The structure of SAICAR synthase: an enzyme in the *de novo* pathway of purine nucleotide biosynthesis

Vladimir M Levdikov[1*], Vladimir V Barynin[1], Albina I Grebenko[1], William R Melik-Adamyan[1], Victor S Lamzin[2] and Keith S Wilson[2,3]

Background: The biosynthesis of key metabolic components is of major interest to biologists. Studies of *de novo* purine synthesis are aimed at obtaining a deeper understanding of this central pathway and the development of effective chemotherapeutic agents. Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase catalyses the seventh step out of ten in the biosynthesis of purine nucleotides. To date, only one structure of an enzyme involved in purine biosynthesis has been reported: adenylosuccinate synthetase, which catalyses the first committed step in the synthesis of AMP from IMP.

Results: We report the first three-dimensional structure of a SAICAR synthase, from *Saccharomyces cerevisiae*. It is a monomer with three domains. The first two domains consist of antiparallel β sheets and the third is composed of two α helices. There is a long deep cleft made up of residues from all three domains. Comparison of SAICAR synthases by alignment of their sequences reveals a number of conserved residues, mostly located in the cleft. The presence of two sulphate ions bound in the cleft, the structure of SAICAR synthase in complex with ATP and a comparison of this structure with that of other ATP-dependent proteins point to the interdomain cleft as the location of the active site.

Conclusions: The topology of the first domain of SAICAR synthase resembles that of the N-terminal domain of proteins belonging to the cyclic AMP-dependent protein kinase family. The fold of the second domain is similar to that of members of the D-alanine:D-alanine ligase family. Together these enzymes form a new superfamily of mononucleotide-binding domains. There appears to be no other enzyme, however, which is composed of the same combination of three domains, with the individual topologies found in SAICAR synthase.

Addresses: [1]Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, Moscow 117333, Russia, [2]European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany and [3]Department of Chemistry, University of York, York YO1 5DD, UK.

*Corresponding author.
E-mail: Vladimir@sunchem.chem.uga.edu

Key words: ATP-binding protein, crystal structure, phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase, purine biosynthesis

Received: **19 November 1997**
Revisions requested: **22 December 1997**
Revisions received: **19 January 1998**
Accepted: **21 January 1998**

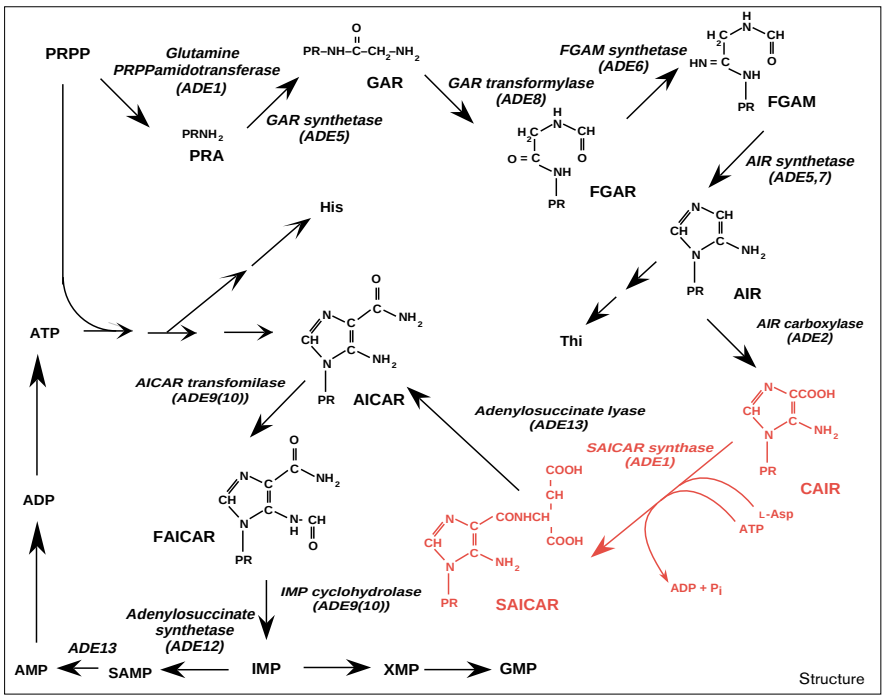
Structure 15 March 1998, **6**:363–376
http://biomednet.com/elecref/0969212600600363

Introduction

There are ten known enzymatic steps, in the *de novo* purine nucleotide pathway from 5-phosphoribosyl 1-pyrophosphate (PRPP) to the inosine monophosphate (IMP) branch point (Figure 1; [1]). IMP is subsequently converted in two further steps either to adenosine monophosphate (AMP) or to guanosine monophosphate (GMP). Purine nucleotide biosynthesis is integrated with the pathways for histidine and thiamine synthesis. Progress in understanding *de novo* purine nucleotide biosynthesis and in particular its regulation, together with the cloning of the genes coding for the enzymes catalysing most of steps in the pathway, has been reviewed [2]. Only one crystal structure of an enzyme involved in purine biosynthesis is known, that of adenylosuccinate synthetase from *Escherichia coli*, determined at 2.8 Å resolution [3], which catalyses the first committed step in the synthesis of AMP from IMP.

SAICAR (phosphoribosylaminoimidazolesuccinocarboxamide) synthase (EC 6.3.2.6) catalyses the seventh step of purine nucleotide synthesis [4]. The scheme of the reaction of phosphoribosylcarboxyaminoimidazole (CAIR) and aspartic acid in the presence of adenosine triphosphate (ATP) and magnesium ions leading to formation of SAICAR, is also shown in Figure 1. The *ADE1* gene, which codes for yeast SAICAR synthase, has been cloned and overexpressed, and the deduced amino acid sequence consists of 306 residues [5]. The overexpression facilitated structural studies of the enzyme and provides the possibility of using site-directed mutagenesis for probing the active site in the future. Biochemical studies indicate that the optimal reaction conditions for SAICAR synthase are pH 8.0 and a magnesium chloride concentration about 5 mM, while a concentration of more than 30 mM inhibits activity. The Michaelis constants for the three substrates of the reaction are 1.6 μM for CAIR, 14 μM for ATP and

Figure 1



Schematic representation of the biosynthesis of purine nucleotides in *S. cerevisiae*. For each step, the name of the enzyme and the structural gene (in parentheses) are given. The reaction catalysed by SAICAR synthase is shown in red. Abbreviations: AICAR, phosphoribosylaminoimidazolecarboxamide; AIR, phosphoribosylaminoimidazole; CAIR, phosphoribosylcarboxyaminoimidazole; FAICAR, phosphoribosylformylamino-imidazolecarboxamide; FGAM, phospho-ribosylformylglycinamidine; FGAR, phosphoribosylformylglycinamide; GAR, phosphoribosylglycinamide; PR, 5-phosphoribosyl; PRA, phosphoribosylamine; PRPP, 5-phosphoribosyl 1-pyrophosphate.

960 μM for aspartic acid. Guanosine triphosphate (GTP) and 2′-dATP can substitute for ATP in the reaction, while cytidine triphosphate (CTP) and uridine triphosphate inhibit (UTP) [6].

The substrate specificity of SAICAR synthase towards a set of carboxyaminoimidazole ribotide analogues with modifications in the imidazole ring, ribose and phosphate moieties, as well as towards aspartic acid analogues, has been studied biochemically. It was shown that the phosphate group of the CAIR molecule is necessary for enzymatic reaction, but in contrast the double charge on neither the phosphate group, the 2′- and 3′-hydroxyl groups on the ribose nor the amino group in the imidazole ring of the CAIR molecule are absolutely required for activity [7].

Here we present the crystal structure of SAICAR synthase from the yeast *Saccharomyces cerevisiae*. To date, the amino acid sequences of eleven SAICAR synthases from various prokaryotic and eukaryotic organisms have been elucidated [5,8–18]. Six of these have previously been compared, revealing that 10% of the residues appeared to be fully conserved [7]. Comparison of eleven sequences, using the knowledge of the 3D structure presented here, reveals the location of the conserved residues and leads to hypotheses for their role in folding or in ATP and CAIR binding. The structure of the SAICAR synthase–ATP complex confirms the position of the active site and indicates that the adenine base of ATP forms H-bonds mostly to the mainchain N atoms and that sidechains (except Glu219) are not involved. This explains why the hydrophobic ATP-binding site is not made up of conserved residues and could not be identified simply by analysis of their location in the native structure alone.

The 3D structure of SAICAR synthase is compared with those of other known nucleotide-binding proteins with the aim of finding common features. Previous analyses of the structures of nucleotide-binding proteins, on the basis of chain fold similarities and amino acid sequence patterns that correlate with nucleotide-binding activity, indicated several families [19–24]. The ATP-binding fold of SAICAR synthase has some features in common with the ATPase fragment of heat-shock protein HSC70 [22,25] and actin [26]. SAICAR synthase also has structural relations to cyclic AMP-dependent protein kinase (cAPK; [27]), glutathione synthetase (GSHase; [28]), D-alanine: D-alanine ligase (DD ligase; [29]) and biotin carboxylase [30]. Topological similarity between the SAICAR synthase N-terminal domain and proteins from the cAPK family, as well as the SAICAR synthase central domain and proteins from the GSHase family, suggests that SAICAR synthase combines the cAPK and GSHase subgroups to form one divergent family.

Results and discussion

Accuracy of the final model

The structure of SAICAR synthase was refined to an R factor of 0.153 at 1.9 Å resolution, see Table 1. The final model consists of 2383 protein atoms, 319 water molecules

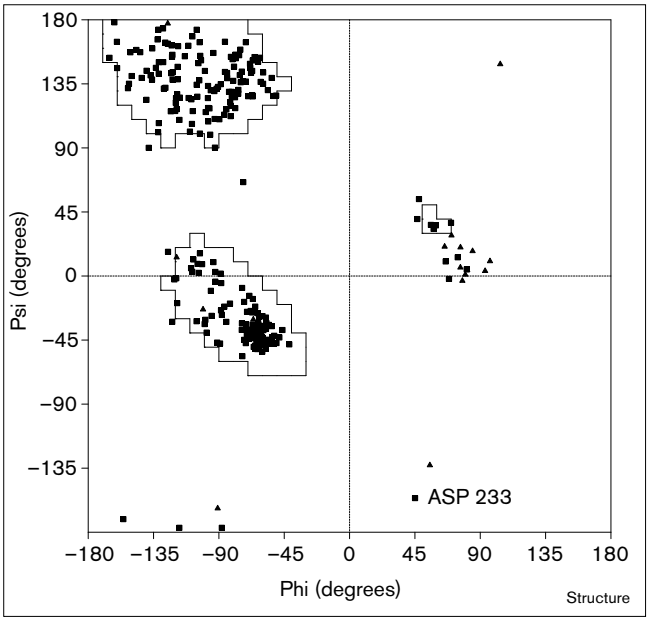
Table 1

Summary of restrained refinement.

	Native SAICAR synthase	SAICAR synthase in in complex with ATP
Resolution (Å)	10.0–1.9	10.0–2.05
Number of reflections	24,879	23,870
R factor*	0.153	0.177
R_{free} †	–	0.226
No. protein atoms	2,383	2,418
No. water molecules	319	322
No. Sulphate anions	2	2
No. ATP molecules	0	2
Average atomic displacement parameters (Å^2)		
protein	20.7	21.0
domain A	17.1	17.8
domain B	22.4	23.2
domain C	24.3	24.1
water molecules	41.3	42.8
Rms deviation from ideality‡		
1–2 distances (Å)	0.020 (0.020)	0.017 (0.020)
1–3 distances (Å)	0.041 (0.030)	0.039 (0.040)
Deviation in isotropic atomic displacement parameters (Å^2)‡		
1–2 mainchain	2.2 (2.0)	3.2 (3.0)
1–3 mainchain	3.3 (3.0)	4.2 (5.0)
1–2 sidechain	5.4 (3.0)	7.8 (7.0)
1–3 sidechain	8.3 (4.0)	10.0 (9.0)

*R factor = $\Sigma \mid F_o - F_c \mid / \Sigma \mid F_o \mid$, where F_o and F_c are the observed and calculated structure factors. †4% of the reflections were used to calculate R_{free}. ‡Target is given in parentheses.

Figure 2



Ramachandran plot [32] of the mainchain dihedral angles for the final atomic model of SAICAR synthase. Glycine residues are indicated as triangles, all other residues are squares. The outlined areas indicate the preferred regions [33].

and two sulphate ions. The average atomic displacement parameter is 23.2 Å^2 for all atoms, 20.7 for protein atoms and 41.3 for solvent molecules. The maximum, minimum and root mean square (rms) values of the final difference density are 0.32, –0.30 and 0.06 el Å^{-3}, respectively. The rms coordinate error estimated from the σ_A plot [31] is 0.17 Å. The ($3F_o–2F_c$, α_c) density map contoured at 1σ above the mean shows continuous density for all mainchains atoms with two exceptions. The first residue that could not be seen in the electron density is the N-terminal residue Met1. This may be a result of post-translational modification and subsequent acetylation of Ser2 (VD Domkin, personal communication). Acetylated Ser2 fits well to the electron density and has been included in the final model. The second segment of chain with no visible density consists of seven residues, Lys164–His170, which form a solvent-exposed surface loop. The presence of several hydrophilic residues and Gly168 probably confer flexibility on this loop. The next two residues, Asp171 and Glu172, also have poorly defined density. This loop may be of functional importance, see below.

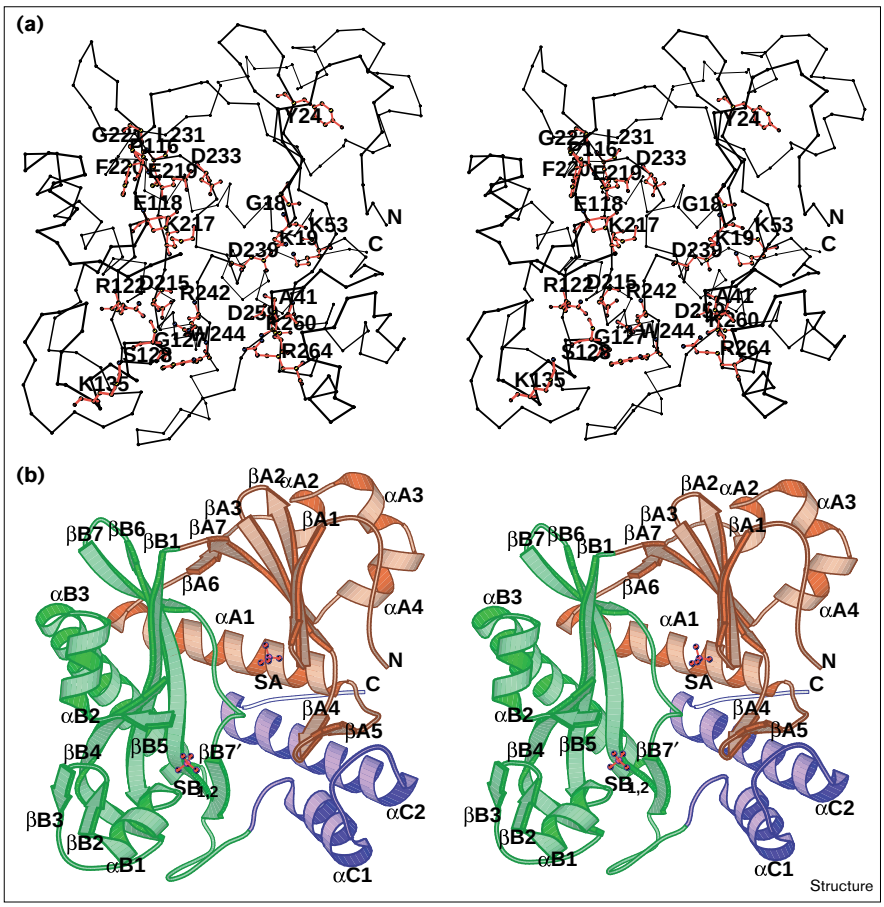
The Ramachandran plot of the mainchain dihedral angles [32] made using PROCHECK [33] is shown in Figure 2. Only one residue, Asp233, has an unusual, although generally allowed, mainchain conformation. Asp233 belongs to a cluster of fully conserved residues located near the ATP-binding site and is expected to play a role in catalysis, see below. Only one residue, Arg17, was modelled in two conformations, clearly indicated by the electron density. Both conformations form two equivalent H-bond networks with Glu181 of the same molecule and Asp22 of a symmetry-related molecule.

The variation of the average mainchain and sidechain atomic displacement parameters along the polypeptide chain show that most of the structure is well ordered and has low mobility. There are 27 residues for which at least one sidechain atom has a displacement parameter higher than 60 Å^2 and poorly defined electron density. Most of them are hydrophilic, located on the surface of the enzyme and exposed to solvent. The displacement parameters are high at the C terminus, at the residues adjacent to the invisible region Lys164–His170, connecting β strands βB4 and βB5, and in the loop Asp223–Glu228, connecting strands βB6 and βB7. Although the loop Lys164–His170 does not contain conserved residues, it becomes more rigid with well-defined electron density in the SAICAR synthase–ATP complex. The loop Asp223–Glu228 has no assigned functional role.

Polypeptide fold

SAICAR synthase has a globular shape with overall dimensions of about 70 × 60 × 40 Å^3, with three domains

Figure 3



Stereo views of SAICAR synthase. **(a)** Cα backbone trace. The conserved residues are in red and labelled. **(b)** Secondary structural elements. The β strands: βA1, residues 13–18; βA2, 20–27; βA3, 30–36; βA4, 39–41; βA5, 44–46; βA6, 75–77; βA7, 105–111; βB1, 113–125; βB2, 137–139; βB3, 142–144; βB4, 151–163; βB5, 172–174; βB6, 210–223; βB7, 227–232; βB7′, 242–246. The helices: αA1, 52–72; αA2, 84–88; αA3, 90–95; αA4, 95–103; αB1, 128–137; αB2, 175–184; αB3, 184–210; αC1, 260–271; αC2, 281–301. The three domains are in different colours: domain A, red; domain B, green; domain C, blue. The positions of the two sulphate ions, SA and SB, are shown.

referred to as A, B and C having a large deep cleft across the central part, see Figure 3. The N-terminal domain A contains 111 residues, Ser2–His112, domain B contains 143 residues, Leu113–Gln255, and the C-terminal domain is composed of 51 residues, Asp256–His306. Residues from all three domains line the cleft. The overall dimensions of the cleft are big enough to accommodate large substrates such as CAIR and ATP simultaneously. This cleft is the obvious candidate for the active site.

Domains A and B have (α+β) structures. The core of both domains is composed of an extensive five-stranded antiparallel right-hand twisted β sheet. Helices are located on one side of each β sheet. The third domain, C, consists simply of two α helices tightly associated with one another. They make contacts with helix αA1 from domain A. There are only three mainchain interdomain H-bonds: Leu231 N to Arg73 O (domains A and B), Tyr258 N to Phe243 O (domains B and C) and Glu47 N to Val278 O (domains A and C).

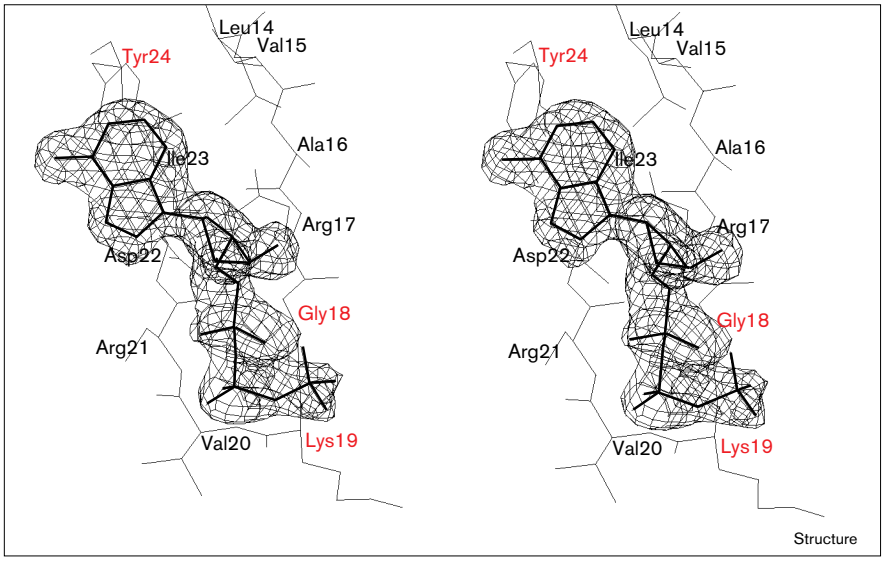
The secondary structural elements were defined as described previously [34]. All α helices in SAICAR synthase except αB2 have regions with 3_{10} H-bonds. β Turns were classified according to [35]. SAICAR synthase contains one type I β turn (237–240) which forms loop in the middle of strand B7, in fact, dividing this strand into two, βB7 and βB7′. This region (232–242) seems to be important for catalysis and Asp233 (the only outlier on the Ramachandran plot) probably interacts with Mg^{2+} ions in the active site. Three type I' β turns (8-11, 41–44, 139–142) lie within or adjacent to β sheets. There are four type II β turns (50–53, 80–83, 149–152, 272–275). Two of them connect strands and helices in domain A, one is in a loop between two strands in domain B and one connects two helices in domain C. Two π turns (206–211, 267–272), making an H-bond O_i to N_{i+5} [34], are located at the ends of helices αB3 and αC1.

Sulphate ions

In the last stages of refinement of native SAICAR synthase, the difference electron-density map clearly revealed two peaks, identified as two sulphate ions, SA and SB (see Figure 3b). The density corresponding to SA was weaker than that for SB. The SA site is located in the interdomain cleft near domain A, close to strands

Figure 4

Stereo view of the Fourier map for the enzyme–ATP complex superimposed onto the refined model, showing the region near the phosphate-binding loop, residues 14–24. The conserved residues are labelled in red. The Fourier coefficients are (F_o–F_c) with the model phases using data in the 10–2.05 Å range. The contour level is 1σ.



βA1 and βA2. One of the SA oxygen atoms makes two H-bonds, one to Arg21 N, the other to Val20. The occupancy for SA was not refined, but judging from the average displacement parameter (50.9 Å^2), the site is probably not fully occupied.

The second sulphate, SB, is located in the interdomain cleft near the largest domain, B, close to strands βB1, βB6, βB7′, helix αB1 and the flexible loop Lys164–Glu170 that connects strands βB4 and βB5. Judging from the directions of these strands, the loop extends over the ion shielding it from bulk solvent. The helix αB1 axis and the resulting dipole originating from the individual moments of the backbone peptide units [36,37] point towards SB. The positively charged helix N terminus interacts favourably with the negatively charged sulphate. SB was modelled in two mutually exclusive subsites, SB_1 and SB_2, which provided a better fit to the electron density. The occupancies of the subsites were assigned as 0.7 and 0.3, respectively, on the basis of the average atomic displacement parameter (33.5 Å^2 for SB_1 position and 35.0 for SB_2). Each subsite has well defined H-bonds to Arg122, Arg242, Ser128 and two water molecules. The two sulphate ions, SA and SB, are likely to bind in positions corresponding to the phosphate moieties of ATP and CAIR. The distance between SA and SB is about 17 Å.

The ATP-binding site

Determination of a series of structures of SAICAR synthase in complex with nucleotides and other ligands is in hand and detailed analysis of the mechanism awaits the results of this work. We plan to collect high resolution data once cryogenic conditions have been established. Here we briefly discuss preliminary experimental identification of the nucleotide-binding site from the structure of the enzyme–ATP complex at 2.05 Å, see Table 1. Refinement of the complex gives an R factor for the current model of 0.177 for data within the 10.0–2.05 Å resolution range. The ATP-binding site is shown in Figure 4. ATP lies along the phosphate-binding loop, in a manner similar to that seen in cAPK and DD ligase, described below, confirming that the SA sulphate ion in the native structure indeed lies in the place occupied by the β-phosphate of ATP in the complex.

In SAICAR synthase, the adenine base of ATP is in an *anti* conformation with respect to the ribose moiety. ATP adenine fits into the hydrophobic pocket formed by the two interdomain fragments (Leu114 and Val232, Leu31 and Phe33). Specificity and tightness of adenine binding is governed by a set of H-bonds: adenine N1 to His112 N, N6 to His100 O and N7 to Asp233 N. The fact that these H-bonds are formed by the mainchain atoms implies that the sidechains in this region do not have to be conserved. The ATP ribose moiety is in the C_3'-*endo* conformation, stabilised by H-bonds O2 and O3 to Lys164 NZ and O3 to Glu219 OE2. The ATP β-phosphate group interacts with the loop Val15–Tyr24. Residues Lys19, Val20 and Arg21, forming the edge of this loop, have their mainchain N atoms pointing towards the ATP phosphate group.

Seven nearby residues, comprising the loop Lys164–His170, are not defined in the electron density of native enzyme and presumably form a flexible loop. In the complex, there is well-defined density for these residues, suggesting that the loop serves as a lid when ATP is bound, to shield the reaction intermediate from solvent and prevent degradation.

As crystals of the complex of SAICAR synthase with ATP were grown in the presence of 10 mM magnesium chloride, we expected to find Mg^{2+} ions in the active site; however, analysis of the electron density, atomic displacement parameters and coordination sphere of solvent molecules in the active site suggested there were no metal ions present. There could be two possible reasons for this. The first is that the active-site residues have high mobility, as judged from their displacement parameters. The metal ions could well occupy the space filled by the alternative conformation of the active-site residue Arg21. The second is that the phosphate moiety of ATP competes for binding with sulphate ions present in the mother liquor. Indeed, the electron density corresponding to the ATP β-phosphate is somewhat higher than that for α- or γ-phosphates. This may suggest that we observe an overlapping of the densities of ATP and sulphate molecules, both not fully occupied. In either case, binding of Mg^{2+} ions would be weakened. This is clear that a crystal structure to a higher resolution and therefore to a higher level of detail is needed to address these questions.

The putative CAIR binding site

The most probable binding site for CAIR is at the location of the SB sulphate ion, which probably occupies a position corresponding to the phosphate moiety of CAIR. According to the mechanism which involves synthesis of the amide bond with simultaneous splitting of a phosphoanhydride bond of nucleoside triphosphate [4], the CAIR carboxyl group and γ-phosphate of ATP should be close to one other. The positions of these two terminal moieties of CAIR define its conformation. CAIR would be situated in the enzyme cleft, so that its carboxyl group oxygen involved in the reaction is positioned about 3 Å from the position of the phosphorous atom of the γ-phosphate of ATP.

These hypotheses on the mode of substrate binding require further experimental verification through studies of substrate–inhibitor complexes. Nevertheless, the evidence is extremely strong for the location of the active site.

Conserved residues

Two amino acid sequences have been published for SAICAR synthase from *S. cerevisiae* [5,38]. The only difference between them is at residue 185, which is glycine or glutamate in the two sequences. In the electron-density map, there is a significant peak corresponding to the Cβ atom of this residue and therefore residue 185 appears to be glutamine, with its carboxyl moiety disordered.

The 11 known sequences of SAICAR synthases from SwissProt [39] were aligned using the program CLUSTAL V [40]. The alignment shows the sequences are clustered into four groups: yeast, plants, bacteria and vertebrae, with different levels of identity within each group. Enzymes from *S. cerevisiae* [5,38], *Candida utilis* [8], *Candida albicans* [39], *Candida maltosa* [9,10] make up the first group with about 60% sequence identity. The second group contains SAICAR synthases from *Arabidopsis thaliana* [39] and *Vigna aconitifolia* [11] with 80% identical residues. Enzymes from *Streptococcus pneumoniae* [12], *Bacillus subtilis* [13] and *E. coli* [14–16] form the third group with 30% identity. The fourth group consists of the N-terminal domains of vertebrate SAICAR synthases from *Gallus gallus* [17] and *Homo sapiens* [18] with an identity of 80%. The 3D structure of SAICAR synthase coupled with the sequence alignment shows that the conserved residues are not randomly spread through the secondary structural elements. The secondary structural elements βA1, βA2, βA3, βA4, βA5, βA7, αB1, βB2, βB3, βB4, βB6, βB7, βB7′ and αC1 are more highly conserved than the others. These elements align best with other nucleotide-binding proteins, see below.

In spite of fairly high identity within the groups, alignment of all SAICAR synthase sequences results in only 24 fully conserved residues, corresponding to about 8% of the yeast enzyme. Eight conserved residues, Lys19, Lys53, Arg122, Lys135, Lys217, Arg242, Lys260 and Arg264, are positively charged. Some of these have been previously proposed to interact with negatively charged phosphate and carboxylate moieties of the substrates, to account for data on the substrate specificity [7]. Six conserved residues, Glu118, Asp215, Glu219, Asp233, Asp239 and Asp259, are negatively charged. As mentioned above, bivalent Mg^{2+} cations are required for enzymatic activity and some of these conserved negatively charged carboxylate groups probably participate in the chelation of metal ion in the SAICAR synthase–Mg·ATP complex.

The other fully conserved residues either play important roles in folding with glycine and proline residues — Gly18, Pro116, Gly127 and Gly221 — having special torsion angles, or form the hydrophobic core of the molecule — Ala41, Phe220, Leu231 and Trp244. The role of the conserved Tyr24 and Ser128 residues is not yet clear.

Almost all the fully conserved residues are located at the surface of the interdomain cleft, Figure 3a. They form four compact clusters in the 3D structure.

One cluster is Gly18 and Lys19, associated with the SA sulphate ion and the ATP-binding site. These residues are in a β-loop-β motif characteristic for the phosphate-binding loop in many other nucleotide-binding proteins, including actin, the ATPase fragment of heat-shock protein HSC70 and cyclic AMP-dependent protein kinase.

A second cluster consists of Leu231, Phe220, Lys217, Glu118, Pro116, Asp233 and Glu219. The sidechains of the last two residues point towards to the SA sulphate ion. The negatively charged groups of these conserved

residues presumably interact with the Mg^{2+} cations associated with ATP. As mentioned above, Asp233 has unusual dihedral angles and might be expected to play a significant role in structure or catalysis.

The third cluster consists of Arg122, Gly127, Ser128, Asp215, Arg242 and Trp244, situated in the native structure near to the second sulphate ion, SB, the putative CAIR-binding site. The N terminus of the αB1 helix, located nearby, may also be involved in binding the negatively charged CAIR phosphate group [36,37]. The phosphate moiety of CAIR could form H-bonds with Arg122, Arg242 and Ser128.

The fourth cluster is far removed from the sulphate ions, with residues involved in an H-bond network between all three domains. Asp239 belonging to domain B forms H-bonds to Lys53 from domain A and to Lys260 from domain C. Asp259, which forms H-bonds to Lys260, also belongs to this cluster. This cluster therefore plays a structural role, supporting mutual orientation of the domains in SAICAR synthase.

Relationship of SAICAR synthase to other nucleotide-binding proteins

A number of 3D structures of nucleotide-binding proteins are known and several attempts to classify them according to nucleotide binding chain fold have been made. One of the most recent classifications divides the nucleotide-binding proteins into four main families [23], hereafter called Families 1, 2, 3 and 4.

Family 1, with classic dinucleotide-binding fold, includes proteins that bind to NAD(P) or FAD. This fold is based on parallel βαβ supersecondary structural elements. A short phosphate-binding loop with the fingerprint sequence GXGXXG (where X represents any amino acid) is typically found in the loop connecting the first β strand and α helix.

Proteins from three other families bind mononucleotides. Family 2 includes proteins with the classic mononucleotide-binding fold and also has βαβ supersecondary structures with parallel β strands. These structures have longer phosphate- binding loops with the fingerprint sequence GXXGXGK, and form a positively charged site to accommodate the phosphates of the mononucleotide.

Family 3 includes a number of protein kinases, which bind ATP as the phosphoryl donor. Proteins of this family have a phosphate-binding loop with the same fingerprint sequence, GXGXXG, as the first family, but the loop connects two adjacent strands of an antiparallel β-pleated sheet.

Family 4 includes the ATPase fragment of heat-shock protein HSC70 [25] and actin [26]. The sugar kinases have also been proposed to be part of this family [41], but are less closely related and not discussed further here. The ATP-binding fold of these proteins is distinct from both the classical folds (Families 1 and 2) and the protein kinase fold (Family 3). Two N-terminal antiparallel β strands form a sharp turn which interacts with the β and γ phosphates of the nucleotide. According to Flaherty *et al.* [22], the fingerprint sequence around this contact is (I/L/V)X(I/L/V/C)DXG(T/S/G)(T/S/G)XX(R/K/C).

The structures of two related proteins, GSHase [28] and DD ligase [29], have been proposed to form a new family, Family 5, of mononucleotide-binding proteins [24], in addition to the four families mentioned above. A structural similarity between biotin carboxylase, GSHase and DD ligase has been discussed [42,30].

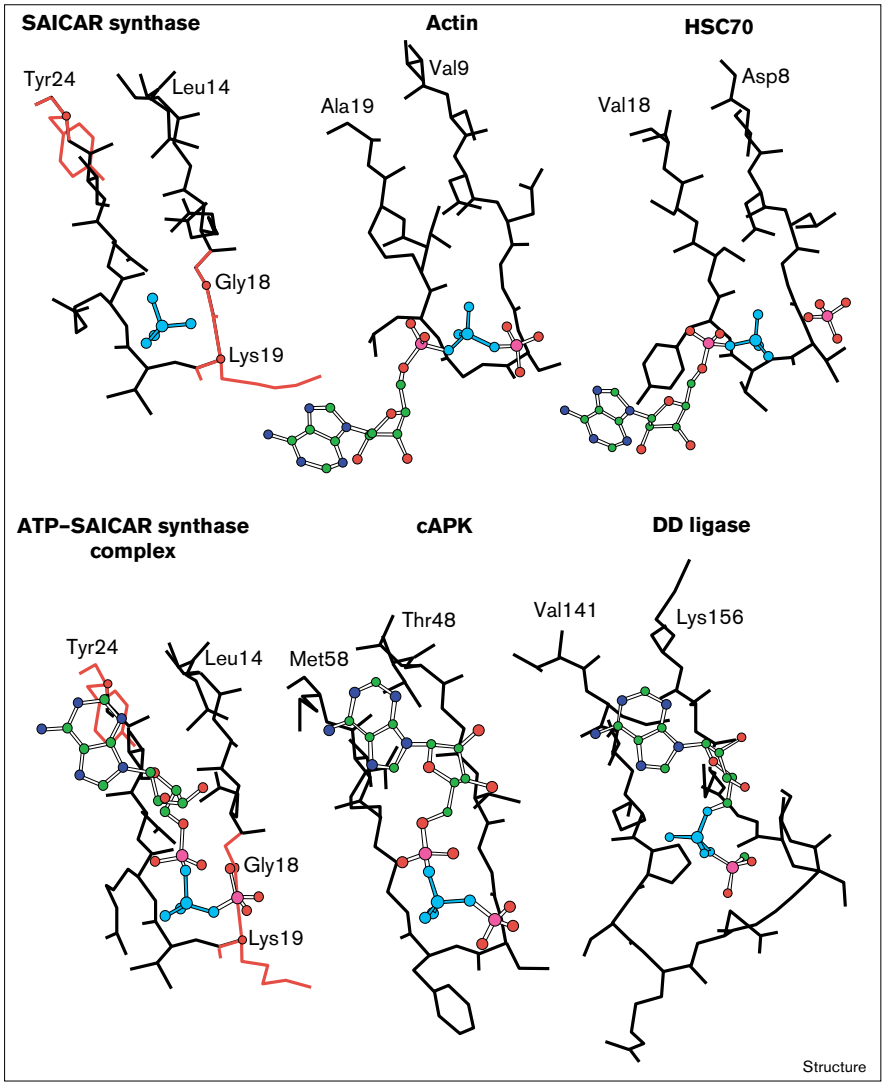
SAICAR synthase does not belong to Families 1 and 2, as it does not contain a classical Rossmann-type βαβ nucleotide-binding motif and has neither a GXGXXG nor a GXXGXGK fingerprint in the sequence. The enzyme does, however, have close relations to Families 3, 4 and 5 of mononucleotide-binding proteins which have two antiparallel β-strands in the phosphate–binding motifs. Here and below secondary structural elements of related proteins are named according to the original publications.

Comparison with actin and the ATPase fragment of a 70 kDa heat shock cognate protein (Family 4)

Actin and the ATPase fragment of a 70 kDa heat shock cognate protein (HSC70) have common characteristic β hairpins that form a β-loop-β phosphate-binding motif with a good 3D similarity [22], Figure 5. A similar loop but with a different sequence has been identified in SAICAR synthase and superimposes well on the actin and HSC70 loops. The rms difference in Cα P-loop positions in native SAICAR synthase and actin/HSC70 is 0.53 Å. In all three proteins, this loop is formed by the first two β strands from the N terminus. The topology of the nucleotide-binding domains in SAICAR synthase is quite different, however.

After spatial superposition of the loops, the position of the SA sulphate in native SAICAR synthase and the β-phosphate in the SAICAR synthase–ATP complex fits closely (rms around 1.0 Å) to the positions of the β-phosphates of ATP in actin and HSC70. In actin, the β-phosphate forms hydrogen bonds with the amide groups of Gly15 and Leu16, in HSC70 equivalent bonds are formed with Thr14 and Tyr15. Two equivalent H-bonds in SAICAR synthase are formed with the mainchain amide groups of the spatially equivalent residues Arg21 and Val20. These observations strongly suggest that: firstly, the location of three conserved residues in SAICAR synthases, Gly18, Lys19 and Tyr24, in this phosphate-binding loop strongly supports its functional importance for ATP binding; secondly, the location of β-phosphate group of ATP in the complex (the site of the SA sulphate in native SAICAR

Figure 5



Phosphate-binding loops (P loops) in nucleotide-binding proteins. SAICAR synthase (^{14}LVARGKVRDIY^{24}), actin (^{9}VCDNGSGLVKA^{19}), HSC70 (^{8}DLGTTYSCV^{18}), cAPK (^{48}TLGTGSFGRVM^{58}) and DD ligase (^{141}VIVKPSREGSSVGMSK^{156}). Superposition of the Cα atoms in the P loop of SAICAR synthase and actin, and of HSC70 and cAPK gives an rms deviation of around 0.5 Å, in both cases. The positions of the ATP/ADP substrate are shown in ball and stick representation in each structure, with the β-phosphate groups (and sulphate ions in native SAICAR synthase) in blue. Residues conserved among SAICAR synthase sequences are shown in red.

synthase) corresponds to the comparable site in actin and HSC70, indicating a clear relationship between SAICAR synthase and Family 4 nucleotide-binding proteins. There is, however, an important difference between the position of ATP in SAICAR synthase, which lies approximately along the loop, and in actin/HSC70, where ATP is perpendicular to the loop, Figure 5.

Comparison with cyclic AMP-dependent protein kinase (Family 3)

The catalytic subunit of cyclic AMP-dependent protein kinase (cAPK) has a bilobal structure with a deep cleft between the lobes. The small lobe, consisting mainly of the N-terminal region, is associated with nucleotide binding. The large lobe is dominated by helical structure with a β sheet at the domain interface; it is involved in peptide binding and catalysis [27].

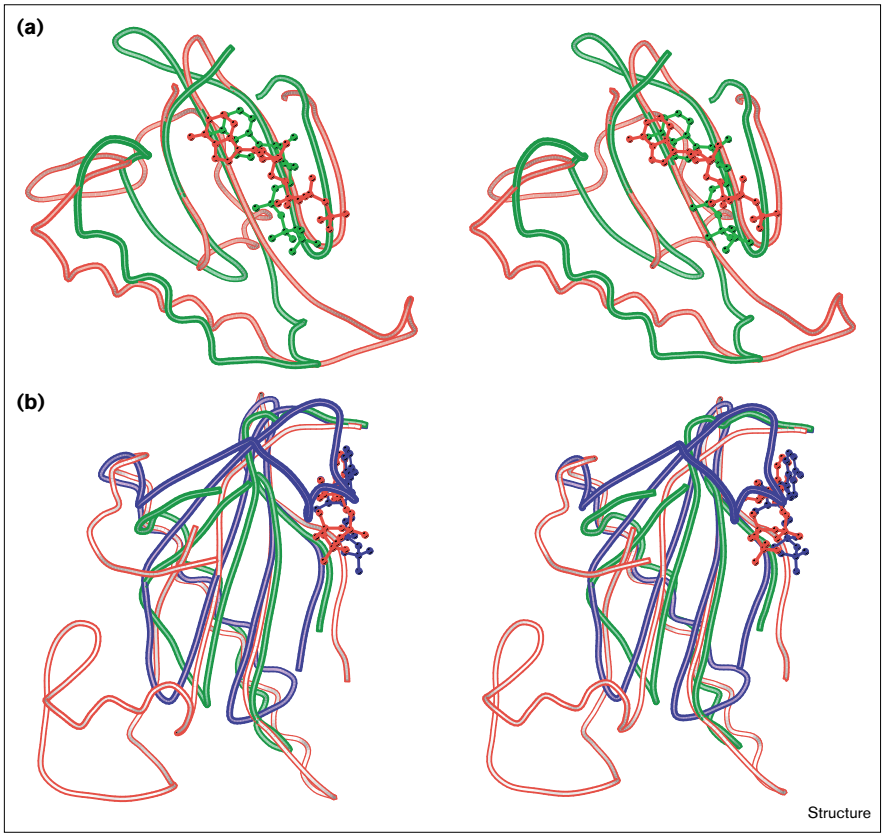
The small domain of cAPK, comprising five antiparallel β strands has a strong similarity to the A domain of SAICAR synthase , Figure 6a, with an rms deviation of 2.0 Å for 38 Cα atoms. The phosphate-binding loop in cAPK, connecting strands β1 and β2, corresponds to the sulphate-binding loop of SAICAR synthase, connecting strands βA1 and βA2, Figures 6 and 7. The chain topology is quite different, however (Figure 7). In particular, this concerns connections between strands in the β-sheet (βA3, βA4 and βA7 in SAICAR synthase). In addition, cAPK has an additional N-terminal helix which is absent in SAICAR synthase.

Comparison with glutathione synthetase and D-alanine:D-alanine ligase (Family 5)

Domain B in SAICAR synthase is topologically very similar to the C-terminal domains of GSHase and D-alanine:

Figure 6

Stereo views of the 3D superposition of SAICAR synthase domains onto those in other nucleotide-binding proteins. Bound nucleotides are shown in ball and stick representation in the same colour as the respective proteins. **(a)** The core of domain A of SAICAR synthase (red) superposed on that of the N-terminal small domain of cAPK (green). **(b)** The core of the B domain of SAICAR synthase (red) superposed on those of the C-terminal domains of GSHase (green) and DD ligase (blue).



D-alanine ligase (DD ligase), Figure 7. Superimposition of these structures onto the structure of SAICAR synthase, Figure 6b, results in rms deviation of 2.25 Å in 83 Cα atoms for GSHase and 2.25 Å in 99 Cα atoms for DD-ligase. There is a clear topological and spatial similarity between the loops 163–171 in SAICAR synthase and 225–242 in GSHase, Figure 7. SAICAR synthase domain A and the central domain in GSHase and DD ligase have a different number of β strands and significant differences in topology.

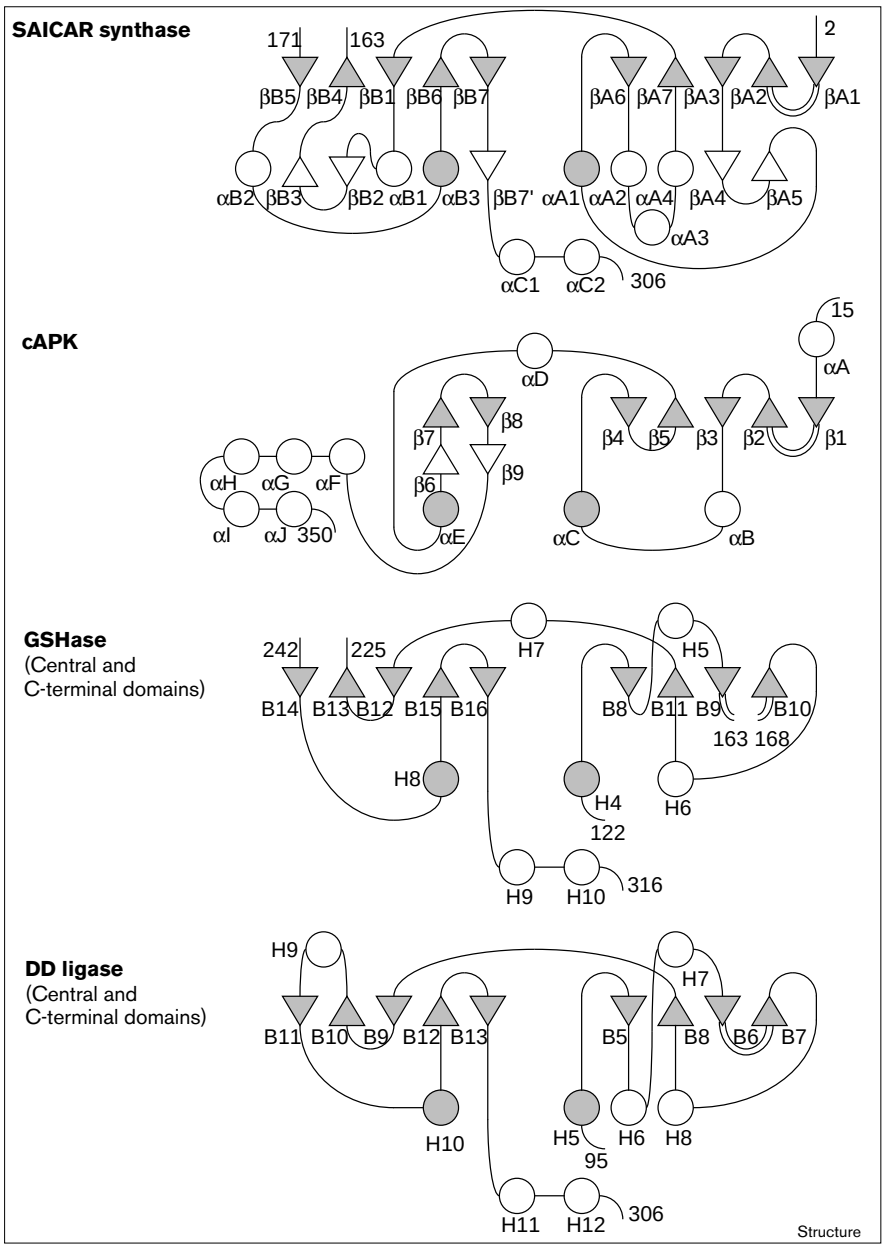
The nucleotide-binding loop is not visible in the structure of GSHase, but it is defined in DD ligase. These loops correspond to the ATP-binding loop in SAICAR synthase, although the DD ligase does not have the fingerprint sequence motif GXGXXG. The loop in DD ligase does not form a sharp bend as in SAICAR synthase, Figure 7, and results in a higher number of H-bonds to the β-phosphate group of ATP — six compared to three in SAICAR synthase. The comparison, given in Figure 5, indicates that the mode of ATP binding in SAICAR synthase and proteins of Family 5 is similar (in both the ATP molecule lies along the loop), although the similarity is not as great as that between SAICAR synthase and Family 3 proteins.

During the preparation of this manuscript a number of structures of the DD ligase and cAPK families have been determined and reviewed [30,42–46]. Close examination of these structures revealed a strong spatial similarity in a four-residue segment surrounding the adenine moiety [43]. These data provide additional support for the conclusions presented here with regard to the relationship between SAICAR synthase and enzymes from Families 3 and 5, as derived from the topological similarities in the fold of these structures.

Overview of structure–function relationships

We present the X-ray crystal structure of SAICAR synthase which catalyses the seventh step of purine nucleotide synthesis. The reaction involves the conversion of CAIR and aspartic acid, in the presence of ATP and magnesium ions, to SAICAR. The enzyme substrate specificity and the optimal reaction conditions are known from extensive biochemical studies. The model of the native enzyme has been refined at 1.9 Å resolution. Although the refinement of the enzyme–ATP structure at 2.05 Å is to be finalised, the experimental evidence of identification of the ATP-binding site in SAICAR synthase and its correspondence to the position of sulphate in the native structure is clear. Detailed analysis of the mechanisms awaits the

Figure 7



Topology diagrams for SAICAR synthase, cyclic AMP-dependent protein kinase (cAPK), glutathione synthase (GSHase) and D-alanine:D-alanine ligase (DD ligase). Circles denote helices, triangles denote β strands. Double lined connections indicate the phosphate-binding loop. The spatially equivalent secondary structural elements of the different proteins are shaded.

determination of structures at higher resolution and of additional ligand-bound complexes.

SAICAR synthase has three domains with a large deep cleft (putative active site) between them. In the native structure , there are two sulphate ions, one corresponding to the position of the β-phosphate group of ATP in the complex and another probably to the phosphate moiety of CAIR. As the reaction requires the carboxyl group of CAIR and the γ-phosphate group of ATP to be close to one other, a hypothetical mode of CAIR binding has been suggested. The flexible loop, which is not visible in the native structure but has been identified in the SAICAR synthase–ATP complex, is expected to shield the reaction intermediate from solvent. This may be essential for transfer of γ-phosphate from ATP to CAIR.

The sequence alignment of SAICAR synthases from different sources shows clustering into four groups: yeast, plants, bacteria and vertebrae, with different levels of sequence identity within the groups. Most of the 24 fully conserved residues are located in the active-site cleft. Eight positively charged residues supposedly interact with substrate phosphate groups, while six negatively charged

ones are probably important for chelation of the Mg^{2+} ions required for catalysis.

Comparison of SAICAR synthase with several nucleotide-binding proteins, belonging to five different families resulted in the following conclusions.

SAICAR synthase has some relation to those families of mononucleotide-binding proteins, which do not contain a classical Rossmann-type βαβ nucleotide-binding motif.

The ATP-binding loop in SAICAR synthase, the sulphate ion binding site in the native structure and the position of the ATP β-phosphate are comparable to those in proteins of Families 3, 4 and, to some extent, Family 5. The position of the whole ATP moiety in SAICAR synthase relates to Family 3 and to a lesser extent to Family 5, but not to Family 4.

SAICAR synthase domain A has strong topological and spatial similarity to the small domain of AMP-dependent protein kinase (Family 3). Domain B has strong similarity to GSHase and DD ligase, proteins from Family 5. This suggests that SAICAR synthase together with the enzymes in Families 3 and 5 arose from a common ancestor.

It seems likely that the fingerprint sequence in the mononucleotide-binding loops is required for their spatial organisation. The major conserved residues are glycines (which is why the loop is often called glycine-rich). The conformation of the residues in the loop is such as to orient mainchain N atoms towards the negatively charged β-phosphate group of ATP. Proteins from all the above families possess this feature, with differences in the number of H-bonds to the β-phosphate.

To provide a fuller understanding of catalysis by SAICAR synthase, it will be necessary to carry out high resolution X-ray analyses of complexes with additional substrates or analogues and to use site-directed mutagenesis to probe the active site. Detailed knowledge of the spatial structure of the enzyme and the residues involved in catalysis are essential for design of inhibitors for this step of purine biosynthesis and creation of effective chemotherapeutic agents.

Biological implications

The biosynthesis of key metabolic components is of central interest to biologists. Studies of *de novo* purine synthesis have been directed towards a deeper understanding of the mechanism and evolution of this vital pathway and development of effective chemotherapeutic agents. SAICAR synthase catalyses the seventh out of ten steps in the biosynthesis of purine nucleotides. The metabolic role of SAICAR synthase in the purine synthesis pathway is the utilisation of the CAIR product of the phosphoribosylaminoimidazolecarboxylase reaction to produce SAICAR. This work provides the first structure of a SAICAR synthase.

Analysis of the position of the two bound sulphate ions, consideration of the conserved residues and the position of ATP, in the SAICAR synthase–ATP complex, indicate the location of the nucleotide-binding site. The likely position of the substrate CAIR-binding site is suggested to be in the interdomain cleft, which probably also harbours the residues involved in catalysis. These findings have allowed the comparison of SAICAR with other nucleotide-binding enzymes.

The essential features of SAICAR synthase include: two nucleotide-binding domains with a characteristic fold; three domains forming the large wide substrate-binding cleft in the middle of the molecule; an ATP phosphate anchor made up of a β-loop-β motif, typical for actin, heat-shock cognate protein and cAMP-dependent protein kinase; and a large domain presumed to be associated with CAIR binding, with a fold also similar to that of these three enzymes. SAICAR synthase has a flexible loop extending over the active-site cleft, topologically equivalent to those in some other nucleotide-binding enzymes and presumed to shield the reaction intermediate from solvent. The structure suggests two previously identified nucleotide-binding protein families, which include enzymes such as glutathione synthase and cAMP-dependent protein kinase, share a common ancestor with SAICAR synthase. The structure of SAICAR synthase has increased our knowledge of the mechanism, evolution and characteristic features of this set of enzymes.

Materials and methods

Enzyme purification

Protein samples containing SAICAR synthase with about 70% purity were kindly provided by AN Myasnikov (Biological Institute, University of St. Petersburg, Russia, where the SAICAR-synthase-producing strain GRF18[pJDB(MSADIR)] [6,47] was created). This strain accumulates SAICAR synthase at a level of about 40% of total soluble protein growing in low P_i medium. The enzyme was purified following the procedure described previously [47,48]. In brief, yeast cells grown in 2% peptone containing 2% glucose were collected, washed with water and broken by shaking with glass beads in two volumes 0.2% SDS containing 2 mM phenylmethylsulphonyl fluoride. Debris was removed by centrifugation at 10,000 g for 30 min and the supernatant was clarified by centrifugation at 30,000 g for 60 min. Solid ammonium sulphate was added to 45% saturation. The precipitate was discarded and the supernatant was brought to 60% saturation. The resulting precipitate was collected by centrifugation and dissolved in 20 mM Tris-HCl buffer, pH 8.0. The protein solution was dialysed against the same buffer. DEAE-Sephadex was added to the solution and the protein eluted in a batchwise fashion with the same buffer containing 0.2 M sodium chloride. The eluted protein was precipitated with ammonium sulphate (65% saturation) and contained SAICAR synthase and about 30% contaminating proteins.

The precipitate containing crystals of SAICAR synthase together with amorphous protein was collected by centrifugation and the protein was extracted using a 1.6 M solution of ammonium sulphate with 20 mM

Table 2

Data collection, processing and phasing.

Crystals	Native enzyme	Derivative $Cs_3UO_2(CN)_5$	Derivative $KAu(CN)_2$	Derivative $KAu(CN)_2$	Enzyme–ATP Complex
Source	Synchrotron	Synchrotron	Synchrotron	CuKα	Synchrotron
Wavelength (Å)	1.01	1.01	1.01	1.54	0.901
Resolution (Å)	1.9	2.5	2.6	2.5	2.0
Data redundancy	2.0	2.0	1.5	3.6	4.1
Overall completeness (%)	97	99	66	92	98
Overall R_{merge}(I) (%)	5.1	5.8	6.0	5.3	7.8
R_c*		0.16	0.13	0.16	
R_s†		0.82	0.66	0.82	
Phasing power (acentric)‡		1.1	2.0	1.7	
Phasing power (centric)		0.8	1.3	1.1	

*$R_c = \Sigma\,|F_{Po}-F_{PHo}| / \Sigma|F_{Po}|$. †$R_s = \Sigma|F_{PHo}-F_{PHc}| / \Sigma|F_{PHo}-F_{Po}|$. F_{Po} and F_{PHo} are the observed structure factors for the native and heavy-atom derivatives, respectively; F_{PHc} is the calculated structure factor for a derivative; ‡Phasing power = f_h/E, where f_h is the rms contribution of the heavy atom and E is the rms lack of closure error.

magnesium chloride, 50 mM Tris-HCl, pH 7.5. Undissolved material was removed by centrifugation at 30,000 g for 30 min and the supernatant (150 mg total protein) was applied to a column (1.6 × 20 cm) of Butyl-Toyopearl 650S (Tosoh Corporation, Tokyo), pre-equilibrated with 1.6 M ammonium sulphate in magnesium chloride/Tris-HCl solution. The column was developed with 400 ml of an inverse gradient (1.5 to 1.0 M ammonium sulphate with magnesium chloride/Tris-HCl buffer), followed by 60 ml of the buffer and 60 ml of water. Fractions of SAICAR synthase were pooled and brought to 75% saturation with ammonium sulphate. The resulting protein was at least 99% pure and was used for crystallisation.

Crystallisation

Crystals were grown by vapour diffusion [49] at room temperature as described in [48]. The hanging drops contained protein at a concentration of 8–15 mg/ml, in 50 mM Tris-HCl buffer, pH 7.0, 40 mM aspartic acid and 1.0–1.25 M ammonium sulphate. The wells contained the same buffer plus 2.25 M ammonium sulphate. Crystals grew up to 1.2 mm in 4–7 days. The crystals belong to space group $P2_12_12_1$ with unit cell dimensions a = 62.6, b = 63.7 and c = 81.2 Å. There is one molecule in the asymmetric unit. The solvent content is about 45%. Heavy-atom derivatives $Cs_3UO_2(CNS)_5$ and $KAu(CN)_2$ were obtained by soaking native protein crystals in solutions of heavy-atom compounds in the above precipitating solution at concentrations of 2.5–5.0 mM for 10–45 days. Crystals of the SAICAR synthase–ATP complex with ATP were obtained by co-crystallisation of native SAICAR synthase with 50 mM ATP and 0.01 M $MgCl_2$.

X-ray data collection

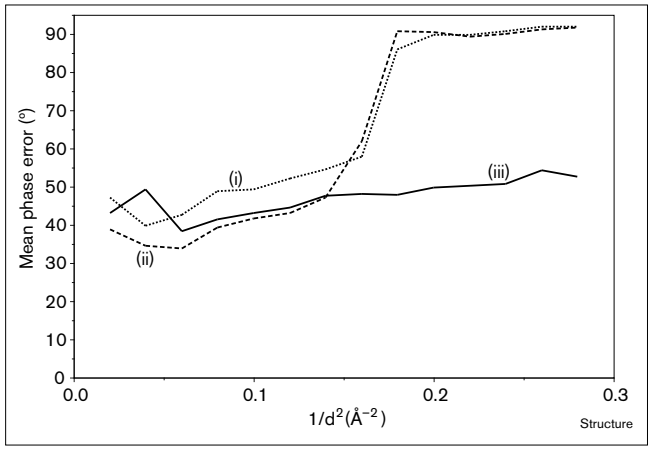
X-ray diffraction data from single crystals of native protein to 1.9 Å resolution, two isomorphous heavy-atom derivatives to 2.5 Å and the complex with ATP to 2.05 Å were collected using synchrotron radiation and a MAR Research imaging plate scanner at the EMBL Hamburg outstation, Table 2. The data were processed using the DENZO program package [50]. A further set for the $KAu(CN)_2$ derivative was collected to 2.5 Å resolution on a KARD-4 area detector diffractometer (Institute of Crystallography, Moscow).

Phasing

Initial phases for the protein structure factors at 2.5 Å resolution were derived from multiple isomorphous replacement (MIR). The heavy-atom parameters were refined jointly taking into account anomalous dispersion using the MLPHARE program [51] from the CCP4 suite [52]. The phases obtained were iteratively used for finding minor sites in difference Fourier syntheses. The MIR phases had a mean figure of merit of 0.59. Statistics for the heavy-atom refinement are given in Table 2. The MIR phases gave a mean phase error of 42° (at 2.5 Å resolution) compared to those calculated from the refined model. The MIR map had 58% correlation to the final map.

MIR phases were improved using solvent flattening and histogram matching using the SQUASH program [53]. This improved the figure of merit to 0.71, the mean phase error to 37° at 2.5 Å resolution and the map correlation to 69%. The phases were further extended and improved by an automated refinement procedure, ARP [54]. 2,800 equal atoms were put into the SQUASH map, keeping the distance between neighbouring atoms in the range 1.2–1.6 Å. 80 cycles of unrestrained least squares minimisation in the whole range of resolution followed by updating the model were run. Up to 50 atoms were allowed to be removed and added every cycle. The resulting ARP model gave phases with a mean error to final phases of 49° in the whole resolution range 10.0–1.9 Å and the map correlation improved to 72%, Figure 8. The density map calculated from the ARP model revealed atomic features, which made it possible to trace about 90% of the polypeptide chain including sidechains. The model was built using FRODO [55]. The structure of the SAICAR synthase–ATP complex with ATP was solved by molecular replacement using the model of native SAICAR synthase.

Figure 8



Mean phase difference compared with the final phases, as determined by (i) MIR, (ii) SQUASH [53] and (iii) ARP [54].

Refinement

All X-ray data were used with no cut-off on amplitude. The refinement was carried out using stereochemically restrained least squares minimisation with the CCP4 version of PROLSQ [56], coupled with automated building and updating of the solvent structure using ARP. The refinement of the native structure was made in seven steps interspaced with manual rebuilding. The last cycles of refinement included the partial contribution from hydrogen atoms placed at their idealised positions, which resulted in improvement of the R factor by about 0.01. The final R factor is 0.153 for all data within the 10.0–1.9 Å resolution range. Preliminary refinement of the structure of complex with ATP was carried out using the CCP4 version of REFMAC [57]. The current R factor is 0.177 at 2.05 Å resolution. The results are summarised in Table 1.

Accession numbers

The atomic coordinates and corresponding structure factors of SAICAR synthase have been deposited in the Brookhaven Protein Data Bank [58] with accession numbers 1A48 and R1A48SF, respectively.

Acknowledgements

We thank AN Myasnikov for kindly providing protein samples, AV Teplyakov for help in data collection and AG Murzin for fruitful discussions. The research was made possible in part by grants No. MFW000 and No. MFW300 from the International Science Foundation, grant No. 96-04-48897 from the Russian Foundation for Basic Research and an EMBO fellowship to VML.

References

1. Miller, R.W. & Buchanan, J.M. (1962). Biosynthesis of the purines: XXVII. *N*-(5-amino-1-ribosyl-4-imidazolylcarbonyl)-L-aspartic acid 5′-phosphate kinosynthetase *J. Biol. Chem.* **237**, 458-490.
2. Zalkin, H. & Dixon, J.E. (1992). *De Novo* purine nucleotide biosynthesis. *Prog. Nucl. Acid Res. Mol. Biol.* **42**, 259-287.
3. Poland, B.W., *et al*., & Honzatko, R.B. (1993) Crystal structure of adenylosuccinate synthetase from *Escherichia coli. J. Biol. Chem.* **268**, 25334-25342.
4. Lukens, L.N. & Buchanan, J.M. (1959). Biosynthesis of the purines: XXIII. The enzymatic synthesis of *N*-(5-amino-1-ribosyl-4-imidazolylcarbonyl)-L-aspartic acid 5′-phosphate. *J. Biol. Chem.* **234**, 1791-1798.
5. Myasnikov, A.N., Plavnik, Y.A., Sasnauskas, K.V., Gedminene, G.K., Janulaitis, A.A. & Smirnov, M.N. (1986). Nucleotide sequence of *ADE1* gene of yeast *Saccharomyces cerevisiae*. *Bioorg. Khimia (USSR)* **12**, 555-558.
6. Ostanin, K.V., Alenin, V.V., Domkin, V.D. & Smirnov, M.N. (1989). Isolation and properties of phosphoribosyl-succinocarboxamide-aminoimidazole synthetase from the yeast *Saccharomyces cerevisiae*. *Biokhimiya (USSR)* **54**, 1265-1273.
7. Alenin, V.V., Ostanin, K.V., Kostikova, T.R., Domkin, V.D., Zubova, V.A. & Smirnov, M.N. (1992). Substrate specificity of Phosphoribosyl-aminoimidazole-succinocarboxamide-synthetase (SAICAR-synthetase) of the yeast *Saccharomyces cerevisiae*. *Biokhimiya (USSR)* **57**, 845-855.
8. Nishiya, Y. (1994). Primary structure of the *ADE1* gene from *Candida utilis*. *Biosci. Biotechnol. Biochem.* **58**, 208-210.
9. Sasnauskas, K., Jomantiene, R., Geneviciute, E., Januska, A. & Lebedys, J. (1991). Molecular cloning of the *Candida maltosa ADE1* gene. *Gene* **107**, 161-164.
10. Kawai, S., Hikiji, T., Murao, S., Takagi, M. & Yano, K. (1991). Isolation and sequencing of a gene, *C-ADE1,* and its use for a host-vector system in *Candida maltosa* with two genetic markers. *Agric. Biol. Chem.* **55**, 59-65.
11. Chapman, K.A., Delauney, A.J., Kim, J.H. & Verma, D.P. (1994). Structural organization of *de novo* purine biosynthesis enzymes in plants: 5-aminoimidazole ribonucleotide carboxylase and 5-aminoimidazole-4-N-succinocarboxamide ribonucleotide synthetase cDNAs from *Vigna aconitifolia*. *Plant Mol. Biol.* **24**, 389-395.
12. Hui, F.M. & Morrison, D.A. (1993). Identification of a purC gene from *Streptococcus pneumoniae*. *J. Bacteriol.* **175**, 6364-6367.
13. Ebbole, D.J. & Zalkin, H. (1987). Cloning and characterization of a 12-gene cluster from *Bacillus subtilis* encoding nine enzymes for *de novo* purine nucleotide synthesis. *J. Biol. Chem.* **262**, 8274-8287.
14. Tiedemann, A.A., Demarini, D.J., Parker, J. & Smith J.M. (1990). DNA Sequence of the *pur*C gene encoding 5′-phosphoribosyl-5-aminoimidazole-4-*N*-succinocarboxamide synthetase and organization of the *dapA–purC* region of *Escherichia coli* K-12. *J. Bacteriol.* **172**, 6035-6041.
15. Bouvier, J., Pugsley, A.P. & Stragier, P. (1991). A gene for a new lipoprotein in the *dap*A–*pur*C interval of the *Escherichia coli* chromosome. *J. Bacteriol.* **173**, 5523-5531.
16. Meyer, E., Leonard, N.J., Bhat, B., Stubbe, J. & Smith J.M. (1992). Purification and characterization of the *pur*E, *pur*K and *pur*C gene products: identification of a previously unrecognized energy requirement in the purine biosynthetic pathway. *Biochemistry* **31**, 5022-5032.
17. Chen, Z., Dixon, J.E. & Zalkin H. (1990). Cloning of a chicken liver cDNA encoding 5-aminoimidazole ribonucleotide carboxylase and 5-aminoimidazole-4-N-succinocarboxamide ribonucleotide synthetase by functional complementation of *Escherichia coli pur* mutants. *Proc. Natl. Acad. Sci. USA.* **87**, 3097-3101.
18. Minet, M. & Lacroute, F. (1990). Cloning and sequencing of human cDNA coding for a multifunctional polypeptide of the purine pathway by complementation of the ADE2-101 mutant in *Saccharomyces cerevisiae*. *Curr. Genet.* **18**, 287-291.
19. Rossmann, M.G., Moras, D. & Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *J. Mol. Biol.* **250**, 194-199.
20. Rossmann, M.G., Liljas, A., Brändén, C.-I. & Banaszak, L.J. (1975). Evolutionary and structural relationships among dehydrogenases. In *The Enzymes. 3rd edition,* (Boyer, P.D., ed.) **11**A, pp. 61-102 Academic Press, New York.
21. Saraste, M., Sibbald, P.R. & Wittinghofer, A. (1990). The P-loop - a common motif in ATP- and GTP-binding proteins. *Trends. Biochem. Sci.* **15,** 430-434.
22. Flaherty, K.M., McKay, D.B., Kabsch, W. & Holmes, K.S. (1991). Similarity of the three-dimensional structure of actin and the ATPase fragment of a 70K heat-shock cognate protein. *Proc. Natl. Acad. Sci. USA* **88**, 5041-5045.
23. Schulz, G.E. (1992). Binding of nucleotides to proteins. *Curr. Opin. Struct. Biol.* **2**, 61-67.
24. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.
25. Flaherty, K.M., DeLuca-Flaherty, C. & McKay, D.B. (1990). Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* **346**, 623-628.
26. Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F. & Holmes, K.C. (1990). Atomic structure of the actin: DNase I complex. *Nature* **347**, 37-44.
27. Knighton, D.R., *et al*., & Sowadski, J.M. (1991). Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407-414.
28. Yamaguchi, H., *et al.,* & Katsube, Y. (1993). Three-dimensional structure of the glutathione synthetase from *Escherichia coli* B at 2.0 Å resolution. *J. Mol. Biol.* **229**, 1083-1100.
29. Fan, C., Moews, P.C., Walsh, C.T. & Knox, J.R. (1994). Vancomycin resistance: structure of D-alanine:D-alanine ligase at 2.5 Å resolution. *Science* **266**, 439-443.
30. Waldrop, G.L., Rayment, I. & Holden, H.M. (1994). Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* **33**, 10249-10256.
31. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* A **42**, 140-149.
32. Ramakrishnan, C. & Ramachandran, G.N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. *Biophys. J.* **5**, 909-903.
33. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.
34. Lamzin, V.S., Dauter, Z., Popov, A.O., Harutyunyan, E.H., & Wilson, K.S. (1994). High resolution structures of holo and apo formate dehydrogenase. *J. Mol. Biol.* **236**, 759-785.
35. Wilmot, C.M. & Thornton, J.M. (1990). β-Turns and their distortions: proposed new nomenclature. *Protein Eng.* **3**, 479-493.
36. Wada, A. (1976) The α-helix as an electric macro-dipole. *Adv. Biophys.* **9**, 1-63.
37. Hol, W.G.J., Van Duijnen, P.T. & Berendsen, H.J.C. (1978). The α-helix dipole and properties of proteins. *Nature* **273**, 443-446.
38. Davies, C.J. & Hutchison, C.A., III. (1991). A directed DNA sequencing strategy based upon Tn3 transposon mutagenesis: application to the *ADE1* locus on *Saccharomyces cerevisiae* chromosome I. *Nucleic Acids Res.* **19,** 5731-5738.

39. Bairoch, A. (1990). *SwissProt Protein Sequence Databank User Manual, Release 14.0,* EMBL Data Library, Heidelberg, FRG.
40. Higgins, D.G., Bleasby, A.J. & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* **8**, 189-191.
41. Bork, P., Sander, C. & Valencia, A. (1992). An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. Natl. Acad. Sci. USA* **89**, 7290-7294.
42. Artymiuk, P.J., Poirrette, A.R., Raice, D.W. & Willett, P. (1995). Biotin carboxylase comes into the fold. *Nat. Struct. Biol.* **3**, 128-132.
43. Kobayashi, N. & Go, N. (1997). ATP binding proteins with different folds share a common ATP-binding structural motif. *Nat. Struct. Biol.* **4**, 6-7.
44. Matsuda, K., Mizugichi, K., Nishioka, T., Kato, H., Go, N. & Oda, J. (1996). Crystal structure of glutathione synthetase at optimal pH: domain architecture and structural similarity with other proteins. *Protein Eng.* **9**, 1083-1092.
45. Walsh, C.T., Fisher, S.L., Park, I.S., Prahalad, M. & Wu, Z. (1996). Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chem. Bio*3, 21-28.
46. Fan, C., Moews, P.C., Shi, Y., Walsh, C.T. & Knox, J.R. (1995). A common fold for peptide synthetases cleaving ATP to ADP: glutathione synthetase and D-alanine:D-alanine ligase of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **92**, 1172-1176.
47. Myasnikov, A.N., Sasnauskas, K.V., Janulaitis, A.A. & Smirnov, M.N. (1991). The *Saccharomyces cerevisiae ADE1* gene: structure, overexpression and possible regulation by general amino acid control. *Gene* **107**, 143-147.
48. Grebenko, A.I., Levdikov, V.M., Barynin, V.V., Melik-Adamyan, W.R. & Myasnikov, A.N. (1992). Crystallization and preliminary X-ray investigation of phosphoribosylaminoimidazolesuccinocarboxamide synthase from the yeast *Saccharomyces cerevisiae*. *J. Mol. Biol.* **228**, 298-299.
49. Davies, D.R. & Segal, D.M. (1971). Protein crystallization: micro techniques involving vapour diffusion. In *Methods in Enzymology.* (Jakoby, W.B., ed.), **22**, pp. 266-269, Academic Press, New York and London.
50. Otwinowski, Z. & Minor, W. (1993). DENZO: a film processing program for macromolecular crystallography. Yale University, New Haven, CT, USA.
51. Otwinowski, Z. (1991). Isomorphous replacement and anomalous scattering. *In Proceedings of the CCP*4 *Study Weekend* (Wolf, W., Evans, P.R. & Leslie, G.W., eds.), pp. 80-86. SERC Daresbury Laboratory, Warrington, UK.
52. CCP4: Collaborative Computational Project, Number 4 (1994). The *CCP*4 suite: programs for protein crystallography. *Acta Cryst.* D **50**, 760-763.
53. Zhang, K.Y.J. & Main, P. (1990). Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Cryst.* A **46**, 41-46.
54. Lamzin, V.S. & Wilson, K.S. (1993). Automated refinement of protein models. *Acta Cryst.* D **49**, 129-147.
55. Jones, T.A. (1978) A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268-272.
56. Konnert, J.H. & Hendrickson, W.A. (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Cryst.* A **36**, 344-350.
57 Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* D **53**, 240-255.
58. Bernstein, F.C., *et al*., & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.