RESEARCH PAPERS

Acta Cryst. (1997). D**53**, 131–142

Refined Structure of Purine Nucleoside Phosphorylase at 2.75 Å Resolution

STHANAM V. L. NARAYANA,[a]* CHARLES E. BUGG[b] AND STEVEN E. EALICK[c]

[a]*Center for Macromolecular Crystallography and School of Optometry, University of Alabama at Birmingham, Birmingham, AL 35294–0005, USA,* [b]*BioCryst Pharmaceuticals, Inc., Birmingham, AL 35244, USA, and* [c]*Cornell University, Department of Biochemistry, Ithaca, NY 14853, USA*

(*Received 30 July 1996; accepted 8 October 1996*)

Abstract

Human purine nucleoside phosphorylase (PNP) catalyzes the reversible phosphorolysis of ribonucleosides and 2′-deoxyribonucleosides to the free base and (2′-deoxy)ribose-1-phosphate. The crystal structure previously determined at 3.2 Å resolution by multiple isomorphous replacement methods [Ealick, Rule, Carter, Greenhough, Babu, Cook, Habash, Helliwell, Stoeckler, Parks, Chen & Bugg (1990). *J. Biol. Chem.* **265**, 1812–1820] has now been refined at 2.75 Å. One important solvent molecule in the active site is found to be hydrogen bonded to Thr242 and Asn243, a second molecule to the Glu210 side chain (rotated out of the substrate-binding pocket), and a third bridges the hydroxyl of Tyr88 and SO_4(290), located in the phosphate-binding subsite. Hydrophobic interactions dominate the structure and many secondary structural elements are held together by hydrophobic clusters. In the low-resolution structure, the active-site residue Lys244 was modeled to be pointing into the active site, and the refined structure revealed that it is pointing away from the active site. Refinement improved the density for residues 244–249; however, loop 250–263 still shows significant disorder in the native structure. Comparison between crystal structures of native and an inhibitor (THDZ) complex reveals that this flexible loop 250–263 is stabilized by the hydrophobic interactions with the bound inhibitor. The refined structure of PNP is structurally homologous to carboxypeptidase A(CPA), an enzyme which cleaves C-terminus peptides in protein degradation. Similarities and differences between the structures of PNP and CPA are discussed.

1. Introduction

Purine nucleoside phosphorylase (PNP) catalyzes the reversible phosphorolysis of ribonucleosides and 2′-deoxyribonucleosides to the free base and ribose-1-phosphate and 2′-deoxyribose-1-phosphate, respectively. PNP catalyzes the phosphorolysis of many nucleoside analogs and is specific for guanosine, inosine and their analogs (Parks & Agarwal, 1972; Stoeckler, Agarwal, Agarwal & Parks, 1978). Although equilibrium favors the synthetic reaction, the enzyme functions biologically in the salvage pathway by catalyzing the cleavage reaction.

PNP has been isolated and purified from a variety of sources including both prokaryotic and eukaryotic organisms. The mammalian PNP is a trimer with a total molecular weight of 100 kDa (Ward, McAndrew & Wallis, 1979). Bacterial PNP is a hexamer with a molecular weight of about 150 kDa (Jensen & Nygaard, 1975). In addition to the differences in quaternary association and subunit size, the two enzymes differ in their substrate specificity. The bacterial enzyme works on adenosine and purine arabinoside whereas the mammalian PNP is more efficient in the synthesis of guanosine, inosine and their analogs (Krenitsky, Elion, Henderson & Hitchens, 1968).

Interest in PNP arises from its role in purine nucleoside metabolism and from its involvement in T-cell development. Absence of PNP activity has been associated with severe T-cell deficiency, while B-cell immunity remains functional. It has been proposed that PNP inhibitors might be used to treat T-cell leukemia, to suppress the host-*versus*-graft response, or to counter autoimmune disease without disturbing the patient's humoral immunity (Giblett, Ammann, Wara & Diamond, 1975; Stoop *et al.*, 1977; Rich, Arnold, Palella & Fox, 1979). The structure of human PNP has been determined to 3.2 Å resolution by multiple isomorphous replacement methods (Ealick *et al.*, 1990). Here were discuss the refinement and describe the detailed structure of PNP at 2.75 Å resolution.

2. Experimental

2.1. *Crystallization and data collection*

PNP crystals were grown from ammonium sulfate solution at room temperature (Cook, Ealick, Bugg, Stoeckler & Parl, 1981). The standard stabilizing

solution for these crystals was 50 m*M* citrate buffer (pH 5.4) containing 65% ammonium sulfate. Crystals of the enzyme are trigonal, space group *R*32, with hexagonal cell parameters $a = 142.9$ and $c = 165.1$ Å. The molecule is a trimer in which subunits are related by the crystallographic threefold axis. With one monomer per asymmetric unit, PNP crystals contain about 76% solvent in the unit cell (Matthews, 1968). A data set to 2.75 Å resolution was collected by oscillation photography (Arndt & Wonacott, 1977) at the synchrotron radiation source facility in Daresbury, England. Ten crystals ($0.4 \times 0.5 \times 0.3$ mm) were used to collect 93 film packs (of three films each) covering 50° of rotation. Most of the crystals were mounted such that the rhombohedral a^* axis was parallel to the oscillation axis. This mounting of crystals oriented the crystallographic threefold axis approximately 45° away from the oscillation axis and resulted in a complete data set with multiple recordings for equivalent reflections. The films were digitized with an Optronics P1000 rotating-drum microdensitometer and processed with the *MOSFLM* set of programs developed by Wonacott (1980) and modified by Greenhough & Helliwell (1982). Final integrated 62 377 observations were scaled and merged into 16 752 unique reflections. Out of the possible 17 038 unique reflections to 2.75 Å resolution, 98% of the data were recorded, including 89.5% of the possible data between 2.85 and 2.75 Å. The overall merging *R* factor on intensities to 2.75 Å resolution was 9.7%. A Wilson plot (Wilson, 1949) gave an overall isotropic temperature factor of 22 Å^2.

2.2. *Starting model*

A detailed account of the work, including enzyme purification, crystallization and various steps involved in building an atomic model of PNP at 3.2 Å resolution, has been described previously by Ealick *et al.* (1990). The original multiple isomorphous replacement (MIR) phases were obtained from the native and four heavy-atom derivative data sets. The electron-density map was easily interpretable and an atomic model for PNP was built using the molecular graphics program *FRODO* (Jones, 1985). Except for the region between residues 249 and 262, where the density was weak, a complete protein model of PNP was built. The *R* factor for the model fitted to the 20.0–3.2 Å resolution solvent-flattened MIR map was 38.5% and this was the starting point for the refinement reported in this paper.

2.3. *Refinement*

The structure was refined in 'stages' consisting of 15–20 cycles of restrained least-squares refinement, using the program *PROLSQ* (Hendrickson & Konnert, 1981). After each stage of *PROLSQ* refinement, the model was manually adjusted using the program *FRODO* (Jones, 1985). Explicit hydrogen bonds were used as restraints from secondary structural elements in the initial stages of the refinement. Data in the resolution range 5.0–3.2 Å were used initially; while extending resolution to 2.75 Å in small steps stereochemical restraints were relaxed gradually. Following each round of least-squares minimization, the model was checked against the maps calculated with combined phases. $2F_o - F_c$ maps were calculated using MIR phases for 20.0–5.0 Å resolution data. MIR and calculated phases were combined for 5.0–3.2 Å resolution data, and calculated phases were used for data beyond 3.2 Å resolution (Rice, 1981; Read, 1986). In the initial stages of refinement and before any solvent molecules were included, calculated structure factors were replaced with observed structure factors for the resolution range 20.0–5.0 Å in the calculation of phase-combined $(2|F_o| - |F_c|)$ maps. The r.m.s. difference between combined phases and MIR phases to 3.2 Å remained about 20° throughout the investigation. The $|F_o|$ maps calculated with combined phases were put through eight cycles of solvent flattening using Wang's *ISIR/ISAS* package of programs (Wang, 1985) and used in model building along with $(2|F_o| - |F_c|)$ maps. When the *R* factor fell below 25.0% (5.0–2.75 Å), omit maps (Bhat & Blow, 1982) were used to adjust the model. For the first nine stages of refinement the atomic coordinates and overall temperature factor were refined and resolution was extended to 2.75 Å in small increments. Starting from round 10, individual thermal parameters were refined. When the *R* factor fell below 22% (5.0–2.75 Å), solvent molecules were added to the model where peaks in $(F_o - F_c)$ maps were greater than three times the r.m.s. deviation of the map and occupied stereochemically reasonable positions. At this stage, a third sulfate ion was identified from the size and shape of the electron-density peak and from the environment. Occupancies and temperature factors for solvent molecules were refined in alternate cycles and solvent molecules with high *B* factors (>60 Å^2) and low occupancies (<0.3) were omitted from the model on several occasions while refinement was carried out. A bulk solvent model (Bolin, Filman, Matthews, Hamlin & Kraut, 1982), with mask parameters $k_2 = 0.335$ e Å^{-3} and $B = 110.0$ Å^2 for 70% of the cell, allowed 15.0–5.0 Å resolution data to be included. The solvent mask parameters were optimized (final values: $k_2 = 0.431$ e Å^{-3} and $B = 122.0$ Å^2 for a mask covering 74% of the cell) while refinement of positional and individual temperature factors was carried out. Details of the refinement are given in Table 1. The final model (288 residues, three sulfate ions and 56 water molecules) gave a residual $[R = \sum(|F_o| - |F_c|)/\sum|F_o|]$ of 21.5% for 15.0–2.75 Å resolution data and 20.3% residual for 5.0–2.75 Å resolution data (13 749 reflections, intensity greater than 1.5σ).

Table 1. *PNP refinement to 2.75 Å resolution*

Cell parameters (Å, °)	
a	142.9
b	142.9
c	165.1
α	90.0
β	90.0
γ	120.0
Data collection	
No. of crystals used	10
Resolution to which data collected (Å)	2.75
No. of reflections collected	62377
No. of unique reflections recorded	16752
No. of unique reflections possible	17517
R_{sym} (%)	9.7
% data recorded to 2.75 Å	95
Latest refinement	
Resolution (Å)	5.0 to 2.75
R (%)	20.3
No. of reflections ($>1.5\sigma$)	13749
Distance restraints	
Bond distance	0.022 (0.025)
Angle distance	0.035 (0.035)
Planar 1–4 distance	0.062 (0.055)
Restraints for peptide plane angle	3.1 (3.5)
Isotropic thermal factor restraints	
Main-chain bonded atoms	1.305 (1.25)
Main-chain non-bonded atoms	1.919 (2.15)
Side-chain bonded atoms	1.362 (1.50)
Side-chain non-bonded atoms	2.200 (2.40)
No. of solvent molecules located	56

3. Results

3.1. *Quality of the refined structure*

The plot of R value as a function of resolution (Luzzati, 1952) gave an estimated error in coordinates of 0.28 Å. The r.m.s. difference in coordinates of main-chain atoms between the starting and final models is 1.8 Å. Fig. 1 shows the coordinate shifts (Å) between the MIR and final refined models for main-chain (mc) and side-chain (sc) atoms. Some loops required considerable rebuilding during the refinement. The fit of the model to the electron density has been quantified by the calculation of a real-space R factor (Fig. 2) as proposed by Brändén & Jones (1990) and implemented in the program *O*. The overall real-space R factor for main-chain atoms is 16.5%, 21.2% for the side-chain atoms and 19.5% for all the atoms. Regions of the highest real-space R factor correlate with regions of high B-factor values (Fig. 3). The Ramachandran plot (Ramakrishnan & Ramachandran, 1965) for the refined coordinates is shown in Fig. 4. Almost all dihedral angles lie in the stereochemically allowed regions (Morris, MacArthur, Hutchinson & Thornton, 1992). The non-Gly residues found outside or near the edge of allowed regions are characterized by high B-factor values and poor density. All these residues are located in the flexible loop (Asp250–Gly 263), which borders the active site. Over the course of refinement and

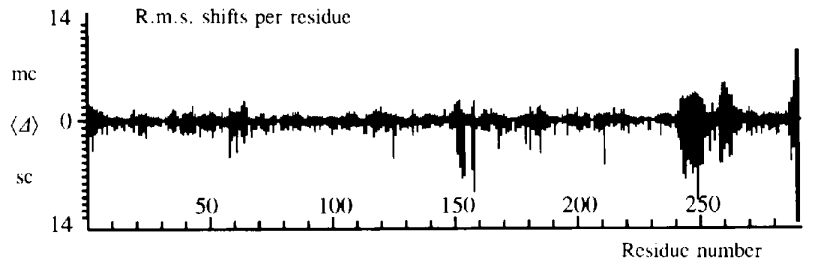

Fig. 1. R.m.s. coordinate shifts (Å) between the MIR model and the refined model; computed for main-chain atoms (mc) and for side-chain atoms including Cα (sc). Plot produced using *RIBBONS* (Carson & Bugg, 1986).

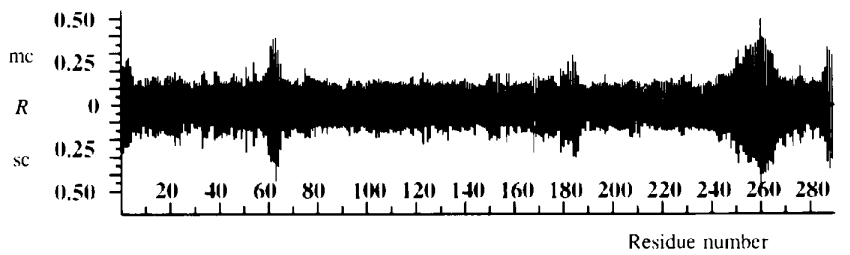

Fig. 2. Real-space electron-density fit per residue (R) computed using $(2F_o - F_c)\exp(i\alpha c)$ map. 20.0–2.75 Å data were used in the map calculation. The R factor as computed for main-chain atoms (mc) and for side-chain atoms including Cα (sc). Plot produced using *RIBBONS* (Carson & Bugg, 1986).

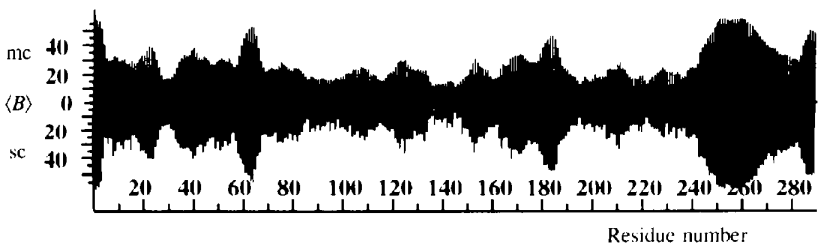

Fig. 3. Average temperature factors (Å^2) per residue, ⟨B⟩, computed for main-chain atoms (mc) and for side-chain atoms including Cα (sc). Plot produced using *RIBBONS* (Carson & Bugg, 1986).

rebuilding, the electron density for residues 244–249 has improved, while the density for residues 250–263 remained poor. The first three residues at the N terminus are poorly defined, as reflected in the *B*-factor plot. Packing of Tyr5 against a hydrophobic cluster including Phe98, Trp94 and Tyr10 stabilizes the N terminus beyond Gly4. Residues 59–65 are part of the loop bordering the active site opposite the purine binding loop. Residues on this loop have higher than average *B* factors, and thus may be flexible enough to allow the substrate to enter and products to exit the active site. Clear density was present up to Asp286 of C terminus and the remaining two residues are poorly defined in electron-density maps.

3.2. *Secondary structure and folding topology*

Although the general fold of PNP presented in this paper does not differ from the originally reported low-resolution structure (Ealick *et al.*, 1990), there are many changes of detail in the refined structure. Hydrogen bonds are assumed whenever the donor–acceptor distance was less than 3.5 Å and the precursor–donor–acceptor angle was between 110 and 180° (Baker & Hubbard, 1984). Fig. 5 illustrates the secondary structural elements of the final refined structure. The most prominent feature of the PNP molecule is the distorted β-barrel made up of an eight-stranded mixed β-sheet (sheet *A*) and a five-stranded mixed β-sheet (sheet *B*) joined at the edges, where four of the five strands of the smaller sheet are extensions of the larger sheet. The hydrogen bonds formed between the β-strands are represented schematically in Fig. 6. The central anchor of PNP, and eight-stranded mixed β-sheet (*A*), is flanked by helices *H*, *K* and *N* on one side and helices *J*, *L*, *M* and *O* on the substrate-binding side.

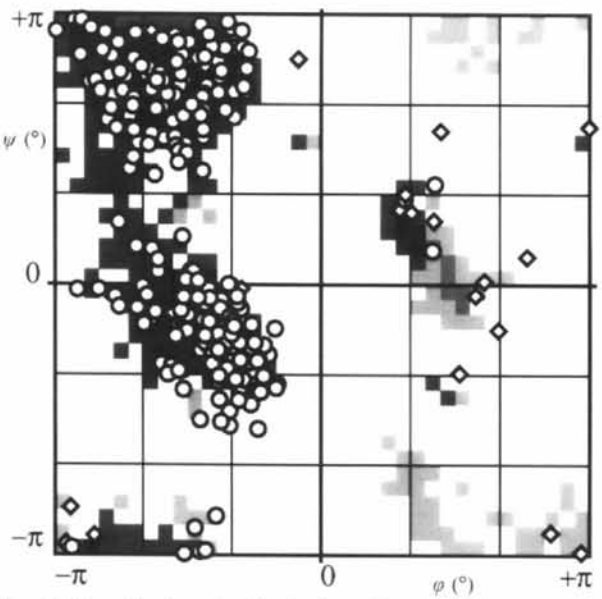

Fig. 4. Ramachandran plot for final coordinates. Circles represent non-glycine residues and diamonds mark glycine residues. The darkest gray levels represent commonly observed distribution ranges of non-glycine residues; the lightest gray regions where glycine residues are observed. The darker the shading, the more likely the observed conformation. The distributions are from an analysis of highly refined structures in the Protein Data Bank. The circles observed at the border and outside the allowed regions in the plot are from the flexible loops 58–63 and 249–265, which show higher temperature factors and weak density. Plot produced using *RIBBONS* (Carson & Bugg, 1986).

Helix *H* is a four-turn amphipathic helix, starting with a type-II β-turn. The N terminus of helix *H* has three tyrosine (Tyr5, Tyr7 and Tyr10) residues, each participating in a hydrophobic cluster involving Tyr5, Tyr7, Tyr10, Tyr50, Phe56, Phe85, Phe98 and Phe 155 (Fig. 7). Similarly, Trp16 and His20 present at the C terminus of helix *H* are packed against hydrophobic cluster Leu17, His23, Phe48, Phe70 and Leu106. The exposed side of this helix has hydrophilic residues Thr6, Glu8, Asp9, Lys11, Asn12, Glu15 and Ser19. Tyr5 OH is hydrogen bonded to the Asn55 carbonyl O atom, Tyr7 OH to His104 Nδ1 through $H_2O(300)$ and Tyr10 OH to $H_2O(317)$ while leaving their aromatic groups to participate in hydrophobic interactions.

Helix *K* (93–106), hydrophobic in nature, completely buried between helices *H* and *L*, is packed against sheet *A* with its helical axis parallel to the sheet plane. Its N terminus has two bulky groups, Tyr91 and Trp94, part of the hydrophobic cluster present at the N terminus of helix *H*. Pro92 ($\psi = -42°$ and $\varphi = -64°$) starts the helix *K* and Pro99 ($\psi = -64°$ and $\varphi = -57°$) breaks the flow of the helix where the carbonyl group of Trp94 points outwards and forms a hydrogen bond with $H_2O(317)$ (Fig. 8). Pro99 also participates in the hydrophobic cluster, along with the neighbouring Trp94, Phe98 and Phe103 and other hydrophobic residues Tyr5, Tyr10, Tyr50, Pyr56, Pro57, Phe85, Ph98 and Met81. Charged residues in helix *K*, Lys95 and Arg101, participate in hydrophobic interactions by packing their long aliphatic side chains parallel against aromatic rings. The Lys95 side chain is packed between Trp94, Tyr91 and Tyr5. Similarly, the Arg101 side chain is packed against Phe155 and Tyr10, allowing their terminal polar atoms to extend towards the surface of the PNP molecule. Thus, Pro99 creates a bend in helix *K*, which allows the residues at the two ends of this helix to participate in two hydrophobic clusters.

Helix *N* (225–231) is the third member of this bundle and packed against helix *K*. Sheet *B* wraps around the C terminus of helix *N* and Ser220, present at its N-terminus end, is hydrogen bonded to the active-site $SO_4(290)$. Pro223 ($\varphi = -71°$ and $\psi = -42°$), present at the N terminus, does not break the flow of the helix *N* and participates in the hydrophobic cluster involving Leu93, Pro140 and Ile226.

Helices J, L, M and O are packed on the substrate-binding side of the central sheet A. Helix J is a short 3_{10} helix. The average φ and ψ values are -67 and $-26°$, respectively, and most of the residues involved in this helix have either none or small side chains. Tyr166 present at the N terminus of helix L has its side chain extended across the hinge of sheet A and B and interacts hydrophobically with the buried helix N on the other side of the central sheet.

Helix M (203–214) lies in a plane perpendicular to the crystallographic threefold axis. The Nζ of the Lys211 side chain, present at the C terminus of helix M, is a donor of a hydrogen bond to the carbonyl O atom of a threefold-related Lys211. The aliphatic portion of the Lys211 side chain and its threefold-related lysines form the edges of a triangle around the threefold axis. Val208, Leu209 and Leu212 from helix M, along with Ile136, Leu138, Phe141, Val 193 and Val 195 from the strand Bi, form a hydrophobic cluster. This group of residues, along with the crystallographically related hydrophobic residues, create a strong hydrophobic environment around the threefold axis and thus seems to hold the biologically active trimer together. The PNP molecule ends with a long amphipathic C-terminal helix O. The N terminus of helix O has no hydrophobic residues whereas the C terminus has Phe274, Leu278,

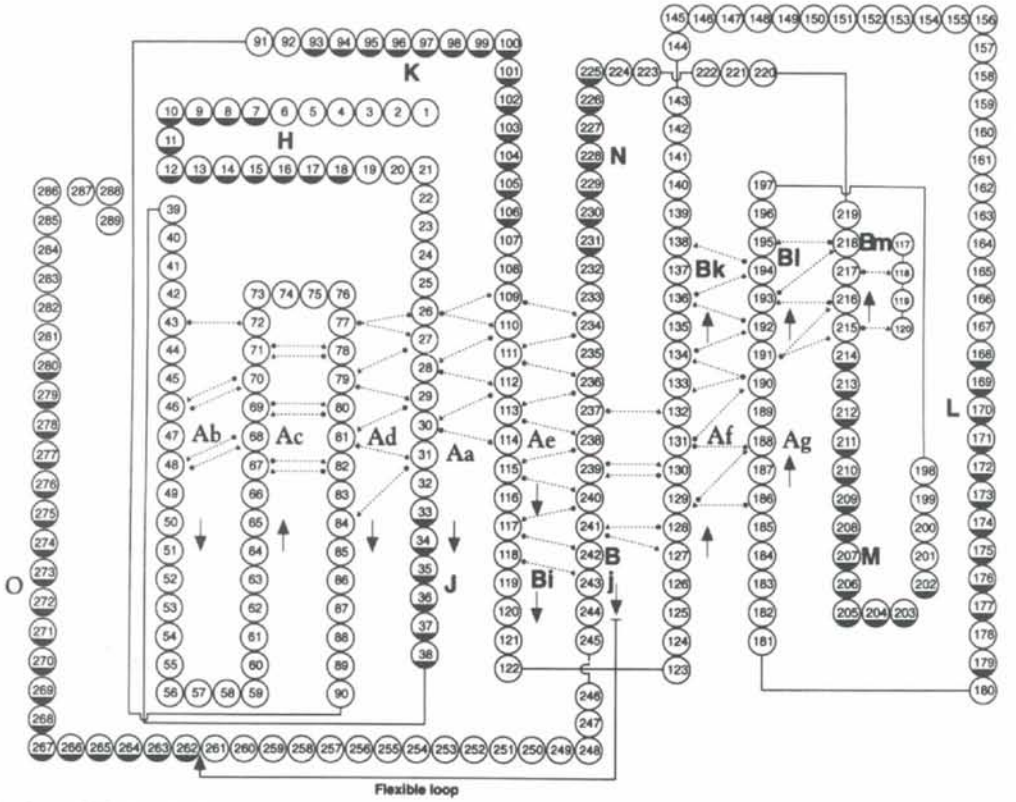

Fig. 5. Topology of PNP molecule. Hydrogen-bond network observed in the central β-sheets.

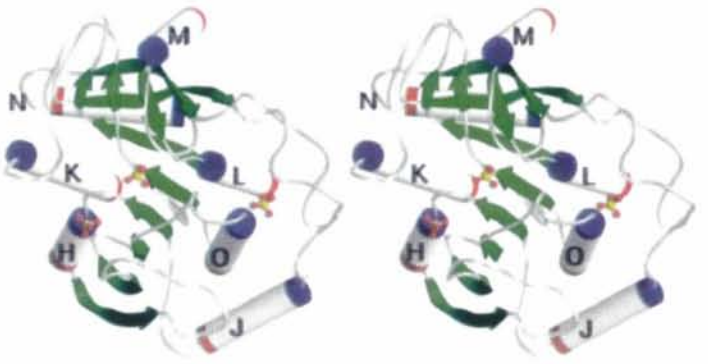

Fig. 6. Secondary structural elements of PNP molecules. Helices are represented as cylinders and β-sheet strands are represented as green arrows. Plot produced using *RIBBONS* (Carson & Bugg, 1986).

Val 275 and Leu271, involved in hydrophobic interactions with the N-terminal residues of helix *L* as well as Phe236 and Phe238 from the strand *Ah*.

The central eight-stranded mixed β-sheet (*A*) is enclosed on both sides by connections, the helices connecting the strands. A small five-stranded β-sheet (*B*) is packed against the larger sheet to form a distorted β-barrel. The larger β-sheet *A* shows the characteristic right-handed twist of about 120°. The central β-sheet (*A*) of PNP has a similar number and orientation of β-strands as other 'doubly wound' mixed sheet structures (Richardson, 1981) and as observed in carboxypeptidase A. Fig. 9 shows the topology diagrams for carboxypeptidase A (CPA) and PNP structures. However, the central β-sheet of CPA is not a true 'doubly wound' sheet as its N terminus starts with a helix and is linked to the first (rather than an adjacent) strand of the β-sheet. However, in PNP the N-terminus helix is connected to the central part of the sheet *A*. Nevertheless, both topologies look similar. Brändén (1980) suggested that α/β proteins are well suited topologically to form binding crevices at the carboxy end of the sheet where the strand order switches, exemplified by lactate dehydrogenase (Adams *et al.*, 1970). The phosphate-binding site of the PNP is observed at the C-terminus end of the sheet where the strand order changes. PNP does not have the classic mononucleotide-binding fold, but it does have the nucleotide-binding site. In CPA (Lipscomb, Reeke, Hartsuck, Quiocho & Bethge, 1970), the predicted binding crevice does not occur at the carboxy end of the sheet but a Zn atom does bind where the strand order switches. Thus, PNP is topologically similar to CPA but also carries a nucleotide-binding site.

Fig. 10 shows the backbone superposition of PNP on CPA. There are striking similarities in the secondary structure except in the surface loops. There is no significant sequence homology between CPA and PNP but the structural homology is noteworthy. 22% of PNP Cα atoms superimpose on CPA Cα atoms with a separation less than 1.5 Å and

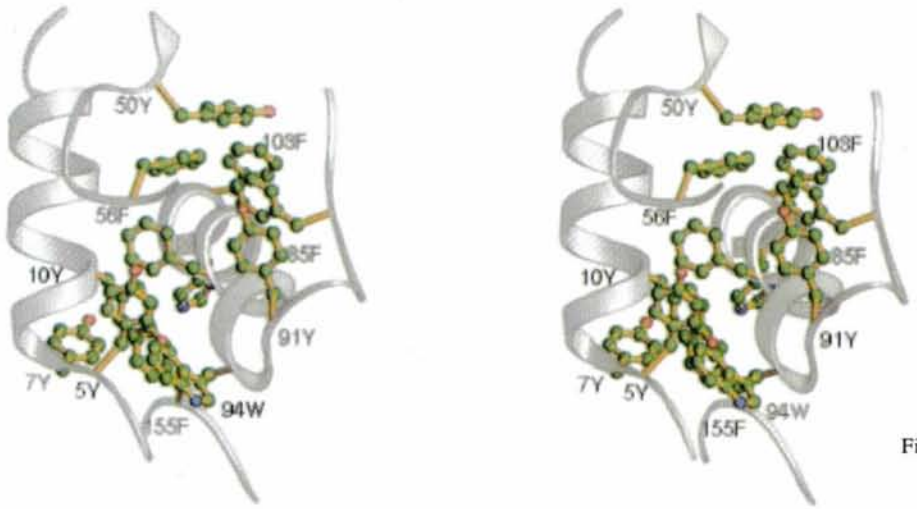

Fig. 7. A representative hydrophobic cluster holding the molecule together. Plot produced using *RIBBONS* (Carson & Bugg, 1986).

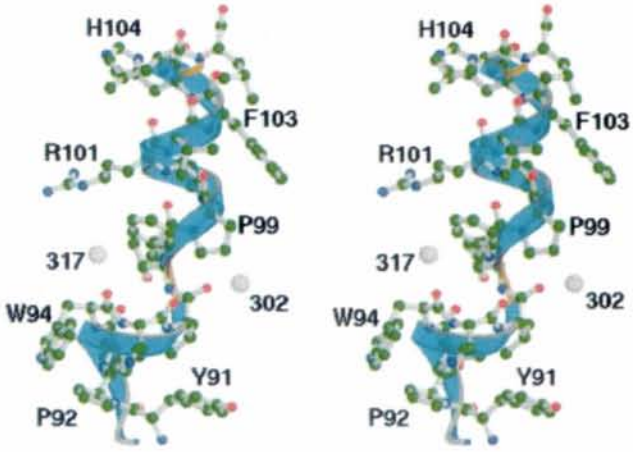

Fig. 8. Helix *K* with two prolines. Pro92 present in the beginning of helix and Pro99 in the middle of helix *K*. Two solvent molecules, 302 and 317, were located having hydrogen bonded to the CO groups of residues 94 and 96. Plot produced using *RIBBONS* (Carson & Bugg, 1986).

this number increases to 34% when the separation limit is extended to 2.0 Å. The substrate-binding site of CPA is a large crevice on the substrate-binding side of sheet *A* in PNP, but sheet *B* of PNP wraps and blocks the access to this space. In addition, a bulky hydrophobic residue Tyr192 occupies this area. With the exception of helices *H* and *M*, helices *J*, *K*, *L*, *N* and *O* superimpose on the corresponding helices of CPA. Helix *H* is the N-terminus helix where CPA and PNP are substantially different, and helix *M* covers the top of the sheet *B* in PNP and participates in intermolecular contacts of the PNP trimer.

3.3. *The active site*

The active site of PNP was characterized initially from the crystal structures of 8-iodoguanine and 5′-iodoformycin B (Ealick *et al.*, 1990) complexed with the enzyme. Three active sites are located near the subunit–subunit boundaries between the monomers of the trimer, and involve five polypeptide segments from one monomer and a short segment (Phe159) from the adjacent $(-y, x-y, z)$ related molecule. The side chains of Phe200, Glu201, Val217, Gly218, Met219, Thr242 and Asn243 appear to create the preferred recognition pocket for the purine base (Fig. 11). A water molecule [$H_2O(294)$] is 2.8 Å from the carbonyl O atom of residue Ser199, and is within hydrogen-bonding distance (2.9 Å) to the Glu201 side chain. Rotation around the Cα—Cβ and Cβ—Cγ bonds of Glu201 would position the carboxylate side chain in the correct orientation to form hydrogen bonds with N1 and N2 amino groups of the guanine base. Rotation around the Cα—Cβ bond of Phe200 from the observed position in the refined structure would maximize the hydrophobic interactions between the aromatic ring of Phe200 and the guanine base. Residues Val217 and Gly218 form the opposite side of this hydrophobic pocket along with Met219, which closes the bottom of the pocket. Residues Thr242 and Asn243 from the loop joining the C-terminus helix and strand *Bj* are in the correct position to form hydrogen bonds with N7 and O6 of the purine base.

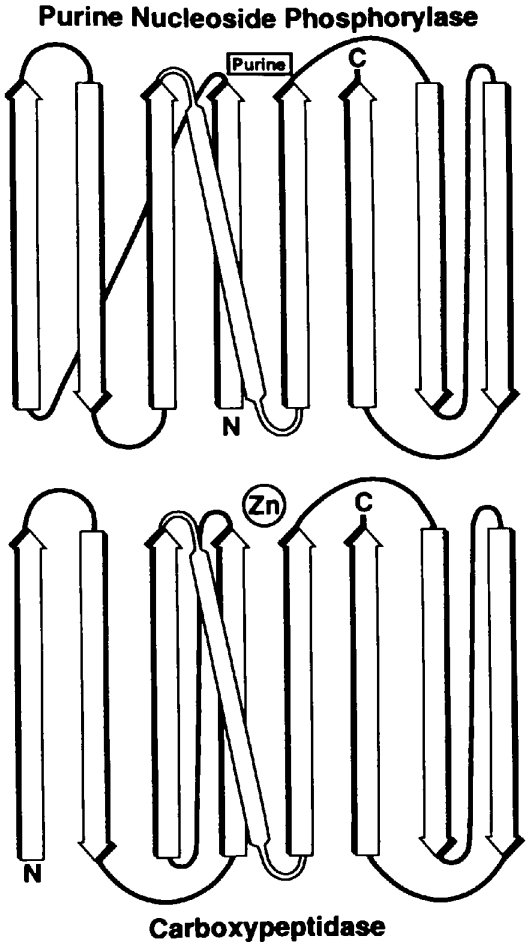

Fig. 9. Topology of the central β-sheet in PNP and CPA where the polarity of the sheets in two molecules is similar.

In the low-resolution MIR structure (Ealick *et al.*, 1990; Babu *et al.*, 1995) it was suggested that the orientation of the Lys244 side chain dictates the preference of PNP towards guanine analogs (Krenitsky, 1967). Guanine analogs have a hydrogen-bond donor at N1 and accept a hydrogen bond at O6, while adenine and its analogs have a hydrogen-bond donor at N6 and an acceptor at N1. Directing the Lys244 side chain into the active-site pocket, as done previously while building the low-resolution structure, would explain the specificity of the enzyme. However, while refining and rebuilding the structure, we realized that there was considerable torsional strain in the main chain and also the presence of unexplained density around Lys244 in difference Fourier maps. This extra density was not suitable for a water molecule and was pointing away from the substrate-binding pocket. Excluding the residues 242–250 from the model and after many cycles of refining and rebuilding, the calculated omit maps (Bhat & Blow, 1982) showed clear density identifying the position of residues 242–249. Rebuilding by pointing Asn243 inwards, Lys244 outwards, and positioning the Val245 side chain into the density originally occupied by the Lys244 side chain, has helped in positioning residues 242–249 much more satisfactorily. Fig. 12 shows density for omitted residues 242–245 in an $(F_o - F_c)\alpha_c$ difference map calculated using 15.0–2.8 Å resolution data. MIR and refined models are superposed on the density. The Thr242 OH group could function as a hydrogen-bond donor to the Asn243 side-chain carbonyl O atom and thus position its Nδ2 group to form a hydrogen bond with O6 of guanine and guanine analogs. This suggests that Thr242 and Asn243 are

more responsible for the PNP selectivity for guanine than Lys244.

3.4. *Sugar-binding site*

The space for the sugar-binding site, between guanine and the phosphate-binding site, is highly hydrophobic in nature (Fig. 11). The Ala116 carboxyl and the hydroxyl group of Tyr88 are pointed into this pocket. Phe159 (of the neighboring subunit), Phe200, His86 and Tyr88 lie on one side of the sugar-binding pocket, creating a hydrophobic patch which may serve to orient the ribose to facilitate nucleophilic attack by phosphate (Ealick *et al.*, 1990). The peptide N atom of Met219 and the His86 side-chain Nδ1 are in close proximity to form hydrogen bonds to hydroxyl O atoms of ribose.

3.5. *Phosphate-binding site/sulfate molecules*

Three sulfate ions have been identified in the PNP crystal structure. Two of these are located on the surface of the molecule and one, SO_4(290), is centrally located in the most accessible region of the active site. The O atoms are bound to the side chains of Arg84, Ser33, Ser220 and to Tyr88 through molecule H_2O(313). Table 2 presents the distances and angles to the neighboring atoms of SO_4(290) and Fig. 13 shows the geometry of its contacts. The Ser33 hydroxyl O

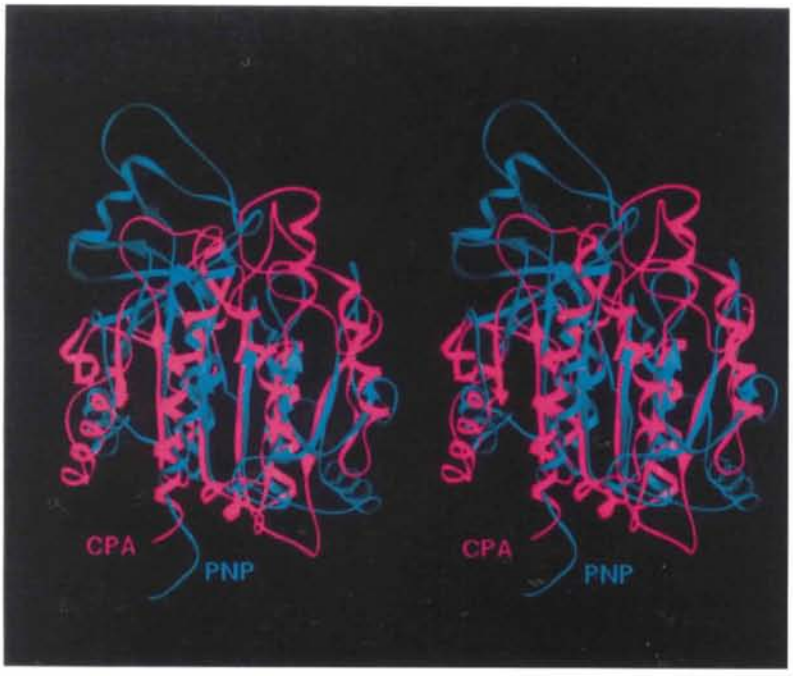

Fig. 10. The structures of PNP (red) and CPA (blue) are compared. The molecules are represented by connecting Cα atoms. Plot produced using *RIBBONS* (Carson & Bugg, 1986).

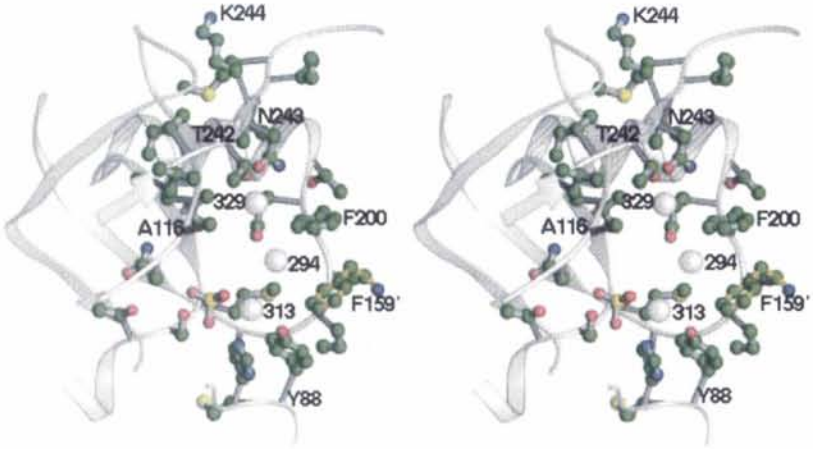

Fig. 11. The active site of PNP. Labeled residues are involved in the binding of purine base and sugar. Solvent molecule 329 is found to be hydrogen bonded to Thr242 and Asn243. Solvent molecule 294 is hydrogen bonded to Glu201 side chain which needs to be rotated around the Cα—Cβ bond to participate in the binding of the purine base. Lys244 is pointing away from the active site. Plot produced using *RIBBONS* (Carson & Bugg, 1986).

Table 2. *Interactions of SO_4(290) present in the catalytic phosphate-binding site with neighboring protein atoms*

SO_4(290) atom	Residue and atom name	Distance (Å)
O1	H_2O(313)	3.01
	Ser220 Oγ	3.01
O2	Ser33 N	2.70
	Ser33 Oγ	2.86
	H_2O(313)	2.48
O3	Ser220 Oγ	2.94
	Ala116 N	3.32
O4	Arg84 Nε	2.93
	His86 NHε2	2.90

atom and its peptide N atom are at proper distances and orientations to form hydrogen bonds with O2 of SO_4(290). The hydroxyl group of Ser220 is at hydrogen-bonding distance with O1 and O4 of SO_4(290). In a recent survey of refined protein structures, Chakrabarti (1993) suggested that, even though sulfate and phosphate are similar structurally, fully ionized sulfates need exclusively donor groups whereas phosphates (HPO_4^{2-} or $H_2PO_4^-$) require acceptor groups to some extent. Hydroxyl side chains of Ser33 and Ser220 and H_2O(313), can function as acceptors of hydrogen bonds for the active-site phosphate, whereas the same ligands can function as donors for the fully ionized sulfate.

A naturally occurring PNP mutant with no functional activity has been cloned and sequenced (Williams, Gekeler, McIvor & Martin, 1987). In this mutant, Glu89 has been replaced with a lysine. In the refined structure it is noticeable that the Glu89 carboxylate O atom interacts with His86 Nδ1, thus positioning its Nε2 to interact with O4 (2.9 Å) of molecule SO_4(290). The loop joining strand *Ad* to helix *K* consists of Arg84, Phe85, His86, Met87, Tyr88, Gly89, Gly90 and Tyr91, in addition to contributing critical residues to the phosphate-binding site, creates a suitable hydrophobic environment for the ribose-binding site. Replacing Gly89 with positively charged lysine would probably destabilize these favorable interactions leading to the observed loss of catalytic activity.

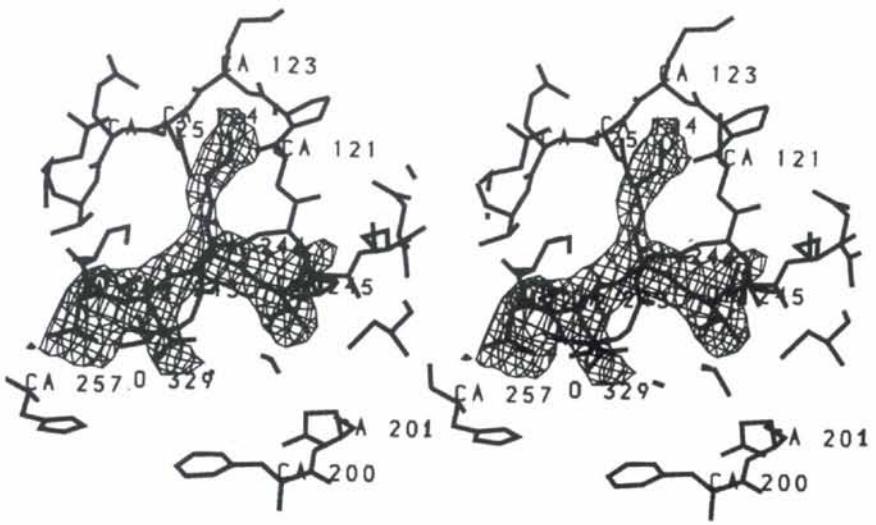

Fig. 12. Difference Fourier map calculated after many cycles of refinement in which residues 242–245 are not included.

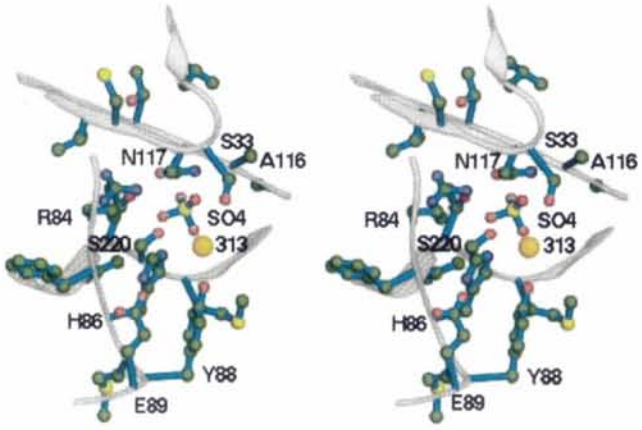

Fig. 13. The phosphate-binding site of PNP. Labeled residues are involved in the binding of SO_4(290). Glu89 side chain is hydrogen bonded to His86, which in turn participates in hydrogen bonding with SO_4(290) O atoms. Plot produced using *RIBBONS* (Carson & Bugg, 1986).

The second sulfate ion, $SO_4(291)$, has lower occupancy compared with the other two and is located on the surface of the molecule. It is held in place by the peptide N atoms of Gly34, Leu35 and Gly36 (Chakrabarti, 1993). The third sulfate ion, $SO_4(292)$, is also present on the surface of the molecule and is primarily held in place by the side chain of Arg284 from the neighboring threefold-related molecule. In addition, its location at the N-terminus end of helix *H* could be facilitated by its interactions with the helix dipole moment. Three symmetry-related $SO_4(292)$ ions form a triangle around the crystallographic threefold axis and interact with each other through $H_2O(298)$ and its symmetry-related water molecules.

3.6. *Flexible loop*

Loop 241–268 is among the most mobile regions of the enzyme, with an average *B* factor exceeding 45 Å^2. Residues 250–263 show highest *B* factors around 58 Å^2. Because of high thermal parameters in this loop, the atomic coordinates in this region are less reliable than elsewhere in the structure. Nevertheless, comparing the crystal structures of 2-thienylmethyl 9-deazaguanine complex (THDZ) of PNP with the native refined PNP structure, we can visualize the concerted nature of this loop movement (Fig. 14). The primary driving force for the conformational changes appears to be the binding of the inhibitor.* The loop containing the residues 250–263 does not participate in the crystal contacts and the root-mean-square (r.m.s.) displacement for all Cα atoms between the PNP and PNP–THDZ structures is 1.32 Å. The r.m.s. deviation for all the atoms of the 241–263 loop is 5.91 Å. The maximum movement, caused by substrate or inhibitor binding (Ealick *et al.*, 1990; Babu *et al.*, 1995), occurs at His257 which is displaced by several ångströms. Enhanced hydrophobic interactions, resulting from the participation of His257, Val260 and Leu261 in a hydrophobic cluster involving 2-thienylmethyl ring substitution on the guanine base and His86, Tyr88 and Phe199, seems to be the force behind this conformational change from a flexible loop to an apparent helical transformation for residues 257–261. Various hydrophobic substitutions at the N7 position of the guanine base have a similar effect, but

* Refinement of the THDZ complex crystal structure ($R = 19.2\%$ to 2.75 Å resolution) along with many more PNP inhibitor crystal structures wil be published separately. All the inhibitors are soaked into crystals and the complex crystals are isomorphous.

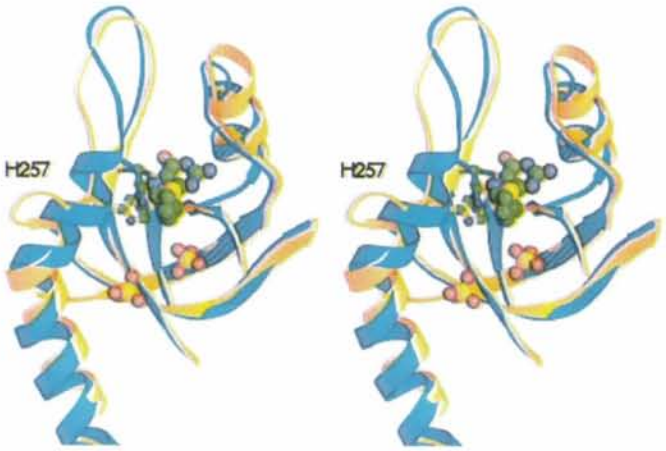

Fig. 14. Movement in the loop 250–263 when a substrate analog is bound in the active site. Plot produced using *RIBBONS* (Carson & Bugg, 1986).

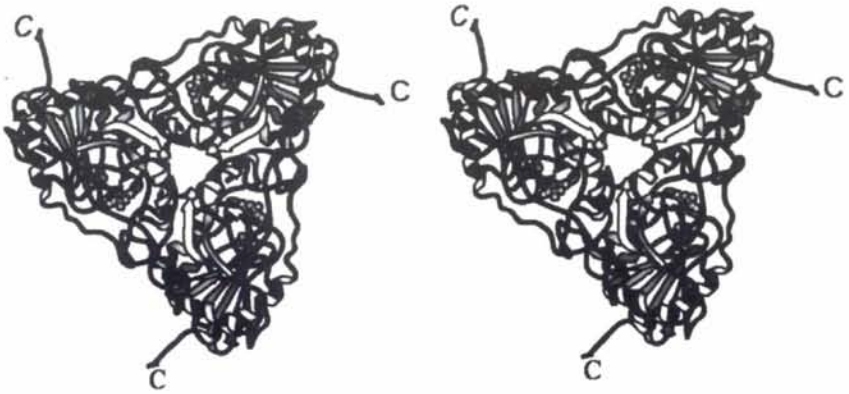

Fig. 15. Biologically active trimer of PNP. Phe159 from the neighboring PNP monomer completes the hydrophobic environment of the bound purine base.

with different magnitudes on the movement of this loop (Ealick *et al.*, 1990; Babu *et al.*, 1995).

3.7. *Solvent structure*

In total, 56 solvent molecules were identified in the difference electron-density maps computed during the refinement. Three of them occupy the empty active site (Fig. 11). Of these, $H_2O(294)$ hydrogen bonded to Glu201 has the highest occupancy of any solvent molecule present in the substrate-binding pocket. Solvent molecule $H_2O(329)$ occupies the position of the guanine base and satisfies the hydrogen-bond network with Thr242 and Asn243 residues. A third solvent molecule, $H_2O(313)$, forms hydrogen bonds with O2 of $SO_4(290)$, the hydroxyl group of Tyr88 and Nε2 of His86. Thus, by occupying the empty subsites of the substrate-binding pocket, the solvent molecules satisfy the spatial requirements of native PNP.

3.8. *Intermolecular packing*

The subunits of the trimer are related to each other by a threefold crystallographic axis. The substrate-binding pocket of a monomer is completed with the introduction of Phe159 from the neighboring monomer (Fig. 15). Phe159 of the neighboring molecule contributes to the hydrophobicity of the substrate-binding pocket along with Phe200, Tyr88 and His86. Adjacent trimers interact across the twofold axis in the crystal packing. Each trimer is rotated by the crystallographic twofold axis perpendicular to the threefold axis in such a way that the N-termini of adjacent molecules pack against each other. These trimers, situated on the threefold axis, are 70 Å in edge length and 35 Å in thickness. The C-terminal ends of the participating molecules protrude from the flattened corners of the trimers and interact with the C-terminus of twofold-rotated (1/6 translated along the threefold axis) neighboring molecules. The loop covering the entrance to the substrate-binding pocket is positioned on the edges of these triangular discs. Interactions between the flattened corners leave the edges exposed and result in solvent channels of 140 Å in diameter running through the crystal.

4. Conclusions

The PNP structure was determined using multiple-isomorphous replacement methods, and refined to 2.75 Å resolution by restrained least-squares methods. The refined structure of PNP is similar to the low-resolution structure reported previously, but reveals major new details about side chains and water structure involved in the substrate-binding site. PNP belongs to the family of α/β proteins, a family comprising a large group of proteins with widely different functions, specificity and substrate binding. This family includes proteases, kinases, dehydrogenases, rhodanases and phosphorylases (Levitt & Chothia, 1976; Cohen, Sternberg & Taylor, 1982). As observed in some of these proteins, the nucleotide-binding site of PNP is found near the center of the β-sheet on its carboxy edge. We also noted that carboxypeptidase A (Lipscomb *et al.*, 1970), a proteolytic enzyme, is topologically similar to PNP and has a Zn atom in this position.

The initial low-resolution MIR structure of PNP displayed weak density for residues 241–260 (Ealick *et al.*, 1990), but during the refinement to 2.75 Å resolution the density improved for residues 241–249, although it remained poor for the remaining residues to 260. X-ray analysis of some complexes (Ealick *et al.*, 1990) revealed that this part of the loop is stabilized and moves outward to accommodate the inhibitors. The refined structure of native PNP in conjunction with some complex structures of PNP, confirms the hypothesis that loop 241–260 is flexible in the native structure, and functions as a swinging gate to allow the substrate in and out of the active site.

References

Adams, M. J., Ford, G. C., Koekoek, R., Lentz, P. J. Jr, McPherson, A. Jr, Rossmann, M. G., Smiley, I. E., Schevitz, R. W. & Wancott, A. J. (1970). *Nature (London)*, **227**, 1098–1103.

Arndt, U. W. & Wonacott, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

Babu, Y. S. Ealick, S. E., Bugg, C. E., Erion, M. D., Guida, W. C., Montogomery, J. A. & Secrist, J. A. (1995). *Acta Cryst.* D**51**, 529–535.

Baker, E. N. & Hubbard, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–109.

Bhat, T. N. & Blow, D. M. (1982). *Acta Cryst.* A**38**, 21–25.

Bolin, T. J., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982). *J. Biol. Chem.* **257**, 13650–13662.

Brändén, C.-I. (1980). *Q. Rev. Biophys.* **13**, 317–338.

Brändén, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.

Carson, M. & Bugg, C. E. (1986). *J. Mol. Graph.* **4**, 121–122.

Chakrabarti, P. (1993). *J. Mol. Biol.* **234**, 463–482.

Cohen, J. E., Sternberg, M. J. E. & Taylor, W. R. (1982). *J. Mol. Biol.* **156**, 821–862.

Cook, W. J., Ealick, S. E., Bugg, C. E., Stoeckler, J. D. & Parl, R. E. Jr (1981). *J. Biol. Chem.* **256**, 4079–4082.

Ealick, S. E., Rule, S. A., Carter, D. C., Greenhough, T. J., Babu, Y. S., Cook, W. J., Habash, J., Helliwell, J. R., Stoeckler, J. D., Parks, R. E., Chen, S. & Bugg, C. E. (1990). *J. Biol. Chem.* **265**, 1812–1820.

Giblett, E. R., Ammann, A. J., Wara, D. W. & Diamond, L. K. (1975). *Lancet*, **1**, 1010–1013.

Greenhough, T. J. & Helliwell, J. R. (1982). *J. Appl. Cryst.* **15**, 493–508.

Hendrickson, W. A. & Konnert, J. (1981). *Biological Structure, Function, Conformation and Evolution*, Vol. 1, edited by R. Srinivasan, pp. 43–47. London: Pergamon Press.

Jensen, K. F. & Nygaard, P. (1975). *Eur. J. Biochem.* **51**, 253–265.

Jones, T. A. (1985). *Methods in Enzymology*, Vol. 115, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 157–171. London: Academic Press.

Krenitsky, T. A. (1967). *Mol. Pharmacol.* **3**, 526–536.

Krenitsky, T. A., Elion, G. B., Henderson, A. M. & Hitchens, G. H. (1968). *J. Biol. Chem.* **243**, 2876–2881.

Levitt, M. & Chothia, C. (1976). *Nature (London)*, **261**, 552–557.

Lipscomb, W. N., Reeke, G. N. Jr, Hartsuck, J. A., Quiocho, F. A. & Bethge, P. H. (1970). *Philos. Trans. R. Soc. London Ser. B*, **257**, 177–214.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.

Parks, R. E. Jr. & Agarwal, R. P. (1972). *The Enzymes*, Vol. 3, edited by P. D. Boyer, pp. 483–514. New York: Academic Press.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Read, J. R. (1986). *Acta Cryst.* A**42**, 140–149.

Rice, D. W. (1981). *Acta Cryst.* A**37**, 491–500.

Rich, K. C., Arnold, W. J., Palella, T. & Fox, I. H. (1979). *Am. J. Med.* **67**, 172–176.

Richardson, J. S. (1981). *Advances in Protein Chemistry*, Vol. 34, edited by C. B. Anfinsen, J. T. Edsall & F. M. Richards, pp. 288–307. London: Academic Press.

Stoeckler, J. D., Agarwal, R. P., Agarwal, K. C. & Parks, R. E. Jr (1978). *Methods in Enzymology*, Vol. 51, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 530–568. London: Academic Press.

Stoop, J. W., Zegers, B. J. M., Hendrick, G. F. M., Siegenbeek van Heukelom, L. H., Stall, C. E. J., de Bree, P. K., Wadman, S. K. & Allieux, R. E. (1977). *N. Engl. J. Med.* **296**, 651–655.

Wang, B. C. (1985). *Methods in Enzymology*, Vol. 115, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 90–112. London: Academic Press.

Ward, R. D., McAndrew, B. J. & Wallis, G. P. (1979). *Biochem. Genet.* **17**, 251–256.

Williams, S. R., Gekeler, V., McIvor, R. S. & Martin, D. W. Jr (1987). *J. Biol. Chem.* **262**, 2332–2338.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–320.

Wonacott, A. J. (1980). *MOSFLM. A Suite of Programs for On-Line Evaluation and Analysis of Integrated Intensities on Small-Angle Rotation/Oscillation Photographs.* Cambridge, England.