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

Preliminary crystallographic studies of glucose dehydrogenase from the promiscuous Entner–Doudoroff pathway in the hyperthermophilic archaeon *Sulfolobus solfataricus*

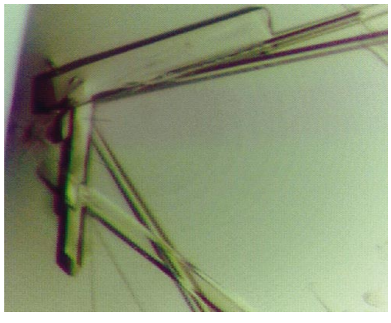
Alex Theodossis,[a]‡ Christine C. Milburn,[a]‡ Narinder I. Heyer,[b] Henry J. Lamble,[b] David W. Hough,[b] Michael J. Danson[b] and Garry L. Taylor[a]*

[a]Centre for Biomolecular Sciences, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, Scotland, and [b]Centre for Extremophile Research, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, England

‡ These authors contributed equally to this work.

Correspondence e-mail: glt2@st-andrews.ac.uk

Received 20 October 2004
Accepted 25 November 2004
Online 24 December 2004



The hyperthermophilic archaeon *Sulfolobus solfataricus* grows optimally above 353 K and can metabolize glucose and its C4 epimer galactose *via* a non-phosphorylative variant of the Entner–Doudoroff pathway involving catalytically promiscuous enzymes that can operate with both sugars. The initial oxidation step is catalysed by glucose dehydrogenase (SsGDH), which can utilize both NAD and NADP as cofactors. The enzyme operates with glucose and galactose at similar catalytic efficiency, while its substrate profile also includes a range of other five- and six-carbon sugars. Crystals of the 164 kDa SsGDH homotetramer have been grown under a variety of conditions. The best crystals to date diffract to 1.8 Å on a synchrotron source, have orthorhombic symmetry and belong to space group $P2_12_12$. Attempts are being made to solve the structure by MAD and MR.

1. Introduction

The hyperthermophilic archaeon *Sulfolobus solfataricus* grows optimally at 353–358 K and pH 2–4 utilizing a wide range of carbon and energy sources (Grogan, 1989). Glucose metabolism in this organism proceeds *via* a non-phosphorylative variant of the classical Entner–Doudoroff pathway, in which one molecule of glucose produces two molecules of pyruvate with no net yield of ATP (De Rosa *et al.*, 1984). Analogous pathways have been detected in the thermoacidophilic archaea *Sulfolobus acidocaldarius* (Selig *et al.*, 1997), *Thermoplasma acidophilum* (Budgen & Danson, 1986) and *Thermoproteus tenax* (Siebers & Hensel, 1993), as well as strains of *Aspergillus* fungi (Elzainy *et al.*, 1973; Elshafei, 1989).

In *S. solfataricus*, the enzymes of this non-phosphorylative Entner–Doudoroff pathway have been shown to possess substrate promiscuity (Lamble *et al.*, 2003) and are therefore able to metabolize both glucose and its C4 epimer galactose. The first reaction step involves the NAD(P)-dependent oxidation of glucose to gluconate, or galactose to galactonate, catalysed by glucose dehydrogenase (GDH; EC 1.1.1.47). Gluconate and galactonate are subsequently dehydrated by gluconate dehydratase (GD) to give 2-keto-3-deoxygluconate (KDG) and 2-keto-3-deoxygalactonate (KDGal), respectively (Lamble *et al.*, 2004). KDG aldolase (KDGA) then catalyses the aldolate cleavage of KDG or KDGal, both giving rise to glyceraldehyde and pyruvate (Lamble *et al.*, 2003). The glyceraldehyde molecule is converted to a second molecule of pyruvate by a further four catalytic steps.

The substrate promiscuity of these enzymes may have physiological significance and when considered alongside other evidence (Grogan, 1991; Cusdin *et al.*, 1996; Albers *et al.*, 1999) suggests that the organism at no point distinguishes between glucose and galactose during uptake and catabolism. *S. solfataricus* is considered to be an opportunistic heterotroph (Schafer, 1996) and the pathway promiscuity observed in its sugar metabolism could be an adaptation to the hostile thermoacidophilic environment of the organism, allowing it to scavenge efficiently for energy substrates. This metabolic promiscuity may also reflect an early evolutionary state and is likely to extend to thermoacidophilic archaea closely related to *S. solfataricus*, such as *Thermoplasma acidophilum*. This is supported by studies of *T. acidophilum* GDH (TaGDH), which also has a high activity with glucose and galactose (Smith *et al.*, 1989).

Kinetic studies of recombinant *S. solfataricus* GDH (SsGDH) have shown it to have indistinguishable properties from enzyme purified from the organism. Dual cofactor specificity for NAD and NADP has been demonstrated and catalytic activity has been measured with a number of five- and six-carbon sugar substrates, in addition to D-glucose and D-galactose for which the enzyme displays a high catalytic efficiency. The substrate profile indicates specificity for the glucose-specific stereo-configuration only at the C2 and C3 carbon positions, while there is no apparent preference for the configuration at the other positions (Lamble *et al.*, 2003). The involvement of a metal ion (most probably Zn^{2+}) in enzyme catalysis is supported by a 60% reduction in activity observed in the presence of EDTA (Lamble *et al.*, 2003).

The SsGDH monomer is composed of 366 amino-acid residues with a molecular weight of 41 kDa and forms tetrameric assemblies in solution (Lamble *et al.*, 2003). The amino-acid sequence has shown that the enzyme is a putative member of the medium-chain alcohol/polyol dehydrogenase/reductase branch of the superfamily of pyridine-nucleotide-dependent alcohol/polyol/sugar dehydrogenases (Edwards *et al.*, 1996). This family of enzymes includes TaGDH, horse liver alcohol dehydrogenase (LADH), yeast alcohol dehydrogenase (YADH) and sheep liver sorbitol dehydrogenase (LSDH). These enzymes are characterized by a chain length of 350–375 residues and conserved zinc-binding and nucleotide-binding sites. Preference for either NAD or NADP is determined by variations within a G*X*G*XX*G/A fingerprint motif and associated charged residues at structurally conserved positions (Wierenga *et al.*, 1985). Sequence alignments have shown that SsGDH possesses a G*X*G*XX*G motif (residues 188–193), as well as residues Asn211 and Arg213 at structurally conserved positions involved in nucleotide binding in other members of the family. Kinetic characterization revealed dual cofactor specificity and Arg213 is expected to help stabilize the adenosine-2-phosphate of NADP (Lamble *et al.*, 2003). SsGDH also possesses conserved cysteines and histidines equivalent to the residues involved in binding Zn atoms with putative structural and catalytic roles in the family (Lamble *et al.*, 2003).

Knowledge of the SsGDH structure would allow rationalization of its dual cofactor specificity and unusual substrate profile. The characterization of this enzyme from *S. solfataricus* forms an integral part of the wider study of the organism's variant Entner–Doudoroff pathway, which is currently under way by the authors. The lack of facial selectivity and substrate specificity of SsKDGA is already the subject of detailed structural investigations (Theodossis *et al.*, 2004). The aim of this research is to achieve a comprehensive understanding of substrate promiscuity in these enzymes, with the possibility of manipulating them to effect specific biotransformations. Such potential is greatly improved by the catalysts' thermostability and specificity for non-phosphorylated substrates.

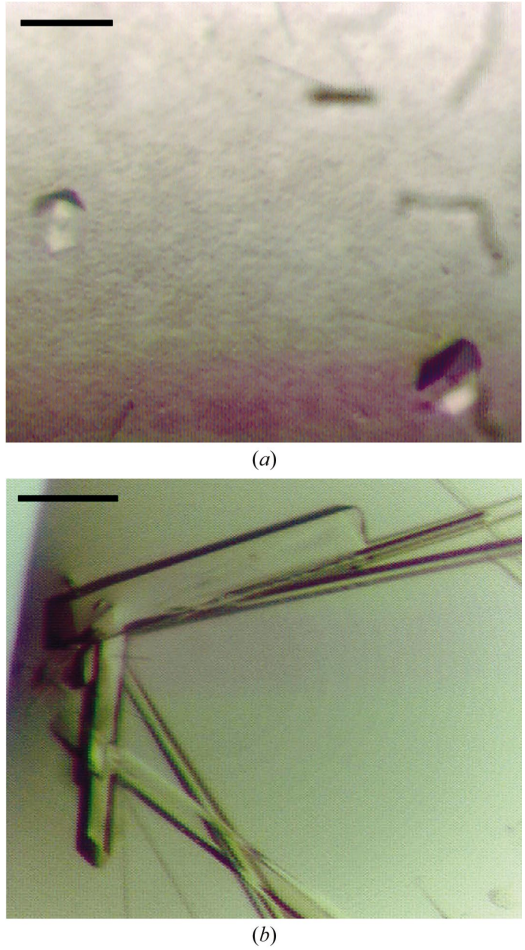


Figure 1
(*a*) Orthorhombic crystals of SsGDH belonging to space group $P2_12_12$, grown from 0.1 *M* HEPES pH 6, 13%(*w*/*v*) PEG 4000, 8%(*v*/*v*) propan-2-ol, 20 m*M* $MgCl_2$ with 3.5 mg ml^{-1} protein. (*b*) Optimized crystals diffracting to 1.8 Å; the largest crystal is approximately 0.3 × 0.1 × 0.1 mm in size. The black bars correspond to 0.1 mm.

2. Expression and purification

The GDH gene from *S. solfataricus* was cloned into the expression vector pREC7/*Nde*I, which was provided by Dr L. C. Kurz (Washington University School of Medicine, St Louis, USA), and transformed into *Escherichia coli* JM109. Cells were grown in the presence of ampicillin for 21 h at 310 K without induction and then lysed with BugBuster (Novagen). Cell extracts were heat-treated at 353 K for 30 min and precipitated proteins were removed by centrifugation. Initially, SsGDH was purified from the supernatant by gel filtration using a Superdex 200 16-60 column (Pharmacia) equilibrated in 50 m*M* Tris HCl pH 7.5 containing 20 m*M* $MgCl_2$ and 0.2 *M* NaCl (prep-A). SsGDH activity was demonstrated spectrophotometrically by the increase in absorbance at 340 nm corresponding to enzyme-dependent NADP reduction, as described in Lamble *et al.* (2003). A revised protocol (prep-B) was used to obtain the best diffracting crystals, in which the gel-filtration buffer was altered to 50 m*M* Tris–HCl pH 7.5, 20 m*M* $MgCl_2$, 5 m*M* NaCl to enable further purification. The protein was then dialysed into 50 m*M* MES pH 5.5 containing 20 m*M* $MgCl_2$ and was subsequently applied to a 5 ml bed volume of Reactive Red-500 dye affinity media (Sigma-Aldrich). SsGDH was eluted using the MES buffer with a 0–1.5 *M* NaCl elution gradient, with SsGDH eluting between 0.5 and 0.7 *M* NaCl and remaining contaminants eluting at lower salt concentrations. Finally, the pure protein was dialysed into the crystallization buffer of 50 m*M* Tris–HCl pH 7.5, 20 m*M* $MgCl_2$.

3. Crystallization and data collection

Early crystallization trials using the hanging-drop vapour-diffusion method yielded crystals diffracting to 7 Å in-house when mounted in thin-walled glass capillary tubes. This primary screen identified PEG and propan-2-ol as successful precipitants and a subsequent sparse-matrix screen of these conditions using the sitting-drop method gave rise to a number of hits. Following optimization of the conditions, improved crystals grew in 0.1 *M* HEPES pH 6.0, with 13%(*w*/*v*) PEG

4000, 8%(*v*/*v*) propan-2-ol and 20 m*M* $MgCl_2$ using protein obtained by prep-A at a concentration of 3.5 mg ml^{-1} (Fig. 1*a*). All sitting drops were formed by mixing 2 µl protein solution with 2 µl mother liquor. Crystals reached maturity within two weeks and were positively identified as being composed of SsGDH by mass spectrometry. Following cryoprotection by stepwise equilibration against mother liquor containing 10%(*v*/*v*) followed by 20%(*v*/*v*) glycerol, crystals were tested in-house at 100 K and diffracted to 3.3 Å. The crystals had orthorhombic symmetry, with unit-cell parameters a = 68.6, b = 88.4, c = 138.8 Å. Data were collected, integrated in *MOSFLM* (Leslie, 1992) and then scaled/merged in *SCALA* (Evans, 1997; Table 1). The observed systematic absences and calculated merging statistics were consistent with space group $P2_12_12$.

Further optimization of these crystals was carried out using protein from prep-B, with the best crystals growing to approximately 0.3 × 0.1 × 0.1 mm in size (Fig. 1*b*) and being formed from a mother liquor containing 12%(*w*/*v*) PEG 4000, 0.1 *M* MES pH 5.75 and 3%(*v*/*v*) propan-2-ol. Although lower concentrations of propan-2-ol led to larger crystals, these were later found to be highly anisotropic. These crystals were cryoprotected using the method described for the previous crystals. Data were collected to 1.8 Å at the ESRF (ID14-3) on a MAR CCD detector and processed in the same way to give the statistics shown in Table 1.

4. Crystal analysis

The unit-cell volume suggests the presence of a GDH dimer in the asymmetric unit with an estimated solvent content of 53% and a Matthews coefficient of 2.6 Å^3 Da^{-1} (Matthews, 1968). Calibrated gel filtration of SsGDH shows the enzyme to exist as a tetramer in solution (Lamble *et al.*, 2003) and SsGDH is likely to form a 222 tetramer, as seen in TaGDH, formed by operation of the crystallographic twofold axis along *c*. A self-rotation function calculated using *POLARRFN* (Collaborative Computational Project, Number 4, 1994; Fig. 2) reveals no obvious NCS twofold axes in the *ab* plane, suggesting that the NCS twofolds are parallel to the crystallographic *a* or *b* axes. This is confirmed by calculation of a native Patterson using data from 10 to 4 Å, which shows a peak 27% the height of the origin at (0.5, 0.5, 0.5). This would suggest that the NCS twofolds are along *a* and *b* and that the origin of the tetramer lies a quarter of the way along *c*.

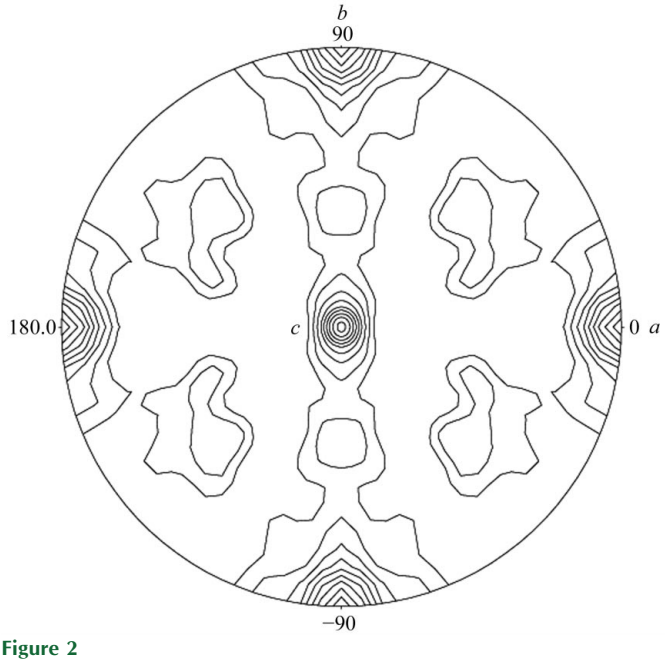


Figure 2
Self-rotation function stereographic projection of the $\kappa = 180^\circ$ section for SsGDH, calculated with ESRF data from 10 to 4 Å resolution and a sphere of 20 Å using the *POLARRFN* program (Collaborative Computational Project, Number 4, 1994). Spherical polar angles are defined as φ, the angle from the Cartesian x axis (a) on the xy plane (ab); ω, the angle from the z axis (c); κ, the rotation around the axis defined by $\varphi\omega$.

Table 1
Data-collection statistics.

Values in parentheses refer to the highest resolution shell.

	In-house	ID14-3
Wavelength (A)	1.54178	0.913
Resolution limits (Å)	50.5–3.3 (3.48–3.30)	148.4–1.8 (1.9–1.8)
Space group	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å)	a = 68.6, b = 88.4, c = 138.8	a = 68.5, b = 90.3, c = 138.9
No. observations	67816	276584
No. unique reflections	13193	77522
Completeness (%)	99.6 (99.2)	96.7 (92.9)
$R_{\rm merge}$† (%)	15.5 (30.1)	6.9 (26.7)
$\langle I/\sigma(I)\rangle$	11.1 (5.3)	12.8 (4.0)
Multiplicity	5.2 (4.9)	3.6 (3.5)
Wilson B (Å^2)	43.8	18.145

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_{hkl,i} - \langle I_{hkl}\rangle| / \sum_{hkl}\langle I_{hkl}\rangle$.

5. Attempts at molecular replacement

The SWISSPROT database was searched for homologues of SsGDH and it was found that the greatest sequence identity (32%) is shared with TaGDH, while only 20% identity is shared with horse LADH. The latter enzyme was used as a model in map improvement for solving the TaGDH structure (John *et al.*, 1994) and despite sharing only 19% sequence identity, the two proteins were shown to have significant structural homology.

Secondary-structure predictions for SsGDH using *PredictProtein* (Rost, 1996) also suggested strong structural similarities with TaGDH, so that SsGDH is also expected to possess a central nucleotide-binding domain that forms a Rossmann-fold structure as well as a catalytic domain formed from both the amino-terminal and carboxy-terminal regions of the protein. Local alignments identified the highest sequence identity (40%) between the first 183 residues of SsGDH and the first 186 residues of TaGDH, which includes the coordination sites for both the catalytic and structural Zn atoms present in TaGDH.

Efforts are being made towards molecular replacement in *AMoRe* (Navaza, 1993) and *CNS* (Brünger *et al.*, 1998) using a variety of TaGDH-derived polyalanine and full residue models, including the full-length chain, Rossmann-fold region and catalytic domain. Multiple-wavelength anomalous dispersion (MAD) phasing using the scattering by the catalytic/structural Zn atoms in the enzyme is also being attempted.

This work was supported by the Engineering and Physical Sciences Research Council, UK and the Biotechnology and Biological Sciences Research Council, UK. Staff at the European Synchrotron Radiation Facility are thanked for their assistance, as is the European Union for funds to access the facility.

References

Albers, S.-V., Elferink, M. G. L., Charlebois, R. L., Sensen, C. W., Driessen, A. J. M. & Konings, W. N. (1999). *J. Bacteriol.* **181**, 4285–4291.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Budgen, N. & Danson, M. J. (1986). *FEBS Lett.* **196**, 207–210.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cusdin, F. S., Robinson, M. J., Holman, G. D., Hough, D. W. & Danson, M. J. (1996). *FEBS Lett.* **387**, 193–195.
De Rosa, M., Gambaccorta, A., Nicolaus, B., Giardina, P., Poerio, E. & Buonocore, V. (1984). *Biochem. J.* **224**, 407–414.
Edwards, K. J., Barton, J. D., Rossjohn, J., Thorn, J. M., Taylor, G. L. & Ollis, D. L. (1996). *Arch. Biochem. Biophys.* **328**, 173–183.
Elshafei, A. M. (1989). *Acta Biotechnol.* **9**, 485–489.
Elzainy, T. A., Hassan, M. M. & Allam, A. M. (1973). *J. Bacteriol.* **114**, 457–459.
Evans, P.R. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances In Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 97–102. Warrington: Daresbury Laboratory.
Grogan, D. W. (1989). *J. Bacteriol.* **171**, 6710–6719.
Grogan, D. W. (1991). *Appl. Environ. Microbiol.* **57**, 1644–1649.
John, J., Crennell, S. J., Hough, D. W., Danson, M. J. & Taylor, G. L. (1994). *Structure*, **2**, 385–393.
Lamble, H. J., Heyer, N. I., Bull, S. D., Hough, D. W. & Danson, M. J. (2003). *J. Biol. Chem.* **278**, 34066–34072.
Lamble, H. J., Milburn, C. C., Taylor, G. L., Hough, D. W. & Danson, M. J. (2004). *FEBS Lett.* **576**, 133–136.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Navaza, J. (1993). *Acta Cryst.* D**49**, 588–591.
Rost, B. (1996). *Methods Enzymol.* **266**, 525–539.
Schafer, G. (1996). *Biochim. Biophys. Acta*, **1277**, 163–200.
Selig, M., Xavier, K. B., Santos, H. & Schönheit, P. (1997). *Arch. Microbiol.* **167**, 217–232.
Siebers, B. & Hensel, R. (1993). *FEMS Microbiol. Lett.* **111**, 1–8.
Smith, L. D., Budgen, N., Bungard, S. J., Danson, M. J. & Hough, D. W. (1989). *Biochem. J.* **261**, 973–977.
Theodossis, A., Walden, H., Westwick, E. J., Connaris, H., Lamble, H. J., Hough, D. W., Danson, M. J. & Taylor G. L. (2004). *J. Biol. Chem.* **279**, 43886–43892.
Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. (1985). *Biochemistry*, **24**, 1346–1357.