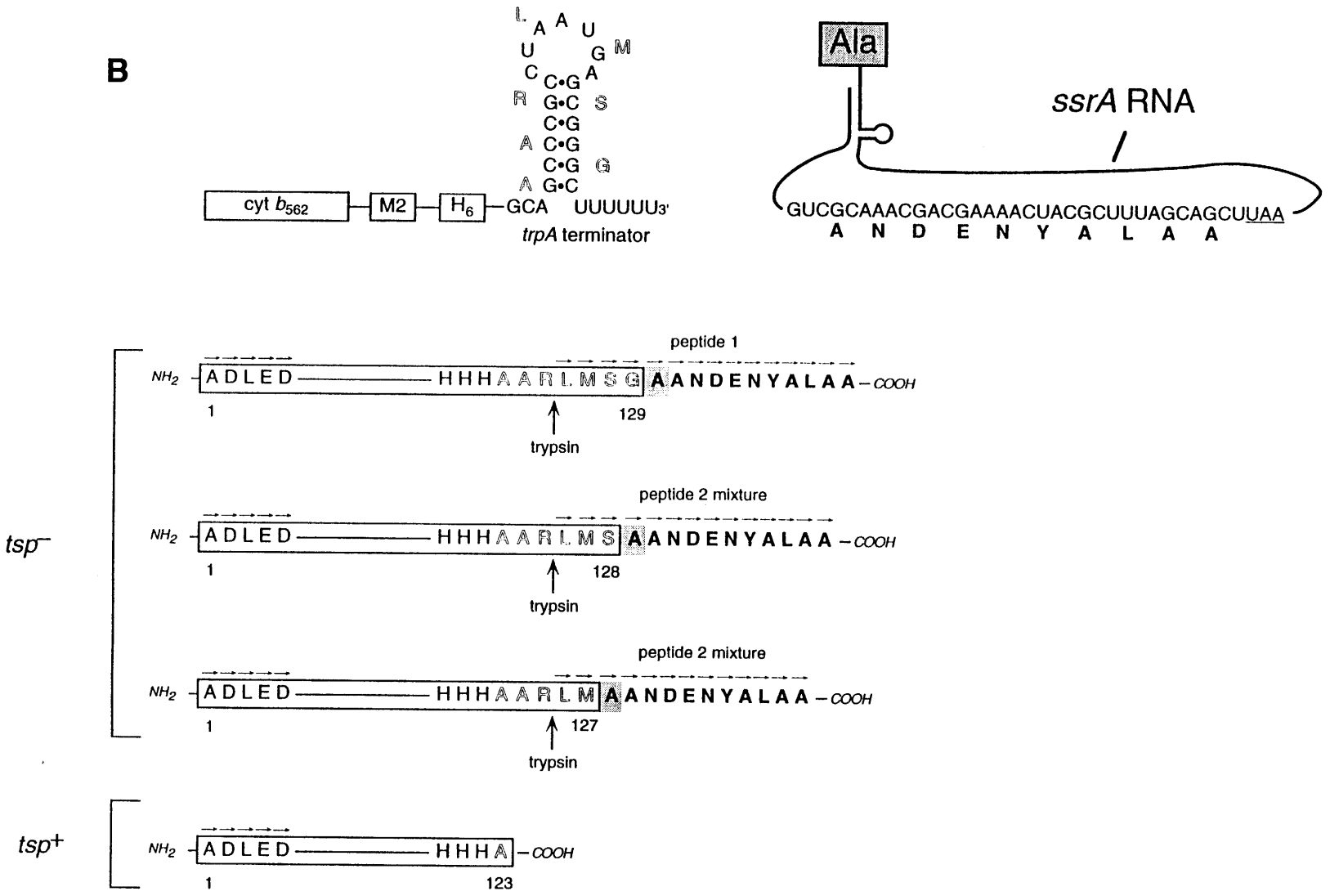

cyt b_{562}
trpAt
cyt b_{562}
M2
H_6
trpAt
tsp +
tsp -
tsp +
tsp -
ime (min)
U 0 60
U 0 60
U 0 60
U 0 60

Figure 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* Diagrams of the mRNA expected if transcription of the cyt-b_{562}-M2-H_6-*trpAt* gene terminates at the *trpA* terminator (the amino acids encoded by the terminator are shown in outline) and 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 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.


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


B
cyt b_{562}
M2
H_6
GCA UUUUUU$_3$'
trpA terminator
Ala
ssrA RNA
GUCGCAAACGACGAAAACUACGCUUUAGCAGCUUAA
A N D E N Y A L A A
peptide 1
ADLED
HHHAARLMSGAANDENYALAA
1
129
trypsin
tsp⁻
peptide 2 mixture
ADLED
HHHAARLMSAANDENYALAA
1
128
trypsin
peptide 2 mixture
ADLED
HHHAARLMAANDENYALAA
1
127
trypsin
tsp+
ADLED
HHHA
1
123