

## Functional Annotation of Two New Carboxypeptidases from the Amidohydrolase Superfamily of Enzymes<sup>†,‡</sup>

Dao Feng Xiang,<sup>§</sup> Chengfu Xu,<sup>§</sup> Desigan Kumaran,<sup>||</sup> Ann C. Brown,<sup>⊥</sup> J. Michael Sauder,<sup>@,#</sup> Stephen K. Burley,<sup>@,#</sup> Subramanyam Swaminathan,<sup>\*||</sup> and Frank M. Raushel<sup>\*§</sup>

<sup>§</sup>Department of Chemistry, P.O. Box 30012, Texas A&M University, College Station, Texas 77845,  
<sup>||</sup>Biology Department, Brookhaven National Laboratory, Upton, New York 11973, <sup>⊥</sup>Medgar Evers College,  
Brooklyn, New York 11225, and <sup>@</sup>SGX Pharmaceuticals, 10505 Roselle Street, San Diego, California 92121  
<sup>#</sup>Current affiliation: Eli Lilly & Co

Received March 16, 2009

**ABSTRACT:** Two proteins from the amidohydrolase superfamily of enzymes were cloned, expressed, and purified to homogeneity. The first protein, Cc0300, was from *Caulobacter crescentus* CB-15 (Cc0300), while the second one (Sgx9355e) was derived from an environmental DNA sequence originally isolated from the Sargasso Sea (gi|44371129). The catalytic functions and the substrate profiles for the two enzymes were determined with the aid of combinatorial dipeptide libraries. Both enzymes were shown to catalyze the hydrolysis of L-Xaa-L-Xaa dipeptides in which the amino acid at the N-terminus was relatively unimportant. These enzymes were specific for hydrophobic amino acids at the C-terminus. With Cc0300, substrates terminating in isoleucine, leucine, phenylalanine, tyrosine, valine, methionine, and tryptophan were hydrolyzed. The same specificity was observed with Sgx9355e, but this protein was also able to hydrolyze peptides terminating in threonine. Both enzymes were able to hydrolyze *N*-acetyl and *N*-formyl derivatives of the hydrophobic amino acids and tripeptides. The best substrates identified for Cc0300 were L-Ala-L-Leu with  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_{\text{m}}$  values of  $37 \text{ s}^{-1}$  and  $1.1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ , respectively, and *N*-formyl-L-Tyr with  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_{\text{m}}$  values of  $33 \text{ s}^{-1}$  and  $3.9 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ , respectively. The best substrate identified for Sgx9355e was L-Ala-L-Phe with  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_{\text{m}}$  values of  $0.41 \text{ s}^{-1}$  and  $5.8 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ . The three-dimensional structure of Sgx9355e was determined to a resolution of 2.33 Å with L-methionine bound in the active site. The  $\alpha$ -carboxylate of the methionine is ion-paired to His-237 and also hydrogen bonded to the backbone amide groups of Val-201 and Leu-202. The  $\alpha$ -amino group of the bound methionine interacts with Asp-328. The structural determinants for substrate recognition were identified and compared with other enzymes in this superfamily that hydrolyze dipeptides with different specificities.

Complete bacterial genomes are now being sequenced at an extraordinary rate (1). This influx of new information has provided an unprecedented view of the evolutionary relationships that exist among enzymes that recognize different substrates and catalyze diverse functional transformations. The pace for the deposition of new three-dimensional structures in the Protein Data Bank (PDB)<sup>1</sup> has also risen sharply in recent

years (2). However, it is apparent that a significant fraction of the newly sequenced genes have an unknown, uncertain, or incorrect catalytic functional assignment (3). This reality suggests that a considerable number of new metabolic transformations remain to be elucidated and that the annotation of function from sequence and structural information alone is a challenging problem. Nevertheless, the vast amount of structure and sequence information provides a unique opportunity for the elucidation of the mechanisms for the evolution of new catalytic functions from preexisting structural templates.

One approach to the assignment of function for enzymes with an unknown substrate profile is the utilization of enzyme superfamilies. Comprehensive amino acid sequence alignments and three-dimensional structural comparisons can provide valuable insights into the identification of the catalytic and structural motifs that dictate the substrate and reaction profiles (4). One such superfamily that has proven to be amenable to this approach

<sup>†</sup>This work was supported by the National Institutes of Health (Grants GM71790 and GM74945) and the Hackerman Advanced Research Program (010366-0034-2007).

<sup>‡</sup>The X-ray coordinates and structure factors for Sgx9355e have been deposited in the Protein Data Bank as entry 2qs8.

<sup>\*</sup>To whom correspondence should be addressed. F.M.R.: telephone, (979)-845-3373; fax, (979)-845-9452; e-mail, raushel@tamu.edu. S.S.: telephone, (631)-344-3187; fax, (631)-344-3407; e-mail, swami@bnl.gov.

<sup>1</sup>Abbreviations: PDB, Protein Data Bank; ATCC, American type Culture Collection; ICP-MS, inductively coupled plasma emission-mass spectrometry.

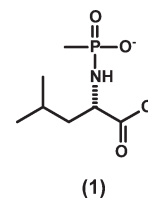
is the amidohydrolase superfamily (5, 6). This superfamily was first identified by Holm and Sander, who recognized the structural and functional relationships among adenosine deaminase, urease, and phosphotriesterase (7). All enzymes from this superfamily adopt a  $(\beta/\alpha)_8$ -barrel structural fold with their active sites located at the C-terminal face of the  $\beta$ -barrel (5). Within the active site is a mono- or binuclear metal center that functions to activate solvent water for nucleophilic attack and/or stabilization of the transition state. Most of the enzymes within this superfamily catalyze the hydrolysis of C–O, C–N, or P–O bonds (5). However, some members at the periphery of the superfamily catalyze decarboxylation, hydration, and isomerization reactions (8).

In this paper, we have interrogated the catalytic functions for two enzymes within the amidohydrolase superfamily. The first of these enzymes, Cc0300 from *Caulobacter crescentus* CB15, is currently annotated by NCBI as an Xaa-Pro dipeptidase. The second enzyme, derived from an environmental DNA sequence isolated from the Sargasso Sea and labeled here as Sgx9355e, is 39% identical in amino acid sequence with Cc0300. We have recently determined the catalytic functions for two other putative Xaa-Pro dipeptidases from *C. crescentus*, Cc2672 and Cc3125, that are ~40% identical in sequence to Cc0300 (9). These two enzymes do not catalyze the hydrolysis of Xaa-Pro dipeptides, but they do catalyze the hydrolysis of L-Xaa-L-Arg/Lys dipeptides; thus, the current functional annotations for these enzymes are incorrect. The X-ray structure of a catalytic homologue to Cc2672 and Cc3125 has unveiled the structural motif responsible for the recognition of the cationic side chains of arginine- and lysine-containing substrates in these enzymes. In Cc0300, this recognition motif is missing, suggesting that the substrate specificity for Cc0300 differs from that determined previously for Cc2672 and Cc3125. The substrate profiles for Cc0300 and Sgx9355e were determined by screening multiple dipeptide libraries containing nearly all possible combinations of L-Xaa-L-Xaa dipeptides made from the 20 common amino acids. The X-ray structure of the methionine-bound form of Sgx9355e has been determined to a resolution of 2.33 Å.

## MATERIALS AND METHODS

**Materials.** Genomic DNA from *C. crescentus* CB15 was purchased from 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 (College Station, TX). The pET-30a(+) expression vector was obtained from Novagen. T4 DNA ligase and various restriction enzymes were acquired from New England Biolabs. Platinum Pfx DNA polymerase was purchased from Invitrogen. The Wizard Plus SV Mini-Prep DNA purification kit was obtained from Promega. Chromatographic columns and resins were from GE Healthcare. Chelex-100 resin was purchased from Bio-Rad. ICP standards were purchased from Inorganic Ventures Inc. The tripeptides, L-Gly-L-Phe-L-Arg and L-Gly-L-Ala-L-Tyr, were purchased from Aroz Technologies LLC. NAD and formate dehydrogenase from *Candida boidinii* were purchased from Sigma. The *N*-methyl phosphonate derivative of L-leucine was synthesized according to the method described by Xu et al. (10). The structure of this compound is presented in Chart 1. All other buffers, purification reagents, and other chemicals used in this work were purchased from Sigma-Aldrich, unless otherwise stated.

Chart 1



**Synthesis of Dipeptides and Dipeptide Libraries.** Syntheses of the L,L-dipeptide libraries were conducted as reported previously (9). Mass spectrometric analysis (ESI, positive and negative mode) was employed to verify the components of each library. The individual dipeptides L-Ala-L-Leu, L-Ala-L-Phe, and L-Asp-L-Leu were prepared in a manner similar to the synthesis of the dipeptide libraries.

**Gene Cloning.** Standard molecular cloning techniques were employed throughout (11). A 1287 bp fragment containing the entire open reading frame of Cc0300 was amplified from the genomic DNA of *C. crescentus* CB15 using the primer pair 5' AGAACTTCCATATGCGTAAATT GATGGCGGG CGCCTGC 3' (forward) and 5' ACGGAATTCTTACTTGTA GACGACGCCGCC TTGATGAC 3' (reverse). The forward and reverse primers were designed to introduce an *Nde*I restriction site and an *Eco*RI restriction site, respectively. The PCR product, purified using a PCR cleanup system (Promega) and digested with *Nde*I and *Eco*RI, was ligated into the same sites of expression vector pET-30a(+). The cloned fragment was completely sequenced to verify the fidelity of the PCR amplification. On the basis of the reported sequence of gi|44371129 deduced from a DNA sample originally isolated from the Sargasso Sea, the New York SGX Research Center for Structural Genomics (NYSGXRC, www.nysgxrc.org) cloned and expressed Sgx9355e as a His-tagged protein following codon optimization and gene synthesis by Codon Devices, Inc. (Cambridge, MA). NYSGXRC will make the clone (9355e1BCt10p) available through the Protein Structure Initiative Material Repository (PSI-MR), accessible from the PSI Knowledgebase ([http://kb.psi-structuralgenomics.org/KB/psi\\_resources.html](http://kb.psi-structuralgenomics.org/KB/psi_resources.html)).

**Protein Purification.** The plasmid for Cc0300 was constructed for expression under control of the IPTG-inducible T7 RNA polymerase. BL21(DE3) Star cells (Novagen) were transformed with the plasmid containing the gene for Cc0300. A single colony was used to inoculate a 5 mL overnight culture of LB medium containing 50  $\mu$ g/mL kanamycin. The 5 mL overnight culture was then used to inoculate 1.0 L of LB medium containing 50  $\mu$ g/mL kanamycin. The cells were grown at 30 °C and induced by adding IPTG to a final concentration of 0.5 mM when an OD<sub>600</sub> of the culture reached 0.6. Zn(OAc)<sub>2</sub> (1.0 mM) was added to the culture, and the cells were harvested by centrifugation (6000 rpm for 15 min) 14 h after IPTG induction and stored at –80 °C. To purify Cc0300, 10 g of frozen cells was thawed and resuspended in 50 mL of 50 mM Hepes (pH 8.0). The cells were disrupted by sonication in an ice bath, and the soluble fraction was isolated by centrifugation. The supernatant solution was treated with 2% (w/v) protamine sulfate to precipitate nucleic acids, which were removed by centrifugation. The supernatant solution was treated with ground ammonium sulfate until 60% saturation. The pellet from the ammonium sulfate fractionation step was dissolved in a minimal volume of 50 mM Hepes (pH 8.0) and then passed through a 0.2  $\mu$ m pore filter prior to being loaded onto a pre-equilibrated Hi Load 26/60 Superdex 200 prep grade

gel filtration column (Amersham Biosciences). Fractions containing the target protein were pooled on the basis of SDS-PAGE analysis. The protein was further purified by being loaded onto a Resource Q anion exchange column (Amersham Biosciences) and eluted with a linear gradient of NaCl in 20 mM Hepes (pH 8.0). The appropriate fractions were pooled after analysis by SDS-PAGE. Protein concentrations were determined using the calculated extinction coefficient of  $31860 \text{ M}^{-1} \text{ cm}^{-1}$  at 280 nm (12).

For purification of Sgx9355e, *Escherichia coli* BL21(DE3) Star cells were transformed with the plasmid encoding the gene for Sgx9355e. One liter of LB medium was inoculated with a 5 mL overnight culture. The inoculated culture was grown with agitation at 30 °C to an  $\text{OD}_{600}$  of 0.6 and then supplemented with 1.0 mM  $\text{Zn}(\text{OAc})_2$  and induced with IPTG at a final concentration of 0.5 mM. The cells were grown at 30 °C for an additional 14 h and then harvested by centrifugation (6000 rpm for 15 min). The cell pellet was resuspended 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 (10000 rpm for 15 min). The clarified cell extract was applied to a 24 mL column of chelating Sepharose Fast Flow resin (Amersham Biosciences) charged with  $\text{Ni}^{2+}$  and pre-equilibrated with binding buffer. The column was washed thoroughly with 1000 mL of binding buffer until the absorbance of the flow-through at 280 nm did not change. The His-tagged protein, Sgx9355e, was eluted with a linear gradient of elution buffer [10 mM Hepes (pH 7.9), 0.25 M NaCl, and 0.5 M imidazole]. The protein obtained was further purified by application to a Resource Q anion exchange column after the NaCl was removed by dialysis. Sgx9355e was eluted with a linear gradient of NaCl in 20 mM Hepes (pH 8.0). The protein was more than 95% pure as judged by SDS-PAGE analysis.

**Amino Acid Sequence Verification.** N-Terminal amino acid sequence analysis of the purified Cc0300 was conducted by the Protein Chemistry Laboratory at Texas A&M University. The amino acid sequence obtained for the first eight amino acid residues of Cc0300 was QATFVQAG, which indicates that either the first 23 amino acid residues of the protein were post-translationally removed via proteolysis or the protein was initially expressed from the 24th amino acid residue of the reported gene sequence.

**Metal Analysis.** Apo-Sgx9355e was made by incubating the enzyme with 2.0 mM 1,10-phenanthroline for 24 h at 4 °C. The chelator was removed by loading the protein/chelator mixture onto a PD-10 column (GE Healthcare) and eluting with 50 mM metal-free Hepes buffer (pH 8.0). The metal-free Hepes buffer was made by passing the sample through a Chelex-100 column to remove possible metal salts present in the buffer. All glassware used for making metal-free buffer was soaked with 10 mM EDTA overnight and then rinsed thoroughly with deionized water. The apo-Sgx9355e was reconstituted with 2 equiv of  $\text{ZnCl}_2$  in 50 mM metal-free Hepes (pH 8.0) in the presence of 10 mM potassium bicarbonate. The metal content of as-purified and reconstituted proteins was determined by inductively coupled plasma emission-mass spectrometry (ICP-MS) (13).

**Assay Methods for Peptidase Activity.** A modified colorimetric ninhydrin assay method was used to screen the dipeptidase activity of Cc0300 and Sgx9355e (14). The same method was used to measure the hydrolysis of tripeptides and the hydrolytic activity with *N*-acetyl- and *N*-formyl derivatives of L-amino acids. The absorbance at 507 nm was recorded with

a SPECTRAMax plate reader from Molecular Devices. Quantitative analysis of the liberated amino acids was conducted by the Protein Chemistry Laboratory at Texas A&M University.

**Screening of Dipeptide Libraries.** Cc0300 and Sgx9355e were tested for dipeptidase activity with the entire dipeptide library (361 compounds in all). Each dipeptide library 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. Screening of the dipeptide libraries was initiated by mixing a fixed concentration of the dipeptide library (containing each dipeptide at  $\sim 0.1$  mM) and various amounts of enzyme over a concentration range of 10–1000 nM. All reactions were conducted in 50 mM Hepes (pH 8.0) in 96-well microplates. The reaction mixtures were incubated at 30 °C for a fixed time period and then the reactions quenched by adding the Cd-ninhydrin reagent. The quenched reaction mixtures were incubated for 5 min at 80 °C for color development. Control reactions without enzyme or without the dipeptide library were conducted simultaneously. The relative rates were determined by a fit of the data to eq 1

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

where  $y$  is the change in absorbance at 507 nm,  $x$  is the concentration of the enzyme, and  $b$  is the relative rate constant. In these measurements, the enzyme concentration was varied but the reaction time was fixed.

**Amino Acid Analysis.** The specific amino acids that were released from the dipeptide libraries were quantitatively identified by HPLC. The substrate specificity for either Cc0300 or Sgx9355e was determined by assay with selected dipeptide libraries. The hydrolysis reactions were conducted in 25 mM ammonium bicarbonate buffer (pH 8.0). For each dipeptide library assay, a mixture of 19 dipeptides at  $\sim 0.1$  mM was subjected to hydrolysis with variable concentrations of enzyme ranging from 1 to 1000 nM. The identity of the amino acids released from the hydrolysis reactions was determined by amino acid analysis. Two dipeptide libraries (L-Ala-L-Xaa and L-Asp-L-Xaa) were assayed with Cc0300, and two dipeptide libraries (L-Ala-L-Xaa and L-Gln-L-Xaa) were used in the assays with Sgx9355e. The enzyme and dipeptide library were incubated at 30 °C for 14 h in a 96-well microplate, and then the enzyme was removed by being passed through a filter. Quantitative amino acid analysis was conducted after the samples were dried under reduced pressure and reconstituted with water. The relative reaction rates were determined by fitting the change in amino acid concentration as a function of enzyme concentration to eq 1.

**Kinetic Measurements for Individual Substrates.** Selected dipeptides (L-Ala-L-Leu, L-Ala-L-Phe, and L-Asp-L-Leu), *N*-formyl-L-amino acids (*N*-formyl-L-Leu, *N*-formyl-L-Ile, *N*-formyl-L-Phe, *N*-formyl-L-Tyr, *N*-formyl-L-Met, and *N*-formyl-L-Val), *N*-acetyl-L-amino acids (*N*-acetyl-L-Leu, *N*-acetyl-L-Ile, *N*-acetyl-L-Phe, *N*-acetyl-L-Tyr, *N*-acetyl-L-Met, and *N*-acetyl-L-Val), and the tripeptide, L-Gly-L-Ala-L-Tyr, were used as substrates for measurement of the kinetic parameters for Cc0300 and Sgx9355e. The kinetic assays were conducted in 50 mM Hepes (pH 8.0), and the products were quantified using the ninhydrin-based assay. The kinetic constants,  $k_{\text{cat}}$ ,  $K_m$ , and  $k_{\text{cat}}/K_m$  were determined by fitting the initial velocity data to eq 2

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

where  $v$  is the initial velocity,  $k_{\text{cat}}$  is the turnover number,  $E_t$  is the enzyme concentration,  $A$  is the substrate concentration, and  $K_m$  is the Michaelis constant.

**Inhibition by a Tetrahedral Analogue.** The inhibitory properties of the *N*-methyl phosphonate derivative of L-leucine (**1**) with Cc0300 were determined using *N*-formyl-L-leucine as the substrate. The assays were conducted by monitoring the rate of formate production using a formate dehydrogenase coupling system in the presence of NAD. The reaction was followed spectrophotometrically at 340 nm, using a SPECTRAMax (Molecular Devices) plate reader. All assays were conducted in a 96-well UV plate in a volume of 250  $\mu\text{L}$ . The buffer used in all assays was 50 mM Hepes (pH 8.0). The concentrations of *N*-formyl-L-leucine and the *N*-methyl phosphonate derivative of L-leucine were varied in the ranges of 50–200 and 1–100  $\mu\text{M}$ , respectively. For data analysis, the initial rates for substrate hydrolysis in the presence of the inhibitor at different concentrations of the substrate were fitted to eq 3 for competitive inhibition.

$$v/E_t = k_{\text{cat}}A/[K_m(1 + I/K_i) + A] \quad (3)$$

where  $v$  is the initial velocity,  $A$  is the substrate concentration,  $I$  is the inhibitor concentration,  $k_{\text{cat}}$  is the turnover number,  $E_t$  is the enzyme concentration,  $K_m$  is the Michaelis constant for the substrate, and  $K_i$  is the competitive inhibition constant.

**Crystallization and Data Collection for Sgx9355e.** Crystals of the selenomethionine form of the Sgx9355e protein were grown using the sitting drop vapor diffusion method at 20 °C. Equal volumes of protein (10–15 mg/mL) and a solution containing 30% MPD, 0.1 M Hepes, and 0.2 M NaCl (pH 7.0) were mixed and equilibrated against a 800  $\mu\text{L}$  reservoir containing the same precipitant solution. A single crystal was flash-frozen via addition of 20% glycerol to the mother liquor before the crystal was loop mounted. SAD data extending to 2.33 Å were collected at the selenium absorption edge ( $\lambda = 0.979$  Å) from a single crystal at liquid nitrogen temperature using the ADSC QUANTUM 315 detector at the National Synchrotron Light Source (NSLS) using beamline X29. X-ray diffraction images were processed with HKL2000, and the data collection statistics are given in Table 1 (15). The crystals belong to monoclinic space group *I4* with the following unit cell dimensions:  $a = b = 144.9$  Å, and  $c = 101.0$  Å (Table 1). Assuming two molecules of 45781 Da per asymmetric unit, the Matthews coefficient is 2.0 Å<sup>3</sup>/Da, corresponding to an estimated solvent content of 47% by volume of the unit cell.

**Structure Determination and Refinement of Sgx9355e.** The structure was determined by the single-wavelength anomalous dispersion method. Twenty-two of the possible twenty-four selenium positions in the asymmetric unit were located using SHELXD (16). These selenium positions were refined and phases calculated with SHARP followed by solvent flattening (17, 18). The resulting experimental electron density map was of excellent quality and revealed practically all of the secondary structural elements comprising the dimer. The automated model building program ARP/wARP was then used to build the model into the electron density map (19). This process built ~87% of each protomeric polypeptide chain, and the remainder of the model was completed using O (20, 21). The structure was refined with Crystallography and NMR system (CNS) (22). Cycles of manual rebuilding and structure refinement were continued until the convergence of  $R$  and  $R_{\text{free}}$ . Phasing and refinement statistics are

Table 1: Crystal Data, Phasing, and Refinement Statistics for Sgx9355e

unit cell dimensions	$a = b = 144.9$ Å, $c = 101.0$ Å, $\alpha = \beta = \gamma = 90^\circ$
space group	<i>I4</i>
data set	peak
wavelength (Å)	0.9791
resolution range (Å)	50.0–2.33 (2.41–2.33)
no. of unique reflections	44488
completeness (%)	99.8 (98.4)
redundancy	19 (12.5)
$R_{\text{merge}}^a$	0.10 (0.29)
phasing power <sup>b</sup> (ano)	2.95
figure of merit <sup>b</sup> /SAD (centric/acentric)	0.18/0.55
after solvent flattening	0.95
Refinement Statistics	
$R^c/R_{\text{free}}$	0.21/0.23
resolution range (Å)	50.0–2.33
no. of atoms	
proteins	6165
waters	285
heterogeneous atoms	20
root-mean-square deviation from ideality	
bonds (Å)	0.007
angles (deg)	1.3
<sup>a</sup> $R_{\text{merge}} = \sum_i ( I_i - \langle I_i \rangle ) / \sum_i I_i$ , where $\langle I_i \rangle$ is the average intensity over symmetry equivalents. <sup>b</sup> As defined in SHARP. <sup>c</sup> R-factor = $\sum  F_{\text{obs}} - F_{\text{calc}}  / \sum  F_{\text{obs}} $ .	

provided in Table 1. The final model was validated with PROCHECK (23). Atomic coordinates and structure factors have been deposited in the Protein Data Bank (entry 2qs8).

## RESULTS

**Purification of Cc0300 and Sgx9355e.** Cc0300 from *C. crecentus* was expressed in reasonable quantities after the plasmid containing the gene for this enzyme was transformed in BL21(DE3) Star cells and allowed to grow in an LB medium supplemented with 1.0 mM Zn(OAc)<sub>2</sub>. The enzyme was purified, and SDS–PAGE analysis showed a single protein band which was slightly smaller than the theoretical molecular mass of 44 kDa calculated from the reported gene sequence. N-Terminal protein sequence analysis of purified Cc0300 indicated that the first 23 amino acids were missing. This result indicates that either these residues were lost through proteolysis or this gene has an alternate initiation site. The as-purified protein was found to contain an average of 1.6 equiv of Zn<sup>2+</sup> per subunit. Sgx9355e was purified as a His-tagged protein and was found to contain 0.2 equiv of Zn per subunit. However, the Zn content could be increased to an average of 1.2 equiv per subunit via incubation of apo-Sgx9355e with 2 equiv of ZnCl<sub>2</sub> in the presence of 10 mM bicarbonate overnight.

**Crystal Structure of Sgx9355e.** The structure of Sgx9355e is shown in Figures 1 and 2. Within the crystal, the protein occurs as a homo-octamer with a buried surface area of approximately 3759 Å<sup>2</sup> (~19% of the total surface area) per protein–protein interaction (Figure 1). Each subunit consists of two domains, a small predominantly  $\beta$ -stranded domain (residues 7–64 and 379–412) and a large TIM barrel domain (residues 65–370) as presented in Figure 2. The active site is located at the bottom of the cone-shaped structure of each subunit and is open to solvent. The residues that are expected to bind the two divalent cations are His-69, His-71, His-235, His-255, Lys-194, and Asp-328 based upon the known homology with other members of the

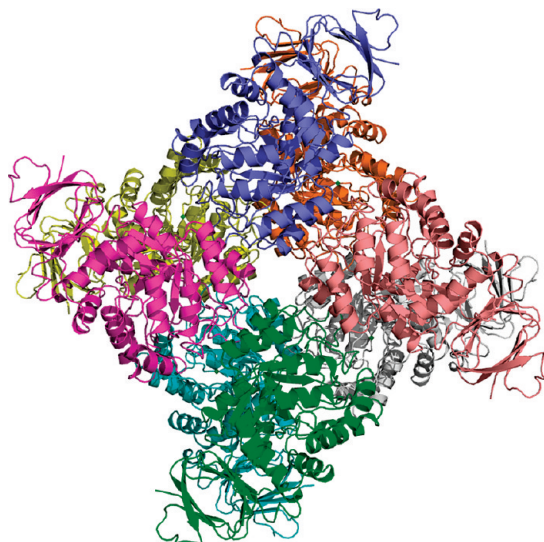


FIGURE 1: Structure of Sgx9355e. Ribbon representation of the Sgx9355e homo-octamer with different subunits colored green, yellow, blue, magenta, cyan, gold, brown, and gray.

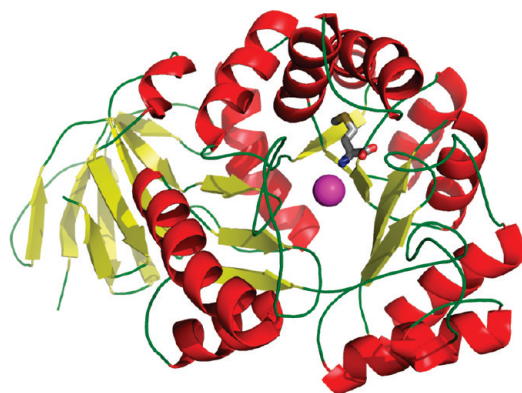


FIGURE 2: Ribbon representation of a single subunit of Sgx9355e with the bound methionine shown as a stick figure. The magnesium ion is shown as a magenta sphere.

amidohydrolase superfamily (5). The relative orientations of these presumptive metal binding residues are presented in Figure 3. However, the enzyme crystallized without zinc, and a single magnesium ion was found ligated to the  $\alpha$ -metal site. The magnesium is coordinated to His-69, His-71, and Asp-328. Interestingly, methionine, from the crystallization medium, is bound in the active site close to the position normally occupied by the more solvent-exposed  $\beta$ -metal. The  $\alpha$ -amino group of the methionine interacts with the side chain carboxylate of Asp-328. The  $\alpha$ -carboxylate of the bound methionine is ion paired with His-237 and hydrogen bonded with the backbone amide groups of Val-201 and Ile-202 as illustrated in Figure 4. The side chain of methionine is found within a pocket formed by the side chains of Thr-277 and Ile-309.

**Screening of Dipeptide Libraries.** To determine the substrate profile of Cc0300 and Sgx9355e, a total of 19 dipeptide libraries were tested for catalytic activity using the Cd-ninhydrin-based assays. The relative preferences for the N-terminal amino acid residue within these dipeptide libraries are depicted in Figure 5. Cc0300 and Sgx9355e exhibit broad catalytic activity for the hydrolysis of these peptide libraries with different N-termini, but the relative rates of hydrolysis vary from library to library. These two enzymes exhibit relatively high reaction

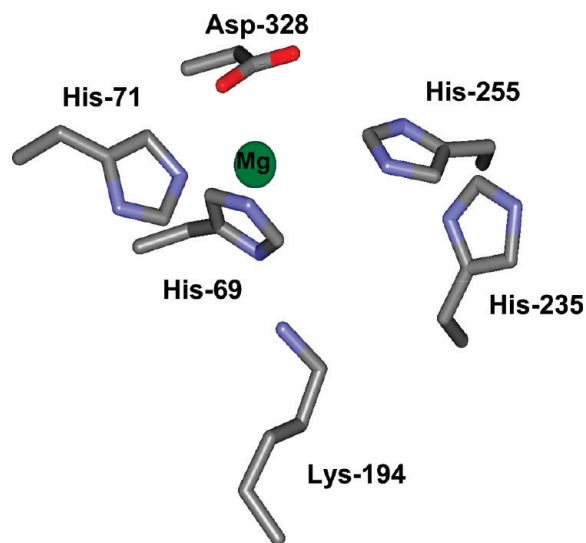


FIGURE 3: Active site geometry of residues expected to form the binuclear metal center of Sgx9355e.

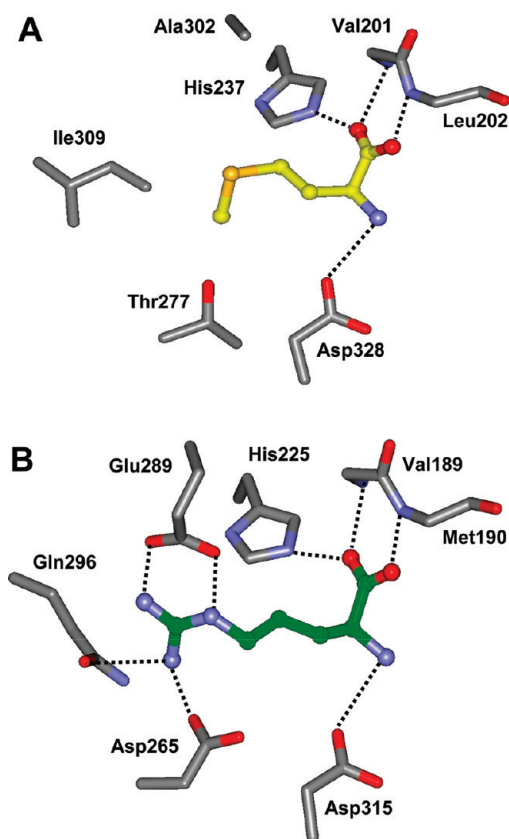


FIGURE 4: (A) Structure of Sgx9355e with L-methionine bound within the active site. (B) Structure of Sgx9359b with L-arginine bound within the active site.

rates with L-Trp-L-Xaa and low reaction rates with L-Ile-L-Xaa and L-Val-L-Xaa. Cc0300 exhibits a higher reaction rate with L-Tyr-L-Xaa and L-His-L-Xaa than Sgx9355e does. Conversely, Sgx9355e exhibits a higher turnover rate with L-Lys-L-Xaa and L-Arg-L-Xaa than Cc0300 does.

**Specificity of Cc0300 and Sgx9355e.** Quantitative amino acid analysis (AAA) was employed to determine the relative rates of hydrolysis for specific dipeptides within selected dipeptide libraries. Cc0300 was assayed with the L-Ala-L-Xaa and L-Asp-L-Xaa dipeptide libraries. The amino acid products, detected from

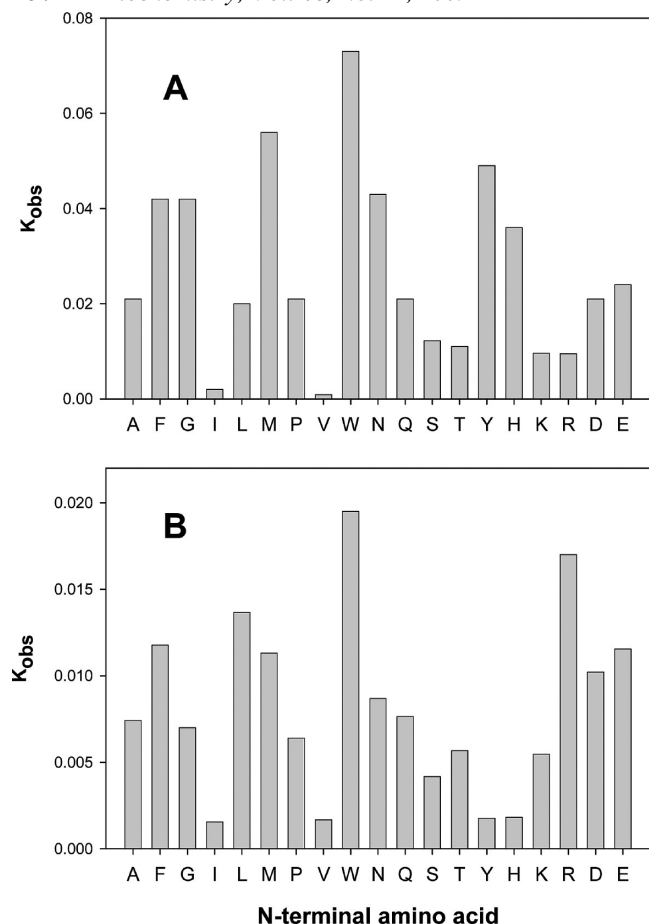


FIGURE 5: Relative substrate activities for hydrolysis of 19 dipeptide libraries of the type L-Xaa-L-Xaa in which a single amino acid was at the N-terminus (defined on the  $x$ -axis) and 19 different amino acids were at the C-terminus (not including L-cysteine). The results for Cc0300 are presented in panel A and those for Sgx9355e in panel B. Additional details are found in the text.

the reaction mixture of Cc0300 with the L-Ala-L-Xaa library, were limited to alanine, leucine, isoleucine, phenylalanine, tryptophan, methionine, tyrosine, and valine. The amino acid products detected from the hydrolysis reaction using the L-Asp-L-Xaa library as a substrate were aspartate, leucine, isoleucine, phenylalanine, tryptophan, methionine, tyrosine, and valine. Since alanine and aspartate were detected only when libraries containing these amino acids were at the N-terminus, it can be concluded that the C-terminal specificity of Cc0300 with dipeptides is limited to the hydrophobic amino acid residues leucine, isoleucine, phenylalanine, tryptophan, methionine, tyrosine, and valine. Two libraries, L-Ala-L-Xaa and L-Gln-L-Xaa, were selected for quantitative analysis with Sgx9355e for specificity at the C-terminus. Amino acid analysis showed that the C-terminal specificity of Sgx9355e with these dipeptide substrates is very similar to that found for Cc0300 except that Sgx9355e can hydrolyze dipeptides with threonine at the C-terminus. The relative reaction rates for Cc0300 and Sgx9355e are presented in Figure 6. The relative rates of hydrolysis for the other dipeptides in these libraries are less than 5% for both Cc0300 and Sgx9355e.

**Hydrolysis of *N*-Substituted Amino Acids and Tripeptides.** Cc0300 and Sgx9355e were screened for catalytic activity with *N*-acetyl-L-Xaa and *N*-formyl-L-Xaa derivatives. These two enzymes exhibited very similar catalytic specificities with these two types of substituted amino acid derivatives. The hydrolyzed

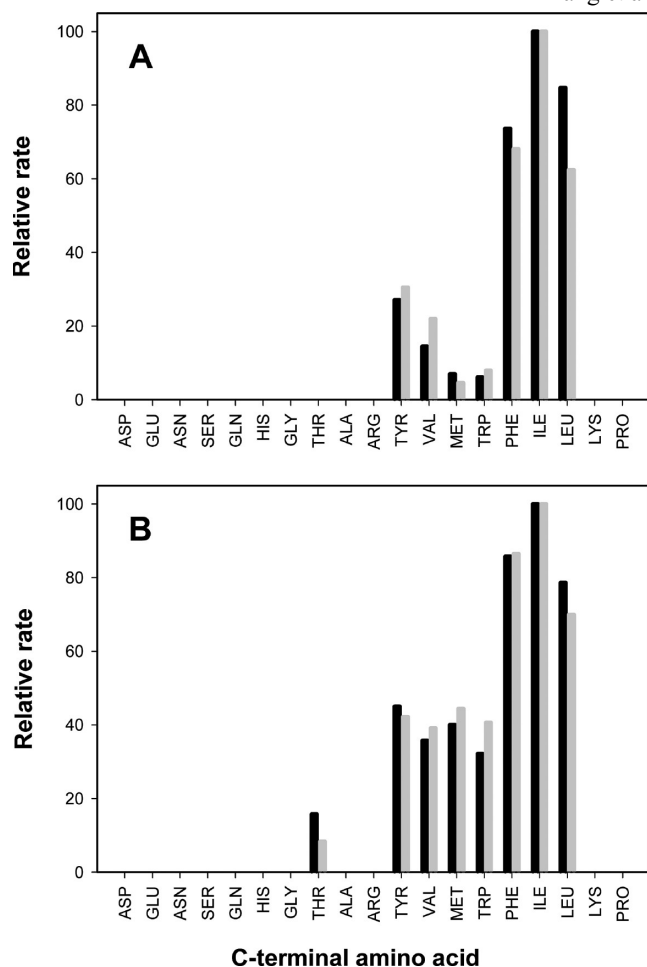


FIGURE 6: Relative rates of hydrolysis for the hydrolysis of selected dipeptide libraries by Cc0300 and Sgx9355e: (A) Cc0300 with L-Ala-L-Xaa (black) and L-Asp-L-Xaa (gray) and (B) Sgx9355e with L-Ala-L-Xaa (black) and L-Gln-L-Xaa (gray). The relative rates were determined by a fit of the data to eq 1. Additional details are found in the text.

substrates were limited to the *N*-acetyl and *N*-formyl derivatives of L-hydrophobic amino acids leucine, isoleucine, phenylalanine, tryptophan, methionine, tyrosine, and valine. Cc0300 and Sgx9355e were also tested with two tripeptides: L-Gly-L-Phe-L-Arg and L-Gly-L-Ala-L-Tyr. These two enzymes were unable to hydrolyze L-Gly-L-Phe-L-Arg but were able to hydrolyze the C-terminal L-tyrosine from L-Gly-L-Ala-L-Tyr.

**Kinetic Constants for Cc0300 and Sgx9355e.** The kinetic constants for Cc0300 were determined using the following substrates: L-Ala-L-Leu, L-Ala-L-Phe, L-Asp-L-Leu, *N*-formyl-L-Leu, *N*-formyl-L-Ile, *N*-formyl-L-Phe, *N*-formyl-L-Met, *N*-formyl-L-Tyr, *N*-formyl-L-Val, *N*-acetyl-L-Leu, *N*-acetyl-L-Ile, *N*-acetyl-L-Phe, *N*-acetyl-L-Met, *N*-acetyl-L-Tyr, and *N*-acetyl-L-Val. Also tested was the tripeptide L-Gly-L-Ala-L-Tyr. The kinetic data were fit to eq 2, and the values of  $k_{cat}$ ,  $K_m$ , and  $k_{cat}/K_m$  for the hydrolysis of the selected substrates are listed in Table 2. The turnover numbers ranged from  $1.7 \text{ s}^{-1}$  for the hydrolysis of *N*-acetyl-L-Met to  $37 \text{ s}^{-1}$  for the hydrolysis of L-Ala-L-Leu. The kinetic constants for the hydrolysis of selected compounds by Sgx9355e are listed in Table 2. In general, the kinetic constants,  $k_{cat}$  and  $k_{cat}/K_m$ , for Cc0300 are greater than those measured for Sgx9355e.

**Inhibition by *N*-Methyl Phosphonate Derivative (1).** Compound 1 was tested as an inhibitor of Cc0300. This

Table 2: Kinetic Parameters for Cc0300 and Sgx9355e

enzyme	substrate	$k_{cat}$ (s <sup>-1</sup> )	$K_m$ (mM)	$k_{cat}/K_m$ (M <sup>-1</sup> s <sup>-1</sup> )
Cc0300	L-Ala-L-Leu	37 ± 2	0.35 ± 0.04	(1.1 ± 0.1) × 10 <sup>5</sup>
	L-Ala-L-Phe	19 ± 1	0.66 ± 0.05	(2.9 ± 0.3) × 10 <sup>4</sup>
	L-Asp-L-Leu	29 ± 1	0.22 ± 0.04	(1.3 ± 0.2) × 10 <sup>5</sup>
	N-acetyl-L-Leu	2.5 ± 0.1	0.15 ± 0.03	(1.6 ± 0.3) × 10 <sup>4</sup>
	N-acetyl-L-Ile	2.6 ± 0.1	0.14 ± 0.02	(1.9 ± 0.2) × 10 <sup>4</sup>
	N-acetyl-L-Phe	5.3 ± 0.2	0.31 ± 0.03	(1.7 ± 0.2) × 10 <sup>4</sup>
	N-acetyl-L-Tyr	4.6 ± 0.2	0.17 ± 0.02	(2.7 ± 0.3) × 10 <sup>4</sup>
	N-acetyl-L-Val	2.0 ± 0.2	1.1 ± 0.1	(1.8 ± 0.2) × 10 <sup>4</sup>
	N-acetyl-L-Met	1.7 ± 0.1	0.66 ± 0.05	(2.6 ± 0.2) × 10 <sup>3</sup>
	N-formyl-L-Leu	39 ± 2	0.13 ± 0.01	(3.0 ± 0.3) × 10 <sup>5</sup>
	N-formyl-L-Ile	20 ± 1	0.07 ± 0.01	(2.9 ± 0.4) × 10 <sup>5</sup>
	N-formyl-L-Phe	52 ± 2	0.26 ± 0.03	(2.0 ± 0.2) × 10 <sup>5</sup>
	N-formyl-L-Tyr	33 ± 1	0.09 ± 0.01	(3.9 ± 0.4) × 10 <sup>5</sup>
	N-formyl-L-Val	18 ± 1	0.08 ± 0.01	(2.2 ± 0.3) × 10 <sup>5</sup>
N-formyl-L-Met	12 ± 1	0.41 ± 0.03	(2.9 ± 0.2) × 10 <sup>4</sup>	
Sgx9355e	L-Gly-L-Ala-L-Tyr	0.87 ± 0.05	0.39 ± 0.02	(2.2 ± 0.2) × 10 <sup>3</sup>
	L-Ala-L-Leu	0.34 ± 0.01	0.25 ± 0.03	(1.3 ± 0.2) × 10 <sup>3</sup>
	L-Ala-L-Phe	0.41 ± 0.01	0.07 ± 0.01	(5.8 ± 0.6) × 10 <sup>3</sup>
	N-acetyl-L-Leu	0.032 ± 0.01	0.11 ± 0.01	(2.9 ± 0.4) × 10 <sup>2</sup>
	N-formyl-L-Leu	0.072 ± 0.003	0.26 ± 0.04	(2.8 ± 0.4) × 10 <sup>2</sup>
L-Gly-L-Ala-L-Tyr	0.10 ± 0.01	0.30 ± 0.04	(3.3 ± 0.4) × 10 <sup>3</sup>	

compound was found to be a competitive inhibitor of this enzyme when *N*-formyl-L-leucine was used as the substrate. An inhibition constant of  $2.3 \pm 0.2 \mu\text{M}$  was obtained from a fit of the data (not shown) to eq 3.

## DISCUSSION

Elucidation of the substrate specificity for any enzyme given the amino acid sequence and/or three-dimensional structure alone is a difficult and demanding problem for most proteins. The rapid sequencing of whole bacterial genomes has provided an explosion of uncharacterized enzymes whose functions cannot be reliably annotated on the basis of a traceable homology to proteins of known function. Our approach to this problem is to address an entire enzyme superfamily simultaneously in an attempt to reveal the evolutionary lineages for the emergence of new catalytic functions and to provide a structural framework for deciphering substrate profiles from the amino acid sequences alone. More than 6000 proteins with unique amino acid sequences have been classified as members of the amidohydrolase superfamily (24). To date, more than 30 distinct reactions have been shown to be catalyzed by members of this superfamily using amino acid-, carbohydrate-, and nucleic acid-derived substrates (5, 7, 24).

The substrate specificity for two related enzymes was examined in this investigation. Cc0300 from *C. crecentus* and Sgx9355e from an unknown bacterium found in the Sargasso Sea are currently annotated as Xaa-Pro dipeptidase in the NCBI. Sequence comparisons have established that both proteins belong to the amidohydrolase superfamily. In this study, the two enzymes were purified to homogeneity and the enzyme from the Sargasso Sea was crystallized and its three-dimensional structure determined. Cc0300 is a zinc metalloprotein, and the active site is likely to be populated by a binuclear metal center, similar to that found previously for urease and phosphotriesterase (25–27). In this binuclear metal center, the two metal ions are coordinated by an aspartic acid and four histidine residues. The metal ions are also bridged by a carboxylated lysine residue and a hydroxide from solvent. The structure of Sgx9355e is consistent with this conclusion, but unfortunately, the metal ions were lost during purification and/or crystallization.

The substrate specificities for Cc0300 and Sgx9355e were determined using small libraries of potential dipeptide substrates. Contrary to initial expectations, neither of these enzymes was able to hydrolyze any peptide that contained proline at the C-terminus. However, both enzymes were able to catalyze the hydrolysis of dipeptides that terminated in a hydrophobic amino acid but were not specific for the amino acid at the N-terminus. *N*-Acetyl and *N*-formyl derivatives of L-hydrophobic amino acids were hydrolyzed by these two enzymes. Limited experiments with tripeptides demonstrated that hydrophobic amino acids from the C-terminus can be hydrolyzed, and thus, these two enzymes are more accurately classified as carboxypeptidases with a requirement for a hydrophobic amino acid at the C-terminus.

The X-ray structure of Sgx9355e was determined with L-methionine bound in the active site. The positioning of methionine in the active site has identified those amino acid residues that are responsible for the structural determinants of the substrate specificity for this enzyme. In this complex, L-methionine is positioned as the product derived from the hydrolysis of the C-terminal end of an oligopeptide. The  $\alpha$ -carboxylate of methionine is ion paired with His-237, a residue that originates from the loop that follows  $\beta$ -strand 5. The  $\alpha$ -carboxylate of L-methionine is also hydrogen bonded to the backbone amide groups of Val-201 and Leu-202. The  $\alpha$ -amino group of the bound L-methionine interacts with Asp-328. This residue is found at the end of  $\beta$ -strand 8 and is also expected to ligate the divalent metal ion in the  $M_\alpha$  position. The positioning of the  $\alpha$ -amino group of the C-terminally derived amino acid product with Asp-328 is consistent with this residue functioning as a general acid/base catalyst in the transfer of the proton from the bridging hydroxide to the leaving group amine (28). The side chain of the L-methionine is found in a small pocket that is formed from Thr-277, Ala-302, Val-305, and Ile-309. Thr-277 is found in a very short loop that immediately follows  $\beta$ -strand 7, whereas Ala-302, Val-305, and Ile-309 are found in a pair of helices that follow this loop. All of these residues, with the exception of Ile-309, are conserved in the amino acid sequence of Cc0300. In Cc0300, the isoleucine is substituted with a methionine.

We have previously determined the substrate profile for two other peptidases from *C. crecentus*, Cc2672 and Cc3125 (9). Cc2672 is specific for the hydrolysis of oligopeptides that terminate in either L-lysine or L-arginine. The specificity of Cc3125 is more restrictive in the sense that only dipeptides (but not longer peptides) that terminate in L-lysine or L-arginine represent productive substrates. In addition to these two proteins, we have determined the structure and substrate specificity of Sgx9359b from the Sargasso Sea (gi|44368820). This enzyme is specific for the hydrolysis of oligopeptides that terminate in L-arginine. The mode of binding of L-arginine to the active site of Sgx9359b is presented in Figure 4B. In this structure, the  $\alpha$ -carboxylate is ion paired to His-225 and hydrogen bonded to the backbone amide groups from Val-189 and Met-190. The  $\alpha$ -amino group interacts with the  $\alpha$ -carboxylate of Asp-315. These residues are homologous to those residues found to interact with L-methionine in the structure of Sgx9355e (Figure 4A). As expected, the structural determinants for the recognition of the guanidino group of the bound L-arginine in Sgx9359b are different. The guanidino group is hydrogen bonded with the side chain amide group of Gln-296 and ion paired with the side chain carboxylates of Glu-289 and Asp-265. These residues occupy positions similar to those of three amino acids that define the substrate recognition in Sgx9355e (Ile-309, Ala-302, and Thr-277).

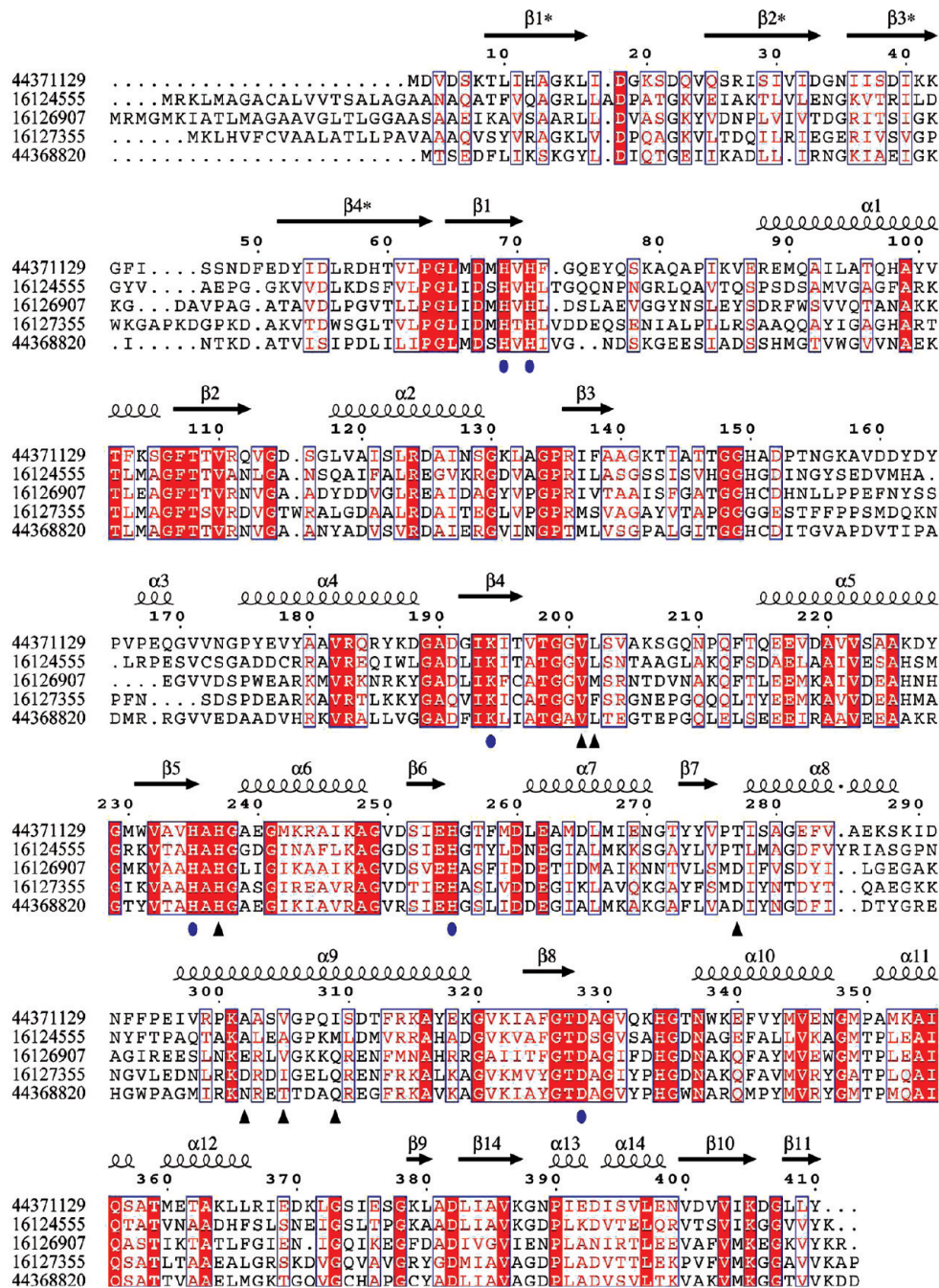


FIGURE 7: Amino acid sequence comparison for Sgx9355e (gi|44371129), Cc0300 (gi|16124555), Cc2672 (gi|16126907), Cc3125 (gi|16127355), and Sgx9359b (gi|44368820). The predicted metal ligands for the two divalent metal ions are indicated with ovals. Residues in Sgx9355e, and the equivalent residues in Cc0300, that interact with the bound L-methionine are indicated with triangles. Residues in Sgx9359b, and the equivalent residues in Cc2672 and Cc3125, that interact with the bound L-arginine are at equivalent positions. Other fully conserved residues in all five proteins are highlighted in red. The N-terminal residues that are removed via proteolysis during the expression of Cc0300, Cc2672, and Cc3125 are ANA/Q, ASA/A, and AAA/Q, respectively. This figure was produced with ESPrnt (29).

A sequence alignment of Sgx9355e, Cc0300, Cc2672, Cc3125, and Sgx9359b is presented in Figure 7. In this alignment, there are a total of 71 residues that are fully conserved among the five proteins. All six of the amino acids expected to bind to the two divalent cations in the active site are conserved, as is the histidine that ion pairs with the C-terminal carboxylate of peptide substrates. The three enzymes that recognize positively charged amino acids at the C-terminus of potential substrates (Cc2672, Cc3125, and Sgx9359b) have either anionic or highly polar residues at the binding site, indicated by black triangles in Figure 7. Conversely, the two enzymes that recognize hydrophobic side chains at the C-terminus of potential substrates

(Sgx9355e and Cc0300) have less polar amino acids at equivalent positions.

The discovery of function for Cc0300 and Sgx9355e, coupled with the structure elucidation of Sgx9355e with a bound product, permits annotation of other proteins in the amidohydrolase superfamily with greater confidence. A search of the NCBI database of completely sequenced bacterial genomes finds 29 other protein sequences that are now predicted by analogy to hydrolyze short oligopeptides that terminate in hydrophobic amino acids. All of these proteins have a conserved histidine equivalent to His-237, threonine and alanine residues equivalent to Thr-277 and Ala-302, and an isoleucine, methionine, leucine,



Table 3: Proteins Predicted To Share Substrate Specificity with Cc0300 and Sgx9355e

Gi number	Organism	Locus tag	Identity of residues equivalent to those in Sgx9355e (Thr277, Ala302, Val305 and Ile309)
44371129	Environmental sequence	Sgx9355e	PTISAGE.....RPKAAASVGPQIS
16124555	<i>C. crescentus</i> CB15	CC_0300	PTLMAGD.....TAKALEAGPKML
219674575	<i>O. alexandrii</i> HTCC2633	OA2633_13760	PTVMAGE.....AAKARVVGPLML
91794386	<i>S. denitrificans</i> OS217	Sden_3036	PTLLAGA.....RAKAIIRVGADMQ
116626946	<i>S. usitatus</i> Ellin6076	Acid_7923	PTLMAAE.....AGKARQAMAALN
114561981	<i>S. frigidimarina</i> NCIMB 400	Sfri_0801	PTLLAGD.....KAKAIIRVGDMM
71280007	<i>C. psychrerythraea</i> 34H	CPS_1348	PTLMAGD.....SAKAIIRVGGDMI
221233241	<i>C. crescentus</i> NA1000	CCNA_00302	PTLMAGD.....TAKALEAGPKML
197106524	<i>P. zucchini</i> HLK1	PHZ_c3063	PTLMAGD.....TAKALQAGPLML
127513920	<i>S. loihica</i> PV-4	Shew_2992	PTLLAGD.....KQKAVRVGADMT
212634704	<i>S. piezotolerans</i> WP3	swp_1881	PTLLAGA.....KAKAIIRVGDML
209964193	<i>R. centenum</i> SW	RC1_0865	PTLLAGK.....KRKALTVGPMQ
114798392	<i>H. neptunium</i> ATCC 15444	HNE_0934	PTVLAGA.....RAKAAQVGPML
196186567	<i>B. sp.</i> BAL3	BBAL3_2700	PTLLAGD.....TAKALEAGPKML
148555139	<i>S. wittichii</i> RW1	Swit_2224	PTLLVQG.....AQKATAIAPMLQ
170728069	<i>S. woodyi</i> ATCC 51908	Swoo_3741	PTLLAGD.....KAKAERVGDMM
83857796	<i>C. atlanticus</i> HTCC2559	CA2559_13368	PTISAGK.....VPKALEIGPKIQ
56461490	<i>I. loihiensis</i> L2TR	IL2390	PTIMAGK.....RPKARAIGPQIQ
163753087	<i>K. algicida</i> OT-1	KAOT1_13042	PTITAGK.....VPKALAVGPKIQ
163786441	<i>F. bacterium</i> ALC-1	FBALC1_14687	PTISAGK.....VPKALDIDGPKIQ
120434642	<i>G. forsetii</i> KT0803	GFO_0275	PTITAGK.....VPKAREIGPKIQ
88711102	<i>F. bacterium</i> HTCC2170	FB2170_03290	PTITAGK.....AKKAAEIGPLIQ
124006459	<i>M. marina</i> ATCC 23134	M23134_02799	PTISAGE.....VPKALKIGSQVK
88803709	<i>P. irgensii</i> 23-P	PI23P_00415	PTITAGK.....VPKALAVGPKIQ
89889513	<i>F. bacterium</i> BBFL7	BBFL7_01328	PTITAGK.....VPKARTVGPQIQ
86141823	<i>L. blandensis</i> MED217	MED217_01790	PTLSAGK.....VPKALEIGPQLQ
86135565	<i>P. sp.</i> MED152	MED152_12694	PTITAGK.....VPKALAVGPKIQ
91215139	<i>P. torquis</i> ATCC 700755	P700755_12617	PTITAGK.....VPKAKAIGPKIQ
109900575	<i>P. atlantica</i> T6c	Pat1_4277	PTVMAGN.....RPKAAATIGPLIL
119773229	<i>S. amazonensis</i> SB2B	Sama_0087	PTLLAGE.....RPKAAAIIGPKIQ
196158836	<i>A. macleodii</i> 'Deep ecotype'	MADE_04041	PTILAGK.....RPKAAAIIGPLIQ

or valine residue equivalent to Ile-309 in the structure of Sgx9355e. A list of these sequences is presented in Table 3.

## ACKNOWLEDGMENT

We thank A. Mahmood and J. Levia for technical assistance.

## REFERENCES

- Wu, C. H., Apweiler, R., Bairoch, A., Natale, D. A., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Mazumader, R., O'Donova, C., Redaschi, N., and Suzek, B. (2006) The universal protein resource (UniProt): An expanding universe of protein information. *Nucleic Acids Res.* 34, D187–D191.
- Berman, H. M. (2008) The protein data bank: A historical perspective. *Acta Crystallogr.* A64, 88–95.
- Gerlt, J. A., and Babbitt, P. C. (2000) Can sequence determine function?. *Genome Biol.* 5, 1–10.
- Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E. V., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function. *Nature* 448, 775–781.
- Seibert, C. M., and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* 44, 6383–6391.
- Pieper, U., Chiang, R., Seffernick, J. J., Brown, S. D., Glasner, M. E., Kelly, L., Eswar, N., Sauder, J. M., Bonanno, J. B., Swaminathan, S., Burley, S. K., Zheng, X., Chance, M. R., Almo, S. C., Gerlt, J. A., Raushel, F. M., Jacobson, M. P., Babbitt, P. C., and Sali, A. (2009) Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamilies. *J. Struct. Funct. Genomics* 10, 107–125.
- Holm, L., and Sander, C. (1997) An evolutionary treasure: Unification of a broad set of amidohydrolase related to urease. *Proteins* 28, 72–82.
- Nguyen, T. H., Brown, S., Fedorov, A. A., Fedorov, E. V., Babbitt, P. C., Almo, S. C., and Raushel, F. M. (2008) At the periphery of the amidohydrolase superfamily: Bh0493 from *Bacillus halodurans* catalyzes the isomerization of D-galacturonate to D-tagaturonate. *Biochemistry* 47, 1194–1206.
- Xiang, D., Patskovsky, Y., Chengfu, X., Meyer, A. J., Sauder, J. M., Burley, S. K., Almo, S. C., and Raushel, F. M. (2009) Functional identification of incorrectly annotated prolidases from the amidohydrolase superfamily of enzymes. *Biochemistry* 48, 3730–3742.
- Xu, C., Hall, R. S., Cummings, J., and Raushel, F. M. (2006) Tight binding inhibitors of N-acyl amino sugar and N-acyl amino acid deacetylases. *J. Am. Chem. Soc.* 128, 4244–4245.
- Russell, P. J. (2005) *iGenetics: A molecular approach*, Pearson Education, San Francisco, ISBN 0-8053-4665-1.
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 4, 2411–2423.
- Newman, A. (1996) Elements of ICPMS. *Anal. Chem.* 68, 46A–51A.
- Doi, E., Shibata, D., and Matoba, T. (1981) Modified colorimetric ninhydrin methods for peptidase assay. *Anal. Biochem.* 118, 173–184.
- Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, M., and Usón, I. (2001) *International tables for macromolecular crystallography* (Arnold, E., and Rossmann, M., Eds.) Vol. F, pp 333–345, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003) Generation, representation and flow of phase information in structure determination: Recent developments in and around SHARP 2.0. *Acta Crystallogr.* D59, 2023–2030.
- De-La-Fortelle, E., and Bricogne, G. (1997) Maximum-likelihood heavy atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* 276, 472–493.
- Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6, 458–463.
- Jones, T. A. (1978) A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* 11, 268–272.
- Jones, T. A., Zou, J., Cowtan, S., and Kjeldgaard, M. (1991) Improved methods in building protein models in electron density map and the location of errors in these models. *Acta Crystallogr.* A47, 110–119.
- Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Somsomson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for

- macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921.
23. 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.
  24. Pegg, S. C., Brown, S. D., Ojha, S., Seffernick, J., Menn, E. C., Morris, J. H., Chang, P. J., Huang, C. C., Ferrin, T. E., and Babbitt, P. C. (2006) Leveraging enzyme structure-function relationships for functional inference and experimental design: The structure-function linkage database. *Biochemistry* **45**, 2545–2555.
  25. Aubert, S. A., Li, Y., and Raushel, F. M. (2004) Mechanism for the hydrolysis of organophosphates by the bacterial phosphotriesterases. *Biochemistry* **43**, 5707–5715.
  26. 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.
  27. Kim, J., Tsai, P. C., Chen, S. L., Himo, F., Almo, S. C., and Raushel, F. M. (2008) Structure of diethyl phosphate bound to the binuclear metal center of phosphotriesterases. *Biochemistry* **47**, 9497–9504.
  28. 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.
  29. Gouet, P., Courcelle, E., Stuart, D. I., and Metz, F. (1999) ESPript: Multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308.