Escherichia coli Signal Peptide Peptidase A Is a Serine-Lysine Protease with a Lysine Recruited to the Nonconserved Amino-Terminal Domain in the S49 Protease Family†

Peng Wang,‡,§ Eunjung Shim,‡,§ Benjamin Cravatt,|| Richard Jacobsen,⊥ Joe Schoeniger,⊥ Apollos C. Kim,# Mark Paetzel,# and Ross E. Dalbey*,‡

Department of Chemistry, The Ohio State University, Columbus, Ohio 43210, Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada, The Skaggs Institute for Chemical Biology, Department of Cell Biology and Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and Sandia National Laboratories, Livermore, California 94551-0969

Received April 14, 2008

ABSTRACT: *Escherichia coli* signal peptide peptidase A (SppA) is a serine protease which cleaves signal peptides after they have been proteolytically removed from exported proteins by signal peptidase processing. We present here results of site-directed mutagenesis studies of all the conserved serines of SppA in the carboxyl-terminal domain showing that only Ser 409 is essential for enzymatic activity. Also, we show that the serine hydrolase inhibitor FP-biotin inhibits SppA and modifies the protein but does not label the S409A mutant with an alanine substituted for the essential serine. These results are consistent with Ser 409 being directly involved in the proteolytic mechanism. Remarkably, additional site-directed mutagenesis studies showed that none of the lysines or histidine residues in the carboxyl-terminal protease domain (residues 326−549) is critical for activity, suggesting this domain lacks the general base residue required for proteolysis. In contrast, we found that *E. coli* SppA has a conserved lysine (K209) in the N-terminal domain (residues 56−316) that is essential for activity and important for activation of S409 for reactivity toward the FP-biotin inhibitor and is conserved in those other bacterial SppA proteins that have an N-terminal domain. We also performed alkaline phosphatase fusion experiments that establish that SppA has only one transmembrane segment (residues 29−45) with the C-terminal domain (residues 46−618) protruding into the periplasmic space. These results support the idea that *E. coli* SppA is a Ser-Lys dyad protease, with the Lys recruited to the amino-terminal domain that is itself not present in most known SppA sequences.

Almost all proteins that are exported to the cell surface of Gram-negative and Gram-positive bacteria are synthesized in a higher-molecular weight precursor form with an amino-terminal cleavable signal peptide. During protein export, the signal peptide is cleaved from the precursor by the membrane protease signal peptidase (*1*). The cleaved signal peptide can be further digested by signal peptide hydrolases (*2, 3*).

In 1984, the protease that degrades the lipoprotein signal peptide, named signal peptide peptidase A (SppA), was purified to homogeneity (*4*) and shown to correspond to the previously identified inner membrane protein, protease IV (*5*). SppA was subsequently cloned and sequenced, and the gene-encoded protein was shown to be a tetramer (*6*). Sequence analysis of *Escherichia coli* SppA reveals three hydrophobic regions that are candidate membrane-spanning domains. SppA is inhibited by serine protease inhibitors such as chymostatin, leupeptin, antipain, and elastinal (*4*). However, unlike typical serine proteases, the SppA family of proteins does not have a conserved histidine residue, suggesting SppA is an atypical serine protease.

The *E. coli* SppA protease has been classified as part of the S49 family of proteases in the MEROPS protease database (*7*), specifically to the S49.001 subfamily. This group, like subfamily S49.004, contains, in addition to the carboxyl-terminal protease domain that is conserved in all the S49 family members, an amino-terminal domain. The other subfamily members (S49.002, S49.003, S49.005, and S49.006) do not contain this amino-terminal domain. Members of these groups include the sohB peptidase (*8*), protein C (*9*), protein 1510-N (*10*), and the archaeal signal peptide peptidase (*11*), respectively.

Recently, site-directed mutagenesis has been utilized (*12*) to identify the catalytically important residues of the archaeal SppA protease from *Thermococcus kodakaraensis* which

† This work was supported by National Science Foundation Grant MCB-0316670 (to R.E.D.), the Canadian Institute of Health Research (to M.P.), a National Science and Engineering Research Council of Canada operating grant (to M.P.), a Michael Smith Foundation for Health Research Senior Scholar award (to M.P.), and a Canadian Foundation of Innovation grant (to M.P.).

* To whom correspondence should be addressed: Department of Chemistry, The Ohio State University, Columbus, OH 43210. Phone: (614) 292-2384. Fax: (614) 292-1532. E-mail: dalbey@chemistry.ohio-state.edu.

‡ The Ohio State University.

§ These authors contributed equally to this work.

|| The Scripps Research Institute.

⊥ Sandia National Laboratories.

Simon Fraser University.

10.1021/bi800657p CCC: $40.75
Published on Web 05/14/2008

contain only the C-terminal protease domain. They showed that Ser 162 and Lys 214 were critical for activity, suggesting that this protease may use a serine-lysine dyad for catalysis like signal peptidase I (*13*), LexA (*14*), UmuD (*15*), Tsp (*16*), Lon protease (*17*), and VP4 protease (*18*). However, *E. coli* SppA lacks a lysine at the position aligned to Lys 214.

In this report, we have employed site-directed mutagenesis to identify possible active site residues for *E. coli* SppA. We show within the carboxyl-terminal protease domain of *E. coli* SppA that Ser 409, which is homologous to the critical Ser 162 of *T. kodakaraensis* SppA (*12*), is essential for activity. Wild-type SppA, but not an S409 mutant where the serine is substituted for alanine, can be covalently modified with the FP-biotin serine hydrolase inhibitor. These results support the notion that Ser 409 is the active site residue in the *E. coli* enzyme. None of the conserved amino acid residues within the protease domain that have ionizable side chains, including all the histidine and lysine residues, are critical for activity. Strikingly, the invariant Lys 209 residue located in the amino-terminal domain is indispensable for activity and important for activation of S409. The alkaline phosphatase fusion approach suggests that *E. coli* SppA, which was predicted to span the membrane three times by most topology programs, spans the membrane only once with its carboxyl terminus localized to the periplasmic space. Taken together, the results provide evidence that the single-spanning *E. coli* SppA carries out catalysis using a serine-lysine dyad with the serine located in the conserved carboxyl-terminal protease domain and the lysine in the nonconserved amino-terminal domain.

EXPERIMENTAL PROCEDURES

Materials. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. *Pfu* DNA polymerase was obtained from Stratagene. Oligonucleotides were ordered from Integrated DNA Technologies, Inc., and the QIAprep Spin Miniprep Kit for the plasmid purification was purchased from Qiagen. TALON metal affinity resins were ordered from Clontech. *N*-Benzyloxycarbonyl-L-valine-*p*-nitrophenyl ester (Cbz-Val-ONP) was purchased from Research Organics.

Bacterial Strains and Plasmids. E. coli strains BL21(DE3), DH5α, and MC1061 [D*lacX74*, araD139, D(*ara, leu*)7697, *galU, galk, hsr, hsm, strA*] were from our collection. Lin 205 [*fhuA22, phoA8, ompF627(T_2^R), glpA24, fadL701(T_2^R), relA1, glpR8(glp^c), glpR7 (glp^n), glpD26, pit-10, spoT1*] was obtained from the *E. coli* Genetic Stock Center. The pCR 2.1 cloning vector was purchased from Invitrogen, and the pET-28(a) expression vector was from our laboratory. The pING-1 vector with the *araB* promoter and the *araC* regulatory elements was from our collection (*19*).

Cloning and Mutagenesis of E. coli SppA. The *sppA* gene was cloned by the PCR method into the pET-28(a) expression vector. The *E. coli* chromosomal DNA from MC1061 [D*lacX74*, araD139, D(*ara, leu*)7697, *galU, galk, hsr, hsm, strA*] was used as the template. The PCR was run for 32 cycles (95 °C for 1 min, 46 °C for 1 min, and 72 °C for 90 s) with the sense primer (5′-AAG TTG GGA GAA CAT ATG CGA ACC CTT TGG CG-3′) and the antisense primer (5′-TCA GTA CAA AAG CTT ACG CAT GTT GGC GCA GGT C-3′) for SppA. The underlined sequences specify the restriction sites for endonucleases *Nde*I and *Hin*dIII, respectively. The PCR product was cloned into the cloning vector, pCR 2.1 (ampicillin resistant). SppA was then subcloned into the *Nde*I and *Hin*dIII sites of the pET-28(a) expression vector (kanamycin resistant). The pCR 2.1 plasmids bearing the SppA gene were digested with *Nde*I and *Hin*dIII, and the DNA fragment containing *sppA* was isolated by excision from an agarose gel. The pET-28(a) vector was prepared in the same way. The ligation product was transformed into *E. coli* host BL21(DE3) cells. The resulting DNA was sent to the Plant-Microbe genomics sequencing facility at The Ohio State University for confirmation.

Site-directed mutagenesis was carried out with the SppA-pET-28(a) vector as a template using the QuickChange mutagenesis PCR method (Stratagene Inc.) to incorporate different amino acid residues. Mutations were verified by DNA sequencing.

Purification of E. coli SppA. E. coli BL21(DE3) cells harboring six-His-tagged SppA encoded by the pET-28(a) vector were used for overexpression of SppA. Wild-type SppA and its mutants were expressed and purified using cobalt affinity chromatography. A single colony was used to inoculate 100 mL of LB[1] broth containing 25 μg/mL kanamycin. Twenty milliliters of this overnight cell culture was diluted into 2 L of LB broth with the same concentration of kanamycin. The cultures were grown at 37 °C until the OD_{600} reached 0.6, at which time they were induced with 1 mM IPTG (final concentration). After further growth for 5 h, the cells were harvested by centrifugation at 5000 rpm for 10 min with a JA-10 rotor. The cells were resuspended in lysis buffer [1 M NaCl, 25% sucrose, and 10 mM sodium phosphate (pH 7.0)] and lysed by ultrasonication. After removal of unbroken cells, the membrane fraction was spun in the centrifuge at 45000 rpm for 3 h using a Ti-70 rotor. The membrane fraction was then resuspended in extraction buffer [0.5 M NaCl, 0.5% Triton X-100, 40% glycerol, and 10 mM sodium phosphate (pH 7.0)] and stirred at 4 °C until the pellet was completely homogenized. After the centrifugation step had been repeated once more, the supernatant was isolated and applied to the affinity column. The Triton X-100 extract was loaded on the column with a 5 mL bed volume of TALON cobalt resin, which was equilibrated with the extraction buffer. The column was washed with wash buffer [300 mM NaCl, 0.5% Triton X-100, 5 mM imidazole, and 50 mM sodium phosphate (pH 7.0)]. The His-tagged protein was eluted with elution buffer [300 mM NaCl, 0.5% Triton X-100, 5 mM imidazole, and 50 mM sodium phosphate (pH 7.0)] with an increasing concentration of imidazole (50, 100, and 150 mM). Following elution, proteins were assayed for purity using a 12% SDS−PAGE gel and selected fractions were dialyzed four times with 1 L of 50 mM sodium phosphate (pH 7.0) with 1% Triton X-100 buffer to remove the imidazole.

Assay of E. coli SppA Activity. The activity of SppA, wild type and mutants, was measured against the chromophoric substrate Cbz-Val-ONP as follows (*5*). A total of 10 μg of purified enzyme was preincubated with 0.2% Triton X-100 and 10 mM sodium phosphate (pH 7.2) for 10 min at room

[1] Abbreviations: ECL, enhanced chemiluminescence; IPTG, isopropyl β-D-thiogalactopyranoside; LB, Luria broth; M9, minimal medium; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate.

A

```
Eco             MRTLWRFIAGFFKWTWRLLNFVREMVLNLFFIFLVLVGVGIWMQVSGGD-SKETASRGALLLDISGVIVDKPDSSQR 76
Sty             MRTLWRFIAGFFKWTWRVLNFVREMVLNLFFIFLVLVGVGIWMQIGNGS-NSEQTARGALLLDISGVIVDKPSTNHR 76
Vch             MKSLFRFVGLILKGIWKAITFIRLALTNLIFLLSIGIIYFIYVHADAPL-PTMDKSS-ALVLNLSGPIVEQSTHINP 75
Hin             -------MFQVLKFCWKVLCFIRDLVMNVVFLGFVLLLVAIISFSSGGKKSTALTSEGALLLNLDGYLADNRDETLR 70
Cte             ---------MADKNRKKGGCFRAGCLTAVVAVLLLVGLGGWFVHQRSNR----LPARFVLSVPLTGELDERPPDAGP 64
Tko             -------------------------------------------------------------------------------

Eco             FSKLSRQLLGASSDRLQENSLFDIVNTIRQAKDDRNITGIVMDLKNFAGGDQPSMQYIGKALKEFRDSGKPVYAVGE 153
Sty             LGALGRQLFGASSDRLQENSLFDIVNAIRQAKDDRNITGIVLDLKNFTGADQPSMRYIGKALREFRDSGKPVFAVGE 153
Vch             MDSFTGSVFGEELP--RENVLFDIVETLRHAKNDNNVTGLVLALGDMPETNLTKLRYIAKAINEFKASGKPVFAVGD 150
Hin             WQDALSELNGEHVP--RKISTFDVVFAIQQAEDDPKIKGLVLDLNYFEGADLPALDFIGGAISHFKDAGKPVIAYAD 145
Cte             FPFGKGRHL---------LSFEELLTILDRAKTDRRVDSVLLRIDGLG-ASPAKIQELRSSIAALRKSGKKVTAFLV 132
Tko             -------------------------------------------------------------------------------

Eco             NYSQGQYYLASFANKIWLSPQGVVDLHGFATNGLYYKSLLDKLKVSTHVFRVGTYKSAVEPFIRDDMSPAAREADSR 230
Sty             NYSQGQYYLASFANKIWLSPQGQVDLHGFATNGLYYKTLLDKLKVSTHVFRVGTYKSAVEPFIRDDMSPAAREADSR 230
Vch             FYNQSQYYLASYADKIYLAPDGAVLLKGYSAYSMYYKTLLEKLDVTTHVFRVGTYKSAIEPFVRDDMSDAARESASR 227
Hin             NYSQGQYYLASFADEIYLNSIGSVDIHGLSQENLYFKEMLDKLAVTPHIFRVGTYKSAVEPFLRNDMSAEAKANMQR 222
Cte             TPEDKDYQLAVACDSIIVQKGSWMTLDGLKAELFFVADPLKKLGVSFQAAQWKKYKSAVETFTRNSASPENLEETNA 208
Tko             ----------------------------------MDDRIWKYVSAVLTLLLALSIVSVAVLYYQVNPPAVPANQTG 42
                                                   : *   :           ::  :  :  .

Eco             WIGELWQNYLNTVAANRQIPAEQVFPGAQGLLEGLTKTGGDTAKYALENKLVDALASSAEIEKALTKEFGWSKTDKN 307
Sty             WIGELWQNYLHTVSANRQISPQQLFPGAQAIIDGLTSVGGDTAKYALDHKLVDALASSADVEKALTKQFGWSKTENN 307
Vch             WLTQLWSAYVDDVAANRQIEIKTLTPSMEQFVAQLKEVNGDLAALSKKVGLVDELATRQQVRQTLAETFGSDGKDS- 303
Hin             WLGEMWNNYVLSVSENRNIKKDRILPNAKQYLAELKALKGNSTAYAQQRGLVTDVVTRLDLDKKLSALFG-KGSDGK 298
Cte             LLDDAWSDYLDSVSRQRRIG-------KDAFRKVVDSLAVLTPEKALGLHLIDRVATERELEQEYARRLNKPAAELL 278
Tko             FVIETPSNFTLVCEG------------------SVNTTSTEVAELRDQIAFLQRLVSSNQGG--------------- 86
                 : :  . :                        :       .       ::  :.:  :

Eco             YRAISYYDYALKTPADTG-----DSIGVVFANGAIMDGEETQGNVGG----DTTAAQ 355
Sty             YRAISYYDYSLKTPADTG-----GTIAVIFANGAIMDGEETPGNVGG----DTTASQ 355
Vch             YNAIGYYEYKTTIKPTTLTD--ANDIAVVVASGAIMDGSQPRGTVGG----DTVAGL 354
Hin             ANLIEFDDYLTQLPDRLEHYNVPNKIAVVNVEGTIIDGESDEENAGG----DTIARI 351
Cte             VGGREYLGATGGMRPQGG----GDRIAVINITGMIVSDGAGGMSEGDGTDVATVKEA 351
Tko             ----------------------TTIAVVPIFGIIDDYTALQ----------VIPL 109
                                     *.*:   * * .
```

B

```
Eco             IRDARLDPKVKAIVLRVNSPGGSVTASEVIRAELAAARAAGKPVVVSMGGMAASGGYWISTPANYIVANPSTLTGSIG 433
Sty             IRDARLDPKVKAIVLRVNSPGGSVNASEVIRAELAAARAAGKPVVVSMGGMAASGGYWISTPANYIVASPSTLTGSIG 433
Vch             LREARNDSNVKAVVLRVDSPGGSAFASEVIRNEIEALKAAGKPVVVSMSSLAASGGYWISMSADKIVAQPTTLTGSIG 432
Hin             LRKAHDDNSVKAVILRVNSPGGSAFASEIIRQETENLQKIGKPVIVSMGAMAASGGYWISSTADYIIADSNTITGSIG 429
Cte             LQTAIDDLKVKAIVLRIDSPGGDALAASTMLELLNEAKAK-KPIVASMSGLAASGGYMVALAGDKIFAEPLTITGSIG 408
Tko             LRNLAMNESVGGVLLWIESPGG-VVGPVINIHSEIKKLSLVKPVVAYSGDIIASGGYYIAVGAQKIVASPLAEVGSIG 186
                ::    : .* .::* ::**** . ..              **::.  . : ***** ::  .: *.*.. : .****

Eco             IFGVITTVENSLDSIGVHTDGVSTSPLADVS-ITRALPPEAQLMMQLSIENGYKRFITLVADARHSTPEQIDKIAQGH 510
Sty             IFGVINTVENSLSSIGVHSDGVSTSPLADIS-MTKALSPEVQQMMQLSIEYGYKRFITLVADARKRTPEQIDKIAQGH 510
Vch             IFSVITTFEKGLNNLGIYTDGVGTTPFSGQG-LTTGLTQGAKDAIQLGIEHGYQRFISLVAEKRGLTLKAVDELAQGR 509
Hin             IFTMFPTFENSIKKIGVHADGVSTTELANTS-AFSPLAKPVQDIYQTEIEHGYDRFLEIVSKGRQLSKTQVDKLAQGQ 506
Cte             VFSLKPDLSSLLEKTGIRREVLIRGRFADAETPFRAFDDASFRKFVELTGTVYEDFIAKVAKGRHMTPAQVDAVAGGR 486
Tko             VIYVHYDLEKNYEMNGIKVNVFKTGKHKDMGAEWRDLTPEEREKITEMVNTYFQAFISAVSEGRNMTIDEVKNFSTGE 264
                :: :   ...  .  *:  : .      .       :               :. *:  *:. *  :   :. .: *.

Eco             VWTGQDAKANGLVDSLGDFDDAVAKAAELAKVKQWHLEYYVDEPTFFDKVMDNMSGSVRAMLPDAFQAMLPAP---LA 585
Sty             VWTGEDAKANGLVDSLGDFDDAVAKAAELAKLKQWHLDYYQDEPTVLDMVMDSMTGSVRAMLPETIQAMLPAP---LV 585
Vch             VWTAQDAQTLGLVDQLGDFDDAVHLAADLAQLDQYNLYWVEEPLTPAQQFLQDLLGQVRVSLGLDVSTLLPKS---LQ 584
Hin             VWLGSDAFQNGLVDEIGSFNEAVNKAEQLVNQRQDTAVQDFSVEWFTDDNVSLISTLLSDTKKGAQEQLVKWLG--LP 582
Cte             VWSGKRALEVGLIDQIGGLGDAVQEAKKLAGVKKEVKPALLYLPAQKTWLEYLLGSDTSDMVSALTAAMVRESLGQLQ 564
Tko             TWFAENVTG-ALVDELGGMDTAIDVLEKLMNVSGAKVVVYKDLETPEEFGVYGSTALYIDPRYLTPLIGGG------- 334
                .* .. .   .*:*.:*.:. *:    .*

Eco             SVASTVKSESDKLAAFNDPQNRYAFCLTCANMR 618
Sty             SAANTVKAEGDKLAAFNDPQNRYAFCLTCANVR 618
Vch             PLAVEWQQQTSLLNQLNDPKGQYAFCLPCQVE- 616
Hin             APIQKLQKELNILTKFNDPKGQYLYCLNCGKVK 615
Cte             PLSRLPGAGIARFLLRTDAPQVLALDPVEVVIK 597
Tko             ---------------------------------
```

FIGURE 1: Alignment of the amino acid sequences of SppA protease members. (A) Amino-terminal domain of SppA proteases within the S49 family. (B) Carboxyl-terminal protease domain within the S49 family. Shown are the sequences for *E. coli* (Eco), *Salmonella typhimurium* (Sty), *Vibrio cholerae* (Vch), *Haemophilus influenzae* (Hin), *Chlorobium tepidum* (Cte), and *T. kodakarensis* (Tko). Shown below are the consensus motifs for aligned SppA proteases. An asterisk denotes that the amino acids are identical in all sequences in the alignment. A colon means that conserved substitutions in the aligned sequences have been observed, according to the nature of the amino acid (hydrophobic, hydrophilic, acidic, or basic). A period means that semiconserved substitutions in the aligned sequences are found. The shaded residues in the carboxyl-terminal domain are invariant in all protease members in the S49 family. The shaded residue in the amino-terminal domain is absolutely conserved when the amino-terminal domain is present in the S49.001 group.

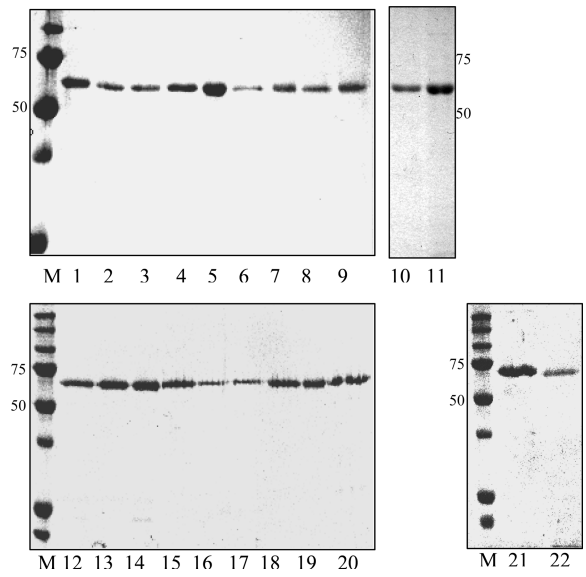

FIGURE 2: SDS−polyacrylamide gel electrophoresis analysis of the purified wild-type and mutant SppA. The purified enzymes were run on a 12% SDS−polyacrylamide gel and stained with Coomassie Blue R-250: lane M, markers; lane 1, WT; lane 2, S374A; lane 3, S402A; lane 4, S409A; lane 5, S431A; lane 6, R496K; lane 7, R496A; lane 8, K366A; lane 9, K397A; lane 10, G410A; lane 11, Y412F; lane 12, H451A; lane 13, H497A; lane 14, H510A; lane 15, K364A; lane 16, K535A; lane 17, K486A; lane 18, K505A; lane 19, K518A; lane 20, K541A; lane 21, K209A; lane 22, D524N.

temperature, and then 1 μL of 100 mM Cbz-Val-ONP was added. The reaction mixture (total volume of 1.0 mL) was incubated for 10 min at room temperature. The rate of *p*-nitrophenol release was monitored at 400 nm on a Perkin-Elmer UV−vis spectrophotometer.

FP-Biotin Reaction. A biotinylated fluorophosphonate (FP-biotin) chemical reagent was used to inactivate SppA. The reaction of wild-type, S409A, K209A, K366A, and H510A SppA with FP-biotin was performed as described by Liu et al. (*20*) with the following modifications. Twenty microliters of 0.1 mg/mL SppA was treated with 40 μM FP-biotin (final concentration) for 30 min at 37 °C. The reaction mixture was then divided into two aliquots. One aliquot was used for the activity assay. The other aliquot was used for the immunoblotting study after the reaction was quenched by the addition of 2× SDS−PAGE loading buffer. The sample was analyzed on a 12% SDS−PAGE gel, and the biotin-modified SppA was detected using the ECL Western blot detection kit (Amersham Biosciences). The SppA modified with biotin was identified with an avidin−horseradish peroxidase conjugate.

Construction of SppA−PhoA Fusion Proteins. To produce the plasmid containing the full-length *sppA−phoA* fusion gene, the *sppA* gene was excised from *sppA*-pET-28(a) by digestion with *Nde*I and *Not*I. A pING plasmid bearing a gene encoding the mature PhoA fusion protein was digested by the same restriction enzymes. The two were then ligated to yield the full-length SppA−PhoA construct. Site-directed mutagenesis was used to create a series of SppA−PhoA fusion proteins in which various lengths of the 3′ end of the s*ppA* gene were deleted. For a negative control, we used a pING construct in which PhoA was attached immediately after the first transmembrane segment (H1) of the leader peptidase in the cytoplasmic domain (*21*).

Alkaline Phosphatase Assay of the SppA−PhoA Constructs and Expression. The alkaline phosphatase activity of the fusion proteins was determined by a plate assay and by a spectroscopic enzyme assay. For the plating assay, cells transformed with the appropriate pING-derived plasmids encoding the fusion protein were grown overnight on a LB agar plate supplemented with 100 μg/mL ampicillin, 0.2% arabinose, and 40 μg/mL 5-bromo-4-chloro-3-indolyl phosphate *p*-toluidine salt. The enzyme assay was performed as described by ref *22*. The cell culture were grown overnight in LB medium supplemented with 100 μg/mL ampicillin, then back diluted to 2 mL of LB medium, and grown at 37 °C until the OD_{600} reached 0.3. The cells were induced with 0.2% arabinose and grown for an additional 2 h at 37 °C. The OD_{600} of the cells was measured, and the cells were placed on ice to inhibit further growth. The 1 mL cell culture sample was pelleted and washed once with 1 mL of LB medium and 1 mL of 1 M Tris-HCl (pH 8.0). The cells were permeabilized by adding one drop of 0.1% SDS and three drops of chloroform and incubating them at 37 °C for 5 min, at which time the OD_{420} was measured. The reaction was initiated by adding 10 μL of *p*-nitrophenyl phosphate (40 μg/mL) and incubating the mixture at room temperature for 5 min. Absorption was measured using a Lambda 20 UV−vis spectrometer (Perkin-Elmer), and the alkaline phosphatase activity was determined by using the following equations: AP total activity = ΔOD_{420}/5 min × 1000, and normalized AP activity = (AP total activity)/(OD_{600} of cells).

To measure the steady state cellular levels of the SppA−PhoA fusion proteins, immunoblot analysis was performed using anti-PhoA antiserum. Cells were normalized from 1 mL cultures and added directly to SDS−PAGE loading buffer. PhoA was detected by immunoblotting using the ECL Western blot detection kit (Amersham Biosciences).

RESULTS

Elucidating the Catalytically Important Conserved Residues of E. coli SppA. The serine protease SppA has a conserved carboxyl-terminal protease domain found in all S49 family members (Figure 1B) and a nonconserved amino-terminal domain (Figure 1A) found in the S49.001 and S49.004 groups. Recently, the archeon *T. kodakaraensis* SppA has been classified as a serine-lysine dyad protease (*12*). However, while the Ser residue found in the *T. kodakaraensis* lines up with the Ser 409 residue in *E. coli* SppA, the *E. coli* enzyme does not have a lysine at the homologous position as the archaeal protease (Figure 1B, *T. kodakaraensis* Lys 214). There are also no conserved histidine residues in the SppA family of proteases (Figure 1; see sequence alignment), which could function as the general base in the Ser/His/Asp triad of conventional serine proteases. However, Arg 496 and Asp 524 are absolutely conserved in the S49 family (Figure 1B) and could be possible members of a variant catalytic triad or a novel serine-arginine dyad.

To identify the serine residue that could function as the active site nucleophile in *E. coli* SppA, we mutated the highly conserved Ser 374, Ser 402, Ser 409, and Ser 431 residues individually to Ala residues by site-directed mutagenesis. Mutant and wild-type sequence constructs with N-terminal six-His tags were cloned into a pET-28a vector, and high

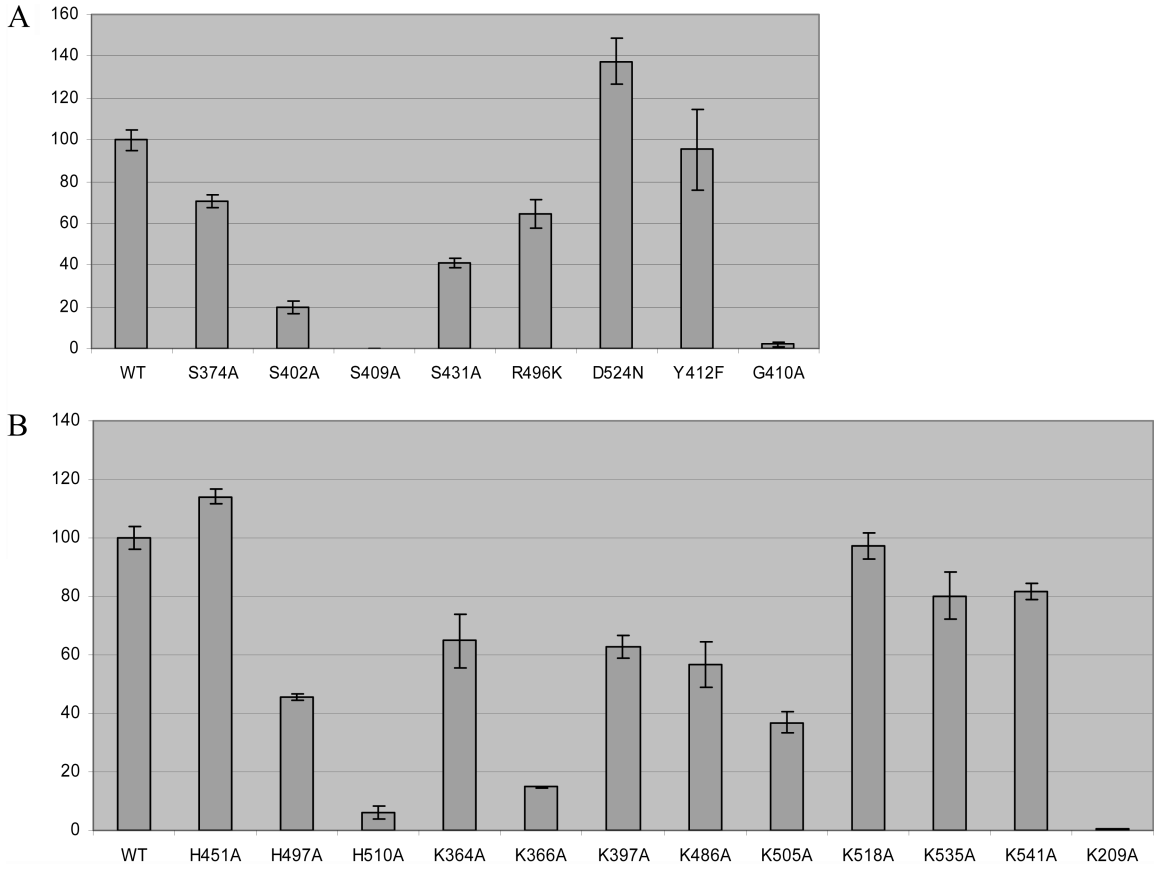

FIGURE 3: Activities of wild-type SppA and its mutants. The relative activity of each mutant was calculated by measuring the hydrolysis of the Cbz-Val-OPN substrate, as described in Experimental Procedures. The activity level of wild-type SppA was designated as 100%. The error bars indicate the standard deviation in triplicate measurements.

levels of expression were observed. Expressed proteins were purified to homogeneity (Figure 2) using cobalt affinity chromatography (see Experimental Procedures). Protease activities were assayed using the chromophoric substrate *N*-benzyloxycarbonyl-L-valine *p*-nitrophenyl ester (Cbz-Val-ONP) (Figure 3). With the wild-type SppA protein, hydrolysis of the Cbz-Val-ONP substrate was rapid and linear with time (data not shown). Activity was also observed with the Ser 374, Ser 402, and Ser 431 to alanine mutants. In contrast, the Ser 409 mutant had background activity, indicating the Ser 409 is critical for enzymatic activity.

To assess the role in proteolysis of the invariant Arg 496 and Asp 524 amino acids of SppA, we altered these residues. Figure 3 shows that R496 is not required for SppA activity; substitution with lysine (Figure 3A) or alanine (data not shown) does not lead to a marked reduction in activity. Neither is Asp 524 important for activity as the SppA D524N mutant had no reduction in activity (Figure 3A). In fact, the activity of this mutant was reproducibly higher than that of the wild-type SppA enzyme.

We also investigated the highly conserved Tyr 412 for its requirement for activity since it could act as a general base. Figure 3A shows that Y412F SppA was fully active, ruling out a role of this residue as a general base. Interestingly, the Gly 410 adjacent to the catalytically important Ser 409 is critical for optimal activity. SppA with the glycine mutated to alanine severely perturbs the SppA activity (Figure 3A). Substitution of the nonconserved histidine at positions 451 and 497 had only a small effect on activity, whereas mutation of histidine at position 510 impairs activity roughly 20-fold, although it still exhibits measurable activity (Figure 3B). The data together indicate that there is no histidine in the C-terminal domain of SppA that is essential for activity.

Since *T. kodakaraensis* SppA has been proposed to use a serine-lysine mechanism, we assessed the catalytic role of lysine residues in *E. coli* SppA by mutating the C-terminal domain Lys 364, Lys 366, Lys 397, Lys 486, Lys 505, Lys 518, Lys 535, and Lys 541 residues to Ala. As one can see in Figure 3B, SppA enzymes with these single mutations all maintained activity, indicating these residues are not essential for catalysis. The K366A mutation had the greatest effect on catalysis, although it was only 5-fold. These results rule out the possibility that a lysine residue within the carboxyl-terminal domain is the general base residue.

We next tested whether the absolutely conserved Lys 209 in the amino-terminal domain of *E. coli* SppA is crucial for activity (Figure 3B). Strikingly, mutation of Lys 209 to alanine inactivates the SppA protease. The data show that Lys 209 in the amino-terminal domain is essential for the proteolytic reaction and is consistent with the *E. coli* SppA employing a lysine in catalysis.

Inhibition of SppA by FP-Biotin. To provide further evidence that Ser 409 is the active site serine residue in the *E. coli* SppA enzyme, we tested whether wild-type SppA, unlike the Ser409Ala mutant, is modified with the biotinylated fluorophosphonate inhibitor (FP-biotin), which has been shown to inhibit serine hydrolases (*20*). First, we tested whether the reagent FP-biotin inhibited the SppA enzyme. Figure 4A shows that the addition of FP-biotin significantly decreased the activity of the wild-type enzyme compared

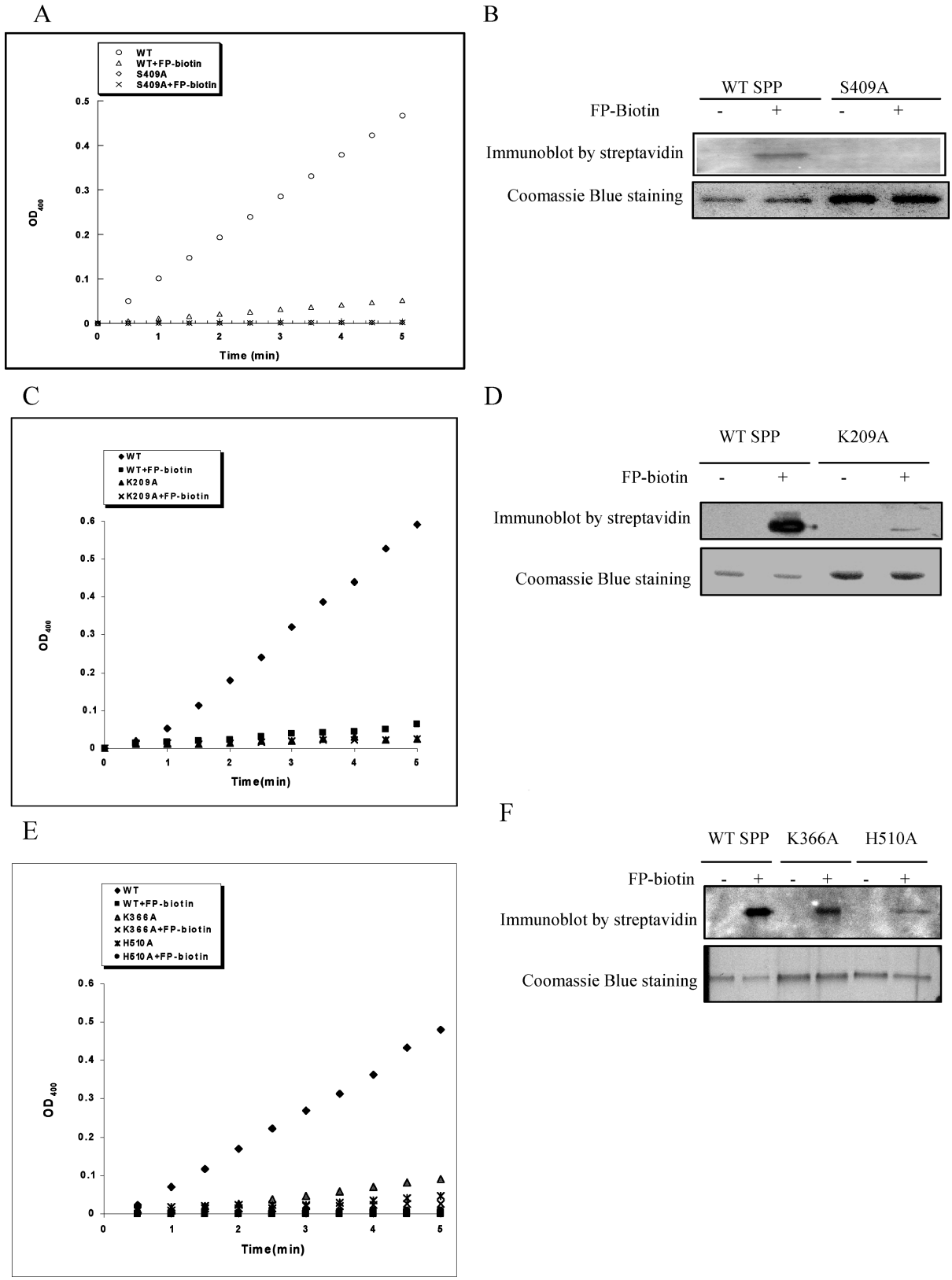

FIGURE 4: Inhibition and modification of wild-type SppA with FP-biotin. Purified wild-type, S409A, K209A, K366A, and H510A SppA proteins were incubated in the presence or absence of FP-biotin (final concentration of 40 μM) for 10 min, followed by activity assays (A, C, and E), or analyzed by SDS−PAGE and immunoblotting with an avidin−horseradish peroxidase conjugate (B, D, and F). Note that in panel F, only one-third of the wild-type SppA sample was used, in comparison to the K366S and H510A samples.

with the mock treated enzyme. Next, we examined whether the FP-biotin treated enzyme is modified with the inhibitor but remains unchanged when the candidate active site Ser 409 residue is mutated. As one can see in Figure 4B, a streptavidin reactive band is detected on the immunoblot with wild-type SppA, but not with the S409A SppA mutant. These results support the notion that serine 409 is indeed the active site residue. Likewise, we examined whether Lys 209 is critical for activation of the Ser such that it can react with the FP-biotin inhibitor. Figure 4D shows that the K209A

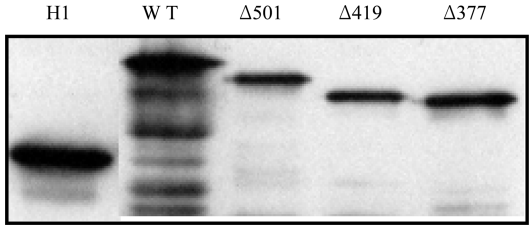

FIGURE 5: Expression of SppA−PhoA fusion proteins in the cell. Lin 205 cells expressing the SppA−PhoA constructs on the pING plasmid were grown and analyzed as described in Experimental Procedures. The cellular levels of the fusion proteins were determined using anti-PhoA antiserum.

Table 1: Characteristics of SppA−PhoA Fusion Proteins

	OD_{600}	total AP activity	normalized AP activity	predicted location
full-length (WT)	0.7877	17	21.5	periplasm
501	0.6672	19	28.5	periplasm
419	0.7647	12.9	16.9	periplasm
377	0.8289	19.3	23.3	periplasm
H1 (negative control)	0.7303	2.7	3.7	cytoplasm

SppA mutant only poorly reacts with the inhibitor in comparison to wild-type SppA. Since K366A and H510A exhibited low activity in Figure 3, we also included these mutants in the FP-biotin reaction assay. Both SppA mutants reacted with the FP-biotin inhibitor, although poorly, compared to the wild-type protein (Figure 4F). Note only one-third of the WT SppA sample was loaded onto the gel for immunoblot analysis compared to K366A and H510A samples (Figure 4F).

Topology of E. coli SppA. Hydropathy analysis and topology programs do not provide a clear indication of the membrane orientation of *E. coli* SppA. The topology program TMHMM predicts one transmembrane segment from residue 21 to 43 (*23*), while TopPred2 predicts residues 29−45, 398−414, and 421−441 are transmembrane segments (*24*). Both programs predict the amino terminus of SppA is localized to the cytoplasm and the carboxyl terminus protrudes into the periplasm.

The membrane topology of SppA was experimentally probed using the well-established PhoA fusion method (*25*). PhoA is enzymatically active only when it is exported to the periplasm; cytoplasmic PhoA is inactive. Therefore, fusion of PhoA to periplasmic domains of SppA would result in high PhoA activity, whereas fusion to cytoplasmic loops of SppA would result in low PhoA activity. To determine whether the C-terminal domain transmembrane segments predicted by TopPred2 were real, the PhoA mature sequence was fused to SppA after residues 377, 419, 501, and 618. As a negative control, we used a construct in which PhoA was fused after the first hydrophobic domain of the leader peptidase. Previous studies with the H1Lep−PhoA fusion (H1) showed that PhoA is localized primarily to the cytoplasm, although the first apolar domain (without the carboxyl-terminal positively charged residues present) can export a small percentage of PhoA to the periplasmic space (*21*). All the PhoA fusion constructs were expressed in the $PhoA^-$ strain to determine the expression level and the PhoA activity of the PhoA fusion proteins. Figure 5 confirms by immunoblotting that the various PhoA constructs are expressed in the cell. We measured the activity of PhoA in liquid using the chromophoric substrate *p*-nitrophenyl phosphate. Table 1 summarizes the PhoA activities of the SppA fusion proteins. All the SppA fusion proteins gave high enzymatic activities of alkaline phosphatase: the normalized AP activities are 21.5, 28.5, 16.9, and 23.3 units for the 618 (full-length), 501, 419, and 377 SppA−PhoA constructs, respectively. The negative control (H1) gave a normalized AP activity of 3.7. The results are consistent with all regions of the C-terminal domain of SppA being periplasmic. Because there are no other candidates for TM segments other than the one at the beginning of the protein, the data support a model in which SppA spans the membrane one time with its large carboxyl-terminal domain and most of its N-terminal domain protruding into the periplasmic space.

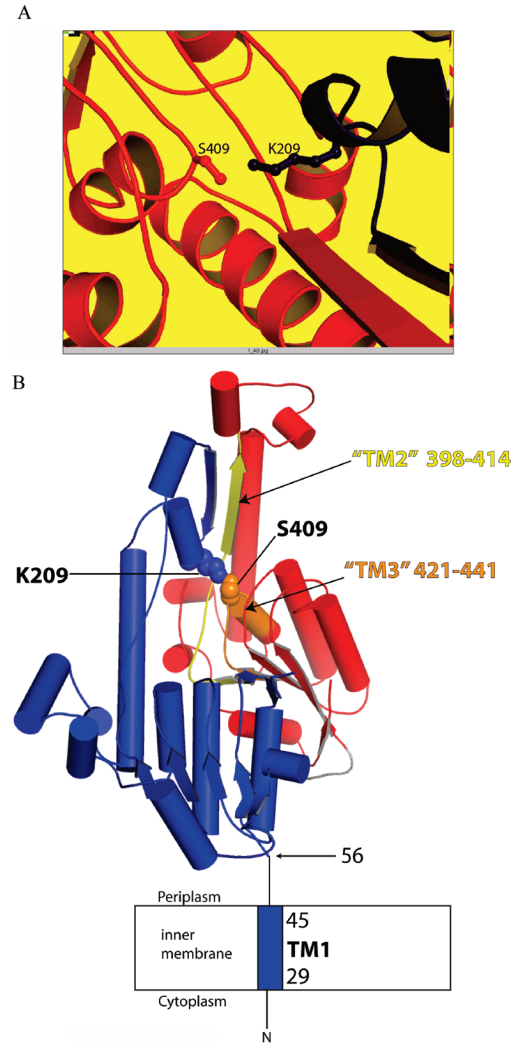

FIGURE 6: SppA is a Ser-Lys dyad protease and spans the membrane once with a large carboxyl-terminal domain exposed to the periplasmic space. (A) Ribbon diagram of *E. coli* SppA highlighting the active site Ser and Lys residues. Panel A was adapted from ref *26*. (B) A ribbon diagram of a SppA monomer (for clarity, the other three monomers of the SppA tetrameric structure are not shown). The N-terminal domain (residues 56−316) is colored blue. The C-terminal domain (residues 326−549) is colored red. The linker region (residues 317−325) is colored white. The previously predicted transmembrane domain 2 (TM2) is colored yellow (residues 398−414). The previously predicted transmembrane domain 3 (TM3) is colored orange (residues 421−441). The serine nucleophile Ser 409 and lysine general base Lys 209 are shown as spheres and meet at the interface of the N-terminal and C-terminal domains. Residues 20−45 are shown schematically, not to scale.

DISCUSSION

In this report, we present two pieces of data that suggest that Ser 409 is directly involved in catalysis of *E. coli* SppA. The first datum is the fact that Ser 409 of *E. coli* SppA, which is homologous to the proposed active site serine residue in *T. kodakaraensis* SppA (*12*), is critical for enzymatic activity. Second, the FP-biotin serine hydrolase inhibitor modifies wild-type SppA but does not label the protein when Ser 409 is mutated, suggesting that Ser 409 is modified by the inhibitor (Figure 4).

Surprisingly, we found that Lys 209 is the best candidate for a general base residue for the *E. coli* SppA peptidase. This Lys 209 residue is localized to the nonconserved amino-terminal domain of *E. coli* SppA peptidase. All other candidates for a general base residue were found not to be essential for catalytic activity (Figure 3). Also we found that the mutagenesis of Lys 209 to an alanine markedly weakened the ability of the SppA to be modified with the FP-biotin inhibitor (Figure 4). This is consistent with Lys 209 being important for activation of the Ser 409 nucleophile such that it can be reactive toward substrates and inhibitors.

A serine and lysine residue is also implicated in the catalytic mechanism from the X-ray structure of *E. coli* SppA that was determined in the course of this work (Figure 6; *26*). The 2.4 Å structure revealed that Ser 409 and Lys 209 are indeed active site residues for the *E. coli* SppA peptidase. The structure reveals that the Oγ atom of Ser 409 is within hydrogen bonding distance of the Nζ atom of Lys 209 and that there are no other titratable functional groups within the vicinity of the Oγ atom of Ser 409 other than the Nζ atom of Lys 209. In fact, given that there is no sequence similarity and no similarity in protein fold, the active site architecture of *E. coli* SppA is strikingly similar to that of the *E. coli* signal peptidase. It appears that bacteria have converged on the same catalytic mechanism (Ser/Lys) to both cleave off the signal peptide (*13*) and hydrolyze the remaining signal peptide. In eukaryotic organisms, these same processes are catalyzed by a Ser/His/Asp protease [within the signal peptidase complex (*27*)] and an aspartic protease mechanism [within the signal peptide peptidase (*28*)].

These data, along with recent studies (*12, 29*), suggest that all SppA proteases are serine-lysine dyad peptidases. Members of this family of proteases all contain an active site Ser in the protease domain. They differ only in the location of the critical lysine residue. Members of two of the subfamilies (S49.001 and S49.004) have a highly conserved lysine in the amino-terminal domain (Figure 1A) that we showed (Figure 3) is crucial for the activity of *E. coli* SppA. Members of the other subfamilies (S49.002, S49.003, and S49.006) lack most of this domain and contain an absolutely conserved Lys residue in the protease domain located 52 residues (except for the 549.005 subfamily which has Lys 41 residues downstream from the Ser) downstream from the invariant Ser [see the MEROPS protease database (*7*)]. Both the lysine and serine residues were shown to be critical for the activity of *T. kodakaraensis* SppA (S49.006) (*12*), and the structure of protein 1510-N from *Pyrococcus horikoshii* (S49.005) revealed a Ser and a Lys residue at the active site region (*29*).

In our mutagenesis work on *E. coli* SppA, we also discovered that mutation of G410, S402, H510, and K366 impairs activity (Figure 3). The structure of *E. coli* SppA shows that the G410 main chain nitrogen contributes to the oxyanion hole (*26*), which explains its importance in activity. If G410 is changed to any other residue, it would sterically compete for space with the carbonyl oxygen of the scissile bond. The structure of the SppA active site reveals that K209 may achieve the depressed pK_a that is required for it to function as a general base at biological pH by providing a hydrophobic microenvironment that shields the lysine from water. His 510 is located close to the K209 residue and has its polar side chain pointing away from the catalytic dyad. Changing H510 to a smaller residue may weaken the ability of the residues to shield the dyad from water, thus impairing activity. It is not clear from the structure why mutation of S402 impaired activity. Nor is it clear why a single mutation of K366 located on the charged patch at the large opening of SppA (*26*) would drastically affect the activity of SppA.

In addition to this work on the catalytic mechanism, we investigated the membrane topology of SppA. In contrast to what is predicted by several topology programs (*23, 24*), our results using alkaline phosphatase fusion analysis indicate that the *E. coli* enzyme spans the membrane only once and has its carboxyl-terminal domain, as well as the portion of the N-terminal domain containing the essential base, Lys 209, in the periplasmic space. Moreover, the recently determined X-ray structure of SppA, lacking the amino-terminal transmembrane segment, shows that previously predicted transmembrane segment 2 with the active site Ser 409 and transmembrane segment 3 are buried within the globular domain of SppA (*26*). The combined data indicate that SppA spans the membrane with an amino-terminal transmembrane segment and its globular domain located in the periplasmic space (Figure 6B).

REFERENCES

1. Paetzel, M., Karla, A., Strynadka, N. C., and Dalbey, R. E. (2002) Signal peptidases. *Chem. Rev. 102*, 4549–4580.
2. Zwizinski, C., and Wickner, W. (1980) Purification and characterization of leader (signal) peptidase from *Escherichia coli*. *J. Biol. Chem. 255*, 7973–7977.
3. Hussain, M., Ozawa, Y., Ichihara, S., and Mizushima, S. (1982) Signal peptide digestion in *Escherichia coli*. Effect of protease inhibitors on hydrolysis of the cleaved signal peptide of the major outer-membrane lipoprotein. *Eur. J. Biochem. 129*, 233–239.
4. Ichihara, S., Beppu, N., and Mizushima, S. (1984) Protease IV, a cytoplasmic membrane protein of *Escherichia coli*, has signal peptide peptidase activity. *J. Biol. Chem. 259*, 9853–9857.
5. Pacaud, M. (1982) Purification and characterization of two novel proteolytic enzymes in membranes of *Escherichia coli*. Protease IV and protease V. *J. Biol. Chem. 257*, 4333–4339.
6. Ichihara, S., Suzuki, T., Suzuki, M., and Mizushima, S. (1986) Molecular cloning and sequencing of the sppA gene and characterization of the encoded protease IV, a signal peptide peptidase, of *Escherichia coli*. *J. Biol. Chem. 261*, 9405–9411.
7. Rawlings, N. D., and Barrett, A. J. (1999) MEROPS: The peptidase database. *Nucleic Acids Res. 27*, 325–331.
8. Baird, L., Lipinska, B., Raina, S., and Georgopoulos, C. (1991) Identification of the *Escherichia coli* sohB gene, a multicopy suppressor of the HtrA (DegP) null phenotype. *J. Bacteriol. 173*, 5763–5770.
9. Liu, J., and Mushegian, A. (2004) Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages. *J. Bacteriol. 186*, 4369–4375.
10. Yokoyama, H., and Matsui, I. (2005) A novel thermostable membrane protease forming an operon with a stomatin homolog from the hyperthermophilic archaebacterium *Pyrococcus horikoshii*. *J. Biol. Chem. 280*, 6588–6594.
11. Matsumi, R., Atomi, H., and Imanaka, T. (2005) Biochemical properties of a putative signal peptide peptidase from the hyper-

thermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol. 187*, 7072–7080.

12. Matsumi, R., Atomi, H., and Imanaka, T. (2006) Identification of the amino acid residues essential for proteolytic activity in an archaeal signal peptide peptidase. *J. Biol. Chem. 281*, 10533–10539.
13. Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (1998) Crystal structure of a bacterial signal peptidase in complex with a β-lactam inhibitor. *Nature 396*, 186–190. (Erratum: (1998) *Nature 396*, 707).
14. Slilaty, S. N., and Little, J. W. (1987) Lysine-156 and serine-119 are required for LexA repressor cleavage: A possible mechanism. *Proc. Natl. Acad. Sci. U.S.A. 84*, 3987–3991.
15. Peat, T. S., Frank, E. G., McDonald, J. P., Levine, A. S., Woodgate, R., and Hendrickson, W. A. (1996) Structure of the UmuD′ protein and its regulation in response to DNA damage. *Nature 380*, 727–730.
16. Silber, K. R., Keiler, K. C., and Sauer, R. T. (1992) Tsp: A tail-specific protease that selectively degrades proteins with nonpolar C termini. *Proc. Natl. Acad. Sci. U.S.A. 89*, 295–299.
17. Rotanova, T. V., Mel'nikov, E. E., and Tsirul'nikov, K. B. (2003) [Catalytic dyad Ser-Lys at the active site of *Escherichia coli* ATP-dependent Lon-proteinase]. *Bioorg. Khim. 29*, 97–99.
18. Feldman, A. R., Lee, J., Delmas, B., and Paetzel, M. (2006) Crystal structure of a novel viral protease with a serine/lysine catalytic dyad mechanism. *J. Mol. Biol. 358*, 1378–1389.
19. Johnston, S., Lee, J. H., and Ray, D. S. (1985) High-level expression of M13 gene II protein from an inducible polycistronic messenger RNA. *Gene 34*, 137–145.
20. Liu, Y., Patricelli, M. P., and Cravatt, B. F. (1999) Activity-based protein profiling: The serine hydrolases. *Proc. Natl. Acad. Sci. U.S.A. 96*, 14694–14699.
21. San Millan, J. L., Boyd, D., Dalbey, R., Wickner, W., and Beckwith, J. (1989) Use of phoA fusions to study the topology of the *Escherichia coli* inner membrane protein leader peptidase. *J. Bacteriol. 171*, 5536–5541.
22. Michaelis, S., Inouye, H., Oliver, D., and Beckwith, J. (1983) Mutations that alter the signal sequence of alkaline phosphatase in *Escherichia coli*. *J. Bacteriol. 154*, 366–374.
23. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol. 305*, 567–580.
24. von Heijne, G. (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol. 225*, 487–494.
25. Manoil, C., and Beckwith, J. (1986) A genetic approach to analyzing membrane protein topology. *Science 233*, 1403–1408.
26. Kim, A. C., Oliver, D. C., and Paetzel, M. (2008) Crystal structure of a bacterial signal peptide peptidase. *J. Mol. Biol. 376*, 352–366.
27. VanValkenburgh, C., Chen, X., Mullins, C., Fang, H., and Green, N. (1999) The catalytic mechanism of endoplasmic reticulum signal peptidase appears to be distinct from most eubacterial signal peptidases. *J. Biol. Chem. 274*, 11519–11525.
28. Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K., and Martoglio, B. (2002) Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science 296*, 2215–2218.
29. Yokoyama, H., Matsui, E., Akiba, T., Harata, K., and Matsui, I. (2006) Molecular structure of a novel membrane protease specific for a stomatin homolog from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *J. Mol. Biol. 358*, 1152–1164.

BI800657P