The crystal structure of *Escherichia coli* purine nucleoside phosphorylase: a comparison with the human enzyme reveals a conserved topology

Chen Mao[1†], William J Cook[2], Min Zhou[1], George W Koszalka[3], Thomas A Krenitsky[3] and Steven E Ealick[1*]

Background: Purine nucleoside phosphorylase (PNP) from *Escherichia coli* is a hexameric enzyme that catalyzes the reversible phosphorolysis of 6-amino and 6-oxopurine (2′-deoxy)ribonucleosides to the free base and (2′-deoxy)ribose-1-phosphate. In contrast, human and bovine PNPs are trimeric and accept only 6-oxopurine nucleosides as substrates. The difference in the specificities of these two enzymes has been utilized in gene therapy treatments in which certain prodrugs are cleaved by *E. coli* PNP but not the human enzyme. The trimeric and hexameric PNPs show no similarity in amino acid sequence, even though they catalyze the same basic chemical reaction. Structural comparison of the active sites of mammalian and *E. coli* PNPs would provide an improved basis for the design of potential prodrugs that are specific for *E. coli* PNP.

Results: The crystal structure of *E. coli* PNP at 2.0 Å resolution shows that the overall subunit topology and active-site location within the subunit are similar to those of the subunits from human PNP and *E. coli* uridine phosphorylase. Nevertheless, even though the overall geometry of the *E. coli* PNP active site is similar to human PNP, the active-site residues and subunit interactions are strikingly different. In *E. coli* PNP, the purine- and ribose-binding sites are generally hydrophobic, although a histidine residue from an adjacent subunit probably forms a hydrogen bond with a hydroxyl group of the sugar. The phosphate-binding site probably consists of two mainchain nitrogen atoms and three arginine residues. In addition, the active site in hexameric PNP is much more accessible than in trimeric PNP.

Conclusions: The structures of human and *E. coli* PNP define two possible classes of nucleoside phosphorylase, and help to explain the differences in specificity and efficiency between trimeric and hexameric PNPs. This structural data may be useful in designing prodrugs that can be activated by *E. coli* PNP but not the human enzyme.

Addresses: [1]Section of Biochemistry, Cell and Molecular Biology, Cornell University, Ithaca, NY 14853, USA, [2]Department of Pathology, University of Alabama at Birmingham, Birmingham, AL 35294, USA and [3]Wellcome Research Laboratories, Research Triangle Park, NC 27709, USA.

[†]Present address: Hughes Institute, Roseville, MN 55112, USA.

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

Key words: enzyme catalysis, nucleoside phosphorylase, nucleosides, phosphorolysis

Received: **23 June 1997**
Revisions requested: **22 July 1997**
Revisions received: **22 August 1997**
Accepted: **3 September 1997**

Structure 15 October 1997, **5**:1373–1383
http://biomednet.com/elecref/0969212600501373

Introduction

Purine nucleoside phosphorylase (PNP) catalyzes the reversible phosphorolysis of (2′-deoxy)purine ribonucleosides to the free base and (2′-deoxy)ribose-1-phosphate. The enzyme has been isolated from both eukaryotic and prokaryotic organisms and functions in the purine salvage pathway. Sequence analysis among PNPs suggests that there are two major categories of this enzyme. Mammalian PNPs, such as those from human erythrocytes and bovine spleen, are trimeric and have a monomeric molecular weight of 31 kDa [1–4], whereas prokaryotic PNPs are hexameric with a monomeric molecular weight of 26 kDa [5–11]. In general, PNPs with a trimeric structure accept only guanosine and inosine as substrates, whereas the hexameric PNPs accept adenosine as well, although in at least one instance a hexameric PNP is specific for adenosine and deoxyadenosine [9]. In addition, *Escherichia coli* PNP is more tolerant of nucleosides that contain modified ribose moieties, and compared to mammalian PNP its activity toward purine arabinosides is ten times higher [12]. Interestingly, both trimeric and hexameric PNPs have been observed in *Bacillus subtilis* [8], *E. coli* [10] and *Bacillus stearothermophilus* [11].

The two types of PNPs have different inhibition profiles. For example, formycin A is completely inactive against human PNP, but it is a good inhibitor of *E. coli* PNP (K_i = 5 μM) [13]. Furthermore, although mammalian PNPs have higher substrate specificity than the hexameric PNPs, they are more efficient in the synthesis of 2-oxo-6-substituted purines and purine analogs than *E. coli* PNP [14].

The structure of *E. coli* PNP is interesting for mechanistic reasons, but it also has practical implications. For instance, a recently proposed gene therapy for human tumors takes advantage of the differences in specificity of human and *E. coli* PNPs [15]. Certain prodrugs, such as 2′-deoxy-6-methylpurine, are not cleaved by human PNP when they are injected into tumor cells. In tumor cells transfected with the *E. coli* PNP gene, however, the nontoxic 2′-deoxy-6-methylpurine is converted by bacterial PNP to the highly cytotoxic base, 6-methylpurine. The free purine base analog can diffuse freely out of the cell, thus causing toxicity to the transfected cells as well as nearby tumor cells.

We describe here the three-dimensional structure of *E. coli* PNP and compare it to the structures of mammalian PNP and *E. coli* uridine phosphorylase (UP). All three enzymes have similar subunit folds, although there is negligible similarity between the amino acid sequences of human and *E. coli* PNP and only about 20% identity between UP and *E. coli* PNP. All three enzymes have a similar geometric arrangement of the purine-, ribose- and phosphate-binding sites, however, *E. coli* and human PNPs use completely different amino acids within the binding sites. In addition, although both PNPs utilize residues from adjacent subunits, the participating residues are different (His4 and Arg43 in *E. coli* PNP and Phe159 in human PNP), owing to differences in quaternary structure. Finally, the base-binding site in *E. coli* PNP is much larger and more accessible than that of human or bovine PNP.

Results and discussion

Quaternary structure

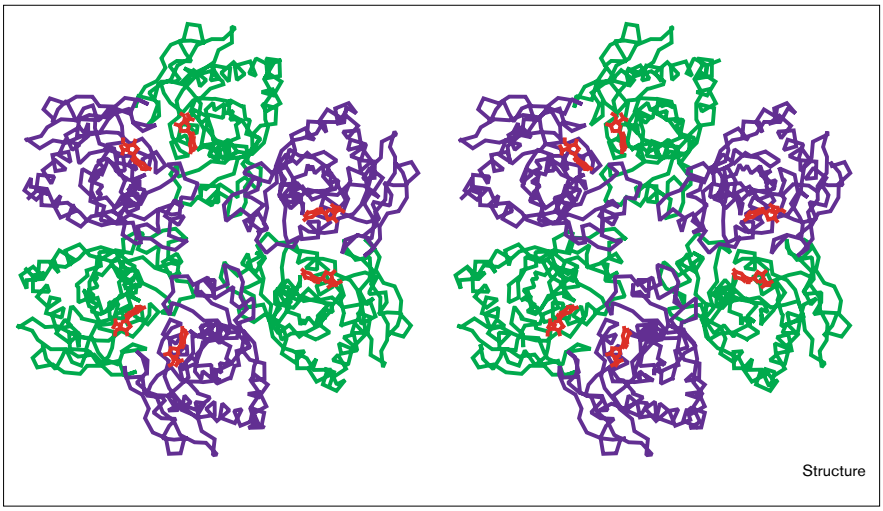
E. coli PNP is a disc-shaped hexamer with D_3 symmetry as shown in Figure 1. The hexamer is approximately 60 Å thick in the direction of the threefold axis and about 100 Å in diameter. The *E. coli* PNP hexamer can be viewed as a trimer of dimers, in which each dimeric pair contains two complete active sites (Figure 2). The contacts between subunits within the dimers are more extensive than the contacts between threefold-related pairs. Two adjacent subunits are involved in the formation of each active site: Arg43 from one subunit probably participates in binding the phosphate via both hydrogen-bonding and electrostatic interactions, and His4 probably forms one or two hydrogen bonds with the hydroxyl groups of the ribose. The mutation of Arg43 to lysine destroys PNP activity, indicating the importance of this subunit contact [16].

In human or calf spleen PNP [1–4], three identical subunits form a triangular prism with C_3 symmetry that is 45 Å thick in the direction of the threefold axis and 75 Å along the sides. In mammalian PNPs, the active site also involves the participation of two subunits. Although subunit contacts are essential to the formation of the active sites in both types of PNP, the differences in the surface loops appear to modulate the quaternary structure. Thus, the position of subunits in the hexamer with respect to the symmetry axis is entirely different from the trimer.

Secondary structural features

E. coli PNP has a central eight-stranded mixed β sheet with strands labeled β1 to β7 and β9 (β8 is a three-residue strand that forms several hydrogen bonds with one end of β7, and β10 is a two-residue strand that forms several hydrogen bonds with one end of β9; Figures 3 and 4). Several α helices pack against the core of the β-sheet structure. Figure 5 shows a sequence alignment according to a manual superposition of the structures from *E. coli* PNP, human PNP and *E. coli* UP. *E. coli* PNP shows no

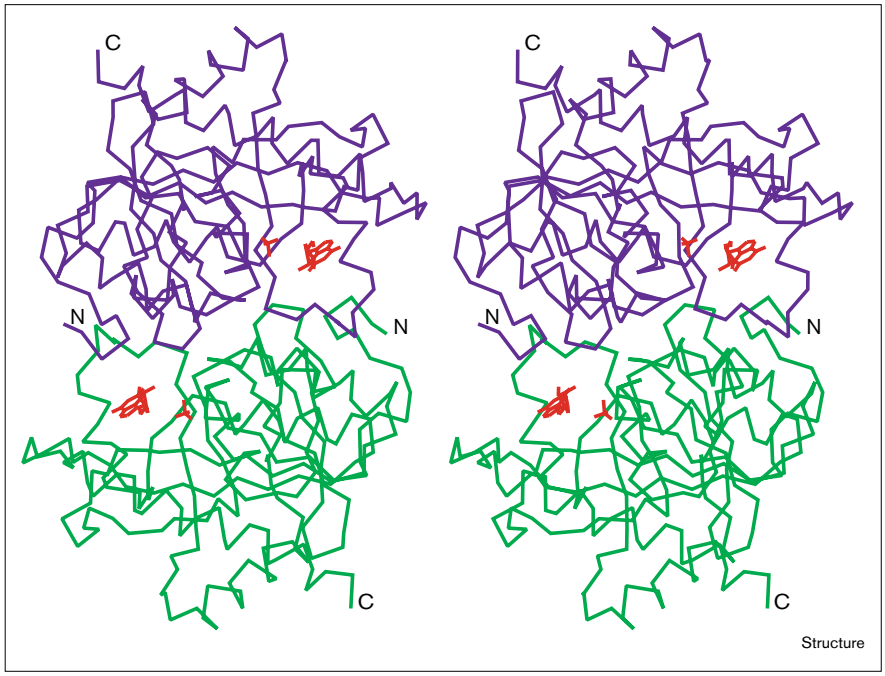
Figure 1



Stereo view of the *E. coli* PNP hexamer viewed down the noncrystallographic threefold axis. Models of inosine and phosphate (red) are shown in the proposed active sites. (The models of inosine and phosphate included in Figures 1–4, 6 and 7 are based on a comparison with structures of human PNP with bound substrates.) The subunits related by the threefold axis have the same color. The subunits in each green and blue dimeric pair are related to each other by horizontal noncrystallographic twofold axes. (The figure was made using the program CHAIN [39].)

Figure 2

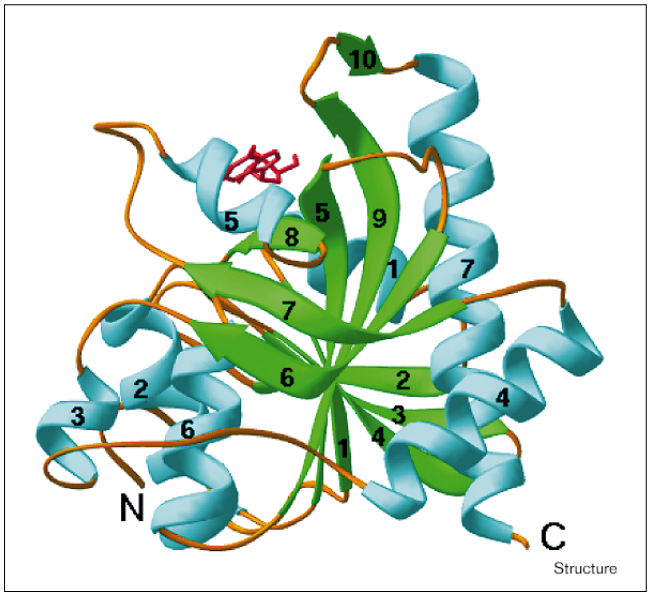
Stereo view of two *E. coli* PNP subunits with bound substrate viewed down the noncrystallographic twofold axis. Models of inosine and phosphate (red) are shown in the proposed active sites; the N and C termini are marked. (The figure was made using the program CHAIN [39].)



apparent sequence similarity to human PNP, but it shows about 20% identity with UP [17], which is also hexameric [18,19]. The similarity of the overall topology of the α helices and the β sheet in the structures of both PNPs and UP is illustrated in Figure 6. As expected, *E. coli* PNP and UP show structural homology. In addition, even though the sequences of *E. coli* and human PNP show little similarity, the secondary structures also correlate quite well with each other. All three structures have a central eight-stranded mixed β sheet, as well as a three-residue strand (β8) that forms several hydrogen bonds with one end of β7. In each molecule, the C-terminal helix is the longest helix. None of the enzymes contain the classical βαβαβ topology characteristic of the Rossmann fold [20], even though they have both phosphate- and nucleoside-binding sites. Many of the residues in the active site are close to the middle four strands in the eight-stranded β sheet. These four strands are parallel to each other with their C-terminal ends pointing towards the active site. In *E. coli* PNP and UP, the central β sheets are flanked by seven α helices, while human PNP has one additional helix near the N terminus (α0 in Figure 5). Neither human PNP or UP have a short β strand corresponding to β10 in *E. coli* PNP. Superposition of the three structures based on the Cα positions of the 64 residues in the nine common β strands gives root mean square (rms) differences of 1.3 Å for human PNP, 1.3 Å for bovine PNP and 1.0 Å for UP. Including the five common α helices in the superposition (a total of 133 residues) gives rms differences of 2.0 Å for human PNP, 2.1 Å for bovine PNP and 1.3 Å for UP.

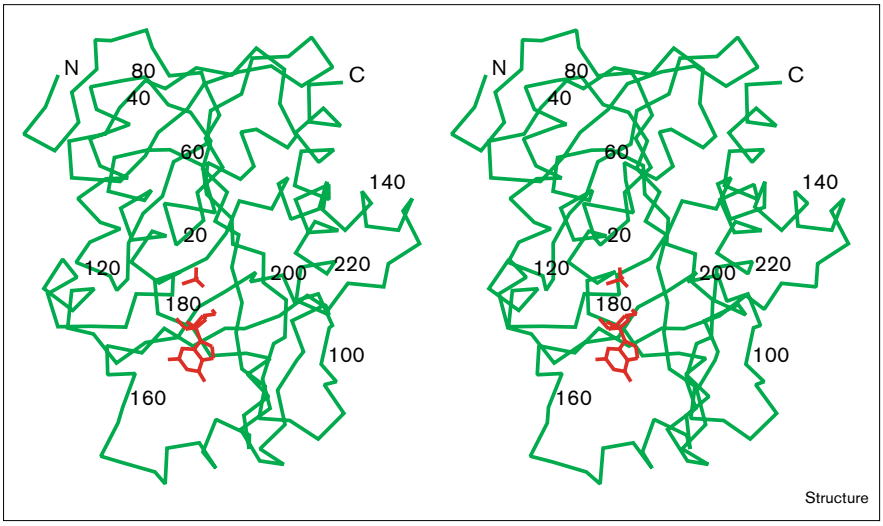
In *E. coli* PNP, the β-strand content is about 27%, compared with 23% in human PNP, and the α-helical content

Figure 3



The overall fold of the *E. coli* PNP subunit. The β strands (green) and α helices (blue) are labeled. Models of inosine and phosphate (red) are shown in the proposed active sites. (The figure was made using the program RIBBONS [40].)

Figure 4



Stereo view of a Cα tracing of the *E. coli* PNP monomer with every twentieth residue labeled. Models of inosine and phosphate are shown in red. (The figure was made using the program CHAIN [39].)

is 34% versus 33% in human PNP. Overall, *E. coli* PNP has 5% more secondary structure than human PNP, even though the *E. coli* subunit has only 238 amino acids compared to 289 in the human PNP subunit. Human PNP has proportionally more loops and turns. As expected, many of these turns and loops are involved in subunit–subunit contacts or contribute to active sites. In other words, the similar β-sheet and α-helical arrangement of these two enzymes has little effect on the details of the active sites. Instead, the variation in active sites comes from the differences in loops and turns. While the core of the secondary structure remains more or less constant, the loops show one of two patterns, depending on the quaternary structure. In the case of the hexameric enzymes, the loops are

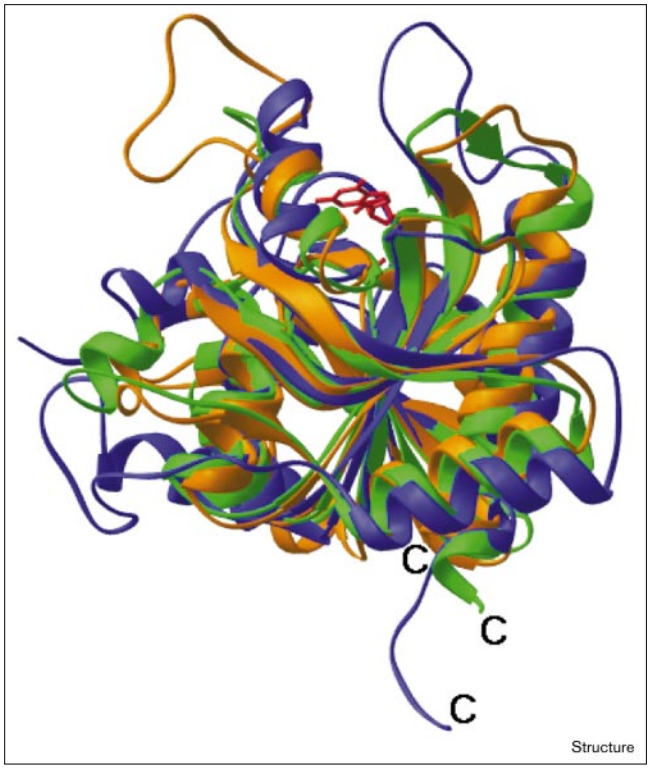
Figure 5

```
            ├─────α0─────┤        ├β1┤  ├─α1─┤    ├─β2─┤
EPNP              ATPHINAEMGDFADVVLMPGDPLRAKYIAETFLEDAREVNNVRGML....  46
HPNP MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGS...GLGGLTDKLTQAQIFDYSEIPNFPRS  59
UP         MSKSDVFHLGLTKNDLQGATLAIVPGDPDRVEKIAALM.DKPVKLASHREFT....  51


             ├─β3─┤ ├β4─┤            ├─────α2──────┤    ├─β5─────┤
EPNP ........GFTGTYKGRKISVMGHGMG......IPSCSIYTKELITDFGVKKIIRVGSCGAV  94
HPNP TVPGHAGRLVFGFLNGRACVMMQGRFHMYEGYPLWKVTFPVRVFHLL.GVDTLVVTNAAGGL 120
UP   ........TWRAELDGKPVIVCSTGIG......GPSTSIALEELAQL.GIRTFLRIGTTGAI  98


             ├──β6──┤  ├α3┤                          ├──────α4──────┤
EPNP LPHVKLRDVVIGMGACTDSKVNRIRFKDHDFAAIAD...........FDMVRNAVDAAKALG 145
HPNP NPKFEVGDIMLIRDHINLPGFSGQNPLRGPNDERFGDRFPAMSDAYDRTMRQRALSTWKQMG 182
UP   QPHINVGDVLVTTASVRLDGASLHFAPLEFPAVAD...........FECTTALVEAAKSIG 148


          ├─β7─┤                            ├─α5─┤   β8 ┤  ├─α6─┤
EPNP ID..ARVGNLFSADLFYSPDGE..............MFDVMEKYGILGVEMEAAGIYGVAAE 191
HPNP EQRELQEGTYVMVAGPSFETVA..............ECRVLQKLGADAVGMSTVPEVIVARH 230
UP   AT..THVGVTASSDTFYPGQERYDTYSGRVVRHFKGSMEEWQAMGVMNYEMESATLLTMCAS 208


        ├──β9──┤     β10          ├──────────α7──────────┤
EPNP FGKALTICTVSDHIRTHEQTT........AAERQTTFNDMIKIALESVLLGDKE          238
HPNP CGLRVFGFSLITNKVIMDYESLEKANHEEVLA AGKQAAQKLEQFVSILMASIPLPDKAS   289
UP   QGLRAGMVAGVIVNRTQQEIPN........AETMKQTESHAVKIVVEAARRLL           253
```

Structure

Sequence alignment of *E. coli* PNP (EPNP), human PNP (HPNP) and *E. coli* uridine phosphorylase (UP) based on a superposition of the corresponding α helices and β strands. Symbols α and β represent the secondary structure elements for the three proteins. The helix α0 is present only in human PNP, α3 is not present in human PNP, and β10 is present only in *E. coli* PNP. Gaps introduced into the sequences for the purpose of alignment are represented by the dotted regions.

Figure 6



Superimposed structures of *E. coli* PNP (green), human PNP (blue) and *E. coli* UP (orange). The orientation is the same as in Figure 3. The superposition is based on 64 residues in the nine common β strands of each structure. Models of inosine and phosphate are shown in red. (The figure was made using the program RIBBONS [40].)

generally of similar length, while in trimeric mammalian PNP the loops are significantly different.

Subunit contacts

The residues of *E. coli* PNP involved in interface contacts are 3–4, 20–23, 42–43, 64–65, 67–68, 71–72, 158–160, 162, 180, and one side of helix 112–119. The nature of these interactions is mainly hydrophobic, although a few hydrogen bonds mediate polar–polar interactions. The active sites of adjacent subunits are separated by a pair of loops: Gly63-Met64-Gly65-Ile66-Pro67-Ser68 and Ile71-Tyr72. The first is a glycine-rich turn from the β4 strand to the α2 helix and passes near the phosphate-binding site. Although the loop is not directly involved in phosphate binding, it is conserved between *E. coli* PNP and UP [18,19]. On the other hand, *E. coli* PNP and UP do differ in the nature of a turn (β2–β3) which bridges each pair of twofold related phosphate-binding sites: Val42-Arg43 in *E. coli* PNP and His47-Arg48 in UP. Arg43 in *E. coli* PNP is able to extend into the phosphate-binding site of the other subunit, whereas the sidechain of Arg48 in UP is disordered and not defined in the crystal structure.

The active site

Complexes of *E. coli* PNP with sulfate, ribose-1-phosphate, and 6-iodopurine have been examined by difference Fourier methods. Although electron density for each of these small molecules was apparent in the difference maps, in each case the electron density was weak and poorly defined so accurate placement of the substrates was impossible. Comparison of these structures with human PNP, bovine PNP and *E. coli* UP, however, has allowed the development of a model for all substrate-binding positions. If the subunits of trimeric PNP, hexameric PNP and UP are superimposed by aligning the secondary structure elements, then the purine- (or pyrimidine in the case of UP), ribose- and phosphate- binding sites also superimpose (Figure 6). The most conserved feature is the secondary structure near the substrate. The residues in *E. coli* PNP that are probably involved in substrate binding are listed in Table 1 and shown in Figure 7a; for comparison, the residues in human PNP that are involved in substrate binding are also given. In each active site, part of the β5 strand stretches from the phosphate-binding site to the base-binding site. The portion of β5 involved includes residues 90–93 (Ser–Cys–Gly–Ala) for *E. coli* PNP, 116–119 (Ala–Ala–Gly–Gly) for human PNP, and 94–97 (Thr–Thr–Gly–Ala) for UP. Interestingly, this subtle structural homology was detected prior to the structure

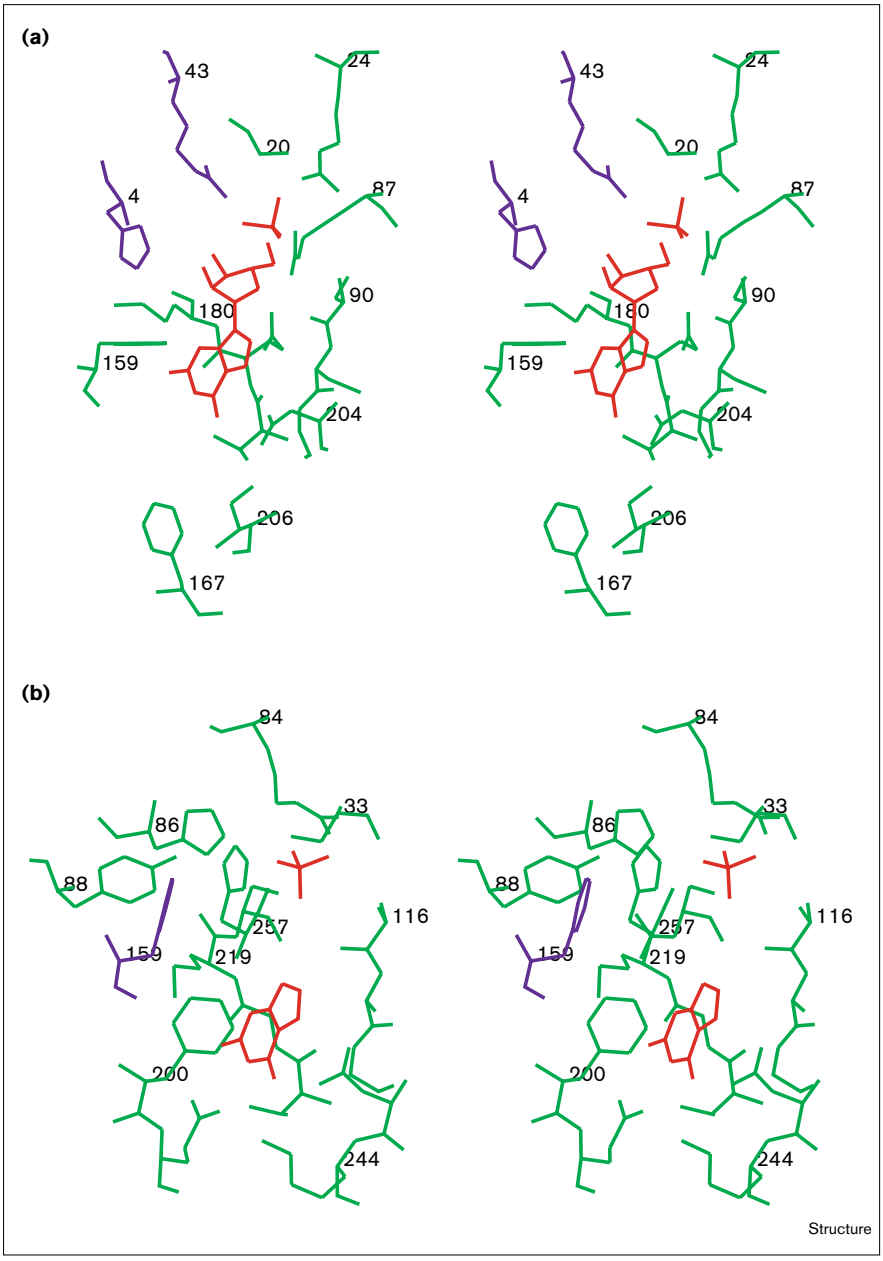
Table 1

Residues in the active sites of *E. coli* and human PNP.

	E. coli PNP	Human PNP
Phosphate	Gly20	Ser33
	Arg24	Arg84
	Arg87	His86
	Ser90	Ala116
	Arg43*	Ser220
Base	Ser90	Ala116
	Cys91	Ala117
	Gly92	Gly118
	Phe159	Phe200
	Phe167	Glu201
	Val178	Val217
	Glu179	Gly218
	Met180	Met219
	Asp204	Asn243
	Ile206	Lys244
		Phe159*
Ribose	Met64	His86
	Phe159	Tyr88
	Val178	Phe200
	Glu179	Val217
	Met180	Gly218
	His4*	Met219
		His257
		Phe159*

*The residue is contributed by an adjacent subunit.

Figure 7



Stereo view of the active sites in *E. coli* and mammalian PNPs. **(a)** In the proposed active site of *E. coli* PNP, fourteen of the sixteen residues are from one subunit (green), and two (His4 and Arg43) are from an adjacent subunit (blue). Models of inosine and phosphate are shown in red. **(b)** In the active site of human PNP with bound phosphate and guanine (PDB entry code 1ULB), sixteen of the seventeen residues are from one subunit (green), and one (Phe159) is from the adjacent subunit (blue). Guanine and phosphate are shown in red. (The figure was made using the program CHAIN [39].)

determination, by Mushegian and Koonin using computer assisted sequence analysis [21].

The relationship between the neighboring active sites is completely different in hexameric *E. coli* PNP and trimeric mammalian PNP. In the *E. coli* PNP hexamer, the distance between the closest phosphate centers is about 22 Å, which is much shorter than the value of 35 Å for trimeric PNP [1–4]. The active sites in the mammalian PNP are situated near the equator of the trimer. In contrast, when *E. coli* PNP is viewed down the threefold axis, three active sites are located on top of the molecule and the other three, related by the twofold axis, lie on the bottom of the molecule. The closest pairs of phosphate sites are related by the twofold axes. It is possible that communication between the phosphate sites in each pair is responsible for the strong negative cooperativity observed for *E. coli* PNP [5].

Phosphate-binding site

According to our model, the phosphate-binding site of *E. coli* PNP includes Arg24 and Arg87 from one subunit

and Arg43 from a neighboring subunit (the corresponding residues in human PNP are Arg84, His86 and Ser220); all of these residues form sidechain interactions with the phosphate oxygen atoms (Table 1; Figure 7). In *E. coli* PNP, both sidechain and backbone atoms from Ser90 also contribute to phosphate binding forming hydrogen bonds with the phosphate ion. In addition, an arginine residue at the beginning of helix α7, Arg217, approaches the phosphate-binding site, but the guanido group is probably at least 8 Å from the closest phosphate oxygen.

Although different residues are used by mammalian and *E. coli* PNPs for phosphate binding, several similarities exist. Both phosphate-binding sites are located in the same region relative to the secondary structure elements and probably use residues from the β1–α1 turn. In mammalian PNP two hydrogen bonds are formed with the phosphate anion by mainchain nitrogen atoms in the β5 strand (Ala116) and the β1–α1 turn (Ser33); in *E. coli* PNP the corresponding residues are Gly20 and Ser90. Similarly, in mammalian PNP one oxygen of the phosphate anion forms a hydrogen bond to a water molecule associated with the carbonyl oxygen of Ala116 in the β5 strand; the carbonyl oxygen of Ser90 in *E. coli* PNP is positioned such that a similar interaction could occur if the phosphate oxygen is protonated. The Ser90 sidechain interaction, however, has no counterpart in mammalian PNP.

Even though there is no typical phosphate-binding site when compared to other phosphate-binding proteins [22], *E. coli* and mammalian PNPs belong to the group of phosphate-binding proteins in which phosphates are found in close proximity to the N terminus of an α helix (α1). Both PNPs utilize the most commonly found residues at phosphate binding sites, which are glycine, arginine, threonine and serine. Because of the presence of histidine residues, the phosphate-binding site in mammalian PNP is unique when compared to phosphate-binding sites that utilize helix-type dipoles [22].

Base-binding site

The base-binding site in *E. coli* PNP appears to consist primarily of hydrophobic residues (Table 1; Figure 7). Three of the hydrophobic residues utilized in *E. coli* PNP (Val178, Met180 and Phe159) correspond to residues Val217, Met219 and Phe159 (from an adjacent subunit) in human PNP. However, two other hydrophobic residues in the *E. coli* PNP base-binding site (Phe167 and Ile206) do not have counterparts in the human PNP active site. Instead, the corresponding residues are hydrophilic (Glu201 and Lys244), and they contribute hydrogen bonds to N1, the amino group at the C3 position, and the carbonyl oxygen at the C6 position of the base. The mainchain atoms of residues 90–92 (within the β5 strand) in *E. coli* PNP also form part of the base-binding pocket. In human PNP similar interactions are provided by residues 116–118, which are also part of the β5 strand. There are two additional hydrophobic residues in the base-binding pocket of *E. coli* PNP (Phe167 and Ile206) and one in human PNP (Phe200) which have no counterparts in the other structure. Asp204 of *E. coli* PNP, which corresponds to Asn243 in human PNP and Asn222 in UP, is located near N7 of the purine base and may be important for catalysis. The arrangement of the active-site residues is illustrated in Figure 7.

This base-binding pocket of *E. coli* PNP differs from bovine or human PNP in several significant respects. Firstly, the strong complementary hydrogen-bonding pattern provided by Glu201 (in the β7–α5 loop) and Asn243 (in the β9 strand) of mammalian PNP (and similar residues in other purine base binding proteins such as adenosine deaminase and GTP-binding protein [23]) is not found in the *E. coli* enzyme. The β7–α5 loop in *E. coli* PNP has swung away from the base-binding site by about 7–8 Å relative to mammalian PNP. Asn243 in the β9 strand of mammalian PNP probably forms hydrogen bonds with N7 of the purine base and OG1 of Thr242. The corresponding residue in *E. coli* PNP (Asp204) would have to be protonated in order to provide similar interactions with N7 and OG of Ser203. Secondly, the base-binding and nucleoside-binding sites are remarkably more open in *E. coli* PNP, as compared with bovine or human PNP. This observation may explain the much broader specificity of the *E. coli* enzyme and the higher efficiency of the mammalian enzyme. Finally, in mammalian PNP residues 257–261 (part of the large disordered β9–α7 loop consisting of residues 241–260) is transformed into an ordered α helix upon base or nucleoside binding, and this conformational change closes the sugar-binding site. This change is not likely to occur in *E. coli* PNP, because the corresponding loop is eight residues shorter and is well ordered in the purine-free native structure.

Ribose-binding site

The proposed mechanism of ribose-binding in *E. coli* PNP is similar to that described for mammalian PNP, in that both enzymes primarily use hydrophobic van der Waals contacts (Table 1; Figure 7). The β8–α6 region (shown in Figures 3 and 4), consists of residues Val178-Glu179-Met180 (*E. coli* PNP) and Val217-Gly218-Met219 (human PNP); the methionine residue lies across the hydrophobic face of the ribosyl group and has been identified in all known PNP sequences.

In human or bovine PNP, Phe159 (located on a loop from an adjacent subunit) participates in the active site by completing a hydrophobic patch near the ribosyl group. In *E. coli* PNP, the absence of such a loop and the different subunit arrangement alter this interaction feature. However, Phe159 in the *E. coli* enzyme, which

coincidentally has the same residue number, occupies the same relative position in the ribose-binding site as Phe159 from the adjacent subunit in mammalian PNP.

As noted above, binding of the substrate causes movement of a large loop (residues 241–260) in mammalian PNPs. The maximum movement occurs at His257, which is displaced outwards by 5–6 Å in human PNP and 2–3 Å in bovine PNP, and it allows the formation of a hydrogen bond with the O5′ hydroxyl group of the ribose. This loop is not present in *E. coli* PNP, but His4 in an adjacent subunit occupies a similar position in the ribose-binding pocket and would probably be able to form a hydrogen bond to the O3′ hydroxyl group of the ribose.

Mechanism of catalysis

The available structural information from *E. coli* PNP complexed with sulfate and ribose-1-phosphate indicates that the phosphate-binding site is similar for both reactant and product. In studies of bovine PNP (CM *et al.*, unpublished data) a ribosyl group was found binding the enzyme in only two situations: as part of the nucleoside (with or without phosphate) or as part of ribose-1-phosphate only in the presence of purine base. In the mammalian enzyme, ribose-1-phosphate binding is probably facilitated by a purine-induced conformational change that brings His257 into position to form a hydrogen bond with O5′ of the ribose. In *E. coli* PNP, however, no evidence exists for a loop→helix transition or other significant conformational changes. Furthermore, the *E. coli* PNP complex with ribose-1-phosphate could be prepared in the absence of a purine base. Binding of either purine base or ribose-1-phosphate alone is consistent with a random bi-bi mechanism in the synthetic direction, as has been proposed from inhibition studies [24,25].

The negatively charged nature of the phosphate has a great impact on the catalytic mechanism in both mammalian and *E. coli* PNPs. In both cases positively charged residues are spread on one hemisphere of the phosphate-binding site, away from the ribose and base-binding sites and interact with the phosphate. The opposite side of the phosphate is left exposed to the rest of the active site. This exposed side of the phosphate ion will certainly alter the electrostatic properties of the nucleoside after binding. In mammalian PNPs, there are six to eight hydrogen bonds between the phosphate oxygen atoms and the protein or water molecules. Although the binding site is not as clear in *E. coli* PNP, three of the four phosphate oxygen atoms probably form a total of at least five hydrogen bonds with Gly20 N, Arg24 NH_2, Ser90 N, Ser90 OG and Arg87 NH1, while the oxygen atom that forms a bond with the ribosyl C1′ has no obvious hydrogen-bond donors or acceptors. The strong interactions around the phosphate ion in each case suggest that the ribosyl group rather than the phosphate ion moves during the chemical reaction.

Bovine PNP hydrolyzes nucleoside with about 100-fold lower efficiency compared with phosphorolysis [26]. Inosine hydrolysis in mammalian PNP is thought to occur when His64 swings away from the active site, thus exposing the ribosyl group [2]. In comparison the phosphate binding site in *E. coli* PNP appears to be rigid with Arg43 of the neighboring subunit occupying the closed position of His64. This observation suggests that the ribosyl group will not be easily accessible by solvent and therefore resistant to hydrolysis. To date, no reports of nucleoside hydrolysis by *E. coli* PNP have appeared in the literature.

Biological implications

Purine nucleoside phosphorylase (PNP) from *E. coli* is a hexameric enzyme that catalyzes the reversible phosphorolysis of 6-amino and 6-oxopurine (2′-deoxy)ribonucleosides to the free base and (2′-deoxy)ribose-1-phosphate. In contrast, mammalian (e.g. human and bovine) PNPs are trimeric and accept only 6-oxopurine nucleosides as substrates.

Sequence alignment between human PNP and *E. coli* PNP shows little similarity with only 11% apparent sequence identity [27]. Nevertheless, even though the quaternary structures of the enzymes differ and there is no obvious sequence similarity, this study has shown that both *E. coli* and mammalian PNPs have the same subunit topology. Furthermore, both enzymes catalyze the same reaction, even though there are differences in the amino acids within the active site as well as differences in substrate specificity. The fact that the chemical reaction type and topology are conserved, while quaternary structure, specificity and reaction mechanism vary between trimeric and hexameric PNPs, is truly remarkable.

It has been shown that all available sequence data for PNPs show similarity to either human PNP or *E. coli* PNP. Consequently, the structures of these two enzymes define two possible classes that probably encompass most PNPs. Furthermore, the crystal structure of *E. coli* PNP described here reveals similar subunit topology and a similar active-site location with *E. coli* uridine phosphorylase, which catalyzes a different reaction.

The structural data reported here may have important implications in the design of prodrugs for use in gene therapy. For example, *E. coli* PNP (but not human PNP) is capable of catalyzing the conversion of several nontoxic deoxyadenosine analogs to highly toxic adenine analogs. Recently, Sorscher *et al.* [15] utilized this difference in specificity to study tumor cell killing *in vitro* by cells transfected with the *E. coli* PNP gene. In their study, the expression of *E. coli* PNP in <1% of the cells in a human colonic carcinoma cell line led to the death of virtually all bystander cells after treatment with 6-methylpurine-2′-deoxyribonucleoside, a deoxyadenosine analog

that is a substrate for *E. coli* PNP but not human PNP. Furthermore, the *E. coli* PNP structure may suggest mutant PNP enzymes with more effective activation properties. Such studies may eventually lead to development of a toxin-mediated gene therapy for human malignancy.

Materials and methods

Crystallization and data collection

Purified protein was provided by Wellcome Research Laboratories. Crystals of the native protein were grown by the hanging-drop method in 12–14% PEG 4000 with 50 mM citrate buffer (pH 4.8–5.0). With the help of micro seeding techniques, crystals as large as $0.6 \times 0.6 \times 0.4$ mm were obtained after one week. The crystals are tetragonal rods with space group $P2_1$ and unit cell constants a = 89.5 Å, b = 111.5 Å, c = 74.8 Å and $\beta = 110.7°$. There is one hexamer per asymmetric unit. Based on a molecular weight of 155,712 Da, as predicted from the cDNA, the value of V_m [28] is 2.24, which corresponds to a solvent volume fraction of 45%.

Prior to data collection the crystals were kept in a stabilizing mother liquor that contained 20–25% PEG 4000. Data were collected at room temperature with a San Diego Multiwire Area Detector using CuKα radiation from a Rigaku RU-200 rotating-anode generator operating at 40 kV and 100 mA. The twin detectors were set at distances of 500 mm and 560 mm with two-theta angles asymmetrically offset by 30° and 23° from the center. Native crystals diffracted beyond 2.0 Å resolution. Potential heavy-atom derivatives were prepared by soaking native PNP crystals in a stabilizing solution containing the heavy atom at a concentration of 1 mM or saturated, whichever was less. Potential substrate–inhibitor complexes were prepared in a similar manner. All soaks were for 24 h. Of a dozen possible heavy-atom derivatives that were analyzed, only 6-iodopurine and K_2PtCl_6 were useful. Table 2 contains a summary of the data collection statistics for the native and heavy-atom derivative crystals.

Initial determination of phases

Calculation of a self-rotation function using the program AMORE [29] demonstrated a noncrystallographic threefold axis and three twofold peaks perpendicular to the threefold axis. These results were consistent with a hexameric structure with D_3 symmetry. As UP from *E. coli* has a hexameric structure with the same symmetry and about 20% sequence identity with *E. coli* PNP, it seemed likely that it would serve as a molecular replacement (MR) model. A translation search on the x and z axes was performed using AMORE and XPLOR [30] and data in the resolution range 15.0 Å – 6 Å. The UP hexamer was used as the search model, after all sidechains were truncated after the Cβ atom. The correct position of the model corresponded to the highest peak in the list of possible solutions (correlation coefficient of 37% and R factor of 50%). The R factor of the MR model increased to a random value around 6.5 Å resolution, so the model phases were only useful at resolutions lower than 6.5 Å. Subsequent analysis with the SIGMAA program [31] gave similar results.

Using the initial phases provided by the UP polyalanine model, difference Fourier analyses were employed on the potential heavy-atom derivatives. For the 6-iodopurine and K_2PtCl_6 derivatives, the top six peaks in difference Fourier maps calculated at 6 Å resolution indicated the same position on each subunit. The heavy-atom positions were refined using the PHASES program [32] and then used to generate the first set of MIR phases, which was very poor even at 6.0 Å. After map modification by symmetry-averaging, however, the protein-solvent boundary was clearly distinguishable.

Phase improvement and phase extension

Because of the poor low-resolution phases and the lack of useful phase information beyond 6.0 Å, a special phase extension procedure was used. Beginning at 6.0 Å resolution the phases were extended to 2.5 Å using symmetry-averaging and solvent-leveling, which led to high-resolution maps for model building. The density modification was performed on $2F_o–F_c$ maps instead of F_o maps. In the later stages, a $(2mF_o–DF_c)$ map weighted by SIGMAA [32] was used for symmetry-averaging when partial models became available. Calculations were performed using the CCP4 package [33] and the RAVE program suite [34,35].

The noncrystallographic symmetry (NCS) matrices were calculated from the self-rotation search result, and the center of the symmetry axis was calculated from the six heavy atom peak positions and confirmed by the UP model. A mask in the $P2_1$ space group was generated by using the UP polyalanine model with a 3.5 Å radius from the center of each atom. Electron densities were set to zero outside the mask, while inside the mask the densities were replaced by their averaged values from the six symmetry-related regions. Using the molecular isomorphous replacement (MIR) map calculated from 6-iodopurine and K_2PtCl_6 derivatives at 6 Å resolution, an extension from 6 to 3 Å resolution was performed in 26 steps. Each phase extension step included a new resolution bin which was calculated based on equally dividing the reciprocal D spacing in the resolution range from 6.0 to 3.0 Å, so that a roughly equal number of reflections was included in each step. The R factor and correlation coefficient were 0.30 and 0.65, respectively, after the extension. Although the phases were calculated to 3 Å resolution, the best map was at 4 Å resolution.

Model building as part of the phase extension procedure

The resulting 4 Å map showed traceable electron densities. The continuous β strands were resolved, α helices were clearly visible, and some large sidechains could be seen. Based on this map, continuous Cα models were easily built from the N terminus to the C terminus for each subunit. Even though some connections in loop regions were not clear, most of the chain direction was defined with confidence. Immediately after building this model, the function lego in the program O was used

Table 2

Summary of data collection for the native crystals and heavy-atom derivatives.

Data set	Resolution (Å)	Total observations	Unique reflections*	Completeness (%)	R_{sym}†	$\langle I/\sigma \rangle$
Native‡	2.0	230,200	70,600	70	8.7	5.3
6-Iodopurine	2.2	241,169	55,185	84	11.7	5.2
K_2PtCl_6	2.2	211,058	52,310	79	15.0	4.0

*The number of unique reflections is based on $F > 2\sigma$.
†$R_{sym} = \Sigma\Sigma \left| I - I_i \right| / \Sigma I$, where I is the mean intensity of the N reflections with intensities I_i and common indices *h,k,l*. ‡The native data were collected on a crystal in a mother liquor containing sodium tungstate in an attempt to make a tungstate derivative. When it became apparent that there was no tungstate in the derivative, the data were treated as native. The completeness of the native data is about 90% in the 2.6 Å resolution bin, but drops to only 67% in the 2.0 Å resolution bin.

Table 3

Refinement statistics.

Refinement	
resolution range (Å)	8.0–2.0
number of reflections (F>2σ)	70,600
R factor (%)	20
R free (%)	29
Model	
number of protein atoms*	10,764
number of water molecules	480
Thermal parameters	
<B> for protein atoms (Å^2)	11
<B> for water molecules (Å^2)	19
Stereochemistry	
rms deviation bond lengths (Å)	0.005
rms deviation bond angles (°)	1.3
rms deviation dihedral angles (°)	23.8
rms deviation improper angles (°)	1.1
Ramachandran plot	
residues in most favored regions (%)	88.3
residues in additional allowed regions (%)	11.5
residues in generously allowed regions (%)	0.2
residues in disallowed regions (%)	0.0

*The sidechain of Lys237 in each subunit was truncated after the Cβ atom.

to complete the backbone from the Cα positions [36]. Cβ carbons were assigned to every Cα carbon, and many major fitting errors were corrected by visual inspection. At this stage, the number of protein atoms included in the model was about half of the total in the molecule.

After one round of positional refinement of this polyalanine model in XPLOR, the R factor dropped to 44%. The partial model from positional refinement was used to calculate a map at 5.0 Å resolution, and the initial MIR phases were abandoned. The phase extension through density modification was applied again for the new map in the resolution range from 5.0 Å to 3.0 Å. This 3 Å map showed many well defined sidechains. The first 100 residues were fit smoothly, but the middle portion of the polypeptide chain was ambiguous. At this point, having developed confidence in this procedure, the phase improvement and extension procedures were repeatedly used with the updated models to get better maps. Model building was carried out conservatively to avoid adding incorrect information, and the protein-solvent mask and NCS matrices were slowly improved during extension to higher resolution. This process led to a steady improvement in the map until the nearly complete sequence was visible.

Refinement

The refinement of *E. coli* PNP in space group $P2_1$ required careful treatment of each individual subunit. The complete molecule contains six subunits with a total of 1422 residues in one asymmetric unit. One might expect that some of the residues in each subunit would be different, especially at high resolution, as a result of different crystallographic environments. Therefore, a mixture of several methods was employed in the refinement. Initially, positional and simulated-annealing refinement was performed on the model, and the averaged difference maps ($2F_o–F_c$ and $F_o–F_c$) were calculated as references for model adjustment. Using the updated refined models and NCS, the phase extension cycles were started at a higher resolution. Better maps were generated in this back and forth manner.

Refinement of the complete model of *E. coli* PNP was begun with strict NCS constraints at 3.0 Å resolution, but after initial refinement the constraints were dropped. After adjusting the model according to the averaged $F_o–F_c$ and $2F_o–F_c$ maps, the model was refined against the native data at 2.0 Å resolution. The R factor at this point was 26%, and the free R factor was 34%. Analysis by PROCHECK [37] indicated that about 10% of the total residues needed to be adjusted. From the refined coordinates of six subunits, NCS matrices were recalculated. The Cα rms deviations were invariably around 0.5 Å, and a deviation of 0.5° from a strict twofold or threefold was found in the subunit–subunit relationship. The R factor distribution showed that the R factor increased around 3 to 3.5 Å. Thus, the model coordinates were used to calculate a map at 3.5 Å, and the map was subjected to phase improvement and phase extension from 3.5 Å to 2.5 Å. After refitting of the model using this map, the model was refined using the data to 2.0 Å resolution. Although group B factors were used at lower resolution (3.0 Å), analysis of the free R factor at high resolution indicated that individual B factors gave a better refinement, so individual B factors were included in the final refinement. Water molecules were added automatically by a series of small programs written to search the peaks in the $F_o–F_c$ map and check distance criteria for reasonable hydrogen-bond donors and acceptors. After refinement, only water molecules with a B factor less than 40 Å^2 and occupancy larger than 50% were included.

The final model includes 1422 residues (237 residues out of a reported 238 per subunit [38]) and 480 water molecules. The C-terminal residue in each subunit (Glu238) was not present in the electron-density maps. The R factor is 20%, and the free R factor is 29%; the free R factor is only 25% at 3.0 Å resolution. Table 3 lists the refinement statistics. The rms deviations between subunits for Cα atoms range from 0.3 to 0.4 Å with the largest differences occurring in the loop that includes residues 210–216. This loop is on the surface of the molecule and is not involved in any intramolecular contacts, nor does it appear to function in the active site. A Ramachandran plot showed that 88.3% of the residues were located in most favored regions (range 87.4% to 90.3% for the six subunits), and only two residues (0.2%) were in generously allowed regions; none were in disallowed regions (Table 3).

Accession numbers

The atomic coordinates of the refined structure have been deposited with the Brookhaven Protein Data Bank (entry number 1ECP).

Acknowledgements

This work was supported by grants GM48874, RR01646 and CA67763 from the National Institutes of Health, the WM Keck Foundation and the Lucille P Markey Charitable Trust.

References

1. Ealick, S.E., *et al.*, & Bugg, C.E. (1990). Three-dimensional structure of human erythrocytic purine nucleoside phosphorylase at 3.2 Å resolution. *J. Biol. Chem.* **265**, 1812–1820.
2. Mao, C. (1995). Structure determination of purine nucleoside phosphorylase from bovine spleen and *Escherichia coli*, elucidation of reaction mechanism. PhD thesis, Cornell University, USA.
3. Bzowska, A., Luic, M., Schroder, W., Shugar, D., Saenger, W. & Koellner, G. (1995). Calf spleen purine nucleoside phosphorylase: purification, sequence and crystal structure of its complex with an N(7)-acycloguanosine inhibitor. *FEBS Lett.* **367**, 214–218.
4. Koellner, G., Luic, M., Shugar, D., Saenger, W. & Bzowska, A. (1997). Crystal structure of calf spleen purine nucleoside phosphorylase in a complex with hypoxanthine at 2.15 Å resolution. *J. Mol. Biol.* **265**, 202–216.
5. Jensen, K.F. & Nygaard, P. (1975). Purine nucleoside phosphorylase from *Escherichia coli* and *Salmonella typhimurium*: purification and some properties. *Eur. J. Biochem.* **51**, 253–265.
6. Koszalka, G.W. & Krenitsky, T.A. (1979). Nucleosidases from *Leishmania donovani*. Pyrimidine ribonucleosidase, purine ribonucleosidase, and a novel purine 2′-deoxyribonucleosidase. *J. Biol. Chem.* **254**, 8185–8193.
7. Cacciapuoti, G., Porcelli, M., Bertoldo, C., De Rosa, M. & Zappia, V. (1994). Purification and characterization of extremely thermophilic and thermostable 5′-methylthioadenosine phosphorylase from the archaeon *Sulfolobus solfataricus.* Purine nucleoside phosphorylase activity and evidence for intersubunit disulfide bonds. *J. Biol. Chem.* **269**, 24762–24769.

8. Senesi, S., Falcone, G., Mura, U., Sgarrella, F. & Ipata, P.L. (1976). A specific adenosine phosphorylase distinct from purine nucleoside phosphorylase. *FEBS Lett.* **64**, 353–357.
9. Jensen, K.F. (1978). Two purine nucleoside phosphorylases in *Bacillus subtilis.* Purification and some properties of the adenosine-specific phosphorylase. *Biochim. Biophys. Acta* **525**, 346–356.
10. Buxton, R.S., Hammer-Jespersen, K. & Valentin-Hansen, P. (1980). A second purine nucleoside phosphorylase in *Escherichia coli* K-12. I. Xanthosine phosphorylase regulatory mutants isolated as secondary-site revertants of a deoD mutant. *Mol. Gen. Genet.* **179**, 331–340.
11. Hori, N., Watanabe, M., Yamazaki, Y. & Mikami, Y. (1989). Purification and characterization of second thermostable purine nucleoside phosphorylase in *Bacillus stearothermophilus* JTS 859. *Agric. Biol.* **53**, 3219–3224.
12. Stoeckler, J.D., Cambor, C. & Parks, R.E., Jr. (1980). Human erythrocytic purine nucleoside phosphorylase: reaction with sugar-modified nucleoside substrates. *Biochemistry* **19**, 102–107.
13. Bzowska, A., Kulikowska, E. & Shugar, D. (1992). Formycins A and B and some analogues: selective inhibitors of bacterial (*Escherichia coli*) purine nucleoside phosphorylase. *Biochim. Biophys. Acta* **1120**, 239–247.
14. Krenitsky, T.A., Koszalka, G.W. & Tuttle, J.V. (1981). Purine nucleoside synthesis, an efficient method employing nucleoside phosphorylases. *Biochemistry* **20**, 3615–3621.
15. Sorscher, E.J., Peng, S., Bebok, Z., Allan, P.W., Bennett, L.L., Jr. & Parker, W.B. (1994). Tumor cell bystander killing in colonic carcinoma utilizing the *Escherichia coli DeoD* gene to generate toxic purines. *Gene Therapy* **1**, 233–238.
16. Hershfield, M.S., Chaffee, S., Koro-Johnson, L., Mary, A., Smith, A.A. & Short, S.A. (1991). Use of site-directed mutagenesis to enhance the epitope-shielding effect of covalent modification of proteins with polyethylene glycol. *Proc. Natl. Acad. Sci. USA* **88**, 7185–7189.
17. Walton, L., Richards, C.A. & Elwell, L.P. (1989). Nucleotide sequence of the *Escherichia coli* uridine phosphorylase (*udp*) gene. *Nucleic Acids Res.* **17**, 6741.
18. Zhao, B.-G. (1991). The crystal structure of uridine phosphorylase from *Escherichia coli* . PhD thesis, University of Alabama at Birmingham, AL, USA.
19. Morgunova, E.Y., *et al*., & Debabov, V.G. (1995) Atomic structure at 2.5 Å resolution of uridine phosphorylase from *E. coli* as refined in the monoclinic crystal lattice. *FEBS Lett.* **367**, 183–187.
20. Rao, S.T. & Rossmann, M.G. (1973). Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76**, 241–256.
21. Mushegian, A.R. & Koonin, E.V. (1994). Unexpected sequence similarity between nucleosidases and phosphoribosyltransferases of different specificity. *Protein Sci.* **3**, 1081–1088.
22. Copley, R.R. & Barton, G.J. (1994). A structural analysis of phosphate and sulphate binding sites in proteins. Estimation of propensities for binding and conservation of phosphate binding sites. *J. Mol. Biol.* **242**, 321–329.
23. Lambright, D.G., Noel, J.P., Hamm, H.E. & Sigler, P.B. (1994). Structural determinants for activation of the alpha-subunit of a heterotrimeric G protein. *Nature* **369**, 621–628.
24. Carlson, J.D. & Fischer, A.G. (1979). Thyroid purine nucleoside phosphorylase. II. Kinetic model by alternate substrate and inhibition studies. *Biochim. Biophys. Acta* **566**, 259–265.
25. Jensen, K.F. (1976). Purine-nucleoside phosphorylase from *Salmonella typhimurium* and *Escherichia coli*: initial velocity kinetics, ligand binding, and reaction mechanism. *Eur. J. Biochem.* **61**, 377–386.
26. Kline, P.C. & Schramm, V.L. (1993). Purine nucleoside phosphorylase. Catalytic mechanism and transition-state analysis of the arsenolysis reaction. *Biochemistry* **32**, 13212–13219.
27. Devereux, J., Haeberli, P.A. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.
28. Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
29. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Cryst. A* **50**, 157–163.
30. Brünger, A.T. (1988). X-PLOR Manual, Version 3.1. Yale University, New Haven, CT.
31. Furey, W. & Swaminathan, S., [Abstract PA33], American Crystallographic Association Meeting, New Orleans, July, 1990, vol. 18, p. 73.
32. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140–149.
33. Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763.
34. Kleywegt, G.J. & Jones, T.A. (1993). Masks made easy. In *The Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography 28*. (Wilson, K.S., ed.), pp. 58–59, Daresbury Laboratory, Warrington, UK.
35. Kleywegt, G.J. & Jones, T.A. (1994). From first map to final model. In *Proceedings of the CCP4 Study Weekend*. (Bailey, S., Hubbard, R. & Waller, D., eds), pp. 59–66, Daresbury Laboratory, Warrington, UK.
36. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
37. 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.
38. Fischer, M. & Short, S.A. (1982). Cloning of the *Escherichia coli* K-12 deoxyribonucleoside operon. *Gene* **17**, 291–298.
39. Sack, J.S. (1988). CHAIN – a crystallographic modeling program. *J. Mol. Graph.* **6**, 244–245.
40. Carson, M. (1987). Ribbon models of macromolecules. *J. Mol. Graph.* **5**, 103–106.