PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Sequence fingerprint and structural analysis of the SCOR enzyme A3DFK9 from *Clostridium thermocellum*

Robert Huether,[1] Zhi-Jie Liu,[2] Hao Xu,[3] Bi-Cheng Wang,[3] Vladimir Z. Pletnev,[4,5] Qilong Mao,[4] William L. Duax,[1,4]* and Timothy C. Umland[1,4]*

[1] Department of Structural Biology, SUNY at Buffalo, Buffalo, New York 14203
[2] National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[3] Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602
[4] Hauptman-Woodward Medical Research Institute, Buffalo, New York 14203
[5] Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117997 Moscow, Russia

ABSTRACT

We have identified a highly conserved fingerprint of 40 residues in the TGYK subfamily of the short-chain oxidoreductase enzymes. The TGYK subfamily is defined by the presence of an N-terminal TGxxxGxG motif and a catalytic YxxxK motif. This subfamily contains more than 12,000 members, with individual members displaying unique substrate specificities. The 40 fingerprint residues are critical to catalysis, cofactor binding, protein folding, and oligomerization but are substrate independent. Their conservation provides critical insight into evolution of the folding and function of TGYK enzymes. Substrate specificity is determined by distinct combinations of residues in three flexible loops that make up the substrate-binding pocket. Here, we report the structure determinations of the TGYK enzyme A3DFK9 from *Clostridium thermocellum* in its apo form and with bound NAD^+ cofactor. The function of this protein is unknown, but our analysis of the substrate-binding loops putatively identifies A3DFK9 as a carbohydrate or polyalcohol metabolizing enzyme. *C. thermocellum* has potential commercial applications because of its ability to convert biomaterial into ethanol. A3DFK9 contains 31 of the 40 TGYK subfamily fingerprint residues. The most significant variations are the substitution of a cysteine (Cys84) for a highly conserved glycine within a characteristic VNNAG motif, and the substitution of a glycine (Gly106) for a highly conserved asparagine residue at a helical kink. Both of these variations occur at positions typically participating in the formation of a catalytically important proton transfer network. An alternate means of stabilizing this proton wire was observed in the A3DFK9 crystal structures.

Proteins 2010; 78:603–613.

Key words: SCOR; SDR; cofactor binding; proton transfer; fingerprint variation; A3DFK9; Cthe_1510.

INTRODUCTION

The short-chain oxidoreductase (SCOR) superfamily has members widely distributed across viruses, bacteria, and eukaryotes.[1,2] The Pfam database identifies more than 18,000 members of the SCOR family within Pfam *adh_short* (PF00106).[3,4] SCOR enzymes are responsible for regulating many vital processes including the oxidation and/or reduction of alcohols, aldehydes, and ketones, and they play key roles in steroid synthesis and metabolism.[2] SCOR enzymes are grouped within the Enzyme Commission (E. C.) class 1.1.1.*, and analysis of these E.C. numbers indicated that the SCOR family used more than 300 different substrates.[5] Members are characterized as containing between 250 and 350 amino acids, binding NAD(H) or NADP(H) cofactors, and exhibiting the classic α/β Rossmann fold with seven parallel β-strands. Despite all SCOR proteins possessing the same overall fold, very low sequence identity (typically ∼15%) exists between nonorthologous proteins.[6] Such low sequence similarity between members plus the wide species distribution indicates extensive evolutionary divergence. It has been postulated that the presence of multiple open reading frames and a severe codon bias within 20% of the SCOR members is an additional evidence of their early ancestry.[7]

Not one amino acid is fully conserved across the entire SCOR family. However, the SCOR family can be divided into several subfamilies based on characteristic N-terminal motifs. The largest SCOR subfamily contains a TGxxxGxG motif near the N-terminus and a YxxxK motif in its catalytic site, and

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: NIH; Grant number: DK026546.
*Correspondence to: William L. Duax, Hauptman-Woodward Medical Research Institute, 700 Ellicott St., Buffalo, NY 14203. E-mail: duax@hwi.buffalo.edu or Timothy C. Umland, Hauptman-Woodward Medical Research Institute, 700 Ellicott St., Buffalo, NY 14203. E-mail: umland@hwi.buffalo.edu
Received 26 March 2009; Revised 24 July 2009; Accepted 7 August 2009
Published online 21 August 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22584

thus will be termed the TGYK subfamily. This subfamily contains ~75% of the entire SCOR family, and 40 highly conserved residues have been previously identified as TGYK fingerprint residues.[8,9] These 40 amino acids are implicated in protein folding, cofactor binding, catalysis, and oligomerization.[8,10–13]

Clostridium thermocellum is a thermophilic bacterium that has an optimum growth temperature of 60°C.[14] This bacterium is of biological and commercial importance because it is capable of directly converting cellulose biomass into sugars, which are then fermented into ethanol though its cellulosome.[14,15] A3DFK9 (UniProt nomenclature) is a 247-residue SCOR enzyme coded by the *C. thermocellum* gene Cthe_1510. It contains 31 of the 40 highly conserved fingerprint residues of the TGYK subfamily. Similar to the majority of the SCOR family members, A3DFK9 has not been biochemically characterized. Rather, database annotations for this protein are inferred and are generic descriptions (e.g., oxidoreductase activity).

The A3DFK9 sequence contains several variations from the TGYK subfamily fingerprint not previously observed in an experimentally determined three-dimensional (3D) structure of a SCOR enzyme. To evaluate the importance the fingerprint variations and to predict the substrate type, we undertook the X-ray crystal structure determination of A3DFK9 and A3DFK9 with its predicted cofactor, NAD^+.

METHODS

Expression and purification

A pET-based maltose-binding protein (MBP) fusion expression vector (pET-MBP) was constructed containing the gene Cthe_1510 from *C. thermocellum* (strain ATCC 27405), and was provided by the SouthEast Collaboratory For Structural Genomics (SECSG Target ID Cth-1068). This expression vector contained a N-terminal $(His)_6$-MBP affinity tag followed by a TEV protease cleavage site. *E. coli* (Rosetta(DE3); Novagen) were transformed with pET-MBP-Cthe_1510, and cells were cultured to an optical density at 600 nm of ~0.6 in LB with ampicillin (0.1 mg/mL). Expression was induced with isopropyl-β-D-1-thiogalactopyranoside (0.1 mg/mL) and occurred overnight at 25°C.

Harvested cells were resuspended in cell-lysis buffer (50 m*M* Tris pH 8.0, 1 m*M* DTT, 500 m*M* NaCl, and 10 m*M* imidazole) plus protease inhibitor cocktail (Sigma) added at 1 mL per cell pellet from 1 L of culture. Following lysis via microfluidizer (Microfluidics), cell debris was removed by centrifugation at 35,000*g* for 1 h at 4°C. The A3DFK9 protein was purified using Ni-NTA Superflow chromatography media (Qiagen) and eluted using imidazole. Separation of the His-MBP tag from the A3DFK9 protein was accomplished through a TEV protease cleavage site following the His-MBP affinity tag. TEV protease was added to a protein:TEV protease molar ratio of 20:1 and incubated overnight at 20°C. A second Ni-NTA column was run to remove MBP and protease from the protein. Polishing was achieved with a Superdex 200 16/60 gel filtration column (GE Healthcare) in 20 m*M* Hepes pH 7.0, 100 m*M* NaCl, 5% glycerol (v/v), and 1 m*M* DTT. The fractions that contained A3DFK9 were pooled. SDS-PAGE (BioRad) was used to test sample purity, and protein concentration was measured by the Bradford assay with BSA as the standard.[16]

Crystallization

The protein solution for crystallization had a concentration of 30 mg/mL in gel filtration buffer. Initial crystallization experiments were conducted using the high throughput crystallization robot at the Hauptman Woodward Institute[17] and manual screens. The optimized condition for growth of diffraction quality crystals by hanging drop vapor diffusion used a reservoir cocktail containing 25% (w/v) polyethylene glycol 4000, 100 m*M* TAPS pH 9.0, and 50 m*M* sodium thiosulfate pentahydrate and a protein droplet containing a 1:1 reservoir cocktail to protein ratio to a final volume of 4 μL. Large crystals (1 mm × 0.5 mm × 0.5 mm) grew within 1 week. The crystals were soaked for 5 min in the reservoir cocktail plus a cryoprotectant of 25% (v/v) 2-methyl-2,4-pentanediol. NAD^+ was introduced into the crystals of the apo form by soaking the crystals in cryoprotectant plus 2 m*M* NAD^+ for 5 min.

Data collection and refinement

Diffraction data were collected via remote control on beam line 11-1 equipped with an ADSC Quantum 315 CCD detector at the Stanford Synchrotron Radiation Laboratory. Data were collected through the Blu-Ice interface[18] with a strategy calculated in Mosflm.[19] The reflections were indexed in the I centered tetragonal space group $I4_122$. Intensities were indexed, integrated, and scaled with HKL2000.[20]

Initial phases for the apo enzyme crystal were obtained through molecular replacement (MR) using AMoRe.[21] The search structure was a hybrid of the conserved Rossmann fold of two SCOR structures 1YDE (29% identity) and 1VL8 (29% identity). After an initial MR solution was identified, a simulated annealing refinement step was preformed using CNS,[22] and further refinement continued with Refmac and PHENIX.[19,23] Model building was performed using Coot.[24] Initial phases for the A3DFK9-NAD^+ enzyme complex were calculated using the isomorphous apo A3DFK9 model. Following rigid body refinement, the A3DFK9-NAD^+ structure was refined similarly to the apo structure. TLS refinement was applied to both structures with each chain treated as a

Table I
Crystal and Diffraction Data Statistics

Crystal data	Apo	Holo
Beam line	SSRL 11-1	SSRL 11-1
Space group	$I4_122$	$I4_122$
Cell dimensions (Å)		
a	124.3	124.3
b	124.3	124.3
c	162.6	161.2
Data collection		
Resolution (Å)[a]	49.69–1.69 (1.74–1.69)	43.90–2.10 (2.16–2.10)
Number of observed reflections	833817	412463
Number of unique reflections	70028	36980
Completeness (%)[a]	97.9 (97.2)	99.8 (100)
R_{merge}[a] (%)	6.4 (30.20)	7.8 (31.70)
Redundancy[a]	11.9 (2.5)	11.2 (5.7)
$I/\sigma(I)$[a]	18.9 (4.2)	17.3 (5.60)

[a]Numbers in parentheses are for the high resolution bin.

separate group. The models were validated using MolProbity.[25] Data collection and refinement statistics are presented in Tables I and II.

Images were created with PyMOL.[26] Ligplot was used to generate protein–NAD^+ interaction plots.[27] Structural comparison was done with the SSM server (http://www.ebi.ac.uk/msd-srv/ssm),[28] DaliLite (http://ekhidna.biocenter.helsinki.fi/dali_server),[29] and PyMOL. SCOR sequences used for comparison to A3DFK9 were selected by PROSITE,[30] using the search query "TGxxxGxGx (60,70)NACx(65,75)YxxxK," where the single letters represent conserved amino acids in the TGYK SCOR subfamily connected by linkers of specified lengths.[8] Sequence alignments were performed using Muscle.[31] Protein Data Bank accession number: 3GED and 3GEG.

RESULTS AND DISCUSSION

Overall A3DFK9 structure

The crystal structure of apo A3DFK9 was determined to a resolution of 1.7 Å, whereas that of the A3DFK9-NAD^+ complex was determined to 2.1 Å resolution (Tables I and II). The asymmetric units of both structures contained the protein as a homodimer of 247 amino acids per chain [Fig. 1(a)]. A3DFK9 exhibits the expected Rossmann fold. It contains an alternating α/β secondary structure topology in the order β3 β2 β1 β4 β5 β6 β7 [Fig. 1(b)].[8] For the apo A3DFK9 structure, electron density was absent for residues Gln182, Asn183, and Asn184. Similarly, the NAD^+ bound A3DFK9 structure has disordered residues in the β6α7 loop (Gln182, Asn183, Asn184, and Gln185). The β6α7 loop is associated with substrate binding and is disordered in many SCOR crystal structures. The N-and C-terminal residues were disordered or exhibited high mobility in both structures. Disordered residues were excluded from the refined model.

The biological unit of A3DFK9 is a tetramer composed of the dimer located within the asymmetric unit plus a second dimer related by a crystallographic twofold rotational axis approximately parallel to the central β-sheet of each monomer [Fig. 1(c)]. Many SCOR proteins have been found to be active as tetramers. Superposition of 241 residues of Monomer A from the apo- and NAD^+-bound structures, respectively, yielded a root-mean-square distance (RMSD) of 0.27 Å. NAD^+ binding imposed localized structural rearrangement to the Cα backbone only in the vicinity of residues Ala83 to Arg85. This rearrangement occurred at the location of one of the TGYK subfamily sequence variations observed within A3DFK9 (Cys84 in place of a highly conserved glycine residue).

Bound ligands

NAD^+ was predicted to bind to A3DFK9 and serve as its cofactor by the presence of an aspartate residue (Asp33) 18 residues downstream from the last glycine in the TGxxxGxG motif.[8,32] The crystal structure of A3DFK9 soaked in NAD^+ confirmed this prediction. The NAD^+ molecule in the coenzyme-binding site has the conformation commonly observed within other members of the SCOR family, with the Asp33 side chain interacting with the adenine ribose hydroxyls facilitating selective NAD^+ binding (Fig. 2). There are 16 contacts between A3DFK9 and the NAD^+ molecule, six weak hydrophobic interactions involving Gly9, Ile34, Asn82, Cys84, Ile131,

Table II
Structural Refinement Statistics

Crystal	Apo	Holo
Structure determination		
Resolution (Å)	50–1.7	44–2.1
Total unique reflections	68407	36874
Test set reflections	3442	1841
R_{cryst} (%)[a]	17.88	16.72
R_{free} (%)[a]	20.79	19.99
Number of protein atoms	3770	3761
Average *B*-factor (Å^2)	30.5	28.9
Water molecules	551	293
Average *B*-factor (Å^2)	47.9	39.7
NAD^+ molecules	–	2
Average *B*-factor (Å^2)	–	54.2
Other ligand molecules[b]	7	5
Average *B*-factor (Å^2)	64.5	47.5
Root-mean-square deviations		
Bond lengths (Å)	0.011	0.010
Bond angles (°)	1.327	1.330
Ramachandran plot		
Residues in most favored regions (%)	97.3	96.8
Residues in disallowed regions (%)	0.0	0.0

[a]$R_{cryst} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$; R_{free} is calculated in the same way but using 5% of the reflections omitted from the refinement procedure.
[b]The other ligand molecules include glycerol, thiosulfate, and sodium ions.

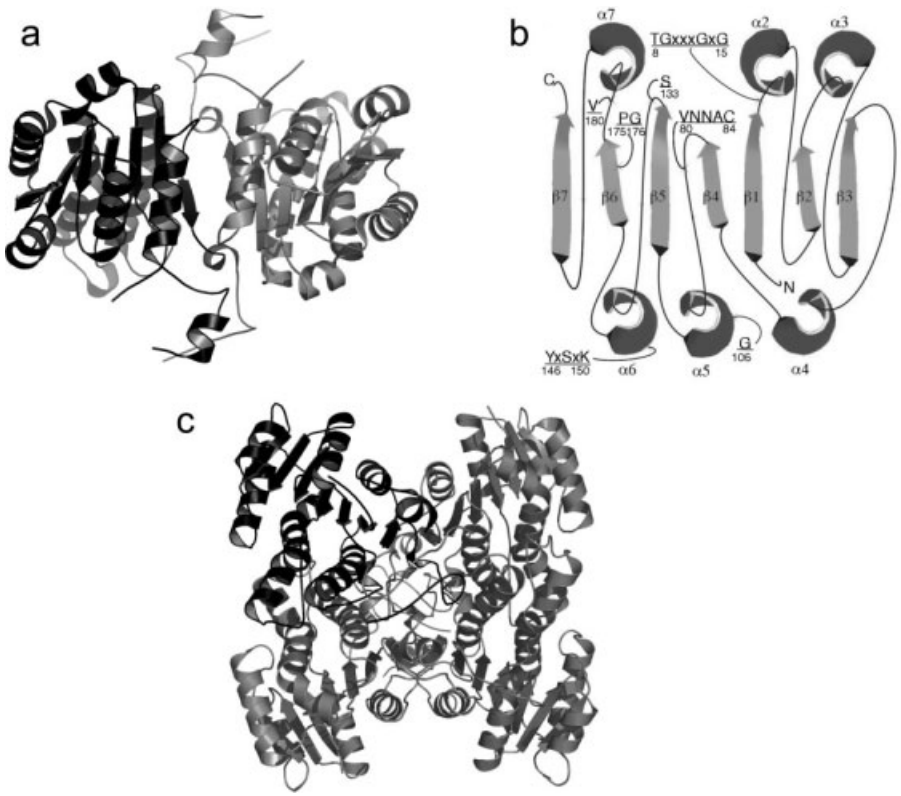

Figure 1

Structure of A3DFK9. (**a**) Cartoon of the asymmetric unit of the apo A3DFK9 crystal structure. (**b**) Classical schematic representation of a SCOR protein with secondary structural elements indicated; solid curves: α-helices, arrows: β-sheets, lines: turns. The β-sheets and α-helices are labeled β1–β7 and α2–α7, respectively, as well as the N and C terminus. Conserved amino acids and variations to the fingerprint positions (A3DFK9 numbering) discussed in the text are underlined. (**c**) Representation of biological tetramer with symmetry-related monomers.

and Ser133, and 10 hydrogen bond contacts involving His12, Ile14, Asp33, Asp55, Val56, Tyr146, Lys150, Pro175, Gly176, and Ile178 [Fig. 2(b)]. The significance of several of these contacts has been reported previously.[8,13] Duax *et al.*[8] observed that in many SCOR enzymes the backbone carbonyls of the conserved Pro-Gly motif present at the N-terminus of the β6 strand accept C=O..H hydrogen bonds from the nicotinamide ring. In A3DFK9, the Pro175 and Gly176 carbonyl oxygen atoms are 3.3 and 3.0 Å from the nicotinamide ring C5 and C4 atoms, respectively, stabilizing the nicotinamide ring for the predicted B-face proton transfer. The carboxamide group on the nicotinamide ring is stabilized by interactions with the backbone amide and carbonyl of Ile178. The catalytic Tyr146 and Lys150 form stabilizing interactions with hydroxyls on the nicotinamide ribose. Tyr146 Oη is 2.6 Å away from O2 of this ribose, whereas Lys150 Nζ participates in two contacts of 2.7 and 3.0 Å from the ribose O2 and O3, respectively. These interaction distances are common among the SCOR family.[32,33] The binding of cofactor to a SCOR protein can induce a conformation shift from an open to a closed form of the enzyme.[33] However, for A3DFK9 very little change was observed in the substrate-binding loops following cofactor binding (RMSD comparison of apo- and cofactor-bound A3DFK9: 0.23 Å over all Cα atoms and 0.17 Å over the 33 Cα atoms present in the loops). The conformations of these loops were similar to the open structures of other SCOR family members.

Two glycerol molecules are present in both A3DFK9 structures. One glycerol is present in the active site of each monomer. In all cases, the glycerol interacts with residues (Arg135, Gly176, and Trp177) on one monomer and with residues (Tyr235 and Asp238) of a second crystallographically independent monomer within the tetramer [Fig. 2(a)]. Three thiosulfate molecules were identified bound to the apo enzyme structure. One thiosulfate molecule is bound in the cofactor-binding site within each A3DFK9 monomer, interacting with residues Ile34, Asn82, Ala83, Cys84, and Arg85. The third thiosulfate ion is present between two monomers related by noncrystallographic twofold symmetry, interacting with the basic side chains of Lys16 and Arg46 contributed from both monomers. A thiosulfate molecule was observed in a similar position within the NAD^+-ligated structure, but the bound NAD^+ cofactor prevented thio-

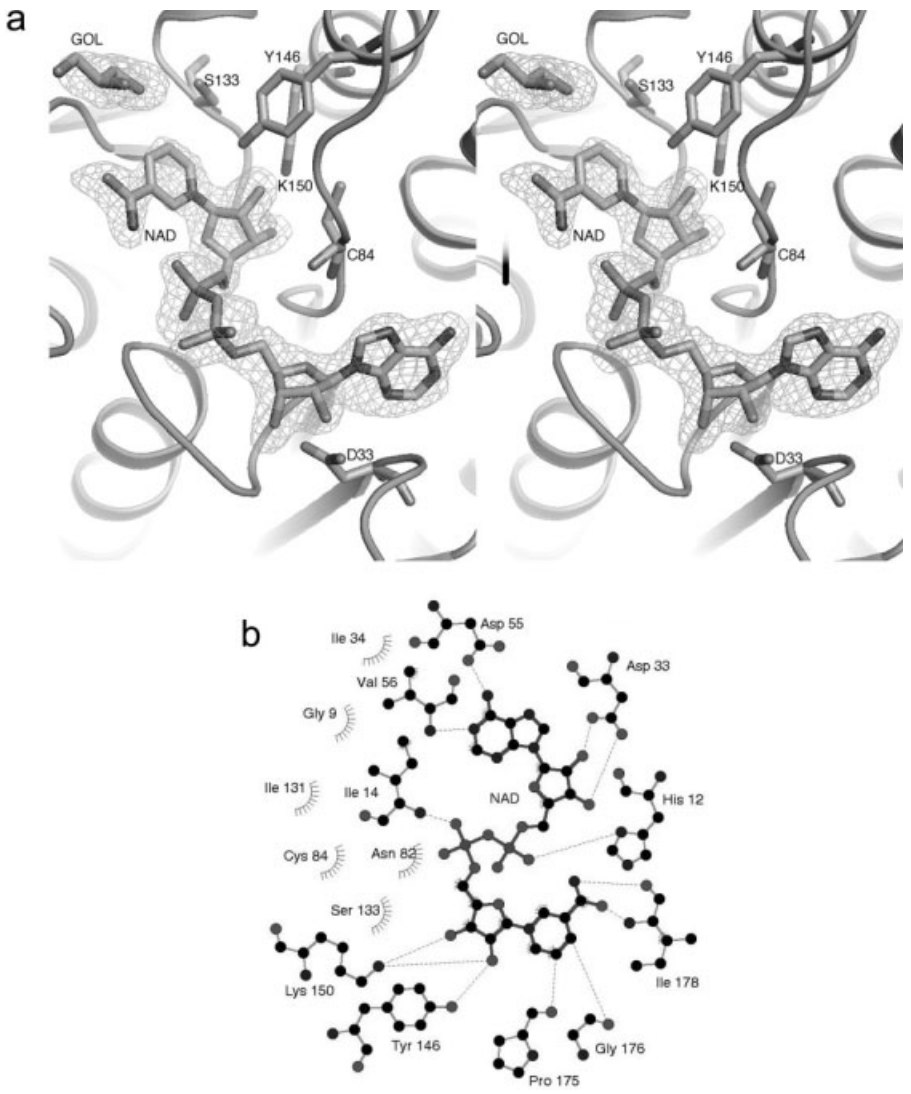

Figure 2

(**a**) Stereoview of $F_o - F_c$ density map at 3σ displaying the bound NAD^+ cofactor in Monomer A plus a glycerol (GOL) molecule bound above the nicotinamide ring in the substrate-binding pocket. (**b**) Two-dimensional representation of the residues within 3.0 Å surrounding the bound NAD^+ molecule.

sulfate from also occupying the cofactor binding site. Additionally, two Na^+ ions were identified in both the apo enzyme and the NAD^+-bound structures. Glycerol, thiosulfate, and NaCl were present in the protein sample solution or the crystallization cocktail.

Unexplained electron density

Unexpected positive electron density was observed extending from the Cys191 γ-sulfur atoms present within the crystallographic-independent A3DFK9 monomers within both the apo and NAD^+ complex structures (Fig. 3). Attempts at modeling a chemical group into this strong density failed. The expression vector was resequenced and the presence of a cysteine residue was confirmed at this position, and so mutation is unlikely the cause of the excess density. All known crystallization cocktail and protein buffer small molecules were placed into this density. Oxidation of Cys191 was considered and sulfenic, sulfinic, and sulfonic acid moieties were modeled. However, in all cases, the results were judged unsatisfactory because of poor shape complementarily to the density and/or significant positive difference density remained following refinement of the modified model. Thus, no atoms were modeled into this additional electron density. Although the modification of Cys191 has not been identified, the crystal structure demonstrated that the modified side chain does not form close contacts

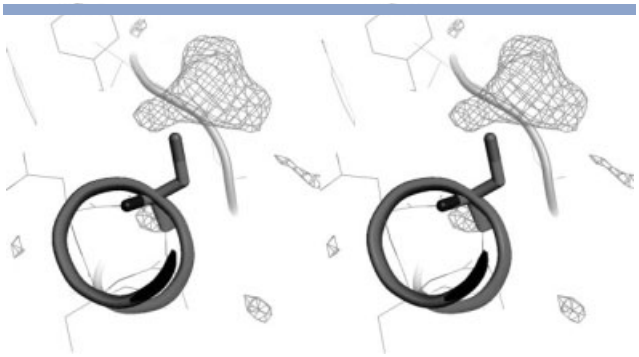

Figure 3

Unidentified modification of Cys191. (A) Stereoview of positive $F_o - F_c$ electron density contoured at 3σ adjacent to Cys191(A), indicating an unidentified modification of the side chain. Similar density was observed in Monomer B. No significant negative $F_o - F_c$ electron density was observed in this region.

(≤3.5 Å) with adjacent residues, is not involved in forming oligomer-stabilizing contacts, and is not located in a position that is likely to participate in substrate binding or the catalytic mechanism. Additionally, unexplained electron density was observed extending from the side chains of both crystallographically independent Cys84 residues in the apo enzyme structure. Although the additional density about Cys191 appeared to be due to a covalent side-chain modification, the extra density associated with Cys84 was not clearly defined and it may represent a noncovalently bound small molecule. Water was modeled into this extra density. In the NAD^+-bound structure, no extraneous electron density was observed in the immediate vicinity of the Cys84 side chains. Cys84 and several neighboring residues underwent significant conformational changes upon NAD^+ binding, and this structural change is likely the reason for the lack of this unexplained density in the presence of NAD^+.

TGYK subfamily fingerprint variations

Alignments of 426 enzymes from the TGYK subfamily identified 40 highly conserved residues (>70% identity), which were termed fingerprint residues.[8] Residues discussed in the text are identified in Figure 1(b). The 40 fingerprint residues are mainly present in loops and have been shown or implicated to be critical to SCOR protein folding, cofactor binding, catalysis, or oligomerization.[8,10–12] A3DFK9 contains the expected residue type at 31 of the 40 fingerprint positions. Six of the nine variations are present in other TGYK members for which crystal structures have been reported (Table III and Supporting Information Fig. 2), and the three previously unobserved fingerprint variations were a significant consideration in choosing A3DFK9 for structural studies. The variations present in A3DFK9 from the conserved fingerprint residues do not significantly alter main-chain geometry compared to structures of other typical TGYK enzymes possessing the fingerprint residues. However, several of the variations alter the side-chain size, hydrophobicity, or charge at the fingerprint positions. Most of the fingerprint variations that result in a significant change in residue properties occur at the protein surface (Supporting Information Fig. 3). One fingerprint residue in the TGYK subfamily is an alanine present between the catalytic tyrosine and lysine resides, to form a YxAxK motif on helix α6. However, A3DFK9 exhibits an alanine → serine (Ser148) fingerprint variation at this position. This variation is located at the dimer interface, where the presence of a small amino acid is required to maintain a tight interface. This variation has a minor effect on hydrophobicity, but serine residues are often interchangeable with alanine residues in positions that require a small side chain at interfaces involving helices.[34] This is an example of a fingerprint variation that occurs at a buried and functionally critical location but does not significantly alter the surrounding structure.

A second example of an A3DFK9 fingerprint variation involves a cofactor-binding residue. Zhou and Tai demonstrated the significance of a fingerprint threonine residue (Thr188) in the SCOR enzyme 15-hydroxyprostaglandin dehydrogenase (15-PGDH).[13] The mutants T188A and T188Y were inactive, and the T188S mutant was active but exhibited a 100-fold higher K_m for NAD^+. The crystal structure of human 15-PGDH showed that the carboxamide group of NAD^+ was hydrogen bonded to the side-chain Oγ atom of Thr188 (PDB: 2GDZ). This cofactor-stabilizing interaction is a common feature in the TGYK subfamily and in SCOR enzymes in general. In A3DFK9, valine (Val180) replaces this fingerprint threonine (Supporting Information Fig. 2), and this substitution has not been previously structurally characterized. This fingerprint variation eliminates this cofactor-stabilizing contact, as the Val180 side chain cannot participate in hydrogen bonding. The A3DFK9-NAD^+ crystal structure verified the loss of this important cofactor–enzyme interaction without replacement by a compensatory interaction involving a different residue. However, the relationship of the nicotinamide ring to the catalytic residues is similar between the A3DFK9-NAD^+ and the 15-PGDH structures.

A3DFK9's proton wire

The positions corresponding to residues Cys84 and Gly106 in A3DFK9 are typically occupied by the fingerprint residues glycine and asparagine, respectively, in the TGYK subfamily and the larger SCOR family. In A3DFK9, Cys84 is present in place of the glycine fingerprint residue (89% conserved) within the VNNAG motif in the β4α5 loop, and Gly106 is present in place of an asparagine fingerprint residue (94% conserved) that

Table III
A3DFK9 Variations from TGYK SCOR Fingerprint Amino Acids

	Variant (A3DFK9 #)	Function	SCOR structure with same variation (PDB code)[a]
1	G74 -> Q	Fold (turns)	1SEP
2	G84 -> C	Catalysis	–
3	N106 -> G	Catalysis	–
4	A148 -> S	Dimer (interface)	1FMC
5	E163 -> S	Dimer (interface)	1HXH
6	T180 -> V	Catalysis, cofactor	–
7	R198 -> K	Fold (turns)	1CYD
8	A206 -> S	Fold (unknown)	1AE1
9	S215 -> Q	Fold (turns)	1SEP

[a]Sequence alignments included as Supporting Information Figure 2.

occurs at a structurally conserved kink in helix α5. These two fingerprint residues have been shown to be critical for cofactor binding and catalysis [Fig. 1(b)].[10,12,13,33] SCOR enzyme catalysis involves the transfer of a proton between substrate and bulk solvent through a buried proton transfer wire, and these two fingerprint residues also participate in forming this proton wire.[10,33] Thus, the structural effects of the nonconservative fingerprint variations Gly → Cys84 and Asn → Gly106 in A3DFK9 were of extreme interest and were first observed in the A3DFK9 structures.

The asparagine fingerprint residue typically present at the position corresponding to A3DFK9 residue Gly106 participates in several important structure–function roles. Helix α5 forms a major dimer interface. This helix exhibits a characteristic kink at the fingerprint asparagine, with the canonical α-helical hydrogen-bonding pattern interrupted at the asparagine's main-chain carbonyl. This carbonyl oxygen cannot participate in a helix-stabilizing hydrogen bond with the $n + 4$ residue's main-chain amide nitrogen because the carbonyl bond is directed roughly normal to the helical axis rather than parallel. The atypical conformation of the asparagine carbonyl makes it available for hydrogen bonding to a water molecule that in turn forms a hydrogen bond to the catalytic lysine's side-chain amide group. This interaction has been postulated to both stabilize the position of the catalytic lysine side chain and to position a water molecule to act as a protein wire component. The asparagine side chain acts as a lever to introduce the helical kink. The conformation of this kink-forming asparagine is locked in place by the presence of hydrogen bonds between its side-chain Oδ1 and Nδ2 atoms and the main-chain amide nitrogen and carbonyl oxygen, respectively, of the residue immediately following the conserved fingerprint glycine residue present at the C-terminus of the VNNAG motif in the β4α5 loop. This asparagine was mutated to a leucine (N111L) in a bacterial 3β/17β-hydroxysteriod dehydrogenase and shown to completely inactivate the enzyme.[10] A serine or threonine residue immediately precedes the catalytic lysine in many SCOR enzymes. The side chain of this serine/threonine has been observed to hydrogen bond to the fingerprint asparagine side chain, further stabilizing the asparagine's lever-like conformation. Mutation of this serine to a glycine dramatically decreases the catalytic rate, presumably by destabilizing the asparagine's conformation and the helical kink.[10,35]

The fingerprint glycine within the VNNAG motif serves many critical functional roles. It has been postulated that a flexible main chain is required at this position in order for the main chain to temporarily accept a proton via a enolic intermediate as part of the proton wire.[10] Main-chain flexibility at this position may also be required for protein conformation changes associated with cofactor binding.[33] Additionally, the presence of a side chain at this position would prevent the side chain of the fingerprint asparagine present at the helix α5 kink by forming dual hydrogen bonds to the main chain of the residue immediately downstream of this fingerprint glycine (residue $n + 1$). A large side chain at this position may also intrude into the active site.

The A3DFK9 crystal structures are the first structural characterization of the TGYK subfamily fingerprint variations Gly → Cys84 and Asn → Gly106. Perhaps most significantly, helix α5 retains the characteristic and functionally important kink, centered at Gly106. This observation was unexpected, as glycine is unable to form side-chain interactions analogous to those that induce the kink at the fingerprint asparagine residue. A3DFK9's α5 helical kink is introduced via an arrangement of residues unique within the SCOR family. In a canonical α-helix, Gly106's carbonyl group would hydrogen bond to the

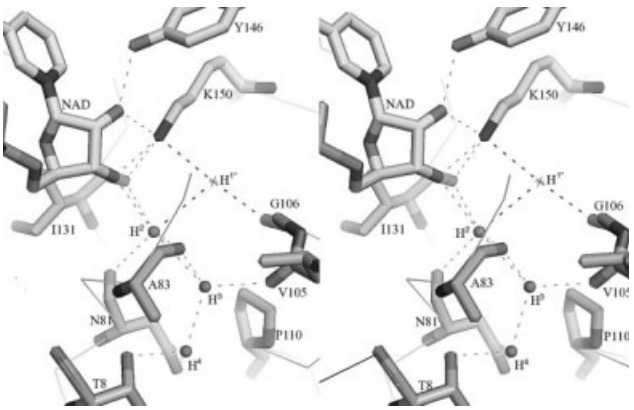

Figure 4
Stereoview of A3DFK9-NAD^+ proton wire generated with conserved waters H^{1*}-H^4. Surrounding the conserved waters are fingerprint residues. H^{1*} is a water present in the apo structure and likely at least partially occupied in the A3DFK9-NAD^+ complex structure. The presence of the Pro110 disrupts the hydrogen-bonding pattern in helix α5 leaving the carbonyl of Gly106 available for hydrogen bonding to water.

amide nitrogen of the $n + 4$ residue. In A3DFK9, this residue is a proline (Pro110) and is incapable of participating in the α-helical hydrogen-bonding scheme (Fig. 4). Furthermore, the steric clash with the Pro110 side chain forces Gly106's carbonyl group to be directed toward the internal solvent tunnel, placing it in a conformation similar to that typically observed for the main-chain carbonyl of the fingerprint asparagine residue. Thus, A3DFK9 exhibits an alternate means of forming a kink in helix α5 to place a main-chain carbonyl group in the correct position for participating in the catalytically important proton wire.

Helix α5, including the kink at Gly106, possesses extremely similar structures in both the apo- and the cofactor-bound A3DFK9 enzymes, as do the catalytic residues Ser133, Tyr146, and Lys150. The apo enzyme contains water positioned between the Gly106 carbonyl and Lys150 side-chain amide groups. A similarly positioned water has been postulated to belong to the proton wire in other SCOR proteins.[10,33] This bridging water was not considered to be observed in the A3DFK9-NAD^+ complex structure. $F_o - F_c$ difference density peaks (2.4σ and 2.9σ maximum height in Monomer A and B, respectively) were present at the expected positions of the bridging water, but both peaks were below the 3.0σ cutoff for the inclusion of waters. The lower resolution data obtained for the cofactor bound structure may account for the failure to observe this water.

A3DFK9 residues Asn82-Ala83-Cys84-Arg85 undergo a significant structural change upon binding of NAD^+. Asn82, Ala83, and Cys84 belong to A3DFK9's VNNAC motif, which is a variation of the TYGK subfamily's conserved VNNAG motif. The Cys84 side chain extends into the cofactor binding site in the apo enzyme structure. This residue rotates ∼130° upon NAD^+ binding, resulting in the side chain being removed from the cofactor-binding site and redirected toward the catalytic Tyr146 and Lys150. A cysteine side chain is of modest size, and so it does not intrude upon the active site. This rearrangement is necessary to avoid a steric clash with the cofactor. Residues Asn82 to Arg85 exhibit mobility in both the apo- and the cofactor-bound structures, based upon the quality of the electron density and the refined temperature factors. These residues in the NAD^+-bound structure closely resemble the active conformation reported in other SCOR structures (e.g., *E. coli* FabG, PDB: 1Q7B; human retinal dehydrogenase/reductase 3, PDB: 1YDE). The Asn → Gly106 fingerprint variation removes restraints upon maintaining the fingerprint glycine within the usually conserved VNNAG motif and allowing it to be replaced by a modestly sized residue: (a) the dual hydrogen-bond interactions between the side chain of the fingerprint asparagine usually present at the helix α5 kink and the main chain of the residue immediately following the VNNAG glycine are not present in A3DFK9; (b) the volume that would be occupied by the fingerprint asparagine side chain is left unoccupied because of the A3DFK9's Gly106 fingerprint variation. Modeling the fingerprint asparagine residue into the A3DFK9-NAD^+ complex structure in place of Gly106 demonstrated that the presence of the Cys84 side chain would interfere with the formation of the hydrogen bonds between the fingerprint asparagine's side chain and the main chain of the residue immediately downstream of Cys84 (i.e., Arg85 in A3DFK9) that stabilize the kink in helix α5 in the majority of SCOR enzymes. It is unknown if the presence of the thiol group proximal to the active site effects the reaction mechanism, as A3DFK9's substrate is unknown. The codon corresponding to Cys84 is UGC. This raises the potential that the Gly → Cys84 fingerprint variation arose following a single base change from the glycine codon GGC. There may be no functional significance to the presence of Cys84 other than its modest side chain volume does not result in steric clashes with functionally critical residues.

Proton wire water structure

A3DFK9 possesses a proton wire composed of a combination of protein atoms, ribose hydroxyls, and waters that is similar to those described for other SCOR enzymes (Fig. 3).[10,33] The unique feature of A3DFK9's proton wire is the manner in which the kink in helix α5 is formed (discussed earlier), rather than the wire itself (Fig. 3). The Tyr146 hydroxyl is hydrogen bonded to the NAD^+ nicotinamide ribose hydroxyl O2D, which in turn is hydrogen bonded to the side-chain amide of Lys150. This moiety is hydrogen bonded to a water (observed in the apo structure and likely present in the cofactor-bound structure) held in place through an interaction with the carbonyl group of Gly106. This water is then connected to the bulk solvent through a linear chain of three more waters connected by hydrogen bonds. The second water interacts with the main-chain carbonyls of Ala83 and Ile131. The third water is stabilized by the main-chain carbonyls of Ala83 and Val105. In other SCOR enzyme structures, an interaction analogous to that with Val105 is absent. Rather, this third water of the proton wire typically interacts with the hydroxyl group of a serine, one turn downstream ($n + 4$) from the fingerprint asparagine residue at the helical kink. However, in A3DFK9 this residue is a proline (Pro110), and so such an interaction is impossible. The fourth water is hydrogen bonded to the side chains of Thr8 and Asn81. These water–protein interactions are analogous to the stabilizing interactions previously observed for analogous SCOR enzyme proton wires.[10,33] Surprisingly, the four waters of the proton wire remain in place in the apo A3DFK9 structure, although several water–protein hydrogen bonds differ. This situation is unlike *E. coli* FabG in which the proton wire is only completely present in the cofactor-bound structure.[33] This observation was unex-

pected, because of the rearrangement of A3DFK9 residues Asn82 to Arg85 upon cofactor binding that is similar to the conformational change described for FabG. A3DFK9 is further evidence that the proton wire is an important conserved feature of SCOR enzymes. Furthermore, the number of TGYK subfamily fingerprint residues (Thr8, Asn81, Ala83, and the Gly → Cys84 and Asn → Gly106 variations in A3DFK9) participating in the formation of the proton wire suggests that the functional definition of these fingerprint residues may require expansion to include this role.

Evolutionary significance of fingerprint variations

No previously reported SCOR crystal structure contains a fingerprint variation analogous to the combined Gly → Cys84/Asn → Gly106 variation present in A3DFK9. To understand the evolutionary significance of these fingerprint variations, we performed a PROSITE pattern search of the key positions in Figure 1(b).[30] A variation analogous to Gly → Cys84 occurs very infrequently across the TGYK subfamily, but this variation is present in all known TGYK proteins within the *Clostridium* genus. The next most frequent substitution at the fingerprint Gly position is alanine (7.5%), compared with <1% for the cysteine substitution. The Gly → Ala substitution has been observed in porcine carbonyl reductase at Ala91, but this enzyme retains the fingerprint asparagine (Asn113) residue at the helix α5 kink (PDB: 1N5D).[36] The small side chain of Ala91 did not interfere with the formation of the kink stabilizing hydrogen bonds between the side chain of Asn113 and the main chain of the residue immediately downstream of the Ala91 fingerprint variation (Ile92). As such, this structural feature of the TGYK subfamily was present, and so this Gly → Ala fingerprint variation can be considered a conservative substitution. A residue larger than alanine would likely disrupt this important hydrogen-bonding interaction and/or impinge upon the active site.

The most frequent fingerprint variation to the asparagine residue located at the helix α5 kink is histidine, which is present in 2.8% of the TGYK subfamily, compared with <1% for the glycine present in A3DFK9. The Asn → His fingerprint variation has been observed in (3R)-hydroxyacyl-CoA dehydrogenase crystal structure (PDB: 1GZ6).[37] This protein's His123 residue mimics interactions formed by the fingerprint asparagine residue normally present at this position. The His123 Nϵ2 atom hydrogen bonds to the main-chain carbonyl of Ile102, which immediately follows the VNNAG motif. The His123 Nδ1 atom hydrogen bonds to the catalytic Lys168 side chain without the need for the bridging water that usually mediates the interaction between the fingerprint asparagine and lysine side chains.

Substrate-binding residues

The substrate-binding pockets for both the apo- and the NAD^+-bound A3DFK9 structures contain a bound glycerol molecule. Glycerol was present in the crystallization medium. In both the apo- and NAD^+-bound A3DFK9 structures, glycerol interacts with residues lining the substrate-binding pocket and is sandwiched between Arg135 and Trp177. Arg 135 is located in the β5α6 loop and lies between the catalytic Ser133 and Tyr146. Trp177 is located in loop β6α7, directly adjacent to the PG motif which serves to position the cofactor's nicotinamide ring in the active site. Within each structure, the two crystallographically independent substrate-binding sites exhibit identical glycerol–protein interactions. However, binding differences were observed between the A3DFK9 apo and NAD^+ complex structures. In apo A3DFK9, glycerol forms a hydrophobic contact with Trp177, and a hydrogen bond between its O3 hydroxyl to Arg135, with its O2 hydroxyl group facing away from this arginine. In the A3DFK9-NAD^+ structure, the glycerol is stabilized through a hydrophobic contact with Trp177 and a hydrogen bond between its O2 hydroxyl and Arg135.

The residues interacting with the glycerols belong to loops typically associated with substrate binding within the SCOR family. For example, the residues in these positions have been previously seen as key substrate-interacting residues from 17β-hydroxy-steroid dehydrogenase and 2,3-dihydro2,3-dihydroxybenzoic acid.[38,39] Thus, the bound glycerol molecule likely mimics at least some aspects of endogenous substrate binding.

Substrate prediction

SCOR proteins contain several fingerprint motifs that can be used to classified proteins into subfamilies or are predictive of cofactor usage.[8] Sequence analysis, however, has proven less useful for substrate prediction. The specific substrate of A3DFK9 is unknown. Simple sequence similarity searches (e.g., BLAST) failed to identify a highly homologous protein having a well-established substrate preference. The Cth_1510 gene encoding A3DFK9 is predicted to be present in an operon containing only one other gene, Cth_1509, which is annotated as encoding a hypothetical protein. Thus, the genetic environment failed to provide insight into the potential function or substrate of A3DFK9. The high conservation of the TYGK subfamily fingerprint residues overwhelms the subtle sequence differences that impart substrate specificity within this large enzyme subfamily. Another sequence analysis method, the PRINTS server, uses a motif-scanning database that clusters proteins by small conserved sequence segments.[40] This server classified A3DFK9 as exhibiting a similarity to the glucose/ribitol dehydrogenase family signature.

Three-dimensional structure comparison was attempted as a method to predict the A3DFK9 substrate. SCOR substrate-binding loops are known to be mobile and are often disordered in crystal structures in the absence of a bound substrate or substrate analog. The bound glycerol in A3DFK9 is likely to stabilize the substrate-binding loops in the "substrate bound" conformation, relaying information about the conformation of the loops upon substrate binding. The higher resolution apo A3DFK9 structure was compared against the entire PDB (November 2008 release). This search identified many structurally homologous proteins, but comparisons of the substrate-binding pockets of these search hits to A3DFK9 revealed conformations and residue identities differed significantly. Thus, it was concluded that the selected proteins were not likely substrate orthologs (data not shown). This failure is due to the overall structural similarity of the conserved SCOR family fold overwhelming the more localized diversity of the substrate-binding pockets. For instance, superposition of A3DFK9 with the top hit (PDB: 2D1Y chain C) had a RMSD of 1.28 Å over 218 Cα atoms, whereas comparing just the substrate loops (4 residues in the β4α5 loop, 18 residues in the β5α6 loop, and 10 residues in the β6α7 loop) had a RMSD of 3.4 Å over the Cα atoms.

No improvement in substrate prediction was observed when the 3D search was limited to the three substrate-binding loops, with the Rossmann fold excluded. No SCOR proteins were among the top hits. Furthermore, visual inspection of the top hits revealed that the protein fragments best matching the main-chain structure of A3DFK9's substrate-binding loops did not belong to ligand-binding pockets, nor did they possess similar sequences. None of the SCOR protein structures deposited in the PDB had substrate-binding loop sequences corresponding to those in A3DFK9. The β6α7 loop exhibited the greatest structural and sequence differences from other SCOR proteins. Glycerol in the A3DFK9 substrate-binding site is sandwiched between Arg135 and Trp177 side chains. Visual examination and manual searching identified a similar arginine–tryptophan substrate-binding sandwich in the structure of the carbohydrate-binding domain of galectin-9 complexed with lactose (PDB: 2EAK) (Supporting Information Fig. 4).

This article described the structure of the SCOR protein from A3DFK9 *C. thermocellum*. The protein contains natural variations from several highly conserved fingerprint residues that have not been previously structurally characterized. These fingerprint variations include several residues that participate in the formation of a catalytically critical proton wire. The crystal structures reported here reveal how this proton wire is retained despite nonconservative sequence changes. Although the endogenous substrate(s) of A3DFK9 is unknown, the presence of a bound glycerol in the substrate-binding pocket in combination with bioinformatic analysis suggests that the substrate may be a polyalcohol or carbohydrate. Structural and sequence analysis of A3DFK9's substrate-binding loops demonstrated that the substrate-binding pocket is unique among SCOR structures previously deposited in the PDB. Further biochemical studies and verification are currently being carried out to further test the substrate hypotheses generated from this structure and bioinformatic analysis. The diversity present within this ancient protein family emphasizes the need for further research to understand substrate specificity.

REFERENCES

1. Reading PC, Moore JB, Smith GL. Steroid hormone synthesis by vaccinia virus suppresses the inflammatory response to infection. J Exp Med 2003;197:1269–1278.
2. Tanaka N, Nonaka T, Nakamura KT, Hara A. SDR structure, mechanism of action, and substrate recognition. Curr Org Chem 2001;5: 89–111.
3. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 2003;31:365–370.
4. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. Pfam: clans, web tools and services. Nucleic Acids Res 2006;34 (Database issue):D247–D251.
5. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), Enzyme Supplement 5 (1999). Eur J Biochem 1999;264:610–650.
6. Jornvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D. Short-chain dehydrogenases/reductases (SDR). Biochemistry 1995;34:6003–6013.
7. Duax WL, Huether R, Pletnev VZ, Langs D, Addlagatta A, Connare S, Habegger L, Gill J. Rational genomics. I. Antisense open reading frames and codon bias in short-chain oxido reductase enzymes and the evolution of the genetic code. Proteins 2005;61:900–906.
8. Duax WL, Pletnev V, Addlagatta A, Bruenn J, Weeks CM. Rational proteomics. I. Fingerprint identification and cofactor specificity in the short-chain oxidoreductase (SCOR) enzyme family. Proteins 2003;53:931–943.
9. Duax WL, Thomas J, Pletnev V, Addlagatta A, Huether R, Habegger L, Weeks CM. Determining structure and function of steroid dehydrogenase enzymes by sequence analysis, homology modeling, and rational mutational analysis. Ann N Y Acad Sci 2005;1061:135–148.
10. Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jornvall H, Oppermann U. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. J Biol Chem 2002;277:25677–25684.
11. Kallberg Y, Oppermann U, Jornvall H, Persson B. Short-chain dehydrogenases/reductases (SDRs). Eur J Biochem 2002;269:4409–4417.
12. Oppermann U, Filling C, Hult M, Shafqat N, Wu X, Lindh M, Shafqat J, Nordling E, Kallberg Y, Persson B, Jornvall H. Short-chain dehydrogenases/reductases (SDR): the 2002 update. Chem Biol Interact 2003;143/144:247–253.
13. Zhou H, Tai HH. Threonine 188 is critical for interaction with NAD^+ in human NAD^+-dependent 15-hydroxyprostaglandin dehydrogenase. Biochem Biophys Res Commun 1999;257:414–417.
14. Demain AL, Newcomb M, Wu JH. Cellulase, clostridia, and ethanol. Microbiol Mol Biol Rev 2005;69:124–154.
15. Doi RH, Kosugi A. Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol 2004;2:541–551.
16. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976;72:248–254.

17. Luft JR, Collins RJ, Fehrman NA, Lauricella AM, Veatch CK, DeTitta GT. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. J Struct Biol 2003; 142:170–179.
18. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, Garman E, Gonzalez A, Sauter NK, Phizackerley RP, Soltis SM, Kuhn P. Blu-Ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. J Synchrotron Radiat 2002;9 (Part 6):401–406.
19. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50 (Part 5):760–763.
20. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter CW, Jr, Sweet RM, editors. Methods in enzymology, Vol. 276. New York: Academic Press; 1997. pp 307–326.
21. Navaza J. Implementation of molecular replacement in AMoRe. Acta Crystallogr D Biol Crystallogr 2001;57 (Part 10):1367–1372.
22. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54 (Part 5):905–921.
23. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 2002; 58 (Part 11):1948–1954.
24. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60 (Part 12 Part 1):2126–2132.
25. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, III, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35 (Web Server issue):W375–W383.
26. DeLano WL. The PyMOL Molecular Graphics System. Palo Alto, CA: DeLano Scientific; 2002. Available at:http://www.pymol.org.
27. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 1995;8:127–134.
28. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 2004;60 (Part 12 Patt 1):2256–2268.
29. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v. 3. Bioinformatics 2008;24:2780–2781.
30. Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJ. The 20 years of PROSITE. Nucleic Acids Res 2008;36 (Database issue):D245–D249.
31. Edgar RC. MUSCLE. multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–1797.
32. Pletnev VZ, Weeks CM, Duax WL. Rational proteomics. II. Electrostatic nature of cofactor preference in the short-chain oxidoreductase (SCOR) enzyme family. Proteins 2004;57:294–301.
33. Price AC, Zhang YM, Rock CO, White SW. Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG. Structure 2004;12: 417–428.
34. Doherty AJ, Serpell LC, Ponting CP. The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. Nucleic Acids Res 1996;24:2488–2497.
35. Obeyesekere VR, Trzeciak WH, Li KX, Krozowski ZS. Serines at the active site of 11 beta-hydroxysteroid dehydrogenase type I determine the rate of catalysis. Biochem Biophys Res Commun 1998; 250:469–473.
36. Ghosh D, Sawicki M, Pletnev V, Erman M, Ohno S, Nakajin S, Duax WL. Porcine carbonyl reductase. Structural basis for a functional monomer in short chain dehydrogenases/reductases. J Biol Chem 2001;276:18457–18463.
37. Haapalainen AM, Koski MK, Qin YM, Hiltunen JK, Glumoff T. Binary structure of the two-domain (3R)-hydroxyacyl-CoA dehydrogenase from rat peroxisomal multifunctional enzyme type 2 at 2.38 Å resolution. Structure 2003;11:87–97.
38. Pletnev VZ, Duax WL. Rational proteomics. IV. Modeling the primary function of the mammalian 17beta-hydroxysteroid dehydrogenase type 8. J Steroid Biochem Mol Biol 2005;94:327–335.
39. Sundlov JA, Garringer JA, Carney JM, Reger AS, Drake EJ, Duax WL, Gulick AM. Determination of the crystal structure of EntA, a 2,3-dihydro-2,3-dihydroxybenzoic acid dehydrogenase from *Escherichia coli*. Acta Crystallogr D Biol Crystallogr 2006;62 (Part 7):734–740.
40. Attwood TK, Bradley P, Flower DR, Gaulton A, Maudling N, Mitchell AL, Moulton G, Nordle A, Paine K, Taylor P, Uddin A, Zygouri C. PRINTS and its automatic supplement, prePRINTS. Nucleic Acids Res 2003;31:400–402.