Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of a short-chain dehydrogenase/reductase from *Bacillus anthracis*

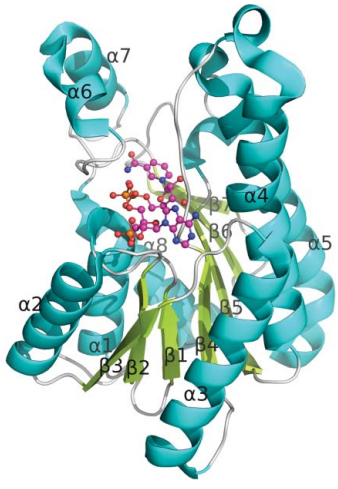
Jing Hou,[a,b] Kamila Wojciechowska,[a] Heping Zheng,[a,b] Maksymilian Chruszcz,[a,b] David R. Cooper,[a,b] Marcin Cymborowski,[a,b] Tatiana Skarina,[b,c] Elena Gordon,[b,c] Haibin Luo,[a,b] Alexei Savchenko[b,c] and Wladek Minor[a,b]*

[a]Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Jordan Hall, Room 4223, Charlottesville, VA 22908, USA, [b]Center for Structural Genomics of Infectious Diseases (CSGID), USA, and [c]Banting and Best Department of Medical Research, University of Toronto, 112 College Street, Toronto, Ontario, Canada

Correspondence e-mail: wladek@iwonka.med.virginia.edu

Received 21 March 2012
Accepted 22 April 2012

PDB Reference: short-chain dehydrogenase/reductase, 3icc.



The crystal structure of a short-chain dehydrogenase/reductase from *Bacillus anthracis* strain 'Ames Ancestor' complexed with NADP has been determined and refined to 1.87 Å resolution. The structure of the enzyme consists of a Rossmann fold composed of seven parallel β-strands sandwiched by three α-helices on each side. An NADP molecule from an endogenous source is bound in the conserved binding pocket in the *syn* conformation. The loop region responsible for binding another substrate forms two perpendicular short helices connected by a sharp turn.

1. Introduction

Bacillus anthracis is a Gram-positive bacterium that can potentially cause the lethal disease anthrax in humans and animals, and also has the potential to be used as a biological weapon (Bouzianas, 2010). *B. anthracis* is classified by the NIAID as a Category A priority pathogen and is one of the target organisms of the Center of Structural Genomics of Infectious Diseases (CSGID; Anderson, 2009). The CSGID selected this target because it is conserved and believed to be essential for survival and involved in sporulation.

The BA1847 gene was cloned from the genome of *B. anthracis* strain 'Ames Ancestor'. It encodes a protein consisting of 252 amino acids (UniProt Q81S30) that we will refer to as BA1847 in this manuscript. It contains the typical motifs of an oxidoreductase of the short-chain dehydrogenase/reductase (SDR) superfamily, which includes NAD(H)- or NADP(H)-dependent reductases with a wide variety of substrates (Kallberg *et al.*, 2002, 2010). Members of the SDR superfamily are ubiquitous in all kingdoms of life (Jörnvall *et al.*, 1999). More than 300 SDR families have been identified, among which about 50% are found only in bacteria (Kallberg *et al.*, 2010). Substrates for SDRs include many important metabolites such as sugars and their derivatives in carbohydrate metabolism (Koropatkin & Holden, 2005; Zhang *et al.*, 2009), steroids in signal transduction (Benach *et al.*, 2002; Svegelj *et al.*, 2011) and keto-acyl-(acyl carrier proteins) and enoyl-acyl-(acyl carrier proteins) in fatty-acid synthesis (Lee *et al.*, 2007; Kim *et al.*, 2010; Price *et al.*, 2004; Zaccai *et al.*, 2008). SDRs usually consist of two domains: a coenzyme-binding domain and a substrate-binding domain. While the coenzyme-binding domain is usually conserved, the substrate-binding domain normally varies and determines the substrate specificity (Kavanagh *et al.*, 2008).

Although the biological substrate of BA1847 is unknown, *BLAST* searches (Altschul *et al.*, 1990) using the NCBI nonredundant protein database showed that the protein is highly conserved among the *Bacillus* genus, with greater than 90% identity among different species of *Bacillus*, including *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. mycoides* and *B. pseudomycoides*, indicating that this protein might play an indispensible role. Searching the *B. anthracis* genome (http://hamap.expasy.org/proteomes/BACAN.html) revealed that 21 genes are annotated as short-chain dehydrogenases/oxidoreductases. The substrate specificities for most of them are unknown, except for a 3-keto-acyl-(acyl carrier protein) reductase (BaFabG; PDB entry 2uvd; Zaccai *et al.*, 2008), which shares 38% sequence identity with BA1847.

Table 1
Diffraction data and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data collection	
Wavelength (Å)	0.9793
Space group	$C222_1$
Unit-cell parameters (Å)	a = 98.7, b = 104.8, c = 103.7
Resolution (Å)	42.00–1.87 (1.92–1.87)
No. of unique reflections	44201 (2186)
Completeness (%)	100.0 (100.0)
Multiplicity	7.3 (7.0)
Mean $I/\sigma(I)$	46.3 (2.9)
Molecules in asymmetric unit	2
Matthews coefficient (Å^3 Da^{-1})	2.43
Solvent content (%)	49.33
$R_{\rm merge}$ (%)	10.0 (68.6)
Structure refinement	
Resolution range (Å)	42.00–1.87 (1.92–1.87)
$R_{\rm work}/R_{\rm free}$ (%)	16.6/20.7 (21.5/26.6)
No. of residues/protein atoms	510/3826
No. of water atoms	288
Average B factors (Å^2)	
Main chain	14.66
Side chains	17.69
Overall	17.17
Waters	23.26
Ligands/ion	36.58
Ramachandran plot (%)	
Most favored	98.00
Allowed	2.00
Disallowed	0.00
R.m.s. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.557

In this paper, we report the crystal structure of BA1847 in complex with NADP. The structure reveals that BA1847 has a Rossmann fold with an endogenous NADP bound to the conserved binding pocket. The binding of NADP also induces the substrate-binding loop to form an ordered helix–turn–helix motif.

2. Materials and methods

2.1. Cloning, expression and purification of BA1847

The ORF of the short-chain dehydrogenase BA1847 (CSGID target ID IDP02573, gene locus GBAA1847, gi:47527139, accession No. YP_018488) was amplified by polymerase chain reaction (PCR) from genomic DNA of *B. anthracis* strain 'Ames Ancestor' using the forward primer 5′-TACTTCCAATCCAATGCGATGTTAAAAG-GAAAAGTAGCATTA-3′ and the reverse primer 5′-TTATCCAC-TTCCAATGTTATAAACAAGATCCTCCGCTA-3′. The gene was cloned into pMCSG7 plasmid encoding an N-terminal tobacco etch virus protease-cleavable hexahistidine tag (MHHHHHHSSGVD-LGTENLYFQSNA; Stols *et al.*, 2002) using ligation-independent cloning. The recombinant protein was overexpressed in *Escherichia coli* strain BL21 (Codon Plus)-RIPL. The cells were grown in M9 minimal medium supplemented with selenomethionine (SeMet) at 310 K until the OD_{600} reached 1.0. Protein expression was then induced by the addition of 0.5 m*M* isopropyl β-D-1-thiogalacto-pyranoside (IPTG). The induced cells were cultured at 293 K for a further 16–18 h.

The harvested cells were sonicated in lysis buffer consisting of 50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 5% glycerol, 5 m*M* imidazole supplemented with non-EDTA protease-inhibitor cocktail (Roche) and clarified by centrifugation. The supernatant was applied onto Ni–NTA affinity resin (Qiagen, USA) pre-equilibrated with lysis buffer and washed with buffer consisting of 50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 5% glycerol, 30 m*M* imidazole. The His-tagged protein was eluted with 250 m*M* imidazole.

Recombinant His-tagged TEV protease (1 mg) was added to 40 mg eluted protein to cleave the hexahistidine tag from BA1847 and was incubated at 277 K overnight. The cleaved protein solution was dialyzed against 50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 5% glycerol and applied onto an Ni–NTA column to remove the TEV, uncleaved protein and free hexahistidine tag from recombinant BA1847. The flowthrough from the second Ni–NTA column was concentrated using an Amicon Ultra 30K concentrator (Millipore, USA), applied onto a Superdex G200 gel-filtration column attached to an ÄKTA purification system (GE Healthcare, USA) and eluted with 50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 5% glycerol. BA1847 eluted as a single peak at an elution volume consistent with a tetramer and appeared as a single band of ∼27 kDa on SDS–PAGE with Coomassie Brilliant Blue staining. Dynamic light-scattering analysis was also consistent with a tetramer in solution (data not shown). The purified protein was stored in 50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 5% glycerol.

2.2. Crystallization

Crystallization was performed by hanging-drop vapour diffusion. 1 µl of BA1847 at 11.5 mg ml^{-1} was mixed with an equal volume of 0.1 *M* MES pH 6.5, 0.2 *M* ammonium sulfate, 20–30%(*w*/*v*) PEG 5000 MME and equilibrated against the same buffer at 289 K. Square-plate-shaped crystals appeared the next day and reached their largest size in two weeks. 40%(*w*/*v*) PEG 5000 MME in 0.1 *M* MES pH 6.5, 0.2 *M* ammonium sulfate was used as a cryoprotectant. The crystals were then flash-cooled in liquid nitrogen.

2.3. Data collection, structure determination and refinement

Data were collected at the Structural Biology Center (Rosenbaum *et al.*, 2006) at the Advanced Photon Source in Argonne National Laboratory (Argonne, Illinois, USA). Data collection on beamline 19ID was facilitated by the *HKL*-3000 program suite (Minor *et al.*, 2006), which incorporates *HKL*-2000 (Otwinowski & Minor, 1997) for data processing and scaling. Data were collected at a single wavelength at 100 K and the structure was determined and refined by *HKL*-3000, which utilized the following programs during the course of the BA1847 project: *SHELXD*, *SHELXE* (Sheldrick, 2008, 2010), *DM* (Cowtan & Main, 1993), *ARP*/*wARP* (Perrakis *et al.*, 1999), *CCP*4 (Winn *et al.*, 2011), *SOLVE* (Terwilliger & Berendzen, 1999), *RESOLVE* (Terwilliger, 2004) and *REFMAC*5 (Murshudov *et al.*, 2011). 12 seleniums were located within the asymmetric unit and a partial model was built using the SAD phases during data collection. This partial model was used for further refinement and manual model building/adjustments. The NCS symmetry was utilized for both phase improvement and refinement. The merged Bijvoet pairs were used for refinement and were deposited in the PDB. Manual model building and refinement were performed in *Coot* (Emsley & Cowtan, 2004). *MolProbity* (Chen *et al.*, 2010) and *ADIT* (Yang *et al.*, 2004) were used for structure validation. The statistics of data collection and structure determination are summarized in Table 1.

The coordinates and experimental structure factors were deposited in the PDB with accession code 3icc and the diffraction images are available at the CSGID website (http://www.csgid.org).

2.4. Catalytic assays for substrate screening

Substrate screening was performed using β-nicotinamide adenine dinucleotide 2′-phosphate reduced tetrasodium salt hydrate

(NADPH) and potential substrates inferred from the *DALI* search (Holm & Rosenström, 2010) results, such as single sugars, tropinone, gluconate, hydroxysteroids, alcohols and acetoacetyl coenzyme A (as a substitute for acetoacetyl-acyl carrier protein). Enzymatic activities were measured by monitoring the decrease in absorbance of NADPH at 340 nm at 298 K in 96-well UV-transmitting half-area microplates (Corning) using a spectrophotometer (SpectraMax, Molecular Devices). Assays were performed in 50 m*M* HEPES pH 7.5 and 200 m*M* NaCl in a total volume of 80 µl with 1.0–10 nmol BA1847, 400 µ*M* NADPH and 0.1–10 m*M* substrate.

3. Results and discussion

3.1. Overall structure and putative substrate-binding site

The crystal structure of BA1847 contains a dimer in the asymmetric unit with a core typical of short-chain dehydrogenases (SDRs). The overall structures of the monomers are very similar. Least-squares superposition of C^{α} atoms gives an r.m.s. deviation of 0.27 Å, with the largest deviation observed for the fragment 193–198 (r.m.s. deviation of 1.95 Å). Each monomer has a Rossmann fold (Rao & Rossmann, 1973) with seven parallel β-strands sandwiched between α1, α2 and α8 on one side and α3, α4 and α5 on the other side (Fig. 1). Substrate binding is mediated by two short helices (α6 and α7) that serve as a lid for the active site.

A *DALI* search (Holm & Rosenström, 2010) was performed using the BA1847 monomer. It identified 209 similar structures with Z score > 15 and r.m.s.d. < 2.0 Å. The sequence identities between BA1847 and these proteins range between 20 and 38%. These proteins have very different substrate specificities; the most popular substrates are sugars and their derivatives, followed by 3-oxoacyl-ACP and hydroxysteroids.

The structure of BA1847 has the highest sequence identity (38%) and the lowest r.m.s.d. (1.4 Å for 244 C^{α} atoms) to that of 3-keto-acyl-(acyl carrier protein) reductase from *B. anthracis* (BaFabG; PDB entry 2uvd; Zaccai *et al.*, 2008; Fig. 2). These two proteins have almost identical N-terminal sequences and well conserved NADP(H)-binding motifs, but the differences described below suggest different functions for the two proteins. It has been hypothesized that acyl carrier protein (ACP) recognizes its functional partners in fatty-acid syntheses, including 3-keto-acyl-(acyl carrier protein) reductase, through interactions between its highly electronegative/hydrophobic helix α2 and the electropositive/hydrophobic surface of its partner enzymes (Zhang *et al.*, 2003). In BaFabG, the ACP-binding residues Arg132 and Arg175 are located at the C-terminal ends of α4 and α5, with both side chains stretching out into the solvent (Zhang *et al.*, 2003). In BA1847, however, major conformational differences are observed for the corresponding residues (Arg135 and Arg179) that limit the solvent exposure of their side chains. A three-residue deletion shortens the α4 helix and positions the Arg135 side chain in the vicinity of the loop following α3. The position corresponding to the main chain of Arg132 in BaFabG is occupied by the side chain of an aspartic acid (Asp136), which forms hydrogen bonds and a salt bridge to the Arg179 side chain. Additional hydrogen bonding is also found between the Arg179 side chain and the Leu134 backbone carbonyl, which is exposed as a result of the deletion. These features in BA1847 might make the electrostatics/hydrophobicity of this area different from that of BaFabG, thus diminishing the affinity towards ACP. Thus, the biological substrate of BA1847 is unlikely to be 3-ketoacyl-ACP. This may explain why it lacks activity toward acetoacetyl coenzyme A, which is used as a substitute for 3-oxoacyl-ACP in catalytic activity assays (Karmodiya & Surolia, 2006; Silva *et al.*, 2006).

The crystal structure of BA1847 in complex with NADP is also very similar to those of tropinone reductase (sequence identity 30%, Z = 35.8; Yamashita *et al.*, 2003), trihydroxynaphthalene reductase (sequence identity 37%, Z = 35.7; Liao *et al.*, 2001), 7-α-hydroxy-steroid dehydrogenase (sequence identity 35%, Z = 35.5; Benach *et al.*, 2002), glucose dehydrogenase (sequence identity 32%, Z = 35; Yamamoto *et al.*, 2001) and many more (Fig. 2). Substrate screening was performed using tropinone, gluconate, hydroxysteroid, glucose and many other potential substrates as inferred from the *DALI*

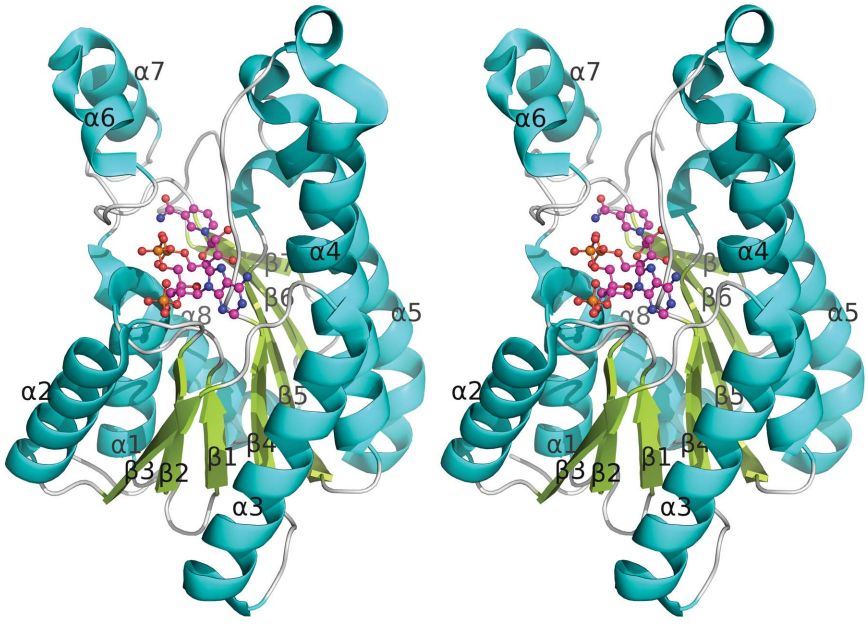


Figure 1
BA1847 monomer shown in cartoon representation. α-Helices are shown in cyan, β-strands in pale green and loops in gray. The NADP molecule is shown in ball-and-stick representation with C atoms in magenta, N atoms in blue, O atoms in red and P atoms in orange. All figures were generated using *PyMOL* (Schrodinger LLC).

results. Nevertheless, BA1847 showed no catalytic activity in these assays, so the actual substrate is likely to be unique, yet still remains unknown.

3.2. Catalytic tetrad

NADP(H)-dependent SDRs have a characteristic catalytic tetrad in their active site. In BA1847, this catalytic tetrad forms a conformation ready for catalysis with residues Asn119 (α4 C-terminus), Ser145 (β5 C-terminus), Tyr158 (α5 N-terminus) and Lys162 (in the middle of α5) (Fig. 3). The presence of NADP in the binding pocket might account for the formation of the active state for the catalytic tetrad. This conformational change was demonstrated in structures determined previously for some other SDR family members. In these proteins, residues corresponding to Ser145, Tyr158 and Lys162 reside far away from the binding pocket in the absence of cofactor (Cohen-Gonsaud *et al.*, 2002, 2005; Price *et al.*, 2001), while NADP(H) binding induces conformational transition into the active state (Price *et al.*, 2004; Yamazawa *et al.*, 2011).

3.3. NADP-binding site

NADP molecules were found in the BA1847 crystal structure, even though no NADP was added to the protein-sample solution during purification or crystallization. There is one NADP in each monomer and both NADP molecules within the asymmetric unit are bound in the same manner. However, only the NADP in monomer *B* has full occupancy, and this will serve as the basis for description of the NADP-binding mode. Detailed interactions between BA1847 and the cofactor NADP in the crystal structure are shown in Fig. 3.

The binding motif for NADP is well conserved in the SDR superfamily and the binding mode seen here is very similar to that observed in other family members (Kallberg *et al.*, 2002; Zhang *et al.*, 2009; Svegelj *et al.*, 2011; Price *et al.*, 2004; Yamashita *et al.*, 2003; Yamazawa *et al.*, 2011). The binding pocket for the adenine ring is mainly formed by residues from the N-terminal part of the protein, while the binding of the nicotinamide moiety involves the catalytic tetrad and the C-terminal part of the protein. The NADP adapts a *syn* conformation. The cleft surrounded by the β1–α1, β2–α2, β3–α3 and β4–α4 loops and helix α4 accommodates the adenine ring and its associated ribose. One side of the adenine ring faces the hydrophobic patch formed by the side chains of Leu63, Ala96 and Val118. On the other side of the adenine ring there are water-mediated hydrogen bonds between the $O^{\delta1}$ atom of Asn37 and O^{3X} of NADP. Two direct hydrogen bonds play crucial roles in holding the adenine ring in place. One is between N^{6A} of NADP and the $O^{\varepsilon2}$ atom of Glu64 and the other is between the ring N^{1} atom of NADP and the amide N atom of Leu63.

There are extensive hydrophilic interactions between the co-enzyme and the N-terminal conserved NADP(H)-binding motif (TGASRGIG). The O^{γ} atom of Ser13 forms two hydrogen bonds to the ribose O^{3B} atom and the ribose 2′-phosphate O^{1X} atom. This phosphate also forms two hydrogen bonds to the amide N atoms of Asn37 and Arg38 and forms a salt bridge and a hydrogen bond to the side chains of Arg14 and Arg38. The two pyrophosphate groups are secured by both direct and water-mediated hydrogen bonds to the amide N atoms of Ile16, Gly15 and Met195, $N^{\delta2}$ of Asn95, $O^{\gamma1}$ of Thr193 and the carbonyl O atom of Gly97. The hydroxyl groups on the nicotinamide-associated ribose form direct hydrogen bonds to the carbonyl O atom of Asn95, the hydroxyl of Tyr158 and the N^{ε} atom of Lys162. The direct interactions between residues in the catalytic tetrad (Tyr158 and Lys162) and NADP are consistent with the active catalytic state of the enzyme (Price *et al.*, 2004; Yamazawa *et al.*, 2011).

Comparison with other family members suggests that residues 192–212 are responsible for substrate binding and that the conserved motif Pro188-Gly189 at the end of β7 might serve as a conformational switch for substrate binding. The presence of NADP is also critical for stabilization of the conformation of the substrate-binding loop and this region is disordered in many SDRs without a ligand (Yamazawa *et al.*, 2011; Tang *et al.*, 2006). Binding of cofactor usually induces this region to form an ordered structure (Yamazawa *et al.*, 2011). In the crystal structure of BA1847 the nicotinamide ring is stacked in the clamp formed between the Pro188 and Met195 side chains (Fig. 2). The O atom on the nicotinamide forms three hydrogen bonds to the carbonyl O atom of Gly189, the amide N atom of Val191 and $O^{\delta2}$ of Asn196. The N atom of nicotinamide forms two hydrogen bonds to the carbonyl O atom of Val191 and $O^{\gamma1}$ of Thr193. O^{1N} of NADP forms a hydrogen bond to $O^{\gamma1}$ of Thr193. There are two water-mediated hydrogen bonds between the pyrophosphate O atoms and both the O and the N atoms of Asp194. These intensive interactions between the nicotinamide ring of NADP and the residues in the

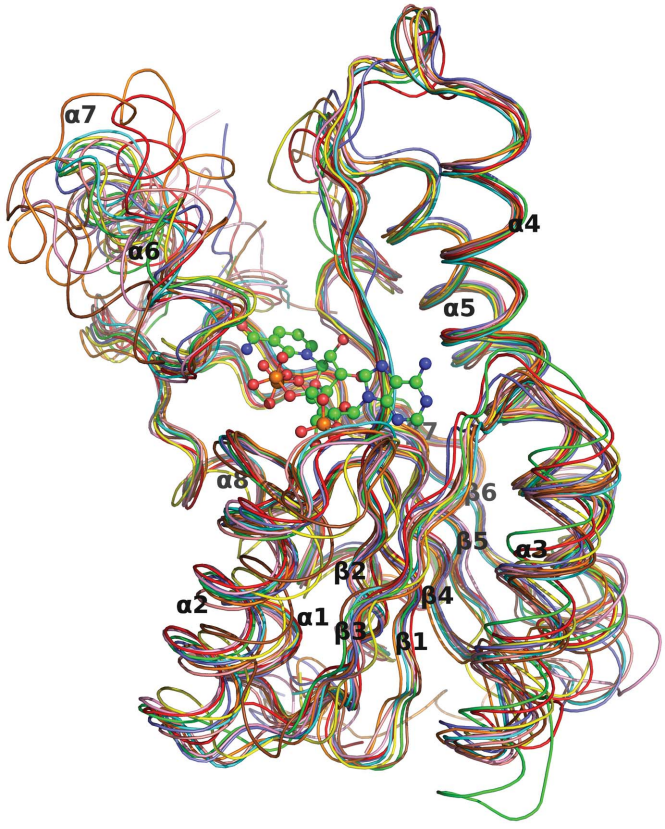


Figure 2
Comparison of the crystal structures of BA1847 and some other representative short-chain dehydrogenases/reductases. The NADP molecule from 3icc is shown in ball-and-stick mode as a reference. 3icc (BA1847 from *B. anthracis*), green; 2uvd [3-keto-acyl-(acyl carrier protein) reductase from *B. anthracis*; Zaccai *et al.*, 2008], cyan; 1ipf (tropinone reductase II from *Datura stramonium*; Yamashita *et al.*, 2003), red; 1vl8 (gluconate 5-dehydrogenase from *Thermotoga maritima*; Joint Center for Structural Genomics, unpublished work), yellow; 1g0n (trihydroxynaphthalene reductase from *Magnaporthe grisea*; Liao *et al.*, 2001), orange; 1fmc (7-α-hydroxysteroid dehydrogenase from *E. coli*; Tanaka *et al.*, 1996), slate; 1edo [3-keto-acyl-(acyl carrier protein) reductase from *Brassica napus*; Fisher *et al.*, 2000], salmon; 1gco (glucose dehydrogenase from *B. megaterium*; Yamamoto *et al.*, 2001), pink; 1q7b [3-keto-acyl-(acyl carrier protein) reductase from *E. coli*; Price *et al.*, 2004], brown.

substrate-binding loop help these residues to stay ordered (Fig. 3). In monomer *B* of the crystal structure of BA1847, fragments 197–201 and 203–211 form two perpendicular α-helices which are connected by a sharp turn (Fig. 1), while in monomer *A* only fragment 203–211 forms an α-helix. The difference in the local structure of the substrate-binding pocket is probably a consequence of the differing NADP occupancies in the two monomers. The NADP molecule in monomer *B* is modelled with full occupancy, while in monomer *A* the occupancy of the NADP molecule is 50%.

3.4. Tetramer of BA1847

The biological assembly of BA1847 is a tetramer, as reported for most other members of the SDR superfamily (Zhang *et al.*, 2009; Zaccai *et al.*, 2008; Yamashita *et al.*, 2003). Although only a dimer is present in the asymmetric unit, the tetrameric biological unit can be generated by a crystallographic symmetry operation (Fig. 4). *PDBePISA* (Krissinel & Henrick, 2007) also indicated that a tetrameric quaternary structure is the functional unit. Two symmetrical dimerization interfaces are responsible for the formation of this tetramer, both of which are structurally conserved in all SDR superfamily members. The first interface involves helices α4 and α5 from two neighboring symmetry-related monomers. These helices run antiparallel to each other, which in turn forms a four-helix bundle. The second interface is formed mainly between the antiparallel α8 helices and β7 strands from each monomer within the asymmetric unit.

In the first interface, hydrophobic interactions play a dominant role. Helix α4 (Ile104, Phe112, Val116, Phe124, Phe125 and Leu131) forms extensive hydrophobic interactions with the same set of residues from $\alpha4'$ (where $\alpha4'$ indicates the counterpart segment from the symmetry-related molecule), while helix α5 (Leu156, Ala157,

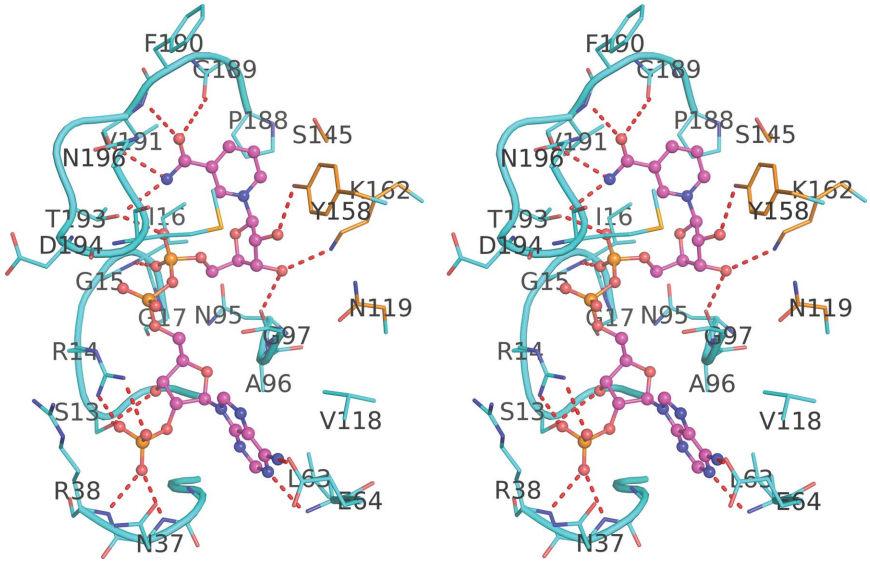


Figure 3
Interaction between BA1847 and NADP. Relevant amino acids are shown in stick representation, while NADP is shown in ball-and-stick representation. The C atoms of amino acids in the binding pocket (except for the catalytic tetrad) are shown in cyan and the C atoms of the catalytic tetrad (Asn119, Ser145, Tyr158 and Lys162) are shown in gold. The C atoms of the NADP molecule are shown in magenta. N atoms are shown in blue, O atoms in red and P atoms in orange. Hydrogen bonds are shown as red dashed lines. Water molecules are not shown for clarity.

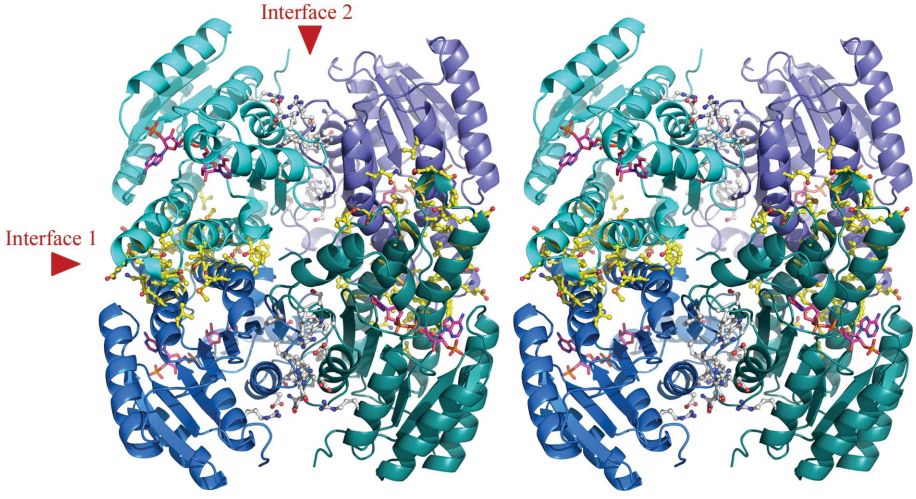


Figure 4
The tetrameric assembly of BA1847. Each monomer is shown in a different color, with symmetry-related monomers in related shading. NADP molecules are shown in stick representation. Side chains of residues in the dimer interfaces are shown in ball-and-stick representation, with C atoms in interface 1 in yellow and those in interface 2 in gray.

Met160, Gly163, Ala164, Met168 and Phe170) forms interactions with $\alpha5'$ in a similar manner. Hydrophilic interactions also contribute to this interface. There is a salt bridge between Lys121 of one monomer and Glu109 and Asp113 of the other monomer. A hydrogen-bond network is formed between residues from one monomer (Ile104, Glu105, Thr107, Gln128, Ser151, Thr171 and Gln175) and their counterpart residues in the other monomer (Gln128, Gln175, Ile104, Thr107, Thr171, Ser151 and Glu105).

A second interface is stabilized by interactions between the C-terminal parts of two monomers. Hydrogen-bonding and salt-bridge interactions are mostly formed by polar residues located on helix $\alpha8$ and the loops before and after helix $\alpha8$. These include the salt bridge and hydrogen bonds formed between Arg237 on loop $\alpha8$–$\beta7$ and Asp226 in $\alpha8'$, between Arg217 on loop $\alpha6$–$\alpha8$ and Asp223 in $\alpha8'$, and between the N atom of Gly248 and the O atom of Trp238 in the neighboring monomer. Three groups of hydrophobic interactions contribute to the second interface: (i) the rings of the two Phe230 residues in $\alpha8$; (ii) the indole ring of Trp238, which protrudes into the pocket formed by residues Gly219 and Gly248 and the side chains of Phe215, Arg217, Glu220, Asp226, Val246 and Ser247 from the neighboring monomer; and (iii) the nonpolar residues on $\beta7$, which interact with those in the other monomer. Arg23 in the N-terminus also contributes to this second interface. The N atoms of its guanidine group both form hydrogen bonds and a salt bridge with the side-chain O atoms of Asp235.

This research was funded by Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under Contract No. HHSN272200700058C. The results shown in this report are derived from work performed at Argonne National Laboratory at the Structural Biology Centre of the Advanced Photon Source. Argonne is operated by University of Chicago Argonne LLC for the US Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.

Anderson, W. F. (2009). *Infect. Disord. Drug Targets*, **9**, 507–517.

Benach, J., Filling, C., Oppermann, U. C., Roversi, P., Bricogne, G., Berndt, K. D., Jörnvall, H. & Ladenstein, R. (2002). *Biochemistry*, **41**, 14659–14668.

Bouzianas, D. G. (2010). *J. Med. Chem.* **53**, 4305–4331.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Cohen-Gonsaud, M., Ducasse, S., Hoh, F., Zerbib, D., Labesse, G. & Quemard, A. (2002). *J. Mol. Biol.* **320**, 249–261.

Cohen-Gonsaud, M., Ducasse-Cabanot, S., Quemard, A. & Labesse, G. (2005). *Proteins*, **60**, 392–400.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fisher, M., Kroon, J. T., Martindale, W., Stuitje, A. R., Slabas, A. R. & Rafferty, J. B. (2000). *Structure Fold. Des.* **8**, 339–347.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Jörnvall, H., Höög, J. O. & Persson, B. (1999). *FEBS Lett.* **445**, 261–264.

Kallberg, Y., Oppermann, U., Jörnvall, H. & Persson, B. (2002). *Eur. J. Biochem.* **269**, 4409–4417.

Kallberg, Y., Oppermann, U. & Persson, B. (2010). *FEBS J.* **277**, 2375–2386.

Karmodiya, K. & Surolia, N. (2006). *FEBS J.* **273**, 4093–4103.

Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. (2008). *Cell. Mol. Life Sci.* **65**, 3895–3906.

Kim, K.-H., Ha, B. H., Kim, S. J., Hong, S. K., Hwang, K. Y. & Kim, E. E. (2010). *J. Mol. Biol.* **406**, 403–415.

Koropatkin, N. M. & Holden, H. M. (2005). *Acta Cryst.* D**61**, 365–373.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Lee, H. H., Moon, J. & Suh, S. W. (2007). *Proteins*, **69**, 691–694.

Liao, D., Basarab, G. S., Gatenby, A. A., Valent, B. & Jordan, D. B. (2001). *Structure*, **9**, 19–27.

Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* D**62**, 859–866.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Price, A. C., Zhang, Y. M., Rock, C. O. & White, S. W. (2001). *Biochemistry*, **40**, 12772–12781.

Price, A. C., Zhang, Y. M., Rock, C. O. & White, S. W. (2004). *Structure*, **12**, 417–428.

Rao, S. T. & Rossmann, M. G. (1973). *J. Mol. Biol.* **76**, 241–256.

Rosenbaum, G. *et al.* (2006). *J. Synchrotron Rad.* **13**, 30–45.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.

Silva, R. G., de Carvalho, L. P., Blanchard, J. S., Santos, D. S. & Basso, L. A. (2006). *Biochemistry*, **45**, 13064–13073.

Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R. & Donnelly, M. I. (2002). *Protein Expr. Purif.* **25**, 8–15.

Svegelj, M. B., Stojan, J. & Rižner, T. L. (2011). *J. Steroid Biochem. Mol. Biol.* **129**, 92–98.

Tanaka, N., Nonaka, T., Tanabe, T., Yoshimoto, T., Tsuru, D. & Mitsui, Y. (1996). *Biochemistry*, **35**, 7715-7730

Tang, Y., Lee, H. Y., Tang, Y., Kim, C.-Y., Mathews, I. & Khosla, C. (2006). *Biochemistry*, **45**, 14085–14093.

Terwilliger, T. (2004). *J. Synchrotron Rad.* **11**, 49–52.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Yamamoto, K., Kurisu, G., Kusunoki, M., Tabata, S., Urabe, I. & Osaki, S. (2001). *J. Biochem.* **129**, 303–312.

Yamashita, A., Endo, M., Higashi, T., Nakatsu, T., Yamada, Y., Oda, J. & Kato, H. (2003). *Biochemistry*, **42**, 5566–5573.

Yamazawa, R., Nakajima, Y., Mushiake, K., Yoshimoto, T. & Ito, K. (2011). *J. Biochem.* **149**, 701–712.

Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst.* D**60**, 1833–1839.

Zaccai, N. R., Carter, L. G., Berrow, N. S., Sainsbury, S., Nettleship, J. E., Walter, T. S., Harlos, K., Owens, R. J., Wilson, K. S., Stuart, D. I. & Esnouf, R. M. (2008). *Proteins*, **70**, 562–567.

Zhang, Q., Peng, H., Gao, F., Liu, Y., Cheng, H., Thompson, J. & Gao, G. F. (2009). *Protein Sci.* **18**, 294–303.

Zhang, Y.-M., Wu, B., Zheng, J. & Rock, C. O. (2003). *J. Biol. Chem.* **278**, 52935–52943.