doi:10.1016/j.jmb.2004.01.043 J. Mol. Biol. (2004) 337, 387–398

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Substrate Specificity and Mechanism from the Structure of *Pyrococcus furiosus* Galactokinase

Andrew Hartley[1], Steven E. Glynn[1], Vladimir Barynin[1] Patrick J. Baker[1], Svetlana E. Sedelnikova[1], Corné Verhees[2] Daniel de Geus[2], John van der Oost[2], David J. Timson[3] Richard J. Reece[4] and David W. Rice[1]*

[1]*Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN UK*

[2]*Laboratory of Microbiology Wageningen University Hesselink van Suchtelenweg 4 NL-6703 CT Wageningen, The Netherlands*

[3]*Medical Biology Centre School of Biology and Biotechnology, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK*

[4]*School of Biological Sciences University of Manchester 2.205 Stopford Building Oxford Road, Manchester M13 9PT, UK*

**Corresponding author*

Galactokinase (GalK) catalyses the first step of the Leloir pathway of galactose metabolism, the ATP-dependent phosphorylation of galactose to galactose-1-phosphate. In man, defects in galactose metabolism can result in disorders with severe clinical consequences, and deficiencies in galactokinase have been linked with the development of cataracts within the first few months of life. The crystal structure of GalK from *Pyrococcus furiosus* in complex with MgADP and galactose has been determined to 2.9 Å resolution to provide insights into the substrate specificity and catalytic mechanism of the enzyme. The structure consists of two domains with the active site in a cleft at the domain interface. Inspection of the substrate binding pocket identifies the amino acid residues involved in galactose and nucleotide binding and points to both structural and mechanistic similarities with other enzymes of the GHMP kinase superfamily to which GalK belongs. Comparison of the sequence of the Gal3p inducer protein, which is related to GalK and which forms part of the transcriptional activation of the *GAL* gene cluster in the yeast *Saccharomyces cerevisiae*, has led to an understanding of the molecular basis of galactose and nucleotide recognition. Finally, the structure has enabled us to further our understanding on the functional consequences of mutations in human GalK which cause galactosemia.

Keywords: galactokinase deficiency; GHMP kinase superfamily; *Pyrococcus furiosus*; X-ray crystallography

Introduction

Pyrococcus furiosus is a hyperthermophilic archaeon, isolated from thermally active regions. Like other saccharolytic organisms, *P. furiosus* utilises the glycolytic pathway as a main catabolic route. Conversion of galactose *via* glycolysis requires an additional metabolic branch, the Leloir pathway,[1] and the first step of this pathway involves the conversion of galactose to galactose-1-phosphate, catalysed by galactokinase (GalK; EC 2.7.1.6), which drives the transfer of the γ-phosphate group from ATP to the O1 position of galactose.

In man, galactosemia is an autosomal recessive disorder caused by a defect in one of the three Leloir pathway enzymes involved in galactose metabolism, galactokinase, galactose-1-phosphate uridyl transferase (GalT) and UDP-galactose 4′ epimerase (GalE).[2] Deficiency in the transferase, GalT, is responsible for classic galactosemia (MIM 230400) and can lead to severe neonatal symptoms

Abbreviations used: F_A, structure factor amplitudes for the anomalous scatterers; GalK, galactokinase; GHMP, galactokinase homoserine kinase, mevalonate kinase phosphomevalonate kinase; GTP, guanosine-5′-triphosphate; HSK, homoserine kinase; MAD, multi-wavelength anomalous dispersion; MMK, *Methanococcus janaschii* mevalonate kinase; NCS, non-crystallographic symmetry; *Pf*, *Pyrococcus furiosus*; RMK, *Rattus norvegicus* mevalonate kinase; rmsd, root-mean-squared deviation; SeMet, selenomethionine; σ, standard deviation.

E-mail address of the corresponding author: d.rice@sheffield.ac.uk

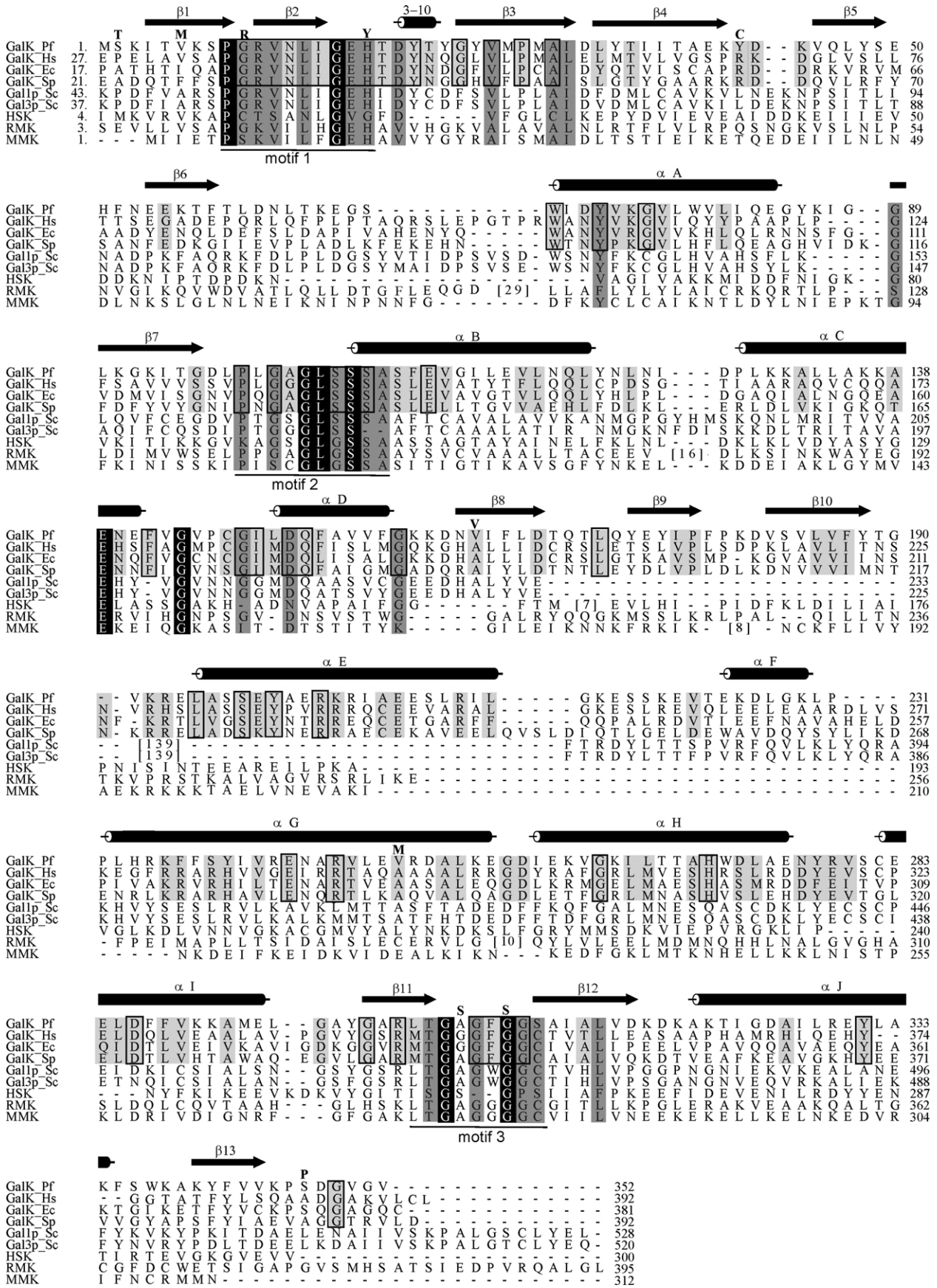

Figure 1. Structure-based sequence alignment of representative members of the GHMP family, galactokinase family and yeast Gal3p. GalK, galactokinase; HSK, *Methanococcus janaschii* homoserine kinase; RMK, *Rattus norvegicus* mevalonate kinase; MMK, *M. janaschii* mevalonate kinase; Pf, *Pyrococcus furiosus*; Hs, *Homo sapiens*; Ec, *Escherichia coli*; Sp,

such as failure to thrive, hepatomegaly and bacterial sepsis. Galactokinase deficiency (galactosemia II; MIM 230200) in man is an inborn error of galactose metabolism and is linked to development of cataracts during the first months of life and also pre-senile cataracts, the onset of which is between 20 and 50 years of age.[3] Over 20 mutations have been identified in human galactokinase that are associated with reduced galactokinase activity in the blood,[4–8] 11 of these mutations are single amino acid substitutions and the biochemical characteristics of some of these have recently been characterised.[9]

Studies on the yeast *Saccharomyces cerevisiae* have led to identification of a galactokinase (Gal1p) in this organism as part of a gene cluster which encodes the enzymes of the Leloir pathway.[10,11] The transcriptional activation of this cluster, which results in Gal1p expression involves three proteins: a transcriptional activator and DNA-binding protein, Gal4p; a repressor, Gal80p; and a ligand sensor and inducer, Gal3p. Induction occurs as a result of an ATP and galactose-dependent interaction between Gal3p and Gal80p with Gal3p acting as the ligand sensor and transducer of galactose signal.[12] Sequence analysis has shown that there is a very high similarity (73% identity) between the yeast proteins Gal1p and Gal3p.[13] However, despite this similarity, Gal3p does not show galactokinase activity and the transcriptional activation of the gene cluster by Gal1p is also observed, although at a much reduced level compared to Gal3p.[13] Determination of the structure of *Pyrococcus furiosus* (*Pf*) GalK may offer insights into the mechanism of galactose and ATP binding in the yeast Gal3p protein and also shed light on the interactions involved in the Gal3p–Gal80p complex.

Sequence similarities have suggested that GalK is a member of the GHMP kinase superfamily[14] for which the structures of several family members have now been reported, including homoserine kinase from *Methanococcus janaschii*[15] (HSK; EC 2.7.1.39), mevalonate kinase from *Rattus norvegicus*[16] (RMK; EC 2.7.1.36), and from *M. janaschii*[17] (MMK), mevalonate diphosphate decarboxylase from *M. janaschii*[18](MDD; EC 2.7.4.2) and phosphomevalonate kinase from *Streptococcus pneumoniae*[19] (PMK; EC 2.7.4.2). Although the overall sequence identity in the family is low[20] (with 16%, 15% and 20% identities between *Pf* GalK and HSK, RMK and MMK, respectively), three conserved glycine-rich motifs can be identified (motifs 1–3; Figure 1). 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 less common, *syn* conformation of the glycosidic bond of the ADP in the active site.[20] However, determination of the structure of another GHMP kinase family member, mevalonate kinase,[16] in complex with ATP, revealed that in some enzymes of the family, the glycosidic bond can adopt an *anti* conformation. Analysis of the structures of the GHMP kinase superfamily has shown that the most conserved motif across the GHMP kinase family, with a consensus of $PX_3GL(G/S)SSA$ (motif 2; Figure 1), forms an atypical phosphate binding P-loop which interacts through hydrogen bonds to the α and β-phosphate groups of ADP. More recently, the three-dimensional structure of the GalK from *Lactobacillus lactis* in complex with galactose and inorganic phosphate has been reported. This structure confirms the general similarity of the GalK fold to the other members of the GHMP kinase superfamily.[21]

Here, we report the three-dimensional structure of *Pf* GalK at 2.9 Å resolution in a complex with ADP and galactose. The structure reveals the mode of nucleotide recognition and its relationship to the galactose-binding site. Comparison of the structure with other members of the GHMP kinase superfamily has permitted the identification of the residues involved in substrate binding and catalysis and led to a better understanding of the lack of galactokinase activity in Gal3p. The sequence similarity between *Pf* GalK and human GalK has facilitated a homology-based modelling study, which has shed light on the structural basis of the galactosemia causing amino acid substitutions in the human GalK protein.

Results and Discussion

Quality of the structure

The GalK structure was solved using multi-wavelength anomalous dispersion (MAD) data collected on a single selenomethionine labelled (SeMet) crystal to 2.9 Å resolution. The final refined model of GalK contains nine independent subunits in the asymmetric unit (labelled A through I in Figure 2(a)), a total of 2999 amino acid residues. Analysis of the packing in the cell reveals subunit I is related by a crystallographic 2-fold axis to its symmetry related partner, I′ and subunits A and D appear to be related by an approximate non-crystallographic 2-fold axis of symmetry within the crystal. However, the

Streptococcus pneumoniae; Sc, *Saccharomyces cerevisiae*. The invariant residues are highlighted in black with identities in 7/9 sequences highlighted in dark grey. Identities in 3/4 of the non-yeast galactokinases are highlighted in light grey. The secondary structural elements in the *Pf* GalK are shown above the sequences. Invariant residues amongst the galactokinase family are boxed. The mutations identified in galactokinase deficiency in man are identified above the sequences with the corresponding substituted amino acid single letter code.

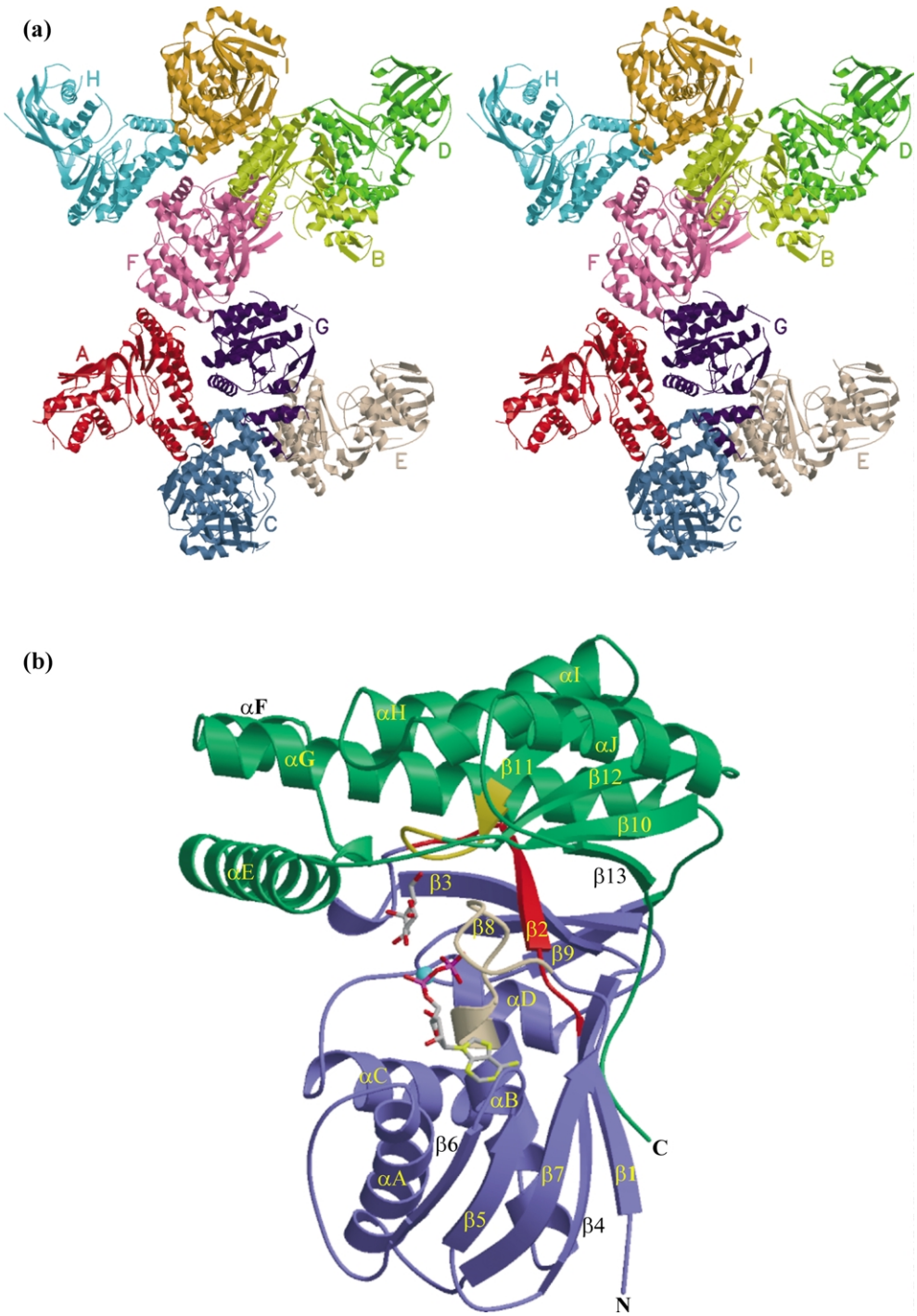

Figure 2. (a) Schematic diagram of the crystal packing arrangement of GalK. The positions of all nine monomers are shown and labelled A to I. (b) Schematic diagram of the GalK monomer. The N-terminal domain is shown in blue and the C-terminal domain in green. The three conserved motifs in the GHMP kinase superfamily are highlighted (1, red; 2, beige; 3, gold). Helices are labelled αA through αJ and β-strands labelled B1 to B13. ADP and galactose are represented with a stick diagram (atom-coloured). Mg^{2+} is coloured cyan. Both the N and C termini are labelled. Label colours are not significant.

interfaces that mediate the interactions between these pairs of subunits are different and the fraction of the solvent accessible surface buried across both interfaces (3% and 4.5% for subunits I–I′ and A–D, respectively) lies below the minimum range found in an analysis of other dimeric proteins.[22] Gel filtration of *Pf* GalK resulted in elution at the approximate molecular mass of the monomer (40–45 kDa compared to the calculated molecular mass of 39.4 kDa, SWISSPROT: Q9HHB6,[23,24]), suggesting that the monomer described here is the biologically relevant oligomerization state. The path of the polypeptide chain is well defined by the electron density map for five subunits.

However, in two of the other subunits 4% and 2% of residues have very poor density for the main-chain and in the other two subunits, 12% and 27% of the residues corresponding to part of the N-terminal domain have very poor density for the main-chain. Superposition of the C^{α} atoms of the subunits using the program LSQKAB[25] gives an average rmsd value for the C^{α} atoms of 0.23 Å showing that there are no major conformational differences in the subunits.

Overall polypeptide fold

The following discussion is based on an analysis of subunit I, the best-defined subunit in the electron density. The GalK monomer is approximately 60 Å × 55 Å × 40 Å in size, comprising two domains separated by a deep cleft. The N-terminal domain (Figure 2(b)) is composed of a mixed, five-stranded β-sheet (β1, β4–β7) packed on one face by a bundle of four helices (αA–αD) with the other face exposed entirely to solvent. A small anti-parallel β-sheet (β2, β3, β8, β9) that lies at the C-terminal end of a five-stranded β-sheet dominates the domain interface. The C-terminal domain (Figure 2(b)) is composed of an antiparallel β-sheet (β10–β13) and five α-helices (αE–αH) with the β-sheet completing the domain interface.

Nucleotide binding site

Crystallisation of *Pf* GalK was achieved in the presence of ADP (5 mM), $MgCl_2$ (10 mM) and galactose (20 mM).[23] A difference Fourier map calculated at a late stage of the refinement revealed two areas of additional electron density in the cleft between the two domains. This density lies in a similar position to the nucleotide and substrate binding sites of the homologous GHMP kinase family members and an ADP and galactose molecule were fitted into this density and refined. The quality of the electron density compared to the surrounding protein is such that we estimate the ADP/Mg^{2+} and galactose sites are fully occupied (Figure 3(a) and (b)). The ADP molecule is bound in an *anti* conformation with the adenine ring situated in a hydrophobic pocket formed by the side-chains of residues Phe52, Trp69, Ile94, Phe110 and Leu100 (Figure 4(a)). This *anti* conformation of the adenine ring contrasts with the *syn* conformation proposed in the recent structure determination of GalK from *L. lactis* on the basis of modelling.[21] The adenine ring of ADP forms a π face-to-face stacking interaction with the aromatic ring of Phe110 on one side with the other face packing against the side-chain of Leu100 (Figure 4(a)). The N1 nitrogen atom on the adenine ring makes a hydrogen bond with the hydroxyl oxygen atom of Ser49. There are no other direct hydrogen bonds between protein atoms and the nitrogen atoms on the adenine ring. However, there is space in the structure for water molecules to be involved in water-mediated hydrogen bonds between the N6 amino group and main-chain

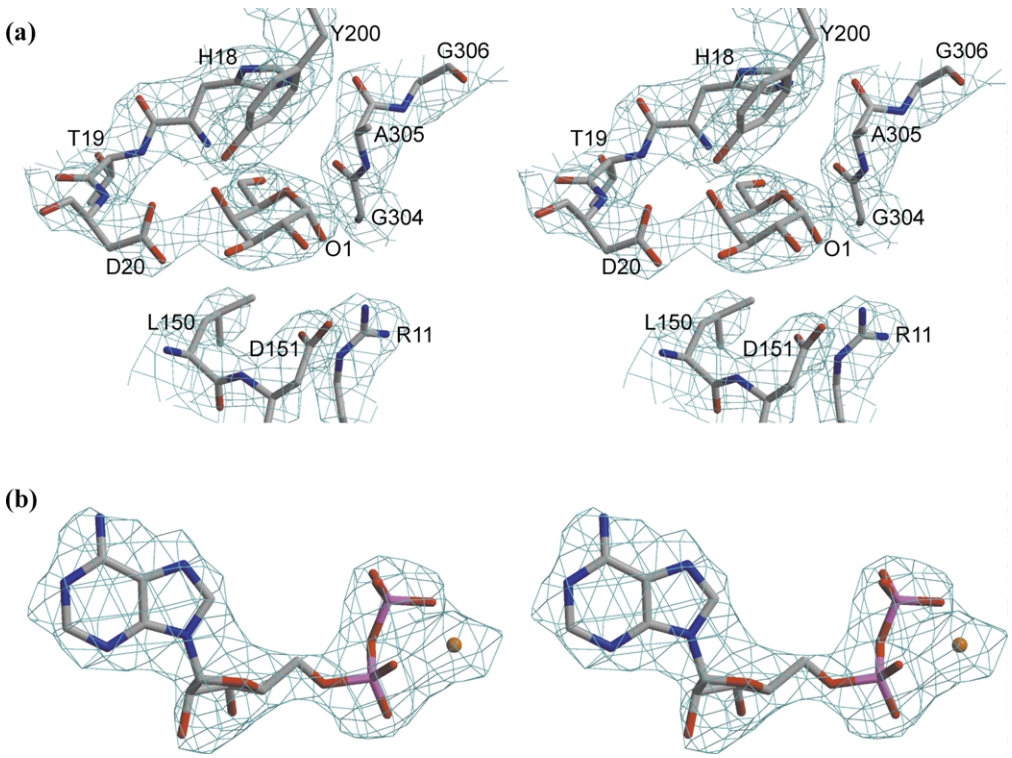

Figure 3. (a) Final electron density map ($2F_o - F_c$) in the vicinity of the galactose-binding site contoured at the 1σ level. The galactose molecule and surrounding protein residues are labelled and shown in stick format. (b) Final electron density map, as in (a), in the vicinity of the enzyme-bound MgADP. The ADP molecule is shown as sticks and the magnesium ion is shown with an orange ball.

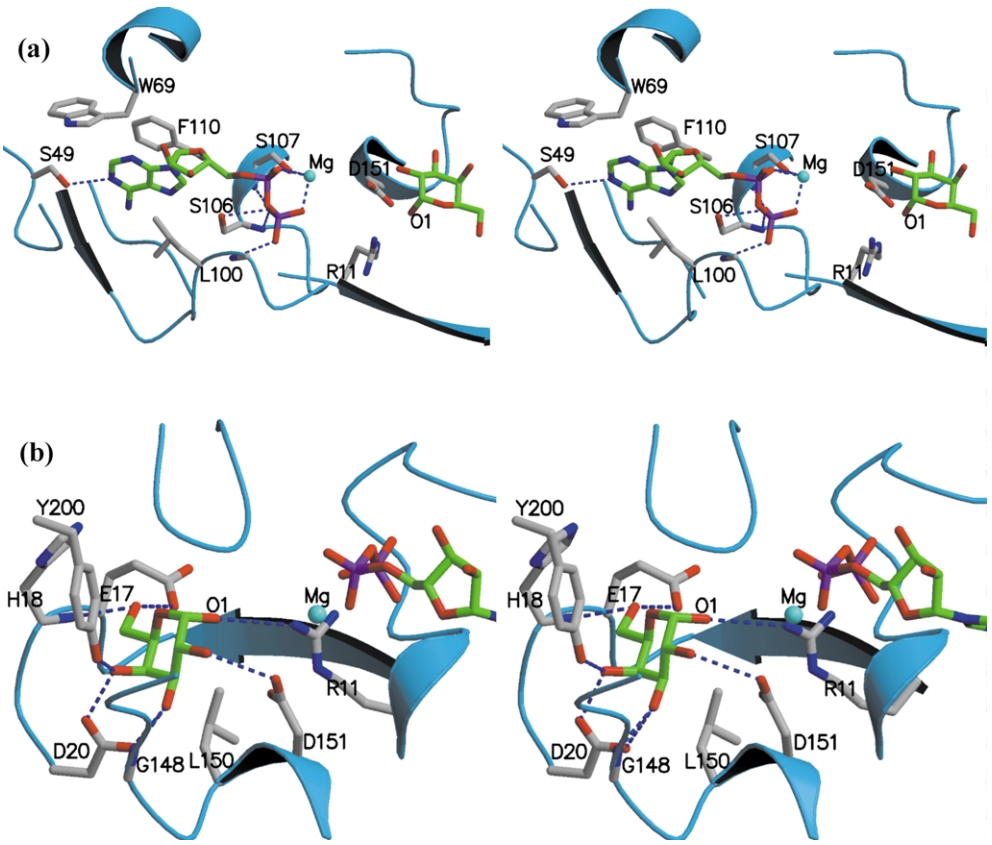

Figure 4. (a) Stereo diagram of the ADP-binding site showing the ADP and galactose (green and atom colours) and the protein backbone of nearby residues in blue. Residues involved in direct interactions to the ADP are shown in stick format (grey and atom colours). Hydrogen bonds to the ADP and ion pair interactions to the Mg^{2+} are shown as blue broken lines. (b) Stereo diagram of the galactose-binding site showing galactose (grey and atom colours), Mg^{2+} and part of the ADP (green and atom colours). The nearby residues involved in direct interactions with the galactose and hydrogen bonds to the galactose are shown as for (a).

atoms. Biochemical studies have shown that *Pf* GalK will also utilise GTP as a phosphoryl donor with only a 40% decrease in activity compared to ATP.[24] This is consistent with the analysis of the purine-binding pocket, which suggests that a GTP molecule could be accommodated with only minor modifications to the position of the guanine ring. The ribose moiety packs against the side-chain aromatic ring of Tyr72 with its 2′ and 3′ hydroxyl groups exposed to the solvent and not forming any direct hydrogen bonds to the enzyme. The residues ^{99}PLGAGLSSSAS^{109} corresponding to motif 2 (Figure 1), form a phosphate-binding loop which wraps around the β-phosphate group, forming hydrogen bonds with the main-chain amide groups of Gly101 and Ser107 and the side-chain hydroxyl groups of Ser106 and Ser107. At this resolution, we cannot unambiguously identify the position of the bound magnesium, which we assume is part of the complex given its inclusion in the crystallisation mixture. However, the electron density shows a bulge between the α and β-phosphate groups (Figure 3(b)), the position of which maps exactly on to that identified as the divalent cation in the structures of the other GHMP family members. The magnesium in the GalK structure is coordinated by the side-chain oxygen atoms of the conserved residues, Glu130 and Ser107 and by oxygen atoms from the α and β-phosphate groups of ADP (Figure 4(a)).

Galactose binding

The galactose sits in a cavity between the two domains, oriented with the phosphoryl acceptor oxygen atom (O1) towards the phosphate-binding loop and the position of the terminal phosphate of the bound ADP nucleotide (Figure 4(b)). The cavity is formed from and surrounded by residues conserved amongst galactokinases (motifs 1 and 3; see Figure 1) and these form extensive and specific interactions with the galactose oxygen atoms. The O6 hydroxyl of galactose forms hydrogen bonds with a carboxyl oxygen atom of Glu17 and the main-chain amide of His18. The O4 atom of galactose forms a hydrogen bond with the hydroxyl group of Tyr200. The carboxyl oxygen atoms of Asp20 hydrogen bond with both the O4 and O3 hydroxyl oxygen atoms. The carboxyl oxygen of Asp151 is in a position to form a hydrogen bond with the O2 hydroxyl of galactose and is 3.4 Å from the O1 hydroxyl group, which is phosphorylated by GalK. Arg11 also forms a hydrogen bond with this phosphoryl acceptor hydroxyl. The

interactions made by galactose in *Pf* GalK are similar to those made by galactose in *L. lactis* GalK.[21]

Comparison with other GHMP kinase family members

The overall structure of GalK is very similar to the structures of the other GHMP kinase superfamily that have been solved to date, even though overall sequence similarity within the family is low. The conserved motifs, 1–3 (Figure 1), show structural similarity to the other GHMP family members with motifs 1 and 3 forming part of the galactose-binding site and motif 2 forming the phosphate-binding loop. However, there are some differences in structure between GalK and the other GHMP kinase family members. The *anti* conformation of the ADP bound to GalK is similar to that observed in the complex of RMK and ATP[16] but is in contrast to the *syn* conformation of the nucleotide found in the equivalent complex with HSK.[15] However, the binding of the adenine ring in GalK is different to the RMK structure as there are no direct hydrogen bonds involving the amino group of the adenine ring in the GalK structure. Instead, it is possible that water-mediated hydrogen bonds are involved. The region corresponding to helices αE and αF in *Pf* GalK (residues 195–231) does not show similarity to any of the other GHMP kinase structures and a DALI search reveals no other structures with a *Z*-score of over 2.5, suggesting that this may be a novel structural motif.

GalK catalytic mechanism

Previous studies on GHMP kinase superfamily members have led to two distinct proposals for the catalytic mechanism. Phosphoryl transfer in HSK has been suggested to occur by direct nucleophilic attack on the γ-phosphate group of ATP by the δ-hydroxyl of homoserine.[20] In this mechanism, the latter is stabilised by the formation of a hydrogen bond to a neighbouring asparagine residue (Asn141), not conserved in the superfamily. Catalysis is proposed to be assisted through activation of the γ-phosphate of ATP by the magnesium ion, which is coordinated by a conserved glutamate residue (Glu130) with the deprotonation of the δ-hydroxyl possibly involving the γ-phosphate.[20]

In contrast, studies on both RMK[16] and MMK[17] have suggested a mechanism involving an aspartic acid residue conserved in these two enzymes as a possible catalytic base (Asp204 in RMK) whose role is to activate the receptor hydroxyl group. A conserved lysine residue (Lys13 in RMK) is then proposed to stabilise the resultant C5 alkoxide group and the penta-coordinated γ-phosphate.

Comparison of the structures of HSK, RMK and GalK reveals the presence of Arg11 and Asp151 of GalK in positions equivalent to the proposed catalytic lysine and aspartate residues, respectively, in the mevalonate kinases (Figure 5). Asp151 is located 3.8 Å from the O1 hydroxyl of galactose, suggesting that this residue is well placed to play a possible role in proton abstraction. Arg11 is located 3.2 Å from the O1 phosphoryl acceptor hydroxyl with its guanidinyl group suitably positioned to stabilise the developing negative charge which accumulates during the transition state. As observed in RMK, MMK and HSK, a glutamate residue is involved in coordination of the magnesium ion (Glu193 in RMK, Glu130 in HSK and Glu139 in GalK). Studies on HSK have suggested that Gln152 in GalK may have a comparable role to Asn141 in HSK, responsible for stabilisation of the acceptor hydroxyl group.[20] However, superposition of the structures shows that Asn141 in HSK is equivalent to the putative catalytic

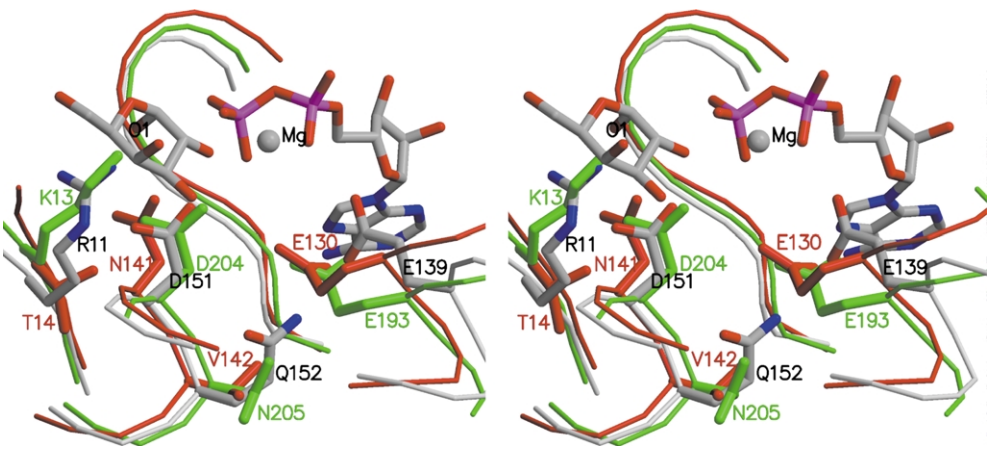

Figure 5. Stereo diagram showing a superposition of the active site residues of GalK (grey), HSK (red) and RMK (green). The galactose and ADP molecules as well as the magnesium ion from the GalK structure are also shown. The superposition shows the equivalence in position of the proposed catalytic residues in GalK and RMK (Asp151/Asp204 and Arg11/Lys13, respectively). Notable differences in these residues seen in HSK, suggest that there may be differences in the catalytic mechanism of the latter. The disparity in the positions of the proposed catalytic asparagine in HSK (Asn141) and Gln152 of GalK is clearly shown.

aspartate residues in GalK and RMK (Asp151 and Asp204, respectively) (Figure 5). Moreover, the amide group of Gln152, being the next residue in this helical section of the molecule, is some 9 Å away from the position of the O1 hydroxyl of galactose and not in a suitable orientation to be considered a potential catalytic residue. In addition, HSK is the only enzyme in the family that does not possess a positively charged residue equivalent to Lys13 (RMK) and Arg11 (GalK) and instead, HSK has a threonine (Thr14) at this position (Figure 5). This would suggest that the mechanistic details of GalK are more closely related to the mevalonate kinases and possibly distinct to HSK.

Comparison with *S. cerevisiae* GalK and Gal3p

The structure-based sequence alignment of the galactokinase family reveals significant sequence similarity between *Pf* GalK and the yeast *S. cerevisiae* galactokinase (Gal1p) and Gal3p inducer protein (Figure 1). For Gal1p and Gal3p, the location of the strongly conserved residues is virtually identical to that seen across the galactokinase family. Thus, of the seven residues intimately involved in galactose binding in the *Pf* enzyme, six can be identified in the Gal1p and Gal3p proteins, the only exception being the replacement of Tyr200 by a lysine in both. Similarly, of the five residues involved in ATP binding and conserved across the galactokinase family, four can be identified in Gal3p. This implies that the recognition of galactose and ATP by these proteins will be closely related. Compared to the *Pf* enzyme both of the yeast proteins are considerably larger with the difference being primarily due to an insertion of 93 residues that is located in the loop between helices αE and αF where it would not interfere with the basic fold of the protein.

Wild-type Gal3p lacks galactokinase activity despite sharing most of the substrate binding motifs with GalK enzymes. However, Gal3p does have a two-residue deletion within motif 3 (Figure 1) compared to other members of the family, corresponding to the loss of Ser107 and Ala108 in the *Pf* structure. In *Pf* GalK Ser107 interacts with both the magnesium ion and β-phosphate of ATP. Gal3p is only functionally active in the presence of a nucleotide (ATP or ADP) and magnesium[26,27] and therefore it is likely that this deletion affects the relative positioning of the γ-phosphate group of the ATP with respect to the O1 hydroxyl of galactose rather than preventing substrate binding. This is consistent with the biochemical data showing that the insertion of serine and alanine residues at this position in Gal3p confers galactokinase activity, although at a reduced level when compared to Gal1p[13].

Insights into galactosemia inducing mutations of GalK

The structure of *Pf* GalK, together with the close sequence similarity amongst the galactokinase family has allowed us to investigate the location and effect of the mutations in the monomeric human galactokinase[9] that cause galactosemia, comparing earlier conclusions drawn from the structure of *L. lactis* GalK.[21]

P28T is a mutation causing galactosemia found amongst Roma gypsies,[5] and corresponds to Ser2 in the *Pf* structure (Figure 6), which is at the start of strand β1 in the N terminus. The human GalK

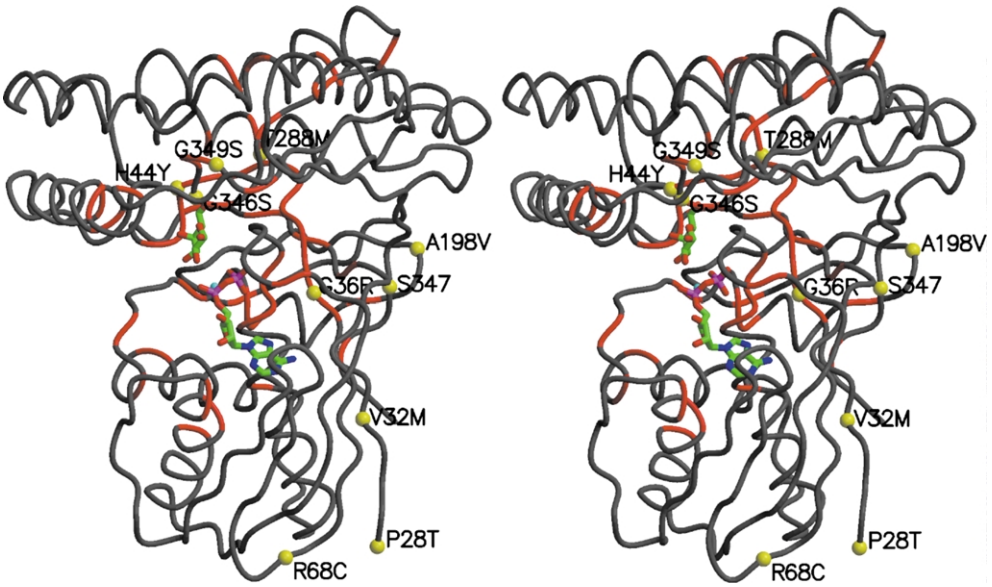

Figure 6. Schematic stereo representation of the C^{α} backbone of the GalK monomer with invariant residues in the galactokinase family coloured red. The positions of the amino acid substitutions in human GalK responsible for galactosemia are mapped onto the *Pf* GalK and shown as yellow spheres and the corresponding labels indicate the mutations in human GalK.

has a 25 amino acid insertion before the *Pf* N terminus so it is likely that this area differs in structure, preventing a detailed comparison.

V32M, another common mutation causing galactosemia,[7] corresponds to Val6 in the *Pf* structure. Val6 is located on strand β1 which makes up the N-terminal β-sheet that is conserved amongst members of the GHMP family, and the side-chain of this residue is buried in a hydrophobic cluster which involves residues from helix αC. This helix is important in providing contacts to the magnesium ion through the side-chain of Ser107 and also provides a dipole charge stabilisation of the negative charge of the pyrophosphate moiety of the nucleotide. The substitution of valine by a bulkier methionine residue in this buried part of the molecule is likely to cause substantial disruption to the structure.

G36R corresponds to Gly10 in *Pf* GalK, a highly conserved residue in motif 1 (Figure 1), which packs against the phosphate-binding loop. Substitution to an arginine residue would disturb this packing interaction and cause alterations in the structure of the nucleotide-binding site. This mutation, along with P28T and V32M, when made in human GalK caused the protein to be insoluble upon expression in *Escherichia coli*, suggesting that misfolding or destabilisation of the folded state is caused by these mutations.[9]

H44Y corresponds to a conserved histidine residue in *Pf* GalK, His18, which is located in motif 1 which forms a loop between β2 and a 3_{10} turn forming part of the galactose binding pocket. The $N^{\varepsilon 1}$ and $N^{\varepsilon 2}$ nitrogen atoms of this histidine form hydrogen bonds to the main-chain carbonyl oxygen atom of Ala305 and the side-chain hydroxyl oxygen of Tyr21, respectively, and the main-chain carbonyl oxygen forms a hydrogen bond with the O6 atom of galactose. Substitution of this buried histidine residue to a bulkier tyrosine would result in loss of these interactions but probably more importantly, the pocket that the histidine side-chain sits in cannot accommodate the larger tyrosine ring. This would cause disruption of the galactose-binding pocket, consistent with the reduced K_m value for galactose and lower k_{cat} value found in the human enzyme with this mutation.[9]

R68C corresponds to Tyr42 in *Pf* GalK, which is situated on a large loop between β3 and β4 on the solvent exposed surface of the N-terminal domain. This residue, however, is located over 30 Å from the active site and whilst it is conserved between the human, *E. coli* and *S. pneumoniae* galactokinases, it is surprising that a substitution so remote can have an impact on the functionality of the enzyme.[9]

A198V corresponds to Val163 in *Pf* GalK and is located on strand β8, which forms part of the antiparallel β-sheet in the N-terminal domain. Human GalK with the A198V mutation has kinetic properties that are almost identical to wild-type human GalK.[9] This is consistent with the observation that the site of this mutation is distant from the active site and the conservative nature of the mutation in a relatively uncrowded environment would probably have little effect on the function of this enzyme. This is further supported by an analysis of the levels of erythrocytic galactokinase in patients with this mutation, which parallels the reduction in k_{cat}[8] value, suggesting that this mutation does not affect the function of galactokinase.

T288M corresponds to Val248 in *Pf* GalK and is located in the middle of αG with its side-chain packed against the C-terminal β-sheet and particularly close to residue Gly158, which is conserved amongst the galactokinase family. Gly158 immediately follows helix αD that forms one part of the active site and contains the potential catalytic residue, Asp151. Creation of the T288M mutation in human GalK caused the protein to be insoluble on expression,[9] suggesting that perturbation of this packing interaction causes misfolding of GalK.

G346S appears in a loop between strands β11 and β12, which is an area of high sequence similarity across the GHMP kinase family where it forms part of motif 3, corresponding to Ala305 in *Pf* GalK. This residue is either an Ala or Gly in the aligned GHMP family members and galactokinase family (Figure 1) and in the *Pf* structure the main-chain carbonyl group of Ala305 forms a hydrogen bond to His18, which itself forms an important interaction with the galactose. The side-chain of Ala305 points towards the solvent but with the β-carbon atom only 4.7 Å from the O1 oxygen of galactose. Whilst it appears from the analysis of the structure that the substitution to serine can be accommodated with the additional hydroxyl not forming any adverse steric interactions, the likely structural changes that accompany the attack of the O1 oxygen atom on the γ-phosphate of ATP may well be affected. This is consistent with the finding that this mutant showed little change in the K_m value for galactose but the $k_{cat}/K_{m,gal}$ value was reduced 100-fold compared to the wild-type GalK.[9]

G349S corresponds to Gly308 in *Pf* GalK, which is a completely conserved residue in motif 3 lying close to the active site. This residue lies in a region of the Ramachandran plot that is unfavourable for residues other than glycine. The observed reduction in k_{cat} resulting from this mutation[9] may arise as a result of the enforced differences in the Ramachandran angles adopted in the mutant structure and a possible reduction in flexibility, affecting important transition state interactions.

A384P corresponds to S347 in *Pf* GalK and is located in the C-terminal tail away from the active site. However, this is an area of low sequence similarity in the galactokinase family and therefore it is not possible to accurately describe the effect that this mutation may have in the human homologue.

Conclusion

The structure of *Pf* GalK presented here represents the first for the ternary complex of this enzyme and the first description of any GalK from a thermophilic organism. The structure has revealed the molecular basis for the recognition of galactose, ADP and Mg^{2+} and has provided insights in to the enzyme mechanism. Analysis of the structure and comparison with other GHMP kinase family members has revealed considerable similarities in nucleotide binding and catalysis. In particular the mode of nucleotide binding is closely related to that seen in RMK and MMK with an *anti* conformation of the glycosidic bond unlike the *syn* conformation observed in HSK. In mechanistic terms, the identification of conserved negatively and positively charged residues (Asp151 and Arg11, respectively) in equivalent positions to those proposed to act as a catalytic base and to stabilise the negative charges generated at the catalytic centre in the mevalonate kinases (Asp204 and Lys13 in RMK), points to clear similarities in mechanism. Finally, the structure of *Pf* GalK has been useful in extending our ideas on the interactions which control ATP and galactose recognition in the yeast protein Gal3p and the consequences of mutations which lead to galactosemia in man.

Materials and Methods

SeMet incorporation, over-expression, purification and crystallisation of *Pf* GalK were performed as described.[23] For data collection, described in detail by De Geus *et al.*,[23] the crystal was transferred to a cryoprotectant solution containing the original precipitant solution plus 20% (v/v) glycerol and flash frozen at 100 K. A single crystal was used for MAD experiments performed at three wavelengths on beamline ID14.4 at the ESRF Grenoble laboratory at 100 K. The data for each wavelength were processed and merged using the DENZO/SCALEPACK package.[28] The crystals belong to space group $C222_1$ with cell dimensions $a = 211.7$ Å, $b = 355.4$ Å, $c = 165.5$ Å, $\alpha = \beta = \gamma = 90°$.

The diffraction data were analysed and F_A values determined using the program XPREP (Bruker AXS, Madison, USA). The anomalous signal was found to extend to 2.9 Å and F_A values were truncated to this resolution. The SeMet substructure was solved using the "half-baked" approach as implemented in the program SHELXD.[29] Twenty-seven sites were found and validated, based on the consistency with the Patterson function and with the non-crystallographic symmetry, by manual inspection of the crossword table provided by SHELXD. This confirms that there are nine independent monomers in the asymmetric unit with a V_M value of 4.4 Å^3/Da, corresponding to a high solvent content of ~70%. Diffraction data were manipulated with the CCP4 suite of programs.[25] The selenium sites were refined and phases calculated using MLPHARE[30] using the inflection (λ_2) as the native data set. Solvent flattening, histogram matching and ninefold non-crystallographic symmetry (NCS) averaging were applied using DM[31] and an initial model was built into the MAD phased map calculated to 2.9 Å using TURBO-FRODO.[32] Iterative cycles of refinement and rebuilding using tight NCS restraints until the final cycle, in REFMAC[33] and including TLS refinement, resulted in a final model of 22,729 atoms, with $R = 0.23$ and $R_{free} = 0.27$ (Table 1). The overall average *B*-factor is 64 Å^2. In the final model there are no non-glycine Ramachandran outliers, 90% of the residues are in the most favoured conformation and all 27 selenium sites correspond to the positions of methionine residues. The Figures shown here have been prepared using the following programs; Figure 1, ALSCRIPT;[34] Figures 2 and 4–6, MOLSCRIPT[35] and Raster3D;[36] Figure 3, BOBSCRIPT[37] and Raster3D[36].

Table 1. Phasing and refinement statistics

Resolution range (Å)	10–2.9
Space group	$C222_1$
Unit cell parameters (Å)	$a = 211.7$, $b = 355.4$, $c = 165.5$
No. of reflections (working/free)	127,499/6821
R_{cullis} acentric (peak/remote)	0.84/0.77
Overall figure-of-merit	0.56
Protein non-H atoms	22,729
Galactose atoms	108
ADP atoms	243
Mg atoms	9
R-factor[a]/R_{free}[b]	0.23/0.27
rms deviations[c]	
Bond lengths (Å)	0.015
Bond angles (deg.)	1.66
Average *B* (Å^2)	64
Ramachandran plot statistics[d]	
Most favoured regions (%)	89.9
Additional allowed regions (%)	10.1
Disallowed regions (%)	0

[a] $R\text{-factor} = \sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$.
[b] R_{free} is the crystallographic *R* value calculated using 5% of the native data witheld from refinement.
[c] Root-mean-square (rms) deviations are from ideal bond lengths and angles, respectively.
[d] Average over all subunits.

Accession numbers

The atomic coordinates for *Pf* GalK in complex with galactose and ADP have been deposited in the RCSB Protein Data Bank with accession number 1S4E.

Acknowledgements

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, the Wellcome Trust and the Leverhulme Trust. The Krebs Institute is a designated BBSRC Biomolecular Science Centre and a member of the North of England Structural Biology Centre. A.H. is a BBSRC CASE student

with British Biotechnology. C.V. was supported by the Earth and Life Science Foundation (ALW), which is subsidized by the Netherlands Organization for Scientific Research (NWO).

References

1. Frey, P. A. (1996). The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB J.* **10**, 461–470.
2. Petry, K. G. & Reichardt, J. K. V. (1998). The fundamental importance of human galactose metabolism: lessons from genetics and biochemistry. *Trends Genet.* **14**, 98–102.
3. Stambolian, D., Scarpino-Myers, V., Hodes, B. & Harris, H. (1986). Cataracts in patients heterozygous for galactokinase deficiency. *Invest. Opthal. Vis. Sci.* **27**, 429–433.
4. Stambolian, D., Ai, Y., Sidjanin, D., Nesburn, K., Sathe, G., Rosenberg, M. & Bergsma, D. J. (1995). Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts. *Nature Genet.* **10**, 307–312.
5. Kalaydjieva, L., Perez-Lezaun, A., Angelicheva, D., Onengut, S., Dye, D., Bosshard, N. U. *et al.* (1999). A founder mutation in the GK1 gene is responsible for galactokinase deficiency in Roma (Gypsies). *Am. J. Hum. Genet.* **65**, 1299–1307.
6. Kolosha, V., Anoia, E., de Cespedes, C., Gitzelmann, R., Shih, L., Casco, T., Saborio, M. *et al.* (2000). Novel mutations in 13 probands with galactokinase deficiency. *Hum. Mutat.* **15**, 447–453.
7. Hunter, M., Angelicheva, D., Levy, H. L., Peuschel, S. M. & Kalaydjieva, L. (2001). Novel mutations in the GALK1 gene in patients with galactokinase deficiency. *Hum. Mutat.* **17**, 77–78.
8. Okano, Y., Asada, M., Fujimoto, A., Ohtake, A., Murayama, K., Hsiao, K. J. *et al.* (2001). A genetic factor for age-related cataract: identification and characterization of a novel galactokinase variant, "Osaka", in Asians. *Am. J. Hum. Genet.* **68**, 1036–1042.
9. Timson, D. & Reece, R. J. (2003). Functional analysis of disease-causing mutations in human galactokinase. *Eur. J. Biochem.* **270**, 1767–1774.
10. Johnston, M. (1987). A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **51**, 458–476.
11. Reece, R. J. & Platt, A. (1997). Signaling activation and repression of RNA polymerase II transcription in yeast. *Bioessays*, **19**, 1001–1010.
12. Yano, K. & Fukasawa, T. (1996). Galactose-dependent reversible interaction of Gal3p with Gal80p in the induction pathway of Gal4p-activated genes of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **94**, 1721–1726.
13. Platt, A., Ross, H. C., Hankin, S. & Reece, R. J. (2000). The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase. *Proc. Natl Acad. Sci. USA*, **97**, 3154–3159.
14. Bork, P., Sander, C. & Valencia, A. (1993). Convergent evolution of similar enzymatic function on different protein folds—the hexokinase, ribokinase and galactokinase families of sugar kinases. *Protein Sci.* **2**, 31–40.
15. Zhou, T., Daugherty, M., Grishin, N. V., Osterman, A. L. & Zhang, H. (2000). Structure and mechanism of homoserine kinase: prototype for the GHMP kinase superfamily. *Struct. Fold. Des.* **8**, 1247–1257.
16. Fu, Z., Wang, M., Potter, D., Miziorko, H. M. & Kim, J. (2002). The structure of a binary complex between a mammalian mevalonate kinase and ATP: insights into the reaction mechanism and inherited disease. *J. Biol. Chem.* **277**, 18134–18142.
17. Yang, D., Shipma, L. W., Roessner, C. A., Scott, A. I. & Sacchettini, J. C. (2002). Structure of the *Methanococcus jannaschii* mevalonate kinase, a member of the GHMP kinase superfamily. *J. Biol. Chem.* **277**, 9462–9467.
18. Bonanno, J. B., Edo, C., Eswar, N., Pieper, U., Romanowski, M. J., Ilyin, V. *et al.* (2001). Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis. *Proc. Natl Acad. Sci. USA*, **98**, 12896–12901.
19. Romanowski, M. J., Bonanno, J. B. & Burley, S. K. (2002). Crystal structure of the *Streptococcus pneumoniae* phosphomevalonate kinase, a member of the GHMP kinase superfamily. *Proteins: Struct. Funct. Genet.* **47**, 568–571.
20. Krishna, S. S., Zhou, T., Daugherty, M., Osterman, A. & Zhang, H. (2001). Structural basis for the catalysis and substrate specificity of homoserine kinase. *Biochemistry,* **40**, 10810–10818.
21. Thoden, J. B. & Holden, H. M. (2003). Molecular structure of galactokinase. *J. Biol. Chem.* **278**, 33305–33311.
22. Jones, S. & Thornton, J. M. (1995). Protein–protein interactions: a review of protein dimer structures. *Prog. Biophys. Mol. Biol.* **63**, 31–65.
23. De Geus, D., Hartley, A. P., Sedelnikova, S. E., Glynn, S. E., Baker, P. J., Verhees, C. *et al.* (2003). Cloning, purification, crystallization and preliminary crystallographic analysis of galactokinase from *Pyrococcus furiosus*. *Acta Crystallog. sect. D*, **59**, 1819–1821.
24. Verhees, C. H., Koot, D. G., Ettema, T. J., Dijkema, C., de Vos, W. M. & van der Oost, J. (2002). Biochemical adaptations of two sugar kinases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochem. J.* **366**, 121–127.
25. Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
26. Zenke, F. T., Engles, R., Vollenbroich, V., Meyer, J., Hollenberg, C. P. & Breunig, K. D. (1996). Activation of Gal4p by galactose-dependent interaction of galactokinase and Gal80p. *Science*, **272**, 1662–1665.
27. Timson, D. J., Ross, H. C. & Reece, R. J. (2002). Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry. *Biochem. J.* **363**, 515–520.
28. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
29. Sheldrick, G. M. (1998). SHELX: applications to macromolecules. In *Direct Methods for Solving Macromolecular Structures* (Fortier, S., ed.), pp. 401–411, Kluwer Academic, Dordrecht.
30. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters in isomorphous replacement and anomalous scattering. In *Proceeding of the CCP4 Study Weekend* (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), pp. 80–86, SERC Daresbury Laboratory, Warrington, UK.
31. Cowtan, K. (1994). DM: an automated procedure for phase improvement by density modification. *Joint*

CCP4 ESF-EACBM Newsletter Protein Crystallog. **31**, 34–38.
32. Roussel, A. & Cambillau, C. (1991). *Silicon Graphics Geometry Partners Directory 86*, Silicon Graphics, Mountain View, CA.
33. Murshodov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
34. Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Proteins Eng.* **6**, 37–40.
35. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
36. Merrit, E. A. & Bacon, D. J. (1994). Raster3D: photo-realistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
37. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallog. sect. D*, **55**, 938–940.

Edited by D. Rees

(Received 15 August 2003; received in revised form 16 January 2004; accepted 24 January 2004)