

## Supporting Information for

# Nature counts to three: Universal Mg-pinch motif polarizes the cleaved bond in NTP-processing enzymes

Dénes Berta<sup>a,b†</sup>, Balint Dudas<sup>a,c†</sup>, Pablo Jambrina<sup>d</sup>, Pedro J. Buigues<sup>e</sup>, Reynier Suardiaz<sup>f</sup>, Silvia Gómez-Coca<sup>g</sup>, Bernard R. Brooks<sup>c</sup>, Beáta G. Vértessy<sup>h,i</sup>, Edina Rosta<sup>a,\*</sup>

† equal contribution

a Department of Physics and Astronomy, University College London, London, UK

b Department of Physical Chemistry and Materials Science, Budapest University of Technology and Economics, Budapest, Hungary

c Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA

b Department of Chemical Physics, University of Salamanca, Salamanca, Spain

e Italian Institute of Technology, Genova, Italy

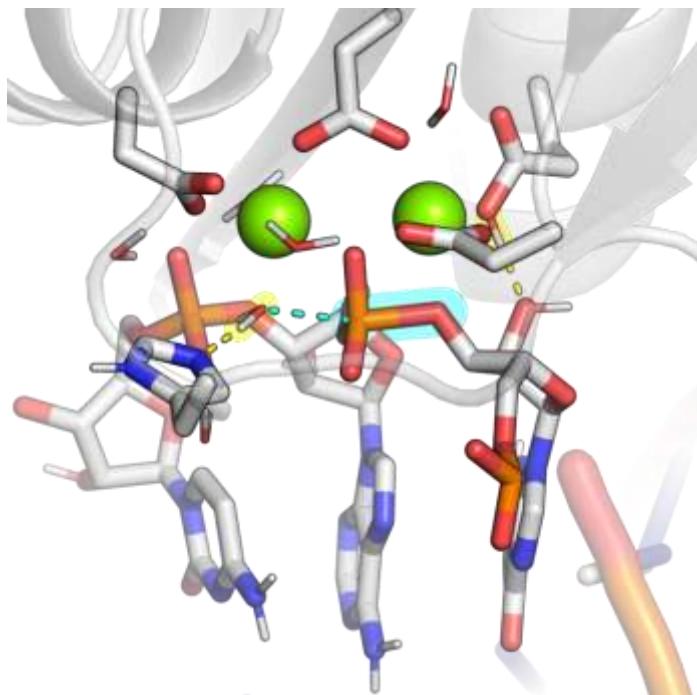
f Department of Physical Chemistry, Complutense University of Madrid, Madrid, Spain

g Department of Inorganic and Organic Chemistry and Institute of Theoretical and Computational Chemistry, University of Barcelona, Barcelona, Spain

h Department of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, Budapest, Hungary

i Genome Metabolism Research Group, Research Centre for Natural Sciences, Hungarian Research Network, Budapest, Hungary

\* [e.rosta@ucl.ac.uk](mailto:e.rosta@ucl.ac.uk)



**Figure S1:** Active site of the HIV-1 ribonuclease H (based on Ref [1]). The cleaved phosphodiester bond is highlighted in cyan, associated proton transfers are highlighted in yellow.

## Methods and data collection

### Analysis of Enzymatic Reactions

The KEGG database was analyzed as of 20<sup>th</sup> of December 2023. There are 6728 distinct EC numbers, of which 6197 carry out 10571 KEGG registered chemical reactions. The reactions were primarily obtained from the KEGG reaction database, substrates and products of the enzymatic reactions were parsed for ECs where it was not assigned to KEGG reaction. Each reaction corresponds to one or more EC numbers. The reactions were categorized as follows:

- If any of the substrates contain phosphates. 2012 phosphate compounds (appendix A1) were identified from the KEGG compound database.
- If the enzymatic activity involves phosphate chemistry. This was determined using KEGG reaction classes (appendix A2). It is noted that reaction class assignments are not exhaustive, so the number of actual phosphate processing reactions and ECs are greater than what we found.
- If one of the substrates was a nucleoside triphosphate (or modified NTP), it was subcategorized into one of these (appendix A3):
  - Phosphatase, if it produced inorganic phosphate and/or NDP
  - Pyrophosphatase, if it produced inorganic pyrophosphate or NMP
  - Triphosphatase, if it produced a nucleoside or inorganic triphosphate.

Notably, there are 36 reactions that could not be categorized against these criteria, these were assigned manually (appendix A4).

The identified NTP-processing ECs were then used to retrieve PDB structures and protein sequences assigned to them. The collected protein chain sequences were then matched against the SUPERFAMILY 2.0 [25, 26] database containing a collection of hidden Markov models that represent structural protein domains at the SCOPe SF level to assign the corresponding SCOPe SFs [34, 35]. Where SUPERFAMILY 2.0

failed to retrieve a match, the InterPro SF (or if absent the InterPro Family) was used instead. Finally, redundant SFs (InterPro entries that resembled any of the identified SCOPe SFs after DALI alignment) were merged. In total, we identified 90 NTP-processing SFs.

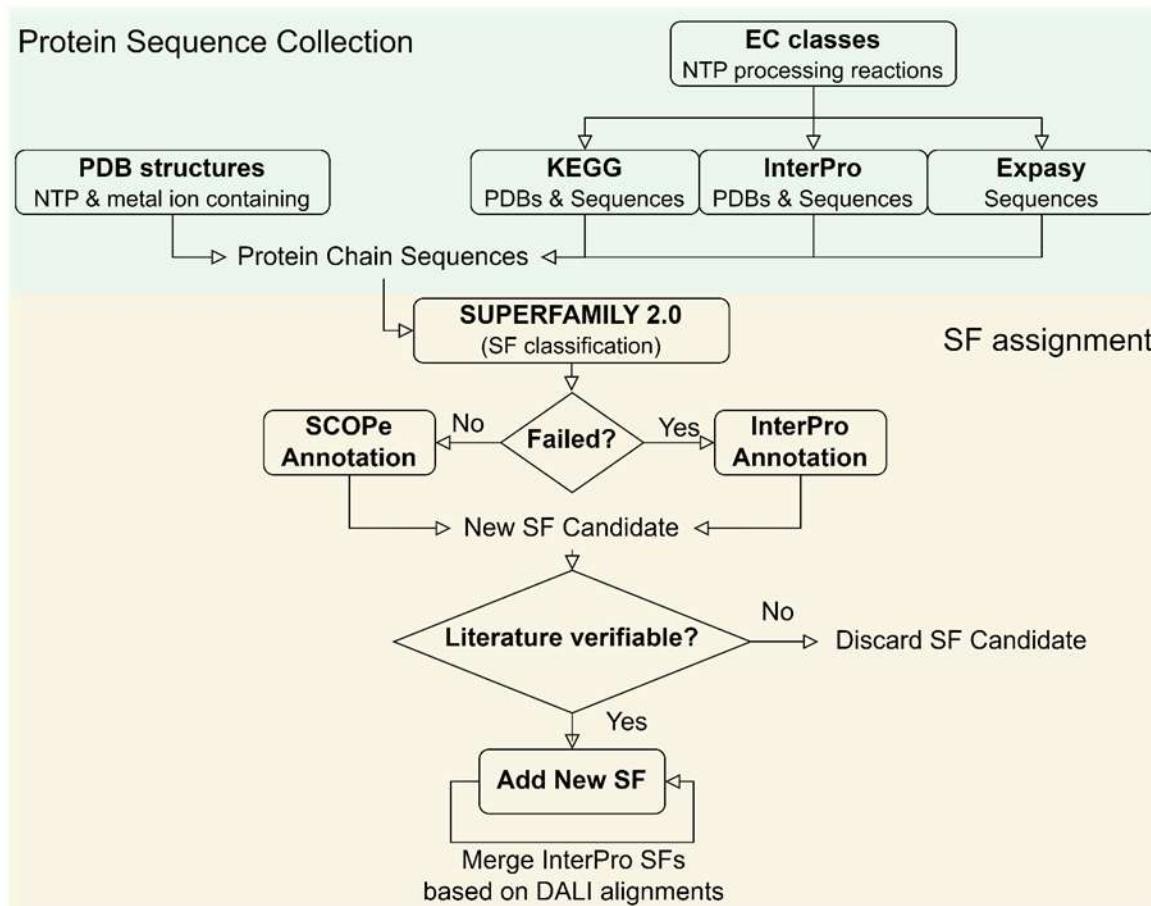
#### PDB accession and analysis of the Mg<sup>2+</sup> coordination

We performed the Mg<sup>2+</sup> coordination analysis on experimental PDB structures [24] that meet the following four criteria: i) have a resolution of at least 2.7 Å for reliable metal-ion-coordination, ii) contain a residue with three phosphorus atoms, iii) feature at least one relevant divalent metal ion (Mg<sup>2+</sup>, Ca<sup>2+</sup>, or Mn<sup>2+</sup>; excluding Zn<sup>2+</sup> due to its different biochemical function) within 5 Å of the three phosphorus atom-containing residue, and iv) possess a coordination number around the metal ion of at least four (defined within a sphere with a radius of 3 Å centered around the cation), ensuring adequate structural quality of the active site.

From the PDB, we downloaded 8641 structures that satisfied criteria i) and ii). Out of these, 4553 structures contained at least one divalent metal ion as per criterion iii). After excluding structures that did not meet any of the four criteria, or displayed ambiguity in their active site (with an occupancy lower than 1.00 for either the metal ion or the coordinating ions), a total of 3118 PDB structures fulfilled all the specified structural and geometric requirements and were used in the subsequent analysis.

Using the SUPERFAMILY 2.0 webserver [25, 26], we classified active sites that bind the three-phosphorus-containing residue and the metal ion(s) into SFs based on their amino acid sequence. Residues were filtered to contain at least three phosphates in a chain, based on geometric criteria. The active site was assigned to the protein chain that had the most atoms around the NTP (or analog) residue.

We subsequently performed a systematic analysis of the active sites to determine: i) the number of coordinating metal ions present in the active sites (one, two, or more) and ii) the coordination configuration of the individual metal ions, assessing whether they are coordinated by nonbridging oxygens of the  $\alpha$ , the  $\beta$ , and/or the  $\gamma$  phosphate groups.



**Figure S2:** Flowchart of the data collection process to identify all NTP processing SFs that have currently available structural data.

#### Consensus structures representing each superfamily

A representative structure in each SF was selected that represents the consensus metal ion coordination observed within the specific superfamily after aligning the NTP-containing active sites of the structures. While the selection of these representative structures is not fully objective, we aimed to demonstrate the consensus coordination pattern of the given SF, therefore manually selected structures where all metal ions are present (if there are more than one), and they have a full octahedral coordination including crystallographic waters, if available. Based on the leaving group of the corresponding catalytic reactions, we categorized the superfamilies into phosphatases ( $P_\gamma$  monophosphate leaving group), pyrophosphatases ( $P_\beta P_\gamma$  pyrophosphate leaving group), and triphosphatases ( $P_\alpha P_\beta P_\gamma$  triphosphate leaving group).

Finally, we examined and compared the positions of metal ions relative to the triphosphate chain in all representative structures from both phosphatase and pyrophosphatase SFs. This systematic analysis allowed us to comprehensively describe all the currently identified catalytically active Mg-coordination modes in NTP processing enzymes.

## QM/MM calculations

Structures for the reaction paths and the definition of the QM regions were taken from Refs [67, 69, 70]. The QM region with electrostatic embedding were simulated with the B3LYP hybrid functional at the 6-31+G(d) split-valence basis set, while the MM region was described by the charmm27 and charmm36m protein, and the TIP3 water force fields, as implemented in Qchem 5.2 [71] and CHARMM 47b1 [72].

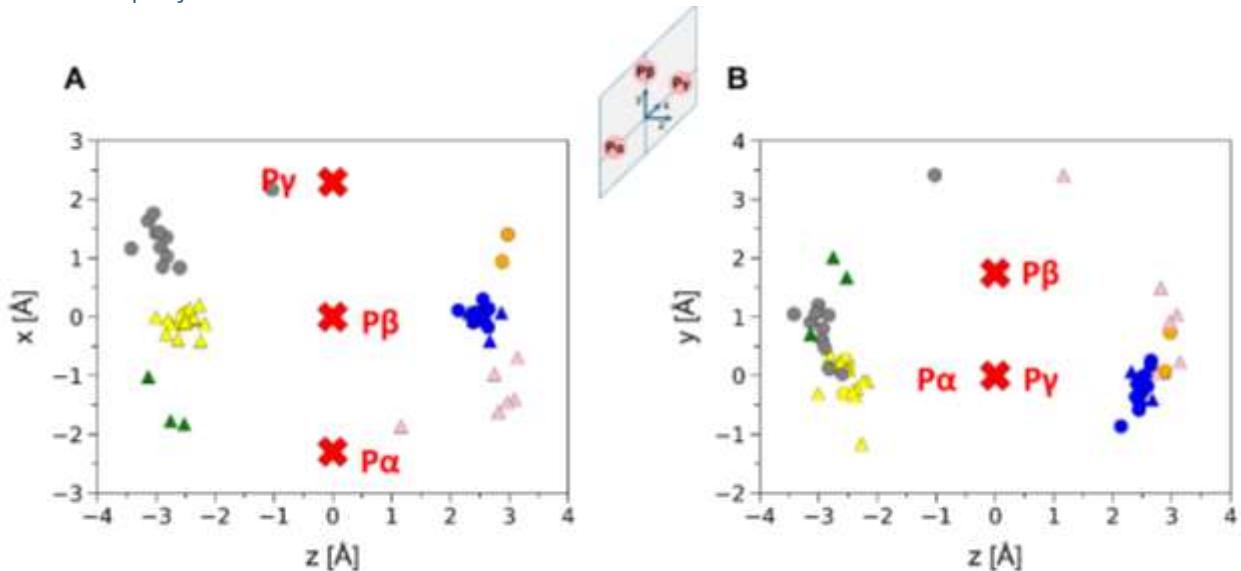
## QM calculations

Minimal models were derived from the QM/MM structures, without any MM representation. Electronic properties and population analysis were carried out at the B3LYP/6-31+G(d) level of theory with the software packages: QChem 5.2, Gaussian 09 Rev. E and NBO 3.0 [71, 73, 74].

For the point charge screening calculations, reactant and transition state structures were aligned. A cartesian grid with 0.35 Å spacing was created for points between 2.0 and 2.1 Å distance from the phosphate oxygens.

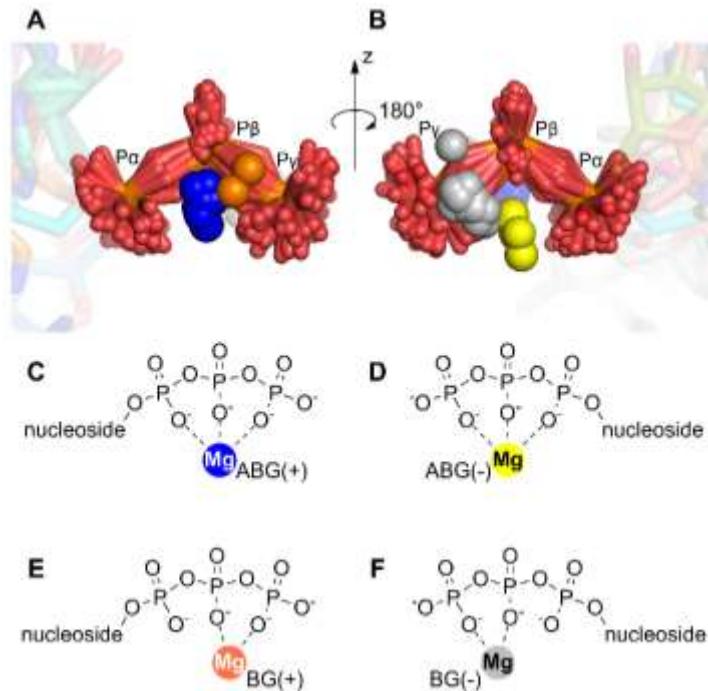
Electron densities from the DFT SCF calculations were processed with Gaussian 09 and interpreted at a 200<sup>3</sup> grid. Density difference isosurfaces are depicted at +/- 0.001 values.

## Alternative projections of the ion clusters



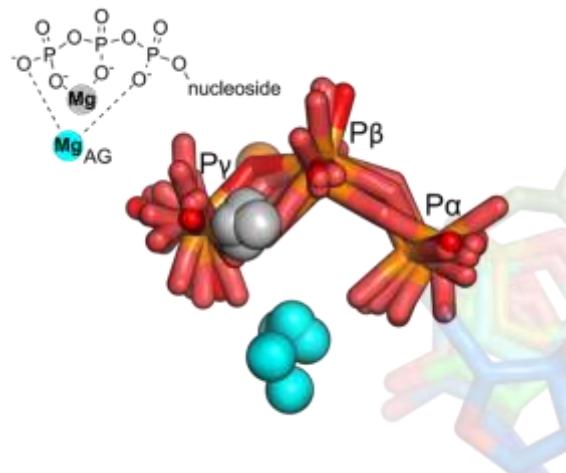
**Figure S3:** Positions of the Mg-pinch-forming metal ions shown with respect to the phosphates in different views. Phosphatase ion positions are denoted by circles, and pyrophosphatase ion positions are represented by triangles, with colors matching those in main figures.

## Phosphatases: Py leaving group



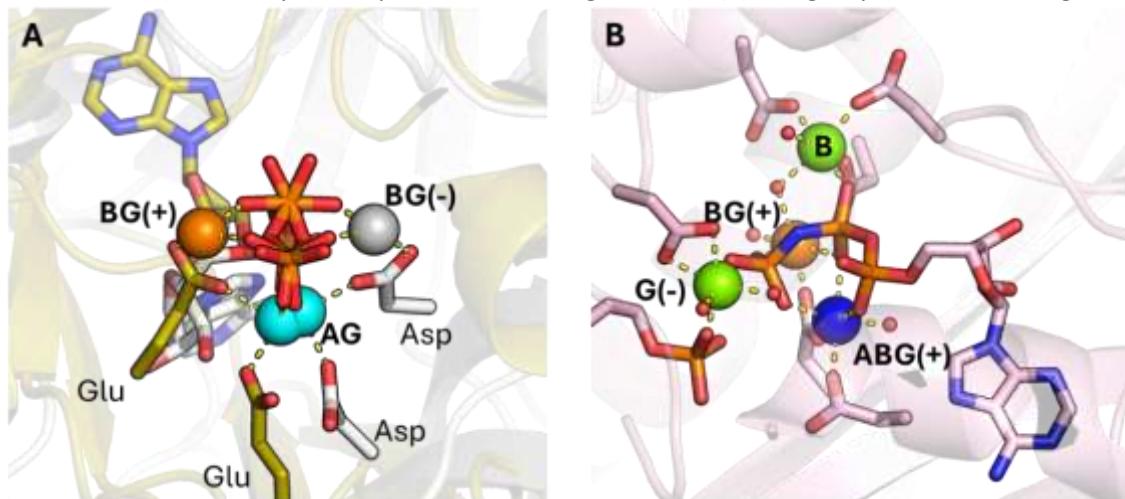
**Figure S4.** Clusters of the catalytic divalent metal ions observed in phosphatases. The Mg-pinch forming metal ions as observed in the representative structures of the phosphatases are shown on **A**) the “A” side of the plane defined by the three phosphorus atoms (left side from a front-view),  $\alpha\beta\gamma$ -coordinated ions are shown in blue and  $\beta\gamma$ -coordinated ions in orange; and on **B**) the “B” side of the same plane (right side from a front-view),  $\alpha\beta\gamma$ -coordinated ions are shown in yellow and  $\beta\gamma$ -coordinated ions in gray. The NTPs of the representative structures are superimposed using their three phosphorus atoms and are shown in sticks.

Besides the pinching  $Mg^{2+}$ , we also analyzed the positions of any additional metal ions if present at the active site and so identified another well-defined cluster formed by a second metal ion in 5 superfamilies (Figure S5). These second metal ions are at the AG site and are located close to the plane defined by the three phosphorus atoms. Interestingly, the cluster of second metal ions at the AG site is only observed as an addition to BG coordinated pinching ions. Additional ions are found in further superfamilies in pinching position (BG, 2 SFs), in AB coordination (2 SFs), and only in contact with the leaving phosphate (G, 2 SFs, see Table 1).



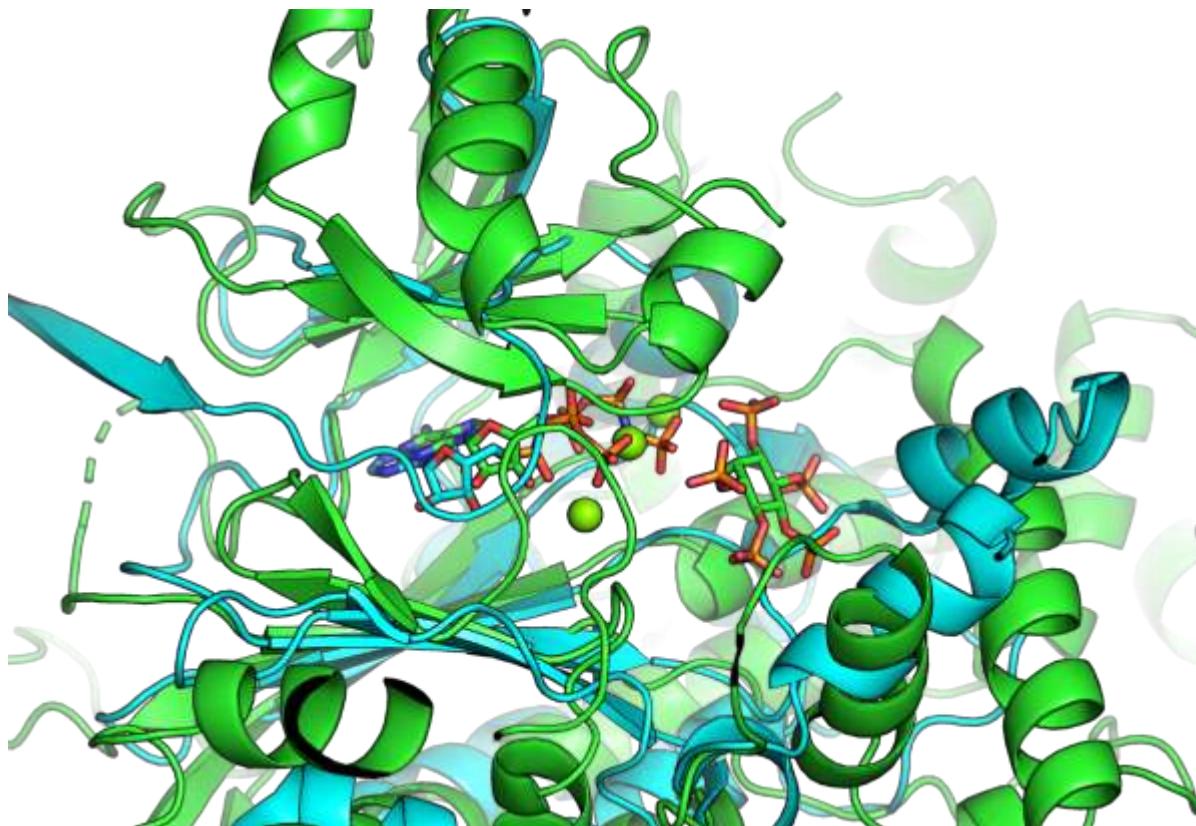
**Figure S5.** The cluster of divalent second metal ions (in cyan) found in 5 phosphatase superfamilies in addition to their Mg-pinch forming ion (shown in gray or orange). The ions of the cyan cluster are  $\alpha\gamma$ -coordinated (AG site) and are located closer to the plane than the Mg-pinch forming ions. The following superfamilies fall into this group: Protein kinase-like, Glutathione synthetase ATP-binding domain-like, SAICAR synthase-like, and Diacylglycerol kinase (DgkA)-like, that all have their Mg-pinch forming ions at the BG(–) site (gray), as well as Glutamine synthetase/guanido kinase, which symmetrically has its Mg-pinch forming ion at the BG(+) site (orange).

We identified metal ions with  $\beta\gamma$  and  $\alpha\beta\gamma$  coordination on both sides of the plane defined by the three phosphorus atoms in phosphatases. Notably, most of the metal ions found on the (+) side are simultaneously coordinated by all three phosphate groups (ABG(+), Figure 2 and Table 1). Among phosphatases having the Mg pinch-forming ion at the ABG(+) site, eight superfamilies feature a single metal ion, while three have additional metal ions at their active site. Among the phosphatases with metal ions on the (+) side, only two superfamilies belong to the BG(+) site group (shown in orange in Figure 2).



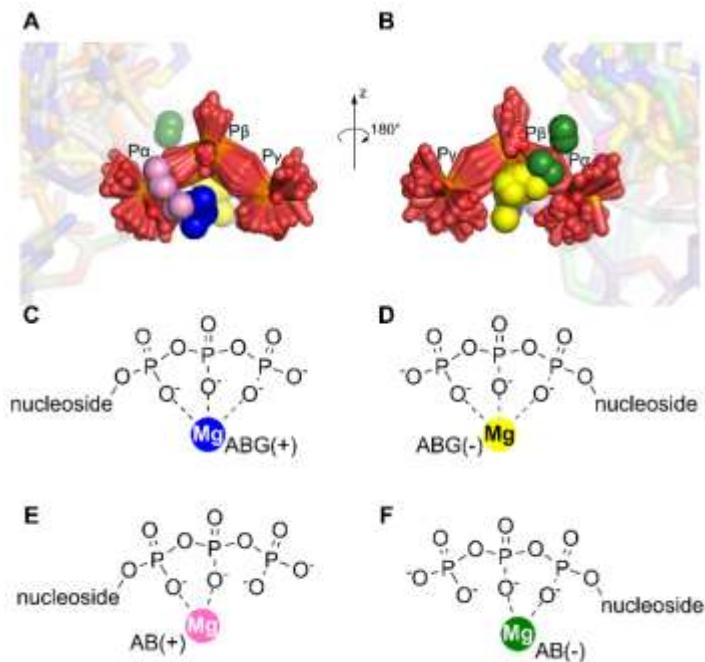
**Figure S6.** **A** Cation coordination of the Glutamine synthetase/guanido kinase (olive, orange and cyan spheres, 7CQQ) and the Glutathione synthetase ATP-binding domain-like (white, gray and cyan spheres, 5DGH) SFs, exhibiting a mirrored two ion arrangement. **B** ATP and substrate binding in PurM N-terminal domain-like SF (pink, 5DD7), with 4 active site metal ions.

Conversely, the *Glutamine synthetase/guanido kinase* superfamily, with very few available high-resolution structures, presents a multi-ion active site. Notably, this superfamily stands alone in having a second metal ion at the AG site (shown in cyan in Figure S5) while having its Mg-pinch forming ion on the (+) side, and a third ion near the nucleophile. Interestingly, the coordinative arrangement bears a striking resemblance to the mirror image of the *Glutathione synthetase ATP-binding domain-like* superfamily (Figure S6A).



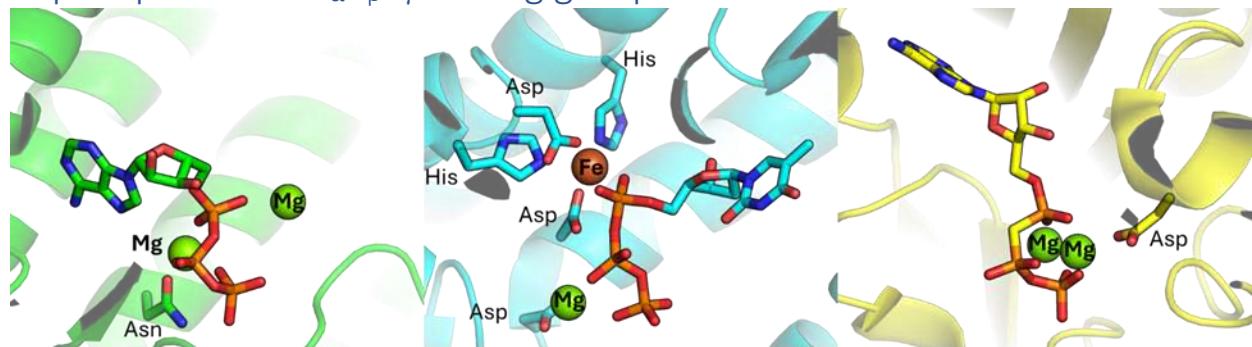
**Figure S7:** Structures of the Inositol-pentaphosphate 2-kinase (green, 2XAN) and the SAICAR synthase-like SFs (cyan, 2AQX).

## Pyrophosphatases: P $\beta$ P $\gamma$ leaving group



**Figure S8.** The four clusters of the Mg-pinch-forming metal ions as observed in the representative structures of pyrophosphatases. A) The (+) side,  $\alpha\beta\gamma$ -coordinated ions are shown in blue (ABG(+)) and  $\alpha\delta$ -coordinated ions in pink (AB(+)). B) The (-) side,  $\alpha\beta\gamma$ -coordinated ions are shown in yellow (ABG(-)) and  $\alpha\delta$ -coordinated ions in green AB(-). The NTPs of the representative structures of the pyrophosphatase superfamilies are superimposed using their three phosphorus atoms and are shown in sticks.

## Triphosphatases: $P_\alpha P_\beta P_\gamma$ leaving group



**Figure S9:** Versatile ion coordination found in triphosphatases: Cobalamin adenosyltransferase-like (green, 6D5K); HD-domain/PDEase-like (cyan, 6TXE); S-adenosylmethionine synthetase (yellow, 6VD0). Ion coloring of coordination is not used in these exceptional superfamilies to show elements instead.

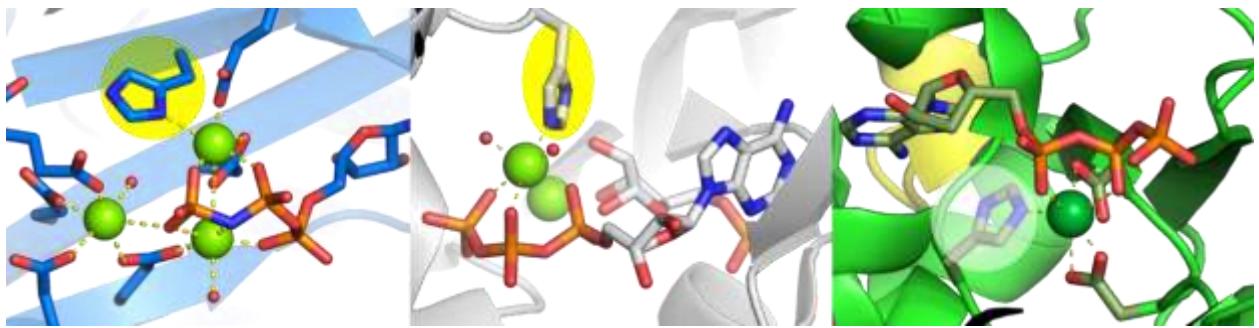
## Unusual metal ions and metal ion coordinating residues

Additionally, we also analyzed the coordinating protein residues around the catalytic metal ions among the 71 identified representative structures (Table S1). We prioritized structures with complete first shell coordination wherever possible, especially when selecting representative structures. SFs not relying on  $Mg^{2+}$  but using transition catalytic metal ions like zinc, iron, or cobalt can often possess exceptional coordinating residues. Coordinating histidine residues were identified in the above discussed *Metallo-dependent phosphatases*, *Alkaline phosphatase-like*, and *Ca-dependent phosphotriesterase* SFs. In the rare case of the *tRNA-splicing ligase RtcB-like superfamily*, a cysteine is found to coordinate the two  $Mn^{2+}$ . Interestingly, a histidine residue even participates in the ion-coordination of a few  $Mg^{2+}$ -dependent SFs: *Glutamine synthetase/guanido kinase* among phosphatases and *PRTase-like*, *Adenylylcyclase toxin (the edema factor)* among pyrophosphatases (Figure S10).

**Table S1.** Protein residues involved in the coordination of the Mg-pinch forming ion.

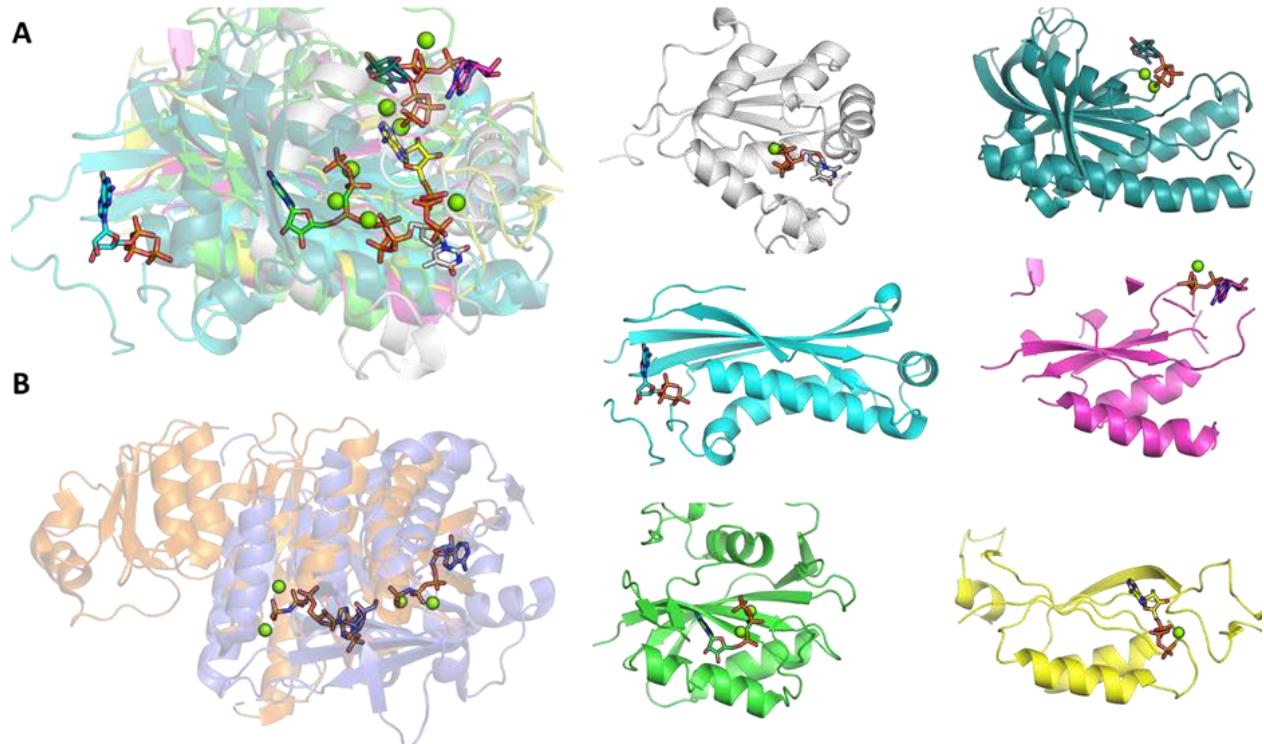
	Asp	Glu	Ser	Thr	Asn	backbone CO	His
Pi	16	12	4	4	4	1	3
PPi	40	8	2	0	0	5	6
total	56	20	6	4	4	6	9
ratio	53%	19%	6%	4%	4%	6%	8%

Carboxylate residues (aspartate and glutamate) are the most prevalent protein residues both for phosphatases and pyrophosphatases. Other oxygen donors are also involved in the coordination of  $Mg^{2+}$ : the alcohol sidechains of serine and threonine are identified 4-4 times in phosphatases, while only 2 serine residues are found to coordinate ions in pyrophosphatase SFs. There is occasional carbonyl donation, 4 asparagine among phosphatase SFs and 6 times the backbone oxygen, mostly in pyrophosphatase structures. Finally, water molecules commonly complete the octahedral  $Mg^{2+}$  coordination (63 water molecules in 33 phosphatases, 59 water molecules in 38 pyrophosphatase structures).

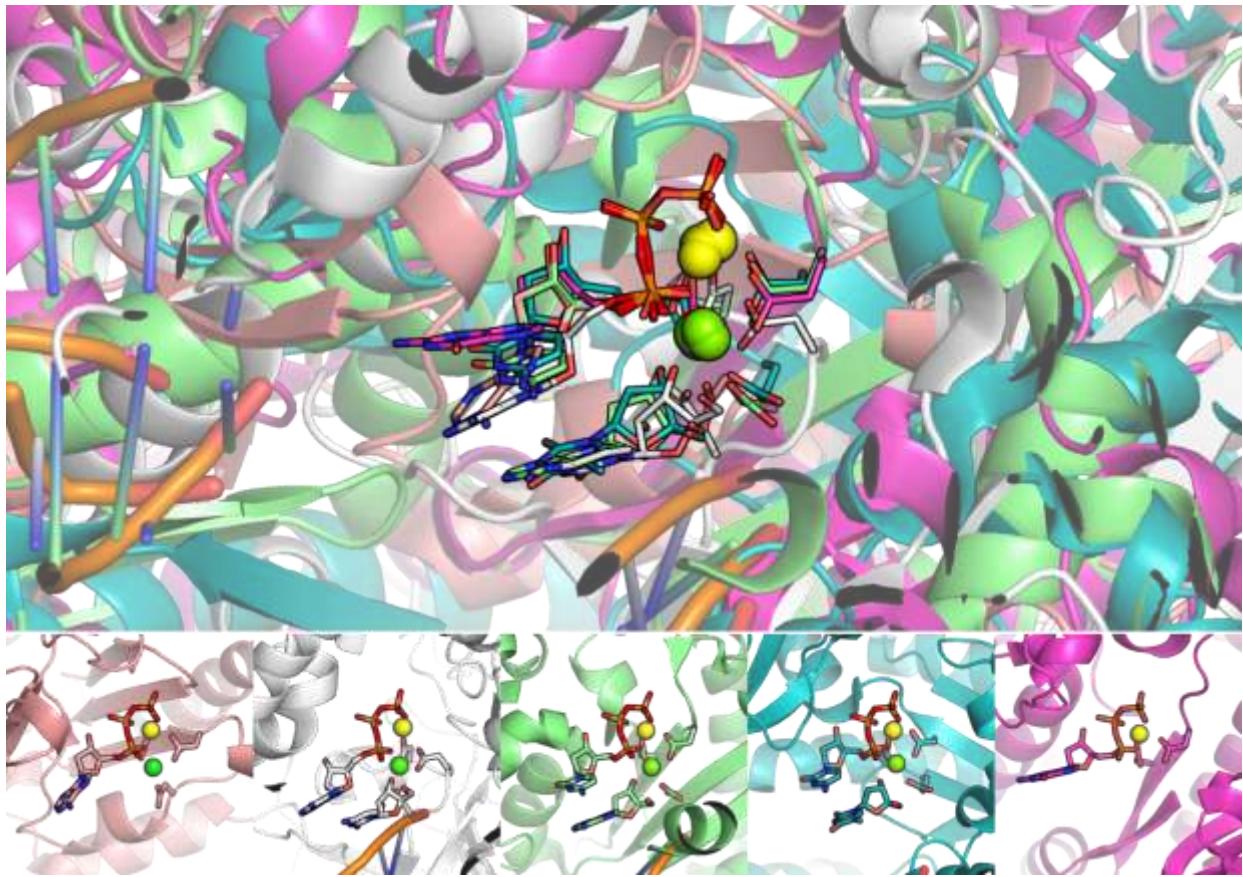


**Figure S10.**  $Mg^{2+}$  coordinating histidine residues (highlighted) in superfamilies: Glutamine synthetase/guanido kinase (blue, 7CQQ); PRTase-like (white, 8DBJ), Adenylylcyclase toxin (the edema factor) (green, 1K90).

### Analysis on the protein fold level

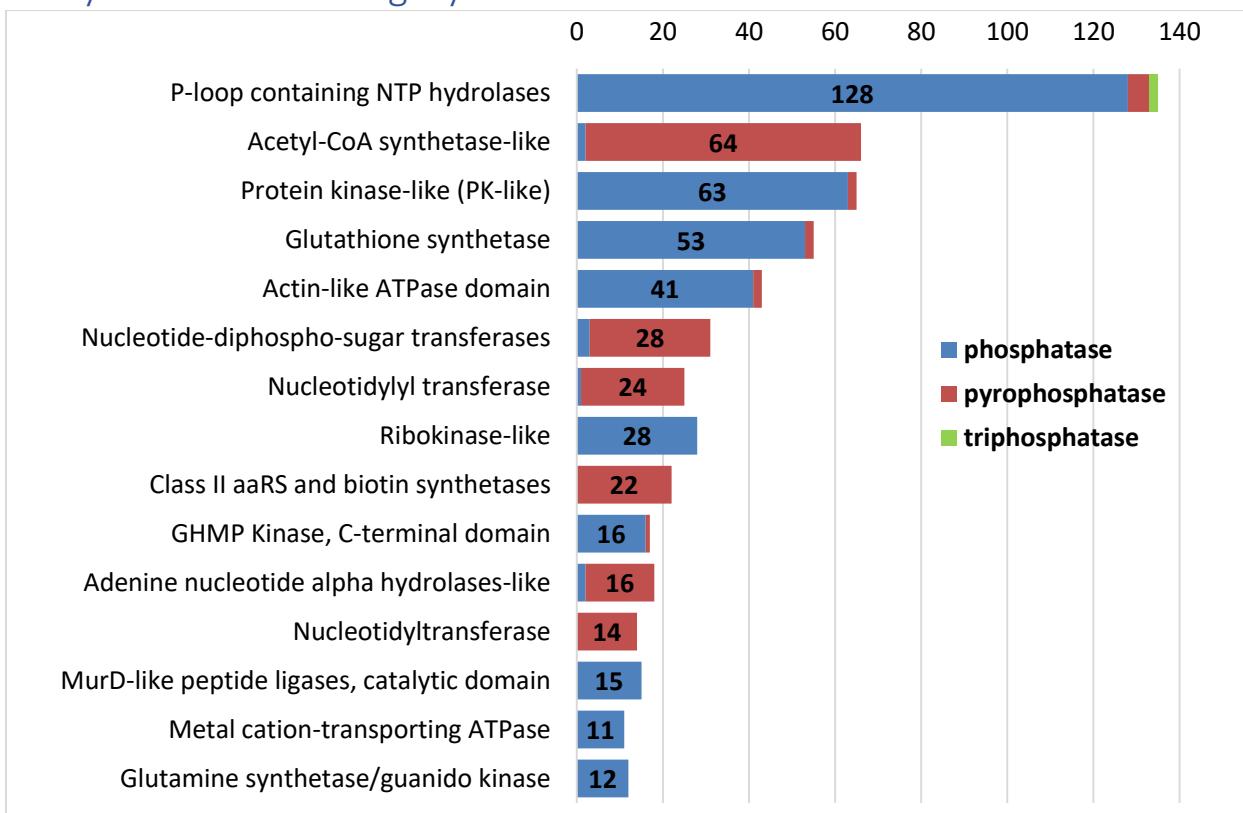


**Figure S11:** Different NTP-binding sites of SFs from the same fold. **A** Six Ferredoxin-like fold SFs: Nucleoside diphosphate kinase, NDK (white, 1F3F); Nucleotide cyclase (teal, 1WC6); Molybdenum cofactor biosynthesis protein C, MoaC (cyan, 3JQM, no structure with ion); GHMP Kinase, C-terminal domain (magenta, 3V2U); 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, HPPK (green, 4CRJ); GlnB-like (yellow, 2J2C) **B** the two members of the Ribokinase-like fold, Ribokinase-like (blue, 2JG1) and MurD-like peptide ligases, catalytic domain (orange, 6CAU) that are both phosphatases with a  $Mg^{2+}$  at the BG(–) position, yet possessing active sites at different locations.



**Figure S12:** Example of 5 polymerases (Prim-pol domain; Nucleotidyltransferase; DNA/RNA polymerases; Bacterial DNA polymerase III, alpha subunit, NTPase domain; and RNA-dependent RNA polymerase, eukaryotic-type) with different folds. Intriguingly, their coordination geometries are strikingly similar even though the coordinating amino acid side chains reside on structurally different elements.

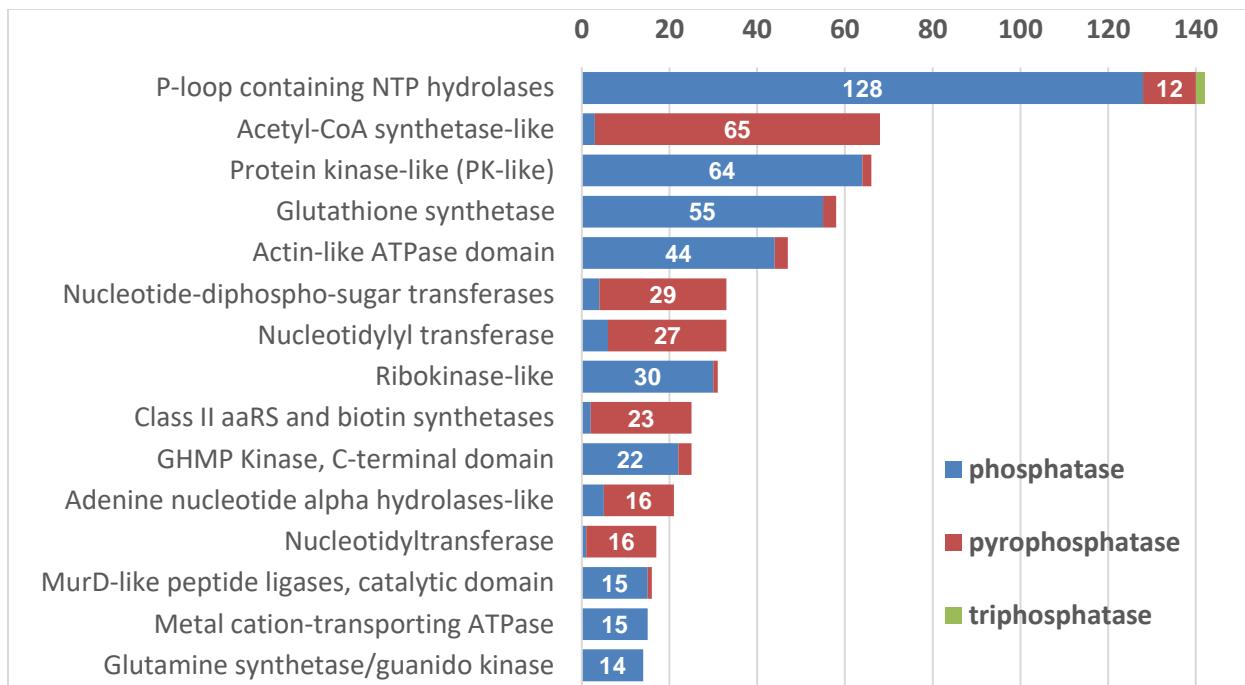
## Analysis of the EC category distributions



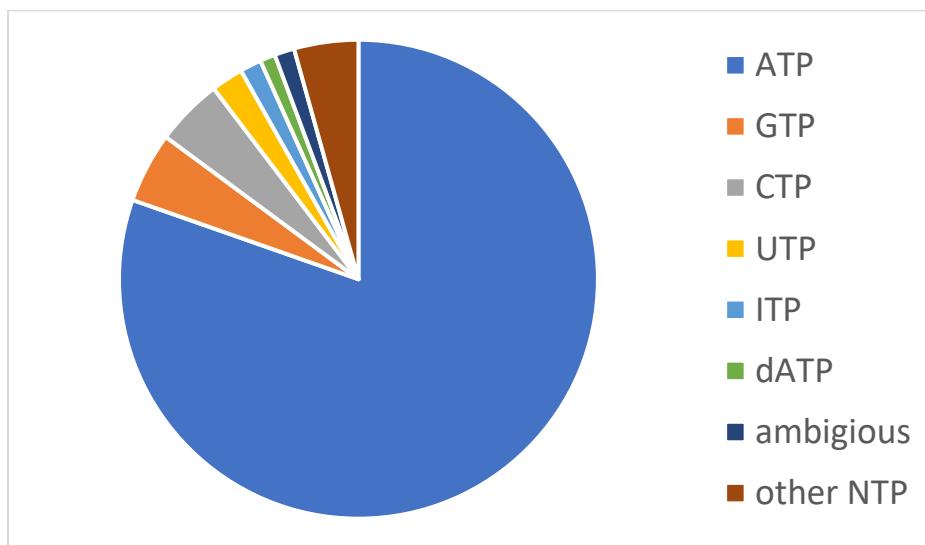
**Figure S13:** Number of EC categories (unambiguously) assigned to structures and sequences of superfamilies, grouped along reaction types: phosphatases (blue), pyrophosphatases (red) and triphosphatases (green). The full lists are provided in Note 1.

To provide an approximate assignment, we used the EC categories associated with the sequences from the sources used in Figure S2. We noticed that the EC assignments coupled reactions may be catalyzed by different domains or even chains, and therefore when multiple domains are present, the reaction is associated with all the domains of a given chain, despite that only one of the domains is responsible for the corresponding EC reaction. We therefore restricted our analysis to sequences that may contain only one selected SF and rank the most prevalent SFs in terms of EC function (Figure S13).

Often, the EC assignments might be extended to the P-loop NTPase domain incorrectly (e.g., 2.7.7.6 polymerase reactivity was assigned because the replication complex contains P-loop NTPase helicase chains [60]), resulting in additional false positive structure-function associations. In this SF, we manually validated all pyrophosphatase reactions and were able to verify that these are all incorrect. In other cases, the series of phosphatase ECs assigned to the *Nucleotidyl transferase* superfamily are separate kinase domains mistakenly associated with the second function of the protein they are part of, for example the N-terminal domain has a kinase function, while the C-terminal is a *Nucleotidyl transferase* in a reported prokaryotic FAD synthetase [61]. Among *CYTH-like phosphatases*, both phosphatase and pyrophosphatase activities are identified in the literature, due to the adenylyl cyclase class IV pyrophosphatases that possess a CYTH domain [62, 63]. Interestingly, the *PK-like* SF is also verifiably associated with adenylyltransferase activity through flipped pseudokinases in the literature [64, 65].



**Figure S14:** Number of EC categories assigned to structures and sequences of superfamilies, without filtering for misclassification based on literature; grouped along reaction types: phosphatases (blue), pyrophosphatases (red) and triphosphatases (green).



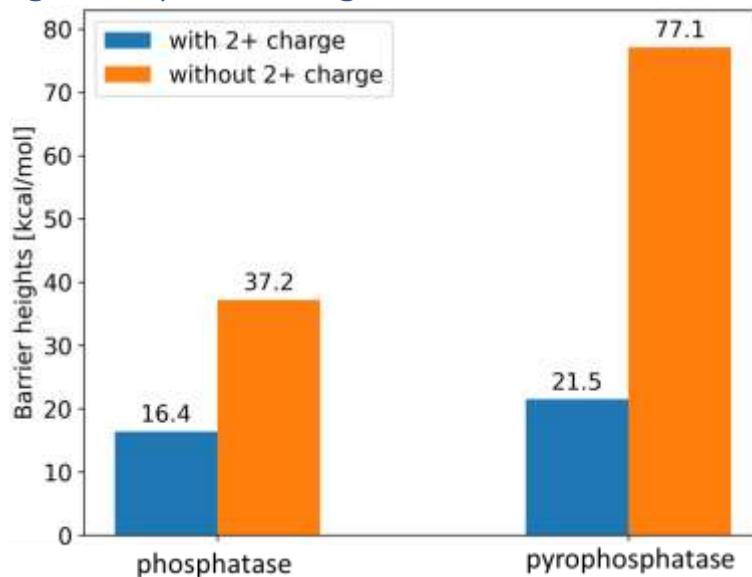
**Figure S15.** Substrate distribution among NTP processing enzymes based on the EC reactions. Note that one EC may be assigned to multiple substrates.

Increased structural diversity (represented by the number of associated SFs), with 5, 4, and 4 SFs, is also observed amongst non-specific kinases (2.7.11.1, 2.7.10.2 and 2.7.13.3, respectively). Some ECs may be associated with both phosphatase and pyrophosphatase SFs, and correspond to general reactivity categories, e.g. 3.6.1.15 Nucleoside-triphosphate phosphatase.

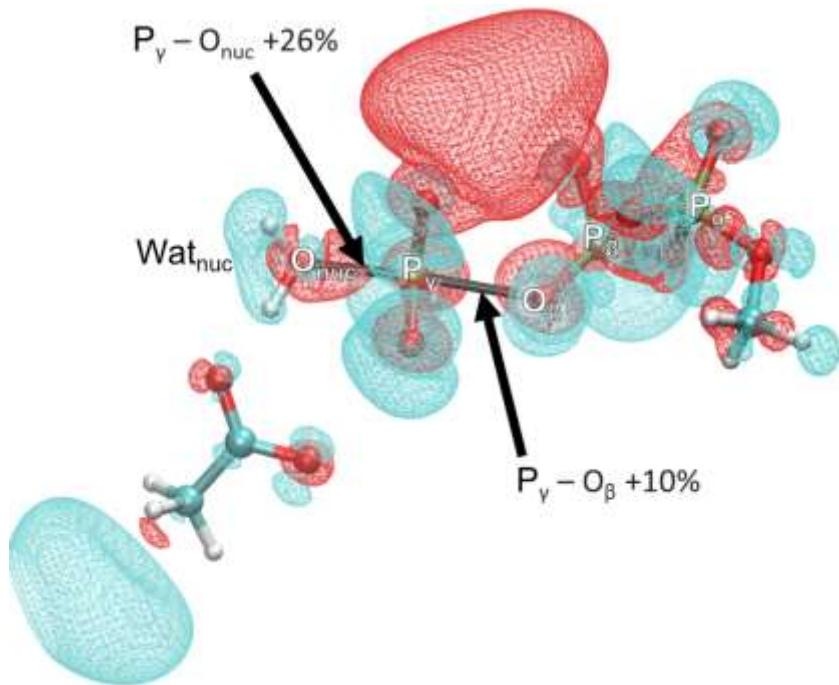
**Table S2.** Enzyme Commission categories with members assigned to 4 or more superfamilies. Groups with polymerase activity are highlighted in purple, kinases are in pink.

EC	name	No. superfamily
2.7.7.19	RNA adenylating enzyme	6
2.7.7.48	RNA-directed RNA polymerase	6
2.7.11.1	non-specific serine/threonine protein kinase	5
2.7.7.7	DNA-directed DNA polymerase	4
2.7.10.2	non-specific protein-tyrosine kinase	4
2.7.13.3	histidine kinase	4

The effect of  $Mg^{2+}$  analyzed through QM/MM and QM calculations



**Figure S16:** Activation energies with and without a 2+ pointcharge representing the  $Mg^{2+}$  in minimal model systems.



**Figure S17:** Isosurface of the electron density change upon the introduction of a 2+ point charge at the  $Mg^{2+}$  position for the phosphatase model system. Increased density is depicted in red, decreased in cyan. The changes in the forming and breaking  $P$ - $O$  bonds (black lines) are highlighted by arrows.

### Note 1: ECs associated to superfamily members

P-loop, containing, nucleoside, triphosphate, hydrolases: 3.6.1.8, 7.1.2.2, 2.5.1.17, 2.7.7.6, 2.7.12.2, 2.7.1.164, 2.7.1.145, 2.7.4.25, 2.7.4.13, 6.6.1.1, 2.7.11.17, 2.7.4.8, 2.7.1.76, 2.7.4.2, 2.7.1.105, 2.7.4.9, 6.1.1.11, 3.6.5.5, 2.7.10.2, 6.3.4.3, 6.3.4.4, 2.7.1.71, 2.7.1.12, 3.6.5.2, 2.7.1.48, 2.7.4.3, 2.7.4.10, 2.7.1.19, 2.7.4.14, 2.7.1.78, 2.7.1.113, 6.3.3.3, 2.7.1.130, 3.6.5.1, 2.7.1.21, 3.1.5.1, 2.7.1.25, 2.7.1.33, 2.7.1.74, 2.7.4.1, 2.7.1.24, 2.7.7.4, 3.6.5.4, 2.7.11.13, 6.3.4.2, 6.5.1.3, 2.7.1.67, 6.3.5.12, 6.3.3.7, 6.3.4.25, 1.3.7.15, 6.3.5.11, 2.7.1.176, 2.7.4.23, 2.7.2.16, 2.7.1.156, 6.5.1.9, 2.7.4.34, 2.7.1.224, 6.3.5.9, 2.7.1.186, 7.3.2.6, 6.6.1.2, 6.3.5.10, 2.7.7.83, 2.7.7.76, 2.7.4.4, 3.6.1.15, 2.7.11.1, 3.6.5.3, 2.7.6.2, 7.2.2.1, 3.6.4.6, 7.6.2.5, 7.4.2.8, 7.6.2.4, 5.6.1.1, 7.2.2.16, 7.5.2.4, 7.5.2.3, 7.6.2.2, 5.6.1.8, 5.6.1.6, 7.5.2.6, 7.4.2.6, 3.6.4.13, 2.7.13.3, 7.6.2.10, 2.7.1.23, 7.3.2.2, 7.6.2.11, 7.6.2.6, 2.7.7.49, 7.4.2.4, 7.6.2.14, 7.3.2.3, 4.6.1.1, 7.6.2.12, 7.3.2.5, 7.2.2.7, 7.2.2.20, 7.5.2.13, 7.5.2.7, 7.2.2.11, 7.6.2.9, 7.5.2.9, 2.3.1.193, 7.4.2.12, 7.6.2.8, 7.6.2.16, 7.4.2.10, 7.2.2.18, 7.4.2.11, 7.2.2.4, 7.6.2.7, 7.5.2.10, 1.18.6.1, 7.3.2.1, 7.4.2.13, 7.5.2.11, 7.2.2.17, 7.3.2.7, 7.5.2.12, 2.7.7.50, 7.4.2.9, 7.6.2.13, 7.5.2.5, 2.7.7.19, 7.4.2.7, 2.7.2.1, 7.6.2.3, 7.4.2.14, 7.5.2.1, 7.3.2.4, 7.2.2.2, 5.6.1.4, 7.6.2.15, 7.4.2.5, 7.4.2.1, 7.5.2.8, 7.2.2.5

Acetyl-CoA, synthetase-like: 1.13.12.7, 6.2.1.8, 6.2.1.12, 6.2.1.1, 6.2.1.57, 6.2.1.20, 6.2.1.2, 6.1.1.13, 6.2.1.33, 6.2.1.26, 5.1.1.11, 6.2.1.32, 6.2.1.30, 6.2.1.25, 6.2.1.3, 6.2.1.53, 6.2.1.51, 6.2.1.72, 6.2.1.70, 6.2.1.50, 6.2.1.35, 6.2.1.46, 6.3.2.40, 2.7.7.97, 6.3.2.50, 6.2.1.66, 6.2.1.27, 6.1.3.1, 6.2.1.48, 6.3.1.15, 6.2.1.44, 6.2.1.7, 6.2.1.47, 6.7.1.1, 6.2.1.62, 6.2.1.15, 6.2.1.40, 6.2.1.17, 6.2.1.65, 6.2.1.61, 6.2.1.68, 6.2.1.63, 6.2.1.24, 6.3.2.20, 6.2.1.76, 6.3.2.14, 6.2.1.37, 6.2.1.16, 6.2.1.71, 6.2.1.60, 6.2.1.43, 6.2.1.42, 6.2.1.41, 6.2.1.13, 6.2.1.75, 6.2.1.69, 1.2.1.95, 6.3.2.46, 6.2.1.31, 6.2.1.23, 6.2.1.19, 1.3.7.8, 6.2.1.67, 6.3.2.26, 6.2.1.59, 6.2.1.49, 6.2.1.54, 6.3.2.52

Protein, kinase-like, (PK-like): 2.7.1.153, 6.5.1.1, 2.7.10.1, 2.7.11.11, 2.7.12.2, 2.7.11.24, 2.7.11.22, 2.7.11.21, 2.7.1.230, 2.7.11.14, 2.7.1.190, 2.7.11.17, 2.7.1.183, 2.7.11.7, 2.7.1.39, 2.7.11.15, 2.7.10.2, 2.7.1.82, 2.7.11.30, 2.7.1.95, 2.7.11.10, 2.7.11.31, 2.7.1.23, 2.7.1.154, 2.7.1.3, 2.7.1.163, 2.7.1.32, 2.7.1.17, 2.7.11.23, 2.7.1.137, 2.7.1.162, 2.7.1.119, 2.7.1.100, 2.7.11.25, 2.7.11.12, 2.7.11.19, 2.7.1.1, 2.7.12.1, 2.7.1.22, 2.7.11.26, 2.7.11.13, 2.7.1.67, 2.7.1.175, 2.7.1.89, 2.7.1.72, 2.7.1.87, 2.7.1.166, 2.7.1.8, 2.7.1.181, 2.7.1.222, 2.7.1.103, 2.7.1.221, 2.7.1.81, 2.7.1.65, 2.7.1.156, 2.7.11.1, 2.7.11.18, 2.7.1.235, 2.7.11.20, 2.7.13.3, 2.7.11.34, 2.7.11.5, 2.7.1.171, 2.7.1.172, 2.7.11.16, 2.7.7.108

Glutathione, synthetase, ATP-binding, domain-like: 2.3.3.8, 6.4.1.4, 6.3.5.5, 6.3.4.16, 6.4.1.1, 6.4.1.3, 6.3.2.3, 6.3.2.2, 6.3.1.9, 2.7.9.2, 6.3.3.1, 6.3.2.23, 2.7.1.159, 6.4.1.2, 6.3.4.6, 6.3.1.8, 6.1.2.1, 6.3.4.13, 2.7.4.21, 6.3.4.14, 6.4.1.7, 2.7.9.1, 6.3.4.18, 2.7.1.134, 2.7.4.24, 6.3.3.5, 6.2.1.56, 6.2.1.5, 6.3.2.4, 6.3.4.23, 2.7.3.13, 6.3.2.32, 6.2.1.9, 6.3.1.17, 6.3.2.29, 6.3.2.48, 6.2.1.39, 6.3.2.47, 6.3.2.49, 6.3.2.43, 6.3.2.11, 6.3.2.30, 6.3.1.21, 6.3.4.24, 6.3.1.12, 6.3.2.59, 6.3.2.42, 6.3.2.41, 6.2.1.13, 6.3.2.33, 6.2.1.18, 6.4.1.5, 2.7.1.58, 6.2.1.4, 2.7.9.6, 2.7.9.5, 2.7.13.3, 2.7.9.4

Actin-like, ATPase, domain: 4.6.1.1, 2.7.1.30, 2.7.1.4, 2.7.1.5, 2.7.2.7, 2.7.1.59, 2.7.1.189, 2.7.1.170, 2.7.2.15, 2.7.1.60, 2.7.1.55, 2.7.1.58, 2.7.1.12, 2.7.2.1, 2.7.1.16, 2.7.1.17, 2.7.1.51, 2.7.1.1, 2.7.1.33, 2.7.1.2, 2.7.1.47, 2.7.1.14, 1.3.7.8, 2.7.1.215, 2.7.1.179, 2.7.1.8, 2.7.1.232, 2.7.1.233, 6.1.2.2, 2.7.1.214, 2.7.1.7, 2.7.1.188, 2.7.1.53, 3.5.2.14, 6.4.1.8, 2.7.1.85, 2.7.1.157, 2.7.1.27, 3.5.2.9, 3.6.1.40, 3.6.1.11, 2.7.10.2, 2.7.11.1, 6.3.4.15, 1.16.99.1, 2.7.4.1, 6.4.1.6

Nucleotide-diphospho-sugar, transferases: 2.7.7.68, 2.7.7.62, 2.7.7.24, 2.7.7.40, 2.7.7.64, 3.6.5.2, 2.7.7.38, 2.7.7.43, 2.7.7.60, 2.7.7.27, 2.7.7.13, 2.7.7.33, 2.7.7.74, 2.7.7.23, 2.7.7.9, 2.7.7.103, 2.7.7.83, 2.7.7.91, 2.7.7.77, 2.7.7.106, 2.7.7.71, 2.7.7.90, 2.7.7.81, 2.7.7.76, 2.7.7.82, 2.7.7.99, 2.7.7.105, 2.7.7.92, 6.4.1.8, 2.7.1.52, 2.7.1.168, 4.6.1.17, 2.7.7.107

Nucleotidylyl, transferase: 6.1.1.1, 6.1.1.4, 6.1.1.10, 6.3.2.1, 6.1.1.16, 6.1.1.6, 2.7.1.26, 2.7.7.3, 6.1.1.5, 2.7.7.1, 2.7.7.15, 2.7.7.2, 2.7.7.39, 6.3.1.13, 2.7.1.25, 2.7.7.18, 6.1.1.9, 6.1.1.2, 6.1.1.19, 2.7.7.14, 2.7.7.4, 6.1.1.17, 6.1.1.18, 2.7.7.70, 2.7.7.104, 2.7.7.93, 6.2.1.22, 6.4.1.8, 6.1.1.24, 2.7.1.237, 3.6.1.73, 6.1.1.15, 2.7.1.24

Ribokinase-like: 2.7.1.64, 2.7.1.4, 2.7.1.11, 4.2.1.93, 2.7.1.35, 2.7.1.83, 2.7.1.20, 2.7.1.73, 2.7.1.15, 2.7.1.3, 2.7.1.49, 2.7.1.50, 2.7.1.184, 2.7.4.7, 2.7.1.187, 2.7.1.144, 2.7.1.45, 2.7.1.2, 2.7.1.218, 2.7.1.223, 2.7.1.92, 2.7.7.70, 2.7.1.178, 2.7.1.213, 2.7.1.13, 2.7.1.101, 6.4.1.8, 2.7.1.58, 6.3.1.12, 2.7.1.56, 2.7.1.239

Class, II, aaRS, and, biotin, synthetases: 6.1.1.4, 6.1.1.14, 6.1.1.15, 6.1.1.20, 6.1.1.6, 6.3.4.15, 6.1.1.11, 6.3.1.1, 6.1.1.7, 6.1.1.27, 6.1.1.21, 6.1.1.26, 6.1.1.22, 6.3.1.20, 6.1.1.17, 6.1.1.23, 6.1.1.12, 6.3.4.11, 6.3.4.9, 6.3.4.10, 6.1.1.3, 2.7.7.75, 2.7.11.1, 6.2.1.14, 2.7.1.33

GHMP, Kinase, C-terminal, domain: 4.1.1.33, 2.7.4.2, 2.7.1.39, 4.6.1.2, 2.7.1.148, 2.7.1.167, 2.7.6.5, 2.7.1.36, 2.7.1.169, 2.7.1.6, 2.7.7.30, 2.7.1.185, 2.7.1.46, 2.7.1.52, 4.1.1.99, 2.7.1.43, 2.7.1.44, 2.7.1.157, 2.7.1.177, 2.7.1.168, 6.3.2.49, 2.7.11.1, 3.6.4.10, 3.6.4.13, 2.7.1.71

Adenine, nucleotide, alpha, hydrolases-like: 2.8.1.13, 6.3.1.5, 2.7.7.2, 6.3.5.1, 6.3.4.5, 6.3.5.4, 6.3.5.2, 2.7.7.4, 6.3.4.19, 6.3.3.4, 6.3.4.20, 4.4.1.37, 6.3.1.14, 5.1.1.23, 6.3.3.6, 2.7.1.108, 6.3.4.24, 2.8.1.15, 2.7.13.3, 2.8.1.14, 2.7.11.1

Nucleotidyltransferase: 2.7.7.7, 4.6.1.1, 2.7.7.52, 2.7.7.72, 2.7.7.47, 2.7.7.42, 2.7.7.84, 2.7.6.5, 2.7.7.19, 2.7.1.24, 2.7.7.59, 2.7.7.46, 2.7.7.65, 2.7.7.31, 2.7.7.85, 2.7.7.108, 2.7.7.86

MurD-like, peptide, ligases, catalytic, domain: 6.3.5.10, 6.3.2.10, 6.3.2.17, 6.3.2.12, 6.3.5.13, 6.3.2.9, 6.3.2.37, 6.3.2.8, 6.3.2.7, 6.3.2.13, 6.3.2.29, 6.3.2.30, 6.3.2.53, 6.4.1.9, 6.3.2.45, 5.1.1.23

Metal, cation-transporting, ATPase, ATP-binding, domain, N: 7.2.2.8, 7.2.2.10, 7.6.2.1, 7.2.2.19, 7.2.2.13, 7.2.2.6, 7.2.2.14, 7.1.2.1, 7.6.2.16, 7.2.2.9, 7.2.2.3, 7.2.2.12, 7.2.2.15, 7.2.2.21, 7.2.2.22

Glutamine, synthetase/guanido, kinase: 6.3.1.11, 6.3.2.3, 6.3.2.2, 2.7.3.2, 2.7.3.5, 2.7.14.1, 2.7.3.3, 6.3.1.2, 2.7.3.4, 2.7.3.1, 6.3.4.12, 6.3.1.6, 6.3.1.18, 6.3.1.19

Nudix: 6.3.2.25, 3.6.1.23, 2.7.7.1, 3.6.1.9, 2.7.7.96, 3.6.1.65, 2.7.6.2, 3.6.1.67, 3.6.1.69, 3.6.1.55, 3.6.1.56

ATPase, domain, of, HSP90

chaperone/DNA, topoisomerase, II/histidine, kinase: 3.6.1.8, 2.7.11.2, 2.7.11.4, 4.6.1.2, 2.7.13.1, 2.7.11.1, 3.6.4.10, 2.7.13.3, 2.7.10.2, 4.6.1.1

Carbamate, kinase-like: 2.7.2.8, 2.7.2.11, 2.7.4.26, 2.7.4.22, 2.7.2.4, 2.7.2.2, 2.7.4.4, 2.7.4.31, 2.7.2.19, 2.7.2.17

SAICAR, synthase-like: 2.7.1.127, 2.7.1.151, 2.7.1.149, 2.7.1.140, 2.7.1.68, 2.7.4.21, 6.3.2.6, 2.7.1.150, 2.7.1.153, 2.7.11.1

DNA/RNA, polymerases: 2.7.7.7, 2.7.7.6, 2.7.7.4, 2.7.7.48, 3.6.1.15, 3.6.4.13, 2.7.7.49, 2.7.7.50, 2.7.7.19

HD-domain/PDEase-like: 6.1.1.14, 2.7.6.5, 3.1.5.1, 2.7.7.59, 3.6.1.40, 3.6.1.11, 2.7.7.18, 2.7.4.1

Nucleotide, cyclase: 4.6.1.1, 2.7.7.65, 4.6.1.2, 4.6.1.6, 4.6.1.26, 2.7.11.1, 3.5.4.29, 2.7.7.79

DNA, ligase/mRNA, capping, enzyme, catalytic, domain: 6.5.1.1, 2.7.7.50, 6.5.1.6, 6.5.1.3, 6.5.1.7, 2.7.7.48, 2.7.1.78

NAD, kinase/diacylglycerol, kinase-like: 2.7.1.107, 2.7.1.86, 2.7.1.91, 2.7.1.23, 2.7.1.138, 2.7.1.94, 2.7.1.93

Phosphoenolpyruvate/pyruvate, domain: 2.7.1.40, 2.7.9.2, 2.7.9.1, 2.7.7.104, 6.2.1.9, 2.7.10.2, 2.7.11.1

ClpP/crotonase: 6.4.1.4, 6.4.1.3, 6.4.1.2, 6.3.4.14, 6.4.1.5, 6.4.1.7, 2.7.13.3

Aerobactin, siderophore, biosynthesis, lucA/lucC-like: 6.3.2.38, 6.3.2.39, 6.3.2.54, 6.3.2.55, 6.3.2.56, 6.3.2.57, 6.3.2.58

Alphavirus-like, methyltransferase, domain: 2.7.7.19, 2.7.7.48, 3.1.3.84, 3.4.19.12, 3.6.1.15, 3.6.1.74, 3.6.4.13

PurM, N-terminal, domain-like: 2.7.9.3, 6.3.3.1, 6.3.5.3, 2.7.4.16, 6.1.1.24, 6.3.4.13

Mitochondrial, carrier: 6.3.2.39, 2.7.1.66, 2.7.7.67, 2.7.1.174, 2.7.1.182, 2.7.1.216

Activating, enzymes, of, the, ubiquitin-like, proteins: 2.7.7.80, 2.7.7.73, 2.7.7.100, 6.2.1.64, 6.2.1.45

DhaL-like: 2.7.1.29, 2.7.1.210, 2.7.1.209, 2.7.1.28

HIT-like: 2.7.7.53, 2.7.7.10, 6.3.2.39, 3.6.1.29

Metallo-dependent, phosphatases: 3.1.3.1, 3.6.1.15, 2.7.11.5, 3.6.1.5

NagB/RpiA/CoA, transferase-like: 6.3.3.2, 3.6.5.3, 2.7.1.29, 6.3.4.22

PRTase-like: 2.7.6.1, 6.3.4.19, 3.5.2.14, 6.3.2.47  
YgbK-like: 2.7.1.219, 2.7.1.231, 2.7.1.217, 2.7.1.220  
Metallo-dependent, hydrolases: 6.3.1.10, 6.3.5.5, 3.6.1.63, 6.3.4.16  
Alkaline, phosphatase-like: 3.6.1.9, 3.1.3.1, 3.6.1.29  
all-alpha, NTP, pyrophosphatases: 3.6.1.8, 3.6.1.12, 3.6.1.23  
beta, and, beta-prime, subunits, of, DNA, dependent, RNA-polymerase: 2.7.7.6, 2.7.7.48, 3.6.1.15  
CYTH-like, phosphatases: 4.6.1.1, 2.7.7.50, 3.6.1.28  
dUTPase-like: 3.6.1.23, 3.5.4.30, 2.7.7.49  
EPT/RTPC-like: 6.5.1.4, 6.5.1.5, 2.7.1.71  
Fic-like: 2.7.7.1, 2.7.11.1, 2.7.7.108  
ITPase-like: 3.6.1.15, 3.6.1.66, 3.6.1.73  
Molybdenum, cofactor, biosynthesis, proteins: 2.7.7.75, 2.7.7.76, 4.6.1.17  
PAP/Archaeal, CCA-adding, enzyme, C-terminal, domain: 2.7.7.72, 2.7.7.19, 6.2.1.14  
Phosphofructokinase: 2.7.1.11, 2.7.1.184, 2.7.1.144  
Riboflavin, kinase-like: 2.7.1.26, 2.7.7.2, 2.7.1.161  
SIS, domain: 2.7.1.2, 6.3.2.48, 2.7.1.168, 6-hydroxymethyl-7,8-dihydropterin, pyrophosphokinase, HPPK: 6.3.2.12, 2.7.6.3  
Calcium-dependent, phosphotriesterase: 3.1.3.1, 2.7.13.3  
CoaB-like: 6.3.2.51, 6.3.2.5  
CofE-like: 6.3.2.34, 6.3.2.31  
DHH, phosphoesterases: 3.6.1.11, 6.1.1.7  
Glycerate, kinase, I: 2.7.1.31, 2.7.1.165  
Molybdenum, cofactor, biosynthesis, protein, C, MoaC: 2.7.7.77, 4.6.1.17  
Nicotinate/Quinolinate, PRTase, C-terminal, domain-like: 6.3.1.10, 3.5.2.14  
Nucleoside, diphosphate, kinase, NDK: 2.7.4.6, 2.7.13.3  
ParB/Sulfiredoxin: 2.7.1.225, 1.8.98.2  
PEP, carboxykinase-like: 4.1.1.32, 4.1.1.49  
Thiamin, pyrophosphokinase, catalytic, domain: 2.7.6.2, 2.7.6.3  
Tubulin, nucleotide-binding, domain-like: 3.6.5.6, 6.3.2.25  
YrdC/RibB: 2.7.7.87, 3.5.4.25  
Adenylylcyclase, toxin, (the, edema, factor): 4.6.1.1  
Apyrase: 3.6.1.5  
F1FO ATP, synthase, subunit, C: 7.1.2.2  
GlnB-like: 6.3.5.10  
GroEL, equatorial, domain-like: 5.6.1.7  
Phosphoglycerate, kinase: 2.7.2.3  
Phospholipase, D/nuclease: 2.7.4.1  
Poly(A), polymerase, catalytic, subunit-like: 2.7.7.19  
Prim-pol, domain: 2.7.7.7  
RibA-like: 3.5.4.25  
YojJ-like: 2.7.7.85  
S-adenosylmethionine, synthetase: 2.5.1.6  
Inositol-pentakisphosphate, 2-kinase: 2.7.1.158  
tRNA-splicing, ligase, RtcB-like, superfamily: 6.5.1.8  
Bacterial, DNA, polymerase, III, alpha, subunit, NTPase, domain: 2.7.7.7  
tRNA(Ile2), 2-agmatinylcytidine, synthetase, TiaS: 6.3.4.22  
Phosphatidate, cytidylyltransferase: 2.7.7.41  
Virion, DNA-directed, RNA, polymerase, domain: 2.7.7.6

GTP, cyclohydrolase, MptA: 3.5.4.39  
Influenza, RNA-dependent, RNA, polymerase, subunit, PB1: 2.7.7.48  
Phosphopantoate/pantothenate, synthetase, superfamily: 6.3.2.36  
RNA-dependent, RNA, polymerase, eukaryotic-type: 2.7.7.48  
DNA, primase, core: 2.7.7.101  
Diacylglycerol, kinase, (DgkA)-like: N/A  
RPB5-like, RNA, polymerase, subunit: N/A  
Transglutaminase, two, C-terminal, domains: N/A  
YcaO-related, McrA-glycine, thioamidation, protein: N/A  
Peptidase, G2, IMC, autoproteolytic, cleavage, domain: N/A

Note 2: NTP processing ECs with no associated superfamily

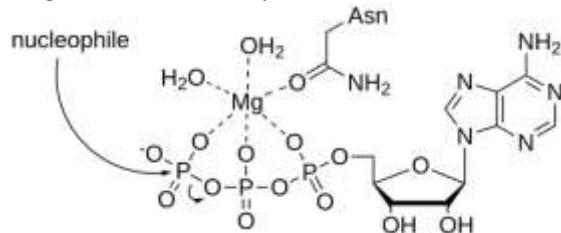
1.18.6.2	2.7.3.6	6.2.1.38
1.19.6.1	2.7.3.7	6.2.1.52
2.7.1.10	2.7.3.8	6.2.1.55
2.7.1.102	2.7.4.11	6.2.1.58
2.7.1.122	2.7.4.12	6.2.1.6
2.7.1.136	2.7.4.15	6.2.1.73
2.7.1.18	2.7.4.18	6.2.1.74
2.7.1.234	2.7.4.19	6.2.2.1
2.7.1.236	2.7.4.32	6.2.2.2
2.7.1.238	2.7.6.4	6.2.2.3
2.7.1.34	2.7.7.11	6.3.1.4
2.7.1.54	2.7.7.28	6.3.1.7
2.7.1.84	2.7.7.32	6.3.2.16
2.7.1.88	2.7.7.34	6.3.2.24
2.7.10.3	2.7.7.44	6.3.2.35
2.7.11.28	2.7.7.45	6.3.2.44
2.7.11.29	2.7.7.57	6.3.2.60
2.7.11.3	3.6.1.39	6.3.2.61
2.7.11.35	3.6.4.7	6.3.2.62
2.7.11.6	4.1.2.50	6.3.4.17
2.7.11.8	4.2.3.12	6.3.4.7
2.7.11.9	5.6.1.2	6.3.4.8
2.7.13.2	5.6.1.3	7.4.2.2
2.7.2.10	5.6.1.5	7.4.2.3
2.7.2.14	5.6.1.9	7.5.2.14
2.7.2.18	6.2.1.10	7.5.2.2
2.7.2.6	6.2.1.11	
2.7.3.10	6.2.1.28	

## Phosphatase superfamilies

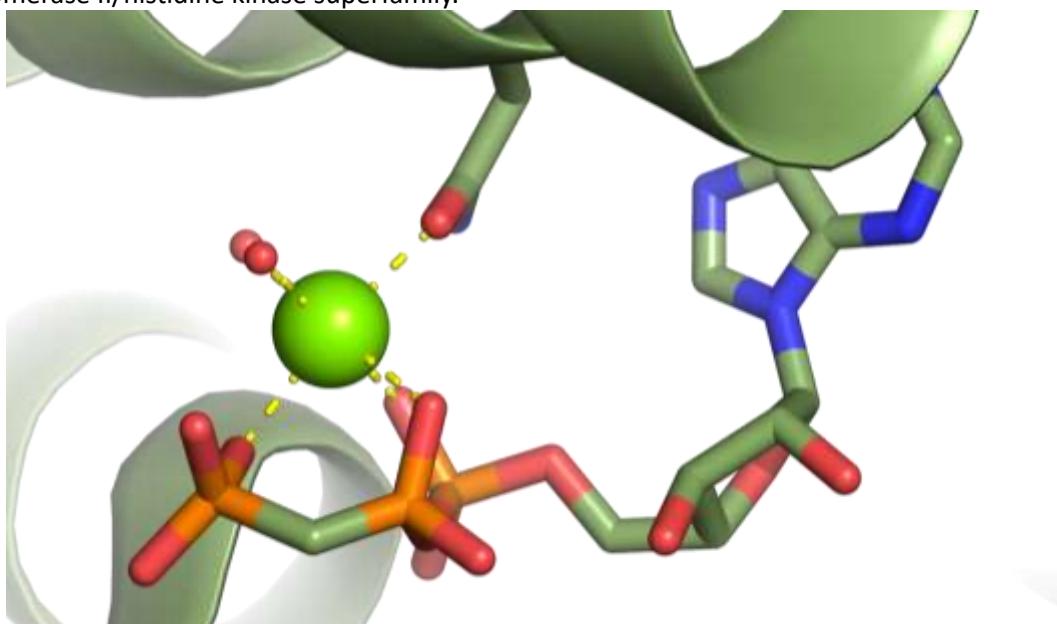
Single metal phosphatases with  $\alpha\beta\gamma$  coordination on the (+) side

### ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase

The  $\alpha\beta\gamma$  coordination is unambiguous in this family and all active sites have one metal ion.



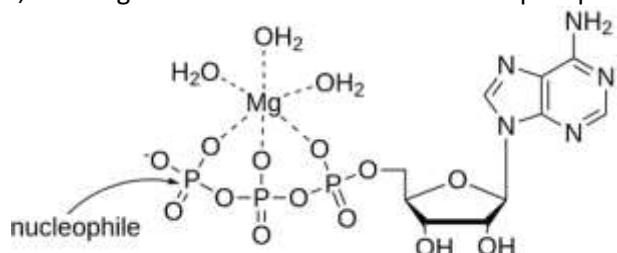
Typical coordination and reaction facilitated by the ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase superfamily.



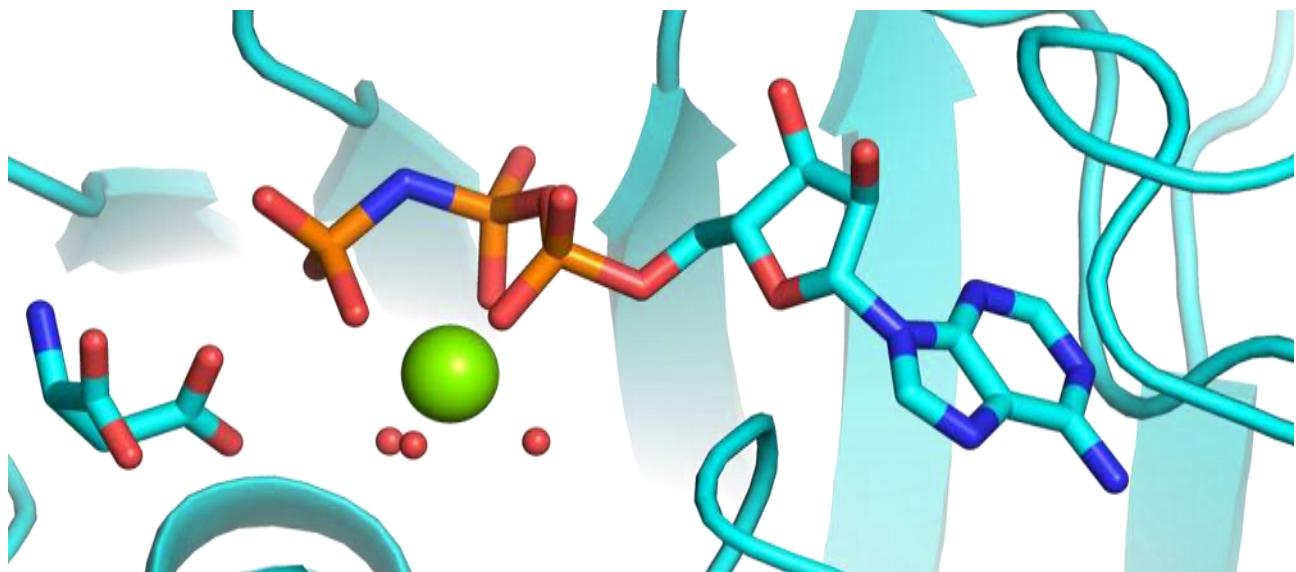
ATP binding represented by PDB structure 1I58-A.

### Carbamate kinase-like

Some enzymes are known to have kinase function on small molecules, however many molybdenum storage protein structures have the same fold. Their ATPase activity is connected to molybdenite release [2]. These structures are not necessarily in their active forms and exhibit only  $\alpha\beta$  phosphate coordination. In the kinases, the magnesium ion coordinates all three phosphates.



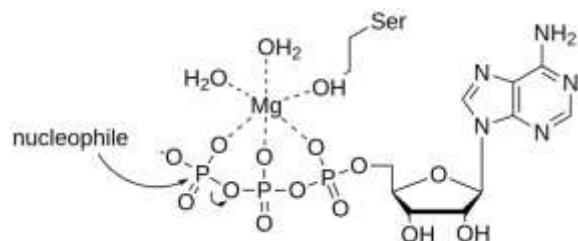
Typical coordination and reaction catalyzed by the Carbamate kinase-like superfamily.



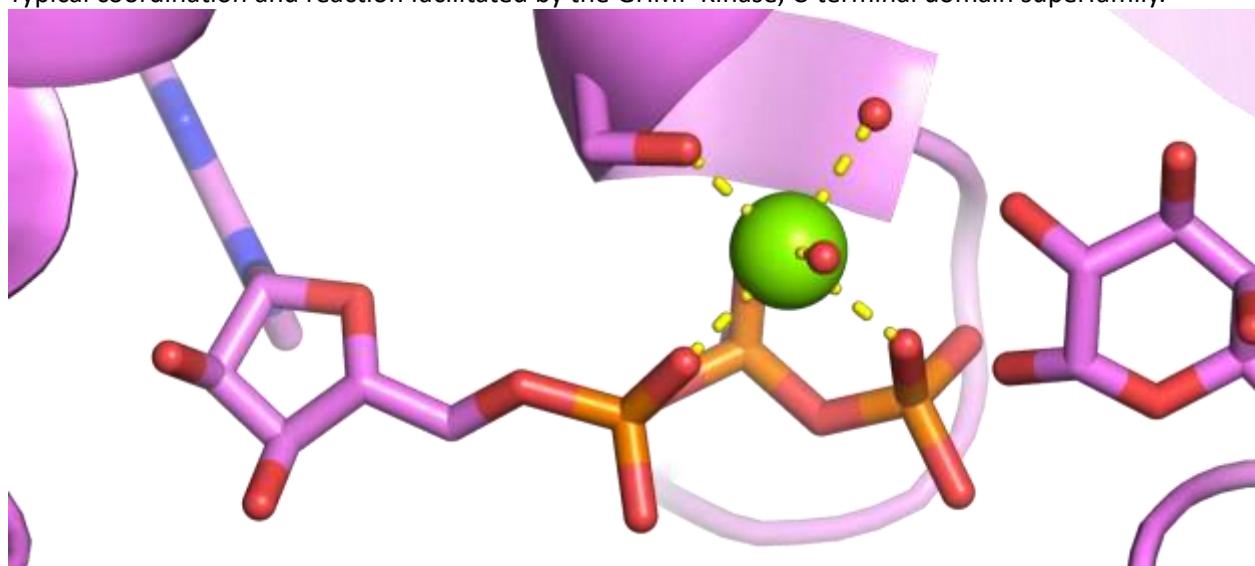
ATP and metal ion coordination in the representative structure 3C1M.

#### **GHMP Kinase, C-terminal domain (Ribosomal protein S5 domain 2-like)**

The collected structures agree in the active site architecture, binding the ATP at the end of a helix, although in two of them the magnesium ion is misplaced, not coordinated by the conserved serine. Interestingly, this superfamily is commonly identified in the same chain as more common ones as 'ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase' and 'P-loop containing nucleoside triphosphate hydrolases'.



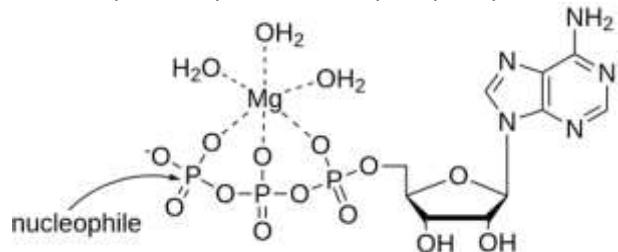
Typical coordination and reaction facilitated by the GHMP Kinase, C-terminal domain superfamily.



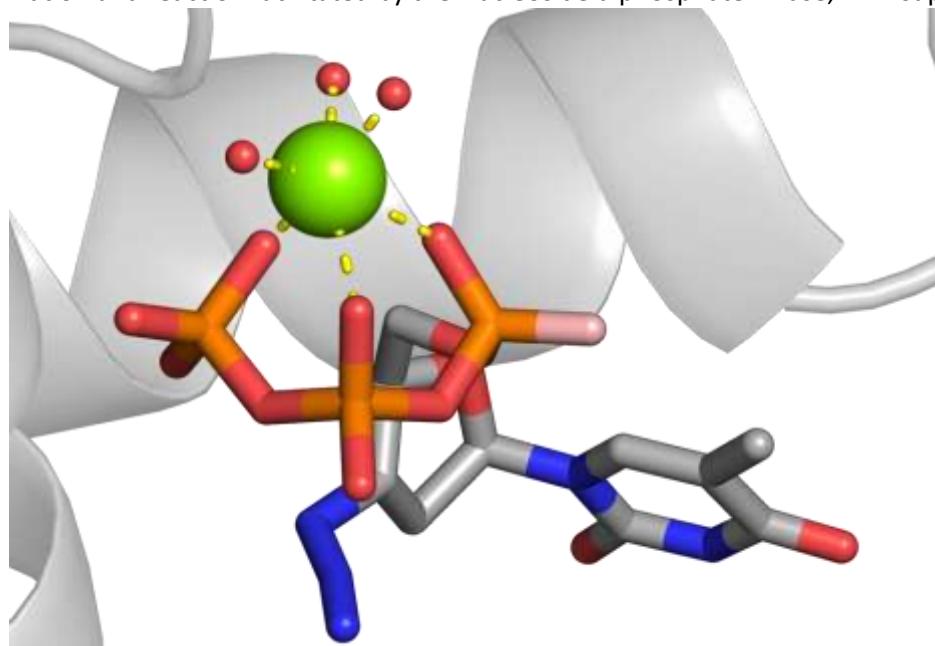
ATP binding represented by PDB structure 3V2U.

### Nucleoside diphosphate kinase, NDK

Based on the EC association, this superfamily uses ATP to phosphorylate am NDP or L-histidine.



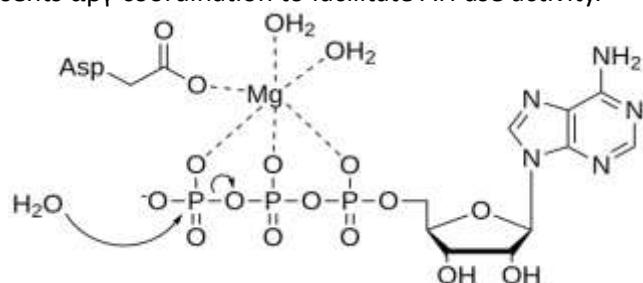
Typical coordination and reaction facilitated by the Nucleoside diphosphate kinase, NDK superfamily.



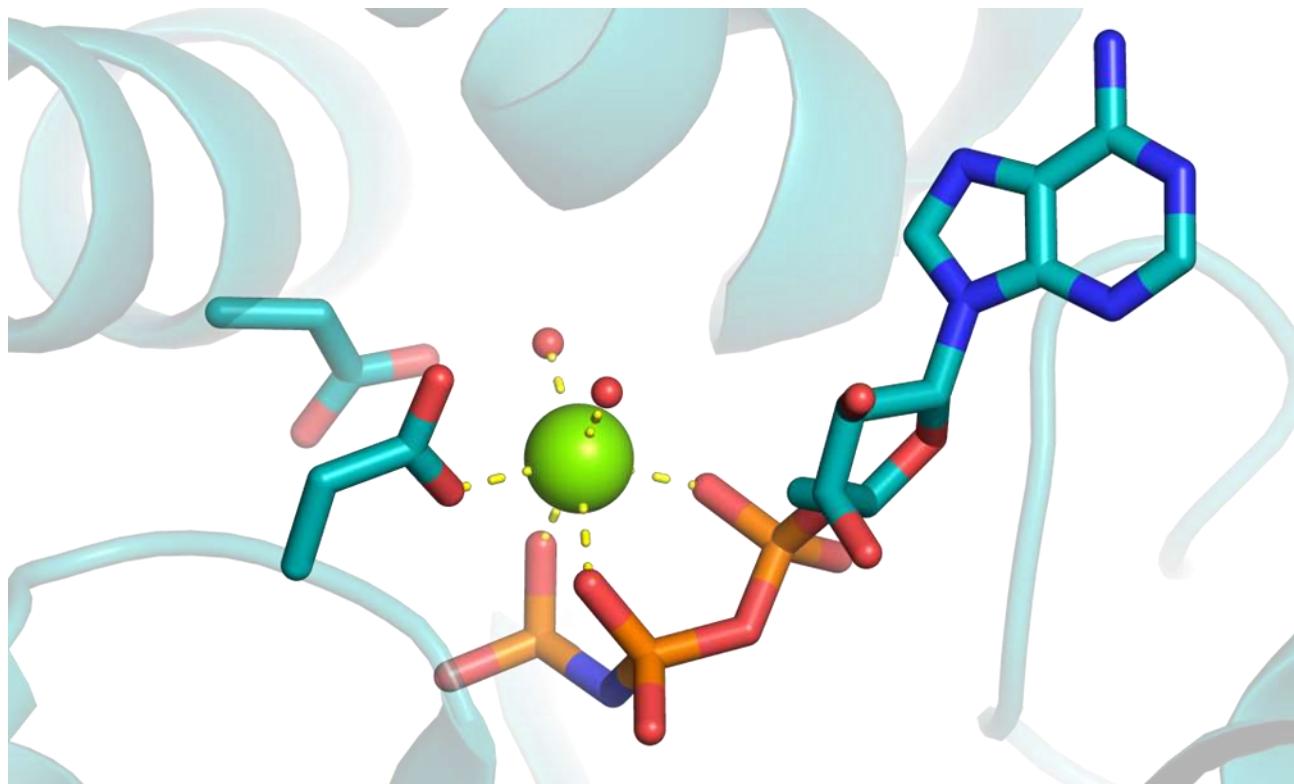
Nucleotide binding represented by PDB structure 1F3F.

### GroEL equatorial domain-like

This superfamily also presents  $\alpha\beta\gamma$  coordination to facilitate ATPase activity.



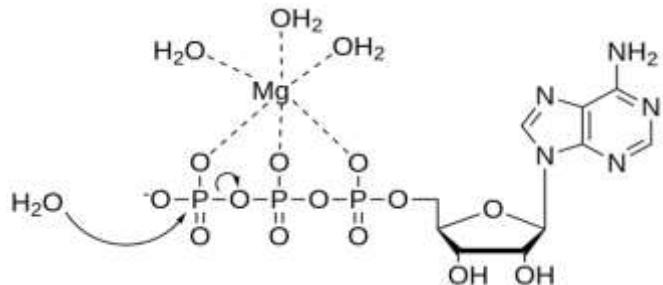
Typical coordination and reaction facilitated by the GroEL equatorial domain-like superfamily.



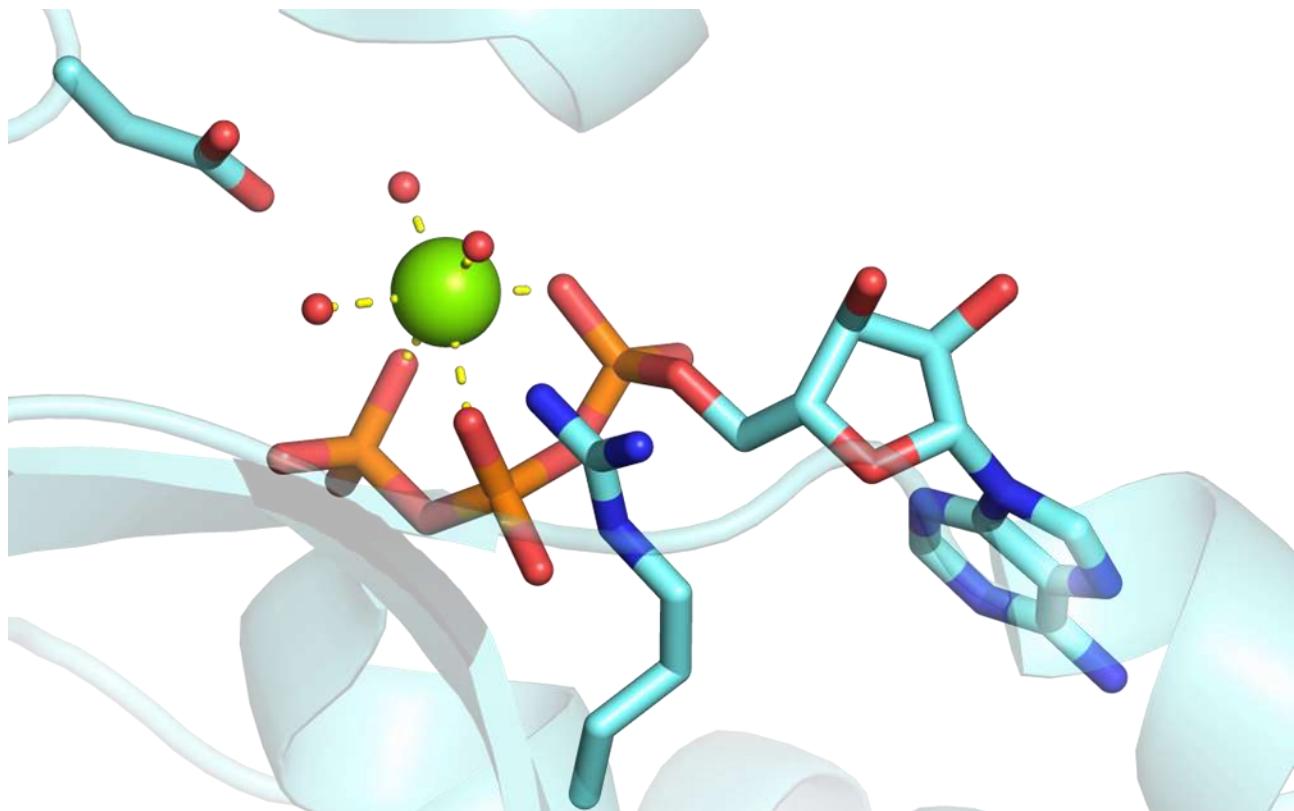
ATP binding represented by PDB structure 3RUV.

### ParB/Sulfiredoxin

Similarly to the NDK, the Sulfiredoxin superfamily contains a few specialized structures, all with clear  $\alpha\beta\gamma$  phosphate coordination.



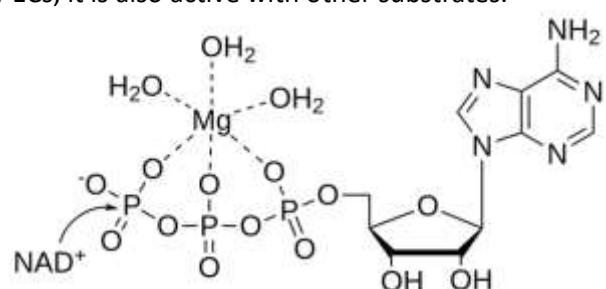
Typical coordination and reaction facilitated by the Sulfiredoxin superfamily.



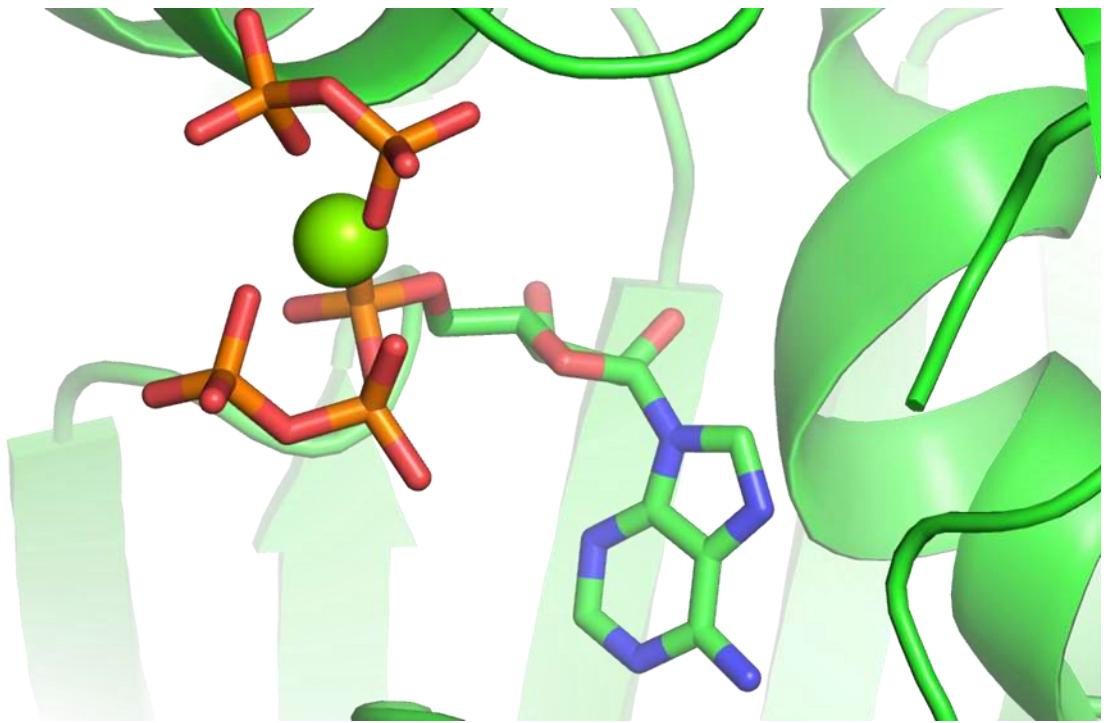
ATP binding represented by PDB structure 3HY2.

#### NAD kinase/diacylglycerol kinase-like

$\alpha\beta$ γ phosphate coordination like the superfamilies above. There are structures with bound NAD<sup>+</sup> and NADP<sup>+</sup>, but the binding pose overlaps with the ATP binding, therefore it is unclear how the catalytic conformation looks like. By ECs, it is also active with other substrates.



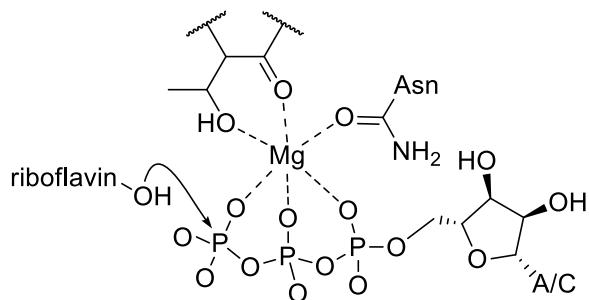
Coordination and reaction facilitated by the NAD<sup>+</sup> kinase superfamily, based on the example 1Z0Z.



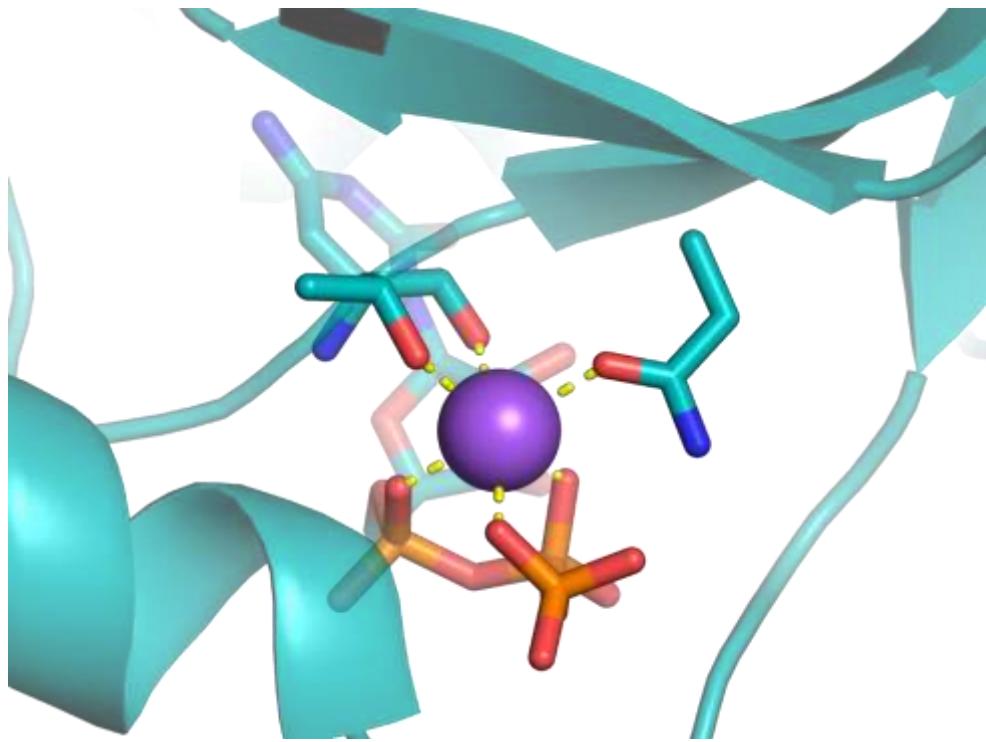
ATP binding represented by PDB structure 1Z0Z.

### Riboflavin kinase-like

The fold often occurs alongside the *Adenine nucleotide alpha hydrolases-like*, but it is associated with phosphatase activity. 2.7.1.26 consumes ATP, 2.7.1.161 is CTP dependent, they align reasonably well. Limited number of NTP bound structures lack ions, 5TRD contains a sodium from which we derive the likely  $\alpha\beta\gamma$  coordination.



Coordination and reaction facilitated by the Riboflavin kinase-like superfamily, based on the example 5TRD.

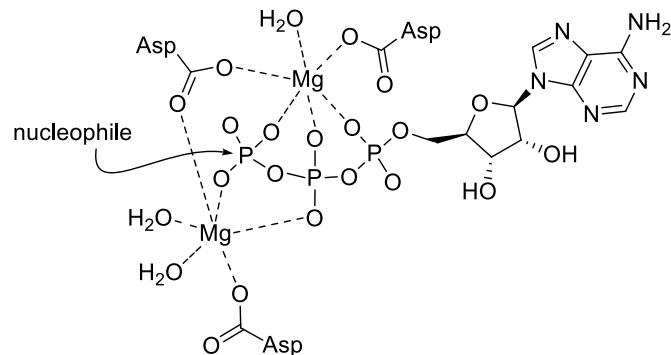


NTP binding and ion coordination represented by PDB structure 5TRD.

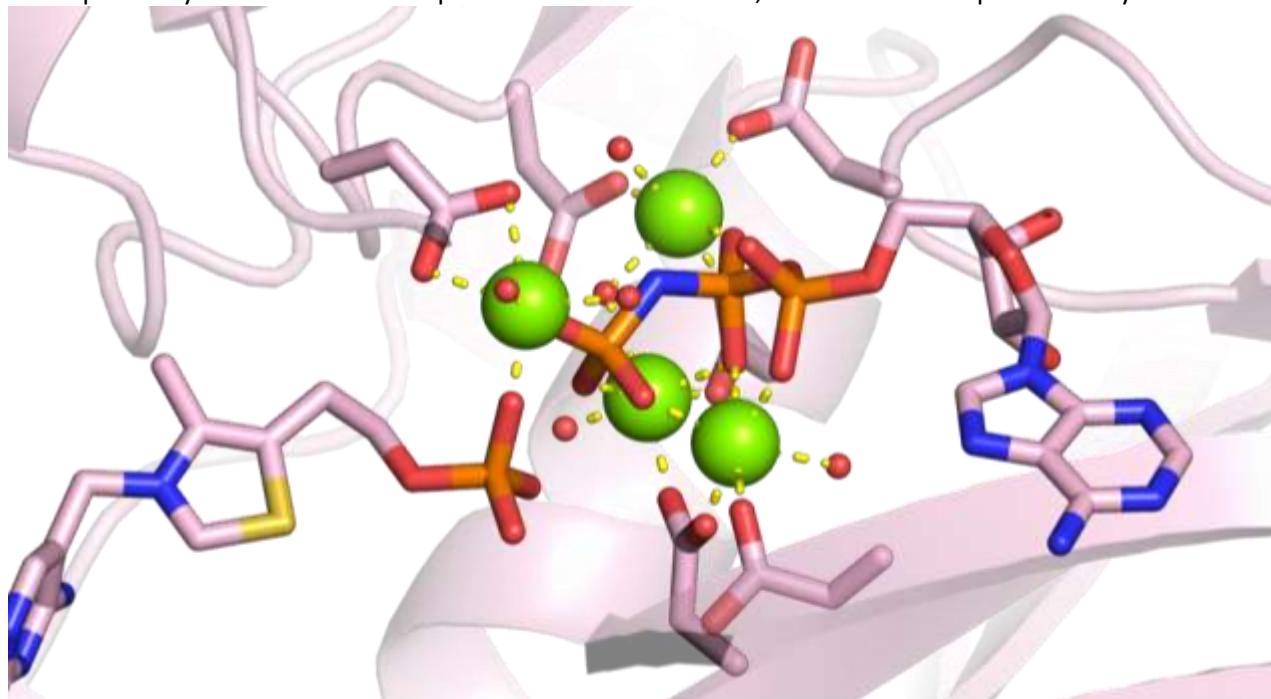
## Multiple metal ion phosphatases with $\alpha\beta\gamma$ coordination on the (+) side

### PurM N-terminal domain-like

Besides the  $\alpha\beta\gamma$  coordinating metal ion, which is always present, members of this fold contain other metal ions as well. One additional  $\beta\gamma$  coordinating ion is resolved in most structures, but still in some three or four metal ions are within the active site.



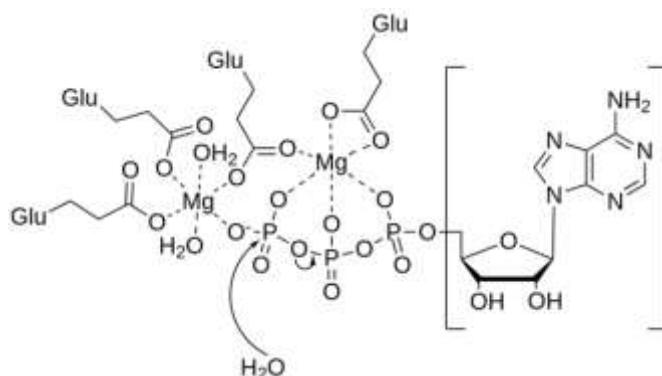
Metal coordination of the two main metal sites and reaction facilitated by the PurM N-terminal domain-like superfamily. Additional ions are present in some structures, coordination is represented by 5DD7.



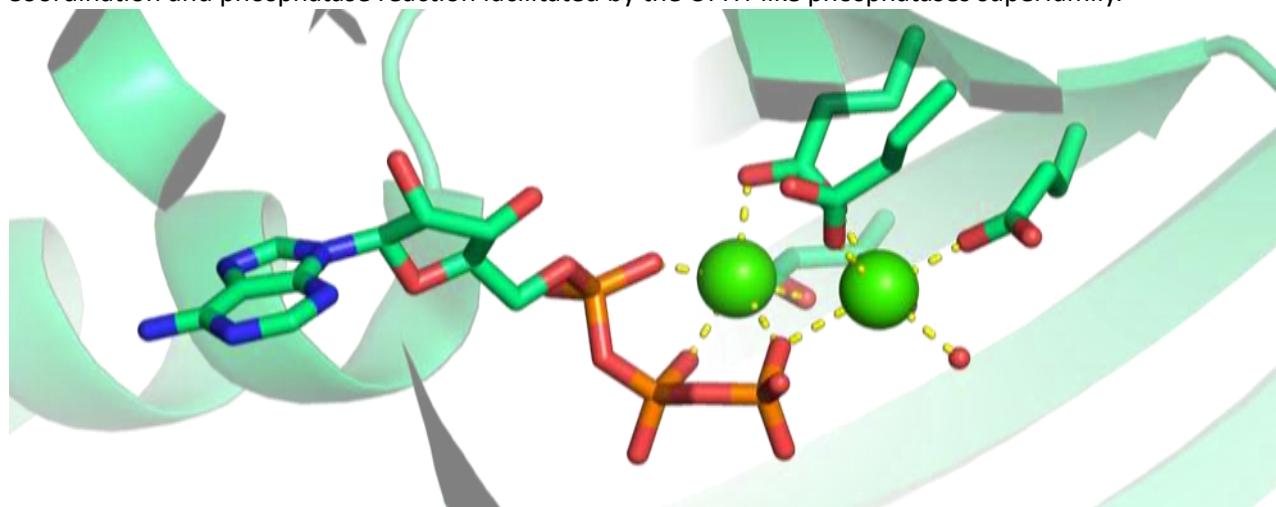
ATP and substrate binding represented by PDB structure 5DD7.

### CYTH-like phosphatases

The main reaction carried out by proteins in the superfamily is hydrolysing inorganic triphosphates. One crystal structure had a bound ATP, hence the similarity of phosphatase enzymes could be identified. Interestingly, class IV adenylyl cyclases are also members of this superfamily [3], despite the cyclase activity cleaving a pyrophosphate.



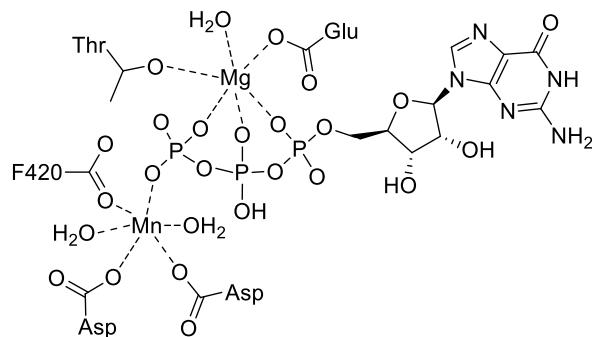
Coordination and phosphatase reaction facilitated by the CYTH-like phosphatases superfamily.



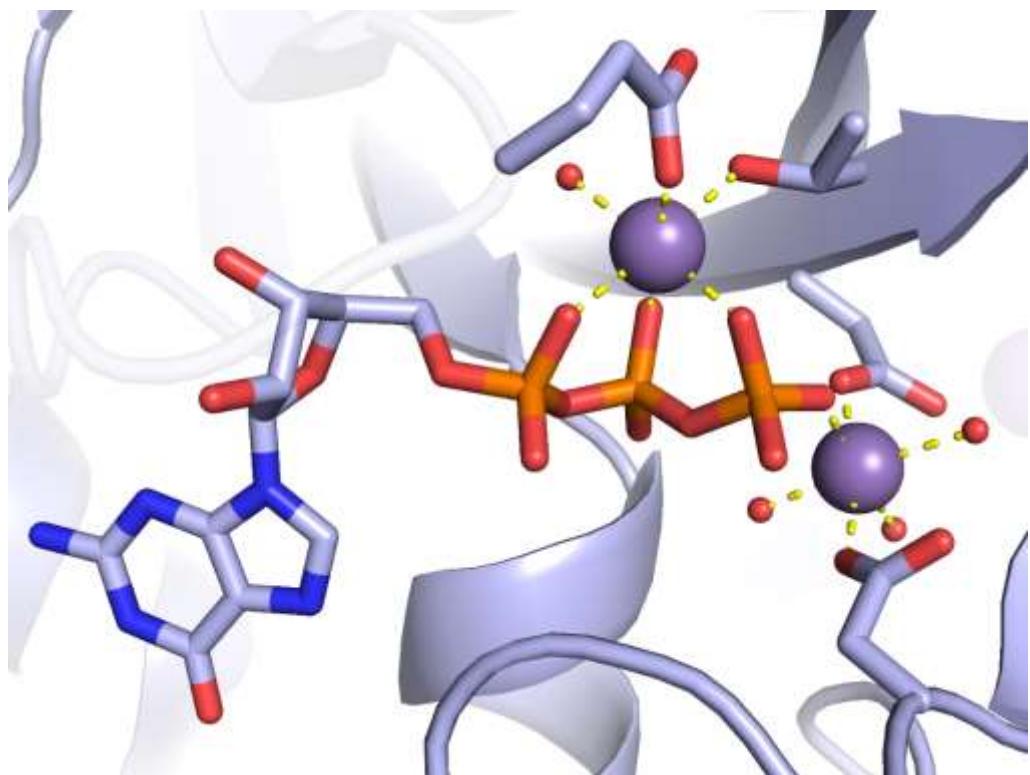
Substrate binding represented by PDB structure 7NSD.

### CofE-like

The study corresponding to the only two deposited structures we identified for this superfamily has not been published at the time of our analysis. They have 2 metal ions; one  $Mn^{2+}$  ion is coordinated by the nonesterified oxygens of the  $\alpha\beta\gamma$  phosphates, a glutamate and a threonine residue, and a water molecule. The other  $Mn^{2+}$  ion is coordinated by an oxygen of the  $\gamma$  phosphate, two aspartate residues, and three water molecules.



By EC, the catalyzed reaction is a two-step amide bond formation, of which the first is a kinase-like reaction by the carboxyl group of coenzyme F420 variants. The  $\alpha\beta\gamma$  coordinating ion is thought to be a magnesium, while a manganese coordinates the nucleophile as well, not resolved in the experimental structure [4].

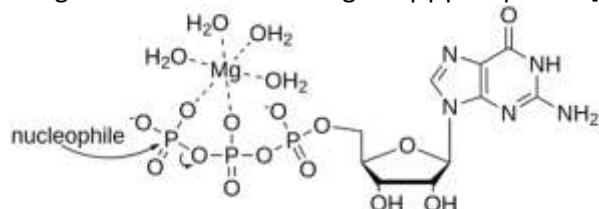


NTP binding and ion coordination represented by PDB structure 7ULD.

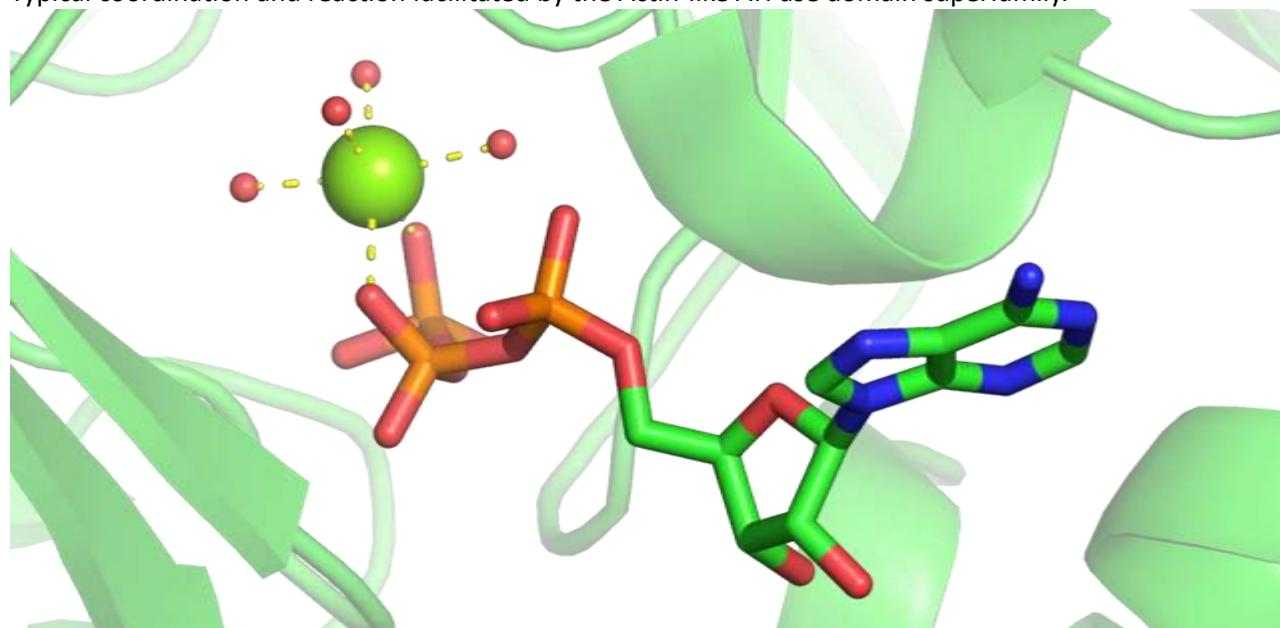
## Phosphatases with $\beta$ coordination from the (+) side

### Actin-like ATPase domain

With well over a hundred structures, the Actin-like ATPase domain superfamily having a well-established ATP active site with a single magnesium ion coordinating the  $\beta\gamma$  phosphates [5].



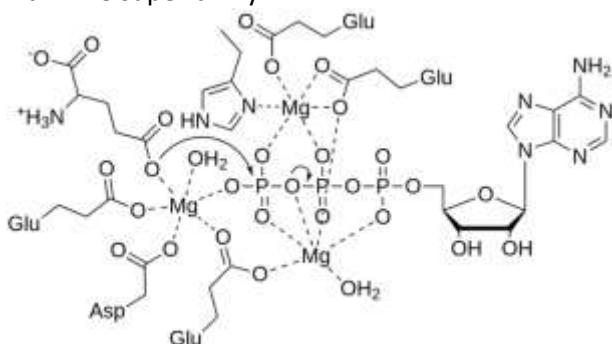
Typical coordination and reaction facilitated by the Actin-like ATPase domain superfamily.



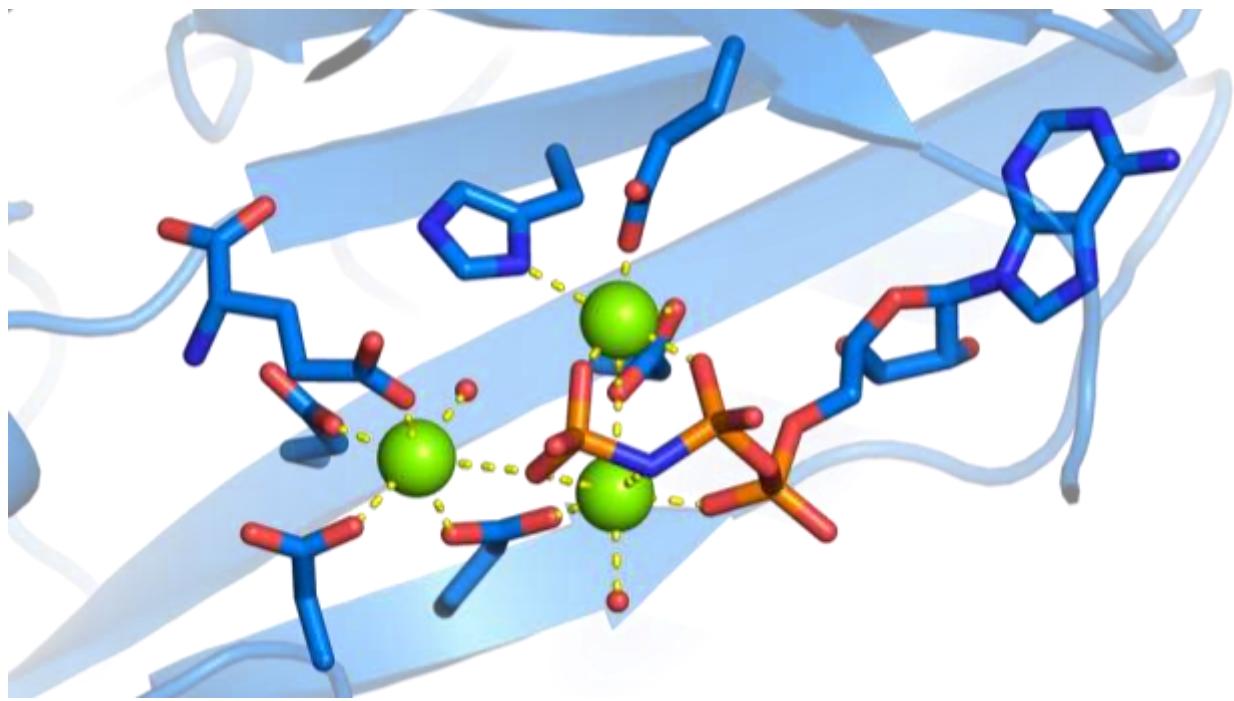
Substrate binding represented by PDB structure 1D4X.

### Glutamine synthetase/guanido kinase

This is the only superfamily which features ions from (+) side of the triphosphate and AG ion in its plane at the same time, the first only pinching the  $\beta$  and  $\gamma$  phosphates. A third ion is present near the nucleophile. This coordinative arrangement is very similar to the mirror image of the one of Glutathione synthetase ATP-binding domain-like superfamily.



Coordination network and reaction facilitated by the Glutamine synthetase/guanido kinase superfamily.

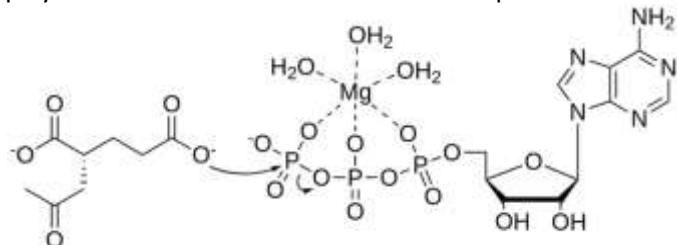


Metal ion coordination represented by the crystal structure 7CQQ.

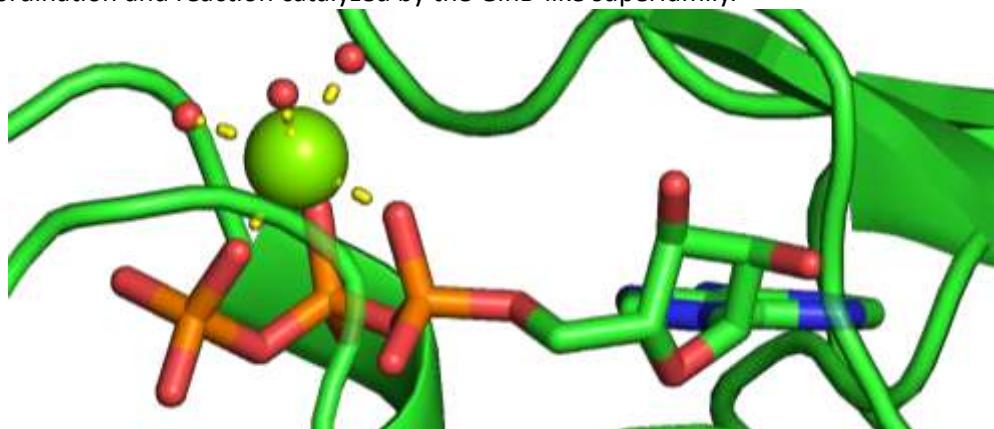
## Phosphatases with $\alpha\beta\gamma$ coordination on the (-) side

### GlnB-like

Although the position of the metal ion is quite well-defined, it is unclear from the structures whether any protein residue is involved in the metal coordination and many often organic substrates chelate the metal ion, hence we displayed waters in coordination as in the representative structure 2J9C.



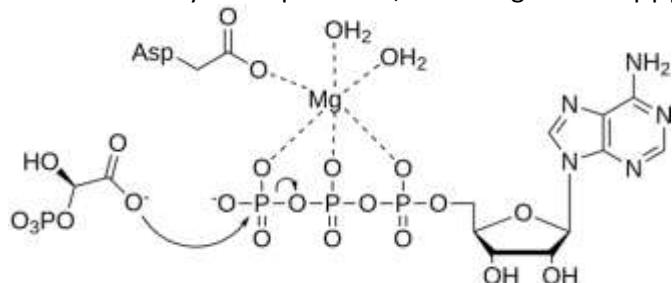
Typical coordination and reaction catalyzed by the GlnB-like superfamily.



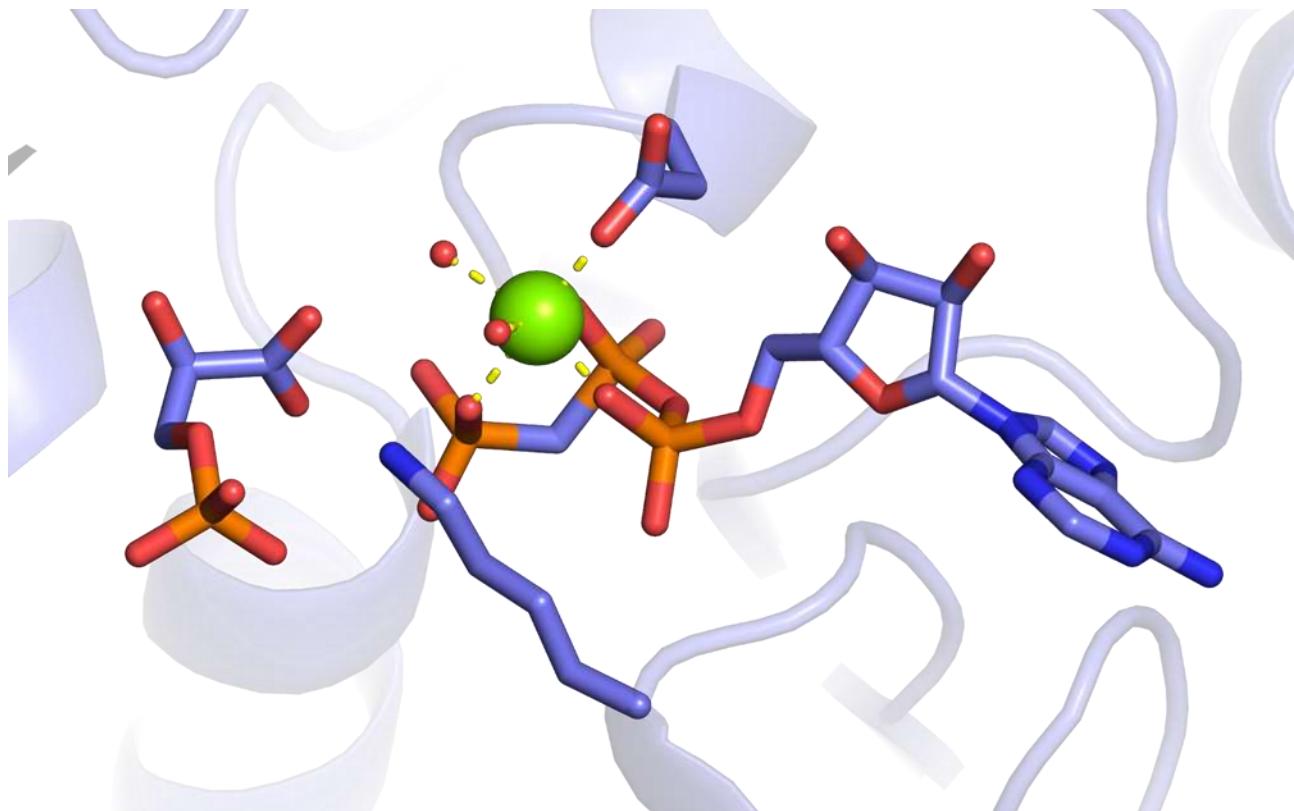
Metal ion and ATP coordination represented by the crystal structure 2J9C.

### Phosphoglycerate kinase

This superfamily contains structures only for a specific EC, exhibiting a clear  $\alpha\beta\gamma$  phosphate coordination.



Typical coordination and reaction catalyzed by the member of the phosphoglycerate kinase superfamily.

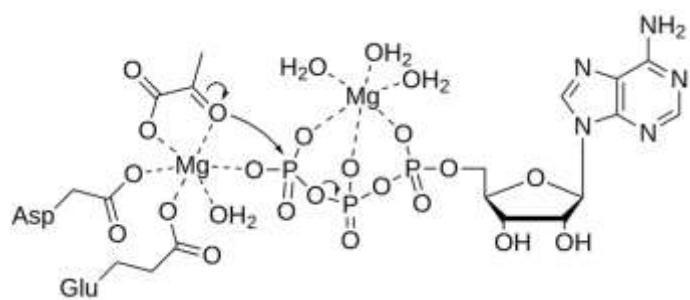


ATP and metal ion coordination in the representative structure 2X14.

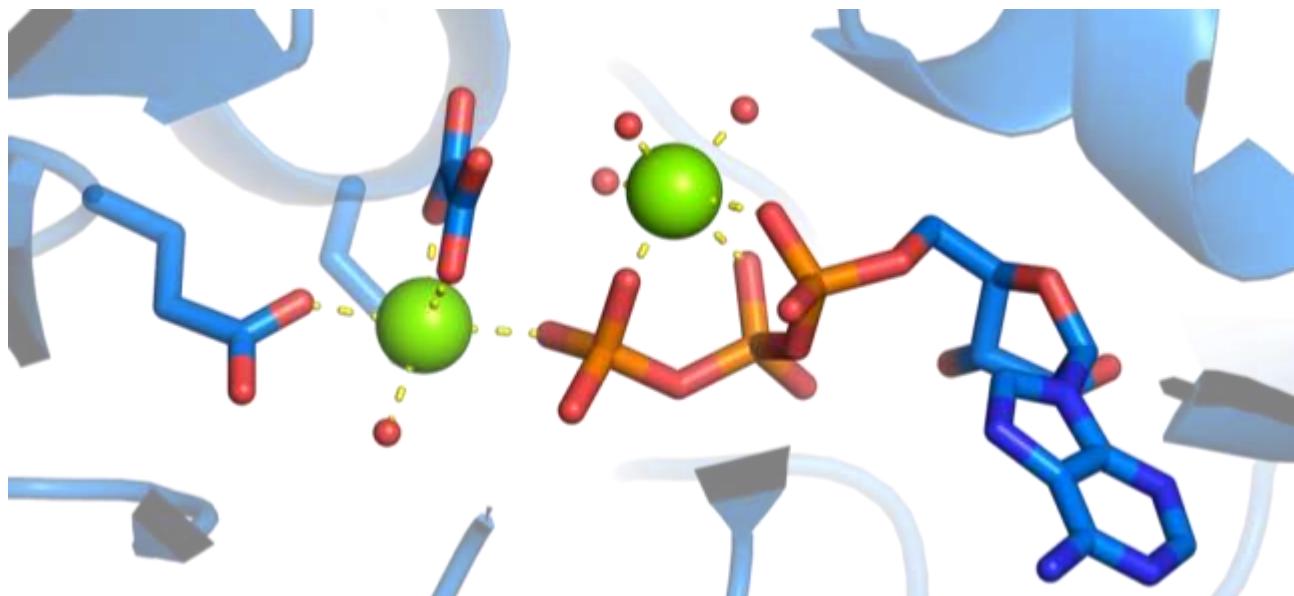
#### Phosphoenolpyruvate/pyruvate domain

A few structures corresponding to EC 4.1.3.25 do not have a NTP residue or phosphatase activity, however their sequences do resemble the superfamily. In the analysis of the coordination, they are omitted. In addition to the magnesium in contact with the  $\alpha\beta\gamma$  phosphates, a second one can be found near the pyruvate.

The superfamily also includes structures of the pyruvate phosphate dikinases (EC 2.7.9.1-2), which perform pyrophosphatase activity [6].



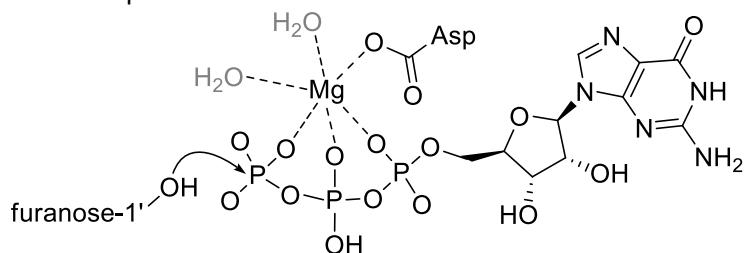
Typical coordination and reaction facilitated by the members of the Phosphoenolpyruvate/pyruvate domain superfamily.



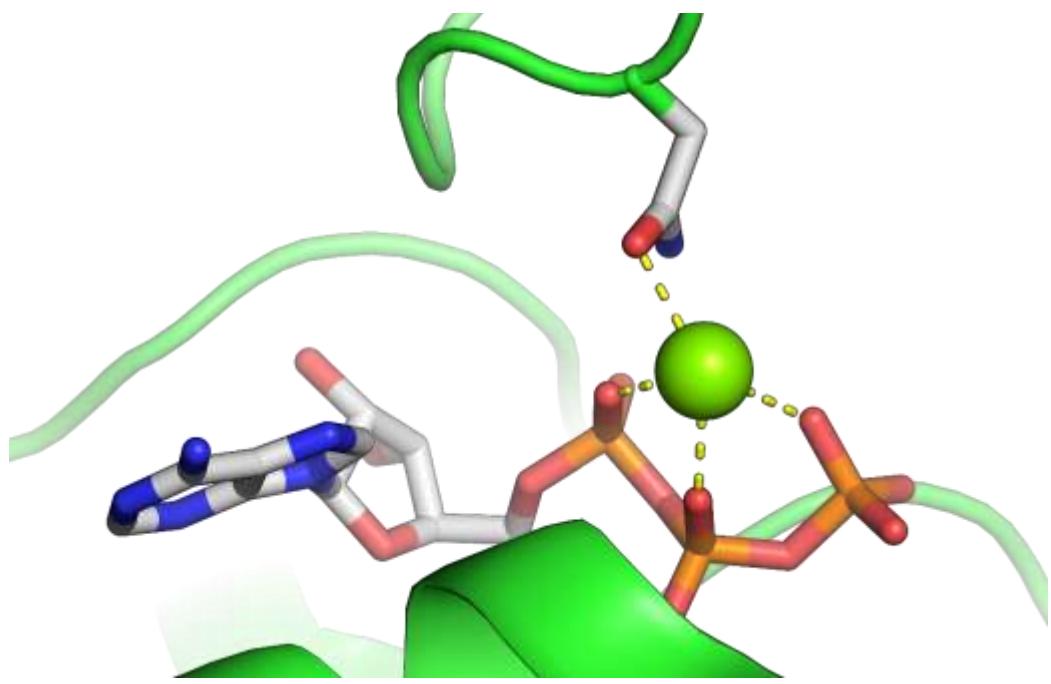
ATP and metal ion binding in the representative structure 6KSH.

### Phosphofructokinase

Phosphofructokinases (PFKs) are enzymes that catalyze the production of fructose 1,6-bisphosphate (F-1,6-BP) early in the glycolytic pathway. In mammals, plants, yeasts, many protists, and bacteria, PFKs use ATP as the phosphate donor in this essentially irreversible reaction [7]. The  $Mg^{2+}$  ion is coordinated by all three phosphate groups, and by the carboxyl group of an aspartate that emerges from a neighboring, large, inserted loop.



Sugar kinase reactivity and Mg binding in the Phosphofructokinase superfamily.

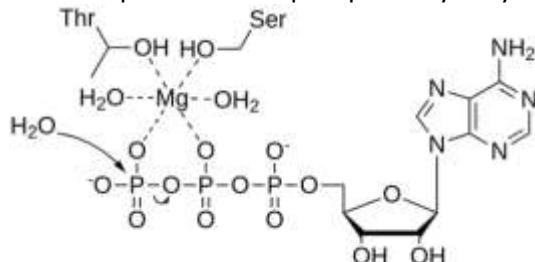


ATP and metal ion binding in the representative structure 3F5M.

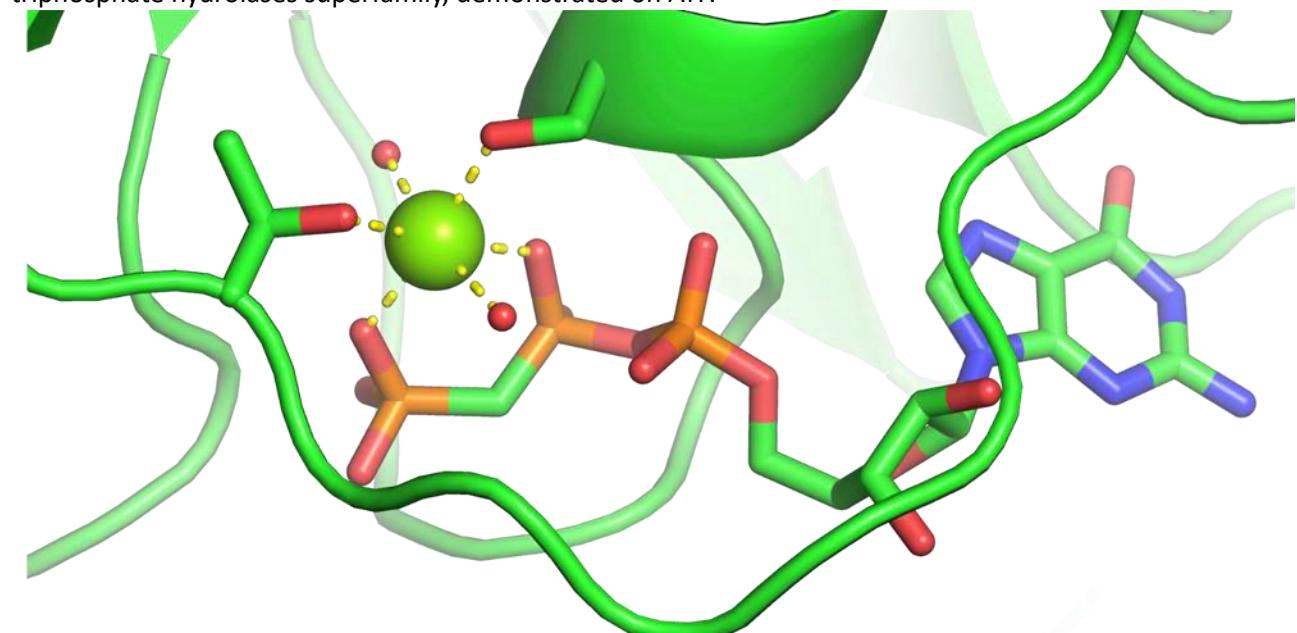
## Single metal phosphatases with $\beta\gamma$ coordination on the (-) side

### P-loop containing nucleoside triphosphate hydrolases

By far the largest superfamily in our dataset. Despite the numbers, the coordination is quite conserved and the structures align well. In a small part of the superfamily, the coordinating threonine is missing, which resembles the conformation adopted after the phosphate hydrolysis in some enzymes.



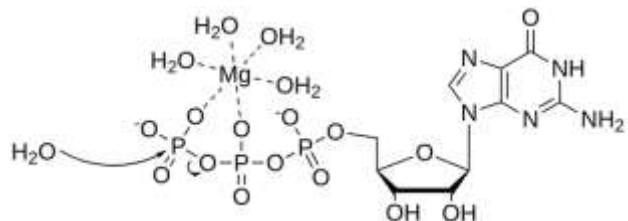
Typical coordination and reaction catalyzed by the members of the P-loop containing nucleoside triphosphate hydrolases superfamily, demonstrated on ATP.



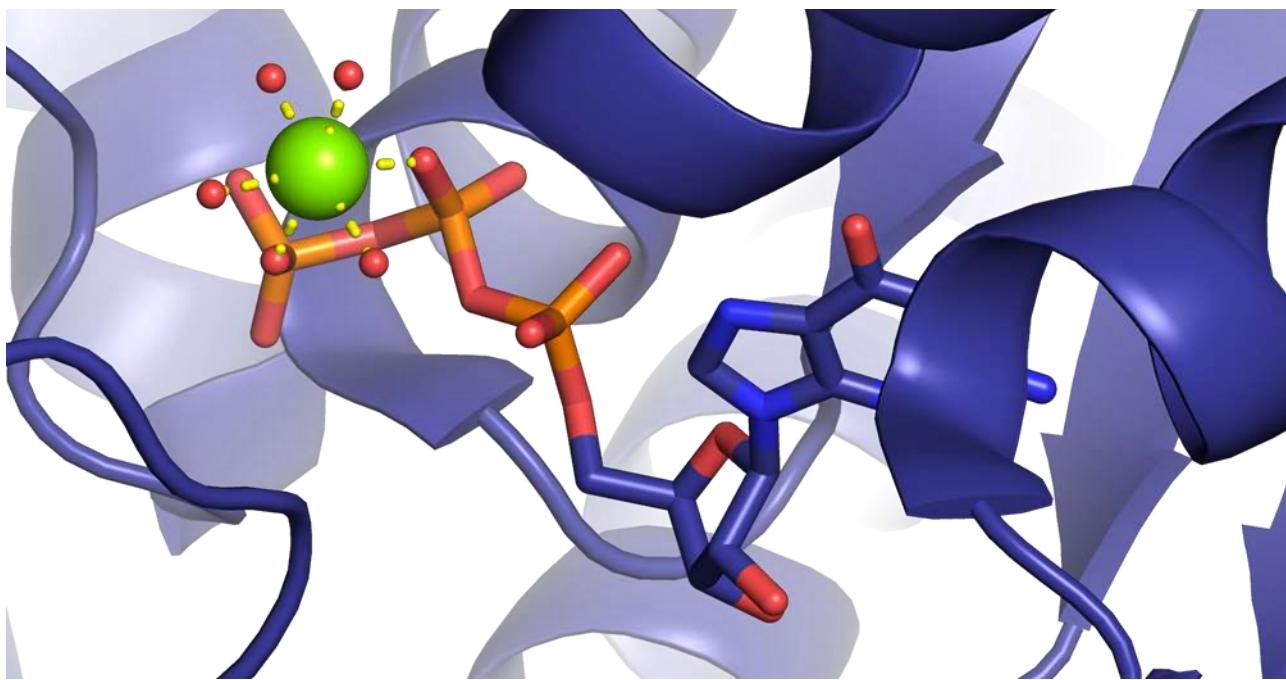
GTP and  $Mg^{2+}$  coordination in the representative structure 121P.

### Tubulin nucleotide-binding domain-like

There is a good agreement in the active site of the PDB structures of the Tubulin nucleotide binding domain superfamily. It is well-known GTPase [8], although not all tubulin isoforms hydrolyse the bound GTP.



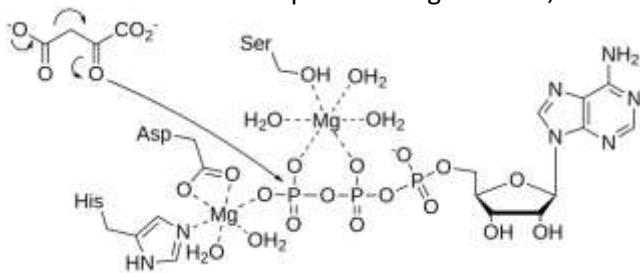
Typical coordination and reaction catalyzed by the members of the Tubulin nucleotide-binding domain-like superfamily



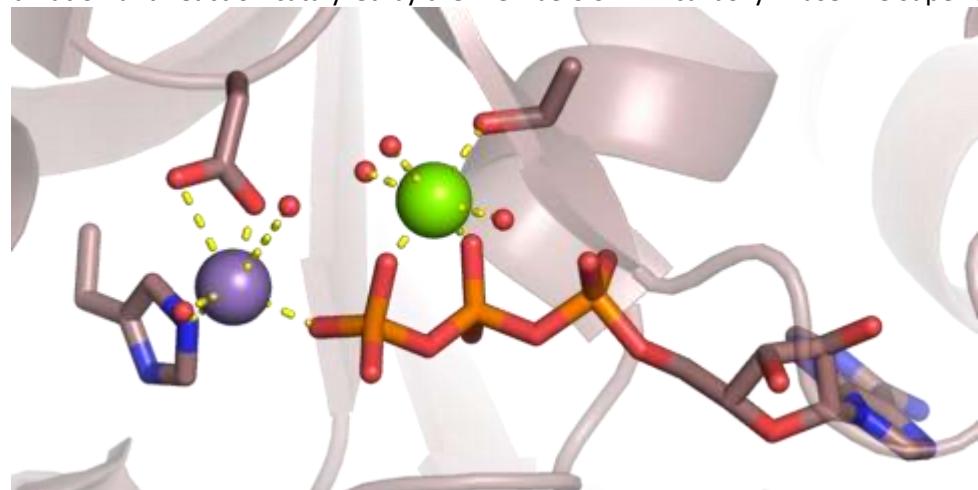
GTP and metal ion binding in the representative structure 7ALR.

### PEP carboxykinase-like

The fold of the PEP carboxykinase-like superfamily resembles the one of P-loop containing nucleoside triphosphate hydrolases, especially the p-loop and the helix it transitions into. However, the structures often contain a second metal ion. There are examples working with ATP, GTP and ITP.



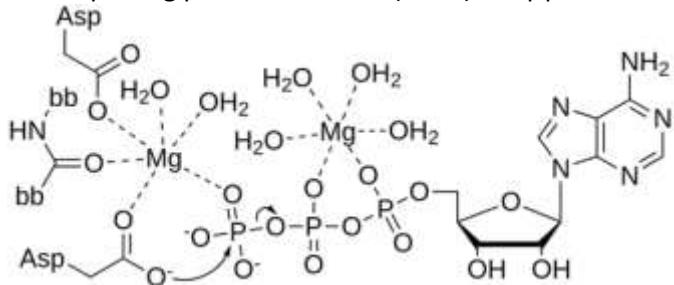
Typical coordination and reaction catalyzed by the members of PEP carboxykinase-like superfamily.



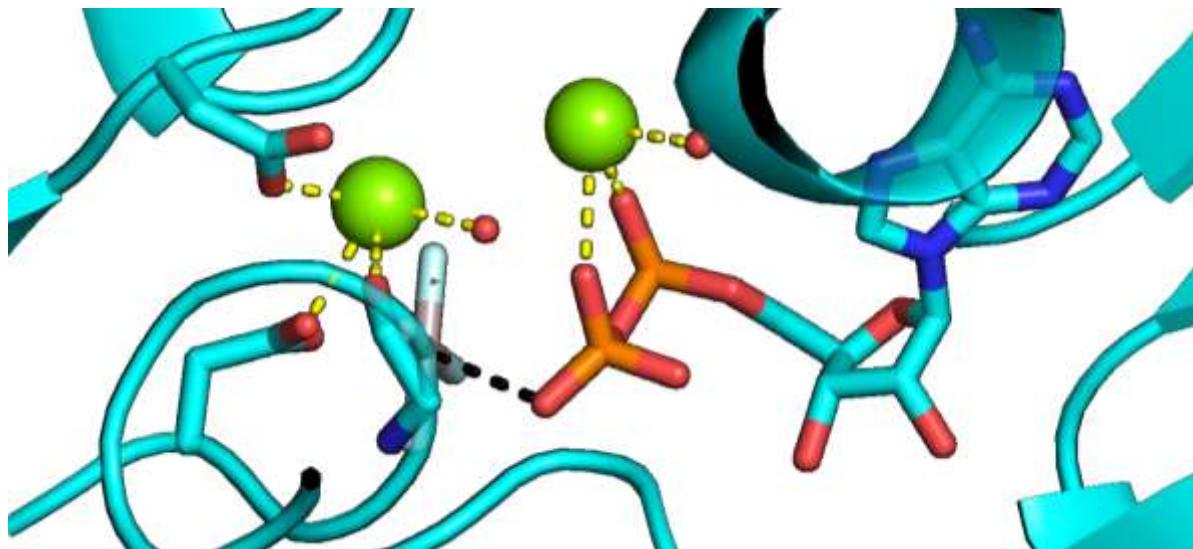
ATP and metal ion binding in the representative structure 2PY7.

### Metal cation-transporting ATPase, ATP-binding domain N

The most problematic coordination amongst all the phosphatases is presented by ATP dependent cation transport proteins. Some of the  $\text{Ca}^{2+}$  transport proteins have a single  $\alpha\beta\gamma$  coordinating ion but the nucleophile is far from the  $\gamma$  phosphate, others lack the  $\text{Mg}^{2+}$  pinch of the breaking phosphate bond. However, the only copper transporting protein structure (3A1C) has  $\beta\gamma$  coordination.



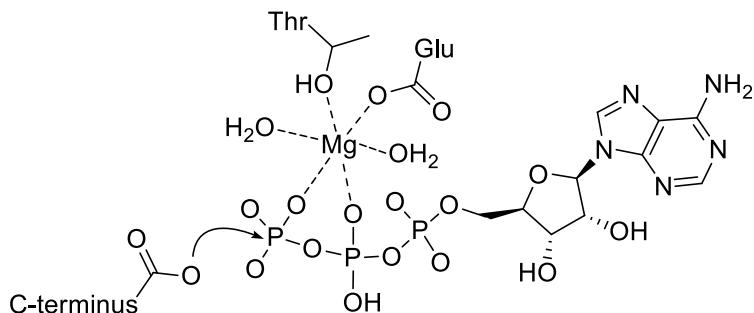
Assumed coordination of ATP in Metal cation transporting ATPases



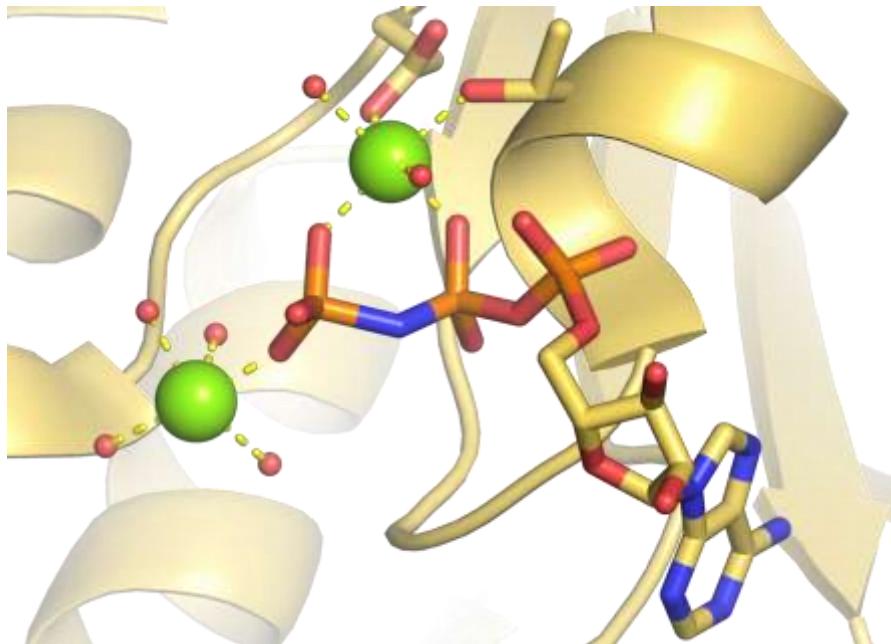
Metal ion coordination represented by the transition state mimicking crystal structure 3WGU.

### MurD-like peptide ligases, catalytic domain

The active site of this superfamily is also similar to the P-loop containing nucleoside triphosphate hydrolases', but the serine is replaced by a threonine and a glutamate is involved in the coordination. It uses ATP to activate carboxylates for peptide synthesis [9]. A second metal ion is also present in some structures, in a position it may be involved in nucleophile coordination. Notably, it is a member of the Ribokinase-like fold, similarly to the Ribokinase-like superfamily.



Typical coordination and reaction catalyzed by the members of MurD-like peptide ligases, catalytic domain superfamily.

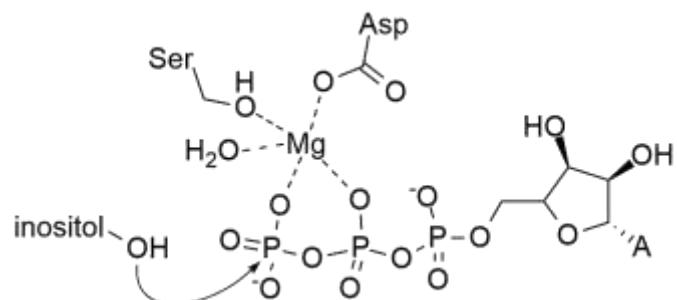


ATP and metal ion binding in the representative structure 6CAU.

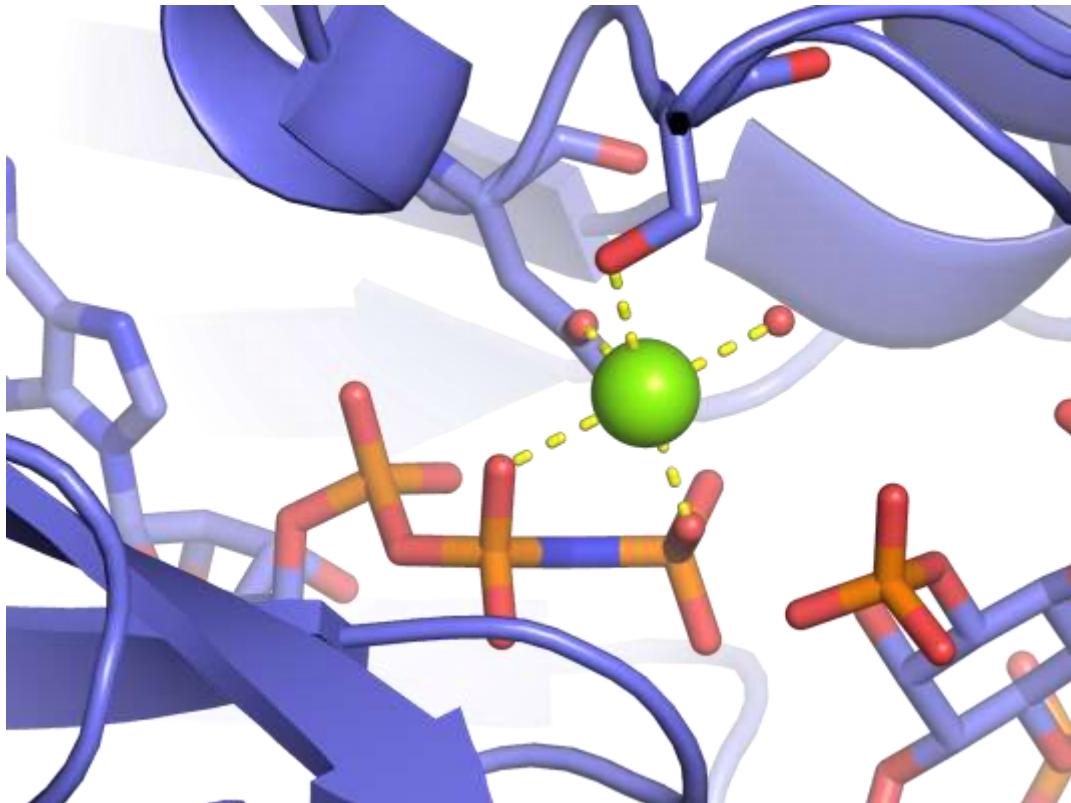
## Inositol-pentaphosphate 2-kinase

Adopted from InterPro entry IPR043001.

Clear phosphatase activity with  $\beta\gamma$  coordinated magnesium cation. The majority of the structures do not contain NTP, but they align very well, the SUPFAM analysis does not assign them to any known fold/superfamily.



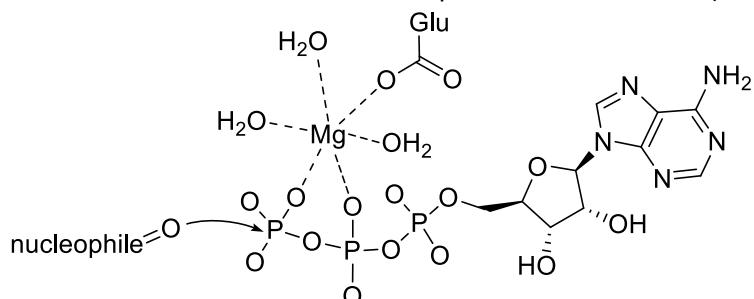
## Ion coordination and kinase activity templated on the structure 2XAN.



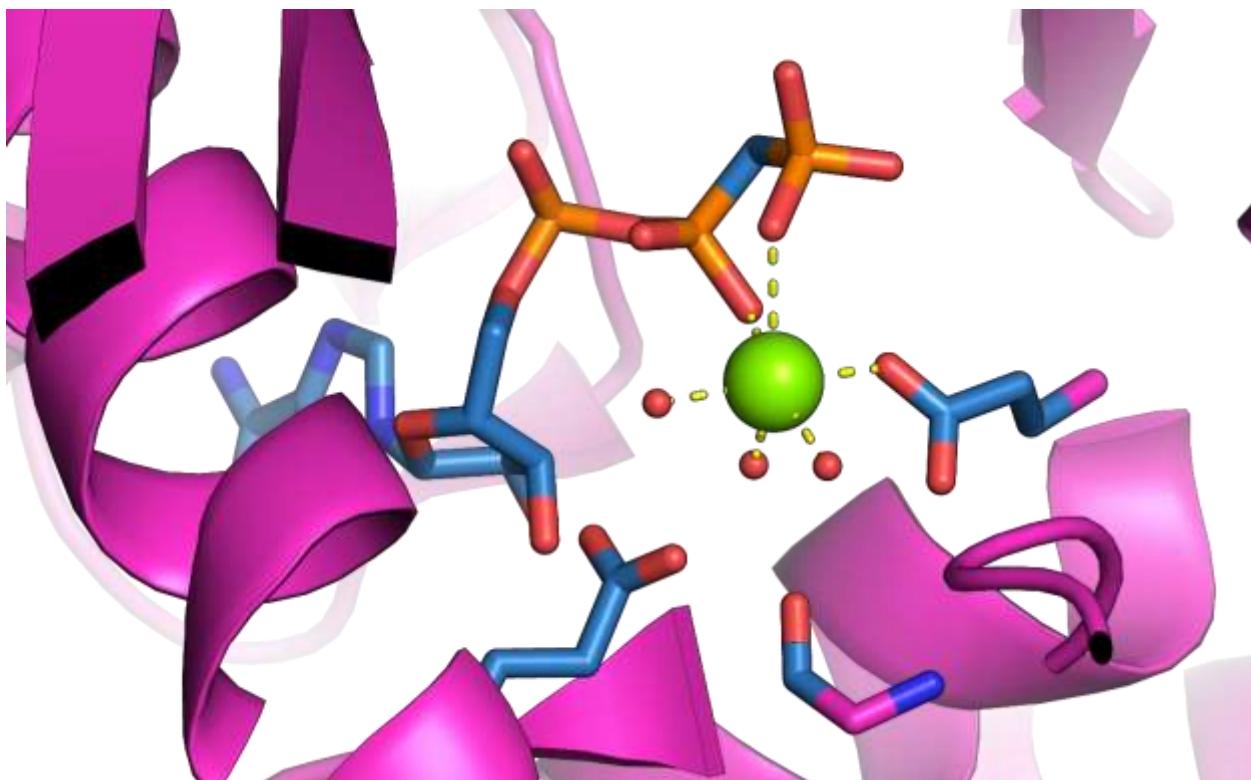
Incomplete ion coordination in the representative structure 2XAN.

#### **YcaO-related McrA-glycine thioamidation protein**

This category is defined on the InterPro family level (IPR017667 YcaO-related McrA-glycine thioamidation protein) extended by homologous YcaO structures. There is neither an associated EC to this family, nor to the YcaO structures. The proteins belonging to this family are universal in and restricted to methanogenic archaea [10] and share homology with YcaO (ribosomal protein S12 methylthiotransferase) and its homologues involved in ATP-dependent formation of heterocycles in thiazole/oxazole-modified peptide antibiotics [11]. Methyl-coenzyme M reductase (MCR) is an essential enzyme found strictly in methanogenic and methanotrophic archaea that catalyzes a reversible reaction involved in the production and consumption of the potent greenhouse gas methane [12]. In the corresponding active site, one  $Mg^{2+}$  ion is coordinated by the  $\beta$  and  $\gamma$  phosphates, as well as the carboxyl group of a glutamate residue, and an additional three crystal waters as found in the selected representative structure (PDB 6C17).



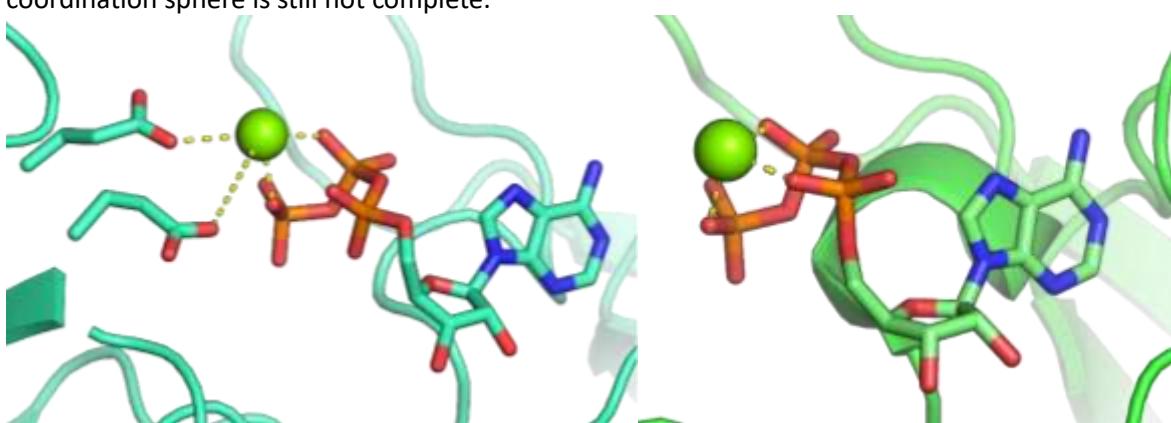
YcaO catalysed phosphorylation according to [11]



ATP and ion coordination in the representative structure 6CI7.

#### Peptidase G2, IMC autoproteolytic cleavage domain

There are 2 PDBs originating from the same publication we identified for this category, which is defined on the InterPro domain level (IPR021865 Peptidase G2, IMC autoproteolytic cleavage domain). There is no associated EC to this domain or to the structures of the pre-mature bacteriophage phi29 gene product 12. Yet, the authors describe that the autocleavage of the C-terminal domain is a post-trimerization event that is followed by a unique ATP-dependent release. In the active site, there are two glutamic acid residues that may be involved in the ATP hydrolysis [13]. Yet, based on the available two structures, a clear coordination could not be established, we are inclined to associate a BG(-) coordination. The PDB 3sucA structure displays an  $\alpha\beta\gamma$ -coordination, yet the coordination sphere around the  $Mg^{2+}$  is not complete. In PDB 3gqkA, it is slightly shifted towards the  $\gamma$  phosphate with a  $\beta\gamma$ -coordination, and the two glutamate residues approach the coordination sphere. However, their distance for the coordination is not optimal and the coordination sphere is still not complete.

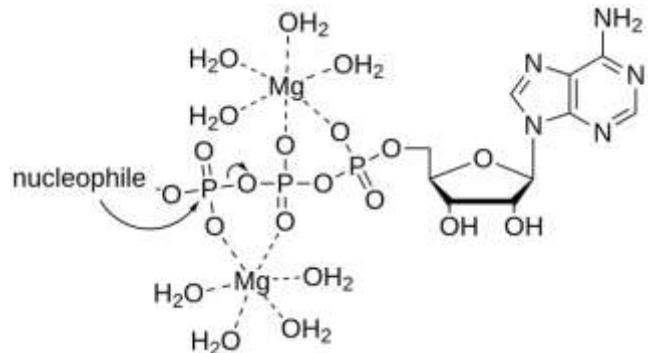


ATP-ion site in structures 3GQK (left) and 3SUC (right).

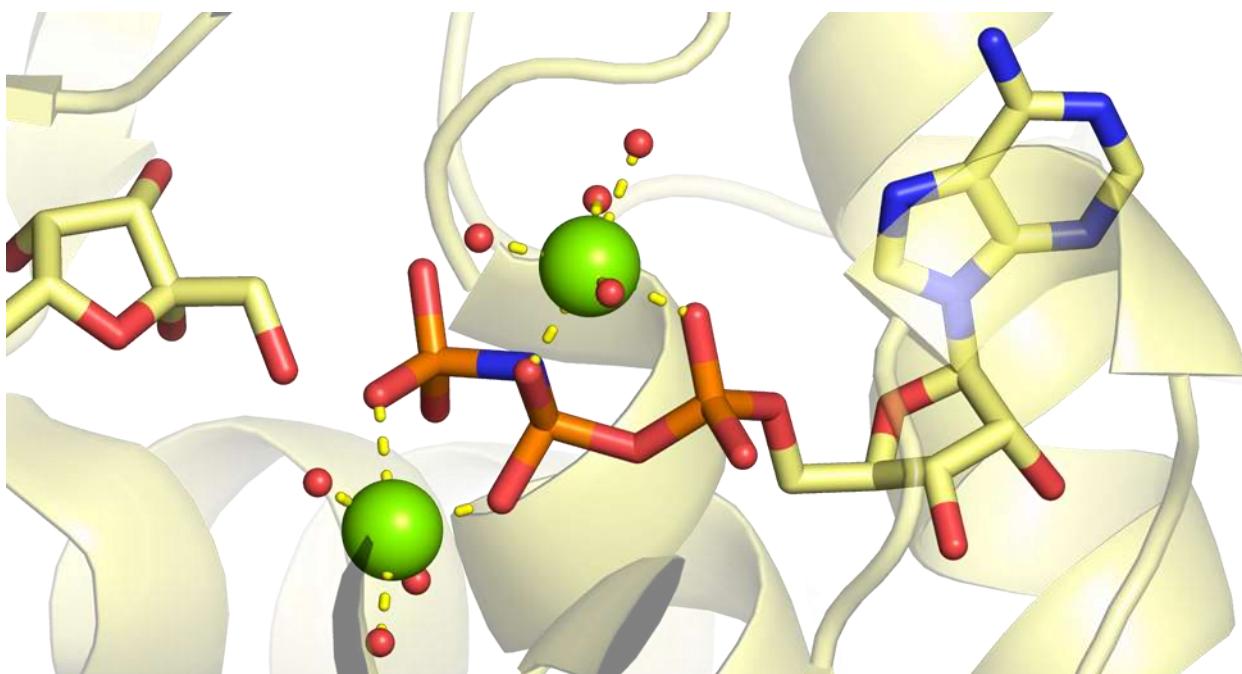
Multiple metal phosphatases with  $\beta$  coordination on the (-) side

#### Ribokinase-like

The Ribokinase-like active site contains a second metal ion in  $\alpha\beta$  coordination in addition to the pinching cation [14-16].



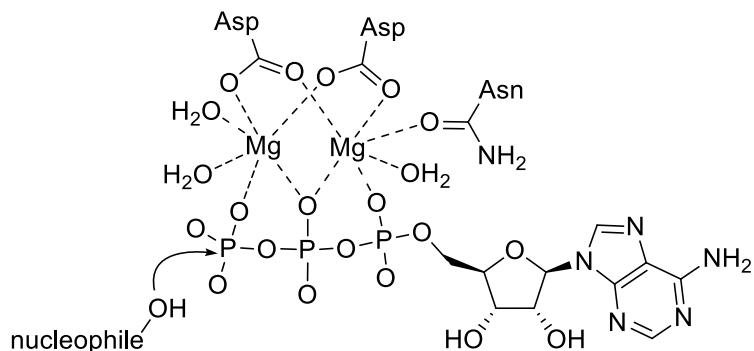
Typical coordination and reaction facilitated by the Ribokinase-like superfamily.



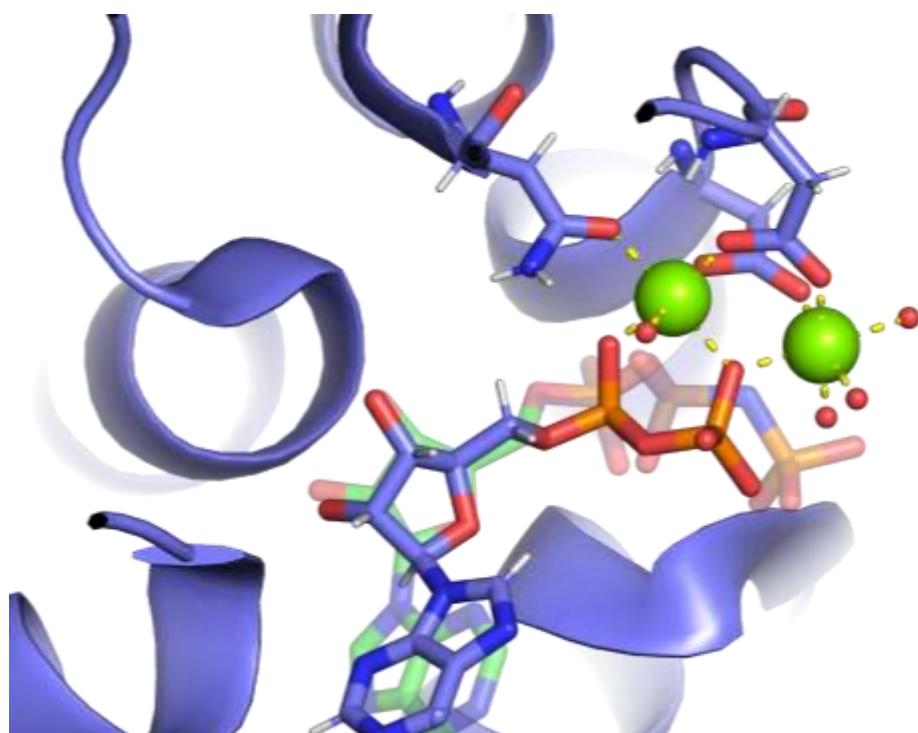
ATP and substrate binding represented by 2JG1.

### DhaL-like

A set of specific kinases are assigned to the same *Dhal-like* superfamily, only one structure is present with ATP and poor two metal ion coordination. Both ions remain in the ADP state, we assume an  $\alpha\beta$  and a  $\beta\gamma$  coordination.



Proposed coordination scheme and catalyzed reaction of Dhal-like kinases.

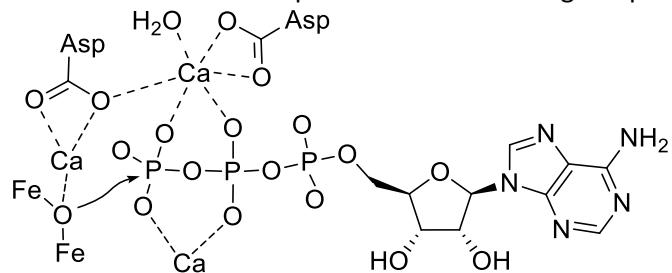


Metal ion coordination in the representative structure 7RM7 with ADP, ATP analog overlay is depicted from 1UN9 in green.

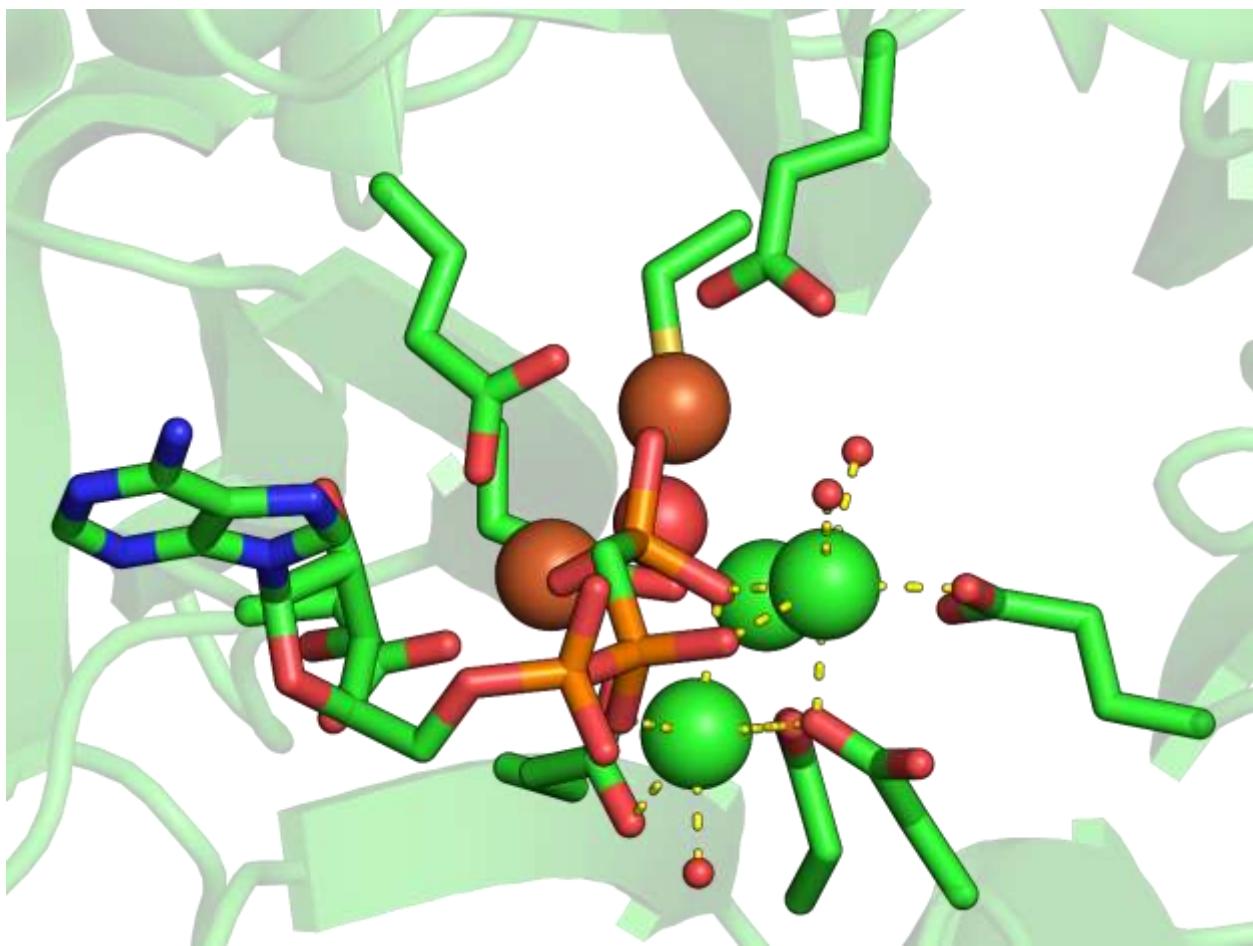
### Calcium-dependent phosphotriesterase

Within this superfamily, we have identified an InterPro family (IPR008557 Alkaline phosphatase PhoX), the corresponding structures of which can cleave ATP [17], while the rest of the superfamily does not process NTP, but these are phosphatases hydrolyzing phosphate ester ligands (EC 3.1.3.1). In the selected representative structure (PDB 4AMF), a nonhydrolyzable analog (AMP-PCP) of the substrate ATP is present. The enzyme contains a complex active-site cofactor comprising two antiferromagnetically coupled ferric iron ions ( $Fe^{3+}$ ), three calcium ions ( $Ca^{2+}$ ), and an oxo group bridging three of the metal ions [17]. A  $Ca^{2+}$  is on each side of the triphosphate chain, both in a  $\beta\gamma$ -coordinated position. Both are

further coordinated by one glutamate and one aspartate residues, and crystal waters. In the proposed mechanism, the phosphate cleavage is initially carried out by the iron oxide moiety, which is recovered by using a water from the calcium coordination sphere to form the inorganic phosphate.



Proposed phosphatase mechanisms in PhoX [17]. Many coordinating residues are omitted for clarity.

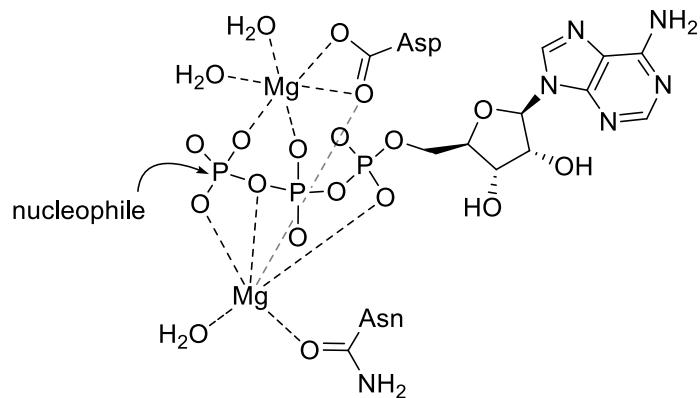


ATP and metal ion coordination in the representative structure 4AMF.

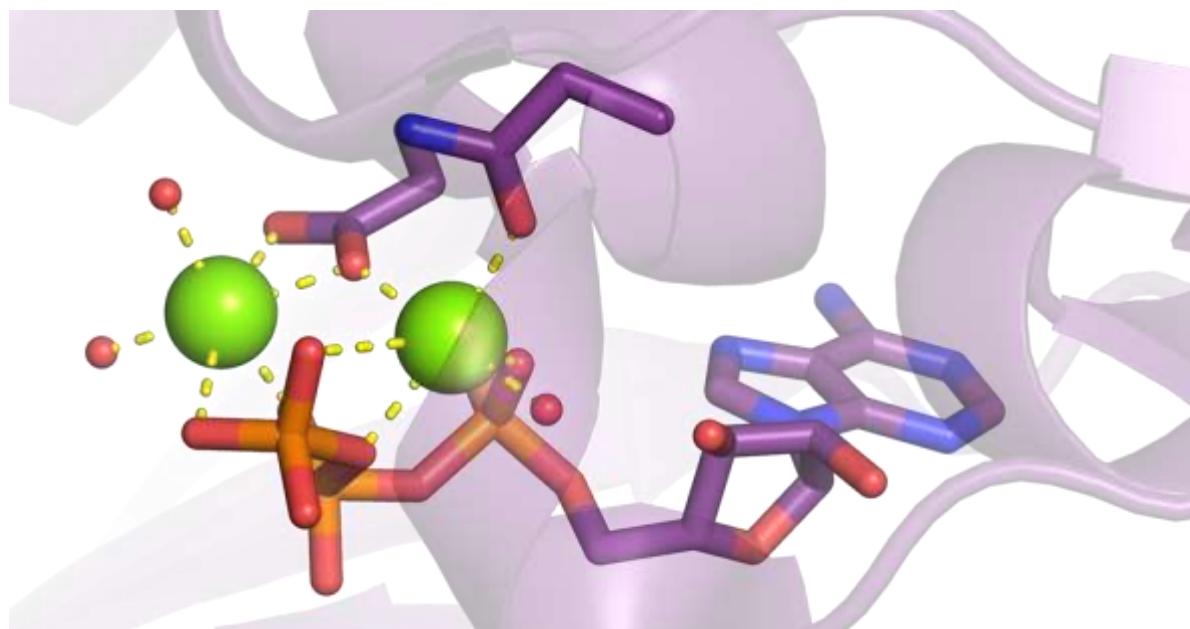
In the following superfamilies, ions feature in BG() position alongside an additional AG ion bridging the  $\alpha$  and  $\gamma$  phosphates (Fig. 5 in the main text).

### Protein kinase-like (PK-like)

There is a variation in the ion coordination within the PK-like superfamily, there are structures missing one ion or the other from the two established positions. In addition, occasional examples are also identified with slightly different active site ion placement, despite being restricted to high resolution structures.



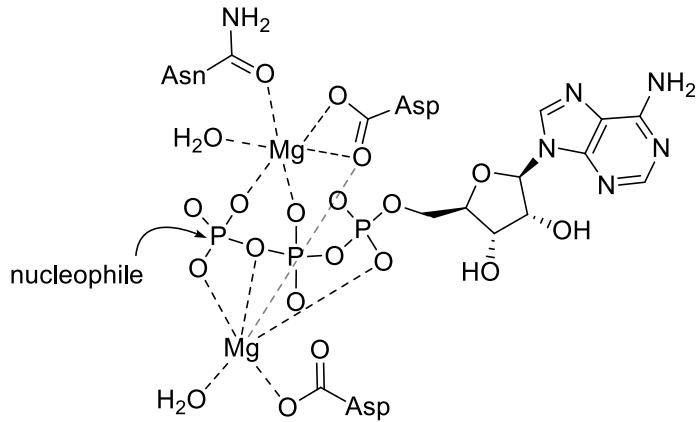
Kinase reactivity and coordination amongst the members of the PK-like superfamily.



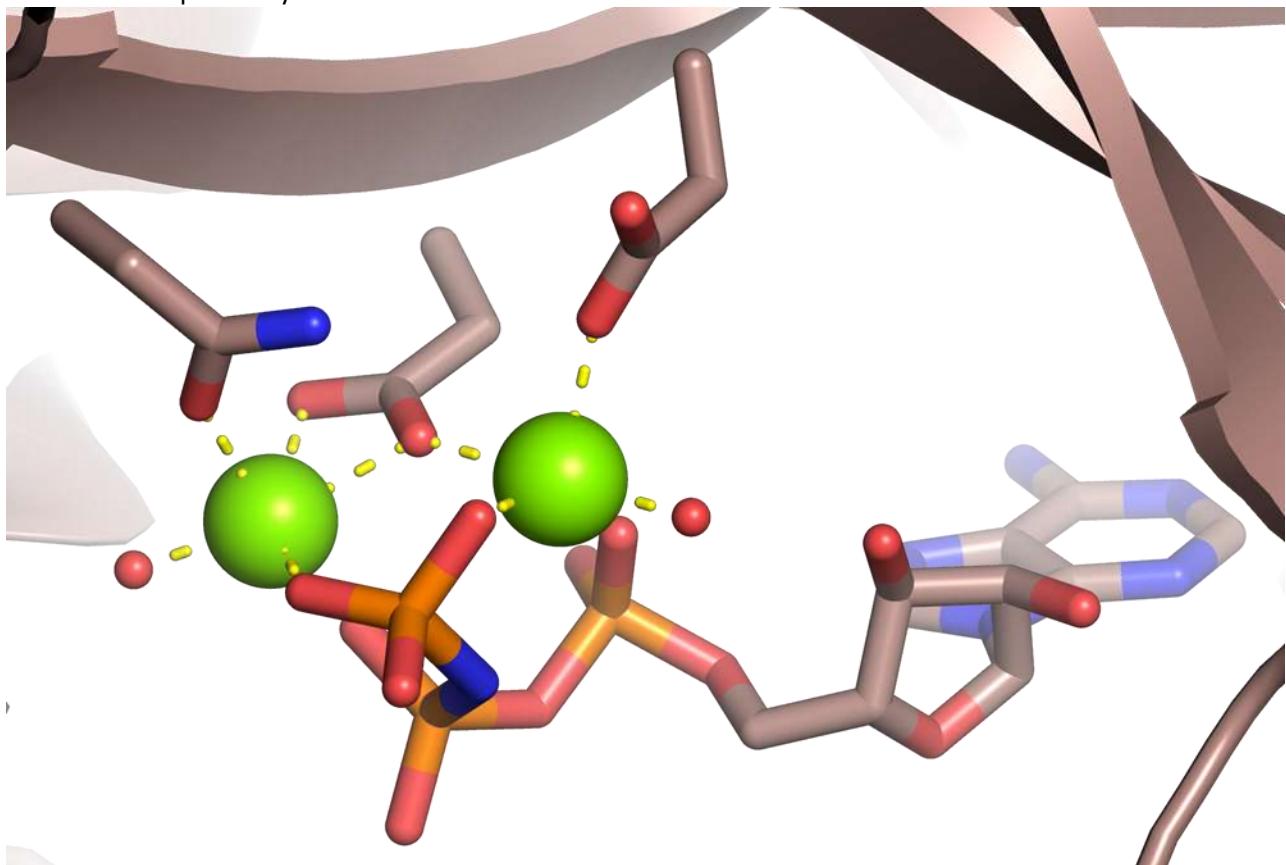
ATP and metal ion binding in the representative structure 3DNT.

### Glutathione synthetase ATP-binding domain-like

There is a good agreement in the coordination of the two magnesium ions in the entire superfamily. The secondary structure around the active site is similar to the SAICAR synthase-like superfamily. The overall architecture is similar to the PK-like superfamily, this superfamily mostly covers ligases creating peptide modes, therefore the phosphate cleavage happens with a carboxylate nucleophile.



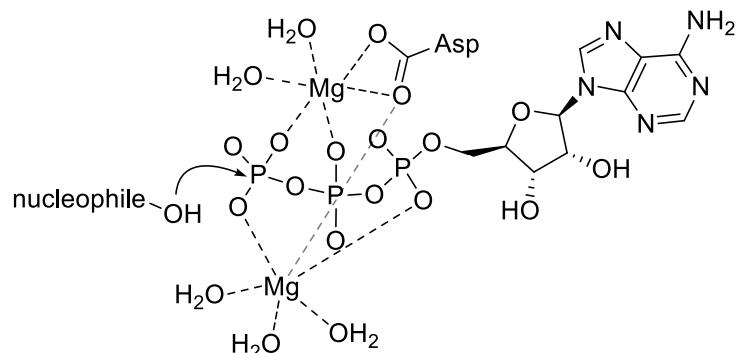
Typical coordination and reaction catalyzed by the members of Glutathione synthetase ATP-binding domain-like superfamily.



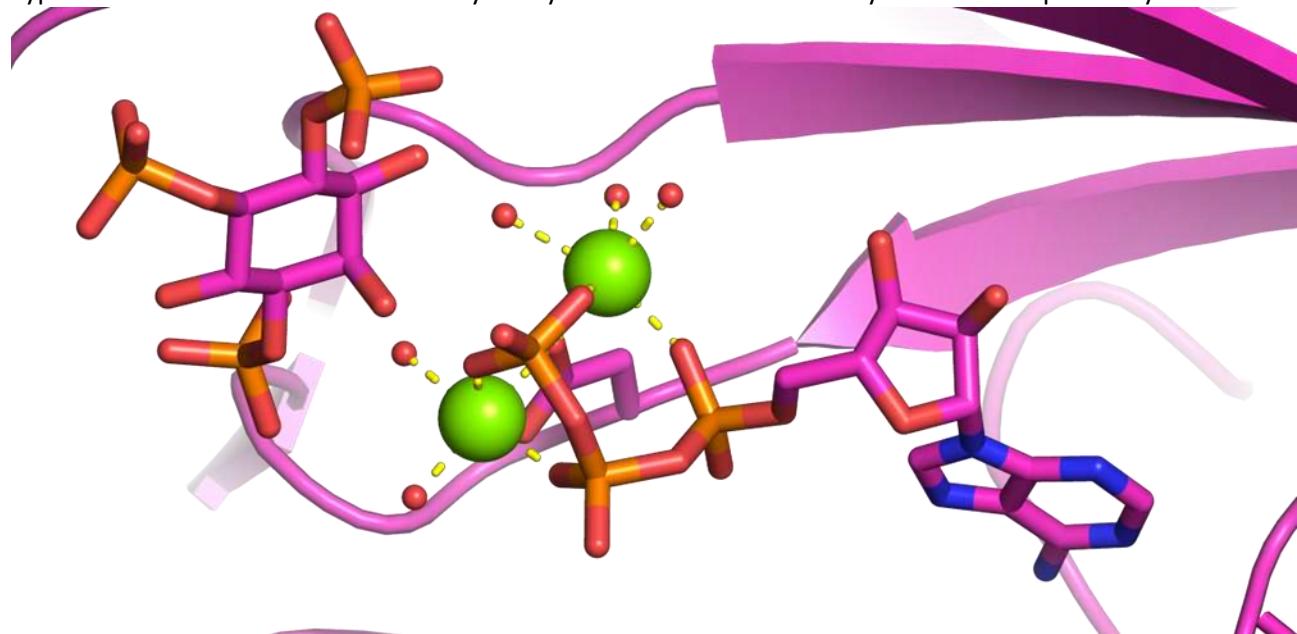
ATP and metal ion binding in the representative structure 5DGH.

### **SAICAR synthase-like**

As mentioned above, both the fold and the coordination is quite similar to the Glutathione synthetase ATP-binding domain-like superfamily, some of the proteins perform the same reaction according to the EC assignments, while mostly SAICAR synthase-like members are involved in phosphorylating inositol derivatives.



Typical coordination and reaction catalyzed by the members of SAICAR synthase-like superfamily.

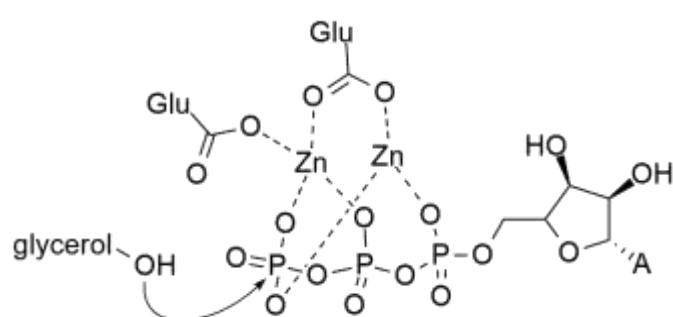


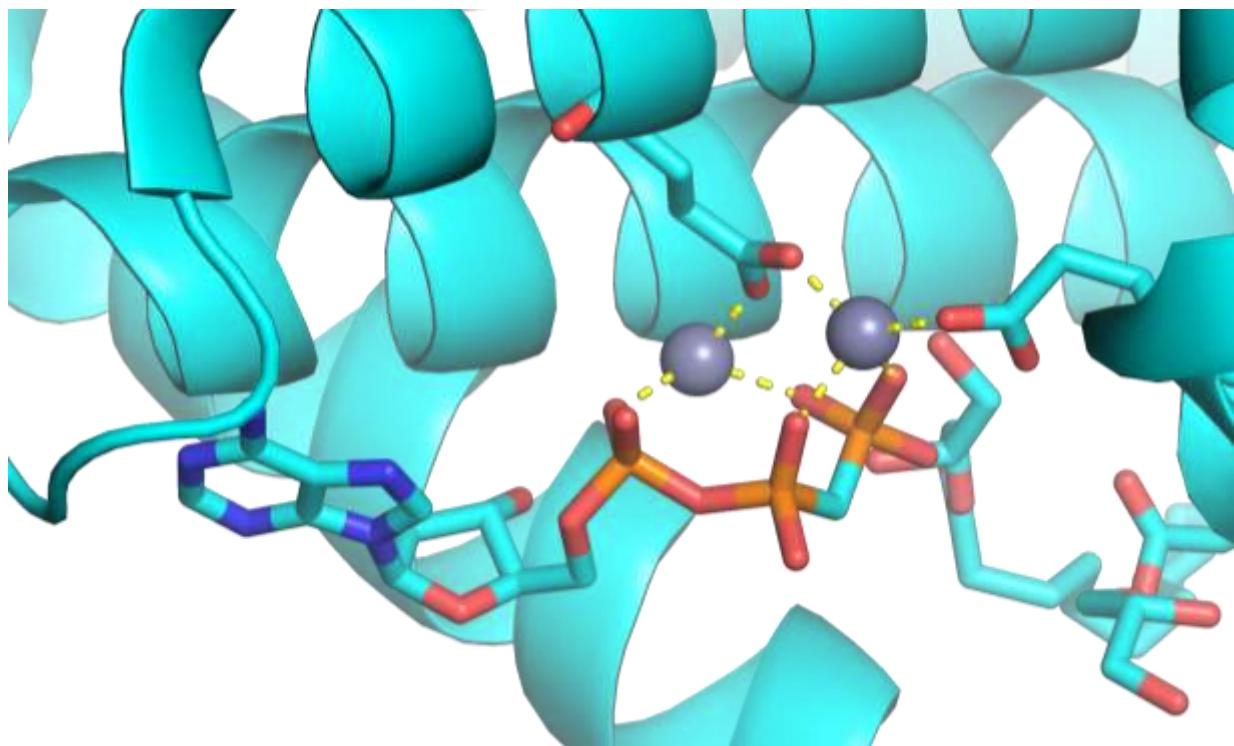
Substrate and metal ion binding in the representative structure 2AQX.

### Diacylglycerol kinase (DgkA)-like

While many structures are categorized into the *NAD kinase/diacylglycerol kinase-like* superfamily, there are two ATP bound membrane proteins, which contain zinc ions and are not assigned by SUPFAM. While magnesium is a common ATP counter-ion in kinases, zinc can serve this role in DgkA [18].

Zinc coordination and glycerol kinase reactivity scheme for the DgkA-like superfamily.

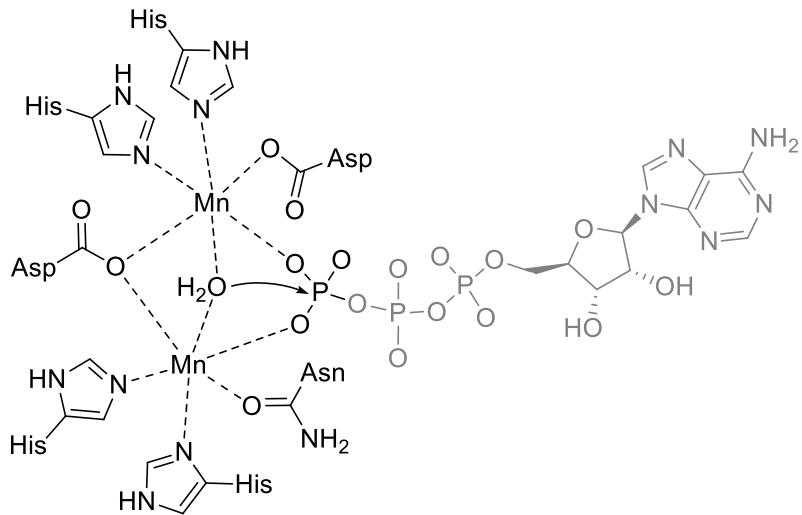




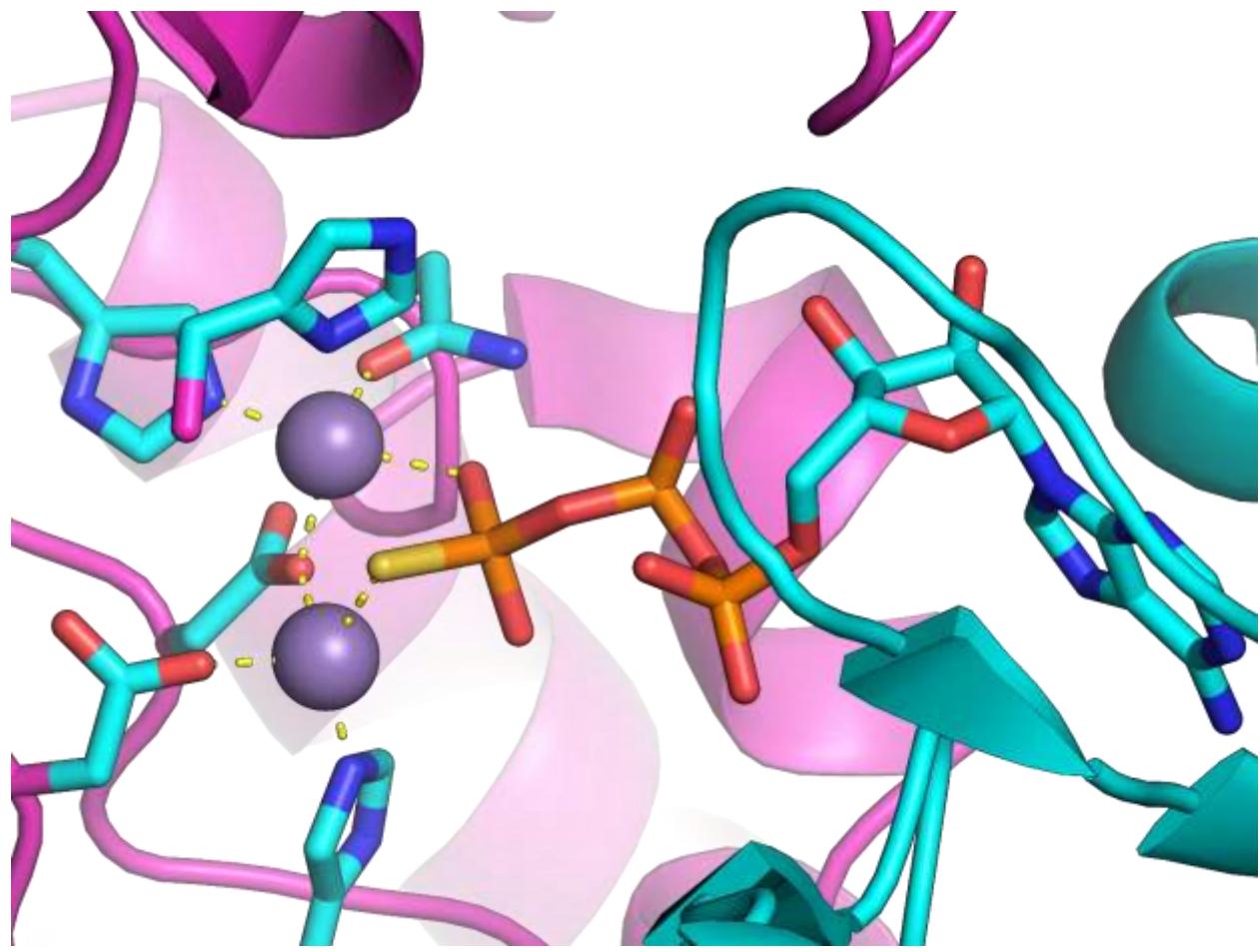
Zinc and ATP binding in the selected DgkA kinase 4UXX.

### Metallo-dependent phosphatases

In the active site, there are two metal ions, usually manganese, iron, or zinc, both coordinated by a cage of histidine, aspartate, and asparagine residues [19]. They are general phosphatases, but may also work on NTPs.



Ion coordination and phosphatase reactivity based on the structure 4LAC.



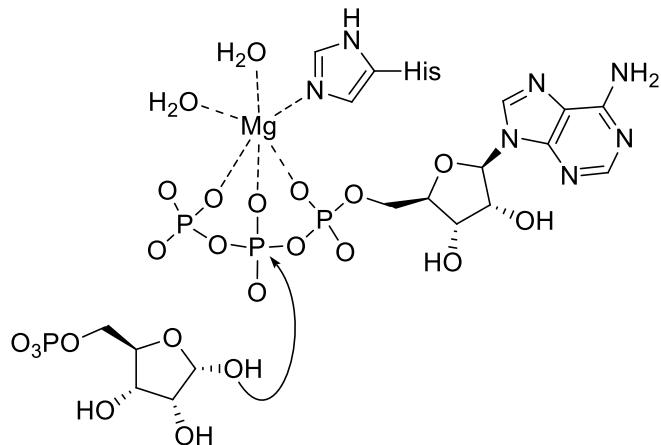
Ion and NTP coordination in the structure 4LAC.

## Pyrophosphatase superfamilies

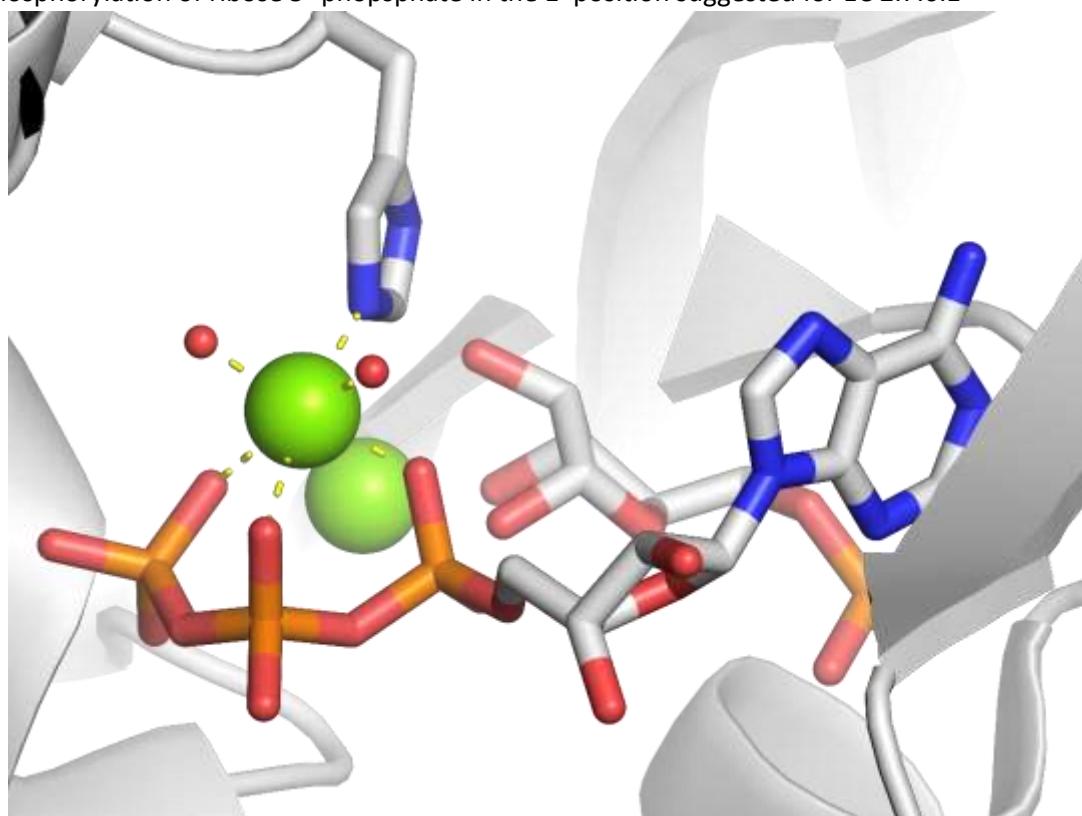
### Pyrophosphatases with $\alpha\beta\gamma$ coordination on the (+) side

#### PRTase-like

Enzymes belonging to EC 2.7.6.1 are phosphoribosyl pyrophosphate synthetases (PRPS), and are key regulators of nucleotide metabolism [20]. For the synthesis of phosphoribosyl pyrophosphate they use ATP. As co-factor, an  $\alpha\beta\gamma$ -coordinated  $Mg^{2+}$  ion is present at the active site. Interestingly, a histidine residue also participates in the  $Mg^{2+}$ -coordination in this superfamily. Two further water molecules also coordinate the  $Mg^{2+}$ .



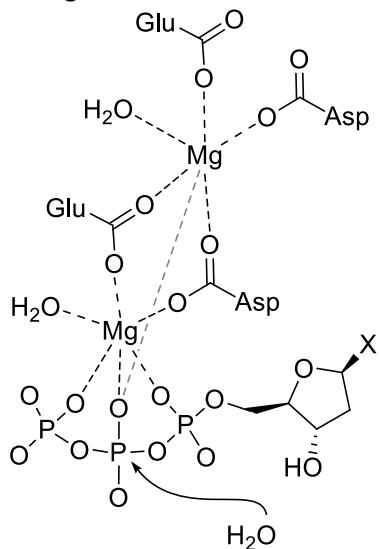
Pyrophosphorylation of ribose 5'-phosphate in the 1' position suggested for EC 2.7.6.1



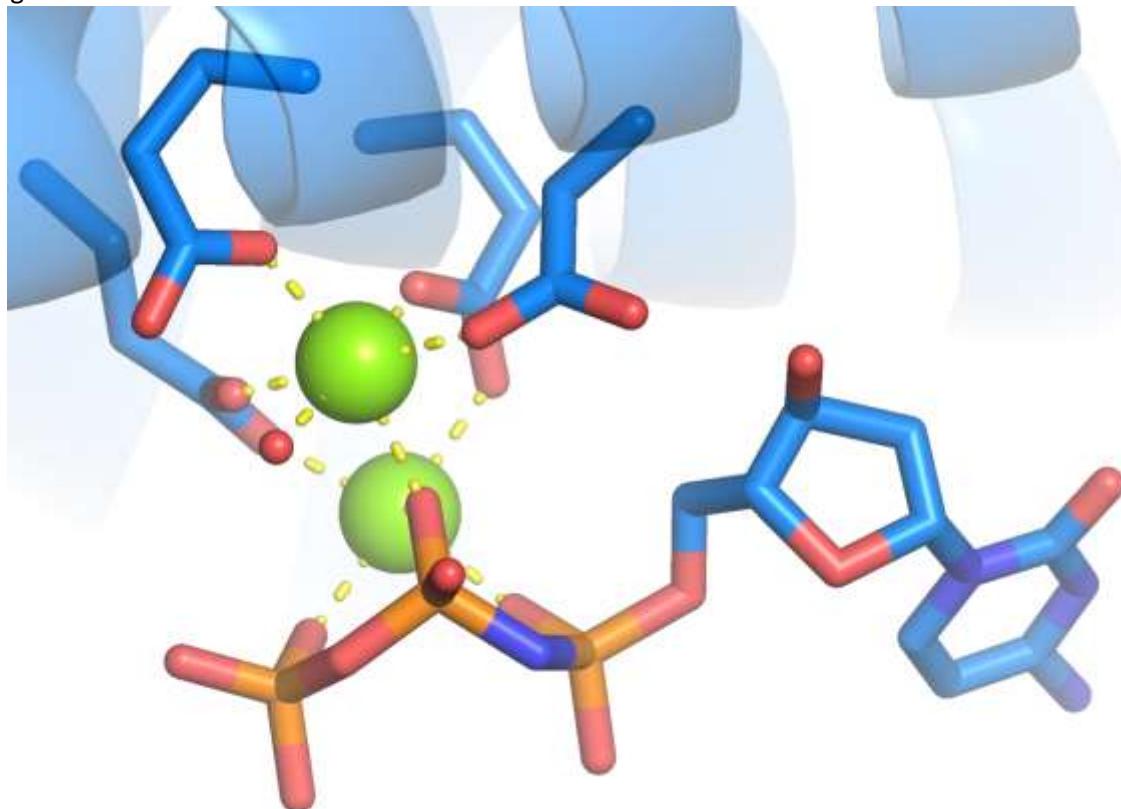
ATP and metal ion coordination in the representative structure 8DBJ.

### All-alpha NTP pyrophosphatases

This fold has 2 catalytic metal ions. One  $Mg^{2+}$  ion is coordinated by the  $\alpha\beta\gamma$  phosphate oxygens, while the other  $Mg^{2+}$  ion by only a  $\beta$  phosphate oxygen atom. Two aspartate and two glutamate residues are also involved in the coordination of the two  $Mg^{2+}$  ions.



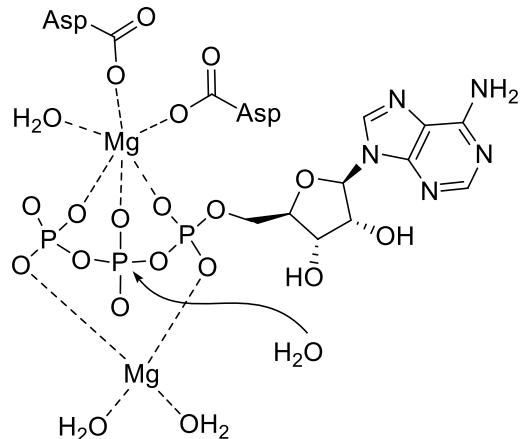
Metal ion coordination and pyrophosphatase reaction based on the dCTPase structure 6SQZ. While the assigned PDB structures are clearly pyrophosphatases, many sequences from this superfamily are linked to the EC 3.6.1.8, which has both phosphatase and pyrophosphatase activity associated and considered ambiguous.



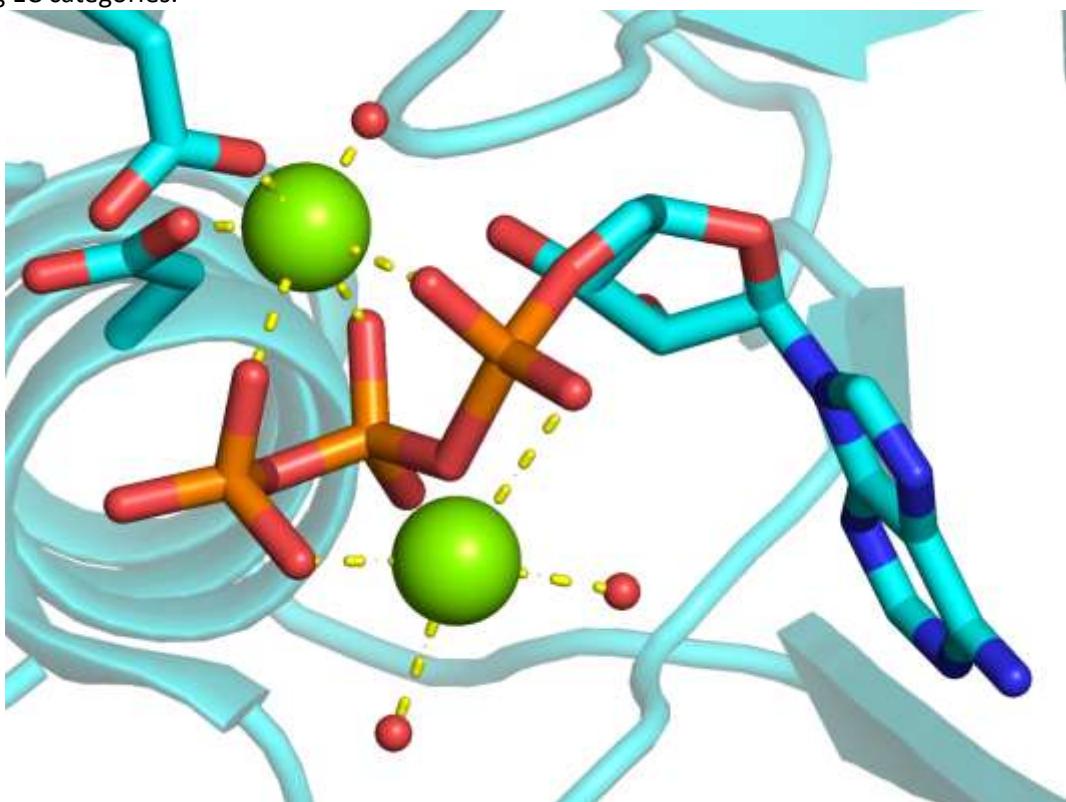
Metal ion and dCTP binding in the representative structure 6SQZ.

### Adenine nucleotide alpha hydrolases-like

This superfamily has members which are stress proteins and do not necessarily catalyze ATP hydrolysis. The group of stress proteins bind a single  $Mg^{2+}$  ion which is octahedrally coordinated by the  $\alpha\beta\gamma$  phosphate oxygen atoms and three additional water molecules [21, 22]. The rest of the members are enzymes forming carbon-nitrogen bonds and have 2 metal ions. One  $Mg^{2+}$  ion is coordinated by the  $\alpha\beta\gamma$  phosphates, whereas the other  $Mg^{2+}$  ion shows an  $\alpha\gamma$  phosphate coordination. Two aspartate residues are also involved in the coordination of the  $Mg^{2+}$  ion at the  $\alpha\beta\gamma$  position.



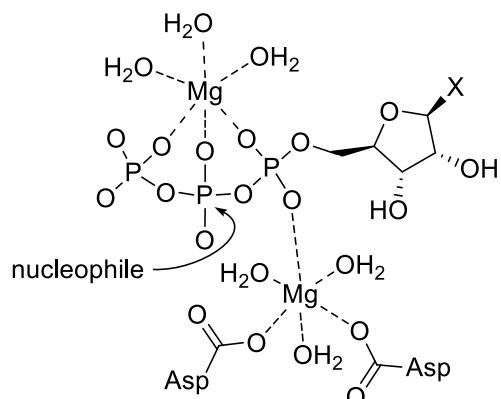
$Mg^{2+}$  coordination and ATP pyrophosphatase reactivity. This superfamily has members from many AMP forming EC categories.



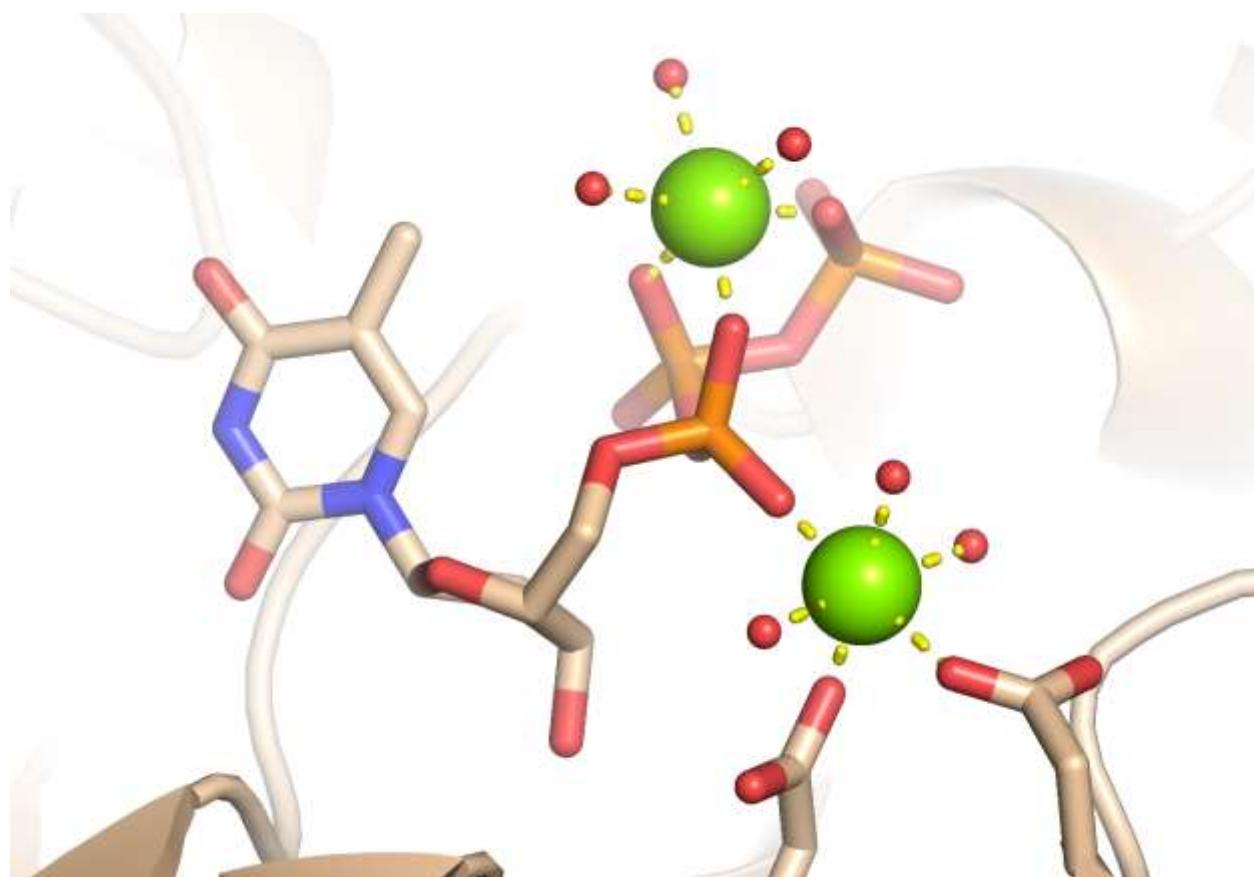
ATP and metal ion coordination in the representative structure 1MB9.

### Nucleotide-diphospho-sugar transferases

Some of the UDP-N-acetylglucosamine diphosphorylases (EC 2.7.7.23) of this fold also function as glucosamine-1-phosphate N-acetyltransferases (EC 2.3.1.157). One  $Mg^{2+}$  ion has a clear  $\alpha\beta\gamma$  phosphate coordination. A second  $Mg^{2+}$  ion is present close to the  $\alpha$  phosphate in some of the structures, however its role could be related to the correct positioning of the nucleotide and promoting the attack of the ligand hydroxyl group, it may not be directly involved in the enzymatic reaction. The  $Mg^{2+}$  ion which is coordinated by an  $\alpha$ -phosphoryl oxygen, is further coordinated by the side chains of two aspartate residues in its vicinity.



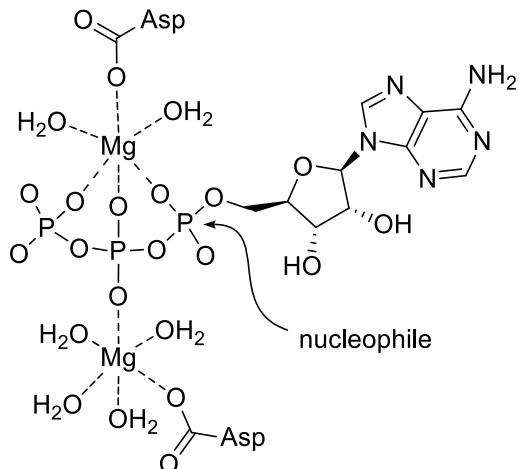
Nucleotide-diphospho-sugar transferases are assigned to many nucleotidyltransferase EC (2.7.7.-) and consequently thought to work with multiple NTPs and nucleophiles.



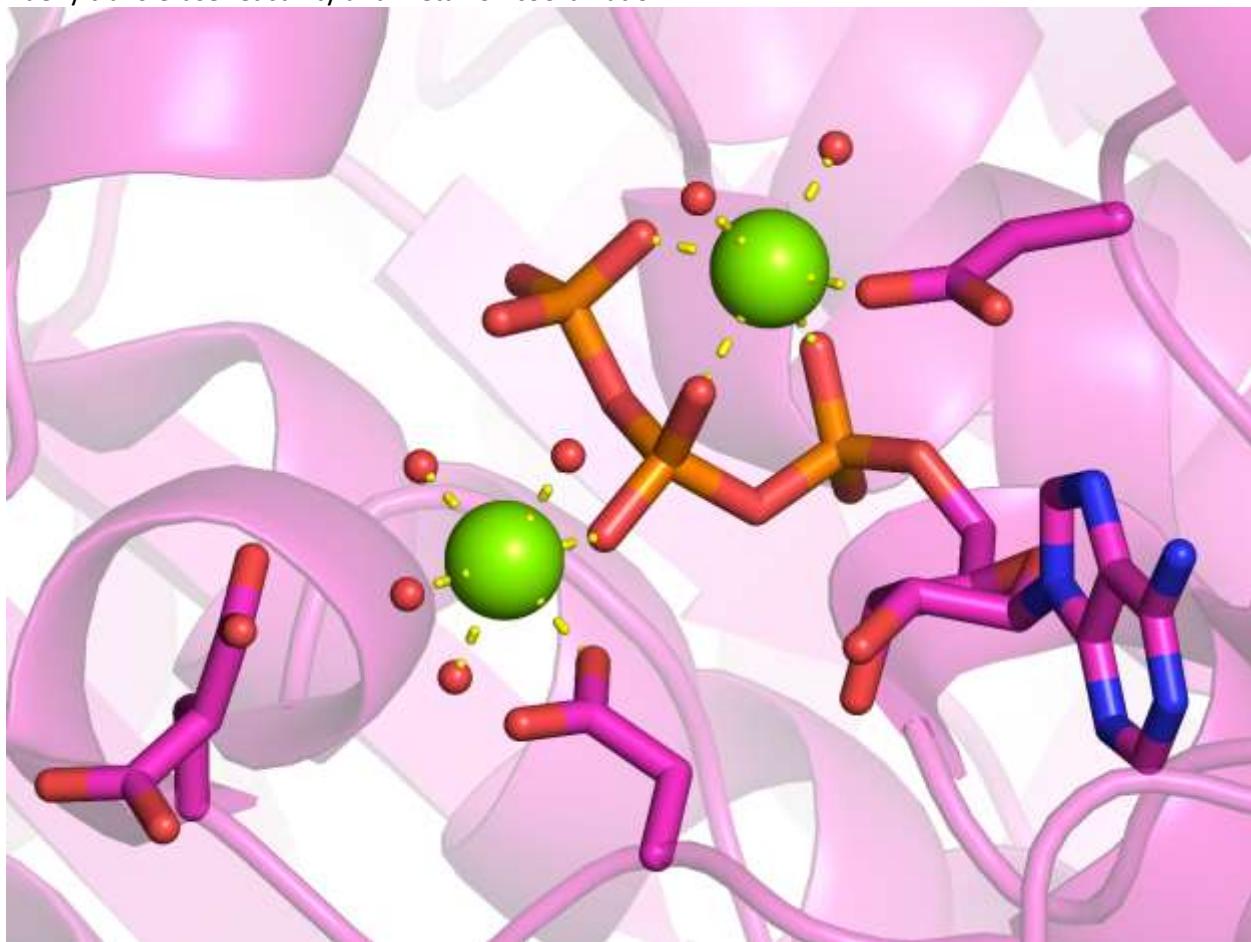
Metal ion and NTP coordination in the representative structure 6B5K.

### Activating enzymes of the ubiquitin-like proteins

This superfamily has two metal ions. One  $Mg^{2+}$  ion is coordinated by the  $\alpha\beta\gamma$  phosphate oxygens, and another one by only a  $\beta$  phosphate oxygen atom. Both  $Mg^{2+}$  ions are further coordinated by a conserved aspartate residue each, and the  $Mg^{2+}$  ion in the  $\beta$  position is further stabilized by a conserved glutamate residue through coordinating water molecules.



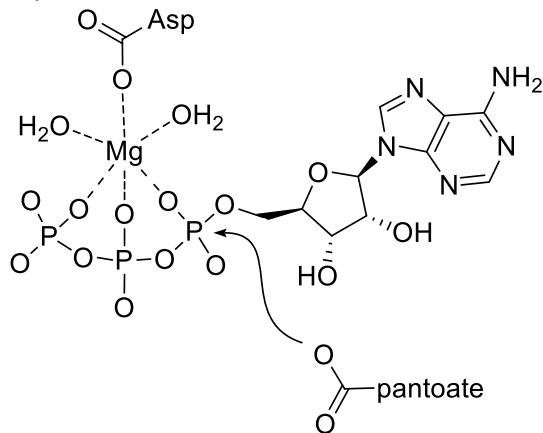
Adenyltransferase reactivity and metal ion coordination.



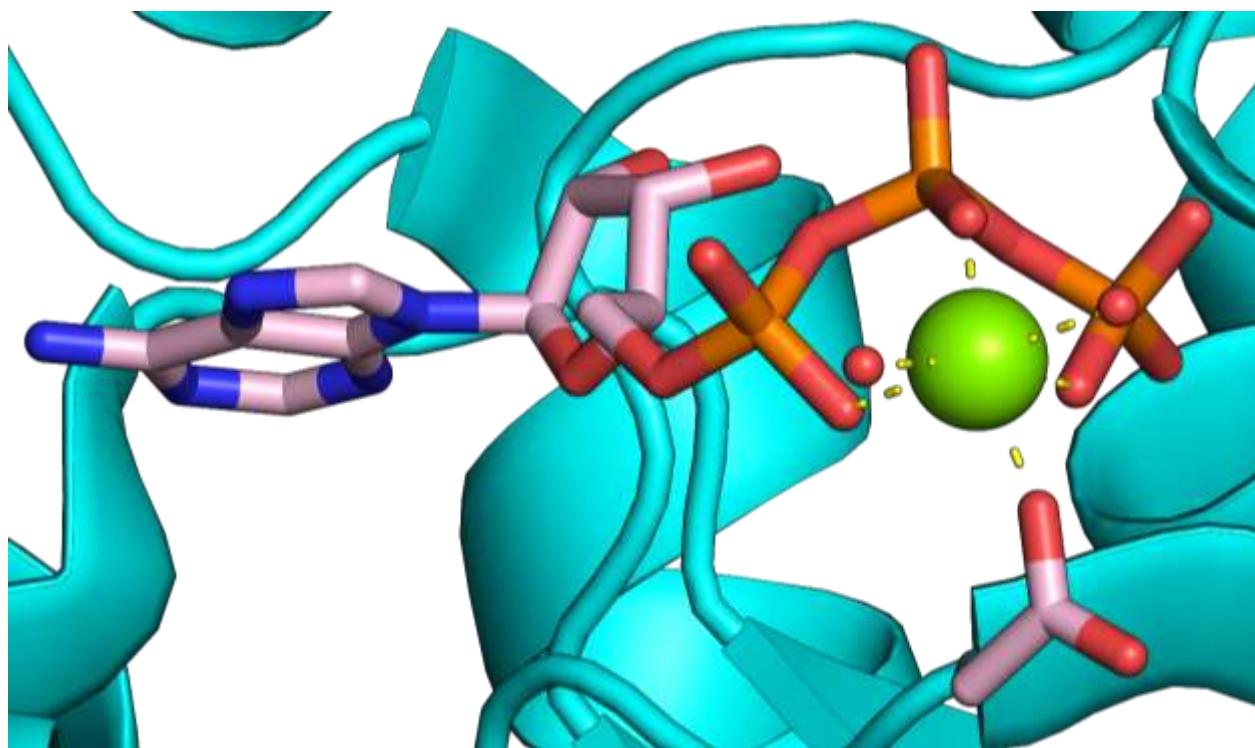
ATP and metal ion coordination in the representative structure 6H77.

### Phosphopantoate/pantothenate synthetase superfamily

There is no corresponding SUPFAM superfamily, this category is named after the InterPro superfamily (IPR038138). The enzymes correspond to 4-phosphopantoate-beta-alanine ligases, also known as phosphopantothenate synthetases (EC 6.3.2.36), catalyzing the conversion of (R)-4-phosphopantoate and beta-alanine to 4'-phosphopantetheine in a two-step reaction, first of which is releasing a pyrophosphate [23] and it is a part of the CoA biosynthesis pathway [24]. The selected representative structure (PDB 3WDL) displays an  $\alpha\beta\gamma$ -coordinated  $Mg^{2+}$  at the active site, which is further coordinated by an aspartate residue and a crystal water.



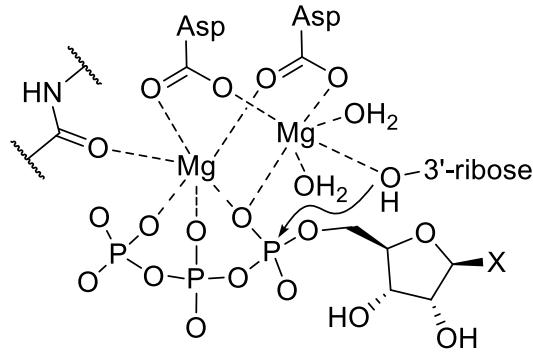
Pantoate carboxyl activation [23] and schematic  $Mg^{2+}$  coordination.



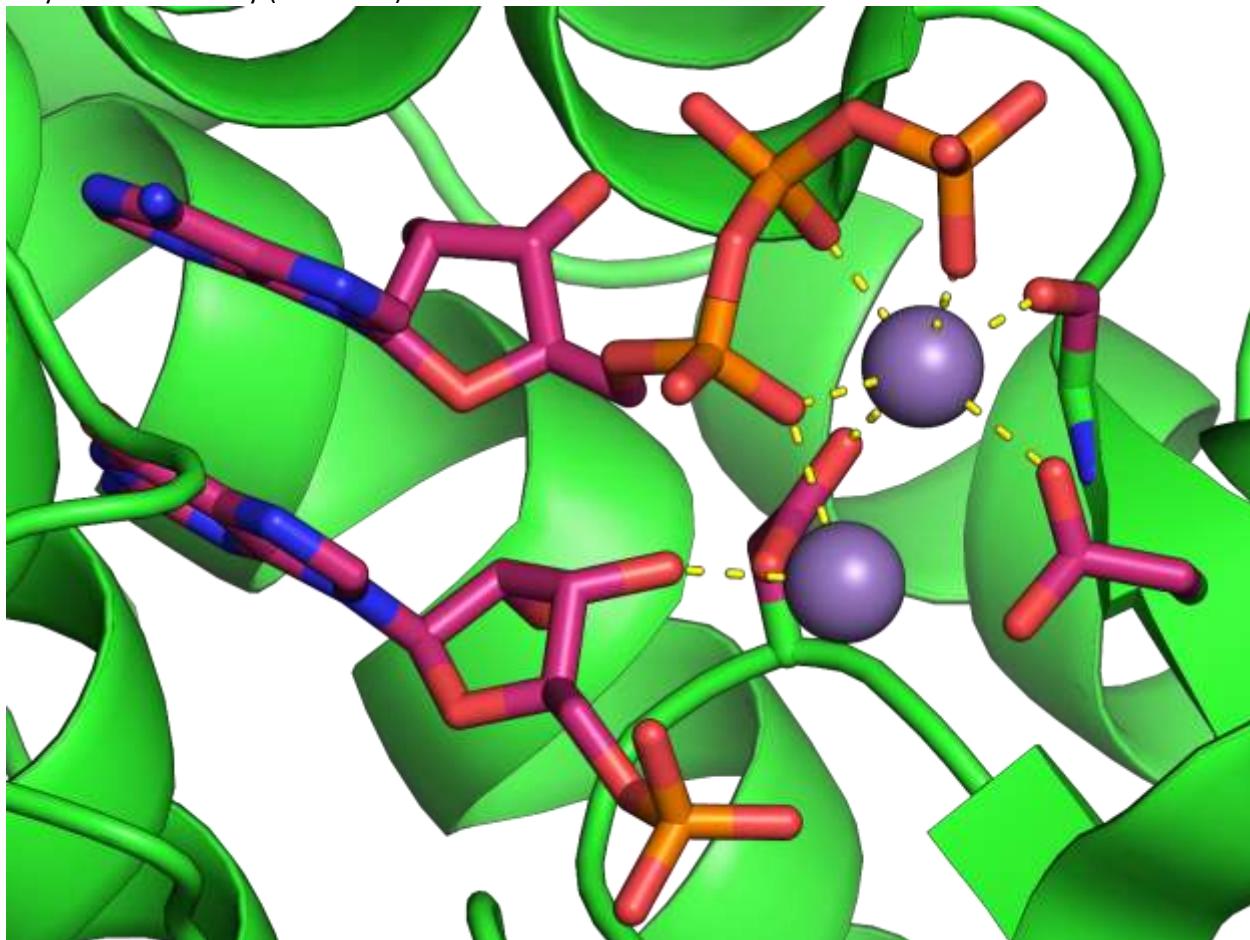
ATP and metal ion coordination in the representative structure 3WDL.

### Virion DNA-directed RNA polymerase domain

There is no corresponding SUPFAM superfamily entry, this category is defined on the InterPro domain level (IPR049432). This domain is found in Virion DNA-directed RNA polymerase from Bacteriophage N4 (vRNAP, Q859P9), responsible for the transcription of the early region of the double-stranded linear DNA genome of the lytic coliphage [25]. It displays two metal ions at the active site in a typical polymerase configuration corresponding to the DNA-directed RNA polymerase function; a  $Mg^{2+}$  is coordinated by all three phosphate groups, and another is coordinated by the 3' hydroxyl of the priming nucleotide and the  $\alpha$ -phosphate of the incoming NTP.



Polymerase reactivity (EC 2.7.7.6) and schematic metal coordination.

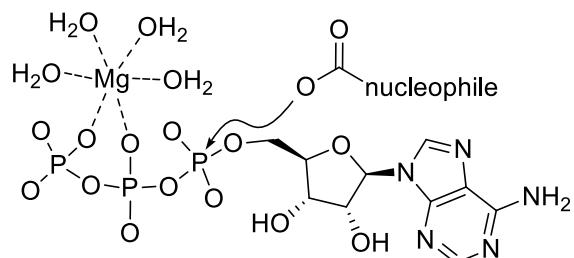


NTP, nucleophile and  $Mn^{2+}$  ion coordination in the representative structure 4FF3.

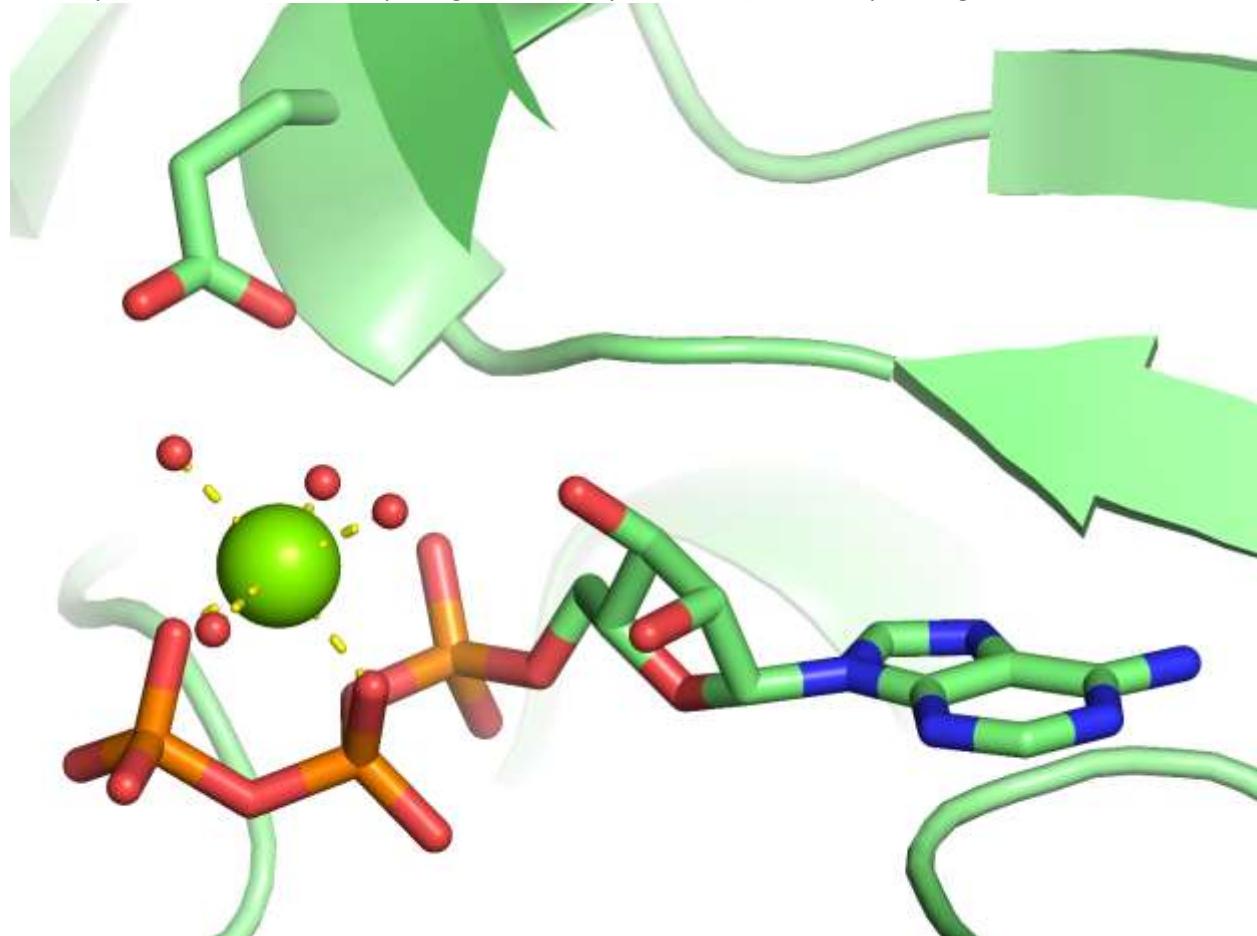
## Pyrophosphatases with two-phosphate coordination on the (+) side

### Acetyl-CoA synthetase-like

The most significant exception from the Mg-pinch role, this superfamily has one  $Mg^{2+}$  ion coordinated by the  $\beta\gamma$  phosphate oxygen atoms (there is no metal cation coordinating the  $\alpha\beta$  phosphate groups). We hypothesize that in the case of this superfamily, due to the negatively charged attacking group, the attack is more easily performed and does not necessitate the coordination between the  $\alpha\beta$  phosphates to hydrolyze ATP to AMP and pyrophosphate. A conserved glutamate residue can contribute to the second-shell coordination of the  $Mg^{2+}$  ion. Also, the superfamily was found to function with various ions [26] suggesting it has a role in the charge balance rather than activation.



Carboxyl activation as a first step of ligase reactivity (EC 6.2.1.-) and non-pinching metal ion coordination.

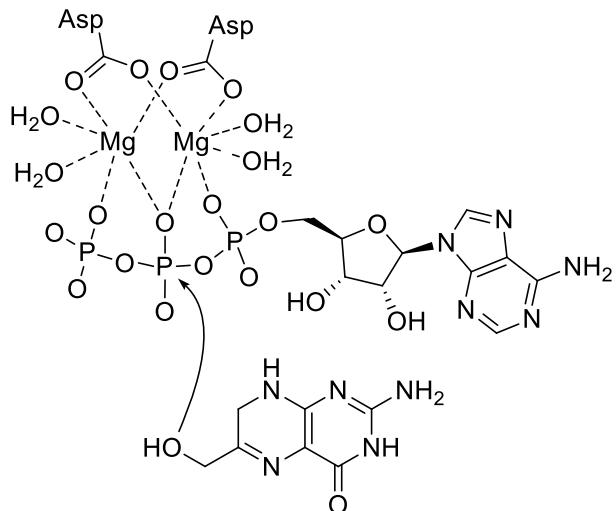


ATP and non-pinching  $Mg^{2+}$  coordination in the representative Ac-CoA synthetase-like structure 5BSM.

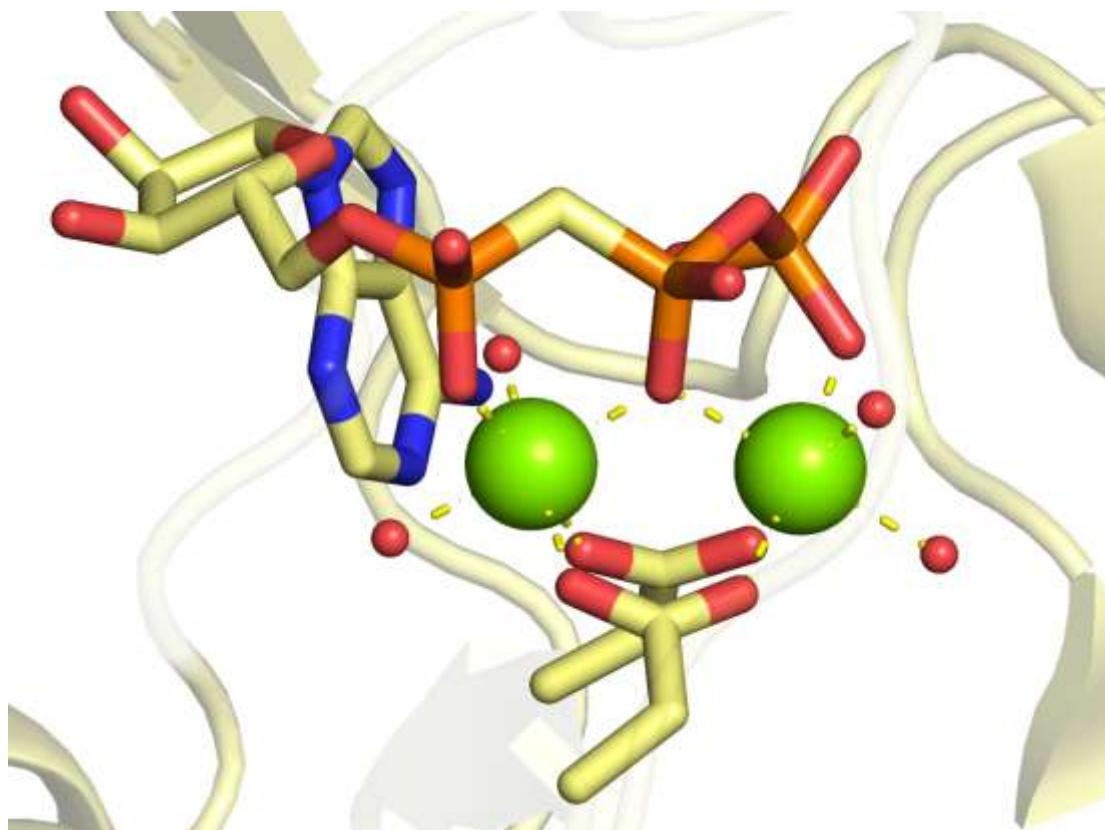
### 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, HPPK

EC: 2.7.6.3 (2-amino-4-hydroxy-6-hydroxymethylidihydropteridine diphosphokinases)

This superfamily has 2 Mg ions, one coordinating the  $\alpha\beta$  and another one coordinating the  $\beta\gamma$  phosphates. In some cases the  $Mg^{2+}$  ion with  $\beta\gamma$  coordination also coordinates the attacking hydroxyl group of the ligand. There are two aspartate residues on the nearby beta-sheet that both coordinate the two Mg ions.



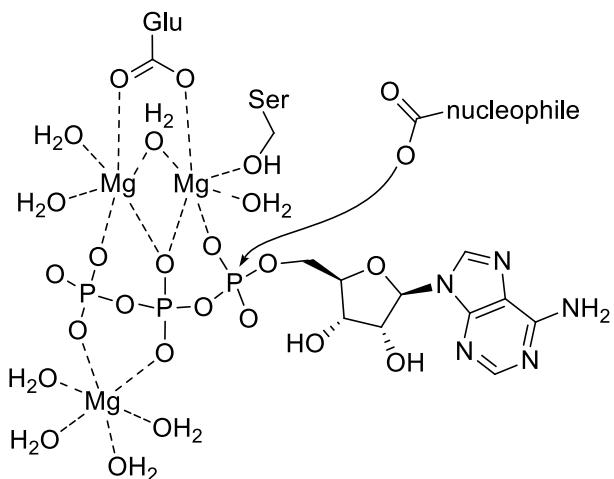
Scheme of ion coordination and specific pyrophosphokinase activity (EC 2.7.6.3).



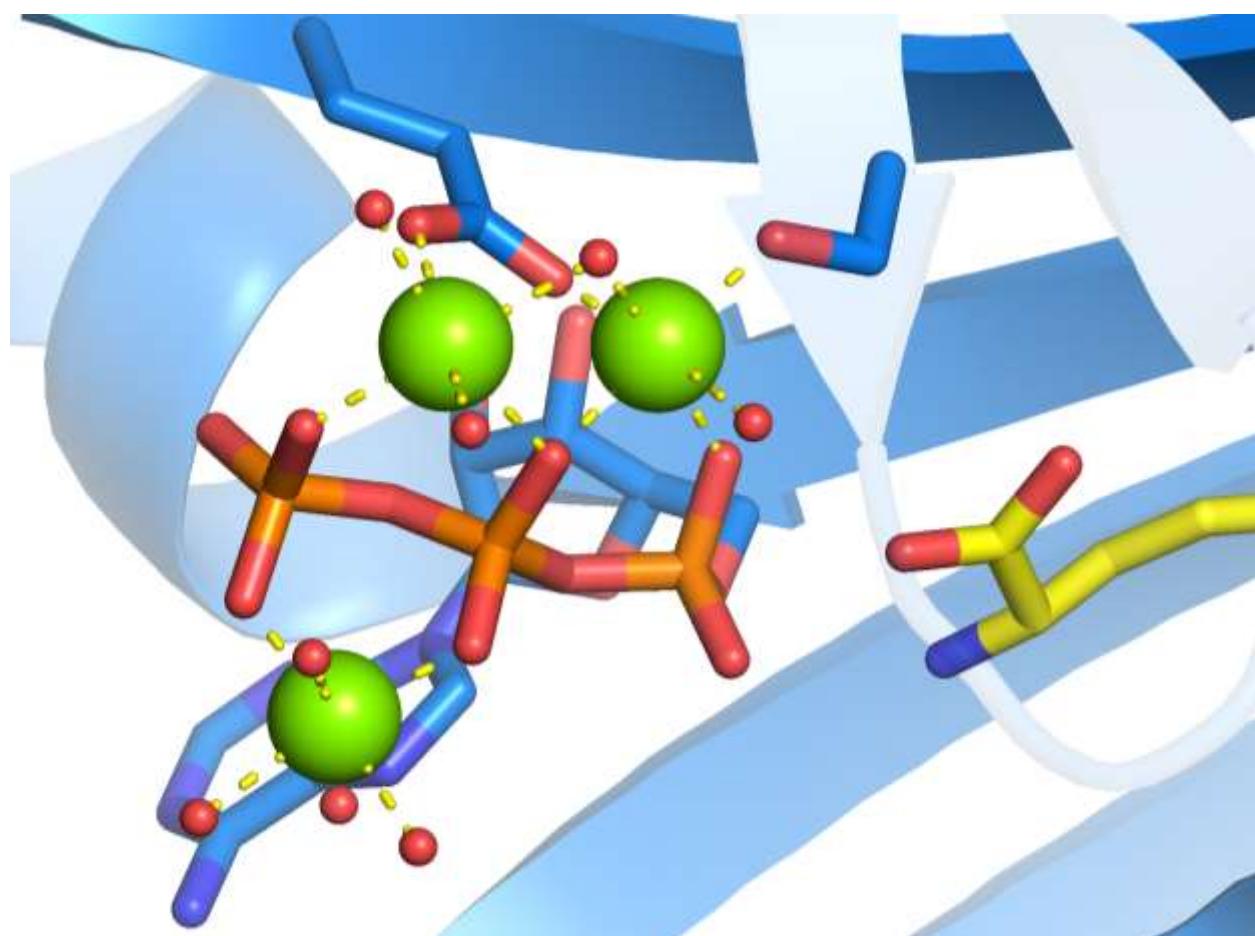
ATP and Mg<sup>2+</sup> coordination in the representative structure 4CRJ.

### Class II aaRS and biotin synthetases

Three metal ion positions are conserved in this superfamily, some resolved structures may be missing one or more of them. Two Mg ions coordinate the  $\beta\gamma$  phosphates while a third Mg<sup>2+</sup> ion coordinates the  $\alpha\beta$  phosphates. A glutamate and a serine residue also take part in the Mg coordination.



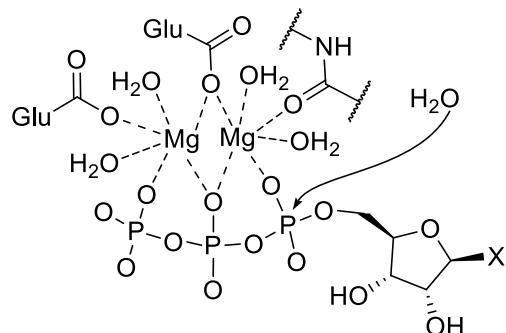
Metal ion coordination and carboxyl activation of typical amino acid tRNA synthetases.



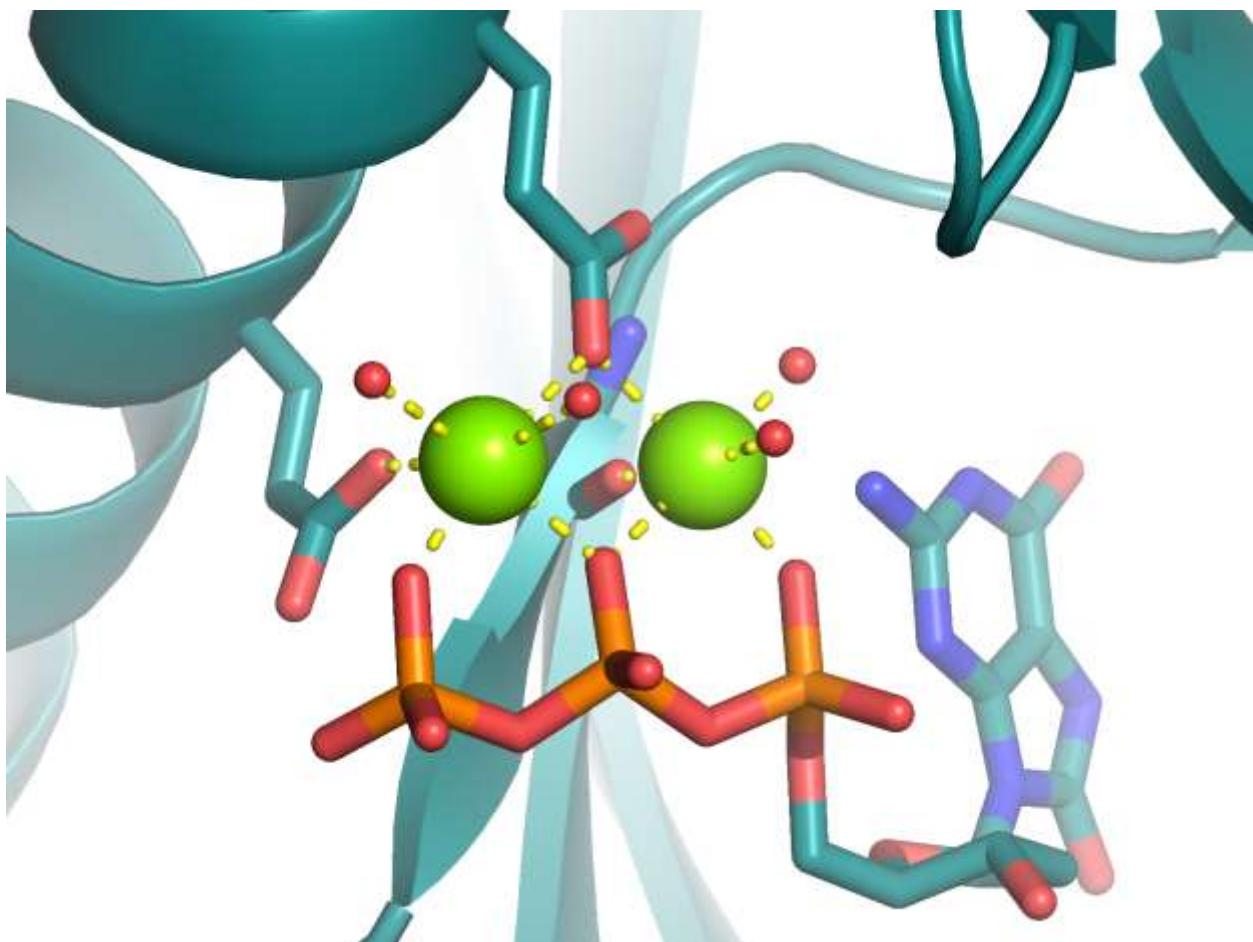
Substrate and ion coordination in the representative structure 6AAZ.

## Nudix

This superfamily has 2 catalytic metal ions. One  $Mg^{2+}$  ion is coordinated by the  $\alpha\beta$ , while another one by the  $\beta\gamma$  phosphate oxygen atoms. The two  $Mg^{2+}$  ions are coordinated by a conserved glutamate residue, while the  $Mg^{2+}$  ion in the  $\beta\gamma$  position is further coordinated by an additional glutamate residue and the  $Mg^{2+}$  ion in the  $\alpha\beta$  position by a main-chain carbonyl group.



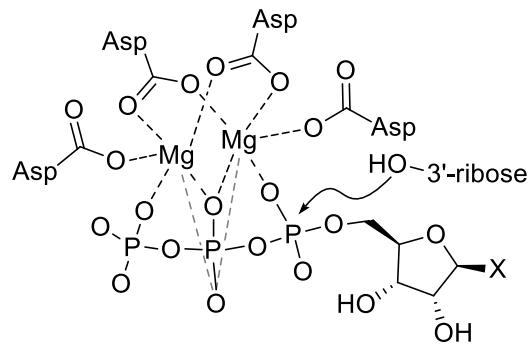
Schematic ion coordination and NTP pyrophosphate hydrolysis (EC 3.6.1.-)



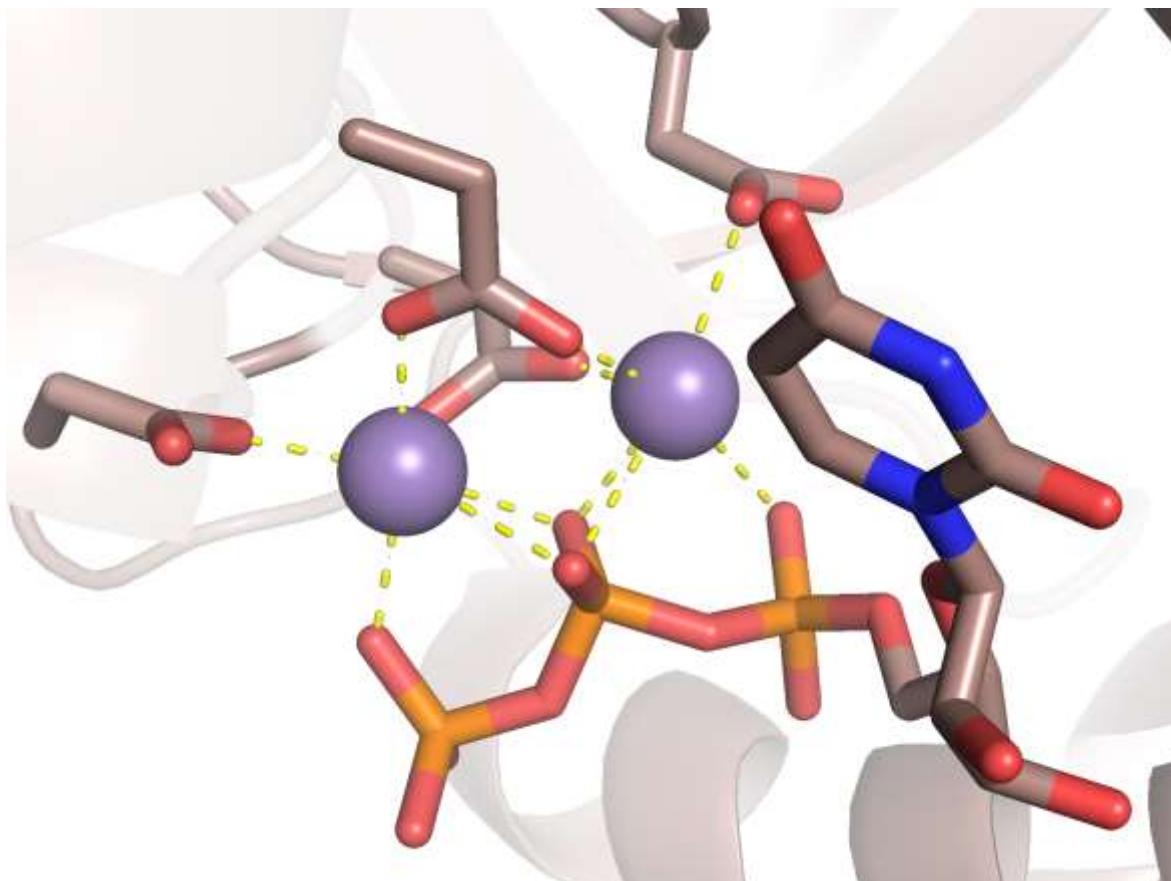
NTP and  $Mg^{2+}$  ion coordination in the representative structure 6FL4.

### Beta and beta-prime subunits of DNA dependent RNA-polymerase

In our dataset only a few structures are found for this superfamily. Despite the excellent agreement upon alignment, only one with an NTP was resolved, while many RNA chains with  $Mg^{2+}$  at the 3' end were also identified. The structure 1TWF was resolved in the presence of two catalytic  $Mn^{2+}$  ions. One  $Mn^{2+}$  ion is coordinated by  $\alpha\beta$ , and the other by the  $\beta\gamma$  phosphate group oxygens. Two aspartate residues further coordinate both metal ions. Additionally, an extra aspartate residue coordinates each of them.



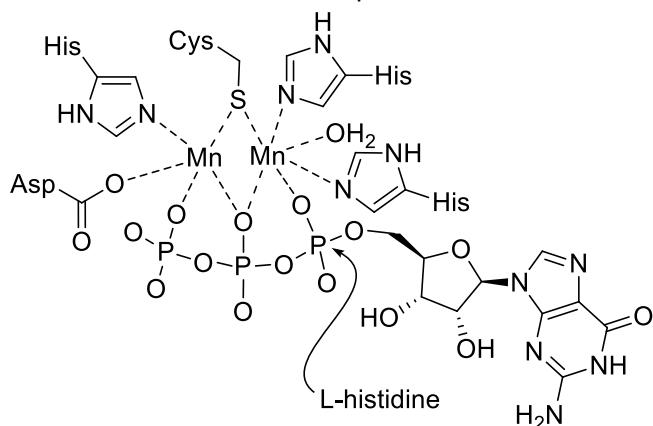
RNA polymerization scheme with  $Mg^{2+}$  ions (EC 2.7.7.6).



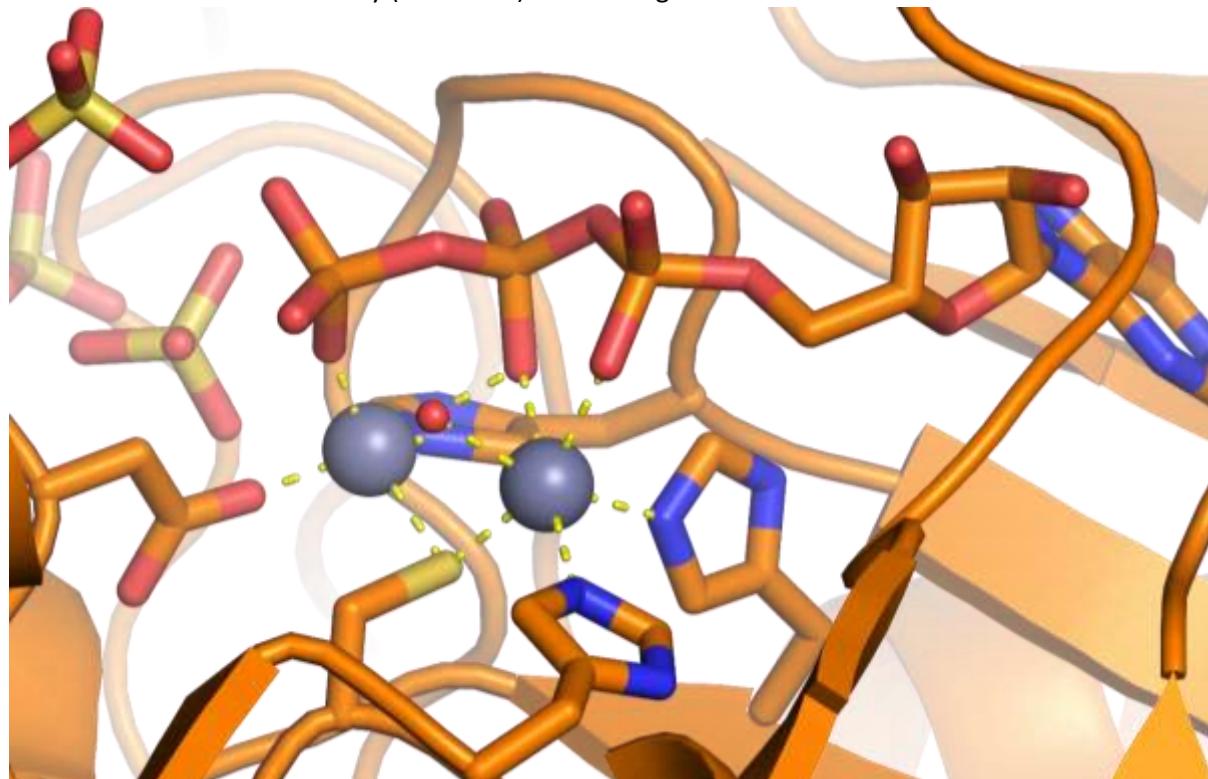
UTP and  $Mn^{2+}$  ion coordination in the representative structure 1TWF.

### tRNA-splicing ligase RtcB-like

This superfamily was identified by InterPro (there is a corresponding SUPFAM entry: Hypothetical protein PH1602, d.261.1; exceptionally, here we use the InterPro name as it is more descriptive). Enzymes belonging to EC 6.5.1.8 make use of GTP, and two catalytic metal cation cofactors. Interestingly, in the case of this superfamily, these metal ions are manganese. The catalytic reaction is less effective with cobalt and nickel, whereas zinc and copper are inactive and potently inhibit manganese-dependent guanylylation [27]. Intriguingly, the  $\alpha\beta$ -coordinated catalytic  $Mn^{2+}$  ion is further coordinated by two histidine and a cysteine residue. An additional water molecule participates in its coordination. A second  $Mn^{2+}$  is in a  $\beta\gamma$  coordination position, and has a coordination number of only 5, which is formed by the previously mentioned cysteine, and an additional histidine and aspartate residues.



Metal coordination and reactivity (EC 6.5.1.8) RtcB-like ligases.

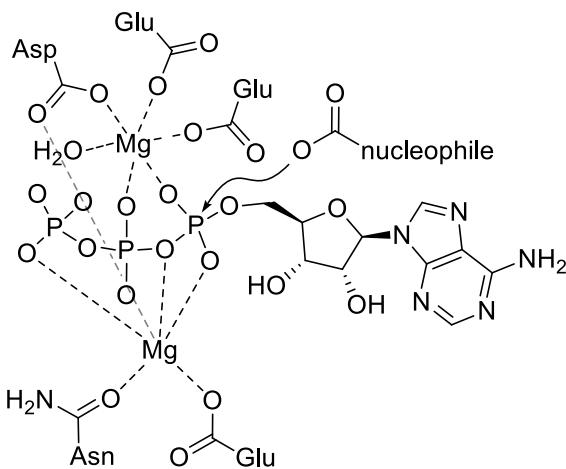


GTP and metal ion coordination in the representative structure 8DCD. The structure is resolved with  $Zn^{2+}$  ions that inhibit the reaction.

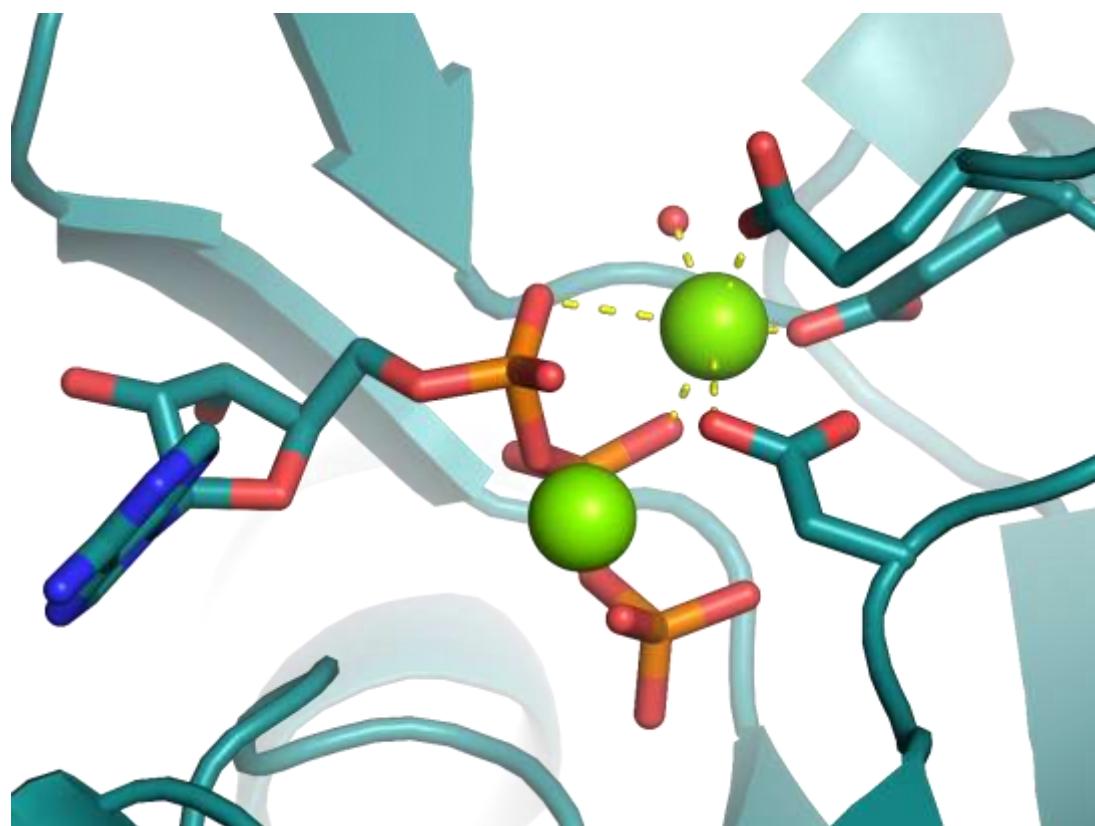
## Aerobactin siderophore biosynthesis, lucA/lucC-like

Known ECs: 6.3.2.39/54-58

This entry is not classified at a superfamily level as no superfamily could be assigned to this particular group by SUPFAM. However, the InterPro database identifies them as belonging to the Aerobactin siderophore biosynthesis, lucA/lucC-like family. Its members exhibit a first  $\alpha\beta$ -coordinated and an additional  $\alpha\gamma$ -coordinated  $Mg^{2+}$ . Interestingly, three carboxyl groups, sidechains of an aspartate and two glutamate residues, participate in the ion coordination. An additional water molecule also coordinates the  $Mg^{2+}$ .



Proposed reactivity based on the associated peptide synthetase ECs 6.3.2.-.

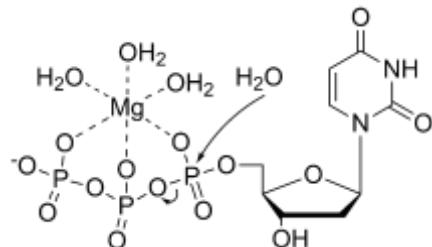


ATP and metal ion coordination in the representative structure 7TGK.

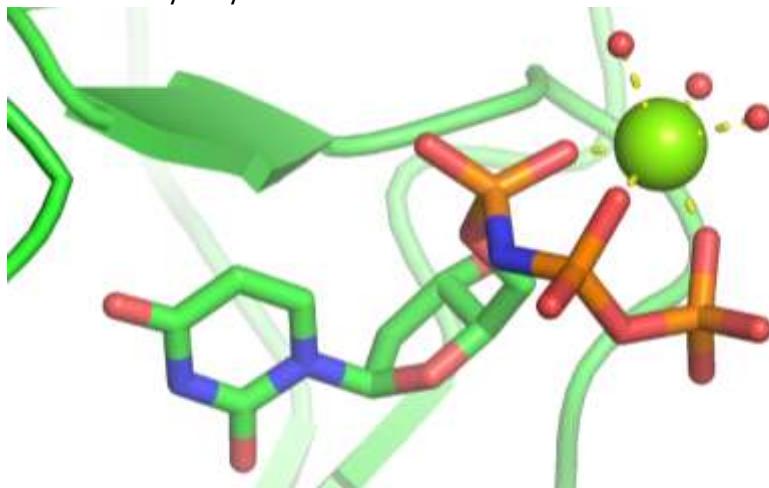
## Pyrophosphatases with $\alpha\beta\gamma$ coordination on the (-) side

### dUTPase-like

Members of this superfamily have one  $Mg^{2+}$  ion that coordinates the  $\alpha\beta\gamma$  phosphates. There are dCTP deaminases in this superfamily (e.g. PDB 1XS1, 2QXX, 2V9X, 4JJC), not all of which are dUMP-forming. In the case where there is no diphosphatase activity of the given deaminase, the spatial exclusion of a potential nucleophilic water molecule is likely to prevent the hydrolysis of the phosphate chain of the bound nucleotide.[28]



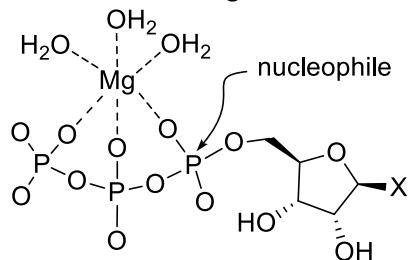
Reaction scheme of dUTP hydrolysis.



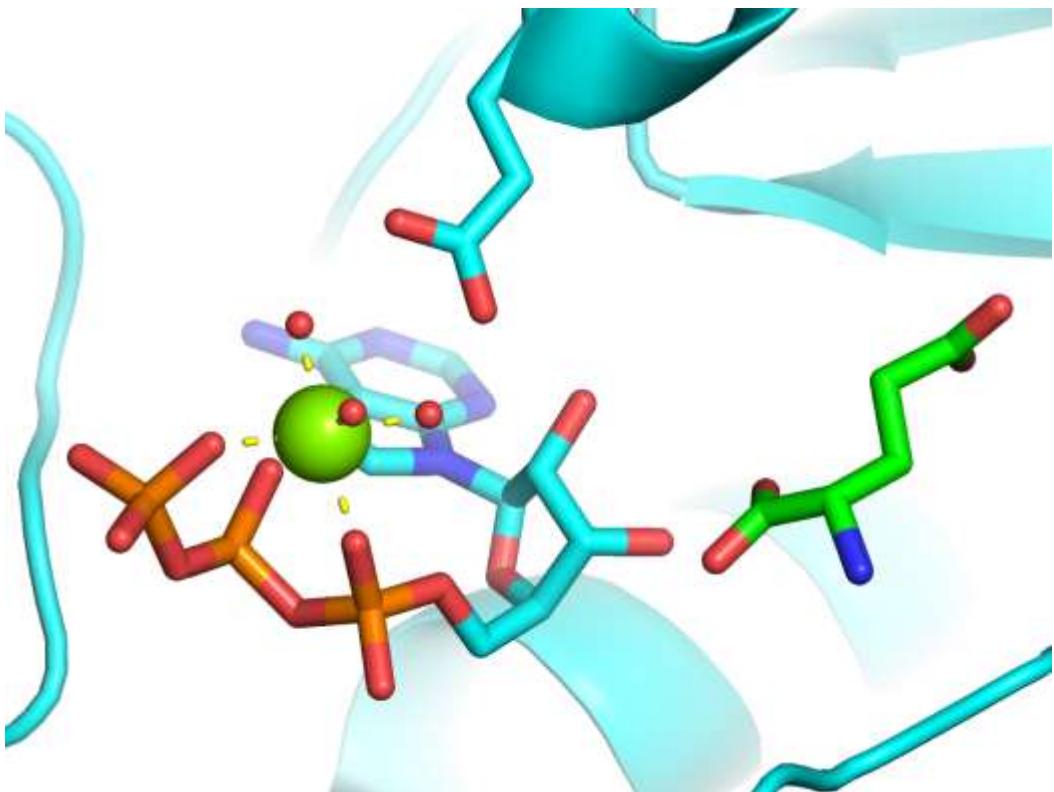
NTP and metal ion coordination in the representative structure 1RN8.

### Nucleotidyl transferase

One structurally conserved aspartate (or in some cases glutamate) residue points towards the  $Mg^{2+}$  ion and can contribute to its coordination mediated through the coordinating water molecules.



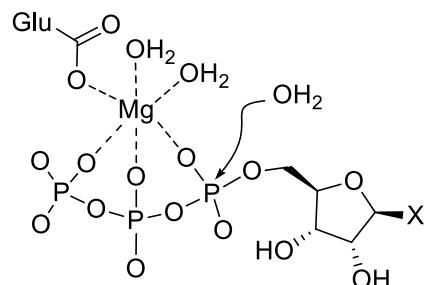
General nucleotidyl transfer reaction scheme with  $Mg^{2+}$  coordination. The superfamily is found among many EC categories, not only in nucleotidyltransferases (2.7.7.-) but also in ligases, like class I tRNA synthases (6.1.1.-).



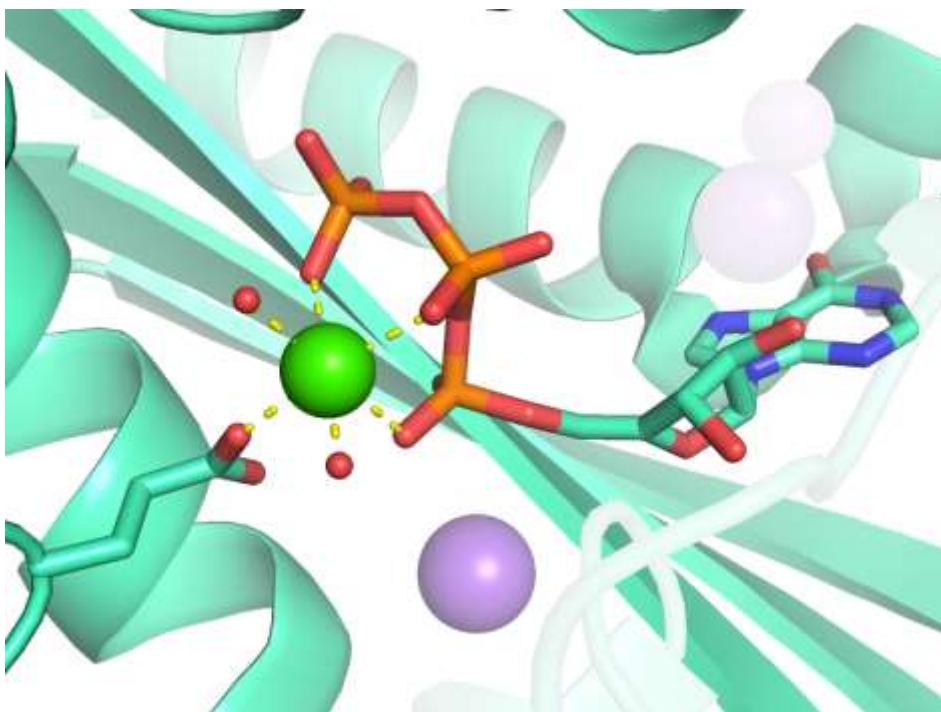
NTP and metal ion coordination alongside an amino acid substrate in the representative structure 1J09.

#### ITPase-like

The superfamily has limited structural data found in our analysis. The representative structure was resolved in the presence of a calcium which is coordinated by  $\alpha\beta\gamma$  phosphate group oxygens as well as a glutamate residue and two water molecules, but for reactivity, the enzyme prefers Mg<sup>2+</sup>. An additional Na<sup>+</sup> ion is present coordinated by the  $\alpha$  phosphate group. The substrate is typically a non-nucleic acid-forming NTP, e.g. XTP or ITP.



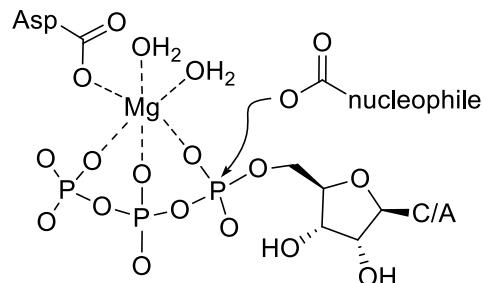
ITP pyrophosphatase reaction and Mg<sup>2+</sup> ion coordination.



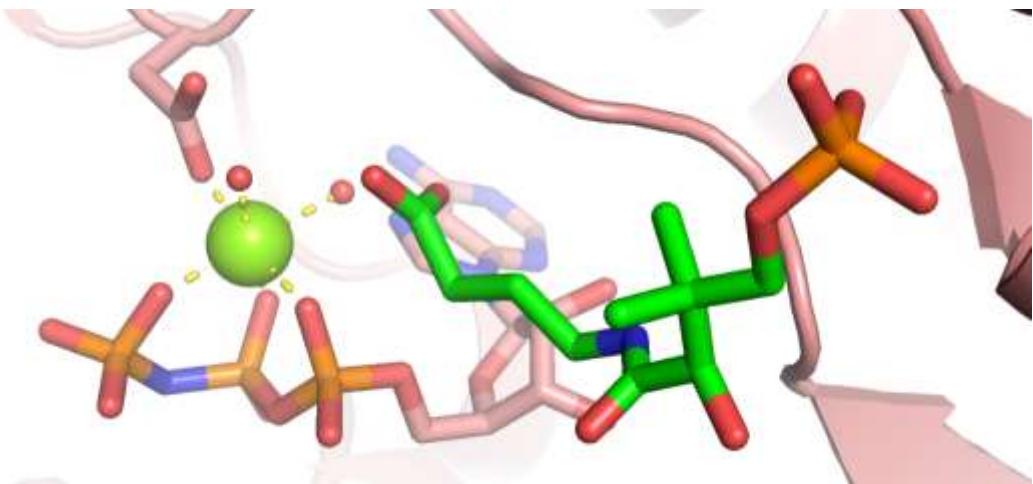
NTP and metal ion coordination in the representative structure 2Q16.

### CoA-like

Members of this superfamily are key enzymes in the production of coenzyme A. The eukaryotic enzymes require ATP, in contrast to the bacterial enzymes which require CTP. A single  $Mg^{2+}$  ion is coordinated by the  $\alpha\beta$  phosphate oxygen atoms as well as by an aspartate residue conserved in all species.[29] The carboxylate group of the ligand is positioned in an orientation towards the  $Mg^{2+}$  ion that is ideal for a nucleophilic attack at the  $\alpha$ -phosphate.



Scheme of the first step of the ligase reaction (EC 6.3.2.5/51) catalyzed by the CoA-like superfamily.

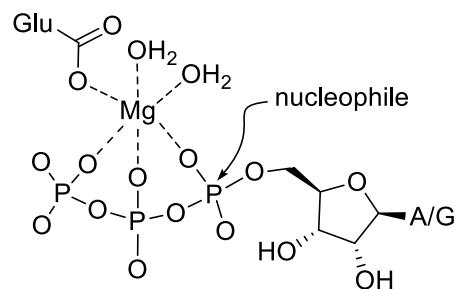


ATP and metal ion coordination in the representative structure 7EDZ.

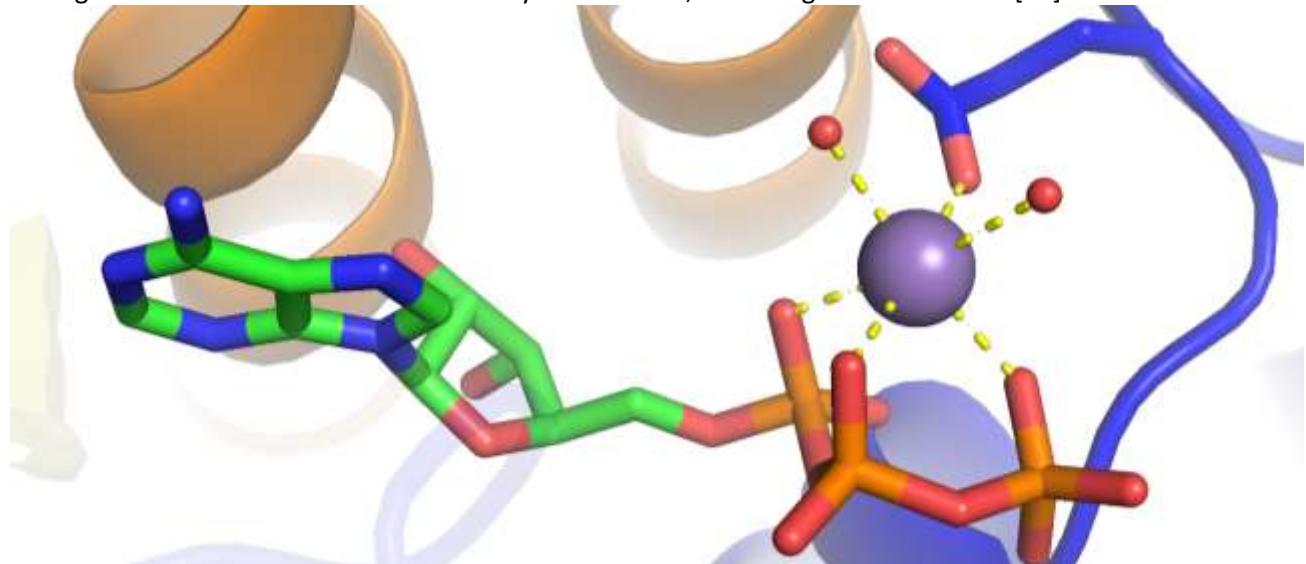
**EPT/RTPC-like or (RNA 3'-terminal phosphate cyclase, RPTC, insert domain)**

Known EC: 6.5.1.4 (RNA 3'-terminal-phosphate cyclase (ATP))

Members of the superfamily may function both with  $Mg^{2+}$  and  $Mn^{2+}$ .<sup>[30]</sup> The ion is coordinated by nonbridging oxygens from each of the ATP phosphates ( $\alpha\beta\gamma$ ), two water molecules, and a glutamate residue.



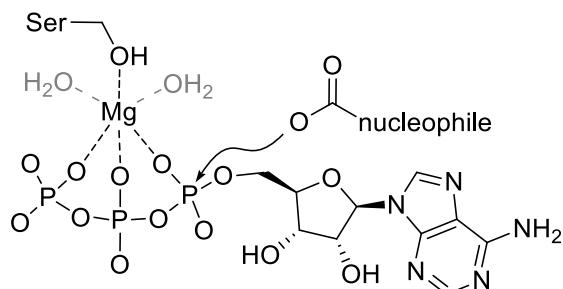
The ligase-like activation can be achieved by ATP and GTP, according to ECs 6.5.1.4-5 [31].



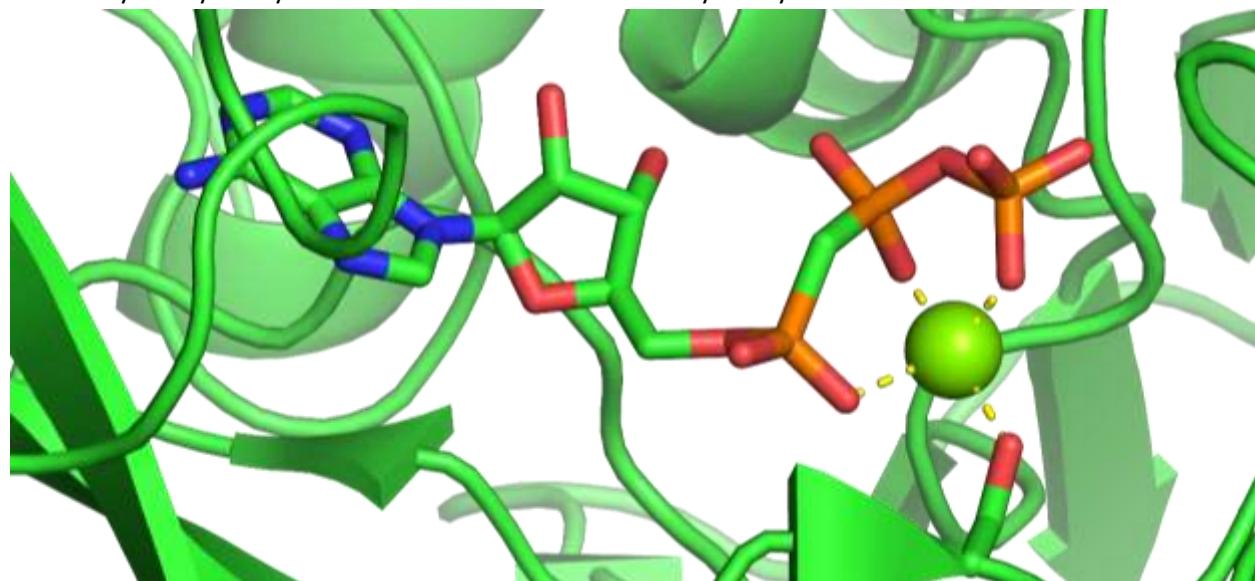
ATP and  $Mn^{2+}$  ion coordination in the representative structure 3TUX.

### **YrdC/RibB**

Enzymes belonging to EC 2.7.7.87 are L-threonylcarbamoyladenylate synthases, while those belonging to EC 3.5.4.25 are GTP cyclohydrolase II, which mainly act on the guanine base but loses phosphates besides. We identified two remarkably different subgroups in this superfamily, those similar to PDB 3VTH and those to 7UF0. Only structures of the PDB 3VTH-like subgroup contained nucleotide (analogs) in our dataset. Our representative structure (3VTH) shows an  $\alpha\beta\gamma$ -coordination of its catalytic  $Mg^{2+}$  ion. This ion is further coordinated by a serine residue, and presumably two water molecules are missing from the active site in this structure.



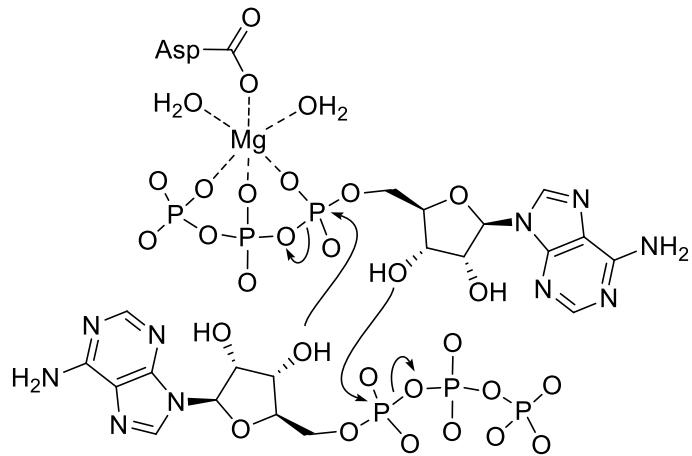
Carbamoyladenylate synthesis based on the reaction catalysed by the structure 3VTH.



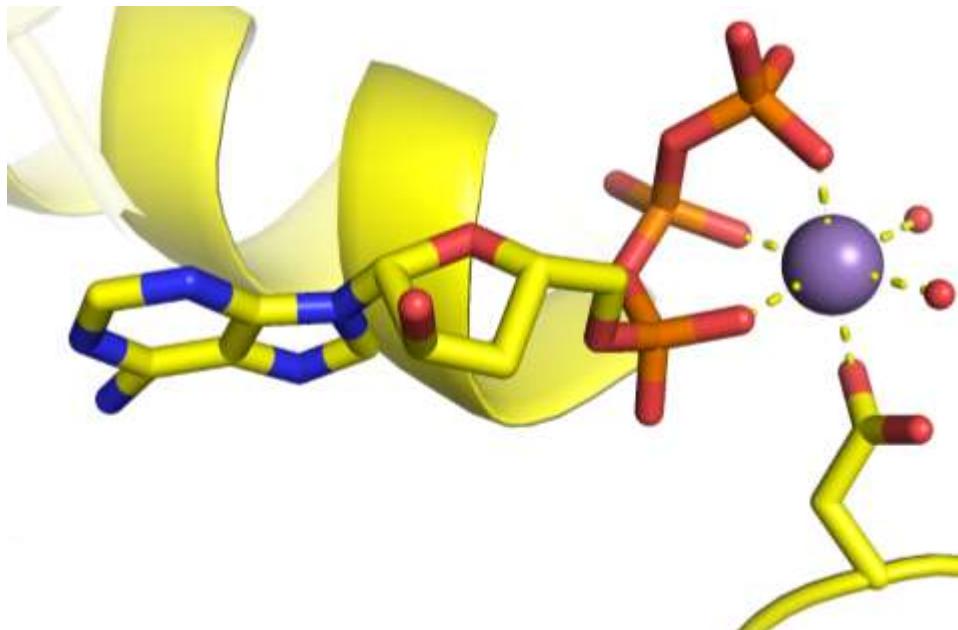
ATP and  $Mg^{2+}$  ion coordination in the representative structure 3VTH.

### **YojJ-like**

The representative structure was solved in the presence of a  $Mn^{2+}$  ion, a substitute for magnesium. The metal ion is octahedrally coordinated by the  $\alpha\beta\gamma$  phosphate groups of the nucleotide together with an aspartate residue and two water molecules. Despite the overall dimerization reaction, only one (d)ATP is present in the active site.



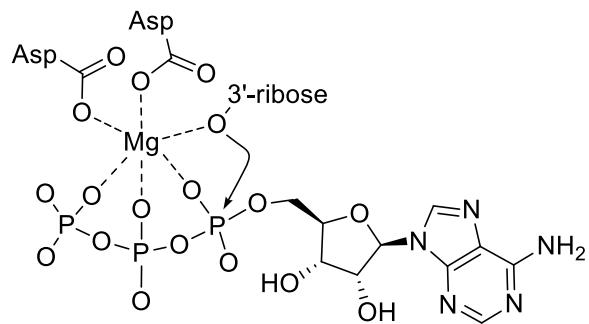
Presumed cyclization reaction and primary ion coordination.



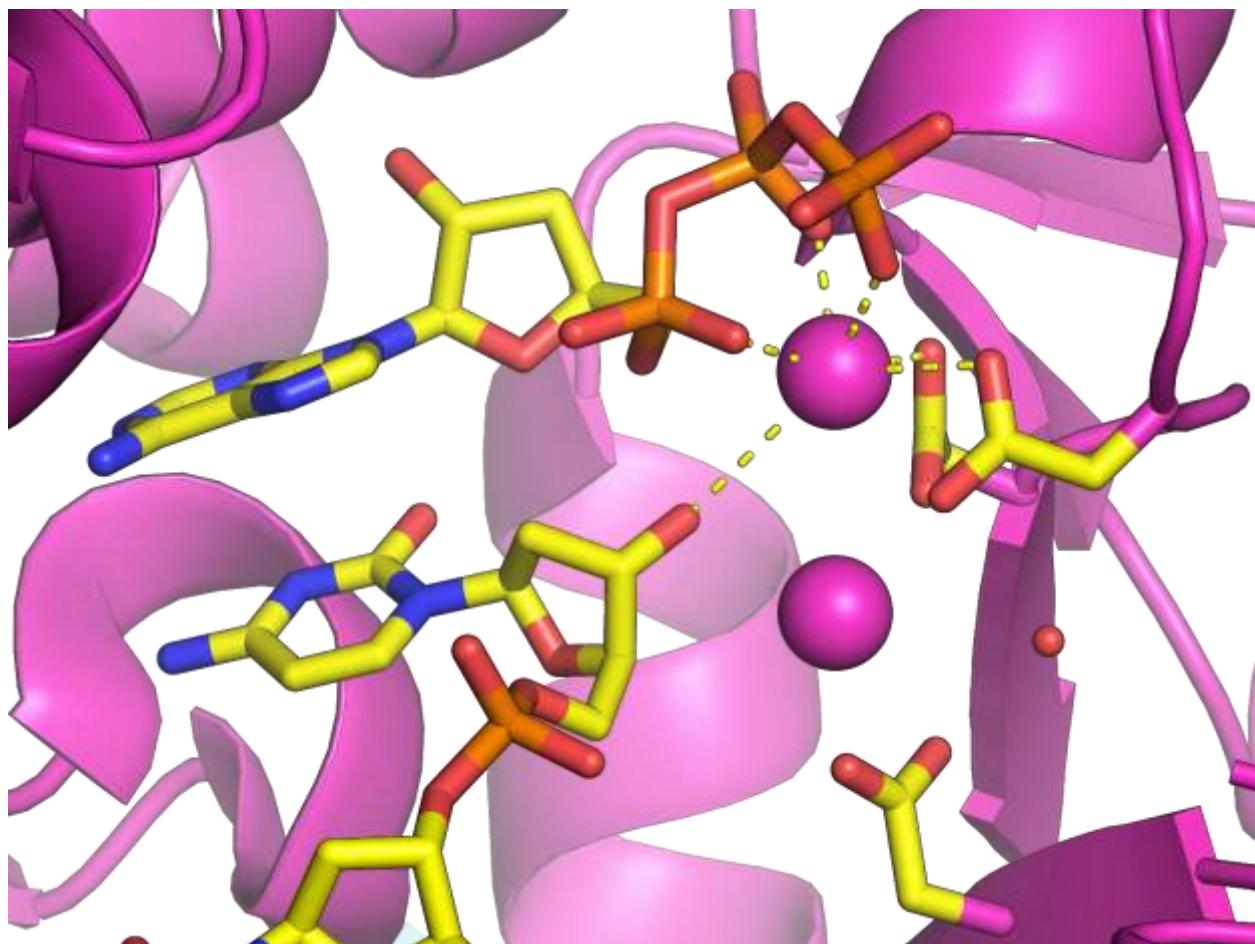
dATP and  $Mn^{2+}$  ion coordination in the representative structure 4YVZ.

#### Poly(A) polymerase catalytic subunit-like

Poly(A) polymerase catalyzes template-independent extension of the 3'-end of a DNA or RNA strand by one nucleotide at a time. The Poxvirus enzyme creates the 3'(poly)A tail of mRNAs, and is a heterodimer of a catalytic and a regulatory subunit, out of which this category corresponds to the catalytic subunit.[32] There are two metal ions at the active site, one is coordinated by all three phosphate groups of the nucleotide, and further coordinated by two aspartate residues, whereas the other is not coordinated by the NTP, but three aspartates and to the oligonucleotide's 3'OH nucleophile.



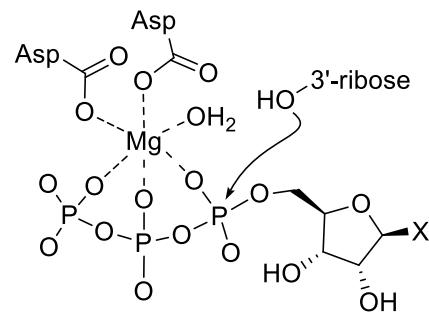
Scheme of the 3' adenylate addition catalyzed by 2.7.7.19.



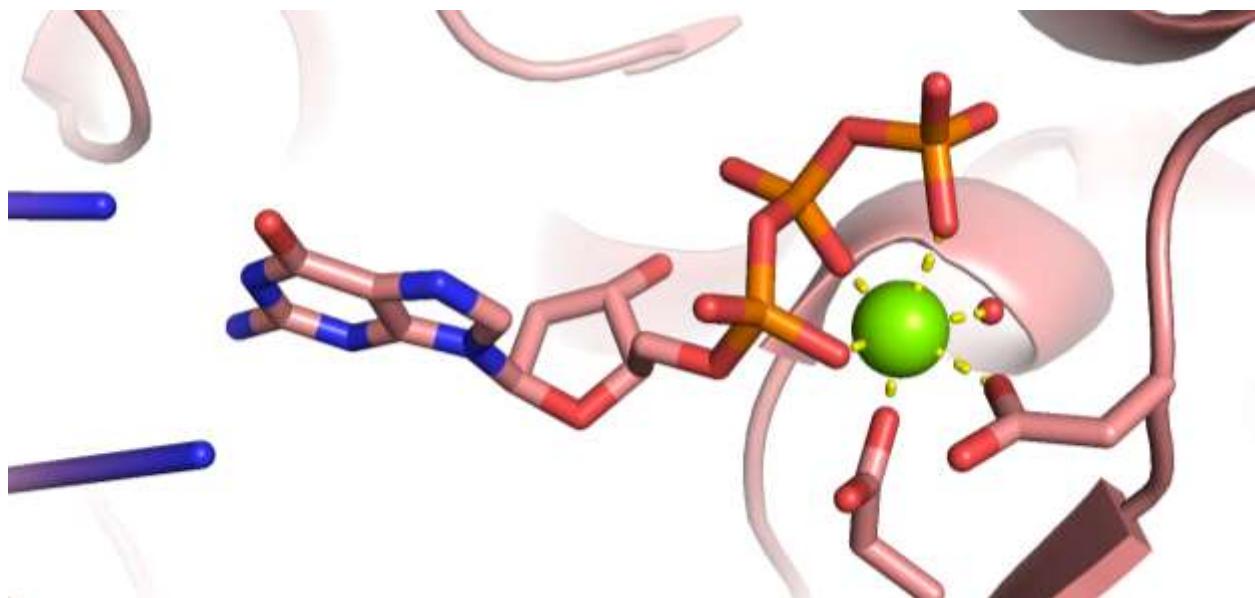
Substrate and metal ion binding in the representative structure 3ERC.

#### Bacterial DNA polymerase III, alpha subunit, NTPase domain (IPR011708)

For three structures originating from the same study, no clear superfamily was assigned for the region that is involved in the nucleotide-binding. They are DNA polymerases, have a single  $Mg^{2+}$  or  $Mn^{2+}$  ion which is coordinated by nonbridging oxygens of the  $\alpha\beta\gamma$  phosphate groups, two aspartate residues and a water molecule. The authors suggest a second metal ion may be missing from their structures coordinated by the  $\alpha$  phosphate group and an additional aspartate residue [33].



DNA polymerase reactivity and ion coordination templated off the representative structure 3F2B.

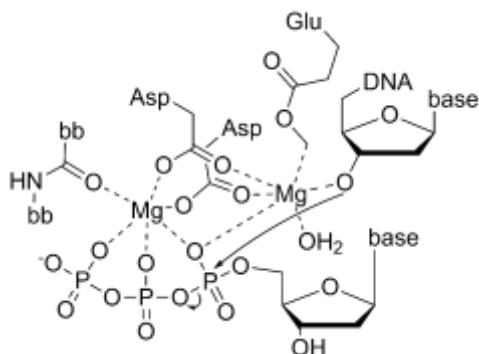


NTP and metal ion coordination in the representative structure 3F2B.

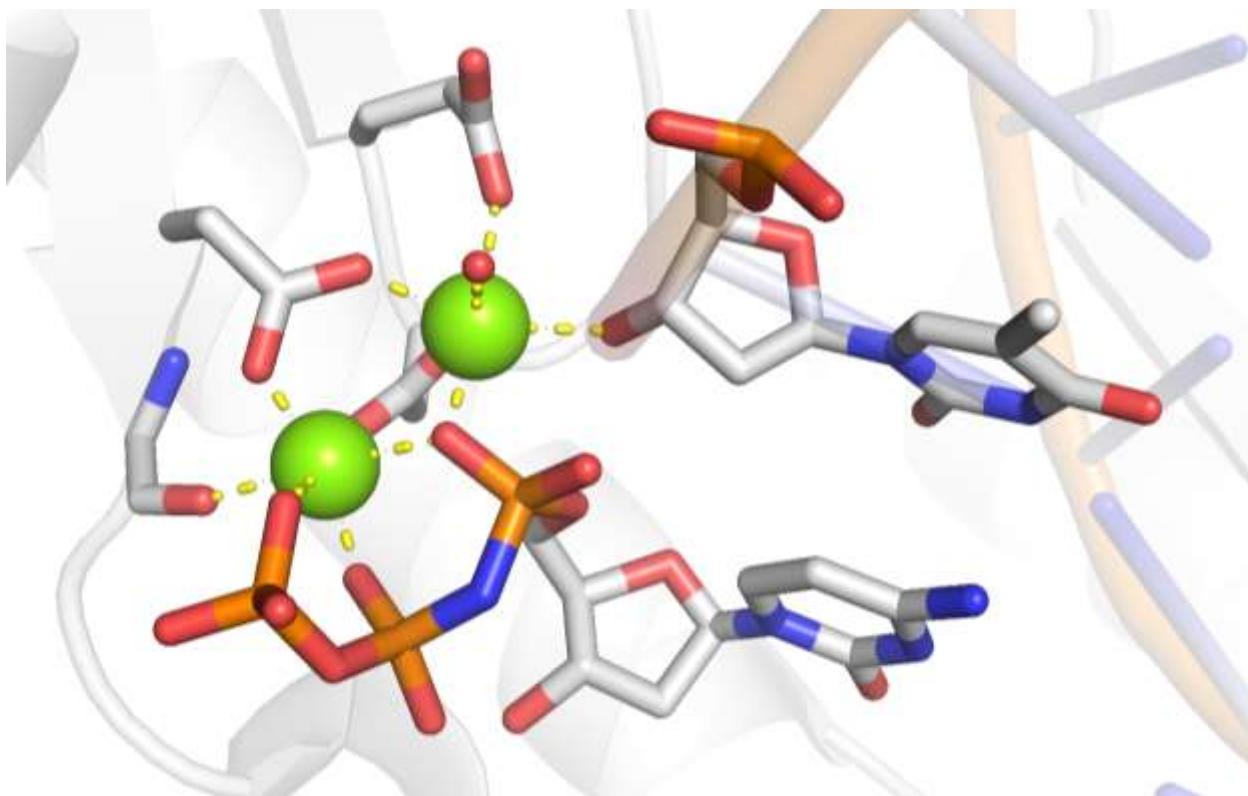
## Pyrophosphatases with two ions on the (-) side: $\alpha\beta\gamma$ and $\alpha$

### DNA/RNA polymerases

It is the most populated pyrophosphatase superfamily in our dataset. It exhibits a clear two-metal-ion coordination. In some of the structures a third metal ion is present (not far from the  $\gamma$  phosphate) which may possess a role in the release of the ligand [34]. All structures have a first  $Mg^{2+}$  ion with a clear  $\alpha\beta\gamma$  phosphate coordination and a second  $Mg^{2+}$  that coordinates the  $\alpha$  phosphate and the attacking hydroxyl group of the ligand nucleotide. Two aspartate residues are located on the nearby beta sheet and are involved in the coordination of both metal ions, whereas an additional glutamate coordinates only the second metal ion.



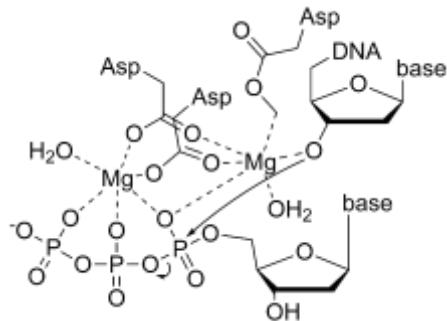
Typical ion coordination depicted in the example of DNA polymerization.



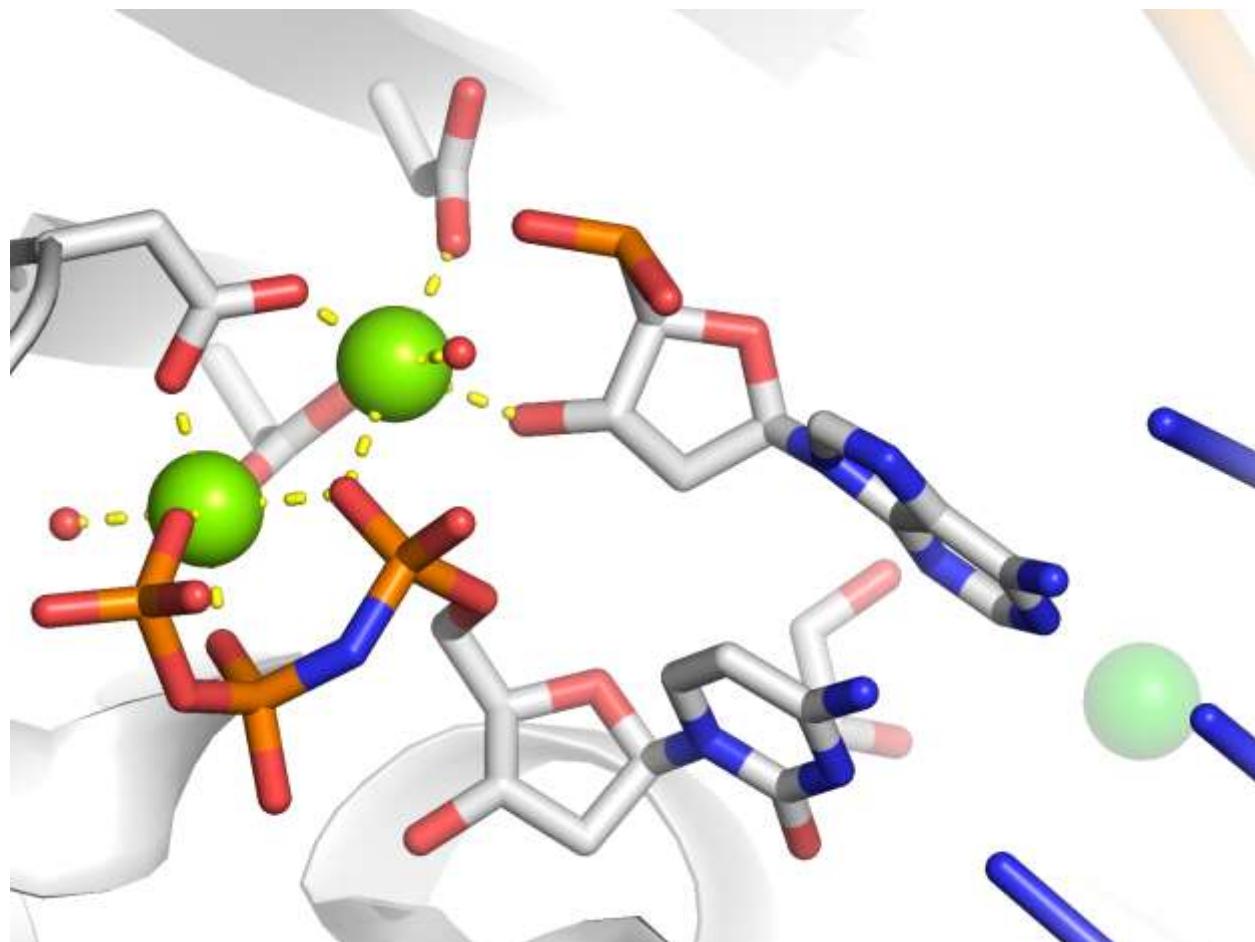
NTP, nucleic acid and metal ion coordination in the representative structure 4RU9.

### Nucleotidyltransferase

Similar to the DNA/RNA polymerases, the majority of this superfamily belongs to the DNA-directed DNA polymerases. Even though the two folds are structurally remarkably different, the nucleotide, the catalytic residues, and the coordinating metal ions are positioned and oriented almost identically. Nucleotidyltransferases also have 2 metal ions at the catalytic site, the first  $Mg^{2+}$  coordinating the  $\alpha\beta\gamma$  phosphates while the second  $Mg^{2+}$  the  $\alpha$  phosphate and the attacking hydroxyl group of the ligand. Instead of the coordinating glutamate residue, all three residues are aspartates in the case of the Nucleotidyltransferases.



Typical ion coordination depicted in the example of DNA polymerization.

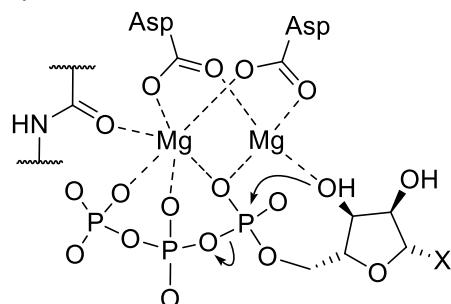


NTP, nucleic acid and metal ion coordination in the representative structure 6P1P.

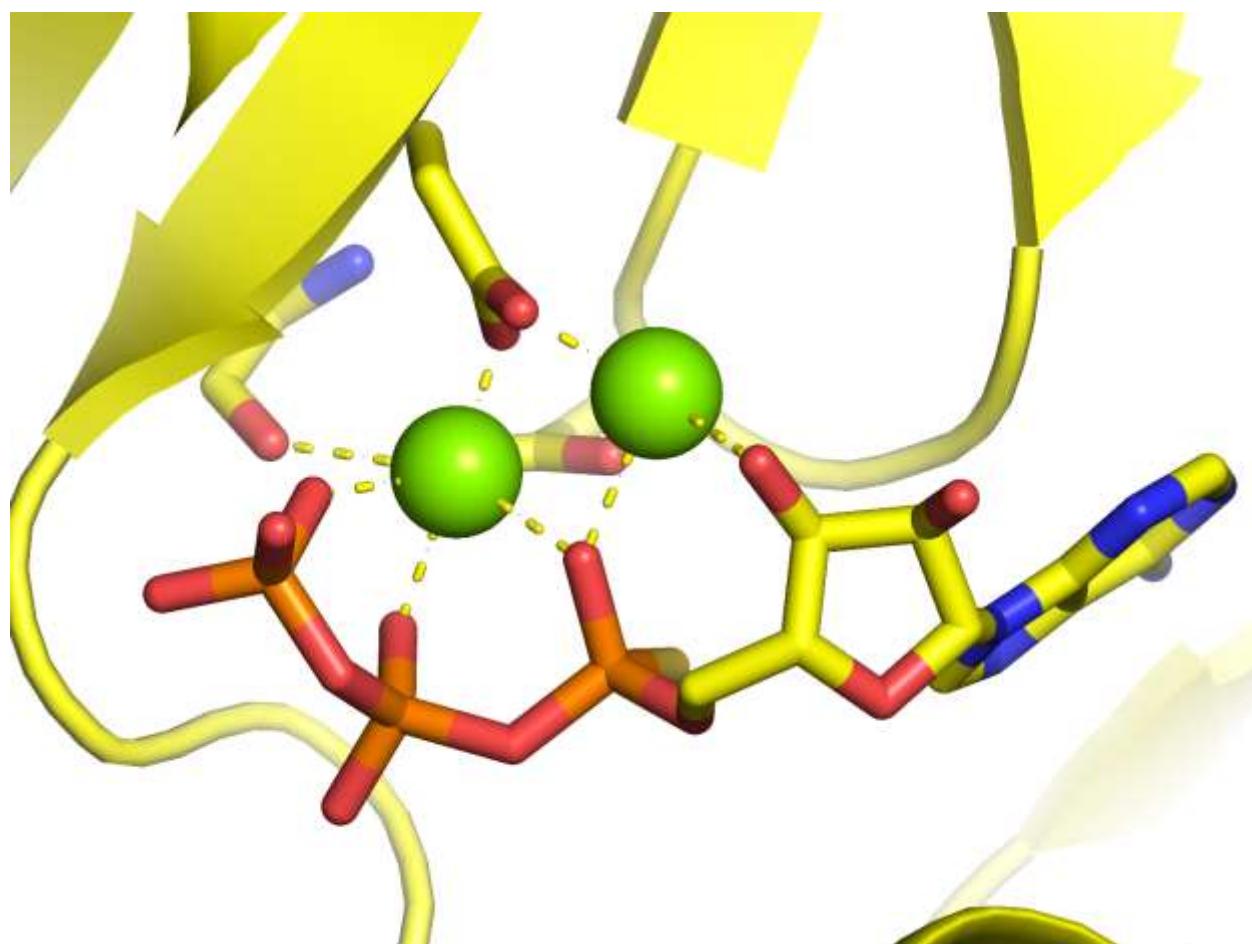
### Nucleotide cyclase

Members of the Nucleotide cyclase superfamily produce 3',5'-cyclic NMP and pyrophosphate while they hydrolyze NTP. Two metal ions are required for catalysis.

Intriguingly, even though  $\text{Ca}^{2+}$  typically inhibits NTP processing enzymes and  $\text{Mg}^{2+}$  is the typical catalytic enzyme, in this superfamily, in the ‘soluble’ adenylyl cyclase (sAC) enzyme active site (*S. platensis*)  $\text{Ca}^{2+}$  plays the pinching metal ion role at the ABG(–) site, while an  $\alpha$ -coordinated  $\text{Mg}^{2+}$  coordinates the attacking hydroxyl group of the ATP, orienting it for the cyclase reaction [35]. Most enzymes of this SF, however, use  $\text{Mg}^{2+}$  in both ion positions. The metal ion coordinated by the  $\alpha\beta\gamma$  phosphates is further coordinated by two aspartate residues in its vicinity.



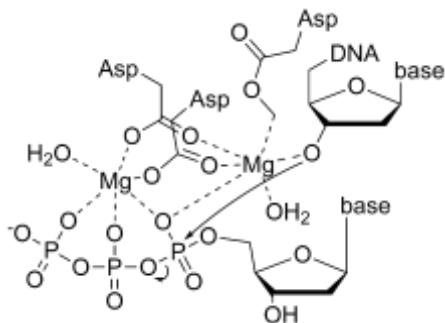
Nucleotide cyclisation and typical metal ion coordination.



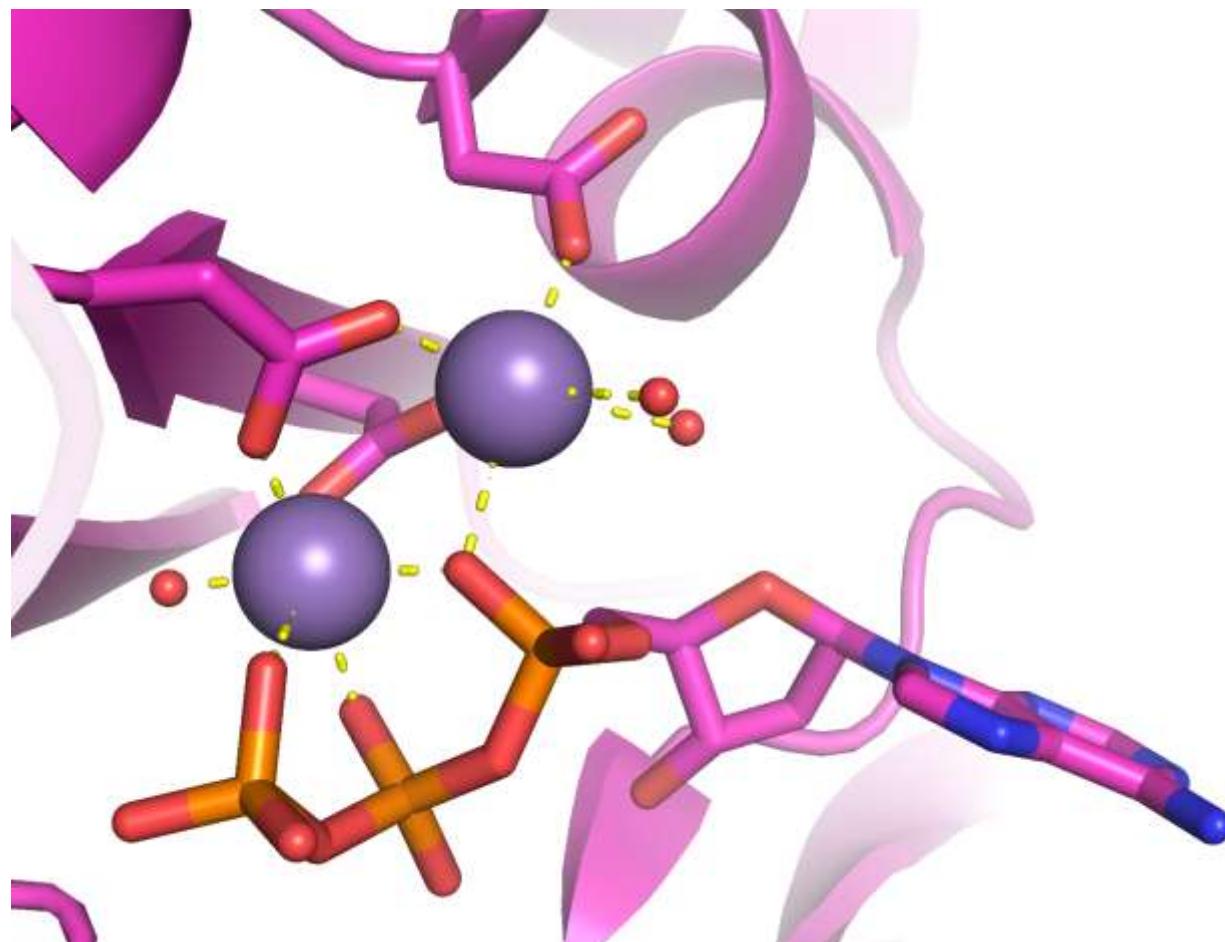
Metal ion and NTP coordination in the representative structure 1WC6.

### Prim-pol domain

Members of this superfamily have two catalytic divalent metal ions. One is coordinated by nonbridging oxygens of the  $\alpha\beta\gamma$  phosphates, and another one of only the  $\alpha$  phosphate. Two conserved aspartate residues are included in the coordination of both Mn ions, and an additional aspartate/glutamate residue coordinates the Mn ion in the  $\alpha$  position. Notably, prim-pol is a Mn ion dependent enzyme that shows significantly improved primase and polymerase activities when binding Mn, rather than Mg ions as cofactors [36].



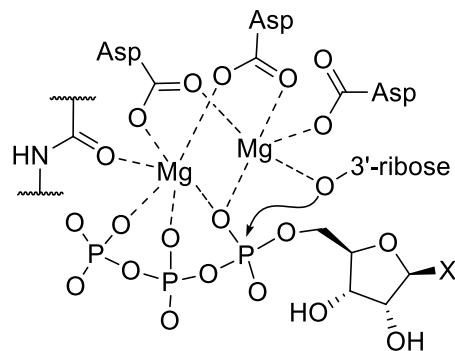
DNA polymerase reactivity and metal ion coordination associated with the Prim-pol domain superfamily.



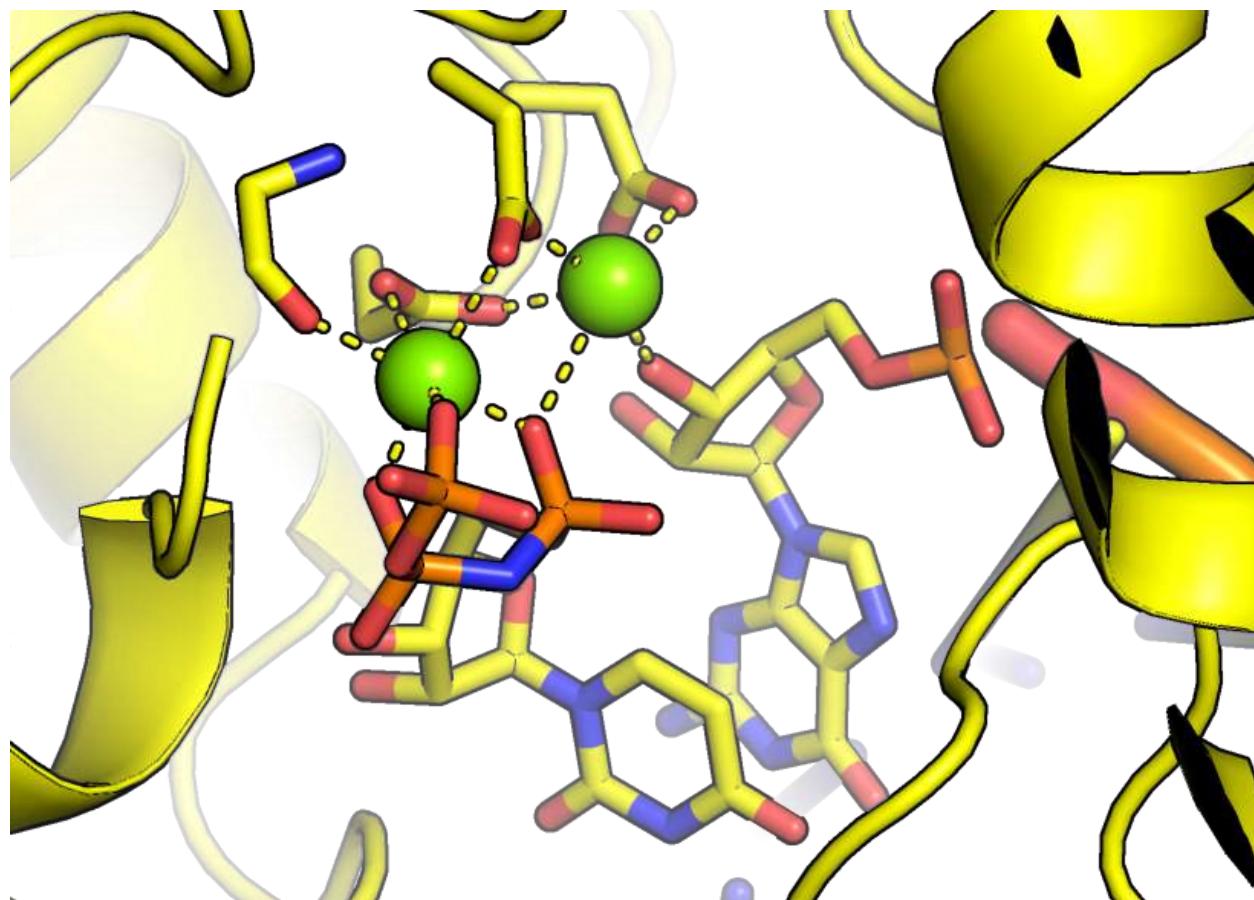
NTP and Mn<sup>2+</sup> coordination in the representative structure 6R5D. The polymerization undergoes with both ions, Mn<sup>2+</sup> is used for enhanced NTP affinity.

### Influenza RNA-dependent RNA polymerase subunit PB1 (IPR001407)

Influenza RNA-dependent RNA polymerase is composed of three subunits; P1 (or PB1), P2 (or PA), and P3 (or PB2). PB1 is the core of the complex and accounts for the polymerase activity [37]. They catalyze RNA-template-directed extension of the 3' end of an RNA strand by one nucleotide at a time and can initiate a chain *de novo*. There is no corresponding SUPFAM category, this group is named at the InterPro family level (IPR001407). The consensus coordination displays a typical polymerase arrangement, one  $Mg^{2+}$  coordinated by all three phosphate groups, and another that is coordinated by the 3' hydroxyl of the priming nucleotide and the  $\alpha$ -phosphate of the incoming NTP.



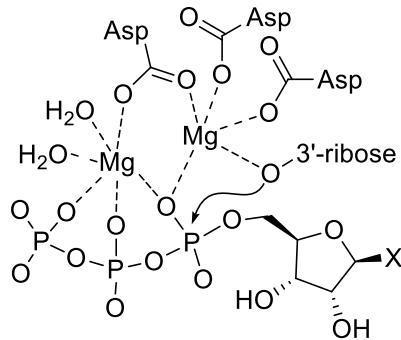
RNA polymerization and typical ion coordination.



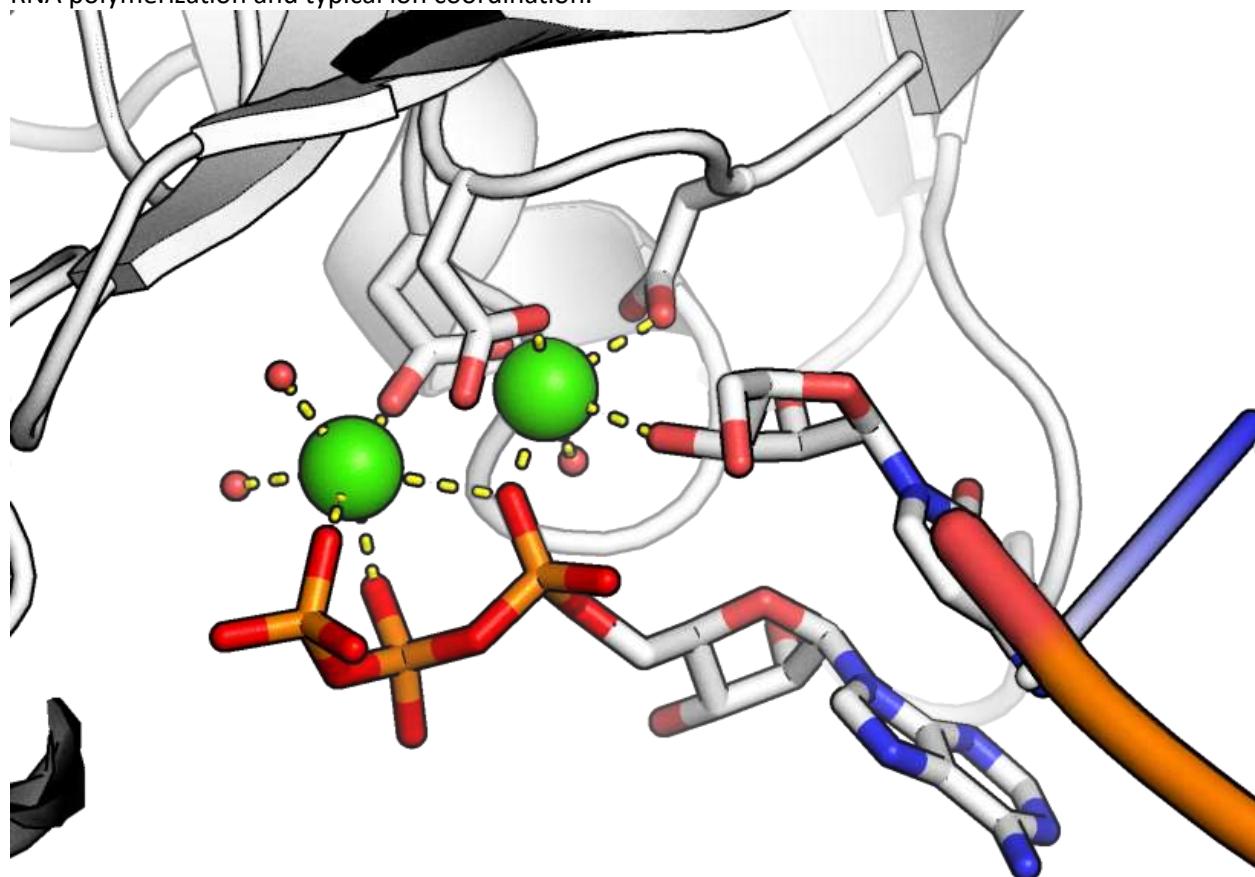
NTP, RNA and metal ion coordination in the representative structure 6T0V.

### RNA-dependent RNA polymerase, eukaryotic-type (IPR007855)

This entry represents various eukaryotic RNA-dependent RNA polymerases, such as RCRP-1, RDRP-2 and RDRP-6. These enzymes are involved in the amplification of regulatory microRNAs during post-transcriptional gene silencing [38]. They catalyze RNA-template-directed extension of the 3'- end of an RNA strand by one nucleotide at a time and can initiate a chain *de novo*. There is no corresponding SUPFAM category, this group is named at the InterPro family level (IPR007855). The consensus coordination displays a typical polymerase arrangement, one  $Mg^{2+}$  coordinated by all three phosphate groups, and another that is coordinated by the 3' hydroxyl of the priming nucleotide and the  $\alpha$ -phosphate of the incoming NTP.



RNA polymerization and typical ion coordination.

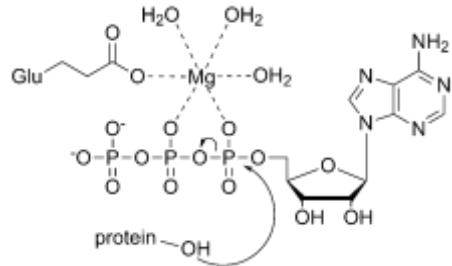


NTP, RNA and metal ion coordination in the representative structure 7Y7P.

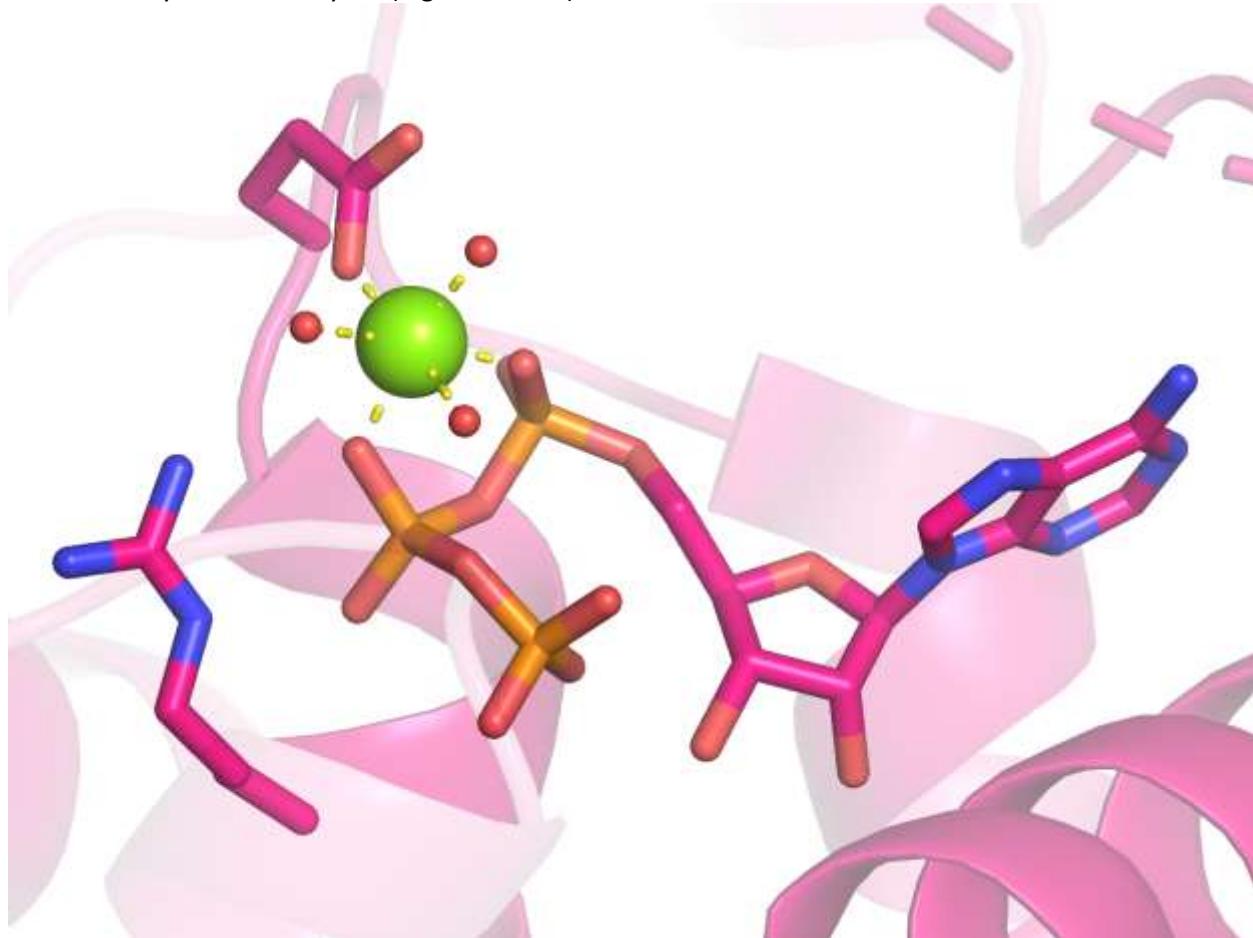
## Pyrophosphatases with $\alpha\beta$ coordination on the (-) side

### Fic-like

This superfamily has one  $Mg^{2+}$  ion which is coordinated by  $\alpha\beta$  phosphate oxygen atoms, and a nearby glutamate (sometimes aspartate) residue.



Protein adenylyltransferase activity, typical to Fic-like enzymes in EC 2.7.7.108. Further adenylation reactions may also be catalysed (e.g. EC 2.7.7.1).

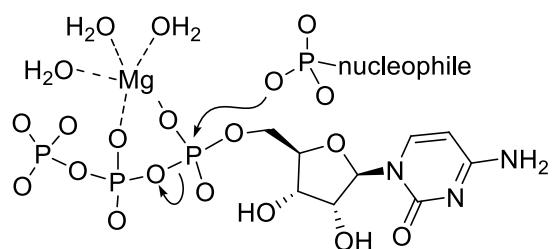


ATP and metal ion coordination in the representative structure 3ZCB.

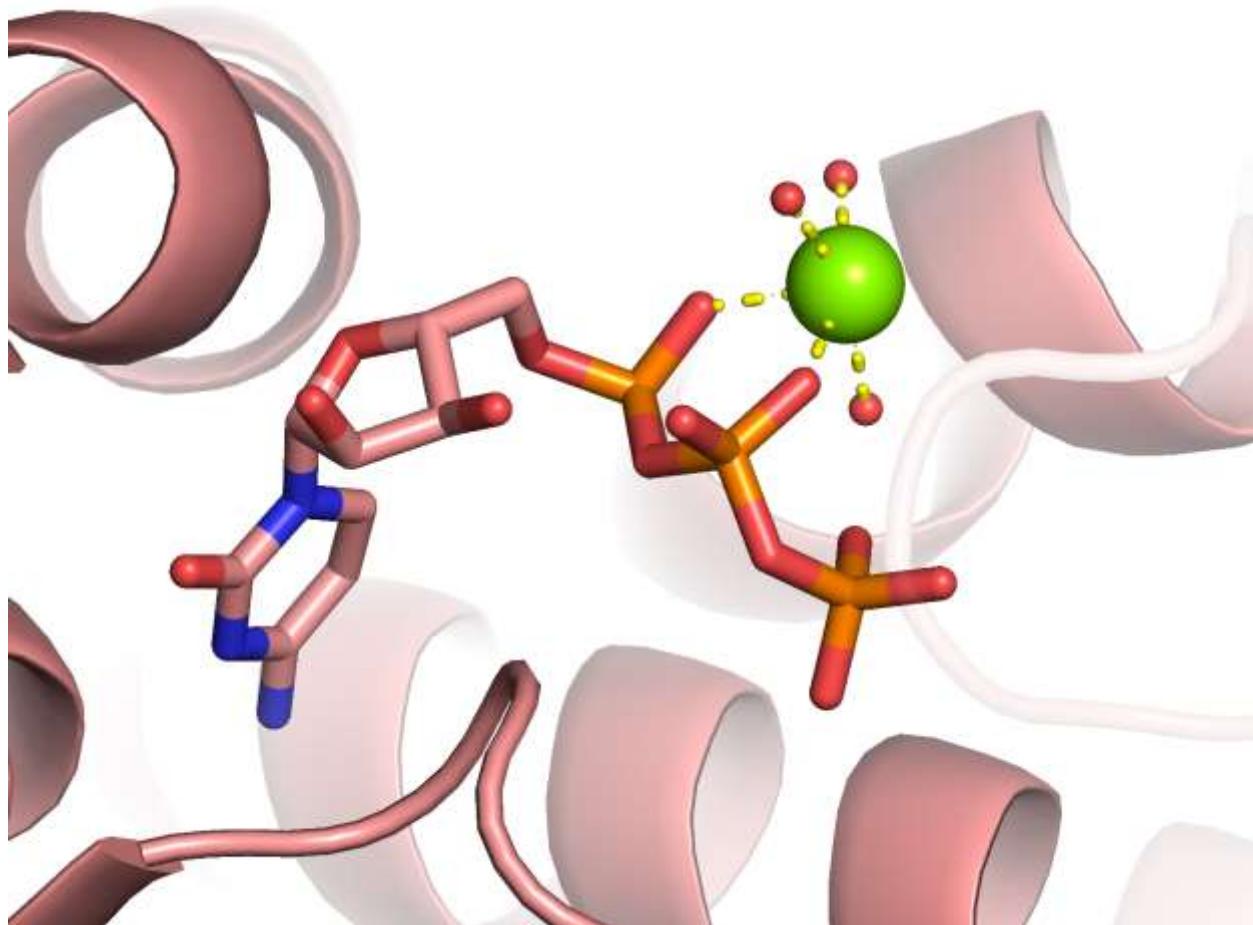
### Mitochondrial carrier

Known EC: 2.7.7.67 (CDP-2,3-bis-(O-geranylgeranyl)-sn-glycerol synthase)

In our dataset only 1 structure is found for this superfamily. The authors describe that a single  $Mg^{2+}$  ion is present that is coordinated by oxygens of the  $\alpha\beta$  phosphate groups.[39]. The substrate is missing from the structure, there are additional 3 water molecules coordinating the  $Mg^{2+}$  ion.



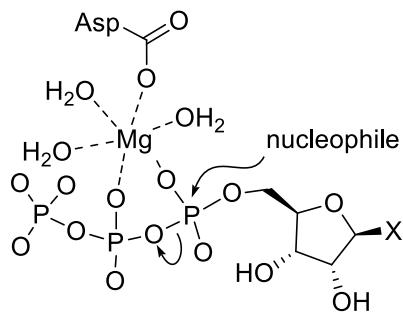
Cytidylyl transfer reaction for EC 2.7.7.67.



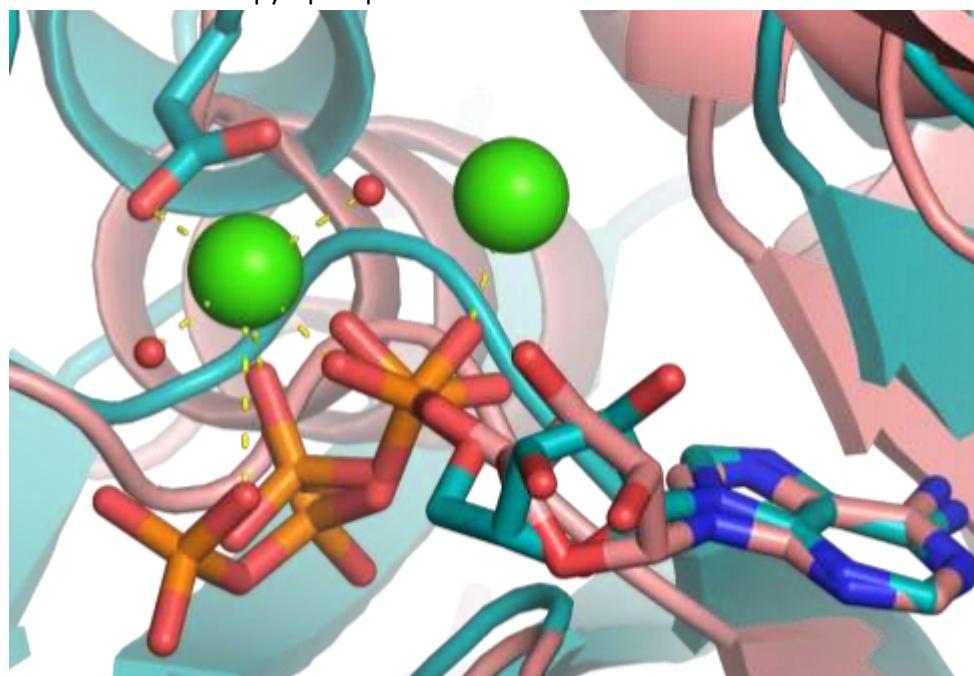
CTP and metal ion coordination in the representative structure 5GUF.

#### Molybdenum cofactor biosynthesis proteins

There is one available (PDB 4CTA) that contains ATP, and several other structures (PDB 5ERR, 5ERT, 5ERM, 5ERV, 6FGD, 6HSU, 6HSO) were resolved in the presence of ADP. Using these structures and the selected representative PDB 4CTA, a clear  $\alpha\beta$ -coordination can be identified for the  $Mg^{2+}$ , which is also coordinated by an aspartate residue and 3 water molecules. The aspartate that binds the  $Mg^{2+}$  is absolutely conserved in CinA sequence alignments.[40]



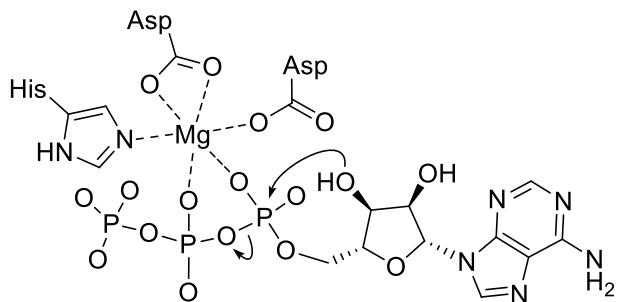
Proposed metal ion coordination and reactivity. In the associated ECs (2.7.7.75, 2.7.7.76, 4.6.1.17), different nucleosides all release pyrophosphate.



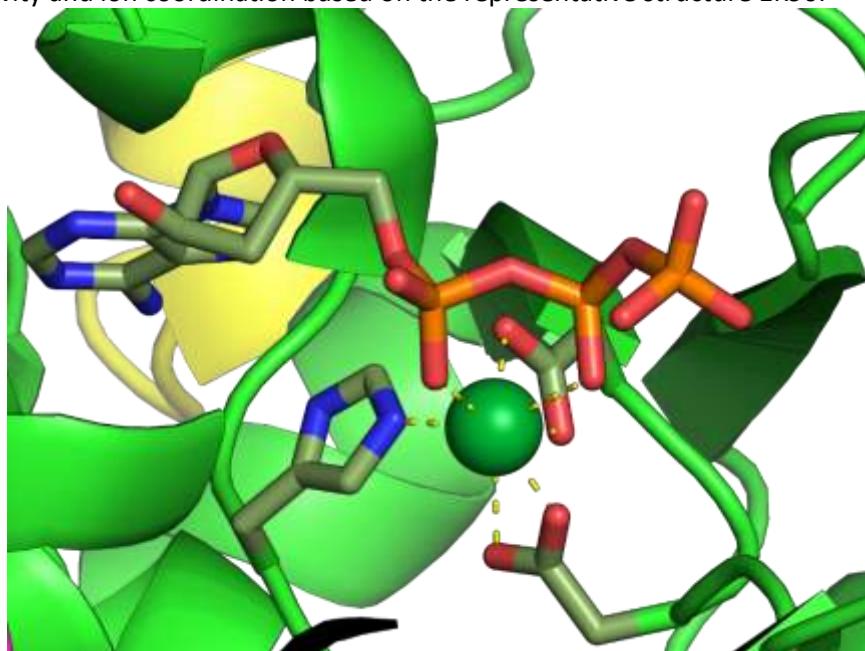
Nucleoside and metal ion coordination in the overlay of the ATP bound structure (4CTA, teal) and an ADP bound structure (salmon, 6FGD).

### **Adenylylcyclase toxin (the edema factor)**

Edema factor (EF) is a component of anthrax toxin produced by *Bacillus anthracis*.[41] It is a calcium- and calmodulin-dependent adenylyl cyclase (EC 4.6.1.1) that significantly increases host intracellular cAMP levels, causing edema (fluid-filled swelling) and interfering with host intracellular signaling. The C-terminal region of EF contains the calmodulin-dependent activation domain and the catalytic site.[42] The cyclase reaction produces inorganic  $PP_i$ . The consensus coordination shows a  $Mg^{2+}$  ion coordinated by the  $\alpha$  and  $\beta$  phosphates, further coordinated by two aspartate and a histidine residues.



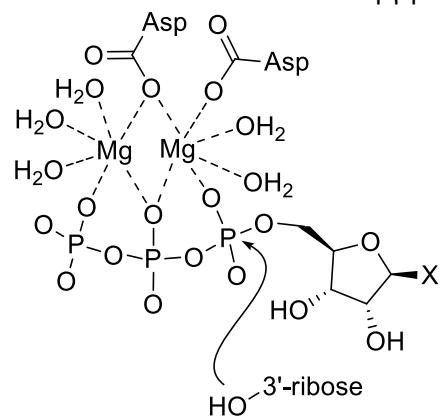
ATP cyclase activity and ion coordination based on the representative structure 1K90.



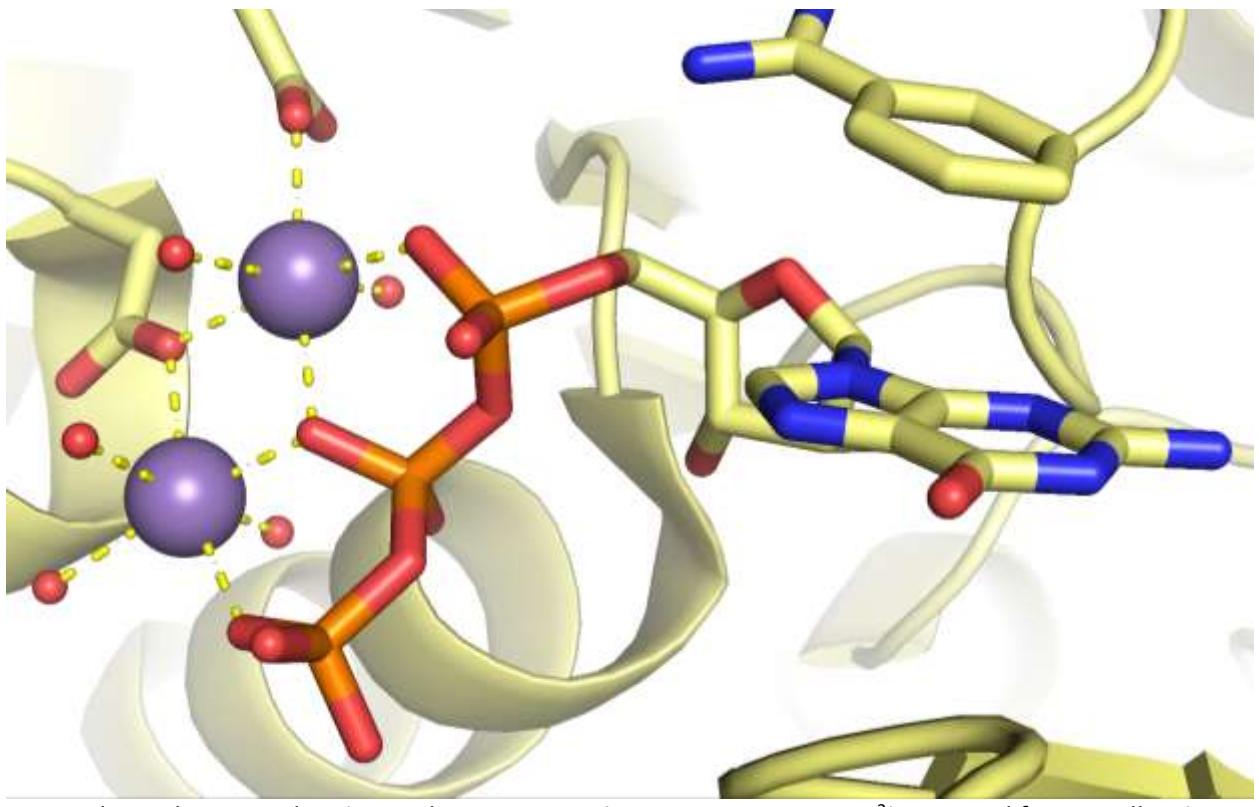
ATP and metal ion coordination in the representative structure 1K90. The structure is resolved with Yb<sup>3+</sup>, which reduces cyclase activity.

#### DNA primase core

In our dataset only a few structures are found for this superfamily, some originating from the same study. The structures have 2 catalytic divalent metal ions. One is coordinated by the  $\alpha\beta$  and the other by the  $\beta\gamma$  phosphate group oxygens. The metal ion in the  $\alpha\beta$  position is coordinated by two aspartates, while one of the aspartates also coordinates the metal ion in the  $\beta\gamma$  position.



Primase activity (EC 2.7.7.101) and metal ion coordination scheme.



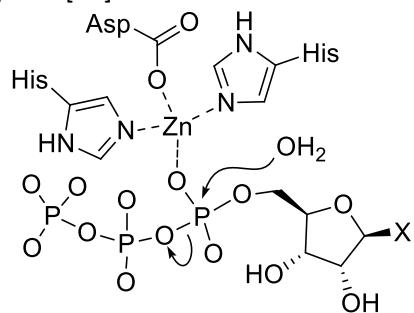
NTP and metal ion coordination in the representative structure 4EDK.  $Mn^{2+}$  was used for crystallization, the enzyme functions with both  $Mg^{2+}$  and  $Mn^{2+}$ .

## Additional pyrophosphatases

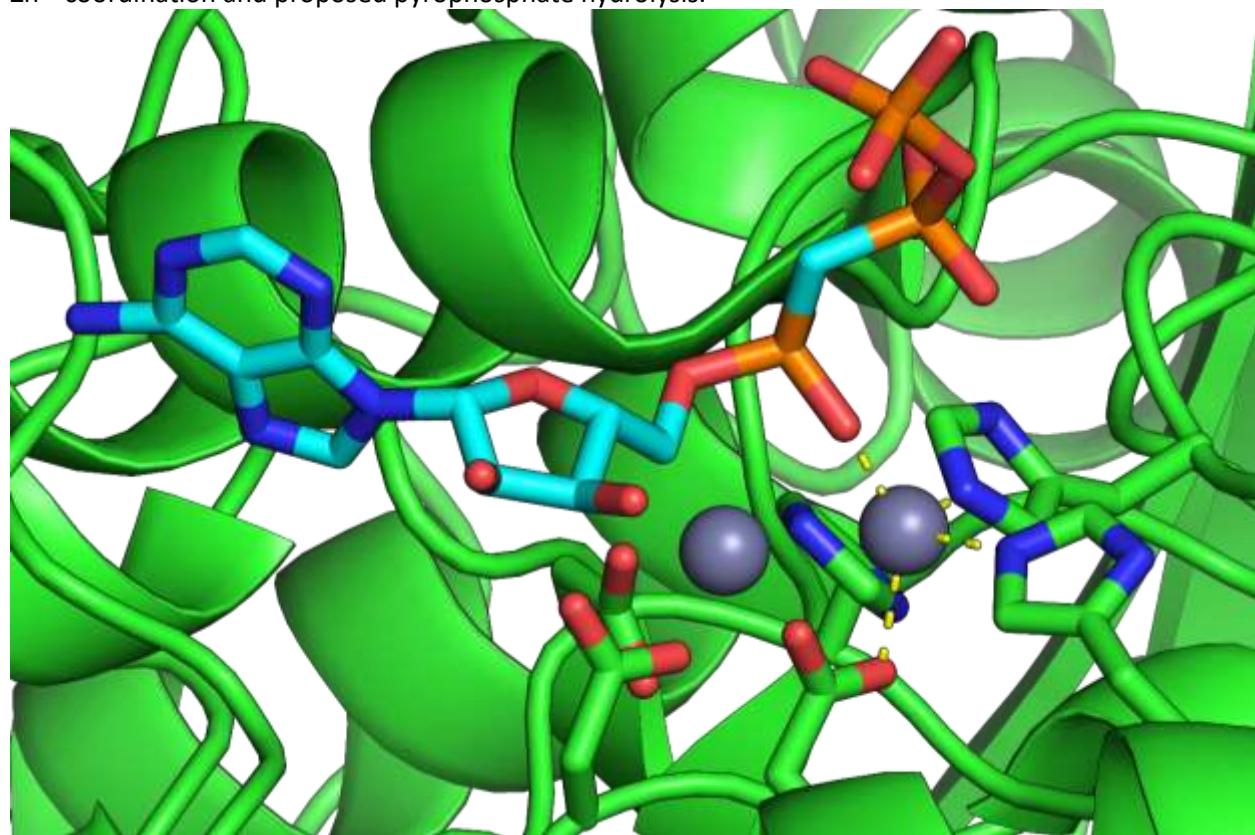
### Alkaline phosphatase-like

Known ECs: 3.6.1.9

There is a family, ecto-nucleotide pyrophosphatase/phosphodiesterase (NPP) within the Alkaline phosphatase-like superfamily the members of which may catalyze ATP hydrolysis,[43] this family corresponds to the InterPro entry IPR002591 (Type I phosphodiesterase/nucleotide pyrophosphatase/phosphate transferase). The human NPP family can be classified into two groups according to their substrate preferences; the nucleotide-degrading proteins NPP1, 3 and 4, and NPP2, 6, and 7 that hydrolyze phospholipids or related molecules.[43] The catalytic center consists of two zinc ions held by seven protein side chains; the catalytic cycle of NPPs is initiated by binding of a substrate (e.g. ATP) to the zinc ions via its  $\alpha$ -phosphate group.[44] Zn2 activates the catalytic threonine by greatly lowering its pKa [45] and allowing it to carry out a nucleophilic attack on the phosphorus atom, followed by departure of a leaving group[46] facilitated by Zn1.[43]



Zn<sup>2+</sup> coordination and proposed pyrophosphate hydrolysis.

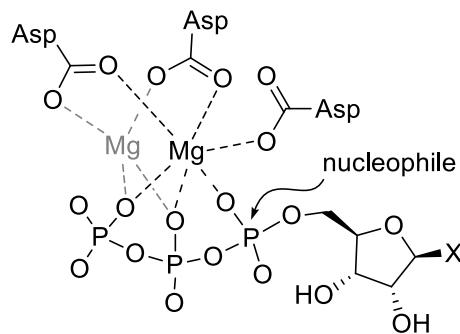


Active site structure of 6C02.

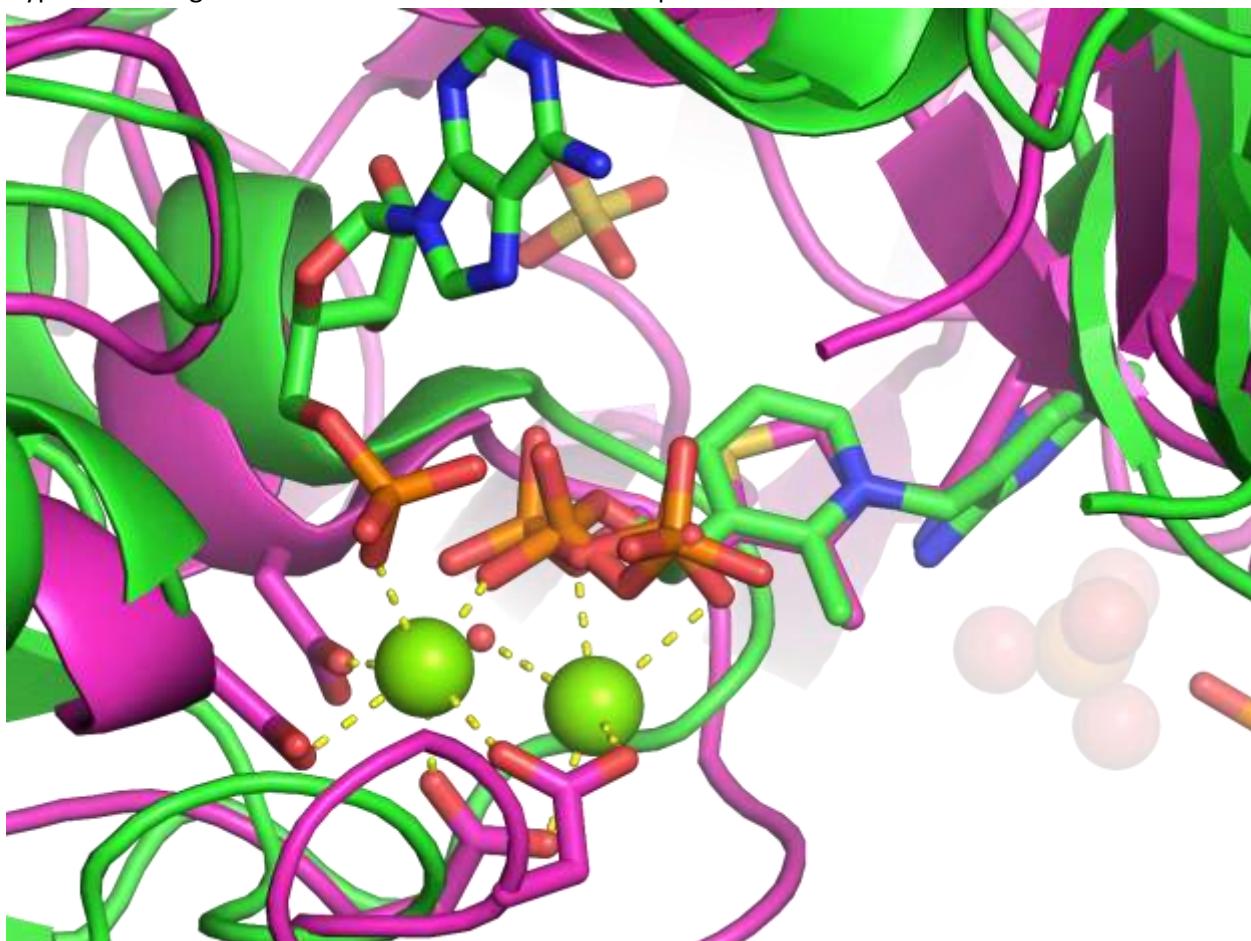
### Thiamin pyrophosphokinase, catalytic domain

Known ECs: 2.7.6.2, 2.7.6.3

In the available structures we identified for this superfamily already the product state is present with thiamine diphosphate and AMP. Yet, based on the position of the  $Mg^{2+}$  the available product structures it seems clear that there is an  $\alpha\gamma$ -coordinated  $Mg^{2+}$  (as seen e.g. in PDB 2F17), which might be  $\alpha\beta\gamma$  before the reaction. Furthermore, based on other structures (where the AMP is missing, yet other  $Mg^{2+}$  ions are present) we hypothesize that a second,  $\beta\gamma$ -coordinated  $Mg^{2+}$  may be present at the active site. There are altogether 4 aspartate residues coordinating the metal ions.



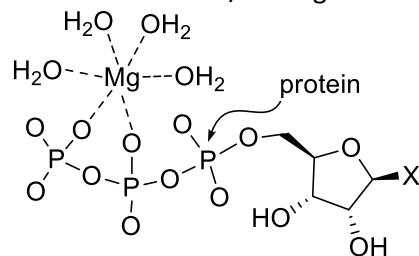
Hypothetical  $Mg^{2+}$  coordination and reaction based on product structures.



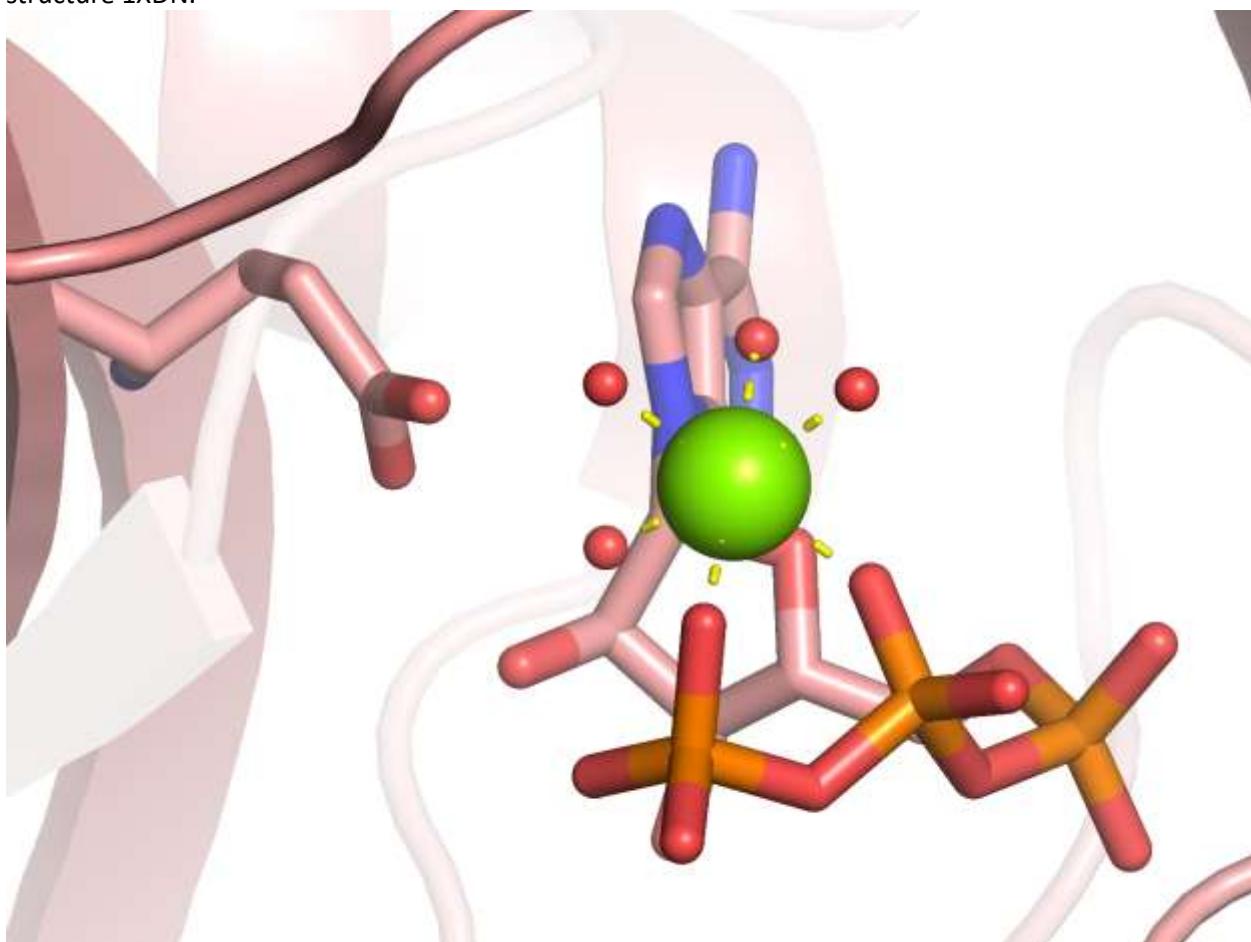
Active site reconstruction overlaying the AMP bound structure 2F17 and the pyrophosphorylated product 3IHK.

### DNA ligase/mRNA capping enzyme, catalytic domain

A two metal-ion mechanism was proposed to occur for the DNA and RNA ligases of this superfamily. A first  $Mg^{2+}$  ion may facilitate the precise positioning of the  $\beta$  and  $\gamma$ -phosphate groups, while another  $Mg^{2+}$  ion, which is not directly coordinating the cleavage of the bond between the departing PPi group and the rest of the ATP. It binds to the  $\alpha$ -phosphate group of the ATP and promotes the formation of a covalent bond with a conserved lysine residue (such a second  $Mg^{2+}$  ion is not present in the representative structure[47-49]) In our dataset, the position of a divalent metal ion binding to the  $\alpha$ -phosphate group of ATP is well conserved, however most probably due to the lack of catalytically active complexes as the DNA or RNA is missing in the structures, the triphosphate chain of the ATP has a large degree of freedom and shows considerable conformational variability together with the other metal ion in the  $\beta\gamma$  position. There is no metal ion neither in the  $\alpha\beta$  or  $\alpha\beta\gamma$  position in the structures in our dataset, however we cannot exclude that this is due to the lack of the presence of the DNA/RNA ligand.



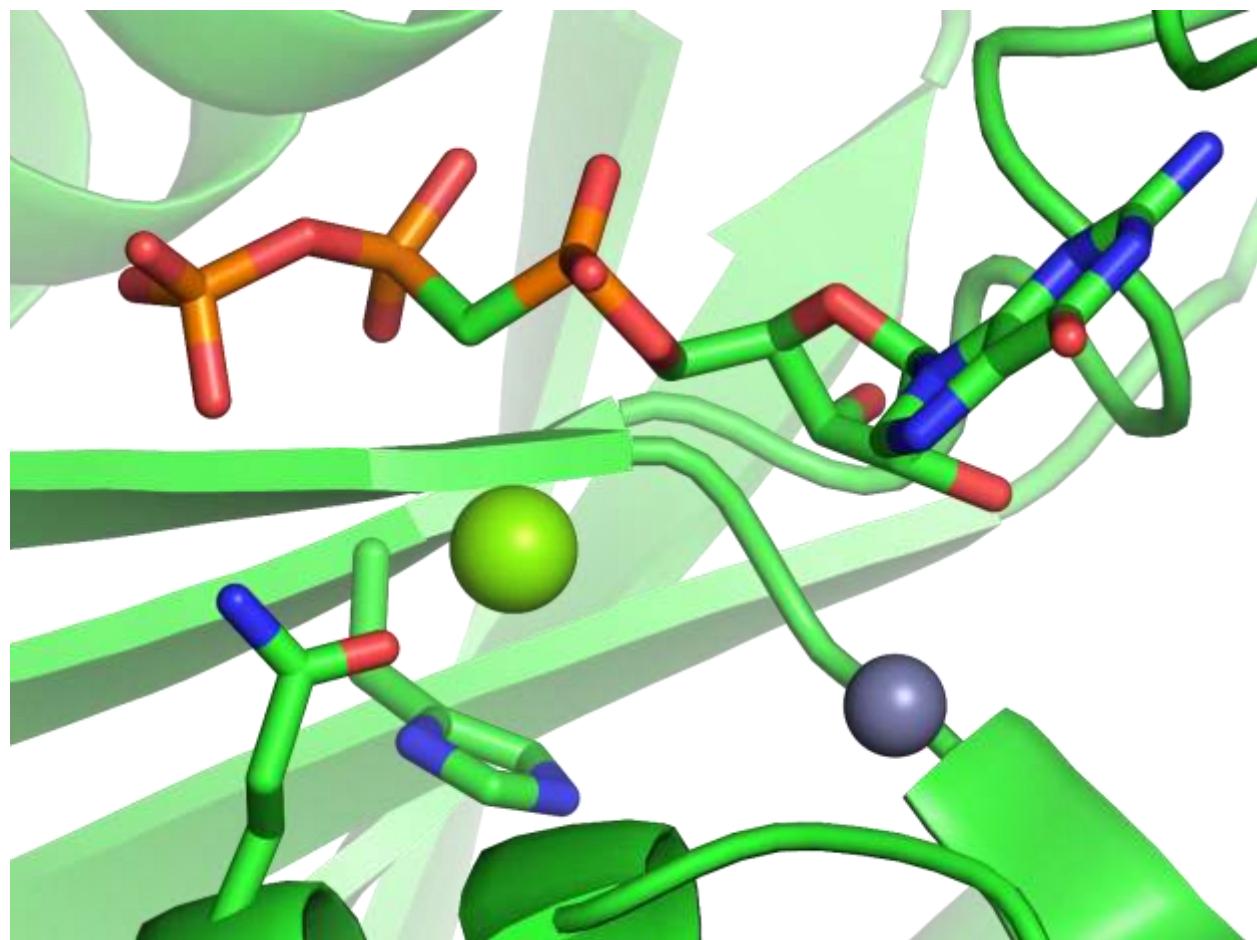
Protein adenylation as the first step of DNA ligase activity. The ion coordination is adopted from structure 1XDN.



ATP and Mg<sup>2+</sup> coordination in the structure 1XDN.

#### RibA-like

There is only one available structure with a triphosphate chain for this superfamily (PDB 2BZ0). Due to the usage of a GTP analog, GMPcPP, it is not evident to deduce the metal ion coordination, yet the most probable coordination seems to be  $\alpha\beta$ .



Active site structure of 2BZ0.

## Other superfamilies without sufficient structural information:

### Glycerate kinase I

Known ECs: 2.7.1.31 (glycerate 3-kinase), 2.7.1.165 (glycerate 2-kinase)

Similar folds, but low sequence identity, no structure with phosphates [50].

### *YgbK*-like

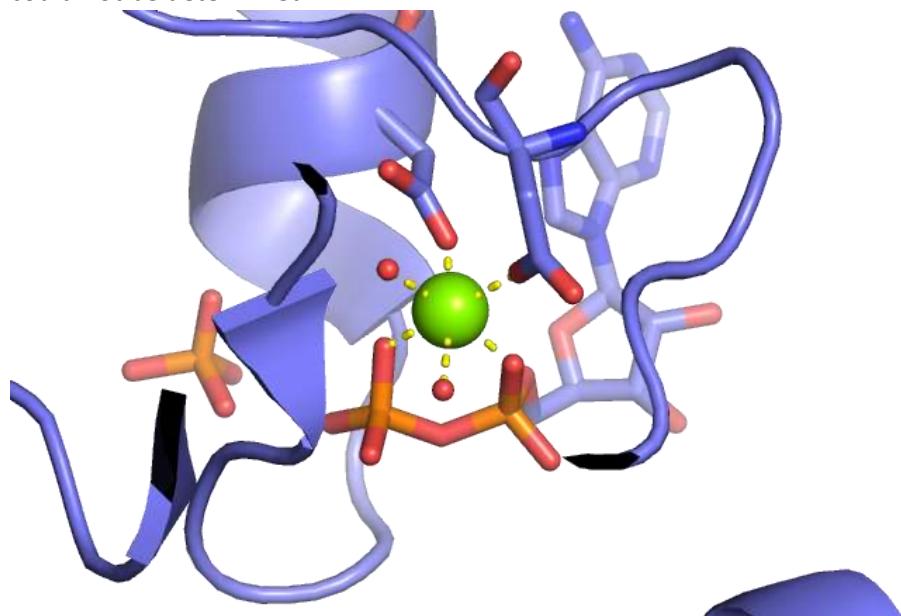
Known ECs: 2.7.1.217, 2.7.1.219, 2.7.1.220, 2.7.1.231 (various kinases)

There are only seven PDB structures associated with the InterPro entry Four-carbon acid sugar kinase, nucleotide binding domain superfamily (IPR042213). They align well, at their active sites an ADP and no ions are present.

### NagB/RpiA/CoA transferase-like

Known EC: 6.3.3.2 (5-formyltetrahydrofolate cyclo-ligase)

The only available structures are in the presence of ADP with  $\alpha\beta$  coordinated ions, hence the ATP binding could not be determined.



Active site structure of 2JCB.

### HIT-like

Known ECs 2.7.7.10, 2.7.7.53, 3.6.1.29, 6.3.2.39

The HIT-like SF covers the synthesis and breakdown of bisadenosyl tri- and tetraphosphates and similar molecules, hardly falling into any category. The PDB files do not contain metal ions, the coordination could not be established.

### tRNA(Ile2) 2-*agmatinyl*cytidine synthetase TiaS

Known EC: 6.3.4.22

Unfortunately for this superfamily the available structures, one in the presence of ATP (PDB 3AMT) and three structures resolved with the ATP analog, AMPCPP (PDBs 3AMU, 4RVZ, 6AGG), do not enable us to

determine the corresponding metal ion coordination. The former two PDBs lack the coordinating metal ion whereas the quality of the latter two is not sufficient for our analysis.

#### **IPR037997 CTP-dependent diacylglycerol kinase 1-like**

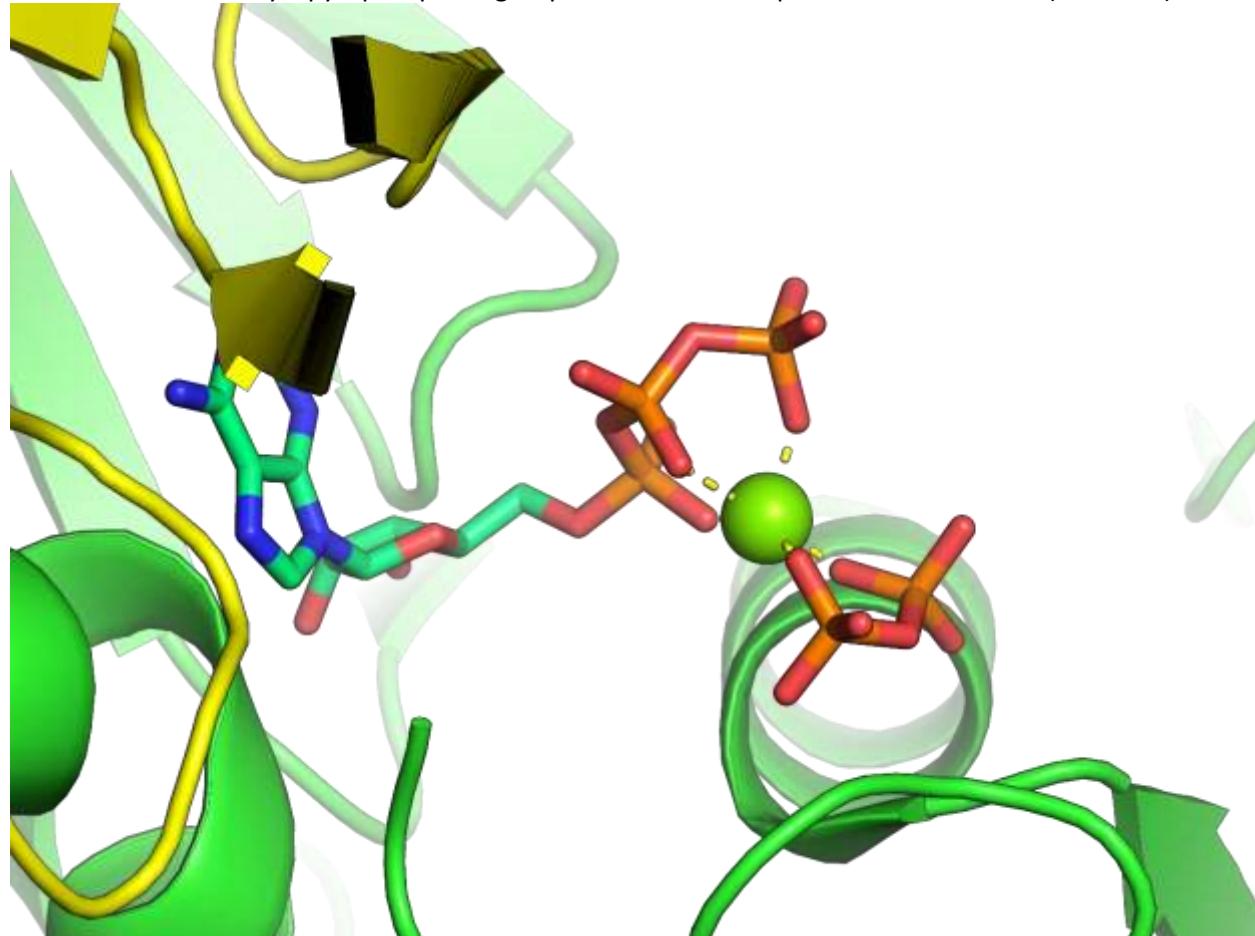
Known ECs: 2.7.1.174 and 2.7.1.182

We have not identified any available experimental structures for this family (defined on the by InterPro), it was not possible to deduce the corresponding metal ion coordination.

#### **NAD kinase/diacylglycerol kinase-like**

Known ECs: 2.7.1.-

ATP-NAD kinases (2.7.1.23) catalyze the phosphorylation of NAD to NADP utilizing ATP and other nucleoside triphosphates as well as inorganic polyphosphate as a source of phosphorus. NADH kinases (2.7.1.86) are specific for NADH where CTP, ITP, UTP and GTP can also act as phosphate donors. ATP-diacylglycerol kinases (2.7.1.107) are involved in synthesis of membrane phospholipids and the neutral lipid triacylglycerol. The active site encompasses one  $Mg^{2+}$  ion which displays an  $\alpha\beta\gamma$ -coordination, and is further coordinated by a pyrophosphate group in the selected representative structure (PDB 1z0s).

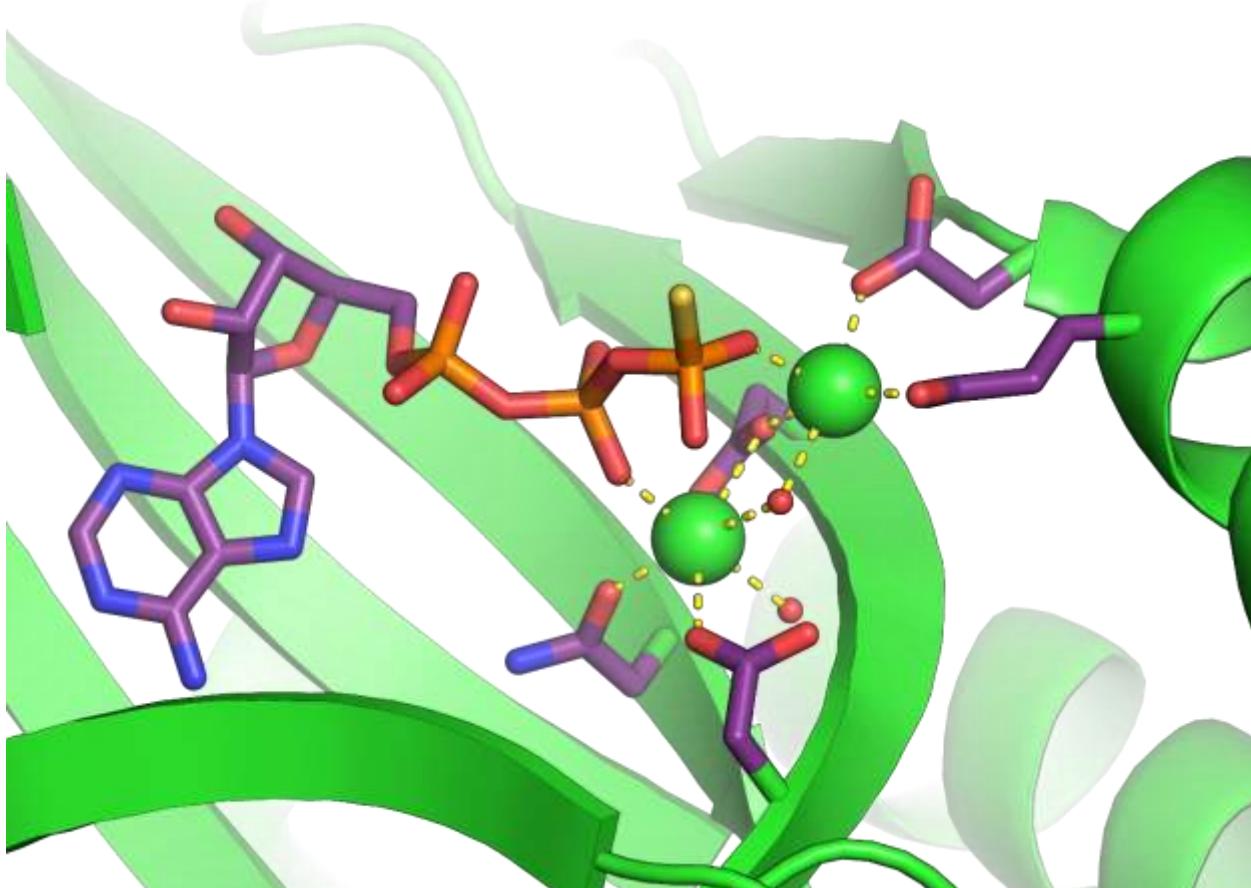


Substrate and  $Mg^{2+}$  binding in the representative structure, 1z0s.

#### **FomD-like**

Known ECs: 3.6.1.15

This group of prokaryotic proteins includes Nucleoside triphosphate/diphosphate phosphatase from *Staphylococcus aureus* (SA1684 or Ntdp), Cytidylyl-2-hydroxypropylphosphonate hydrolase from *Streptomyces wedmorensis* (FomD), and Probable ribonuclease FAU-1 from *Pyrococcus abyssi* [51]. We identified two metal ions at the active site, one which is coordinated by the  $P_\beta$  and the other by the  $P_\gamma$  [51]. Yet, the available structures do not make it possible to establish a clear consensus on the coordination.



Substrate and metal ion binding in the representative structure , 7d8iA.

#### **Phospholipase D/nuclease**

There is one structure available with in the presence of NTP and metal ions at the active site (PDB 1xdp). Two  $Mg^{2+}$  ions are very (likely erroneously) close to each other (3.1 Å apart), one close to the  $\alpha$  while the other close to the  $\gamma$  phosphate. Their coordination sphere is not complete.

#### **Nicotinate/Quinolinate PRTase C-terminal domain-like**

Known ECs: 6.3.4.21 (Nicotinate phosphoribosyltransferase)

The nicotinate phosphoribosyltransferase catalyzes the synthesis of beta-nicotinate D-ribonucleotide from nicotinate and 5-phospho-D-ribose 1-phosphate at the expense of ATP, even though it can also form  $\beta$ -nicotinate D-ribonucleotide and diphosphate from nicotinate and 5-phospho- $\alpha$ -D-ribose 1-diphosphate in the absence of ATP. Yet, the utilization of ATP results in a much lower  $K_m$  for nicotinate and shifts the reaction towards the products [52]. Unfortunately, there are no available structures in the presence of ATP, the coordination could not be established for this superfamily.

## **Apyrase**

Known ECs: 3.6.1.5 (acting on NTPs), 3.6.1.6 (acting on NDPs)

Apyrases are active against both di- and triphosphate nucleotides (NDPs and NTPs) and hydrolyse NTPs to nucleotide monophosphates (NMPs) in two distinct successive phosphate-releasing steps, with NDPs as intermediates. The eukaryotic enzymes require  $\text{Ca}^{2+}$ , but  $\text{Mg}^{2+}$  can substitute. [53, 54]. Unfortunately, there are no available structures in the presence of NTP, the coordination could not be established. The DALI alignment of the SF representatives revealed a surprisingly close resemblance to the *Alkaline phosphatase PhoX* family (InterPro entry IPR008557), the only protein family within the *Calcium-dependent phosphotriesterase* SF (b.67.3) capable of NTP processing. The *Alkaline phosphatase PhoX* family has a complex active-site with  $\text{Ca}^{2+}$  ions playing the pinching roles in BG(+) and BG(–) positions.

## **Transglutaminase, two C-terminal domains**

Known ECs: 2.3.2.13

The calcium-dependent multi-functional transglutaminase 2 possesses protein cross-linking and GTP hydrolysis activities [55, 56]. Unfortunately, there are no structures available in the presence of metal ions, hence the coordination could not be established.

## **S-adenosyl-L-methionine-dependent methyltransferases**

This category consists of S-adenosyl-L-methionine-dependent methyltransferases (SAM MTase), which is not an NTP processing reaction. However, some viral proteins that belong to this superfamily were suggested to possibly possess guanylyl-transferase (GTase) activity [57], where during the guanylyl-transfer process GMP is transferred to a 5'-end diphosphate viral RNA acceptor producing a pyrophosphate by-product [PDB 7fgg, 6z0u (Chikungunya Virus): [58, 59]; PDB 4v03, 5dto, 8gzp (Dengue Virus):[60-62]; PDB 7v1h (Omsk hemorrhagic fever virus): [63]]. The exact coordination could not be established for this superfamily, as the coordination spheres of the bound  $\text{Mg}^{2+}$  only involve the phosphate groups. PDB 4v0r and 5dto suggest a  $\beta$ -coordination, whereas 7v1h suggests an  $\alpha\beta$ -coordination, which we think would be the favorable arrangement for its putative pyrophosphatase activity.

## **RPB5-like RNA polymerase subunit**

Known ECs: 2.7.7.6

This category can be found in prokaryotic subunit H and the C terminus of eukaryotic RPB5. Prokaryotes contain a single DNA-dependent RNA polymerase (RNAP) that is responsible for the transcription of all genes, while eukaryotes have three classes of RNAPs (I-III). They catalyze the DNA-template-directed extension of the 3'- end of an RNA strand by one nucleotide at a time, while releasing pyrophosphate as byproduct. There are no structures available with nucleotides and ions that could help us definitely conclude on the metal ion coordination. Yet since all other polymerases exhibit the same typical metal coordination, we predict that this superfamily will also share the similarity in its active site, with a  $\text{Mg}^{2+}$  coordinated by all three phosphate groups, and another  $\text{Mg}^{2+}$  that is coordinated by the 3' hydroxyl of the priming nucleotide and the  $\alpha$ -phosphate of the incoming NTP.

## **Phosphatidate cytidylyltransferase**

Known ECs: 2.7.7.41

There is no corresponding SUPFAM superfamily for this category, it is named matching the InterPro family IPR000374. Phosphatidate cytidylyltransferases are also known as CDP-diacylglycerol synthases catalyze the synthesis of CDP-diacylglycerol from CTP and phosphatidate. There is only one available structure for this family (PDB 4q2g), however it does not contain nucleotide, the coordination cannot be

established. However, the authors suggest a two-metal-ion catalytic mechanism for the Cds-mediated synthesis of CDP-DAG at the membrane–cytoplasm interface [64].

#### **Molybdenum cofactor biosynthesis protein C, MoaC**

Known ECs: 2.7.7.77, 4.6.1.17

This superfamily corresponds to the molybdenum cofactor biosynthesis protein MoaC from prokaryotes and eukaryotes, also known as cyclic pyranopterin monophosphate synthases. Molybdenum cofactor guanylyltransferases (EC 2.7.7.77) catalyze the guanylation of the molybdenum cofactor, which occurs only in prokaryotes. In bacteria the cyclic pyranopterin monophosphate synthase reaction (EC 4.6.1.17) is catalyzed by MoaC [65]. There are no available structures with metal ions to comment on the metal ion coordination of this superfamily.

#### **GTP cyclohydrolase MptA**

Known ECs: 3.5.4.39

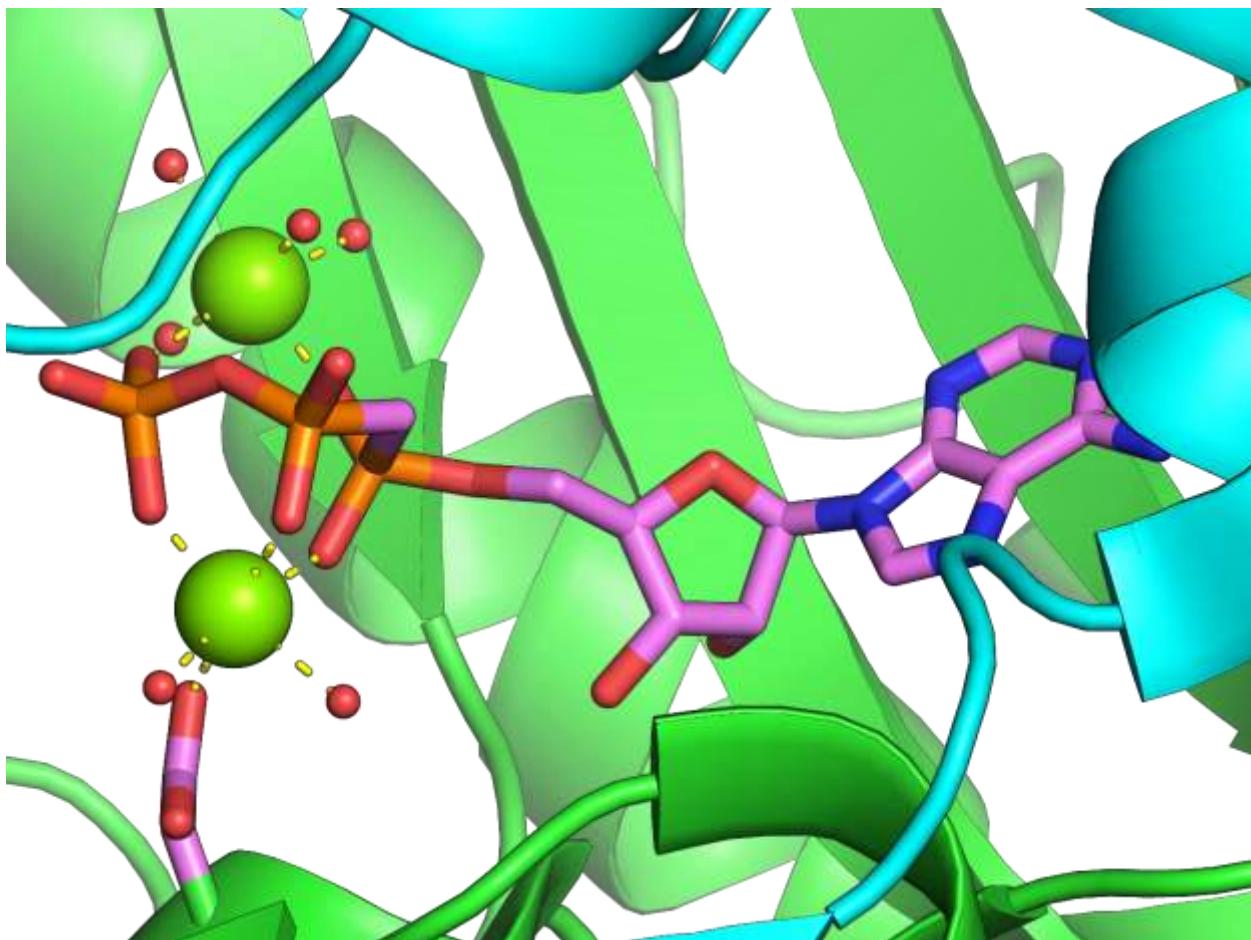
There are no available structures for this category which is defined as an InterPro family (IPR022840). MptA requires  $\text{Fe}^{2+}$  for activity [66]. Unfortunately, due to the lack of available structures, we are unable to comment on the metal ion coordination.

## Triphosphatase superfamilies

### S-adenosylmethionine synthetase

Known ECs: 2.5.1.6

S-adenosylmethionine synthetase is the enzyme that catalyzes the formation of S-adenosylmethionine from methionine and ATP, the sequence of S-adenosylmethionine synthetases is highly conserved throughout isozymes and species [67]. The reaction is suggested to initiate by the cleavage of C5'-O5' bond of ATP, caused by the action of a histidine residue; simultaneously C5' of ATP is the target of a nucleophilic attack by S<sub>6</sub> of Met, leading to the bond formation between Met and C5' of ATP to produce SAM and the release of triphosphate. The triphosphate is then hydrolyzed, producing pyrophosphate and P<sub>i</sub> [68]. There are two Mg<sup>2+</sup> ions in the active site, one is coordinated by all three phosphate groups on the (-) side, which may facilitate the second step of the reaction when the PPP<sub>i</sub> is hydrolyzed, while another is coordinated by P<sub>α</sub> and P<sub>γ</sub> on the opposite side.



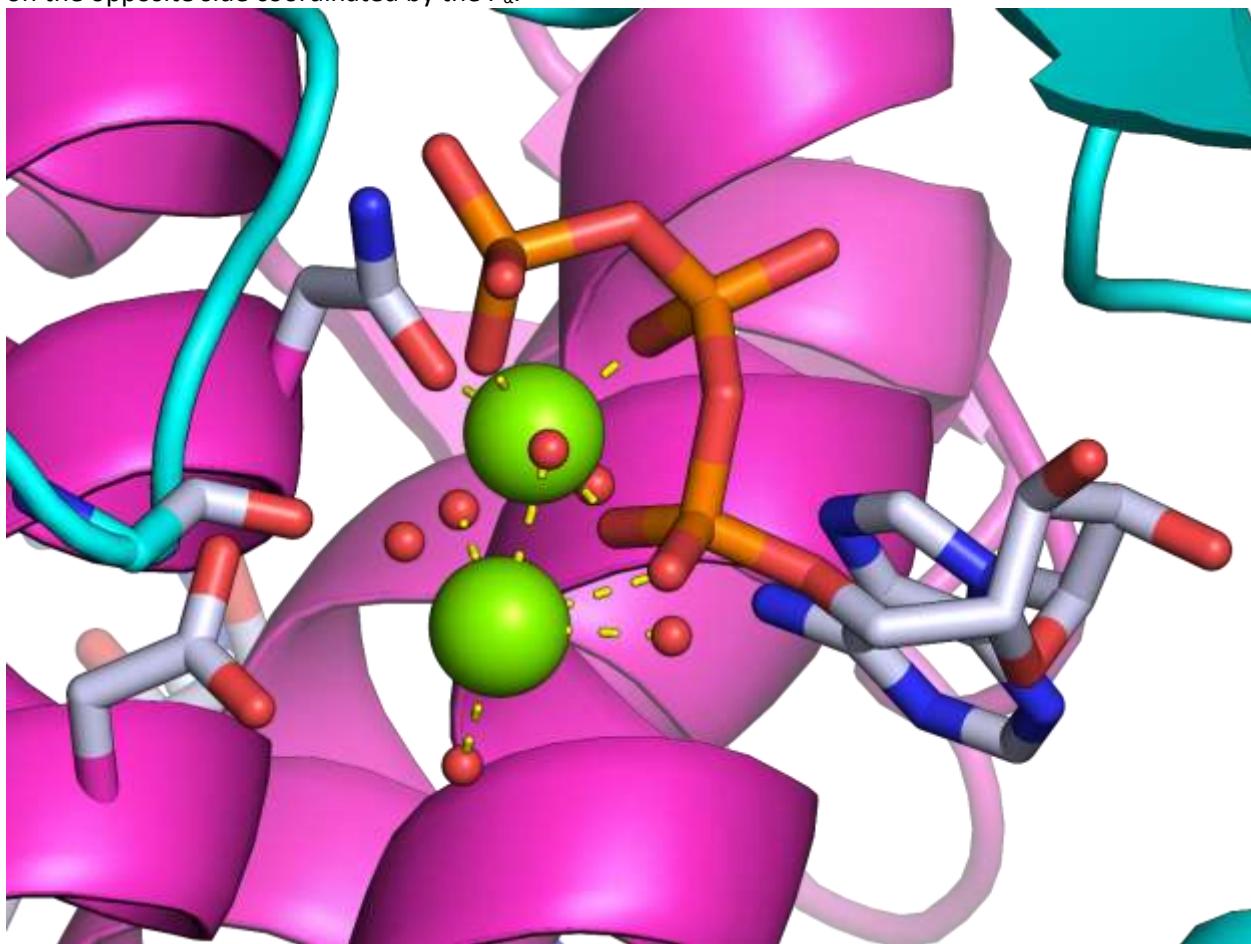
ATP and Mg<sup>2+</sup> binding in the representative structure, 6vd0.

### Cobalamin adenosyltransferase-like

Known ECs: 2.5.1.154, 2.5.1.17

Enzymes belonging to this superfamily catalyze the conversion of cobalamin (vitamin B12) into its coenzyme form, adenosylcobalamin (AdoCbl) or coenzyme B12 [69]. Depending on the EC, the triphosphate may be hydrolyzed into P<sub>i</sub> and PP<sub>i</sub> during catalysis (EC 2.5.1.154). We identified two metal

ions at the active site, one is coordinated by all three phosphate groups on the (+) side, while another is on the opposite side coordinated by the  $P_\alpha$ .



Substrate and metal ion binding in the representative structure , 6d5k.

#### Tetrahydrobiopterin biosynthesis enzymes-like

Known ECs: 4.2.3.12, 4.1.2.50

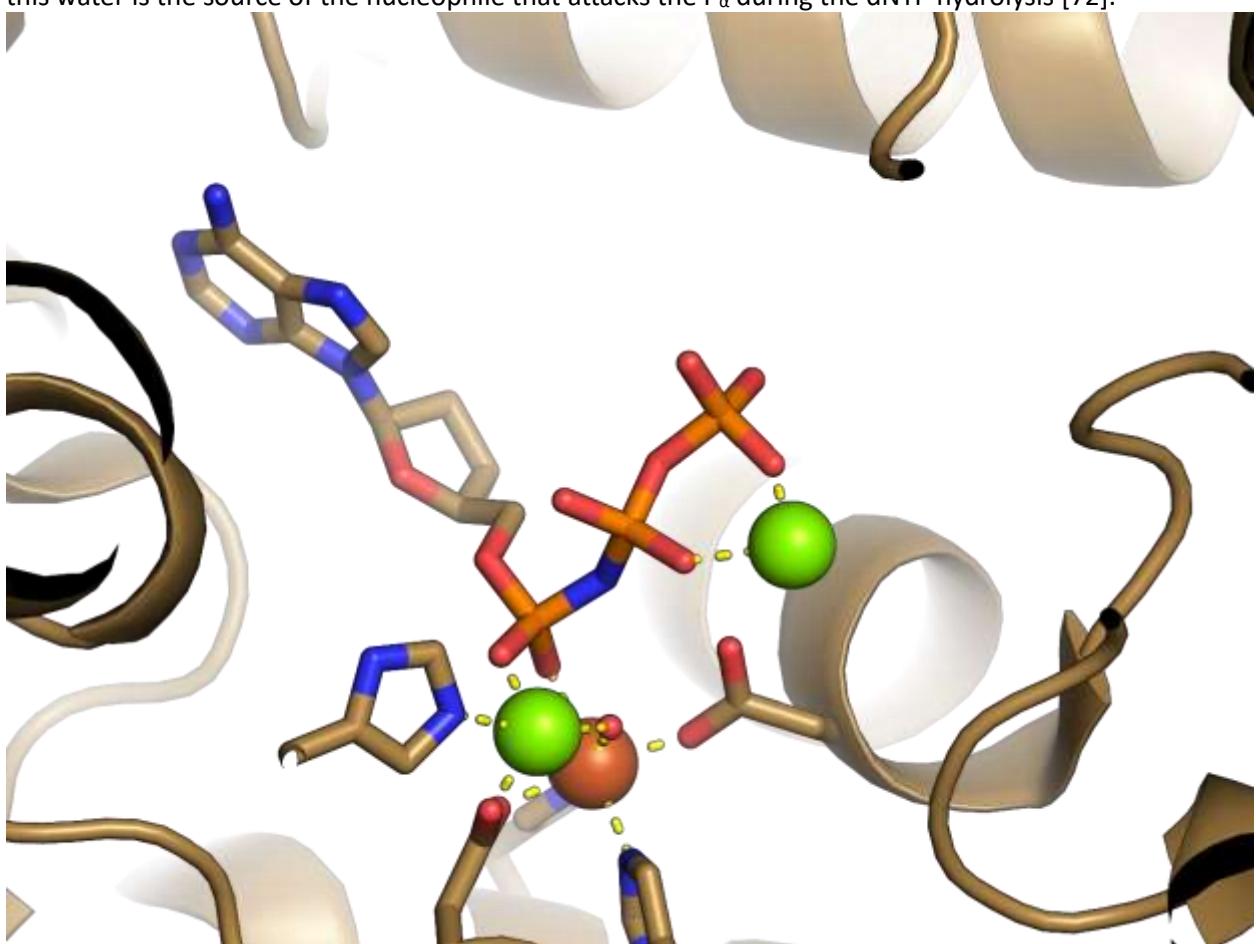
This category is not strictly NTP processing, as it processes 7,8-Dihydronoopterin 3'-triphosphate. During the reaction  $PPP_i$  is released, while 6-Pyruvyltetrahydropterin is produced. 6-Pyruvyl tetrahydrobiopterin synthase (PTPS) catalyzes the conversion of dihydronoopterin triphosphate to 6-pyruvyl tetrahydropterin, the second of three enzymatic steps in the synthesis of tetrahydrobiopterin from GTP [70]. A bound Zn(II) responsible for the enzymatic activity, its binding site is believed to be formed by three histidine residues [71]. Some members of this superfamily are GTP cyclohydrolase I (EC 3.5.4.16), which does not process the triphosphate chain of the GTP, hence is irrelevant for our analysis. All available structures in the presence of NTP and metal ion correspond to GTP cyclohydrolase I (they have a Zn ion coordinated by the  $P_\alpha$ ), however, we did not identify such structures for EC 4.2.3.12 or 4.1.2.50 hence the coordination could not be established.

#### HD-domain/PDEase-like

Known ECs: 3.1.5.1

Members of this superfamily having an HD-domain catalyze the hydrolysis of dNTPs into their constituent triphosphate and 2'-deoxynucleoside [72]. Based on the selected representative structure PDB 6txe, there are three octahedrally coordinated metal ions (2  $Mg^{2+}$  and an Fe) that support catalysis.

A  $Mg^{2+}$  is coordinated by  $P_\beta$  and  $P_\gamma$ , and both the second  $Mg^{2+}$  and the Fe are  $\alpha$ -coordinated. A crystal water bridges the two metal ions with an approximately tetrahedral Fe–O–Mg bond angle, and possibly this water is the source of the nucleophile that attacks the  $P_\alpha$  during the dNTP hydrolysis [72].



Substrate and metal ion binding in the representative structure, 6tx0.

## References

- 1 Dürr, S.L., *et al.* (2021) The Role of Conserved Residues in the DEDDh Motif: the Proton-Transfer Mechanism of HIV-1 RNase H. *ACS Catalysis* 11, 7915-7927
- 2 Brunle, S., *et al.* (2019) Molybdate pumping into the molybdenum storage protein via an ATP-powered piercing mechanism. *Proc Natl Acad Sci U S A* 116, 26497-26504
- 3 Gallagher, D.T., *et al.* (2011) Active-Site Structure of Class IV Adenylyl Cyclase and Transphyletic Mechanism. *Journal of Molecular Biology* 405, 787-803
- 4 Li, H., *et al.* (2003) CofE Catalyzes the Addition of Two Glutamates to F420-0 in F420 Coenzyme Biosynthesis in *Methanococcus jannaschii*. *Biochemistry* 42, 9771-9778
- 5 Vorobiev, S., *et al.* (2003) The structure of nonvertebrate actin: implications for the ATP hydrolytic mechanism. *Proc Natl Acad Sci U S A* 100, 5760-5765
- 6 Chastain, C.J., *et al.* (2011) Functional evolution of C4 pyruvate, orthophosphate dikinase. *Journal of Experimental Botany* 62, 3083-3091
- 7 McNae, I.W., *et al.* (2009) The Crystal Structure of ATP-bound Phosphofructokinase from *Trypanosoma brucei* Reveals Conformational Transitions Different from those of Other Phosphofructokinases. *Journal of Molecular Biology* 385, 1519-1533
- 8 Nogales, E., *et al.* (1998) Tubulin and FtsZ form a distinct family of GTPases. *Nature Structural Biology* 5, 451-458
- 9 Boneca, I.G., *et al.* (2016) Crystallographic Study of Peptidoglycan Biosynthesis Enzyme MurD: Domain Movement Revisited. *Plos One* 11
- 10 Basu, M.K., *et al.* (2011) ProPhylo: partial phylogenetic profiling to guide protein family construction and assignment of biological process. *BMC Bioinformatics* 12
- 11 Nayak, D.D., *et al.* (2017) Post-translational thioamidation of methyl-coenzyme M reductase, a key enzyme in methanogenic and methanotrophic Archaea. *eLife* 6
- 12 Mahanta, N., *et al.* (2018) Enzymatic reconstitution of ribosomal peptide backbone thioamidation. *Proceedings of the National Academy of Sciences* 115, 3030-3035
- 13 Xiang, Y., *et al.* (2009) Crystallographic Insights into the Autocatalytic Assembly Mechanism of a Bacteriophage Tail Spike. *Molecular Cell* 34, 375-386
- 14 Kim, S.H., *et al.* (2021) Structural basis for the substrate specificity and catalytic features of pseudouridine kinase from *Arabidopsis thaliana*. *Nucleic Acids Res* 49, 491-503
- 15 Shumilin, I.A., *et al.* (2012) Identification of unknown protein function using metabolite cocktail screening. *Structure* 20, 1715-1725
- 16 Kang, P.A., *et al.* (2019) Crystal structure and mutational analyses of ribokinase from *Arabidopsis thaliana*. *J Struct Biol* 206, 110-118
- 17 Yong, S.C., *et al.* (2014) A complex iron-calcium cofactor catalyzing phosphotransfer chemistry. *Science* 345, 1170-1173
- 18 Walsh, J.P. and Bell, R.M. (1986) sn-1,2-Diacylglycerol kinase of *Escherichia coli*. Mixed micellar analysis of the phospholipid cofactor requirement and divalent cation dependence. *Journal of Biological Chemistry* 261, 6239-6247
- 19 Keppetipola, N. and Shuman, S. (2008) A Phosphate-binding Histidine of Binuclear Metallophosphodiesterase Enzymes Is a Determinant of 2',3'-Cyclic Nucleotide Phosphodiesterase Activity. *Journal of Biological Chemistry* 283, 30942-30949
- 20 Hvorecny, K.L., *et al.* (2023) Human PRPS1 filaments stabilize allosteric sites to regulate activity. *Nature Structural & Molecular Biology* 30, 391-402

- 21 Schweikhard, E.S., *et al.* (2010) Structure and Function of the Universal Stress Protein TeaD and Its Role in Regulating the Ectoine Transporter TeaABC of Halomonas elongata DSM 2581T. *Biochemistry* 49, 2194-2204
- 22 Bangera, M., *et al.* (2015) Structural and functional analysis of two universal stress proteins YdaA and YnaF from *Salmonella typhimurium*: possible roles in microbial stress tolerance. *Journal of Structural Biology* 189, 238-250
- 23 Kishimoto, A., *et al.* (2014) Crystal structure of phosphopantothenate synthetase from *Thermococcus kodakarensis*. *Proteins: Structure, Function, and Bioinformatics* 82, 1924-1936
- 24 Yokooji, Y., *et al.* (2009) Pantoate Kinase and Phosphopantothenate Synthetase, Two Novel Enzymes Necessary for CoA Biosynthesis in the Archaea. *Journal of Biological Chemistry* 284, 28137-28145
- 25 Gleghorn, M.L., *et al.* (2008) Structural Basis for DNA-Hairpin Promoter Recognition by the Bacteriophage N4 Virion RNA Polymerase. *Molecular Cell* 32, 707-717
- 26 Webster, L.T. (1965) Studies of the Acetyl Coenzyme A Synthetase Reaction. *Journal of Biological Chemistry* 240, 4164-4169
- 27 Jacewicz, A., *et al.* (2022) Structures of RNA ligase RtcB in complexes with divalent cations and GTP. *Rna* 28, 1509-1518
- 28 Johansson, E., *et al.* (2005) Structures of dCTP Deaminase from Escherichia coli with Bound Substrate and Product. *Journal of Biological Chemistry* 280, 3051-3059
- 29 Mostert, K.J., *et al.* (2021) The Coenzyme A Level Modulator Hopantenate (HoPan) Inhibits Phosphopantenoylecysteine Synthetase Activity. *ACS Chemical Biology* 16, 2401-2414
- 30 Chakravarty, A.K., *et al.* (2011) Structures of RNA 3'-phosphate cyclase bound to ATP reveal the mechanism of nucleotidyl transfer and metal-assisted catalysis. *Proceedings of the National Academy of Sciences* 108, 21034-21039
- 31 Chakravarty, A.K. and Shuman, S. (2011) RNA 3'-phosphate cyclase (RtcA) catalyzes ligase-like adenylylation of DNA and RNA 5'-monophosphate ends. *J Biol Chem* 286, 4117-4122
- 32 Li, C., *et al.* (2009) Polymerase Translocation with Respect to Single-Stranded Nucleic Acid: Looping or Wrapping of Primer around a Poly(A) Polymerase. *Structure* 17, 680-689
- 33 Evans, R.J., *et al.* (2008) Structure of PolC reveals unique DNA binding and fidelity determinants. *Proceedings of the National Academy of Sciences* 105, 20695-20700
- 34 Berta, D., *et al.* (2020) Cations in motion: QM/MM studies of the dynamic and electrostatic roles of H(+) and Mg(2+) ions in enzyme reactions. *Curr Opin Struct Biol* 61, 198-206
- 35 Steegborn, C., *et al.* (2005) Bicarbonate activation of adenylyl cyclase via promotion of catalytic active site closure and metal recruitment. *Nat Struct Mol Biol* 12, 32-37
- 36 Tokarsky, E.J., *et al.* (2017) Significant impact of divalent metal ions on the fidelity, sugar selectivity, and drug incorporation efficiency of human PrimPol. *DNA Repair* 49, 51-59
- 37 Gonzalez, S. (1999) Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *The EMBO Journal* 18, 3767-3775
- 38 Iyer, L.M., *et al.* (2003) *BMC Structural Biology* 3
- 39 Ren, S., *et al.* (2017) Structural and mechanistic insights into the biosynthesis of CDP-archaeol in membranes. *Cell Research* 27, 1378-1391
- 40 Karuppiah, V., *et al.* (2014) Structure and Mechanism of the Bifunctional CinA Enzyme from *Thermus thermophilus*. *Journal of Biological Chemistry* 289, 33187-33197
- 41 Collier, R.J. and Young, J.A.T. (2003) Anthrax Toxin. *Annual Review of Cell and Developmental Biology* 19, 45-70

- 42 Guo, Q., et al. (2004) Structural and Kinetic Analyses of the Interaction of Anthrax Adenylyl Cyclase Toxin with Reaction Products cAMP and Pyrophosphate. *Journal of Biological Chemistry* 279, 29427-29435
- 43 Gorelik, A., et al. (2018) Structural basis for nucleotide recognition by the ectoenzyme CD203c. *The FEBS Journal* 285, 2481-2494
- 44 Gijsbers, R., et al. (2001) Structural and Catalytic Similarities between Nucleotide Pyrophosphatases/Phosphodiesterases and Alkaline Phosphatases. *Journal of Biological Chemistry* 276, 1361-1368
- 45 Zalatan, J.G., et al. (2006) Structural and Functional Comparisons of Nucleotide Pyrophosphatase/Phosphodiesterase and Alkaline Phosphatase: Implications for Mechanism and Evolution. *Biochemistry* 45, 9788-9803
- 46 Hausmann, J., et al. (2016) Structural snapshots of the catalytic cycle of the phosphodiesterase Autotaxin. *Journal of Structural Biology* 195, 199-206
- 47 Deng, J., et al. (2004) High Resolution Crystal Structure of a Key Editosome Enzyme from Trypanosoma brucei: RNA Editing Ligase 1. *Journal of Molecular Biology* 343, 601-613
- 48 Cherepanov, A.V. and de Vries, S. (2002) Kinetic Mechanism of the Mg<sup>2+</sup>-dependent Nucleotidyl Transfer Catalyzed by T4 DNA and RNA Ligases. *Journal of Biological Chemistry* 277, 1695-1704
- 49 Shuman, S., et al. (2020) Caveat mutator: alanine substitutions for conserved amino acids in RNA ligase elicit unexpected rearrangements of the active site for lysine adenylylation. *Nucleic Acids Research* 48, 5603-5615
- 50 Schwarzenbacher, R., et al. (2006) Crystal structure of a glycerate kinase (TM1585) from Thermotoga maritima at 2.70 Å resolution reveals a new fold. *Proteins: Structure, Function, and Bioinformatics* 65, 243-248
- 51 Wang, Z., et al. (2021) The structural mechanism for the nucleoside tri- and diphosphate hydrolysis activity of Ntdp from Staphylococcus aureus. *The FEBS Journal* 288, 6019-6034
- 52 Vinitsky, A. and Grubmeyer, C. (1993) A new paradigm for biochemical energy coupling. *Salmonella typhimurium* nicotinate phosphoribosyltransferase. *Journal of Biological Chemistry* 268, 26004-26010
- 53 Wang, T.-F. and Guidotti, G. (1996) CD39 Is an Ecto-(Ca<sup>2+</sup>,Mg<sup>2+</sup>)-apyrase. *Journal of Biological Chemistry* 271, 9898-9901
- 54 Ashihara, H., et al. (2018) Purine salvage in plants. *Phytochemistry* 147, 89-124
- 55 Permyakov, E.A., et al. (2018) Structure of natural variant transglutaminase 2 reveals molecular basis of gaining stability and higher activity. *Plos One* 13
- 56 Han, B.-G., et al. (2010) Crystal structure of human transglutaminase 2 in complex with adenosine triphosphate. *International Journal of Biological Macromolecules* 47, 190-195
- 57 Issur, M., et al. (2009) The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. *Rna* 15, 2340-2350
- 58 Zhang, K., et al. (2022) Molecular basis of specific viral RNA recognition and 5'-end capping by the Chikungunya virus nsP1. *Cell Reports* 40
- 59 Jones, R., et al. (2020) Capping pores of alphavirus nsP1 gate membranous viral replication factories. *Nature* 589, 615-619
- 60 Rey, F.A., et al. (2015) A Crystal Structure of the Dengue Virus NS5 Protein Reveals a Novel Inter-domain Interface Essential for Protein Flexibility and Virus Replication. *PLOS Pathogens*

- 61 Zhao, Y., et al. (2015) Molecular basis for specific viral RNA recognition and 2'-O-ribose methylation by the dengue virus nonstructural protein 5 (NS5). *Proceedings of the National Academy of Sciences* 112, 14834-14839
- 62 Osawa, T., et al. (2023) Structures of dengue virus RNA replicase complexes. *Molecular Cell* 83, 2781-2791.e2784
- 63 Jia, H., et al. (2022) Crystal Structures of Flavivirus NS5 Guanylyltransferase Reveal a GMP-Arginine Adduct. *Journal of Virology* 96
- 64 Liu, X., et al. (2014) Structure and mechanism of an intramembrane liponucleotide synthetase central for phospholipid biosynthesis. *Nature Communications* 5
- 65 Hover, B.M., et al. (2015) Mechanism of pyranopterin ring formation in molybdenum cofactor biosynthesis. *Proceedings of the National Academy of Sciences* 112, 6347-6352
- 66 Grochowski, L.L., et al. (2007) Characterization of an Fe<sup>2+</sup>-Dependent Archaeal-Specific GTP Cyclohydrolase, MptA, from Methanocaldococcus jannaschii. *Biochemistry* 46, 6658-6667
- 67 Horikawa, S., et al. (1990) Molecular cloning and nucleotide sequence of cDNA encoding the rat kidney S-adenosylmethionine synthetase. *Journal of Biological Chemistry* 265, 13683-13686
- 68 Sekula, B., et al. (2020) S-adenosylmethionine synthases in plants: Structural characterization of type I and II isoenzymes from *Arabidopsis thaliana* and *Medicago truncatula*. *International Journal of Biological Macromolecules* 151, 554-565
- 69 Sheppard, D.E., et al. (2004) Evidence that a B
- 12
- Adenosyl Transferase Is Encoded within the Ethanolamine Operon of *Salmonella enterica*. *Journal of Bacteriology* 186, 7635-7644
- 70 Nar, H., et al. (1994) Three-dimensional structure of 6-pyruvoyl tetrahydropterin synthase, an enzyme involved in tetrahydrobiopterin biosynthesis. *The EMBO Journal* 13, 1255-1262
- 71 Bürgisser, D.M., et al. (1995) 6-Pyruvoyl Tetrahydropterin Synthase, An Enzyme With a Novel Type of Active Site Involving Both Zinc Binding and an Intersubunit Catalytic Triad Motif; Site-directed Mutagenesis of the Proposed Active Center, Characterization of the Metal Binding Site and Modelling of substrate Binding. *Journal of Molecular Biology* 253, 358-369
- 72 Morris, E.R., et al. (2020) Crystal structures of SAMHD1 inhibitor complexes reveal the mechanism of water-mediated dNTP hydrolysis. *Nature Communications* 11