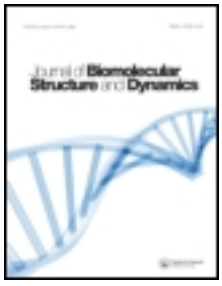

Journal of Biomolecular Structure and Dynamics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tbsd20

Crystal structure of the CN-hydrolase SA0302 from the pathogenic bacterium Staphylococcus aureus belonging to the Nit and NitFhit Branch of the nitrilase superfamily

Roni D. Gordon [a], Wei Qiu [a], Vladimir Romanov [a], Kim Lam [a], Maria Soloveychik [a], Diana Benetteraj [a], Kevin P. Battaile [b], Yuri N. Chirgadze [c], Emil F. Pai [a d] & Nickolay Y. Chirgadze [a e]

[a] Campbell Cancer Research Institute, Ontario Cancer Institute, Princess Margaret Hospital, University Health Network, Toronto, Ontario, M5G 2C4, Canada

[b] Hauptman-Woodward Medical Research Institute, IMCA-CAT, Advanced Photon Source, Argonne National Laboratory, Argonne, IL, 60439, USA

[c] Institute of Protein Research, Russian Academy of Sciences, 142292, Pushchino, Moscow Region, Russia

[d] Departments of Biochemistry, Molecular Genetics, and Medical Biophysics, University of Toronto, Toronto, Ontario, M5S 1A8, Canada

[e] Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, M5S 1A8, Canada

Version of record first published: 02 Oct 2012.

To cite this article: Roni D. Gordon, Wei Qiu, Vladimir Romanov, Kim Lam, Maria Soloveychik, Diana Benetteraj, Kevin P. Battaile, Yuri N. Chirgadze, Emil F. Pai & Nickolay Y. Chirgadze (2012): Crystal structure of the CN-hydrolase SA0302 from the pathogenic bacterium Staphylococcus aureus belonging to the Nit and NitFhit Branch of the nitrilase superfamily, Journal of Biomolecular Structure and Dynamics, DOI:10.1080/07391102.2012.719111

To link to this article: http://dx.doi.org/10.1080/07391102.2012.719111

Journal of Biomolecular Structure and Dynamics
iFirst, 2012, 1–9

Crystal structure of the CN-hydrolase SA0302 from the pathogenic bacterium *Staphylococcus aureus* belonging to the Nit and NitFhit Branch of the nitrilase superfamily

Roni D. Gordon[a], Wei Qiu[a], Vladimir Romanov[a], Kim Lam[a], Maria Soloveychik[a], Diana Benetteraj[a], Kevin P. Battaile[b], Yuri N. Chirgadze[c], Emil F. Pai[a,d] and Nickolay Y. Chirgadze[a,e]*

[a]*Campbell Cancer Research Institute, Ontario Cancer Institute, Princess Margaret Hospital, University Health Network, Toronto, Ontario M5G 2C4, Canada;* [b]*Hauptman–Woodward Medical Research Institute, IMCA-CAT, Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA;* [c]*Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia;* [d]*Departments of Biochemistry, Molecular Genetics, and Medical Biophysics, University of Toronto, Toronto, Ontario M5S 1A8, Canada;* [e]*Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada*

Communicated by Anna Panchenko

(*Received 22 March 2012; final version received 2 August 2012*)

The nitrilases include a variety of enzymes with functional specificities of nitrilase, amidase, and hydrolase reactions. The crystal structure of the uncharacterized protein SA0302 from the pathogenic microorganism *Staphylococcus aureus* is solved at 1.7 Å resolution. The protein contains 261 amino acids and presents a four-layer αββα sandwich with a chain topology similar to that of a few known CN-hydrolase folds. In the crystal, the proteins are arranged as dimers whose monomers are related by a pseudo twofold rotation symmetry axis. Analysis of the sequences and structures of CN-hydrolases with known 3D structures shows that SA0302 definitely is a member of Branch 10 (Nit and NitFhit) of the nitrilase superfamily. Enzyme activities and substrate specificities of members of this branch are not yet characterized, in contrast to those of the members of Branches 1–9. Although the sequence identities between Branch 10 members are rather low, less than 30%, five conserved regions are common in this subfamily. Three of them contain functionally important catalytic residues, and the two other newly characterized ones are associated with crucial intramolecular and intermolecular interactions. Sequence homology of the area near the active site shows clearly that the catalytic triad of SA0302 is Glu41-Lys110-Cys146. We suggest also that the active site includes a fourth residue, the closely located Glu119. Despite an extensive similarity with other Nit-family structural folds, SA0302 displays an important difference. Protein loop 111–122, which follows the catalytic Lys110, is reduced to half the number of amino acids found in other Nit-family members. This leaves the active site fully accessible to solvent and substrates. We have identified conservative sequence motifs around the three core catalytic residues, which are inherent solely to Branch 10 of the nitrilase superfamily. On the basis of these new sequence fingerprints, 10 previously uncharacterized proteins also could be assigned to this hydrolase subfamily.

An animated interactive 3D complement (I3DC) is available in Proteopedia at http://proteopedia.org/w/Journal:JBSD:19

Keywords: nitrilase; CN-hydrolase; Nit-domain; thioredoxin-like activity; protein family

Introduction

The nitrilase superfamily, which has in total over 180 known members, includes a variety of thiol amidase enzymes involved in biosynthesis in plants, animals, fungi, and prokaryotes (Pace & Brenner, 2001). All members of this superfamily have conserved the active site residues Glu-Lys-Cys, believed to form a catalytic triad. Consensus sequences flanking the catalytic residues supply the conserved motifs distinctive for all of the branches (Pace & Brenner, 2001). The superfamily can be classified into 13 branches, nine of which have known or predicted specificity for nitrilase, amidase, and CN-hydrolase reactions. Although the entire family has been considered "nitrilase-related," only members of Branch 1 have demonstrated nitrilase activity. The remaining branches include enzymes with amidase or amide-condensation activity including aliphatic amidase, amino-terminal amidase, biotinidase, β-ureidopropionase,

*Corresponding author. Email: nchirgad@uhnresearch.ca

ISSN 0739-1102 print/ISSN 1538-0254 online

http://dx.doi.org/10.1080/07391102.2012.719111
http://www.tandfonline.com

carbamylase, prokaryotic and eukaryotic NAD-synthetase, and apolipoprotein N-acyltransferase (Brenner, 2002). The domain structure of members of all branches includes a common Nit-domain with the catalytic triad; some members can have up to three additional domains. The common structural nitrilase-like domain has a unique αββα four-layer sandwich protein fold, which in most cases forms an eight-layer dimer. Such dimers have been observed both in solution and in the crystalline state (Bork & Koonin, 1994).

At present, the structures of about 30 proteins that belong to the nitrilase superfamily are deposited in the Protein Data Bank (PDB) (see, e.g. Andrade, Amin Karmali, Carrondo, & Frazao, 2007; Barglow et al., 2008; Chin et al., 2007; Chiu-Lien et al., 2007; Kimani, Agarkar, Cowan, Sayed, & Sewell, 2007; Kumaran et al., 2003; Nakai et al., 2000; Nel, Tuffin, Sewell, & Cowan, 2011; Pace et al., 2000; Sakai, Tajika, Yao, Watanabe, & Tanaka, 2004). Most of the corresponding papers describe the protein structure and discuss its catalytic mechanism. All these structures have a common nitrilase domain with a similar main chain topology differing only in small details. Essential differences are caused by the addition of domains which carry some other function. The main sequence fingerprint of nitrilases is the presence of a Glu-Lys-Cys catalytic triad, which functions as a charge relay system. These conserved residues are recruited from various parts of the primary structure to form an active site. Functional substrate specificity of these enzymes is closely related to the surface side chains around the catalytic residues.

Recently, more details of their catalytic activity have been gleaned from newly solved nitrilase structures. In fact, the structure of AmiF formidase from *Helicobacter pylori* showed that residue Glu141 is located very close to the catalytic triad and facilitates the correct docking of the substrate into the site (Chiu-Lien et al., 2007). Further detailed analysis of the active site in the structure of aliphatic amidase from *Geobacillus pallidus* RAPc8 reveals that the corresponding residue Glu142 could be the fourth necessary participant of the enzyme reaction process, and its position is conserved in homologous sequences (Kimani et al., 2007). These observations argue that one should consider a fourth residue, namely a Glu, when discussing the reaction mechanism. In fact, the acyl transfer intermediate was observed near the nucleophilic amino acid Cys166 in the structure of the amidase from *Pseudomonas aeruginosa* (Andrade et al., 2007). Recently, the crystal structure of a rather unique aliphatic amidase from a psychrotrophic and haloalkaliphilic *Nesterenkonia* species was obtained (Nel et al., 2011). The species was isolated using a screening program for nitrile- and amide-hydrolyzing microorganisms in Antarctic desert soil samples. The enzyme has its optimal activity in the alkaline pH range, at higher salt, and at low temperature. This example demonstrates that there is value in studying amidase enzymes in environmental microbiology.

Here, we report the structure of the uncharacterized protein SA0302 from the pathogenic bacterium *Staphylococcus aureus* subsp. *aureus* COL. We provide evidence that places the protein in Branch 10 (Nit and NitFhit) of the nitrilase superfamily. This work was carried out as part of a project devoted to the structural investigation of essential enzymes of pathogenic bacteria. The main goal of the work was to establish the possible type of enzyme function for gene product *nit sa0302* on the basis of the high-resolution crystal structure of this protein.

Data and methods

Protein expression

The gene coding for SA0302 was amplified by PCR from genomic DNA of *S. aureus* and cloned into a modified version of the plasmid pET15b (Novagen) producing an N-terminal hexahistidine tag fusion protein (http://www.sgc.utoronto.ca/SGC-WebPages/Vector_PDF/p15TV-L.pdf). Selenomethionine-labeled protein was expressed in *Escherichia coli* BL21 (λDE3) RIPL + in minimal selenomethionine media supplied by Medicilon (California). The cells were grown at 37 °C until OD_{600} reached 0.6, then Isopropyl β-D-1-thiogalactopyranoside was added to a final concentration of 1 mM and the expression was continued overnight at 16 °C. The cells were harvested by centrifugation, flash-frozen in liquid nitrogen, and stored at −80 °C.

Purification

Bacterial cells were thawed on ice, resuspended in binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5% CHAPS, 0.2 mM TCEP, and 25 mM imidazole), lysed by sonication and subjected to centrifugation at 60,000*g* for 40 min. The cell-free extract was passed by gravity flow through a DE-52 cellulose column (Whatman) preequilibrated in the binding buffer. The flow-through fraction was loaded directly on nickel–nitrilotriacetic acid-agarose resin (Qiagen) and washed sequentially with 10 column volumes of the binding buffer and 20 column volumes of the binding buffer without CHAPS. The purified protein was eluted with 50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.2 mM TCEP, and 300 mM imidazole and concentrated to 3–5 mg/ml using Vivaspin 20 with 5 kDa molecular weight cut-off concentrators (GE Healthcare). As the final purification step, gel-filtration was performed on a 26/60 Superdex-200 column using the AKTA explorer system (GE Healthcare) with a buffer of 10 mM HEPES pH 7.5, 50 mM NaCl, and 0.2 mM TCEP. Protein-containing fractions were pooled and concentrated to 20 mg/ml as described above. Then, the sample was aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C. The

molecular weight of the protein sample was determined by electrospray ionization time-of-flight mass spectrometry analysis as 32,139 Da matching the calculated mass within the error of measurement.

Crystallization

SA0302 crystals were obtained by sitting-drop vapor diffusion at room temperature against a reservoir solution containing 25% PEG 3350, 200 mM $MgCl_2$, 100 mM TRIS buffer, pH 8.2. Crystals grew to $75 \times 75 \times 100$ µm in one week. Crystals were cryoprotected by transfer to a mixture of Paratone-N and mineral oil (50/50), soaked for 30–60 s, and flash-frozen in liquid nitrogen.

Data collection, processing, and refinement

X-ray diffraction data were collected at 100 K using the Advanced Photon Source IMCA-CAT beam line 17-ID (Argonne National Laboratory). The single-wavelength anomalous dispersion (SAD) data to 1.7 Å resolution were processed using the program package *XDS* (Kabsch, 1993). The initial phases were obtained by the program *autoSHARP* (Vonrhein, Blanc, Roversi, & Bricogne, 2006). Phase improvement by density modification with the use of *CCP4* (Cowtan, 1994) generated an interpretable experimental SAD map, which allowed building an initial model using *ARP*/wARP (Perrakis, Morris, & Lamzin, 1999). The model was subsequently improved through alternate cycles of manual rebuilding using *COOT* (Emsley & Cowtan, 2004) and restrained refinement against a maximum likelihood target with 5% of the reflections randomly excluded as an R_{free} test set.

Table 1. Crystallographic data and refinement statistics.

Data collection	
Wavelength (Å)	0.9794
Resolution (Å)	20.0–1.7 (1.8–1.7)
Space group	$P2_12_12_1$
Unit cell parameters (Å)	58.6, 62.0, 155.2
Molecules per asymmetric unit	2
Unique reflections	60,527 (7,994)
Multiplicity	9.6 (5.8)
Average $I/\sigma(I)$	18.1 (2.8)
R_{merge}	0.08 (0.458)
Completeness (%)	97.7
Refinement and structure statistics	
R_{cryst}	0.20
R_{free}	0.23
RMSD from ideal geometry	
Bond lengths (Å)	0.01
Bond angles (°)	1.14
Numbers of atoms	
Protein nonhydrogen atoms	4318
Water oxygen atoms	555
Ligand atoms[a]	59
PDB ID	3P8 K

[a]Ligands: four Chloride ions, 1 TRIS, 3 Di(Hydroxyethyl) Ether (PEG), 2 Triethylene Glycol (PGE) and one glycerol.

All refinement steps were performed using the *BUSTER/TNT* program (Bricogne & Irwin, 1996, 1997; Tronrud, 1997). The data processing and refinement statistics are presented in Table 1. The atomic coordinates have been deposited in the RCSB PDB, accession code 3P8 K.

Sequence and structural homology search

Sequence and structural homology of SA0302 to a set of homologous proteins with known structure was found by running its sequence in a *protein BLAST* search (Altschul et al., 1997, 2005; Basic Local Alignment Search Tool [BLAST], 2000). Structural comparisons between SA0302 and each of its significant matches were performed using a *DALI* search (Holm & Sander, 1993).

Results

Protein structure

The crystal structure of SA0302 was solved at 1.7 Å resolution by the SAD method using Se-Met substituted protein. The protein contains 261 amino acids and consists of a four-layer αββα sandwich, consistent with the expected CN-hydrolase fold (Figure 1(a)). The monomer consists of two similar subdomains related by a pseudo twofold rotational symmetry (Figure 1(b)). These related parts differ in length: the N-subdomain consists of 110 residues while the C-subdomain includes 151 residues. Consequently, a small portion of the protein chain undergoes a structural rearrangement, resulting in improved spatial symmetry but violating topological symmetry (Figure 1(c)). For example, helix α2 (residues 49–55) is symmetry-related to deformed helix α4 (residues 115–121), which must be assigned now as belonging to the C-subdomain of the protein molecule. The topology of SA0302 resembles that of the putative CN-hydrolase from *Saccharomyces cerevisiae*, PDB code 1F89 (Kumaran et al., 2003). In the crystal lattice, the protein molecule exists as a dimer with protein monomers related by a pseudo twofold symmetry (Figure 1(d)). The subdomain nature of the monomer is directly related to the evolutionary precursor of this gene that seems to arise from a gene fusion as a result of the joining of two subdomains into one protein molecule. The dissimilarity of the subdomains can be seen in Figure 1(b) and accurately evaluated by topology in Figure 1(c). Note that the active site is formed from the residues of both the N- and C-terminal subdomains. The dimer contains two active sites, which are positioned far from each other and easily accessible to the solvent.

Sequence and structural homology

A set of homologous proteins with known structure was found by running the *protein BLAST* search against the SA0302 sequence. Thirteen structures from the PDB with significant similarity were identified (*Z*-scores 127–

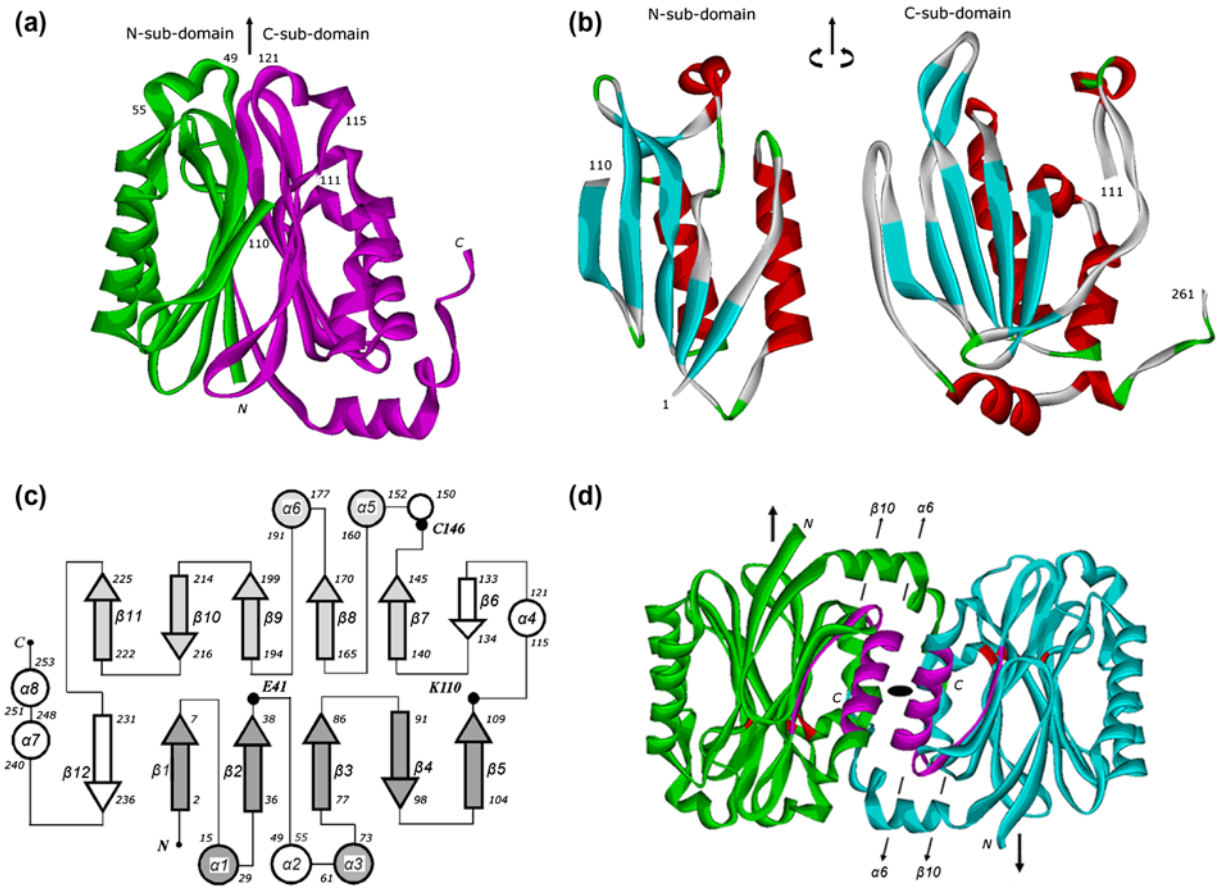

Figure 1. Crystal structure, chain topology and dimer structure of the putative hydrolase SA0302 from *S. aureus*. (a) The protein main chain includes two subdomains marked in green and magenta; both are related by a pseudo twofold symmetry rotation axis. (b) Symmetry related N- and C-parts of the protein are rotated at 90° along the twofold rotation axis. Color-coding: red – α-helix, cyan – β-strand, green – turn. (c) Main chain topology of the N- and C-subdomains; the similar fragments are shown in gray color; the catalytic residues are indicated by black circles. (d) The dimer found in the crystal; it consists of two protein molecules related by a twofold rotation axis. Positions of catalytic triad residues are marked in red, and the conserved α6 and β10 fragments are shown in magenta.

61; sequence identity 32–15%). These matching structures exhibited two groups of *Z*-score values: the first consists of only four proteins with PDB codes 1F89, 2W1V, 2E11, and 1EMS with score 127–104, and the second with lower score values 82–27. The best match 1F89 was a putative CN-hydrolase from *S. cerevisiae* (Kumaran et al., 2003) with a pairwise sequence identity of 31.5%. Structural comparisons between SA0302 and each of these matches were performed using a *DALI* search. Strong structural matches with *Z*-scores 35.6–30.0 and C_α RMSD: 1.5–2.3 Å were revealed for many nitrilase superfamily enzymes: CN-hydrolases 1F89, 2W1V, 2E11, 1EMS, amidohydrolases 1J31, 3HKX, 1ERZ, 2VHH, and other amidases 2PLQ, 2UXY, and 2DYU.

Sequence alignments of SA0302 with four known nitrilase structures are presented in Figure 2. They clearly show the conserved catalytic triad Glu41-Lys110-Cys146. These catalytic residues are located within three semi-conserved regions corresponding to their relative locations in the SA0302 protein sequence: region 1 – flanking E41; region 2 – following K110; and region 3 – flanking C146, which also has a protein signature consistent with the 21-residue "nucleophile elbow" motif (Kumaran et al., 2003). These flanking catalytic sequences of SA0302 are in good accordance with consensus sequences of CN-hydrolase of Branch 10 (Nit and NitFhit) of the nitrilase superfamily (Pace & Brenner, 2001). On the basis of semi-conserved regions flanking the catalytic triad, we have assigned 10 other uncharacterized protein sequences to Branch 10 of the nitrilase superfamily (Figure 3).

We have also identified additional semi-conservative regions 4 (residues 179–195) and 5 (residues 211–220), completing all of the conserved segments in the Nit-nitrilase protein sequence (Figure 2). As we can see below, these two regions represent unique structural elements of Nit-nitrilase-like proteins that supply the external and internal monomer specific interactions. Semi-conservative region 4 provides stability of the dimer. It is connected by

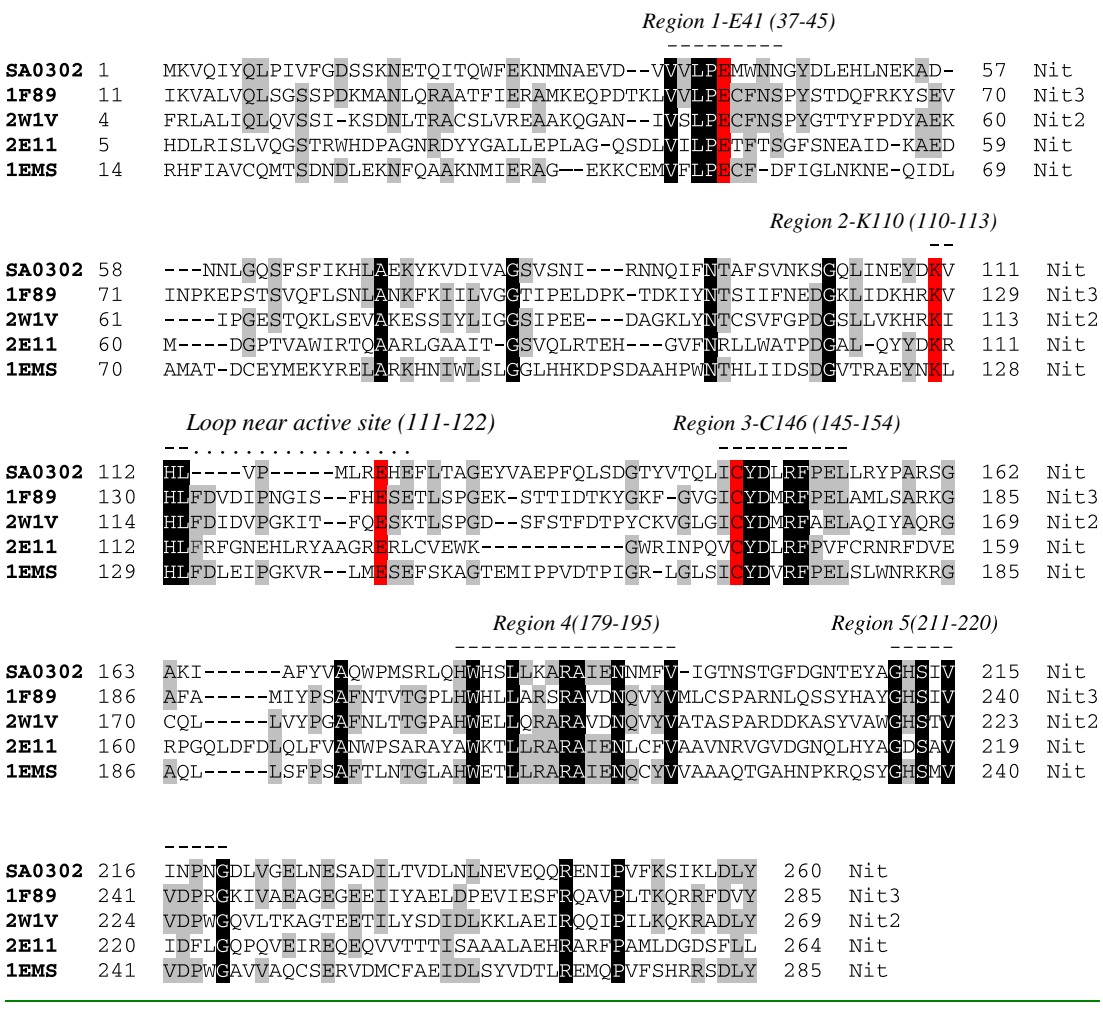

Figure 2. Sequence of SA0302 aligned with Nit-domains of other CN-hydrolases having similar crystal structures and sequence identities of 28–30%. Conserved catalytic triad residues Glu41, Lys110, Cys146, and the additional Glu119 are highlighted with a red background. Fully conserved amino acids are shown in white color against the black background. Semi-conserved residues are shown against a gray background. PDB codes of proteins, with exception of SA0302, are indicated in the left column.

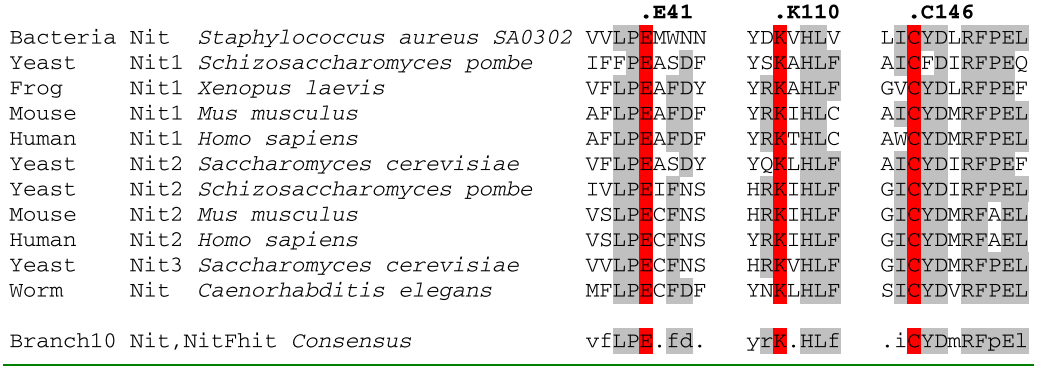

Figure 3. Conserved regions, flanking the catalytic triad, and assignment of uncharacterized proteins to Branch 10 of the nitrilase superfamily. Catalytic triad residues Glu41-Lys110-Cys146 are highlighted against a red background. Conserved residues are shown against a gray background. Sequence conservation of positions is indicated by letter case: upper case corresponds to more than 90% conservation and lower case to 50–90%.

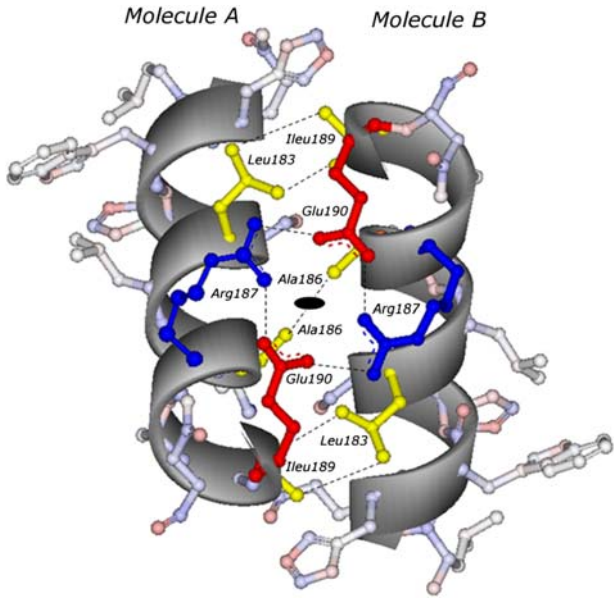

Figure 4. Interactions between molecules A and B in the dimer interface of hydrolase SA0302. The projection is shown along the twofold rotation symmetry axis of the dimer molecule. Distinctive charge-alternating hydrophilic and nonpolar hydrophobic contacts between external sides of two α6-helixes of molecules A and B are shown by dotted lines.

several symmetry related hydrophilic and hydrophobic contacts between external sides of two α6-helixes of molecules A and B, as shown in Figure 4. The polar interaction is formed by four charge-alternating cyclic chain contacts

… Gln190 (A): Arg187(A): Gln190(B): Arg187(B) …

A specific type of contacts similar to the known "leucine-zipper" interaction forms the nonpolar contacts

Leu183(A): Ile 189(B), Ala186(A): Ala186(B) and Ile189(A): Leu183(B).

Semi-conservative region 5 provides the internal core contacts, which are important for molecular stability of the protein monomer. Two nonpolar residues form contacts between the internal surface of helix α6 and short strand β10. These nonconservative residues form two contacts as follows:

Leu184: Ile214 and Ala188: Ile216.

Active site of hydrolase SA0302 from S. aureus

Analysis of the protein sequence and structure homology indicates that SA0302 belongs to the nitrilase superfamily, and more exactly to the CN-hydrolases of Branch 10 (Nit and NitFhit). This conclusion is strongly supported by the sequence alignment in the vicinity of the catalytic residues, which coincides with other members of this Branch (Figure 3).

Despite extensive similarity with other Nit-family structural folds, there is an important structural difference that distinguishes SA0302 from at least four known Nit structures of Branch 10. It is related directly to the region of the active center of SA0302, which is presented in Figure 5(a) and (b). In the vicinity of the active site pocket, one can see loop 111–122, which is marked in dark blue in Figure 5(b). This loop follows catalytic Lys110 and it is fully accessible to the solvent. It also contains deformed helix α4 (115–121). Temperature factors (B) for the main chain atoms of this loop are within the 12–20 $Å^2$ range. These B-values indicate that this fragment has a rather stable main chain conformation. It is significantly shorter when compared to similar loop regions of other Nit-family members (Figure 2). Similar Nit-nitrilases assigned to Branch 10 include an elongated loop of 15–17 residues while in SA0302 this loop is only 8 residues, about half as long. This loop was found in the open position in yeast Nit3 (Kumaran et al., 2003, pdb 1F89) and worm NitFhit (Pace et al., 2000, pdb 1EMS), but in the closed position in mouse Nit2 (Barglow et al., 2008, pdb 2M1V). The dynamic nature of this region implies a possible regulatory role for the active site loop in catalysis (Barglow et al., 2008). However, the apparent shortening of this loop in SA0302 suggests another role in the catalysis process. For example, it could provide specificity to substrate binding. The functional importance of this loop is also connected to the fourth conservative residue Glu119 in protein homologs, which may also participate in the catalysis process.

The catalytic triad of SA0302 is located around a deep and solvent-accessible pocket in both molecules of the dimer (Figure 1(d) and Figure 5(a)). As shown in Figure 5(a), the active site is viewed clearly along a pseudo twofold symmetry rotation axis of the monomer. In the dimer molecule, two active sites are situated far enough from each other so as not to interfere. In the dimer, however, the reliability of the whole enzyme system appears to increase due to strong interactions between the monomers, as seen in Figure 4 and described above. Details of the surface region of SA0302 in the vicinity of catalytic residues Glu41, Lys110, and Cys146 are presented in Figures 5(b) and 6. It is important to note that one observes a central water molecule, W272, which is located 2.9 Å from Glu41 OE2, 3.7 Å from Cys146 SG, and 2.6 Å from Lys110 NZ (Figure 6). It is also interesting that the functionally important loop, which follows catalytic Lys110, contains residue Glu119 and that atom Glu119 OE1 is located 2.7 Å from atom Lys110 NZ. That makes residue Glu119 the prime candidate for the fourth conserved residue in the charge relay system in the process of catalysis (Kimani et al., 2007).

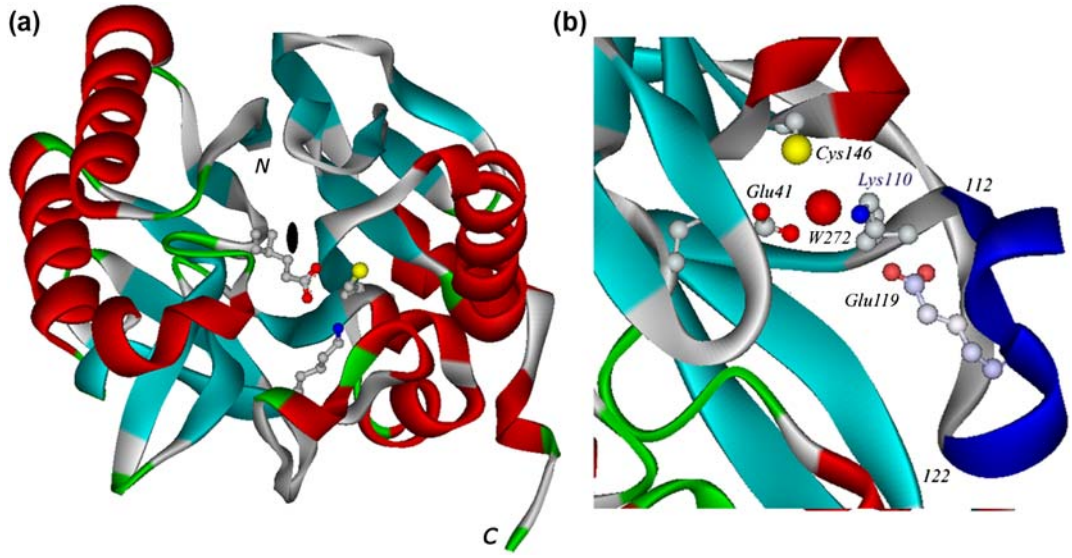

Figure 5. Active site of hydrolase SA0302. (a) General view of the active site with the catalytic triad as seen along the pseudo twofold rotation axis. (b) Detailed view of the active site at a larger scale. The catalytic charge relay system Glu41-Water272-Cys146-Lys110 is located along a deep and accessible pocket. The water molecule is colored in red and the cysteine SH group is colored in yellow, the "loop" located near the active site is marked in a dark blue color. The location of conserved residue Glu119 is shown here; this residue could be a possible fourth residue involved in the process of catalysis.

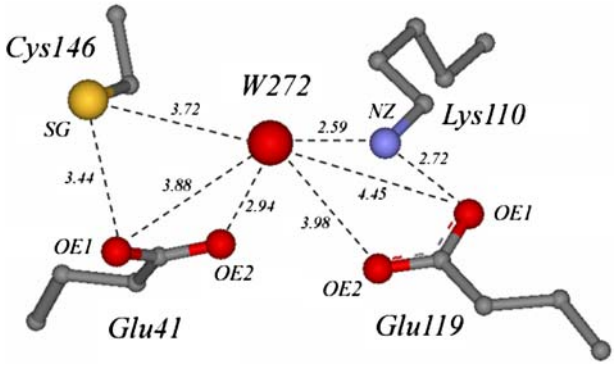

Figure 6. Disposition of catalytic atomic groups and water molecule in hydrolase SA0302, which are possible participants in the charge relay system.

Discussion and conclusion

As we have shown, the high-resolution protein structure SA0302 from *S. aureus* assigns the protein as a CN-hydrolase of Branch 10 of the nitrilase superfamily. Despite rather low pairwise sequence identity values of 25–31%, against the already known structures of this Branch, many structural features of SA0302 are very similar to those of the other members of the subfamily. However, we have also revealed a number of specific features inherent solely to SA0302. The most significant new observations are as follows:

- The protein monomer of SA0302 consists of two quasi-symmetrical N- and C-subdomains. Both are built from the fragments. This feature was not considered earlier for other similar structures and it is clearly seen in the common scheme of main chain topology.
- We have observed three distinctive conserved regions 1, 2, and 3 for the sequences of Branch 10 members, which are related to the functionally important catalytic residues Glu41-Lys110-Cys146. In addition, we have found two new reasonably conserved regions 4 and 5, which are related to the unique structural elements of dimeric protein associates. These two structural elements impart crucial intramonomer and intermonomer interactions providing stability of dimer Nit structures.
- Of special interest is the nonconserved loop, which is situated near the active centre, followed by catalytically active Lys110. The structure of hydrolase SA0302 is remarkable because it is the only member of the Branch 10 Nit subfamily where loop 111–122 is approximately half as long as in other members. We believe that this functionally very important loop relates directly to the substrate–enzyme relationship.
- We have used three short sequence regions, flanking three catalytic residues, as a test to reveal the sequence homologs of the Branch 10 nitrilase superfamily. On this basis, we have assigned 10 new, yet uncharacterized proteins to the hydrolase subfamily of this Branch.

- We have studied the disposition of catalytic triad Glu41-Lys110-Cys146 in details and found in its vicinity an important structural water molecule and the carboxyl group of residue Glu119, which seem to be the potential additions to the catalytic apparatus (Figure 6). This suggests a total five participants closely related to the catalytic charge transfer process. It is interesting to note that residue Glu119 is also in the middle of peptide loop 111–122.

In spite of growing interest about the details of the enzymatic mechanism of the members of Branch 10, at present little is known about the specificity of possible substrates or inhibitors. This is a very challenging biochemical problem that is still far from being resolved. At the moment, we can cite one important related reference (Barglow et al., 2008). In this paper, two murine nitrilases, including Nit1 and Nit2, were identified as targets for a dipeptide-chloroacetamide activity-based probe (dipeptide -CA probe). Gel analysis of binding of Nit with dipeptide -CA probes shows definite selectivity of labeling inside the Nit subfamily. Experimental data of positive labeling are as follows:

Nit1: Leu-Tyr, Leu-His, Leu-Leu, Asp-Leu, Glu-Leu, Tyr-Leu, D-Leu-Asp, Leu-D-Asp, D-Leu-D-Asp, D-Leu-D-Asp

Nit2: Leu-Tyr, Leu-His, Leu-Leu, Leu-Arg, Leu-Glu, Leu-Asp, D-Leu-Asp,

where common dipeptide -CA probes are underlined.

These data implicate the specificity of substrate binding for a definite member of Branch 10 of the Nit-nitrilase superfamily. The present solution of the substrate-free bacterial protein structure of SA0302 forms the basis for substrate binding studies with any potential substrates or inhibitors.

Acknowledgments

The authors acknowledge a grant from the Ontario Research and Development Challenge Fund (99-SEP-0512). We also would like to thank Affinium Pharmaceuticals, Inc., Toronto, for their contribution at earlier stages of the project. Emil F. Pai acknowledges support from the Canadian Research Chairs program. Use of the IMCA-CAT beamline 17-ID at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Hauptman-Woodward Medical Research Institute. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search program. *Nucleic Acids Research, 25*, 3389–3402.

Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schaffer, A.A., & Yu, Y.K. (2005). Protein database searches using compositionally adjusted substitution matrices. *FEBS Journal, 272*, 5101–5109.

Andrade, J., Amin Karmali, A., Carrondo, M.A., & Frazao, C. (2007). Structure of amidase from *Pseudomonas aeruginosa* showing a trapped acyl transfer reaction intermediate state. *Journal of Biological Chemistry, 282*, 19598–19605. pdb 2UXY.

Barglow, K.T., Saikatendu, K.S., Bracey, M.H., Huey, R., Morris, G.M., Olson, A.J., … Cravatt, B.F. (2008). Functional proteomic and structural insights into molecular recognition in the nitrilase family enzymes. *Biochemistry, 47*, 13514–13523. pdb 2W1V.

Basic Local Alignment Search Tool [BLAST]. (2000). Retrieved from http://www.ncbi.nlm.nih.gov/blast/Blast.cgi

Bork, P., & Koonin, E.V. (1994). A new family of carbon–nitrogen hydrolases. *Protein Science, 3*, 1344–1346.

Brenner, C. (2002). Catalysis in the nitrilase superfamily. *Current Opinion in Structural Biology, 12*, 775–782.

Bricogne, G., & Irwin, J.J. (1996). Macromolecular refinement. In E. Dodson, M. Moore, A. Ralph, & S. Bailey (Eds.), *Proceedings of the CCP4 Study Weekend* (pp. 85–92). Warrington: SERC Daresbury Laboratory.

Bricogne, G., & Irwin, J.J. (1997). The BUSTER environment. In P. Bourne & K. Watenpaugh (Eds.), *Crystallographic computing*, *Vol. 7* (pp. 1–9). Oxford: Oxford University Press.

Chin, K.H., Tsai, Y.D., Chan, N.L., Huang, K.F., Wang, A.H.J., & Chou, S.H. (2007). The crystal structure of XC1258 from *Xanthomonas campestris*: A putative procaryotic Nit protein with an arsenic adduct in the active site. *Proteins, 69*, 665–671pdb 2E11.

Chiu-Lien, H., Jia-Hsin, L., Wei-Chun, C., Shao-Wei, H., Jenn-Kang, H., & Wen-Ching, W. (2007). Crystal structure of *Helicobacter pylori* formamidase AmiF reveals a cysteine–glutamate–lysine catalytic triad. *Journal of Biological Chemistry, 282*, 12220–12229. pdb 2E2D.

Cowtan, K. (1994). Phase improvement by density modification. Joint project CCP4 and ESF-EACBM. *Newsletter on Protein Crystallography, 31*, 34–38.

Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica, D60*, 2126–2132.

Holm, L., & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *Journal of Molecular Biology, 233*, 123–138.

Kabsch, W. (1993). XDS program package. *Journal of Applied Crystallography, 26*, 795–800.

Kimani, S.W., Agarkar, V.B., Cowan, D.A., Sayed, M.F.R., & Sewell, B.T. (2007). Structure of an aliphatic amidase from *Geobacillus pallidus* RAPc8. *Acta Crystallographica, D63*, 1048–1058. pdb 2PLQ.

Kumaran, D., Eswaramoorthy, S., Gerchman, S.E., Kycia, H., Studier, F.W., & Swaminathan, S. (2003). Crystal structure of a putative CN hydrolase from yeast. *Proteins, 52*, 283–291. pdb 1F89.

Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., … Tsukihara, T. (2000). Crystal structure of N-carbamyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases. *Structure, 8*, 729–739. pdb 1ERZ.

Nel, A.J., Tuffin, I.M., Sewell, B.T., & Cowan, D.A. (2011). Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic nesterenkonia isolate. *Applied Environmental Microbiology, 77*, 3696–3702. pdb 3HKX.

Pace, H.C., & Brenner, C. (2001). The nitrilase superfamily: Classification, structure and function. *Genome Biology, 2*, 001–009.

Pace, H.C., Hodawadekar, S.C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., … Brenner, C. (2000). Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers. *Current Biology, 10*, 907–917. pdb 1EMS.

Perrakis, A., Morris, R., & Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Structural Biology, 6*, 458–463.

Sakai, N., Tajika, Y., Yao, M., Watanabe, N., & Tanaka, I. (2004). Crystal structure of hypothetical protein PH0642 from *Pyrococcus horikoshii* at 1.6 Å resolution. *Proteins, 57*, 869–873. pdb 1J31.

Tronrud, D.E. (1997). The TNT refinement package, in "macromolecular crystallography, part B". *Methods in Enzymology, 277*, 306–319.

Vonrhein, C., Blanc, E., Roversi, P., & Bricogne, G. (2006). Automated structure solution with autoSHARP. *Methods in Molecular Biology, 364*, 215–230.