



# The structure of His-tagged *Geobacillus stearothermophilus* purine nucleoside phosphorylase reveals a ‘spanner in the works’

Fiona M. Given, Fuchsia Moran, Ashleigh S. Johns, James A. Titterington, Timothy M. Allison, Deborah L. Crittenden and Jodie M. Johnston\*

Received 20 October 2022  
Accepted 16 November 2022

Edited by F. T. Tsai, Baylor College of Medicine, Houston, USA

**Keywords:** purine nucleoside phosphorylase; *Geobacillus stearothermophilus*; His tags; affinity tags.

**PDB reference:** purine nucleoside phosphorylase, 8d38

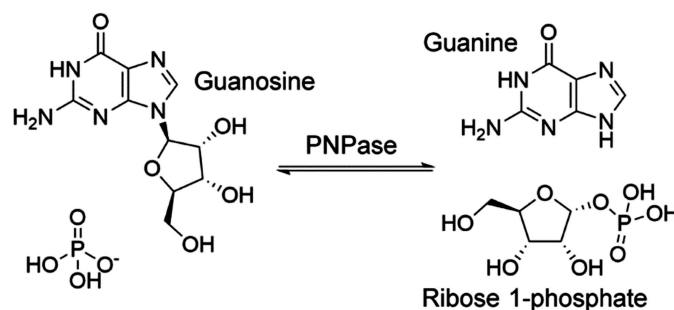
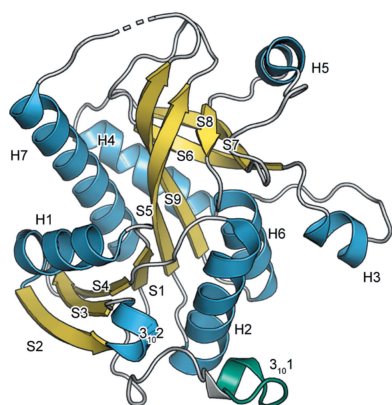
**Supporting information:** this article has supporting information at journals.iucr.org/f

School of Physical and Chemical Sciences, Biomolecular Interaction Centre, University of Canterbury, New Zealand.  
\*Correspondence e-mail: jodie.johnston@canterbury.ac.nz

The 1.72 Å resolution structure of purine nucleoside phosphorylase from *Geobacillus stearothermophilus*, a thermostable protein of potential interest for the biocatalytic synthesis of antiviral nucleoside compounds, is reported. The structure of the N-terminally His-tagged enzyme is a hexamer, as is typical of bacterial homologues, with a trimer-of-dimers arrangement. Unexpectedly, several residues of the recombinant tobacco etch virus protease (rTEV) cleavage site from the N-terminal tag are located in the active site of the neighbouring subunit in the dimer. Key to this interaction is a tyrosine residue, which sits where the nucleoside ring of the substrate would normally be located. Tag binding appears to be driven by a combination of enthalpic, entropic and proximity effects, which convey a particularly high affinity in the crystallized form. Attempts to cleave the tag in solution yielded only a small fraction of untagged protein, suggesting that the enzyme predominantly exists in the tag-bound form in solution, preventing rTEV from accessing the cleavage site. However, the tagged protein retained some activity in solution, suggesting that the tag does not completely block the active site, but may act as a competitive inhibitor. This serves as a warning that it is prudent to establish how affinity tags may affect protein structure and function, especially for industrial biocatalytic applications that rely on the efficiency and convenience of one-pot purifications and in cases where tag removal is difficult.

## 1. Introduction

Purine nucleoside phosphorylase (PNPase) is an enzyme in the purine salvage pathway that is present in some organisms. This enzyme performs the reversible phosphorolysis of nucleosides to form ribose 1-phosphate and a purine base (Fig. 1; Bzowska *et al.*, 2000). There are two major structural forms of PNPase: a trimeric form primarily found in eukaryotes and a hexameric ‘trimer-of-dimers’ form primarily found in bacteria (Štefanić *et al.*, 2017). The trimeric and hexameric



**Figure 1**  
The enzymatic reaction performed by PNPase. The purine base and nucleoside can be modified to create different base/nucleoside pairs.

forms of PNPase have similar subunit structures, namely an  $\alpha/\beta$  fold (Mao *et al.*, 1997), and differ primarily in substrate specificity, with most studied trimeric PNPases restricted to 6-oxopurine ribonucleosides, while hexameric PNPases can use both 6-oxopurine ribonucleosides and 6-aminopurine ribonucleosides (Bennett *et al.*, 2003; Ducati *et al.*, 2009). Whilst the hexameric form is primarily found in bacteria, many bacteria have both forms encoded in their genome, including *Geobacillus stearothermophilus*, which is a Gram-positive thermophile bacterium and a common cause of spoilage in food products (Hamamoto *et al.*, 1997).

PNPases are potential antimicrobial drug targets (Pant *et al.*, 2021; Madrid *et al.*, 2008) and industrial catalysts for the production of antiviral nucleoside compounds (Nannemann *et al.*, 2010; Xie *et al.*, 2011). The substrate promiscuity of PNPases is relevant for the latter application, as the wider the range of purine bases that a PNPase can accommodate, the broader the scope of antiviral nucleoside candidates that it may be used to produce. As many of the potential purine bases have limited solubility at ambient temperature (Zhu *et al.*, 2013), PNPases from thermophilic organisms, such as *G. stearothermophilus*, could be used at higher temperatures with consequently improved substrate solubility and thus allow the production of a wider range of these compounds. As part of our exploration of the biocatalytic potential of thermostable PNPases, we solved the structure of the N-terminally His-tagged PNPase from *G. stearothermophilus* (His-*Gse*PNPase).

## 2. Materials and methods

### 2.1. Macromolecule production

The His-*Gse*PNPase gene from *G. stearothermophilus* (KEGG GT50\_19015[TA1]) was purchased as a synthetic gene from GenScript. The gene was amplified and inserted into a pET-30a expression vector by In-Fusion cloning (Park *et al.*, 2015) using NcoI and XhoI restriction sites. The resulting plasmid encodes an N-terminal 6 $\times$ His tag, a thrombin cleavage site, an S-tag followed by a tobacco etch virus protease (TEV) recognition sequence, and the opening reading frame of interest.

The purified plasmid was transformed into chemically competent *Escherichia coli* BL21 (DE3) cells for protein expression. The cultures were grown at 37°C in lysogeny broth medium with 100  $\mu$ M kanamycin until the optical density was in the range 0.4–0.6; protein expression was then induced using isopropyl  $\beta$ -D-1-thiogalactopyranoside at a final concentration of 100  $\mu$ M and the cells were grown overnight at 18°C. The cells were harvested by centrifugation at 12 000g for 15 min and the resultant cell pellets were stored at –80°C until use.

The cell pellet was resuspended in lysis buffer (50 mM HEPES pH 7.6, 100 mM NaCl, 20 mM imidazole) and lysed by sonication. Following centrifugation, the supernatant was filtered through a 0.22  $\mu$ m filter and then purified by immobilized metal-affinity chromatography (IMAC) on a HisTrap

Table 1

Macromolecule-production information.

Source organism	<i>Geobacillus stearothermophilus</i>
DNA source	Synthetic DNA
Cloning vector	pET-30a
Expression vector	pET-30a
Expression host	<i>Escherichia coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced	MHHHHHSSGLVPRGSGMKETAALKFERQH MDSPDLGTGSENLYFQGAMMSVHIGAKA GEIAERILLPGDPLRAKYIAETFLEGAV CYNEVRGMLGFTGTYKGERISVQGTGMG VPSISIVNELIQSYGVKTLIRVGTGGA IQPDVVRVDVILAMSASTDSNMNRLIFR GRDYAPTADFHLLRTAYEVGVEKGLALK VGNVFTADMFYNDPEPNWETWARYGVLA EMETAALYTLAAKFGCRALSVLTVSDHI LTGEETTAERQMTFNEMIEVALEAAIR NGA

Table 2

Crystallization.

Method	Vapour diffusion
Plate type	Sitting drop
Temperature (K)	291
Protein concentration (mg ml <sup>-1</sup> )	25
Buffer composition of protein solution	50 mM HEPES pH 7.6, 100 mM NaCl
Composition of reservoir solution	0.06 M magnesium chloride hexahydrate, 0.06 M calcium chloride dihydrate, 0.1 M Tris (base), 30%(v/v) Bicine pH 8.5, 40%(v/v) ethylene glycol, 20%(w/v) PEG 8000
Volume and ratio of drop	300 nl; 2:1 protein:reservoir solution
Volume of reservoir ( $\mu$ l)	40

FF column (Cytiva). The protein was eluted from the column using a gradient of elution buffer (50 mM HEPES pH 7.6, 100 mM NaCl, 500 mM imidazole). Size-exclusion chromatography on a Superdex 10/300 column run in storage buffer (50 mM HEPES pH 7.6, 100 mM NaCl) was then used to further purify the protein prior to crystallography. Purified protein was stored at –80°C until use. Protein concentration was determined from the UV absorbance in a NanoDrop spectrophotometer (ThermoFisher) using a theoretical molar extinction coefficient of 27 515 M<sup>-1</sup> cm<sup>-1</sup> as determined by *ProtParam* (Gasteiger *et al.*, 2005).

When cleavage by recombinant TEV (rTEV) was attempted, 1 mM tris(2-carboxyethyl)phosphine (TCEP) was added. Purified rTEV was added at a ratio of 1:10 rTEV:protein. The sample was left overnight at 4°C or incubated at 37°C for 4 h and then incubated at 4°C overnight before a second IMAC purification was performed. Macromolecule-production information is summarized in Table 1.

### 2.2. Crystallization

His-*Gse*PNPase was crystallized using 25 mg ml<sup>-1</sup> protein at a 2:1 protein:well solution ratio (drop volume 300 nl; well volume 40  $\mu$ l) using the commercial MORPHEUS protein crystallization screen (Gorrec, 2009) and a Mosquito robot (SPT Labtech). Crystals were harvested and cooled in liquid nitrogen. Crystallization information is summarized in Table 2.

**Table 3**

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	MX2 beamline, Australian Synchrotron
Wavelength (Å)	0.95374
Detector	Dectris EIGER 16M
Total rotation range (°)	720
Space group	$P4_32_12$
$a, b, c$ (Å)	102.96, 102.96, 167.76
$\alpha, \beta, \gamma$ (°)	90, 90, 90
Resolution range (Å)	49.220–1.720 (1.750–1.720)
Total No. of reflections	5180774 (256581)
No. of unique reflections	96171 (4713)
Completeness (%)	100.000 (99.900)
Multiplicity	53.900 (54.400)
Mn( $I$ ) half-set correlation ( $CC_{1/2}$ )	1.00 (0.599)
$\langle I/\sigma(I) \rangle$	18.500 (0.5)†
$R_{r.i.m.}$	0.157 (9.227)
$R_{p.i.m.}$	0.021
Overall $B$ factor from Wilson plot (Å <sup>2</sup> )	34.170

† The mean  $I/\sigma(I)$  in the outer shell falls below 2.0 at resolutions above 1.92 Å. Analyses of merged  $CC_{1/2}$  correlations between intensity estimates from half data sets (Karplus & Diederichs, 2015) were used to influence the high-resolution cutoff for data processing.

### 2.3. Data collection and processing

Diffraction data were collected at the Australian Synchrotron on the MX2 macromolecular crystallography beamline equipped with a Dectris EIGER 16M detector (Aragão *et al.*, 2018). From each crystal that diffracted, 720° of data were collected and processed using *X-ray Detector Software (XDS; Kabsch, 2010)*. The reflections were merged using *AIMLESS* (Evans & Murshudov, 2013) and the most likely space group,  $P4_32_12$ , was determined using *POINTLESS* from the *CCP4* suite (Winn *et al.*, 2011). An  $R_{free}$  set corresponding to 5% of reflections was assigned. Data-collection and processing statistics are summarized in Table 3.

### 2.4. Structure solution and refinement

A molecular-replacement search model (PDB entry 2ac7; adenosine phosphorylase from *Bacillus cereus*; A. Rinaldo-Matthis, S. Allegrini & F. Sgarrella, unpublished work), with the His-*Gse*PNPase sequence docked onto it, was used in molecular replacement to solve the structure of His-*Gse*PNPase. The His-*Gse*PNPase sequence was docked onto the model using *CHAINS AW* (Stein, 2008) and nonmatching regions were removed. Analysis of the Matthews coefficient (Matthews, 1968) revealed that there were likely to be three molecules in the asymmetric unit. Molecular replacement was completed using *Phaser* (McCoy *et al.*, 2007) in *CCP4* and was run in expert mode to find all possible space groups within the chosen point group ( $P422$ ), as well as to refine the solution and reduce the  $R$  factor. *REFMAC5* (Murshudov *et al.*, 2011) from the *CCP4* suite was then used to refine the solution further. The electron-density maps and model generated by the above programs were used in *Coot* (Emsley *et al.*, 2010) to manually build and refine the model. After each round of manual building, the model was refined further in *REFMAC5* or *Phenix* (Afonine *et al.*, 2012).  $R$  and  $R_{free}$  were monitored to determine the model quality. The structure has been deposited

**Table 4**

Structure refinement.

Resolution range (Å)	46.0500–1.7200 (1.7400–1.7200)
Completeness (%)	99.5
$\sigma$ Cutoff	$F > 1.330\sigma(F)$
No. of reflections, working set	90727 (2976)
No. of reflections, test set	4907 (149)
Final $R_{cryst}$	0.196 (0.4449)
Final $R_{free}$	0.224 (0.4695)
No. of non-H atoms	
Total	5755
Protein	5487
Ions	2
Solvent	266
R.m.s. deviations	
Bond lengths (Å)	0.013
Angles (°)	1.177
Average $B$ factors (Å <sup>2</sup> )	
Protein	42.58
Ions	34.29
Solvent	43.23
Ramachandran plot	
Most favoured (%)	97.45
Allowed (%)	2.55

in the PDB as entry 8d38. Refinement statistics are summarized in Table 4.

## 3. Results and discussion

### 3.1. His-*Gse*PNPase adopts a hexameric quaternary structure

His-*Gse*PNPase crystallized in the absence of ligands (Table 2) in space group  $P4_32_12$  (Table 3) and diffracted to a resolution of 1.72 Å. Three chains ( $a$ – $c$ ) were present in the asymmetric unit, from which the hexameric biological assembly can be generated through the application of crystallographic symmetry operations (Fig. 2*a*). The hexameric structure shows threefold symmetry through the ring-shaped hexamer and His-*Gse*PNPase can be viewed as being a trimer of dimers ( $a/b$ ,  $d'/b'$  and  $c/c'$ ), like other hexameric PNPases (Narczyk *et al.*, 2021). The observed electron density for each chain is continuous, apart from residues 209–210 in chain  $c$  and portions of the N-terminal tag.

### 3.2. Monomeric fold

Analysis of the fold of the subunits suggests that His-*Gse*PNPase is structurally similar to other hexameric PNPases, such as the hexameric PNPase from *E. coli* (*Ec*PNPase;  $C^\alpha$  r.m.s.d. of 0.614 Å; Mao *et al.*, 1997; Rinaldo-Matthis *et al.*, 2007; Grenha *et al.*, 2005). More generally, the His-*Gse*PNPase subunit structure possesses a conserved fold, with over 702 matches ( $C^\alpha$  r.m.s.d. of 0–2.5 Å) in a *PDB eFold* search (Krissinel & Henrick, 2004*a,b*, 2005; Krissinel *et al.*, 2004). Its structure is most like those of *Helicobacter pylori* PNPase (Narczyk *et al.*, 2018) and the well studied *Ec*PNPase (Narczyk *et al.*, 2021), with r.m.s.d.s of 0.42–0.51 Å. The subunits of His-*Gse*PNPase are similar to each other, with r.m.s.d.s of 0.136–0.160 Å on pairwise comparison. Each subunit is comprised of seven  $\alpha$ -helices (H1–H7) and two  $\beta$ -sheets (Fig. 2*b*). The first  $\beta$ -sheet (S1–S4) has four strands, with S3 antiparallel, and the

second sheet forms a sheet-like barrel roll (S5–S10) with S5 and S10 antiparallel. Based on analysis in SCOP, His-GsePNPase is part of the phosphorylase/hydrolase-like fold superfamily, as are other PNPases (Fox *et al.*, 2014).

### 3.3. High conservation between the His-GsePNPase and EcPNPase active sites

The active sites are formed close to the interfaces between two subunits (for example subunits *a* and *b*; Fig. 3*a*) and there are six active sites in the hexamer. The active site is mostly formed from residues from one subunit, with additional contributing residues from the other subunit in each dimer pair (Fig. 3*b*). There is high sequence and structural conservation between the active sites of *EcPNPase* and His-GsePNPase. In *EcPNPase*, the residues that interact with the purine base were identified as Ala156, Phe159, Val178, Met180, Asp204 and Ile206 (Bennett *et al.*, 2003), which are equivalent to Ala156, Phe159, Val177, Met179, Asp203 and Ile205, respectively, in His-GsePNPase (Fig. 3*b*). The ribose-binding site in *EcPNPase* is primarily formed by interactions with Glu181 from one subunit and His4 from an adjacent subunit (Bennett *et al.*, 2003), which are equivalent to Glu180 and His4, respectively, in His-GsePNPase (Fig. 3*b*). The phosphate-binding site of *EcPNPase* is comprised of two arginine residues, Arg87 from one subunit and Arg43 from an adjacent subunit, and these residues are also conserved in His-GsePNPase (Fig. 3*b*).

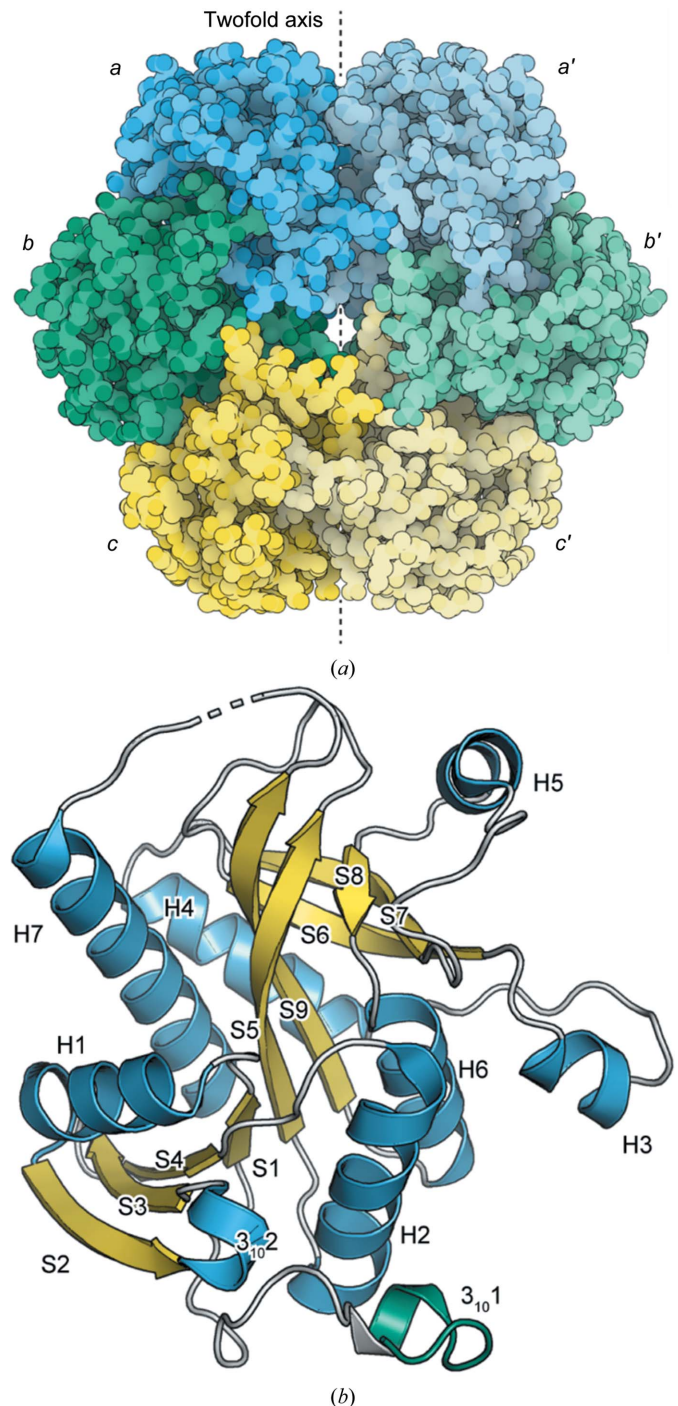
### 3.4. Occupation of the active site by the N-terminal tag from an adjacent subunit

One surprising aspect of the His-GsePNPase structure is the presence in the active site of several residues (NLYFQ) from the nine-residue rTEV site (ENLYFQGAM) in the N-terminal tag. This portion of the rTEV site was observed to be bound in the active site of the adjacent subunit in each dimer (Fig. 3*a*) consistently across all dimers in the hexameric structure. The interactions between the active site and the N-terminal tag are dominated by the Tyr residue in the rTEV recognition site. This residue is oriented in a similar way to the purine-base part of the substrate (Fig. 3*c*). This similarity in binding pose suggests that the hydrophobic stacking interactions that this residue makes with active-site residues may contribute to holding the N-terminal tag in its position bound to the active site and thus also potentially prevent access to and cleavage by the protease which cuts between Gln and Gly in the sequence.

### 3.5. N-terminal tag occupation prevents the His-GsePNPase active site from closing

PNPase structures feature an  $\alpha$ -helix (H7, using the  $\alpha$ -helix numbering for *EcPNPase*; Mao *et al.*, 1997) that can segment into two parts to close over the active site (Štefanić *et al.*, 2017). In His-GsePNPase H7 consists of residues Val200–Gln217. In *EcPNPase* two different active-site conformations have been observed, denoted as ‘open’ (nonsegmented form) and ‘closed’ (segmented form) (Štefanić *et al.*, 2018). In the closed conformation, this loop abrogates access to the active

site (Koellner *et al.*, 2002). In the structure of His-GsePNPase, the conformation of this loop in all three subunits is most consistent with the *EcPNPase* ‘open’ conformation (Fig. 3*d*), with the occupation of the active site by the tag precluding the



**Figure 2**  
The structure of His-GsePNPase. (a) The hexameric biological unit of PNPase. The asymmetric unit contains one-and-a-half dimers (subunits *a*–*c*) and the hexamer is generated by a twofold symmetry operation that produces subunits *a'*–*c'*. (b) The tertiary structure of a His-GsePNPase subunit. Secondary-structure elements are labelled and helices are coloured blue,  $\beta$ -sheets yellow and loop regions white. The portion of the structure corresponding to the N-terminal tag is coloured green and features a  $3_{10}$ -helix ( $3_{10}1$ ).

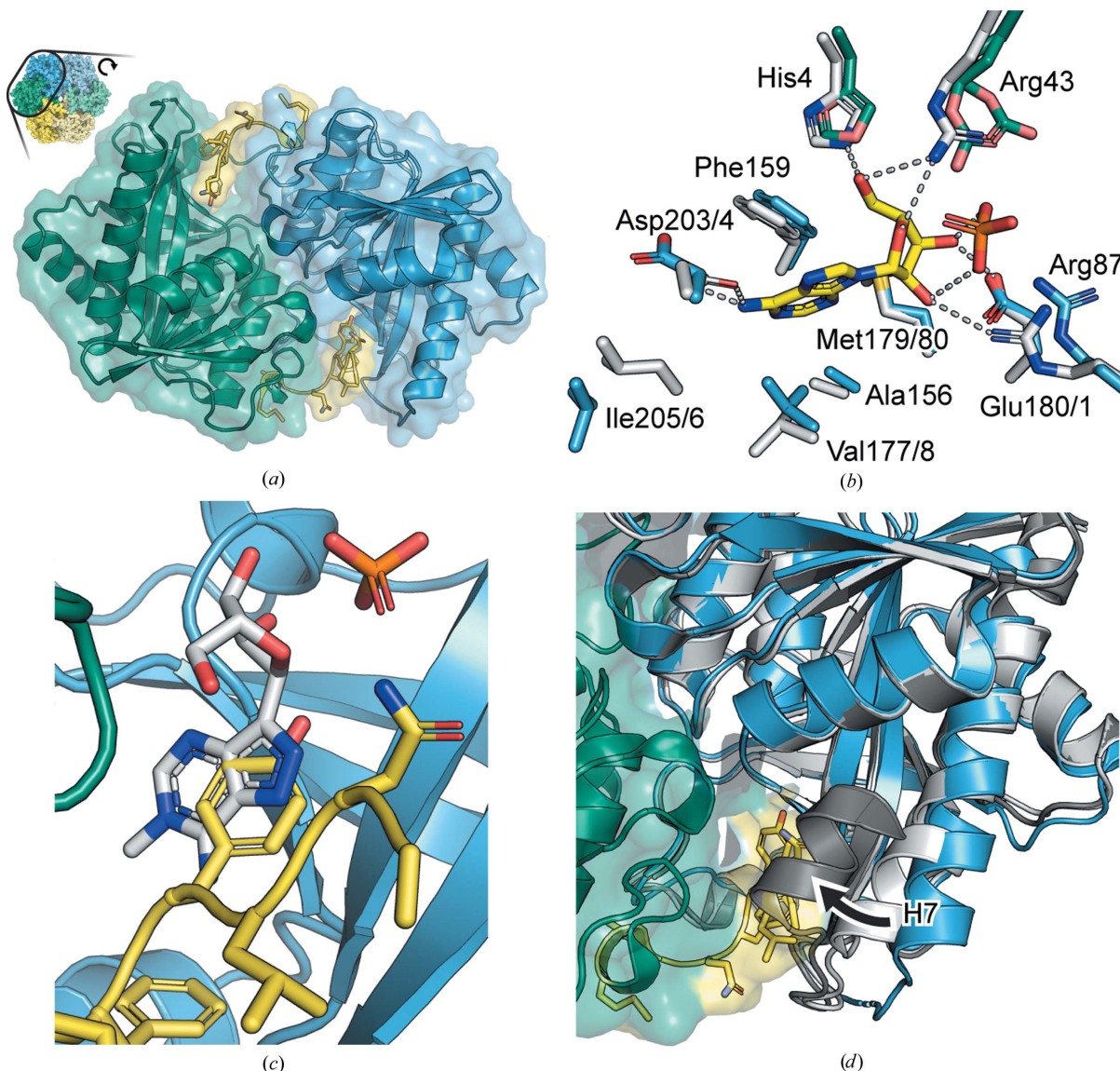
ability of the helix to segment and adopt the ‘closed’ conformation.

### 3.6. Interactions of the N-terminal tag with the active site may affect enzyme function

Before considering the functional implications of N-terminal tag binding in the active site, it is first necessary to investigate whether this is simply a crystallization artefact driven by proximity effects and entropic factors. We therefore attempted to cleave the tag and purify the untagged form, with a view to performing joint structural and functional characterization of the tagged and untagged forms. However, regardless of the

experimental conditions, only mixed populations of cleaved and uncleaved subunits were obtained, which were predominantly uncleaved. This result suggests that the tag may be difficult to remove because His-*Gse*PNPase exists in a tag-bound form in solution, precluding access of the protease to the cleavage site.

The strong affinity of the tag for the active site in the crystallized form is supported by the results of crystallization screening experiments in the presence of high concentrations of the PNPase inhibitor acyclovir and the substrate 7-methyl-guanine, which show that neither substrate is capable of displacing the tag from the active site. However, UV–Vis activity assays tell a different story, demonstrating that



**Figure 3**  
 (a) The dimeric substructure of His-*Gse*PNPase. The two subunits are independently coloured and the regions of each corresponding to the N-terminal tag are coloured yellow. The position of the N-terminal tag indicates the approximate position of the active sites in each dimer, which sit at the interface between the subunits. (b) A structural overlay of *Ec*PNPase (PDB entry 1pk7, white) and His-*Gse*PNPase (subunit *a* in blue and subunit *b* in green) highlighting the conservation of the active-site residues. Adenine and phosphate from the *Ec*PNPase structure are coloured yellow. Putative hydrogen-bonding interactions between these molecules and the *Ec*PNPase residues are shown as dashed lines. (c) Overlay of the Tyr5 residue from the rTEV site (yellow) and the adenine base from PDB entry 1k9s (white). (d) The segmenting H7 in the two different positions found in *Ec*PNPase (closed, dark grey; open, white, from PDB entry 1k9s chains *A* and *D*) and the position of this region in His-*Gse*PNPase (blue; the N-terminal tag is shown in yellow).

N-terminally tagged His-*Gse*PNPase is catalytically active in solution (data not shown). The enzymatic activity implies that the active site must be at least somewhat accessible to the substrate under standard assay conditions. However, it is important to keep in mind that while substrates can compete with the tag for active-site binding in solution, the reverse may also be true, *i.e.* the tag may compete with substrates, potentially affecting the activity of the enzyme.

Taken together, this evidence suggests that it is most likely that His-*Gse*PNPase exists in a tag-bound form in both the solution and crystalline phases, but with a lower affinity in solution due to entropic favourability of the disordered unbound form partially counterbalancing the binding enthalpy. This situation, along with the more dynamic nature of proteins in solution in general, may explain why substrate binding is observed in activity assays but not in crystallization screening.

### 3.7. Comparison with related superfamilies reveals a fold that can accommodate peptide binding

Although proximity effects undoubtedly contribute to promoting tag binding, *Gse*PNPase may also have a native affinity for peptide substrates and an ability to accommodate peptide binding. Analysis in SCOP (Andreeva *et al.*, 2014, 2020) of the purine/uridine phosphorylase superfamily that has the same phosphorylase/hydrolase-like fold as the PNPases reveals members that cleave peptide substrates (for example zinc-dependent exopeptidases, pyrrolidone carboxyl peptidase and the peptidyl-tRNA hydrolase-like and HybD-like superfamilies). Structural alignments between His-*Gse*PNPase and superfamily members, including carboxypeptidase T in complex with a protein inhibitor (PDB entry 1tdt; Reverter *et al.*, 2000), suggest that the location of the active site is conserved and the fold generally accommodates peptide binding. Notably, peptides generated from snake venom have been shown to be effective inhibitors of PNPase from *Plasmodium falciparum* (Martins *et al.*, 2019), suggesting that not only can PNPases bind peptides by that they may also be a potential route for the design of new PNPase inhibitors.

### Acknowledgements

This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, which is part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector. Access to the Australian Synchrotron was supported by the New Zealand Synchrotron Group Ltd.

### Funding information

Financial support from the Biomolecular Interaction Centre at the University of Canterbury is gratefully acknowledged.

### References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.

Andreeva, A., Howorth, D., Chothia, C., Kulesha, E. & Murzin, A. G. (2014). *Nucleic Acids Res.* **42**, D310–D314.

Andreeva, A., Kulesha, E., Gough, J. & Murzin, A. G. (2020). *Nucleic Acids Res.* **48**, D376–D382.

Aragão, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N. P., Ericsson, D. J., Gee, C. L., Macedo, S., Mudie, N., Panjikar, S., Price, J. R., Riboldi-Tunncliffe, A., Rostan, R., Williamson, R. & Caradoc-Davies, T. T. (2018). *J. Synchrotron Rad.* **25**, 885–891.

Bennett, E. M., Li, C., Allan, P. W., Parker, W. B. & Ealick, S. E. (2003). *J. Biol. Chem.* **278**, 47110–47118.

Bzowska, A., Kulikowska, E. & Shugar, D. (2000). *Pharmacol. Ther.* **88**, 349–425.

Ducati, R. G., Santos, D. S. & Basso, L. A. (2009). *Arch. Biochem. Biophys.* **486**, 155–164.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.

Fox, N. K., Brenner, S. E. & Chandonia, J.-M. (2014). *Nucleic Acids Res.* **42**, D304–D309.

Gasteiger, E., Hoogland, C., Duvaud, S., Gattiker, A., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.

Gorrec, F. (2009). *J. Appl. Cryst.* **42**, 1035–1042.

Grenha, R., Levnikov, V. M., Fogg, M. J., Blagova, E. V., Brannigan, J. A., Wilkinson, A. J. & Wilson, K. S. (2005). *Acta Cryst.* **F61**, 459–462.

Hamamoto, T., Okuyama, K., Noguchi, T. & Midorikawa, Y. (1997). *Biosci. Biotechnol. Biochem.* **61**, 272–275.

Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.

Karplus, P. A. & Diederichs, K. (2015). *Curr. Opin. Struct. Biol.* **34**, 60–68.

Koellner, G., Bzowska, A., Wielgus-Kutrowska, B., Luić, M., Steiner, T., Saenger, W. & Stepniński, J. (2002). *J. Mol. Biol.* **315**, 351–371.

Krissinel, E. & Henrick, K. (2004a). *Acta Cryst.* **D60**, 2256–2268.

Krissinel, E. & Henrick, K. (2005). *Computational Life Sciences*, edited by M. R. Berthold, R. C. Glen, K. Diederichs, O. Kohlbacher & I. Fischer, pp. 67–78. Berlin, Heidelberg: Springer.

Krissinel, E. B. & Henrick, K. (2004b). *Softw. Pract. Exper.* **34**, 591–607.

Krissinel, E. B., Winn, M. D., Ballard, C. C., Ashton, A. W., Patel, P., Potterton, E. A., McNicholas, S. J., Cowtan, K. D. & Emsley, P. (2004). *Acta Cryst.* **D60**, 2250–2255.

Madrid, D. C., Ting, L.-M., Waller, K. L., Schramm, V. L. & Kim, K. (2008). *J. Biol. Chem.* **283**, 35899–35907.

Mao, C., Cook, W. J., Zhou, M., Koszalka, G. W., Krenitsky, T. A. & Ealick, S. E. (1997). *Structure*, **5**, 1373–1383.

Martins, G. G., de Jesus Holanda, R., Alfonso, J., Gómez Garay, A. F., dos Santos, A. P., de Lima, A. M., Francisco, A. F., Garcia Teles, C. B., Zanchi, F. B. & Soares, A. M. (2019). *Biochimie*, **162**, 97–106.

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

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.

Nannemann, D. P., Kaufmann, K. W., Meiler, J. & Bachmann, B. O. (2010). *Protein Eng. Des. Sel.* **23**, 607–616.

Narczyk, M., Bertoša, B., Papa, L., Vuković, V., Lešćić Ašler, I., Wielgus-Kutrowska, B., Bzowska, A., Luić, M. & Štefanić, Z. (2018). *FEBS J.* **285**, 1305–1325.

Narczyk, M., Mioduszewski, Ł., Oksiejuk, A., Winiewska-Szajewska, M., Wielgus-Kutrowska, B., Gojdz, A., Cieśla, J. & Bzowska, A. (2021). *Sci. Rep.* **11**, 11144.

- Pant, P., Pathak, A. & Jayaram, B. (2021). *J. Phys. Chem. B*, **125**, 2856–2862.
- Park, J., Throop, A. L. & LaBaer, J. (2015). *Curr. Protoc. Mol. Biol.* **110**, 3.
- Reverter, D., Fernández-Catalán, C., Baumgartner, R., Pfänder, R., Huber, R., Bode, W., Vendrell, J., Holak, T. A. & Avilés, F. X. (2000). *Nat. Struct. Biol.* **7**, 322–328.
- Rinaldo-Matthis, A., Wing, C., Ghanem, M., Deng, H., Wu, P., Gupta, A., Tyler, P. C., Evans, G. B., Furneaux, R. H., Almo, S. C., Wang, C. C. & Schramm, V. L. (2007). *Biochemistry*, **46**, 659–668.
- Štefanić, Z., Mikleušević, G., Luić, M., Bzowska, A. & Lešćić Ašler, I. (2017). *Int. J. Biol. Macromol.* **101**, 518–526.
- Štefanić, Z., Narczyk, M., Mikleušević, G., Kazazić, S., Bzowska, A. & Luić, M. (2018). *Sci. Rep.* **8**, 1–13.
- Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst. D* **67**, 235–242.
- Xie, X., Xia, J., He, K., Lu, L., Xu, Q. & Chen, N. (2011). *Biotechnol. Lett.* **33**, 1107–1112.
- Zhu, S., Song, D., Gong, C., Tang, P., Li, X., Wang, J. & Zheng, G. (2013). *Appl. Microbiol. Biotechnol.* **97**, 6769–6778.