Protein Science (1999), *8*:2533–2536. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Common protein architecture and binding sites in proteases utilizing a Ser/Lys dyad mechanism

MARK PAETZEL AND NATALIE C.J. STRYNADKA

Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

(RECEIVED May 3, 1999; ACCEPTED July 15, 1999)

Abstract: *Escherichia coli* signal peptidase (SPase) and *E. coli* UmuD protease are members of an evolutionary clan of serine proteases that apparently utilize a serine-lysine catalytic dyad mechanism. Recently, the crystallographic structure of a SPase inhibitor complex was solved elucidating the catalytic residues and the substrate binding subsites. Here we show a detailed comparison of the *E. coli* SPase structure to the native *E. coli* UmuD′ structure. The comparison reveals that despite a very low sequence identity these functionally diverse enzymes share the same protein fold within their catalytic core and allows by analogy for the assignment of the cleavage-site orientation and substrate binding subsites in the UmuD(D′) protease. The structural alignment of SPase and UmuD′ predicts important mechanistic and structural similarities and differences within these newly characterized families of serine proteases.

Keywords: clan SF proteases; lambda repressor; LexA repressor; protein fold; Ser/Lys dyad; serine protease; signal peptidase; UmuD protease

Based on their three-dimensional structures, serine proteases can be grouped into six evolutionary clans (SA, SB, SC, SE, SF, and SH) (Rawlings & Barrett, 1994). The clans SA, SB, SC, and SH although having very different protein folds, all utilize in their mechanism a histidine general base. The unique serine proteases in clans SE and SF utilize a lysine general base. The evolutionary clans can be further subcategorized into families. *Escherichia coli* type 1 signal peptidase (SPase), a member of the serine protease family S26, and *E. coli* UmuD protease, a member of the serine protease family S24, are the proteases within the clan SF that have been most thoroughly characterized at the structural level (Peat et al., 1996; Paetzel et al., 1998).

SPases are membrane-bound endopeptidases that function to cleave off the N-terminal signal peptide extension from proteins that are translocated across the lipid bilayers of cellular membranes. *E. coli* SPase has two transmembrane segments and a small cytoplasmic domain at its N-terminus (residues 1–74) and a larger C-terminal catalytic region (residues 75–323) located in the periplasmic space. Extensive site-directed mutagenesis (Tschantz et al., 1993) and chemical modification experiments (Paetzel et al., 1997) are consistent with bacterial SPases utilizing a Ser/Lys dyad mechanism whereby the lysine serves as a general base to activate the nucleophilic serine (Paetzel & Dalbey, 1997). The recently solved crystal structure of the catalytic domain of *E. coli* SPase with a β-lactam inhibitor bound to the nucleophilic Ser90 Oγ supports the general base role of Lys145 (Paetzel et al., 1998).

The C-terminal proteolytic domain of the UmuD protein (UmuD′) has extensive sequence similarity to the proteolytic domain of the large family of cI-λ and LexA-like repressors. UmuD is involved in damage inducible SOS mutagenesis in *E. coli*. The UmuD protein is activated for its role in the mutagenesis by an intermolecular or intramolecular (McDonald et al., 1999) self-cleavage reaction that removes its N-terminal 24 residues (Walker, 1995). This reaction requires the binding of UmuD to RecA* (RecA protein bound to single-stranded DNA and ATP). The active C-terminal fragment (UmuD′) works with UmuC, RecA*, and DNA polymerase III to facilitate translesion DNA synthesis. Although UmuD′ has been studied crystallographically (Peat et al., 1996), by NMR (Ferentz et al., 1997), and through site-directed mutagenesis (Nohmi et al., 1988), to date it has been difficult to assign a cleavage site orientation or specific substrate binding pockets on UmuD(D′).

In this paper we propose a substrate orientation and define the S1 and S3 substrate binding pockets (Schechter & Berger, 1967) for the *E. coli* UmuD(D′) proteases based on our structural comparison to the *E. coli* SPase-inhibitor complex. Our structural alignment gives mechanistic and structural insights into the UmuD(D′) proteases as well as the large and well-studied family of autocatalytic repressor/proteases such as cI-λ and LexA.

Results and discussion: The crystallographic structure of the catalytic region of *E. coli* type 1 SPase revealed a mainly β-sheet protein fold consisting of two antiparallel β-sheet domains (Paetzel et al., 1998). Domain I (Fig. 1A, in green) is the catalytic core region containing the catalytic residues Ser90 and Lys145 as well

Reprint requests to: Natalie C.J. Strynadka, University of British Columbia, Department of Biochemistry and Molecular Biology, Faculty of Medicine, 2146 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada; e-mail: natalie@byron.biochem.ubc.ca.

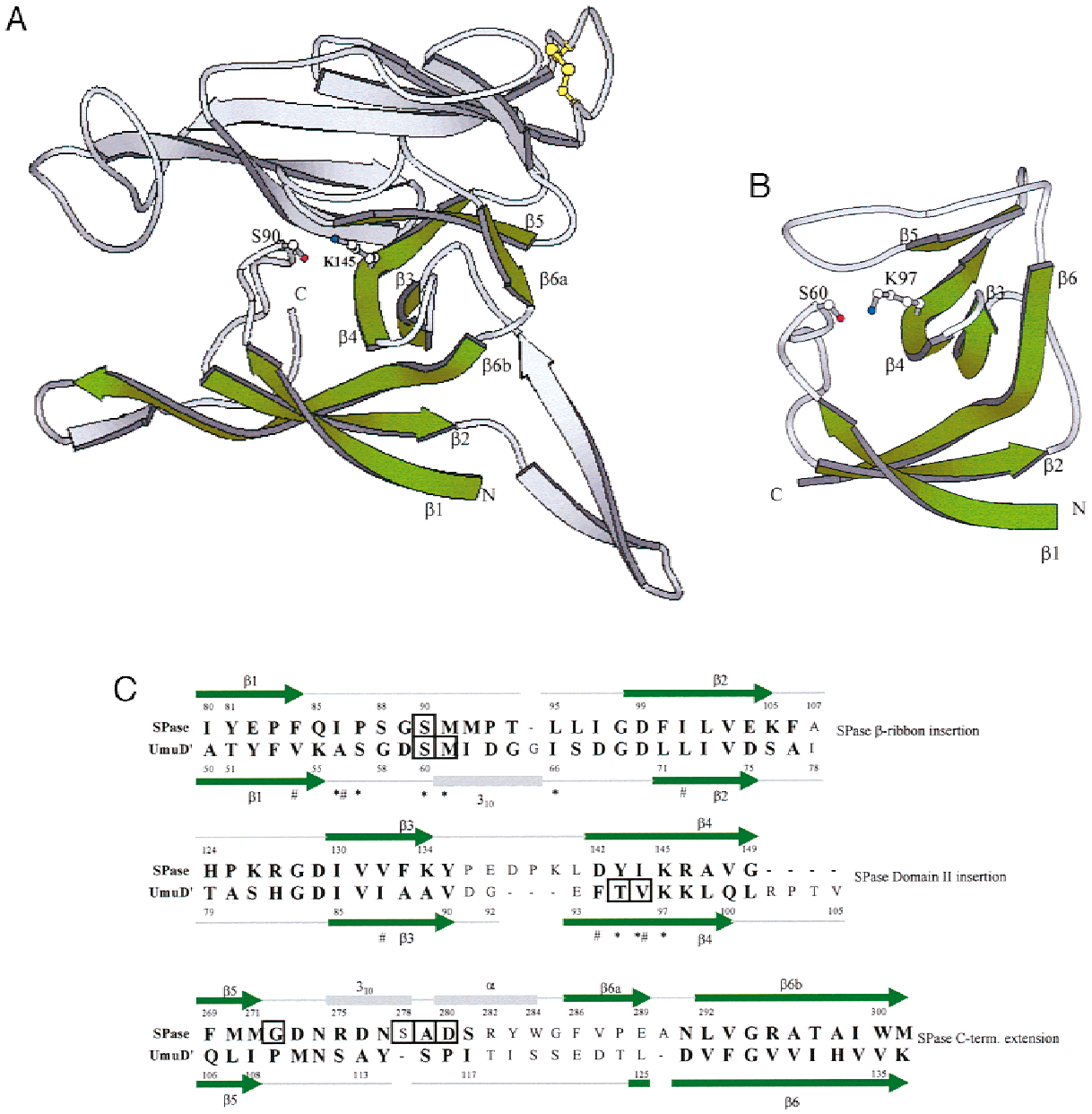

Fig. 1. A: A MOLSCRIPT (Kraulis, 1991) ribbon diagram of the catalytic domain of *E. coli* SPase (residues 80–323). The region of SPase that shares similarity in structure to UmuD′ is shown in green. **B:** A MOLSCRIPT (Kraulis, 1991) ribbon diagram of the *E. coli* UmuD′ protease (residues 50–136). **C:** A structure based sequence alignment of *E. coli* SPase (987643, PDB 1b12) and *E. coli* UmuD′ protease (137027, PDB 1umu). The secondary structure in *E. coli* SPase is indicated above the alignment. Only those regions of SPase sequence that structurally align with UmuD′ (domain I) are shown. The residues that structurally align are in bold. The secondary structural elements of *E. coli* UmuD′ are shown below the UmuD′ sequence. The residues that contribute atoms to the S1 and S3 substrate binding pockets of UmuD′ are indicated with the symbols * and #, respectively. Residues that contribute atom to the environment of the general base Lys Nζ are boxed.

as the residues making up the S1 and S3 substrate binding sites. Domain II (residues 150–268) (Fig. 1A, in gray) lies adjacent to domain I and is formed from a single contiguous insertion between the major strands four and five within domain I (Fig. 1A).

An overlap of the UmuD′ structure (residues 50–136) with *E. coli* SPase (residues 80–323) shows a striking and unexpected level of conservation of the overall topology of domain I. The least-squares superposition of the two molecules have an RMS of 1.6 Å for 69 common Cα atoms including the analogous residues to the SPase Ser90 and Lys145 (Ser60/Lys97). Despite limited sequence identity (17.4% for structurally aligned residues) (Fig. 1C) the protein fold of the catalytic core region of the serine protease families S26 (type 1 SPases) and S24 (UmuD(D′), and self-cleaving repressor/proteases) are the same.

UmuD′ has no structural counterpart for the SPase domain II (150–268) (Fig. 1A, in gray), the extended β-ribbon protruding

from domain I of SPase (residues 106–123) (Fig 1A, in gray) the SPase C-terminus (302–323) (Fig 1A, in gray) or a short loop region (138–140). The fact that the UmuD and LexA-like proteases lack the entire β-sheet domain II of SPase is interesting given that these enzymes require the binding of a second protein, the coprotease RecA*, for catalytic activity (Slilaty & Little, 1987; Nohmi et al., 1988). Intriguingly, the eukaryotic SPases, which also lack domain II, also require a second protein for catalytic activity (Fang et al., 1997).

Using the binding site information elucidated from the structure of the *E. coli* SPase-inhibitor complex (Paetzel et al., 1998) along with the structural superposition of the UmuD′ structure onto the SPase catalytic core structure, we are now able to assign the S1 and S3 substrate binding pockets of UmuD(D′) as well as define the orientation of the cleavage site relative to the binding sites.

The UmuD(D′) self-cleavage reaction cleaves the peptide bond between Cys24 and Gly25 (Shinagawa et al., 1988). Alignment of the scissile bond of the UmuD cleavage sites in an orientation similar to that of the analogous β-lactam amide of the SPase inhibitor (Paetzel et al., 1998) identifies Cys24 of UmuD as the P1 residue and Glu25 as the P1′ residue. This orientation of the cleavage site is consistent with Ser60 NH contributing to the oxyanion stabilization.

There is a large pocket adjacent to the active site (Ser60/Lys97) that correspond to the S1 and S3 binding pockets in the SPase structure (Paetzel et al., 1998). Docking of a tetrapeptide in an extended β-conformation corresponding to residues Cys24 to Leu21 of UmuD into this pocket (Fig. 2) allows the Cys24 (P1) and Val22 (P3) side chains to fit easily into the S1 and S3 binding pockets. The side chains of Gln23 (P2) and Leu21 (P4) point into the solvent. The modeled peptide (Cys24 to Leu21) makes β-sheet type hydrogen bonding contacts with the conserved β-strand (93–100) that contain the lysine general base (Fig. 2). Similar β-sheet interactions were also proposed in the modeling of the SPase/signal peptide complex (Paetzel et al., 1998). The residues contributing atoms to the UmuD′ S1 binding site (accommodating the P1 Cys24 side chain) are polar and hydrophobic in nature and include: Ala56, Ser57, Ser60, Met61, Ile66, Thr95, Val96, and the aliphatic portion of Lys97. The residues making up the UmuD′ S3 binding site (accommodating the P3 Val22 side chain) are hydrophobic in nature and include: Val54, Ala56, Leu72, Ile87, Phe94, and Val96 (Figs. 1C, 2).

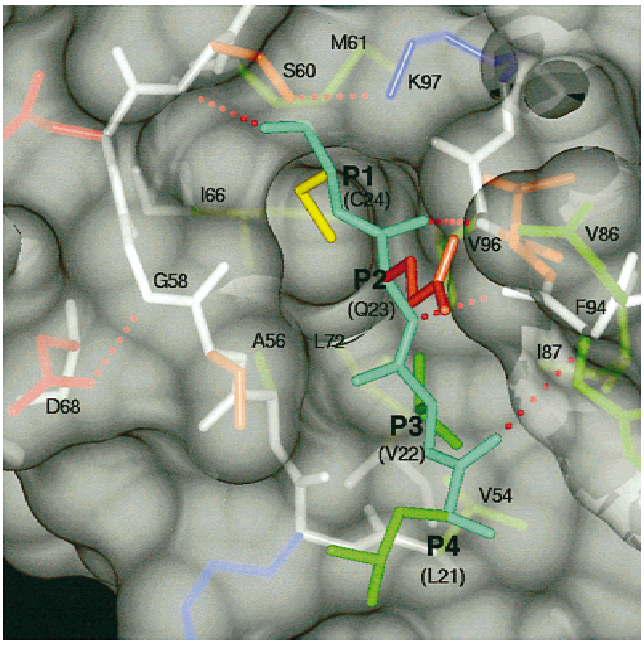

Fig. 2. The residues contributing to the UmuD(D′) S1 and S3 substrate binding subsites as seen behind a semitransparent molecular surface. The cleavage site is modeled such that the P1 Cys24 and P3 Val22 are pointing into the S1 and S3 substrate binding subsites, respectively. β-strand-type hydrogen bonds are shown between the main chain of the peptide substrate and the β-strand containing the general base Lys97. This figure was made with the program PREPI (http://www.bonsai.lif.icnet.uk/people/suhail/prepi.html).

An analytical analysis of the molecular surface of UmuD′ using the program CAST (Liang et al., 1998) reveals that the largest pocket on the UmuD′ surface corresponds to the substrate binding pocket (Fig. 2). It is a common feature of enzymes that the largest surface pocket corresponds to the active site binding pocket (Liang et al., 1998). The volume of the pocket was calculated to be 219.1 Å^3. The binding of a residue side chain into this pocket would provide significant binding energy during the reaction. Analysis of the SPase surface also revealed a correspondence of the largest pocket with the binding pocket. The calculated volume of the SPase pocket is 206.9 Å^3. Both UmuD′ and SPase S1 binding pockets are of sufficient size to accommodate a Cys or Ala/Gly/Ser/Cys side chain, respectively, which is consistent with the SPase substrate profile (Neilsen et al., 1997). The second largest pocket on the UmuD′ surface has a volume of 19.3 Å^3 and is located just across the other side of the active site from the S1 binding site. The residues contributing atoms to this pocket include Ile117, Pro116, Ser115, Pro109, Leu107, and Lys97.

It would appear from the position of the binding pockets and the orientation of the substrate relative to the Ser60 and Lys97 that the Ser60 of UmuD′, analogous to the SPase Ser90, attacks the amide bond of its substrate from the *si*-face. This is a unique feature of the clan SF proteases. All other characterized serine proteases appear to attack their substrates from the *re*-face of the amide bond (James, 1994).

An interesting common feature of the type 1 bacterial SPases, the autocatalytic repressors (e.g., cI-λ and LexA), and the SOS mutagenesis proteases such as UmuD(D′) is that they appear to utilize a Lys Nζ as their general base. In both the SPase acyl-enzyme inhibitor complex structure (Paetzel et al., 1998) and the native UmuD′ structure (Peat et al., 1996), the Lys145 Nζ and Lys97 Nζ are the only titratable group within the vicinity of the nucleophilic Ser90 Oγ and Ser60 Oγ, respectively. This is consistent with extensive site-directed mutagenesis and chemical modification experiments (Nohmi et al., 1988; Tschantz et al., 1993; Paetzel et al., 1997). To utilize a Lys as a general base an enzyme must provide an environment such that the pK_a of that Lys is depressed. Neither the SPase (Paetzel et al., 1998) nor the UmuD′ (Peat et al., 1996) structures are consistent with a local positive charge contributing to a depressed pK_a of the general base Lys. The other environmental scenario that can result in a depressed pK_a of a Lys Nζ is the burial of the Nζ within the enzyme/substrate complex. The Nζ of SPase Lys145 is almost completely buried in the SPase inhibitor complex. The Nζ is 90.8% occluded by other atoms as calculated by the program OS (Pattabiramin et al., 1995). The atoms surrounding Lys145 Nζ include Ser90 (Oγ, Cβ, C, and O), Gly272 (Cα), Ser278 (Oγ, Cβ, Cα, and C), Ala279 (Cβ, N, and C), and Asp280 (Cβ) as well as atoms from the β-lactam inhibitor. Lys145 Nζ is within hydrogen bonding distance to Ser90

Oγ (3.1 Å), Ser278 Oγ (2.6 Å), and the leaving group amide N of the inhibitor (3.1 Å). With the inhibitor removed the percent occluded surface on the Lys145 Nζ drops to 63.5%. The UmuD′ Lys97 Nζ, the proposed general base in UmuD′, is 48.0% occluded by surrounding atoms. The atoms surrounding UmuD′ Lys97 Nζ include Ser60 (Oγ, Cβ), Met61 (Cγ), Thr95 (Cγ2, Cβ, and Cα), and Val96 (O). Lys97 Nζ is within hydrogen bonding distance to Ser60 Oγ (2.8 Å) and Val96 O (3.1 Å). Much of the difference between the environments of the Lys Nζ general base in UmuD′ and SPase stems from the structural difference in the region which connects strand 5 and 6 (Fig. 1). UmuD′ lacks two small helices (275–278 and 280–284) in this region as well as Ser278 (*E. coli* numbering), which makes a hydrogen bond to Lys145 Nζ in SPase (Fig. 1C).

In summary, the detailed comparison of the SPase and UmuD′ three-dimensional structures have revealed new information regarding the substrate binding site of UmuD as well as possible mechanistic and structural differences and similarities in this newly characterized serine proteases clan.

Materials and methods: The structural alignment between *E. coli* SPase and *E. coli* UmuD′ was performed using the structural alignment module LSQ explicit within the program O (Jones et al., 1991). Secondary structure elements were assigned using the program PROMOTIF (Hutchinson & Thorton, 1996). The atomic coordinates for the *E. coli* SPase (1b12) and *E. coli* UmuD′ protease (1umu) are in the Protein Data Bank (PDB). The atomic coordinates for the *E. coli* SPase (1b12) will be released simultaneous to the publication of this paper.

Molecular docking of a tetrapeptide corresponding to the P1 to P4 residues of the UmuD cleavage site (Cys24 to Leu21) was performed using the program O (Jones et al., 1991). The program CAST (Liang et al., 1998) was used for the identification and characterization of the pockets on the surface of the UmuD′ and SPase structures. The molecular surface was defined using a water molecule with a radius of 1.4 Å. CAST is based on the computational geometry methods of alpha shape and discrete flow theory. The occluded surface area was calculated using the program OS (Pattabiramin et al., 1995).

Acknowledgments: This work was supported by the Medical Research Council of Canada and the Canadian Bacterial Diseases Network of Excellence. M.P. is funded by a Medical Research Council of Canada postdoctoral fellowship and NCJS by a Medical Research Council of Canada Scholarship. We thank Dr. Steve Mosimann (University of British Columbia) and Dr. Roger Woodgate (NIH) for many enthusiastic discussions.

References

Fang H, Mullin C, Green N. 1997. In addition to SEC11, a newly identified gene, SPC3, is essential for signal peptidase activity in the yeast endoplasmic reticulum. *J Biol Chem 272*:13152–13158.

Ferentz AE, Opperman T, Walker GC, Wagner G. 1997. Dimerization of the UmuD′ protein in solution and its implications for regulation of SOS mutagenesis. *Nature Struct Biol 4*:979–983.

Hutchinson EG, Thorton JM. 1996. PROMOTIF: A program to identify and analyze structural motifs in proteins. *Protein Sci 5*:212–220.

James MNG. 1994. Convergence of active center geometries among the proteolytic enzymes. In: Bond JS, Barrett AJ, eds. *Proteolysis and protein turnover*. Brookfield, Vermont: Portland Press. pp 1–8.

Jones TA, Zou JY, Cowan SW, Kieldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A47*:110–119.

Kraulis PG. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Liang J, Edelsbrunner H, Woodward C. 1998. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci 7*:1884–1897.

McDonald JP, Peat TS, Levine AS, Woodgate R. 1999. Intermolecular cleavage by UmuD-like enzymes: Identification of residues required for cleavage and substrate specificity. *J Mol Biol 285*:2199–2209.

Neilsen H, Engelbrecht J, Brunak S, von Heijne G. 1997. Identification of prokaryotic signal peptides and prediction of their cleavage sites. *Protein Eng 10*:1–6.

Nohmi T, Battista JR, Dodson LA, Walker GC. 1988. RecA-mediated cleavage activates UmuD for mutagenesis: Mechanistic relationship between transcriptional depression and post-translational activation. *Proc Natl Acad Sci USA 85*:1816–1820.

Paetzel M, Dalbey RE. 1997. Catalytic hydroxyl/amine dyads with serine proteases. *Trends Biochem Sci 22*:28–31.

Paetzel M, Dalbey RE, Strynadka NCJ. 1998. Crystal structure of a bacterial signal peptidase in complex with a β-lactam inhibitor. *Nature 396*:186–190.

Paetzel M, Strynadka NCJ, Tschantz WR, Casareno R, Bullinger PR, Dalbey RE. 1997. Use of site-directed chemical modification to study an essential lysine in *E. coli* leader peptidase. *J Biol Chem 272*:9994–10003.

Pattabiramin N, Ward KB, Fleming PJ. 1995. Occluded molecular surface: Analysis of protein packing. *J Mol Recognition 8*:334–344.

Peat TS, Frank EG, McDonald JP, Levine AS, Woodgate R, Hendrickson WA. 1996. Structure of the UmuD′ protein and its regulation in response to DNA damage. *Nature 380*:727–730.

Rawlings ND, Barrett AJ. 1994. Families of serine peptidases. *Methods Enzymol 244*:19–61.

Schechter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun 27*:157–162.

Shinagawa H, Iwasaki H, Kato T, Nakata A. 1988. RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. *Proc Natl Acad Sci USA 85*:1806–1810.

Slilaty SN, Little JW. 1987. Lysine 156 and Serine 119 are required for LexA repressor cleavage: A possible mechanism. *Proc Natl Acad Sci USA 84*:3987–3991.

Tschantz WR, Sung M, Delgado-Partin VM, Dalbey RE. 1993. A serine and a lysine residue implicated in the catalytic mechanism of the *E. coli* leader peptidase. *J Biol Chem 268*:27349–27354.

Walker GC. 1995. SOS-regulated proteins in translesion DNA synthesis and mutagenesis. *Trends Biochem Sci 20*:416–420.