Title

Structure-Guided Functional Characterization of DUF1460 Reveals a Highly Specific NlpC/P60 Amidase Family

Journal

Structure, 22(12)

ISSN

1359-0278

Authors

Xu, Qingping
Mengin-Lecreulx, Dominique
Patin, Delphine
et al.

Publication Date

2014-12-01

DOI

10.1016/j.str.2014.09.018

Peer reviewed

Published in final edited form as:
Structure. 2014 December 2; 22(12): 1799–1809. doi:10.1016/j.str.2014.09.018.

Structure-Guided Functional Characterization of DUF1460 Reveals a New, Highly Specific NlpC/P60 Amidase Family

Qingping Xu[1,2], **Dominique Mengin-Lecreulx**[3,4,*], **Delphine Patin**[3,4], **Joanna C. Grant**[1,5], **Hsiu-Ju Chiu**[1,2], **Lukasz Jaroszewski**[1,6,7], **Mark W. Knuth**[1,5], **Adam Godzik**[1,6,7], **Scott A. Lesley**[1,5,8], **Marc-André Elsliger**[1,8], **Ashley M. Deacon**[1,2], and **Ian A. Wilson**[1,8,*]

[1]Joint Center for Structural Genomics, http://www.jcsg.org

[2]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA

[3]Laboratoire des Enveloppes Bactériennes et Antibiotiques, Université Paris-Sud, IBBMC, UMR 8619, Orsay, F-91405, France

[4]Centre National de la Recherche Scientifique (CNRS), Orsay, F-91405, France

[5]Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, USA

[6]Center for Research in Biological Systems, University of California, San Diego, La Jolla, California 92093, USA

[7]Program on Bioinformatics and Systems Biology, Sanford-Burnham Institute for Medical Research, La Jolla, California 92037, USA

[8]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, USA

SUMMARY

GlcNAc-1,6-anhydro-MurNAc-tetrapeptide is a major peptidoglycan degradation intermediate and a cytotoxin. It is generated by lytic transglycosylases and further degraded and recycled by various enzymes. We have identified and characterized a novel, highly specific *N*-acetylmuramoyl-L-alanine amidase (AmiA) from *Bacteroides uniformis*, a member of the DUF1460 protein family,

*Correspondence: Dominique Mengin-Lecreulx (dominique.mengin-lecreulx@u-psud.fr); Ian A. Wilson (wilson@scripps.edu).

Accession numbers

The structure factors and atomic coordinates are deposited in the RCSB Protein Data Bank (http://www.rcsb.org) with PDB IDs 4h4j (apo AmiA), 4q68 (AmiA-GlcNAc), and 4q5k (AmiA-GlcNAc-anhMurNAc). The plasmid was deposited in PSI:Biology-Materials Repository (http://dnasu.asu.edu) with clone ID BuCD00332345.

Author Contributions

Conceived and designed the experiments: QX and DML. Performed the experiments: QX and DML. Analyzed the data: QX and DML. Contributed reagents/materials/analysis tools: DP, JG, HJC, LJ, and MWK. Wrote the paper: QX, DML, MAE, AMD, and IAW. AG, SAL, MAE, AMD supervised various stages of the JCSG structural genomics pipeline, and IAW is the PI of the JCSG.

that hydrolyzes GlcNAc-1,6-anhydro-MurNAc-peptide into disaccharide and stem peptide. The high-resolution apo-structure at 1.15 Å resolution shows that AmiA is related to NlpC/P60 γ-D-Glu-*meso*-diaminopimelic acid amidases and shares a common catalytic core and cysteine peptidase-like active site. AmiA has evolved structural adaptations that reconfigure the substrate recognition site. The preferred substrates for AmiA were predicted *in silico* based on structural and bioinformatics data, and were subsequently characterized experimentally. Further crystal structures of AmiA in complexes with GlcNAc-1,6-anhydro-MurNAc and GlcNAc have enabled us to elucidate substrate recognition and specificity. DUF1460 is highly conserved in structure and defines a new amidase family.

Keywords

N-acetylmuramoyl-L-alanine amidase; NlpC/P60 amidases; structure-based function prediction

Introduction

Peptidoglycan (PG) forms a protective layer around bacteria that is essential for its survival. PG is formed by linear glycan chains consisting of alternating *N*-acetylmuramic acid (MurNAc) and *N*-acetylglucosamine (GlcNAc) moieties that are cross-linked by short stem peptides. The stem peptide typically consists of the tetrapeptide L-Ala-γ-D-Glu-*meso*-A_2pm (or Lys)-D-Ala (A_2pm, diaminopimelic acid), which is attached to MurNAc *via* a D-lactyl group. A peptide cross-link is formed between A_2pm (or Lys) and D-Ala from adjacent glycan strands. PG is continuously degraded by cell-wall specific lytic enzymes (van Heijenoort, 2011), and the products are recycled (Park and Uehara, 2008). Lytic transglycosylases (LTs) play a major role in cell-wall degradation and produce the main degradation product GlcNAc-1,6-anhydro-MurNAc-tetrapeptide (abbreviated as GlcNAc-anhMurNAc-tetrapeptide hereafter, Fig 1A). Lytic transglycosylases catalyze cleavage of the glycosidic linkage between GlcNAc and MurNAc followed by a concomitant intramolecular transglycosylation reaction to form the 1,6-anhydro ring at the MurNAc residue of the product. Alternatively, lysozymes can hydrolyze the same linkage as LTs to generate GlcNAc-MurNAc-tetrapeptide, with MurNAc in a reduced form. GlcNAc-anhMurNAc-tetrapeptide, also known as tracheal cytotoxin (TCT), plays a significant role in bacterial pathogenesis, when it escapes from the cell (Cloud-Hansen et al., 2006).

In *Escherichia coli*, the GlcNAc-anhMurNAc-tetrapeptide is transported by the integral membrane transporter AmpG into the cytoplasm, where NagZ removes the *N*-acetylglucosamine moiety. The resulting product serves as the substrate for the amidase AmpD generating 1,6-anhydro-MurNAc and tetrapeptide. The tetrapeptide can be degraded further by the LdcA L,D-carboxypeptidase that removes the D-Ala. The MpaA zinc carboxypeptidase then removes A_2pm from for the resulting tripeptide (Maqbool et al., 2012) and epimerase YcjG converts L-Ala-D-Glu into L-Ala-L-Glu, which is subsequently degraded by peptidase PepD. Amidases AmpD and MpaA are both metal hydrolases requiring zinc for activity (Fig. 1B), but are not related in structure. Alternatively, the L-Ala-γ-D-Glu-*meso*-A_2pm tripeptide can reenter the PG biosynthetic pathway via the cell wall recycling protein, murein peptide ligase (Mpl). The main PG recycling intermediate,

GlcNAc-anhMurNAc-tetrapeptide, and the overall degradation pathway are believed to be similar in other bacteria. However, underlying enzymes for carrying out individual steps are often not conserved. For example, *B. subtilis* has no MpaA ortholog, but instead possesses a functionally equivalent enzyme YkfC (Fig. 1B), an NlpC/P60 papain-like cysteine peptidase that is not related to MpaA (Xu et al., 2010). Members of the NlpC/P60 superfamily, often functioning as cell-wall hydrolases, contain a prototypical papain-like catalytic core (Anantharaman and Aravind, 2003). We have previously determined several crystal structures of NlpC/P60 γ-D-Glu-A_2pm amidases (Xu et al., 2010; Xu et al., 2014; Xu et al., 2009) and other structurally related enzymes (Xu et al., 2011).

DUF1460 defines a large protein family (>400 members) of unknown function and its members are widely distributed across the bacterial kingdom, particularly in the phyla Bacteroides (~25%) and Proteobacteria (~68%). It is classified as a member of the Peptidase_CA clan (CL0125), a large collection of proteins evolutionarily related to the papain cysteine peptidase. Two proteobacteria DUF1460 proteins are implicated in plant pathogenesis: *psa9* in the pathogenicity island 2 of *Pantoea stewartii* subsp. stewartii DC283 (Correa et al., 2012), and *ORF2* in the exchangeable effector locus of *Pseudomonas syringae* pv. tomato DC3000 (Fouts et al., 2002). Mobile gene elements in *E. coli* DEC5E and *Klebsiella pneumoniae* CG43 also carry genes encoding DUF1460 proteins (Uniprot: C0J1I1_ECOLX and Q6U5S1_KLEPN), suggesting that horizontal gene transfer is likely involved in the evolution of the DUF1460 family. The crystal structures of two DUF1460 proteins were previously determined by the New York SGX Research Center for Structural Genomics; BF2036 from *Bacteroides fragilis* YCH46 (PDB ID 2p1g) and LPG0564 from *Legionella pneumophila* (PDB ID 2im9), but were not functionally characterized. Some DUF1460 proteins are annotated as xylanases in various databases, but with no supporting experimental evidence.

Structural information is of significant value in assessing protein function; however, accurate functional assignment based on structure data alone is often difficult. Here, we report structure-based functional evaluation of DUF1460. We first determined the high resolution crystal structure of *Bacteroides uniformis* AmiA, a DUF1460 family member encoded by a putative operon that is conserved in the Bacteroides genera (Fig. 1C), as a part of our structural genomics effort to characterize the secretome of the human gut microbiome. The structure reveals that AmiA is evolutionarily related to NlpC/P60 proteins. Combining bioinformatics analysis and ligand docking, we were able to predict the protein function and its substrate specificity, which were then confirmed by biochemical characterization and co-crystallization studies. Furthermore, we demonstrated that AmiA is a novel GlcNAc-anhMurNAc-peptide-specific *N*-acetylmuramoyl-L-alanine amidase. Structure-sequence analysis of the DUF1460 family members suggests a common amidase function.

Results

Structure determination of apo AmiA and model quality

AmiA (262 aa) from *B. uniformis* is a secreted protein with a predicted N-terminal signal peptide (residues 1–23). The apo AmiA structure was determined using the semi-automated,

high-throughput pipeline of the Joint Center for Structural Genomics (Lesley et al., 2002). The selenomethionine derivative of the mature protein (residues 24–262) was expressed in *E. coli* with an N-terminal, TEV-cleavable, His-tag and purified by metal affinity chromatography. The purification tag was removed prior to crystallization trials. Crystals were harvested and screened for diffraction to identify the best crystals for structure determination.

The crystal structure of apo AmiA was determined in orthorhombic space group $P2_12_12_1$ using the SAD method, and was refined using data to 1.15 Å resolution with an R_{cryst} of 13.1 % and an R_{free} of 16.4 %. The asymmetric unit (asu) contains one monomer and 430 water molecules. Backbone conformations for all residues are within allowed region of the Ramachandran plot, and 98.2 % in the most favorable regions. The electron density is well-defined for most residues except for a short disordered loop (residues 80–83), which was not included in the final model. The high-resolution maps also allowed modeling of multiple conformations for 22.5 % of the residues. The quality of the final model (PDB ID 4h4j) compares favorably to other structures with similar resolutions, with an overall MolProbity score (Chen et al., 2010) that ranks in the 93th percentile. A summary of data collection, processing and refinement statistics is provided in Table 1.

AmiA is evolutionarily related to NlpC/P60 proteins

Analytical size-exclusion chromatography indicates that AmiA is a monomer in solution (data not shown), which is consistent with the lack of extensive inter-molecular packing interfaces in the crystal lattice. The AmiA monomer has an α/β/α three-layered sandwich fold. The catalytic core consists of a central five-stranded β-sheet and four α-helices with a topology of α1-α2-α3-β1-β2-β3-β4-α4-β5, which most closely resembles that of the NlpC/P60 γ-D-Glu-A_2pm amidase (Fig. 2A). For example, AmiA can be superimposed onto the NlpC/P60 domain of *Clostridium difficile* CwlT (PDB ID 4hpe), a bifunctional cell-wall hydrolase that we recently determined (Xu et al., 2014), with an RMSD of 2.8 A and sequence identity of 19% for 119 equivalent Cα atoms. Compared to a prototypical NlpC/P60 domain such as CwlT, AmiA contains three insertions: insert 1 between α1 and α2 (residues 44–60), insert 2 between α2 and α3 (residues 84–111), and insert 3 between α3 and β1 (residues 124–193) (Fig. 2A). Insert 3 is the longest (~70 aa) and is stabilized by extensive interactions with the core domain. All three inserts converge on one side of AmiA and contribute to the formation of a large pocket aside the catalytic cysteine (Cys63).

The catalytic triad Cys63/His212/His227 and a nearby Tyr41, which presumably functions to stabilize the reaction intermediate (Xu et al., 2009), are identical to those of CwlT (Fig. 2B) and other NlpC/P60 enzymes (Xu et al., 2010; Xu et al., 2014; Xu et al., 2009). Two discrete conformations of the active site residues are observed in the electron density (Fig. 2B). The catalytic Cys63 is located on the N-terminus of α2, which is a common characteristic of the papain superfamily of enzymes. A feature of NlpC/P60 proteins is the frequent substitution of the third polar residue of the catalytic triad by a histidine, rather than asparagine or glutamine typical of other papain-related peptidases. AmiA also shares this feature (His227). Thus, AmiA appears to share the same ancestral origin as NlpC/P60 γ-D-Glu-A_2pm amidases.

Computational prediction of AmiA function

Based on the above analysis and on its homology to NlpC/P60 proteins, we inferred that AmiA might be a cysteine hydrolase, but its specific substrate was unknown. To find potential substrates, we examined the genetic context of DUF1460 family members since functionally associated genes are often clustered together as operons in bacteria. In Bacteroides genomes, DUF1460 genes are often proximal to genes encoding an oligopeptide transporter and another α/β hydrolase (Fig. 1C). More intriguingly in *Mycobacterium sp.* and *Fibrobacter succinogenes*, the DUF1460 domain is fused to either a lytic transglycosylase (Uniprot: Q1B6F7_MYCSS) or a D-alanyl-D-alanine carboxypeptidase (Uniprot: C9RRN1_FIBSS), suggesting that DUF1460 is likely to cooperate with these two enzyme activities at the substrate level. These clues suggest that DUF1460 members may be involved in the degradation of PG and generation of oligopeptides, consistent with the role of NlpC/P60 enzymes as cell-wall hydrolases.

The active site (Fig. 2C) is composed of a large deep pocket on one side of the active site Cys63 (referred to as the non-primed side, borrowing the terminology for peptidases) and a shallow groove on the other side (primed side), which would complement the shape of saccharide-peptide PG fragments (a large head connected to a thin tail, Fig. 1A). To identify the best possible substrates for AmiA, we fitted various PG fragments into the active site using *in silico* docking. We started by docking mono- and di-saccharides (GlcNAc, MurNAc, GlcNAc-anhMurNAc, and GlcNAc-MurNAc) into the non-primed side pocket to access its relative size. This modeling process suggested that the pocket could best accommodate two saccharide moieties. Next, we tested docking of representative peptide-linked disaccharides (GlcNAc-MurNAc-L-Ala, GlcNAc-anhMurNAc-tripeptide, and GlcNAc-MurNAc-tripeptide). The best docking solution was selected by visual inspection from top-scored poses to best satisfy: 1) shape complementarity between ligand and pocket, *i.e.* saccharides filling the large pocket on the non-primed side along with fitting of the stem peptide into the groove on the primed side; 2) presence of a substrate carbonyl group near Tyr41 so that the tyrosine can fulfill its role in stabilizing the reaction intermediate; 3) favorable interactions between protein and ligand (including van der Waals contacts and hydrogen bonds). Overall, we concluded that the active site of AmiA could best accommodate the GlcNAc-anhMurNAc-tripeptide (Fig. 2D). The lactic acid-L-Ala linkage is located above the catalytic cysteine, suggesting that AmiA is an *N*-acetylmuramoyl-L-alanine amidase.

Enzymatic activity and substrate specificity

The activity of AmiA was then tested on different PG precursors as well as on various characteristic PG fragments (muropeptides) that are known to be released by hydrolases during cell growth and division and subsequently recycled for *de novo* PG synthesis. The main cell-wall degradation product, GlcNAc-anhMurNAc-L-Ala-γ-D-Glu-*meso*-A_2pm-D-Ala (TCT), was readily hydrolyzed by AmiA (Fig. 3). HPLC analysis of the reaction mixture unambiguously identified GlcNAc-anhMurNAc and tetrapeptide as the two reaction products. Thus, AmiA clearly exhibited *N*-acetylmuramoyl-L-alanine amidase-type activity. We estimated the K_m value for TCT to be 16 μM. Inhibition of the enzyme by excess of substrate was observed for substrate concentrations higher than 0.2 – 0.4 mM.

Consistent with the involvement of an essential nucleophilic cysteine residue in the catalytic process, the AmiA activity was dramatically affected by thiol reagents: it was totally abolished following preincubation for 5 min with 60 μM 2,4-dinitrothiocyanobenzene (DTNB) and *p*-hydroxymercuribenzoate (pHMB), and was decreased by 60% and 85% with 2-nitro-5-thiocyanobenzoic acid (NTCB) and N-ethylmaleimide (NEM), respectively. Iodoacetamide, however, had no detectable inhibitory effect in the same assay conditions. Neither $MgCl_2$ (2.5 mM) nor EDTA (5 mM) showed any effect on the enzyme activity.

The substrate specificity of this enzyme was then analyzed by incubating AmiA with a panel of other compounds, including PG precursors, muropeptides and the polymer itself. In a first series of assays that used a high amount of AmiA (1 μl of undiluted stock, *i.e.* 20 μg of protein per assay), total hydrolysis was observed for the GlcNAc-anhMurNAc-L-Ala-D-Glu, GlcNAc-anhMurNAc-L-Ala-γ-D-Glu-*meso*-A_2pm, TCT-dimer (two TCT cross-linked via their tetrapeptide chains), GlcNAc-MurNAc-tetrapeptide and its dimer. AmiA did not exhibit any activity towards any other tested compounds, in particular anhMurNAc-tetrapeptide and MurNAc-tetrapeptide (Table 2). These results indicated that AmiA absolutely required the presence of a disaccharide but accepted either a MurNAc or anhMurNAc moiety at the second position of the disaccharide, as well as di-, tri- and tetra-peptide chains.

The specific activity of AmiA for all its identified substrates was then precisely determined in the same assay conditions but with appropriate enzyme dilutions (Table 2). The activity of AmiA towards TCT was estimated at 8500 nmol/min/mg, as compared to only 1.4 nmol/min/mg for the non-anhydro GlcNAc-MurNAc-tetrapeptide, with the same trend for the corresponding peptide dimers. The length of the stem peptide (di-, tri- or tetrapeptide) did not affect significantly the AmiA activity, but a slight preference (2-fold increased activity) for the dipeptide was observed (Table 2). Overall, we conclude that AmiA is an *N*-acetylmuramoyl-L-alanine amidase that is not active towards intact PG but exhibits a high activity and specificity towards GlcNAc-anhMurNAc-peptides (monomer and dimer forms) derived from the PG polymer.

Structural basis for substrate specificity

Docking of the GlcNAc-anhMurNAc-tripeptide substrate into our crystal structure along with biochemical analysis suggest that the non-primed side binding sites play a more significant role in substrate specificity. To understand the structural basis for substrate specificity, we determined co-crystal structures of AmiA with GlcNAc (PDB ID 4q68) and GlcNAc-anhMurNAc (a reaction product, PDB ID 4q5k). AmiA-GlcNAc complex crystallized in the same crystal form as apo AmiA with only small changes of the unit cell, whereas the AmiA-GlcNAc-anhMurNAc complex crystallized in the same space group but in a different crystal form. Both complex crystals diffracted to high resolution (1.07 Å and 1.3 Å, respectively), and the data and the final models are of similar quality to that of the apo AmiA structure (Table 1). As expected, these structures are very similar to that of the apo protein (RMSDs <0.5 Å for all Cα atoms). GlcNAc is bound to AmiA but with an occupancy of ~0.5 (Fig. 4A), while GlcNAc-anhMurNAc fully occupies the binding site of AmiA (Fig. 4B), suggesting that the disaccharide may have higher binding affinity than

GlcNAc alone. In each case, the quality of electron density for the ligands was excellent and allowed unambiguous identification of the mode of sugar recognition.

As expected, the GlcNAc-anhMurNAc saccharide occupies the non-primed pocket. GlcNAc and anhMurNAc moieties have a ~90° difference in their orientation. GlcNAc is stabilized by extensive hydrogen bonding interactions, while anhMurNAc does not form any specific polar interactions with AmiA except for one hydrogen bond between the lactic acid group and Tyr41OH (Fig. 4C–D). The conformation of GlcNAc moiety in both complexes is identical and is sandwiched between His111 and the α2 helix carrying the catalytic Cys63. The side chains of Asp62, Arg109, and His111 form hydrogen bonds with 6-OH, 4-OH, and 3-OH of GlcNAc, respectively. The *N*-acetyl amine group of GlcNAc is stabilized by hydrogen bonds with His111O (carbonyl) and Tyr152OH. These observations are consistent with the requirement of GlcNAc for enzyme activity. The GlcNAc-anhMurNAc structure explains why GlcNAc-MurNAc-peptide is a poor substrate for AmiA. The sugar moiety of anhMurNAc is located in a small hydrophobic pocket formed by Thr64, Ile113, Met213, Val210, and Ala211, which is sterically and chemically unfavorable to accommodate the polar groups of 1-OH and 6-OH of MurNAc. In our modeled complex (Fig. 2D), the saccharides occupy the same locations but are ~90° rotated in orientation compared to the crystal structures (Fig. 4). Although it is challenging to computationally predict accurate poses for large, flexible substrates, in this case, we were able to correctly predict substrate identities by narrowing down possible substrates significantly via bioinformatics analysis, and exploring differences among substrates (*e.g.* one sugar *vs.* two sugars, MurNAc *vs* anhMurNAc) via docking.

The structural data above suggested that GlcNAc or GlcNAc-anhMurNAc may inhibit the activity of AmiA. Using TCT as a substrate, we tested the activity of AmiA in the presence of excess GlcNAc, MurNAc, or GlcNAc-anhMurNAc. AmiA retained 15%, 25%, and 98% of its activity at 1 mM concentration of GlcNAc-anhMurNAc, MurNAc, and GlcNAc, respectively. At 5 mM inhibitor concentration, >85% of AmiA activity was inhibited by MurNAc, while only ~8% was inhibited by GlcNAc. Thus, the disaccharide is a more efficient inhibitor than MurNAc or GlcNAc, while MurNAc is also a good inhibitor as compared to GlcNAc. Modeling studies suggest that MurNAc can bind AmiA in the same manner as GlcNAc, except that the lactic acid group of MurNAc might increase the binding affinity by forming additional contacts with the protein, as compared to that of GlcNAc.

DUF1460 defines a new amidase family

The AmiA insights derived from our structural and functional studies can be applied to other members of DUF1460, such as, the two structures previously determined by structural genomics efforts, BF2036 and LPG0654. BF2036 is almost identical to AmiA in overall structure (RMSD of only 0.8 Å with sequence identity of 67% for 228 equivalent Cα atoms). LPG0654 is significantly more divergent in sequence and 95 aa longer than AmiA. However, its overall structure is still similar to that of AmiA (RMSD for 2.4 Å and sequence identity of 27% for 219 equivalent Cα atoms). The most significant differences are located in insert or loop regions, such as insert 3, insert 1 and the β1-β2 loop. LPG0654 also has an additional N-terminal helix. However, these structural differences do not impact the active

site of BF2036 and LPG0654, which are highly conserved (Fig. 5A–B). The conformation of the substrate disaccharide in the AmiA structure fits well into the active sites of both BF2036 and LPG0654 (Fig. 5A–B). Most notably, the residues involved in GlcNAc binding are strictly conserved in these three proteins as well as in other DUF1460 members (Fig. 5C). One side of the saccharide binding pocket mainly consists of catalytic residues with a DCXT+H+H/N/Q+Y motif (where X is any amino acid), while the other side is anchored by an RXH motif (Arg105 and His107 of BF2036, Arg156 and His158 of LPG0654, Fig. 5). The arginine side chain in the RXH motif is stabilized by one or more acidic residues, while Nε2 of the histidine hydrogen bonds with a main-chain carbonyl group (Tyr102 of BF2036 and Tyr153 of LPG0654) to orient Nδ1 towards the pocket. Thus, these observations suggest that DUF1460 likely consists of a family of amidases with similar substrate specificities despite significant sequence divergence.

Discussion

In the GlcNAc-anhMurNAc complex structure, the carboxylate carbon atom of the lactic acid is located above Cys63 (distance to Sγ 3.58 A, Fig. 4B). This model is consistent with a papain cysteine peptidase-like mechanism (Storer and Menard, 1994; Xu et al., 2009) for AmiA where the thiol group of Cys63 is deprotonated by the polarizing imidazole group of His212 to form a nucleophilic thiolate/imidazolium ion pair, and the thiolate anion then attacks the carbonyl carbon of the amide scissile bond of the substrate to form a tetrahedral intermediate. The transition state is stabilized by an oxyanion hole that may consist of the hydroxyl group of Tyr41, and potentially the helix dipole of α2 and the carbonyl group of Ala211. His212 then acts as a general acid and protonates the nitrogen in the amide bond, resulting in release of the stem peptide and concomitant formation of the acyl-enzyme intermediate. His212 abstracts a proton from a water molecule while it attacks the carbonyl carbon, forming a second oxyanion hole-stabilized tetrahedral intermediate, which then decomposes with the Cys63 sulfur functioning as a leaving group, releasing GlcNAc-anhMurNAc and regenerating the active enzyme.

PG fragments mediate a range of microbial-host interactions (Cloud-Hansen et al., 2006). As a result, studying PG fragments released by human pathogens and symbioses are of great biological interest. GlcNAc-anhMurNAc-tetrapeptide is a cytotoxin and a potent elicitor of innate immune responses. However, little is currently known about the production and recycling of PG fragments other than in a few bacteria. Here, we identified a new highly specific gut bacterium amidase involved in the degradation of GlcNAc-anhMurNAc-peptide. The clustering of functionally related genes in Bacteroides (Fig. 1B) suggests a new pathway for GlcNAc-anhMurNAc-peptide degradation and recycling, in which AmiA, likely located in periplasm, hydrolyzes GlcNAc-anhMurNAc-peptide into a disaccharide and stem peptide. The stem peptide may be further hydrolyzed by an α/β hydrolase (*e.g.* by removing a D-Ala) and other enzymes, resulting in smaller peptides which are eventually transported by the oligopeptide permease into the cell where they can be further degraded or recycled. This proposed pathway differs from that in *E. coli*, where GlcNAc-anhMurNAc-peptide is degraded in the cytoplasm by NagZ and then AmpD. Although AmiA and AmpD cleave the same bond, they differ in substrate specificities: AmpD can use both GlcNAc-anhMurNAc-peptide and anhMurNAc-peptide equally well and thus do not have specificity for the

GlcNAc moiety (Jacobs et al., 1995), whereas AmiA absolutely requires the GlcNAc moiety for its function. AmpD is a zinc-dependent amidase in the amidase_2 family, and thus is not related to AmiA in sequence and structure. Thus, AmpD and AmiA represent another example of independent evolution of amidases from a zinc peptidase or cysteine peptidase (NlpC/P60) ancestor, respectively, similar to that of MpaA and YkfC (Fig. 1B). Our study of AmiA reinforces the notion that bacteria have evolved more diverse systems for PG recycling, in contrast to the highly conserved PG biosynthesis pathway.

Our structure analysis established that DUF1460 is most closely related to the NlpC/P60 γ-D-Glu-A_2pm amidases. NlpC/P60 amidases have broad-spectrum substrate specificity. Prototypical NlpC/P60 enzymes often have broader substrate specificity and are able to cleave any PG, while NlpC/60 enzymes with more complex architectures, evolved from structural adaptions of the same fold, are more specific and cleave only certain type of PG fragments. Our structural studies identified two modes of such structural adaptations. For the γ-D-Glu-A_2pm YkfC amidase, an N-terminal bacterial SH3 domain is attached to the non-primed side of the active site groove, thereby restricting access to stem peptides with a free L-alanine (Xu et al., 2010; Xu et al., 2009). For AmiA described here, the substrate specificity is achieved through several insertions within the scaffold. In both cases, however, the structural adaptions result in modifications of the non-primed binding sites. These highly specific enzymes have evolved with the specific purpose of salvaging PG components, while avoiding the negative consequences of comprising the integrity of the cell wall or disrupting the biosynthesis of PG.

DUFs account for more than 20% of all protein domains, and many of them have important biological functions (Goodacre et al., 2013). Identifying the function of DUFs aids in uncovering unexpected similarities to other characterized proteins or in identification of novel biological processes and reactions, thereby promoting new frontiers of biological discovery. Structural genomics efforts have contributed a large number of DUF protein structures. However, it remains a significant challenge to leverage this enormous wealth of structural information and achieve a better understanding of their molecular function. Here, we have applied a structure-driven, combined approach to uncover the function of DUF1460. The structural data were indispensable in revealing the relationship with NlpC/P60 proteins, and provided a platform for virtual screening of potential substrates, which significantly reduced the number of biochemical experiments to validate our functional hypotheses. We expect this approach can also be applied to understand the structure and function of other DUFs and accelerate the discovery of new and unexpected biological activities.

Experimental Procedures

Protein production

AmiA was cloned, expressed and purified using a protocol similar as reported previously for CwlT (Xu et al., 2014) (see Supplemental Information, related to Experimental Procedures). The purified protein was concentrated to 20 mg/ml (0.75 mM) by centrifugal ultrafiltration (Millipore) for crystallization trials or biochemical characterizations.

Crystallization

Apo-AmiA was crystallized using the nanodroplet vapor diffusion method (Santarsiero et al., 2002) with standard JCSG crystallization protocols (Lesley et al., 2002). Sitting drops composed of 100 nl protein solution mixed with 100 nl crystallization solution in a sitting drop format were equilibrated against a 35 μl reservoir at 277 K for 29 days prior to harvest. The crystallization reagent consisted of 30% polyethylene glycol (PEG) 1500. PEG 200 was added to a final concentration of 10% (v/v) as a cryoprotectant. Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM)(Cohen et al., 2002) at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA). Apo-AmiA diffraction data were indexed in orthorhombic space group $P2_12_12_1$ with unit cell a=46.3, b=63.7, and c=73.0 Å.

Crystals of the AmiA-ligand complexes were obtained by co-crystallization using the hanging drop method at 277 K. The hanging drops contained 1 μl protein solution mixed with 1 μl reservoir. For AmiA-GlcNAc, 1.3 mM GlcNAc was added to the reservoir (250 μl) containing 30% PEG 1500 and the drops were set up using the apo protein. For AmiA-GlcNAc-anhMurNAc, 0.93 mM of co-crystallization compound GlcNAc-anhMurNAc was added to the protein solution, and the reservoir (500 μl) contained 29% PEG 4000 and 0.1 M Tris pH 8.5. No cryoprotectant was added. The AmiA-GlcNAc diffraction data were indexed in the same space group as the apo-AmiA with only small changes in the unit cell (a=46.5, b=63.6, and c=74.0 Å). AmiA-GlcNAc-anhMurNAc crystals had the same space group but exhibited differences in the unit cell dimensions (a=49.5, b=60.9, and c=80.1 Å).

Data collection, structure solution, and refinement

SAD data for apo-AmiA were collected at the selenium edge using a Pilatus 6M detector (DECTRIS) at SSRL beamline BL12-2. Single-wavelength data were collected for the AmiA-GlcNAc-anhMurNAc and AmiA-GlcNAc complexes using MarCCD325 detector (Rayonix) at SSRL beamline 14-1. Data processing and structure solution were carried out using the automated structure determination protocols developed at the JCSG (van den Bedem et al., 2011; Xu et al., 2010). The data were processed using XDS (Kabsch, 2010). The apo structure was determined using the SAD method, where the location of selenium sites, initial phasing, and identification of the space group were carried out using SHELXD (Sheldrick, 2008). Phase refinement and initial model building were performed using autoSHARP (Bricogne et al., 2003) and ARP/wARP (Langer et al., 2008). The AmiA-GlcNAc-anhMurNAc and AmiA-GlcNAc complex structures were solved by the molecular replacement method implemented in MOLREP (Vagin and Teplyakov, 2010) using the apo structure as a template. Model completion and refinement were performed manually with COOT (Emsley and Cowtan, 2004) and REFMAC5 (Murshudov et al., 2011). All refinements included anisotropic temperature factors and, for the apo structure, experimental phase restraints in the form of Hendrickson-Lattman coefficients. Molecular graphics were prepared with PyMOL (Schrödinger, LLC, USA).

Bioinformatics and molecular modeling

Protein family information for DUF1460 and domain architectures were obtained from Pfam (Punta et al., 2012). Genomic context was studied using KEGG (Ogata et al., 1999). Initial

conformations of various ligands in random orientations were obtained from the PDB databank or generated using ChemBioOf ce (CambridgeSoft Corporation, Cambridge, MA, USA) or JLigand (Lebedev et al., 2012) of CCP4 (Winn et al., 2011) and COOT (Emsley and Cowtan, 2004). Flexible ligand docking was performed with AutoDock Vina (Trott and Olson, 2010) and Glide (Schrödinger LLC, USA) to generate a list of poses with highest scores. The final solutions were obtained by visual inspection to satisfy geometrical and chemical restraints, manual adjustments in COOT to fine tune local interactions, and then energy minimization.

Activity assays

The AmiA activity assay consisted of monitoring the hydrolysis of PG-related compounds in a reaction mixture (50 µl) containing 50 mM Tris-HCl, pH 7.8, 0.2 mM substrate, and purified AmiA enzyme. 1 µl of the protein stock used for structure determination (*i.e.* 20 µg of AmiA protein) was used first in preliminary assays aiming at identifying compounds that were or were not substrates of this enzyme. Subsequent assays aiming at precisely determining the specific activity of AmiA for these substrates used appropriately adjusted amounts of enzyme (from 5 ng to 2 µg; protein dilutions being performed in 20 mM potassium phosphate buffer pH 8). After 30 min of incubation at 37°C, reactions were stopped by flash freezing in liquid nitrogen. In each case, the substrate and reaction products were separated by HPLC, using appropriate column and elution conditions. For all tested compounds, a Nucleosyl 100 5 µ C_{18} column (250 × 4.6 mm, Alltech France) was used and elution was with 50 mM sodium phosphate, pH 4.5, with or without application of a linear gradient of MeOH (from 0 to 25%) between 0 and 60 min, at a flow rate of 0.6 ml/min. Peaks were detected by measuring the absorbance at 262 nm (nucleotide PG precursors) or 207 nm (other compounds). Retention times of AmiA substrates and products observed in these HPLC conditions are reported in Supplementary Table S1 (related to Table 2). Identification of these compounds was based on their retention time as compared to authentic standards as well as on their amino acid and amino sugar composition determined with a Hitachi model L8800 analyzer (ScienceTec, Les Ulis, France) after hydrolysis of samples in 6 M HCl at 95°C for 16 h. For determination of the K_m for TCT, the same assay was used at various concentrations of this substrate in the 2 µM – 400 µM range and 1.3 ng of enzyme per assay.

Inhibition of AmiA activity

AmiA (0.75 µM) was pre-incubated for 5 min at 37°C with various thiol reagents (DTNB, NTCB, pHMB, iodoacetamide and NEM) at a 60 µM final concentration in 20 mM potassium phosphate buffer, pH 7.4. Then, 2-µl aliquots of these mixtures were added to 50-µl standard assay reaction mixtures containing GlcNAc-anhMurNAc-tetrapeptide (TCT) as the substrate. After 30 min of incubation at 37°C, reactions were stopped and mixtures were analyzed by HPLC as described above. The inhibitory effects of GlcNAc and MurNAc at 1 mM and 5 mM, and GlcNAc-anhMurNAc at 0.3 mM and 1 mM were tested using 0.1 mM of TCT as substrate.

Peptidoglycan precursors and muropeptides

UDP-MurNAc-peptides were generated as described earlier (Flouret et al., 1981) and their MurNAc-peptides derivatives were obtained by mild acid hydrolysis (0.1 M HCl, 100°C, 30 min) (Hervé et al., 2007). Lactoyl-pentapeptide and free pentapeptide were produced by treatment of MurNAc-L-Ala-γ-D-Glu-*meso*-A_2pm-D-Ala-D-Ala with 4 M ammonium hydroxide (Stenbak et al., 2004) and *E. coli* AmiD *N*-acetylmuramoyl-L-alanine amidase (Pennartz et al., 2009), respectively. Peptidoglycan (PG) was purified from an *E. coli* Δ*lpp* mutant strain that does not express the Lpp lipoprotein (Leulier et al., 2003). GlcNAc-anhMurNAc-L-Ala-γ-D-Glu-*meso*-A_2pm-D-Ala (TCT) and its dimer was produced by digestion of PG with *E. coli* SltY lytic transglycosylase (Stenbak et al., 2004). The TCT derivative carrying only a L-Ala-γ-D-Glu-*meso* -A_2pm tripeptide chain was produced by treatment of TCT with LdcA L,D-carboxypeptidase (Das et al., 2013) and the TCT derivative carrying a L-Ala-γ-D-Glu side chain was generated by treatment of TCT with a γ-D-Glu-A_2pm amidase activity purified from *Desulfovibrio vulgaris* (data not shown). AnhMurNAc-tetrapeptide was produced by treatment of TCT with *E. coli* NagZ β-*N*-acetylglucosaminidase (Stenbak et al., 2004). GlcNAc-MurNAc-tetrapeptide and its dimer were generated by treatment of PG with muramidase (mutanolysin, Sigma) (Stenbak et al., 2004). GlcNAc-anhMurNAc was prepared by cleaving TCT with pure *B. uniformis* AmiA activity. All these compounds were purified by HPLC and their composition was controlled by amino acid and amino sugar content analysis and/or by MALDI-TOF mass spectrometry.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Irimpan I. Mathews and all of the members of the JCSG high-throughput structural biology pipeline for their contribution to this work. This work was supported by the NIH, National Institute of General Medical Sciences (NIGMS), Protein Structure Initiative [U54 GM094586], CNRS and Université Paris-Sud. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. Genomic DNA from *Bacteroides uniformis* was extracted from cells (ATCC 8492) obtained from the American Type Culture Collection (ATCC).

References

Anantharaman V, Aravind L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol. 2003; 4:R11. [PubMed: 12620121]

Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr D Biol Crystallogr. 2003; 59:2023–2030. [PubMed: 14573958]

Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010; 66:12–21. [PubMed: 20057044]

Cloud-Hansen KA, Peterson SB, Stabb EV, Goldman WE, McFall-Ngai MJ, Handelsman J. Breaching the great wall: peptidoglycan and microbial interactions. Nat Rev Microbiol. 2006; 4:710–716. [PubMed: 16894338]

Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP. An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact samples cassettes and a small-scale robot. J Appl Cryst. 2002; 35:720–726. [PubMed: 24899734]

Correa VR, Majerczak DR, Ammar el D, Merighi M, Pratt RC, Hogenhout SA, Coplin DL, Redinbaugh MG. The bacterium *Pantoea stewartii* uses two different type III secretion systems to colonize its plant host and insect vector. Appl Environ Microbiol. 2012; 78:6327–6336. [PubMed: 22773631]

Das D, Hervé M, Elsliger MA, Kadam RU, Grant JC, Chiu HJ, Knuth MW, Klock HE, Miller MD, Godzik A, et al. Structure and function of a novel LD-carboxypeptidase a involved in peptidoglycan recycling. J Bacteriol. 2013; 195:5555–5566. [PubMed: 24123814]

Emsley P, Cowtan K. COOT: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]

Flouret B, Mengin-Lecreulx D, van Heijenoort J. Reverse-phase high-pressure liquid chromatography of uridine diphosphate *N*-acetylmuramyl peptide precursors of bacterial cell wall peptidoglycan. Anal Biochem. 1981; 114:59–63. [PubMed: 7283154]

Fouts DE, Abramovitch RB, Alfano JR, Baldo AM, Buell CR, Cartinhour S, Chatterjee AK, D'Ascenzo M, Gwinn ML, Lazarowitz SG, et al. Genomewide identification of Pseudomonas syringae pv. tomato DC3000 promoters controlled by the HrpL alternative sigma factor. Proc Natl Acad Sci USA. 2002; 99:2275–2280. [PubMed: 11854524]

Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, et al. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. Infect Immun. 2011; 79:4286–4298. [PubMed: 21896772]

Goodacre NF, Gerloff DL, Uetz P. Protein domains of unknown function are essential in bacteria. MBio. 2013; 5:e00744–00713. [PubMed: 24381303]

Hervé M, Boniface A, Gobec S, Blanot D, Mengin-Lecreulx D. Biochemical characterization and physiological properties of *Escherichia coli* UDP-*N*-acetylmuramate:L-alanyl-γ-D-glutamyl-*meso*-diaminopimelate ligase. J Bacteriol. 2007; 189:3987–3995. [PubMed: 17384195]

Jacobs C, Joris B, Jamin M, Klarsov K, Van Beeumen J, Mengin-Lecreulx D, van Heijenoort J, Park JT, Normark S, Frére JM. AmpD, essential for both β-lactamase regulation and cell wall recycling, is a novel cytosolic N-acetylmuramyl-L-alanine amidase. Mol Microbiol. 1995; 15:553–559. [PubMed: 7783625]

Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132. [PubMed: 20124692]

Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc. 2008; 3:1171–1179. [PubMed: 18600222]

Lebedev AA, Young P, Isupov MN, Moroz OV, Vagin AA, Murshudov GN. JLigand: a graphical tool for the CCP4 template-restraint library. Acta Crystallogr D Biol Crystallogr. 2012; 68:431–440. [PubMed: 22505263]

Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, et al. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA. 2002; 99:11664–11669. [PubMed: 12193646]

Leulier F, Parquet C, Pili-Floury S, Ryu JH, Caroff M, Lee WJ, Mengin-Lecreulx D, Lemaitre B. The Drosophila immune system detects bacteria through specific peptidoglycan recognition. Nat Immunol. 2003; 4:478–484. [PubMed: 12692550]

Maqbool A, Hervé M, Mengin-Lecreulx D, Wilkinson AJ, Thomas GH. MpaA is a murein-tripeptide-specific zinc carboxypeptidase that functions as part of a catabolic pathway for peptidoglycan-derived peptides in γ-proteobacteria. Biochem J. 2012; 448:329–341. [PubMed: 22970852]

Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011; 67:355–367. [PubMed: 21460454]

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 1999; 27:29–34. [PubMed: 9847135]

Park JT, Uehara T. How bacteria consume their own exoskeletons (turnover and recycling of cell wall peptidoglycan). Microbiol Mol Biol Rev. 2008; 72:211–227. [PubMed: 18535144]

Pennartz A, Généreux C, Parquet C, Mengin-Lecreulx D, Joris B. Substrate-induced inactivation of the *Escherichia coli* AmiD *N*-acetylmuramoyl-L-alanine amidase highlights a new strategy to inhibit this class of enzyme. Antimicrob Agents Chemother. 2009; 53:2991–2997. [PubMed: 19237650]

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. The Pfam protein families database. Nucleic Acids Res. 2012; 40:D290–301. [PubMed: 22127870]

Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, Uber DC, Cornell EW, Nordmeyer RA, Kolbe WF, et al. An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr. 2002; 35:278–281.

Sheldrick GM. A short history of SHELX. Acta Crystallogr A Found Crystallogr. 2008; 64:112–122.

Stenbak CR, Ryu JH, Leulier F, Pili-Floury S, Parquet C, Hervé M, Chaput C, Boneca IG, Lee WJ, Lemaitre B, et al. Peptidoglycan molecular requirements allowing detection by the *Drosophila* immune deficiency pathway. J Immunol. 2004; 173:7339–7348. [PubMed: 15585858]

Storer AC, Menard R. Catalytic mechanism in papain family of cysteine peptidases. Methods Enzymol. 1994; 244:486–500. [PubMed: 7845227]

Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455–461. [PubMed: 19499576]

Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr. 2010; 66:22–25. [PubMed: 20057045]

van den Bedem H, Wolf G, Xu Q, Deacon AM. Distributed structure determination at the JCSG. Acta Crystallogr D Biol Crystallogr. 2011; 67:368–375. [PubMed: 21460455]

van Heijenoort J. Peptidoglycan hydrolases of *Escherichia coli*. Microbiol Mol Biol Rev. 2011; 75:636–663. [PubMed: 22126997]

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011; 67:235–242. [PubMed: 21460441]

Xu Q, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Cai X, Carlton D, Chen C, Chiu HJ, Chiu M, et al. Structure of the γ-D-glutamyl-L-diamino acid endopeptidase YkfC from *Bacillus cereus* in complex with L-Ala-γ-D-Glu: insights into substrate recognition by NlpC/P60 cysteine peptidases. Acta Crystallogr F Struct Biol Cryst Commun. 2010; 66:1354–1364.

Xu Q, Chiu HJ, Farr CL, Jaroszewski L, Knuth MW, Miller MD, Lesley SA, Godzik A, Elsliger MA, Deacon AM, et al. Structures of a bifunctional cell wall hydrolase CwlT containing a novel bacterial lysozyme and an NlpC/P60 DL-endopeptidase. J Mol Biol. 2014; 426:169–184. [PubMed: 24051416]

Xu Q, Rawlings ND, Chiu HJ, Jaroszewski L, Klock HE, Knuth MW, Miller MD, Elsliger MA, Deacon AM, Godzik A, et al. Structural analysis of papain-like NlpC/P60 superfamily enzymes with a circularly permuted topology reveals potential lipid binding sites. PLoS ONE. 2011; 6:e22013. [PubMed: 21799766]

Xu Q, Sudek S, McMullan D, Miller MD, Geierstanger B, Jones DH, Krishna SS, Spraggon G, Bursalay B, Abdubek P, et al. Structural basis of murein peptide specificity of a γ-D-glutamyl-l-diamino acid endopeptidase. Structure. 2009; 17:303–313. [PubMed: 19217401]

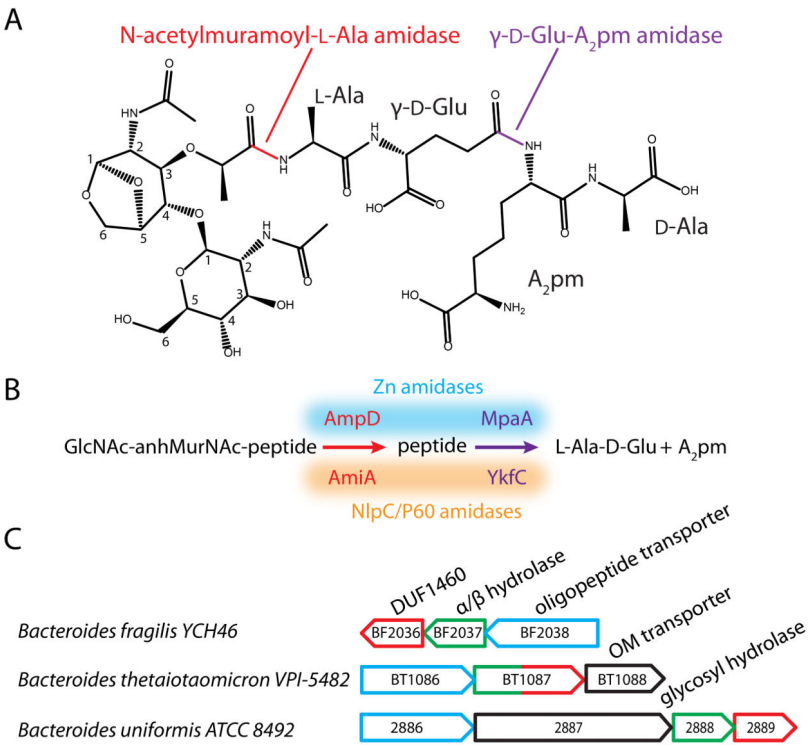

Fig. 1.

The main PG degradation product and putative operons in Bacteroides involved in its degradation. (A) Chemical structure of GlcNAc-anhydroMurNAc-tetrapeptide (TCT), produced by lytic transglycosylases and D-Ala-D-Ala carboxypeptidase. The amide bonds cleaved by *N*-acetylmuramoyl-L-alanine amidase and γ-D-Glu-A_2pm amidase (also known as endopeptidase) are highlighted in red and purple, respectively. (B) Amidases involved in degradation of GlcNAc-anhydroMurNAc-peptide in *E. coli* and other bacteria. The *E. coli* enzymes AmpD and MpaA are shown at the top, while *Bacteroides uniformis* AmiA (this study) and *Bacillus subtilis* YkfC are shown at the bottom. (C) Putative operons encoding DUF1460 members in Bacteroides. Genes encoding DUF1460 members are colored red, α/β hydrolase orthologs green, and oligopeptide transporter orthologs cyan. The PATRIC (Gillespie et al., 2011) annotation is used for *B. uniformis* ATCC 8492.

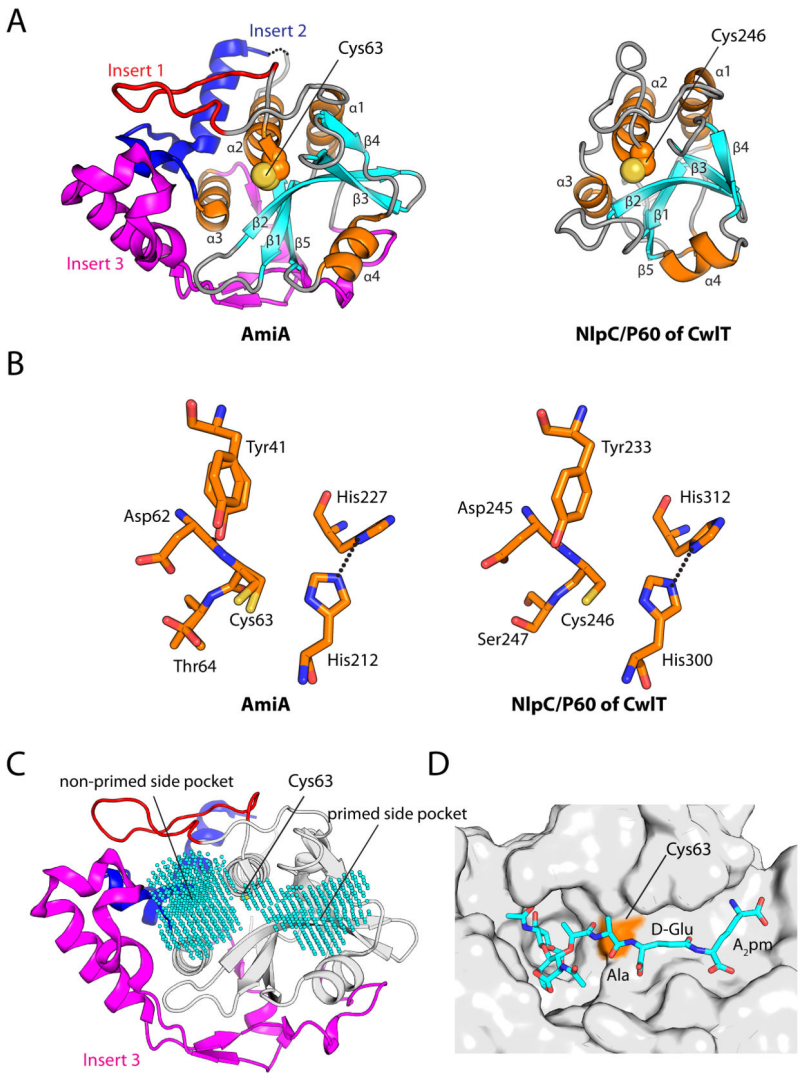

Fig. 2.

AmiA structure and active site. (A) AmiA structure and comparison with a prototypical NlpC/P60 γ-D-Glu-A_2pm amidase. The fold of a typical γ-D-Glu-A_2pm-specific amidase (PDB 4hpe) is shown on the right. AmiA contains the same NlpC/P60 catalytic core (orange/cyan) except for three long insertions (magenta). These insertions define a large pocket that is absent in the prototypical NlpC/P60 domains (see panel C). (B) The active site of AmiA closely resembles that of γ-D-Glu-A_2pm amidase. The catalytic triad consists of Cys/His/His. (C) Architecture of the AmiA substrate-binding site. Grid points filling the binding pocket are shown as cyan dots. (D) A predicted binding mode of GlcNAc-anhMurNAc-tripeptide in the active site of AmiA, which differs for the actual crystal structure (see Fig. 4).

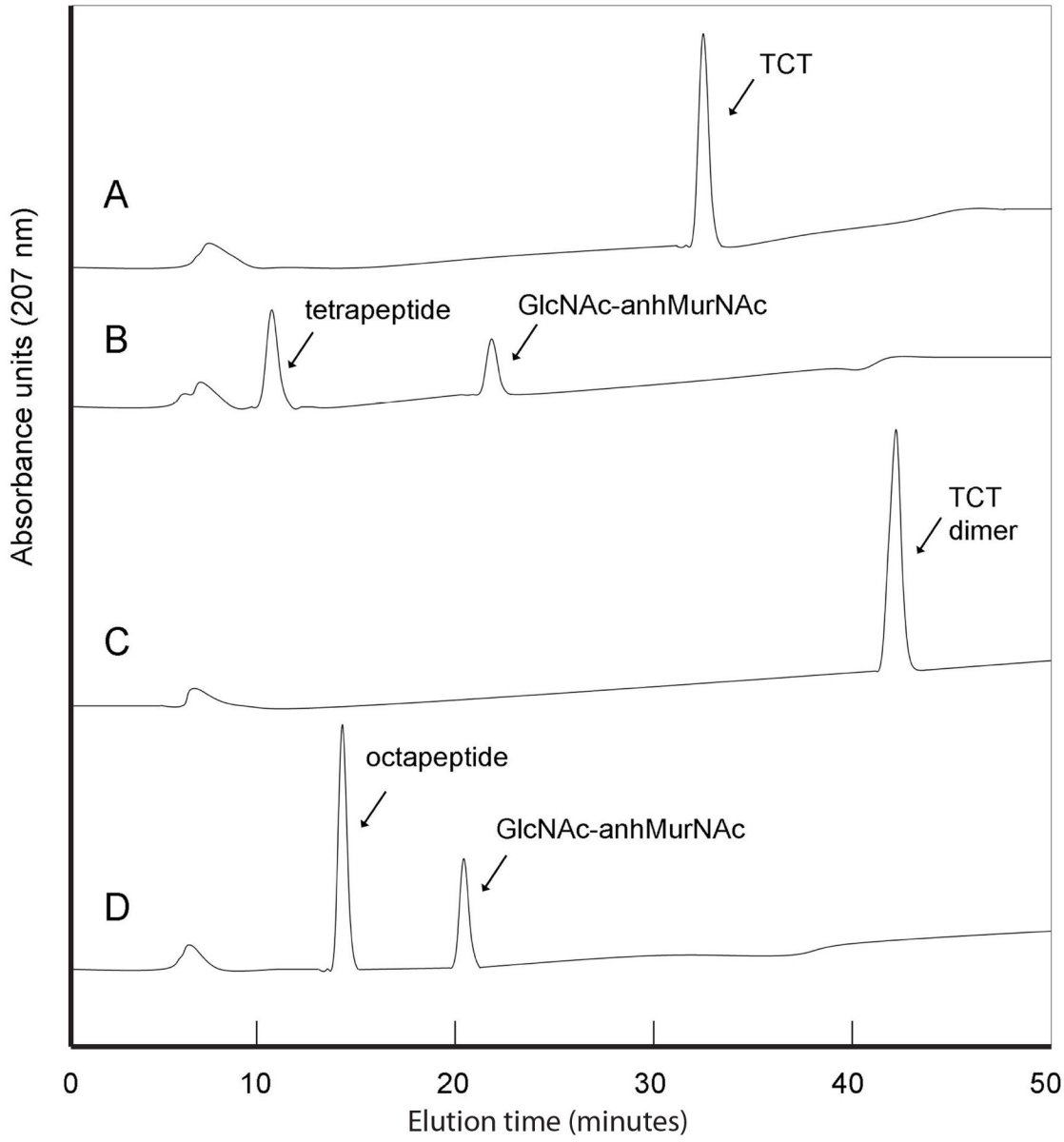

Fig. 3.

Enzymatic activity of AmiA. TCT (A, B) and TCT dimer (C, D) were incubated with (B, D) or without (A, C) purified AmiA (0.1 μg) for 30 min at 37°C. Reaction mixtures were analyzed by HPLC on a column of Nucleosyl 100 5 μ C18. Elution was performed at 0.6 ml/min with 50 mM sodium phosphate, pH 4.5, and a linear gradient of MeOH from 0 to 25% applied over 60 min. Peaks were detected at 207 nm. Both TCT and TCT dimer were cleaved, generating in each case two products that were identified as tetrapeptide and GlcNAc-anhMurNAc, and octapeptide and GlcNAc-anhMurNAc, respectively, consistent with an *N*-acetylmuramoyl-L-alanine amidase activity of AmiA.

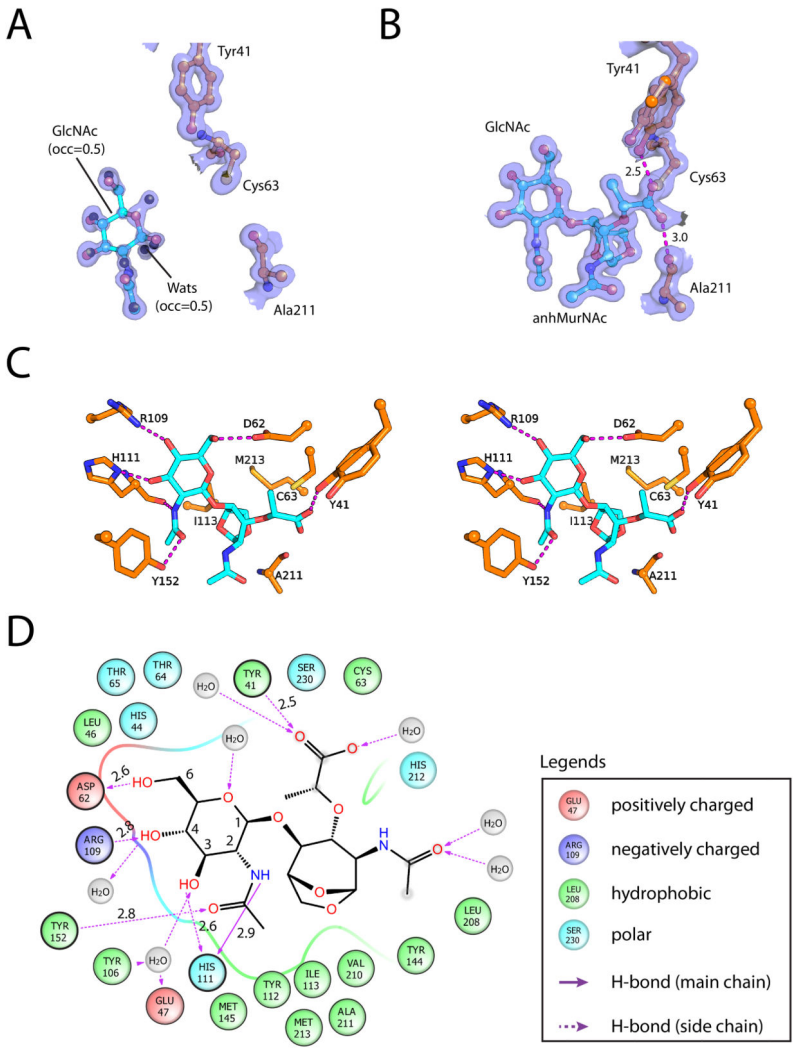

Fig. 4.

Crystal structures of AmiA in complex with GlcNAc or GlcNAc-anhMurNAc. (A) 2Fo-Fc density map (blue surface, 1.0 σ) of the AmiA-GlcNAc active site. The protein (orange/red/blue) and ligand (cyan/red/blue) are shown as sticks. Water molecules that are present in the absence of GlcNAc, which is present at 50% occupancy, are shown as black spheres. (B) 2Fo-Fc density map (blue surface, 1.0 σ) of AmiA-GlcNAc-anhMurNAc active site. (C) Stereoview of interactions between AmiA (orange/red/blue) and GlcNAc-anhMurNAc (cyan/red/blue). Hydrogen bonds are shown as dashed lines. (D) Schematic diagram of interactions between AmiA and GlcNAc-anhMurNAc.

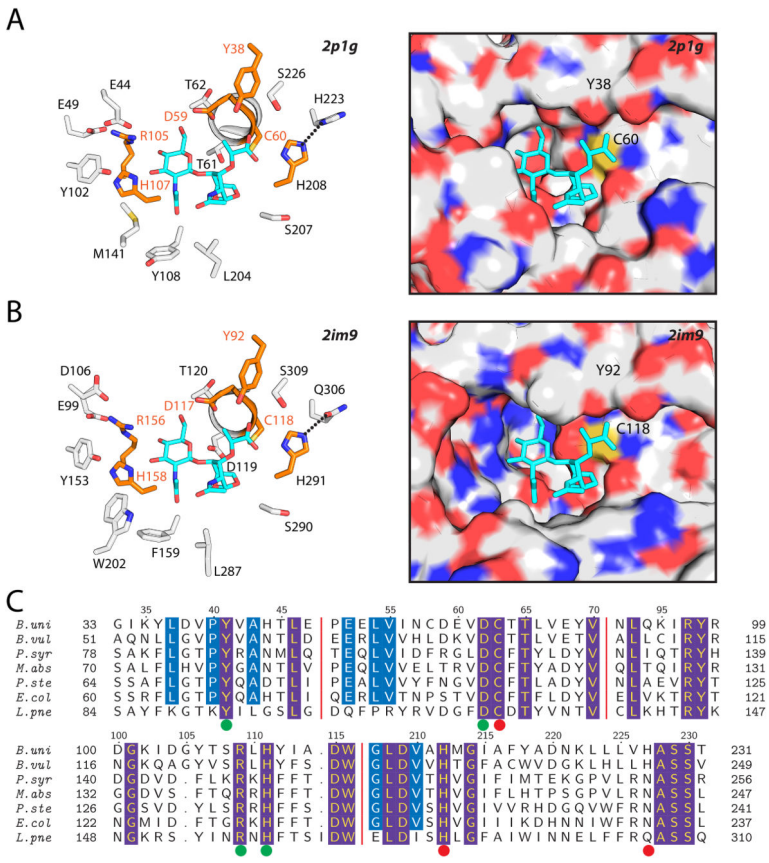

Fig. 5.

The catalytic residues and sugar-binding pocket are highly conserved in the DUF1460 family. (A) The active site of BF2036 (PDB 2p1g) in stick (left) and surface representation (right). The strictly conserved residues in the DC+H+Y+RXH motif are highlighted in orange/red/blue. A modeled GlcNAc-anhMurNAc ligand is in cyan/red/blue sticks. (B) The active site of LPG0654 (PDB 2im9). (C) The conservation of active sites among representative DUF1460 members. Sequence numbering for AmiA is shown at the top row. Residues at the catalytic center, non-primed and primed side of the substrate-binding pocket are marked by red, green and cyan dots on the bottom row, respectively. B.uni: *Bacteroides uniformis*, B.vul: *Bacteroides vulgatus*, P.syr: *Pseudomonas syringae*, M.abs: *Mycobacterium abscessus*, P.ste: *Pantoea stewartii*, E.col: *Escherichia coli* DEC5E, L.pne (PDB ID 2im9): *Legionella pneumophila*.

Table 1

Crystallization, data collection and refinement statistics

Crystallization and data collection			
Crystal (PDB ID)	AmiA apo (4h4j)	AmiA-GlcNAc (4q68)	AmiA-GlcNAc-anhMurNAc (4q5k)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell	*a*=46.3, *b*=63.7, *c*=73.0 Å	*a*=46.5, *b*=63.6, *c*=74.0 Å	*a*=49.5, *b*=60.9, *c*=80.1 Å
Beamline	SSRL 12-2	SSRL 14-1	SSRL 14-1
Wavelength (Å)	0.9794	1.0000	1.0000
Resolution range (Å)	28.7-1.15	46.5-1.07	40.1-1.30
No. observations	935,696	366,347	240,318
No. unique reflections	71,198	93,219	60,273
Completeness (%)[a]	92.3 (66.6)	96.2 (88.3)	99.9 (100.0)
Mean I/σ (I)[a]	17.5 (2.8)	22.8 (3.5)	20.4 (2.7)
R_{merge} on I[a] (%)[a]	9.5 (92.9)	3.4 (32.2)	3.9 (53.4)
R_{meas} on I[a](%)[a]	9.9 (97.2)	3.9 (38.5)	4.5 (61.7)
R_{pim} on I[a] (%)[a]	2.7 (28.2)	1.9 (20.5)	2.2 (30.3)
CC(1/2) on I (%)[a]	99.9 (79.3)	100 (86.9)	99.9 (80.4)
Highest resolution shell	1.21-1.15	1.13-1.07	1.37-1.30
Model and refinement statistics			
No. reflections (total)	71,147	93,156	60,203
No. reflections (test)	3,572	4,687	2,940
R_{cryst} (%)	13.1	11.0	11.6
R_{free} (%)	16.4	13.8	14.9
Stereochemical parameters			
Restraints (RMSD observed)			
Bond lengths (Å)	0.017	0.009	0.011
Bond angles (°)	1.76	1.36	1.37
MolProbity score (percentile)	93	96	99
Ramachandran plot (%)[b]	98.2 (0)	98.7 (0)	98.3 (0)
<B> (all, protein, ligand, $Å^2$)	18.3/16.1	12.9/9.6/9.1	17.8/14.9/13.6
ESU based on R_{free} (Å)	0.039	0.026	0.041
No. protein residues	237	235	237
Solvent	431 H_2O	576 H_2O and 3 Na	455 H_2O and 1 Na

[a] Highest resolution shell in parentheses.

[b] Percentage of residues in favored regions of Ramachandran plot (No. outliers in parentheses) as calculated by MolProbity.

ESU = Estimated Standard Uncertainty in coordinates.

$R_{merge}=\Sigma_{hkl}\Sigma_i|I_i(hkl)-<I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, R_{meas}(redundancy-independent R_{merge}) $=\Sigma_{hkl}[N_{hkl}/(N_{hkl}-1)]^{1/2}\Sigma_i|I_i(hkl)-<I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, and R_{pim}(precision-indicating R_{merge})$=\Sigma_{hkl}[1/(N_{hkl}-1)]^{1/2}\Sigma_i|I_i(hkl)-<I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$. CC(1/2) values for <I> are calculated by splitting the data randomly in half.

$R_{cryst} = \Sigma| |F_{obs}| - |F_{calc}| |/\Sigma|F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.

R_{free} = as for R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from refinement.

Table 2

Enzymatic activity of AmiA

Substrate	Activity[a]	Specific activity[b] (nmol/min/mg)
GlcNAc-anhMurNAc-tetrapeptide (TCT)	+	8,500
GlcNAc-anhMurNAc-tripeptide	+	6,200
GlcNAc-anhMurNAc-dipeptide	+	18,600
Dimer of TCT	+	10,500
AnhMurNAc-tetrapeptide	−	
GlcNAc-MurNAc-tetrapeptide	+	1.4
Dimer of GlcNAc-MurNAc-tetrapeptide	+	1.8
MurNAc-tetrapeptide	−	
MurNAc-tripeptide	−	
UDP-MurNAc-tetrapeptide	−	
UDP-MurNAc-pentapeptide	−	
Pentapeptide	−	
Lactoyl-pentapeptide	−	
Peptidoglycan	−	

[a] − and + indicate “undetected” (below the limitations of the analytical method used; < 2% substrate transformation, specific activity < 0.2 nmol/min/mg) and “complete substrate transformation”, respectively, when 1 μl of undiluted stock of AmiA, *i.e.* 20 μg of protein, was used per assay.

[b] Determined as described in Experimental Procedures using appropriate dilutions of enzyme and substrate concentrations of 0.2 mM.