Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Cloning, purification, crystallization and preliminary crystallographic analysis of galactokinase from *Pyrococcus furiosus*

Daniel de Geus,[a] Andrew P. Hartley,[b] Svetlana E. Sedelnikova,[b] Steven E. Glynn,[b] Patrick J. Baker,[b] Corné H. Verhees,[a] John van der Oost[a] and David W. Rice[b]*

[a]Laboratory of Microbiology, Wageningen University, The Netherlands, and [b]Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, England

Correspondence e-mail: d.rice@sheffield.ac.uk

Received 7 May 2003
Accepted 21 July 2003

Galactokinase catalyses the conversion of galactose to galactose-1-phosphate as the first step in the Leloir pathway, a metabolic route that eventually enables the degradation of galactose *via* the glycolytic pathway. Galactokinases have been isolated from a wide range of prokaryotic and eukaryotic organisms and the enzyme has been identified as a member of the GHMP kinase (galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase) superfamily. *Pyrococcus furiosus* galactokinase was cloned, expressed in *Escherichia coli*, purified and crystallized using the hanging-drop method of vapour diffusion with ammonium sulfate as the precipitant. The crystals belong to the space group $C222_1$, with more than eight subunits in the asymmetric unit and with approximate unit-cell parameters $a = 211.7$, $b = 355.4$, $c = 165.5$ Å, $\alpha = \beta = \gamma = 90^\circ$. The crystals diffract X-rays to 2.9 Å resolution on a synchrotron-radiation source. Determination of the structure will provide insights into the molecular basis of substrate recognition and catalysis of this enzyme, for which no structures are currently available.

1. Introduction

Pyrococcus furiosus is a hyperthermophilic archaeon isolated from thermally active regions. Enzymes purified from this species generally show extreme thermal stability, with half-lives at 373 K that are often in excess of 24 h. Such proteins provide an ideal opportunity to study the intramolecular interactions involved in maintaining structure and function under such extreme conditions.

A main catabolic route in saccharolytic organisms such as *P. furiosus* is glycolysis, in which glucose is converted to pyruvate. Conversion of galactose *via* glycolysis requires an additional metabolic branch, the Leloir pathway (Frey, 1996). The first step of this pathway involves the conversion of galactose to galactose-1-phosphate. The latter reaction is catalysed by galactokinase (GalK), which drives the transfer of the γ-phosphate group from ATP to the C-1 position of galactose. The *galK* gene in *P. furiosus* encodes a protein of 39.4 kDa molecular weight and has been shown to have an unusually high affinity for ATP (Verhees *et al.*, 2002).

Sequence similarities have suggested galactokinase to be a member of the GHMP kinase superfamily (Bork *et al.*, 1993). Other members of the family for which structures have recently become available include homoserine kinase (Zhou *et al.*, 2000), mevalonate kinase (Yang *et al.*, 2002) and mevalonate decarboxylase (Bonanno *et al.*, 2001). Although the overall sequence identity in the family is low (Krishna *et al.*, 2001), three common glycine-rich motifs characterize the family and these can be clearly identified in *P. furiosus* galactokinase. The three-dimensional structure of homoserine kinase was the first to be solved for any member of this family and reveals a novel nucleotide-binding fold and a rare *syn* conformation of the glycosidic bond of the ADP in the active site (Krishna *et al.*, 2001). Analysis of this structure has shown that the most conserved glycine-rich motif across the GHMP kinase family, with a consensus of PX_3GL-(G/S)SSA, forms a novel phosphate-binding loop which interacts through hydrogen bonds to the α- and β-phosphates of ADP. Comparison with other structures of GHMP kinase family members, those of mevalonate kinase (Fu *et al.*, 2002) and mevalonate decarboxylase (Bonanno *et al.*, 2001), show that many structural elements are conserved between GHMP kinase family members. One difference between the mevalonate kinase structure and that of the homoserine kinase is the adoption of an *anti* conformation of the glycosidic bond in the nucleotide.

There is a very high (72%) sequence identity between yeast GalK and yeast Gal3p, an inducer protein involved in transcriptional activation and repression of the *Saccharomyces cerevisiae* GAL operon (Platt & Reece, 1998). Activation is dependent on ATP and galactose and occurs as part of a tripartite protein complex. However, despite this simi-

larity, Gal3p does not show galactokinase activity and *S. cerevisiae* contains a separate galactokinase enzyme (Platt & Reece, 1998). Determination of the structure of GalK may offer insights into the mechanism of galactokinase and the interactions involved in the homologous yeast activation complex. This paper describes the cloning, purification, crystallization and preliminary X-ray analysis of the GalK protein.

2. Materials and methods

The GalK gene was PCR amplified and cloned into pET9d plasmid (Novagen), creating the vector pLUW570 (Verhees *et al.*, 2002) which was transformed into the overexpression strain *Escherichia coli* BL21 (DE3). To produce selenomethionine-labelled protein for crystallization trials and subsequent structural analysis, a 250 ml flask with 50 ml of LB medium containing 50 µg ml^{-1} kanamycin was used as the seed batch culture and grown overnight, inoculated from a single colony of *E. coli* BL21 (DE3)/pLUW570.

Subsequently, 4 ml of this seed culture was pelleted and the cells were resuspended in 4 ml of minimal medium [M9 salts, 0.1 *M* $MgSO_4$, 20%(*w*/*v*) glucose, 1 µ*M* thiamine, 4.2 mg l^{-1} $FeSO_4$] at 310 K and 1 ml of this resuspension was used to inoculate each of four 2 l baffled flasks containing 250 ml of the same medium. The cultures were then grown on a shaking tray at 310 K until the absorbance at 600 nm reached 0.3. At this point in growth, the amino acids lysine, phenylalanine and threonine (100 mg l^{-1}) and valine, isoleucine and leucine (50 mg l^{-1}) were added in order to suppress methionine biosynthesis (Duyne *et al.*, 1993) and in addition, selenomethionine (40 mg l^{-1}) was also added to allow incorporation of selenomethionine into the protein. The culture was left to grow overnight (approximately 16 h) without the addition of IPTG because the vector used showed sufficient overexpression of *galK* in *E. coli* without IPTG induction. To harvest the cells, the cultures were centrifuged at 5000*g* for 20 min at 277 K and the pellets were frozen immediately for storage prior to purification.

To purify SeMet-incorporated GalK, approximately 5 g of cell paste was suspended in buffer *A* (40 m*M* Tris–HCl pH 8.0, 2 m*M* EDTA) and disrupted by sonication. The homogenate was made up to 0.4 *M* NaCl and heated to 348 K for 15 min. Debris was removed by centrifugation at 45 000*g* for 20 min. The supernatant fraction was made up to 1.5 *M* $(NH_4)_2SO_4$ and loaded onto a butyl Toyopearl column (Toyo Soda) and washed with 1.2 *M* $(NH_4)_2SO_4$ in buffer *A*. The proteins were eluted with a gradient of 1.2–0 *M* $(NH_4)_2SO_4$ in buffer *A*. Fractions containing GalK were combined and concentrated on a Vivaspin concentrator with molecular-weight cutoff of 10 kDa (Vivascience) to a volume of 1.5 ml and loaded onto a Hi-Load Superdex-200 column (Pharmacia) equilibrated with 0.1 *M* NaCl in buffer *A*; this solution was run through the column until the protein was eluted.

The protein was concentrated to 13 mg ml^{-1} using a Vivaspin concentrator and crystallization attempts were made using Hampton Crystal Screens I and II and PEG/Ion screen (Hampton Research). Trials were performed in both the presence and absence of a range of substrates using the standard hanging-drop vapour-diffusion technique by mixing 2 µl of the protein solution with an equal volume of precipitant and equilibrating against the same precipitant at 290 K.

Data collection at cryogenic temperatures utilized crystals soaked in a solution containing 2.5 *M* ammonium sulfate, 0.1 *M* Tris–HCl and 20% glycerol and flash-frozen in a stream of nitrogen gas at 100 K using an Oxford Cryosystems Cryostream device. Multiple wavelength anomalous dispersive (MAD) data were collected by the rotation method for 120 frames with 0.75° rotation per frame using an ADSC Quantum 4 CCD detector on beamline ID14.4 at the ESRF, Grenoble, France. Three wavelengths were chosen near the selenium-absorption edge based on an X-ray fluorescence scan of the frozen crystal in order to maximize the f'' component (λ_1, peak), to minimize the f' component (λ_2, inflection) and to maximize $\Delta f'$ (λ_3, remote).

3. Results

SDS–PAGE analysis of the soluble fraction of the BL21 (DE3) cells containing the overexpressed GalK showed a large band corresponding to a protein of approximately 40 kDa molecular weight. This high level of expression facilitated purification by the simple three-step protocol described above. GalK elutes from the gel-filtration column with an approximate weight of 40 kDa consistent with a monomeric quaternary structure. The purity of the protein estimated by SDS–PAGE (4–12% bis-tris NuPAGE gel, Novex) was 90–95%, with a typical yield of about 8 mg of protein from 5 g of cell paste as estimated by the method of Bradford (1976) using the Bio-Rad Protein Assay reagent.

Crystallization trials of the selenomethionine-incorporated protein indicated that the presence of galactose, ADP and Mg^{2+} was essential for crystallization. Conditions were optimized following the initial trials and rectangular plate-shaped crystals up to 0.4 × 0.6 × 0.1 mm in size appeared within several days using 2.1–2.3 *M* ammonium sulfate and 0.1 *M* Tris–HCl pH 8.0 as the precipitant in the presence of galactose (20 m*M*), ADP (5 m*M*) and $MgCl_2$ (10 m*M*). A single selenomethionine-derivative crystal was used for the MAD data collection, which was performed to a resolution of 2.9 Å. An example of the diffraction images obtained is shown in Fig. 1. The data for each wavelength were processed using the *DENZO*/*SCALEPACK* package (Otwinowski & Minor, 1997); data-processing statistics are given in Table 1. Preliminary analysis of the X-ray diffraction

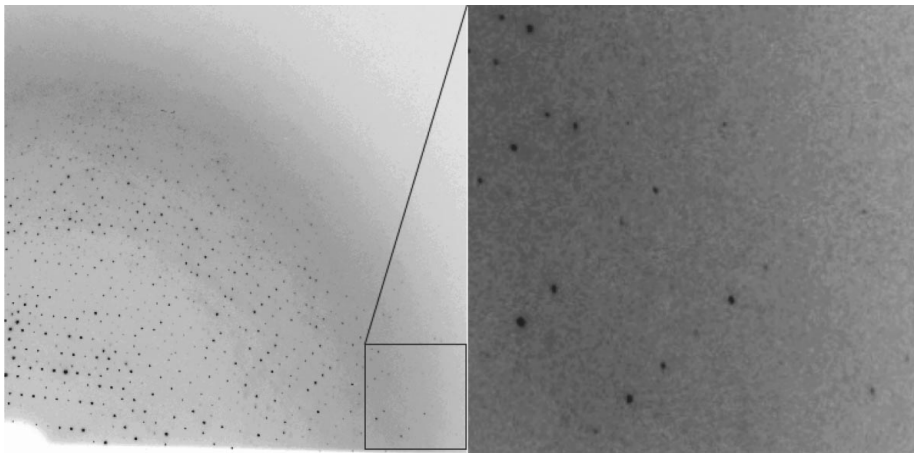

Figure 1
A 0.75° oscillation frame of a cryocooled crystal of GalK. Diffraction data are observed to 2.9 Å.

Table 1
Data-processing statistics for the single SeMet-incorporated GalK crystal.

Values in parentheses are for the highest resolution shell.

Data set	Peak (λ_1)	Inflection (λ_2)	Remote (λ_3)
Wavelength (Å)	0.97950	0.97973	0.93928
Resolution (Å)	25–3.0 (3.07–3.0)	25–2.9 (2.97–2.90)	25–3.1 (3.17–3.10)
Completeness	99.7 (99.7)	99.5 (98.5)	99.7 (99.3)
Reflections	449569	485409	406970
Unique reflections	124196	137024	112794
Redundancy	3.62	3.54	3.61
$\langle I/\sigma(I)\rangle$	16.47 (4.0)	17.81 (3.4)	15.70 (3.3)
$R_{\rm sym}$† (%)	7.5 (33)	6.0 (38)	7.1 (41)

† $R_{\rm sym} = \sum|I - \langle I\rangle|/\sum I$, where I is the integrated intensity of a given reflection.

data using the autoindexing routine in *DENZO* indicated that the crystal belonged to the centred system, point group 222, with unit-cell parameters a = 211.7, b = 355.4, c = 165.5 Å, $\alpha = \beta = \gamma = 90^\circ$. Analysis of the X-ray diffraction pattern showed that along the l axis reflections were only present if $l = 2n$, identifying the space group as $C222_1$. Using the subunit molecular weight of 40 kDa, considerations of $V_{\rm M}$ (Matthews, 1977) suggest that the asymmetric unit could contain between eight and 20 subunits. Given that GalK is a monomeric protein, the large number of copies in the asymmetric unit is somewhat unusual. Therefore, since the crystals were grown in the presence of galactose, ADP and Mg^{2+}, the gel filtration was repeated under equivalent conditions. This had no effect on the elution profile compared with the free enzyme and thus it appears that the number of subunits in the asymmetric unit is an artefact of crystal packing rather than reflecting a higher order quaternary structure for the enzyme. A full structural determination is currently under way.

We thank the support staff and Dr H. Belrhali of beamline ID14.4 at the ESRF Grenoble Laboratory for assistance with station alignment. This work was supported by the BBSRC and the Wellcome Trust. The Krebs Institute is a designated BBSRC Biomolecular Science Centre and a member of the North of England Structural Biology Centre. AH is a BBSRC CASE student with British Biotechnology Plc. CV was supported by the Earth and Life Science Foundation (ALW), which is subsidized by the Netherlands Organization for Scientific Research (NWO).

References

Bonanno, J. B., Edo, C., Eswar, N., Pieper, U., Romanowski, M. J., Ilyin, V., Gerchman, S. E., Kycia, H., Studier, F. W., Sali, A. & Burley, S. K. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 12896–12901.

Bork, P., Sander, C. & Valencia, A. (1993). *Protein Sci.* **2**, 31–40.

Bradford, M. M. (1976). *Anal Biochem.* **72**, 248–254.

Frey, P. A. (1996) *FASEB J.* **10**, 461–470.

Fu, Z., Wang, M., Potter, D., Miziorko, H. M. & Kim, J. P. (2002). *J. Biol. Chem.* **277**, 18134–18142.

Krishna, S., Zhou, T., Daugherty M., Osterman, A. & Zhang, H. (2001). *Biochemistry*, **40**, 10810–10818.

Matthews, B. W. (1977). *X-ray Structure of Proteins*, 3rd ed., edited by H. Neurath & R. L. Hill, Vol. 3, pp. 468–477. New York: Academic Press.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Platt, A. & Reece, R. J. (1998). *EMBO J.* **17**, 4086–4091.

Van Duyne, G. D., Standeart, R. F., Karplus, A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.

Verhees, C. H., Koot, D. G. M., Ettema, T. J. G., Dijkema, C., De Vos, W. M. & Van der Oost, J. (2002). *Biochem. J.* **366**, 121–127.

Yang, D., Shipman, L. W., Roessner, C. A., Scott, I. & Sacchettini, J. C. (2002). *J. Biol. Chem.* **277,** 9462–9467.

Zhou, T., Daugherty, M., Grishin, N. V., Osterman, A. L. & Zhang, H. (2000). *Structure Fold. Des.* **8**, 1247–1257.