Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of zinc-independent sorbitol dehydrogenase from *Rhodobacter sphaeroides* at 2.4 Å resolution

Ansgar Philippsen,[a]* Tilman Schirmer,[a] Martin A. Stein,[b] Friedrich Giffhorn[b] and Jörg Stetefeld[c]

[a]Division of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland, [b]Lehrstuhl für Angewandte Mikrobiologie, Universität des Saarlandes, Im Stadtwald, 66123 Saarbrücken, Germany, and [c]Division of Biophysical Chemistry, Biozentrum, University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland

Correspondence e-mail: ansgar.philippsen@unibas.ch

Received 23 August 2004
Accepted 28 December 2004

PDB References: sorbitol dehydrogenase, experimentally derived model, 1k2w, r1k2wsf; theoretical model, 1uzo.

Recombinant sorbitol dehydrogenase (SDH) from *Rhodobacter sphaeroides* has been crystallized in the absence of the cofactor NAD(H) and its structure determined to 2.4 Å resolution using molecular replacement (refined R and $R_{\rm free}$ factors of 18.8 and 23.8%, respectively). As expected from the sequence and shown by the conserved fold, SDH can be assigned to the short-chain dehydrogenase/reductase protein family. The cofactor NAD and the substrate sorbitol have been modelled into the structure and the active-site architecture, which displays the highly conserved catalytic tetrad of Asn-Ser-Tyr-Lys residues, is discussed in relation to the enzyme mechanism. This is the first structure of a bacterial SDH belonging to the SDR family.

1. Introduction

Sorbitol dehydrogenase (SDH) from the bacterium *Rhodobacter sphaeroides* has been characterized as an NAD(H)-dependent dimeric protein with a subunit molecular weight of about 29 000 Da (Schauder *et al.*, 1995). The enzyme catalyzes the interconversion of sorbitol (D-glucitol) and D-fructose, of L-iditol and L-sorbose and of galactitol and D-tagatose (Fig. 1), with a preference for the substrate pair sorbitol/D-fructose. In accordance with the biochemical data, the cloned SDH gene encodes a peptide of 256 amino-acid residues with a derived molecular weight of 27 012 Da. The peptide sequence contains an N-terminal Gly-X_3-Gly-X-Gly motif as part of the NAD(H)-binding region and, on the basis of multiple sequence alignments, the conserved residues Asn110, Ser139, Tyr152 and Lys156 that are characteristic of most proteins of the short-chain dehydrogenase/reductase (SDR) family (Stein *et al.*, 1997). SDRs form a large functionally heterogeneous protein family with about 3000 primary and 30 three-dimensional structures deposited in databases (Oppermann *et al.*, 2003). Although SDRs display low sequence identities ranging from 15 to 30%, their three-dimensional structures have similar and characteristic structural patterns. In addition, most of the SDRs contain a set of highly conserved residues, including the almost invariant Tyr152 and Lys156, which are essential for catalysis and constitute the architecture of the active site. In order to validate the predicted classification of SDH, the recombinant enzyme was overproduced for crystallization trials and subsequent X-ray structure analysis. Judged on the polyol projections (Fig. 1), the stereospecific arrangements of the hydroxyl groups at the chiral C atoms C2 and C3 may define the substrate specificity. The goal of this work was to solve the first crystal structure of a bacterial sorbitol dehydrogenase, thereby contributing to a better understanding of the enzymatic mechanism.

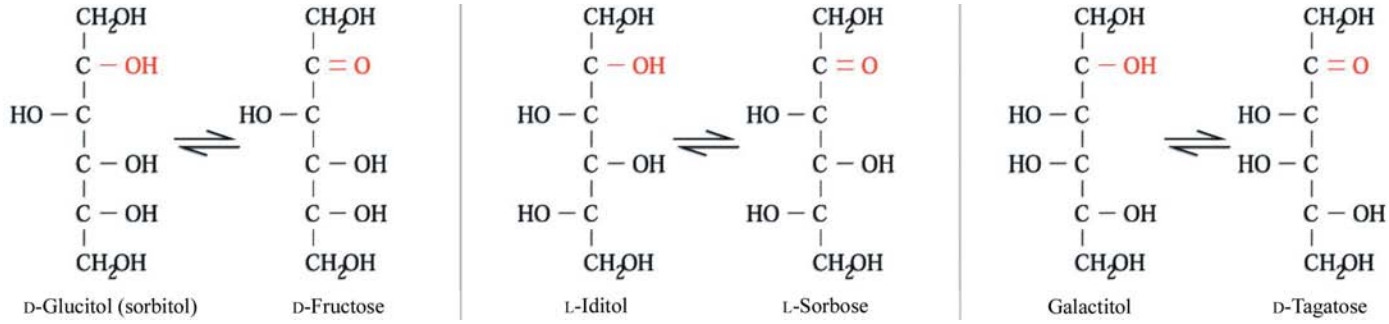

Figure 1
R. sphaeroides SDH-catalyzed reactions. These three reactions were identified biochemically as being carried out by sorbitol dehydrogenase from *R. sphaeroides*. Based on these projections, the substrate specificity may be defined by the stereospecific arrangements of the hydroxyl groups at the chiral C atoms C2 and C3.

Table 1
Data-collection and refinement parameters.

Values in parentheses are for the last shell.

Data-collection statistics	
Space group	$P3_121$
Unit-cell parameters (Å)	$a = b = 67.46$, $c = 191.04$
Resolution range (Å)	30–2.4 (2.57–2.42)
Unique reflections	21038
Completeness (%)	91.5 (92.1)
Redundancy	3.0 (2.4)
R_{sym}† (%)	7.8 (22.5)
$\langle I/\sigma(I)\rangle$	8.0 (3.2)
Refinement statistics	
R factor‡ (%)	18.8
R_{free} (%)	23.8
No. of protein atoms	3780
No. of water molecules	170
R.m.s. deviations from ideal values	
Bonds (Å)	0.005
Angles (°)	1.14
Ramachandran plot (%)	
Most favoured region	91.9
Additional allowed region	8.1
Generously allowed region	0
Disallowed region	0

† $R_{sym} = \sum|I - \langle I\rangle|/\sum I$. ‡ R factor $= \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$.

2. Results and discussion

2.1. Quality of the model

Here, we present the crystal structure of *R. sphaeroides* SDH at 2.4 Å resolution. The structure was solved by the Patterson search method using *meso*-2,3-butanediol dehydrogenase (amino-acid residues 6–252, with an r.m.s.d. of 3.3 Å for the main-chain atoms; PDB code 1geg; Otagiri *et al.*, 2001) as a search template. The asymmetric unit contains two monomers of *R. sphaeroides* SDH ($V_M = 2.7$ Å^3 Da^{-1}). All 256 residues are resolved in the well defined electron density, a region of which is shown in Fig. 2, phased from the initial molecular-replacement solution as well as the final model. The final model was refined in the resolution range 24–2.4 Å, resulting in crystallographic R factors of 18.8% (working set) and 23.8% (free set), including 512 amino-acid residues and 170 water molecules. The parameters of the final model are given in Table 1.

The r.m.s. coordinate difference between the two monomers is 0.14 Å for the main-chain atoms and 0.70 Å for the side-chain atoms. The average crystallographic temperature factor (B factor) is 33 Å^2 for monomer I and 31 Å^2 for monomer I′. This difference of temperature factor between the monomers is most prominent in region 29–80 (40.2 Å^2 in I and 27.6 Å^2 in I′), which forms crystal contacts in monomer I′ but not in monomer I, explaining the observed additional flexibility for monomer I (data not shown).

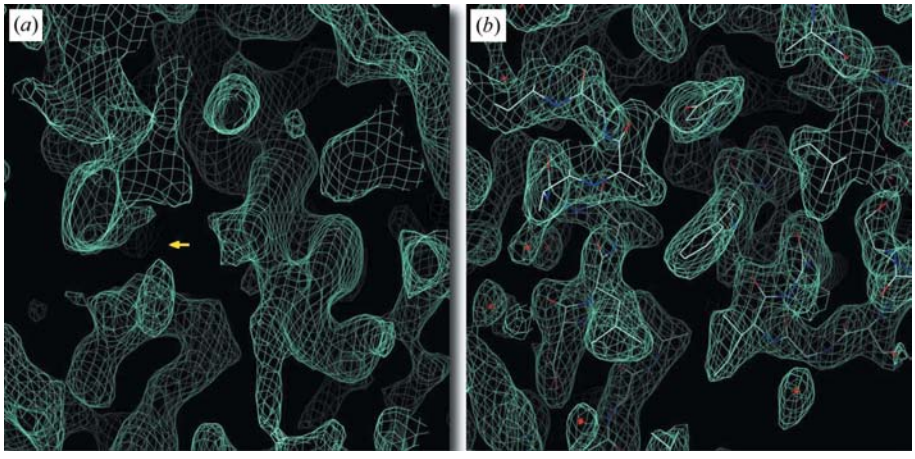

Figure 2
Electron-density comparison. $2F_{obs} - F_{calc}$ electron density contoured at 1σ using phase information from the initial molecular-replacement solution (*a*) and from the final refined model (*b*). Although the connectivity was not completely defined (as indicated by the arrow), the molecular-replacement model served as a good starting point to build the final structure.

2.2. Overall fold

A monomeric subunit of the dimeric SDH from *R. sphaeroides* is composed of a NAD-binding domain that is formed by seven parallel β-strands and six parallel α-helices, three on each side of the β-sheet (Fig. 3*b*). This fold is characteristic of the SDR family. The substrate-binding domain is formed by two clamp-like extensions from the NAD-binding domain (Fig. 3*a*). The

larger extension (residues 188–221) is hinged to the NAD-binding domain *via* two loops and exhibits higher average temperature factors than the remaining structure (56.7 Å^2 for residues 188–221 compared with an overall temperature factor of 31.9 Å^2), which is indicative of a flexibility required for enzymatic activity.

The crystallographic dimer, whose monomers are denoted I and I′, does not correspond to the biological dimer (Fig. 3*c*, left), whose monomers are named I and II. Upon application of the appropriate crystallographic symmetry operations, both monomers of the asymmetric unit (I and I′) form tight tetramers (I-II-III-IV and I′-II′-III′-IV′) with 222 symmetry. The biological dimer was deduced by comparison with the molecular-replacement model, which forms a very similar dimer (1geg; Otagiri *et al.*, 2001).

The dimerization interface between monomers I and II is formed by the two long helices extending from the NAD-binding domain to the sugar-binding domain (Fig. 3*a*). Another dimer, packed closely to this one, is crystallographically related (Fig. 3*c*). In both the dimer and tetramer the active site and the NAD-binding groove are freely accessible (not shown). Upon dimerization of monomers I and II, 3244 Å^2 of surface area becomes buried. Upon dimerization of the dimer I-II with III-IV, an additional 3086 Å^2 per monomer becomes buried, resulting in a total buried surface area of 12 660 Å^2. Based on these numbers, the *in vivo* oligomerization state could be tetrameric. Sedimentation-centrifugation analysis performed previously (Schauder *et al.*, 1995) suggests *R. sphaeroides* SDH to be dimeric.

2.3. NAD-binding region

No electron density was found in the NAD-binding pocket. The following side chains border the NAD-binding site: 9, 12–14, 33–35, 58–60, 86–88, 138, 182–185 and 187–190, as well as the active-site residues Ser138, Tyr152 and Lys156. These residues were used for the least-squares superposition with the molecular-replacement structure (Fig. 4*a*) and the resulting fit is remarkably good. NAD modelled in this way fits nicely into the surface groove generated by *R. sphaeroides* SDH (Figs. 4*a* and 5*a*), despite its absence from the *R. sphaeroides* SDH structure. Therefore, no conformational change is expected upon cofactor binding. The binding

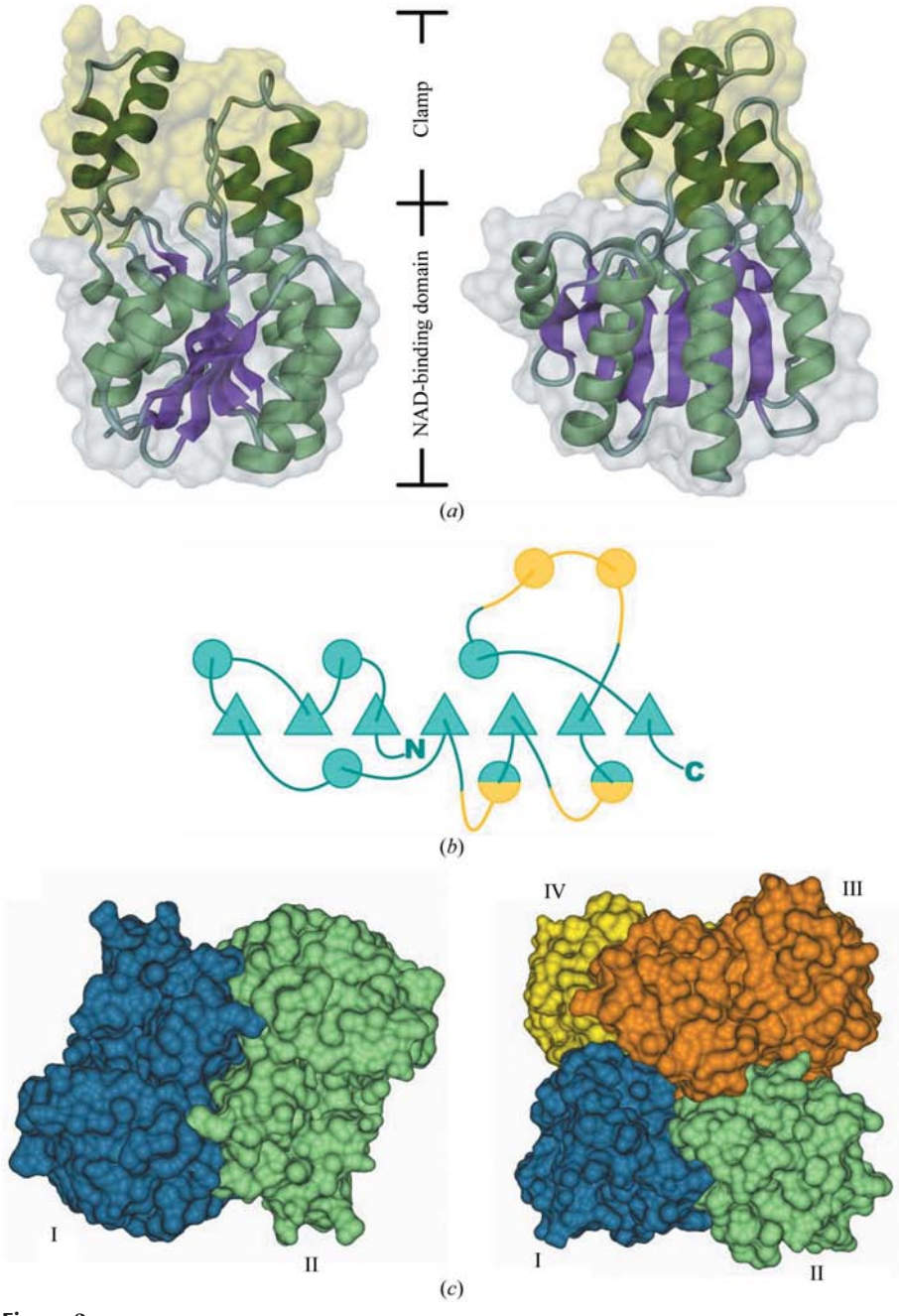

Figure 3
Overall structure. (*a*) Secondary-structure cartoon of *R. sphaeroides* SDH: the two domains of *R. sphaeroides* SDH are presented as secondary-structure cartoons with the molecular surface overlaid transparently. The surface of the NAD-binding domain is shown in white, enveloping its typical seven-stranded β-sheet and flanking α-helices. The substrate-binding domain (clamp), shown as a yellow surface, is an extension from the NAD domain. The left view is approximately along the β-sheet, while the right is a 90° rotation along the vertical axis. (*b*) Schematic diagram of secondary-structure elements and their consecutive arrangement, colour-coded by domains (blue, NAD-binding; yellow, substrate binding). (*c*) Dimeric and tetrameric packing. On the left are the dimers that bury the largest surface area between them (3244 Å^2); on the right the tetramers as found in the crystal packing. Since the surface buried upon tetramerization is considerable (12 660 Å^2) and all four active sites are still accessible, *R. sphaeroides* SDH could function as a tetramer *in vivo*.

pocket is specific to NAD, not NADP, since the characteristic groove is missing that could accommodate the additional phosphate moiety (Lesk, 1995).

2.4. Active-site and substrate modelling

A prominent substrate tunnel is present in the structure, as indicated in Fig. 5(*b*). The tunnel is flanked by the nicotinamide ring at one end, which is also the location of the active site. A molecule of sorbitol was placed manually in the substrate tunnel. The resulting ternary complex of *R. sphaeroides* SDH, NAD and sorbitol was subjected to energy minimization, yielding a model with a main-chain r.m.s.d. of 0.10 Å and an overall r.m.s.d. for all heavy protein atoms of 0.19 Å.

The resulting active-site architecture (Fig. 4*b*) supports the reaction mechanism proposed by Tanaka *et al.* (1996) and extended by Filling *et al.* (2002). The deprotonated tyrosine (Tyr152) accepts the proton from the sorbitol C2 hydroxyl group and the C4 carbon of the nicotinamide accepts the hydride ion. The deprotonated state of the tyrosine is facilitated by an electron-sink or proton-relay system formed by an intricate hydrogen-bonding network consisting of Asn110, Lys156 and the ribose hydroxyl groups, as well as several waters and other protein residues (see Fig. 4*b*). In particular, the previously stated hypothesis of a catalytic tetrad (Filling *et al.*, 2002), which includes Asn110 as an essential component in the proton-relay system of SDRs, is verified.

The sorbitol binding is mediated mainly by hydrogen bonding of the hydroxyl groups at the C2, C3 and C5 position to O atoms in Tyr152 (OH), Ser139 (OG), NAD (NO7), Gly183 (O) and Gln140 (OE1).

Well defined water molecules (not shown) are found within the active site of the experimentally determined apo form of the enzyme, *e.g.* at hydrogen-bonding distance to the hydroxy group of Tyr152 as well as the amino group of Lys156.

While the active-site residues Ser139, Tyr152 and Lys156 of our model could be superimposed almost perfectly onto the equivalent residues in the molecular-replacement model 1geg, the active-site region, in particular the sugar-binding pocket, bore no similarity, as expected from the different substrates (sorbitol and 2,3-butanediol, respectively; Otagiri *et al.*, 2001).

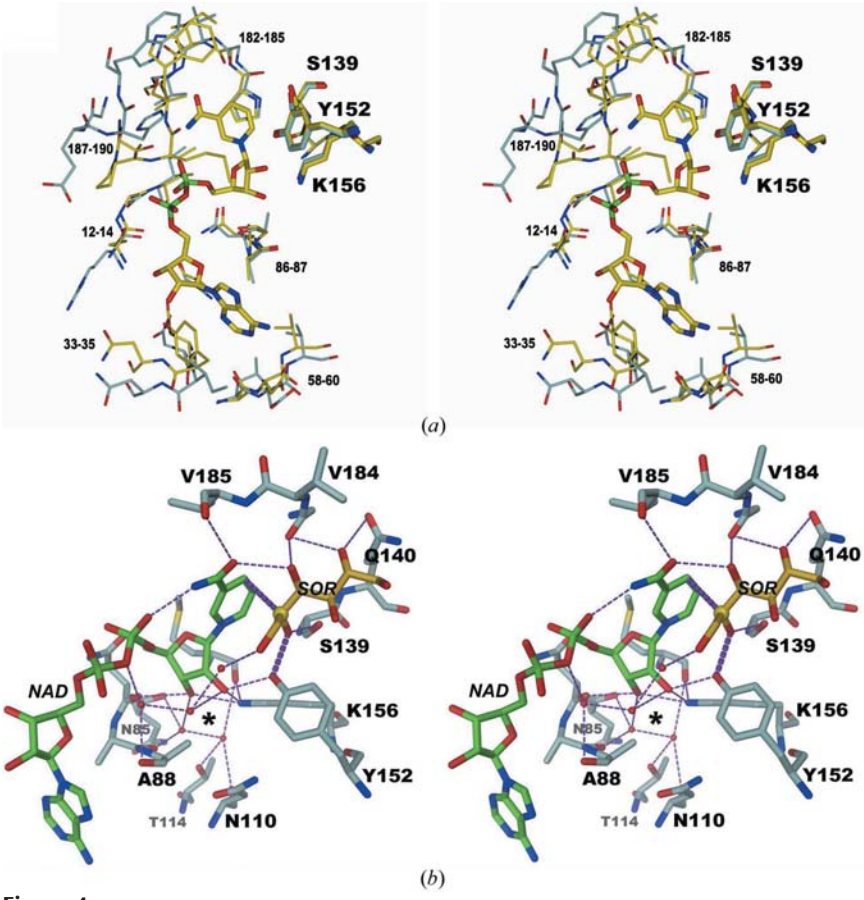

Figure 4
Active site of *R. sphaeroides* SDH with modelled NAD and sorbitol. (*a*) Stereo representation of NAD-binding residues in the molecular-replacement model 1geg (yellow) fitted to the corresponding residues in *R. sphaeroides* SDH (white). The almost perfect overlap of the NAD-binding residues is the basis for the modelling of the cofactor NAD into the active site of *R. sphaeroides* SDH. (*b*) Stereo representation of the active site of *R. sphaeroides* SDH based on the apo form of the enzyme with modelled and energy-minimized NAD (green) and sorbitol (yellow). The thick purple bonds indicate the main interaction of the C2 carbon and hydroxyl group of the sugar sandwiched between the nicotinamide ring and Tyr152. The intricate hydrogen-bonding network is indicated by thin purple lines; the region of the electron sink is denoted by an asterisk.

2.5. Biotechnological implications

Sorbitol (D-glucitol) is an acyclic polyol occurring naturally in many plants and fruits. It is produced industrially in amounts of $>500\,000$ t y^{-1} and used for various applications. The main use is as an additive and sweetener in food and pharmaceutical products and as the starting material for the synthesis of vitamin C (Silveira & Jonas, 2002). Specific determination of sorbitol in foods and other complex materials is required in food chemistry and can be accomplished by enzymatic analysis with SDH, which oxidizes sorbitol to D-fructose in the presence of NAD. However, the SDH used commercially is unspecific, acting similarly on xylitol, a polyol that is used like sorbitol. Differentiation between the polyols is possible but requires the use of additional auxiliary enzymes in the assay (Beutler, 1984*a*,*b*). In order to minimize interference with assays and in view of an increasing demand for automation, specific assays are required that depend on one single enzyme. *R. sphaeroides* SDH meets these requirements because it is practically unaffected by xylitol (Schauder *et al.*, 1995) and the side substrates L-iditol and galactitol are not expected to interfere with the assay

as they do not occur in nature and are not permitted as food additives, respectively. The structural analysis of *R. sphaeroides* SDH gives insight into the active site of this enzyme and the modelling of the substrates into this site provides a rational basis for improving the enzyme's specificity by molecular evolution.

3. Materials and methods

3.1. Crystallization and data collection

Crystallization trials were carried out on purified and lyophilized material (Stein *et al.*, 1997). A 10 mg ml^{-1} solution was prepared by dissolving 4.18 mg lyophilized *R. sphaeroides* SDH (5.52 mg total) in 418 µl water. Successful sitting-drop crystallization conditions centred around 100 m*M* MES pH 7.0, 18–21% PEG 4000, 6–11% methanol. Large rhombohedral shaped crystals of dimensions around 500 µm grew readily after a few days. A data set was collected at 277 K on the in-house system, a rotating-anode generator with a MAR 345 image-plate detector. Processing and scaling was performed with *IPMOSFLM* and *SCALA* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994; see Table 1). 10% of the reflections were flagged for use as a test for further processing. The space group was identified to be *P*321, $P3_121$ or $P3_221$. It was not possible to determine the translational component of the threefold axis because of missing data and therefore all processing was performed in *P*321. The true space group was found to be $P3_121$ during molecular replacement. The presence of two monomers per asymmetric unit was assumed, leading to a Matthews coefficient (Matthews, 1968) of 2.7 Å^3 Da^{-1}.

3.2. Structure determination and refinement

To identify a candidate structure for molecular replacement, a *BLASTP* (Altschul *et al.*, 1997) query using the sequence of *R. sphaeroides* SDH against all sequences in the Protein Data Bank (Berman *et al.*, 2000) was used. The search returned *meso*-2,3-butanediol dehydrogenase from *Klebsiella pneumoniae* (PDB code 1geg; Otagiri *et al.*, 2001), with 38.0% sequence identity using pairwise alignment. The molecular-replacement model was obtained by stripping each residue in 1geg to the common atoms shared with the corresponding residue in *R. sphaeroides* SDH based on sequence alignment. A Perl script was written for this purpose and is available on request. Molecular replacement was performed with *AMoRe* (Navaza, 2001). The model obtained with molecular replacement was first subjected to a simulated-annealing procedure with subsequent energy minimization using strict NCS constraints. Upon inspection of the resulting model and electron density, 64 residues per monomer located in ill-defined or fragmented regions of electron density were removed, whereas 32 incomplete side chains (from the stripping procedure) per monomer were completed. At this point, 72% of all protein atoms were included in the model. The remaining parts of the model were completed in a cyclic fashion: residues were built into regions of well defined electron density and the resulting extended model was used to regenerate the electron density, in turn allowing further residues to be built into emerging electron-density regions. Strict NCS constraints were applied during this procedure. Model building was performed with *O* (Jones *et al.*, 1991); refinement and water placement was performed with *CNS* (Brünger *et al.*, 1998). After completion of the model, it became apparent that the two monomers (I and I′) were not identical, in particular for some side-chain conformations. NCS constraints were therefore removed. The different conformations in the monomers were adjusted and refined subsequently. As a last step, 170 waters were added using the water-picking routine from *CNS* (Brünger *et al.*, 1998) and manual verification of the suggested positions.

3.3. Active-site modelling

The NAD molecule from the molecular-replacement model was placed straightforwardly into the *R. sphaeroides* SDH structure. Sorbitol was fitted into the approximate position by hand, orienting the C2 carbon according to the proposed reaction mechanism. The subsequent steps were performed with *GROMACS* (Lindahl *et al.*, 2001): the structure was completed with H atoms, placed into a water-filled box and subjected to iterative energy minimization; all these steps were performed with default settings. Convergence was reached after about 100 cycles.

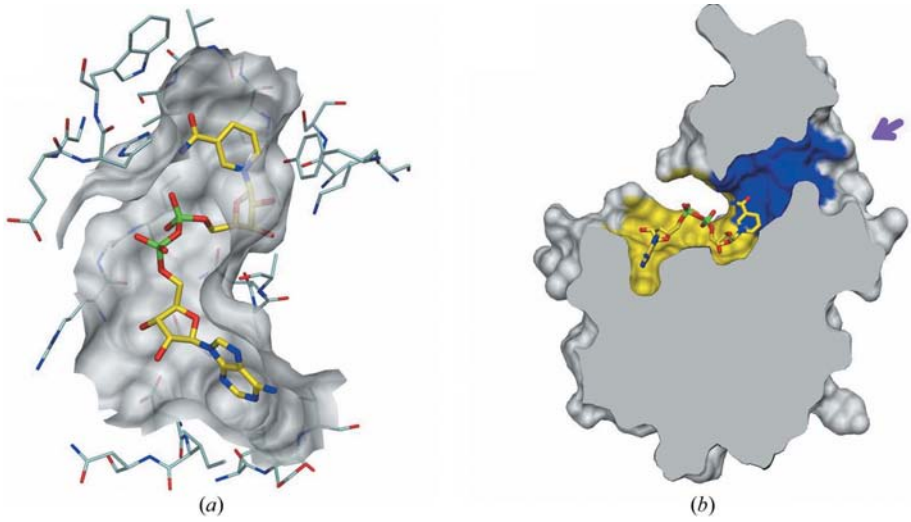

Figure 5
NAD-binding pocket and substrate-access tunnel. (*a*) Binding pocket for NAD formed by *R. sphaeroides* SDH. (*b*) Side view of the surface of *R. sphaeroides* SDH with bound NAD. The surface is cut to reveal the NAD-binding site (yellow) and substrate-binding site (blue), as well as the active-site access direction (arrow).

3.4. Structure analysis

The quality of the resulting structure was analyzed with *PROCHECK* (Laskowski *et al.*, 1998). The secondary structure was obtained with *DSSP* (Kabsch & Sander, 1983). All multiple sequence alignments were performed with *BLASTP* (Altschul *et al.*, 1997). Fold-similarity searches were performed with *DALI* (Holm & Sander, 1993). R.m.s. deviations were calculated with *LSQKAB* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). All visualization was performed with *DINO* (http://www.dino3d.org), a comprehensive visualization program developed and maintained by one of the authors (AP). Molecular surfaces and solvent-accessible surface areas were calculated with *MSMS* (Sanner *et al.*, 1996).

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Beutler, H. O. (1984*a*). *Methods of Enzymatic Analysis*, 3rd ed., edited by H. U. Bergmeyer, Vol. 6, pp. 356–362. Weinheim: Verlag Chemie.

Beutler, H. O. (1984*b*). *Methods of Enzymatic Analysis*, 3rd ed., edited by H. U. Bergmeyer, Vol. 6, pp. 484–490. Weinheim: Verlag Chemie.

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.

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

Filling, C., Berndt, K. D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jörnvall, H. & Oppermann, U. (2002). *J. Biol. Chem.* **277**, 25677–25684.

Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. (1998). *Curr. Opin. Struct. Biol.* **8**, 631–639.

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

Lindahl, E., Hess, B. & van der Spoel, D. (2001). *J. Mol. Mod.* **7**, 306–317.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B. & Jornvall, H. (2003). *Chem. Biol. Interact.* **143–144**, 247–253.

Otagiri, M., Kurisu, G., Ui, S., Takusagawa, Y., Ohkuma, M., Kudo, T. & Kusunoki, M. (2001). *J. Biochem (Tokyo)*, **129**, 205–208.

Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). *Biopolymers*, **38**, 305–320.

Schauder, S., Schneider, K. H. & Giffhorn, F. (1995). *Microbiology*, **141**, 1857–1863.

Silveira, M. M. & Jonas, R. (2002). *Appl. Microbiol. Biotechnol.* **59**, 400–408.

Stein, M. A., Schäfer, A. & Giffhorn, F. (1997). *J. Bacteriol.* **179**, 6335–6340.

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