

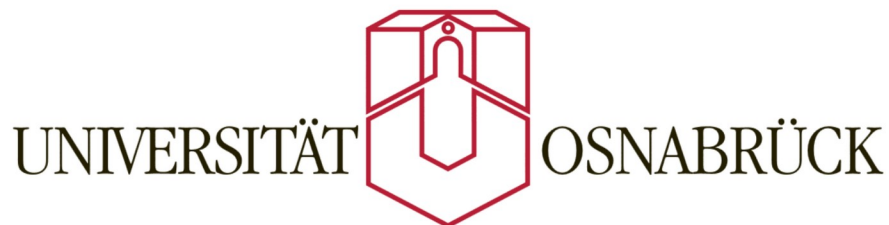
CATALYTIC MECHANISM OF  
P-LOOP FOLD NUCLEOSIDE TRIPHOSPHATASES  
UNRAVELED BY LARGE-SCALE COMPARATIVE  
STRUCTURE ANALYSIS

**Dissertation**

*submitted for the degree of  
Dr. rer. nat. (Doctor of Natural Sciences)  
in the Department of Biology/Chemistry  
University of Osnabrueck, Germany*

*by*

**Dipl.-Biol. Maria Kozlova**



*April 2023*



PRÜFUNGSAUSSCHUSS:

PD Dr. Armen Mulkidjanian — **Erstgutachter**

Prof. Dr. Roland Brandt — **Zweitgutachter**

Prof. Dr. Christian Kost

Dr. Dovilė Janulienė



## Table of Contents

1. INTRODUCTION	4
1.1. ACCUMULATION OF INFORMATION IN PROTEIN STRUCTURE DATABASES OVER TIME	4
1.2. P-LOOP FOLD NUCLEOSIDE TRIPHOSPHATASES	7
1.2.1. Catalysis by P-loop NTPases	9
1.2.2. Generic numbering of key amino acid residues for P-loop NTPases	13
1.2.3. Survey of P-loop NTPases of different classes	14
1.2.3.1. Coordination of the Mg-triphosphate moiety	14
1.2.3.2. Catalytically relevant amino acids, activating partners, stimulatory patterns, and coordination of $W_{\text{cat}}$	16
<i>Kinase-GTPase division</i>	16
<i>ASCE division</i>	21
1.2.3.3. Asymmetric charge distribution in catalytic pockets of P-loop NTPases	26
1.3. AIMS OF THE CURRENT WORK	27
2. METHODS	29
2.1. EVOLUTIONARY BIOPHYSICS APPROACH	29
2.2. METHODS DEVELOPED FOR GLOBAL STRUCTURAL ANALYSIS OF P-LOOP NTPASES	29
2.2.1. Automated analysis of binding sites	29
2.2.1.1. Structure sampling	30
2.2.1.2. Protein family mapping	30
2.2.1.3. Identification of the catalytic sites and of Lys <sup>WA</sup>	31
2.2.1.4. Assessment of the nucleotide integrity and $\gamma$ -phosphate state	31
2.2.1.5. Identification of the residues of Walker A motif	31
2.2.1.6. Identification of the residues of Walker B motif	32
2.2.1.7. Assessment of the Mg <sup>2+</sup> -binding site	32
2.2.1.8. Evaluation of Mg <sup>2+</sup> binding in the $\text{AlF}_4^-$ -containing structures	32
2.2.1.9. Identification of stimulatory moieties and residues interacting with $\gamma$ -phosphate	34
2.2.1.10. Description of individual catalytic sites	35
2.2.1.11. Structure visualization	37
2.2.1.12. Calculation of solvent-accessible surface area	38
3. RESULTS	39
3.1. COMPUTATIONAL COMPARATIVE STRUCTURAL ANALYSIS OF P-LOOP NTPASES	39
3.2. ANALYSIS OF BINDING OF TRANSITION STATE ANALOGS TO P-LOOP NTPASES	41
3.2.1. Variability in the $\text{AlF}_4^-$ binding to catalytic sites of P-loop NTPases	41
3.2.1.1. Different modes of $\text{AlF}_4^-$ interaction with the cofactor Mg <sup>2+</sup> ion	41
3.2.1.2. Untypical binding of $\text{AlF}_4^-$ in RecA ATPases	44
3.3. COMPARATIVE STRUCTURAL ANALYSIS OF STIMULATORY PATTERNS OF P-LOOP NTPASES	46
3.3.1. Stabilization of the O2G atom of $\gamma$ -phosphate by the backbone HNK-3 group	46
3.3.2. Pre-catalytic configurations in NTP-containing structures.	47
3.3.3. Interactions of stimulating moieties with the triphosphate chain	48
3.3.3.1. Stimulatory patterns of Arg fingers	49
3.3.3.2. Stimulatory patterns of Lys fingers	53
3.3.3.3. Interaction patterns of Asn fingers	53

3.3.3.4. Summary on quantitative analysis of stimulatory interactions of Arg, Lys, and Asn fingers in P-loop NTPases	54
3.3.3.5. Stimulatory interactions in ABC ATPases	55
3.3.3.6. Modes of interactions of the stimulators with phosphate chain	55
<i>Linking of <math>\alpha</math>- and <math>\gamma</math>-phosphates by the stimulator</i>	56
<i>Interaction of the stimulator only with <math>\gamma</math>-phosphate</i>	56
3.3.4. Shared features in P-loop NTPase stimulation	57
3.4. ANALYSIS OF HYDROGEN BONDING PATTERNS AND PROTON PATHWAYS IN CATALYTIC SITES OF P-LOOP NTPASES	58
3.4.1. Proton paths and water networks in the catalytic sites of P-loop NTPases	58
3.4.2. The H-bond between Asp <sup>WB</sup> and [Ser/Thr] <sup>K+1</sup> is shorter in presence of TS analogs	62
3.4.3. The solvent accessible surface area of Asp <sup>WB</sup> is diminished in TS-like structures	63
3.4.4. Proton path connects Asp <sup>WB</sup> with W <sub>cat</sub> in all classes of P-loop NTPases	64
4. DISCUSSION	66
4.1. COMMON FEATURES AS OBSERVED FOR TS ANALOG-CONTAINING STRUCTURES	66
4.1.1. Comparison of metal fluorides as TS analogs	66
4.1.1.1 Non-physiological binding of AlF <sub>4</sub> <sup>-</sup> to the cofactor Mg <sup>2+</sup> ion	66
4.1.1.3. Untypical binding of AlF <sub>4</sub> <sup>-</sup> in RecA ATPases	67
4.1.2. Role of mechanistic bonding in the common stimulation mechanism of P-loop NTPases	69
4.1.3. Constriction of the catalytic site in the transition state	71
4.1.4. Transition state analogs and energetics of P-loop NTPases	72
4.1.5. Summary on common structural traits of P-loop NTPases	75
4.2. COMMON MECHANISM OF P-LOOP NTPASES AS INFERRED FROM GLOBAL COMPARATIVE STRUCTURE ANALYSIS	75
4.2.1. Requirements for catalysis	75
4.2.2. Does the [Ser/Thr] <sup>K+1</sup> – Asp <sup>WB</sup> pair accept a proton from W <sub>cat</sub> ?	78
4.2.2.1. Comparison with known proton-transfer modules	81
4.2.2.2. Evidence for [Ser/Thr] <sup>K+1</sup> – Asp <sup>WB</sup> as a transient proton trap	83
4.2.2.3. Compensation of the negative charge on the bridging O <sup>3B</sup> atom upon hydrolysis	88
4.2.2.4. Assignment of novel functions to the Walker motifs	88
4.2.2.5. Minimal mechanistic model of NTP hydrolysis by P-loop NTPases	90
5. CONCLUSIONS	93
6. OUTLOOK	95
7. BIBLIOGRAPHY	96
8. SUMMARY	120
9. ZUSAMMENFASSUNG	121
10. ABBREVIATIONS	122
11. LIST OF FIGURES	124
12. LIST OF TABLES	126
13. PUBLICATIONS	127
14. ACKNOWLEDGMENTS	129

---

15. APPENDIX	130
A. SUPPLEMENTARY FIGURES	131
B. TABLE B.1: DESCRIPTION OF REPRESENTATIVE STRUCTURES OF P-LOOP NTPASES	137
C. TABLE C.1: RESULTS OF THE COMPUTATIONAL ANALYSIS OF ALL AVAILABLE STRUCTURES OF THE P-LOOP NTPASES WITH FULL-FLEDGED CATALYTIC SITES CONTAINING MG-NTP SUBSTRATE OR ITS ANALOGS	139
D. TABLE D.1: COORDINATION OF THE $Mg^{2+}$ ION IN THE $AlF_4^-$ -CONTAINING STRUCTURES OF P-LOOP NTPASES	140
E. TABLE E.1: RELATIVE OCCURRENCE OF DISTINCT STIMULATORY PATTERNS FOR ARG, LYS, AND ASN "FINGERS"	141
F. NOTES ON RARITY OF Y-PATTERN OF STIMULATORY ARG INTERACTION	142
G. EVIDENCE FROM IR SPECTROSCOPY	150
H. SOFTWARE AVAILABILITY	152
16. ANLAGEN	153
Lebenslauf	153
Erklärung über die Eigenständigkeit der erbrachten wissenschaftlichen Leistung	154

# 1. Introduction

## 1.1. Accumulation of information in protein structure databases over time

The determination of structures of biopolymers and their analysis plays a key importance in biological research. It is well established that the functions of proteins, including their ability to accelerate chemical reactions by many orders of magnitude are fully determined by their structural features. Any significant distortion of a protein structure inevitably leads to the loss of catalytic function. Structure (including sequence, the primary protein structure) seems to also determine protein dynamics [1]. In living systems, the information appears to be stored in the structure: it is possible to cool a small animal to  $-271^{\circ}\text{C}$ , dramatically reducing the rates of dynamic processes, and it still comes back to life after thawing [2].

As protein structures are of key importance for the understanding of biological processes, structure determination has long been one of the primary objectives of the biological research. Since the announcement of Protein Data Bank (PDB) [3–5] in 1971 at Brookhaven National Laboratory, the number of deposited structures kept growing to reach 10000 structures in 1999, 100'000 in 2014 and 200'000 structures in January 2023. Since the early days of protein structure determination, majority of structures were derived from X-ray crystallography experiments; and while cryo-electron microscopy (cryo-EM) is used to solve 3D structures of biopolymers since 1991, the recent advances in the field has brought the attainable resolution of EM structures down to atomic resolution [6,7], and the number of structures determined with this method now grows exponentially (see Fig. 1.1). The paradigm shift in structural biology is aided by application of deep neural networks for exploration of the sequence – structure relationship in proteins [8–11]. Such approaches have allowed to perform predictions of secondary and tertiary (3D) structures up to atomic resolution. However, these methods still have limited applicability for predictions of bound ligands, interactions with the solvent, disordered regions and functional information on the higher order structure levels (multimeric proteins and large protein complexes).



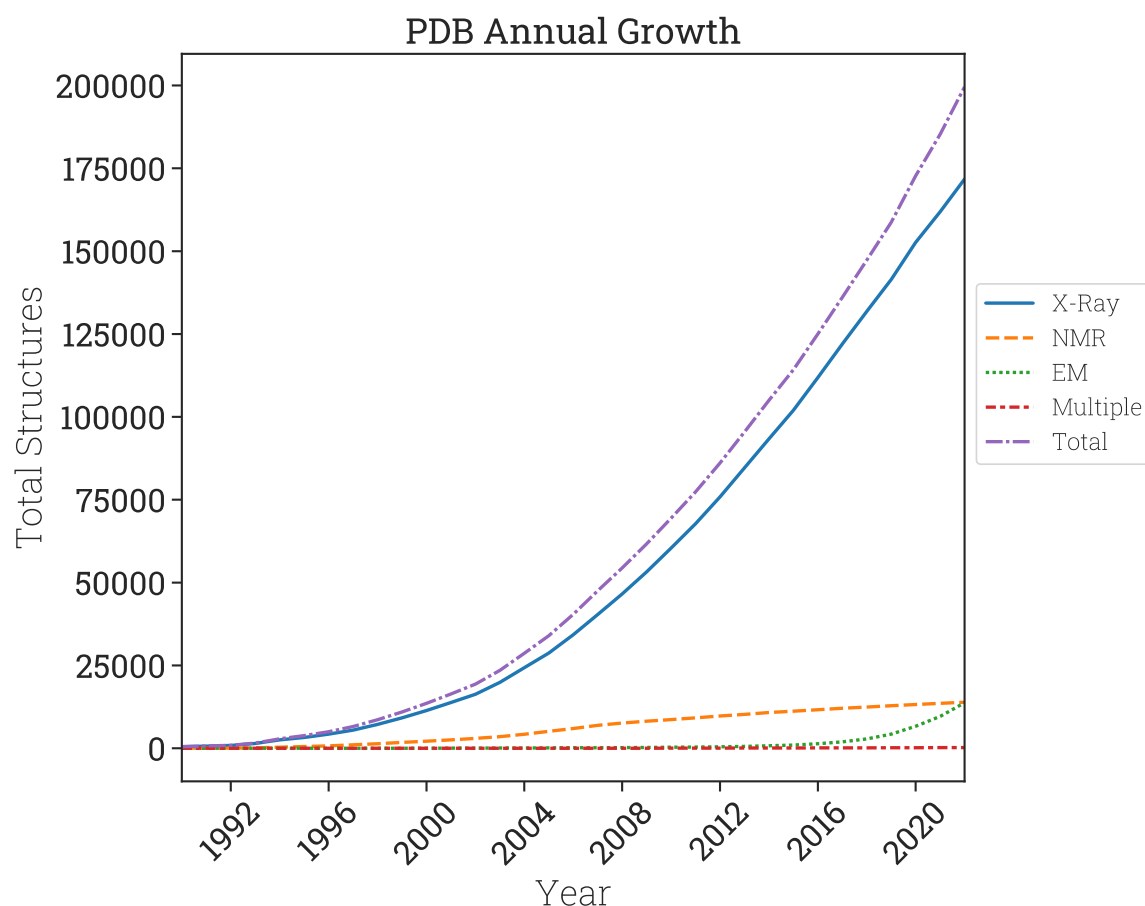


Figure 1.1: Protein Data Bank annual growth.

Number of PDB structures released annually by technique, data retrieved from <https://www.rcsb.org/stats> on 11.05.2022. Structure determination methods listed here include X-Ray crystallography (X-Ray), nuclear magnetic resonance spectroscopy (NMR, here comprising both solution NMR and solid state NMR) and electron microscopy methods (EM, including cryo-electron microscopy).

While last years were marked by a revolution in protein structure prediction methods [8–11], studies involving functional comparative analysis of experimentally determined structures still rely mostly on a small number of so called “representative” structures while connecting these with larger datasets of sequence data (see for instance [12–14]). While programs for structure superimposition allow alignment of multiple structures, this is a computationally intensive task and the number of structures to be superimposed is usually limited (i.e., for DALI, the maximum number of input structures is 64) [15]. Meanwhile, as much as 74 of the protein superfamilies (“clans” and domains not assigned to any Pfam clan) represented in Pfam database [16] release 35.0 include over a thousand unique solved structures, with four superfamilies having more than five thousand PDB structures mapped (see Table 1.1). High numbers of structures solved for these protein families reflect not only their diversity and widespread distribution, but also the abundance of medically relevant proteins among them.

To integrate this overwhelming amount of structural data and gain insights in the mechanism of hydrolysis of nucleoside triphosphates by P-loop fold nucleoside triphosphatases, hereafter P-loop NTPases, we decided to ground our analysis of their catalytic sites in comparisons of the positioning of amino acid residues relative to the atoms of substrate triphosphate chain that is known to attain the same configuration across the whole enzyme superfamily. Similar approach might be used for identification of the common determinants of enzymatic activity also in proteins that share a substrate but are not known to be homologous.

*Table 1.1: Most prevalent protein folds among solved structures in PDB (data retrieved on 07.11.2022)*

	Pfam clan ID	Clan name	Description	Structures in PDB	Families in Pfam
1	CL0063	NADP_Rossmann	FAD/NAD(P)-binding Rossmann fold superfamily, dinucleotide-binding $\alpha/\beta$ domain characterized by an alternating motif of $\beta\alpha\beta$ units [17].	8534	<b>209</b>
2	CL0023	P-loop_NTPase	P-loop NTPases	8411	245
3	CL0016	PKinase	Protein kinase superfamily, serine/threonine- and tyrosine- protein kinases [18]	7446	40
4	CL0123	HTH	Helix-turn-helix clan, a common DNA-binding domain [19].	5406	381
5	CL0011	Ig	Immunoglobulin superfamily. Mainly $\beta$ proteins that include antibodies, the giant muscle kinase titin and receptor tyrosine kinases [20].	3943	34
6	CL0124	Peptidase_PA	Peptidases with the trypsin fold, all $\beta$ proteins.	3927	26
7	CL0020	TPR	Tetratricopeptide repeat superfamily. The 34aa repeats have a loosely conserved motif and form amphipathic $\alpha$ -helices that serve as protein interaction modules [21,22]	3794	252
8	CL0159	E-set	Ig-like fold superfamily (E-set)	3394	257
9	CL0036	TIM_barrel	Proteins with eightfold ( $\beta\alpha$ ) barrel fold that contain a common phosphate binding site [23,24].	3387	61
10	CL0021	OB	The OB (oligonucleotide/oligosaccharide binding) fold is a five-stranded $\beta$ -sheet forming a closed $\beta$ -barrel. This barrel is capped by an $\alpha$ -helix located between the third and fourth strands [25]	3158	113

## 1.2. P-loop Fold Nucleoside Triphosphatases

Hydrolysis of nucleoside triphosphates (NTPs), such as ATP or GTP, by P-loop fold NTPases (also known as Walker NTPases after John Walker and colleagues, who first discovered similar amino acid motifs in very diverse NTPases [26]) is one of the key enzymatic reactions. P-loop NTPases are coded by up to 20% gene products in a typical cell. P-loop NTPase domains drive the activity of rotary ATP synthases, DNA and RNA helicases, kinesins and myosins, ABC transporters, as well as most GTPases, including ubiquitous translation factors,  $\alpha$ -subunits of signaling heterotrimeric G-proteins and oncogenic Ras-like GTPases [26–37]. In the ECOD database [38], the topology-level entry “P-loop\_NTPase” contains 193 protein families. In the Pfam database [39], the P-loop NTPase clan CL0023 contains 217 families. The main classes of P-loop NTPases were already present in the Last Universal Cellular Ancestor (LUCA) [32,40,35,37,41–45].

The P-loop fold domain is a 3-layer  $\alpha\beta\alpha$  sandwich, see Fig.1.2A,B and [32,36,46–49]. In small P-loop NTPases (Fig. 1.2B), the eponymous *p*hosphate-binding loop (P-loop) usually connects the first  $\beta$ -strand ( $\beta_1$ -strand) with the first  $\alpha$ -helix ( $\alpha_1$ -helix); the P-loop together with the two first residues of the  $\alpha_1$ -helix have the [G/A]xxxxGK[S/T] sequence, known as the Walker A motif [26] (Fig 1.2C). This motif is responsible for the binding of the triphosphate chain and cofactor  $Mg^{2+}$  ion, see [26,27,50] and Fig. 1.2C-E. The Walker B motif *hhhhD*, where ‘*h*’ denotes a hydrophobic residue, is the other shared motif of P-loop NTPases, see Fig. 1.2C-E and [26,27,50]. The aspartate residue of the Walker B motif (Asp<sup>WB</sup>) is thought to contribute to the stabilization of the  $Mg^{2+}$  binding site. Asp<sup>WB</sup> makes a hydrogen bond (H-bond) with one of the water molecules in the coordinating sphere of the  $Mg^{2+}$  ion (see Fig 1.2D-E). In small P-loop NTPases, this residue is at the C-terminal tip (hereafter C-cap) of the  $\beta_3$ -strand [40], just opposite of the  $\alpha_1$ -helix.

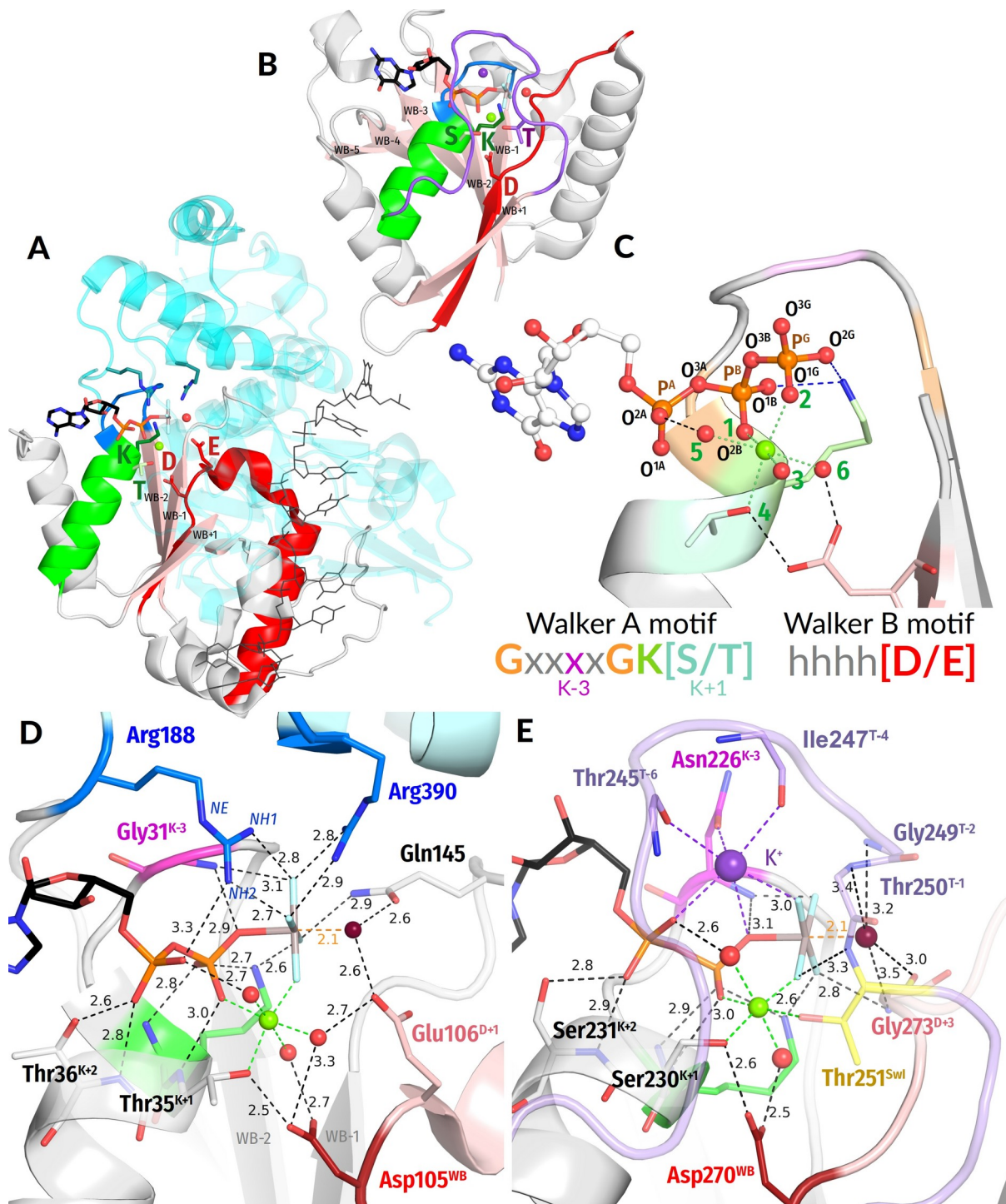


Figure 1.2: P-loop NTPases.

**A,B** Typical folds of P-loop NTPases: **A**, Superfamily 1 helicase Pif1p with a transition state (TS) analog ADP:AlF<sub>4</sub><sup>-</sup> (PDB ID 5FHD [61]); **B**, K<sup>+</sup>-dependent GTPase MnmE with a TS analog GDP:AlF<sub>4</sub><sup>-</sup> bound (PDB ID 2GJ8 [62]). **C**, conserved motifs in P-loop NTPases and Mg:NTP binding by P-loop, naming of atoms according to IUPAC recommendations for nucleoside triphosphates [84] and typical Mg<sup>2+</sup> coordination. **D, E**, crystal structures of typical catalytic sites of P-loop NTPases: **D**, Pif1p helicase, structure shown in panel **A**; **E**, GTPase MnmE, structure shown in panel **B**. Color code A-B: Polypeptide chains are shown as gray cartoons, β-strands, pink; α<sub>1</sub>-helix, green; P-loop, blue; Walker B motif and the following residues, red; Switch I loop, lilac; the activating partners are colored cyan. Color code D-E: nucleotide analogs and important amino acid residues are shown as sticks, water molecules – as red spheres, Mg<sup>2+</sup> ions -- as lime spheres, K<sup>+</sup> – as a purple sphere; the conserved Lys residue of Walker A motif is shown in green, and the residue three residues before it is highlighted in magenta; the conserved Asp residue of Walker B motif is shown in dark red, and the Arg fingers are shown in blue. In those amino acid residues that are shown as sticks, the oxygen atoms are colored red, and the nitrogen atoms are colored blue. In the AlF<sub>4</sub><sup>-</sup> moiety, the Al atom is colored gray, and the fluoride atoms are colored light blue. Switch I/K-loop motif is shown in lilac, the conserved Thr<sup>SwI</sup> is highlighted in yellow.

### 1.2.1. Catalysis by P-loop NTPases

P-loop NTPases catalyze a plethora of phosphoryl transfer reactions from ATP or GTP (Fig. 1.3). Vast majority of these enzymes hydrolyze ATP or GTP to  $P_i$  and ADP or GDP, respectively. P-loop kinases, on the other hand, transfer a phosphoryl group from ATP to diverse substrates (hereafter “second substrates”), such as sugars, nucleosides, nucleotides, polynucleotides, coenzyme precursors and multiple other compounds [40]. As shown in Fig. 1.3, the catalytic reaction includes deprotonation of the attacking moiety (the second substrate in kinases and “catalytic” water in all other P-loop NTPases) yielding a nucleophile that attacks the phosphorus atom of  $\gamma$ -phosphate [51–54]. It is thought that the proton taken from the attacking moiety ultimately compensates for the negative charge that develops on  $\beta$ -phosphate group concomitantly with breaking away of  $\gamma$ -phosphate. This negative charge is thought to be responsible for the activation barrier of hydrolysis [55–57].

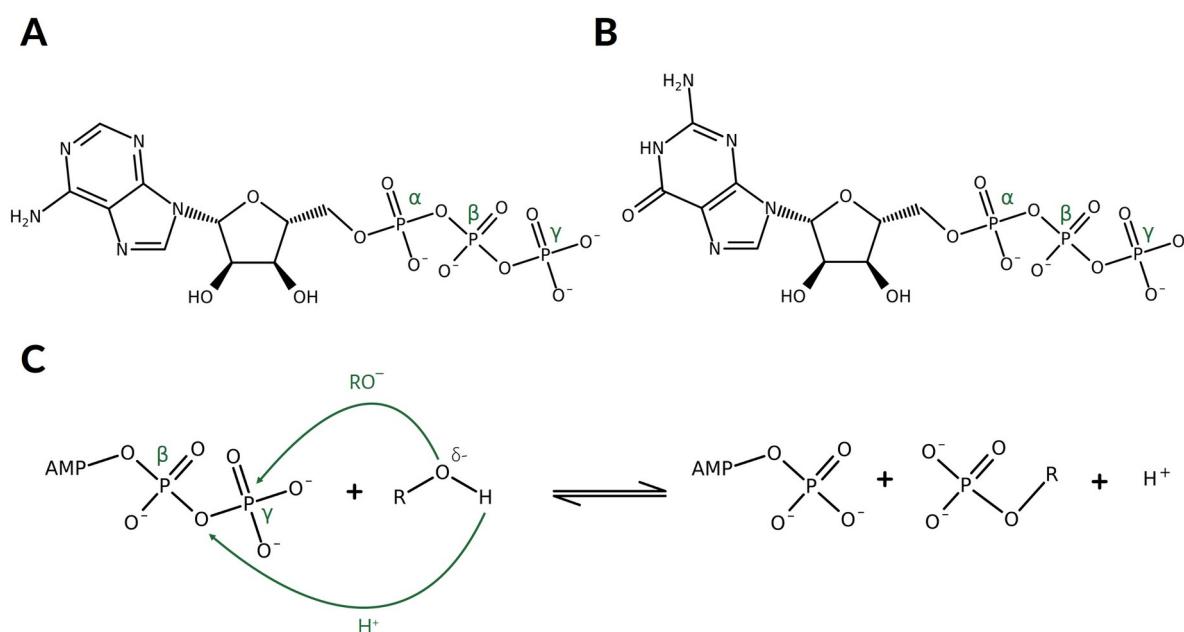


Figure 1.3: Nucleoside triphosphates (NTPs) and NTP hydrolysis.

**A.** Adenosine triphosphate (ATP). **B.** Guanosine triphosphate (GTP). **C.** Phosphoryl transfer from ATP to a nucleophile. In case of ATP hydrolysis,  $R=H$ . GTP hydrolysis occurs in a similar manner.

As a rule, large-scale conformational changes accompany both the binding of NTP by P-loop domain (“closing” of the catalytic pocket) and the NTP hydrolysis itself (“opening” of the catalytic pocket). In most cases, these structural changes are coupled to useful mechanical work and drive most biomolecular motors, see, e.g. [29,58–60].

Since uncontrolled NTP hydrolysis would be dangerous for cell survival, a feature of P-loop NTPases is that they have to be activated before each turnover. Usually, an ATP- or GTP-bound P-loop domain interacts with its cognate activating partner, which could be a domain of the same protein, a separate protein and/or a DNA/RNA molecule. Upon this interaction, specific stimulatory moieties, usually Arg or Lys residues (“Arg/Lys

finger(s)", Fig. 1.2C), are inserted into the catalytic site and initiate the cleavage of  $\gamma$ -phosphate [29,34,61–69].

Important hints for clarifying the catalytic mechanism of P-loop NTPases are provided by their structures with bound transition state (TS) analogs, such as NDP:AlF<sub>4</sub><sup>-</sup> or NDP:MgF<sub>3</sub><sup>-</sup> or NDP:VO<sub>4</sub><sup>-</sup> complexes (see Fig 1.4A,B for compound structures and geometry) [29,61–63,68,70–74]. Particularly many structures were obtained with metal fluorides as transition state analogs. Beginning in 1987, that is, even before the first GDP:AlF<sub>4</sub><sup>-</sup>-containing structures were resolved in 1994 (see [62,63]) metal fluorides were shown to promote binding of various P-loop NTPases to their activating partners. Generally, many NTPases interacted with their activators only in the presence of metal fluorides [68,69,71,74,64,75–81].

Therefore, it is anticipated that these analogs promote the formation of a catalytically productive configuration of the active site. The crystal structures with NDP:MgF<sub>3</sub><sup>-</sup> or NDP:AlF<sub>4</sub><sup>-</sup> bound (see Fig. 1.4C and 1.4D, and [70,62,63,68,72,74,64,82,83] respectively) revealed a "catalytic" water molecule  $W_{\text{cat}}$  located apically to the P<sup>G</sup> atom and almost in-line with its bond with the O<sup>3b</sup> oxygen atom (see Fig. 1.2C for the atom names according to the IUPAC recommendations for nucleoside triphosphates [84]). In the NDP-VO<sub>4</sub><sup>-</sup> complexes (Fig. 1.4A), one of the four oxygen atoms of vanadate occupies the position of the catalytic water molecule [29,73]. These structures, in support to earlier suggestions [51–53], indicate that the ultimate cleavage of  $\gamma$ -phosphate is triggered by the apical nucleophilic attack on the terminal phosphorus atom P<sup>G</sup> by a deprotonated nucleophile that is  $\text{OH}_{\text{cat}}^-$  in case of GTPases and ATPases, see Fig. 1.2D-E, 1.4C-D and [34,51,52,54,57,68,74,85–87]. Furthermore, the structures with bound TS analogs often disclose how a stimulator residue interacts with the triphosphate chain, see Fig. 1.4C-D.

In spite of availability of  $W_{\text{cat}}$  containing, TS-like structures, it has remained unclear how the deprotonation of  $W_{\text{cat}}$  is triggered and what is the fate of the proton that is released upon this deprotonation. The first TS-like crystal structures were obtained for GTPases, such as shown in Fig. 1.4C, and their relatives. These structures did not reveal potential bases among the ligands of  $W_{\text{cat}}$  [29,62,63,71]. Thus, it was suggested that the proton is accepted directly by  $\gamma$ -phosphate [88–90]. In contrast, the subsequently obtained crystal structures of many other P-loop NTPases showed that  $W_{\text{cat}}$  interacts with Glu or Asp residues; these "catalytic" residues were suggested to serve as proton acceptors in these NTPases, see e.g. Glu106 in Fig. 1.2D, Glu253 in Fig. 1.4D and [31,72,87,91]. It remains, however, unclear whether the proton is transferred from these bases to the triphosphate and, if so, how this transfer takes place.

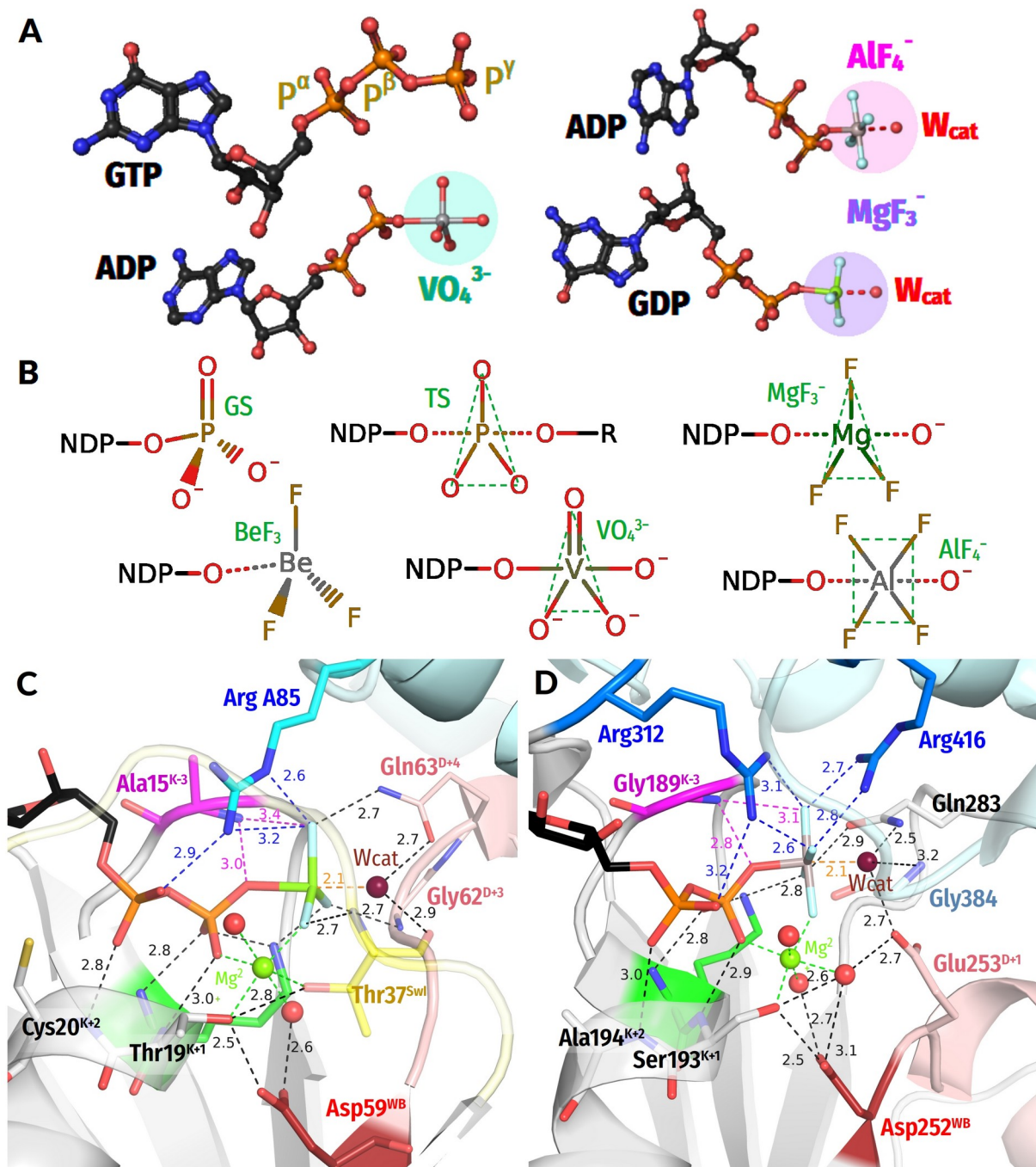


Figure 1.4: Transition state-like complexes of P-loop NTPases.

**A**, Transition states analogs compared to native NTPs. For the sake of clarity, each type of NTP analog is shown either for ATP or GTP-mimicking version. **B**, Geometry of metal complexes used as  $\gamma$ -phosphate mimics upon structure determination. For comparison, the ground state (GS) tetrahedral geometry and the transition state (TS) geometry is shown. The  $AlF_3$  moiety has geometry and charge similar to the  $BeF_3$  and is used as a mimic of the pre-catalytic state. **C, D**, Examples of catalytic sites of P-loop NTPases complexed with TS analogs. **C**, GTPase Rho with its activator RhoGAP and the transition state analog GDP- $MgF_3^-$ , (PDB ID 1OW3 [70]); **D**, Chikungunya virus nsP2 SF1 helicase complexed with transition state analog ADP- $AlF_4^-$  (PDB ID 6JIM [371]).

There is also no consensus on how do stimulators initiate the hydrolysis. Warshel and colleagues suggested that stimulators, owing to their positive charge, compensate the negative charges of oxygen atoms of triphosphate and make the  $\gamma$ -phosphorus atom ( $P^\gamma$ )

more prone to the nucleophilic attack by a water molecule [92–94]. It was also suggested that the pushing of bound water molecules out of the binding pocket into the bulk solution by the stimulatory arginine finger provides an entropic gain [95]. Jin and colleagues proposed reorientation of the  $W_{\text{cat}}$  molecule into the attack position by the stimulator [96–98]. On a later reaction step, stimulators were implied in compensating for the negative charge that develops at  $\beta$ -phosphate upon the breakaway of  $\gamma$ -phosphate [56,57]. Recently, Gerwert and colleagues proposed, for small GTPases, that the stimulatory Arg finger rotates the  $\alpha$ -phosphate group towards an eclipsed conformation with respect to  $\beta$ - and  $\gamma$ -phosphates, which would destabilize the triphosphate chain [99–101].

Elsewhere [102] our group has focused on P-loop NTPases that are stimulated not by Arg/Lys residues but by  $K^+$  and  $Na^+$  ions, see [61,66,103–106]. Molecular dynamics (MD) simulations of the  $Mg^{2+}$ -GTP-containing dimer of  $K^+$ -dependent tRNA-modifying enzyme MnmE (see Fig. 1.2E) were performed and compared with MD simulations of water-dissolved  $Mg^{2+}$ -GTP and  $Mg^{2+}$ -ATP complexes in the presence of monovalent cations. In the latter case, one of the monovalent cations got bound between the  $O^{2A}$  and  $O^{3G}$  atoms of  $\alpha$ - and  $\gamma$ -phosphates, in the so-called AG site. This binding was accompanied by rotation of  $\alpha$ -phosphate - as the least constrained phosphate group - yielding a fully eclipsed configuration of the triphosphate chain [102]. These data agreed with the results of Gerwert and colleagues who used a  $Mg^{2+}$ -methyltriphosphate complex - as a simple mimic of  $Mg^{2+}$ -GTP - in their modeling the interaction with an Arg finger [99].

However, MD simulations of the GTP-containing MnmE protein dimer showed that  $\alpha$ -phosphate was fully immobilized by bonds between the GTP molecule and the residues of the P-loop: these interactions stabilized the triphosphate chain in an extended, supposedly catalytically prone conformation. In this case, only the terminal  $\gamma$ -phosphate retained some mobility. The  $K^+$  ion, by binding between the  $O^{2A}$  and  $O^{3G}$  atoms of the protein-bound GTP molecule, caused twisting of the  $\gamma$ -phosphate group by approximately  $30^\circ$  [102]. It is important that the  $K^+$  ion can hardly link these two oxygen atoms without twisting  $\gamma$ -phosphate. The rotated  $\gamma$ -phosphate was stabilized by a new H-bond between the  $O^{2G}$  atom of  $\gamma$ -phosphate and the backbone amide group three residues before the conserved lysine residue of Walker A motif (see Supplementary Figure A.1 in Appendix A). Both the twisting of  $\gamma$ -phosphate and its stabilization by an additional H-bond could potentially promote hydrolysis.



### 1.2.2. Generic numbering of key amino acid residues for P-loop NTPases

Historically, different families of P-loop NTPases were studied by different research communities, each developing its own terminology. Therefore, no generic amino acid numbering for P-loop NTPases exists. For convenience, and by analogy with the Ballesteros-Weinstein numbering scheme for the G-protein coupled receptors [107], we introduce here a generic residue numbering for the conserved regions of P-loop NTPases. Without this generic nomenclature, the structural comparison of different classes of P-loop NTPases would be hardly possible.

First, we chose highly conserved residues as benchmark references. In the Walker A motif, this is the conserved Lys residue ( $K^{WA}$ ) that forms H-bonds with the  $O^{1B}$  oxygen atom of  $\beta$ -phosphate and  $O^{2G}$  atom of  $\gamma$ -phosphate, see Fig. 1.2. In the Walker B motif, we have chosen the conserved Asp<sup>WB</sup> residue ( $D^{WB}$ ) that is involved in the coordination of the  $Mg^{2+}$  ion (see Fig. 1.2).

In addition to universal Walker A and Walker B motifs, we also invoked the Switch I motif that is located between the Walker A and Walker B motifs in NTPases of TRAFAC (TRANslation FACtors) class, see Fig. 1.2B,E and [40,108]. The Switch I has only a single strictly conserved [Thr/Ser]<sup>SwI</sup> residue ( $[T/S]^{SwI}$ , colored yellow in Fig. 1.2E) that can be used as a reference. In NTPases of the TRAFAC class, the side chain of this  $[T/S]^{SwI}$  residue coordinates the  $Mg^{2+}$  ion, its backbone amino group (hereafter **HN**) forms a H-bond with  $\gamma$ -phosphate, and its backbone carbonyl group (hereafter **CO**) stabilizes  $W_{cat}$ , see Fig. 1.2E.

In the following, we number the amino acids of the Walker A, Switch I, and Walker B motifs relatively to the reference residues, as shown in Fig. 1.2. Then the “catalytic” glutamate residue that follows Asp<sup>WB</sup> in the sequence of many P-loop NTPases is labeled as Glu<sup>D+1</sup>. Although a few P-loop NTPases have a glutamate residue as a conserved carboxylic residue of Walker B motif, this residue is hereafter denoted as Asp<sup>WB</sup> and not as [Asp/Glu]<sup>WB</sup> for the sake of simplicity (see Fig. 1.2D-E), unless a specific role of a Glu residue in this position is addressed. In such rare cases the subsequent residue is denoted as Xxx<sup>E+1</sup>.

Historically, different numbering was used for  $\beta$ -sheets in different families of P-loop NTPases due to variations of the common fold (see [40,50]). For ease of reference, we also introduced a novel generic numbering for the strands of the  $\beta$ -pleated sheet (see [109]). For comparison, see Figure A.2 in Appendix A for classical numbering of  $\beta$ -strands [32,40]. The  $\beta$ -strand which is positioned in-line with the triphosphate chain is referred as “Walker B strand” or *WB-strand*. This strand is easy to find in a structure because it carries an Asp<sup>WB</sup> or, rarely, a Glu<sup>WB</sup> residue at its C-cap; the carboxyl of this residue usually interacts with  $Mg^{2+}$ -coordinating water molecules, as shown in Fig. 1.2D-

E. In addition, this carboxylic residue is usually preceded by four non-polar amino acids of the Walker B motif, see Fig. 1.2C and [26]. Other strands of the same  $\beta$ -pleated sheet, independently of their position in the amino acid sequence, can be numbered by their position relatively to WB-strand as WB-1, WB-2, or WB+1, WB+2 and so on, see Fig. 1.2.

In those cases where the crystal structures of NTPases with their cognate activating partners are available, according to our observations, these partners usually interact with the P-loop domain via the stretch of amino acids that follows the Walker B motif and approximately corresponds to residues from Asp<sup>WB+1</sup> (aka D+1) to Asp<sup>WB+12</sup> (aka D+12) [109]; this stretch is colored red in Fig. 1.2A-B or pale red in Fig. 1.2D-E. In the case of small NTPases, this region corresponds to their  $\alpha_3$ -helices. The amino acids beyond the conserved Asp<sup>WB</sup> residue show no conservation throughout P-loop NTPases; hence, they make no conserved motif. By its shape, this stretch resembles a cock's crest and usually rises above the neighboring  $\beta$ -strands as clearly seen in Fig. 1.2A-B. Therefore, hereafter we will name this structural element "Walker B crest" or WB-crest.

To distinguish between (i) the interaction of the P-loop domain with its activating partner from (ii) the following insertion of, say, an Arg finger into the catalytic site, we call those elements that are inserted into catalytic sites to stimulate hydrolysis 'stimulatory moieties' or 'stimulators'; see, for instance, the dark blue Arg fingers in Fig 1.2D or the K<sup>+</sup> ion in Fig. 1.2E. By using these terms we follow Wittinghofer and colleagues who wrote about "GAP stimulated GTPase activity" in one of their pioneering works [110].

### 1.2.3. Survey of P-loop NTPases of different classes

The P-loop NTPases are thought to form two divisions, namely the Kinase-GTPase division and ASCE (Additional Strand Catalytic E (glutamate)) division, with both divisions containing several enzyme classes, see Fig. A.2 in Appendix A and [32,40,111,112,37].

Below we describe the best studied classes. For classes of P-loop NTPases with reliable TS-like structures available, we have selected representatives out of a set of proteins with available crystal structures in the Protein Data Bank (PDB) at [www.rcsb.org](http://www.rcsb.org) [4,5]. These structures are listed in Table B.1 in Appendix B. Each selected structure contains a Mg<sup>2+</sup> cation and an NTP molecule or its analog bound; we selected the structures that contained TS analogs, if available.

#### 1.2.3.1. Coordination of the Mg-triphosphate moiety

The Figure 1.5 shows representative structures of P-loop NTPases from different classes that were superposed with the structure of AlF<sub>4</sub><sup>-</sup>-containing, K<sup>+</sup>-dependent GTPase MnME (PDB ID 2GJ8, resolution 1.7 Å [61]). We aligned 20 amino acids of the  $\beta_1$ -strand, P-loop, and  $\alpha_1$ -helix with the corresponding amino acids 217-236 of MnME.

The whole shape of the P-loop found to be strictly conserved across all classes of P-loop NTPases (Fig. 1.5A). Accordingly, the binding mode of the triphosphate chain, which is described below, appears to be conserved throughout P-loop NTPases.

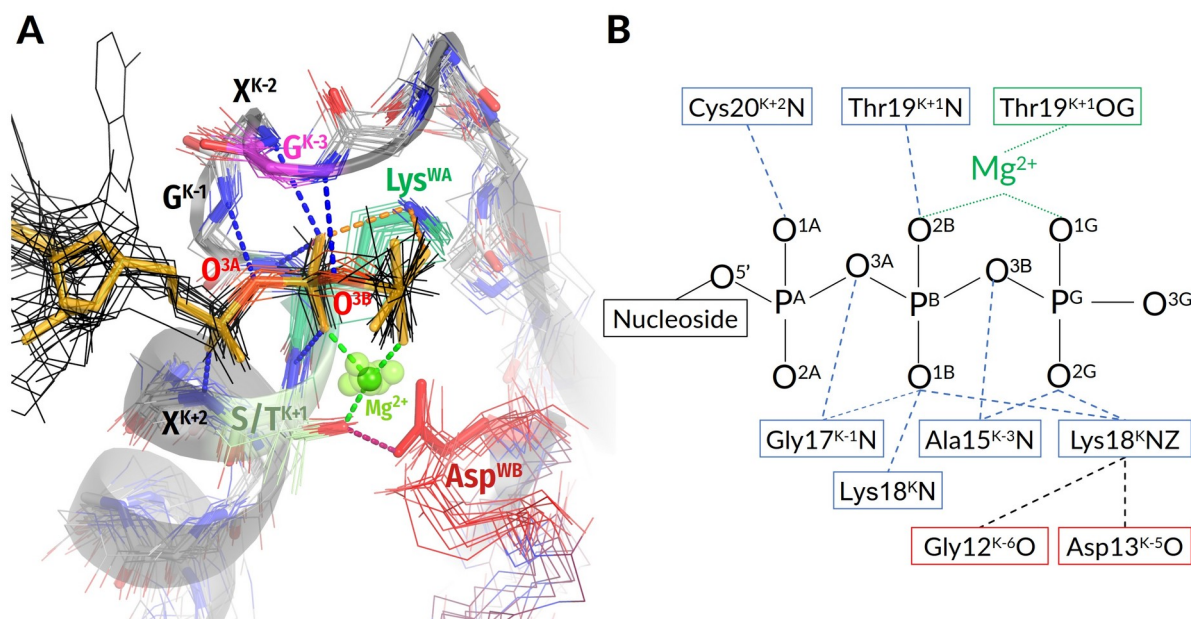


Figure 1.5: Conserved features in  $Mg^{2+}$ :NTP binding in P-loop NTPases.

A. Structures of proteins, representing different classes of P-loop NTPases and described in Table B.1 in Appendix B are shown superimposed on the P-loop region of the MnmeE GTPase with a TS analog  $GDP:AlF_4^-$  bound (PDB ID: 2GJ8 [61]). For the MnmeE GTPase, the P-loop region is shown as a cartoon. The bonding pattern is shown only for the MnmeE GTPase. For all the proteins, the last seven residues from the motif  $GxxxxGK[IS/T]x$ , as well as the Asp residue of the Walker B motif and NTP analogs are shown by lines and colored as follows:  $Lys^{WA}$  in green,  $Ser/Thr^{K+1}$  in pale green,  $Gly/Ala/Asn^{K-3}$  in purple,  $Asp^{WB}$  in red; NTP molecules or their analogs are shown in black with bridging oxygen atoms in red.  $GDP:AlF_4^-$  from the MnmeE structure is shown in orange. Other protein residues are shown in gray with backbone nitrogen and oxygen atoms shown in blue and red, respectively.  $Mg^{2+}$  ions are shown as green spheres. The phosphate chain is involved in numerous bonds with backbone HN groups of Walker A motif residues (the bonds are highlighted in blue), conserved  $Lys^{WA}$  contacts  $O^{2G}$  and  $O^{1B}$  atoms (orange bonds) and  $O^{2B}$  and  $O^{1G}$  atoms are part of the  $Mg^{2+}$  coordination shell (green bonds). Structures were superimposed by  $\alpha_1$  helix, P-loop and the following  $\beta$ -strand in Pymol v. 2.5.0. B. Coordination bonds and H-bonds between the Walker A motif and  $Mg$ -NTP moiety in the MnmeE GTPase (PDB ID: 2GJ8).

In all manually inspected structures, the NTP molecule (or its analog) is bound to the P-loop of the Walker A motif in an extended conformation which is supposedly catalytically prone (Fig. 1.2D-E, 1.5A); this conformation is similar to those of  $Mg$ -ATP and  $Mg$ -GTP in water in the presence of large monovalent cations such as  $K^+$  or  $NH_4^+$  [102]. The configuration of the catalytic site is enforced by a plethora of conserved bonds that mostly involve the backbone amino group (hereafter **HN** groups) of the Walker A motif, see Fig. 1.2, 1.5A, B and [34,113,114]. The partial positive charges of these groups compensate for negative charges of the triphosphate chain oxygens. The structural elements responsible for stabilization of the triphosphate chain by amino acids of the Walker A motif are almost universally conserved (Fig. 1.5). The conservation of invariant residues in the  $[G/A]XXXXGK[T/S]$  motif has straightforward reasons: the  $Gly^{K-1}$  and

Gly<sup>K-6</sup> residues mark the beginning and end of the P-loop and enable the bending of the backbone; Gly<sup>K-1</sup>, in addition, electrostatically stabilizes the O<sup>3A</sup> atom of  $\alpha$ -phosphate. The Lys<sup>WA</sup> residue interacts with O<sup>1B</sup> and O<sup>2G</sup> atoms and additionally appears to stabilize the P-loop by interacting with backbone carbonyl oxygens of K-5 and K-6 residues (Fig. 1.5).

The side chain of the conserved [Ser/Thr]<sup>K+1</sup> serves as the ligand #4 of Mg<sup>2+</sup>. As shown in Fig. 1.2B, other strictly conserved ligands are O<sup>2B</sup> and O<sup>1G</sup> atoms of the triphosphate chain as ligands #1 and #2, respectively, and water molecules as ligands #5 (W5) and #6 (W6). The position #3 is taken either by water W3 or by diverse amino acid residues in different enzyme families, see [50] for details. The positions of Mg<sup>2+</sup> ligands are similar over the entire superfamily of P-loop NTPases.

Fig. 1.5A, B show that the O<sup>2A</sup> and O<sup>3G</sup> atoms of the triphosphate are "free", not bound to P-loop residues. Not surprisingly, in many TS-like structures, they interact with stimulators (Fig. 1.2D-E). For the MnmE GTPase, molecular dynamics (MD) simulations showed that the insertion of a K<sup>+</sup> ion and its simultaneous interaction with O<sup>2A</sup>, O<sup>3B</sup> and O<sup>3G</sup> atoms rotates  $\gamma$ -phosphate and leads to formation of a new H-bond between the O<sup>2G</sup> atom and HN of Asn226<sup>K-3</sup>, see [102]. Notably, in the inspected representative structures from Table B.1, the distances between HN<sup>K-3</sup> and the closest oxygen atom of  $\gamma$ -phosphate (or its structural analog) were shorter in the case of TS analogs (see Section 3.3.1 of Results for quantitative analysis).

### 1.2.3.2. Catalytically relevant amino acids, activating partners, stimulatory patterns, and coordination of W<sub>cat</sub>

Other structural elements of P-loop NTPases, as well as their activating partners and stimulating moieties, vary among enzyme families; each class of P-loop NTPases is characterized by its specific constellation(s) of stimulators and W<sub>cat</sub>-coordinating residues. Since no detailed, comparative survey of catalysis-relevant structural attributes in all the multitude of P-loop NTPases is available, we provide here such a survey for the major classes of P-loop NTPases. Here we consider only those families and classes of P-loop NTPases for which the available structures allowed us to obtain unambiguous structural information. The results of this analysis are summarized in the Table B.1 in Appendix B. Those classes of P-loop NTPases for which the structural information is still ambiguous we have considered elsewhere [109].

#### Kinase-GTPase division

The Kinase-GTPase division unites three classes of NTPases: the TRAFAC (from *translation factors*) class of translational factors and regulatory NTPases, SIMIBI (*signal recognition, MinD, and BioD*) class of regulatory dimerizing ATPases and GTPases, and the class of nucleotide kinases (Fig. A.2 in Appendix A).

In the NTPases of the **TRAFAC class**, the  $\alpha_1$ -helix is followed by an elongated Switch I loop that is specific to the class. The elongated Switch I loop goes into the  $\beta$ -strand, which is antiparallel to all other  $\beta$ -strands of the core  $\beta$ -pleated sheet (Fig. 1.2B,E). Apart from the TRAFAC class, other classes of P-loop NTPases have predominantly all-parallel core  $\beta$ -pleated sheets.

The Switch I loop contains the only strictly conserved [Thr/Ser]<sup>SwI</sup> reference residue with its side chain coordinating Mg<sup>2+</sup> as the ligand #3 (Fig. 1.2E and 1.6). In most TRAFAC NTPases, the **HN** group of [Thr/Ser]<sup>SwI</sup> forms a H-bond with  $\gamma$ -phosphate (Fig. 1.2E, 1.6). The backbone carbonyl group (hereafter **CO** group) of [Thr/Ser]<sup>SwI</sup> interacts with the  $W_{\text{cat}}$  molecule seen in TS-like crystal structures (Fig. 1.2E, 1.6).

The TRAFAC class NTPases show a remarkably broad variety of stimulatory interactions; they are shown in Fig. 1.2E, 1.6, listed in Table B.1 and described below.

**Monovalent cation-dependent NTPases**, which have been previously examined in our research group [102], are usually stimulated by K<sup>+</sup> ions, as the aforementioned GTPase MnmE [61,66]. The K<sup>+</sup> ion in MnmE GTPase is coordinated by O<sup>2A</sup>, O<sup>3B</sup>, and O<sup>3G</sup> atoms of the triphosphate chain, two **CO** groups of the K-loop, and the side chain of Asn<sup>K-3</sup>, see Fig. 1.2E and [61,102]. In the unique eukaryotic protein family of dynamins, the NTP hydrolysis can be stimulated by either K<sup>+</sup> or Na<sup>+</sup> ions. Here a Na<sup>+</sup> or a K<sup>+</sup> ion interacts only with the O<sup>3B</sup> and O<sup>3G</sup> atoms but does not reach the O<sup>2A</sup> atom, see Fig. 1.6A and [66,102] for details. These K<sup>+</sup>/Na<sup>+</sup>-stimulated TRAFAC class NTPases belong to the family of HAS (hydrophobic amino acid substitution) NTPases [115]. No proton-accepting side chains are present in the vicinity of  $W_{\text{cat}}$ ; as indicated in Table B.1,  $W_{\text{cat}}$  interacts only with the nearby atoms of the protein backbone, e.g. **CO**<sup>SwI</sup>, **HN**<sup>SwI</sup>, and **HN**<sup>D+2</sup> in the case of the MnmE GTPase. In dynamins (Fig. 1.6A), these are the side chain of Gln<sup>K-4</sup>, **CO**<sup>SwI</sup> and **HN**<sup>D+3</sup>.

The family of **translation factors**, after which the whole TRAFAC class is denominated, includes ribosome-dependent GTPases directly involved in the translation, such as elongation factors EF-Tu and EF-G. These GTPases display K<sup>+</sup>-dependence both in the absence of the ribosome [116,117] and under conditions of protein synthesis [118–121]. The earlier suggestion from our group that these translational GTPases, alike K<sup>+</sup>-dependent GTPases, are stimulated by a K<sup>+</sup>-ion bound between the side chain of Asp<sup>K-3</sup> and **CO** of Gly<sup>T-2</sup> [122], was recently validated by crystal structures with monovalent cations in the predicted position [123]. In the translational GTPases,  $W_{\text{cat}}$  is uniquely coordinated by the side chain of His<sup>D+4</sup> residue in addition to **HN** of Gly<sup>D+3</sup> and **CO** of Thr<sup>SwI</sup>; the side chain of His turns towards  $W_{\text{cat}}$  in response to the activating interaction of the WB-crest with the small ribosomal subunit and tRNA [124].

In **GB1/RHD3-type GTPases** (e.g. atlastins) the stimulatory Arg finger is in the K-3 position of the P-loop and links the O<sup>2A</sup> and O<sup>3G</sup> atoms of GTP when two protein

monomers dimerize in response to the interaction with the activating partner (PDB ID 4IDQ [125]). In this case, Arg fingers stimulate GTP hydrolysis in the very same P-loop domain they belong to. In the case of atlastins,  $W_{\text{cat}}$  is seen stabilized by  $\text{CO}^{\text{SwI}}$  and  $\text{HN}^{\text{D}+3}$  (Table B.1, Fig. 1.6B).

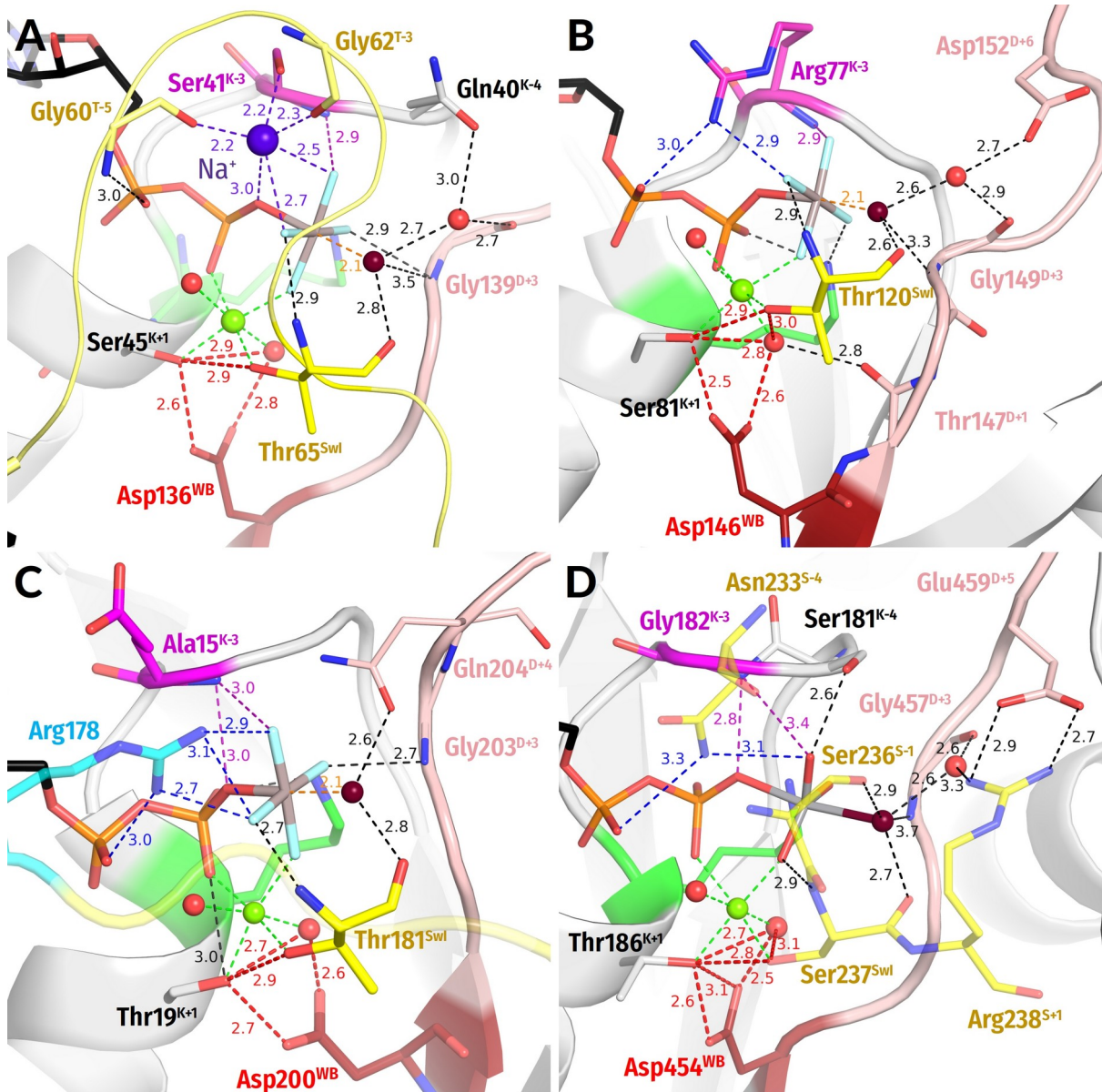


Figure 1.6: Representative NTPases of the TRAFAC class.

The residues of Switch I/K-loop are shown in yellow, nucleotides, their analogs, and functionally relevant residues are shown as sticks, water molecules and cations are shown as spheres: water in red,  $\text{Mg}^{2+}$  in lime,  $\text{Na}^+$  in blue. Other colors as in Fig. 1.2E. All distances are given in ångströms. A, Dynamin (PDB ID 2X2E, [369]); B, Atlastin-1 (PDB ID 6B9F, [370]); C,  $G\alpha_{12}$  subunit of a heterotrimeric G-protein (PDB ID 2ODE, [126]); D, Myosin II, (PDB ID 1VOM [215]).

In the GTPase domains of  $\alpha$ -subunits of heterotrimeric G-proteins, the intrinsic Arg finger, as provided by a family-specific insertion domain, links  $\text{O}^{2\text{A}}$  and  $\text{O}^{3\text{G}}$  atoms of GTP and is on a H-bond compatible distance from the  $\text{O}^{3\text{B}}$  atom, see Fig. 1.6C and

[63,126]. The  $W_{\text{cat}}$  is stabilized by  $\text{Gln}^{\text{D}+4}$  and  $\text{HN}^{\text{D}+3}$  of the WB-crest, as well as by  $\text{CO}^{\text{SwI}}$ , see Fig. 1.6C, Table B.1 and [63,126].

The family of *Ras-like GTPases* named after its oncogenic members (from *rat sarcoma*), is one of the best-studied groups in the TRAFAC class [34,58,69,127]. In these proteins, the Switch I loop interacts with diverse physiological modulators of activity whereas the specific GTPase activating proteins (GAPs) bind to the WB-crest (also called Switch II in these proteins), see Fig. 1.4C, and [34,69,128,129]. As in  $\alpha$ -subunits of heterotrimeric G-proteins shown in Fig. 1.6C, the NH2 group of the stimulatory Arg finger links the  $\text{O}^{2\text{A}}$  and  $\text{O}^{3\text{G}}$  atoms (Fig. 1.4C). The  $W_{\text{cat}}$  molecule is stabilized by the side chain of  $\text{Gln}^{\text{D}+4}$ ,  $\text{HN}$  of  $\text{Gly}^{\text{D}+3}$ , and  $\text{CO}^{\text{SwI}}$ , see Table B.1 and [34,64,128].

In P-loop NTPases of the *kinesin and myosin families* the conserved residue of Switch I is a Ser residue ( $\text{Ser}^{\text{SwI}}$ ). In these proteins, the  $\text{Asn}^{\text{S}+4}$  residue inserts between  $\alpha$ - and  $\gamma$ -phosphates [130–132], so that the side chain amino group of  $\text{Asn}^{\text{S}+4}$  links the  $\text{O}^{2\text{A}}$  and  $\text{O}^{3\text{G}}$  atoms, see Fig. 1.6D and [133]. Additional coordination of the  $\gamma$ -phosphate appears to be provided by the side chain of  $[\text{Ser/Thr}]^{\text{K}+4}$  of the P-loop and  $\text{Ser}^{\text{S}+1}$  of Switch I. The  $\text{CO}^{\text{SwI}}$  group and, likely, the side chain of  $\text{Ser}^{\text{S}+1}$  form H-bonds with  $W_{\text{cat}}$ . One more H-bond with  $W_{\text{cat}}$  is provided by  $\text{HN}$  of  $\text{Gly}^{\text{D}+3}$  (Table B.1, Fig. 1.6D).

**The SIMIBI class NTPases** include ATPases and GTPases that dimerize upon interaction with the activating partner in such a way that catalytic sites of the monomers interact “face to face”, see [134,135] for reviews. Each monomer inserts either a Lys (Fig. 1.7A) or an Arg residue (Fig. 1.7B) into the catalytic site of the other monomer [134,135]. Many SIMIBI class ATPases and GTPases contain the so-called “deviant” Walker A motif  $\text{KGGxGK[S/T]}$  with an additional conserved  $\text{Lys}^{\text{K}+5}$  residue that is inserted into the catalytic site of the partner subunit in a dimer (see Fig. 1.7A and [135]). Lysine fingers, as used by SIMIBI proteins, form H-bonds with both  $\text{O}^{2\text{A}}$  and  $\text{O}^{3\text{G}}$  atoms (Fig. 1.7A).

Signal recognition particles (*SRPs*) and their cognate receptors (*SRs*) stand separately within the SIMIBI class. Their GTPase domains form a pseudo-homodimer, where the two GTP binding sites interact face-to-face, but GTP hydrolysis occurs in only one of the two monomers. These proteins employ Arg residues that are inserted reciprocally so that the guanidinium group interacts with the  $\alpha$ - and  $\gamma$ -phosphate of the GTP molecule bound by “its” subunit and the  $\alpha$ -phosphate of the GTP molecule bound by the other subunit (Fig. 1.7B). Many SIMIBI proteins have a  $\text{Gly}^{\text{D}+3}$  residue which appears to provide its  $\text{HN}^{\text{D}+3}$  for coordination of the  $\gamma$ -phosphate, similarly to  $\text{Gly}^{\text{D}+3}$  in TRAFAC class proteins, see Fig. 1.7A-B and [32]. Additional coordination of  $\gamma$ -phosphate is provided by amino acids located outside of the conserved motifs. Such residues are protein family-specific and are often introduced into the catalytic site from the adjacent monomer upon the interaction with the activating partner and dimerization (Fig. 1.7A-B). The  $W_{\text{cat}}$  molecule in SIMIBI NTPases is routinely stabilized by  $\text{HN}$  of  $\text{Gly}^{\text{D}+3}$  and the Asp/Glu residue at the C-cap of the WB+1 strand; the carboxy group of this “catalytic”

residue links  $W_{\text{cat}}$  with the  $\text{Mg}^{2+}$ -coordinating  $W3$  molecule (Fig. 1.7A-B). Also, the residues of the other monomer and even the ribose 3' hydroxyl group of the "other" GTP molecule can contribute to the coordination of  $W_{\text{cat}}$ , see Table B.1, Fig. 1.7A-B. The NTP-bound SIMIBI dimer is believed to require the interaction with an activating protein (or RNA in SRP/SR complexes) to bring these residue(s) closer to the catalytic site where they can contribute to the H-bond network around  $W_{\text{cat}}$  [134,136].

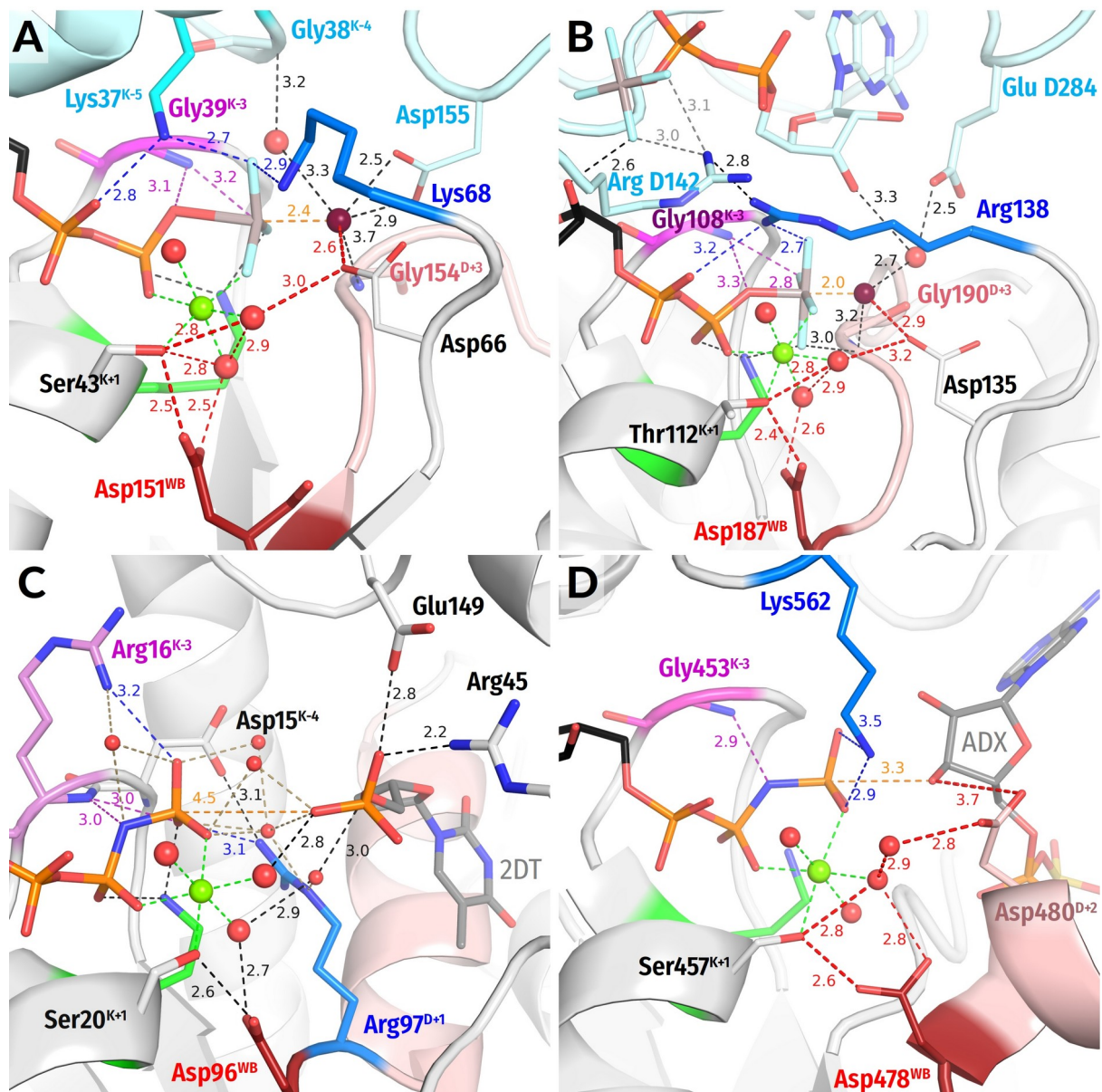


Figure 1.7: Representatives of the SIMIBI and kinase classes.

The residues of the adjacent monomers in dimers are shown in cyan. Arg and Lys residues belonging to the same protein chain as Walker A motif, i.e. to the LID domains in kinases are shown in blue. Red dashed lines mark protonic connections between  $W_{\text{cat}}$  and  $\text{Asp}^{\text{WB}}$ . Other colors as in Fig. 1.6. All distances are indicated in ångströms. A. ATP-binding component of the dark-operative protochlorophyllide reductase (PDB ID 2YNM, [166]). B. Signal recognition particle (FtsY/Efh) complex (PDB ID 2CNW, [372]), C. Human thymidylate kinase (PDB ID 1NN5 [373]). D. Adenosine 5'-phosphosulfate kinase (PDB ID 4BZX, [252]).



**Nucleotide Kinases** are ubiquitous enzymes that usually transfer a  $\gamma$ -phosphoryl residue from ATP to a wide range of “second” substrates, primarily small molecules [40,137]. The key roles in the catalysis by P-loop kinases are usually played by Arg or Lys residue(s) located in the so-called “LID” domain, a small helical segment that “covers” the catalytic site and often carries several positively charged residues [40,137,138], see Fig. 1.7C-D. In thymidylate kinases, these fingers are assisted by Arg<sup>K-3</sup>, similarly to atlastines, cf Fig. 1.6B and 1.7C. Thus, P-loop kinases do not need other proteins or domains to provide stimulatory Arg/Lys fingers. Instead, binding of their second substrate is enough to trigger the LID domain rearrangement that results in the insertion of Arg/Lys finger(s). While Arg residues serve as stimulatory moieties in most kinases (Fig. 1.7C), a Lys residue appears to be involved in adenylylsulfate kinases (Fig. 1.7D) [139].

One of the stimulatory fingers usually inserts between the  $\alpha$ - and  $\gamma$ -phosphates. In the case of the adenosine-5'-phosphosulfate (APS) kinase, the LID domain inserts a Lys finger that interacts directly with the  $\gamma$ -phosphate and is connected to  $\alpha$ -phosphate via a water molecule (Fig. 1.7D). As in many other P-loop NTPases, the catalysis in kinases is usually assisted by auxiliary arginine and lysine fingers that position the interacting molecules and neutralize the negative charges of phosphate groups, see Fig. 1.7C-D, Table B.1 and [74,140]. Some families of kinases use also an Arg residue of the WB-crest, see Fig. 1.7C and [40].

In kinases, the hydrolysis of the phosphoryl donor molecule is mediated by the acceptor molecule, which, similarly to  $W_{\text{cat}}$  in other reactions of NTP hydrolysis, appears to initiate the nucleophilic attack and formation of the pentavalent intermediate. In nucleotide monophosphate kinases, the attacking group is a negatively charged phosphate moiety requiring no specific proton acceptor (Fig. 1.7C). In those kinase families, where the deprotonation of the attacking molecule is needed to produce a nucleophile, the [Asp/Glu]<sup>D+2</sup> residue appears to serve as an immediate proton acceptor, see Fig. 1.7D and Table B.1. This residue links the would-be nucleophilic group with the  $\text{Mg}^{2+}$  ligand in position 3, usually a water molecule (Fig. 1.7D).

#### ASCE division

The NTPases of the ASCE division have all-parallel  $\beta$ -pleated sheets with an inserted, as compared to the sequences of the Kinase-GTPase division NTPases,  $\beta$ -strand in the WB+1 position, see Fig. 1.2A,C, A.2. In many enzyme classes of this division, further additional  $\beta$ -strands were identified, see Fig. A.2 and [36,37,112]. For coordination of  $W_{\text{cat}}$  in addition to class-specific residues, ASCE NTPases typically use a “catalytic” glutamate residue that either directly follows the conserved Asp<sup>WB</sup> or is located at the C-cap of the WB+1 strand, similarly to SIMIBI NTPases. These two features define the name of this division: *additional strand, catalytic E* (ASCE) [32,33,111,141]. According to

current views [36,37,112] and the most recent phylogenetic scheme depicted in Fig. A.2, the ASCE NTPases are divided into two clades that differ by the number of  $\beta$ -strands in their P-loop domains. The group of “middle-size” domains with up to five-six  $\beta$ -strands includes AAA+ ATPases, helicases of superfamily 3 (SF3), as well as STAND and KAP ATPases. The clade of “large” ATPases domains, with many  $\beta$ -strands, includes helicases of superfamilies 1 and 2 (SF1/2), ABC ATPases, RecA/F<sub>1</sub> ATPases, VirD/PilT-like ATPases, and FtsK-HerA-like ATPases.

**AAA+ ATPases** are *ATPases associated with various cellular activities*, where “+” stands for “extended”, see Fig. 1.8A and A.2. These enzymes contain an N-terminal P-loop domain and an additional  $\alpha$ -helical C-terminal domain, see [30,65,67,111,142–144] for comprehensive reviews. The P-loop domain of the AAA+ ATPases carries conserved Arg/Lys residue(s) from the side that is opposite to the P-loop. The P-loop domains interact upon oligomerization (most often a hexamer is formed), so that the nucleotide-binding site of one subunit receives the Arg/Lys finger(s) from the neighboring subunit (Fig. 1.8A-B) and/or, in some protein families, an additional Arg/Lys residue from its own C-terminal helical domain (Fig. 1.8A), see [67]. One of the stimulatory residues interacts with  $\gamma$ -phosphate and is called “finger” whereas the other one occupies the space between  $\alpha$ - and  $\gamma$ -phosphates and is called “sensor 2”. We could not find structures of AAA+ ATPases with bound TS analogs; based on available data on site-specific mutants [67],  $W_{\text{cat}}$  appears to be coordinated by the “catalytic” [Glu/Asp]<sup>D+1</sup> residue, Arg/Lys finger, and, perhaps, Asn/Ser/Thr residue at the C-cap of the WB-1 strand (“sensor 1”, see Fig. 1.8A). As shown in Fig. 1.8A, Glu<sup>D+1</sup> connects a would-be  $W_{\text{cat}}$  with Mg<sup>2+</sup>-coordinating W3 in the structure of N-ethylmaleimide sensitive factor (PDB ID 1NSF [145]). AAA+ NTPases are rather diverse, they can use further polar residues to interact with  $W_{\text{cat}}$  or  $\gamma$ -phosphate. Also, the topology of domains that interact with the P-loop domain can vary. For instance, the helicases of superfamily 3 (SF3 helicases) have different topology of their C-terminal helical domain [142]. Their other specific feature is the presence of a Glu<sup>WB</sup>–Asp<sup>E+1</sup> pair at the C-cap of the WB strand; the Asp<sup>E+1</sup> residue builds a link to the Mg<sup>2+</sup>-coordinating W3 molecule (Fig. 1.8B).

**SF1/2 class helicases** are mostly monomeric or dimeric with each polypeptide chain containing two P-loop fold domains, see [28,146] for reviews. While SF1 and SF2 helicases differ in the structural motifs that couple their DNA/RNA binding sites with ATP-hydrolyzing catalytic pockets, the pockets proper are quite similar (cf Fig. 1.2A,D and 1.8C, respectively). Both in SF1 and SF2 helicases, the ATP molecule binds to the functional Walker A and B motifs of the N-terminal domain; the C-terminal domain, although it has a P-loop-like fold, lacks the Walker A and B motifs. Following the interaction with an RNA or a DNA molecule, two Arg residues of the C-terminal domain are usually inserted into the ATP-binding site [147]. One of these Arg residues forms H-bonds with both  $\alpha$ - and  $\gamma$ -phosphates, whereas the other Arg residue interacts with

$\gamma$ -phosphate (or its analog) and  $W_{\text{cat}}$  (Fig. 1.2D, 1.8C). The  $W_{\text{cat}}$  molecule is also coordinated by the “catalytic”  $\text{Glu}^{\text{D}+1}$  residue and class-specific Gln and Arg residues (Fig. 1.2D, 1.8C). The same  $\text{Glu}^{\text{D}+1}$  residue links  $W_{\text{cat}}$  with the  $\text{Mg}^{2+}$ -coordinating  $W3$  molecule (Fig. 1.2D, 1.8C, Table B.1).

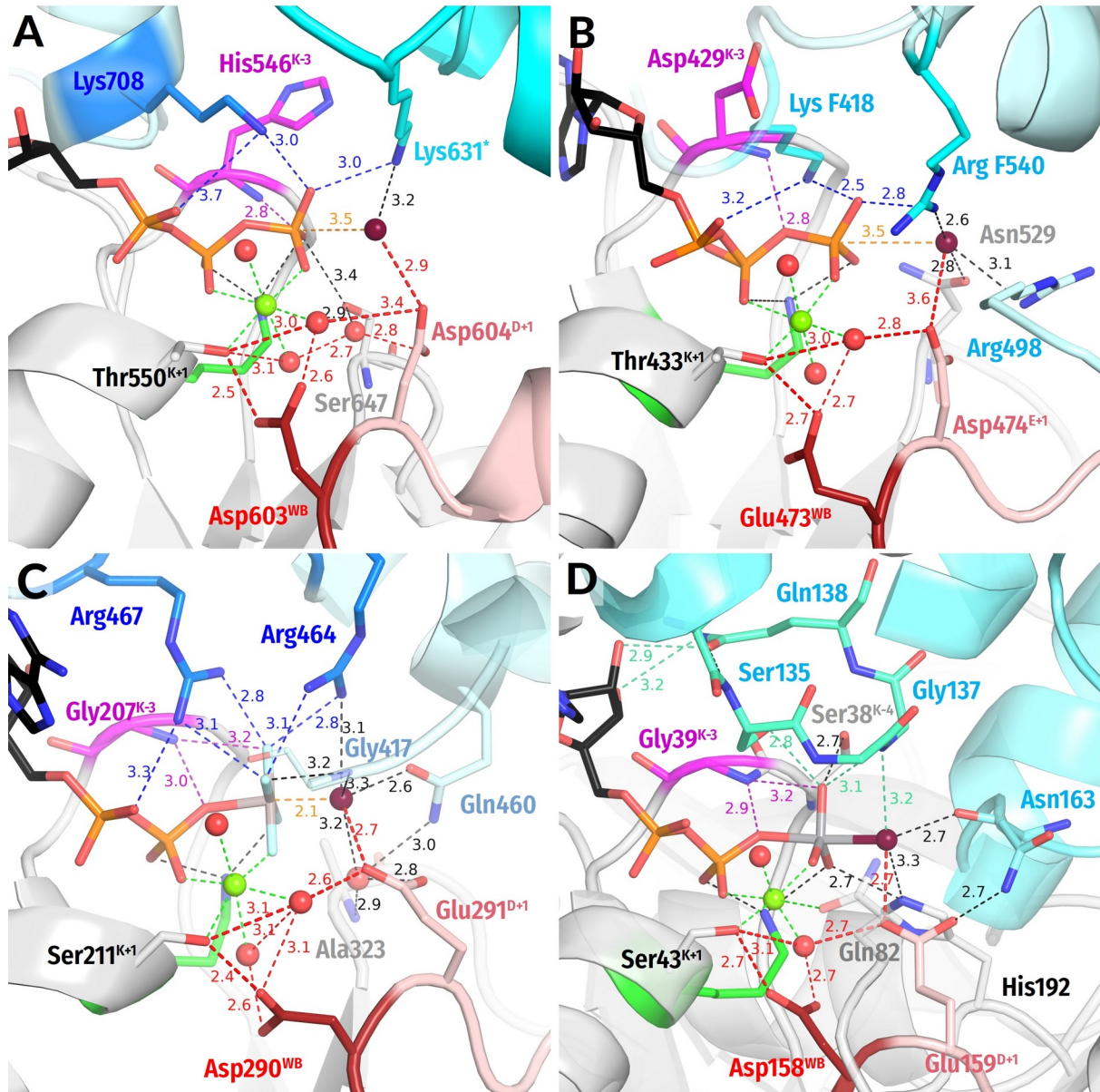


Figure 1.8: Representatives of AAA+ ATPases, SF3 helicases, SF2 helicases and ABC ATPases. The adjacent monomers and their Arg and Lys residues are shown in cyan, residues of adjacent monomers labelled in light blue. The Arg residues from the C-terminal helical domain of the same monomer are shown in deep blue (panel C). Red dashed lines mark protonic connections between  $W_{\text{cat}}$  and  $\text{Asp}^{\text{WB}}$ . Other colors as in Fig. 1.6. All distances are given in ångströms. **A**, N-ethylmaleimide sensitive factor (PDB ID 1NSF [145]); **B**, Replicative hexameric helicase of SV40 large tumor antigen (PDB ID 1SVM [199]); **C**, Hepatitis C virus NS3 SF2-helicase (PDB 5E4F [193]); **D**, Outward-facing maltose ABC transporter complex with  $\text{ADP}:\text{VO}_4^-$  bound (PDB ID 3PUV [150]). The LSGGQ motif is shown in green-cyan, the  $\text{VO}_4^-$  moiety by grey and red.

**ABC (ATP-binding cassette) ATPases** are multidomain proteins that usually operate as homo- or heterodimers, see Fig. 1.8D, Fig. A.2 and [148,149]. Members of the ABC

class make several families named alphabetically from A to I [148]. Most of these families contain members that possess transmembrane domains and operate as genuine ATP-driven membrane transporters where the P-loop domains hydrolyze ATP [150,151]. However, the members of ABCE and ABCF families have no transmembrane domain(s) [152].

Crystal structures with TS analogs bound were obtained for the maltose transporter complex, see Fig. 1.8D and [150]. More recently, structures with bound TS analogs were solved by cryo-EM microscopy for heterodimeric ABC exporter TmrAB of the *Thermus thermophilus*, which consists of multi-drug-resistance proteins A and B [151].

In dimers of ABC ATPases, the nucleotide-binding sites of P-loop domains are located on the interface between the monomers, in the same way, as in dimers of SIMIBI NTPases, cf. Fig. 1.8D with Fig. 1.7A-B. Instead of an Arg or Lys residue, each monomer inserts a whole signature motif **LSGGQ** into the catalytic pocket of the other monomer (Fig. 1.8D, [150,151]).

Several amino acids commonly found in the catalytic sites of ABC transporters can stabilize  $W_{\text{cat}}$ , see Fig. 1.8D and Table B.1. In the case of maltose transporter and the TmrAB exporter, these are histidine residues at the C-cap of the WB-1 strand and  $\text{Glu}^{\text{D}+1}$  [150,151]; the latter, in addition, links  $W_{\text{cat}}$  with the  $\text{Mg}^{2+}$ -coordinating  $W_6$  molecule. In both enzyme complexes, the activating monomer in a dimer contributes to the coordination of  $W_{\text{cat}}$  by providing a backbone **CO** group of a residue that is located outside of the signature motif, see Fig. 1.8D and [150,151].

**RecA/F<sub>1</sub> NTPases** class encompasses oligomeric ATP-dependent motors involved in homologous recombination and DNA repair (RecA and RadA/Rad51), catalytic subunits of rotary F/N- and A/V-type ATP synthases, helicases of superfamilies 4 and 5 (SF4 and SF5), as well as several other protein families, see Fig. A.2 and [28,33,153].

In NTPases of this class, similarly to most AAA+ ATPases, the stimulatory moiety(is) is/are inserted in the catalytic site by the P-loop domain of the adjacent monomer (Fig. 1.9). Our structural analysis, however, showed that the stimulation mechanism appears to differ between rotary F<sub>1</sub> ATPases/SF5 helicases, on the one hand, and recombinases/SF4 helicases, on the other hand [109].

In the **F<sub>1</sub> ATPases**, the stimulatory Arg residue of the adjacent monomer interacts with both  $\alpha$ - and  $\gamma$ -phosphates, whereas an additional, intrinsic Arg residue coordinates  $\gamma$ -phosphate, approaching it apically (see Fig. 1.9A, and Table B.1). A similar stimulation mechanism is realized in **SF5 helicases**, see e.g. the Rho helicase (PDB ID 6DUQ [154]).

The  $W_{\text{cat}}$  molecule is seen in the catalytic apical position only in the ADP:AlF<sub>4</sub><sup>-</sup>-containing structure of the bovine F<sub>1</sub>-ATPase (PDB ID 1H8E, Fig. 1.9A) [72]. While in most ASCE ATPases the  $W_{\text{cat}}$ -coordinating  $\text{Glu}^{\text{D}+1}$  residue directly follows the Asp<sup>WB</sup> of the Walker B motif, the  $W_{\text{cat}}$ -coordinating Glu188 of the F<sub>1</sub>-ATPase, similarly to SIMIBI NTPases, is at the C-cap of the WB+1 strand; within the ASCE division this feature is

specific for the RecA/F<sub>1</sub> class members [31]. This glutamate residue also links W<sub>cat</sub> with the Mg<sup>2+</sup>-coordinating W3 molecule (Fig. 1.9A). In F<sub>1</sub>-ATPase, W<sub>cat</sub> is also stabilized by Arg260<sup>D+4</sup>; this residue concurrently interacts with the activating neighboring monomer, which also provides a CO group to stabilize W<sub>cat</sub> (Fig. 1.9A).

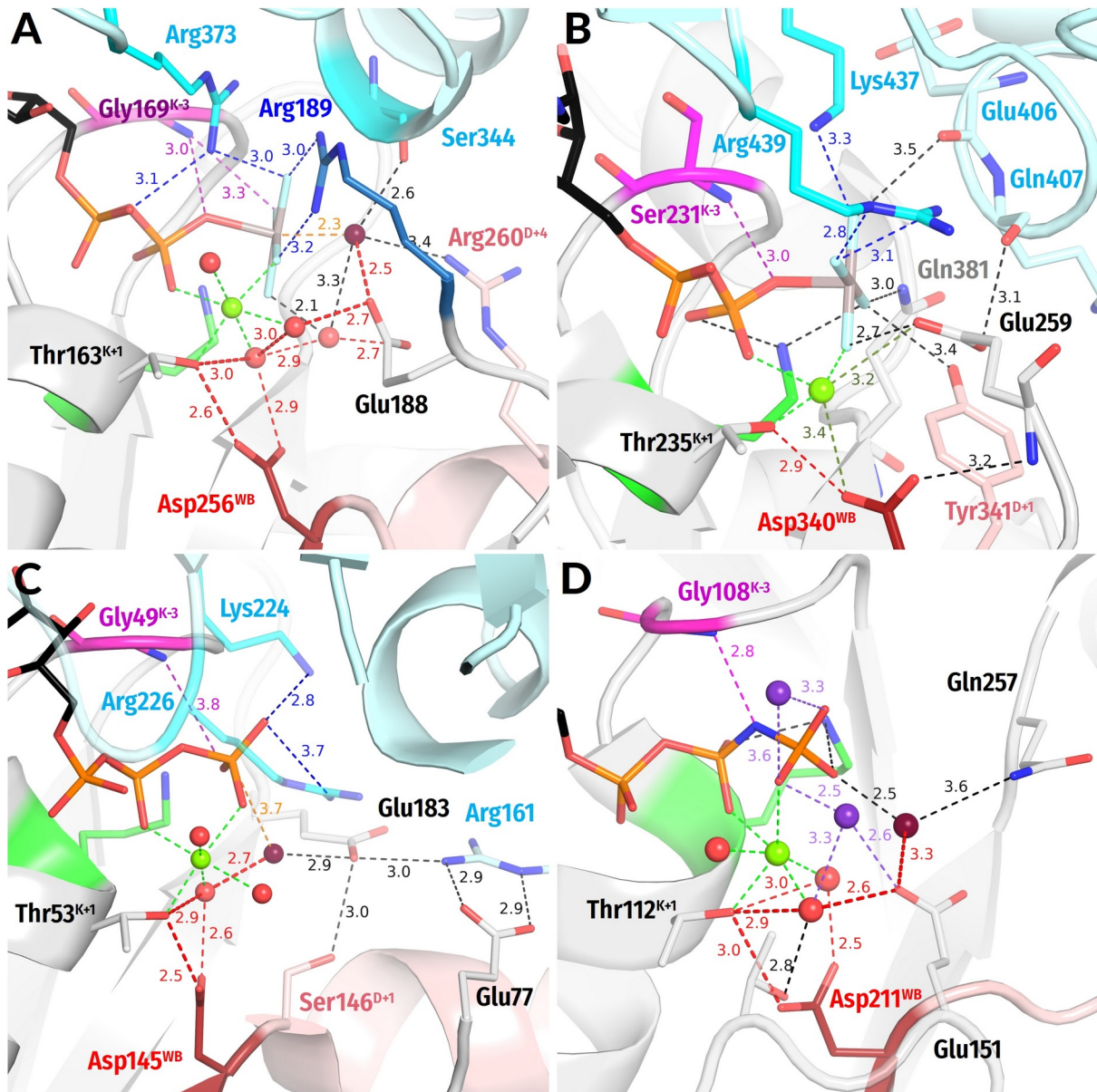


Figure 1.9: Representative proteins of the RecA/F<sub>1</sub>-like class of the P-loop NTPases.

The adjacent monomers and their Arg and Lys residues are shown in cyan, K<sup>+</sup> ions are shown as purple spheres. The Arg residue from the P-loop domain proper is shown in deep blue (panel A). Red dashed lines mark protonic connections between W<sub>cat</sub> and Asp<sup>WB</sup>. Other colors as in Fig. 1.6.

All distances are in Ångströms. **A**, Bovine F<sub>1</sub>-ATPase (PDB ID 1H8E [72]); **B**, DnaB replicative helicase from *Vibrio cholerae* (PDB ID 6T66 [155]); **C**, circadian clock protein KaiC (PDB ID 4TL7 [156]); **D**, Rada recombinase (PDB ID 3EW9 [157]).

In *RecA*-like recombinases and *SF4* helicases, two stimulatory positively charged moieties interact only with the  $\gamma$ -phosphate group, in contrast to F<sub>1</sub>-ATPases. In several families, the adjacent subunit provides one Lys residue and one Arg residue that form a short KxR motif [158], with both residues reaching only  $\gamma$ -phosphate. These are, for

instance, the bacterial helicase DnaB (Fig. 1.9B, see also [159]), circadian clock protein KaiC (Fig. 1.9C [156]), and gp4d helicase from the T7 bacteriophage (PDB ID 1E0J, [146]). In bacterial RecA recombinases, the adjacent monomer in the homooligomer provides two Lys residues which approach the phosphate chain laterally and interact with  $\gamma$ -phosphate as in the RNA recombinase of *E. coli* (see Table B.1 and PDB ID 3CMX, [160]). In archaeal and eukaryotic RadA/Rad51-like recombinases, the positions of terminal groups of stimulatory Lys/Arg residues are occupied by two  $K^+$  ions, which appear to interact with  $\gamma$ -phosphate, see Fig. 1.9D and [158,161].

In the case of RecA-like ATPases, the  $AlF_4^-$ -containing structures of the bacterial DnaB helicase of *Vibrio cholerae* (Fig. 1.9B), *Bacillus stearothermophilus* (PDB ID 4ESV, [162]) and RecA recombinase of *E. coli* (PDB ID 3CMX, [160]) are available. However, the apical water molecule is absent from these structures, the possible reasons of which are discussed in the Results and Discussion sections in relation to the *Bacillus stearothermophilus* structures. Therefore, the full set of  $W_{cat}$ -coordinating groups remains unknown for RecA-like ATPases; the available structures imply the participation of glutamate residues at the C-terminus of the WB+1 strand (Fig. 1.9B-D), as in other RecA/ $F_1$  class enzymes. The same residue appears to link the would-be  $W_{cat}$  with the  $Mg^{2+}$ -coordinating W3 molecule (Fig. 1.9B-D). Generally, the H-bond networks around  $\gamma$ -phosphate and  $W_{cat}$  seem to be richer in RecA/ $F_1$ -NTPases than in other classes of P-loop NTPases. Specifically, the coordination of  $W_{cat}$  may potentially also involve the D+1 residue of the WB strand, which is usually a H-bonding Ser/Thr/Asn/Tyr, as well as other polar residues at the C-caps of the WB-1 and WB+1 strands [146,156,163].

### 1.2.3.3. Asymmetric charge distribution in catalytic pockets of P-loop NTPases

In structures of P-loop NTPases that are shown in Fig. 1.5-1.9, the positively charged groups that stabilize the triphosphate chain, namely the Lys<sup>WA</sup> residue, the backbone HN groups and the stimulatory moiety(ies) are opposed by acidic residues that interact with  $W_{cat}$  either directly or via water bridges. Together with Asp<sup>WB</sup>, these catalytic acidic residues form negatively charged clusters. Fig. 1.10 shows, for diverse P-loop NTPases, these positively and negatively charged clusters, which should produce a strong local electric field.

The positive charges of amino group of Lys<sup>WA</sup>, the  $Mg^{2+}$  ion, and several HN groups of the P-loop compensate for the negative charges of phosphate oxygen atoms (see Fig. 1.5B). In addition to these common positively charged moieties, class-specific auxiliary residues could be involved, as, for instance, "sensors 3" in some AAA+ ATPases [67]. Notably, the phosphate chain "sits" on the last N-terminal turn of the  $\alpha_1$ -helix, which generally carries a dipole positive charge of about 0.5 [164,165]. The positions of the

groups involved are strictly conserved (Fig. 1.5), so that such a compensation is common to all major families of P-loop NTPases.

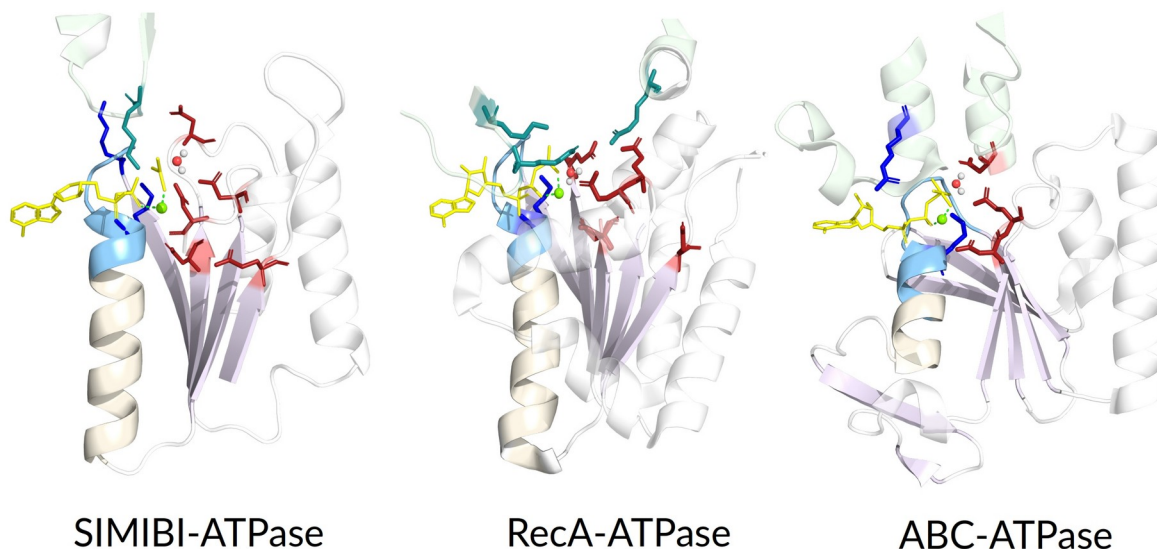


Figure 1.10: Uneven distribution of charged residues around nucleotide binding pockets in different classes of P-loop NTPases.

Nucleotide analogs are shown in yellow, negatively charged residues are shown in red, positively charged residues are shown in blue and teal, the  $\alpha_1$  helix is shown in beige with its N-terminus and the P-loop shown in light blue. **A**, SIMIBI class, light-independent protochlorophyllide reductase iron-sulfur ATP-binding protein chlL (PDB ID 2YNM, [166]); **B**, RecA class, circadian clock protein kinase KaiC (PDB ID 4TL7, [156]); **C**, ABC ATPase class, MBP-maltose transporter complex (PDB ID 3PUW, [150]).

On activation, the negative charges of the  $\gamma$ -phosphate oxygen atoms are additionally compensated by the positive charges provided by most stimulators, see Fig. 1.2,1.4-1.9.

Furthermore, Fig. 1.10 shows that the electrostatic potential at the catalytic sites of diverse P-loop NTPases is distributed unevenly, which has already been noted for particular enzymes of this family, see, e.g., [167]. The strength of the local electric field can be roughly estimated using the Coulomb's equation as  $\sim 10^8$  V/m under the very modest assumptions that (i) the resulting electric charge difference at a distance of 10 Å (between the acid residues and the P-loop) corresponds to one elementary charge and (ii) the effective dielectric permittivity of the catalytic pocket is about 10.0 [168,169]. Such an electric field strength, albeit large, is compatible with those measured in the catalytic pockets of other enzymes [170]. Hence, the catalytic pocket is strongly polarized in P-loop NTPases, which also can contribute to catalysis, as discussed in Section 4.2.

### 1.3. Aims of the current work

P-loop NTPases constitute an ancient, ubiquitous, and diverse superfamily of proteins involved in the key energy transformation reaction in the cell. The central evolutionary

and medical importance of these enzymes and the extensive set of their interacting partners prompted many research groups to solve thousands of their 3D structures. However, to our knowledge, no comparative structure analysis encompassing the whole superfamily has yet been performed. Moreover, the chain of events following the insertion of a stimulating moiety into the catalytic site of P-loop NTPases remains unclear and controversial. The reasons for the remarkable conservation of some amino acid residues of the catalytic center also remain obscure.

Therefore, the aims of the current work were:

1. To develop a methodical approach for a quantitative structural comparison of thousands of catalytic sites in evolutionary divergent proteins, as a complementary approach to the traditional visual comparison of a smaller number of structures.
2. To apply the developed approach to the available structures of P-loop NTPases in order to elucidate:
  - a. How widespread are the features previously identified in our group upon MD simulations of a  $K^+$ -stimulated GTPase MnmE [102]? The simulations revealed the twisting of  $\gamma$ -phosphate by the stimulating moiety and stabilization of  $\gamma$ -phosphate by an additional H-bond to one of the backbone **HN** groups of the P-loop; both of these changes could potentially promote hydrolysis.
  - b. How is the catalytic water molecule deprotonated upon NTP hydrolysis in P-loop NTPases, and what is the fate of the proton? We wanted to build on the expertise in our group, as gathered upon studies of proton transfer in bacterial photosynthetic reaction centers [171], cytochrome bc complex [172,173], ferredoxin I [174], and biological interfaces [175]. Our goal was to trace the proton pathways in active sites of P-loop NTPases from the available body of structural data and to identify those amino acid residues that are key to proton transfer in the P-loop NTPases.



## 2. Methods

### 2.1. Evolutionary biophysics approach

Our approach can be described as based on the logic of evolutionary biophysics, which assumes that certain structural elements are preserved by evolution because of their functional importance and that the degree of preservation of the corresponding structural elements should be considered when choosing among several possible biophysical mechanisms corresponding to the structural data, see [158,176–180]. Evolutionary biophysics brings structural or sequence commonalities in connection with catalytic mechanisms: a common mechanism would imply common executors. Such approach would benefit from larger scale comparative analysis; however, quantitative scrutinization of multiple binding sites remains rare. We were not aware of quantitative, comparative sequence-independent analyses of thousands catalytic sites in homologous, but far-diverged proteins. Therefore we developed the tools ourselves.

### 2.2. Methods developed for global structural analysis of P-loop NTPases

#### 2.2.1. Automated analysis of binding sites

It is well established that structures tend to be more conserved than amino acid sequences on longer evolutionary distances [181–183]. For instance, P-loop NTPases of long-diverged families, such as families belonging to different classes, have relatively low sequence conservation. In our approach, we, instead of relying on amino acid sequences, identify catalytic site elements based on their geometric positions relative to each other and the substrate. This approach is suitable for P-loop NTPases due to the rigidity of the P-loop and similarity of the bound triphosphate chain configuration across the whole superfamily (see Section 1.2.3.1 of Introduction). To compare binding sites across the superfamily, we have developed a Python-based routine for identifying the nature of amino acid residues in the key positions of each catalytic site and calculation of distances between relevant residues as well as between the residues and the atoms of the triphosphate chain. This routine also incorporates a set of quality estimates and filters for catalytic sites sampled. The same approach can also be generalized to other protein families given that (1) the substrates have some shared structural elements that are recognized by the enzymes under consideration and (2) some minimal information on enzymatic activity determinants and residue conservation in such a protein family is available.

### 2.2.1.1. Structure sampling

P-loop NTPases are also remarkable for the availability of structures believed to represent different stages of catalysis and complexed both with native substrate molecules and different types of their structural analogs. We selected structures among those PDB entries that matched the following criteria: (1) the entry is assigned to InterPro record IPR027417 “P-loop containing nucleoside triphosphate hydrolase”; (2) contains an ATP/GTP molecule, or non-hydrolyzable analog of NTP, a ground-state analog, or a transition-state analog; (3) contains at least one  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$  or  $\text{Ca}^{2+}$  ion. The NTP analogs sampled included adenosine 5'-[ $\beta,\gamma$ -imido]triphosphate, (AMP-PNP, “ANP” in PDB), guanosine 5'-[ $\beta,\gamma$ -imido]triphosphate (GMP-PNP, “GNP” in PDB), adenosine 5'-[ $\beta,\gamma$ -methylene]triphosphate (AMP-PCP, “ACP” in PDB), guanosine 5'-[ $\beta,\gamma$ -methylene]triphosphate (GMP-PCP, “GCP” in PDB), adenosine 5'-[ $\gamma$ -thio]triphosphate (ATP- $\gamma$ -S, “AGS” in PDB), and guanosine 5'-[ $\gamma$ -thio]triphosphate (GTP- $\gamma$ -S, “GSP” in PDB). We also sampled structures containing (an) ADP or GDP molecule(s) and metal fluoride complexes such as  $\text{BeF}_3$  (“BEF”),  $\text{AlF}_3$  (“AF3”),  $\text{AlF}_4^-$  (“ALF”) and  $\text{NDP}:\text{MgF}_3^-$  (“MGF”); as well as vanadate complexes:  $\text{VO}_4^-$  (“VO4”). As we relied on the RCSB PDB API and retrieved structures containing compounds specified by ligand names mentioned above, ATP or GTP analogs stored under less common ligand names in PDB might have evaded our consideration.

We considered X-ray structures as well as cryo-EM and NMR structures. This search yielded 1474 structures with 3666 catalytic sites in them. For the NMR structures, only the first frame containing an NTP analog bound to the P-loop region was included. For X-ray and cryo-EM structures an additional criterion was applied: structures with resolution less than 5 Å were excluded. In the case of structures containing multiple subunits or multiple copies of the same protein, each interaction between a protein and an NTP-like molecule was treated separately, as an individual complex. The pipeline that was applied to each complex is depicted in Fig.1.9.

### 2.2.1.2. Protein family mapping

Proteins were assigned to major classes of P-loop NTPases according to membership in Pfam families that was determined from PDB-to-Pfam mapping (retrieved 10.10.2020 from [ftp://ftp.ebi.ac.uk/pub/databases/Pfam/mappings/pdb\\_pfam\\_mapping.txt](ftp://ftp.ebi.ac.uk/pub/databases/Pfam/mappings/pdb_pfam_mapping.txt)). Each chain was considered separately and only the best scoring Pfam domain included in Pfam clan CL0023 (“P-loop\_NTPase”) was used for the assignment. Each nucleotide binding site had a separate Pfam mapping using the corresponding Lys<sup>WA</sup> sequence position as some proteins have multiple P-loop domains in one protein chain. Since many Pfam domains describing P-loop NTPases were described before a coherent classification of P-loop NTPases was developed [32,40,111], some domain names may not be accurate.

### 2.2.1.3. Identification of the catalytic sites and of Lys<sup>WA</sup>

Each NTP-like molecule bound in each of the structures sampled was considered separately and the coordinates of the atoms of  $\alpha$ -,  $\beta$ - and  $\gamma$ -phosphates (or  $\gamma$ -phosphate mimic) were used to search for elements of the binding site in each case.

To identify complexes with a substrate or its analog bound to the P-loop motif, we applied the following filter: an NTP(-like) molecule was considered bound to the P-loop if both  $\text{Mg}^{2+}/\text{Ca}^{2+}/\text{Mn}^{2+}$  cation and the NZ atom of Lys<sup>WA</sup> could be found within 5Å of the oxygen atoms of the  $\beta$ -phosphate. Lys<sup>WA</sup> was defined as the closest Lys found in the vicinity of  $\beta$ -phosphate that is followed in the sequence by a Ser or Thr residue. Manual check of a number of structures has shown that this strategy works well for non-distorted catalytic sites.

### 2.2.1.4. Assessment of the nucleotide integrity and $\gamma$ -phosphate state

We have checked the nucleotide integrity for each complex evaluated. Complexes with molecules annotated as non-hydrolyzable NTP analogs or native ATP/GTP molecules that were lacking  $\beta$ - and/or  $\gamma$ -phosphate or their mimicking groups were not included in the analysis. For structures annotated as containing fluoride complexes or vanadate, only ADPs and GDPs possessing a fluoride complex or vanadate ion in vicinity were analyzed further. P-loop domains containing NDPs complexed with  $\text{P}_i$  were not further used for automated analysis; we considered them as corresponding to post-catalytic state. The P-loop domains from multisubunit complexes that possessed an NDP but no  $\gamma$ -phosphate mimic were also not considered further (i.e. catalytic sites with a bare ADP molecule bound in structures of ATP synthases).

Catalytic sites not fulfilling the conditions listed above were discarded from the sample. In total, we have selected 3136 complexes in 1383 structures with various substrates: ATP and GTP, non-hydrolyzable analogs of ATP and GTP, and ADP or GDP molecules associated with  $\gamma$ -phosphate-mimicking moieties (see Fig.1.9.) for further analysis.

### 2.2.1.5. Identification of the residues of Walker A motif

The Lys<sup>WA</sup> residue were identified as described in section 2.2.1.3 above. In each catalytic site under consideration, the K-3 residue was identified by its position relative to the P-loop Lys residue. Distances (in Å) were measured from the HN of the K-3 residue to the nearest oxygen or fluorine atom of the  $\gamma$ -phosphate or its mimic. We accounted for differences in phosphate oxygen atoms numeration among studied structures. The [Ser/Thr]<sup>K+1</sup> residue was identified by sequence position relative to the Lys<sup>WA</sup>.

### 2.2.1.6. Identification of the residues of Walker B motif

Putative [Asp/Glu]<sup>WB</sup> residue was identified as follows: distances from all Asp and Glu residues to the [Ser/Thr]<sup>K+1</sup> were measured, and the closest residue that was preceded by at least three non-ionizable residues (Glu, Asp, Ser, Thr, Tyr, Lys, Arg and His were considered as ionizable) was chosen as the partner of [Ser/Thr]<sup>K+1</sup>. If this failed, a closest Asp/Glu was selected without hydrophobicity check. Residues located further than 5 Å were not considered. In those few cases where [Asp/Glu]<sup>WB</sup> residue is indicated as “absent at the threshold distance”, the catalytic site is likely to be fully “open”. We have not checked all these cases manually.

### 2.2.1.7. Assessment of the Mg<sup>2+</sup>-binding site

We have also measured distances from [Ser/Thr]<sup>K+1</sup> to Mg<sup>2+</sup>, to ensure the correct binding of the Mg<sup>2+</sup> and general reliability of the structure resolution at the binding site (i.e, very long distance would indicate a disturbed catalytic site or resolution at the site that is insufficient for purposes of comparative analysis), and from [Asp/Glu]<sup>WB</sup> to Mg<sup>2+</sup>, to identify cases of direct coordination of Mg<sup>2+</sup> by the acidic residue (short distances) or disassembled binding sites (long distances).

### 2.2.1.8. Evaluation of Mg<sup>2+</sup> binding in the AlF<sub>4</sub><sup>-</sup>-containing structures

For AlF<sub>4</sub><sup>-</sup>-containing structures, coordination of Mg<sup>2+</sup> was also evaluated in more details. Ligands of Mg<sup>2+</sup> present within 2.8 Å were listed and assigned to categories. Each site was expected to feature following Mg<sup>2+</sup> ligands: the oxygen atom of β-phosphate (#1), the fluoride atom of AlF<sub>4</sub><sup>-</sup> (#2), the side chain hydroxyl of [Ser/Thr]<sup>K+1</sup> (#4), and water molecules (#5, #6). The position #3 is taken by the side chain hydroxyl of Ser/Thr<sup>SwI</sup> in TRAFAC proteins. Some families of P-loop NTPases possess another amino acid ligand at #3 position, such as Gln in ABC ATPases. In the majority of the remaining ASCE NTPases, this position (#3) is taken by a water molecule. All sites were expected to have 6 Mg<sup>2+</sup> ligands falling into the categories above. Sites missing ligands of any of the required types and sites possessing 7 or more ligands were not used for further quantitative analysis.

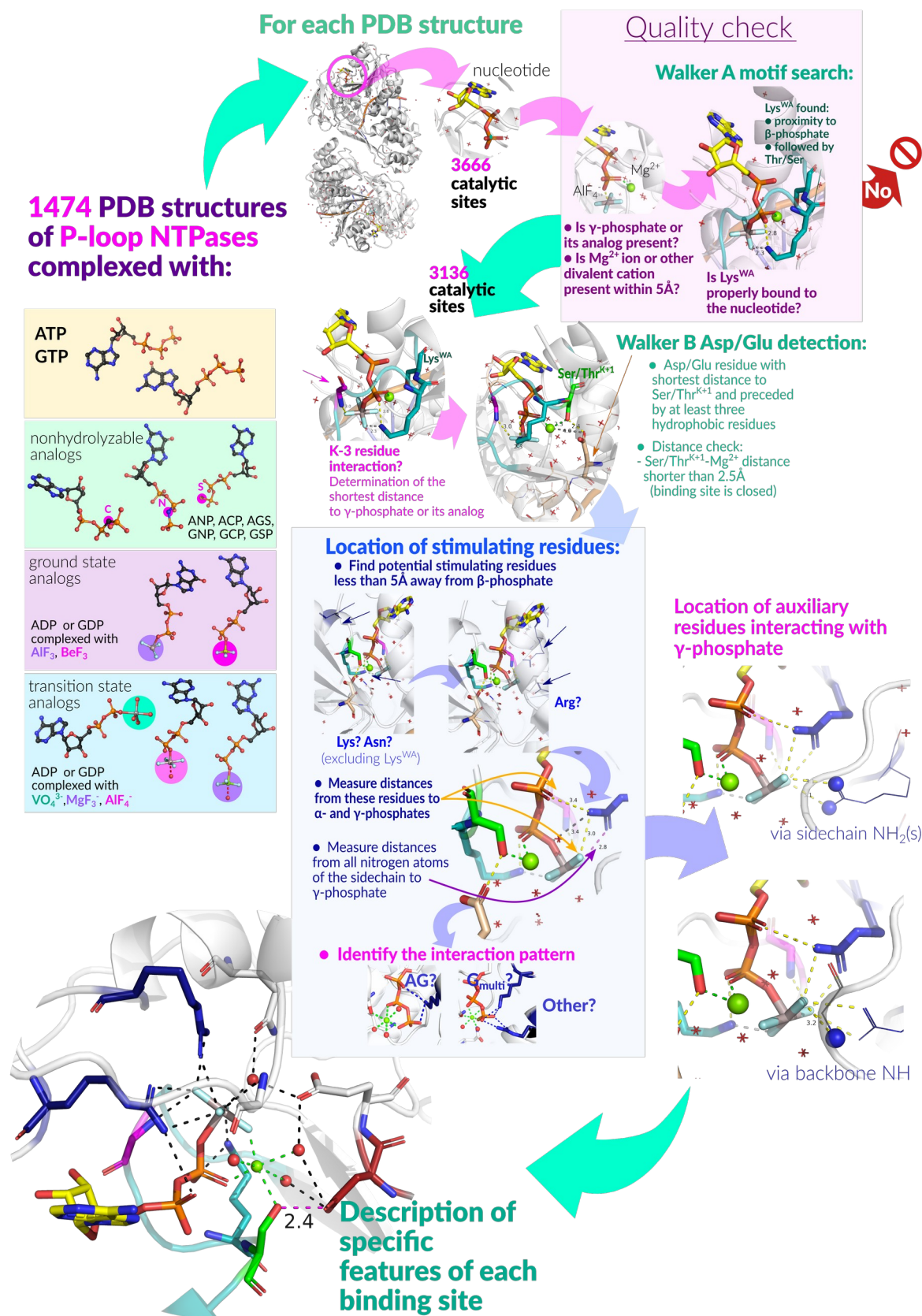


Figure 2.1: Automated comparative structural analysis of P-loop NTPases.

The examination pipeline, as applied to the structure of each selected nucleotide-binding site, is shown. We checked for (i) the presence of the  $\text{HN}^{\text{K-3}} - \text{O}^{2\text{G}}$  bond, (ii) the length of the  $[\text{Asp/Glu}]^{\text{WB}} - [\text{Ser/Thr}]^{\text{K+1}}$  H-bond and (iii) the type of interaction of stimulatory residues. Scripts used to download and analyze the structures are available at [github.com/servalli/pyplloop](https://github.com/servalli/pyplloop)

### 2.2.1.9. Identification of stimulatory moieties and residues interacting with $\gamma$ -phosphate

We had also inspected the presence of positively charged stimulatory residues near the phosphate chains in all complexes. To automatically identify such residues, we selected Arg and Lys residues (excluding the P-loop Lys) nearest to the  $\beta$ -phosphate group oxygen atoms or its structural analogs. Distances were then measured from the guanidinium group nitrogen atoms (NE, NH1, NH2) of the Arg residue or from the NZ atom of the Lys residue to the closest oxygen atom of  $\alpha$ -phosphate moiety and to the nearest fluorine or oxygen atom of  $\gamma$ -phosphate or its mimicking group. Similarly, we checked possible interactions of the phosphate chain with the ND2 atom of Asn. Possible additional interactions of  $\gamma$ -phosphate oxygens with N atoms of protein backbone and side chains in vicinity were also listed for each complex. For all these interactions the distance threshold was 4 Å.

In the systematic analysis of all available structures, the patterns of Arg finger binding were assigned automatically, based on the composition of H-bonds between the Arg residue and the substrate molecule. We had considered each pair of possible donor and acceptor atoms, where donors were NE/NH1/NH2 atoms of Arg residue (see Fig. 1.2) and the acceptors were oxygen/fluorine atoms of the substrate. The presence of an H-bond was inferred from the atomic distance: an H-bond was stated at distances less or equal 3.2Å, a weak H-bond stated at distances between 3.2 and 4Å, and no possibility of a H-bond was stated for distances over 4 Å [184–186]).

To assign the Arg finger types, several sets of criteria were applied consecutively, as described in the Results. Structures that did not fit any of the criteria were additionally inspected and Arg binding patterns assigned manually. After all the Arg fingers were categorized, the frequency of each interaction type was determined together for the automatically and manually assigned types.

The lysine residues were assumed to be present in AG site if both NZ-O $\alpha$  and NZ-O $\gamma$  distances are shorter than 4Å (finger type “AG”), and to interact only with  $\gamma$ -phosphate if only the second distance met the criteria (finger type “G”). Otherwise, no interaction with a Lys finger was presumed (finger type “None”).

For each complex, AG site was interpreted as occupied by an Arg residue if the closest Arg residue was assigned interaction type “NH1”, “NH1 weak”, “NH2”, “NH2 weak” or “Y-TYPE”, or by a Lys residue if the closest Lys was assigned type “AG”.

All other Arg or Lys residues located within 4Å from  $\gamma$ -phosphate/ $\gamma$ -mimic were listed as present in “G-site”.

## 2.2.1.10. Description of individual catalytic sites

For each of the 3136 analyzed catalytic sites, we have recorded the information on the residues described above and the distances between them, these data are available in Table C.1 in Appendix C, hereafter Table C.1. The properties that were recorded for every catalytic site analyzed are listed in Table 2.1.

Table 2.1: Properties recorded for each catalytic site analyzed

Field	Subfield	Description
Site_id		Internal numeration for all binding sites. The site id is unique
site rel		Site reliability (only for $AlF_4^-$ - containing sites). Possible values:
		<ul style="list-style-type: none"> <li>• y – assessed as reliable</li> <li>• * - probably reliable, see notes</li> <li>• mw – “missing water” – the site might be reliable but there is no water in the structure to evaluate <math>Mg^{2+}</math> binding</li> <li>• n – No, <math>Mg^{2+}</math> ligands missing or incorrect</li> <li>• na – was not evaluated</li> </ul>
Class		Protein assignment to a P-loop NTPase class (i.e. “TRAFAC” or “SF1/SF2”).
Pfam	accession	Pfam assignment for binding site, based on Pfam-PDB mapping, domain that includes $Lys^{WA}$ residue is listed (if simple sequence mapping was possible)
	family name	Pfam family name
Uniprot		Uniprot Id (if PDB is mapped)
Protein name		Protein name from Uniprot assignment. Left empty for structures or chains absent from UP mapping
PDBID		PDB ID
structure properties	method	Structure determination method
	resolution	Resolution in Ångströms
nucleotide	type	Nucleotide analog type bound in this site. One row in this table stands for one nucleotide analog
	id	Chain id and residue number of the nucleotide analog. In case of ADP/GDP complexes with $\gamma$ -phosphate-mimics, the id is for the NDP moiety
model		Number of the model in PDB structure being described. For structures with multiple models (i.e. NMR), only the first one with $Mg^{2+}$ and $Lys^{WA}$ bound to nucleotide is listed.
Cation		Type of divalent cation bound in the site. Allowed values: MG, CA, MN, SR
AG	AG-site	Positively charged group bound between $\alpha$ and $\gamma$ phosphates. $K^+$ and $Na^+$ ions were not accounted for in this study, so in case of a site with a bound $K^+$ the column value is “NONE”
G	G-site	Positively charged group(s) contacting $\gamma$ phosphates. Calculated from “Closest Arg”, “Closest Lys” and “Contacting, Side chain nitrogen atoms” sections

Closest LYS	ID	
	TYPE	Information about the Lys residue that is the closest one to the $\beta$ -phosphate and is not the Lys <sup>WA</sup> .
	nz-alpha-atom	Distances from Lys NZ atom to nearest oxygen atoms of $\alpha$ -phosphate and $\gamma$ -phosphate/its mimic are listed, in a format "atom being reached", then "distance in A" in the next column.
	nz-alpha-dist	A Lys residue is listed for any structure, so it might not be interacting with the phosphate chain, such cases are listed as TYPE = "NONE".
	nz-gamma-atom	
	nz-gamma-dist	
Closest ARG	ID	
	TYPE	
	nh1-alpha-atom	
	nh1-alpha-dist	
	nh1-gamma-atom	Information about the Arg residue closest to the $\beta$ -phosphate. Distances from Arg NH1, NH2 and NE atoms to nearest oxygen atoms of $\alpha$ -phosphate and $\gamma$ -phosphate/its mimic are listed, in a format "atom being reached", then "distance in A" in the next column.
	nh1-gamma-dist	
	nh2-alpha-atom	
	nh2-alpha-dist	An Arg residue is listed for any structure, so it might not be interacting with the phosphate chain, such cases are listed as TYPE = "NONE".
	nh2-gamma-atom	
	nh2-gamma-dist	
	ne-alpha-atom	
	ne-alpha-dist	
	ne-gamma-atom	
ne-gamma-dist		
Closest ASN	asn_ID	
	asn_TYPE	Information about the Arg residue closest to the $\beta$ -phosphate. Distances from Asn ND2 atom to nearest oxygen atoms of $\alpha$ -phosphate and $\gamma$ -phosphate/its mimic are listed, in a format "atom being reached", then "distance in A" in the next column.
	asn_nd2-alpha-atom	
	asn_nd2-alpha-dist	An Asn residue is listed for any structure, so it might not be interacting with the phosphate chain, such cases are listed as TYPE = "NONE".
	asn_nd2-gamma-atom	
	asn_nd2-gamma-dist	
Other residues contacting gamma-phosphate/mimic (excluding closest ARG, closest LYS, P-loop LYS and G13 analog)	Contacting, backbone N	Backbone nitrogen atoms in close proximity (4 Å) from $\gamma$ -phosphate, residues described in other columns not listed
	Contacting, side chain nitrogen atoms	Side chain nitrogen atoms in close proximity (4 Å) from $\gamma$ -phosphate, residues described in other columns not listed.
G13 ANALOG	RESIDUE	
	Nucleotide atom	The interaction of K-3 residue with $\gamma$ -phosphate via backbone nitrogen atom
	distance to N	
P-loop Lys	ID	Information about conserved P-loop Lys residue, identified from



	Distance to beta-phosphate O	close proximity to $\beta$ -phosphate and sequence check (next residue is a Ser/Thr). Sites with no such Lys residue in 5Å from $\beta$ -phosphate were deemed incomplete/open and excluded from the table. Some kinases possess a substitution in Walker A motif and no S/T immediately following Lys <sup>WA</sup> , these structures were added back to the table manually, see "Notes"
	Ser <sup>K+1</sup> -Mg distance	Distance from [Ser/Thr] <sup>K+1</sup> to Mg <sup>2+</sup> ion, used to estimate general reliability of structure determination in the binding site. No sites were excluded on this criterium.
	WB-Asp/Glu	Presumed [Asp/Glu] <sup>WB</sup> , identified as nearest acidic residue in the relation of [Ser/Thr] <sup>K+1</sup> with hydrophobicity check. In case there were no residues found, search was repeated with no hydrophobicity check. Such cases were marked with "False" in "is_hydro" column. If there was no acidic residue within 5Å from [Ser/Thr] <sup>K+1</sup> , NOT_FOUND error is shown.
Ser/Thr <sup>K+1</sup> and Walker B	Asp/Glu <sup>WB</sup> -Ser <sup>K+1</sup> _dist	Distance from [Ser/Thr] <sup>K+1</sup> to Asp <sup>WB</sup> .
	Asp/Glu <sup>WB</sup> -Mg-dist	Distance from Asp <sup>WB</sup> to Mg <sup>2+</sup> ion. Short distances indicate direct coordination of Mg <sup>2+</sup> by the acidic residue (typically Glu), long distances indicate open binding site, low structure resolution at the binding site or problems with correct identification of residues at the binding site by our tool.
	is_hydro	Hydrophobicity check for three residues preceding Asp <sup>WB</sup> . Typically, Asp <sup>WB</sup> is preceded by four hydrophobic residues. Polar residues unable to assist proton transfer were allowed.
	preceding_res	Three preceding residues, sequence
	Asp/Glu <sup>WB</sup> SASA	Solvent-accessible surface area for the identified Asp <sup>WB</sup> .
	Asp/Glu <sup>WB</sup> SASA,%	Relative solvent-accessible surface area for the identified Asp <sup>WB</sup> .
water	water_present	Are water molecules present in the structure? This does not guarantee that there are water molecules identified in the catalytic site.
Comment	Notes	Manually added comments

### 2.2.1.11. Structure visualization

For each catalytic site analyzed, we have automatically generated a quick visualization script for Pymol [187] (catalytic site elements displayed and colored as representative structures in the Introduction section). Elements for which named selections are displayed and defined include:

- nucleotide and each of its phosphate groups,
- Lys<sup>WA</sup>, X<sup>K+3</sup> and their bonds to the phosphate chain,
- [Ser/Thr]<sup>K+1</sup>,
- [Asp/Glu]<sup>WB</sup>, and, if available, its bonds to Ser/Thr<sup>K+1</sup> and water molecules (distance threshold 3.4 Å),

- $\text{Mg}^{2+}$ , its coordinating water molecules (numbered), Ser/Thr<sup>SI</sup> (if available), coordination bonds,
- putative Arg/Lys stimulators and their bonds to the phosphate chain,
- $W_{\text{cat}}$  and its interactors, if available,
- other  $\gamma$ -phosphate interactors in vicinity of  $3\text{\AA}$ , if available,

Structure is automatically superimposed with a previously opened structure with the properties defined in a similar script. Catalytic site figures in this manuscript were produced using the resulting structure visualizations.

#### 2.2.1.12. Calculation of solvent-accessible surface area

Solvent-accessible surface area values (SASA) were calculated using Shrake-Rupley method (1) implemented in BioPython [188] (method Bio.PDB.SASA.ShrakeRupley) and (2) implemented in Pymol v. 2.5.0 [187]. The method [189] involves creating a mesh of points representing the surface of each atom, at a distance of the van der Waals radius plus the probe radius from the nuclei. To generate the points on the sphere, the golden section spiral algorithm is used [190]. This involves creating a spiral that traces out the unit sphere and placing points equidistant along the spiral. The number of points on the molecular surface, i.e. not within the radius of another atom, is then counted. Assuming the points are evenly distributed, the number of points is directly proportional to the accessible surface area, which can be calculated by multiplying the fraction of the points that are accessible by  $4\pi r^2$ .

We have calculated SASA and relative SASA for identified Asp<sup>WB</sup> residues. To speed up the calculation, only the residues having at least one atom within  $8\text{\AA}$  from the query Asp<sup>WB</sup> were included in the Biopython calculation. The SASA values were stored only for Asp<sup>WB</sup> residues. The number of points was set to 200, solvent radius was set to 1.4.

## 3. Results

### 3.1. Computational comparative structural analysis of P-loop NTPases

The ever-rising four-digit numbers of P-loop NTPase structures with ATP, GTP, and their analogs bound, as deposited in the PDB, demanded computational approaches. The search in the Protein Data Bank (PDB) at [www.rcsb.org](http://www.rcsb.org) [5,191] for proteins assigned to the entry IPR027417 “P-loop containing nucleoside triphosphate hydrolase” of the Interpro database [192] yielded as many as 1484 structure entries with 3666 catalytic sites with NTP or NTP-mimicking molecules bound (as of 11.09.2019; many of the structures contained several catalytic sites). As discussed in Section 2.2.1 and depicted in Fig. 1.5, the shape of the P-loop and the spatial position of Asp<sup>WB</sup> relative to it are strictly conserved among P-loop NTPases. Building on that, we were able to use the atomic coordinates of the Mg-triphosphate to pinpoint the catalytically relevant residues in an almost sequence-agnostic way. We applied this approach to perform analysis of nucleotide-binding sites in thousands of structures of P-loop NTPases. The criteria for selection of full-fledged catalytic sites from this set and the routine of their subsequent structural analysis are described and depicted in the Section 2.2.1 of Methods. After filtering, we obtained 1383 structure entries with 3136 catalytic sites containing complexes of Mg<sup>2+</sup> ions with NTPs or NTP-like molecules, these structures were subjected to further analysis. The relevant data for all these catalytic sites are presented in Excel Tables C.1, D.1, E.1 in Appendices C, D, E, respectively. These tables contain the main results of this work.

Based on the type of the molecule bound, the catalytic sites could be sorted into four groups (see counts and compound names in Table 3.1): 1043 sites contained native ATP/GTP molecules; 1612 sites contained bound non-hydrolyzable NTP analogs, 234 sites contained NDP:fluoride complexes mimicking the substrate state, such as NDP:BeF<sub>3</sub> and NDP:AlF<sub>3</sub>, and 247 sites contained NDP:AlF<sub>4</sub><sup>-</sup> (204), NDP:MgF<sub>3</sub><sup>-</sup> (10) and ADP:VO<sub>4</sub><sup>-</sup> (33) thought to be TS analogs [68,70,71,73,96].

Table 3.1: Sampled catalytic sites of P-loop NTPases by type of NTP analog bound

Type	Compound name	PDB ligand id	Total structures <sup>†</sup>	Structures, qualified	Complexes, total	Complexes, selected
Nucleotides	ATP		268	241	842	744
	GTP		205	186	326	299
	<b>Total</b>		<b>473</b>	<b>427</b>	<b>1168</b>	<b>1043</b>
NON-HYDROLYZABLE NTP ANALOGS						
ATP non-hydrolyzable analogs	Adenosine 5'-[ $\beta,\gamma$ -imido] triphosphate (AMP-PNP)	ANP	169	159	488	446
	Adenosine 5'-[ $\beta,\gamma$ -methylene] triphosphate (AMP-PCP)	ACP	21	18	39	35
	Adenosine 5'-[ $\gamma$ -thio] triphosphate (ATP- $\gamma$ -S)	AGS	62	55	257	236
GTP non-hydrolyzable analogs	Guanosine 5'-[ $\beta,\gamma$ -imido] triphosphate (GMP-PNP)	GNP	633	383	952	685
	Guanosine 5'-[ $\beta,\gamma$ -methylene] triphosphate (GMP-PCP)	GCP	108	58	163	109
	Guanosine 5'-[ $\gamma$ -thio] triphosphate (GTP- $\gamma$ -S)	GSP	74	73	102	101
<b>Total</b>			<b>1067</b>	<b>746</b>	<b>2001</b>	<b>1612</b>
TS MIMICS						
	$\gamma$ -mimic	$\gamma$ -mimic PDB ligand id	Total structures	Structures qualified	Complexes, total	Complexes, selected
ADP* $\gamma$ -mimic	$\text{AlF}_4^-$	ALF	39	37	107	104
	$\text{MgF}_3^-$	MGF	1	1	2	2
	$\text{VO}_4^{3-}$	VO4	22	21	35	33
GDP* $\gamma$ -mimic	$\text{AlF}_4^-$	ALF	48	48	103	100
	$\text{MgF}_3^-$	MGF	5	5	8	8
<b>Total</b>			<b>115</b>	<b>112</b>	<b>255</b>	<b>247</b>
GROUND STATE MIMICS						
ADP* $\gamma$ -mimic	$\text{AlF}_3$	AF3	15	13	25	23
	$\text{BeF}_3$	BEF	61	61	156	150
GDP* $\gamma$ -mimic	$\text{AlF}_3$	AF3	21	21	41	40
	$\text{BeF}_3$	BEF	8	8	21	21
<b>Total</b>			<b>105</b>	<b>103</b>	<b>243</b>	<b>234</b>
<b>TOTAL, ALL COMPOUNDS</b>			<b>1474</b>	<b>1383</b>	<b>3667</b>	<b>3136</b>

<sup>†</sup> structures sampled as assigned to IPR027417, resolution  $<5\text{\AA}$  and listed as containing a divalent cation (Mg, Ca, Sr, Mn) and one of the NTP mimics listed in the table above.

To ensure detailed analysis and to cover as many structures as possible, we combined the systematic computational analysis of all available structures of P-loop NTPases with bound Mg-NTP-like molecules with the manual inspection of catalytic machinery in those representatives from the major classes of P-loop NTPases that were pinpointed as remarkable by computational analysis.

## 3.2. Analysis of binding of transition state analogs to P-loop NTPases

### 3.2.1. Variability in the $\text{AlF}_4^-$ binding to catalytic sites of P-loop NTPases

As already mentioned, TS analogs, namely  $\text{NDP}:\text{AlF}_4^-$ ,  $\text{NDP}:\text{MgF}_3^-$ , and  $\text{NDP}:\text{VO}_4^{3-}$  complexes, “fix” the catalytic pocket in a particular conformation that is same in P-loop NTPases of distinct classes and is thought to mimic the transition state. In addition, TS analogs promote formation of complexes between the P-loop NTPase and its activator, specifically in case of monomeric NTPases, see Section 1.2.1 of Introduction and [71,78,77,76,68,74,64,79,80,73,81,75].

It is well established that, upon hydrolysis of ATP or GTP, a pentavalent trigonal bipyramidal (tbp) transition state is formed (Fig. 1.4). As  $\text{VO}_4^{3-}$  and  $\text{MgF}_3^-$  are similar to the planarized  $\gamma$ -phosphate in its anticipated TS both in geometry and electric charge, their performance as TS mimics is not surprising.

The  $\text{AlF}_4^-$  moiety, on the other hand, is not a TS mimic in a strict sense as it has a different geometry, it is tetragonal instead of trigonal, see Figure 1.4A-B. Still, its ability in assisting with the assembly of activated complexes is unmatched,  $\text{NDP}:\text{AlF}_4^-$  complexes are more potent analog of the TS than  $\text{NDP}:\text{MgF}_3^-$  complexes as it follows from functional studies [64,69,71,76–78,80] and also from the distance between  $\mathbf{W}_{\text{cat}}$  and metal atom being shorter (2.0-2.1Å) in  $\text{NDP}:\text{AlF}_4^-$  than in  $\text{NDP}:\text{MgF}_3^-$  complexes (approx. 2.5 Å) [68,74]. The reason of why  $\text{NDP}:\text{AlF}_4^-$  performs better than  $\text{NDP}:\text{MgF}_3^-$  in all these cases has remained obscure, see e.g. [75]. We, therefore, decided to take a closer look at  $\text{AlF}_4^-$  binding to the catalytic sites of P-loop NTPases from our sample of structures.

#### 3.2.1.1. Different modes of $\text{AlF}_4^-$ interaction with the cofactor $\text{Mg}^{2+}$ ion

After inspecting the distances between the atoms of the  $\text{AlF}_4^-$  moieties and their surroundings we have found out that  $\text{AlF}_4^-$  moieties can interact with the cofactor  $\text{Mg}^{2+}$  ion in two different ways, at least, see Fig. 3.1A-D and Table D.1 in Appendix D. In most structures (77% of all  $\text{AlF}_4^-$  complexes), only one fluorine atom interacts with  $\text{Mg}^{2+}$ , like its structural counterpart, the  $\text{O}^{\text{IG}}$  atom of  $\gamma$ -phosphate. In this case, the next closest

fluorine atom is on  $> 3.0 \text{ \AA}$  from  $\text{Mg}^{2+}$ , see Fig. 3.1A, 3.1C and Table D.1. In some structures, however, the two fluoride atoms are at similar distances of 2.0-2.7  $\text{Å}$  from the  $\text{Mg}^{2+}$  ion and both appear to interact with it, which is only possible when the  $\text{AlF}_4^-$  is rotated by approx.  $45^\circ$  around the  $\text{O}^{3\text{B}}-\text{Al}$  bond, see Fig. 3.1B, 3.1D and Table D.1. This kind of interaction is non-physiological and in some cases prevents  $\text{Mg}^{2+}$  from making a bond with one of its physiological ligands (see Fig. 3.1B and Table D.1). Our manual inspection of the  $\text{AlF}_4^-$  structures showed that, in general, structures with a second fluorine atom found within 2.7  $\text{Å}$  from  $\text{Mg}^{2+}$  (these include structures with different degrees of  $\text{AlF}_4^-$  rotation) often have distortions in the  $\text{Mg}^{2+}$  coordination sphere. The most dramatic distortion is the absence of coordination bond between  $\text{Ser}^{\text{K}+1}$  and  $\text{Mg}^{2+}$ , which usually leads to a partial opening of the catalytic site and an increased distance between  $\text{Ser}^{\text{K}+1}$  of the Walker A motif and  $\text{Asp}^{\text{W}+1}$  of the Walker B motif, see Fig. 3.1B and Table C.1. In other cases where the  $\text{AlF}_4^-$  moieties are twisted, the distortions manifest themselves in incorrect bonding to  $\beta$ -phosphate or missing bonds to other  $\text{Mg}^{2+}$  ligands. Remarkably, no major distortions are seen in the structure of transducin  $G_{\text{tu}}$  complexed with  $\text{Ca}^{2+}$  instead of  $\text{Mg}^{2+}$  (PDB ID 1TAD, Fig. 3.1D). Here, in spite of the twisted  $\text{AlF}_4^-$  the interactions in the coordination sphere are preserved, supposedly due to a larger ionic radius of  $\text{Ca}^{2+}$  and its ability to bind up to 8 ligands.

Accordingly, all structures with two bonds between fluoride atoms and  $\text{Mg}^{2+}$  (indicated in Table D.1) appear to be suspicious as TS-state analogs due to non-physiological coordination of the  $\text{Mg}^{2+}$  ion.

The data on all found non-physiologically bound  $\text{AlF}_4^-$  moieties are highlighted pink in Table C.1 and separately summarized in Table D.1.

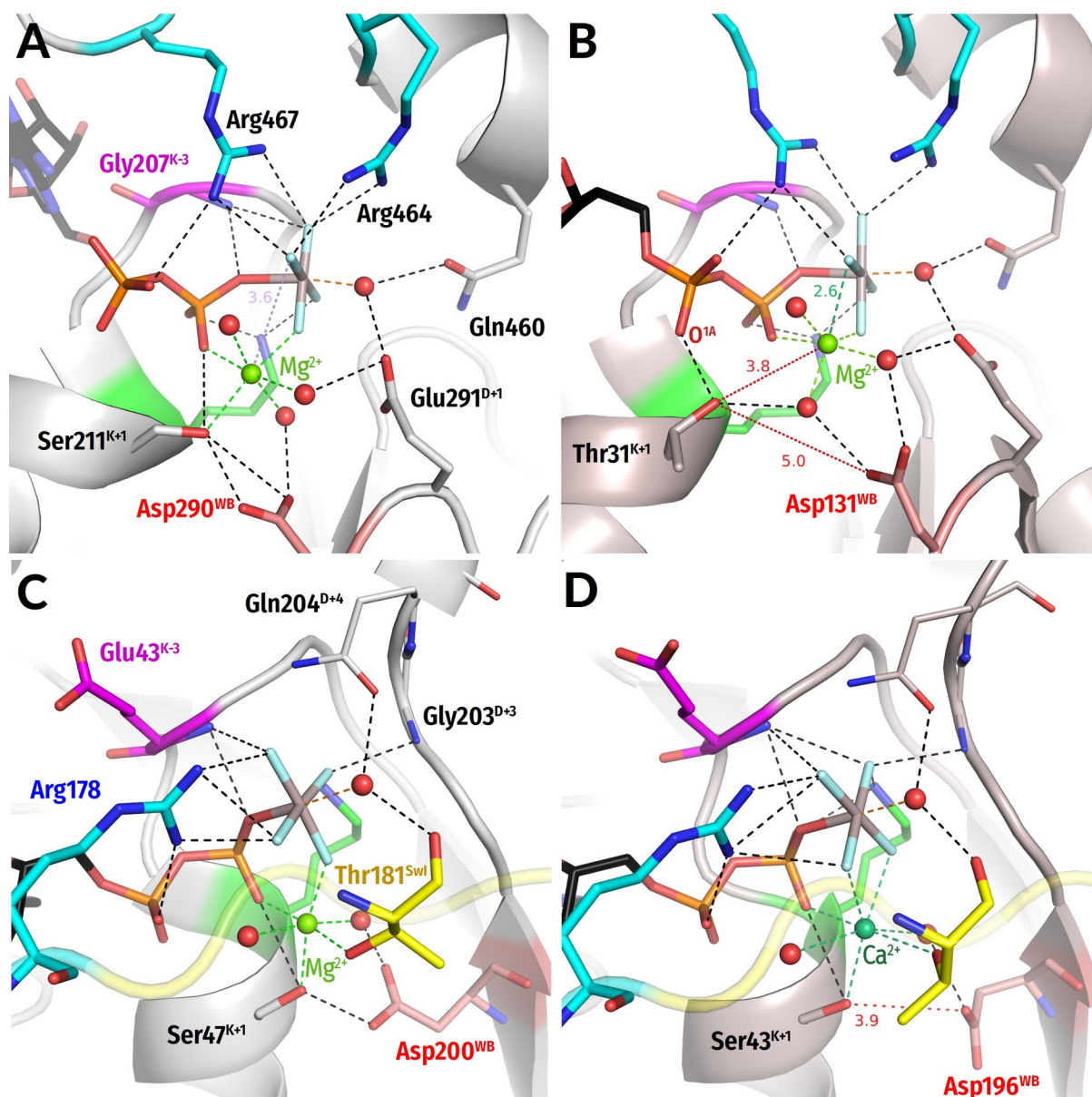


Figure 3.1: Variation in the  $\text{AlF}_4^-$  interaction with  $\text{Mg}^{2+}$  ions.

Residues are highlighted as in Fig.1.6. The distances that are too long for H-bonds are indicated by red dashed lines. All distances are given in ångströms. **A.** The  $\text{ADP}:\text{AlF}_4^-$  binding in the structure of HCV (Hepatitis C Virus) NS3 helicase, PDB 5E4F, chain A [193]. The  $\text{Mg}^{2+}$  ion is coordinated by the  $\text{O}^{2\text{B}}$  atom of ADP, one fluorine atom of  $\text{AlF}_4^-$ , the conserved Ser residue of Walker A motif, and three water molecules. **B.** The  $\text{ADP}:\text{AlF}_4^-$  binding in the structure of double-stranded RNA (dsRNA)-dependent helicase LGP2 from chicken (PDB 5JAJ [194]). Here,  $\text{AlF}_4^-$  is rotated counterclockwise and makes two bonds with the  $\text{Mg}^{2+}$  ion; the  $\text{Ser}^{\text{K}+1}$  residue is no longer directly coordinating  $\text{Mg}^{2+}$  and, instead, is forming an unusual bond with the  $\text{O}^{1\text{A}}$  atom of ADP. **C.** The  $\text{AlF}_4^-$  binding in a structure of guanine nucleotide-binding protein  $\text{G}\alpha_{\text{i}3}$  complexed with the regulator of G-protein signaling 8 (PDB 2ODE [126]). The  $\text{Mg}^{2+}$  ion is coordinated by the  $\text{O}^{2\text{B}}$  atom of GDP, one fluorine atom of  $\text{AlF}_4^-$ , the conserved Ser residue of Walker A motif, the conserved Thr of Switch I, and two water molecules. **D.** In the structure of transducin  $\text{G}\alpha$  complexed with  $\text{Ca}^{2+}$  instead of  $\text{Mg}^{2+}$  (PDB 1TAD [63]),  $\text{AlF}_4^-$  is rotated and forms two bonds with  $\text{Ca}^{2+}$ , but all interactions in the coordination sphere are preserved due to a larger ionic radius of  $\text{Ca}^{2+}$  and its ability to bind up to 8 ligands. However, the catalytic site is more open, as evidenced by a distance of 3.9 Å between  $\text{Ser}^{\text{K}+1}$  and  $\text{Asp}^{\text{WB}}$ .

Still, the comparison of the same or closely related NTPases with differently bound  $\text{NDP}:\text{AlF}_4^-$  complexes has proven useful. Fig. 3.1A, 3.1B shows two structures of the Family 2 helicases (ASCE division, SF1/SF2 class, see Fig. A.2) with different

coordination of the  $\text{AlF}_4^-$  moiety. One can see that the non-physiological coordination in Fig. 3.1B is achieved via counterclockwise rotation of the more “physiological” configuration of the  $\text{AlF}_4^-$  moiety in Fig. 3.1A, whereby the interactions of  $\text{AlF}_4^-$  with  $\text{Lys}^{\text{WA}}$  and the stimulatory Arg residues are retained in both structures. The two stimulatory fingers retain their H-bonds with  $\text{AlF}_4^-$  in both configurations. This comparison shows that the residues which bind  $\gamma$ -phosphate appear to be adapted to the counterclockwise rotation of  $\gamma$ -phosphate by 30-40°. Similar counterclockwise rotation of  $\gamma$ -phosphate can be inferred from the comparison of  $G_\alpha$  protein structures (Kinase-GTPase division, TRAFAC class) in Fig. 3.1C-D.

### 3.2.1.2. Untypical binding of $\text{AlF}_4^-$ in RecA ATPases

Another type of untypical  $\text{ADP}:\text{AlF}_4^-$  binding is observed in RecA NTPases; in these, the  $\text{AlF}_4^-$  moiety is oriented differently as compared to other TS analog-containing structures, and the apical water molecule is not resolvable. We scrutinized this kind of untypical  $\text{AlF}_4^-$  binding in the framework of a collaboration with the research groups of B. H. Meier at the Laboratory of Physical Chemistry, ETH Zurich, Switzerland and A. Böckmann at the Université de Lyon, France. The two groups had measured, in the presence of diverse ATP analogs, the solid state NMR spectra of SF4 helicase DnaB from *Helicobacter pylori* and ABC transporter ATPase BmrA from *Bacillus subtilis* ([159,195]). In the case of BmrA ABC ATPase, the NMR spectrum measured in the presence of  $\text{ADP}:\text{AlF}_4^-$  completely overlapped with the spectrum measured in the presence of the  $\text{ADP-VO}_4^{3-}$  where the  $\text{ADP-VO}_4^{3-}$  moiety is covalently bound to ADP ([195]). In the case of the SF4 helicase DnaB from *Helicobacter pylori*, in contrast, the averaging of the  $^{27}\text{Al}$  quadrupolar coupling constant in combination with the single  $^{19}\text{F}$  resonance indicated that the  $\text{AlF}_4^-$  moiety did not form tight coordination bonds in contrast to the tight H-bond network that was observed for ADP ([159]). The solid state NMR data showed that all six catalytic sites of the *Helicobacter pylori* DnaB hexamer were fully occupied by  $\text{ADP}:\text{AlF}_4^-$  complexes capable of free rotational diffusion.

Trying to understand the possible reasons for such untypical behavior of  $\text{AlF}_4^-$  in the *Helicobacter pylori* DnaB helicase, we contributed by analyzing the available, TS analogs-containing structures of homologs of the *Bacillus subtilis* BmrA and *Helicobacter pylori* DnaB, respectively.

The observed overlap of solid state NMR spectra of *Bacillus subtilis* BmrA ABC ATPase with bound  $\text{ADP}:\text{AlF}_4^-$  and  $\text{ADP-VO}_4^{3-}$ , respectively ([195]), is in perfect agreement with the relevant crystallographic data on *E. coli* maltose ABC transporter, showing that the configuration of the catalytic site with  $\text{ADP}:\text{AlF}_4^-$  was indistinguishable from that with  $\text{ADP-VO}_4^{3-}$  bound [150]. In this case, the  $\text{AlF}_4^-$  moiety makes many bonds with surrounding atoms and is tightly fixed in the site.



As a counterpart of DnaB from *Helicobacter pylori* we took the crystal structure of the DnaB helicase of *Geobacillus stearothermophilus* which was obtained in the presence of GDP:AlF<sub>4</sub><sup>-</sup> instead of ADP:AlF<sub>4</sub><sup>-</sup> and in the presence of Ca<sup>2+</sup> ions instead of Mg<sup>2+</sup> ions. This structure resembled the DnaB from *Helicobacter pylori* in that five of the six catalytic sites contained GDP:AlF<sub>4</sub><sup>-</sup> complexes bound.

As it follows from the superposition of two crystal structures in Fig. 3.2, the position of AlF<sub>4</sub><sup>-</sup> in the structure of *Geobacillus stearothermophilus* DnaB (Figure 3.2B) differs from that typical for other P-loop NTPases, which in the figure is represented by the crystal structure of the BmrA ABC ATPase from *Bacillus subtilis*. In the DnaB structure, no catalytic water molecules are present apically to the plane of AlF<sub>4</sub><sup>-</sup> moieties and their position does not correspond to that of the  $\gamma$ -phosphate group (see Figure 3.2). This unusual position of the AlF<sub>4</sub><sup>-</sup> moiety in DnaB helicases may account for its free rotational diffusion in the site, as elaborated elsewhere [159] and deliberated in the section 4.1.1 of Discussion.

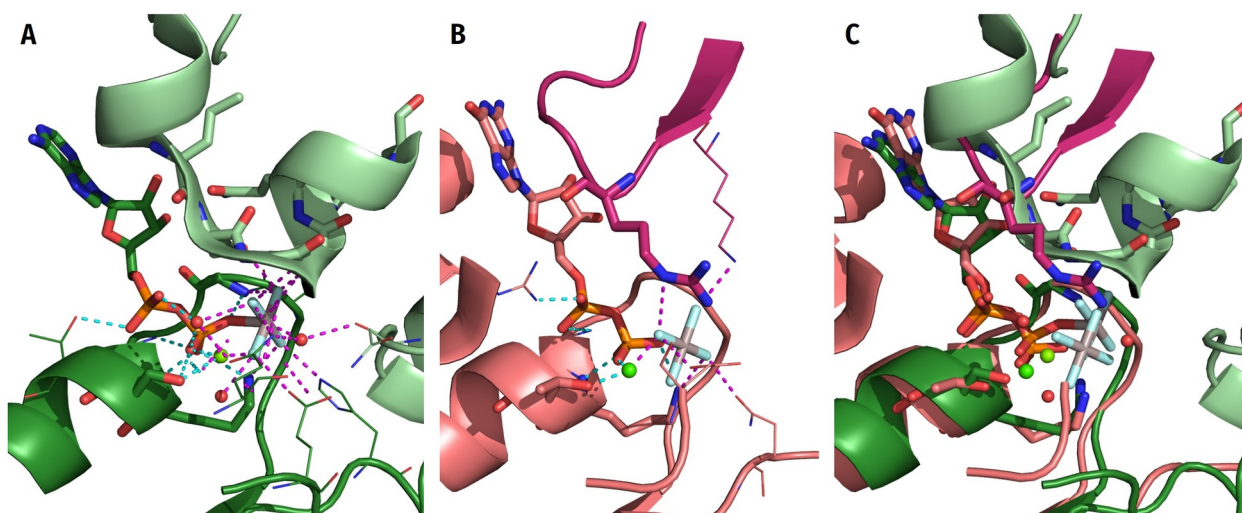


Figure 3.2: Structural comparison of nucleotide-binding sites in SF4 helicase DnaB from *Geobacillus stearothermophilus* and ABC transporter MalK from *Escherichia coli* K-12.

Different protein subunits are colored in different shades of the same color. Mg<sup>2+</sup> or Ca<sup>2+</sup> ions are shown as green spheres, water molecules are shown as red spheres; H-bonds and metal interactions involving  $\alpha$ - and  $\beta$ -phosphates are shown as dashes in cyan, interactions with AlF<sub>4</sub><sup>-</sup> are shown as dashes in magenta. Nucleotide analog, P-loop motif residues and activating residues (Arg residue or LSGGQ motif) shown as thick sticks, other interacting amino acid residues as thin sticks. A, Maltose/maltodextrin import ATP-binding protein MalK (PDB ID 3PUW [150], chain B) from *Escherichia coli* K-12. B, Replicative helicase DnaB from *Geobacillus stearothermophilus* (PDB ID 4ESV [162], chain E). C, Structures 3PUW and 4ESV superimposed by phosphate chain and ribose atoms of NDP moieties.

### 3.3. Comparative structural analysis of stimulatory patterns of P-loop NTPases

#### 3.3.1. Stabilization of the $O^{2G}$ atom of $\gamma$ -phosphate by the backbone $HN^{K-3}$ group

For the MnmE GTPase, MD simulations have shown that the insertion of the stimulatory  $K^+$  ion and its simultaneous interaction with  $O^{2A}$ ,  $O^{3B}$ , and  $O^{3G}$  atoms triggered the twist of  $\gamma$ -phosphate, leading to the formation of a new H-bond between the  $O^{2G}$  atom and **HN** of Asn226<sup>K-3</sup>, see [102] and Fig. 1.2E. Generally, the position of **HN**<sup>K-3</sup> in the vicinity of the  $O^{2G}$  atom (or the corresponding atom of an NTP analog) is structurally conserved across P-loop NTPases, being determined by the highly conserved H-bond of **HN**<sup>K-3</sup> with the bridging  $O^{3B}$  oxygen (see Fig. 1.2D,E).

To assess the possibility of a transient H-bond formation between **HN**<sup>K-3</sup> and  $O^{2G}$  of  $\gamma$ -phosphate, as found in MD simulations of MnmE GTPases [102], we measured corresponding distances in the available structures of P-loop NTPases with bound substrates or their analogs as described and depicted in Methods. The data obtained are presented in Table C.1 and Fig. 3.3, where the H-bond-compatible distance range is highlighted in amber. For simplicity, we used the same threshold of 3.4 Å for the H—F and H—O bonds. On the one hand, this value is somewhat lower than the threshold of 3.5 Å, as suggested for H-bonds in protein structures by Martz [196]. On the other hand, this distance corresponds to the longest F-H-N bond reported for crystalized L-cysteine-hydrogen fluoride [184]. Still, this threshold is rather arbitrary; according to Jeffrey, weak H-bonds in proteins can have donor-acceptor distances up to 4.0 Å long [197].

In all groups of complexes, there is a fraction with distances shorter than 3.4 Å between **HN**<sup>K-3</sup> and the nearest  $O^G$  atom or its structural analog (hereafter **HN**<sup>K-3</sup>— $O^G$  distance). For the ATP- and GTP-containing structures, this fraction makes 31% of complexes (326 of 1043 binding sites); for the non-hydrolyzable analogs, it makes 24% of complexes (392 out of 1612 binding sites). Among complexes with NDP:BeF<sub>3</sub>, the **HN**<sup>K-3</sup>— $O^G$  distance is < 3.4 Å in 20.5% of cases (35 catalytic sites out of 171), whereas in the case of structures containing NDP:AlF<sub>3</sub> this fraction makes 28.5% of catalytic sites.

However, the distances between **HN**<sup>K-3</sup> and  $O^G$  analog are shorter than 3.4 Å in most structures with TS analogs as shown in Fig. 3.3 where the data for the three TS analogs are plotted separately for clarity. Distances of < 3.4 Å are observed in 61% of complexes with ADP:VO<sub>4</sub><sup>3-</sup> (out of 33), in seven out of ten NDP:MgF<sub>3</sub><sup>-</sup> complexes and in 80% of NDP:AlF<sub>4</sub><sup>-</sup> complexes (163 out of 204). Hence, the TS-like structures of catalytic sites correlate with shortening of the distance between **HN**<sup>K-3</sup> and  $O^G$  analog.

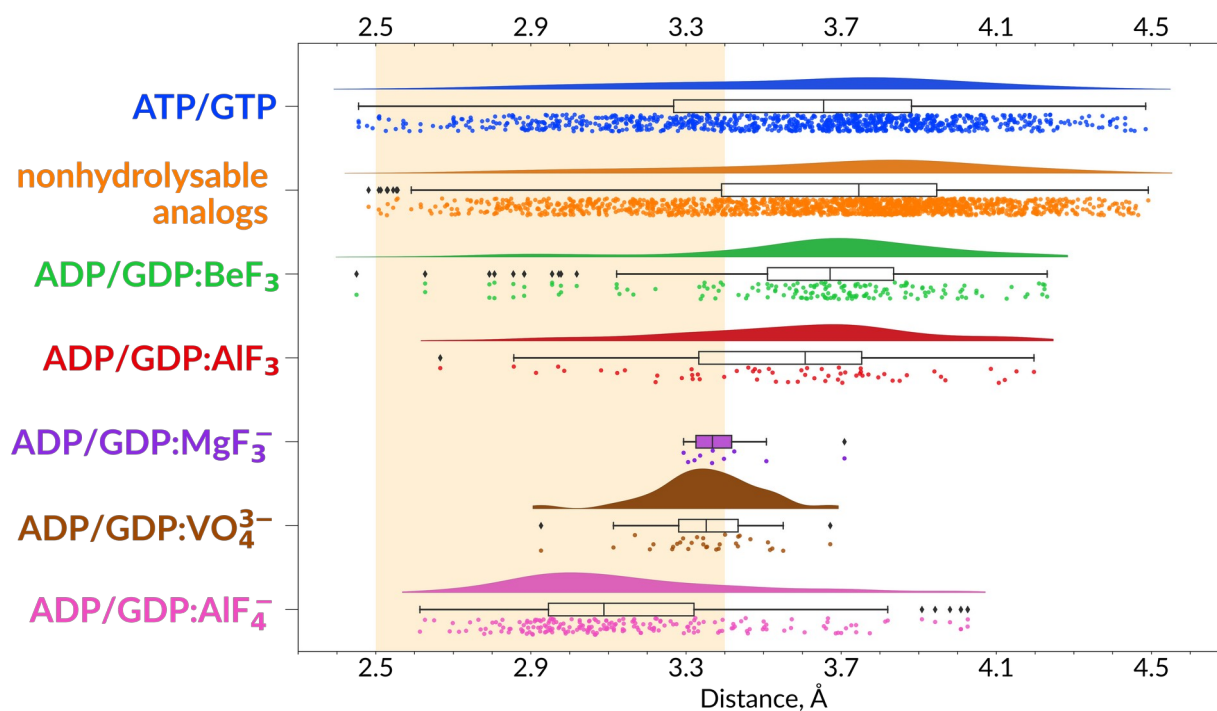


Figure 3.3: Distances between  $\text{HN}^{\text{K-3}}$  and the  $\text{O}^{2\text{G}}$  atom or its analog in the catalytic sites of P-loop NTPases. For each type of complexes, distances are visualized as a kernel density estimate (KDE) plot, a boxplot, and individual data points, each point representing one catalytic site in one structure. For  $\text{ADP/GDP:MgF}_3^-$  complexes the density plot is not shown because of scarcity of data. The range of H-bond-compatible lengths is highlighted in amber.

### 3.3.2. Pre-catalytic configurations in NTP-containing structures.

While H-bond-compatible distances between  $\text{HN}^{\text{K-3}}$  and  $\text{O}^{\text{G}}$  (or its analog in the case of TS-like structures, Fig. 3.3) confirmed the earlier suggestion on the importance of this H-bond for stabilization of the TS [102], we were rather surprised to see that the  $\text{HN}^{\text{K-3}}-\text{O}^{\text{G}}$  distances were H-bond compatible also in 31% of ATP- and GTP-containing structures (Fig. 3.3, Table C.1). Therefore, we inspected the top 100 high-resolution NTP-containing structures with H-bond compatible  $\text{HN}^{\text{K-3}}-\text{O}^{2\text{G}}$  distances to clarify their origin.

In principle, an ATP or GTP molecule could be crystallized within an NTPase only if the latter is inactive. Not surprisingly, the ATP or GTP-containing P-loop NTPases which were crystallized in the absence of their cognate activators made the majority in the inspected set. Although in these proteins the  $\text{HN}^{\text{K-3}}-\text{O}^{2\text{G}}$  distances were indeed H-bond compatible, we did not explore these structures further because they could hardly help in clarifying the mechanisms of hydrolysis stimulation.

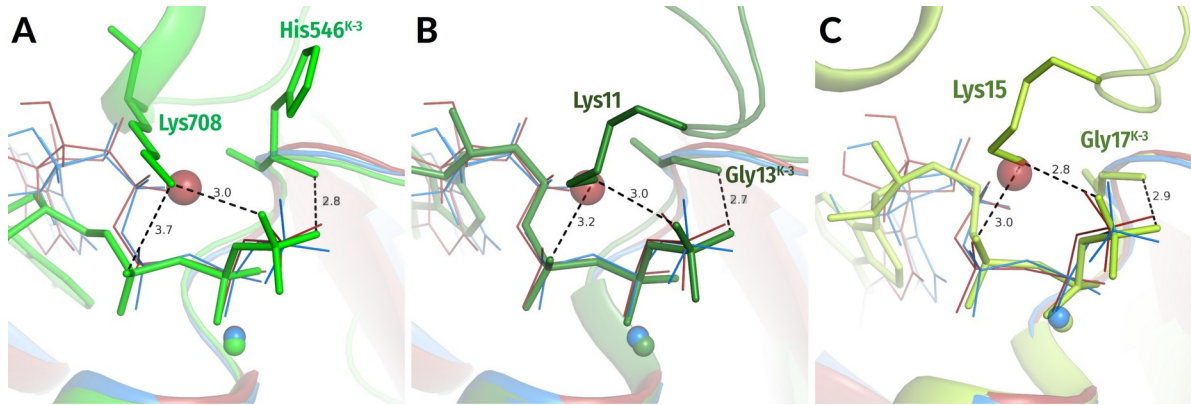


Figure 3.4: Conformation of native ATP molecules in selected crystal structures as compared to MD simulations of the MnmE GTPase.

Crystal structures of P-loop ATPases in complex with ATP molecules are shown in lime or green colors; they are superposed with conformations of GTP bound to the MnmE GTPase, as sampled from MD simulations of inactive monomer (blue) and active,  $K^+$ -bound dimer (red), see Fig. A.1 and [102]. All distances are given in ångströms. **A.** N-ethylmaleimide-sensitive factor (PDB ID 1NSF, [145]); **B.** ATPase MinD (PDB ID 3Q9L, [377]); **C.** Chromosome segregation protein Soj (PDB ID 2BEK, [378]).

Still, we could identify a set of structures where NTP molecules remained not hydrolyzed despite the presence of activating partner(s). In many such complexes, the interaction of  $W_{\text{cat}}$  with  $\gamma$ -phosphate was hindered by the incompleteness of  $W_{\text{cat}}$  ligands, specifically caused by mutations, so that the NTP-binding sites were trapped in precatalytic configurations. Some such structures are shown in Fig. 3.4A-C. In these enzymes, the stimulatory fingers are inserted into the catalytic sites,  $\gamma$ -phosphates are twisted, and the triphosphate chains are in a configuration like that observed upon the MD simulations of MnmE GTPase in the presence of a bound stimulatory  $K^+$  ion, see Fig. A.1 and [102]; this configuration is superimposed as a dark-red contour. One can see from Fig. 3.4A-C that H-bond compatible  $\text{HN}^{\text{K-3}}-\text{O}^{2\text{G}}$  distances correlate with linking of  $\text{O}^{2\text{A}}$  and  $\text{O}^{3\text{G}}$  atoms by the stimulator and twisting  $\gamma$ -phosphate.

### 3.3.3. Interactions of stimulating moieties with the triphosphate chain

In most families of P-loop NTPases, Arg, Lys, or Asn residues serve as stimulatory moieties. We used computational approach to inspect the patterns of their interactions with phosphate chain atoms (or their analogs) in 3136 catalytic sites of P-loop NTPases with nucleoside triphosphates or their analogs bound. The interactions of monovalent cations were quantified in our group earlier [102]. For ABC ATPases, there was not much to quantify because there are only four available TS-like structures [150,151]. The aim of the comparative structure analysis was to find common features in their stimulatory patterns and to use this information for elucidating the general mechanism of these enzymes.

For this purpose, we analyzed, as described in the Methods, the same 3136 PDB structures of full-fledged catalytic sites that contain complexes of  $\text{Mg}^{2+}$  ions with NTPs or NTP-like molecules. For each complex, we measured the distances between oxygen atoms of the triphosphate chain (or their structural counterparts in the NTP analogs) and the amino groups of Arg, Lys, and Asn side chains within 4 Å radius, for further details see the Methods section and the scheme of the analysis pipeline, Fig. 2.1. The distances were measured towards the NE/NH1/NH2 atoms of Arg residues; the NZ atom of Lys residues; and the ND2 atom of Asn residues (see Fig. 3.5 and 3.6 for the atom naming scheme). The data on all atom pairs and corresponding distances are summarized in Table C.1.

We have found that more than half of analyzed Mg-NTP complexes (60%) have none of the inspected residue types within the 4Å radius around oxygen atoms. Those are structures of P-loop NTPases that were crystallized in the absence of their activating partners or are stimulated by moieties other than Asn/Arg/Lys (e.g. a monovalent cation or the signature motif of ABC ATPases, see below).

In the remaining 1380 catalytic sites of P-loop NTPases, at least one Arg, Asn, or Lys residue (other than the reference Lys residue of the Walker A motif) was found in the proximity of the phosphate chain and categorized as a stimulator. The analysis of interactions between Arg, Lys, and Asn fingers and the phosphate chains has revealed several distinct types of configurations, which hereafter are called "stimulatory patterns". The Table E.1 shows how many proteins were assigned to each stimulatory pattern.

### 3.3.3.1. Stimulatory patterns of Arg fingers

Arginine fingers are the most widespread stimulatory moieties among P-loop NTPases. In an arginine side chain, the positive charge is distributed over three nitrogen atoms of the guanidinium group. In principle, each of these atoms can interact with the phosphate chain. Consequently, we observed a variety of interactions for the Arg fingers.

In most cases, the type of stimulatory pattern was assigned automatically based on the H-bond compatibility of distances between the NTP molecule or its analog and nearby Arg residue(s). Here we relied on Jeffrey who categorized H-bonds with donor-acceptor distances of 2.2-2.5 Å as "strong, mostly covalent", those with distances of 2.5-3.2 Å as "moderate, mostly electrostatic", and H-bonds of 3.2-4.0 Å as "weak, electrostatic" [197].

Due to inconsistencies in the atom numbering and differences among NTP analogs, we measured the distances from the Arg side chain nitrogen atoms to the nearest oxygen atom of  $\alpha$ -phosphate (hereafter  $\text{O}\alpha$ ) and the nearest oxygen of  $\gamma$ -phosphate (hereafter

O $\gamma$ ) or the corresponding atom in NTP analogs. Hereafter, for simplicity, we will use “ $\gamma$ -phosphate” both for  $\gamma$ -phosphate proper and for its analogs.

Several sets of criteria were applied consecutively, with each following criterion applied only to the cases that did not match any of the previous criteria:

- 1) If both distances NH1-O $\alpha$  and NH1-O $\gamma$  do not exceed 3.2Å, the interaction type “NH1” was assigned, meaning that NH1 atom forms H-bonds with both  $\alpha$ - and  $\gamma$ -phosphates. Similarly, “NH2” interaction type was assigned if both distances NH2-O $\alpha$  and NH2-O $\gamma$  were less than 3.2Å.
- 2) If both distances NH1-O $\alpha$  and NH1-O $\gamma$  do not exceed 4Å, whereas both distances NH2-O $\alpha$  and NH2-O $\gamma$  are longer than 4Å, the interaction type “NH1 weak” was assigned, meaning that NH1 atom forms weak interactions with both  $\alpha$ - and  $\gamma$ -phosphates. Analogous criteria were used to assign the “NH2 weak” interaction type.
- 3) If at least one of the distances NH1-O $\gamma$  and NH2-O $\gamma$  do not exceed 3.2Å, whereas both distances NH1-O $\alpha$  and NH2-O $\alpha$  are longer than 4Å, the interaction type “only gamma” was assigned. Similarly, if at least one of the distances NH1-O $\gamma$  and NH2-O $\gamma$  do not exceed 4Å, whereas both distances NH1-O $\alpha$  and NH2-O $\alpha$  are longer than 4Å, the interaction type “Only gamma weak” was assigned.
- 4) If all distances between NH1/NH2 atoms and the nearest oxygen (or fluorine) atoms of  $\alpha$ - and  $\gamma$ -phosphates exceed 4Å, the Arg residue was considered not to be a stimulatory finger (interaction type “none”).

The remaining cases, which did not match any of these criteria, were inspected manually (see below). After the interaction types were assigned to all structures under investigation, the interaction types were attributed to particular stimulatory patterns and their frequencies were assessed, see Table E.1.

As documented in Table E.1, Arg residues in the proximity of the phosphate chain were identified as stimulatory moieties in 981 cases. Majority of Arg fingers link  $\alpha$ - and  $\gamma$ -phosphates by their NH1 or NH2 groups and fall into NH1, “NH1 weak”, NH2, “NH2 weak” interaction types, which together are grouped into the stimulatory pattern “AG” seen in the case of 63% of all identified Arg fingers (Tables C.1, E.1, Fig. 3.5A-C). Among the structures with TS analogs, the fraction of this stimulatory pattern reaches 94%. In contrast, in complexes with ATP or GTP molecules, only 56% of interactions could be categorized in this way.

In most remaining structures, Arg fingers show interaction types “only gamma”/“only gamma weak” and interact only with oxygen atom(s) of  $\gamma$ -phosphate or their analogs (Fig. 3.5D, stimulatory pattern “G”). This stimulatory pattern was identified in 39% of complexes with ATP/GTP, 47% of complexes with non-hydrolyzable NTP analogs and only 6% of complexes with TS analogs (Table E.1).

The remaining 33 complexes, which account for 3% of all Arg fingers, did not match any of these patterns. In these 33 cases, one NH1/NH2 atom of the Arg residue forms an H-bond with  $\alpha$ -phosphate, whereas the other NH2/NH1 atom forms another H-bond with  $\gamma$ -phosphate. Further, we refer to such Y-shaped interactions as “Y-interactions” or “Y-patterns”. Since such Y-interactions are seen only in a small fraction of catalytic sites, we had inspected each of these sites manually; the results of this inspection are presented in the Appendix F. As argued and illustrated in the Appendix F, there are reasons to consider all cases of Y-interactions as structure determination/crystallization artifacts of diverse nature.

Since the guanidinium group of Arg residues can donate several H-bonds, further H-bonds are seen between amino groups of the Arg finger and the oxygen atoms of the  $\gamma$ -phosphate (or its mimicking group). There are two types of such additional bonds: formed by the NE atom and formed by NH1/NH2 groups not involved in the main stimulatory interaction, as exemplified by Fig. 3.5.

The NE atoms of Arg fingers are often located at the H-bond distances from the  $\gamma$ -phosphate. Such interactions are documented for 10% of all Arg fingers, both for those Arg fingers that interact only with the  $\gamma$ -phosphate (Fig. 3.5D), and for those fingers that coordinate both  $\alpha$ - and  $\gamma$ -phosphate with the NH2 atom (Fig. 3.5A).

An additional H-bond can be formed also by an NH1/NH2 atom which is not involved in the main stimulatory interaction. Usually, this occurs when one NH1/NH2 atom coordinates both  $\alpha$ - and  $\gamma$ -phosphates; in 51% of such complexes, the other atom (NH2/NH1 correspondingly) forms a H-bond with  $\gamma$ -phosphate (Fig. 3.5B, C). This interaction is particularly common in complexes with TS analogs (77%). Finally, when the Arg finger interacts only with  $\gamma$ -phosphate, it can accept H-bonds both from NH1 and NH2 atoms, as observed in 13% of complexes with such interaction pattern (Table E.1). In these cases, the longer H-bond was categorized as the “auxiliary” interaction, see Tables C.1 and E.1 for the complete data set.

Overall, the NH2/NH1 groups of Arg fingers that interact only with  $\gamma$ -phosphate or its analog are often assisted – in 40% of such complexes additional bonds are provided by the second NH1/NH2 atom, NE atom, or an additional Arg/Lys finger (Fig. 3.5D). In this case, one can speak about the stimulatory pattern  $G_{\text{multi}}$ . The Arg fingers in the AG position can also receive assistance (Fig 3.5C). For example, in the FtsK DNA translocase structure (PDB ID 6T8B, chain C [198]), the Arg residue interacts with  $\alpha$ - and  $\gamma$ -phosphates via the NH2 atom, while the Lys finger reaches the  $\gamma$ -phosphate. Arg fingers reaching only  $\gamma$ -phosphate often contact residues involved in the coordination of  $W_{\text{cat}}$  (see Fig. 3.5C, D).

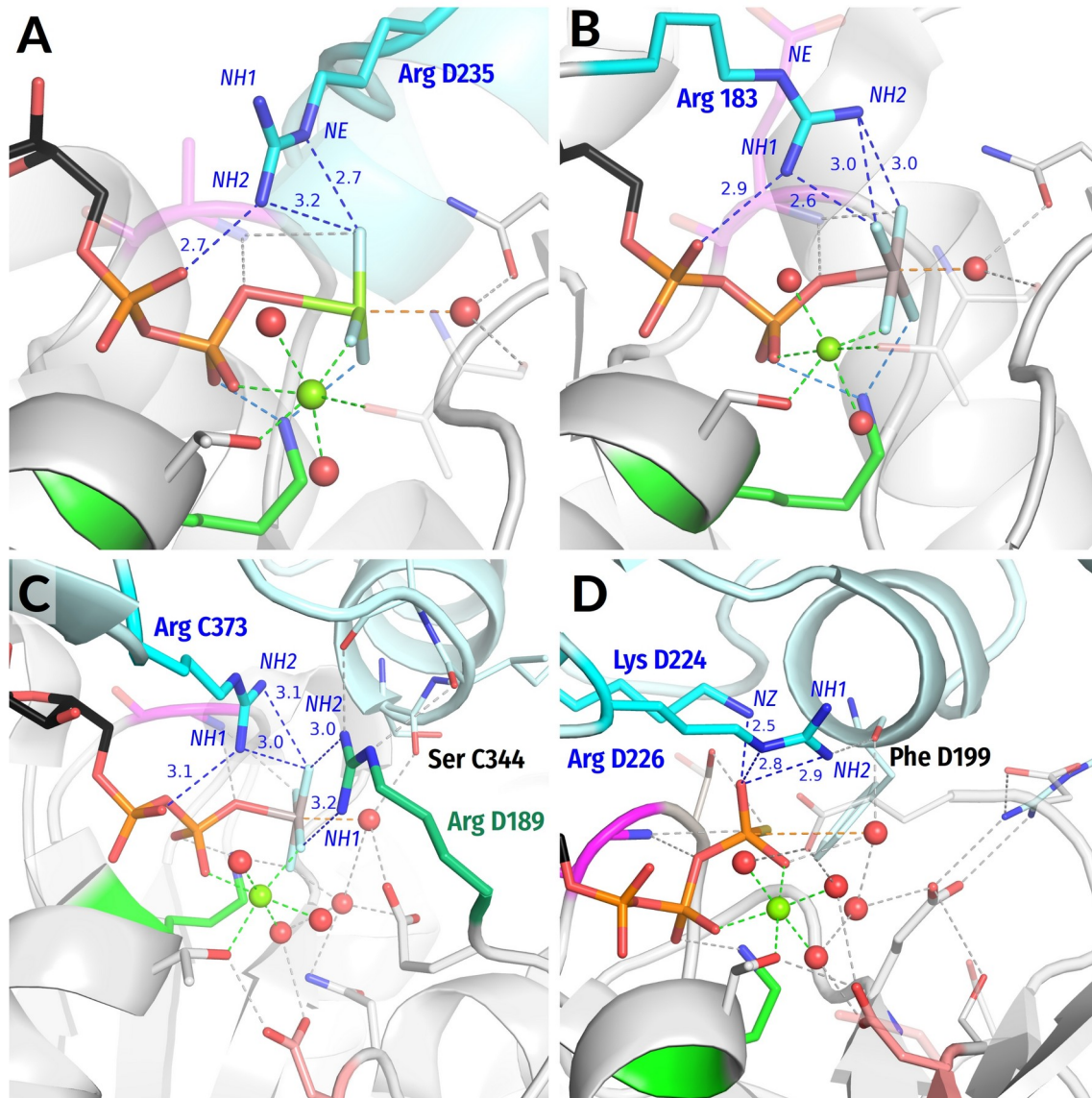


Figure 3.5: Examples of different interaction types/stimulatory patterns involving Arg residues.

Protein fragments including P-loop domain are shown as grey cartoons, other subunits carrying Arg/Lys fingers are shown in cyan; functionally relevant residues are shown as sticks: P-loop Lys residue is shown in green, Arg and Lys fingers are shown in cyan; Mg cations are shown as green spheres. All distances are given in ångströms. **A.** Both  $\alpha$  and  $\gamma$  phosphates are coordinated by the NH2 atom, an additional H-bond is formed by NE atom (stimulatory pattern AG); the structure of the Ras-like GTPase RhoA (PDB ID 5HPY, chain B [375]); **B.** Both  $\alpha$  and  $\gamma$  phosphates are coordinated by NH1 atom, an additional H-bond is formed by the NH2 atom (stimulatory pattern AG); the structure of GTP-binding protein G(q) subunit  $\alpha$ , (PDB ID 5DO9, chain A [376]); **C.** Both  $\alpha$  and  $\gamma$  phosphates are coordinated by the NH1 atom, an additional H-bond with  $\gamma$ -phosphate is formed by NH2 atom (stimulatory pattern AG); the structure of bovine ATP synthase (PDB ID 1H8E [72]). An additional Arg residue from the same chain provides more interactions with  $\gamma$  phosphate. Notably, this residue also interacts with a backbone atom of the activating  $\alpha$ -subunit); **D.** Only the  $\gamma$ -phosphate is coordinated by the NH2 atom of the Arg finger, the NE atom and Lys residue provide additional H-bonds (stimulatory pattern  $G_{multi}$ ); the structure of circadian clock protein KaiC, (PDB 4TL8, chain C [156]).



### 3.3.3.2. Stimulatory patterns of Lys fingers

Lys residues were assumed to be present in AG site if both NZ-O $\alpha$  and NZ-O $\gamma$  distances are shorter than 4Å (stimulatory pattern “AG” see Fig. 3.6A and Table C.1 in Appendix C) and to interact only with  $\gamma$ -phosphate if only the second distance met the criteria (stimulatory pattern “G” in Table C.1). Otherwise, no interaction with a Lys finger was presumed (pattern “None”), see Table C.1.

Lys fingers were identified in 141 structures. One typical pattern is with the NZ atom of Lys interacting with both  $\alpha$ - and  $\gamma$ -phosphates, similarly to a K<sup>+</sup> ion in K<sup>+</sup>-dependent P-loop NTPases, cf Fig. 1.2D and [102]. Although a Lys finger interacts both with  $\alpha$ - and  $\gamma$ -phosphates in 22% of all cases (we categorize these cases as stimulatory patterns AG, Fig. 3.6A), the fraction of such interactions was as high as 84% in complexes with TS analogs (Tables C.1, E.1). When the NZ atom of Lys interacts only with the  $\gamma$ -phosphate (pattern “G”), another Arg residue is also often involved in the interaction with the  $\gamma$ -phosphate (in 78% of cases, see Fig. 3.5D). Six of inspected structures had Lys finger coordinating both  $\alpha$ - and  $\gamma$ -phosphate and an additional Arg residue in the proximity of  $\gamma$ -phosphate. All these structures are subunits of the Large T antigen (PDB ID 1SVM, in complex with ATP [199]), see Table C.1 for details.

### 3.3.3.3. Interaction patterns of Asn fingers

Asn residues were classified as stimulatory fingers when both ND2-O $\alpha$  and ND2-O $\gamma$  distances were shorter than 4Å (stimulatory pattern “AG”, see Fig. 3.6B, Table C.1). The Asn residues were found to be in contact with both  $\alpha$ - and  $\gamma$ -phosphate groups in 67 complexes. All these structures belong to myosin or kinesin families (PF00063, PF00225).

More common are auxiliary Asn residues, which were found in the proximity of  $\gamma$ -phosphates in 248 catalytic sites, in addition to the “main” stimulators, these auxiliary Asn residues are indicated in Table C.1.

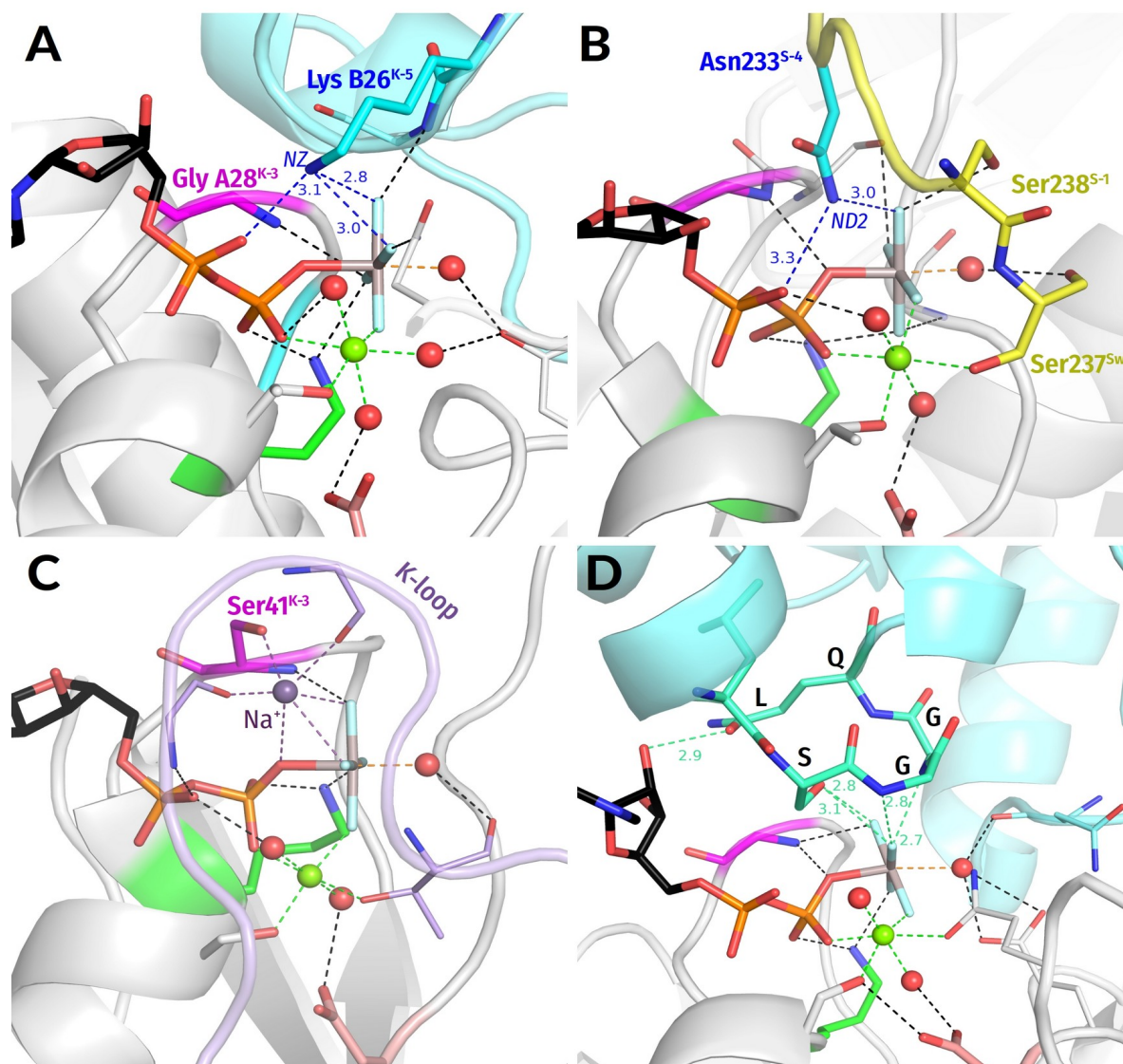


Figure 3.6: Examples of stimulatory interactions involving moieties other than arginine residues as stimulators. Residues are highlighted as in Fig. 3.5.

**A.** Interaction of a Lys finger, as provided by another subunit of the homodimer, with  $\alpha$  and  $\gamma$  phosphates; the structure of ATPase GET3, (PDB 2WOJ, chain B [374]). **B.** Asn finger in myosin II from *Dictyostelium discoideum* (PDB 1W9J). The The Asn<sup>S-4</sup> finger is provided by the Switch I motif which is shown in yellow. **C.** Na<sup>+</sup> ion as a stimulator in dynamin, (PDB 2X2E [369]). The switch I/K-loop motif is shown in lilac. Notably, a Gly residue coordinating Na<sup>+</sup> is also contacting the O<sup>2A</sup> atom of GDP. **D.** The LSGGQ motif in a structure of ABC ATPase MBP-Maltose transporter MalK, (PDB 3PUW, chain A [150]). While Ser and Gly residues are coordinating  $\gamma$ -phosphate via the backbone HN atom and the OG atom of the conserved serine residue, the Gln residue is contacting the O2' atom of GDP.

### 3.3.3.4. Summary on quantitative analysis of stimulatory interactions of Arg, Lys, and Asn fingers in P-loop NTPases

As summarized in Tables C.1, E.1 and shown in Fig. 3.5 and 3.6A-B, most of the analyzed P-loop NTPase complexes with stimulatory Arg, Lys, or Asn residue(s) positioned next to the triphosphate chain (or its structural analog) possess either a

residue providing an amino group to interact with both  $\alpha$ - and  $\gamma$ -phosphates (stimulatory pattern AG, 56,6% with Y-interactions), or (2) an amino-group-providing residue(s) forming multiple bonds with  $\gamma$ -phosphate (stimulatory pattern  $G_{\text{multi}}$ , 25,6%). In the rest of the cases (17,8%), only one amino group is contacting  $\gamma$ -phosphate (stimulatory pattern  $G_{\text{one}}$ ).

In the case of TS analogs, the fraction of catalytic sites with stimulatory pattern AG is remarkably high: all structures with  $\text{ADP-VO}_4^{3-}$  and  $\text{NDP:MgF}_3^-$  that possess Arg, Lys or Asn as a stimulatory residue have it interacting with both  $\alpha$ - and  $\gamma$ -phosphates. The same interaction is observed in 75% of  $\text{NDP:AlF}_4^-$  complexes whereas 22% of them display the stimulatory pattern  $G_{\text{multi}}$ . Only 3% of  $\text{NDP:AlF}_4^-$  complexes show stimulatory pattern  $G_{\text{one}}$  with a single amino group contacting  $\gamma$ -phosphate.

The additional data on diverse kinds of auxiliary residues additionally stabilizing the negatively charged  $\gamma$ -phosphate and/or  $W_{\text{cat}}$  in all the studied structures are presented in Tables C.1.

### 3.3.3.5. Stimulatory interactions in ABC ATPases

Activation of ABC ATPases appears to be triggered by the transported substrate and accompanied by dimerization of P-loop domains. Upon dimerization, each monomer, instead of an amino acid “finger”, inserts a whole signature motif LSGGQ into the catalytic pocket of the other monomer (Fig. 3.6D). Some soluble ABC ATPases have a non-canonical signature motif (e.g. CSAGQ in Rad50 [200] and xSTFx in MutS [201]). In some cases, the last residue of the motif is Glu or even Trp [202]. Thus, the serine residue is the most conserved member of the signature motif.

Crystal structures with TS analogs bound were obtained for the maltose transporter complex, see Fig. 1.8D and [150]. More recently, structures were solved by cryo-EM for heterodimeric ABC exporter TmrAB of the *Thermus thermophilus*, which consists of multi-drug-resistance proteins A and B [151]. In the  $\text{ADP:AlF}_4^-$ - and  $\text{ADP-VO}_4^{3-}$ -containing structures of maltose transporter complex, the side chain of serine and HN of the second glycine residue of the signature motif are at  $< 3.0 \text{ \AA}$  from the atom that is the counterpart of the  $\text{O}^{3\text{G}}$  atom of  $\gamma$ -phosphate. In the  $\text{ADP-VO}_4^{3-}$ -containing structures of the TmrAB exporter, the distance between the counterpart of the  $\text{O}^{3\text{G}}$  atom of  $\gamma$ -phosphate and signature motif serine is longer, but still  $< 4 \text{ \AA}$  [151].

### 3.3.3.6. Modes of interactions of the stimulators with phosphate chain

As shown above, P-loop NTPases exhibit many different stimulatory patterns (see Fig.1.2-1.9,3.5,3.6 and Tables C.1, E.1) involving diverse atoms of stimulators and different phosphate groups. These interactions can play an instrumental role in twisting the  $\gamma$ -phosphate group upon catalytic transition. Still, our comparative structural

analysis with emphasis on TS-like structures showed that highly diverse stimulatory moieties interact with the triphosphate chain only in two distinct ways that may complement each other, as discussed below.

#### Linking of $\alpha$ - and $\gamma$ -phosphates by the stimulator

In most classes of P-loop NTPases, at least one stimulator, provided either by the same P-loop domain or by another domain/protein, gets inserted between  $\alpha$ - and  $\gamma$ -phosphates, which implies the possibility of simultaneous interaction with  $O^{2A}$  and  $O^{3G}$  atoms (stimulatory pattern AG). As seen in Fig. 1.2, 1.6-1.9, 3.5A-C, 3.6A-B, in many such cases, the stimulator in the AG site is complemented by a second auxiliary stimulator (finger) that interacts with  $\gamma$ -phosphate, see also Tables C.1, E.1. The AG pattern was observed in 56.6% of cases.

Quantification of the interaction types for computationally analyzed catalytic sites (Tables C.1, E.1) showed that the AG stimulatory pattern dramatically prevailed in the structures containing TS analogs. This pattern was observed in all structures with ADP- $VO_4^{3-}$  and NDP:MgF $_3^-$  and in 75% of NDP:AlF $_4^-$ -containing structures. The fractions of this stimulatory pattern in structures containing ATP, GTP, or their non-hydrolyzable analogs are smaller (Table E.1). In many cases, the Arg residue is in the AG-site in the presence of a TS analog but "outside" when the substrate or its analog are bound. Apparently, catalytic sites of P-loop NTPases constrict additionally in the transition state, as it has been earlier shown for myosin [203].

Earlier, MD simulations showed that  $\gamma$ -phosphate, after its twisting by the stimulatory  $K^+$  ion, was stabilized by a new H-bond between  $HN^{K-3}$  and  $O^{2G}$  [102]. Our structural analysis strongly indicates that the counter-clockwise twist of  $\gamma$ -phosphate correlates with the formation of a new H-bond between the backbone of the P-loop and  $O^{2G}$  indeed. Specifically, H-bond-compatible distances between  $HN^{K-3}$  and the nearest oxygen/fluorine atom are seen in most structures with TS analogs bound (Fig. 3.3, Table C.1). Twisted  $\gamma$ -phosphates and H-bond compatible  $HN^{K-3} - O^{2G}$  distances are also seen in diverse NTP-containing proteins which were crystalized after being trapped in their pre-transition configurations (because of mutated  $W_{cat}$ -stabilizing residues and/or their impaired interaction with  $W_{cat}$ , see Fig. 3.4A-C, Table C.1).

Hence, linking the  $O^{2A}$  and  $O^{3G}$  atoms by the stimulator causes a counter-clockwise twist of  $\gamma$ -phosphate that correlates with the formation of a new H-bond between the  $O^{2G}$  atom (or its counterpart in NTP analogs) and the backbone  $HN^{K-3}$  group of the P-loop.

#### Interaction of the stimulator only with $\gamma$ -phosphate

In the remaining 43,3% of P-loop NTPase structures with determined stimulatory pattern, the stimulators interact only with  $\gamma$ -phosphate (stimulatory pattern G). In most such structures (25,6%),  $\gamma$ -phosphate is involved in several interactions with distinct

amino groups of an Arg finger and/or with auxiliary stimulators (stimulatory pattern  $G_{\text{multi}}$ ), see Fig. 3.5D, 3.6D and [204].

In 17.7% of structures, our computational analysis has reported only one H-bond between the stimulatory residue and  $\gamma$ -phosphate (stimulatory pattern  $G_{\text{one}}$ ). However, in many cases, the crystal structure does not contain all the partners involved in the activation, or additional stimulatory fingers are present but are too remote because of crystallization artifacts. In fact, many catalytic sites exhibiting a  $G_{\text{one}}$  pattern have counterparts with a richer network of H-bonds around  $\gamma$ -phosphate in other subunits of homooligomers of the same protein (47.6%) or in proteins belonging to the same Pfam domain (75%). Consequently, the value of 17.7% should be at least halved.

In sum, our structural analysis shows that in all cases where stimulators reach only  $\gamma$ -phosphate, these stimulators, independently on whether they are Arg or Lys residues,  $K^+$  or  $Na^+$  ions, or signature motifs of ABC ATPases, are in position to twist or pull the  $\gamma$ -phosphate group.

### 3.3.4. Shared features in P-loop NTPase stimulation

Our analysis shows that the common denominator of stimulatory patterns in diverse P-loop NTPases is the mechanistic interaction of stimulators with the  $\gamma$ -phosphate group. As summarized in Fig. 3.7, this kind of interaction is observed in TS-like structures of P-loop NTPases of different classes (see also Figure 1.2D,E and Figures 1.4–1.9, 3.5, 3.6).

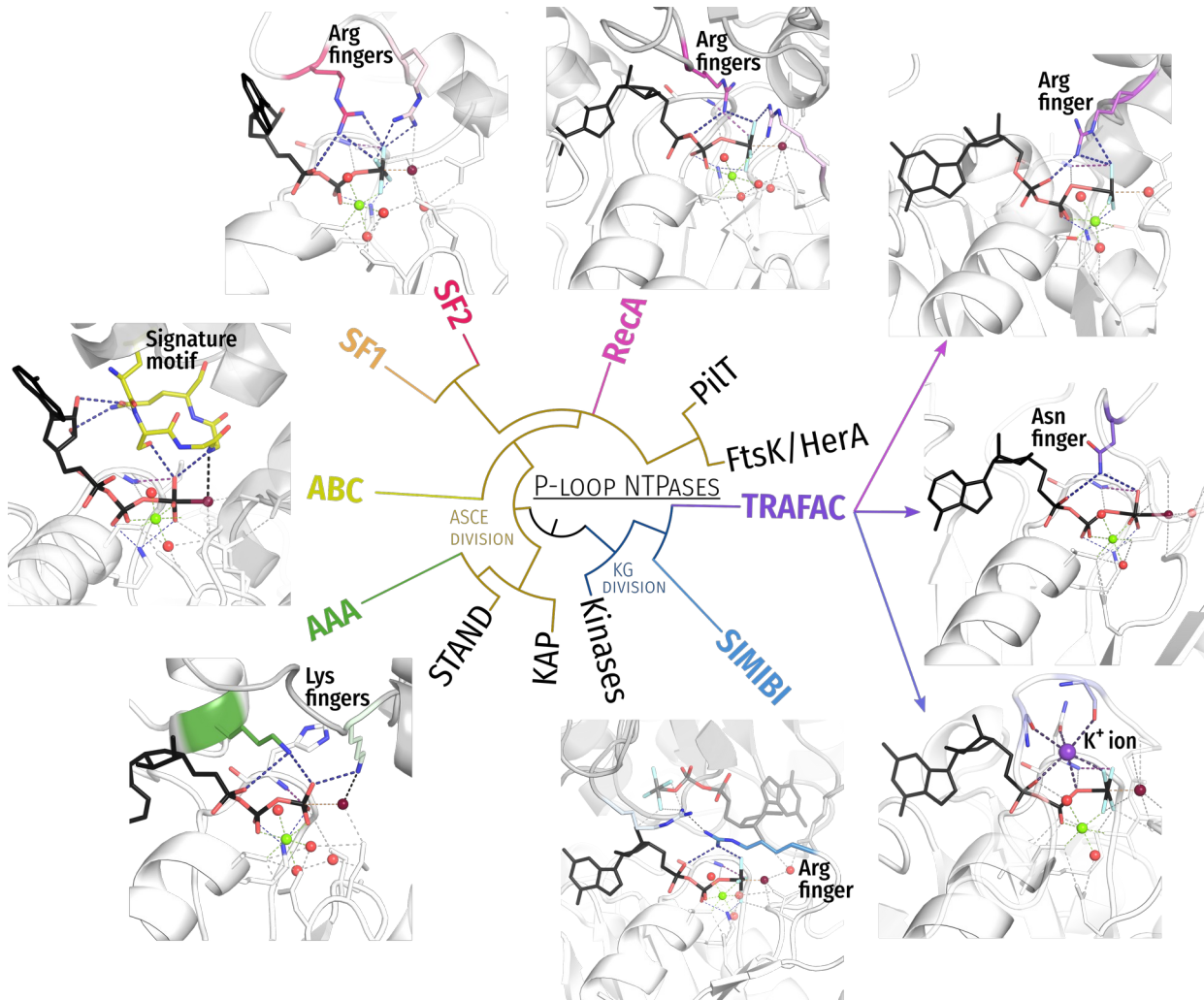


Figure 3.7: Stimulators mechanically interact with the  $\gamma$ -phosphate in different classes of P-loop NTPases. As shown, the arginine, lysine, or asparagine “fingers”, potassium or sodium ions, as well as signature amino acid motifs serve as stimulators in the various classes of P-loop NTPases. The only shared feature of these stimulators turned out to be their mechanistic interaction with the oxygen atoms of  $\gamma$ -phosphate group, capable of causing its rotation. One of these oxygen atoms is a ligand of the cofactor  $Mg^{2+}$  ion. The rotation appears to distort its coordination shell and thereby may initiate hydrolysis.

### 3.4. Analysis of hydrogen bonding patterns and proton pathways in catalytic sites of P-loop NTPases

P-loop NTPases (except P-loop kinases), as many other hydrolases, are thought to use a deprotonated water molecule as a nucleophile (see Fig. 1.3 and [86]). In such enzymes, protons are usually directly involved in catalysis [57,86,205], which justifies investigation of the H-bonded networks at the catalytic centers of P-loop NTPases.

#### 3.4.1. Proton paths and water networks in the catalytic sites of P-loop NTPases

The  $W_{\text{cat}}$  molecule should be deprotonated on some stage of the catalytic cycle, see Fig. 1.3 and [34,51,52,54,57,68,74,85,87,94]. However, the  $pK_a$  value (hereafter  $pK$  for

simplicity) of water is 14.0, which hinders the deprotonation both thermodynamically and kinetically.

The P-loop NTPases, in general, are not alone in their need to deprotonate a water molecule and to use the resulting hydroxyl for hydrolyzing the O-P bond. In various other, evolutionarily unrelated Mg-dependent enzymes,  $Mg^{2+}$  ions are thought to decrease the pK value of potential nucleophilic water molecules [86,206–209]. In such cases, usually, the would-be nucleophilic water molecule belongs to the first or second coordination shell of the metal cation. In the case of P-loop NTPases, the  $Mg^{2+}$  ion is remote from  $W_{cat}$  and can incite its deprotonation only if there is a protonic pathway between  $W_{cat}$  and the coordination shell of  $Mg^{2+}$ . Therefore, we searched for connections between  $W_{cat}$  (or the would-be  $W_{cat}$ ) and the  $Mg^{2+}$ -coordinating ligands.

In many high-resolution structures of P-loop NTPases with bound substrate molecules or their analogs, water molecules are present in the vicinity of  $\gamma$ -phosphate and interact with the  $O^{1G}$ ,  $O^{2G}$ ,  $O^{3G}$  atoms of  $\gamma$ -phosphate, as seen in Fig. 3.8. In principle, a water molecule can interact with up to two oxygen atoms, which allows categorizing the  $\gamma$ -phosphate-bound water molecules as, for example,  $W_{12}$ ,  $W_{13}$ , or  $W_{23}$ . Most likely, one of these molecules becomes the  $W_{cat}$  upon activation. For instance, Fig. 3.8A shows a structure of RadA from *Pyrococcus furiosus* that was post-soaked with ATP, see PDB ID 4A6X and [210]. The structure clearly shows  $W_{12}$  and  $W_{13}$  and resolves two conformations of the “catalytic” Glu174 of the WB+1 strand. In both conformations, Glu174 interacts with  $W_{13}$ , the apparent would-be  $W_{cat}$  in this enzyme. However, only in the Glu174A conformation, the side chain of Glu174 connects  $W_{cat}$  with the  $Mg^{2+}$ -coordinating W3 molecule.

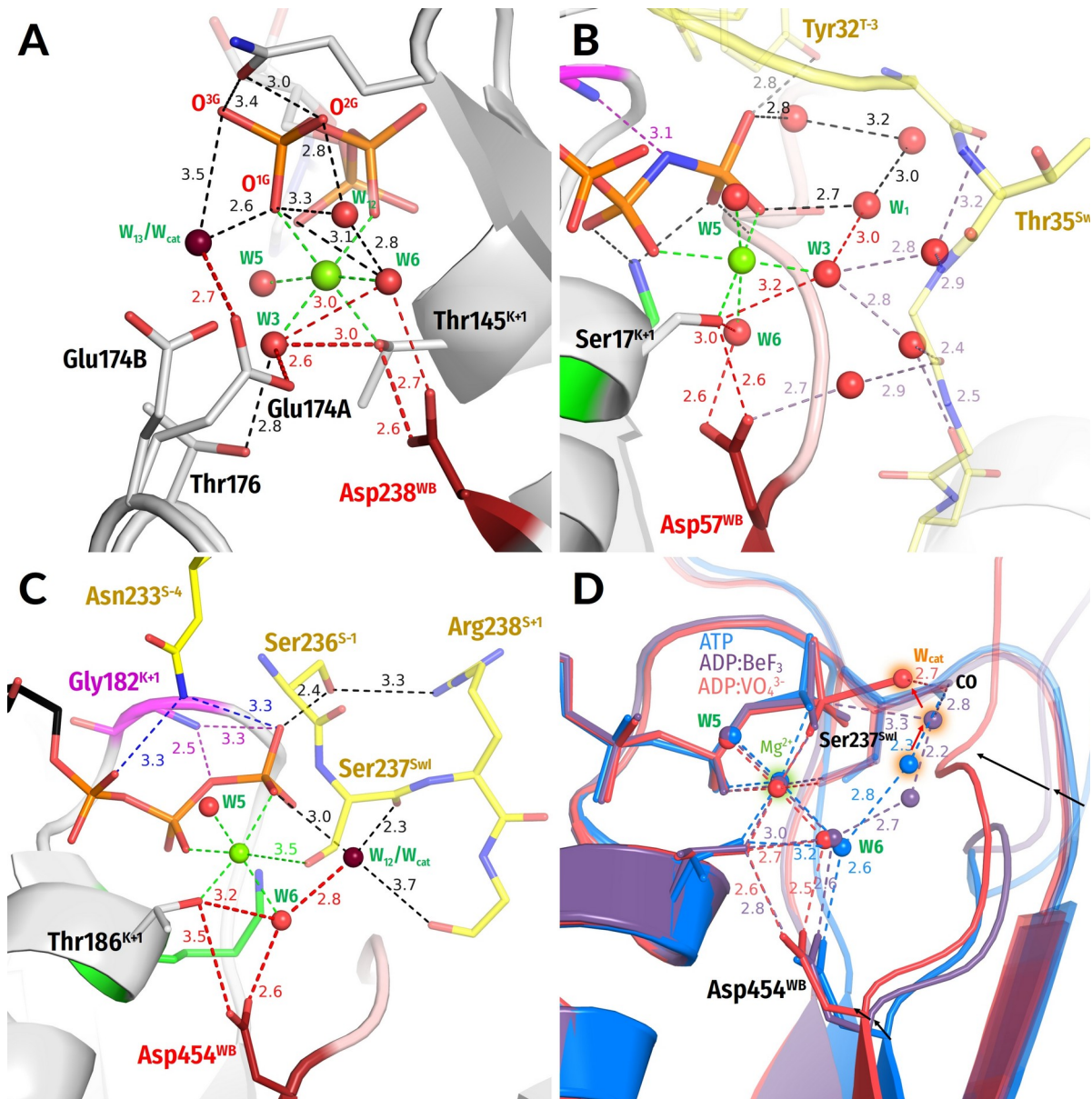


Figure 3.8: Networks of water molecules in the catalytic sites of P-loop NTPases.

**A**, ATP-soaked ATPase domain of RadA from *Pyrococcus furiosus* (PDB ID 4A6X [210]); **B**, human H-Ras GTPase (PDB ID 5B30 [211]). **C**, ATP-soaked Myosin II from *Dictyostelium discoideum* (PDB ID 1FMW [214]; Other colors as in Figures 1.6-1.9. **D**, Three overlapped crystal structures of myosin II from *Dictyostelium discoideum*, the same as in panel C; **blue**, ATP-soaked structure (PDB ID 1FMW); **violet**, ADP:BeF<sub>3</sub>-containing structure (PDB ID 1W9K); **red**, ADP:VO<sub>4</sub><sup>-</sup>-containing, TS-like structure (PDB ID 1VOM [215]); the would-be W<sub>cat</sub> water molecules are surrounded by an orange halo. Other colors as in Figures 1.6-1.9.

Similarly, as it has been repeatedly noted in Section 1.2.3, the “catalytic” Glu or Asp residues link W<sub>cat</sub> with Mg<sup>2+</sup>-ligands #3 or #6 in TS-like structures of NTPases of all classes (except the TRAFAC class). The tentative proton pathways from W<sub>cat</sub> towards the Mg<sup>2+</sup>-ligands, as inferred upon the structural analysis provided here, are shown by red dashed lines in Fig. 1.5-1.9.

In the case of TRAFAC class NTPases, which lack “catalytic” Glu/Asp residues, connections between W<sub>cat</sub> and the Mg<sup>2+</sup>-coordinating ligands are not so evident. The



water molecules around  $\gamma$ -phosphate appear to be highly mobile and hardly crystallizable. In the absence of TS analogs, those water molecules that interact with  $\gamma$ -phosphate are usually resolvable in the structures only if constrained by additional interactions. For instance, in TS-like structures of small GTPases of the TRAFAC class, the closest to  $W_{\text{cat}}$   $\text{Mg}^{2+}$ -coordinating ligand is [Ser/Thr]<sup>SwI</sup> of Switch I at about 4–5 Å (Fig. 1.2E, 1.6A-D, 3.8B). In TRAFAC proteins, however, the Switch I motif is very dynamic so that the protein fluctuates between “open” and “closed” conformations even with a bound substrate or its analogs [211–213]. While in the closed conformation [Ser/Thr]<sup>SwI</sup> serves as ligand #3 for  $\text{Mg}^{2+}$ , in the open conformation the whole Switch I moves away from  $\text{Mg}^{2+}$  and a water molecule serves as ligand #3 (Fig. 3.8B).

We have analyzed how the water network changes in response to the displacements of Switch I in a well-studied TRAFAC class protein myosin. In the ATP-soaked myosin structure (PDB ID 1FMW [214], see Fig. 3.8C),  $W_{12}$  is resolved because it is additionally bound to CO of Ser<sup>S-1</sup>. This  $W_{12}$  molecule is at a H-bond-compatible distance from the  $\text{Mg}^{2+}$ -coordinating  $W_6$  molecule. Figure 3.8D shows three overlapped structures of *Dictyostelium discoideum* myosin—namely, the red ATP-soaked myosin structure that is the same as in panel 7C (PDB ID 1FMW [214]), the violet structure with a substrate analog ADP:BeF<sub>3</sub> bound (PDB ID 1W9K), and the red TS-like structure with ADP:VO<sub>4</sub><sup>-</sup> bound (here an oxygen atom of VO<sub>4</sub><sup>-</sup> mimics  $W_{\text{cat}}$  in the apical position (PDB ID 1VOM [215])).

One can see how a water molecule, which is stabilized by CO of Ser237<sup>SwI</sup> in all three structures, “moves” towards the apical position. It is noteworthy that an additional water molecule links this would-be  $W_{\text{cat}}$  with the  $\text{Mg}^{2+}$  ligand sphere in the “intermediate” violet structure. This linking water is seen in several myosin structures and appears to be stabilized by invariant backbone groups of the D+2 residue. In the red TS-like structure, the WB-crest comes closer to the nucleotide and interrupts the water connection with the ligands of  $\text{Mg}^{2+}$  (Figure 3.8D).

Notably, Matsumoto and colleagues succeeded to control the open to close transition in H-Ras GTPases by changing the humidity [211]. The crystal structure of H-Ras that crystallized in the open conformation at low humidity (PDB ID 5B30) reveal an additional water molecule next to  $W_3$  (Fig. 3.8B). It is depicted as  $W_1$  in Fig. 3.8B and occupies approx. the same position as  $W_{12}$  in the myosin structure in Fig. 3.8C. In the closed conformation (Fig. 3.8D, 1.6A-D), Switch I reaches  $\text{Mg}^{2+}$  and ousts  $W_1$  and  $W_3$ ; only  $W_{\text{cat}}$  is present in the TS-like structure (cf Fig. 1.2E and 3.8B).

### 3.4.2. The H-bond between Asp<sup>WB</sup> and [Ser/Thr]<sup>K+1</sup> is shorter in presence of TS analogs

We have noted that the strictly conserved [Ser/Thr]<sup>K+1</sup> of the Walker A motif is often involved in a H-bond with the similarly strictly conserved Asp<sup>WB</sup>. Furthermore, we noted that this H-bond is shorter in TS analog-containing structures. This bond is particularly short in the NDP:AlF<sub>4</sub><sup>-</sup>-containing representative structures (see Fig. 1.6-1.9). In contrast, this H-bond may be even absent in structures where the AlF<sub>4</sub><sup>-</sup> moiety is improperly bound, see Section 3.2.1.1. and Table C.1. The H-bond between Asp<sup>WB</sup> and [Ser/Thr]<sup>K+1</sup> is special as it links the Walker A and Walker B motifs.

Therefore, we measured the distances between the two residues to characterize their H-bonding in all the available structures of P-loop NTPases with bound substrates or their analogs, as described in the Methods section. The results of these measurements for all the analyzed structures are presented in Table C.1. Fig. 3.9 summarizes the data on best resolved structures (with resolution of 2.5 Å or less). The data for the three TS analogs are plotted separately for the sake of clarity. We consider as short the H-bonds of < 2.7 Å [216]; the range of short H-bonds is highlighted by cyan.

As documented in Section 3.2.1.1, improper binding of AlF<sub>4</sub><sup>-</sup> can distort the coordination sphere of Mg<sup>2+</sup>. Therefore, for the plot in Fig. 3.9, we selected only the sites with a correct Mg<sup>2+</sup> coordination; the cut-off for the distance between Mg<sup>2+</sup> and [Ser/Thr]<sup>K+1</sup> was 2.5 Å. When comparing only TS analogs, fluoride complexes were superior to vanadate complexes in their ability to shorten Asp<sup>WB</sup>-[Ser/Thr]<sup>K+1</sup> bonds, as can be judged from Fig. 3.9. From the same data it follows that the NDP:AlF<sub>4</sub><sup>-</sup> complexes are characterized by shortest Asp<sup>WB</sup>-[Ser/Thr]<sup>K+1</sup> bond lengths among metal fluorides.

As seen in Fig. 3.9, the length of the H-bond between Asp<sup>WB</sup> and [Ser/Thr]<sup>K+1</sup> is, on average, < 2.7 Å in all groups of structures and < 2.6 Å in structures with ATP/GTP, non-hydrolyzable NTP analogs and NDP:MgF<sub>3</sub><sup>-</sup> or NDP:AlF<sub>4</sub><sup>-</sup> as TS analogs. Specifically, in NDP:AlF<sub>4</sub><sup>-</sup>-containing structures the H-bond distance is on average as short as 2.5 Å.

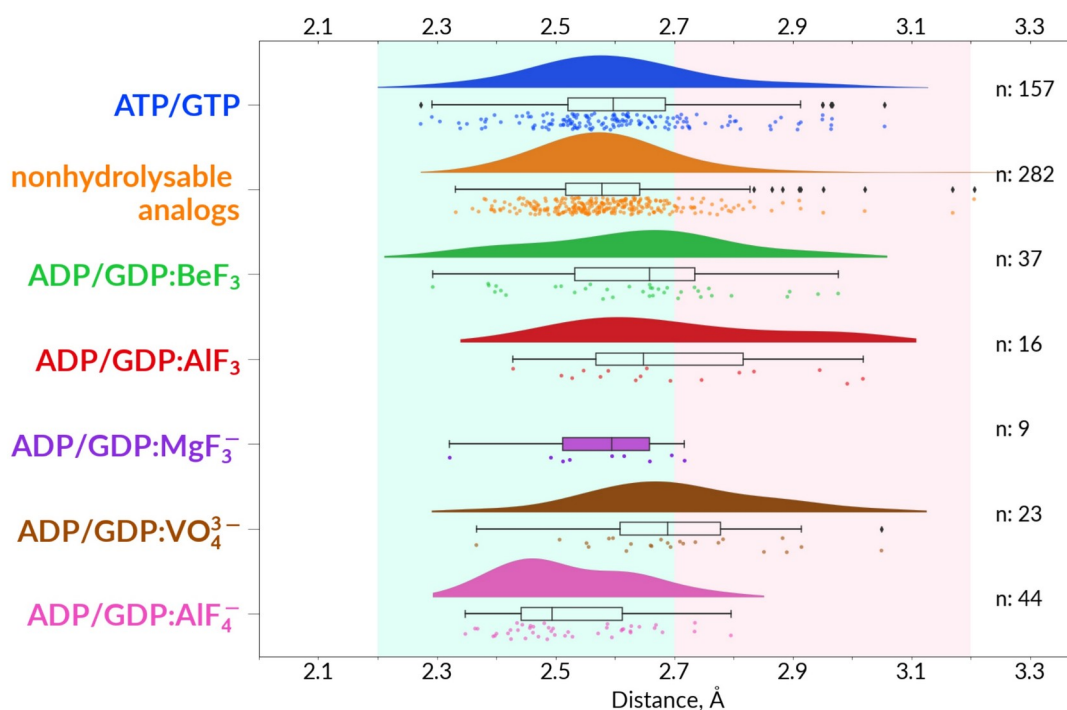


Figure 3.9: Asp<sup>WB</sup>-[Ser/Thr]<sup>K+1</sup> distances in high-quality P-loop NTPase structures with bound Mg-NTP molecules or their analogs.

Resolution threshold 2.5 Å; see Table C.1 for the entire data set. For each type of complexes, distances are visualized as a kernel density estimate (KDE) plot, a boxplot and individual data points, each point representing one catalytic site in one structure. For ADP/GDP:MgF<sub>3</sub><sup>-</sup> complexes the density plot is not shown owing to the scarcity of data. The integrity of catalytic sites was checked by two criteria: the [Ser/Thr]<sup>K+1</sup> - Mg<sup>2+</sup> distance ≤ 2.5 Å and Mg<sup>2+</sup> - Asp<sup>WB</sup> distance ≤ 6 Å. The range of short H-bonds (2.2 Å-2.7 Å) is highlighted in cyan, the typical H-bond range (2.7 Å-3.2 Å) in pink.

### 3.4.3. The solvent accessible surface area of Asp<sup>WB</sup> is diminished in TS-like structures

It is well known that short H-bonds readily form between buried amino acid residues [205]. Therefore, we determined the solvent accessible surface area (SASA) of Asp<sup>WB</sup> for the same structures for which the Asp<sup>WB</sup>-[Ser/Thr]<sup>K+1</sup> bond lengths were measured. As it can be seen in Figure 3.10, Asp<sup>WB</sup> is well buried in the case of TS-like structures, its SASA values are <6%.

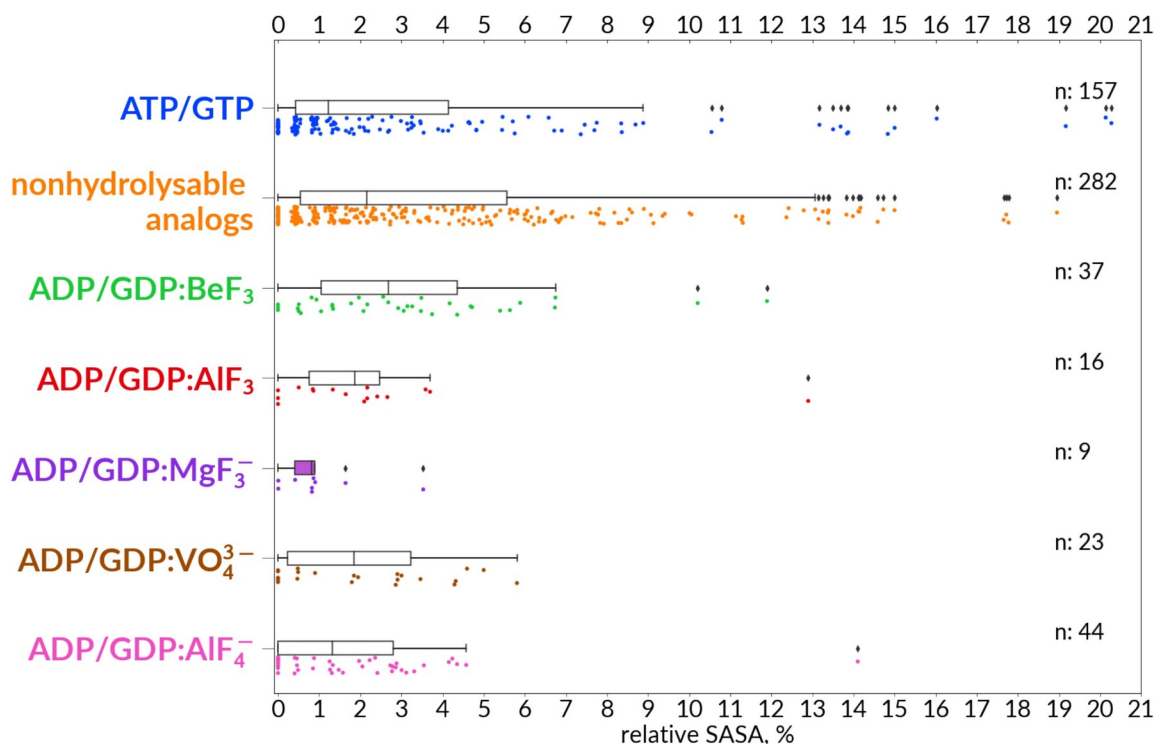


Figure 3.10: Relative solvent-accessible surface area (SASA) values for Asp<sup>WB</sup> residues. The calculated SASA values are shown for the same catalytic sites as shown in Fig. 3.9.

### 3.4.4. Proton path connects Asp<sup>WB</sup> with $W_{\text{cat}}$ in all classes of P-loop NTPases

As noted in Section 3.4.1., our analysis indicates that  $W_{\text{cat}}$  is connected by a distinct proton pathway to one of the Mg<sup>2+</sup> ligands in all TS-like structures (Fig. 1.7-1.9) except those of TRAFAC class NTPases, where such a protonic connection is not evident (Fig. 1.6). For the latter, however, a transiently formed proton path between  $W_{\text{cat}}$  and the coordination shell of Mg<sup>2+</sup> could be inferred from the comparative analysis of myosin structures with different substrates or their analogs bound, see Fig. 3.8C-D. Noteworthy, the distance between the neighboring ligands of Mg<sup>2+</sup> is 2.9 Å, which enables proton exchange between them. In Section 3.4.2., also building upon results of our analysis, we argue that the Mg<sup>2+</sup>-coordinating [Ser/Thr]<sup>K+1</sup> is H-bonded to Asp<sup>WB</sup> in the TS-like structures with “properly” bound AlF<sub>4</sub><sup>-</sup> moiety (as well as in most other structures of P-loop NTPases with bound substrates or their analogs, see Fig. 3.9 and Table C.1 in Appendix C). Together, these observations point to the existence of a protonic connection from  $W_{\text{cat}}$  to Asp<sup>WB</sup> via [Ser/Thr]<sup>K+1</sup> in the TS; the respective proton pathways are depicted in Fig. 3.11 for the main classes of P-loop NTPases.

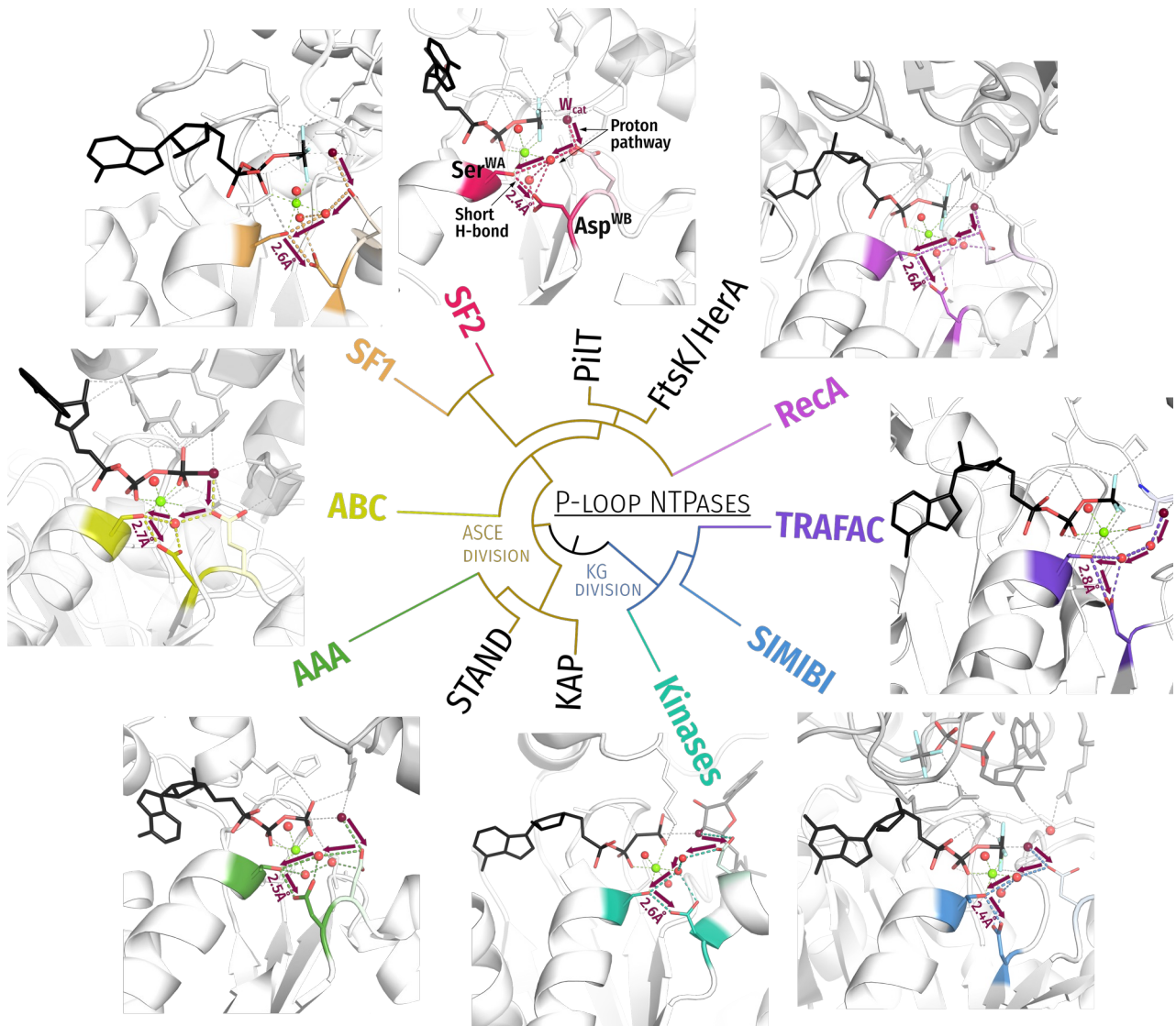


Figure 3.11: Putative protonic connections from  $W_{cat}$  to  $Asp^{WB}$  via short H-bonds between  $[Ser/Thr]^{K+1}$  and  $Asp^{WB}$  can be seen in different classes of P-loop NTPases.

The arrows show putative proton pathways that we have found to connect the catalytic water molecule  $W_{cat}$  with the strictly conserved  $[Ser/Thr]^{K+1}$  of Walker A motif in transition-state-like structures of ATPases and GTPases of different classes. In these structures,  $[Ser/Thr]^{K+1}$  is linked with the strictly conserved  $Asp^{WB}$  of Walker B motif by a short hydrogen bond.

## 4. Discussion

### 4.1. Common features as observed for TS analog-containing structures

In the Results section, we presented some findings from our comparative structure analysis of 3136 catalytic sites of P-loop ATPases and GTPases with substrates or their analogs bound. The entire wealth of quantitative data obtained is presented in the Table C.1 in Appendix C.

In Section 4.1. below, we discuss in more detail some of the common features of P-loop NTPases. We focus particularly on the structures with TS analogs bound as being most informative concerning the enzyme mechanisms. In the following Section 4.2 we build upon the data obtained to infer the tentative common mechanism of P-loop fold NTPases.

#### 4.1.1. Comparison of metal fluorides as TS analogs

##### 4.1.1.1 Non-physiological binding of $\text{AlF}_4^-$ to the cofactor $\text{Mg}^{2+}$ ion

Upon inspection of  $\text{NDP}:\text{AlF}_4^-$  containing structures, we have noted that the  $\text{Asp}^{\text{WB}} - [\text{Ser/Thr}]^{\text{K}+1}$  distances are too long (of  $\geq 4.0 \text{ \AA}$ ) in some of them (see Fig. 3.1B,D and Table C.1). Further inspection of these structures, however, has shown that  $\text{AlF}_4^-$  moiety is “improperly” bound in these structures; it makes two bonds with  $\text{Mg}^{2+}$  (Table C.1). When  $\text{AlF}_4^-$  makes two bonds with  $\text{Mg}^{2+}$ , the bond between  $\text{Mg}^{2+}$  and  $[\text{Ser/Thr}]^{\text{K}+1}$  can get lost and the whole catalytic site undergoes distortion so that Walker A and Walker B motifs stay apart, see Fig. 3.1 B,D. In these cases, (all marked pink in Table C.1), the structure of the catalytic site cannot be considered as analogous to the TS.

After we excluded these cases, the configurations of the remaining catalytic sites were found to be the same and the  $\text{Asp}^{\text{WB}} - [\text{Ser/Thr}]^{\text{K}+1}$  distances were found to be H-bond compatible in all TS analog-containing structures (about 200) but one, which is depicted in Fig. A.4 and discussed in its caption (Appendix A). Hence,  $\text{AlF}_4^-$ -containing P-loop NTPases can be used as reliable models of transition states, in support of earlier suggestions [68,71,75,96], provided that the correctness of the interaction between  $\text{AlF}_4^-$  and  $\text{Mg}^{2+}$  is checked in each particular case.

Although the discovery of twisted  $\text{AlF}_4^-$  was rather unpleasant as such, it provided some useful information. The “wrong” binding of  $\text{AlF}_4^-$  was accompanied by its counterclockwise rotation as compared to “properly bound”  $\text{AlF}_4^-$  moieties, whereby the interactions of the fluoride atoms with  $\text{Lys}^{\text{WA}}$  and the stimulator were preserved. It

could be inferred that the ligands of  $\gamma$ -phosphate oxygens are adapted to the twist of the  $\gamma$ -phosphate group.

#### 4.1.1.3. Untypical binding of $\text{AlF}_4^-$ in RecA ATPases

Another case where the orientation of the  $\text{AlF}_4^-$  differs both from those observed in most other TS-like structures and of the orientation of  $\gamma$  phosphate is the case of RecA ATPases and, specifically, DnaB SF4 helicases, as considered in Section 3.2.1.2.

The subunits of DnaB helicases, depending on the presence of substrate analogs and their nature, can either form rings of distinct shapes [146,217–220] or arrange themselves as a hexameric ladder along a DNA strand [162,221]. The latter type of the structure was reported for DnaB from *Geobacillus stearothermophilus*, which was crystalized, in the presence of a DNA strand, with  $\text{GDP}:\text{AlF}_4^-$  in five of its six catalytic sites [162]. In this structure, each monomer of DnaB interacts in a similar way with two nucleotides of DNA; together, the subunits make a kind of a spiral ladder. It is noteworthy, that the position of  $\text{AlF}_4^-$  in the structure of *Geobacillus stearothermophilus* DnaB (Fig. 4.1) differs from that in other P-loop fold NTPases. No catalytic water molecule is present apically to the plane of  $\text{AlF}_4^-$  (see Figure 4.1A as a typical example), and the position of the  $\text{AlF}_4^-$  moiety does not correspond to that of the  $\gamma$ -phosphate group; the  $\text{AlF}_4^-$  moieties are distinctly tilted (Fig. 3.2). Concurrently, the NMR data on DnaB from *Helicobacter pylori* point to a full occupation of all six catalytic sites by  $\text{AlF}_4^-$  moieties capable of free rotational diffusion [159,195].

The uniform spectra of  $\text{AlF}_4^-$  in DnaB from *Helicobacter pylori*, however, comes in contradiction with the sequential operation of catalytic subunits, as observed in several well-studied oligomeric P-loop ATPases [142,217,222–224]. Their subunits operate one after another so that the catalysis in one subunit is thermodynamically promoted by the substrate binding to the other subunit [142,223]. Correspondingly, neighbouring P-loop domains are in different conformations in any moment. Based upon comparative structural analysis of diverse RecA ATPases, the unusual behavior of  $\text{AlF}_4^-$  moieties in DnaB helicases might be explained by a scenario in which the catalytic sites sequentially get trapped in their post-transition states, as depicted in Fig. 4.1, and considered in more details elsewhere [159].

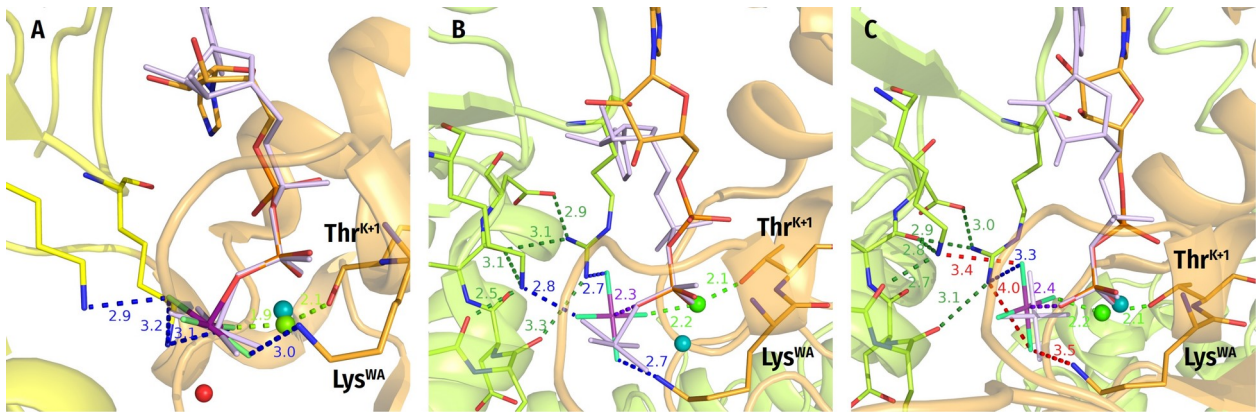


Figure 4.1: The transition-state analog  $\text{AlF}_4^-$  can adopt different orientations in diverse P-loop ATPases of the RecA class.

**A**  $\text{AlF}_4^-$  binding in RecA from *E. coli* (PDB ID 3CMW [253]). The P-loop domain is shown in orange, the P-loop domain of the adjacent activating subunit that provides the stimulating “fingers” is colored yellow. The  $\text{Mg}^{2+}$  ion is shown as a green sphere. To show the catalytic water molecule, the  $\text{ADP}:\text{AlF}_4^-$  complex is superimposed with the structure of the  $\text{ADP}:\text{AlF}_4^-:\text{W}_{\text{cat}}$  complex from the ABC ATPase of the maltose transporter MalK (PDB ID 3PUW [150]). The ADP molecule and  $\text{AlF}_4^-$  of MalK are shown in lilac,  $\text{W}_{\text{cat}}$  as a red sphere,  $\text{Mg}^{2+}$  as a teal sphere. ADP moieties were superimposed by atoms  $\text{O}^{3\text{A}}$ ,  $\text{P}^{\text{B}}$  and  $\text{O}^{3\text{B}}$  (see Fig. 1.2 for the atom notation used for ADP) in Pymol; **B** Coordination of  $\text{AlF}_4^-$  in the DnaB from *Geobacillus stearothermophilus* structure (PDB ID 4ESV [162]). The orange, nucleotide-binding chain and the green activating chain correspond to the chains C and B of the 4ESV. To show the displacement of  $\text{AlF}_4^-$ , the  $\text{GDP}:\text{AlF}_4^-$  complex is superimposed, as described for panel A, with  $\text{ADP}:\text{AlF}_4^-$  bound to the RecA protein (PDB ID 3CMW, see panel A). The H-bonds formed by  $\text{AlF}_4^-$  are shown in blue, the additional interactions that stabilize the position of stimulating side chains of K418 and R420 are shown in green. The  $\text{AlF}_4^-$  moiety is twisted in comparison to the transition-state-mimicking complex shown on panel A; **C** Coordination of  $\text{AlF}_4^-$  in the structure of DnaB, different interface. The orange, nucleotide-binding chain and the green activating chain correspond to the chains F and E of the PDB 4ESV [162]. To show the further displacement of  $\text{AlF}_4^-$ , the  $\text{GDP}:\text{AlF}_4^-$  complex of subunit F is superimposed, as described for panel A, with the same complex bound to subunit C of the 4ESV PDB structure. Bonding interactions that are observed for the  $\text{GDP}:\text{AlF}_4^-$  complex trapped at the B/C interface (see panel B), but not in this complex trapped at the E/F interface are shown as red dashed lines.

Specifically, the exergonic binding of the first  $\text{ADP}:\text{AlF}_4^-$  moiety to a DnaB subunit (subunit 1) brings it into its catalytically active, DNA-bound configuration with the Arg (R446) and Lys (K444) fingers of the adjoining subunit 2 interacting with the  $\text{ADP}:\text{AlF}_4^-$  complex in the “catalytic” position (Fig. 4.1A). Binding of the  $\text{ADP}:\text{AlF}_4^-$  to the subunit 2 transforms it in a similar way and, simultaneously, provides free energy for pulling the  $\gamma$ -phosphate-mimicking  $\text{AlF}_4^-$  out of its catalytic position in the subunit 1—by K444 and R446 fingers of subunit 2—into one of the post-TS positions as seen in the structure of DnaB from *Geobacillus stearothermophilus*, see Fig. Fig. 4.1B,C and [162]. After this sequence of events repeats six times, all six protein subunits are in the same catalytic configuration being tightly fixed on the DNA strand (as seen from the solid NMR spectra, Fig. 3 in [159]) whereas their six  $\text{AlF}_4^-$  moieties are, most likely, in positions similar to those taken by  $\text{AlF}_4^-$  moieties in two of six catalytic sites of DnaB from *Geobacillus stearothermophilus*, namely those on the subunit interfaces B/C and F/A, see Fig. 4.1B,C, and [162]. In this state, the  $\text{AlF}_4^-$  moieties are detached both from the ADP moiety and the Arg and Lys fingers and can, potentially, freely rotate. The position of



$\text{AlF}_4^-$  resembles to some extent the position of  $\text{H}_2\text{PO}_4^-$  in several post-TS structures [1,2], which suggests that the mobile  $\text{AlF}_4^-$  moiety in DnaB from *Helicobacter pylori*, may mimic the phosphate group during a late transition state of ATP hydrolysis. To which extent the trapped transition state using a metal fluoride resembles the physiological transition state remains an open question at this stage.

Summarizing the section on improper binding of  $\text{AlF}_4^-$  moieties in some catalytic centers, we would like to note that the finding per se was alarming. These fluoride complexes are deservedly vaunted as powerful TS analogs, and structures containing them are commonly interpreted as TS-like [68,71,96]. Still, our data show that the integrity of  $\text{Mg}^{2+}$  coordination in the presence of  $\text{NDP}:\text{AlF}_4^-$  should be evaluated separately for each enzyme structure before linking the corresponding structural data to the catalytic mechanism. Notably,  $\text{NDP}:\text{AlF}_4^-$  complexes are also used as TS analogs in other enzyme superfamilies [68,96]. It is important to check whether a similar distortion of catalytic sites by  $\text{NDP}:\text{AlF}_4^-$  complexes is possible in enzymes other than P-loop NTPases.

#### 4.1.2. Role of mechanistic bonding in the common stimulation mechanism of P-loop NTPases

Here provided information on the stimulatory interactions within hundreds of catalytic sites provides a bird's-eye view of activation in P-loop NTPases. Looking together at all types of identified stimulatory patterns provides some additional clues about the mechanisms of hydrolysis stimulation. Without challenging the previously proposed tentative stimulatory effects as referred to in the Introduction section, our structure analysis indicates that none of so far suggested mechanisms is common to all P-loop NTPases. Indeed, it is beyond doubt that the positive charges of Arg/Lys fingers or  $\text{K}^+/\text{Na}^+$  ions, by interacting with oxygen atoms of  $\gamma$ -phosphate, would make the  $\text{P}^{\text{G}}$  atom more prone to the nucleophilic attack, as suggested by Warshel and colleagues [92–94] and as calculated by Rudack et al. [99]. The positive charge of stimulators could also compensate for the negative charge that develops at  $\beta$ -phosphate upon the breakaway of  $\gamma$ -phosphate [56,57]. Nevertheless, the absence of a positive charge on the stimulatory signature motifs of ABC ATPases (Fig. 3.6D) does not stop them from triggering ATP hydrolysis. Also, expelling water molecules out of the catalytic pocket by stimulatory Arg fingers may provide an entropic gain, as suggested by Kotting and colleagues [95]. However, such effects are hardly to be expected when tiny  $\text{K}^+$  or  $\text{Na}^+$  ions act as stimulators and immobilize water molecules in the catalytic pocket (Fig. 1.2E, 3.6C). Reorientation of the  $W_{\text{cat}}$  molecule into the attack position and its polarization, as suggested by Jin and colleagues, can be realized by those Arg or Lys fingers that reach  $W_{\text{cat}}$  (Fig. 3.5D), but not by most other stimulators.

Our analysis points to the importance of mechanistic interaction of stimulators with the  $\gamma$ -phosphate group, which is the only feature shared by all inspected stimulators. Importance of this interaction is exemplified by NTPases that are stimulated by moieties with only a minute positive charge. These are ABC ATPases, where  $\gamma$ -phosphate interacts with the signature LSGG[Q/E] motif via the side chain of its serine residue and the backbone HN of the second glycine residue, see Fig. 3.6D and [150]. Other examples are the kinesin and myosin families, where the Asn finger inserts its NH group between  $\alpha$ - and  $\gamma$ -phosphates, see Fig. 3.6B and [130–132]. It is unlikely that small partial electric charges of Ser or Asn side chains could be decisive for catalysis in these cases; rather, their mechanistic H-bonding to the  $O^{3G}$  atom appears to be the key.

The mechanistic nature of the stimulatory interaction is consistent with the predominance of Arg residues as stimulators (Tables C.1, E.1). First, a guanidinium group could donate up to three H-bonds for interaction with the oxygens of triphosphate. Second, the strength of H-bonds between guanidinium groups and phosphate anions has been shown to be comparable to that of covalent bonds [225–227].

The importance of mechanistic binding rationalizes the preference for multiple stimulatory fingers (Table E.1), as well as the choice of the stimulatory signature motif by omnipresent ABC ATPases. This motif is electrically neutral but donates several H-bonds to the  $O^{3G}$  atom.

The notable feature is the apparent scarcity – if not complete absence – of Y-patterns with NH1 and NH2 groups of an Arg finger separately interacting with  $\alpha$ - and  $\gamma$ -phosphates, respectively (see Table C.1). The Y-pattern is not observed in a single structure with a bound TS analog, and it is such structures that enable us to judge with certainty the stimulatory pattern in a particular ATPase. Our analysis has shown that the few structures with the Y-pattern are likely to be artifacts either of crystallization or of structure determination, as substantiated in the Appendix F.

Outside of P-loop NTPases, however, the Y-pattern of Arg interaction is very common, especially in protein-DNA complexes, where one Arg residue often donates its NH1 and NH2 groups to neighboring phosphate groups of the DNA backbone [228,229]. In the case of P-loop NTPases, however, a Y-linked Arg residue would fix the  $O^{2A}$  and  $O^{3G}$  atoms approx. 6.1 Å apart and prevent the twist of  $\gamma$ -phosphate. The apparent absence of Y-patterns in the examined structures of P-loop NTPases can be seen as further evidence in favor of the  $\gamma$ -phosphate twist as the key pre-catalytic configuration change in P-loop NTPases.

In most cases, other auxiliary positively charged groups, such as HN groups and/or additional Arg/Lys residues, are involved in the coordination of the oxygen atoms of  $\gamma$ -phosphate and/or  $W_{cat}$  in addition to the “main” stimulator. This is observed in AAA+ NTPases (Fig. 1.8A), many helicases (Fig. 1.2D, 1.8B,C), ABC NTPases (Fig. 1.8D),

hexamers of F<sub>1</sub>/RecA-like ATPase (Fig. 1.9), PilT-like proteins [109]. These auxiliary residues, indicated for the various classes of P-loop NTPases in Tables B.1 and C.1, are poorly conserved even within individual families of P-loop NTPases.

Still, the common denominator of stimulatory patterns in diverse P-loop NTPases is the mechanistic interaction of stimulators with the  $\gamma$ -phosphate group; this interaction is observed in all analyzed TS-like structures (Fig. 1.2D,E, 1.4-1.9, Tables C.1, E.1) and can be inferred from many other structures, specifically those that can be related to post-transition states, see [159].

While linking of O<sup>2A</sup> and O<sup>3G</sup> atoms by a stimulator enforces a counterclockwise twist of  $\gamma$ -phosphate, it is not clear yet, what conformational changes are caused by stimulators that interact only with  $\gamma$ -phosphate. As it follows from comparative structure analysis (Section 3.2.1.2),  $\gamma$ -phosphate may be twisted clockwise in RecA NTPases, see also [159,195]. Even interacting only with  $\gamma$ -phosphate, the stimulator is often located between  $\alpha$ - and  $\gamma$ -phosphates, as in dynamins (Fig. 3.6C) or ABC transporters (Fig. 3.6D) and connected to the “head” of the NTP molecule. For instance, the Na<sup>+</sup> ion in dynamins, while not reaching the  $\alpha$ -phosphate directly, is connected to it via two noncovalent bonds (Fig. 3.6C). The signature motif of ABC ATPases is H-bonded via conserved Ser and Gln residues to the O2' atom of the ribose (Fig. 3.6D). Such connectivity may strengthen the mechanistic impact on  $\gamma$ -phosphate.

Mechanistic interaction with the  $\gamma$ -phosphate group may promote hydrolysis in different ways; for instance, it may destabilize the O<sup>3B</sup> – P<sup>G</sup> bond and/or make the triphosphate chain almost fully eclipsed, and/or facilitate the inversion of  $\gamma$ -phosphate, see the discussion in [102]. Notably, any rotation of  $\gamma$ -phosphate inevitably disturbs the coordination sphere of Mg<sup>2+</sup>, since the O<sup>1G</sup> atom of  $\gamma$ -phosphate is one of the Mg<sup>2+</sup> ligands (Fig. 1.2C). In P-loop NTPases, the O<sup>1G</sup> atom is negatively charged, so that its displacement, by affecting the proton affinity of the other five Mg<sup>2+</sup> ligands, may trigger the deprotonation of W<sub>cat</sub>.

#### 4.1.3. Constriction of the catalytic site in the transition state

Binding of TS analogs leads to greater constriction of catalytic sites than binding of ATP, GTP, or their analogs. The highest constriction of the catalytic site in the presence of ADP-VO<sub>4</sub><sup>3-</sup> as a TS analogue was earlier reported from the comparative analysis of myosin structures. In this work, the volume of the catalytic pocket was assessed from SASA measured with ADP, ADP:BeF<sub>3</sub>, and ADP-VO<sub>4</sub><sup>3-</sup> as ligands [203]. The constriction of the catalytic point of myosin is obvious from 3.8. Also the constriction is evident from comparison of the open catalytic pocket of a small GTPase (Fig. 3.8B) and the closed, TS-containing pockets of such enzymes (Fig. 1.2E, 1.4C).

From the literature, it is known that the distance between  $W_{\text{cat}}$  and the analogues of the  $P^G$  atom is also the shortest in the complexes with  $\text{NDP}:\text{AlF}_4^-$  bound [68,74], which corresponds to the WB-crest moving closer to  $\gamma$ -phosphate in P-loop NTPases. The movement of the WB-crest closer to  $\gamma$ -phosphate follows also from the structure comparison (Fig. 3.8D and Fig. A.3). Furthermore, the Switch I of TRAFAC class NTPases undergoes dramatical conformational changes during the catalytic cycle (see Fig. 3.8D, cf Fig 1.2E and 3.8C).

The constriction of the catalytic pocket also manifests itself in the convergence of the Walker A and Walker B motifs. As it follows from Fig. 3.9 and Table C.1 in Appendix C, the most H-bonds between the side chains of  $\text{Asp}^{\text{WB}}$  and  $[\text{Ser/Thr}]^{\text{K}+1}$  residues can be categorized as short in those structures that contain ATP/GTP or their non-hydrolyzable analogs. Apparently, these short H-bonds form upon “closing” of the catalytic site in response to the substrate binding. Still, these H-bonds are even shorter in the presence of  $\text{NDP}:\text{AlF}_4^-$  and  $\text{NDP}:\text{MgF}_3^-$  (Fig. 3.9, Table C.1), which indicates a further constriction of the catalytic pocket in the presence of these TS analogs.

In addition, our structural analysis strongly indicates that the counter-clockwise twist of  $\gamma$ -phosphate by stimulator(s) correlates with the formation of a new H-bond with  $\text{O}^{2G}$ . Specifically, H-bond-compatible distances between  $\text{HN}^{\text{K}-3}$  and the nearest oxygen/fluorine atom are seen in most structures with TS analogs bound, see Table C.1 and Fig. 3.3.

Twisted  $\gamma$ -phosphates and H-bond compatible  $\text{HN}^{\text{K}-3} - \text{O}^{2G}$  distances are also seen in diverse NTP-containing NTPases which were crystallized in their pre-transition configurations because of mutated  $W_{\text{cat}}$ -coordinating residues and/or their impaired interaction with  $W_{\text{cat}}$ , see Table C.1 and Fig. 3.4.

In sum, as it follows from Fig. 3.3, 3.9 and Table C.1, the structures with TS analogs bound have particularly short H-bonds between  $\text{HN}^{\text{K}-3}$  and  $\text{O}^{2G}$ , as well as between  $\text{Asp}^{\text{WB}}$  and  $[\text{Ser/Thr}]^{\text{K}+1}$ , which suggests an additional constriction of the catalytic site in the TS.

#### 4.1.4. Transition state analogs and energetics of P-loop NTPases

The observation of the shortest H-bonds within the catalytic sites in the presence of properly bound  $\text{NDP}:\text{AlF}_4^-$  (see Fig. 3.9, Table C.1) helps to clarify why these complexes are the most potent functional TS-mimics in P-loop NTPases [70,62,63,68,69,71,72,74,64,75–78,82,79,230–232,83]. While the shape and electric charge of  $\text{MgF}_3^-$  is the same as that of  $\gamma$ -phosphate in the anticipated TS (see Fig. 1.4), the shape of  $\text{AlF}_4^-$  differs significantly, it is tetragonal instead of trigonal. Nevertheless,  $\text{NDP}:\text{AlF}_4^-$  complexes are more potent functional analogs of the TS than  $\text{NDP}:\text{MgF}_3^-$  or  $\text{NDP}:\text{AlF}_3$ , as follows from their superiority in promoting binding of P-loop NTPases to their

activators [68,69,71,74,64,75–81]. In addition, the distances between  $W_{\text{cat}}$  and the metal atom are shorter (2.0–2.1 Å) in the case of  $\text{NDP}:\text{AlF}_4^-$  complexes than  $\text{NDP}:\text{MgF}_3^-$  (approx. 2.5 Å) or  $\text{NDP}:\text{AlF}_3$  complexes (approx. 3.0 Å) [68,74]. The reason why  $\text{NDP}:\text{AlF}_4^-$  performs better than other substrate and TS analogs in all these cases has remained obscure, see e.g. [75].

The reason could be, however, understood by considering the energetics of P-loop NTPases. Earlier MD simulations performed in our group have shown that the linking of  $\alpha$ - and  $\gamma$ -phosphates by an inserted stimulatory cationic moiety is an endergonic reaction; about 20–25 kJ/mol appear to be needed to twist the  $\gamma$ -phosphate by straining its bonds with  $\text{Mg}^{2+}$  and  $\text{Lys}^{\text{WA}}$  [102,233]. This activation barrier prevents haphazard NTP hydrolysis due to accidental insertion of a  $\text{K}^+$  ion or an Arg residue in the catalytic site, as argued elsewhere [102,233]. These estimates from MD simulations corroborate earlier considerations of Warshel and colleagues that the stimulation of hydrolysis in P-loop NTPases by their cognate activators requires an input of about 20–25 kJ/mol of free energy, see e.g. [92,94].

A specific problem of P-loop NTPases is the source of a such free energy input. In most enzymes, the free energy for lowering the activation barrier – by destabilizing the substrate-bound ground state and/or stabilizing the TS – is usually provided by substrate binding [234]. In P-loop NTPases, where the substrate binding step is separated from the ultimate catalytic step in most cases, the energy of substrate binding is partly used to bring the NTP molecule into its extended, catalytically prone conformation with eclipsed  $\beta$ - and  $\gamma$ -phosphates [164,235,236]. Another part of the binding energy is used to pre-organize the catalytic site, with the conformational changes that accompany catalytic site closure often being coupled with useful mechanical work, e.g. in myosins [131,237]. While increasing the rate of NTP hydrolysis as compared to that in water (by five orders of magnitude in the case of Ras GTPase [238,239]), the free energy of NTP binding to the P-loop is spent without achieving physiologically relevant hydrolysis rates. For the further acceleration of hydrolysis, additional source(s) of free energy is/are needed.

The source of additional free energy is evident in the case of ring-forming oligomeric P-loop NTPases, such as many AAA+ ATPases (see Fig. 1.8A and [143,240]), helicases (see Fig. 1.8B–C and [112,146]), or rotary ATPases/synthases of the RecA/ $F_1$  class (see Fig. 1.9A and [223,241,242]). In such complexes, the free energy for the cleavage of a bound ATP molecule in one monomer is provided by the binding of another ATP molecule to another monomer. Part of the substrate binding energy at this other site drives a molecular motion that is transferred, for example, to the Arg finger; in the case of rotational  $F_1$ -ATPases, Paul Boyer called this kind of coupling "binding change mechanism" [223].

The same reasoning can be applied to ABC transporters (Fig. 1.8D), where ATP-hydrolysis appears to be coupled to the binding of the translocated molecule. Also, the activation of kinases (Fig. 1.7C-D) appears to be driven by binding of their phosphate-accepting substrates.

In other cases, however, P-loop NTPases, with an NTP molecule bound, do not bind further small molecules, but only the cognate activating partners, which are protein/RNA/DNA molecules (Table B.1). These NTPases, per exclusionem, can harness only the free energy of binding to their activating partner(s). Indeed, there is evidence that the strength of such protein-protein binding does affect the catalytic activity of some P-loop NTPases [29,243–245], making them easy to regulate via protein-protein interactions, as observed with small GTPases.

In all these cases, we are dealing with thermodynamic coupling where the free energy of binding is used for endergonic constriction of the catalytic site. Such a thermodynamic coupling is reversible by definition. Consequently, if we find a way to exergonically constrict the catalytic site, it will lead to the spontaneous assembly of the complex between P-loop NTPase and its cognate activator(s). Apparently, the binding of TS analogs such as  $\text{NDP:AlF}_4^-$  or  $\text{NDP:MgF}_3^-$ , which form multiple strong bonds within the catalytic site, is an exergonic reaction that leads to a TS-like configuration of the catalytic site by "constricting it from the inside", which can happen even in the absence of an activating partner (as in the  $\text{NDP:AlF}_4^-$ -containing separately crystallized dimer of GTPase domains of the MnME protein (Fig. 1.5A) or SIMIBI NTPases (Fig. 1.7A-B). If cognate activators are also present, the "from-the-inside constriction" drives their interaction with the NTPase domain yielding a full-fledged activated complex, as described in the literature [68,69,71,74,64,75–81].

If this assumption is correct, then the assembling "efficiency" of the TS analogs will depend on the amount of free energy that becomes available upon their binding. As it follows from Fig. 3.3 and Fig. 3.9) fluoride complexes are superior to vanadate complexes in their ability to constrict the catalytic site. This is consistent with the fact that the more electronegative fluorine atoms form stronger H-bonds than oxygen atoms. Also  $\text{NDP:AlF}_4^-$  complexes are superior to  $\text{NDP:MgF}_3^-$  complexes (Fig. 3.3 and Fig. 3.9) because they enter into more bonds, as we specified elsewhere [109]. Not surprisingly,  $\text{NDP:AlF}_4^-$  complexes, which "raise" more binding energy despite their "unphysiological" geometry, are the most efficient in assembling activated complexes *in vitro*.

The ability of TS analogs to induce self-assembly of activated complexes appear to correlate with the total strength of the bonds that TS analogs can form in the catalytic pocket [68,69,71,74,64,75–81], suggesting that these H-bonds stabilize the TS upon hydrolysis of ATP or GTP. On the one hand, because of the additional fluorine atom, the

NDP:AlF<sub>4</sub><sup>-</sup> complexes “overperform” other metal fluorides complexes in the ability to constrict the catalytic pocket. On the other hand, because of this additional fluorine atom, AlF<sub>4</sub><sup>-</sup> molecules are always in danger of being involved in a non-physiological bonding that can deform the catalytic pocket. This trade-off must always be taken into account.

#### 4.1.5. Summary on common structural traits of P-loop NTPases

Comparative structural analyses unraveled the following common structural features of P-loop NTPases:

- i. in P-loop NTPases of different classes, the NTP molecules (or their analogs) are bound by the Walker A motif in a same extended conformation (Fig. 1.5A); this conformation is likely to be catalytically prone [102];
- ii. the common trait of all inspected stimulators is their mechanistic interaction with the oxygen atom(s) of  $\gamma$ -phosphate (Fig. 3.5, 3.6, Table E.1);
- iii. comparing the structures with analogs of ATP/GTP and TS, respectively, we noticed that the binding of TS analogs results in greater constriction of catalytic sites than the binding of ATP or GTP. The constriction manifests itself in the shorter distances between Asp<sup>WB</sup> and [Ser/Thr]<sup>K+1</sup> which are as short as 2.5 Å on average in the presence of ADP:AlF<sub>4</sub><sup>-</sup> (Fig. 1.2, 1.4-1.9, 3.9). Also the distances between HN<sup>K-3</sup> and the analogs of  $\gamma$ -phosphate are shorter in the TS-like structures, see Fig. 1.2, 1.4-1.9, 3.3;
- iv. in TS-like configurations of P-loop NTPases of all major classes, except the TRAFAC NTPases, the W<sub>cat</sub>-coordinating “catalytic” Glu or Asp residue links W<sub>cat</sub> with Mg<sup>2+</sup>-ligands in positions 3 and/or 6, see Fig. 1.2D,E, 1.7-1.9, 3.8A. For the TRAFAC class NTPases we could also reconstruct tentative proton pathways by comparing substrate- and TS analog-containing structures, Fig. 3.8D.

These common structural traits, which are shared by P-loop NTPases of different classes, have been mostly overlooked so far. These features, however, might be of key importance for understanding the common mechanism of P-loop NTPases.

## 4.2. Common mechanism of P-loop NTPases as inferred from global comparative structure analysis

### 4.2.1. Requirements for catalysis

The cleavage of  $\gamma$ -phosphate by P-loop NTPases is thought to involve a nucleophilic attack of W<sub>cat</sub>/OH<sub>cat</sub><sup>-</sup> on the P<sup>G</sup> atom (Fig. 4.2). However, it is yet unclear whether the reaction proceeds in two steps separated by formation of a metastable intermediate (Fig. 4.2A-B), or in one concerted transition with the pentavalent trigonal bipyramidal (tbp)

transition state, see Fig. 4.2C and [34,51,52,54,57,68,74,85–87,94]. In the case of two-step mechanism, it is also unclear whether the reaction follows the dissociative  $S_N1$  pathway (Fig. 4.2A) or the associative  $S_N2$  pathway (Fig. 4.2B). In the  $S_N1$  mechanism, the rate-limiting step is dissociation of the terminal  $O^{3B}-P^G$  bond to form a metaphosphate intermediate, which interacts with  $OH^-_{cat}$  yielding inorganic phosphate ( $P_i$ ), see Fig. 4.2A.

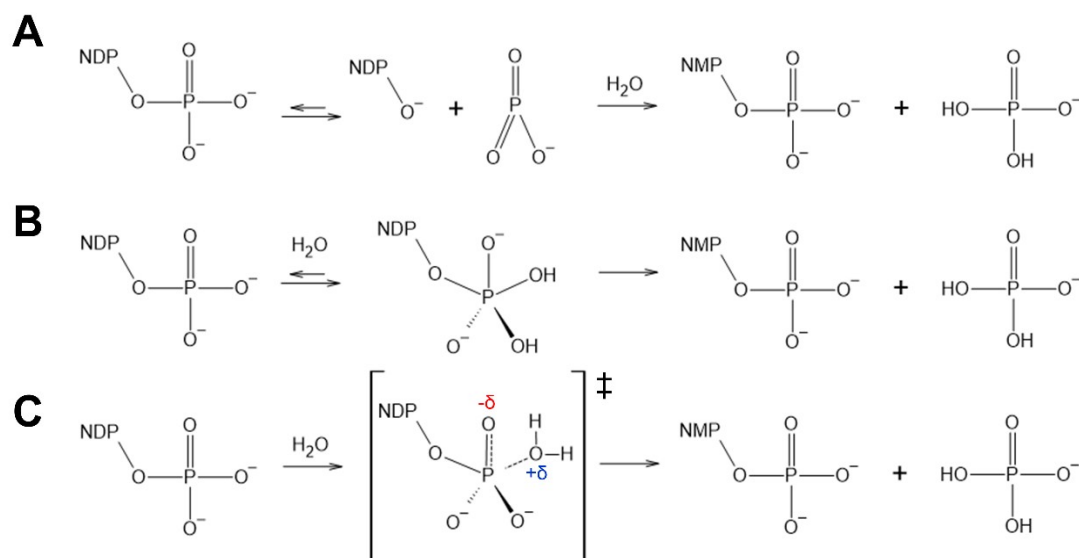


Figure 4.2: Mechanisms of NTP hydrolysis. Dissociative (A), associative (B), and concerted pathways (C).

In the  $S_N2$  mechanism, the rate-limiting reaction is the nucleophilic attack by  $OH^-_{cat}$  on  $P^G$  and the formation of the pentavalent *tbp* intermediate, which then dissociates into NDP and  $P_i$ , see Fig. 4.2B.

The research community is roughly evenly divided between proponents of these three mechanisms, see e.g. [87,94,246]. However, it is not clear whether the difference between the mechanisms is fundamental in this case. It was experimentally shown that the mechanism of ATP hydrolysis varies even in water; it changes from dissociative in pure water to associative in the presence of positively charged chelators for  $\gamma$ -phosphate [247].

Therefore, instead of focusing on these differences, we would like to relate our findings to the universally recognized requirements for NTP hydrolysis, over which there is a general consensus in the literature, namely:

1. **in any model, the covalent bond between  $O^{B3}$  and  $P^G$  must be destabilized** [51,52,85].

Comparative structure analysis showed that the NTP molecule is bound within the closed catalytic site in a conformation more extended compared to its conformation in water, as noted earlier for particular enzymes [102,238,248,249]. As seen in Figure 1.5A, this extended conformation of the bound NTP molecule is common to all major



families of P-loop NTPases. In enzyme-bound NTP molecules, the  $\beta$ - and  $\gamma$ -phosphates are in an almost perfectly eclipsed conformation due to interactions with  $\text{Mg}^{2+}$  and  $\text{Lys}^{\text{WA}}$  (see Fig. 1.2), in contrast to their staggered configuration in water [102,249]. The repulsion of the eclipsed oxygen atoms has the potential to destabilize the  $\text{O}^{3\text{B}} - \text{P}^{\text{G}}$  bond [57,164,235].

Our finding that stimulators always interact with  $\gamma$ -phosphate indicates that this interaction may contribute further to destabilization of the  $\text{O}^{3\text{B}} - \text{O}^{3\text{G}}$  bond. For example, the twisted  $\gamma$ -phosphate may become more eclipsed relative to the  $\alpha$ -phosphate, so that the oxygen atoms of the  $\alpha$ - and  $\gamma$ -phosphate repel each other, as suggested by Rudack and colleagues [99]. Moreover, any twisting or pulling  $\gamma$ -phosphate can destabilize the entire  $\text{Mg}$ -NTP system, inevitably disturbing the coordination sphere of  $\text{Mg}^{2+}$ , since the  $\text{O}^{1\text{G}}$  atom of  $\gamma$ -phosphate is one of the  $\text{Mg}^{2+}$  ligands (Fig. 1.2C, 1.5).

2. **The  $\gamma$ -phosphate group undergoes a steric inversion when forming a covalent bond with the nucleophile, see Fig. 4.2 and [51]. Therefore, catalytic interactions should planarize the  $\gamma$ -phosphate [248,250].**

Already in the absence of the stimulator, the oxygen atoms of the  $\gamma$ -phosphate are already "pulled up" to the  $\beta$ -phosphate by  $\text{Lys}^{\text{WA}}$ , which interacts with  $\text{O}^{1\text{B}}$  and  $\text{O}^{2\text{G}}$ , and by  $\text{Mg}^{2+}$ , which interacts with  $\text{O}^{2\text{B}}$  and  $\text{O}^{1\text{G}}$ , see Fig. 1.2C and [251]. The stimulatory moiety in the AG site further planarizes  $\gamma$ -phosphate by drawing the  $\text{O}^{3\text{G}}$  atom toward  $\text{O}^{2\text{A}}$  and enabling the interaction between  $\text{HN}^{\text{K-3}}$  and  $\text{O}^{2\text{G}}$  (Fig. 1.2, 1.4-1.9, Table C.1). Even when interacting only with  $\gamma$ -phosphate, the stimulator is often located between  $\alpha$ - and  $\gamma$ -phosphates, as in dynamins (Fig. 1.6A) or ABC transporters (Fig. 1.8D) and can planarize  $\gamma$ -phosphate by tying  $\text{O}^{3\text{G}}$  to the "head" of the NTP molecule. For instance, the  $\text{Na}^+$  ion in dynamins, while not reaching the  $\alpha$ -phosphate directly, is connected to it via two noncovalent bonds (Fig. 1.6A). The signature motif of ABC ATPases is H-bonded via conserved Ser and Gln residues to the  $\text{O}2'$  atom of the ribose (Fig. 1.8D).

Notably, the twisted or tilted  $\gamma$ -phosphate becomes more eclipsed relative to  $\alpha$ -phosphate, but *less* eclipsed relative to  $\beta$ -phosphate [102], which might be a prerequisite for inversion of the  $\gamma$ -phosphate group (otherwise, in the case of an ideal eclipse, the oxygen atoms of  $\beta$ -phosphate would prevent the inversion of  $\gamma$ -phosphate).

3. **Negative charges of the oxygen atoms of  $\gamma$ -phosphate must be compensated to increase the electrophilicity of the  $\text{P}^{\text{G}}$  atom and make it prone to nucleophilic attack [34,51,52,54,57,68,85,87,94,248];**

Our analysis confirms for P-loop NTPases of different classes that the positive charges of amino group of  $\text{Lys}^{\text{WA}}$ , the  $\text{Mg}^{2+}$  ion, and several  $\text{HN}$  groups of the P-

loop compensate for the negative charges of phosphate oxygen atoms (see Fig. 1.5B) so that electrons are “pulled away” from the  $P^G$  atom. The positions of the groups involved are strictly conserved (Fig. 1.5), so that such a compensation is common to all major families of P-loop NTPases. This may, at least partly, explain why the binding of the GTP molecule to the catalytic site of Ras GTPase accelerated hydrolysis by five orders of magnitude, as compared to hydrolysis in water, even in the absence of an activating partner [238,239]. In addition to these common positively charged moieties, class-specific auxiliary residues could be involved, as, for instance, “sensors 3” in some AAA+ ATPases.

Here we show, for P-loop NTPases of different classes, that stimulators mostly, albeit not always, carry a positive charge, see Figures 1.2, 1.4–1.9.

The additional electrostatic compensation significantly increases the electrophilicity of the  $P^G$  atom. Rudack and colleagues showed in their QM/MM calculations that the insertion of the arginine finger alone increases the partial positive charge on  $P^G$  to 1.46 elementary charges [99]. The interaction of  $O^{2G}$  with  $HN^{K-3}$  should further increase the positive charge on  $P^G$  in the pre-transition state. Also we provide evidence that a new H-bond is formed between  $HN^{K-3}$  and  $O^{2G}$  or its counterpart in case of TS-state analogs (Fig. 3.3), as well as ATP-containing structures that were trapped in their pre-catalytic states (Fig. 3.4). This additional interaction of  $O^{2G}$  with  $HN^{K-3}$ , should further increase the positive charge on  $P^G$  in the pre-transition state. Remarkably, the phosphate chain “sits” on the last N-terminal turn of the  $\alpha_1$ -helix, which generally carries a dipole positive charge of about 0.5 [164,165]. This positive charge should also contribute to electrostatic compensation.

As already noted in Section 1.2.3.3, the electrostatic potential at the catalytic sites of diverse P-loop NTPases is distributed unevenly. Notably, the local electric field is directed approximately from the positively charged cluster around the P-loop to the negatively charged cluster around  $Asp^{WB}$ . Consequently, in those cases where the stimulator is positively charged, its positive charge not only secures the bonding with particular oxygen atom(s) of triphosphate and increases the positive charge on  $P^G$  but additionally polarizes the whole catalytic pocket.

#### 4.2.2. Does the $[Ser/Thr]^{K+1} - Asp^{WB}$ pair accept a proton from $W_{cat}$ ?

One more requirement of catalysis, as already mentioned in Sections 1.2.1, 3.4.1 is that the conditions in the catalytic site should enable deprotonation of  $W_{cat}$  by a suitable proton acceptor [51,52,54,85,86];

In their comprehensive review on enzymatic mechanisms of phosphate transfer, Cleland and Hengge wrote: “The problem for an ATPase ... is thus to position a water

molecule so that it is in a position to attack the  $\gamma$ -phosphorus. This requires steric restraints as well as organized hydrogen bonding networks. And more specifically, there must be a path for one proton of the attacking water molecule to reach a suitable acceptor" (quoted from [86]). In P-loop NTPases, the proton from  $W_{\text{cat}}$  is commonly thought to be taken up either by  $\gamma$ -phosphate in TRAFAC class NTPases [88,94] or by  $W_{\text{cat}}$ -coordinating, „catalytic“ Glu or Asp residues in other classes of P-loop NTPases [31,72,150,157,166,199,252–256]. Still, our comparative structure analysis showed that such Glu/Asp residues, present in all classes of NTPases except the TRAFAC class, are non-homologous. In SIMIBI NTPases,  $W_{\text{cat}}$ -coordinating Glu or Asp residues are at the C-tip of their WB+1  $\beta$ -strands (Fig. 1.7A-B). In most classes of ASCE ATPases, the “catalytic” Glu<sup>D+1</sup> follows Asp<sup>WB</sup>, see Fig. 1.8, and [50,67,112]. In F<sub>1</sub>/RecA ATPases, the “catalytic” Glu residue is located at the C-tip of the WB+1  $\beta$ -strand, see Fig. 1.9 and [31,37,112]. Also, those kinases that require deprotonation of the second substrate have usually an Asp<sup>WB</sup>-Leu<sup>D+1</sup>-[Asp/Glu]<sup>D+2</sup> motif where the D+2 residue interacts with the prospective nucleophilic group, see Fig. 1.7D and sequence alignments in [40]. Such an absence of homology is unusual for catalytically critical residues.

More importantly, neither  $\gamma$ -phosphate, nor  $W_{\text{cat}}$ -coordinating Glu/Asp residues can hold the proton from  $W_{\text{cat}}$ . Indeed, the inverse solvent isotope effect on the GAP-activated hydrolysis by the Ras GTPase upon H/D substitution [257] indicates that deprotonation is likely to occur before the rate-limiting step of bond cleavage [86]. The bond cleavage time by P-loop NTPases is in the millisecond range (100 ms at 260°K in a Ras GTPase activated by the RasGAP [258]). During this time, the proton must be “detained” in such a way as to prevent its return to  $\text{OH}^-_{\text{cat}}$  and the reversal of the reaction. In general, the ability to keep the proton is seen as a prerequisite for the completion of enzymic reactions with participation of deprotonated nucleophiles, as discussed e.g. for the eukaryotic cAMP-dependent protein kinase [259,260].

The common belief of proton trapping from water ( $\text{pK}_a=14.0$ ) either by a “catalytic” Glu or Asp residue in a polar environment (with expected  $\text{pK}_a$  in the range of 2.0-5.0) or by  $\gamma$ -phosphate (with  $\text{pK} < 3.0$  [261,262]) is at odds with the basic rules of proton transfer as formulated by Eigen [263,264]. After Eigen, if the donor and acceptor of the proton are connected by a “hydrogen bridge”, proton transfer between them is “practically unhindered *provided the difference  $\text{pK}_{\text{acceptor}} - \text{pK}_{\text{donor}}$  is positive*” (quoted from [264]). In the case of  $\gamma$ -phosphate and  $W_{\text{cat}}$ -coordinating Glu/Asp residues, the respective differences are strongly negative so the proton will promptly (in picoseconds) return to  $W_{\text{cat}}$ .

It was speculated that the  $W_{\text{cat}}$ -coordinating groups may facilitate its catalytic deprotonation by decreasing the proton affinity of  $W_{\text{cat}}$ , see e.g. [98]. Our structural analysis of TS-like structures showed that positively charged side chains of Lys or Arg,

which could significantly lower the pK value of  $W_{\text{cat}}$  interact with it only in a few classes of P-loop NTPases, e.g. in some AAA+ ATPases, see Section 1.2.3 and Table B.1. Even in these cases, their positive charges can hardly decrease the apparent pK of  $W_{\text{cat}}$  by >10 pH units, which is needed for sustainable protonation of a “catalytic” Glu/Asp residue or  $\gamma$ -phosphate. Hence, neither “catalytic” acidic residues, nor  $\gamma$ -phosphate can trap a proton from  $W_{\text{cat}}$  per se.

As early as in 2004, some of these problems were recognized by Frick and colleagues [265,266] who calculated the pK values for ionizable residues of the SF2 class Hepatitis C virus (HCV) NS3 helicase by using the MCCE software [267]. Frick and colleagues wrote about supposedly  $W_{\text{cat}}$ -coordinating Glu291<sup>D+1</sup> and its preceding Asp290<sup>WB</sup>: “to function as a catalytic base, the pKa of Glu291 would need to be much higher than that of a typical Glu in a protein. However, electrostatic analysis of all HCV helicase structures reveals that neither Glu291, nor any nearby Glu, has an abnormally high pKa. In contrast, Asp290 has a pKa as high as 10 in some structures and as low as 3 in others. Interestingly, in structures in the open conformation (such as 8OHM), the pKa of Asp290 is low, and in the closed conformation (ex. 1A1V), the pKa of Asp290 is higher than 7, suggesting that Asp290 picks up a proton (like a catalytic base) when the protein changes from the open to the closed conformation. Thus, Asp290 could serve as a catalytic base...” (quoted from [266]). Frick and colleagues made their calculations with crystal structures of helicases that contained neither Mg-ATP nor its analogs. Therefore, they were unaware of the exact positions of Asp290<sup>D</sup> and Glu291<sup>D+1</sup> relative to the bound substrate and could only guess by analogy with known structures of related P-loop NTPases. The two subsequently resolved ADP:AlF<sub>4</sub><sup>-</sup>-containing, TS-like crystal structures of HCV NS3 helicase (PDB IDs 3KQL [268] and 5E4F [193]) show a typical for SF2 helicases arrangement of Asp290<sup>WB</sup> and Glu291<sup>D+1</sup> (see Fig. 1.8C). Furthermore, both structures show short 2.4 Å H-bonds between Ser211<sup>K+1</sup> and Asp290<sup>WB</sup>, as seen in Fig. 1.8C.

Unlike “catalytic”,  $W_{\text{cat}}$ -coordinating Glu/Asp residues, Asp<sup>WB</sup> is strictly conserved in P-loop NTPases. Equally strictly conserved is the [Ser/Thr]<sup>K+1</sup> residue [32,40,111,269]. Both these residues can translocate protons. Therefore, building on our comparative structure analysis and following Frick and colleagues who suggested Asp290<sup>WB</sup> as a catalytic base in the HCV NS3 helicase [265,266], we propose here that the buried, H-bonded pair of strictly conserved [Ser/Thr]<sup>K+1</sup> and Asp<sup>WB</sup> serves as a universal module that accepts the proton from  $W_{\text{cat}}$  and holds it as long as needed in P-loop NTPases of all classes.

#### 4.2.2.1. Comparison with known proton-transfer modules

It can be countered that the [Ser/Thr]<sup>K+1</sup> – Asp<sup>WB</sup> pair does not interact directly with  $W_{\text{cat}}$ . However, direct tracing of intra-protein displacements of protons in energy converting enzymes and chemical models (see [263,264,270–282] for reviews) showed that fast proton transfer over a distance of up to 20 Å can be mediated by water bridges, provided that the distance between the groups involved is  $\leq 3.0$  Å, in accordance with Eigen [264]. Thereby it does not matter that water molecules are equally poor proton acceptors (with  $pK_b$  of 0.0) and proton donors (with  $pK_a$  of 14.0) because protons pass water by so-called von Grothuss mechanism [283–285]. According to the modern version of this relay mechanism, the bridging molecule receives an external proton simultaneously with the transfer of its own proton to the next carrier at picoseconds [284,285]. Apart from water molecules, the side chains of serine, threonine, tyrosine, and neutral histidine are suitable for proton transfer by the von Grothuss mechanism: they have a proton-accepting lone pair of electrons and also their own proton to transfer on [286]. In addition, protons are transferred very fast between carboxy groups when they are bridged by single water molecules [263,264,276,280]. Zundel and colleagues showed that such systems are highly polarized [270,287], which lowers the activation barriers for proton transfer.

The best-studied examples of such proton-conducting chains are known from the photochemical reaction center (PRC) of  $\alpha$ -proteobacterium *Rhodobacter sphaeroides* (as shown in Fig. 4.3A and described in its caption) and bacteriorhodopsin of an archaeon *Halobacterium salinarum* (as shown in Fig. 10B and described in its caption).

The two proton transfer routes shown in Fig. 4.3A-B were reconstructed from direct electrometric tracking of flash-induced proton displacements [272,288–291], kinetic IR and UV/Vis spectroscopy data [272,273,292–297], kinetic ESR measurements [298–301], and comparative structure analyses [171,281,302,303]. These are the two best-understood cases of intra-protein proton transfer by far.

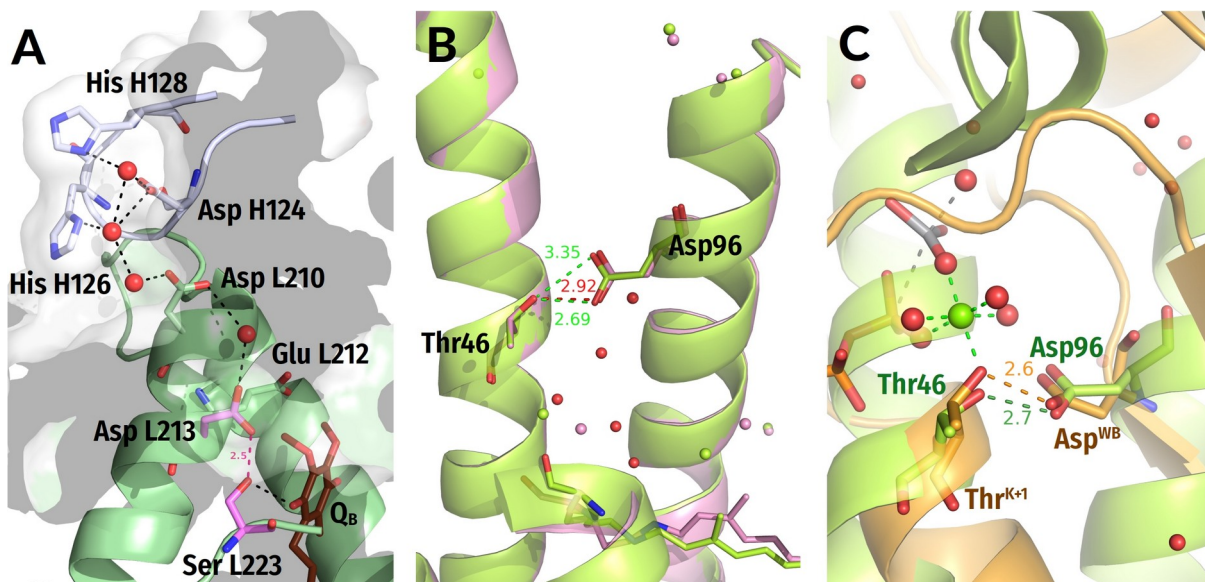


Figure 4.3: Proton traps in the photochemical reaction center and bacteriorhodopsin as compared with the Thr<sup>K+1</sup>-Asp<sup>WB</sup> pair of P-loop NTPase.

Distances are given in ångströms. **A**, a high-pK proton trap in the photosynthetic reaction center (PRC) from *Rhodobacter sphaeroides*. Shown is the PRC in the charge-separated state (PDB ID 1DV3 [383]). Here, a proton path connects the protein surface with the buried binding site of the secondary ubiquinone acceptor Q<sub>B</sub>. The proton is translocated over more than 20 Å via amino acid residues of the two PRC subunits, L and H. The proton pathway begins at the two His residues on the surface (His H126 and His H128); they are thought to have pK values of about 7.0 and to harvest protons which are ejected onto the photosynthetic membrane surface by the ATP synthase. From these histidine residues, the proton goes, via water-bridged carboxy groups of Asp H124 and Asp L210, to the Q<sub>B</sub>-binding pocket. Proton transfer from the surface to the more hydrophobic interior of the protein becomes possible after the Q<sub>B</sub> molecule docks to the Glu L212 – Asp L213 – Ser L223 cluster [302] and turns it into a buried, high-affinity proton acceptor (a membrane proton trap according to Mitchell [391]), with a functional pK of about 9.5 – 10.0 [292,293,306,380,392]. This high functional pK is apparently due to the expulsion of water upon Q<sub>B</sub> binding, the negative charge on the Glu L212 – Asp L213 pair, and the H-bond between Asp L213 and Ser-L223, through which the proton passes to Q<sub>B</sub>. The functional pK value of the trap increases to > 12.0 after the appearance of an electron at Q<sub>B</sub> and the formation of the Q<sub>B</sub><sup>-</sup> anion radical, which attracts further protons into the buried catalytic site [275]. Mutations of either Glu L212, or Asp L213, or Ser L223 block proton transfer from the surface [380–382]. For further details, see [171,279,383–385].

**B**, a high-pK proton trap in bacteriorhodopsin (BR), a membrane protein that pumps a proton across the membrane in response to the photoexcitation of its retinal pigment, see [272,273,298,363] for reviews. Shown is the superposition of femtosecond X-ray laser-captured structures of bacteriorhodopsin from [386] in a closed resting state (PDB ID 6G7H, light-green) and in an “open” state 8.3 ms after illumination (PDB ID 6G7L, pink). Additional red-colored water molecules were taken from a high-resolution crystal structure of the V49A bacteriorhodopsin mutant that was crystallized in an “open” state (PDB ID 1P8U, [387]). In the BR, the key proton carrier Asp-96 has pK of ~12.0 owing to the absence of water in the vicinity and a H-bond with Thr46 of the nearby α-helix. Replacement of Thr46 by a valine decreased the apparent pK of Asp96 by approximately two pH units [388], which may characterize the contribution of the H-bond between Thr46 and Asp96 to the unusually high apparent pK of the latter. The photoisomerization of the retinal twists the α-helices and lets water molecules in the space between them, see the pink structure [272,281,282,299,300,303,337,389]. The pK value of Asp-96 shifts to 7.1, which allows its deprotonation and proton transfer, via a transiently formed water chain, to the Schiff base of retinal at about 10 Å, see [294,303,390]. Later, the cleft between helices closes again, the pK of Asp96 returns to 12.0, and it is reprotonated from the surface [272,273,363,303].

**C**, the Thr46-Asp96 pair of bacteriorhodopsin in its resting closed state (the green structure from panel 10B, PDB ID 6G7H) superimposed with the Thr<sup>K+1</sup>-Asp<sup>WB</sup> pair from the transition-state like, VO<sub>4</sub><sup>3-</sup>-containing structure of myosin shown in Fig. 1.6D, orange (PDB ID 1VOM [215]).

In the case of PRC (Fig. 4.3A), the buried proton acceptor must have a functional pK value greater than 9.0 to compensate for the desolvation penalty of about three pH units, which must be “paid” for delivery of a proton from a pK-neutral “antenna” group at the protein surface into the hydrophobic membrane [304,305]. However, if such a buried

group is equilibrated with the protein surface, its pK value can hardly exceed the ambient pH, which is usually neutral, because the groups with even higher pK values are already protonated. Hence, the strong proton acceptor shown in Fig. 4.3A is not fully equilibrated with the surrounding solution. In this case, it is better to speak about the *apparent/functional* pKa or high *proton affinity*. Such very strong proton acceptors usually emerge only transiently [306,307].

The proton traps in Fig. 4.3A, B include H-bonded [Ser/Thr] – Asp pairs (modules). These examples show that proper H-bonding of a protein-buried Asp residue with a nearby Ser/Thr residue in a hydrophobic environment can generate a trap with a high proton affinity; such traps can, in principle, accept protons even from water. Specifically, in the absence of pH-buffers and protons from the ATP synthase, the protons that pass through the PRC to compensate the negative charge at the Q<sub>B</sub> site (Fig. 4.3A) appear to detach from surface water molecules, presumably polarized by surface charges [297].

#### 4.2.2.2. Evidence for [Ser/Thr]<sup>K+1</sup> – Asp<sup>WB</sup> as a transient proton trap

Our suggestion that a similarly H-bonded [Ser/Thr]<sup>K+1</sup> – Asp<sup>WB</sup> pair accepts the proton from W<sub>cat</sub> in P-loop NTPases of different classes, to our best knowledge, has not been considered before. Therefore, we further substantiate the suggested mechanism below:

1. In diverse P-loop NTPases, mutations of Asp<sup>WB</sup> or [Ser/Thr]<sup>K+1</sup> diminished the enzyme activity dramatically, see e.g. [266,308–315]. More specifically, mutations of Asp<sup>WB</sup> to Asn retarded the activated hydrolysis without affecting the NTP binding [266,308–312]. In the case of *E.coli* F<sub>1</sub>-ATPase, the mutation even increased the affinity for ATP [310]. The Asp<sup>WB</sup> to Asn mutation mimics the charge state of a protonated Asp<sup>WB</sup>. Hence, the protonation of Asp<sup>WB</sup> is unlikely to distort the catalytic pocket and be the cause of the universal catalytic incompetence of the Asp<sup>WB</sup> to Asn mutants. We attribute this incompetence to the inability of Asn<sup>WB</sup> to trap a proton from [Ser/Thr]<sup>K+1</sup>.
2. In TS-like structures of P-loop NTPases of various classes (except the TRAFAC class), the “catalytic” Glu/Asp residues were found to connect W<sub>cat</sub> with ligands of Mg<sup>2+</sup> in positions #3 or #6, see Fig. 1.2D, 1.7–1.9, 3.8A. For TRAFAC NTPases, we could also reconstruct a tentative proton route via [Ser/Thr]<sup>Swi</sup> from comparative analysis of structures with different substrates or their analogs bound, Fig. 3.8D.
3. Notably, the six ligands of Mg<sup>2+</sup> form a regular octahedron with edges 2.9–3.0 Å long, so that the ligands # 3 and #6 are on a H<sup>+</sup>-transfer distance from [Ser/Thr]<sup>K+1</sup> (Fig. 4.4). The short H-bond between [Ser/Thr]<sup>K+1</sup> and Asp<sup>WB</sup> completes the proton-conducting pathway that connects W<sub>cat</sub> with Asp<sup>WB</sup>. The proton pathways from W<sub>cat</sub> to Asp<sup>WB</sup>, which resemble proton translocation systems of PRC and BR (cf Fig.

4.3), are shown by the red dashed lines for P-loop NTPases of different classes in Fig. 1.7–1.9 and by dashed arrows in Fig. and 4.4.

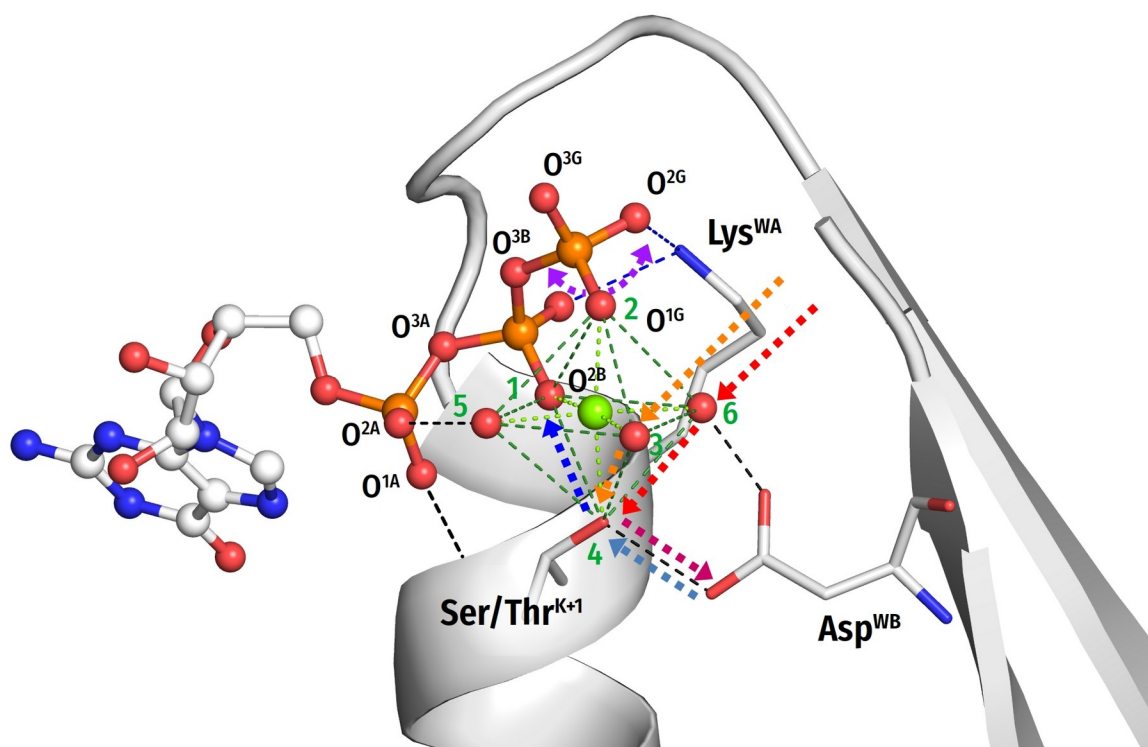


Figure 4.4: Schematic presentation of tentative proton routes along the edges of the octahedral coordination shell of  $Mg^{2+}$  ion.

The  $Mg^{2+}$  ligand #3 is assumed to be a water molecule. Proton entry points via ligand #6 or ligand #3 are shown by red and orange arrows, respectively. Protonic connection between  $[Ser/Thr]^{K+1}$  and  $Asp^{WB}$  are shown by magenta and light blue arrows, the route from  $[Ser/Thr]^{K+1}$  to  $O^{2B}$  of  $\beta$ -phosphate is shown as a dark blue arrow. The movement of the  $O^{1G}$  atom as a result of  $\gamma$ -phosphate twist is shown by purple arrows.

- The  $pK_a$  of an aspartate residue in water is about 4.0, much lower than that of water (14.0). However, unlike the “catalytic” Glu/Asp residues or  $\gamma$ -phosphate surrounded by charged residues (Fig. 1.2D-E, 1.6-1.9),  $Asp^{WB}$  is in a nonpolar environment and its functional  $pK$  is likely to be high when the catalytic site is closed.  $Asp^{WB}$  is in the middle of an  $\alpha\beta\alpha$  sandwich, on the interface between the  $\beta$ -pleated sheet and the  $\alpha_1$ -helix; such interfaces are stabilized by hydrophobic interactions [165]. In addition,  $Asp^{WB}$  is preceded by four hydrophobic residues of the Walker B motif (Fig. 1.2); the adjacent  $\beta$  strands, as well as the  $\alpha_1$ -helix, also contain many hydrophobic residues, see the sequence alignments in [32,40,111,269]. Upon constriction of the catalytic pocket and expulsion of eventually present water molecules, the hydrophobic environment should elevate the proton affinity of the H-bonded  $Asp^{WB}$ , as it happens with similarly H-bonded Asp96 which has a functional  $pK_a$  of  $\sim 12.0$  in a hydrophobic environment of the ground-state BR. Fig. 4.3C shows that the structure of the Ser186<sup>K+1</sup> – Asp454<sup>WB</sup> pair of myosin overlaps nicely with Thr46 – Asp96 pair of BR.



5. In contrast, the pK of [Ser/Thr]<sup>K+1</sup>, which is about 13.0 in water, is likely to be reduced when [Ser/Thr]<sup>K+1</sup> serves as a Mg<sup>2+</sup> ligand. Coordination of a Zn<sup>2+</sup> ion by a serine side chain is known to decrease the pK value of the latter up to 5.5 yielding a serine anion (alkoxide) at neutral pH, see [86] and references therein. The impact of a Mg<sup>2+</sup> ion should be weaker; still, within a closed/constricted catalytic site, the low dielectric permittivity would enhance electrostatic interactions. [Ser/Thr]<sup>K+1</sup> is the most deeply buried of the Mg<sup>2+</sup> ligands (see Fig. 1.2, 1.4-1.9, 4.4), so it should be the most sensitive to electrostatic effects. As a result, the functional pK value of [Ser/Thr]<sup>K+1</sup> may dramatically decrease upon constriction of the catalytic site.
6. The found shortening of H-bonds in the TS-like structures (Fig. 3.3, 3.9) enables estimation of the difference in functional pK values of [Ser/Thr]<sup>K+1</sup> and Asp<sup>WB</sup> in the TS. It is well established, based on ample experimental evidence, that hydrogen bonds “generally shorten as ΔpK<sub>a</sub>, the difference in the donor and acceptor pK<sub>a</sub> values, decreases” (quoted from [205]). Specifically, Herschlag and colleagues observed, on various systems, that the ΔpK<sub>a</sub> decreases linearly from 20 to 0 with the decrease in the O—H••••O distance from 2.9 Å to 2.4 Å, with a slope of 0.02 Å/pK<sub>a</sub> unit, [205,316]. In NDP:AlF<sub>4</sub><sup>-</sup>-containing structures with constricted catalytic site, the length of the H-bonds between [Ser/Thr]<sup>K+1</sup> and Asp<sup>WB</sup> varies around 2.5 Å (Fig. 1.2E-F, 1.6, 1.7A-B, 1.7C-D, 1.9A, 3.9), which corresponds to ΔpK < 3.0 and indicates a low-barrier hydrogen bond [205,316]. Hence, [Ser/Thr]<sup>K+1</sup> and Asp<sup>WB</sup> may have comparably high proton affinities in a constricted catalytic site.
7. In the octahedral coordination shell of Mg<sup>2+</sup>, the O<sup>1G</sup> atom of γ-phosphate is the ligand opposite to [Ser/Thr]<sup>K+1</sup> (Figure 4.4). Therefore, the stimulator-induced rotation of γ-phosphate, by moving O<sup>1G</sup> in *any* direction (as shown by dashed purple arrows in Fig. 4.4), would inevitably increase the distance between O<sup>1G</sup> and the hydroxyl of [Ser/Thr]<sup>K+1</sup>. Pulling away the negatively charged O<sup>1G</sup> will increase the cumulative positive charge at [Ser/Thr]<sup>K+1</sup> prompting the relocation of its proton to Asp<sup>WB</sup>, e.g. in response to a thermal fluctuation [317] (Fig. 4.5A,D).
8. We suggest that the resulting Mg<sup>2+</sup>-coordinated Ser/Thr anion (alkoxide), used as a proton acceptor from water by many enzymes [86,234,318,319], withdraws the proton from W<sub>cat</sub> (or the sugar moiety in some kinases) via proton pathways shown in Figs. 1.7–1.9, 3.8A, 4.4, 4.5B,D; this proton transfer is additionally driven by strong local electric field (see Fig. 1.10 and Section 1.2.3.3). The resulting state where both [Ser/Thr]<sup>K+1</sup> and Asp<sup>WB</sup> are protonated corresponds to the ground state of the Thr46-Asp96 pair in the BR (see the light-green structures in Fig. 4.3B-C).

9. The formed anionic nucleophile (e.g.  $\text{OH}^-_{\text{cat}}$ ) stabilized and polarized by its ligands, is attracted by the electrophilic  $\text{P}^{\text{G}}$  atom (Fig. 4.5B). The proton affinity of the anionic nucleophile decreases as it gets closer to  $\text{P}^{\text{G}}$ , so that proton return from the  $[\text{Ser/Thr}]^{\text{K}+1} - \text{Asp}^{\text{WB}}$  couple becomes increasingly unfavorable, eventually satisfying the Eigen's condition for proton transfer and making it complete.
10. In Appendix G we present and discuss IR-spectroscopic evidence for transient protonation of a carboxyl residue in a P-loop NTPase, as available from the literature [320,321].

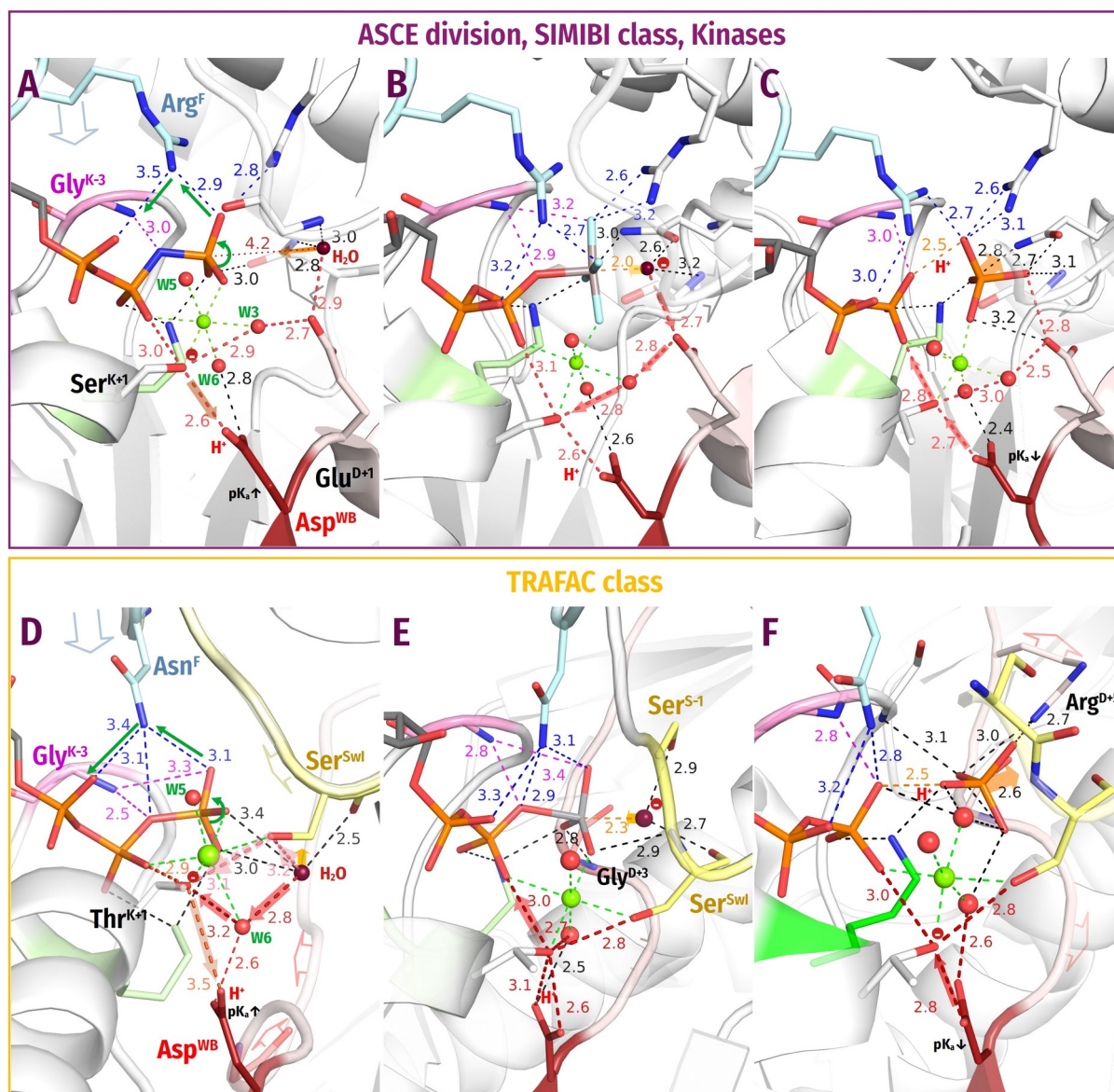


Figure 4.5: Two tentative mechanisms of proton transfer from  $W_{cat}$  to  $Asp^{WB}$  in P-loop NTPases.

The shade of the red arrows of proton transfers varies through panels A-C and D-F to emphasize that the kinetic coupling of these steps with each other and reactions in the catalytic site could vary depending on the thermodynamics of particular enzymes. Other colors as in Fig. 1.6-1.9. **Top (A-C): Tentative mechanism of proton transfer via the  $W_{cat}$ -coordinating Glu/Asp residue in all classes of P-loop NTPases but TRAFAC class enzymes**, as illustrated by structures of the SF1 helicase Pif1p. **A**, In the pre-catalytic state, which is exemplified by a structure with a bound non-hydrolyzable ATP analog ANP (PDB ID 6HPH [393]), the binding of ATP and closure of the catalytic site increases the proton affinity of  $Asp^{WB}$ . Activation-induced constriction of the catalytic site and twist of  $\gamma$ -phosphate prompt the proton redistribution from  $[Ser/Thr]^{K+1}$  to  $Asp^{WB}$ , **B**, in the TS, which is exemplified by a structure with a bound TS analog  $ADP:AlF_4^-$  (PDB ID 5O6B [256]), a proton from  $W_{cat}$  repletes the vacancy at the anionic  $[Ser/Thr]^{K+1}$  by going via the  $W_{cat}$ -coordinating  $Glu^{D+1}$  and W3; the resulting  $OH^-_{cat}$  attacks  $\gamma$ -phosphate; **C**, in the post-TS, which is exemplified by a structure with  $ADP:MgF_4^{2-}$  bound (PDB ID 6S3I [322]), where we replaced the  $H_2PO_4^{2-}$  mimic  $MgF_4^{2-}$  by  $H_2PO_4^{2-}$ , the proton goes from  $Asp^{WB}$ , via  $Ser^{K+1}$ , to  $\beta$ -phosphate to compensate its negative charge (see the main text). **Bottom (D-F): Tentative mechanism of proton transfer from  $W_{cat}$  to  $Asp^{WB}$  in TRAFAC class NTPases**, as illustrated by structures of myosin. **D**, For the pre-catalytic state we used the ATP-soaked crystal structure of myosin II (PDB ID 1FMW [214]) where the Switch I loop was taken from the structure of Myosin V complexed with  $ADP-BeF_3$  (PDB ID 1W7J [379]) and superimposed via  $Lys^{WA}$  and  $Mg^{2+}$ . In this state, the binding of ATP and closing of the catalytic site increases the proton affinity of  $Asp^{WB}$ . Constriction of the catalytic site upon activation and twist of  $\gamma$ -phosphate prompt the proton redistribution from  $Thr^{K+1}$  to  $Asp^{WB}$  (purple arrow). The anionic  $Thr^{K+1}$  accepts a proton from W6 that, in turn, takes a proton from  $W_{12}$  (red arrows, cf. Fig. 3.8B-C); alternatively, the proton route from  $W_{12}$  to  $Thr^{K+1}$  may involve the conserved  $Ser^{Sw1}$  (dashed red arrows). The resulting  $OH^-_{cat}$  is brought into the apical position by residues of Switch I (colored yellow) and WB-crest; **E**, in the transition state (TS), which is exemplified by a structure with a bound TS analog  $ADP:VO_4^-$  (PDB ID 1VOM [215]),  $OH^-_{cat}$  attacks  $P^{\gamma}$ . In this constricted TS, the protonic connection between  $OH^-_{cat}$  and the  $Mg^{2+}$ -coordinating water molecules appears to be broken. In the subsequent stage of the catalytic transition, the proton goes from  $[Ser/Thr]^{WA}$  to  $\beta$ -phosphate and compensates for the large negative charge that builds up upon the detachment of  $\gamma$ -phosphate (the red arrow); **F**, in the post-transition state, which is exemplified by a structure with bound ADP and  $H_2PO_4^{2-}$  (PDB ID 4PJJK [60]), the negative charge on  $\beta$ -phosphate is compensated by a hydrogen bond between  $\beta$ - and  $\gamma$ -phosphate.  $Asp^{WB}$  repletes  $Thr^{K+1}$  concomitantly with the opening of the catalytic pocket.

#### 4.2.2.3. Compensation of the negative charge on the bridging O<sup>3B</sup> atom upon hydrolysis

According to Herschlag and colleagues [56,57,87], the activation barrier of NTP hydrolysis is due to the strong negative charge that develops at O<sup>3B</sup> atom of  $\beta$ -phosphate as  $\gamma$ -phosphate breaks away; this charge should be compensated to cleave the O<sup>B3</sup> – P<sup>G</sup> bond. The comparative analysis of the post-TS structures of the RhoA/RhoA-GAP complex (PDB ID 6R3V [98]) and activated myosin (PDB IDs 4PFP, 4PJK [60]) with NDP and detached P<sub>i</sub> (Fig. 4.5F), as well as the structure of SF1 helicase Pif1p with a H<sub>2</sub>PO<sub>4</sub><sup>2-</sup> mimic ADP:MgF<sub>4</sub><sup>2-</sup> bound (Fig. 4.5C) show that the negative charge of  $\beta$ -phosphate is compensated by the joint action of Mg<sup>2+</sup>, Lys<sup>WA</sup>, Arg/Lys/Asn fingers and HN<sup>K-3</sup>. Also, at this stage the H-bond between HN<sup>K-3</sup> and O<sup>2G</sup> is lost so that HN<sup>K-3</sup> do not compensate the negative charge of  $\gamma$ -phosphate (see Fig. 4.5C,F), but is fully involved in compensating for the negative charge at  $\beta$ -phosphate. In addition, the structures show a short H-bond of 2.4–2.5 Å between the oxygen atoms of  $\beta$ -phosphate and P<sub>i</sub> (see Fig. 4.5C, F and [60,98,322,323]); this H-bond apparently also compensates for the negative charge at  $\beta$ -phosphate, the proton for this bond is thought to stem from W<sub>cat</sub> [60,98,323].

In the octahedral coordination shell of Mg<sup>2+</sup>, [Ser/Thr]<sup>K+1</sup> is 2.9 Å away from  $\beta$ -phosphate, so that its proton can directly pass on and compensate the negative charge that develops at  $\beta$ -phosphate upon cleavage of its bond with  $\gamma$ -phosphate (as shown in Fig. 4.4, 4.5C,E), which is a requirement for catalysis. The proton vacancy on [Ser/Thr]<sup>K+1</sup> could be then refilled by a proton from Asp<sup>WB</sup> (Fig. 4.4, 4.5C, F), thus restoring the initial protonic configuration. Since the octahedral arrangement of Mg<sup>2+</sup> ligands (Fig. 4.4) is similar in all P-loop NTPases, the proton can relocate from Asp<sup>WB</sup> to  $\beta$ -phosphate, via [Ser/Thr]<sup>K+1</sup>, in all of them.

#### 4.2.2.4. Assignment of novel functions to the Walker motifs

In the framework proposed here, the Walker B motif serves as a trap for the proton from W<sub>cat</sub> whereas the coordination shell of the Mg<sup>2+</sup> ion functions as an octahedral proton transfer hub with almost all Mg<sup>2+</sup> ligands involved (Fig. 4.4). The proton is transferred from W<sub>cat</sub> to Asp<sup>WB</sup> in two steps: first, the proton shifts from [Ser/Thr]<sup>K+1</sup> to Asp<sup>WB</sup> in response to the constriction of the catalytic site and stimulatory interaction, and then a proton from W<sub>cat</sub> fills the proton vacancy at [Ser/Thr]<sup>K+1</sup>, as shown in Fig. 4.5. Totally, one proton relocates from W<sub>cat</sub> to Asp<sup>WB</sup>. Still, the existence of separate proton transfer steps, as shown in Fig. 4.4, 4.5 by different shades, provides flexibility; the mechanism could be adapted to the dissociative, associative, or concerted mechanism of hydrolysis. For instance, if the reaction is concerted, the proton from W<sub>cat</sub> can pass directly through the Mg<sup>2+</sup>-coordinating ligands to  $\beta$ -phosphate, with the proton vacancy

at [Ser/Thr]<sup>K+1</sup> being filled by Asp<sup>WB</sup> later, concomitantly with the loosening of the catalytic site and decrease in proton affinity of Asp<sup>WB</sup>, as shown in Fig. 4.5D-F.

The Asp<sup>WB</sup> residue is almost strictly conserved throughout P-loop NTPases, see Table B.1 and [32,33,40,50,111]. Only in several cases its function is performed by Glu, see [50] for details. Some of such cases are shown in Fig. A.5 from Appendix A and described in its extended caption.

Notably, the cause for the strict conservation of the entire Walker B motif has remained obscure so far; to our best knowledge, no function has been attributed to the motif as a whole. The here suggested scheme invokes Asp<sup>WB</sup> as the common terminal acceptor of a proton from  $W_{\text{cat}}$ . Other four hydrophobic residues of the same motif serve as “hydrophobic protonic insulators”. They may be needed to increase the proton affinity of Asp<sup>WB</sup> in the constricted catalytic site and to prevent an eventual unwanted proton escape from Asp<sup>WB</sup>.

In our model, [Ser/Thr]<sup>K+1</sup> also acquires new key functions as a catalytic nucleophilic alkoxide and a von Grotthus-type proton carrier, which may explain its strict conservation, see Table B.1. A few known exceptions are shown in Fig. A.6 in Appendix A and described in its extended caption. Most common outliers are glycine residues which substitute for [Ser/Thr]<sup>K+1</sup> in several distinct families of P-loop nucleotide monophosphate kinases [40] that catalyze reactions not accompanied by deprotonation of the nucleophilic second substrate, see Fig. A.6B and its caption for details.

Neither of the exceptions in Fig. A.5, A.6 calls into question the suggested mechanism where the buried [Ser/Thr]<sup>K+1</sup> – [Asp/Glu]<sup>WB</sup> module turns into a deep proton trap after the constriction of the catalytic site in most P-loop NTPases.

Walker A and Walker B motifs — together — contain only two strictly conserved residues capable of proton transfer, namely [Ser/Thr]<sup>K+1</sup> and Asp<sup>WB</sup>. It appears natural to consider this H-bonded pair in relation to proton transfer from  $W_{\text{cat}}$ .

In contrast, the still prevalent notion that chemically different catalytic bases (phosphates vs glutamates) are used in different classes of apparently homologous P-loop NTPases is surprising. Furthermore, neither phosphates nor glutamates are common as catalytic bases in the other families of phosphate transferases. In one of the most comprehensive reviews on their mechanisms, Cleland and Hengge even pointed out the oddness of anticipated catalytic bases in P-loop ATPases: “... more specifically, there must be a path for one proton of the attacking water molecule to reach a suitable acceptor. ATPases appear not to use general bases such as the aspartates usually found in kinase active sites....” (quoted from [86]). Our message is that P-loop NTPases are no exception, they do use aspartates as catalytic bases just like most other hydrolases.

#### 4.2.2.5. Minimal mechanistic model of NTP hydrolysis by P-loop NTPases

Building on the comparative structure analysis of over 3100 catalytic sites presented here, as well as on the available experimental and theoretical data, we describe the activated catalysis typical for P-loop NTPases by a simple mechanistic model depicted in Fig. 4.6. We show the minimalistic version of the model that includes ubiquitous Walker A and Walker B motifs,  $Mg^{2+}$ -NTP, a simple stimulator, such as a  $K^+$  ion or a Lys/Asn residue, and a few water molecules one of which serves as  $W_{cat}$ . The model, however, can be easily expanded/adapted to fit distinct NTPase families by adding further stimulatory interactions, auxiliary and  $W_{cat}$ -coordinating residues, as well as by replacing  $W_{cat}$  by other nucleophiles. According to the model, the catalytic transition proceeds in the following steps:

- A)  $Mg$ -NTP complex binds to the Walker A and Walker B motifs; the binding energy is used to bring the NTP molecule into an elongated conformation with eclipsed  $\beta$ - and  $\gamma$ -phosphates and to close the catalytic site by surrounding the triphosphate chain by positively charged groups that are provided by the P-loop, WB-crest and, only in TRAFAC NTPases, Switch I loop. The accompanying protein conformational changes can power useful mechanical work. The catalytic site is further stabilized by the H-bond between  $Asp^{WB}$  and  $[Ser/Thr]^{K+1}$ . The H-bond length is about 2.6-2.7 Å (Fig. 3.9); in this state  $Asp^{WB}$  is negatively charged.
- B) An exergonic interaction between the activating partner (another protein domain and/or an RNA/DNA molecule) and the WB-crest (i) shields and constricts the catalytic site, (ii) moves the WB-crest residues closer to the  $\gamma$ -phosphate, (iii) inserts the stimulator(s) next to the phosphate chain. The constriction of the catalytic site shortens the H-bond between  $[Ser/Thr]^{K+1}$  and  $Asp^{WB}$  to 2.4–2.5 Å, turning  $Asp^{WB}$  into a potent proton trap. In most cases, (one of) the stimulator(s) links the  $O^{2A}$  and  $O^{3G}$  atoms of the triphosphate (Fig. 1.2D-E, 1.6B-D, 1.7A-D, 1.7A-C, 1.9A, 3.5-3.7) and twists  $\gamma$ -phosphate counterclockwise; the rotated  $\gamma$ -phosphate is stabilized by a new H-bond between  $O^{2G}$  and  $HN^{K-3}$ . In other cases, the stimulators drag only  $\gamma$ -phosphate and, supposedly, twists it in some direction, see Fig. 1.6A, 1.7D, 1.9B-D, 3.5-3.7. The interaction of stimulators with  $\gamma$ -phosphate (i) increases the electrophilicity of the  $P^G$  atom, (ii) weakens the  $O^{3B}-P^G$  bond, (iii) promotes the transition of  $\gamma$ -phosphate to a more planar conformation, and (iv) inevitably affects the coordination of the  $Mg^{2+}$  ion by displacing the  $O^{1G}$  atom. The increase of local positive charge at  $[Ser/Thr]^{K+1}$  – after  $O^{1G}$  is moved aside by the stimulator – promotes the relocation of proton from  $[Ser/Thr]^{K+1}$  to  $Asp^{WB}$ .

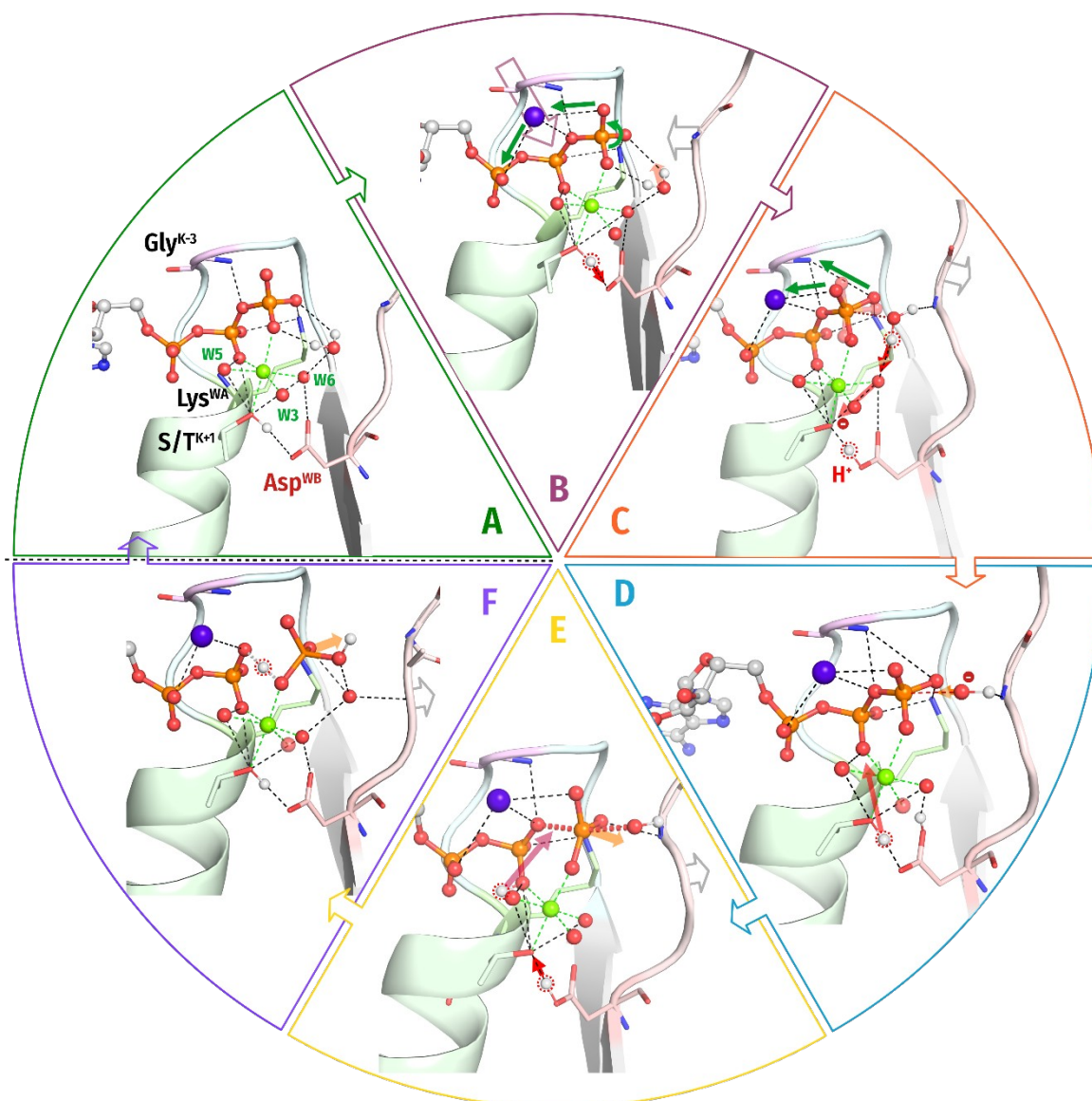


Figure 4.6: The common scheme of stimulated hydrolysis in P-loop NTPases.

Empty purple arrow, movement of the stimulator (shown as a purple sphere); gray arrow, movement of the WB-crest; green arrows, twist and planarization of  $\gamma$ -phosphate; red arrows, proton displacements; orange arrows, detachment of  $P_i$ . The following crystal structures were used as templates: panels A, B – PDB ID 1FMW [214]; panels C-E – PDB ID 1VOM [215]; panel F – PDB ID 4PFP [60].

- C) The anionic  $[\text{Ser/Thr}]^{\text{K}+1}$  alkoxide withdraws a proton from the polarized  $W_{\text{cat}}$  molecule via intermediate proton carriers. Here we depicted the simplest proton route as envisioned for TRAFAC NTPases (see Fig. 4.5D-F and Section 4.2.2.2). More complex proton routes via  $W_{\text{cat}}$ -coordinating Glu/Asp residues, as found in other classes of NTPases, are indicated by red dashed lines in Fig. 1.7-1.9, 3.8A, 3.11 and differently shaded red arrows in Fig. 4.2, 4.5A-C.
- D) The resulting  $\text{OH}^-_{\text{cat}}$ , stabilized/polarized by its ligands, attacks the  $P^{\text{G}}$  atom. Although the simplified diagram in Fig. 4.6D shows only one stabilizing interaction of  $\text{OH}^-_{\text{cat}}$  with the HN group of the WB-crest residue, several ligands are usually involved in the stabilization, see Fig. 1.2D-E, 1.4C-D, 1.6-1.9. During

this step, the proton stays on Asp<sup>WB</sup>. The formation of a covalent bond between OH<sup>-cat</sup> and P<sup>G</sup> increases the planarization of  $\gamma$ -phosphate; its oxygen atoms repel the  $\beta$ -phosphate oxygen atoms, resulting in a lengthening of the O<sup>3B</sup>-P<sup>G</sup> bond. With the inversion of  $\gamma$ -phosphate, increase in the O<sup>3B</sup>-P<sup>G</sup> distance, and  $\gamma$ -phosphate moving away from  $\beta$ -phosphate, HN<sup>K-3</sup> detaches from  $\gamma$ -phosphate and, together with Lys<sup>WA</sup>, Mg<sup>2+</sup> and the stimulator, neutralizes the negative charge appearing on the O<sup>3B</sup> atom, thereby lowering the activation barrier. In addition, the negative charge on  $\beta$ -phosphate attracts a proton from [Ser/Thr]<sup>K+1</sup>.

- E) The proton that comes from [Ser/Thr]<sup>K+1</sup> forms a new short H-bond between  $\beta$ - and  $\gamma$ -phosphate [60,98,323], which further stabilizes the negative charge on the  $\beta$ -phosphate. The proton at Asp<sup>WB</sup> relocates to [Ser/Thr]<sup>K+1</sup>.
- F) The H-bond between  $\beta$ - and  $\gamma$ -phosphate gradually dissociates as H<sub>2</sub>PO<sub>4</sub><sup>2-</sup> leaves the catalytic site. The departure of H<sub>2</sub>PO<sub>4</sub><sup>2-</sup> is an exergonic reaction that may be coupled to conformational changes, detachment of the activating partner from the WB-crest, and useful mechanical work.



## 5. Conclusions

1. A comparative structural analysis of more than 3,100 available structures of catalytic sites of P-loop NTPases with bound Mg-NTPs or their analogs was carried out using an originally developed computational approach. This approach allows a nearly sequence-independent analysis of thousands related or similar catalytic sites by focusing on coordinates of substrate atoms and a few key conserved residues.
2. A comparative structural analysis of more than 200 catalytic sites of P-loop NTPases of various families with bound TS analogs showed that the configurations of most of these catalytic sites coincide, which suggests the existence of a common catalytic mechanism. However, the same analysis also revealed that in some of the structures containing NDP:AlF<sub>4</sub><sup>-</sup> as a TS analog, the AlF<sub>4</sub><sup>-</sup> moiety forms not one, but two bonds with the Mg<sup>2+</sup> cofactor. As a result, the coordination bond between Mg<sup>2+</sup> and its physiological ligand, the strictly conserved [Ser/Thr]<sup>K+1</sup> residue of the Walker A motif, might be lost. As a consequence, the [Ser/Thr]<sup>K+1</sup> residue is distanced from the strictly conserved Asp<sup>WB</sup> of the Walker B motif, which prevents the formation of the H-bond between them. In several first-resolved, paradigmatic structures of P-loop NTPases, the AlF<sub>4</sub><sup>-</sup> moieties are incorrectly bound and the catalytic sites are distorted, which may explain why the comprehensive structural similarity of the TS-like structures of P-loop NTPases of different classes has not been noticed before.
3. To assess the possibility of a transient H-bond formation between **HN**<sup>K-3</sup> and O<sup>2G</sup> of  $\gamma$ -phosphate (or its structural analog), as observed in previous MD simulations [102], we have measured the respective distances in all the analyzed 3100 catalytic sites of P-loop NTPases. In all groups of complexes, there is a fraction with distances shorter than 3.4 Å between **HN**<sup>K-3</sup> and the nearest O<sup>G</sup> atom or its structural analog. However, in the case of the structures with TS analogs bound, H-bond compatible **HN**<sup>K-3</sup>—O<sup>G</sup> distances were observed in most of the structures, which suggests involvement of this bond in stabilization of the transition state.
4. Concerning the possibility of twisting the gamma-phosphate group by the stimulator(s), as revealed by earlier MD simulations [102], we found that seemingly different interactions between completely distinct stimulatory moieties (Arg/Asn/Lys residues, or K<sup>+</sup>/Na<sup>+</sup> ions, or LSGGQ/E motifs) come down to only two stimulatory patterns. In most cases, at least one stimulator links the O<sup>2A</sup> atom of  $\alpha$ -phosphate with the O<sup>3G</sup> atom of  $\gamma$ -phosphate, which requires counterclockwise twist of the  $\gamma$ -phosphate group. In other cases, stimulators only interact with the  $\gamma$ -phosphate group. In general, the only common feature of all the

identified stimulators seems to be the ability to enter such a mechanistic interaction with the  $\gamma$ -phosphate group, which may enable its twist/rotation. Hence, the insertion of a stimulator appears to lead to the twist of  $\gamma$ -phosphate and shortening of the  $\text{HN}^{\text{K}-3}\text{-O}^{2\text{G}}$  distance in diverse NTP-containing P-loop NTPases in support of the earlier MD simulation data obtained in our group ([102]).

5. We have found that in structures with properly bound TS analogs,  $\text{Asp}^{\text{WB}}$  and  $[\text{Ser/Thr}]^{\text{K}+1}$  are connected by short H-bonds, which are as short as 2.5 Å on average in the presence of  $\text{ADP}:\text{AlF}_4^-$ . Apart from this, potential proton pathways connect  $W_{\text{cat}}$  with one of the  $\text{Mg}^{2+}$  ligands in TS-like structures of all classes except the TRAFAC class. For the TRAFAC class NTPases we could also reconstruct tentative proton pathways by comparing substrate- and TS analog-containing structures. The distance between neighboring ligands in the coordination shell of  $\text{Mg}^{2+}$  is 2.9-3.0 Å, which implies the possibility of proton exchange between all of them. Hence, in P-loop fold NTPases,  $W_{\text{cat}}$  is connected by a proton pathway to  $\text{Asp}^{\text{WB}}$  via  $[\text{Ser/Thr}]^{\text{K}+1}$ , the  $\text{Mg}^{2+}$  ligand.
6. Building on the common structural elements found, we propose a novel common catalytic mechanism for P-loop NTPases. The key feature of the proposed mechanism is the activation-induced convergence of proton affinities of  $\text{Asp}^{\text{WB}}$  and  $[\text{Ser/Thr}]^{\text{K}+1}$  that enables proton transfer from  $[\text{Ser/Thr}]^{\text{K}+1}$  to  $\text{Asp}^{\text{WB}}$ . The resulting alkoxide, withdraws a proton from  $W_{\text{cat}}$  and converts it into a strong nucleophile  $\text{OH}^-_{\text{cat}}$  that attacks the  $\gamma$ -phosphate of NTP. When the  $\gamma$ -phosphate breaks away, the trapped proton at  $\text{Asp}^{\text{WB}}$  passes by the Grotthuss relay via  $[\text{Ser/Thr}]^{\text{K}+1}$  to  $\beta$ -phosphate and compensates for its developing negative charge that is thought to be the culprit of the activation barrier of hydrolysis. The proposed mechanism explains the strict evolutionary conservation of  $[\text{Ser/Thr}]^{\text{K}+1}$  and  $\text{Asp}^{\text{WB}}$ , neither of which has been ascribed a specific catalytic function so far.

## 6. Outlook

As argued elsewhere, the ability to provide strong acids and bases timely is important for enzyme catalysis [306]. Specifically, enzymes were suggested to generate strong bases or acids – precisely when required – by transiently altering the length of relevant H-bonds [205,316,324–330]. In the case of P-loop NTPases, the mechanism is quite simple: the free energies of (i) substrate binding and (ii) interaction with the activation partner are used to close and constrict the catalytic site, which shortens the H-bond between  $[\text{Ser/Thr}]^{\text{K}+1}$  and  $\text{Asp}^{\text{WB}}$  and evens their functional pK values. Eventually, after  $\gamma$ -phosphate is rotated by the stimulator, proton relocates from  $[\text{Ser/Thr}]^{\text{K}+1}$  to  $\text{Asp}^{\text{WB}}$  yielding an alkoxide of serine/threonine as a strong nucleophile for  $W_{\text{cat}}$ .

The proposed mechanism of proton transfer from  $W_{\text{cat}}$  to  $\text{Asp}^{\text{WB}}$  via  $\text{Mg}^{2+}$ -coordinating ligands brings the P-loop NTPases into the general context of other Mg-dependent hydrolases and transferases [51,52,86,318,331,332] as P-loop NTPases do use aspartates as catalytic bases just like most of these enzymes. The full range of theoretical approaches developed for such enzymes can now be applied to P-loop NTPases.

Our tentative identification of the anionic  $[\text{Ser/Thr}]^{\text{K}+1}$  alkoxide as the proton acceptor from  $W_{\text{cat}}$  brings P-loop NTPases into another broad context of enzymes that generate a strong nucleophile by stripping a Ser/Thr residue of its proton [86,234,318,319]. These are numerous families of serine proteases where, within apparently similar, but non-homologous “catalytic triads”, proton is transferred from the catalytic Ser to proton-accepting Asp via a histidine residue that serves as a von Grothuss-type proton carrier. It is less known that the proton-accepting Asp residue of serine proteases is usually H-bonded to another conserved Thr/Ser residue [325,333,334]. Also, the catalytic Ser/Thr of eukaryotic protein kinases gives its proton to the conserved Asp [260,335,336], which, in turn, is also H-bonded to another conserved Ser/Thr residue.

In essence, we argue that P-loop NTPases use the same (bio)chemical strategies to produce a strong nucleophile as many other enzymes do; there is no reason to consider them different or special in this respect.

Notably, serine proteases, serine-threonine kinases, Mg-dependent hydrolases and transferases, as well as P-loop NTPases considered here, form the largest known enzyme families. Many of them use H-bonded  $[\text{Asp/Glu}]$ – $[\text{Ser/Thr}]$  functional modules in different roles, which feature they share with membrane-embedded PRC and BR, see Fig. 4.3 and [337]. We believe that further search for structural features common to distinct enzyme families is a very promising task.

## 7. Bibliography

- [1] H.-B. Guo et al., *AlphaFold2 Models Indicate That Protein Sequence Determines Both Structure and Dynamics*, *Sci Rep* **12**, 1 (2022).
- [2] W. Erdmann and Ł. Kaczmarek, *Tardigrades in Space Research - Past and Future*, *Orig Life Evol Biosph* **47**, 545 (2017).
- [3] *Crystallography: Protein Data Bank*, *Nature New Biology* **233**, 42 (1971).
- [4] H. Berman, K. Henrick, and H. Nakamura, *Announcing the Worldwide Protein Data Bank*, *Nat Struct Biol* **10**, 980 (2003).
- [5] S. K. Burley et al., *RCSB Protein Data Bank: Powerful New Tools for Exploring 3D Structures of Biological Macromolecules for Basic and Applied Research and Education in Fundamental Biology, Biomedicine, Biotechnology, Bioengineering and Energy Sciences*, *Nucleic Acids Res* **49**, 437 (2021).
- [6] T. Nakane et al., *Single-Particle Cryo-EM at Atomic Resolution*, *Nature* **587**, 7832 (2020).
- [7] K. M. Yip, N. Fischer, E. Paknia, A. Chari, and H. Stark, *Atomic-Resolution Protein Structure Determination by Cryo-EM*, *Nature* **587**, 7832 (2020).
- [8] J. Jumper et al., *Highly Accurate Protein Structure Prediction with AlphaFold*, *Nature* **596**, 7873 (2021).
- [9] M. Baek et al., *Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network*, *Science* **373**, 871 (2021).
- [10] A. Perrakis and T. K. Sixma, *AI Revolutions in Biology*, *EMBO Reports* **22**, e54046 (2021).
- [11] Z. Lin et al., *Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model*, *Science* **379**, 1123 (2023).
- [12] D. A. Suplatov, V. K. Arzhanik, and V. K. Svedas, *Comparative Bioinformatic Analysis of Active Site Structures in Evolutionarily Remote Homologues of  $\alpha,\beta$ -Hydrolase Superfamily Enzymes*, *Acta Naturae* **3**, 93 (2011).
- [13] D. J. Blum, Y. H. Ko, and P. L. Pedersen, *Mitochondrial ATP Synthase Catalytic Mechanism: A Novel Visual Comparative Structural Approach Emphasizes Pivotal Roles for Mg(2)(+) and P-Loop Residues in Making ATP*, *Biochemistry* **51**, 1532 (2012).
- [14] D. Suplatov, D. Shalaeva, E. Kirilin, V. Arzhanik, and V. Svedas, *Bioinformatic Analysis of Protein Families for Identification of Variable Amino Acid Residues Responsible for Functional Diversity*, *J Biomol Struct Dyn* **32**, 75 (2014).
- [15] L. Holm, *Using Dali for Protein Structure Comparison*, in *Structural Bioinformatics: Methods and Protocols*, edited by Z. Gáspári (Springer US, New York, NY, 2020), pp. 29–42.
- [16] J. Mistry et al., *Pfam: The Protein Families Database in 2021*, *Nucleic Acids Research* **49**, D412 (2021).
- [17] M. G. Rossmann, D. Moras, and K. W. Olsen, *Chemical and Biological Evolution of a Nucleotide-Binding Protein*, *Nature* **250**, 194 (1974).
- [18] S. K. Hanks, A. M. Quinn, and T. Hunter, *The Protein Kinase Family: Conserved Features and Deduced Phylogeny of the Catalytic Domains*, *Science* **241**, 42 (1988).

- [19] R. G. Brennan and B. W. Matthews, *The Helix-Turn-Helix DNA Binding Motif*, J Biol Chem **264**, 1903 (1989).
- [20] P. Bork, L. Holm, and C. Sander, *The Immunoglobulin Fold. Structural Classification, Sequence Patterns and Common Core*, J Mol Biol **242**, 309 (1994).
- [21] J. R. Lamb, S. Tugendreich, and P. Hieter, *Tetratrico Peptide Repeat Interactions: To TPR or Not to TPR?*, Trends Biochem Sci **20**, 257 (1995).
- [22] N. Zeytuni and R. Zarivach, *Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module*, Structure **20**, 397 (2012).
- [23] R. R. Copley and P. Bork, *Homology among (Betaalpha)(8) Barrels: Implications for the Evolution of Metabolic Pathways*, J Mol Biol **303**, 627 (2000).
- [24] N. Nagano, C. A. Orengo, and J. M. Thornton, *One Fold with Many Functions: The Evolutionary Relationships between TIM Barrel Families Based on Their Sequences, Structures and Functions*, Journal of Molecular Biology **321**, 741 (2002).
- [25] A. G. Murzin, *OB(Oligonucleotide/Oligosaccharide Binding)-Fold: Common Structural and Functional Solution for Non-Homologous Sequences*, EMBO J **12**, 861 (1993).
- [26] J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay, *Distantly Related Sequences in the Alpha- and Beta-Subunits of ATP Synthase, Myosin, Kinases and Other ATP-Requiring Enzymes and a Common Nucleotide Binding Fold*, The EMBO Journal **1**, 945 (1982).
- [27] M. Saraste, P. R. Sibbald, and A. Wittinghofer, *The P-Loop--a Common Motif in ATP- and GTP-Binding Proteins*, Trends Biochem Sci **15**, 430 (1990).
- [28] A. E. Gorbalenya and E. Koonin, *Helicases: Amino Acid Sequence Comparisons and Structure-Function Relationships*, Current Opinion in Structural Biology **3**, 419 (1993).
- [29] C. A. Smith and I. Rayment, *Active Site Comparisons Highlight Structural Similarities between Myosin and Other P-Loop Proteins*, Biophys J **70**, 1590 (1996).
- [30] A. F. Neuwald, L. Aravind, J. L. Spouge, and E. V. Koonin, *AAA +: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes*, 27 (1999).
- [31] E. Muneyuki, H. Noji, T. Amano, T. Masaike, and M. Yoshida, *F<sub>0</sub>F<sub>1</sub>-ATP Synthase: General Structural Features of "ATP-Engine" and a Problem on Free Energy Transduction*, Biochimica et Biophysica Acta - Bioenergetics **1458**, 467 (2000).
- [32] D. D. Leipe, Y. I. Wolf, E. V. Koonin, and L. Aravind, *Classification and Evolution of P-Loop GTPases and Related ATPases*, Journal of Molecular Biology **317**, 41 (2002).
- [33] L. M. Iyer, K. S. Makarova, E. V. Koonin, and L. Aravind, *Comparative Genomics of the FtsK-HerA Superfamily of Pumping ATPases: Implications for the Origins of Chromosome Segregation, Cell Division and Viral Capsid Packaging*, Nucleic Acids Research **32**, 5260 (2004).
- [34] A. Wittinghofer and I. R. Vetter, *Structure-Function Relationships of the G Domain, a Canonical Switch Motif*, Annual Review of Biochemistry **80**, 943 (2011).
- [35] A. M. Burroughs and L. Aravind, *The Origin and Evolution of Release Factors: Implications for Translation Termination, Ribosome Rescue, and Quality Control Pathways*, Int J Mol Sci **20**, (2019).
- [36] L. M. Longo, J. Jabłońska, P. Vyas, M. Kanade, R. Kolodny, N. Ben-Tal, and D. S. Tawfik, *On the Emergence of P-Loop NTPase and Rossmann Enzymes from a Beta-Alpha-Beta Ancestral Fragment*, ELife **9**, e64415 (2020).

- [37] A. Krishnan, A. M. Burroughs, L. M. Iyer, and L. Aravind, *Comprehensive Classification of ABC ATPases and Their Functional Radiation in Nucleoprotein Dynamics and Biological Conflict Systems*, *Nucleic Acids Research* **48**, 10045 (2020).
- [38] R. D. Schaeffer, Y. Liao, H. Cheng, and N. V. Grishin, *ECOD: New Developments in the Evolutionary Classification of Domains*, *Nucleic Acids Res* **45**, D296 (2017).
- [39] R. D. Finn et al., *The Pfam Protein Families Database: Towards a More Sustainable Future*, *Nucleic Acids Res* **44**, D279 (2016).
- [40] D. D. Leipe, E. V. Koonin, and L. Aravind, *Evolution and Classification of P-Loop Kinases and Related Proteins*, *Journal of Molecular Biology* **333**, 781 (2003).
- [41] A. N. Lupas, C. P. Ponting, and R. B. Russell, *On the Evolution of Protein Folds: Are Similar Motifs in Different Protein Folds the Result of Convergence, Insertion, or Relics of an Ancient Peptide World?*, *J Struct Biol* **134**, 191 (2001).
- [42] C. P. Ponting and R. R. Russell, *The Natural History of Protein Domains*, *Annual Review of Biophysics and Biomolecular Structure* **31**, 45 (2002).
- [43] J. Söding and A. N. Lupas, *More than the Sum of Their Parts: On the Evolution of Proteins from Peptides*, *BioEssays* **25**, 837 (2003).
- [44] J. A. Ranea, A. Sillero, J. M. Thornton, and C. A. Orengo, *Protein Superfamily Evolution and the Last Universal Common Ancestor (LUCA)*, *J. Mol. Evol* **63**, 513 (2006).
- [45] V. Alva, J. Soding, and A. N. Lupas, *A Vocabulary of Ancient Peptides at the Origin of Folded Proteins*, *Elife* **4**, e09410 (2015).
- [46] T. F. Cour, J. Nyborg, S. Thirup, and B. F. Clark, *Structural Details of the Binding of Guanosine Diphosphate to Elongation Factor Tu from E. Coli as Studied by X-Ray Crystallography*, *EMBO J* **4**, 2385 (1985).
- [47] F. Journak, *Induction of Elongation Factor Tu-GDP Crystal Polymorphism by Polyethylene Glycol Contaminants*, *J Mol Biol* **185**, 215 (1985).
- [48] V. N. Hingorani and Y. K. Ho, *A Structural Model for the Alpha-Subunit of Transducin. Implications of Its Role as a Molecular Switch in the Visual Signal Transduction Mechanism*, *FEBS Lett* **220**, 15 (1987).
- [49] U. Krengel, I. Schlichting, A. Scherer, R. Schumann, M. Frech, J. John, W. Kabsch, E. F. Pai, and A. Wittinghofer, *Three-Dimensional Structures of H-Ras P21 Mutants: Molecular Basis for Their Inability to Function as Signal Switch Molecules*, *Cell* **62**, 539 (1990).
- [50] M. Kanade, S. Chakraborty, S. S. Shelke, and P. Gayathri, *A Distinct Motif in a Prokaryotic Small Ras-Like GTPase Highlights Unifying Features of Walker B Motifs in P-Loop NTPases*, *Journal of Molecular Biology* **432**, 5544 (2020).
- [51] J. R. Knowles, *Enzyme-Catalyzed Phosphoryl Transfer Reactions*, *Annu Rev Biochem* **49**, 877 (1980).
- [52] F. H. Westheimer, *Why Nature Chose Phosphates*, *Science* **235**, 1173 (1987).
- [53] Z. A. Shabarova and A. A. Bogdanov, *Advanced Organic Chemistry of Nucleic Acids*, (1994).
- [54] M. W. Bowler, M. J. GCliff, J. P. Waltho, and G. M. Blackburn, *Why Did Nature Select Phosphate for Its Dominant Roles in Biology?*, *New Journal of Chemistry* **34**, 784 (2010).
- [55] S. J. Admiraal and D. Herschlag, *Mapping the Transition State for ATP Hydrolysis: Implications for Enzymatic Catalysis*, *Chemistry & Biology* **2**, 729 (1995).

- [56] K. A. Maegley, S. J. Admiraal, and D. Herschlag, *Ras-Catalyzed Hydrolysis of GTP: A New Perspective from Model Studies.*, Proc. Natl. Acad. Sci. U.S.A. **93**, 8160 (1996).
- [57] J. K. Lassila, J. G. Zalatan, and D. Herschlag, *Biological Phosphoryl-Transfer Reactions: Understanding Mechanism and Catalysis*, Annu. Rev. Biochem. **80**, 669 (2011).
- [58] I. R. Vetter and A. Wittinghofer, *Signal Transduction - The Guanine Nucleotide-Binding Switch in Three Dimensions*, Science **294**, 1299 (2001).
- [59] K. Nam, J. Pu, and M. Karplus, *Trapping the ATP Binding State Leads to a Detailed Understanding of the F1-ATPase Mechanism*, Proc Natl Acad Sci U S A **111**, 17851 (2014).
- [60] P. Llinas et al., *How Actin Initiates the Motor Activity of Myosin*, Dev Cell **33**, 401 (2015).
- [61] A. Scrima and A. Wittinghofer, *Dimerisation-Dependent GTPase Reaction of MnmE: How Potassium Acts as GTPase-Activating Element.*, The EMBO Journal **25**, 2940 (2006).
- [62] D. E. Coleman, A. M. Berghuis, E. Lee, M. E. Linder, A. G. Gilman, and S. R. Sprang, *Structures of Active Conformations of G $\alpha$ 1 and the Mechanism of GTP Hydrolysis*, Science **265**, 1405 (1994).
- [63] J. Sondek, D. G. Lambright, J. P. Noel, H. E. Hamm, and P. B. Sigler, *GTPase Mechanism of Gproteins from the 1.7-Å Crystal Structure of Transducin  $\alpha$  - GGDP AIF-4*, Nature **372**, 276 (1994).
- [64] K. Scheffzek, M. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz, and A. Wittinghofer, *The Ras-RasGAP Complex: Structural Basis for GTPase Activation and Its Loss in Oncogenic Ras Mutants*, Science **277**, 333 (1997).
- [65] T. Ogura, S. W. Whiteheart, and A. J. Wilkinson, *Conserved Arginine Residues Implicated in ATP Hydrolysis, Nucleotide-Sensing, and Inter-Subunit Interactions in AAA and AAA+ ATPases*, Journal of Structural Biology **146**, 106 (2004).
- [66] M.-R. Ash, M. J. Maher, J. Mitchell Guss, and M. Jormakka, *The Cation-Dependent G-Proteins: In a Class of Their Own.*, FEBS Letters **586**, 2218 (2012).
- [67] P. Wendler, S. Ciniawsky, M. Kock, and S. Kube, *Structure and Function of the AAA+ Nucleotide Binding Pocket*, Biochimica et Biophysica Acta - Molecular Cell Research **1823**, 2 (2012).
- [68] Y. Jin and G. M. B. R.W. Molt Jr., *Metal Fluorides: Tools for Structural and Computational Analysis of Phosphoryl Transfer Enzymes*, Top, Curr. Chem. (Cham **375**, 36 (2017).
- [69] R. Gasper and F. Wittinghofer, *The Ras Switch in Structural and Historical Perspective*, Biol Chem **401**, 143 (2019).
- [70] D. L. Graham, P. N. Lowe, G. W. Grime, M. Marsh, K. Rittinger, S. J. Smerdon, S. J. Gamblin, and J. F. Eccleston, *MgF(3)(-) as a Transition State Analog of Phosphoryl Transfer*, Chem Biol **9**, 375 (2002).
- [71] A. Wittinghofer, *Signaling Mechanistics: Aluminum Fluoride for Molecule of the Year*, Curr. Biol **7**, 682 (1997).
- [72] R. I. Menz, J. E. Walker, and A. G. Leslie, *Structure of Bovine Mitochondrial F(1)-ATPase with Nucleotide Bound to All Three Catalytic Sites: Implications for the Mechanism of Rotary Catalysis*, Cell **106**, 331 (2001).

- [73] D. R. Davies and W. G. Hol, *The Power of Vanadate in Crystallographic Investigations of Phosphoryl Transfer Enzymes*, FEBS Lett **577**, 315 (2004).
- [74] Y. Jin, N. G. Richards, J. P. Waltho, and G. M. Blackburn, *Metal fluorides as analogues for studies on phosphoryl transfer enzymes*, Angew, Chem. Int. Ed. Engl **56**, 4110 (2017).
- [75] N. J. Baxter, G. M. Blackburn, J. P. Marston, A. M. Hounslow, M. J. Cliff, W. Bermel, N. H. Williams, F. Hollfelder, D. E. Wemmer, and J. P. Waltho, *Anionic Charge Is Prioritized over Geometry in Aluminum and Magnesium Fluoride Transition State Analogs of Phosphoryl Transfer Enzymes*, J Am Chem Soc **130**, 3952 (2008).
- [76] N. Zhang and M. Buck, *Formation of MgF<sub>3</sub><sup>-</sup>-Dependent Complexes between an AAA(+) ATPase and Sigma(54)*, FEBS Open Bio **2**, 89 (2012).
- [77] M. Chaney et al., *Binding of Transcriptional Activators to Sigma 54 in the Presence of the Transition State Analog ADP-Aluminum Fluoride: Insights into Activator Mechanochemical Action*, Genes Dev **15**, 2282 (2001).
- [78] T. Higashijima, K. M. Ferguson, P. C. Sternweis, E. M. Ross, M. D. Smigel, and A. G. Gilman, *The Effect of Activating Ligands on the Intrinsic Fluorescence of Guanine Nucleotide-Binding Regulatory Proteins*, J Biol Chem **262**, 752 (1987).
- [79] L. Gremer, B. Gilsbach, M. R. Ahmadian, and A. Wittinghofer, *Fluoride complexes of oncogenic Ras mutants to study the Ras-RasGap interaction*, Biol, Chem **389**, 1163 (2008).
- [80] R. Mittal, M. R. Ahmadian, R. S. Goody, and A. Wittinghofer, *Formation of a Transition-State Analog of the Ras GTPase Reaction by Ras-GDP, Tetrafluoroaluminate, and GTPase-Activating Proteins*, Science **273**, 115 (1996).
- [81] P. P. Chakrabarti, O. Daumke, Y. Suveyzdis, C. Kotting, K. Gerwert, and A. Wittinghofer, *Insight into Catalysis of a Unique GTPase Reaction by a Combined Biochemical and FTIR Approach*, J Mol Biol **367**, 983 (2007).
- [82] I. Schlichting and J. Reinstein, *PH Influences Fluoride Coordination Number of the AlFx Phosphoryl Transfer Transition State Analog*, Nat Struct Biol **6**, 721 (1999).
- [83] B. Antonny, M. Sukumar, J. Bigay, M. Chabre, and T. Higashijima, *The mechanism of aluminum-independent G-protein activation by fluoride and magnesium. <sup>31</sup>P NMR spectroscopy and fluorescence kinetic studies*, J Biol Chem **268**, 2393 (1993).
- [84] G. M. Blackburn, J. Cherfils, G. P. Moss, N. G. J. Richards, J. P. Waltho, N. H. Williams, and A. Wittinghofer, *How to Name Atoms in Phosphates, Polyphosphates, Their Derivatives and Mimics, and Transition State Analogues for Enzyme-Catalysed Phosphoryl Transfer Reactions (IUPAC Recommendations 2016)*, Pure Appl. Chem **89**, 653 (2017).
- [85] Z. A. Shabarova and A. A. Bogdanov, *Advanced Organic Chemistry of Nucleic Acids* (VCH, Weinheim, 1994).
- [86] W. W. Cleland and A. C. Hengge, *Enzymatic Mechanisms of Phosphate and Sulfate Transfer*, Chem Rev **106**, 3252 (2006).
- [87] F. A. Kiani and S. Fischer, *Comparing the Catalytic Strategy of ATP Hydrolysis in Biomolecular Motors*, Phys Chem Chem Phys **18**, 20219 (2016).
- [88] R. Langen, T. Schweins, and A. Warshel, *On the mechanism of guanosine triphosphate hydrolysis in ras p21 proteins*, Biochemistry **31**, 8691 (1992).
- [89] T. Schweins, R. Langen, and A. Warshel, *Why Have Mutagenesis Studies Not Located the General Base in Ras P21*, Nat Struct Biol **1**, 476 (1994).



- [90] T. Schweins, M. Geyer, K. Scheffzek, A. Warshel, H. R. Kalbitzer, and A. Wittinghofer, *Substrate-Assisted Catalysis as a Mechanism for GTP Hydrolysis of P21ras and Other GTP-Binding Proteins*, *Nat Struct Biol* **2**, 36 (1995).
- [91] K. Braig, R. I. Menz, M. G. Montgomery, A. G. Leslie, and J. E. Walker, *Structure of Bovine Mitochondrial F(1)-ATPase Inhibited by Mg(2+) ADP and Aluminium Fluoride*, *Structure* **8**, 567 (2000).
- [92] T. M. Glennon, J. Villa, and A. Warshel, *How Does GAP Catalyze the GTPase Reaction of Ras?: A Computer Simulation Study*, *Biochemistry* **39**, 9641 (2000).
- [93] B. R. Prasad, N. V. Plotnikov, J. Lameira, and A. Warshel, *Quantitative Exploration of the Molecular Origin of the Activation of GTPase*, *Proc, Natl. Acad. Sci. USA* **110**, 20509 (2013).
- [94] S. C. Kamerlin, P. K. Sharma, R. B. Prasad, and A. Warshel, *Why Nature Really Chose Phosphate*, *Q. Rev. Biophys.* **46**, 1 (2013).
- [95] C. Kötting, A. Kallenbach, Y. Suveyzdis, A. Wittinghofer, and K. Gerwert, *The GAP Arginine Finger Movement into the Catalytic Site of Ras Increases the Activation Entropy*, *Proceedings of the National Academy of Sciences* **105**, 6260 (2008).
- [96] G. M. Blackburn, Y. Jin, N. G. Richards, and J. P. Waltho, *Metal Fluorides as Analogs for Studies on Phosphoryl Transfer Enzymes*, *Angew Chemistry International Edition* **1** (2016).
- [97] Y. Jin, R. W. Molt, J. P. Waltho, N. G. Richards, and G. M. Blackburn, *<sup>19</sup>F NMR and DFT Analysis Reveal Structural and Electronic Transition State Features for RhoA-Catalyzed GTP Hydrolysis*, *Angew. Chem. Int. Ed. Engl.* **55**, 3318 (2016).
- [98] E. P. R.W. Molt Jr. and Y. Jin, *A GAP-GTPase-GDP-Pi Intermediate Crystal Structure Analyzed by DFT Shows GTP Hydrolysis Involves Serial Proton Transfers*, *Chemistry* **25**, 8484 (2019).
- [99] T. Rudack, F. Xia, J. Schlitter, C. Kötting, and K. Gerwert, *Ras and GTPase-Activating Protein (GAP) Drive GTP into a Precatalytic State as Revealed by Combining FTIR and Biomolecular Simulations*, *Proceedings of the National Academy of Sciences* **109**, 15295 (2012).
- [100] D. Mann, C. Teuber, S. A. Tennigkeit, G. Schroter, K. Gerwert, and C. Kötting, *Mechanism of the Intrinsic Arginine Finger in Heterotrimeric G Proteins*, *Proc. Natl. Acad. Sci. USA* **113**, E8041 (2016).
- [101] K. Gerwert, D. Mann, and C. Kötting, *Common Mechanisms of Catalysis in Small and Heterotrimeric GTPases and Their Respective GAPs*, *Biological Chemistry* **398**, 523 (2017).
- [102] D. N. Shalaeva, D. A. Cherepanov, M. Y. Galperin, A. V. Golovin, and A. Y. Mulikidjanian, *Evolution of Cation Binding in the Active Sites of P-Loop Nucleoside Triphosphatases in Relation to the Basic Catalytic Mechanism*, *Elife* **7**, (2018).
- [103] K. Yamanaka, J. Hwang, and M. Inouye, *Characterization of GTPase Activity of TrmE, a Member of a Novel GTPase Superfamily, from Thermotoga Maritima*, *J Bacteriol* **182**, 7078 (2000).
- [104] S. Meyer, S. Bohme, A. Kruger, H.-J. Steinhoff, J. P. Klare, and A. Wittinghofer, *Kissing G Domains of MnmE Monitored by X-Ray Crystallography and Pulse Electron Paramagnetic Resonance Spectroscopy*, *PLoS Biol* **7**, e1000212 (2009).

- [105] S. Bohme, S. Meyer, A. Kruger, H. J. Steinhoff, A. Wittinghofer, and J. P. Klare, *Stabilization of G Domain Conformations in the TRNA-Modifying MnmE-GidA Complex Observed with Double Electron Electron Resonance Spectroscopy*, *J Biol Chem* **285**, 16991 (2010).
- [106] B. Anand, P. Surana, and B. Prakash, *Deciphering the Catalytic Machinery in 30S Ribosome Assembly GTPase YqeH*, *PloS One* **5**, e9944 (2010).
- [107] J. A. Ballestros and H. Weinstein, *Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors*, *Methods Neurosci.* **25**, 366 (1995).
- [108] M. V. Milburn, L. Tong, A. M. deVos, A. Brunger, Z. Yamaizumi, S. Nishimura, and S. H. Kim, *Molecular Switch for Signal Transduction: Structural Differences between Active and Inactive Forms of Protooncogenic Ras Proteins*, *Science* **247**, 939 (1990).
- [109] M. I. Kozlova, D. N. Shalaeva, D. V. Dibrova, and A. Y. Mulkidjanian, *Common Mechanism of Activated Catalysis in P-Loop Fold Nucleoside Triphosphatases – United in Diversity*, *Biomolecules* **12**, 10 (2022).
- [110] M. Frech, J. John, V. Pizon, P. Chardin, A. Tavitian, R. Clark, F. McCormick, and A. Wittinghofer, *Inhibition of GTPase Activating Protein Stimulation of Ras-P21 GTPase by the Krev-1 Gene Product*, *Science* **249**, 169 (1990).
- [111] L. M. Iyer, D. D. Leipe, E. V. Koonin, and L. Aravind, *Evolutionary History and Higher Order Classification of AAA+ ATPases.*, *Journal of Structural Biology* **146**, 11 (2004).
- [112] N. D. Thomsen and J. M. Berger, *Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases*, *Mol Microbiol* **69**, 1071 (2008).
- [113] E. F. Pai, U. Krengel, G. A. Petsko, R. S. Goody, W. Kabsch, and A. Wittinghofer, *Refined Crystal Structure of the Triphosphate Conformation of H-Ras P21 at 1.35 Å Resolution: Implications for the Mechanism of GTP Hydrolysis*, *EMBO J* **9**, 2351 (1990).
- [114] M. P. Mueller and R. S. Goody, *Review: Ras GTPases and Myosin: Qualitative Conservation and Quantitative Diversification in Signal and Energy Transduction*, *Biopolymers* **105**, 422 (2016).
- [115] R. Mishra, S. Gara, S. Mishra, and B. Prakash, *Analysis of GTPases Carrying Hydrophobic Amino Acid Substitutions in Lieu of the Catalytic Glutamine: Implications for GTP Hydrolysis*, *Proteins* **59**, 332 (2005).
- [116] O. Fasano, E. De Vendittis, and A. Parmeggiani, *Hydrolysis of GTP by Elongation Factor Tu Can Be Induced by Monovalent Cations in the Absence of Other Effectors*, *J Biol Chem* **257**, 3145 (1982).
- [117] C. Maracci, F. Peske, E. Dannies, C. Pohl, and M. V. Rodnina, *Ribosome-Induced Tuning of GTP Hydrolysis by a Translational GTPase*, *Proc Natl Acad Sci U S A* **111**, 14418 (2014).
- [118] G. Chinali and A. Parmeggiani, *The Coupling with Polypeptide Synthesis of the GTPase Activity Dependent on Elongation Factor G*, *J. Biol. Chem* **255**, 7455 (1980).
- [119] R. Ivell, G. Sander, and A. Parmeggiani, *Modulation by Monovalent and Divalent Cations of the Guanosine-5'-Triphosphatase Activity Dependent on Elongation Factor Tu*, *Biochemistry* **20**, 6852 (1981).
- [120] A. Parmeggiani and G. Sander, *Properties and Regulation of the GTPase Activities of Elongation Factors Tu and G, and of Initiation Factor 2*, *Mol Cell Biochem* **35**, 129 (1981).

- [121] J. Voigt, G. Sander, K. Nagel, and A. Parmeggiani, *Effect of NH<sub>4</sub><sup>+</sup> and K<sup>+</sup> on the Activity of the Ribosomal Subunits in the EF-G- and EF-T-Dependent GTP Hydrolysis*, *Biochem Biophys Res Commun* **57**, 1279 (1974).
- [122] A. Y. Mulkidjanian, A. Y. Bychkov, D. V. Dibrova, M. Y. Galperin, and E. V. Koonin, *Origin of First Cells at Terrestrial, Anoxic Geothermal Fields.*, *Proceedings of the National Academy of Sciences of the United States of America* **109**, E821 (2012).
- [123] B. Kuhle and R. Ficner, *A Monovalent Cation Acts as Structural and Catalytic Cofactor in Translational GTPases*, *EMBO J* **33**, 2547 (2014).
- [124] R. M. Voorhees, T. M. Schmeing, A. C. Kelley, and V. Ramakrishnan, *The Mechanism for Activation of GTP Hydrolysis on the Ribosome*, *Science* **330**, 835 (2010).
- [125] L. J. Byrnes, A. Singh, K. Szeto, N. M. Benvin, J. P. O'Donnell, W. R. Zipfel, and H. Sondermann, *Structural Basis for Conformational Switching and GTP Loading of the Large G Protein Atlastin*, *EMBO J* **32**, 369 (2013).
- [126] M. Soundararajan et al., *Structural Diversity in the RGS Domain and Its Interaction with Heterotrimeric G Protein Alpha-Subunits*, *Proc Natl Acad Sci U S A* **105**, 6457 (2008).
- [127] N. J. Gay and J. E. Walker, *Homology between Human Bladder Carcinoma Oncogene Product and Mitochondrial ATP-Synthase*, *Nature* **301**, 262 (1983).
- [128] S. R. Sprang, *G proteins, effectors and GAPs: structure and mechanism*, *Curr Opin Struct Biol* **7**, 849 (1997).
- [129] A. Wittinghofer, *GTP and ATP hydrolysis in biology*, *Biopolymers* **105**, 419 (2016).
- [130] A. J. Fisher, C. A. Smith, J. B. Thoden, R. Smith, K. Sutoh, H. M. Holden, and I. Rayment, *X-Ray Structures of the Myosin Motor Domain of Dictyostelium Discoideum Complexed with MgADP.BeFx and MgADP.AlF<sub>4</sub>*, *Biochemistry* **34**, 8960 (1995).
- [131] M. A. Geeves, *Review: The ATPase Mechanism of Myosin and Actomyosin*, *Biopolymers* **105**, 483 (2016).
- [132] R. A. Cross, *Review: Mechanochemistry of the Kinesin-1 ATPase*, *Biopolymers* **105**, 476 (2016).
- [133] M. I. Kozlova, D. N. Shalaeva, D. V. Dibrova, and A. Y. Mulkidjanian, *Common Patterns of Hydrolysis Initiation in P-Loop Fold Nucleoside Triphosphatases*, *Biomolecules* **12**, 10 (2022).
- [134] G. Bange and I. Sinning, *SIMIBI Twins in Protein Targeting and Localization*, *Nature Structural and Molecular Biology* **20**, 776 (2013).
- [135] J. Lutkenhaus, *The ParA/MinD Family Puts Things in Their Place*, *Trends Microbiol* **20**, 411 (2012).
- [136] S. F. Ataide, N. Schmitz, K. Shen, A. Ke, S. Shan, J. A. Doudna, and N. Ban, *The Crystal Structure of the Signal Recognition Particle in Complex with Its Receptor*, *Science* **331**, 881 (2013).
- [137] S. Cheek, K. Ginalska, H. Zhang, and N. V. Grishin, *A Comprehensive Update of the Sequence and Structure Classification of Kinases*, *BMC Structural Biology* **5**, 1 (2005).
- [138] C. P. Kenyon, R. L. Roth, C. W. Van Der Westhuyzen, and C. J. Parkinson, *Conserved Phosphoryl Transfer Mechanisms within Kinase Families and the Role of the C8 Proton of ATP in the Activation of Phosphoryl Transfer*, *BMC Research Notes* **5**, 131 (2012).

- [139] E. B. Lansdon, I. H. Segel, and A. J. Fisher, *Ligand-Induced Structural Changes in Adenosine 5'-Phosphosulfate Kinase from Penicillium Chrysogenum*, *Biochemistry* **41**, 13672 (2002).
- [140] S. J. Kerns et al., *The Energy Landscape of Adenylate Kinase during Catalysis*, *Nat Struct Mol Biol* **22**, 124 (2015).
- [141] L. Aravind, L. M. Iyer, D. D. Leipe, and E. V. Koonin, *A Novel Family of P-Loop NTPases with an Unusual Phyletic Distribution and Transmembrane Segments Inserted within the NTPase Domain.*, *Genome Biology* **5**, R30 (2004).
- [142] M. Ammelburg, T. Frickey, and A. N. Lupas, *Classification of AAA+ Proteins*, *J Struct Biol* **156**, 2 (2006).
- [143] A. N. Lupas and J. Martin, *AAA Proteins*, *Curr Opin Struc Biol* **12**, 746 (2002).
- [144] J. M. Miller and E. J. Enemark, *Fundamental Characteristics of AAA+ Protein Family Structure and Function*, *Archaea* 2016, (2016).
- [145] R. C. Yu, P. I. Hanson, R. Jahn, and A. T. Brunger, *Structure of the ATP-Dependent Oligomerization Domain of N-Ethylmaleimide Sensitive Factor Complexed with ATP*, *Nat Struct Biol* **5**, 803 (1998).
- [146] M. R. Singleton, M. S. Dillingham, and D. B. Wigley, *Structure and Mechanism of Helicases and Nucleic Acid Translocases*, *Annual Review of Biochemistry* **76**, 23 (2007).
- [147] T. Sengoku, O. Nureki, A. Nakamura, S. Kobayashi, and S. Yokoyama, *Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa*, *Cell* **125**, 287 (2006).
- [148] M. Dean, A. Rzhetsky, and R. Allikmets, *The Human ATP-Binding Cassette (ABC) Transporter Superfamily*, *Genome Res* **11**, 1156 (2001).
- [149] D. C. Rees, E. Johnson, and O. Lewinson, *ABC Transporters: The Power to Change*, *Nat Rev Mol Cell Biol.* **10**, 218 (2009).
- [150] M. L. Oldham and J. Chen, *Snapshots of the Maltose Transporter during ATP Hydrolysis*, *Proc Natl Acad Sci U S A* **108**, 15152 (2011).
- [151] S. Hofmann et al., *Conformation Space of a Heterodimeric ABC Exporter under Turnover Conditions*, *Nature* **571**, 7766 (2019).
- [152] I. D. Kerr, *Sequence analysis of twin ATP binding cassette proteins involved in translational control, antibiotic resistance, and ribonuclease L inhibition*, *Biochem Biophys Res Commun* **315**, 166 (2004).
- [153] D. V. Dibrova, K. A. Konovalov, V. V. Perekhvatov, K. V. Skulachev, and A. Y. Mulkidjanian, *COGcollator: A Web Server for Analysis of Distant Relationships between Homologous Protein Families*, *Biol Direct* **12**, 29 (2017).
- [154] M. R. Lawson, W. Ma, M. J. Bellecourt, I. Artsimovitch, A. Martin, R. Landick, K. Schulten, and J. M. Berger, *Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein*, *Mol Cell* **71**, 911 (2018).
- [155] S. Marsin et al., *Study of the DnaB:DciA Interplay Reveals Insights into the Primary Mode of Loading of the Bacterial Replicative Helicase*, *Nucleic Acids Res* **49**, 6569 (2021).
- [156] J. Abe et al., *Circadian Rhythms. Atomic-Scale Origins of Slowness in the Cyanobacterial Circadian Clock*, *Science* **349**, 312 (2015).
- [157] Y. Li, Y. He, and Y. Luo, *Conservation of a Conformational Switch in RadA Recombinase from Methanococcus Maripaludis*, *Acta Crystallogr D Biol Crystallogr* **65**, 602 (2009).

- [158] D. V. Dibrova, M. Y. Galperin, E. V. Koonin, and A. Y. Mulkidjanian, *Ancient Systems of Sodium/Potassium Homeostasis as Predecessors of Membrane Bioenergetics*, *Biochemistry (Mosc)* **80**, 495 (2015).
- [159] A. A. Malär et al., *Spectroscopic Glimpses of the Transition State of ATP Hydrolysis Trapped in a Bacterial DnaB Helicase*, *Nat Commun* **12**, 5293 (2021).
- [160] X. Yang et al., *Mechanism of ATP hydrolysis by the Zika virus helicase*, *Faseb J* **32**, 5250 (2018).
- [161] X. Qian, Y. He, Y. Wu, and Y. Luo, *Asp302 Determines Potassium Dependence of a RadA Recombinase from Methanococcus Voltae*, *J Mol Biol* **360**, 537 (2006).
- [162] O. Itsathitphaisarn, R. A. Wing, W. K. Eliason, J. Wang, and T. A. Steitz, *The Hexameric Helicase DnaB Adopts a Nonplanar Conformation during Translocation*, *Cell* **151**, 267 (2012).
- [163] Y. Yamaichi and H. Niki, *Active Segregation by the Bacillus Subtilis Partitioning System in Escherichia Coli*, *Proc. Natl. Acad. Sci. USA* **97**, 14656 (2000).
- [164] A. Matte, L. W. Tari, and L. T. Delbaere, *How Do Kinases Transfer Phosphoryl Groups?*, *Structure* **6**, 413 (1998).
- [165] A. Finkelstein and B. Ptitsyn, *Protein Physics* (Academic Press, 2016).
- [166] J. Moser, C. Lange, J. Krausze, J. Rebelein, W. D. Schubert, M. W. Ribbe, D. W. Heinz, and D. Jahn, *Structure of ADP-aluminium fluoride-stabilized protochlorophyllide oxidoreductase complex*, *Proc Natl Acad Sci U S A* **110**, 2094 (2013).
- [167] D. delToro et al., *Functional Dissection of a Viral DNA Packaging Machine's Walker B Motif*, *J Mol Biol* **431**, 4455 (2019).
- [168] B. Garcia-Moreno, J. J. Dwyer, A. G. Gittis, E. E. Lattman, D. S. Spencer, and W. E. Stites, *Experimental measurement of the effective dielectric in the hydrophobic core of a protein*, *Biophys Chem* **64**, 211 (1997).
- [169] P. Kukic, D. Farrell, L. P. McIntosh, E. B. Garcia-Moreno, K. S. Jensen, Z. Toleikis, K. Teilum, and J. E. Nielsen, *Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations*, *J Am Chem Soc* **135**, 16968 (2013).
- [170] S. D. Fried, S. Bagchi, and S. G. Boxer, *Extreme Electric Fields Power Catalysis in the Active Site of Ketosteroid Isomerase*, *Science* **346**, 1510 (2014).
- [171] D. A. Cherepanov, S. I. Bibikov, M. V. Bibikova, D. A. Bloch, L. A. Drachev, O. A. Gupta, D. Oesterhelt, A. Y. Semenov, and A. Y. Mulkidjanian, *Reduction and Protonation of the Secondary Quinone Acceptor of Rhodobacter Sphaeroides Photosynthetic Reaction Center: Kinetic Model Based on a Comparison of Wild-Type Chromatophores with Mutants Carrying Arg→Ile Substitution at Sites 207 and 217 in the L-Subunit*, *Biochim Biophys Acta* **1459**, 10 (2000).
- [172] S. S. Klishin, W. Junge, and A. Y. Mulkidjanian, *Flash-Induced Turnover of the Cytochrome Bc1 Complex in Chromatophores of Rhodobacter Capsulatus: Binding of Zn<sup>2+</sup> Decelerates Likewise the Oxidation of Cytochrome b, the Reduction of Cytochrome C1 and the Voltage Generation*, *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1553**, 177 (2002).
- [173] A. Y. Mulkidjanian, *Proton Translocation by the Cytochrome Bc1 Complexes of Phototrophic Bacteria: Introducing the Activated Q-Cycle*, *Photochem Photobiol Sci* **6**, 19 (2007).

- [174] D. A. Cherepanov and A. Y. Mulkidjanian, *Proton Transfer in Azotobacter Vinelandii Ferredoxin I: Entatic Lys84 Operates as Elastic Counterbalance for the Proton-Carrying Asp15*, *Biochim. Biophys. Acta* **1505**, 179 (2001).
- [175] A. Y. Mulkidjanian and D. A. Cherepanov, *Probing Biological Interfaces by Tracing Proton Passage across Them*, *Photochem. Photobiol. Sci.* **5**, 577 (2006).
- [176] A. Y. Mulkidjanian, M. Y. Galperin, K. S. Makarova, Y. I. Wolf, and E. V. Koonin, *Evolutionary primacy of sodium bioenergetics*, *Biol. Direct* **3**, 13 (2008).
- [177] D. V. Dibrova, D. A. Cherepanov, M. Y. Galperin, V. P. Skulachev, and A. Y. Mulkidjanian, *Evolution of Cytochrome Bc Complexes: From Membrane-Anchored Dehydrogenases of Ancient Bacteria to Triggers of Apoptosis in Vertebrates*, *Biochim Biophys Acta* **1827**, 1407 (2013).
- [178] D. V. Dibrova, D. N. Shalaeva, M. Y. Galperin, and A. Y. Mulkidjanian, *Emergence of Cytochrome Bc Complexes in the Context of Photosynthesis*, *Physiol Plant* **161**, 150 (2017).
- [179] D. N. Shalaeva, D. A. Cherepanov, M. Y. Galperin, G. Vriend, and A. Y. Mulkidjanian, *G Protein-Coupled Receptors of Class A Harness the Energy of Membrane Potential to Increase Their Sensitivity and Selectivity*, *Biochim Biophys Acta Biomembr* **1861**, 183051 (2019).
- [180] D. N. Shalaeva, M. Y. Galperin, and A. Y. Mulkidjanian, *Eukaryotic G Protein-Coupled Receptors as Descendants of Prokaryotic Sodium-Translocating Rhodopsins*, *Biology Direct* **10**, 25 (2015).
- [181] C. Chothia and A. M. Lesk, *The Relation between the Divergence of Sequence and Structure in Proteins.*, *The EMBO Journal* **5**, 823 (1986).
- [182] K. Illergård, D. H. Ardell, and A. Elofsson, *Structure Is Three to Ten Times More Conserved than Sequence--a Study of Structural Response in Protein Cores*, *Proteins* **77**, 499 (2009).
- [183] J. Gu and P. E. Bourne, *Structural Bioinformatics* (John Wiley & Sons, 2009).
- [184] V. S. Minkov, V. V. Ghazaryan, E. V. Boldyreva, and A. M. Petrosyan, *Unusual Hydrogen Bonding in L-Cysteine Hydrogen Fluoride*, *Acta Crystallogr C Struct Chem* **71**, 733 (2015).
- [185] I. K. McDonald and J. M. Thornton, *Satisfying Hydrogen Bonding Potential in Proteins*, *J Mol Biol* **238**, 777 (1994).
- [186] B. van Beusekom, W. G. Touw, M. Tatineni, S. Somani, G. Rajagopal, J. Luo, G. L. Gilliland, A. Perrakis, and R. P. Joosten, *Homology-Based Hydrogen Bond Information Improves Crystallographic Structures in the PDB*, *Protein Sci* **27**, 798 (2018).
- [187] Schrödinger, LLC, *The PyMOL Molecular Graphics System, Version 2.5.0*, (unpublished).
- [188] P. J. Cock et al., *Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics*, *Bioinformatics* **25**, 1422 (2009).
- [189] A. Shrake and J. A. Rupley, *Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin*, *Journal of Molecular Biology* **79**, 351 (1973).
- [190] E. B. Saff and A. B. Kuijlaars, *Distributing Many Points on a Sphere*, *The Mathematical Intelligencer* **19**, 5 (1997).

- [191] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *The Protein Data Bank*, *Nucleic Acids Research* **28**, 235 (2000).
- [192] M. Blum et al., *The InterPro Protein Families and Domains Database: 20 Years On*, *Nucleic Acids Research* **49**, D344 (2021).
- [193] M. Gu and C. M. Rice, *The Spring  $\alpha$ -Helix Coordinates Multiple Modes of HCV (Hepatitis C Virus) NS3 Helicase Action*, *Journal of Biological Chemistry* **291**, 14499 (2016).
- [194] E. Uchikawa, M. Lethier, H. Malet, J. Brunel, D. Gerlier, and S. Cusack, *Structural Analysis of DsRNA Binding to Anti-Viral Pattern Recognition Receptors LGP2 and MDA5*, *Mol Cell* **62**, 586 (2016).
- [195] D. Lacabanne, T. Wiegand, N. Wili, M. I. Kozlova, R. Cadalbert, D. Klose, A. Y. Mulikidjanian, B. H. Meier, and A. Bockmann, *ATP Analogues for Structural Investigations: Case Studies of a DnaB Helicase and an ABC Transporter*, *Molecules* **25**, (2020).
- [196] E. Martz, *Help, Index & Glossary for Protein Explorer* ([Http://Www.Umass.Edu/Microbio/Chime/Pe\\_beta/Pe/Protexpl/Igloss.Htm?](http://www.umass.edu/microbio/chime/pe_beta/pe/protexpl/igloss.htm?Q=microbio/Chime/Explorer/Igloss.Htm) Q=microbio/Chime/Explorer/Igloss.Htm, 2001).
- [197] G. A. Jeffrey, *An Introduction to Hydrogen Bonding* (Oxford University Press, 1997).
- [198] N. L. Jean, T. J. Rutherford, and J. Lowe, *FtsK in Motion Reveals Its Mechanism for Double-Stranded DNA Translocation*, *Proc Natl Acad Sci U S A* **117**, 14202 (2020).
- [199] D. Gai, R. Zhao, D. Li, C. V. Finkelstein, and X. S. Chen, *Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen*, *Cell* **119**, 47 (2004).
- [200] K. P. Hopfner, A. Karcher, D. S. Shin, L. Craig, L. M. Arthur, J. P. Carney, and J. A. Tainer, *Structural Biology of Rad50 ATPase: ATP-Driven Conformational Control in DNA Double-Strand Break Repair and the ABC-ATPase Superfamily*, *Cell* **101**, 789 (2000).
- [201] J. A. Eisen, *A Phylogenomic Study of the MutS Family of Proteins*, *Nucleic Acids Res* **26**, 4291 (1998).
- [202] A. Decottignies and A. Goffeau, *Complete Inventory of the Yeast ABC Proteins*, *Nat Genet* **15**, 137 (1997).
- [203] J. D. Lawson, E. Pate, I. Rayment, and R. G. Yount, *Molecular Dynamics Analysis of Structural Factors Influencing Back Door Pi Release in Myosin*, *Biophysical Journal* **86**, 3794 (2004).
- [204] A. V. Afonin, I. V. Sterkhova, A. V. Vashchenko, and M. V. Sigalov, *Estimating the Energy of Intramolecular Bifurcated (Three-Centered) Hydrogen Bond by X-Ray, IR and  $^1\text{H}$  NMR Spectroscopy, and QTAIM Calculations*, *J Mol Struct* **1163**, 185 (2018).
- [205] D. Herschlag and M. M. Pinney, *Hydrogen Bonds: Simple after All?*, *Biochemistry* **57**, 3338 (2018).
- [206] M. Fothergill, M. F. Goodman, J. Petrushka, and A. Warshel, *Structure-Energy Analysis of the Role of Metal Ions in Phosphodiester Bond Hydrolysis by DNA Polymerase I*, *J Amer Chem Society* **117**, 11619 (1995).
- [207] R. J. P. Williams and J. J. R. Frausto da Silva, *The Biological Chemistry of the Elements* (Clarendon Press, Oxford, 1991).

- [208] D. Berta, P. J. Buigues, M. Badaoui, and E. Rosta, *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 (2020).
- [209] T. P. Silverstein, *How Enzymes Harness Highly Unfavorable Proton Transfer Reactions*, *Protein Sci* **30**, 735 (2021).
- [210] D. E. Scott, M. T. Ehebauer, T. Pukala, M. Marsh, T. L. Blundell, A. R. Venkitaraman, C. Abell, and M. Hyvonen, *Using a Fragment-Based Approach to Target Protein-Protein Interactions*, *Chembiochem* **14**, 332 (2013).
- [211] S. Matsumoto et al., *Molecular Mechanism for Conformational Dynamics of Ras·GTP Elucidated from In-Situ Structural Transition in Crystal*, *Sci Rep* **6**, 1 (2016).
- [212] Y. Li, Y. Zhang, F. Grosseruschkamp, S. Stephan, Q. Cui, C. Kottling, F. Xia, and K. Gerwert, *Specific Substates of Ras To Interact with GAPs and Effectors: Revealed by Theoretical Simulations and FTIR Experiments*, *J Phys Chem Lett* **9**, 1312 (2018).
- [213] J. S. Johansen, D. Kavaliauskas, S. H. Pfeil, M. Blaise, B. S. Cooperman, Y. E. Goldman, S. S. Thirup, and C. R. Knudsen, *E. Coli Elongation Factor Tu Bound to a GTP Analogue Displays an Open Conformation Equivalent to the GDP-Bound Form*, *Nucleic Acids Res* **46**, 8641 (2018).
- [214] C. B. Bauer, H. M. Holden, J. B. Thoden, R. Smith, and I. Rayment, *X-Ray Structures of the Apo and MgATP-Bound States of Dictyostelium Discoideum Myosin Motor Domain*, *J Biol Chem* **275**, 38494 (2000).
- [215] C. A. Smith and I. Rayment, *X-Ray Structure of the Magnesium(II).ADP.Vanadate Complex of the Dictyostelium Discoideum Myosin Motor Domain to 1.9 Å Resolution*, *Biochemistry* **35**, 5404 (1996).
- [216] S. Zhou and L. Wang, *Unraveling the Structural and Chemical Features of Biological Short Hydrogen Bonds*, *Chem. Sci.* **10**, 7734 (2019).
- [217] M. R. Singleton, M. R. Sawaya, T. Ellenberger, and D. B. Wigley, *Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides*, *Cell* **101**, 589 (2000).
- [218] S. Bailey, W. K. Eliason, and T. A. Steitz, *Structure of Hexameric DnaB Helicase and Its Complex with a Domain of DnaG Primase*, *Science* **318**, 459 (2007).
- [219] E. J. Enemark and L. Joshua-Tor, *On Helicases and Other Motor Proteins*, *Current Opinion in Structural Biology* **18**, 243 (2008).
- [220] G. Wang, M. G. Klein, E. Tokonzaba, Y. Zhang, L. G. Holden, and X. S. Chen, *The Structure of a DnaB-Family Replicative Helicase and Its Interactions with Primase*, *Nat Struct Mol Biol* **15**, 1 (2008).
- [221] Y. Gao, Y. Cui, T. Fox, S. Lin, H. Wang, N. de Val, Z. H. Zhou, and W. Yang, *Structures and Operating Principles of the Replisome*, *Science* **363**, eaav7003 (2019).
- [222] T. Elston, H. Wang, and G. Oster, *Energy Transduction in ATP Synthase*, *Nature* **391**, 510 (1998).
- [223] P. D. Boyer, *The ATP Synthase—a Splendid Molecular Machine*, *Annu Rev Biochem* **66**, 717 (1997).
- [224] C. Zhao, E. C. Smith, and S. W. Whiteheart, *Requirements for the Catalytic Cycle of the N-Ethylmaleimide-Sensitive Factor (NSF)*, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 159 (2012).



- [225] A. S. Woods and S. Ferre, *Amazing Stability of the Arginine-Phosphate Electrostatic Interaction*, *J Proteome Res* **4**, 1397 (2005).
- [226] L. Shimoni and J. P. Glusker, *Hydrogen Bonding Motifs of Protein Side Chains: Descriptions of Binding of Arginine and Amide Groups*, *Protein Sci* **4**, 65 (1995).
- [227] C. A. Seipp, N. J. Williams, M. K. Kidder, and R. Custelcean, *CO<sub>2</sub> Capture from Ambient Air by Crystallization with a Guanidine Sorbent*, *Angew Chem Int Ed Engl* **56**, 1042 (2017).
- [228] K. A. Schug and W. Lindner, *Noncovalent Binding between Guanidinium and Anionic Groups: Focus on Biological- and Synthetic-Based Arginine/Guanidinium Interactions with Phosph[on]ate and Sulf[on]ate Residues*, *Chem Rev* **105**, 67 (2005).
- [229] B. J. Calnan, B. Tidor, S. Biancalana, D. Hudson, and A. D. Frankel, *Arginine-Mediated RNA Recognition: The Arginine Fork*, *Science* **252**, 1167 (1991).
- [230] M. Chabre, *Aluminumfluoride and Beryllfluoride Complexes: New Phosphate Analogs in Enzymology*, *Trends in Biochemical Sciences* **15**, 6 (1990).
- [231] B. Antonny, J. Bigay, and M. Chabre, *A Novel Magnesium-Dependent Mechanism for the Activation of Transducin by Fluoride*, *FEBS Lett* **268**, 277 (1990).
- [232] D. L. Graham, J. F. Eccleston, C. W. Chung, and P. N. Lowe, *Magnesium Fluoride-Dependent Binding of Small G Proteins to Their GTPase-Activating Proteins*, *Biochemistry* **38**, 14981 (1999).
- [233] D. N. Shalaeva, D. A. Cherepanov, M. Y. Galperin, and A. Y. Mulkidjanian, *Comparative Analysis of Active Sites in P-Loop Nucleoside Triphosphatases Suggests an Ancestral Activation Mechanism*, *BioRxiv* **439992 [preprint]**, (2018).
- [234] W. P. Jencks, *Catalysis in Chemistry and Enzymology* (Courier Corporation, 1987).
- [235] L. T. J. Delbaere, A. M. Sudom, L. Prasad, Y. Leduc, and H. Goldie, *Structure/Function Studies of Phosphoryl Transfer by Phosphoenolpyruvate Carboxykinase*, *Biochimica et Biophysica Acta - Proteins and Proteomics* **1697**, 271 (2004).
- [236] C. Allin and K. Gerwert, *Ras Catalyzes GTP Hydrolysis by Shifting Negative Charges from Gamma- to Beta-Phosphate as Revealed by Time-Resolved FTIR Difference Spectroscopy*, *Biochemistry* **40**, 3037 (2001).
- [237] J. J. Frye, V. A. Klenchin, C. R. Bagshaw, and I. Rayment, *Insights into the Importance of Hydrogen Bonding in the Gamma-Phosphate Binding Pocket of Myosin: Structural and Functional Studies of Serine 236*, *Biochemistry* **49**, 4897 (2010).
- [238] C. Kotting and K. Gerwert, *Time-Resolved FTIR Studies Provide Activation Free Energy, Activation Enthalpy and Activation Entropy for GTPase Reactions*, *Chem Phys* **307**, 227 (2004).
- [239] A. Shutes and C. J. Der, *Real-Time in Vitro Measurement of Intrinsic and Ras GAP-Mediated GTP Hydrolysis*, *Methods Enzymol* **407**, 9 (2006).
- [240] D. A. Hattendorf and S. L. Lindquist, *Analysis of the AAA Sensor-2 Motif in the C-Terminal ATPase Domain of Hsp104 with a Site-Specific Fluorescent Probe of Nucleotide Binding*, *Proc, Natl. Acad. Sci. USA* **99**, 2732 (2002).
- [241] J. E. Walker, *The Regulation of Catalysis in ATP Synthase*, *Curr Opin Struct Biol* **4**, 912 (1994).
- [242] A. E. Senior, S. Nadanaciva, and J. Weber, *The molecular mechanism of ATP synthesis by F1F0-ATP synthase*, *Biochim. Biophys. Acta* **1553**, 188 (2002).

- [243] P. Gideon, J. John, M. Frech, A. Lautwein, R. Clark, J. E. Scheffler, and A. Wittinghofer, *Mutational and Kinetic Analyses of the GTPase-Activating Protein (GAP)-P21 Interaction: The C-Terminal Domain of GAP Is Not Sufficient for Full Activity*, *Mol, Cell Biol* **12**, 2050 (1992).
- [244] J. Codina and L. Birnbaumer, *Requirement for intramolecular domain interaction in activation of G protein alpha subunit by aluminum fluoride and GDP but not by GTP gamma S*, *J. Biol. Chem* **269**, 29339 (1994).
- [245] J. Cherfils and M. Zeghouf, *Regulation of Small GTPases by GEFs, GAPs, and GDIs*, *Physiological Reviews* **93**, 269 (2013).
- [246] M. Klahn, E. Rosta, and A. Warshel, *On the Mechanism of Hydrolysis of Phosphate Monoesters Dianions in Solutions and Proteins*, *J Am Chem Soc* **128**, 15310 (2006).
- [247] N. H. Williams, *Magnesium ion catalyzed ATP hydrolysis*, *JACS* **122**, 12023 (2000).
- [248] T. Rudack, F. Xia, J. Schlitter, C. Kotting, and K. Gerwert, *The Role of Magnesium for Geometry and Charge in GTP Hydrolysis, Revealed by Quantum Mechanics/Molecular Mechanics Simulations*, *Biophys J* **103**, 293 (2012).
- [249] J. C. Liao, S. Sun, D. Chandler, and G. Oster, *The Conformational States of Mg.ATP in Water*, *Eur Biophys J* **33**, 29 (2004).
- [250] G. Li and X. C. Zhang, *GTP Hydrolysis Mechanism of Ras-like GTPases*, *Journal of Molecular Biology* **340**, 921 (2004).
- [251] H. Cheng, S. Sukal, H. Deng, T. S. Leyh, and R. Callender, *Vibrational Structure of GDP and GTP Bound to RAS: An Isotope-Edited FTIR Study*, *Biochemistry* **40**, 4035 (2001).
- [252] O. Poyraz, K. Brunner, B. Lohkamp, H. Axelsson, L. G. Hammarstrom, R. Schnell, and G. Schneider, *Crystal Structures of the Kinase Domain of the Sulfate-Activating Complex in Mycobacterium Tuberculosis*, *PLoS One* **10**, 0121494 (2015).
- [253] Z. Chen, H. Yang, and N. P. Pavletich, *Mechanism of Homologous Recombination from the RecA-SsDNA/DsDNA Structures*, *Nature* **453**, 484 (2008).
- [254] F. Voigts-Hoffmann, N. Schmitz, K. Shen, S. O. Shan, S. F. Ataide, and N. Ban, *The Structural Basis of FtsY Recruitment and GTPase Activation by SRP RNA*, *Mol Cell* **52**, 643 (2013).
- [255] Y. Dong, S. Zhang, Z. Wu, X. Li, W. L. Wang, Y. Zhu, S. Stoilova-McPhie, Y. Lu, D. Finley, and Y. Mao, *Cryo-EM Structures and Dynamics of Substrate-Engaged Human 26S Proteasome*, *Nature* **565**, 49 (2019).
- [256] K. Y. Lu, W. F. Chen, S. Rety, N. N. Liu, W. Q. Wu, Y. X. Dai, D. Li, H. Y. Ma, S. X. Dou, and X. G. Xi, *Insights into the Structural and Mechanistic Basis of Multifunctional S. Cerevisiae Pif1p Helicase*, *Nucleic Acids Res* **46**, 1486 (2018).
- [257] T. Schweins and A. Warshel, *Mechanistic Analysis of the Observed Linear Free Energy Relationships in P21ras and Related Systems*, *Biochemistry* **35**, 14232 (1996).
- [258] C. Allin, M. R. Ahmadian, A. Wittinghofer, and K. Gerwert, *Monitoring the GAP Catalyzed H-Ras GTPase Reaction at Atomic Resolution in Real Time*, *Proc Natl Acad Sci U S A* **98**, 7754 (2001).
- [259] M. Valiev, R. Kawai, J. A. Adams, and J. H. Weare, *The Role of the Putative Catalytic Base in the Phosphoryl Transfer Reaction in a Protein Kinase: First-Principles Calculations*, *J. Am. Chem. Soc.* **125**, 9926 (2003).

- [260] M. Valiev, J. Yang, J. A. Adams, S. S. Taylor, and J. H. Weare, *Phosphorylation Reaction in CAPK Protein Kinase-Free Energy Quantum Mechanical/Molecular Mechanics Simulations*, *J. Phys. Chem. B* **111**, 13455 (2007).
- [261] H. Cheng, S. Sukal, R. Callender, and T. S. Leyh, *Gamma-Phosphate Protonation and PH-Dependent Unfolding of the Ras.GTP.Mg<sup>2+</sup> Complex: A Vibrational Spectroscopy Study*, *J Biol Chem* **276**, 9931 (2001).
- [262] D. Mann, J. Guldenhaupt, J. Schartner, K. Gerwert, and C. Kottling, *The Protonation States of GTP and GppNHp in Ras Proteins*, *J Biol Chem* **293**, 3871 (2018).
- [263] M. Eigen and W. Kruse, *Über den Einfluß von Wasserstoffbrücken-Struktur und elektrostatischer Wechselwirkung auf die Geschwindigkeit protolytischer Reaktionen*, *Z, Naturforsch. B* **18b**, 857 (1963).
- [264] M. Eigen, *Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES*, *Angew. Chem. Int. Ed. Engl.* **3**, 1 (1964).
- [265] D. N. Frick, R. S. Rypma, A. M. Lam, and C. M. Frenz, *Electrostatic Analysis of the Hepatitis C Virus NS3 Helicase Reveals Both Active and Allosteric Site Locations*, *Nucleic Acids Res* **32**, 5519 (2004).
- [266] D. N. Frick, *The Hepatitis C Virus NS3 Protein: A Model RNA Helicase and Potential Drug Target*, *Curr Issues Mol Biol* **9**, 1 (2007).
- [267] E. G. Alexov and M. R. Gunner, *Incorporating Protein Conformational Flexibility into the Calculation of PH-Dependent Protein Properties*, *Biophys J* **72**, 2075 (1997).
- [268] M. Gu and C. M. Rice, *Three Conformational Snapshots of the Hepatitis C Virus NS3 Helicase Reveal a Ratchet Translocation Mechanism*, *Proc Natl Acad Sci U S A* **107**, 521 (2010).
- [269] D. D. Leipe, L. Aravind, N. V. Grishin, and E. V. Koonin, *The Bacterial Replicative Helicase DnaB Evolved from a RecA Duplication*, *Genome Res* **10**, 5 (2000).
- [270] G. Zundel, *Proton polarizability of hydrogen bonds: infrared methods, relevance to electrochemical and biological systems*, *Methods Enzymol* **127**, 439 (1986).
- [271] B. Brzezinski and G. Zuñdel, *The Role of Water and Proton-Transfer Processes in Hydrogen-Bonded Chains with Large Proton Polarizability*, *Faraday Discussions* **103**, 363 (1996).
- [272] A. D. Kaulen, *Electrogenic Processes and Protein Conformational Changes Accompanying the Bacteriorhodopsin Photocycle*, *Biochim Biophys Acta* **1460**, 204 (2000).
- [273] J. Heberle, *Proton Transfer Reactions across Bacteriorhodopsin and along the Membrane*, *Biochim Biophys Acta* **1458**, 135 (2000).
- [274] T. E. Decoursey, *Voltage-Gated Proton Channels and Other Proton Transfer Pathways*, *Physiological Reviews* **83**, 475 (2003).
- [275] C. A. Wraight, *Chance and Design—Proton Transfer in Water, Channels and Bioenergetic Proteins*, *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1757**, 886 (2006).
- [276] A. Y. Mulkidjanian, *Proton in the Well and through the Desolvation Barrier*, *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1757**, 415 (2006).
- [277] A. Y. Mulkidjanian, J. Heberle, and D. A. Cherepanov, *Protons @ Interfaces: Implications for Biological Energy Conversion*, *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1757**, 913 (2006).

- [278] V. A. Lorenz-Fonfria and J. Heberle, *Proton Transfer and Protein Conformation Dynamics in Photosensitive Proteins by Time-Resolved Step-Scan Fourier-Transform Infrared Spectroscopy*, *J Vis Exp* **88**, 51622 (2014).
- [279] D. Kaur, U. Khaniya, Y. Zhang, and M. R. Gunner, *Protein Motifs for Proton Transfers That Build the Transmembrane Proton Gradient*, *Front Chem* **9**, 660954 (2021).
- [280] D. Pines, E. T. J. Nibbering, and E. Pines, *Monitoring the Microscopic Molecular Mechanisms of Proton Transfer in Acid-base Reactions in Aqueous Solutions*, *Israel Journal of Chemistry* **55**, 1240 (2015).
- [281] V. Borshchevskiy, *True-Atomic-Resolution Insights into the Structure and Functional Role of Linear Chains and Low-Barrier Hydrogen Bonds in Proteins*, *Molecular Biology* **27** (n.d.).
- [282] L. S. Brown, *Light-Driven Proton Transfers and Proton Transport by Microbial Rhodopsins – A Biophysical Perspective*, *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1864**, 183867 (2022).
- [283] C. J. T. de Grotthuss, *Mémoire Sur La Décomposition de l'eau et Des Corps Qu'elle Tient En Dissolution à l'aide de l'électricité Galvanique* (Rome, 1805).
- [284] D. Marx, *Proton Transfer 200 Years after von Grotthuss: Insights from Ab Initio Simulations*, *Chemphyschem* **7**, 1848 (2006).
- [285] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos, *Nuclear Quantum Effects and Hydrogen Bond Fluctuations in Water*, *Proc Natl Acad Sci USA* **110**, 15591 (2013).
- [286] D. Kaur, Y. Zhang, K. M. Reiss, M. Mandal, G. W. Brudvig, V. S. Batista, and M. R. Gunner, *Proton Exit Pathways Surrounding the Oxygen Evolving Complex of Photosystem II*, *Biochim Biophys Acta Bioenerg* **1862**, 148446 (2021).
- [287] C. Nadolny and G. Zundel, *Fourier Transform Infrared Spectroscopic Studies of Proton Transfer Processes and the Dissociation of Zn<sup>2+</sup>-Bound Water in Alcohol Dehydrogenases*, *Eur J Biochem* **247**, 914 (1997).
- [288] L. A. Drachev, A. D. Kaulen, L. V. Khitrina, and V. P. Skulachev, *Fast stages of photoelectric processes in biological membranes, I. Bacteriorhodopsin*, *Eur J Biochem* **117**, 461 (1981).
- [289] P. Brzezinski, M. L. Paddock, M. Y. Okamura, and G. Feher, *Light-Induced Electrogenic Events Associated with Proton Uptake upon Forming QB<sup>-</sup> in Bacterial Wild-Type and Mutant Reaction Centers*, *Biochim Biophys Acta* **1321**, 149 (1997).
- [290] L. Drachev, B. Kaurov, M. Mamedov, A. Y. Mulkidjanian, A. Y. Semenov, V. Shinkarev, V. Skulachev, and M. Verkhovsky, *Flash-Induced Electrogenic Events in the Photosynthetic Reaction Center and Bc1 Complexes of Rhodospirillum rubrum Chromatophores*, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **973**, 189 (1989).
- [291] M. A. Kozlova, H. D. Juhnke, D. A. Cherepanov, C. R. Lancaster, and A. Y. Mulkidjanian, *Proton Transfer in the Photosynthetic Reaction Center of Blastochloris Viridis*, *FEBS Lett* **582**, 238 (2008).
- [292] R. Hienerwadel, S. Grzybek, C. Fogel, W. Kreutz, M. Y. Okamura, M. L. Paddock, J. Breton, E. Navedryk, and W. Mantele, *Protonation of Glu L212 Following QB<sup>-</sup> Formation in the Photosynthetic Reaction Center of Rhodospirillum rubrum: Evidence from Time-Resolved Infrared Spectroscopy*, *Biochemistry* **34**, 2832 (1995).

- [293] E. Nabedryk, J. Breton, R. Hienerwadel, C. Fogel, W. Mantele, M. L. Paddock, and M. Y. Okamura, *Fourier Transforms Infrared Difference Spectroscopy of Secondary Quinone Acceptor Photoreduction in Proton Transfer Mutants of Rhodobacter Sphaeroides*, *Biochemistry* **34**, 14722 (1995).
- [294] C. Zscherp, R. Schlesinger, J. Tittor, D. Oesterhelt, and J. Heberle, *In Situ Determination of Transient PKa Changes of Internal Amino Acids of Bacteriorhodopsin by Using Time-Resolved Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy*, *Proceedings of the National Academy of Sciences* **96**, 5498 (1999).
- [295] L. Drachev, A. Kaulen, and V. Skulachev, *Correlation of Photochemical Cycle, H<sup>+</sup> Release and Uptake, and Electric Events in Bacteriorhodopsin*, *FEBS Lett* **178**, 331 (1984).
- [296] M. Engelhard, K. Gerwert, B. Hess, and F. Siebert, *Light-Driven Protonation Changes of Internal Aspartic Acids of Bacteriorhodopsin: An Investigation of Static and Time-Resolved Infrared Difference Spectroscopy Using [4-<sup>13</sup>C] Aspartic Acid Labeled Purple Membrane*, *Biochemistry* **24**, 400 (1985).
- [297] O. A. Gupta, D. A. Cherepanov, W. Junge, and A. Y. Mulikidjanian, *Proton Transfer from the Bulk to the Bound Ubiquinone Q<sub>B</sub> of the Reaction Center in Chromatophores of Rhodobacter Sphaeroides: Retarded Conveyance by Neutral Water*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13159 (1999).
- [298] J. P. Klare, E. Bordignon, M. Engelhard, and H. J. Steinhoff, *Sensory Rhodopsin II and Bacteriorhodopsin: Light Activated Helix F Movement*, *Photochem Photobiol Sci* **3**, 543 (2004).
- [299] H. J. Steinhoff, R. Mollaaghababa, C. Altenbach, K. Hideg, M. Krebs, H. G. Khorana, and W. L. Hubbell, *Time-Resolved Detection of Structural Changes during the Photocycle of Spin-Labeled Bacteriorhodopsin*, *Science* **266**, 105 (1994).
- [300] T. Rink, M. Pfeiffer, D. Oesterhelt, K. Gerwert, and H. J. Steinhoff, *Unraveling Photoexcited Conformational Changes of Bacteriorhodopsin by Time Resolved Electron Paramagnetic Resonance Spectroscopy*, *Biophys J* **78**, 1519 (2000).
- [301] A.-A. Wegener, I. Chizhov, M. Engelhard, and H.-J. Steinhoff, *Time-Resolved Detection of Transient Movement of Helix F in Spin-Labelled Pharaonis Sensory Rhodopsin III11* Edited by W. Baumeister, *Journal of Molecular Biology* **301**, 881 (2000).
- [302] M. H. Stowell, T. M. McPhillips, D. C. Rees, S. M. Soltis, E. Abresch, and G. Feher, *Light-Induced Structural Changes in Photosynthetic Reaction Center: Implications for Mechanism of Electron-Proton Transfer*, *Science* **276**, 812 (1997).
- [303] H. Luecke, B. Schobert, J. P. Cartailler, H. T. Richter, A. Rosengarth, R. Needleman, and J. K. Lanyi, *Coupling Photoisomerization of Retinal to Directional Transport in Bacteriorhodopsin*, *J Mol Biol* **300**, 1237 (2000).
- [304] A. Warshel and S. T. Russell, *Calculations of Electrostatic Interactions in Biological Systems and in Solutions*, *Q Rev Biophys* **17**, 283 (1984).
- [305] M. E. Bozdaganyan, A. V. Lokhmatikov, N. Voskoboynikova, D. A. Cherepanov, H. J. Steinhoff, K. V. Shaitan, and A. Y. Mulikidjanian, *Proton Leakage across Lipid Bilayers: Oxygen Atoms of Phospholipid Ester Linkers Align Water Molecules into Transmembrane Water Wires*, *Biochim Biophys Acta Bioenerg* **1860**, 439 (2019).
- [306] A. Y. Mulikidjanian, *Conformationally Controlled PK-Switching in Membrane Proteins: One More Mechanism Specific to the Enzyme Catalysis?*, *FEBS Letters* **6** (1999).

- [307] T. K. Harris and G. J. Turner, *Structural Basis of Perturbed PKa Values of Catalytic Groups in Enzyme Active Sites*, IUBMB Life **53**, 85 (2002).
- [308] A. E. Senior and M. K. al-Shawi, *Further Examination of Seventeen Mutations in Escherichia Coli F1-ATPase Beta-Subunit*, J Biol Chem **267**, 21471 (1992).
- [309] J. P. Wolk, M. Klose, J. G. Wit, T. Blaauwen, R. Freudl, and A. J. Driessen, *Identification of the Magnesium-Binding Domain of the High-Affinity ATP-Binding Site of the Bacillus Subtilis and Escherichia Coli SecA Protein*, J Biol Chem **270**, 18975 (1995).
- [310] S. Lobau, J. Weber, S. Wilke-Mounts, and A. E. Senior, *F1-ATPase, Roles of Three Catalytic Site Residues*, J Biol Chem **272**, 3648 (1997).
- [311] A. Frelet and M. Klein, *Insight in Eukaryotic ABC Transporter Function by Mutation Analysis*, FEBS Lett **580**, 1064 (2006).
- [312] J. John, H. Rensland, I. Schlichting, I. Vetter, G. D. Borasio, R. S. Goody, and A. Wittinghofer, *Kinetic and structural analysis of the Mg(2+)-binding site of the guanine nucleotide-binding protein p21H-ras*, J Biol Chem **268**, 923 (1993).
- [313] D. N. Frick, S. Banik, and R. S. Rypma, *Role of Divalent Metal Cations in ATP Hydrolysis Catalyzed by the Hepatitis C Virus NS3 Helicase: Magnesium Provides a Bridge for ATP to Fuel Unwinding*, Journal of Molecular Biology **365**, 1017 (2007).
- [314] T. Yamashita, H. Unno, Y. Mori, H. Tani, K. Moriishi, A. Takamizawa, M. Agoh, T. Tsukihara, and Y. Matsuura, *Crystal Structure of the Catalytic Domain of Japanese Encephalitis Virus NS3 Helicase/Nucleoside Triphosphatase at a Resolution of 1.8 Å*, Virology **373**, 426 (2008).
- [315] R. Yang, R. Scavetta, and X. Chang, *Interaction between the Bound Mg·ATP and the Walker A Serine Residue in NBD2 of Multidrug Resistance-Associated Protein MRP1 Plays a Crucial Role for the ATP-Dependent Leukotriene C4 Transport*, Biochemistry **47**, 8456 (2008).
- [316] P. A. Sigala, E. A. Ruben, C. W. Liu, P. M. Piccoli, E. G. Hohenstein, T. J. Martinez, A. J. Schultz, and D. Herschlag, *Determination of Hydrogen Bond Structure in Water versus Aprotic Environments To Test the Relationship Between Length and Stability*, J Am Chem Soc **137**, 5730 (2015).
- [317] S. Pylaeva, C. Allolio, B. Koeppe, G. S. Denisov, H. H. Limbach, D. Sebastiani, and P. M. Tolstoy, *Proton Transfer in a Short Hydrogen Bond Caused by Solvation Shell Fluctuations: An Ab Initio MD and NMR/UV Study of an (OHO)(-) Bonded System*, Phys Chem Chem Phys **17**, 4634 (2015).
- [318] A. Fersht, *Enzyme structure and mechanism*, (1985).
- [319] V. N. Uversky, *The Intrinsic Disorder Alphabet*, III. Dual Personality of Serine, Intrinsically Disord Proteins **3**, 1027032 (2015).
- [320] E. J. Wojcik, N. A. Dalrymple, S. R. Alford, R. A. Walker, and S. Kim, *Disparity in Allosteric Interactions of Monastrol with Eg5 in the Presence of ADP and ATP: A Difference FT-IR Investigation*, Biochemistry **43**, 9939 (2004).
- [321] B. Jun and S. Kim, *Real-Time Structural Transitions Are Coupled to Chemical Steps in ATP Hydrolysis by Eg5 Kinesin*, J Biol Chem **285**, 11073 (2010).
- [322] Y. X. Dai, W. F. Chen, N. N. Liu, F. Y. Teng, H. L. Guo, X. M. Hou, S. X. Dou, S. Rety, and X. G. Xi, *Structural and Functional Studies of SF1B Pif1 from Thermus Oshimai Reveal Dimerization-Induced Helicase Inhibition*, Nucleic Acids Res **49**, 4129 (2021).

- [323] S. Pasqualato and J. Cherfils, *Crystallographic Evidence for Substrate-Assisted GTP Hydrolysis by a Small GTP Binding Protein*, *Structure* **13**, 533 (2005).
- [324] W. W. Cleland, *Low-Barrier Hydrogen Bonds and Low Fractionation Factor Bases in Enzymic Reactions*, *Biochemistry* **31**, 317 (1992).
- [325] P. A. Frey, S. A. Whitt, and J. B. Tobin, *A Low-Barrier Hydrogen Bond in the Catalytic Triad of Serine Proteases*, *Science* **264**, 1927 (1994).
- [326] C. L. Perrin and J. B. Nielson, *Strong" hydrogen bonds in chemistry and biology*, *Annu Rev Phys Chem* **48**, 511 (1997).
- [327] C. L. Perrin, *Are Short, Low-Barrier Hydrogen Bonds Unusually Strong?*, *Acc Chem Res* **43**, 1550 (2010).
- [328] W. Y. Wahlgren, G. Pál, J. Kardos, P. Porrogi, B. Szenthe, A. Patthy, L. Gráf, and G. Katona, *The Catalytic Aspartate Is Protonated in the Michaelis Complex Formed between Trypsin and an in Vitro Evolved Substrate-like Inhibitor: A REFINED MECHANISM OF SERINE PROTEASE ACTION \**, *Journal of Biological Chemistry* **286**, 3587 (2011).
- [329] P. Kumar, P. K. Agarwal, M. B. Waddell, T. Mittag, E. H. Serpersu, and M. J. Cuneo, *Low-Barrier and Canonical Hydrogen Bonds Modulate Activity and Specificity of a Catalytic Triad*, *Angewandte Chemie* **131**, 16406 (2019).
- [330] P. Kumar, P. K. Agarwal, and M. J. Cuneo, *On the Case of the Misplaced Hydrogens*, *ChemBioChem* **22**, 288 (2021).
- [331] G. N. Nagy, R. Suardiaz, A. Lopata, O. Ozohanics, K. Vékey, B. R. Brooks, I. Leveles, J. Tóth, B. G. Vértessy, and E. Rosta, *Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase DUTPases*, *J. Am. Chem. Soc.* **138**, 15035 (2016).
- [332] B. Grigorenko, I. Polyakov, and A. Nemukhin, *Mechanisms of ATP to CAMP Conversion Catalyzed by the Mammalian Adenylyl Cyclase: A Role of Magnesium Coordination Shells and Proton Wires*, *J Phys Chem B* **124**, 451 (2020).
- [333] W. W. Cleland, P. A. Frey, and J. A. Gerlt, *The Low Barrier Hydrogen Bond in Enzymatic Catalysis*, *J Biol Chem* **273**, 25529 (1998).
- [334] P. D. Ngo, S. O. Mansoorabadi, and P. A. Frey, *Serine Protease Catalysis: A Computational Study of Tetrahedral Intermediates and Inhibitory Adducts*, *J Phys Chem B* **120**, 7353 (2016).
- [335] B. A. Reikhardt and P. D. Shabanov, *Catalytic Subunit of PKA as a Prototype of the Eukaryotic Protein Kinase Family*, *Biochemistry Moscow* **85**, 409 (2020).
- [336] J. A. Endicott, M. E. M. Noble, and L. N. Johnson, *The Structural Basis for Control of Eukaryotic Protein Kinases*, *Annu. Rev. Biochem.* **81**, 587 (2012).
- [337] M. Lazaratos, M. Siemers, L. S. Brown, and A.-N. Bondar, *Conserved Hydrogen-Bond Motifs of Membrane Transporters and Receptors*, *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1864**, 183896 (2022).
- [338] S. Mehmood et al., *Structural and Functional Basis for Lipid Synergy on the Activity of the Antibacterial Peptide ABC Transporter McjD*, *J Biol Chem* **291**, 21656 (2016).
- [339] S. A. Ferguson, G. M. Cook, M. G. Montgomery, A. G. Leslie, and J. E. Walker, *Regulation of the Thermoalkaliphilic F1-ATPase from *Caldalkalibacillus Thermarum**, *Proc Natl Acad Sci U S A* **113**, 10860 (2016).

- [340] M. Podobnik, T. F. Weitze, M. O'Donnell, and J. Kuriyan, *Nucleotide-Induced Conformational Changes in an Isolated Escherichia Coli DNA Polymerase III Clamp Loader Subunit*, *Structure* **11**, 253 (2003).
- [341] A. Lavie, N. Ostermann, R. Brundiers, R. S. Goody, J. Reinstein, M. Konrad, and I. Schlichting, *Structural Basis for Efficient Phosphorylation of 3'-Azidothymidine Monophosphate by Escherichia Coli Thymidylate Kinase*, *Proc Natl Acad Sci U S A* **95**, 14045 (1998).
- [342] F. A. Tezcan, J. T. Kaiser, D. Mustafi, M. Y. Walton, J. B. Howard, and D. C. Rees, *Nitrogenase Complexes: Multiple Docking Sites for a Nucleotide Switch Protein*, *Science* **309**, 1377 (2005).
- [343] Y. Lin, S. Lu, J. Zhang, and Y. Zheng, *Structure of an Inactive Conformation of GTP-Bound RhoA GTPase*, *Structure* **29**, 553 (2021).
- [344] M. Del Campo and A. M. Lambowitz, *Structure of the Yeast DEAD Box Protein Mss116p Reveals Two Wedges That Crimp RNA*, *Mol Cell* **35**, 598 (2009).
- [345] P. J. Porebski et al., *Structural Characterization of Helicobacter Pylori Dethiobiotin Synthetase Reveals Differences between Family Members*, *Febs J* **279**, 1093 (2012).
- [346] K. J. Gibson, *Isolation and Chemistry of the Mixed Anhydride Intermediate in the Reaction Catalyzed by Dethiobiotin Synthetase*, *Biochemistry* **36**, 8474 (1997).
- [347] K. J. Gibson, G. H. Lorimer, A. R. Rendina, W. S. Taylor, G. Cohen, A. A. Gatenby, W. G. Payne, D. C. Roe, and B. A. Lockett, *Dethiobiotin Synthetase: The Carbonylation of 7,8-Diaminononanoic Acid Proceeds Regiospecifically via the N7-Carbamate*, *Biochemistry* **34**, 10976 (1995).
- [348] A. Dikfidan, B. Loll, C. Zeymer, I. Magler, T. Clausen, and A. Meinhart, *RNA Specificity and Regulation of Catalysis in the Eukaryotic Polynucleotide Kinase Clp1*, *Mol Cell* **54**, 975 (2014).
- [349] J. Baranwal, S. Lhospice, M. Kanade, S. Chakraborty, P. R. Gade, S. Harne, J. Herrou, T. Mignot, and P. Gayathri, *Allosteric Regulation of a Prokaryotic Small Ras-like GTPase Contributes to Cell Polarity Oscillations in Bacterial Motility*, *PLoS Biol* **17**, e3000459 (2019).
- [350] A. Matte and L. T. J. Delbaere, *ATP-Binding Motifs*, in *Encyclopedia of Life Sciences (ELS)* (John Wiley & Sons, Ltd, Chichester, 2010).
- [351] A. E. Rose, R. S. Brown, and C. Schlieker, *Torsins: Not Your Typical AAA+ ATPases*, *Crit Rev Biochem Mol Biol* **50**, 532 (2015).
- [352] C. Zhao, R. S. Brown, A. R. Chase, M. R. Eisele, and C. Schlieker, *Regulation of Torsin ATPases by LAP1 and LULL1*, *Proc Natl Acad Sci U S A* **110**, E1545 (2013).
- [353] F. E. Demircioglu, B. A. Sosa, J. Ingram, H. L. Ploegh, and T. U. Schwartz, *Structures of TorsinA and Its Disease-Mutant Complexed with an Activator Reveal the Molecular Basis for Primary Dystonia*, *Elife* **5**, (2016).
- [354] R. Xu, Y. Yang, and X. Zheng, *Unique Structural Features of the Adenylate Kinase HCINAP/AK6 and Its Multifaceted Functions in Carcinogenesis and Tumor Progression*, *FEBS Letters* **595**, 2071 (2021).
- [355] C. E. Drakou, A. Malekkou, J. M. Hayes, C. W. Lederer, D. D. Leonidas, N. G. Oikonomakos, A. I. Lamond, N. Santama, and S. E. Zographos, *HCINAP Is an*



- Atypical Mammalian Nuclear Adenylate Kinase with an ATPase Motif: Structural and Functional Studies*, *Proteins: Structure, Function, and Bioinformatics* **80**, 206 (2012).
- [356] J. Gu, L. Zhang, S. Zong, R. Guo, T. Liu, J. Yi, P. Wang, W. Zhuo, and M. Yang, *Cryo-EM Structure of the Mammalian ATP Synthase Tetramer Bound with Inhibitory Protein IF<sub>1</sub>*, *Structure* **27**, 9 (2019).
- [357] B. J. Murphy, N. Klusch, J. Langer, D. J. Mills, Ö. Yildiz, and W. Kühlbrandt, *Rotary Substates of Mitochondrial ATP Synthase Reveal the Basis of Flexible F<sub>1</sub>-F<sub>o</sub> Coupling*, *Science* **364**, eaaw9128 (2019).
- [358] N. Kapoor, S. T. Menon, R. Chauhan, P. Sachdev, and T. P. Sakmar, *Structural Evidence for a Sequential Release Mechanism for Activation of Heterotrimeric G Proteins*, *Journal of Molecular Biology* **393**, 882 (2009).
- [359] K. Rittinger, P. A. Walker, J. F. Eccleston, S. J. Smerdon, and S. J. Gamblin, *Structure at 1.65 Å of RhoA and Its GTPase-Activating Protein in Complex with a Transition-State Analogue*, *Nature* **389**, 758 (1997).
- [360] E. Amin, M. Jaiswal, U. Derewenda, K. Reis, K. Nouri, K. T. Koessmeier, P. Aspenstrom, A. V. Somlyo, R. Dvorsky, and M. R. Ahmadian, *Deciphering the Molecular and Functional Basis of RHO GAP Family Proteins: A SYSTEMATIC APPROACH TOWARD SELECTIVE INACTIVATION OF RHO FAMILY PROTEINS*, *J Biol Chem* **291**, 20353 (2016).
- [361] H. Resat, T. P. Straatsma, D. A. Dixon, and J. H. Miller, *The Arginine Finger of RasGAP Helps Gln-61 Align the Nucleophilic Water in GAP-Stimulated Hydrolysis of GTP*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6033 (2001).
- [362] R. P. Joosten et al., *PDB\_REDO: Automated Re-Refinement of X-Ray Structure Models in the PDB*, *J Appl Crystallogr* **42**, 376 (2009).
- [363] J. K. Lanyi, *Proton Transfers in the Bacteriorhodopsin Photocycle*, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1757**, 1012 (2006).
- [364] C. Kottling and K. Gerwert, *Proteins in Action Monitored by Time-Resolved FTIR Spectroscopy*, *Chemphyschem* **6**, 881 (2005).
- [365] A. Barth and C. Zscherp, *What vibrations tell us about proteins*, *Q Rev Biophys* **35**, 369 (2002).
- [366] C. L. Parke, E. J. Wojcik, S. Kim, and D. K. Worthylake, *ATP hydrolysis in Eg5 kinesin involves a catalytic two-water mechanism*, *J Biol Chem* **285**, 5859 (2010).
- [367] Y. Yan, V. Sardana, B. Xu, C. Homnick, W. Halczenko, C. A. Buser, M. Schaber, G. D. Hartman, H. E. Huber, and L. C. Kuo, *Inhibition of a Mitotic Motor Protein: Where, How, and Conformational Consequences*, *J Mol Biol* **335**, 547 (2004).
- [368] J. C. Cochran, Joseph. E. Gatial III, T. M. Kapoor, and S. P. Gilbert, *Monastrol Inhibition of the Mitotic Kinesin Eg5*, *J Biol Chem* **280**, 12658 (2005).
- [369] J. S. Chappie, S. Acharya, M. Leonard, S. L. Schmid, and F. Dyda, *G Domain Dimerization Controls Dynamin's Assembly-Stimulated GTPase Activity*, *Nature* **465**, 435 (2010).
- [370] J. P. O'Donnell, L. J. Byrnes, R. B. Cooley, and H. Sonderrmann, *A Hereditary Spastic Paraplegia-Associated Atlastin Variant Exhibits Defective Allosteric Coupling in the Catalytic Core*, *J Biol Chem* **293**, 687 (2018).

- [371] Y.-S. Law, A. Utt, Y. B. Tan, J. Zheng, S. Wang, M. W. Chen, P. R. Griffin, A. Merits, and D. Luo, *Structural Insights into RNA Recognition by the Chikungunya Virus NsP2 Helicase*, Proc. Natl. Acad. Sci. U.S.A. **116**, 9558 (2019).
- [372] P. J. Focia, J. Gawronski-Salerno, Js. Coon, and D. M. Freymann, *Structure of a GDP:Alf4 Complex of the SRP GTPases Ffh and FtsY, and Identification of a Peripheral Nucleotide Interaction Site*, J Mol Biol **360**, 631 (2006).
- [373] N. Ostermann, D. Segura-Pena, C. Meier, T. Veit, C. Monnerjahn, M. Konrad, and A. Lavie, *Structures of Human Thymidylate Kinase in Complex with Prodrugs: Implications for the Structure-Based Design of Novel Compounds*, Biochemistry **42**, 2568 (2003).
- [374] A. Mateja, A. Szlachcic, M. E. Downing, M. Dobosz, M. Mariappan, R. S. Hegde, and R. J. Keenan, *The Structural Basis of Tail-Anchored Membrane Protein Recognition by Get3*, Nature **461**, 361 (2009).
- [375] F. Yi, R. Kong, J. Ren, L. Zhu, J. Lou, J. Y. Wu, and W. Feng, *Noncanonical Myo9b-RhoGAP Accelerates RhoA GTP Hydrolysis by a Dual-Arginine-Finger Mechanism*, J Mol Biol **428**, 3043 (2016).
- [376] V. G. Taylor, P. A. Bommarito, and J. J. Tesmer, *Structure of the Regulator of G Protein Signaling 8 (RGS8)-Galpaq Complex: MOLECULAR BASIS FOR Galpa SELECTIVITY*, J Biol Chem **291**, 5138 (2016).
- [377] W. Wu, K.-T. Park, T. Holyoak, and J. Lutkenhaus, *Determination of Structure of the MinD-ATP Complex Reveals the Orientation of MinD on the Membrane and the Relative Location of the Binding Sites for MinE and MinC*, Molecular Microbiology **79**, 1515 (2011).
- [378] T. A. Leonard, P. J. Butler, and J. Löwe, *Bacterial Chromosome Segregation: Structure and DNA Binding of the Soj Dimer - A Conserved Biological Switch*, EMBO Journal **24**, 270 (2005).
- [379] P. D. Coureux, H. L. Sweeney, and A. Houdusse, *Three Myosin V Structures Delineate Essential Features of Chemo-Mechanical Transduction*, EMBO J **23**, 4527 (2004).
- [380] M. L. Paddock, S. H. Rongey, G. Feher, and M. Y. Okamura, *Pathway of Proton Transfer in Bacterial Reaction Centers: Replacement of Glutamic Acid 212 in the L Subunit by Glutamine Inhibits Quinone (Secondary Acceptor) Turnover*, Proc Natl Acad Sci U S A **86**, 6602 (1989).
- [381] M. L. Paddock, P. H. McPherson, G. Feher, and M. Y. Okamura, *Pathway of Proton Transfer in Bacterial Reaction Centers: Replacement of Serine-L223 by Alanine Inhibits Electron and Proton Transfers Associated with Reduction of Quinone to Dihydroquinone*, Proc Natl Acad Sci U S A **87**, 6803 (1990).
- [382] M. L. Paddock, S. H. Rongey, P. H. McPherson, A. Juth, G. Feher, and M. Y. Okamura, *Pathway of Proton Transfer in Bacterial Reaction Centers: Role of Aspartate-L213 in Proton Transfers Associated with Reduction of Quinone to Dihydroquinone*, Biochemistry **33**, 734 (1994).
- [383] H. L. Axelrod, E. C. Abresch, M. L. Paddock, M. Y. Okamura, and G. Feher, *Determination of the Binding Sites of the Proton Transfer Inhibitors Cd<sup>2+</sup> and Zn<sup>2+</sup> in Bacterial Reaction Centers*, Proc Natl Acad Sci U S A **97**, 1542 (2000).
- [384] M. L. Paddock, G. Feher, and M. Y. Okamura, *Proton Transfer Pathways and Mechanism in Bacterial Reaction Centers*, FEBS Lett **555**, 45 (2003).

- [385] A. Y. Mulkidjanian, *Ubiquinol Oxidation in the Cytochrome Bc1 Complex: Reaction Mechanism and Prevention of Short-Circuiting*, *Biochim Biophys Acta* **1709**, 5 (2005).
- [386] P. Nogly et al., *Retinal Isomerization in Bacteriorhodopsin Captured by a Femtosecond X-Ray Laser*, *Science* **361**, eaat0094 (2018).
- [387] B. Schobert, L. S. Brown, and J. K. Lanyi, *Crystallographic Structures of the M and N Intermediates of Bacteriorhodopsin: Assembly of a Hydrogen-Bonded Chain of Water Molecules between Asp-96 and the Retinal Schiff Base*, *J Mol Biol* **330**, 553 (2003).
- [388] T. Marti, H. Otto, T. Mogi, S. J. Rosselet, M. P. Heyn, and H. G. Khorana, *Bacteriorhodopsin Mutants Containing Single Substitutions of Serine or Threonine Residues Are All Active in Proton Translocation*, *J Biol Chem* **266**, 6919 (1991).
- [389] J. Heberle, J. Fitter, H. J. Sass, and G. Buldt, *Bacteriorhodopsin: the functional details of a molecular machine are being resolved*, *Biophys Chem* **85**, 229 (2000).
- [390] Y. Cao, G. Varo, A. L. Klinger, D. M. Czajkowsky, M. S. Braiman, R. Needleman, and J. K. Lanyi, *Proton Transfer from Asp-96 to the Bacteriorhodopsin Schiff Base Is Caused by a Decrease of the PKa of Asp-96 Which Follows a Protein Backbone Conformational Change*, *Biochemistry* **32**, 1981 (1993).
- [391] P. Mitchell, *Epilogue: from energetic abstraction to biochemical mechanism*, *Symp, Soc. Gen. Microbiol* **383** (1977).
- [392] L. Drachev, M. Mamedov, A. Y. Mulkidjanian, A. Y. Semenov, V. Shinkarev, and M. Verkhovskiy, *Electrogenesis Associated with Proton Transfer in the Reaction Center Protein of the Purple Bacterium Rhodobacter Sphaeroides*, *FEBS Lett* **259**, 324 (1990).
- [393] S. Dehghani-Tafti, V. Levnikov, A. A. Antson, B. Bax, and C. M. Sanders, *Structural and Functional Analysis of the Nucleotide and DNA Binding Activities of the Human PIF1 Helicase*, *Nucleic Acids Res* **47**, 3208 (2019).

## 8. Summary

Though P-loop fold nucleoside triphosphatases (also known as Walker NTPases) are ubiquitous, their catalytic mechanism remains unclear. In these proteins, hydrolysis of ATP or GTP is initiated by interaction with an activating partner (usually another protein domain), which is accompanied by insertion of stimulatory moiety(ies) (usually arginine or lysine residues) into the catalytic site. Based on our automatized comparative structure analysis of 3136 Mg-NTP-containing catalytic sites, we identified those with stimulator(s) inserted into catalytic sites and analysed the patterns of stimulatory interactions. In most cases, at least one stimulator twists gamma-phosphate counter-clockwise by linking the oxygen atoms of alpha- and gamma-phosphates; the twisted gamma-phosphate is stabilized by a hydrogen bond with the backbone amino group of the fourth residue of the Walker A motif. In the remaining cases, the stimulators only interact with gamma-phosphate. The ubiquitous mechanistic interaction of diverse stimulators with the gamma phosphate group suggests its twist/rotation as the trigger for NTP hydrolysis. Based on our comparative structure analysis, we propose a common scheme of activated catalysis for P-loop NTPases. In this scheme, a hydrogen bond (H-bond) between the strictly conserved, Mg-coordinating Ser/Thr of the Walker A motif ([Ser/Thr]<sup>WA</sup>) and the conserved aspartate of the Walker B motif (Asp<sup>WB</sup>) plays the key role. We found that this H-bond is very short in the structures with bound transition state analogs. Given that a short hydrogen bond (also known as a low-barrier hydrogen bond) implies parity of pK values of the H-bond partners, we suggest that the proton affinities of these two residues reverse upon activation so that the proton relocates from [Ser/Thr]<sup>WA</sup> to Asp<sup>WB</sup>. The anionic [Ser/Thr]<sup>WA</sup> alkoxide withdraws then a proton from the would-be nucleophile (either a water molecule or a sugar moiety in some P-loop kinases), and the nascent anion attacks the gamma-phosphate group. When gamma-phosphate breaks away, the trapped proton relays from Asp<sup>WB</sup>, via [Ser/Thr]<sup>WA</sup>, to beta-phosphate and compensates for its developing negative charge.

## 9. Zusammenfassung

Obwohl die P-Loop-Nukleosidtriphosphatasen (auch Walker-NTPasen genannt) allgegenwärtig sind, bleibt ihr katalytischer Mechanismus unklar. Bei diesen Proteinen wird die Hydrolyse von ATP oder GTP durch eine Interaktion mit einem aktivierenden Partner veranlasst, die mit der Einführung einer oder mehrerer stimulierender Einheiten (in der Regel Arginin- oder Lysinreste) in das katalytische Zentrum einhergeht. Auf der Grundlage unserer automatisierten vergleichenden Strukturanalyse von 3136 Mg-NTP-haltigen katalytischen Zentren haben wir diejenigen identifiziert, bei denen ein oder mehrere Stimulatoren sich in den katalytischen Zentren befinden, und das Pattern der stimulierenden Interaktionen analysiert. In den meisten Fällen verdreht mindestens ein Stimulator das Gamma-Phosphat gegen den Uhrzeigersinn, indem er die Sauerstoffatome der Alpha- und Gamma-Phosphate miteinander verbindet; das verdrehte Gamma-Phosphat wird durch eine Wasserstoffbrückenbildung (H-Brücke) mit der Rückgrat-Aminogruppe des vierten Rests des Walker-A-Motivs stabilisiert. In den übrigen Fällen interagieren die Stimulatoren nur mit dem Gamma-Phosphat. Die ubiquitäre mechanistische Interaktion verschiedener Stimulatoren mit der Gamma-Phosphatgruppe lässt vermuten, dass deren Verdrehung als Auslöser für die NTP-Hydrolyse dient. Auf der Grundlage unserer vergleichenden Strukturanalyse beschreiben wir die aktivierte Katalyse durch P-Loop-NTPasen mit folgendem generellen Schema, in dem die H-Brücke zwischen dem streng konservierten, Mg-koordinierenden Ser/Thr des Walker-A-Motivs ( $[\text{Ser/Thr}]^{\text{WA}}$ ) und dem ebenso konservierten Aspartat des Walker-B-Motivs ( $\text{Asp}^{\text{WB}}$ ) eine Schlüsselrolle spielt. Wir haben festgestellt, dass diese H-Brücken in den Strukturen mit gebundenen Übergangszustandsanaloga sehr kurz sind. Da eine kurze H-Brücke (auch bekannt als H-Brücke mit niedriger Energiebarriere) die Parität der pK-Werte (Protonenaffinitäten) der H-Brücke-Partner voraussetzt, ist davon auszugehen, dass sich die Protonenaffinitäten dieser beiden Reste bei der Aktivierung der Katalyse umkehren, so dass das Proton von  $[\text{Ser/Thr}]^{\text{WA}}$  nach  $\text{Asp}^{\text{WB}}$  verlagert wird. Das anionische  $[\text{Ser/Thr}]^{\text{WA}}$ -Alkoxid entzieht dann dem potenziellen Nukleophil (entweder einem Wassermolekül oder einer Zuckereinheit in einigen P-Loop-Kinasen) ein Proton, und das entstehende nukleophile Anion greift die Gamma-Phosphatgruppe an. Wenn das Gamma-Phosphat abbricht, geht das gefangene Proton von  $\text{Asp}^{\text{WB}}$  via  $[\text{Ser/Thr}]^{\text{WA}}$  auf das Beta-Phosphat über und kompensiert seine entstehende negative Ladung.

## 10. Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AAA+	ATPases associated with diverse cellular activities ( <i>P-loop class</i> )
aa-tRNA	aminoacyl-tRNA
ACP	Adenosine 5'-[ $\beta$ , $\gamma$ -methylene]triphosphate
ADP	Adenosine-5'-diphosphate
AGS	Adenosine 5'-[ $\gamma$ -thio]triphosphate
ANP	Adenosine 5'-[ $\beta$ , $\gamma$ -imido]triphosphate
ASCE	"Additional Strand Catalytic E" ( <i>P-loop division</i> )
Asp <sup>WB</sup>	Conserved Asp/Glu residue of Walker B motif
ATP	Adenosine-5'-triphosphate
DNA	Deoxyribonucleic acid
DARR	dipolar assisted rotational recoupling
FRET	Fluorescence resonance energy transfer
GAP	GTPase-activating protein
GCP	Guanosine 5'-[ $\beta$ , $\gamma$ -methylene]triphosphate
GDP	Guanosine-5'-diphosphate
GNP	Guanosine 5'-[ $\beta$ , $\gamma$ -imido]triphosphate
GSP	Guanosine 5'-[ $\gamma$ -thio]triphosphate
GTP	Guanosine-5'-triphosphate
H-bond	Hydrogen bond
HMM	Hidden Markov model
LUCA	Last universal cellular ancestor
MD	Molecular dynamics
mRNA	Messenger RNA
MSA	Multiple sequence alignment
NDP	Nucleoside diphosphate
NMR	Nuclear magnetic resonance
NTP	Nucleoside triphosphate
P-loop	Phosphate-binding loop ( <i>motif</i> )
P-loop NTPases	P-loop containing nucleoside triphosphate hydrolase superfamily
PDB	Protein Data Bank
P <sub>i</sub>	Inorganic phosphate
RGS	Regulators of G-protein signaling
RMSD	Root mean square deviation
RNA	Ribonucleic acid

---

rRNA	Ribosomal RNA
SIMIBI	“Signal recognition particle, MinD and BioD” ( <i>P-loop class</i> )
SR	Signal recognition particle receptor
SRP	Signal recognition particle
TRAFAC	“Translation factors” ( <i>P-loop class</i> )
tRNA	Transfer RNA
WB	Walker B ( <i>motif</i> )

## 11. List of figures

- Figure 1.1: Protein Data Bank annual growth. 5
- Figure 1.2: P-loop fold NTPases. 8
- Figure 1.3: Nucleoside triphosphates (NTPs) and NTP hydrolysis. 9
- Figure 1.4: Transition state-like complexes of P-loop NTPases. 11
- Figure 1.5: Conserved features in  $Mg^{2+}$ :NTP binding in P-loop NTPases. 15
- Figure 1.6: Representative NTPases of the TRAFAC class. 18
- Figure 1.7: Representatives of the SIMIBI and kinase classes. 20
- Figure 1.8: Representatives of AAA+ ATPases, SF3 helicases, SF2 helicases and ABC ATPases. 23
- Figure 1.9: Representative proteins of the RecA/F1-like class of the P-loop NTPases. 25
- Figure 1.10: Uneven distribution of charged residues around nucleotide binding pockets in different classes of P-loop NTPases. 27
- Figure 2.1: Automated comparative structural analysis of P-loop NTPases. 33
- Figure 3.1: Variation in the  $AlF_4^-$  interaction with  $Mg^{2+}$  ions. 43
- Figure 3.2: Structural comparison of nucleotide-binding sites in SF4 helicase DnaB from *Geobacillus stearothermophilus* and ABC transporter MalK from *Escherichia coli* K-12. 45
- Figure 3.3: Distances between  $HN^{K-3}$  and the  $O^{2G}$  atom or its analog in the catalytic sites of P-loop NTPases. 47
- Figure 3.4: Conformation of native ATP molecules in selected crystal structures as compared to MD simulations of the MnmE GTPase. 48
- Figure 3.5: Examples of different interaction types/stimulatory patterns involving Arg residues. 52
- Figure 3.6: Examples of stimulatory interactions involving moieties other than arginine residues as stimulators. Residues are highlighted as in Fig. 3.5. 54
- Figure 3.7: Stimulators mechanically interact with the  $\gamma$ -phosphate in different classes of P-loop NTPases. 58
- Figure 3.8: Networks of water molecules in the catalytic sites of P-loop NTPases. 60
- Figure 3.9:  $Asp^{WB} - [Ser/Thr]^{K+1}$  distances in high-quality P-loop NTPase structures with bound Mg-NTP molecules or their analogs. 63
- Figure 3.10: Relative solvent-accessible surface area (SASA) values for  $Asp^{WB}$  residues. 64
- Figure 3.11: Putative protonic connections from  $W_{cat}$  to  $Asp^{WB}$  via short H-bonds between  $[Ser/Thr]^{K+1}$  and  $Asp^{WB}$  can be seen in different classes of P-loop NTPases. 65
- Figure 4.1: The transition-state analog  $AlF_4^-$  can adopt different orientations in diverse P-loop ATPases of the RecA class. 68
- Figure 4.2: Mechanisms of NTP hydrolysis. Dissociative (A), associative (B), and concerted pathways (C). 76



Figure 4.3: Proton traps in the photochemical reaction center and bacteriorhodopsin as compared with the Thr<sup>K+1</sup>-Asp<sup>WB</sup> pair of P-loop NTPase. 82

Figure 4.4: Schematic presentation of tentative proton routes along the edges of the octahedral coordination shell of Mg<sup>2+</sup> ion. 84

Figure 4.5: Two tentative mechanisms of proton transfer from W<sub>cat</sub> to Asp<sup>WB</sup> in P-loop NTPases. 87

Figure 4.6: The common scheme of stimulated hydrolysis in P-loop NTPases. 91

## 12. List of tables

Table 1.1: Most prevalent protein folds among solved structures in PDB (data retrieved on 07.11.2022) 6

Table 2.1: Properties recorded for each catalytic site analyzed 35

Table 3.1: Sampled catalytic sites of P-loop NTPases by type of NTP analog bound 40

## 13. Publications

Publications in peer-reviewed journals:

1. Lacabanne, D.; Wiegand, T.; Wili, N.; Kozlova, M. I.; Cadalbert, R.; Klose, D.; Mulkidjanian, A. Y.; Meier, B. H.; Bockmann, A.  
**ATP analogs for Structural Investigations: Case Studies of a DnaB Helicase and an ABC Transporter.**  
*Molecules* **2020**, *25* (22). <https://doi.org/10.3390/molecules25225268>.
2. Malär, A. A.; Wili, N.; Völker, L. A.; Kozlova, M. I.; Cadalbert, R.; Däpp, A.; Weber, M. E.; Zehnder, J.; Jeschke, G.; Eckert, H.; Böckmann, A.; Klose, D.; Mulkidjanian, A. Y.; Meier, B. H.; Wiegand, T.  
**Spectroscopic Glimpses of the Transition State of ATP Hydrolysis Trapped in a Bacterial DnaB Helicase.** *Nature Communications* **2021**, *12* (1), 5293.  
<https://doi.org/10.1038/s41467-021-25599-z>.
3. Kozlova, M. I.; Shalaeva, D. N.; Dibrova, D. V.; Mulkidjanian, A. Y.  
**Common Patterns of Hydrolysis Initiation in P-Loop Fold Nucleoside Triphosphatases.**  
*Biomolecules* **2022**, *12* (10), 1345. <https://doi.org/10.3390/biom12101345>.
4. Kozlova, M. I.; Shalaeva, D. N.; Dibrova, D. V.; Mulkidjanian, A. Y.  
**Common Mechanism of Activated Catalysis in P-Loop Fold Nucleoside Triphosphatases—United in Diversity.**  
*Biomolecules* **2022**, *12* (10), 1346. <https://doi.org/10.3390/biom12101346>.

## Conference talks and posters:

1. [BEST POSTER PRIZE] Shalaeva DS, Kozlova MI, Dibrova DV, Mulkidjanian AY (2020). **Multiple roles of the Walker A motif in the catalysis by P-loop fold NTPases.**  
*55th Winter Manfred Eigen's Seminar - Biophysical Chemistry, Molecular Biology and Cybernetics of Cell Functions, Klosters, Switzerland.*
2. [TALK AND ORGANIZATION] Kozlova MI. **P-loop fold NTPases: universal catalytic mechanism and enzyme-activator coevolution.**  
*Evocell Workshop — Cellular Mechanisms of Evolutionary Innovation*
3. [TALK] Kozlova MI (2021). **Proton transfer routes in P-loop fold nucleoside triphosphatases as inferred from comparative structure analysis**  
*Telluride Workshop on Proton Transfer in Biology, June 22-25<sup>th</sup> 2021, Telluride, CO, USA*
4. [TALK] Kozlova MI (2023). **Common catalytic mechanism of P-loop fold NTPases unraveled by large-scale comparative structural analysis**  
*56th Winter Manfred Eigen's Seminar - Biophysical Chemistry, Molecular Biology and Cybernetics of Cell Functions, Klosters, Switzerland.*

## 14. Acknowledgments

Firstly, I would like to express my deepest gratitude to my supervisor PD Dr Armen Y. Mulikidjanian for the opportunity to pursue a number of exciting projects in biophysics and early evolution, for guidance, invaluable life lessons, patience, continuous support, and sharing the understanding of how the scientific landscape is built in real life.

I would like to thank Osnabrück University for financial support throughout the years, as well as DAAD for short-term funding in 2021.

My deep appreciation goes to Dr. Daria Shalaeva that started the structural investigation of P-loop NTPases in our research group, for the introduction to the Osnabrück life, profound collaboration and friendship.

I am thankful to Dr. Daria Dibrova, for laying foundation for this project and for her mentorship since my times at the Moscow University.

I am very grateful to the members of the EvoCell graduate college, Prof. Dr. Roland Brandt, Prof. Dr. Christian Kost, Prof. Dr. Sabine Zachgo, Prof. Dr. Karin Frank and Prof. Dr. Florian Fröhlich, for mentoring, creating a collaborative and welcoming environment, introduction to interdisciplinary approaches and fruitful discussions. Additionally, I want to thank Prof. Dr. Christian Kost for the organization of EvoCell and its activities.

I would like to thank the EvoCell program graduate students, Dr. André Bogdanowski, Dr. Cilian Kock, Linea Muhsal, Sergej Limar, for navigating the research waters together during this years, and, specifically, my long-time friend and colleague Nataliya Trushina for all the adventures we had together.

I am very grateful to Prof. Dr. Heinz-Jürgen Steinhoff for his expertise and the possibility to use the infrastructure of his research group; to Dr. Johann Klare, for his support in the last years of the project; and to Dr. Natalia Voskoboinikova, Katharina Rudi, Alex Colbasevici, Dr. Magdalena Schumacher, Dr. Sabrina Dunkel and other members of our department, past and present, for the nice atmosphere and research environment, and for our cultural experiences in Osnabrück.

I must also thank Prof. Dr. Thomas Wiegand (MPI for chemical energy conversion, Mülheim) for collaboration and the chance to compare our knowledge derived from bioinformatics with the experimental NMR data.

Many thanks to Prof. Dr. Joachim Heberle (FU Berlin), Prof. Dr. Markus Sauer (Uni Würzburg) and Dr. Ramona Schlesinger (FU Berlin), for the invitations to the Winter Manfred Eigen's Seminar and the ability to discuss our findings there, for the great time at Klosters and valuable exchange on ongoing research and career paths in science.

Many thanks to Prof. Dr. A. N. Lupas (MPI for Developmental Biology, Tübingen) for discussions and input on the evolution of P-loop NTPases.

Special thanks to Prof. Dr. Jacob Piehler for the ability to perform measurements on the equipment at his department.

Thanks also to Dr. Kira S. Makarova (NCBI, Bethesda, USA) for sharing her knowledge on phylogenetic reconstructions.

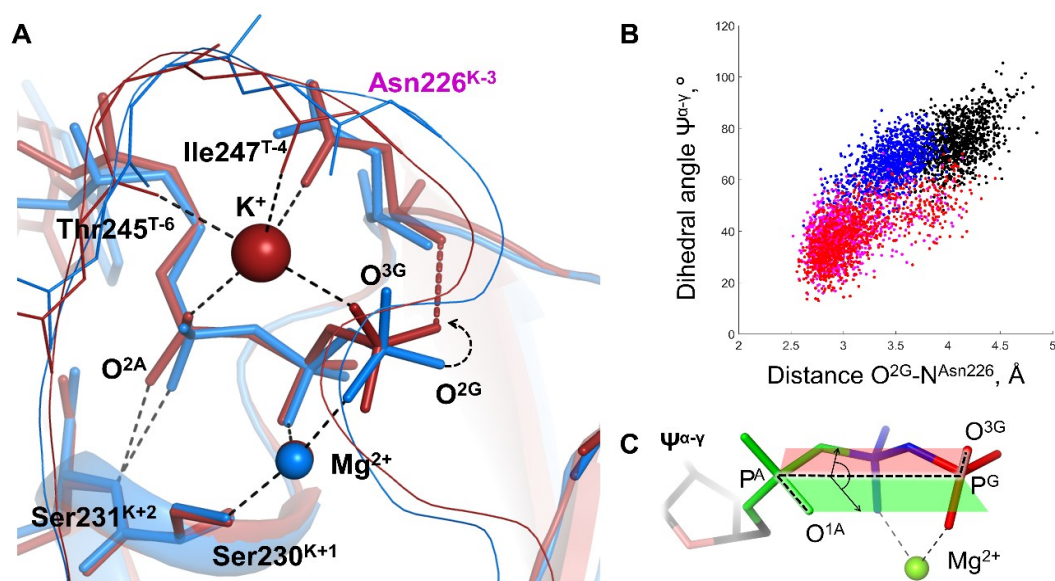
I am grateful to the secretaries of the Physics and Biology departments, Barbara Gunkel, Rita Krautwald, Michaela Horstkott, Yvonne Kirchhoff, for continuous help with the administrative matters.

I am thankful to my friends, Fernanda Gonçalves, João Ricardo Sater, Daniele Valentini, Dr. Franciele da Rocha, Mina Bakharzi, Nicolle Olsen, Alexandre Ederer and many others; without whom my years in Osnabrück wouldn't be the same.

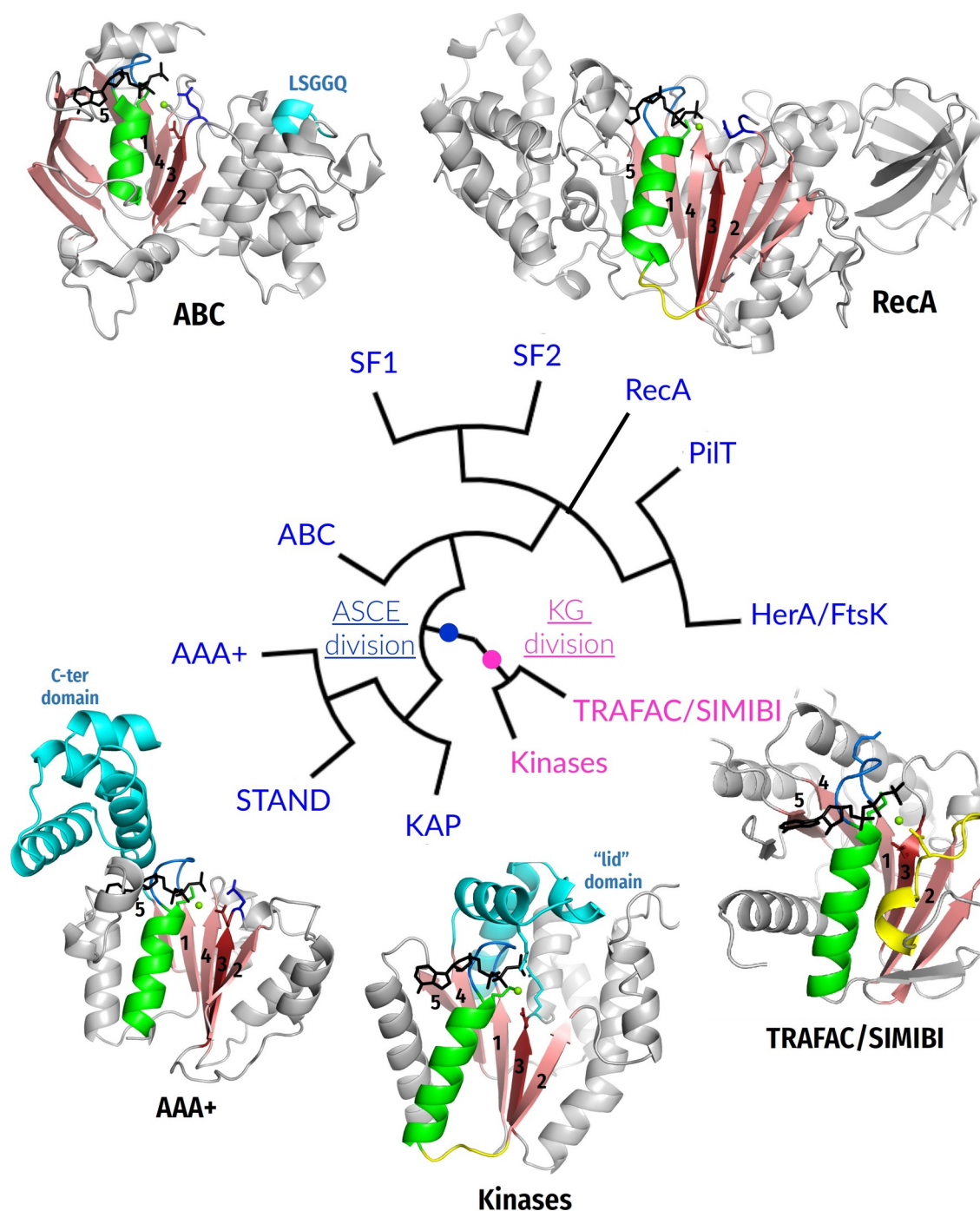
Finally, I can't find words to express my gratitude to my mother Tatiana V. Kozlova, for everything she has done for me and all the continuing trust in me, and my grandfather, Dr. Vladimir V. Verzbitsky, for starting the path of natural sciences in our family and being an inspiration for me.

# 15. APPENDIX

## A. Supplementary figures



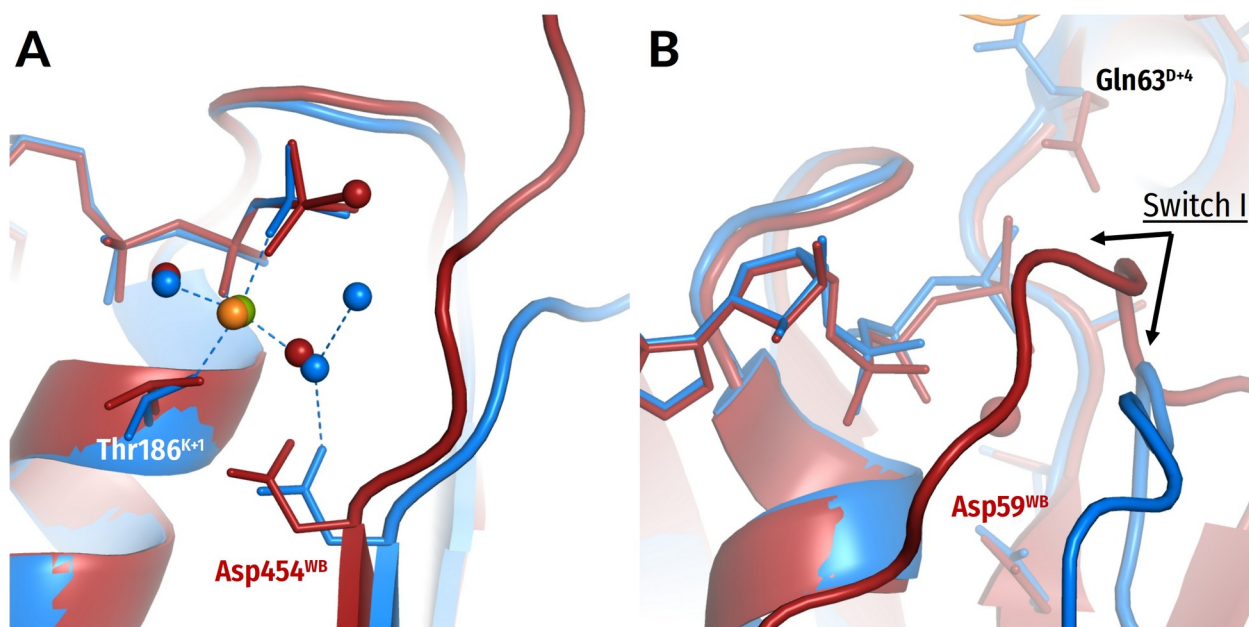
**Figure A.1. Molecular dynamics of the MnmE GTPase** (The figure is taken from [102] and modified). **A.** Superposition of the GTP-binding sites of the inactive, monomeric G-domain of MnmE (blue) and active,  $K^+$ -bound G-domain in a dimer (red); the representative structures were sampled from 100 ns simulations as described in [102]. The protein backbones are shown as cartoons; GTP and surrounding amino acid residues are shown as sticks;  $Mg^{2+}$  and  $K^+$  ions are shown as spheres. Black dashed lines indicate hydrogen bonds and coordination bonds for cations that are present in both structures; the red dashed line indicates the H-bond between  $NH^{K-3}$  and  $O^{2G}$  that is present only in the  $K^+$ -containing dimer. **B.** Conformational space of GTP in different states of MnmE GTPase. Scatter plot of the  $\Psi^{\alpha-\gamma}$  dihedral angle (Y-axis) against the distance between the  $O^{2G}$  atom and  $NH$  of  $Asn226^{K-3}$  (X-axis) as sampled from the MD simulations of three systems: (1) active dimer of G-domains with  $K^+$  ions bound (red and magenta for individual monomers); (2) monomeric G-domain of MnmE with the  $K^+$  ion replaced by a water molecule, blue; and (3) inactive monomer G-domain of MnmE without a full-fledged K-loop, black. **C.** The dihedral angle  $\Psi^{\alpha-\gamma}$  in the phosphate chain of GTP as measured for the plot on panel B.



**Figure A.2.** Cladogram of higher-order relationships between major divisions/classes of P-loop NTPases with depicted typical structures, taken from [109].

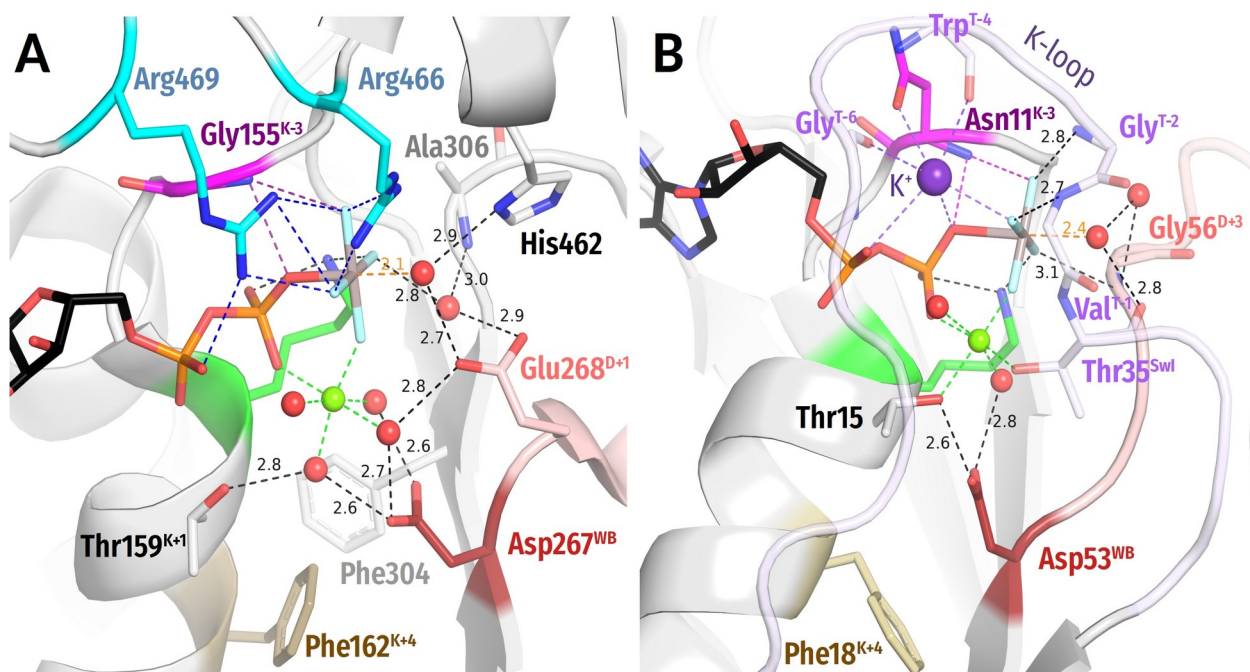
The cladogram (as modified from [37]) shows the two major divisions: the kinase-GTPase (KG) division and the ASCE (additional strand, catalytic E) division. The  $\beta$ -strands forming the cores of P-loop domains are numbered in a traditional way [32,37]. These  $\beta$ -strands are colored pink, the P-loop is shown in blue, the following  $\alpha_1$ -helix is shown in green, the K-loop/Switch I in TRAFAC class NTPases as well as the corresponding loops in other NTPases are colored in yellow, NTP analogs are shown as black sticks,  $Mg^{2+}$  ions are shown as green spheres, the rest of proteins is shown as gray cartoons. The reference residues of Walker A and Walker B motifs (lysine (K) and aspartate (D), respectively), as well as the catalytic glutamate (E) in ASCE NTPases are shown as sticks. The following reference crystal structures are depicted: ABC – antibacterial peptide ABC transporter McjD of *Escherichia coli*, PDB ID 5EG1 [338]; RecA/F<sub>1</sub> –F<sub>1</sub>-ATPase of *Caldalakilabacillus thermarum*, PDB ID 5HKK [339]; AAA+ –clamp loader  $\gamma$ -subunit *E. coli*, PDB ID 1NJG [340]; Kinases –thymidylate kinase of *E. coli*, PDB ID 4TMK [341]; TRAFAC/SIMIBI –nitrogenase ATPase subunit of *Azotobacter vinelandii*, PDB ID 4WZB [342].





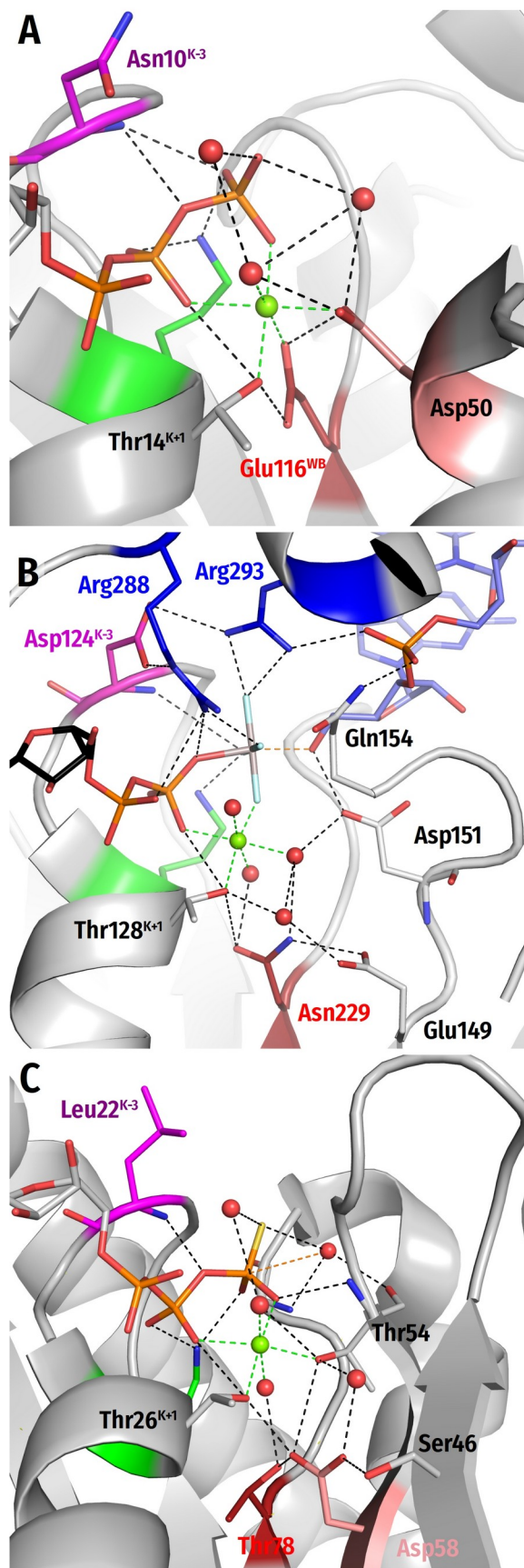
**Figure A.3.** Structures of same P-loop NTPases with transition state analogs and substrate/substrate analogs bound, respectively, as modified from [109].

The structures with bound substrate or substrate analogs are shown in blue with their Mg<sup>2+</sup> ions in green; the structures with transition state analogs bound are shown in dark red with their Mg<sup>2+</sup> ions in orange. Cations and water molecules are shown as spheres. The following crystal structures are depicted: **A**, Myosin II from *Dictyostelium discoideum* in complex with a bound ATP molecule (blue, PDB ID 1FMW [214]) and with a transition state analog ADP:VO<sub>4</sub> bound (dark red, PDB ID 1VOM [215]); **B**, Ras-like GTPase RhoA with a bound non-hydrolyzable substrate analog GNP (blue, PDB 6V6M, [343]) and a complex of Rho with RhoGAP and a transition state analog GDP:MgF<sub>3</sub><sup>-</sup> (dark red, ODB 1OW3, [70]).



**Figure A.4. Interference of Phe<sup>K+4</sup> with the Asp<sup>WB</sup> – [Ser/Thr]<sup>K+1</sup> bonding.**

**A.** The only detected structure with properly bound  $\text{AlF}_4^-$  and, still, long distance between Asp<sup>WB</sup> and [Ser/Thr]<sup>K+1</sup> is the mitochondrial yeast DEAD box protein Mss116p where this distance is about 4 Å, see Panel A and (PDB 3I62 [344]). The structure analysis has shown that the formation of the H-bond between Asp<sup>WB</sup> and [Ser/Thr]<sup>K+1</sup> is prevented by the Phe162<sup>K+4</sup> residue. It interacts with Phe304 of the WB-1 strand so that these two residues intercalate between Walker A and Walker B motifs and prevent the formation of the H-bond (Panel A). Remarkably, the Thr<sup>K+1</sup> is not directly coordinating Mg<sup>2+</sup>, pointing to a possibility of a crystallization artefact or an inactive configuration of the binding site with the Phe<sup>K+4</sup> acting as a regulatory switch. **B.** A phenylalanine residue is not unique in K+4 position, it is also found in the well-studied K<sup>+</sup>-stimulated GTPase FeoB. In its GDP:AlF<sub>4</sub><sup>-</sup>-containing structure (PDB ID 7BWV), the homologous Phe18<sup>K+1</sup> is, however, in a different rotameric position and do not interfere with the Asp<sup>WB</sup> – [Ser/Thr]<sup>K+1</sup> bond (Panel B). Notably, in this structure the residue is not involved in a  $\pi$ - $\pi$  interaction with another phenylalanine residue that would orient it towards the Asp<sup>WB</sup> – Thr<sup>K+1</sup> bond (cf Panel A).



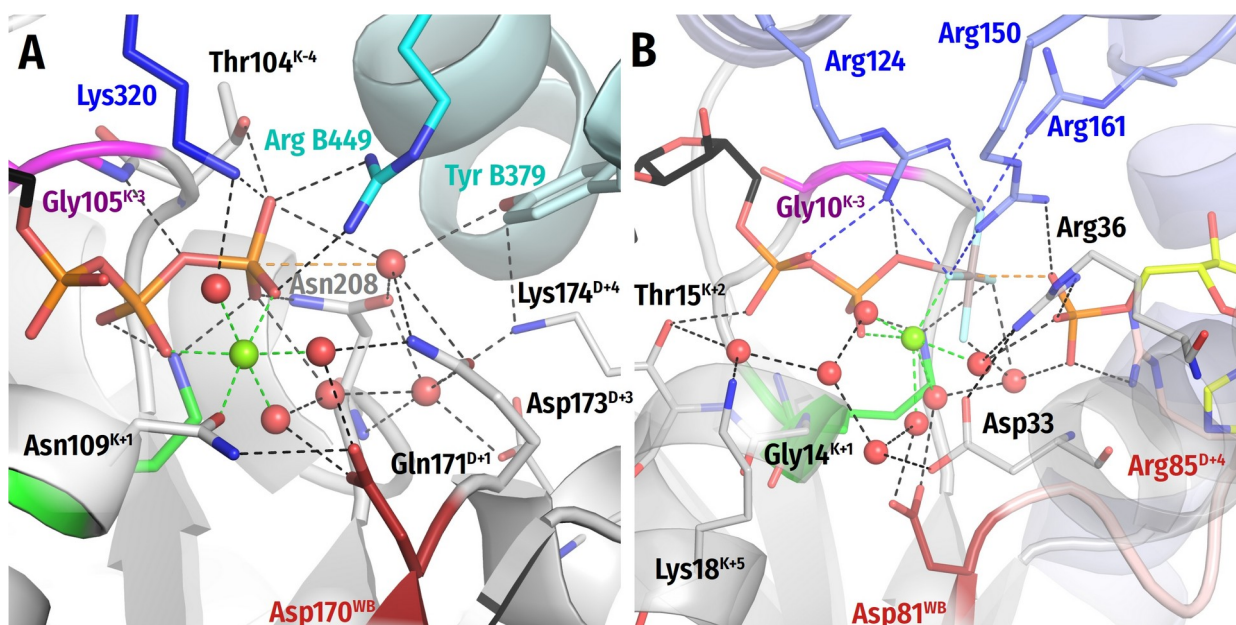
**Figure A.5. Deviations from a typical Walker B motif. Color code as in Fig. 1.2E,F.**

**A.** In a few cases, which are listed in [50], a glutamate residue occupies the position of Asp<sup>WB</sup>. In dethiobiotin synthetase, the long side chain of Glu<sup>WB</sup> reaches Mg<sup>2+</sup> and directly coordinates it, forming also a H-bond with [Ser/Thr]<sup>K+1</sup>, as shown in Panel A for the enzyme from *Helicobacter pylori* complexed with GTP (PDB 3QXJ [345]). Dethiobiotin synthetase catalyzes the formation of dethiobiotin from (7R,8S)-diaminononanoic acid (7,8-diaminopelargonic acid, DAPA) and CO<sub>2</sub>, in course of the reaction the  $\gamma$ -phosphate group of ATP is transferred to the carboxy group of the reaction intermediate, 7-carbamate of DAPA, producing a mixed anhydride intermediate that then cyclizes to dethiobiotin while the phosphate is expelled [346]. This carboxy group stems from CO<sub>2</sub> and is likely to be deprotonated at a neutral pH, so that no proton trapping is needed. The carbamate intermediate is thought to be stabilized by the neighboring protonated aminogroup (at the position 8 of the substrate) [347]. In a few more conventional cases, the Glu<sup>WB</sup> residue, similarly to Asp<sup>WB</sup>, does not reach Mg<sup>2+</sup>, interacts with Mg<sup>2+</sup>-coordinating water and makes an H-bond with [Ser/Thr]<sup>K+1</sup>, see [50,348].

**B.** An Asn229 residue links W3, W6 and Thr131<sup>K+1</sup> in the wild-type polynucleotide kinase of *Caenorhabditis elegans*, see Panel B for its structure with bound ssRNA dinucleotide GC and ADP:AlF<sub>4</sub><sup>-</sup> (PDB ID 4OI1 and [50,348]). In this case, however, the adjacent WB+1 strand has a <sup>149</sup>ELD<sup>151</sup> triad at its C-cap. In the TS-like AlF<sub>4</sub><sup>-</sup>-containing structure, the Glu149 residue is connected via a bridging water molecule to Thr131<sup>K+1</sup> and to W3, whereas Asp151 makes a short H-bond (2.48 Å) with the hydroxyl of the RNA ribose. Deprotonation of this hydroxyl yields the nucleophile that attacks the  $\gamma$ -phosphate. We believe that the functions of Asp<sup>WB</sup> are divided between Asn229 and Glu149 in this polynucleotide kinase. The Asn229<sup>WB</sup> residue serves as a structural linker to the Walker A motif, whereas Glu149 serves as a trap for the proton that is taken by Asp151 from W<sub>cat</sub>.

While the same Asn-Glu-Asp triad is found in other Metazoa enzymes, Asp substitutes for Asn229 and Asn substitutes for Glu149 in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans*. Hence, in evolutionarily primary primitive microorganisms, a single Asp<sup>WB</sup> appears to interact with W6 and Thr<sup>K+1</sup> and to obtain a proton from the catalytic Asp residue at the C-cap of the adjacent WB+1 strand; in the *C. elegans* structure (PDB ID 4OI1 [348]), Asn229 and Asp 151 are linked by W6. The “primordial” dyad of aspartates is functionally similar but structurally distinct from the Asp<sup>WB</sup>L<sup>D+1</sup>Asp<sup>D+2</sup> aspartate dyads found in some other kinase families, see Fig. 1.7D and [40]. Generally, the overall diversity of proton routes within kinases [40] might reflect the diversity of second substrates in these enzymes, see Section 2.2.

**C.** A natural mutation in an otherwise typical small TRAFAC class GTPase MglA changed Asp<sup>WB</sup> to Thr. The structure of MglA GTPase of *Myxococcus xanthus* with a bound non-hydrolyzable analog of GTP is shown (PDB ID: 6IZW and [349]). The enzyme retained its function owing to the appearance of an Asp residue at the N-terminus of the adjacent antiparallel WB+1 strand [50]. The resulting topology of the catalytic site is typical for P-loop NTPases.



**Fig. A.6. Residues other than [Ser/Thr] in the K+1 position of the Walker A motif.**

Colors as in Fig. 1.7-1.9 in the main text.  $Lys^{WA}$  shown in green,  $Asp^{WB}$  in dark red, adjacent monomer and its residues in cyan, Arg and Lys finger from the same polypeptide chain in blue. **A**, an asparagine residue replaces  $[Ser/Thr]^{K+1}$  in such distantly related AAA+ ATPases as torsinA of metazoa and DnaC helicase loader of some enterobacteria including *E. coli* [50,331,350]. In addition, the potentially  $W_{cat}$ -stabilizing “catalytic”  $[Glu/Asp]^{D+1}$  – ubiquitous in AAA+ ATPases [111] – is replaced by  $Asn^{D+1}$  in torsinA (but not in DnaC). The ATPase activity of both these proteins are very low, so it is not yet clear if they are true ATPases or simply work as switches that change their conformation depending on whether ATP or ADP is bound in the catalytic site [331,350–352]. Shown is the high-resolution ATP-containing structure of human torsinA in complex with its activator LULL1 (PDB ID 5J1S [353]). It can be seen that  $Asn109^{K+1}$  makes the canonic H-bond with  $Asp170^{WB}$  which could increase the proton affinity of the latter. While  $Asn171^{D+1}$  cannot transfer a proton from  $W_{cat}$ , a water chain – similar to one observed in TRAFAC class NTPases (Fig. 7B, 12D-F) – connects  $W_{23}$  with  $Asp170^{WB}$  via the  $Mg^{2+}$ -coordinating  $W6$  molecule. In AAA+ NTPases, a Lys/Arg residue always interacts with  $\gamma$ -phosphate and, supposedly,  $W_{cat}$  (Fig. 1.8A and [65,67]). In torsinA, this is the Arg449 residue of LULL1. In the absence of negatively charged  $Asp/Glu^{D+1}$ , the interaction of a positively charged Arg449 with  $W_{cat}$  by dramatically decreasing its proton affinity, may trigger proton transfer from  $W_{cat}$  via  $W6$ , to  $Asp170^{WB}$ . Subsequently, the proton can move to  $\beta$ -phosphate via  $W6$ . Hence, the structure is compatible with involvement of  $Asp170^{WB}$  as a proton trap in torsinA and could explain its, albeit slow, ATPase activity.

**B**, in Adenylate kinase of *Aquifex aeolicus* (PDB 3SR0 [140]) a water molecule serves as the  $Mg^{2+}$  ligand #4. On the one hand, the successful involvement of a water molecule as the 4<sup>th</sup> ligand of  $Mg^{2+}$  in these kinases indicates that the conservation of  $[Ser/Thr]^{K+1}$  in all other classes of P-loop NTPases may be unrelated to its function as a  $Mg^{2+}$  ligand. On the other hand, these multiple losses of  $[Ser/Thr]^{K+1}$  provide additional support for here proposed mechanism. It is in these P-loop kinases that the attacking nucleophile – the anionic phosphate moiety of a nucleotide monophosphate – needs no deprotonation and, therefore, does not need a catalytic alkoxide. As noted earlier, kinases do not interact with separate activator proteins; they are activated by binding the second substrate, which causes covering of the catalytic site by the Lid domain and the insertion of stimulatory finger(s) (Fig. 1.7C-D). Therefore, there is always some danger that the insertion of the finger(s) will stimulate a nucleophilic attack on  $\gamma$ -phosphate not by the anionic phosphate group of the second substrate but by a haphazard water molecule, leading to a futile ATP hydrolysis. Within our proposed scheme, an unwanted ATP hydrolysis can be prevented by replacing  $[Ser/Thr]^{K+1}$  with a residue incapable of accepting a proton from water. And that is what independently happened in several lineages of nucleotide monophosphate kinases, see also multiple alignments in [40]. In particular, all the human adenylate kinases have glycine residues in the K+1 position except adenylate kinase 6, which has a threonine residue [354]. And it is for this kinase that both kinase and ATPase activities have been shown [355]. Hence, the consistent loss of  $[Ser/Thr]^{K+1}$  – independently in several families of nucleotide monophosphate kinases [40] – finds a plausible explanation.

## B. Table B.1: Description of representative structures of P-loop NTPases

Table B.1: Structural traits of representative P-loop NTPases of different classes.

P-Loop Class, Activation Mechanism	Representative Protein Structure	Site ID in Table C.1	Figure in the Text	PDB Entry ID, Resolution	NTP or NTP Analog	Walker A Motif (K-3 and K+1 Residues)		Walker B Asp/Glu	Stimulatory Moieties	Coordination of $W_{cat}^a$
<b>Kinase-GTPase Division</b>										
<b>TRAFAC</b> Interaction with the activating partner or dimerization in the presence of the activating partner leads to the stabilization of the Switch I loop and insertion of diverse stimulatory moieties into the catalytic site	RhoA	3508	1.4C	1OW3, 1.8 Å	GDP:MgF <sub>3</sub> <sup>-</sup>	Ala15	Thr19	Asp59	Arg85 <sup>+</sup> -NH2 (AG)	Gln63 <sup>D+4</sup> -OE1, Thr37 <sup>Sw1</sup> -CO
	MnmE	236	1.2B,E	2GJ8, 1.7 Å	GDP:AlF <sub>4</sub> <sup>-</sup>	Asn226	Ser230	Asp270	K <sup>+</sup> ion (AG)	Thr251 <sup>Sw1</sup> -CO, Gly249 <sup>T-2</sup> -HN, Thr250 <sup>T-1</sup> -HN, Gly273 <sup>D+3</sup> -HN, Gly273 <sup>D+3</sup> -CO*
	Dynammin	3550	1.6A	2X2E, 2.0 Å	GDP:AlF <sub>4</sub> <sup>-</sup>	Ser41	Ser45	Asp136	K <sup>+</sup> or Na <sup>+</sup> ion (G)	Thr65 <sup>Sw1</sup> -CO, Gly139 <sup>D+3</sup> -HN, Gly139 <sup>D+3</sup> -CO*, Gln40 <sup>K-4</sup> -OE1*
	Atlastin	3661	1.6B	6B9F, 1.9 Å	GDP:AlF <sub>4</sub> <sup>-</sup>	Arg77	Ser81	Asp146	Arg77 <sup>K-3</sup> -NH2 (AG)	Gly149 <sup>D+3</sup> -HN, Thr120 <sup>Sw1</sup> -CO, Gly149 <sup>D+3</sup> -CO*, Asp152 <sup>D+6</sup> -OD2*
	Gα <sub>3</sub>	3545	1.6C	2ODE, 1.9 Å	GDP:AlF <sub>4</sub> <sup>-</sup>	Glu43	Ser47	Asp200	Arg178 <sup>T-3</sup> -NH1 (AG)	Thr181 <sup>Sw1</sup> -CO, Gln204 <sup>D+4</sup> -OE1, Gly203 <sup>D+3</sup> -HN
Myosin II	42	1.6D	1VOM, 1.9 Å	ADP-VO <sub>4</sub> <sup>3-</sup>	Gly182	Thr186	Asp454	Asn233 <sup>S-4</sup> -ND2 (AG)	Ser237 <sup>Sw1</sup> -CO, Ser236 <sup>S-1</sup> -OG, Gly457 <sup>D+3</sup> -HN, Gly457 <sup>D+3</sup> -CO, Arg238 <sup>S+1*</sup> -NH1, Glu459 <sup>D+5*</sup> -OE1	
<b>SIMIBI</b> Monomeric domains dimerize and provide activating Lys or Arg fingers for each other in response to the interaction with the activating partner	GET3	64	1.7A	2YNM, 2.1 Å	ADP:AlF <sub>4</sub> <sup>-</sup>	Gly39	Ser43	Asp151	Lys37 <sup>†</sup> -NZ (AG)	Asp66 <sup>WB+1</sup> -OD2, Asp155 <sup>†</sup> -OD1, Lys68 <sup>WB+1</sup> -NZ, Gly154 <sup>D+3</sup> -HN, Lys37 <sup>†</sup> -HN*
	Signal recognition particle	3522	1.7B	2CNW, 2.39 Å	GDP:AlF <sub>4</sub> <sup>-</sup>	Gly108	Thr112	Asp187	Arg138-NH1 (AG)	Gly190 <sup>D+3</sup> -HN, Asp135 <sup>WB+1</sup> -OD2, GTP <sup>†</sup> -O3 <sup>*</sup> , Gly190 <sup>D+3</sup> -CO*, Glu284 <sup>†</sup> -OE1*
<b>Kinases</b> Rearrangement of the lid domain upon binding of the second substrate leads to the insertion of Arg/Lys fingers into the catalytic site	Thymidylate kinase	1207	1.7C	1NN5, 1.5 Å	ANP	Arg16	Ser20	Asp96	Arg16 <sup>K-3</sup> -NH1 (AG?), Arg97 <sup>D+1</sup> -NH2 (G)	The second substrate is coordinated by Arg45 <sup>WB+1</sup> -NH2, Arg-97 <sup>D+1</sup> -NE, Glu149 <sup>Lid</sup> -OE1, Pro43-CO*
	Adenosine 5'-phosphosulfate kinase	1490	1.7D	4BZX, 1.7 Å	ANP	Gly453	Ser457	Asp478	Lys562 <sup>Lid</sup> -NZ (G)	The second substrate is coordinated by Asp41 <sup>D+2</sup> -OD1, Lys562 <sup>Lid</sup> -NZ, Arg483 <sup>D+5</sup> -NH2, Arg497-NH1
	Adenylate kinase	xxx0	A.6B	3SR0, 1.6 Å	ADP:AlF <sub>4</sub> <sup>-</sup>	Gly10	Gly14	Asp81	Arg124 <sup>Lid</sup> -NH1 (AG), Arg124 <sup>Lid</sup> -NH2 (G), Arg150 <sup>Lid</sup> -NH1 (G), Arg161 <sup>Lid</sup> -NH1 (G)	The second substrate (phosphate acceptor) is coordinated by Arg150 <sup>Lid</sup> -NH2, Arg85 <sup>D+4</sup> -NH1, Arg85 <sup>D+4</sup> -NH2, Arg36 <sup>WB+1</sup> -NH1, Arg36 <sup>WB+1</sup> -NH2

Table B.1: Description of representative structures of P-loop NTPases

P-Loop Class, Activation Mechanism	Representative Protein Structure	Site ID in Table C.1	Figure in the Text	PDB Entry ID, Resolution	NTP or NTP Analog	Walker A Motif (K-3 and K+1 Residues)		Walker B Asp/Glu	Stimulatory Moieties	Coordination of $W_{cat}^a$
<b>ASCE Division</b>										
<b>AAA+/SF3 ATPases</b> In a hexamer, the binding /hydrolysis of ATP in one subunit causes conformational changes activating the adjacent subunit; the activation involves class-specific helical domain	N-ethylmaleimide sensitive factor	359	1.8A	1NSF, 1.9 Å	ATP	His546	Thr550	Asp603	Lys708†-NZ (AG) Lys631†-NZ (G)	Asp604 <sup>D+1</sup> -OD2, Lys631†-NZ, Ser647 <sup>WB-1</sup> -OG
	SV40 large T antigen helicase (SF3)	372	1.8B	1SVM, 1.9 Å	ATP	Asp429	Thr433	Glu473	Lys418†-NZ (AG), Arg540†-NH2 (G)	Asp474 <sup>E+1</sup> -OD1, Asn529 <sup>WB-1</sup> -OD1, Arg498†-NH1, Arg540†-NH2
	Chikungunya virus nsP2 helicase (SF1)	N/A	1.4D	6JIM, 2.0 Å	ADP:AIF <sub>4</sub> <sup>-</sup>	Gly189	Ser193	Asp252	Arg312‡-NH2 (AG), Arg416‡-NH1 (G), Arg416‡-NH2 (G), Arg467‡-NH2 (AG), Arg467‡-NH1 (G), Arg467‡-NH2 (G), Arg464‡-NH1 (G), Arg464‡-NH2 (G),	Glu253 <sup>D+1</sup> -OE2, Gln283 <sup>D+1</sup> -OE2, Gly384‡-HN
<b>Helicases SF1/2:</b> Rearrangement of the C-terminal domain upon DNA or RNA binding leads to the insertion of an Arg finger into the nucleotide-binding domain	HCV NS3 helicase (SF2)	136	1.8C	3KQL, 2.5 Å	ADP:AIF <sub>4</sub> <sup>-</sup>	Gly207	Ser211	Asp290	Arg467‡-NH2 (AG), Arg467‡-NH1 (G), Arg467‡-NH2 (G), Arg464‡-NH1 (G), Arg464‡-NH2 (G),	Glu291 <sup>D+1</sup> -OE1, Gln460‡-OE1, Arg464‡-NH2, Gly417‡-NH, Ala323-NH <sup>†WB+1*</sup>
	Multifunctional helicase Pif1p (SF1)	233	1.2A, D	5O6B, 2.0 Å	ADP:AIF <sub>4</sub> <sup>-</sup>	Gly261	Ser265	Asp-341	Arg417‡-NH2 (AG)	Glu342 <sup>D+1</sup> -OE1, Gln381 <sup>WB-1</sup> -OE1, Arg734‡-NH1, Gly709‡-HN
<b>ABC ATPases:</b> Monomeric domains dimerize and provide activating LSGGQ motifs for each other in response to the substrate binding	Maltose transporter	154	1.8D	3PUW, 2.2 Å	ADP:AIF <sub>4</sub> <sup>-</sup>	Gly39	Ser43	Asp158	Ser135†-OG (G), Gly137†-HN (G)	Gln82 <sup>WB+1</sup> -NE2, Glu159 <sup>D+1</sup> -OE1, Glu159 <sup>D+1</sup> -OE2, Asn163†-CO, His192 <sup>WB-1</sup> -NE2
	F <sub>1</sub> -ATPase	18	1.9A	1H8E, 2.0 Å	ADP:AIF <sub>4</sub> <sup>-</sup>	Gly159	Thr163	Asp256	Arg373†-NH1 (AG), Arg189 <sup>WB-1</sup> -NH1 (G), Arg189 <sup>WB-1</sup> -NH2 (G),	Glu188 <sup>WB-1</sup> -OE1, Arg260 <sup>D+4</sup> -NH2, Ser344†-CO
<b>RecA/F1-like:</b> In an oligomer, ATP binding/hydrolysis in one subunit causes conformational changes that activate the adjacent subunit by inserting Arg/Lys fingers	Replicative helicase DnaB	N/A	1.9B	6T66, 3.9 Å	GDP:AIF <sub>4</sub> <sup>-</sup>	Ser231	Thr235	Asp340	Arg439†-NE (G), Lys437†-NZ (G)	Glu259 <sup>WB+1**</sup> , Tyr341 <sup>D+1**</sup> Gln381 <sup>WB-1**</sup> Arg439†
	Circadian clock protein KaiC	683	1.9C	4TL7, 1.9 Å	ATP	Gly49	Thr53	Asp145	Lys224†-NZ (G) Arg226†-NH2 (G)	Glu183 <sup>WB-1</sup> -OE1**, Ser146 <sup>D+1**</sup> , Phe199†-CO
	Recombinase RadA	1376	1.9D	3EW9, 2.4 Å	ANP	Gly108	Thr112	Asp211	K <sup>+</sup> -503 (G), K <sup>+</sup> -504 (G)	Glu151 <sup>WB+1**</sup> , Ser212 <sup>D+1**</sup> , Gln257 <sup>WB-1**</sup>
	RecA	100	N/A	3CMX, 3.4 Å	ADP:AIF <sub>4</sub> <sup>-</sup>	Ser69	Thr73	Asp144	Lys248†-NZ (G) Lys250†-NZ (G)	Glu96 <sup>D+1**</sup> , Phe216 <sup>**†</sup> -CO, Gln194 <sup>**D-1</sup>

Residue numbers are as in the listed PDB structures; <sup>a</sup>—polar atoms, located within 3.6 Å from the catalytic water molecule; AG—the stimulatory residue inserts between  $\alpha$ -phosphate and  $\gamma$ -phosphate (or its mimic). G—the stimulatory residue is only coordinating  $\gamma$ -phosphate or its mimic. †—residue from a polypeptide chain other than the P-loop containing one; ‡—residues from a domain other than the described P-loop domain; \*—residue coordinates the attacking water molecule via another water molecule (e.g., Figure 1.7B); \*\*—catalytic water molecule not resolved, the coordinating residue(s) were inferred from structure superposition and literature data.

## C. Table C.1: Results of the computational analysis of all available structures of the P-loop NTPases with full-fledged catalytic sites containing Mg-NTP substrate or its analogs

The results of automated survey of 3136 catalytic sites of P-loop NTPases are presented in Table C.1, provided as a Microsoft Excel table due to large table size. The table is available from <https://www.mdpi.com/article/10.3390/biom12101345/s1> as Table S1.

The Sheet A of the Microsoft Excel table contains the list with characteristics of all analyzed structures, together with indicated key functional residues of the Walker A and Walker B motifs, Arg, Lys, and Asn fingers, as well as distances from (1) the respective atoms of NTPs/their analogs to the K-3 residues and Arg/Lys fingers, as well as (2) from [Asp/Glu]<sup>WB</sup> to [Ser/Thr]<sup>K+1</sup>. Each row contains the data for one catalytic site in one structure. Catalytic sites containing “properly bound” NDP:AlF<sub>4</sub><sup>-</sup> complexes that we deemed to be reliable TS analogs (see Table D.1 for detailed description of AlF<sub>4</sub><sup>-</sup> and Mg<sup>2+</sup> binding in such structures) are marked with “y” or “\*” in column “site rel”; they are colored green. The sites with “improperly bound” NDP:AlF<sub>4</sub><sup>-</sup> complexes are colored pink. All columns present in the Sheet A of the Table C.1 (data) are described in the Sheet B.

## D. Table D.1: Coordination of the $\text{Mg}^{2+}$ ion in the $\text{AlF}_4^-$ -containing structures of P-loop NTPases

This Excel Table contains the results of the evaluation of  $\text{Mg}^{2+}$  coordination in all analyzed structures that contain  $\text{NDP}:\text{AlF}_4^-$ , such as list of distances between  $\text{Mg}^{2+}$  and its ligands and notes on the coordination sphere properties in each catalytic site in each structure.

Available from <https://www.mdpi.com/article/10.3390/biom12101345/s1> as Table S3.



## **E. Table E.1: Relative occurrence of distinct stimulatory patterns for Arg, Lys, and Asn “fingers”**

This Excel Table contains the data on occurrence of different stimulatory patterns clustered by type of the stimulatory residue and nucleotide mimic.

Available from <https://www.mdpi.com/article/10.3390/biom12101345/s1> as Table S2.

## F. Notes on rarity of Y-pattern of stimulatory Arg interaction

Our analysis revealed only 33 complexes with Y-like interaction of an Arg finger with an NTP molecule or its analog (see also Table C.1). The Y-pattern is not observed in a single structure with a bound TS analog, and it is such structures that enable us to judge with certainty the stimulatory pattern in a particular ATPase. Therefore, we inspected these 33 complexes manually.

- I. Seventeen of these binding sites belong to the structures obtained via electron microscopy (EM), mostly with resolution worse than 3.5 Å. Eleven of these NTP binding site structures are of catalytically inactive ATP-binding sites of  $\alpha$ -subunits of F1-type ATPases, from *Sus scrofa* (PDB 6J5J), chain A [356] and *Polytomella sp.* Pringsheim 198.80 (all structures from [357]). In other 190 structures of non-catalytic sites, a Gln residue in K-3 position links the O<sup>2A</sup> and O<sup>3G</sup> atoms, see Table F.1 below. However, in these eleven structures, Gln<sup>K-3</sup> does not reach the O<sup>2A</sup> atom so that the “Y-type” Arg residue of the adjoining monomer takes the canonical “finger position” and enters a Y-interaction with O<sup>2A</sup> and O<sup>3G</sup>. However, the high-resolution X-ray structures of the same non-catalytic sites reveal the amino group of the Gln<sup>K-3</sup> residue in the AG position, as discussed in the main text. We do not know why these eleven structures show a different interaction than the rest 190 structures of non-catalytic sites. Anyhow, both the Gln “plug” in 190 structures, as well as the Y-interacting Arg in eleven deviating structures are fully compatible with the major task of non-catalytic ATP binding sites, which is not to catalyze ATP hydrolysis. The remaining six EM-derived complexes with Y-type interaction come from structures of oligomeric complexes where other subunits display a typical stimulatory interaction via a single NH<sub>2</sub> group. Only one of these structures has a resolution better than 3.5 Å.
- II. We manually inspected the remaining 15 binding sites with Y-pattern (as obtained by X-ray crystallography of 13 crystal structures) and found out that they can be attributed at least to one of the following five cases (see details for each binding site in Table F.1):
  1. Presence of other complexes of the same protein with other, more common stimulatory patterns, with only one NH<sub>2</sub> group of an Arg finger interacting with both  $\alpha$ - and  $\gamma$ -phosphates or stimulators interacting with only  $\gamma$ -phosphate. Table F.1 contains five oligomeric structures where one protein subunit exhibits Y-like interaction of the Arg finger, whereas other subunits of the same protein have a different configuration of the catalytic site. No structures display two or more catalytic sites with a Y-like interaction for the

same oligomeric protein. For some of the proteins with Y-pattern in Table F.1, there are many other structures either from the same (ten complexes) or different organisms (six complexes) that display a different binding pattern. For instance, the Y-pattern is seen in one catalytic site of the rotary ATP-synthase (PDB 3OEE, chain L). This binding pattern is not displayed either in the other catalytic sites of the same structure, or in 69 other catalytic sites from  $\beta$ -subunits with either ATP or its analogs bound and an Arg finger present (see Table F.1).

2. Arg finger residue is listed as an outlier in wwPDB structure quality assessment reports.

In five cases, the Arg finger residues involved in Y-interactions were reported as outliers in regard to side chain geometry. Four of them are reported to possess a non-rotameric side chain, while one residue is a bond angle outlier (NE-CZ-NH<sub>2</sub> angle), see Table F.1. The reports also list too-close contacts; eight sites possess an Arg finger engaged in interatomic clashes either with other amino acid residues or with atoms of nucleotide moiety that are not expected to form H-bonds. In the structure of Guanine nucleotide-binding protein G (PDB 3FFA [358]) the side chain atoms of Arg178 are even clashing with neighboring atoms of the same residue.

Specifically, this set of cases contains the stimulatory Arg789 residue of H-RasGAP that is Y-linked with  $\alpha$ - and  $\gamma$ -phosphates in the first prototypical structure of the H-Ras/RasGAP complex (PDB ID 1WQ1 [64]), which has been widely used for MD and QM/MM modeling. The PDB X-ray Structure Validation Report for this structure (accessible at <https://www.rcsb.org/structure/1WQ1>) indicates that the conformation of the side chain of the Arg789 finger contains at least one outlier for two of the geometric quality criteria; it is noted also that Arg789 has a non-rotameric side chain conformation, which may point to a crystallization artefact.

Notably, in earlier reported structures of highly homologous G $\alpha$ -proteins the Arg finger interacted with  $\alpha$ - and  $\gamma$ -phosphates via a single NH<sub>2</sub> group (PDB ID 1GFI, 1GIL [62]). Also in the all subsequently obtained structures of Ras-like GTPases crystallized with TS analogs and cognate activators (with resolution < 2.0 Å, see, for instance, the high-resolution structures with PDB ID 1OW3 [70], 1TX4 [359], 3MSX, 5IRC [360]), only one NH<sub>2</sub> group interacts both with  $\alpha$ - and  $\gamma$ -phosphates. This prompts the suggestion that the Arg residue in the very first structure of the Ras/RasGAP complex (PDB ID 1WQ1 [4]) should have the same orientation. This suggestion is corroborated by MD simulations of the Ras/RasGAP which, after starting from the crystal

- structure, promptly yielded a conformation where a single NH<sub>2</sub> group interacted with  $\alpha$ - and  $\gamma$ - phosphates, see e.g. [361].
3. The distances between the NH<sub>2</sub> group and O<sub>2</sub>A, O<sub>3</sub>G atoms that are too short for an H-bond (O..H-N distance less than 2.4 Å) are observed in five complexes, see Table F.1.
  4. Electron densities (ED) are inconsistent with the positioning of the residue side chains. We have manually evaluated electron density maps (2Fo-Fc) for all 15 sites, see examples in Fig F.1A-F. Four Arg fingers lack ED entirely (example: Fig F.1A, F), while in one site the side chain is poorly fitted to the available density (Fig F.1E). In two cases electron density distribution is more consistent with the same NH<sub>2</sub> group bonded both to  $\alpha$ - and  $\gamma$ -phosphates (Fig F.1E, B). Six complexes display some electron density, but it is poorly resolved in the terminal region of Arg residue, thus not allowing determination of the exact location of the guanidinium group (examples: Fig F.1C, D).
  5. Optimized structures in PDB REDO depict a different configuration of the stimulator. We also checked the structures with Y-pattern in the PDB REDO structure databank, which contains automatically optimizes crystallographic structure models [362]. In three cases, the Y-pattern is absent from the re-refined structure, in two other sites the Arg is still inserted in a Y-like manner, however, the distance between the respective NH<sub>2</sub> group and the more distant phosphate group is shortened. It is worth noting that side chains lacking electron density cannot be expected to shift considerably in a re-refined structure.
  6. Finally, eight complexes with Y-like patterns belong to SF1/SF2 helicases (see Fig 1.2D for an example of a typical binding site in a SF1 helicase). Remarkably, most (105) other complexes of this class, including all the complexes with TS analogs bound, depict a single NH<sub>2</sub> group interacting with  $\alpha$ - and  $\gamma$ -phosphates, and in seven remaining cases Arg/Lys residue(s) contacts only  $\gamma$ -phosphate. Almost all these complexes harbor a second Arg residue contacting the  $\gamma$ -phosphate moiety. We would suggest that the catalytic site is not properly arranged in the absence of TS analog in these eight structures with Y pattern.

In sum, these findings indicate that a Y-like stimulatory pattern is unlikely to be inherent to P-loop fold NTPases; its presence in a few experimental structures might be due to poor resolution of particular residues or crystallization artifacts.

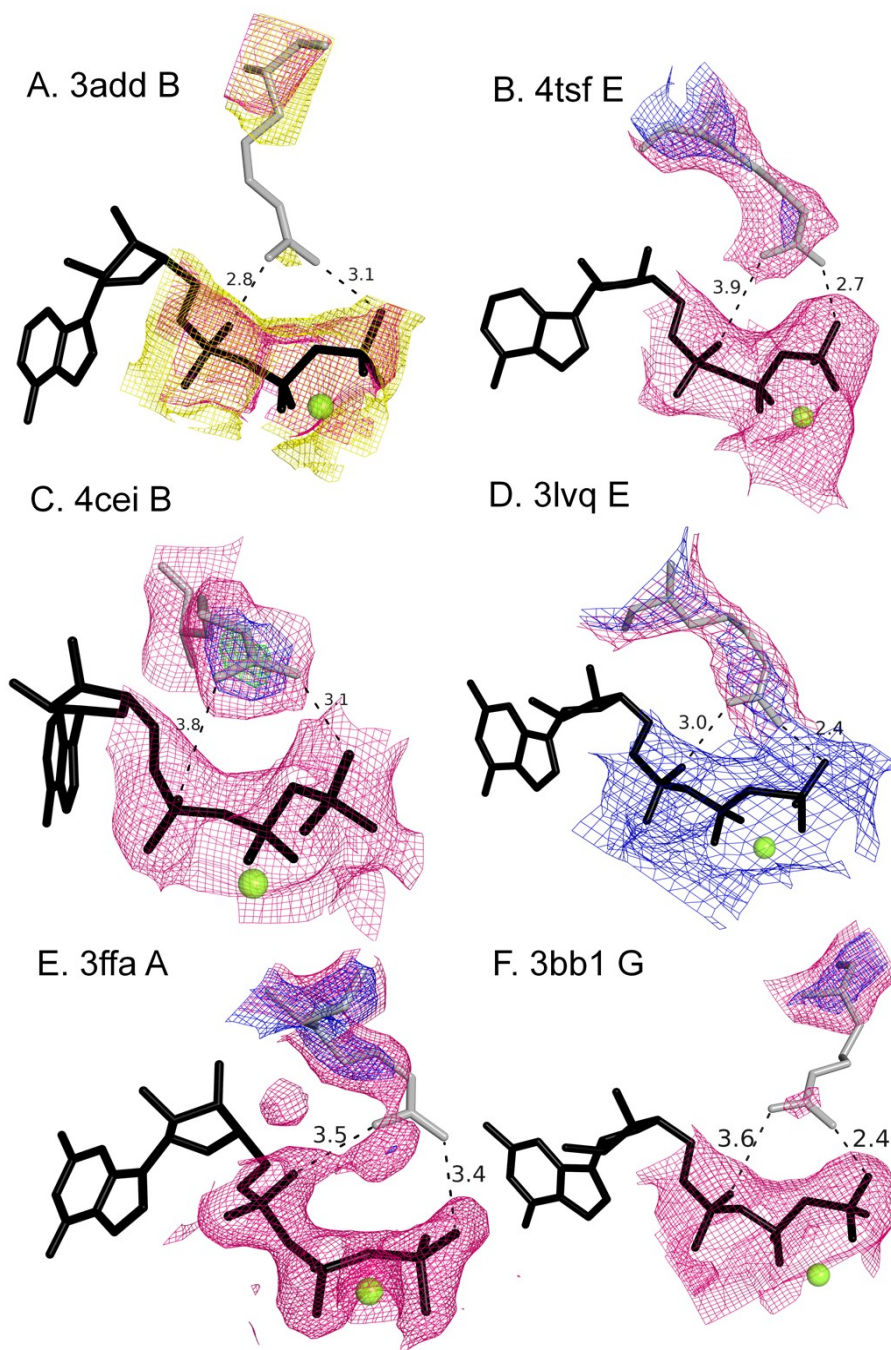


Figure F.1: 2Fo-Fc Electron density maps for Arg residues contacting  $\alpha$ - and  $\gamma$ -phosphate groups in a Y-like mode.

Nucleotides and their analogs are shown as black sticks, interacting Arg residues as gray sticks.  $Mg^{2+}$  ions are shown as green spheres. Density map is colored according to the contouring level ( $1\sigma$  in pink,  $2\sigma$  in blue,  $3\sigma$  in green,  $0.5\sigma$  in yellow). **A.** PDB 3ADD, chain B, Arg 116. Density contoured at  $0.5\sigma$  and  $1\sigma$  shown for phosphate chain and Arg residue. **B.** PDB 4TSF, Arg E356. Electron density shown for phosphate chain is contoured at  $1\sigma$  and at  $1\sigma$  and  $2\sigma$  for Arg residue. **C.** PDB 4CEI B 283. Electron density shown for phosphate chain is contoured at  $1\sigma$  and at  $1\sigma$ ,  $2\sigma$ ,  $3\sigma$  for Arg residue. **D.** PDB 3LVQ, Arg E469. Electron density shown for phosphate chain is contoured at  $2\sigma$  and at  $1\sigma$  and  $2\sigma$  for Arg residue. **E.** PDB 3FFA, Arg A178. Density shown as in (B). **F.** PDB 3BB1, Arg G133. Density shown as in (B).

Table F.1: Arginine residues contacting phosphate chain in a Y-like mode

P-LOOP CLASS	PROTEIN NAME	PDB ID	RESOLUTION, Å	NUCLEOTIDE	ARG FINGER		NUCLEOTIDE-BINDING POCKETS IN OTHER STRUCTURES OF THE SAME PROTEIN			DIFFERENCES IN OPTIMIZED STRUCTURE (PDB-REDO)	STRUCTURE QUALITY, wwPDB		ED EVALUATION	DISTANCES
					RESIDUE ID	ADDITIONAL ARG	OTHER SUBUNIT/COPY***	OTHER STRUCTURES, SAME ORGANISM***	OTHER STRUCTURES, DIFFERENT ORGANISMS***,#		SIDE CHAIN*	INTERATOMIC CLASHES **		
TRAFAC	Guanylate-binding protein 1	2BC9	2.8	GNP-A593	A48	-	-	NH2: 2B92 A; NH2 weak: 2B92 B.	-	-	NR	-	-	Ultra-short distance to O1G (2.3 Å)
Kinases	L-seryl-tRNA(Sec) kinase	3ADD	2.4	ANP-B2002	B116	-	NH2 weak: A	NH1 weak: 3ADB; NH2 weak: 3A4L, 3ADC AB.	-	-	-	3: T-N_SUG, T-N_bridgN, SideCh-AAB	No density	-
TRAFAC	Translocase of chloroplast 34	3BB1	2.8	GNP-H281	G133	-	NH2 weak: A,E; NH1 weak: C; ONLY GAMMA: B,D,E,H.	-	-	NH2: O2A -3.82Å, O1G- 2.84Å	-	-	Almost no density	Ultra-short distance to O1G (2.3 Å)
SF1_SF2	Eukaryotic initiation factor 4A-III	3EX7	2.3	ADP:AlF <sub>3</sub> -C414	C370	C367: NH2-F1 (2.63)	NH2 weak: H.	NH2: 2HYI CI; NH2 weak: 2J0Q AB, 2J0S A, 2XB2 AX.	-	-	-	-	-	-
TRAFAC	Guanine nucleotide-binding protein G	3FFA	2.3	GSP-A1	A178	-	-	NH1: 1AGR AD, 1GFI, 1SHZ AD, 1SVK, 3D7M; NH1 weak: 1BH2; NH2 weak: 2ZJY; ONLY GAMMA: 4N0D.	HUMAN: NH1 (2G83 A, 2GTP AB; 2IK8 AC); NH1 weak (2G83 B)	NH1: O1A 3.58, O3G 3.44; NH2 additional: O3G 3.27	NR	12: Internal(5), T-AAS(7)	Bad fit, by density: NH1 bonding α and gamma	-
TRAFAC	ADP-ribosylation factor 6 (fusion with Arf-GAP)	3LVQ	3.38	GDP:AlF <sub>3</sub> -E682	E469	-	-	NH2: 5JCP; NH2 weak: 3LVR.	MOUSE: NH2 weak: (3BH6, 3BH7)	AlF <sub>3</sub> floated away (before: O3B-Al: 2 Å, after: 2.42Å), both distances are longer	-	-	Uncertain group position	Ultra-short distance to F3 (2.36 Å)

P-LOOP CLASS	PROTEIN NAME	PDB ID	RESOLUTION, Å	NUCLEOTIDE	ARG FINGER		NUCLEOTIDE-BINDING POCKETS IN OTHER STRUCTURES OF THE SAME PROTEIN			DIFFERENCES IN OPTIMIZED STRUCTURE (PDB-REDO)	STRUCTURE QUALITY, wwPDB		ED EVALUATION	DISTANCES
					RESIDUE ID	ADDITIONAL ARG	OTHER SUBUNIT/COPY***	OTHER STRUCTURES, SAME ORGANISM***	OTHER STRUCTURES, DIFFERENT ORGANISMS***,#		SIDE CHAIN*	INTERATOMIC CLASHES **		
RecA_F1	ATP synthase subunit beta, mitochondrial	3OEE	2.74	ANP-M600	L375	M190 NH2-O3G (2.58)	<b>NH1:</b> X; <b>NH1 weak:</b> B,C,K,U375.	<b>NH1:</b> 2HLD BTU, 2XOK B, 3OE7 T, 3OE7 BCKLU, 3OFN L, 6B8H BW; <b>NH1 weak:</b> 2HLD CKL, 2XOK C, 3OE7 BCKLU, 3OFN BKT, 6B8H CX. <b>ONLY GAMMA:</b> 3OFN V, 3OE7 U.	<b>BOVIN:</b> <b>NH1</b> (1COW B, 1E1R C, 1H8E C, 1H8H B, 1NBM B, 1OHH AC, 1W0JC, 2JDI C, 2JJ2 I, 2XND C, 4YXW BC); <b>NH1 weak</b> (1BMF B, 1E1Q B, 1E1R B, 1EFR B, 1H8E B, 1OHH BC, 1W0J B, 2CK3 B, 2JDI B, 2JIZ BI, 2JJ1 BI, 2JJ2 B, 2XND B, 2WSS K); <b>PARDP:</b> <b>NH1 weak</b> (5DN6 C); <b>ONLY GAMMA</b> (5DN6 F)	both distances are longer	-	<b>4:</b> T-AAS, T-N_SUG(3)	Uncertain group position	-
SF1_SF2	ATP-dependent helicase/nuclease subunit A	4CEI	2.8	ANP-A2233	A479	A873 NH2-O3G (2.81)	-	<b>NH1 weak:</b> 4CEJ B.	-	-	NR	-	Uncertain group position	-
SF1_SF2	ATP-dependent helicase/deoxyribonuclease subunit B	4CEI	2.8	ANP-B2161	B283	-	-	-	-	-	NR	<b>1:</b> T-N_SUG	Uncertain group position	-
SF1_SF2	ATP-dependent helicase/nuclease	4CEJ	3	ANP-A2233	A479	A873 NH2-O3G (3.13)	-	-	-	<b>NH2:</b> O2A 3.70, O3G 3.26; <b>NH1</b>	-	<b>4:</b> T-N_bridgN (2), SideCh-AAS	Uncertain group position	-

P-LOOP CLASS	PROTEIN NAME	PDB ID	RESOLUTION, Å	NUCLEOTIDE	ARG FINGER		NUCLEOTIDE-BINDING POCKETS IN OTHER STRUCTURES OF THE SAME PROTEIN			DIFFERENCES IN OPTIMIZED STRUCTURE (PDB-REDO)	STRUCTURE QUALITY, wwPDB		ED EVALUATION	DISTANCES
					RESIDUE ID	ADDITIONAL ARG	OTHER SUBUNIT/COPY***	OTHER STRUCTURES, SAME ORGANISM***	OTHER STRUCTURES, DIFFERENT ORGANISMS***,#		SIDE CHAIN*	INTERATOMIC CLASHES **		
	subunit A									additional O3G 3.16		(2)		
RecA_F1	ATP synthase subunit alpha, mitochondrial  (see section I on EM structures)	4TSF	3.2	ATP-B600	E356	-	<b>ONLY GAMMA WEAK: F;</b> NONEweak: D.	<b>NH2 weak:</b> (4TT3, 4Z1M) <b>ONLY GAMMA</b> (2JJ2 EL, 2V7Q) <b>ONLY GAMMA WEAK</b> (1E1Q, 1E1R, 1NBM, 2CK3, 2JDI, 2JIZ EL, 2JJ1 EL, 2WSS EN, 2XND, 4TT3, 4YXW)	<b>ECOLI:</b> <b>NH2 weak</b> 3OAA cEMU <b>ONLY GAMMA WEAK</b> 3OAA bDdLT	<b>NH2 – O2A</b> shortened to 4.04 Å (previously 5.10 Å)	-		NH2 with no density, by density: same atom bonding both alpha and gamma	-
SF1_SF2	Zika virus NS3 helicase	5GJC	2.2	ATP-A702	A462	A459 NH2-O1G (3.52)	-	NH2 (5K8I, 5K8T, 5Y4Z)	<b>DEN4T:</b> <b>NH1 weak</b> (2LJR, 2LJV AB), <b>HCVCO:</b> <b>NH2</b> (5E4F B) <b>NH2 weak</b> (5E4F A,3KQL AB,3KQN,3KQU ABCDE) <b>9HEPC:</b> <b>NH2 weak:</b> (3O8D AB, 3O8R)	<b>NH1 – O1G</b> shortened to 3.81 Å from 4.41 Å	-	<b>1:</b> SideCh-N_SUG	Uncertain group position	Very short distance to O1G (2.39 Å)
SF1_SF2	Pre-mRNA-splicing factor ATP-dependent RNA helicase PRP43	5I8Q	4.2	ANP-A802	A430	A427 NH1-O1G (3.20)	-	-	-	-	-	<b>2:</b> T-N_bridgN, T-AAS	No density	-
SF1_SF2	Pre-mRNA-splicing	5I8Q	4.2	ANP-B802	B430	B427 NH1-O1G	-	-	-	-	-	<b>2:</b> T-N_bridgN,	No density	-



P-LOOP CLASS	PROTEIN NAME	PDB ID	RESOLUTION, Å	NUCLEOTIDE	ARG FINGER		NUCLEOTIDE-BINDING POCKETS IN OTHER STRUCTURES OF THE SAME PROTEIN			DIFFERENCES IN OPTIMIZED STRUCTURE (PDB-REDO)	STRUCTURE QUALITY, wwPDB		ED EVALUATION	DISTANCES
					RESIDUE ID	ADDITIONAL ARG	OTHER SUBUNIT/COPY***	OTHER STRUCTURES, SAME ORGANISM***	OTHER STRUCTURES, DIFFERENT ORGANISMS***,#		SIDE CHAIN*	INTERATOMIC CLASHES **		
	factor ATP-dependent RNA helicase PRP43					(2.91)						T-AAS		
SF1_SF2	Zika virus helicase	5Y6M	2	ADP:AIF <sub>3</sub> <sup>-</sup> A702	A462	A459 NH2-F3 (2.96)	-	NH2: 5K8I, 5K8T, 5Y4Z.	DEN4T: NH1 weak: 2LJR, 2LJV AB; HCVCO: NH2: 5E4F B; NH2 weak: 5E4F A, 3KQL AB, 3KQN, 3KQU ABCDE; 9HEPC: NH2 weak 3O8D AB, 3O8R.	-	bond angle outlier (NE-CZ-NH2)	-	-	Ultra-short distance to F3(2.22 Å); O3A is a bridging atom, shortest distance to non-bridging is 4Å

\*NR – non-rotameric

\*\* Types of interatomic clashes:

- Internal: contacts within Arg side chain

Clashes of terminal atoms:

- T-AAS: contacts of Arg guanidinium group with side chain of other amino acid residue
- T-N\_SUG: contacts of Arg guanidinium group with ribose group of nucleotide
- T-N\_bridgN: contact of one of the terminal amino groups with NH group replacing ester O3B atom in non-hydrolyzable analogs ANP and GNP

Clashes of atoms other than NH1, NH2 and corresponding H atoms:

- SideCh-AAB: contacts of Arg side chain with backbone atoms of other amino acid residue
- SideCh-AAS: contacts of Arg side chain with side chain of other amino acid residue
- SideCh-N\_SUG: contacts of Arg side chain with ribose group of nucleotide

\*\*\* only for corresponding Arg residues

# Organism names given as Uniprot organism mnemonics: BOVIN - *Bos taurus*, ECOLI - *Escherichia coli* (strain K12), PARDP - *Paracoccus denitrificans* (strain Pd 1222), DEN4T - *Dengue virus* type 4 (strain Thailand/0348/1991), HCVCO - *Hepatitis C virus genotype 1b* (isolate Con1), 9HEPC - *Hepacivirus C* (unspecified), MOUSE - *Mus musculus*.

## G. Evidence from IR spectroscopy

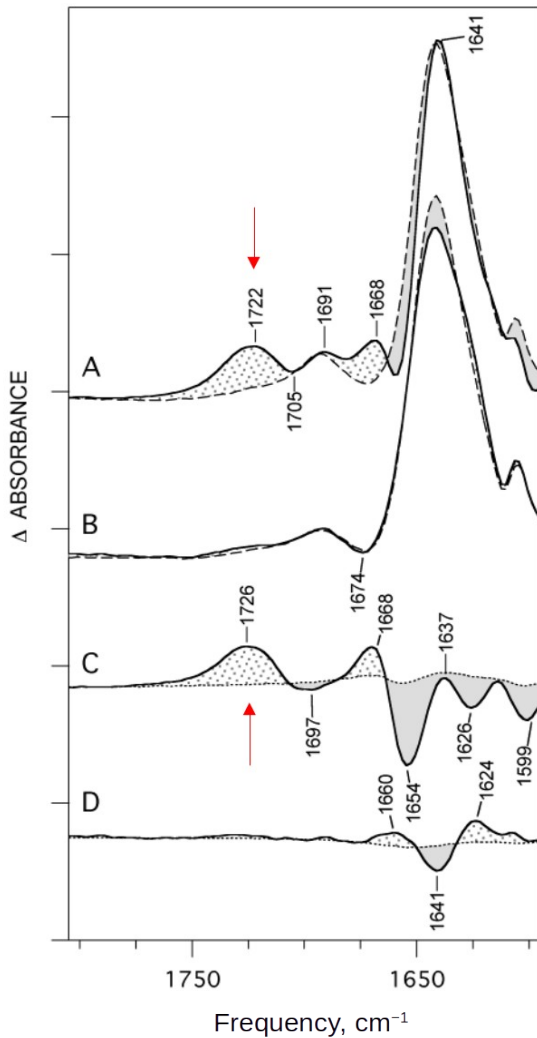


Figure G.1: Difference infrared spectra of caged nucleotide photolysis in the presence of monastrol and Eg5, modified from Fig. 3 in [320].

Shown is the 1800-1600  $\text{cm}^{-1}$  region. **A.** Caged ATP photolysis in the presence of Eg5. Monastrol was either preincubated with the sample (solid line), prior to photolysis, or absent (dashed line). **B.** Caged ADP photolysis in the presence of Eg5. Monastrol was either preincubated with the sample (solid line), prior to photolysis, or absent (dashed line). **C.** Eg5-monastrol difference spectrum upon photolytic cleavage of caged ATP. These data (s) is a double-difference spectrum of Eg-monastrol-caged ATP minus Eg5-caged ATP. The control double-difference spectrum is shown as a dotted line. **D.** Eg5-monastrol difference spectrum upon photolytic cleavage of caged ADP. These data (s) is a double-difference spectrum of Eg5-monastrol-caged ADP minus Eg5-caged ADP. The control double-difference spectrum is shown as a dotted line.

In PRC and BR shown in Fig. 4.3, the protonation of Asp/Glu residues was directly followed in the infrared (IR) spectral range [273,292–294,296,363,364]. In general, Asp and Glu residues are unique because the  $\nu(\text{C}=\text{O})$  vibration of their protonated carboxyl groups absorbs in the 1710-1760  $\text{cm}^{-1}$  spectral region that is free from overlap with absorption of other protein components [365]. Specifically, the protonated Asp-96 of BR has an absorption maximum at 1741  $\text{cm}^{-1}$  [294], whereas the protonated GluL212-AspL213-Ser223 complex of PRC absorbs at 1728-1725  $\text{cm}^{-1}$  [292,293].

Application of the IR-spectroscopy to P-loop NTPases is complicated by the transient nature of  $\text{Asp}^{\text{WB}}$  protonation – proton passes through  $\text{Asp}^{\text{WB}}$  on the time scale of the NTPase turnover. Nevertheless, IR measurements in this spectral range were performed by Kim and colleagues who studied human Eg5, a kinesin-like motor protein of TRAFAC class, where  $\text{Asp}265^{\text{WB}}$  and  $\text{Thr}112^{\text{K}+1}$  are connected by a short H-bond of 2.59 Å (PDB ID 3HQD [366]). Kinesin Eg5, as TRAFAC class ATPase, lacks a  $\text{W}_{\text{cat}}$ -coordinating catalytic carboxylic residue, which simplifies the attribution of the signal in the “carboxylic” spectral region.

In their steady state experiments [320], Kim and colleagues investigated the action of monastrol, an allosteric inhibitor that binds some 12 Å away from the catalytic site [367] and increases the reversals in this enzyme by hampering the release of  $\text{H}_2\text{PO}_4^{2-}$  [368]. In the presence of monastrol, but not in its absence, an absorption maximum at 1726-1722  $\text{cm}^{-1}$  was recorded and attributed to the protonation of a carboxylic group [320], see Fig. G.1. These data are consistent with proton trapping at  $\text{Asp}265^{\text{WB}}$  when the overall equilibrium of ATP hydrolysis shifts to

the left, as it happens in the presence of monastrol [368].

In kinetic experiments, the FTIR spectra of human Eg5 were monitored in real time and in response to the photorelease of caged ATP [321], see Fig. G.2. In the absence of microtubules, the hydrolysis by Eg5 proceeded at a time scale of seconds, which facilitated the IR measurements. A sharp absorption maximum at  $1743\text{ cm}^{-1}$  appeared at approx. 3 s and then decayed, whereas  $\text{H}_2\text{PO}_4^{2-}$ , as measured at  $1049\text{ cm}^{-1}$ , appeared at approximately 5 s and reached its maximum at 10 s (see Fig. G.2, modified from Fig. 1 in [321]).

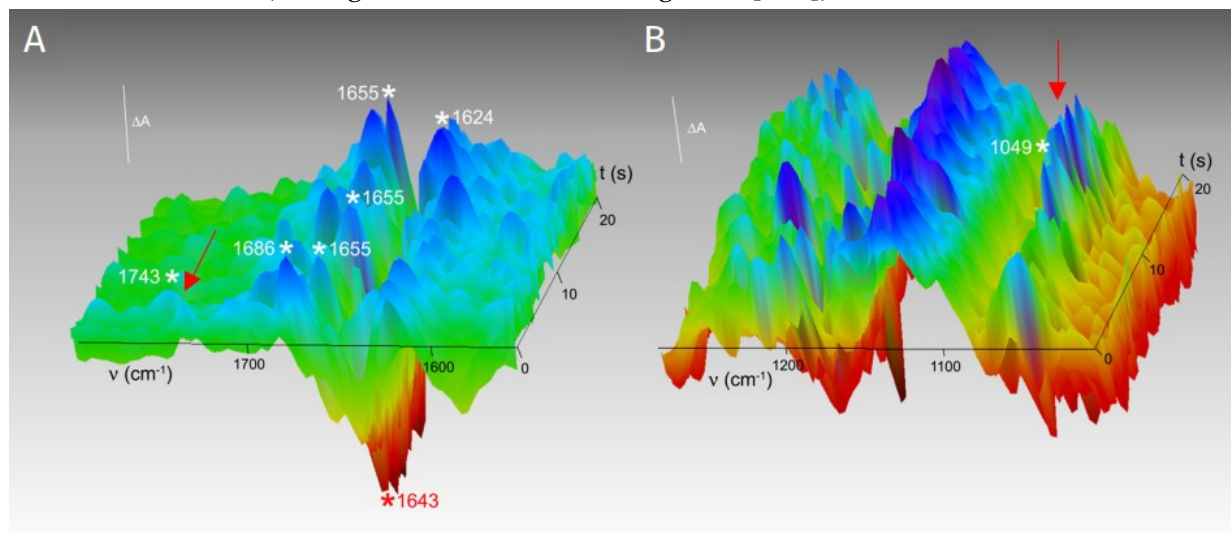


Figure G.2: Time-resolved IR spectroscopy of Eg5 catalysis in solution, modified from Fig. 1 in [321].

**A,B**, the amide I region (A) and the phosphate regions (B) of the time-resolved hydrolytic spectra. Alterations in absorbance are visualized by color, in which blue is the greatest amplitude and red is the lowest negative amplitude. The bar represents  $0.5 \times 10^{-4}$  absorbance units. All spectral data have  $12\text{-cm}^{-1}$  resolution.

It is tempting to suggest that the well-defined, sharp transient maximum at  $1743\text{ cm}^{-1}$  was due to the transient protonation of the carboxylic residue closest to the ATP molecule, namely Asp265<sup>WB</sup>, that forms a H-bond with Asp265<sup>WB</sup> next to Thr112<sup>K+1</sup> in human Eg5 (cf with the maximum at  $1741\text{ cm}^{-1}$  of the Thr46-bound, protonated Asp96 in BR [294]).

## H. Software availability

All software and scripts developed for this project are available at <https://github.com/servalli/pyploop>.

## 16. Anlagen

### LEBENS LAUF

#### Maria Kozlova

Geboren am 05.01.1995 in Moskau, Russland

#### SCHULBILDUNG

2001-2004	Schule Nr. 812, Moskau
2004-2007	Schule Nr.1329, Moskau
2007-2012	Staatliche Schule für Hochbegabte "Intellektual", Moskau <i>Schwerpunkte: Chemie und Biologie</i>

#### HOCHSCHULBILDUNG

01.09.2012-30.06.2018	Diplomstudium an der Fakultät für Bioengineering und Bioinformatik der Staatliche Lomonossow Universität Moskau <i>Hauptfach: Bioinformatik</i> <i>Nebenfach: Bioengineering</i>
21.06.2018	Diplom mit Auszeichnung, Thema der Diplomarbeit: "Halbautomatische Klassifikation der Bindungsmodi der Protein-Cofaktoren"
seit 01.10.2018	Promotionstudium, Graduiertenkolleg "EvoCell – Zelluläre Mechanismen der evolutionären Innovation", Universität Osnabrück (bei PD Dr. A.Y. Mulkidjanian)

#### BERUFLICHER WERDEGANG

01.10.2018-30.09.2021	Wissenschaftliche Mitarbeiterin (Entgeltgruppe 13 TV-L) am Fachbereich Biologie/Chemie der Universität Osnabrück.
01.10.2022-31.03.2023	Wissenschaftliche Hilfskraft am Fachbereich Physik der Universität Osnabrück

#### SPRACHKENNTNISSE

Russisch (Muttersprache)  
Englisch (fließend)  
Portugiesisch (fließend)  
Deutsch (Grundkenntnisse)  
Französisch (Grundkenntnisse)

#### PROGRAMMIERSPRACHENKENNTNISSE

Python (gute Kenntnisse)  
R (gute Kenntnisse)  
C++ (Grundkenntnisse)

Osnabrück, den \_\_\_\_\_

\_\_\_\_\_

## ERKLÄRUNG ÜBER DIE EIGENSTÄNDIGKEIT DER ERBRACHTEN WISSENSCHAFTLICHEN LEISTUNG

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Weitere Personen waren an der inhaltlichen materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

.....  
Osnabrück, Deutschland

.....  
Maria Kozlova