

Altered –3 Substrate Specificity of *Escherichia coli* Signal Peptidase 1 Mutants as Revealed by Screening a Combinatorial Peptide Library*

Received for publication, September 11, 2006, and in revised form, October 27, 2006 Published, JBC Papers in Press, October 31, 2006, DOI 10.1074/jbc.M608779200

Özlem Doğan Ekici[‡], Andrew Karla[‡], Mark Paetzel[§], Mark O. Lively[¶], Dehua Pei[‡], and Ross E. Dalbey^{‡1}

From the [‡]Department of Chemistry, The Ohio State University, Columbus, Ohio 43210, the [¶]Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, and the [§]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

Signal peptidase functions to cleave signal peptides from pre-proteins at the cell membrane. It has a substrate specificity for small uncharged residues at –1 (P1) and aliphatic residues at the –3 (P3) position. Previously, we have reported that certain alterations of the Ile-144 and Ile-86 residues in *Escherichia coli* signal peptidase I (SPase) can change the specificity such that signal peptidase is able to cleave pro-OmpA nuclease A *in vitro* after phenylalanine or asparagine residues at the –1 position (Karla, A., Lively, M. O., Paetzel, M. and Dalbey, R. (2005) *J. Biol. Chem.* 280, 6731–6741). In this study, screening of a fluorescence resonance energy transfer-based peptide library revealed that the I144A, I144C, and I144C/I86T SPase mutants have a more relaxed substrate specificity at the –3 position, in comparison to the wild-type SPase. The double mutant tolerated arginine, glutamine, and tyrosine residues at the –3 position of the substrate. The altered specificity of the I144C/I86T mutant was confirmed by *in vivo* processing of pre- β -lactamase containing non-canonical arginine and glutamine residues at the –3 position. This work establishes Ile-144 and Ile-86 as key P3 substrate specificity determinants for signal peptidase I and demonstrates the power of the fluorescence resonance energy transfer-based peptide library approach in defining the substrate specificity of proteases.

Proteins destined for secretion are synthesized in a precursor form with an amino-terminal extension peptide that targets the exported protein to the Sec machinery (1) or the Tat machinery (2) in bacteria. During the export process, the signal peptide is cleaved from the precursor protein by a signal peptidase that is embedded in the plasma membrane.

In *Escherichia coli*, signal peptidase (SPase I)² consists of a single polypeptide chain of 37 kDa (3). This enzyme spans the

membrane twice with a small cytoplasmic segment (residues 29–58) and a large carboxyl-terminal catalytic domain located in the periplasm (residues 77–323) (4–6). Catalysis by SPase I is carried out by a Ser-Lys dyad (7–10). In the case of the *E. coli* SPase I, Ser-90 is the nucleophilic residue that attacks the scissile bond of the precursor substrate and lysine 145 is the general base that deprotonates the serine residue (for review, see Ref. 11). A critical serine and lysine residue is also present in SPases from other species of bacteria (12), and members of the signal peptidase I family in mitochondria (13).

With the exception of the mitochondrial inner membrane peptidase I (Imp1), all type I signal peptidases carry out processing with a specificity for small aliphatic residues at the –1 (P1) and –3 (P3) positions (11). Alanine is usually the preferred amino acid residue at the –1 and –3 positions and results in the frequently observed “Ala-X-Ala” motif for signal peptide cleavage (14–16). The residues of SPase I that comprise the substrate binding site have been identified by solving the x-ray structure of the soluble catalytic domain with a covalently attached 5S penem inhibitor (10) and a structure with a non-covalent lipohexapeptide inhibitor (17). The three-dimensional structure of SPase I with no inhibitor bound (apo-structure) revealed that there is some variation in the binding pocket volume when compared with the inhibitor-bound structures (18). The *E. coli* SPase residues making direct van der Waals contact with the P1 methyl group are Met-91, Ile-144, Leu-95, and Ile-86. Those making contact with the P3 residues are Phe-84, Ile-144, Val-132, and Ile-86. The substrate binding to SPase I occurs in an extended conformation. Recently, we have made mutations of the *E. coli* SPase I in the S1 and S3 pockets that bind the P1 and P3 residues of the substrate to identify the residues that control the substrate specificity (19). We found that alterations of the Ile-144 and Ile-86 residues to alanine residues could alter the substrate specificity and lead to cleavage after a –1 Phe residue *in vitro*.

Defining the SPase residues that control the substrate specificity is important because it provides insight into how SPase is

* This work was supported by National Science Foundation Grant MCB-0316670 (to R. E. D.), National Institutes of Health GM062820 (to D. P.), the Canadian Institute of Health Research (to M. P.), a National Science and Engineering Research Council of Canada operating grant (to M. P.), the Michael Smith Foundation for Health Research Scholar award (to M. P.), and a Canadian Foundation of Innovation grant (to M. P.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ To whom correspondence should be addressed: 100 West 18th Ave, Columbus, OH 43210. Tel.: 614-292-2384; Fax: 614-292-1532; E-mail: dalbey@chemistry.ohio-state.edu.

² The abbreviations used are: SPase, signal peptidase; AcOH, acetic acid; CH₃CN, acetonitrile; DABCYL, *N*-4-[4'-(dimethylamino)phenylazo]benzoic

acid; DMF, *N,N*-dimethylformamide; EDANS, 5-((2-aminoethyl)amino)naphthalene-1-sulfonic acid; Fmoc, *N*-9-fluorenylmethoxycarbonyl; Fmoc-OSu, *N*-(9-fluorenylmethoxycarbonyloxy)succinimide; HBTU, 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HOBt, 1-hydroxybenzotriazole; MeOH, methanol; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; WT, wild type; FRET, fluorescence resonance energy transfer; PED/MS, partial Edman degradation/mass spectrometry.

able to locate the site of cleavage at the cell membrane surface. Recognition of the substrate by SPase at the correct site is challenging because the substrate specificity determinant Ala-X-Ala is a common motif in proteins. In addition to the importance of the SPase S1 and S3 subsites in recognition of the correct cleavage site, it is likely that the interaction of the SPase catalytic domain with the membrane is important for high fidelity of the enzyme (10, 20).

In this paper we have systematically determined the substrate specificity of various signal peptidase Ile-144 single mutants and Ile-144/Ile-86 double mutants by using a powerful fluorescence resonance energy transfer (FRET)-based peptide library approach. These mutants exhibited a more relaxed substrate specificity at the -3 position, with the double mutant tolerating glutamine and arginine as P3 residues. In a cellular assay, the SPase I144C/I86T mutant efficiently cleaved a pre- β -lactamase mutant with a glutamine or arginine at the -3 position. These results show that Ile-144 and Ile-86 play critical roles in controlling the substrate specificity at the -3 position and demonstrate the power of this new combinatorial library method for determining the substrate specificity of proteases.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Plasmids—The *E. coli* strain DH5 α was obtained from our laboratory collection although the *E. coli* temperature-sensitive SPase I strain, IT41, was obtained from Dr. Yoshikazu Nakamura (21). The plasmids pRD8, which contains the SPase I gene in the pING vector, and pUC19 were obtained from our collection. The plasmid pGZ119HE was generously provided by Dr. Andreas Kuhn.

Construction of Plasmids—To examine the ability of various SPase binding pocket mutants to process β -lactamase mutants, a two-plasmid system was employed requiring the preparation of two constructs. The construction of the two plasmids was accomplished as follows. First, the SPase mutants were subcloned from the pET23b vector (19) to avoid the concomitant expression of WT β -lactamase from this vector in these studies. To this end, SPase was subcloned into the SmaI/SalI site of the pGZ119HE vector (22). The pGZ119HE is suitable for this study because it possesses the ColD origin of replication and confers chloramphenicol resistance. The pET23b plasmids bearing the various SPase mutants were digested with SalI and SmaI in the same reaction vessel using 1.5 \times Universal Buffer (Stratagene), and the DNA fragment containing SPase was purified by excision from an agarose gel. The pGZ119HE vector was prepared in the same way and the two were ligated to produce the pGZ119HE-SPase expression vector. The resulting DNA was sequenced to confirm successful subcloning of SPase. Second, a plasmid capable of expressing *E. coli* TEM-1 β -lactamase (UniRef90_P62593) was needed that could be simultaneously transformed with the SPase mutants. For this vector, we modified pRD8 by removing the SPase gene (23). The pRD8 plasmid contains the ColE1 replication origin and the *bla* gene for ampicillin resistance and is thus compatible with the pGZ119HE-SPase expression vector. Additionally, β -lactamase can be highly expressed from this plasmid by the addition of 0.2% arabinose. The pRD8 plasmid was digested with the SalI and SmaI enzymes as described above and the products were

separated on an agarose gel. The DNA fragment corresponding to the doubly cut vector was excised and purified from the gel. The 5' overhangs left by SalI digestion were filled in with Klenow and the resulting DNA was then ligated to produce what is essentially the original pING vector (24). Additionally, pUC19 empty vector was also used in the pulse-chase studies to express β -lactamase. The plasmid, pUC19, contains the pMB1 origin of replication and can be co-transformed with the pGZ119HE-SPase constructs. In addition, it expresses β -lactamase constitutively at high level. The pING and pUC19 plasmids were then modified using the QuikChange (Stratagene Inc.) site-specific mutagenesis method to incorporate different amino acid residues at the position -3 to the cleavage site.

Purification of β -Lactamase— β -Lactamase was purified using the PheBo system from MoBiTec (Goettingen, Germany) that utilizes phenylboronate-agarose resin for the specific purification of β -lactamase. The purification was conducted as described in the manual. First, DH5 α cells were transformed with the pUC19 vector containing either WT or mutant β -lactamase and were grown to saturation at 37 $^{\circ}$ C. Cells were isolated by low speed centrifugation (4000 \times *g*, Beckman JA-10) and the periplasmic fraction was isolated by osmotic shocking of the *E. coli* cells. Cells were resuspended in ice-cold STE buffer (20% sucrose, 200 mM Tris/HCl, 100 mM EDTA, pH 9.0) with gentle shaking for 20 min and then pelleted (10,000 \times *g*, Beckman JA-10). The pellet was then resuspended in ice-cold 10 mM Tris/HCl, pH 9.0, with gentle shaking for 20 min and then centrifuged (10,000 \times *g*, Beckman JA-10). At this step, the supernatant containing the periplasmic cell fraction was isolated. β -Lactamase was precipitated from the periplasmic fraction with ammonium sulfate (130 g into a 200-ml periplasmic fraction) and the precipitated proteins were dissolved in 20 mM triethanolamine, 0.5 M NaCl, pH 7.0, buffer and applied to the phenylboronate column. Elution was performed with borate buffer to elute the β -lactamase.

Mass Spectrometry—Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry was used to determine the masses of the SPase cleavage products. The sites of cleavage of wild-type and the -3 Trp β -lactamase were determined by comparison of the predicted masses of the product proteins to the observed masses. Theoretical molecular masses of the proteins were calculated using the PeptideMass program (25). The proteins were purified from intact cells as described above and the samples precipitated for mass spectrometry analysis.

β -Lactamase fractions purified by phenylboronate-agarose chromatography were first precipitated with a final concentration of 12% trichloroacetic acid for 1 h on ice. Precipitated proteins were then centrifuged at maximum speed at 4 $^{\circ}$ C in an Eppendorf microcentrifuge for 20 min. The trichloroacetic acid supernatant was carefully removed and the pellet was washed with ice-cold 90% acetone, vortexed, and centrifuged again at maximum speed for 15 min. The acetone supernatant was carefully removed and the acetone wash was repeated once. The protein pellet was finally dried of all residual acetone.

The precipitated β -lactamase protein pellets were dissolved by addition of 5 μ l of 10% formic acid. The dissolved proteins were mixed with the MALDI matrix solution: 5 μ l of saturated

solution of α -cyano-4-hydroxycinnamic acid (10 mg dissolved in 500 μ l of 0.1% trifluoroacetic acid and 500 μ l of CH_3CN). The aluminum MALDI-TOF target plate was spotted with 1 μ l of each reaction sample containing matrix and analyzed using a Bruker Daltonics Autoflex mass spectrometer in the linear mode. The instrument was calibrated with a mixture of protein standards including insulin (5,734.6 Da); cytochrome *c* (12,361.1 Da), and myoglobin (16,952.6 Da). The mass accuracy in the 30-kDa mass range is approximately ± 25 Da.

Pulse-Chase Assay of β -Lactamase Processing—Competent cells of the IT41 strain were prepared using the CaCl_2 method and co-transformed with the various mutants of the pGZ119HE-SPase and pING vectors. In the later studies that incorporate the P2F mutation in addition to the -3 mutations in β -lactamase, pUC19 was used to express these mutants of β -lactamase alongside the pGZ119HE-SPase vector. For culturing IT41, all media were prepared with a reduced salt concentration of 2.5 g of NaCl /liter (LS2.5 media). IT41 cells carrying the plasmids for SPase and β -lactamase were then grown at 30 °C on solid LS2.5 media until colonies were 1–2 mm in diameter. These were then transferred to liquid LS2.5 media and grown to a cell density of 0.3–0.5 A_{600} . At this point, the cells were transferred to M9 media containing 0.5% fructose plus 19 amino acids minus methionine and incubated at 30 °C. Even at this permissive temperature, the chromosomal SPase I activity is strongly impaired. After 30 min, arabinose was added to a final concentration of 0.2% for 10 min to increase the expression of β -lactamase from the pING vector. For co-transformants utilizing the pUC19 vector for β -lactamase expression, the arabinose induction was omitted. The cells were pulsed with Trans ^{35}S -label for 1 min and then chased with non-radioactive methionine for the indicated times. At each time point, samples were quenched with a 10% final concentration of trichloroacetic acid in preparation for immunoprecipitation with rabbit anti- β -lactamase polyclonal antibody (Chemicon Int.).

Materials for Peptide Library—PL-PEGA resin (0.2 mmol/g, 300–500 μ m) was purchased from Polymer Laboratories Ltd. (Amherst, MA). All of the reagents for peptide synthesis were purchased from Advanced ChemTech (Louisville, KY), Novabiochem (San Diego, CA), or Bachem (Torrance, CA). Sodium 5-((2-aminoethyl)amino)naphthalene-1-sulfonate (EDANS) was purchased from Invitrogen Molecular Probes (Carlsbad, CA). All other chemicals were purchased from Aldrich and Acros Organics (Belgium).

Synthesis of Peptide Library I—Peptide library I, Fmoc-K(D-ABCYL)ATXXXXATE(EDANS)BBRM-resin, was synthesized using the standard Fmoc/HBTU/HOBt synthesis protocol ($X = 18$ natural amino acids except cysteine and methionine, $B = \beta$ -alanine). The PL-PEGA resin was used as the solid support. The synthesis was carried out on 1.5 g of PEGA resin in a reaction vessel specifically designed for manual peptide synthesis. First, the 7-residue constant region, ATE(Allyl)BBRM, was synthesized with 4 eq of reagents. Next, the random region was generated using the split-pool synthesis method (26). The resin was evenly divided into 18 aliquots and placed into 18 separate reaction vessels. Each aliquot was coupled with a different amino acid (4 equivalents of reagent, 30 min) and the coupling

reaction was repeated once. Then, the resin from all the vessels was combined, mixed, washed exhaustively with DMF, and the Fmoc group was removed by treatment with 20% piperidine/DMF twice (5 + 15 min). The resin was then redistributed into 18 reaction vessels, and this process was repeated until a library with four randomized positions was generated. After the construction of the entire peptide chain was completed, the allyl group was selectively removed using 1 eq of $\text{Pd}(\text{PPh}_3)_4$ in $\text{CHCl}_3/\text{AcOH}/N$ -methylmorpholine (37:2:1) under argon at room temperature. The reaction was quenched with 0.5% N,N -diisopropylethylamine/DMF and 0.5% diethyldithiocarbamate/DMF after 3 h and the resin was washed exhaustively with DMF. The resin was then treated twice with 5-fold excess of a EDANS-sodium salt/HBTU/HOBt mixture for 4 h at room temperature and washed with DMF and methanol until the white precipitate that had formed throughout the reaction completely disappeared. Finally, the side chain protecting groups were removed with a cleavage mixture containing 4.75 ml of trifluoroacetic acid, 0.2 ml of thioanisole, 0.1 ml of anisole, and 0.1 ml of ethanedithiol for 1 h at room temperature. The resin was washed with CH_2Cl_2 (5 \times 10 ml) and stored in the same solvent in the swollen form at 4 °C.

Synthesis of Peptide Library II—Fully protected peptide library I from above, Fmoc-K(DABCYL)ATXXXXATE(Allyl)BBRM-resin, was treated with 20% piperidine/DMF twice (5 + 15 min) to remove the NH_2 -terminal Fmoc group. The resulting resin was washed with DMF and water, and soaked in water overnight. The water was drained and the resin was treated with 1 eq of Fmoc-OSu and N,N -diisopropylethylamine in CH_2Cl_2 while vigorously shaking for 30 min to achieve the Fmoc group coupling only on the surface of the beads. This procedure spatially segregated the beads into two layers; the peptides on the surface layer were NH_2 terminally blocked by Fmoc group, whereas the interior peptides contained a free NH_2 terminus (27). Next, the interior peptides were NH_2 terminally blocked with a Boc group by the treatment of 4 eq of Boc-Gly-OH and HBTU/HOBt/ N,N -diisopropylethylamine for 30 min twice. After removal of the NH_2 -terminal Fmoc group, the peptide sequence KKKKLLLLLLLLLLLL was coupled to the beads sequentially. The last lysine residue introduced was Ac-Lys(Boc)-OH with the same conditions. Removal of the allyl group from glutamic acid, addition of the EDANS group, and peptide deprotection were carried out as described above.

On-bead Screening of Peptide Library—A typical screening reaction involved $\sim 30,000$ beads of peptide libraries I or II. The beads were washed with water and the *E. coli* SPase reaction buffer (50 mM Tris, 10 mM CaCl_2 , 1% Triton X-100, pH 8.0), and treated with 250 μ l of WT (0.4 mg/ml) or 400–800 μ l of mutant *E. coli* SPase (I144A, 0.13 mg/ml; I144C, 0.061 mg/ml; I144C/I86T, 0.36 mg/ml) for 18 h at 37 °C in a 60 \times 15-mm Petri dish (Baxter Scientific Products). The dish was viewed under a fluorescence microscope (Olympus SZX12) using the appropriate filter set for the EDANS group (filter set for 1,5-IAEDANS group: exciter 360 nm, emitter 460 nm). Positive beads were identified by their intense turquoise color and removed from the library by a micropipette. A control screening was carried out under the same conditions with the exclu-

Signal Peptidase Binding Site Mutants

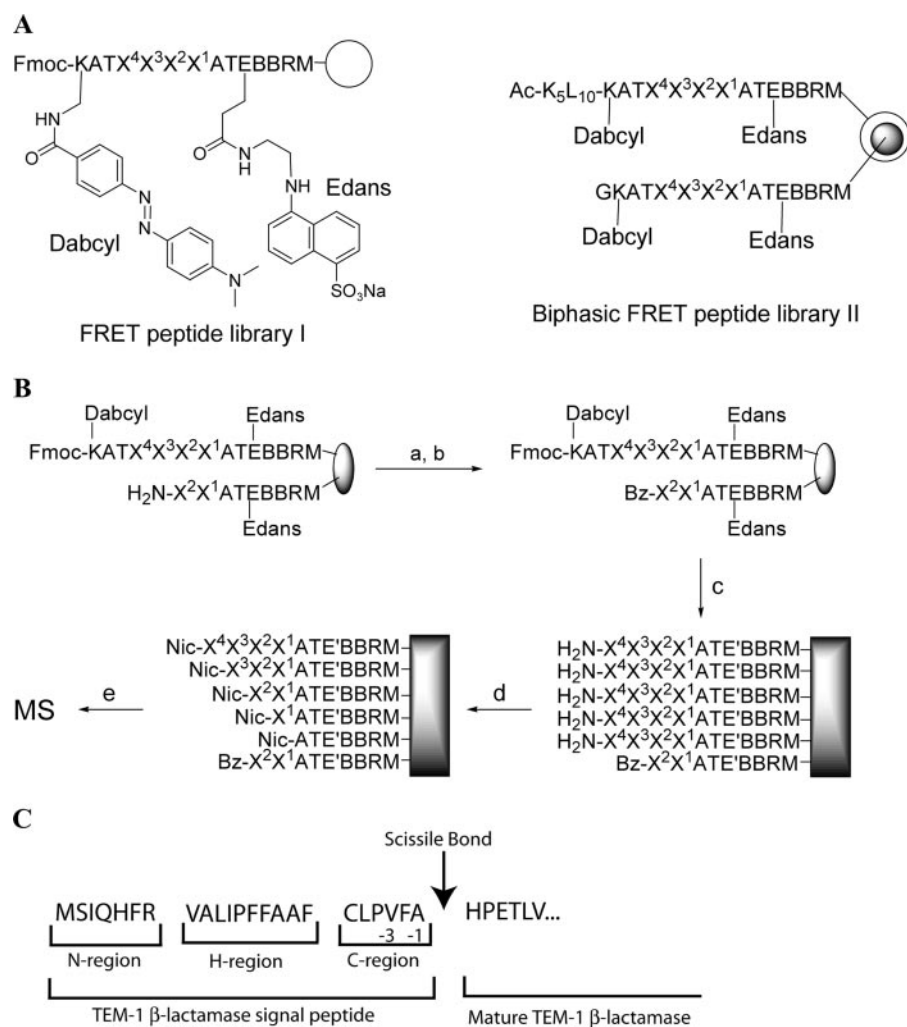


FIGURE 1. A, FRET peptide library design. 1) FRET peptide library and 2) biphasic FRET Lys/Leu chain peptide library. B, identification of the sequence and the cleavage site on the FRET library using the partial Edman degradation method ($E' = \text{EDANS}$). a, Bz-OSu; b, 20% piperidine/DMF; c, phenylisothiocyanate followed by trifluoroacetic acid (3 cycles); d, phenylisothiocyanate/Nic-OSu, 5:1, followed by trifluoroacetic acid (4 cycles), *N*-hydroxysuccinimidyl nicotinate (*Nic-Osu*) (1 cycle); e, general CNBr cleavage. C, the amino acid sequence surrounding the pre- β -lactamase cleavage site. The amino acid sequence surrounding the site of cleavage is depicted in *single letter code*. WT cleavage site (scissile bond) is indicated with a *vertical arrow* and the -1 and -3 residues are labeled. The N-region, H-region, and C-regions of the TEM-1 β -lactamase signal peptide are assigned based on analysis using the SignalP 3.0 server (38).

sion of the enzyme. This screening resulted in no fluorescent beads.

Peptide Sequencing—Positive beads from above were sequenced by the method of partial Edman degradation/mass spectrometry (PED/MS) as previously described (28) with the following modifications. After washing with water and pyridine, the beads were suspended in 160 μl of pyridine and treated with 10% *N*-hydroxysuccinimidyl benzoate (Bz-OSu) for 12 min at room temperature. This blocked any free NH_2 terminus including those resulting from enzymatic cleavage. The beads were then treated with 20% piperidine in DMF to remove the NH_2 -terminal Fmoc group and subjected to PED reactions and MS (28).

Molecular Modeling—A model of the wild-type SPase in complex with the c-region of the β -lactamase signal peptide was built using the atomic coordinates from the crystal structure of signal peptidase in complex with a lipopeptide inhibitor

(Protein Data Bank code 1T7D) as the template (17). The model for the *E. coli* TEM-1 β -lactamase signal peptide (Sequence Swiss-Prot accession number Q1WBW3) was built into the active site of SPase using the acyl-enzyme penem inhibitor (29) and noncovalently bound lipopeptide inhibitor complexes (17) as a guide to position the P1 and P3 residues into the S1 and S3 binding pockets. The paths taken by the signal peptide along the surface of SPase and the parallel β -sheet-type hydrogen bonding interactions with SPase are similar to those observed in the lipopeptide-inhibitor complex structure. The model of the SPase double mutant (I144C/I86T) in complex with the -3 Arg β -lactamase signal peptide was built from the modeled wild-type complex. The mutations were made with the program COOT (30). The models of the SPase/ β -lactamase signal peptide complexes were energy minimized using the program CNS (31). Figures were prepared using the program PyMol (32). The surface and binding site analysis was performed using the CASTp server (33).

RESULTS

Design, Synthesis, and Sequencing of FRET-based Peptide Library—A peptide library, Fmoc-K(DABCYL)-ATXXXATE(EDANS)BBRM-resin ($X = 18$ amino acids), was designed to probe the sequence specificity of WT and mutant *E. coli* SPase I (I, Fig. 1A). The peptide library was synthesized

on a solid support using a split-pool methodology to generate a one bead-one sequence library (26). A polyethylene glycol and acrylamide-based amino PEGA₁₉₀₀ resin was chosen as the solid support due to its ability to swell in hydrophilic conditions thus allowing relatively large biomolecules such as enzymes to easily permeate the resin (34).

Prior to enzymatic reaction, the resin beads are non-fluorescent due to efficient quenching of EDANS fluorescence by the DABCYL group within the same peptide. Upon enzymatic cleavage, however, the quencher group DABCYL is released into the solution and the beads carrying the cleaved substrates containing the EDANS group become intensely fluorescent. The fluorescent beads are easily detected and manually collected under a fluorescence microscope. The selected beads were then subjected to PED/MS, a sensitive and reliable method for sequence determination of support-bound peptides from combinatorial libraries (28). In this method, a support

bound peptide is converted into a series of sequence-related truncation products by treating the peptide with a 5:1 mixture of phenylisothiocyanate and *N*-hydroxysuccinimidyl nicotinate (Fig. 2B). The resulting peptide ladder was analyzed by MALDI-MS and the sequence of the original peptide is identified.

In this study, the PED/MS procedure was slightly modified to reveal the site of enzymatic cleavage. The NH₂ termini of the

resin-bound peptides were protected with Fmoc groups during incubation with the enzyme (Fig. 1B). After treatment with SPase, the isolated fluorescent beads were treated with *N*-hydroxysuccinimidyl benzoate to cap the new amino termini produced by SPase. Thus attachment of the benzoyl group (Bz) marked the sites of cleavage (a, Fig. 1B). The remaining Fmoc groups on the uncleaved peptides were then removed and "partial" Edman degradation were performed to create the peptide ladder (Fig. 1B, see legend for details). The amino termini of the ladder were capped with the nicotinoyl group by treatment with *N*-hydroxysuccinimidyl nicotinate. In the resulting MALDI-TOF spectrum, the NH₂-terminal benzoylated peptides produced by enzymatic cleavage of the resin-bound peptide have a mass that is 0.995 Da less than the corresponding *N*-nicotinoylated counterpart formed during the four cycles of partial Edman degradation. The mass spectrum allows the interpretation of the amino acid sequence present on the bead by calculating the differences between the nicotinoyl-labeled peptide fragments. The cleavage site is identified as the one benzoyl-labeled fragment with a mass 0.995 Da less than the corresponding nicotinoyl-labeled form. This approach permits the identification of the peptidase cleavage site as well as the amino acid sequence of the randomized peptide region present on a single bead.

Sequence Specificity of WT and Mutant SPases—Treatment of 75 mg of library I with detergent solubilized (full-length) WT SPase I produced 28 fluorescent beads. PED/MS analysis gave 25 unambiguous sequences. All of these peptides contained an alanine at the P1 position (Table 1). At position -3, SPase also has a strong preference for alanine (19 of 25 sequences), although other small residues such as valine and threonine were occasionally observed. A variety of residues was observed at the -2 position including alanine, arginine, asparagine, glutamic acid, glutamine, histidine, lysine, phenylalanine, serine, threonine, tryptophan, and tyrosine. These results are largely in line with the results of Rosse *et al.* (35), although they also observed leucine and lysine at the -3 position, in addition to alanine, valine, and threonine.

Next, I144A and I144C single mutants were analyzed by the FRET-based library. The I144A mutant efficiently cleaved peptides with -1 residues of alanine, serine, threonine, glycine, and asparagine (Table 1). Like the wild-type SPase, alanine was still the most preferred residue. At the -3 position, the cleaved peptides most frequently contained alanine, but also serine, valine, leucine, glutamic, and aspartic acid residues. Similar

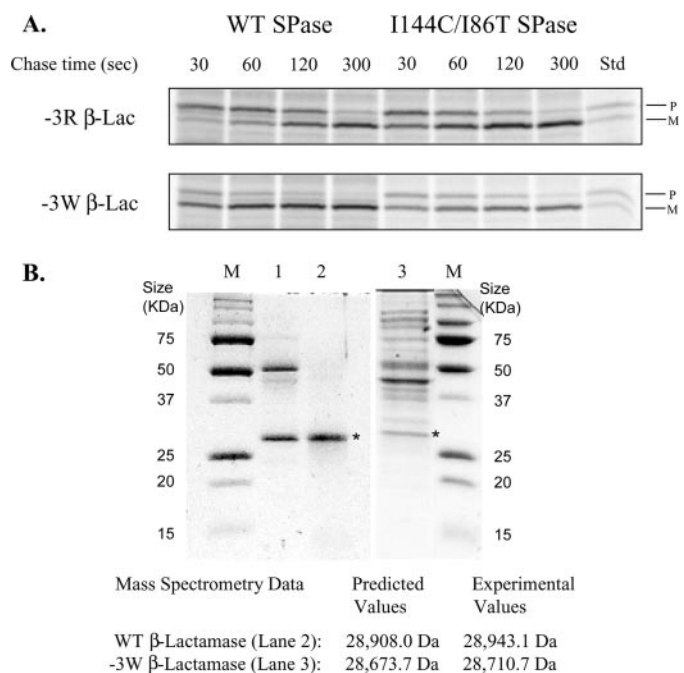


FIGURE 2. Processing of -3 β-lactamase mutants at an alternate site by wild-type SPase *in vivo*. *A*, *in vivo* cleavage of pre-β-lactamase mutants by wild-type and I144C/I86T SPase. Pulse-chase studies showing processing of -3 Arg and -3 Trp β-lactamase by WT and I144C/I86T SPase *in vivo*. Cells were pulse-labeled with [³⁵S]methionine and chased with cold methionine for the indicated times, as described under "Experimental Procedures." Radio-labeled β-lactamase was immunoprecipitated with anti-β-lactamase serum, and analyzed by SDS-PAGE and phosphorimaging. Pre- and mature forms of β-lactamase are indicated by P and M, respectively, and a partially processed β-lactamase standard (Std) is shown as a control. *B*, purification of WT and -3 Trp β-lactamases with the predicted and experimental mass data. Mature β-lactamase was isolated from the periplasmic fraction of DH5α containing plasmids expressing either WT or -3 Trp β-lactamase. Periplasmic fractions were further purified by phenylboronate chromatography and run on SDS-PAGE to assess purity. Lanes 1 and 2 are two fractions of the purified WT β-lactamase, and lane 3 is the partially purified -3 Trp β-lactamase. Mature bands are indicated by the asterisk. The results of the mass spectrometry analysis are listed with the predicted values expected for cleavage following the -1 alanine for the WT substrate and the +2 proline for the -3 Trp β-lactamase.

TABLE 1
Substrate specificity of WT and mutant SPases with FRET peptide substrate libraries I and II

Enzyme	Enzyme subsite ^a		
	-1	-2	-3
FRET library I			
WT SPase (25 beads)	25 Ala	X 19 Ala, 3 Val, 2 Thr	X 63 Ala, 11 Ser, 3 Val, 2 Leu, 1 Glu, 1 Asp, 1 U
I144A SPase (82 beads)	66 Ala, 10 Ser, 4 Thr, 1 Gly, 1 Asn	X	X 136 Ala, 26 Ser, 8 Val, 7 Thr, 5 Glu, 4 Leu, 3 Gly, 2 Asp, 2 Lys, 1 His, 1 Trp, 1 Tyr, 6 U
I144C SPase (200 beads)	176 Ala, 15 Ser, 9 Thr, 2 Gly	X	
FRET library II			
I144C/I86T (35 beads) ^b	35 Ala	X 8 Ala, 7 Arg, 7 Leu, 3 Pro, 2 Tyr, 2 Gln, 1 Glu, 1 His, 1 Gly, 1 Ser, 1 Val, 1 Trp	

^a X, any natural L-amino acid except cysteine or methionine; U, unidentified residue, the notation of the sequences will be: NH₂ terminus-T/XXXX/A-COOH terminus.

^b The sequences obtained are: T/URAF/A, T/AAFA/A, T/AFAF/A, T/WHAP/A, T/FPAR/A, T/NRAW/A, T/RNAY/A, T/ALAA/A, T/ULRL/A, T/RTRN/A, T/RTRQ/A, T/PARV/A, T/PYRA/A, T/RRFR/A, T/RPLR/A, T/TYYE/A, T/NPVR/A, T/VPQH/A, T/WRQP/A, T/PRSL/A, T/RVLA/A, T/LVLP/A, T/ALPV/A, T/RLYR/A, T/RWRF/A, T/RRPG/A, T/RHGR/A, T/LVLL/A, 2 X T/RNLR/A, T/ARWY/A, T/GQPQ/A, T/RPLY/A, T/RWHE/A, T/TPER/A. With alternate cleavage sites the distribution of the residues is as follows: (-1) 35 Ala, (-2) X, (-3) 2 Ala, 6 Arg, 7 Leu, 3 Pro, 3 Tyr, 2 Gln, 2 Trp, 2 Val, 2 Phe, 1 Asn, 1 Thr, 1 His, 1 Glu, 1 Gly, 1 U.

Signal Peptidase Binding Site Mutants

results were observed with the I144C mutant except that the -1 position did not include an asparagine residue, whereas the -3 position also included glycine, histidine, tryptophan, tyrosine, and lysine residues. These results suggest the mutant enzymes have a more relaxed specificity at both the -1 and -3 positions.

Treatment of the above library with I144C/I86T mutant SPase yielded no fluorescent beads, likely due to the reduced catalytic activity of the double mutant. Stein and co-workers (36) have previously shown that the addition of a K_5L_{10} sequence to the NH_2 terminus of a peptide substrate increased its rate of cleavage by 30,000-fold (36). Therefore, we modified library I (Fig. 1A) by adding a K_5L_{10} peptide sequence to the NH_2 terminus of each library member (II, Fig. 1A). To facilitate later peptide sequencing by PED/MS, each resin bead was spatially segregated into outer and inner layers by a biphasic synthesis strategy (27). The K_5L_{10} sequence was added only to peptides on the bead surface, making them better substrates for the mutant SPase and thus providing more sensitive detection of residual catalytic activities. A glycine residue was added only to the NH_2 termini of peptides located in the bead interiors; these peptides are not substrates for the mutant SPase but serve as encoding tags that can be readily sequenced by PED/MS. Unfortunately, this strategy was not compatible with the NH_2 -terminal benzoylation approach used to mark cleaved sites. Consequently, the enzymatic cleavage site could not be experimentally determined using library II (Fig. 1A). Incubation of library II with the I144C/I86T mutant produced weak to moderately fluorescent beads (Table 1). Inspection of the cleaved peptides suggests that the double mutant still cleaved predominantly after an alanine at the P1 position. However, these peptides contained a wide variety of amino acids at the P3 position, including alanine, arginine, leucine, proline, tyrosine, glutamine, glutamic acid, histidine, glycine, serine, valine, and tryptophan. Thus, the P3 site no longer plays a determining role in cleavage for the I144C/I86T double mutant. It is worth noting, however, that this SPase still requires an extended NH_2 -terminal hydrophobic tail on the substrate to obtain cleavage.

In Vivo Processing of -3 Pre- β -lactamase Mutants by Wild-type and SP Mutants at an Alternative Site—The altered specificity of mutant SPases at P3 site was further tested *in vivo*. We developed a system in which SPase and its substrate were encoded by two separate plasmids. The TEM-1 β -lactamase protein was chosen as a substrate because this periplasmic protein possesses a cleavable signal peptide. Inspection of the sequence of the cleavage region of pre- β -lactamase reveals a single SPase I processing site (Fig. 1C, see arrow). The bacterial host used in these studies, *E. coli* IT41 (21), contained a chromosomally encoded, temperature-sensitive mutant SPase. Deletion of SPase is lethal (37) so use of the temperature-sensitive strain is necessary to assess the effects of mutations on SPase itself. The IT41 strain does not grow at 42 °C but does grow slowly at 30 °C. Even at the permissive temperature of 30 °C the activity of the chromosomal signal peptidase is sharply reduced.

We first examined the ability of the I144C/I86T mutant to process a β -lactamase mutant containing an arginine at the P3 site. As observed *in vitro*, the I144C/I86T SPase cleaved the

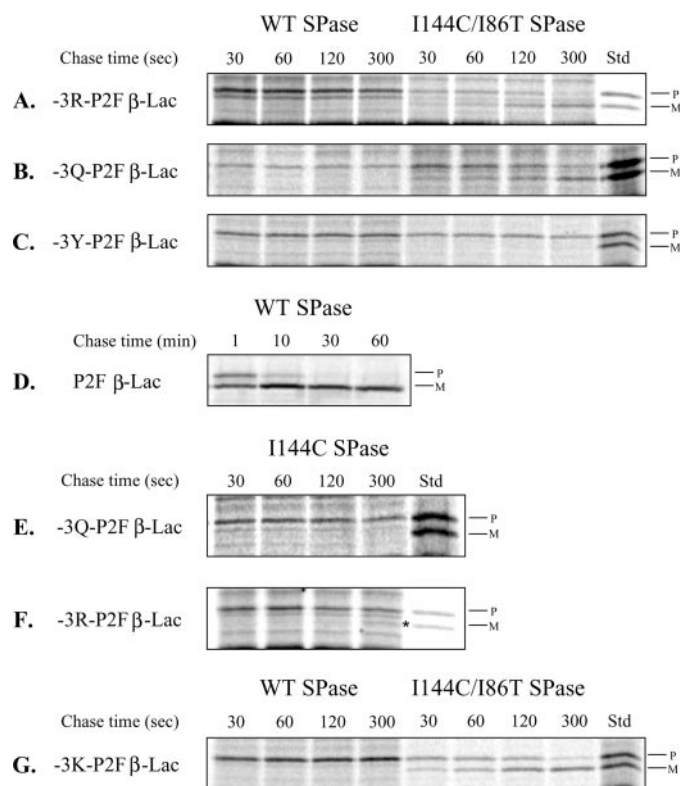


FIGURE 3. Processing of various -3 β -lactamase mutants by wild-type and I144C/I86T SPase *in vivo* where alternate cleavage after proline +2 is prevented. Pulse-chase studies are shown for the -3 β -lactamase mutants in which alternate processing is prevented by mutating the +2 proline to phenylalanine (P2F). The mutants $-3R$ -P2F (A), $-3Q$ -P2F (B), or $-3Y$ -P2F (C) were studied in the presence of WT SPase or the I144C/I86T SPase *in vivo*. The pulse-chase experiment was conducted as described in Fig. 2A. Pre- and mature forms of β -lactamase are indicated by P and M, respectively, and a partially processed β -lactamase standard (Std) is shown as a control. Pulse-chase studies were also performed with the P2F β -lactamase (D) when wild-type SPase was expressed. Pulse-chase studies are shown for the -3 Gln (E), and the -3 Arg (F) β -lactamase mutant that showed some level of processing by the I144C SPase enzyme. The faint processed band in the 300-s time point for the $-3R$ -P2F substrate is indicated by asterisk. Pulse-chase studies of the $-3K$ -P2F (G) β -lactamase substrates in the presence of WT SPase or the I144C/I86T mutant *in vivo*. The pulse-chase experiment was conducted as described in Fig. 2A.

$-3R$ β -lactamase (Fig. 2A, upper panel). Surprisingly, the wild-type enzyme also cleaved the mutant substrate (Fig. 2A). Both forms of SPase also cleaved the -3 Trp β -lactamase mutant (Fig. 2A, lower panel). Because WT SPase has never been observed to cleave substrates with a -3 arginine or tryptophan (38) we considered the possibility that the observed processing was occurring at an alternative site other than the normal processing site.

To determine the actual cleavage site by the wild-type SPase, we purified the wild-type and -3 Trp β -lactamases from DH5 α cells by affinity chromatography on a phenylboronate-agarose resin. Fig. 2B shows that the WT β -lactamase can be isolated in a pure form (lane 2). In contrast, we were unable to isolate the -3 Trp β -lactamase protein in pure form by this procedure (see Fig. 2B, lane 3 and asterisk for position of β -lactamase protein). Nevertheless, the impure preparation was analyzed by MALDI-TOF mass spectrometry. The average mass of the SPase cleavage product of the -3 Trp β -lactamase mutant was 28,711 Da, compared with 28,943 Da for the processed WT β -lactamase

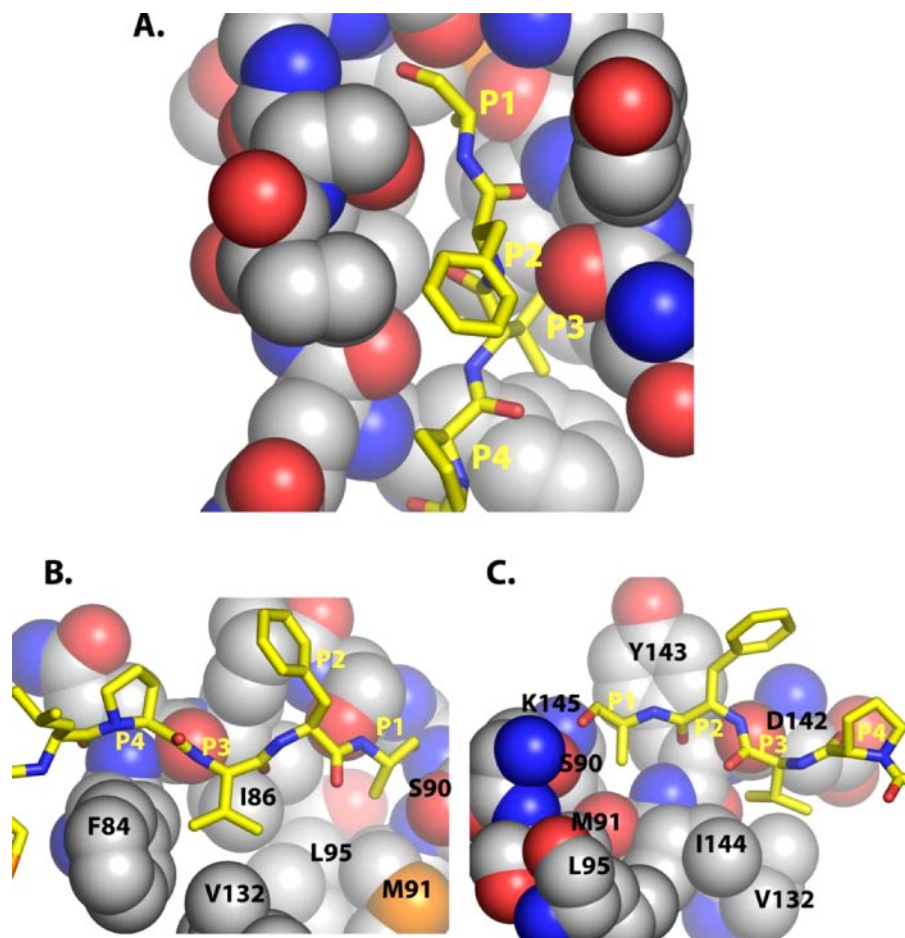


FIGURE 4. Modeled complex between wild-type SPase and wild-type β -lactamase signal peptide (cleavage recognition sequence, c-region). The residues of SPase that contribute atoms to the binding site are shown in van der Waals spheres (carbon, white; oxygen, red; nitrogen, blue). The signal peptide is rendered as stick (carbon, yellow; oxygen, red; nitrogen, blue). *A*, the top view looking down into the binding site. The P1–P4 residues of the signal peptide are labeled. The figure shows the P1 and P3 side chains pointing into the binding site and the P2 and P4 side chains pointing out into the solvent. *B*, the side view of the complex with residues Asp-142, Tyr-143, and Ile-144 of the SPase binding site removed to show the steric fit at the P3 position. *C*, the view from the opposite side of the binding pocket with residues Ser-88, Pro-87, Ile-86, Gln-85, and Phe-84 of the SPase binding site removed to display the steric fit at the P3 position of the peptide with the binding pocket.

(theoretical = 28,908 Da). The difference in the observed masses (232.4 Da) is consistent with the loss of His-Pro (234.3 Da), suggesting cleavage occurred after the NH_2 -terminal His-Pro sequence in β -lactamase. At this new processing site, the -3 residue is an alanine and the -1 residue is a proline. Proline is an acceptable -1 residue using the M13 procoat protein as a substrate (39).

The SPase I144C/I86T Mutant Can Cleave Pre- β -lactamase Mutants with Non-canonical -3 Residues—Because our data are consistent with cleavage at the alternative site after the $+2$ proline, we substituted β -lactamase Pro-2 with a phenylalanine to prevent cleavage at that position. As shown in Fig. 3A, the wild-type SPase could not process the -3 Arg P2F β -lactamase. Interestingly, the I144C/I86T SPase cleaves the -3 Arg P2F β -lactamase substrate, reproducing the substrate specificity determined by the library method. Similarly, the I144C/I86T SPase could process the -3 Gln β -lactamase mutant (Fig. 3B), but not the -3 Tyr β -lactamase mutant (Fig. 3C). The wild-type SPase could not process the -3 Gln β -lactamase (Fig. 3B). In a

control experiment, we confirmed that wild-type SPase could cleave the P2F β -lactamase (see Fig. 3D).

Whereas the I144C mutant failed to process the -3 Gln β -lactamase substrate (Fig. 3E), a small amount of processing was observed with the -3 arginine (Fig. 3F, see *asterisk*). Thus, whereas both Ile-144 and Ile-86 residues must be altered to allow processing with a glutamine residue at the -3 position, mutation of Ile-144 alone is sufficient to allow some cleavage of the -3 arginine substrate (Fig. 3F), although not as efficient as the double mutant (see Fig. 3A). We also tested whether the I144C/I86T mutant could process substrates with other basic amino acids at the P3 site. Fig. 3G shows that the double mutant was able to process the pre- β -lactamase with a -3 lysine residue, whereas no cleavage was observed with the wild-type SPase. Interestingly, we did not observe any substrates with a -3 lysine residue by the FRET signal peptide-substrate library (Table 1).

Molecular Modeling—Molecular modeling was performed to gain structural insight into the observed changes in SPase specificity. An energy minimized model of the wild-type SPase in complex with a wild-type β -lactamase signal peptide (Fig. 4) was compared with that of the double mutant (I86T/I144C) SPase in complex with a β -lactamase signal peptide with an arginine at the -3 position (Fig. 5).

Analysis of the surface by the program CASTp (33) of both the WT and mutant enzymes (with the signal peptides removed) identified a similar cleft that incorporates both the SPase S1 and S3 binding pockets. The wild-type SPase binding site includes atoms from residues (Fig. 4): Phe-84, Gln-85, Ile-86, Pro-87, Ser-88, Gly-89, Ser-90, Met-91, Leu-95, Val-132, Asp-142, Tyr-143, Ile-144, and Lys-145. This cleft covers 189.3 \AA^2 of surface area and has a volume of 222.6 \AA^3 . The same binding cleft in the double mutant (I86T/I144C) (Fig. 5) includes atoms from the same residues except it also includes atoms from residues Ile-101 and Val-103, which are exposed at the bottom of the deeper pocket in the double mutant (Fig. 5). The double mutant binding cleft has a surface area of 282.5 \AA^2 and a volume of 385.4 \AA^3 , which represents an increase in surface area of 93.2 \AA^2 and an increase in volume of 162.8 \AA^3 with respect to the wild-type structures. The energy minimization in the presence of the modeled β -lactamase signal peptide in the binding cleft results in a complex with no steric clashes. Other

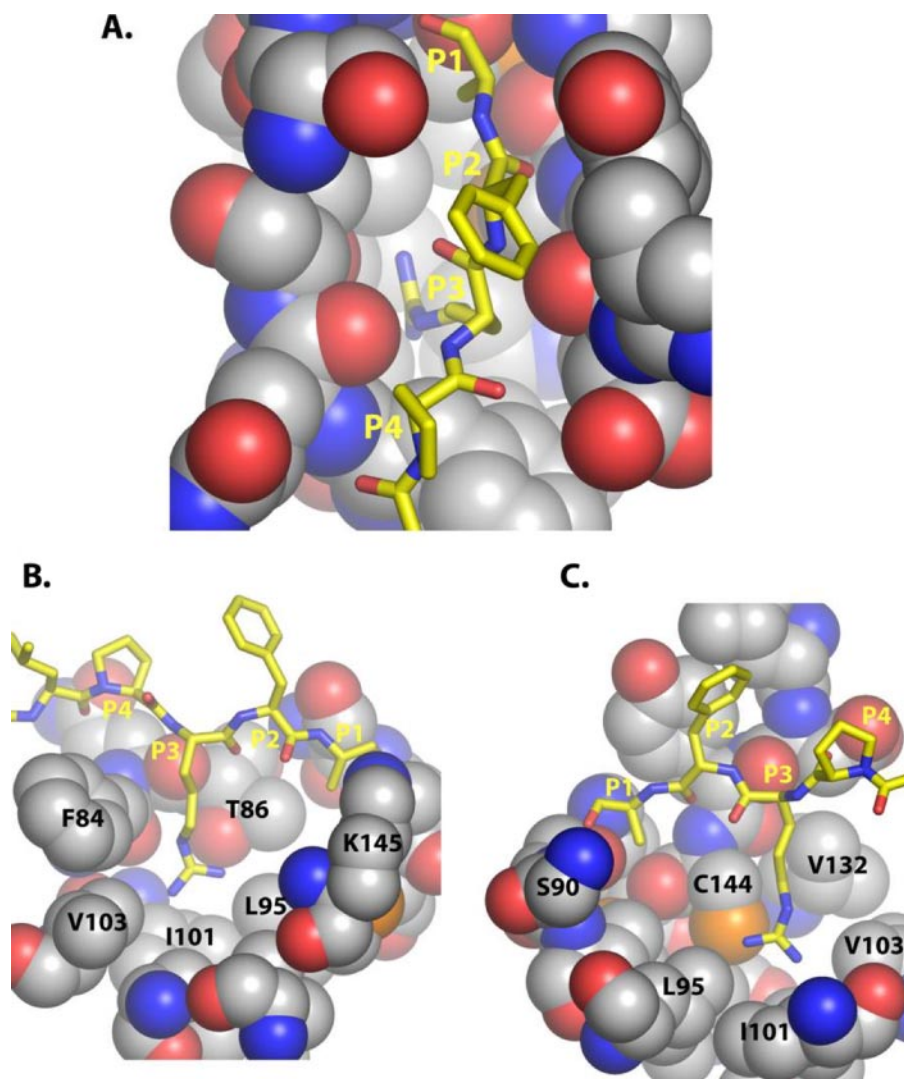


FIGURE 5. Modeled complex between SPase double mutant (I144C/I86T) and the β -lactamase signal peptide mutant $-3R$. *A*, the top view looking down into the binding site. The P1–P4 residues are labeled. The figure shows the P1 and P3 side chains pointing into the binding site and the P2 and P4 side chains pointing out into the solvent. *B*, the side view of the complex with residues Asp-142, Tyr-143, and Ile-144 removed to show the steric fit at the P3 position of the peptide with the binding pocket. *C*, the view from the opposite side of the binding pocket with residues Ser-88, Pro-87, Ile-86, Gln-85, and Phe-84 removed to display the steric fit at the -3 position of the peptide with the binding pocket. With the smaller side chains at the 86 and 144 positions in the SPase double mutant (I144C/I86T) the binding site has increased in size and is capable of accepting an arginine at the -3 position of a signal peptide. The model of the double mutant (I144C/I86T) SPase shows that residues Ile-101 and Val-103 make up part of the bottom of the binding side, which is not the case in the wild-type.

than the mutational differences only the side chains of Ile-101 and Val-132 showed any significant adjustment from the starting wild-type structure. The modeling shows that the binding site pocket of the double mutant can accommodate an arginine at the -3 position. The introduction of a thiol and hydroxyl group on opposing sides of the binding pocket could also function to make the binding of the guanidinium group of the arginine more energetically favorable.

DISCUSSION

In this study, we have identified residues Ile-86 and Ile-144 of the *E. coli* SPase as important determinants of its substrate specificity at the -3 position. Using the FRET library approach, we found that the single Ile-144 mutants could cleave peptide

substrates with valine, serine, leucine, glutamic acid, and aspartic acid at the -3 position, in addition to the alanine (Table 1). The I144C mutant could also cleave substrates with threonine, glycine, histidine, lysine, tryptophan, or tyrosine in addition to the other amino acids listed above at the -3 position. The wild-type SPase only cleaved peptides with alanine, valine, and threonine at the -3 position (Table 1), which are residues found at this position in pre-proteins (14, 15, 40). Strikingly, the I144C/I86T mutant tolerated a more diverse set of residues at the -3 position, including arginine and glutamine. This surprising finding was verified by the observation that this SPase double mutant cleaved the β -lactamase pre-protein with these atypical residues *in vivo* (Fig. 3). In addition, the finding that a -3 arginine was tolerated by the double mutant led us to predict that a -3 lysine would be tolerated as well (Fig. 3G). Molecular modeling studies (Figs. 4 and 5) reveal that the I144C/I86T mutant has a much larger and more polar pocket than the wild-type SPase that would allow it to accept a wider variety of -3 substrate residues.

This work also demonstrated the utility of our combinatorial library approach in defining the substrate specificity of endoproteases. Although a number of methods have previously been developed to identify optimal substrates including phage display (41), position-scanning library (42), and FRET assays (34), these methods each have some drawbacks. Briefly, the phage display method (41) is limited to the 20 natural amino acids found in proteins and cannot be used to identify the sites of cleavage. The position-scanning method of Ellman and co-workers (42), although able to evaluate both natural and unnatural amino acids, can only assess the specificity of the P residues of the substrate (not P' residues). Furthermore, it cannot give individual sequences or reveal any sequence coverage. The previous FRET method (34) can determine the preferred sequences that are cleaved by proteases but Edman sequencing is expensive and cannot determine the site of cleavage.

The FRET-based method described in this work is advantageous over all existing methods. It can evaluate both natural and unnatural amino acids and provides individual sequences, whereas covering both S and S' subsites. It is high

throughput and inexpensive because PED/MS can sequence up to 20–30 peptides in 1 h at a cost of less than \$1/peptide (cost in reagents and instrument time). In addition, our method can unambiguously identify the protease cleavage site. Hence, it should be generally applicable to any endoproteases.

REFERENCES

- Emr, S. D., Hanley-Way, S., and Silhavy, T. J. (1981) *Cell* **23**, 79–88
- Stanley, N. R., Palmer, T., and Berks, B. C. (2000) *J. Biol. Chem.* **275**, 11591–11596
- Wolfe, P. B., Silver, P., and Wickner, W. (1982) *J. Biol. Chem.* **257**, 7898–7902
- Wolfe, P. B., Wickner, W., and Goodman, J. M. (1983) *J. Biol. Chem.* **258**, 12073–12080
- Moore, K. E., and Miura, S. (1987) *J. Biol. Chem.* **262**, 8806–8813
- Whitley, P., Nilsson, L., and von Heijne, G. (1993) *Biochemistry* **32**, 8534–8539
- Black, M. T. (1993) *J. Bacteriol.* **175**, 4957–4961
- Tschantz, W. R., Sung, M., Delgado-Partin, V. M., and Dalbey, R. E. (1993) *J. Biol. Chem.* **268**, 27349–27354
- Paetzel, M., Strynadka, N. C., Tschantz, W. R., Casareno, R., Bullinger, P. R., and Dalbey, R. E. (1997) *J. Biol. Chem.* **272**, 9994–10003
- Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (1998) *Nature* **396**, 186–190
- Paetzel, M., Karla, A., Strynadka, N. C., and Dalbey, R. E. (2002) *Chem. Rev.* **102**, 4549–4580
- van Dijl, J. M., de Jong, A., Venema, G., and Bron, S. (1995) *J. Biol. Chem.* **270**, 3611–3618
- Chen, X., Van Valkenburgh, C., Fang, H., and Green, N. (1999) *J. Biol. Chem.* **274**, 37750–37754
- von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21
- Perlman, D., and Halvorson, H. O. (1983) *J. Mol. Biol.* **167**, 391–409
- von Heijne, G. (1985) *J. Mol. Biol.* **184**, 99–105
- Paetzel, M., Goodall, J. J., Kania, M., Dalbey, R. E., and Page, M. G. (2004) *J. Biol. Chem.* **279**, 30781–30790
- Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (2002) *J. Biol. Chem.* **277**, 9512–9519
- Karla, A., Lively, M. O., Paetzel, M., and Dalbey, R. (2005) *J. Biol. Chem.* **280**, 6731–6741
- 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**, 75–79
- Inada, T., Court, D. L., Ito, K., and Nakamura, Y. (1989) *J. Bacteriol.* **171**, 585–587
- Lessl, M., Balzer, D., Lurz, R., Waters, V. L., Guiney, D. G., and Lanka, E. (1992) *J. Bacteriol.* **174**, 2493–2500
- Dalbey, R. E., and Wickner, W. (1985) *J. Biol. Chem.* **260**, 15925–15931
- Johnston, S., Lee, J. H., and Ray, D. S. (1985) *Gene (Amst.)* **34**, 137–145
- Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J. C., Hochstrasser, D. F., and Appel, R. D. (1997) *Electrophoresis* **18**, 403–408
- Lam, K. S., Salmon, S. E., Hersh, E. M., Hrubby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) *Nature* **354**, 82–84
- Liu, R., Marik, J., and Lam, K. S. (2002) *J. Am. Chem. Soc.* **124**, 7678–7680
- Sweeney, M. C., and Pei, D. (2003) *J. Comb. Chem.* **5**, 218–222
- McRee, D. E. (1999) *J. Struct. Biol.* **125**, 156–165
- Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921
- DeLano, W. L. (2002) *PyMol*, 0.96 Ed., DeLano Scientific, San Carlos, CA
- Liang, J., Edelsbrunner, H., and Woodward, C. (1998) *Protein Sci.* **7**, 1884–1897
- Meldal, M., Svendsen, I., Breddam, K., and Auzanneau, F. I. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 3314–3318
- Rosse, G., Kueng, E., Page, M. G., Schauer-Vukasinovic, V., Giller, T., Lahm, H. W., Hunziker, P., and Schlatter, D. (2000) *J. Comb. Chem.* **2**, 461–466
- Stein, R. L., Barbosa, M. D., and Bruckner, R. (2000) *Biochemistry* **39**, 7973–7983
- Date, T. (1983) *J. Bacteriol.* **154**, 76–83
- Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) *J. Mol. Biol.* **340**, 783–795
- Shen, L. M., Lee, J. I., Cheng, S. Y., Jutte, H., Kuhn, A., and Dalbey, R. E. (1991) *Biochemistry* **30**, 11775–11781
- von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690
- Matthews, D. J., and Wells, J. A. (1993) *Science* **260**, 1113–1117
- Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7754–7759