Structure and Function of a Novel LD-Carboxypeptidase A Involved in Peptidoglycan Recycling

Debanu Das, a,b Mireille Hervé, c,d Marc-André Elsliger, a,b Rameshwar U. Kadam, a Joanna C. Grant, a,f Hsiu-Ju Chiu, a,b Mark W. Knuth, a,f Heath E. Klock, a,f Mitchell D. Miller, a,b Adam Godzik, a,b,h Scott A. Lesley, a,f Ashley M. Deacon, a,b Dominique Mengin-Lecreulx, c,d, Ian A. Wilson a,b

Join Joint Center for Structural Genomics, a; Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California, USA; Laboratoire des Enveloppes Bactériennes et Antibiotiques, Université Paris-Sud 11, J.B.M.C., UMR 8619, Orsay, France; Centre National de la Recherche Scientifique (CNRS), Orsay, France; Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California, USA; Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California, USA; Center for Research in Biological Systems, University of California, San Diego, La Jolla, California, USA; Program on Bioinformatics and Systems Biology, Sanford-Burnham Institute for Medical Research, La Jolla, California, USA.

Received 29 July 2013. Accepted 25 September 2013
Published ahead of print 11 October 2013

Approximately 50% of cell wall peptidoglycan in Gram-negative bacteria is recycled with each generation. The primary substrates used for peptidoglycan biosynthesis and recycling in the cytoplasm are GlcNAc-MurNAc (anhydrotetrapeptide) and its degradation product, the free tetrapeptide. This complex process involves ~15 proteins, among which the cytoplasmic enzyme LD-carboxypeptidase A (LdcA) catalyzes the bond between the last two L- and D-amino acid residues in the tetrapeptide to form the tripeptide, which is then utilized as a substrate by murein peptide ligase (Mpl). LdcA has been proposed as an antibacterial target. The crystal structure of Novosphingobium aromaticivorans DSM 12444 LdcA (NaLdcA) was determined at 1.89-Å resolution. The enzyme was biochemically characterized and its interactions with the substrate modeled, identifying residues potentially involved in substrate binding. Unaccounted electron density at the dimer interface in the crystal suggested a potential site for disrupting protein–protein interactions should a dimer be required to perform its function in bacteria. Our analysis extends the identification of functional residues to several other homologs, which include enzymes from bacteria that are involved in hydrocarbon degradation and destruction of coral reefs. The NuLdcA crystal structure provides an alternate system for investigating the structure-function relationships of LdcA and increases the structural coverage of the proteins in bacterial cell wall recycling.

Peptidoglycan (murein) is an essential bacterial cell wall polymer consisting of long glycan chains with an alternating arrangement of N-acetylmuramic acid (MurNAc) and N-acetylmuramic acid (MurNAc) that are cross-linked by short peptides. The peptidoglycan barrier surrounds the cytoplasmic membrane, and its main functions are to maintain the shape and turgidity of the cell and to serve as a scaffold for anchoring other cell envelope components, such as proteins (1). In de novo peptidoglycan synthesis, the first steps of assembly involve MurA and MurB, which form UDP-GlcNAc-enolpyruvate and UDP-MurNAc, respectively. The Mur ligases, MurC, MurD, MurE, and MurF, are then involved in sequential addition of L- and D-amino acid residues (L-Ala, D-Glu, meso-diaminopimelate [meso-A2pm], and D-Ala-D-Ala, respectively) to form the pentapeptide derivative UDP-MurNAc-1-Ala-γ-D-Glu-meso-A2pm-D-Ala-D-Ala (2). Subsequent biosynthetic steps catalyzed by several membrane and periplasmic enzymes create the final peptidoglycan (3).

In Gram-negative bacteria, ~30 to 60% of the cell wall peptidoglycan is recycled (4), in a process that involves numerous transglycosidases, amidases, endopeptidases, and carboxypeptidases (5). During the recycling process, LD-carboxypeptidase A (LdcA) specifically cleaves the LD-peptide bond between meso-A2pm and D-Ala in the 1-Ala-γ-D-Glu-meso-A2pm-D-Ala tetrapeptide to produce the 1-Ala-γ-D-Glu-meso-A2pm tripeptide, which is the substrate for the recycling enzyme, murein peptidase ligase (Mpl) (6, 7), whose crystal structure was determined recently (8). Mpl then produces UDP-MurNAc-1-Ala-γ-D-Glu-meso-A2pm, which reenters the peptidoglycan synthesis pathway as the substrate for MurF. Loss of LdcA activity results in cells with heightened sensitivity to lysis in the stationary phase (9). Functional characterization of LdcA includes enzymatic studies (10, 11) and inhibitor experiments (9, 12, 13).

Structural studies of LdcA started with the elucidation of the crystal structure of PaLdcA (Protein Data Bank [PDB] ID 1zl0; ordered locus name P5A198) (14), a protein of unknown function from Pseudomonas aeruginosa, by the Midwest Center for Structural Genomics (http://www.mcsg.anl.gov/) as part of the NIH Protein Structure Initiative (PSI) (15). This work occurred almost simultaneously with the independent structural and biochemical characterization of PaLdcA (PDB IDs 1zrs, 2aum, and 2aua) (16). This study provided the first description of the peptidase activity of LdcA, with a Ser-His-Glu catalytic triad in the active site, which subsequently led to the classification of LdcA enzymes into the Peptidase_S66 (PF02016) family in Pfam (17). PF02016 contains 2,826 sequences (Pfam, version 27.0; March 2013), among which 2,765 are bacterial proteins. This family also includes the microcin C7 self-immunity protein (MccF; plasmid-encoded protein) (18). The crystal structure of an MccF protein

Received 29 July 2013. Accepted 25 September 2013
Published ahead of print 11 October 2013

Address correspondence to Dominique Mengin-Lecreulx, dominique.mengin-lecreulx@u-psud.fr, or Ian A. Wilson, wilson@scripps.edu.
†For this virtual institution, see http://www.jcsg.org.
Copyright © 2013, American Society for Microbiology. All Rights Reserved. doi:10.1128/JB.00900-13
from *Bacillus anthracis* was recently described (PDB ID 3gjz; Center for Structural Genomics of Infectious Diseases [19]) and has the same fold and catalytic triad as LdcA (a genome-encoded protein).

We determined the crystal structure of the LdcA protein from *Novosphingobium aromaticivorans* DSM 12444 (NaLdcA) (UniProt accession no. Q2GBF3; GenBank accession no.YP_496704.1; ordered locus name, Sarto_1426) and identified residues likely involved in substrate recognition. Modeling of the binding of substrate in the active site demonstrates that the catalytic residues are poised for nucleophilic attack on the scissile bond. We also tested the specific activity and substrate specificity of the enzyme and identified a novel site that could potentially be involved in protein–protein interactions. Our results serve as a basis for further structure-based biochemical analyses of LdcA enzymes.

**MATERIALS AND METHODS**

**Protein production and crystallization.** Clones were generated using the polymerase incomplete primer extension (PIPE) cloning method (20). The gene encoding NaLdcA was amplified by PCR from *N. aromaticivorans* DSM 12444 genomic DNA by using *Pfu* Turbo DNA polymerase (Stratagene) and I-PIPE (insert) primers that included sequences for the predicted 5' and 3' ends. The Joint Center for Structural Genomics (JCSG) NaLdcA expression plasmid was deposited in the PSI Biology Materials Repository (DNASU ID NaCD00087837) (21). The expression vector pSpg2, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSKHH HHHHHENLYFQ/G), was PCR amplified with V-PIPE (vector) primers (forward primer, 5'-TAACCCGACTAAATACCTGTTAAGACGTGCT TCCACGTGAGTGATGATGATG-3'; reverse primer, 5'-GCCCTGAGAACTGACGATTTTGGT CGTATGATGATG-3'), V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *Escherichia coli* GeneHogs (Invitrogen) competent cells were transformed with the I-PIPE–V-PIPE mixture and dispensed on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Expression was performed in a selenomethionine-containing medium at 37°C. Sel enomethionine was incorporated via inhibition of methionine biosynthesis (22), which does not require a methionine-auxotrophic strain. At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 μg/ml, and the cells were harvested and frozen. After one freeze-thaw cycle, the cells were homogenized in lysis buffer [50 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM imidazole, 1 mM Tris (2-carboxyethyl)phosphine–HCl (TCEP)] and passed through a microfluidizer. The lysate was clarified by centrifugation at 32,500 × g for 30 min and loaded onto a nickel-chelating affinity column (GE Healthcare) preequilibrated with lysis buffer, the column was washed with wash buffer (50 mM HEPES, pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% [vol/vol] glycerol, 1 mM TCEP), and the protein was eluted with elution buffer (20 mM HEPES, pH 8.0, 300 mM imidazole, 10% [vol/vol] glycerol, 1 mM TCEP). The eluate was buffer exchanged with TEV buffer (20 mM HEPES, pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) by use of a PD-10 column (GE Healthcare) and was incubated with 1 mg of TEV protease per 15 mg of eluted protein for 2 h at room temperature (20°C to 25°C) and then overnight at 4°C. The protease-treated eluate was passed over nickel-chelating resin (GE Healthcare) preequilibrated with HEPES crystallization buffer (20 mM HEPES, pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP), and the column was washed with the same buffer. The flow-through and wash fractions were combined and concentrated to 9.9 mg/ml by centrifugal ultrafiltration (Millipore) for crystallization trials. NaLdcA was crystallized using the nanodrop vapor diffusion method (23) with standard JCSG crystallization protocols (24, 25). Drops composed of 200 nl protein solution mixed with 200 nl crystallization solution in a sitting-drop format were equilibrated against a 50-μl reservoir at 277 K for 28 days prior to harvest. The crystallization reagent consisted of 1.6 M (NH₄)₂SO₄ and 0.1 M Tris-HCl, pH 8.0. Glyceral was added to a final concentration of 20% (vol/vol) as a cryoprotectant. Initial screening for diffraction was carried out using the Stanford automated mounting (SAM) system (26) and an X-ray microsource (27) installed at the Stanford Synchrotron Radiation Lightsource (SSRL; SLAC National Accelerator Laboratory, Menlo Park, CA). The diffraction data were indexed in orthorhombic space group P2₁2₁2₁. The oligomeric state of NaLdcA in solution was determined by analytical size-exclusion chromatography, using a 0.8-cm by 30-cm Shodex protein KW-803 size-exclusion column (Thomson Instruments) (20).

**X-ray data collection, structure determination, and refinement.** Multiwavelength anomalous diffraction (MAD) data were collected at SSRL on beamline 9-2 at wavelengths corresponding to the high-energy remote (λ₁), inflection point (λ₂), and peak (λ₃) of a selenium MAD experiment using the BLU-ICE (28) data collection environment. The data sets were collected at 100 K, using a MarMosaic 325 charge-coupled device (CCD) detector (Rayonix). The MAD data were integrated and reduced using MOSFLM (29) and scaled with the program SCALA from the CCP4 suite (30). The heavy atom substructure was determined with SHELXD (31). Phasing was performed with autoSHARP (32), SOLOMON (33) (implemented in autoSHARP) was used for density modification, and ARP/wARP (34) was used for automatic model building to 1.89-Å resolution. Model completion and crystallographic refinement were performed with the λ₁ data set, using COOT (35) and REFMAC5 (36). The refinement protocol included the experimental phase restraints in the form of Hendrickson–Lattman coefficients from autoSHARP and translation, libration, and screw (TLS) refinement with one TLS group per molecule in the crystallographic asymmetric unit (asu). Data and refinement statistics are summarized in Table 1 (37–40).

**Validation and deposition.** The quality of the crystal structure was analyzed using the JCSG quality control server (http://smb.slac.stanford.edu/jcsg/QC). This server verifies the stereochemical quality of a model by using AutoDepInputTool (41), MolProbity (42), and Phenix (43), the agreement between the atomic model and the data by using RESOLVE (44), the protein sequence by using CLUSTALW (45), the ADP distribution by using Phenix, and differences in $R_{	ext{cryst}}$/Rfree, expected $R_{	ext{free}}$/ $R_{	ext{cryst}}$ and various other items, including atom occupancies, consistency of noncrystallographic symmetry (NCS) pairs, ligand interactions, special positions, etc., by using in-house scripts to analyze the refinement log file and PDB header. Protein quaternary structure analysis was performed using the PISA server (46). Figure 1B was adapted from an analysis using PDBeSum (47), and other figures were prepared with PyMOL (48).

**Sequence analysis.** The analysis of conserved residues (see Fig. 3) was performed using CONSURF (49–51) with default parameters, with homologs collected from UniProt (http://www.uniprot.org) and PSI-BLAST (52), the E value set to 0.001, the number of PSI-BLAST iterations set to 1, the sequence identity range set to 35 to 95%, and the maximum number of homologs set to 100 based on conservation scores, with a Bayesian method of calculation and using the JTT model for substitution of proteins (53, 54). For Fig. 4, the multiple-sequence alignment was generated using CLUSTAL W (45).

**Modeling of ligand-bound NaLdcA.** The model of ligand-bound NaLdcA was obtained by manually positioning the coordinates of MLD (PDB ligand ID MLD, from PDB ID 2ch3) in the substrate-binding cleft by using COOT. The main constraints imposed during this docking were the position of the peptide bond to be cleaved facing the active site residues and the lowest number of clashes with protein residues. The manual docking of MLD was further validated computationally by using Glide (Schrodinger, LLC). This procedure was done in 3 stages. First, the MLD model was subjected to geometry optimization using the Maestro interface and to optimization by Macromodel v9.1 (Schrodinger, LLC), using the optimized potentials for liquid simulations–all-atom (OPLS-AA) force field (55) with the truncated Newton conjugate gradient protocol. Partial atomic charges were computed using the OPLS-AA force field. Second, the protein structure was prepared for calculations. Crystallographic water molecules within 5 Å of the MLD cocrystallized ligand from

---

PDB ID 2cb3 were kept, and the NaLdcA structure was optimized for docking using the protein preparation and refinement utility provided in the Schrödinger suite. Partial atomic charges were assigned according to the OPLS-AA force field. Third, docking calculation was performed using the "extra precision" (XP) mode of Glide (56, 57). The receptor grid was generated with a box size of 20 Å around MLD and the potential for the nonpolar part of the receptor softened, keeping a van der Waals radius scaling factor of 1.0 and a partial charge cutoff of 0.25. For the ligand, the scaling factor was set to 0.8 and the partial charge cutoff to 0.15. The root mean square deviation (RMSD) between the docked conformations and the manual docking of MLD was 4.5 Å (see Fig. 5).

Protein production for biochemical characterization. (i) NaLdcA. DH5α(pTrcHis30::EcldcA) cells were grown in 2YT medium supplemented with ampicillin (2-liter cultures). When the optical density at 600 nm reached 0.2, isopropyl-β-D-thiogalactopyranoside (IPTG) was added (1 mM), and incubation was continued overnight at 20°C. Cells were harvested at 4°C, and the pellet was washed with cold 20 mM phosphate buffer, pH 7.4, containing 0.1% 2-mercaptoethanol (buffer A). Cells were suspended in buffer A (12 ml) and disrupted by sonication in the cold, using a Bioblock Vibracell 72412 sonicator. The resulting suspension was centrifuged at 4°C for 30 min at 200,000 × g with a Beckman TL100 apparatus, and the pellet was discarded.

(ii) EcLdcA. DH5α(pTrcHis30::EcldcA) cells were grown in 2YT medium supplemented with ampicillin (2-litter cultures). When the optical density at 600 nm reached 0.2, isopropyl-β-D-thiogalactopyranoside (IPTG) was added (1 mM), and incubation was continued overnight at 20°C. Cells were harvested at 4°C, and the pellet was washed with buffer A. Cells were suspended in buffer A (30 ml) and disrupted by sonication, and the resulting suspension was centrifuged at 4°C for 30 min at 200,000 × g. The His6-tagged LdcA proteins were purified from the soluble fractions on Ni2+-nitrilotriacetic acid (Ni2+-NTA)-agarose following the manufacturer’s recommendations (Qiagen). Washing and elution steps were performed with a discontinuous gradient of imidazole (200 to 400 mM) in buffer A containing 300 mM KCl. Protein contents were analyzed by
sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and relevant fractions were pooled and dialyzed against 100 volumes of buffer A. Glycerol (10% final concentration) was added for storage of the protein at \(-20\)°C. Protein concentrations were determined by quantitative amino acid analysis with a Hitachi L8800 analyzer (ScienceTec) after hydrolysis of samples in 6 M HCl at 105°C for 24 h.

**Assays for biochemical characterization.** The standard \(\alpha\)-carboxy-peptidase activity assay consisted of following the hydrolysis of the tetrapeptide \(\text{L-Ala-γ-D-Glu-meso-A2pm-D-Ala}\) in a reaction mixture (25 \(\mu\)l) containing 100 mM Tris-HCl buffer, pH 8, 0.2 mM tetrapeptide, and LdcA enzyme (0.2 ng of protein in 5 \(\mu\)l of buffer A). After 30 min of incubation at 37°C, the reaction was stopped by adding 5 \(\mu\)l of glacial acetic acid. The substrate and reaction product (the tripeptide \(\text{L-Ala-γ-D-Glu-meso}\)) were separated by high-pressure liquid chromatography (HPLC) on a Hypersil 3 \(\mu\) column (250 \(\times\) 4.6 mm; Alltech France), using 0.05% trifluoroacetic acid (TFA) and 2% acetonitrile as the elution buffer, at a flow rate of 0.6 ml/min. Peaks were detected by measuring the absorbance at 214 nm (under these conditions, the tetrapeptide and tripeptide were eluted in 18 and 11 min, respectively). For determination of the kinetic constants, the same assay was performed with various concentrations of tetrapeptide. In all cases, the substrate consumption was <20%, and the linearity was ensured within this interval even at the lowest substrate concentration. The data were fitted to the equation \(v = \frac{V_{\text{max}}S}{(K_m + S)}\) using the MDFit software developed by M. Desmadril (IBBMC, Orsay, France). When other compounds were tested as substrates, the assay conditions were similar, except that the amount of LdcA enzyme used was appropriately adjusted (from 0.2 ng to 5 \(\mu\)g). Also, HPLC conditions were modified so that the separation of product and substrate was optimal: 50 mM ammonium formate, pH 3.9, was used for the separation of UDP-MurNAc-tetrapeptide and UDP-MurNAc-tripeptide (retention times of 24 and 12 min, respectively); the same buffer at pH 4.4 was used to separate UDP-MurNAc-pentapeptide (21 min) and UDP-MurNAc-tripeptide (8 min), as well as the A2pm-containing MurNAc-tetrapeptide (retention times for the two anomers [\(\alpha/\beta\)], 14/26 min) and MurNAc-tripeptide (\(\alpha/\beta\) retention times, 7/11 min). Ammonium acetate, pH 6.0 (50 mM), was used instead for the separation of Lys-containing MurNAc-
tetrapeptide (α/β retention times, 14/28 min) and MurNAc-tripeptide (α/β retention times, 10/16 min). For the separation of GlcNAc-MurNAc-c(anhdro)-tetrapeptide from its tripeptide derivative, 0.05% TFA was used as the elution buffer, together with a gradient of acetonitrile (from 0 to 20%) applied between 0 and 40 min; the latter compounds were eluted in 34 and 29 min, respectively.

Protein structure accession number. Atomic coordinates and experimental structure factors for NaLdCA to 1.89-Å resolution have been deposited in the Protein Data Bank (www.wwpdb.org) under PDB ID 3g23.

RESULTS

Crystal structure of NaLdCA. The cloning, expression, purification, and crystallization of NaLdCA were carried out using standard protocols at the Joint Center for Structural Genomics (http://www.jcsg.org), as detailed in Materials and Methods. The crystal structure of NaLdCA was determined by multiwavelength anomalous diffraction (MAD) phasing to a resolution of 1.89 Å. Data collection, model, and refinement statistics are summarized in Table 1 (37–40). There are two NaLdCA molecules in the asu. Each molecule (Fig. 1) contains residues 1 to 271 of the 273-residue full-length protein, in addition to Gly0, which remains after cleavage of the expression and purification tag. Residues 159 and 160 and residues 272 and 273 in both molecules were omitted from the model due to the lack of interpretable electron density. Seventeen sulfate ions from the crystallization condition, 8 glycerol molecules from the cryoprotectant, 671 water molecules, and 1 unidentified ligand (UNL) were also identified in the asu. The Matthews coefficient (58) is ~2.4 Å³/Da, with an estimated solvent content of ~50%. A Ramachandran plot produced by MolProbity (42) shows that 98.9% of the residues are in the favored regions and none are in disallowed regions. The catalytic Ser104 residue in both copies in the asu is in an energetically unfavorable conformation and appears in the additionally allowed regions (lower right quadrant) of the Ramachandran plot, which is typical for serine proteases.

The NaLdCA monomer is composed of α-helices H1 to H11 and β-strands β1 to β10 (Fig. 1). Both molecules in the asu are very similar in structure: chain A can be superimposed onto chain B, with an RMSD of 0.2 Å over 270 Ca atoms. According to SCOP (59), the LdCA architecture has an N-terminal domain with a flavodoxin-like fold (residues ~3 to 169 in PaLdCA) and a C-terminal domain with a “swiveling” β/β/α fold (residues ~170 to 307 in PaLdCA). The active site comprises a Ser104-His261-Glu191 catalytic triad located in a cleft between the two domains.

Structural comparison with PaLdCA and MccF. The overall structure of NaLdCA is similar to those of PaLdCA and MccF. NaLdCA can be superimposed on PaLdCA (PDB ID 1zrs; 20% sequence identity), with an RMSD of 2.4 Å over 254 Ca atoms, and on MccF (PDB ID 3gjz; 16% sequence identity), with an RMSD of 2.5 Å over 242 Ca atoms (Fig. 2). However, there are several notable differences. The counterpart of one of the regions that is disordered in the NaLdCA structure (residues 159 and 160) contains an insertion in both PaLdCA (residues 173 to 185) and MccF (residues 166 to 204) that forms a loop which, in the case of MccF, extends out toward the active site and forms a lid (residues 168 to 195). This lid partially covers and restricts access to the substrate-binding cavity and could have important functional consequences. In addition, both PaLdCA and MccF have longer C-terminal extensions than that of NaLdCA.

Sequence and structure analysis. The majority (2,084 proteins) of the members of Pfam PF02016 contain a single Peptidase_S66 domain of ~300 residues, which encompasses both the N-terminal and C-terminal domains of LdCA. The crystal structure of NaLdCA is only the third structure determined from this family (note that while the manuscript was in preparation, structures of three other MccF proteins from different bacterial species were determined, with PDB codes 3tla, 4e5s, and 4e94). In order to identify conserved regions of the protein, sequences of the following 9 proteins, annotated as either putative LdCA proteins or proteins of unknown function, with sequence identities ranging from 35 to 95% compared to NaLdCA as identified in a PSI-BLAST search, were selected for comparison (UniProt [60] IDs are given): B4W8U2 from Brevundimonas sp. BAL3, D9QNU0 from Brevundimonas subvibrioides, A3WBK2 from Erythrobacter sp. NA1, Q2N9T0 from Erythrobacter litoris, A5P850 from Erythrobacter sp. SD-21, Q1GRW2 from Sphingopyxis alaskensis RB2256, Q2WBO4 from Magnetospirillum magneticum, Q0C337 from Hyphomonas...
neptunium, and B8FC74 from Desulfatibacillum alkenivorans AK-01 (see Materials and Methods for more details). The residue conservation was mapped onto the NaLdcA structure to identify potential areas of functional importance. One side of NaLdcA displays significantly more conserved residues than the other, and as expected, these highly conserved regions include putative substrate-binding site residues (Asp52, Arg55, Tyr78, Arg82, Ser104, Asp105, Asn167, Leu174, Thr177, Glu191, Val193, Tyr198, Asp201, Arg202, His261, and Arg248), the catalytic triad (Ser104-His261-Glu191), and residues involved in dimer interactions (Glu195 and Arg202) (Fig. 3). The Ser(nucleophile)-His (general base)-Glu catalytic triad is characteristic of other LdcA proteins (16), thereby placing LdcA into the category of nonclassical serine proteases with a catalytic triad that is also seen in lipases and esterases (61). LdcA binds a large substrate that has a specific peptide portion and a less-specific sugar moiety. The residues that bind the peptide moiety of the substrate are the most conserved, and residues in the sugar-binding pocket are more variable. This distribution is consistent with data reported previously (62) and here (see "Substrate specificity") showing that the enzyme prefers tetrapeptide substrates and can discriminate

FIG 3 Residue conservation analysis. The surface representation of NaLdcA is color coded (magenta, most conserved; cyan, most variable) according to amino acid conservation based on comparison to 9 other proteins annotated as either putative LdcA enzymes or proteins of unknown function, with sequence identities of 35 to 95% compared to NaLdcA. (A) The left panel (same view as in Fig. 1) shows that one side of NaLdcA has more highly conserved residues in and around the putative active site than the other side (right panel, viewed at 180° around a vertical axis from the left panel). (B) Stick representation of the most conserved residues, which include Asp52, Arg55, Tyr78, Arg82, Ser104, Asp105, Asn167, Leu174, Thr177, Glu191, Val193, Tyr198, Asp201, Arg202, His261, and Arg248. The Ser104-Glu191-His261 catalytic triad is shown in blue.
between tetrapeptide and pentapeptide substrates (Table 2, rows 2 and 4) but is less selective toward the N-terminal sugar moiety, i.e., GlcNAc-MurNAc, UDP-MurNAc, or MurNAc (Table 2, compare row 2 to rows 3, 5, and 6).

Substrate recognition residues and site mechanism. To elucidate protein-ligand interactions, GlcNAc-MurNAc-\(\gamma\)-D-Glu-meso-A2pm-D-Ala (PDB ligand ID MLD, from PDB ID 2cb3), which represents one of the substrates of LdcA, was manually docked into the substrate-binding site such that the C-terminal of Ala of the tetrapeptide is positioned for attack by the manu-ally docked into the substrate-binding site such that the C-

| TABLE 2 Specific activities and substrate specificities of N. aromaticivorans and E. coli 1D-carboxypeptidases |
|-----------------------------------------|-------------------|-------------------|
| Substrate                              | \(\text{Sp act} \ (\text{nmol/min/mg})\) |
| L-Ala-\(\gamma\)-D-Glu-meso-A2pm-D-Ala | 230,000 250,000   |
| UDP-MurNAc-L-Ala-\(\gamma\)-D-Glu-meso-A2pm-D-Ala | 24,500 65,500    |
| UDP-MurNAc-L-Ala-\(\gamma\)-D-Glu-meso-A2pm-D-Ala | 7 51             |
| GlcNAc-MurNAc(anhydro)-L-Ala-\(\gamma\)-D-Glu-meso-A2pm-D-Ala | 14,300 26,500  |
| MurNAc-L-Ala-\(\gamma\)-D-Glu-meso-A2pm-D-Ala | 30,000 35,500    |
| MurNAc-L-Ala-\(\gamma\)-D-Glu-L-Lys-D-Ala | 1,000 450        |

\(a\) The standard enzyme assay conditions described in Materials and Methods were used. The amount of enzyme varied from 0.2 ng to 0.5 \(\mu\)g, depending on the substrate used. 
To ensure linearity, the consumption of substrate was monitored for 5-10 min, and the specific activities were calculated.

Enzymatic activity. The expression and purification of NaLdcA and EcLdcA for biochemical analysis were carried out as detailed in Materials and Methods. Both proteins appeared to be >95% pure, as judged by SDS-PAGE analysis, with yields of 12 and 4 mg of purified protein/liter of culture, respectively. The specific activities of purified NaLdcA and EcLdcA, determined under standard assay conditions and using the tetrapeptide L-Ala-\(\gamma\)-D-Glu-meso-A2pm-D-Ala as the reference substrate, had similar values of 230 and 250 \(\mu\)mol/min/mg, respectively. Analogs of the studies performed on PaLdcA to test whether the enzyme pos-
sessed metallopeptidase activity, the effect of cations was also tested on NaLdcA. Neither enzyme required a cation (Mg$^{2+}$/Hg$^{1001}$ and Zn$^{2+}$/Hg$^{11001}$ were tested). EDTA at 2.5 mM did not inhibit the reaction. However, addition of 2 mM ZnSO$_4$ dramatically decreased the activity of NaLdcA (3.8 versus 2300 mol/min/mg). The optimal pH for activity of NaLdcA was 7.4 to 7.6. The $K_m$ for the tetrapeptide of NaLdcA was estimated to be 0.17 ± 0.01 mM, and the $k_{cat}$ was 160 ± 25 s$^{-1}$.

**Substrate specificity.** The activity of the NaLdcA and EcLdcA enzymes was tested on a series of potential substrates, including peptidoglycan precursors and intermediates generated during maturation of the cell wall polymer (Table 2). The free tetrapeptide was clearly the best substrate. The presence of either a MurNAc, UDP-MurNAc, or GlcNAc-MurNAc(anhydro) moiety at the N terminus of the peptide significantly reduced the activity (8- to 16-fold). The replacement of diaminopimelate (A$_2$pm) by Lys at position 3 of the peptide reduced the enzyme activity 30-fold (tested with MurNAc-peptide substrates), suggesting that the ε-carboxyl group at the D center of A$_2$pm is important for substrate binding within the catalytic site as a result of stabilizing H-bonding interactions between the ε-carboxyl group and Glu195. UDP-MurNAc-pentapeptide also appeared to be a

**FIG 4** Model of protein-ligand complex and conservation of putative functional residues. (A) The binding of NaLdcA to the ligand GlcNAc-MurNAc-$\gamma$-$\delta$-Glu-meso-$\delta$-Ala (PDB ligand ID MLD, from PDB ID 2c63) was modeled such that the carboxy-terminal $\delta$-Ala is positioned for catalysis near the active site Ser-His-Glu triad. The molecular orientation is similar to that in Fig. 3, and the helices are labeled. (B) Magnified view of the binding site. Oxygen atoms from the carboxylate groups of A$_2$pm and $\delta$-Glu of the ligand are superimposed on the two sulfate ions (SO$_4^{-274}$ and SO$_4^{-276}$, respectively) present in the substrate-binding pocket and present similar interactions to the active site residues Arg75 and Glu195, thereby suggesting that the sulfate ions mimic portions of the substrate. (C) Sequence logo representation of relative frequencies of occurrence of residues in NaLdcA and other homologs, obtained using Weblogo (68; [http://weblogo.berkeley.edu/](http://weblogo.berkeley.edu/)). The same set of proteins and multiple-sequence alignment listed in Fig. 3 were used in this analysis. Panels A, B, and C represent different sequence ranges in the enzyme.
substrate, although at a much lower rate than that with UDP-MurNAc-tetrapeptide (0.007 versus 24.5 μmol/min/mg), indicating that the LdcA enzyme also exhibits a very low endopeptidase activity. Comparison of the two LdcA (NaLdcA/EcLdcA) orthologs shows that their properties are quite similar in terms of both specific activity and substrate specificity (Table 2).

DISCUSSION
Substrates originating from peptidoglycan biosynthesis or recycling that are normally present in the cytoplasm are primarily GlcNAc-MurNAc(anhydro)-tetrapeptide and its degradation product, the free tetrapeptide. If the tetrapeptide is not rapidly cleaved by LdcA, it is added to UDP-MurNAc by the Mpl enzyme, and the UDP-MurNAc-tetrapeptide generated accumulates (as an “end product,” which is not a normal intermediate in this pathway) until it is processed by LdcA (7, 9). In this study, we confirmed that these compounds are substrates for NaLdcA, with a preference for the free tetrapeptide. An endopeptidase secondary activity of NaLdcA was also observed with UDP-MurNAc-pentapeptide, which was very low and therefore should not have any deleterious effect (or regulatory function) on peptidoglycan metabolism (note that EcLdcA activity on free tetrapeptide, UDP-MurNAc-pentapeptide, and MurNAc-tetrapeptide, as well as (to a lesser extent, ca. 25%) on disaccharide tetrapeptide and its anhydro derivative.

Our results extend previous biochemical and structural work on LdcA proteins by providing insights into how the substrate is likely to bind and the residues that may be involved in substrate interactions. Functional characterization reveals how the activity of NaLdcA compares with that of EcLdcA. Prior work on LdcA has indicated that it may be a promising antibacterial target, with LdcA inhibitors being useful in the late stages of replication in bacteria or in combination therapy with drugs acting in earlier stages (12). Although crystal structures of LdcA show a consistent dimer, it is not known whether LdcA is functionally active as a dimer and, if so, whether dimers are more or less active than monomers. Our structure reveals a putative binding site for molecules that may stabilize or disrupt protein-protein interactions in dimer formation, and future studies will be needed to address this question. Modulation of LdcA activity by ligands could find applications in medicine for destabilizing pathogenic Gram-negative bacteria, in general, and NaLdcA in particular. For example, N. aromaticivorans has been linked to primary biliary cirrhosis (64, 65), and Novosphingobium or Sphingomonas species have been associated with environmental damage resulting from the
death of coral reefs (66). Applications in industrial biotechnology may also be feasible, such as hydrocarbon bioremediation, where B8FC74 from Desulfatibacillum alkenivorans AK-01 has been proposed as a model organism for anaerobic alkane biodegradation (67).

ACKNOWLEDGMENTS

This work was conducted at the JCSG and the CNRS (UMR 8619). The JCSG is supported by the NIH, National Institutes of General Medical Sciences, Protein Structure Initiative (grant U54 GM094586). R.U.K. is grateful to the Swiss National Science Foundation for a postdoctoral fellowship.

Genomic DNA from Novosphingobium aromaticivorans DSM 12444 (ATCC 700278D) was obtained from the American Type Culture Collection (ATCC). We thank all members of the JCSG for their contribution to the development and operation of our HTP structural biology pipeline and for bioinformatic analyses, protein production, and structure determination.

Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a directorate of the SLAC National Accelerator Laboratory and an Office of the SLAC National Accelerator Laboratory. The Advanced Photon Source is a U.S. Department of Energy Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract DE-AC02-06CH11357. The Structural Molecular Biology Program at the Stanford Synchrotron Radiation Lightsource is supported by the U.S. Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393).

The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

REFERENCES

8. Das D, Hervé M, Feuerhelm J, Farr CL, Chiu HJ, Elsiger MA, Knuth MW, Klock HE, Miller MD, Godzik A, Lesley SA, Deacon AM, Mengin-


