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

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

† 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

Keywords
Phosphate processing enzymes, structural database, NTP, metal-ion coordination, QM/MM

Abstract

Phosphates are essential for all forms of life, playing key roles in DNA/RNA, signaling, energy storage and transfer, and biosynthetic processes. We conducted a quantitative analysis of nucleoside triphosphate (NTP) processing enzymes across all enzymatic reactions, revealing their dominance in phosphate reactivity with ATP as the most prevalent substrate. Two main reaction types occur predominantly: cleavage resulting in (i) pyrophosphate or (ii) phosphate release/transfer. The large majority of NTP processing enzymes require divalent Mg^{2+} ions in a mechanistically analogous manner. Despite their importance, the precise coordination of metal ions in the catalytically competent active site structure in many NTP processing enzymes remains elusive.

We hypothesized that efficient phosphate processing requires specific metal ion coordination, facilitating the catalytic reaction. We postulate that the Mg^{2+} ion should coordinate both phosphates that are involved in the cleaved P-O bond. We present a comprehensive analysis of NTP processing superfamilies across all species, determining distinct enzyme active site structures. By examining a vast dataset of crystallographic structures, we identified a universal "Mg-pinch" motif, confirming our structural hypothesis for almost all NTP processing enzymes. We highlight exceptional cases and propose challenging superfamilies that lack sufficient structural data to determine precise active site coordination.

Our quantum mechanical calculations revealed the electrostatic polarization effects of the Mg^{2+} ion as key determinants of their role in the enzyme catalysis. The Mg-pinch motif provides a mechanistic framework for understanding the catalytic role of metal ions in NTP processing. Our findings offer insights into enzyme evolution, provide a basis for rational enzyme engineering, and could inform the development of novel therapeutics targeting NTP processing enzymes.

1 Introduction

Life, as we know it, is dependent on phosphate chemistry, essential in all major biological processes of living organisms, including the dynamic regulation and maintenance of genetic material; signaling and regulation; energy storage and transfer, and in activating and regulating biosynthetic pathways [1, 2].

The most basic building blocks involving phosphates are nucleotides. Interestingly, while molecules have the ability to form longer linear and cyclic phosphate anhydrides [3], nucleoside-5'-triphosphates (NTPs) with three phosphates are the longest phosphate chains that dominate biological processes, central to enzymatic phosphate chemistry. Oligophosphates beyond NTPs only rarely occur [4, 5].

As essential co-factors for phosphate catalysis, divalent metal ions are required in the large majority of phosphate catalytic enzymes, not only in proteins, but also in RNA-based ribozymes. Mg^{2+} ions are almost always the natural co-factors, however they can be substituted by Mn^{2+} in most cases. Ca^{2+} is also crucial in cell signaling, its concentration is heavily regulated, and, interestingly, Ca^{2+} often inhibits phosphate catalysis, consistent with its key role in apoptosis [6, 7]. The metal ion-aided phosphate catalytic mechanism is, therefore, considered a foundational element in the molecular basis of life [8, 9].

However, the structure-function relationship concerning the metal ion coordination is not well understood for phosphate catalysis. Bioinformatics tools can be used to predict NTP binding sites from as little as sequence information [10, 11]. These are, however, not always reliable in identifying the active sites, and they only provide labels for interacting residues, accurate structural information is not available,

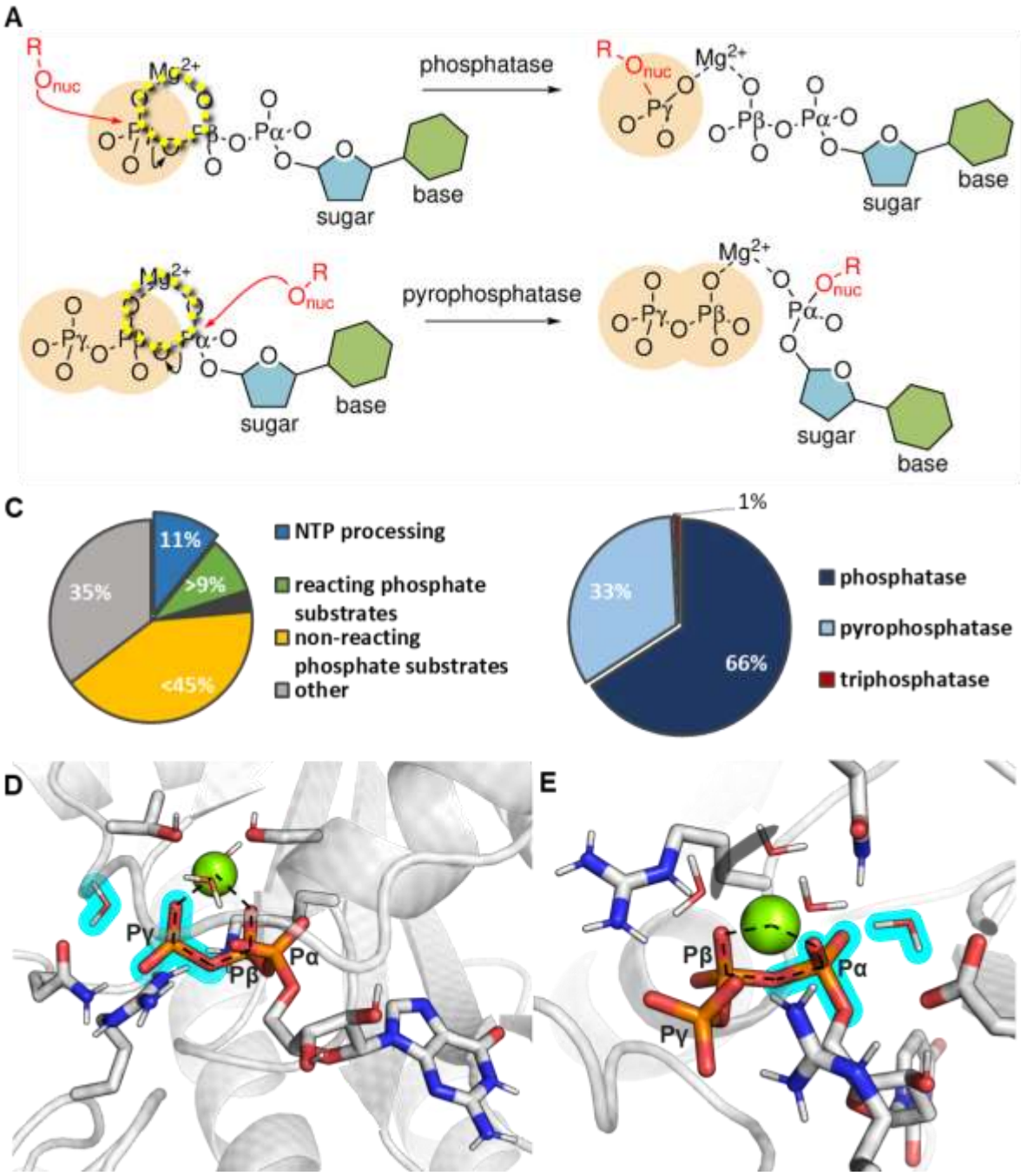

Figure 1. **A-B:** Schematics of γ-phosphate (**A**, purple) and βγ-diphosphate (**B**, cyan) cleavages by an oxygen nucleophile (O_{nuc}). "Mg-pinch" is highlighted by yellow hexagons. **C**: Phosphate chemistry is key in enzymatic reactions, with the NTP-processing enzymes dominating biological processes. Statistical analysis was performed using Enzyme Commission Numbers. **D, E:** Active site for phosphate hydrolysis in the KRas.p120GAP enzyme complex (**D**) and in dUTPase (**E**). Nucleophilic waters and the attached phosphates are highlighted in cyan.

particularly regarding the metal ions. The coordination geometry, number of ions required varies among different types of enzymes, and there is a lack of systematic analysis. The precise active site structure of

the catalytically active enzymes often remains elusive [12-14]. Furthermore, quantitative sub-atomic resolution insights are experimentally lacking, which could provide comprehensive understanding of the catalytic role of these essential cofactors in modulating the kinetics of phosphate reactions in biology.

Here we aim to systematically reveal the functional role of the divalent metal ions in NTP processing enzymes. We hypothesize that the catalytically active coordination of Mg^{2+} ions is determined by the chemistry performed by these enzymes: Mg^{2+} coordinates both phosphate groups involved in the cleaved P-O bond. We focused on the two most common reactions of NTP catalytic enzymes: hydrolysis or transfer (i) with phosphate and a nucleoside diphosphate product, or (ii) with pyrophosphate and nucleoside monophosphate products (Figure 1A-B). To challenge this hypothesis, we identified NTP-processing superfamilies (SFs) based on all available high-resolution crystallographic structures of NTP hydrolysis enzyme active sites deposited in the Protein Data Bank (PDB) [15] to date. We found that almost all exhibit a universal metal ion coordination, which we termed the "Mg-pinch motif", where an active site divalent metal ion coordinates both the attacked and the cleaved phosphate groups of the NTP simultaneously.

To quantitatively evaluate geometrical, electrostatic, and polarization factors, we additionally performed quantum mechanical (QM) and quantum mechanical/molecular mechanical (QM/MM) calculations on three representative enzymes: a phosphatase, Ras (Figure 1D), a pyrophosphatase, dUTPase (Figure 1E), and a general phosphate cleaving enzyme, RNase H (Figure S1), revealing the key polarization effects essential to biocatalysis.

2 Results and Discussion

2.1 EC analysis

To systematically identify examples of enzyme-catalyzed NTP processing reactions, we obtained reactivity data from the *Kyoto Encyclopedia of Genes and Genomes* (KEGG) [16]. We analyzed the substrates and products of the reactions categorized by the Enzyme Commission (EC) number assigned to enzymes. We evaluated (i) whether the reacting chemical species contain phosphates, and (ii) whether the reaction involves the phosphate groups directly, and finally (iii) whether the reaction uses NTP as a substrate.

Phosphates are present in at least one of the substrates in as much as 65% of the EC numbers (4348 out of 6728), and more than 20% of all enzyme reaction types (1469) carry out phosphorus chemistry directly, as measured by the prevalence of the EC classes enzymatic reactions using our KEGG-analysis (Figure 1C).

Demonstrating the importance of the triphosphate chain, about half of the phosphate processing enzymes uses NTP (732). Of all NTP reacting EC reactions, two-thirds of the enzymes (484) perform phosphatase activity, removing the γ-phosphate (Figure 1A), and one third are pyrophosphatases (244), cleaving between the α and β phosphates (Figure 1B). Only six EC categories correspond to reactions cleaving all three phosphates of the nucleotide (2.5.1.6, 2.5.1.17, 5.1.154, 3.1.5.1, 4.1.2.50 and 4.2.3.12). The members of EC 3.6.1.8 are capable of both phosphatase and pyrophosphatase activity, while EC 3.6.1.29 hydrolyses P1,P3-bisnuleoside-triphosphates, where the conventional numbering of the phosphates is ambiguous. Two enzyme categories (6.3.1.34 and 6.3.2.18) consume ATP, but their product is not yet established by IUBMB [17]. Occasionally, more than three phosphates are attached to a nucleoside (e.g., in 3.6.1.17). Here we primarily focus on the two prominent NTP reactions, the phosphatase and pyrophosphatase activities, which are distinct and well defined in all cases except for the above mentioned two exceptions.

2.2 Data collection

To identify superfamilies (SFs) capable of NTP processing, we collected i) PDB structures [15] containing NTPs and bound divalent metal ions, as well as ii) PDB structures and protein sequences that are associated with NTP-processing ECs in protein databases (KEGG [18-20], InterPro [21], and Expasy [22]) (Figure S2). The identified protein chain sequences were then fetched to SUPERFAMILY 2.0 [23, 24] to assign the corresponding SCOPe SFs [25, 26] by matching them against a collection of hidden Markov models that represent structural protein domains at the SCOPe SF level (a superfamily is a group of protein domains arisen from a common ancestor). For those cases where SUPERFAMILY did not identify any SCOPe SFs, we used the InterPro annotation [21]. The identified SFs were then validated against the literature to ensure that the corresponding protein chains are indeed capable of NTP-processing. Finally, redundant SFs (if an InterPro entry closely resembled an existing SCOPe SF based on structural alignments with the distance matrix alignment (DALI) software [27]) were merged. For each SF, we were able to select a representative structure by establishing consensus metal ion coordination within the SF using the good structural alignment of all members.

We hypothesized that members of a given SF share a common active site arrangement with a corresponding consensus metal ion coordination, and that NTP-processing active sites exhibit a universal metal ion coordination strategy, which we termed the “Mg-pinch” motif. We hypothesized that the Mg-pinch, whereby both phosphates that are involved in the cleaved P-O bond are simultaneously coordinated, is universal in NTP-processing enzymes and is energetically advantageous for the catalytic reaction to polarize the reacting bonds. To validate our hypothesis, we analyzed the metal ions and their specific coordination geometries within the active sites for each SF. We introduced a reference plane, defined by the three phosphorus atoms of the NTP, and labelled the two half-spaces (+) and (-) in a way, that looking from position of the α phosphorus, the righthand side of the phosphates is (+), the left is (-) (see inset in Figure 2 top left corner). The main ion positions coordinated by two or three phosphates are color-coded (Figure 2) and labelled by which phosphates coordinate them with capital Latin letters (ABG for αβγ, respectively). We found that an Mg-pinch is formed in the vast majority of NTP-processing enzymes (96%, in 65 of 68 SFs with identified catalytic Mg^{2+}), where the Mg^{2+} coordinates both phosphate groups between which the bond is cleaved during the catalytic reaction. This configuration results in the characteristic hexagonal arrangement depicted in Figure 2.

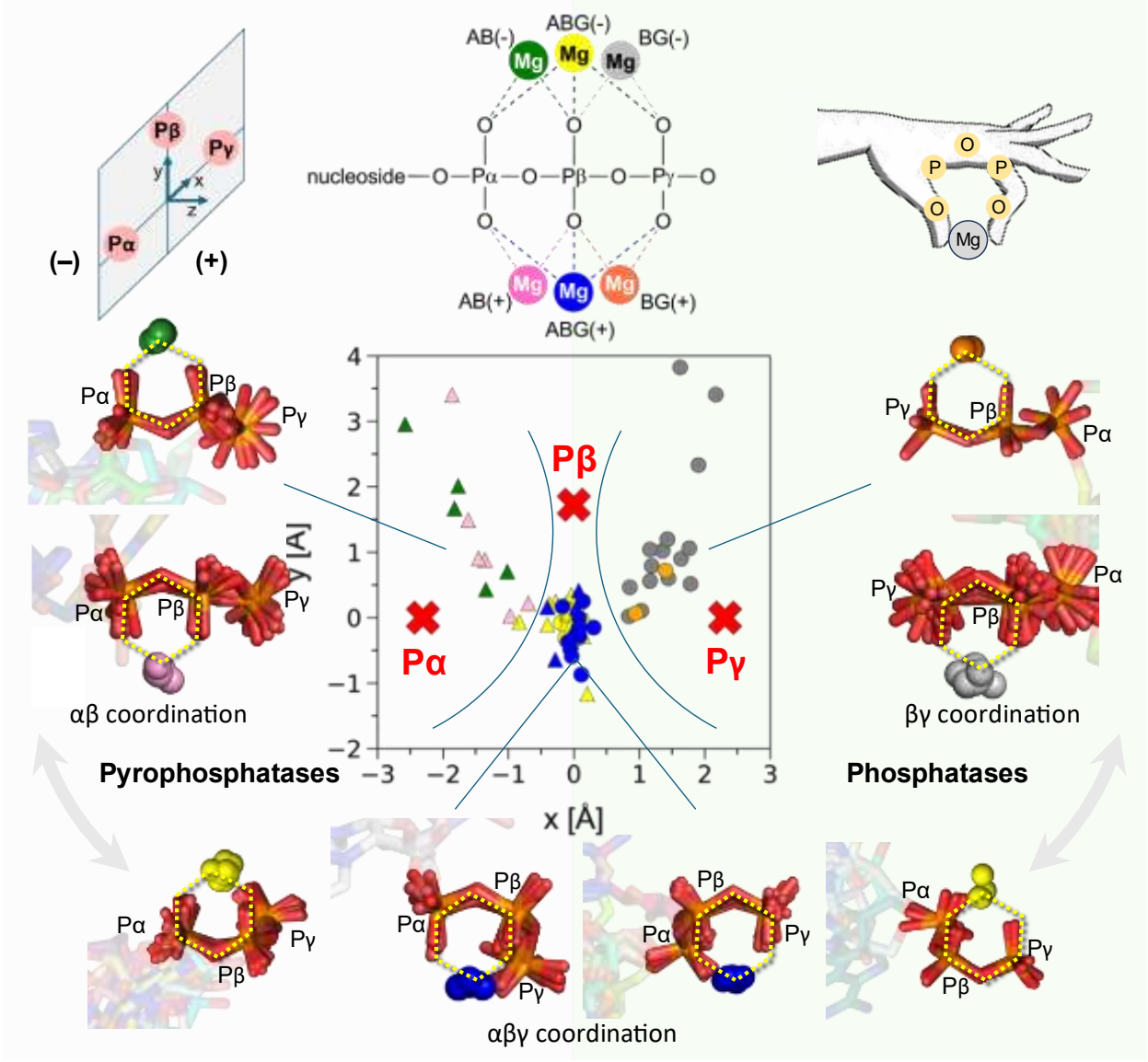

Figure 2. Mg-pinch configurations in the phosphatase (right, green shading) and pyrophosphatase (left, gray shading) SFs shown in the corresponding representative structures superimposed using their three phosphorus atoms (sticks). In addition to the two phosphorus atoms and the Mg^{2+}, the Mg-pinch includes the bridging oxygen between the two phosphorus atoms and one non-bridging oxygen from both coordinated phosphate groups. The different ion clusters are color coded; αβγ-coordinated metal ions are shown in blue on the (+) side (ABG(+)) and yellow on the (–) side (ABG(–)), βγ-coordinated metal ions in orange (BG(+)) and gray (BG(–)), and the αβ-coordinated metal ions in pink (AB(+)) and dark green (AB(–)). In the middle panel: projection of different Mg^{2+} positions for phosphatases and pyrophosphatases. The positions of the Mg-pinch-forming metal ions are shown with respect to the phosphates (middle). Phosphatase ion positions are depicted by circles, and pyrophosphatase ion positions are represented by triangles (color code as introduced earlier). The αβγ-coordination is common for both phosphatases and

pyrophosphatases, whereas the βγ coordination is prevalent in phosphatases and the αβ coordination is characteristic of pyrophosphatases. The Mg-pinch is organized in a hexagonal configuration (indicated by yellow dashed hexagons). Alternative projections are depicted in Figure S3.

2.3 Phosphatases: Pγ leaving group

We identified a total of 42 phosphatase enzyme SFs, that form four groups in terms of the Mg-pinch involving the β and the γ phosphate groups: ABG(+), ABG(–), BG(+), and BG(–) (Figure 2 right, Figure S4). The majority of phosphatases are either ABG(+) (11 SFs) or BG(–) (15 SFs). An additional cluster of auxiliary ions are identified in the AG position (Figures S5-6).

Table 1: *Phosphatase superfamilies grouped by their Mg-pinch-forming metal ions. The colors of the four groups match those of the corresponding ion positions shown in* Figure 2*.*

Mg Coordination and binding site	Phosphatase SF	SF ID	SCOPe Fold (or InterPro ID)	Additional ion(s)
αβγ ABG(+)	GroEL equatorial domain-like	a.129.1	GroEL equatorial domain-like	
	Riboflavin kinase-like	b.43.5	Reductase/isomerase/elongation factor common domain	
	Carbamate kinase-like	c.73.1	Carbamate kinase-like	
	Nucleoside diphosphate kinase, NDK	d.58.6	Ferredoxin-like	
	GHMP Kinase, C-terminal domain	d.58.26	Ferredoxin-like	
	CYTH-like phosphatases	d.63.1	CYTH-like phosphatases	G(+)
	PurM N-terminal domain-like	d.79.4	Bacillus chorismate mutase-like	BG(+)
	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase	d.122.1	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase	
	ParB/Sulfiredoxin	d.268.1	ParB/Sulfiredoxin	
	CofE-like	d.340.1	CofE-like	G(–)
	NAD kinase/diacylglycerol kinase-like	e.52.1	NAD kinase/diacylglycerol kinase-like	
αβγ ABG(-)	Phosphoenolpyruvate/pyruvate domain	c.1.12	TIM beta/alpha-barrel	
	Phosphoglycerate kinase	c.86.1	Phosphoglycerate kinase	
	Phosphofructokinase	c.89.1	Phosphofructokinase	
	GlnB-like	d.58.5	Ferredoxin-like	
βγ BG(+)	Actin-like ATPase domain	c.55.1	Ribonuclease H-like motif	
	Glutamine synthetase/guanido kinase	d.128.1	Glutamine synthetase/guanido kinase	AG
βγ	DhaL-like	a.208.1	DhaL-like	AB(–)

BG(-)	Calcium-dependent phos-photriesterase	b.68.6	6-bladed beta-propeller	BG(+)
	Tubulin nucleotide-binding domain-like	c.32.1	Tubulin nucleotide-binding domain-like	
	P-loop containing nucleoside triphosphate hydrolases	c.37.1	P-loop containing nucleoside triphosphate hydrolases	
	Ribokinase-like	c.72.1	Ribokinase-like	AB(+)
	MurD-like peptide ligases, catalytic domain	c.72.2	Ribokinase-like	
	PEP carboxykinase-like	c.91.1	PEP carboxykinase-like	
	Glutathione synthetase ATP-binding domain-like	d.142.1	ATP-grasp	AG
	SAICAR synthase-like	d.143.1	SAICAR synthase-like	AG
	Protein kinase-like (PK-like)	d.144.1	Protein kinase-like (PK-like)	AG
	Metal cation-transporting ATPase, ATP-binding domain N	d.220.1	Metal cation-transporting ATPase, ATP-binding domain N	
	Diacylglycerol kinase (DgkA)-like	f.62.1	Diacylglycerol kinase (DgkA)-like	AG
	Inositol-pentaphosphate 2-kinase	N/A	(IPR009286)	
	YcaO-like domain	N/A	(IPR003776)	
	Peptidase G2, IMC autopro-teolytic cleavage domain	N/A	(IPR021865)	

The *P-loop containing nucleoside triphosphate hydrolase* SF, by far the largest in our dataset with nearly 1000 corresponding high-resolution PDB structures, showcases a highly conserved βγ phosphate coordination, besides a conserved serine and threonine residue in the first coordination sphere of the Mg^{2+} (Figure 3). Both the *PEP carboxykinase-like* and the *MurD-like peptide ligase, catalytic domain* SFs exhibit nearly identical metal ion coordination configurations to the *P-loop containing NTP hydrolase* SF, as well as the *MurD-like peptide ligases* SF where the serine is replaced by a threonine and an additional glutamate is involved in the Mg^{2+} coordination. A very similar coordination is also present in the *Tubulin nucleotide-binding domain-like* SF, with the corresponding ~250 structures exhibiting consistent agreement in their active sites. Four SFs: *Protein kinase-like*, the *Glutathione synthetase ATP-binding domain-like*, the *SAICAR synthase-like*, and the *Diacylglycerol kinase (DgkA)-like* SFs have similar BG(–) coordination accompanied by an additional αγ-coordinated metal ion. This coordination is also referred to as the ATP-grasp motif [28]. In the case of the *Inositol-pentaphosphate 2-kinase* group, no SCOPe SF could be assigned, and we termed it after EC 2.7.1.158 (InterPro SF, Inositol-pentakisphosphate 2-kinase, N-terminal lobe, IPR043001). This family was previously described to share some level of similarity to the *Inositol polyphosphate kinase* family belonging to the *SAICAR synthase-like* SF [29]. A clear phosphate coordination with a Mg^{2+} at the BG(–) site can be identified for the corresponding structures. Even though the structural (DALI) alignment of the representative structures of all superfamilies also revealed a good structural overlap with the *SAICAR synthase-like* SF where the nucleotides are positioned similarly, we introduced them as distinct categories due to some differences observed at the active site (Figure S7). The

Diacylglycerol kinase (DgkA)-like SF has a similar ATP configuration and Mg^{2+} coordination to those discussed above, but the overall secondary structure is rather different. The *Actin-like ATPase domain* is the most abundant phosphatase superfamily having a metal ion on the (+) side – with ~200 corresponding high-resolution PDB structures uncovered – and features a one metal-ion at its active site.

Some other SFs had less structural data available or lacked a SCOPe annotation. For the *DhaL-like* superfamily, only one structure with ATP and two metal ions were identified (PDB 1UN9 [30]), which were not part of the high-resolution dataset. This SF was identified as the corresponding enzymes for three EC categories (SI Note 1). Other ADP-containing structures suggest that it most probably exhibits an αβ and a βγ coordinated Mg^{2+}. The *YcaO-like domain* SF – for which we identified one Mg^{2+} at the BG(–) site – is defined on the InterPro domain level (IPR003776) and has been shown to use ATP to activate amide backbones during peptide cyclodehydrations [31]. However, there is no associated EC to this domain and its corresponding structures. Similarly, no associated EC exists for the *Peptidase G2, IMC autoproteolytic cleavage domain* SF (defined based on the InterPro domain level, IPR021865) and the corresponding structures of the pre-mature bacteriophage phi29 gene product 12 originate from a single publication [32]. The authors suggest that autocleavage of the C-terminal domain is a post-trimerization event that is followed by a unique ATP-dependent cleavage and release [32]. Based on the structures we propose a BG(–) Mg^{2+} coordination.

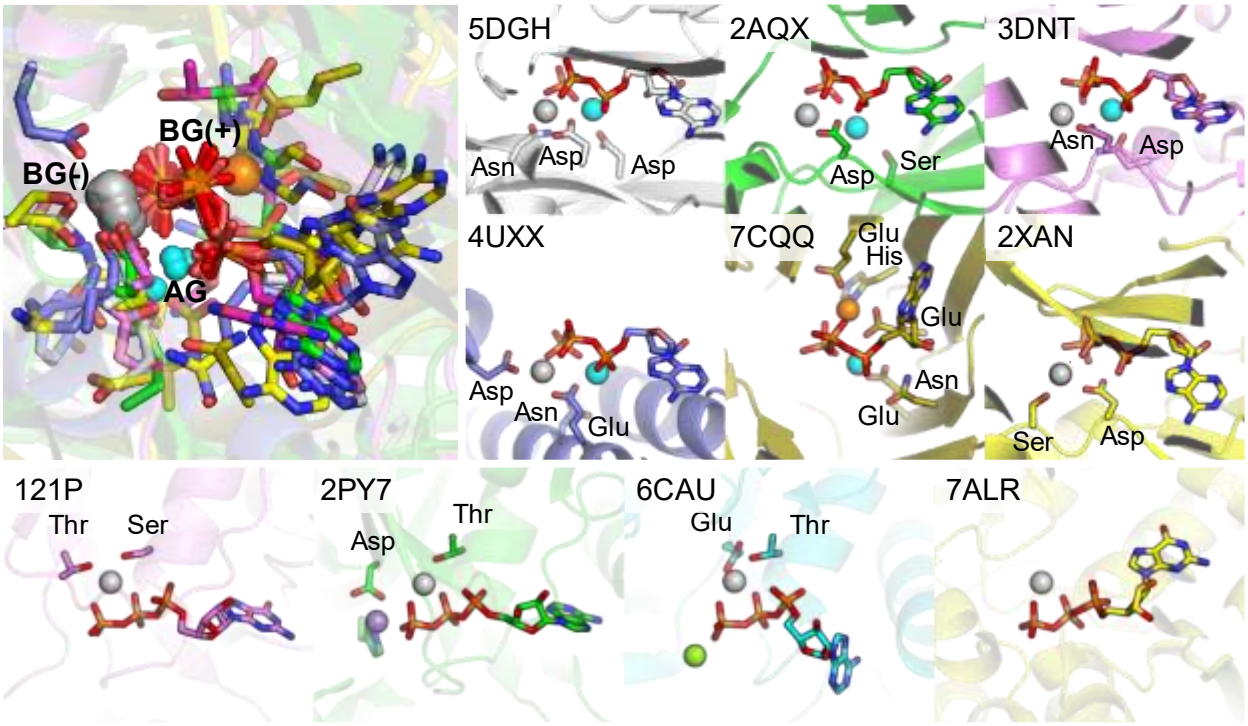

Figure 3. SFs with a BG(-) pinching ion. Several among them exhibit an additional AG second ion (shown in cyan): Glutathione synthetase ATP-binding domain-like (white, 5DGH); SAICAR synthase-like (green, 2AQX); Protein kinase-like (pink, 3DNT); Diacylglycerol kinase (DgkA)-like (blue, 4UXX) and one exception that has its pinching ion at the BG(+) position: Glutamine synthetase/guanido kinase (olive, 7CQQ). Structurally similar Inositol-pentaphosphate 2-kinase (yellow, 2XAN) does not have an additional ion in the AG site. In the bottom four panels:

similarities in Mg^{2+} coordination among: P-loop containing NTP hydrolase (pink, 121P); PEP carboxykinase-like (green, 2PY7, second ion is a Mn^{2+}, shown in purple); MurD-like peptide ligase, catalytic domain (cyan, 6CAU); Tubulin nucleotide-binding domain-like (yellow, 7ALR).

2.4 Pyrophosphatases: PβPγ leaving group

Among the 90 SFs in our dataset, 44 enzymes have pyrophosphatase activity, 33 of which have available structures enabling us to clearly determine the corresponding metal ion coordination. The Mg-pinch in pyrophosphatases has an αβ or an αβγ metal ion coordination, forming four groups: ABG(+), ABG(–), AB(+), and AB(–) (Figure 2, left, Figure S8).

20 SFs possess a Mg-pinch-forming metal ion on the (–) side, the majority (15 SFs) showcasing αβγ-coordination. Amongst these, the *DNA/RNA polymerase* and the *Nucleotidyltransferase* SFs display a remarkably similar nucleotide configuration, orientation, and ion coordination. These are the two most populated pyrophosphatase SFs in our high-resolution PDB dataset, with ~475 and ~340 structures, respectively. Despite their structurally distinct folds, the two SFs exhibit nearly identical positioning and orientation of the nucleotide, catalytic residues, and coordinating metal ions. Both demonstrate a clear two-metal-ion catalytic coordination [14], with a third metal ion present in some structures, located not far from the γ phosphate, potentially playing a role in ligand release [33]. Alongside the αβγ-coordinated Mg^{2+} that forms the Mg-pinch, the second Mg^{2+} is coordinated by the α phosphate and the attacking hydroxyl group of the ligand nucleotide (Figure 4).

Table 2: *Pyrophosphatase superfamilies grouped by their Mg-pinch-forming metal ions. The colors of the four groups match those of the corresponding ion positions shown in Figure 2.*

Mg Coordination and binding site	**Pyrophosphatase SF**	**SF ID**	**SCOPe Fold** (or InterPro ID)	**Additional ion(s)**
αβγ ABG(+)	all-alpha NTP pyrophosphatases	a.204.1	all-alpha NTP pyrophosphatases	B(+)
	Adenine nucleotide alpha hydrolases-like	c.26.2	Adenine nucleotide alpha hydrolase-like	AG(–)
	PRTase-like	c.61.1	PRTase-like	
	Nucleotide-diphospho-sugar transferases	c.68.1	Nucleotide-diphospho-sugar transferases	A(–)
	Activating enzymes of the ubiquitin-like proteins	c.111.1	Activating enzymes of the ubiquitin-like proteins	B
	Phosphopantoate/pantothenate synthetase superfamily	N/A	(IPR038138)	
αβγ ABG(–)	dUTPase-like	b.85.4	beta-clip	
	Nucleotidylyl transferase	c.26.1	Adenine nucleotide alpha hydrolase-like	
	ITPase-like	c.51.4	Anticodon-binding domain-like	
	CoaB-like	c.72.3	Ribokinase-like	
	Nucleotide cyclase	d.58.29	Ferredoxin-like	A(–)

	EPT/RTPC-like	d.68.2	IF3-like	
	YrdC/RibB	d.115.1	YrdC/RibB	
	Nucleotidyltransferase	d.218.1	Nucleotidyltransferase	A(–)
	Prim-pol domain	d.264.1	Prim-pol domain	A(–)
	YojJ-like	d.320.1	YojJ-like	
	DNA/RNA polymerases	e.8.1	DNA/RNA polymerases	A(–)
	Poly(A) polymerase catalytic subunit-like	e.69.1	Poly(A) polymerase catalytic subunit-like	
	Bacterial DNA polymerase III, alpha subunit, NTPase domain	N/A	(IPR011708)	
	Influenza RNA-dependent RNA polymerase subunit PB1	N/A	(IPR001407)	A(–)
	RNA-dependent RNA polymerase, eukaryotic-type	N/A	(IPR007855)	A(–)
	Virion DNA-directed RNA polymerase domain	N/A	(IPR049432)	A(–)
αβ AB(+)	6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, HPPK	d.58.30	Ferredoxin-like	BG(+)
	Class II aaRS and biotin synthetases	d.104.1	Class II aaRS and biotin synthetases	BG(+), BG(–)
	Nudix	d.113.1	Nudix	BG(+)
	tRNA-splicing ligase RtcB-like superfamily (Hypothetical protein PH1602)	d.261.1	Hypothetical protein PH1602	BG(+)
	beta and beta-prime subunits of DNA dependent RNA-polymerase	e.29.1	beta and beta-prime subunits of DNA dependent RNA-polymerase	BG(+)
	Aerobactin siderophore biosynthesis, IucA/IucC-like	N/A	(IPR037455)	AG
αβ AB(–)	Fic-like	a.265.1	Fic-like	
	Molybdenum cofactor biosynthesis proteins	c.57.1	Molybdenum cofactor biosynthesis proteins	
	DNA primase core	e.13.1	DNA primase core	BG(–)
	Adenylylcyclase toxin (the edema factor)	e.41.1	Adenylylcyclase toxin (the edema factor)	
	Mitochondrial carrier	f.42.1	Mitochondrial carrier	

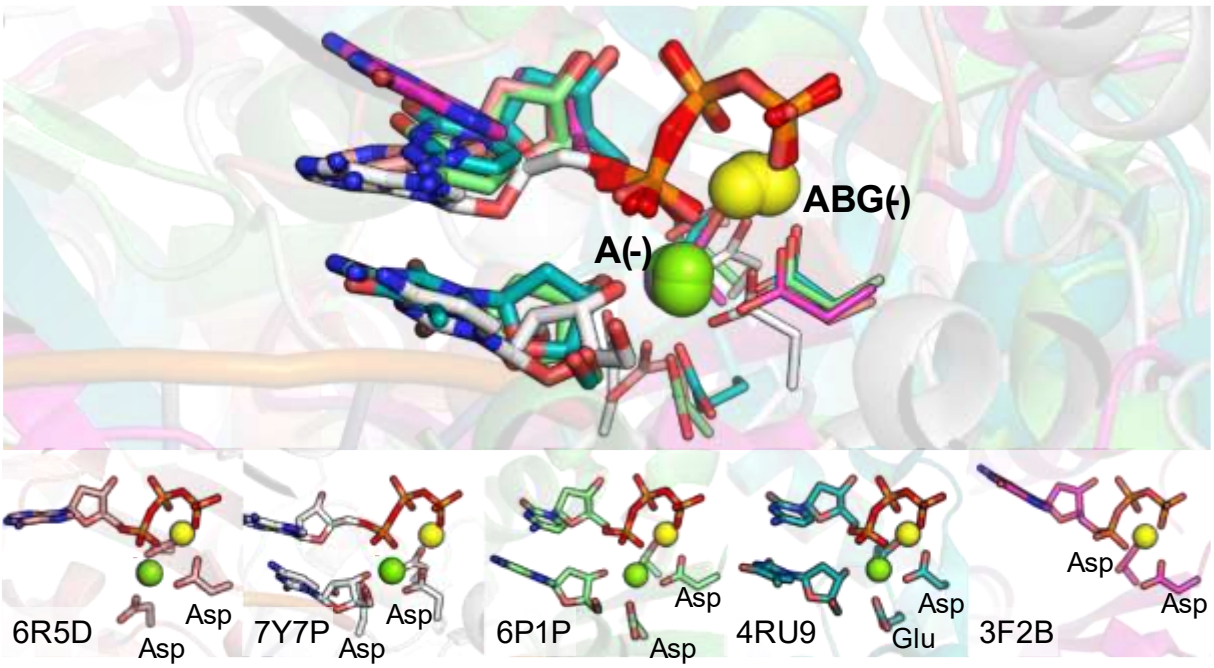

Figure 4. Pyrophosphatases with additional ion between the α phosphate and the nucleophile alongside the ABG(-) pinching ion, typical to polymerases: Prim-pol domain (salmon, 6R5D); RNA-dependent RNA polymerase, eukaryotic-type (white, 7Y7P); Nucleotidyltransferase (green, 6P1P); DNA/RNA polymerases (teal, 4RU9); Bacterial DNA polymerase III, alpha subunit, NTPase domain (magenta; 3F2B).

SCOPe classification did not always give a result, or sometimes it was incorrect. We identified a group of PDB structures, exemplified by PDB 2IRX [34] that were classified as *PLP-dependent transferases* (yet with a weak score) or were not classified at all by SUPFAM 2.0. By using the InterPro database [21] we clearly identified this domain to be the *DNA ligase D, polymerase domain* (the *PLP-dependent transferase* was a misclassification). After comparing the corresponding PDB structures to our dataset, we found that these belong to the *Prim-pol* superfamily, yet this domain remains unannotated in the SCOPe database [25, 26].

No SCOPe annotation is available for the structures of the *Influenza RNA-dependent RNA polymerase subunit PB1* (defined at the InterPro family level, IPR001407), the *Virion DNA-directed RNA polymerase domain* (defined on the InterPro domain level, IPR049432), and the *RNA-dependent RNA polymerase, eukaryotic-type* (defined at the InterPro family level, IPR007855) SFs. All three SFs exhibit a typical two-metal ion catalytic active site configuration highly alike polymerases, yet they result in a poor overall DALI alignment suggesting that they represent distinct SFs. A second active site Mg^{2+} at the A(–) site may be missing from the few structures we identified for the *Bacterial DNA polymerase III, alpha subunit, NTPase domain* SF [35]. This SF does not have an available SCOPe annotation either (defined at the InterPro domain level, IPR011708), and one Mg^{2+} is found in their active site at the ABG(–) site.

Some of the (+) side SFs also do not have SCOPe annotation, we identified the *Phosphopantoate/pantothenate synthetase* and *Aerobactin siderophore biosynthesis, IucA/IucC-like* SFs that have a unique active site organization and are clearly distinct from other SFs based on DALI alignments.

2.5 Exceptional coordination and superfamilies challenging to classify

The *Acetyl-CoA synthetase-like* SF, even though it uses Mg^{2+} as cofactor, its BG(+) coordination is exceptional considering its pyrophosphatase activity (Figure 5A-C). A potential reason for the deviation from our hypothesis is due to the negatively charged attacking group, whereby the nucleophilic attack can occur more readily and does not necessitate the metal coordination between P_α and P_β to hydrolyze ATP to AMP and pyrophosphate. In this case, the resulting products are high energy intermediates that are used for subsequent enzyme-catalyzed reactions, which might also be the reason for the exceptional coordination.

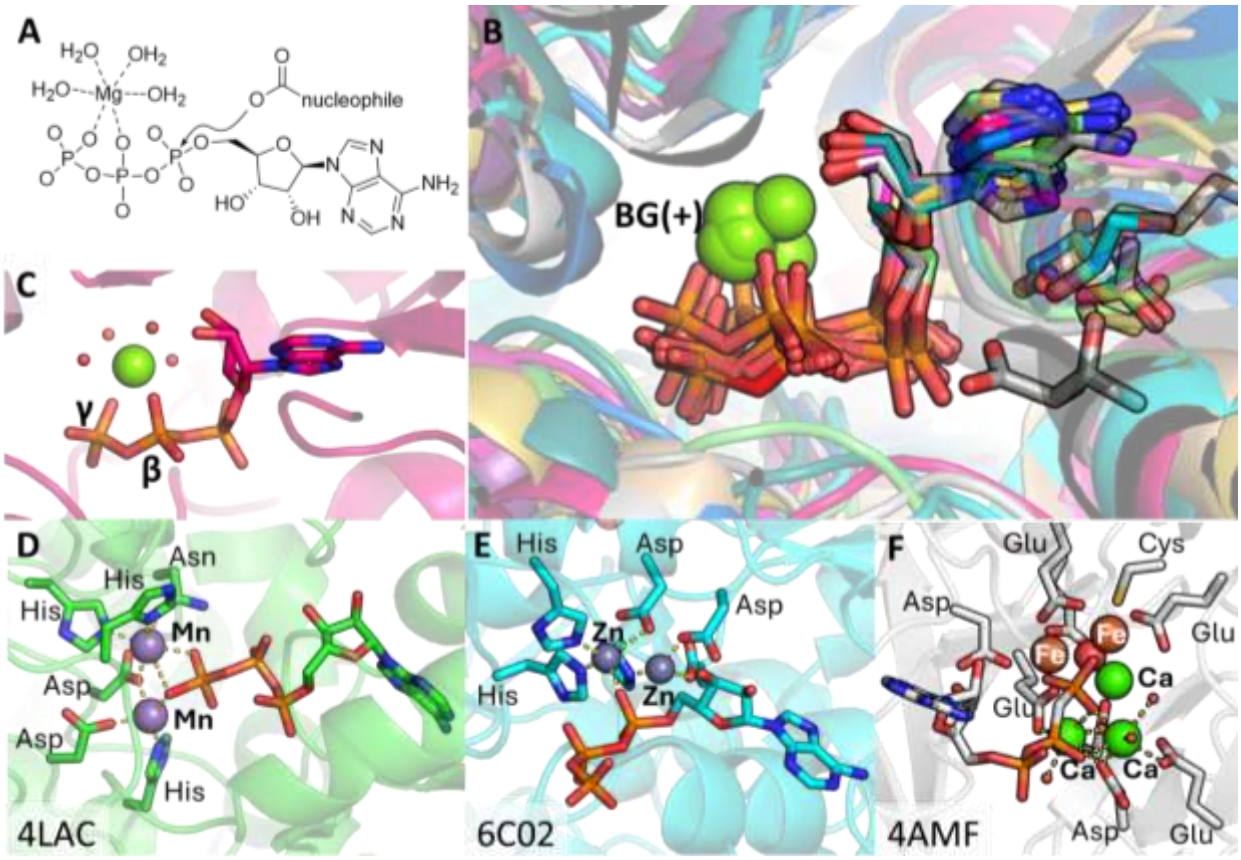

Figure 5. **A** Reaction scheme of carboxylate activation in peptide synthesis catalyzed by the Acetyl-CoA synthetase-like SF. **B** Multiple structural alignment agrees on metal coordination and secondary structure. **C** Metal ion coordination in the Acetyl-CoA synthetase-like SF (PDB 5BSM). Mg^{2+} ions are not colored based on their coordination. Additionally, exceptional metal ion coordination in the **D** Metallo-dependent phosphatases SF (green, 4LAC) **E** Alkaline phosphatase-like SF (cyan, 6C02) and **F** Ca-dependent phosphotriesterase (white, 4AMF).

A decisive conclusion regarding the metal coordination and corresponding representative structures could not be drawn for further 9 phosphatase and 9 pyrophosphatase SFs due to the lack of available structures with NTP and metal ions in their active site (Table 3). Some of SFs had only structures in the product state – e.g. the *Thiamin pyrophosphokinase, catalytic domain* SF. Using these structures with thiamine diphosphate and AMP, we hypothesize an αβγ coordination. NTP analogues can also distort the active site – e.g. the *RibA-like* SF, where we hypothesize an αβ coordination. In some cases high resolution structures do not represent catalytically active complexes – e.g. the DNA or RNA is missing in the structures of the

DNA ligase/mRNA capping enzyme, catalytic domain SF for which a two metal-ion mechanism was proposed [36-38], yet it cannot be validated based on the available structures. There are further SFs, for which no substrate-bound active site structures are available to date.

While the majority of SFs are ubiquitous in all kingdoms of life, lack of structural information sometimes also corresponds to the restricted prevalence of the enzymes in various species, when primarily prokaryotic or microorganisms use the corresponding enzymes. For example, the *tRNA(Ile2) 2-agmatinylcytidine synthetase TiaS* SF was found to be present in extremophile archaea species only, or the *Molybdopterin cofactor biosynthesis C, (MoaC)* SF is identified in bacteria, archaea and plants only.

In three linked kinase categories: 2.7.1.174, 2.7.1.182 and 2.7.1.216, although SFs could not be identified due to lack of structures, we were able to infer a SF similarity. In this case, we used the available sequences to generate AlphaFold structure predictions, which suggest excellent structural agreement with the *Mitochondrial carrier* pyrophosphatase SF. The only SF for which literature verifiable NTP processing takes place, but no structural information exists at all, is the *GTP cyclohydrolase MptA* SF [39]. In this case, sequences are available and using those, the AlphaFold predicted structure did not align to any other SFs in our database, therefore this is likely a novel, yet structurally unknown NTP processing SF. We also encountered some historically inaccurate EC assignments. For example, recent Cryo-EM structures of FAST kinases were not found to possess any kinase activity [40], despite the original proposition [41] and EC annotation. A further 82 ECs could not be categorized structurally due to the absence of any protein sequence information (see SI Note 2).

Table 3: *Superfamilies the coordination of which could not be determined with certainty due to the lack of structures with high-resolution active sites, in the presence of NTP and metal ions.*

Superfamily	**Reaction**	**SF ID**	**SCOPe Fold** (or InterPro ID)
Transglutaminase, two C-terminal domains	Pi	b.1.5	Immunoglobulin-like beta-sandwich
Apyrase	Pi	b.67.3	5-bladed beta-propeller
FomD-like	Pi	b.175.1	FomD barrel-like
Nicotinate/Quinolinate PRTase C-terminal domain-like	Pi	c.1.17	TIM beta/alpha-barrel
NagB/RpiA/CoA transferase-like	Pi	c.124.1	NagB/RpiA/CoA transferase-like
Glycerate kinase I	Pi	c.141.1	Glycerate kinase I
YgbK-like	Pi	c.146.1	YgbK-like
HIT-like	Pi	d.13.1	HIT-like
Phospholipase D/nuclease	Pi	d.136.1	Phospholipase D/nuclease
S-adenosyl-L-methionine-dependent methyltransferases	PPi	c.66.1	S-adenosyl-L-methionine-dependent methyltransferases
Thiamin pyrophosphokinase, catalytic domain	PPi	c.100.1	Thiamin pyrophosphokinase, catalytic domain
RibA-like	PPi	c.144.1	RibA-like
Molybdenum cofactor biosynthesis protein C, MoaC	PPi	d.58.21	Ferredoxin-like
RPB5-like RNA polymerase subunit	PPi	d.78.1	RPB5-like RNA polymerase subunit
DNA ligase/mRNA capping enzyme, catalytic domain	PPi	d.142.2	ATP-grasp

tRNA(Ile2) 2-agmatinylcytidine synthetase TiaS	PPi	N/A	(IPR024913)
Phosphatidate cytidylyltransferase	PPi	N/A	(IPR000374)
GTP cyclohydrolase MptA	PPi	N/A	(IPR022840)

Finally, while a well-defined consensus structure can be identified for all distinct SFs that carry out a well-defined reaction, this might not always be the case in some very rare examples. MutT1 reportedly carries out phosphatase activity [42], although being a member of the *Nudix* SF, confirmed both by sequences and alignment using AlphaFold predicted structures, as experimental structures are not available. *Nudix* members, however, generally hydrolyze pyrophosphates. Two distinct groups are present in the *Adenine nucleotide alpha hydrolases-like* SF: ATP hydrolysis of the stress proteins within this SF that bind a single Mg^{2+} coordinated by the αβγ phosphates is unlikely [43, 44], whereas the remaining enzymes of this SF have a second Mg^{2+} at the AG(–) site and possess pyrophosphatase activity.

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

Only five EC categories correspond to reactions cleaving all three phosphates of the nucleotide (2.5.1.6, 2.5.1.17, 5.1.154, 3.1.5.1, 4.1.2.50) and one (4.2.3.12) that similarly processes 7,8-Dihydroneopterin 3'-triphosphate. We identified four SFs that catalyze these reactions (Table 4), a minority compared with 90 phosphatase and pyrophosphatase SFs. Their coordination is very variable (Figure S9). The *S-adenosylmethionine synthetase* SF has a Mg^{2+} in the ABG(–) position and another in the AG(+), the *Cobalamin adenosyltransferase-like* SF one at ABG(+) and another at A(–), the *HD-domain/PDEase-like* SF one at BG(+) and two additional ions coordinated by the α phosphates. The coordination could not be determined for the 7,8-Dihydroneopterin 3'-triphosphate processing *Tetrahydrobiopterin biosynthesis enzymes-like* SF, due to the lack of structural information. The phosphate chemistry of triphosphatases differs from those of the phosphatases and pyrophosphatases, as for triphosphatases the bond is cleaved between C5'-O5', as opposed to the P–O bond cleavage.

Table 4: *Triphosphatase superfamilies and their metal ion coordination.*

Triphosphatase SF	SF ID	Fold	Mg^{2+} coordination	Additional ion(s)
Cobalamin adenosyltransferase-like	a.25.2	Ferritin-like	ABG(+)	A(–)
HD-domain/PDEase-like	a.211.1	HD-domain/PDEase-like	BG(+)	A, A
Tetrahydrobiopterin biosynthesis enzymes-like	d.96.1	T-fold	N/A	N/A
S-adenosylmethionine synthetase	d.130.1	S-adenosylmethionine synthetase	ABG(–)	AG(+)

2.7 Unusual metal ions and metal ion coordinating residues

The vast majority of NTP-processing enzymes rely on Mg^{2+} as the divalent catalytic metal ion cofactor, with the coordination of six 2-electron donors in an octahedral arrangement. Some enzymes may utilize Mn^{2+} as their preferred metal ion cofactors. Manganese was identified to be optimal for the guanylylation reaction of the RtcB enzyme belonging to the *tRNA-splicing ligase RtcB-like* SF [45]. Similarly, the terminal transferase activity of the Mycobacterium tuberculosis DNA ligase belonging to the *PLP-dependent*

transferases SF is preferentially activated by Mn^{2+} versus Mg^{2+} [34]. Likewise, PrimPol of the *Prim-pol domain* SF is Mn^{2+} dependent that shows significantly improved primase and polymerase activities when binding Mn^{2+} rather than Mg^{2+} as cofactors [46].

Intriguingly, although Ca^{2+} inhibits NTP processing enzymes in almost all cases, however, the of the 'soluble' adenylyl cyclase (sAC) enzyme (*S. platensis*) is an exception in the *Nucleotide cyclase* SF. In the active site, Ca^{2+} plays the pinching metal ion role at the ABG(–) position, while an α-coordinated Mg^{2+} coordinates the attacking hydroxyl group of the ATP, orienting it for the cyclase reaction [47]. The unusual Ca^{2+} activation is suggested to increase the ATP binding and leading to more effective overall kinetics [48]. Most enzymes of this SF, however, use Mg^{2+} in both ion positions.

Very few enzymes do not work with Mg^{2+}. The *Metallo-dependent phosphatases* SF has an atypical coordination with two metal ions at its active site, usually manganese, iron, or zinc (Figure 5**Error! Reference source not found.**D) [49]. It is however also more promiscuous, and can hydrolyze different lengths phosphate chains besides NTPs, whereas typical NTP catalytic enzymes have a clear specificity for triphosphates. Similarly, the nucleotide-degrading members of the *Alkaline phosphatase-like* SF contain two Zn^{2+} ions instead of Mg^{2+} at their catalytic center (Figure 5**Error! Reference source not found.**E) for their pyrophosphatase activity.

The only family belonging to the *Calcium-dependent phosphotriesterase* SF (b.67.3) that processes NTP as substrate, is the *Alkaline phosphatase PhoX* family (InterPro entry IPR008557), containing proteins predominantly found in bacteria. Its complex active-site comprising of two antiferromagnetically coupled ferric iron ions (Fe^{3+}), three calcium ions (Ca^{2+}), and an oxo group bridging three of the metal ions [50]. Interestingly, two of the Ca^{2+} ions have pinching positions in this group, in BG(+) and BG(–) coordination (**Error! Reference source not found.**Figure 5F). The DALI alignment of the SF representatives revealed a surprisingly close resemblance to the *Apyrase* SF (b.68.6) for which the coordination could not be deduced based on the available structures.

2.8 Analysis on the protein fold level

For those SFs that are defined in the SCOPe database, we collected their corresponding fold information. The fundamental unit in the SCOPe database is the domain found in experimentally determined protein structures, which are organized in several levels of hierarchy. While SFs bridge together protein families with common functional and structural features inferred to be from a common evolutionary ancestor, a fold, which corresponds to the level above SFs, groups structurally similar SFs solely relying on structural features; hence sharing a common fold does not imply sequence homology [25].

NTP hydrolysis enzymes cover a wide range of protein folds defined in the SCOPe database, and even the same fold does not imply similar metal ion coordination geometries. In our dataset, 77 SFs originate from the SCOPe database, and cover a total of 67 folds. We identified a maximum of six NTP-processing SFs within the same *Ferredoxin-like* fold. Three SFs belong to the *Ribokinase-like* fold, and two SFs belong to the *TIM beta/alpha-barrel*, the *ATP-grasp*, and the *Adenine nucleotide alpha hydrolase-like* folds each. The remaining 62 SFs all belong to distinct folds.

Moreover, SFs in the same fold may even catalyze distinct reaction classes. In the *Ferredoxin-like* fold, three SFs are phosphatases and three are pyrophosphatases, with very diverse coordination geometries (Figure S11A**Error! Reference source not found.**). Similarly, both the *Glutathione synthetase ATP-binding domain-like* and the *DNA ligase/mRNA capping enzyme, catalytic domain* SFs belong to the *ATP-grasp* fold,

yet the former is a phosphatase, and the latter is a pyrophosphatase. Even if the NTP-processing is similar, as in the *Ribokinase-like* fold, where both the *Ribokinase-like* and the *MurD-like peptide ligases, catalytic domain* SFs are phosphatases and they both have a Mg^{2+} at the BG(–) position, when overlapped, their active sites are located distantly from each other (Figure S11B**Error! Reference source not found.**).

On the other hand, very different secondary structural arrangements may create similar active sites to tackle similar reaction mechanisms. This is observed for five polymerases (*Prim-pol domain*; *Nucleotidyltransferase*; *DNA/RNA polymerases*; *Bacterial DNA polymerase III, alpha subunit, NTPase domain*; and *RNA-dependent RNA polymerase, eukaryotic-type*) that belong to different folds (d.264, d.218, e.8) or are not defined in SCOPe, yet they share high similarities in their active sites. They all exhibit a Mg^{2+} binding site at ABG(–) and another at A(–). Both ions are coordinated by two aspartates, and the A(–) is further coordinated by a side chain carboxyl group, as well as by the 3' hydroxyl of the priming nucleotide. The three coordinating sidechains can reside on different structural elements, yet their relative positions and orientations are strikingly similar among the different SFs that catalyze the polymerase reaction (Figure S12).

Identical metal coordination (at the BG(–) site and an additional αγ-coordinated ion is also presented by four SFs that include the ATP grasp motif SFs. All belong to different folds (d.142-144 and f.62), nevertheless three of them share structural similarities. However, the *Diacylglycerol kinase (DgkA)-like* SF is structurally different from the others, yet its NTP coordination strikingly resembles the rest of the group. Interestingly, there is an aspartate that coordinates both Mg^{2+} ions in a conserved position with respect to the NTP for all four SFs, which either resides on a β-sheet (*Glutathione synthetase ATP-binding domain-like*), on an α-helix (*Diacylglycerol kinase (DgkA)-like*) or on a loop (*Protein kinase-like (PK-like)* and *SAICAR synthase-like*). Consequently, the Mg^{2+} coordination geometries seem to be determined by the reaction mechanism rather than the fold itself.

2.9 Analysis of the EC category distributions

We analyzed the distribution of the identified 733 NTP processing EC categories (Figure 1C) in terms of the structural motifs found. ECs are unevenly covered by the identified 90 SFs, the most common eight SFs perform more than half (439/732) of the NTPase reactions. We note that assignment of function in terms of EC categories may be ambiguous and in cases incorrectly linked to SFs (cf Figures S13-14), resulting in SFs seemingly perform both phosphatase and pyrophosphatase activities. Interestingly, the phosphatase *PK-like* SF is verifiably associated with adenylyltransferase activity through flipped pseudokinases [64, 65].

The top SFs are often used to fuel molecular machines *via* inorganic phosphate release (*P-loop containing nucleoside triphosphate hydrolases* or *Actin-like ATPase domain*), including AAA+ ATPases [51, 52]. While ATP is by far the most common NTP substrate (Figure S15), some SFs are specific to other nucleosides, e.g. *dUTPase-like* SF for dUTP. The *P-loop containing nucleoside triphosphate hydrolases* SF is identified most frequently, in ~130 EC categories (Figure S13), ubiquitous in all forms of life. While ATP is the most common substrate, *P-loop NTPases* also include GTP processing enzymes, e.g. signaling GTPases.

Acetyl CoA synthetase-like and *Glutathione synthetase ATP-binding domain-like* SF enzymes are involved in biosynthetic pathways and specifically use ATP. *Acetyl CoA synthetase-like* is abundant among carbon-sulfur ligases (6.2.-), typically with a carboxylate nucleophile being adenylated. *Glutathione synthetase ATP-binding domain-like* are phosphatases on the other hand, mostly found among peptide ligases (6.3.-).

Kinases of the *Protein kinase-like* and *Actin-like ATPase domain* SFs also cover many EC categories, as they are often associated with distinct ECs corresponding to a range of nucleophiles being phosphorylated.

For a given function, the same structure with a single SF is typically used, and the distribution of the number of SFs present in the same ECs is far less ambiguous. Only six EC categories are linked to four or more SFs (Table S2). Two EC categories 2.7.7.48, RNA-directed RNA polymerase and 2.7.7.19 RNA adenylating enzyme, are the most structurally diverse, both include six SFs. While the *DNA/RNA polymerases* SF, with its many members, covers all organisms from viruses to mammals, other SFs were identified with the same functionality, but specific for negative-sense RNA viruses (*Virion DNA-directed RNA polymerase domain* or *Influenza RNA-dependent RNA polymerase subunit PB1*) or bacteria (*Bacterial DNA polymerase III, alpha subunit, NTPase domain* SF). Four SFs are associated with 2.7.7.7 DNA-directed DNA polymerase.

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

We demonstrate the catalytic function of the cofactor Mg^{2+} on three prototype phosphate-catalytic enzymes: Ras, dUTPase, and the ribonuclease H (RNase H), studied with QM/MM simulations previously [53-55]. Ras is a GTPase, which belongs to the largest identified superfamily, the *P-loop containing NTP hydrolases*, and serves as a representative model for phosphatase enzymes. A representative of the *dUTPase-like* fold, dUTPase is our prototype for a pyrophosphatase enzyme. To demonstrate that our findings are more general beyond NTP processing enzymes, we also included a phosphate cleaving enzyme, RNase H, which uses the classical two-metal ion catalysis.

We investigated changes in the activation energy barrier in the presence and absence of the catalytic Mg^{2+}. In Ras, the mechanism of the hydrolysis is assisted by Gln61 functioning as a transient weak base [53]. Therefore, the mechanism consists of two steps: the phosphate cleavage and a proton transfer to obtain the inorganic dihydrogen phosphate. We found that the absence of the Mg^{2+} at the active site of Ras significantly impairs the associated proton transfer step, resulting in an activation energy increase from 21.4 kcal/mol to 36.4 kcal/mol (see Figure 6**A**). A similarly large difference is observed for the dUTPase, where pyrophosphate cleavage occurs in a single step with Asp95 deprotonating the nucleophilic water [54]. Removing the Mg-pinch-forming, αβγ-coordinated Mg^{2+} elevates the activation energy of the associated catalytic reaction from 18.0 kcal/mol to 39.0 kcal/mol.

Additionally, we included the HIV-1 RNase H in our analysis, as a more general example of phosphate catalytic enzymes. While polymerases incorporate new bases from NTPs into the nucleic acid chain, ribonucleases are cleaving such chains to produce 3'-OH and 5'-P-terminated products (Figure S1). The RNase H active site shares similarities with DNA/RNA polymerases [55, 56], hosting two Mg^{2+} ions by conserved carboxylate motifs. Removing the catalytic ion near the attacking water increases the activation energy, from 19.1 kcal/mol to 39.4 kcal/mol (Figure 6A), demonstrating importance of catalytic Mg^{2+}.

Additional QM calculations were carried out on minimal phosphate catalytic QM models, where a +2 point-charge in the position of the Mg^{2+} is sufficient to emulate its catalytic contribution (Figure S16). We analyzed the differences in electron density upon the replacement of the Mg^{2+} with a +2 charge (Figure 6B and S17). In the case of the pyrophosphatase model originating from the dUTPase, the positive charge draws a significant electron density, reducing it near the phosphate oxygens which are not in contact with the positive charge (or the cation in the full system). The rearrangement in polarization increases the electron density along the forming P-O bond and decreases it along the breaking one, effectively shifting

the transition state towards the product state. Quantifying this effect, the Wiberg bond index of the breaking $P_\alpha - O_\beta$ bond decreases from 0.125 to 0.108, meaning the presence of the Mg^{2+} ion helps cleaving the pyrophosphate. At the same time, the $P_\alpha - O_{nuc}$ bond index increases from 0.330 to 0.356.

We also evaluated the reaction barriers while varying the position of the point-charge around the triphosphate (see Figure 6C-D). The blue regions indicate beneficial ion placement, and concentrate on a torus around the cleaved P–O bond, where a Mg-pinch can be formed (yellow arrow and gray highlight). This also matches the position of arginine fingers [57].

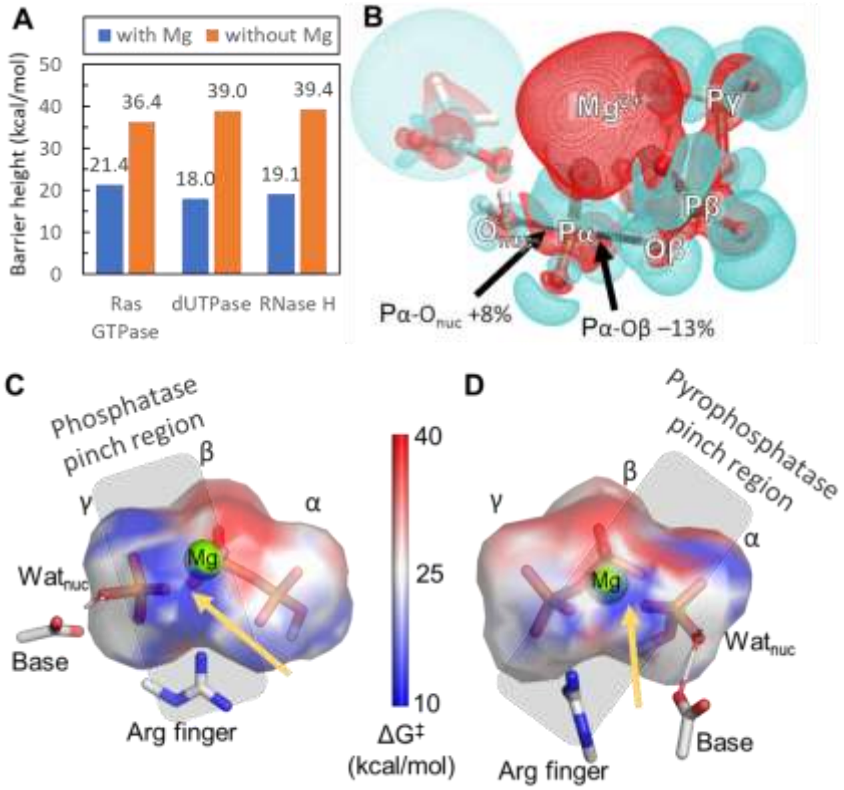

Figure 6. ***A*** *Activation energies in the presence and absence of the catalytic Mg^{2+} ion in QM/MM simulation of three phosphate catalytic enzymes: Ras, dUTPase, and RNase H.*
B *Isosurface of the electron density change upon the introduction of a 2+ point charge at the Mg^{2+} position for the dUTPase 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. Map of the 2+ charge positions around the triphosphate colored by the activation energy to carry out* ***C*** *phosphate,* ***D*** *pyrophosphate hydrolysis. The approximate region where the cation can form a Mg-pinch is highlighted in gray, optimal position by yellow arrows. The original Mg^{2+} (green spheres) and arginine fingers (white sticks) are depicted for orientation only.*

3 Conclusions

Phosphate chemistry is at the center of biological processes. Triphosphates have a unique dominating role in all forms of biological processes, as NTPs are used most frequently as reactants of phosphate processing enzyme reactions. We classify two major categories of NTP reactivity: phosphatase and pyrophosphatase

activities. Both reactivities require metal cations for activation, predominantly Mg^{2+} will function as the catalytic divalent ion.

We hypothesized that the metal ion coordination has a specific structural requirement for its functional role in the catalytic mechanism. In particular, that the Mg^{2+} ion should be coordinated by the phosphates that are involved in the cleaved P-O bond, ie. for phosphatases at least one Mg^{2+} must show an AB (or ABG) coordination while for pyrophosphatases a BG (or ABG) coordination is expected. To test our hypothesis, as well as to enable analysis of future structure-function relationships, we built a comprehensive structural database of distinct NTP processing enzyme superfamilies. We associated SCOPe superfamilies for each NTP processing enzyme when this was possible, and identified InterPro IDs otherwise. We compared the active site geometries within each superfamily using currently available structural data, and highlighted challenging cases where this is currently missing.

We analyzed the experimental structures of enzyme-NTP-cation complexes by their Mg – triphosphate chain interaction and position. We established that the members of a superfamily exhibit a consensus binding mode including conserved motifs of the ion coordination of protein residues. Furthermore, the NTP-Mg^{2+} coordination forms a universal Mg-pinch motif to activate the P-O bond for cleavage, consistently among phosphatases (between the β and γ phosphates) and pyrophosphatases (between the α and β phosphates).

Our analysis of NTP binding sites provides templates for less studied enzymes, when only their reactivity (registered as an EC number) or their sequence information (through assigned superfamilies) is known, as well as if the available structures are incomplete (e.g. missing substrates or ions). Indeed, several examples we identified correspond to cases where standard nucleoside binding prediction fails, even where the binding site is experimentally known. The established structural templates and insights into the Mg-pinch motif can be used in simulations to study enzyme reactivity and guide the development of new inhibitors or activators.

The strategic placement of the activating cation is captured by QM and QM/MM calculations. Our results demonstrate that electrostatic polarization, promoting the transition and product states, is the primary factor responsible for the high catalytic efficiency in phosphate cleavage accomplished by the Mg-pinch motif.

Our comprehensive database matches reactivity with structural alignment, advances our understanding of these ubiquitous biological processes. The insights into phosphate chemistry and NTP processing can contribute to a deeper understanding of fundamental biological mechanisms. The structural database and analysis of NTP processing enzymes can also inform the design of new drugs targeting these enzymes, which are crucial for many diseases, as well as help identify off target toxicity for NTP-analogue inhibitors and inform protein engineering efforts to enhance or modify the activity of NTP processing enzymes.

4 Acknowledgement

Authors acknowledge funding from ERC Starting Grant (Project 757850 BioNet) and EPSRC (grant no. EP/R013012/1). The work is supported by the Ministry for Innovation and Technology and the National Research, Development, and Innovation Office (K135231, NKP-2018-1.2.1-NKP-2018-00005, TKP2021-EGA-02, to B.G.V. and FK142489 to D.B.). D.B. acknowledges funding from the University Research Excellence Program (EKÖP-24-4-II-BME-321).

References

1 Westheimer, F.H. (1987) Why nature chose phosphates. *Science* 235, 1173-1178
2 Kamerlin, S.C*., et al.* (2013) Why nature really chose phosphate. *Q Rev Biophys* 46, 1-132
3 Shepard, S.M*., et al.* (2022) Beyond Triphosphates: Reagents and Methods for Chemical Oligophosphorylation. *J Am Chem Soc* 144, 7517-7530
4 Jakubowski, H. (1986) Sporulation of the yeast Saccharomyces cerevisiae is accompanied by synthesis of adenosine 5'-tetraphosphate and adenosine 5'-pentaphosphate. *Proc Natl Acad Sci U S A* 83, 2378-2382
5 Azevedo, C*., et al.* (2015) Protein polyphosphorylation of lysine residues by inorganic polyphosphate. *Mol Cell* 58, 71-82
6 Knape, M.J*., et al.* (2015) Divalent Metal Ions Mg(2)(+) and Ca(2)(+) Have Distinct Effects on Protein Kinase A Activity and Regulation. *ACS Chem Biol* 10, 2303-2315
7 Rosta, E*., et al.* (2014) Calcium inhibition of ribonuclease H1 two-metal ion catalysis. *J Am Chem Soc* 136, 3137-3144
8 Maguire, M.E. and Cowan, J.A. (2002) Magnesium chemistry and biochemistry. *Biometals* 15, 203-210
9 Krzywoszynska, K*., et al.* (2020) General Aspects of Metal Ions as Signaling Agents in Health and Disease. *Biomolecules* 10
10 Chen, K*., et al.* (2012) Prediction and analysis of nucleotide-binding residues using sequence and sequence-derived structural descriptors. *Bioinformatics* 28, 331-341
11 Hu, J*., et al.* (2018) ATPbind: Accurate Protein-ATP Binding Site Prediction by Combining Sequence-Profiling and Structure-Based Comparisons. *J Chem Inf Model* 58, 501-510
12 Gao, Y. and Yang, W. (2016) Capture of a third Mg(2)(+) is essential for catalyzing DNA synthesis. *Science* 352, 1334-1337
13 Yang, W. (2008) An equivalent metal ion in one- and two-metal-ion catalysis. *Nat Struct Mol Biol* 15, 1228-1231
14 Steitz, T.A. and Steitz, J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* 90, 6498-6502
15 Berman, H.M*., et al.* (2000) The Protein Data Bank. *Nucleic Acids Res* 28, 235-242
16 Kanehisa, M. (2017) Enzyme Annotation and Metabolic Reconstruction Using KEGG. *Methods Mol Biol* 1611, 135-145
17 (1992) *Enzyme Nomenclature - Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*. Academic Press
18 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28, 27-30
19 Kanehisa, M. (2019) Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 28, 1947-1951
20 Kanehisa, M*., et al.* (2023) KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 51, D587-D592
21 Paysan-Lafosse, T*., et al.* (2023) InterPro in 2022. *Nucleic Acids Res* 51, D418-D427
22 Gasteiger, E*., et al.* (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31, 3784-3788
23 Gough, J*., et al.* (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 313, 903-919
24 Pandurangan, A.P*., et al.* (2019) The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res* 47, D490-D494

25 Fox, N.K., *et al.* (2014) SCOPe: Structural Classification of Proteins--extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res* 42, D304-309
26 Chandonia, J.M., *et al.* (2022) SCOPe: improvements to the structural classification of proteins - extended database to facilitate variant interpretation and machine learning. *Nucleic Acids Res* 50, D553-D559
27 Holm, L., *et al.* (2023) DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci* 32, e4519
28 Fawaz, M.V., *et al.* (2011) The ATP-grasp enzymes. *Bioorg Chem* 39, 185-191
29 Gonzalez, B., *et al.* (2010) Inositol 1,3,4,5,6-pentakisphosphate 2-kinase is a distant IPK member with a singular inositide binding site for axial 2-OH recognition. *Proc Natl Acad Sci U S A* 107, 9608-9613
30 Siebold, C., *et al.* (2003) Crystal structure of the Citrobacter freundii dihydroxyacetone kinase reveals an eight-stranded alpha-helical barrel ATP-binding domain. *J Biol Chem* 278, 48236-48244
31 Dunbar, K.L., *et al.* (2012) YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat Chem Biol* 8, 569-575
32 Xiang, Y., *et al.* (2009) Crystallographic Insights into the Autocatalytic Assembly Mechanism of a Bacteriophage Tail Spike. *Molecular Cell* 34, 375-386
33 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
34 Pitcher, R.S., *et al.* (2007) Structure and function of a mycobacterial NHEJ DNA repair polymerase. *J Mol Biol* 366, 391-405
35 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
36 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
37 Cherepanov, A.V. and de Vries, S. (2002) Kinetic Mechanism of the Mg2+-dependent Nucleotidyl Transfer Catalyzed by T4 DNA and RNA Ligases. *Journal of Biological Chemistry* 277, 1695-1704
38 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
39 Grochowski, L.L., *et al.* (2007) Characterization of an Fe(2+)-dependent archaeal-specific GTP cyclohydrolase, MptA, from Methanocaldococcus jannaschii. *Biochemistry* 46, 6658-6667
40 Liu, S., *et al.* (2023) Structural basis of gRNA stabilization and mRNA recognition in trypanosomal RNA editing. *Science* 381, eadg4725
41 Jourdain, A.A., *et al.* (2017) The FASTK family of proteins: emerging regulators of mitochondrial RNA biology. *Nucleic Acids Res* 45, 10941-10947
42 Patil, A.G., *et al.* (2013) Mycobacterium tuberculosis MutT1 (Rv2985) and ADPRase (Rv1700) proteins constitute a two-stage mechanism of 8-oxo-dGTP and 8-oxo-GTP detoxification and adenosine to cytidine mutation avoidance. *J Biol Chem* 288, 11252-11262
43 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 2581(T). *Biochemistry* 49, 2194-2204
44 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. *J Struct Biol* 189, 238-250
45 Tanaka, N., *et al.* (2011) Novel mechanism of RNA repair by RtcB via sequential 2',3'-cyclic phosphodiesterase and 3'-Phosphate/5'-hydroxyl ligation reactions. *J Biol Chem* 286, 43134-43143
46 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 (Amst)* 49, 51-59

47 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
48 Litvin, T.N*., et al.* (2003) Kinetic properties of "soluble" adenylyl cyclase. Synergism between calcium and bicarbonate. *J Biol Chem* 278, 15922-15926
49 Keppetipola, N. and Shuman, S. (2008) A phosphate-binding histidine of binuclear metallophosphodiesterase enzymes is a determinant of 2',3'-cyclic nucleotide phosphodiesterase activity. *J Biol Chem* 283, 30942-30949
50 Yong, S.C*., et al.* (2014) A complex iron-calcium cofactor catalyzing phosphotransfer chemistry. *Science* 345, 1170-1173
51 Snider, J*., et al.* (2008) The AAA+ superfamily of functionally diverse proteins. *Genome Biol* 9, 216
52 Jessop, M*., et al.* (2021) AAA+ ATPases: structural insertions under the magnifying glass. *Curr Opin Struct Biol* 66, 119-128
53 Berta, D*., et al.* (2023) Mechanism-Based Redesign of GAP to Activate Oncogenic Ras. *J Am Chem Soc* 145, 20302-20310
54 Lopata, A*., et al.* (2015) Mutations Decouple Proton Transfer from Phosphate Cleavage in the dUTPase Catalytic Reaction. *ACS Catalysis* 5, 3225-3237
55 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
56 Yang, W*., et al.* (2006) Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Mol Cell* 22, 5-13
57 Nagy, G.N*., et al.* (2016) Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase dUTPases. *J Am Chem Soc* 138, 15035-15045