X-Ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the Escherichia coli purine biosynthetic pathway at 2.5 Å resolution

ARTICLE *in* STRUCTURE · OCTOBER 1999

Impact Factor: 5.62 · DOI: 10.1016/S0969-2126(99)80182-8 · Source: PubMed

X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution

Chenglong Li[1], T Joseph Kappock[2], JoAnne Stubbe[2], Todd M Weaver[1†] and Steven E Ealick[1*]

Background: The purine biosynthetic pathway in procaryotes enlists eleven enzymes, six of which use ATP. Enzymes 5 and 6 of this pathway, formylglycinamide ribonucleotide (FGAR) amidotransferase (PurL) and aminoimidazole ribonucleotide (AIR) synthetase (PurM) utilize ATP to activate the oxygen of an amide within their substrate toward nucleophilic attack by a nitrogen. AIR synthetase uses the product of PurL, formylglycinamidine ribonucleotide (FGAM) and ATP to make AIR, ADP and P_i.

Results: The structure of a hexahistidine-tagged PurM has been solved by multiwavelength anomalous diffraction phasing techniques using protein containing 28 selenomethionines per asymmetric unit. The final model of PurM consists of two crystallographically independent dimers and four sulfates. The overall R factor at 2.5 Å resolution is 19.2%, with an R_{free} of 26.4%. The active site, identified in part by conserved residues, is proposed to be a long groove generated by the interaction of two monomers. A search of the sequence databases suggests that the ATP-binding sites between PurM and PurL may be structurally conserved.

Conclusions: The first structure of a new class of ATP-binding enzyme, PurM, has been solved and a model for the active site has been proposed. The structure is unprecedented, with an extensive and unusual sheet-mediated intersubunit interaction defining the active-site grooves. Sequence searches suggest that two successive enzymes in the purine biosynthetic pathway, proposed to use similar chemistries, will have similar ATP-binding domains.

Addresses: [1]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA and [2]Departments of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

[†]Present address: Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN 55455, USA.

*Corresponding author.
E-mail: see3@cornell.edu

Keywords: AIR synthetase, FGAR amidotransferase, purine biosynthesis, PurL, PurM, trifunctional enzyme

Received: 18 March 1999
Revisions requested: 13 April 1999
Revisions received: 30 April 1999
Accepted: 5 May 1999

Published: 1 September 1999

Structure September 1999, 7:1155–1166
http://biomednet.com/elecref/0969212600701155

0969-2126/99/$ – see front matter

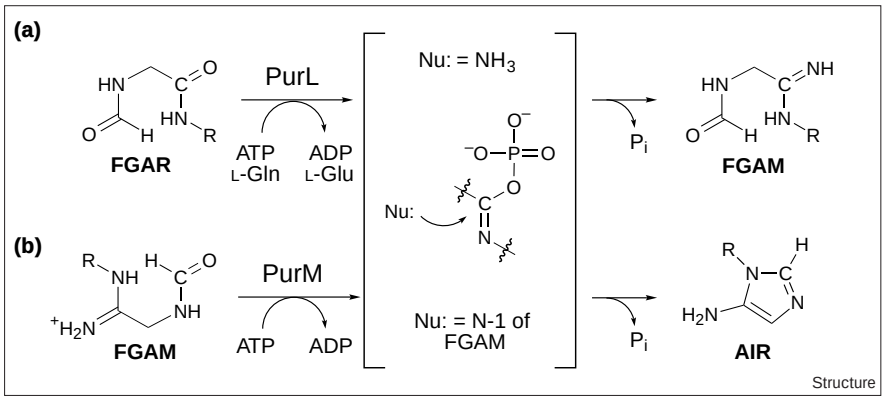
Introduction

The purine biosynthetic pathway serves as an excellent paradigm for thinking about the evolution of a biosynthetic pathway, including the importance of conserved strategies for substrate binding and/or chemical reactivity in this process. Each enzyme in this pathway binds 1′-substituted ribose 5-phosphate derivatives [1]. Six of twelve enzymes require ATP as a dehydrating agent, and four of these six enzymes are proposed to involve phosphoanhydride intermediates (PurD [GAR synthetase; EC 6.3.4.13], PurT [GAR transformylase, formate-dependent; EC 6.3.4.-], PurK [N^5-carboxyaminoimidazole ribonucleotide synthetase; EC 6.3.4.-], and PurC [4-[N-succinocarboxamide]-5-aminoimidazole ribonucleotide synthetase; EC 6.3.2.6]). The structures of PurD, PurK, and PurC that have recently become available demonstrate that PurD [2] and PurK [3] are part of the ATP grasp superfamily [4–8], while PurC has a related but distinct ATP-binding motif [9]. The high degree of sequence homology between PurT and PurK guarantees that the former is also part of the ATP grasp superfamily [10,11]. The relationship of PurD to PurK and PurT is of evolutionary interest as the latter two enzymes seem to be missing from avian and mammalian systems. The possibility that PurK and PurT have evolved from PurD has been the subject of recent discussion [3].

The two additional ATP-requiring enzymes in this pathway are formylglycinamide ribonucleotide (FGAM) synthetase (PurL) and aminoimidazole ribonucleotide (AIR) synthetase (PurM). PurL catalyzes the conversion of formylglycinamide ribonucleotide (FGAR), ATP and glutamine to FGAM, glutamate, ADP and P_i. PurM catalyzes the conversion of FGAM and ATP to AIR, ADP, and P_i. Both enzymes utilize very similar substrates and are proposed to utilize very similar chemistry (Figure 1).

In both cases ATP is proposed to phosphorylate an amide oxygen, generating an intermediate activated for nucleophilic attack [12,13]. PurL uses NH_3 derived from glutamine as the nucleophile, whereas PurM utilizes the N-1 of FGAM to catalyze an intramolecular ring formation. Iterated PSI-BLAST searches of sequence databases [14]

Figure 1



Reactions catalyzed by (a) PurL and (b) PurM. Both are proposed to generate a phosphorylated intermediate (R is ribose 5-phosphate).

using *E. coli* PurM as a query sequence reveal some very interesting hits. Notably, the searches show that those forms of PurL that have an independent glutamine-binding protein (designated PurQ) seem to have sequence homology with all forms of PurM in the putative ATP-metal-binding domain, as well as in the putative substrate-nucleotide-binding domain. An intriguing and unsolved question is the relationship between these two enzymatic activities and their structures. Toward understanding this relationship, we now report the first structure of an enzyme, *E. coli* PurM, catalyzing this reaction type.

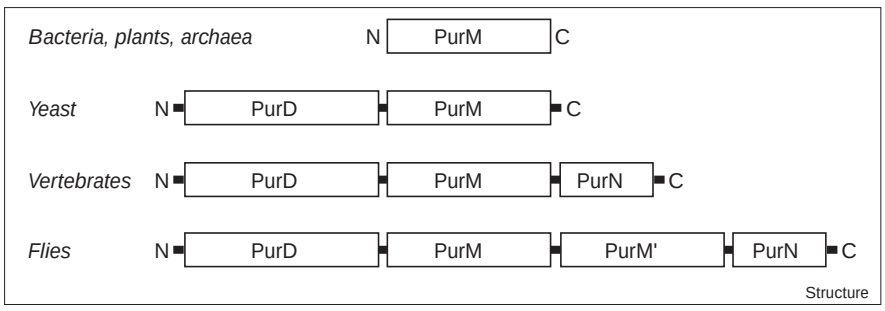
The structure of *E. coli* PurM is thus far unprecedented. Although the protein is not bound to any organic ligand, it contains a sulfate from crystallization conditions that is proposed to reside in the FGAM-binding domain. Sequence conservation of 23 residues among 24 available PurM sequences and recent ATP affinity labeling experiments allow us to identify the active-site region and to speculate on both the ATP- and FGAM-binding sites.

PurM catalyzes reaction six of the purine pathway (Figure 1). The proteins from *E. coli*, avian, and mammalian sources have been best characterized to date. The *E. coli* enzyme is a dimer with identical subunits (M_d = 38 kDa). Mechanistic studies have shown that the oxygen of ^{18}O-labeled formamide is transferred to P_i [12], consistent with the mechanism shown in Figure 1. In addition, studies of prochirally labelled FGAM suggest that either proton can be removed to form the aromatized aminoimidazole ring of the ribonucleotide [15]. These results suggest that this step happens non-enzymatically. Efforts to determine if a phosphorylated enzyme intermediate, in addition to a phosphorylated substrate intermediate, is involved in this reaction have been thus far unsuccessful [16].

In addition to its unusual reaction chemistry, PurM has attracted attention because it is a component of the mammalian trifunctional protein containing glycinamide ribonucleotide (GAR) synthetase (PurD, activity 2) and GAR transformylase (PurN, activity 3) [17]. The latter enzyme has been the target of the antifolates 5-deazacyclotetrahydrofolate [18] and N-[4-[2-2(amino-3,4-dihydro-4-oxo-7H-pyrrolo [2,3-d]pyrimidin-5-yl)ethyl]-benzoyl]-L-glutamic acid, each of which has anti-tumour activity [19]. In fact, PurM has been found in several multifunctional proteins within different sequence contexts (Figure 2).

The recent structure of PurD [2], the availability of the structure of PurN [20], and the structure presented herein allow us to speculate on the structure of the mammalian

Figure 2



Organization of the PurM genes in different organisms.

trifunctional protein. The first detailed structure of a PurM and a model for the trifunctional protein are the subject of this report. This structure provides a starting point for thinking about the unusual chemistry of this reaction and the structural relationship of PurL to PurM.

Results and discussion

Purification and crystallization

$(His)_6$-tagged PurM from *E. coli* was purified using a one-step nickel column affinity procedure, yielding 21 mg of 95% homogeneous protein from a 1.2 l culture. The enzyme was active in both wild-type and selenomethionine-substituted forms, with specific activities of 6.7 and 1.4 U/mg, respectively, at 37°C. These values are comparable to the 3.9 U/mg wild-type PurM activity determined using enzyme purified by ATP-affinity chromatography [12]. The substitution of methionine by selenomethionine (SeMet), required for the structure solution (see below), lowers the activity of $(His)_6$-tagged PurM to 0.9 s^{-1}. The solution aggregation state of $(His)_6$-tagged PurM was found to be a mixture of dimers and tetramers using a BioSilect size-exclusion column at room temperature. The molar ratio of dimer:tetramer is approximately 6:1 (data not shown). A comparable distribution of species was observed for the SeMet-substituted His-tagged PurM by Sephacryl 300HR size-exclusion column chromatography at 4°C (data not shown). These results contrast to those observed with the wild-type *E. coli* PurM, which behaves as a dimer in solution. The difference in aggregation state may therefore result from the $(His)_6$ tag. As none of the N-terminal $(His)_6$-tag residues are detectable by crystallography, it is not possible to determine the structural basis for the different aggregation states. Sequence alignments reveal that *E. coli* PurM has 2–6 additional N-terminal residues compared to most other bacterial PurMs, which in turn suggests that the disordered residues of the $(His)_6$ tag and first four amino acids of the protein crystallized here are not important in catalysis. PurM crystallizes from ammonium sulfate in space group $P2_12_12_1$ with unit cell dimensions of a = 71.2 Å, b = 211.7 Å, and c = 94.4 Å. Each asymmetric unit contains four monomers; however, the PurM dimer is the biologically active form [12].

Structure determination and refinement

The structure of PurM was determined by multiwavelength anomalous diffraction (MAD) phasing techniques using SeMet-incorporated protein. The crystallographic asymmetric unit contains two dimers, which corresponds to 28 Se-atoms. The Se atom positions were located using the computer program SnB [21], which utilizes direct methods based on Hauptman's minimal function [22]. E values derived from the anomalous-scattering signal corresponding to the peak wavelength alone were the most effective in locating the Se atoms. As described in the Materials and methods section, 23 of 28 Se-atom positions were located using anomalous differences to 3.0 Å resolution. The results we report here along with several other recent results [23] illustrate that MAD-phasing methods can be effectively applied to large structures.

The final model of PurM consists of two crystallographically independent dimers, 538 water molecules, and four sulfate ions. Dimer 1 contains residues 5–345 for monomer A and 21–345 for monomer B and dimer 2 contains residues 21–345 for each monomer A and B. The overall R factor at 2.5 Å resolution is 19.2% (R_{free} 26.4%). Examination of the final model using PROCHECK [24] suggests that the model is of very high quality.

Structure of the monomer

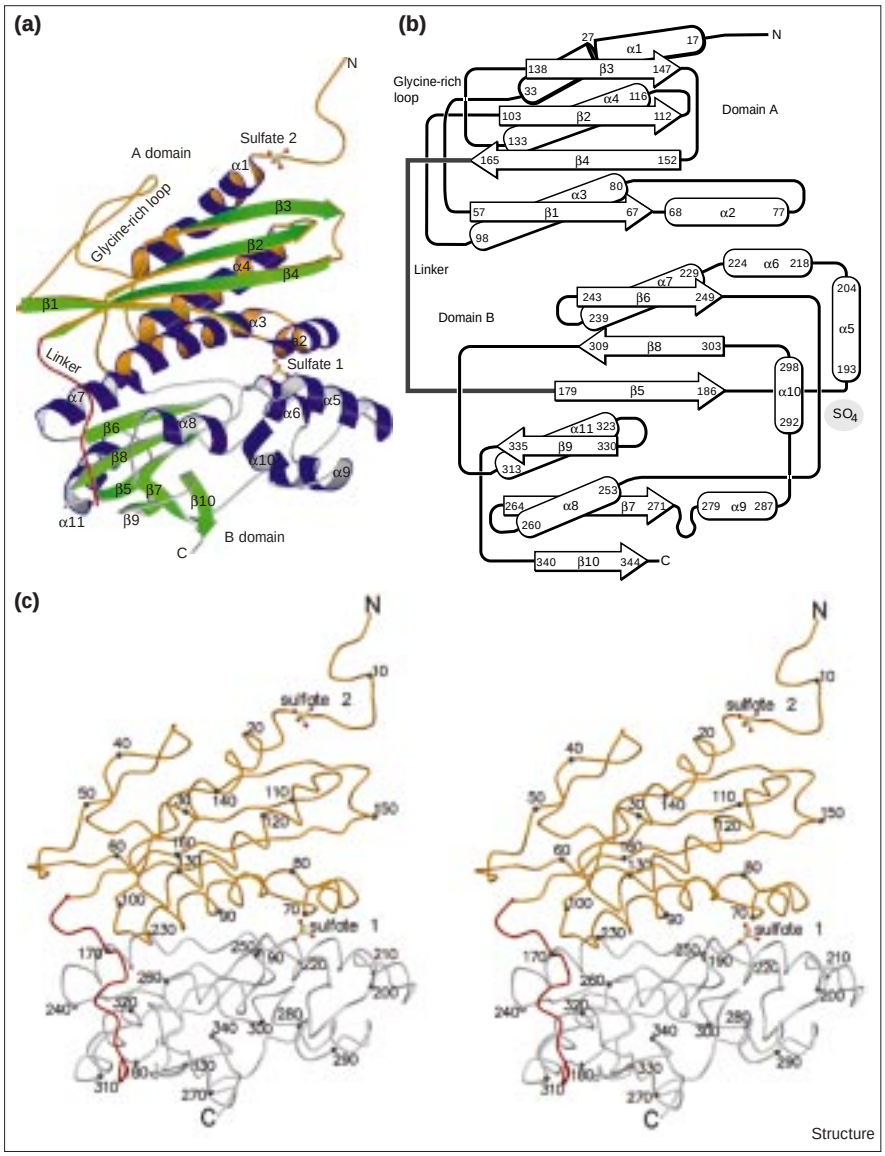
The PurM monomer consists of 345 amino acid residues in a single chain that can be divided into two clear structural domains. Domain A includes residues 16–166 and domain B includes residues 178–345. The two domains are connected by an 11-residue, well-defined loop from residue 167 to 177. The overall fold of PurM is illustrated in Figure 3.

Both domain A and domain B have an α/β-type fold. The N-terminal domain A is 47% α helix, 28% β sheet and 25% loops. The first 16 residues of domain A include the polyhistidine tag used for purification and are disordered. Domain A contains a four-stranded mixed β sheet ordered as β1β4β2β3. Strands β1, β2 and β3 are parallel to each other while strand β4 is antiparallel to the other three. All four β strands are unusually long, having 10–11 residues each. The sheet is flanked on one side by four α helices while the other side of the sheet forms the dimer interface. Three of the helices are long, with about 20 residues, while the shorter α2 is only half as long as the others. Domain A contains a long, poorly ordered glycine-rich loop between helix α1 and strand β1. This loop contains residues 40–46 (GGLGGFG), which have comparatively high temperature factors.

The C-terminal domain B is 36% α helix, 28% β sheet, and 46% loops. Both the α helices and the β strands in domain B are shorter than those in domain A. Domain B contains a six-stranded mixed β sheet ordered as β6β8β5β9β7β10. All six β strands are seven to nine residues long. The first five strands form an antiparallel sheet with the final strand, β10, parallel to β7. Seven α helices flank the β sheet, with α7 and α11 on one side of the sheet, α8 on the other side of the sheet and α5, α6, α9 and α10 clustered along one edge of the sheet. Helices α5–α10 are about 9–11 residues long while α11 is a little longer, with 13 residues.

The two domains are connected by a hinge-like loop consisting of 11 residues (red trace in Figure 3a). All of the residues in the loop region (S_{167}EIIDGSKVSD$_{177}$) are well ordered. Helix α3 of domain A interacts strongly with

Figure 3



Structure of the PurM monomer. **(a)** Ribbon diagram showing key structural features. Helices are shown in blue and strands in green. Domain A is outlined in gold and domain B is outlined in silver. **(b)** Topology diagram. Secondary structural elements are labeled with the first and last amino acid sequence number. **(c)** Stereoview of the Cα backbone. Every tenth Cα position is labeled. Domain A is shown in gold and domain B is shown in silver.

helices α6 and α7 of domain B and helix α2 of domain A interacts with helices α5 and α6 of domain B.

PurM displays a novel fold

DALI database searches [25] using the whole monomer or domains A and B separately reveal no topologically similar protein structure, indicating a novel protein fold. Furthermore, no ATP-binding protein could be identified with structural features similar to PurM, suggesting that PurM contains a novel ATP-binding domain.

Structure of the dimer

The PurM dimer is formed from identical monomers related by a noncrystallographic twofold axis with the N and C termini of different subunits about 20–25 Å apart. The monomeric subunit is about 55 Å long, 30 Å wide and 15 Å thick on the domain-A side and 25 Å thick on the domain-B side. The dimer is shaped like a prolate ellipsoid with two prominent grooves on opposite sides (Figure 4). These grooves, or clefts, are defined by portions of each monomer. The overall dimensions of the dimer are about $85 \times 45 \times 60$ Å.

The dimer interface is formed by joining two four-stranded β sheets face to face, one from each subunit, resulting in an eight-stranded β-barrel-like structure flanked by eight α helices (Figure 4). Most of the backbone distances between the two sheets are in the range of 6–7 Å and the intersubunit interactions are mostly hydrophobic. Essentially all of the dimer contacts occur

between domain A of one subunit and domain A of a twofold-related subunit. Thus, PurM displays a core made from two copies of domain A, resulting in a four-layer αββα sandwich. The core structure has a length of about 50 Å along the barrel axis and about 35 Å in the direction perpendicular to the barrel axis. Two copies of domain B are attached to opposite ends of the core structure using a single connecting loop for each (Figure 4).

Proposed active site

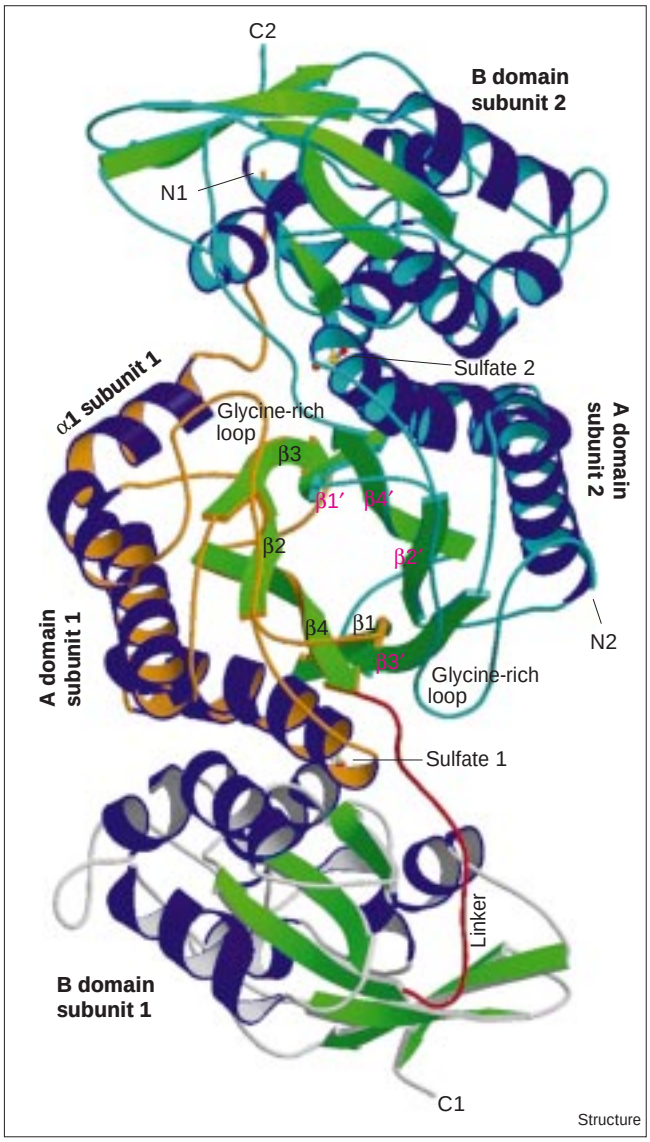
Attempts to soak AIR, ADP or ATP into the crystals have thus far been unsuccessful, resulting in either damaged crystals or no bound ligand. In addition, attempts to co-crystallize PurM with products or reactants have not yet succeeded. For these reasons, other methods were sought to identify the active site. Three pieces of evidence support the proposal that the active site of PurM is located within a large cleft formed primarily from β1, α2 and α3 of domain A, α5 and β6 of domain B from one monomer and β3 of the domain A of the second monomer (Figure 5). The cleft is about 30 Å long, about 15 Å wide and about 7 Å deep, about the right size and shape to contain the mononucleotide substrate, FGAM, and an extended ATP molecule with its terminal phosphate positioned adjacent to the mononucleotide-binding site.

First, an alignment of 23 PurM sequences indicates that most of the conserved residues line the putative active-site cleft. Second, the sulfate that co-crystallizes with PurM is buried at one end of the cleft. This site, as discussed below, is proposed to represent the mononucleotide-binding site. Third, affinity-labeling studies with fluorosulfonylbenzoyladenosine (FSBA), an ATP analog, acylated K_{27}, which is located at the other end of the cleft from the sulfate-binding site [26]. The labeling by FSBA was potentiated by the presence of FGAM and inhibited by the presence of ADP, suggesting that this end of the cleft binds ATP.

Sequence alignments

An alignment of nine PurMs showing the greatest sequence diversity, but excluding the extreme thermophiles, is shown in Figure 6. The locations of the conserved residues are shown in Figures 5a and b. Most striking is the alignment of 19 glycines. All of the sequences shown contain a glycine-rich loop located between residues 40 and 47 (GGLGGFGA [single-letter amino acid code] in *E. coli*). G_{43} and F_{45} are completely conserved in mesophiles; this region has a different sequence in extreme thermophiles, a modification that may reduce its flexibility at elevated temperatures. This loop exists in multiple conformations (Figure 5a) and is at the end of the cleft near the K_{27} that was labeled with FSBA. A second glycine-rich region including G_{66}VGxK_{70} (where x is any amino acid) near the beginning of α2 is also conserved, as is the glycine-rich region at the beginning of α5 (S_{186}xGxHSNGxxL_{196}). These last two regions, as described below, form the

Figure 4

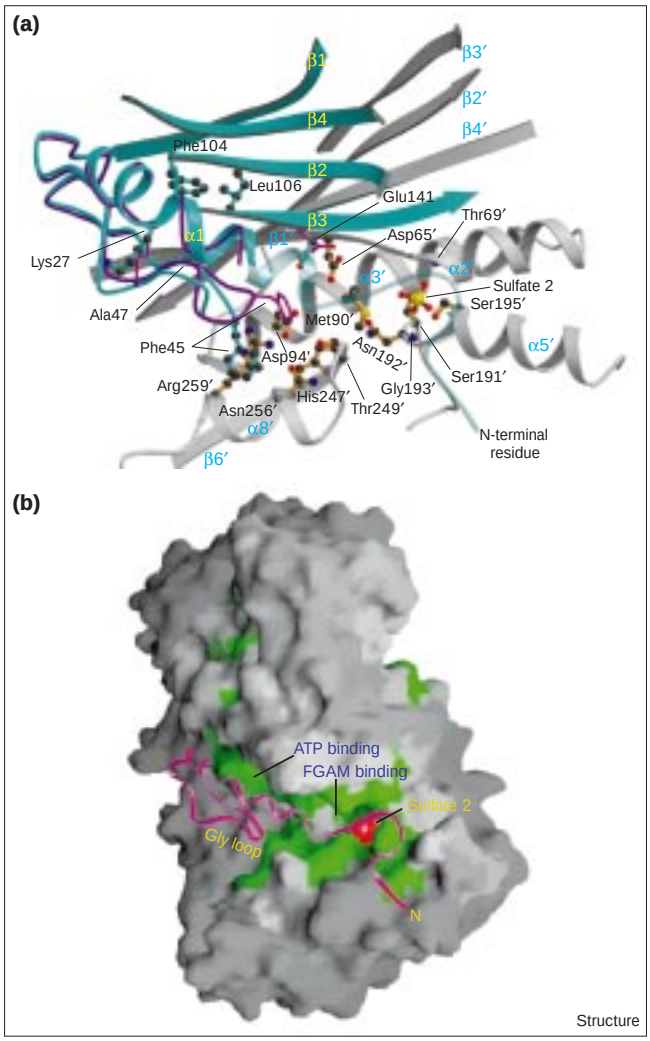


Structure of the PurM dimer. Ribbon diagram viewed down the twofold axis. Major structural features and the strands of the central β barrel, which form most of the dimer interface, are labeled. For subunit 1, the domain-A ribbon is highlighted in gold, the domain-B ribbon is highlighted in silver and the linker is shown in red. For subunit 2, the ribbon is highlighted in light blue.

sulfate-binding site. Finally, a number of conserved polar residues reside in the putative active-site cleft: D_{94}, H_{247}, R_{259}, T_{249} (Figure 5a). These residues are potential candidates for metal-ion ligands and for general acid and base catalysts. D_{86}, D_{155} and H_{190} are also conserved.

Two other conserved residues adopt unusual buried conformations near the cleft: D_{107} is buried in a hydrophobic pocket (and is presumably protonated) in the central core defined by the A domains. K_{70} extends away from the

Figure 5



Proposed active site for PurM. (a) Ribbon diagram of the active-site region. Conserved residues, shown in ball-and-stick representation, are labeled. Portions of domain A from one subunit are shown in blue. For the other subunit (labeled with primes), portions of domain A are shown in gray and portions of domain B are shown in silver. The sulfate site for the second subunit is labeled sulfate 2. (b) Surface representation showing the proposed binding sites for ATP and FGAM. The N-terminal residues (starting at position 20) for subunits 2–4 are shown in magenta. The ordered N-terminal residues (starting at position 5) for subunit 1 are shown in the transparent gray.

groove through the interface between each subunit's A and B domains to make a hydrogen bond with the backbone oxygen of A_{224}. K_{70} is part of a group of buried, conserved residues underneath the sulfate-binding site that may contribute to defining its structure (see below).

Sulfate-binding site

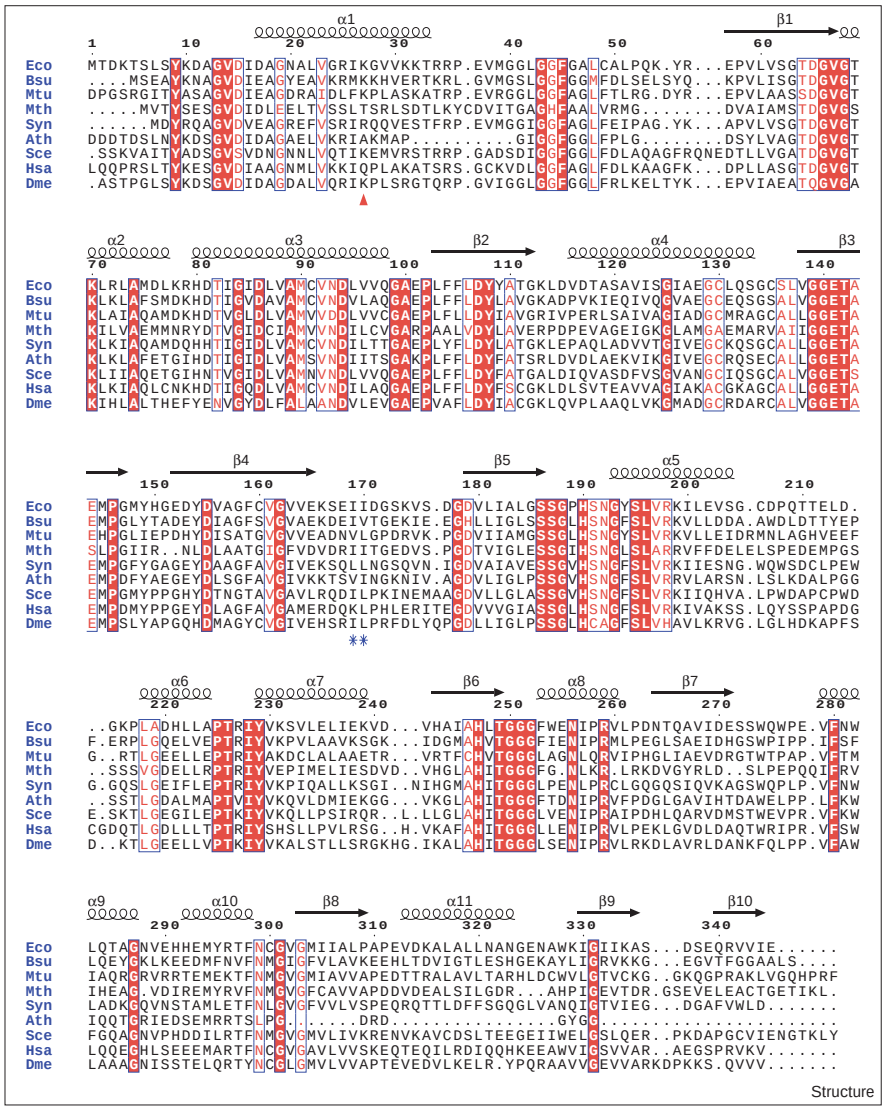
A sulfate ion is bound near the N-terminal end of helix α2 from domain A and the N-terminal end of helix α5 from domain B (Figure 3a and Figure 5a). In one of the two monomers, residues 5–20 extend across the groove and coil over the sulfate-binding site (Figure 5). Residues 5–11 are bound in a small groove on the B domain of the partner subunit. Hydrogen bonds are formed between the sulfate and the amide groups of Thr69 and Gly193 and with the sidechain and/or amide group of Ser195. The reaction catalyzed by PurM (Figure 1) suggests that this sulfate-binding site could represent either a site for ATP binding, for the product P_i binding, or for mononucleotide binding. As the sulfate site is deeply buried and consequently inaccessible to interaction with a mononucleotide, it is unlikely that this site represents one of the phosphate groups of ATP. It is also unlikely that this is the binding site of the P_i produced during the reaction, for the same reason. In addition, as noted above, affinity labeling with an ATP analog suggests that ATP binds at the other end of the cleft, near the glycine-rich loop (Figure 5). Thus, it seems more likely that the sulfate occupies the phosphate-binding site of the nucleotide reactant (FGAM) and product (AIR). However, if this is in fact the case, then it differs from the mononucleotide-binding site in other enzymes involved in the purine biosynthetic pathway. PurF, PurD, PurN, PurK, PurE, and PurC all seem to bind the phosphate of the common ribose-5-P moiety at a loop between a β strand and an α helix [2,3,20,27]. In PurM, the sulfate/phosphate-binding site is located at the N-terminal ends of two helices, one each from domain A and domain B.

Proposed ATP-binding motif

On the basis of FSBA labeling and the sequence conservation of the glycine-rich loop, we propose that the ATP-binding region is between this loop, which extends into the active-site cleft, and the central core of the protein, defined by the two A domains [26]. The flexibility of this region of the molecule, which is typical of ATP-binding loops that close over chemically reactive intermediates (e.g., the ATP grasp motif [28]), is consistent with relatively high temperature factors and the observed conformational differences between the glycine-rich loop in monomers 1 and 2. An important requirement for the ATP-binding site is conserved residues that could ligate Mg^{2+} ions, which are absolutely required for *E. coli* and chicken PurM activity [12]. Asp65, Asp94, and Glu141 are all positioned near the tip of the glycine-rich loop, where any or all of them could be involved in Mg^{2+} binding when ATP binds at the gap between the glycine-rich loop and the central core of PurM (Figure 5a). K^+ is also required for PurM activity [12], but it is not known if the ion participates in ATP binding.

As conserved residues from both subunits are positioned in this putative active-site groove, a dimer seems to be the minimal functional unit of PurM. This should also be true of other forms of PurM (Figure 2). The fly PurD–PurM–PurM′–PurN fusion protein could be a dimer in solution, but because it contains an internal

Figure 6



Sequence alignment for PurMs from diverse organisms. Locations of the secondary structural elements are labeled above the sequence. Conserved residues are highlighted in red. The sequences shown are *E. coli* (Eco), *B. subtilis* (Bsu), *M. tuberculosis* (Mtu), *Methanobacterium thermoautotrophicum* (Mth), *Synechococcus* sp. (Syn), *Arabidopsis thaliana* (Ath), *Saccharomyces cerevisiae* (Sce), *Homo sapiens* (Hsa) and *Drosophila melanogaster* (Dme). The red arrow indicates the position of K_{27}, which is modified by FSBA, and the pair of asterices indicates the site of the PurM′ domain insertion in the Dme sequence. Thirteen additional sequences from mesophiles are not shown. Overall sequence identity ranges from 38 to 55%.

duplication, it can potentially fold back upon itself to form a monomer.

Comparison of PurM with other ATP-dependent enzymes

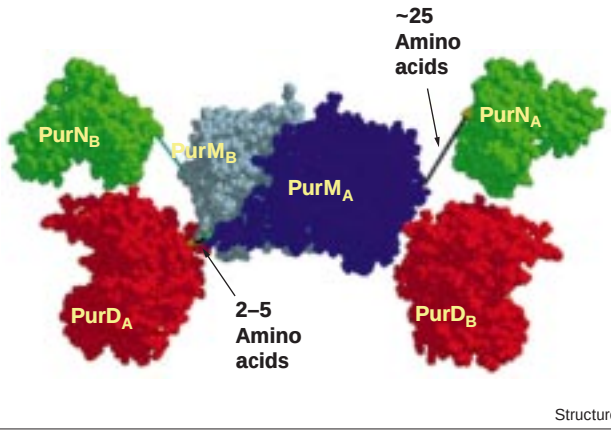
The PurM fold has not been previously observed, but sequence alignments suggest that it may resemble several other ATP-dependent enzymes whose structures are not yet known. PSI-BLAST database searches with the *E. coli* PurM protein sequence reveal similarities to several hydrogenase maturation proteins (HypE), thiamine monophosphate kinase (ThiL), selenophosphate synthetase (SelD), and most interesting, FGAR amidotransferase (PurL). With the exception of PurL, all of these enzymes are the same size as PurM (~340 amino acids). Even though ThiL and SelD use ATP [29,30], their chemistry must be mechanistically different [31] from that postulated for PurM. The relationship between HypEs [32], which are not known to be ATP-dependent enzymes, and PurMs remains a mystery. We will thus focus specifically on the similarities between the PurLs and PurM. As noted in the introduction, PurL precedes PurM in the purine biosynthetic pathway and seems to be mechanistically similar. Importantly, additional enzymes thought to function by similar catalytic mechanisms [33] (CTP synthetase [34], 5-oxoprolinase [35]) were not detected by these sequence searches.

PurL is the fourth enzyme in the purine pathway, generating the PurM substrate FGAM from FGAR (Figure 1). PurL sequences can be divided into two classes: the large PurLs have a glutaminase domain within a polypeptide of ~1450 amino acids, whereas the small PurLs (~740 amino acids) have a separate glutaminase protein, PurQ (~230 amino acids). PurQ and small PurL FGAR amidotransferase

forms are found in archaebacteria and several Gram-positive bacteria (*Bacillus subtilis*, *Mycobacterium tuberculosis*, *Lactobacillus casei* [36–38]), whereas large PurLs are found in most other organisms [1,13]. There is a remarkably low degree of sequence identity (6–7%) between the small and large PurL classes, which have 16% and 15% sequence identities within their respective groups. To date, only the large PurLs have been biochemically characterized [13,39]. The A domain of all PurM sequences, whether PurM is monomeric or part of a multifunctional protein, resembles the small PurL family. No alignments to the large PurL family, including the *E. coli* protein, are detected in BLAST searches, although they do appear at low similarity levels in the more sensitive iterated PSI-BLAST searches.

Sequence alignments using CLUSTALW [40] were carried out between the PurMs and small PurLs, excluding all of the hyperthermophiles. The A domain of PurM aligns with a 100 amino acid residue region at the N terminus of the small PurLs (residues 70–170). The conserved residues are not extensive and both deletions and insertions are required to make these alignments. However, the structure of PurM provides a guide to thinking about these alignments and their significance. First, the alignments reveal that eight glycine residues are conserved in strategic positions: there is a conserved glycine in the flexible glycine-rich loop of PurM (residues 40–43), and there are numerous glycine residues in the regions of putative metal-binding ligands located in the active-site cleft of PurM. Glycine residues are common at the active sites of proteins, where they confer sufficient flexibility for proper functioning of catalytic regions [41]. The most convincing regions of homology reside in a sequence $D_{94}x_4GAxP_{102}$ (*E. coli* PurM numbering; residues 121–130 in *B. subtilis* PurL). This region is conserved among all the PurMs and small PurLs and includes D_{94}, which may be a ligand to Mg^{2+}ATP. A second region of limited, but significant, homology is $G_{125}x_{13}(G/A)GE_{141}$ (residues 157–173 in *B. subtilis* PurL). E_{141} may also be a ligand to a metal in the ATP-binding site. Examination of the structure of PurM (Figure 5) reveals that both D_{94} and E_{141} are present in the putative active-site cleft. Also clear in the figure is that these aspartate and glutamate residues are on separate subunits. Thus this alignment makes a clear prediction: if PurLs are structurally homologous to PurMs, the small PurLs must function as dimers to form their ATP-binding sites, unless a single subunit of PurL is able to fold back upon itself.

Figure 7



Model of the trifunctional enzyme of mammalian purine biosynthesis (PurD–PurM–PurN) composed from structures of the *E. coli* homologs. PurD (red) is unliganded (PDB identifier 1gso [2]), PurM (blue, subunit 1 and light blue, subunit 2) is bound to sulfate, and PurN (green) is bound to GAR and a folate analog (PDB identifier 1cde [20]). Note the head-to-tail arrangement of the dimer and the proximity of the PurD and PurN domains. Linkers between domains are shown as bars, colored black for the subunit 1 fusion and light blue for the subunit 2 fusion. The dimensions of the model are 200 × 100 × 85 Å.

Limited homology and alignment by metal ion binding motifs has recently been used to identify the enolase superfamily of enzymes, which catalyze reactions involving carbanion intermediates [42]. Once one knows where to look, based on the structure and chemistry, one can find other members of a superfamily [43]. In the case of PurM and PurL, we need a structure of a nucleotide-bound PurM and the structure of a small PurL to confirm this speculation.

Possible structure of trifunctional GARS–PurM–GART

One intriguing aspect of the purine biosynthetic pathway is that multifunctional enzyme fusions are common in higher organisms and rare in bacteria [1]. PurM is the central component of the vertebrate trifunctional PurD–PurM–PurN [17] and the fly PurD–PurM–PurM′–PurN [44] fusions (Figure 2). As all three prokaryotic forms of these highly conserved enzymes are now crystallographically characterized [2,20], it is possible to use sequence alignments to estimate the lengths of the two linkers that separate the enzyme modules, and to model the trifunctional enzyme based on the shapes of its individual components.

Using sequence alignments, the PurD–PurM and the PurM–PurN linker lengths have been estimated to arrive at a model for the trifunctional protein (Figure 7). Using sequence alignments and the structures of *E. coli* PurD and PurM, we estimate the linker length may be as short as two to seven residues between PurD and PurM. Similarly, the linker length between PurM and PurN is estimated to be in the range of 17–28 residues. The linker sequences suggest that they are positively charged at neutral pH.

The crystallographically visible N-terminal residues of PurM contact the B domain of the partner subunit (Figure 5b). Thus, the results of the linker analysis predict that in the trifunctional protein the PurD domain is in intimate contact with the B domain of the other subunit in

the PurM dimer, to which the PurN domain is covalently fused by a longer tether. Thus, the PurM B domain will contact both fused enzymes. It is further evident that PurD and PurN, catalyzing steps 2 and 3 in the purine biosynthetic pathway, are constrained to be within 20–30 Å of each other, raising the possibility that the product of PurD, GAR, may be transferred directly, or channeled, between the active sites of these domains of the trifunctional enzyme.

The elongated conformation of the vertebrate trifunctional protein depicted in Figure 7 would also explain the anomalously large apparent molecular weight reported for the avian forms of the enzyme. Sucrose-gradient ultracentrifugation and size-exclusion chromatography using globular proteins as molecular-weight standards [13] allowed calculation of a solution size of 330 kDa for the trifunctional enzyme, an apparent trimer of the 108 kDa gene sequence [45,46]. Our prediction from the PurM structure is that this protein is really a dimer. Using the crystal-structure coordinates, we calculated radii of gyration for the PurM dimer (34 Å) and the trifunctional model in Figure 7 (~60 Å). These radii indicate that the trifunctional protein is roughly five times larger than the 70 kDa PurM dimer, in qualitative agreement with the observed trifunctional protein molecular weight.

Biological implications

The first crystal structure of an AIR synthetase, which is also the first member of a distinct class of ATP-utilizing enzymes, reveals that the dimer is the minimal structural and functional unit. It completes the structure determination of all three components of the human trifunctional enzyme (PurD–PurM–PurN) of purine biosynthesis, an emerging chemotherapeutic target. Recent studies have suggested that a pyrrolo [2,3-d]pyrimidine folate analog inhibits a variety of human folate-requiring enzymes, including PurN and PurH [19]. The success of an agent that inhibits multiple folate-requiring enzymes in the purine pathway suggests that the purine pathway should be revisited as a therapeutic target. Perhaps PurM and PurD inhibitors in combination with other nucleic acid synthesis inhibitors could function as anti-tumor agents. The structure of PurM is a starting point for designing inhibitors of this interesting reaction.

The structure of PurM and sequence alignments with PurL suggest that enzymes 4 and 5 in the purine pathway, which share a common product/substrate, will also share a common ATP-binding structural motif. Furthermore, identification of a number of additional enzymes that utilize Mg^{2+}ATP with this proposed motif, through their conserved glycines and metal ion binding residues, suggests that we have uncovered a new superfamily of ATP binders.

Materials and methods

Protein expression and isolation

The N-terminal $(His)_6$-tagged PurM overexpression vector (pHAS) was prepared by cloning a polymerase chain reaction (PCR) product derived from the wild-type *E. coli* PurMN expression plasmid pJS119 [15] into the *Eco*RI–*Hind*III sites of a modified pET28 vector [47]. The oligonucleotide primers used for PCR were 5′-CCGAATTCGTGACCGATAAAACCTCTCT and 5′-CAAAAGCTTTTATTCGATAACCACGCGTTGTT, where the *Eco*RI or *Hind*III sites are underlined.

FGAM was prepared and quantitated as described previously [12,16]. The $(His)_6$-tagged PurM was quantitated using the Bradford assay and crystalline bovine serum albumin as a standard [48]. PurM activity was monitored using the Bratton–Marshall assay for diazotizable amines [12,49]. A typical reaction mixture in a final volume of 1 ml contained 50 mM HEPES (pH 7.5), 20 mM $MgCl_2$, 150 mM KCl, 0.1 mM β-FGAM, 2 mM ATP and was carried out at 37°C. The reaction was initiated with 1.1 μg of protein and 0.15 ml aliquots were quenched at various times with 20% trichloroacetic acid in 1.33 M potassium phosphate (pH 1.4). One unit (U) is defined as one μmol AIR formed per min.

BL21(DE3) cells transformed with pHAS were selected on LB plates with 70 mg/l kanamycin. A single colony was transferred into 1.2 l LB with kanamycin and grown with shaking at 37°C to an $A_{600} \approx 1$, at which time solid isopropyl-β-D-thiogalactoside (Boehringer Mannheim) was added to 0.3 mM. After growth for an additional 3 h at 37°C, the cells (6 g) were harvested by centrifugation at 4°C at 5000 × g for 10 min. A 9% sodium dodecyl sulfate polyacrylamide gel electrophoresis gel showed induction of a 38 kDa band that accounted for 60% of the total protein (data not shown). Cells were resuspended in 60 ml buffer (0.1 M KCl, 20 mM Tris, pH 8.0, in 10% glycerol) and disrupted by two passes through a French press at 16,000 psi. Cell debris was removed by centrifugation at 4°C at 24,000 × g for 45 min. Iminodiacetic acid 6B FF resin (Sigma, 5 ml) was charged with 0.1 M $NiSO_4$ in water (10 ml), washed with water, and equilibrated in buffer. A slurry of the charged resin was poured into the soluble extract, which was stirred at 4°C for 30 min. The resin was centrifuged at 4°C at 10,000 × g for 10 min, the supernatant was decanted, and the resin was poured into a glass column (1.5 × 3.0 cm). The column was washed with buffer (30 ml) and buffer plus 10 mM imidazole (50 ml). Bound proteins were eluted with buffer plus 400 mM imidazole (50 ml); EDTA and βME were added from concentrated stocks to the pooled eluate to final concentrations of 10 mM. The eluate was concentrated by ultrafiltration (Amicon YM30 membrane) to 5 ml, which was dialyzed overnight against buffer plus 10 mM βME (1 l, two changes) at 4°C. The dialyzed enzyme was concentrated by ultrafiltration, flash-frozen in aliquots in liquid N_2, and stored at –80°C.

$(His)_6$-tagged PurM (0.1 mg in 0.1 ml of 0.2 M KCl, 50 mM Tris, pH 8.0) was injected onto a BioRad BioSilect SEC 250-5 column (7.8 × 300 mm) equilibrated in and eluted isocratically with the same buffer at 23°C at 0.75 ml/min. A mixture of molecular-weight standards (thyroglobulin [669 kDa], immunoglobulin G [158 kDa], ovalbumin [44 kDa], myoglobin [17 kDa], and vitamin B12 [1.35 kDa] from BioRad) was injected and eluted from the same column under the same conditions. The elution times were used to construct a calibration curve.

PurM containing SeMet was prepared using the same cell growth and protein-purification procedure described for PurD containing SeMet [47], in which pHAS was substituted for the PurD overexpression construct.

Crystallization

Crystals of SeMet PurM were prepared by the hanging-drop vapor-diffusion technique. The SeMet (His)6-tagged PurM was stored at a concentration of 12 mg/ml in 15 mM Tris at pH 7.5, with 10 mM DTT to protect the SeMet $(His)_6$-tagged PurM from oxidation. The protein solution was stored at –70.0°C and a freshly thawed sample was used for all crystallizations. Optimized crystallization conditions consisted of 2.10–2.25 M ammonium sulfate and 3–6% (v/v) isopropanol.

Table 1

X-ray data-collection statistics.

Dataset	Edge	Peak	Remote
Wavelength (Å)	0.9792	0.9790	0.9638
Resolution (Å)	20–2.5	20–2.5	20–2.5
Unique reflections	45,728	45,754	45,650
Redundancy	4.2	4.2	4.2
Completeness(%)	91.4	91.4	91.4
I/σ ratio	10.1	8.8	9.7
R_{sym} (%)	5.4	5.6	4.7
R_{sym} (2.64–2.50 Å; %)	18.3	19.9	14.3

No additional buffer was added. The hanging drops consisted of 3 µl of protein solution plus 3 µl of well solution and were equilibrated against 1 ml of well solution. Crystals appeared after 7 days and grew to a maximum size of $0.1 \times 0.1 \times 0.5$ mm in 10–14 days. Crystals of SeMet $(His)_6$-tagged PurM belong to space group $P2_12_12_1$ with unit-cell dimensions of a = 71.2 Å, b = 211.7 Å and c = 94.4 Å. The Matthew's number is 2.38 $Å^3$/Da, assuming four molecules per asymmetric unit, which corresponds to 47.5% solvent volume.

X-ray data collection

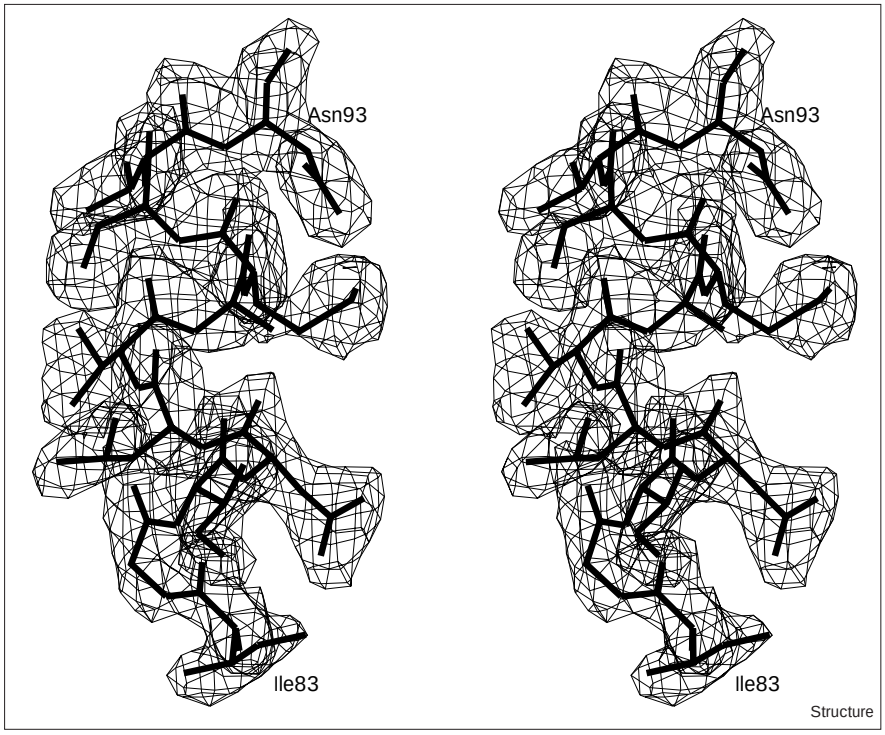
The MAD data were collected at CHESS station F2. The data-collection crystal was flash-frozen in a liquid N_2 stream at 170K. The stabilizing solution contained 25% (v/v) glycerol to prevent crystal damage during freezing. All data were recorded with a Quantum 4 mosaic charge-coupled device (CCD) detector. A crystal scan was used to determine the wavelengths for data collection. Data were collected at the Se absorption edge (12.661 keV), the absorption peak (12.664 keV) and at a remote reference energy (12.864 keV). The crystals showed a low mosaic spread (0.200°) and 2.5 Å data could be measured using a 2.0° oscillation angle and an exposure time of 40 s. A total of 66.0° of data were collected plus inverse beam images for each of the three energies. The data were processed with MOSLFM [50] and scaled using SCALA [51]. Because of technical problems during data collection, only the fully recorded reflections were used, resulting in a completeness of about 91% and a redundancy of greater than four. The R_{sym} values for the edge, peak and remote energies were 5.4%, 5.6% and 4.7%, respectively. Data-collection statistics are summarized in Table 1.

Structure determination

Integrated intensities were processed using the programs DREAR and DIFFE to calculate normalized structure factors, E, from the anomalous differences [52]. The 3.0 Å data from the peak absorption were input into the direct methods program SnB [53], which implements the Shake-and-Bake procedure of Hauptman and coworkers [54]. The Shake-and-Bake procedure uses the minimal principle and real-space filtering to refine randomly generated phase angles for the largest E values. A total of 500 random trials were used with 1120 E values and 11,200 triple phase relationships. For each trial, the phases were refined through 150 cycles of the Shake-and-Bake procedure. The results for the 500 trials showed a bimodal distribution of the minimal function and indicated that three of the random trials had converged to the correct phase angles. Each of the three phase sets consistently showed 23 of the 28 possible Se-atom sites. Technical problems during data collection limited the data quality and may be the reason that five Se-atom sites were not found by SnB.

The Se sites were refined and phases calculated with the program MLPHARE [55]. At 3.0 Å, the overall figure of merit (FOM) was 0.568. The maps calculated directly from unmodified MLPHARE phases showed a clear protein/solvent boundary, including the major protein domains and many secondary structure features. However, the electron density contained many breaks and the mainchain tracing was not possible at this stage. The electron density was then modified by noncrystallographic-symmetry (NCS) averaging using the four monomers in one asymmetric unit. The transformation matrices between monomers were estimated using the corresponding Se atoms in each of the

Figure 8



Stereoview of a section of electron density corresponding to an α-helical region of the model. The MAD-phased electron-density map has been modified by solvent flattening and fourfold-symmetry averaging. The electron density is contoured at 1σ.

Table 2

Refinement statistics.

Resolution (Å)	20.0–2.5
Unique reflections	44,742
Completeness (%)	89.2
R factor	19.2
R_{free}	26.4
Average B factor (Å^2)	16.6
Deviation from ideals (rmsd)	
bonds (Å)	0.009
angle (°)	1.4
dihedrals (°)	25
impropers (°)	1.3
Atoms in an asymmetric unit	
protein	9990
water	538
sulfate ion	20
Coordinate error from Luzzati plot (Å)	0.15
Ramachandran plot	
most favored region(%)	90.1
additionally allowed region(%)	8.3
generally allowed region(%)	0.7
unfavored region(%)	0.8

Rmsd is root mean square deviation.

monomers and refined by the program RAVE_IMP [56]. The NCS averaging, solvent flattening and histogram matching were performed with the program DM [57]. Solvent flattening and histogram matching alone resulted in a drop in the DM R factor from 51.6% to 43.7%. NCS averaging improved the phases further and the DM R factor indicator dropped to 35.8%. Using this procedure, the maps at both 3.0 Å and 2.5 Å became traceable. Figure 8 depicts a section of α helix in an initial F_o map at 3.0 Å resolution.

The first round of chain tracing allowed the fitting of 291 out of 345 residues, leaving 20 residues at the N terminus, residues 40–55 and residues 203–216 not built for all four monomers. The N terminus contained the polyhistidine tag that was used for purification. The initial crystallographic R factor based on this model was 44.2%. The first round of refinement was carried out with strict NCS constraints and resulted in an R factor of 31.5% (R_{free} 33.6%) at 2.5 Å resolution. Residues 203–216 were added to the model in a second round of model building. For the second round of refinement, the NCS restraints were relaxed, which resulted in a R factor 27.0% and R_{free} 31.0%. Residues 47–55 were added to the model for each monomer. In parallel, torsion-angle refinements were performed using CNS [58], yielding almost the same results. At this point, the N-terminal 20 residues and a glycine-rich loop, 40–46 (GGLGGFG), were missing from the model. Subsequently, refinements without NCS restraints helped build residues 40–46, which showed conformational variations among the four monomers. The N-terminal residues, 5–20, were visible only in one monomer. In the final stages of refinement, 538 water molecules were included. In addition, a sulfate ion was located at the same position in each of the four monomers. The final refinement statistics are summarized in Table 2. The final R factor was 19.2% (R_{free} 26.4%). The final Ramachandran plot [59] showed good geometry for the final model, with 90.1% of the residues in the most favored region.

Prediction of the molecular weight of the trifunctional (PurD–PurM–PurN) fusion protein based on individual structures and the model for the fusion protein structure

The radii of gyration for monomeric PurD (22 Å), dimeric PurM (34 Å) and monomeric PurN (17 Å) were calculated using MOLEMAN2 (GJ Kleywegt, unpublished program) on all non-water atoms in each PDB file (PurD, 1gso.pdb [2]; PurN, 1cde.pdb [20]). A model of the trifunctional protein shown in Figure 7 has a radius of gyration of ~60 Å, indicating that it is roughly $(60/34)^3$ = 5.4-fold larger than the PurM dimer or about 380 kDa. The PurM dimer has been used to compare size because it has an elongated shape, similar to that proposed for our trifunctional model.

Figure generation

Programs used to prepare figures included MolScript [60], GRASP [61], Raster3D [62], and ESPript (P Gouet, unpublished program).

Accession numbers

The coordinates for the PurM model have been deposited with the Protein Data Bank with accession code 1CLI.

Acknowledgements

The authors would like to acknowledge the support of the WM Keck Laboratory for Molecular Structure (SEE), the Lucille P Markey Charitable Trust (SEE), and the National Institutes of Health grant GM32191 (JS). TJK is supported by NIH cancer training grant CA09112. This work is based upon research conducted at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the National Science Foundation under award DMR-9311772, using the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by award RR-01646 from the National Institutes of Health. We thank Ayla Aygun for technical assistance.

References

1. Zalkin, H. & Dixon, J.E. (1992). *De novo* purine nucleotide biosynthesis. *Prog. Nucleic Acid Res. Mol. Biol.* 42, 259-287.
2. Wang, W., Kappock, T.J., Stubbe, J. & Ealick, S.E. (1998). X-ray crystal structure of glycinamide ribonucleotide synthetase from *Escherichia coli*. *Biochemistry* 37, 15647-15662.
3. Thoden, J.B., Kappock, T.J., Stubbe, L. & Holden, H.M. (1999). Three-dimensional structure of N5-carboxyaminoimidazole ribonucleotide (N5-CAIR) synthetase: a member of the ATP-Grasp protein superfamily. *Biochemistry*, in press.
4. 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.
5. Fan, C., Moews, P.C., Walsh, C.T. & Knox, J.R. (1994). Vancomycin resistance: Structure of D-alanine:D-alanine ligase at 2.3 Å resolution. *Science* 266, 439-443.
6. Yamaguchi, H., *et al*., & Katsuube, Y. (1993). 3-Dimensional structure of the glutathione synthetase from *Escherichia coli* B at 2.0 Å resolution. *J. Mol. Biol.* 229, 1083-1100.
7. Esser, L., Wang, C.-R., Hosaka, M., Smagula, C.S., Sudhof, T.C. & Deisenhofer, J. (1998). Synapsin I is structurally similar to ATP-utilizing enzymes. *EMBO J.* 17, 977-984.
8. Thoden, J.B., Holden, H.M., Wesenberg, G., Raushel, F.M. & Rayment, I. (1997). Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product. *Biochemistry* 36, 6305-6316.
9. Levdikov, V.M., Barynin, V.V., Grebenko, A.I., Melik-Adamyan, W.R., Lamzin, V.S. & Wilson, K.S. (1998). The structure of SAICAR synthase: An enzyme in the *de novo* pathway of purine nucleotide biosynthesis. *Structure* 6, 363-376.
10. Marolewski, A., Smith, J.A. & Benkovic, S.J. (1994). Cloning and characterization of a new purine biosynthetic enzyme: A non-folate glycinamide ribonucleotide transformylase from *Escherichia coli*. *Biochemistry* 33, 2531-2537.
11. Galperin, M.Y. & Koonin, E.V. (1997). A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci.* 6, 2639-2643.
12. Schrimsher, J.L., Schendel, F.J., Stubbe, J. & Smith, J.M. (1986). Purification and characterization of aminoimidazole ribonucleotide synthetase from *Escherichia coli*. *Biochemistry* 25, 4366-4371.
13. Schendel, F.J., Mueller, E., Stubbe, J., Shiau, A. & Smith, J.A. (1989). Formylglycinamide ribonucleotide synthetase from *Escherichia coli*: cloning, sequencing, overexpression, isolation, and characterization. *Biochemistry* 28, 2459-2471.
14. Altschul, S.F., *et al.,* & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of database search programs. *Nucleic Acids Res.* 25, 3389-3402.
15. Schendel, F.J. (1986). *Mechanistic studies on de novo purine biosynthesis.* Ph.D. dissertation, University of Wisconsin-Madison.
16. Mueller, E.J. (1994) *Mechanistic studies on de novo purine biosynthesis: Probing for intermediates of PurM, FGAR amidotransferase and AIR carboxylase.* Ph.D. dissertation, Massachusetts Institute of Technology.

17. Daubner, S.C., *et al.*, & Benkovic, S.J. (1985). A multifunctional protein possessing glycinamide ribonucleotide synthetase, glycinamide ribonucleotide transformylase, and aminoimidazole ribonucleotide synthetase activities in *de novo* purine biosynthesis. *Biochemistry* 24, 7059-7062.
18. Pizzorno, G., *et al.,* & Beardsley, G.P. (1995). Multifactorial resistance to 5,10-dideazatetrahydrofolic acid in cell lines derived from human lymphoblastic leukemia CCRF-CEM. *Cancer Res.* 55, 566-573.
19. Shih, C., *et al.,* & Schultz, R.M. (1997). LY231514, a pyrrolo[2,3-d]pyrimidine-based antifolate that inhibits multiple folate-requiring enzymes. *Cancer Res.* 57, 1116-1123.
20. Almassy, R.J., Janson, C.A., Kan, C.-C., & Hostomska, Z. (1992). Structures of apo and complexed *Escherichia coli* glycinamide ribonucleotide transformylase. *Proc. Natl Acad. Sci. USA* 89, 6114-6118.
21. Miller, R., Gallo, S.M., Khalak, H.G. & Weeks, C.W. (1994). SnB: crystal structure determination via shake-and-bake. *J. Appl. Crystallogr.* 27, 613-621.
22. Hauptmann, H.A. (1991). A minimal principle in the phase problem. In *Crystallographic Computing 5:* From *Chemistry to Biology*. (Moras, D., Podjarny, A.D. & Thierry J.C., eds), pp. 324-332, IUCr and Oxford University Press, Oxford.
23. Cassetta, A., Deacon, A.M., Ealick, S.E., Helliwell, J.R. & Thompson, A.W. (1999). Development of instrumentation and methods for MAD and structural genomics at the SRS, ESRF, CHESS and Elettra facilities. *J. Synchrotron Radiation* 6, 822-833.
24. 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. Crystallogr.* 26, 283-291.
25. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123-138.
26. Mueller, E.J., *et al.*, & Stubbe, J. (1999). Investigation of the ATP binding site of *Escherichia coli* aminoimidazole ribonucleotide synthetase using affinity labelling and site-directed mutagenesis. *Biochemistry* 38, 9831-9839.
27. Krahn, J.M., Kim, J.H., Burns, M.R., Parry, R.J., Zalkin, H. & Smith, J.L. (1997). Coupled formation of an amidotransferase interdomain ammonia channel and a phosphoribosyltransferase active site. *Biochemistry* 36, 11061-11068.
28. Kato, H., *et al.,* & Oda, J. (1994). Flexible loop that is novel catalytic machinery in a ligase. Atomic structure and function of the loopless glutathione synthetase. *Biochemistry* 33, 4995-4999.
29. Imamura, N. & Nakayama, H. (1982). thiK and thiL loci of *Escherichia coli*. *J. Bacteriol.* 151, 708-717.
30. Forchhammer, K. & Böck, A. (1991). Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. *J. Biol. Chem.* 266, 6324-6328.
31. Mullins, L.S., Hong, S.B., Gibson, G.E., Walker, H., Stadtman, T.C. & Raushel, F. M. (1997) Identification of a phosphorylated enzyme intermediate in the catalytic mechanism for selenophosphate synthetase. *J. Am. Chem. Soc.* 119, 6684-6685.
32. Maier, T. & Böck, A. (1996) Generation of Active [NiFe] Hydrogenase *in vitro* From a Nickel-Free Precursor Form. *Biochemistry* 35, 10089-10093.
33. Satterthwait, A.C. & Westheimer, F.H. (1980). Monomeric methyl metaphosphate: reactions with carbonyl groups. *J. Am. Chem. Soc.* 102, 4464-4472.
34. Lewis, D.A. & Villafranca, J.J. (1989). Investigation of the mechanism of CTP synthetase using rapid quench and isotope partitioning methods. *Biochemistry* 28, 8454-8459.
35. Griffith, O.W. & Meister, A. (1981) 5-Oxo-L-prolinase (L-pyroglutamate hydrolase). Studies of the chemical mechanism. *J. Biol. Chem.* 256, 9981-9985.
36. 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 biosynthesis. *J. Biol. Chem.* 262, 8274-8287.
37. Jackson, M., *et al.,* & Guilhot, C. (1996). The *Mycobacterium tuberculosis* purine biosynthetic pathway: Isolation and characterization of the *purC* and *purL* genes. *Microbiology* 142, 2439-2447.
38. Gu, Z.-M., Martindale, D.W. & Lee, B.H. (1992). Isolation and complete sequence of the *purL* gene encoding FGAM synthase II in *Lactobacillus casei*. *Gene* 119, 123-126.
39. Buchanan, J.M. (1982). Covalent reaction of substrates and antimetabolites with formylglycinamide ribonucleotide amidotransferase. *Methods Enzymol.* 87, 76-84.
40. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.
41. Yan, B.X. & Sun, Y.Q. (1997). Glycine residues provide flexibility for enzyme active sites. *J. Biol. Chem.* 272, 3190-3194.
42. Babbitt, P.C., *et al.,* & Gerlt, J.A. (1996). The enolase superfamily: A general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids. *Biochemistry* 24, 16489-16501.
43. Gerlt, J.A. & Babbitt, P.C. (1998). Mechanistically diverse enzyme superfamilies: The importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol.* 2, 607-612.
44. Henikoff, S. (1986). The *Saccharomyces cerevisiae* ADE5,7 protein is homologous to overlapping *Drosophila melanogaster* Gart polypeptides. *J. Mol. Biol.* 190, 519-528.
45. Schrimsher, J.L., Schendel, F.J. & Stubbe, J. (1986). Isolation of a multifunctional protein with aminoimidazole ribonucleotide synthetase, glycinamide ribonucleotide synthetase, and glycinamide ribonucleotide transformylase activities: Characterization of aminoimidazole ribonucleotide synthetase. *Biochemistry* 25, 4356-4365.
46. Aimi, J., Qiu, H., Williams, J., Zalkin, H. & Dixon, J.E. (1990). *De novo* purine nucleotide biosynthesis: Cloning of human and avian cDNAs encoding the trifunctional glycinamide ribonucleotide synthetase-aminoimidazole ribonucleotide synthetase-glycinamide ribonucleotide transformylase by functional complementation in *E. coli*. *Nucleic Acids Res* 18, 6665-6672.
47. Weaver, T.M., Wang, W. & Ealick, S.E. (1999). Purification, crystallization and preliminary X-ray diffraction data from selenomethionine glycinamide ribonucleotide synthetase. *Acta. Crystallogr. D* 55, 518-521.
48. Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248-254.
49. Bratton, A.C. & Marshall, E.K., Jr. (1939). A new coupling component for sulfanilimide determination. *J. Biol. Chem.* 128, 537-550.
50. Leslie, A.G.W. (1997). *MOSFLM version 5.40.* MRC Laboratory of Molecular Biology.
51. Evans, P.R. (1993). *SCALA version 3.3.* MRC Laboratory of Molecular Biology.
52. Smith, D.G., Nagar, B., Rini, J.M., Hauptman, H.A. & Blessing, R.H. (1997). The use of SnB to determine an anomalous scattering substructure. *Acta Crystallogr. D* 54, 799-804.
53. Weeks, C.M. & Miller, R. (1999). The design and implementation of SnB version 2.0. *J. Appl. Crystallogr.* 32, 120-124.
54. Miller, R., DeTitta, G.T., Jones, R., Langs, D.A., Weeks, C.M. & Hauptman, H.A. (1993). On the application of the minimal principle to solve unknown structures. *Science* 259, 1430-1433.
55. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters, In *Isomorphous Replacement and Anomalous Scattering,* (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80-86, Daresbury Laboratory, Warrington, UK.
56. Kleywegt, G.J. & Jones, T.A. (1997) Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallogr. D* 53, 179-185.
57. Cowtan, K. (1994). An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31, 34-38.
58. Brünger, A.T., *et al.,* & Warren, G.L. (1998). Crystallography and NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54, 905-921.
59. Ramachandran, C. & Ramakrishnan, G.N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* 5, 909-933.
60. Kraulis, P.J. (1991) MolScript: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.
61. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281-296.
62. Merritt, E.A. & Bacon, D.J. (1997). Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* 277, 505-524.