Structural analysis of two enzymes catalysing reverse metabolic reactions implies common ancestry

Olga Mayans^{1,2}, Andreas Ivens^{3,4},
L.Johan Nissen¹, Kasper Kirschner³ and
Matthias Wilmanns^{1,5}

¹EMBL Hamburg Outstation c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany and ³Department of Biophysical Chemistry, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

²Present address: Department of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

⁴Present address: Universität zu Köln, Institut für Biochemie, Otto-Fischer-Strasse 12–14, D-50674 Köln, Germany

⁵Corresponding author
e-mail: wilmanns@embl-hamburg.de

The crystal structure of the dimeric anthranilate phosphoribosyltransferase (AnPRT) reveals a new category of phosphoribosyltransferases, designated as class III. The active site of this enzyme is located within the flexible hinge region of its two-domain structure. The pyrophosphate moiety of phosphoribosylpyrophosphate is co-ordinated by a metal ion and is bound by two conserved loop regions within this hinge region. With the structure of AnPRT available, structural analysis of all enzymatic activities of the tryptophan biosynthesis pathway is complete, thereby connecting the evolution of its enzyme members to the general development of metabolic processes. Its structure reveals it to have the same fold, topology, active site location and type of association as class II nucleoside phosphorylases. At the level of sequences, this relationship is mirrored by 13 structurally invariant residues common to both enzyme families. Taken together, these data imply common ancestry of enzymes catalysing reverse biological processes—the ribosylation and deribosylation of metabolic pathway intermediates. These relationships establish new links for enzymes involved in nucleotide and amino acid metabolism.

Keywords: enzyme evolution/nucleoside phosphorylase/nucleotide salvage/phosphoribosyltransferase/tryptophan biosynthesis

Introduction

The transfer of a ribosyl group between an aromatic base and phosphate groups is one of the most fundamental biochemical reactions in the metabolism of nucleotides and amino acids (Craig and Eakin, 2000; Pugmire and Ealick, 2002). This transfer is generally reversible, leading to either the addition or the removal of a ribosyl unit from metabolic compounds. The addition of ribosyl groups is catalysed by phosphoribosyltransferases (PRTs), which

displace the 1'-pyrophosphate moiety from the substrate 5-phosphoribosyl-1'-pyrophosphate (PRPP), forming a 1'-glycosidic–nitrogen bond between a nitrogenated base and a phosphoribosyl group. The reverse process is catalysed by nucleoside phosphorylases (NPs) and is associated specifically with the removal of ribose from nucleosides by the phosphorolytic cleavage of the N-1'-glycosidic bond (Figure 1). PRT catalysis generally requires the presence of a divalent metal ion, whereas no metal ion is needed for NP catalysis. Both types of reactions are believed to follow a sequential S_N1 mechanism, although alternative mechanisms are not ruled out (Craig and Eakin, 2000; Pugmire and Ealick, 2002).

PRTs are known to be involved in nucleotide salvage and synthesis pathways as well as in the biosynthesis of the amino acids histidine and tryptophan, and the co-factor NAD. Most of the PRTs with known three-dimensional structure share the same two-domain architecture, known as the PRT-I fold (Craig and Eakin, 2000; Sinha and Smith, 2001), the only exception being quinolate PRT, which has been classified as the PRT-II fold (Eads *et al.*, 1997). The currently available NP structures are also categorized into two unrelated folds. The first class is involved in the cleavage of all purine and some pyrimidine nucleosides, whereas the second class (referred as NP-II throughout this text) catalyses the lysis of pyrimidine nucleosides only (Pugmire and Ealick, 2002). Despite the close mechanistic relationship between PRT and NP catalysis, there has been no direct evidence for the common evolutionary ancestry of these enzymes to date.

Anthranilate phosphoribosyltransferase (AnPRT) is involved in the biosynthesis pathway of the aromatic amino acid tryptophan. It catalyses the formation of a carbon–nitrogen bond between PRPP and anthranilate, thus establishing the precursor skeleton of tryptophan. Subsequently, the anthranilate group is converted into an indole side chain by two isomerization reactions and a cyclization reaction. Depending on the microorganism, AnPRT is either a monofunctional enzyme or is fused to other enzymes of the tryptophan biosynthesis pathway. Recently, it has been demonstrated that AnPRT mutants of *Mycobacterium tuberculosis* present attenuated virulence, allowing protection in the mouse model that is equal to vaccination with the virulent strain BCG (Smith *et al.*, 2001). Therefore, AnPRT has gained considerable interest as a target with which to develop superior strains of *M.tuberculosis* that could be used for vaccination.

Previous studies have noted that, in contrast to other PRTs, AnPRT shares similarity within three sequence segments with pyrimidine nucleoside phosphorylases (PyNPs) that comprise the NP-II fold (Mushegian and Koonin, 1994; Pugmire and Ealick, 2002). Lacking an experimental three-dimensional structure for AnPRT, it

Fig. 1. Reactions catalysed by (A) AnPRT and (B) TNP.

has remained uncertain whether members of the NP-II family and AnPRT share the same overall structure.

Here we present the crystal structure of AnPRT from the hyperthermophile *Sulfolobus solfataricus* (ssAnPRT), revealing that indeed AnPRT has an NP-II-like fold. With this structure, we are able to show that there could be an evolutionary link between members of the NP-II family and AnPRT. The structure of ssAnPRT, however, is unrelated to those of other PRTs, thus establishing a third PRT fold, designated here as PRT-III. The new ssAnPRT structure completes the structural inventory of all seven enzymes of the tryptophan biosynthesis pathway.

Results

Overall architecture of ssAnPRT

The crystal structure of the dimeric AnPRT from *S. solfataricus* (Ivens *et al.*, 2001) has been determined at 2.0 Å resolution using multiple anomalous dispersion (MAD) phases from a selenomethionine variant. This crystal form contains in its asymmetric unit four ssAnPRT protomers as two pairs of dimers. Each ssAnPRT protomer is composed of two domains (Figures 2, 3 and 4). The small α -helical domain comprises six helices ($\alpha 1$ – $\alpha 4$ and $\alpha 8$ – $\alpha 9$). The large α/β domain is formed by a central β -sheet (strands $\beta 1$ – $\beta 7$) and a C-terminal cluster of eight α -helices ($\alpha 5$ – $\alpha 7$ and $\alpha 10$ – $\alpha 15$), of which helix $\alpha 10$ is largely positioned towards the small α -helical domain (Figure 3A and B). All strands in the β -sheet are parallel except for $\beta 6$ (Figures 3A and 4). The two domains within this bilobal fold are not separated topologically. The inter-domain connections are formed by three loops $\alpha 4$ – $\beta 1$, $\beta 3$ – $\alpha 8$ and $\alpha 9$ – $\beta 4$ (Figure 4C).

The structure shows that ssAnPRT dimerizes via the small α -helical domain in a head-to-head fashion (Figure 4B). The dimer interface follows an approximate 2-fold symmetry and is formed predominantly by surface

patches of hydrophobic residues from helices $\alpha 1$ and $\alpha 3$ and the C-terminus of $\alpha 8$. Symmetrical contacts occur: (i) between helices $\alpha 3$ of both protomers; and (ii) between helix $\alpha 1$ from one protomer and helix $\alpha 8$ from the other. The only specific interaction observed at the dimer interface is one hydrogen bond between the main chain oxygen of L37 from the first molecule and the side chain hydroxyl group of S39 from the second. Due to this scarcity of directed interactions, dimer formation imposes few restraints on sequence conservation for the residues involved.

The 2-fold symmetry observed at the dimer interface is maintained by the two small α -helical domains, but does not extend throughout the entire dimer due to the presence of an intramolecular hinge region. Both ssAnPRT dimers comprise one protomer in the open and another in the closed conformation. Taking the structure of the larger α/β domain as reference, the small α -helical domain is rotated by $\sim 8^\circ$ in one conformation with respect to the other, as calculated with the program DynDom (Hayward *et al.*, 1997). The hinge opening is of the same magnitude in both dimers in this crystal form. This rotational difference translates only into small coordinate deviations within the hinge region but into amplified differences in the order of 4.5 Å at the peripheral regions of the two ssAnPRT domains (Figure 3C). Reflecting the observed flexibility in the domain–domain arrangement, there are only a small number of residue-specific interactions at the hinge interface between the two ssAnPRT domains (Figure 3B). One hydrogen bond is formed between a conserved residue pair, D76 (strand $\beta 1$) and N181 (helix $\alpha 9$). Both residues are located at different segments connecting the small α -helical domain and the large mixed α/β domain via the hinge region. However, the other, non-conserved inter-domain interactions are from K52 (helix $\alpha 4$) to E207 (helix $\alpha 10$) and from R60 (helix $\alpha 4$) to the main chain carbonyl group of E207 (helix $\alpha 10$). These contacts are

Fig. 2. Structure-based sequence alignment, using the sequences of the structures of AnPRT from *S. solfataricus* (ssAnPRT, PDB code: 1K8E), pyrimidine nucleoside phosphorylase from *B. stearothermophilus* (bsPyNP, PDB code: 1BRW) and thymidine phosphorylase from *E. coli* (ecTNP, PDB code: 1OTP). The numbers above the alignment correspond to the sequence of ssAnPRT. The positions of the α-helices and β-strands of the structure of ssAnPRT are shown by cylinders and arrows, respectively, and are labelled as in Figure 3A. The shared secondary structural elements of the small α-helical domain are in yellow, and those of the large mixed α/β domain are in red (α-helices) and cyan (β-strands). Strand β6, which is not topologically equivalent in the structures of ssAnPRT and bsPyNP/ecTNP but is structurally superimposable, is shown in magenta. Strand β7, which is unique in the ssAnPRT structure, is in grey. For comparison, the locations of the secondary structural elements of all corresponding structures are indicated by an underline. Residues that are conserved among each of the available sequence sets of AnPRT and NP-II families (data not shown) are highlighted in green if identical in at least 80% of the sequences. Conserved residues in structurally equivalent positions in the AnPRT and NP-II sequences are shown in bold, and are indexed below the alignment. The following ssAnPRT residue positions are labelled: *, specific inter-domain contacts between the small α-helical domain and the large mixed α/β domain; #, specific dimer contacts; A, predicted anthranilate-binding site; P, pyrophosphate-binding site; M, metal-binding site.

found in all four copies of ssAnPRT. Their different conformations indicate that there is inherent flexibility within the hinge region and that this property could be involved in the catalytic function of ssAnPRT or its regulation, possibly allowing even larger domain motions.

AnPRT has an NP-II like fold

There is no structural relationship between ssAnPRT and any other PRT of known structure. Thus, the structure of ssAnPRT constitutes a new PRT fold, designated here PRT-III. In contrast, the ssAnPRT structure reveals that its fold is related to that of the NP-II family (Figure 4), thus confirming a previous prediction from sequence data (Mushegian and Koonin, 1994). These data were based on the identification of amino acid similarities within three

sequence segments in the NP-II family and AnPRT, including an N-terminal 60 residue segment and the TGGxG phosphate-binding motif, whereas no significant sequence similarity was detectable for the large α/β domain. Comparison of the new ssAnPRT structure with the two currently available NP-II structures, thymidine phosphorylase from *Escherichia coli* (ecTNP) (Walter *et al.*, 1990; Pugmire *et al.*, 1998) and pyrimidine nucleoside phosphorylase from *Bacillus stearothermophilus* (bsPyNP) (Pugmire and Ealick, 1998), demonstrates that the fold similarity indeed extends to this domain. Although large, overall structural deviations originate from the different orientations of the small α-helical and large mixed α/β domains, the separated domains unambiguously show structural similarity. The small α-helical

Fig. 3. The structure of the protomeric unit of ssAnPRT. (A) Ribbon presentation coloured in a blue-to-red gradient. The secondary structural elements and the termini are labelled as in Figure 2. (B) Ribbon, using the colour code of Figure 2. The side chains of the residues that are involved in specific inter-domain interactions are shown in blue. Those residues involved in specific interactions that are conserved among the structures of ssAnPRT, PyNP and TNP are shown in green and numbered as in Figure 2. (C) Superposition of two monomeric units of ssAnPRT, where one is in the closed hinge conformation (magenta) and the other is in the open hinge conformation (green).

domains of ssAnPRT and bsPyNP superimpose with a root-mean-square deviation (r.m.s.d.) of 1.7 Å, using all matching main chain atoms of the coordinate sets in the program ALIGN_PDB (Cohen, 1997). The structures of the large mixed α/β domains deviate by ~ 2.0 Å. Despite the structural resemblance, a structure-based sequence alignment confirms the lack of overall sequence similarity within the large α/β domain (Figure 2).

The common fold translates into an identical topology of the shared secondary structural elements and their inter-domain connections, except for two inserts, which are specific to each one of the two families (Figures 2 and 4C). In ssAnPRT, the large mixed α/β domain contains a 40 residue insertion between strand $\beta 5$ and helix $\alpha 11$. It folds into two additional β -strands, $\beta 6$ and $\beta 7$, which extend the shared β -sheet $\beta 1$ – $\beta 5$. Conversely, the members of the NP-II family contain a C-terminal 100 residue extension beyond helix $\alpha 15$, which folds into an additional subdomain of unknown function (Figure 4C). Remarkably, the very C-terminus in the NP-II fold forms a β -strand, $\beta 6$, which extends the shared β -sheet $\beta 1$ – $\beta 5$ by one strand, instead of two ($\beta 6$ – $\beta 7$) as observed in the structure of the ssAnPRT. Strand $\beta 6$ is the only one in antiparallel orientation with respect to the shared β -sheet and it is structurally equivalent in NP-II and AnPRT.

A structure-based sequence alignment reveals that 13 residues are structurally identical between these PRT and NP-II members (Figure 2). These residues are confined to segments covering helices $\alpha 3$, $\alpha 4$ and $\alpha 8$ from the small α -helical domain and to two loop regions, $\beta 1$ – $\alpha 5$ and $\beta 2$ – $\alpha 6$, from the large α/β domain. Interestingly, none of them are involved in dimer formation. One of the structurally conserved residues is E54 (helix $\alpha 4$), which is involved in salt bridge formation in all structures available. Another residue (D76), located at the hinge

interface, is also conserved across ssAnPRT and the NP-II members of known structure. In the bsPyNP structure, this residue forms a salt bridge to K183 (numbering as for ssAnPRT, cf. Figure 2), a side chain, which is structurally equivalent to N181 in ssAnPRT.

Furthermore, the structures of members of the NP-II family and ssAnPRT share the same type of dimeric arrangement (Figure 4B), by head-to-head interactions between equivalent secondary structure elements of the small α -helical domain. In the dimeric structure of bsPyNP, asymmetric binding of a substrate analogue to only one of the two protomers is observed. The small α -helical subunit of the liganded protomer is rotated by $\sim 21^\circ$ when compared with the open conformation of the free protomer (Pugmire and Ealick, 1998). In contrast, the uncomplexed ecTNP structure presents a symmetric dimer (Pugmire *et al.*, 1998), although differences in the hinge conformation of a few degrees were observed among different crystal forms of this structure, confirming limited hinge flexibility.

The active site of ssAnPRT

The active site of ssAnPRT needs to accommodate both PRPP and anthranilate, in a structurally appropriate arrangement to catalyse the replacement of the 1'-pyrophosphate group of PRPP by anthranilate (Figure 1). In the reaction catalysed by NP-II, the nucleoside base is reversibly replaced by a phosphate group in the 1'-position. The leaving group, RP, is closely related to the PRPP substrate of AnPRT. The similarity between the aromatic bases, namely pyrimidines and anthranilate, is more tenuous. Only the PRT reaction requires divalent metal ions (Bauerle *et al.*, 1987; Yanofsky *et al.*, 1999). By mapping residues conserved among members of the AnPRT family onto the crystal structure of ssAnPRT, it

Fig. 4. Structural and topological similarity in the structures of ssAnPRT (left) and bsPyNP (right). The colour code is as in Figure 2. (A) Monomeric unit; (B) asymmetric dimer; (C) topology diagram (for ease of comparison, only major secondary structure elements are included). The α -helices and β -strands are shown as circles and triangles, respectively.

is possible to locate the active site to the hinge region between the small α -helical domain and the large mixed α/β domain. A deep groove opens to the protein surface, similar to the location of the active site in the available NP-II structures (Figures 5, 6 and 7).

PRPP/metal-binding site

To determine the precise binding site of PRPP, we have soaked crystals from ssAnPRT with an excess of PRPP, magnesium chloride and an anthranilate analogue. The diffraction limit of these complexed crystals was 2.7 Å resolution. A difference Fourier map calculated using X-ray data from complexed and unliganded crystals reveals the presence of a pyrophosphate group associated with a metal ion within the binding groove of the four

Fig. 5. Location of the active site in the AnPRT structure. (A) Ribbon diagram. The colour code is as in Figures 2 and 3B. Side chains of residues that may be involved in PRPP/metal or anthranilate binding are shown in blue and magenta, respectively. The pyrophosphate group of PRPP is shown in green, the bound metal-binding ion as a yellow sphere and the modelled solvent molecule as a red sphere. (B) Surface presentation. Ligands follow the same colour code as in (A). The residues involved in binding of the ligands are mapped in the respective colours on the interior protein surface.

molecule copies of this crystal form (Figures 6 and 7). The ribose-5-phosphate moiety of PRPP has, however, remained invisible, as well as the base analogue, suggesting that PRPP hydrolysis has occurred.

A magnesium ion is bound between the pyrophosphate moiety and the D223–E224 motif, which is invariant among all AnPRT sequences (Figure 2), and S91 from helix $\alpha 5$. In the unliganded structure at 2.0 Å resolution, no solvent molecules are located between the carboxylate groups of D223 and E224, providing further indirect evidence for specific metal binding. In addition, the two ssAnPRT copies with a closed hinge conformation present another electron density peak between the carboxylate groups of D223 and E224. In the absence of further experimental evidence, this density has been interpreted to originate from a solvent molecule, although a second magnesium-binding site is possible. This solvent molecule is not involved in specific contacts with the pyrophosphate group. In one of the ssAnPRT copies, in the open conformation, there is only weak electron density at this site (data not shown). Co-ordination of the pyrophosphate group via metal binding is in agreement with observations from other PRTs for which complexed structures are available (Craig and Eakin, 2000; Sinha and Smith, 2001). This suggests that in AnPRT catalysis as well, the metal ion plays a role in activation of the pyrophosphate. The DE binding motif is located in the loop connecting strands $\beta 5$ and $\beta 6$, which is part of the AnPRT-specific insert (Figures 2 and 4). Thus, the presence of this insert establishes a specific function in AnPRT, PRPP-associated metal binding, which is neither present nor required in NP catalysis.

In ssAnPRT, the pyrophosphate group is bound by the side chains of T92 (loop $\beta 1$ – $\alpha 5$) and K106 (strand $\beta 2$) (Figure 7A). S118 (loop $\beta 2$ – $\alpha 6$), albeit in close proximity to pyrophosphate, is not involved in direct binding. In the structure of bsPyNP, a lysine residue structurally equiva-

Fig. 6. $(2F_{\text{obs}} - F_{\text{calc}})\alpha_{\text{calc}}$ electron density map corresponding to the pyrophosphate/metal-binding site of the Mg^{2+} /pyrophosphate complex in the ssAnPRT protomer C at 2.7 Å resolution. The magnesium ion (yellow) is coordinated by D223, E224, S91 (not shown) and the two phosphate groups of pyrophosphate. Phosphate groups are bound further by the side chains of T92 and K106. In the molecule copies in the closed hinge conformation (protomers C and D), additional electron density bridging residues D223 and E224 have been modelled as a solvent molecule (red). The electron density associated with the residues involved in binding of the metal/pyrophosphate group is contoured at 1.0σ in dark violet. The density associated with the metal/pyrophosphate group and the solvent molecule is contoured at 1.5σ in dark green.

lent to K106 also provides a salt bridge to the phosphate substrate. In addition, a threonine, equivalent to S118 of ssAnPRT, provides a hydrogen bond to the phosphate (Figures 2 and 7). This comparison indicates that the pyrophosphate- and phosphate-binding sites in ssAnPRT and in the NP-II family, respectively, are shared, in particular, with respect to two loops involving their flanking secondary structural elements $\beta 1$ – $\alpha 5$ and $\beta 2$ – $\alpha 6$ from the large α/β domain (Figures 2, 4 and 7). The first loop contains a glycine-rich sequence motif DxxxTGG that was identified previously as the phosphate-binding motif in other unrelated enzymes (Kinoshita *et al.*, 1999).

Predicted anthranilate-binding site

To date, no experimental data on the binding of anthranilate by AnPRT are available. However, in the structures of ecTNP and bsPyNP, the known binding sites of the respective thymidine or uridine substrate analogue base moieties (Walter *et al.*, 1990; Pugmire and Ealick, 1998) allow conclusions on anthranilate binding by AnPRT to be drawn by analogy (Figure 7).

In members of the NP-II family, one of the key residues in base recognition is an arginine on helix $\alpha 8$ that interacts with the lateral O-4 group of the pyrimidine (Figure 7B). The equivalent residue position in the AnPRT sequences, R164 (Figure 2), is invariant, suggesting a similar role in binding of anthranilate by possibly forming a salt bridge with its *o*-carboxylate group. This prediction permits the approximate localization of the anthranilate-binding pocket in ssAnPRT. Other conserved residues involved in pyrimidine binding in NP-II are, however, different in the AnPRT sequences. On the other hand, there are a number of residues in this pocket, such as H107, H152 and N174, which are invariant among the AnPRT sequences (Figure 2), implying a role in recognition of anthranilate (Figure 7).

In the current crystal form of ssAnPRT, the binding sites for PRPP and anthranilate are separated by 14 Å (approximate distance between the guanidinium group of R164 and the proximal phosphate group of pyrophosphate) and are, most likely, too distant for catalysis. We, therefore, speculate that ssAnPRT may undergo further domain movements via the flexible hinge region during productive substrate binding, exceeding the conformational flexibility observed in the current structures.

Discussion

Common evolutionary ancestry of AnPRT and the NP-II family

Comparison of the structures of ssAnPRT and the NP-II family demonstrates that they are related, confirming previous predictions (Mushegian and Koonin, 1994). Their relationship is manifested by the following: (i) the overall structural similarity of the two domains; (ii) their identical topologies, except strands $\beta 6$ and $\beta 7$; (iii) the identical location of 13 invariant residues across NP-II and AnPRT sequences, providing a number of inter- and intra-domain interactions common to both families; (iv) the observation of structurally similar and topologically related binding sites for the phosphate substrate of NP-II and the pyrophosphate moiety within the PRPP substrate of AnPRTs; and (v) the prediction of the same binding site for the base substrate, in which the invariant R164 (ssAnPRT sequence numbering) is expected to play a key role in anthranilate binding, indicating an equivalent function to the structurally identical arginine in the NP-II structures. Taken together, the multilevel similarities in topology, structure and sequence unambiguously indicate that AnPRT and members of the NP-II family are related, and thus imply a common ancestry. With these data, AnPRT constitutes the first evolutionary link between

Fig. 7. Substrate-binding sites in ssAnPRT (left) and bsPyNP (right). **(A)** Pyrophosphate/metal- and phosphate-binding sites, respectively (Pugmire and Ealick, 1998). **(B)** Predicted base-binding site in ssAnPRT and as observed in the bsPyNP–pseudouridine complex (Pugmire and Ealick, 1998). The binding site for anthranilate is predicted to be formed by R164 and by three further residues (H107, H152 and N174) that are conserved within the AnPRT sequences (Figure 2). The invariant arginine in the PyNP–pseudouridine complex aligns with R164 of the ssAnPRT sequence (Figure 2). Numbering is as in Figure 2.

reverse, but distinct, metabolic ribosylation and deribosylation processes.

In contrast, there are no detectable overall relationships between ssAnPRT and other PRTs. Members of the PRT-I family are involved in nucleotide biosynthesis or salvage pathways and consist of a major mixed α/β domain and a flanking hood domain, which is poorly conserved in sequence and structure (Craig and Eakin, 2000; Sinha and Smith, 2001). These PRTs contain a conserved 13 residue metal/PRPP-binding motif. Remarkably, this motif contains a loop with the sequence signature DxxxTGG (Phillips *et al.*, 1999), which is identical to that of the first PRPP-binding loop in ssAnPRT as well as to a related glycine-rich motif in the NP-II family. Variations of this motif are also found in other phosphate group-binding enzymes (Kinoshita *et al.*, 1999). In members of the PRT-I family and AnPRT sequences, even the less conserved residue positions 'xxx' within this loop are shared. In both sequence families, the first two residue positions are generally hydrophobic, whereas the third position is largely conserved as an alanine in PRT-I and as a glycine in the AnPRT sequences (Figure 2).

Another PRT of known structure is quinolinic acid PRT, which is involved in *de novo* NAD biosynthesis and

quinolinic acid metabolism. This enzyme comprises a two-domain structure, which was designated as PRT-II, and includes an N-terminal mixed α/β domain and an incomplete β/α -barrel (Eads *et al.*, 1997). Its active site is located at the C-terminal face of the β/α -barrel, which contains a phosphate group-binding motif that is identical to those found in many other β/α -barrel enzymes (Wilman *et al.*, 1991; Höcker *et al.*, 2001). Apart from the shared enzymatic activity, there is no detectable relationship of quinolinic acid PRT to other PRTs.

In summary, the present data indicate that the catalytic activities of enzymes involved in the ribosylation or deribosylation of metabolic compounds have evolved independently of each other and even at different times in the course of evolution. In contrast to this, there are established evolutionary relationships for enzymes involved in the catalysis of closely related substrates, namely among the members of the purine PRT-I, NP-I and NP-II families. However, unlike these previous examples, the structure of ssAnPRT herein provides evidence to suggest that two enzyme families, ssAnPRT and NP-II, that catalyse reverse metabolic processes in unrelated pathways, share a common ancestry. Thus, ssAnPRT and NP-II could serve as a paradigm to study

Table I. X-ray data statistics

Space group	<i>P</i> 2				
Unit cell	<i>a</i> = 91.7 Å, <i>b</i> = 65.8 Å, <i>c</i> = 116.8 Å, β = 107.5°				
	Native data set	MAD data on Se–Met derivative			PRPP/Mg complex
Beamline	BW7B	BW7A			X11
Wavelength/energy	0.834 Å	λ1 (peak) 12 655 eV	λ2 (infl.) 12 651 eV	λ3 (remote) 13 775 eV	0.85
Overall resolution (Å)	18.0–2.05	18.0–2.85	18.0–2.85	18.0–2.85	20.0–2.70
Unique reflections	77 715	30 897	30 733	30 787	36 578
<i>R</i> _{sym} (<i>I</i>)	0.053	0.063	0.059	0.065	0.043
Multiplicity	3.6	7.4	3.5	3.9	4.6
Completeness (%)	92.7	99.9	99.8	99.8	99.8
< <i>I</i> /σ(<i>I</i>)>	10.8	11.1	11.8	11.4	30.1
Outer resolution shell (Å)	2.09–2.05	2.95–2.85	2.95–2.85	2.95–2.85	2.75–2.70
Unique reflections	3382	2041	2025	2042	1795
<i>R</i> _{sym} (<i>I</i>)	0.242	0.270	0.286	0.281	0.288
Multiplicity	1.9	7.2	3.5	3.8	4.7
Completeness (%)	81.2	100.0	99.9	100.0	100.0
< <i>I</i> /σ(<i>I</i>)>	1.8	2.1	2.1	2.1	4.5
SHARP phasing statistics (28 Se–Met sites/asymmetric unit)					
Phasing power					
Ano		1.7	0.9		1.3
Iso		2.1	2.2		Reference
SHARP figure of merit		0.42 (acentric reflections only)			

further the origins of enzyme evolution, which could have developed from a common biochemistry of the catalysed reactions, substrate similarity or retrograde pathway evolution (Gerlt and Babbitt, 2001; Henn-Sax *et al.*, 2001). Such an analysis, however, will require the complete biochemical and structural determination of ssAnPRT in the presence of substrates or their analogues.

Structural basis for PRT-function of ssAnPRT with NP-II-fold

Our data from sequence (Figure 2) and structure comparisons (Figure 4) provide indications as to why ssAnPRT and members of the NP-II family, sharing the same overall fold, are associated with distinct catalytic functions. The presence of a metal-binding site in ssAnPRT but not in the NP-II structures reveals the most obvious difference. Taking into account available data, metal-assisted catalysis is a strict requirement in all PRT reactions (Bauerle *et al.*, 1987; Yanofsky *et al.*, 1999) but not in NP catalysis. Despite some distant relationships in phosphate binding in the structures of ssAnPRT and members of the NP-II family, sharing the same location within the overall fold, structural comparison indicates distinct geometrical differences in the way in which inorganic phosphate versus PRPP are bound as substrates within the corresponding sites. For a quantitative comparison, however, a complex structure will be required revealing the full structural detail of PRPP binding in an AnPRT. In contrast, there are no indications that binding of the substrate base leads to functional discrimination. The observed distant structural

relationships for the base-binding pockets in ssAnPRT and NP-II seem to reflect distinct binding requirements for specific residual groups of each base substrate.

Relationships of tryptophan biosynthesis and other metabolic pathways

To date, the complete structural characterization of the enzymes of a major metabolic pathway is only available for glycolysis, encompassing 10 catalytic activities. Tryptophan biosynthesis is one of the most thoroughly studied metabolic pathways (Yanofsky *et al.*, 1999). With the new structure of ssAnPRT, tryptophan biosynthesis is the first amino acid biosynthesis pathway for which the atomic structures of all seven enzymes have been determined.

Remarkably, three of the seven reactions in this pathway are catalysed by enzymes with the (β/α)-barrel fold, implying evolutionary relationships of enzymes within this pathway (Wilmanns *et al.*, 1991). The structures of the enzymes of the tryptophan biosynthesis pathway may also serve as tools to reveal possible evolutionary relationships to enzymes from other metabolic pathways. Indeed, the structures of two (β/α)₈-barrel enzymes of histidine biosynthesis, HisA and HisF, compare well with the (β/α)₈-barrel enzymes of tryptophan biosynthesis, TrpF and TrpC (Lang *et al.*, 2000). Another relationship between the two amino acid biosynthesis pathways has been established recently by the new structure of the imidazole glycerol phosphate synthase bienzyme complex (Chaudhuri *et al.*, 2001; Douangamath

Table II. Refinement statistics

	Apo-ssAnPRT	ssAnPRT-PRPP-Mg ²⁺
No. of reflections	77 715 ^a	36 578 ^a
Resolution	18.0–2.05 Å	20.0–2.7 Å
R-factor (%)	21.3	23.3
R _{free} (%)	26.0	29.7
Content of asymmetric unit		
No. of protein molecules	4	4
No. of protein residues/atoms	1365/10 499	1364/10 494
No. of solvent atoms	940	172
R.m.s.d. bond length (Å)	0.007	0.006
R.m.s.d. bond angles (°)	1.4	1.3
Ramachandran plot		
No. of residues in core regions	1117	1095
No. of residues in disallowed regions	1	2
Average B-factors (Å ²)		
Main chain atoms	34.2	60.7
All protein atoms	35.5	61.5
Solvent atoms	45.8	62.4

^aAll observed reflections were used in the refinement.

et al., 2002) and its comparison with the available anthranilate synthase structures (Knöchel *et al.*, 1999; Morollo and Eck, 2001; Spraggon *et al.*, 2001). These established relationships in histidine and tryptophan biosynthesis are remarkable since these pathways start with unrelated reactions, chorismate/glutamate and ATP/PRPP, respectively. In light of these established relationships, it remains to be seen whether the PRT activities in these two pathways are also related. Like AnPRT, ATP PRT from the histidine pathway does not adhere to the sequence pattern of the canonical PRTs-I (Lohkamp *et al.*, 2000). Since AnPRT and ATP PRT do not share significant sequence similarity (data not shown), the experimental structure of the latter needs to be determined to detect possible relationships between these two enzymes.

Materials and methods

AnPRT from *S.solfataricus* was overexpressed, purified and crystallized as previously reported (Ivens *et al.*, 2001). X-ray data were collected on the wiggler beamline BW7B (EMBL/DESY, Hamburg, Germany) at a wavelength of 0.834 Å under cryo-conditions. The crystals were frozen in the mother liquor supplemented with 22% (v/v) glycerol as cryo-protectant. Reflections up to a resolution of 2.05 Å were recorded and processed on a 345 mm Mar image plate scanner in two sweeps, high and low resolution, at $\Delta\Phi = 1.0^\circ$ rotation angle. The crystals belong to the space group *P2* with unit cell dimensions of $a = 92.3$ Å, $b = 65.8$ Å, $c = 116.4$ Å and $\beta = 107.5^\circ$. The HKL suite of programs (Otwinowski and Minor, 1997) was used for data processing and reduction. Data statistics are included in Table I. Four ssAnPRT monomers were present in the asymmetric unit, resulting in a solvent content of 45%. A native Patterson map revealed two peaks corresponding to the translation vectors ($\Delta x = 0.01$, $\Delta y = 0.5$, $\Delta z = 0.48$) and ($\Delta x = 0.02$, $\Delta y = 0.5$, $\Delta z = 0.51$) of 24.7σ and 9.5σ height, respectively. These peaks result from screw, non-crystallographic symmetry (NCS) axes parallel to the crystallographic 2-fold axis leading to A2 pseudosymmetry.

Experimental phases were obtained from three-wavelengths anomalous dispersion measurements using crystals from a selenium–methionine derivative of ssAnPRT (Ivens *et al.*, 2001). X-ray data were collected at the wiggler beamline BW7A (EMBL/DESY, Hamburg, Germany) from a crystal grown and cryo-cooled as those from native ssAnPRT. MAD data statistics are listed in Table I. The positions of the anomalous scatterers (28 selenium sites in the asymmetric unit, seven sites per ssAnPRT

monomer) were determined and refined using the software RSPS (Knight, 2000) and SHARP (LaFortelle and Bricogne, 1997). Initial phases were improved by density modification, including NCS domain averaging in DM (Cowtan and Main, 1998). Mask creation and modification was performed with the MAMA software (Kleywegt and Jones, 1999). Calculation of the NCS matrices was aided by locating the NCS axes in real space using GETAX (Vornrhein and Schulz, 1999). The model was built with the graphics software O (Jones *et al.*, 1991) and the intermediate models were refined with the CNS package v1.0 (Brünger *et al.*, 1998). For refinement, the data were divided into a working and a test set (1.5% of the total number of reflections) using the program FREERFLAG (Brünger, 1992). NCS restraints were applied throughout refinement on both geometry and temperature factors and were released at final stages. Refinement was carried out by overall anisotropic B-factor scaling, bulk solvent correction and restraint atomic B-factor minimization as implemented within the CNS package. The geometrical properties of intermediate models were assessed using PROCHECK (Laskowski *et al.*, 1993). Water molecules were built automatically using ARP/wARP (Perrakis *et al.*, 1999) and employing the ‘water pick’ routine in CNS. Waters with elevated temperature factors were validated by visual inspection in $2F_o - F_c$ and $F_o - F_c$ electron density maps using O. The protein model includes all residues of the four polypeptide chains, except for residues 80–83 and 344–345 in copy A, 80–82 and 343–345 in copy B, and 80–83 in copy C. All protein residues are present in copy D. Geometrical parameters for the final model and refinement statistics are listed in Table II. The structure coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 1K8E and 1GXB.

For structural investigation of the metal/PRPP-binding site, crystals from ssAnPRT were soaked in 10 mM PRPP, 10 mM MgCl₂ and 5 mM anilin-2-arsenic acid for 15 min. X-ray diffraction data were collected at the bending magnet, X11 beamline (EMBL/DESY) under the same cryo-conditions as those from native crystals. Manual modifications of the model were carried out using the graphics software O. Crystallographic refinement was performed with the CNS software and followed the same protocol as for the native structure, although, in this case, strict NCS was applied throughout refinement for non-divergent parts between copies.

Acknowledgements

Galina Obmolova was involved in the early stages of this project. We thank Reinhard Sterner and Gavin Murphy for critical comments on this manuscript. This work was supported by a Marie-Curie postdoctoral fellowship by the European Union to O.M. and by grant No. 31-4585595 of the Swiss National Science Foundation to K.K.

References

- Bauerle, R., Hess, J. and French, S. (1987) Anthranilate synthase–anthranilate phosphoribosyltransferase complex and subunits of *Salmonella typhimurium*. *Methods Enzymol.*, **142**, 366–386.
- Brünger, A.T. (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
- Brünger, A.T. *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Chaudhuri, B.N., Lange, S.C., Myers, R.S., Chittur, S.V., Davisson, V.J. and Smith, J.L. (2001) Crystal structure of imidazole glycerol phosphate synthase. A tunnel through a (β/α)(8) barrel joins two active sites. *Structure*, **9**, 987–997.
- Cohen, G.E. (1997) A program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.*, **30**, 1160–1161.
- Cowtan, K. and Main, P. (1998) Miscellaneous algorithms for density modification. *Acta Crystallogr. D*, **54**, 487–493.
- Craig, S.P., III and Eakin, A.E. (2000) Purine phosphoribosyltransferases. *J. Biol. Chem.*, **275**, 20231–20234.
- Doungamath, A., Walker, M., Beismann-Driemeyer, S., Vega-Fernandez, C.M., Sterner, R. and Wilmanns, M. (2002) Structural evidence for ammonia tunnelling across the (β/α)8-barrel of the imidazole glycerol phosphate synthase bienzyme complex. *Structure*, **10**, 185–193.
- Eads, J.C., Ozturk, D., Wexler, T.B., Grubmeyer, C. and Sacchettini, J.C. (1997) A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. *Structure*, **10**, 185–193.
- Gerlt, J.A. and Babbitt, P.C. (2001) Divergent evolution of enzymatic

- function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.*, **70**, 209–246.
- Hayward, S., Kitao, A. and Berendsen, H.J. (1997) Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins*, **27**, 425–437.
- Henn-Sax, M., Höcker, B., Wilmanns, M. and Sterner, R. (2001) Divergent evolution of ($\beta\alpha$)8-barrel enzymes. *Biol. Chem.*, **382**, 1315–1320.
- Höcker, B., Jürgens, C., Wilmanns, M. and Sterner, R. (2001) Stability, catalytic versatility and evolution of the ($\beta\alpha$)(8)-barrel fold. *Curr. Opin. Biotechnol.*, **12**, 376–381.
- Ivens, A., Mayans, O., Szadkowski, H., Wilmanns, M. and Kirschner, K. (2001) Purification, characterization and crystallization of thermostable anthranilate phosphoribosyltransferase from *Sulfolobus solfataricus*. *Eur. J. Biochem.*, **268**, 2246–2252.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
- Kinoshita, K., Sadanami, K., Kidera, A. and Go, N. (1999) Structural motif of phosphate-binding site common to various protein superfamilies: all-against-all structural comparison of protein–mononucleotide complexes. *Protein Eng.*, **12**, 11–14.
- Kleywegt, G.J. and Jones, T.A. (1999) Software for handling macromolecular envelopes. *Acta Crystallogr. D*, **55**, 941–944.
- Knight, S.D. (2000) RSPS version 4.0: a semi-interactive vector-search program for solving heavy-atom derivatives. *Acta Crystallogr. D*, **56**, 42–47.
- Knöchel, T., Ivens, A., Hester, G., Gonzalez, A., Bauerle, R., Wilmanns, M., Kirschner, K. and Jansonius, J.N. (1999) The crystal structure of anthranilate synthase from *Sulfolobus solfataricus*: functional implications. *Proc. Natl Acad. Sci. USA*, **96**, 9479–9484.
- LaFortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.
- Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. and Wilmanns, M. (2000) Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science*, **289**, 1546–1550.
- Laskowski, R.A., MacArthur, M.W. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lohkamp, B., Coggin, J.R. and Laphorn, A.J. (2000) Purification, crystallization and preliminary X-ray crystallographic analysis of ATP-phosphoribosyltransferase from *Escherichia coli*. *Acta Crystallogr. D*, **56**, 1488–1491.
- Morollo, A.A. and Eck, M.J. (2001) Structure of the cooperative allosteric anthranilate synthase from *Salmonella typhimurium*. *Nature Struct. Biol.*, **8**, 243–247.
- Mushegian, A.R. and Koonin, E.V. (1994) Unexpected sequence similarity between nucleosidases and phosphoribosyltransferases of different specificity. *Protein Sci.*, **3**, 1081–1088.
- Otwinowski, Z. and Minor, V. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Perrakis, A., Morris, R. and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.*, **6**, 458–463.
- Phillips, C.L., Ullman, B., Brennan, R.G. and Hill, C.P. (1999) Crystal structures of adenine phosphoribosyltransferase from *Leishmania donovani*. *EMBO J.*, **18**, 3533–3545.
- Pugmire, M.J. and Ealick, S.E. (1998) The crystal structure of pyrimidine nucleoside phosphorylase in a closed conformation. *Structure*, **6**, 1467–1479.
- Pugmire, M.J. and Ealick, S.E. (2002) Structural analyses reveal two distinct families of nucleoside phosphorylases. *Biochem. J.*, **361**, 1–25.
- Pugmire, M.J., Cook, W.J., Jasanoff, A., Walter, M.R. and Ealick, S.E. (1998) Structural and theoretical studies suggest domain movement produces an active conformation of thymidine phosphorylase. *J. Mol. Biol.*, **281**, 285–299.
- Sinha, S.C. and Smith, J.L. (2001) The PRT protein family. *Curr. Opin. Struct. Biol.*, **11**, 733–739.
- Smith, D.A., Parish, T., Stoker, N.G. and Bancroft, G.J. (2001) Characterization of auxotrophic mutants of *Mycobacterium tuberculosis* and their potential as vaccine candidates. *Infect. Immun.*, **69**, 1142–1150.
- Spraggon, G., Kim, C., Nguyen-Huu, X., Yee, M.C., Yanofsky, C. and Mills, S.E. (2001) The structures of anthranilate synthase of *Serratia marcescens* crystallized in the presence of (i) its substrates, chorismate and glutamine and a product, glutamate and (ii) its end-product inhibitor, L-tryptophan. *Proc. Natl Acad. Sci. USA*, **98**, 6021–6026.
- Vonrhein, C. and Schulz, G.E. (1999) Locating proper non-crystallographic symmetry in low-resolution electron-density maps with the program GETAX. *Acta Crystallogr. D*, **55**, 225–229.
- Walter, M.R., Cook, W.J., Cole, L.B., Short, S.A., Koszalka, G.W., Krenitsky, T.A. and Ealick, S.E. (1990) Three-dimensional structure of thymidine phosphorylase from *Escherichia coli* at 2.8 Å resolution. *J. Biol. Chem.*, **265**, 14016–14012.
- Wilmanns, M., Hyde, C.C., Davies, D.R., Kirschner, K. and Jansonius, J.N. (1991) Structural conservation in parallel β/α -barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry*, **30**, 9161–9169.
- Yanofsky, C., Miles, E., Bauerle, R. and Kirschner, K. (1999) The trp operon and tryptophan biosynthesis in *Escherichia coli* and *Salmonella enterica*. In Creighton, T.E. (ed.), *The Encyclopedia of Molecular Biology*. Vol. 4. Wiley, New York, NY, pp. 2276–2689.
- Zalkin, H. and Smith, J.L. (1998) Enzymes utilizing glutamine as an amide donor. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **72**, 87–144.

Received January 17, 2002; revised April 18, 2002;
accepted April 22, 2002