Acta Cryst. (1998). D**54**, 671–674

Characterization, crystallization and preliminary X-ray investigation of glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*

T. M. Fleming,[a] C. E. Jones,[b] P. W. Piper,[b] D. A. Cowan,[b] M. N. Isupov[a] and J. A. Littlechild[a]* at [a]*Departments of Chemistry and Biological Sciences, University of Exeter, Stocker Road, Exeter EX4 4QD, England, and* [b]*Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WCIE 6BT, England. E-mail: j.a.litlechild@exeter.ac.uk*

(*Received 25 June 1997; accepted 27 November 1997*)

Abstract

Recombinant *Sulfolobus solfataricus* glyceraldehyde-3-phosphate dehydrogenase has been purified and found to be a tetramer of 148 kDa. The enzyme shows dual cofactor specificity and uses $NADP^+$ in preference to NAD^+. The sequence has been compared with other GAPDH proteins including those from other archaeal sources. The purified protein has been crystallized from ammonium sulfate to produce crystals that diffract to 2.4 Å with a space group of $P4_32_12$ or $P4_12_12$. A native data set has been collected to 2.4 Å using synchrotron radiation and cryocooling.

1. Abbreviations

DTT, dithiothreitol; EDTA, ethylene diamine tetracetic acid; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; MIR, multiple isomorphous replacement; NAD^+, nicotinamide adenine dinucleotide; $NADP^+$, nicotinamide adenine dinucleotide phosphate; PIPES, piperazine -*N*,*N*′-bis [2-ethanesulfonic acid]; SDS–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis.

2. Introduction

Glyceraldehyde-3-phosphate dehydrogenase (E.C. 1.2.1.12; GAPDH) catalyzes the phosphate-dependent oxidative phosphorylation of D-glyceraldehyde-3-phosphate to form 1,3-diphosphoglycerate. This enzyme usually utilizes the cofactor NAD^+, although the GAPDH enzymes from the Archaea *Methanothermus fervidus* (Fabry & Hensel, 1987) and *Pyrococcus woesei* (Zwickl *et al.*, 1990) have been found to exhibit dual cofactor specificity, utilizing both NAD^+ and $NADP^+$. Another form of GAPDH (E.C. 1.2.1.13) is found in chloroplasts and has been shown to exhibit dual coenzyme specificity, preferentially using $NADP^+$ (Creff, 1978; Ferri *et al.*, 1978). In all reported cases to date, the GAPDH enzyme is tetrameric, consisting of four identical subunits with a total M_r of around 145 000.

We have previously reported the cloning of the *gap* gene, encoding GAPDH, from the hyperthermophilic archaeon *Sulfolobus solfataricus* and overexpressed the gene product to high levels in *E. coli* (Jones *et al.*, 1995). The *S. solfataricus* GAPDH enzyme was shown to be very stable at high temperatures, exhibiting an activity half life of 17 h at 353 K (Jones *et al.*, 1995). The crystal structures of GAPDH from a number of species have been reported including the moderate thermophile *Bacillus stearothermophilus* (Skarzynski *et al.*, 1987), the thermophile *Thermus aquaticus* (Tanner *et al.*, 1996), the hyperthermophile *Thermotoga maritima* (Korndorfer *et al.*, 1995), *Bacillus coagulans* (Griffith *et al.*, 1983), *Escherichia coli* (Duce *et al.*, 1996), human muscle (Mercer *et al.*, 1976; Vellieux *et al.*, 1993), lobster (Moras *et al.*, 1975) and *Trypanosoma brucei* (Vellieux *et al.*, 1993). Comparisons of the structures of the eubacteria thermophilic GAPDH enzymes from *T. aquaticus* (Tanner *et al.*, 1996) and *T. maritima* (Korndorfer *et al.*, 1995) with the corresponding mesophilic enzymes in order to understand the molecular mechanisms involved in protein thermostability have been reported. There is evidence to suggest that thermostability is achieved in different ways in different species and in the different evolutionary kingdoms (Fleming & Littlechild, 1997). It would, therefore, be of interest to see how thermostability is achieved in the thermophilic Archaea. Here we report the crystallization of the recombinant *S. sulfolobus* GAPDH and some preliminary X-ray data. This is the first archael GAPDH to be studied at a structural level.

3. Materials and methods

3.1. *Purification of recombinant S. solfataricus GAPDH*

S. solfataricus GAPDH was expressed and purified as described in Jones *et al.* (1995). The protein was purified using a two-step process of a heat-treatment step which inactivates the majority of *E. coli* proteins including *E. coli* GAPDH, followed by affinity chromatography using the dye ligand Reactive red 120. An additional gel-filtration step was included before crystallization or sedimentation analysis. The sample for analysis by sedimentation equilibrium was brought to 80% ammonium sulfate saturation and the precipitated protein harvested by centrifugation and 17 000*g* for 30 min at 277 K. The sample was resuspended in 50 m*M* potassium phosphate pH 7.0, 0.15 *M* KCl, 1 m*M* dithiothreitol (DTT) before applying to a HiLoad 16/60 Superdex 200 column (Pharmacia) and eluted in the same buffer. Fractions were assayed for GAPDH activity in the standard manner at 323 K as described previously (Jones *et al.*, 1995). The sample was concentrated to 2.2 mg ml^{-1} using a Centricon-10 microconcentrator (Amicon) using the conditions described by the manufacturer.

3.2. *N-terminal amino-acid sequencing*

Purified protein was diluted to 200 µg ml^{-1} and dialyzed against 5%(*v*/*v*) acetic acid and the N-terminal amino-acid sequence of the first 15 residues was obtained by gas-phase sequencing at Aberdeen University.

3.3. *Mass spectrophotometry*

Purified GAPDH was concentrated to 100 µg ml^{-1} and dialyzed against 10 m*M* Tris–HCl pH 7.5, 1 m*M* DTT. Analysis

by laser desorption mass spectrophotometry was carried out at Aberdeen University.

3.4. *Sedimentation equilibrium analysis*

Sedimentation equilibrium analysis of purified GAPDH was carried out by the National Centre for Macromolecular Hydrodynamics at the University of Leicester. The GAPDH sample was monitored by sedimentation equilibrium at concentrations of 0.2, 0.4, 0.6 and 0.8 mg ml^{-1}.

3.5. *Crystallization*

The GAPDH protein was crystallized by the hanging-drop vapour-diffusion technique. The protein concentration was 10 mg ml^{-1} in the initial droplet in the presence of 10 m*M* PIPES pH 6.5, 5 m*M* EDTA, 10 m*M* NAD^+. The precipitant was 35% $(NH_4)_2SO_4$ and crystallization was at 290 K. The crystals were harvested into mother liquor containing 45% saturated ammonium sulfate, 10 m*M* PIPES pH 6.5, 5 m*M* EDTA, 10 m*M* NAD^+. Data were collected at the DORIS storage ring DESY, Hamburg Synchrotron source using a MAR image plate on the EMBL wiggler line BW7B. Data were collected at a wavelength of 0.88 Å using a crystal frozen at 100 K.

3.6. *Preliminary crystallographic analysis*

The data was processed using the *DENZO* and *SCALEPACK* software package (Otwinowski & Minor, 1995). The rotation- and translation-function calculations were performed using *AMoRe* (Navaza, 1994), implemented in the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994).

4. Results and discussion

4.1. *Enzyme characterization*

GAPDH from *S. solfataricus* was overexpressed in *E. coli*. The recombinant protein can be purified by a two-step process. This yielded a pure protein of M_r of 39 500 as judged by SDS–PAGE. The predicted molecular weight of the recombinant enzyme is 37 581. By laser desorption mass spectrophotometry the M_r was seen to be 37 611 ± 0.1%. Sedimentation analysis gave a molecular weight of 148 330 Da ± 830.8 indicating a tetramer. The recombinant GAPDH protein also behaved as a tetramer by gel filtration on a Superose 12 (35/32) column. This is in common with other GAPDH's from other species. The N-terminal amino-acid sequence of the first 15 residues of the recombinant GAPDH matched the deduced amino-acid sequence from the gene construct.

4.2. *Nucleotide binding*

We have previously shown that *S. solfataricus* GAPDH can utilize both NAD^+ (K_m = 2.2 m*M*) and $NADP^+$ (K_m = 0.067 m*M*) as cofactors but it exhibits a marked preference for $NADP^+$ (Jones *et al.*, 1995). Other archaeal GAPDH enzymes have also been shown to exhibit this dual cofactor specificity, with a marked preference for $NADP^+$ (Fabry & Hensel, 1987; Zwickl *et al.*, 1990).

Previous studies have established that in many NAD^+/$NADP^+$ binding proteins the nucleotide-binding domain exists as a $\beta\alpha\beta$ fold centred around a highly conserved consensus sequence (Wierenga *et al.*, 1985). NAD^+ binding proteins have been found to contain a G*X*G*XX*G sequence motif with a negatively charged residue at the end of the second β-strand. In contrast $NADP^+$ binding proteins have a G*X*G*XX*A sequence motif with a positively charged residue at the end of the second β-strand. *S. solfataricus* GAPDH possesses the NAD^+ characteristic motif, G*X*G*XX*G, close to the N-terminus (Fig. 1) even though it preferentially utilizes $NADP^+$. However, there is a basic Lys residue at the end of the second β-strand, as found in other $NADP^+$ binding enzymes. Detailed analysis of the different cofactor binding sites will be carried out when the structure of the *S. solfataricus* enzyme is solved.

4.3. *Sequence analysis*

Over 40 primary amino-acid sequences are now known for GAPDH enzymes from all three kingdoms. The sequences of five archaeal GAPDH's have been reported, including the mesophilic enzymes of *Methanobacterium bryantii* and *Methanobacterium formicicum* (Fabry *et al.*, 1989) and the thermophilic GAPDH's of *P. woesei* (Zwickl *et al.*, 1990), *M. fervidus* (Fabry & Hensel, 1988) and *S. solfataricus* (Jones *et al.*, 1995). The GAPDH sequence alignment shows high conservation throughout the Archaea. Sequence alignments using the GCG program *GAP* (Devereux *et al.*, 1984). indicated that the *S. solfataricus* GAPDH is 49% identical to the GAPDH's of *M. formicicum, M. bryantii, M. fervidus* and *P. woesei*. This is in contrast to the low similarity found between the sequence of *S. solfataricus* GAPDH and those of the thermophilic bacteria *B. stearothermophilus*, *T. maritima* and *T. aquaticus* which were found to be 16, 18 and 15%, respectively. An active-site Cys residue, thought to be involved in forming a covalent phosphoglycerol thioester intermediate is also conserved throughout all of the GAPDH sequences.

4.4. *Crystallization, preliminary X-ray characterization and molecular-replacement studies*

Crystals of GAPDH grow reproducibly in 10–14 d at 290 K from 35% ammonium sulfate at 10 mg ml^{-1} protein. The

```
                    //////  +++++++++            ////    +++      residue no.
                                                  *
Yeast                MVRVAINGFGRIGRLVMRIAL--SRPNVEVVALNDPFITND      38
B. stearo           A-VKVGINGFGRIGRNVFRAAL--KNPDIEVVAVNDL-TDAN      38
T. maritima         A--RVAINGFGRIGRLVYRIIYERKNPDIEVVAINDL-TDTK      39
T. aquat            --MKVGINGFGRIGRQVFRILHSR---GVEVALINDL-TDNK      36
P. woesei           MKIKVGINGYGTIGKRVAYAVT--KQDDMELIGVTKT-KPDF      39
S. solf             V-INVAVNGYGTIGKRVADAII--KQPDMKLVGVAKT-SPNY      38
M. fervidus         MKA-VAINGYGTVGKRVADAIA--QQDDMKVIGVSKT-RPDF      38
```

Fig. 1. Alignment of GAPDH sequences in the region of the NAD(P). binding motif. * indicates the charged residue at the end of the second β-strand. The β-strands and α-helix which are taken from the *B. stearothermophilus* GAPDH crystal structure (Skarzynski *et al.*, 1987) are shown as / and +, respectively, above the sequence. The G*X*G*XX*G motif characteristics of NAD^+ binding proteins is indicated in bold.

Table 1. *Statistics of the X-ray data*

Resolution ranges (Å)	$\langle I \rangle$	$\langle I/\sigma(I) \rangle$	Redundancy†	$R_{\rm merge}$‡	Completeness (%)
15.0–5.11	18383	21.5	14.2	0.051	92.9
5.11–4.08	20109	20.1	15.3	0.061	94.8
4.08–3.57	12852	13.2	5.2	0.062	94.8
3.57–3.25	8169	11.2	5.0	0.073	95.3
3.25–3.02	4837	8.3	5.1	0.110	95.9
3.02–2.84	3014	5.2	5.1	0.142	96.2
2.84–2.70	2080	3.8	5.1	0.221	96.4
2.70–2.58	1661	2.9	5.0	0.272	96.6
2.58–2.49	1326	2.5	5.0	0.315	97.1
2.49–2.40	1099	2.0	4.9	0.394	97.4
Total	7419	8.5	7.0	0.073	95.7

† The average number of observations of the same reflection. ‡ The value of the merging R factor between equivalent measurements of the same reflection $R_I = \sum |I - \langle I \rangle| / \sum I$.

crystals appear bi-pyrimidial and are 0.35 mm in their largest dimension (Fig. 2), are stable to X-ray radiation and diffract to 2.4 Å resolution at room temperature.

A native data set was collected as previously described to 2.4 Å resolution from a crystal frozen at 100 K. 253 416 individual measurements were reduced to 35 828 unique reflections with an overall $R_{\rm merge}$ of 0.073 (Table 1). The data were 95.7% complete. The space group was determined as $P4_12_12$ or its enantiomorph $P4_32_12$ with cell dimensions $a = b = 101.57$, $c = 179.81$ Å. The asymmetric unit contains two subunits with molecular weight of 37 611 Da. This results in a specific volume of 3.07 Å^3 Da^{-1} and a solvent content of 59.7% (Matthews, 1968), assuming protein density 1.34 g cm^{-3}. The asymmetric unit cannot contain a whole tetrameric molecule of GAPDH, as it would result in a solvent content of 20% which is impossible for protein crystals. Since in the space group $P4_12_12$ ($P4_32_12$). there are no intercepting crystallographic twofold axes, with which molecular dyads of the tetramer could coincide the asymmetric unit cannot contain one subunit.

Molecular-replacement studies using *B. stearothermophilus* GAPDH structure (Skarzynski *et al.*, 1987; PDB code 1GD1; Bernstein *et al*., 1977) as a starting model were conducted using *AMoRe* (Navaza, 1994). The sequence identity between the model and the structure is only about 16%. However, the cross-rotation function solutions were quite clear when the integration radius was increased to 40 Å. Both the full protein model and the polyalanine model with trimmed surface loops were used. The best results were obtained with the trimmed polyalanine model. The cross-rotation function for a tetrameric model calculated in 10–5 Å resolution range gave one clear solution with height of 7.5σ with no other peaks over 3.75σ. When the model tetramer is rotated by this solution, its molecular dyad P is parallel to crystallographic dyad, which is consistent with proposed dimer in the asymmetric unit. The cross-rotation function for such a dimer, formed by subunits O and Q calculated in similar conditions gave the same solution as for the tetramer with a height 6.3σ with noise peaks not higher than 4.2σ. The cross-rotation function for one subunit gave peaks consistent with a solution for the tetramer at 5.3σ and 3.5σ with noise peak at 3.9σ.

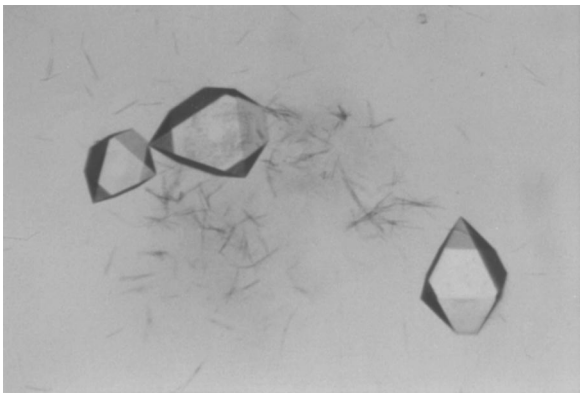

Fig. 2. Crystals of recombinant *S. solfataricus* GAPDH. The crystals are 0.35 mm in their largest dimension and grow in 10–14 d at 290 K from 35% $(NH_4)_2SO_2$. The initial droplet contained 10 m*M* PIPES pH 6.5, 5 m*M* EDTA, 10 m*M* NAD$^+$.

As one of the molecular dyads of the tetramer coincides with crystallographic dyad the translation search is a one-dimensional task. A translation search with both F's and E's was conducted. So far the best translation-function solution has been obtained for a trimmed tetramer model at 10–5.5 Å resolution using F's. The best solution was in space group $P4_12_12$ with a correlation $C = \sum(|F_o| - \langle|F_o|\rangle)(|F_c| - \langle|F_c|\rangle) / [\sum(|F_o| - \langle|F_o|\rangle)^2 \sum(|F_c| - \langle|F_c|\rangle)^2]^{1/2}$ of 14.7% and an R factor $R = \sum \big||F_o| - |F_c|\big| / \sum |F_o|$ of 52.8%. The next solution in the same space group had a correlation of 12.5% and an R factor of 53.5%, while the best solution in the enantiomorphic space group $P4_32_12$ has had a correlation of 12.7% and a R factor of 53.8%. The above solution has not been confirmed by the results of the translation search for a dimeric or a monomeric model. All attempts to refine the molecular replacement model with either *PROLSQ* (Hendrickson & Konnert, 1980) or *REFMAC* (Murshudov *et al*., 1997) have failed.

The structure of this GAPDH (to be published elsewhere) has been solved recently by multiple isomorphous replacement (MIR) and non-crystallographic symmetry averaging. It has shown the correctness of the space-group enantiomorph and molecular-replacement solution, which was used as a reference in the course of model building into the averaged MIR map. The GAPDH's from *S. solfataricus* and *B. stearothermophilus* superimpose with an r.m.s. fit of 2.04 Å for 224 matching Cα atoms out of 340 in the *S. solfataricus* sequence, while the rest of the atoms belong to the loop regions which are a different length or to additional secondary-structure elements. It would appear that the molecular-replacement solution discussed above could not be refined since the starting model was outside the convergence limits of existing refinement programs.

Primary sequence alignment (GCG, 1994). using prokaryotic and eukaryotic *versus* archaeal GAPDH's have failed to give a result that is consistent with the three-dimensional structure of archaeal GAPDH. This is not surprising at a level of sequence identity below 20% and would also have contributed to the failure of refinement of the molecular-replacement solution. It seems that molecular replacement will sometimes produce a successful result even at low level of sequence identity between proteins. However, its success will depend on the development of improved refinement programs.

The detailed structural comparison of *S. solfataricus* GAPDH with its eubacterial and euckaryotic counterparts is in progress and will hopefully shed some light onto the dramatically increased thermal stability of the enzyme.

We thank Dr Alexei Teplyakov for the assistance with the data collection. We also thank the European Union for support of the work at EMBL Hamburg through the HCMP to Large Installations Project, contract No. CHGE-CT93–0040. This work was supported by grants (to TMF, CEJ and MNI). from the Chemical and Pharmaceutical Directorate of the BBSRC.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. J., Brice, M. D., Rogers, J. K., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol*. **112**, 535–542.

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

Creff, R. (1978). *Phytochemistry,* **17**, 2061–2067.

Devereux, J., Haerbeli, P. & Smithies, O. (1984). *Nucleic Acids Res*. **12**, 387–395.

Duce, E., Olivierdeyris, L., Fanchon, E., Corbier, C., Bralant, G. & Dideberg, O. (1996). *J. Mol. Biol*. **257**, 814–838.

Fabry, S. & Hensel, R. (1987). *Eur. J. Biochem*. **165**, 147–155.

Fabry, S. & Hensel, R. (1988). *Gene,* **64**, 189–197.

Fabry, S., Lang, J., Niermann, T., Vingran, M. & Hensel, R. (1989). *Eur. J. Biochem*. **179**, 405–413.

Ferri, G., Comerio, G., Iandarola, P., Zapponi, M. C. & Speranza, M.L. (1978). *Biochim. Biophys. Acta,* **522**, 19–31.

Fleming, T. & Littlechild, J. (1997). *Comp. Biochem. Phys. A*, **118**, 439–451.

GCG (1994). *Program manual for the Wisconsin package, version 8.* Genetics Computer Group, Madison, Wisconsin, USA.

Griffith, J. B., Lee, B., Murdock, A. L. & Amelunxen, R. E. (1983). *J. Mol. Biol*. **169**, 963–974.

Hendrickson, W. A. & Konnert, J. H. (1980). In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

Jones, C. E., Fleming, T. M., Cowan, D. A., Littlechild, J. A. & Piper, P. W. (1995*). Eur. J. Biochem*. **233**, 800–808.

Korndorfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995). *J. Mol. Biol*. **246**, 511–521.

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

Mercer, W. D., Winn, S. I. & Watson, H. C. (1976). *J. Mol. Biol*. **104**, 277–283.

Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. & Rossmann, M. G. (1975). *J. Biol. Chem*. **250**, 9137–9162.

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

Navaza, J. (1994). *Acta Cryst*. A**50**, 157–163.

Otwinowski, Z. & Minor, W. (1995). *The HKL manual. A description of the programs DENZO, XDisplayF and SCALEPACK.* Yale University, New Haven, Connecticut, USA.

Skarzynski, T., Moody, P. C. E. & Wonacott, J. A. (1987). *J. Mol. Biol*. **193**, 171–187.

Tanner, J. J., Hecht, R. M. & Krause, K. L. (1996). *Biochemistry*, **35**, 2597–2609.

Vellieux, F. M. D., Hadju, J., Verlinde, C. L. M. J., Groendijk, H., Read, R. J. Greenhough, T. J., Campbell, J. W., Kalk, K. H., Littlechild, J. A., Watson, H. C. & Hol, W. G. J. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 2355–2359.

Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. (1985). *Biochemistry*, **24**, 1346–1357.

Zwickl, P., Fabry, S., Bogedain, C., Haas, A. & Hensel, R. (1990). *J. Bacteriol*. **172**, 4329–4338.