

Crystal structures of a new class of pyrimidine/purine nucleoside phosphorylase (ppnP) revealed a Cupin fold

Yan Wen¹, Xiaojia Li¹, Wenting Guo¹, and Baixing Wu¹

¹Sun Yat-Sen Memorial Hospital

November 23, 2021

Abstract

Nucleotides metabolism is a fundamental process in all organisms. Two families of nucleoside phosphorylases (NP) that catalyze the phosphorolytic cleavage of the glycosidic bond in nucleosides have been found, including the trimeric or hexameric NP-I and dimeric NP-II family enzymes. Recently studies revealed another class of NP protein in *E. coli* named Pyrimidine/purine nucleoside phosphorylase (ppnP), which can catalyze the phosphorolysis of diverse nucleosides and yield D-ribose 1-phosphate and the respective free bases. Here, we solve the crystal structures of ppnP from *E. coli* and the other three species. Our studies revealed that the structure of ppnP belongs to the Rlmc-like cupin fold and showed as a rigid dimeric conformation. Detail analysis revealed a potential nucleoside binding pocket full of hydrophobic residues. And the residues involved in the dimer and pocket formation are all well conserved in bacteria. Since the cupin fold is a large superfamily in the biosynthesis of natural products, our studies provide the structural basis for understanding and the directed evolution of NP proteins.

Crystal structures of a new class of pyrimidine/purine nucleoside phosphorylase (ppnP) revealed a Cupin fold

Yan Wen^{1,2,#}, Xiaojia Li^{1,3,#}, Wenting Guo¹, Baixing Wu^{1, *}

¹ Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Guangdong-Hong Kong Joint Laboratory for RNA Medicine, RNA Biomedical Institute, Medical Research Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China 510120.

² Breast Tumor Center, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, 510120, China.

³ Department of Obstetrics and Gynecology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, Guangdong, China

* Co-corresponding author: Baixing Wu (E-mail: wubx28@mail.sysu.edu.cn)

These authors contributed equally to this work.

ABSTRACT

Nucleotides metabolism is a fundamental process in all organisms. Two families of nucleoside phosphorylases (NP) that catalyze the phosphorolytic cleavage of the glycosidic bond in nucleosides have been found, including the trimeric or hexameric NP-I and dimeric NP-II family enzymes. Recently studies revealed another class of NP protein in *E. coli* named Pyrimidine/purine nucleoside phosphorylase (ppnP), which can catalyze the phosphorolysis of diverse nucleosides and yield D-ribose 1-phosphate and the respective free bases. Here, we solve the crystal structures of ppnP from *E. coli* and the other three species. Our studies revealed that the structure of ppnP belongs to the Rlmc-like cupin fold and showed as a rigid dimeric conformation. Detail analysis revealed a potential nucleoside binding pocket full of hydrophobic residues. And the residues involved in the dimer and pocket formation are all well conserved in bacteria. Since the cupin

fold is a large superfamily in the biosynthesis of natural products, our studies provide the structural basis for understanding and the directed evolution of NP proteins.

KEYWORDS

Nucleotides metabolism; ppnP; Cupin

INTRODUCTION

Nucleotides metabolism is a fundamental process involving a series of chemical reactions to the nitrogenous base, pentose sugar, and phosphate. The differences of pyrimidine or purine nucleotides are determined by their nitrogenous bases, which are surveilled by several enzymes to keep the metabolic balance¹. The biosynthesis of nucleotides mainly proceeds through two pathways in the cell: 1) the *de novo* pathway, which uses a variety of amino acids and other precursors to produce nucleotides; and 2) the salvage pathway, which uses preformed nucleobases and nucleosides as precursors to produce nucleotides¹⁻⁴. The salvage pathway enables the cell to avoid the energy-costly *de novo* pathway when appropriate precursors are available. In the salvage pathway, one of the important biochemical reactions is the reversible phosphorolysis of purine and pyrimidine nucleosides. Two families of nucleoside phosphorylases that catalyze the phosphorolytic cleavage of the glycosidic bond in nucleosides have been discovered, the trimeric or hexameric nucleoside phosphorylase-I (NP-I) family enzymes share a common α/β -subunit fold and accept a range of purine nucleosides, as well as uridine nucleotide; and 2) the nucleoside phosphorylase-II (NP-II) family enzyme that display a dimeric structure and accept both thymidine and uridine in lower organisms, but are specific for thymidine in higher species¹. The proposed enzymic reaction for the purine/pyrimidine nucleosides was established previously, whereby the C-N glycosidic bond is cleaved by a phosphate ion⁵.

Recently, a combinative approach based on non-targeted metabolomics and activity-based enzyme discovery at the proteome scale discovered 241 potential novel enzymes in *E. coli*, 12 of which were experimentally validated⁶. One of these enzymes is Pyrimidine/purine nucleoside phosphorylase (ppnP), which is further identified to catalyze the phosphorolysis of diverse nucleosides such as uridine, adenosine, guanosine, cytidine, thymidine, inosine, and xanthosine as substrates, and yielding D-ribose 1-phosphate and the respective free bases⁶. Therefore, the reaction scheme of ppnP is indicated as *p* urine/*p* yrimidin*e*n nucleoside Phosphorylase and showed non-nucleotides specific and reversible activity⁶.

Here we determine the crystal structures of ppnP in many species of bacteria, our structural studies show that the ppnP family proteins present dual-layer beta-fence cupin fold, and adopt dimeric quaternary structure through a set of hydrophobic interactions. Detailed analysis also identified the potential substrate-binding pocket. Our structures uncover the structural basis for the new class nucleoside phosphorylase and shed light on deeply understanding the nucleotides' metabolism in bacteria and further usage of biocatalysts for industrial applications.

MATERIALS AND METHODS

2.1 Cloning, Expression, and Purification

The cDNA encoding *Escherichia coli* (strain K12) *Ec* ppnP, *Pseudomonas aeruginosa* (strain PA0750) *Pa* ppnP, *Vibrio cholerae* (strain SO5Y) *Vc* ppnP, *Salmonella enterica* (subsp. enterica serovar Typhimurium strain SL7207) *St* ppnP are ordered from Tsingke Biotechnology Co., Ltd. All the coding sequences of these genes are inserted into a modified pET-28a vector with 6xHis-Sumo tag and a ULP1 protease site. The vectors are transformed into *E. coli* BL21(*DE3*) competent cells and coated plates with Kana antibiotic. After incubation overnight, a single colony was picked to the LB medium supplied with 50 mg/L kanamycin. The *E. coli* cells were cultured in LB medium at 37 until the OD₆₀₀ reached 0.6-0.8, then the bacteria were induced with 0.2 mM IPTG at 18 for 12 h. The culture was collected by centrifugation, resuspended in buffer containing 20 mM Tris HCl pH8.0, 500 mM NaCl, 20 mM imidazole pH 8.0, and lysed by high pressure. Cell extracts were centrifuged at 17,000 rpm for 1h at 4. Supernatants were purified with Ni-NTA (GE), the target protein was washed with lysis buffer, and then eluted with a buffer containing 20mM Tris-HCl, pH8.0, 500mM NaCl, and 500mM imidazole. Ulp1 protease was added to remove the N-terminal tag and

fusion protein of the recombinant protein and dialyzed with lysis buffer for 3 hours. The mixture was applied to another Ni-NTA resin to remove the protease and uncleaved proteins. Eluted proteins were concentrated by centrifugal ultrafiltration, loaded onto a pre-equilibrated HiLoad 16/60 Superdex 75-pg column (GE Healthcare), eluted at a flow rate of 1 ml/min with the same buffer containing 10mM Tris-HCl pH8.0, 100 mM NaCl. Peak fractions were analyzed by SDS-PAGE (15%, w/v) and stained with Coomassie Brilliant Blue R-250. Purified fractions were pooled together and concentrated by centrifugal ultrafiltration. The concentration was determined by A280. The protein was concentrated to 10 mg/ml for crystallization trials.

2.2 Crystallization

All ppnP proteins were crystallized using the sitting drop vapor diffusion method by mixing 0.2 μ l of protein and 0.2 μ l of reservoir solution at 20 .

- 1) The crystal of EcppnP suitable for x-ray diffraction was grown in a reservoir solution consisting of 2.0 M Ammonium sulfate (dibasic), 0.1M CAPS/Sodium hydroxide pH 10.5, 0.2M Lithium sulfate (RIGAKU).
- 2) Selenomethionine-derived EcppnP suitable for x-ray diffraction was grown in reservoir solution consisting of 1.26 M Ammonium sulfate (dibasic), 0.1M Sodium cacodylate/Hydrochloric acid pH 6.5 (RIGAKU).
- 3) The crystal of StppnP suitable for x-ray diffraction was grown in reservoir solution consisting of 0.2 M Lithium sulfate monohydrate, 0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350 (Hampton Research).
- 4) The crystal of VcppnP suitable for x-ray diffraction was grown in a reservoir solution consisting of 0.2 M Ammonium sulfate, 30% w/v Polyethylene glycol 8,000 (Hampton Research).
- 5) The crystal of PappnP suitable for x-ray diffraction was grown in a reservoir solution consisting of 0.1 M BIS-TRIS pH 5.5, 2.0 M Ammonium sulfate. (Hampton Research)

2.3 X-ray data collection

Data collection was performed at 100K with cryoprotectant solution (reservoir solution supplemented with an additional 25% (v/v) glycerol). Diffraction data were collected at beamline BL18U1 of the Shanghai Synchrotron Radiation Facility (SSRF).

2.4 Structure solution and refinement

For *Ec* ppnP, the diffraction data set was processed and scaled using HKL3000⁷. The phase was determined by Selenomethionine-derived *Ec* ppnP using the program hkl2map⁸. And all the native ppnP structures are solved by Phaser with the structure of Se-*Ec* ppnP as the search model⁹. Cycles of refinement and model building were carried out using REFMAC5¹⁰, COOT¹¹, and PHENIX¹² until the crystallography reached a reasonable range. The details of data collection and processing are presented in **Table 1**. All structure figures were prepared with PyMOL (DeLano Scientific). The coordinates that support the findings of this study have been deposited in the Protein Data Bank with accession codes for 7EYJ for Native *Ec* ppnP, 7EYK for Se-*Ec* ppnP, 7EYP for *Pa* ppnP, 7EYL for *St* ppnP, and 7EYM for *Vc* ppnP, respectively.

3. RESULTS AND DISCUSSION

The overall structure of *E*cppnP shows RmlC-like cupins fold

The recombinant native *Ec* ppnP protein was crystallized in space group *P* 62 diffracted to 1.38 Å resolution, and the initial phase was determined by the single-wavelength anomalous dispersion (SAD) method using selenomethionine-labeled crystal, which is also diffracted to 1.38 Å resolution belongs to the same space group (**Table 1**). The final R_{work} and R_{free} values of the high-resolution structure are 0.1522 and 0.1780, respectively (**Table 1**). The electron density is continuous for the full-length protein; thus, we can model the full-length *Ec* ppnP structure. There is one *Ec* ppnP molecule in the asymmetric unit and the overall structure of *Ec* ppnP contains eleven beta-strands composing a dual-layer β -fence (**Figure 1A**). The β 1- β 2- β 3- β 10- β 5- β 8 compose an antiparallel β -sheet and the other side contains an antiparallel β 4- β 9- β 6- β 7 sheet (**Figure 1A, B**). Therefore, the structure of *Ec* ppnP we solved belongs to the cupin fold superfamily,

which is among the most functionally diverse superfamily and comprises both enzymatic and non-enzymatic members containing either one or two cupin domains¹³. Within the conserved tertiary structure, the variety of biochemical functions is provided by minor variation of the residues in the active site and the identity of the bound metal ion¹⁴.

The *Ec*ppnP act as a non-canonical dimerization mode with a conserved dimer interface

We then analyzed the symmetric unit and found that the *Ec* ppnP adopt a pseudodimer conformation (**Figure 1C**). In the homodimer of *Ec* ppnP, the interactions between each subunit are mediated mainly by the anti-parallel 6-stranded β -sheet ($\beta 1$ - $\beta 2$ - $\beta 3$ - $\beta 10$ - $\beta 5$ - $\beta 8$), herein termed the inner sheet. The anti-parallel 5-stranded β -sheet ($\beta 7$ - $\beta 6$ - $\beta 9$ - $\beta 4$ - $\beta 11$) is located outside of the dimer and termed the outer sheet (**Figure 1C**). The interactions between the inner sheets are mostly mediated by the four β -strands $\beta 1$ - $\beta 2$ - $\beta 3$ - $\beta 10$ and are largely mediated by hydrophobic interactions. Residues involved in the dimer formation are symmetric in the *Ec* ppnP structure including Leu2, Ile15, Phe17, Val27, Val29, Val31, and Pro86 (**Figure 1D**). These residues mainly located in $\beta 2$ - $\beta 3$ composed a hydrophobic surface region for the dimer formation. Size exclusion chromatography (SEC) is a commonly used approach for the identification of a protein oligomeric state. According to our SEC results, the wildtype *Ec* ppnP protein is present as a dimer in the solution same as we observed in the crystal (**Figure 2A**). And then, we select to mutate the Val27 and Val29 residues to aspartate (V27D/V29D) and test the in-solution status of *Ec* ppnP with SEC. However, the mutate *Ec* ppnP is not stable enough and quickly degraded (**Figure 2B**).

To further validate this dimer status of *Ec* ppnP, we solved another three ppnP structures in different species including *Pseudomonas aeruginosa* (strain PA0750) *Pa* ppnP, *Vibrio cholerae* (strain SO5Y) *Vc* ppnP, *Salmonella enterica* (subsp. enterica serovar Typhimurium strain SL7207) *St* ppnP. The diffraction data of *Pa* ppnP, *Vc* ppnP, and *St* ppnP were also determined with high resolution to 1.50 Å, 1.4 Å, and 1.2 Å, respectively (**Table 1**). Interestingly, all of the three structures present dimer conformation in one asymmetric unit (**Figure 2C, 2D, 2E**). We then superimposed each structure of them to *Ec* ppnP, the superposition results showed that all the three structures have minimal RMSD to *Ec* ppnP, with 0.446 Å of *Pa* ppnP to *Ec* ppnP, 0.476 Å of *St* ppnP to *Ec* ppnP and 0.406 Å of *Vc* ppnP to *Ec* ppnP. Therefore, the homodimer conformation of *Ec* ppnP is well-conserved in other species both from the quaternary structure (**Figure 2C, 2D, 2E**).

Potential substrate-binding pocket

Ec ppnP protein was suggested to be the enzyme catalyzing the phosphorolysis of diverse nucleosides such as uridine, adenosine, guanosine, cytidine, thymidine, inosine, and xanthosine as substrates⁶. Thus, we further analyzed the potential substrate-binding region in ppnP proteins. All the four structures we obtained showed a groove with a relative hydrophobic pocket constitute by the bilayer β -sheet (**Figure 3A, 3B, 3C, 3D**). Detailly, the *Ec* ppnP contains many residues involved in the pocket formation, including Leu81, Met30, Tyr89, Met45, Phe79, Phe37, Phe8, Cys91, Tyr7 as the bottom of the pocket, and Glu43, Ser26, Ala40, Glu41, Gln3, Arg24, Met1 as the wall (**Figure 3E**). These residues, together with the ones involved in the dimer formation are all well conserved in many ppnP proteins from different species (**Figure 3F**). These results suggested the conservative conformation in bacteria.

ppnP lack the divalent metal-binding site

We then analyze the homolog proteins with the Dali lite server, the search result with monomeric *Ec* ppnP revealed many structures with high similarity but with low sequence identity (**Figure 4A**)¹⁵. However, many of these proteins are unknown functions. Therefore, based on the DALI results, we selected two published structures, which are complexed with their substrates, to further dissect the substrate-binding properties of *Ec* ppnP. The HppE from *Streptomyces wedmorensis* (4J1X) and BacB from *Bacillus subtilis* are selected to compare with *Ec* ppnP^{16,17}. HppE is a non-heme-dependent dioxygenase that catalyzes the oxidative epoxidation of (S)-2-hydroxypropylphosphonate into (1R,2S)-epoxypropylphosphonate and is shown as a homotetramer¹⁷. BacB is indicated to be responsible for the biosynthesis of dipeptide antibiotic bacilysin with double cupin fold¹⁶. Superposition of *Ec* ppnP with HppE and BacB showed the r.m.s.d.

about 0.996 Å and 1.105 Å with the whole chain of *Ec* ppnP (**Figure 4B,4C**). We then detailly compared the substrate-binding pocket region, both the HppE and BacB contains a metal chelated by histidine and other residues to promote the catalysis process, however, no metal can be modeled in the counterpart in *Ec* ppnP, furthermore, the EcppnP lack the corresponding amino acids that can be used to chelate the metal ion (**Figure 4D,4E,4F**).

The members of the NP-I family include PNPs (EC 2.4.2.1), UP (EC 2.4.2.3), and MTAP (EC 2.4.2.28). PNPs from a variety of species revealed either the trimeric form specific for guanine and hypoxanthine (2'-deoxy) ribonucleosides, or the hexameric form accepts adenine as well as guanine and hypoxanthine (2'-deoxy) ribonucleosides^{1,18-20}. UP functions as a hexamer specific for uridine nucleosides in bacteria and also accepts 2'-deoxypyrimidine nucleosides in higher organisms²¹. MTAP is reported to function as a trimer in all species except *Sulfolobus solfataricus*²², in which it functions as a hexamer and catalyzes the phosphorolysis of inosine, guanosine, and adenosine²³. Compared with the NP-I family, TP (EC 2.4.2.4) and pyrimidine nucleoside phosphorylases (PyNPs; EC 2.4.2.2) belong to the NP-II family and function as a dimer in all cases^{1,24,25}. Our structure studies revealed that the ppnP family proteins showed dimeric quaternary conformations like NP-II class, but with a different dimerization mode belonging to the cupin fold.

Trimeric PNPs (e.g. bPNP and hPNP) are specific for 6-oxo purine nucleosides, whereas the hexameric PNPs accept adenosine in addition to 6-oxo purine nucleosides. MTAP is specific for 5'-methylthioadenosine but will accept various substrate analogs^{26,27}. UP accepts pyrimidine nucleosides and lacks specificity at the 2'-ribose position in higher organisms. TP is the high specificity of TP for the 2'-deoxyribose moieties in these studies, interestingly, UP will not cleave cytidine, which is not cleaved by any known nucleoside phosphorylase²¹. The ppnP family proteins are validated to catalyze the reaction involving both purine and pyrimidine, indicating the broadest substrate selectivity. Although we tried our best to obtain the complex structure to illustrate the binding properties for these nucleotides, we cannot capture the substrates in our structures. Therefore, further studies are needed to illustrate the mechanism for ppnP in catalyzing various ribonucleosides. On the other hand, the main feature of the ppnP family proteins are members of the cupin fold superfamily, the structural comparison of ppnP with *Streptomyces wedmorensis* HppE and *Bacillus subtilis* BacB revealed a conserved pocket that has the potential to bind the substrate, however, in ppnP, it lacks the metal-binding residues. How the catalytic scheme is achieved needs further investigation.

CONCLUSIONS

The NP-I and NP-II families represent two structurally distinct families of enzymes that catalyze the fundamental biochemical reaction. These families represent the only two folds known to be involved in the phosphorolysis of a wide range of purine and pyrimidine nucleosides previously. Here, we solved multiple crystal structures of another NP family protein ppnP in bacteria. Our structures showed a conserved dimeric cupin fold with a high hydrophobic dimer interface. The conformational analysis suggested that the ppnP proteins may act as a third nucleoside phosphorylase family to catalyze the reactions. Further complex structures are needed to uncover the catalytic scheme for ppnP proteins to various nucleosides.

ACKNOWLEDGMENT

We thank the staffs from BL18U1 beamline of National Facility for Protein Science in Shanghai (NFPS) at Shanghai Synchrotron Radiation Facility, for assistance during data collection. This work was supported by grants from the National Natural Science Foundation of China (31900435 to B.W.), and the Guangdong Science and Technology Department (2020B1212060018 and 2020B1212030004 to B.W.).

AUTHOR CONTRIBUTIONS

Yan Wen and Xiaojia Li expressed, purified, and grew crystals of the four ppnP proteins. Yan Wen and Baixing Wu collected X-ray diffraction data. Baixing Wu solved the structures, wrote and revised the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Baixing Wu: 0000-0003-2502-9785

REFERENCES

1. Pugmire MJ, Ealick SE. Structural analyses reveal two distinct families of nucleoside phosphorylases. *The Biochemical journal*.2002;361(Pt 1):1-25.
2. Pedley AM, Benkovic SJ. A New View into the Regulation of Purine Metabolism: The Purinosome. *Trends in biochemical sciences*.2017;42(2):141-154.
3. Yehia H, Kamel S, Paulick K, Wagner A, Neubauer P. Substrate spectra of nucleoside phosphorylases and their potential in the production of pharmaceutically active compounds. *Curr Pharm Des*. 2017.
4. Kamel S, Yehia H, Neubauer P, Wagner A. Enzymatic Synthesis of Nucleoside Analogues by Nucleoside Phosphorylases. In: *Enzymatic and Chemical Synthesis of Nucleic Acid Derivatives*. %U <https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527812103.ch1>; 2019:1-28.
5. Il'icheva IA, Polyakov KM, Mikhailov SN. Strained Conformations of Nucleosides in Active Sites of Nucleoside Phosphorylases.*Biomolecules*. 2020;10(4).
6. Sevin DC, Fuhrer T, Zamboni N, Sauer U. Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in Escherichia coli. *Nat Methods*. 2017;14(2):187-194.
7. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta crystallographica Section D, Biological crystallography*. 2006;62(Pt 8):859-866.
8. Pape T, Schneider TR. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *Journal of Applied Crystallography* %@ 0021-8898. 2004;37(5):843-844.
9. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *Journal of applied crystallography*. 2007;40(Pt 4):658-674.
10. Murshudov GN, Skubak P, Lebedev AA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica Section D, Biological crystallography*. 2011;67(Pt 4):355-367.
11. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta crystallographica Section D, Biological crystallography*. 2004;60(Pt 12 Pt 1):2126-2132.
12. Liebschner D, Afonine PV, Baker ML, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*. 2019;75(Pt 10):861-877.
13. Dunwell JM, Purvis A, Khuri S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry*. 2004;65(1):7-17.
14. Dunwell JM. Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins.*Biotechnol Genet Eng Rev*. 1998;15:1-32.
15. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D.*Nucleic acids research*. 2010;38(Web Server issue):W545-549.
16. McLuskey K, Cameron S, Hammerschmidt F, Hunter WN. Structure and reactivity of hydroxypropylphosphonic acid epoxidase in fosfomycin biosynthesis by a cation- and flavin-dependent mechanism.*Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(40):14221-14226.

17. Chang WC, Dey M, Liu P, et al. Mechanistic studies of an unprecedented enzyme-catalysed 1,2-phosphono-migration reaction. *Nature*. 2013;496(7443):114-118.
18. Kalckar HM. The enzymatic synthesis of purine ribosides. *The Journal of biological chemistry*. 1947;167(2):477-486.
19. Senesi S, Falcone G, Mura U, Sgarrella F, Ipata PL. A specific adenosine phosphorylase, distinct from purine nucleoside phosphorylase. *FEBS letters*. 1976;64(2):353-357.
20. Seeger C, Poulsen C, Dandanell G. Identification and characterization of genes (xapA, xapB, and xapR) involved in xanthosine catabolism in Escherichia coli. *J Bacteriol*.1995;177(19):5506-5516.
21. Krenitsky TA, Mellors JW, Barclay RK. Pyrimidine Nucleosidases. Their Classification and Relationship to Uric Acid Ribonucleoside Phosphorylase. *The Journal of biological chemistry*.1965;240:1281-1286.
22. Cacciapuoti G, Porcelli M, Bertoldo C, De Rosa M, Zappia V. 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. *The Journal of biological chemistry*. 1994;269(40):24762-24769.
23. Appleby TC, Mathews, II, Porcelli M, Cacciapuoti G, Ealick SE. Three-dimensional structure of a hyperthermophilic 5'-deoxy-5'-methylthioadenosine phosphorylase from Sulfolobus solfataricus. *The Journal of biological chemistry*.2001;276(42):39232-39242.
24. Paegle LM, Schlenk F. Bacterial uracil riboside phosphorylase. *Arch Biochem Biophys*. 1952;40(1):42-49.
25. Kamel S, Thiele I, Neubauer P, Wagner A. Thermophilic nucleoside phosphorylases: Their properties, characteristics and applications. *Biochim Biophys Acta Proteins Proteom*. 2020;1868(2):140304.
26. Ferro AJ, Wrobel NC, Nicolette JA. 5-methylthioribose 1-phosphate: a product of partially purified, rat liver 5'-methylthioadenosine phosphorylase activity. *Biochimica et biophysica acta*.1979;570(1):65-73.
27. Zappia V, Oliva A, Cacciapuoti G, Galletti P, Mignucci G, Carteni-Farina M. Substrate specificity of 5'-methylthioadenosine phosphorylase from human prostate. *The Biochemical journal*.1978;175(3):1043-1050.

Table 1. Data collection and refinement statistics

	EcppnP	EcppnP-SeMet	PappnP	StppnP	VcppnP
Data collection					
Wavelength	0.97915	0.97915	0.97915		
Space group	P62	P62	P3 ₂ 2 ₁	P3 ₁	P4 ₁ 2 ₁ 2
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	59.89, 59.89, 52.15	59.66, 59.66, 52.65	61.60, 61.60, 96.48	34.57, 34.57, 124.89	55.24, 55.24, 96.48
α , β , γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 90
Resolution (Å)	50.0-1.38 (1.43-1.38)*	50-1.38 (1.43-1.38)	30.0-1.50 (1.55-1.50)	50.0-1.20(1.24-1.20)	50.0-1.38(1.43-1.38)
<i>R</i> _{merge}	0.084 (0.196)	0.077 (0.505)	0.103 (0.740)	0.070 (0.198)	0.112 (0.505)
<i>I</i> / σ <i>I</i>	37.3 (18.7)	36.41 (7.00)	30.88 (6.75)	19.558 (7.824)	26.33 (18.7)
Completeness (%)	99.9 (99.5)	100.0 (100.0)	97.6 (96.3)	99.7 (98.8)	99.6 (97.6)
Redundancy	19.9 (19.1)	19.8 (19.4)	20.0 (19.1)	4.3 (4.2)	12.5 (6.3)
Refinement					
Resolution (Å)	20.0-1.38	20.0-1.38	20.0-1.50	20.0-1.20	20.0-1.38
No. reflections	21,825	21,787	33, 684	51507	35447
<i>R</i> _{work} / <i>R</i> _{free}	0.1522/0.1780	0.1507/0.1688	0.1794/0.2127	0.1738/0.1893	0.2054/0.2127
No. atoms	891	904	1,801	1,649	1,601
Protein	760	753	1,494	1,449	1,491
Ligand/Water	131	151	307	200	110
<i>B</i> -factors	15.61	16.32	17.32	17.69	17.35

	EcppnP	EcppnP-SeMet	PappnP	StppnP	VcppnP
R.m.s. deviations					
Bond lengths (Å)	0.005	0.005	0.008	0.004	0.006
Bond angles (°)	0.867	0.874	1.019	0.841	0.952
Ramachandran plot					
Favored/allowed (%)	96.77/3.23	96.77/3.23	98.91/1.09	97.85/2.15	98.39/1.61

* Highest resolution shell is shown in parentheses.

Figure legend

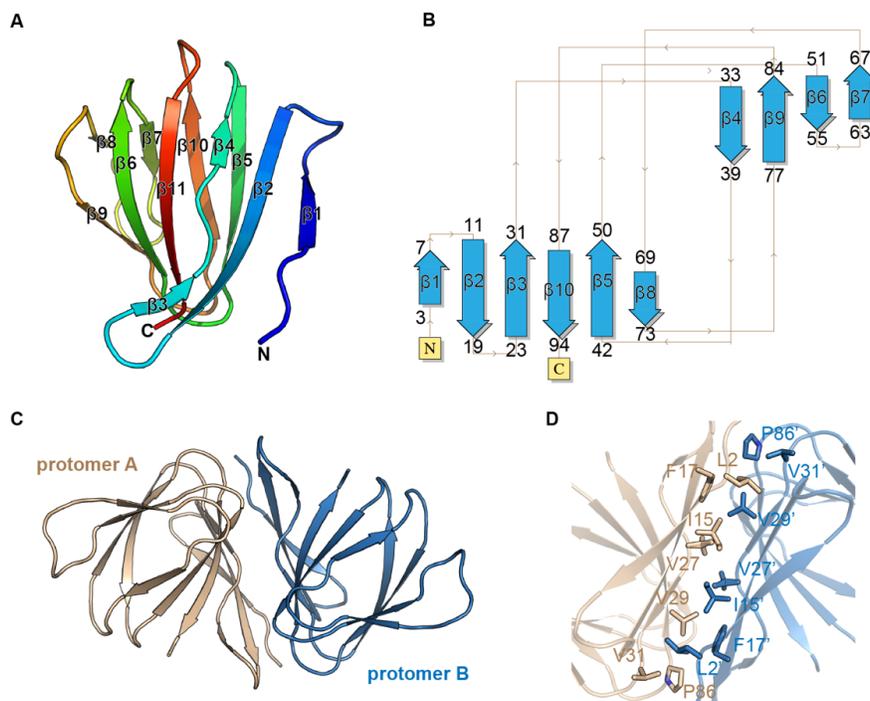


Figure 1 Structure of *Ec*ppnP shows a cupin fold. (A) The overall structure of monomeric *Ec* ppnP is shown by ribbon and colored by spectrum. (B) Topological diagram of the structure of *Ec* ppnP. (C) The dimeric structure of *Ec* ppnP. The two protomers in the dimer conformation are colored as wheat and blue, respectively. (D) Dimer interface of *Ec* ppnP. The residues involved in dimerization are shown as the stick.

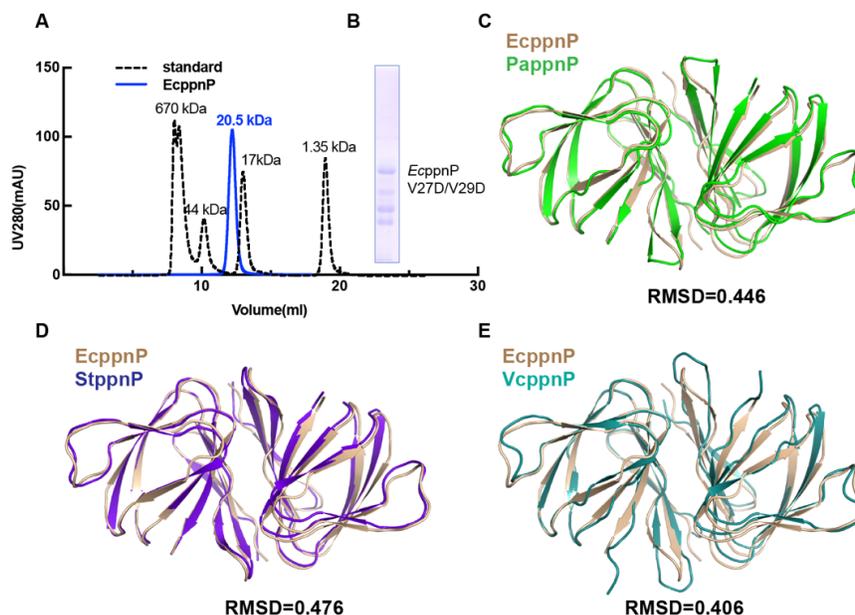


Figure 2 Dimerization of ppnP is conserved in bacteria. (A) In solution validation the aggregation status of the *Ec* ppnP by gel filtration analysis. (B) SDS-PAGE showed that the double mutant V27D/V29D of *Ec* ppnP is not stable. (C-E) The dimeric structural comparison of *Ec* ppnP to *Pa* ppnP, *St* ppnP, and *Vc* ppnP. The *Pa* ppnP is colored green, the *St* ppnP is colored in purple-blue, and the *Vc* ppnP is colored in teal.

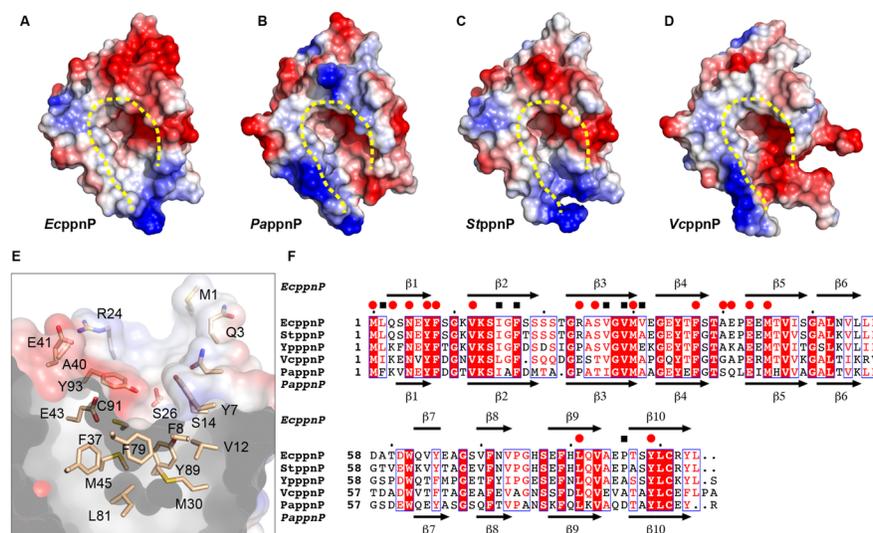


Figure 3 ppnP proteins have conserved pocket (A-D) Electrostatics surface representation of the ppnP structures. The potential substrate-binding pocket is indicated by yellow dash lines. (E) Residues composed of the pocket of *Ec* ppnP are shown as the stick. (F) Sequence alignment of *Ec* ppnP, *Pa* ppnP, *St* ppnP, and *Vc* ppnP, and *Yersinia pestis* ppnP. The residues involved in dimerization are indicated by black squares. And the residues composed of the potential pocket are indicated by red circles.

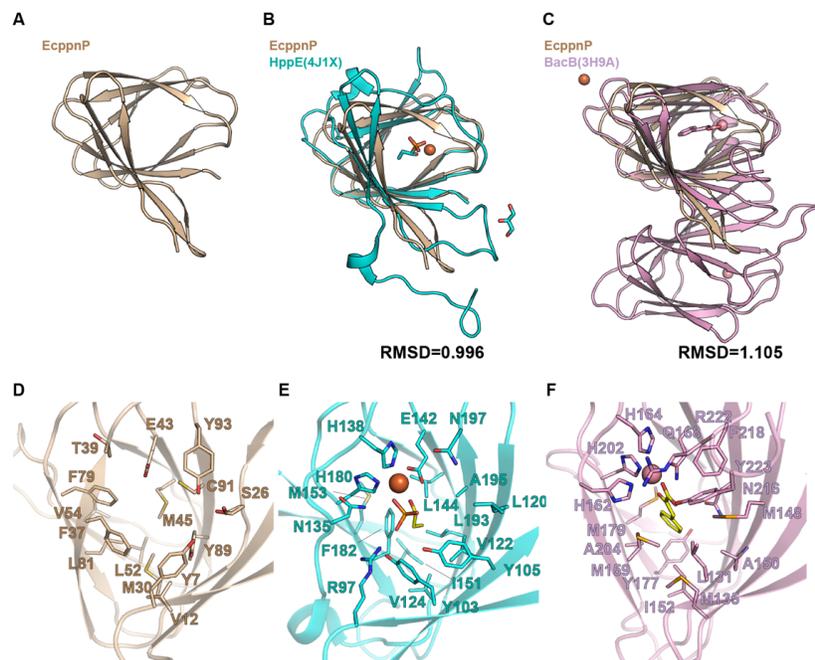


Figure 3 Comparisons of *Ec* ppnP with HppE and BacB(A) The monomeric structure of *Ec* ppnP. (B) Superposition of *Ec* ppnP with HppE (PDB code: 4J1X). The HppE is colored in cyan. The substrate or ligand of HppE is shown as the stick. The metal ion is shown as a sphere. (C) Superposition of *Ec* ppnP with BacB (PDB code: 3H9A). The HppE is colored in pink. The substrate of BacB is shown as the stick. The metal ion is shown as a sphere. (D) Potential substrate binding residues in *Ec* ppnP are shown as the stick. (E) Residues in HppE for substrate binding and catalytic reaction. The substrate is shown as a stick colored yellow. (F) Residues in BacB for substrate binding and catalytic reaction. The substrate is shown as a stick colored yellow.