Tracing functional evolution in P-loop NTPases

A Thesis submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by Ranjana Nataraj

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

8 April, 2023

Under the guidance of

Supervisor: Dr.Gayathri Pananghat,

Indian Institute of Science Education and Research Pune

From May 2022 to March 2023

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Declaration

I hereby declare that the matter embodied in the report entitled **Tracing functional evolution in P-loop NTPases** are the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, under the supervision of **Dr. Gayathri Pananghat** and the same has not been submitted elsewhere for any other degree

Ranjana Nataraj

08.04.2023

Certificate

This is to certify that this dissertation entitled '**Tracing functional evolution in P-loop NTPases'** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents work carried out by **Ranjana Nataraj** at the Indian Institute of Science Education and Research Pune under the supervision of **Dr.Gayathri Pananghat**, Associate Professor, Department of Biology, during the academic year 2022-2023.

Gayathri

Dr. Gayathri Pananghat

Committee:

Dr. Gayathri Pananghat

Dr. Amrita Hazra

Acknowledgments

I would like to express my heartfelt gratitude to my thesis supervisor, Dr. Gayathri Pananghat, for her invaluable guidance and unwavering support throughout my project. My heartfelt thanks also extends to Dr. Amrita Hazra, who graciously served as my thesis advisor during the tenure of my project.

The constant support and assistance from members of both G3 and SK labs, particularly Sukanya Chakraborty and Puja Kumari, were instrumental in the successful completion of my project. I also wish to thank Dr. Jyoti Baranwal, Shekhar Jadhav, Sanket Shelke, and previous members of the lab who had made pivotal contributions to the initial conceptualization and execution of part of the analysis.

Lastly and most importantly, I am indebted to my best friend Anantha.S.Rao and my family for their unconditional love and encouragement that fuels me to be the best version of myself.

Table of Contents

List of figures

Abstract

The simplest members of the P-loop NTPase superfamily- the Ras family of GTPases act as molecular switches that regulate many cellular processes in all domains of life.
The GTPase domain in small Ras-like GTPases is functionally minimalistic and hence the completion of the GTPase cycle in these GTPases requires additional proteins - GTPase Activating Proteins (GAPs) and Guanine nucleotide Exchange factors (GEFs). The project reveals that the interface in Ras-GAP and Ras-GEF complexes most often involves the nucleotide-binding pocket. Hence, these interactors predominantly employ a direct mode of regulation of the Ras GTPase. The GAPs facilitate the hydrolysis of GTP to GDP in Ras by properly positioning two catalytic residues, while GEFs reduce Ras's affinity for GDP by transiently opening up the nucleotide-binding pocket. An examination of more complex P-loop NTPase members revealed that these P-loop NTPases avoid dependence on GAPs by either modifying the existing minimal Ras core or evolving insertions to carry the stimulatory residues. The strategy adopted seems to depend on the degree of relatedness to the Ras GTPase, as closely related members incorporate point mutations in the core while further distantly related members evolve insertions that carry the stimulatory residues. The analysis of adenine binding environments revealed that adenine could position itself atleast in four different ways in P-loop ATPases by altering the tilt of the adenine plane or the relative positioning of ribose and purine planes. In all four modes of adenine positioning, insertions majorly stabilize the adenine. Analyzing residues that stabilize the adenine revealed that there is prominent utilization of main chain and water-based interactions unlike guanine recognition in Ras GTPases with a strictly conserved side-chain based recognition. The results from the project offer insights into specific parts of the small Ras-like GTPase core that have been modified by evolutionarily related NTPases to possess gains in function like hydrolytic competency and altered substrate specificity.

Contributions

Contributor name	Contributor role
Dr. Gayathri Pananghat	Conceptualization Ideas
Dr. Gayathri Pananghat, Dr. Jyoti Baranwal, Ranjana Nataraj	Methodology
Ranjana Nataraj, Shekhar Jadhav, Sanket Shelke	Software
Dr. Gayathri Pananghat	Validation
Ranjana Nataraj	Formal analysis
Ranjana Nataraj	Investigation
	Resources
Ranjana Nataraj	Data Curation
Ranjana Nataraj	Writing - original draft preparation
Dr. Gayathri Pananghat	Writing - review and editing
Ranjana Nataraj	Visualization
Dr. Gayathri Pananghat	Supervision
Dr. Gayathri Pananghat	Project administration
	Funding acquisition

Chapter 1

Introduction

The utilization of nucleotides as energy intermediates, building blocks for nucleic acids, or regulatory signals is at the center of all fundamental processes in biochemistry. One of the most common nucleotide-binding protein folds is the P-loop NTPase fold and is found in proteins encoded by the genomes of nearly all cellular life forms. Members of this fold carry out critical functions within the cellular environment like powering muscle contraction, regulating cell growth and division, controlling chromosome segregation and catalyzing metabolic reactions that involve phosphate transfer through nucleotide triphosphate (NTP) hydrolysis.

The P-loop NTPase fold has a structural framework of α/β proteins with repeating α-β units (Kinoshita et al.,1999). The β strands form a central, mostly parallel β-sheet that is surrounded on both sides by α-helices. At a sequence level, the fold is characterized by two highly conserved sequence motifs: an N-terminal Walker A motif and a distal Walker B motif (Leipe et al., 2003; Kinoshita et al., 1999).

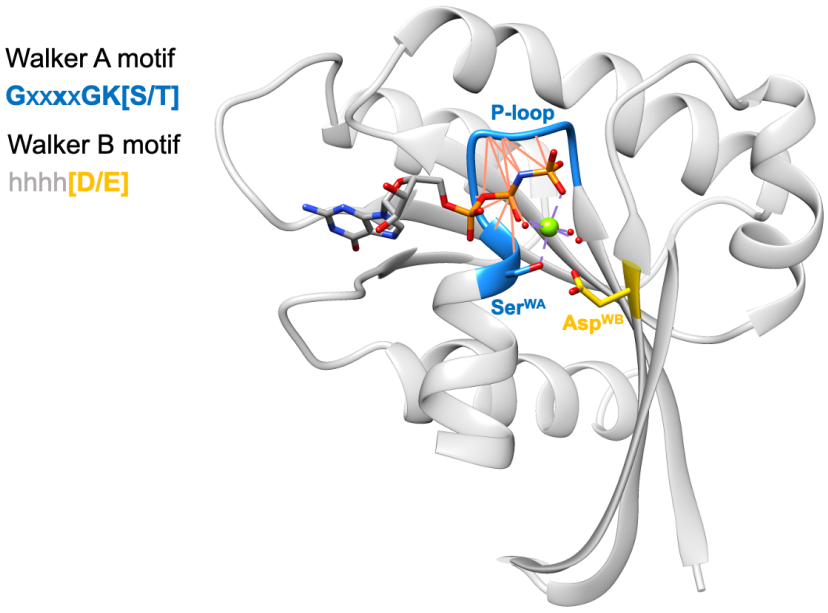


Fig. 1.1. Characteristic motifs in P-loop NTPases
The P-loop, Ser Walker A, and Aspartate Walker B motifs are the characteristic features of all P-loop NTPases. In this representative structure of a Ras-like GTPase, the Walker A motif is highlighted in blue, while the Walker B motif is highlighted in yellow. The coral lines represent the stabilization interactions between the phosphate moiety and the P-loop, a critical component of the Walker A motif. The Mg^{2+} ion, essential for NTP hydrolysis, is shown in lime, and its coordination sphere is depicted as purple dashed lines. Walker A Ser and Walker B Asp are highlighted.

1.1 Walker A and Walker B motifs in P-loop NTPases

The Walker A motif includes the Phosphate binding-loop (P-loop) with a sequence motif of (GxxxxG) that connects a beta-strand to an alpha-helix in the NTPase, and the first two residues of the helix (K[S/T]). In smaller P-loop NTPases, the P-loop connects the first beta-strand to the first alpha-helix. The P-loop adopts a characteristic rigid conformation in all P-loop NTPases that allows it to accommodate the triphosphate moiety of a bound nucleotide. The conserved Lys residue (K^{WA}) in the Walker A motif forms H-bonds with the O1B oxygen atom of β-phosphate and O2G atom of γ-phosphate. The Ser/Thr of Walker A ($[S/T]^{WA}$) plays a role in coordinating the Mg^{2+} ion, usually directly but in some cases indirectly through a water molecule.

The other signature motif in P-loop NTPase - the Walker B motif is located on the C-terminal end of the beta-strand, opposite the $[S/T]^{WA}$, and is characterized by the conserved hhhh[D/E], where "h" denotes a hydrophobic residue. The $[D/E]^{WB}$ residue in the Walker B motif makes an H-bond with the $[S/T]^{WA}$ residue and is involved in the indirect coordination of the Mg^{2+} through a water molecule.

Kozlova et al.'s recent comparative structural analysis suggests additional functional significance of the conserved Walker A and Walker B motifs, extending beyond their established roles in nucleotide and Mg^{2+} coordination. Specifically, they hypothesize that the Walker A - Walker B module functions as the distantly located base in P-loop NTPases, ensuring the sustained presence of the nucleophilic hydroxyl radical (OH^-). This model provides a rationale behind the strict conservation of the entire hhhhD Walker B motif, as the $[D/E]^{WB}$ terminal acceptor of the proton relay, along with the other

four hydrophobic residues of the motif acting as "hydrophobic protonic insulators." This strict conservation of the hydrophobic residues is necessary to increase the proton affinity of $[D/E]^{WB}$ within the narrow catalytic site, thereby preventing any unwanted proton escape from $[D/E]^{WB}$. Furthermore, in this model, the $[S/T]^{WA}$ plays a critical role as a catalytic nucleophilic alkoxide that causes the abstraction of a proton from the catalytic water (W_{cat}). The model explains the loss of catalytic activity in P loop NTPases with disruptions in the Walker A or Walker B motifs as a result of the nucleophilic OH^- not forming in these NTPases.

1.2 Members of P-loop NTPase superfamily

The P-loop NTPase superfamily can be divided into two main classes based on the arrangement of the P-loop and Walker B strands in their nucleotide-binding domains: the Kinase GTPase division (KG division) and the Additional Strand Catalytic Glutamate (ASCE) division (Leipe *et al.,* 2002).

Members of the KG division share the structural similarity of adjacent placement of the P-loop and Walker B strands, which form part of the core beta-sheet that interacts with the phosphate groups of the nucleotide. The KG division can be further subdivided into two classes based on distinct sequence and structural signatures (Leipe et al., 2002). The first class, known as TRAFAC (after translation factors), includes enzymes that play essential roles in translation, signal transduction (especially in the extended Ras-like family), cell motility, intracellular transport, and other processes. The second class, designated SIMIBI (after signal recognition particle, MinD, and BioD), includes GTPases involved in protein localization, chromosome partitioning, membrane transport, and metabolic enzymes with kinase or related phosphate transferase activity. Notably, the KG division contains most of the P-loop GTPases of the P-loop NTPase superfamily (Leipe *et al*., 2003)

The ASCE division includes a diverse set of proteins, such as ATPase Binding Cassette (ABC) transporters, DExD/H-like helicases, 4Fe-4S iron-sulfur cluster binding proteins

of NifH family, RecA-like F1-ATPases, and ATPases Associated with a wide variety of Activities (AAA). These proteins are crucial elements in multiple cellular events, including DNA replication and repair, RNA processing, protein degradation, and ion transport across cellular membranes. The ABC transporters, for instance, are responsible for the transport of a wide range of substrates across cellular membranes, while DExD/H-like helicases are involved in unwinding RNA duplexes during transcription and RNA processing. Members of the ASCE division are characterized by an additional strand in the core sheet, located between the P-loop strand and the Walker B strand (Leipe et al., 2003). The ASCE division contains mostly P-loop ATPases.

The evolutionary origins of P-loop NTPases, and how they have acquired their remarkable structural and functional diversity, remain an area of ongoing research. Although a complete understanding has not been achieved, some possible scenarios have been proposed based on sequence and structural analysis.

One such proposal comes from Aravind and co-workers where they compared the sequences and available structures of various P-loop GTPase representatives that are widely distributed, with the goal of constructing a structure-based phylogeny (Leipe *et al.,* 2003). Their evolutionary classification revealed that there are 57 subfamilies within the KG division, with 10 of them showing a universal phyletic distribution compatible with the presence of their ancestors in the last universal common ancestor of current life forms (LUCA). Among the ten subfamilies of the KG division, there are four translation factors, two OBG-like GTPases, two signal-recognition-associated GTPases, and one MRP subfamily of MinD-like ATPases. The majority of GTPases that can be traced back to LUCA have functions that are related to translation. This observation led them to hypothesize that the original function of the ancestral GTPase was to have a general regulatory role in translation. This common ancestor of all P-loop NTPases is thought to have evolved into the additional strand catalytic E (ASCE) division by acquiring an additional strand and a catalytic glutamate residue in the P-loop domain. The TRAFAC

and SIMIBI classes diverged from each other within the KG division by developing different features.

1.3 Small Ras-like GTPase – a minimalistic P-loop NTPase

Among the P-loop NTPases currently in existence, the small Ras-like GTPases which belong to the KG class have the simplest NTPase architecture. These GTPases have a single domain which serves as the GTPase domain (G-domain). The G-domain of the small Ras-like has a typical size of 25 KDa and is comprised of a six-stranded β-sheet and five α-helices, with conserved sequence motifs called G1, G2, G3, G4, and G5 responsible for binding guanine nucleotides (Milburn et.al, 1990).

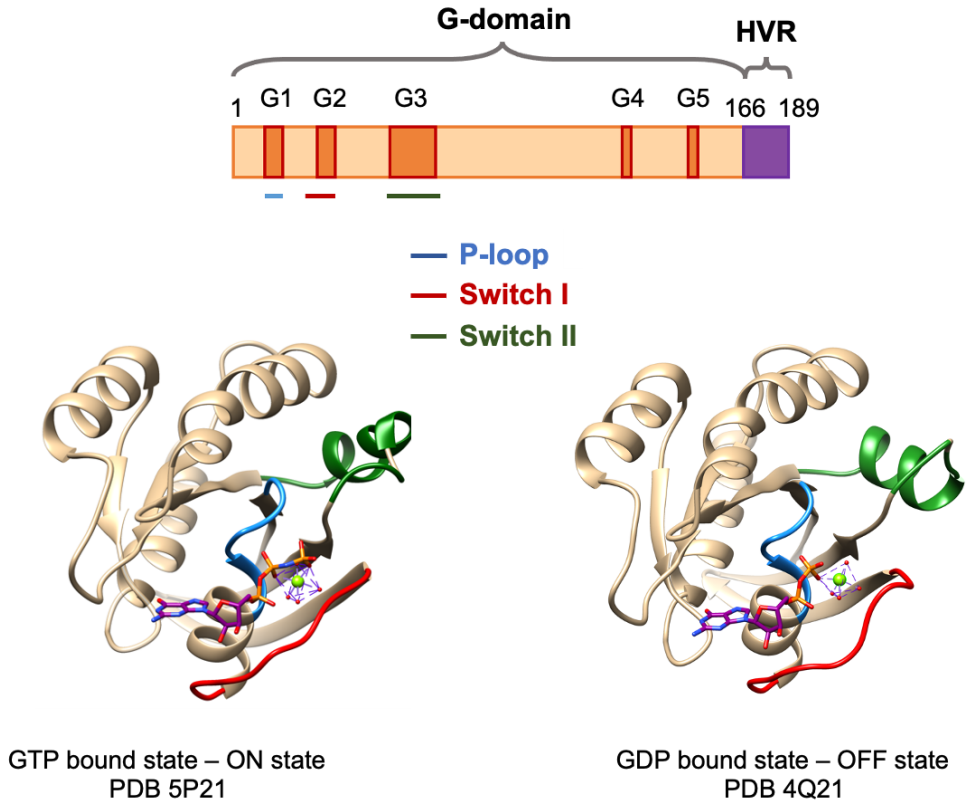


Fig. 1.2. Small Ras-like GTPase fold – the structurally minimal P loop NTPase
The G domain of the small Ras-like GTPase is depicted with the five signature motifs, G1-G5, highlighted to show the relative primary sequence arrangement. The Hyper Variable region (HVR) is also highlighted, as this part of the protein structure exhibits variability among members of this family. The bottom panel shows the GTP and GDP bound conformations of the small Ras-like GTPase, with the P-loop, Switch I,

and Switch II regions highlighted according to the color scheme used above. Notably, the Switch regions undergo pronounced conformational changes upon binding to GTP compared to GDP

Functionally, an NTPase is expected to bind NTP, hydrolyze it to NDP and inorganic phosphate (P_i), and release the NDP, allowing the next NTP to bind to the pocket and complete one NTPase cycle. However, the intrinsic GTP hydrolysis rate in small Ras-like GTPases is negligible, making the transition from the NTP to the NDP-bound form unlikely, unless additional activating proteins form a complex with the GTPase. The topology around the guanine in small Ras-like GTPases is such that it exhibits a much higher affinity for the nucleotide, requiring an explicit exchange factor to remove GDP post-hydrolysis.

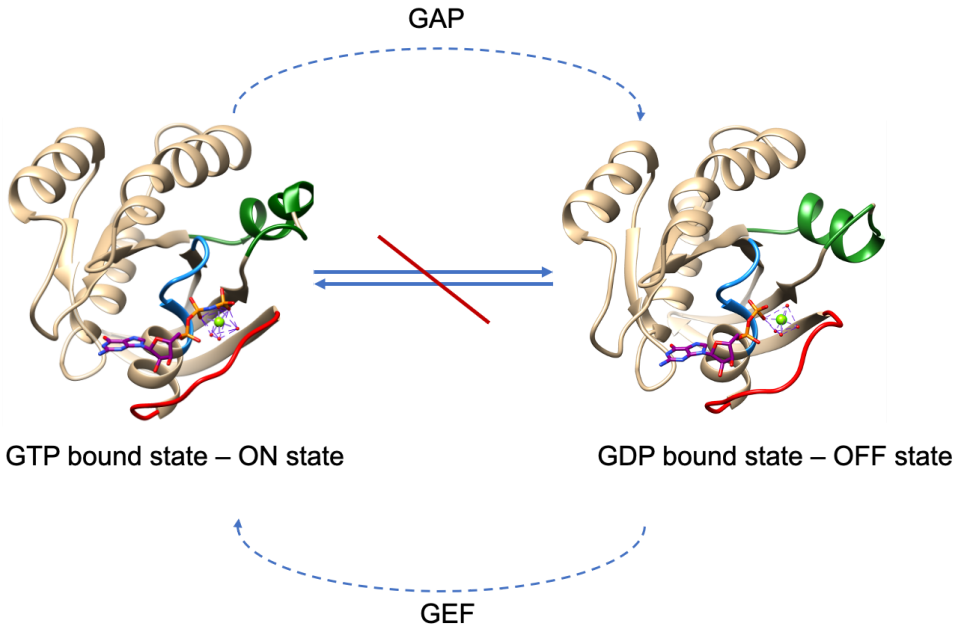


Fig. 1.3. Small Ras-like GTPase fold – the functionally minimal P loop NTPase
GAP - GTPase Activating Protein, GEF - Guanine nucleotide Exchange Factor

Thus, the P-loop NTPase core in small Ras-like GTPases incorporates only guanine binding motifs, in addition to the phosphate binding P-loop fold, without having complete hydrolytic machinery in place. In contrast, other P-loop NTPases have modified the NTPase core to show considerable intrinsic NTPase activity. Other P-loop GTPases exhibit much lower affinity for the nucleotide, suggesting alternate modes of stabilizing guanine. A vast majority of the P-loop NTPases are ATPases, which implies an entirely different purine-stabilizing environment.

It is possible to redefine the structure of P loop NTPase as the minimal NTPase core similar to that in small Ras-like GTPases with structural modifications that might contribute to its gains in function relative to small Ras-like GTPases. By tracing the structural origin of the additional functionalities, such as hydrolytic competency and changes in substrate specificity, we can identify the modifications that make the requirement of a GAP or GEF obsolete in these P-loop NTPases.

1.4 Objectives

This thesis attempts to trace the functional evolution in P-loop NTPases from the simple minimal Ras architecture by redefining each P-loop NTPase as a Ras-like GTPase core with insertions. The project has three specific objectives.

1. Unravel the unified basis of interaction in GAPs and GEFs to identify how these interactors influence the activity of the Ras GTPase at the level of the residues through a comprehensive structural analysis of available structures.

2. Extend the comparative structural analysis to identify the structural basis of hydrolytic competence of P-loop NTPases. Through this, we aim to identify structural modifications within the minimal NTPase that conferred the activity of GAPs, resulting in hydrolytic competence.

3. Examine the base recognition pocket in P-loop ATPases to trace the modifications within the Ras core that enabled a switch in substrate affinity from guanine to adenine nucleotides.

Chapter 2

Structural analysis of small Ras-like GTPases and their regulators

2.1 Introduction

The small Ras-like GTPase fold is adopted by six subfamilies - Arf, Rab, Ran, Rag, Ras, and Rho. Of these families, all except Rag GTPases are monomeric. Members of the Ras-like superfamily of small GTPases have been implicated in fundamental cellular processes like vesicular and membrane trafficking, and differentiation. In all these processes, the small Ras-like GTPases associate with interacting partners for linking extracellular cues to direct appropriate cellular responses and thus represent key signal-transducing molecular switches. These small Ras-like GTPases function as molecular switches because of two loops termed Switch I and II (Wittinghofer *et al.,* 2011). The Switch regions in these GTPases are loops near the nucleotide-binding pocket that show pronounced conformational changes in the GTP and GDP-bound states. Downstream effectors read these changes in the GTPase conformation and influence cell behavior. The GTP-bound state is considered the active state or ON state, while the GDP-bound conformation is the OFF state or inactive state. The small Ras-like GTPases require additional interacting partners - the GTPase activating Protein (GAP) and Guanine nucleotide Exchange Factors (GEF) to transition between the ON and OFF states. Despite the similarity in GTPase architecture across the small Ras-like GTPase subfamilies, the interactors of these GTPases show pronounced diversity in their global fold.

In this chapter, we will explore how the two major interactors - GAPs and GEFs of monomeric small Ras-like GTPases influence the GTPase activity. The emphasis will be on uncovering the common basis of these interactions at the GTPase interacting surface.

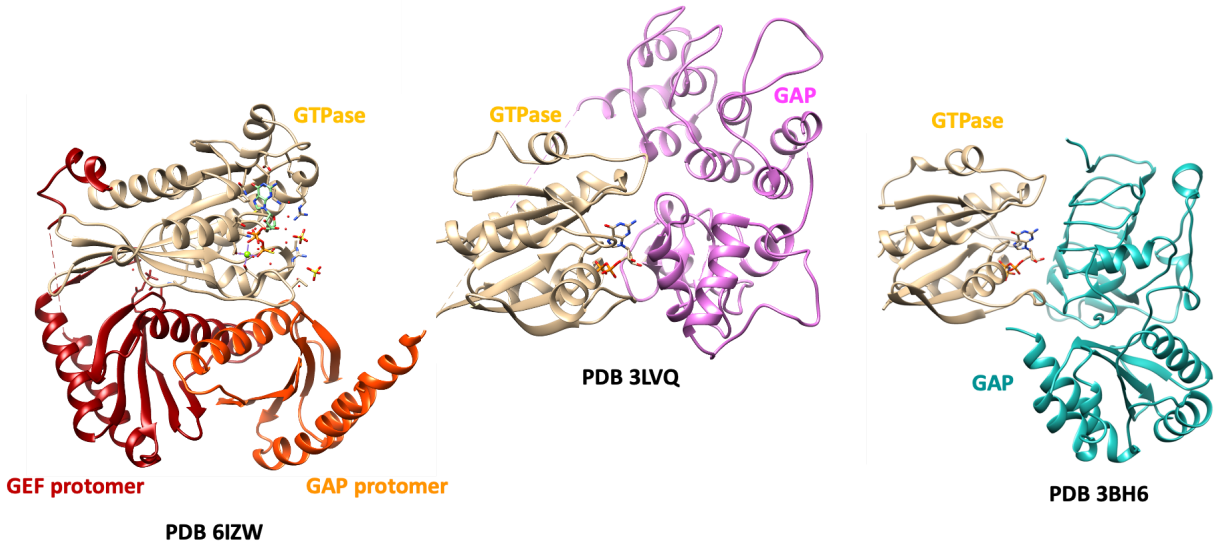


Fig. 2.1 Relative positioning of interactor hints at the nature of its influence on the GTPase
A. Prokaryotic small Ras-like GTPase MglA (tan) in complex with its bifunctional interacting partner - MglB. One protomer of MglB (shown in red) acts as a GEF, while the other protomer acts as a GAP. Since the GAP protomer positions near the nucleotide-binding pocket, it influences the GTPase activity of MglA in a direct mechanism. The GTPase-GEF interface excludes the nucleotide-binding pocket suggesting that the GEF activity of MglB is allosteric. **B, C.** The GTPase-GAP interface includes the nucleotide-binding pocket of the GTPase, and therefore these interactors likely influence the GTPase activity in a direct mechanism.

The interactor's mechanism heavily relies on its positioning relative to the GTPase. If at the interface of the complex, the interactor is located at the nucleotide-binding pocket of the GTPase, this indicates a direct mechanism (Fig 2.1.B; Fig 2.1.C). Conversely, if the GTPase-interactor interface of the complex excludes the nucleotide-binding pocket, this implies an allosteric influence from the interactor (Fig 2.1.A). Thus, the development of a framework that accurately captures information on how the interactor positions itself relative to the GTPase is critical in recognizing common modes of interactor activity despite variations in their global fold.

Face assignment of small Ras-like GTPases

A face assignment description of small Ras-like GTPases is an appropriate framework for the above analysis and also to extend this further to the functional evolution of P-loop NTPases. The interacting surface of monomeric small Ras-like GTPases (Arf, Rab, Ran, Ras, and Rho) is asymmetric and can be conceptualized as a pyramid with

four faces and a base. The secondary structures in small Ras-like GTPase can then be grouped into one of the five faces. The face assignment is done such that the main elements of the nucleotide-binding pocket – Switch I and II loops are contained within one face, which is Face 2.

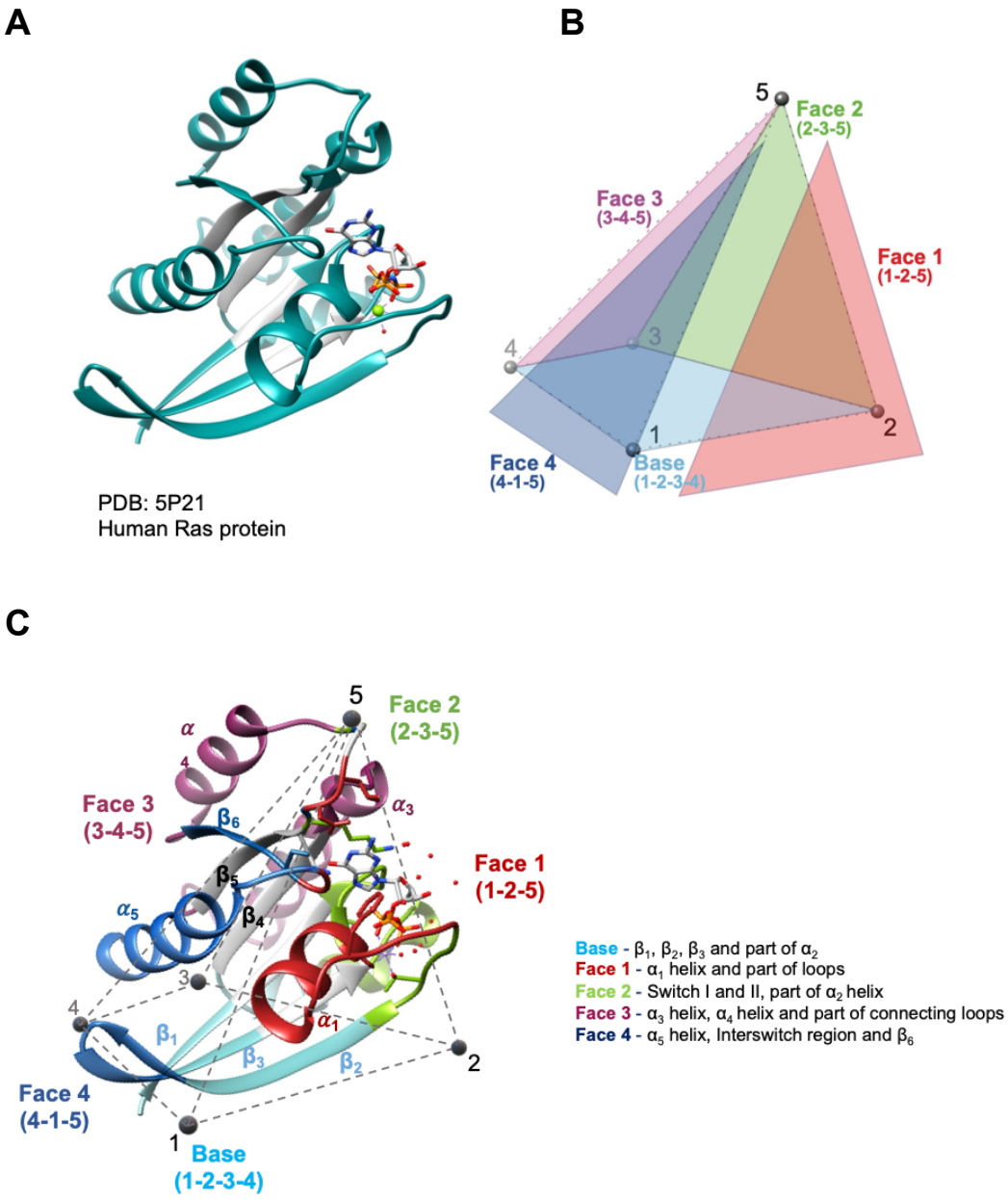


Fig. 2.2. Modelling the interacting surface of monomeric small Ras-like GTPases as an asymmetric pyramid

A. Interacting surface in small Ras-like GTPase highlighted in teal, the interior of the GTPase which is not accessible to interactors in a complex shown in light gray **B.** An asymmetric pyramid with 4 faces and a base. **C.** Dissecting the interacting surface of the GTPase into one of the 4 faces and base. The

breakdown of the interacting secondary structures in the small Ras-like GTPase into the faces is shown on right.

This concept and analysis based on the description were initiated as part of previous work carried out in the lab (Jyoti Baranwal, Ph.D. thesis, 2019). However, in this iteration, two main changes were made. Firstly, the face assignment of the GTPase secondary structures was refined so that the positioning of the interactor could be unambiguously described based on the contacts made to the faces. Specifically, in the earlier version, alpha helix 3 was assigned to Face 2, but in this project, it is considered a part of Face 3. A statistic that captures the overall positioning of the interactor at the surface of the GTPase (Face interaction frequency) was also developed in this project.

Additionally, computational scripts were developed in Python and incorporated into the analysis presented in this thesis. These scripts automate various parts of the analysis, including database curation, face assignment of the GTPases, and contact identification at the GTPase faces. This automation reduces the potential for errors and makes the pipeline scalable for future analyses as the number of protein structures deposited in the RCSB PDB server continues to increase over time.

2.2 Methods

Dataset generation

The following three approaches were attempted to obtain a comprehensive database of small Ras-like GTPase structures and complexes in PDB.

1. DALI server

Five PDB IDs were chosen as queries to probe structurally similar entries from the PDB database using the DALI server: 4JVS (Rho), 3LVR (Arf), 3MSX (Ras), 1RRP (Ran), 1XD2 (Rab). These PDBs were selected at random and do not hold any other significance. The DALI server returns a list of PDBs that are structurally similar for all the chains in the query structure. Since some of these queries were GTPases in complex with other proteins, the DALI results include PDBs that are structurally similar

to the non-Ras protein. Therefore, the results from the DALI search were passed through the PFAM server to get the family classification for each chain in a PDB. Only those PDBs that had atleast one chain that belonged to the small Ras-like GTPase family was retained. The family classification in PFAM for every PDB in the DALI output was retrieved using a webscrapping script written in Python. The BeautifulSoup module was used to access the HTML tag corresponding to the family information within the PFAM web result for each PDB. Once the PDBs that had atleast one Ras GTPase chain were identified, a resolution cut-off of 4 Å was also enforced.

2. Advanced sequence search from the PDB database

A PDB sequence-based search for each of the five representatives was adopted as a secondary method to identify the PDB IDs of structures belonging to the small Ras-like GTPase. This approach looks for PDB entries that show at least 50% sequence similarity to the query structure and have a resolution ≤ 4 Å. This relatively high sequence identity cut-off can be justified by the fact that proteins belonging to the small Ras-like GTPases show a high degree of sequence and structure similarity.

3. Family-based search in PFAM

A family-based search for Ras-like GTPases on the PFAM database was also conducted. The list retrieved was redundant as the different chains of a PDB ID are documented as separate hits for the Ras-like GTPases. After removing duplicate information and chain-based redundant entries, the list of unique entries belonging to the small Ras-like GTPase can be identified.

Visualization of contact maps

The contacts made at the interface in every Ras GTPase-interactor complex was found using PyMol (version 2.5.2). The secondary structures belonging to each face were defined as separate selections using the **sel** command. The contacts made by the interactor chain to each of these selections were found by looking for atomic contacts within 4 Å. The inbuilt “pairwisedistances” script in PyMol was modified to append the

contact information for multiple PDBs. This process of contact identification was automated using a custom script written in Python.

The results from contact identification at the interface of GTPase-interactor complexes were visualized in Cytoscape (version 3.9.1) as a contact map with nodes representing each of the five faces of the GTPase. Each Ras-GTPase complex was represented as a point, and the GTPase-interactor interface was mapped such that number of faces contacted by the interactor was indicated by the number of edges originating from the point. The number of atomic contacts made at the particular face of the GTPase was represented by the color of the edge connecting the point to the node corresponding to that particular face.

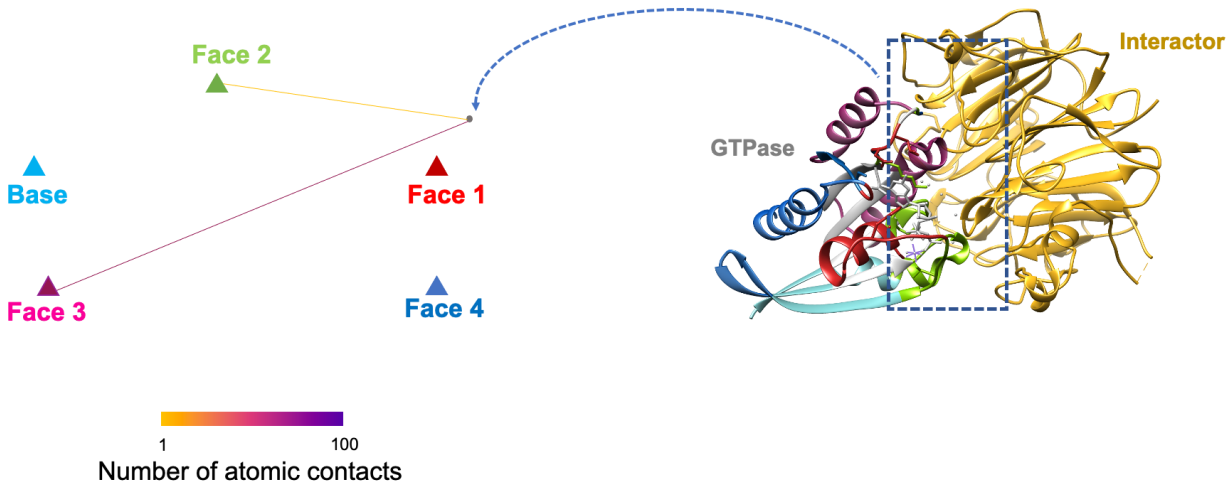


Fig. 2.3. Contact map captures the interface of a small Ras-like GTPase - interactor complex.
A complex between a Ras GTPase and an interactor shown in right. The secondary structures of the Ras GTPase are colored according to the faces to which they belong. The interactor's secondary structure is colored yellow. The interface of the complex is enclosed in a dotted box. The contact map, shown on the left, visualizes the atomic contacts within 4 Å made within the interface of the complex. The point in the map represents the complex, and the edges from the point indicate the faces of the GTPase the interactor contacts.

In the complex depicted in Fig 2.3, the interactor contacts Face 2 and Face 3 of the GTPase. The color of the edge reflects the degree of involvement of each face in the interface, with Face 3 being more heavily involved than Face 2 in this case. This is can be inferred from the edge color, with the number of atomic contacts made to Face 3

being closer to the violet end of the continuous color spectrum for the number of atomic edges.

The initial layout and representation of the plots were formulated by Shekhar Jadav, a former member of the lab.

Quantifying the contact map

The contact map generated captured the interface of the GTPase-interactor complex for each PDB in the dataset. To gain insight into the overall trend in interactor positioning it is useful to have a summary statistic indicating the percentage of contacts made to a specific face of the GTPase by an interactor. This number indicates the extent to which this particular face of the GTPase is contacted on average, when it is in complex with the interactor. The combination of faces that are most frequently involved in the interface describes the interactor's positioning relative to the GTPase. To calculate the interaction frequency (IF) at Face X by a particular interactor type for each PDB in the dataset, we use the following definition.

$$\mathbf{IF_x\ (PDB) = \frac{No:\ contacts}{(No:\ faces\ contacted)\ X\ (Fold\ population)}}$$

While developing this statistic for mapping the positioning of the interactor relative to the GTPase, the following possibilities of biasing were eliminated.

- Any two interactors of the Ras GTPase (in this example Interactor A and B) could be making the same number of contacts say 100 at a given face, X. Interactor A could however make comparable atomic contacts at other faces, that is act more diffuse when compared to interactor B that makes 100 contacts only at a specific face. To distinguish these cases of diffuse vs localized interactions, while determining IF_X we divide the number of atomic contacts made at Face X by the number of other faces the interactor contacts. This would reduce the weightage given to an interactor that makes contacts at multiple

faces. Incorporating this would convey that for interactor B the positioning is localized at Face X while for interactor A, despite similar number of atomic contacts made at Face X as interactor B, the positioning and influence is more diffuse.

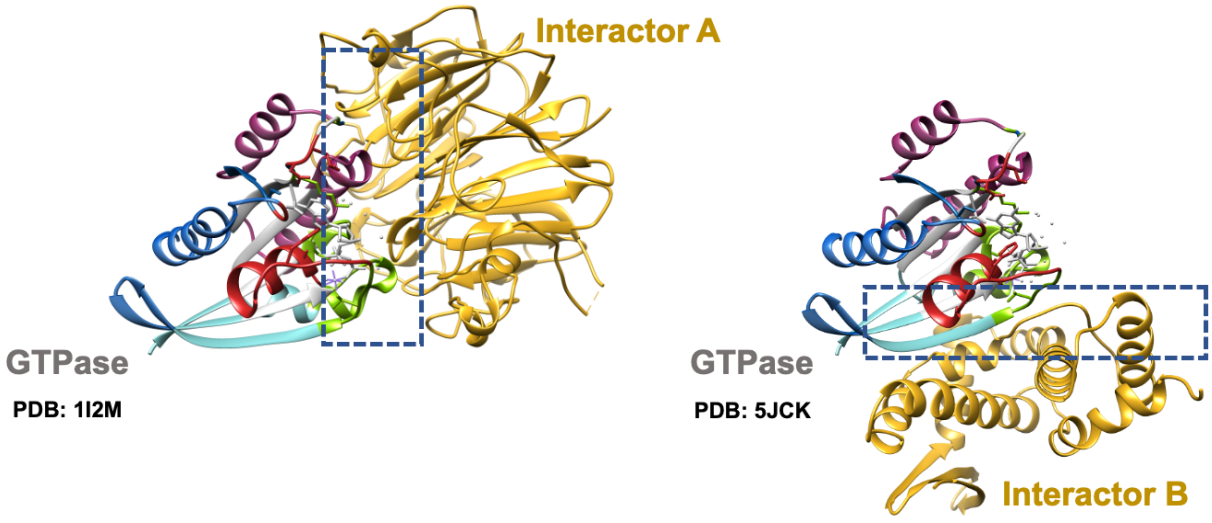


Fig. 2.4. Diffuse vs localized interactions to the GTPase
The secondary structures of the Ras GTPase are colored according to the faces to which they belong. The interactor's secondary structure is colored yellow. The interface of the complex is enclosed in a dotted box. Interactor A contacts 3 faces of the GTPase thus represents a diffuse interaction while interactor B contacts a single face and hence represents a localized interaction to the GTPase

- To prevent bias towards the interactor placement with the most representation, we divide by the number of interactors with the same fold. This ensures that folds with greater representation are weighed down compared to those with fewer representations.

Once the normalization for the size of the interactor and representation from each fold is made, the IF at a face for all members of a family that are bound to a particular interactor was found by:

$$IF_x\ (interactor) = \left[\frac{\sum IF_x\ (PDB)}{\text{Total no: contacts across faces}} \right] X\ 100$$

Dividing the number by the total number of contacts ensures that the number is between 0 and 1 and that the interaction frequencies for all faces add upto one. This is then multiplied by 100 to get the percentage.

The face interaction frequency for a given interactor was represented as a radar plot using the scatterpolar function in the Plotlyexpress module of Python.

Ras type	Base	Face 1	Face 2	Face 3	Face 4
Arf (4,2)	12.5	12.7	48.7	15.2	10.9
Rab (9,3)	41.5	5.2	40.9	3.5	8.9
Ran (57,2)	6.2	60.9	4.6	17.9	10.4
Ras (3,2)	19.3	5.7	38.9	31.4	4.7
Rho (14,2)	27.7	8.5	49.9	9.1	4.8

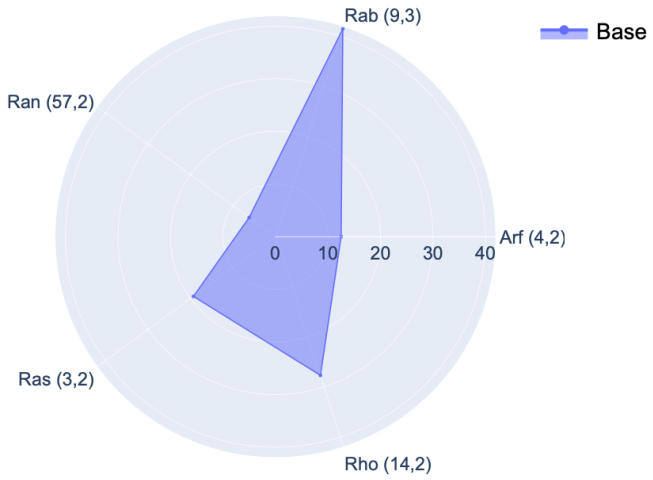


Fig. 2.5. Radar plot showing the interaction frequency for Base in Ras-GAP complexes
The frequency with which the GAP contacts each face of the GTPase is calculated using the face interaction frequency equation. The frequency with which GAP contacts base across GTPase subfamilies is plotted on right as a radar plot. The radial axis represents the percentage of contacts. The angular axis represents the family of the GTPase.

The radar plot conveys two main pieces of information on the general trend of interactor positioning in GTPase-interactor complexes:

1. **The dominance of a particular face to be in the interface can be understood by examining how far out the trace extends**

   The radar plot depicted in Fig 2.5 shows the interaction frequency at Base for all the small Ras-like GTPase - GAP complexes. An important attribute of this trace is its extension along the radial axis which denotes the face interaction frequency. Greater the extension of the trace at a specific GTPase family, the greater the extent to which the face is used up in the interface of the GTPase and interactor. Examining the radar plot of Fig 2.5, we observe that the trace for interaction frequency at Base extends out much more in the Rab family while the

trace does not extend as much for Arf and Ran. That is, in Rab-GAP complexes, the base is contacted much more frequently than the base in Arf and Ran GAP complexes.

2. **The preference for the particular face to be in the interface is shared across families or is unique to GTPases of a given subfamily**

   If the preference for a particular face is shared across families, the trace for that particular face will extend outward for multiple points along the angular axis. This would increase the area of the trace making it look broad. Conversely, if the usage of a particular face is unique to a family of GTPases, then this will result in a narrow trace or equivalently a trace with reduced area.

2.3 Results

2.3.1 Dataset statistics

Of the three approaches used, the results from DALI search contained the most number of unique PDBs of small Ras GTPases. Therefore, the analysis in this chapter was carried out using the dataset obtained from DALI.

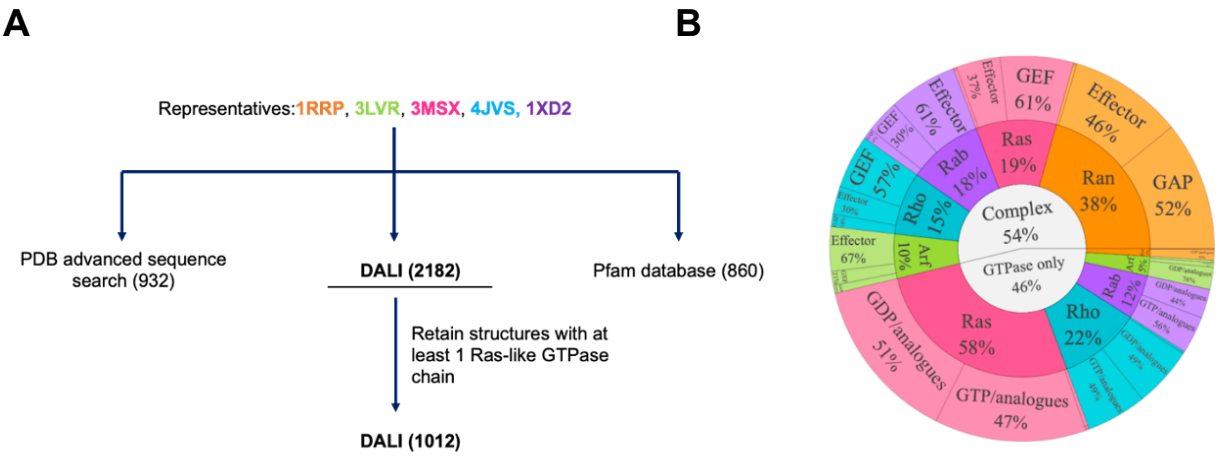


Fig. 2.6. Database for Ras-like GTPase analysis
A. Pipeline adopted to get the 1012 structures of small Ras-like GTPases. The color of PDB is the same as that of the family it represents. The number in brackets denotes the number of unique Ras GTPase containing PDBs at the end of the step **B.** Sunburst plot depicts the statistics of the 1012 PDBs studied in this analysis. The GTPases in complex with an interactor are classified based on the family of GTPase and the function of the interactor. Structures that contain only Ras-like GTPase are classified based on nucleotide bound.

From the dataset of 1012 unique PDB IDs retrieved from DALI, it was observed that 492 structures were uncomplexed GTPases, while 519 structures were bound to regulatory proteins. Among the uncomplexed GTPase structures, the Ras sub-family had the highest number of uncomplexed structures, totaling 307. Furthermore, it was observed that 95% of uncomplexed GTPases were in a nucleotide-bound state, reflecting their unstable nature in the absence of both nucleotide and interactors. Overall, only 16 structures were found to be captured in the transition state, highlighting the challenge of capturing GTPases in transition states without stabilizing interacting partners, typically GAPs.

Among GTPases in complex with other proteins, the Ran sub-family had the most structures in GAP complexes. The Rho sub-family dominated in GEF complexes with 68 structures. Ran family accounted for the majority of GTPase-effector complexes. Notably, about 97% of all Ras-like GTPase structures were bound to nucleotides, including various GTP analogs with distinct phosphates geometries.

The analysis presented in this chapter utilizes the dataset of 230 Ras GTPases in complex with either a GAP(88) or GEF (142). Table 2.1 shows the statistics on the global folds of the interactor to the subfamily and its representation in the database.

Table 2.1. Statistics of GTPase interactor folds

Ras sub-family	Fold	Count
Arf GAP	Fold 1	2
	Fold 2	2
Rab GAP	Fold 1	3
	Fold 2	3
	Fold 3	3
Ran GAP	Fold 1	3
	Fold 2	54
Ras GAP	Fold 1	2
	Fold 2	1
Rho GAP	Fold 1	1
	Fold 2	13

Ras sub-family	Fold	Count
Arf GEF	Fold 1	8
	Fold 2	1
Rab GEF	Fold 1	1
	Fold 2	2
	Fold 3	6
	Fold 4	6
	Fold 5	3
	Fold 6	4
Ran GEF	Fold 1	2
	Fold 2	1
Ras GEF	Fold 1	63
Rho GEF	Fold 1	6
	Fold 2	4
	Fold 3	32
	Fold 4	2

From Table 2.1, we can see that the small Ras-like GTPases can have interactors of multiple folds. There is greater fold diversity in GEFs that influence Rab and Rho while there is only one GEF fold that interacts with the Ras GTPase. Even in cases where more than one fold of interactor was documented, their representation in the dataset is not homogeneous. This is best illustrated by the Ran GAPs. In the dataset used for this analysis, GAPs with two different global folds influence Ran GTPase activity. One of these folds dominates the dataset as 54 out of the 57 Ran GAP complexes have the GAP of this particular fold. A similar observation can be made for Rho GEFs, where one

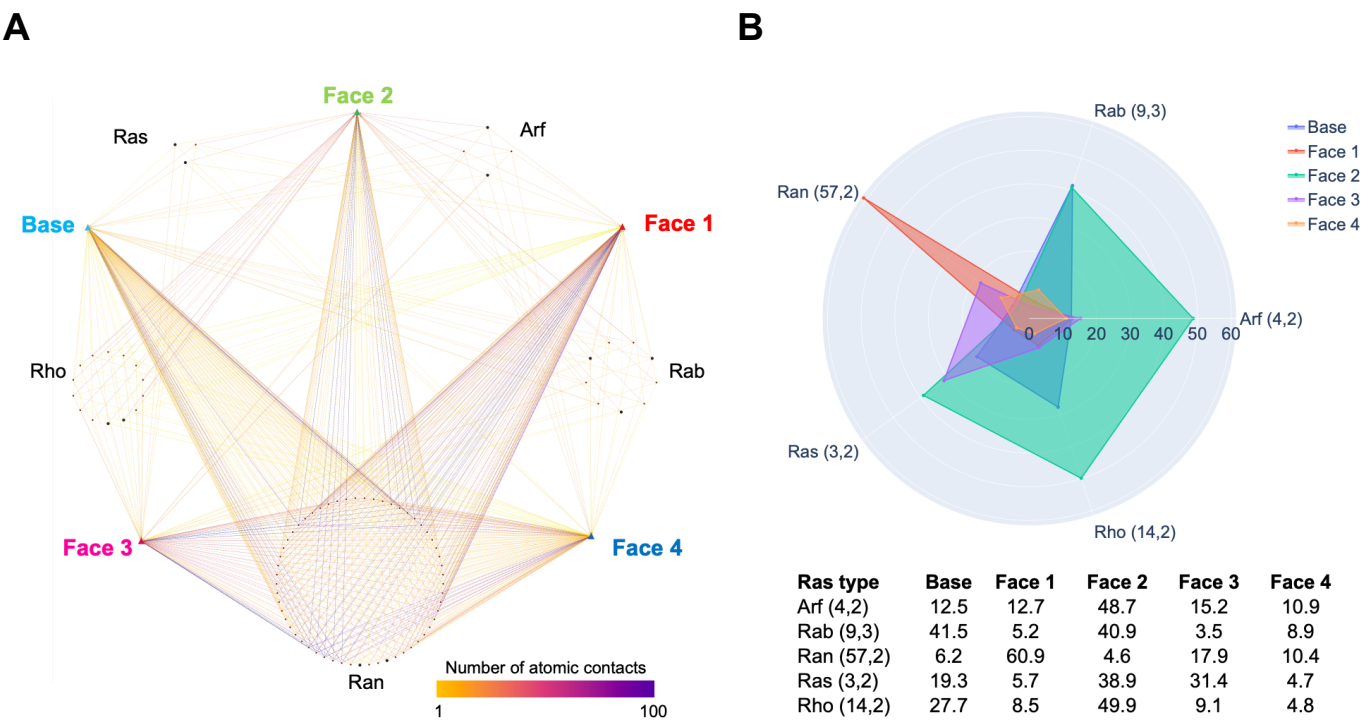
of the folds - Fold 3 is more represented in the database when compared to GEFs of other folds that influence Rho GTPase activity.

In the remaining sections, we aim to map the relative positioning of these interactors in GTPase-interactor complexes. Once the main interfaces are identified, the structural features at this interface will be explored to understand how these interactors influence the GTPase at these interfaces.

2.3.2 Structural and functional analysis of GTPase-GAP complexes

The contact map in Fig 2.7 reveals that Face 2 is the most frequently contacted in Ras GAP complexes across subfamilies. In the accompanying radar plot, we observe that the trace corresponding to interaction frequency at Face 2 in GTPase-GAP complexes is broad, which reveals that the preference for Face 2 is shared across subfamilies. The outward reach of this trace further reveals that the interface in GTPase-GAP complexes is dominated by contacts made to Face 2. In Rho and Arf, for instance, Face 2 accounts for up to 50% of all contacts made by GAPs to the GTPase. The faces that contribute in addition to Face 2 vary across subfamilies, with Rho GAPs mainly concentrating contacts at the Base, while Arf GAPs show more diffuse interactions with the remaining faces. Rab GAPs position themselves relative to the GTPase, much like Rho GAPs, but invest more in Base interactions, with other faces receiving fewer contacts. Ras GAPs, on the other hand, use Face 3 in addition to the dominant Face 2, with Base accounting for a smaller percentage of contacts.

The GAPs of Arf, Rab, Ras, and Rho seem to majorly contact Face 2 of the GTPase. Since Face 2 harbors the catalytically important regions - Switch I, II, the observation suggests that the GAPs in these subfamilies employ a direct mode of influence on GTPase activity.



Ras type	Base	Face 1	Face 2	Face 3	Face 4
Arf (4,2)	12.5	12.7	48.7	15.2	10.9
Rab (9,3)	41.5	5.2	40.9	3.5	8.9
Ran (57,2)	6.2	60.9	4.6	17.9	10.4
Ras (3,2)	19.3	5.7	38.9	31.4	4.7
Rho (14,2)	27.7	8.5	49.9	9.1	4.8

Fig. 2.7 Mapping contact interface in small Ras-like GTPase-GAP complexes
A Contact map for small Ras-like GTPase -GAP complexes created as explained in the methods section. The points are grouped on basis of the subfamily of GTPase. PDBs that have interactor with same global fold can be identified by tracing anti-clockwise from the dots with a bolder outline within each group having a GTPase of the same family. **B** The contact map in A is quantified using the face interaction frequency equation described in the methods section. The top figure visualizes the interaction frequency at each face across subfamilies of the GTPase as a radar plot. The numbers in brackets convey information on the total number of GAP complexes with this subfamily of GTPase, and the different types of GAP folds among these complexes are written after the comma. The table below the radar plot shows the exact face interaction frequency value for each face of the GTPase.

Interestingly, the dominant face of interaction in Ran GAPs is Face 1 of the GTPase. The narrow peak in the radar plot (Fig 2.7.B) reveals that this preference for contacts at Face 1 in GAP complexes is unique to Ran GAPs and is not shared by other subfamilies of small Ras-like GTPases. The lack of involvement of Face 2 suggests a different mechanism by which Ran GAPs activate the GTPase. The contact map (Fig 2.7.A) reveals inhomogeneity in the major face contacted within families, as some Ran GAP complexes show differences in the faces they contact. This variability may be due to the presence of other proteins, such as exportin and nuclear cargo, which can either block or make certain faces more accessible to the GAP.

Mechanism of small Ras-like GTPase GAP activity

This part of the analysis builds upon the work initiated by Dr. Jyoti Baranwal in the previous iteration of this project.

The results from contact map (Fig 2.7.A) reveal that there are two main interfaces in a small Ras-like GTPase - GAP complex. One involves the GAP positioning itself to make most contacts at Face 2 of the GTPase while using other faces variably. The other strategy does not involve Face 2, but rather involves Face 1 of the GTPase. The Face 2 based interaction is adopted by GAPs of Arf, Rab, and Ras, while Ran GAPs employ the non Face 2 based interaction.

Mechanistically, a GAP completes the nucleotide-binding pocket of small Ras-like GTPases, enabling it to hydrolyze GTP to GDP and inorganic phosphate. A complete active site in small Ras-like GTPases has a stimulatory finger, most often an arginine and residues that stabilize the catalytic water molecule which is usually a glutamine residue (Wittinghofer *et al.,* 2011). An examination of the GTPase-GAP complexes revealed that the particular faces contacted by the GAP depend on the specific role it plays in completing the Ras active site geometry. In cases where the stimulatory residue, typically Arg, comes from the GAP, it is necessary for the GAP to position itself near the nucleotide-binding pocket and make significant contacts at Face 2. This mechanism of action is observed in GAPs of Arf, Rab, Ras, and Rho, explaining why Face 2 of these GTPases is strongly contacted by the GAPs.

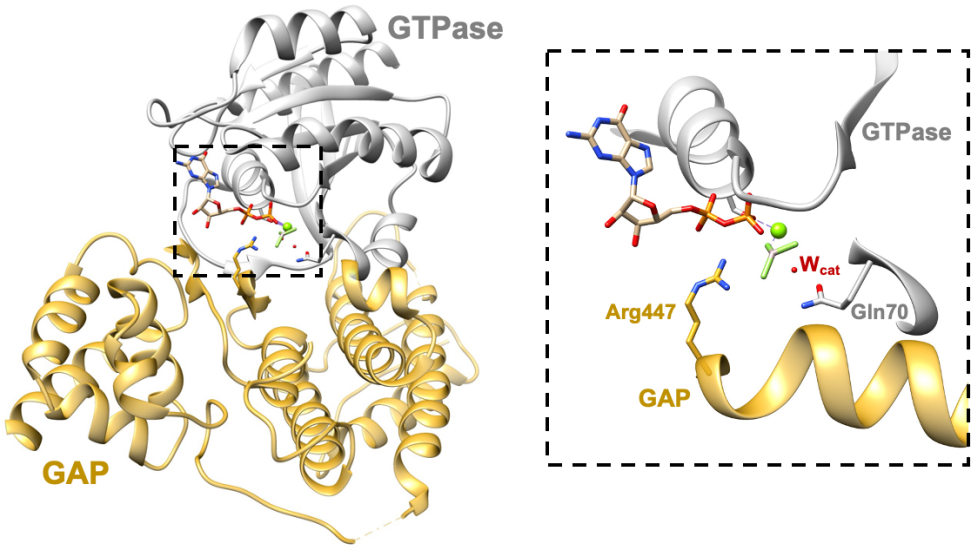


Fig 2.8. GAP activity involves completing the active site of small Ras-like GTPases
A. Left panel shows a small Ras-like GTPase in complex with GAP (PDB: 4JVS). The right panel shows a zoomed perspective of the active site region and shows that one of the stimulatory residues - Arg 447 is provided by the GAP while the W_{cat} stabilizing Gln70 is part of the small Ras-like GTPase core. The GAP positions at the nucleotide-binding pocket and hence at Face 2 since it has to provide one of the two catalytically critical residues.

In contrast, Ran GTPases have both catalytic residues, so GAPs activate the GTPase through conformational changes that optimally position the catalytic residues for hydrolysis. As a result, GAPs in Ran GTPases do not interact directly with the nucleotide-binding pocket, which explains why Face 2 is not the primary face of interaction in these GTPases.

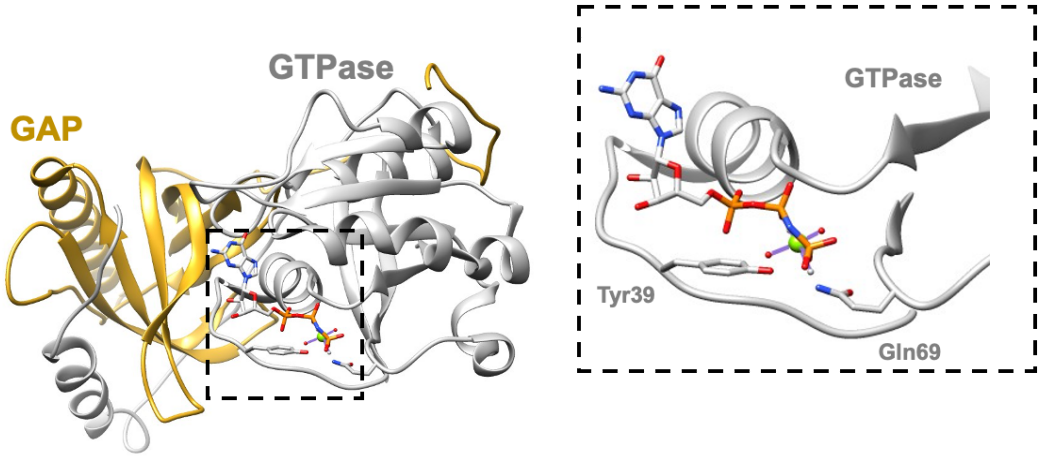


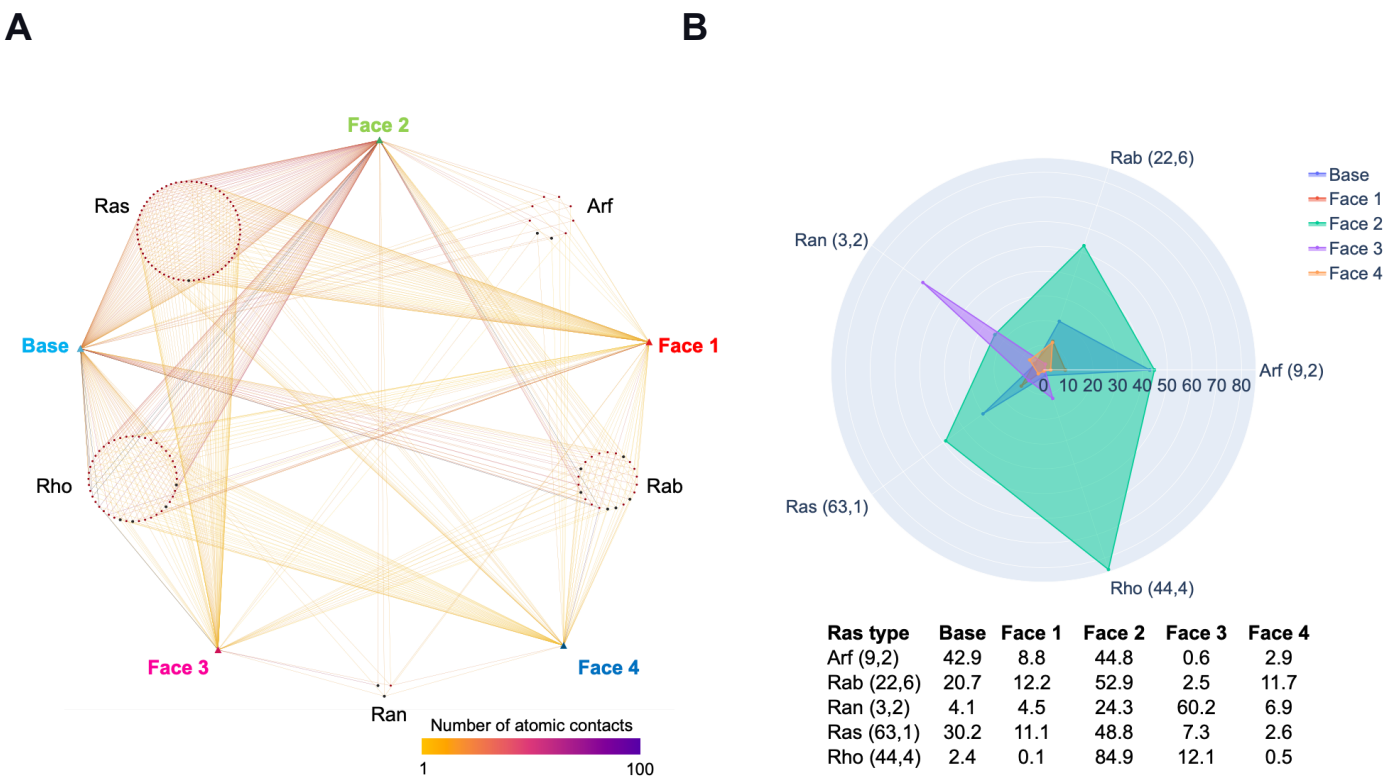
Fig. 2.9. GAP activity involves orienting the existing catalytic residues of the GTPase

A. Left panel shows a Ran GTPase in complex with GAP (PDB: 1K5G). The right panel shows a zoomed perspective of the active site region and shows that both of the stimulatory residues (Tyr39, Gln69) exist within the GTPase.

An inspection of the catalytic site geometry also revealed that irrespective of differences in the structural origin of the stimulatory finger, its positioning relative to the gamma phosphate is identical. That is, even when the stimulatory finger comes from the GAP or exists within the GTPase it positions itself in the same way with respect to the gamma phosphate.

2.3.5 Structural and functional analysis of GTPase-GEF complexes

The contact map and radar plot of Fig 2.10 reveal that the majority of atomic contacts in GTPase-GEF complexes happen at Face 2 of the GTPase. The dominance of Face 2 in the interface is most pronounced for Arf and Ras GEFs. While Rab GAPs displayed a more localized action as per section 2.3.2, the radar plot conveys that Rab GEFs interact more diffusely with Face 1, 2, and the base in addition to making dominant contacts at Face 2. Rho-GEF complexes, on the other hand, have a highly concentrated interface at Face 2, with about 85% of all contacts made to this face, indicating that Rho GEFs function at the Face 2-Face 3 combinations of the GTPase. Rho GAPs, however, position themselves at the Base-Face 2 junction. Despite variations in exact positioning, the GEFs of Arf, Rab, Ras, and Rho subfamilies are all similar in that they primarily interact with Face 2 to reduce the nucleotide affinity in GTPases by directly interacting with the regions of the nucleotide-binding pocket.



Ras type	Base	Face 1	Face 2	Face 3	Face 4
Arf (9,2)	42.9	8.8	44.8	0.6	2.9
Rab (22,6)	20.7	12.2	52.9	2.5	11.7
Ran (3,2)	4.1	4.5	24.3	60.2	6.9
Ras (63,1)	30.2	11.1	48.8	7.3	2.6
Rho (44,4)	2.4	0.1	84.9	12.1	0.5

Fig. 2.10 Mapping contact interface in small Ras-like GTPase-GEF complexes
A. Contact map for small Ras-like GTPase-GEF complexes created similar to the map generated for the GAP complexes **B.** The contact map in Figure A is quantified using the face interaction frequency equation similar to the GAP complexes.

Ran GEFs, however, are an exception to this trend, preferring other faces over Face 2. The radar plot (Fig. 2.10.B) clearly shows that Ran GEFs prefer Face 3 of the GTPase, with a Face 2-Face 3 combinations being the dominant point of interaction. This preference for Face 3 is unique to Ran GEFs as evident from the narrowess of the Face 3 trace. This is similar to Ras GEFs, where the dominant face is Face 2, but in Ran, it is Face 3. The emphasis on Face 3 of the GTPase over Face 2 suggests that Ran GEFs function to reduce nucleotide affinity in a different way than GEFs of the other subfamilies.

Mechanism of small Ras-like GTPase GEFs

An analysis of GTPase-GEF complexes, where the GEF primarily contacts Face 2 of the GTPase, has revealed three potential mechanisms. All three mechanisms involve the GEF acting on the phosphate end of the nucleotide-binding pocket.

The first mechanism involves direct occlusion of the Mg^{2+} binding site in Ras GTPases, which is essential for nucleotide binding in P-loop NTPases. In such cases, the GEF inserts a residue into the nucleotide-binding pocket and pushes out the Mg^{2+} from the pocket. The second mechanism involves transiently opening up the nucleotide-binding pocket by reorienting the switch loops, which contain the Thr residue that is one of the Mg^{2+} ligands. By interfering with Mg^{2+}coordination, this mechanism reduces nucleotide binding affinity. These two mechanisms are the most common modes of action for Ras-like GTPase GEFs. Both mechanisms interfere with nucleotide binding either by directly occluding the Mg^{2+} or reorienting its ligands.

In a few cases, the interaction of the GEF to Face 2 of the GTPase results in an altered positioning of the P-loop. The P-loop, or phosphate-binding loop, is crucial for stabilizing the phosphate moiety in Ras-like GTPases. The Lys residue that follows the P-loop stabilizes the beta phosphate group. Interfering with the P-loop-based stabilization of the beta phosphate could be another way in which GEFs reduce the affinity of the GTPase for nucleotide.

Alternate modes of GEF activity

The guanine nucleotides in Ras-like GTPases are stabilized by two crucial motifs – the SAK motif and the NKxD motif. These motifs provide stacking interactions and bidentate H bonds to the atoms of the Guanine and are responsible for the high specificity to Guanine. Examining the interface of the GTPase GEF complexes reveal that in cases where the GEF acts at faces other than Face 2 (Ran GEFs) they function by pulling away crucial residues within the NKxD and or SAK motifs away from the Guanine. The loss of stabilization from these residues results in reduced affinity for the guanine nucleotide.

2.4 Conclusion

- GAPs and GEFs of Arf, Rab, Ras and Rho position themselves at Face 2 of the GTPase and thus influence the GTPase activity directly. The interface in Ran-GAP and Ran-GEF complexes exclude Face 2 which suggested a different mechanism of interaction in these complexes.

- Face 2 interaction in small Ras-like GTPase - GAP complexes occur in instances where the GAP provides atleast one of the two catalytic residues required for GTP hydrolysis. Ran GAPs function by positioning catalytic residues that already exist within the GTPase and hence do not need to position themselves at Face of the GTPase.

- Small Ras-like GTPase GAPs function by ensuring the proper placement of two catalytic residues in the small Ras-like GTPase nucleotide-binding pocket. The placement of the stimulatory finger is conserved in small Ras-like GTPases relative to the gamma phosphate, irrespective of whether it originates from the GAP or exists within the GTPase. This mode of activity is similar across small Ras-like GTPase GAPs irrespective of variations in their global fold.

- Small Ras-like GTPase GEFs that contact Face 2 of the GTPase reduce the nucleotide affinity by destabilizing the nucleotide at the phosphate binding end. In the case of Ran GEFs, the GEFs reduce the nucleotide affinity by disrupting the guanine stabilization and do so by interacting at Face 3 of the GTPase.

Chapter 3

Structural description of P-loop NTPases based on Ras-like GTPase core

3.1 Introduction

The central NTPase fold in all P-loop NTPases is a conserved architecture consisting of the P-loop and the core beta-sheet sandwiched between alpha helices. While small Ras-like GTPases have minimal modifications to this core architecture, other P-loop NTPases exhibit additional domains that modify their functions. They can oligomerize, bind different substrates etc. The remaining chapters of this thesis aim to explore the structural modifications at the nucleotide-binding pocket of the P-loop NTPases that are responsible for hydrolytic competency and the ability to bind adenine nucleotides instead of guanine.

3.2 Methods

Dataset generation

The analysis in subsequent chapters examines the atomic environment of the catalytic site, and the presence of a reliable nucleotide binding is essential to ensure meaningful interpretations. Thus while obtaining the dataset for this analysis, we enforced the following quality check criteria that would evaluate the reliability of the catalytic site and the binding of the nucleotide within it.

The P-loop NTPase dataset was obtained through a series of steps outlined in the following flowchart.

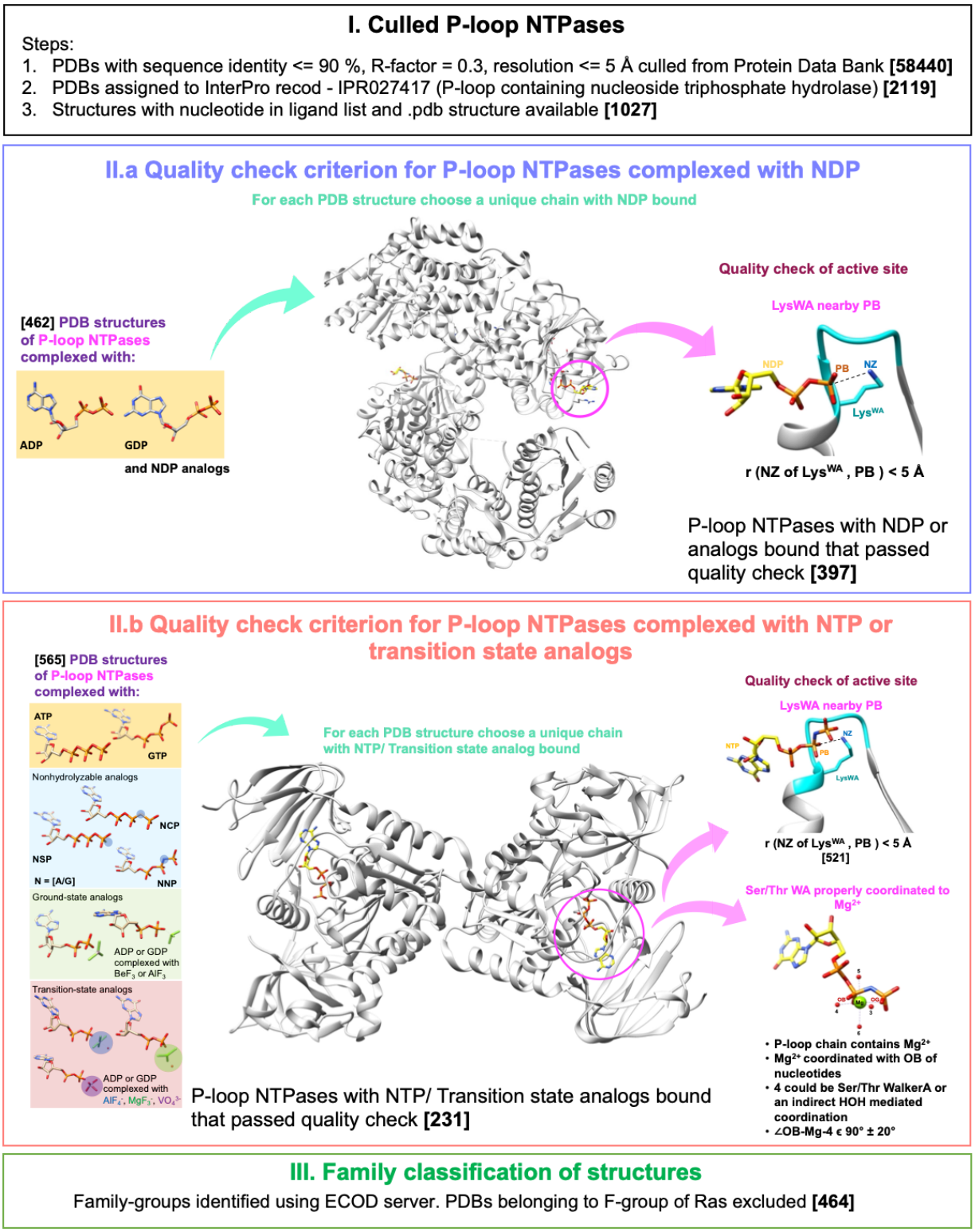


Fig. 3.1. Dataset generation for P-loop NTPase structural analysis

Flowchart showing the steps for shortlisting structures of P-loop NTPases based on Mg^{2+} coordination and Lys^{WA} binding to determine the integrity of the nucleotide-binding pocket. The number of structures at each step is given within square brackets.

The integrity of the nucleotide bound was analyzed by ensuring the presence of Lys^{WA} within 5 Å of the beta phosphate. If the Lys^{WA} was not within this distance, this would mean that the nucleotide is not bound properly by the Walker A motif making the structure unreliable for the analysis. An additional quality check criterion done for NTP/TS bound structures was to ensure that hexa-coordination geometry of Mg^{2+} is proper in the NTPase. In P-loop NTPases, Mg^{2+} is coordinated by O2B (and O2G in NTP bound structures), the conserved $[Ser/Thr]^{WA}$, while the remaining ligand positions are variable (Kanade et.al, 2020). To check the octahedral geometry of the Mg coordination sphere, the angle between OB, Mg, and $[Ser/Thr]^{WA}$ was calculated. If the angle falls within 90 ± 20°, it indicates that the catalytic site is not distorted. However, for NDP bound structures, we have only done the Lys^{WA} check, as Mg^{2+} is not necessary for binding NDP, for example, prokaryotic small Ras-like GTPase MglA is found bound to GDP even when not complexed with Mg^{2+}.

Core – Insertion mapping of P-loop NTPases

To distinguish the minimal functioning core architecture from the insertions that confer gain of function, the P-loop NTPase was superimposed with the Human small Ras-like GTPase (PDB: 5P21) along the P-loop, and a structure-based sequence alignment was generated using UCSF Chimera. A regular sequence-based alignment would fail to capture the structural conservation of the NTPase core since the overall sequence identity among these proteins is very low. The structure-based sequence alignment generated from Chimera was passed onto a custom Python class that added secondary structure information to each residue in the alignment.

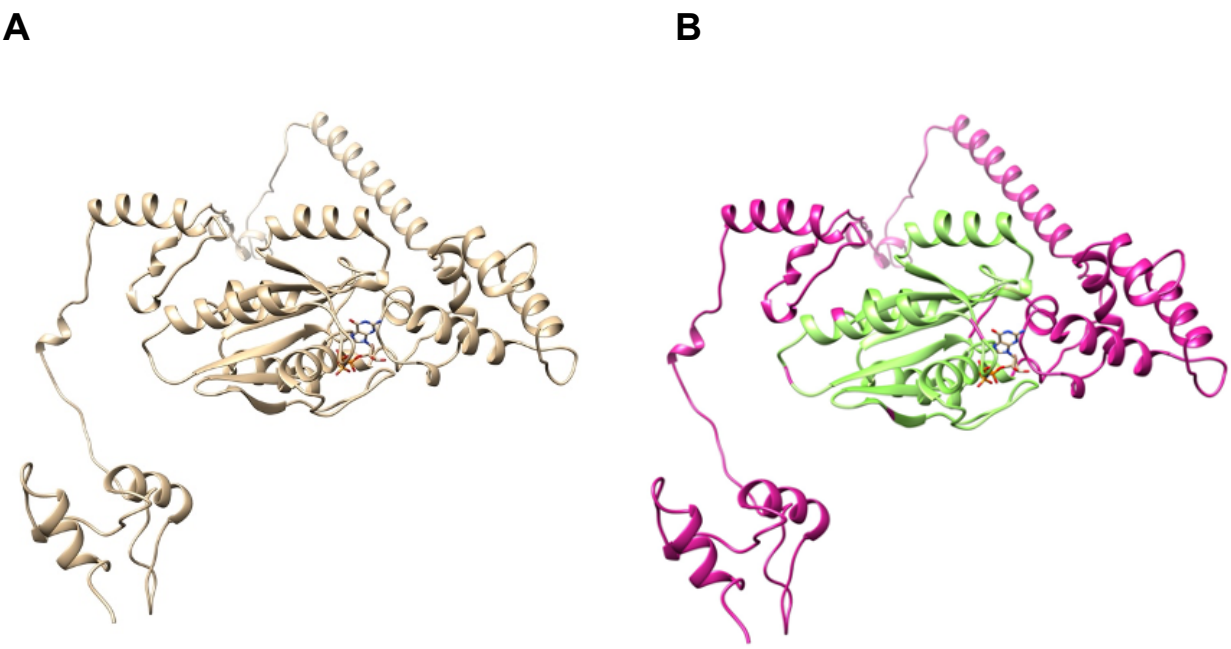


Fig. 3.2. Core-insertion mapping in P-loop NTPases
A. The secondary structure of Nog2 P-loop NTPase is depicted in tan, with the nucleotide bound in the active site. B. Secondary structures in Nog2 are color-coded based on the secondary structure-based sequence alignment with HRas (PDB:5P21). The residues that have an equivalent in 5P21 are highlighted in green, while those that have no equivalent are marked in magenta. The green-colored residues are considered part of the minimal NTPase core, while the magenta-colored ones are regarded as insertions. Instances of magenta within an otherwise green-colored secondary structure represent point insertions within the core.

By analyzing the output, the core of the P-loop NTPase was defined as those residues that have an equivalent in 5P21. The residue in Ras may be the same or a different residue (in which case it would mean that the P-loop NTPase made a point mutation at that point). If a residue from P-loop NTPase has no equivalent in Ras then this would be an insertion. This insertion residue could be a part of a secondary structure that otherwise aligns with the core (point insertion within a core structure) or part of a secondary structure that has no equivalent in small Ras-like GTPase. This framework enabled the identification of the minimal NTPase core in each P-loop NTPase while assigning the remainder of the protein structure as structural modifications.

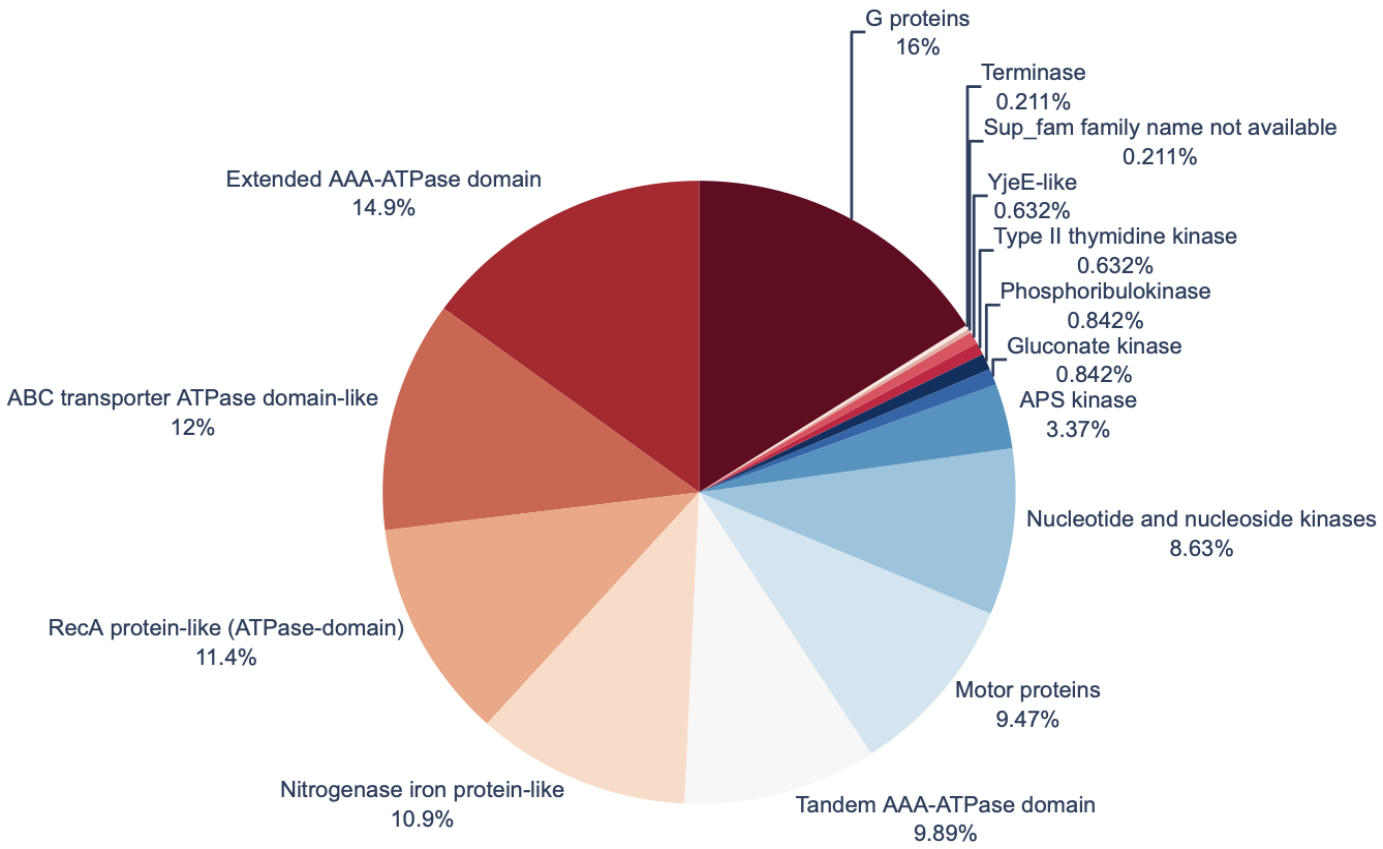
The Python script for core-insertion mapping was initially developed by Sanket Shelke, a former member of the lab. In this analysis, the script was modified to reduce the number of inputs the user had to provide and restructured to improve execution time.

3.3 Results

Dataset statistics

Fig 3.4 represents the entire dataset that contains P-loop NTPases with reliable binding of nucleotide in their active site. The family distribution reveals a mostly uniform representation of all major families of P-loop NTPases. There is not one single family that dominates the dataset obtained in this section of the thesis.

A



B

Family	Count
G proteins	76
Motor proteins	45
Phosphoribulokinase	4
Nucleotide and nucleoside kinases	41
Gluconate kinase	4
APS kinase	16
Type II thymidine kinase	3
Nitrogenase iron protein-like	52
Terminase	1
YjeE-like	3
ABC transporter ATPase domain-like	57
Extended AAA-ATPase domain	71
Tandem AAA-ATPase domain	47
RecA protein-like (ATPase-domain)	54
Sup_fam family name not available	1

Total PDBs = 464

Fig. 3.3. Family-wise classification of P loop NTPase dataset
Of the 464 PDBs of P-loop NTPases that passed the quality check criterion, the SUPFAM-based family classification is depicted as fractions in a pie chart. The number of PDBs belonging to each family is shown in the table below.

The specific question addressed in each chapter will further impose quality restraints on the dataset that will be used.

Chapter 4

Examining hydrolytic competence in P-loop NTPases

4.1 Introduction

P-loop NTPases are a class of enzymes that catalyze the hydrolysis of NTP to NDP and inorganic phosphate. Despite their ubiquity in cellular processes, the mechanism by which these enzymes use water to remove the terminal phosphate of NTP has remained unclear. A prerequisite for nucleotide hydrolysis is the generation of a potent nucleophile, which in the case of P-loop NTPases, is the hydroxide ion (OH^-) derived from water. However, to maintain the stability of the highly reactive OH^- ion, which has a strong affinity for protonation, it is essential to abstract a proton from the surrounding environment. The consensus in the field was that a glutamine residue in the immediate vicinity of the active site could act as a proton acceptor and serve as the base in NTP hydrolysis. However, not all P-loop NTPases possess such a glutamine residue, and in such cases, it was postulated that one of the oxygens of the phosphate group itself could abstract the proton from the catalytic water molecule, generating the nucleophilic OH^- (Langen *et.al,* 1992; Kamerlin *et.al,* 2013).

Recent research by Kozlova *et al.* (2022) has put forth a novel, unified mechanism of nucleotide hydrolysis in P-loop NTPases. Contrary to the previous notion that the proton acceptor is a non-conserved residue near the active site, Kozlova et al. proposes that the strictly conserved $[Ser/Thr]^{WA}$ and Asp^{WB} , located distally from the active site, act as the proton acceptors in all P-loop NTPases. This mechanism involves the relocation of a proton from the $[Ser/Thr]^{WA}$ residue to Asp^{WB} , triggered by the interaction of a P-loop NTPase with an activating partner such as a protein or RNA/DNA molecule. The resulting anionic $[Ser/Thr]^{WA}$ alkoxide then removes a proton from the catalytic water (W_{cat}) molecule, and the newly formed hydroxyl group attacks the gamma phosphate of

the NTP substrate. As the gamma-phosphate breaks away, the proton trapped at Asp^{WB} travels via a von Grotthuss relay through $[Ser/Thr]^{WA}$ to the beta-phosphate, thereby compensating for its developing negative charge (Kozlova *et.al,* 2022).

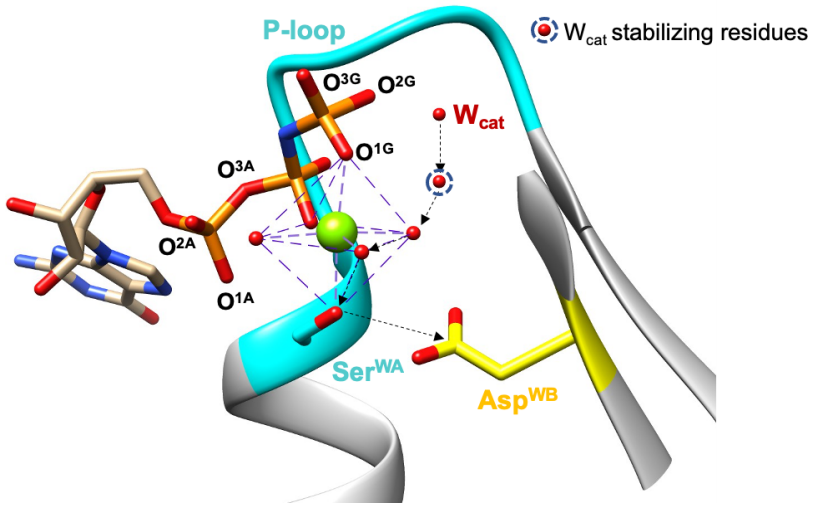


Fig. 4.1. Generation of nucleophilic hydroxyl ion in P loop NTPases
Black dashed arrows indicate a possible route of proton transfer from W_{cat} to the conserved base module of Ser^{WA}- Asp^{WB}. P-loop, Ser^{WA} highlighted in cyan, Asp^{WB} in yellow. The Mg^{2+} coordination is shown in dashed purple lines. Only the atoms of residues involved in Mg^{2+} coordination and Wcat stabilization are shown.

As per this mechanism, three components are crucial for NTP hydrolysis in P-loop NTPases:

1. Stimulatory residue - a positively charged amino acid (or a monovalent cation in some P-loop NTPases) that interacts with the free oxygens of beta or gamma phosphates, rotating it to disrupt the Mg^{2+} coordination shell and initiate hydrolysis by affecting the properties of the other Mg^{2+} ligands
2. W_{cat} stabilizing residues - amino acids or water molecules that hold the catalytic water molecule in place
3. Mg^{2+} coordinating residues - these residues, together with the W_{cat} stabilizing residues, provide a proton transfer hub for abstracting the proton away from Wcat and to the conserved base module.

In small Ras-like GTPases, the stimulatory residue is often an arginine provided in trans from a GAP while a glutamine residue is often implicated in stabilizing the catalytic water molecule.

While Ras GAPs provide the necessary residues for catalysis in Ras-like GTPases, higher P-loop NTPases lack an explicit GAP, possibly because they have structurally modified the minimal NTPase core to carry these crucial residues. This can be achieved through point mutations within the core or the evolution of insertions that contain the necessary residues. The aim of this section is to identify the structural modifications that enable higher P-loop NTPases to catalyze nucleotide hydrolysis with such competence.

4.2 Methods

Identifying stimulatory residue(s)

Given a P-loop NTPase structure, the stimulatory residues were identified as positively charged groups within 4 Å of the oxygens of PA, PB or PG (equivalent F in transition state structures). These groups included monovalent cations like Na^+ or K^+, NH1/NH2 groups of arginine, ND2 of asparagine, or NZ of lysine. A script that utilized the BioPython module was used to identify these groups. Once identified, the residue number, chain, and point of origin with respect to core or insertion were noted for the stimulatory residue. When a lysine residue was detected, an extra step was taken to confirm that it was not the Walker A lysine by checking the residue that occurred just after the lysine. If the residue following the lysine was a serine or a threonine, and the $[Ser/Thr]^{WA}$ was directly coordinated (in a few cases indirect coordination mediated through water) to the Mg, then the lysine would be the Walker A lysine near the beta phosphate group and not considered stimulatory.

Identifying catalytic water and its direct neighbors

Among all nucleotide-bound states, the catalytic water (W_{cat}) are best resolved in the Transition state (TS) complexes. In these complexes, W_{cat} was identified as the water

molecule in line with both the TS and the nucleotide, providing the OH^- for the apical nucleophilic attack required for hydrolysis. To determine W_{cat}, water molecules within 3.5 Å of the Fluorines or Oxygen of the TS analog are analyzed. The angle PB-TS(central atom)-HOH is measured for each water molecule, and W_{cat} is defined as the molecule with a $theta_{purine}$ value that falls within 180 ± 20°. After W_{cat} is identified, the residues that are within 3.5 Å of the water molecule are examined using the NeighborSearch class of the BioPython module. The residue number, chain, and point of origin with respect to core or insertion were noted for these direct neighbors. In instances where the direct neighbors of the W_{cat} include another water molecule, the residues that hold this water in place were identified in a similar way.

4.3 Results

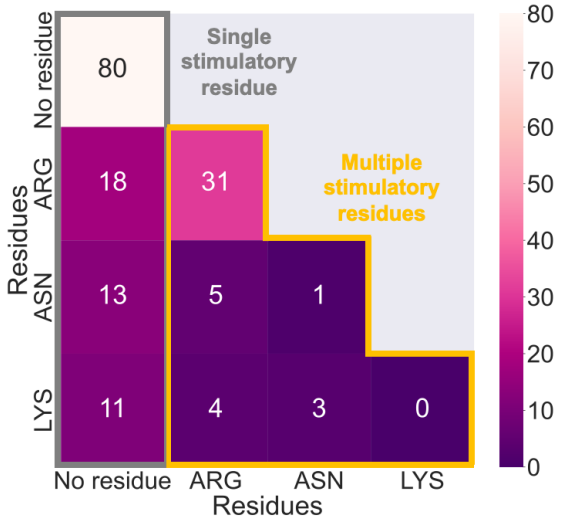
Of the 464 structures that passed the quality check criterion of section 3.2, this analysis will utilize the structures with NTP or transition state structures bound to it as the stimulatory fingers are properly oriented in these nucleotide-bound states of the NTPase. Among the 464 PDBs in the dataset, 168 were bound to NTP, NTP analogs, or TS analogs. The analysis in the following sections was carried out on this subset of 168 structures.

4.3.1 Identity of stimulatory residue in P-loop NTPases

Fig 4.2.A displays the frequency distribution of positively charged residues in the vicinity of the triphosphate moiety in the set of 168 structures. 80 of the 168 structures were not found to have positively charged residues in the vicinity of the phosphates. An inspection of these structures revealed that 15 instances were cation-dependent GTPases crystallized without the monovalent cation. A few members of P-loop NTPases insert an LSGGQ motif into the active site of adjacent subunit instead of relying on stimulatory fingers like Arg, Asn, Lys (Kozlova *et al.,* 2022). In the dataset this was observed in 33 PDBs. There were 20 PDBs that contained monomeric P loop NTPases that possibly require an oligomeric assembly to be catalytically active. There were 12 instances of a positively charged residue that was not oriented properly with respect to the oxygens of the phosphate and hence not captured by the NeighborSearch function.

Among the 88 structures that have at least one positively charged group near the beta and/or gamma phosphates, Fig.4.2 reveals that arginine is the most common stimulatory residue, being present in 49 structures. Interestingly, when arginine is the stimulatory residue, it more frequently appears paired with another arginine. Asparagine and lysine are less common stimulatory residues, and when present, they tend to act as singular stimulators.

A



B

Stimulatory residue		Family
ARG	-	ABC transporter ATPase (1), Extended AAA-ATPase (1), G proteins (6), Nitrogenase iron protein-like (5), RecA protein-like (3)
ARG	ARG	ABC transporter ATPase domain-like (4), Extended AAA-ATPase (7), G proteins (1), Nitrogenase iron protein-like (1), Nucleotide kinases (1), RecA protein-like (3), Tandem AAA-ATPase(11)
ARG	ASN	APS kinase (1), Extended AAA-ATPase (2), Nucleotide kinases (1), Motor proteins (1)
ARG	LYS	Extended AAA-ATPase (1), RecA protein-like (3)
ASN	-	ABC transporters(7), Extended AAA ATPase (1), Motor proteins (2), Nitrogenase iron protein-like (1), Tandem AAA-ATPase (2)
ASN	ASN	Nitrogenase iron protein-like (1)
ASN	LYS	Nitrogenase iron protein-like (3)
LYS	-	APS kinase (1), G proteins (1), Nitrogenase iron protein-like (7),RecA protein-like (2)

Fig. 4.2. Stimulatory residues in P-loop NTPases
A. The heatmap depicts how often positively charged residues appear near the triphosphate moiety. Each cell represents a combination of stimulatory residues, and the number in the cell indicates the number of PDBs in which that combination was found. The gray column with an x-axis value of "No residue" indicates instances where only a single stimulatory residue was found. The cells within the yellow border represent combinations of at least two positively charged stimulatory residues within the vicinity of the phosphates **B.** The table provides a family-wise breakdown of the stimulatory residues identified in Fig.4.2.A. The number of PDBs within each family that exhibit the combination of stimulatory residues is indicated in brackets.

Fig 4.2.A reveals a clear trend in P-loop NTPases of incorporating multiple stimulatory residues near the phosphate moiety. This trend contrasts with small Ras-like GTPases, where a single arginine residue, usually provided by the GAP, acts as the stimulatory residue. The family composition table in Fig 4.2.B shows that even families closely related to small Ras-like GTPases, such as G proteins and motor proteins, exhibit multiple stimulatory residues in the vicinity of the phosphate moiety. The nitrogenase iron protein NTPases exhibit almost all possible combinations of stimulatory residues, suggesting that this family does not strictly prefer a particular combination. ABC transporters seem to use either a single arginine or asparagine or a pair of arginine as the stimulatory residue. Extended and Tandem AAA ATPases more often have a pair of arginine near the phosphate moiety than a single stimulatory residue. The RecA family of NTPases contains members with either a single arginine, lysine, or a combination of the two as stimulators.

The stimulatory residues identified in Fig 4.2 can be categorized according to their interaction with specific phosphate groups. The distance between the positively charged side group and all free oxygens of the nucleotide is measured, and the group within 4 Å of the positively charged group is considered to be the interacting group. In some instances, the positively charged group may be within 4 Å of free oxygens of multiple phosphate groups. In such cases, the residue is considered to be interacting with all these groups.

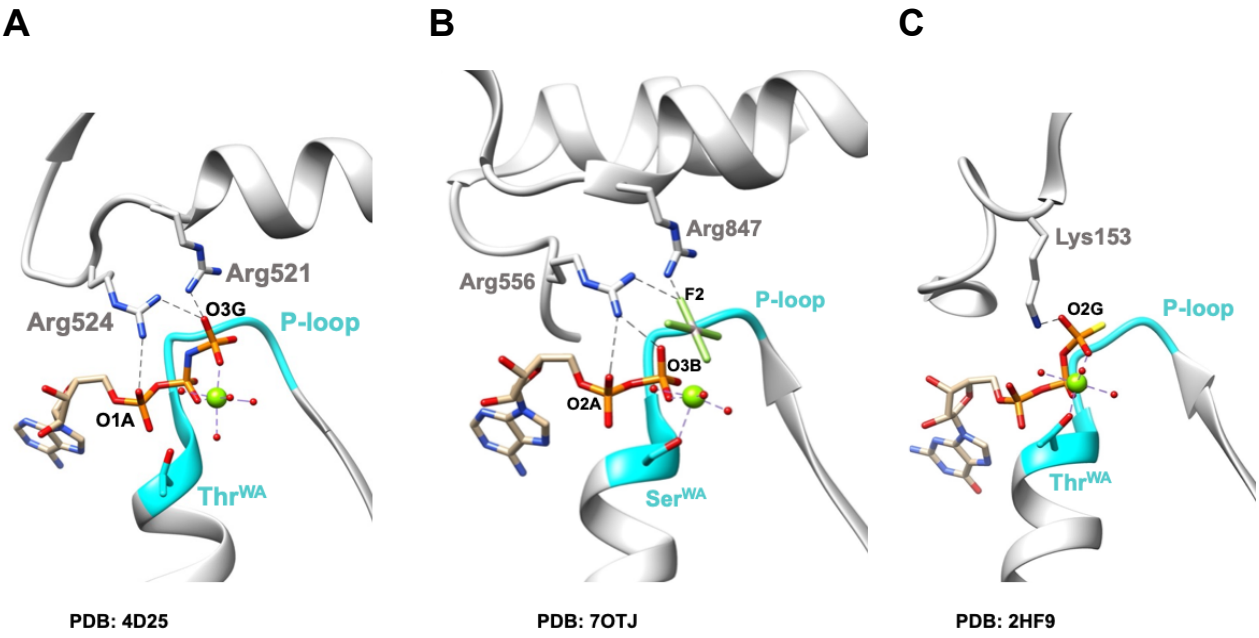


Fig. 4.3 Phosphate groups contacted by stimulatory fingers in P-loop NTPases
A. Arg524 of PDB 4D25 is within 4Å of free oxygens in both the PA and PG groups. While Arg521 contacts the free oxygen of PG alone. B. Arg556 of PDB 7OTJ contacts the free oxygens of all three phosphate groups, Arg847 contacts the free oxygen of PG group alone. C. Lys153 contacts the free oxygen of PG group alone. PDB 4D25 belongs to Tandem AAA+ ATPase family, PDB 7OTJ belongs to ABC transporter family, PDB 2HF9 belongs to G proteins. The family classification is based on SUPFAM server.

Fig 4.4 shows the frequency of interactions between stimulatory residues and phosphate groups in different types of P-loop NTPases. Fig 4.4.A reveals that out of the 42 single stimulatory residues identified from Fig 4.2, 40 of them interact with free oxygens of PG. Only in two cases is a stimulatory residue (both cases being Arg) found near the oxygens of PB and PG. These structures are however transition state

structures, and the O3B oxygen, (which would be bonded to PG in an NTP structure but is free in a TS structure) is being contacted. The orientation of the single stimulatory residue is reminiscent of the Arg from GAP in Ras GTPases, which was revealed to interact with the free oxygens of PG from the results of section 2.3.2.

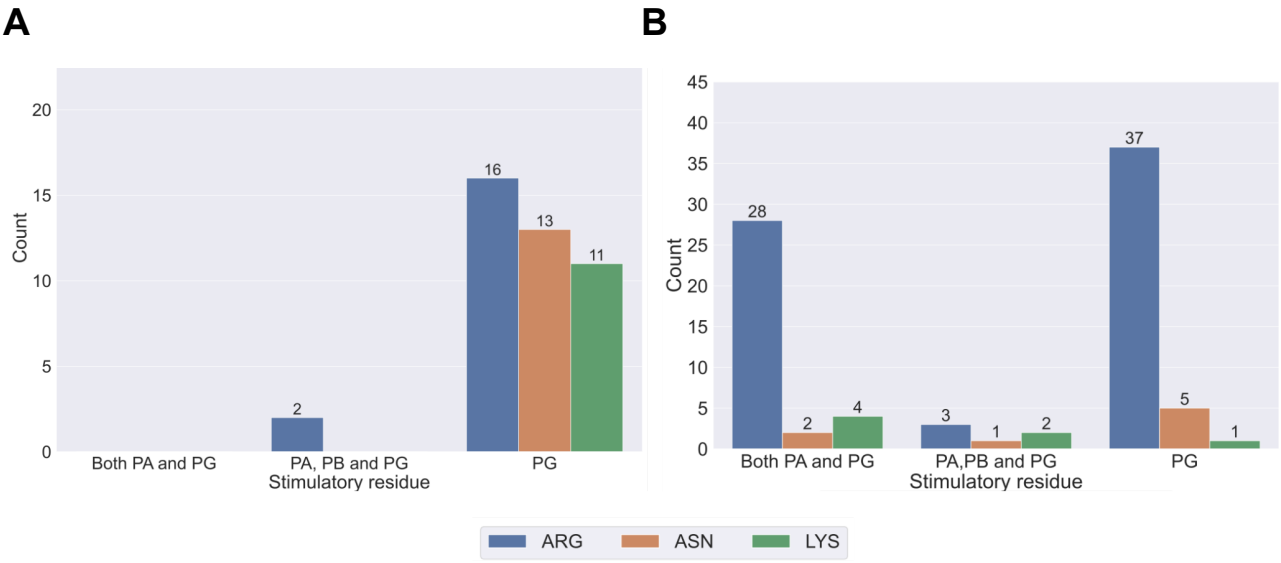


Fig. 4.4. Interactions of stimulatory residues with phosphate groups of the nucleotide
Stimulatory residues identified in Fig 3.2 are grouped based on the phosphate group they interact with. Residues are color-coded by type: blue for Arg, orange for Asn, and green for Lys. The numbers on top of each bar represent the number of occurrences of the corresponding stimulatory residue-phosphate group interaction. Panel A shows data for PDBs with a single stimulatory residue, while panel B shows data for PDBs with multiple stimulatory residues.

The frequency distribution for P-loop NTPases that carry multiple stimulatory residues depicted in Fig 4.4.B reveals the presence of a major population that interacts with oxygens of PG and hence behaves as the stimulator residue provided by GAP in small Ras-like GTPases. Fig 4.4.B also reveals the presence of positively charged residues that complement this conventional GAP stimulatory residue by interacting either with oxygens of PA or PB in addition to an oxygen of PG. The auxiliary stimulatory residue is most often an Arg that interacts with the oxygens of PA and PG while the role of the conventional stimulator residue is most often taken up by an Arg.

4.3.2 Structural origin of stimulatory residue

Next, we proceeded to map the location of the conventional and auxiliary stimulatory residues with respect to the core and insertion based structural framework (described in Chapter 3) of P-loop NTPases.

From Fig 4.5 it can be inferred that the location of the singular stimulatory residue in P-loop NTPases appears to be partially influenced by its identity. Specifically, when a single Arg serves as the stimulatory residue, it tends to be located in an insertion region, whereas Asn and Lys residues with stimulatory roles are typically found in the protein core. Notably, Lys residues exhibit the strongest preference for originating from the core.

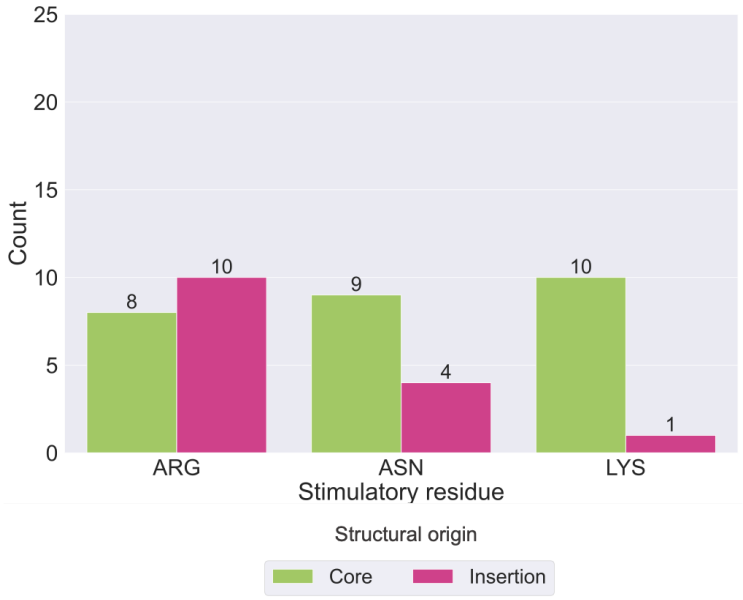


Fig 4.5. Structural origin of single stimulatory residues
The structural origin of the single stimulatory residues identified are traced back to the core or insertion. The bar plot illustrates the frequency of occurrence of stimulatory residues originating from either the core or insertion. The residues are grouped according to the type of amino acid: Arg, Asn, and Lys. The bars are colored based on their origin: violet for the core and orange for the insertion. The numbers on top of each bar indicate the total number of amino acids falling into the corresponding category.

Based on the current dataset, it appears that the conventional stimulatory residue Arg (residue 1 in Fig 4.6.A) is most commonly carried by insertions in P loop NTPases. Furthermore, the auxiliary stimulatory residue to the GAP, which is often another Arg, also tends to come from an insertion. Though the data on Lys being the auxiliary

residue is limited, it appears to originate more often from the core than insertion in P loop NTPases.

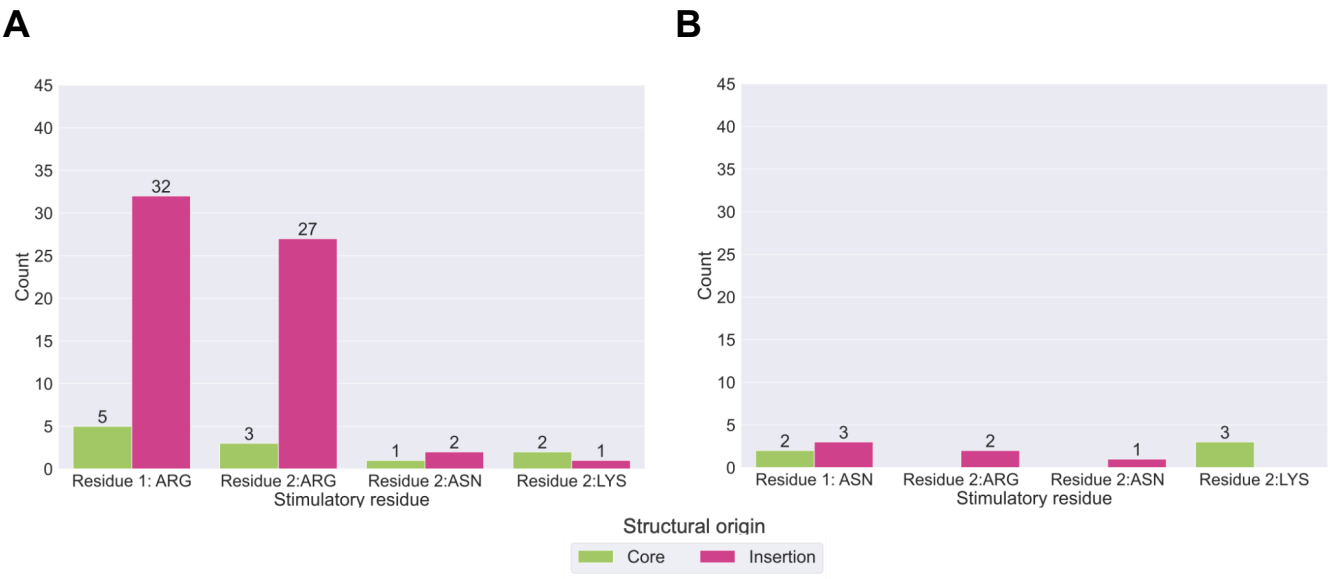


Fig. 4.6. Structural origin of multiple stimulatory residues.
The structural origin of the multiple stimulatory residues identified is traced back to the core or insertion. The bar plot illustrates the frequency of occurrence of stimulatory residues originating from either the core or insertion. The residues are grouped based on both the type of amino acid (Arg, Asn, and Lys) and the phosphate group contacted. Residue 1 refers to the residue that contacts the phosphate group alone, while Residue 2 refers to those that contact other phosphate groups (PA and PG, or PB and PG). The bars are color-coded based on their origin: violet for the core and orange for the insertion. The numbers above each bar indicate the total number of amino acids falling into the corresponding category. Panel A shows the structural origin of stimulatory residues when residue 1 is Arg. Panel B shows the same when residue 1 is Asn.

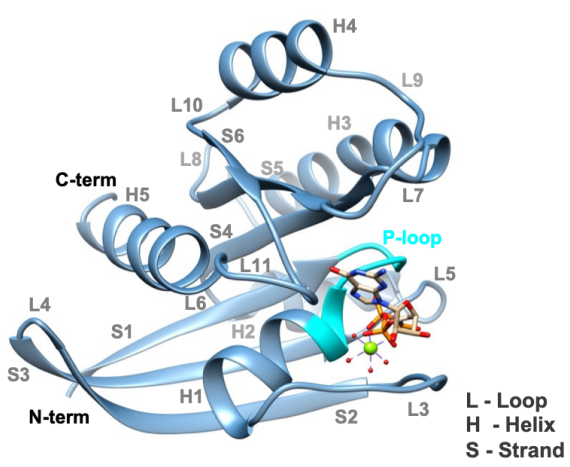
Taken together, the results in this section suggest that in P-loop NTPases with multiple stimulatory residues, insertions are likely to carry both the GAP equivalent and the auxiliary stimulatory residue.

4.3.3 Modification of core to carry catalytically critical residues

Table 4.1 reveals that the location within the core where stimulatory residues are introduced dictates how the stimulation proceeds. When stimulatory residues are introduced in secondary structures near the nucleotide-binding pocket (P-loop, Loop 3, Loop 5, Strand 5), the activation happens in cis. The P-loop in particular is revealed to

be a site that is prone to carrying cis-acting stimulatory residues across families regardless of the identity of the residue. Mutations occurring away from the nucleotide binding pocket result in stimulation being received in trans. Families that carry the stimulatory residues in cis would be active in their monomeric form while those that get the stimulatory residue in trans would be active only in an oligomeric state.

Stimulatory residue	Ras structure equivalent	Ras residue equivalent	Cis vs trans activity	Family
Arg (18)	P-loop (5)	G13 (5)	Cis	G proteins (3), Nucleotide kinases (1), Extended AAA-ATPase (1)
	Loop 3 (3)	Y32 (2), T35(1), D33(1)	Cis	G proteins (2), Motor proteins (1), RecA protein-like (1)
	Helix 2 (2)	T74 (1), M72 (1)	Trans	Extended AAA-ATPase (1), RecA protein-like (1)
	Helix 5 (2)	I163 (2)	Trans	RecA protein-like (2)
	Strand 3 (3)	T58 (3)	Cis	G proteins (2), Nucleotide kinases (1)
	Loop 5 (1)	G60 (1)	Cis	Extended AAA-ATPase(1)
	Strand 1 (1)	Y4 (1)	Trans	ABC transporter ATPase (1)



Stimulatory residue	Ras structure equivalent	Ras residue equivalent	Cis vs trans activity	Family
Lys (15)	P-loop (10)	G13 (1), A11 (9)	Cis, Trans	RecA protein-like (1), Nitrogenase iron protein-like (9)
	Helix 3 (1)	E91 (1)	Trans	G proteins (1)
	Loop 8 (2)	V109 (2)	Trans	RecA protein-like (3)
	Strand 2 (1)	D38 (1)	Cis	Nitrogenase iron protein-like (1)

Stimulatory residue	Ras structure equivalent	Ras residue equivalent	Cis vs trans activity	Family
Asn (12)	P-loop (8)	G12 (8)	Cis	ABC transporter ATPase (6), Tandem AAA-ATPase (2)
	Loop 3 (1)	T35 (1)	Cis	Nitrogenase iron protein-like (1)
	Loop 5 (3)	G60 (3)	Cis	Extended AAA-ATPase (3)

Table 4.1. Mutation hotspots that contain stimulatory residues in P-loop NTPases
The tables present information on the secondary structure of the Ras core within P loop NTPases that have been mutated to carry stimulatory residues. The "Ras residue equivalent" column indicates the Ras GTPase residue position that is now occupied by the stimulatory residue. The ”Cis vs Trans activity" column indicates whether the stimulatory residue is in the same subunit (Cis) or a different subunit (Trans) vicinity of the nucleotide. The "Family" column shows the P-loop NTPase family displaying the particular stimulatory residue. The numbers in brackets represent the number of PDBs that exhibit the corresponding characteristic. A small Ras-like GTPase structure (PDB:5P21) with secondary structures labeled shown on the right top corner.

Motor proteins and G proteins seem to incorporate Arg fingers in regions near the nucleotide-binding pocket like P-loop, Loop 3, strand 3 and hence enable cis activation. On the other hand, in families such as RecA proteins, stimulatory residues like Arg and Lys are incorporated both near and away from the nucleotide-binding pocket, resulting in cis or trans activation depending on the location of the mutation. In families like ABC transporters and Nitrogenase iron proteins, the identity of the stimulatory residue

influences which part of the core is mutated. For instance, if the stimulatory residue is Arg, regions further from the nucleotide-binding pocket are mutated, allowing the Arg to act in trans from an adjacent subunit. However, if the stimulatory residue is Asn, regions of the core near the nucleotide-binding pocket are mutated, allowing the Asn to act in cis.

While modifications of the core account for all instances of G proteins, distantly related families such as APS kinase do not seem to use point mutations to carry the stimulatory residues.

The frequency distribution of residues within the Ras core that gets mutated into stimulatory residues (Fig 4.7) reveals that mutations to catalytic residues occur throughout the core.

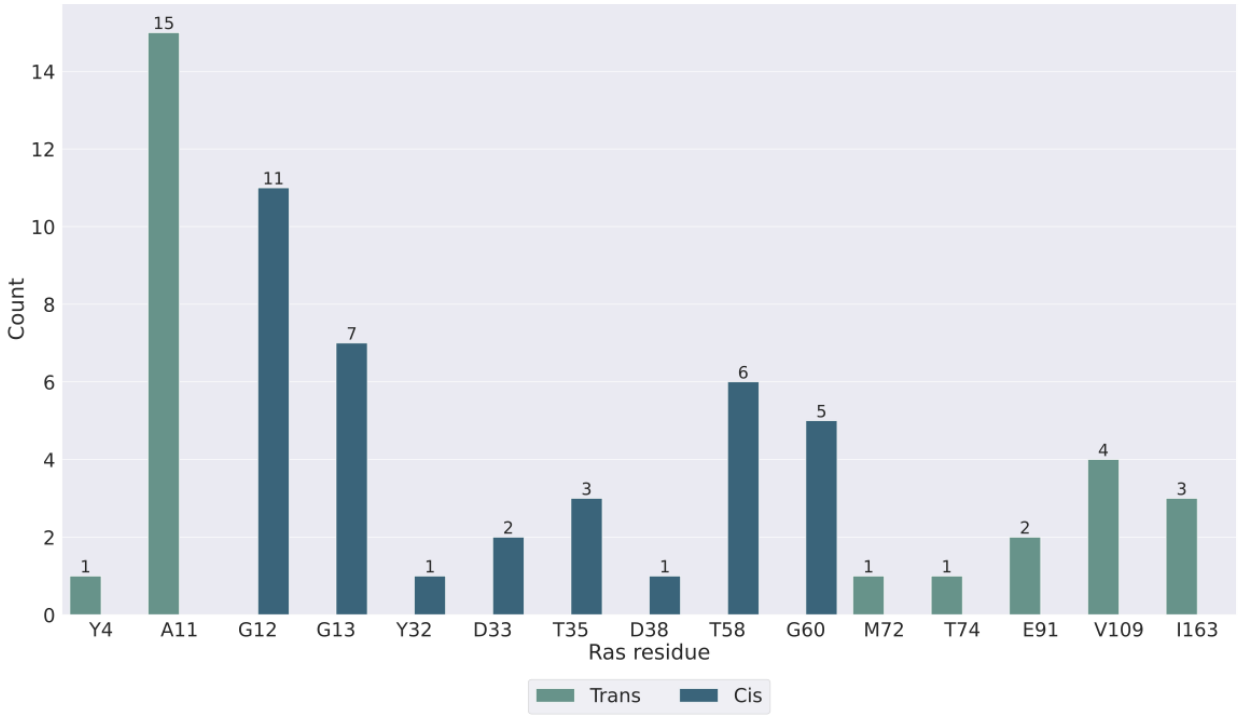


Fig. 4.7 Frequency of mutations at specific residue positions within the Ras GTPase core
The single letter code of residues followed by their residue number in 5P21. Denotes how often residue in Ras core gets mutated to become a stimulatory residue. That is in 11 cases when the stimulatory residue appears from the core it takes the position of G12 in Ras-GTPases. The bars are colored by the mode of activity of the stimulatory residue introduced at that position. That is all 11 instances when a stimulatory residue is introduced at G12 the stimulatory residue acts in cis.

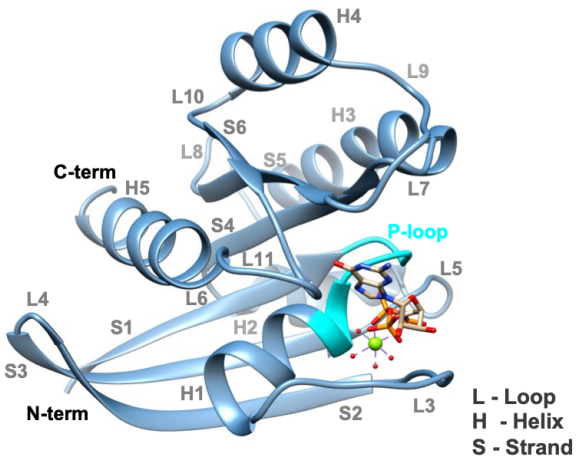
It is interesting to note that the residues that get mutated to carry the stimulatory residue in cis have also been identified by experimental studies to be sites of oncogenic significance. In oncogenic mutants of Ras, the residues identified above, G12, G13, and G60 have been mutated to acidic residues, most often Aspartate (Bos et.al, 1987; Prior et.al 2012). Mutating the glycine at these positions to acidic residues like aspartate shifts the GTPase to a non-hydrolytic state, even in the presence of a GAP. The results from this section suggest that evolutionarily related P-loop NTPases also mutate these residues but mutate them into positively charged residues that serve as stimulatory residues.

Therefore, we can distinguish between three cases:
Case 1 - a mutated Ras GTPase with aspartate at these positions, which cannot hydrolyze GTP to GDP even in the presence of GAP
Case 2 - an unmutated Ras GTPase with glycine in these positions, which can hydrolyze GTP to GDP in the presence of GAP
Case 3 - a mutated Ras GTPase with arginine, lysine, or asparagine in these positions, which does not require GAP to hydrolyze GTP to GDP.

4.3.4 Evolution of insertions to carry catalytically critical residues

Table 4.2 illustrates that insertions containing arginine are prevalent throughout the core of Ras GTPase, with 11 secondary structures harboring such insertions. Certain secondary structures, such as loops 6, 7, 8, helix 2, and strand 4, appear to carry insertions that contain arginine exclusively, without any asparagine or lysine residues that position near the nucleotide phosphates. Insertions containing asparagine or lysine are found in fewer secondary structures. Helix 3 and strand 2 seem to be particularly prone to insertions, as insertions to these structures have been found in all cases, regardless of the type of stimulatory residue.

Stimulatory residue	Ras structure equivalent	Cis vs Trans activity	Family
ARG	N-term	Trans (2), Cis (2)	Extended AAA-ATPase (2), RecA protein-like (2)
	Helix 1	Cis (1)	RecA protein-like (1)
	Helix 2	Trans (6)	Extended AAA-ATPase (6)
	Helix 3	Cis (17)	ABC transporter ATPase (4), APS kinase (4), Tandem AAA-ATPase (9)
	Loop 3	Cis (5)	Extended AAA-ATPase (2), Nitrogenase iron protein-like (3)
	Loop 6	Trans (2)	Extended AAA-ATPase (2)
	Loop 7	Cis (19)	ABC transporter ATPase (4), Extended AAA-ATPase (5), Nitrogenase iron protein-like (2), Tandem AAA-ATPase (8)
	Loop 8	Trans(2)	RecA protein-like (2)
	Loop 9	Trans (10)	Extended AAA-ATPase (4), RecA protein-like (1), Tandem AAA-ATPase (5)
	Strand 2	Cis(4)	Extended AAA-ATPase (1), Nitrogenase iron protein-like (2), RecA protein-like (1)
	Strand 4	Trans(1)	RecA protein-like (1)



Stimulatory residue	Ras structure equivalent	Cis vs Trans activity	Family
LYS	Helix 3	Cis (1)	APS kinase (1)
	Loop 9	Trans (1)	RecA protein-like (1)
	Strand 2	Cis (1)	Extended AAA-ATPase (1)

Stimulatory residue	Ras structure equivalent	Cis vs Trans activity	Family
ASN	N-term	Cis (1)	ABC transporter ATPase (1)
	Helix 1	Cis (3)	Motor proteins (3)
	Helix 3	Cis (1)	Nucleotide kinases (1)
	Loop 3	Cis (2)	Nitrogenase iron protein-like (2)
	Strand 2	Cis (5)	APS kinase (1), Nitrogenase iron protein-like (4)

Table 4.2. Insertion hotspots that contain stimulatory residues in P loop NTPases
The tables provide information on the secondary structure of the Ras core within P loop NTPases into which insertions carrying stimulatory residues were made. The "cis vs trans" column indicates whether the stimulatory residue is in the same subunit (cis) or a different subunit (trans) vicinity of the nucleotide. The "Family" column shows the P loop NTPase family displaying the particular stimulatory residue. The numbers in brackets represent the number of PDBs that exhibit the corresponding characteristic. A small Ras-like GTPase structure (PDB: 5P21) with secondary structures labeled shown on right top corner.

Certain families, such as ABC transporters, Motor proteins, Nitrogenase Iron, and APS kinase, do not exhibit a strict preference for any particular type of stimulatory residue within insertions. Rather, they incorporate these residues in a way that allows them to act in cis. Other families, such as Extended, Tandem AAA+, and RecA proteins, do not show a preference for either cis- or trans-acting stimulatory residues within insertions.

The results suggest that the evolution of insertions to contain stimulatory residues is more commonly employed by evolutionarily distant members, as evidenced by the paucity of members from closely related G-protein families and the dominance of distantly related ATPase families in the panel. Sections 4.3.3 and 4.3.4, taken together,

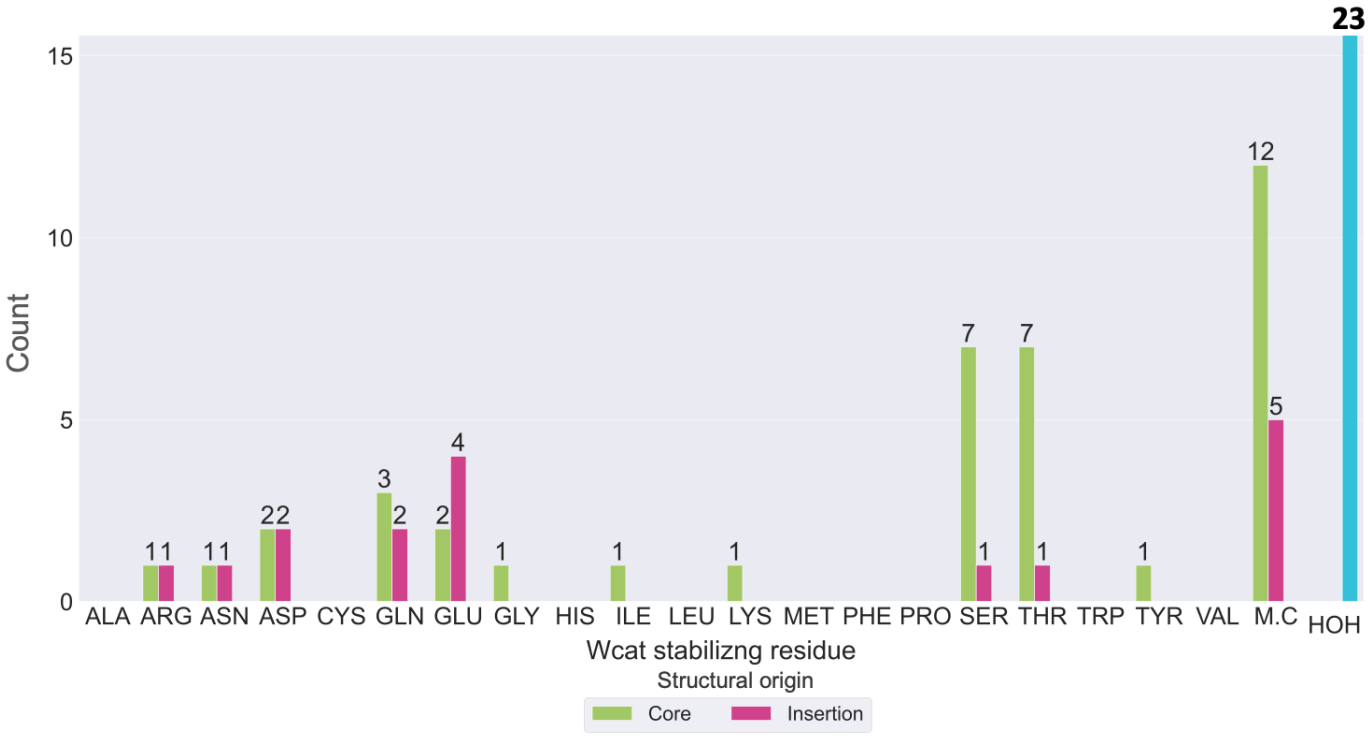
indicate that distantly related P-loop NTPase families use the evolution of insertions to carry stimulatory residues as a mechanism, while closely related G protein families modify their existing core architecture to carry these residues.

4.3.5 Catalytic water stabilization in P-loop NTPases

To ensure the unambiguous identification of the catalytic water molecule (W_{cat}), the dataset used in this section includes only those P-loop NTPases bound to TS analogs and with the catalytic water molecule resolved within the nucleotide-binding pocket. Therefore a subset of 25 P-loop NTPases bound to TS analogs from the 464 dataset was utilized for this part of the analysis.

In small Ras-like GTPases, there is an emphasis on specific residues, such as glutamine, to stabilize the W_{cat}. However, in P-loop NTPases, the analysis of direct neighbors of the W_{cat} from Fig 4.8. A reveals a much reduced focus on specific residues. Instead, main chain interactions and other water molecules dominate the H bonding network around the W_{cat}. The family distribution from Fig 4.8.B suggests that this shift in the mechanism of stabilizing the W_{cat} happens early on in the P-loop NTPase history as closely related members like G proteins adopt this trend. Interestingly, the few instances in which a glutamine stabilizes the W_{cat} happen in the distantly related family of Tandem AAA-ATPases and not families closely related to small Ras-like GTPases like G-proteins or Motor proteins.

A



B

Direct neighbour of catalytic water	Structural origin	Family
GLU (6)	Core (2)	Tandem AAA-ATPase domain (2)
	Insertion (4)	ABC transporter ATPase domain-like (1), Motor proteins (1), RecA protein-like (1), Tandem AAA-ATPase domain (1)
GLN (5)	Core (3)	G proteins (1), Tandem AAA-ATPase domain (2)
	Insertion (2)	Tandem AAA-ATPase domain (2)
Main chain (17)	Core (12)	G proteins (4), Motor proteins (4), Tandem AAA-ATPase domain (2), Extended AAA-ATPase domain (2), ABC transporter ATPase domain-like (1)
	Insertion (5)	G proteins (2), Nitrogenase iron protein-like (1), Tandem AAA-ATPase domain (2)
HOH	-	G proteins (6), Motor proteins (4), Nitrogenase iron protein-like (5), RecA protein-like (1), Tandem AAA-ATPase domain (4), Extended AAA-ATPase domain (2), ABC transporter ATPase (1)

Fig. 4.8. Direct neighbors of catalytic water in P-loop NTPases

A. Frequency distribution of residues within 3.5 Å of catalytic water in P-loop NTPases. The three letter codes of the amino acid residues, M.C - for Main chain interactions, HOH - water molecules. The bars are colored on the basis of their origin. The number on top of the bar denotes the number of occurrences of the particular residue near the catalytic water. B. The frequency distribution is broken down to identify the P-loop NTPase families that use the particular residue to stabilize the catalytic water. The data is shown only for the major stabilizers which is Main chain and HOH. The family distribution for GLU and GLN is also shown as these residues are the ones that are utilized in small Ras-like GTPases to stabilize the catalytic water.

Analyzing the indirect stabilization of W_{cat} from Fig 4.9 reveals involvement from residues like serine, threonine, aspartate, glutamine, and main chain interactions. The prevalence of serine, and threonine in the vicinity of the W_{cat}, either as direct or indirect neighbors can be reasoned by their ability to partake in the von Grotthuss relay given the presence of side groups have a proton-accepting lone pair of electrons and also their own proton to transfer (Kozlova *et.al,* 2022). These residues come more often from the core than the insertion.

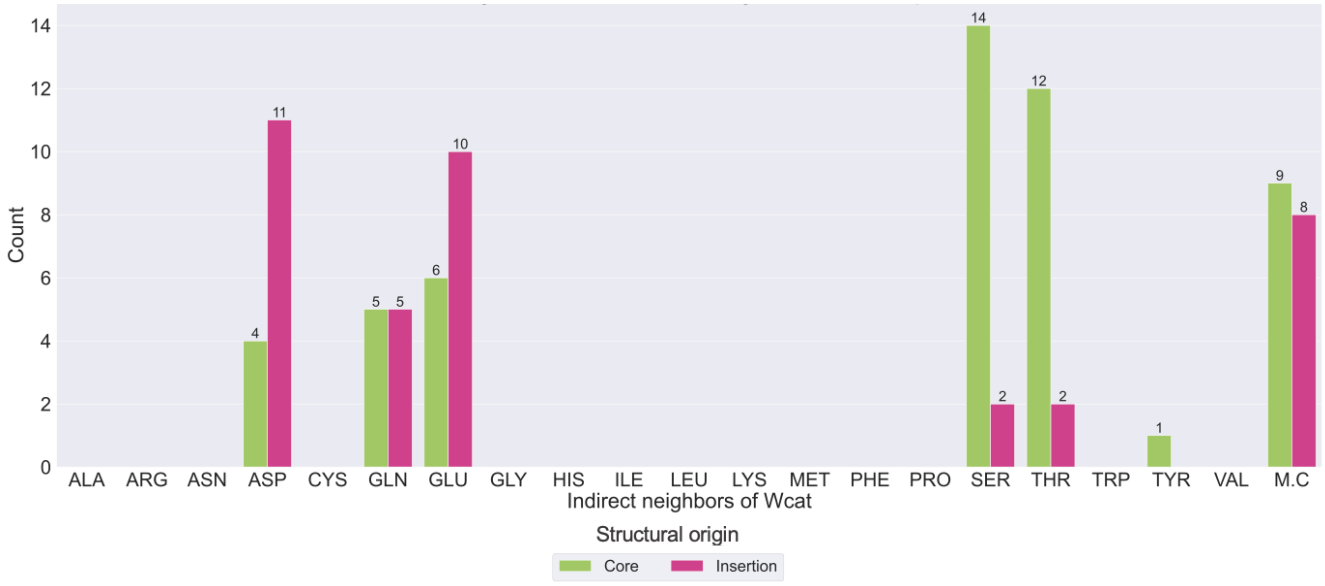


Fig. 4.9. Indirect stabilization of catalytic water in P-loop NTPases. Frequency distribution of residues within 3.5 Å of water molecules that directly stabilize Wcat in P-loop NTPases. The three letter codes of the amino acid residues, M.C - for Main chain interactions. The bars are colored on the basis of their origin. The number on top of the bar denotes the number of occurrences of the particular residue near the water molecules that directly stabilize the W_{cat}.

This shift in fulfilling H bonding requirements from Glutamine to main chain interactions and other water molecules could be a key reason why P-loop NTPases do not require an explicit GAP to achieve catalytic competency.

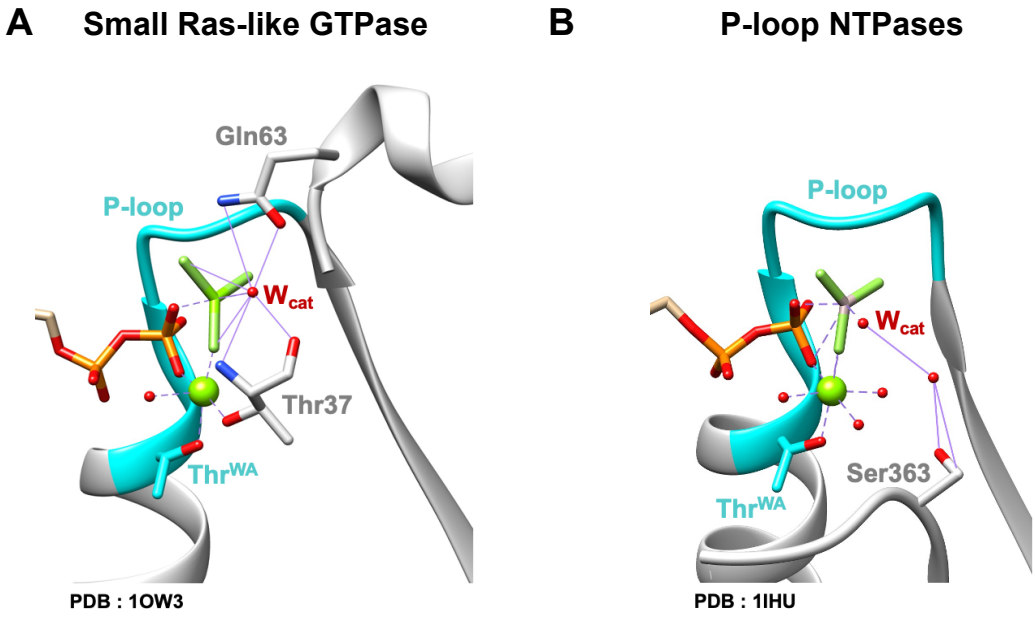


Fig. 4.10. Differences in the mechanism of stabilizing W_{cat} in P-loop NTPases
A. In small Ras-like GTPases, a specific Glutamine residue is involved in the stabilization, as depicted by the example of the Rho GTPase (PDB 1OW3). **B.** In other P-loop NTPases, the mode of stabilizing the catalytic water shifts from specific residues to incorporate indirect contacts, exemplified by Ser363 in the AAA ATPase (PDB 1IHU). The protein secondary structure is shown in gray, with only the P-loop highlighted in both A and B

4.4 Conclusion

- Families closely related to Ras GTPases, such as G-proteins, may adopt a strategy of modifying their existing core architecture to carry the stimulatory residue. Families distantly related to small Ras-like GTPases, such as Extended AAA+ ATPases and Tandem AAA+ ATPases, seem to be evolving insertions to gain catalytic competence.
- The insertion point was not limited to any specific portion of the small Ras-like GTPase core, as insertions carrying the stimulatory residue were observed at the core's N-term, interior, and C-term.

- The analysis of direct neighbors of W_{cat} revealed that the mechanism by which catalytic water is stabilized differs between Ras-like GTPases and higher P-loop NTPases. There seems to be a greater utilization of main chain atoms both from the core and insertions and other water molecules to stabilize the W_{cat}. There however seems to be indirect coordination of W_{cat} prevalent in P-loop NTPases using serine, threonine from the core and aspartic acid, glutamine from insertions.
- The catalytic geometry in P-loop NTPases seems reminiscent of that in small Ras-like GTPases as the stimulatory residue comes from an insertion while the W_{cat} stabilizing residues (though indirect) come from the core.

Chapter 5

Examining adenine binding environment in P-loop NTPases

5.1 Introduction

The P-loop NTPase superfamily encompasses a diverse range of proteins that play critical roles in various biological processes. The conserved P-loop unites all members of the superfamily in their ability to bind and stabilize the phosphate moiety, yet family members do show variations in overall nucleotide specificities and affinities. The protein's specificity and affinity for a nucleotide is determined by the degree to which the local environment surrounding the nucleotide stabilizes it. Therefore, differences in the purine environment in P-loop NTPases could account for variations in affinity and specificity between members. For instance, the local purine binding environment in Ras-like GTPases is specifically primed to recognize and stabilize guanine. Hence, if an adenine is placed in this pocket, there wouldn't be sufficient stabilization to hold it in place. The fact that the P-loop NTPase superfamily encompasses a large number of ATPases indicates that there have been modifications to the minimal NTPase core, which allowed the shift in substrate specificity from guanine to adenine.

This chapter aims to examine the modifications in the minimal GTPase that may have underpinned this transition in substrate specificity. Specifically, it will explore whether the switch in substrate specificity was a singular concerted event, whereby all P-loop ATPases bind and stabilize adenine in a similar fashion, or whether it involved a series of independent events, with different ATPases choosing their unique modes of modifying the core to stabilize adenine.

5.2 Methods

Characterizing purine positioning

Our analysis began by characterizing the positioning of purine within the nucleotide-binding pocket for the various P-loop NTPase families. Similar purine positioning hints at similarity in the local environment stabilizing it. Thus identifying P-loop ATPases that place the adenine similarly would help group the strategies utilized to modify the Ras core architecture for binding adenine instead of guanine. To accurately capture the positioning of the purine plane relative to the guanine in small Ras-like GTPases, we define two angular parameters that will be calculated for the nucleotide in each of the P-loop NTPase in the dataset from section 3.3.

The first angular parameter is $theta_{purine}$, which is the angular tilt of the purine plane relative to guanine in Ras-like GTPases. To calculate $theta_{purine}$, we performed a superposition of 5P21 (as reference Ras GTPase structure) and the PDB of the P-loop NTPase member by matching the conserved P-loop residues. After the P-loops are superimposed, planes passing through the purine ring atoms common to both adenine and guanine (namely N1, C2, N3, C4, C5, C6, N7, C8, N9) were defined, followed by defining the two normal vectors to the planes using UCSF Chimera. The crossing angle between the two vectors was then recorded, resulting in a $theta_{purine}$ value that ranges from 0°- 90°, where a value of 0° indicates no relative tilt between the guanine and purine planes, while a value of 90° signifies that the purine plane is perpendicular to the orientation of the guanine plane in Ras GTPases.

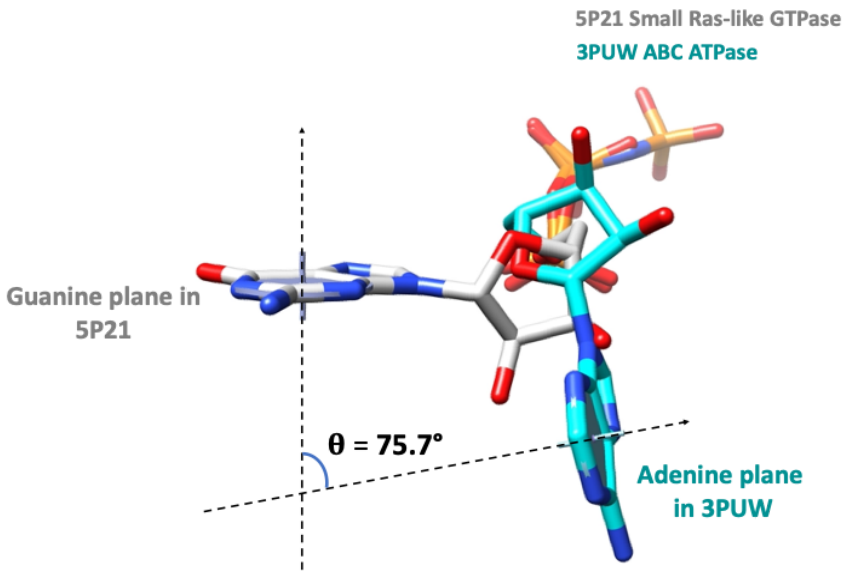


Fig. 5.1. Measuring angular tilt of purine plane relative to guanine in 5P21
P-loop ATPase (PDB 3PUW) is superposed with 5P21 along the P-loop, which allowed perfect superposition of the phosphate moieties. The angular tilt of adenine in 3PUW relative to guanine in 5P21 is measured as described in text.

The second parameter we used to characterize purine placement is phi_{purine}, which is defined as the torsion angle (O4′, C1′, N9, and C4) that describes the orientation of the purine in relation to its ribose. Unlike $theta_{purine}$, phi_{purine} is an absolute parameter that establishes the *syn* and *anti*-conformations of the nucleotide. To compute phi_{purine} for a given PDB, we extracted the coordinates of four atoms within the purine (O4′, C1′, N9, and C4) from the PDB file using Python and passed them into a custom function that calculates the dihedral angle. The stereochemical arrangements corresponding to torsion angles between 0° and ±90° are called *syn*, while those corresponding to torsion angles between ±90° and 180° are called *anti*. The value of phi_{purine} for the guanine in small Ras-like GTPase 5P21 is -112°, corresponding to the anti-conformation.

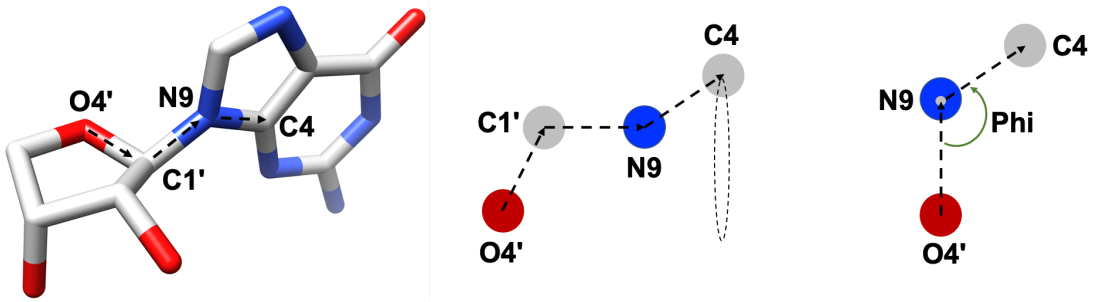


Fig. 5.2. Relative placement of purine and ribose in a nucleotide

phi_{purine} was chosen as the second parameter to characterize purine placement because when considered together with $theta_{purine}$, it can capture instances where the relative placement of the 6 and 5-member rings of purines change. In cases where the value of phi corresponds to the *syn* conformation and the $theta_{purine}$ is within 20 degrees, the 6-member and 5-member rings of adenine switch places relative to their orientation in guanine of 5P21. It must be noted that both factors must be satisfied for this switch in ring positions.

Clustering purine placements

Once the purine placement in the P-loop NTPases were mapped onto the $theta_{purine}$, phi_{purine} space, we grouped these points into clusters based on similarity in the $theta_{purine}$, phi_{purine} values. This clustering was performed using the K-means clustering algorithm, specifically utilizing the sklearn package in Python. K-means is a widely used unsupervised machine learning algorithm that clusters data points into groups based on their similarity in the features examined. In this context, the features are the values of the $theta_{purine}$ and phi_{purine} parameters. However, one common issue with K-means, and other feature-similarity based algorithms, is that when features are measured on different scales, those with larger scales may dominate the similarity calculation, resulting in biased clustering results. In our case, the $theta_{purine}$ lies between 0°-90° while the phi_{purine} value range is between -360° to 360°. Implementing the standard K means clustering algorithm on its own lead to a clustering that places a greater emphasis on phi_{purine} value over $theta_{purine}$ value owing to this difference in scales of the features. That is, the clustering would be more sensitive to phi_{purine} while being less sensitive to the $theta_{purine}$.

To mitigate this issue, we applied a standard preprocessing technique known as Z-score normalization. This transformation involves subtracting the mean of a feature ($theta_{purine}$, phi_{purine}) and dividing it by its standard deviation, which results in a new distribution with a mean of 0 and a standard deviation of 1. By standardizing the features using Z-score normalization, we ensured that each feature contributes equally to the similarity metric and is scaled to a common range. This helped prevent phi_{purine} from dominating the clustering process and improved the accuracy of the resulting clusters.

Identifying atomic contacts stabilizing adenine

For every P-loop NTPase in the dataset, the atomic contacts to the ten adenine atoms (N1, C2, N3, C4, C5, C6, N6, N7, C8, N9) were identified using a custom script that utilized the NeighborSearch module of the BioPython class in Python. The residue number and chain id of the atom making contact were also retrieved from the program. Once the residues making atomic contact with adenine were identified, structural information, including core, insertion, and secondary structure details, were added using the core insertion mapping script developed. Atoms that are polarized, such as oxygen (O) and nitrogen (N), and lie within a distance of 3.5 Å to N6 or N9 will be considered as hydrogen (H) accepting atoms. On the other hand, if they are in contact with N1, N3, and N7, they will be considered as H donors. In cases where any of the interacting atoms are non-polar, the interaction will be considered as a van der Waals interaction.

5.3 Results

Out of the 464 PDBs that passed the quality check in section 3.1, 302 were bound to adenine nucleotides. This section aims to describe the positioning of adenine in these 302 P-loop ATPases.

5.3.1 Clusters in adenine positioning

Interpreting the purine position map

The intersection of the two dashed lines in Figure 5.3.A represents the positioning of guanine in small Ras-like GTPases. Points near this intersection indicate positioning of the adenine in P-loop ATPases that would be similar to the guanine in Ras-like GTPases. The separation of the clusters in the horizontal (theta) axis, indicates variations in the tilt of the adenine plane relative to the guanine in Ras. In contrast, vertical separation indicates differences in the relative orientation of purine and ribose.

A

B

Clusters	$Theta_m$ ± std	Phi_m ± std	Families
Cluster 1	14.61 ± 6.92	-121.55 ± 20.22	Nitrogenase iron protein-like (32), Tandem AAA- ATPase (29), RecA protein (24), Nucleotide kinases (24), ABC transporters (9), Extended AAA ATPase (6), Phosphoribulokinase (3), G proteins (2), Motor proteins (1), APS kinase (1), Terminase (1)
Cluster 2	41.59 ± 9.20	-151.10 ± 21.77	Extended AAA ATPase (60), Motor proteins (13), RecA protein (10), Tandem AAA- ATPase (8), Nitrogenase iron protein-like (2), Nucleotide kinases (1), ABC transporters (1), G proteins (1)
Cluster 3	77.10 ± 6.72	-124.38 ± 21.43	ABC transporters (46), RecA protein (7), Extended AAA ATPase (3), Tandem AAA ATPase (2), G proteins (1)
Cluster 4	19.24 ± 6.91	64.31 ± 8.10	APS kinase (3).YjeE-like (3),G proteins (1), Extended AAA- ATPase (1)

Fig. 5.3. Mapping purine positioning in the nucleotide-binding pocket of P-loop ATPases
A. Positioning of the purine moiety within the nucleotide-binding pocket of P-loop ATPases. The x-axis represents the relative angle between the purine plane of 5P21 and adenine plane of a P-loop ATPase ($theta_{purine}$), while the y-axis shows phi_{purine}, the torsion angle (O4'-C1'-N9-C4) for adenine. The black dashed line at *phi* = -112 represents the mean phi value for a set of 28 Ras-like GTPases, and the vertical dashed line at $theta_{purine}$ = 5 represents the mean $theta_{purine}$ value for the same dataset. Values corresponding to a total of 302 P-loop ATPases are plotted, with each dot corresponding to a specific purine orientation. The points are color-coded according to the cluster they belong to. The shape of the points represents the P-loop NTPase family they belong to **B.** Statistics of purine positioning clusters. Each cluster's mean $theta_{purine}$ and phi_{purine} values are presented in columns, along with their respective standard deviations. The "Family" column provides a breakdown of each cluster based on the P-loop NTPase family its members belong to. The number in brackets after each family name indicates the number of PDBs from that family in the dataset that fall into the corresponding cluster.

Among the clusters of adenine positioning in Figure 5.3.A, members of cluster 1 show the greatest similarity to the guanine positioning in Ras-like GTPases. Points near either of the dashed lines represent instances where one parameter of purine positioning is similar to that in Ras. For example, clusters 1 and 3 fall close to the horizontal dashed line representing the phi value of Ras GTPases, indicating a similar orientation of the purine relative to the ribose in these clusters. While members of cluster 3 retain the relative orientation of purine and ribose that Ras GTPase guanines have, these P-loop ATPases show a pronounced tilt in their adenine plane. The family composition table indicates that members of the ABC transporter subfamily dominate this cluster.

Members of cluster 2 in Figure 5.3.A show moderate changes in phi_{purine} and $theta_{purine}$ indicating a distinct positioning of adenine relative to the guanine in Ras-like GTPases (Fig 5.4). The adenine in cluster 4 ATPases is in syn conformation, even though the adenine plane is not tilted much. The syn conformation of the adenine together with the $theta_{purine}$ value being small manifests as "purine inversion," such that the 5-membered ring of adenine falls in the place of the 6-membered ring of guanine when the structures are superposed along their P-loops (Fig 5.4).

The composition of P-loop ATPase families varies across the clusters defined by differences in $theta_{purine}$ and phi_{purine} values. For instance, Cluster 1, which has the $theta_{purine}$ and phi_{purine} values most similar to small Ras-like GTPases, includes members from a diverse range of families while cluster 3 has representation from mainly a single family. Interestingly, some families, such as the Extended AAA+ ATPase family, are present in all clusters, while others are more specific to certain clusters. These observations highlight the limitations of clustering based solely on family, as it may overlook important features shared between members of different families.

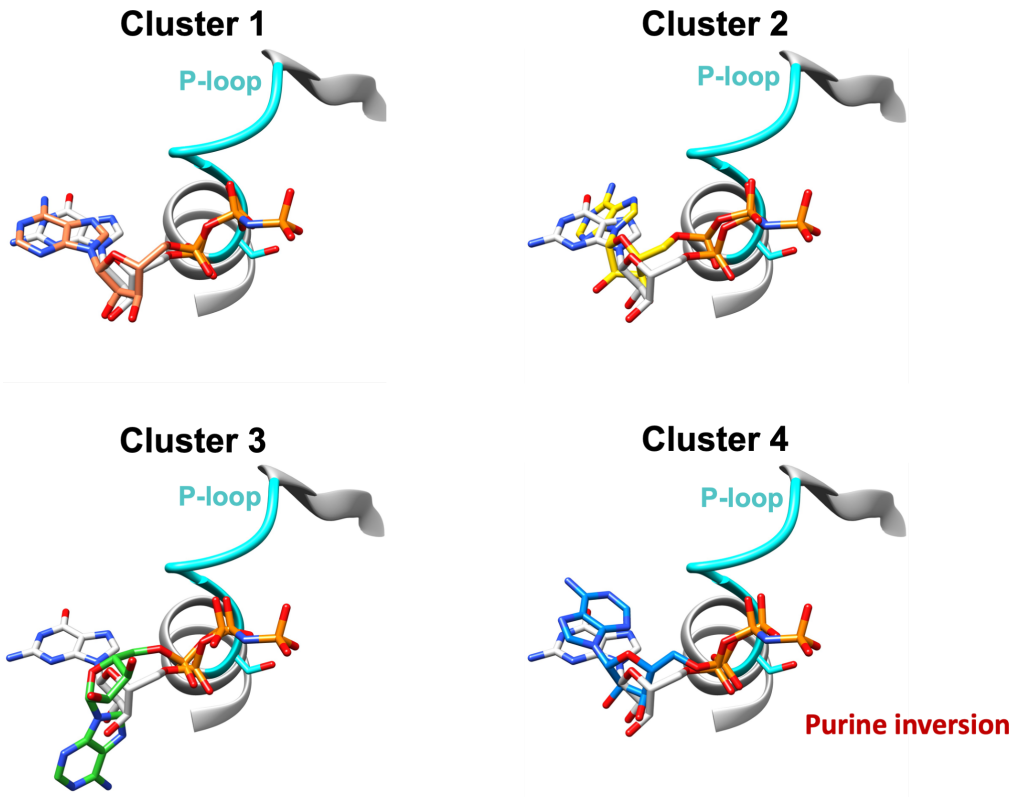


Fig. 5.4. Adenine positioning in P-loop ATPases
Representative members from each cluster of P-loop ATPases, with $theta_{purine}$ and phi_{purine} values that closely match the mean of their respective cluster. The structures have been superposed with 5P21 along the P-loop, resulting in a perfect superposition of the phosphate moieties. For clarity, the structure of the P-loop ATPase has been hidden, and only the guanine, P-loop, part of strand 1, and alpha 1 helix of HRas GTPase (PDB:5P21) are shown in gray.

The presence of distinct clusters in the $theta_{purine}$ - phi_{purine} space indicates that adenine can be accommodated in the nucleotide-binding pocket of P-loop NTPases through multiple modes. This implies that the guanine binding architecture in the minimal P-loop NTPase, such as small Ras-like GTPases, can be modified in various ways to facilitate adenine binding.

The following sections will examine how these strategies vary in terms of :

1. The total number of atomic contacts and the number of direct neighbors to adenine
2. The stabilization of adenine atoms - as evidenced by percent investment to each purine atom, nature of the interaction (van der Waals, H bond)
3. Structural composition of the adenine binding pocket - in terms of

a. extent of core, insertion, water contacts utilization
b. secondary structure within core that stabilizes the adenine atoms
c. point of insertion within the core that stabilizes the adenine
d. residues that stabilize adenine

5.3.2 Variation in number of atomic contacts and the number of direct neighbors to adenine across clusters

Modifying the binding environment of guanine to accommodate adenine is expected to involve adjustments in the hydrogen bonding network that are specifically tailored to meet the unique hydrogen bonding requirements of adenine, which are distinct from those of guanine.

The data from Fig 5.5.A reveals a lowered median value of atomic contacts made to adenine in all four clusters when compared to the median number of atomic contacts guanine in small Ras-like GTPases receive. This observation suggests that the transition from a guanine-binding architecture to one that binds adenine involved an additional change that reduced the overall atomic contacts made to the purine ring in the adenine-binding environment. The decline in overall atomic contacts made to adenine could have consequences in its stabilization within the nucleotide-binding pocket, making adenine loosely held in the pocket compared to guanine in small Ras-like GTPases. In fact, this could partially explain the cause of lowered nucleotide affinity in P-loop ATPases when compared to small Ras-like GTPases that require GEFs to remove the bound GDP.

Comparing the median number of atomic contacts within the adenine clusters we can observe that the adenine receives similar overall atomic stabilization with values ranging from 24 to 26 in all the different positions. This suggests that, despite the variation in purine positioning among P-loop ATPases, they maintain a similar number of contacts to the purine. However, there is considerable variability in the minimum and maximum number of contacts in clusters 1, 2, and 3. Nonetheless, most atomic contacts fall within

the range of 17-30, as evident from the lines depicting the 75th and the 25th quartile of the distribution.

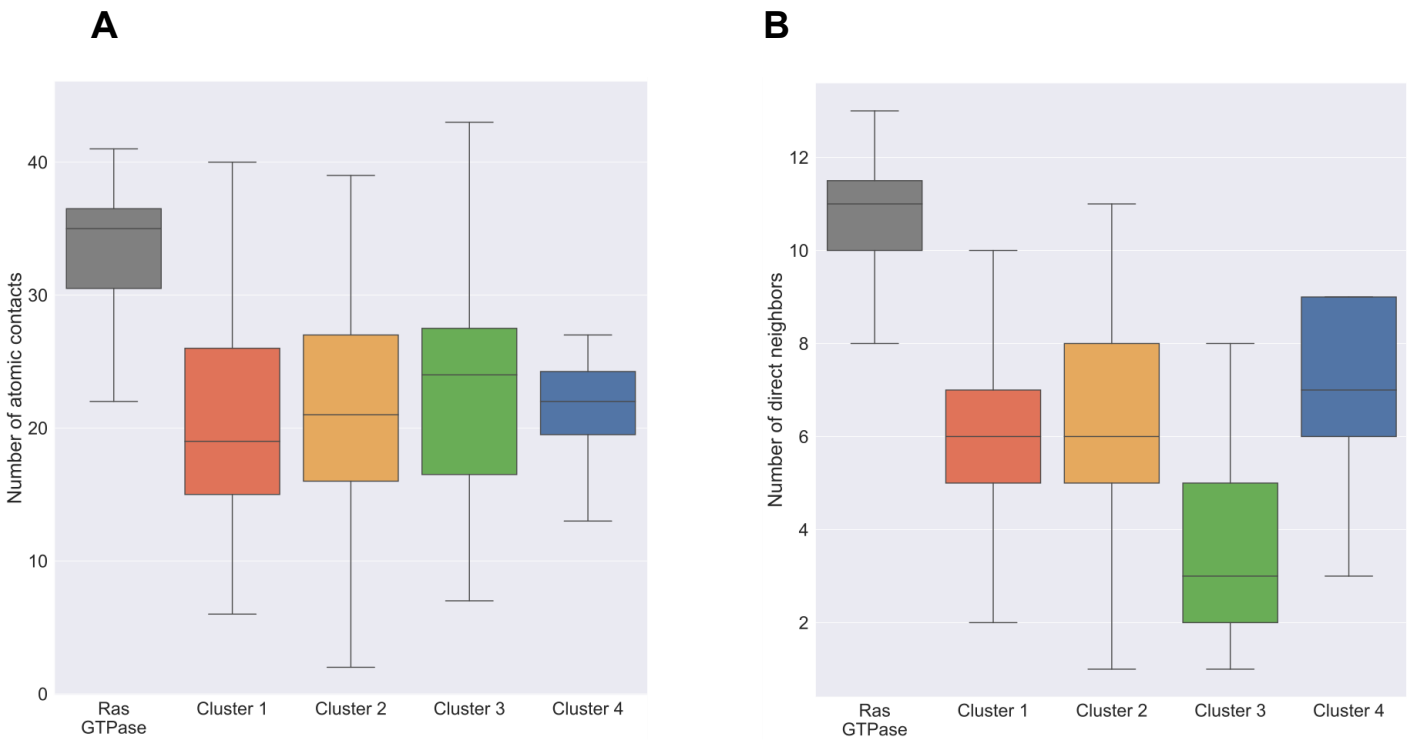


Fig. 5.5. Analysis of the purine environment in P loop ATPases: Variation across clusters based on number of contacts and direct neighbors.
A. The number of contacts made to purine in each PDB of a given cluster is identified and the distribution is plotted as a box plot. The median is represented by a line in the middle of the box, the top and bottom lines represent the 75th and 25th quartile, respectively, and the whiskers represent the standard deviation of the distribution. Each boxplot is colored according to the coloring in Fig 5.3.A. **B.** The number of direct neighbors to the purine in each PDB of a given cluster is also plotted as a boxplot. The distribution is shown in a similar way to Fig 5.5.A.

Examining the number of direct neighbors to the adenine (i.e., residues within 4 Å of each adenine atom) from Fig 5.5.B reveals variability across clusters, with values ranging from 5 to 9 residues. Fewer residues take up the responsibility of stabilizing the adenine when compared to guanine in small Ras-like GTPases. Therefore, despite the similarity in the overall number of atomic contacts to adenine across clusters, the responsibility for stabilization is distributed among a variable number of residues (and water molecules). The number of direct neighbors contacting the adenine is important because it determines the resistance of P-loop ATPases to mutations affecting nucleotide binding. When the responsibility of stabilizing the purine is taken up by a limited number of residues, the impact of a single-point mutation to these residues is

much greater. For example, in cluster 3, where the number of residues contacting the adenine is the lowest but the number of atomic contacts is the highest, a single mutation to one of these crucial residues could completely abrogate the purine stabilization and nucleotide affinity.

5.3.3 Consequences of altered purine positioning on stabilization of adenine atoms

Fig 5.6 depicts the variation in the overall atomic contacts made to the 6 vs 5-membered rings of purine across the clusters. The percentage of atomic contacts made to purine's 6-membered and 5-membered rings is calculated for each PDB in a given cluster by dividing the number of contacts made to each ring by the total number of contacts made to the purine [(x/(x+y)) * 100 where x is the number of contacts made to the 6-membered rings and y is the number of contacts made to the 5-membered ring]. Using the percentage instead of the raw number of contacts helps to account for potential variations in the number of contacts and neighbors among PDBs within a cluster. Plotting the mean percent of contacts to the 6 and 5-membered rings for all members of a cluster also normalizes the differences in the number of PDBs across each cluster, allowing for comparison between clusters.

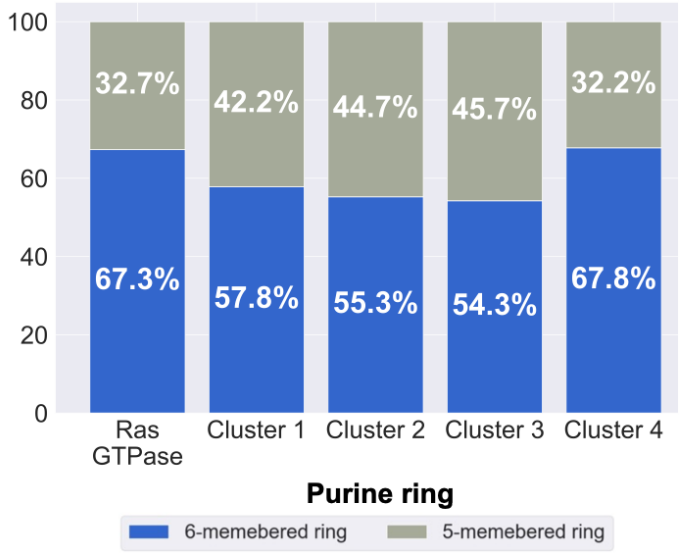


Fig. 5.6. Comparing investments in stabilizing the atoms of 6 vs 5 member ring across clusters

The percentage of atomic contacts made to the 6-membered and 5-membered rings of Adenine is calculated for each PDB in a given cluster. The mean percent of contacts made to Adenine's 6 and 5-member rings by members of each cluster is plotted. The blue bars represent the mean percentage of contacts made to the 6-membered ring, while the gray bars represent the mean percentage of contacts made to the 5-membered ring

The bar graph in Fig 5.6 reveals a reduced relative stabilization of the 6-membered ring of adenine in clusters 1, 2 and 3 when compared to the guanine's 6-membered ring. Cluster 4, which contains P-loop ATPases that show purine inversion is most similar to small Ras-like GTPases in how they invest in the overall stabilization of the 6-membered ring over the 5-membered ring. This suggests that the strategy of switching the orientation of 6 and 5-membered rings allowed these ATPases to retain the feature of prioritizing 6-membered ring stabilization over the 5-membered ring. The similarity in overall atomic investment to the 6 vs. 5 membered rings of adenine in the other clusters means that a difference in the positioning of purines across these clusters does not necessarily result in a significant reallocation of contacts to either of the rings as a whole.

The results suggest that preserving the relative placement of the 6 and 5-membered rings in clusters 1, 2 and 3 seems to have led to a compromise on the stabilization of the 6-membered ring compared to the 5-membered ring. On the other hand, ATPases that choose to invert this orientation of the purine rings (Cluster 4) continue to exhibit a more pronounced preference for stabilizing the 6-membered ring, similar to Ras GTPases.

The reason for analyzing the percentage of contacts made to the 6 and 5-membered rings was to gain a preliminary understanding of the specificity of the P-loop ATPase for adenine over guanine. At this level of inspection, cluster 4 appears to be more discerning than the other adenine binding clusters as it invests about 67% of all atomic contacts to the 6-membered ring of the adenine. All other clusters seem to invest a comparable number of contacts to both the 6 and 5-membered rings. The extent to which the greater investment in stabilizing the 6-membered ring by P-loop ATPases of cluster 4 results in more specificity for adenine will depend on the specific atoms of the

6-membered ring that are contacted and the nature of this interaction. These aspects will be investigated further in the subsequent analysis.

In all four adenine clusters, there is a reduced emphasis on stabilizing the discriminatory atoms of the 6-membered ring when compared to small Ras-like GTPases (Fig 5.7.B). This is apparent as the overall number of atomic contacts made to the discriminatory atoms drops significantly in the adenine clusters compared to the small Ras-like GTPase. There seems to be a greater focus on stabilizing the non-discriminatory atoms in the 6-membered ring (C2, N3, C6) in the adenine clusters when compared to the guanine in small Ras-like GTPase. This observation suggests that the P-loop ATPases may be less selective in their ability to distinguish between adenine and guanine nucleotides, when compared to small Ras-like GTPases. It can be inferred that the modifications in the guanine-binding environment to accommodate adenine may have also led to a reduction in discriminatory abilities.

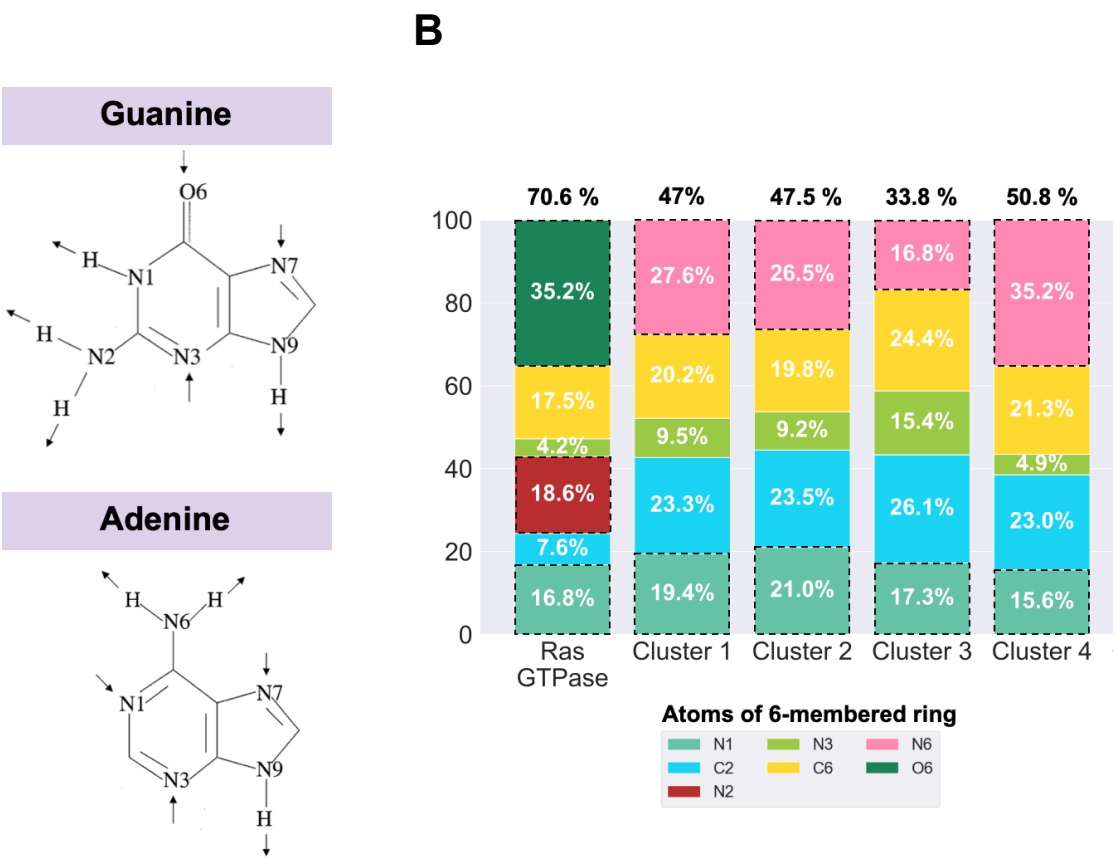


Fig. 5.7. Investments in atom-wise stabilization of 6-membered ring of purine

A. Schematic of adenine and guanine. The outward arrows in the diagram represent the ability of the purine atoms to donate hydrogen ions, while the inward arrows represent the ability of purine atoms to accept hydrogen ions. **B.** For each PDB the contacts made to each atom in the 6-membered ring (N1, C2, N3, C6, N6/O6, N2) is calculated as a percentage of contacts made to the 6-membered ring as a whole. The mean percent of contact made to each atom of the 6-membered ring in a given cluster is plotted. The bars corresponding to the discriminatory atoms - N1, N2, N6 and O6 is enclosed within dashed box. The number on top of the stacked bar for each cluster shows the total percentage of atomic contacts made to the discriminatory atoms.

Comparing adenine clusters, we can observe that clusters 1 and 2 display a similar degree of atomic stabilization to the 6-membered ring, with the discriminatory nitrogens (N1, N6) getting much more stabilization when compared to the non-discriminatory nitrogen (N3). In these clusters, the non-polar atoms also receive substantial atomic contact, but the emphasis seems to be on stabilizing the polar nitrogens.

In contrast, cluster 3 prioritizes the stabilization of the non-polar carbon atoms of the 6-membered ring over the discriminatory nitrogen atoms. This is supported by the observation that half of the contacts made to the 6-membered ring are to the carbons, while the remaining half is distributed among the three nitrogen atoms. Even among contacts made to the polar nitrogens, the non-discriminatory N3 gets greater investment in this cluster when compared to clusters 1 and 2. This greater stabilization of N3 in cluster 3 seems to come at the expense of reduced stabilization of the discriminatory nitrogens N1 and N6.

Cluster 4, which contains P-loop ATPases exhibiting purine inversion, exhibits maximum investment in stabilizing the N6 atom compared to all other clusters.

Taken together, these results suggest that P-loop ATPases with adenine in the orientation of cluster 4 may be more discriminatory than other P-loop ATPases, with those in cluster 3 appearing to be the least competent at distinguishing between purines. However, this conclusion is based only on the percent atomic contacts, and further analysis is necessary to explore the nature of the interactions (van der Waals, hydrogen bonds) and the distance between atoms (fewer contacts at a shorter distance could be as stabilizing as many contacts at a greater distance) for proper reasoning.

Nature of interactions with the adenine atoms

Figure 5.8 demonstrates that adenine clusters show a higher utilization of van der Waals interactions compared to Ras GTPases. This observation aligns with the previous section, which suggests that P-loop ATPases have a reduced discriminatory ability relative to small Ras-like GTPases.

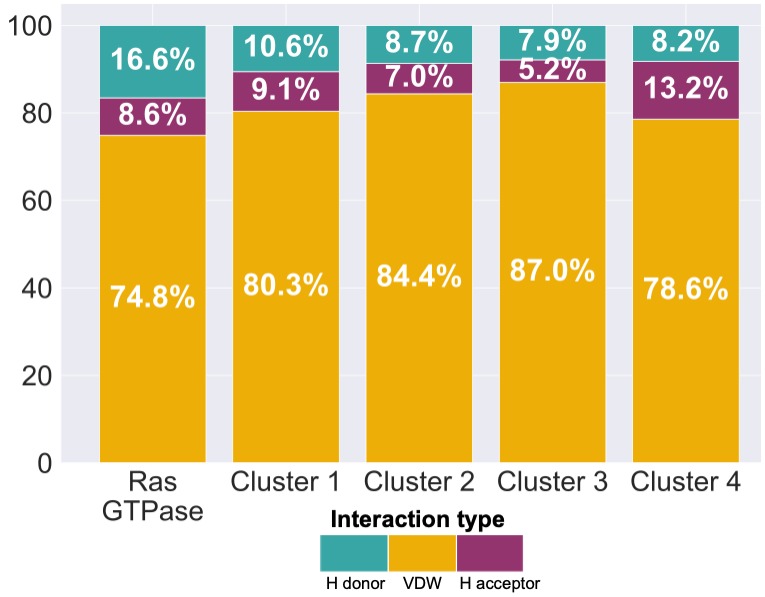


Fig. 5.8. Nature of interactions in purine binding environments
The percentage of interaction type (van der waals, hydrogen donor, and hydrogen acceptor) made to adenine was calculated. These percentages are visualized as a stacked bar plot.

In guanine, three atoms (N1, N2, and N9) can receive stabilization from hydrogen-accepting entities, while N1, N3, and O6 engage in interactions with hydrogen-donating entities. Despite having the same number of atoms that can interact with hydrogen donors or hydrogen acceptors, Fig 5.8 reveals that the guanine binding environment in small Ras-like GTPases is dominated by interactions from hydrogen donors.

In adenine, three atoms (N1, N3, N7) can participate in hydrogen bonding with hydrogen-donating entities, while two atoms (N6 and N9) interact with hydrogen-accepting entities from the P-loop NTPase. Analyzing the overall percentage

of contacts received from hydrogen bonding interactors from Fig 5.8, we find that adenine binding environments in clusters 1 and 2 show comparable involvement of both hydrogen donating and hydrogen accepting interactions. However, in cluster 4, hydrogen acceptor interactions dominate.

Fig 5.9 examines the type of interactions made by the P-loop NTPases with each atom in the purine ring, and allows us to compare the behavior of the different clusters.

Fig. 5.9. Nature of atomic interactions in purine binding environments
The percentage of interactions type (van der Waals, hydrogen donor, and hydrogen acceptor) made to each adenine atom was calculated. These percentages are visualized as a stacked bar plot. Non-polar atoms of adenine (C2, C4, C5, C6, and N6) were excluded from the plot as they only receive van der Waals contacts. The bars represent the polar atoms that can receive both van der Waals and hydrogen bonding interactions.

From Fig. 5.9, we can observe that the guanine in small Ras-like GTPases is stabilized such that 40% of contacts made to the discriminatory N1 and N2 comes from H acceptors. Stabilization of O6 involves about 50% H donating interactions. It is interesting to note that the non-discriminatory atom N3 receives significant H-based

stabilization while the other non discriminatory atom - N9, receives only van der Waals based stabilization.

When comparing guanine and adenine clusters, it is interesting to note that N9 starts receiving hydrogen-based stabilization in adenine clusters. The stabilization of N7 appears to be similar in both guanine and adenine clusters. However, N3 receives fewer hydrogen bonding-based interactions in adenine clusters compared to guanine clusters.

All adenine clusters share the similarity that about 50% of all contacts made to N6 is H based. The clusters do show differences in the mode of stabilizing the other discriminatory atom - N1. Clusters 1 and 2 devote almost one-third of atomic contacts made to N1 to be Hydrogen bond based. Cluster 4 seems to be investing the least in an H-based stabilization of N1. Fig 5.9 reveals that despite the greater overall investment in H bonding interaction cluster 4 shows, these interactions appear to be concentrated on the atoms of the 5-membered ring (N7 and N9), which are not specific to adenine and also present in guanine. That is, the greater investment in H bonding interactions does not translate into ATPases in cluster 4 being more discriminatory when compared to the other clusters.

5.3.4 Variation in core, insertion, water involvement in stabilizing adenine

Fig 5.10 reveals that insertions dominate over the core in terms of adenine stabilizing environment in all clusters. This observation suggests that instead of point-mutating the existing Ras-like architecture, P-loop ATPases more often evolve insertions to stabilize the adenine ring.

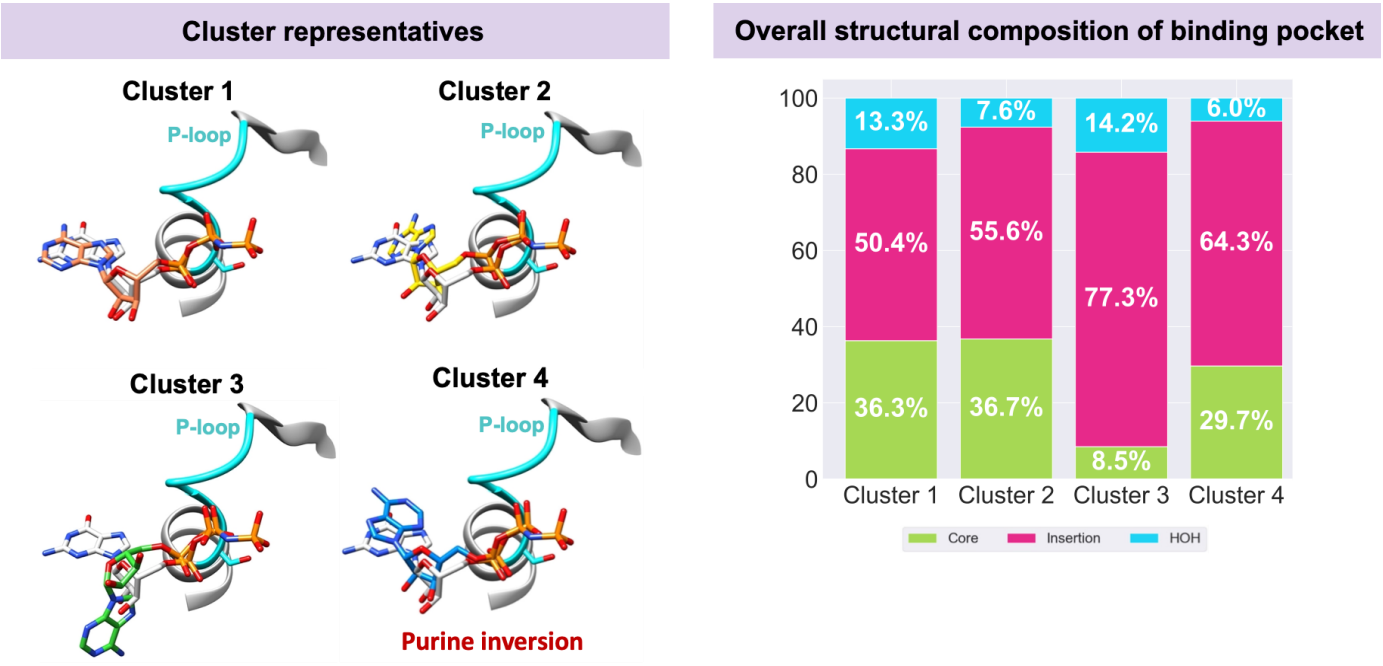


Fig. 5.10 Overall utilization of core or insertion architecture in the adenine environment across clusters

The percent of atomic contacts originating from core, insertion, water (HOH) is calculated for every PDB in a cluster. The mean percent of contacts is plotted for each cluster.

The clusters that position the purine plane similarly (that is, similar $theta_{purine}$ values), such as clusters 1 and 2 utilize the same amount of core architecture but differ in their allocation to insertion and water molecules. Cluster 1 relies more heavily on water molecules than cluster 2 to stabilize the purine environment. This suggests that differences in the positioning of the purine can lead to differences in the allocation of stabilizing responsibilities between insertion and water molecules while the contribution from the core remains similar.

Cluster 3 ATPases show the least involvement of core and the maximum reliance on insertion. Even water molecules are utilized more than core to stabilize the adenine in these P-loop ATPases. This cluster has an almost similar purine ribose orientation but a pronounced tilt in the purine plane, relative to guanine in 5P21. Cluster 4, which shows purine inversion, still utilizes the same amount of core compared to clusters 1 and 2. The purine inversion could have allowed the greater utilization of existing architecture in

Ras to stabilize the adenine allowing these P-loop ATPases to avoid a scenario like cluster 3, whose members needed to rely more on insertions and water molecules to fulfill the hydrogen bonding of adenine. That is, having a smaller tilt but different placement of purine relative to ribose (cluster 4) allowed greater utilization of the core architecture when compared to having greater tilt but similar relative placement of purine and ribose (cluster 3).

While Fig.5.10 demonstrated similarities in the overall use of core architecture in clusters 1, 2, and 4 Fig.5.11 point out differences in the atoms of adenine being stabilized by elements of the core.

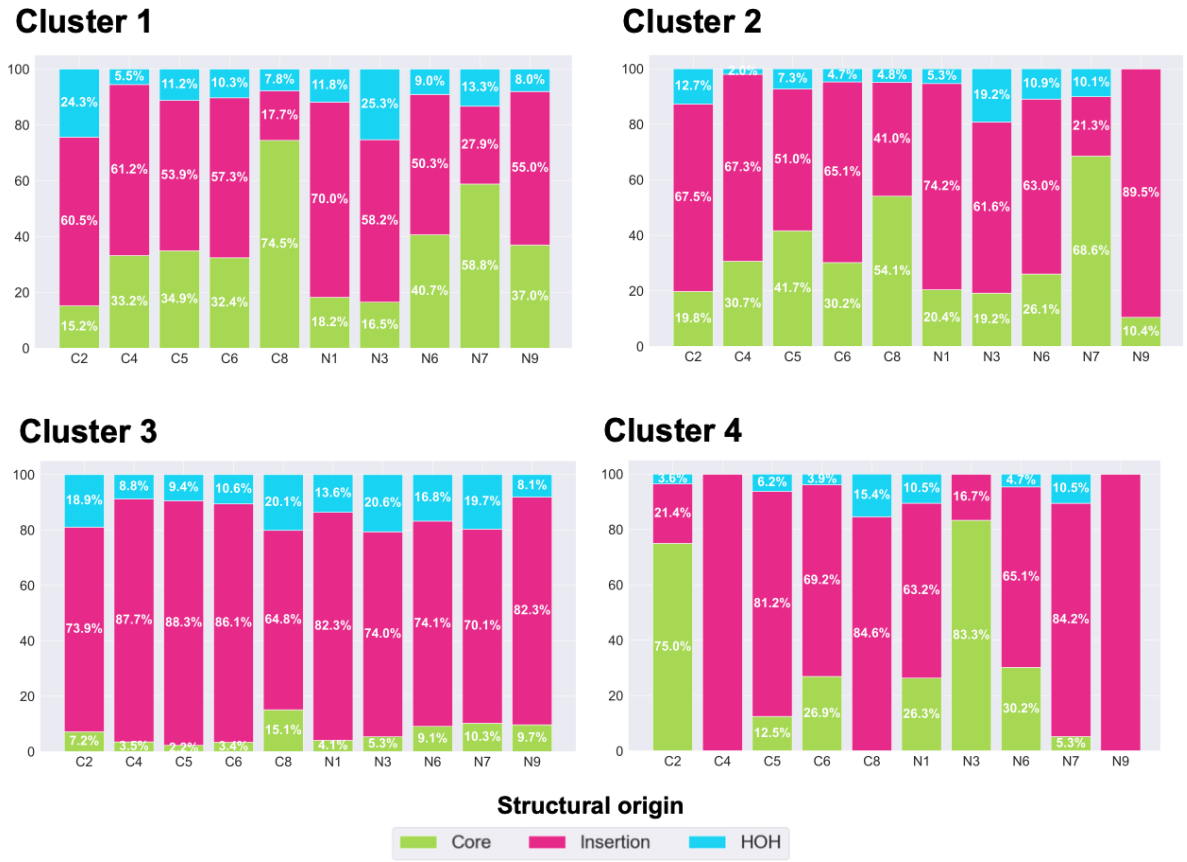


Fig. 5.11. Structural composition of adenine binding environments
The percentage of contacts made to each adenine atom in each PDB of a cluster arising from the core, insertion, and water molecules was calculated. The mean value plotted as a stacked bar plot and colored according to their structural origin

Clusters 1 and 2 follow a trend of utilizing insertions and HOH to stabilize the 6-membered ring while using the core to stabilize the atoms of the 5-membered ring. However, there are some variations among the clusters in terms of the specific atoms within the 5-membered ring that the core stabilizes. Specifically, Cluster 2 uses the core to stabilize N7 of the 5-membered ring, while Cluster 1 uses the core to stabilize N9 of the 5-membered ring.

The atom-wise distribution of contacts made to adenine in Cluster 4 differs from the trend observed in the other clusters. In Cluster 4, the atoms of the 5-membered ring (C4, C5, C8, N7, and N9) are primarily stabilized by insertions and HOH, with minimal involvement of the core. The atoms of the 6-membered ring (C2, N3, N1, C6 and N6), on the other hand, are stabilized predominantly by the core. This suggests that, unlike other clusters, which have evolved insertions to stabilize the 6-membered ring, cluster 4 is able to utilize much more of the existing Ras architecture to stabilize the 6-membered ring as a result of purine inversion.

In the remainder of this section, we will analyze the core and insertion contacts to adenine separately. When investigating core stabilization, we will delve into the secondary structures present in the core that contribute to adenine stabilization. In the case of insertions, we will concentrate on the points within the Ras core where these insertions are made.

Secondary structures within the core that stabilize adenine in P-loop ATPases

Fig 5.12 reveals that all clusters utilize certain secondary structures, such as Loop 11, Helix 1, P-loop, Loop 9, and Strand 5, to stabilize adenine. However, other secondary structures like Loop 3, Strand 4, and Loop 7 only appear in select clusters. Interestingly, the degree to which different secondary structures stabilize the core changes as the tilt of the purine plane increases from the orientation of guanine in Ras-like GTPases. For instance, the P-loop provides greater stabilization as $theta_{purine}$ increases, whereas Helix 1 and Loop 11 offer less stabilization.

Clusters	Purine angular parameters	Equivalent Ras structure	Percentage utilization
1	$Theta_m$ = 14.61 Phi_m = -121.55	Loop 11 Helix 1 Loop 9 Strand 5 P-loop	32.52 24.86 20.47 9.91 7.94
2	$Theta_m$ = 41.59 Phi_m = -151.10	Helix 1 P-loop Loop 11 Loop 9 Loop 7	33.97 25.19 14.5 10.43 5.85
3	$Theta_m$ = 77.10 Phi_m = -124.38	P-loop Loop 3 Helix 1 Loop 11 Loop 7 Strand 5 Strand 4	23.81 21.09 15.65 12.93 8.16 8.16 7.48
4	$Theta_m$ = 19.24 Phi_m = 64.31	Loop 11 Helix 1 Strand 5 P-loop	33.33 29.63 18.52 18.52

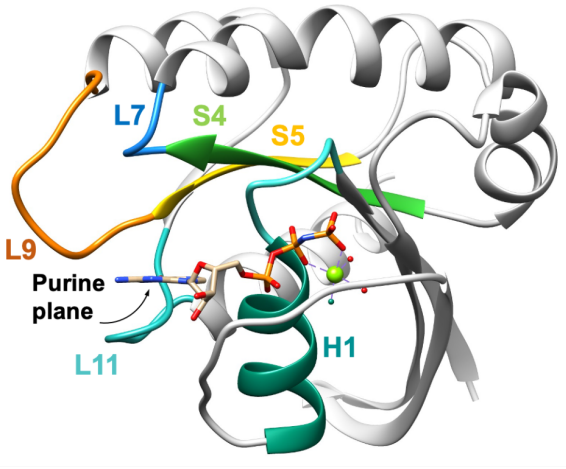


Fig. 5.12. Secondary structure within the Ras core that contacts the adenine in P loop ATPases
The table presents information on the secondary structure within the Ras core of P loop ATPases that contact the atoms of the adenine. The rows are grouped according to the cluster the ATPase belongs to. The percentage utilization column denotes the percentage of contacts made to the secondary structure while considering all contacts coming from the core. That is of 100 contacts made to adenine by the core of a P loop ATPase in cluster 1 32.52 of the contacts would come from Loop 11. This does not mean that 32.52 of all contacts made to the adenine would come from Loop 11. Rather, of all contacts made by the core to the adenine 32.52 of them come from Loop 11. On right, the core secondary structures responsible for stabilizing Adenine are represented within the Ras GTPase structure - 5P21. The secondary structures (H - Helix, L - Loop, S -Strand) are highlighted in the legend's color scheme, while the remaining parts of 5P21 are shown in light gray.

This reveals that even though the overall percent utilization of the core is similar in clusters 1 and 2 the specific secondary structure from the core used to make the contacts differs with altered placements of the purine.

Insertion hotspots that carry adenine stabilizing elements

From Fig 5.13 we observe that N-terminal insertions stabilize the adenine in P-loop the most. Greater the tilt in the purine plane (greater $theta_{purine}$) greater the utilization of N-term insertions to stabilize adenine. C-term regions carry insertions only because of their vicinity to the nucleotide-binding pocket.

Clusters with similar $theta_{purine}$ values seem to have chosen the same points within Ras core to evolve insertions that maximally stabilize adenine as evidenced from clusters 1

and 4 using N-term and Loop 7 for most stabilization, while cluster 2 uses N-term and Loop 9.

Clusters	Purine angular parameters	Equivalent Ras structure	Percentage utilization
1	$Theta_m$ = 14.61 Phi_m = -121.55	N-term Loop 7 Loop 11 Helix 3 Loop 9	45.59 19.39 16.43 13.06 3.57
2	$Theta_m$ = 41.59 Phi_m = -151.10	N-term Loop 9 Loop 7 Loop 11 Loop 3	54.12 15.8 15.21 6.47 2.35
3	$Theta_m$ = 77.10 Phi_m = -124.38	N-term Loop 3 Strand 2 Loop 7 Helix 3 Loop 9	67.86 13.5 5.67 3.13 2.83 2.31
4	$Theta_m$ = 19.24 Phi_m = 64.31	N-term Loop 7 Helix 3 Strand 4 Helix 4	46.15 26.5 12.82 5.98 4.27

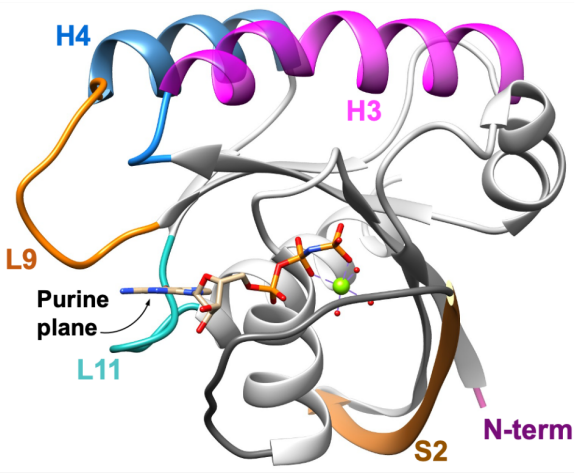


Fig 5.13. Points of insertion within the Ras core that stabilizes the adenine in P-loop ATPases
The table presents information on the secondary structure within the Ras core of P loop ATPasesinto which insertions that stabilize atoms of the adenine are made in P-loop ATPases. The rows are grouped according to the cluster the ATPase belongs to. The percentage utilization column denotes the percentage of contacts made to the secondary structure while considering all contacts coming from the insertion. That is of 100 contacts made to adenine by the insertions of a P loop ATPase in cluster 1 19.39 of the contacts would come from an insertion into Loop 7. This does not mean that 19.39 of all contacts made to the adenine would come from insertions in Loop 7. On right, the insertion points in the Ras core responsible for stabilizing Adenine are represented in the Ras GTPase structure (5P21), and their corresponding secondary structures are highlighted in the legend's color scheme. The remaining parts of 5P21 are shown in light gray.

Similar to the trend from stabilization from the core, the extent to which insertions stabilize adenine atoms changes with $theta_{purine}$. This becomes apparent while examining the extent of adenine stabilization offered by Loop 11 dropping from 16.43 to 6.47 while comparing clusters 1 and 2. Other clusters seem to have completely lost stabilization from insertions originating from Loop 11.

Residues involved in the recognition of adenine

This section aims to investigate the residues comprising the atomic environment of adenine within four identified clusters. The primary objective is to decipher the stabilizing elements and determine the percentage of contacts they constitute, illuminating the degree to which these residue types stabilize adenine in each cluster.

Cluster 1

Residue type	Percentage
Main chain	20.25
PHE	16.14
ARG	15.56
TYR	7.55
GLN	6.22
THR	5.16
ASN	3.87
LEU	3.79
ILE	3.64

Cluster 2

Residue type	Percentage
Main chain	27.98
TYR	11.86
ARG	9.02
PHE	8.11
ILE	7.35
PRO	6.39
VAL	6.03
ASN	5.37

Cluster 3

Residue type	Percentage
Main chain	51.43
ILE	16.67
LEU	10.24
PRO	5.24

Cluster 4

Residue type	Percentage
Main chain	40.94
ARG	25.15
THR	11.69
ILE	4.09

Table 5.1 Residue composition in adenine binding environment across clusters
Table displays the percentage composition of residues in five clusters of adenine binding environments. The table for each cluster shows the residue type in column 1 and the corresponding percentage of atomic contacts made by that residue type in column 2. The residues are sorted in descending order of their contribution to atomic contacts, and only those with a cumulative sum of about 85% are included. The percentage values represent only the contacts made by residues and exclude those made by water molecules (HOH). This approach allows for a more specific examination of the residues that play a crucial role in stabilizing the adenine moiety within each cluster of adenine binding environments. A,B,C,D,E represent the residue composition table for clusters 1-4.

All clusters show at least 20% of contacts from the main chain, with the percentage increasing in clusters that have a reduced contribution from charged polar residues. Despite the scarcity of contacts from polar residues in these environments, results from section 5.3.3 show that the adenine atoms continue to receive H-bonding interactions, suggesting that the main chain takes up the responsibility of H-bonding interactions in

these specific adenine positions. This is most apparent in Cluster 3, where there are no charged or polar residues that make significant side-chain contacts. Therefore, in this cluster, the responsibility of H-bonding is taken up entirely by the main chain atoms and HOH molecules.

Residues like ARG, ASN, and TYR appear in most clusters, while other residues like LEU, ILE, VAL, GLN, and TRP appear in only a few clusters.

5.4 Conclusion

- The shift from a guanine-binding environment to one that recognizes adenine doesn't seem to be a single concerted event as we do not see just a single way that the adenine is positioned within the pocket. There seem to be multiple independent events that allowed the conversion of the guanine binding environment to adenine binding by choosing the extent of involvement of core, insertion, and water in the adenine stabilization.
- All four strategies exhibit a reduced stabilization of the purine relative to the guanine in small Ras GTPases
- Effect of varying $theta_{purine}$ but not phi_{purine} (ATPases of cluster 3) led to the least utilization of core architecture and maximum reliance on insertion.
- Effect of varying phi_{purine} but not $theta_{purine}$ (ATPases of cluster 4) led to purine inversion, which allowed utilization of existing Ras architecture to stabilize the 6-membered ring of adenine, unlike other purine positioning strategies that utilized the core to stabilize the 5-membered ring
- Transitioning from an architecture that recognized guanine to one that recognizes adenine involved a transition from specific residue motifs that made side chain-based interactions in guanine, such as the NKxD and SAK motifs, to a greater emphasis on main chain interactions. The main chain interactions could be involved in both H-bonding stabilizations and van der Waals interactions.

Chapter 6

Discussion

P-loop NTPases are a class of enzymes that play crucial roles in powering muscle contraction, regulating cell growth, controlling cell division and chromosome segregation, and catalyzing metabolic reactions involving phosphate transfer. Initiation of hydrolysis in these NTPases requires stimulatory fingers like Arg, Lys, Asn, and residues that stabilize the catalytic water. In the simplest members of P-loop NTPases – the small Ras-like GTPases, the stimulatory residue is most often a singular Arg residue, and the catalytic water molecule is stabilized by a glutamine residue. In small Ras-like GTPase, the catalytic site is incomplete, and it requires the GAP to provide one or both of these residues. The contact analysis at the faces of the GTPase revealed that across subfamilies, the GAP functioned by appropriately positioning the stimulatory Arg finger with respect to the gamma phosphate of the GTPase. In some cases, the GAP provides the catalytic water-stabilizing glutamine as well. This is the case despite variations in the overall fold of GAPs.

The P-loop NTPases are hydrolytically competent in that they do not require additional activating proteins to hydrolyze the NTP to NDP. The active site analysis points out that this circumvention of GAP in P-loop NTPases happened by either point-mutating parts of the existing Ras-like GTPase core or evolving insertions that carry the stimulatory residue. The choice of where the stimulatory residue comes from appears to be influenced by family relatedness. Families like G-proteins and Motor proteins harbor their stimulatory residues within the core and thus have resorted to point-mutating the existing Ras-like GTPase architecture to become catalytically competent. Families that are more distantly related to Ras-like GTPases, such as the members of the ASCE division, carry their stimulatory residues within the insertions and thus seem to have evolved insertions instead of utilizing the Ras architecture.

It is worth noting that not requiring an explicit activating protein to complete the NTPase cycle can be seen as the cell relinquishing a degree of control over the enzyme. Having the stimulatory residue within the NTPase comes with the risk of unnecessary self-activation that could deplete the NTP reserves in the cell. One possible way of avoiding this scenario is by introducing stimulatory residues within the NTPase that act in trans. This ensures that the NTPase is catalytically active and can hydrolyze NTP only when an oligomeric assembly is formed. This is common in P-loop NTPases, as evidenced by numerous members within the SIMIBI, RecA, and AAA+ families, which receive the stimulatory finger in trans (section 4.3.4) and hence are catalytically active only in their oligomeric state. The nature of the oligomeric assembly can be determined based on the part of the Ras core from which the stimulatory residue comes. If the trans-acting stimulatory residue comes from regions near the nucleotide-binding pocket of the neighboring subunit, then a dimer would be the catalytically active state. In such a dimer, there would be reciprocal activation, with one subunit providing the other with the stimulatory fingers. In cases where the stimulatory residue comes from parts away from the nucleotide-binding pocket, a higher-order oligomer will be the active state. The mode of activation wouldn't be reciprocal but would be such that each subunit, say x, provides the stimulatory residue to one of the neighbors, x+1, while receiving stimulation from another subunit, x-1.

Adopting the nomenclature of faces in the minimal GTPase-like core from Chapter 2 (Fig 2.2) would mean that if trans stimulatory residue originates from Face 2 of the neighbor, then a reciprocally activated dimer would be the catalytic state. When the trans stimulatory residue is part of other faces, then a higher-order oligomeric state would be the active form of the NTPase.

There is currently no evidence of the existence of Purine exchange factors, similar to GEFs in Ras GTPases, in members of the P-loop NTPase family. It is plausible that these ATPases have assimilated a domain that functions akin to the GEF in small Ras-like GTPases. The mechanistic analysis of GEFs in small Ras-like GTPases, as

discussed in Chapter 2, revealed that the most common mode of action in these GEFs involved disrupting the stabilization at the phosphate. Due to the lower affinity of GTPases for the GEF when bound to GTP, there is no instance of GEF interfering with GTP binding and hydrolysis in these enzymes.

If P-loop NTPases do not require an explicit GEF partner because they have incorporated a GEF domain within themselves, it could create issues of NTP binding in the NTPase as this domain could obstruct the Mg binding site or reposition the P-loop even when NTP must be bound. One possibility is that the GEF domain would be mobile and would contact the nucleotide-binding pocket only when the site has NDP. The analysis also revealed that, while less common, GEF activity could be concentrated at the purine-binding region of the nucleotide-binding pocket. Therefore, it is possible that the absence of a GEF in P-loop ATPases involves modifying the topology at the purine-binding end of the nucleotide-binding pocket instead of the phosphate-binding end. This modification of the topology at the purine binding site could lead to a reduction in adenine stabilization and, thus, lead to an intrinsically lower nucleotide affinity.

This phenomenon appears to be widespread among P- loop ATPases, as none of the adenine molecules sit in the exact same manner as the guanine molecule in Ras-like GTPases. We also observed that there seems to be an overall reduction in the number of atomic contacts adenine receives in P-loop ATPases when compared to guanine in small Ras-like GTPases. Work in progress investigates the guanine positioning in P-loop GTPases, and preliminary findings suggest that the guanine molecule in these GTPases is not positioned in the same way as the guanine in Ras GTPases. The altered purine positioning relative to Ras-like GTPases results from the change in topology at the purine binding end, and one consequence may be a weakened pocket that no longer binds the purine molecule as firmly, thus eliminating the need for a GEF in these P loop NTPases to remove the NDP. Furthermore, the modified purine positioning in ATPases has the added benefit of stabilizing the discriminatory atoms of the adenine molecule using pre-existing architecture elsewhere in the core. In conclusion, the absence of GEF in P-loop ATPases may be attributed to the

incorporation of a domain that functions similarly to a GEF in small Ras-like GTPases or the alteration of the topology at the purine binding end of the nucleotide-binding pocket, which reduces the stabilization and avoids the need for a GEF.

Ongoing projects

- **Guanine binding environment**

An ongoing analysis, similar to the one presented in Chapter 5, is being conducted to investigate the guanine binding environment of P-loop GTPases. Preliminary findings indicate that the positioning of guanine in P-loop GTPases differs from that in small Ras-like GTPases. This suggests that other P-loop GTPases have undergone modifications in their guanine binding environment compared to small Ras-like GTPases. The objective of this analysis is to unveil how guanine recognition has evolved in P-loop NTPases, specifically identifying which aspects of the Ras architecture are conserved and which elements of guanine recognition have been lost. These modifications may influence the affinity of P-loop GTPases, shedding light on why they do not require GEFs.

References

1. 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, 309

2. 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, 1345.

3. Prior, I.A.; Lewis, P.D.; Mattos, C. A comprehensive survey of Ras mutations in cancer. *Cancer Res.* **2012**, *72*, 2457–2467

4. Wey, M.; Lee, J.; Jeong, S.S.; Kim, J.; Heo, J. Kinetic mechanisms of mutation-dependent Harvey Ras activation and their relevance for the development of Costello syndrome. *Biochemistry* **2013**, *52*, 8465–8479

5. Bos, J.L.; Fearon, E.R.; Hamilton, S.R.; Verlaan-de Vries, M.; van Boom, J.H.; van der Eb, A.J.; Vogelstein, B. Prevalence of ras gene mutations in human colorectal cancers. *Nature* **1987**, *327*, 293–297

6. Wittinghofer, A.; Vetter, I.R. Structure-function relationships of the G domain, a canonical switch motif. *Annu. Rev. Biochem.* **2011**, *80*, 943–971

7. Anantharaman, V.; Aravind, L.; Koonin, E.V. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr. Opin. Chem. Biol.* **2003**, *7*, 12–20.

8. Walker, J.E.; Saraste, M.; Runswick, M.J.; Gay, N.J. 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. *EMBO J.* **1982**, *1*, 945–951

9. Saraste, M.; Sibbald, P.R.; Wittinghofer, A. The P-loop—A common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **1990**, *15*, 430–434.

10. Rees, D.C.; Johnson, E.; Lewinson, O. ABC transporters: The power to change. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 218–227

11. Oldham, M.L.; Chen, J. Snapshots of the maltose transporter during ATP hydrolysis. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15152–15156

12. Kamerlin, S.C.; Sharma, P.K.; Prasad, R.B.; Warshel, A.Why nature really chose phosphate. Q.Rev.Biophys.2013,46,1–132.

13. Langen, R.; Schweins, T.; Warshel, A. On the mechanism of guanosine triphosphate hydrolysis in ras p21 proteins. Biochemistry **1992**, 31, 8691–8696.

14. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research **2003**, 13, 2498–2504.

15. Pandit, S.B.; Bhadra, R.; Gowri, V.; et al. SUPFAM: A database of sequence superfamilies of protein domains. BMC Bioinformatics **2004**, 5, 28\

16. Leipe, D.D.; Wolf, Y.I.; Koonin, E.V.; Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **2002**, *317*, 41–72

17. Leipe, D.D.; Koonin, E.V.; Aravind, L. Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* **2003**, *333*, 781–815.

18. Iyer, L.M.; Leipe, D.D.; Koonin, E.V.; Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **2004**, *146*, 11–31.