Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Determination of the crystal structure of EntA, a 2,3-dihydro-2,3-dihydroxybenzoic acid dehydrogenase from *Escherichia coli*

Jesse A. Sundlov,[a] Julie A. Garringer,[a] Jill M. Carney,[a] Albert S. Reger,[a,b] Eric J. Drake,[a,b] William L. Duax[a,b] and Andrew M. Gulick[a,b]*

[a]Hautpman–Woodward Medical Research Institute, Buffalo, USA, and [b]Department of Structural Biology, State University of New York at Buffalo, Buffalo, USA

Correspondence e-mail: gulick@hwi.buffalo.edu

Received 6 February 2006
Accepted 28 April 2006

PDB Reference: EntA, 2fwm, r2fwmsf.

The *Escherichia coli* enterobactin synthetic cluster is composed of six proteins, EntA–EntF, that form the enterobactin molecule from three serine molecules and three molecules of 2,3-dihydroxybenzoic acid (DHB). EntC, EntB and EntA catalyze the three-step synthesis of DHB from chorismate. EntA is a member of the short-chain oxidoreductase (SCOR) family of proteins and catalyzes the final step in DHB synthesis, the NAD^+-dependent oxidation of 2,3-dihydro-2,3-dihydroxybenzoic acid to DHB. The structure of EntA has been determined by multi-wavelength anomalous dispersion methods. Here, the 2.0 Å crystal structure of EntA in the unliganded form is presented. Analysis of the structure in light of recent structural and bioinformatic analysis of other members of the SCOR family provides insight into the residues involved in cofactor and substrate binding.

1. Introduction

Under conditions of low iron, bacteria synthesize and secrete into their environment low-molecular-weight siderophores (Meyer, 2000; Braun & Braun, 2002; Clarke *et al.*, 2001; Quadri, 2000). These compounds bind to Fe^{3+} with very high affinity. The Fe–siderophore complex is then actively transported back into the bacterial cell, where the Fe^{3+} ion is released to be used as a structural or catalytic functional group in a wide variety of proteins.

Escherichia coli synthesizes the enterobactin siderophore (Raymond *et al.*, 2003; Quadri, 2000) that consists of three cyclically linked amides of 2,3-dihydroxybenzoic acid (DHB) and serine (Fig. 1*a*). The three serine molecules are joined through ester linkages, while the DHB residues are free to coordinate the Fe^{3+} ion through catecholate O atoms. The genetic cluster that is responsible for the synthesis of enterobactin contains six genes, *entA–entF*. EntF, EntE and EntB function as non-ribosomal peptide synthetases, large modular enzymes that frequently contain multiple catalytic domains joined as a single large polypeptide (Gehring *et al.*, 1997, 1998; Ehmann *et al.*, 2000; Rusnak *et al.*, 1989; Shaw-Reid *et al.*, 1999). The EntA, EntB and EntC proteins are responsible for the synthesis of DHB from chorismic acid (Liu *et al.*, 1989, 1990; Rusnak *et al.*, 1989, 1990). We have recently determined the structure of EntB, a bifunctional enzyme that catalyzes a step in DHB synthesis and also plays a role in enterobactin synthesis as a carrier-protein domain (Gehring *et al.*, 1997; Drake *et al.*, 2006).

In the formation of DHB, EntC first serves as a isochorismate synthase, converting chorismate to isochorismate (Fig. 1*b*). The isochorismate lyase domain of EntB then catalyzes the conversion of isochorismate into 2,3-dihydro-2,3-

dihydroxybenzoic acid (2,3-DHDHB). This molecule then serves as the substrate for EntA, which oxidizes 2,3-DHDHB to form DHB. Biochemical studies with alternate substrates demonstrate that the EntA reaction proceeds through oxidation of the C3-hydroxyl of the 2*S*,3*S* conformer of 2,3-DHDHB to form the 3-keto group, which then tautomerizes to form the final aromatic DHB product (Sakaitani *et al.*, 1990).

EntA is a member of the short-chain oxidoreductase (SCOR) family of enzymes (Duax *et al.*, 2000). Enzymes from this family catalyze the oxidation, reduction and isomerization of a wide variety of biologically important molecules. The SCOR proteins share a nicotinamide dinucleotide-binding motif which is used in the catalysis of a diverse array of chemical reactions (Lesk, 1995). In this regard, the SCOR enzymes are similar to the enolase family of enzymes that use a divalent-cation interaction to stabilize the removal of the proton from the α-carbon of a carboxylate (Babbitt *et al.*, 1996; Gerlt & Babbitt, 2001; Gulick *et al.*, 2000); the differing fates of the initial enolate intermediate can then be dictated by additional parts of the protein which share less similarity. In the SCOR family of enzymes, the abstraction of the hydride from the substrate creates an intermediate from which epimerization, isomerization, aromatization or additional redox reactions will proceed.

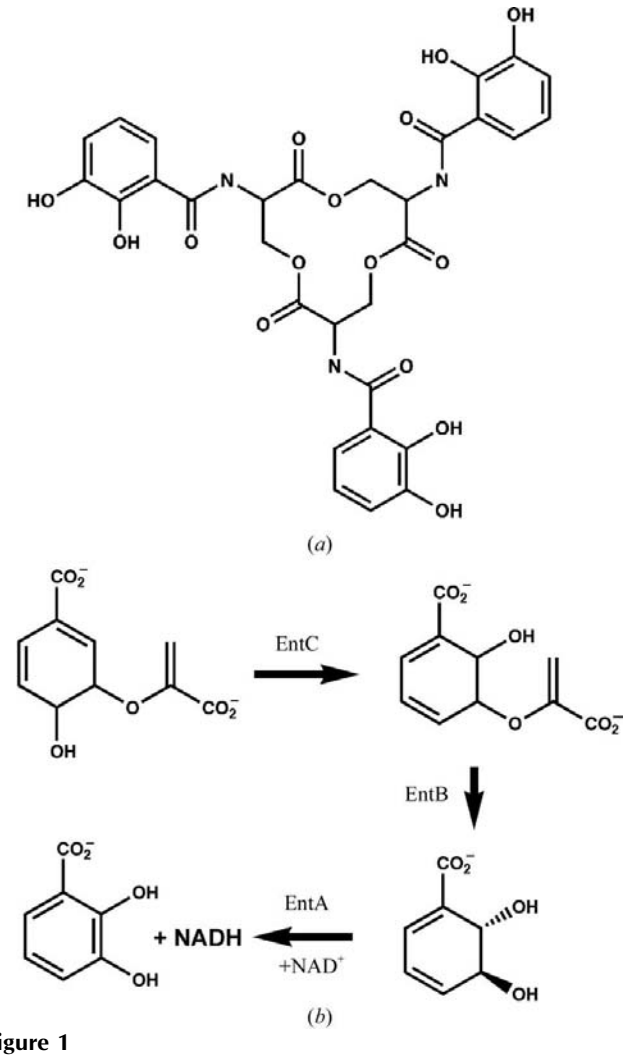

Figure 1
Enterobactin and enzymatic production of 2,3-dihydroxybenzoic acid. (*a*) The structure of the enterobactin molecule is shown, consisting of three serine residues joined by ester linkages. The three 2,3-dihydroxybenzoate groups joined through amide bonds to the serine amino groups coordinate the Fe^{3+} ion. (*b*) The reactions catalyzed by EntC, the isochorismate lyase domain of EntB and EntA catalyze the three-step conversion of chorismate to 2,3-dihydroxybenzoate.

The SCOR enzymes, which range from 250 to 350 residues in length, exhibit a typical Rossmann fold consisting of a central seven-stranded parallel β-sheet surrounded on both sides by α-helices (Duax *et al.*, 2000). The specificity of NAD(H) or NADP(H) is dictated in the majority of proteins by the identity of a residue on the loop following the second strand of the β-sheet (Duax *et al.*, 2003; Pletnev *et al.*, 2004; Thomas *et al.*, 2003). In proteins that utilize NAD(H), a conserved Asp residue is present at position 37 (using the sequence numbering of 3α20β-hydroxysteroid dehydrogenase; 3α20βHSD). In contrast, proteins that prefer NADP(H) contain an Arg residue that is conserved at position 38 which interacts with the 2′-phosphate of the NADP cofactor. Analysis of the family shows that ∼39% of the proteins contain an Asp at position 37, while 34% of the proteins contain an Arg at position 38 (Duax *et al.*, 2003). Indeed, it is possible to alter the cofactor preference while retaining dehydrogenase activity by mutagenic replacement of the residues at positions 37 and 38 as demonstrated in the human 3β-hydroxysteroid dehydrogenase (Thomas *et al.*, 2003).

An analysis of a subfamily of SCOR proteins with a TG*XXX*GIG signature at this position has recently been used to identify a fingerprint of 40 residues that are conserved at greater than 70% within this subfamily (Duax *et al.*, 2003). Analysis of these fingerprint residues allows placement of an unknown sequence into the SCOR family and serves as a starting point for the identification of biochemical functions. The SCOR family is an example of a large family of related proteins that share certain features, yet exhibit the ability to catalyze a variety of reactions. A combination of bioinformatic, functional and structural analyses will in some cases allow the prediction of function for uncharacterized members of the family.

To provide additional insight into the structures of SCOR proteins, we have solved the 2.0 Å crystal structure of EntA, a member of a structurally uncharacterized subfamily of SCOR proteins, by multiwavelength anomalous dispersion phasing methods. Although a number of SCOR proteins have been crystallographically studied, the closest homolog to EntA shares only 32% sequence identity. We present here the structure of the unliganded EntA protein, the interface used to form the tetrameric protein and analysis of the fingerprint residues conserved in the cofactor-binding site. We have modeled the cofactor into the EntA active site based on homologous structures. Additionally, analysis of the covariance of residues in the SCOR family has identified conserved residues that are proposed to play a role in substrate specificity (Duax *et al.*, in preparation; Pletnev & Duax, 2005). This analysis has been extended to EntA and allows us to predict

Table 1
Crystallographic data.

Values for the highest resolution shell are given in parentheses.

	Peak	Inflection	Remote	Native
Wavelength (Å)	0.97910	0.97934	0.964108	0.979331
Resolution (Å)	2.0	2.0	2.0	2.0
Unit-cell parameters				
a (Å)	58.77	58.75	58.75	58.59
b (Å)	66.60	66.62	66.63	66.64
c (Å)	109.86	109.89	109.90	109.73
$R_{\rm merge}$† (%)	5.6 (12.6)	4.7 (11.3)	4.8 (13.2)	5.9 (29.8)
Completeness (%)	99.8 (100.0)	99.8 (100.0)	99.8 (100.0)	95.4 (70.0)
$I/\sigma(I)$	23.5 (5.8)	26.5 (6.5)	23.0 (5.2)	13.6 (1.9)
No. of observations	72109	74750	75077	61570
No. of reflections	14864	14807	14826	14485
Phasing power				
Anomalous	4.72	3.08	3.04	
Isomorphous	1.01	2.67		

† $R_{\rm merge} = \sum(|I_{hi} - I_h|)/\sum I_{hi}$, where I_{hi} and I_h are individual and mean intensities of all equivalent reflections, respectively. † Anomalous and isomorphous phasing powers are defined as the mean value of F_h''/E and F_h/E, where F_h'' is the calculated anomalous scattering structure factor, F_h is the calculated heavy-atom structure factor and E is the lack-of-closure error.

the residues that form the 2,3-DHDHB substrate-binding pocket.

2. Materials and methods

2.1. Cloning, protein expression and purification

The *entA* gene was cloned from genomic DNA from *Escherichia coli* JM109. 10 µl of an overnight culture of bacteria were diluted to 100 µl with H_2O and boiled for 5 min followed by centrifugation for 5 min. The soluble material was used as template for a PCR reaction with forward and reverse primers 5′-CGACCG**CATATG**GATTTCAGCGGTAAAA-ATG-3′ and 5′-GCG**CTCGAG**TTATGCCCCCAGCGTT-GAG-3′, respectively, where the sequences in bold represent the *Nde*I and *Xho*I restriction sites used for cloning purposes. The PCR product was subcloned into a modified pET15b plasmid that encoded a TEV protease site to replace the thrombin site (Kapust *et al.*, 2001). The entire coding sequence was confirmed by DNA sequencing.

The EntA protein was produced in BL21(DE3) cells by growing cells to an OD_{600} of ∼0.6 and inducing with 1.0 m*M* IPTG. Induction continued overnight at 290 K. Cells were harvested by centrifugation and the wet paste was frozen in liquid nitrogen and stored at 193 K. Protein was purified by two IMAC chromatography steps using 5 ml Ni^{2+} HisTrap columns and an intervening proteolysis step to remove the His tag. Cells were thawed and sonicated in lysis buffer (50 m*M* HEPES pH 8.0 at 277 K, 150 m*M* NaCl, 10 m*M* imidazole, 0.2 m*M* TCEP) and loaded onto the column. The unbound material was washed with 10 column volumes of lysis buffer followed by a wash with lysis buffer containing 50 m*M* imidazole. Finally, the protein was eluted with the same buffer containing 300 m*M* imidazole. The EntA protein was dialyzed overnight into TEV cleavage buffer (50 m*M* HEPES pH 8.0 at 277 K, 150 m*M* NaCl, 1 m*M* NaN_3, 0.5 m*M* EDTA); fresh buffer was added the next day along with TEV protease at a ∼1:100 mass ratio of total protein. The cleavage reaction continued at 277 K in the dialysis bag for 24 h. The digested protein was supplemented with 10 m*M* imidazole and then passed over the Ni^{2+} column a second time; material in the flowthrough fractions was pooled and dialyzed against final crystallization buffer (50 m*M* HEPES pH 8.0 at 277 K, 50 m*M* NaCl, 0.5 m*M* TCEP, 1 m*M* NaN_3). The protein was concentrated to ∼10 mg ml^{-1} and frozen directly in liquid nitrogen.

To produce SeMet-labeled protein, cells were grown in M9 minimal media and SeMet incorporation was induced by the metabolic inhibition method (Doublié, 1997; Van Duyne *et al.*, 1993). The purification procedure was identical to the procedure used for production of native protein.

2.2. Crystallization and data collection

Crystallization conditions for EntA were identified by sparse-matrix screening (Carter & Carter, 1979; Jancarik & Kim, 1991) using an in-house screen. Crystals of EntA were grown by vapor diffusion using the hanging-drop protocol. SeMet-labeled protein was combined in a 1:1 ratio with mother liquor and incubated above a precipitant consisting of 9% PEG 4000, 250–350 m*M* $(NH_4)_2SO_4$, 2% ethylene glycol, 50 m*M* sodium succinate pH 5.0. Native crystals of the same morphology were grown in 2% PEG 4000, 2% ethylene glycol, 250–350 m*M* $(NH_4)_2SO_4$, 50 m*M* sodium succinate pH 5.0. The crystals were transferred to solutions of mother liquor containing increasing concentrations of ethylene glycol and frozen directly in a stream of cryocooled nitrogen gas; crystals diffracted to better than 2 Å resolution on a home source. The space group was determined to be *I*222.

Data were collected on native and SeMet crystals at CHESS beamline F2. Diffraction was strong to 2 Å. Three-wavelength MAD data were collected. At each wavelength, 135 frames of data were collected at 1° and 30 s per frame using an ADSC Quantum 210 CCD detector set at a distance of 170 mm. Data were integrated and scaled using *HKL*2000 (Otwinowski & Minor, 1997). For the native data set, the detector distance was set to 210 mm and 125 frames were collected with 1° oscillation and 60 s exposure time per frame. Final data-collection statistics are shown for the MAD and native data in Table 1.

2.3. Structure determination and refinement

The positions of the Se heavy atoms were determined with *BnP* running in the auto mode (Weeks *et al.*, 2002). Five Se positions were located (out of a possible six). Phase refinement and solvent flattening from the *PHASES* (Furey & Swaminathan, 1997) portion of *BnP* was used to improve the final phases (mean figure of merit = 0.74). The output phases were imported into *RESOLVE* (Terwilliger, 2000, 2001) for automated chain tracing. Of the 248 residues in the EntA protein, *RESOLVE* was able to identify 175 residues with side chains; an additional seven residues were inserted as poly-alanine. The remainder of the chain was modeled through manual model building and refinement with *REFMAC*

Table 2
Refinement statistics.

Resolution range (Å)	40.0–2.0
R_{cryst} (working set) (%)	18.6 (12933)†
R_{cryst} (highest resolution shell) (%)	18.8 (727)†
R_{free} (overall) (%)	21.4 (694)†
R_{free} (highest resolution shell) (%)	18.4 (41)†
Wilson *B* factor (Å^2)	21.5
Average *B* factor, overall (Å^2)	21.5
Average *B* factor (Å^2)	
Protein	21.0
Main chain	20.3
Side chain	21.9
Average *B* factor, solvent (Å^2)	28.0 (109)‡
R.m.s. deviations	
Bond lengths (Å)	0.008
Angles (°)	1.08

† The number of reflections used is shown in parentheses. ‡ The total number of atoms used in calculation is given in parentheses.

(Collaborative Computational Project, Number 4, 1994; Murshudov *et al.*, 1997).

The final model contains residues 1–177 and 214–248. The model also contains 109 water molecules. The final Ramachandran plot (Laskowski *et al.*, 1993) contains 92.3% residues in the most favored region, 6.6% in the allowed region and 1.1% in the generously allowed region. There are no residues in the disallowed region. Surface-area calculations were determined with *CCP*4 using the Lee and Richards algorithm (Lee & Richards, 1971) and a 1.4 Å molecular probe. Modeling of the cofactor into the active site was performed by aligning the structure for 7α-hydroxysteroid dehydrogenase (PDB code 1fmc) on the structure of EntA using homologous residues. The r.m.s. deviation between EntA and 1fmc over 176 homologous C$^\alpha$ atoms is 1.4 Å.

3. Results and discussion

3.1. Structure determination of the EntA protein

The crystal structure of the EntA protein was determined by multi-wavelength anomalous dispersion methods using SeMet-labeled protein. The positions of five of the six heavy atoms were determined with *BnP* (Weeks *et al.*, 2002). The resultant electron-density map was of excellent quality (Fig. 2). Automated residue tracing with *RESOLVE* was used to place ~70% of the chain; the remainder of the refinement was performed using iterative cycles of manual model building and refinement. The final model contains residues 1–177 and 214–248; the side chains of Lys15, Glu42, Thr177 and Glu217 are disordered and these residues are modeled as Ala. The large disordered loop at residues 178–213 is a region that is partly responsible for the binding of substrates (see below). This region, located between secondary-structural elements β6 and α7, has been seen to adopt different orientations in the presence or absence of substrates and also adopts different orientations in different proteins (Tanaka *et al.*, 1996; Zhang *et al.*, 2005; Thoden *et al.*, 2000). In the EntA structure determined in the absence of ligands, the region is completely disordered. Crystallographic and refinement data are presented in Tables 1 and 2.

3.2. Structure of the EntA monomer

The final structure demonstrates the typical Rossmann fold observed for other members of the SCOR family (Fig. 3*a*). The structure contains a central seven-stranded parallel β-sheet with β3-β2-β1-β4-β5-β6-β7 topology. As has been observed for other members of the SCOR family, there is some variability in the helices joining the seven strands. In EntA, three helices are positioned on one face of the β-sheet (α2, α4 and α7), while the remaining two helices (α5 and α6) are on the opposite face of the sheet and form one subunit interface. Helix α2 is located on the loop joining strands β1 and β2; helix α4 is located between strands β3 and β4. Finally, helix α7 immediately follows the disordered substrate-binding loop joining strands β6 and β7.

3.3. Oligomeric structure of the EntA protein

The EntA protein forms a tetramer with dimensions of 65 × 69 × 43 Å that displays crystallographic 222 symmetry (Fig. 3*b*), a tertiary arrangement observed for other members of the SCOR family. In forming the complete tetramer, each subunit buries 2413 Å^2 of surface, which represents ~24% of the total monomeric surface area. Tetrameric members of this family have been described as dimers of dimers, with each subunit forming a strong interaction with one other subunit. To form these dimers, each subunit interacts with another subunit by forming a four-helix bundle composed of helices 5 and 6. The surface area of each subunit in this interaction is 1322 Å^2. Residues that contribute to this interaction reside almost entirely on these two helices; however, Thr88 and Asp89, which lie on the loop immediately preceding α5, also contribute to the interaction. The dimer formed around these helices interacts with another dimer to form the complete tetramer. Residues that form these latter interactions reside on α7, β7, the loops that join these elements and the C-terminal residues that follow β7. The surface area of this interface is 995 Å^2 for each subunit.

3.4. Analysis of proposed cofactor- and ligand-binding sites

A core fingerprint of 40 residues has been identified in 426 proteins that exhibit greater than 30% sequence identity to 3α20βHSD (Ghosh *et al.*, 1994), a bacterial SCOR protein with hydroxysteroid dehydrogenase activity. Each of the fingerprint residues is conserved in greater than 70% of this subfamily of SCOR proteins (Duax *et al.*, 2003). Of the 40 identified residues, 32 are conserved in the EntA sequence. The key cofactor-binding residues are all present (Table 3). EntA contains an aspartic acid residue, Asp36, that is consistent with the biochemical demonstration that EntA uses NAD$^+$ as the cofactor (Sakaitani *et al.*, 1990; Young & Gibson, 1969).

Of the 40 SCOR family fingerprint residues, eight residues are not conserved in EntA. The EntA residues are listed here, with the predicted fingerprint residue shown in parentheses:

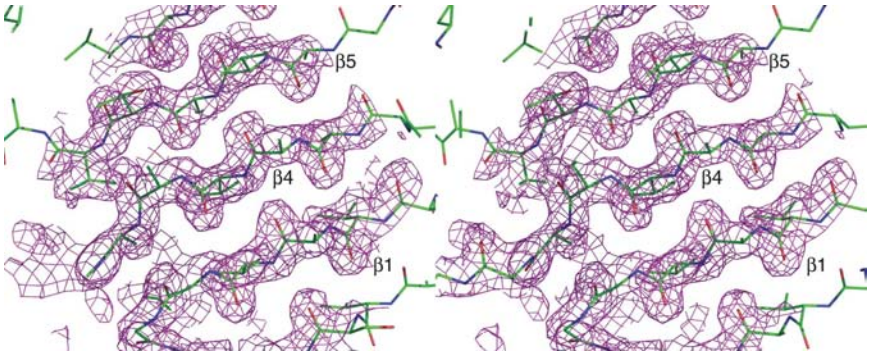

Figure 2
Representative experimental electron density. Density is shown from the refined solvent-flattened experimental map using observed remote structure factors and phases derived from *BnP*. The map, contoured at 1σ, shows a region of the central β-sheet.

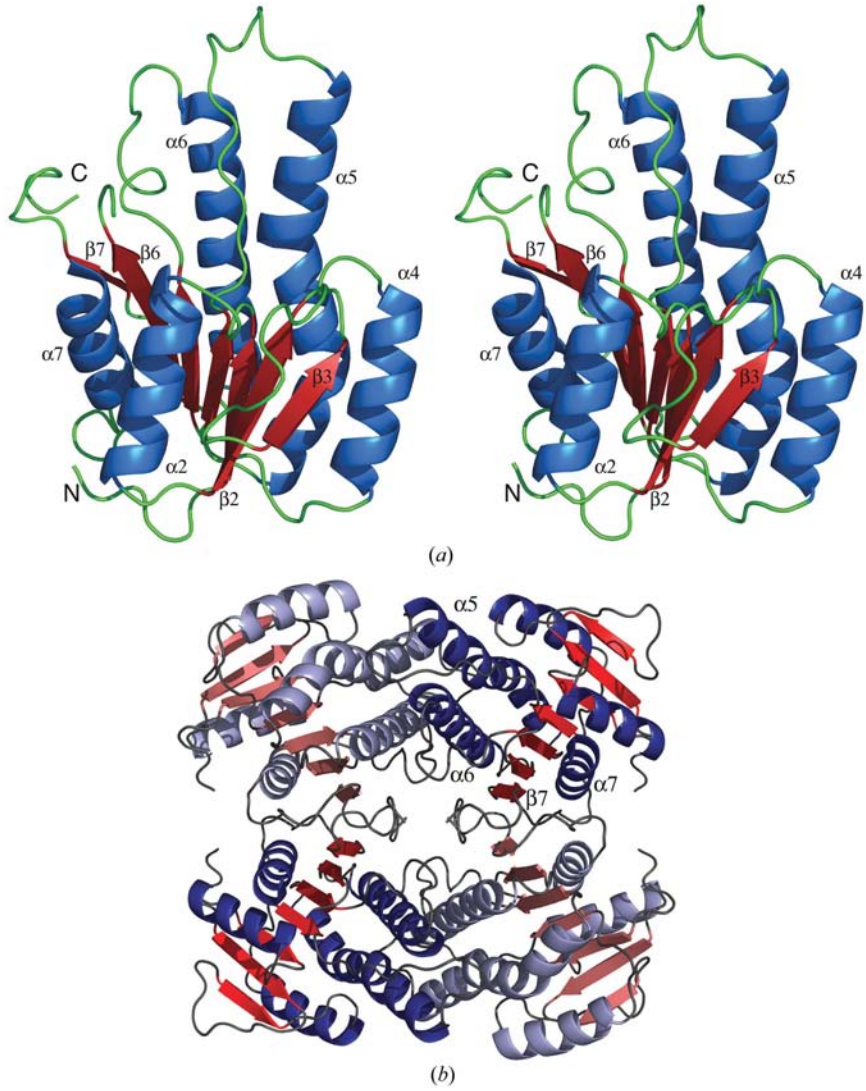

Figure 3
Structure of the EntA monomer and tetramer. (*a*) Ribbon diagram of the EntA monomer. The central β-sheet is shown in red, while the surrounding helices are shown in blue. The secondary-structural elements are labeled. Note that helices α1 and α3 are not present in the EntA structure; in accordance with secondary-structure nomenclature for prior SCOR proteins (Duax *et al.*, 2000), the helices are labeled α2, α4, α5, α6, and α7. (*b*) The EntA tetramer is shown viewed down one twofold axis of 222 crystallographic symmetry. The two top and two bottom subunits interact around the four-helix bundle formed by helices α5 and α6. The C-terminal loops and β7 form interactions between dimers.

Glu71 (Gly), Ala79 (Asn), Thr128 (Asn), Gly159 (Ala) Val167 (Ile), Cys169 (Val), Lys211 (Arg) and Leu235 (Gly). We compared the structure of EntA with 3α20βHSD to provide insight into the role of allowable substitutions within the SCOR family.

Glu71 is normally a glycine residue in other SCOR family members. This residue occurs on a surface loop of the protein that is directed towards solvent in the tetramer. In the structure of 3α20βHSD, the glycine residue at this position is necessary to exhibit φ, ψ angles of 73, -21°. This loop is shorter in EntA by one residue and adopts a slightly different orientation, allowing Glu71 to adopt the more preferred φ, ψ angles of -78, -35°.

Ala79 of EntA is predicted to be an Asn residue. This Asn residue makes water-mediated interactions with the cofactor phosphate oxygen as well as a 3′-hydroxyl. It is unclear why EntA does not retain this seemingly important binding residue. It is worth noting, however, that some SCOR family proteins copurify with a dinucleotide cofactor bound and crystal structures of binary complexes have in several instances been determined when no exogenous cofactor was added (He *et al.*, 1996; Zhang *et al.*, 2005). The absence of cofactor in the purified EntA protein may suggest that it has a lower affinity for the cofactor than other SCOR family members, perhaps as a result of this Ala79 replacement.

Thr128 of EntA is located on β5 of the central sheet; however, it is pointed away from the active site and the standard fingerprint residue (Asn) makes no specific contacts in the structure of 3α20βHSD.

Two residues that differ from the standard fingerprint, Gly159 and Leu235 in EntA, are located in close proximity to each other in the folded protein. The standard fingerprint residue for position Gly159 is an alanine and for Leu235 is a glycine. The side chains of these two residues are directed toward each other and thus the Leu235 side chain partly occupies the space normally occupied by Ala159 in other SCOR proteins. This is an interesting case of compensatory mutations occurring to fill the same space in three dimensions.

The remaining residues that differ from the standard fingerprint identity are either minor conserved replacements (Val167 is normally an isoleucine and Cys169 is

Table 3
Conserved residues that are likely to interact with the cofactor.

3α20βHSD residue	EntA residue	Location	Interaction with cofactor
Asp60	Asp52	β3 loop	O^{δ} to N6 of NAD adenine ring
Asn86	Asn78	β4 loop	Side chain to phosphate oxygen
Ser139	Ser131	β5 loop	O^{γ} to C5 of nicotinamide ring
Tyr152	Tyr144	α6	Side-chain OH to 2′ OH of nicotinamide ribose
Lys156	Lys148	α6	Side-chain amino to 2′ and 3′ OH of nicotinamide ribose
Thr187	Thr179	β6 loop	Side-chain OH to phosphate oxygen and NAD amide

normally a valine) or are located on the disordered loop (Lys211), making any discussion of the effects of the change rather difficult.

3.5. Residues likely to be involved in substrate binding

A recent analysis of residue covariance has allowed the SCOR proteins to be divided into subfamilies on the basis of their substrate-binding site (Duax *et al.*, in preparation; Pletnev & Duax, 2005). This procedure involved the comparison of residues that co-vary within the SCOR superfamily to identify subfamilies that share substrate specificity. This analysis is similar in some respects to the analysis of amino acyl substrate-binding specificity within the NRPS adenylation domains; in this protein family, construction of ten-residue 'pseudo-peptides' using crystallographically determined substrate-binding residues allowed the authors to cluster adenylation domains on the basis of substrate specificity (Stachelhaus *et al.*, 1999). What makes the SCOR family covariance analysis more powerful, however, is that the residues can be identified with a bare minimum of preliminary structural information. The residues that co-vary within the large family can be clustered on the basis of shared substrate-binding residues and then correlated with known functional data.

This analysis of the SCOR family identified a subfamily which includes the EntA protein. The residues predicted to form the 2,3-DHDHB-binding pocket are Leu83, Met85, Arg138, Gly140, Met141, Ser176, Met181, Gln182 and Leu185. While the latter three residues are located on the disordered substrate-binding loop between β6 and α7, the remainder of the residues do cluster around the binding pocket seen in other SCOR family structures (Fig. 4). Arg138, the only positively charged residue, is a likely candidate to interact with the carboxylate of 2,3-DHDHB. Interactions with the carboxylate group provide significant contributions to the reaction and the enzyme is able to catalyze the reaction with the decarboxy substrate analog; however, K_m is increased by ∼50-fold and k_{cat} is decreased by ∼90-fold (Sakaitani *et al.*, 1990). Biochemical evidence demonstrates that the EntA-catalyzed reaction proceeds with stereospecific oxidation at the C3 position (Sakaitani *et al.*, 1990). Potential polar side chains that may interact with the substrate 2-hydroxyl from this reaction-specific signature sequence include Ser176 and Gln182. Alternatively, it is possible that the residues that interact with the 2-hydroxyl may be conserved residues that are part of the SCOR family fingerprint residues. Residues that have been implicated as playing multiple roles in other enzymes include the catalytic tyrosine (Tyr144), Lys148 and the conserved Ser (Ser131) at the end of strand β5.

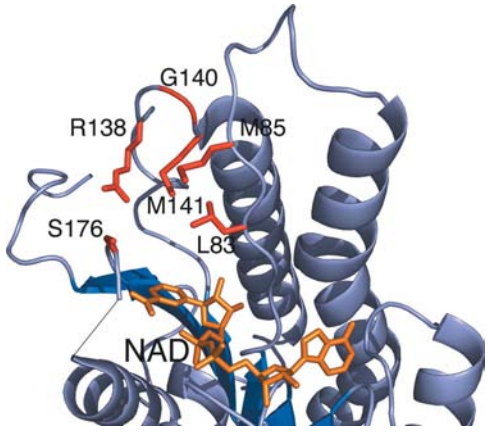

Figure 4
Proposed substrate-binding pocket of EntA. The residues of the EntA protein predicted by covariance analysis to contribute to 2,3-DHDHB-specific binding are labeled. The cofactor molecule (shown in orange) was modeled into the unliganded EntA structure by superimposing the 7α-hydroxysteroid dehydrogenase (PDB code 1fmc) structure onto the EntA molecule.

4. Summary

The structure of EntA provides further insight into the SCOR family of proteins. While further structural analysis of binary and tertiary complexes will continue to provide additional information, the use of the structure of the unliganded protein and bioinformatic analysis of prior family members provides insight into the residues that form the active site and provides insight into the diversity of amino-acid residues that are compatible within the active site. Analysis of the substrate-binding residues within EntA, a member of a newly characterized subfamily of SCOR enzymes, will be useful to further characterize newly defined members with as yet undetermined biochemical function.

This research was supported in part by NIH grant GM-068440 (AMG) and NIH grant DK-026546 (WLD). JAG was supported by a summer undergraduate research fellowship from the Max and Victoria Dreyfus Foundation Inc. This work is based upon research conducted at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the National Science Foundation under award DMR 0225180 and

the National Institutes of Health through its National Center for Research Resources under award 5 P41 RR001646-23.

References

Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L. & Gerlt, J. A. (1996). *Biochemistry*, **35**, 16489–16501.

Braun, V. & Braun, M. (2002). *Curr. Opin. Microbiol.* **5**, 194–201.

Carter, C. W. Jr & Carter, C. W. (1979). *J. Biol. Chem.* **254**, 12219–12223.

Clarke, T. E., Tari, L. W. & Vogel, H. J. (2001). *Curr. Top. Med. Chem.* **1**, 7–30.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.

Drake, E. J., Nicolai, D. A. & Gulick, A. M. (2006). *Chem. Biol.* **13**, 409–419.

Duax, W. L., Ghosh, D. & Pletnev, V. (2000). *Vitam. Horm.* **58**, 121–148.

Duax, W. L., Pletnev, V., Addlagatta, A., Bruenn, J. & Weeks, C. M. (2003). *Proteins*, **53**, 931–943.

Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C. & Walsh, C. T. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 2509–2514.

Furey, W. & Swaminathan, S. (1997). *Methods Enzymol.* **277**, 590–620.

Gehring, A. M., Bradley, K. A. & Walsh, C. T. (1997). *Biochemistry*, **36**, 8495–8503.

Gehring, A. M., Mori, I. & Walsh, C. T. (1998). *Biochemistry*, **37**, 2648–2659.

Gerlt, J. A. & Babbitt, P. C. (2001). *Annu. Rev. Biochem.* **70**, 209–246.

Ghosh, D., Wawrzak, Z., Weeks, C. M., Duax, W. L. & Erman, M. (1994). *Structure*, **2**, 629–640.

Gulick, A. M., Hubbard, B. K., Gerlt, J. A. & Rayment, I. (2000). *Biochemistry*, **39**, 4590–4602.

He, X., Thorson, J. S. & Liu, H. W. (1996). *Biochemistry*, **35**, 4721–4731.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). *Protein Eng.* **14**, 993–1000.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

Lesk, A. M. (1995). *Curr. Opin. Struct. Biol.* **5**, 775–783.

Liu, J., Duncan, K. & Walsh, C. T. (1989). *J. Bacteriol.* **171**, 791–798.

Liu, J., Quinn, N., Berchtold, G. A. & Walsh, C. T. (1990). *Biochemistry*, **29**, 1417–1425.

Meyer, J. M. (2000). *Arch. Microbiol.* **174**, 135–142.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

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

Pletnev, V. Z. & Duax, W. L. (2005). *J. Steroid Biochem. Mol. Biol.* **94**, 327–335.

Pletnev, V. Z., Weeks, C. M. & Duax, W. L. (2004). *Proteins*, **57**, 294–301.

Quadri, L. E. (2000). *Mol. Microbiol.* **37**, 1–12.

Raymond, K. N., Dertz, E. A. & Kim, S. S. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 3584–3588.

Rusnak, F., Faraci, W. S. & Walsh, C. T. (1989). *Biochemistry*, **28**, 6827–6835.

Rusnak, F., Liu, J., Quinn, N., Berchtold, G. A. & Walsh, C. T. (1990). *Biochemistry*, **29**, 1425–1435.

Sakaitani, M., Rusnak, F., Quinn, N. R., Tu, C., Frigo, T. B., Berchtold, G. A. & Walsh, C. T. (1990). *Biochemistry*, **29**, 6789–6798.

Shaw-Reid, C. A., Kelleher, N. L., Losey, H. C., Gehring, A. M., Berg, C. & Walsh, C. T. (1999). *Chem. Biol.* **6**, 385–400.

Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. (1999). *Chem. Biol.* **6**, 493–505.

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

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Terwilliger, T. C. (2001). *Acta Cryst.* D**57**, 1755–1762.

Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L. & Holden, H. M. (2000). *Biochemistry*, **39**, 5691–5701.

Thomas, J. L., Duax, W. L., Addlagatta, A., Brandt, S., Fuller, R. R. & Norris, W. (2003). *J. Biol. Chem.* **278**, 35483–35490.

Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.

Weeks, C. M., Blessing, R. H., Miller, R., Mungee, R., Potter, S. A., Rappleye, J., Smith, G. D., Xu, H. & Furey, W. (2002). *Z. Kristallogr.* **217**, 686–693.

Young, I. G. & Gibson, F. (1969). *Biochim. Biophys. Acta*, **177**, 401–411.

Zhang, J., Osslund, T. D., Plant, M. H., Clogston, C. L., Nybo, R. E., Xiong, F., Delaney, J. M. & Jordan, S. R. (2005). *Biochemistry*, **44**, 6948–6957.