Specificity and structural characterization of the PDZ domain from DegS, an extracytoplasmic *E. coli* protease

by

Nathan P. Walsh

Sc.B. Biochemistry
Brown University, 1995

Submitted to the Department of Biology
on January 28, 2002 in Partial Fulfillment of the
Requirements for the Degree of

Ph.D. in Biology

at the

Massachusetts Institute of Technology
[February 2002]

Signature of Author: Signature redacted
Department of Biology
January 28, 2002

Certifed by: Signature redacted
Robert T. Sauer
Department of Biology
Thesis Supervisor

Accepted by: Signature redacted
Department of Biology
Chairman, Committee for Graduate Students

This page was intentionally left blank.

Specificity and structural characterization of the PDZ domain from DegS, an extracytoplasmic *E. coli* protease

by

Nathan P. Walsh

Submitted to the Department of Biology
on January 28, 2002 in Partial Fulfillment of the
Requirements for the Degree of Ph.D. in Biology

Abstract

DegS is a membrane-bound bacterial protease that is involved in the extracytoplasmic-stress response. The C-terminal domain has limited homology to PDZ domains and was thought to be involved in regulation or substrate recognition. A model of this PDZ domain was generated from NMR solution studies and homology modeling. Peptide selection studies identified the sequence Tyr-Tyr-Phe (YYF) as a C-terminal motif that binds to the PDZ domain. Possible targets were identified including many of the outer-membrane proteins (OMPs), which contain both a conserved terminal YxF and internal YYF sequences. The binding of the DegS PDZ domain to a YYF peptide and OMP derivatives were confirmed using microcalorimetry. Because stress signaling can be triggered by over-expression of some of the outer-membrane proteins, I propose that DegS may receive a signal from unassembled OMPs and transmit it to the σ^E transcription factor by increasing proteolysis of RseA.

Thesis Supervisor: Robert T. Sauer
Title: Salvador E. Luria Professor of Biology, Head of Department

Biographical Note

Education

1995-present	Ph.D. in Biology, Massachusetts Institute of Technology Advisor: Robert T. Sauer "Specificity and structural characterization of the PDZ domain from DegS, an extracytoplasmic *E. coli* protease".
1991-1995	Sc.B. in Biochemistry (with Honors), Brown University Advisor: Edward Hawrot "Mutagenesis of α-Bungarotoxin".

Research experience

1995-present	Graduate Thesis work at MIT -Structure and Function of PDZ domains/Folding and Stability of Arc
1993-1995	Undergraduate Thesis work at Brown University -α-Bungarotoxin Specificity for Neuronal Receptors
Summer 1993	Summer Researcher at Worcester Foundation for Experimental Biology -Proteases as molecular switches
Summer 1990	Summer Researcher at Tufts University -Performed ovariectomies and brain histologies on hamsters

Honors and awards

1999-present	Lester Wolfe Fellow for Spectroscopy
1995-1999	NIH Institutional Training Grant
1993-1995	Howard Hughes Undergraduate Fellowship Researcher
1991	Digital Equipment Corporation Scholarship
1991	ThermoElectron, Inc. Scholarship (Fred N. Huffman Award)

Teaching experience

Fall 1998	Teaching Assistant for "Biochemistry for Graduate Students" at MIT TA for course required of all first year Biology graduate students.
Fall 1996	Teaching Assistant for "Undergraduate Introductory Biology" at MIT Taught and advised 50 undergraduates in two sections.
Fall 1993	Exam Grader for "Inorganic Chemistry" at Brown University.
1990-1991	Lab Assistant for "Advanced Placement Biology" at Lincoln-Sudbury Regional High School.

Publications

Nathan P. Walsh and Robert T. Sauer "Structure and Specificity of Bacterial Protease Recognition Domains" Manuscript in preparation.

Matthew H. J. Cordes, Randall E. Burton, **Nathan P. Walsh**, C. James McKnight, and Robert T. Sauer "An Evolutionary Bridge to a New Protein Fold", *Nature Structural Biology* 2000, 7, pp. 1129-1132.

Matthew H. J. Cordes, **Nathan P. Walsh**, C. James McKnight, and Robert T. Sauer "Evolution of a Protein Fold *in Vitro*", *Science* 1999, 284, pp. 325-327.

Igor Levchenko, Catherine K. Smith, **Nathan P. Walsh**, Robert T. Sauer, and Tania A. Baker "PDZ-like Domains Mediate Binding Specificity in the Clp/Hsp 100 Family of Chaperones and Protease Regulatory Subunits", *Cell* 1997, 91, pp. 939-947.

Table of Contents

This page was intentionally left blank.

Table of Figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

This page was intentionally left blank.

Chapter 1. Introduction

Background

Proteases are essential to degrade proteins that are no longer needed, to remove misfolded proteins, to modulate precisely the concentration of signaling proteins, and to allow recycling of amino acids. They are also essential to activate pre-proteins, to process secreted proteins, and to generate peptide hormones. Most proteases need to discriminate between proteins that should be cleaved and those that should not. They do this in several ways. Classical proteases, like trypsin, recognize specific amino acids at or near the cleavage site within unfolded proteins. The mammalian 26S protease complex recognizes and degrades proteins conjugated to multiple ubiquitin molecules (Ciechanover, 1994, Hershko & Ciechanover, 1992). Other proteases recognize specific N and C-terminal targeting sequences in their substrates (Bowie & Sauer, 1989, Parsell *et al.*, 1990, Tobias *et al.*, 1991).

Recognition via C-terminal sequences

In bacteria, recognition of non-polar C-terminal sequence motifs is important for degradation of some proteins. Certain bacterial proteins, like CtrA in *Caulobacter crescentus,* must be degraded before the cell can enter S-phase in the cell cycle because active CtrA prevents initiation of DNA replication (Domian *et al.,* 1999). Degradation of CtrA by the ClpXP protease (Jenal & Fuchs, 1998) is dependent on the CtrA C-terminal sequence, which ends in Ala-Ala. Changing this sequence to Asp-Asp prevents CtrA from being degraded (Domian *et al.,* 1997).

During translation, bacterial proteins that are synthesized from damaged mRNA are frequently prematurely truncated. Such truncated proteins may have dominant-negative or other undesirable effects. To prevent deleterious consequences, a mechanism exists to ensure the degradation of these abnormal proteins. When the ribosome stalls on damaged mRNA, ssrA (tmRNA) acts as a tRNA to facilitate addition of a non-coded alanine, then acts as an mRNA to

replace the damaged mRNA, to direct the synthesis of the sequence ANDENYALAA, and to mediate ribosome release with a subsequent termination-codon (Keiler *et al.,* 1996, Tu *et al.,* 1995). This process results in addition of the ssrA tag (AANDENYALAA) to the truncated proteins' C-terminus. This tag targets the modified protein for degradation by the ClpXP or ClpAP proteases in the cytoplasm (Gottesman *et al.,* 1998), by the FtsH protease in the inner membrane (Herman *et al.,* 1998), and by tail specific protease (Tsp) in the periplasm (Keiler & Sauer, 1996). Tsp recognizes and degrades proteins with small hydrophobic residues at their C-terminus (Keiler *et al.,* 1995), of which ssrA tagged proteins are just one example.

Some eukaryotes, like spinach and *Scenedesmus obliquus*, contain an enzyme with homology to Tsp called CtpA or D1 protease. CtpA proteolytically removes a C-terminal extension of the precursor D1 protein while or just after the precursor D1 protein is inserted into the thylakoid membrane of the chloroplast (Diner *et al.,* 1988). Depending on the species, an eight to sixteen residue sequence that ends in A-P-(A/S)-(V/T/I)-(N/I/E)-(G/A) is specifically cleaved from D1 by CtpA. D1 is a component of the membrane-bound complex, photosystem II (PSII). PSII catalyzes water oxidation to evolve O_2 and transfer electrons to plastiquinone for photosynthesis. D1 is ligated to the manganese cluster and protects the rest of the PSII complex by scavenging radicals and thus absorbing oxidative damage. The oxidized D1 is subsequently degraded and new D1 is inserted. The crystal structure of CtpA reveals three domains (called A to C) (Liao *et al.,* 2000). The A domain (residues 78-147, 401-415) combined with the other two domains helps form a cleft where the substrate may bind, the B domain (residues 160-249) folds as a PDZ domain, and the C domain (254-400, 416-463) contains the active site. The PDZ domain is thought to be responsible for binding and recognizing the C-terminal sequence of D1. CtpA appears to be a serine-protease with serine 372 and lysine 397 being part of the active site.

Tail Specific Protease

Tsp, also known as protease Re and Prc (PRocesses C-terminus of FtsI), is a periplasmic, bacterial protease that functions to degrade ssrA-tagged proteins and to process other substrates (Keiler & Sauer, 1996). Tsp was identified independently by two different strategies. One group screened mutants for the inability to remove the C-terminus of FtsI (Hara *et al.,* 1991). Another group purified Tsp based on its ability to degrade a λ repressor variant with the C-terminal sequence WVAAA, but not one ending with RSEYE (Silber *et al.,* 1992). They sequenced part of the protein, and cloned the gene using degenerate primers based on the protein sequence. Cells with a deletion of Tsp are conditionally lethal and cannot grow at 42 ^{o}C on salt free media (Hara *et al.*, 1991, Silber & Sauer, 1994). Tsp appears to be a serine-protease, as mutagenesis experiments have identified serine 430, aspartate 441, and lysine 455 as likely active-site residues (Keiler & Sauer, 1995). Tsp is an endoprotease, cleaving target proteins at sites with alanine, serine, or valine at the P1 position and alanine, serine, valine, methionine, tyrosine, or tryptophan at the P1' position (Keiler *et al.,* 1995). However, the substrate specificity of Tsp depends on the target protein's C-terminal sequence as well as the P1/P1' positions. Proteins and peptides with non-polar residues at the C-terminus (e.g. AAA, AYA, LAA) are recognized and cleaved by Tsp (Keiler & Sauer, 1996). Because many of the characterized C-terminal sequences that target substrates to Tsp appear to be unstructured in solution (Silber *et al.,* 1992), it appears that these degradation tags act mainly to provide a binding or tethering site for the protease.

A domain in Tsp that recognizes appropriate C-terminal sequences has been identified (Beebe *et al.,* 2000). Photo affinity cross-linking of peptides ending in LAA to truncation mutants of Tsp demonstrated that the centrally located PDZ-like domain was responsible for

recognition. As described in Chapter 3, the isolated Tsp PDZ-like domain is independently folded and can bind proteins containing the ssrA tag. Peptide-recognition domains homologous to the Tsp PDZ-like domain are also found in two other periplasmic proteases, DegP (HtrA) and DegQ (HhoA), and in two membrane-bound proteases, DegS (HhoB) and EcfE (YaeL). A plasmid expressing DegQ and DegS complements a tsp-defective mutant for growth on salt-free media at 42 °C (Bass *et al.,* 1996) and over-expression of DegQ complements a degP-defective mutant for growth at high temperatures (Waller & Sauer, 1996). These results suggest that a number of these proteases may have overlapping substrate specificities or functions.

Cytoplasmic Heat-Shock Response

Regulated proteolysis is involved in signaling pathways. Bacteria such as *Escherichia coli*, *Pseudomonas aeruginosa,* and *Haemophilus influenzae* share highly conserved cytoplasmic and extracytoplasmic heat-shock responses. These two responses share several common features When bacterial cells are grown at elevated temperatures, like 42 °C, many proteins undergo complete or partial denaturation and aggregate. In response, bacteria undergo a change in gene expression and express a set of heat-shock proteins, including chaperones and proteases, to degrade, disaggregate or refold denatured and/or aggregated proteins.

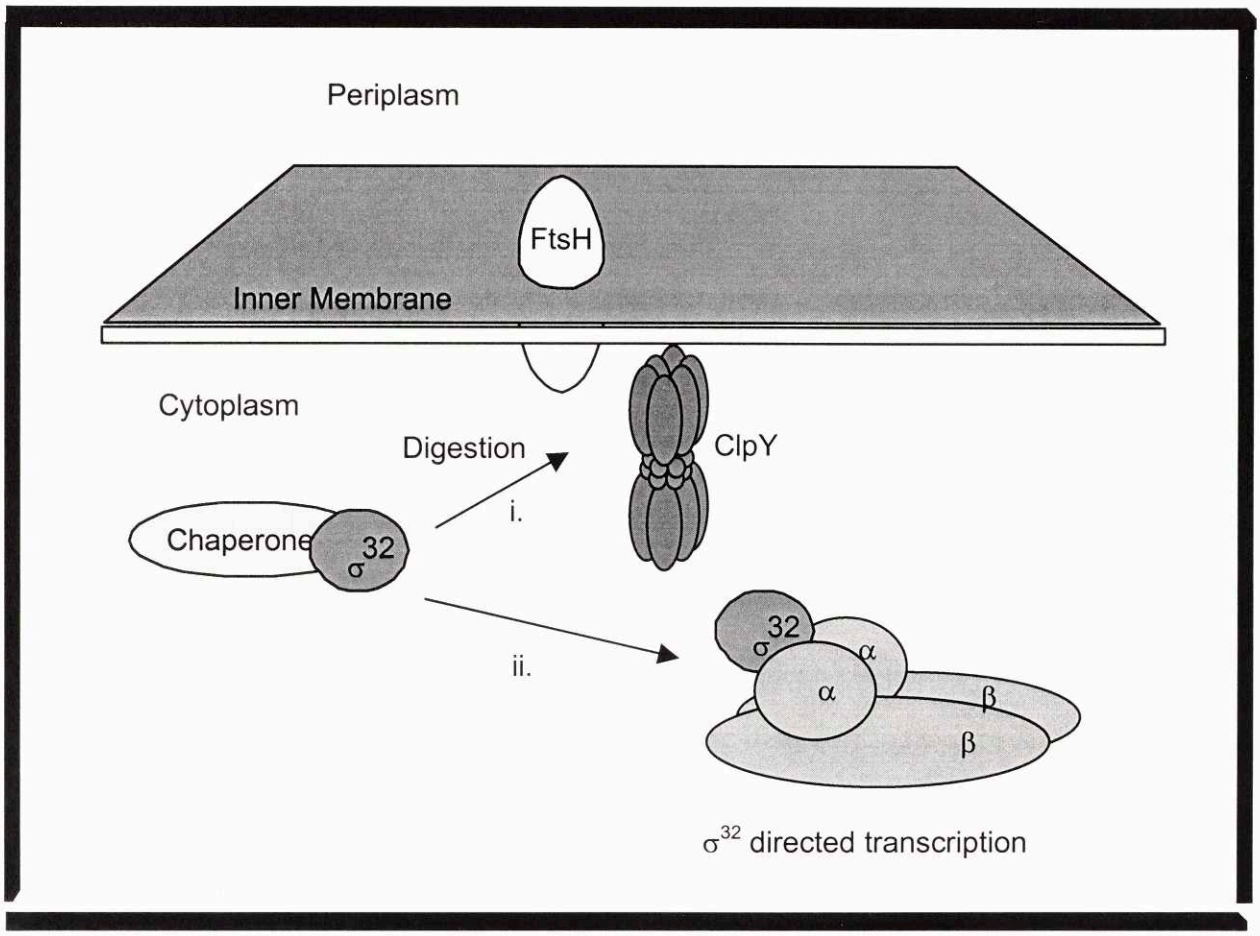

Figure 1-1. Model for the role of σ^{32} in the sensing of cytoplasmic stress. In the absence of heat shock, i, σ^{32} is bound to the DnaK chaperone system in an inactive state and it is targeted for degradation by FtsH, ClpYQ, and Lon. During heat shock, ii, other proteins unfold and preferentially bind chaperones and free σ^{32} can bind the RNA polymerase $\alpha_2\beta\beta$' subunits. This $E\sigma^{32}$ holoenzyme turns on the heat-shock genes.

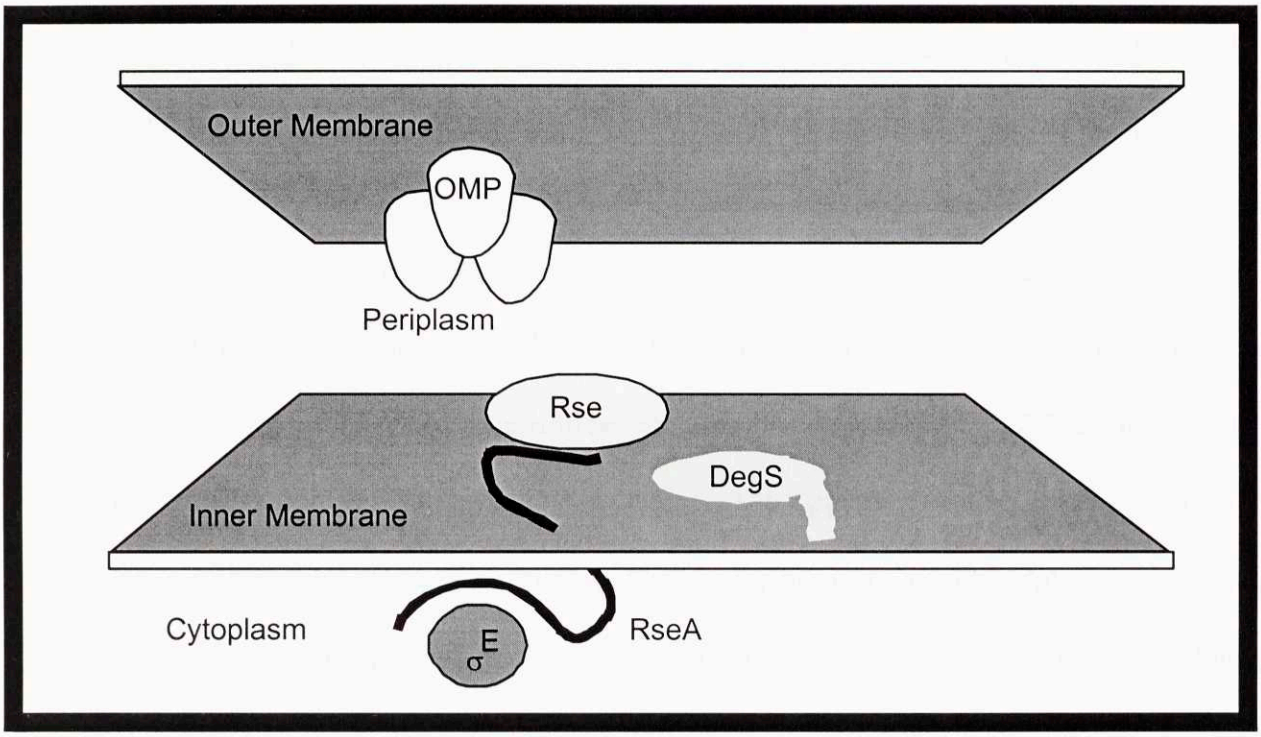

Figure 1-2. Model for the role of σ^E in the sensing of extracytoplasmic stress. In the absence of unfolded proteins in the periplasm, σ^E is bound to the antisigma factor RseA in an inactive state. RseB binds to RseA, increasing RseA inhibition of σ^E. Heat shock or OMP over-expression/misfolding leads to RseA degradation by DegS, freeing σ^E, which then binds RNA polymerase in the cytoplasm and activates the extracytoplasmic-stress genes.

In the cytoplasmic heat-shock response (Figure 1-1), σ^{32} (σ^{H}) is a transcription factor that activates heat-shock genes under appropriate conditions. Normally σ^{32} is degraded rapidly and consequently does not accumulate to high levels. The DnaK chaperone system (DnaK with the DnaJ and GrpE co-chaperones) binds σ^{32} (Liberek *et al.*, 1992, Liberek & Georgopoulos, 1993, Straus *et al.*, 1990, Tomoyasu *et al.*, 1998) and both blocks its activity and stimulates its rapid degradation by proteases such as FtsH (Herman *et al.,* 1995, Tomoyasu *et al.,* 1995), ClpYQ (Kanemori *et al.*, 1997, Kanemori *et al.*, 1999), and Lon (Kanemori *et al.*, 1999, Smith *et al.*, 1999). During heat shock, the unfolding of other proteins titrates the DnaK chaperone away from σ^{32}, leading to a decrease in σ^{32} degradation. Hence, σ^{32} is transiently stabilized (Straus *et al.,* 1987) and can bind RNA polymerase and direct transcription of heat-shock genes. The kinetics of this signaling pathway are fast because *de novo* synthesis of transcription factors is not required. Simple release of σ^{32} from DnaK and suppression of its degradation following heat shock quickly gives rise to a large pool of active σ^{32}. When more DnaK (itself under heat-shock control) is synthesized and/or the amount of competing denatured protein drops, σ^{32} is once again bound, inactivated, and degraded resulting in a rapid return to equilibrium.

Extracytoplasmic-Stress Response

The extracytoplasmic-stress response in *Escherichia coli* (Ades *et al.*, 1999, Mecsas *et al.*, 1993) (Figure 1-2) can be activated by the accumulation of outer-membrane protein (OMP) intermediates. Misfolded periplasmic proteins, heat shock, and osmotic stress can also activate this response, although to a lesser extent (Collinet *et al.*, 2000, Rouviere *et al.*, 1995). The extracytoplasmic-stress response is composed of two pathways: one of the pathways involves

Cpx proteins (reviewed by Raivio & Silhavy, 2001) and the other pathway, described here, uses σ^E. σ^E is a cytosolic transcription factor that activates many of the genes (like DegP, a periplasmic protease/chaperone) needed to cope with periplasmic stress (Erickson & Gross, 1989, Wang & Kaguni, 1989). The levels of σ^E are maintained by a feedback loop consisting of σ^E, RseA, RseB, RseC, and DegS (Ades *et al.*, 1999, De Las Penas *et al.*, 1997, Missiakas *et al.*, 1997, Rouviere *et al.*, 1995). σ^E positively autoregulates itself and the three rse genes, all of which are part of a single operon. In *E. coli*, DegP is not part of this operon but its expression is controlled by σ^E, whereas in *Pseudomonas aeruginosa* the DegP homologue (MucD) is part of the operon (Martinez-Salazar *et al.*, 1996). RseA is an inner-membrane protein that binds and inactivates σ^E under normal periplasmic conditions (De Las Penas *et al.*, 1997). RseB and RseC are periplasmic proteins that modulate the activity of RseA (De Las Penas *et al.*, 1997). When misfolded proteins accumulate in the periplasm, RseA is degraded by DegS (Ades *et al.*, 1999), a membrane-bound endoprotease (Bass *et al.*, 1996, Waller & Sauer, 1996). This degradation of RseA leads to a rapid increase in cytoplasmic σ^E activity, causing increased transcription of rpoE, rseA, rseB, rseC, chaperone and protease genes (fkpA, surA, dsbC, skp, degP and ecfE), lipid biosynthesis genes (htrM, lpxD, lpxA, and ecfA), genes for other sigma factors (rpoH and rpoD), and several genes of unknown function (mdoG, cutC, nlpB, ecfD (yfiO), ecfF (yggN), ecfG (htrG), ecfH (yraP), ecfI (yidQ), ecfJ (ytfJ), ecfK (UP0), ecfL (yqjA)) (Dartigalongue *et al.*, 2001). When misfolded proteins are no longer present in the periplasm, degradation of RseA by DegS ceases, the levels of RseA increase, and RseA inactivates σ^E, bringing the cell to its pre-stress state. What is not known is how extracytoplasmic-stress signals are recognized and passed to σ^E via RseA. Because over-expression of the outer-membrane proteins OmpC, OmpF, and OmpX leads to the accumulation of folding intermediates in the periplasm, which activates the

pathway, it has been proposed that chaperones are titrated away from RseA by misfolded proteins, making RseA susceptible to proteolysis (Mecsas *et al.,* 1993). Alternatively, it has been proposed that RseB may sense misfolded protein and subsequently dissociate from RseA, allowing release of σ^E (Collinet *et al.,* 2000). Both of these prevailing theories place the critical sensing role with one of the Rse proteins and suggest a passive role for DegS. Chapter 2 presents evidence that suggests DegS might play a more active role in sensing misfolded OMPs. Specifically, the PDZ-like domain in DegS binds a sequence found in outer-membrane proteins. This sequence is not exposed in native OMPs, but would likely be exposed in misfolded or aggregated OMPs.

Outer-membrane proteins require numerous steps for proper localization, folding, and assembly. Problems with these processes can result in a large population of unassembled protein in the periplasm. Native OMPs are trimers of β-barrel subunits each with a solvent exposed core that allows passage of selected molecules through the outer-membrane. In each beta-barrel, one face is hydrophobic and packs against the membrane lipids. The other face is polar. This results in an alternating pattern of hydrophobic and hydrophilic residues over most of the sequence. OMPs are completely translated in the cytoplasm, where they bind trigger factor co-translationally, preventing premature folding. The N-terminal signal sequence of the OMPs binds to SecB, which targets them to the secretory machinery involving SecA, SecY, SecE, and SecG. OMPs and Sec-dependent periplasmic proteins are targeted to the periplasm in the same way and the signal sequences are interchangeable (Tommassen *et al.,* 1983). Long stretches of hydrophobic residues are not common in the OMPs as these may act as stop-transfer sequences and prevent them from reaching the periplasm. As the OMPs are exported into the periplasm, the signal sequence is removed and the translocated OMPs bind to Skp. Skp is a periplasmic

chaperone that binds both lipopolysacharides and several distinct portions of unfolded OMPs. Other periplasmic proteins involved in protein folding are also important for proper assembly of the OMPs. These include peptidyl-prolyl isomerases (FkpA, SurA, RotA, and PpiD) that catalyze inter-conversion of the *cis* and *trans*-peptidyl-prolyl bond and disulphide-bond isomerases (DsbA, DsbB, DsbC, DsbD, DsbE, and DsbG) that catalyze formation of disulphide bonds as well as their isomerization. Once the OMP monomers are folded, they bind lipopolysacharide and are inserted into the outer membrane as trimers. The signal for targeting to the outer membrane may be coded in the primary sequence, the three-dimensional structure, or in some aspect of a folding intermediate. Removal of the C-terminal β-strand of OMPs prevents trimerization and outer-membrane localization (Bosch *et al.,* 1989). The C-terminal sequence of most OMPs is highly conserved with a tyrosine at the antepenultimate position and a phenylalanine at the C-terminus. Mutation of the phenylalanine to less hydrophobic residues reduces the incorporation of the OMPs into the outer membrane *in vivo* (Struyve *et al.,* 1991), but not *in vitro* (de Cock *et al.,* 1997, Jansen *et al.,* 2000), suggesting that proteins not present in the *in vitro* reactions may specifically recognize the C-terminus to aid in folding, insertion, or degradation.

Domains in the DegS family of proteases

DegS belongs to a family of proteases containing a protease domain and one or more peptide-recognition domains (Figure 1-3). The protease domains of DegS, DegP, and DegQ contain a conserved serine at position 201, a conserved histidine at position 96, and a conserved aspartate at position 126 (numbered according to the DegS sequence), which appear to form a classical serine-protease catalytic triad (Waller & Sauer, 1996). DegP and DegQ also contain a signal sequence and two peptide-recognition domains, whereas DegS has a membrane-anchor

sequence and a single peptide-recognition domain. The peptide-recognition (PDZ-like) domains found in DegS, DegP and DegQ are homologous to domains present in the Tsp, CtpA, and EcfE proteases. Only some of these proteases share similar active-site domains. For instance, in EcfE, the active site domain appears to use a mechanism common in Site 2 proteases (S2P), which are zinc metallo-proteases that cleave substrates at membrane-embedded positions. Interestingly, the peptide-recognition domain appears to be inserted into the middle of the protease domain in EcfE. In Tsp, the peptide-recognition domain is on the N-terminal side of the protease domain, whereas the peptide-recognition domains are on the C-terminal side of the protease domains of the DegS, DegP and DegQ enzymes. These observations emphasize the modularity of these peptide-recognition domains. The peptide-recognition domains of DegS, DegP, DegQ, Tsp, EcfE, and CtpA are also functionally and sequentially homologous to eukaryotic PDZ domains (Ponting, 1997, Ponting *et al.*, 1997).

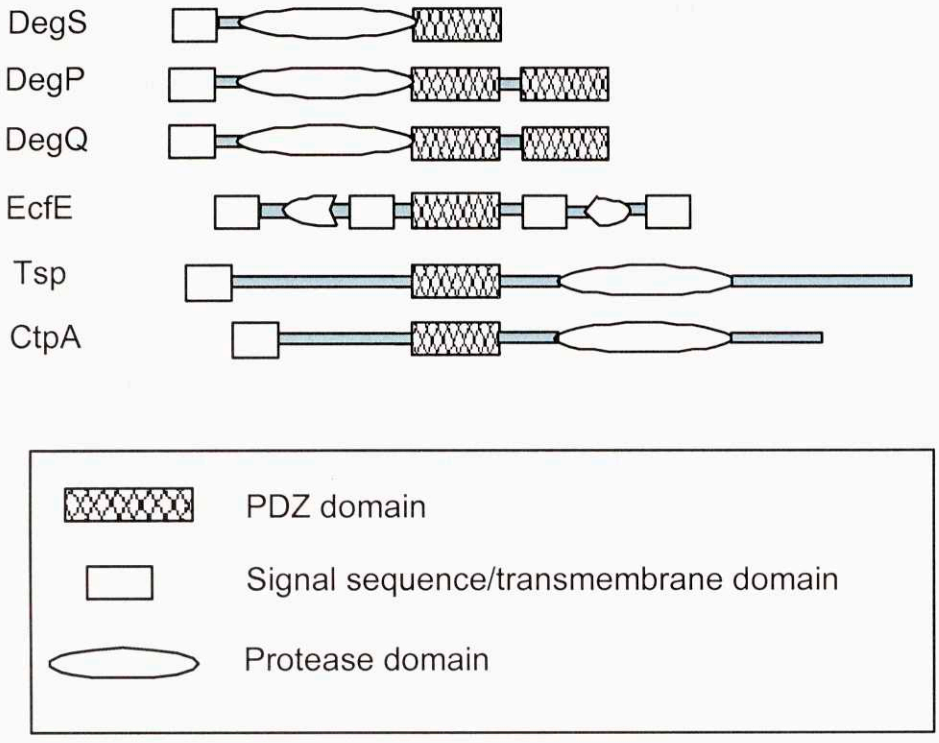

Figure 1-3. Domain structures of proteases containing a PDZ domain. DegS, DegP, and DegQ are highly related. DegS has a transmembrane domain, a protease domain, and a PDZ domain and is located in the inner membrane with the PDZ and protease active site exposed to the periplasm. DegP and DegQ are periplasmic proteins and have a signal sequence that is cleaved in the mature form, a protease domain and two PDZ domains. EcfE is thought to be held in the inner membrane by four transmembrane domains and positioned so the PDZ domain is facing the periplasm and the protease domain is facing the cytoplasm. Tsp and CtpA are extracytoplasmic, but have protease domains that are very different from the other proteases.

PDZ domains

Proteins containing PDZ domains play important roles in cell signaling. PDZ domains were first identified as an approximately 100 amino-acid sequence repeated nine times in PSD95 and also found in Discs-large and ZO-1 (Cho *et al.*, 1992, Woods & Bryant, 1993). PDZ domains have subsequently been identified in hundreds of proteins including some from plants and bacteria (Ponting, 1997, Ponting *et al.*, 1997) (http://www.ocms.ox.ac.uk/~ponting/PDZ/pdz_align.html).

In eukaryotes, PDZ domains are usually found in membrane-associated proteins and are responsible for binding and clustering receptors and other regulatory molecules (reviewed by Fanning & Anderson, 1996, Saras & Heldin, 1996). PDZ domains are capable of recognizing short peptide sequences in other proteins (usually at the C-terminus) and acting as scaffolds to allow the formation or localization of large complexes (Figure 1-4). In humans, NHERF1 is a Na^+/H^+ exchange regulatory factor, which contains two PDZ domains. Both PDZ domains can bind to the C-terminus of the cystic fibrosis transmembrane conductance regulator, CFTR (Raghuram *et al.*, 2001). The CFTR is a Cl^- channel whose activity depends on its phosphorylation state. CFTR is phosphorylated on multiple sites by cAMP-dependent kinase and protein kinase C (Riordan *et al.*, 1989). NHERF1 has a C-terminal ERM binding domain. One of the ERM proteins, ezrin, links NHERF1 to the actin cytoskeleton, and may be responsible for both the CFTR localization at the apical membrane and its proximity to cAMP-dependent kinase (Moyer *et al.*, 2000). Moreover, NHERF1 is able to bind other membrane proteins. In a PDZ independent manner, NHERF1 binds Na^+/H^+ exchanger 3, NHE3, (Yun *et al.*, 1997) and using the first PDZ domain NHERF1 binds the β2-adrenergic receptor (Hall *et al.*, 1998b). As many of these components can bind several different targets, the size and composition of the

complexes that form can be large and varied. The PDZ domains act as adapters, which bring together proteins that need to interact.

Most PDZ domains have been shown to bind to the C-terminal sequences of their target proteins, although some can bind to internal target sequences (Hillier *et al.,* 1999). PDZ domains can have different target recognition sequences. Many PDZ domains, including the third PDZ domain of PSD-95, recognize the C-terminal sequence (S/T)-x-V-COOH (Kornau *et al.,* 1995). The first PDZ domain of NHERF recognizes (S/T)-x-L-COOH (Hall *et al.,* 1998a). Using peptide libraries to identify PDZ binding specificity, a number of non (S/T) recognition sequences were identified (Songyang *et al.,* 1997). For example, the PDZ domains of p55, human LIN-2, Tiam-1, and AF-6 were found to bind to aromatic C-terminal sequences such as F-x-x, F-F-(V/F/A), (Y/F)-(Y/H)-(F/A), and (F/I/M/Y)-(Y/F)-(V/F/I/L) respectively. Most PDZ domains have nanomolar affinity for their preferred target peptide and roughly 1000-fold lower affinities for other peptides. For those PDZ domains that are known to bind internal sequences , the internal sequence must be in a beta-sheet and be followed by a tight turn. Sequences in a beta-sheet not followed by a hairpin turn would have a steric clash between the carboxylate-binding loop of the PDZ domain and the continuing residues of the peptide, which should prevent binding.

Proteases that contain PDZ-like domains are thought to use these domains for targeting and localization of substrates. Unlike their cytoplasmic counterparts, Tsp, DegS, and other extracytoplasmic proteases are ATP-independent proteases. Hence, they cannot use the energy of ATP hydrolysis to help to unfold proteins targeted for degradation. This suggests that extracytoplasmic proteases must wait for native substrates to unfold before they can be degraded.

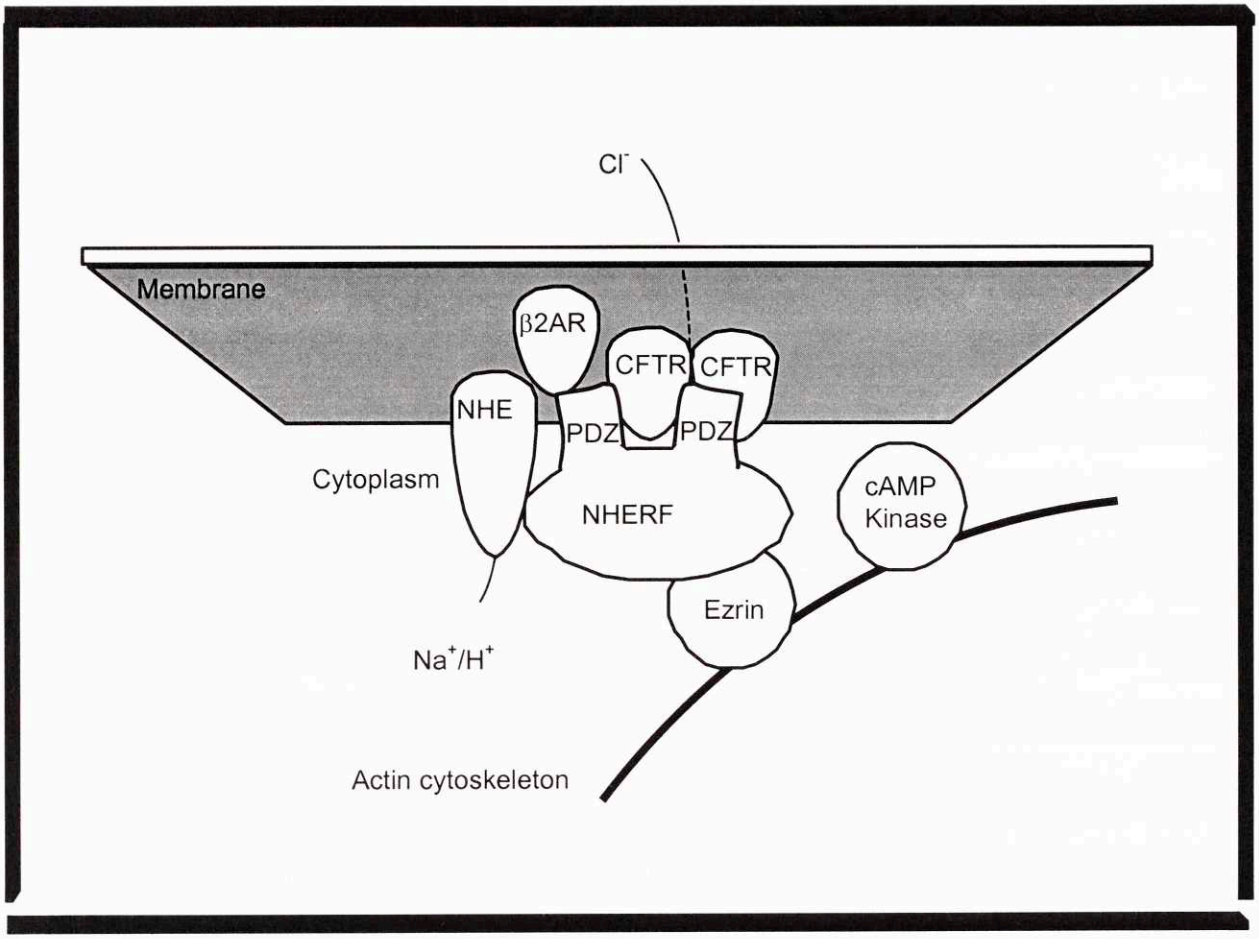

Figure 1-4. Model of a typical eukaryotic complex involving a protein with PDZ domains. NHERF1 is involved in activating the CFTR by bringing together inputs from several proteins. The CFTR is activated by dimerization and by being localized near cAMP kinase, which phosphorylates many sites on the CFTR. NHERF1, like most eukaryotic PDZ containing proteins, binds several membrane-bound receptors including the NHE3 and the β2-adrenergic receptor. These receptors may influence the activity of the CFTR.

By tethering native substrates, the protease has a much greater chance of degrading the substrate as soon as it unfolds, because there is no requirement for an intermediate, bimolecular binding step. Tsp uses its PDZ-like domain to tether substrates bearing unstructured, hydrophobic C-terminal sequences (Beebe & Pei, 1998, Beebe *et al.*, 2000, Keiler & Sauer, 1996, Keiler *et al.*, 1995). For proteases like DegP, the PDZ-like domains may act as substrate tethers (Pallen & Wren, 1997), but also appear to play some role in stabilization of the DegP hexamer (Sassoon *et al.,* 1999).

PDZ Structure

Table 1-1 lists PDZ domains or proteins containing them whose structures are known. The basic PDZ structure has two beta-sheets that form a sandwich and two alpha-helices at adjacent edges of the sandwich (Figure 1-5). Depending on the criteria used to assign secondary structure, the top beta-sheet has two to three strands and the bottom beta-sheet has four strands.

One strand of the bottom beta-sheet (B0 in Figure 1-5) can be formed by either the N or the C-terminal sequences. This represents a circular permutation of the sequence in the three-dimensional structure. The GLGF loop, which leads into strand B1, is a hallmark of PDZ domains. The first glycine of the GLGF motif is highly conserved in most PDZ domains, although serine and proline are occasionally found at this position. This glycine allows a water molecule to enter the binding pocket and form a hydrogen bond to the α-carboxylate of bound peptides. When serine is present instead of glycine, the serine hydroxyl forms a hydrogen bond directly with the peptide C-terminus. The leucine and phenylalanine at the second and fourth positions of the GLGF motif pack into the peptide-binding pocket. Some sequences, like those belonging to DegP, DegQ, and DegS, have a glycine at one of these positions. A variable length

Table 1-1. Known PDZ domain structures

Protein Name	PDB identifier	Date
third PDZ of post-synaptic density protein 95	1BE9, 1BFE (Doyle *et al.*, 1996)	1996
third PDZ of human Discs Large Protein	1PDR (Morais Cabral *et al.*, 1996)	1996
human cask/Lin-2	1KWA (Daniels *et al.*, 1998, Doyle *et al.*, 1996)	1996, 1998
interleukin 16	1I16 (Muhlhahn *et al.*, 1998)	1998
syntrophin	1QAV, 2PDZ (Hillier *et al.*, 1999, Schultz *et al.*, 1998)	1998, 1999
neuronal nitric oxide synthase	1B8Q, 1QAU, 1QAV (Hillier *et al.*, 1999, Tochio *et al.*, 1999)	1999
second PDZ of post-synaptic density protein 95	1QLC (Tochio *et al.*, 2000)	2000
photosystem II D1 protease	1FC6, 1FC7, 1FC9, 1FCF (Liao *et al.*, 2000)	2000
second PDZ of human phospho-tyrosine phosphotase 1E	3PDZ (Kozlov *et al.*, 2000)	2000
first PDZ of the Na^+/H^+ exchanger regulatory factor	1G9O, 1I92 (Karthikeyan *et al.*, 2001a, Karthikeyan *et al.*, 2001b)	2001

loop (L1) from 3 to 14 residues links strands B1 and B2. Strand B2 forms hydrogen bonds with B1 and B3, although both B2 and B3 have a single residue kink, or aspartate beta bulge, which breaks the usual binary pattern of alternating polar and hydrophobic residues. The residue before the bulge usually has a psi angle of –35 to –50 degrees instead of the normal value of 110 to 180 degrees for beta-sheets. The first alpha-helix (α1) connects B2 and B3. This helix is short (5

residues) and consists almost entirely of alanine, proline, and glycine. A type II turn before the alpha-helix leads into a glycine N-cap with proline occupying the N1 position of the alpha-helix. The α1 helix is terminated by a type I turn. A type II turn connects beta-strands B3 and B4. The most conserved residue in PDZ domains is an aspartate at the i+3 position of this turn. The aspartate side chain helps form the turn by hydrogen bonding to the amide backbone of the i^{th} position of the turn.

The bottom beta-sheet of the PDZ domains also has an anti-parallel arrangement, with B4 and B6 as the internal beta-strands, and B5 and either B7 or B0 as edge beta-strands. A type I' turn connects beta-strands B4 and B5. Beta-strands B5 and B6 are connected by a long alpha-helix (α2) flanked by loops of variable lengths. This region near the α2 helix shows the greatest structural divergence among different PDZ domains, including different types and lengths of turns both before and after the helix. For example, the 18 PDZ structures listed in Table 1-1 use three different prototypical turns (type I, II, and VIb) along with a large number of non-prototypical turns that do not form complete turns within four residues. The second alpha-helix is also responsible for the many of the contacts with bound peptides. Because of the flanking variable-length loops, the loop-helix-loop regions of PDZ domains are aligned differently by different alignment algorithms.

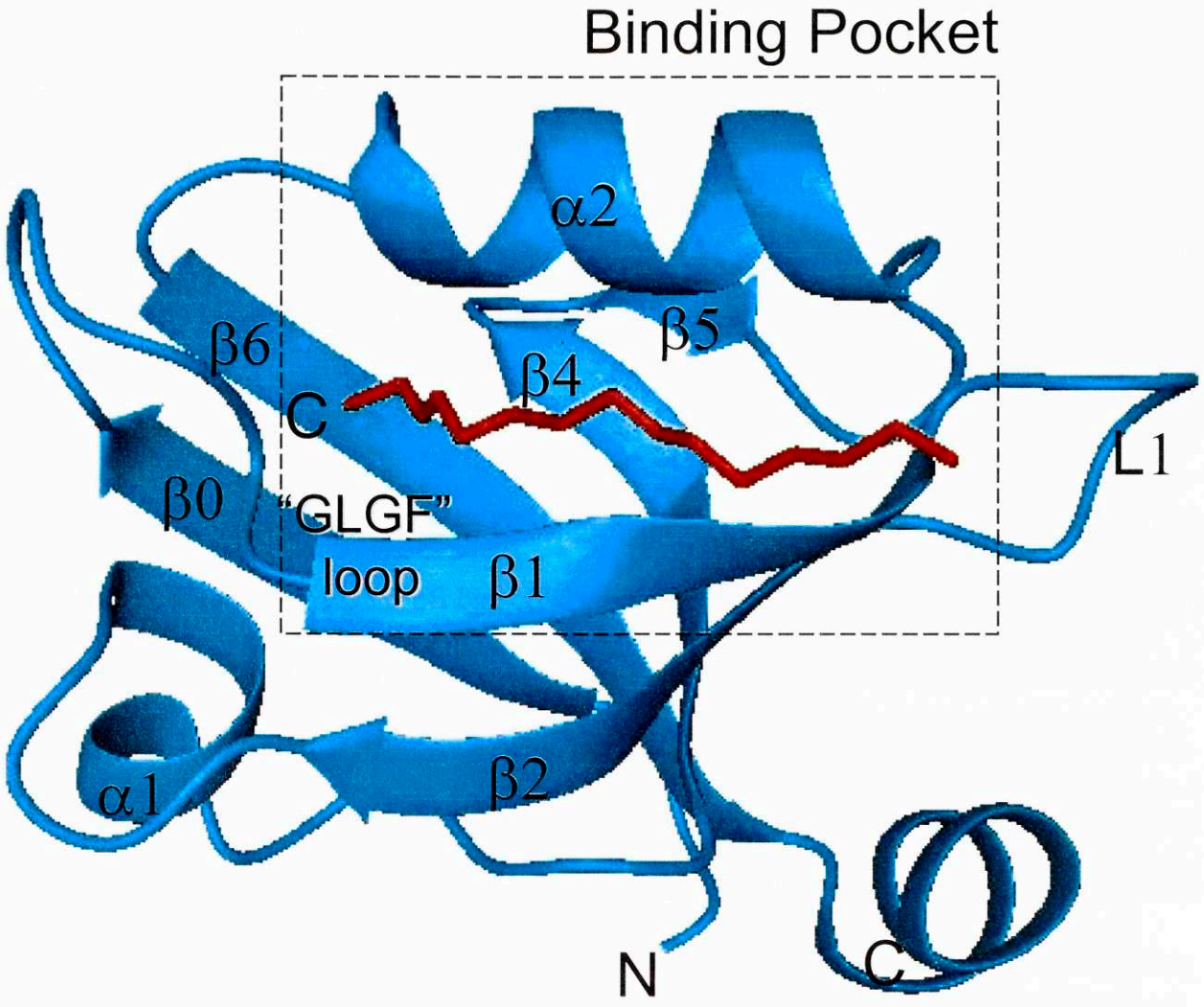

Figure 1-5. Structure of the third PDZ domain of PSD-95 by Doyle *et al.* PDZ domains have two beta-sheets and several alpha-helices. The top beta-sheet of this PDZ domain is composed of beta-strands B1 and B2, while the bottom sheet is composed of beta-strands B0, B6, B4, and B5. The peptide-binding pocket is comprised of the GLGF loop, beta-strand B1, and helix α2. The bound peptide is shown in red. In this structure, there is a beta-strand B0 and no B7, the N and C-termini are close together and located at the bottom of the structure.

Goals

DegS is a component of the extracytoplasmic stress-signaling pathway found in many bacteria. When I began the work described in Chapters 2 and 3, there was no structural information about the PDZ-like domain of DegS, its close relatives, or Tsp. Moreover, it was far from certain whether these domains had the PDZ fold, as the homology and sequence identity were poor. In addition, there was no information about possible target sequences of the DegS PDZ-like domain. In Chapter 2, I characterize many of the biophysical and structural properties of the DegS PDZ-like domain. I also use a peptide-library method to identify peptide sequences that bind the DegS PDZ domain and determine the binding properties of some of these sequences. This data combined with informatic studies, suggest that the PDZ domain of DegS may bind to the C-terminal sequences of OMPs. The implications of this finding for signaling are discussed. In Chapter 3, I characterize the biophysical properties of the Tsp PDZ domain, including protease mapping and peptide binding.

Chapter 2. DegS PDZ domain

Introduction

DegS is a membrane-bound protease involved in the response to extracytoplasmic stress (Ades *et al.*, 1999). The PDZ-like domain of DegS has homology to peptide-binding domains in other proteases. In one of those proteases, Tsp, the PDZ-like domain has been shown to be responsible for substrate specificity (Beebe *et al.*, 2000 and Chapter 3). Very little is known about the biochemical properties of DegS, its substrate specificity, or its role in regulation of the extracytoplasmic stress response. As an initial step in addressing some of these questions, the PDZ-like domain of DegS was purified and its structure was modeled using a combination of homology modeling and NMR restraints. This domain appeared to have a circularly permuted PDZ fold, similar to that observed in the CtpA structure. Peptide libraries were used to determine the specificity of the DegS PDZ-like domain, revealing a preference for C-terminal sequences that end in tyrosine-tyrosine-phenylalanine. Peptides containing this C-terminal sequence bind the DegS domain with sub-μM affinity. Using bioinformatics, outer-membrane proteins were revealed as likely targets of the DegS PDZ-like domain. Unassembled outer-membrane proteins are known to increase the proteolysis of RseA by DegS and several potential mechanisms of this signaling are discussed.

Materials and Methods

Alignments

Sequence alignments of PDZ domains were initially created based on super-positions of the known three-dimensional structures of 1FCF, 1B8Q, 1PDR, 1BE9, 1I16, 1KWA, 1QAV, 1QLC, and 3PDZ. From this initial alignment, hydrophobic residues whose side chains were buried or partially buried were identified as "core" positions. In addition, positions where a single amino acid or a small subset of residues was found were labeled as "unique" positions.

Requiring the conservation of core and unique positions led to unambiguous alignment of most new sequences of unknown structure, with the exceptions of the termini and the second alpha-helix which has loops of variable lengths at either end. Initially these regions were examined by eye, threaded onto the known structures, and made to conform to expectations that hydrophobic residues should be buried. Straight pair-wise alignments between known structures and new sequences often gave alignments that put large insertions into secondary structure elements or resulted in conflicting alignments depending on which known structure was used.

Expression and stability

The DegS PDZ domain (residues 245-355) was expressed from a pET expression plasmid (pLC263 provided by Carol Gross, UCSF) with a C-terminal 6-His tag in strain BL21(DE3). A 1 L minimal media prep (M9 plus 3 g L^{-1} glucose, 10 mL 0.1 M $MgSO_4$, 0.2 mL 0.5 M $CaCl_2$, 1 mL 0.01 M $FeCl_3*6H_20$, 2 mL 0.5 mg mL^{-1} thiamine-B1*H_2O, 0.8 g L^{-1} NH_4Cl) typically yielded about 2-3 μmoles of the DegS PDZ domain, when induced with 1 mL 100 mM IPTG at an OD_{600} of 0.6 for 4 hours. Harvested cells were resuspended in 20 mL of buffer S (300 mM NaCl, 10 mM imidazole, and 50 mM $NaPO_4$, pH 8.5) per L of culture and lysed by sonication. After centrifugation to remove unlysed cells and debris, the supernatant was loaded onto a 2 mL column of Ni-NTA resin (Qiagen) and the flow-through fraction was reapplied. The column was washed with 100 mL of buffer S and bound protein was eluted with 10 mL of this buffer plus 200 mM imidazole. The eluted protein was dialyzed against buffer A (10 mM sodium acetate, pH 5) and was loaded onto a Mono S HR5/5 column (Pharmacia Biotech) on an Äkta FPLC system, at a rate of 1 mL min^{-1}. A 30 min linear gradient from 0 to 1 M NaCl in buffer A was run beginning 20 min after injection. The DegS PDZ domain eluted at 300-500 mM NaCl. Fractions were pooled, dialyzed against two changes of 6 L of either MiliQ H_2O or 50 mM

NH_4HCO_3, frozen at -80 °C, and lyophilized. The lyophilized protein was stored at -20 °C. For experiments requiring ^{15}N and/or ^{13}C labeled protein, the DegS PDZ domain was expressed in media containing either 0.8 g L^{-1} of $^{15}NH_4Cl$ and/or 3 g L^{-1} of ^{13}C glucose, respectively.

At pH values less than 4.5 or at temperatures less than 10 °C or higher than 30 °C, the DegS PDZ domain aggregated and could not be resolubilized in native form. At room temperature, a typical sample of the DegS PDZ domain could be used for a period of days to months before aggregating. At 30 °C, by contrast, precipitation often occurred immediately. This solubility problem was not improved by using low or high salt conditions or by varying cations and anions (Na^+, Mg^{++}, Ca^{++}, Cl^-, SO_4^{--}, and PO_4^{---}).

The DegS PDZ domain contains two tyrosines and concentrations were determined by OD_{278} using an extinction coefficient of 2520 M^{-1} cm^{-1} at neutral pH in an 8452A Diode Array Spectrophotometer (Hewlett Packard). Circular dichroism experiments were performed using a 60DS Circular Dichroism Spectrometer (Aviv), typically at a protein concentration of 25 μM in 50 mM $NaPO_4$ (pH 7.6) with a 1 minute equilibration time and 30 second averaging. Thermal melts were fit using software (Nathan Walsh, unpublished) that utilized the gnuplot program (Dartmouth).

Peptide Selections

Two peptide libraries (NH_2-KNxxxxxx-COOH and NH_2-DNxxxxYxF-COOH; where x denotes all 20 amino acids except cysteine) were synthesized (MIT Biopolymers Laboratory). All 19 amino acids were simultaneously added to resin at each coupling step, although different concentrations of specific amino acids were used because of high or low coupling efficiencies. The selection experiments were based on the general method by Songyang *et al.* (1997) and were performed at 4 °C. Two columns were made from 1 mL syringes using glass wool to form a plug

and 200 μL of 50% isopropanol / Affigel-15 slurry (Biorad). Air pressure from a large syringe connected by plastic tubing was used to force 3 x 1 mL aliquots of water through the resin. To the experimental column, 100 μL of 4 mg mL^{-1} solution of the DegS PDZ domain in 100 mM $NaHCO_3$ (pH 8.4) was added. To the control column, 100 μL of 500 mM glycine was added. In both cases, the N-hydroxysuccinimide ester reactive group on the Affigel-15 resin was reacted with these solutions for 4 hours. After this time, 1 mL of 500 mM glycine was added to both columns to block any remaining sites. Both columns were washed with 4 x 1 mL aliquots of wash buffer (100 mM $NaHCO_3$, pH 8.4, 100 mM NaCl, and 0.5% Triton X-100). A portion of the peptide library (2 mg in 90 μL of 100 mM $NaHCO_3$, pH 8.4) was centrifuged to remove insoluble peptides and half of the supernatant was added to each column. The peptide mixture was allowed to bind for 15 min and 2-4 x 1 mL washes were performed rapidly (peptide analysis was contaminated with glycine when this wash was only 2 mL). Both the columns were washed rapidly with an additional 1 mL of water, and eluted with 1 mL of 30% acetic acid under gravity flow. The elutants were dried in a Speed-Vac Concentrator (Savant) overnight, resuspended in 20% acetonitrile/1% TFA and adsorbed onto a filter made of micro-TFA glass fibers treated with biobrene, and subjected to N-terminal Edman degradation (MIT Biopolymer Laboratory). PTH-amino acid derivatives were identified and quantified by HPLC analysis. HPLC traces from the two columns were compared to determine binding preferences at each randomized position.

Peptide synthesis and purification

Peptides sol-ssrA (NH_2-NKKGRHGAANDENYALAA-COOH), OmpC C-terminus (NH_2-DNVGLVYQF-COOH), solYQM (NH_2-DNRDGNVYQM-COOH), solYQF (NH_2-DNRDGNVYQF-COOH), and solYYF (NH_2-DNRDGNVYYF-COOH) were synthesized

and the expected molecular weights were confirmed by MALDI mass spectroscopy (MIT Biopolymers Laboratory). The peptides were purified over a Vydac 218TP C18 reverse phase column (4.6 x 250 mm) using a Shimadzu VP series HPLC. A linear gradient from 0.06% TFA to 80% acetonitrile/0.052% TFA was applied and the peptides were found to elute between 20 and 30% acetonitrile. Fractions were pooled, dried under vacuum, and stored at room temperature.

NMR experiments

NMR studies were performed using a Bruker AMX-500 (Whitehead Institute), a Bruker DMX-500 (Boston University Medical School, Department of Biophysics), or a Bruker DMX-600 (MIT Francis Bitter Magnet Lab). Data processing was performed using NMRPipe1.8 (Delaglio *et al.,* 1995). Spectral analysis was carried out using NMRview 4.1.3 (Johnson & Blevins, 1994). Structural models were calculated using CNS1.0 (Brunger *et al.,* 1998), and molecular graphics used a custom version of MOLMOL version 2.6 (Koradi *et al.,* 1996) called MOLMOL2k.1.

Five samples of DegS PDZ domain (one unlabeled, two ^{15}N labeled, and two $^{15}N/^{13}C$ labeled) at 298 K were used for all of the experiments. A non-isotopically labeled sample of the DegS PDZ domain at 2.5 mM (pH of 5.8) was used for 2D-NOESY (50 msec), 2D-DQF-COSY, and 2D-TOCSY (69 and 100 msec) experiments. An ^{15}N labeled sample at 2.1 mM (pH 5.8) was used for 2D-^{15}N-HSQC and 3D-^{15}N-NOESY (150 msec) experiments. The second ^{15}N labeled sample at 0.5 mM (pH 6.9) was used for the 2D-^{15}N-relaxation experiments and the solYYF peptide titration. A $^{13}C/^{15}N$ labeled sample at 1.8 mM (pH 5.8) was used for 2D-^{13}C-HSQC, 3D-^{13}C-NOESY (80 msec), 3D-HCCH-TOCSY (23.3 msec), 3D-HNCACB, and 3D-CBCA(CO)NH experiments. The other $^{13}C/^{15}N$ labeled sample at 2.8 mM (pH 5.8) was used for the solssrA

peptide titration. Table 2-1 lists the acquisition parameters and Table 2-2 lists the processing parameters that were used for the three-dimensional NMR experiments. The pulse lengths and delays were all within normal settings. The ^{1}H 90° was calibrated to 12.3 microseconds for the HSQCs, CBCA(CO)NH, HNCACB, ^{13}C-NOESY, and HCCH-TOCSY. Gradient water suppression was employed in all 3D experiments and either gradients or pre-saturation was used to suppress water in the 2D experiments.

Peaks were identified interactively using NMRview. Peaks were selected only if they were above background in at least 3 successive planes in all dimensions, were located in an expected chemical-shift range, and were sufficiently distinct from overlapping peaks to allow a center to be determined.

Three types of ^{15}N-relaxation experiments — T1, T2 (Farrow *et al.,* 1994) and heteronuclear-NOE — were performed. ^{15}N T1-relaxation experiments were performed with delays of 20, 60, 140, 240, 370, 570, 750, and 1140 milliseconds. ^{15}N T2-relaxation experiments were performed with delays of 8.4, 25, 40, 58.6, 75, 108, 142, and 246 milliseconds. The error was estimated by repeated trials of selected delays. The heteronuclear-NOE was calculated from spectra with and without NOE saturation. Values larger than 110 milliseconds for the T2-relaxation and smaller than 0.5 for the heteronuclear-NOE indicated amides in disordered regions.

Table 2-1. Acquisition Parameters for 3D NMR experiments

Experiment	CBCA(CO)NH [1]	HNCACB [2]	^{13}C-NOESY [3]	HCCH-TOCSY [4]	^{15}N-NOESY [5]
Field Strength	600	600	600	600	500
F1 nuclei	N	N	C	H(C)	N
Points	21 complex	21 complex	32 complex	78 complex	64 complex
Mode	Echo-antiecho	Echo-antiecho	Echo-antiecho	Complex	Echo-antiecho
Sweep width	1824.068 Hz	1824.069 Hz	12070.006 Hz	5296.610 Hz	1419.142 Hz
Center	120.480 ppm	118.48 ppm	42.38 ppm	4.463 ppm	117.008 ppm
F2 nuclei	C	C	H	C	H
Points	142 complex	75 complex	128 complex	64 complex	256 real
Mode	States-TPPI	States	States	Complex	TPPI
Sweep width	10562.450 Hz	10562.450 Hz	6613.757 Hz	12070.006 Hz	5580.377 Hz
Center	42.38 ppm	42.38 ppm	4.771 ppm	42.38 ppm	4.789
F3 nuclei	H	H	H(C)	H	H(N)
Points	512 complex	512 complex	512 complex	512 complex	512 complex
Mode	Digital Quad Detection	Digital Quad Detection	Complex	Complex	Complex
Sweep Width	6613.757 Hz	6613.757 Hz	6613.757 Hz	5296.610 Hz	5580.357 Hz
Center	4.772 ppm	4.771 ppm	4.771 ppm	4.463 ppm	4.789

1: (Grzesiek *et al.*, 1992, Muhandiram & Kay, 1994)
2: (Kay *et al.*, 1992, Muhandiram & Kay, 1994, Palmer *et al.*, 1991, Wittekind & Mueller, 1993)
3: (Muhandiram *et al.*, 1993)
4: (Bax *et al.*, 1990, Bax & Pochapsky, 1992, Fesik & Zuiderweg, 1990, Kay *et al.*, 1990, Kay *et al.*, 1993)
5: (Davis *et al.*, 1992, Kay *et al.*, 1992, Palmer *et al.*, 1991, Schleucher *et al.*, 1993)

Table 2-2. Processing Parameters for 3D NMR experiments

Experiment	CBCA(CO)NH	HNCACB	^{13}C-NOESY	HCCH-TOCSY	^{15}N-NOESY
F3 solvent filter	Length = 64	Length = 64	Length = 64	None	Length = 32
Window	Shifted sine[1]	Shifted sine[1]	Shifted sine[1]	Shifted sine[1]	Shifted sine[1]
Zero-fill	Auto	Auto	Auto	Auto	Auto
Phase	86.0, 0.0	-68.0, 36.4	114.2, -41.0	24.3, 64.2	123.0, 0.0
F1 linear-prediction[3]	10	10	First point and 15	15	15
Window	Hamming[2]	Hamming[2]	Hamming[2]	Hamming[2]	Hamming[2]
Zero-fill	Auto	Auto	Auto	512 points	256 points
Phase	0.0, 0.0	-2.5, 4.0	86.1, 15.0	-24.5, 48.6	-86.5, 0.0
F2 linear prediction[3]	10	20	20	15	20
Window	Hamming[2]	Hamming[2]	Shifted sine[1]	Hamming[2]	Shifted sine[1]
Zero-fill	Auto	Auto	Auto	256	Auto
Phase	-82.2, 180.0	0.0, 0.0	-26.2, 57.0	8.5, 33.6	-94.2, 216.0

1: The first point was scaled by 0.5, and a sine window shifted by 0.33 (60° shifted sine-bell) was applied.
2: A user macro (Appendix B) that implements a hamming function was used (Cavanagh *et al.*, 1996).
3: Forward-backward linear prediction was used with the indicated order.

NMR titrations

Binding of the peptides to the ^{15}N-labeled DegS PDZ domain was tested by monitoring the ^{15}N-HSQC spectrum of the DegS PDZ domain during titration experiments. For the solYYF peptide titration, the DegS PDZ domain was at a concentration of 470 μM in a volume of 540 μL. An amount of solYYF equimolar to DegS was split into three tubes and lyophilized. An HSQC spectrum was acquired on the DegS PDZ domain solution. Three additional HSQC spectra were acquired after addition of each of the three equivalents of lyophilized solYYF peptide.

For the sol-ssrA peptide, 660 μL of the ^{13}C/^{15}N double-labeled DegS PDZ domain at 2.8 mM was used. 10.1 mg of the sol-ssrA peptide was dried and resuspended in 100 μL of the DegS PDZ domain solution to be used as a titrant. 20 μL at a time were added back, with an HSQC spectrum acquired after each addition for a total of six spectra at a peptide:protein ratio of 0, 0.29, 0.57, 0.84, 1.10, and 1.35. Peaks that have a different resonance in the bound form may either show a decrease in intensity of the old resonance and an increase in the new resonance (slow exchange) or the observed resonance may be a weighted average of the bound and unbound state (fast exchange). In intermediate exchange, the resonance observed for each peak also becomes broader near 50% bindng. For fast exchange, peaks were identified by plotting their position with respect to the amount of substrate added. For slow exchange, some of the new peaks were assigned based on the identity of the nearest peak that disappeared, whereas other new peaks were left unassigned.

Binding Experiments

Isothermal-titration calorimetry (ITC) was performed using a MicroCal instrument and data were analyzed using Origin software with scripts provided by MicroCal. Experiments were

conducted in overflow mode at 25 °C. Lyophilized samples of the DegS PDZ domain and the C-terminal of OmpC, solYQM, solYQF, or solYYF peptides were resuspended in 60 mM NaCl with 50 mM Tris (pH 7.5) or 50 mM $NaPO_4$ (pH 7.5). The protein and peptide solutions were adjusted to within 0.01 pH units of each other in range of 7.5-7.9. ITC experiments were initiated with a 1.4 mL solution of DegS PDZ domain in the temperature controlled cell, a 0.3 mL solution of peptide in the titrator/mixing syringe, and stirring at 310 rpm. 30-50 injections of 2-10 μL peptide were dispensed into the cell at a rate of 2 μL sec^{-1}. A 300 sec equilibration time was used between injections. The heats of peptide dilution were determined by measurements of peptide titrated into buffer alone and/or by averaging the peaks present after saturation. Heats of peptide dilution were subtracted from each injection prior to data analysis. The first point of each titration was also removed because of dilution during the initial equilibration. For most samples at least two ITC experiments were performed.

Binding may involve charged atoms, like the oxygens of the carboxylate. To determine whether there is any net change in protons being released or absorbed during complex formation, ITC was done using different buffers. Tris has a different heat of protonation (ΔH^{b}_{i}) at 25 °C than phosphate. The total heat evolved (ΔH_{obs}) in each buffer was compared and the number of

$$\Delta H_{obs} = \Delta H^{o} + N_{H^{+}} \bullet \Delta H^{buffer}_{ionization}$$

protons released by the buffer (N_{H+}) was calculated to be essentially zero.

Results

Alignments

Figure 2-1 shows an alignment of the protease PDZ-like domains with eukaryotic PDZ domains of known structures. The N-terminal portion of the sequence, labeled B0 in the alignment, forms the first beta-strand of the bottom beta-sheet in most of the solved structures. Sequences with homology to the B0 sequence are missing from most if not all protease PDZ-like domains, which instead contain a C-terminal sequence labeled B7.

Expression and Properties

The PDZ-like portion of DegS was expressed and purified to homogeneity. Circular-dichroism and thermal-denaturation experiments showed that this region forms an independently folded domain. Minima near 208 and 218 nm in the CD spectrum are suggestive of a mostly beta-sheet protein with some alpha-helix and random coil (Figure 2-2A). The thermal melt showed a single, highly cooperative transition ($T_m \approx 68\ ^{o}C$) (Figure 2-2B) that was irreversible because of aggregation.

NMR assignments and structural calculations

Resonances corresponding to $^{1}HN_i$, $^{15}NH_i$, $^{13}CA_{i-1}$, and $^{13}CB_{i-1}$ atoms were first identified using a CBCA(CO)NH experiment (Grzesiek *et al.*, 1992, Muhandiram & Kay, 1994), allowing subsequent assignment of $^{13}CA_i$ and $^{13}CB_i$ atoms using a HNCACB experiment (Kay *et al.*, 1992, Muhandiram & Kay, 1994, Palmer *et al.*, 1991, Wittekind & Mueller, 1993). ^{1}HNs and ^{15}NHs were ordered sequentially in temporary clusters based on overlapping ^{13}CA and ^{13}CB resonances. The HCCH-TOCSY experiment (Bax *et al.*, 1990, Bax & Pochapsky, 1992, Fesik & Zuiderweg, 1990, Kay *et al.*, 1990, Kay *et al.*, 1993) also used the ^{13}CA and ^{13}CB resonances to

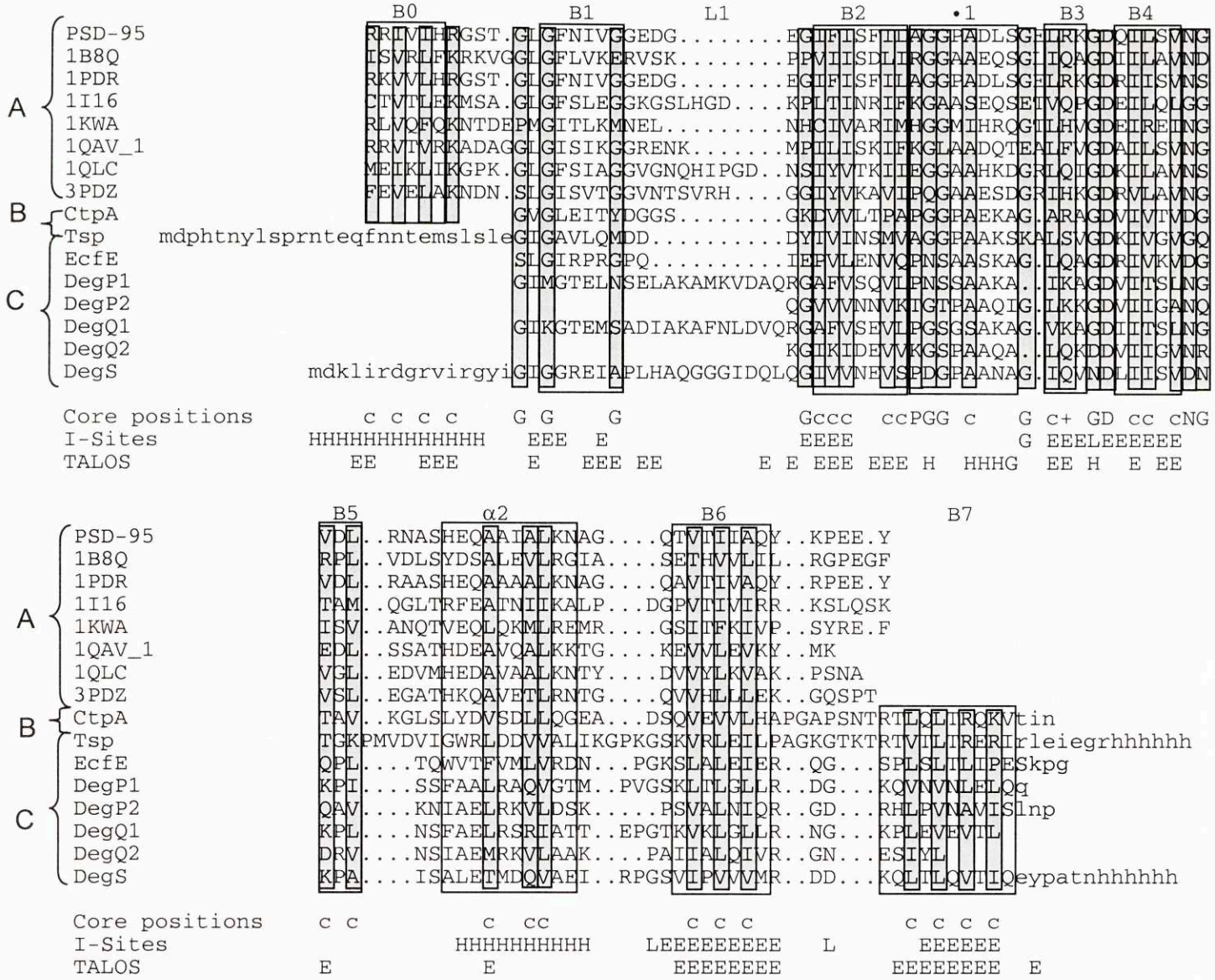

Figure 2-1. Alignment of PDZ domains of known structure with those from bacterial and plant proteases. The bacterial PDZ domains (C) are aligned with the sequences from the structures of both the known plant, CtpA, PDZ domain(B) and the known mammalian PDZ domains (A). The constructs used in this study extend beyond the PDZ domain and the extra sequence is shown in lower case letters. The bacterial PDZ domains were aligned using the hydrophobic core positions "c" and highly conserved glycine "G", proline "P", aspartate "D", asparagine "N", and charged "+" residues. The secondary structure of the DegS PDZ domain is predicted by the programs TALOS and I-SITES.

"H" = Helix = –180 < phi < 0 & -100 < psi < -10 "G" = turn = 0 < phi < 180 & -90 < psi < 90
"E" = Ext'd = -200 < phi < 20 & 40 < psi < 260 "L" = turn = 20 < phi < 160 & 90 < psi < -90

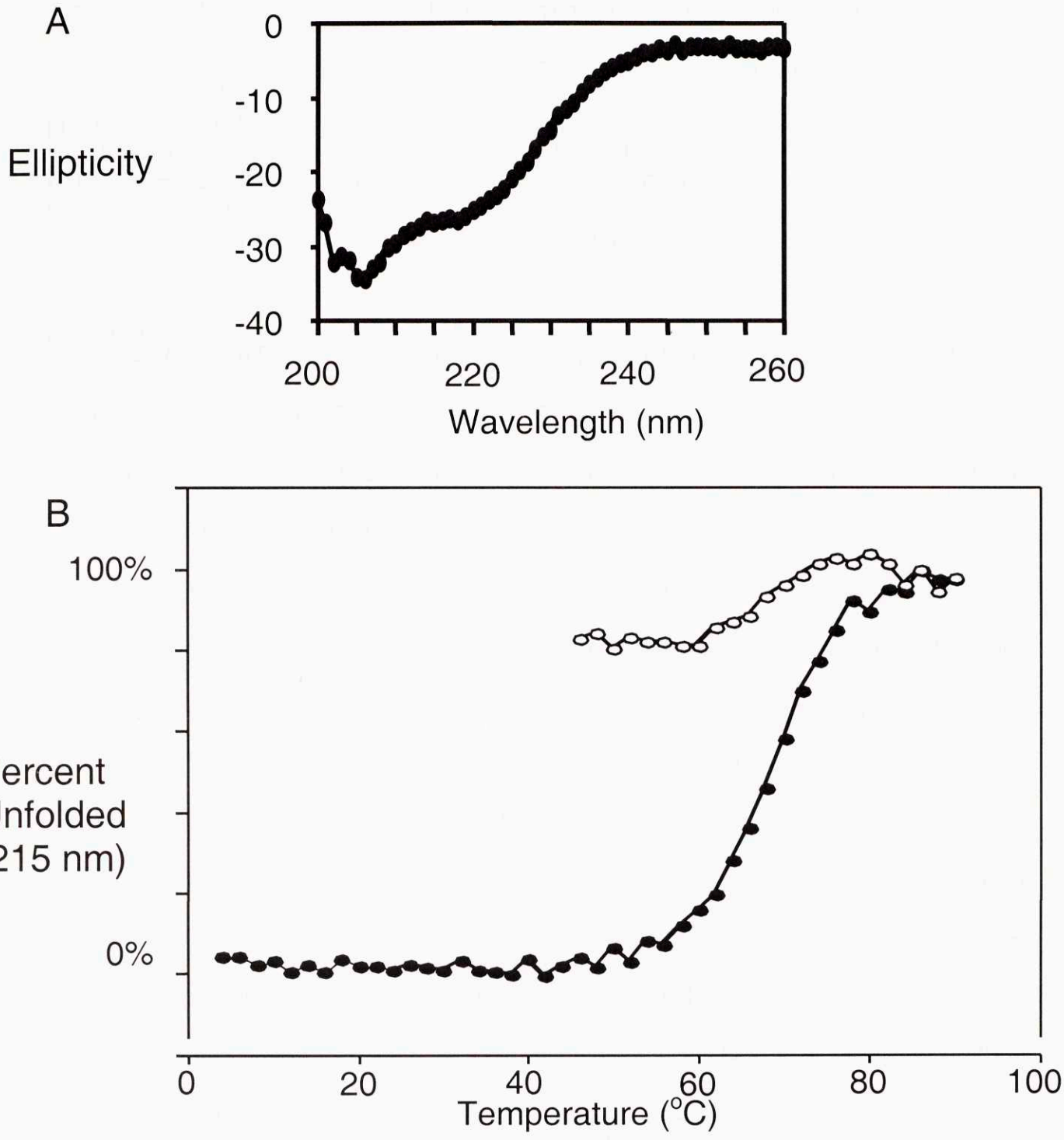

Figure 2-2. Circular Dichroism of DegS PDZ domain. (A) The CD spectra of purified DegS PDZ domain (2 μM) in 0.3 mM $NaPO_4$ (pH 5.0), 2 mM NaCl was measured at 25 °C. (B) Thermal denaturation of the DegS PDZ domain was followed by CD at 215 nm under similar conditions. Filled circles are from the initial melt, while open circles are from the reverse melt.

link the side chain resonances into the temporary clusters. Clusters were accepted if they were consistent with the amino-acid sequence of DegS. From these NMR experiments, 73 of the main-chain amide ^{1}H and ^{15}N resonances were sequentially assigned out of 88 (111 residues – 7 prolines – 6 histidine tag), 84 of the $^{13}CA_i$ and $^{13}CB_i$ positions were sequentially assigned, and 72 of the side chains were mostly or completely assigned (Table 2-3). There was considerable overlap that precluded assignment of the missing atoms.

Initial structure calculations based on the assignments and 489 ^{1}H NOEs from the 3D-^{13}C-NOESY experiment (414 short-range and 75 long-range (Table2-4)) were consistent with a PDZ-like structure in that two of the seven beta-strands (B6 and B7) and the first of two alpha-helices were well ordered and in the expected orientations. The other five beta strands were in the expected location, but were not well ordered. However, one of the beta-strands (B7) was from the C-terminus, even though all known PDZ structures at the time had a strand from the N-terminus in that position. The same circular permutation was subsequently observed in the crystal structure of the PDZ-like domain from CtpA (Liao *et al.,* 2000). The RMSD among structures in this initial set of calculations was very high. The regions of greatest divergence included the N-terminal 16 residues, the C-terminal 13 residues, the loop between residues 268 and 281, and the region which normally forms the second alpha-helix in known PDZ domains. In NMR relaxation experiments (Figure 2-3) most of these regions had a large T2-relaxation and small heteronuclear NOEs and are therefore likely to be unstructured. The exception was the region expected to form the second alpha-helix, where few relaxation measurements were made due to the lack of assignments.

A range was assigned to each dihedral angle using two algorithms, TALOS (Cornilescu *et al.,* 1999) and I-SITES (Bystroff & Baker, 1997). Both methods had

Table 2-3. NMR assignments

Residue	(Atom	Chemical Shift)								
2	CA	54.595	HA	4.721	CB	41.568	HB2	2.772	HB1	2.633
3	N	122.071	HN	8.465	CA	56.541	HA	4.318	CB	33.383
	HB2	1.849	HB1	1.774	CG	28.705	HG2	1.734	CD	24.643
	HD2	1.502	HD1	1.461	CE	42.143	HE2	3.049		
4	N	123.980	HN	8.537	CA	55.284	HA	4.385	CB	42.446
	HB1	1.637								
5	N	123.666	HN	8.170	CA	60.553	HA	4.382	CB	38.896
	HB	1.821	CG1	27.455	HG12	1.474	HG11	1.144	CD1	12.142
	HD11	0.944	CG2	17.455	HG21	0.844				
6	N	126.248	HN	8.623	CA	55.835	HA	4.454	CB	31.799
	HB2	1.844								
8	CA	45.656								
9	N	120.771	HN	8.044	CA	55.674	HA	4.487	CB	31.530
	HB2	1.829	HG2	1.625						
10	N	122.589	HN	8.325	CA	62.906	HA	4.156	CB	32.537
	HB	2.024	CG2	21.205	HG21	0.984	CG1	21.205	HG11	0.920
11	N	128.118	CA	60.268	HA	4.271	CB	37.455	HB	1.876
	CG1	27.143	HG12	1.450	HG11	1.279	CD1	11.517	HD11	0.822
	CG2	18.080	HG21	0.922						
14	HB2	2.572	HD1	6.716	HE1	6.847				
17	CA	59.455	HA	5.066	CB	43.302	HB	1.730	CG1	26.205
	HG12	1.290	HG11	0.741	CD1	14.330	HD11	0.622	CG2	18.705
	HG21	0.744								
18	N	110.303	HN	8.696	CA	44.708	HA2	4.672	HA1	3.771
19	N	111.145	HN	8.667	CA	44.900	HA2	4.478	HA1	3.507
20	N	117.555	HN	8.267	CA	54.899	HA	4.159	CB	33.742
21	N	124.152	HN	9.223	CA	56.456	HA	4.765	CB	30.465
	HB2	2.147	CG	35.580	HG2	2.551	HG1	2.479		
22	N	124.251	HN	7.927	CA	61.044	HA	3.955	CB	39.526
	HB	1.983	CG1	27.455	HG12	1.014	CD1	14.330	HD11	0.854
	CG2	18.393	HG21	0.957						
23	N	128.304	HN	8.272	CA	50.189	HA	4.772	CB	19.248
	HB1	1.410								
24	CA	63.106	HA	4.490	CB	32.196	HB2	2.288	HB1	1.872
	CG	27.143	HG2	2.049	CD	50.268	HD2	3.810	HD1	3.702
25	N	122.104	HN	8.376	CA	55.479	CB	42.774		
26	N	119.323	HN	8.470	CA	55.340	HA	4.730	CB	29.371
	HB2	3.284	HB1	3.202						
27	N	125.477	HN	8.440	CA	52.735	CB	19.466		

Table 2-3. NMR assignments (continued)

Residue	(Atom	Chemical Shift)								
28	N	118.852	HN	8.542	CA	56.330	HA	4.384	CB	29.447
	HB2	2.186	HB1	2.057	CG	33.705	HG2	2.425	NE2	111.672
	HE21	7.575	HE22	6.904						
29	N	110.372	HN	8.570	CA	45.528				
30	N	108.778	HN	8.333	CA	45.285	HA2	4.011		
31	CA	45.304								
32	N	119.431	HN	8.041	CA	61.370	HA	4.225	CB	39.015
	HB	1.898	CG1	27.143	HG12	1.449	HG11	1.185	CD1	13.080
	HD11	0.874	CG2	17.455	HG21	0.921				
33	N	123.720	HN	8.463	CA	54.611	HA	4.620	CB	41.210
	HB2	2.696	HB1	2.624						
34	N	120.157	HN	8.191	CA	55.474	HA	4.365	CB	30.068
	HB2	2.085	HB1	1.963	HG2	2.314				
35	N	124.418	HN	8.239	CA	55.221	CB	42.971		
36	N	119.960	HN	8.036	CA	55.200	HA	4.600	CB	31.568
	HB2	2.129	HB1	1.999	CG	33.915	HG2	2.409	HG1	2.361
	NE2	112.030	HE21	7.668	HE22	6.927				
37	N	108.759	HN	7.952	CA	45.074	HA2	4.478	HA1	4.316
38	N	117.880	HN	8.468	CA	59.479	HA	4.378	CB	41.365
	HB	1.503	CG1	26.830	HG12	1.414	HG11	0.732	CD1	13.080
	HD11	0.633	CG2	16.205	HG21	0.542				
39	N	129.180	HN	8.952	CA	62.008	HA	4.476	CB	32.535
	HB	2.014	CG2	22.067	HG21	0.951	CG1	20.580	HG11	0.901
40	N	127.813	HN	8.828	CA	63.943	HA	3.876	CB	31.736
	HB	2.282	CG2	23.393	HG21	0.768	CG1	22.455	HG11	0.919
41	N	128.190	HN	9.266	CA	53.482	HA	5.053	CB	40.474
	HB2	2.806	ND2	109.271	HD21	7.616	HD22	6.624		
42	N	120.152	HN	8.493	CA	56.043	HA	4.464	CB	33.098
	HB2	1.949	HB1	1.849	CG	35.810	HG2	2.158	HG1	2.045
43	N	123.447	HN	8.721	CA	60.778	HA	4.465	CB	34.245
	HB	1.876	CG2	21.755	HG21	0.781	CG1	21.442	HG11	0.692
44	N	124.128	HN	8.007	CA	57.360	HA	4.674	CB	62.500
	HB2	3.964	HB1	3.719						
45	CA	64.438	HA	4.271	CB	31.937	HB2	2.329	HB1	1.953
	CG	27.455	HG2	2.129	HG1	2.080	CD	51.206	HD2	4.104
	HD1	3.796								
46	N	117.162	HN	8.826	CA	54.962	HA	4.527	CB	40.510
	HB2	2.940	HB1	2.832						
47	N	105.768	HN	7.544	CA	45.243	HA2	4.622	HA1	3.830
49	CA	55.444	HA	3.896	CB	19.540	HB1	1.287		
50	N	123.502	HN	8.569	CA	55.123	HA	4.047	CB	18.317

Table 2-3. NMR assignments (continued)

Residue (Atom Chemical Shift)

	HB1	1.533								
51	N	116.882	HN	8.439	CA	55.636	HA	4.437	CB	38.243
	HB2	2.879	HB1	2.821	ND2	110.691	HD21	7.306	HD22	6.755
52	N	119.570	HN	7.663	CA	52.600	HA	4.528	CB	21.545
	HB1	1.571								
53	N	103.697	HN	7.650	CA	45.352	HA2	4.702	HA1	4.014
54	N	122.704	HN	7.850	CA	63.172	HA	3.633	CB	37.124
	HB	1.510	CG1	29.643	HG12	1.613	HG11	0.971	CD1	12.455
	HD11	0.706	CG2	16.830	HG21	0.906				
55	N	125.658	HN	8.805	CA	53.589	HA	4.674	CB	32.755
	HB2	2.352								
56	N	119.789	HN	8.354	CA	65.290	HA	3.359	CB	31.615
	HB	1.895	CG2	22.455	HG21	0.961	CG1	21.205	HG11	0.989
57	N	120.101	HN	9.420	CA	56.285	HA	4.233	CB	37.313
	HB2	3.430	HB1	3.078	ND2	113.782	HD21	7.716	HD22	6.967
58	N	122.700	HN	8.107	CA	56.363	HA	4.626	CB	41.152
	HB2	2.842	HB1	2.279						
59	N	121.588	HN	8.082	CA	52.936	HA	5.055	CB	44.282
	HB2	1.996	HB1	1.438	CG	27.455	HG	1.208	CD1	25.893
	HD11	0.771	CD2	24.330	HD21	0.671				
60	N	127.107	HN	9.741	CA	62.105	HA	4.239	CB	38.426
	HB	1.796	CG1	27.455	HG12	1.565	HG11	0.861	CD1	13.705
	HD11	0.589	CG2	17.768	HG21	0.676				
61	N	122.931	HN	9.000	CCA	61.30	HA	4.707	CB	39.219
	HB	1.929	CG1	26.518	HG12	1.308	HG11	1.118	CD1	12.767
	HD11	0.791	CG2	17.768	HG21	0.948				
62	N	115.501	HN	7.768	CA	57.509	HA	5.005	CB	64.698
	HB2	3.651	HB1	3.503						
63	N	122.287	HN	8.598	CA	60.893	HA	4.253	CB	34.330
	HB	1.873	CG2	23.080	HG21	0.896	CG1	21.830	HG11	0.719
64	N	129.960	HN	9.583	CA	55.481	HA	4.394	CB	40.475
	HB2	3.185	HB1	2.106						
65	N	107.231	HN	9.105	CA	55.361	HA	3.993	CB	37.902
	HB2	3.076	HB1	3.026	ND2	115.909	HD21	7.791	HD22	7.043
66	N	120.222	HN	7.718	CA	53.159	HA	4.927	CB	33.467
	HB2	1.867	HG2	1.517	HG1	1.439	CD	28.705	HD2	1.694
	CE	42.143	HE2	3.082						
67	CA	62.768	HA	4.393	CB	32.143	HB2	2.360	HB1	1.848
	CG	27.455	HG2	2.148	HG1	2.047	CD	50.893	HD2	3.938
	HD1	3.700								
70	HA	4.792	CB	64.560	HB2	4.036	HB1	3.762		
73	CA	52.596	CB	19.384						

Table 2-3. NMR assignments (continued)

Residue		(Atom	Chemical Shift)							
74	N	113.617	HN	8.233	CA	60.893	HA	4.293	CB	69.331
	HB	4.189	CG2	21.205	HG21	1.166				
84	N	113.954	HN	9.144	CA	44.701	HA2	4.495	HA1	3.796
85	N	117.159	HN	7.725	CA	59.549	HHA	4.43	CB	63.565
	HB2	3.839	HB1	3.682						
87	CA	56.206	HA	5.061	CB	41.205	HB	1.701	CG1	25.268
	HG12	1.607	HG11	1.026	CD1	15.580	HD11	0.695	CG2	19.018
	HG21	0.859								
88	CA	61.550	HA	5.151	CB	32.146	HB2	2.162	HB1	1.789
	CG	27.455	HG2	2.225	HG1	2.162	CD	49.956	HD2	4.032
	HD1	3.875								
89	N	126.492	HN	9.372	CA	61.780	CB	35.176	HG21	0.904
90	N	127.813	HN	8.396	CA	61.458	HA	5.054	CB	32.636
	HB	1.984	CG2	21.205	HG21	0.954	CG1	21.205	HG11	0.836
91	N	121.931	HN	9.237	CA	57.890	HA	5.583	CB	35.848
	HB	1.860	CG2	21.517	HG21	0.847	CG1	19.643	HG11	0.809
92	N	119.809	HN	9.397	CA	53.370	HA	5.429	CB	33.267
	HB2	2.139	HB1	1.947	CG	31.518	HG2	2.553	HG1	2.460
93	N	128.627	HN	9.168	CA	55.336	HA	4.679	CB	33.183
	HB2	2.124	CG	25.526	HG2	1.821	HG1	1.573	CD	43.711
	NE	84.937	HE	9.434						
94	N	129.183	HN	9.400	CA	56.561	HA	4.239	CB	39.631
	HB2	2.965	HB1	2.754						
95	N	113.285	HN	8.708	CA	55.449	HA	4.365	CB	40.161
	HB2	2.928	HB1	2.855						
96	N	120.305	HN	7.953	CA	55.075	HA	4.627	CB	35.041
	HB2	1.847	HG2	1.495	HD2	1.298	HE2	2.622	HE1	2.262
97	N	122.515	HN	8.495	CA	55.424	HA	5.274	CB	30.875
	HB2	1.987	HB1	1.907	CG	35.366	HG2	2.213	HG1	2.041
	NE2	110.899	HE21	7.144	HE22	6.825				
98	N	128.567	HN	9.536	CA	54.487	HA	4.787	CB	45.055
	HB2	1.565	HB1	1.467	CG	27.067	HG	1.552	CD1	24.330
	HD11	0.874	CD2	25.268	HD21	0.822				
99	N	118.494	HN	8.602	CA	62.627	HA	4.873	CB	69.175
	HB	4.040	CG2	21.205	HG21	1.110				
100	N	129.598	HN	9.392	CA	53.217	CB	43.836		
101	N	119.792	HN	8.549	CA	54.533	HA	4.835	CB	28.726
	CG	33.047	NE2	110.058	HE21	7.304	HE22	6.577		
102	N	125.870	HN	9.212	CA	60.974	CB	34.251		
103	N	124.637	HN	8.901	CA	61.999	HA	4.871	CB	68.670
	HB	4.082	CG2	20.268	HG21	1.059				
104	N	127.883	HN	8.752	CA	59.263	HA	3.842	CB	35.808

Table 2-3. NMR assignments (continued)

Residue (Atom Chemical Shift)

	HB	2.116	CG1	26.518	HG12	1.774	HG11	1.310	CD1	8.392
	HD11	0.578	CG2	17.768	HG21	1.001				
108	CA	63.003	CB	31.860						
109	N	124.979	HN	8.266	CA	52.628	CB	19.409	HB1	1.429
110	N	112.389	HN	8.034	CA	61.862	HA	4.369	CB	69.935
	HB	4.288	CG2	21.518	HG21	1.236				
111	N	120.530	HN	8.380	CA	53.060	HA	4.677	CB	38.928
	HB2	2.782	HB1	2.753	ND2	112.641	HD21	7.617	HD22	6.939

Amino acid 1 is the first amino acid of the DegS PDZ domain from pLC263.

Table 2-4. NOEs used in the structure calculations.

Residue 1	Atom 1	Residue 2	Atom 2	Min	Range
NOE's (j > i+2)					
17	HA	44	HN	1.8	3.2
21	HG2	40	HA	2.0	3.0
39	HB	59	HG	2.0	3.0
40	HA	21	HA	1.8	3.2
40	HA	21	HB2	2.0	3.0
40	HA	56	HG11	1.8	3.2
43	HA	19	HN	2.0	3.0
43	HA	19	HA1	2.0	3.0
43	HA	54	HG21	1.8	1.6
44	HB1	19	HA1	1.8	3.2
44	HB2	19	HA1	2.0	3.0
44	HB1	19	HA2	2.0	3.0
45	HA	50	HB1	1.8	1.6
45	HB	50	HB1	2.0	3.0
45	HA	51	HB1	1.8	1.0
49	HA	52	HB1	1.8	1.6
49	HA	53	HN	1.8	3.2
49	HA	53	HA2	1.8	3.2
49	HA	53	HA1	1.8	1.0
49	HA	54	HG21	1.8	3.2
50	HA	43	HG21	1.8	1.0
50	HB1	43	HG21	2.0	3.0
50	HB1	45	HA	2.0	3.0
50	HB1	46	HN	2.0	3.0
50	HA	53	HN	1.8	3.2
50	HA	54	HG21	1.8	3.2
54	HB	49	HB1	1.8	3.2
54	HA	91	HG11	2.0	3.0
56	HA	40	HG21	1.8	1.0
56	HA	40	HG11	2.0	3.0
56	HA	91	HG21	1.8	3.2
57	HA	39	HG11	1.8	1.6
57	HA	54	HG11	1.8	3.2
59	HA	39	HA	2.0	3.0
59	HA	39	HG11	2.0	3.0
59	HA	39	HG21	1.8	1.6

Residue 1	Atom 1	Residue 2	Atom 2	Min	Range
59	HA	40	HN	1.8	3.2
61	HB	38	HG11	1.8	1.6
62	HA	65	HN	1.8	3.2
62	HA	67	HA	1.8	3.2
62	HB1	67	HA	1.8	3.2
62	HB2	67	HA	1.8	3.2
62	HB1	92	HB1	1.8	3.2
62	HB2	92	HB1	1.8	3.2
62	HB2	92	HB2	1.8	3.2
63	HA	90	HN	1.8	1.6
63	HA	91	HG11	1.8	1.6
65	HA	62	HA	1.8	3.2
85	HB1	103	HG21	1.8	3.2
88	HA	101	HA	1.8	3.2
88	HA	102	HN	2.0	3.0
90	HA	63	HG11	1.8	3.2
90	HA	99	HA	1.8	1.6
90	HA	100	HN	1.8	3.2
91	HA	63	HA	1.8	1.6
91	HA	63	HG11	1.8	3.2
91	HB	63	HG11	1.8	1.6
91	HA	63	HG21	2.0	3.0
91	HA	98	HN	1.8	3.2
92	HA	97	HA	1.8	1.6
92	HA	97	HB2	1.8	1.6
92	HG1	97	HG11	1.8	1.0
92	HA	98	HN	1.8	3.2
97	HA	90	HG11	1.8	1.6
97	HA	92	HA	1.8	1.6
97	HA	93	HN	1.8	3.2
98	HA	55	HN	1.8	3.2
99	HA	90	HN	1.8	3.2
99	HA	90	HA	1.8	1.6
99	HA	90	HG21	1.8	1.6
99	HA	91	HN	1.8	1.6
99	HB	101	HN	1.8	3.2
101	HA	88	HA	1.8	1.6
104	HA	87	HG21	1.8	3.2
104	HA	87	HD11	1.8	3.2

reasonable agreement over most of the protein (Figure 2-1). TALOS uses NMR derived chemical shifts and amino-acid neighbor information to predict dihedral angles. I-SITES is entirely sequence dependent and uses a window of amino acids in the sequence to identify what conformations those amino acids are likely to form. Both methods predict dihedrals similar to PDZ domains.

To improve the overall model, homology-based hydrogen bonds (Table 2-5) and conserved dihedral angles were included in the structural calculations, in addition to the NOE distance constraints. The NMR/alignment model (Figure 2-4) generated from the combined data was used for all structural descriptions. In 67 topologically acceptable structures, the backbone RMSD was 1.7 Å and the heavy atom RMSD was 2.6 Å for the ordered regions (261-268, 280-313, and 331-346). The final energies ranged from 12,000 to 20,000 kcal/mol. The seven beta-strands and the first alpha-helix were moderately well-ordered. The region from residue 314 to 323 formed the second alpha-helix, but it lacked restraints to other residues in the structure and was not packed consistently in all structures.

Peptide Selection

Because PDZ domains generally bind to C-terminal peptide sequences, the ability of the DegS domain to retain members of a peptide library with randomized positions at the C-terminus was determined. In experiments in which a KNxxxxxx library was bound to columns made from resin conjugated to the DegS PDZ domain or conjugated to glycine, only the last three randomized positions showed differences between the two columns (Figure 2-5). This selection was performed twice with similar results. At the penultimate and antepenultimate positions (Figure 2-5D and E), there was a strong preference for tyrosine. At the last position (Figure 2-5F), phenylalanine was preferred and methionine was next best.

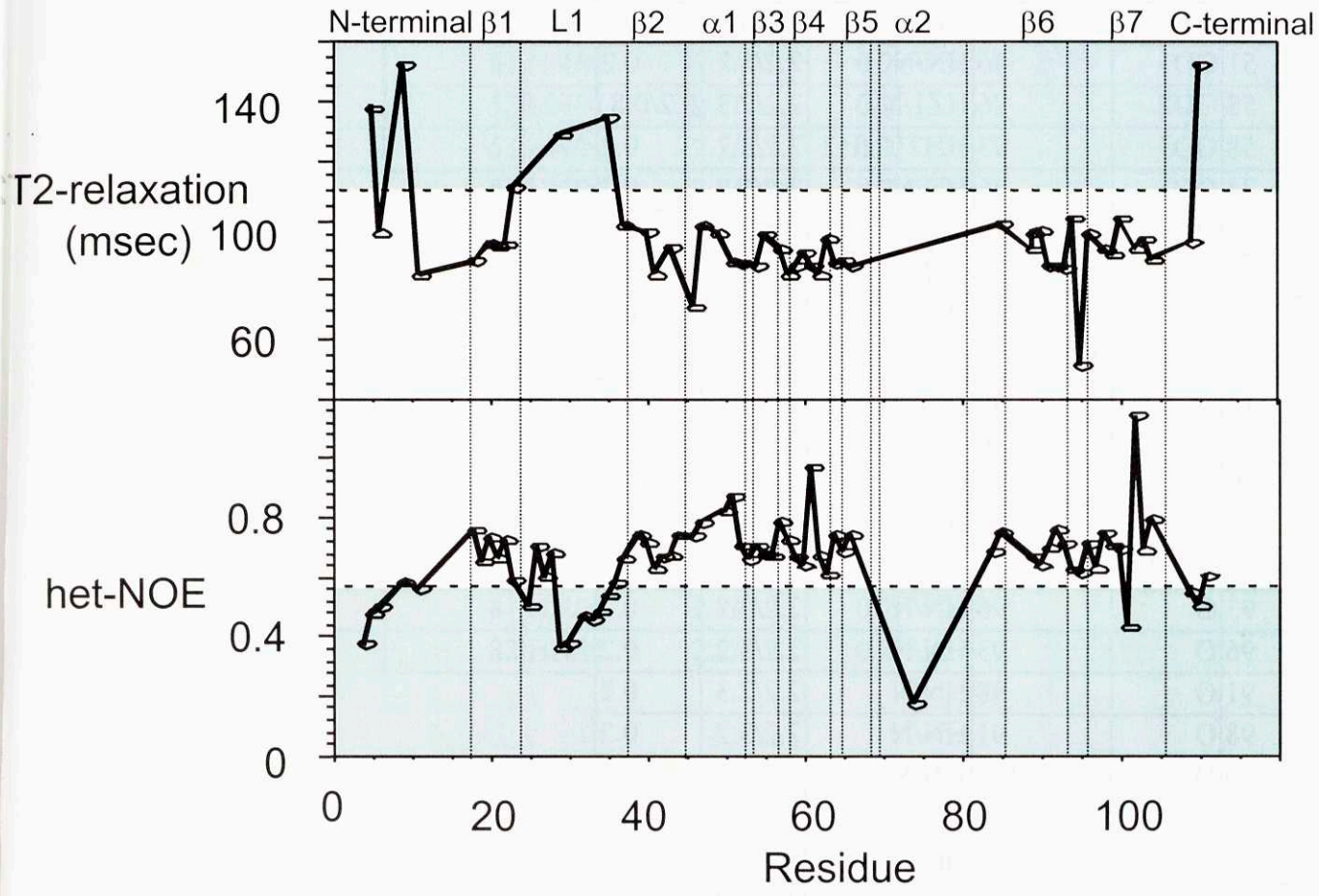

Figure 2-3. NMR relaxation experiments. Amides in ^{15}N-labeled DegS PDZ domain that show large T2-relaxations and small heteronuclear-NOEs are likely to be disordered. By this criterion, the N and C-terminal regions and the L1 loop are disordered. There are not enough assigned resonances in the α2 helix to determine whether it is structured.

Table 2-5. Homology based hydrogen bonds.

Residue 1	Atom 1	Residue 2	Atom 2	Min	Range
Conserved Hydrogen Bonds					
21	OE2	82	HH1/NH1	2.2/3.2	0.2
51	OD1	46	HN/N	2.2/3.2	0.2
58	OD2	96	HZ1/NZ	2.2/2.3	0.2/0.8
58	OD1	93	HH21/NH2	2.2/3.2	0.2
74	OG1	76	HN/N	2.2/3.2	0.2
74	OG1	77	HN/N	2.2/3.2	0.2
73	OE1	78	HN/N	2.2/3.2	0.2
80	OE1	66	HZ1/NZ	2.2/3.2	0.2
Beta Strand 1-1 to 1-2					
18	O	44	HN/N	2.2/3.2	0.2
42	O	20	HN/N	2.2/3.2	0.2
39	O	22	HN/N	2.2/2.3	0.2
22	O	39	HN/N	2.2/3.2	0.2
Beta Strand 2-1 to 2-2					
93	O	96	HN/N	2.2/3.2	0.2
96	O	93	HN/N	2.2/3.2	0.2
91	O	98	HN/N	2.2/2.3	0.2
98	O	91	HN/N	2.2/3.2	0.2
89	O	100	HN/N	2.2/3.2	0.2
100	O	89	HN/N	2.2/3.2	0.2
87	O	102	HN/N	2.2/3.2	0.2
102	O	87	HN/N	2.2/3.2	0.2
85	O	104	HN/N	2.2/3.2	0.2
104	O	85	HN/N	2.2/2.3	0.2
Beta Strand 2-2 to 2-3					
90	O	64	HN/N	2.2/3.2	0.2
62	O	92	HN/N	2.2/3.2	0.2
Beta Strand 2-3 to 2-4					
63	O	66	HN/N	2.2/3.2	0.2
66	O	63	HN/N	2.2/3.2	0.2
61	O	68	HN/N	2.2/3.2	0.2

Table 2-5. Homology based hydrogen bonds (continued).

Residue 1	Atom 1	Residue 2	Atom 2	Min	Range
Alpha Helix 1					
44	O	50	HN/N	2.2/3.2	0.2
47	O	51	HN/N	2.2/3.2	0.2
48	O	52	HN/N	2.2/2.3	0.2
49	O	53	HN/N	2.2/3.2	0.2
50	O	53	HN/N	2.2/3.2	0.2
Alpha Helix 2					
69	O	73	HN/N	2.2/3.2	0.2
70	O	74	HN/N	2.2/3.2	0.2
71	O	75	HN/N	2.2/2.3	0.2
72	O	76	HN/N	2.2/3.2	0.2
73	O	77	HN/N	2.2/3.2	0.2
74	O	78	HN/N	2.2/3.2	0.2
75	O	79	HN/N	2.2/3.2	0.2
76	O	80	HN/N	2.2/2.3	0.2
77	O	81	HN/N	2.2/3.2	0.2
78	O	82	HN/N	2.2/3.2	0.2

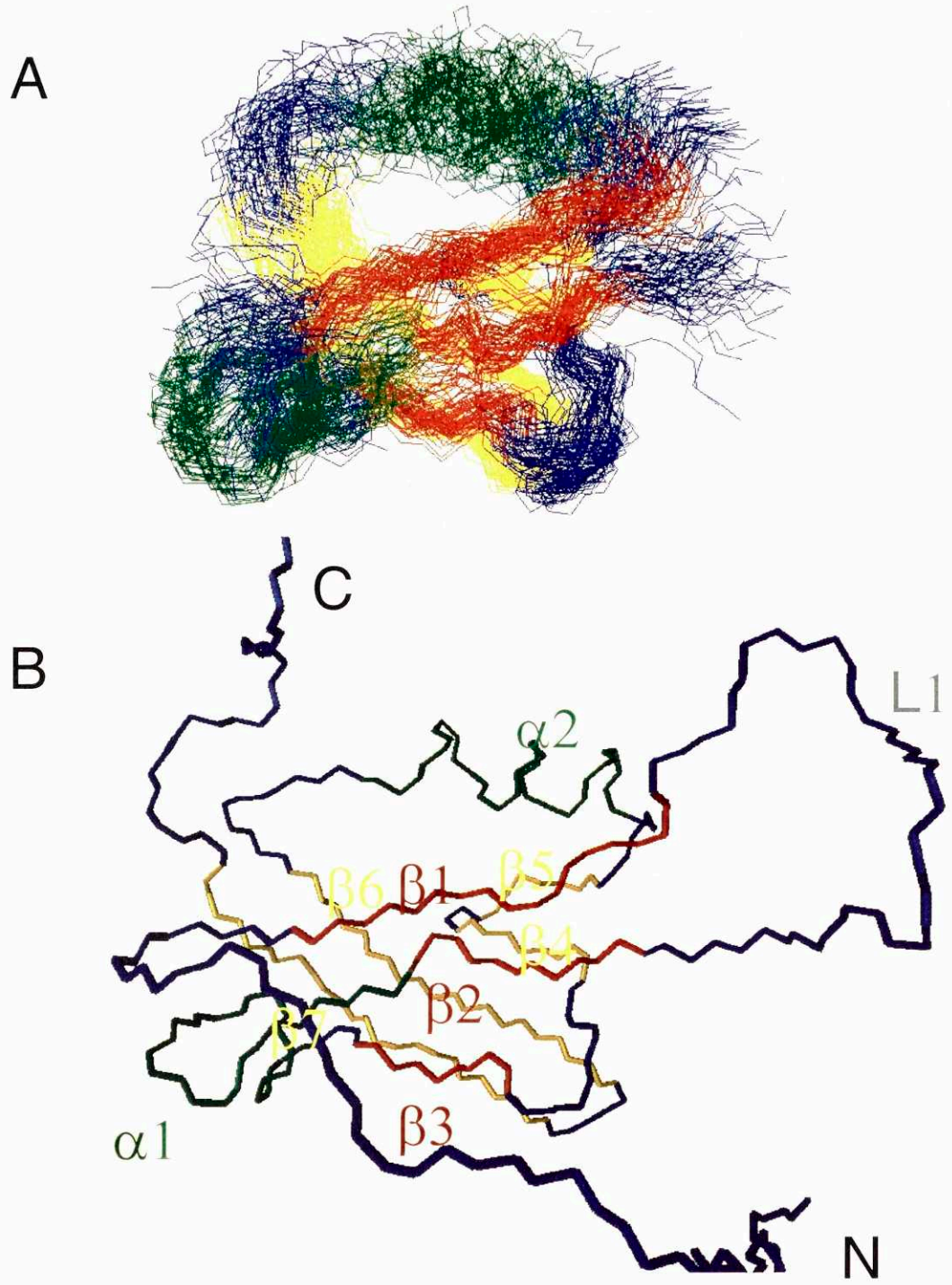

Figure 2-4. DegS PDZ model. The DegS PDZ model, an ensemble of structures, was generated by NMR distance restraints, dihedral restraints, and homology based hydrogen bonds. Portions of sixty-seven calculated structures are shown (A); the N and C-termini and the loop L1 are not displayed. One representative structure has the individual secondary structure elements labeled (B). The left edge of the bottom beta-sheet (beta-strand B7) is formed by the C-terminus, which differs from the majority of the eukaryotic PDZ domains.

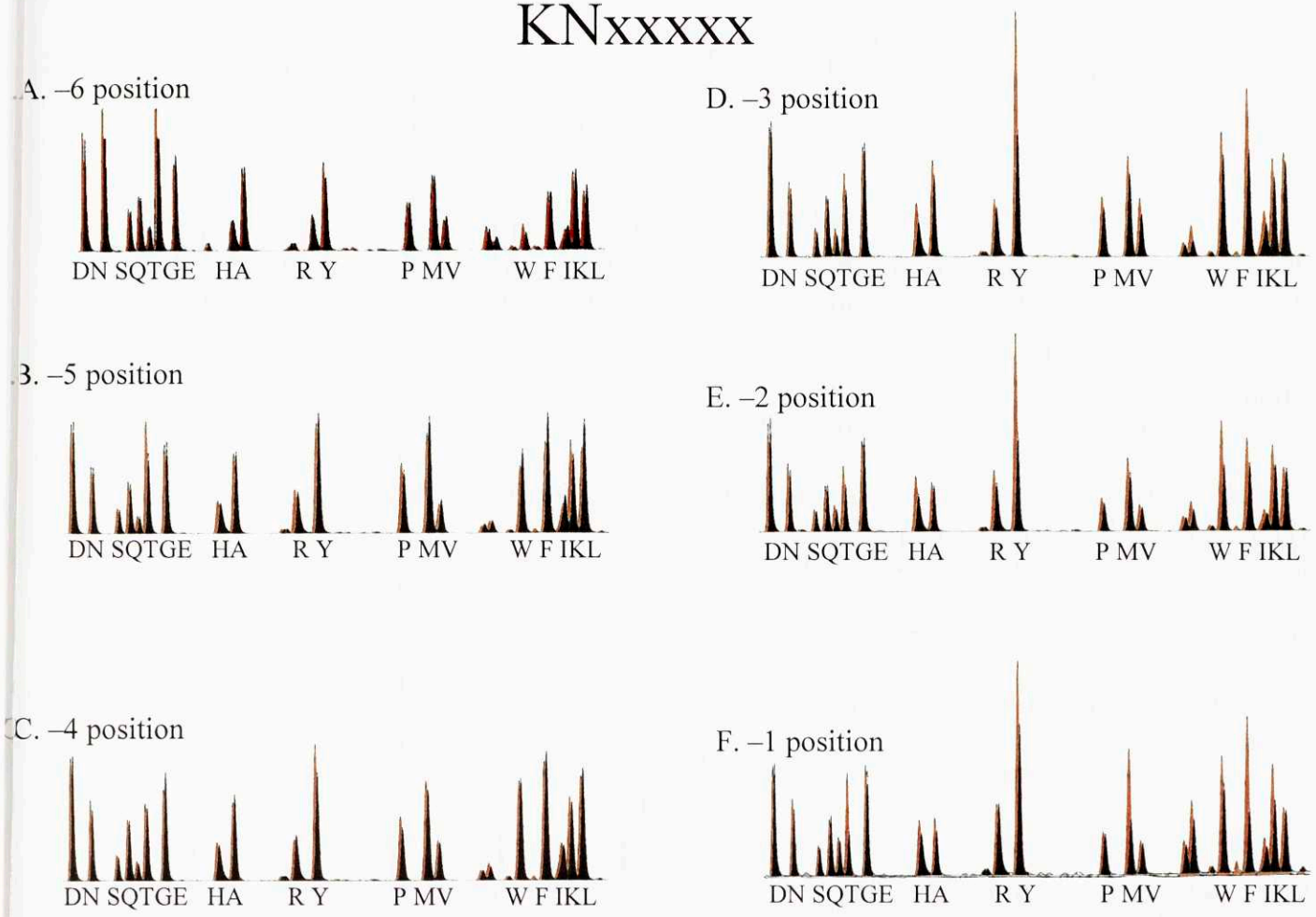

Figure 2-5. KNxxxxxx peptide library. Peptides from the KNxxxxxx library that bound the DegS PDZ domain or a control column were sequenced. Each panel A-F represents one of the randomized positions in the libray. In black is the HPLC chromatogram of the PTH-amino-acids from the N-terminal sequencing of the peptides retaned by the control column, whereas in red are the results from the column containing the DegS PDZ domain. Places where the red is larger than the black indicate a preference for that amino acid at that position.

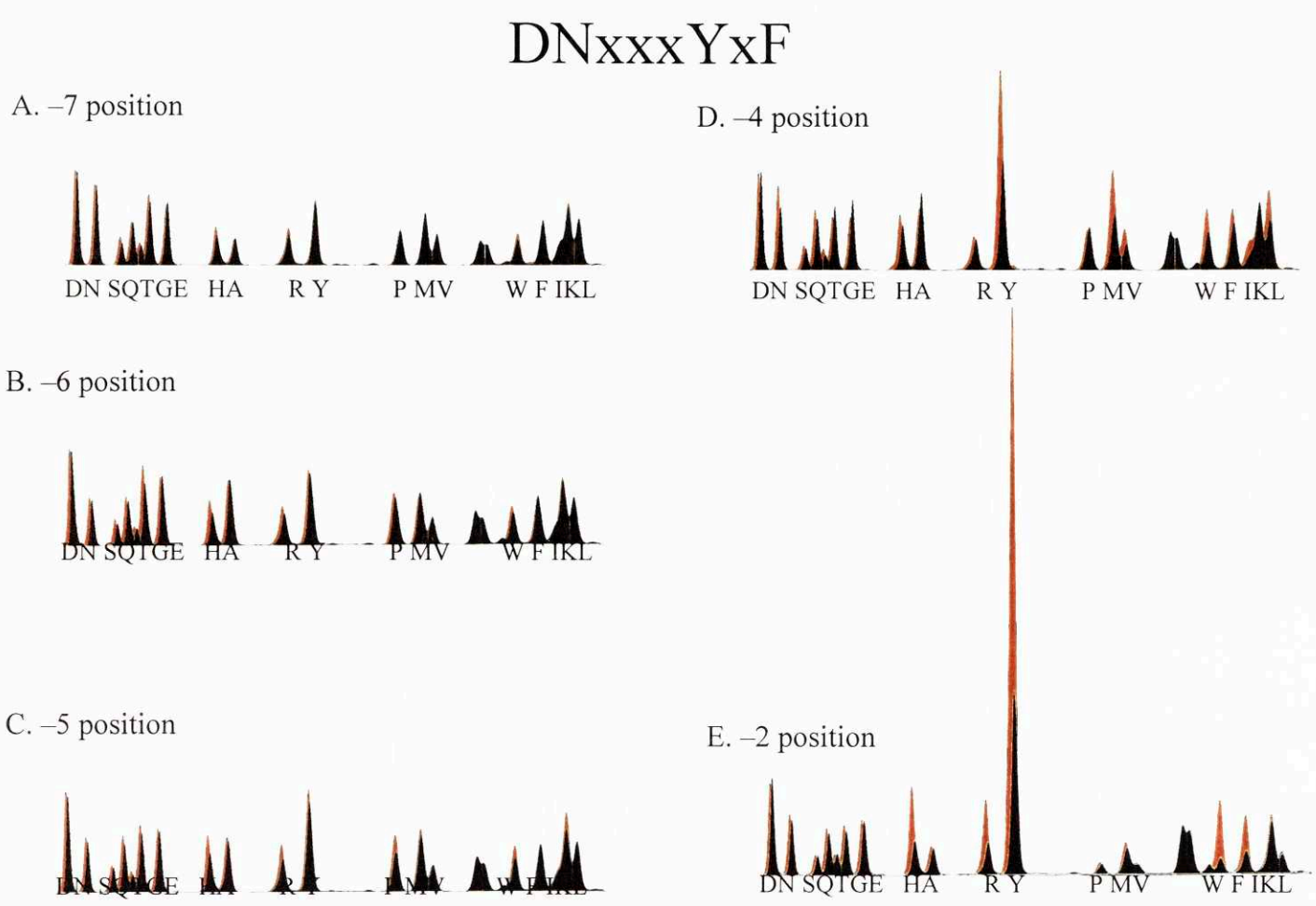

Figure 2-6. DNxxxxYxF peptide library. Peptides from the DNxxxxYxF library were treated in the same manner as Figure 2-5. Each panel A-E represents on of the randomized positions in the libray. The results from control column is shown in black, while those from the column containing the DegS PDZ domain are shown in red. Positions where the red is larger than the black indicate a preference for that amino acid at that position.

Proteins sequences encoded in the *E. coli* genome were searched for Tyr-Tyr-Phe-COOH or Tyr-Tyr-Met-COOH motifs. No exact matches were found. However, a number of proteins matched two of the three positions in these motifs. Interestingly, a set of outer-membrane proteins ending in a Tyr-x-Phe made up the largest group of these proteins (Table 2-7).

In similar studies, fixing several positions of the library as the preferred amino acid made the selection more sensitive to weaker preferences at positions that showed no preference in the original library (Aramburu *et al.*, 1999). Because the OMP proteins appeared to be good candidates for potential DegS interacting proteins, a second peptide library (NH_2-DNxxxxYxF-COOH) was used to test whether the penultimate position might prefer amino acids in addition to tyrosine. Compared to the first library, the length of the peptide was increased by one residue to nine, the base charge on the invariant region was changed from +1 to -1, and the C-terminal and antepenultimate positions were fixed. For the DNxxxxYxF library, no preferences were observed for the first three randomized positions and tyrosine was again strongly preferred at the penultimate position (Figure 2-6). In addition a weak preference for aromatics can be seen at the penultimate position including histidine, tryptophan, phenylalanine and tyrosine. A slight preference for tyrosine was seen at the –4 position, although the preference may be an artifact of fixing the –3 position to tyrosine. These results together with the first selection suggest that only the three C-terminal residues play a significant role in DegS peptide-binding specificity.

Peptide Binding Assayed by NMR

NMR was used to test for binding of a peptide ending with YYF (solYYF) to the isotopically-labeled PDZ-like domain of DegS (Figure 2-7). During the titrations, at least 25 of the resonances moved more than 0.05 ppm, but the majority of the DegS domain resonances remained unchanged during binding. The peaks that moved during the titration experienced a

drop in intensity at their original position and an increase in intensity at new positions, indicating that peptide binding was in slow exchange. After three additions of the peptide (approximately 1:1 ratio), the peaks that moved showed no intensity at their original position and no further changes in the spectrum were seen. The resonances affected by peptide binding included those within the top beta-sheet where beta-strand B1 (residues 261-265) and beta-strand B2 (residues 280-285,287) exhibited changes in chemical shift (Figure 2-8). Changes were also seen in individual residues in the first alpha-helix, and in the core. Several peaks that were not assigned also exhibited a change in chemical shift and are likely to be from the second alpha-helix. These results are consistent with a PDZ-like model of binding in which the peptide would be in an extended conformation hydrogen bonding with the top beta-sheet to form an additional beta-strand and to pack side chains into the core and against the second alpha-helix.

In similar experiments, we also tested the binding of the sol-ssrA peptide to the DegS domain (data not shown). The sol-ssrA peptide bound weakly with a dissociation constant in the millimolar range, but the resonances that changed were again mostly localized to the top beta-sheet. During the sol-ssrA peptide titrations, the resonances that moved exhibited a gradual change in chemical shifts that was proportional to the amount of bound DegS. This result, and the broadening seen when the DegS PDZ domain was approximately half bound, indicated that the sol-ssrA peptide and the DegS domain were in fast to intermediate exchange.

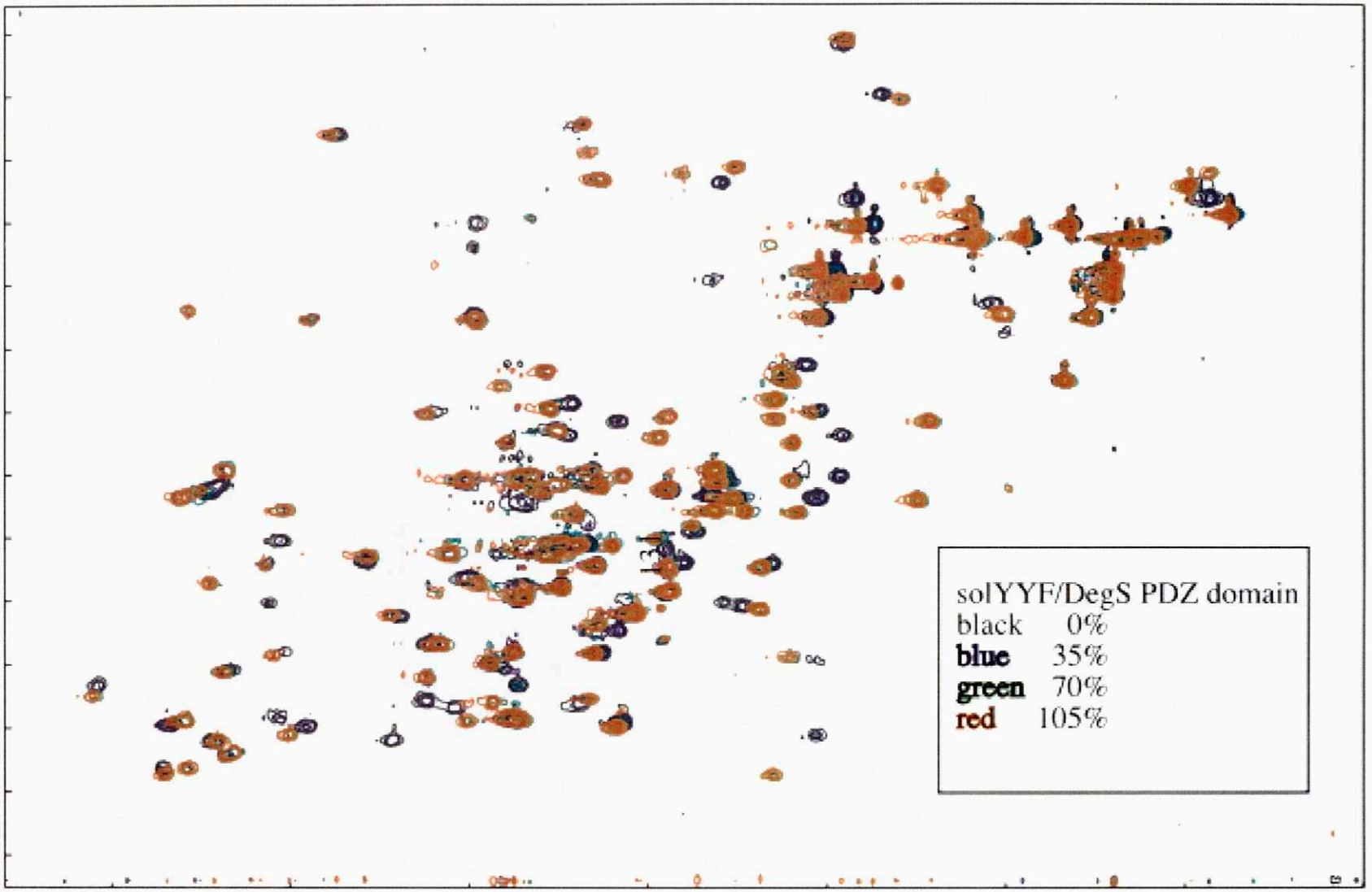

Figure 2-7. HSQC of DegS PDZ domain with solYYF peptide. A series of HSQC experiments were taken while adding solYYF peptide to a solution of DegS PDZ domain. The first HSQC spectrum black) had no solYYF peptide present. In subseuent HSQC spectra (blue, green, and red) solYYF peptide was added. While most peaks showed no change in position or shape, some of the peaks present in the first HSQC disappeared, while new peaks appeared in subsequent spectrum.

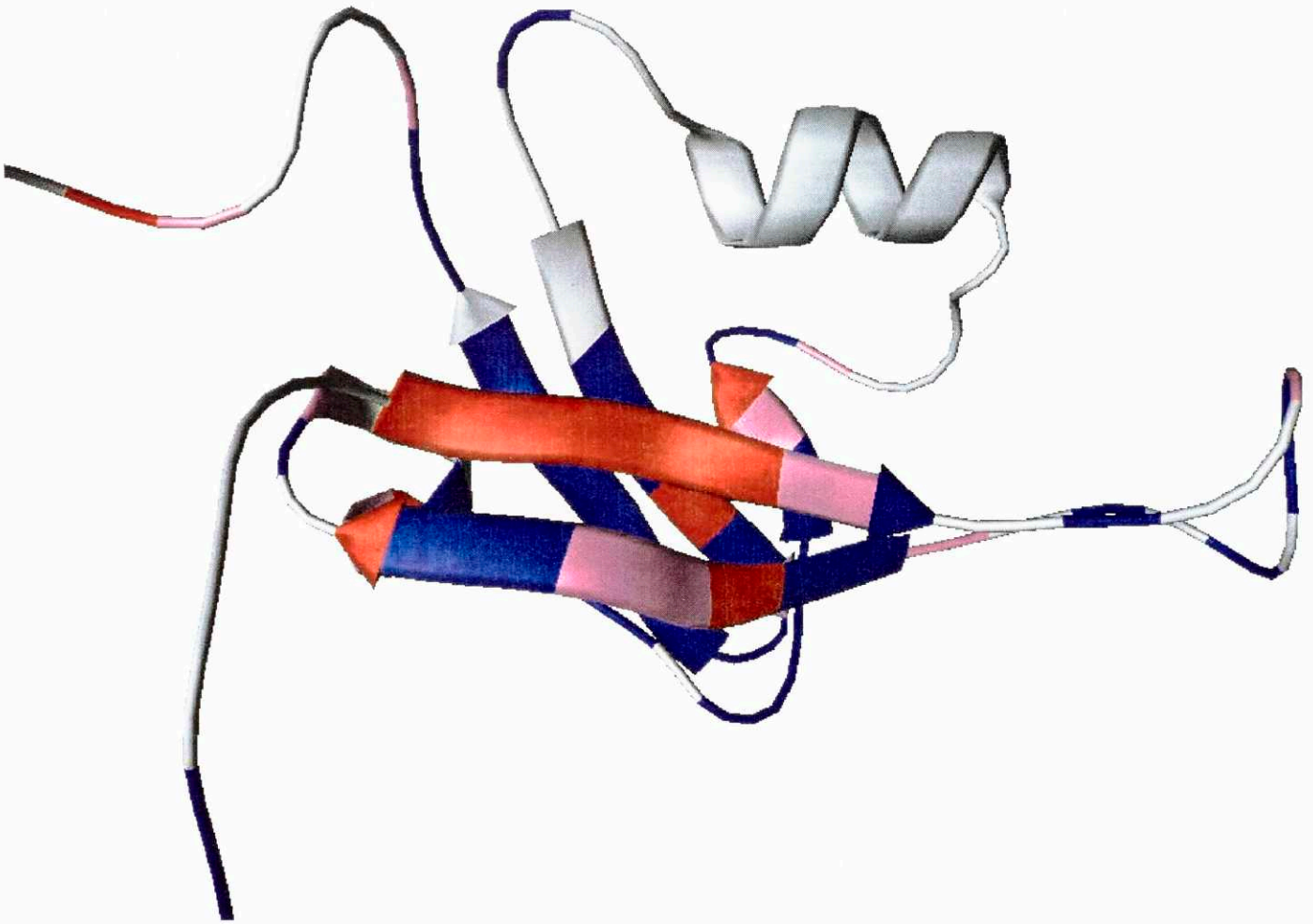

Figure 2-8. DegS PDZ model showing interactions with solYYF peptide. A series of HSQCs was taken while adding solYYF peptide to a solution of DegS PDZ domain. Residues that showed a significant change in chemical shift upon binding are hilighted in red. Those regions that show a small change in chemical shift are shown in magenta. Regions in blue showed no change upon binding, and regions in grey could not be accurately determined. The major changes lie in the upper beta-strand of the top sheet, which in other PDZ domains is known to interact with peptides.

Peptide binding assayed by isothermal calorimetry

Four peptides were tested for binding to the DegS domain using ITC experiments (Table2-6). One peptide (NH_2-DNVGLVYQF-COOH) had the last six residues of OmpC. This peptide aggregated to some extent at the concentrations used in the ITC experiments. Although binding was observed, a K_d could not be determined (data not shown). The remaining peptides were synthesized with an N-terminal sequence (DNRDGNV) designed to increase solubility. The solYYF peptide bound with a K_d of 0.63 ± 0.1 μM (Figure 2-9A), the solYQF peptide had a K_d of 15 ± 1.2 μM (Figure 2-9B), and the solYQM peptide had a K_d of 34 ± 3 μM (Figure 2-9C). In each of these cases, binding had a favorable enthalpy (ΔH) and an unfavorable entropy (ΔS).

Table 2-6. Binding of DegS PDZ domain to peptides by ITC

Peptide	[Peptide] (μM)	[DegS PDZ] (μM)	Temp (°C)	K_d (μM)	ΔH (cal mol^{-1})
OmpC	397	44.8	25	>30[1]	ND
solYYF	1180	76.6	25	0.54	-10940
solYYF	133.7	12.9	25	0.71	-12770
solYQF	306	39.5	25	13.9	-11550
solYQF	1260	122.7	25	16.2	-8805
solYQM	788.8	113.9	25	36.9	-11270
solYQM	1290	70.18	25	31.3	-13690

1 The peptide aggregates and the K_d could not be measured precisely.

Each of the experiments shown in Figure 2-9 are at approximately the same concentrations of peptide. The K_d of the solYYF peptide was 2000 fold lower than the initial concentration and this results in a very sharp sigmoidal transition with well defined upper and lower baselines. The K_d of the solYQF peptide was 80 fold lower than the initial concentration and shows a sigmoid with a more gradual transition. The K_d of the solYQM peptide was 40 fold

lower than the initial concentration and has a binding curve that is almost linear at low peptide:protein ratios and begins to plateau without ever reaching saturation.

The sub-micromolar dissociation constant between DegS PDZ domain and the solYYF peptide is typical of PDZ-peptide interactions. The K_d, as measured by BIAcore, between specific peptides and the PDZ domains of mDlg-2 (Songyang *et al.*, 1997), PTPbas-3 (Songyang *et al.*, 1997), CFTR-1 (Raghuram *et al.*, 2001) and CFTR-2 (Raghuram *et al.*, 2001) were 42 nM, 154 nM, 14 nM and 75 nM respectively.

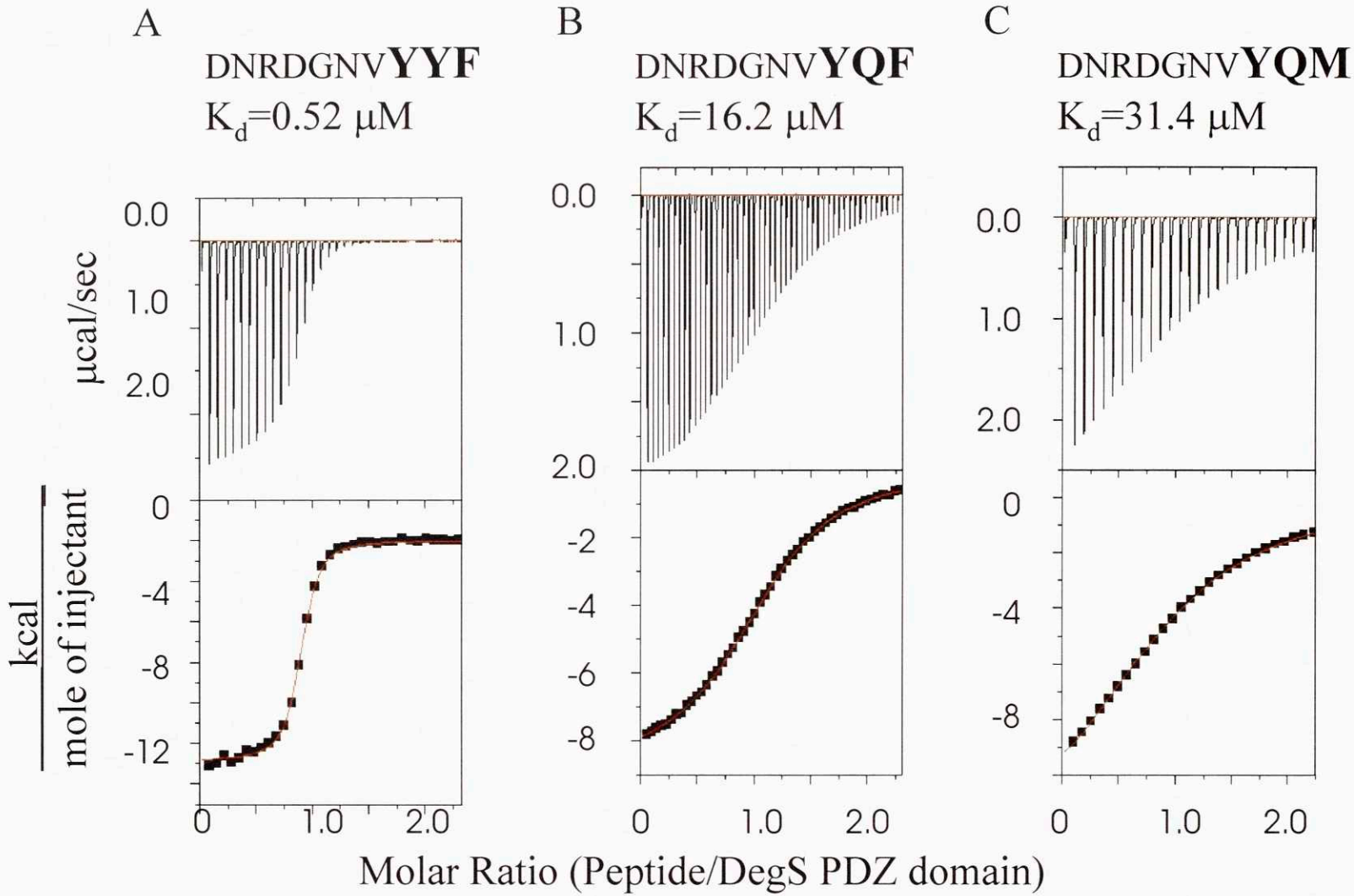

Figure 2-9. DegS PDZ domain + peptide assayed by isothermal titration calorimetry. Data (black squares) were obtained from titrating peptide in 6 μL aliquots into DegS PDZ domain. All solutions contained 50 mM $NaPO_4$ (pH 7.5). This data was fit (red line) to obtain a K_d. (A) A solution of 1.18 mM solYYF was titrated into a solution of 76.6 μM DegS PDZ domain. (B) A solution of 1.26 mM solYQF was titrated into a solution of 122.7 μM DegS PDZ domain. (C) A solution of 129 mM solYQM was titrated into a solution of 70.2 μM DegS PDZ domain.

Discussion

The region of DegS thought to correspond to a PDZ domain folded into a stable structure and displayed a circular-dichroism spectrum consistent with the secondary structure of a PDZ domain. NMR analysis, while not providing a complete structure, supported this domain being part of the PDZ family. Two strands of the top beta-sheet (B2 and B3), the first alpha helix (α1), and three strands of the bottom beta-sheet (B4, B6, and B7) were clearly identified. Information about the other regions (B1, L1, B5, and α2) were dependent upon homology based dihedral and hydrogen bond restraints. The DegS PDZ domain had several other features in common with the PDZ family, including binding to C-terminal peptide sequences. NMR studies indicated that the peptide-binding pocket of the DegS PDZ domain was relatively well ordered. Upon binding, resonances in beta-strand B1, B2, and α1 were shifted, whereas resonances in most of the domain were unaffected. The DegS PDZ domain binds peptides ending in YYF with 0.63 μM affinity and peptides ending with YQF with 15 μM affinity. This sequence specificity indicates that outer-membrane proteins are likely to bind DegS. Such an interaction, if physiological, would provide a link between OMPs and the extracytoplasmic-stress response pathway.

DegS peptide-recognition domain is a PDZ domain

Sequence homology between the C-terminal domain of DegS and PDZ domains is weak with less than 20% identity. However, multiple alignments of the family of bacterial PDZ domains, including DegS, DegP, DegQ, Tsp, and EcfE, showed reasonable homology in terms of residues conserved in the PDZ domains of known structure. The modeled structure suggests that the DegS domain is divergent from known PDZ domains in several ways. For example, the L1 loop (13 residues) is much longer than in other PDZ domains and seems to be largely disordered in the isolated domain. Additional differences occur at the termini, the binding pocket, and the

regions surrounding the binding pocket. One strand of the bottom beta-sheet is formed by the C-terminus of the PDZ domains from DegS and the D1 protease, but is formed by the N-terminus in other PDZ domains of known structure. This helps to explain some of the wide divergence seen at the termini between DegS and other PDZ containing proteins. All of the PDZ domains that were discovered initially had a GLGF sequence in the binding loop, whereas the DegS sequence has a GIGG. The leucine and the phenylalanine of the GLGF sequence pack together and create a hydrophobic pocket that accommodates the C-terminal residue of the bound peptide, which is usually a valine or leucine. However, some PDZ domains that bind peptides with aromatic C-termini (Songyang *et al.*, 1997) have "GLGF" sequences with smaller side chains at positions 2 and 4 creating a larger hydrophobic pocket for the terminal residue of the bound peptide. The glycine at the last position of the DegS GIGG motif is consistent with this domain binding peptides with aromatic residues at the C-terminal position.

Peptide-binding specificity

Peptide libraries were used to identify C-terminal sequences that bind to the DegS PDZ domain, resulting in identification of a Tyr-Tyr-Phe-COOH motif.

Binding of potential ligands to the DegS PDZ domain was assayed by NMR and/or isothermal titration calorimetry. Two different substrates were tested in NMR experiments, a peptide ending in Tyr-Tyr-Phe (solYYF) and a peptide with the ssrA degradation-tag sequence (sol-ssrA). The solYYF peptide bound stoichiometrically to the DegS domain and showed slow exchange with an exchange rate of less than 300 sec^{-1} at 298 K. In ITC experiments, this peptide bound the DegS domain with a K_d of 0.6 μM. The sol-ssrA peptide bound very weakly with a millimolar K_d and was in fast to intermediate exchange.

ITC experiments showed that a peptide ending with GNVYQF, a sequence found in several outer-membrane proteins, bound the DegS PDZ domain with a K_d of approximately 15 μM. This result is highly suggestive, especially as over-expression of outer-membrane proteins is known to result in DegS-mediated degradation of RseA (Ades *et al.*, 1999). A peptide ending with GNVYQM bound to the DegS domain about 2-fold more weakly than the GNVYQF peptide, supporting the peptide selection data, which suggested that Phe was preferred to Met at the last position. It is not clear, however, that unassembled OMPs reach concentrations in the periplasm that would be required to saturate a binding site with a K_d of 15 μM. If the periplasm is 10 nm thick and the surface area of the bacteria is 6 μm^2 then the total periplasmic space is 0.06 μm^3. The outer-membrane proteins with the C-terminal sequence, YxF, are present at levels below those of the most common outer membrane protein (OmpA), which is at 100,000 copies/cell. Assuming 1000 copies of these potentially interacting OMPs per cell and that 90% of them are stably folded, yields a concentration of unfolded OMPs of 28 μM. The levels of DegS are thought to be much lower and the DegS protease and PDZ domains are located in the periplasmic space although DegS is anchored in the inner membrane. If the DegS protein is present at 50 copies per cell and the entire periplasmic space is used as the volume, then a concentration of 1.4 μM is obtained. Using a K_d of 15 μM, the amount of bound DegS is calculated to be about 36% of the total DegS population, under these conditions.

It is possible, however, that DegS is oligomeric in the membrane and that OMP trimers, which have not been incorporated into the outer membrane, bind significantly tighter because of the chelate affect. A second possibility is that additional sequences in an OMP monomer might contribute to a stronger interaction (see below and Table 2-7).

DegS may recognize unfolded OMPs

No proteins encoded in the *E. coli* genome (Blattner *et al.*, 1997) end in YYF or YYM. However, 34 proteins end with two out of three residues of these motifs (i.e. YxF) (Table 2-7). The most striking feature of these sequences is that about half of them are annotated as outer-membrane proteins (OMPs). In addition, more than 70% of these proteins are extracytoplasmic by the criterion that they contain signal sequences. If only sequences that end in YxF or YxM are considered, then 22 of 24 proteins are extracytoplasmic and all of the OMPs in Table 2-7 are included. Use of profiles that included the relative weight of each amino acid at each position did not improve the discrimination between cytoplasmic and extracytoplasmic proteins and yielded a largely overlapping set of proteins.

Assuming an equal distribution of amino acids at each position in the 4275 genes present in the *E. coli* K12 genome, then calculating the expected number of matching genes is a matter of multiplying the number of genes by the probability of finding the amino acid(s) at each position. For the sequence tyrosine-tyrosine-phenylalanine at the C-terminus there are expected to be 4275*(1/20)*(1/20)*(1/20) genes that match that sequence (Table 2-8).

Table 2-7. *E. coli* sequences that match two of the three residues in the C-terminal motif YYF.

Gene	Signal Sequence	C-termini	internal sequences
Adhesin	ss	GIKYSF	
btub	ss	SGSYTF	YGY,YYS,YGF,YYN,YDF
cirA	ss	AVDYRF	YGF,YYG,YYM,YYN,RYF
fadL	ss	NFNYAF	YYY,YAF
fhuE	ss	TGTYQF	YYS,TYF,RYF,YNF
lomR	ss	GVGYKF	
mipA	ss	GITYKF	YYL,LYF,YYY,YNF
nmpC	ss	GLVYQF	HYF,YEF,YQF,YYF
ompC	ss	GLVYQF	HYF,YQF,YYF
ompF	ss	GIVYQF	HYF,YNF,YQF,NYF,YYF
ompG	ss	GVNYSF	VYF,YYQ,YQF,YKF,YTF, YYL,YDF,YAF
ompN	ss	GLVYQF	HYF,YQF,YYF
ompT	ss	GLKYTF	YSF,YYS,YYV,YNF
ompP	ss	GLKYTF	
ompW	ss	SAGYRF	WYF
ompX	ss	GVGYRF	YYG,YGF
phoE	ss	GMTYQF	YDF,YQF,YYF
tsx	ss	VVGYNF	WYF,YRF,KYF
ycgH	ss	GTTYKF	
yedS	ss	GLVYQF	YQF,YYF
yneA	ss	IGKYDF	YYI
yshA	ss	GTKYFF	YNF,YNF,YKF,HYF,PYF
ubiE	-	HRGYKF	YSF
kdgR	?	YHGYPF	YRF
inaA	-	AYYYAM	YPF
ydiY	ss	SLGYSM	YLF
yibS	-	CFRYSM	YYN,YYR
cysD	-	KRQGYF	YEF,YGF,YSF
ybck	-	DGEIYF	YYP,YYV,IYF
yceO	-	IYVFYF	YLF,YVF
ydjA	-	PFVTYF	
hemK	-	LGRYYQ	
yadK	ss	VDIYYE	GYF
yajL	?	IYNYYE	

Table 2-8. Expected and actual C-terminal sequences.

Pattern	Expected number	Actual number
YYF	1	0
YYM	1	0
Yx_1F	10	24
x_1YF	10	4
Yx_1M	10	3
x_1YM	10	0
YYx_2	10	3

x_1 is all amino acids except tyrosine, and x_2 is all amino acids except phenylalanine and methionine.

The pattern Yx_1F is more than two-fold over-represented, while all other patterns are at least two-fold under-represented.

The conserved C-terminal phenylalanine has been noted on many, although not all, outer-membrane proteins and alignment of the C-termini of the majority of outer-membrane proteins revealed a hxhxhxYxF-COOH (where h is a hydrophobic and x is any residue) motif (Struyve *et al.,* 1991). Of the18 *E. coli* genes that end with GxxYxF or GxYxF, 14 of them are known to be outer-membrane proteins. The other four have no known function or annotation. However, closer inspection reveals that they have probable signal sequences and are likely to be outer-membrane proteins. This same C-terminal OMP motif is also conserved in other gram-negative bacteria. Intriguingly, many of the OMPs also contain internal YYx, xYF, or YxF sequences (Table 2-7). Six positions within OmpF and its close homologues are shown (Table 2-9). In several of the OMPs, whose structures are known, including OmpF, many of the internal YYFs occur in beta-sheets followed by beta-hairpins. Internal sequences of this type could plausibly also bind to the DegS PDZ domain.

Table 2-9. Sequences in OmpF homologues that may bind PDZ domains.

Escherichia coli - OmpF		HYF	YNF	YQF	NYF	YYF	YQF
Enterobacter cloacae	78%[1]	HYF	YNF	YQF	NYI	YYF	YQF
Serratia marcescens	61%[1]	HYF	YNV	YQF	NYV	YYF	YQF
Yersinia pestis	59%[1]	HYF	YNF	YLF	KYV	YYF	YQF
Escherichia coli - PhoE	60%[1]	HYM	AEF	YQF	NYI	YYF	YQF

1: Percent identity to *E. coli* OmpF.

Role of DegS PDZ domain in extracytoplasmic-stress response

Outer-membrane protein over-expression activates the σ^E pathway (Mecsas *et al.*, 1993), leading to degradation of RseA by DegS and release of σ^E. Previous models have focused on RseA and RseB as sensors of OMP-mediated stress signals (Ades *et al.*, 1999, Collinet *et al.*, 2000, Missiakas *et al.*, 1997), with DegS functioning largely to recognize and degrade RseA. However, the results presented here suggest that the DegS PDZ domain may bind outer-membrane proteins directly, raising the possibility that DegS itself may be a sensor of OMP-mediated stress. This model predicts that over-expression of just the C-terminal region of an outer-membrane protein should activate the stress response.

In collaboration with Ben Alba and Carol Gross (UCSF), two proteins were constructed and tested. The first was β-lactamase with an OmpC tag (DNVGLVYQF) at the C-terminus. The second was the last 40 amino acids of OmpC fused to a pelB leader sequence. The β-lactamase construct (with only one potential binding site) did not induce σ^E activity, while the OmpC construct (with two potential binding sites) was able to induce σ^E activity about two-fold (personal communication). The OmpC construct with a mutated internal YYF was still able to induce two-fold, while mutation of the terminal YQF led to no induction. Changing the terminal YQF to YYF led to an eight-fold increase in induction as would be expected on the basis of the peptide library results. Future experiments are planned to try to identify the minimal sequence

needed to induce the system and whether this minimal fragment signaling depends upon a functional PDZ domain.

Alba and Gross have also constructed and studied DegS mutants that have the PDZ domain deleted (DegS(–PDZ)), have an active site mutation (DegS*), or have both the deletion and the active-site mutation (DegS*(–PDZ)). They found that DegS(–PDZ) complements a degS null and supported σ^E activity following OMP over-expression. The inducibility of σ^E in these unpublished experiments depended on the active site of DegS, but not on the presence of its PDZ domain. Taken at face value, this result seems to rule out a role for the DegS PDZ domain in OMP detection. One confounding factor, however, was that the levels of OMPs in these experiments were high enough to potentially block transport into the periplasm and to affect the folding and processing of other exported proteins, which in turn, might give rise to alternative-stress signals. In addition, Alba and Gross found that basal σ^E levels in the degS null strains were very low, but could be returned to normal by increasing the expression of either the DegS(–PDZ) or the DegS*(–PDZ). Two mechanisms might explain these results. In the first model, the unbound PDZ domain inactivates DegS, with this auto-inhibitory repression being relieved when the PDZ domain binds an outer-membrane protein. This would allow OMP-bound DegS to interact with RseA and activate σ^E. Because the active-site mutant restores σ^E levels, the RseA-DegS interaction alone would have to be sufficient for σ^E release. Auto-inhibition could occur because the DegS PDZ domain binds to the protease domain of DegS (although there are no obvious internal YYF or related target sequences in this region) or could occur because of conformational occlusion.

A second model is that DegS may equilibrate between inactive monomers and active oligomers, with the DegS PDZ domain helping to stabilize oligomers when bound to OMPs in a

multivalent fashion. In fact, DegP is reported to be at least a hexamer, but becomes a dimer or trimer when its PDZ domains are deleted (Sassoon *et al.*, 1999). If the PDZ domain enhances, but is not required for DegS oligomerization, then the near wild-type basal activity of over-expressed DegS(–PDZ) could result from its increased concentration alone.

Application to other organisms

The extracytoplasmic-stress signaling in *Escherichia coli* is analogous to the alginate signaling in *Pseudomonas aeruginosa*. *Escherichia coli* uses the σ^E pathway to respond to extracytoplasmic stress, whereas *Pseudomonas* can alter this pathway to set up lethal mucoid infections. Formation of the alginate exopolysaccharide forms a barrier that reduces the amount of antibiotics that reach the bacteria and protect it from our own immune response. Recurring *Pseudomonas* infection and the inflammation that accompanies it gradually damage the lungs, causing respiratory failure, which is the leading cause of death among cystic fibrosis patients (Schurr *et al.*, 1995). Cystic fibrosis is the most common human genetic disease of its severity in the United States and the second most common overall; hereditary hemochromatosis is the most common although it does not present until the patient is in their fifties (Edwards *et al.*, 1988). The mucoid colonies in *Pseudomonas* collect mutations in the RseA homologue, MucA, which prevent the inhibition of an alternative sigma factor homologous to σ^E. This suggests that one target for antibiotics should be the alternative sigma factor itself, which should affect the alginate coating. The *Escherichia coli* and the *Pseudomonas aeruginosa* alternative sigma factors are so highly conserved with *E. coli* that a lethal deletion of the alternative sigma factor can be rescued by the *Pseudomonas* alternative sigma factor. This strain could be easily screened for drugs that block the extracytoplasmic response.

Chapter 3. Tsp PDZ domain

Introduction

Tail-specific protease (Tsp) degrades substrates with hydrophobic C-termini (Keiler & Sauer, 1996). Full-length Tsp was subjected to proteolysis and stable fragments were characterized. A Tsp fragment encompassing residues 223 to 341 forms a stable domain and binds ssrA-tagged substrates, specifically and strongly.

Materials and Methods

Purification of Tsp

Full-length Tsp with a C-terminal 6-His tag was expressed from a plasmid, pKK101, in *E. coli* strain KS1000 (Silber *et al.*, 1994). A 1 L LB prep was grown at 37 °C until an OD_{600} of 0.4-0.5. The cells were induced with 1 mL 100 mM IPTG for 2-4 hours. The cells were centrifuged and the pellet was resuspended in five volumes of 100 mM Tris HCl (pH 8), 200 mM KCl, 1 mM EDTA, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 1.4 mM 2-mercapto ethanol. The cells were lysed by sonication and insoluble protein and membranes were removed by centrifugation. After centrifugation, the supernatant was loaded onto a 2 mL column of Ni-NTA agarose resin (Qiagen) and the flow-through fraction was reapplied. The column was washed with 100 mL equilibration buffer (100 mM NaH_2PO_4, 10 mM Tris adjusted to pH 8). The column was washed with 50 mL of equilibration buffer adjusted to pH 6.1 and 50 mL equilibration buffer adjusted to pH 5.46. Bound protein was eluted with 6 mL of equilibration buffer adjusted to pH 2.8. The protein eluted from the Ni-NTA agarose resin was further purified by chromatography on SP-sephadex or Superdex-75 (Pharmacia). Ni-NTA purified TspHis6 for determining digestion conditions was loaded onto an SP-sephadex column, washed with equilibration buffer adjusted to pH 4.0, and eluted with equilibration buffer adjusted to pH 7.0. Similar to a selective ammonium

sulphate precipitation, most contaminating proteins were precipitated by concentrating the protein in a centriprep-30 and adjusting the pH to 8.0. The TspHis6 remains soluble. For large-scale purification of the proteolytically stable domain, Ni-NTA purified TspHis6 was dialyzed against 10 mM NaH_2PO_4, 1 mM Tris (pH 8) and chromatographed on a 24 mL Superdex-75 column at 0.4 mL/min in the same buffer. Peak fractions were identified by SDS-gel electrophoresis and pooled.

Domain mapping and cloning

TspHis6 (20 μg) was digested in 20 μL of 100 mM Tris (pH 8) using chymotrypsin (0.4-100 μg) or trypsin (0.1-250 μg) for 1 hr at 37 °C. The digestion reactions were electrophoresed on an SDS gel and transferred to 0.45 μm pore PVDF membrane (Millipore Immobilon-P) at 173 mAmps for 30 min. The PVDF membrane was stained with coomassie blue, destained in 50% methanol, and bands corresponding to TspHis6 fragments were excised and subjected to N-terminal Edman degradation (MIT Biopolymers Laboratory).

In a separate experiment, TspHis6 (3 mg) was digested with 30 μg of chymotrypsin or trypsin in 3 mL 100 mM Tris (pH 8). The digestion was quenched with 50 μL of 0.2 M PMSF and 10 μL of DTT and chromatographed on a 120 mL Sephadex G-50 column at 0.7 mL/min with 50 mM ammonium bicarbonate. Fractions corresponding to the proteolytically stable domains were pooled and dialyzed twice against 6 L of 10 mM ammonium bicarbonate (pH 7.8). The domains were lyophilized and analyzed by matrix-assisted laser desorption/ionization (MALDI) mass spectroscopy.

Primer NPW-36 (GCATACGTCCCATGGACCCGCATACCAACTACTTTTC-C) contains an NcoI site, an ATG start site, and the region coding for amino acid 223-230 of Tsp.

NPW-35 (GGATACGTCCCATGGTCTCGAGGCGCCCTTCAATTTCGAGA-CGAATACG) contains an XhoI site, a stop codon, and the region coding for amino acid 337-341 of Tsp. The region coding for amino acids 223-341 of Tsp was PCR amplified from plasmid pKK101 plasmid, cut with NcoI and XhoI, and ligated into plasmid pET24d to create pNPW3635, using standard molecular biology procedures (Current Protocols in Molecular Biology © 2000 John Wiley & Sons).

PDZ domain purification

pNPW3635 was transformed into BL21(DE3) pLysS (Novagen) and the Tsp PDZ domain was expressed. Five 1 L LB preps were induced with 1 mL 100 mM IPTG / L at an OD_{600} of 0.3-0.5 for 3-5 hours. Cells were resuspended in 40 mL of buffer S (300 mM NaCl, 10 mM imidazole, and 50 mM $NaPO_4$, pH 8.2) and lysed by sonication. After centrifugation, the supernatant was loaded onto a 2 mL column of Ni-NTA agarose resin and the flow-through fraction was reapplied. The column was washed with 56 mL buffer S. 10 x 2 mL of buffer S plus 0, 0, 0, 50, 100, 200, 300, 400, 500, and 500 mM imidazole respectively was added to the column to wash and elute the bound protein. The eluted protein was dialyzed against 50 mM $NaPO_4$ (pH 5.5) and was loaded onto a Mono S HR5/5 (Pharmacia) column at a rate of 1 mL min^{-1}. A 40 min 0 M to 250 mM NaCl gradient in 50 mM $NaPO_4$ (pH 5.5) was run, yielding protein of > 95% purity when assayed by comassie-stained SDS-PAGE.

Fluorescence and CD assays

Circular-dichroism experiments were performed in a manner similar to the DegS experiments described in Chapter 2, typically at a protein concentration of 4.5 μM in 50 mM $NaPO_4$ (pH 6), 55 mM NaCl. Fluorescence spectra of the Tsp PDZ domain were collected using

a Fluoromax-2 (Instruments S. A., Inc.). Following excitation at 278 nm, the center of mass of the resulting tryptophan fluorescence from 300-450 nm was calculated. Urea denaturation was assayed by following the shift in the center of mass.

Far-ELISA assays

Binding assays were performed using a modified ELISA. The Tsp PDZ domain (9 μg / 50 μL) was incubated overnight in PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4 , pH 7.3) in a 96 well plate. The plate was washed twice with PBS and blocked with 200 μL of 0.3% BSA in PBS for 2 hours. The plate was washed twice with TBSTM (10 mM Tris, 150 mM NaCl, 0.1% Tween 20, 0.1% dried non-fat milk pH7.2). 50 μL of phage P22 Arc Repressor (Arc) containing an ssrA tag or Arc containing an ssrA tag with the last two residues changed to aspartate, at varying concentrations (0.0-2.8 μM), was added to the plate, incubated for 1.5 hours, and washed twice with TBSTM without milk (TBST). 50 μL of a 1:5000 dilution of rabbit anti-arc antibodies was added to the plate, incubated for 1 hour, and washed twice with TBST. 50 μL of a 1:5000 dilution of goat anti-rabbit antibodies conjugated to horseradish peroxidase was added to the plate, incubated for 45 min, and washed twice with TBST. Immuno-complexes were reacted with 100 μL TMB substrate (Kirkegaard & Perry Labs) for 4.5 minutes before adding 100 μL of Stop solution (Kirkegaard & Perry Labs). Absorbance at 450 nm was measured using a SpectraMAX plus plate reader.

NMR

The purified Tsp PDZ domain was concentrated in a centricon-10 and the buffer was exchanged with 18 mM sodium phosphate (pH 5.0). 50 μL of D_2O was added to 420 μL of the

solution containing the Tsp PDZ domain. 1D spectra were acquired with presaturation during the relaxation delay and the NOE mixing time to improve water suppression. The Tsp PDZ domain was 180 μM with 18 mM $NaPO_4$ (pH 5) at 277 K with 256 scans and referenced to TMSP.

Results and Discussion

Initial digestion with trypsin and chymotrypsin revealed that Tsp contains a proteolytically stable domain (Figure 3-1). N-terminal Edman degradation identified the starting positions of the fragments and the mass spectroscopy gave C-terminal boundaries. The trypsin-generated fragment starts at position 233 and ends near position 339. The chymotrypsin generated fragments start at position 229 and 242 and also end near position 339. The Tsp PDZ domain originally predicted by Ponting encompassed residues 239 to 325 (Ponting, 1997). A slightly larger fragment (residues 223-341) of Tsp containing the PDZ domain was cloned, expressed, and purified.

This Tsp PDZ domain had a cooperatively folded structure in urea-denaturation studies monitored by tryptophan fluorescence (Figure 3-2A). The tryptophan was partially buried at 0 M urea and became red-shifted (solvent-exposed) during the urea titration. The center of mass moved from 355 nm to 360 nm. The CD spectrum was consistent with a structure containing a mix of alpha-helix, beta-sheet, and random coil (data not shown). The dispersion of chemical shifts in 1D-NMR spectra can be used as a probe for a native protein. Proton resonances below 0.5 ppm usually result from aliphatic methyl protons that are packed near aromatic protons in the core. Resonances between 5 and 6 ppm are usually from alpha protons that are involved in beta-sheets. Resonances above 9 ppm are usually from structured amide or aromatic protons. The

Tsp PDZ domain has proton resonances in all three of these regions (Figure 3-2B), supporting the interpretation of a stable, native structure.

Using a Far-ELISA assay (Figure 3-3), the Tsp PDZ domain bound arc-repressor with a C-terminal ssrA tag with a half-maximal binding of less than 20 nM but did not bind arc-repressor with an ssrA tag ending with Asp-Asp. This binding mirrors the substrate specificity demonstrated by full length Tsp (Keiler & Sauer, 1996), and indicates that this PDZ domain is responsible for Tsp's substrate specificity.

After these experiments were performed, another group independently published a report using a Tsp fragment from amino acid 206 to 334 that could be cross-linked to photo-reactive peptides (label-GRGYALAA) and has a dissociation constant of 1.9 μM (Beebe *et al.*, 2000). They also tested the sensitivity of full length Tsp to proteolysis and found a major site of cleavage between amino acid 211 and 212. Both my results and those published by Beebe *et al.* indicate that Tsp has a PDZ domain and that this domain binds substrate.

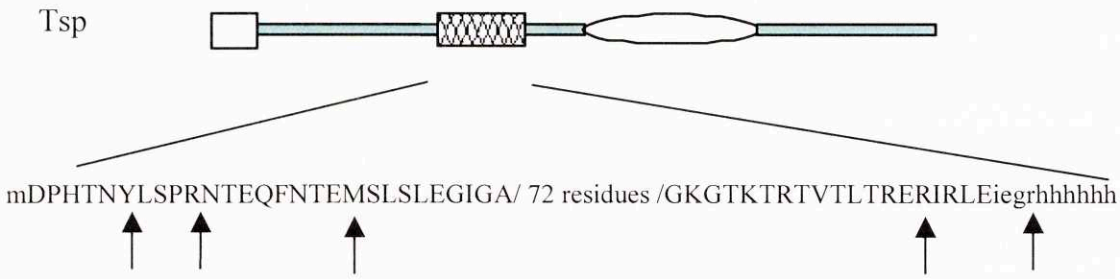

Figure 3-1. Proteolysis of Tsp. After digestion of full length Tsp with trypsin or chymotrypsin there were stable fragments. Characterization of the fragments by Edman degradation led to the identification of the N-termini (arrows). Using a smaller region of Tsp and similar digestion conditions, MALDI mass spectroscopy led to identification of the C-termini (arrows). Sequence in lower case letters are from the expression construct and are not found in full-length Tsp.

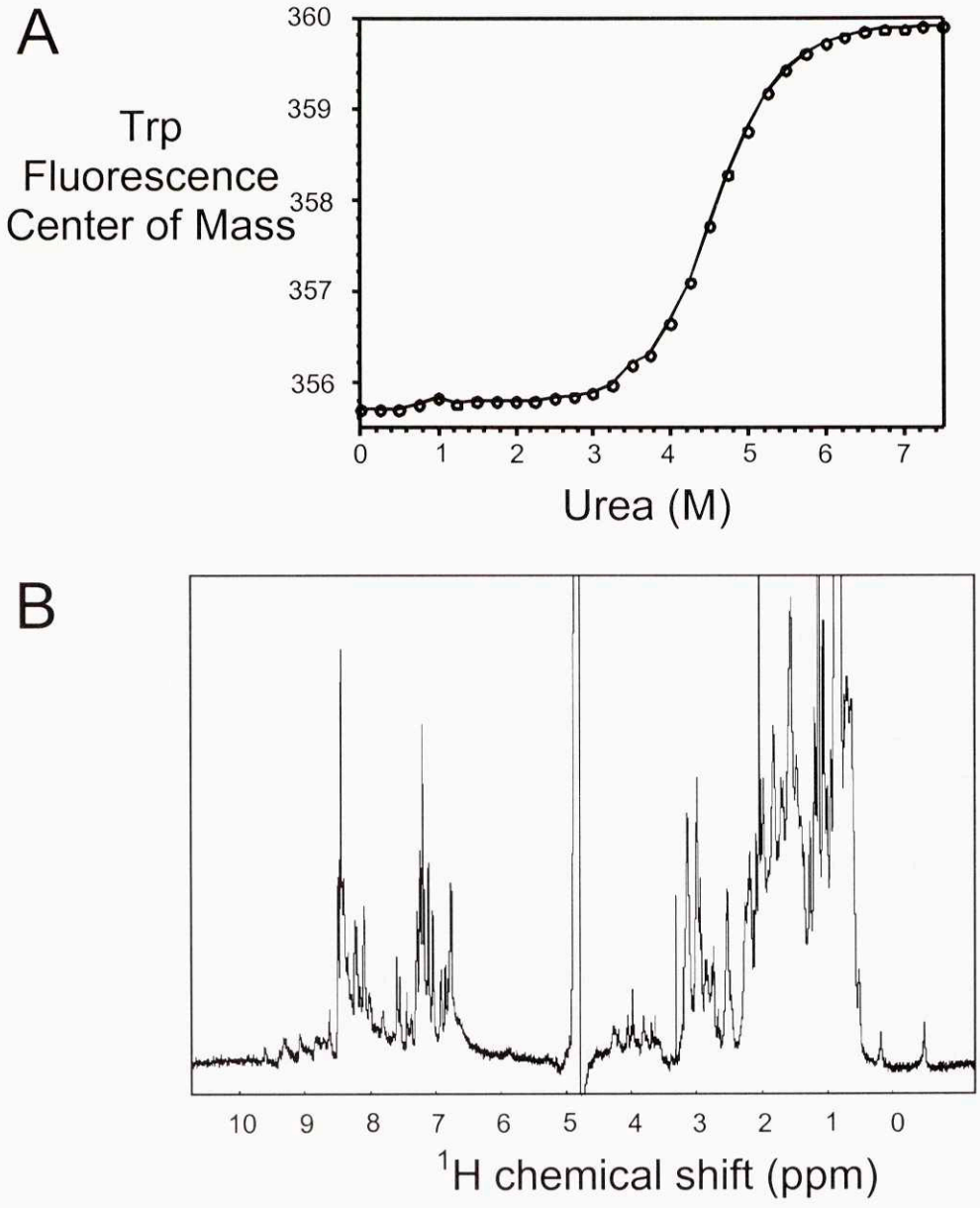

Figure 3-2. Biochemical characterization of the Tsp PDZ domain. Urea denaturation (A) was monitored by following the center of mass of the tryptophan fluorescence and showed a single, cooperative unfolding. A 1D-NMR spectrum (B) of the Tsp PDZ domain shows chemical shift dispersions expected of native proteins.

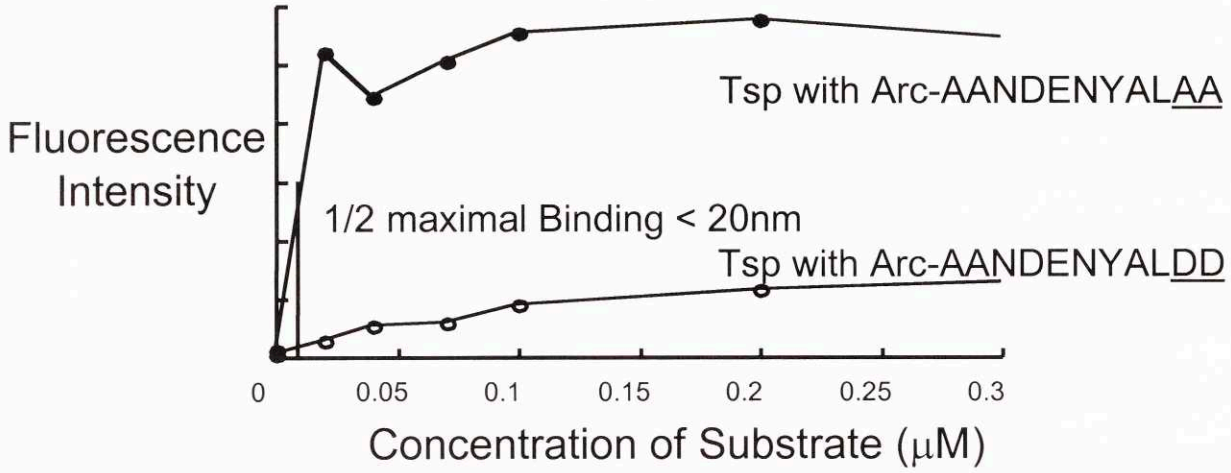

Figure 3-3. Far-ELISA of Tsp PDZ domain with Arc variants. The Tsp PDZ domain selectively bound Arc variants tagged with an ssrA tag (AANDENYALAA) (filled circles), but not Arc variants tagged with an ssrA-DD tag (AANDENYALAA) (open circles).

Appendix

A

A construct containing the PDZ domains from DegP (Spiess *et al.*, 1999) was expressed and attached to a column in the same manner as DegS. Using the KNxxxxxx library there were no significant differences between the DegP column and the glycine control column. All six randomized positions looked similar to the first three randomized positions from the DegS PDZ domain experiment (Figure 2-5A-C).

B

To improve the shape of NMR data from experiments with a limited number of increments I wrote a script that implements a Hamming function (Cavanagh *et al.*, 1996) in NMRPipe.

```
/** Hamming window function                          **/
/** coded by Nathan P Walsh (c) 2000                 **/
/** "nmrPipe -fn MAC -macro hamming.M"               **/

for (i=0;i<size;i++)
   {
    temp     = rdata[i]*(0.54+0.46*cos(PI*i/size));
    rdata[i] = temp;
    temp     = idata[i]*(0.54+0.46*cos(PI*i/size));
    idata[i] = temp;
   }
```

Acknowledgements

Gratitude is due to my mentor, Dr. Robert Sauer, for his leadership, guidance, and support. I wish to thank all of the members that have made up my thesis committee during my time at MIT for their aide and direction: Drs. Tania Baker, David Bartel, Peter Kim, C. James McKnight, and Mike Yaffe. I am grateful to both Drs. C. James McKnight and Sue Pochapsky for their help with NMR experiments. Dr. Carol Gross and members of her lab were kind enough to provide the original DegS PDZ construct I used and to share pre-publication data with me. I also want to thank members of the Sauer lab for the training and interaction they provided, especially Dr. Matthew Cordes who introduced me to NMR and Dr. Arie Berggrun who helped construct plasmid pNPW3635.

No acknowledgement of mine would be complete without mentioning my wife, Alison, whose patience and tolerance I have tested on innumerable occasions. Through this period she has also been my best friend and helped me to finish by keeping me focused.

Bibliography

Ades, S. E., Connolly, L. E., Alba, B. M., and Gross, C. A. 1999. The *Escherichia coli* sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. *Genes Dev* 13:2449-61.

Aramburu, J., Yaffe, M. B., Lopez-Rodriguez, C., Cantley, L. C., Hogan, P. G., and Rao, A. 1999. Affinity-driven peptide selection of an NFAT inhibitor more selective than cyclosporin A. *Science* 285:2129-33.

Bass, S., Gu, Q., and Christen, A. 1996. Multicopy suppressors of prc mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated R1pA. *J Bacteriol* 178:1154-61.

Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M., and Kay, L. E. 1990. Practical Aspects of Proton-Carbon-Carbon-Proton Three-Dimensional Correlation Spectroscopy of ^{13}C-Labeled Proteins. *J. Magn. Reson.* 87:620-627

Bax, A., and Pochapsky, S. S. 1992. Optimized recording of heteronuclear multi-dimensional NMR spectra using pulsed field gradients. *J Magn Reson*:638-643

Beebe, K. D., and Pei, D. 1998. A continuous fluorimetric assay for tail-specific protease. *Anal Biochem* 263:51-6.

Beebe, K. D., Shin, J., Peng, J., Chaudhury, C., Khera, J., and Pei, D. 2000. Substrate recognition through a PDZ domain in tail-specific protease. *Biochemistry* 39:3149-55.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453-74.

Bosch, D., Scholten, M., Verhagen, C., and Tommassen, J. 1989. The role of the carboxy-terminal membrane-spanning fragment in the biogenesis of *Escherichia coli* K12 outer membrane protein PhoE. *Mol Gen Genet* 216:144-8.

Bowie, J. U., and Sauer, R. T. 1989. Identification of C-terminal extensions that protect proteins from intracellular proteolysis. *J Biol Chem* 264:7596-602.

Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., M., R. L., Simonson, T., and Warren, G. L. 1998. Crystallography and NMR System (CNS): A New Software System for Macromolecular Structure Determination. *Acta Cryst* D54:905-921.

Bystroff, C., and Baker, D. 1997. Blind predictions of local protein structure in CASP2 targets using the I-sites library. *Proteins* Suppl:167-71.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. 1996. *Protein NMR Spectroscopy Principles and Practice* Boston: Academic Press, Inc.

Cho, K. O., Hunt, C. A., and Kennedy, M. B. 1992. The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. *Neuron* 9:929-42.

Ciechanover, A. 1994. The ubiquitin-mediated proteolytic pathway: mechanisms of action and cellular physiology. *Biol Chem Hoppe Seyler* 375:565-81.

Collinet, B., Yuzawa, H., Chen, T., Herrera, C., and Missiakas, D. 2000. RseB binding to the periplasmic domain of RseA modulates the RseA:SigmaE Interaction in the cytoplasm and the availability of SigmaE-RNA polymerase. *The Journal of Biological Chemistry* 275:33898-33904.

Cornilescu, G., Delaglio, F., and Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289-302.

Daniels, D. L., Cohen, A. R., Anderson, J. M., and Brunger, A. T. 1998. Crystal structure of the hCASK PDZ domain reveals the structural basis of class II PDZ domain target recognition. *Nat Struct Biol* 5:317-25.

Dartigalongue, C., Missiakas, D., and Raina, S. 2001. Characterization of the *Escherichia coli* sigma E regulon. *J Biol Chem* 276:20866-75.

Davis, A. L., Keeler, J., Laue, E. D., and Moskau, D. 1992. Experiments for recording pure-absorption heteronuclear correlation spectra using pulse field gradients. *J Magn Reson* 98:207-16.

de Cock, H., Struyve, M., Kleerebezem, M., van der Krift, T., and Tommassen, J. 1997. Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of *Escherichia coli* K-12. *J Mol Biol* 269:473-8.

De Las Penas, A., Connolly, L., and Gross, C. A. 1997. The sigmaE-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of sigmaE. *Mol Microbiol* 24:373-85.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX Pipes. *J Biomol NMR* 6:277-293.

Diner, B. A., Ries, D. F., Cohen, B. N., and Metz, J. G. 1988. COOH-terminal processing of polypeptide D1 of the photosystem II reaction center of *Scenedesmus obliquus* is necessary for the assembly of the oxygen-evolving complex. *J Biol Chem* 263:8972-80.

Domian, I. J., Quon, K. C., and Shapiro, L. 1997. Cell type-specific phosphorylation and proteolysis of a transcriptional regulator controls the G1-to-S transition in a bacterial cell cycle. *Cell* 90:415-24.

Domian, I. J., Reisenauer, A., and Shapiro, L. 1999. Feedback control of a master bacterial cell-cycle regulator. *Proc Natl Acad Sci U S A* 96:6648-53.

Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. 1996. Crystal structures of a complexed and peptide-free membrane protein- binding domain: molecular basis of peptide recognition by PDZ. *Cell* 85:1067-76.

Edwards, C. Q., Griffen, L. M., Goldgar, D., Drummond, C., Skolnick, M. H., and Kushner, J. P. 1988. Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. *N Engl J Med* 318:1355-62.

Erickson, J. W., and Gross, C. A. 1989. Identification of the sigma E subunit of *Escherichia coli* RNA polymerase: a second alternate sigma factor involved in high- temperature gene expression. *Genes Dev* 3:1462-71.

Fanning, A. S., and Anderson, J. M. 1996. Protein-protein interactions: PDZ domain networks. *Curr Biol* 6:1385-8.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33:5984-6003.

Fesik, S. W., and Zuiderweg, E. R. P. 1990. Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules. *Quarterly Reviews of Biophysics* 23:97-131.

Gottesman, S., Roche, E., Zhou, Y., and Sauer, R. T. 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* 12:1338-47.

Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A. M., and Bax, A. 1992. ^{1}H, ^{13}C, and ^{15}N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry* 31:8180-90.

Hall, R. A., Ostedgaard, L. S., Premont, R. T., Blitzer, J. T., Rahman, N., Welsh, M. J., and Lefkowitz, R. J. 1998. A C-terminal motif found in the beta2-adrenergic receptor, P2Y1 receptor and cystic fibrosis transmembrane conductance regulator determines binding to

the Na^+/H^+ exchanger regulatory factor family of PDZ proteins. *Proc Natl Acad Sci U S A* 95:8496-501.

Hall, R. A., Premont, R. T., Chow, C. W., Blitzer, J. T., Pitcher, J. A., Claing, A., Stoffel, R. H., Barak, L. S., Shenolikar, S., Weinman, E. J., Grinstein, S., and Lefkowitz, R. J. 1998. The beta2-adrenergic receptor interacts with the Na^+/H^+-exchanger regulatory factor to control Na^+/H^+ exchange. *Nature* 392:626-30.

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* 173:4799-813.

Herman, C., Thevenet, D., Bouloc, P., Walker, G. C., and D'Ari, R. 1998. Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Dev* 12:1348-55.

Herman, C., Thevenet, D., D'Ari, R., and Bouloc, P. 1995. Degradation of sigma 32, the heat shock regulator in *Escherichia coli*, is governed by HflB. *Proc Natl Acad Sci U S A* 92:3516-20.

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

Hillier, B. J., Christopherson, K. S., Prehoda, K. E., Bredt, D. S., and Lim, W. A. 1999. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science* 284:812-5.

Jansen, C., Heutink, M., Tommassen, J., and de Cock, H. 2000. The assembly pathway of outer membrane protein PhoE of *Escherichia coli*. *Eur J Biochem* 267:3792-800.

Jenal, U., and Fuchs, T. 1998. An essential protease involved in bacterial cell-cycle control. *Embo J* 17:5658-69.

Johnson, B. A., and Blevins, R. A. 1994. NMRView: A computer program for the visualization and analysis of NMR data. *J Biomolecular NMR* 4:603-614.

Kanemori, M., Nishihara, K., Yanagi, H., and Yura, T. 1997. Synergistic roles of HslVU and other ATP-dependent proteases in controlling *in vivo* turnover of sigma32 and abnormal proteins in *Escherichia coli*. *J Bacteriol* 179:7219-25.

Kanemori, M., Yanagi, H., and Yura, T. 1999. Marked instability of the sigma(32) heat shock transcription factor at high temperature. Implications for heat shock regulation. *J Biol Chem* 274:22002-7.

Karthikeyan, S., Leung, T., Birrane, G., Webster, G., and Ladias, J. A. 2001. Crystal structure of the PDZ1 domain of human Na^+/H^+ exchanger regulatory factor provides insights into the mechanism of carboxyl- terminal leucine recognition by class I PDZ domains. *J Mol Biol* 308:963-73.

Karthikeyan, S., Leung, T., and Ladias, J. A. 2001. Structural basis of the Na^+/H^+ exchanger regulatory factor PDZ1 interaction with the carboxyl-terminal region of the cystic fibrosis transmembrane conductance regulator. *J Biol Chem* 276:19683-6.

Kay, L. E., Ikura, M., and Bax, A. 1990. Proton proton correlation via carbon carbon couplings--A 3-dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with C-13. *J Am Chem Soc* 112:888-889.

Kay, L. E., Keifer, P., and Saarinen, T. 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114:10663-5.

Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. 1993. A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain ^1H and ^{13}C correlations in H_20 Samples of Protein. *J Magn Reson Ser B*:333-337.

Keiler, K. C., and Sauer, R. T. 1995. Identification of active site residues of the Tsp protease. *J Biol Chem* 270:28864-8.

Keiler, K. C., and Sauer, R. T. 1996. Sequence determinants of C-terminal substrate recognition by the Tsp protease. *J Biol Chem* 271:2589-93.

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 Sci* 4:1507-15.

Keiler, K. C., Waller, P. R., and Sauer, R. T. 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* 271:990-3.

Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: A program for displaying and analysis of macromolecular structures. *J Mol Graphics* 14:51-55.

Kornau, H. C., Schenker, L. T., Kennedy, M. B., and Seeburg, P. H. 1995. Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. *Science* 269:1737-40.

Kozlov, G., Gehring, K., and Ekiel, I. 2000. Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptor. *Biochemistry* 39:2572-80.

Liao, D. I., Qian, J., Chisholm, D. A., Jordan, D. B., and Diner, B. A. 2000. Crystal structures of the photosystem II D1 C-terminal processing protease. *Nat Struct Biol* 7:749-53.

Liberek, K., Galitski, T. P., Zylicz, M., and Georgopoulos, C. 1992. The DnaK chaperone modulates the heat shock response of *Escherichia coli* by binding to the sigma 32 transcription factor. *Proc Natl Acad Sci U S A* 89:3516-20.

Liberek, K., and Georgopoulos, C. 1993. Autoregulation of the *Escherichia coli* heat shock response by the DnaK and DnaJ heat shock proteins. *Proc Natl Acad Sci U S A* 90:11019-23.

Martinez-Salazar, J. M., Moreno, S., Najera, R., Boucher, J. C., Espin, G., Soberon-Chavez, G., and Deretic, V. 1996. Characterization of the genes coding for the putative sigma factor AlgU and its regulators MucA, MucB, MucC, and MucD in *Azotobacter vinelandii* and evaluation of their roles in alginate biosynthesis. *J Bacteriol* 178:1800-8.

Mecsas, J., Rouviere, P. E., Erickson, J. W., Donohue, T. J., and Gross, C. A. 1993. The activity of sigma E, an *Escherichia coli* heat-inducible sigma- factor, is modulated by expression of outer membrane proteins. *Genes Dev* 7:2618-28.

Missiakas, D., Mayer, M. P., Lemaire, M., Georgopoulos, C., and Raina, S. 1997. Modulation of the *Escherichia coli* sigmaE (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. *Mol Microbiol* 24:355-71.

Morais Cabral, J. H., Petosa, C., Sutcliffe, M. J., Raza, S., Byron, O., Poy, F., Marfatia, S. M., Chishti, A. H., and Liddington, R. C. 1996. Crystal structure of a PDZ domain. *Nature* 382:649-52.

Moyer, B. D., Duhaime, M., Shaw, C., Denton, J., Reynolds, D., Karlson, K. H., Pfeiffer, J., Wang, S., Mickle, J. E., Milewski, M., Cutting, G. R., Guggino, W. B., Li, M., and Stanton, B. A. 2000. The PDZ-interacting domain of cystic fibrosis transmembrane conductance regulator is required for functional expression in the apical plasma membrane. *J Biol Chem* 275:27069-74.

Muhandiram, D. R., Farrow, N. A., Xu, G. Y., Smallcome, S. H., and Kay, L. E. 1993. A gradient ^{13}C NOESY-HSQC experiment for recording NOESY spectra of ^{13}C-labeled proteins dissolved in H_2O. *J Magn Reson Ser B* 102:317-321.

Muhandiram, D. R., and Kay, L. E. 1994. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *J. Magn. Reson. Series B* 103:203-216.

Muhlhahn, P., Zweckstetter, M., Georgescu, J., Ciosto, C., Renner, C., Lanzendorfer, M., Lang, K., Ambrosius, D., Baier, M., Kurth, R., and Holak, T. A. 1998. Structure of interleukin 16 resembles a PDZ domain with an occluded peptide binding site. *Nat Struct Biol* 5:682-6.

Pallen, M. J., and Wren, B. W. 1997. The HtrA family of serine proteases. *Mol Microbiol* 26:209-21.

Palmer, A. G., Cavanagh, J., Wright, P. E., and Rance, M. 1991. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J Magn Reson*:151-170.

Parsell, D. A., Silber, K. R., and Sauer, R. T. 1990. Carboxy-terminal determinants of intracellular protein degradation. *Genes Dev* 4:277-86.

Ponting, C. P. 1997. Evidence for PDZ domains in bacteria, yeast, and plants. *Protein Sci* 6:464-8.

Ponting, C. P., Phillips, C., Davies, K. E., and Blake, D. J. 1997. PDZ domains: targeting signalling molecules to sub-membranous sites. *Bioessays* 19:469-79.

Raghuram, V., Mak, D. D., and Foskett, J. K. 2001. Regulation of cystic fibrosis transmembrane conductance regulator single-channel gating by bivalent PDZ-domain-mediated interaction. *Proc Natl Acad Sci U S A* 98:1300-5.

Raivio, T. L., and Silhavy, T. J. 2001. Periplasmic stress and ecf sigma factors. *Annu Rev Microbiol* 55:591-624

Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., *et al.* 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245:1066-73.

Rouviere, P. E., De Las Penas, A., Mecsas, J., Lu, C. Z., Rudd, K. E., and Gross, C. A. 1995. rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*. *Embo J* 14:1032-42.

Saras, J., and Heldin, C. H. 1996. PDZ domains bind carboxy-terminal sequences of target proteins. *Trends Biochem Sci* 21:455-8.

Sassoon, N., Arie, J. P., and Betton, J. M. 1999. PDZ domains determine the native oligomeric structure of the DegP (HtrA) protease. *Mol Microbiol* 33:583-9.

Schleucher, J., Sattler, M., and Griesinger, C. 1993. Coherence Selection via Gradients without Signal Attenuation: Application to 3D HNCO. *Angew Chem* 105:1518-21.

Schultz, J., Hoffmuller, U., Krause, G., Ashurst, J., Macias, M. J., Schmieder, P., Schneider-Mergener, J., and Oschkinat, H. 1998. Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. *Nat Struct Biol* 5:19-24.

Schurr, M. J., Yu, H., Boucher, J. C., Hibler, N. S., and Deretic, V. 1995. Multiple promoters and induction by heat shock of the gene encoding the alternative sigma factor AlgU (sigma E) which controls mucoidy in cystic fibrosis isolates of *Pseudomonas aeruginosa*. *J Bacteriol* 177:5670-9.

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 U S A* 89:295-9.

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* 242:237-40.

Smith, C. K., Baker, T. A., and Sauer, R. T. 1999. Lon and Clp family proteases and chaperones share homologous substrate- recognition domains. *Proc Natl Acad Sci U S A* 96:6678-82.

Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., Crompton, A., Chan, A. C., Anderson, J. M., and Cantley, L. C. 1997. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* 275:73-7.

Spiess, C., Beil, A., and Ehrmann, M. 1999. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* 97:339-47.

Straus, D., Walter, W., and Gross, C. A. 1990. DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of sigma 32. *Genes Dev* 4:2202-9.

Straus, D. B., Walter, W. A., and Gross, C. A. 1987. The heat shock response of *E. coli* is regulated by changes in the concentration of sigma 32. *Nature* 329:348-51.

Struyve, M., Moons, M., and Tommassen, J. 1991. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J Mol Biol* 218:141-8.

Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. 1991. The N-end rule in bacteria. *Science* 254:1374-7.

Tochio, H., Hung, F., Li, M., Bredt, D. S., and Zhang, M. 2000. Solution structure and backbone dynamics of the second PDZ domain of postsynaptic density-95. *J Mol Biol* 295:225-37.

Tochio, H., Zhang, Q., Mandal, P., Li, M., and Zhang, M. 1999. Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide. *Nat Struct Biol* 6:417-21.

Tommassen, J., van Tol, H., and Lugtenberg, B. 1983. The ultimate localization of an outer membrane protein of *Escherichia coli* K-12 is not determined by the signal sequence. *Embo J* 2:1275-9.

Tomoyasu, T., Gamer, J., Bukau, B., Kanemori, M., Mori, H., Rutman, A. J., Oppenheim, A. B., Yura, T., Yamanaka, K., Niki, H., *et al.* 1995. *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor sigma 32. *Embo J* 14:2551-60.

Tomoyasu, T., Ogura, T., Tatsuta, T., and Bukau, B. 1998. Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli*. *Mol Microbiol* 30:567-81.

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* 270:9322-6.

Waller, P. R., and Sauer, R. T. 1996. Characterization of degQ and degS, Escherichia coli genes encoding homologs of the DegP protease. *J Bacteriol* 178:1146-53.

Wang, Q. P., and Kaguni, J. M. 1989. A novel sigma factor is involved in expression of the rpoH gene of *Escherichia coli*. *J Bacteriol* 171:4248-53.

Wittekind, M., and Mueller, L. 1993. *J. Magn. Reson. Series B* 101:201.

Woods, D. F., and Bryant, P. J. 1993. ZO-1, DlgA and PSD-95/SAP90: homologous proteins in tight, septate and synaptic cell junctions. *Mech Dev* 44:85-9.

Yun, C. H., Oh, S., Zizak, M., Steplock, D., Tsao, S., Tse, C. M., Weinman, E. J., and Donowitz, M. 1997. cAMP-mediated inhibition of the epithelial brush border Na^+/H^+ exchanger, NHE3, requires an associated regulatory protein. *Proc Natl Acad Sci U S A* 94:3010-5.