2

Bacterial Type I Signal Peptidases

JOSEPH L. CARLOS* • MARK PAETZEL† •
PHILIP A. KLENOTIC‡ • NATALIE C. J. STRYNADKA† •
ROSS E. DALBEY*

**Department of Chemistry*
The Ohio State University
Columbus, Ohio 43210

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

‡*The Cleveland Clinic*
Center for Molecular Genetics
Cleveland, Ohio

THE ENZYMES, Vol. XXII

I. Introduction: Bacterial Signal Peptidase, Signal Peptides and Protein Targeting

In the bacterial secretory pathway, proteins that are exported across the plasma membrane are synthesized as higher molecular weight precursors with an N-terminal extension peptide. This extension peptide, called a signal or leader peptide, is proteolytically removed by type I signal peptidase (SPase I or leader peptidase, Lep). Unrelated to type I signal peptidases are the lipoprotein-specific signal peptidases (type II), which recognize lipid-modified eubacterial preproteins, and the (type IV) prepilin signal peptidases (*1, 2*). In the secretory pathway in bacteria, the apparent natural function of signal peptide processing by SPase I is the release of export-targeted and translocated proteins from the cytosolic membrane. Genetic studies in a number of bacteria have shown that SPase I is essential for cell viability (*3–7*). Mechanistic and structural analyses have helped explain how this enzyme binds substrate and how catalysis occurs by a unique mechanism. In this chapter the type I signal peptidase enzymes found in eubacteria with particular emphasis on the most thoroughly studied enzymes in this field, *E. coli* and *B. subtilis* SPase I, will be discussed. The focus will be on the biological and functional enzymology of SPase I. Finally, the three-dimensional X-ray crystal structure of *E. coli* SPase I will also be presented.

A. Bacterial Signal Peptides Are Essential for Preprotein Export

Proteins that are exported across the plasma membrane of bacteria typically use cleavable signal peptides. In bacteria, two preprotein translocation machineries are currently known, the Sec translocase and the Tat translocase systems (see Fig. 1). The Sec machinery in *E. coli* is comprised of the membrane-embedded protein components SecYEGDF and YajC (*8*), a peripheral membrane component SecA (*9–11*), and the cytoplasmic molecular chaperone SecB (*12*). SecB helps target some preproteins to the membrane by interaction the mature regions of the preprotein, as well as with membrane-bound SecA (*13*). The Tat pathway components consist of the integral membrane components TatA, TatB, TatC, and TatE (*14, 15*). TatA, TatB, and TatE have been found to be homologous to the Hcf106 protein (*14, 16, 17*) that is involved in the ΔpH-dependent protein export pathway of plant thylakoids. Resident bacterial inner membrane proteins typically lack cleavable signal peptides (*18*), whereas transient inner membrane proteins encoded by some phage genomes contain signal peptides that are removed during the membrane insertion process. A membrane-embedded protein, YidC, has been found to be essential for the proper insertion of the *M13* procoat protein [Fig. 1, (*19*)]. *M13* procoat was previously thought

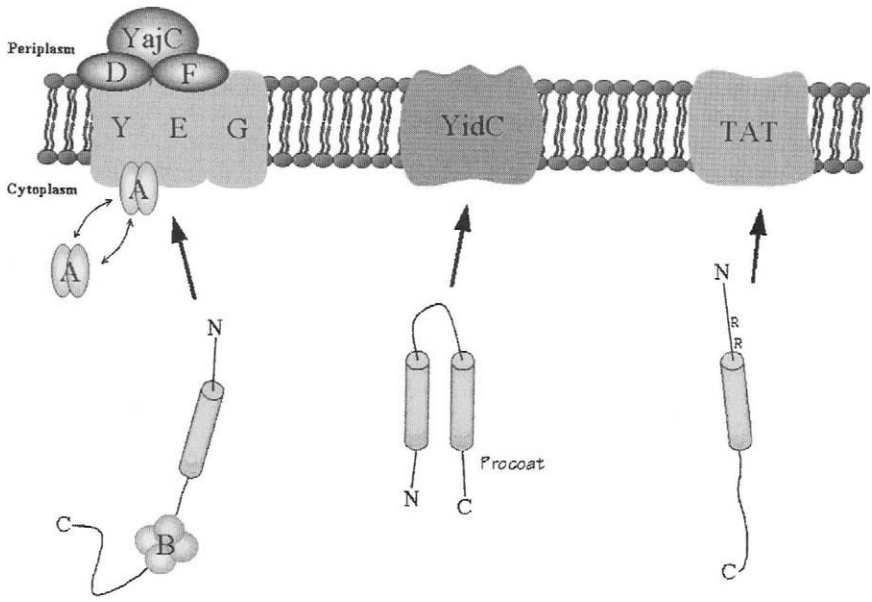
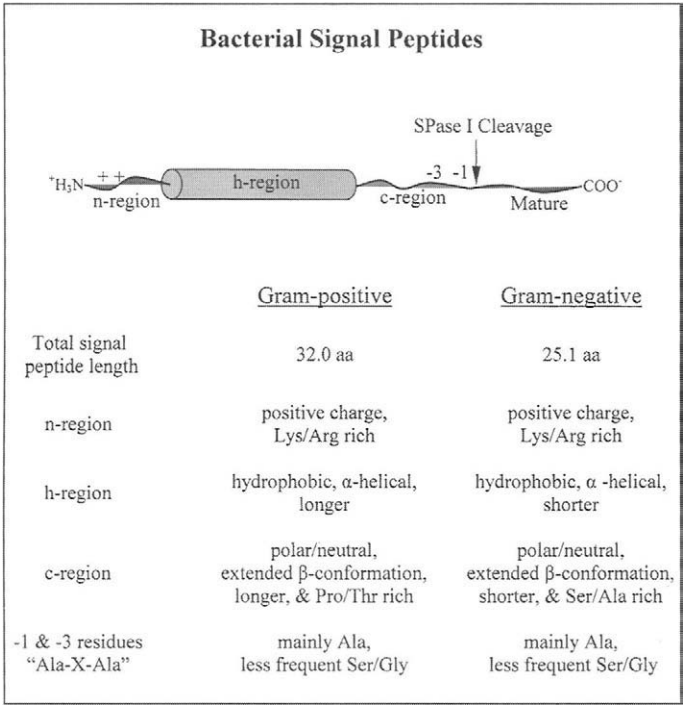


FIG. 1. Schematic representation of the known bacterial protein translocation systems.

to insert into the membrane by a spontaneous and Sec machinery–independent mechanism.

Although signal peptides do not maintain any overall sequence homology, they do contain some regions that have conserved features (*20, 21*) (Fig. 2). At the N-terminus of preproteins involved in the secretory pathway (destined for the outer membrane or cellular export) signal peptides typically contain a basic amino-terminal 1- to 5-residue N-region, a 7- to 15-residue central core hydrophobic H-region, and a polar carboxyl-terminal C-region of 3–7 amino acids (Fig. 2). These regions have been shown to be important elements for cleavage *in vivo* by signal peptidase (*22*). In particular, the C-region has been shown to harbor important elements of substrate specificity such as the "Ala-X-Ala" motif that is prevalent at the −1 to −3 position with respect to the cleavage site (*20, 21, 23*). As shown in Fig. 2, statistical analyses indicate that signal peptides of gram-positive bacteria are on average longer than those found in gram-negative bacteria (*24–26*). These differences are manifested mainly in longer H- and C-regions for gram-positive bacteria.

Signal peptides can interact directly with the peripheral translocation component SecA (*27*) and the SecY/E complex (*28*). Signal peptides also interact with membrane phospholipids. The positively charged N-region of the signal



	Gram-positive	Gram-negative
Total signal peptide length	32.0 aa	25.1 aa
n-region	positive charge, Lys/Arg rich	positive charge, Lys/Arg rich
h-region	hydrophobic, α-helical, longer	hydrophobic, α -helical, shorter
c-region	polar/neutral, extended β-conformation, longer, & Pro/Thr rich	polar/neutral, extended β-conformation, shorter, & Ser/Ala rich
-1 & -3 residues "Ala-X-Ala"	mainly Ala, less frequent Ser/Gly	mainly Ala, less frequent Ser/Gly

FIG. 2. A summary of the characteristics of bacterial signal peptides.

peptide is believed to function at an early step in the protein export process by interacting with the acidic phospholipid headgroups in the membrane (*29*). On the other hand, the signal peptide H-region likely interacts with the hydrophobic acyl groups of membrane phospholipids. Functional signal peptides have been shown to adopt α-helical conformations in detergent and model membrane systems (*30, 31*).

Early work by the Beckwith and Silhavy laboratories (*32, 33*) demonstrated that signal peptides in bacteria are essential for export of proteins across the inner membrane. Interruption of the H-region with either a charged residue or a deletion led to an export-defective signal peptide. The positively charged N-terminal region also can be important for the efficiency of translocation. Though the positively charged N-region is not as critical for export as the H-region, proteins with signal peptides that contain an acidic N-terminal region are typically exported more slowly. Interestingly, the positive charge(s) in the N-terminal region play a more decisive role when the hydrophobic core is less than optimal (*34*). Defects in the N-region can be compensated for by increased length and/or increased overall hydrophobicity of the corresponding H-region. The C-region contains the

sequence elements important for processing, as determined by site-directed mutagenesis experiments (*35, 36*). The sequence elements are small, uncharged amino acids at the −1 and −3 positions relative to the cleavage site.

In the case of signal peptides involved in the TAT translocation pathway (Fig. 1), the twin arginine motif is a critical determinant for export. In addition to a two-arginine consensus (*37*), the export of proteins by this pathway requires a highly hydrophobic residue at the +2 or +3 position (relative to the two arginines) (*38*).

B. Signal Peptidases in Gram-Negative and Gram-Positive Bacteria

The enzyme now commonly referred to as signal peptidase SPase cleaves signal peptides from preproteins in a wide variety of organisms. The type I SPase from gram-negative *Escherichia coli,* first detected more than two decades ago by Chang *et al.* (*39*), was first purified by Zwizinski and Wickner (*40*). The protease was assayed by its ability to cleave the signal peptide by using *M13* procoat protein as a substrate. The gene encoding *E. coli* type I signal peptidase (lepB) was eventually cloned (*41*) and sequenced (*42*). *Escherichia coli* SPase I has been found to be essential for cell growth (*3*) and its amino acid sequence reveals three apolar stretches: H1 (residues 1–22), H2 (residues 62–77), and H3 (residues 83–91).

The membrane topology of *E. coli* SPase I was determined through several studies. Wolfe *et al.* (*42*) found that treatment of inside-out inner membrane vesicles with trypsin yields a protected *E. coli* SPase I fragment of approximately 32 kDa. The likely cleavage at Lys-57 in the cytoplasmic domain suggests that SPase I spans the membrane *in vivo* from the cytoplasm to the periplasm after this residue. Moreover, Moore and Miura (*43*) found that treatment of right-side-out inner membrane vesicles and spheroplasts with proteinase K or trypsin yields two protected fragments of approximately 80 and 105 amino acids, respectively. Both protected fragments are derived from the amino terminus of the protein and the shorter one is derived from the larger one. Additionally, the same treatment of an amino acid 82–98 deletion mutant (lacking the H3 domain) does not change the size of the smaller protected peptide, but does decrease the size of the larger peptide (*43*). These combined data suggest that the second (H2) apolar domain (residues 62–76) is a membrane-spanning region, whereas the third apolar domain, H3 (residues 82–98) is exposed to the periplasm. Apolar domain 1 also spans the membrane and interacts with the second transmembrane segment. The helix–helix interface was determined by analyzing disulfides formed between pairs of cysteines engineered at the periplasmic ends of the transmembrane regions (*44*). The resulting membrane topology model of *E. coli* SPase I (H3 domain not shown) is shown in Fig. 3.

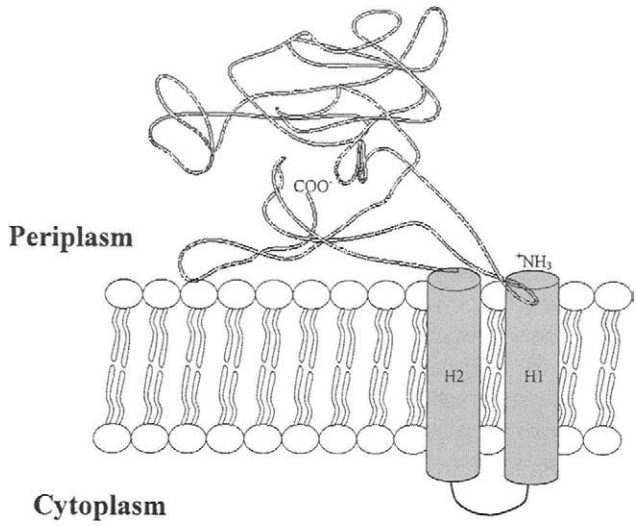


FIG. 3. The membrane topology of *E. coli* type I SPase.

After *E. coli,* the next type I signal peptidase to be cloned was from the gram-negative organism *Salmonella typhimurium* (*45*). *Salmonella typhimurium* type I SPase was found to be homologous to the *E. coli* signal peptidase. As with the *E. coli* enzyme, hydropathy and sequence analyses indicated that the enzyme spans the membrane twice with the catalytic domain in the extracellular medium. From whole genome sequencing projects, homologous type I signal peptidase open reading frames have now been and continue to be identified in a wide variety of gram-negative and gram-positive organisms.

Alignment of the known (archeal, eubacterial, and eukaryal) type I SPase sequences revealed that there were five conserved domains in the signal peptidase family. These domains were designated as A–E (*46*). However, our current alignment of sequences retrieved from the results of an *E. coli* SPase I BLAST homology search (http://www.ncbi.nlm.nih.gov/BLAST/) (*47, 48*) reveals that there is a second C-like homology domain. Similar alignments by Tjalsma and co-workers also reveal this sixth domain and they have designated it as C′ (*49, 50*). Additionally, based on differing conservation patterns and the putative general base residue (Lys or His) found in the D domain, the alignments of all currently known type I SPases have resulted in their sub-classification of either P-type (Lys) or ER-type (His) (*49, 50*). The alignment patterns of the sequences found in our BLAST search are shown in Fig. 4. The A domain (not shown), preceding the B domain, corresponds to a region that is within the transmembrane region of the SPase I enzyme.

The P-type B domain contains the putative catalytic Ser in an SM_PTL motif (Fig. 4A). This catalytic Ser (Ser 90 in the *E. coli* signal peptidase) is invariant in the entire family. The C, C′, and D domains contain the residues reported to be important in substrate binding and catalysis (*51–53*).

The putative general base (K145 of domain D in *E. coli*) found in mitochondrial and chloroplast signal peptidases is invariant in all the gram-positive and gram-negative bacteria surveyed. It is replaced by a His residue in the ER-type SPases. Finally, the E domain contains the strictly conserved GDN motif. Domain E has been reported to be important in active site architecture (*54*). The functions of specific residues in the conserved domains will be further discussed below. It is intriguing that SipW, a characterized *B. subtilis* type I SPase (*49*), and Spc21 of *Clostridium perfringens* are gram-positive SPases classified as ER-type signal peptidases (Fig. 4B).

While most bacteria have only one signal peptidase gene, there are exceptions. *B. subtilis,* for example, has five chromosomally encoded type I signal peptidase genes—SipS, SipT, SipU, SipV, and SipW (*55*)—and two plasmid-encoded signal peptidases (*56*). These chromosomal signal peptidases have overlapping substrate specificities. SipS through SipV resemble bacterial, P-type, signal peptidases. But, as we have mentioned earlier, SipW is more like the eukaryotic, ER-type, signal peptidase; it lacks the conserved Lys general base and, instead, has a His at the homologous position. Similarly, multiple signal peptidase genes are found in *Streptomyces lividans* (*57*). These include SipW, SipX, SipY, and SipZ. Three of these genes are in a single operon. Finally, *Staphylococcus aureus* has two signal peptidase genes, one of which is thought to be inactive as it lacks homologous catalytic Ser and Lys residues (*5*).

C. Signal Peptide Processing Is Required for Cell Growth

Signal peptidase is critical for cell viability. In the Date experiments (*3*), integration of an ampicillin-resistant plasmid bearing a promoter and signal peptidase–deficient sequence into the chromosome of a *pol*A mutant (plasmid replication deficient) *E. coli* strain was attempted. However, this technique did not result in the isolation of any signal peptidase–deficient and ampicillin-resistant strains. This suggested that signal peptidase is essential for *E. coli* viability. In another study, construction of an *E. coli* strain with signal peptidase expression under the control of the araB promoter led to arabinose-dependent cell growth (*4*). Signal peptidase I has also been shown to be essential in human pathogens such as *Staphylococcus aureus* (*5*) and *Streptococcus pneumoniae* (*7*). Finally, though the genes encoding SipS or SipT by themselves are not essential for cell viability, deletion of both SipS and SipT genes prevents cell growth in *B. subtilis* (*49*).

A

	B		C		C'		D		E	
Gram Negative										
Sip_Eco_	IPSGSMMPTLL	96	IGDFILVEKFA	107	RGDIVVFKYP	136	YIKRAVGLPGDKV	155	VPPGQYFMMGDNRDNSADSR	282
Sip_Ctr_	VPTGSMRPTIL	119	EQDRILVSKTT	130	RGELVVFTVG	160	YIKRCMGKPGDTV	193	IPEGHVLVLGDNCPMSADSR	562
Sip_Bbu_	IPSGSMENTLQ	78	IGDFLFVDKFS	89	ESDIIIFENP	118	LVKRGAFADGKIV	169	VPDGYILPIGDNRDNSHDGR	299
Sip_Cpn_	VPTGSMRPTIL	114	EQDRILVSKTT	125	RGGLVVFTVG	155	YIKRCMGRPGDFL	188	VPKGHVLVLGDNYPMSADSR	561
Sip_Hin_	IPSGSMESTLR	121	VGDFLVVNKYA	132	RGDVIVFKAP	161	LIRTGLGATRAAF	177	VPEGQYFVMGDHRDHSDDSR	307
Sip_Hpy_	IPSRSMVGTLY	44	EGDMLFVKKFS	55	RGEVVVFIPP	97	YVKRNFAIGGDEV	116	INHDEFFMIGDNRDNSSDSR	227
Sip_Pfl_	IPSGSMKPTLD	96	VGDFILVNKFS	107	RGDVMVFRYP	136	YIKRVVGLPGDVV	155	VPAGHYFMMGDNRDNSNDSR	235
Sip_Pla_	IPSESMLPTLE	65	VNDRLIVEKIS	76	RGDIIVFHPT	93	FIKRVIGLPGETV	119	VPADSFLVLGDNRNNSYDSH	170
Sip_Rca_	IPSGSMKDTLL	46	IGDFLFVNKMA	57	RGDVVVFRHP	97	FIKRLIGLPGDRI	115	VPEGQYFFMGDNRDNSEDSR	217
Sip_Sty_	IPSGSMMPTLL	97	IGDFILVEKFA	108	RGDIVVFKYP	137	YIKRAVGLPGDKI	156	VPPGQYFMMGDNRDNSADSR	283
Sip_Tpa_	IPSESMVPSFM	93	VGDRLLVFKTA	104	RGDIVVFSNP	133	LVKRIVALPGEKV	187	LPEHNYFMMGDNRLNSTDMR	460
Sip_Tma_	VPTGSMIPTIQ	54	IGDRLFVEKIT	55	IGEIVVFWSP	72	YVKRLVGKGGDVL	115	VPEGFYFLMGDNTKESLDCR	276
SipF_Bja_	IPSGSMKATLL	50	VGDYLFVSKYS	61	RGDIVVFRLP	96	YIKRVIGLPGDRV	115	VPAGHFFMMGDNRDNSTDSR	204
Sip_Bja_	VPSGSMEPTLL	55	TRDALLASKFP	66	QGDVVVFRWP	102	WVKRVVGLPGDRI	121	VPAGHLFVLGDNRDNSADSR	205
Sip_Mle_	IPSESMEPTLH	90	VGDRIMVDKIT	106	PGDVIVFKGP	123	LVKRVIAVGGQTV	172	VPQGRLWVMGDNRIHSADSR	235
Sip_Mtu_	IPSESMEPTLH	102	VGDRINVDKLS	118	PGDVIVFRGP	135	LVKRVIAVGGQTV	184	VPFGRVWVMGDNRTHSADSR	247
Sip_Sli_	IPSGSMEQTIR	91	IGDRVLVDKLT	102	RGDVVVFRDP	120	LIKRVVGVGGDHV	171	VPEGRLWVMGDHRSNSADSR	226
Sip_Sco_	IPSGSMERGLR	83	IGDRVLVNKLA	94	RGDIVVFDGT	112	YIKRVVGVGGDHV	132	VPDGTLFVLGDHRSDSSDSR	188
Sip_Ssp_	IPSSSMEPTLQ	58	INDRLIIEKIS	69	RGEIVVFNPT	86	FIKRIIGLPGDEV	110	VPDDQYLVLGDNRNNSYDSH	161
Spi_Spn_	VEGHSMDPTLA	44	DGEILFVVKH-	54	RFDIVVAHEE	68	IVKRVIGMPGDTI	86	VPEGEYLLLGDDRLVSSDSR	177
Sip_Smi_	VEGHSMDPTLA	19	DGEILFVVKH-	29	RFDIVVAHEE	43	IVKRVIGMPGDTI	61	VPEGEYLLLGDDRLVSSDSR	152
Sip_Bca_	VEGKSMMPTLE	44	SGNLLIVNKLS	55	RFDIIVFHAN	72	YVKRVIGLPGDRI	89	VPPGCIFVLGDNRLSSWDSR	154
SpsB_Sau_	IKGESMDPTLK	42	DGERVAVNIIG	53	KGNVVVFHAN	70	YVKRVIGVPGDKV	87	IPKGKYLVLGDNREVSKDSR	156
SipB_Sca_	VRGDSMYPTLK	42	DGEKVIVNMIG	53	KGNVIVFHAT	70	YVKRVIGMPGDSI	87	IPKNKLLVLGDNREVSKDSR	155
Sip_Bam_	VEGSSMYPTLH	57	DGERLFVNKSV	68	RGDIVIINGD	85	YVKRLIGKPGETV	103	VPKGKYFVMGDNRLNSMDSR	164
Sip2_Bam_	VDGESMEPTLH	51	DRERIFVNMTV	62	RGQIVVLNGE	79	YVKRIIGLPGDTV	95	VPDDKYFVMGDNRRNSMDSR	156
Sip_Bli_	VEGTSMDPTLH	50	DGERLFVYKTV	61	RGDIVIIDGD	78	YVKRLIGLPGDTV	96	VPEGKYFVMGDNRQRSMDSR	157
Sip_Aae_	IPSASMEPTLL	38	VGDFILVNKLV	49	RGDMIVFKYP	66	FIKRIIARGGDTV	85	VPEGYYFVMGDNRDNSQDSR	197
Sip_Bsu_	VEGVSMNPTFQ	40	EGNELLVNKFS	51	RFDIVLFKGP	68	LIKRVIGLPGETI	85	VPKGKYFVVGDNRIYSFDSR	150
SipS_Bsu_	VDGDSMYPTLH	49	NRERVFVNMTV	60	RGDIVVLNGD	77	YVKRIIGLPGDTV	93	VPDNKYFVMGDNRRNSMDSR	155

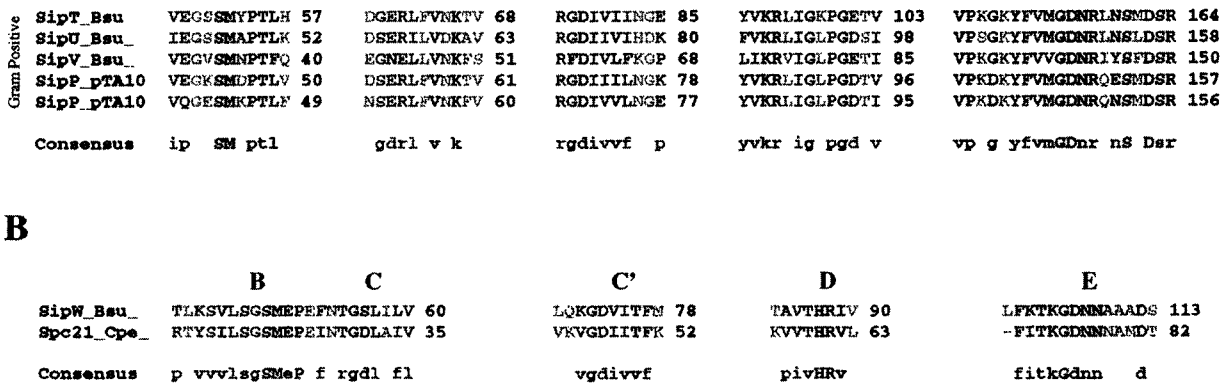


FIG. 4. Amino acid sequence alignment of bacterial SPase I enzymes. This alignment is generated from ClustalX v.1.8 (*103*) amino acid alignments of all P-type (A.) and ER-type (B.) SPases (from a GenBank BLAST search) and are grouped as described by Tjalsma *et al.* (*49*) and Dalbey *et al.* (*46*). The catalytic Ser 90 and Lys 145 residues (*E. coli*) are indicated by a '*'. In the consensus sequence, uppercase indicates strictly conserved residues and lowercase indicates conserved residues appearing in greater than 50% of the aligned sequences. (A) Alignment of conserved residues in P-type SPase enzymes. I. Eco, *Escherichia coli;* Sty, *Salmonella typhimurium;* Hin, *Haemophilus influenzae;* Pfl, *Pseudomonas fluorescens;* Bja, *Bradyrhizobium japonicum;* Rca, *Rhodobacter capsulatus;* Mle, *Mycobacterium leprae;* Mtu, *Mycobacterium tuberculosis;* Sli, *Streptomyces lividans;* Sco, *Streptomyces coelicolor;* Pla, cyanobacterium *Phormidium laminosum;* Ssp, *Synechocystis* sp. strain PCC6803 II; Aae, *Aquifex aeolicus;* Tpa, *Treponema pallidum;* Hpy, *Helicobacter pylori* J99; Bbu, *Borrelia burgdorferi;* Spn, *Streptococcus pneumoniae;* Smi, *Streptococcus mitis;* Bca, *Bacillus caldolyticus;* Sau, *Staphylococcus aureus;* Sca, *Staphylococcus carnosus;* Bsu, *Bacillus subtilis;* Bam, *Bacillus amyloliquefaciens;* Bli, *Bacillus licheniformis;* Tma, *Thermotoga maritima;* Cpn, *Chlamydophila pneumoniae;* Ctr, *Chlamydia trachomatis;* Sip, SPase; Spi, *Streptococcus pneumoniae* SPase I; Sps, SPase *Staphylococcus aureus.* (B) Alignment of conserved residues in Eubacterial ER-type SPase I. Bsu, *Bacillus subtilis;* Cpe, *Clostridium perfringens;* Sip, signal peptidase; Spc, signal peptidase complex.

In bacteria, the apparent function of the signal peptidase enzyme is to release translocated preproteins from their membrane attached signal sequences. This allows periplasmic or outer membrane destined proteins to enter the periplasmic space. Using an *E. coli* strain in which synthesis of signal peptidase is under the control of the arabinose operon, it was shown that precursor proteins of *M13* procoat, maltose binding protein (MBP), and outer membrane protein A (OmpA), accumulate in the cell when the synthesis of signal peptidase is repressed (*4*). Protease accessibility assays demonstrated that, with attenuated signal peptidase production, procoat, pre-MBP and pro-OmpA proteins were still found translocated across the plasma membrane. This result is in line with studies showing that noncleavable signal sequence mutants of β-lactamase (*58*) and *M13* procoat (*59*) were also translocated. Interestingly, pre-MBP and pro-OmpA were found remaining on the outer surface of the plasma membrane, suggesting that the proteins were tethered to the membrane by the membrane embedded uncleaved signal sequences (*4*).

It is fascinating that *Mycoplasma genitalium,* a bacteria with the smallest genome known, does not appear to have a gene with any significant homology to type I signal peptidase (*60*). This is the only bacterium known to lack a signal peptidase.

D. Type I Signal Peptidase Is an Antibacterial Target

Signal peptidase is a potential target for antibacterial compounds and is currently being actively investigated by pharmaceutical companies (*61*). What makes type I signal peptidase an attractive drug target is that it is essential for cell viability for all bacteria. Also, the SPase I enzyme should be readily accessible by small molecules since the protease active site is located in the periplasmic space of gram-negative and in the outside surface of the plasma membrane in gram-positive bacteria. The recently solved 3D crystal structure of the catalytic domain of *E. coli* signal peptidase should provide a useful model for the rational design of potential inhibitors (*52*).

The practicality of using signal peptidase as a drug target has been questioned because SPase I is also an enzyme found in eukaryotic cells. However, there are notable differences between the bacterial versus the eukaryotic paralogs. The prokaryotic signal peptidases are single polypeptide chains, whereas the endoplasmic reticulum signal peptidases are multimeric complexes and contain some nonhomologous polypeptides (*46*). Also, the signal peptidase complex in *Saccharomyces cerevisiae* most likely carries out catalysis using a Ser-His dyad, rather than the Ser-Lys dyad found in prokaryotic signal peptidases (*53*). In mitochondria, signal peptidase is postulated to use a Ser-Lys mechanism, but it is a dimer consisting of the Imp1/Imp2 polypeptides. These differences suggest that it may be possible to design

inhibitors that specifically inhibit the bacterial signal peptidases only and not the ER or mitochondrial enzymes.

II. Type I Signal Peptidase Enzymology

A. Enzyme Purification and Substrate Assays

The early isolation work on *E. coli* type I signal peptidase resulted in the characterization of some of its physical properties. Cell extracts overproducing signal peptidase indicate that signal peptidase activity is sensitive to high salt, Mg^{2+} concentration, and pH (*62*). Using purified substrate and purified *E. coli* signal peptidase, a pH optimum of about 9.0 has the greatest level of activity (*63, 64*). Also, a profile of activity versus temperature indicates that the enzyme is stable up to 40°C (*63*).

Overexpression and purification of *E. coli* type I SPase has been accomplished by recombinant techniques. Typically *E. coli* strains bearing plasmids engineered to overexpress the *E. coli lepB* gene are used. One such strain takes advantage of a plasmid (pPS9) bearing the signal peptidase gene under transcriptional control of the *lambda* promoter (*65*). This plasmid also codes for a temperature-sensitive *lambda* repressor causing reduced expression at 30°C and rapid expression at 42°C. Another plasmid, pRD8, expresses the *lepB* gene under the control of the *araB* promoter (*4*). In this system, expression is induced by the addition of arabinose. With both these plasmids the protein is purified by a protocol involving membrane isolation, Triton X-100 detergent extraction, DEAE ion-exchange chromatography (Pharmacia), and final isolation to homogeneity by a polybuffer chromatofocusing technique (Pharmacia) (*65, 66*).

A difficulty with *E. coli* SPase I purification results from an apparent autoproteolysis of the enzyme. Talarico *et al.* (*67*) demonstrated that the purified SPase gets cleaved after an Ala-Gln-Ala sequence (residues 38–40), which is consistent with the "−3, −1" or "Ala-X-Ala" motif of SPase I signal peptide substrate specificity (see Section II,B). With this information, a more efficient scheme for *E. coli* SPase I purification was devised (using pRD8) by inserting a 6-His tag after amino acid residue 35 (cytosolic domain) in the protein sequence. This eliminates yield losses from self-cleavage and enables the use of a nickel-chelate affinity chromatography purification system (*64*). The most productive system, however, utilizes the plasmid pET23Lep (*54*). This method also uses a 6-His/nickel-chelate approach, but also takes advantage of the very high expression levels of the pET vector system (Novagen). With this system, milligram quantities of purified *E. coli* SPase I protein are generated from a few liters of culture in a relatively short amount of time.

The enzymatic activity of the *E. coli* SPase has been assayed with a number of different substrates. These include peptides (*68–71*) and preprotein

substrates (*40, 64, 72, 73*). One of the peptide substrates, Phe-Ser-Ala-Ser-Ala-Leu-Ala-Lys-Ile, is based on the cleavage site region of pre-MBP. It is cleaved by SPase to generate the Phe-Ser-Ala-Ser-Ala-Leu-Ala and Lys-Ile fragments. HPLC is used to separate and quantitate the two products. The resulting k_{cat} and K_m values with this substrate are 114 hr^{-1} and 0.8 m*M*, respectively (*71*). On the other hand, using the preprotein pro-OmpA nuclease A as a substrate for SPase results in much better kinetic constants. In this assay, cleavage of pro-OmpA nuclease A is analyzed by resolving the preprotein from the mature protein with SDS–PAGE. This substrate results in a k_{cat} of 44 s^{-1}, a K_m of 19.2 μM, and k_{cat}/K_m of 2.3×10^6 s^{-1} M^{-1} at pH 8.0. This catalytic efficiency is comparable to that of other Ser proteases such as trypsin and chymotrypsin (*74*). The reason for the dramatic increase in k_{cat}/K_m for the preprotein substrate compared to the peptide substrate is that the preprotein/SPase interactions not available with synthetic peptides may lead to optimal substrate positioning and increased processing efficiency.

Fluorogenic substrates have also been developed as continuous assays of SPase activity (*68, 69*). One example is the internally quenched fluorescent substrate Tyr(NO2)-Phe-Ser-Ala-Ser-Ala-Leu-Ala-Lys-Ile-Lys(Abz) (*68*). This conjugate peptide is also based on the cleavage site region of the preprotein pre-MBP, and *E. coli* SPase cleavage is able to generate the expected products Tyr(NO2)-Phe-Ser-Ala-Ser-Ala-Leu-Ala and Lys-Ile-Lys(Abz). This results in a fluorescence increase that is monitored during the course of the reaction. Unfortunately, like the Dev peptide (*71*), the resulting k_{cat}/K_m for this substrate is also very low (71.1 M^{-1} s^{-1}), indicating it is also a poor substrate. A hydrophobic H-region is a common motif in signal peptides (Fig. 2). The poor catalytic efficiency for this substrate is most likely due to the lack of a hydrophobic core in the primary sequence of the peptide substrate itself (*68*).

With the development of a new fluorogenic substrate, Stein and co-workers (*75*) have addressed some aspects of the function of the H-region in the signal peptide. In this work, the insertion of 10 Leu residues into the N terminus of the peptide used by Zhong and Benkovic (*68*) results in a substrate displaying a dramatic 10^4 increase in k_{cat}/K_m. Stein and co-workers suggest that this increase most likely results not only from the proximity effects gained from anchoring the substrate to micelles (also containing micelle-anchored SPase), but also from specific interactions achieved between the SPase enzyme and the new "H-region-like" domain of the substrate signal peptide itself.

B. Substrate Specificity

Statistical analyses of preprotein sequences from gram-negative and gram-positive bacteria have been very useful in the determination of the

conservation patterns found in the C-region of the signal peptide. The data indicates patterns that are obligatory for signal peptide processing (*24*) and has led to the so called "−3, −1" or "Ala-X-Ala" rule (*20, 21, 23*), which states that mainly small uncharged residues are found at the −3 and −1 positions relative to the cleavage site. In both gram-negative and gram-positive bacteria, Ala is almost exclusively located at the −1 residue (Gly and Ser are the next most frequent). Ala is also the most common residue found at the −3 position followed by Val and Ser (less frequent). Also, Ala is common at +1 position, while Asp and Glu residues are found in the first few positions of the mature region of prokaryotic secretory proteins.

Using *in vivo* assays, the determinants of substrate specificity have been examined extensively for several preprotein substrates. The results of the studies on *M13* procoat, pre-phoA (pre-phosphatase A), and pre-MBP substrates are summarized in Fig. 5. Site-directed mutagenesis was used to substitute various residues at the +1, −1, −2, −3, −4, −5, and −6 positions of the *M13* procoat protein (*35*). The critical positions in the signal peptide for substrate processing are at the −1, −3, and −6 positions. Processing of procoat only occurs with small residues at the −1 position (Ala, Gly, Ser, and Pro). Some small, uncharged (Ser, Gly, Thr), or aliphatic (Leu and Val) residues at the −3 position result in processing, but others such as Pro, Gln, Lys, or Arg result in no processing. The results also indicate no distinct requirements for *in vivo* processing for residues at the +1, −2, −4, and −5 positions. Almost any residue is tolerated except for a Pro at the +1 position.

As shown in Fig. 5, similar findings were observed in studying the *in vivo* processing of pre-phoA (*36*) and pre-MBP (*76*). As suggested by statistical analyses, the critical positions for processing of these substrates is at the −1 and −3 positions. *In vivo* processing is maintained (*36*) with almost any residue at the +1, −2, −4, −5 positions of pre-phoA.

In addition to the −1 and −3 residue requirements, the presence of a helix breaker or a beta turn residue in the −4 to −6 region has been shown to be important for SPase processing. A Pro and Gly residue is frequently present in this region of bacterial signal peptides (*26*). It is intriguing that almost any residue besides a Pro at the −6 position of the *M13* procoat blocks *in vivo* processing (*35*). Perhaps a helix breaker prevents the C-region from forming a long helix with the hydrophobic core region of the signal peptide and allows the signal peptide C-region to bind to SPase in an extended conformation.

Jain *et al.* (*77*) analyzed the SPase cleavage of a number of phoA signal sequence mutants differing only in the length of the C-region. C-region lengths varied from 3 to 13 residues, and it was found that lengths ranging from 3 to 9 residues are completely and efficiently cleaved whereas those of 11 to 13 residues are not. One interpretation of this data is that since the active site

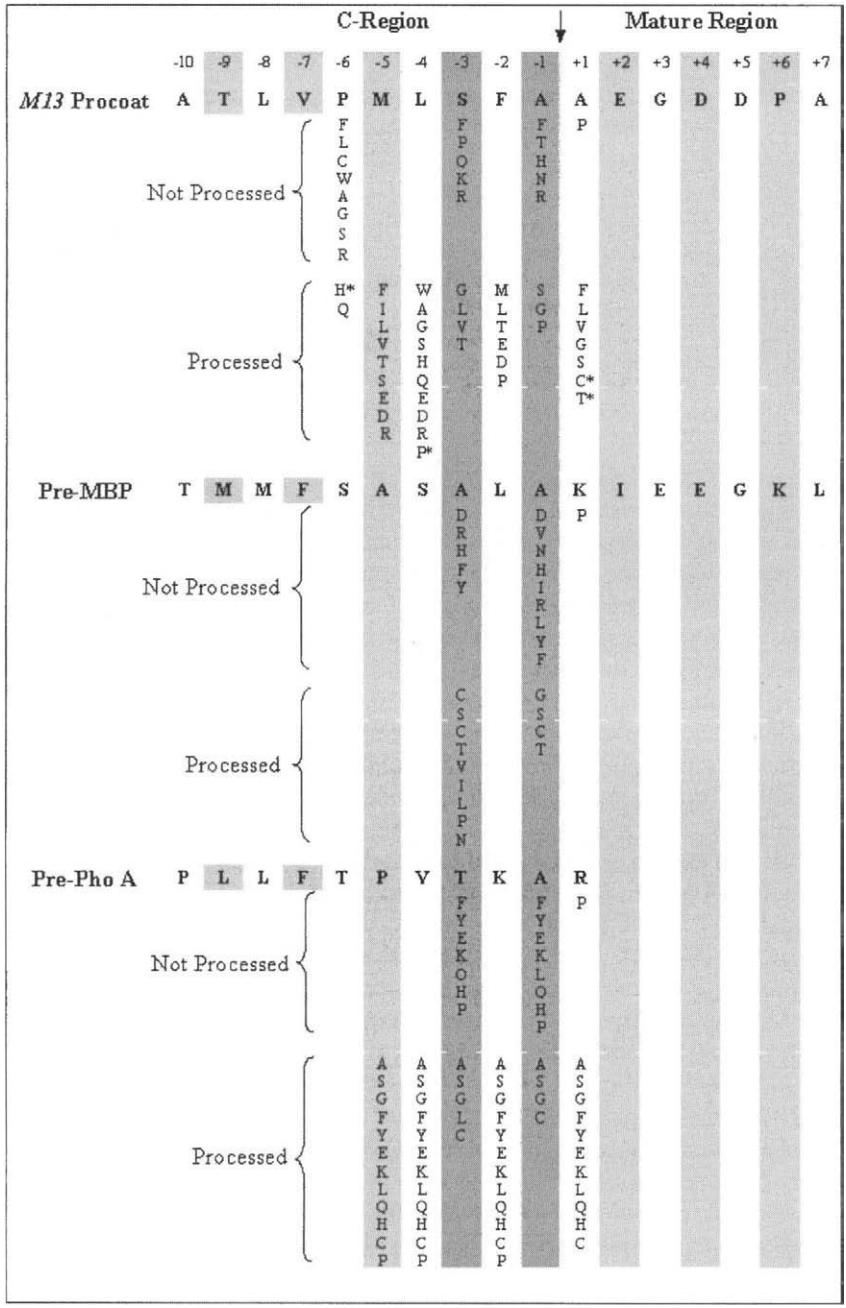


FIG. 5. Signal peptide C-region point mutations affect SPase I processing. The tabulated results of *in vivo* processing by *E. coli* SPase I of C-region signal sequence point mutants of M13 procoat (*35*), pre-MBP (*76*), and pre-Phosphatase A (*36*) substrates are shown. The bold sequences correspond to the wild-type sequences. * indicates processing is <16%.

of *E. coli* SPase is close to the periplasmic surface of the inner membrane, when the C-region of the signal peptide exceeds nine residues, the cleavage site of the preprotein is presented too far away from the active site of the enzyme. This results in the sudden drop-off in activity seen with the insertion of nine or more residues [*77*].

After von Heijne and co-workers identified the −3, −1 ("Ala-X-Ala") substrate specificity requirement for type I signal peptidases, a computational method was developed to predict whether signal peptides are located within biological sequences using a weight matrix method (*26*). This method has been improved using the neural network and hidden Markov model-based prediction model of Nielsen *et al.* (*24, 25, 78*). Biosequence analysis using this program is now available on the SignalP server (http://www.cbs.dtu.dk/services/SignalP/) to identify signal peptide cleavage sites. The algorithm even discriminates between cleaved signal peptides and uncleaved signal anchors in both prokaryotic and eukaryotic models. Also, the effectiveness of prediction programs such as these have enabled the *de novo* design of artificial signal peptides with demonstrated biological activity. In both the studies of Nilsson and von Heijne (*79*) and Wrede *et al.* (*80*), computer engineered signal peptides, located N-terminal to fusion protein constructs, were shown to be translocated and processed effectively in *E. coli*.

C. The Active Site and Catalytic Mechanism: Site-Directed Mutagenesis Studies

Type I SPase has the unusual property of being resistant to inhibitors of the classical Ser, Cys, Asp, and metallo-protease classes (*62, 69, 81*). Thus, there is great interest in the protease community to pinpoint its catalytic mechanism. To date, most of the work in this area has been on the *E. coli* enzyme but there has also been some work on the *B. subtilis* type I signal peptidase (SipS). Initial clues to the proteolytic mechanism of SPase were determined using site-directed mutagenesis of the *E. coli* enzyme. SPase maintains activity if each of the Cys and His residues are mutated (*81, 82*), demonstrating that neither of these residues is catalytically important. Substitution of Ser-90 with an Ala completely inactivates the enzyme (*82*) and Lys-145 is also important for activity (*83, 84*). Mutation of Lys-145 to His, Asn, or Ala results in an inactive protease. These data show that *E. coli* SPase has a critical Ser-90 and Lys-145 residue, and support the notion that these are the catalytic residues. Consistent with this is that these residues are conserved in all bacterial type I signal peptidases.

Complementing the loss of function mutagenesis studies, the catalytic roles of Ser-90 and Lys-145 in *E. coli* type I SPase were further substantiated by chemical modification studies. Replacement of Ser-90 with a Cys residue

produces an active enzyme that can then be inactivated by the addition of a Cys-specific reagent (*84*). In addition, an inactive Cys-145 *E. coli* SPase mutant can regain activity by modification with bromoethylamine to generate an enzyme with a γ-thia-Lys (*64*). Enzyme activity is also restored, to a lesser extent, by modification of the thio-145 SPase with either bromopropylamine or 2-mercaptoethylamine to generate other Lys analogs. There is no recovery of activity when the Cys-145 mutant derivative is reacted with (2-bromoethyl)trimethylammonium-Br. This finding supports the role of Lys as a general base rather than a positive charge donor in the mechanism.

Guided by amino acid conservation patterns among bacterial, ER, and mitochondrial type I signal peptidases, site-directed mutagenesis studies of the *Bacillus subtilis* type I SPase enzyme (SipS) indicated similar critical roles for some of the homologous residues found to be critical for function in *E. coli* (*85*). *B. subtilis* Ser-43 and Lys-83, homologous to the *E. coli* Ser-90 and Lys-145 residues, are critical for *in vivo* enzymatic activity. In addition, the amino acid homologous to *E. coli* Asp-280, Asp-153, is also essential. Two other residues, Asp-146 and Arg-84 (*E. coli* Asp-273 and Arg-146, respectively), appear to be structural determinants for the *B. subtilis* SPase (*85*).

The crystal structure of the inhibitor-bound, truncated, soluble form of *E. coli* type I SPase (Δ2–75) (*52*) has contributed much to the SPase field. Using this structure as a guide, the work of Klenotic *et al.* (*54*) has shed light onto the roles of most of the conserved residues in the homology domain E-region (see Fig. 4A) of *E. coli* SPase. Most of these residues are in the active site region, as shown in Fig. 6 (see color plate). In contrast to the *B. subtilis* experiments, mutagenesis of *E. coli* Arg-146 does not result in a dramatic loss of function. The crystal structure shows an ionic interaction between Asp-273 and Arg-146, but the mutation of Arg-146 to Ala in *E. coli* results in no reduction in enzymatic activity (*54*). However, like the *B. subtilis* counterpart, there is a marked reduction in activity for Asp-273 mutations. The salt bridge interaction of Asp-280 with Arg-282, also evident in Fig. 6, supports the loss of function resulting from mutagenesis of these residues.

Other conserved *E. coli* Box E residues are Gly-272 and Ser-278 (Fig. 4A). The active site structure (Fig. 6) of the *E. coli* SPase reveals that both Gly-272 and Ser-278 are in close proximity to the catalytic Lys-145. In fact, Ser-278 is within H-bonding distance to Lys-145 (*54*). Mutagenesis studies demonstrate that indeed the Ser-278 is important for activity, as changing this residue to an Ala causes a reduction in activity of approximately 300-fold in processing the substrate pro-OmpA nuclease A (*54*). This suggests that Ser-278 may actually help orient the proposed general base Lys-145. Changing the Gly-272 to Ala reduces the activity of SPase 750-fold relative to the wild-type enzyme (*54*). Consistent with a Gly-272 to Ala mutant with reduced activity, modeling studies suggest that changing the side chain at amino acid 272 from a hydrogen to a methyl, or any other group, causes steric crowding and

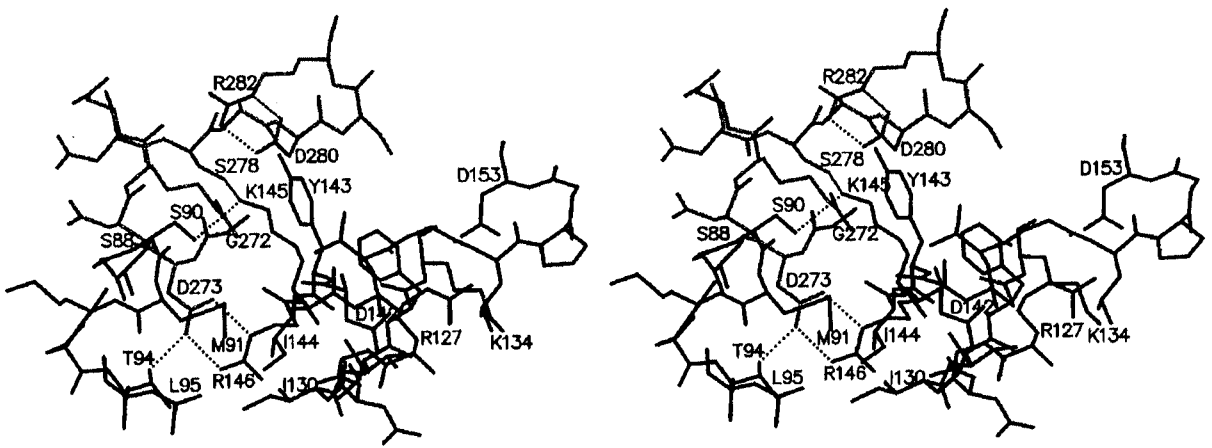


FIG. 6. (See color plate.) The residues conserved among type I SPases (boxes B-E) make up the active site region of *E. coli* SPase I. A ball and stick stereographic representation of the residues that make up box B (88–95, black), box C (127–134, red), box D (142–153, blue), and box E (272–282, green). This figure was made with the program MOLSCRIPT (*104*).

perturbs the positioning of the Lys-145 side chain. In total, the conserved Box E residues may help stabilize the enzyme and are responsible for maintaining the overall architecture of the active site.

In other recent work, the Ser-88 residue has also been shown to be important in the catalytic mechanism of *E. coli* SPase (*86*). The crystal structure of Δ2–75 SPase I reported by Paetzel *et al.* (*52*) indicates that the catalytic Ser-90 amide backbone nitrogen and the Ser-88 hydroxyl may be involved in forming an oxyanion hole, stabilizing a tetrahedral oxyanion transition state intermediate that forms during the course of catalysis. Mutagenesis of Ser-88 to Ala leads to a greater than 2000-fold reduction in the k_{cat} with very little effect on K_m (*86*). Interestingly, sequence alignment studies show that only Ser, Thr, and Gly residues occur at this position in other signal peptidases (Box B of Fig. 4). In signal peptidases surveyed, the Gly residues were present at this homologous position only in gram-positive bacteria (Box B of Fig. 4). It is possible that with this subset of gram-positive bacterial signal peptidases, oxyanion stabilization is mediated by a more conventional backbone amide rather than a side-chain interaction.

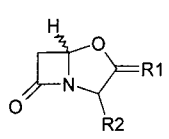
From an evolutionary perspective, it is very intriguing that ER- and Archaea-like signal peptidases have been identified in gram-positive bacteria such as *Bacillus subtilis* and *Clostridium perfringens* (Fig. 4B). SipW of *B. subtilis* has been characterized as one of seven signal peptidases found in *B. subtilis* (*49*) while Spc-21 of *C. perfringens* is a putative SPase identified from genome sequencing (accession CAA60213). The overall amino acid conservation patterns and a putative His general base in place of the Lys general base (Lys-145 of *E. coli*) differentiates signal peptidases into the P-type (eubacterial and mitochondria/chloroplast) and ER-type [eukaryal (ER) and archeal] signal peptidases (*49*). Mutagenesis studies show that the conserved Ser and His residues (at the same positions as Ser-90 and Lys-145 in *E. coli*) are critical for the functioning of SipW (*50*). In contrast to similar studies done with the Sec11 homologous subunit in the ER SPase of *S. cerevisiae* where the His cannot be substituted with a lysine residue (*53*), this work on SipW showed that the putative His general base can be substituted by a Lys and still maintain enzyme function.

D. Inhibitors of Type I Signal Peptidase

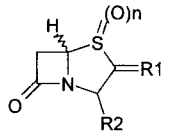
It has been very challenging for chemists to synthesize effective inhibitors against bacterial type I signal peptidases because of their unusual mechanism. As mentioned previously, protease inhibitors against the Ser-, Cys-, Asp-, and metalloenzyme classes were ineffective against SPase. Though SPase is a Ser protease, very high concentrations of [^{3}H]diisopropyl fluorophosphate do not inhibit the enzyme (*63*). The first report of an inhibitor

was by Kuo and co-workers (*69*), where they showed that certain β-lactams could inhibit the enzyme. β-Lactams had been shown previously to inhibit other Ser proteases and β-lactamases (*87–91*).

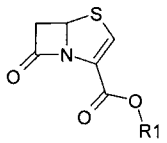
The observation that the catalytic mechanism of SPase occurs by a Ser-Lys dyad (*83, 84*) is a significant breakthrough. This mechanism is similar to β-lactamase enzymes, which use Lys as a general base in the acylation step (*92*). With this information, researchers at Smithkline Beecham Pharmaceuticals focused on β-lactam type compounds. Several types of effective compounds were identified with an IC_{50} in the 0.260 to 50 μM region (*61*). As shown in Fig. 7, the best inhibitors found include clavams, thioclavams,



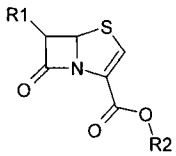
Clavams



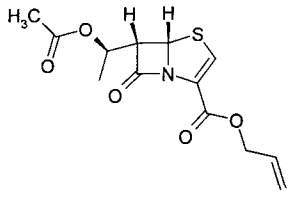
Thioclavams



Penem Carboxylates



Penem Carboxylate
C6 Substituted Esters



Allyl (5S,6S)-6-[(R)-acetoxyethyl]-penem-3-carboxylate

FIG. 7. Inhibitors of *E. coli* SPase I. A listing of some of the types of inhibitors that have been designed to inhibit *E. coli* SPase I (*61*). The most effective inhibitor, synthesized by researchers at SmithKline Beecham (*61*), is the C6 substituted penem carboxylate ester, allyl (5*S*, 6*S*)-6-[(*R*)-acetoxyethyl]penem-3-carboxylate as shown.

and penem carboxylates. The 5*S*-penem derivatives are the most potent (*61*). The compound allyl(5*S*,6*S*)-6-[(*R*)-acetoxyethyl]penem-3-carboxylate (Fig. 7) is the most potent inhibitor developed to date and has been shown to inhibit *E. coli* and *S. aureus* as well as chloroplast signal peptidases. From a mechanistic point of view, it is interesting that the penem inhibitors of signal peptidases are of the 5*S* stereochemistry. This is the opposite to that of the 5*R* β-lactams that are recognized by β-lactamases and penicillin binding proteins (*61*).

Besides small molecule inhibitors, *E. coli* type I SPase is also inhibited by the signal peptide of the *M13* procoat protein (*72*). Also, the substitution of a Pro residue into the +1 position of pre-MBP prevents its processing by *E. coli* SPase (*93*). Expression of this pre-MBP mutant *in vivo* leads to the accumulation of preproteins normally processed by type I SPase but not proteins processed by lipoprotein (type II) SPase (*93*). This suggests that the pre-MBP +1 Pro mutant acts as a competitive inhibitor of type I SPase.

III. Three-Dimensional Structure

A. A Novel Protein Fold

The SPase crystal structure of the soluble, catalytic domain fragment of *E. coli* SPase I, Δ2–75 (*52*), reveals a unique protein fold (see Fig. 8). It consists mainly of two large antiparallel β-sheet domains (I and II), two small 3_{10} helices (residues 246–250 and 315–319), and one small α-helical region (residues 280–285). There is one disulfide bond between Cys-170 and Cys-176 located immediately before a beta turn between the outer strands of β-sheet domain II. An extended β ribbon protrudes from domain I. In conjunction with the N-terminal strand, this ribbon gives the overall molecule a conical shape with dimensions of 60 Å × 40 Å × 70 Å.

Another protease that has been proposed to use a Ser-Lys dyad is UmuD, a member of the LexA family of proteases (*94*). UmuD maintains 23.4% sequence identity (residues 40–139) with *E. coli* type I SPase (75–202). The crystal structure of the fragment of UmuD, UmuD′, reveals a fold similar to SPase and is mostly β sheet (*94, 95*). In UmuD′, however, there are no structural counterparts to the β-sheet domain II and the extended hairpin (between residues 108 and 124) found in SPase. In fact, sequence homology and modeling studies of other signal peptidases in gram-positive bacteria, mitochondria, and endoplasmic reticulum indicate that the extended hairpin is also missing and that most of β-sheet domain II is missing. Whether these differences manifest themselves through variations in substrate binding or specificity is yet to be determined.

FIG. 8. General fold of *E. coli* Δ2–75 type I SPase. A MOLSCRIPT ribbon diagram of *E. coli* Δ2–75 type I SPase. The domain that appears [from sequence alignments (*105*)] to be conserved across all type I SPases is shown in black.

B. VIEW OF A UNIQUE ACTIVE SITE AND CATALYTIC MECHANISM

A GRASP molecular surface representation of the Δ2–75 crystal structure is shown in Fig. 9. The substrate binding pockets S1 and S3 are labeled. The dark gray areas in Fig. 9 represent the exposed hydrophobic surfaces. The large exposed surface is formed by antiparallel β strands consisting of residues 81–85, 99–105, 292–307, and 321–314, and includes the hydrophobic residues Tyr-81, Phe-100, Leu-102, Trp-300, Met-301, Trp-310, and Leu-314 within the β strand, and the nearby residues Phe-79, Ile-80, Leu-316, and Ile-319. Studies have shown that Δ2–75 can bind to the inner membrane vesicles of *E. coli* and insert into membrane monolayers (*96*). The insertion of the catalytic domain into the lipid phase suggests that the active site may be partially buried in the membrane. Thus the extended hydrophobic patch of Δ2–75 seen in the crystal structure may constitute the membrane association surface (*52*).

The crystal structure also revealed a covalent bond from the active site Ser-90 Oγ to the carbonyl carbon, C7 of the 5*S*,6*S*-penem inhibitor (Fig. 10). This is the first direct evidence of the nucleophilic nature of the catalytic Ser-90. In addition, the Ser-90 Oγ oxygen is within 2.9 Å of the Nε of Lys-145.

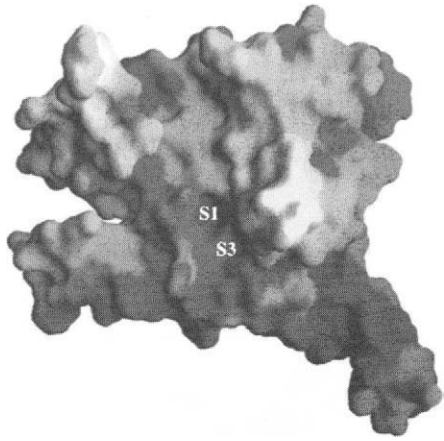


FIG. 9. A representation of the molecular surface of *E. coli* Δ2–75 type I SPase made with the program GRASP. The dark gray areas indicate hydrophobic surfaces. The location of the S1 and S3 substrate binding sites are indicated.

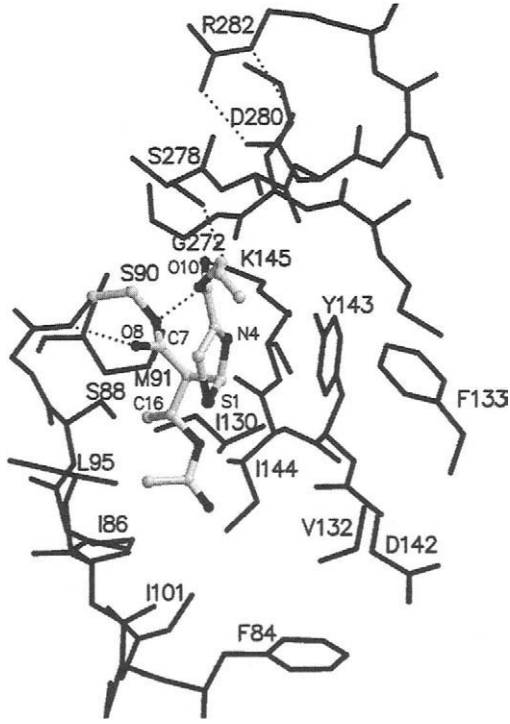


FIG. 10. A MOLSCRIPT (*104*) ball-and-stick representation of the active site of *E. coli* Δ2–75 type I SPase that is bound to inhibitor.

Lys-145 is the only titratable amino acid residue within H-bonding distance of Ser-90 (Fig. 10). This is further evidence that it may act as a general base during catalysis. Also, the main chain nitrogen of Ser-90 forms a hydrogen bond (2.9 Å) to the carbonyl oxygen of the cleaved β-lactam ring within the penem inhibitor. This suggests that the Ser-90 N stabilizes the tetrahedral transition state oxyanion intermediate. The Ser-88 side-chain hydroxyl is also a likely candidate for oxyanion stabilization provided it is able to rotate about its χ_1 angle. Because of steric conflicts, the covalently bound inhibitor prevents the Ser-88 hydroxyl from being in position to contribute to the oxyanion stabilization. As described previously, the mutational studies of Carlos *et al.* (*86*) suggest that, as with the classical serine proteases, an oxyanion hole is operational in *E. coli* SPase.

The crystal structure shows that *E. coli* SPase I Lys-145 is buried and makes van der Waals contacts with the side chains of Phe-133, Tyr-143, and Met-270, as well as the main-chain atoms of Met-270, Met-271, Ala-279, and Gly-272. This hydrophobic environment may altogether be responsible for lowering of the Lys-145 pK_a. The results of studies on temperature-sensitive mutants of *B. subtilis* SPase I (SipS) are consistent with these observations of the *E. coli* SPase I structure. The mutations of *B. subtilis* SPase I (SipS) Tyr-81 (*E. coli* Tyr-143) to an Ala or Leu-74 to Ala (*E. coli* Phe-133) result in reduced SipS SPase I activity at 37°C and almost no activity at 42°C (*51*). These results are consistent with the hypothesis that the side chains of these residues reduce the pK_a of the general base Lys that is important for catalysis.

Also with *B. subtilis* SipS, the homologs of *E. coli* Arg-146 and Asp-273 (*B. subtilis* Arg-84 and Asp-146) are important for activity (*51*). SipS R84A and SipS D146A mutants display little activity at 37°C but sufficient to replace the chromosomally encoded enzyme (*49*). However, these SPase I mutants are temperature-sensitive for growth with evidence that they are prone to proteolytic degradation *in vivo*. In the crystal structure of *E. coli* SPase I, Arg-146 and Asp-273 form an ionic salt bridge interaction and the results from *B. subtilis* SipS indicate an analogous interaction is essential for optimal activity. In contrast, mutation of the Arg-146 in the *E. coli* full-length SPase I does not impair activity at all, suggestive of a much less catalytically important interaction in the *E. coli* enzyme (*54*).

The active site geometry from the crystal structure also reveals that Lys-145 is hydrogen bonded to Ser-278 which in turn is also hydrogen bonded to Asp-280 (see Fig. 6). Ser-278 is also held firmly in place by an interaction with the main chain amide of Gly-272. Ser-278 may help orient the Lys-145 residue similar to the manner in which the Asp residue functions to orient the His residue in the classical Ser/His/Asp catalytic triads of serine proteases. Asp-280 may also help orient Ser-278 in addition to playing a structural role by forming a salt bridge with Arg-282. Asp-273 maintains a bifurcated interaction by forming a salt bridge with Arg-146 and a hydrogen

bond with Thr-94 (Fig. 6). In total, these residues may help stabilize the enzyme and are perhaps responsible for maintaining the overall architecture of the active site (*54*).

Finally, it is noteworthy that Ser-278 is invariant in not just bacterial but all P-type SPase I enzymes [bacterial (see Fig. 4), mitochondrial (Imp1/Imp2), and chloroplast type I signal peptidases (thylakoid processing peptidases)] that are proposed to utilize a Ser-Lys dyad catalytic mechanism. The ER-type (endoplasmic reticulum and archaeal) SPase I enzymes, proposed to use a Ser-His dyad, instead do not have a homologous *E. coli* Ser-278 residue at this position. This implies that the Ser-278–Lys-145 interaction mentioned earlier is critical for the functioning of the Ser-90–Lys-145 dyad mechanism.

C. A Binding Site Consistent with General Substrate Specificity

The structure of the Δ2–75 SPase-penem inhibitor complex suggests an S1 pocket [Schechter and Berger notation (*97*)] that binds the P1 residue of a preprotein substrate (*52*). This hypothesis is based on the observation that the methyl group (C16) on the 6-[acetoxethyl] side chain of the penem inhibitor is critical for the effectiveness of the inhibitor and presumably mimics the P1 (−1 relative to the cleavage site) Ala side chain of a preprotein substrate (*61*). The residues making direct van der Waals contact with the P1 methyl group of the inhibitor in the crystal structure are Met-91, Ile-144, Leu-95, and Ile-86, which are all conserved residues (Fig. 10).

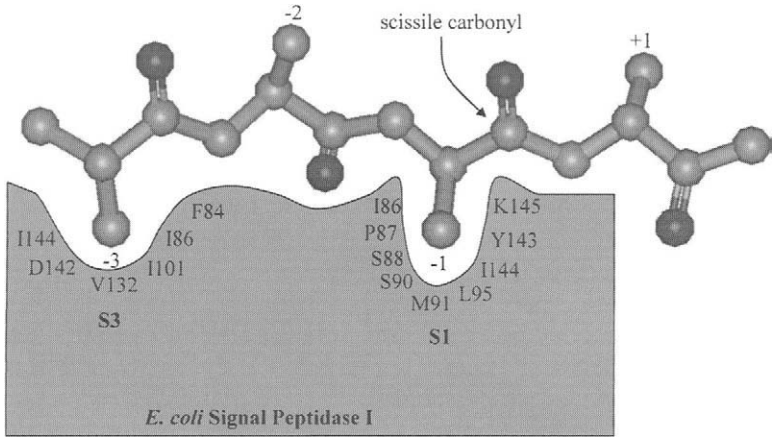


FIG. 11. *E. coli* SPase I substrate binding. A schematic representation of the S1 and S3 subsite interactions of *E. coli* SPase I with the −1 (P1) and −3 (P3) residues of a modeled tetraalanine peptide substrate. The substrate is shown in an extended (β) conformation.

An S3 subsite is also extrapolated by modeling an extended tetra-Ala peptide into the active site of SPase. This model uses the methyl group of the inhibitor and the carbonyl group of the cleaved inhibitor as a template. Hydrogen-bonding interactions between the peptide and the SPase I β strand containing the Lys145 are obtained with this model. The residues that form the S3 site are Phe-84, Ile-144, Val-132, and Ile-86. A schematic diagram of the S1 and S3 interactions with the modeled tetra-Ala peptide is shown in Fig. 11. As was revealed from surface analysis (Fig. 9), S3 is shallower and broader than the S1 hydrophobic depression. This result is consistent with computational analyses of naturally occurring bacterial signal peptide substrates indicating that, although Ala is most frequently found at both -1 (P1) and -3 (P3) positions, larger aliphatic residues are sometimes observed at the -3 position (*24, 25*).

IV. Other Ser-Lys Dyad Proteases and Amidases

LexA, a protein involved in the SOS response in DNA repair (*98*), was the first protease suggested to carry out catalysis using a Ser-Lys dyad (*99*). UmuD is another member of the LexA family and also utilizes a Ser-Lys dyad catalytic mechanism. In the crystal structure of the UmuD′ fragment, the homologous critical Ser and Lys SPase I residues are within hydrogen bonding distance of each other (*94*). Interestingly, members of the LexA family of proteases undergo proteolysis at sites that follow the "-3, -1" rule (*85*).

Sauer and colleagues (*100*), using site-directed mutagenesis techniques, showed that Tsp protease also has Ser and Lys residues critical for catalysis. Similarly, mutagenesis studies suggest that a noncanonical (viral) Lon protease as well as the bacterial and organeller Lon proteases most likely utilize a Ser-Lys dyad catalytic mechanism (*101*).

Although they are not proteases, some amidases are also reported as utilizing a Ser-Lys dyad mechanism. An example is *E. coli* RTEM-1 β-lactamase. The crystal structure of a complex with penicillin G shows a Lys residue found within H-bonding distance of a Ser Oγ group of the cleaved penicillin complex (*92*). Mutagenesis studies of another amidase, mammalian fatty acid amide hydrolase (FAAH), strongly suggest a catalytic Ser residue and a Lys acid/base catalyst are present (*102*). FAAH is a hydrolase that is widespread in many different organisms, including mammals.

V. Conclusions and Perspective

Type I SPase belongs to a family of membrane-bound proteases that remove signal sequences from exported proteins after they are translocated

across the membrane. These proteases are found in all domains of life: eubacteria, archaea, and eukarya. Because they appear to be essential for bacterial viability, there is a great deal of interest in studying SPase as a drug target. To date, efforts to obtain effective inhibitors have been mixed, but researchers at SmithKline Beecham have isolated an inhibitor with an IC50 of 260 n*M*. SPase is a very challenging target for medicinal chemists focusing on drug design, especially given the fact that SPase I homologs are found in mitochondria and in the endoplasmic reticulum.

It is striking that the proteolytic mechanism used by type I bacterial signal peptidases is not the Ser/His/Asp triad of prototypical serine proteases, but rather, a Ser-Lys dyad. In this paradigm, the general base Lys, in a reduced pK_a environment, deprotonates the catalytic Ser hydroxyl group creating a strong nucleophile for subsequent catalysis. Evidence for this Ser–Lys mechanism has also been reported in other proteases such as LexA, Tsp, Lon, as well as amidases such as β-lactamase and fatty acid amide hydrolase.

Though this field has advanced considerably in recent years, other questions regarding the mechanism and substrate specificity of type I signal peptidases remain unanswered. For example, the pK_a of the putative general base Lys residue has never been directly measured and more precise studies to pinpoint microenvironmental factors responsible for its pK_a shift are necessary. How can SPase I accurately cleave its substrate when other potential sites are available nearby within the signal peptide and mature region of the substrate? Additional interactions likely occur during catalysis such as interactions between the H- or C-region of the substrate signal peptide and SPase enzyme. Also, studies have shown that there is a detergent or phospholipid requirement for the optimal catalytic activity of SPase I, but it is not known precisely which regions of the catalytic domain of the *E. coli* SPase physically interact with detergents or the actual membrane bilayer. Continuing efforts to address these and other issues will undoubtedly enlighten our understanding of this novel class of enzyme.

Acknowledgments

Ross E. Dalbey thanks his students who have contributed to the signal peptidase work over the years. Work from RED has been supported by a grant from the National Institutes of Health (GM 08512) and NSF (MCB-9808843). Joe Carlos was funded by the National Institutes of Health Predoctoral Training Grant GM 08512; Mark Paetzel was supported by a Medical Research Council of Canada postdoctoral fellowship. Natalie Strynadka was supported by the Medical Research Council of Canada, the Canadian Bacterial Diseases Network of Excellence, and the Burroughs Wellcome Foundation grant.

REFERENCES

1. Pugsley, A. P. (1993). *Microbiol. Rev.* **57,** 50.
2. LaPointe, C. F., and Taylor, R. K. (2000). *J. Biol. Chem.* **275,** 1502.
3. Date, T. (1983). *J. Bacteriol.* **154,** 76.
4. Dalbey, R. E., and Wickner, W. (1985). *J. Biol. Chem.* **260,** 15925.
5. Cregg, K. M., Wilding, I., and Black, M. T. (1996). *J. Bacteriol.* **178,** 5712.
6. Klug, G., Jager, A., Heck, C., and Rauhut, R. (1997). *Mol. Gen. Genet.* **253,** 666.
7. Zhang, Y. B., Greenberg, B., and Lacks, S. A. (1997). *Gene* **194,** 249.
8. Duong, F., and Wickner, W. (1997). *EMBO J.* **16,** 2756.
9. Cabelli, R. J., Chen, L., Tai, P. C., and Oliver, D. B. (1988). *Cell* **55,** 683.
10. Cabelli, R. J., Dolan, K. M., Qian, L. P., and Oliver, D. B. (1991). *J. Biol. Chem.* **266,** 24420.
11. Oliver, D. B., and Beckwith, J. (1982). *J. Bacteriol.* **150,** 686.
12. Kumamoto, C. A., and Beckwith, J. (1983). *J. Bacteriol.* **154,** 253.
13. Hartl, F. U., Lecker, S., Schiebel, E., Hendrick, J. P., and Wickner, W. (1990). *Cell* **63,** 269.
14. Sargent, F., Bogsch, E. G., Stanley, N. R., Wexler, M., Robinson, C., Berks, B. C., and Palmer, T. (1998). *EMBO J.* **17,** 3640.
15. Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H., Cole, J. A., and Turner, R. J. (1998). *Cell* **93,** 93.
16. Chanal, A., Santini, C., and Wu, L. (1998). *Mol. Microbiol.* **30,** 674.
17. Santini, C. L., Ize, B., Chanal, A., Muller, M., Giordano, G., and Wu, L. F. (1998). *EMBO J.* **17,** 101.
18. Boyd, D., Schierle, C., and Beckwith, J. (1998). *Protein Sci.* **7,** 201.
19. Samuelson, J. C., Chen, M., Jiang, F., Moller, I., Wiedmann, M., Kuhn, A., Phillips, G. J., and Dalbey, R. E. (2000). *Nature* **406,** 637.
20. von Heijne, G. (1983). *Eur. J. Biochem.* **133,** 17.
21. Perlman, D., and Halvorson, H. O. (1983). *J. Mol. Biol.* **167,** 391.
22. von Heijne, G. (1990). *J. Membr. Biol.* **115,** 195.
23. von Heijne, G. (1985). *J. Mol. Biol.* **184,** 99.
24. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). *Protein Eng.* **10,** 1.
25. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). *Int. J. Neural Syst.* **8,** 581.
26. von Heijne, G. (1986). *Nucleic Acids Res.* **14,** 4683.
27. Akita, M., Sasaki, S., Matsuyama, S., and Mizushima, S. (1990). *J. Biol. Chem.* **265,** 8164.
28. van der Wolk, J. P., Fekkes, P., Boorsma, A., Huie, J. L., Silhavy, T. J., and Driessen, A. J. (1998). *EMBO J.* **17,** 3631.
29. Nesmeyanova, M. A., Karamyshev, A. L., Karamysheva, Z. N., Kalinin, A. E., Ksenzenko, V. N., and Kajava, A. V. (1997). *FEBS Lett.* **403,** 203.
30. Hoyt, D. W., and Gierasch, L. M. (1991). *J. Biol. Chem.* **266,** 14406.
31. Rizo, J., Blanco, F. J., Kobe, B., Bruch, M. D., and Gierasch, L. M. (1993). *Biochemistry* **32,** 4881.
32. Michaelis, S., Hunt, J. F., and Beckwith, J. (1986). *J. Bacteriol.* **167,** 160.
33. Stader, J., Benson, S. A., and Silhavy, T. J. (1986). *J. Biol. Chem.* **261,** 15075.
34. Puziss, J. W., Fikes, J. D., and Bassford, P. J., Jr. (1989). *J. Bacteriol.* **171,** 2303.
35. Shen, L. M., Lee, J. I., Cheng, S. Y., Jutte, H., Kuhn, A., and Dalbey, R. E. (1991). *Biochemistry* **30,** 11775.
36. Karamyshev, A. L., Karamysheva, Z. N., Kajava, A. V., Ksenzenko, V. N., and Nesmeyanova, M. A. (1998). *J. Mol. Biol.* **277,** 859.
37. Chaddock, A. M., Mant, A., Karnauchov, I., Brink, S., Herrmann, R. G., Klosgen, R. B., and Robinson, C. (1995). *EMBO J.* **14,** 2715.

38. Brink, S., Bogsch, E. G., Edwards, W. R., Hynds, P. J., and Robinson, C. (1998). *FEBS Lett.* **434,** 425.
39. Chang, C. N., Blobel, G., and Model, P. (1978). *Proc. Natl. Acad. Sci. USA* **75,** 361.
40. Zwizinski, C., and Wickner, W. (1980). *J. Biol. Chem.* **255,** 7973.
41. Date, T., and Wickner, W. (1981). *Proc. Natl. Acad. Sci. USA* **78,** 6106.
42. Wolfe, P. B., Wickner, W., and Goodman, J. M. (1983). *J. Biol. Chem.* **258,** 12073.
43. Moore, K. E., and Miura, S. (1987). *J. Biol. Chem.* **262,** 8806.
44. Whitley, P., Nilsson, L., and von Heijne, G. (1993). *Biochemistry* **32,** 8534.
45. van Dijl, J. M., van den Bergh, R., Reversma, T., Smith, H., Bron, S., and Venema, G. (1990). *Mol. Gen. Genet.* **223,** 233.
46. Dalbey, R. E., Lively, M. O., Bron, S., and van Dijl, J. M. (1997). *Protein Sci.* **6,** 1129.
47. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). *Nucleic Acids Res.* **25,** 3389.
48. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998). *Trends Biochem. Sci.* **23,** 403.
49. Tjalsma, H., Bolhuis, A., van Roosmalen, M. L., Wiegert, T., Schumann, W., Broekhuizen, C. P., Quax, W. J., Venema, G., Bron, S., and van Dijl, J. M. (1998). *Genes Dev.* **12,** 2318.
50. Tjalsma, H., Stover, A. G., Driks, A., Venema, G., Bron, S., and van Dijl, J. M. (2000). *J. Biol. Chem.* **275,** 25102.
51. Bolhuis, A., Tjalsma, H., Stephenson, K., Harwood, C. R., Venema, G., Bron, S., and van Dijl, J. M. (1999). *J. Biol. Chem.* **274,** 15865.
52. Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (1998). *Nature* **396,** 186.
53. Van Valkenburgh, C., Chen, X., Mullins, C., Fang, H., and Green, N. (1999). *J. Biol. Chem.* **274,** 11519.
54. Klenotic, P. A., Carlos, J. L., Samuelson, J. C., Schuenemann, T. A., Tschantz, W. R., Paetzel, M., Strynadka, N. C., and Dalbey, R. E. (2000). *J. Biol. Chem.* **275,** 6490.
55. Tjalsma, H., Noback, M. A., Bron, S., Venema, G., Yamane, K., and van Dijl, J. M. (1997). *J. Biol. Chem.* **272,** 25983.
56. Meijer, W. J., de Jong, A., Bea, G., Wisman, A., Tjalsma, H., Venema, G., Bron, S., and van Dijl, J. M. (1995). *Mol. Microbiol.* **17,** 621.
57. Parro, V., Schacht, S., Anne, J., and Mellado, R. P. (1999). *Microbiology* **145,** 2255.
58. Koshland, D., Sauer, R. T., and Botstein, D. (1982). *Cell* **30,** 903.
59. Kuhn, A., and Wickner, W. (1985). *J. Biol. Chem.* **260,** 15914.
60. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrman, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P.-C., and Lucier, T. S. (1995). *Science* **270,** 397.
61. Black, M. T., and Bruton, G. (1998). *Curr. Pharm. Des.* **4,** 133.
62. Zwizinski, C., Date, T., and Wickner, W. (1981). *J. Biol. Chem.* **256,** 3593.
63. Kim, Y. T., Muramatsu, T., and Takahashi, K. (1995). *J. Biochem. (Tokyo)* **117,** 535.
64. Paetzel, M., Strynadka, N. C., Tschantz, W. R., Casareno, R., Bullinger, P. R., and Dalbey, R. E. (1997). *J. Biol. Chem.* **272,** 9994.
65. Wolfe, P. B., Silver, P., and Wickner, W. (1982). *J. Biol. Chem.* **257,** 7898.
66. Tschantz, W. R., and Dalbey, R. E. (1994). *Methods Enzymol* **244,** 285.
67. Talarico, T. L., Dev, I. K., Bassford, P. J., Jr., and Ray, P. H. (1991). *Biochem. Biophys. Res. Commun.* **181,** 650.
68. Zhong, W., and Benkovic, S. J. (1998). *Anal. Biochem.* **255,** 66.
69. Kuo, D., Weidner, J., Griffin, P., Shah, S. K., and Knight, W. B. (1994). *Biochemistry* **33,** 8347.

70. Dierstein, R., and Wickner, W. (1986). *EMBO J.* **5,** 427.
71. Dev, I. K., Ray, P. H., and Novak, P. (1990). *J. Biol. Chem.* **265,** 20069.
72. Wickner, W., Moore, K., Dibb, N., Geissert, D., and Rice, M. (1987). *J. Bacteriol.* **169,** 3821.
73. Chatterjee, S., Suciu, D., Dalbey, R. E., Kahn, P. C., and Inouye, M. (1995). *J. Mol. Biol.* **245,** 311.
74. Suciu, D., Chatterjee, S., and Inouye, M. (1997). *Protein Eng.* **10,** 1057.
75. Stein, R. L., Barbosa, M. D. F. S., and Bruckner, R. (2000). *Biochemistry* **39,** 7973.
76. Fikes, J. D., Barkocy-Gallagher, G. A., Klapper, D. G., and Bassford, P. J., Jr. (1990). *J. Biol. Chem.* **265,** 3417.
77. Jain, R. G., Rusch, S. L., and Kendall, D. A. (1994). *J. Biol. Chem.* **269,** 16305.
78. Nielsen, H., Brunak, S., and von Heijne, G. (1999). *Protein Eng.* **12,** 3.
79. Nilsson, I., and von Heijne, G. (1991). *J. Biol. Chem.* **266,** 3408.
80. Wrede, P., Landt, O., Klages, S., Fatemi, A., Hahn, U., and Schneider, G. (1998). *Biochemistry* **37,** 3588.
81. Black, M. T., Munn, J. G., and Allsop, A. E. (1992). *Biochem J.* **282,** 539.
82. Sung, M., and Dalbey, R. E. (1992). *J. Biol. Chem.* **267,** 13154.
83. Black, M. T. (1993). *J. Bacteriol.* **175,** 4957.
84. Tschantz, W. R., Sung, M., Delgado-Partin, V. M., and Dalbey, R. E. (1993). *J. Biol. Chem.* **268,** 27349.
85. van Dijl, J. M., de Jong, A., Venema, G., and Bron, S. (1995). *J. Biol. Chem.* **270,** 3611.
86. Carlos, J. L., Klenotic, P. A., Paetzel, M., Strynadka, N. C., and Dalbey, R. E. (2000). *Biochemistry* **39,** 7276.
87. Chabin, R., Green, B. G., Gale, P., Maycock, A. L., Weston, H., Dorn, C. P., Finke, P. E., Hagmann, W. K., Hale, J. J., MacCoss, M., Shah, S. K., Underwood, D., Doherty, J. B., and Knight, W. B. (1993). *Biochemistry* **32,** 8970.
88. Faraci, W. S., and Pratt, R. F. (1986). *Biochemistry* **25,** 2934.
89. Knight, W. B., Maycock, A. L., Green, B. G., Ashe, B. M., Gale, P., Weston, H., Finke, P. E., Hagmann, W. K., Shah, S. K., and Doherty, J. B. (1992). *Biochemistry* **31,** 4980.
90. Pratt, R. F., Surh, Y. S., and Shaskus, J. J. (1983). *J. Am. Chem. Soc.* **105,** 1006.
91. Wilmouth, R. C., Kassamally, S., Westwood, N. J., Sheppard, R. J., Claridge, T. D., Aplin, R. T., Wright, P. A., Pritchard, G. J., and Schofield, C. J. (1999). *Biochemistry* **38,** 7989.
92. Strynadka, N. C., Adachi, H., Jensen, S. E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K., and James, M. N. (1992). *Nature* **359,** 700.
93. Barkocy-Gallagher, G. A., and Bassford, P. J., Jr. (1992). *J. Biol. Chem.* **267,** 1231.
94. Peat, T. S., Frank, E. G., McDonald, J. P., Levine, A. S., Woodgate, R., and Hendrickson, W. A. (1996). *Nature* **380,** 727.
95. Paetzel, M., and Strynadka, N. C. (1999). *Protein Sci.* **8,** 2533.
96. van Klompenburg, W., Paetzel, M., de Jong, J. M., Dalbey, R. E., Demel, R. A., von Heijne, G., and de Kruijff, B. (1998). *FEBS Lett.* **431,** 75.
97. Schechter, I., and Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27,** 157.
98. Shinagawa, H. (1996). *Exs* **77,** 221.
99. Little, J. W. (1993). *J. Bacteriol.* **175,** 4943.
100. Keiler, K. C., and Sauer, R. T. (1995). *J. Biol. Chem.* **270,** 28864.
101. Birghan, C., Mundt, E., and Gorbalenya, A. E. (2000). *EMBO J.* **19,** 114.
102. Patricelli, M. P., and Cravatt, B. F. (2000). *J. Biol. Chem.* **275,** 19177.
103. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). *Nucleic Acids Res.* **25,** 4876.
104. Kraulis, P. K. (1991). *J. Appl. Cryst.* **24,** 946.
105. Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (2000). *Pharmacol. Ther.* **87,** 27.