

Methanoarchaeal sulfolactate dehydrogenase: prototype of a new family of NADH-dependent enzymes

Adriana Irimia^{1,3}, Dominique Madern^{1,3},
Giuseppe Zaccai^{1,2} and
Frédéric MD Vellieux^{1,*}

¹Laboratoire de Biophysique Moléculaire, Institut de Biologie Structurale J-P Ebel CEA CNRS UJF, Grenoble, France and

²Institut Laue Langevin, Grenoble, France

The crystal structure of the sulfolactate dehydrogenase from the hyperthermophilic and methanogenic archaeon *Methanocaldococcus jannaschii* was solved at 2.5 Å resolution (PDB id. 1RFM). The asymmetric unit contains a tetramer of tight dimers. This structure, complexed with NADH, does not contain a cofactor-binding domain with 'Rossmann-fold' topology. Instead, the tertiary and quaternary structures indicate a novel fold. The NADH is bound in an extended conformation in each active site, in a manner that explains the pro-S specificity. Cofactor binding involves residues belonging to both subunits within the tight dimers, which are therefore the smallest enzymatically active units. The protein was found to be a homodimer in solution by size-exclusion chromatography, analytical ultracentrifugation and small-angle neutron scattering. Various compounds were tested as putative substrates. The results indicate the existence of a substrate discrimination mechanism, which involves electrostatic interactions. Based on sequence homology and phylogenetic analyses, several other enzymes were classified as belonging to this novel family of homologous (S)-2-hydroxyacid dehydrogenases.

The EMBO Journal (2004) 23, 1234–1244. doi:10.1038/sj.emboj.7600147; Published online 11 March 2004

Subject Categories: structural biology; cellular metabolism

Keywords: coenzyme M; dehydrogenase; hyperthermostable; methanogens; pro-S hydrogen transfer

Introduction

The hyperthermophilic methanoarchaeon *Methanocaldococcus jannaschii* (*Mj*) lives in deep sea marine hydrothermal vents located on the East Pacific rise at depths ca. 2600 m (Vieille and Zeikus, 2001). The conditions in this biotope are anaerobic, with pH ranging from 5.5 to 6.5 and temperatures between ca. 50 and 100°C. Anaerobic carbon dioxide respiration is characteristic of such hyperthermophiles (Stetter,

1999). The anaerobic reduction of fermentation products (acetate, formate or single-carbon compounds) by these methanogens leads to a large production of methane (Kasting and Siefert, 2002). Since the release of its full genomic sequence (Bult *et al*, 1996), *M. jannaschii* is one of the most extensively studied methanoarchaea. Structural genomics investigations on methanoarchaeal proteins (including those on *Mj* proteins) have allowed to identify several structures with new folds (e.g. Hwang *et al*, 1999; Wang *et al*, 2000).

Since the beginning of *Mj* genome annotation, the functional assignment of gene *Mj1425* has been subject to controversy. According to sequence comparisons and data bank annotation, *Mj1425* was first classified as a gene encoding for a new type of 'Archaeal malate dehydrogenase' with no homology to 'canonical' malate dehydrogenases (Graupner *et al*, 2000). Using only oxaloacetate/malate as substrates during characterization, the homologous protein from *Methanothermobacter fervidus* was classified as an NAD-dependent malate dehydrogenase (MalDH, EC 1.1.1.37; Honka *et al*, 1990). This inappropriate annotation was reinforced when it was claimed that two MalDHs are present in *Methanothermobacter thermoautotrophicus* (Thompson *et al*, 1998), one of which is homologous to the [LDH-like] MalDH enzymes. Here again, only oxaloacetate/malate were used to probe enzymatic activity. More than 150 genomic sequences have been deposited, and this has allowed to identify several putative MalDHs such as *Mj1425*, which are stored at the NCBI data bank with the 'cluster of orthologous group' (COG) number 2055.

The sequence attribution for *Mj1425* was misleading because MalDHs are very well-known enzymes. Phylogenetic and structural analyses have shown that, within the 2-hydroxyacid dehydrogenases, 'canonical' NAD(P)-dependent malate and lactate dehydrogenases (LDHs) form a homologous system (Madern, 2002 and reference therein). They possess a coenzyme-binding domain adopting the well-known 'Rossmann fold', which is a six-stranded parallel β -sheet core surrounded on both sides by helices (Rossmann *et al*, 1975). Based on a set of more complete biochemical experiments, an improved attribution was proposed for *Mj1425*. The enzyme uses NAD(P)H as coenzyme (with higher affinity for NADH), and uses the pro-4S hydrogen of the cofactor (Graupner and White, 2001) to catalyze the reduction of a broad range of α -keto acids (e.g. sulfolactate, oxaloacetate and α -ketoglutarate) to the [S]-isomers of the corresponding 2-hydroxy acids (sulfolactate, malate and 2-hydroxyglutarate). In the reverse reaction, only the (S)-isomers of the 2-hydroxy acids are accepted as substrates. The preference for sulfolactate/sulfolactate over the other substrates has led to the classification of this enzyme as a sulfolactate dehydrogenase (SLDH, EC 1.1.1.272, Swiss-Prot entry Q58820; Graupner *et al*, 2000). The SLDH enzyme is thought

*Corresponding author. Laboratoire de Biophysique Moléculaire, Institut de Biologie Structurale J-P Ebel CEA CNRS UJF, UMR-5075, 41 rue Jules Horowitz, 38027 Grenoble Cedex 01, France.
Tel.: +33 438 789 605; Fax: +33 438 785 494;
E-mail: vellieux@ibs.fr

³These authors contributed equally to this work

Received: 19 November 2003; accepted: 6 February 2004; published online: 11 March 2004

to be involved in the biosynthesis of both methanogenic cofactors coenzyme M and methanopterin through the oxidation of sulfolactate and α -ketoglutarate, respectively (Graupner *et al*, 2000; White, 2001).

We describe here the 2.5 Å resolution crystal structure of *Mj*SLDH complexed with NADH. In contrast to the oligomeric state in room temperature solutions and at 56°C (a homodimer), the three-dimensional (3-D) structure shows an octameric aggregate in the asymmetric unit. The tertiary structure of the subunit is different from that found in the LDH/MalDH enzymes, and does not contain a dinucleotide-binding domain with Rossmann-fold topology (Rossmann *et al*, 1975). Further, unlike in other dehydrogenases, each NADH-binding site comprises residues from the two subunits that make up the tight dimer. We tested the recognition of 13 substrates by *Mj*SLDH, and describe a discrimination mechanism that precedes catalysis. The phylogenetic distribution of 31 additional homologous sequences is also discussed. Our results show that *Mj*SLDH belongs to a new class of NAD(P)-dependent dehydrogenases with a novel fold and unique stereospecificity.

Results

The asymmetric unit of the monoclinic crystals contains a tetramer of tight dimers with overall 222 symmetry (Supplementary Figure 1S). The octamer fits in a parallelepiped of dimensions ca. 129 Å × 118 Å × 94 Å, with a quaternary structure organization similar to that of the 'classical' tetrameric dehydrogenases with 222 symmetry (Rossmann *et al*, 1973). Within each tight dimer, the surface area per monomer that is buried in the subunit interface is 3120 Å² (i.e. 18.7% of the total monomer surface). The interface areas in the tetramer of tight dimers are far smaller (Supplementary Figure 1S). This and several other data suggest that the tight dimer is the smallest entity that could be biologically relevant (*vide infra*).

The refined molecular model has proper stereochemistry, which reflects the quality of the recorded diffraction data (Table I). The ϕ/ψ plot (Ramakrishnan *et al*, 1965) shows 2060-amino-acid residues in the most favored regions and 240 in the additionally allowed regions. Four residues (E296A, E296C, K191E and K202F) are in disallowed regions of conformational space. K191E and K202F are located in mobile domains (*vide infra*), which are less ordered than the rest of the polypeptide. E296A and E296C are located in nine-residue-long loops at the protein surface that are not involved in protein contacts and are also less ordered.

The tight dimer and topology of the monomer

The tight dimer of *Mj*SLDH is an ovoid-shaped globular assembly of two identical subunits, with length ca. 90 Å and diameter ca. 54 Å (Figure 1A). The two subunits face each other, with a proper two-fold noncrystallographic symmetry axis relating the two halves. The area of interaction comprises mostly four α -helices (two per subunit, α H and α I), in addition to facing mixed β -sheet segments (Figure 1B). The interacting surfaces at the monomer–monomer interface are essentially hydrophobic. This hydrophobic interface is surrounded by electrostatic interactions (seven ion pairs and 14 hydrogen bonds that involve lysine and glutamate residues, plus six hydrogen bonds involving residues other than these

two types). The protein contains a high fraction of charged residues (predominantly lysines and glutamates, 11.9 and 9.3%, respectively), which are found on the surface of the tight dimers where they form ion pairs or ion-pair networks (Supplementary Table IS). This correlates well with observations made for other hyperthermostable proteins (Cambillau and Claverie, 2000).

Each polypeptide chain comprises 344 amino-acid residues organized in two domains (Figures 1C and D): (i) the largest domain, a dinucleotide-binding domain that contains both chain termini and most of the active site residues; and (ii) a mobile domain that covers the active site. The latter has the highest isotropic temperature factors and shows the largest location deviations within the eight identical subunits present in the asymmetric unit.

The dinucleotide-binding domain (residues 66–178 and 219–316; Figure 1C and Supplementary Figure 3S) is a triangular wedge made up of a central, seven-stranded mixed β -sheet, sandwiched between three helices on one side and two helices on the other side. The topology is a three-layer $\alpha\beta\alpha$ sandwich (Figure 1D, CATH fold 3.40.91; Orengo *et al*, 1997), which most closely resembles the topology of the cleavage domain of the *FokI* restriction endonuclease (Wah *et al*, 1997). This dinucleotide-binding domain does not include the classical 'Rossmann fold' seen in most dehydrogenases (e.g. LDH, MalDH, glyceraldehyde-3-phosphate dehydrogenase; Rossmann *et al*, 1975).

The domain is extended by a smaller subdomain (residues 1–65 and 317–344; Figure 1C and Supplementary Figure 3S) that comprises both the N- and C-termini (located ca. 8.0 Å apart) and has a mixed four-helix barrel topology: helices α A, α B, α C, contiguous in the sequence, form an antiparallel pack, while helix α K at the C-terminus is parallel to both α A and α C. This four-helix bundle topology (Figure 1D) is very different from typical four-helix bundles, where the four helices are consecutive and antiparallel to each other (Weber and Salemme, 1980), and is most closely related to the orthogonal α -helical bundle (domain III) present in the DNA recombination protein RuvA (CATH topology 1.10.8; Rafferty *et al*, 1996).

The second domain (residues 179–218; Figure 1C and Supplementary Figure 3S) is termed a 'mobile domain' by analogy to the 3-D structures of enzymes with mobile segments covering their active sites (e.g. triose phosphate isomerase, MalDH): the electron density is less well defined for the side chains in the mobile domains of monomers A, C, F, E and G. These regions also have higher isotropic temperature factors than the rest of the structure (ca. 80 Å², compared to an average isotropic temperature factor ca. 48 Å²; Supplementary Figure 1S, Table I). Further, these regions do not obey the eight-fold noncrystallographic symmetry. All this suggests high flexibility for these segments of the polypeptide chains, which are located above the cofactor-binding clefts. Such domain mobility might therefore be important for an assumed function of closure over the *Mj*SLDH active sites. The mobile domain has a two-layer sandwich $\alpha\beta$ topology (Figure 1D), which resembles that of the DNA-binding domain of the transcription factor GATA-1 (CATH topology 3.30.50; Omichinski *et al*, 1993).

Overall, the topology of the SLDH monomer (Figure 1D) bears resemblance to that of the putative glycerate kinase from *Thermotoga maritima* (PDB id. 1O0U; Joint Center for

Table 1 Crystallographic parameters, data processing and refinement statistics

Data sets	Inflexion	Peak ^a	Remote	Merged ^a
<i>Diffraction data</i>				
X-ray source	ESRF CRG beamline BM30A (FIP)			
Temperature (K)	100			
No. of crystals	1			
Space group	P2 ₁			
Unit cell (<i>a</i> , <i>b</i> , <i>c</i> ; Å)	81.0; 203.8; 100.2			
(α , β , γ ; deg)	90.0; 112.0; 90.0			
Wavelength (Å)	0.979682	0.979558	0.977769	—
Resolution range (Å)	32.8–2.5	32.8–2.5	32.8–2.5	32.8–2.5
Highest resolution shell	(2.58–2.50)	(2.58–2.50)	(2.58–2.50)	(2.58–2.50)
No. of measurements	549 366 (2475)	595 976 (3064)	446 590 (2515)	1 430 402 (6966)
No. of unique reflections ^b	169 111 (2259)	164 505 (2815)	151 340 (2473)	93 855 (3587)
Multiplicity	3.3 (1.1)	3.6 (1.1)	3.0 (1.0)	15.2 (1.9)
Completeness (%)	86.3 (17.0)	85.5 (22.5)	80.7 (20.3)	90.1 (34.5)
R_{sym} ^c (%)	4.1 (34.1)	5.22 (18.1)	4.74 (33.6)	7.5 (22.8)
R_{ano} ^d (%)	8.8 (38.4)	10.4 (68.1)	7.96 (59.9)	—(—) ^b
Signal to noise ($\langle I/\sigma_I \rangle$)	24.0 (2.5)	22.3 (4.8)	19.6 (4.4)	30.5 (4.3)
<i>Phasing</i>				
Overall figure of merit		0.64		
R_{cullis} ^e (%)		0.41		
Phasing power ^f		3.32		
Lack of closure (deg)		29.12		
<i>Refinement</i>				
Asymmetric unit content		8 monomers		
Solvent content (%)		50		
Model composition				
No. of amino-acid residues (8 chains)		2752		
No. of protein atoms (non-H)		20 984		
No. of water atoms		506		
R_{f} ^g		0.198		
R_{free} ^h		0.242		
Temperature factor from Wilson plot (Å ²)		24.1		
Mean isotropic temperature factor (Å ²)		46.6		
r.m.s. deviation in temperature factors				
Bonded main chain atoms (Å ²)		1.37		
Bonded side-chain atoms (Å ²)		2.45		
r.m.s. standard deviation from ideal values				
Bond lengths (Å)		0.009		
Bond angles (deg)		1.14		
Dihedral angles (deg)		21.6		
Improper angles (deg)		0.86		

The values in parentheses correspond to the highest resolution shells.

^aThe data set corresponding to the wavelength of the peak was used for refinement, except in the final refinement round where the merged data set (three wavelengths) was used.

^bIncludes the Bijvoet pairs, except for the merged data set, where all data (including Bijvoet pairs) were merged.

^c $R_{\text{merge}} = \sum_{hkl} \sum_{j=1, N} |I_{hkl} - I_{hkl,j}| / \sum_{hkl} \sum_{j=1, N} I_{hkl,j}$, where the outer sum (*hkl*) is taken over the unique reflections.

^d $R_{\text{ano}} = \sum_{hkl} |I_+ - I_-| / [(\frac{1}{2}) \sum_{hkl} |I_+ + I_-|]$.

^e $R_{\text{cullis}} = \langle \text{phase-integrated lack of closure} \rangle / \langle |F_{\text{PH}} - F_{\text{P}}| \rangle$.

^fPhasing power = $\langle |F_{\text{H}}(\text{calc})| / \text{phase-integrated lack of closure} \rangle$.

^g $R_{\text{f}} = \sum_{hkl} ||F_{\text{ohkl}} - k|F_{\text{c}hkl}| / \sum_{hkl} |F_{\text{ohkl}}|$, where $|F_{\text{ohkl}}|$ and $|F_{\text{c}hkl}|$ are the observed and calculated structure factor amplitudes, respectively.

^h R_{free} : same as R_{f} , for the test set of reflections (5% of the total) omitted from the refinement target.

Structural Genomics, unpublished results; A Murzin, personal communication), even though the architecture and disposition of secondary structure elements in the two proteins differ. Clearly, the architecture and topology of the SLDH subunit are organized in a different fashion (Figures 1C and D) than those of the LDH/MalDH family (Holbrook *et al.*, 1975; Rossmann *et al.*, 1975).

Structural resemblance of MjSLDH to the *Escherichia coli yiaK* gene product

Using the refined model of MjSLDH, a search of structural resemblance to other 3-D structures available in the PDB was

performed, using the secondary structure matching program SSM (<http://www.ebi.ac.uk/msd-srv/ssm>). The search produced a single match, to the 3-D structure of the homodimeric *E. coli yiaK* gene product (PDB id. 1NXU; Forouhar *et al.*, unpublished results), a putative gulonate oxidoreductase. The superposition of the 3-D structures of the compact dimer results in an r.m.s. deviation of 1.77 Å between 401 equivalent C α atoms (a similar value, 1.74 Å, is obtained when the superposition is carried out using only one monomer and 216 equivalent C α positions). The 3-D structure-based sequence alignment (Supplementary Figure 3S) shows a sequence identity of 27% for these two proteins. It therefore

appears, based on structural similarity alone, that the two enzymes should bind similar ligands and carry out similar enzymatic reactions.

Cofactor binding in *Mj*SLDH

The 3-D structure is that of a complex of the SLDH enzyme with its cofactor, NADH. In the structure, one NADH molecule is bound per monomer, thereby defining the approximate location of the active site. The cofactor binds in a cleft, delineated by secondary structure elements belonging to both monomers that make up the tight dimer (Figures 1C and 2, and Supplementary Figure 2S). Three helices (α D, α E, α J),

and the small helical turn α B', form one side of the pocket. The other side is made up by the central β -sheet of monomer A, and by helix α H and a short antiparallel β -sheet (β 6 and β 8) from monomer B. Each dimer therefore contains two identical active sites.

The adenine moiety of NADH binds in a pocket formed by residues from both monomer A (the external side of helix α J, strand β 10 and loop β 16- α J) and monomer B (helix α H and loop β 6- β 7). The adenine ring is confined in a hydrophobic pocket delineated by L301A, P303A, G304A and I306A on one side, and by F147B, P224B and Y227B on the other side (Figures 2A and C). The diol group of the ADP ribose interacts with the side chain of E307A, and with main-chain atoms of G304A and D173A. The diphosphate group is held by the side chain of K225B located in helix α H (this is the glycine-rich motif discussed later), and by the main chain of D173A and A175A from strand β 10. Contrary to the malate and lactate dehydrogenases (Holbrook *et al*, 1975; Rossmann *et al*, 1975; see also Baker *et al*, 1992a for an example of a B-side-specific Rossman-type dehydrogenase), the A side of the nicotinamide ring is stacked onto a small hydrophobic patch made up of A120A, P158A and A175A. The B side (that contains the pro-4S hydrogen) is thus exposed for catalysis. The nicotinamide O7 and N7 atoms interact with the side chain of T140A and with the main chain of G118A, respectively. The nicotinamide ribose diol oxygens are held by hydrogen bonds to the side chain of D173A and to the main chain of A120A. H44A is hydrogen bonded to the O5* atom of the nicotinamide ribose. Residues P158A, D173A, K225B, P303A and G304A, which are all involved in binding the NADH, are found in the same locations in the 3-D structure of the *E. coli yiaK* gene product. Residue E307A, which interacts with the diol group of the ADP ribose, is shifted by one residue in the sequence of the *E. coli* enzyme (Supplementary Figure 3S), but the location of its side chain in the *Yiak* protein indicates that it should perform the same function for NADH binding. Residues H44A, R48A, H116A and T156A, located in the vicinity of the nicotinamide ring (Figure 2), have equivalent locations in the two enzymes. In the *E. coli* enzyme, T140A is replaced by a serine residue,

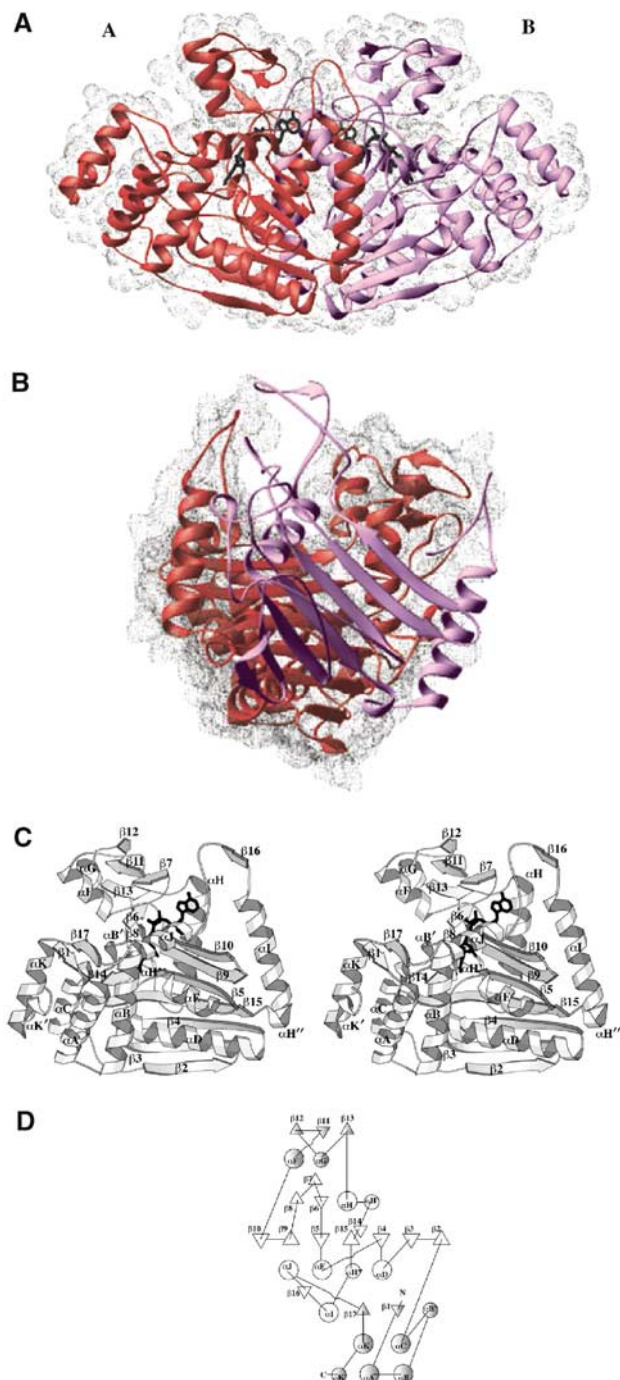


Figure 1 Tight dimer of *Mj*SLDH and topology of the monomer. (A) Ribbon representation of the tight dimer. The α -helices are shown in thick helical coil, the β -strands are represented by arrows and the coil regions by thin ribbon. The NADH cofactor is shown as black sticks. The monomers A and B are colored in red and mauve, respectively. The surface of the dimer is shown by a dotted Van der Waals surface calculated using all nonhydrogen atoms. (B) Ribbon representation of the secondary structure elements at the interface between two monomers within the tight dimer. The dotted Van der Waals surface is shown for monomer A only (depicted in red ribbon). The hydrophobic external surfaces of the central mixed β -sheets of each monomer face each other. This hydrophobic core is surrounded by the two α -helices α H and α I from monomers A and B. (C) Stereo view of the *Mj*SLDH monomer showing the secondary structure elements. The 11 α -helices are labelled α A- α K, the 17 β -strands are named β 1- β 17 and four α -helical turns are labelled α B', α H', α H'' and α K'. The central mixed β -sheet is made up of strands β 2, 3, 4, 15, 5, 9 and 10. There, only strands β 3 and 4 are parallel. The NADH cofactor is shown as dark sticks. (D) Projection topology cartoon of the *Mj*SLDH subunit. α -Helices are represented by circles and β -strands are shown as triangles. The lines connecting the secondary structure elements indicate the direction taken by these secondary structure elements. The two chain termini are indicated by N and C.

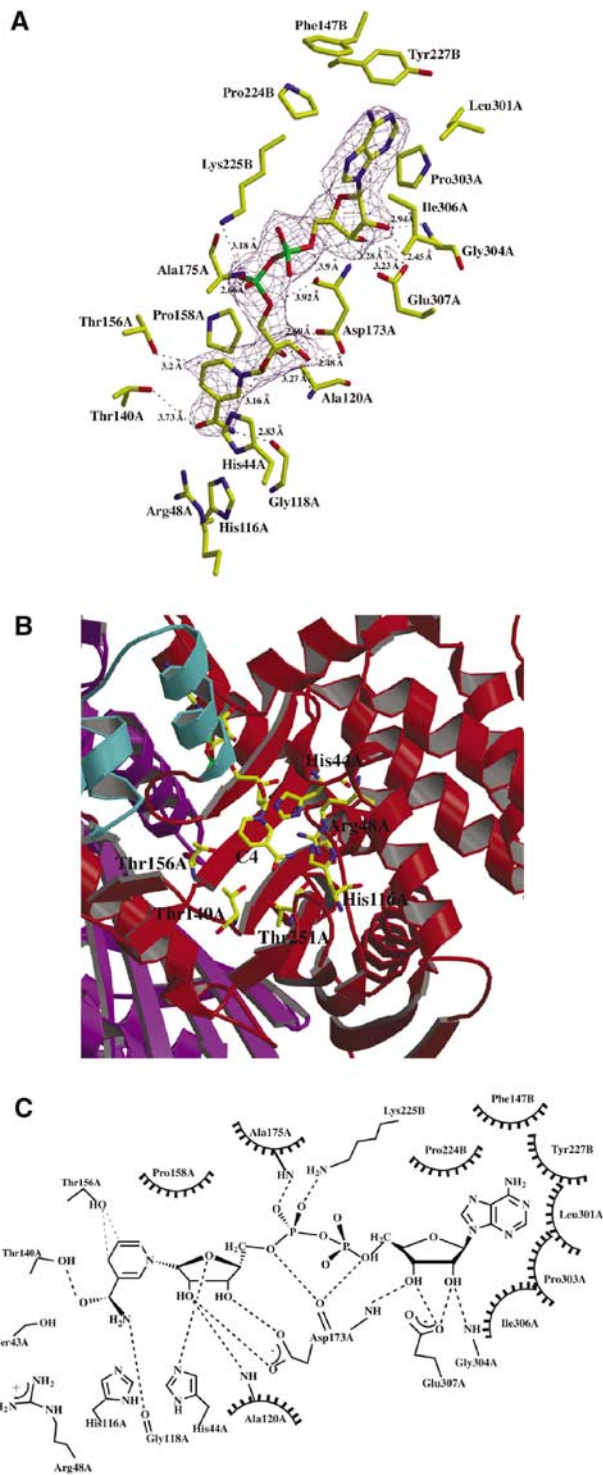


Figure 2 Active site of *MjSLDH*. (A) The NADH cofactor is shown in sticks inside the final omit map (magenta chicken wire) contoured at 1.0σ . The residues involved in cofactor binding, and the presumed residues of the catalytic site are labelled, with an indication of their subunit of origin. Broken lines represent the hydrogen bonds between cofactor atoms and those of the protein. A stereo version of this figure is available as Supplementary Figure 2S. (B) Close-up view of the active site in monomer A. Monomers A and B are colored in red and mauve, respectively, except for the mobile domain of monomer A, which is shown in cyan. The NADH and several potential active site residues are shown in stick representation. (C) Schematic representation of the active site. Hydrogen bonds are represented by thick dashed lines, while thin dashed lines show the interaction of T156A with the C4 and C5 atoms of the nicotinamide ring of NADH.

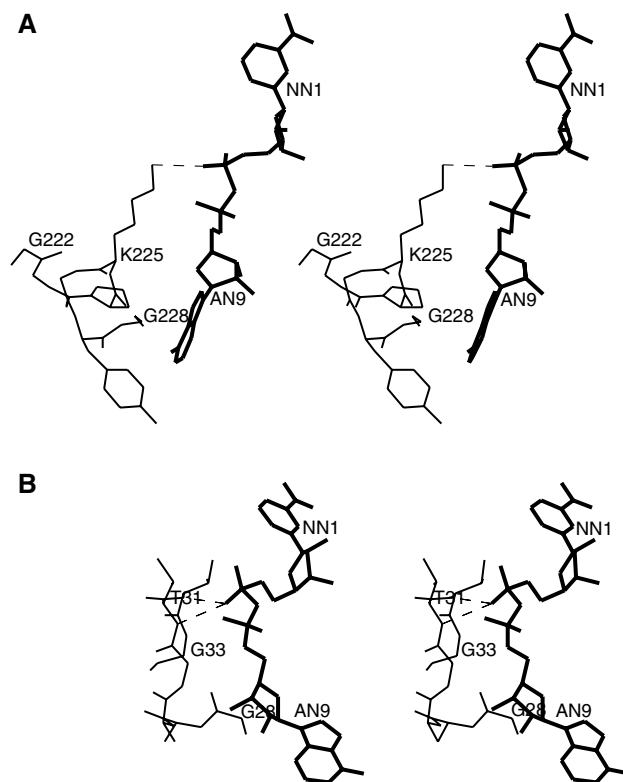


Figure 3 Comparison of the glycine-rich motif in the 3-D structure of *MjSLDH* to the 'classical' motif present in dehydrogenases with a Rossman-fold motif. Hydrogen bonds between cofactor and protein are shown by dashed lines. (A) Stereo view of NADH interacting at the level of the diphosphate moiety with residue K225 in the glycine-rich motif of *MjSLDH*. This glycine-rich sequence signature is located at the amino-terminal end of helix α H. (B) Stereo view of NAD interacting with the glycine-rich motif found in a representative member of Rossman-fold dehydrogenases, the (R207S, R292S) mutant of *HmMalDH* (PDB id. 1O6Z; Irimia *et al*, 2003).

which probably assumes the same function. Further, residues S43A and T138A, which are also located in the neighborhood of the nicotinamide ring, have identical locations in the two structures (not shown). The active site pocket is situated in the immediate vicinity of the nicotinamide ring of the cofactor. There, an asymmetric distribution of protein side chains is observed (Figure 2B): the side chains of the basic amino acids H44A, R48A and H116A are located close to the C2 atom of the nicotinamide ring, while the side chains of the polar residues T140A, T156A and T251A are located on the opposite side, next to the reactive C4 atom of the cofactor.

NADH binds in the cofactor-binding sites of *MjSLDH* in an extended conformation (Figures 2A and 3A), different from the 'boomerang' conformation observed in the LDH/MalDH enzymes (Figure 3B; Rossmann *et al*, 1975). As a consequence, the distance between the AN9 and NN1 atoms of the cofactor in *MjSLDH*, 19.3 \AA , is increased by more than 3 \AA compared to the corresponding distance measured in the cofactor of *Haloarcula marismortui* MalDH (16.1 \AA ; Richard *et al*, 2000; Irimia *et al*, 2003).

MjSLDH is a dimer in solution

The interdimer contacts observed in the octameric aggregate that make up the asymmetric unit of the *MjSLDH* crystals (Supplementary Figure 1S) involve predominantly charged

residues. In addition, the small surface contacts between the tight dimers that make up the octamer suggest the existence of lower oligomeric states of the enzyme: ca. 6.3% of the monomer surface takes part in the dimer-dimer interactions within the tetramers, and only 1% of the monomer surface links two tetramers to form the octameric particle.

We therefore carried out gel filtration experiments at room temperature in order to investigate the oligomeric state of *Mj*SLDH in solution. The elution volume in 0.1 M NaCl, 50 mM Tris-HCl buffer pH 7.6 suggested that *Mj*SLDH behaves as a dimer at 20°C. Efforts to detect higher oligomeric states by incubating the protein with 1 mM NADH, or with components of the crystallization buffer before gel filtration were unsuccessful.

We also performed velocity sedimentation analysis by analytical centrifugation at various protein concentrations (0.5, 1.0, 4.0 and 7.0 mg/ml) in 0.1 M NaCl, 50 mM Tris buffer pH 7.6. The expected values of $s_{20,w}$ and M_r are 11.8 S, 298 kDa and 4.6 S, 74.7 kDa for an octamer and a dimer, respectively. The velocity sedimentation data, over the protein concentration range tested, were fitted by a single species with an $s_{20,w}$ value of 4.54 S and a calculated M_r of 75 kDa. This indicates that *Mj*SLDH is a dimer in solution at the temperature of the experiment (20°C). Small-angle neutron scattering experiments, performed in the same buffer as above, at room temperature and at 56°C also showed the particle in solution to be a homodimer with a molecular mass of ca. 75 kDa and a radius of gyration of 25 Å.

Because of changes that occur in the drops during crystallization time, it is impossible to know the exact conditions when nucleation takes place. We estimate that, after 1 week, the concentrations of enzyme and polyethylene glycol 8000 (PEG 8000) are higher than 10 mg/ml and 10% (w/v), respectively. In order to mimic as closely as possible the initial crystallization conditions, the analytical ultracentrifugation experiments were repeated in buffer solutions containing 2.5, 5.0 and 9.0% PEG 8000 (the latter being half of the final PEG concentration in the crystallization droplets). Here again, we were not able to detect a shift in the oligomeric state of *Mj*SLDH toward either tetrameric or octameric species. However, the accretion of octameric particles that occurs during the formation of the crystals of *Mj*SLDH agrees with the current view that hyperthermostable proteins possess a high oligomerization state (Vieille and Zeikus, 2001).

Sequence alignment

The set of 32 aligned homologous sequences (Supplementary Figure 3S) shows 11 residues that are strictly conserved. The 3-D structure of the *Mj*SLDH enzyme indicates that four of these (T156, P158, D173 and K225) are probably involved in cofactor binding in all enzymes. Both H44 and H116, which are inside the catalytic cleft (Figures 2A and B), are probably involved in the enzymatic mechanism. The other residues (S43, G45, N157, G204 and G226) are all located in the immediate vicinity of the active site/cofactor-binding pocket and thus should play a role in the enzyme-assisted catalysis. Additional structure determination studies of complexes between *Mj*SLDH and potential substrate analogs will be required to reveal the precise role of these conserved residues in the enzyme's catalytic mechanism.

With the lack of a classical dinucleotide-binding domain in the 3-D structure of *Mj*SLDH, the multiple sequence align-

ment shown in Supplementary Figure 3S was used to locate a sequence signature analogous to that proposed by Wierenga *et al* (1985) and Baker *et al* (1992b): dehydrogenases with a Rossmann-fold topology contain a glycine-rich motif, which is involved in binding the diphosphate moiety of the dinucleotide cofactor (Figure 3B). In *Mj*SLDH, we have noted the presence of a conserved motif, GXXXGXG (GGPK₂₂₅GYG; Figure 3A), where the first glycine may be replaced by an alanine or by a serine, and the last glycine by an alanine. The lysine residue central to this motif is involved in binding the pyrophosphate moiety of NADH (Figures 2A and 3A), so that this sequence should be involved in binding of a dinucleotide cofactor in all the homologous enzymes listed in Supplementary Figure 3S.

Phylogenetic repartition, distribution and evolution of SLDH homologs

An unrooted neighbor-joining tree (Figure 4) was inferred from the structure-based amino-acid sequence alignment. With respect to available biochemical data and from the results of genomic analysis, one can suppose that a process of evolution (that includes a paralogous gene duplication event) has taken place within this family of homologous enzymes.

Six sequences (underlined in Figure 4 and Supplementary Figure 3S) are found in the archaeal domain. The sequences from the methanogenic Archaea (*M. kandleri*, *M. thermoautotrophicus*, *M. fervidus* and *M. jannaschii*) are all grouped together in the tree. This grouping indicates that the enzymes should act as true sulfolactate dehydrogenases, as is required by the specific metabolisms of the organisms. The two other archaeal sequences are from Pyrococcales lineages (*P. abyssi* and *P. horikoshii*, with NCBI Entrez accession codes

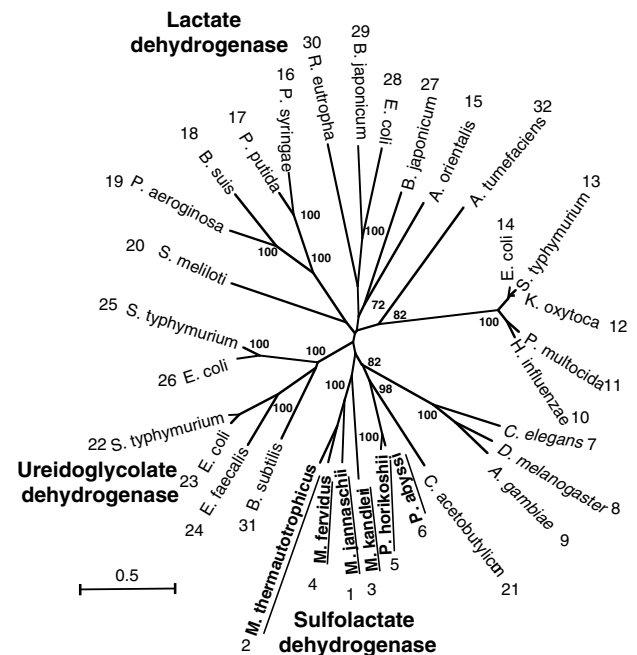


Figure 4 Unrooted neighbor-joining tree of sequences homologous to that of *Mj*SLDH. The scale bar represents 0.5 amino-acid substitution per site. Bootstrap values are indicated. Archaeal sequences are shown underlined in bold, eucaryotic sequences are in italics and those from bacteria are in nonbold roman. Numbering is according to the sequence numbers listed in Supplementary Figure 3S.

NP_126538 and NP_143166). The distribution of sequences homologous either to the [LDH-like] MalDHs (ORF 0490; Madern, 2000, 2002) or to SLDH (ORF 1425; Graupner *et al*, 2000) in various archaeal genomes shows that both genes are only found in the methanogenic Archaea (Supplementary Table IIS). In the Pyrococcales, the gene encoding an orthologous [LDH-like] MalDH is absent. However, MalDH activity has been measured in these cells (Selig *et al*, 1997). SLDH enzymes are known to use oxaloacetate as a substrate, with relatively high efficiency (Graupner *et al*, 2000). These observations strongly suggest that, in these organisms, the enzymes encoded by sequences NP_126538 and NP_143166 were selected during evolution to act as analogous MalDHs to compensate the lack of a specific orthologous [LDH-like] MalDH. This would therefore be a case of nonhomologous gene replacement. It should be interesting to find out if the two Pyrococcales enzymes still possess a broad range of substrate recognition (including that of oxaloacetate), or if they have evolved to use only oxaloacetate and discriminate against the other substrates.

Other homologous sequences are found in both bacterial and eucaryotic domains. The precise function of the corresponding enzymes remains to be elucidated, since it makes no sense that they should act as SLDHs in these organisms. A set of five sequences, numbered 10–14 in Figure 4 and Supplementary Figure 3S, includes the *E. coli yiaK* putative gulonate oxidoreductase enzyme (number 14, PDB id. 1NXU). Unfortunately, no biochemical information is available concerning these five enzymes.

Another set of six sequences (numbers 22–26 and 31) includes a ureido-glycolate dehydrogenase (number 23; Cusa *et al*, 1999), whose substrate is oxaluric acid. Oxaluric acid possesses two carboxyl groups, like oxaloacetate and sulfolactate (see the next section). Further, it should be noticed that various gene duplication events have taken place within bacteria. In particular, four homologous sequences were detected in *E. coli* (Figure 4). The *Ralstonia eutropha* enzyme (number 30, located within the large group of bacterial homologs) was characterized as a new type of LDH, unrelated to the 'classical' LDHs, due to its ability to oxidize pyruvate. *ReLDH* recognizes different oxoacids, and does not use oxaloacetate (Steinbüchel and Schlegel, 1983), contrary to *MjSLDH*, which uses oxaloacetate and does not recognize pyruvate. There are thus at least two cases of paralogous evolution that can be identified. These observations suggest the presence in the active sites of side chains responsible for substrate discrimination.

Substrate recognition and discrimination by *MjSLDH*

In order to gain insight into the substrate recognition mechanism within the SLDH family of enzymes, we tested if different substrates of various sizes and chemical structures can be accommodated by the enzyme (Table II). We also compared the results to those obtained with *ReLDH* (Steinbüchel and Schlegel, 1983). A previous study (Graupner *et al*, 2000) had shown that *MjSLDH* is able to use both 2-oxo pentanedioate and oxaloacetate as substrates, in addition to its preferred substrate sulfofpyruvate. We found that two shorter substrates, glyoxylate and oxo propanedioic acid, are also recognized by *MjSLDH*, albeit with lower efficiency. Further, we tested pyruvate, 2-oxobutyrate, 2-oxovalerate, ketoisovalerate, 2-oxohexanedioate, 2-ketohexa-

nedioate, 3-methyl 2-oxovalerate and 4-methyl 2-oxovalerate as potential substrates: none were recognized.

The data presented in Table II show that oxoacids containing up to five carbon atoms are recognized by *MjSLDH*. Above this size, recognition is abolished. The presence of two negatively charged groups appears essential to allow recognition of the substrate by the *MjSLDH* enzyme. Strikingly, the *ReLDH* enzyme shows the opposite trend: the presence of two carboxyl moieties abolishes substrate recognition. This analysis strongly suggests that substrate discrimination in *MjSLDH* occurs via the second negatively charged carboxylate of the substrate. The fact that sulfofpyruvate (which contains a single carboxylate group) is the preferred substrate of the enzyme also agrees with this hypothesis: in sulfofpyruvate, the second carboxylate is replaced by a sulfate group, that is, by an acidic moiety that possesses the same negative character as the carboxylic acid function. According to the structural information available for the catalytic cleft in this enzyme (Figure 2), a careful inspection of the sequence alignment between *MjSLDH* and *ReLDH* (Supplementary Figure 3S) identifies position 48 as the likely substrate-discriminating position: there, *MjSLDH* contains an arginine (whose side chain is held by a stacking interaction with the side chain of the conserved H116; Figure 2A), which may provide the compensatory positive charge required for the binding of a negatively charged substrate. At this location, an uncharged alanine side chain is found in *ReLDH*. Site-directed mutagenesis studies are underway to test the validity of this hypothesis.

Conclusion

The 3-D structure of *MjSLDH* indicates that this enzyme belongs to a new class of dinucleotide cofactor-dependent dehydrogenases that do not contain a Rossmann-fold motif. The ability of the enzyme to catalyze the interconversion of (S)-2-hydroxyacids into the corresponding 2-oxoacids is due to a novel mode of binding the dinucleotide cofactor: NADH binds to *MjSLDH* in an extended conformation, with its nicotinamide ring in the *anti* conformation. This different binding mode leaves the B side of the cofactor exposed to provide the pro-S hydrogen for catalysis. In addition, the biochemical and phylogenetic data strongly suggest that the introduction of a mutually exclusive substrate discrimination mechanism took place during a process of divergent evolution within the SLDH family of enzymes. This has led to the current presence of at least two classes of this enzyme, which recognize two types of charged substrates differing by the presence or absence of an additional negative charge. Further characterization is now required to assign the functional distribution of the various paralogous genes within this enzyme family.

Materials and methods

Protein expression, purification and crystallization

SeMet-labelled *MjSLDH* was expressed in *E. coli* BL21 (DE3), using the pT7-7 plasmid and the protocol described by Van Duyne *et al* (1993) for the production of SeMet protein. The expressing cells were lysed by sonication. The extract was heated for 25 min at 75°C and cleared by centrifugation. The supernatant was loaded onto a Q Sepharose Fast Flow chromatography column (Pharmacia) equilibrated with 50 mM NaCl in standard buffer (10 mM DTT, 50 mM Tris-HCl, pH 8.0). The column was washed with standard buffer and the protein was eluted with a gradient from 50 to 500 mM NaCl in the same buffer. The enzyme-containing fractions were

Table II Comparative substrate recognition by *Mj*SLDH and *Re*LDH

Name	Formuh	V_m U mg ⁻¹	K_m mM	V_m/K_m min ⁻¹ mg ⁻¹	<i>Mj</i> SLDH	<i>Re</i> LDH
1 Glyoxylic acid		14	46	0.3	+	
2 Oxo propanedioic acid		12	3.4	3.5	+	
3 Pyruvic acid					-	+
4 Sulfolactic acid		370	0.04	10000	+	
5 Oxaloacetic acid		47	0.13	356	+	-
6 2-oxo butyric acid					-	+
7 2-oxo pentanedioic acid		49	1.9	26	+	
8 2 oxo valeric acid					-	+
9 4-methyl 2 oxo valeric acid					-	+

Substrates with their chemical formulae are numbered 1–9. Substrate recognition is indicated by the symbol +, and by the value of V_{max}/K_m for the *Mj*SLDH enzyme (Graupner *et al*, 2000; this work). Absence of recognition is indicated by a hyphen. Recognition by the homologous enzyme from *Ralstonia eutropha* (*Re*LDH) is taken from the results of Steinbüchel and Schlegel (1983). The results are grouped according to the carbon content of each substrate.

selected by measuring enzyme activity (*vide infra*), and were revealed by SDS-PAGE with Coomassie blue staining. Active fractions were loaded onto a Red Sepharose CL-6B chromatography column (Pharmacia) equilibrated with standard buffer. After washing the column with this buffer, the protein was eluted with a gradient from 0 to 3 M NaCl in the standard buffer. The pure, active fractions were concentrated to a final concentration of 20 mg/ml. The buffer was changed prior to crystallization to 0.1 M NaCl, 5 mM DTT, 50 mM Tris-HCl pH 8.0. SeMet incorporation was assessed by electrospray mass spectrometry.

All crystallization was performed using sitting drops at room temperature. Initial crystallization conditions were obtained for the native protein using the Jena Bioscience crystallization screening kits (Jena, Germany). The additive screens from Hampton Research (Laguna Niguel, CA) were used to optimize crystal size and morphology. Equal volumes of protein (2 μ l) and of precipitant solution (2 μ l of 18% PEG 8000, 0.1 M Tris-HCl pH 7.6) were equilibrated against 1 ml of precipitant. Prior to crystallization, NADH was added to the protein solution, to reach a concentration of 2.5 mM in the final drops. The droplets were supplemented with 0.5 μ l of spermine tetra-HCl. Small plate crystals of SeMet-labelled *Mj*SLDH appear within 24 h, and reach final dimensions ca. 0.3 \times 0.15 \times 0.07 mm³ in 2 weeks.

X-ray diffraction data collection, structure determination and refinement

The 3-D structure of *Mj*SLDH was solved at 2.5 Å resolution by a combination of MAD phasing (Hendrickson and Ogata, 1997), molecular replacement (Rossmann, 1972) and density modification.

Diffraction data for the SeMet enzyme were collected at three wavelengths from a single crystal using a 165 mm MarCCD detector on the BM30A beam line at the ESRF. Prior to cryocooling at 100 K, the crystal was soaked for 20 min in mother liquor supplemented with 18% glycerol. The three wavelengths were selected using a fluorescence spectrum recorded from the same crystal. For the peak wavelength (0.979558 Å), 935 data frames of 0.5° rotation each were collected, and 720 data frames were recorded for each of the remote (0.977769 Å) and inflection wavelengths (0.979682 Å). The data were reduced using XDS (Kabsch, 1988). Autoindexing and systematic absences indicated space group P2₁ (Table I). The data were further processed using BIOMOL software (Groningen protein crystallography group, unpublished) and the CCP4 program TRUNCATE (French and Wilson, 1978).

The determination of initial selenium atom positions and their refinement was carried out with SnB (Weeks *et al*, 1993), MLPHARE (Otwinowski, 1991) and auto-SHARP (de la Fortelle and Bricogne, 1997). The final model for the heavy-atom substructure contains 69 out of the 72 selenium sites (with nine SeMet per subunit). The correct hand of the heavy-atom model was selected from automated runs of density modification by solvent flipping (Abrahams and Leslie, 1996) within auto-SHARP. Statistics of data processing and MAD phasing are reported in Table I.

The electron density map obtained at this stage was used to build a poly-Ala model corresponding to one *Mj*SLDH subunit (with SeMet residues included in their proper locations). A model for the tight dimer was generated by application of the two-fold non-crystallographic symmetry (NCS) operator, previously determined from the selenium positions. The complete model was then generated by molecular replacement with *AMoRe* (Navaza, 1994),

using the tight dimer as search probe and data between 3.5 and 15 Å. The initial model, comprising eight subunits, gave a correlation coefficient between observed and calculated structure factor amplitudes of 0.26.

Phase combination (using equally contributing heavy atom and model phases; Vellieux and Read, 1997) and electron density improvement by NCS averaging were performed with DEMON (Vellieux *et al*, 1995), first using four-fold NCS constraints, then with eight-fold improper NCS. During the procedure, the resolution was gradually increased from 3.58 Å to the high-resolution limit (2.5 Å). The resultant eight-fold averaged electron density map was used to rebuild the model for one subunit, including one NADH molecule. The asymmetric unit (eight subunits) was generated from this model by application of the NCS operators. The starting model had residual indices $R_f = 0.275$ and $R_{free} = 0.324$.

Model rebuilding was carried out into σ_A (Read, 1986) and OMIT electron density maps (Bhat, 1988; Vellieux and Dijkstra, 1997) using TURBO-FRODO (Roussel and Cambillau, 1991). The model was refined using maximum likelihood targets with CNS (Brünger *et al*, 1998), first using NCS constraints, then with NCS restraints (only over equivalent segments in each monomer), and finally removing all NCS restraints in the ultimate refinement rounds. Initially, simulated annealing refinement in cartesian space was used, with grouped temperature factor refinement. Toward the end of refinement, only energy minimization runs (including refinement of the individual isotropic temperature factors) were carried out. In the course of model building and refinement, well-defined water atoms were introduced. The refined model comprises 2752 amino-acid residues, eight NADH and 506 water atoms.

The quality and stereochemistry of the models was monitored using PROCHECK (Laskowski *et al*, 1993), standard procedures in CNS (Brünger *et al*, 1998) as well as the WHATCHECK validation server (Hoofst *et al*, 1996). The figures were prepared using Bobscript (Esnouf, 1997), ChemDraw (CambridgeSoft Corporation, Cambridge, MA), Molscrip (Kraulis, 1991), Raster3D (Merritt and Bacon, 1997), Ribbons (Carson, 1997) and TURBO-FRODO (Roussel and Cambillau, 1991).

Size-exclusion chromatography

In all, 15 μ l of purified protein (20 mg/ml) was applied to a Superose 12 HR10/30 column (Pharmacia) equilibrated with 0.1 M NaCl, 50 mM Tris buffer pH 7.6 (injection buffer). A flow rate of 0.2 ml/min was maintained throughout. The elution profile was monitored by measuring the absorbance at 280 nm, and fractions were tested for activity using oxaloacetate as substrate. The column was calibrated using ferritin, aldolase, LDH, MalDH and cytochrome c. To test the effect of the cofactor on the oligomeric state of MjSLDH, the protein was incubated with 1 mM NADH and was injected in the column, equilibrated with the injection buffer containing 1 mM NADH.

Analytical ultracentrifugation and small-angle neutron scattering experiments

The ultracentrifugation experiments were performed on a Beckman XLA analytical ultracentrifuge equipped with a UV-visible scanning system. A total of 200 sedimentation velocity profiles of MjSLDH were obtained every 5 min, at 20°C and 42 000 rpm, in various solvent conditions as described in Results. The data were corrected for density and viscosity, and analyzed using the software packages dcdt+ and Sedfit (Philo, 2000; Schuck, 2000). The calculated partial specific volume was $\bar{v}_2 = 0.75 \text{ ml/g}$ for MjSLDH.

Small-angle neutron scattering experiments were performed using ca. 6 mg/ml enzyme solutions in the standard buffer on the D22 camera at the Institut Laue Langevin. The data were analyzed in the Guinier approximation to yield the molecular mass and radius of gyration (Jacrot and Zaccai, 1981).

References

- Abrahams JP, Leslie AGW (1996) Methods used in the structure determination of bovine mitochondrial F_1 ATPase. *Acta Crystallogr D* **52**: 30–42
- Baker PJ, Britton KL, Engel PC, Farrants GW, Lilley KS, Rice DW, Stillman TJ (1992a) Subunit assembly and active site location in the structure of glutamate dehydrogenase. *Proteins* **12**: 75–86

Sequence alignment and phylogenetic analysis

We have used the amino-acid sequence of MjSLDH to identify homologous sequences in the SWISSPROT data bank by performing a BLAST search. The selected sequences were aligned using the CLUSTALW software package (Thompson *et al*, 1994) and manually adjusted using structural information. Long extensions were found at the N-termini of eucaryotic ORFs. These might be artifacts introduced by the automated sequence annotation. They were therefore trimmed out during manual alignment. The sequence alignment (without gaps) was subject to phylogenetic analyses, carried out using the phylogeny inference software package (PHYLIP) available online at <http://www.pasteur.fr>. The JTT option was used to generate distance matrices with PROTDIST. Phylogenetic trees were generated with BIONJ and drawn using NJPlot (Perrière and Gouy, 1996). The Bootstrap option was used to estimate the confidence limits of nodes.

Determination of enzymatic activities

The enzymatic activity of MjSLDH was measured in 1.0 ml of 0.1 M NaCl, 50 mM Tris-HCl, pH 8.0, supplemented with various putative substrates and with a saturating concentration of NADH, 0.3 mM. The oxidation of NADH was followed at 70°C by measuring the decrease in absorbance at 340 nm using a Beckman DU 7500 spectrophotometer. The data were corrected for NADH decrease without enzyme. For the determination of the K_m and V_{max} values for the different substrates, the putative substrate was assayed at various concentrations. The results were analyzed with the Lineweaver-Burk or Eadie-Hofstee representations, and the given K_m is the mean of the values obtained. One unit of enzyme activity refers to 1 μ mol of NADH oxidized per minute. If the time-dependent decrease in NADH absorbance, recorded for 5 min in the presence of the substrate to be tested, was found equal to the value obtained without substrate, then the chemical compound under investigation was considered not to be recognized by the enzyme.

Protein data bank accession number

The atomic coordinates and measured structure factor amplitudes for MjSLDH have been deposited in the Protein Data Bank with accession code 1RFM.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

Note added in proof

A manuscript reporting the 3-D structure of the E.Coli YiaK protein, by Foreuhar *et al*, has appeared recently in the Papers in Press section of the *Journal of Biological Chemistry* online, <http://www.jbc.org>

Acknowledgements

We thank Prof. R White (Virginia Polytechnic Institute) for having provided the gene encoding MjSLDH. We also thank Drs R Kahn and P Carpentier (IBS) for helpful discussions, the BM-30 and BM14 beam line staff at ESRF (Grenoble), and Prof. Rolf Hilgenfeld for access to the CDSB facilities at IMB-Jena (EC contract FMGE-CT98-0121). We thank Dr C Ebel for help with analytical ultracentrifugation, Dr R May for access to the D22 instrument at ILL, and Dr F Metz, Mr L Nauton, Mr D Lascoux (IBS) and Dr S Richard (the Salk Institute) for technical help. We are indebted to Dr A Murzin (MRC Cambridge) for pointing out the topological relationships between 3-D structures. AI holds a doctoral fellowship from the MRT. This work was funded in part by the GEOMEX Programme of the CNRS. Financial support of the CEA and the CNRS is acknowledged.

- Baker PJ, Britton KL, Rice DW, Rob A, Stillman TJ (1992b) Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold. *J Mol Biol* **228**: 662–671
- Bhat TN (1988) Calculation of an OMIT map. *J Appl Crystallogr* **21**: 279–281

- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**: 905–921
- Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, Fitzgerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**: 1058–1073
- Cambillau C, Claverie J-M (2000) Structural and genomics correlates of hyperthermostability. *J Biol Chem* **275**: 32383–32386
- Carson M (1997) Ribbons. In *Methods in Enzymology*, Carter CW, Sweet RM (eds) Vol. CCLXXVII, pp 493–505. San Diego, CA: Academic Press
- Cusa E, Obradors N, Baldomà L, Badia J, Aguilar J (1999) Genetic analysis of a chromosomal region containing genes required for assimilation of allantoin nitrogen and linked glyoxylate metabolism in *Escherichia coli*. *J Bacteriol* **181**: 7479–7484
- Esnouf RM (1997) An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J Mol Graph* **15**: 132–134
- de la Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multi-wavelength anomalous diffraction methods. In *Methods in Enzymology*, Carter CW, Sweet RM (eds) Vol. CCLXXVI, pp 472–494. San Diego, CA: Academic Press
- Forouhar F, Lee I, Benach J, Kulkarni K, Xiao R, Acton TB, Montelione GT, Tong L. A Novel NAD binding protein revealed by the Crystal structure of 2, 3-diketogulonate reductase (YiaK). JBC Papers in Press Published on January 12, 2004; doi:10.1074/jbc.M313580200
- French GS, Wilson KS (1978) On the treatment of negative intensity observations. *Acta Crystallogr A* **34**: 517–525
- Graupner M, White RH (2001) The first example of (S)-2-hydroxyacid dehydrogenases catalysing the transfer of the pro-4S hydrogen of NADH are found in the archaea. *Biochem Biophys Acta* **1548**: 169–173
- Graupner M, Xu H, White RH (2000) Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing reactions involved in coenzyme biosynthesis in methanoarchaea. *J Bacteriol* **182**: 3688–3692
- Hendrickson WA, Ogata CM (1997) Phase determination from multiwavelength anomalous diffraction measurements. In *Methods in Enzymology*, Carter CW, Sweet RM (eds) Vol. CCLXXVI, pp 494–523. San Diego, CA: Academic Press
- Holbrook JJ, Liljas A, Steindel SJ, Rossmann MG (1975) Lactate dehydrogenase. In *The Enzymes*, Boyer PD (ed) Vol. XI, pp 191–292. New York: Academic Press
- Honka E, Fabry S, Niermann T, Palm P, Hensel R (1990) Properties and primary structure of the L-malate dehydrogenase from the extremely thermophilic archaeobacterium *Methanothermus fervidus*. *Eur J Biochem* **188**: 623–632
- Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* **381**: 272
- Hwang KY, Chung JH, Kim SH, Han YS, Cho Y (1999) Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. *Nat Struct Biol* **6**: 691–696
- Irimia A, Ebel C, Madern D, Richard SB, Cosenza LW, Zaccai G, Vellieux FMD (2003) The oligomeric state of *Haloarcula marismortui* malate dehydrogenase are modulated by solvent components as shown by crystallographic and biochemical studies. *J Mol Biol* **326**: 859–873
- Jacrot B, Zaccai G (1981) Determination of molecular weight by neutron scattering. *Biopolymers* **20**: 2414–2426
- Kabsch W (1988) Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J Appl Crystallogr* **21**: 916–924
- Kasting JF, Siefert JL (2002) Life and the evolution of earth's atmosphere. *Science* **296**: 1066–1068
- Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of structures. *J Appl Crystallogr* **24**: 946–950
- Laskowski LA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**: 283–291
- Madern D (2000) The putative L-lactate dehydrogenase from *Methanococcus jannaschii* is an NADPH-dependent L-malate dehydrogenase. *Mol Microbiol* **37**: 1515–1520
- Madern D (2002) Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *J Mol Evol* **54**: 825–840
- Merritt EA, Bacon DJ (1997) Raster3D photorealistic molecular graphics. In *Methods in Enzymology*, Carter CW and Sweet RM (eds) Vol. CCLXXVII, pp 505–524. San Diego, CA: Academic Press
- Navaza J (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr A* **50**: 157–163
- Omichinski JG, Clore GM, Schaad O, Felsenfeld G, Trainor C, Appella E, Stahl SJ, Gronenborn AM (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **261**: 438–446
- Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997) CATH—a hierarchic classification of protein domain structures. *Structure* **5**: 1093–1108
- Otwinowski Z (1991) Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering, Proceedings of the CCP4 Study Weekend*, Wolf W, Evans PR, Leslie AGW (eds) pp 80–86. Warrington, UK: SERC Daresbury Laboratory
- Perrière G, Gouy M (1996) WWW-query: an on line retrieval system for biological sequences data banks. *Biochimie* **78**: 364–369
- Philo JS (2000) A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. *Anal Biochem* **279**: 151–163
- Rafferty JB, Sedelnikova SE, Hargreaves D, Artymiuk PJ, Baker PJ, Sharples GJ, Mahdi AA, Lloyd RG, Rice DW (1996) Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* **274**: 415–421
- Ramakrishnan C, Ramachandran GN, Sasikharan V (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys J* **5**: 909–933
- Read RJ (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr A* **42**: 140–149
- Richard SB, Madern M, Garcin E, Zaccai G (2000) Halophilic adaptation: novel solvent protein interactions observed in the 2.9 and 2.6 Å resolution structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui*. *Biochemistry* **39**: 992–1000
- Rossmann MG (1972) *The Molecular Replacement Method*. New York: Gordon and Breach
- Rossmann MG, Adams MJ, Buehner M, Ford GC, Hackert ML, Liljas A, Rao ST, Banaszak LJ, Hill E, Tsernoglou D, Webb L (1973) Molecular symmetry axes and subunit interfaces in certain dehydrogenases. *J Mol Biol* **76**: 533–537
- Rossmann MG, Liljas A, Brändén C-I, Banaszak LJ (1975) Evolution and structural relationships among dehydrogenases. In *The Enzymes*, Boyer PD (ed) Vol. XI, pp 61–102. New York: Academic Press
- Roussel L, Cambillau C (1991), Silicon Graphics Directory, Silicon Graphics, Mountain View, CA, USA
- Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**: 1606–1619
- Selig M, Xavier KB, Santos H, Schönheit P (1997) Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. *Arch Microbiol* **167**: 217–232
- Steinbüchel A, Schlegel HG (1983) NAD-linked L(+)-lactate dehydrogenase from the strict aerobic *Alcaligenes eutrophus*. 1. Purification and properties. *Eur J Biochem* **130**: 321–328
- Stetter KO (1999) Extremophiles and their adaptation to hot environments. *FEBS Lett* **452**: 22–25
- Thompson H, Tersteegen A, Thauer RK, Hedderich R (1998) Two malate dehydrogenases in *Methanobacterium thermoautotrophicum*. *Arch Microbiol* **170**: 38–42
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680
- Van Duyn GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229**: 105–124

- Vellieux FMDAP, Hunt JF, Roy S, Read RJ (1995) DEMON/ANGEL: a suite of programs to carry out density modification. *J Appl Crystallogr* **28**: 347–351
- Vellieux FMD, Dijkstra BW (1997) Computation of Bhat's OMIT maps with different coefficients. *J Appl Crystallogr* **30**: 396–399
- Vellieux FMD, Read RJ (1997) Noncrystallographic symmetry averaging in phase refinement and extension. In *Methods in Enzymology*, Carter CW, Sweet RM (eds) Vol. CCLXXVII, pp 18–53. San Diego, CA: Academic Press
- Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* **65**: 1–43
- Wah DA, Hirsch JA, Dorner LF, Schildkraut I, Aggarwal AK (1997) Structure of the multimodular endonuclease *FokI* bound to DNA. *Nature* **388**: 97–100
- Wang H, Boisvert D, Kim KK, Kim R, Kim S-H (2000) Crystal structure of a fibrillar homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *EMBO J* **19**: 317–323
- Weber PC, Salemme FR (1980) Structural and functional diversity in 4- α -helical proteins. *Nature* **287**: 82–84
- Weeks CM, DeTitta GT, Miller R, Hauptman HA (1993) Applications of the minimal principle to peptide structures. *Acta Crystallogr D* **49**: 179–181
- White RH (2001) Biosynthesis of the methanogenic cofactors. *Vitam Horm* **61**: 299–337
- Wierenga RK, De Maeyer MCH, Hol WGJ (1985) Interactions of pyrophosphate moieties with α -helices in dinucleotide binding proteins. *Biochemistry* **24**: 1346–1357