Functional Identification of Incorrectly Annotated Prolidases from the Amidohydrolase Superfamily of Enzymes[†,‡]

Dao Feng Xiang,[§] Yury Patskovsky,[‖] Chengfu Xu,[§] Amanda J. Meyer,[‖] J. Michael Sauder,[⊥] Stephen K. Burley,[⊥] Steven C. Almo,*,[‖] and Frank M. Raushel*,[§]

Department of Chemistry, Texas A&M University, P.O. Box 30012, College Station, Texas 77842-3012, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, and SGX Pharmaceuticals, Inc., 10505 Roselle Street, San Diego, California 92121

ABSTRACT: The substrate profiles for two proteins from *Caulobacter crescentus* CB15 (Cc2672 and Cc3125) and one protein (Sgx9359b) derived from a DNA sequence (gi|44368820) isolated from the Sargasso Sea were determined using combinatorial libraries of dipeptides and *N*-acyl derivatives of amino acids. These proteins are members of the amidohydrolase superfamily and are currently misannotated in NCBI as catalyzing the hydrolysis of L-Xaa-L-Pro dipeptides. Cc2672 was shown to catalyze the hydrolysis of L-Xaa-L-Arg/Lys dipeptides and the *N*-acetyl and *N*-formyl derivatives of lysine and arginine. This enzyme will also hydrolyze longer peptides that terminate in either lysine or arginine. The *N*-methyl phosphonate derivative of L-lysine was a potent competitive inhibitor of Cc2672 with a K_i value of 120 nM. Cc3125 was shown to catalyze the hydrolysis of L-Xaa-L-Arg/Lys dipeptides but will not hydrolyze tripeptides or the *N*-formyl and *N*-acetyl derivatives of lysine or arginine. The substrate profile for Sgx9359b is similar to that of Cc2672 except that compounds with a C-terminal lysine are not recognized as substrates. The X-ray structure of Sgx9359b was determined to a resolution of 2.3 Å. The protein folds as a $(\beta/\alpha)_8$-barrel and self-associates to form a homooctamer. The active site is composed of a binuclear metal center similar to that found in phosphotriesterase and dihydroorotase. In one crystal form, arginine was bound adventitiously to the eight active sites within the octamer. The orientation of the arginine in the active site identified the structural determinants for recognition of the α-carboxylate and the positively charged side chains of arginine-containing substrates. This information was used to identify 18 other bacterial sequences that possess identical or similar substrate profiles.

In 2008 there were nearly 800 completely sequenced bacterial genomes reported in the NCBI Web site. A critical assessment of the annotations for the more than 4 million genes contained within these organisms indicates that a significant fraction of the derived enzymes and proteins have an uncertain, incorrect, or ambiguous catalytic function. This observation suggests that substantial parts of the metabolic landscape remain to be identified and that the rules for deciphering catalytic activity from protein sequences and three-dimensional structures are quite challenging. Our approach to a limited portion of this difficult problem has been to combine three-dimensional structure determination, computational docking, high-throughput screening, and genomic context toward the assignment of function to members of the amidohydrolase superfamily (*1−3*).

The amidohydrolase superfamily (AHS)[1] is a complex cluster of enzymes that has been shown to catalyze a variety of chemical transformations (*4, 5*). Well-characterized members of this superfamily include phosphotriesterase (*6*), urease (*7*), dihydroorotase (*8*), and adenosine deaminase (*9*). The predominant reactions catalyzed by the AHS include isomerizations (*10*), decarboxylations (*11*), hydrations (*12*), and hydrolysis of C−O (*13*), C−N (*14*), and P−O (*15*) bonds. All members of the AHS adopt a $(\beta/\alpha)_8$-barrel structural fold and contain a mononuclear or binuclear metal center embedded at the C-terminal end of the β-barrel. *In vivo*, the metal centers can be occupied by Ni^{2+}, Zn^{2+}, Fe^{2+}, or Mn^{2+} (*4*). However, certain members of the AHS can be activated by Co^{2+} and Cd^{2+} via reconstitution of the apoenzyme (*16*). The metal centers found within the active sites of the amidohydrolase superfamily function in catalysis by activating solvent water for nucleophilic attack and/or stabilization of transition state structures (*4*).

Enzymes within the AHS have been annotated by NCBI as prolidases that presumably catalyze the hydrolysis of dipeptides that contain proline as the C-terminal amino acid. Approximately 2−3% of the known members of the ami-

---

* To whom correspondence may be addressed. F.M.R.: telephone, (979) 845-3373; fax, (979) 845-9452; e-mail, raushel@tamu.edu. S.C.A.: telephone, (718) 430-2746; fax, (718) 430-8565; e-mail, almo@aecom.yu.edu.

§ Texas A&M University.

‖ Albert Einstein College of Medicine.

⊥ SGX Pharmaceuticals, Inc.

[1] Abbreviations: AHS, amidohydrolase superfamily; ATCC, American Type Culture Collection; IPTG, isopropyl thiogalactopyranoside; rmsd, root mean squared deviation; IAD, isoaspartyl dipeptidase.

dohydrolase superfamily have been annotated as prolidases. However, none of these enzymes has been adequately characterized, and the breadth of the substrate specificity for variation in the amino acids that can occupy the amino and carboxy termini of potential substrates is unknown. Also unknown are the structural determinants and highly conserved amino acid residues within the active sites of these enzymes that dictate the substrate profile.
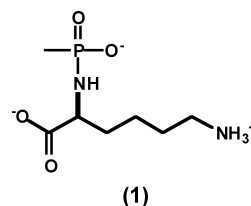
In this paper we present the discovery of the catalytic activities for two enzymes of ambiguous function from *Caulobacter crescentus* CB15 and a related protein derived from an environmental DNA sequence isolated from the Sargasso Sea. *C. crescentus* naturally inhabits an aquatic environment and plays an important role in biochemical cycling of organic nutrients (*17*). Based upon amino acid sequence alignments with structurally characterized members of the AHS, it is expected that all three enzymes will contain a binuclear metal center at the active site. The two enzymes from *C. crescentus*, Cc2672 and Cc3125, are 47% identical in protein sequence to one another and are 47% and 37% identical, respectively, to the enzyme from the Sargasso Sea, Sgx9359b (gi| 44368820). The substrate specificities were determined for these enzymes via the synthesis and characterization of multiple dipeptide libraries containing nearly all possible combinations of L-Xaa-L-Xaa dipeptides using the 20 common amino acids. Contrary to our initial expectations none of the three enzymes was able to hydrolyze dipeptides that contained L-proline at the C-terminus. Cc2672 can hydrolyze dipeptides with either L-lysine or L-arginine at the C-terminus. It is also able to hydrolyze the *N*-acetyl or *N*-formyl derivatives of L-lysine or L-arginine and short polypeptides ending in these amino acids. The substrate specificity of Cc3125 is restricted to the hydrolysis of dipeptides with either L-lysine or L-arginine at the C-terminus. The substrate specificity of Sgx9359b is similar to that of Cc2672, except that the C-terminus must be L-arginine. The X-ray structure of Sgx9359b has been determined in the presence of the hydrolysis product, L-arginine. Sgx9359b and Cc2672 are more properly described as carboxypeptidases whereas Cc3125 appears to be a true dipeptide hydrolase.

MATERIALS AND METHODS

Materials. All chemicals were obtained from Sigma or Aldrich, unless otherwise stated. The genomic DNA from *C. crescentus* CB15 was purchased from the American Type Culture Collection (ATCC). The synthesis of oligonucleotides and DNA sequencing reactions were performed by the Gene Technology Laboratory of Texas A&M University. The pET-30a(+) expression vector was acquired from Novagen. T4 DNA ligase and various restriction enzymes were purchased from New England Biolabs. Platinum *Pfx* DNA polymerase and the Wizard Plus SV Mini-Prep DNA purification kit were obtained from Invitrogen and Promega, respectively. The *N*-methyl phosphonate derivative of L-lysine (**1**) was synthesized according to the method described by Xu et al., and the structure is presented in Scheme 1 (*18*).

Synthesis of Dipeptide Libraries and Dipeptides. The dipeptide libraries were made as described below. Nineteen preloaded *N*-Fmoc-protected (or unprotected) L-amino acid Wang resins (0.02 mmol each of *N*-Fmoc-L-Ala, *N*-Fmoc-L-

Scheme 1

(1)

Arg(Mtr), *N*-Fmoc-L-Asn(Trt), *N*-Fmoc-L-Asp(OtBu), *N*-Fmoc-L-Glu(OtBu), *N*-Fmoc-L-Gln(Trt), *N*-Fmoc-L-Gly, *N*-Fmoc-L-His(Trt), *N*-Fmoc-L-Ile, *N*-Fmoc-L-Leu, *N*-Fmoc-L-Lys(Boc), *N*-Fmoc-L-Met, *N*-Fmoc-L-Phe, *N*-Fmoc-L-Pro, *N*-Fmoc-L-Ser-(Trt), *N*-Fmoc-L-Thr(Trt), *N*-Fmoc-L-Trp(Boc), *N*-Fmoc-L-Tyr(tBu), and *N*-Fmoc-L-Val) and DMF (5 mL) were shaken in a syringe for 30 min. The DMF was removed by filtration, and then 6 mL of 20% piperidine in DMF was added and the mixture shaken for 30 min. This process was repeated. The beads were washed with DMF (4 × 5 mL), and then *N*-Fmoc-L-Ala-OH (177.4 mg, 0.57 mmol), HOBt·H_2O (87.2 mg, 0.57 mmol), and *N,N'*-diisopropylcarbodiimide (71.8 mg, 0.57 mmol) in DMF (6 mL) were added and shaken overnight. The reagents were removed, and the beads were washed with DMF (4 × 5 mL), dichloromethane (4 × 5 mL), and methanol (4 × 5 mL) and dried for several hours. To the dried beads was added cocktail R (4 mL, TFA/thioanisole/EDT/anisole (90/5/3/2 v/v)), and the mixture was shaken for 3 h. The *N*-Fmoc-L-Ala-L-Xaa amino acid library was obtain after removal of the solvent under reduced pressure, washing with EtOAc/Et_2O (1/5 v/v), and then drying overnight at 50 °C. The library was stirred with 20% piperidine in DMF (5 mL) for 30 min to obtain the L-Ala-L-Xaa dipeptide library after removal of the solvent and washing with EtOAc/Et_2O (1/5 v/v). In the same way, the libraries of L-Arg-L-Xaa, L-Asn-L-Xaa, L-Asp-L-Xaa, L-Glu-L-Xaa, L-Gln-L-Xaa, Gly-L-Xaa, L-His-L-Xaa, L-Ile-L-Xaa, L-Leu-L-Xaa, L-Lys-L-Xaa, L-Met-L-Xaa, L-Phe-L-Xaa, L-Pro-L-Xaa, L-Ser-L-Xaa, L-Thr-L-Xaa, L-Trp-L-Xaa, L-Tyr-L-Xaa, and L-Val-L-Xaa were prepared. Mass spectrometry analysis (ESI, positive and negative mode) was used to verify the presence of the 19 members in each of the 19 dipeptide libraries constructed for this investigation.

The dipeptides L-Ala-L-Lys and L-Ala-L-Arg were prepared in a similar fashion using single preloaded *N*-Fmoc-L-Lys(Boc) and *N*-Fmoc-L-Arg(Mtr) beads. The tripeptide L-Gly-L-Phe-L-Arg was purchased from Aroz Technologies LLC.

Gene Cloning. The genes for Cc2672 and Cc3125 were amplified from the genomic DNA of *C. crescentus* CB15. The PCR product for the isolation of the gene for Cc2672 utilized the primers 5′-AGAACTTCCATATGCGTATGGG-GATGAAGATCGCGACGC-3′ and 5′-AGCGAATTCCTA-CGGGGCCTTCACCACCGCGC-3′. The PCR product for the amplification of the gene for Cc3125 utilized the primer pair 5′-AGAACTTCCATATGAAACTGCACGTGTTTTG-CGTCGCCG-3′ and 5′-ACGGAATTCCTAGTCGTCCT-TGACCACGGTCCCG-3′. The PCR products were gel purified, digested with *Nde*I and *Eco*RI, ligated to the expression vector pET-30a(+), and then transformed into XL1-blue cells. Individual colonies containing the plasmid were selected on LB plates containing 50 *μ*g/mL kanamycin and then used to inoculate 5 mL of LB. The entire coding

region of the plasmids containing the genes for Cc2672 and Cc3125 was sequenced to confirm the fidelity of the PCR amplification. Based upon the reported sequence of gi|44368820 deduced from a DNA sample originally isolated from the Sargasso Sea, the New York SGX Research Center for Structural Genomics (NYSGXRC) cloned and expressed Sgx9359b as a His-tagged protein following codon optimization and gene synthesis by Codon Devices, Inc. (Cambridge, MA). NYSGXRC has made the clone (9359b1BCt7p1) available through the Protein Structure Initiative Material Repository (PSI-MR) at the Harvard Institute of Proteomics (http://www.hip.harvard.edu/PSIMR/).

Protein Purification. Cc2672 and Cc3125 were purified by a combination of gel filtration and ion-exchange chromatography. BL21(DE3) Star cells (Novagen) were transformed with a pET-30a(+) plasmid containing the gene for either Cc2672 or Cc3125. A single colony was used to inoculate a 5 mL overnight culture of LB broth containing 50 μg/mL kanamycin. The overnight culture was then used to inoculate 1.0 L of LB medium containing 50 μg/mL kanamycin. The culture was grown at 30 °C and induced by the addition of 0.5 mM isopropyl D-thiogalactopyranoside (IPTG) after an A_{600} of 0.6 had been reached. At this time 1.0 mM Zn(OAc)$_2$ was added to the culture. Cells were harvested by centrifugation (5000 rpm for 15 min) 18 h after IPTG induction and stored at −80 °C. For the purification of Cc2672, the frozen cells were resuspended in 50 mM Hepes, pH 8.0, and lysed by sonication (5 s pulses for 30 min) at 0 °C. After centrifugation, the nucleic acids were removed by adding 2% (w/v) protamine sulfate dropwise. The supernatant solution after centrifugation was fractionated using ammonium sulfate. The precipitated protein (60−75% ammonium sulfate saturation) was resuspended in a minimum volume of 50 mM Hepes, pH 8.0, and loaded onto a Superdex 200 gel filtration column (Amersham Pharmacia) and eluted at a flow rate of 2.0 mL/min. The fractions containing the desired protein were pooled based on the results from SDS−PAGE. The protein was loaded onto a Resource Q anion-exchange column (Amersham Pharmacia) and eluted with a linear gradient of NaCl in 20 mM Hepes, pH 8.0. The purity of the protein was verified by SDS−PAGE. The protein Cc3125 was purified in a manner similar to that of Cc2672 except that the protein was initially fractionated with ammonium sulfate between 0 and 65% of saturation. Cc3125 was eluted from the Resource Q column at approximately 70 mM NaCl.

For purification of Sgx9359b, *Escherichia coli* BL21(DE3) Star cells were transformed with the plasmid encoding the gene for this protein. One liter of LB was inoculated with a 5 mL overnight culture. The inoculated culture was grown with agitation at 30 °C to an OD$_{600}$ of 0.6, supplemented with 1.0 mM Zn(OAc)$_2$, induced with 0.5 mM IPTG, and then allowed to grow at 30 °C for 14 h. The pellet was harvested by centrifugation and suspended in binding buffer (20 mM Hepes, pH 7.9, 0.5 M NaCl, and 5 mM imidazole). The cells were disrupted by sonication, and the insoluble debris was removed by centrifugation (15 min at 10000 rpm). The clarified cell extract was applied to a 24 mL column of chelating Sepharose Fast Flow (Amersham Biosciences) charged with Ni^{2+} and precalibrated with binding buffer. The column was washed with 1000 mL of binding buffer until the absorbance of the flow-through at 280 nm was constant.

The His-tagged Sgx9359b protein was eluted with a gradient of elution buffer (10 mM Hepes, pH 7.9, 0.25 M NaCl, and 0.5 M imidazole). The protein thus obtained was further purified by a Resource Q anion-exchange column after NaCl was removed by dialysis. Sgx9359b was eluted from the column with a gradient of NaCl in 20 mM Hepes, pH 8.0. The protein was >95% pure based upon SDS−PAGE.

Amino Acid Sequence and Metal Content. The purified proteins Cc2672 and Cc3125 were subjected to N-terminal amino acid sequence analysis by the Protein Chemistry Laboratory at Texas A&M University. The N-terminal amino acid sequence analysis for the first 11 amino acids of Cc2672 gave the sequence as AEIKAVSAARL. This result indicates that the protein was either initially expressed from amino acid 27 or that the first 26 amino acids were lost to proteolysis. For Cc3125 the first six amino acid residues had the sequence QVSYVR. This result indicates that the first 21 amino acid residues were lost to proteolysis. The metal content of the purified proteins was determined with a Perkin-Elmer Analyst 700 atomic absorption spectrometer and by inductively coupled plasma emission−mass spectrometry (ICP-MS). Cc2672 contained 2.0 equiv of Zn^{2+} and Cc3125 contained an average of 0.7 equiv of Zn^{2+}. Sgx9359b did not contain any metal after completion of the initial purification. The active site was reconstituted with metal by adding 2 equiv of ZnCl$_2$ and 10 mM potassium bicarbonate overnight in 50 mM Hepes, pH 7.5, to the purified Sgx9359. The protein was subsequently passed through a PD-10 desalting column equilibrated with metal-free 50 mM Hepes, pH 7.5, to remove loosely bound Zn^{2+} ions. The reconstituted protein was found to contain 1.8 equiv of Zn^{2+} as determined by ICP-MS.

Peptidase Activity Measurements. A modification of a ninhydrin-based colorimetric assay was adopted for the peptidase activity of Cc2672, Cc3125, and Sgx9359b (*19*). The Cd−ninhydrin reagent solution was prepared by dissolving 0.4 g of ninhydrin into 40 mL of 99.5% ethanol and 5 mL of acetic acid. To this solution was added 0.5 g of CdCl$_2$ dissolved in 0.5 mL of water. The peptide hydrolysis reactions were conducted at 30 °C in 50 mM Hepes, pH 8.0. Routinely, 120 μL of the reaction solution and 240 μL of the Cd−ninhydrin reagent were mixed and then heated in a 96-well block for 5 min in an 80 °C water bath. After being cooled to room temperature, 250 μL of the reaction solution from each well was transferred to a 96-well UV−visible microplate, and the absorbance at 507 nm was recorded with a SPECTRAmax plate reader from Molecular Devices. The measurement of the hydrolysis of *N*-formyl- and *N*-acetyl-L-amino acid derivatives was performed using the same procedure. A quantitative analysis of the liberated amino acids was conducted by the Protein Chemistry Laboratory at Texas A&M University.

Screening of Dipeptide Libraries. Each of the L-Xaa-L-Xaa dipeptide libraries consisted of a mixture of dipeptides with an identical L-amino acid at the N-terminus but 19 different amino acids at the C-terminus (L-cysteine was not included in any of the libraries). A total of 19 dipeptide libraries containing 361 different dipeptides were used to screen the dipeptidase activity of Cc2672, Cc03125, and Sgx9359b. For each dipeptide library, 120 μL of a reaction mixture containing ~0.1 mM each of dipeptide in 50 mM Hepes, pH 8.0, was incubated at 30 °C for variable lengths

of time with multiple concentrations of enzyme. The reactions were quenched by adding 240 μL of the ninhydrin reagent, and then the samples were incubated for 5 min at 80 °C before rapid cooling to room temperature. For each dipeptide library, 10–1000 nM enzyme was used in the initial screening assays with each dipeptide library. Control reactions with no enzyme or added substrate were conducted simultaneously. The hydrolytic rates for the enzymatic turnover of the 19 dipeptide libraries were determined based on the change in absorbance at 507 nm after subtraction of the background for the two control reactions. The relative rates were determined by a fit of the data to eq 1, where y is the change in absorbance at 507 nm, x is concentration of enzyme, and b is the relative rate constant. These experiments measure the rate of formation of free amino acids within the entire library. In these measurements we have elected to vary the enzyme concentration for a fixed period of time rather than vary the time at a fixed enzyme concentration.

$$y = a(1-e^{-bx}) \qquad (1)$$

Amino Acid Analysis. HPLC methods were employed to identify the specific amino acids that were released from the dipeptide libraries after the addition of enzyme. Selected dipeptide libraries were used to determine the substrate specificity for each of the three enzymes under study. The hydrolysis reactions were conducted in 25 mM ammonium bicarbonate buffer, pH 8.0. For each dipeptide library, variable amounts of enzyme (1–1000 nM) were used to quantify the products for each enzyme reaction. Two dipeptide libraries (L-Ala-L-Xaa and L-Phe-L-Xaa) were assayed with Cc2672, two dipeptide libraries (L-Met-L-Xaa and L-Leu-L-Xaa) were used for Cc31250, and two dipeptide libraries (L-Ala-L-Xaa and L-Leu-L-Xaa) were used for Sgx9359b. Routinely, 30 μL of a reaction mixture containing variable concentrations of enzyme and ~0.1 mM each of dipeptide in a 96-well block was incubated at 30 °C for 2 h (in the case of Cc2672) or overnight (in the case of Cc3125 and Sgx9359b). A 25 μL aliquot was removed from each reaction and diluted to 200 μL with water. The enzyme was removed by filtration through an Amicon centrifuge filter column (Ultracel YM-10). The samples were dried under reduced pressure and then reconstituted with water prior to quantitative amino acid analysis. Relative reaction rates were determined by fitting the change in amino acid concentration as a function of enzyme concentration according to eq 1.

Kinetic Measurements for Individual Substrates. Selected peptides, *N*-formyl-L-Arg, *N*-formyl-L-Lys, *N*-acetyl-L-Arg, and *N*-acetyl-L-Lys, were used as substrates to measure the kinetic parameters for Cc2672, Cc3125, and Sgx9359b. The kinetic parameters, k_{cat}, K_m, and k_{cat}/K_m were determined using the ninhydrin assay in 50 mM Hepes, pH 8.0. The kinetic constants were determined by fitting the initial velocity data to eq 2, where v is the initial velocity, k_{cat} is the turnover number, E_t is the enzyme concentration, A is the substrate concentration, and K_m is the Michaelis constant. Data conforming to competitive inhibition were fit to eq 3, where K_i is the competitive inhibition constant and the other constants have been previously defined.

$$v/E_t = k_{cat}A/(K_m + A) \qquad (2)$$

Table 1: Data Collection and Refinement Statistics for Sgx9359b

	Sgx9359b	
	Zn free (3BE7)	Zn bound (3DUG)
Data Collection		
space group	$P2_12_12_1$	$P2_12_12_1$
unit cell length (Å)	a = 113.42	a = 113.32
	b = 146.58	b = 146.04
	c = 255.96	c = 255.64
unit cell angle (deg)	α = 90.00	α = 90.00
	β = 90.00	β = 90.00
	γ = 90.00	γ = 90.00
diffraction protocol	single wavelength	single wavelength
wavelength used (Å)	0.9793	0.9793
resolution range (Å)	2.15–50.0	2.620–50.0
no. of unique reflections	230541	127892
highest resolution bin (Å)	2.15–2.23	2.62–2.71
no. of total reflections	5199028	2635392
redundancy	4.200	3.600
completeness (%)	94.8 (87.4)	99.4 (98.4)
R_{merge}	0.13 (0.52)	0.11 (0.79)
$\langle I/\sigma I\rangle$	3.7 (0.8)	4.1 (0.7)
Refinement		
$\langle FOM\rangle$ (overall)	0.938	0.92
resolution range (Å)	20.00–2.30	20.00–2.62
reflections used for refinement (obsd)	175832	122571
completeness (%)	99.6	99.1
reflections used for R_{free} (3%)	5451	3808
R_{factor}	0.22602	0.23008
R_{free}	0.27981	0.26364
rms bonds (Å)	0.006	0.008
rms angles (deg)	1.050	1.175
average B factor	48.35	75.91
Ramachandran plot statistics (PROCHECK)		
residues in most favored region (%)	90.2	90.6
residues in additionally allowed region (%)	9.1	8.6
disallowed region (%)	0.5	0.6

$$v/E_t = k_{cat}A/(K_m(1 + (I/K_i)) + A) \qquad (3)$$

Structure Determination and Refinement. Diffraction quality crystals of Sgx9359b were obtained by sitting drop vapor diffusion after mixing 1 μL of protein solution (21 mg/mL, containing 10% glycerol) with 1 μL of mother liquor composed of 100 mM succinic acid, pH 7.0, and 13–17% of PEG3350. Rod-shaped crystals appeared in 3–4 days and continued to grow for an additional 2–3 days. Protein crystals were soaked for 1–10 min in mother liquor supplemented with 15–30% glycerol as a cryoprotectant. To obtain a Zn-bound complex, crystals were soaked in cryoprotectant supplemented with 5 mM zinc sulfate for 5–10 min before freezing in liquid nitrogen. X-ray diffraction data were collected from crystals cooled in a nitrogen stream (100 K) at the National Synchrotron Light Source beamline X29A (Brookhaven National Laboratory, Upton, NY). Data were indexed, integrated, and scaled with the HKL2000 software package (*20*). Statistics are provided in the Table 1.

The crystal structures were determined by molecular replacement using MOLREP (CCP4 package suit) and the PDB file 2R8C as the search model (*21*). The Zn-free and Zn-substituted crystals were isomorphous and exhibited diffraction consistent with the orthorhombic space group $P2_12_12_1$. The asymmetric unit contains eight protein monomers arranged as a homo-octamer with 422 point symmetry,

allowing for the application of noncrystallographic symmetry restraints (NCS) throughout refinement. Models were refined with REFMAC 5.3 (CCP4 package suit) (*22*). The Zn-free structure was refined to a resolution of 2.3 Å, with R_{work} and R_{free} of 22.6% and 28.0%, respectively, and the Zn-bound structure was refined to a resolution of 2.62 Å with R_{work} and R_{free} of 23.0% and 26.3%, respectively (Table 1). All interactive model building was performed with COOT (*23*); the stereochemistry of the models was verified by PROCHECK (*24*).
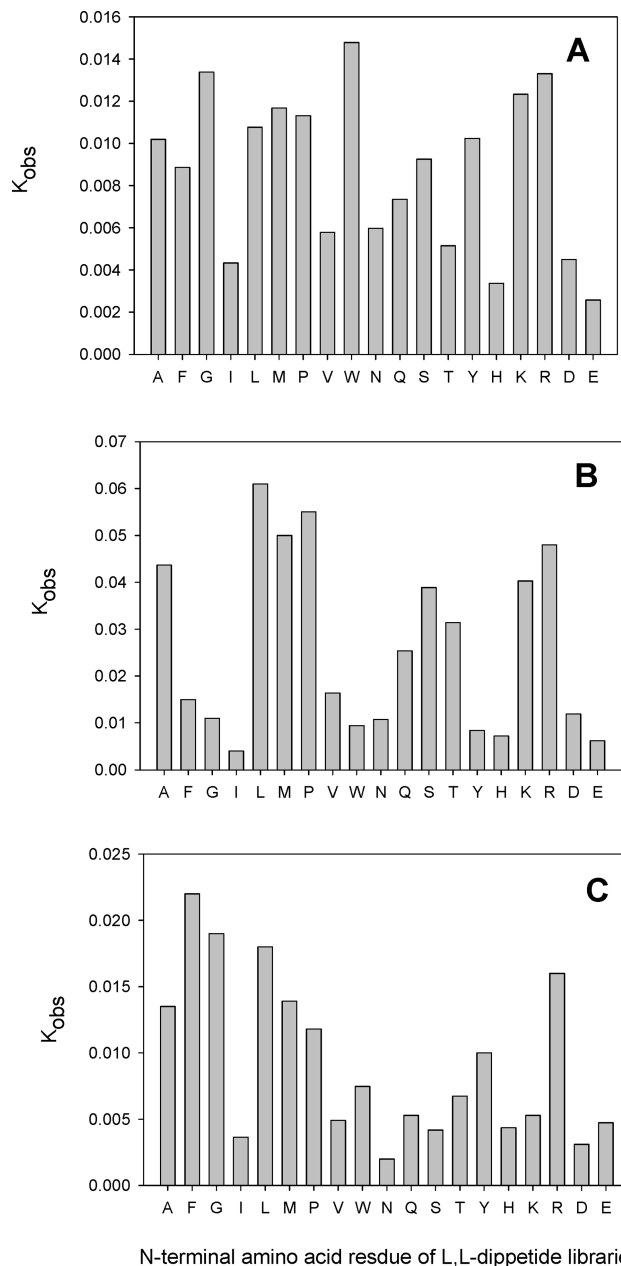
CLUSTALW (http:www.ebi.ac.uk/Tools/clustalw2) was used for determining sequence alignments. Rendering of sequence alignments was performed with ESpript 2.2 (http//expript.ibcp.fr/ESPript/ESPript). Molecular modeling was performed using CPHmodels online software (http://www.cbs.dtu.dk/services/CPHmodels/) and the ligands (arginine or lysine) were fitted manually. Figures were prepared with PyMOL (*25*) and Adobe Photoshop 7.0.

Differential Scanning Fluorometry (ThermoFluor). Fluorescence-monitored thermal denaturation was performed with 20 μL samples per well in a 96-well plate format at a protein concentration of 10 μM in the presence of SYPRO Orange (Invitrogen), screening buffer (100 mM Hepes, pH 7.5, and 150 mM NaCl), and an arginine concentration range of 1−2500 μM (*26*). The 96-well plate was sealed with optical-quality sealing tape (Bio-Rad) and the temperature ramped in an iCycler iQ real time PCR system (Bio-Rad) over a range of 20−95 °C in 1 °C increments. The excitation and emission wavelengths were 490 and 575 nm, respectively. Changes in fluorescence intensity were monitored by the iCycler's internal CCD camera, and the data were fit to a two-state transition to determine T_m.

RESULTS

Purification and Properties of Cc2672, Cc3125, and Sgx9359b. The genes for Cc2672 and Cc3125 were amplified from *C. crescentus* CB15 and cloned into pET-30a(+) for expression in *E. coli*. The protein Cc2672 expressed quite well when the plasmid harboring the gene for Cc2672 was transformed into BL21(DE3) Star cells. The protein was purified to homogeneity and found to contain 2.0 equiv of Zn^{2+} per subunit. The purified protein gave a single band on SDS−PAGE with an apparent molecular mass of 43 kDa, which is slightly smaller than the theoretical molecular mass of 45 kDa calculated from the reported gene sequence. The isolated enzyme was missing 26 amino acid residues from the N-terminus. The protein Cc3125 was purified using a nearly identical protocol as for the isolation of Cc2672. However, the yield of the isolated protein was lower due to poor solubility. The purified protein contained an average of 0.7 equiv of Zn^{2+} per subunit and was missing the first 21 amino acids from the N-terminus. After the initial purification of Sgx9359b, it was determined that no metals were bound to the protein. However, the protein was reconstituted with Zn^{2+}, and 1.8 equiv of Zn^{2+} was bound to the protein per subunit after overnight incubation with 10 mM bicarbonate and Zn^{2+}.

Screening of Dipeptide Libraries. Nineteen dipeptide libraries were used to screen the dipeptidase activity of Cc2672, Cc3125, and Sgx9359b using the ninhydrin assay for measurement of free amino acids. The relative preferences

FIGURE 1: Relative substrate activities for the hydrolysis of 19 dipeptide libraries of the type L-Xaa-L-Xaa, where a single amino acid is at the N-terminus (defined on the *x*-axis) and 19 different amino acids were at the C-terminus. (A) Cc2672; (B) Cc3125; (C) Sgx9359b.

for the first amino acid residue within each of these dipeptide libraries for the three proteins were determined and illustrated in Figures 1. These experiments measure the average rate of formation of free amino acids within each of the dipeptide libraries. All three proteins are promiscuous for the amino acid at the N-terminus, but the relative rates of hydrolysis vary within each of the dipeptide libraries. Cc2672 shows the highest activity with the L-Trp-L-Xaa library whereas Cc3125 shows the most pronounced activity with the L-Leu-L-Xaa library. The L-Phe-L-Xaa library was the best substrate for Sgx9359b. All three proteins exhibit very low activity with the L-Ile-L-Xaa, L-Asp-L-Xaa, and L-Glu-L-Xaa libraries.

Substrate Specificity of Cc2672, Cc3125, and Sgx9359b. Amino acid analysis was used to determine the relative rates of hydrolysis for specific dipeptides within selected dipeptide libraries. Cc2672 was assayed with the L-Ala-L-Xaa and
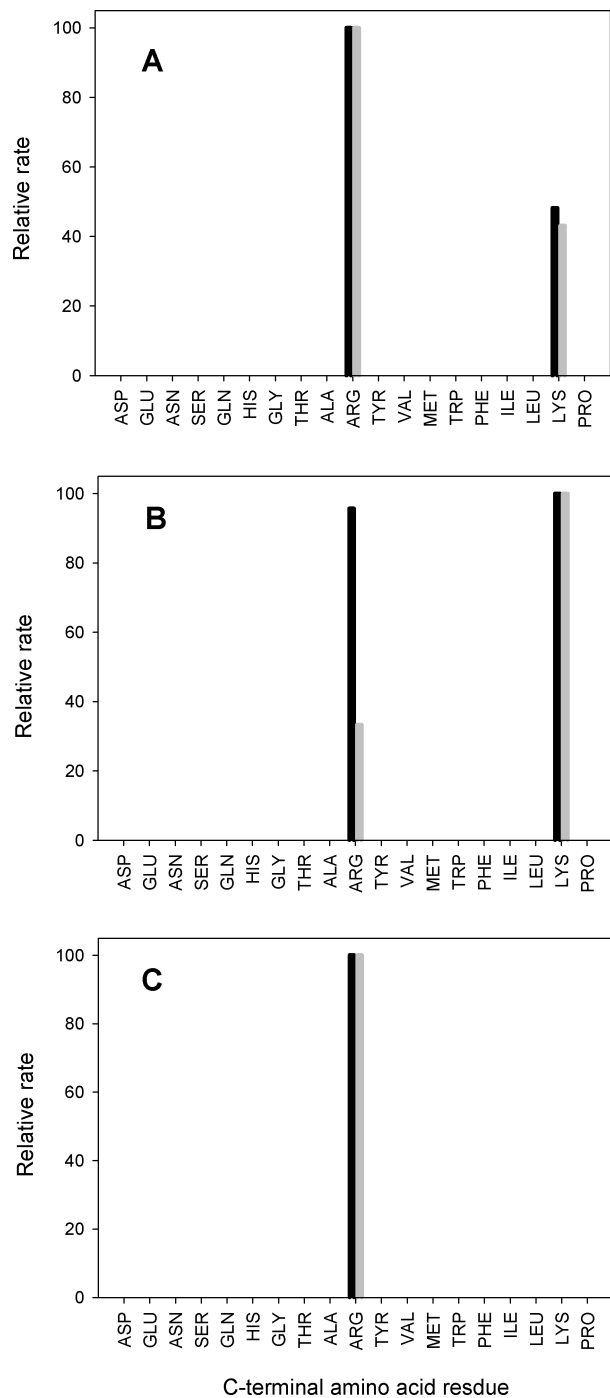
FIGURE 2: Amino acid analysis after the hydrolysis of selected dipeptide libraries by Cc2672, Cc3125, and Sgx9359b. (A) Cc2672 with L-Ala-L-Xaa (black) and L-Phe-L-Xaa (gray); (B) Cc3125 with L-Met-L-Xaa (black) and L-Leu-L-Xaa (gray); (C) Sgx9359b with L-Ala-L-Xaa (black) and L-Leu-L-Xaa (gray). Additional details are provided in the text.

Table 2: Kinetic Parameters of Cc2672, Cc3125, and Sgx9359b[a]

enzyme	substrate	k_{cat}/s^{-1}	k_m/mM	k_{cat}/K_m $(M^{-1}\,s^{-1})$
Cc2672	L-Ala-L-Arg	44 ± 2	0.42 ± 0.03	$(1.1 \pm 0.1) \times 10^5$
	L-Ala-L-Lys	16 ± 1	0.26 ± 0.02	$(6.2 \pm 0.5) \times 10^4$
	N-acetyl-L-Arg	6.0 ± 0.2	0.45 ± 0.04	$(1.3 \pm 0.1) \times 10^4$
	N-acetyl-L-Lys	3.0 ± 0.1	0.21 ± 0.02	$(1.4 \pm 0.1) \times 10^4$
	N-formyl-L-Arg	29 ± 1	0.36 ± 0.03	$(8.0 \pm 0.7) \times 10^4$
	N-formyl-L-Lys	6.3 ± 0.2	0.15 ± 0.02	$(4.2 \pm 0.5) \times 10^4$
	L-Gly-L-Phe-L-Arg	27 ± 3	1.4 ± 0.2	$(1.9 \pm 0.3) \times 10^4$
Cc3125	L-Ala-L-Arg	1.3 ± 0.1	0.9 ± 0.1	$(1.5 \pm 0.2) \times 10^3$
	L-Ala-L-Lys	5.3 ± 0.2	0.10 ± 0.01	$(5.3 \pm 0.5) \times 10^4$
Sgx9359b	L-Ala-L-Arg	3.2 ± 0.2	0.26 ± 0.05	$(1.2 \pm 0.2) \times 10^4$
	N-acetyl-L-Arg	0.7 ± 0.04	0.62 ± 0.10	$(1.1 \pm 0.2) \times 10^3$
	N-formyl-L-Arg	1.4 ± 0.1	0.53 ± 0.06	$(2.6 \pm 0.3) \times 10^3$
	L-Gly-L-Phe-L-Arg	1.9 ± 0.1	0.98 ± 0.1	$(1.9 \pm 0.2) \times 10^3$

[a] The kinetic constants were determined in 50 mM Hepes, pH 8.0, from fits of the data to eq 2.
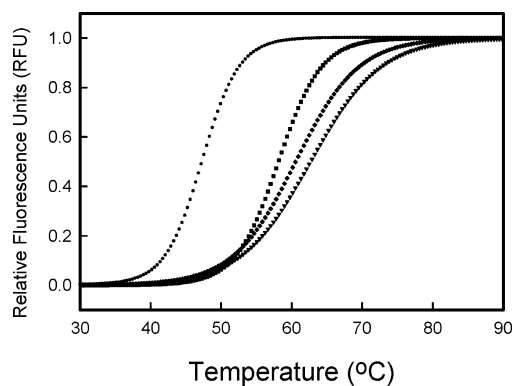

FIGURE 3: Thermofluor analysis for the stabilization of Sgx9359b by added arginine. The concentrations of arginine are as follows: (●) 0 μM; (■) 25 μM; (◆) 500 μM; (▼) 2500 μM. Additional details are provided in the text.

L-Phe-L-Xaa libraries, we assume that all of the other dipeptide libraries will exhibit the same specificity for the C-terminus. Under these reaction conditions the catalytic activity of Cc2672 for L-Xaa-L-Arg is approximately twice that of L-Xaa-L-Lys. The relative reaction rates are presented in Figure 2a. The relative rates of hydrolysis for the other dipeptides in these libraries are less than 1%.

The enzyme Cc3125 was assayed with the dipeptide libraries L-Met-L-Xaa and L-Leu-L-Xaa. With the L-Met-L-Xaa library only arginine, lysine, and methionine were detected after amino acid analysis. When the L-Leu-L-Xaa library was used as a substrate, only arginine, lysine, and leucine were identified as the reaction products. Therefore, this enzyme is specific for the release of arginine and lysine from the C-terminus of dipeptides. The relative rates are illustrated in Figure 2B. Under these reaction conditions Cc3125 exhibited a higher activity with L-Leu-L-Lys than for L-Leu-L-Arg, but with the L-Met-L-Xaa library the relative rates were approximately the same. The upper limit for the hydrolysis of other amino acids at the C-terminus is less than 1%.

The Sgx9359b enzyme was assayed with the L-Ala-L-Xaa and L-Leu-L-Xaa dipeptide libraries. With the L-Ala-L-Xaa library only alanine and arginine were identified as free amino acids whereas with the L-Leu-L-Xaa library only leucine and arginine were found. Therefore, this enzyme is specific for the hydrolysis of L-Xaa-L-Arg dipeptides. The upper limit for the hydrolysis of L-Ala-L-Lys relative to the hydrolysis of L-Ala-L-Arg is 1%. The relative rates for the hydrolysis of the two libraries are shown in Figure 2C.

L-Phe-L-Xaa dipeptide libraries. The amino acids detected from the hydrolysis of the L-Ala-L-Xaa library were limited to alanine, arginine, and lysine. The amino acid products detected from the hydrolysis of the L-Phe-L-Xaa library were limited to phenylalanine, arginine, and lysine. Since alanine and phenylalanine were only detected when libraries containing these amino acids were at the N-terminus, the substrate specificity of Cc2672 at the C-terminus is limited to arginine and lysine. All of the other dipeptides libraries have not been tested in this manner. However, based upon the finding of only arginine and lysine residues from the L-Ala-L-Xaa and
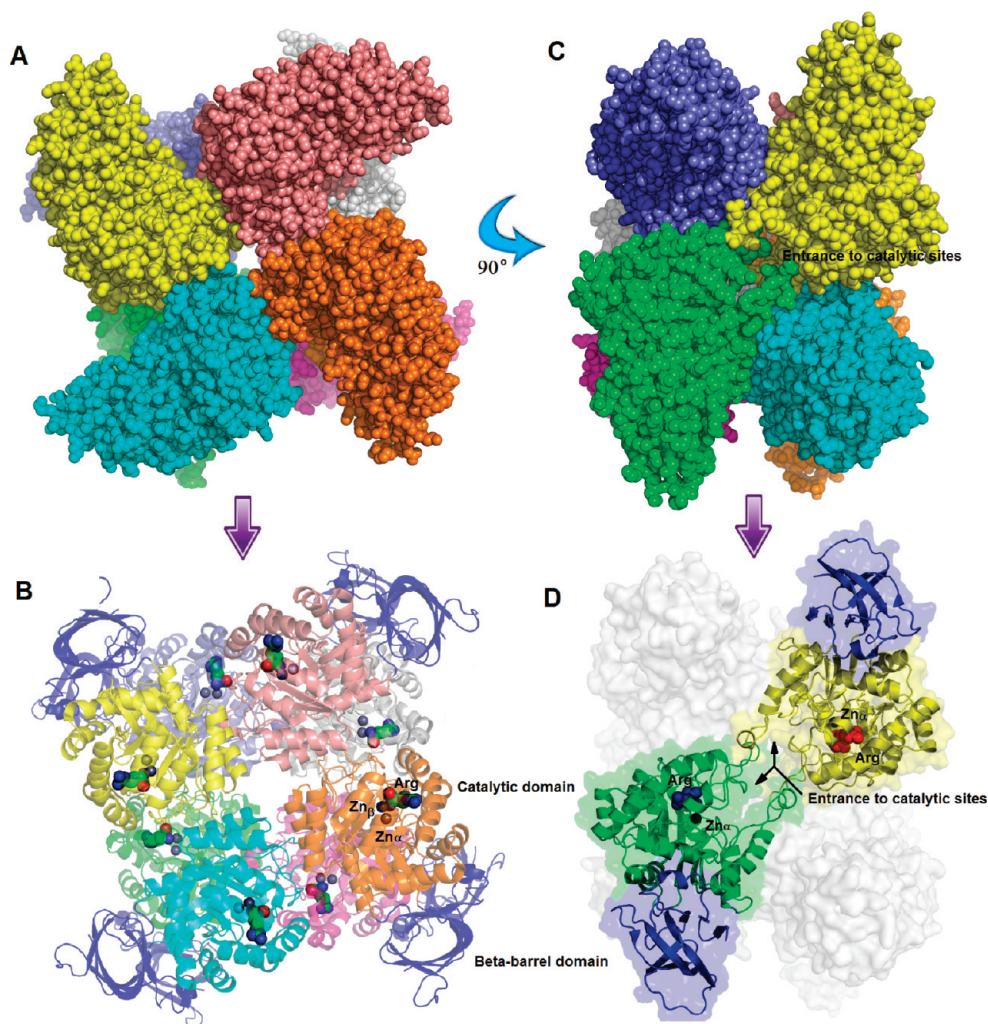
FIGURE 4: Conformation of the Sgx9359b octamer. (A) The space-filled model of Sgx9359b as a homooctamer composed of two identical tetramers related by C2-fold symmetry. (B) The same model shown as a cartoon with Zn and bound arginine atoms drawn as spheres. (C) The model shown in panel A rotated 90° counterclockwise around the C2-fold symmetry axis. (D) The same view as in panel C. Two monomers are shown as cartoons and represent one of the four pseudodimers within the same octamer. Each dimer has one overlapping entrance to both active sites. Arginine and zinc atoms are drawn as spheres.
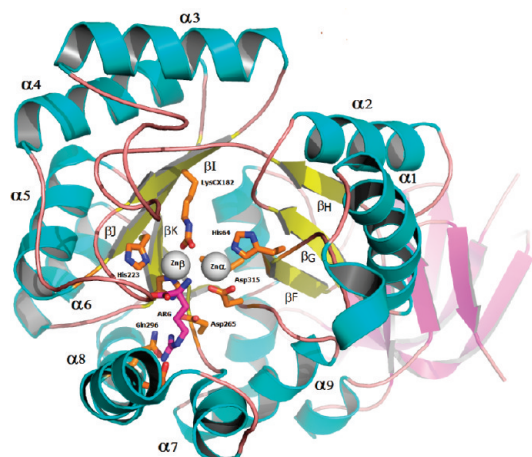

FIGURE 5: Conformation of the Sgx9359b monomer drawn as a cartoon model. The catalytic TIM-barrel domain is painted in blue (α-helices) and yellow (β-sheets). The side chains of active site residues are shown as sticks, and the two zinc atoms are shown as spheres.

Deacetylase and Tripeptidase Activity. Cc2672 was tested for the ability to hydrolyze other substituted amino acid derivatives. The enzyme was tested with 19 N-formyl-L-Xaa

and N-acetyl-L-Xaa (except L-cysteine) derivatives. The only compounds hydrolyzed were the N-formyl- and N-acetyl derivatives of L-lysine and L-arginine. Cc2672 was also tested with two tripeptides, L-Gly-L-Phe-L-Arg and L-Gly-L-Ala-L-Tyr. This enzyme was unable to hydrolyze L-Gly-L-Ala-L-Tyr, but it was able to hydrolyze the terminal L-arginine from L-Gly-L-Phe-L-Arg. Cc3125 was unable to hydrolyze any of the N-formyl- or N-acetyl-L-amino acid derivatives. It was also not able to hydrolyze the L-Gly-L-Phe-L-Arg tripeptide. Sgx9359b was found to hydrolyze N-formyl-L-Arg, N-acetyl-L-Arg, and L-Gly-L-Phe-L-Arg but does not exhibit any hydrolytic activity when the C-terminal amino acid is any amino acid other than L-arginine.

Kinetic Constants for Cc2672, Cc3125, and Sgx9359b. The kinetic constants for the three enzymes were tested with L-Ala-L-Arg, L-Ala-L-Lys, N-formyl-L-Arg, N-formyl-L-Lys, N-acetyl-L-Arg, and N-acetyl-L-Lys. Also tested was the tripeptide L-Gly-L-Phe-L-Arg. The kinetic data were fit to eq 2, and the catalytic constants are presented in Table 2. The N-methyl phosphonate derivative of L-lysine (**1**) was tested as an inhibitor for the hydrolysis of N-acetyl-L-lysine by the enzyme Cc2672. It was a potent competitive inhibitor with a K_i value of 120 nM from a fit of the data to eq 3.
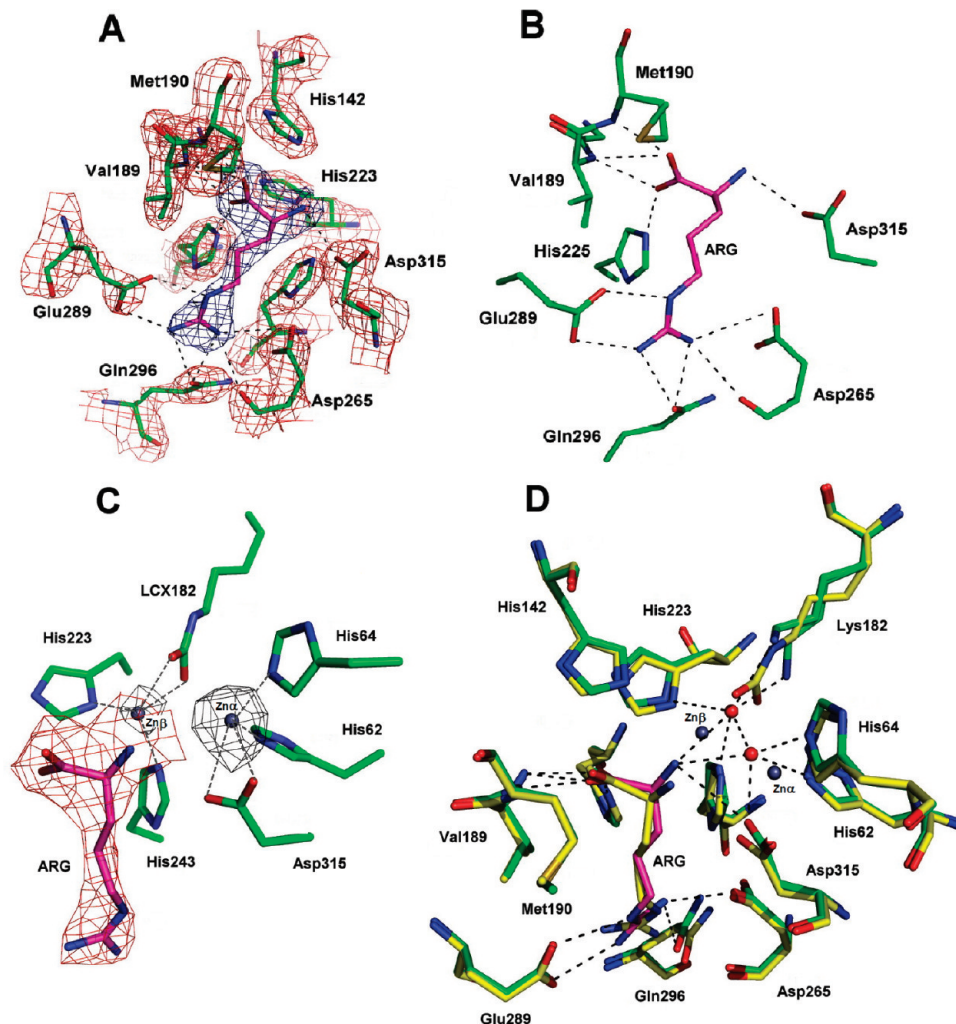
FIGURE 6: (A) $2F_o - F_c$ electron density map (σ cutoff 1.0) drawn around residues at the active site of a Zn-free Sgx9359b monomer. Arginine atoms (C-carbons are pink) were omitted from map calculations. (B) Similar to panel A where polar interactions between ligand and protein atoms are shown as black dashes. (C) $2F_o - F_c$ electron density map (red colored, σ cutoff 1.0) around arginine at the active site of the Zn-bound Sgx9359b monomer. $2F_o - F_c$ (gray, σ cutoff 4.0) is drawn around the two zinc atoms coordinated by active site residues. (D) Superimposition of coordinates for active site residues of the two models: Zn free (C-carbons in green and water molecules in red) and Zn bound (C-carbons in yellow). Atomic coordinates of monomer A were used for making images A through D.

ThermoFluor. The purified Sgx9359b exhibited a clean two-state melting transition (Figure 3), characterized by a T_m of 47.6 °C. A significant increase in T_m (i.e., greater than 10 °C stabilization) was observed at arginine concentrations of 25 μM or higher. This stabilizing effect appeared to plateau at approximately 500 μM arginine (data not shown).

Structure of Sgx9359b. The crystal structure of Sgx9359b reveals a homooctamer in the asymmetric unit (Figure 4), which buries in excess of 30000 Å^2 of accessible surface area, suggesting that this represents the physiologically relevant assembly. Consistent with this assignment, gel filtration chromatography indicates that Sgx9359b has an apparent molecular weight of over 250000 (data not shown). The two structures, Zn-free (PDB code 3BE7) and Zn-bound (PDB code 3DUG), are nearly identical with rmsd (root mean squared deviation) less than 0.3 Å for all C$_\alpha$. These structures can be described as two tightly associated tetramers related by 2-fold rotational symmetry with overall 422 point symmetry (Figure 4). Each protein monomer consists of two structural domains, the catalytic TIM-barrel domain and a smaller β-barrel domain (see Figures 4B and Figure 5). The β-barrel domain is composed of an outer layer formed by

the extreme N-terminal segment (residues 1−55) and an inner layer formed by the extreme C-terminal segment (residues 358−398) of each monomer. The former is exposed to solvent and is the most flexible part of the molecule, with relatively weak electron density and high B factors. In contrast, the residues of the inner layer are well ordered, being sandwiched between the outer layer of the β-barrel domain and the TIM-barrel domain.

Within the octamer, each monomer makes contacts with four other monomers. This arrangement separates the catalytic sites by ∼40 Å, and notably each catalytic site is formed exclusively by residues contributed by a single monomer (Figure 4B,D). A particularly interesting feature is depicted in Figure 4C,D, where two adjacent monomers are interconnected by extended loops, resulting in the formation of a single narrow continuous entrance that provides access to both active sites. This "pseudodimer" consists of monomers related by a local 2-fold symmetry, with four such pseudodimers per octamer.

The catalytic domain is a typical TIM-barrel structure (Figure 5), composed of 8 inner β-strands surrounded by 11 outer α-helices. A single binuclear zinc binding site is
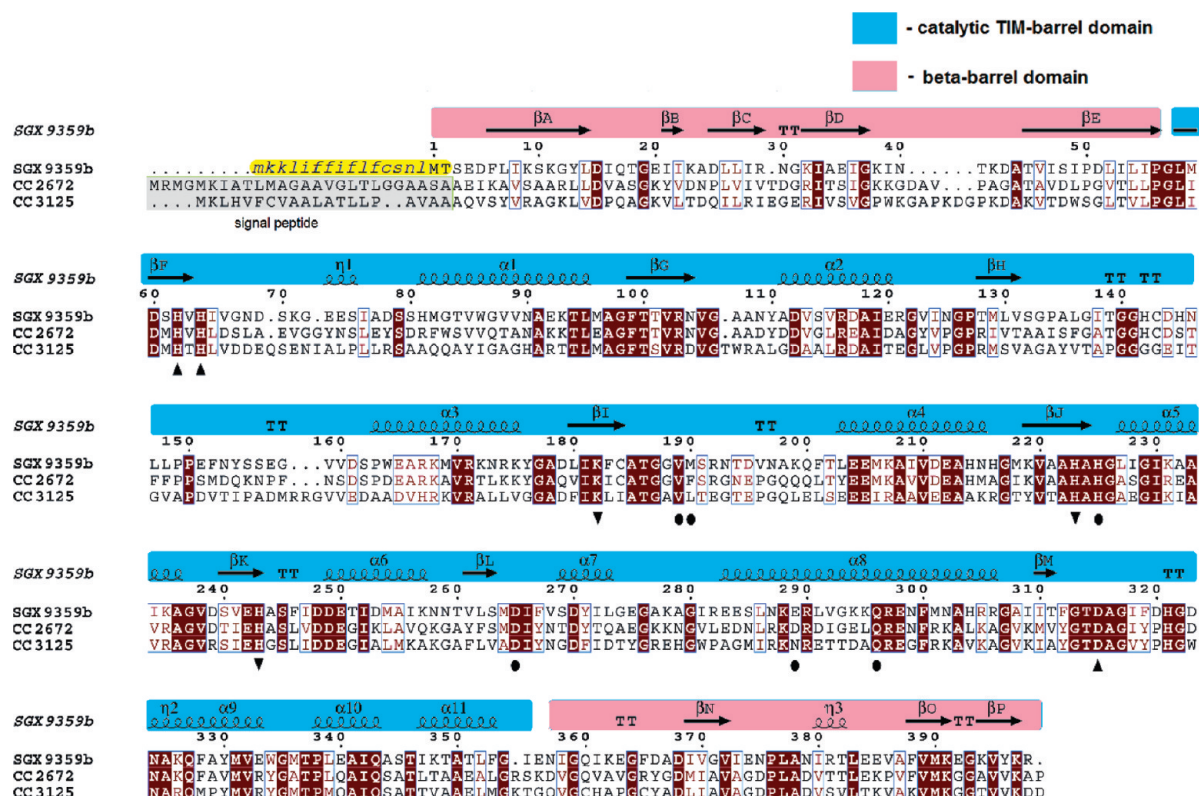
FIGURE 7: Sequence alignment of Sgx9359b, Cc2672, and Cc3125. The zinc binding residues are marked by filled triangles, and the arginine binding residues are marked with dots. The Sgx9359b sequence is labeled according to the protein database entry EAI78147. The sequence in italic (yellow) was translated from the DNA database entry AACY01081974. This sequence likely represents a signal peptide missing from the protein entry EAI78147. Shown in light brown shading are the regions of Cc2672 and Cc3125 that were processed in *E. coli* prior to purification.

composed of six conserved residues, including four histidines (His-62, His-64, His-223, His-243), Asp-315, and Lys-182. The latter is carboxylated in the presence of Zn but appears unmodified in the absence of Zn (see below). The zinc binding site is located in a cavity formed by one short and two longer loops connecting $\alpha7$ and $\alpha8$, βF and $\alpha1$, and βH and $\alpha3$, respectively. The short loop is relatively flexible and solvent accessible, while the two long loops are rigid due to numerous interactions along the octamer interface. Openings between the loops create access to the active site.

Active Site of Sgx9359b. In the process of refining the Zn-free structure, residual electron density that could not be attributed to protein persisted near the putative Zn binding site in all eight monomers, with subunit A exhibiting the most prominent features. This density is well accommodated by arginine, which also affords a highly plausible hydrogen-bonding network (Figure 6A,B). As arginine was not added to either the purified protein or the crystallization buffers, this ligand must possess sufficiently high affinity to be retained during purification and subsequent crystallization. The high complementarity of arginine to its binding site leaves little space for solvent; only two water molecules are located in the vicinity of the guanidine group, although there are no direct interactions. In the A monomer, weak density is present that is suggestive of a short peptide containing a C-terminal arginine.

The Zn-soaked structure at 2.62 Å exhibited two very prominent features, modeled as Zn atoms, in the active site (Figure 6C); strong corresponding features (greater than 10σ) were also present in anomalous difference Fourier maps (not shown). The refined occupancies for the two Zn atoms are slightly different. Zn_α is coordinated by the side chains of His-62, His-64, and Asp-315 and refined to an occupancy close to 1.0. Zn_β is coordinated by the side chains of His-223, His-243, and the carboxylated Lys-182 and refined to an occupancy of about 0.8. Additional Zn binding sites were identified in each monomer but are located on the protein surface and are distant from the active site. It is notable that a soak of 5 min in 5.0 mM zinc sulfate was sufficient for almost complete carboxylation of Lys-182. In the Zn-free structure, Lys-182 was unmodified and likely adopts at least two alternate conformations (Figure 6D). Interestingly, the replacement of water by zinc does not change the overall conformation of the active site, and the relative orientations of bound arginines and hydrogen bond patterns are quite similar in both the Zn-free and Zn-bound structures.

DISCUSSION

Three enzymes with significant sequence identity to one another from the amidohydrolase superfamily have been purified and their substrate profiles determined for the first time. The genes for Cc2672 and Cc3125 were cloned from *C. crescentus*, and the gene for Sgx9359b was chemically synthesized from a DNA sequence originally harvested from the Sargasso Sea. All three enzymes have previously been annotated as L-Xaa-L-Pro dipeptidases, but a literature lineage to this postulated catalytic activity is obscure. These enzymes have now been found experimentally to catalyze the hydrolysis of dipeptides, but none of them is able to hydrolyze an L-Xaa-L-Pro dipeptide, and thus the current biochemical annotations for these three enzymes are incorrect. The
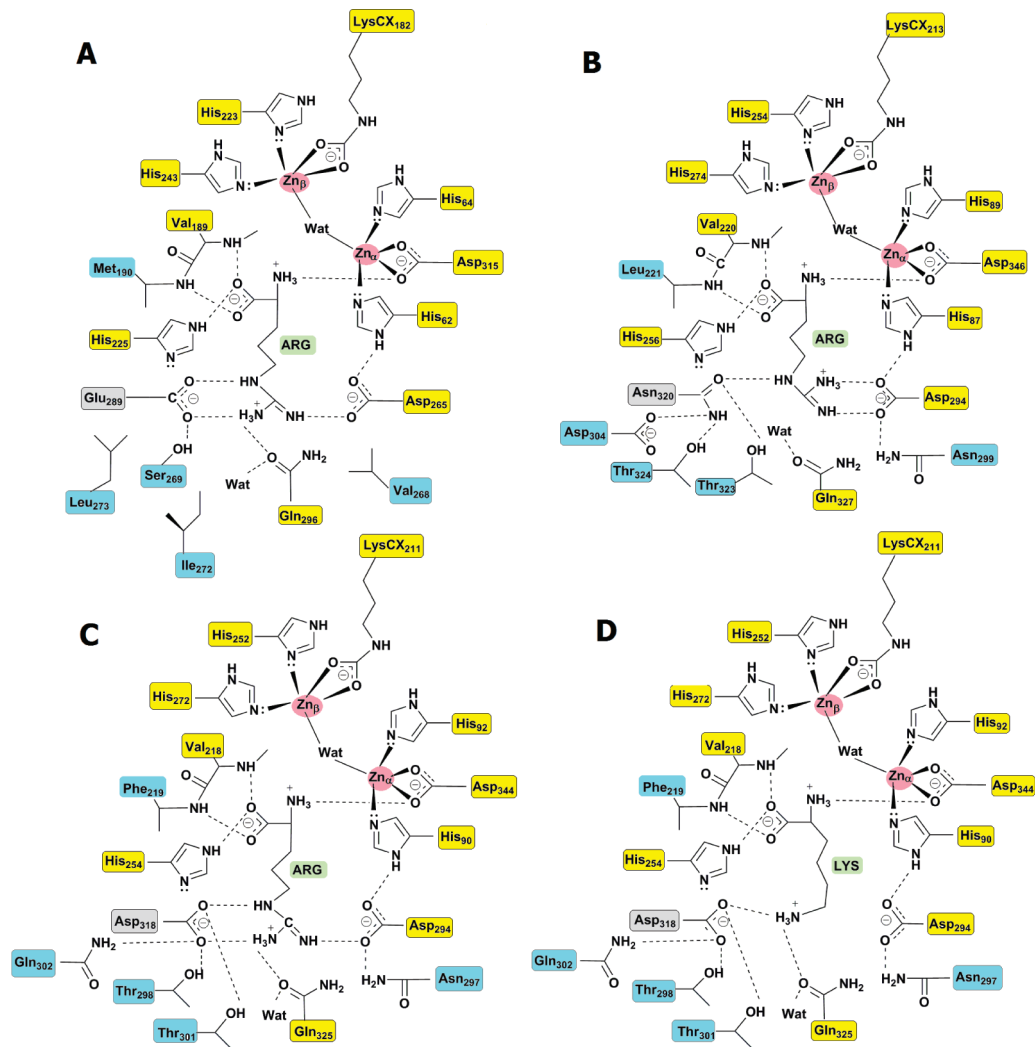
FIGURE 8: (A) Scheme of polar interactions between bound arginine and protein in the active site of Sgx9359b. Hydrogen bonds (distance cutoff 3.2 Å) are shown as dashes. Schemes of polar interactions at the active sites of Cc3135 (panel B) and Cc2672 (panels C and D) complexed with either arginine or lysine are based on molecular modeling. Conserved residues are marked in yellow, and variable residues are marked in blue. Glu-289 (Sgx9359b) and its counterparts, Asp-318 (Cc2672) and Asn-320 (Cc3125), are marked in gray. The larger number of polar groups in Cc2672 and Cc3125 relative to the number found in Sgx9359b allows for alternate patterns of hydrogen bonding.

substrate specificities for the three proteins are similar but not identical.

Cc2672 is able to hydrolyze dipeptides that contain either an L-arginine or L-lysine at the C-terminus. The best substrate determined thus far is L-Ala-L-Arg with a k_{cat} of 44 s^{-1} and a k_{cat}/K_m of 10^5 M^{-1} s^{-1}. In addition, this enzyme is also able to hydrolyze the *N*-acetyl and *N*-formyl derivatives of L-lysine and L-arginine and longer peptides that terminate in amino acids with a positively charged functional group. Thus, the substrate specificity is more correctly annotated as a carboxypeptidase with a requirement for a C-terminal L-arginine or L-lysine. The substrate specificity of Cc3125 is more restrictive in that only dipeptides terminating in either L-lysine or L-arginine are hydrolyzed. Simple *N*-acyl derivatives of these amino acids are not substrates nor are longer peptides hydrolyzed. The best substrate determined thus far is L-Ala-L-Lys with a k_{cat} of 5 s^{-1} and a k_{cat}/K_m of 5×10^4 M^{-1} s^{-1}. Compounds terminating in lysine are somewhat better substrates than those terminating in arginine.

The substrate profile for Sgx9359b is similar in many respects to that of Cc2672. However, the C-terminal residue

is restricted to L-arginine, and it will not catalyze the hydrolysis of compounds that terminate in L-lysine. Dipeptides are hydrolyzed with a broad tolerance for changes in the amino acid at the N-terminus. Tripeptides are also hydrolyzed in addition to the *N*-acetyl and *N*-formyl derivatives of L-arginine. The best substrate discovered thus far is L-Ala-L-Arg with a k_{cat} of 3 s^{-1} and a k_{cat}/K_m of 1×10^4 M^{-1} s^{-1}.

Sgx9359b is the only one of these three proteins to be successfully crystallized and its three-dimensional structure determined. Other proteins of different specificities share some structural similarity to Sgx9359b. For example, isoaspartyl dipeptidase (IAD) from *E. coli* has a TIM-barrel domain with a binuclear Zn binding site, and it also forms a homooctamer with considerable similarity to Sgx9359b (*27*). Quite remarkably, the Sgx9359b protein that was expressed and purified from *E. coli* was found to contain eight arginines bound to the eight catalytic sites of the octamer. This retention of ligand and the substrate specificity of Sgx9359b are consistent with the ThermoF-
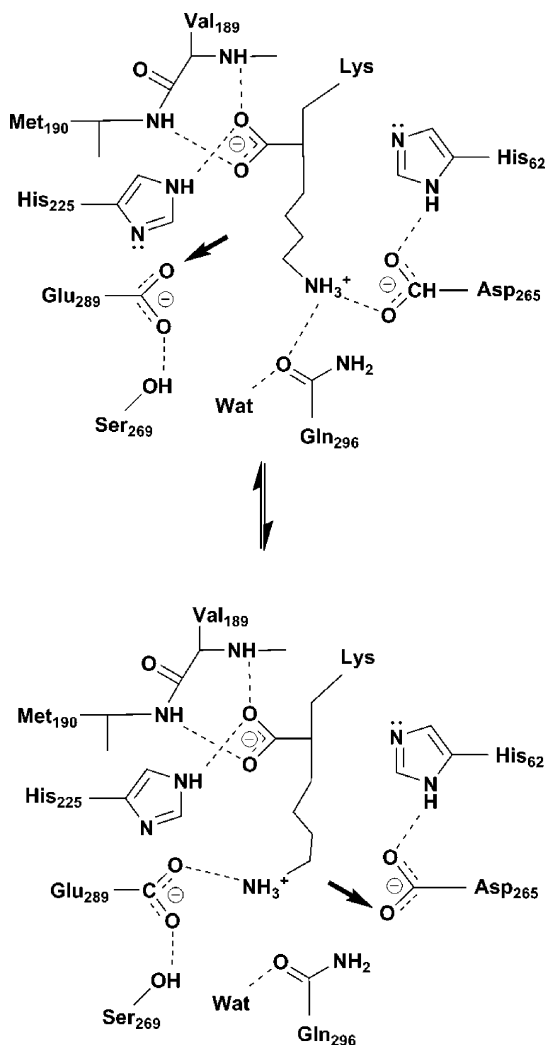
FIGURE 9: Fitting lysine as a ligand at the active site of Sgx9359b most likely leaves at least one hydrogen bond acceptor unsatisfied (shown by arrows). This may explain why this protein prefers arginine as a C-terminal residue of a substrate.

luor results demonstrating significant stabilization of the protein in the presence of low concentrations of arginine.

The structure of Sgx9359b with arginine bound to the active site as a product complex has revealed those amino acids in the active site that help to define the substrate specificity for these three enzymes. The guanidinium group of arginine is ion paired to the side chain carboxylates of Glu-289 and Asp-265. Asp-265 is found at the end of β-strand 7 and is conserved in all three enzymes examined in this paper. Glu-289 is located in the loop that follows β-strand 7. In Cc2672 this residue is an aspartate whereas in Cc3125 this residue is an asparagine. It would thus appear that the aspartate at the end of β-strand 7 is more important for recognition of the positively charged side chains of lysine and arginine. The α-carboxylate of the arginine product is ion paired with the side chain imidazole groups of His-142 and His-225. His-142 is found in the loop that follows β-strand 3, and His-225 is found at the end of β-strand 5. The histidine at the end of β-strand 5 is conserved in all three proteins, but the histidine in the loop that follows β-strand 3 is not conserved in Cc3125. This would suggest that the histidine after β-strand 5 is more significant.
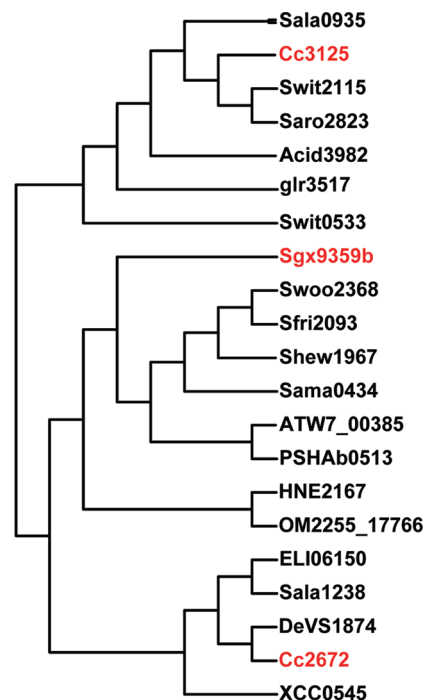

FIGURE 10: Dendrogram of bacterial homologues to Cc2672, Cc3125, and Sgx9359b. The gi numbers for the proteins in this dendrogram are as follows: Sala0935 (gi|103486425), Cc3125 (gi|16127355), Swit2115 (gi|148555030), Saro2823 (gi|87200836), Acid345_3982 (gi|94971008), glr3517 (gi|37523086), Swit0533 (gi|148553457), Sgx9359b (gi|44368820), Swoo2368 (gi|170726716), Sfri2093 (gi|170726716), Shew1967 (gi|114563264), Sama0434 (gi|127512895), ATW7_00385 (gi|119773574), PSHAb0513, (gi|77362422), HNE2167 (gi|114797205), OM2255_17766 (gi|114772900), ELI06150 (gi|85374057), Sala1238 (gi|103486725), DeVS1874 (gi|163813067), Cc2672 (gi| 16126907), XCC0545 (gi|77747758).

The α-amino group of the bound arginine in the active site of Sgx9359b is 3.2 Å away from the side chain carboxylate of Asp-315 and 3.3 Å away from the β-metal ion. Asp-315 is an invariant residue in all members of the amidohydrolase superfamily at the end of β-strand 8 that also ligates the α-metal ion. The position of this residue is consistent with previous proposals (*28*) that this aspartate is critical for the transfer of a proton from the attacking hydroxide or water to the leaving group (in this case the α-amino group of arginine).

Molecular Modeling of Cc2672 and Cc3125. All three proteins studied in this investigation share high amino acid sequence identities among the catalytic site residues. The sequence alignment is presented in Figure 7. This alignment has enabled the application of molecular modeling on the basis of the Sgx9359b structures to rationalize the overlapping yet distinct substrate preferences. As seen in Figure 8, hydrogen bond networks around the active sites of all three proteins look very similar with very few exceptions. The binuclear Zn binding site consists of four histidines, one aspartate, and one lysine. The arginine-interacting residues are almost identical, with one exception. The side chain from Glu-289 maintains two hydrogen bonds with the guanidinium group of arginine in the Sgx9359b structure. From the sequence alignment (Figure 7) and molecular modeling data (Figure 8) Cc2672 has Asp-318 and Cc3125 has Asn-320 in this position, although the adjacent residues are identical for all three

proteins. Another important feature of the ligand binding site is that the Sgx9359b structure has only four polar amino acid side chains in the vicinity of the guanidinium group (Asp-265, Ser-269, Glu-289, and Gln-298), three of which are in direct interactions with the ligand. By contrast, the remaining residues inside the active site pocket are hydrophobic (Figure 8), including Val-268, Ile-272, and Leu-273 adjacent to arginine-interacting residues. Attempts to dock lysine in the active site of Sgx9359b results in an unsatisfied hydrogen bond acceptor (Figure 9), suggesting a plausible explanation for the observed substrate specificity. In contrast, the Cc2672 counterparts of the hydrophobic residues in Sgx9359b are replaced by more polar residues (Figures 7 and 8), such as Asn-297, Thr-301, and Gln-302. As a result, the Cc2672 catalytic site has a more complex hydrogen bond network with all hydrogen bond donors and acceptors satisfied even in the absence of bound arginine or lysine. For example, Asp-294 may be hydrogen-bonded to His-90 and Asn-297, Asp-318 is hydrogen-bonded to Gln-302, Thr-298, and Thr-301, and the carbonyl oxygen of Gln-325 is connected to water. The extensive hydrogen bond network may create more flexibility that results in an equal opportunity for both arginine and lysine to bind into the active site of Cc2672. The same can be concluded about the active site of Cc3125 which is almost 80% identical to that of Cc2672, thus explaining their overlapping substrate specificities. One important difference between these two enzymes is that Cc3125 has Asn-320 instead of Asp-318 and Asp-304 instead of Gln-302. These pairs of residues are in close vicinity to one another, and most likely their side chains maintain hydrogen bond contact. However, since Asn-320 has only one hydrogen bond acceptor, the binding mode of the guanidino group of arginine will likely be different from that found in the Sgx9359b structure (Figure 8).

The discovery of function for these three enzymes represents the first demonstration of dipeptidase or carboxypeptidase activity for members of the amidohydrolase superfamily where the terminal amino acid in the substrate is restricted to either L-arginine and/or L-lysine. The critical residues located within the active sites of these three enzymes were used to identify other enzymes within the AHS with similar or identical catalytic functions. In particular, the specific residues that were shown to be involved in recognition of the side chain for either arginine or lysine at the C-terminus and the α-carboxylate of substrates were incorporated into the search model. For the three proteins characterized in this paper, the second histidine after β-strand 5 and the aspartate after the end of β-strand 7 are invariant (see Figure 7). A search of the nonredundant protein sequences in NCBI via a simple BLAST protocol identified 18 other sequences that are homologous to Cc2672, Cc3125, and Sgx9359b. A dendrogram that identifies these sequences is presented in Figure 10. Of these protein sequences, 3 of them, Sala0935 (from *Sphingopyxis alaskensis* RB2256), Swit2115 (from *Sphinomonas wittichii* RW1), and Saro2823 (from *Novosphinogobium aromaticivorans* DSN 12444) are more closely related to Cc3125. The remaining 15 sequences are closer to Cc2672 and Sgx9359b. In these sequences the histidine in the loop after β-strand 3 and the glutamate/

aspartate in the loop after β-strand 7 are also fully conserved. All of these proteins are expected to catalyze the hydrolysis of N-substituted derivatives of L-lysine and/or L-arginine.

A BLAST search against the *C. crescentus* genomic database shows that in addition to Cc2672 and Cc3125 there is at least one other protein that is closely related to Sgx9359b. This protein, Cc0300, is 38% identical in sequence to Sgx9359b. A sequence alignment (not shown) also demonstrated that all three proteins from *C. crescentus* are closely related. However, all of the residues from Cc0300 that potentially involve the binding of the C-terminal residue of substrates are hydrophobic, suggesting that Cc0300 is a probable carboxypeptidase with a specificity toward hydrophobic residues at the C-terminus.

REFERENCES

1. Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function. *Nature (London) 448*, 775–779.
2. Marti-Arbona, R., Xu, C., Steele, S., Weeks, A., Kuty, G. R., Seibert, C. M., and Raushel, F. M. (2006) Annotating enzymes of unknown function: *N*-formimino-L-glutamate deminase is a member of the amidohydrolase superfamily. *Biochemistry 45*, 1997–2005.
3. Nguyen, T. T., Brown, S., Fedorov, A. A., Fedorov, E. V., Babbitt, P. C., Almo, S. C., and Raushel, F. M. (2007) At the periphery of the amidohydrolase superfamily: Bh0493 from *Bacillus halodurans* catalyzes the isomerization of D-galcuronate to D-tagaturonate. *Biochemistry 47*, 1194–1206.
4. Seibert, C. M., and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry 44*, 6383–6391.
5. Holm, L., and Sander, C. (1997) An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins 28*, 72–82.
6. Aubert, S. D., Li, Y., and Raushel, F. M. (2004) Mechanism for the hydrolysis of organophosphates by the bacterial phosphotriesterase. *Biochemistry 43*, 5705–5715.
7. Jabri, E., Carr, M. B., Hausinger, R. P., and Karplus, P. A. (1995) The crystal structure of urease from *Klebsiella aerogenes*. *Science 268*, 998–1004.
8. Porter, T. N., Li, Y., and Raushel, F. M. (2004) Mechanism of the dihydroorotase reaction. *Biochemistry 43*, 16285–16292.
9. Wang, Z., and Quiocho, F. A. (1998) Complexes of adenosine deaminase with two potent inhibitors: X-ray structures in four independent molecules at pH of maximum activity. *Biochemistry 37*, 8314–8324.
10. Williams, L., Nguyen, T., Li, Y., Porter, T., and Raushel, F. M. (2006) Uronate isomerase: a nonhydrolytic member of the amidohydrolase superfamily with an ambivalent requirement for a divalent metal ion. *Biochemistry 45*, 7453–7462.
11. Li, T., Iwaki, H., Fu, R., Hasegawa, Y., Zhang, H., and Liu, A. (2006) α-Amino-β-carboxymuconic-ε-semialdehyde decarboxylase (ACMSD) is a new member of the amidohydrolase superfamily. *Biochemistry 45*, 6628–6634.
12. Hara, H., Masai, E., Katayama, Y., and Fukuda, M. (2000) The 4-oxalomesaconate hydratase gene, involved in the protocatechuate 4,5-cleavage pathway, is essential to vanillate and syringate degradation in *Sphingomonas paucimolbilis* SYK-6 hydration reaction. *J. Bacteriol. 182*, 6950–6957.
13. Afriat, L., Roodveldt, C., Manco, G., and Tawfik, D. S. (2006) The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase. *Biochemistry 45*, 13677–13684.
14. Marti-Arbona, R., Fresquet, V., Thoden, J. B., Davis, M. L., Holden, H. M., and Raushel, F. M. (2005) Mechanism of the reaction catalyzed by isoaspartyl dipeptidase from *Escherichia coli*. *Biochemistry 44*, 7115–7124.
15. Donarski, W. J., Dumas, D. P., Heitmeyer, D. P., Lewis, V., and Raushel, F. M. (1989) Structure-activity relationships in the

hydrolysis of substrates by the phosphotriesterase from *Pseudomonas diminuta*. *Biochemistry 28*, 4650–4655.

16. Shim, H., and Raushel, F. M. (2000) Self-assembly of the binuclear metal center of phosphotriesterase. *Biochemistry 39*, 7357–7364.

17. Stahl, D. A., Key, R., Flesher, B., and Smit, J. (1992) The phylogeny of marine and freshwater caulobacters reflect their habitat. *J. Bacteriol. 174*, 2193–2198.

18. Xu, C., Hall, R., Cummings, J., and Raushel, F. M. (2006) Tight binding inhibitors of *N*-acyl amino sugar and amino acid deacetylases. *J. Am. Chem. Soc. 128*, 4244–4255.

19. Doi, E., Shibata, D., and Matoba, T. (1981) Modified colorimetric ninhydrin methods for peptidase assay. *Anal. Biochem. 118*, 173–184.

20. Otwinowski, Z., and Minor, A. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, in *Methods in Enzymology, Macromolecular Crystallography, Part A* (Carter, C. W., Jr., and Sweet, R. M., Eds.) Vol. 276, pp 307−326, Academic Press, New York.

21. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr. 30*, 1022–1025.

22. Murshudov, G. M., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.

23. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.

24. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283–291.

25. DeLano, W. L. (2002) The PyMOL Molecular Graphics System (http://www.pymol).

26. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem. 357*, 289–298.

27. Thoden, J. B., Marti-Arbona, R., Raushel, F. M., and Holden, H. M. (2003) High resolution X-ray structure of isoaspartyl dipeptidase from *Escherichia coli*. *Biochemistry 42*, 4874–4882.

28. Hall, R. S., Xiang, D. F., Xu, C., and Raushel, F. M. (2007) *N*-acetyl-D-glucosamine-6-phosphate: substrate activation via a single divalent metal ion. *Biochemistry 46*, 7942–7952.