Structural and biochemical characterization of the β-N-acetylglucosaminidase from Thermotoga maritima: toward rationalization of mechanistic knowledge in the GH73 family

Alexandra Lipski, Mireille Hervé, Vincent Lombard, Didier Nurizzo, Dominique Mengin-Lecreulx, Yves Bourne, Florence Vincent

Glycobiology Advance Access published October 24, 2014

Structural and biochemical characterization of the β-*N*-acetylglucosaminidase from *Thermotoga maritima*: toward rationalization of mechanistic knowledge in the GH73 family

Alexandra Lipski [1,4], Mireille Hervé [2], Vincent Lombard [3,4], Didier Nurizzo [5], Dominique Mengin-Lecreulx [2], Yves Bourne [3,4], Florence Vincent [3,4§]

[1] Laboratory for Biocrystallography and Structural Biology of Therapeutic Targets, Molecular and Structural Bases of Infectious Diseases, UMR 5086 CNRS and University of Lyon, 7 passage du Vercors, F-69367 Lyon Cedex 07, France

[2] Laboratoire des Enveloppes Bactériennes et Antibiotiques, Institut de Biochimie et Biophysique Moléculaire et Cellulaire, UMR 8619 CNRS, Université de Paris-Sud, 91405 Orsay, France

[3] Aix-Marseille University, AFMB UMR7257, 163 avenue de luminy, 13288 Marseille cedex 09, France

[4] CNRS, AFMB UMR7257, 163 avenue de luminy, 13288 Marseille cedex 09, France

[5]European Synchrotron Radiation Facility, Polygone Scientifique Louis Néel, 6 rue Jules Horowitz, 38000 Grenoble, France

[§]To whom correspondence should be addressed: Florence Vincent, laboratoire AFMB, CNRS, 163 avenue de Luminy, 13288 Marseille cedex 09, France, Tel.: (33) 4 921825566; Fax: (33)4 91266720; Email: florence.vincent@afmb.univ-mrs.fr

ABSTRACT

Members of the GH73 glycosidase family cleave the β-1,4-glycosidic bond between the *N*-acetylglucosaminyl (GlcNAc) and *N*-acetylmuramyl (MurNAc) moieties in bacterial peptidoglycan. A catalytic mechanism has been proposed for members FlgJ, Auto, AcmA and Atl(WM) and the structural analysis of FlgJ and Auto revealed a conserved α/β fold reminiscent of the distantly-related GH23 lysozyme. Comparison of the active site residues reveals variability in the nature of the catalytic general base suggesting two distinct catalytic mechanisms: an inverting mechanism involving two distant glutamate residues and a substrate-assisted mechanism involving anchimeric assistance by the C2-acetamido group of the GlcNAc moiety. Herein, we present the biochemical characterization and crystal structure of TM0633 from the hyperthermophilic bacterium *Thermotoga maritima*. TM0633 adopts the α/β fold of the family and displays β-*N*-acetylglucosaminidase activity on intact peptidoglycan sacculi. Site-directed mutagenesis identifies Glu34, Glu65 and Tyr118 as important residues for catalysis. A thorough bioinformatic analysis of the GH73 sequences identified five phylogenetic clusters. TM0633, FlgJ and Auto belong to a group of three clusters that conserve two carboxylate residues involved in a classical inverting acid-base mechanism. Members of the other two clusters lack a conserved catalytic general base supporting a substrate-assisted mechanism. Molecular modeling of representative members from each cluster suggests that variability in length of the β-hairpin region above the active site confers ligand binding specificity and modulates the catalytic mechanisms within the GH73 family.

Keywords: β-N-acetylglucosaminidase / peptidoglycan / X-ray structure / catalytic mechanism

Introduction

Peptidoglycan is the major component of the cell wall in Gram-positive bacteria. It is composed of muropeptides made of two repeated sugar moieties, *N*-acetylmuramic acid (MurNAc) and *N*-acetylglucosamine (GlcNAc), forming glycan chains interconnected by short peptides (Vollmer et al. 2008). Peptidoglycan plays an important role in preserving the cell wall integrity by maintaining the osmotic pressure and it is also intimately involved in the processes of cell growth and cell division (Vollmer et al. 2008). Therefore, the peptidoglycan cell wall is a target for several enzymes called autolysins, which are responsible for peptidoglycan hydrolysis throughout bacterial growth to ensure insertion of new precursors and separation of daughter cells after division (Typas et al. 2012). Among the autolysin family, two types of glycoside hydrolases (GHs), *N*-acetylglucosaminidases and *N*-acetylmuraminidases, cleave the GlcNAc-β1,4-MurNAc bonds/linkages to release GlcNAc and MurNAc reducing ends respectively. In the CAZy database (http://www.cazy.org), *N*-acetylmuraminidases or lysozymes are found in families GH22, GH23, GH24, GH25 and GH18 (Lombard et al. 2014). In contrast, all autolysins harboring *N*-acetylglucosaminidase activity are classified in family GH73 with the exception of a peptidoglycan *N*-acetylglucosaminidase found in family GH18 (Bokma et al. 1997).

Currently, family GH73 contains more than 2800 members, among which several have been biochemically characterized and shown to specifically hydrolyze the GlcNAc-β-1,4-MurNAc linkage in peptidoglycan to release muropeptides carrying GlcNAc reducing ends. GH73 autolysins are involved in daughter cell separation during vegetative growth and they hydrolyze the septum after cell division (Camiade et al. 2010; Eckert et al. 2006). Occasionally, GH73 enzymes are active during host cell invasion such as the virulence-associated peptidoglycan hydrolases Auto and IspC from *Listeria monocytogenes* (Bublitz et al. 2009;Ronholm et al. 2012). GH73 enzymes are surface-located or membrane anchored, and they show a high degree of modularity. For instance, several GH73 members are bifunctional enzymes harboring an additional *N*-acetylmuramoyl-L-Ala amidase module (Bourgeois et al. 2009; Yokoi et al. 2008; Foster 1995; Rashid et al. 1995). They can also display repeated sequences that could be involved in binding bacterial cell wall, *e.g*. CBM50, also known as LysM domains, which target peptidoglycan (Buist et al. 2008 Eckert et al. 2006; Mesnage et al. 2014), or have unknown function (Bublitz et al. 2009; Ronholm et al. 2012; Rashid et al. 1995; Horsburgh et al. 2003; Huard et al. 2003).

Beside their diversity in enzyme modularity, members of the GH73 family also diverge in terms of catalytic mechanism. Among the four biochemically characterized GH73 enzymes, FlgJ, Auto, AcmA and Atl(WM), the catalytic general acid has been identifed as an invariant glutamate (Maruyama et al. 2010; Bublitz et al. 2009; Inagaki et al. 2009; Yokoi et al. 2008). The crystal structures of FlgJ and Auto show a typical lysozyme-like α/β fold,

consisting of an α-lobe and a β-hairpin that create an extended substrate binding groove, as found in the GH19, GH22 and GH23 enzymes (Bublitz et al. 2009; Hashimoto et al. 2009). Therefore, the general acid residue, e.g. Glu185 and Glu122 in FlgJ and Auto, respectively, occupies a position analogous to the catalytically essential carboxylic residue in other structurally-related GHs enzymes (Bublitz et al. 2009; Hashimoto et al. 2009). However, FlgJ and Auto lack the second catalytic carboxylate that acts as the catalytic general base located near the position of Asp52 in hen egg white GH22 lysozyme (HEWL) (Vocadlo et al. 2001). Interestingly, an equivalent general base residue could be located 13Å away from the general acid residue in the active site of Auto (Bublitz et al. 2009). Mutations of this putative distant general base in Auto and FlgJ are associated with a marked decrease in catalytic activity, supporting the role of a catalytic base activating a water molecule for the nucleophilic attack of the anomeric center in an inverting mechanism (Bublitz et al. 2009; Maruyama et al. 2010). However, in AcmA and AtlWM, a significant residual activity (32-44%) was observed when the corresponding putative base was converted into glutamine or asparagine, consistent with a substrate-assisted catalysis involving anchimeric assistance by the acetamido group of the GlcNAc moiety (Mark et al. 2001) (Figure 1).

To support the existence of two catalytic mechanisms in the GH73 family, we present the structural and functional study of non-modular flagellar-like TM0633 from the hyperthermophilic bacterium *Thermotoga maritima* coupled to a bioinformatic analysis. TM0633 has a β-*N*-acetylglucosaminidase activity typically found in GH73 enzymes. Two glutamate and one aromatic residue have been identified as important residues for catalysis, by site-directed mutagenesis. Sequence analysis of the GH73 family identified five phylogenetic clusters harbouring characteristic sequence motifs, supporting existence of two distinct catalytic mechanisms. Moreover, molecular modeling of representative members from each cluster suggests that length variability of the β-hairpin region placed above the active site should contribute to modulation of the catalytic mechanisms in the GH73 family.

Results and Discussion

TM0633 is an N-acetylglucosaminidase. To identify the nature of the glycosidic bond cleaved by TM0633, purified peptidoglycan polymers from *E. coli* were incubated with the enzyme and the resulting digestion products were separated by HPLC. Fractions corresponding to the main peaks (A1 to A4) were collected and analyzed by mass spectrometry (Figure 2B/D). Their amino acid and hexosamine compositions as well as their *m/z* values (940.6 and 1862.1 for $(M+H)^+$ ions of A1/A2 and A3/A4, respectively), indicated that the digestion products were the monomeric (A1, A2) and dimeric (A3, A4) forms of a disaccharide-tetrapeptide containing GlcNAc, MurNAc, and a L-Ala-γ-D-

Glu-*meso*-A_2pm-D-Ala peptide (where A_2pm represents diaminopimelic acid) (Figure 2D). Due to the presence of a free hydroxyl group at the reducing end, two distinct peaks, corresponding to the characteristic α and β anomers, were observed for each of these products. Moreover, the enzymatic digestion pattern of TM0633 clearly revealed a profile different from that obtained from the commercial muramidase enzyme "mutanolysin" from *Streptomyces globisporus* (Figure 2B/C), confirming that these two enzymes act on distinct glycosidic linkages. After reduction by sodium borohydride, identification of the amino acid and hexosamine contents in these compounds reveals that glucosamine was efficiently converted into glucosaminitol while muramic acid was not affected (Figure 2D). This observation demonstrates that GlcNAc is the reducing sugar and consequently that TM0633 is an *N*-acetylglucosaminidase.

Furthermore, TM0633 is inactive on peptide-free peptidoglycan chains generated by digestion of the polymers with an *N*-acetylmuramoyl-L-Ala amidase (data not shown), indicating that either presence of an amide-linked peptide to the lactoyl group of MurNAc or absence of a free carboxylic group at the C3 position on this sugar is required for its activity. A similar observation has been evidenced for lysozyme in which a weak activity was reported on peptidoglycan bearing D-Glu or A2pm residues (position 2 and 3 on the peptide) harboring an amide bond instead of a carboxylic group (Veiga et al. 2009; Figueiredo et al. 2012). Although TM0633 has not been assayed on low cross-linked peptidoglycan, the higher activities of *Enterococcus faecalis* AtlA and *Listeria monocytogenes* IspC on respectively low cross-linked peptidoglycan and fully N-acetylated peptidoglycan chains (Eckert et al. 2006;Ronholm et al. 2012) indicate that GH73 enzymes behave differently toward peptidoglycan.

Overall structure of TM0633

The crystal structure of TM0633 is composed of 136 residues and adopts the typical lysozyme α/β fold but only 115 residues, corresponding to the dominant α-lobe, could be built. This α-lobe is composed of 6 α-helices numbered 1 to 6, with a $\alpha3_{10}$ helix, which form a pseudo-barrel (Figure 3A). Surface loop (Leu55-Asp73) that connects α3 to α4 is disordered and could not be modeled. This region is predicted to fold as a β-hairpin motif consisting of two short adjacent antiparallel β-strands and corresponds to the β-lobe in GH23 GEWL (Weaver et al. 1995;Cole et al. 2008). In Auto, however, the corresponding loop region is well ordered and adopts a β-hairpin motif. In fact, this loop region is unusually long in Auto compared to the GH23 homologs and contributes to the formation of an extended lid above the substrate-binding groove (Figure 3C)(Bublitz et al. 2009). A lack of electron density of this region in TM0633 reflects high flexibility, as it was also observed for the FlgJ homolog (Hashimoto et al. 2009). Attempts to find a different crystal form with the aim to stabilize this region failed.

Unlike the Auto and FlgJ structural homologs, TM0633 is a dimer in solution as evidenced by gel filtration and dynamic light scattering (data not shown). TM0633 is also a dimer in the crystal, with a buried surface area of 935 $Å^2$ per subunit (16% of the total surface area). The two active sites are directed on opposite faces, a topology that could favor access to a large substrate (Figure 3B).

TM0633 contains an extended active site groove

Structural superimposition of TM0633 with FlgJ, and Auto, combined with sequence alignment, allowed us to localize the active site groove, 18 Å long and 7 Å wide, and the putative functional and catalytic residues in TM0633 (Figure 3C). Glu34 is strictly conserved among all members of the GH73 family, located at the tip of α2, where it superimposes with Glu122 in Auto, Glu185 in FlgJ and Glu73 in GH23 GEWL (Figure 3C/D).

Among the crystal structures available in the GH73 family, only the Auto structure shows a complete molecule with well-ordered α and β lobes. In Auto, Glu156 has been identified as the putative general base to fulfill a classical type 2 nucleophilic substitution with an inverting mechanism (Bublitz et al. 2009). However, Glu156, located within the β-hairpin, stands 13Å away from the general acid, and its implication in the catalytic mechanism requires large conformational changes of the β-hairpin (Figure 3C). While we could not model the corresponding Glu65 in the β-hairpin region of TM0633, conservation of Glu65 in the sequences of the GH73 family along with the high structural similarity between TM0633 and Auto (see the Experimental procedures) suggests Glu65 as the catalytic general base. In addition, three aromatic residues, Phe 47, Tyr 118 and Tyr 124, line the active site of TM0633 and superimpose well with their counterparts in Auto and FlgJ (Figure 3D).

A MPD (2-methyl-2,4-pentanediol) molecule was identified in the active site of TM0633. The MPD molecule interacts with the phenol ring of Tyr118, which is highly conserved in the GH73 family with only few exceptions where a Phe residue can be found (Figure 4). The corresponding tyrosine residue in AcmA and Atl(WM) is also essential for the catalytic activity but its functional role is preserved with a Phe or Trp substitution (Yokoi et al. 2008; Inagaki et al. 2009). In TM0633, Tyr118 faces Glu34 across the active site groove and corresponds to Tyr220 and Tyr147 in Auto and GEWL, respectively (Figure 3D). Hence, this tyrosine residue could facilitate substrate accommodation through stacking interaction into the catalytic site, a task that any aromatic residue present at this position would fulfill.

To identify the catalytic residues, we mutated Glu34 to Gln and Ala, Glu65 to Gln and the strictly conserved Tyr118 to Ala. Each mutant protein was incubated with *E. coli* peptidoglycan and the resulting digestion products were analyzed by HPLC. The chromatograms obtained with the wild-type enzyme and the Glu34Gln, Glu34Ala,

Glu65Gln and Tyr118Ala mutants pointed to a drastic decrease of more than 85%, in the enzyme activity (data not shown), indicating that those residues contribute to catalysis, consistent with previous findings on FlgJ, AcmA and Al(WM) (Maruyama et al. 2010; Yokoi et al. 2008; Inagaki et al. 2009).

Attempts to obtain structure of native TM0663 and the two Glu34Gln and Glu34Ala mutants in complex with several ligands, e.g. $(GlcNAc)_2$, $(GlcNAc)_4$, $(GlcNAc)_6$, TCT or $(GlcNAc\text{-}MurNAc)_2$, either by co-crystallization or by crystal soaking, yielded structures devoid of bound ligand. Considering the catalytic efficiency of TM0633 on large muropeptide substrates, consistent with its extended shallow groove as found in Auto, it is possible that these ligands are too small to efficiently bind into the active site of TM0633.

Bioinformatic analysis of the GH73 protein sequences

To explore the evolution of this large family, a phylogenic tree was built with more than 2800 sequences from the GH73 family. The tree clearly evidences five sequence clusters distributed among three different phyla. All sequences within cluster 3 belong to Bacteroidetes, clusters 2 and 4 harbor sequences from Firmicutes and clusters 1 and 5 host sequences from Proteobacteria including TM0633 (Figure 5, Table 2). In cluster 2, enzymes are highly modular, appended exclusively to CBM50 domains also known as LysM domains, which bind peptidoglycan. In cluster 4, many enzymes are also modular, but can harbor S-layer homology domains (SLH) (Engelhardt Peters 1998), unknown domains as well as muramoyl-amidase modules. In contrast, in clusters 1 and 5, which regroup sequences from Proteobacteria including TM0633, the GH73 catalytic modules do not carry additional domains. Comparative sequence analysis within each cluster, with particular attention to the various characterized GH73 enzymes, points to the molecular determinants of sequence divergence. Among the invariant residues lining the active site, Tyr118 is found in a YATD motif, which is a signature of the GH73 and GH23 families (Figure 4). This motif is conserved in all clusters except cluster 5, indicating enzyme peculiarities within this particular cluster. Another aromatic residue, Tyr124 in TM0633, is highly conserved among the GH73 sequences but can be substituted with a Trp in cluster 4. Tyr124 is also located in the vicinity of the active site and it corresponds to Tyr226 and Tyr169 in Auto and GEWL, respectively (Figures 3D and 4). This tyrosine has also been predicted to contribute to substrate positioning but its implication in catalysis has not been studied in the GH73 family. Regarding the catalytic mechanism prevailing in cluster 1, a Glu general base has been characterized in FlgJ and is strictly conserved (Figure 4). In cluster 2, which contains nine characterized enzymes, a conserved Glu general base, Glu156, has been characterized only in Auto (Figure 3C). However mutation of the equivalent Glu128 into Gln in AcmA did not significantly decrease enzyme

activity, suggesting that this cluster may employ two distinct catalytic mechanisms (Inagaki et al. 2009). Cluster 4 sequences are from the same phylum as those in cluster 2 (Firmicutes) but they do not show a conserved general base residue. This cluster hosts several characterized enzymes and mutation to Ala of a highly conserved Asp, Asp1275 in Atl(WM), does not affect enzyme activity, indicating that this residue is not essential for catalysis (Yokoi et al. 2008). This feature argues for catalysis via a substrate-assisted mechanism. In cluster 3, the only cluster hosting Bacteroidetes species, an Asp is highly conserved and is often found in a DDD motif (Figure 4). A catalytic role for this Asp has not yet been assessed, but this residue is most likely located in the β-hairpin region as observed for the corresponding general base residue in clusters 1 and 2 (Figure 4). Finally, in cluster 5, this aspartate is absent and no other putative general base could be identified. For members of this cluster, an inversion mechanism seems therefore unlikely and the conserved tyrosine, equivalent to Tyr118 in TM0633, might play a role in positioning the substrate to facilitate a substrate-assisted mechanism.

Molecular modeling

To support the bioinformatic analysis, we generated homology models of representative members from each cluster. The models for cluster 1 and 2 harbor a long β-hairpin domain as found in Auto (Figure 6). This long βhairpin region distinguishes GH73 enzymes from their lysozyme homologs by contributing to formation of a long extended binding groove (Bublitz et al. 2009). In contrast, models for clusters 3, 4 and 5, display a short β-hairpin domain, which could even be modeled as a small loop region for clusters 4 and 5 (Figure 6C/D and E). Comparative analysis of the models from each group clearly shows that the β-hairpin region harboring the putative Glu general base in clusters 1 and 2 is absent in clusters 4 and 5. In cluster 3, however, a highly conserved Asp could be located at the tip of this short β-hairpin. This Asp position, which matches that of Glu156 in Auto, obeys to the same distance criteria (~12Å) as the Glu general acid (Figure 6C).

Collectively, these models suggest that enzymes belonging to clusters 4 and 5 should follow a substrate-assisted mechanism while those from cluster 3 could utilize a "classical" inverting mechanism (Table 2). The models also suggest that enzymes from clusters 1 and 2 share an inverting catalytic mechanism and that large conformational changes should occur upon substrate binding. Such a conformational adaptation to functional requirements has been proposed for Ra-ChiC, a chitinase from the structurally related GH23 family (Arimori et al. 2013), in which loop L7, located between α5 and α6 and harboring the Asp226 general base at its tip, should undergo a large displacement to accommodate the $(NAG)_4$ substrate. In Ra-ChiC, the authors highlight the unique topology of the Asp226 general base acting from the roof of a tunnel-shaped active site (Figure 7). In fact, loop L7 in Ra-ChiC corresponds to a short loop

containing the YATD motif highly conserved in the GH73 sequences, except for cluster 5. Comparison of the Ra-ChiC structure and a TM0633 model harboring a complete β-hairpin reveals similar positions of the Glu general acids (Glu34 in TM0633 and Glu141 in Ra-ChiC). Similarly, the general base residues (Asp226 in Ra-ChiC and Glu65 in TM0633), which originate from distinct loop regions, are only separated by 3Å. Given the width of the active site groove in TM0633, it is unlikely that rearrangements of the βhairpin region will modify the shape of the active site into a tunnel. However based on the striking similarities in the general base position in Ra-ChiC and TM0633, we propose that enzymes from clusters 1 and 2 act through an inverting mechanism, a scenario that would require significant structural adaptation of the β-hairpin region.

Conclusions

The present study clearly emphasizes the diversity of the GH73 enzymes that cleave peptidoglycan. TM0633 appears to be the only characterized member active on intact *E coli* peptidoglycan and is not able to hydrolyze a peptide-depleted glycan chain. In contrast, AltA shows greater activity on low cross-linked peptidoglycan (Eckert et al. 2006) while IspC is more active on fully acetylated peptidoglycan (Ronholm et al. 2012). Moreover, LytG is an exo-acting *N*-acetylglucosaminidase, which requires the presence of a Mg ion for full activity (Horsburgh et al. 2003). Such peptide-sensitive activity in the GH73 family is probably linked to length variability of the functional loop associated to conformational plasticity of the active site groove. GH73 is the third GH family to host two putative distinct catalytic mechanisms as only the two other GH23 and GH97 families, out of the 132 GH families in the CAZy database, host inverting and retaining enzymes (Gloster et al. 2008; Blackburn Clarke 2001; Kuroki et al. 1999). The present study provides a solid framework to document further the sequence and biochemical diversity of the GH73 family.

Experimental procedures

Cloning and protein production

The coding sequence of TM0633 was amplified by polymerase chain reaction (PCR) from *Thermotoga maritima* MSB8 genomic DNA using pfx DNA polymerase (Invitrogen) and complementary gene-specific primers to which were appended sequences to facilitate Gateway™ cloning (Invitrogen). *attB* sequences were added at both ends as well as a ribosome binding site at the N-Terminal and a His_6tag coding sequence at the C-terminal. The amplicons were cloned by recombination in a Gateway™ pDEST14 vector (Invitrogen), carrying an ampicillin resistance gene. This vector was used to transform *E. coli* DH10B cells (Invitrogen). The DNA sequence of the gene insert was

confirmed, and the plasmid pDEST14-*TM0633* was transformed into *E. coli* Rosetta (Novagen) to confer additional tRNA genes and enhance protein overexpression.

For protein production, cells were cultivated to an OD_{600} of 0.7, at 37°C in LB medium, supplemented with 35 $\mu g \cdot ml^{-1}$ of kanamycin and chloramphenicol, then protein expression was induced using 1 mM IPTG. After overnight incubation at 25°C, cells were harvested by centrifugation and disrupted by sonication. The lysate was clarified by centrifugation and applied on a Ni-chelating Sepharose (5ml column) using an Äktaprime system (GE Healthcare Life Sciences). Bound proteins were eluted using a step gradient of imidazole (50 and 250 mM). The TM0633-containing fractions were pooled and concentrated prior to further purification by gel filtration on Superdex 200 (HR10/30 column, GE Healthcare Life Sciences) in CHES 10 mM pH 9, NaCl 150 mM, DTT 2 mM. Pure fractions of TM0633 were pooled and concentrated to 10-15 $mg \cdot ml^{-1}$ using a 10 kDa cut-off filtration unit (Millipore).

Site-directed mutagenesis

The Quick Change mutagenesis kit (Stratagene) was used to introduce mutations at positions 34, 65 and 118. Subsititutions of Glu34 by Gln or Ala, Glu65 by Gln, and Tyr118 by Ala were performed by PCR using pDest14-*TM0633* as a template. Pairs of oligonucleotides were designed for E34Q: 5'-CCAGTCCGCCCTCCAGACTGGCTGGGGAA-3' and 5'-TTCCCCAGCCAGTCTGGAGGGCGGACTGG-3'; E34A: 5'-CCAGTCCGCCCTCGCAACTGGCTGGGG-3' and 5'-CCCCAGCCAGTTGCGAGGGCGGACTGG-3'; E65Q: 5'-GCAGAAACGAAACAATTCGACGGTGTGAAGACG-3' and 5'-CGTCTTCACACCGTCGAATTGTTTCGTTTCTGC-3'; and for Y118A : 5'-CAAAGATATGGGGCCGCAACAGATCCAATGTACGCGG-3' and 5'-CCGCGTACATTGGATCTGTTGCGGCCCCATATCTTTG-3' as sense and anti-sense primers, respectively. The mutated genes were sequenced to ensure absence of unwanted alteration. The mutant proteins were expressed and purified using the same procedures as for the wild-type enzyme.

Crystallization, data collection and processing

Crystals of TM0633 were grown by vapor diffusion in sitting drops made of 100, 200 or 300 nL of protein at 10-15 $mg \cdot ml^{-1}$ and 100 nL of reservoir solution composed of monobasic ammonium phosphate 0.2 M, Tris-HCl 0.1 M, pH 8.5, and MPD 50% (v/v). Soaking with MurNAc and GlcNAc was performed by transferring a crystal into a reservoir solution containing 15 mM of each ligand. Crystals were flash-cooled at 120 K. The diffraction potency and quality of the crystals were assessed using a home source. A native data set was collected at 1.7Å resolution on the

ID29 beamline at the European Synchrotron Radiation Facility (ESRF) in Grenoble. The crystals belong to space group $C222_1$ with unit cell parameters: a= 46.3Å, b= 61.5Å, c= 101.1Å, corresponding to a solvent content of 45% assuming 1 molecule of TM0633 in the asymmetric unit (Matthews 1968). A sulfur-SAD experiment was conducted on the ID23-EH1 beamline (ESRF) at 100 K using an ADSC CCD detector. A highly redundant data set was collected on a single crystal to the maximum resolution limit of 2.06 Å. Data were integrated with XDS (Kabsch 2010) and intensities were scaled, merged and reduced with SCALA (CCP4 1994).

Phasing, model building and refinement

SHELXD was used for sulfur substructure determination (Schneider Sheldrick 2002). The top solution identified 6 sites with high occupancies. SHARP was used for positional refinement and phase calculations with data in the 15 to 2.2 Å resolution range (Bricogne et al. 2003). The phases were further improved by solvent flattening and histogram matching using SOLOMON and DM (CCP4 1994). A free-atom model was built into the electron-density maps using ARP/wARP(Langer et al. 2008) as implemented in SHARP. Subsequent auto-tracing of the amino-acid main and side chains using the warpNtrace procedure of ARP/wARP resulted in an initial model of high quality containing 111 residues out of 134 (Langer et al. 2008). Cycles of manual building in COOT (Emsley Cowtan 2004) and refinement in Phenix (Adams et al. 2002) against the native 1.7 Å-resolution data set, led to a final model containing 1008 protein atoms, 70 water molecules, a MPD molecule and 1 phosphate ion. Data collection and refinement statistics are reported in Table 1. The structure of the E34Q mutant was solved by molecular replacement with MOLREP (Vagin Teplyakov 1997), using the S-SAD refined structure as template (data not shown). Comparison of TM0633 with structural homologs gave an rsmd of 1.72 Å with FlgJ from *Sphingomonas sp.* for an alignment of 110 Cα□□□□ □□□□.. 1.84 Å when superimposed to Auto (110 Cα PDB: 3FI7) and 2.4 Å for 100 Cα of goose egg white lysozyme (GEWL, PDB: 153L).

Peptidoglycan hydrolase activity

Peptidoglycan was purified from the *E. coli* mutant strain BW25113 □*lpp ::Cm*^{r} that does not express the Lpp lipoprotein (Leulier et al. 2003). Cells were grown overnight at 37°C in 2YT medium (0.8 L). Cultures were stopped by rapid chilling to 0-4°C, harvested in cold conditions and washed with a cold NaCl 145.4 mM solution. The cells resuspended in 2 ml of this NaCl solution were injected with vigorous stirring into 40 ml of a hot (95-100°C) aqueous 4% SDS solution and boiled for 30 min. The mixture was kept overnight at room temperature and subsequently centrifuged at 200,000 × *g* for 30 min. The pellet containing peptidoglycan was washed several times with water to

remove traces of SDS. Final suspension of peptidoglycan was made in 5 ml of water and homogenized by brief sonication. Aliquots were hydrolyzed (16 h at 95°C in 6 M HCl) and analyzed with a Hitachi model 8800 amino acid analyzer. About 9 µmoles of peptidoglycan (expressed in terms of its muramic acid content), *i.e.* 9 mg of pure peptidoglycan, was prepared from 0.8 L of stationary-phase culture (OD_{600} = 4). 100-150 µg of peptidoglycan were incubated with 80 µg of purified TM0633 in 200 µl of 20 mM potassium phosphate buffer, pH 6.5, overnight at 37°C or 60°C. For comparison, 125 units of mutanolysin from *Streptomyces globisporus* (Sigma) were incubated with *E. coli* peptidoglycan in the same conditions. After centrifugation at 14,000 rpm, the supernatant containing soluble muropeptides was loaded on a HPLC Nucleosil 100 C18 column (3 µm, 4.6 × 250 mm, Altech-France) and eluted with a 0 to 20%(v/v) gradient of acetonitrile in 0.05 % TFA (v/v) applied for 100 min at 0.6 $ml.min^{-1}$. The muropeptide fractions were collected and analyzed. Aliquots of the samples were reduced in 0.25 M sodium borohydride for 30 min prior to analysis. Amino acid and hexosamine contents of muropeptides were determined after samples submission to acidic hydrolysis in 6 M HCl for 16 h at 96°C. The mass of muropeptides was determined by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF). For the mutant proteins TM0633-E34Q, TM0633-E34A, TM0633-E65Q and TM0633-Y118A, due to aggregation problems, 100-120 nmoles of peptidoglycan were incubated with only 17 µg of mutant proteins individually, for 3 h at 60°C or overnight at 37°C.

Phylogenetic analysis

Clustering within the GH73 family takes place in three steps: (1) sequences of the GH73 family were edited to isolate the catalytic domains to avoid interference from the presence or absence of additional modules (Henrissat Davies 2000). (2) The catalytic domains were then subjected to a multiple sequence alignment using MAFFT (Katoh Standley 2014). (3) The resulting alignment was used to infer an approximate-maximum-likelihood phylogenetic tree with FASTTREE 2.1 (Price et al. 2010), a program adapted to the analysis of large sets of sequences, using the Whelan Goldman model of amino acid evolution, the gamma option to rescale the branch lengths and compute a Gamma20-based likelihood, a total of four rounds of minimum-evolution moves, and options to make the maximum-likelihood nearest-neighbour interchanges more exhaustive. Consistency of the visually identified clusters was verified using a similar procedure (Stam et al. 2006).

Homology/Molecular modeling

3D models of 19 randomly selected GH73 sequences among the five clusters were generated using MODELLER (Sali Blundell 1993; Söding et al. 2005) with default parameters and, as templates, the structures of FlgJ

(PDB=2ZYC) and Auto (PDB=3FI7) (32 and 17% sequence identity with TM0633, respectively), selected using the TM-score from the HHpred server (http://toolkit.tuebingen.mpg.de/modeller).

References

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. 2002. PHENIX: building new software for automated crystallographic structure determination *Acta Crystallogr D Biol Crystallogr.* 58:1948-1954

Arimori T, Kawamoto N, Okazaki N, Nakazawa M, Miyatake K, Shinya S, Fukamizo T, Ueda M, Tamada T. 2013. Crystal structures of the catalytic domain of a novel glycohydrolase family 23 chitinase from Ralstonia sp. A-471 reveals a unique arrangement of the catalytic residues for inverting chitin hydrolysis. *J Biol Chem.* 288:18696-706

Blackburn NT, Clarke AJ. 2001. Identification of four families of peptidoglycan lytic transglycosylases. *J Mol Evol.* 52:78-84

Bokma E, van Koningsveld GA, Jeronimus-Stratingh M, Beintema JJ. 1997. Hevamine, a chitinase from the rubber tree Hevea brasiliensis, cleaves peptidoglycan between the C-1 of N-acetylglucosamine and C-4 of N-acetylmuramic acid and therefore is not a lysozyme. *FEBS Lett.* 411:161-163

Bourgeois I, Camiade E, Biswas R, Courtin P, Gibert L, Götz F, Chapot-Chartier MP, Pons JL, Pestel-Caron M. 2009. Characterization of AtlL, a bifunctional autolysin of Staphylococcus lugdunensis with N-acetylglucosaminidase and N-acetylmuramoyl-l-alanine amidase activities. *FEMS Microbiol Lett.* 290:105-113

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

Bublitz M, Polle L, Holland C, Heinz DW, Nimtz M, Schubert WD. 2009. Structural basis for autoinhibition and activation of Auto, a virulence-associated peptidoglycan hydrolase of Listeria monocytogenes. *Mol Microbiol.* 71:1509-1522

Buist G, Steen A, Kok J, Kuipers OP. 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol.* 68:838-847

Camiade E, Peltier J, Bourgeois I, Couture-Tosi E, Courtin P, Antunes A, Chapot-Chartier MP, Dupuy B, Pons JL. 2010. Characterization of Acp, a peptidoglycan hydrolase of Clostridium perfringens with N-acetylglucosaminidase activity that is implicated in cell separation and stress-induced autolysis. *J Bacteriol.* 192:2373-2384

CCP4 C. 1994. Collaborative Computational Project Number 4 *Acta Crystallogr.* D50:760-763

Cole C, Barber JD, Barton GJ. 2008. The Jpred 3 secondary structure prediction server *Nucleic Acids Res.* 36:W197-W201

Dhalluin A, Bourgeois I, Pestel-Caron M, Camiade E, Raux G, Courtin P, Chapot-Chartier M-P, Pons J-L. 2005. Acd, a peptidoglycan hydrolase of Clostridium difficile with N-acetylglucosaminidase activity. *Microbiology.* 151:2343-2351

Eckert C, Lecerf M, Dubost L, Arthur M, Mesnage S. 2006. Functional analysis of AtlA, the major N-acetylglucosaminidase of Enterococcus faecalis. *J Bacteriol.* 188:8513-8519

Eckert C, Magnet S, Mesnage S. 2007. The Enterococcus hirae Mur-2 enzyme displays N-acetylglucosaminidase activity. *FEBS Lett.* 581:693-696

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 60:2126-2132

Engelhardt H, Peters J. 1998. Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions. *J Struct Biol.* 124:276-302

Figueiredo TA, Sobral RG, Ludovice AM, Almeida JM, Bui NK, Vollmer W, de Lencastre H, Tomasz A. 2012. Identification of genetic determinants and enzymes involved with the amidation of glutamic acid residues in the peptidoglycan of Staphylococcus aureus. *PLoS Pathog.* 8:e1002508

Foster SJ. 1995. Molecular characterization and functional analysis of the major autolysin of Staphylococcus aureus 8325/4. *J Bacteriol.* 177:5723-5725

García P, González MP, García E, López R, García JL. 1999. LytB, a novel pneumococcal murein hydrolase essential for cell separation *Molecular microbiology.* 31:1275-1277

Gloster TM, Turkenburg JP, Potts JR, Henrissat B, Davies GJ. 2008. Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora. *Chem Biol.* 15:1058-1067

Hashimoto W, Ochiai A, Momma K, Itoh T, Mikami B, Maruyama Y, Murata K. 2009. Crystal structure of the glycosidase family 73 peptidoglycan hydrolase FlgJ. *Biochem Biophys Res Commun.* 381:16-21

Henrissat B, Davies GJ. 2000. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol.* 124:1515-1519

Horsburgh GJ, Atrih A, Williamson MP, Foster SJ. 2003. LytG of Bacillus subtilis is a novel peptidoglycan hydrolase: the major active glucosaminidase. *Biochemistry.* 42:257-264

Huard C. 2003. Characterization of AcmB, an N-acetylglucosaminidase autolysin from Lactococcus lactis *Microbiology.* 149:695-705

Huard C, Miranda G, Redko Y, Wessner F, Foster SJ, Chapot-Chartier M-P. 2004. Analysis of the peptidoglycan hydrolase complement of Lactococcus lactis: identification of a third N-acetylglucosaminidase, AcmC. *Appl Environ Microbiol.* 70:3493-3499

Huard C, Miranda G, Wessner F, Bolotin A, Hansen J, Foster SJ, Chapot-Chartier MP. 2003. Characterization of AcmB, an N-acetylglucosaminidase autolysin from Lactococcus lactis. *Microbiology.* 149:695-705

Inagaki N, Iguchi A, Yokoyama T, Yokoi KJ, Ono Y, Yamakawa A, Taketo A, Kodaira K. 2009. Molecular properties of the glucosaminidase AcmA from Lactococcus lactis MG1363: mutational and biochemical analyses. *Gene.* 447:61-71

Kabsch W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr.* 66:133-144Katoh K, Standley DM. 2014. MAFFT: iterative refinement and additional methods. *Methods Mol Biol.* 1079:131-146

Kuroki R, Weaver LH, Matthews BW. 1999. Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. *Proc Natl Acad Sci U S A.* 96:8949-8954

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

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

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42:D490-D495

Mark BL, Vocadlo DJ, Knapp S, Triggs-Raine BL, Withers SG, James MN. 2001. Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. *J Biol Chem.* 276:10330-10337

Maruyama Y, Ochiai A, Itoh T, Mikami B, Hashimoto W, Murata K. 2010. Mutational studies of the peptidoglycan hydrolase FlgJ of Sphingomonas sp. strain A1. *J Basic Microbiol.* 50:311-317

Matthews BW. 1968. Solvent content of protein crystals *J Mol Biol.* 33:491-497

Mesnage S, Dellarole M, Baxter NJ, Rouget JB, Dimitrov JD, Wang N, Fujimoto Y, Hounslow AM, Lacroix-Desmazes S, Fukase K, Foster SJ, Williamson MP. 2014. Molecular basis for bacterial peptidoglycan recognition by LysM domains. *Nat Commun.* 5: 4269

Oshida T, Sugai M, Komatsuzawa H, Hong YM, Suginaka H, Tomasz A. 1995. A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-alanine amidase domain and an endo-beta-N-acetylglucosaminidase domain: cloning, sequence analysis, and characterization. *Proc Natl Acad Sci U S A.* 92:285-289

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One.* 5:, e9490

Rashid MH, Mori M, Sekiguchi J. 1995. Glucosaminidase of Bacillus subtilis: cloning, regulation, primary structure and biochemical characterization. *Microbiology.* 141:2391-2404

Rolain T, Bernard E, Courtin P, Bron PA, Kleerebezem M, Chapot-Chartier MP, Hols P. 2012. Identification of key peptidoglycan hydrolases for morphogenesis, autolysis, and peptidoglycan composition of Lactobacillus plantarum WCFS1. *Microb Cell Fact.* 11: 137

Ronholm J, Wang L, Hayashi I, Sugai M, Zhang Z, Cao X, Lin M. 2012. The Listeria monocytogenes serotype 4b autolysin IspC has N-acetylglucosaminidase activity. *Glycobiology.* 22:1311-1320

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 234:779-815

Schneider TR, Sheldrick GM. 2002. Substructure solution with SHELXD *Acta Crystallographica Section D-Biological Crystallography.* 58:1772-1779

Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33:W244-W248

Stam MR, Danchin EG, Rancurel C, Coutinho PM, Henrissat B. 2006. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. *Protein Eng Des Sel.* 19:555-562

Typas A, Banzhaf M, Gross CA, Vollmer W. 2012. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Microbiol.* 10:123-136

Vagin AA, Teplyakov A. 1997. MOLREP: an autmated program for molecular replacement *J. Appl. Cryst.* 30:1022-1025

Veiga P, Erkelenz M, Bernard E, Courtin P, Kulakauskas S, Chapot-Chartier MP. 2009. Identification of the asparagine synthase responsible for D-Asp amidation in the Lactococcus lactis peptidoglycan interpeptide crossbridge. *J Bacteriol.* 191:3752-3757

Vocadlo DJ, Davies GJ, Laine R, Withers SG. 2001. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature.* 412:835-838

Vollmer W, Blanot D, de Pedro MA. 2008. Peptidoglycan structure and architecture. *FEMS Microbiol Rev.* 32:149-167

Weaver LH, Grütter MG, Matthews BW. 1995. The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the "goose-type" lysozymes lack a catalytic aspartate residue. *J Mol Biol.* 245:54-68

Yokoi KJ, Sugahara K, Iguchi A, Nishitani G, Ikeda M, Shimada T, Inagaki N, Yamakawa A, Taketo A, Kodaira K. 2008. Molecular properties of the putative autolysin Atl(WM) encoded by Staphylococcus warneri M: mutational and biochemical analyses of the amidase and glucosaminidase domains. *Gene.* 416:66-76

Acknowledgments- We thank Didier Blanot (IBBMC, Orsay) for mass spectrometry experiments and amino-acid content analyses. We thank the ID29 staff at ESRF (Grenoble) for providing beamtime and support; Bernard Henrissat and Pascale Marchot (AFMB, Marseille) for fruitful discussion and careful reading of the manuscript. This work was supported in part by the CNRS and by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01.

FIGURE LEGENDS

FIGURE 1: Schematic representation of the proposed catalytic mechanisms for the GH73 enzyme family A- The inverting mechanism involves the assistance of two acid carboxylic (Glu or Asp). The general acid protonates the glycosidic oxygen, favoring the aglycon departure, while a water molecule activated by the general base makes a nucleophile attack on the sugar anomeric carbon. B- the neighboring group participation mechanism uses an acid/base residue to protonate the glycosidic oxygen. The 2-acetamido group of the substrate plays the role a nucleophile to generate an oxazolinium ion intermediate. The cyclic intermediate is then hydrolyzed by the base-catalyzed attack of water at the anomeric center.

FIGURE 2: TM0633 has *N*-acetylglucosaminidase activity. A- Schematic representation of the peptidoglycan hydrolysis linkage targeted by TM0633 and of the main monomer muropeptide released by this enzyme, which is designated as A1/A2 (two alpha and beta anomer forms of this compound) in panels B and D. RP-HPLC profile of soluble muropeptides obtained after peptidoglycan hydrolysis by (B) TM0633 or (C) mutanolysin. (D) Characterization of the released peptidoglycan fragments by MALDI-TOF mass spectrometry. The *m/z* values, 940.6 (A1, A2) and 1862.1 (A3, A4) respectively, of $(M+H)^+$ ions indicate that they are monomer (A1, A2) and dimer (A3, A4) forms of the disaccharide-tetrapeptide (GlcNAc-MurNAc-L-alanyl-γ-D-glutamyl-*meso*-diaminopimelyl-D-alanine).

FIGURE 3: **Structure of TM0633**. (A) Overall structure of TM0633 (rainbow color) in complex with a MPD molecule (yellow backbone) and a phosphate molecule (orange), showing the dominant α-helical domain. Helices are numbered α1 to α6 and dashes indicate missing amino acids in the β-hairpin. (B) In the TM0633 dimer; the active site of each monomer is indicated by an arrow. (C) Close up view of the overlay of TM0633 (yellow), Auto (white, PDB code: 3FI7) and FlgJ (dark grey, PDB code: 2ZYC). The general acid in TM0633, Glu34, is labeled in red; the equivalent general acid residues in Auto (Glu122) and FlgJ (Glu185) are in black and grey, respectively. The general base located in the β-hairpin region of Auto, Glu156, is shown above the general acid residue. The shape of the active site groove in Auto is highlighted with a partial transparent light blue surface. (D) Overlay of TM0633 (white cartoon and yellow side chains), Auto (white side chains) and GEWL (PDB code: 154L, purple side chains) showing the conserved Glu general acid and aromatic residues in the active site.

FIGURE 4: Clustering and sequence alignment of 20 GH73 sequences. Sequence alignment with 5 representative sequences from each cluster. The conserved general acid residue (corresponding to Glu34 in TM0633) within α2 is indicated with a pink circle. The putative general base (corresponding to Glu65 in TM0633) located in the β-hairpin region (blue background) is indicated with a pink circle for clusters 1 and 2 and with a red circle for cluster 3. The conserved aromatic residues (corresponding to Tyr118 and Tyr124 in TM0633) are indicated with yellow circle. Secondary structure elements of TM0633 are drawn at the top of the sequences; the predicted β-sheets are drawn with a grey arrow.

FIGURE 5: Phylogenetic tree of the GH73 sequences. The branches for cluster 1, 2, 3, 4 and 5 are shown in four shades of grey. Within each cluster, the sequences characterized by a phylum, and the characterized enzymes are labeled and highlighted with a black circle; the structurally characterized GH73 enzymes are indicated with a grey rectangle.

FIGURE 6: Structural models of the five identified clusters

Overlay of the available crystal structures with structural models belonging to clusters 1 to 5 viewed in a similar orientation. When unambiguously identified from the sequence alignment the general acid and the general base side chains are shown in white. Overall view of (A) TM0633 and FlgJ (blue) with a modeled β-hairpin, belonging to cluster 1, and two representative models from *Thermotoga petrophila* and *Frateuria aurantia*, (B) Auto from cluster 2 (yellow) with three models of *Lactococcus lactis* AcmA, *Enterococcus feacalis* AltA and *Enterococcus hirae* Mur2, (C) four models (orange) of the GH73 domain from *Croceibacter atlanticus, Bacteroides fragilis, Gramella forsetii* and *Flavobacteriia bacterium* found in cluster 3, (D) four models (green) of *Clostridium. difficile* Acd ,*Stretococcus pneumoniae* LytB, *Bacillus subtilis* LytD and *Staphylococcus aureus* Atl belonging to cluster 4, (E) four models (red) of the GH73 domain from *Campylobacter jejuni*, *Saccharophagus degradans*, *Thermotoga Thermarum* and *Vibrio fisheri* in cluster 5, (F) superimposition of the functional loop from a member of each phylogenetic cluster, TM0633 (in blue with modeled β-hairpin), Auto (yellow) and three models from *Saccharophagus degradans* (red), *Gramella forsetii* (orange) and Acd from *Clostridium difficile* (green). A general base residue is only present for Auto, TM0633 and the model from *Gramella forsetii*.

FIGURE 7: Comparison of the general base residue location in TM0633 and Ra-ChiC. Overlay of TM0633 harboring a modeled loop (white) and GH23 Ra-ChiC (PDB code: 3W6B, pink) with the conserved general acid

Glu34 from TM0633 and Glu141 from Ra-ChiC shown in liquorice. In Ra-ChiC, the general base Asp226 is enclosed in loop L7 above the active site while in TM0633 the corresponding Glu65 general base originates from the facing β-hairpin.

(Rolain et al. 2012)

Figure 1

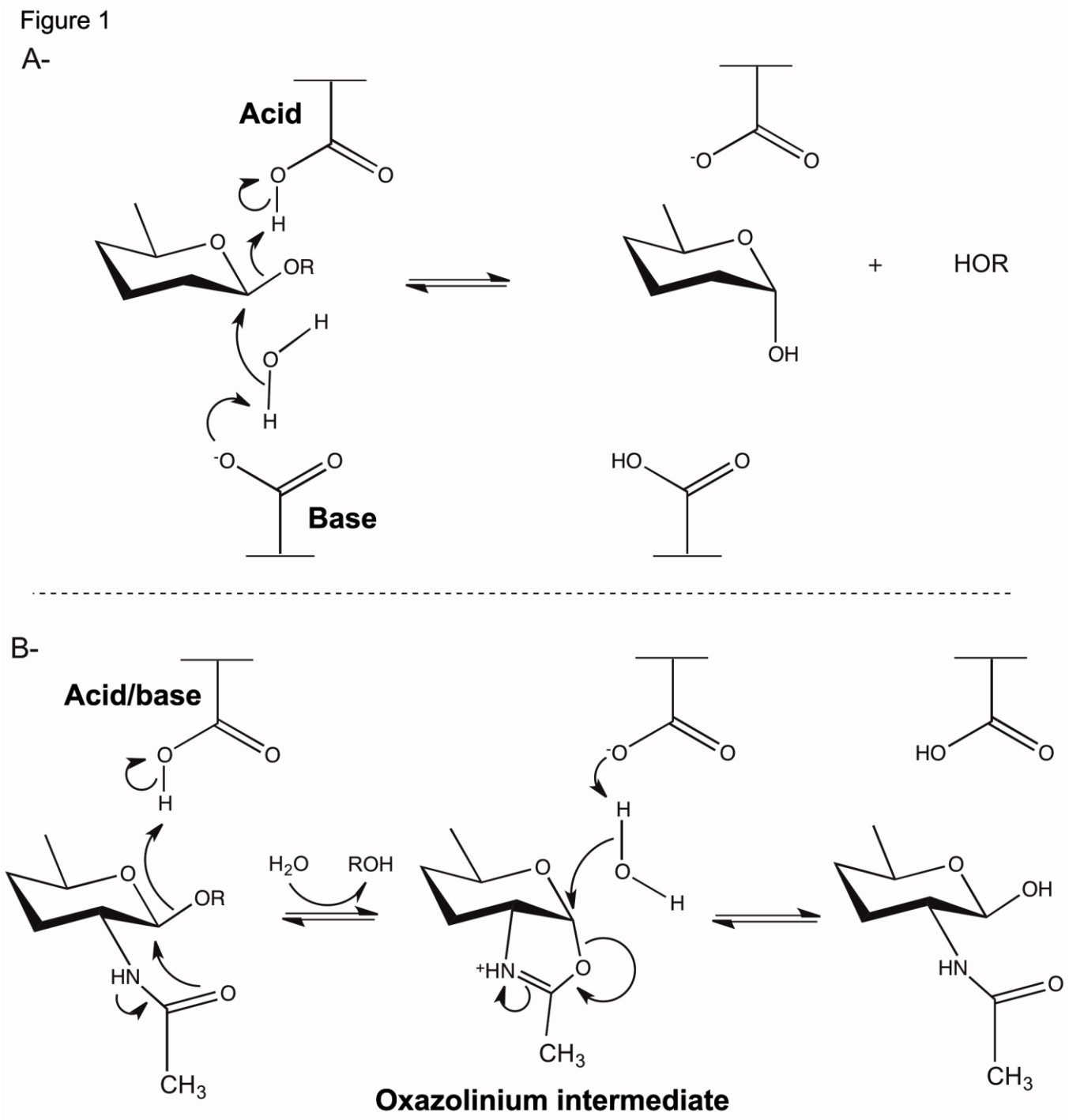
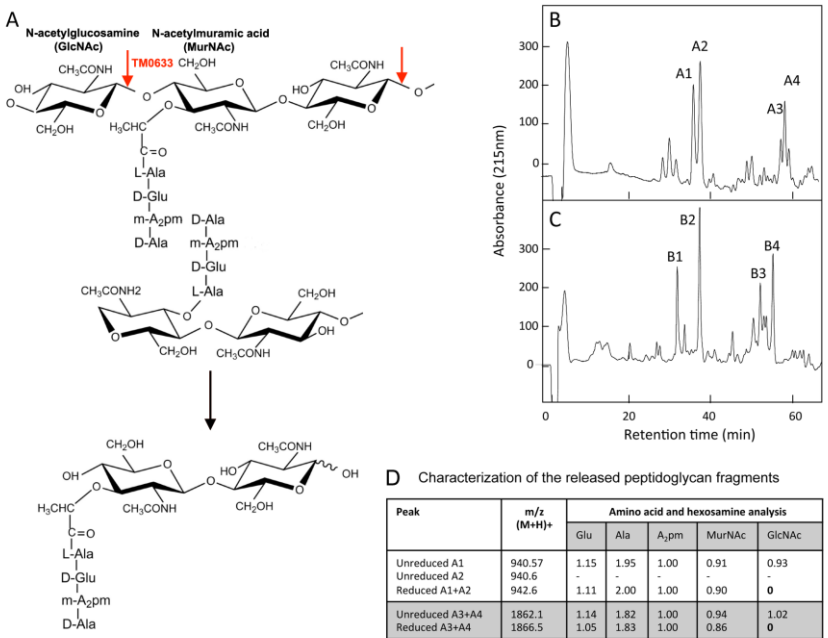


Figure 2



D Characterization of the released peptidoglycan fragments

Peak	m/z (M+H)+	Amino acid and hexosamine analysis				
		Glu	Ala	A_2pm	MurNAc	GlcNAc
Unreduced A1	940.57	1.15	1.95	1.00	0.91	0.93
Unreduced A2	940.6	-	-	-	-	-
Reduced A1+A2	942.6	1.11	2.00	1.00	0.90	**0**
Unreduced A3+A4	1862.1	1.14	1.82	1.00	0.94	1.02
Reduced A3+A4	1866.5	1.05	1.83	1.00	0.86	**0**
Unreduced B1+B2	940.9	1.11	1.78	1.00	0.89	0.96
Reduced B1+B2	942.7	1.08	1.89	1.00	**0**	0.91
Unreduced B3+B4	1862.5	-	-	-	-	-

Figure 3

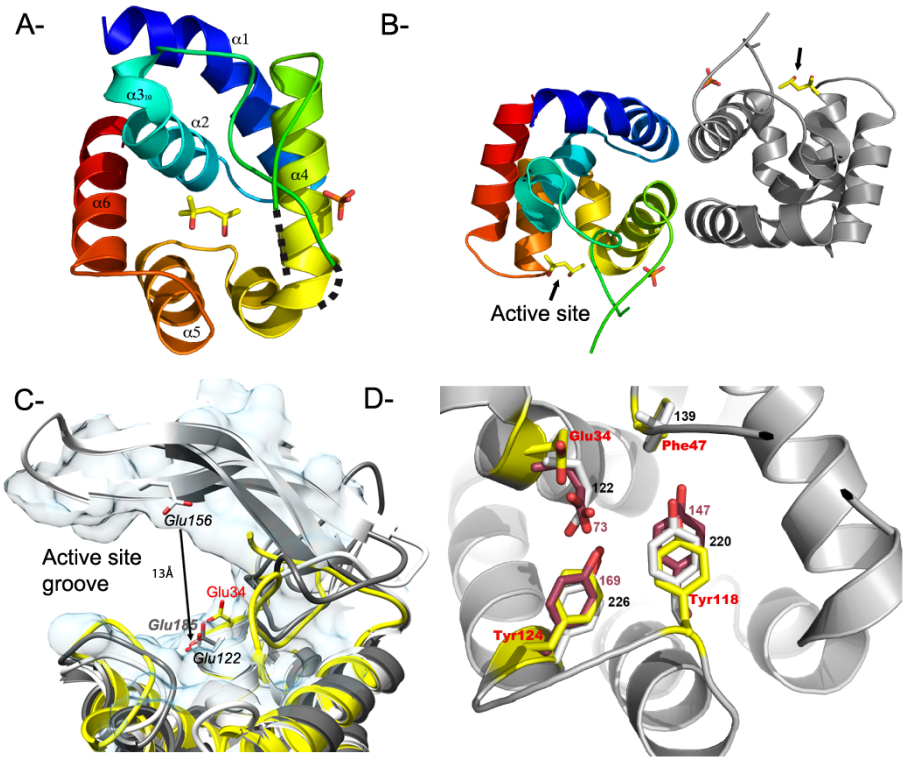


Figure 4

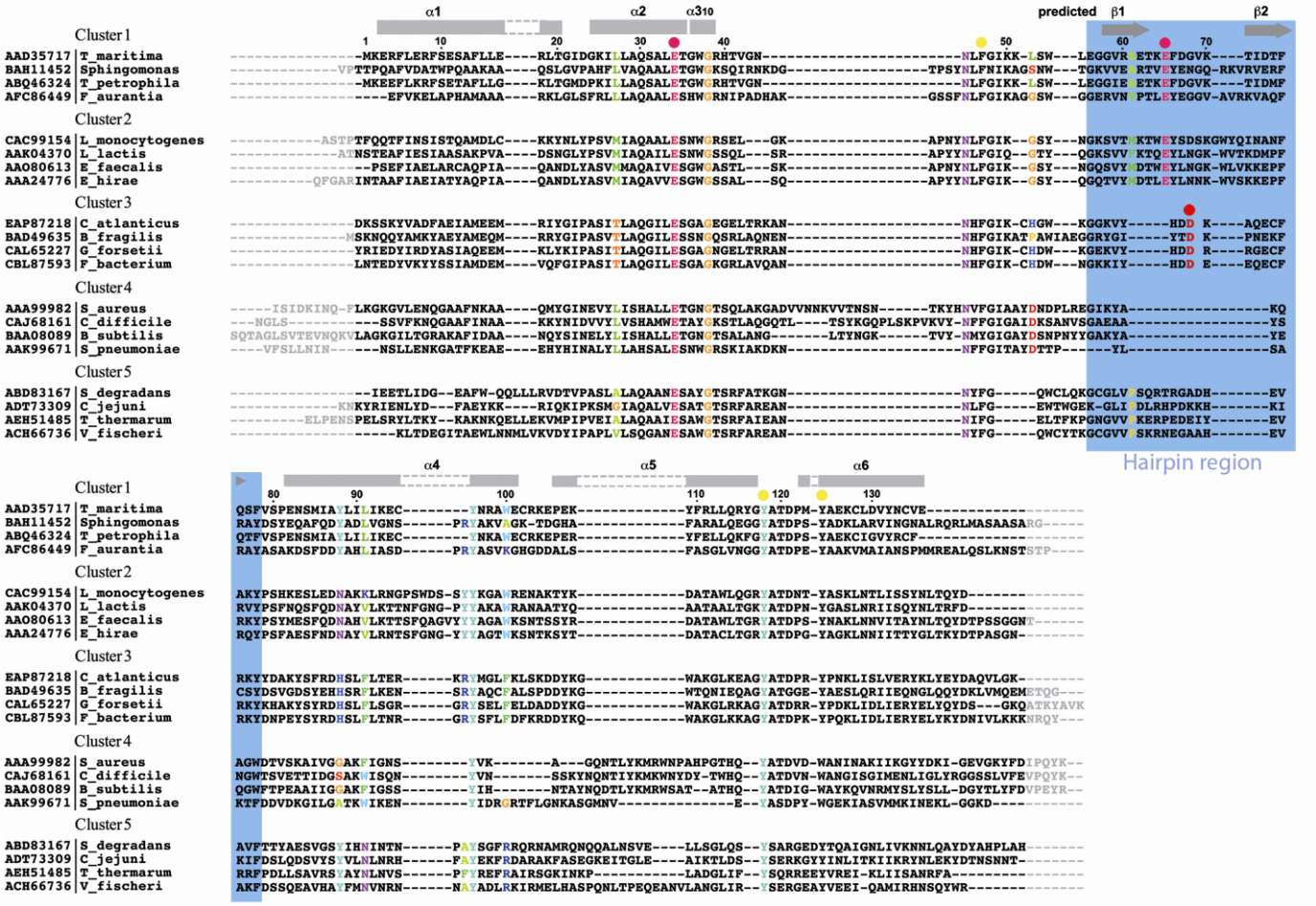


Figure 5

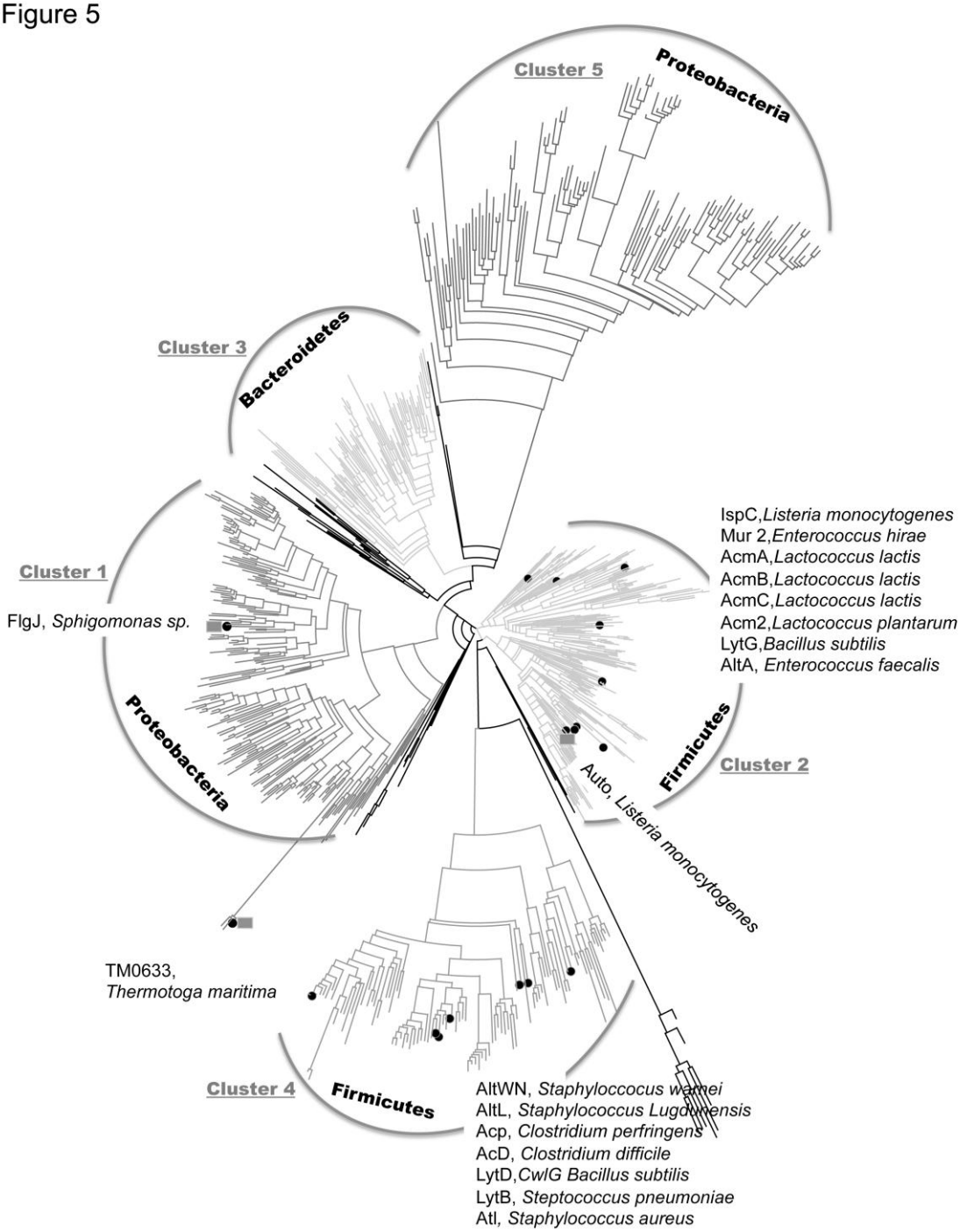


Figure 6

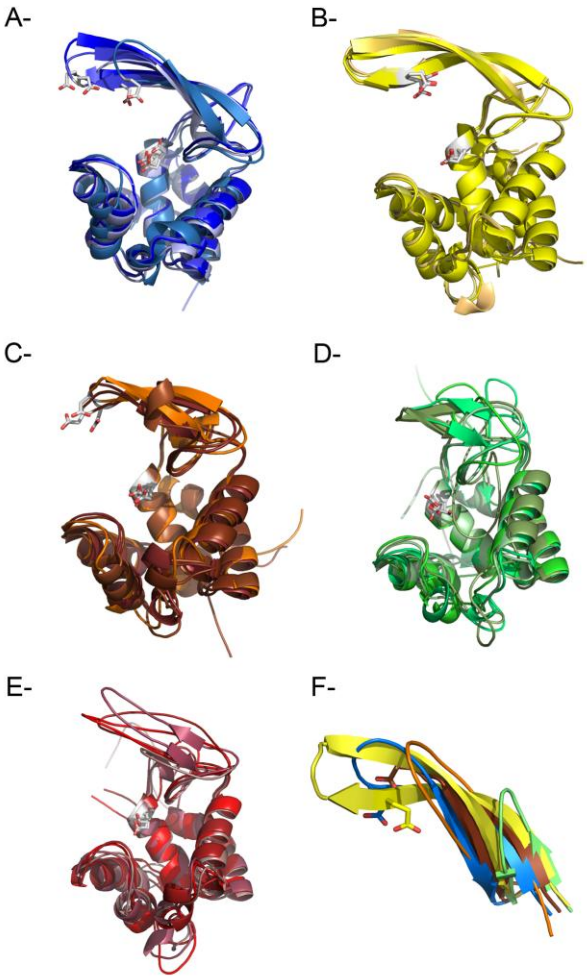


Figure 7

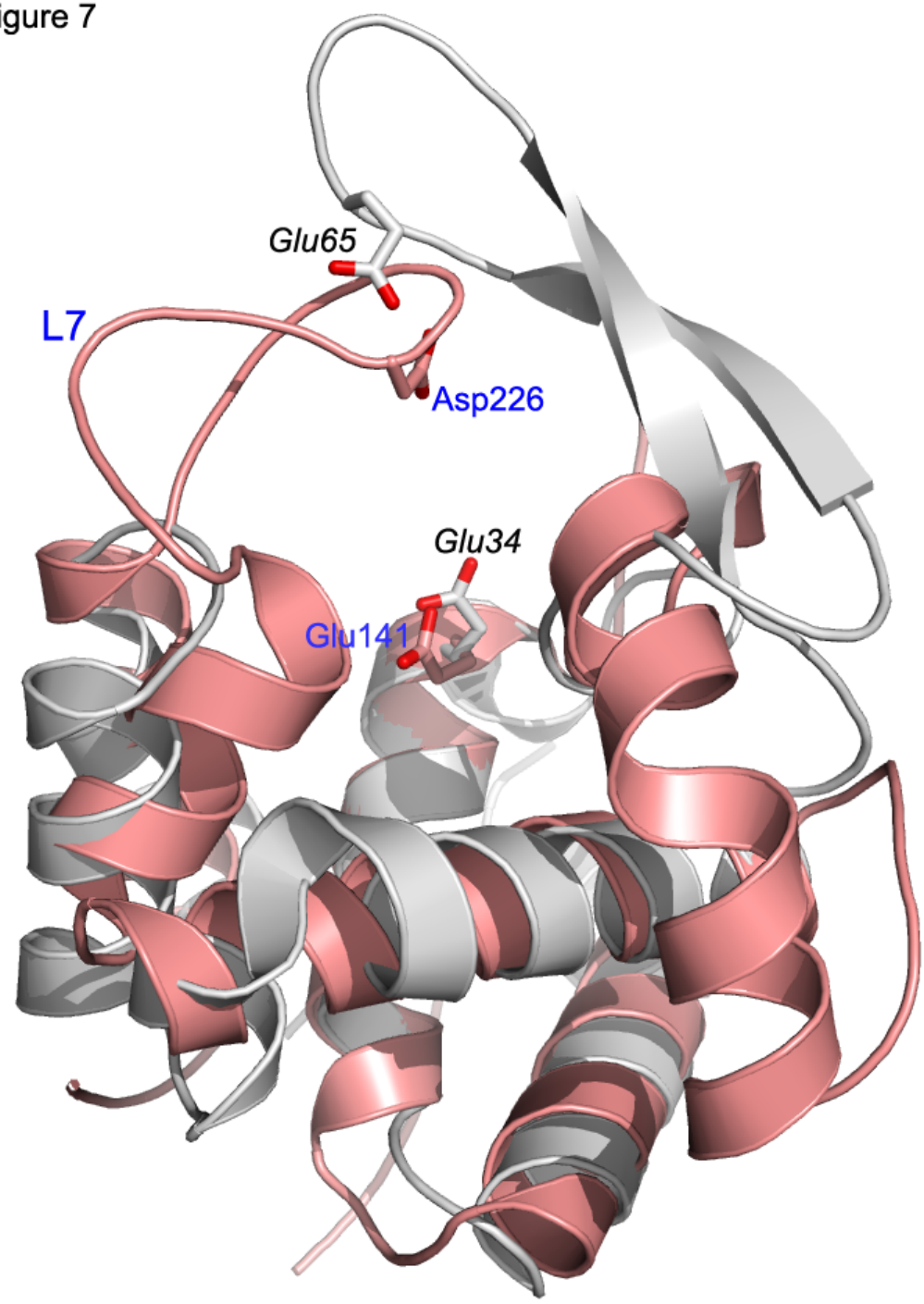


Table1: Data collection and refinement statistics

Crystal parameters	Native	S-SAD
Space group	$C222_1$	$C222_1$
Cell parameters (Å)	a=46.56, b=61.19, c=101.20	a=46.28 b=62.14 c=101.68
Data quality		
Wavelength (Å)	0.9537 (ID29)	2.06 (ID23-EH1)
Resolution of data (Å)	50.6-1.7	15-2.2
Unique reflections	16084	14269
R_{merge} (outer shell)[a]	0.08 (0.58)	0.04 (0.22)
Mean $I/\sigma I$ (outer shell)	13.5 (2.3)	32 (5.1)
Completeness (outer shell) %	98.6 (91)	99 (95.5)
Multiplicity (outer shell)	7.2 (4.9)	31 (18.7)
Refinement		
Protein atoms	10008	-
Solvent waters	70	-
R_{cryst}[b]	0.167	-
R_{free}[c]	0.193	-
Root mean square deviation 1-2 bonds (Å)	0.006	-
Root mean square deviation 1-3 angles (degrees)	0.960	-

[a] $R_{merge} = (\Sigma_{hkl} \Sigma_i \left| I_{hkl} - (I_{hkl}) \right| / \Sigma_{hkl} \Sigma_i [I_{hkl}])$.
[b] $R_{cryst} = \Sigma_{hkl} |Fo| - |Fc| / \Sigma_{hkl} |Fo|$
[c] R_{free} is calculated for randomly selected reflections excluded from refinement

Table 2: Clustering of members of the GH73 enzyme family.

GH73 SUBFAMILY AND CLUSTERS	GH73 ENZYMES	PHYLUM	CATALYTIC RESIDUES	PROPOSED CATALYTIC MECHANISM
CLUSTER 1	**TM0633** (*Thermotoga maritima*)* (this work) *s*	**Proteobacteria**	**Glu** general acid **Glu** general base	Classic inversion mechanism
	FljG (*Sphingomonas sp.)** (Hashimoto et al. 2009)*s*			
	Thermotoga petrofila			
	Frateuria aurantia			
CLUSTER 2	**Auto** (*Listeria monocytogenes*)*(Bublitz et al. 2009) *s*	**Bacteroidetes**	**Glu** general acid **Glu** general base	Classic inversion mechanism + Substrate assisted mechanism (AcmA)
	Mur2 (*Enterococcus hirae*)(Eckert et al. 2007)			
	AcmA (*Lactococcus lactis*)* (Inagaki et al. 2009)			
	AcmB (*Lactotcoccus lactis*)(Huard 2003)			
	AltA (*Enterococcus. faecalis*)(Eckert et al. 2006)			
	AcmC (*Lactococcus lactis)* (Huard et al. 2004)			
	Acm2 (*Lactobacillus plantarum*) (Rolain et al. 2012)			
	LytG (*Bacillus subtilis*) (Horsburgh et al. 2003)			
	IspC (*Listeria monocytogenes*) (Ronholm et al. 2012)			
CLUSTER 3	*C. atlanticus*	**Firmicutes**	**Glu** general acid **Asp** general base	Classic inversion mechanism
	F. bacterium			
	P. heparinus			
	G. forseti			
CLUSTER 4	**AtlWM** (*Staphyloccocus warnei*)* (Yokoi et al. 2008)	**Firmicutes**	**Glu** general acid No putative general base	Substrate assisted mechanism
	AltL (*Staphylococcus Lugdunensis*) (Bourgeois et al. 2009)			
	Acp (*Clostridium perfringens*) (Camiade et al. 2010)			
	AcD (*Clostridium difficile*) (Dhalluin et al. 2005)			
	LytD;CwlG (*Bacillus subtilis*) (Rashid et al. 1995)			
	LytB (*Streptococcus pneumoniae*) (García et al. 1999)			
	Atl (*Staphylococcus aureus*) (Foster 1995;Oshida et al. 1995)			
CLUSTER 5	*Saccharophagus degradans*	**Proteobacteria**	**Glu** general lacid No putative general base	Substrate assisted mechanism
	Campylobacter jejuni			
	Thermus thermophilus			
	Vibrio fisheri			

(*) Characterized catalytic residues (s) 3D structure available