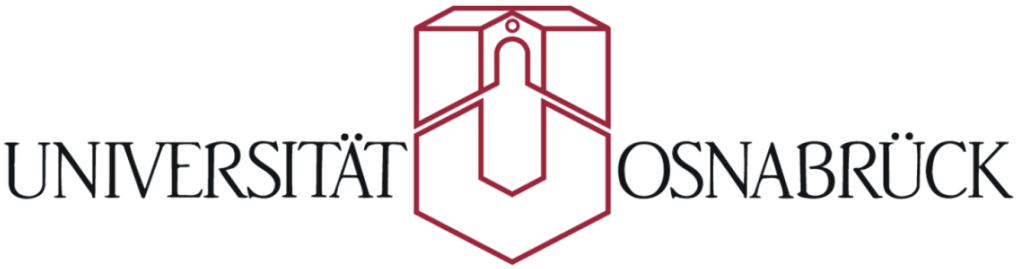


# Molecular Modelling of Monovalent Cations in Energy-Converting Proteins

**Dissertation**

Presented to
the Department of Physics,
University of Osnabrück,
in partial fulfillment of the requirements for the degree of

*"Doctor Rerum Naturalium"*

**Daria N. Shalaeva**

Osnabrück
2021

*"Life is the mode of existence of protein bodies, the essential element of which consists in continual metabolic interchange with the natural environment outside them, and which ceases with the cessation of this metabolism, bringing about the decomposition of the protein."*

**Frederick Engels, "Dialectics of Nature", 1883**

# *Contents*

# 1. Introduction

The elemental composition of living cells is very similar across different tissues and species but differs drastically from the elemental composition of their usual habitats [1, 2]. The presence of particular elements in the living cells is determined by their physicochemical properties. Life, as we know it, rests on the construction of polymers, so the cells are composed primarily of atoms that form stable covalent bonds at temperatures that occur on Earth. Polymers composed of carbon, nitrogen, and oxygen are accompanied by ions that participate in the management of electric charges and redox states in the cell [3]. Phosphate-containing molecules are universally used as an energy depot due to their ability to form stable high-energy bonds, hydrolysis of which can be triggered by proteins [4]. Finally, sulfur atoms are also widespread in cellular molecules owing to their abundance on Earth and ability to act as potent electron transfer carriers [5].

The abundant presence of negatively charged molecules (particularly with carboxyl and phosphate groups) in cells demands positively charged groups or ions to balance the overall charge of the cytoplasm. In extant living cells, this role is primarily fulfilled by potassium ions. Omnipresent divalent metal cations are primarily designated to more chemically specific roles: $Mg^{2+}$ often acts as a co-factor in reactions involving nucleotides, $Zn^{2+}$ - as a co-factor in many enzymes and regulatory proteins, transition metals ($Fe^{2+}$, $Mn^{2+,}$ and $Ni^{2+}$) are crucial for many redox reactions [6]. $K^+$ and $Na^+$ ions serve as cofactors for several enzymes as well, but despite their chemical similarities and omnipresence in living organisms, all cells maintain a high concentration of $K^+$ inside the cytoplasm, while $Na^+$ ions are continuously expelled into the cell bathing fluids (see [7, 8] and references therein). This is especially puzzling since $Na^+$ ions are much more abundant than $K^+$ ions in most aquatic environments and on Earth in general [2]. Specific preference of cells for $K^+$ over $Na^+$ was ascertained as early as 1926, when Archibald Macallum noted that while there are similarities in the total ion concentrations between cellular fluids and the sea water, concentrations of the specific ions differ between the cell cytosol and outside fluids, such as sea water or organismal fluids (e.g. blood and lymph) [2].

Thus, most cells maintain a concentration of the $K^+$ ions higher than 100 mM, all while the surrounding media has a similar or higher concentration of $Na^+$ ions [2]. Both cations leak through the membrane into the cell owing to the negative charge of the cytoplasm as compared to the extracellular phase, but larger $K^+$ ions have lower desolvation penalty so they leak faster. Thus, the cell can maintain the $K^+/Na^+$ disequilibrium by just actively and selectively expelling $Na^+$ ions out of the cell [9]. This process must be performed by transporters that can provide enough energy to push the ion against concentration gradients and can discriminate between $K^+$ and $Na^+$ ions. The

average cell spends up to half of all the energy it generates on maintaining concentration gradients of the monovalent cations [10].

Potassium involvement in the translation system is often credited as one of the main reasons for a cell to accumulate these cations [7]. In addition to the translation, many other cellular processes require monovalent cations and exhibit a strong preference for $K^+$ and $NH_4^+$ ions over $Na^+$ ions. One of such processes in the protein folding, as GroEL/GroES complex and Hsc70 complex, both require $K^+$ ions [11, 12]. In the case of GeoEL/GroES $K^+$ ions can be partially substituted by $NH_4^+$ and $Rb^+$, but smaller ions such as $Na^+$ and $Li^+$ are inactive [11]. Another key $K^+$-dependent process is splicing, which again, goes much faster in the presence of larger cations than in the presence of smaller cations [13]. The innate demand for $K^+$ ions of major cellular processes dictates the accumulation of $K^+$ ions in the cell and the expulsion of $Na^+$ ions to the outside medium.

Translation is performed by a ribosome, a ubiquitous large RNA-containing multimolecular assembly that requires specific metal cations to uphold its structure and to support its functions. Both $K^+$ and $Mg^{2+}$ ions contribute to the stability and function of the ribosome, as well as other RNA structures (see [14] and references therein). The significant difference in physical properties of $K^+$ and $Mg^{2+}$ ions precludes competition between them and together the two metals exhibit a more pronounced synergistic effect on the RNA stability and activity [15]. Potassium ions bind to ribosomal RNA at specific positions, although only a few of such binding sites have been determined with certainty [14]. The peptidyl transfer center (PTC) of the ribosome, where the formation of peptide bonds occurs certainly requires potassium. A recent study attributed 7 $K^+$ ions to the PTC inner shell and 23 more to the surrounding region [14]. Furthermore, two $K^+$ ions play important role in conformational rearrangements of the decoding center associated with the binding of tRNA [14].

The translation machinery requires the energy of GTP hydrolysis not only for the protein synthesis itself, but also to support and control accompanying processes, such as ribosome assembly and construction of aminoacyl-tRNAs. GTP hydrolysis for such needs is performed by various GTPases, many of which were shown to be $K^+$-dependent. One such case is the GTPase MnmE, which contributes modifications to certain tRNA molecules [16]. Potassium dependence was shown also for many GTPases associated with ribosome assembly and maintenance, such as Era, Nug1, RbgA, and others [17]. Furthermore, translation factors, which are directly involved in the translation also show increased activity when $K^+$ ions are present, even when the ribosome is absent [18-21]. The majority of these proteins are ubiquitous across all species [22, 23].

Translation factors, as well as other GTPases catering for the ribosomal machinery, belong to a large superfamily of P-loop (phosphate-binding loop) nucleotide triphosphatases (NTPases) [24]. Proteins of this superfamily catalyze ATP and GTP hydrolysis to fuel multiple processes in the cell and can amount to 18% of all cellular gene products [24-26]. The catalytic activity of P-loop NTPases usually requires an introduction of an activating Lys or Arg residue, which is inserted into the active site by another protein, adjacent monomer in the oligomer, or a different domain of the same protein [27]. In $K^+$-dependent P-loop NTPases, the activating $K^+$ ion is located similarly to the positively charged side-chain nitrogen of the Arg/Lys finger [28, 29]. Notably, in most $K^+$-dependent P-loop NTPases $NH_4^+$ ions can functionally replace $K^+$ ions, but $Na^+$ ions cannot [18, 28, 30-40]. Furthermore, the processing of the phosphoanhydride bonds in the absence of any enzymes also has a preference for larger ions, such as $K^+$, $Rb^+$, and $NH_4^+$ over smaller ions, such as $Na^+$ and $Li^+$ [41]. This suggests that the cation size determines, in some way, how monovalent cations interact with the phosphate chain and facilitate its hydrolysis.

In this work, a combination of molecular modeling and evolutionary approaches is used to address the NTP hydrolysis catalyzed by monovalent cations and by the P-loop NTPases. First, the distinctive effect of cations on the NTP hydrolysis is investigated by molecular dynamics (MD) simulations of ATP and GTP in the water in the presence of different monovalent cations, and in the absence of them. Second, two cases of the key $K^+$-dependent proteins are also investigated by MD simulations: tRNA modification GTPase MnmE and the translation factor EF-Tu, accompanied by comparative structure analysis of several other P-loop proteins. Further, the mechanism of NTP hydrolysis with the participation of cations is used as a template to assess the common features of the catalytic mechanism across all classes of P-loop proteins, using both comparative analysis of particular representative structures and statistical survey of all available structures.

Since cells use energy to expel $Na^+$ ions and maintain their concentration gradient, these ions contribute strongly to the electrochemical potential across the cell membrane. The gradient of $Na^+$ ions can be used as an energy source in different processes (see [9] and references therein). Particularly, a plethora of so-called “secondary transporters” facilitate translocation of substrates across the membrane by utilizing the energy of $Na^+$ gradient [42].

Extracellular sodium ions are known to bind to the eukaryotic G protein-coupled receptors (GPCR), particularly of class A [43]. These proteins govern a plethora of cellular processes by triggering the appropriate response to various signaling molecules. It was suggested, that upon activation of the receptor, the $Na^+$ ion could be transported into the cytoplasm by membrane potential, thus fueling the necessary conformational changes [43]. Here, $Na^+$-binding in class A

GPCRs are examined using the evolutionary biochemistry approach. Available structures of the GPCRs in the $Na^+$-bound state were compared with the recently resolved structures of Na-pumping microbial rhodopsins, revealing the shared origin of the two superfamilies and common features in their activation-associated conformational changes. The suggested ability of GPCRs to translocate $Na^+$ ions is supported by further similarities with other heptahelical proteins, transporting ions across the membrane. The ability of GPCRs to translocate $Na^+$ and subsequent dependence of the receptor activity on the membrane voltage is further explored by activation modeling and comparative structure and sequence analysis of class A GPCRs.

# 2. Methodology of the Evolutionary Biophysics Approach

This chapter describes the strategy behind the combination of computational biology and bioinformatic methods applied here towards understanding the fundamental mechanisms employed in P-loop containing nucleoside triphosphatases (P-loop NTPases) and G-protein coupled receptors (GPCRs).

While all organisms function in accordance with the laws of molecular biochemistry, they are also subjects to the laws of evolution. Advances in genomics, protein sequence analysis, and protein structure determination lead to the emergence of evolutionary genomics (phylogenomics) and provided enormous insights into the evolution and structure-function relationships in proteins. Comparative analysis of protein sequences and structures can be used to reveal two types of information. First, it brings forward conserved residues and structure elements in proteins. Features, conserved within entire (super)families of proteins are likely to be crucial for the general structure or function, attributed to that (super)family; and are usually considered to be inherited by the proteins from their common ancestor. Second, comparative analysis of protein (super)families can reveal features that are conserved only within particular subfamilies or subgroups of proteins within one large (super)family. Such features are likely to be associated with more specific functions, underlying the functional diversity of protein families within larger groups. Positions in the protein sequence that exhibit such subgroup-specific conservation must be functionally relevant, as follows from their partial conservation, but not crucial for the main protein function, as follows from their variability between the subgroups. In this work, comparative structure analysis and phylogenomics were applied to both P-loop NTPases and GPCRs.

Advances in X-Ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and cryogenic electron microscopy (Cryo-EM) have made available a plethora of 3D structures of diverse protein and protein complexes. The abundance of experimental structures of different proteins within a particular (super)family makes it possible to identify universal features, common for all proteins in this group. The task of retrieving and analyzing great volumes of data from biological databases is a central component of bioinformatics. For instance, over 7000 structures in the Protein Data Bank depict proteins, containing P-loop domains. Comparative analysis of such volume of data is only possible with a systematic automated approach, fulfilled by bioinformatics tools. However, the design of the protocol for such an approach and interpretation of the resulting data is only possible after manual inspection of a smaller representative set of structures. Together, manual inspection and automatic survey of all available structures reveal features that are ubiquitous and/or conserved among all P-loop proteins, thus establishing the fundamental basic

mechanism of NTP hydrolysis in those proteins. For the GPCRs, only 218 experimental structures are available, which is still an impressive number for membrane proteins that are notoriously hard to crystallize. Since the majority of those structures contain mutated proteins and cover proteins with very high sequence similarity, manual analysis of all relevant structures was possible.

Living cells are comprised of a diverse range of complex and simple molecules involved in intricate interactions with each other. When such molecules are isolated and examined individually, they conform to all the physical and chemical laws that describe the behavior of an inanimate matter, as do all the processes occurring in living organisms. Thus, the very nature of life itself is a complex network of interactions between molecules that obey the physical and chemical laws that govern the universe in general. It may seem that any process in the living cell can be described with a series of equations, corresponding to the appropriate laws of physics and chemistry. And indeed, in many cases, it is so. One of such cases is the method of molecular dynamics (MD) simulations. It applies the basic laws of Newtonian physics and structural biochemistry to describe the structure and physical movements of atoms and molecules. When applied to biological molecules, MD simulations describe internal motions and resulting conformational changes within such molecules, as well as interactions between them, providing insight into various biomolecular mechanisms. Such analysis, however, is only possible when experimental structures can provide a reliable starting point for the simulations. Availability and quality of crystal structures almost always inform the choices of the specific objects for the MD simulations.

In this work, MD simulations were applied to two types of systems: small molecules in water and proteins in complex with other molecules. First, MD simulations of the ATP and GTP molecules in water were performed, in the presence of different monovalent cations. These simulations reveal the effect those cations had on the phosphate chain shape of the ATP and GTP molecules, contributing to the explanation of the effect different monovalent cations have on the NTP hydrolysis in water. Second, MD simulations were performed for two cation-dependent GTPases MnmE and EF-Tu. Comparative analysis of conformations attained by the GTP molecules bound to the proteins in the presence and the absence of cations pointed at the particular role that cations play in the NTP hydrolysis in such proteins. These findings, however, only describe corresponding particular cases, and while they provide an insight into the mechanism underlying NTP hydrolysis, general claims could not be made based on those results alone.

Described here combination of the physicochemical methods, bioinformatic tools, and phylogenetic analysis comprises the evolutionary biophysics approach to the study of protein (super)families. Investigation of the proteins (super)families with this approach provides a deep

understanding of the respective molecular mechanisms together with the context of their emergence and evolution.

## 2.1. Molecular Dynamics Simulations

### 2.1.1. Principles of Molecular Dynamics Simulations

Computer simulations are carried out in the hope of understanding the properties of biomolecules and their assemblies, particularly in terms of their structure and the microscopic interactions between them. Molecular dynamics (MD) simulations and molecular modeling are essential research instruments in biochemistry and cell biology that complement experimental data, guide rational approaches in the design of experiments, provide access to complex data and models, and yield an atomic level understanding of cellular processes. MD is a method of computer simulation for studying the physical movements of atoms and molecules and their behavior in time.

The method of molecular dynamics simulations was first used by Alder and Wainwright in the late 1950s [44, 45] to study the interaction between solid particles, which resembled the behavior of simple liquids. The next milestone in MD development was 1964 when Rahman carried out the first simulation of liquid argon [46]. The first realistic system simulation was done by Rahman and Stillinger in their simulation of liquid water in 1974 [47]. The first protein simulations appeared in 1977 for the bovine pancreatic trypsin inhibitor [48]. The latest advances in MD simulation methods that combine molecular-mechanics and quantum mechanics methods were awarded the Nobel prize in chemistry in 2013 which was granted jointly to Martin Karplus, Michael Levitt, and Arieh Warshel "for the development of multiscale models for complex chemical systems".

Today one routinely finds in the literature molecular dynamics simulations of solvated proteins, protein-DNA complexes as well as lipid systems addressing a variety of issues including the thermodynamics of ligand binding and the folding of small proteins. The number of simulation techniques has greatly expanded, and many specialized techniques aimed at particular problems are available now. Molecular dynamics simulation techniques are widely used to accompany experimental procedures such as X-ray crystallography and NMR structure determination.

During MD simulation the atoms and molecules are allowed to interact for a fixed period of time, giving a view of the dynamical evolution of the system. In the most common version, the

trajectories of atoms and molecules are determined by numerically solving Newton's equations of motion for the system of interacting particles, where forces between the particles and their potential energies are calculated using molecular mechanics force fields.

In MD simulations Newton's equations of motion are solved for a system of N atoms:

(2.1.1) $$F_i = m_i \frac{d^2 r_i}{dt^2}, i = 1 \dots N$$

The forces are obtained from a potential function as the negative derivatives:

(2.1.2) $$F_i = -\frac{dV}{dr_i}$$

Newton's second law of motion describes the acceleration of a particle *i* of mass $m_i$ along a coordinate $r_i$. The force $F_i$ acting upon particle *i* depends on the positions of all other particles in the system which makes the potential energy surface extremely complex and impedes direct integration. Instead, tedious numerical integrations in small time steps are needed, making MD simulations computationally costly.

The forcefield for MD simulations includes the potential energy function equation and all constants in that equation which describe a set of standard biological molecules (e.g. all standard amino acids, solvent molecules, membrane lipids, etc.). The potential function connects the atomic coordinates of a system to the potential energy. While a single particular structure refers to one configuration in the conformational space of the system, the potential function gives a complete description of the entire potential energy surface. The potential function consists of two components, the Bonded terms and the Non-bonded terms (eq. 2.1.3, Figure 2.1.1). Bonded terms are those describing bond, angle, and dihedral interactions, while Non-bonded terms account for distant interactions and describe electrostatic and Van der Waals interactions.

(2.1.3)

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\varphi(1 + \cos(n\varphi - \varphi_0))$$

$$+ \sum_{impropers} k_\psi(\psi - \psi_0) + \sum_{\substack{non-bonded \\ pairs\ (i,j)}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\substack{non-bonded \\ pairs\ (i,j)}} \frac{1}{4\pi\varepsilon_D}\frac{q_i q_j}{r_{ij}}$$

These equations are solved simultaneously with small (on the femtosecond scale) time steps. This process is repeated to describe the system for some time, provided that the pressure and temperature remain at the pre-set values, and the coordinates are saved to an output file at regular time intervals. The main result of the MD simulation is the trajectory - the set of coordinates as a function of time. Provided the system was at equilibrium, the resulting set of coordinates (the

trajectory) represents the ensemble of conformations, attained by the system (e.g. by a protein in a solution).

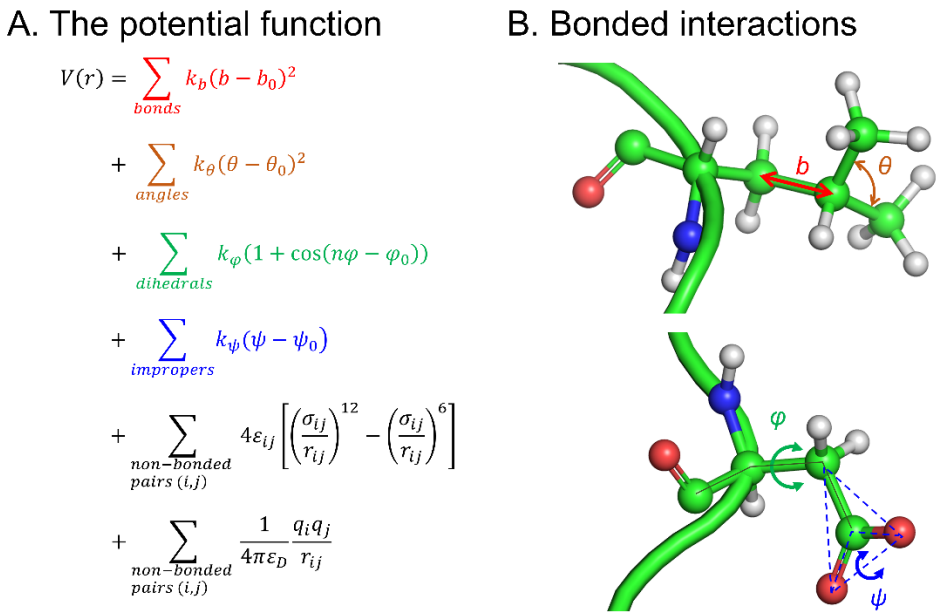


**Figure 2.1.1** Interactions and forces are taken into account in MD simulations calculations. The total potential energy of any molecule is the sum of simple allowing for bond stretching, bond angle bending, bond twisting, van der Waals interactions, and electrostatics. Many properties of a biomolecule can be simulated with such an empirical energy function.

Forcefields provide the means for calculating forces during MD. They are not rigid within the simulation method and the parameters can be modified to include new data or adapt to a specific system or task. Still, the incorporation of particular forcefields in the MD software is subject to limitations. In this study, the CHARMM force field [49] that is incorporated in GROMACS [50] software was used. In version 5.0.15 used here, the force field does not contain fine-tuning of bonded interactions, cannot incorporate polarizabilities, and is pair-additive (except for the long-range Coulomb forces). The number of atoms in the system and the topology (the connectivity between atoms and partial charges) remains unchanged all the time during energy minimization and classical MD simulation.

Before the MD simulation, the object of interest structure must be prepared in the following manner: 3D structure of a molecule(s) under investigation is put in the simulation box providing at least 10Å between the molecule of interest and the edges of the box. This construct is referred to as a unit cell. Since the system size is usually small relative to realistic systems, atoms of the system will have a lot of undesired boundaries with the environment (vacuum). This must be avoided in the simulation of a bulk system. To avoid real phase boundaries periodic boundary conditions were applied. Periodic boundary conditions (PBCs) are typically applied to the simulation box during MD to approximate a large system (realistic solution in case of biological

objects) by using a small part – a unit cell. When a molecule passes through one side of the unit cell, it re-appears on the opposite side with the same velocity. During the simulation, only the properties of the original simulation box need to be recorded and propagated. These conditions allow fast simulations since only a small area of solution around the molecule of interest has to be calculated, instead of realistic volumes of solution.

The unit cell containing molecule(s) of interest is then filled with water molecules (this procedure is called the solvation of the system). Water molecules are placed everywhere in the unit cell except for areas already taken by some atoms. To create realistic water molecule placement additional procedures will be taken at later stages of the simulation. At the next step, some water molecules are replaced by ions, most often $Na^+$ and/or $Cl^-$. Ionization of the system is done to create a realistic concentration of salt in the system and different amounts of cations and anions are added to compensate for the existing charge of the system. The neutral total charge of the system is a crucial requirement for MD simulations.

Prepared in the described manner system is then followed to the energy minimization (EM) procedure, and then for the MD simulation itself. The initial conformation of the system before the MD can contain unnatural geometries, for example, as a result of manual structure editing, crystallization artifacts, automatic solvation, etc., thus the energy minimization (EM) is usually the first step preceding the MD simulation.

GROMACS [50] software provides a simple form of local energy minimization, the steepest descent method. For biologically relevant systems the potential energy function is a complex landscape (or hyper surface) in numerous dimensions. In addition to the one global minimum, a large number of local minima are usually present, where all derivatives of the potential energy function with respect to the coordinates are zero and all second derivatives are nonnegative. Given a starting configuration, it is possible to find the nearest local minimum. Nearest here does not necessarily mean nearest in geometrical terms (i.e., the least sum of square coordinate differences), but the minimum that can be arrived at by systematically following the steepest local gradient.

The *steepest descent* method uses derivative information. As in MD programs the partial derivatives of the potential energy can be calculated with respect to all coordinates this method is very popular. This algorithm repeatedly shifts the system coordinates following the direction of the negative gradient, ignoring the history of previous steps. The step size can be adjusted so that the search goes faster but the downhill motion is guaranteed. This is a reliable, but somewhat blunt, method: the convergence can be rather slow, especially near the local minimum.

Before the first step of the simulation proper, the size and shape (usually a box) of the simulation cell have to be defined, and the coordinates and velocities of all atoms in the system have to be set. The box size is determined simply by the three basis vectors. Unless the simulation is a continuation of a previous run, it starts at $t = t_0$, and the coordinates at $t = t_0$ have to be determined by the user. Then the leap-frog algorithm is used to propagate the time step with $\Delta t$ which means that the velocities at $t = t_0 - \Delta t/2$ also have to be established. Velocities could be available from previous simulations, but if not, the initial atomic velocities $v_i$, $i = 1 \ldots 3N$ are generated with a Maxwellian distribution at a set temperature T:

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}} e^{-\frac{m_i v_i^2}{2kT}}, \quad (2.1.4)$$

where *k* is Boltzmann's constant. This is done by generating normally distributed random numbers, which are then multiplied by the standard deviation of the velocity distribution $\sqrt{kT/m_i}$. The resulting total energy will not correspond precisely to the set temperature T, so a correction must be made by, first, removing the center-of-mass motion, and, second, scaling all velocities until the total energy corresponds precisely to T.

After that following steps of the MD can be calculated. At every step, forces acting on each atom are calculated from the forcefield, which provides new coordinates that each atom will take after a time step. At certain intervals, the geometry of the system is written to the trajectory file, which is analyzed as a main result of the MD simulation. The general scheme for the MD algorithm (often called "the integrator") is given in Figure 2.1.2. The main result of the MD simulation study is a trajectory – a set of consecutive "frames", conformations of the system written with a certain timestep. Before doing any analysis procedures simulation results (particularly the trajectory) were inspected manually using VMD [51].

Some pre-processing of the trajectory files is often required to speed up the analysis. This includes such procedures as merging several files into one, reassigning the center of coordinates and/or unit cell walls positions to prevent "jumps" of molecules between periodic cells and to fix molecules broken by cell walls, removal of water molecules in some cases to decrease the size of the files and provide faster analysis. All manipulations with trajectories were done using VMD [51].

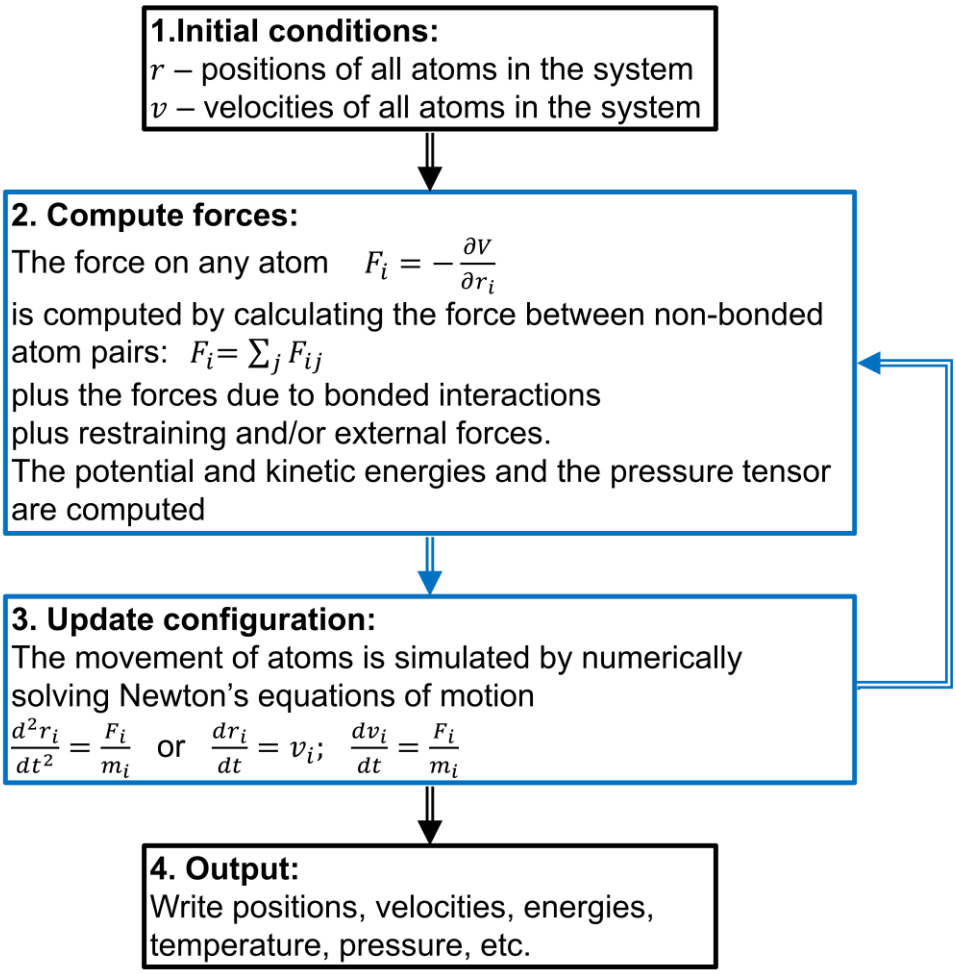


**Figure 2.1.2.** The global MD algorithm. Figure from GROMACS [50] User Guide.

To estimate the overall flexibility of the system and/or its parts the root mean square fluctuation (RMSF) analysis is performed. The RMSF is a measure of the deviation between the position of particle *i* and some reference position:

$$RMSF_i = \left[\frac{1}{T}\sum_{t_j=1}^{T}\left|r_i(t_j) - r_i^{ref}\right|^2\right]^{1/2}, \quad (2.1.5)$$

where T is the time over which one wants to average and $r_i^{ref}$ refers to the reference position of particle *i*. Typically this reference position will be the time-averaged position of the same particle *i*. For comparison of particular structures, the root mean square deviation (RMSD) can be also applied. The difference between RMSD and RMSF is that the RMSF is averaged over time, giving a value for each particle *i*. For the RMSD the average is taken over the particles, giving time-specific values. RMSF analysis was performed with the corresponding plug-ins for VMD [51].

Interaction between small molecules in solution can be evaluated using radial distribution function (RDF). The RDF is an example of a pair correlation function, which describes how (on average) the atoms in a system are packed around each other. This proves to be a particularly

effective way of describing the average structure of disordered molecular systems such as water solutions. For example, in a simulation of a mixture of two components A and B, the radial distribution function $g_{AB}(r)$ would be calculated as follows:

(2.1.6) $$4\pi r^2 g_{AB}(r) = V \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} P(r) \,,$$

where V is the volume and P(r) is the probability of finding an atom B at the distance *r* from an atom A. In any system, two groups of atoms A and B can be defined by a user by assigning particular atoms by their numbers to each group. The RDF is usually plotted as a function of the interatomic distance *r*. A typical RDF plot shows several important features. First, at short distances, the RDF is zero. This indicates the effective width of the atoms since they cannot approach any more closely. Second, one or more obvious peaks can appear, which indicates that the atoms pack around each other in "shells". The location of this peak shows the distance of atomic non-covalent interaction and the height – the probability of such interaction. Usually, at high temperatures the peaks are broad, indicating thermal motion, while at the low temperature they are sharp. At a very long range, every RDF tends to a value of 1, which happens because the RDF describes the average density at this range. In this work, MATLAB scripts were used for all RDF calculations.

VMD was used to visualize the mobility of residues forming particular interactions. MATLAB scripts were used to measure particular distances, angles, and dihedral angles, analyze and plot the results. Biomolecular processes, such as folding or complex formation, can be described in the terms of the molecule's free energy:

(2.1.7) $$\Delta G(r) = -k_B T \cdot [ln(P(r) - ln(P_{max})]),$$

or

(2.1.8) $$G(r) = -k_B T \cdot ln(P(r)),$$

where $k_B$ is the Boltzmann constant, P is the probability distribution of the molecular system along some coordinate *r* (called the order parameter), and $P_{max}$ denotes its maximum, which is often subtracted to ensure $\Delta G = 0$ for the lowest free energy minimum. Order parameter (*r*) is chosen to represent a reaction coordinate for the process under investigation. Free energy values are plotted along the coordinates which correspond to the significant changes in the molecule conformation. Typically, the free energy is plotted along two such order parameters, giving rise to a (reduced) free energy surface. In this work, 2D plots of ATP conformations in solution directly show the probabilities of conformations instead of the free energy surface due to incomplete coverage of conformational space during MD simulations. MATLAB scripts were used to calculate all probability distributions and free energy plots.

Thus, MATLAB software was used for statistical analysis of geometrical features of the molecule during simulation. A set of scripts were written to calculate the distances between particular atoms during MD and the probability distribution of said distances. VMD software was used to visualize the conformation of molecules during simulations and to plot RMSF data for large protein systems. PyMol software [52] was used for protein structures superposition, analysis, and visualization of separate structures obtained after simulations.

### *2.1.2. Simulations of ATP and GTP in Water*

To examine the effects the cation binding has on the conformation of the Mg-NTP complexes, MD simulations were conducted for Mg-NTP complexes in a pure water solution and with the addition of $M^+$ ions ($K^+$, $Na^+$, or $NH_4^+$). In addition to these ions, $Cl^-$ ions were added to bring the sum of charges in the system to zero. In the simulations of Mg-NTP (Mg-ATP and Mg-GTP) complexes in a pure water solution (without cations, other than $Mg^{2+}$), two dummy positive charges were used to neutralize the system. These atoms carried each a single positive charge and were affixed to their locations with positional restraints, preventing any interactions with the ATP or GTP molecule. In all simulations, the ATP or GTP molecule position was restrained in the center of the simulation cell with positional restraints, which were applied to the N1 atom.

For simulations, $ATP_4^-$, $GTP_4^-$ and $NH_4^+$ molecules were described with the parameters from CGenFF v.2b8 [49]. The TIPS3P water model was used, as compared to other classical models TIPS3P has an additional set of van der Waals parameters that describe interactions between water molecules [53]. For $Na^+$ and $K^+$ ions, parameters by Joung and Cheatham [54] were used. The $Mg^{2+}$ ion was described with parameters developed by Callahan et al. [55].

Non-bonded interactions were computed using the particle mesh Ewald method, the real space cutoff for electrostatic interactions was set at 10Å, the van der Waals interactions cutoff was set as switching functions between 10 and 12Å. The multiple time-step method was used to describe electrostatic forces. Finally, the non-bonded interaction list was composed using a cutoff of 14Å, updated every 20 steps. The SHAKE algorithm was used to constrain covalent bonds that include hydrogen atoms [56] (the MD integration step, 1 fs). The total ionic strength in the system was 0.2 M, after the addition of water molecules, $Na^+$ or $K^+$, and/or neutralizing ions.

All productive runs were set in the NPT ensemble. The Berendsen thermostat was used to maintain the temperature at T= 298 K with a coupling parameter of 5 $ps^{-1}$ [57]. The Langevin piston method was used to maintain the pressure at one atm with the piston mass 100 amu, while the Langevin collision frequency was set at 500 $ps^{-1}$ [58].

Prior to productive runs, each system was optimized by performing an energy minimization run, and then a 20 ns equilibration run. MD simulations were executed in three independent runs of 170 ns (totaling at 500 ns) for each system ($K^+$, $Na^+$, $NH_4^+$, no $M^+$). Additional series of short (20–25 ns) simulation runs were performed for both ATP and GTP complexes to further investigate the effect of monovalent cations binding on $Mg^{2+}$ coordination. Finally, two simulations of the Mg-ATP complex with $K^+$ ions bound to the phosphate chain and restrained to their positions were conducted. All simulations are listed in Table 2.1 MD simulation calculations were performed with Gromacs v.4.5.5 [59] software with MPI implementation using the computational resources of the supercomputer SKIF 'Chebyshev' of the Moscow State University Computational Center.

MD data analysis was performed using MATLAB software [60]. For the visualization of the trajectories obtained from MD simulations, the VMD software [51] was used.

**Table 2.1.** Molecular dynamics simulations of NTPs in water

| No. | System | Simulation time | Number of repetitions |
|---|---|---|---|
| 1 | Mg-ATP | 167 ns | 3 |
| 2 | Mg-ATP, $K^+$ | 167 ns | 3 |
| 3 | Mg-ATP, $Na^+$ | 167 ns | 3 |
| 4 | Mg-ATP, $NH_4^+$ | 167 ns | 3 |
| 5 | Mg-ATP | 20 ns | 25 |
| 6 | Mg-ATP, $K^+$ | 20 ns | 25 |
| 7 | Mg-ATP, $Na^+$ | 20 ns | 25 |
| 8 | Mg-ATP, $NH_4^+$ | 20 ns | 25 |
| 9 | Mg-GTP | 20 ns | 20 |
| 10 | Mg-GTP, $K^+$ | 20 ns | 20 |
| 11 | Mg-GTP, $Na^+$ | 20 ns | 20 |
| 12 | Mg-GTP, $NH_4^+$ | 20 ns | 20 |
| 13 | Mg-ATP, $K^+$, with positional restraints | 10 ns | 2 |

### *2.1.3. Simulations of K-dependent GTPase MnmE*

The role of the monovalent cations in the GTPase MnmE was studied by MD simulations of the Mg-GTP/MnmE complex. The following three states have been modelled: (1) the active state of the dimer of MnmE G-domains, with $K^+$ ions bound in both domains (PDB ID 2GJ8, resolution 1.7 Å, source: E.coli), (2) the inactive state (monomer) with the K-loop in its active conformation and a water molecule occupying the $K^+$-binding site, (PDB ID 2GJ8, resolution 1.7 Å, source: E.coli), and, finally, (3) the inactive state (monomer) with disordered K-loop (PDB ID 3GEI, resolution 3.4 Å, source: *Chlorobium tepidum*).

As mentioned previously, the availability and quality of crystal structures almost always inform the choice of objects for MD simulations. In the case of $K^+$-dependent P-loop NTPases, MnmE is the only protein for which both active state of the P-loop domain with resolved $K^+$ ion structure and an inactive state structure are available. Still, these structures are not perfect. Specifically, since the K-loop is disordered in the inactive state, it is unresolved in the corresponding crystal structure. For MD simulations, this loop must be reconstructed. A similar modeling of disordered protein loops was performed earlier in a collaboration with Prof. Dr. Karin Busch to model fluorescent sensor proteins fused to membrane subunits of respiratory supercomplex [61]. Existing experimental structures of the green fluorescent protein and cytochrome oxidase subunits were used as templates to model the respective parts of the fusion construct, but the mobile disordered linkers between these two parts had to be constructed *ab initio*. To model these loops each construct structure was first optimized with the variable target function method with conjugate gradients, and then refined using molecular dynamics (Modeller v.9.25 build-in extension [62]) with simulated annealing. The resulting models were used to describe the mobility and environment of the constructs within the respiratory supercomplex, which corresponded well with the experimental data on fluorescence lifetimes of the sensors [61]. Thus, the same method, as described in more detail in [61], was implemented here to reconstruct the missing loops in the structure of the inactive P-loop domain of MnmE (PDB ID 3GEI).

Each protein complex was placed in a cubic cell filled with TIP3P water with standard periodic boundary conditions. The minimal distance between any atom of the protein and the periodic cell wall was set at 12 Å. In each case, $K^+$ and $Cl^-$ ions were added to the surrounding media. Each productive MD simulation run was 100 ns long. All simulations of GTPase MnmE are listed in Table 2.2. Simulation conditions were the same, as for simulations of ATP and GTP in water, as described in Chapter 2.1.2.

**Table 2.2.** Molecular dynamics simulations of GTPase MnmE

| No. | System | Simulation time | Number of repetitions |
|---|---|---|---|
| 1 | Mg-GTP-MnmE, inactive, no K-loop, 3GEI | 100 ns | 1 |
| 2 | Mg-GTP-MnmE, inactive, K-loop, no $K^+$, $2GJ8_W$ | 100 ns | 1 |
| 3 | Mg-GTP-MnmE, active dimer with $K^+$, $2GJ8_K$ | 100 ns | 1 |

### *2.1.4. Simulations of EF-Tu with GTP and RNA*

Cation binding and conformational mobility of GTP in EF-Tu/ribosome complex were studied by molecular dynamics (MD) simulations. Simulations were performed with Gromacs v.4.5.5 software [59]. Protein/RNA complex was placed in a cubic cell filled with TIP3P water with standard periodic boundary conditions. The minimal distance between any atom of the protein and the periodic cell wall was set at 12 Å. Ions ($K^+$, $Mg^{2+}$, $Cl^-$) were added to the solution to create a physiological concentration of cations and neutralize the system, as described in Section 3.2.5.

For simulations, the CGenFF v.2b8 force field was used for the GTP, and the CHARMM36 force field was used for the protein and RNA fragments [49]. For the $Mg^{2+}$ ion parameters designed by Calahan et al. [55] were used. For $K^+$ ions parameters by Joung and Cheatham [54] were used.

Two systems were prepared for MD simulations: one with a $K^+$ ion placed manually in the AG site, and the other without. Each system was first optimized by an energy minimization run followed by an equilibration run of 20 ns. Equilibration runs were performed with constant temperature and cell volume (NVT ensemble); positional restraints were applied to the RNA, protein, and the GTP molecule. Productive runs were performed with constant temperature and pressure (NPT ensemble) with the temperature of 303.15 K, controlled by Nose-Hoover thermostat, and the pressure of 1 atm, controlled by Parrinello-Rahman barostat. Long-range electrostatic interactions were calculated by implementing the particle mesh Ewald method with Verlet cutoff-scheme [63] and a 12 Å distance cutoff for direct electrostatic and van der Waals interactions. The switching function was set to 10 Å to gradually reduce van der Waals potentials, reaching 0 at the cutoff distance. The geometry of bonds between hydrogens and heavy atoms was constrained to the lengths and angles defined by the force field using LINCS (LINear Constraint Solver) algorithm [64].

Six independent simulations of 100 ns each (productive runs) were performed: a single run for the first system (with $K^+$ ion placed manually in the AG site), and five runs for the second system (with $K^+$ ions randomly distributed at the start of the simulation). All simulations of the EF-Tu/tRNA/SRL complex are listed in Table 2.3. The positional restraints were applied to all RNA fragments in the system for the duration of the production runs. The entire EF-Tu, GTP-$Mg^{2+}$, as well as all surrounding ions and water molecules, were free of positional restraints. During simulations, conformations of the whole system were saved every 0.1 ns. Productive runs were used to extract characteristic frames to represent the geometry of the GTP binging site with visual molecular dynamics (VMD, [51]) and to calculate statistics of structure movement using MATLAB R2017a [60].

**Table 2.3.** Molecular dynamics simulations of EF-Tu/tRNA/SRL complex

| Simulation number | Molecules in the GTP-binding-site | Molecules in the solution | Length |
|---|---|---|---|
| 1 | EF-Tu, SRL fragment, tRNA fragment, GTP-$Mg^{2+}$, $K^+$ (AG site), $Mg^{2+}$ (SRL) | 70 $K^+$ ions<br>10 $Mg^{2+}$ ions<br>51 $Cl^-$ ions<br>19878 water molecules | 100 ns |
| 2-6 | EF-Tu, SRL fragment, tRNA fragment, GTP-$Mg^{2+}$, $Mg^{2+}$ (SRL) | 70 $K^+$ ions<br>10 $Mg^{2+}$ ions<br>50 $Cl^-$ ions<br>19880 water molecules | 100 ns |

## *2.2. Protein Sequence and Structure Analysis*

In the course of evolution, proteins advanced from a common ancestor by sequence modifications – substitutions, deletions, or insertions of residues – giving rise to families of homologous proteins. Not all residues in a protein are equally susceptible to mutations, since some of them may be crucial to maintain structure or function and thus corresponding positions in the sequence are constrained in the allowed residue types [65]. Such positions with residues conserved across the entire families of proteins were the first to attract attention [66]. Some residues can be conserved within a smaller group of proteins, united by a specific function, while not being conserved across the entire (super)family of proteins. Those residues can be associated with the functional diversity between particular protein families within a superfamily, or generally between smaller groups of proteins within a larger group. Finally, variable residues exhibit no detectable conservation across groups of homologous proteins and thus are likely to not be directly responsible for the protein function and structure. The degree of conservation of each particular residue, or even larger parts of a protein (e.g. particular elements or groups of elements of the secondary structure – loops, α-helices, or β-strands) can be obtained from multiple sequence alignments and structure superpositions of homologous proteins. While consulting multiple sequence alignments on many occasions, this work primarily utilizes structure superposition for the identification of residues, that are conserved (universally or partially) within (super)families of proteins.

Comparison of protein structures is an essential step in establishing the evolutionary relationships between proteins and protein families. While high sequence similarity almost always

implies structural similarity, the opposite is not true. Specifically, remote homologs can have a similar structure and other common features, such as functional residues, but share no detectable sequence similarity [67]. It is therefore expected that a three-dimensional structure alignment will provide more clues towards protein evolution and properties than a sequence alignment alone, by revealing similarities between extremely distant homologs. The similarity analysis of protein structures can also provide insight into the protein function mechanism and the roles of particular structural elements.

### *2.2.1. Protein Families and Multiple Sequence Alignments*

Protein sequence analysis typically involves the construction of pairwise or multiple sequence alignments (MSAs) of relevant sequences. Protein sequence alignment is a notion of sequences that places sequences in rows one under another so that identical or similar residues are placed in the same columns. This is achieved by the addition of "gaps", which represent "missing" residues or stretches of residues so that the overall similarity of residues across all columns is maximized. In the case of pairwise alignment, all possible variations can be explored to identify alignment with the best possible score. Alignment scores higher when it matches identical residues and residues with similar sidechains, while the introduction of gaps and their length reduce the score. However, such an approach is too computationally demanding to be applied for the multiple sequence alignments. MSAs are an essential part of protein family analysis. The MSA is constructed in the hope that it will reflect the structural and functional similarity between the proteins, positioning in the same column residues that have similar locations in the structure and presumably have the same function or similar functions.

Before constructing an MSA one must determine the set of sequences that are to be aligned. The standard routine for this is a similarity search, which is performed to retrieve from a sequence database(a) all protein sequences that share at least some similarity with the one or sever sequences of interest (query sequence(s)).

Sequence similarity search is typically performed by the **BLAST** software [68, 69]. This set of tools constructs pairwise sequence alignments between the query sequence and all protein sequences in the target database. Although BLAST tool set can be used also for DNA and RNA sequences, in this work it was applied only to protein sequences and corresponding databases. BLAST stands for the Basic Local Alignment Search Tool (BLAST), which signifies, that rather than aligning entire sequences, the algorithm finds regions of local similarity between sequences. Each time a short region of a query sequence matches a sequence in the target database, the local

alignment between the two fragments is extended to cover as much of the sequence length as possible and the score of the resulting alignment is calculated, based on its length and the number of identical and similar residue pairs. The significance of a match is estimated by an "expect value" (*e-value*) which shows how many matches would have occurred with the same score by chance when searching an entirely random sequence of the same length in an entirely random database. Protein matches detected by BLAST with high significance can be used to infer functional and evolutionary relationships between sequences as well as help identify members of protein families.

The MSAs in this work were constructed using **T-Coffee** software (Tree-based Consistency Objective Function for Alignment Evaluation) [70]. This software constructs a library of pairwise alignments for all pairs of sequences in the input set, which are then used to guide the construction of the MSA. It can also use data from MSAs obtained by various other methods. In the end, T-Coffee produces an MSA that is consistent with most of the intermediate data – pairwise alignments and MSAs produced by other methods. In this work, the online version of the software was used, either directly via http://tcoffee.crg.cat/ or the web service implemented in **JalView** [71]. The latter was also used for visualization and manual inspection of all alignments.

Alternatively, MSA containing protein(s) of interest can be obtained from protein family databases, which already contain such alignments representing various protein families. In this work, the protein family database **Pfam** [72] was used. Each entry in the Pfam database contains a "seed" MSA of proteins representing a particular protein family or domain. This alignment forms the basis for a profile hidden Markov model (HMM) which is used to automatically recognize the presence of particular domains/families in protein sequences. The HMM profile is used to search for matching protein sequences in a *pfamseq* database, which is specially created and curated for this use. All protein sequences in the *pfamseq* database that match the profile well enough are also aligned with each other and "seed" sequences to comprise a "full" MSA, also available for each Pfam entry. In addition to the "seed" and "full" MSAs, Pfam provides the HMM profiles, visualized as sequence logos that highlight individual conserved residues and motifs that are characteristic for a particular family/domain. Generally, Pfam aims to cover as many protein sequences as possible with the fewest number of models. This sometimes leads to Pfam entries uniting proteins/domains which have varying activities or functions, so that many Pfam families can be further subdivided into sub-families. For example, Pfam family PF00006 ("ATP synthase alpha/beta subunits") includes "ATP-binding" domains of both regulatory and catalytic subunits of the rotary ATP synthases. On the opposite side, some superfamilies consist of proteins with similar structure and function, but the sequences are so different that a single profile HMM cannot cover them all. Such cases are addressed by Pfam clans, which unite multiple Pfam entries that are

known to be evolutionarily related to each other. Specifically, all P-loop containing families/domains in Pfam are united in the clan CL0023 “P-loop NTPase”. In this work Pfam database (available at https://pfam.xfam.org/) was used to access MSA of particular families as well as logo representations of HMM profiles of particular protein families.

### *2.2.2. Protein Structure Analysis*

The main source of structural data used in this work is the Protein Data Bank (**PDB**) [73]. Since 1971, the PDB archive has served as the main repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies. The data, typically obtained by X-ray crystallography, NMR spectroscopy, or fitting models into cryo-electron microscopy is submitted by scientists from around the world and is freely accessible on the Internet via the websites of its member organizations (PDBe at http://www.ebi.ac.uk/PDBe/, PDBj at http://PDBj.org/, and RCSB at http://www.rcsb.org/). The PDB is overseen by the Worldwide Protein Data Bank, wwPDB [74].

Currently, every newly determined protein structure has to be deposited with the protein data bank before the scientific paper reporting on the structure can be published. The number of structures in the PDB has exceeded 150 000, as indicated on at http://www.rcsb.org/. However, one should remember that this number does not reflect the number of unique proteins. In many cases, there are many entries of the same protein in the database - some variants with amino acid mutations, some complexes with different bound molecules (substrate analogs, inhibitors, co-factors), molecules crystallized in different conditions, etc.

In the analysis of crystal structures from PDB, electron density data should be considered as well. Such data is available at the Uppsala Electron Density Server (**EDS**; http://eds.bmc.uu.se/) [75]. EDS is a web-based facility that provides access to electron-density maps and statistics concerning the fit of crystal structures and their maps. Maps are available for approximately 87% of the crystallographic PDB entries for which structure factors have been deposited and for which straightforward map calculations succeed in reproducing the published R-value to within five percentage points. This data can be used to inspects electron density in poorly resolved parts of the structure and observe the electron density of particles with low occupancy which were not resolved in the PDB file.

Over the last decades, there has been a huge increase in the numbers of protein sequences and structures determined. In parallel, many methods have been developed for recognizing similarities between these proteins, arising from their common evolutionary background, and for

clustering such relatives into protein families. Since protein function is intrinsically linked to its precise shape, a 3D structure can often be used to detect very remote evolutionary relationships, even after the amino acid sequence has changed beyond recognition. This work had relied on the InterPro database as a source of structure-based protein classification.

**InterPro** provides functional analysis of proteins by classifying them into families and predicting domains and important sites. This is achieved by combining protein signatures from several member databases into a single searchable resource, capitalizing on their individual strengths to produce a powerful integrated database and diagnostic tool. InterPro is used by research scientists interested in the large-scale analysis of whole proteomes, genomes, and metagenomes, as well as researchers seeking to characterize individual protein sequences. In this work, InterPro was used to identify and assess all structures containing P-loop domains.

Structure analysis of transmembrane proteins, such as microbial rhodopsins (MRs) and G-protein coupled receptors (GPCRs), requires information on their relative position and orientation in the membrane. Due to the specifics of purification and crystallization methods, crystal structures of membrane proteins usually do not contain membrane lipids, but often contain molecules of detergent used in protein purification and crystallization experiments. The position of these molecules, together with the distribution of polar and hydrophobic residues on the protein surface can point to the position of the protein in the membrane. In this work "Orientation of Proteins in Membranes" (**OPM**) database [76] was used to obtain protein structures with assigned membrane limits. OPM provides spatial arrangements of membrane proteins relative to the hydrocarbon core of the lipid bilayer. This database includes all unique experimental structures of transmembrane proteins as well as some peripheral proteins and membrane-active peptides. Each protein is positioned in the lipid bilayer of appropriate thickness by minimizing its transfer energy from water to the membrane. OPM database is available at http://opm.phar.umich.edu/, where PDB-format files with dummy-atoms indicating membrane borders can be downloaded.

While OPM provides a visual representation of the boundaries of the hydrophobic layer of the membrane, structure analysis of transmembrane proteins can benefit from a more explicit representation of the locations of lipid molecules around the protein. Particularly in this study, the relative locations of the hydrophobic and hydrophilic parts of the membrane were crucial for the discussion of the suggested sodium ion pathway in GPCR. For this purpose, an all-atom structure of a GPCR molecule in the membrane was constructed using CHARMM-GUI [77]. CHARMM-GUI is a web-based platform for the interactive building of complex systems, which can be used for biomolecular simulations. This platform is particularly useful for the studies of membrane

proteins, as it supports a wide variety of lipid molecules as well as different types of lipid assemblies (e.g. micelles and nanodiscs).

Specialized databases can be an extremely useful source of additional data. Such databases arise in response to a particular interest in the scientific community towards a specific group of proteins, type of interactions, or a type of small molecules present in experimental structures. GPCRs are one of such targets of special interest since they make up the largest family of human membrane proteins and drug targets. The GPCR database, **GPCRdb** serves the wide GPCR community since 1998, providing reference data, web server analysis tools, and dynamic visualization of data and statistics [78, 79]. Available at https://gpcrdb.org/, GPCRdb provides contains data, diagrams, and web tools for GPCRs, including all GPCR structures and the largest collections of receptor mutants. In this work, GPCRdb was used to survey all available structures of GPCRs and the reference alignment between class A GPCRs. Reference structure-based sequence alignments of GPCRs take into account helix bulges and constrictions.

Analysis of the abundance of available protein structures requires rapid construction of protein structure superpositions. In contrast to the sequence alignments, three-dimensional alignment is based on the comparison of geometrical positions of amino acid residues in the structure, rather than the biochemical properties. Here several methods for protein structure superposition are discussed as they are used for different tasks in this study. For the (super)family of P-loop NTPases, comparative structure analysis was performed, which required superposition of multiple proteins, often sharing only local structure similarity. For the (super)families of heptahelical transmembrane proteins structure analysis required extensive search for the structure similarity to ascertain possible relations between the two (super)families under investigation – microbial rhodopsins and G-protein coupled receptors. Structure superposition methods used to address these tasks are described below.

For the comparative structure analysis of the P-loop NTPases multiple structures, representatives of particular protein families were selected manually. All proteins of this group have an easily identifiable and highly conserved structural motif of a β-strand and following α-helix, that flank the titular P-loop. Hence, superposition of even the most distant homologs was possible via direct superposition of their P-loop regions. This procedure was performed with the PyMol build-in function "super". This function aligns two selections, each including a molecule or its fragment. It does a sequence-independent structure-based dynamic programming alignment followed by a series of refinement cycles intended to improve the fit by eliminating pairing with high relative variability. This function was used to construct the superposition of multiple P-loop-

containing proteins by matching the P-loop regions of all structures under investigation onto the single reference structure.

On a larger scale, all available structures of P-loop NTPases were analyzed systematically using an automated procedure. The list of PDB IDs for this analysis obtained from the InterPro entry IPR027417 and narrowed down using the search engine of the RCSB PDB [80] to include only structures containing $Mg^{2+}$ ion and at least one of the following NTP-like molecules (RCSB PDB chemical IDs): ATP or GTP, non-hydrolyzable ATP/GTP analogs (ANP, ACP, AGS, GNP, GCP, and GSP), or transition state analogs (ADP/GDP complexes with $AlF_3$/$AlF_4^-$, $VO_4^{3-}$, or $BeF_3$/$BeF_4^-$). MATLAB software [60] was used to evaluate the distances between the NTPs or analogous molecules and the surrounding Lys/Arg residues. This evaluation was used to identify the structures where the NTPs or analogous molecules were bound to a Lys residue (implying that the molecule is not only present but bound to the P-loop with the P-loop Lys residue present). MATLAB software [60] was used also to evaluate the phosphate chain shapes in each NTP-like molecule or the transition state-mimicking complex. This automated procedure produced data that reflects the shape and binding patterns of different NTP-like molecules that occur in the P-loop NTPase structures.

The second group of proteins, addressed here, are the heptahelical transmembrane proteins – G-protein coupled receptors and microbial rhodopsins. Diverse structure superposition methods were applied to search for structure similarity between members of the two (super)families. This task was addressed by using a structure of sodium-pumping microbial rhodopsin to perform a PDB-wide search for similar structures. The best tools for this task are the jFATCAT-rigid algorithm [81] and the PDBeFOLD server [82] because both of these methods are comparatively fast and can perform similarity search across all structures in the entire PDB database. Furthermore, both these algorithms produced sequence alignments that matched the results of the respective structural superpositions.

The advantage of the jFATCAT-rigid algorithm is the availability on the PDB web-server [80] of pre-calculated superpositions for a set of representative structures [81, 83]. The java-version of the jFATCAT algorithm is also available at http://www.rcsb.org/PDB ID/workbench/workbench.do. The FATCAT (Flexible structure AlignmenT by Chaining Aligned Fragment Pairs with Twists) approach simultaneously addresses the two major goals of flexible structure alignment; optimizing the alignment and minimizing the number of rigid-body movements (twists) around pivot points (hinges) introduced in the reference protein. In contrast, currently existing flexible structure alignment programs treat hinge detection as a post-process of a standard rigid-body alignment.

PDBeFold service performs the PDB-wide similarity search using the SSM (Secondary Structure Match) algorithm [82] implemented at http://www.ebi.ac.uk/msd-srv/ssm/. The SSM algorithm of protein structure comparison in three dimensions includes an original procedure of matching graphs built on the protein's secondary-structure elements, followed by an iterative 3D alignment of protein backbone Cα atoms. PDBeFold service provides pairwise and multiple comparisons and 3D alignment of protein structures, as well as the examination of a protein structure for similarity with the whole PDB archive or SCOP archive. In most cases, the SSM algorithm works very fast. A typical query - matching a protein of a few hundred residues against the whole PDB or SCOP archive - takes less than a minute, a significant part of which falls on the network communication.

For pairwise superposition of structures with high similarity, another algorithm implemented in the PyMol package [52] was used. For proteins with decent sequence similarity (identity >30%) the "align" function produces the best structure superposition results. This function performs a sequence alignment followed by a structural superposition and then carries out zero or more cycles of refinement in order to reject structural outliers found during the fit. This function was used to superpose 7TM proteins within one family (e.g. microbial rhodopsins with other microbial rhodopsins). Finally, PyMol [52] was used for the visualization of all structures and structure superpositions presented here.

# 3. P-loop Nucleoside Triphosphatases

## 3.1. Background

### 3.1.1. Nucleoside Triphosphatases of the P-loop Fold

Living organisms can acquire energy from photons (phototrophic organisms) or chemical reactions (chemotrophic organisms). In the cell, this energy can be then transformed into one of three forms: as an electrochemical ion gradient (e.g. of protons), as macroergic compounds, or as the reducing potential of electron carriers. Hydrolysis of macroergic (high-energy) compounds releases a lot of energy and can be coupled with certain reactions in the cell, shifting the chemical equilibrium of those reactions in favor of the products [84].

Nucleoside triphosphates (NTPs), particularly adenosine triphosphate (ATP) and guanosine triphosphate (GTP) are the most important and universal macroergic compounds [85]. A nucleoside triphosphate (NTP) is a molecule containing a nucleobase (adenine, guanine, or other), a 5-carbon sugar (ribose in ATP and GTP), and three phosphate residues. NTPs and their derivatives also serve as the building blocks for nucleic acids (DNA and RNA) and have a plethora of other roles in cell metabolism and regulation.

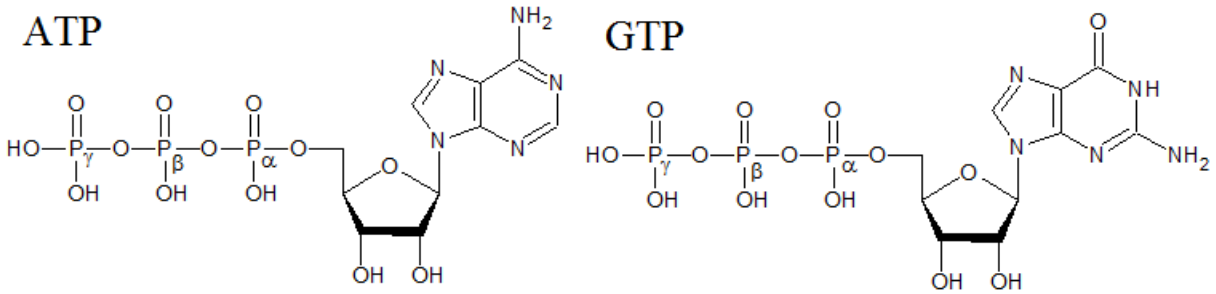


**Figure 3.1.1**. Structure of NTPs: ATP and GTP. Figure created with ChemSketch [86].

NTP molecules mainly serve as the storage and the transfer route for the free energy in cells. Macroergic compounds able to release a greater amount of energy than NTP exist, but they are not as widely used. One of the reasons for the ubiquitous use of ATP and GTP is the relative stability of the phosphate anhydride bond, which is resistant to spontaneous hydrolysis (unlike other anhydrides) and undergoes splitting only in the presence of enzymes.

Although many different nucleotide triphosphates can be found in a cell, the majority of energy-converting systems operate with either ATP, or GTP, or both [87, 88]. Some processes, however, require a specific type of nucleotide, so the equilibrium between the concentrations of

different nucleoside triphosphates must be maintained. This is achieved by the exchange of terminal phosphate between different nucleoside diphosphates (NDP) and triphosphates (NTP) in a reaction

$$XDP + YTP \longleftrightarrow XTP + YDP,$$

where X and Y each represent different nucleobase, e.g. adenine or guanine. This reaction is catalyzed by the nucleoside-diphosphate kinases [84].

Many properties of NTPs are well studied on the example of the ATP, but since the GTP has a very similar structure (see Figure 3.1.1) majority of properties are shared in equal measure by ATP and GTP [87, 88].

Hydrolysis of NTPs to nucleoside diphosphates is crucially important for many processes in the cell [24, 89]. The reaction itself is exothermic and can be coupled with endothermic, energetically unfavorable reactions to allow their progress [90]. NTP binding and hydrolysis can also cause conformational changes in protein performing "switching" the protein from one state to another and causing different signaling from it [91-95]. Finally, direct transfer of phosphate from NTP to a substrate (e.g. in creatine kinase reaction, recently reviewed in [96]) or a protein itself (e.g. in histidine kinases, for a recent review see [97]) can occur.

The most widespread NTPases are so-called P-loop fold NTPases that make up to 10-20% of gene products in a typical cell [24-26]. They are found in rotary ATP synthases, DNA and RNA helicases, proteins that hydrolyze ATP to perform mechanical work, such as kinesin, myosin, and dynein, many GTPases, including ubiquitous translation factors and α-subunits of signaling heterotrimeric G-proteins, as well as in other enzymes. P-loop fold domains appear to be some of the most ancient protein domains dating back to the Last Universal Cellular Ancestor (LUCA) [24, 26, 65, 98-102]. The P-loop fold (CATH annotation 3.40.50.300; SCOP superfamily c.37.1, PFAM clan CL0023) is a 3-layer αβα sandwich [24, 103-106]. It contains the GxxxxGK[S/T] sequence motif, commonly denoted the Walker A motif [107], see Figure 3.1.2. This motif binds the triphosphate chain of an NTP molecule and is also referred to as the P-loop (*p*hosphate-binding loop) motif [108] or, in GTPases, as G1 motif [109]. In the P-loop fold, the conserved Lys residue forms hydrogen bonds (H-bonds) with the $O^{1B}$ oxygen atom of β-phosphate and $O^{2G}$ atom of γ-phosphate groups of the NTP (hereafter, the atom nomenclature follows the recent IUPAC recommendations [110], see Figure 3.1.2). Usually, the P-loop connects the first β-strand and the first α-helix of the P-loop domain. The Walker B motif *hhhh*D, where '*h*' stands for a hydrophobic residue, provides the conserved Asp residue that usually serves as an additional $Mg^{2+}$ ligand [107]; in G-proteins (P-loop GTPases) this motif is also called Switch II [111] or G3 [109]. One more functionally important motif in G-proteins is the Switch I or G2 motif [109, 111], located between

the Walker A and B motifs with only a single Thr/Ser amino acid conserved. In G-proteins, the sidechain of this Thr/Ser residue coordinates $Mg^{2+}$ ion, while its backbone nitrogen forms an H-bond with the γ-phosphate group.

In this work amino acids of conserved motifs Walker A (P-loop, G1), Switch I (G2), and Walker B (Switch II, G3) are referred to by their positions relative to the signature residues, namely Lys (K) of the Walker A (P-loop) motif, Thr (T) or Ser (S) of the Switch I (G2) motif, and Asp (D) of the Walker B (Switch II, G3) motif, respectively, see Figure 3.1.2. For example, the Ser/Thr residue in the K+1 position coordinates the $Mg^{2+}$ ion in all P-loop fold NTPases (Figure 3.1.2).

An uncontrolled NTP hydrolysis would be detrimental for the cell survival. Therefore, the specific feature of most P-loop fold ATPases is their activation upon each turnover. To achieve that, the NTP hydrolysis reaction is divided into two steps. First, an NTP molecule binds to the P-loop domain and attains a strained, catalytically prone conformation that is characterized by the eclipsed orientation of the β- and γ-phosphates enforced by the signature Lys residue and the cofactor $Mg^{2+}$ ion, as shown in Figure 3.1.2 [106, 112-116]. Second, the P-loop fold domain interacts with a proper physiological partner, which could be a domain of the same protein, a separate protein, and/or a DNA/RNA molecule. Upon this interaction, a stimulating moiety, e.g. an Arg residue ("finger"), is inserted into the catalytic site and the cleavage of the NTP molecule takes place [28, 29, 115-123]. The need for a specific activating interaction ensures that the fast NTP hydrolysis proceeds not spontaneously, but in a controlled manner. Both the NTP binding and its hydrolysis can be accompanied by large-scale conformational changes that could be used for performing mechanical work, see, e.g. [91, 124].

To provide a distinction from the term “activating partner” which refers to a domain of the protein, a separate protein, or a DNA/RNA molecule, the term “stimulating moiety” is used hereafter for those moieties that activate the NTP hydrolysis by poking into the catalytic site. Noteworthy, the original Latin meaning of “stimulus” – “a sharp stick used to poke cattle to get them to keep moving” (quoted from https://www.dictionary.com/browse/stimulus) - nicely describes the function of Lys and Arg fingers in biological motors.

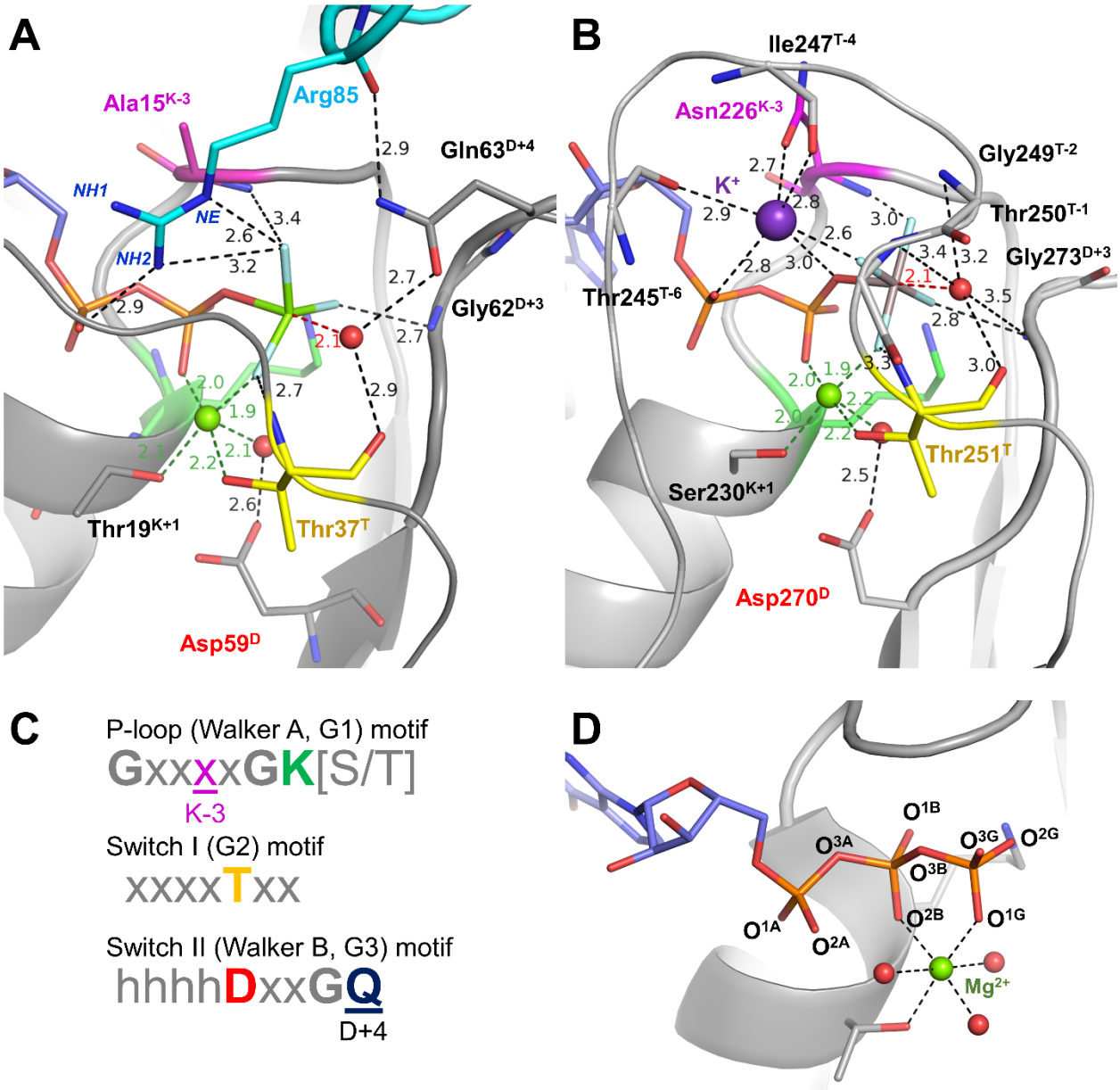


**Figure 3.1.2**. Phosphate chain binding in P-loop NTPases complexes. A. Crystal structure of the transition state analog GDP:$MgF_3^-$ bound to the RhoA/RhoAGAP complex (PDB ID 1OW3 [125]). B. Crystal structure of the transition state analog GDP:$AlF_4^-$ bound to the $K^+$-dependent GTPase MnmE (PDB ID 2GJ8 [28]). Proteins are shown as grey cartoons, conserved residues of signature motifs are shown as sticks and colored as on panel C. $Mg^{2+}$ ions are shown in green, water molecules are shown as red spheres. C. Conserved motifs in G-proteins and other P-loop NTPases. As an example, motifs in Rho GTPase and related G-proteins are shown. D. Naming of atoms according to IUPAC recommendations for nucleoside triphosphates [110] and typical $Mg^{2+}$ coordination

It is customary to refer to the conformations of P-loop fold NTPases in the presence of their activating partners with stimulating moieties inserted as "catalytically active" conformations [106, 117, 126, 127]. Correct identification of the active (catalytically productive) conformation in distinct P-loop fold NTPases is important because enzymes attain these conformations

immediately before reaching their elusive transitional states so that the active conformations are usually used as a starting point in the QM modeling, see [128] for a recent review. Important hints for clarifying the catalytic mechanism of P-loop NTPases are provided by their structures with bound transition state analogs, such as NDP:$AlF_4^-$ or NDP:$MgF_3^-$ or NDP:$VO_4^-$ complexes [28, 123, 125, 129-134]. The structures with NDP:$AlF_4^-$ or NDP:$MgF_3^-$ bound revealed a "catalytic" water molecule $W_{cat}$ near the $P^G$ atom and almost in line with the bond between $O^{3B}$ and $P^G$ atoms, see Figure 3.1.2A, B. In the NDP:$VO_4^-$ complexes, one of the four oxygen atoms of vanadate occupies the position of the catalytic water molecule [133, 135]. These structures, in support of earlier suggestions, indicate that the ultimate cleavage of γ-phosphate is triggered by the apical nucleophilic attack of a polarized water molecule/hydroxyl $W_{cat}$) on the terminal $P^G$ phosphorus [4, 121-123, 134, 136-140].

### *3.1.2. Cation-dependent P-loop Nucleoside Triphosphatases*

Some P-loop NTPases functionally depend on monovalent cations instead of the Arg/Lys fingers (see Figure 3.1.3, Table 3.1). Many of these GTPases are functionally associated with ribosomes. These include the tRNA modification GTPase MnmE [28], Nug1 GTPase, essential for 60S subunit assembly and nuclear export [31]; RbgA, the ribosome biogenesis GTPase A [32]; YjeQ/ RsgA, the ribosome-associated GTPase of unknown physiological function [141]; YchF, a universally conserved ribosome-binding ATPase [35]; HflX, the ribosome-associated GTPase possibly involved in ribosome assembly [36]; Era (Escherichia coli Ras-like protein), a highly conserved GTPase involved in ribosomal biogenesis [36]; EngA/Der, a protein involved in ribosome biogenesis with two GTPase domains, each with $K^+$-dependent activity [37, 38]; EngB/YsxC, a ribosome biogenesis GTPase [36]; YqeH, GTPase crucial for the maturation of small (30S) ribosomal subunit [39]; and Drg1, a GTPase that interacts with translating ribosomes [142]. These and other cation-dependent P-loop proteins are listed in Table 3.1. Notably, many ubiquitous translation factors require $K^+$ or $NH_4^+$ ions for both intrinsic and ribosome-dependent GTPase activity [18-21, 33, 143-145]. The $K^+$-dependence of multiple ATPases and GTPases of ancient families prompted phylogenetic analysis, which indicated that the $K^+$-dependence was an ancestral trait, lost in some families that employ arginine or lysine fingers instead [7, 8].

Crystal structures with potassium cation bound in the active site are available for two such GTPases, namely MnmE – PDB ID 2GJ8 [28] and FeoB – PDB ID 3SS8 [146]). As seen in Figure 3.1.3, the $K^+$ ion occupies the same position as the positively charged amino group of the Arg/Lys

finger in other TRAFAC NTPases, interacting with the oxygen atoms of phosphate groups. Characteristic features of $K^+$-dependent NTPases include an elongated Switch I region (the K-loop) and two conserved Asn residues, $Asn^{K-3}$ and $Asn^{K+5}$ (see [7, 29, 39] and Figure 3.1.3). The side chain of $Asn^{K-3}$ serves as the major $K^+$ ligand whereas $Asn^{K+5}$ stabilizes the K-loop. The K-loop provides two backbone oxygen atoms (from T-4 and T-6 residues) as further coordinating ligands for the $K^+$ ion [28, 29, 36, 39, 146]. Another common feature of $K^+$-dependent NTPases is the absence of a polar, "catalytic" residue; they belong to HAS NTPases (from Hydrophobic Amino Acid Substituted) [147]. In MnmE and FeoB GTPases, catalytic water molecules are stabilized by the backbone carboxyl of signature Thr residue of Switch I (K-loop) and the backbone amino group of the $Gly^{D+3}$ residue of Switch II.

In most $K^+$-dependent P-loop NTPases, $Na^+$ ions cannot functionally replace $K^+$ ions [32, 35, 38, 148-150]. The dependence on $K^+$ ions of multiple ubiquitous NTPases and the translation system are often credited as one of the major reasons for the cells to maintain $[K^+]/[Na^+]$ ratio >>1.0 in cytoplasm [7]. Since natural habitats usually have more $Na^+$ than $K^+$, cells can spend up to half of their energy to maintain the necessary $[K^+]/[Na^+]$ ratio [151]. It was suggested that the first cells originated in $K^+$-rich environments, which would explain the $K^+$-dependence of the ancient cellular systems [7, 8]. However, it is unclear why the cellular machinery did not switch its preference from $K^+$ to $Na^+$ over the course of evolution, despite the abundance of $Na^+$ in various habitats [152]. The switch to Na-dependent systems instead of K-dependent would have been extremely beneficial, particularly for the marine organisms, since they must expend a lot of energy to deal with the $[K^+]/[Na^+]$ ratio of around 0.02 in sea water [153]. For P-loop NTPases, the adaptation to $Na^+$ ion as a stimulating cofactor is, in principle, possible: NTP hydrolysis in certain dynamins and dynamin-like proteins are stimulated by $Na^+$ as well as by $K^+$ ions [30, 154]. The structures of such proteins show that $Na^+$ ions bind similarly to the $K^+$ ions in $K^+$-dependent NTPases [29]. Thus, the pervasive preference for $K^+$ ions in other NTPases remains unclear.

Reactions dealing with phosphoanhydride bonds exhibit a preference for $K^+$ ions even in the absence of proteins. As early as in 1960, larger monovalent cations, specifically $Rb^+$ and $K^+$, were shown to stimulate the transphosphorylation reaction more efficiently than similar ions of the smaller size, such as $Na^+$ and $Li^+$ ions [41]. Notably, even the $NH_4^+$ ions have this effect, stimulating transphosphorylation on par with $K^+$ ions, more effectively than smaller ions. These findings suggest that the stimulating effect of the positively charged Arg/Lys fingers and $K^+$ ions could have a common mechanism and be defined mainly by the cation size.

**Table 3.1.** Monovalent cation dependence of P-loop GTPases and ATPases

| Protein name | UniProt ID | Cation dependence | Refs. |
|---|---|---|---|
| **TRAFAC class** | | | |
| Dynamin-1 | DYN1_HUMAN | $K^+$>$Na^+$ | [30] |
| Dynamin-related protein 1A | DRP1A_ARATH | $K^+$, $Na^+$ | [154] |
| GTPase Nug1 | G0SEW3_CHATD | $K^+$>$Na^+$ | [31] |
| Ribosome biogenesis GTPase A | RBGA_BACSU | $K^+$, no $Na^+$ | [32] |
| Ribosome biogenesis GTPase RsgA(YjeQ) | RSGA_ECOLI | $K^+$ | [141] |
| Elongation Factor Tu, *E. coli* | EFTU1_ECOLI | $K^+$>$Na^+$ | [18] |
| Elongation Factor Tu, *H. marismortui* | EF1A_HALMA | $K^+$>$Na^+$ | [33] |
| Eukaryotic translation initiation factor 5B | IF2P_CHATD | $Na^+$, $K^+$ | [21] |
| Initiation factor IF-2 | IF2_ECOLI | $K^+$ | [20] |
| tRNA modification GTPase MnmE | MNME_ECOLI | $K^+$, no $Na^+$ | [28] |
| Ferrous iron transporter B | Q5M586_STRT2 | $K^+$, no $Na^+$ | [34] |
| Ribosome-binding ATPase YchF | YCHF_ECOLI | $K^+$, no $Na^+$ | [35] |
| GTPase HflX* | HFLX_BACSU | $K^+$ | [36] |
| GTPase Era | ERA_BACSU | $K^+$, no $Na^+$ | [36] |
| GTP-binding protein EngA** *B. subtilis* | DER_BACSU | $K^+$, no $Na^+$ | [36, 37] |
| GTP-binding protein EngA** *T. maritima* | DER_THEMA | $K^+$, no $Na^+$ | [38] |
| NO-associated protein 1 | NOA1_ARATH | $K^+$ | [149] |
| Ribosome Assembly GTPase YqeH | YQEH_BACSU | $K^+$, no $Na^+$ | [39] |
| Developmentally-regulated GTP-binding protein 1 | DRG1_HUMAN | $K^+$ | [142] |
| GTP-binding protein EngB | ENGB_BACSU | $K^+$* | [36] |
| Human GTPBP3 | GTPB3_HUMAN | $K^+$ | [155] |
| **RecA-like family** | | | |
| Human meiotic recombinase Dmc1 | DMC1_HUMAN | $K^+$ | [156] |
| Human DNA repair protein RAD51 | RAD51_HUMAN | $K^+$ | [157] |
| | | $K^+$, no $Na^+$ | [40] |
| Yeast DNA repair protein RAD51 | RAD51_YEAST | $K^+$ | [158] |
| DNA repair protein RadA from *M. voltae* | RADA_METVO | $K^+$ | [159] |
| DNA repair protein RadA from *M. maripaludis* | RADA_METMI | $K^+$, no $Na^+$ | [160] |

In the 'Cation dependence' column, '$K^+$' indicates that only $K^+$-dependence has been shown; '$K^+$, no $Na^+$' indicates hydrolysis stimulation by $K^+$ ions and a lack of stimulation by $Na^+$ ions; '$K^+$>$Na^+$' denotes more effective stimulation by $K^+$ than by $Na^+$ ions; '$K^+$, $Na^+$' and '$Na^+$, $K^+$' is used when both cations have similar effects, with the more effective one listed first.

* The GTPase activity was measured at the same concentrations of KCl and NaCl of 200 mM, and for some proteins (the second GTPase domain of EngA, HflX, EngB, all from *B. subtilis*), the lack of stimulation by cations has been reported [36]. However, higher concentrations of ions may be required for these proteins in the absence of their activating partners, as has been shown for the second GTPase domain of EngA [37].

** This protein has two P-loop GTPase domains, activity measurements were reported for the whole protein.

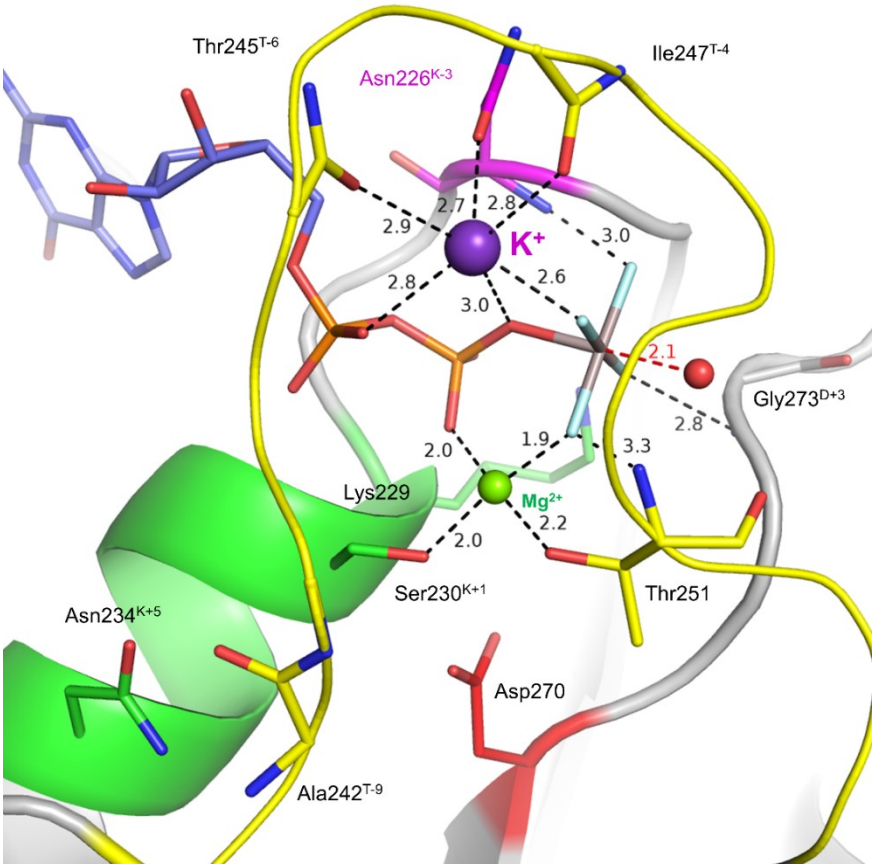


**Figure 3.1.3.** The active site of cation-dependent GTPase MnmE (PDB ID 2GJ8). Protein surrounding the NTP-binding site is shown as a cartoon, functionally relevant residues are shown as sticks, water and cations are shown as spheres: water in red, Mg in green, K in purple. The phosphate chain is shown as sticks with oxygen atoms in red and phosphorus in orange; γ-phosphate mimicking group $AlF_4^-$ is shown in gray and light blue. P-loop lysine and following helix are shown in green, $Asn^{K-3}$ is shown in purple, Switch I/K-loop is shown in yellow, Walker B motif aspartate is shown in red. All distances measured in Å.

### *3.1.3. Translation Factors*

Most known cation-dependent P-loop NTPases belong to the TRAFAC class, named after the translation factors. Protein synthesis on a ribosome is assisted by a plethora of proteins, many of which contain GTPase domains with a characteristic P-loop fold and comprise several protein families within the TRAFAC class, see [127, 161-171] for reviews.

Four translation factors are ubiquitous in bacteria, archaea, and eukaryotes. One of these factors is the elongation factor EF-Tu (named eEF1A in eukaryotes and aEF1A in archaea) that, in its GTP-bound state, is able to bind an aminoacyl-tRNA (aa-tRNA) by clasping it between the GTPase domain from one side and the C-terminal domain from the other side. The resulting ternary

complex binds to the ribosome; the GTP hydrolysis, however, takes place only if the aa-tRNA corresponds to the mRNA codon. One more ubiquitous elongation factor EF-G (eEF2 in eukaryotes, aEF2 in archaea) promotes the further translocation of tRNA and mRNA. Two other ubiquitous factors have slightly differing functions in bacteria and archaea/eukarya, respectively. In bacteria, the initiation factor IF2 recruits the Met-tRNA to the ribosome whereas its orthologs eIF5B in eukaryotes and aIF5B in archaea promote binding of initiator tRNA to the small ribosomal subunit (SSU) and its subsequent assembly with the large ribosomal subunit (LSU) into a ribosome. Accordingly, in archaea and eukaryotes, eIF-2 recruits Met-tRNA to the ribosome whereas its bacterial ortholog SelB is specialized in the delivery of selenocysteine (Sec)-tRNA to the ribosome [167, 172]. In archaea and eukaryotes, the latter function is performed by EFsec, an apparent paralog of EF1A [172, 173]. Similar GTPase domains are also found in a plethora of domain-specific translation factors, such as bacterial EF4, LepA, BipA, RF3, archaeal EF-1α, eukaryotic eRF3, Hbs1, and even viral translational GTPases [165, 166, 174].

Translational GTPases are believed to have common ancestry [24, 171]. The four ubiquitous factors are attributed to the Last Universal Cellular Ancestor (LUCA) [22, 171, 175]. Their genes stem from gene duplication events that preceded the LUCA [171, 176, 177].

By analogy with other P-loop fold NTPases, it was initially speculated that some Arg residues might be involved in the hydrolysis stimulation in translation factors [178]. However, the first crystal structure of the large ribosomal subunit revealed no Arg residues in the vicinity of the GTP molecule [179]. Instead, the structure indicated that the activation of elongation factors EF-Tu and EF-G might be mediated by one of the RNA loops of the large ribosomal subunit, namely the so-called sarcin-ricin loop (SRL) [179]. SRL is one of the longest universally conserved ribosomal RNA sequences (nucleotides 2653–2667, in *E. coli* numbering) [180]. The SRL loop interacts with the GTP-binding site of translation factors immediately before GTP hydrolysis [127, 161, 167, 169, 179, 181, 182], see Figure 3.1.4.

Ramakrishnan and colleagues scrutinized the decoding-related structural changes in the EF-Tu from *Thermus thermophilus* [127]. They showed that the ternary complex, formed by EF-Tu, Mg-GTP, and aa-tRNA, binds to the "shoulder" of SSU. The recognition of cognate codon takes place on the same SSU, but some 80 Å away from the GTP-binding site of EF-Tu. Recognition of the proper, "cognate" codon triggers a chain of conformational changes, including conformational changes in the aa-tRNA and a "closing" motion of the SSU shoulder which moves towards the LSU and shifts the GTPase domain of EF-Tu by 8 Å enabling its binding to the SRL of the LSU. The phosphate group of the invariant A2662 residue of the SRL forms a hydrogen bond (H-bond) with a $His^{D+4}$ residue ($His85^{D+4}$ in *T.thermophilus,* $His84^{D+4}$ in *E.coli*) of EF-Tu,

which corresponds to the "catalytic" $Gln^{D+4}$ in other TRAFAC GTPases, and orients the side chain of $His^{D+4}$ so that an H-bond is formed between $His^{D+4}$ and a water molecule in the pre-attack position [127]. In this way, recognition of the cognate codon enables the GTP hydrolysis, which drives the translocation of the aa-tRNA towards the peptidyl transferase center. This coupling ensures that the energy of GTP hydrolysis is released when it could be properly used. A similar relation between codon recognition, the interaction of EF-Tu with SRL, and activation of the catalytic water molecule was envisioned also for other translational GTPases [127, 161, 167, 169, 181-183].

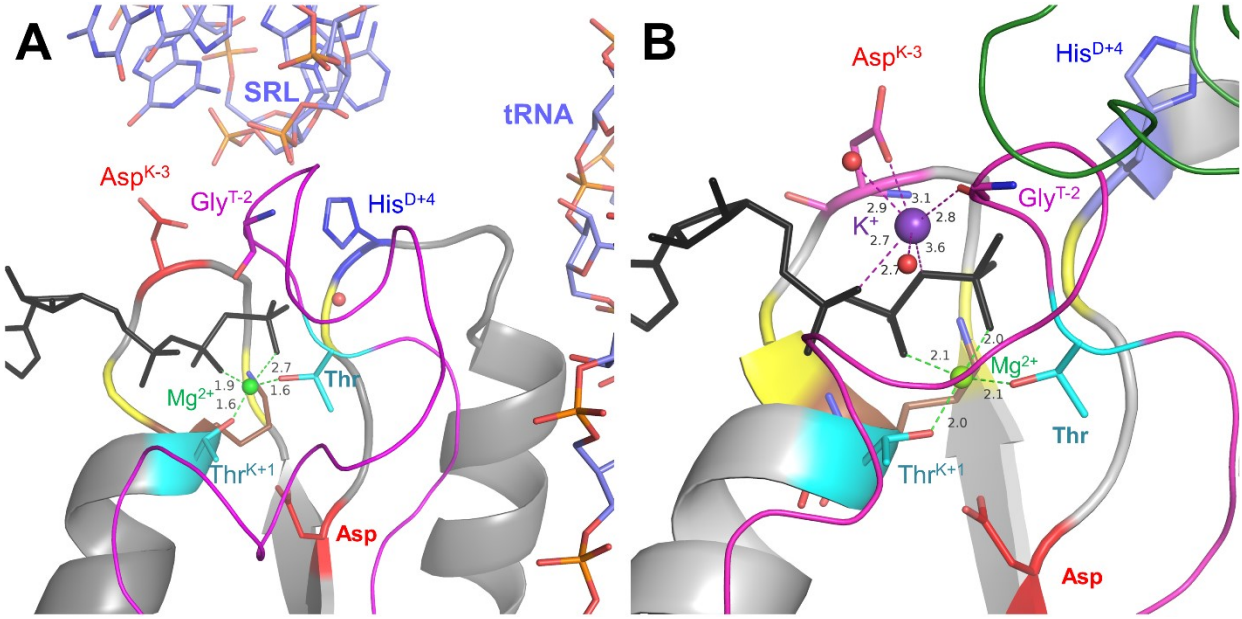


**Figure 3.1.4.** The active site of translation factors. A. Translation factor EF-Tu bound to tRNA and ribosome (PDB ID 4V5L). B. Translation initiation factor eIF5B (517-858 fragment) (PDB ID 4TMZ). Proteins surrounding the NTP-binding site are shown as a cartoon, functionally relevant residues are shown as sticks, water and cations are shown as spheres: water in red, Mg in green, K in purple. Non-hydrolyzable NTP analogs are shown as black sticks. P-loop lysine is shown in brown, signature residues of cation-dependent G-proteins K-3 and Switch I loops are shown in magenta, conserved glycine residues are shown in yellow, conserved threonine residues are shown in cyan, Walker B motif aspartates are shown in red. RNA fragments (tRNA and SRL loop) are shown as blue sticks, with oxygen and phosphorus atoms colored red and orange, respectively. All distances measured in Å.

These seminal findings, however, did not clarify how the fast GTP hydrolysis could take place without positively charged stimulating moiety that interacts with the phosphate chain oxygen atoms in all other classes of P-loop fold NTPases [115, 119, 121]. Indeed, the backbone of the SRL is negatively charged. According to Adamczyk and Warshel, the minor reorientation of $His^{D+4}$ is unlikely to provide the same electrostatic impact as the insertion of an external Arg or Lys finger [184]. And anyway, the $His^{D+4}$ residue does not reach the phosphate chain [127, 167,

183]. Hence, the counterpart of a stimulating Arg or Lys finger in the EF-Tu and other translation GTPases has remained obscure.

Even earlier, it was noted that the $K^+$-binding $Asn^{K-3}$ residue of the $K^+$-dependent GTPases has an Asp as a counterpart in translational GTPases [7]. In $K^+$-dependent GTPase YqeH, the replacement of $Asn^{K-3}$ with $Asp^{K-3}$ did not affect the enzyme activity [39]. Translation factors have been repeatedly shown to have $K^+$ or $NH_4^+$ -activated GTPase activity starting from the 60s [18-20, 143-145, 185-187], see also Table 3.1. Based on these observations, the available structure of EF-Tu [127], and sequence comparison, translational GTPases were suggested to be $K^+$-dependent GTPases and a $K^+$ ion could be coordinated by $Asp^{K-3}$, $Gly^{T-2}$, and phosphate oxygen atoms of GTP [7].

More recently, Kuhle and Ficner succeeded in crystallizing the GTPase domain of eIF5B with $K^+$ or $Na^+$ ions bound, see Figure 3.1.4B, and [21, 188]. These particular structures were obtained with a truncated protein that formed non-physiological dimers in the crystal, so that the Switch I loop of one protein was stabilized by a similar adjoining protein; in these non-physiological dimers, the $His^{D+4}$ residue was turned away of the phosphate chain, see Figure 3.1.4B and [21]. Still, this protein-protein interaction - by chance - shaped Switch I in a such way that it could bind a monovalent cation, see [21, 188] for details. These authors also showed that GTP hydrolysis by free eIF5B and aEF1A (archaeal analog of EF-Tu) was stimulated by monovalent ions, whereby the hydrolysis by aEF1A was specifically stimulated by $K^+$ ions but not $Na^+$ ions. Based on their data, Kuhle and Ficner suggested that monovalent cations serve as structural and catalytic cofactors in translational GTPases [21]. In the structures of eIF5B, the monovalent ions were coordinated by $O^{2A}$, $O^{3B,}$ and $O^{3G}$ atoms of the phosphate chain, the side chain of the $Asp533^{K-3}$ residue of the P-loop, and the carbonyl oxygen of $Gly555^{T-2}$ of the Switch I loop [21], in exact accordance with the earlier prediction for EF-Tu [7].

With free EF-Tu from *E. coli,* it was shown that its intrinsic GTPase activity (in the absence of ribosomes and tRNA) was manifold stimulated by $K^+$ ions in the WT, but not in the mutant where the $Asp21^{K-3}$ residue was replaced by alanine [21, 189]. In the case of programmed (protein-synthesizing) ribosomes, the $Asp21^{K-3}$ to Ala mutation caused a 640-fold drop in the GTP hydrolysis rate, whereas the $Asp21^{K-3}$ to Asn mutations slowed the reaction only by the factor of ten [189].

Still, even despite these spectacular data, the idea of translation factors being $K^+$-dependent GTPases [7, 21] has not obtained common acceptance. Specifically, Rodnina and colleagues argued that the $K^+$ ion was not seen next to the phosphate chain in available ribosome structures. They suggested that $Asp^{K-3}$ and $His^{D+4}$ residues and water molecules in the catalytic pocket form

an H-bonded network so that "the million-fold rate acceleration brought about by the ribosome is achieved solely by electrostatic stabilization and shielding effects, which were collectively denoted as “allosteric”" (quoted from [189]). In the most recent comparative analysis of the mechanisms of the EF-Tu and EF-G, it was noted that the activation of these GTPases involves a structural allosteric effect; the nature of this effect was not specified [190].

Not only in the case of translation factors, but also in general there is no consensus on how the interaction with an activating partner initiates the attack by $W_{cat}$. Specifically, some authors emphasized, based on structural analyses and mutant studies, that the insertion of the stimulating group leads to reorientation of a “catalytic” polar residue towards $W_{cat}$ and its participation in the catalysis as a base that accepts a proton from $W_{cat}$, see [27, 140] for reviews. Jin and colleagues coupled the activation with the reorganization of the hydrogen-bonded (H-bonded) network in the catalytic site and increasing the nucleophilic properties of $W_{cat}$ by positively charged stimulating Arg residue [123, 191, 192]. Warshel and colleagues argued that the common feature in P-loop NTPases is the electrostatic stabilization of the transition state by cationic stimulating moieties with the catalytic residue(s) playing only a secondary role of positioning $W_{cat}$ [122, 193]. Gerwert and colleagues proposed, for small GTPases, that the stimulating Arg finger rotates the α-phosphate towards an eclipsed conformation with respect to β- and γ-phosphates, which would destabilize the triphosphate chain [116, 194, 195]. Important hints are provided by protein structures that were obtained in the presence of the NDP:$AlF_4^-$ or NDP:$MgF_3^-$ complexes. In such structures, a water molecule is usually seen in the apical "attack" position next to the Al/Mg atom, so that the NDP:$AlF_4^-$ or NDP:$MgF_3^-$ complexes are believed to be the most exact mimics of the transition states [123, 131, 134, 196].

## 3.2. Results and Discussion

### 3.2.1. Interactions of Monovalent Cations with NTPs in Water

The preference for large monovalent cations is an attribute not only of P-loop NTPases but of the ATP transphosphorylation reaction that occurs in solution in the absence of proteins [41]. Notably, both $K^+$ and $NH_4^+$ ions have a stronger effect on the reaction, than the smaller $Na^+$ ions [41]. This indicates that the catalytic effect is mostly determined by a positively charged sphere of a certain size, rather than by intrinsic properties of metals. Thus, molecular dynamics simulations can be applied to this system, to investigate the binding of different monovalent cations to the phosphate chain of NTPs and explain the effect cation of different sizes may have on the hydrolysis reaction.

Molecular dynamics (MD) simulations were conducted, describing the behavior of the $Mg^{2+}$-ATP and $Mg^{2+}$-GTP complexes (hereafter Mg-ATP and Mg-GTP). The simulations were conducted separately in pure water and in water with the addition of $K^+$, $NH_4^+$, or $Na^+$ ions (hereafter $M^+$), see Chapter 2.1.2 for details. For the initial structure for the MD simulations, the conformation of the Mg-ATP complex was generated, with the $Mg^{2+}$ ion placed between $O^{1B}$ and $O^{2G}$ oxygens of the β- and γ-phosphate groups and four water molecules placed manually around $Mg^{2+}$ to complete its coordinational sphere (see Figure 3.1.2D). The chosen mode of $Mg^{2+}$ coordination is referred to as bidentate or βγ coordination. Such coordination was reported for the Mg-ATP complex in water by NMR studies [197-200]; and can be seen in experimental structures of P-loop NTPases with NTPs or NTP-like molecules (see [108, 117, 201, 202] and Figure 3.1.2). Before the MD simulations, the Mg-ATP complex conformation was optimized in a vacuum with the PM3 Hamiltonian. To prepare the simulation cell, the optimized structure was placed in a box, then six $M^+$ ions, 4 $Cl^-$ ions, and 1200 water molecules were added. The total ionic strength of the final solution was 0.2 M.

Since the P-loop NTPase structures contain specifically the βγ conformation of the Mg-ATP complex, to gather enough data on this specific conformation additional MD simulations were conducted, in 25 independent 20-ns long runs, with and without $M^+$ ions. The simulations were conducted for both Mg-GTP and Mg-ATP complexes (Table 2.1). The conformations sampled from MD simulations were used to analyze the shapes attained by the phosphate chain depending on the presence of particular $M^+$ ions. Generally, the results for the ATP and GTP complexes were very similar, therefore hereafter only the Mg-ATP data is described. The corresponding data for Mg-GTP complexes are shown in Supplementary Figures S1-S2.

#### 3.2.1.1 Binding of monovalent cations to $Mg^{2+}$-NTP complexes in water

MD simulation data was used to obtain distance distributions, which reveal that $M^+$ ions can form different coordination bonds with ATP phosphate chain oxygen atoms. The respective lengths of these bonds were 2.2 Å for $Na^+$, 2.6 Å for $K^+$, and 2.7 Å for $NH_4^+$ ions (Figure 3.2.2). These distances match well with the distances observed in crystallographic structures [203-205]. Cation binding was also evaluated by calculating the number of ions within the 4 Å distance of the phosphate chain, on time average. In the case of both $Na^+$ and $NH_4^+$, 1.5 cations were present, and in the case of $K^+$ - 0.75 cations (see Figure 3.2.1). For each of the three ions, the first interaction occurs at distances shorter than 4 Å and a less prominent second interaction is at around 6 Å. Monovalent cations can be seen interacting with all free oxygens of the phosphate chain (Figure 3.2.2). The phosphate chain contains two ester bond oxygens, but only the $O^{3B}$ between β- and γ-phosphates is involved in interactions with cations. This atom interacted less often with $NH_4^+$ than with $K^+$ and $Na^+$. All $M^+$ ions occur more often near oxygen atoms of γ-phosphate than near β- or α-phosphates (Figure 3.2.2). The peak distances from the cations to the oxygens were the same 2.7 Å for $K^+$ and $NH_4^+$ ions, and for $Na^+$ the peak occurs at 2.2 Å (Figure 3.2.2).

Two distinctive binding sites for $M^+$ ions were identified from the radial distributions of ions around specific oxygen atoms (Figure 3.2.2) and the superposition of the system conformations, sampled from MD simulations (Figure 3.2.3). Cations in the first site are coordinated by the oxygen atoms of β- and γ-phosphates, and the second site - by oxygens of α- and γ-phosphates. Hereafter these binding sites are referred to as BG and AG sites, correspondingly. An additional site of $M^+$ ions binding can be seen at the distal end of the ATP triphosphate chain, as the ions are coordinated by an oxygen atom(s) of the γ-phosphate (the G site(s), Figure 3.2.3). This aggregation of ions near the γ-phosphate is referred to as the G site.

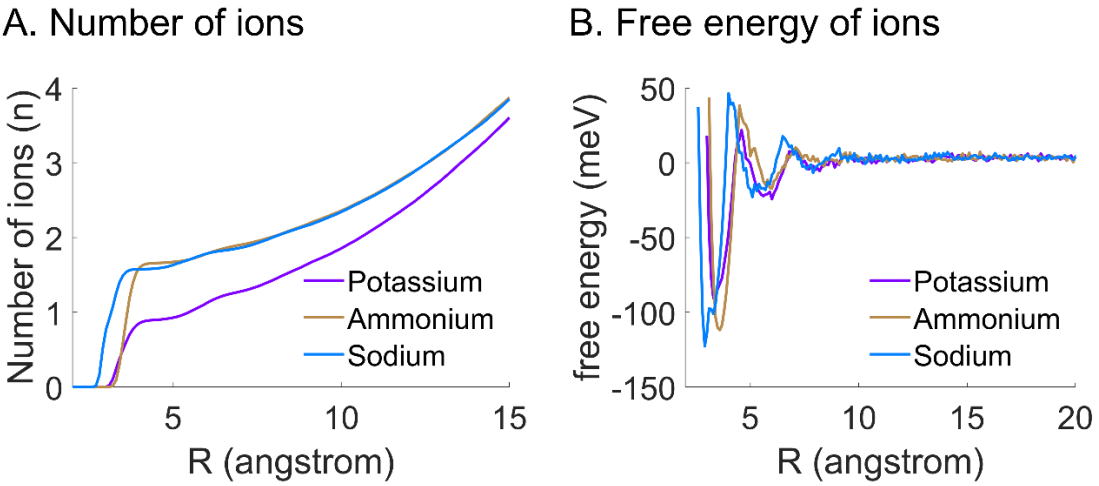


**Figure 3.2.1.** Quantitative evaluation of cations binding to the ATP in the MD simulations. A. Probability distribution functions plotted for $M^+$ ions around the phosphate chain. The plots show the number of ions within an area around NTP phosphorus atoms as a function of the area radius. B. Free energy of the $M^+$ binding plotted as a function of the distance from the phosphate chain, estimated from the probability data shown on panel A.

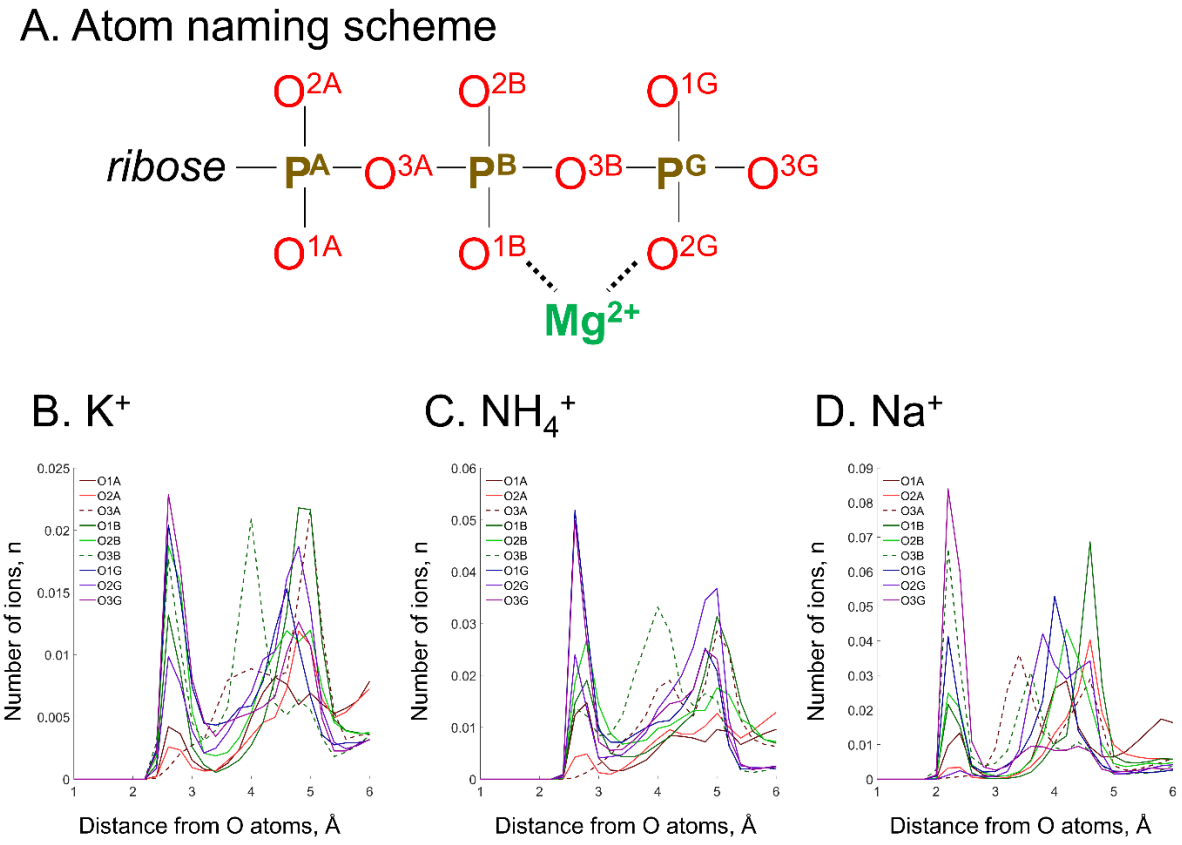


**Figure 3.2.2.** Radial distribution of cations around phosphate chain oxygen atoms.
A. Atom names, according to the CHARMM naming scheme [49] and the IUPAC recommendations [110]. B-D. Radial distributions of $K^+$, $NH_4^+$, and $Na^+$ ions, respectively around specific oxygen atoms. Dashed lines show graphs for the oxygen atoms of ester bonds ($O^{3A}$ and $O^{3B}$). For the $NH_4^+$ ions, the distances were measured to the nitrogen atom.

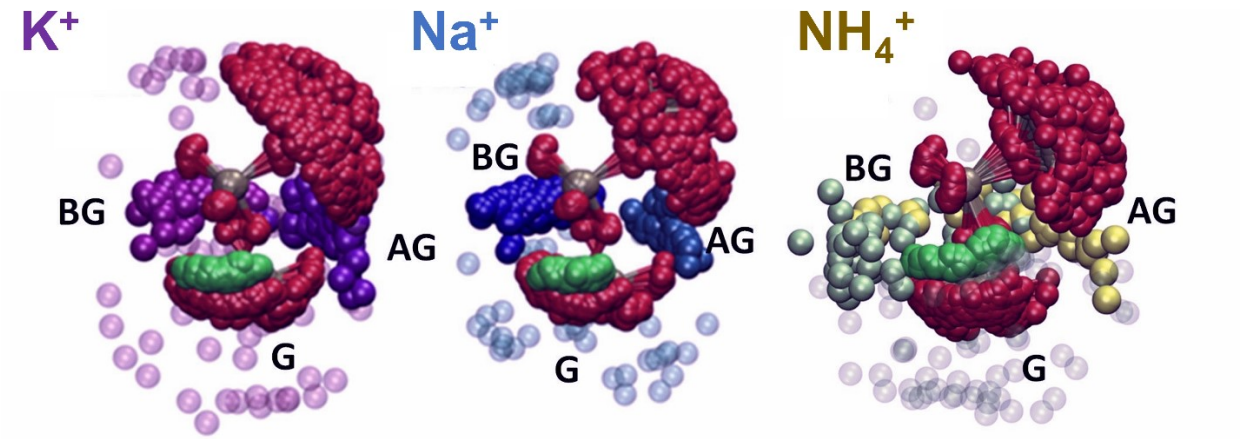


**Figure 3.2.3.** ATP phosphate chain conformations sampled from the MD simulations, superposed. The triphosphate chain is shown with α-phosphate on the top and γ-phosphate on the bottom. Cations are colored differently to differentiate between the AG and BG sites. Ions outside these sites are shown as transparent spheres. Conformations were sampled with a 5-ns interval and superposed to align the positions of the phosphate chain atoms.

To describe $M^+$ binding separately in the AG site and the BG site, the distances were measured between each $M^+$ ion and the nearest oxygen atom of α-, β-, and γ-phosphates (see $R^{AG}$ and $R^{BG}$ in Figure 3.2.4). The occupancy of each site was calculated, as shown in Figure 3.2.4B–D, as the number of $M^+$ ions present in or near each binding site on time average. All ions exhibit a prominent peak in the $R^{BG}$ distribution. Notably, the $R^{BG}$ values peak at the same distances as the distances to individual oxygens (Figure 3.2.2B-D). This shows that the $M^+$ ions in the BG site formed coordination bonds simultaneously with two oxygen atoms. In the AG site, both $NH_4^+$ and $Na^+$ ions exhibit peaks in the $R^{AG}$ distribution with the maxima at 2.7 Å and 2.3 Å, matching the values for individual distances as well. The same is true for $K^+$ ions, however, the peak in a $R^{AG}$ value of 2.6 Å was wide, as compared to other cations. Still, the distance distributions plotted for individual oxygen atoms show that oxygens of γ-phosphate had the most frequent contact with $K^+$ ions, see Figure 3.2.2.

While $M^+$ ions occupy the same binding sites, they exhibit different affinity, in decreasing order: $Na^+ > NH_4^+ > K^+$ (Table 3.2). This corresponds to experimental studies, showing that ATP affinity to $Na^+$ ions is higher than to $K^+$ and $NH_4^+$ ions, although the monovalent cation affinity data is only available for ATP without $Mg^{2+}$ (Table 3.2). Occupancy observed in the AG site was much lower than in the BG site for all $M^+$ ions; on time average the occupancy of the BG site was calculated as 0.95 for $Na^+$, for $NH_4^+$ it was 0.72, and 0.5 for $K^+$, while for the AG site it was 0.15 for $Na^+$, 0.2 for $NH_4^+$, and 0.05 for $K^+$ (see Figure 3.2.4B–D).

In the MD simulations of $Mg^{2+}$-GTP complexes, monovalent cation binding exhibits the same features as in the simulations of $Mg^{2+}$-ATP complexes, see Supplementary Figure S1.

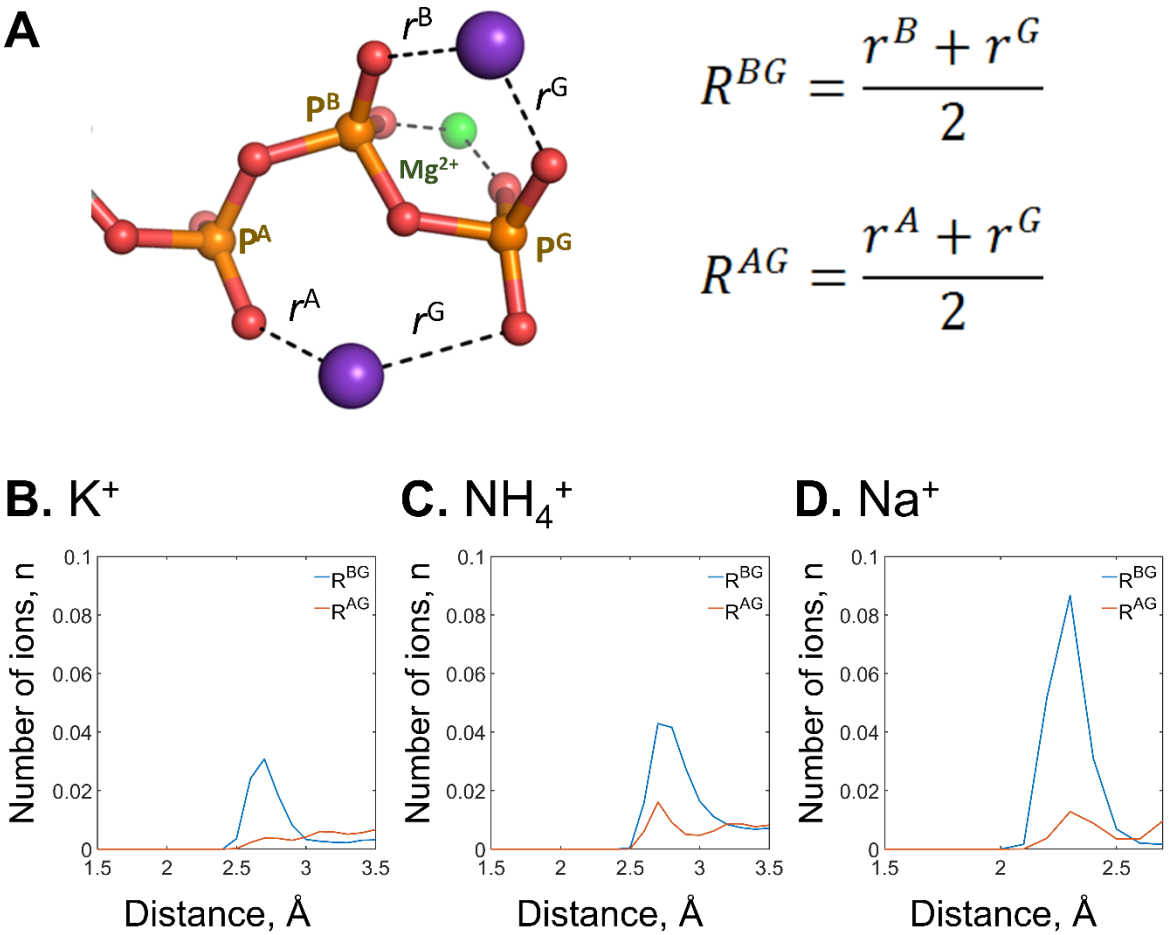


**Figure 3.2.4.** Characterization of cation binding sites on the ATP phosphate chain. A. Mg-ATP complex with two $M^+$ ions bound, in AG and BG sites, respectively. Distances from $M^+$ ions to the binding sites (hereafter $R^{AG}$ and $R^{BG}$, respectively) were estimated as an average of two distances to the corresponding oxygens, as shown on panel A. B-D. $R^{AG}$ and $R^{BG}$ distributions for potassium, ammonium, and sodium ions, respectively.

**Table 3.2.** Experimental data on monovalent cation binding to ATP and $Mg^{2+}$-ATP complex.

| Cation | Ionic radius (Å)[b] | Stimulation of transphospho-rylation, %, [a] | Binding to ATP in the absence of $Mg^{2+}$ ($\log(K_B)$), 25°C | | | Binding to Mg-ATP ($\log(K_B)$)[f] |
|---|---|---|---|---|---|---|
| $Na^+$ | 1.02 | 28 | 1.31±0.03[c] | 1.989±0.007[d] | 1.93[e] | 2.76 |
| $K^+$ | 1.38 | 64-73* | 1.17±0.03[c] | 1.873±0.005[d] | 1.99[e] | 0.88 |
| $NH_4^+$ | 1.44 | 27 | N/A | | | 1.76 |

* measured for different salts: 64% with KCl and 73% with $K_2SO_4$.
*a* –data from [41]; stimulation of transphosphorylation by 100 mM $M^+$ in the presence of 50 μM $MnCl_2$; *b* – data from [205] ; *c* – data from [206] ; *d* – data from [207]; *e* – data from [208]; *f* – calculated from MD simulations

Comparatively weak $K^+$ binding in the AG site can be explained by thermodynamic and structural analysis of the phosphate chain shape in the Mg-NTP complex with two $K^+$ ions bound. However, this analysis was hindered by the low population of such complexes (with two $K^+$ ions

bound) in the simulations. Therefore, additional MD simulations were conducted with the cations fixed by positional restraints (see the full list of all simulations in Table 2.1). These simulations consisted of 10-ns simulations of either Mg-ATP complex in the βγ coordination and $K^+$ ion placed in the BG site; or Mg-ATP complex in the βγ coordination with two monovalent cations, placed one in each site. All $K^+$ and $Mg^{2+}$ ions and the N1 atom of the adenine base were secured in their relative locations by positional restraints. The simulations show that when the second $K^+$ ion binds in the AG site it stabilizes all the three phosphate groups in a conformation close to fully eclipsed so that the phosphorus-oxygen bonds are almost coplanar between α-, β-, and γ-phosphate groups (see Figure 3.2.5; Table 3.3). Only in such conformation, the distances between the oxygen atoms of α- and γ-phosphates were short enough to allow the second $K^+$ ion binding in the AG site.

**Table 3.3.** Dihedral angles values in the Mg-ATP phosphate chain with $K^+$ ions bound.

| Structure | $\Psi^{\alpha-\beta}$ | $\Psi^{\beta-\gamma}$ | $\Psi^{\alpha-\gamma}$ |
|---|---|---|---|
| Mg-ATP (MD simulation) | +69±31° | +10±25° | N/A* |
| Mg-ATP-$K^+$ (MD simulation) | +23±40° | -4±18° | +9±65° |
| Mg-ATP-2$K^+$ (MD simulation) | +13±24° | -27±8° | +1±26° |

A dihedral angle is defined as an angle between two planes that is defined by four atoms. The dihedral angles between phosphates were defined as follows: $\Psi^{\alpha-\beta} = \angle O^{2A}\text{-}P^{A}\text{-}P^{B}\text{-}O^{2B}$; $\Psi^{\beta-\gamma} = \angle O^{1B}\text{-}P^{B}\text{-}P^{G}\text{-}O^{1G}$; and $\Psi^{\alpha-\gamma} = \angle O^{1A}\text{-}P^{A}\text{-}P^{G}\text{-}O^{3G}$, see also Figure 3.2.5. During the analysis of MD simulation data, the average and standard deviation values for dihedral angles were obtained by fitting the angle distribution histograms with normal functions, using the MATLAB function "fit". All angle distributions were fitted with one-term Gaussians, with the exception of the $\Psi^{\beta-\gamma}$ angle in the case of the Mg-ATP with two $K^+$ ions; this distribution was fitted with two-term Gaussian, and parameters are shown for the highest peak. Distribution histograms and fitted curves are shown in Figure 3.2.5.

* The rotation of α-phosphate is unrestricted and the corresponding dihedral angles can take any values between -180° and 180°.

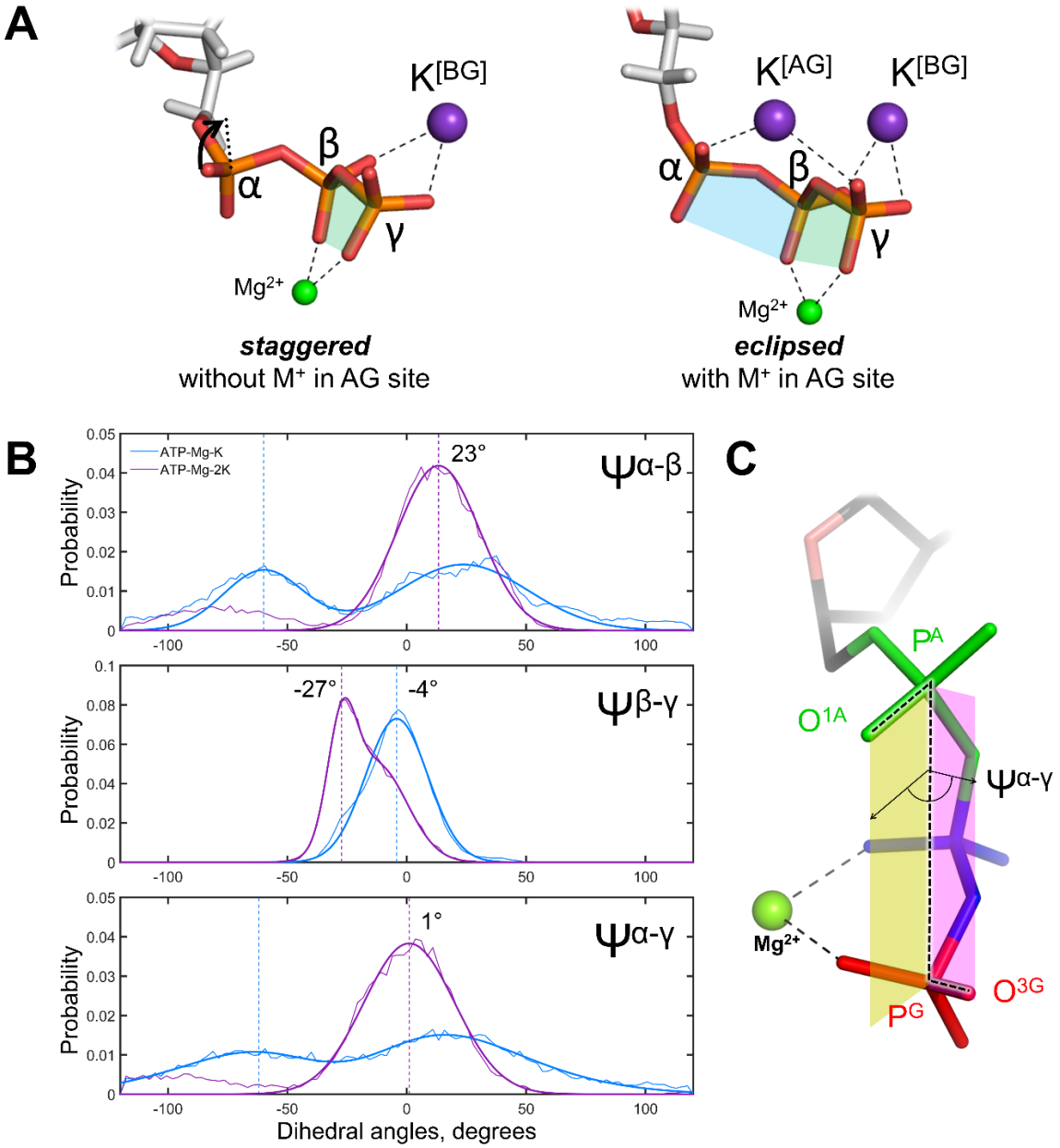


**Figure 3.2.5**. Eclipsed conformation of the phosphate chain, induced by $M^+$ ion binding. A. Mg-ATP complexes bound to one and two $K^+$ ions as sampled from MD simulations. B. Dihedral angle distribution histograms, describing phosphate chain conformations obtained from MD simulations, blue – ATP-Mg-K complex, purple – ATP-Mg-2K complex. Normalized histograms are shown as thin lines, fitted normal distribution functions are shown as thick lines. Dashed lines point to centroid values of each fit. All dihedral angle distributions were fitted with one-term Gaussians, except for the $\Psi^{\beta\text{-}\gamma}$ angle in the Mg-ATP-2K complex, which distribution was fitted with the two-term Gaussian. C. Scheme illustrating the definition of the dihedral angle on the example of $\Psi^{\alpha\text{-}\gamma}$.

### 3.2.1.2 Phosphate chain shape in $Mg^{2+}$-NTP complexes in water as observed in MD simulations

Hydrolysis of the bond between β- and γ-phosphates proceeds via a planar transition complex, so that the $P^B$-$O^{3B}$-$P^G$ angle widens when approaching this transition state [122, 123, 134, 140, 209-212]. Thus, the $P^B$-$O^{3B}$-$P^G$ angle can indicate how susceptible is a particular conformation to the cleavage of this bond. Another important structural characteristic of Mg-ATP complex is the overall shape of the phosphate chain, which could be straight or curled/bent in a bow. The $P^A$-$P^G$ distance was used as an indicator of the general phosphate chain shape – longer distances indicate that the phosphate chain is straight, while shorter distances indicate that the phosphate chain is curled. During the MD simulations, distances $P^A$-$P^G$ and angles $P^B$-$O^{3B}$-$P^G$ attain values that indicate several distinct conformations of the Mg-ATP complex (Figure 3.2.7). These conformations correspond to different types of $Mg^{2+}$ ion coordination by the triphosphate chain (see Figure 3.2.6). The ATP molecules were observed switching between the βγ conformation and the tridentate so-called αβγ conformations. In the latter, the $Mg^{2+}$ ion is coordinated by three oxygens of the phosphate chain, one from each phosphate. This type of $Mg^{2+}$ coordination, as well as the bidentate βγ coordination, was reported in $^{31}$P NMR studies [199, 213] and can be seen in some protein structures [214, 215]. In the MD simulations reported here, slightly different variations of the αβγ conformation were observed, as the $Mg^{2+}$ ion was coordinated by different specific oxygen atoms of the triphosphate. Additional series of short simulations were performed to evaluate the effect of monovalent cations on the lifetime of the βγ conformation (simulations 5–8 in Table 2.1). These simulations did not show significant differences in the stability of the βγ conformation of the Mg-ATP complex depending on the type of monovalent cation present in the simulation (Supplementary Table S1).

The distance $P^A$-$P^G$ and angle $P^B$-$O^{3B}$-$P^G$ were used to evaluate the conformation of the ATP phosphate chain and its dependence on the type of $M^+$ ions present. These measurements were taken from each conformation sampled from the MD simulation (Figure 3.2.7, Table 3.4). All $M^+$ ions seemed to induce shorter conformations of the phosphate chain, indicated by shorter $P^A$-$P^G$ distances, as compared to the Mg-ATP complex in the absence of $M^+$. However, the $P^A$-$P^G$ distances were longer in the simulations with $K^+$ ions, than in the simulations with $NH_4^+$ or $Na^+$ ions. Additionally, in the presence of $Na^+$ and $NH_4^+$, the phosphate chain was attaining an even shorter, curled conformation. These curled conformations were not present at any point in the simulations with $K^+$ ions or in simulations without additional $M^+$ ions (Figure 3.2.7, Table 3.4).

Conformations sampled from the MD simulations were plotted on heatmaps using the $P^B$-$O^{3B}$-$P^G$ angle and $P^A$-$P^G$ distance as coordinates (Figure 3.2.8). The coloring of each cell represents the probability (estimated as normalized frequency) of the Mg-ATP complex structure that fits the corresponding measurements. Since the βγ coordination of $Mg^{2+}$ corresponds to the NTP conformations seen in the structures of P-loop proteins, these conformations were also sampled from additional simulations and plotted separately. Among those, the longest $P^A$-$P^G$ distances, up to 5.5 Å, were seen in the absence of monovalent cations (Figure 3.2.8). In the presence of $M^+$ ions, the longest $P^A$-$P^G$ distances were observed in the simulations with $K^+$ ions. The addition of $M^+$ ions to the system prompted a notable decrease in the $P^A$-$P^G$ distances (Figure 3.2.8, Table 3.4). However, there were no notable differences in the $P^B$-$O^{3B}$-$P^G$ angles between the simulations with different $M^+$ ions or without such ions (Figure 3.2.8, Table 3.4).

The conformations of triphosphate chains in Mg-GTP complexes observed in the simulations with and without monovalent cations were similar to those observed for Mg-ATP complexes, see Supplementary Figure S2.

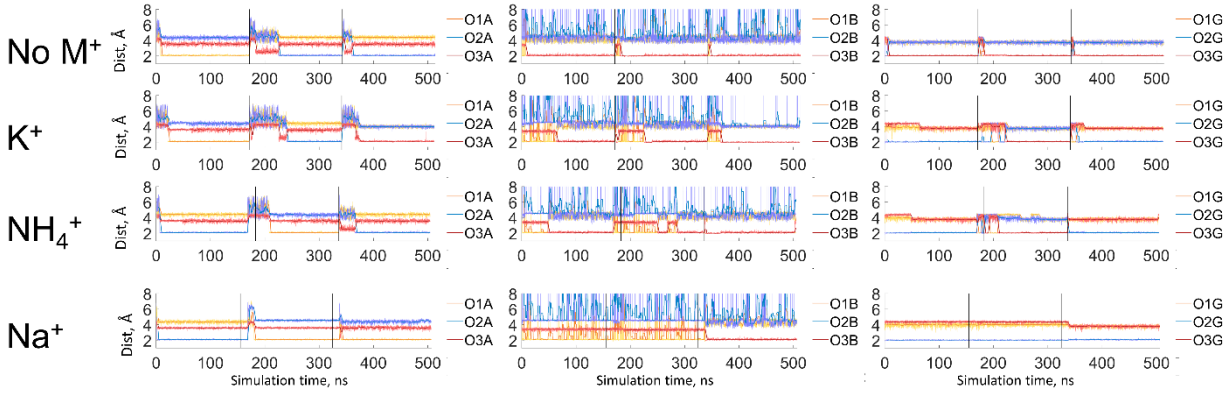


**Figure 3.2.6.** Coordination bonds between $Mg^{2+}$ ion and the phosphate chain oxygen atoms in MD simulations of Mg-ATP complex. Black vertical lines mark borders between individual simulation runs. The names of oxygen atoms are shown in Figure 3.2.2.

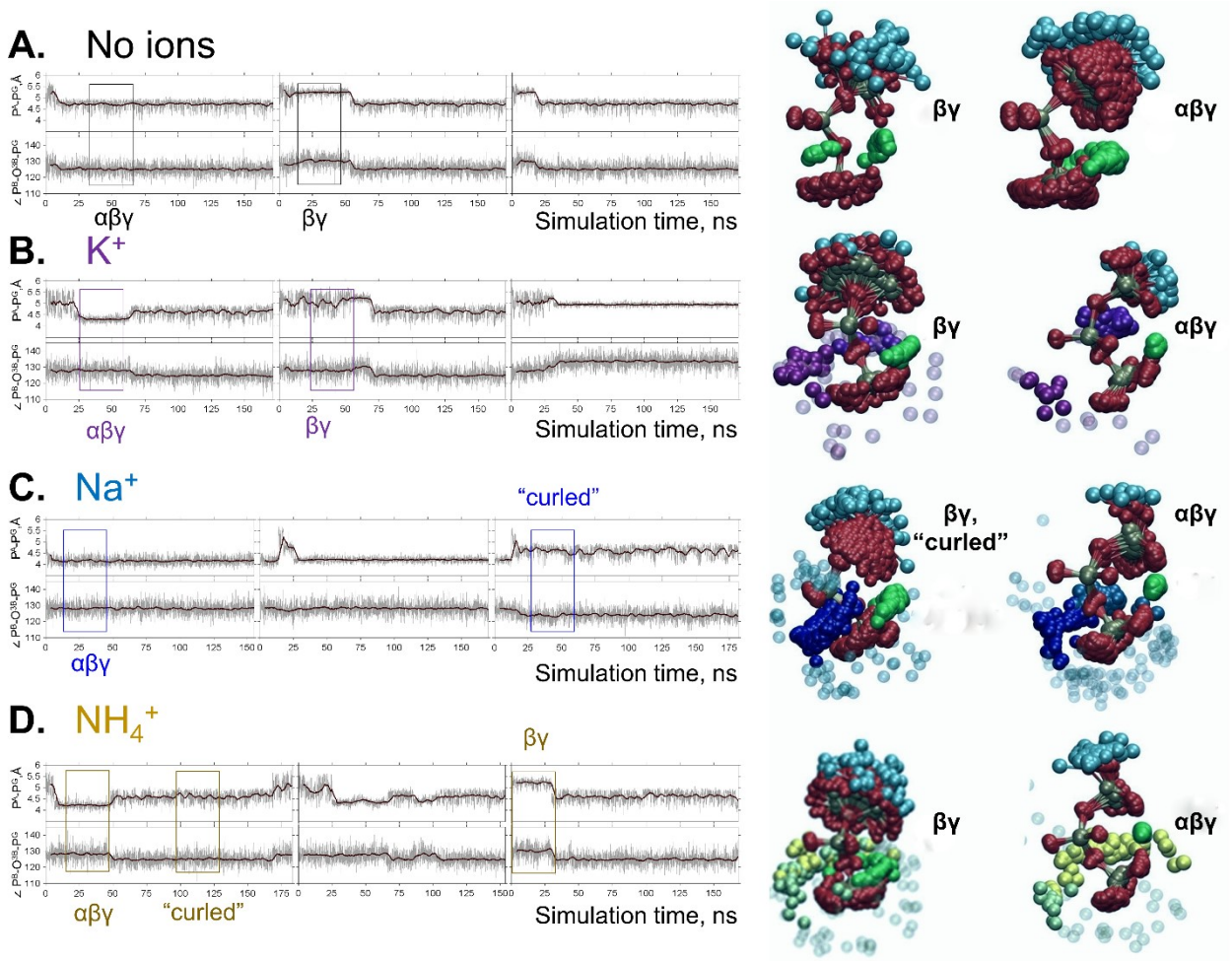


**Figure 3.2.7.** Phosphate chain geometry of Mg-ATP complex tracked throughout MD simulations. Left – tracks of $P^A$-$P^G$ distances (top) and the $P^B$-$O^{3B}$-$P^G$ angle (bottom). Measurements observed in MD simulations are shown in gray, moving average with a 2 ps window is shown in red. Boxes show fragments of the simulations, which were chosen for further analysis of particular conformations of the Mg-ATP complex; corresponding conformations of Mg-ATP and surrounding $M^+$ ions are shown on the right. The color scheme is the same as in Figure 3.2.3.

**Table 3.4.** Effects of $M^+$ ions on the triphosphate chain shape of Mg-ATP complex in water.

| Added cation | Phosphate chain conformation Mg-ATP[a] | | | | | |
|---|---|---|---|---|---|---|
| | βγ-coordination | | βγ-coordination, "curled" phosphate chain | | αβγ-coordination | |
| | $P^A$-$P^G$ distance, Å | $P^B$-$O^{3B}$-$P^G$ angle | $P^A$-$P^G$ distance, Å | $P^B$-$O^{3B}$-$P^G$ angle | $P^A$-$P^G$ distance, Å | $P^B$-$O^{3B}$-$P^G$ angle |
| None | 5.46 ± 0.34 | 122.3 ± 3.5 | N/A | | 4.76 ± 0.18 | 124.9 ± 3.3 |
| $K^+$ | 4.91 ± 0.24 | 122.0 ± 3.3 | N/A | | 4.32 ± 0.24 | 128.0 ± 3.5 |
| $Na^+$ | 4.69 ± 0.22 | 122.9 ± 3.2 | 4.60 ± 0.22 | 124.0 ± 3.3 | 4.26 ± 0.37 | 127.7 ± 3.6 |
| $NH_4^+$ | 4.85 ± 0.22 | 122.3 ± 3.3 | 4.56 ± 0.21 | 124.6 ± 3.3 | 4.22 ± 0.16 | 127.8 ± 3.9 |

[a] -The conformations of the $Mg^{2+}$-ATP complex (see Figure 3.2.7). The mean values and the standard deviations of distance $P^A$-$P^G$ (in Å) and angle $P^B$-$O^{3B}$-$P^G$ (in degrees) were calculated over the corresponding parts of the MD simulations. Time periods corresponding to βγ and αβγ coordination of $Mg^{2+}$ were determined by measuring distances between $Mg^{2+}$ and nearest non-bridging oxygens of the phosphate chain; time periods corresponding to the "curled" conformation were determined from $P^A$-$P^G$ distance tracks (Figure 3.2.7). Distances and angles for the αβγ coordination of the $Mg^{2+}$ and for the curled phosphate chain conformations were obtained from simulations 1-4 in Table 2.1; measurements of the βγ-coordination were obtained from simulations 5-8 in Table 2.1.

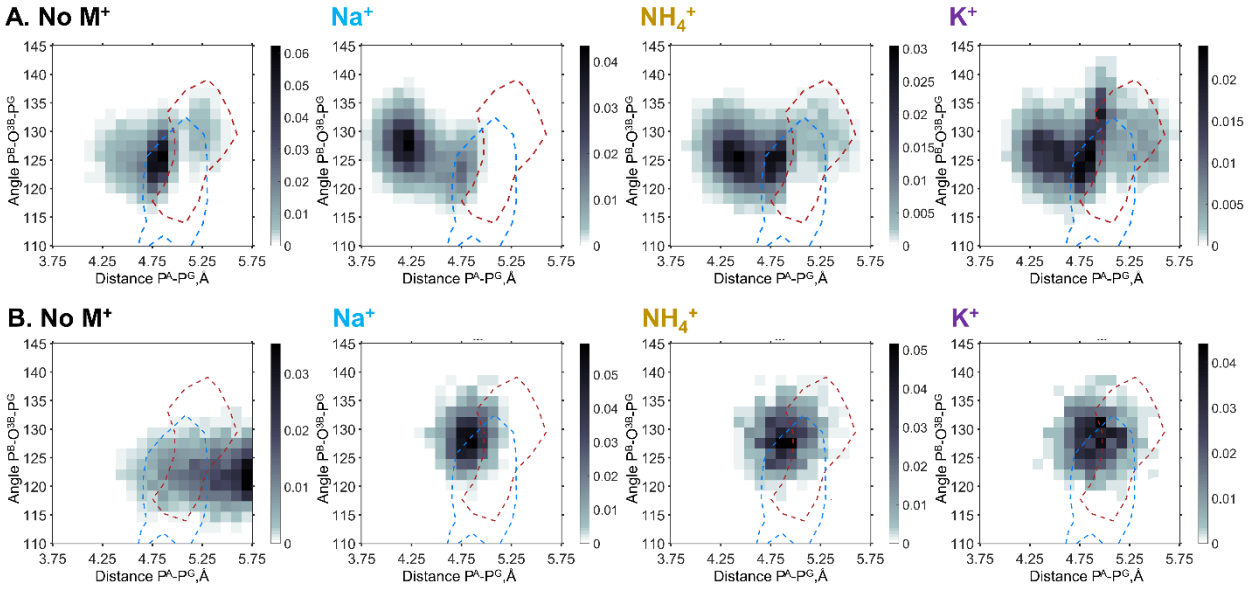


**Figure 3.2.8.** Heat maps of the phosphate chain conformation distributions of the Mg-ATP complex as observed in MD simulations. Conformations were sorted based on distances $P^A$-$P^G$ and angles $P^B$-$O^{3B}$-$P^G$ and plotted as heatmaps. Heatmaps for systems with $M^+$ ions show only those conformations of Mg-ATP complexes, that have at least one $M^+$ ion present within a 4 Å radius of the phosphate chain. The colors reflect the probability (estimated as normalized frequency) of the corresponding conformation. Dashed lines outline areas of the conformational space of NTP-mimicking molecules observed in X-ray structures of P-loop proteins: blue for the non-hydrolyzable analogs, red for the transition state analogs (see Figure 3.2.9). A. Data from simulations no. 1–4 in Table 2.1. B. Data from 4 × 20 ns simulations ATP exclusively with the βγ coordination of $Mg^{2+}$ ion (no. 5–8 in Table 2.1).

### 3.2.1.3 Phosphate chain shape in the P-loop NTPase structures

The binding of the Mg-NTP complex to a P-loop NTPase inflicts constraints on the phosphate chain so that only certain conformations are allowed. The conformations of the substrate, as seen bound in the catalytic site of an enzyme appear to be catalytically prone. Specifically, for the P-loop NTPases, binding of an NTP molecule to a P-loop domain alone, without an activating partner was shown to increase the hydrolysis rate by several orders of magnitude relative to the hydrolysis of NTP in water [216, 217].

Crystal structures of P-loop NTPases were analyzed to describe the conformations of triphosphate chains and the relative locations of positively charged groups in the vicinity. This data was compared with the conformations of Mg-ATP complexes observed in MD simulations. The InterPro [218] entry for 'P-loop containing nucleoside triphosphate hydrolase' (IPR027417) lists 2,899 X-ray and 55 NMR structures of proteins with P-loop domains. From that list only those X-ray structures were selected, that contain $Mg^{2+}$ ions, resulting in a list of 1,333 PDB IDs. Selected structures were analyzed with MATLAB scripts to identify and select structures that contain an NTP molecule, or a non-hydrolyzable analog, or an NDP complex mimicking the transition state. Additionally, NTP-like molecules were only considered if at least one Lys residue was present in their vicinity so that the distance from the NZ atom of Lys sidechain to the nearest of the phosphate chain phosphorus atoms (or the corresponding atoms in transition state-like complexes) was less than 4.5 Å. In total, 1357 complexes of NTP-like molecules with proteins have been identified and used in the phosphate chain shape measurements. Protein structures with proper ATP/GTP molecules, non-hydrolyzable NTP analogs, and analogs of the transition state were analyzed separately. Structures with the transition state analogs, such as ADP and GDP complexes with $AlF_3$/$BeF_3^-$/$MgF_3^-$ or $AlF_4^-$ moieties that mimic the γ-phosphate group [123, 131, 132] were taken as the closest approximations of the most catalytically prone phosphate chain shapes.

Structures with native ATP/GTP and different analogs have been analyzed separately. ATP and GTP molecules are usually present in the structures of inactive proteins, so the majority of such structures do not represent productive phosphate chain conformations. Non-hydrolyzable analogs cover exhibit smaller values of the $P^B$-$O^{3B}$-$P^G$ angle since, in such molecules, the ester oxygen between $P^B$ and $P^G$ is replaced with a different atom (N in ANP, GNP; C in ACP, GCP). In GSP and ASP instead, one of the oxygens of the γ-phosphate is replaced with S. Transition state analogs represent a variety of conformations from hydrolysis-prone conformations of the substrate, via conformations that correspond to different steps of hydrolysis, to completely separated

products. In the latter case, the distance between $P^A$ and $P^G$ (or corresponding atom in analogs) exceeds 5.5Å, indicating complete separation of the γ-phosphate-mimicking group from ADP/GDP; whereas the "$P^B$-$O^{3B}$-$P^G$" angle decreases due to the displacement of the former ester oxygen atom that becomes the free terminal oxygen of β-phosphate. For comparison with the MD data, only the major cluster of true transition state-like conformations was considered.

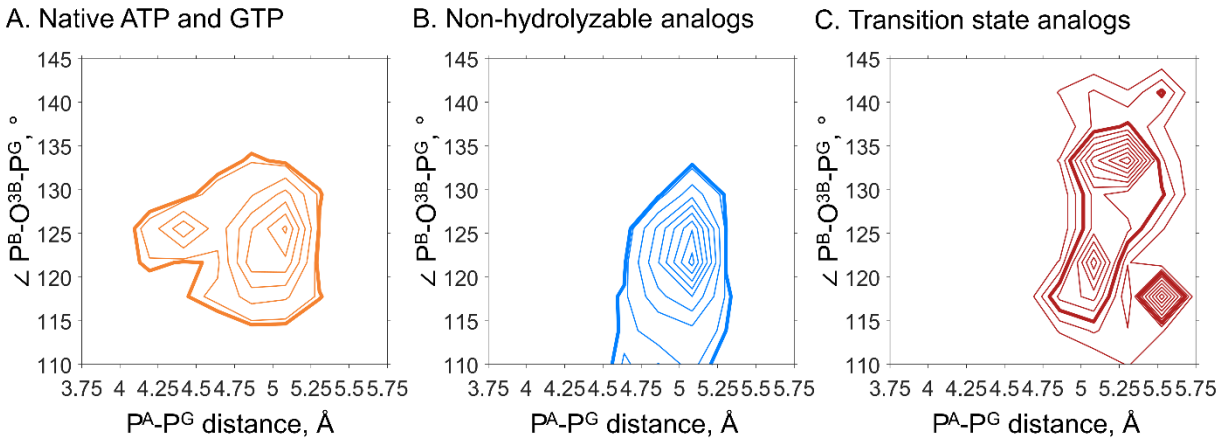


**Figure 3.2.9.** Phosphate chain conformation of NTP analogs in X-ray structures of P-loop NTPases. The most populated areas as shown by the heat map isotherms of the conformations distribution. Isotherms selected to represent crystallographic data for the comparison with the conformations sampled from MD simulations are shown as bold lines. A. Conformations of ATP and GTP molecules. B. Conformations of non-hydrolyzable analogs (PDB IDs: ANP, ACP, AGS, GNP, GCP, GSP). C. Conformations of complexes, mimicking the transition state (ADP or GDP in complex with $AlF_3$/$AlF_4^-$, $BeF_3$/$BeF_4^-$, or $VO_4^{3-}$). While ADP:$BeF_3$ and ADP:$BeF_4^-$ complexes can attain both the shape of the transition state and the ground state, depending on the geometry of the active site [219], they were grouped with the transition state analogs.

To investigate the phosphate chain shapes of NTP and NTP-like molecules bound to the P-loop proteins, the same parameters were used as for the analysis of MD simulation results, namely the distance $P^A$-$P^G$ and the angle $P^B$-$O^{3B}$-$P^G$ (or the analogous distance and angle in NTP-like molecules). These two parameters were used as coordinates, in which the geometries of NTP-like molecules from the X-ray structures were plotted (see Figure 3.2.9). The resulting plots were superimposed onto heatmaps of the Mg-ATP phosphate chain conformations, obtained from MD simulations (Figure 3.2.8). Heat maps of the top row of Figure 3.2.8, include all conformations of Mg-ATP, including those not observed in X-ray structures of P-loop NTPases, but only in MD simulations in water. Particularly, this includes all conformations with αβγ coordination of $Mg^{2+}$, see Figure 3.2.7. Thus, the conformational space of Mg-ATP complexes observed in MD simulations overlaps only partially with the conformations of NTP-like molecules bound to the P-loop NTPases (Figure 3.2.8A). The scope of the overlap is different depending on the nature of

the $M^+$ ion used in MD simulations: with $K^+$ it was the highest and with $Na^+$ - the lowest. Notably, conformations of Mg-ATP complex sampled from MD simulations correspond better to the shapes of non-hydrolyzable analogs, than to the shapes of the transition state analogs (see Figure 3.2.8A). However, phosphate chain geometries, obtained from simulations with $K^+$ ions overlap with both types of experimental structures (non-hydrolyzable analogs and transition state analogs) better, than the geometries observed in the simulations in the presence of sodium cations. Furthermore, the overlap between the conformational spaces of transition state analogs, sampled from X-ray structures, and Mg-ATP as observed in the MD simulations with $Na^+$ ions is completely absent.

For further analysis of the phosphate chain conformation with βγ-coordinated $Mg^{2+}$ ion more conformations of Mg-ATP and Mg-GTP complexes were sampled from short MD simulations (simulations 5-12 in Table 2.1). These conformations were compared with the data from X-ray structures, as described above. Conformation spaces of the phosphate chain in MD simulations plotted as heat maps show that without any $M^+$ ions the phosphate chain attains remarkably elongated geometry, indicated by large $P^A$-$P^G$ distances. Such shapes could not be observed in the simulations with monovalent cations and are not present in the X-ray structures. In the simulations with added $M^+$ ions, the $P^A$-$P^G$ distances were shorter. Particularly in the presence of sodium cations, the phosphate chain of Mg-ATP complex was notably shorter than in X-ray structures of P-loop proteins (Figure 3.2.8B, C). In contrast, MD simulations with $K^+$ or $NH_4^+$ ions produced the phosphate chain shapes that match almost exactly to the conformational space of the NTP analogs obtained from the crystal structures of P-loop proteins. Phosphate chain shapes sampled from these simulations overlap well with the conformations of non-hydrolyzable NTP-like molecules and even cover partially the space of the transition state analogs (Figure 3.2.8B, C). Still, some of the transition state analog conformations have severely widened $P^B$-$O^{3B}$-$P^G$ angle (>135°) and did not overlap with any of the structures sampled from MD simulations.

Altogether, Figure 3.2.8B/C and Figure S2 show that in molecular dynamics simulations of Mg-ATP and Mg-GTP the phosphate chain attains conformations, similar to those present in X-ray structures of P-loop proteins when with $K^+$ or $NH_4^+$ ions are present in the system, while in the presence of $Na^+$ ions such overlap between protein-bound experimental structures and simulated solution structures is noticeably smaller.

### *3.2.2. Cation Binding in P-loop NTPases*

For further specific analysis of $M^+$ ions role in P-loop NTPases, 10 X-ray structures of P-loop proteins from different families were selected as representatives. All selected structures contained bound NTP-like molecules (or complexes) so that positions of all three phosphates could be seen. Selected structures were superposed onto the structure of the $K^+$-dependent P-loop GTPase MnmE (PDB ID 2GJ8) [28] by aligning their P-loop regions (see Figure 3.2.10). Finally, positively charged residues (Lys and Arg), as well as cations (where present), around the phosphate chain were shown. Figure 3.2.10 reveals that the $M^+$ ions binding sites seen in the MD simulations (Figure 3.2.10A) are also taken by the positively charged residues or ions in X-ray structures of P-loop proteins (Figure 3.2.10B, C). Specifically, the BG site is always inhabited by the side chain amino group of the signature P-loop lysine. The AG site in different proteins is taken by either an $M^+$ ion (Figure 3.2.10B) or a positively charged side chain of the stimulating residue, e.g. an arginine finger (Figure 3.2.10C). In some structures of $K^+$ dependent proteins, this space is taken by a water molecule.

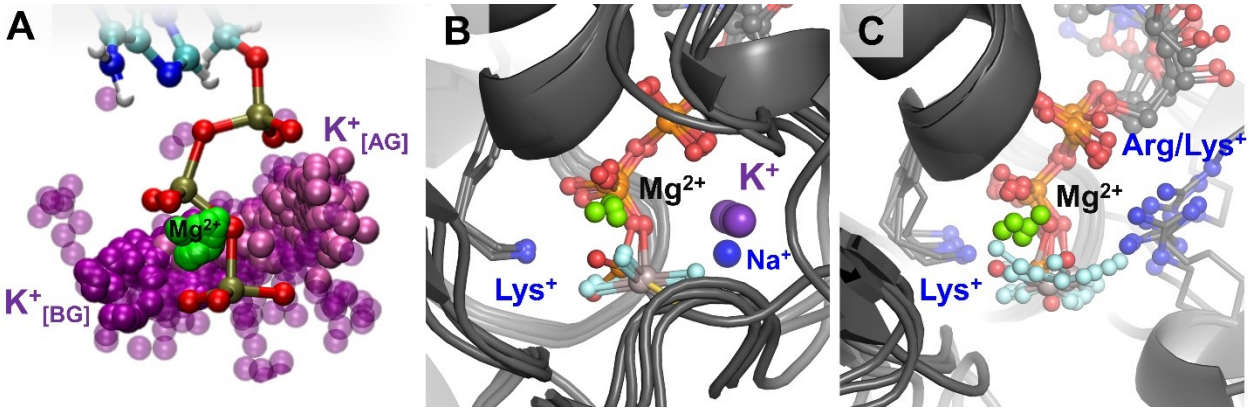


**Figure 3.2.10.** Distribution of positively charged moieties near the phosphate chain of Mg-GTP and Mg-ATP complexes. A. Cation positions sampled from MD simulations with $K^+$ ions. Only one conformation of the Mg-ATP complex is shown, all $K^+$ positions were sampled from complexes with βγ coordination of $Mg^{2+}$. B. Superposition of cation-dependent P-loop NTPases: dynamin-like protein (PDB ID 2X2E), translation factor eIF-B5 (PDB ID 4TMZ), Fe transporter FeoB (PDB ID 3SS8), and GTPase MnmE (PDB ID 2GJ8), see Table **3.6** for details. C. Superposition cation-independent P-loop NTPases: Rho/RhoGAP complex (PDB ID 5JCP), atlastin (PDB ID 4IDQ), $F_1$-ATPase (PDB ID 1H8E), GTPase of the nitrogenase complex (PDB ID 1N2C), ATPase subunit of human 26S proteasome (PDB ID 6MSB), DNA helicase Pif1 (PDB ID 5O6B).Atoms are colored the same as in Figure 3.2.3; proteins are shown as grey cartoons, Lys are Arg residues are shown as sticks, with sidechain nitrogen atoms shown as blue spheres.

**Table 3.5.** Activation mechanisms within TRAFAC and RecA/F1-like classes of P-loop NTPases

| Superfamily | Family | Stimulating charge | Activation mechanism |
|---|---|---|---|
| Kinase-GTPase division, TRAFAC class | | | |
| Classic translation factor GTPases | EF-G/EF-2 | $K^+$ | Functional interaction with ribosomal RNA/other protein(s)/other domain(s) of the same protein [18, 20, 21, 31-33, 35-38, 141, 142, 149, 155] |
| | EF-Tu/EF-1A | $K^+$ | |
| | EIF2G | $K^+$ | |
| | ERF3 | $K^+$ | |
| | IF-2 | $K^+$ | |
| | LepA | $K^+$ | |
| OBG-HflX-like GTPases | HflX | $K^+$ | |
| | OBG | $K^+$ | |
| | NOG | $K^+$ | |
| | YchF/OLA1 | $K^+$ | |
| YlqF/YawG GTPases | NOG2 | $K^+$ | |
| | RsgA | $K^+$ | |
| TrmE-Era-EngA-EngB-Septin-like GTPases | EngA (Der) | $K^+$ | |
| | EngB | $K^+$ | |
| | Era | $K^+$ | |
| | FeoB | $K^+$ | Dimerization (e.g. mRNA-associated in the case of MnmE) [30, 220, 221] |
| | MnmE | $K^+$ | |
| | Septin | Arg finger | |
| | Toc34-like | Arg finger | |
| Dynamin-like GTPases | hGBP | Arg finger | |
| | Dynamin | $K^+/Na^+$ | |
| Extended Ras | Ras family | Arg finger | Interaction with a specialized activating protein or domain [27, 222] |
| | Gα subunits | Arg finger | |
| Myosin/kinesin | Myosin | Arg finger | |
| | Kinesin | Arg finger | |
| ASCE division, RecA/F1-like class | | | |
| DNA-repair and recombination ATPases | RecA | Lys finger | DNA/RNA-dependent oligomerization [223] |
| | RadA | $K^+$ | |
| Rho helicases | Rho | Arg finger | Interaction an adjascent subunit in a hexamer [224-227] |
| T3SS ATPases | YscN | Arg finger | |
| | Flil | Arg finger | |
| F-/V-type ATPases | V-type A | Arg finger | |
| | F-type β | | |
| | V-type B | | |
| | F-type α | | |

In all selected representative P-loop NTPase structures, the phosphate chain attains the extended conformation, of a similar shape to the conformations present in MD simulations with larger $K^+$ and $NH_4^+$, but not smaller $Na^+$ ions (Figure 3.2.8). This elongated shape is enforced by numerous interactions between the phosphate groups and the P-loop motif residues, see [121].

Two classes of P-loop NTPases include both $M^+$-dependent and independent proteins. The activation mechanisms for these classes are summarized in Table 3.5. Generally, a wide variety of activation mechanisms has been described for different families of P-loop NTPases. These mechanisms involve interactions of the P-loop domain with other domains within one protein, or with other proteins (e.g. an adjacent subunit in an oligomer), or nucleic acids. These interactions commonly result in the introduction of a positively charged moiety - an $M^+$ ion or an Arg/Lys residue into the catalytic site [16, 27, 28, 113, 146, 150, 201, 227, 228]. NTP hydrolysis stimulation by the insertion of Arg/Lys residues or other groups into the AG sites in different proteins of the P-loop NTPases (super)family is discussed further in Chapter 3.2.6. Here, mainly the structures of P-loop ATPases and GTPases that are dependent on $M^+$ ions are discussed.

Crystal structures are available for only a few P-loop NTPases with experimentally shown $K^+$-dependent activity. Such structures were investigated to describe $M^+$ ions binding to the NTP phosphate chain and compare with the $Mg^{2+}$-ATP-$2K^+$ and $Mg^{2+}$-ATP-$2Na^+$ complexes observed in MD simulations. Totally, 17 structures of $M^+$-dependent P-loop proteins bound to NTP analogs and monovalent cations were analyzed (Table 3.6). For each structure, the same measurements used earlier to characterize the conformation of the phosphate chain were taken. Additionally, the coordination spheres of the $M^+$ ions in the AG site were examined. In all analyzed structures, the distances $P^A$-$P^G$ (or corresponding distances between analogous atoms in NTP analogs) were between 4.9 and 5.3 Å for the non-hydrolyzable analogs, and between 5.3 and 5.6 Å for transition state analogs (Table 3.6). These $P^A$-$P^G$ distances are similar to the values obtained from MD simulations with $K^+$ ions (see Figure 3.2.8 and Table 3.4).

The majority of $K^+$-dependent NTPases, and the special case of the $Na^+$-adapted family of dynamin-related GTPases, are members of the TRAFAC class [24]. Un these proteins $M^+$ ion binding involves the so-called K-loop [29]. This loop lies over the NTP-binding site and contributes two backbone carbonyl groups as $M^+$ ligands (see Figure 3.1.3). Unfortunately, only seldom structures of $M^+$-dependent proteins contain $K^+$ ions in the AG site (*cf.* Table 3.1 and Table 3.6). Moreover, in most cases, the K-loops are either distorted or unresolved (Figure 3.2.11).

**Table 3.6.** Binding of $M^+$ ions in X-ray structures of P-loop ATPases and GTPases.

| Protein | PDB ID entry | Bound NTP analog | Occupation of the AG site | | | Phosphate chain shape | |
|---|---|---|---|---|---|---|---|
| | | | Cation | Distance to the closest O atom of $P^A$, Å[a] | Distance to the closest O atom of $P^G$, Å [a,b] | $P^A$-$P^G$ distance, Å [a] | $P^B$-$O^{3B}$-$P^G$ angle, degrees[a] |
| Kinase-GTPase division, TRAFAC class | | | | | | | |
| GTPase MnmE (TrmE) | 2gj8 | GDP $AlF_4^-$ | $K^+$ | 2.8 | 2.6 | 5.4 | 136.3 |
| | 2gja | GDP $AlF_4^-$ | $NH_4^+$ | 2.9 | 2.5 | 5.4 | 136.9 |
| | 2gj9 | GDP $AlF_4^-$ | $Rb^+$ | 2.9 | 2.8 | 5.5 | 131.6 |
| GTPase FeoB | 3ss8 | GDP $AlF_4^-$ | $K^+$ | 2.8 | 2.6 | 5.4 | 144.9 |
| Dynamin - like proteins | 2x2e | GDP $AlF_4^-$ | $Na^+$ | 4.0 | 2.5 | 5.3 | 131.2 |
| | 2x2f | GDP $AlF_4^-$ | $Na^+$ | 4.1 | 2.6 | 5.3 | 133.6 |
| | 3w6p | GDP $AlF_4^-$ | $Na^+$ | 4 | 2.4 | 5.5 | 135.3 |
| | 3t34 | GDP $AlF_4^-$ | $Na^+$ | 3.8 | 2.4 | 5.6 | 149.3 |
| GTPase Era | 3r9w | GNP | $H_2O$[c] | 3 | 3.4 | 5.1 | 129.2 |
| Eukaryotic translation initiation factor eIF5B | 4ncn | GTP | $Na^+$ | 2.4 | 2.4 | 5.0 | 126.6 |
| | 4tmv | GSP | $Na^+$ | 2.4 | 2.8 (S)[d] | 4.9 | 126.3 |
| | 4tmw | GTP | $Na^+$ | 2.4 | 2.4 | 4.9 | 125.9 |
| | 4tmz | GSP | $K^+$ | 2.7 | 3.3 (S)[d] | 4.9 | 122.1 |
| ASCE division, RecA/F1-like class | | | | | | | |
| DNA recombinase RadA | 3ew9 | ANP | $K^+$ | 6.2 | 3.3 | 5.1 | 124.5 |
| | 2f1h | ANP | $K^+$ | 6.6 | 3.5 | 5.3 | 125.3 |
| | 2fpm | ANP | $K^+$ | 5.9 | 2.6 | 5.1 | 124.2 |
| | 1xu4 | ANP | $K^+$ | 6.1 | 2.7 | 5.2 | 125.0 |

[a] – distances and angles measured in PyMOL;
[b] – distance to the closest F atom was taken for the structures with $AlF_4^-$ complex;
[c] – While GTPase Era is $K^+$-dependent [36, 229], it was crystallized in a medium without $K^+$ ions, but in the presence of $Na^+$, thus a water molecule occupies the cation-binding site and forms H-bonds with $K^+$-binding residues; still, as the K-loop is fully resolved, the structure was included in this list;
[d] –GDP-monothiophosphate (GSP), non-hydrolyzable GTP analog, has $O^{1G}$ atom of γ-phosphate replaced with a sulfur atom

Crystal structures both with and without stimulating $K^+$ ion resolved were available only for MnmE, the tRNA modification GTPase, see Table 3.6 and Figure 3.2.12. Active and inactive structures of MnmE were compared to further the determinants of $K^+$-binding. MnmE functions as a dimer, where monomers are connected by their N-terminal domains, and the P-loop fold GTPase domains (G-domains) only come in contact during the catalytic turnover. Accordingly, the crystal structure of inactive MnmE shows interactions between the N-terminal domains, which

form a central hinge of the dimer, while the helical domains and P-loop domains are located on the opposite sides of the hinge (PDB ID 3GEI, Figure 3.2.12). In this structure, the distance between the P-loop domains (with bound non-hydrolyzable GTP analogs) is about 20 Å [28, 156]. The K-loops are not resolved in this structure, and $K^+$ binding is absent. The activation of MnmE involves conformational changes, which allow dimerization of their P-loop domains [16, 150]. Crystal structure of isolated P-loop domains of MnmE can be used as a representation of the active state of this protein (PDB ID 2GJ8, Figure 3.2.12). In this structure P-loop domains are bound to GDP:$AlF_4^-$ (transition state analog), and dimerized via the K-loop regions (see Figure 3.1.2 for K-loop definition), both K-loops and $K^+$ ions in both sites of the dimer are resolved (PDB ID 2GJ8, Figure 3.2.12). Unresolved disordered K-loop in the structure of MnmE in the inactive state and resolved, stabilized K-loop in the P-loop domain dimer suggest that the GTP hydrolysis could be stimulated by dimerization via stabilization of the K-loops, which should lead to the binding of $K^+$ ion in the active site.

One of the structures was included in the survey of $K^+$-bound NTPase structures despite not having a resolved cation in the AG site (see Table 3.6). It is the structure of the GTPase Era, which is $K^+$-dependent, but was crystallized in the presence of $Na^+$ ions, and not $K^+$ ions (PDB ID 3R9W [230]). Still, the structure depicts a fully resolved K-loop, while the AG site is taken by a water molecule (id 624) which is located within 2.9–3.4 Å of six potential $K^+$ ion ligands (see also Figure 3.2.13).

Apart from the TRAFAC class proteins, very few cases of $M^+$-dependent P-loop NTPases are known, all of them are RecA-like recombinases (Table 3.1). Together with the rotary ATPases, these proteins belong to the RecA/F1-like class of ASCE division (Additional Strand, Catalytic E), owing to the presence of an additional strand in the core β-sheet and a conserved Glu residue [24]. RecA-like recombinases are drastically different from the proteins of the TRAFAC class and do not possess loops corresponding to Switch I/K-loop or Switch II. The X-ray structure of the $K^+$-activated recombinase RadA (PDB ID 3EW9 [160]) shows two $K^+$ ions bound (Figure 3.2.14). One of these ions binds roughly in the AG site, although it is shifted closer to the γ-phosphate and further away from the α-phosphate. The other cation is situated between the catalytic Glu residue and the γ-phosphate, in the location that corresponds to the G-site, the relatively low-occupancy site seen in MD simulations of ATP and GTP complexes in water (Figure 3.2.3).

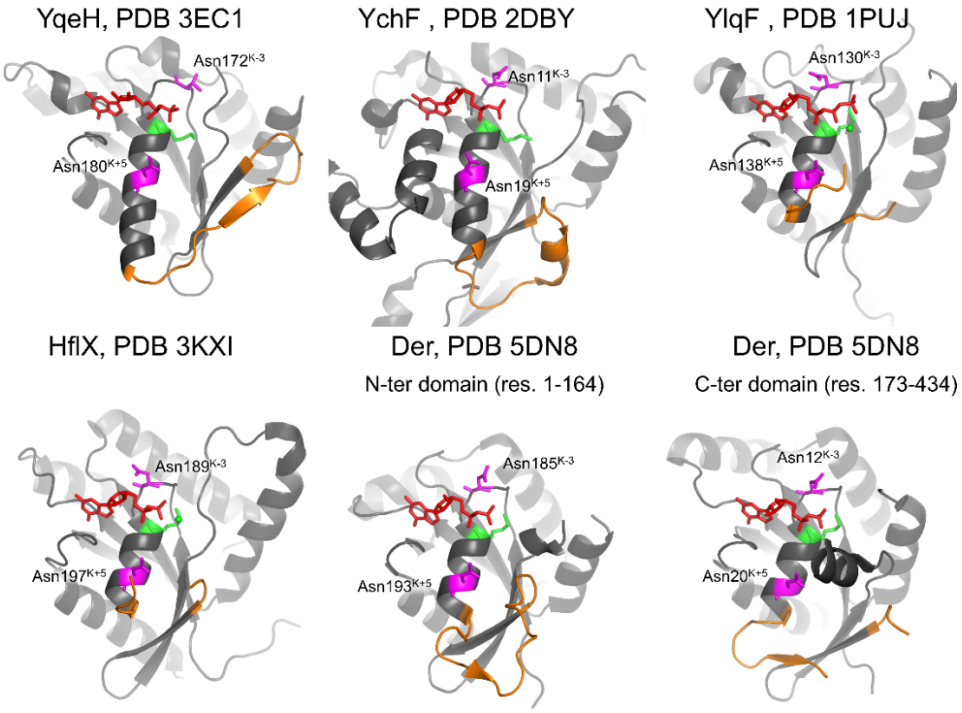


**Figure 3.2.11.** Crystal structures of $K^+$-dependent P-loop GTPases. In each protein, both Asn residues associated with the binding of monovalent cations are present [29]. Residues, associated with $K^+$-dependent activity are shown in magenta, P-loop Lys, and $Mg^{2+}$ ions (where present) - in green, K-loop region – in orange, GDP and GNP molecules – in red.

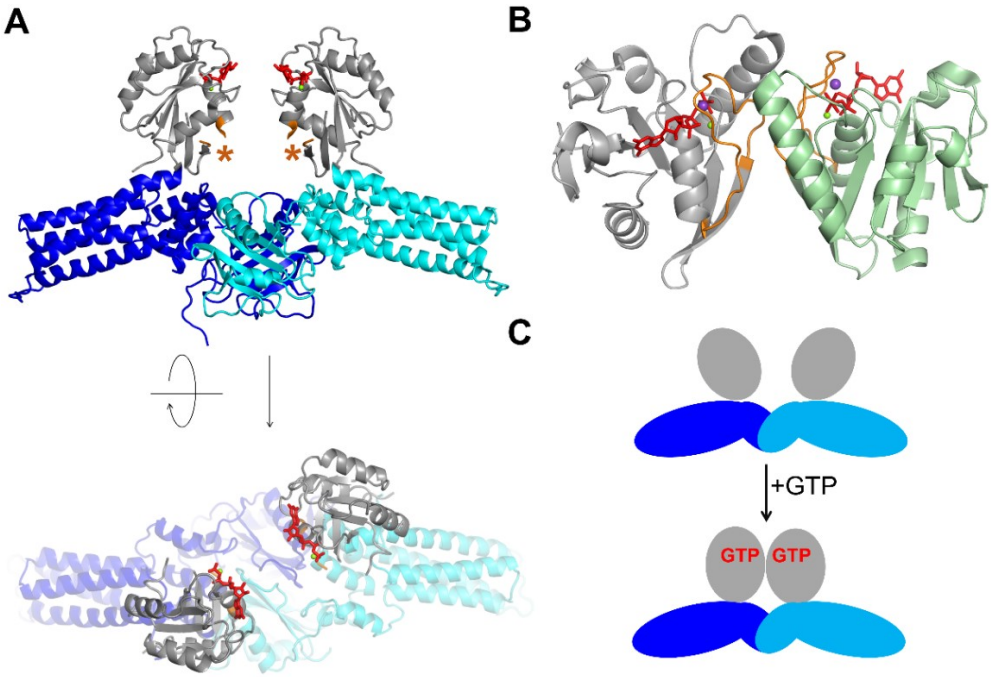


**Figure 3.2.12.** Activation of the GTPase MnmE. A. Inactive MnmE dimer bound to non-hydrolyzable GTP analog (PDB ID 3GEI) The P-loop domains are shown in grey, the absent K-loop is indicated by orange asterisks, the N-terminal and helical domains of the two monomers are colored in different shades of blue. B. Active dimer of isolated P-loop domains of MnmE, in complex with a transition state analog GDP:$AlF_4^-$ and $K^+$ ion (PDB ID 2GJ8). The K-loops are shown in orange, $K^+$ ions are shown in purple. C. Schematic representation of the conformational changes upon activation in MnmE dimers, domains are colored as on panel A (reproduced after [231]).

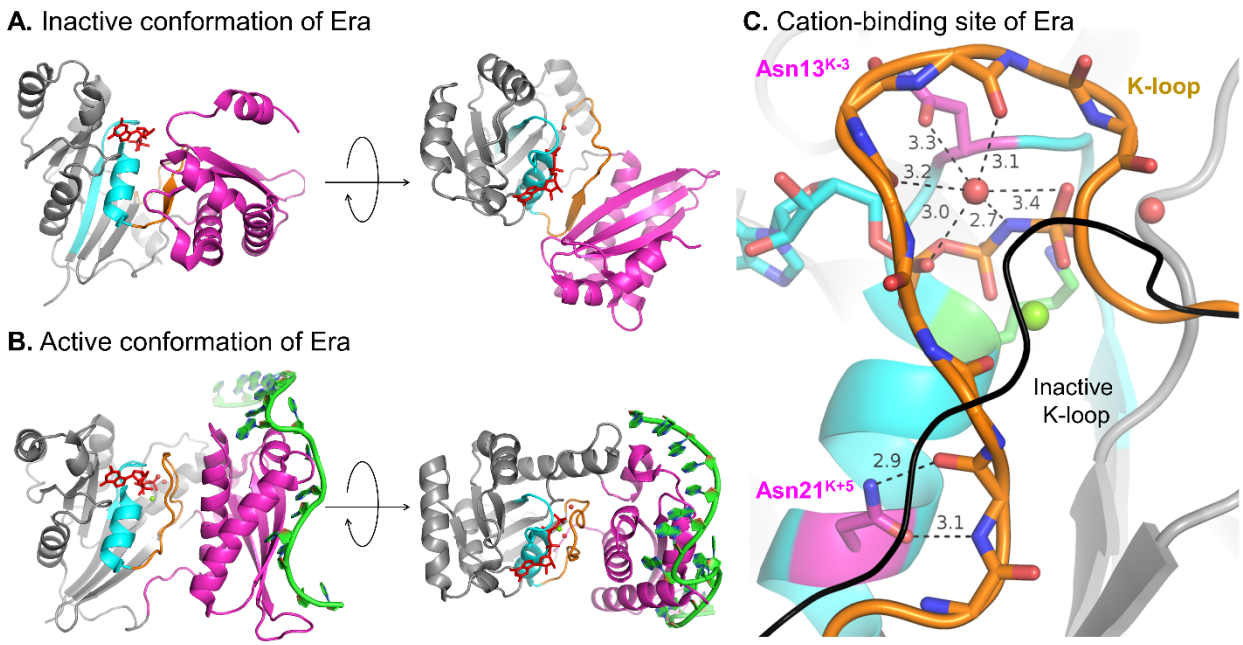


**Figure 3.2.13.** Conformational changes upon RNA binding and activation of the GTPase Era. A. RNA-free Era bound to GDP (PDB ID 3IEU [232]), protein is shown in two projections. B. RNA-bound Era bound to GNP, a non-hydrolyzable analog of GTP (PDB ID 3R9W [230]), the complex is shown in two projections. C. Cation-binding site of RNA-bound Era, catalytic water molecule, and water molecule in the AG site are shown as red spheres (PDB ID 3R9W [230]). The black line indicates the relative position and conformation of the K-loop in the inactive, RNA-free GDP-bound structure (PDB ID 3IEU [232]). The P-loop region is in light blue, the K-loop region is shown in orange, the RNA-binding domain is shown in magenta, GDP and GNP molecules are shown in red.

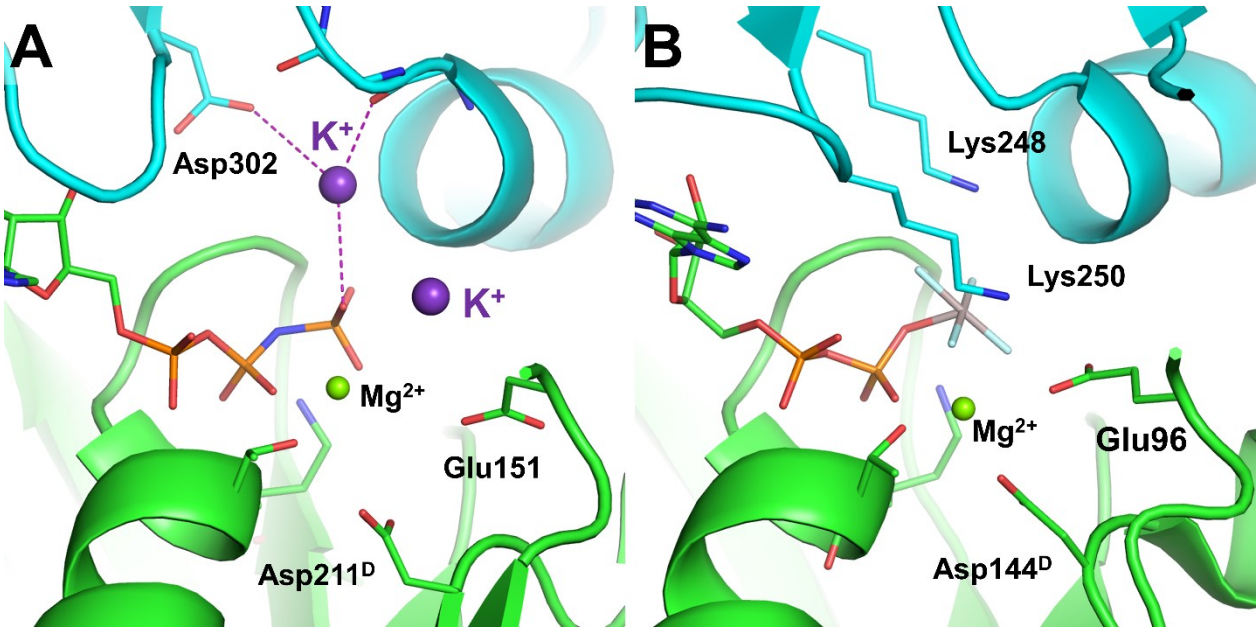


**Figure 3.2.14.** Potassium ions and Lys fingers in RecA-like recombinases.
A. $K^+$-dependent recombinase RadA from *Methanococcus voltae* (PDB ID 2F1H [233]). B. Cation-independent recombinase RecA from *E. coli* (PDB ID 3CMX [223]). NTP binding site is shown in green, adjacent monomer is shown in light blue, adjacent domain in the RadA recombinase was reconstructed from superposition with the RecA complex, $K^+$ ions are shown as purple spheres $Mg^{2+}$ ions - as green spheres.

### *3.2.3. Molecular Dynamics Simulations of GTPase MnmE*

The only $K^+$-dependent-NTPase with both $K^+$- bound and $K^+$-free experimental structures available is MnmE. Structure PDB ID 2GJ8 contains dimers of $K^+$-bound P-loop domains, and PDB ID 3GEI contains inactive, $K^+$-free dimer of the full-length protein, albeit with K-loops unresolved (Figure 3.2.12). To investigate the effect of $K^+$ ion binding on the phosphate chain of the Mg-GTP bound to MnmE, MD simulations of the protein in both its active and inactive states were executed. The active state was modeled after the X-ray structure PDB ID 2GJ8. Cation positions ($Mg^{2+}$ and $K^+$) were taken directly from the crystal structure, transition state analogs (complexes GDP:$AlF_4^-$) in both monomers were replaced by GTP molecules. Hereafter this system is referred to as $2GJ8_K$. To model $K^+$-free, inactive state two systems were created. The first was modelled as a monomer from the 2GJ8 X-ray structure, with the transition state analog replaced by GTP and the $K^+$ ion replaced with a water molecule. Hereafter this system is referred to as $2GJ8_W$. This system was modeled as a monomer since in the absence of $K^+$ ions dimerization of MnmE P-loop domains does not occur [16]. The second inactive system was modeled as a P-loop domain monomer as well, but PDB ID 3GEI (inactive full-length protein) was used as a starting structure. The non-hydrolyzable GTP analog GNP (guanosine 5′-imidotriphosphate) was changed to a GTP molecule and the absent K-loop was reconstructed. The resulting three protein complexes were placed in the simulation boxes, filled with water and KCl. Simulations of all systems were 100 ns long (see Table 2.2).

The main features to characterize the active site of MnmE during the MD simulations are the location and shape of the K-loop, coordination of $Mg^{2+}$ by the GTP, and stability of the H-bonds between GTP and the P-loop residues. In the simulation of the 3GEI system, the movement of the K-loop broke the coordinating bond between the conserved Switch I.K-loop Thr and the $Mg^{2+}$ ion. This led to a distortion of the βγ coordination of $Mg^{2+}$ so that the phosphate chain switched between the βγ- and αβγ-coordination of $Mg^{2+}$ ion. Despite these fluctuations, the H-bond between the backbone nitrogen atom of $Thr248^{K+2}$ ($Ser231^{K+2}$ in the 2GJ8 structure) and the $O^{1A}$ atom of the α-phosphate remained remarkably stable during the whole simulation (Figure 3.2.15). The corresponding H-bond in the $2GJ8_K$ and $2GJ8_W$ systems between the backbone nitrogen of $Ser231^{K+2}$ and the $O^{1A}$ atom remained stable as well (Figure 3.2.15B, C). In the $2GJ8_W$ system, the K-loop remained in its place throughout the MD simulation, maintaining the coordination of $Mg^{2+}$ and its βγ coordination by the phosphate chain, despite the absence of $K^+$ ion in the AG site.

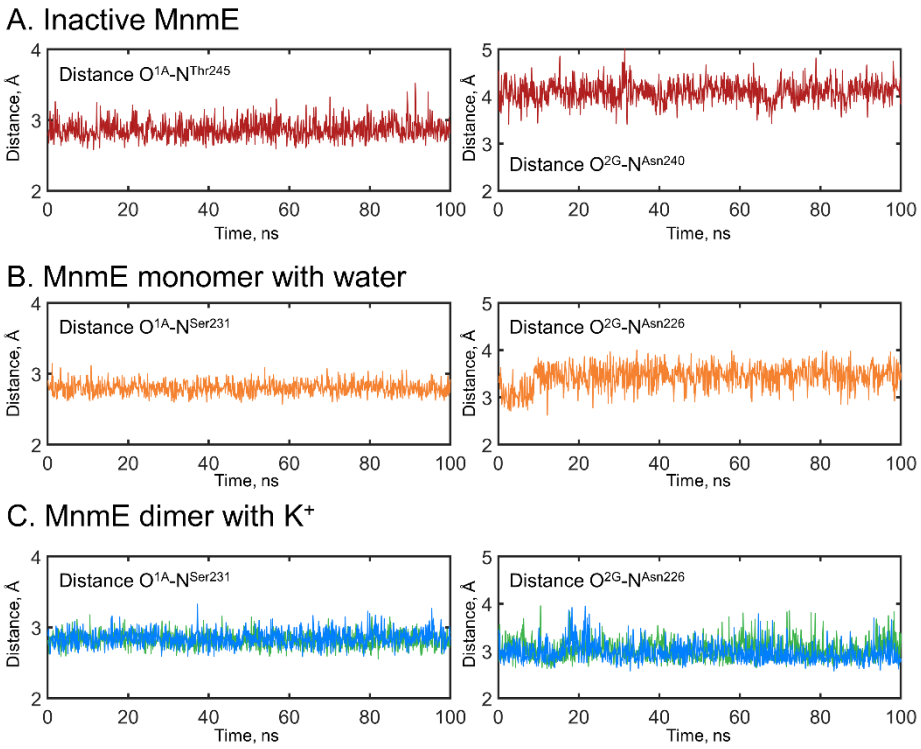


**Figure 3.2.15.** Length of H-bonds measured during MD simulations of the GTPase MnmE. A. Inactive monomer with disordered K-loop (system 3GEI). B. Inactive monomer MnmE with ordered K-loop, but in the absence of $K^+$ ion in the active site (system $2GJ8_W$). C. Dimer with ordered K-loops and bound $K^+$ ions (system $2GJ8_K$, distances in the two monomers are plotted separately in blue and green.

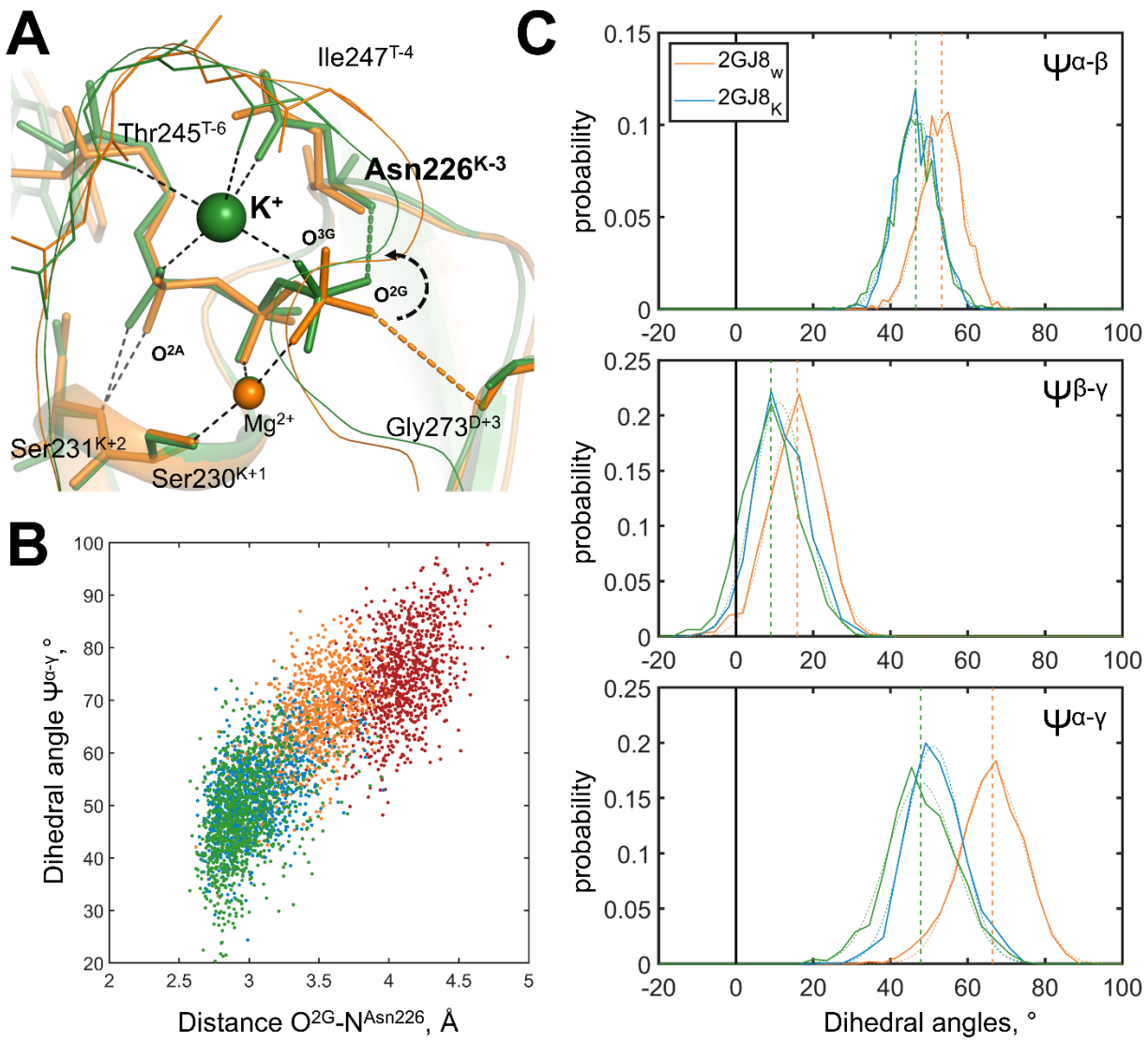


**Figure 3.2.16.** Conformational changes observed in MD simulations of MnmE. A. Superposition of the active sites of monomeric MnmE (system $2GJ8_W$, orange) and the active dimer of MnmE with bound $K^+$ ions (system $2GJ8_K$, green); representative structures sampled from MD simulations are shown. Proteins are shown as cartoons; ions are shown as spheres; GTP and relevant amino acid residues are shown as sticks. Black dashed lines show H-bonds and cation coordination that are present in both systems. B. Scatter plot of the $\Psi^{\alpha-\gamma}$ dihedral angle (Y-axis) against the length of the H-bond between the backbone nitrogen of $Asn226^{K-3}$ and the $O^{2G}$ atom (X-axis) as obtained from the MD simulations: (1) red, inactive monomeric P-loop domain of MnmE with disordered K-loop (system 3GEI), (2) orange, monomeric P-loop domain of MnmE with a water molecule bound in the AG site (system $2GJ8_W$); green/blue, active dimer of P-loop domains with $K^+$ ions bound in the respective AG sites (system $2GJ8_K$, green and blue colors correspond to individual monomers). C. Dihedral angles between the GTP phosphate groups, as obtained from MD simulations of the MnmE dimer with $K^+$ ions in AG sites (system $2GJ8_K$ system, green and blue colors correspond to two individual monomers) and the monomeric MnmE with a water molecule replacing $K^+$ ion in the active site (system $2GJ8_W$, orange). Solid lines show dihedral angle distribution plotted as normalized histograms, dashed lines show fitted normal distribution functions. Vertical lines show the centroid values of the fits. Black vertical lines mark $\Psi = 0°$, which indicates a fully eclipsed conformation, while $\Psi = \pm 60°$ indicates a fully staggered conformation.

Due to unstable $Mg^{2+}$ coordination in the 3GEI system, the two systems with a stable K-loop ($2GJ8_W$ and $2GJ8_K$) were used to compare the shapes of the GTP phosphate chain with and without $K^+$ ion in the AG site. The distances between $O^{2A}$ and $O^{3G}$ atoms indicate, that in the presence of $K^+$ and its coordinational bonds these two atoms are pulled closer together (Figure 3.2.16, Figure 3.2.17). In the $2GJ8_W$ system, the average distance between $O^{2A}$ and $O^{3G}$ was 5.3 Å, while, in the $2GJ8_K$ system, the decreased to 4.7 Å (see Figure 3.2.17). By forming coordinational bonds with $O^{3G}$ and $O^{2A}$, the $K^+$ ion pulls the atoms closer together. While α-phosphate is secured by the bond with the K+2 residue ($Thr248^{K+2}$ in 3GEI, $Ser231^{K+2}$ in 2GJ8), the γ-phosphate seems to be less restricted in its movement. Thus, the $K^+$ ion pulling on the $O^{3G}$ atom forces the γ-phosphate to turn. As a result, the $O^{2G}$ atom of the γ-phosphate is pulled closer to the backbone nitrogen of $Asn226^{K-3}$ and forms a new H-bond with this atom (Figure 3.2.16). Notably, the $Asn226^{K-3}$ residue is directly involved in the coordination of the $K^+$ ion (Figure 3.2.16A). The new H-bond between $Asn226^{K-3}$ and $O^{2G}$ atom was not observed in the systems lacking the $K^+$ ion in the AG site ($2GJ8_W$ and 3GEI systems) (Figure 3.2.15A, B, Figure 3.2.16B).

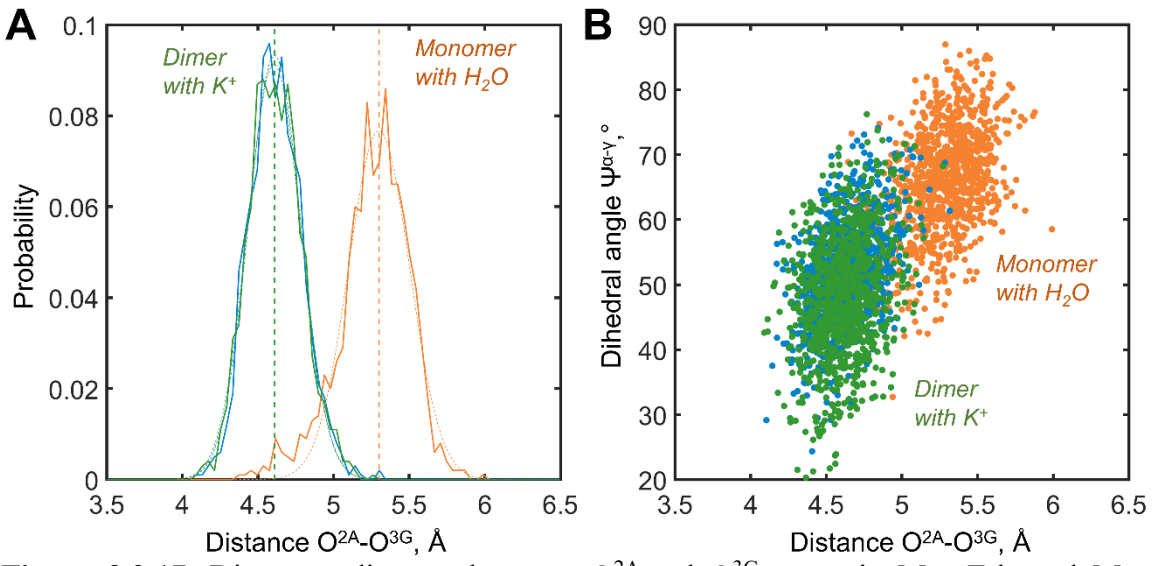


**Figure 3.2.17.** Distances distance between $O^{2A}$ and $O^{3G}$ atoms in MnmE-bound Mg-GTP as obtained from MD simulations. Data for the active dimer of P-loop domains with $K^+$ ions bound (system $2GJ8_K$, blue and green for individual monomers) and inactive, monomeric P-loop domain of MnmE with a water molecule in the AG site (system $2GJ8_W$ system, orange). A. Distributions of the $O^{2A}$-$O^{3G}$ distances during MD simulations. Distribution histograms are shown as solid lines, Gaussian fits are shown as dotted lines, vertical dashed lines indicate centroid values of the corresponding fits. B. Scatter plot of the dihedral angle $\Psi^{\alpha-\gamma}$ (Y-axis) plotted against the $O^{2A}$-$O^{3G}$ distance (X-axis).

To further investigate the changes in the phosphate chain conformation triggered by the presence of $K^+$ ion in the active site, the dihedral angles between the GTP phosphate groups were evaluated, namely $\Psi^{\alpha-\beta} = \angle O^{2A}\text{-}P^A\text{-}P^B\text{-}O^{2B}$, $\Psi^{\beta-\gamma} = \angle O^{1B}\text{-}P^B\text{-}P^G\text{-}O^{1G}$, and $\Psi^{\alpha-\gamma} = \angle O^{1A}\text{-}P^A\text{-}P^G\text{-}$

$O^{3G}$ (see Figure 3.2.16). As in the 3GEI system, $Mg^{2+}$ coordination was not stable, it was not taken into further comparison. Oxygen atoms of α- and β-phosphates remained in the positions close to the staggered conformation, which is more energetically favorable, in both systems. The angle $\Psi^{\alpha-\beta}$ angle only slightly decreased from 53° to 47° with the introduction of the $K^+$ ion (Figure 3.2.16C). Angle $\Psi^{\beta-\gamma}$ was close to the eclipsed conformation in both systems, supported by the βγ coordination of $Mg^{2+}$ and the H-bonds between β- and γ-phosphates and the P-loop Lys229. The average $\Psi^{\beta-\gamma}$ angle in the $2GJ8_K$ system was 6°, and slightly wider in the $2GJ8_W$ system - 16° (Figure 3.2.16A, C). Angle $\Psi^{\alpha-\gamma}$ changed most drastically between the two systems. In the $2GJ8_W$ system α- and γ-phosphates were close to the staggered conformation, with an average $\Psi^{\alpha-\gamma}$ of 66° (Figure 3.2.16A, C). In the $2GJ8_K$ system, with $K^+$ ions in the AG sites, the γ-phosphate was rotated by ~30°, bringing the $\Psi^{\alpha-\gamma}$ angle to an average of 38°, so that relative positions of oxygen atoms of α- and γ-phosphates moved closer towards the eclipsed conformation (Figure 3.2.16A, C). Overall, the presence of the $K^+$ ion bound by the $O^{2A}$ and $O^{3G}$ atoms leads to a more eclipsed conformation of the triphosphate chain. This near-eclipsed state was achieved mainly due to the turn of the γ-phosphate group, with relatively minor shifts of the α- and β-phosphates (Figure 3.2.16A, C).

### *3.2.4. Cation Binding in Translation Factors*

TRAFAC class of P-loop NTPases is unique in its diversity of activation mechanisms, employed by the members of this class. Previous chapters were focused on cation-dependent members of this protein group, with their unique features, responsible for $K^+/Na^+$ binding. Remarkably, while translation factors are a titular (super)family within the TRAFAC class and share a lot of features similar to those of cation-dependent GTPases (such as MnmE), involvement of $K^+$ ions in the stimulation of GTP hydrolysis by the translation factors on the ribosome remains debatable. This chapter describes available structural data on the translation factors and the existing evidence of cation binding. In the next chapter, this information is used to model translation factor EF-Tu in its active state, bound to Mg-GTP, tRNA, and ribosome, to further investigate cation-binding sites in this complex.

In order to gather evidence for cation binding next to catalytic sites of translational GTPases, all their structures available from the Protein Data Bank (PDB) were inspected. Crystallographic and cryo-EM structures of translational GTPases with the resolution of < 4Å were selected for the structure survey. Since the focus of this study was on the mechanism of GTP hydrolysis, the analysis was limited to those structures that contained a Mg-GTP complex or its

analog bound to the P-loop. In total, 84 structures were analyzed, which included 37 structures of translational GTPases bound either to an entire ribosome or a ribosomal subunit. All inspected structures with relevant data are listed in Table 3.7. In each structure, the GTP binding site was inspected for the presence of metal cations and/or water molecules. Water molecules were investigated as potential hubs of H-bonding networks.

Figure 3.2.18A and B show representative structures of translational GTPases that were determined in the absence of ribosomes or their parts (hereafter "solo structures", see also Table S2). $K^+$ and $Na^+$ ions are seen in the aforementioned structures of eIF5B, see Figure 3.2.18, Table 3.6 and [21, 188]. As already noted, and as seen in Figure 3.2.18, the monovalent cations are coordinated by $O^{2A}$, $O^{3B}$, and $O^{3G}$ atoms of the phosphate chain, the side chain of the $Asp533^{K-3}$ residue of the P-loop, and the carbonyl oxygen of $Gly555^{T-2}$ of the Switch I loop [21]. Both $K^+$ and $Na^+$ ions were bound by the same ligands; however, because coordination bonds for a $Na^+$ ion (2.0-2.5 Å) are shorter than for a $K^+$ ion (2.6-3.2 Å) [115, 203, 234, 235], the $Na^+$ ion dwells closer to its ligands than the $K^+$ ion [21].

Representative structures of translational GTPases in Figure 3.2.18B disclose water molecules that occupy the same positions in different structures, see Table 3.7. Specifically, the position between the γ-phosphate and $His^{D+4}$ residue of Switch II is often occupied by a water molecule. Hereafter a water molecule in this position will be denoted as $w^{HIS}$. One more water molecule is found between the side chain of $Asp^{K-3}$ and the $O^{3G}$ atom of γ-phosphate, the water molecule in this position is hereafter denoted as $w^{ASP}$. Several structures show a water molecule approximately in the position that is taken by a $K^+$ ion in $K^+$-dependent NTPases (cf. Figure 3.2.18A). For instance, the structures of EF-Tu with the highest resolution (PDB ID 2C78 with the resolution of 1.4 Å and PDB ID 2C77 with the resolution of 1.6 Å [236]) report a water molecule bound between $Asp21^{K-3}$, $Gly60^{T-2,}$ and $O^{3G}$ atom of the phosphate chain. Finally, two water molecules can be seen coordinating the $Mg^{2+}$ ion bound to the GTP analog (Figure 3.2.18B). All water molecules found in these positions are listed in Table 3.7.

It is noteworthy that the $His^{D+4}$ residue is turned away from the phosphate chain and does not interact directly with the $w^{HIS}$ molecule in all solo structures of translational GTPases (Figure 3.2.18A, B).

**Table 3.7.** Cation binding sites in structures of translation factors

| PDB ID | Uniprot name | NTP | Complex | BG site | SRL site | AG site | Water molecules near the phosphate chain | Other water molecules |
|---|---|---|---|---|---|---|---|---|
| 1EFT | EFTU_THEAQ | GNP | protein | Mg | - | w-482 | w-426 in position $w^{HIS}$ | w-490 and w-491 coordinate $Mg^{2+}_{GTP}$ |
| 1B23 | EFTU_THEAQ | GNP | protein-tRNA | Mg | - | - | w-412 in position $w^{HIS}$ | w-410 and w-419 coordinate $Mg^{2+}_{GTP}$ |
| 1TTT | EFTU_THEAQ | GNP | protein-tRNA | Mg | - | - | None | w-409 coordinates $Mg^{2+}_{GTP}$ |
| 1OB5 | EFTU_THEAQ | GNP | protein-tRNA | Mg | - | - | None | None |
| 4PC7 | EFTU1_ECOLI | GNP | protein | Mg | - | w-501 | None | None |
| 5UYM | EFTU1_ECOLI | GCP | ribosome | Mg | Mg | - | None | None |
| 5UYQ | EFTU1_ECOLI | GCP | ribosome | Mg | - | - | None | None |
| 5UYQ | EFTU1_ECOLI | GCP | ribosome | Mg | - | - | None | None |
| 2C78 | EFTU1_THET10 | GNP | protein | Mg | - | - | w-2046 in position $w^{HIS}$; w-2320 in position $w^{ASP}$ | w-2010 coordinates $Mg^{2+}_{GTP}$ |
| 1EXM | EFTU1_THETH | GNP | protein | Mg | - | w-473 | w-411 in position $w^{HIS}$ | w-414 and w-417 coordinate $Mg^{2+}_{GTP}$ |
| 4H9G | EFTU1_THETH | GNP | protein | Mg | - | - | w-617 in position $w^{HIS}$; w-664 in position $w^{ASP}$ | w-612 and w-628 coordinate $Mg^{2+}_{GTP}$; w-698 near $O^{1A}$ atom |
| 4LBV | EFTU1_THETH | GNP | protein | Mg | - | - | w-607 in position $w^{HIS}$; w-628 in position $w^{ASP}$ | w-604 and w-608 coordinate $Mg^{2+}_{GTP}$; w-783 near $O^{1A}$ atom |
| 4LBW | EFTU1_THETH | GNP | protein | Mg | - | - | w-894 in position $w^{HIS}$; w-692 in position $w^{ASP}$ | w-602 and w-906 coordinate Mg2+(GTP); w-769 near $O^{1A}$ atom |
| 4LBY | EFTU1_THETH | GNP | protein | Mg | - | w-628 | w-601 in position $w^{HIS}$ | w-657 and w-662 coordinate $Mg^{2+}_{GTP}$ |
| 4LBZ | EFTU1_THETH | GNP | protein | Mg | - | w-795 | w-607 in position $w^{HIS}$ | w-747 and w-796 coordinate $Mg^{2+}_{GTP}$; w-722 near $O^{1A}$ atom |
| 4LC0 | EFTU1_THETH | GNP | protein | Mg | - | w-635 | w-606 in position $w^{HIS}$ | w-720 and w-721 coordinate $Mg^{2+}_{GTP}$ |
| 2BVN | EFTU2_ECOLI | GNP | protein | Mg | - | - | w-2082 in position $w^{ASP}$ | w-2003 coordinates $Mg^{2+}_{GTP}$ |
| 6EZE | EFTU2_ECOLI | GNP | protein | Mg | - | - | w-603 in position $w^{HIS}$ | w-511, w-515 and w-517 coordinate $Mg^{2+}_{GTP}$ |
| 1OB2 | EFTU2_ECOLI | GNP | protein/-tRNA | Mg | - | - | None | None |
| 5WDT | EFTU2_ECOLI | GNP | ribosome | Mg | Mg | - | None | w-501 near ribose atom O2' |
| 5WE4 | EFTU2_ECOLI | GNP | ribosome | Mg | Mg | - | None | None |
| 5WE6 | EFTU2_ECOLI | GTP | ribosome | Mg | Mg | - | None | None |
| 5WFS | EFTU2_ECOLI | GTP | ribosome | Mg | Mg | - | None | None |
| 2C77 | EFTU2_THET8 | GNP | protein | Mg | - | - | w-2057 in position $w^{HIS}$; w-2015 in position $w^{ASP}$ | w-2042 and w-2082 coordinate $Mg^{2+}_{GTP}$ |
| 4V5L | EFTU2_THET8 | GCP | ribosome | Mg | - | - | w-2001 in position $w^{HIS}$ | None |
| 3VMF | EF1A_AERPE | GTP | protein | Mg | - | w-619 | w-602 in position $w^{HIS}$ | w-656 and w-657 coordinate $Mg^{2+}_{GTP}$ |
| 3WXM | EF1A_AERPE | GTP | protein | Mg | - | w-605 | w-616 in position $w^{HIS}$ | w-712 and w-792 coordinate $Mg^{2+}_{GTP}$ |
| 6JI2 | EF1A_AERPE | GTP | protein | Mg | - | Na | None | w-602 coordinates $Mg^{2+}_{GTP}$ |
| 4V9O | EFG_ECOLI | GCP | ribosome | Mg | Mg | - | w-901 in position $w^{HIS}$ | None |
| 4V9P | EFG_ECOLI | GCP | ribosome | Mg | - | - | w-901 in position $w^{HIS}$ | None |
| 4V9J | EFG_THET2 | GNP | ribosome | Mg | - | - | None | None |
| 4V9K | EFG_THET2 | GNP | ribosome | Mg | - | - | None | None |
| 4V90 | EFG_THET8 | GCP | ribosome | Mg | Mg | - | w-2002 in position $w^{HIS}$; w-2001 in position $w^{ASP}$ | None |
| 4V9H | EFG_THET8 | GCP | ribosome | Mg | Mg | - | w-903 in position $w^{HIS}$; w-804 in position $w^{ASP}$ | w-801 and w-802 coordinate $Mg^{2+}_{GTP}$ |
| 2BV3 | EFG_THETH | GNP | protein | Mg | - | - | None | w-2157 near $O^{1A}$ atom |
| 2J7K | EFG_THETH | GCP | protein | Mg | - | - | None | None |
| 3JCJ | IF2_ECOLI | GNP | ribosome | - | Mg | - | None | None |
| 4B48 | IF2_THET8 | GTP | protein | Mg | - | - | w-2004 in position $w^{HIS}$ | None |
| 1KK1 | IF2G_PYRAB | GNP | protein | Mg | - | - | w-1020 in position $w^{HIS}$ | w-1024, w-1038, w-1032 coordinate $Mg^{2+}_{GTP}$ |
| 4RD4 | IF2G_SACS2 | GNP | protein | Mg | - | - | w-605 in position $w^{HIS}$; w-656 in position $w^{ASP}$ | w-608 and w-985 coordinate $Mg^{2+}_{GTP}$ |
| 4RD2 | IF2G_SACS2 | GNP | protein | Mg | - | - | w-606 in position $w^{HIS}$; w-886 in position $w^{ASP}$ | w-601 and w-612 coordinate $Mg^{2+}_{GTP}$ |
| 4RCZ | IF2G_SACS2 | GNP | protein | Mg | - | - | w-609 in position $w^{HIS}$; w-650 in position $w^{ASP}$ | w-601 and w-604 coordinate $Mg^{2+}_{GTP}$ |
| 4RD1 | IF2G_SACS2 | GTP | protein | Mg | - | w-648 | w-601 in position $w^{HIS}$; w-887 in position $w^{ASP}$ | w-606 and w-611 coordinate $Mg^{2+}_{GTP}$ |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4RJL | IF2G_SACS2 | GCP | protein | Mg | - | - | w-606 in position $w^{HIS}$; w-625 in position $w^{ASP}$ | w-921 and w-604 coordinate $Mg^{2+}_{GTP}$;w-886 near $O^{1A}$ atom; w-981 near N2 |
| 4RCY | IF2G_SACS2 | GTP | protein | Mg | - | w-768 | w-644 in position $w^{HIS}$ | w-610 and w-604 coordinate $Mg^{2+}_{GTP}$ |
| 4M53 | IF2G_SACS2 | GCP | protein | Mg | - | w-889 | w-618 in position $w^{HIS}$; w-625 in position $w^{ASP}$ | w-602 and w-673 coordinate $Mg^{2+}_{GTP}$ |
| 6H6K | IF2G_SACS2 | GCP | protein | w-606 | - | w-617 | w-654 in position $w^{HIS}$ ; w-718 in position $w^{ASP}$ | w-774 and w-686 coordinate w-606; w-724 is coordinated by $O^{1G}$ atom |
| 4QFM | IF2G_SACS2 | GCP | protein | Mg | - | - | w-617 in position $w^{HIS}$; w-811 in position $w^{ASP}$ | w-940 and w-810 coordinate $Mg^{2+}_{GTP}$ |
| 4NBS | IF2G_SACS2 | GCP | protein | Mg | - | - | w-608 in position $w^{HIS}$; w-603 in position $w^{ASP}$ | w-601 and w-797 coordinate $Mg^{2+}_{GTP}$ |
| 2AHO | IF2G_SACS2 | GNP | protein | Mg | - | - | None | w-513 coordinates $Mg^{2+}_{GTP}$ |
| 4QHY | IF2G_SACS2 | GCP | protein | Mg | - | - | None | w-604 and w-605 coordinate $Mg^{2+}_{GTP}$ |
| 6SWC | IF2G_SACS2 | GNP | ribosome | Mg | - | - | None | None |
| 6SW9 | IF2G_SACS2 | GNP | ribosome | Mg | - | - | None | None |
| 6FYX | IF2G_YEAST | GCP | ribosome | Mg | - | - | None | Zn-301 near ribose oxygen O2' |
| 3JAP | IF2G_YEAST | GCP | ribosome | Mg | - | - | None | None |
| 6GAZ | IF2M_HUMAN | GSP | ribosome | Mg | - | Na | None | w-1001 and w-1002 coordinate $Mg^{2+}_{GTP}$ |
| 6GAW | IF2M_HUMAN | GSP | ribosome | Mg | Mg | Na | None | w-1001 and w-1002 coordinate $Mg^{2+}_{GTP}$ |
| 6GB2 | IF2M_HUMAN | GSP | ribosome | Mg | Mg | Na | None | w-1001 and w-1002 coordinate $Mg^{2+}_{GTP}$ |
| 6RW5 | IF2M_HUMAN | GNP | ribosome | Mg | - | - | None | w-901 and w-902 coordinate $Mg^{2+}_{GTP}$ |
| 4TMX | IF2P_CHATD | GTP | protein | Mg | - | Na | w-1089 in position $w^{HIS}$ | w-1085 and w-1086 coordinate $Mg^{2+}_{GTP}$ |
| 4TMV | IF2P_CHATD | GSP | protein | Mg | - | Na | w-1076 in position $w^{HIS}$ | w-1074 and w-1075 coordinate $Mg^{2+}_{GTP}$ |
| 4TMW | IF2P_CHATD | GTP | protein | Mg | - | Na | w-1059 in position $w^{HIS}$ | w-1060 and w-1061 coordinate $Mg^{2+}_{GTP}$ |
| 4TMT | IF2P_CHATD | GSP | protein | Mg | - | w-1070 | w-1066 in position $w^{HIS}$ | w-1064 and w-1065 coordinate $Mg^{2+}_{GTP}$ |
| 4NCN | IF2P_CHATD | GTP | protein | Mg | - | Na | w-1103 in position $w^{HIS}$ | w-1101 and w-1102 coordinate $Mg^{2+}_{GTP}$ |
| 4TMZ | IF2P_CHATD | GSP | protein | Mg | - | K | w-1015 in position $w^{HIS}$ | w-1013 and w-1014 coordinate $Mg^{2+}_{GTP}$; w-1016 and w-1104 coordinate $K^{+}$ |
| 4TN1 | IF2P_CHATD | GSP | protein | Mg | - | - | w-1028 in position $w^{HIS}$ | w-1029 and w-1030 coordinate $Mg^{2+}_{GTP}$ |
| 1G7T | IF2P_METTH | GNP | protein | Mg | - | - | None | w-844 coordinates $Mg^{2+}_{GTP}$; w-737 near $O^{3G}$ atom, but not in $w^{HIS}$ position |
| 4V85 | RF3_ECOLI | GNP | ribosome | Mg | - | - | None | w-532 takes place of the P-loop lysine amino group; w-531 coordinated by $Asp^{K-3}$ outside of the AG site |
| 4V89 | RF3_ECOLI | GNP | ribosome | Mg | - | - | None | None |
| 5LZD | SELB_ECOLI | GNP | ribosome | Mg | Mg | - | w-801 in position $w^{HIS}$ | w-802 coordinates $Mg^{2+}_{GTP}$ |
| 5IZL | SELB_HUMAN | GCP | protein | Mg | - | - | w-1110 in position $w^{HIS}$ | w-1104 and w-coordinate $Mg^{2+}_{GTP}$ |
| 2YWF | LEPA_AQUAE | GNP | protein | Mg | - | - | None | w-901, w-981 and w-993 coordinate $Mg^{2+}_{GTP}$ |
| 3JCE | LEPA_ECOLI | GNP | ribosome | Mg | - | - | None | None |
| 5J8B | LEPA_THET8 | GCP | ribosome | Mg | Mg | - | None | None |
| 5LZT | ERF3A_HUMAN | GCP | ribosome | Mg | - | - | None | None |
| 5M1J | HBS1_YEAST | GNP | ribosome | Mg | Mg | - | None | w-806 near $His^{D+4}$, but outside of the active site |
| 5LZZ | HBS1L_HUMAN | GCP | ribosome | Mg | - | - | None | None |
| 5LZW | HBS1L_HUMAN | GCP | ribosome | Mg | - | - | None | None |
| 5LZX | HBS1L_HUMAN | GCP | ribosome | Mg | - | - | None | None |
| 5LZY | HBS1L_HUMAN | GCP | ribosome | Mg | - | - | None | None |
| 6GZ3 | *N/A* | GNP | ribosome | - | Mg | - | None | None |
| 6GZ4 | *N/A* | GNP | ribosome | - | Mg | - | None | None |
| 6GZ5 | *N/A* | GNP | ribosome | - | Mg | - | None | None |
| 4ZKE | SKI7_YEAST | GTP | protein | Mg | - | Na | w-916 in position $w^{HIS}$ | w-928 and w-951 coordinate $Mg^{2+}_{GTP}$ |

"w-" followed by a number indicates a water molecule

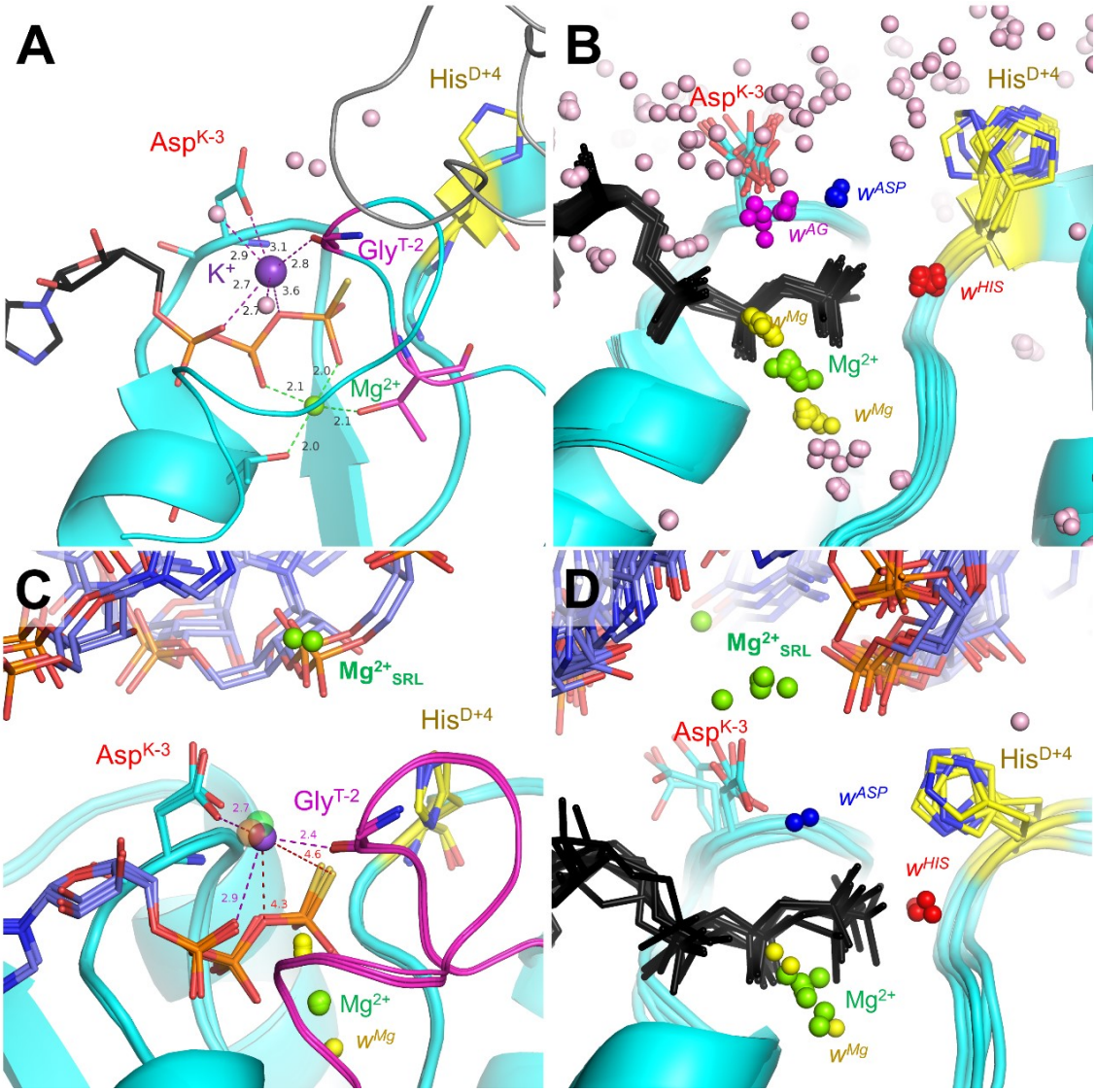


**Figure 3.2.18.** Cation binding in cryo-EM and crystallographic structures of translation factors. Proteins are shown as light blue cartoons, monovalent cations are shown as purple spheres, $Mg^{2+}$ ions - as green spheres. Water molecules outside common binding sites are shown as pink spheres, water molecules within common binding sites are colored as follows: $w^{HIS}$ – red, $w^{ASP}$ – blue, coordinating $Mg^{2+}$ bound to GTP analog – yellow. GTP analogs are shown as black sticks, catalytic $His^{D+4}$ residues are shown in yellow, SRL regions of the ribosome are shown as blue sticks with oxygen and phosphorus atoms colored red and orange, respectively. A. $K^+$ ion bound to the separately crystallized translation initiation factor eIF5B (517-858 fragment) (PDB ID 4TMZ). B. Water molecules in the AG site of separately crystalized translation factors (PDB IDs 1EFT, 4PC7, 1EXM, 4LBY, 4LBZ, 4LC0, 3VMF, 3WXM, 4RCY, 4M53, 6H6K, 2C77, 4RD4, 4RJL). C. Cations binding in the AG site of translation factors bound to the ribosome. The green, orange, and magenta cations bound in the AG site correspond to the structures with PDB IDs 6GAW, 6GAZ, and 6GB2, respectively). D. Water molecules in the GTP-binding sites of ribosome-bound translation factors (PDB IDs 4V9O, 4V90, 4V9H, 5LZD, 5UYM, 5J8B, 5M1J). The full list of analyzed structures is available in Table 3.7.

The solo structures of eIF5B with $K^+$ and $Na^+$ ions bound (see Figure 3.2.18, Table 3.7 and [21, 188]) demonstrate that translational GTPases provide five ligands for $K^+$ ion: these are the $O^{2A}$, $O^{3B}$, and $O^{3G}$ atoms of the phosphate chain, the side chain of the $Asp^{K-3}$ residue of the P-loop,

and the carbonyl oxygen of $Gly^{T-2}$ of the Switch I loop [21]. Four ligands are provided by the P-loop and tightly bound GTP molecule, and one ligand, $Gly^{T-2}$, is provided by the flexible Switch I loop. Hence, the formation of a functional $K^+$-binding site specifically requires stabilization of the Switch I loop, which could be physiological, within a ternary complex docked to the SRL, or non-physiological, being caused by intracrystal contacts as in the eIF5B structures. It is noteworthy that some of the eIF5B structures contain non-hydrolyzed GTP molecules in spite of a full-fledged ion binding site (with a $Na^+$ ion bound). One of the reasons might be the non-active conformation of $His^{D+4}$ of the Switch II motif in all these structures. Hence, activation of hydrolysis requires a proper interaction of P-loop with Switch I and Switch II loops.

Figure 3.2.18C and D show representative structures of translational GTPases that were determined in the presence of ribosomes; these structures provide information on the potentially activating interactions between the GTPase domain, tRNA, and SRL, see also Table 3.7 for the list of analyzed structures. In these structures, the $His^{D+4}$ residue is turned towards the phosphate chain because of the H-bond with the $O^{1P}$ atom of the phosphate group of A2662 nucleotide [127, 167, 183].

A $Na^+$ ion was recently reported in the AG position next to the phosphate chain for three cryo-EM structures of the mitochondrial initiation factor 2 (mtIF2) bound to the ribosome [237]. These ions are bound by $Asp190^{K-3}$, $Gly212^{T-2}$, and the $O^{2A}$ atom of the phosphate chain. However, the respective coordination bonds, as indicated in Figure 3.2.18, are too long for a $Na^+$ ion, cf. refs. [203, 235, 238] and Table 3.2. Furthermore, the Methods section of the respective article reports that only $K^+$ and $Mg^{2+}$ ions were present in the solutions used for dissolving the samples [237]. Since $Mg^{2+}$-coordinating bonds (2.0-2.3 Å) are even shorter than for a $Na^+$ ion [235], the respective atom in these three structures could be either a $K^+$ ion or an oxygen atom of a water molecule. The latter possibility is less likely because the atom in the AG position makes bonds with the side chain oxygen atom of $Asp190^{K-3}$, carbonyl oxygen of $Gly212^{T-2}$, and $O^{2A}$ atom of the phosphate chain (see the 6GB2 structure in Figure 3.2.18C). A water oxygen atom can hardly interact with three surrounding oxygen atoms. Hence, it is plausible that these three structures contain the long-sought $K^+$ ion almost in the AG site of ribosome-bound mtIF2s. Unfortunately, this ion/atom could not make the key bond with the $O^{3G}$ atom of γ-phosphate because the co-crystallized non-hydrolyzable GTP analog had a sulfur atom instead of $O^{3G}$. Further efforts would be needed to obtain a translation factor structure with a fully bound $K^+$ ion.

The superposition of many representative structures in Fig. 2D shows a $Mg^{2+}$ ion between the $Asp21^{K-3}$ of the translation factor and the phosphate of A2662 nucleotide, i.e. the same phosphate group that interacts with the $His^{D+4}$ residue via its $O^{1P}$ atom, see Table 3.7 and [167,

183, 239, 240] for previous reports on a $Mg^{2+}$ ion in this position. Hereafter this $Mg^{2+}$ ion is denoted as $Mg^{2+}_{SRL}$ to distinguish it from the $Mg^{2+}$ ion interacting with the β- and γ-phosphates of GTP (hereafter $Mg^{2+}_{GTP}$). Such $Mg^{2+}_{SRL}$ ions were identified in 17 structures of ribosome-bound translational GTPases, as indicated in Table 3.7. As it is seen from Figure 3.2.18D, the position of this $Mg^{2+}$ ion varies in different structures; the $Mg^{2+}$ ion is sometimes closer to the A2662 nucleotide residue, whereas, in other structures, it is closer to the $Asp^{K-3}$ residue. The distance between the $Mg^{2+}_{SRL}$ ion and the nearest oxygen atom of the A2662 phosphate varies between 3.2Å and 4.7Å, whereas the distance to the nearest oxygen of the $Asp^{K-3}$ carboxyl group varies from 2.7Å to 7Å owing to different orientations of the $Asp^{K-3}$ side chain.

Because of the lower resolution as compared to the solo structures (cf. Figure 3.2.18A, B), only a few water molecules could be identified in the catalytic sites of translational GTPases bound to ribosomes. Several structures contain a $w^{HIS}$ molecule between the $His^{D+4}$ residue and γ-phosphate (Figure 3.2.18D); this water molecule was suggested to be the "catalytic" one [127, 183]. In two structures (PDB IDs 4V90 and 4V9H), a $w^{ASP}$ molecule is bound between the $Asp^{K-3}$ residue and the $O^{3G}$ atom of γ-phosphate, see Table 3.7 and Figure 3.2.18D. Both these structures also report a $Mg^{2+}_{SRL}$ ion bound (Figure 3.2.18D).

The EF-Tu and its close homologs appear to be the best-studied translational GTPases; they were crystallized both solo and in different ribosomal complexes many times. This abundance of structures allowed us to compare the shapes of the Switch I loop in differently prepared samples. In $K^+$-dependent NTPases, Switch I has a shape of an elongated loop (K-loop) and coordinates the $K^+$ ion by backbone carbonyl oxygen atoms of the residues in positions T-4 and T-6 (Figure 3.1.4). In translational GTPases, this segment folds into short α-helix(es) so that only backbone oxygen of $Gly^{T-2}$, owing to a backbone protrusion, is close enough to coordinate a cation in the AG position, as seen in the structures shown in Figure 3.2.18.

The Switch I region is disordered and usually not resolved in the structures of solo translational GTPases; however, in few structures of EF-Tu, the full length of this region is resolved, most likely owing to stabilization by intracrystal contacts, even in the absence of activating partners, see gray cartoons/structures in Figure 3.2.18A. The structures of aa-tRNA-bound GTPases show that tRNA stabilizes the Switch I closer to the catalytic site (cyan cartoons/structures in Figure 3.2.18A). Finally, in the structures of ternary complexes bound to ribosomes, the Switch I is pushed even closer to the active site with the backbone oxygen of $Gly^{T-2}$ approaching the phosphate chain of GTP at a distance of up to 5.3 Å (green cartoons/structures in Figure 3.2.19). This data illustrates how the gradual formation of a functional EF-Tu/tRNA/ribosome complex brings the Switch I loop closer to the phosphate chain. Specifically,

the codon recognition shortens the distance between the two $K^+$-ligands – the $O^{3B}$ atom of GTP and the carbonyl oxygen of $Gly^{T-2}$ – enabling the binding of a $K^+$ ion between them (Figure 3.2.19). These data are in agreement with earlier evidence on the involvement of aa-tRNA in activation of GTP hydrolysis [241, 242].

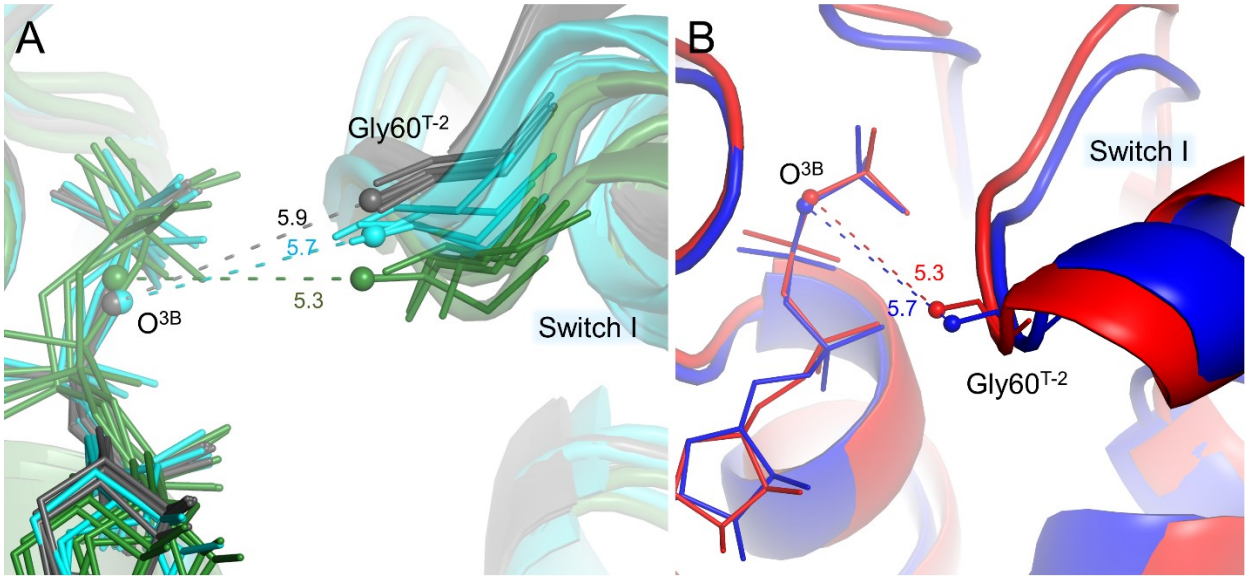


**Figure 3.2.19.** Conformation changes of the Switch I loop of the EF-Tu. A. Superposition of EF-Tu structures shows the difference in Switch I position relative to the triphosphate chain. Structures of translation factors that were crystallized solo are shown in gray (PDB IDs 4LC0, 4LBZ, 4LBY, 4LBW, 4LBV, 4H9G, 1EXM, 1EFT); structures of EF-Tu-tRNA complexes are shown in cyan (PDB IDs 1OB2, 1OB5, 1TTT, 1B23); structures of ternary complexes bound to the ribosome are shown in green (PDB IDs 5WE4, 5WDT, 5UYQ, 5UYM, 4V5L). The EF-Tu binding to the tRNA and subsequent binding to the ribosome causes tightening of the K-binding site, as indicated by shortening of the distance between $O^{3B}$ atom of GTP (or the corresponding atom of GTP analog) and the backbone oxygen of $Gly60^{T-2}$. B. Superposition of EF-Tu/aa-tRNA/ribosome complex at different stages of the elongation cycle obtained from time-resolved cryogenic electron microscopy. In blue: cognate structure II-A, compact EF-Tu/GTP complex on the ribosome with the SSU in the open conformation (PDB ID 6WD2 [182]). In red: cognate structure III-A, EF-Tu/GTP complex pre-GTP-hydrolysis state on the ribosome with the SSU in the closed conformation (PDB ID 6WD8 [182]).

Recently Korostelev and colleagues, by using time-resolved cryogenic electron microscopy, revealed 33 states through which programmed ribosomes were passing [182]. These structures were not included in the comparative analysis because $Mg^{2+}{}_{GTP}$ ions were not resolved in them. Still, several of these structures contained GTP so they were analyzed separately. These structures were expected to capture the dynamics of secondary structure elements such as helices or loops. Figure 3.2.19B shows the overlap of a structure obtained immediately before the codon recognition (Step II-A, PDB ID 6WD2, blue) and that obtained after it, but before GTP hydrolysis (Step III-A, PDB ID 6WD8, red). The Switch I loop with $Gly^{T-2}$ residue appears to get closer to the phosphate chain after the codon recognition.

### *3.2.5. Molecular Dynamics Simulations of EF-Tu*

To check whether a $K^+$ ion could bind in the GTP-binding site of a ribosome-bound translational GTPase, MD simulations of EF-Tu/GTP complex bound to the fragments of aa-tRNA and ribosome in the presence of potassium ions were performed. As a starting system for simulations crystallographic structure of ribosome-bound EF-Tu from *Thermus thermophilus* was chosen (PDB ID 4V5L [127], Figure 3.1.4A).

MD simulations of an entire ribosome complex would require an enormous amount of computational resources. Since this study is focused on the active site of the translation factor, the simulation system included only relevant parts of the ribosomal complex: the entire EF-Tu, the SRL region of the large ribosomal subunit (23S rRNA, nucleotides 2655-2664, here and further numbering is given for PDB ID 4V5L) and the fragment of Trp-tRNA in the proximity of the P-loop domain of EF-Tu (nucleotides 1-6 and 64-77). Non-hydrolyzable substrate analog GCP was changed into a GTP molecule by replacing a carbon atom between β and γ phosphates with an oxygen atom. The $Mg^{2+}{}_{GTP}$ ion and the $w^{HIS}$ molecule (number 2001) that were present in the crystal structure were retained (Figure 3.1.4A). One more $Mg^{2+}$ ion was placed near the SRL loop to serve as $Mg^{2+}{}_{SRL}$ seen in 17 structures of ribosome-bound translational GTPases, see Table 3.7 and Figure 3.2.18D. Finally, two water molecules were placed manually near the $Mg^{2+}{}_{GTP}$ ion, in the same positions as in the experimental structures (see Table 3.7 and Figure 3.2.18B, D). In all simulations, $K^+$ and $Mg^{2+}$ ions were added to the solution to mimic their physiological concentrations, and $Cl^-$ ions were added to bring the total sum of charges in the system to zero. The resulting solvated and ionized complexes were used to run the productive MD simulations, after energy minimization and equilibration of each system, see the Methods section for further details.

MD simulations were performed for two distinct systems. One system (see simulation #1 in Table 2.3 of the Methods section) had a $K^+$ ion placed manually in the AG site to mimic the ion seen in some crystal structures (Figure 3.2.18). In the second system (simulations #2-6 in Table 2.3), a $K^+$ ion was not initially put in the AG site; all $K^+$ ions were placed randomly in the solution around the protein/tRNA/SRL complex. Simulations of the second system were performed five times (simulations #2-6 in Table 2.3). The duration of each productive MD simulation run was 100 ns.

### 3.2.5.1 Cation binding in the catalytic pocket

In the simulation where a $K^+$ ion was initially placed in the space between the $O^{2A}$ and $O^{3G}$ oxygen atoms of the phosphate chain (the AG site, see Figure 3.2.20), this ion remained there for the whole duration of the simulations (tracks #1 in Figure 3.2.21). In the simulations where all $K^+$ ions were placed randomly in the solution, one of these free cations entered the catalytic site and occupied the AG site within the first 12 ns of the simulation; the $K^+$ ion remained bound throughout each of the five simulations (tracks #2-6 in Figure 3.2.21). Potassium ion in the AG site ($K^+_{AG}$) was coordinated by $O^{2A}$, $O^{3B}$, and $O^{3G}$ atoms of the GTP phosphate chain, the side chain oxygen atom of $Asp21^{K-3}$, the backbone oxygen of $Gly60^{T-2}$ and a water molecule (hereafter $w^K$, see Figure 3.2.20). These bonds mostly remained stable during simulations (Figure 3.2.21).

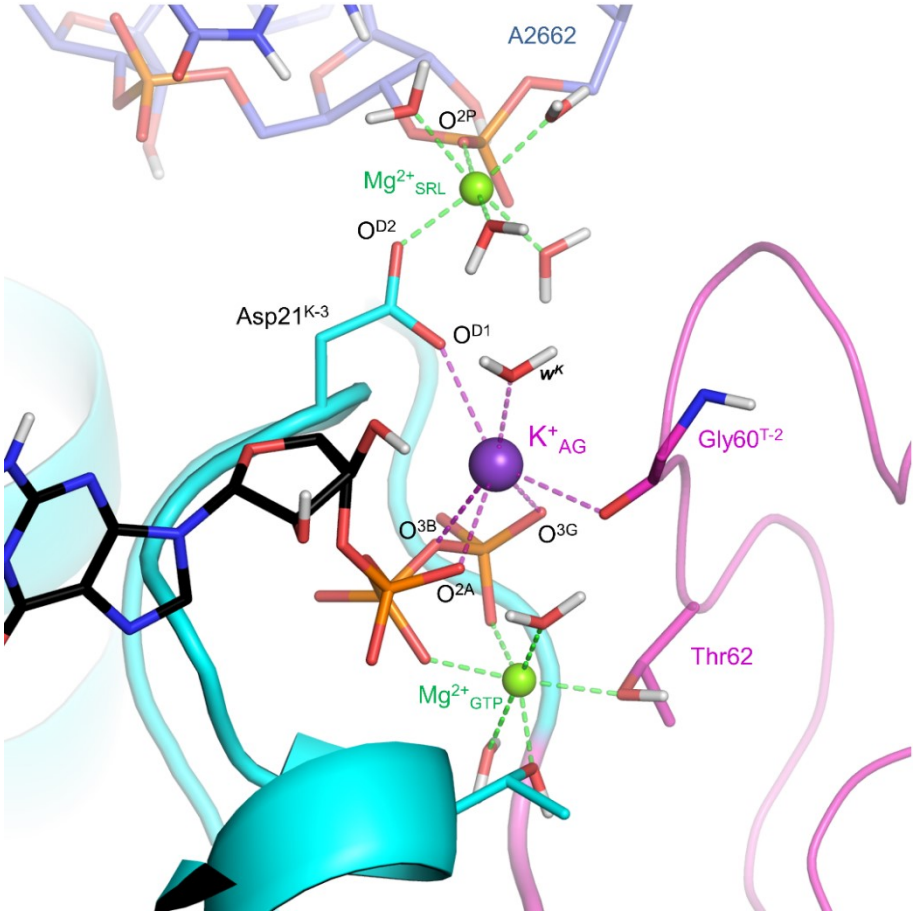


**Figure 3.2.20.** Cation binding in the P-loop domain of EF-Tu as revealed from MD simulations (the shown conformation was sampled from simulation #3 in Table 2.3). Protein is shown as a light blue cartoon, protein loop containing the Switch I motif and its residues are shown in magenta, $K^+$ ion - purple, $Mg^{2+}$ ions - green, SRL fragment is shown as blue sticks with oxygen and phosphorus atoms colored red and orange, respectively. Functionally relevant protein residues are shown as sticks.

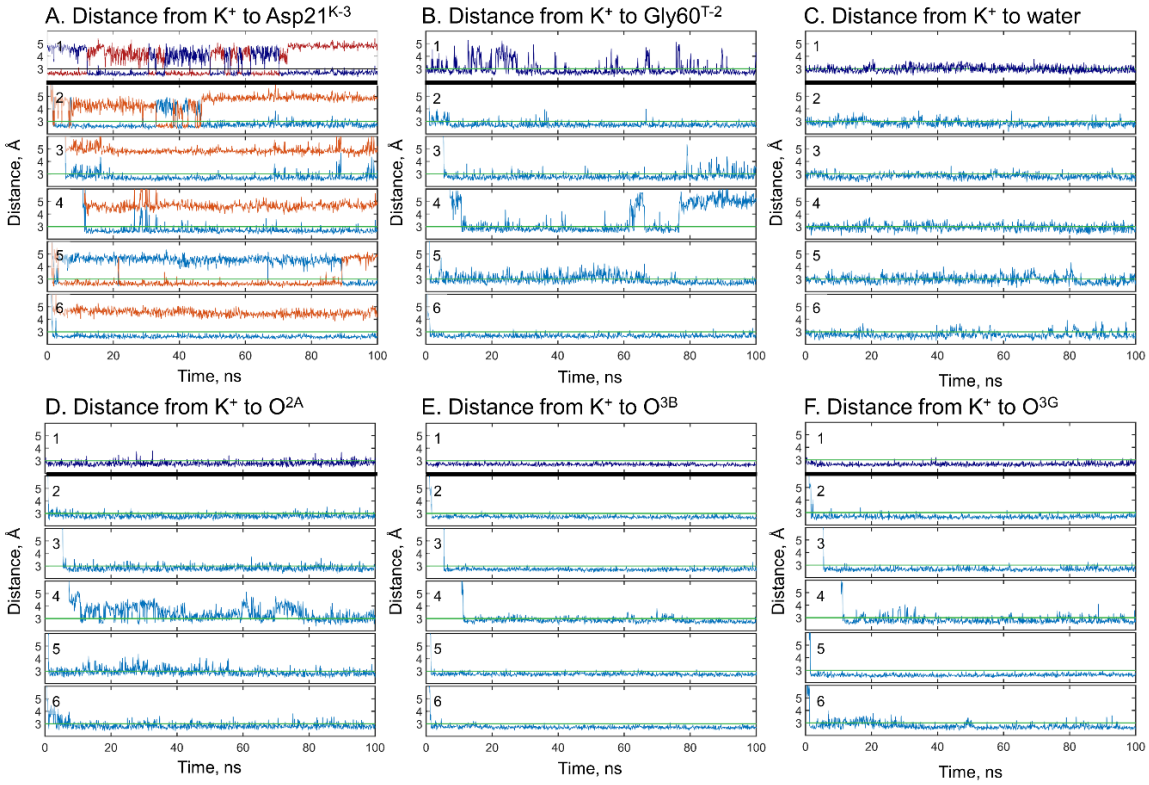


**Figure 3.2.21.** Interactions of the $K^+$ ion during MD simulations, see Figure 3.2.20 for the structural information. A. Distances from the $K^+$ ion to the $O^{D1}$ (blue) and $O^{D2}$ (red/orange) atoms of the $Asp21^{K-3}$ of the P-loop. B. Distance from the $K^+$ ion to the backbone oxygen atom of $Gly60^{T-2}$. C. Distances from the $K^+$ ion to the nearest water molecule ($w^K$). D, E, F – Distances from the $K^+$ ion to the oxygen atoms $O^{2A}$, $O^{3B}$, and $O^{3G}$ of the GTP phosphate chain. Green lines mark the distance of 3Å. Track numbers correspond to simulations in Table 2.3.

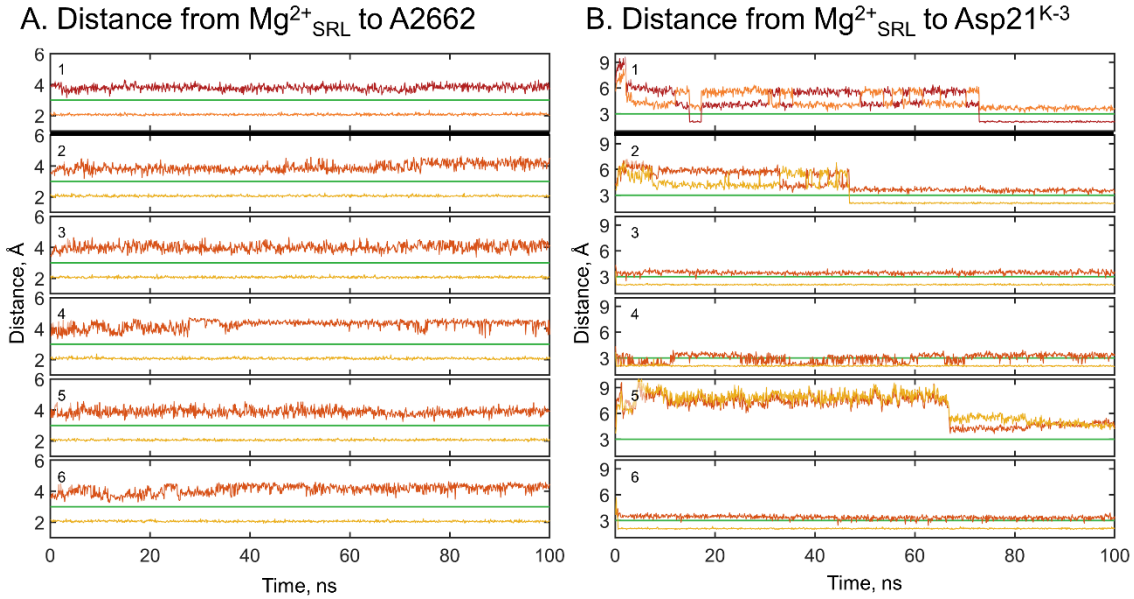


**Figure 3.2.22.** Distances between $Mg^{2+}_{SRL}$ and surrounding oxygen atoms, see Figure 3.2.20 for the structural information. A. Distances from $Mg^{2+}_{SRL}$ to the $O^{1P}$ (red) and $O^{2P}$ (orange) atoms of the A2662 residue of the SRL. B. Distances from $Mg^{2+}_{SRL}$ to the $O^{D1}$ (red) and $O^{D2}$ (orange) atoms of the $Asp21^{K-3}$ of the P-loop. Green lines mark the distance of 3Å. Track numbers correspond to simulations in Table 2.3.

In all six MD simulations, the $Mg^{2+}_{SRL}$ ion remained bound to the $O^{2P}$ atom of the phosphate group of the A2662 residue (Figure 3.2.20, Figure 3.2.22A). In five simulations out of six, this $Mg^{2+}_{SRL}$ ion becomes simultaneously coordinated by the side chain of $Asp21^{K-3}$ thus linking the phosphate group of SRL with the EF-Tu (Figure 3.2.20, Figure 3.2.22B). In simulations #1 and #2, however, this bond was only established after 73 and 48 ns of simulation, respectively, this bond never formed, and in simulation #5. It is noteworthy that the $Mg^{2+}_{SRL}$ ion remained bound and did not exchange with $Mg^{2+}$ ions from the solution during the simulations.

During the simulations, additional cations from the solution (both $Mg^{2+}$ and $K^+$) are bound to the RNA fragments. Several additional $Mg^{2+}$-binding sites can be seen next to the SRL fragment (Figure 3.2.23). One such site is next to the C2658 nucleotide residue, the other site is between the phosphate group of G2663 and the side chain of Glu55 of the Switch I region. Furthermore, a $K^+$ ion was often seen bound to the SRL near the G2661 residue. Further $Mg^{2+}$ and $K^+$ ions are seen next to the tRNA fragment, but all these sites are located further than 12Å from the GTP-binding site. Finally, a single $K^+$ ion binds between the backbone and the side chain oxygens of Asp51 of the Switch I region (Figure 3.2.23).

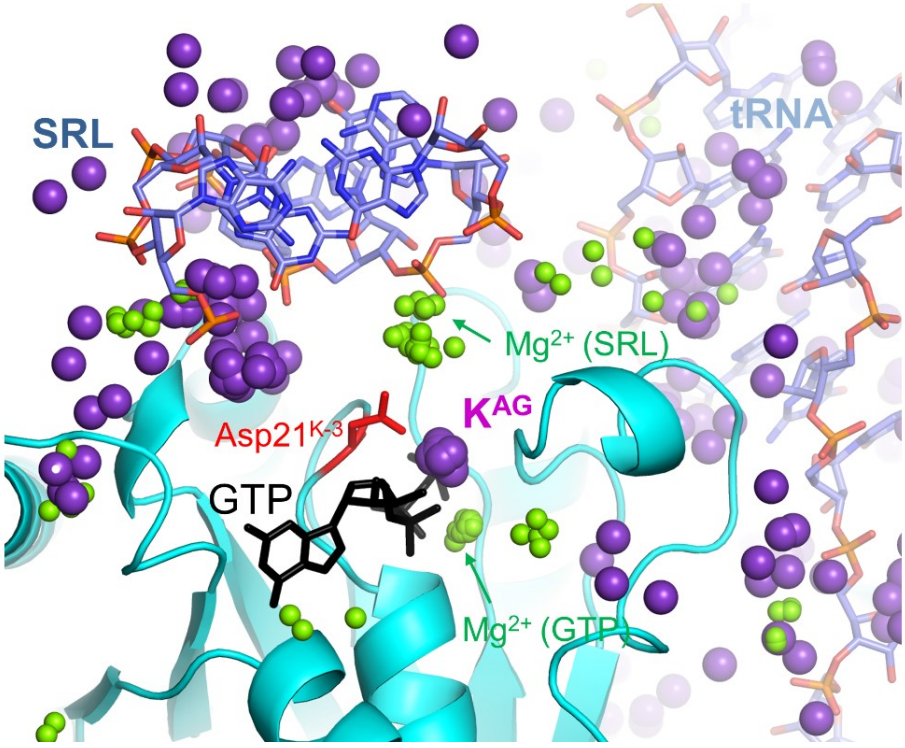


**Figure 3.2.23**. Occupancy of $K^+$- and $Mg^{2+}$-binding sites during MD simulations. Cation positions from 20 structures are shown, structures were samples from simulations 1, 2, 3, and 6 (as shown in Table 1) with 20 ns period. Protein is shown as a light blue cartoon, $K^+$ ions are shown as purple spheres, $Mg^{2+}$ ions - as green spheres, SRL, and tRNA fragments are shown as blue sticks with oxygen and phosphorus atoms colored red and orange, respectively. GTP molecule is shown as black sticks, $Asp21^{K-3}$ is shown as red sticks.

### 3.2.5.2 Hydrogen bonded network in the catalytic site and its changes in response to the $K^+$ binding

The MD simulations revealed a complex bonding network that involves the GTP molecule, A2662 residue of SRL, $Mg^{2+}$ and $K^+$ ions, as well as many amino acids and water molecules (Figure 3.2.27). The core of this network is formed by the aforementioned bonds which stabilize the triphosphate chain in its extended, strained conformation. Conserved hydrogen and coordinational bonds in the NTP-binding site of P-loop NTPases are listed in Table 3.8 and some of them are shown in Figure 3.2.24. Specifically, the $O^{1A}$ atom of α-phosphate forms H-bonds with the backbone amino group of the $Thr26^{K+2}$ residue. The $O^{1B}$ atoms form H-bonds with the backbone amino groups of the signature Lys24 residue and the $Gly23^{K-1}$ residue, an invariant residue of the P-loop motif. The $O^{3A}$ atom between α- and β-phosphates forms a weak H-bond with the backbone amino group of $Gly23^{K-1}$. The $O^{2B}$ atom makes an H-bond with the backbone amino group of the $Thr25^{K+1}$ residue. The $O^{3B}$ atom between β- and γ-phosphates forms an H-bond with the backbone amino group of the $Asp21^{K-3}$ residue. In addition, the $O^{1B}$ and $O^{2G}$ atoms form H-bonds with the side chain amino group of the signature Lys24 residue; its amino group is additionally stabilized by an H-bond with the carbonyl oxygen atom of $His19^{K-5}$ and a weak H-bond with the carbonyl oxygen atom $Gly18^{K-6}$. The $O^{2B}$ and $O^{1G}$ atoms are coordinated by the $Mg^{2+}{}_{GTP}$ ion, which is, in its turn, coordinated by further protein residues. Most of these interactions are common for all P-loop fold NTPases, as they involve universally conserved residues or backbone amino and carbonyl groups of the P-loop; this bonding pattern is seen in high-resolution structures of EF-Tu [236]. In all six simulations, all these bonds remained stable, indicating a proper GTP binding throughout the simulations (Figure 3.2.25).

In addition to these "universal" bonds, two more H-bonds are common for the proteins of the TRAFAC class. First, the backbone amino group of the signature Thr residue of Switch I forms an H-bond with the $O^{3G}$ atom of γ-phosphate whereas its side chain interacts with the $Mg^{2+}$ ion, see Figure 3.2.27. In MD simulations of EF-Tu, the H-bond between the $O^{3G}$ atom of γ-phosphate and the backbone amino group of the signature Thr62 of Switch I was mostly stable and broke only in the middle of simulation # 4 (Figure 3.2.26). Second, an H-bond is often formed between the $O^{2G}$ atom of γ-phosphate and the backbone amino group of $Gly^{D+3}$ residue of the Walker B motif, see Figure 3.2.24, Figure 3.2.26 and [121]. This bond between $Gly84^{D+3}$ and $O^{2G}$ atom of γ-phosphate was present only in simulation runs #2 and # 6 and was absent in other simulations (Figure 3.2.26), where the interaction between $Gly84^{D+3}$ and $O^{2G}$ appeared to be mediated by the $w^{HIS}$ molecule (Figure 3.2.27).

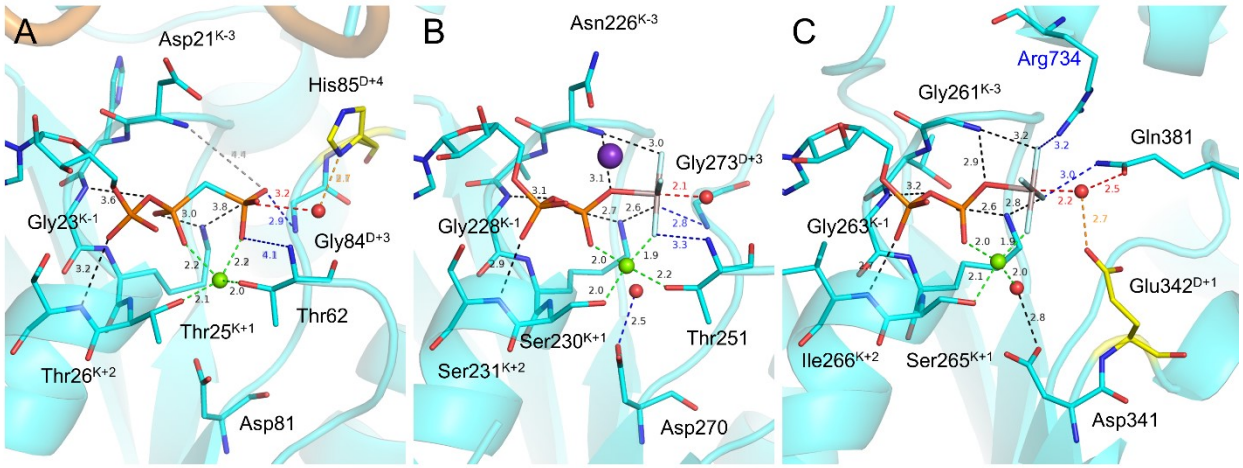


**Figure 3.2.24.** Universally conserved and class-specific bonds in P-loop GTPases and ATPases. Universal bonds are shown as follows: hydrogen bonds between protein and GTP/ATP are shown in black, $Mg^{2+}$ coordination bonds are shown in green, distances from catalytic water molecule to the $P^G$ or corresponding atom are shown in red. Blue dashes show hydrogen bonds, different between protein families, yellow dashes show bonds with the catalytic residues, also different between protein families. Proteins are shown as blue cartoons, GTP, ATP, and functionally relevant protein residues are shown as sticks, catalytic residues are highlighted in yellow. Ions and water molecules are shown as spheres, water molecules in red, $K^+$ in purple, $Mg^{2+}$ in green. A. EF-Tu (PDB ID 4V5L). B. GTPase MnmE (PDB ID 2GJ8). C. Helicase Pif1 (PDB ID 5O6B).

**Table 3.8**. Conserved hydrogen and coordinational bonds in the NTP-binding site of P-loop NTPases

| **Bond** | | | **EF-Tu (PDB 4V5L)** | | **MnmE (PDB 2GJ8)** | | **Ras (PDB 1WQ1)** | | **Pif1 (PDB 5O6B)** | |
|---|---|---|---|---|---|---|---|---|---|---|
| Residue position | atom | (NTP) partner atom | protein residue | Dist. (Å) | protein residue | Dist. (Å) | protein residue | Dist. (Å) | protein residue | Dist. (Å) |
| N/A | MG | $O^{2B}$ | N/A | 2.2 | N/A | 2.0 | N/A | 1.8 | N/A | 2.0 |
| N/A | MG | $O^{1G}$ | N/A | 2.2 | N/A | 1.9* | N/A | 2.3* | N/A | 1.9* |
| K | NZ | $O^{1B}$ | Lys24 | 3.0 | Lys229 | 2.8 | Lys16 | 2.8 | Lys264 | 2.6 |
| K | NZ | $O^{2G}$ | Lys24 | 3.8 | Lys229 | 2.9* | Lys16 | 2.8* | Lys264 | 2.8* |
| K+2 | N | $O^{1A}$ | Thr26 | 3.2 | Ser231 | 2.9 | Ala18 | 2.8 | Ile266 | 2.7 |
| K+2 | $O^{G1}$ | $O^{1A}$ | Thr26 | 2.8 | Ser231 | 2.8 | Ala18 | N/A | Ile266 | N/A |
| K-1 | N | $O^{3A}$ | Gly23 | 3.6 | Gly228 | 3.2 | Gly15 | 3.3 | Gly263 | 3.2 |
| K-1 | N | $O^{1B}$ | Gly23 | 3.2 | Gly228 | 3.0 | Gly15 | 3.0 | Gly263 | 3.1 |
| K | N | $O^{1B}$ | Lys24 | 2.7 | Lys229 | 2.9 | Lys16 | 2.8 | Lys264 | 3.0 |
| K+1 | N | $O^{2B}$ | Thr25 | 3.5 | Ser230 | 2.9 | Ser17 | 3.4 | Ser265 | 3.0 |
| K | NZ | K-6 (O) | Lys24-Gly18 | 3.3 | Lys229-Gly223 | 3.7 | Lys16-Gly10 | 3.3 | Lys264-Ser259 | 3.0 |
| K | NZ | K-5 (O) | Lys24-His19 | 2.8 | Lys229-Arg224 | 2.9 | Lys16-Ala11 | 2.7 | Lys264-Gly258 | 3.2 |
| K-3 | N | $O^{3B}$ | Asp21 | 2.7* | Asn226 | 3.0 | Gly13 | 2.7 | Gly261 | 2.9 |
| K-3 | N | $O^{2G}$ | Asp21 | 4.4 | Asn226 | 3.1 | Gly13 | 3.9* | Gly261 | 3.2* |

* - Distances measured to carbon or fluoride atoms that replace oxygen atoms in NTP analogs

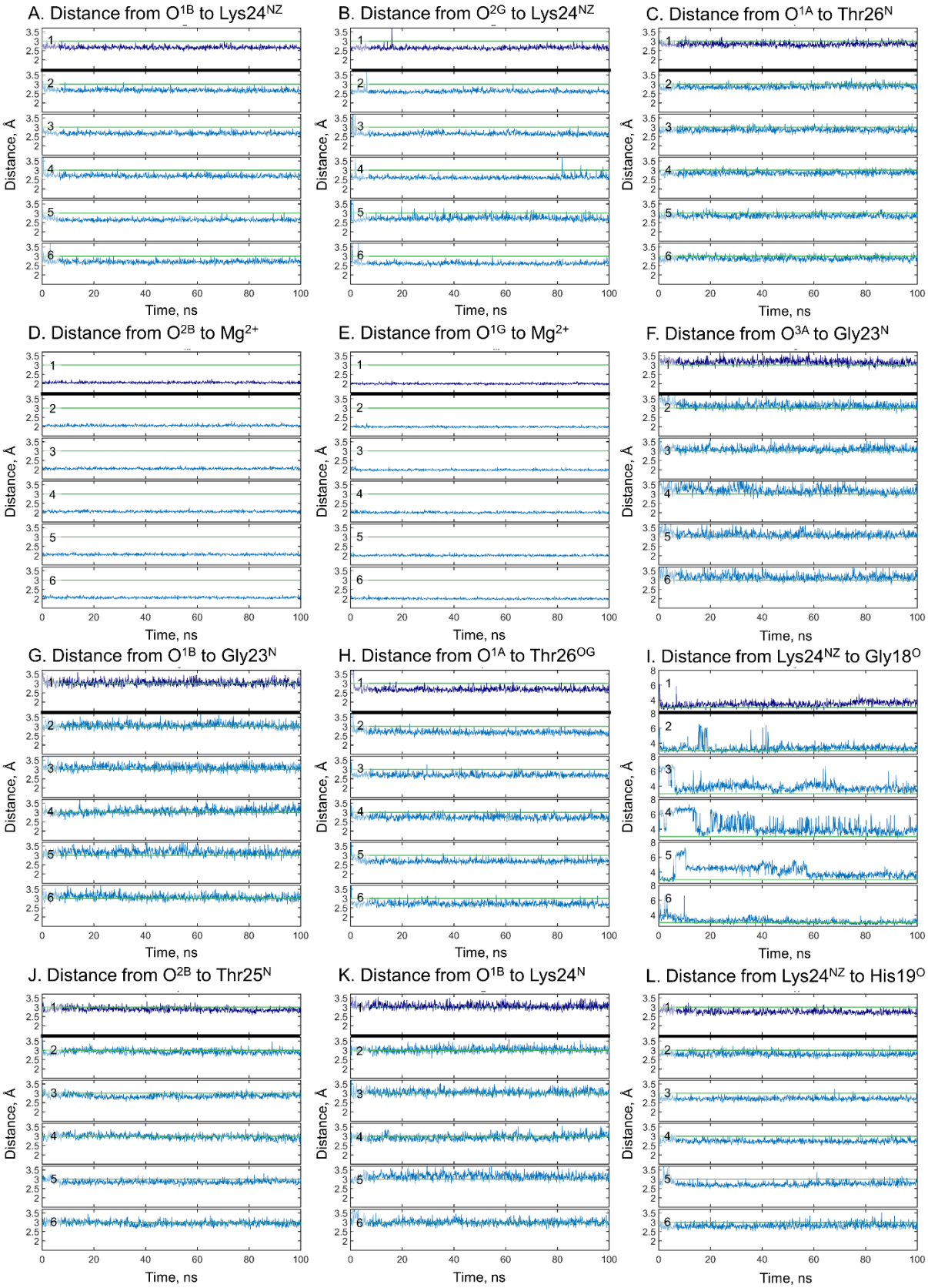


**Figure 3.2.25**. Conserved hydrogen bonds in the active site of EF-Tu. Green lines mark the distance of 3Å. Track numbers correspond to simulations in Table 2.3.

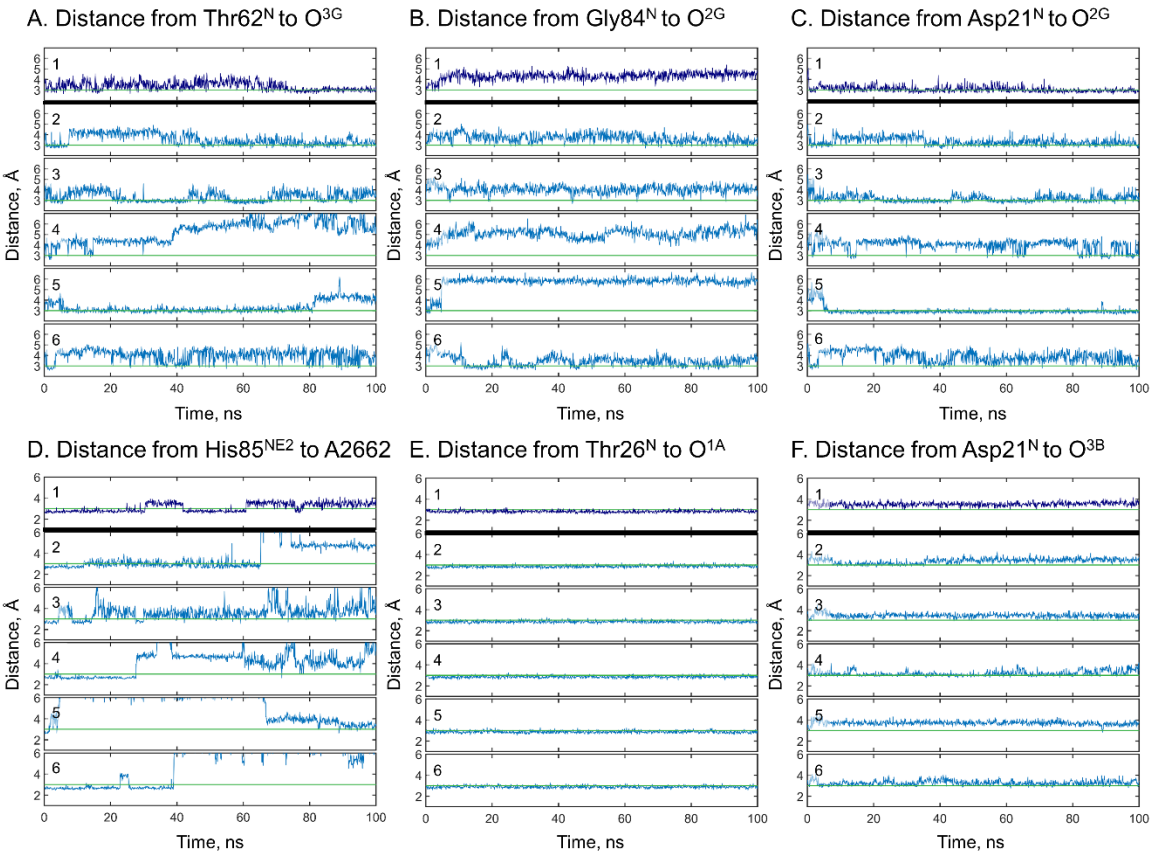


**Figure 3.2.26.** Hydrogen bond network in the catalytic site of EF-Tu. A. Distance from the backbone nitrogen of Thr62 of the Switch I to the $O^{3G}$ atom of γ-phosphate. B. Distance from the backbone nitrogen of $Gly84^{D+3}$ to $O^{2G}$ atom of γ-phosphate. C. Distance from the backbone nitrogen of $Asp21^{K-3}$ to the $O^{2G}$ atom of γ-phosphate. D. Distance from the $N^{E2}$ atom of $His85^{D+4}$ to the $O^{1P}$ atom of A2662 residue of SRL. E. Distance from the backbone nitrogen of $Thr26^{K+2}$ to the $O^{1A}$ atom of α-phosphate. F. Distance from the backbone nitrogen of $Asp21^{K-3}$ to the $O^{3B}$ atom of β-phosphate. Green lines mark the distance of 3Å. Track numbers correspond to simulations in Table 2.3.

In the MD simulations, $His^{D+4}$ of EF-Tu was facing the γ-phosphate and forming an H-bond between the $N^{E2}$ atom of its side chain and $O^{1P}$ oxygen of the A2662 phosphate (Figure 3.2.27, Figure 3.2.26). In this conformation, the $N^{D1}$ atom of the $His^{D+4}$ formed an H-bond with the "catalytic" water molecule $w^{HIS}$, which, in its turn, made an H-bond with the $O^{2G}$ atom of the γ-phosphate. In most simulations, $His^{D+4}$ remained oriented towards the phosphate chain, but its imidazole ring rotated several times (Figure 3.2.26D).

MD simulations #2-6, where the $K^+$ ion was not initially placed in the AG site, reveal the changes in the catalytic site that happened in response to the $K^+$ binding. The $K^+$ ion, by linking $O^{2A}$ and $O^{3G}$ atoms of α- and γ-phosphates, caused rotation of the γ-phosphate group (Figure 3.2.28). The same $K^+$-induced rotation made possible a new H-bond between the $O^{2G}$ atom and the backbone amino group of $Asp21^{K-3}$ residue. This bond was present in all MD simulations

(Figure 3.2.26C). Also, the rotation of γ-phosphate resulted in an almost eclipsed conformation of the triphosphate chain (Figure 3.2.28A).

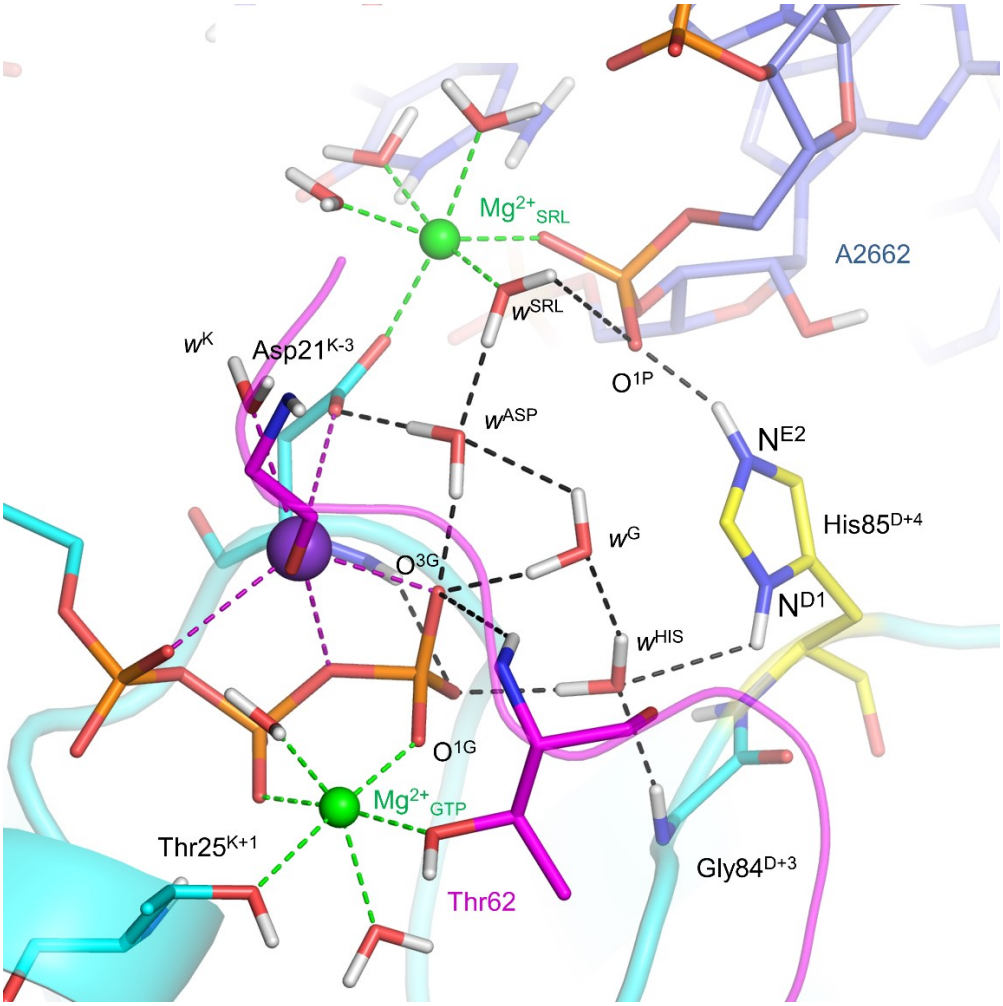


**Figure 3.2.27.** Bonding network and water molecules in the catalytically active conformation of EF-Tu as revealed from MD simulations (the shown conformation was sampled from simulation #3 in Table 2.3). Protein is shown as a light blue cartoon, $K^+$ ion is shown in purple, $Mg^{2+}$ ions are shown in green, SRL fragment is shown as blue sticks with oxygen and phosphorus atoms colored red and orange, respectively. Functionally relevant protein residues and water molecules are shown as sticks, the residues of the Switch I motif are shown in magenta, the catalytic $His^{D+4}$ residue is shown in yellow.

The $K^+$ ion, by "fixing" the conformation of $Asp21^{K-3}$ residue and of the Switch I loop, specifically stabilized the entire H-bonded network of water molecules in the catalytic pocket (Figure 3.2.27). To visualize the stability of individual water molecules in each position, the distances from each water molecule to the surrounding moieties were traced; Figure 3.2.29 shows the data from simulation #3, whereas data from other simulations are shown in Supplementary Figures S3-S7. Remarkably, despite the high mobility of $His^{D+4}$, the molecule $w^{HIS}$ remained stable through MD simulations (Figure 3.2.29, S3-S7). It was additionally H-bonded with the backbone amino group of $Gly84^{D+3}$ so that even when $His^{D+4}$ was not in a position to provide an H-bond, the

$w^{HIS}$ molecule stayed in its position. That said, the $w^{HIS}$ molecule was exchanged with the bulk several times during the 100 ns simulation (Figure 3.2.29A, S3-S7).

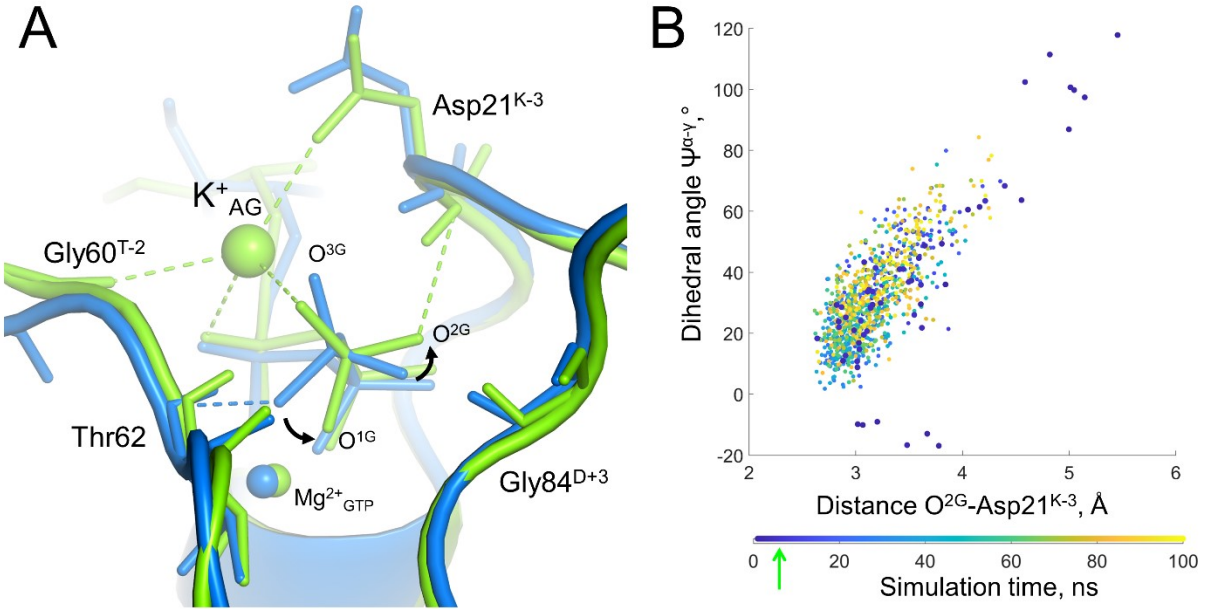


**Figure 3.2.28.** Rotation of γ-phosphate upon $K^+$ binding and formation of an H-bond between the $O^{2G}$ atom and backbone amino group of $Asp21^{K-3}$. A. Coordinational and hydrogen bonds in the active site of EF-Tu with $K^+$ bound, as sampled from simulation #3 (see Table 1). Superposition of two conformations is shown: in blue – conformation sampled before $K^+$ binding in the AG site, in green – conformation sampled shortly after. B. Conformational space of GTP during the MD simulation of EF-Tu. Scatter plot of the $\Psi^{\alpha-\gamma}$ dihedral angle ($O^{2A}$-$P^A$-$P^G$-$O^{3G}$) against the length of the H-bond between the $O^{2G}$ atom and the backbone nitrogen of $Asp21^{K-3}$ as sampled from the MD simulation #3; the color of each dots corresponds to the point of time during the simulation, see the scale below the X-axis. Larger blue dots correspond to conformations, sampled before $K^+$ binding in the AG site. The green arrow on the color scale indicates the moment of $K^+$ binding. Data from other simulations are shown in Supplementary Figure S8.

In addition to the $w^{HIS}$ molecule, three more water molecules, at least, appear to be involved in the H-bond network around the γ-phosphate of GTP. One of them is the $w^{ASP}$ molecule seen in the X-Ray structures (Figure 3.2.18B, D). In the MD simulations, $w^{ASP}$ molecule was located close to the $K^+_{AG}$ ion but stayed too far for a coordination bond. Instead, $w^{ASP}$ formed H-bonds with $O^{D2}$ atom of $Asp^{K-3}$ and the $O^{3G}$ atom of γ-phosphate, as also seen in the crystal structures (Figure 3.2.18B, D). The $w^{ASP}$ molecule is rarely exchanged with the solution (Figure 3.2.29B).

One more water molecule dwelled between the $w^{HIS}$ and $w^{ASP}$ molecules (Figure 3.2.27). Since this molecule appears to form a weak H-bond with the $O^{3G}$ atom of γ-phosphate, hereafter denoted as $w^G$. Since $w^G$ did not form H-bonds with amino acid residues, this water molecule was the least stable one and frequently exchanged with the bulk (Figure 3.2.29C, S3-S7). Finally, one of the water molecules, coordinating $Mg^{2+}_{SRL}$ ($w^{SRL}$ in Figure 3.2.27) made an H-bond with $w^{ASP}$. Interactions between $Mg^{2+}$ ions and coordinating water molecules were very stable; as long as

$Mg^{2+}_{SRL}$ stayed bound to the phosphate group of A2662, the $w^{SRL}$ molecule was also present providing an H-bond to the $w^{ASP}$ molecule and stabilizing the latter (Figure 3.2.29D, S3-S7).

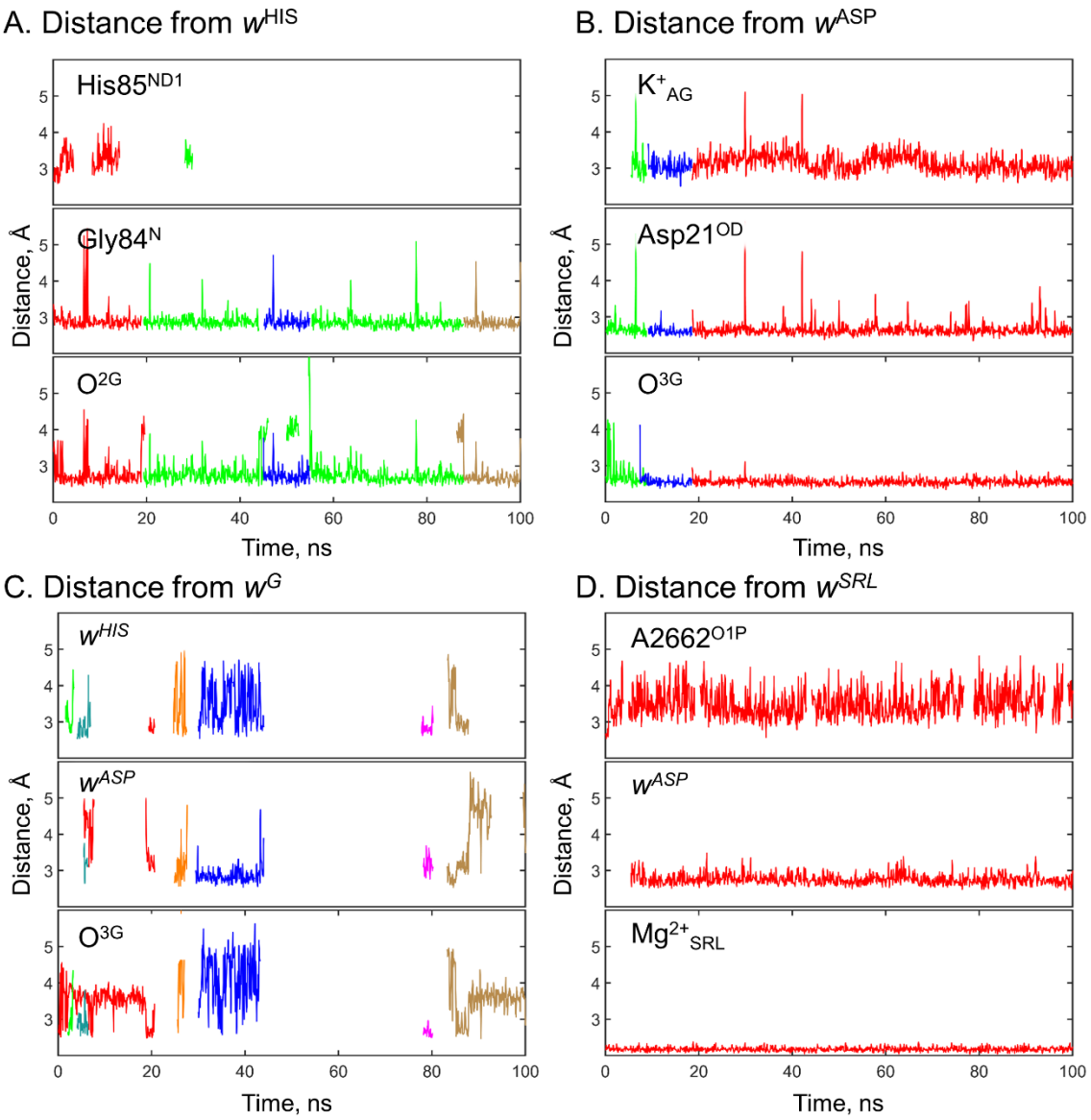


**Figure 3.2.29.** Stability of water molecules in the active site of EF-Tu as seen in MD simulations. Data from simulation #3 is shown, data from other simulations are shown in Figs. S3-S7. Water molecules nomenclature as shown in Figure 3.2.27. A. Distances measured from oxygen atom of water molecules in position $w^{HIS}$ to $N^{D1}$ atom of $His^{D+4}$, backbone nitrogen of $Gly^{D+3}$, and the $O^{2G}$ atom of GTP. B. Distances measured from the oxygen atom of water molecules in position $w^{ASP}$ to the $K^{+}_{AG}$ ion, the nearest $O^{D1}/O^{D2}$ oxygen atoms of $Asp^{K-3}$ sidechain, and the $O^{3G}$ atom of GTP. C. Distances measured from the oxygen atom of the $w^{G}$ water molecule to oxygen atoms of water molecules $w^{HIS}$ and $w^{ASP}$, and the $O^{3G}$ atom of GTP. D. Distances measured from the oxygen atom of $w^{G}$ water molecule to the $O^{1P}$ atom of A2662, the oxygen atom of the $w^{ASP}$ water molecule and the $Mg^{2+}_{SRL}$ ion. All distances were measured to oxygen atoms of water molecules; different colors correspond to different individual water molecules, within each panel same colors correspond to the same molecules. Data are shown only for the cases when a short distance (<5Å) was maintained for at least 1 ns.

The overall pattern of bonding in the catalytic site of EF-Tu resembled that described earlier for the FeoB GTPase [146]) and MnmE GTPase [28, 115], see Figure 3.2.30. Presented here data clarifies the role of the A2662 phosphate group of SRL in the activation of GTP hydrolysis by translation factors. While the $O^{1P}$ atom of this group, by making an H-bond with the $His85^{D+4}$ residue, kept it in the "catalytic" orientation, the $O^{2P}$ atom laid a $Mg^{2+}{}_{SRL}$-mediated bridge towards the $Asp21^{K-3}$ residue of EF-Tu and $K^+$ ion (Figure 3.2.20, Figure 3.2.27). Hence, SRL interacts with EF-Tu via two routes, at least, namely the H-bond with $His85^{D+4}$ and $Mg^{2+}{}_{SRL}$-mediated bridge to $Asp21^{K-3}$.

Earlier it was shown that the recognition of the cognate codon by aa-tRNA enables the interaction of the GTP binding site of EF-Tu with SRL and, in addition, affects the conformation of tRNA [127, 167, 170, 181, 182, 241-243]. Figure 3.2.19 shows that the position of the Switch I loop depends on the interaction of EF-Tu with its partners. It appears that the interaction with the aa-tRNA alone affects the conformation of the Switch I loop and pushes the $Glu60^{T-2}$ residue closer to the phosphate chain, towards the $K^+$-binding position (green structures Figure 3.2.19), which is in agreement with the earlier observation that aa-tRNA directly sends a signal of its acceptance during decoding to the GTP-binding site [241, 242]. Upon the codon recognition, the $Glu60^{T-2}$ residue comes on the coordination distance to the $K^+$ ion due to the Switch I interaction with SRL and tRNA (blue cartoons, Figure 3.2.19). Concurrently, the $His85^{D+4}$ and $Asp21^{K-3}$ attain their "catalytic" positions after the docking of the GTP-binding site to SRL. Hence, the $K^+$ ion "integrates" several conformational signals, which decreases the probability of erroneous stimulation of the GTP hydrolysis. Owing to the coordination by a $K^+$ ion, the codon recognition is precisely synchronized with the GTP hydrolysis.

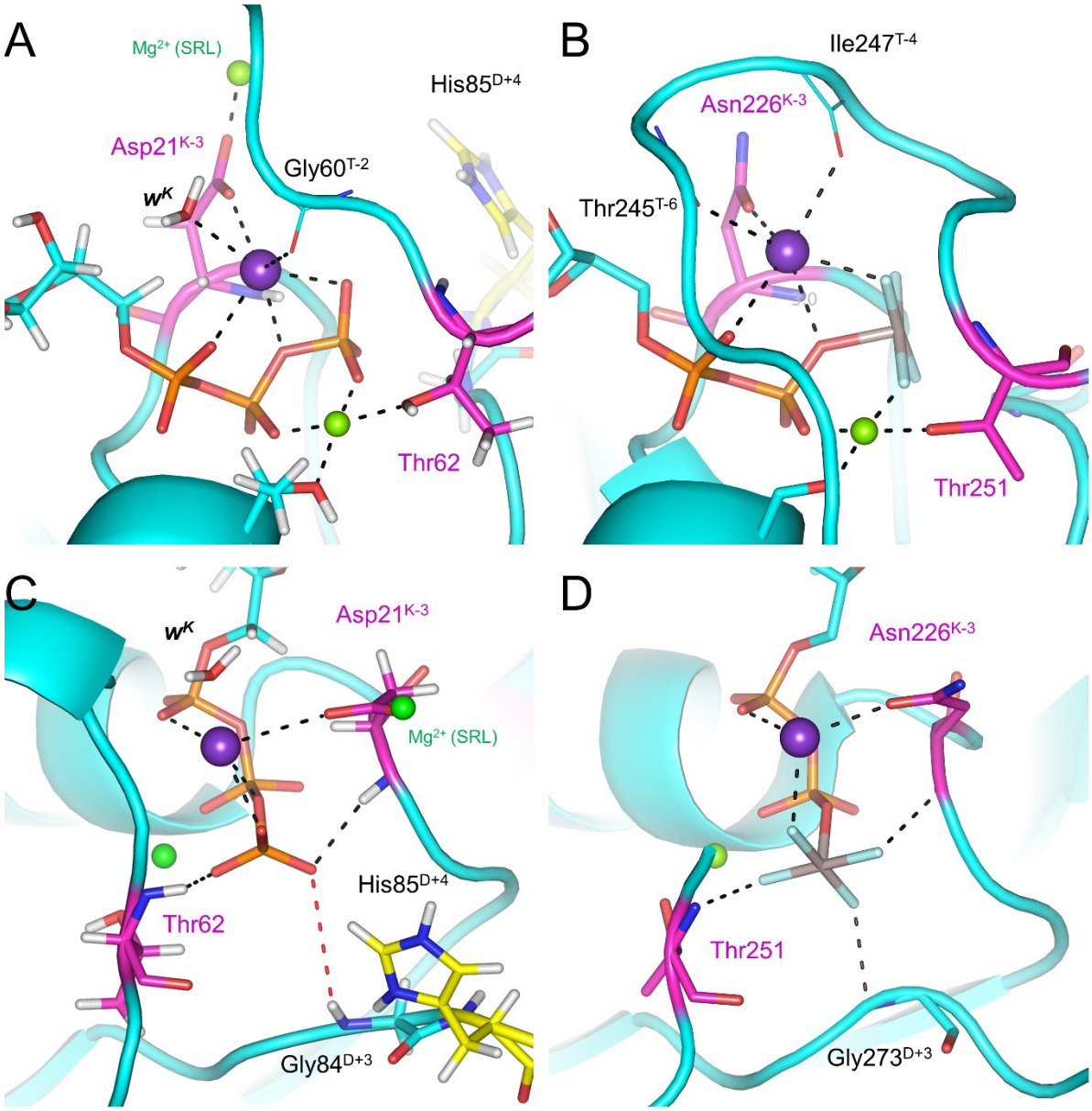


**Figure 3.2.30.** Comparison of the H-bond networks in the GTP-binding sites of EF-Tu and MnmE. A, C - $K^+$-binding site (A) and hydrogen bonds around γ-phosphate (C) in EF-Tu, as sampled from simulation #3 (see Table 2.3). B, D – $K^+$-binding site (B) and hydrogen bonds around the γ-phosphate mimicking $AlF_4^-$ group, as seen in the crystal structure of MnmE (PDB ID 2GJ8). Proteins are shown as blue cartoons, functionally relevant residues are shown as sticks, K-3 residues and conserved Thr of the Switch I are shown in magenta, catalytic $His85^{D+4}$ – in yellow. Cations are shown as spheres, $K^+$ in purple, $Mg^{2+}$ in green.

### *3.2.6. Comparative Analysis of P-loop NTPases*

Structural analysis and molecular dynamics simulations of MnmE and EF-Tu revealed significant similarities in the activation mechanism of these enzymes. Both proteins follow the same steps, that lead to the nucleophilic attack by the catalytic water molecule:

i. Mg-NTP binding to the P-loop motif drives the triphosphate chain into an extended conformation, with β- and γ-phosphates in a near-eclipsed conformation;
ii. Cationic moiety (monovalent cation in MnmE and EF-Tu) is inserted into the catalytic site and interacts with the γ-phosphate, which is rotated towards an almost eclipsed conformation;
iii. The new position of γ-phosphate is stabilized by an additional H-bond with the backbone amino group of the K-3 amino acid of the P-loop motif.

Potassium-dependent NTPases appear to be evolutionary old; most of them interact with ribosomes, arguably, the oldest cellular endowment [7, 8, 22, 175, 176, 244-249]. Modeled here proteins MnmE and EF-Tu both have been attributed to the Last Universal Common Ancestor [175, 250]. Hence, the catalytic mechanism, as suggested for evolutionary old, $K^+$-dependent P-loop NTPases could be envisioned as basic.

In this chapter available structures and activation mechanisms of P-loop NTPases of different classes are analyzed to identify ubiquitous features that underlie the common basic mechanism of NTP hydrolysis stimulation. First, the conservation of the P-loop motif structure was analyzed to illustrate that in all P-loop proteins the phosphate chain is bound in the same conformation. Second, representative proteins from each class of P-loop NTPases were analyzed in the context of available literature data, to determine the stimulating moiety, analogous to the monovalent cations in MnmE and EF-Tu, as well as residues, characterized as “catalytic” or “essential” and other residues involved in the coordination of γ-phosphate and/or catalytic water molecule. Finally, all available structures of P-loop NTPases were analyzed to investigate the possibility of the new H-bond between the $O^{2G}$ oxygen atom and the backbone amino group of the K-3 residue of the P-loop motif.

The abundance and diversity of the P-loop fold domains in NTPases with solved 3D structures make their structural analysis a complex and arduous task. In the SCOP database [251], the P-loop fold entry contains 24 families. In the CATH database, the P-loop superfamily 3.40.50.300 includes 140 structural clusters and 2,438 functional families [252]. In the InterPro database [218], the “homologous superfamily” IPR027417 has 698 family and domain entries. In

the ECOD database [253], the topology-level "P-loop_NTPase" entry contains 193 families. In the Pfam database [254], the P-loop NTPase clan CL0023 contains 217 families. To both provide detailed analysis and cover all available structures, in this work manual inspection of the selected representatives from diverse classes of P-loop NTPases was combined with systematic computational analysis of all available structures of P-loop fold NTPases with bound NTP-like molecules.

For each class of P-loop NTPases, one or several proteins with available crystal structures were selected as representatives. Each selected structure contained a $Mg^{2+}$ cation and either a metal fluoride transition state analog (when available) or a non-hydrolyzable substrate analog. To evaluate the conservation of the P-loop motif structure, all chosen representative structures were superposed with the structure of $AlF_4^-$-containing, $K^+$-dependent GTPase MnmE (PDB ID 2GJ8) by aligning 20 amino acids of the P-loop region with the corresponding amino acids 217-236 of MnmE. Not only the sequence motif but also the shape of the P-loop is strictly conserved across all classes of P-loop NTPases (Figure 3.2.31A). Accordingly, the following interactions of the triphosphate chain with amino acids and the $Mg^{2+}$ ion appear to be conserved across the whole superfamily (Table 3.8).

In all manually inspected structures, the NTP molecule (or its analog) is bound to the P-loop domain in catalytically prone extended conformation that is characterized by the eclipsed orientation of the β and γ phosphates as seen, for instance, in Figure 3.1.2A, B, see also [115, 216, 255, 256]. This strained, extended conformation of the NTP molecule is enforced by a plethora of highly conserved bonds that mostly involve the amino acids of the P-loop motif [115, 121], see Figure 3.2.31 and Figure 3.2.24. Specifically, the $O^{1A}$ atom of α-phosphate forms an H-bond with the backbone amino group of the K+2 amino acid, while the $O^{3A}$ atom between α- and β-phosphates forms an H-bond with the backbone amino group of $Gly^{K+1}$, another invariant amino acid of the P-loop motif. The $O^{3B}$ atom between β- and γ-phosphates forms an H-bond with the backbone amino group of the K-3 amino acid. The $O^{1B}$ and $O^{3G}$ atoms, respectively, form H-bonds with the signature Lys residue of the P-loop motif, while $O^{2B}$ and $O^{1G}$ atoms coordinate $Mg^{2+}$ ion, which is, in its turn, is coordinated by protein residues. All these interactions are common for all P-loop fold NTPases, as they involve universally conserved residues or backbone amino groups of the P-loop. Most of these interactions involve backbone nitrogens of the P-loop motif residues, namely K-3, K-1, K, K+1, and K+2 (Figure 3.2.31B). The consistent presence of H-bonds formed between the backbone groups of K-3, K-1, K, and K+1 residues and α- and β-phosphate oxygen atoms was noted earlier for ATPase motor proteins [140]. Three of these residues (K-1, K, and

K+1) are universally conserved within the P-loop motif GxxxxGK[S/T], so these bonds are likely to be universally present in all P-loop/NTP complexes (Figure 3.2.24).

While the K-3 position is not universally conserved among the P-loop proteins, it is conserved within many families and almost always is represented by a Gly or another small residue (Figure 3.2.32). The notable exceptions are cation-dependent proteins in the TRAFAC class, which, as mentioned earlier, carry a $K^+$-binding residue in this position: Asp in translation factors, and Asn in MnmE and others. It was suggested, that the backbone amid of K-3 residue plays a pivotal role in the catalysis by stabilizing the $O^{3B}$ atom of the substrate in Ras-like GTPases [257]. Conservation of not only the position of the backbone, but the residue itself in the K-3 position indicates, that this residue may play an even bigger role in the catalysis, specifically by stabilizing one of the oxygen atoms of γ-phosphate right before and during the transition state.

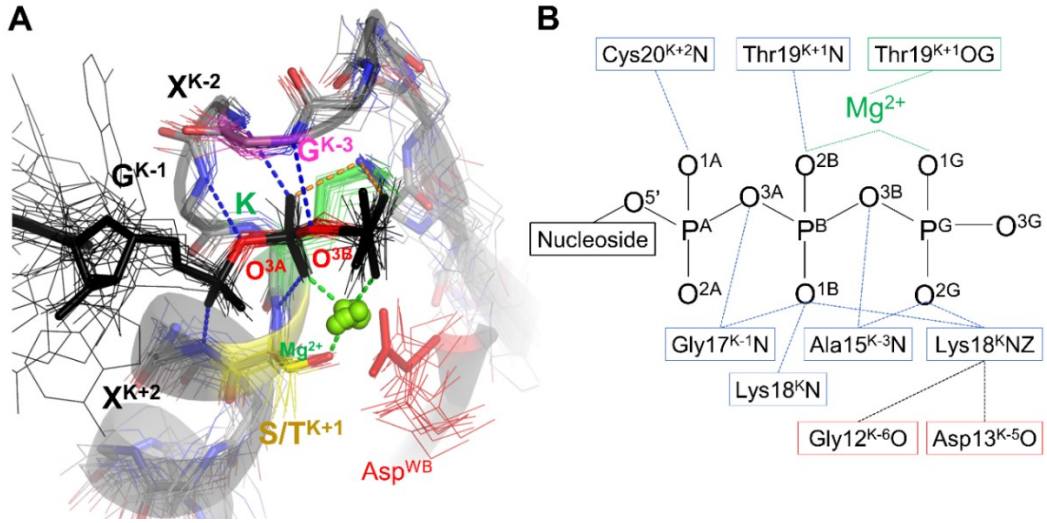


**Figure 3.2.31.** Conservation of the P-loop motif structure across protein families. **A**. Structures of proteins, representing different classes of P-loop NTPases and described in Table 3.9 are shown superimposed on the P-loop region of GTPase MnmE (PDB ID 2GJ8 [28]). For the MnmE GTPase, the P-loop region is shown as a cartoon. For all the proteins, the last seven residues from the motif GxxxxGK[S/T]x, as well as the Asp residue of the Walker B motif and NTP analogs are shown in lines and colored as follows: $Lys^{WA}$ in green, $Gly/Ala/Asn^{K-3}$ in purple, $Asp^{WB}$ in red; NTP analogs are shown in black with ester oxygen atoms in red. Other protein residues are shown in gray with backbone nitrogen and oxygen atoms shown in blue and red, respectively. $Mg^{2+}$ ions are shown as yellow-green spheres. **B**. Coordination and H-bonds between the Walker A motif residues and Mg-NTP moiety. Specific residues and numbers are given for the RhoA GTPase (PDB ID 1OW3 [125]).

Additionally, an H-bond is formed between the backbone of K+2 residue and the $O^{1A}$ atom of α-phosphate (Figure 3.2.24). While this residue is not conserved across all P-loop proteins, the location of its backbone nitrogen group in the structure is invariable, so this bond always can be formed. Finally, the sidechain amino group of the signature lysine residue is anchored by two

backbone oxygen atoms of residues K-5 and K-6. In crystal structures, the distance between P-loop Lys and the backbone oxygen of K-6 does not always support the H-bond formation (Table 3.8). Still, the formation of both bonds with K-5 and K-6 is possible in all P-loop proteins.

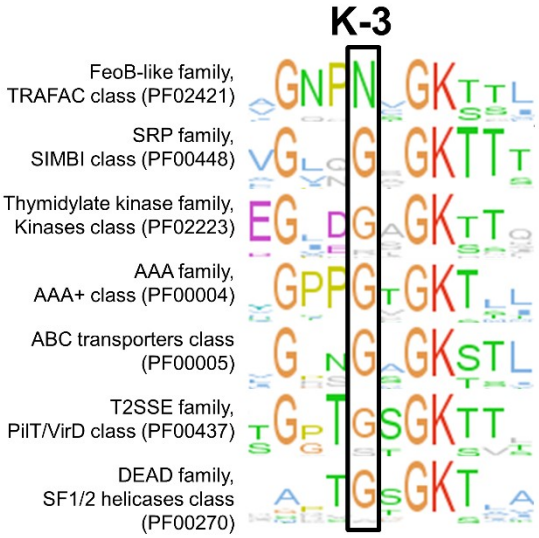


**Figure 3.2.32.** Conservation of the K-3 residue in the P-loop motif of some protein families. Sequence logo representations were obtained using JalView [258] for the "seed" alignments of individual protein families from the Pfam database [254] and superposed to match the conserved lysine residue. The black box indicates the position corresponding to the K-3 residue.

### *3.2.6.1 Activation mechanisms in representative P-loop NTPases*

For the comparative structure analysis, representative structures were manually selected for each class of P-loop proteins based on literature data and structure availability. When available, structures with transition state analogs, e.g. NDP:$AlF_4^-$ or NDP:$MgF_3^-$ complexes, were analyzed. For classes that did not have such structures available, structures with non-hydrolyzable analogs of ATP or GTP were used. In total 21 structures have been selected to represent diverse classes of P-loop NTPases (Table 3.9).

For each class one or several representative structures were checked for the following features: (i) amino acids that coordinate the Mg-triphosphate moiety; (ii) positively charged, potentially stimulating moiety(ies) that are introduced into the catalytic site during the activation and interact with the phosphate chain; (iii) amino acid that interacts with the catalytic water molecule; and (iv) other auxiliary amino acids that interact with oxygen atoms of the γ-phosphate group (or its analogs).

The generic numbering scheme for residues in conserved motifs introduced in Chapter 3.1.1 is used here as well to refer to functionally relevant residues across diverse classes of P-loop

proteins. Additionally, a generic numbering is introduced and employed here for the strands in the β-pleated sheet. In some P-loop NTPases, the $Mg^{2+}$-stabilizing Walker B motif aspartate residue is located at the C-terminal tip (hereafter C-cap) of the β-strand. However, in other NTPases, the Walker B motif could be at the C-cap or even at the N-terminus of other β-strands [259]. This diversity results in a rather confusing numbering of β-strands that is used for major classes of P-loop NTPases. Otherwise in this work, for simplicity, the β-strand, which is in-line with the triphosphate chain is referred to as "Walker B strand" or WB-strand. This strand is easy to find in a structure because it stabilizes $Mg^{2+}$ by an Asp or, rarely, a Glu residue as described in [259]. Other strands of the same β-pleated sheet, independently of their position in the amino acid sequence, are numbered by their position relatively to WB-strand as WB-1, WB-2, etc., or WB+1, WB+2, etc., as shown in Fig. 1A. For the sake of brevity, "WB" could be omitted, and the strands could be numbered -1, -2, 0 (for the WB strand), +1, +2, and so on, as depicted in Figure 3.2.33 and Figure 3.2.36.

### *Kinase-GTPase division*

According to [24, 26], P-loop NTPases form two divisions: Kinase-GTPase division and ASCE (Additional Strand Catalytic E (glutamate)) division [24, 26]. The Kinase-GTPase division unites three classes of NTPases: the TRAFAC (from *tra*nslation *fac*tors) class of translational factors and regulatory NTPases, SIMIBI (*si*gnal recognition, *M*inD, and *Bi*oD) class of regulatory dimerizing ATPases and GTPases, and the class of nucleotide kinases (Figure 3.2.33).

In the **NTPases of the TRAFAC class,** the $\alpha_1$-helix is followed by an elongated Switch I loop that contains a single conserved Thr residue and is specific to the class (Figure 3.1.2, Figure 3.2.33). This elongated loop goes into a β-strand, which is antiparallel to all other β-strands in the core β-pleated sheet of the P-loop domain (Figure 3.2.33). Apart from the TRAFAC class, all other classes of P-loop NTPases have predominantly all-parallel core β-pleated sheets. In the most well studied Ras-like GTPases, Switch I interacts with diverse physiological modulators of activity [260], whereas the activating domains bind to the Switch II loop, which follows the Walker B motif (Figure 3.1.2A). In most TRAFAC NTPases, the HN group of the signature Thr of Switch I motif forms an H-bond with γ-phosphate (Figure 3.1.2A, B, Figure 3.2.34). The $W_{cat}$ molecule, when seen next to the transition-state analogs in crystal structures, is usually stabilized by the backbone oxygen atom (CO) of the same Thr residue whereas its side chain oxygen atom interacts with the $Mg^{2+}$ ion (Figure 3.1.2A, B, Figure 3.2.34). In addition, the D+1 and D+4 residues are usually involved in the stabilization of $W_{cat}$, see e.g. Figure 3.2.34.

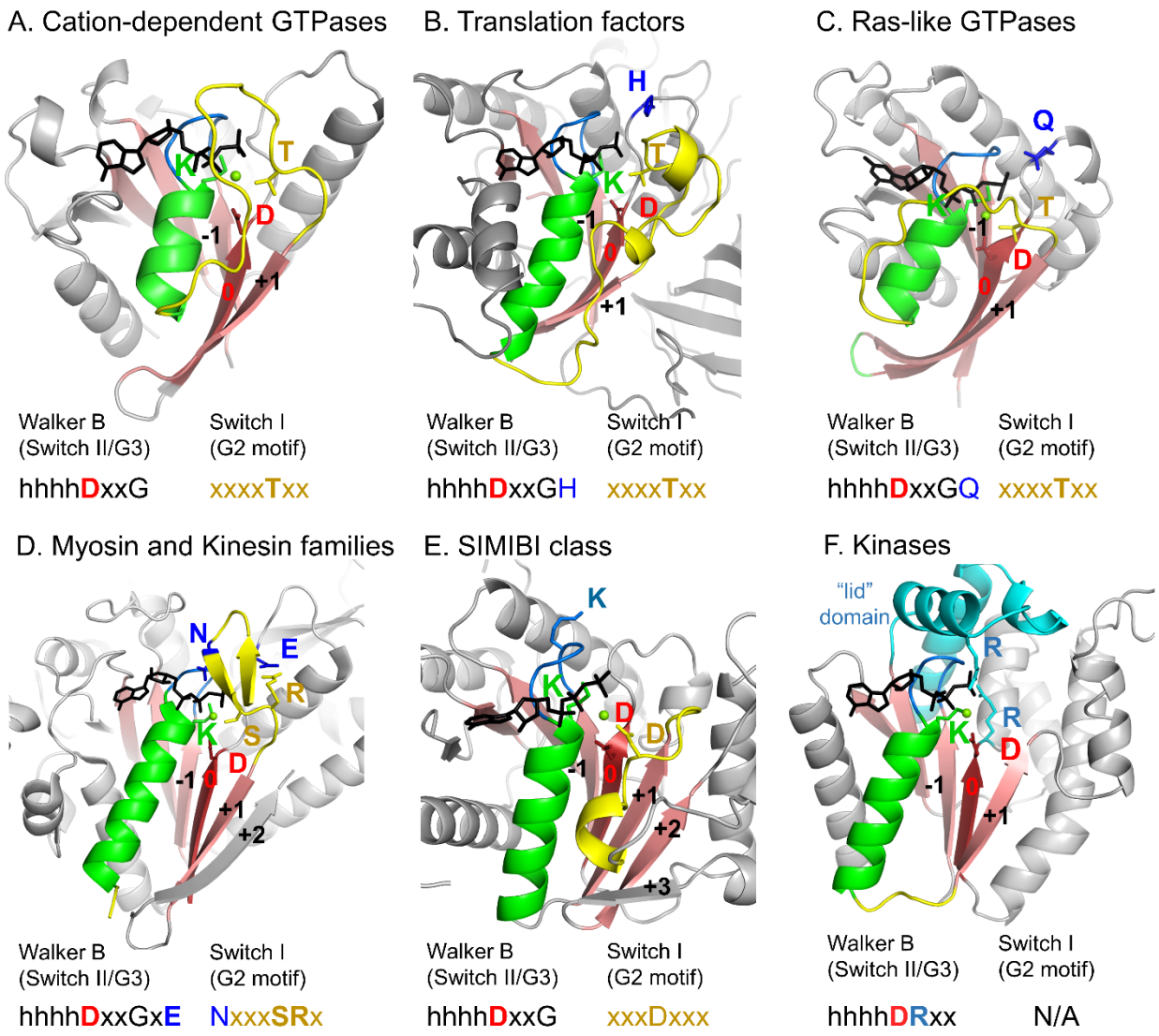


**Figure 3.2.33.** P-loop domain organization and conserved motifs of P-loop proteins in Kinase-GTPase division. Proteins are shown as gray cartoons, strands forming the core of the P-loop domains are numbered and colored pink, P-loop is shown in blue, the following α-helix is shown in green, K-loop/Switch II and corresponding regions are shown in yellow. GTP analogs are shown as black sticks, $Mg^{2+}$ ions – as green spheres. Strands are numbers relative to the Walker B motif-containing strand.

Previous chapters have already covered the activation mechanisms in cation-dependent GTPases of the TRAFAC class. Here, it is important to note the differences in the coordination of the $W_{cat}$. Most cation-dependent GTPases, including MnmE and dynamin, belong to the family of HAS (hydrophobic aminoacid substitution) GTPases [147]. The $W_{cat}$ molecule is stabilized not by "catalytic" side chain(s), but only by nearby atoms of protein backbone, e.g. $CO^{T}$, $HN^{T}$, and $HN^{D+2}$ in the case of the MnmE GTPase (Figure 3.1.2B). In dynamins (Figure 3.2.34A), these are $CO^{T}$ and $HN^{D+3}$. In the classical translational GTPases, $W_{cat}$ is uniquely stabilized by the side chain of $His^{D+4}$ residue in addition to NH of $Gly^{D+3}$ and CO of $Thr^{T}$, whereby the sidechain of His turns

towards $W_{cat}$ in response to the activating interaction of the Switch II with the small ribosomal subunit and tRNA [127].

Septins and septin-like proteins, which also belong to TRAFAC class, are stimulated by Arg fingers, reciprocally inserted upon dimerization of the NTPase domains as, supposedly, induced by interaction with the activating partner(s). In septins, the $NH_2$ group of the Arg finger of the Switch I loop enters the AG site and links the $O^{2A}$ and $O^{3G}$ atoms (see e.g. PDB ID 3FTQ [261]). Another example of TRAFAC proteins that utilize Arg fingers is GB1/RHD3-type GTPases (e.g. atlastin), where the Arg finger is in the K-3 position of the P-loop and links $O^{2A}$ and $O^{3G}$ atoms of GTP when two protein monomers dimerize in response to the interaction with an activating partner (PDB ID 4IDQ [262], Figure 3.2.34B). In both these cases, Arg fingers stimulate GTP hydrolysis in the very same P-loop domain they belong to, but the P-loop domain dimerization is required for the interaction of Arg residues with "their" phosphate chains. In atlastin, $W_{cat}$ is seen stabilized by $CO^{T}$ and $HN^{D+3}$ (Table 3.9, Figure 3.2.34B).

One of the best-studied groups in the TRAFAC class is the extended Ras-like (from *rat sarcoma*) protein family, which includes Ras-GTPases believed to be among the most powerful oncogenes. These small regulatory GTPases are usually stimulated by an Arg finger provided by a specific partner protein GAP (*GTPase activating protein*). Similar Ras-like proteins can have fully unrelated GAPs, see [121, 260, 263] for reviews. In most structures of Ras-like GTPases, the stimulating Arg finger is inserted between α- and γ-phosphates and links $O^{2A}$ and $O^{3G}$ atoms of GTP, see Figure 3.1.2A.

The specific feature of Ras GTPases is that the stimulating Arg residue, in addition, mobilizes a further catalytically important residue in the D+4 position. The backbone carbonyl oxygen of the Arg finger makes an H-bond with the side chain amino group of $Gln^{D+4}$, which brings this residue ($Gln61^{D+4}$ in Ras GTPase, $Gln63^{D+4}$ in Rho GTPase) in a position where its side chain carbonyl interacts with $W_{cat}$, see Figure 3.1.2A and [92, 117, 121, 122, 191, 192, 264].

The GTPase domains of the α-subunits of heterotrimeric G-proteins are closely related to Ras-like GTPases. Here, the intrinsic Arg finger, as provided by a family-specific insertion domain, links $O^{2A}$ and $O^{3G}$ atoms of GTP as well, and the side chain of $Gln^{D+4}$ residue and $HN^{D+3}$ are involved in $W_{cat}$ stabilization in addition to $CO^{T}$. Notably, the Arg finger is also directly involved in the interaction with the D+4 residue, providing a backbone CO group to orient the side chain of $Gln^{D+4}$ towards the γ-phosphate [130](PDB ID 1TAD, Figure 3.2.34C).

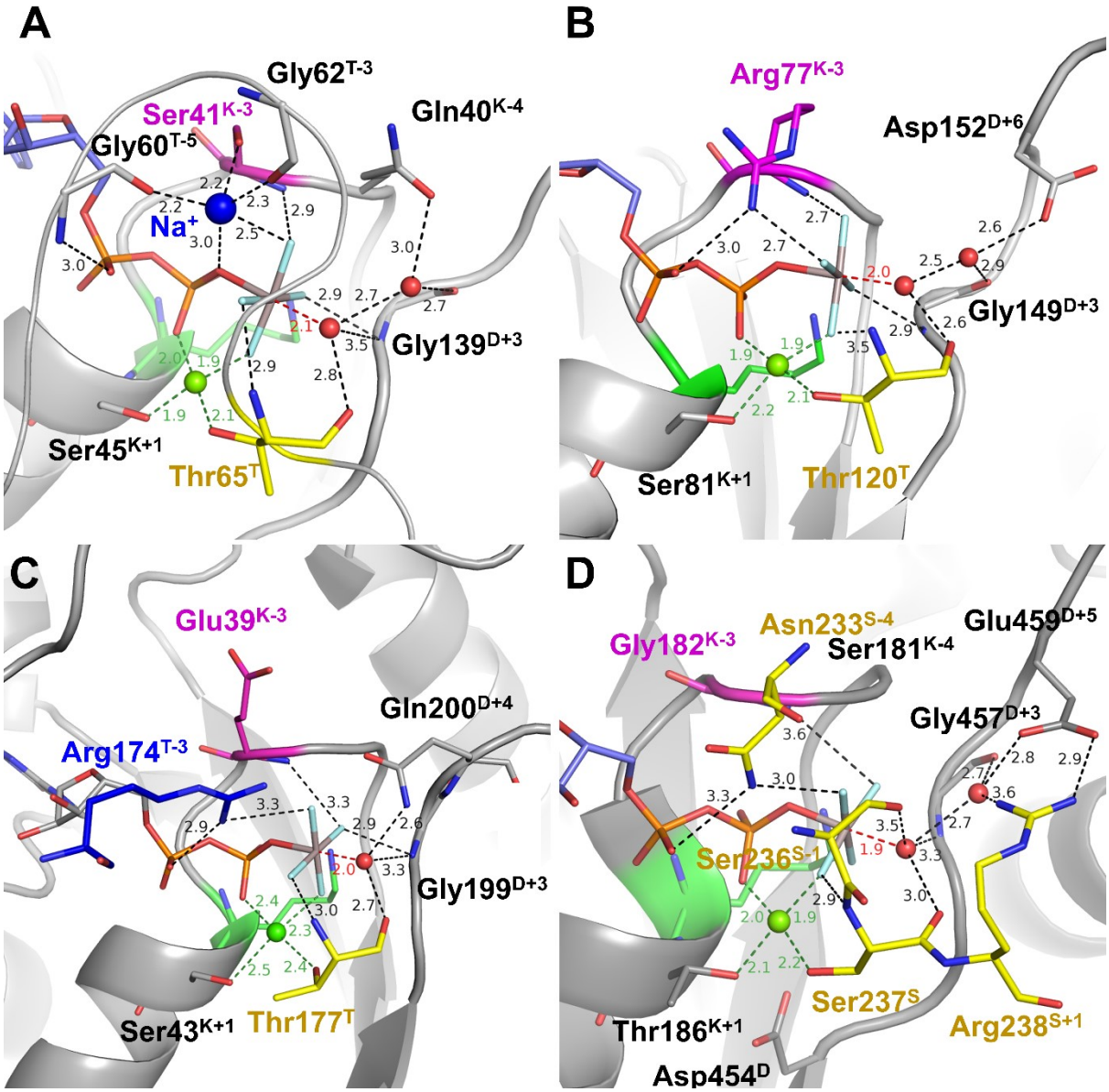


**Figure 3.2.34.** Representative NTPases of the TRAFAC class. The protein backbone is shown as a gray cartoon, nucleotides, their analogs, and functionally relevant residues are shown as sticks, water, and cations are shown as spheres: water in red, $Mg^{2+}$ in yellow-green, $Na^{+}$ in blue. P-loop lysines are shown in green, K-3 residues are shown in magenta, conserved residues from Switch I/K-loop are shown in yellow. All distances are measured in Å. A. Dynamin (PDB ID 2X2E [30]). B. Atlastin-1 (PDB ID 4IDQ [262]). C. $G\alpha_{12}$ subunit of heterotrimeric G-protein (PDB ID 1TAD [130]). D. Myosin II (PDB ID 1W9J, Morris C.A. et al., to be published).

Finally, a unique stimulation mechanism, that employs neither monovalent cations nor Arg fingers is realized in the kinesin and myosin families. In these proteins, the position of conserved Thr in Switch I motif is taken by the signature Ser residue, which provides its side chain oxygen atom to coordinate the cofactor $Mg^{2+}$ ion (Figure 3.2.34D). The $Asn^{S-4}$ residue gets inserted between α- and γ-phosphates [265-267], so that the side chain amino group of $Asn^{S-4}$ links the $O^{2A}$ and $O^{3G}$ atoms. Additional coordination of the γ-phosphate appears to be provided by the sidechain

of Ser/Thr$^{K-4}$ of the P-loop and Ser$^{S-1}$ of Switch I The CO$^{S}$ and, likely, the side chain of Ser$^{S-1}$ form H-bonds with $W_{cat}$. One more H-bond with $W_{cat}$ is provided by HN of Gly$^{D+3}$ (Figure 3.2.34D).

The **SIMIBI** class NTPases include ATPases and GTPases that dimerize upon interaction with the activating partner in a such way that the catalytic sites of monomers interact "face to face". Each monomer inserts, into the catalytic site of the other monomer, either a stimulating Lys residue (Figure 3.2.35A) or an Arg residue (in complexes of signal recognition particles (SRPs) with their respective receptors (SRs)), see Figure 3.2.35B and [268, 269] for reviews. Many SIMIBI class ATPases and GTPases contain so-called "deviant" Walker A motifs KGGxGK[S/T] with an additional conserved Lys$^{K-5}$ residue that is inserted into the active site of the partner subunit in a dimer [269]. Lysine fingers, as used by SIMIBI proteins, form H-bonds with both O$^{2A}$ and O$^{3G}$ atoms (Figure 3.2.35A). Many SIMIBI proteins have a Gly$^{D+3}$ residue [24], which can provide its backbone amide for additional coordination of the γ-phosphate, similarly to Gly$^{D+3}$ in TRAFAC class proteins (Figure 3.2.35A). Additional coordination of γ-phosphate is provided by amino acids located outside of conserved motifs. Such residues are protein family-specific and are often introduced into the active site from the adjacent monomer upon the interaction with an activating partner and dimerization.

Signal recognition particle (SRP) and its receptor (SR) stand separately within the SIMIBI class. Their GTPase domains form a pseudo-homodimer, where the two GTP binding sites interact face-to-face, but GTP hydrolysis occurs in only one of the two monomers. Instead of a Lys finger, these proteins employ Arg residues that are inserted reciprocally between the GTP binding sites so that the guanidinium group interacts with the α- and γ-phosphate of the GTP molecule bound by "its" subunit and the α-phosphate of the GTP molecule bound by another subunit (Figure 3.2.35B). Still, the formation of such a GTPase dimer alone is not enough to trigger the GTP hydrolysis. It was suggested that only upon he interaction with the RNA tetraloop, the "catalytic" Glu residue of SRP (Glu277 in Figure 3.2.35B) is rotated towards the γ-phosphate group of the SR-bound GTP and contributes to the stabilization of $W_{cat}$ so that hydrolysis can occur, see Figure 3.2.35B and [270]. This Glu residue is specific to the SRP family and is located outside of the common Walker A/Walker B motifs.

Many other SIMIBI class NTPases display similar activation pattern, where an NTP-bound (homo)dimer demands the interaction with an activating protein (or RNA in SRP/SR complexes) to bring specific "catalytic" residues in a position next to the γ-phosphate group where they can additionally contribute to the H-bond network around $W_{cat}$ [268].

The $W_{cat}$ molecule is stabilized by HN of the Gly$^{D+3}$ residue and residues of the SIMIBI-specific Switch I motif. Although SIMIBI NTPases lack the antiparallel β-strand of the TRAFAC

class, they have a loop that is located similarly to Switch I and is also called Switch I (Figure 3.2.33). Conserved Asp from this region is involved in the stabilization of $W_{cat}$ (see Table 3.9, $Asp66^{SwI}$ in Figure 3.2.35A, and $Asp136^{SwI}$ in Figure 3.2.35B).

**Nucleotide Kinases** are ubiquitous enzymes that usually transfer the γ-phosphoryl residue of ATP to a wide range of substrates, primarily nucleotides and other small molecules [26, 271]. The key role in the catalysis by P-loop kinases is played by Arg or Lys residue(s) located in the so-called “lid” domain, a small helical bundle domain that "covers" the catalytic site and often carries several positively charged residues [26, 271, 272], see Figure 3.2.35C, D, Figure 3.2.33. Thus, P-loop kinases do not need other proteins or domains to provide stimulating Arg/Lys fingers. Instead, the binding of their substrate molecules is enough to trigger the lid domain rearrangement that results in the introduction of Arg/Lys finger(s). Whereas Arg residues serve as stimulating moieties in most kinases (Figure 3.2.35C), a Lys residue appears to perform this function in adenylyl sulfate kinases (Figure 3.2.35D) [273].

Arg finger of the lid domain binds in the AG site between the α- and γ-phosphates, forming H-bonds with $O^{2A}$ and $O^{3G}$ atoms. In those cases, where several Arg residues are involved around the substrate-binding site, one of them occupies the AG site whereas others interact with its γ-phosphate and the second substrate molecule (e.g., in adenylate kinase, Figure 3.2.35C). In the case of the adenosine-5'-phosphosulfate (APS) kinase, the lid domain inserts a Lys finger that interacts directly with the γ-phosphate and is connected with α-phosphate via a water molecule (Figure 3.2.35D).

In kinases, the hydrolysis of the phosphoryl donor molecule is mediated by the acceptor molecule, which, similarly to $W_{cat}$ in other reactions of NTP hydrolysis, initiates the nucleophilic attack and formation of the pentavalent intermediate. As in other P-loop NTPases, this reaction is assisted by arginine and lysine fingers that position the interacting molecules and neutralize the negative charges of phosphate groups [123, 274]. Some families of kinases have an Arg residue in positions D+1 - D+5 [26]. Such Arg residue may assist in binding the phosphate acceptor molecule (Figure 3.2.35C).

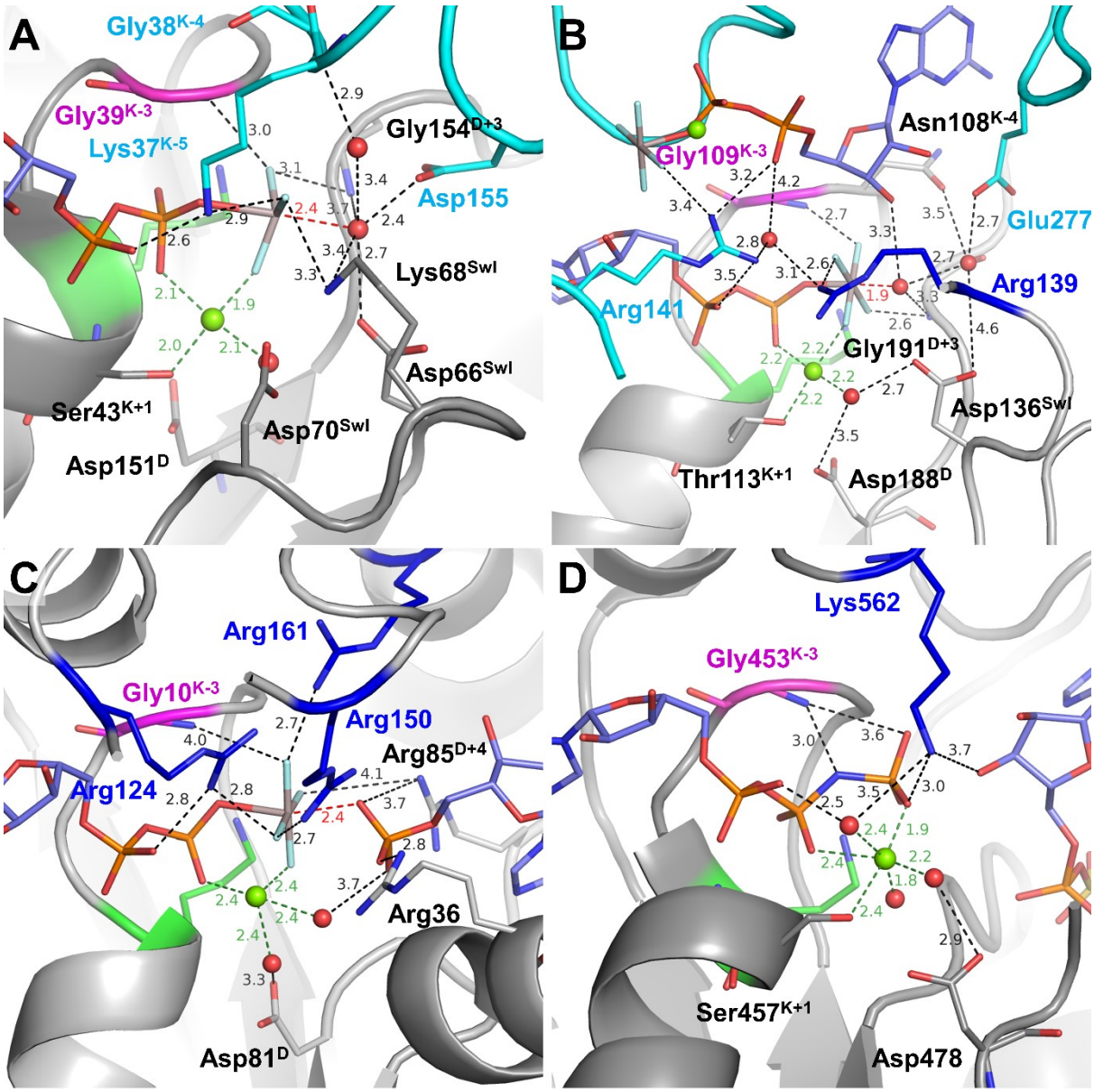


**Figure 3.2.35.** Representatives of the SIMIBI and kinase classes. The residues of the adjacent monomers in dimers are shown in light blue. Arg and Lys residues of the lid domains in kinases are shown in blue. Other colors as in Figure 3.2.34. All distances are measured in Å. A. ATP-binding component of the dark-operative protochlorophyllide reductase (PDB ID 2YNM [275]). B. Signal recognition particle (FtsY/Ffh) complex (PDB ID 4C7O [276]). C. Adenylate kinase (PDB ID 3SR0 [274]). D. Adenosine 5'-phosphosulfate kinase (PDB ID 4BZX [277]).

### *ASCE division*

The NTPases of the ASCE division have all-parallel β-pleated sheets in the core of the P-loop domain with an additional, as compared to the NTPases of the Kinase-GTPase division, β-strand in the WB+1 position (Figure 3.2.36). In many classes of this division, further inserted β-strands were identified (Figure 3.2.36), see also [278-280]. Typically, ASCE NTPases use a catalytic glutamate residue on the C-cap either of the Walker B strand or on the C-cap of the WB+1 strand to stabilize $W_{cat}$, in addition to protein backbone groups. These features define the name of this division: *additional strand, catalytic E* (ASCE) [24, 281-283]. According to current views [278-280], the ASCE NTPases are divided into two clades that roughly differ by the number of β-strands in their P-loop domains (Figure 3.2.36). The group of "middle-size" domains with up to five-six β-strands includes AAA+ ATPases, helicases of superfamily 3 (SF3), as well as STAND and KAP ATPases. The clade of "large" ATPase domains, with many β-strands, includes classes of VirD/PilT-like ATPases, FtsK-HerA-like ATPases, helicases of superfamilies 1 and 2 (SF1/2), ABC-ATPases, and RecA/F1 ATPases.

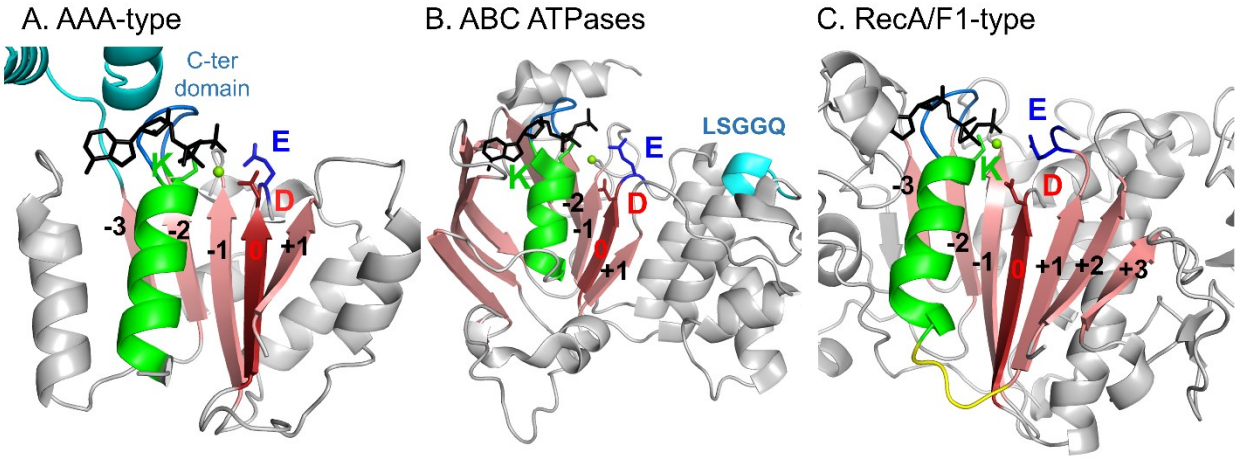


**Figure 3.2.36.** P-loop domain organization and conserved motifs of P-loop proteins in ASCE division. Proteins are shown as gray cartoons, strands forming the core of the P-loop domains are numbered and colored pink, P-loop is shown in blue, the following α-helix is shown in green, K-loop/Switch II and corresponding regions are shown in yellow. GTP analogs are shown as black sticks, $Mg^{2+}$ ions – as green spheres. Strands are numbers relative to the Walker B motif-containing strand, see the main text.

**AAA+ ATPases** are *ATPases associated with various cellular activities*, where "+" stands for 'extended'. These enzymes contain an N-terminal P-loop domain and an additional α-helical C-terminal domain, see [119, 120, 281, 284] for comprehensive reviews. The P-loop domain of the AAA+ ATPases carries conserved Arg residue(s) from the side that is opposite to the P-loop

[285]. P-loop domains of AAA+ proteins interact upon oligomerization (most often a hexamer is formed [120]), so that the nucleotide-binding site of one subunit receives Arg finger(s) from the neighboring subunit and, in some protein families, an additional Arg residue from its own C-terminal helical domain ( [120], Figure 3.2.37A). One of the stimulating residues occupies the space between α- and γ-phosphates (Figure 3.2.37A). This stimulation pattern is also typical for superfamily 6 helicases (SF6 helicases) [120]. No structures with transition state analogs are available for AAA+ ATPases; therefore, the coordination of $W_{cat}$ remains unclear; but $Glu^{D+1}$ residue is believed to be involved [120].

**Helicases of superfamily 3 (SF3 helicases)** are very similar to AAA+ proteins but compose a separate group/class owing to the different topology of their C-terminal helical domain [284]. While the overall structural organization and activation mechanism of SF3 helicases follow those of AAA+ proteins, these NTPases, instead of one or two Arg residues, insert one Lys and one Arg residue into the active site, see Figure 3.2.37B and [284]. Both these residues are provided by the adjacent monomer. The Lys residue forms H-bonds with both α- and γ-phosphates, whereas the Arg residue interacts with the γ-phosphate and also is in a position to stabilize $W_{cat}$ together with $Asp474^{D+1}$ and Asn529 of the WB-1 strand (Figure 3.2.37B).

Two more classes, **STAND** (*signal transduction ATPases with numerous domains)*) and **KAP** (named after Kidins220/ARMS and PifA) are likely to have activation mechanisms similar to the AAA+ proteins, based on their sequence similarities [283, 286]. However, the exact moieties involved in the hydrolysis stimulation in these proteins remain uncertain. Even though several structures of STAND ATPases are available and contain bound NTP analogs, all of them depict Apaf-1 (Apoptotic protease activating factor 1). This multi-domain ATP-binding protein does not exhibit any ATPase activity, and serves only as a signaling protein, undergoing large conformational changes upon binding to ATP and cytochrome *c* in the cell cytoplasm, which lead to heptamerization of Apaf-1/cytochrome c complexes and formation of the apoptosome, launching the caspase cascade of the apoptosis [286-290]. While the ATP-binding site of Apaf-1 highly resembles that of other ASCE proteins, it cannot serve as a representative for the entire STAND class. For the KAP class, no structures are available.

**SF1/2 class helicases** are mostly monomeric or dimeric with each polypeptide chain containing two P-loop domains [291]. In these proteins, the N-terminal domain carries the functional Walker A and B motifs and binds the nucleotide molecule, whereas the C-terminal domain, although having a P-loop-like fold, lacks the Walker A and B motifs and does not bind NTPs. Upon interaction with an RNA or a DNA, the C-terminal domain inserts two Arg residues into the nucleotide-binding site of the N-terminal domain [292]. One of these Arg residues forms

H-bonds with both α- and γ-phosphates, whereas the other Arg residue interacts with γ-phosphate (or its analog) and $W_{cat}$ (Figure 3.2.37C). In addition, $W_{cat}$ is stabilized by the catalytic $Glu342^{D+1}$ and class-specific Gln and Arg residues (Figure 3.2.37C).

**ABC (*ATP-binding cassette*) ATPases** are multidomain proteins that usually operate as homo- or heterodimers [293]. Members of the ABC class are divided into families named alphabetically from A to I [294]. Most of these families contain transmembrane domains and operate as genuine ATP-driven membrane transporters where the P-loop domain performs ATP hydrolysis. However, the members of ABCE and ABCF families have no transmembrane domain(s) [295].

In dimers of ABC-NTPases, the nucleotide-binding sites of P-loop domains are located on the interface between the monomers, in the same way as in dimers of SIMIBI NTPases, cf. Figure 3.2.35A, B. Instead of an Arg or Lys residue, each monomer inserts a whole signature motif LSGGQ into the catalytic pocket of the other monomer (Figure 3.2.37D, Figure 3.2.36B). In this motif, the last residue could be also Glu(E) or even Trp(W) [296]. Some soluble ABC-NTPases have a non-canonical signature motif (e.g. CSAGQ in Rad50 [297] and xSTFx) in MutS [298]). Thus, the serine residue appears to be the most conserved in the motif.

In the structure of the maltose transporter complex, the only structure available with a transition state analog bound (Figure 3.2.37D), the serine side chain and the backbone amino group of the second glycine residue interact with the $O^{3G}$ atom of γ-phosphate [299]. In this interaction, the serine side chain is located between the α- and γ-phosphates of the phosphate chain, approximately in the position of a $Na^+$ ion in dynamin-like proteins, cf. Figure 3.2.37D with Figure 3.2.34A.

Several amino acids commonly found in the active sites of ABC transporters can stabilize $W_{cat}$, see Figure 3.2.37D. These include the $Glu^{D+1}$ residue, and a histidine residue on the C-cap of the WB-1 strand in the case of the maltose transporter [299, 300]. Additionally, the activating monomer in a dimer contributes to the coordination of $W_{cat}$ by providing a backbone CO group of a residue that is located outside of the signature motif (Figure 3.2.37D).

Two more classes of the ASCE Division are likely to have a similar to AAA+ stimulation mechanism. In **VirB/PilT-like** class the class-specific PAS-like domain provides the Arg finger to the AG site of the bound NTP molecule of the same subunit; an additional Arg finger can interact with the γ-phosphate, see [301-304]. In **FtsK-HerA** pumping ATPases recent structures report an Arg residue analogous to the arginine finger of the AAA+ superfamily, that interacts with the NTP-binding site of the adjacent subunit [305]. However, in the absence of structures with transition-state analogs, the exact stimulation mechanisms in these two classes remain uncertain.

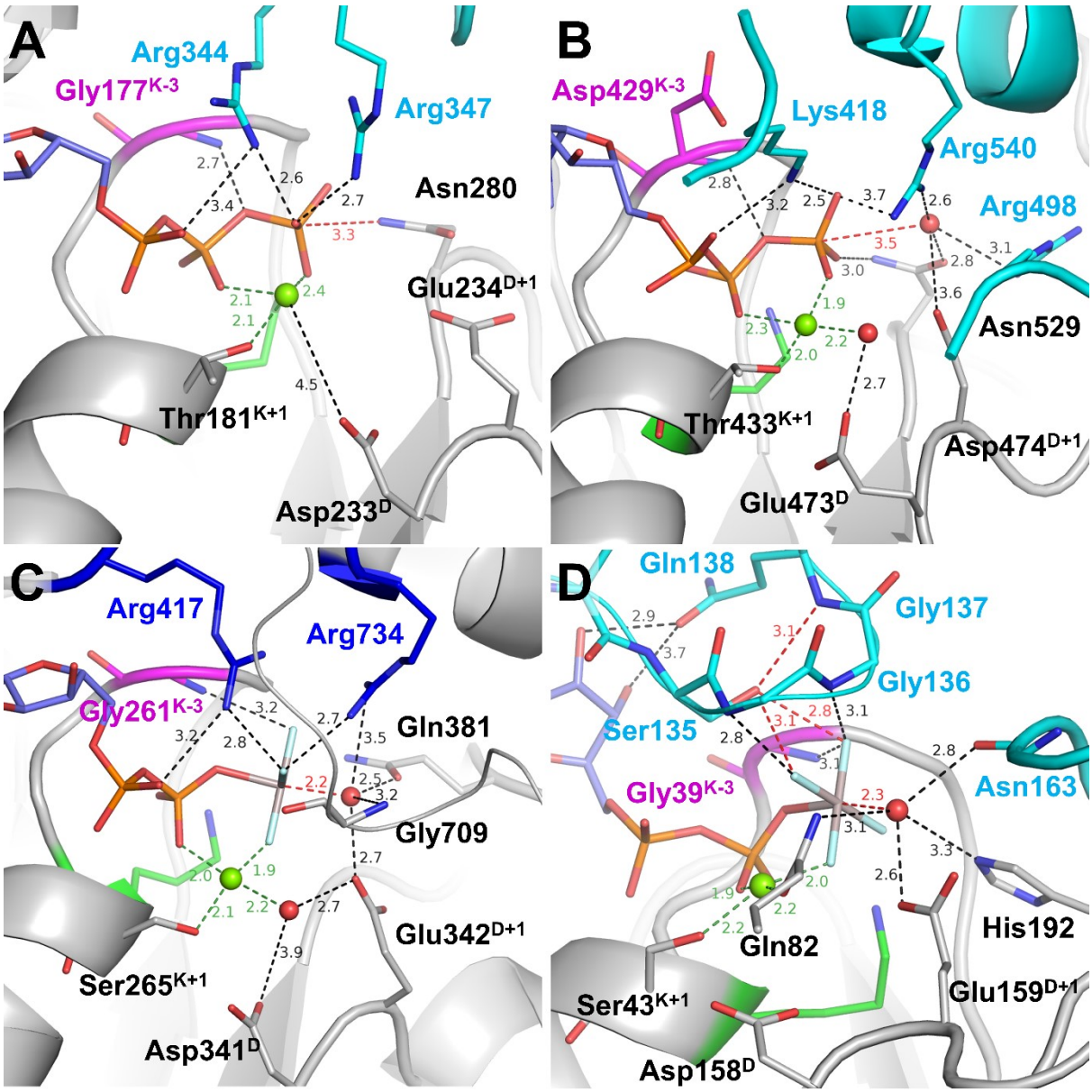


**Figure 3.2.37.** Representatives of AAA+ ATPases, SF3 helicases, SF1/SF2 helicases, and ABC ATPases. The protein backbone around the NTP-binding sites is shown as a gray cartoon, functionally relevant residues are shown as sticks, the $Mg^{2+}$ cations are shown as green spheres, $W_{cat}$ is shown as a red sphere. The signature lysines of the Walker A motif are shown in green, the K-3 residues are shown in magenta. The adjacent monomers and their Arg and Lys residues are shown in light blue. The Arg residues from the C-terminal helical domain are shown in deep blue (panels A, C). All distances are measured in Å. A. ATPase of human 26S proteasome (PDB ID 6MSB [306]). B. Replicative hexameric helicase of SV40 large tumor antigen (PDB ID 1SVM [307]). C. ATP-dependent DNA helicase PIF1 (PDB ID 5O6B [308]). D. Structure of an outward-facing maltose transporter complex (PDB ID 3PUW [299]).

**RecA/$F_1$ NTPases class** encompasses oligomeric ATP-dependent motors, involved in homologous recombination and DNA repair (RecA and RadA/Rad51), catalytic subunits of rotary F- and V-type ATP synthases, helicases of superfamilies 4 and 5 (SF4 and SF5), as well as several other protein families. In NTPases of this class, similarly to some AAA+ ATPases, the stimulating moiety(is) is/are inserted in the catalytic site by the P-loop domain of the adjacent monomer.

The stimulation mechanism appears to differ between rotary $F_1$-ATPases/SF5 helicases, on the one hand, and recombinases/SF4 helicases, on the other hand. In $F_1$-ATPases, the stimulating Arg residue of the adjacent monomer is inserted into the AG site and interacts with both α- and γ-phosphates, whereas an additional, intrinsic Arg residue coordinates the γ-phosphate, approaching it apically (see Figure 3.2.38A). A similar stimulation mechanism is realized in SF5 helicases, see e.g. Rho helicase (PDB ID 6DUQ [309]).

The $W_{cat}$ molecule is seen in the catalytic position only in the ADP:$AlF_4^-$-containing structure of the bovine F1-ATPase (PDB ID 1H8E, Figure 3.2.38A) [132]. While in most ASCE ATPases the $W_{cat}$-stabilizing $Glu^{D+1}$ residue directly follows the Asp of the Walker B motif, the $W_{cat}$ -stabilizing Glu188 is on the C-cap of the WB+1 strand. In addition, $W_{cat}$ is also stabilized by $Arg260^{D+4}$; this residue concurrently interacts with the activating neighboring monomer, which also provides a backbone CO to stabilize $W_{cat}$ (Figure 3.2.38A).

In recombinases/SF4 helicases, two stimulating positively charged moieties interact only with the γ-phosphate group, in contrast to F1-ATPases. In several families, the adjacent subunit provides one Lys residue and one Arg residue that form a short KxR motif [8], which interacts only with γ-phosphate. These are, for instance, the circadian clock protein KaiC (Figure 3.2.38B), bacterial helicase DnaB ([310] and PDB ID 4ESV), and gp4d helicase from bacteriophage T7 ([311] and PDB ID 1E0J). In bacterial RecA recombinases, the adjacent monomer in the homooligomer provides two Lys residues which approach the phosphate chain laterally from the adjacent subunit and interact only with γ-phosphate, see Figure 3.2.38C and [223]. In archaeal and eukaryotic RadA/Rad51-like recombinases, the positions of terminal groups of Lys/Arg are occupied by two $K^+$ ions, which appear to interact only with the γ-phosphate of ATP, see Figure 3.2.38D and [8, 233].

In the case of RecA-like ATPases, $AlF_4^-$-containing structures of the RecA recombinase of *E. coli* [223] and the bacterial DnaB helicase [310] are available. However, $W_{cat}$ is not seen in these structures and the position of $AlF_4^-$ moieties corresponds not to the transition state but rather to a post-transition state, as discussed elsewhere (manuscript in preparation). Therefore, the mechanism of $W_{cat}$ stabilization in RecA-like ATPases (Figure 3.2.38B-D) remains obscure; it potentially could involve polar residues at the C-caps of the WB-1, WB, and WB+1 strands, as

well as one or both of stimulating positively charged moieties [291, 312]. Generally, the H-bonded networks around γ-phosphate and $W_{cat}$ seem to be richer in RecA/F1-NTPases than in other classes of P-loop NTPases.

Overall, the activation mechanisms in all classes of P-loop NTPases involve the interaction between the P-loop domain and another domain, protein, or a subunit in the complex that introduces the stimulating moiety into the active site. The types of stimulating moieties display a remarkable variety, they could be monovalent cations; Arg, Lys, or even Asn residues; or groups of residues, like the characteristic LSGGQ motif of the ABC ATPases. All these moieties have cationic nature and are inserted in the active site in a similar manner by attacking the triphosphate chain from its exposed side, approximately corresponding to the $K^+$-binding AG side that was defined above. While most stimulating moieties interact with α- and γ-phosphates, in some families, the interaction involves only the γ-phosphate. Still, all types of stimulating moieties, described above have the potential to rotate/tilt the γ-phosphate.

Another prominent similarity between diverse classes of P-loop NTPases is the abundance of polar residues, coordinating the catalytic water molecule and additional, auxiliary residues, that provide backbone NH groups or positively charged sidechains to coordinate γ-phosphate. However, there is little to no conservation of these features across classes, and sometimes even within particular classes, these features are diverse.

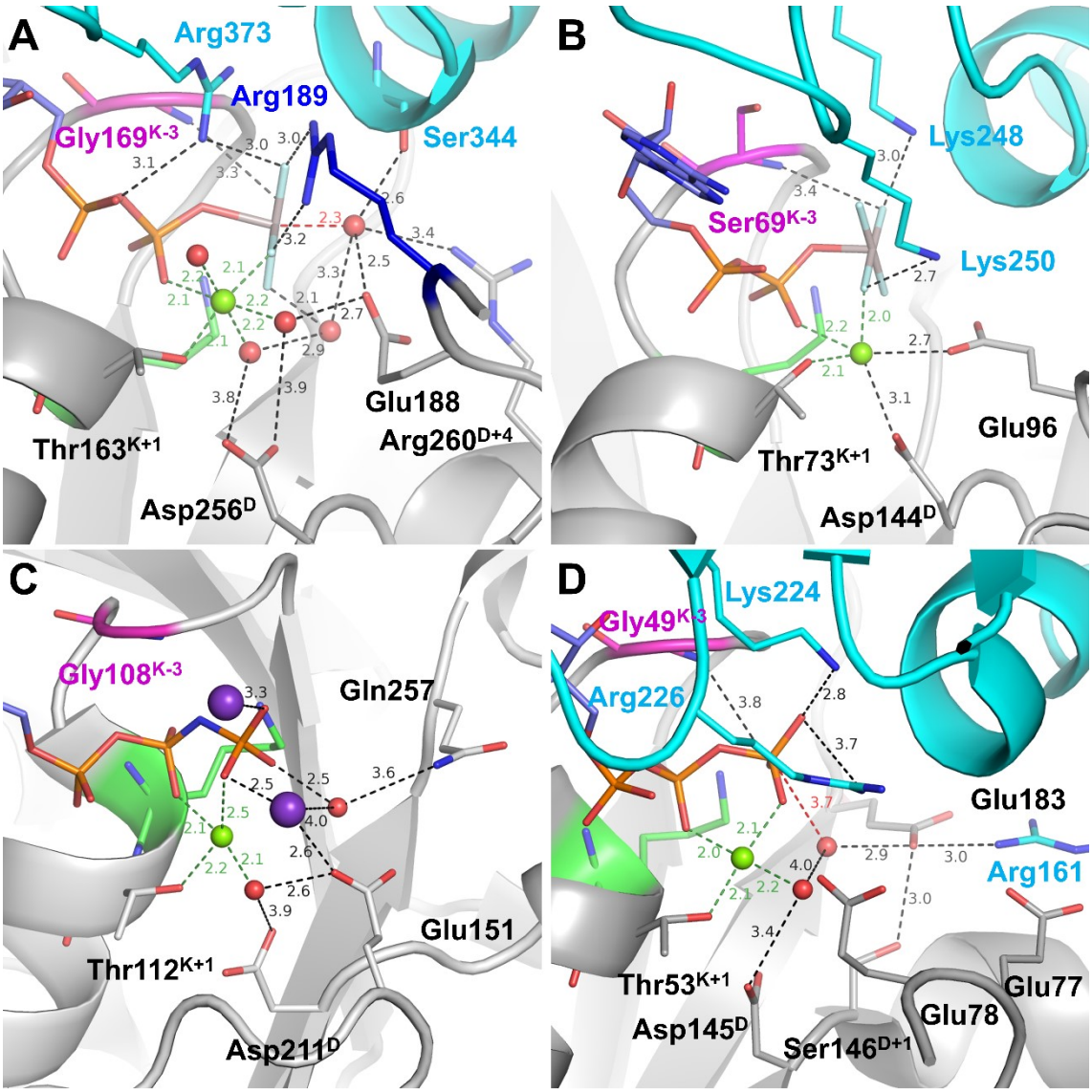


**Figure 3.2.38.** Representative proteins of the RecA/F1-like class of the P-loop NTPases.
The protein backbone around NTP-binding sites is shown as a cartoon, functionally relevant residues are shown as sticks, the $Mg^{2+}$ cations are shown as green spheres, $K^{+}$ ions are shown as purple spheres, water molecules are shown as red spheres. The signature lysines of the Walker A motif are shown in green, the K-3 residues are shown in magenta. The adjacent monomers and their Arg and Lys residues are shown in light blue. Arg residue from the P-loop domain proper is shown in deep blue (panel B). All distances are measured in Ångströms. A. Bovine F1-ATPase (PDB ID 1H8E [132]). B. RecA recombinase (PDB ID 3CMX [223]). C. RadA recombinase (PDB ID 3EW9 [160]). D. Circadian clock protein kinase KaiC (PDB ID 4TL6 [313]).

**Table 3.9.** Activation mechanisms in different classes of P-loop fold NTPases

| P-loop class, activation mechanism | Representative protein structure | PDB ID | NTP analog | Key catalytic moieties | | | |
|---|---|---|---|---|---|---|---|
| | | | | Walker A motif (K and K-3 residues) | Stimulating moiety in the AG site | | $W_{cat}$ stabilization[a] |
| **Kinase-GTPase division** | | | | | | | |
| **TRAFAC** Stabilization of the Switch I loop upon interaction with GAP, ribosome, or dimerization | MnmE | 2GJ8 | GDP:$AlF_4^-$ | Lys229 | Asn226 | ($K^+$) | Thr251$^{T}$-O, Gly249$^{T-2}$-HN, Thr250$^{T-1}$-HN, Gly273$^{D+3}$-HN, Gly273$^{D+3}$-O* |
| | Dynamin | 2X2E | GDP:$AlF_4^-$ | Lys44 | Ser41 | ($Na^+$) | Thr65$^{T}$-O, Gly139$^{D+3}$-HN, Gly139$^{D+3}$-O*, Gln40$^{K-4}$-OE1* |
| | Rho | 1OW3 | GDP:$MgF_3^-$ | Lys18 | Ala15 | Arg85†-NH2 | Thr37$^{T}$-O, Gln63$^{D+4}$-OE1 |
| | $G\alpha_{12}$ | 1TAD | GDP:$AlF_4^-$ | Lys42 | Glu39 | Arg174-NH1 | Thr177$^{T}$-O, Gly199$^{D+3}$-HN, Gln200$^{D+4}$-NE1 |
| | Atlastin | 4IDQ | GDP:$AlF_4^-$ | Lys80 | Arg77 | Arg77$^{K-3}$-NH2 | Gly149$^{D+3}$-HN, Thr120$^{T}$-O, Gly149$^{D+3}$-O*, Asp152$^{D+6}$-OD2* |
| | Myosin II | 1W9J | ADP:$AlF_4^-$ | Lys185 | Gly182 | Asn233-ND2 | Ser237$^{S}$-O, Ser236$^{S-1}$-OG, Gly457$^{D+3}$-HN, Gly457$^{D+3}$-O, Arg238$^{S+1}$*-NH1, Glu459$^{D+5}$*-OE1 |
| **SIMBI** In a homodimer, monomers provide stimulating Lys or Arg fingers for each other | GET3 | 2YNM | ADP:$AlF_4^-$ | Lys42 | Gly49 | Lys37†-NZ | Asp66$^{SwI}$-OD2, Asp155†-OD1, Lys68$^{SwI}$-NZ, Gly154$^{D+3}$-HN, Lys37†-HN* |
| | Signal recognition particle | 4C7O | GDP:$AlF_4^-$ | Lys112 | Gly109 | Arg141†-NH2 | Gly191$^{D+3}$-HN, Asp136$^{SwI}$-OD2, GTP†-O3', Asn108$^{K-4}$-OD1*, Gly191$^{D+3}$-O*,Glu277†-OE1* |
| **Kinases** Rearrangement of the lid domain upon substrate binding | Adenylate kinase | 3SR0 | ADP:$AlF_4^-$ | Lys13 | Gly10 | Arg124‡-NH1 | The second substrate (phosphate acceptor) is coordinated by Arg150‡-NH2, Arg85-NH1, Arg85-NH2, Arg36-NH1, Arg36-NH2 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Adenosine 5'-phosphosulfate kinase | 4BZX | ANP | Lys456 | Gly453 | Lys562‡-NZ | The second substrate (phosphate acceptor) is coordinated by Lys562‡-NZ |
| **ASCE division** | | | | | | | |
| **AAA+ / SF3 Helicases** In a hexamer, ATP binding/hydrolysis in one subunit causes conformational changes activating the following subunit; involves class-specific helical domain | 26S proteasome regulatory subunit 10B | 6MSB | ATP | Lys180 | Gly177 | Arg344†-NH1 | $Glu234^{D+1}$**, $Asn280^{SnI}$** |
| | SV40 large T antigen helicase (SF3) | 1SVM | ATP | Lys432 | Asp429 | Lys418†-NZ | $Asp474^{D+1}$-OD1, $Asn529^{SnI}$-OD1, Arg498†-NH1, Arg540†-NH2, |
| **Helicases SF1/2** Rearrangement of the C-terminal domain upon DNA or RNA binding | Multifunctional helicase Pif1p | 5O6B | ADP:$AlF_4^-$ | Lys264 | Gly261 | Arg417‡-NH2 | $Glu342^{D+1}$-OE1, $Gln381^{SnI}$-OE1, Arg737‡-NH1, Gly709‡-HN |
| **ABC ATPases** In a homodimer, monomers provide stimulating LSGGQ motifs for each other | Maltose transporter | 3PUW | ADP:$AlF_4^-$ | Lys42 | Gly39 | Gly136†-HN Gly137†-HN | Gln82-NE2, $Glu159^{D+1}$-OE1, $Glu159^{D+1}$-OE2, Asn163†-O, $His192^{SnI}$-NE2 |
| **$F_1$/RecA-like** In an oligomer, ATP binding/hydrolysis in one subunit causes conformational changes that activate the next subunit | RecA E. coli | 3CMX | ADP:$AlF_4^-$ | Lys72 | Ser69 | Lys248†-NZ | Glu96** |
| | RadA M. voltae | 3EW9 | ANP | Lys111 | Gly108 | $K^+$-503 | Glu151** |
| | $F_1$-ATPase | 1H8E | ADP:$AlF_4^-$ | Lys162 | Gly159 | Arg373†-NH1 | Glu188-OE1, $Arg260^{D+4}$-NH2, Ser344†-O |
| | KaiC | 4TL7 | ATP | Lys52 | Gly49 | Lys224†-NZ | $Glu183^{SnI}$-OE1 |

Residue numbers are as in the listed PDB ID structures; [a] – polar atoms, located within 3.6Å from the catalytic water molecule;
† - residues from polypeptide chains other than P-loop containing ones; ‡ - residues from domains other than the described P-loop domain;
* residue coordinates the attacking water molecule via another water molecule (e.g. Figure 3.2.35); ** catalytic water molecule not resolved, coordinating residue(s) were inferred from structure superposition and literature data; SwI – residues from Switch I region in SIMIBI, that lacks the conserved Thr; SnI – residues from Sensor I region in AAA+ and related proteins

### 3.2.7. Universal Activation Mechanism of P-loop NTPases

Comparative structure analysis and molecular dynamics simulations, reported in Chapters 3.2.2-3.2.5, revealed the details of the activation mechanism in cation-dependent P-loop NTPases. These findings served as a basis for the wider analysis of all P-loop NTPases which unveils the basic mechanism, universal for the entire superfamily. This mechanism is based on the ubiquity of certain features (i.e. their presence in all classes of P-loop NTPases) and certain features being conserved (i.e. not only present but represented by same or similar residues in all classes of P-loop NTPases).

#### 3.2.7.1 P-loop/Walker A and Walker B motifs: ubiquitous and conserved

The structural elements responsible for the stabilization of the triphosphate chain by the amino acids of the Walker A motif are almost universally conserved (Figure 3.2.31). The conservation of the invariant residues in the [G/A]xxxxGK[T/S] motif has straightforward reasons: the $Gly^{K-1}$ and $Gly^{K-6}$ residues mark the beginning and the end of the P-loop and enable the bending of the backbone; $Gly^{K-1}$, in addition, electrostatically stabilizes the $O^{3A}$ atom of α-phosphate. The signature Lys residue electrostatically interacts with $O^{1B}$ and $O^{2G}$ atoms and additionally appears to stabilize the P-loop by interacting with backbone carbonyl oxygens of K-5 and K-6 residues (Figure 3.2.31). The $T/S^{K+1}$ residue is involved in $Mg^{2+}$ coordination.

Invariant residues that stabilize the $Mg^{2+}$ ion are provided by the Walker A motif ($Thr/Ser^{K+1}$) and the Walker B motif (the signature Asp or Glu residue) see Fig. 1, 3-6 and a recent comprehensive analysis [259, 280]. The $Asp^{D}$ residue interacts with $Mg^{2+}$ via water molecules of the first coordination shell, whereas a $Glu^{D}$ residue, in some cases, can reach $Mg^{2+}$ directly. In addition, $Mg^{2+}$ is coordinated by $O^{2B}$ and $O^{1G}$ atoms of the triphosphate chain. The rest three ligands are either water molecules or family-specific residues [280]. Overall, the coordination of the $Mg^{2+}$ ion is very similar among all classes, as it involves the substrate itself, the conserved $Thr/Ser^{K+1}$ of the Walker A, and the signature Asp/Glu of the Walker B motif.

A special role among the residues of the Walker A/P-loop motif is attributed to the K-3 residue. In many families, this position contains a conserved Gly or other small residues (Ala or Ser). In some cases, this residue is explicitly involved in the activation process: in cation-dependent proteins of the TRAFAC class, this position contains the $K^+$-coordinating Asp/Asn (or $Na^+$-coordinating Ser). Furthermore, in atlastins, $Arg^{K-3}$ directly acts as a stimulating moiety

(Figure 3.2.34B). Still, in all P-loop proteins, the backbone nitrogen of the K-3 residue occupies the same position relative to the phosphate chain of the substrate. Before activation and in the early pre-transition stages this group compensates the negative charge on the bridging $O^{3B}$ atom between β- and γ-phosphates [257]. As seen in MD simulations of MnmE and EF-Tu, this residue can also form an H-bond with the γ-phosphate, provided the latter is already involved with the stimulating moiety. Finally, crystal structures of post-transition state P-loop NTPases indicate that the backbone nitrogen of K-3 residue again forms the H-bond with the $O^{3B}$ atom of the leaving group [192, 314].

#### *3.2.7.2 Stimulating moieties: ubiquitous but not conserved*

MD simulations reported in Chapters 3.2.3 and 3.2.5 revealed that the interaction with stimulating moiety twists γ phosphate, which is then stabilized by a new H-bond between $NH^{K-3}$ and $O^{2G}$. It is noteworthy that in structures where the stimulating residues interact only with γ-phosphate or its analog, as in dynamin (Figure 3.2.34) and ABC-NTPases (Figure 3.2.37) the position of stimulating moiety between α- and γ-phosphates is compatible with twisting of the γ-phosphate group.

Comparative structural analysis showed that each class of P-loop NTPases can be characterized by its signature mechanism(s) of introducing the stimulating cationic moiety into the active site (see Figures 3.2.23-26). In most cases, the interaction is driven by specific interactions with an activation partner - another domain(s), or protein(s), or RNA/DNA molecules. Activation mechanisms typical for P-loop NTPases of different classes are shown in Figures 3.2.23-26 and summarized in Table 3.9. The highly diverse stimulating moiety(ies) interact with the phosphate chain in only two distinct ways that may complement each other. In most classes of P-loop NTPases, at least one cationic moiety, as provided either by the same P-loop domain or by another domain/protein, gets inserted between α-and γ-phosphates and interacts with $O^{2A}$ and $O^{3G}$. A simultaneous interaction with the two oxygen atoms is possible only on the condition of twisting the γ-phosphate group.

In most proteins, other positively charged groups – backbone amides and/or additional Arg/Lys residues – are involved in coordinating the oxygen atoms of γ-phosphate in addition to the stimulating moieties. Interactions of such auxiliary Arg/Lys residues are often promoted by the same protein-protein interaction that inserts the stimulating moiety. This is observed in AAA+ NTPases (Figure 3.2.37A), SF1/2 helicases (Figure 3.2.37B), ABC NTPases (Figure 3.2.37D), hexamers of $F_1$/RecA-like ATPase (Figure 3.2.38), and so on. In some hexameric AAA+ ATPases,

the Arg finger in the AG site comes from the similar domain of the adjacent subunit in hexamer, whereas an additional Arg residue next to γ-phosphate is provided by the C-terminal domain of the same subunit (Figure 3.2.37A) [120]. Similarly, the lid domains of kinases and C-terminal domains in SF1/2 helicases often (but not always) carry two Arg residues on the interface directed towards the P-loop domain. One Arg residue is inserted into the AG site, while the other one interacts only with γ-phosphate (Figure 3.2.37C). These auxiliary basic residues are poorly conserved even within individual families of P-loop NTPases.

#### *3.2.7.3 Auxiliary moieties: not ubiquitous and not conserved*

In most cases, additional positively charged groups – backbone amides and/or additional Arg/Lys residues – are involved in coordinating the oxygen atoms of γ-phosphate in addition to the "main" stimulating moiety. This is observed in AAA+ NTPases (Figure 3.2.37A), SF1/2 helicases (Figure 3.2.37C), ABC NTPases (Figure 3.2.37D), hexamers of $F_1$/RecA-like ATPase (Figure 3.2.38), and so on. In some hexameric AAA+ ATPases, the Arg finger in the AG site comes from the P-loop domain of the adjacent subunit in hexamer, whereas an additional Arg residue next to γ-phosphate is provided by the C-terminal domain of the same subunit (Figure 3.2.37A) [120]. Similarly, the lid domains of kinases and the C-terminal domains in SF1/2 helicases often (but not always) carry two Arg residues on the interface directed towards the P-loop domain. These auxiliary residues are poorly conserved even within individual families of P-loop NTPases.

Such auxiliary residues are close to γ-phosphate and some of them are often involved in the stabilization of $W_{cat}$. However, sequence alignments show no polar residues in this position in most TRAFAC NTPases (see Fig. 2 in [24]). Specifically, many members of the HAS GTPase family of the TRAFAC class lack any catalytic polar sidechains around $W_{cat}$ [147], so that $W_{cat}$ is stabilized only by $HN^{D+3}$ and $CO^{T}$ of Switch I (e.g. Figure 3.1.2A, B). Hence, the participation of the polar D+4 residue in many Ras-like GTPases is family-specific and appears to be optional for the mechanism. By structure analysis, some other optional solutions for stabilizing $W_{cat}$ could be found. For instance, the MnmE GTPase, additionally, uses a Glu residue of its second α-helix to stabilize the H-bonded network around $W_{cat}$ (Figure 3.1.2A). Although this Glu is not conserved within the family, its replacement by Gln or Ala results in the loss of catalytic activity [28]. In other classes of the Kinase-GTPase division, not only the position but also the nature of $W_{cat}$-

stabilizing polar residues (if present) varies between classes and families (Figure 3.2.33, Figure 3.2.36).

In contrast, in the ASCE division, the catalytic residue, which gives the name to the whole division, is usually Glu, but its exact position varies between classes, see Figure 3.2.36. In most enzyme classes, the $Glu^{D+1}$ residue serves as catalytic, see Figure 3.2.37 and [120, 259, 278]. In $F_1$/RecA ATPases, the catalytic Glu residues could be located on the adjacent WB-1 and/or WB+1 strand, so that three β-strands appear to be involved in the $W_{cat}$ stabilization (Figure 3.2.38) [278, 279, 315]. Still, their replacement by non-polar residues usually affects the catalytic function.

In many cases, the stabilization of $W_{cat}$ by specific protein groups is coupled with the introduction of the stimulating moiety. In Ras-like NTPases, the backbone carbonyl oxygen of the stimulating Arg residue reorients this residue towards $W_{cat}$ by making an H-bond with the side chain amino group of the catalytic $Gln^{D+4}$ (Figure 3.1.2A). Similarly, in translation factors, the activating interaction with the sarcin-ricin loop of the large ribosomal subunit rotates the catalytic $His^{D+4}$ towards $W_{cat}$ ( see Chapter 3.1.3 and [127]).

In some cases, the activating partner directly provides not only the stimulating moiety but also a polar group interacting with $W_{cat}$. In some Ras-like NTPases, GAPs can provide both the Arg finger and the catalytic Gln [316]. In SIMIBI NTPases, the "activating" monomer of the dimer provides not only a Lys finger but also an Asp residue that stabilizes $W_{cat}$ (Figure 3.2.35A). In the ABC ATPase of the maltose transporter (Figure 3.2.37D) and F1-ATPases (Figure 3.2.38A), the adjacent monomer that provides the stimulating moiety also provides a backbone CO atom to coordinate the catalytic water molecule (Figure 3.2.37D, Figure 3.2.38A). Such behavior, where both the stimulating moiety and the $W_{cat}$-stabilizing residue(s) are controlled by the interaction with an activating partner, might indicate a kind of a "two-key mechanism" for better control and preventing incidental undesired NTP hydrolysis.

Hence, in most P-loop NTPases, $W_{cat}$ is stabilized by a combined action of evolutionary conserved and variable moieties. The somewhat conserved part is represented by residue(s) following Walker B motif (e.g. Switch II in TRAFAC NTPases) that bind either $W_{cat}$ alone or $W_{cat}$ and γ-phosphate, such as $Gly^{D+3}$ and $Gln^{D+4}$ of small GTPases or $Glu^{D+1}$ in most ASCE ATPases. The variable residues are provided by C-caps of WB-1 and WB+1 β-strands, family-specific protein loops (e.g. Sensor I in AAA+ proteins), or even external activating proteins/domains.

### 3.2.7.4 The minimal mechanistic model of NTP hydrolysis by P-loop NTPases

Building on the results of comparative structural analysis and MD simulations, as well as on available experimental and theoretical data, the activation of a P-loop NTPase could be described by a simple mechanistic model that is shown in Figure 3.2.39. The minimalistic version of the model includes ubiquitous Walker A and B motifs, $Mg^{2+}$-NTP, a single, simple stimulating moiety, such as a $K^+$ ion or a Lys residue, and a catalytic water molecule. The model, however, can be easily expanded/modified to fit distinct NTPase families by adding further stimulating moieties, auxiliary residues, and further water molecules and $W_{cat}$-stabilizing residues.

According to the model, the catalytic transition proceeds in the following steps shown in Figure 3.2.39:

A) Upon Mg-NTP binding to the P-loop, the energy of binding is used to bring the Mg-NTP molecule into an extended configuration with eclipsed β- and γ-phosphates. The interaction of the phosphate chain with multiple positively charged moieties makes it more prone to hydrolysis.

B) An activating interaction between the P-loop domain and the activation partner (another protein or a separate domain of the same protein and/or an RNA/DNA molecule) leads to the insertion of a stimulating moiety next to the phosphate chain and also brings auxiliary/catalytic residues closer to γ-phosphate.

C) In most cases, the stimulating moiety links the $O^{2A}$ and $O^{3G}$ atoms as depicted in Figure 3.2.39C. Otherwise, the stimulating moiety interacts only with γ-phosphate. Either way, the activating interaction, via twisting/pulling of γ-phosphate brings the triphosphate chain into a near-eclipsed configuration which is stabilized by a novel H-bond between $O^{2G}$ and $NH^{K-3}$. This interaction weakens the $O^{3B}$-$P^G$ bond, promotes the transition of γ-phosphate into a more planar conformation.

D) Concurrently with the twisting of γ-phosphate, the $W_{cat}$- site is formed with obligatory participation of residues of auxiliary/catalytic residues. In any case, $H_2O_{cat}$ in the apical position attacks $P^G$ without notable delay [317].

E) Nucleophilic attack results in the formation of a covalent bond between $OH_{cat}^-$ and $P^G$, the planar γ-phosphate repulses the oxygen atoms of β- and α-phosphates, which leads to elongation of the bonds between $P^G$ and $P^B$ and its cleavage.

The steps (C) – (E) are separated for the sake of clarity, but most likely these steps proceed in a concerted way.

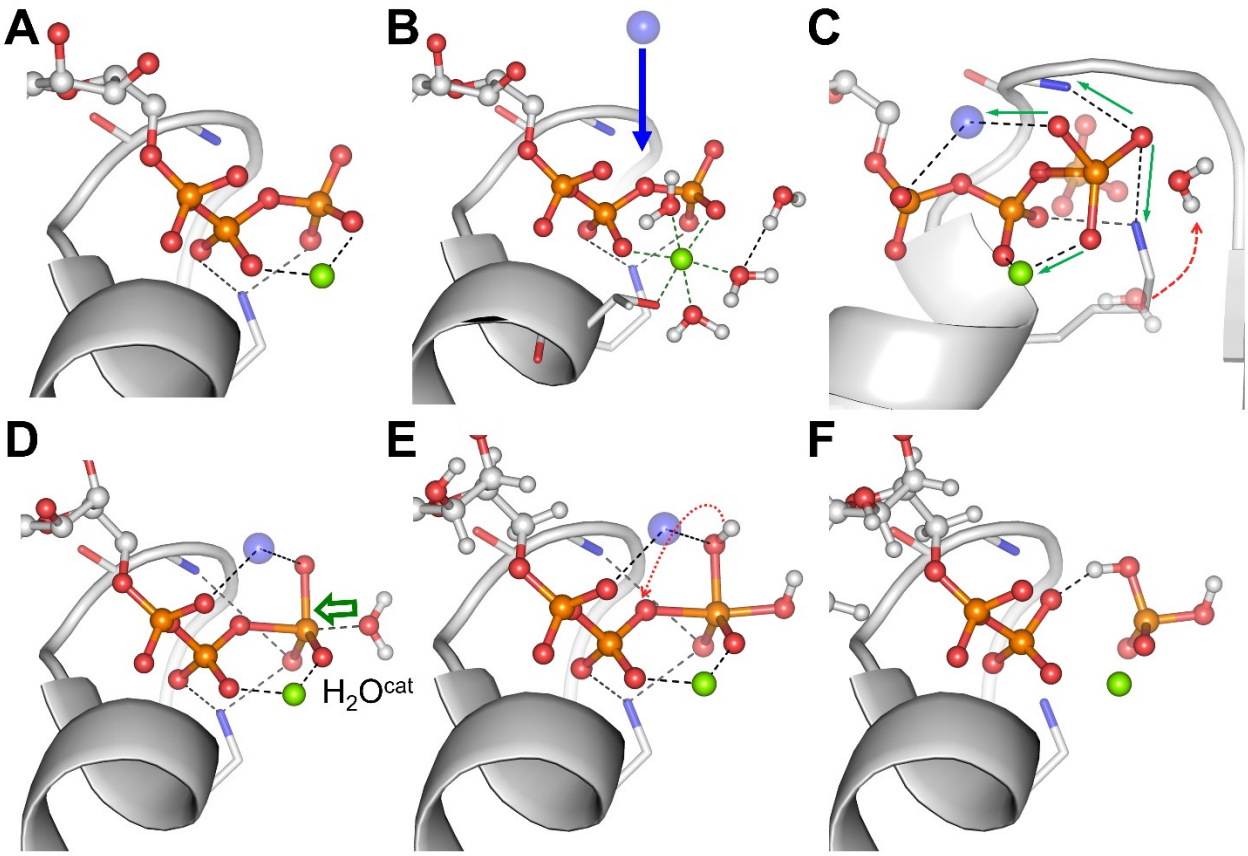


**Figure 3.2.39**. Basal activation mechanism of NTP hydrolysis by P-loop NTPases.
**A**. Triphosphate binding to the P-loop ensures catalytically prone conformation with eclipsed β- and γ-phosphates. **B**. Interaction with the activating partner brings a stimulating moiety into the AG site between α- and γ-phosphates. **C**. Interaction with the stimulating moiety leads to rotation of the γ-phosphate and its planarization due to the new H-bond with the backbone of K-3 residue of the P-loop. **D**. Planarized and conformationally strained γ-phosphate is susceptible to the nucleophilic attack by a polarized catalytic water molecule. **E**. Pentavalent intermediate precedes the bond cleavage. **F**. The phosphoester bond between β- and γ-phosphates is broken. See the main text for more details.

# 4. Eukaryotic G Protein-Coupled Receptors

## 4.1. Background

G-protein coupled receptors (GPCRs) are heptahelical transmembrane proteins (7TM) that contain seven transmembrane helices around a relatively polar core [318-320] (see Figure 4.1.1A-C). In the majority of GPCRs, binding of an endogenous ligand (agonist) by an "inactive" conformation of the protein triggers conformational changes in the protein (Figure 4.1.1). The GPCR in its activated conformation interacts with a G-protein, which then launches the signaling cascade in the cell. GPCRs are widespread among eukaryotic organisms and are exhaustively studied for their involvement in various cellular processes. Human GPCRs are targets for around half of all known drugs [318].

GPCRs are grouped into several classes: class A includes rhodopsin-like receptors, class B - secretin receptors, class C - glutamate receptors, class D - fungal mating pheromone receptors, class E - cAMP receptors, and class F includes frizzled receptors. In humans, class A GPCRs constitute the largest protein family, encoded by more than 700 genes, while only about 150 genes have been attributed to all other classes of GPCRs combined [321].

The recent surge of new high-resolution X-ray structures, depicting GPCRs without ligands, as well as with agonists and antagonists bound, uncovered many crucial features of their functioning (see [318-320, 322-331] and references therein). One such feature is that the activation of GPCRs is accompanied by a large motion of the cytoplasmic end of the helix 6. Conserved Trp residue serves as a pivot for this movement motion (see Figure 4.1.1C and [319, 323, 332-338]).

For GPCRs a generic scheme of residue numbering is used in most studies to ease the comparison of the findings across different GPCRs and ensure uniformity of the nomenclature [339, 340]. This scheme exploits highly conserved residues present in each of the seven transmembrane helices. Within this scheme each residue is denoted by two numbers: first one indicates the number of the helix (1-7) and the second number denotes the exact position of the residue. In each helix the most conserved residue was identified and assigned the number 50, and the rest of the residues within the same helix are numbered relatively to this residue [339, 340]. For example, the mentioned above conserved Trp in the helix 6 is located in the middle of the helix, and is referred to as Trp6.48, because the most conserved residue in helix 6 – among GPCRs - is Pro 6.50. This numbering scheme is used throughout this study to refer to particular residues in GPCRs.

Another feature revealed by crystal structures is the binding of $Na^+$ ions near the ligand-binding sites in class A GPCRs (see [43, 326, 329, 330, 341, 342] and references therein). The $Na^+$-binding site in GPCRs is located between the helices, in the center of the protein and is open to the extracellular side but is isolated from the cytoplasm by a hydrophobic residue cluster (Figure 4.1.1A, B and [43, 338]). Residues, involved in $Na^+$ binding are highly conserved across class A GPCRs [43, 342], which indicates that the $Na^+$ ion binding must be functionally important. This has been confirmed by mutagenesis experiments: replacement of the key $Na^+$-binding residue Asp2.50 (see Figure 4.1.1) can lead to an increase of the association constants with the agonist by 2–3 orders of magnitude (see [43, 342-347] and references therein). Furthermore, $Na^+$ depletion was shown to increase the basal activity of several receptors even in the absence of agonists, indicating that the $Na^+$ ion is responsible for stabilization of the inactive state of the protein ( [43, 342-348] and references therein). Finally, $Na^+$ ions cannot be observed in the structures of active GPCRs (bound with agonists. As compared to the inactive structures, upon activation the $Na^+$-binding pocket shrinks from ~200 to 70 $Å^3$ due to TM helices movement (see Figure 4.1.1, [43, 326, 342, 349-351] and references therein). Since the active structures do not seem to have the space for a $Na^+$ ion to bind, it is likely that the $Na^+$ ion leaves the GPCR upon activation.

The evolutionary relationships between GPCRs and another large group of sensory transmembrane proteins, namely microbial rhodopsins (MRs), have been pondered by biologists for decades. Both groups unite proteins which consist of 7 transmembrane helices surrounding a polar core [352-354]. MRs (also referred to as type I rhodopsins) contain retinal and function either as light sensors or as light-driven ion pumps primarily in bacteria and archaea, but also in some eukaryotes [353, 355-359]. MRs are a distinct family that belongs to a larger group of bacterial 7TM receptors [360, 361], distinguished by the binding of retinal in the center of their heptahelical bundle. The GPCR (super)family unites primarily eukaryotic receptors that bind various small molecules, but also includes visual rhodopsins, which contain retinal and serve as light sensors in animals [322-324, 353, 357].

The comparative analysis of MRs and GPCRs is complicated by the uncertainty of the monophyletic origin of the eukaryotic 7TM-receptors, including GPCRs [321, 360-363]. Additionally, some groups of eukaryotic 7TM-receptors, so-called GPCR-like proteins, have been shown to operate independently from G-proteins [361, 364]. Finally, no statistically significant sequence similarity could be found between the GPCRs of class C (glutamate receptors) and other classes. Phylogenetic analyses were able to trace the class C and class E (cAMP receptors) to a eukaryotic root, and class E has was proposed to be the ancestral group for the GPCR classes A, B, D, and F [321, 362]. However, the first two resolved structures of glutamate receptors have

revealed their striking similarity to the structures of other GPCRs [365, 366], encouraging the suggestion, that all GPCR-like proteins did have a common origin.

In 2013 the first sodium-translocating microbial rhodopsins were discovered [367, 368] and soon later the structures of one such $Na^+$-transporting rhodopsin from *Krokinobacter eikastus* (hereafter KR2) have been reported [369, 370]. This data opens a new opportunity to address the long-standing question of the evolutionary relationships between GPCRs and MRs. In many GPCRS, particularly of class A, sodium ions are known to affect the agonist binding and signaling [43]. Furthermore, binding of a $Na^+$ ion in the center of these receptors has been characterized in detail [43, 326, 329, 330, 371]. Thus, in this work, the shared ability of class A GPCRs and Na-transporting MRs was used as a starting point for the comparative structure analysis of these two protein families. Results of this comparison and its discussion are presented in Chapters 4.2.1 and 4.2.2, accordingly.

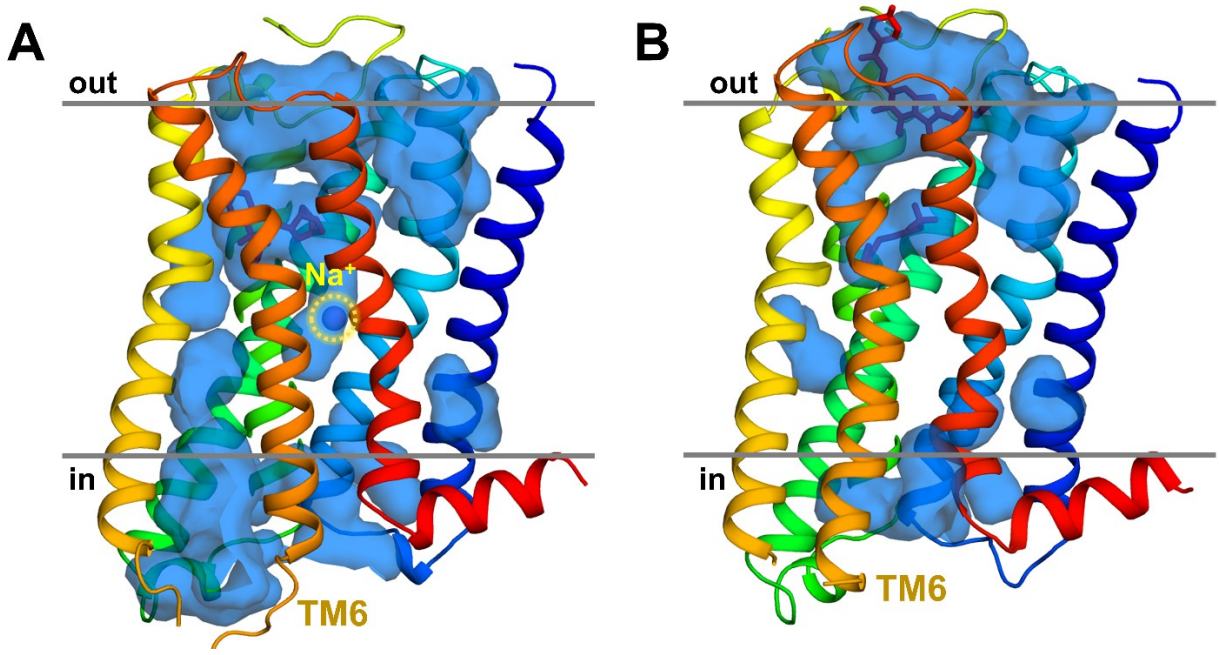


**Figure 4.1.1.** Structures of GPCR in active and inactive conformations. A, B. Cavities in the structures active and inactive states of the muscarinic receptor $M_2$, A - inactive state, PDB ID 3UON [328], B - active state PDB ID 4MQT [350]. Cavities are shown as blue volume, yellow dashed circle highlights the location of the $Na^+$ ion, GPCR helices are colored in rainbow order.

In 2014, a seminal suggestion was made by Katritch and colleagues. They suggested, that upon GPCR activation instead of returning to the extracellular space, the $Na^+$ ion in GPCRs is expelled into the cytoplasm [43]. This would require at least a temporary opening of a channel for the $Na^+$ ion. Available structures of activated GPCRs show no $Na^+$ binding within the heptahelical bundle and no evidence of such channel (see Figure 4.1.1) [349-351]. Still, some computer simulations have shown the formation of a transient water-filled channel leading from the $Na^+$-binding site to the cytoplasmic side [336, 347, 372, 373]. Since the cytoplasm is negatively charged

in relation to the extracellular medium, the translocation of $Na^+$ ion to the cytoplasm would give an energy gain, beneficial for the GPCR activation [43]. If GPCRs benefit from the membrane potential, depolarization of the membrane would notably affect the activation of GPCRs. Indeed, such effects have been observed for multiple receptors. However, while some GPCRs were inhibited by depolarization, others show no such effect (see [374, 375] for reviews and Table 4.1 for further references). Thus, the relationship between the transmembrane gradient of $Na^+$ ions and the activation of GPCRs requires further investigation.

**Table 4.1.** Experimental data on voltage-sensitive activation of GPCRs

| Receptor | Agonist | Agonist type | Voltage effect on activation | Method of activity measurement | Refs. |
|---|---|---|---|---|---|
| M2 muscarinic receptor | Acetylcholine | Full agonist (endogenous) | Enhanced | GIRK currents | [376, 377] |
| M2 muscarinic receptor | Oxotremorine | Full/strong agonist | Enhanced | | |
| M1 muscarinic receptor | Acetylcholine | Full agonist (endogenous) | Decreased | | |
| M2 muscarinic receptor | Acetylcholine | Full agonist (endogenous) | Enhanced | GIRK currents | [378] |
| M2 muscarinic receptor | Acetylcholine | Full agonist (endogenous) | Enhanced | ACh-activated $K^+$ current | [377, 379] |
| M2 muscarinic receptor | Pilocarpine | Partial agonist | Decreased | | |
| M2 muscarinic receptor | Acetylcholine | Full agonist (endogenous) | Enhanced | ACh-activated $K^+$ current | [377, 380] [381] |
| M2 muscarinic receptor | Pilocarpine | Partial agonist | Decreased | | |
| M2 muscarinic receptor | Bethanechol | Agonist (low affinity) | No effect | | |
| M1 muscarinic receptor | Carbachol | Full/strong agonist | Decreased | FRET-based assays | [377, 382] |
| M3 muscarinic receptor | Carbachol | Full/strong agonist | Enhanced | | |
| mGluR3 glutamate receptor | Glutamate | Full agonist (endogenous) | Enhanced | $K^+$ currents and Cl- currents | [383] |
| mGluR1a glutamate receptor | Glutamate | Full agonist (endogenous) | Decreased | | |
| α2A-AR adrenergic receptor | Noradrenaline | Full agonist (endogenous) | Enhanced | FRET-based assays | [384] |
| β1-AR adrenergic receptor | Isoprenaline | Full agonist | Enhanced | FRET-based assays | [385, 386] |
| β1-AR adrenergic receptor | Adrenaline | Full agonist (endogenous) | Enhanced | | |

| | | | | | |
|---|---|---|---|---|---|
| Dopamine D2L receptor | Dopamine | Full agonist (endogenous) | Enhanced | GIRK currents | [173, 387] |
| Dopamine D2L receptor | Quinpirole | Full agonist | Enhanced | | |
| Dopamine D2S receptor | Dopamine | Full agonist (endogenous) | Enhanced | GIRK currents | [388] |
| Dopamine D2S receptor | Dopamine | Full agonist (endogenous) | Enhanced | GIRK currents | [387, 389] |
| Dopamine D2S receptor | P-tyramine | Partial agonist | Decreased | | |
| Dopamine D2S receptor | M-tyramine | Partial agonist | Decreased | | |
| dopamine D2S receptor | Phenylamine | Partial agonist | Decreased | | |
| Dopamine D2S receptor | S-5-oh-dpat | Full agonist | No effect | GIRK currents | [387, 389] |
| Dopamine D2S receptor | R-5-oh-dpat | Full/strong agonist [387] | No effect | | |
| Dopamine D2S receptor | R-7-oh-dpat | Full/strong agonist | No effect | | |
| Dopamine D2S receptor | Ris-oh-dpat | Full/strong agonist [387] | No effect | | |

## 4.2. Results and Discussion

### 4.2.1. Comparative Analysis of G-protein Coupled Receptors and Microbial Rhodopsins

For the initial analysis, likely $Na^+$-binding ligands of KR2 (PDB ID 4XTL [369]), identified in refs. [358, 367-370], and the $Na^+$-binding ligands of $Na^+$-bound δ-opioid receptor (hereafter δ-OR, PDB ID 4N6H [329]) were superposed manually. In both proteins, the entire sets of potential $Na^+$-coordinating residues are located in helices 3/C and 7/G. Manual superposition of the (predicted) sodium-binding sites of the two structures was constructed in PyMol [52] and was able to match the locations of these ligands (not shown). Following this finding, several different software packages for sequence alignment and structural superposition were applied to align the entire proteins. Structure superposition methods used here are discussed in more detail in Chapter 2.2.4.

The "3D-similarity" option on the web server of the Protein Data Bank (PDB) [80] provides protein structure alignments, pre-calculated by the jFATCAT-rigid algorithm [81, 83]. In the ranked list of proteins with detected structure similarity to KR2, MRs were followed by structures

of diverse GPCRs including several $Na^+$-bound GPCRs, namely the proteinase-activated receptor 1 (PDB ID 3VW7, RMSD 4.27 Å, P-value 1.53E-8), A2A adenosine receptor (PDB ID 4EIY, RMSD 4.16 Å, P-value 1.84E-8), β1 adrenergic receptor (PDB ID 4BVN, RMSD 4.6 Å, P-value 2.37E-8), and the δ-OR (PDB ID 4N6H, RMSD 4.49 Å, P-value 2.38E-8).

A better superposition was produced for KR2 (PDB ID 4XTL [369]) and the δ-OR (PDB ID 4N6H [329]) by the SSM (secondary structure match) method available at the PDBeFold server [82]. This superposition had an RMSD of 3.8 Å, aligned all helices (189 residues in total), and matched all $Na^+$ ligands. Additionally, for visualization purposes only, the PDBeFold algorithm [82] was used to construct a superposition that included only the most conserved parts of the structures, excluding helices 4/D and 5/E. The resulting superposition matched the same residues, as the full-protein alignment, but provided better overlap of the structures in the $Na^+$-binding site and surrounding area with RMSD of 2.97 Å across 135 residues (Figure 4.2.1).

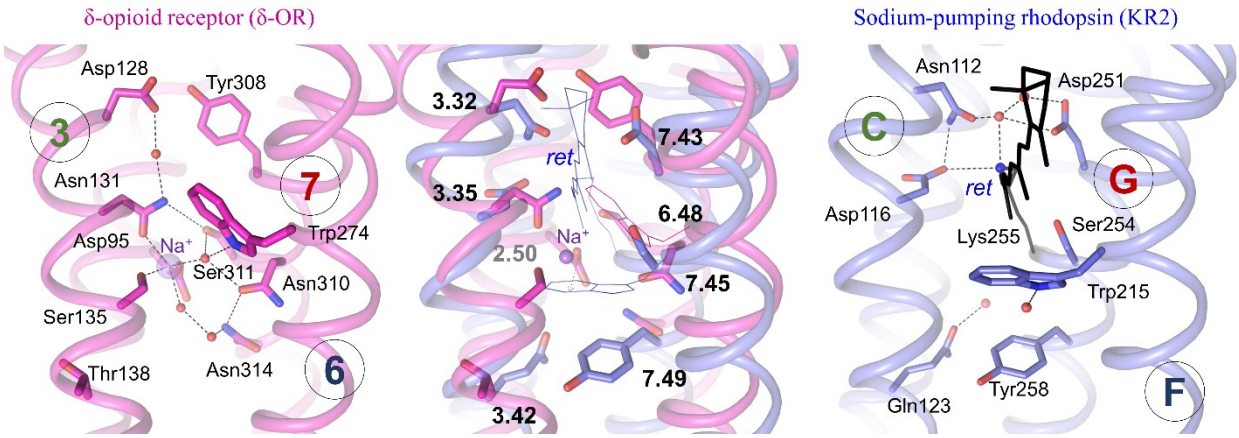


**Figure 4.2.1.** Superposition of $Na^+$-binding sites in δ-OR and KR2. Left panel - $Na^+$-binding site of δ-OR (PDB ID 4N6H), right panel - $Na^+$-binding site of KR2 (PDB ID 4XTL), middle panel – superposition the proteins. Proteins are shown with extracellular side on the top. Only three helices are shown (3,6,7 and C, F, G). The $Na^+$ ion is shown as a transparent pink sphere. Retinal bound to lysine in KR2 is shown in grey with the Schiff base nitrogen shown as a blue sphere; water molecules are shown as red spheres. Residue numbers are given in accordance with Ballesteros–Weinstein nomenclature [339, 340].

In this structural superposition, helices 3/C and 7/G were matched, and KR2 residues that were predicted to coordinate the $Na^+$ ion as it passes through the middle of the membrane [367-370, 390] overlapped with the residues that coordinate $Na^+$ in its binding site of δ-OR. In GPCRs, $Na^+$ ligands form the $Na^+$-binding pocket not only in δ-OR (Figure 4.2.1), but also in the protease-activated receptor 1 (PAR1) [371], $A_{2A}$ adenosine receptor [326], and in $\beta_1$-adrenoreceptor [43, 330]. Specifically, Asp116 of the characteristic NDQ motif of $Na^+$-transporting MRs [368] in the helix C of KR2 matched with $Asn131^{3.35}$ of δ-OR (superscripts give Ballesteros–Weinstein residue numbering [339, 340]), while residues Ser254 and Tyr258 of KR2 helix G matched with

Asn310$^{7.45}$ and Asn314$^{7.49}$ of δ-OR, respectively. In addition, Asn112 of the NDQ motif of KR2 (helix C) matched Asp128$^{3.32}$ of δ-OR, whereas Asp251 of KR2 matched Tyr308$^{7.43}$ of δ-OR.

The pairwise structural alignment of KR2 (PDB ID 4XTL [369]) and δ-OR (PDB ID 4N6H [329]) was used to construct superposition of multiple representative MR and GPCR structures, as well as the corresponding sequence alignment. Representative structures of GPCRs and MRs used in this superposition are listed in Table 4.2. Microbial rhodopsins (MR) structures and GPCR structures for this superposition were selected manually from the PDB ID. All MR structures were aligned to the KR2 structure (PDB ID 4XTL [369]), and all GPCR structures were aligned to δ-OR (PDB ID 4N6H [329]) using the results produced by the jFATCAT-rigid algorithm [81] (Figure 4.2.2). The resulting sequence alignments were inspected manually and corrected in a few cases to ensure the best matching to the respective structural superposition (Figure 4.2.3).

**Table 4.2.** Structures of MRs and GPCRs used for the superposition. All GPCRs listed belong to class A, except for the glutamate receptor (Class C).

| PDB | RMSD, Å * | Number of aligned residues* | Protein name | Ref. |
|---|---|---|---|---|
| Microbial rhodopsins, aligned to KR2 ( PDB ID 4XTL) | | | | |
| 4XTL | N/A | N/A | Sodium pumping rhodopsin (KR2) | [369] |
| 3QBG | 3.30 | 231 | Halorhodopsin | [391] |
| 3QAP | 2.97 | 216 | Sensory rhodopsin II | [392] |
| 6EID | 3.00 | 220 | Channelrhodopsin | [393] |
| 2JAF | 3.20 | 223 | Halorhodopsin | [394] |
| 5AX0 | 3.22 | 225 | Rhodopsin I | unpublished |
| 4JQ6 | 2.02 | 196 | Blue-light absorbing proteorhodopsin | [395] |
| 4HYJ | 2.39 | 233 | Proton-pumping bacteriorhodopsin | [396] |
| 3DDL | 2.43 | 247 | Xanthorhodopsin | [397] |
| G-protein coupled receptors, aligned to δ-OR (PDB ID 4N6H) | | | | |
| 4N6H | N/A | N/A | δ-opioid receptor (δ-OR) | [329] |
| 4DKL | 0.67 | 216 | μ-opioid receptor | [398] |
| 4BVN | 1.96 | 169 | β1-adrenoceptor | [330] |
| 2RH1 | 1.55 | 184 | β2-adrenoreceptor | [399] |
| 3PBL | 3.12 | 218 | Dopamine D3 receptor | [400] |
| 3VW7 | 2.39 | 197 | Protease-activated receptor 1 | [371] |
| 4EIY | 2.99 | 204 | A(2A) adenosine receptor | [326] |
| 4BUO | 2.67 | 223 | Neurotensin receptor 1 | [401] |
| 4IAR | 1.69 | 191 | Serotonin receptor | [402] |
| 3V2Y | 3.54 | 198 | Lipid G protein-coupled receptor | [403] |
| 4MQS | 1.98 | 181 | M2 muscarinic acetylcholine receptor | [350] |
| 4MBS | 1.75 | 226 | CCR5 chemokine receptor | [404] |
| 1U19 | 2.85 | 193 | Visual pigment rhodopsin | [405] |
| 4OR2 | 3.01 | 202 | Glutamate receptor 1 (Class C GPCR) | [365] |

* Alignment properties are given as provided by the jFATCAT algorithm on the PDB website [81, 83].

The alignment in Figure 4.2.3 offers a reason as to why earlier attempts to establish relatedness between MRs and GPCRs from sequence similarity have not yielded convincing results: these studies targeted animal visual rhodopsins, which, evidently share the least similarity with MRs. Meanwhile, other GPRCs, particularly those with $Na^+$-binding sites [43, 318, 326, 329, 330, 371], exhibit much more similarity of structures and sequences to MRs (Figure 4.2.3). The major divergence of sequences between animal rhodopsins and non-opsin GPCRs seems to be associated with the residue $Lys^{7.43}$ which is located on the same helix 7/G as the retinal-binding Lys residues in MRs [353] albeit in a different position within the helix (Figure 4.2.3). Hence, results presented here, on the one hand, strongly indicate homology of MRs and GPCRs, while on the other hand, still support the notion that the use of retinal in MRs and GPCRs is a result of convergent evolution (see e.g., [406]).

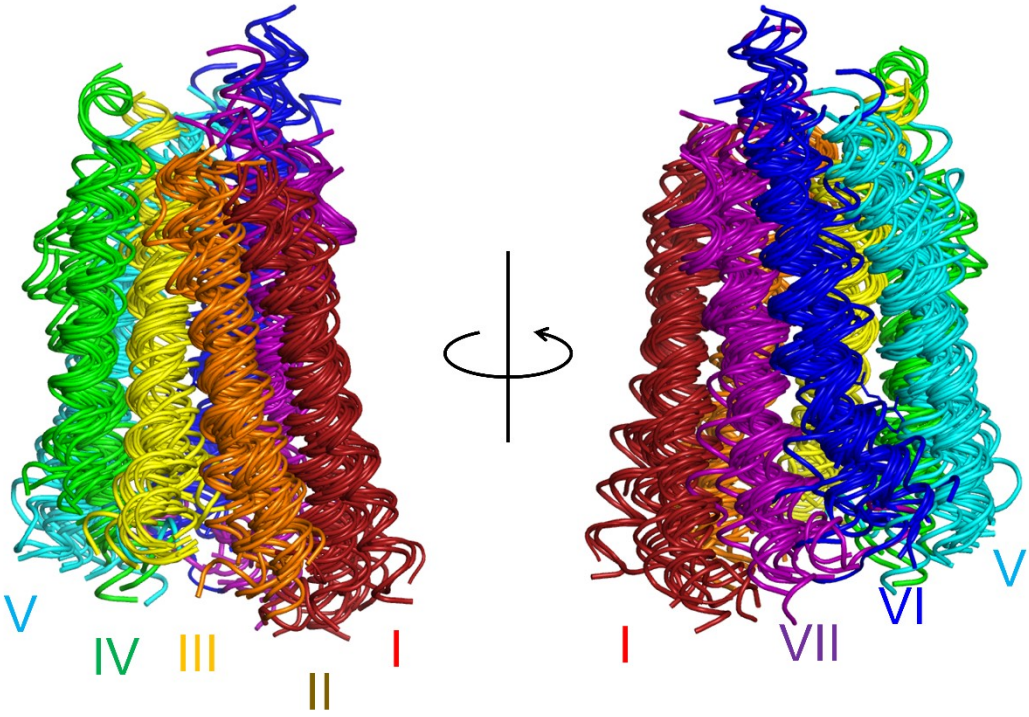


**Figure 4.2.2.** Superposition of representative structures of MRs and GPCRs. Only transmembrane helices are shown, helices 1-7 and A-G are colored in the rainbow color order.

According to the alignment in Figure 4.2.3, only one residue is conserved throughout MRs and GPCR: Trp in the middle of helix 6/F, e.g. Trp215 of KR2 or $Trp274^{6.48}$ of δ-OR. Remarkably, in both groups of proteins, these residues have been shown to play an integral role in protein function. In MRs, this residue couples retinal photoisomerization with the conformational changes in helix F (see Figure 4.2.4), which lead to either ion translocation (e.g., in bacteriorhodopsin [407-410], halorhodopsin [411], $Na^+$-translocating rhodopsin [412] and channelrhodopsin [413, 414]) or signal transduction (e.g. in sensory rhodopsins [353, 415]). In GPCRs, $Trp^{6.48}$ also mediates the

signal transduction as it interacts with agonists (see Figure 4.2.4), and coordinates the $Na^+$ ion via a water molecule, (see Figure 4.2.1 and [43, 319, 371]).

Conformational change of the conservative Trp residue in helix 6/F in response to either agonist binding (see Figure 4.2.4 and [319]) or photoisomerization of the retinal in rhodopsin [323, 324] triggers the tilting and rotation of helix 6/F. The conservation of this "pivotal" Trp residue [323, 324] likely reflects an already mentioned [407] underlying similarities of the molecular mechanisms of MRs and GPCRs. Most of these proteins seem to rely on forcing a reorientation of this large and restrained by hydrogen bonds residue, which leads to the conformational changes in the helix 6/F. The pivotal role of the aminoacid in this position remains even in those rare cases when the Trp is replaced, e.g. in the human PAR1, this position is taken by Phe, which can still fulfill this role [371].

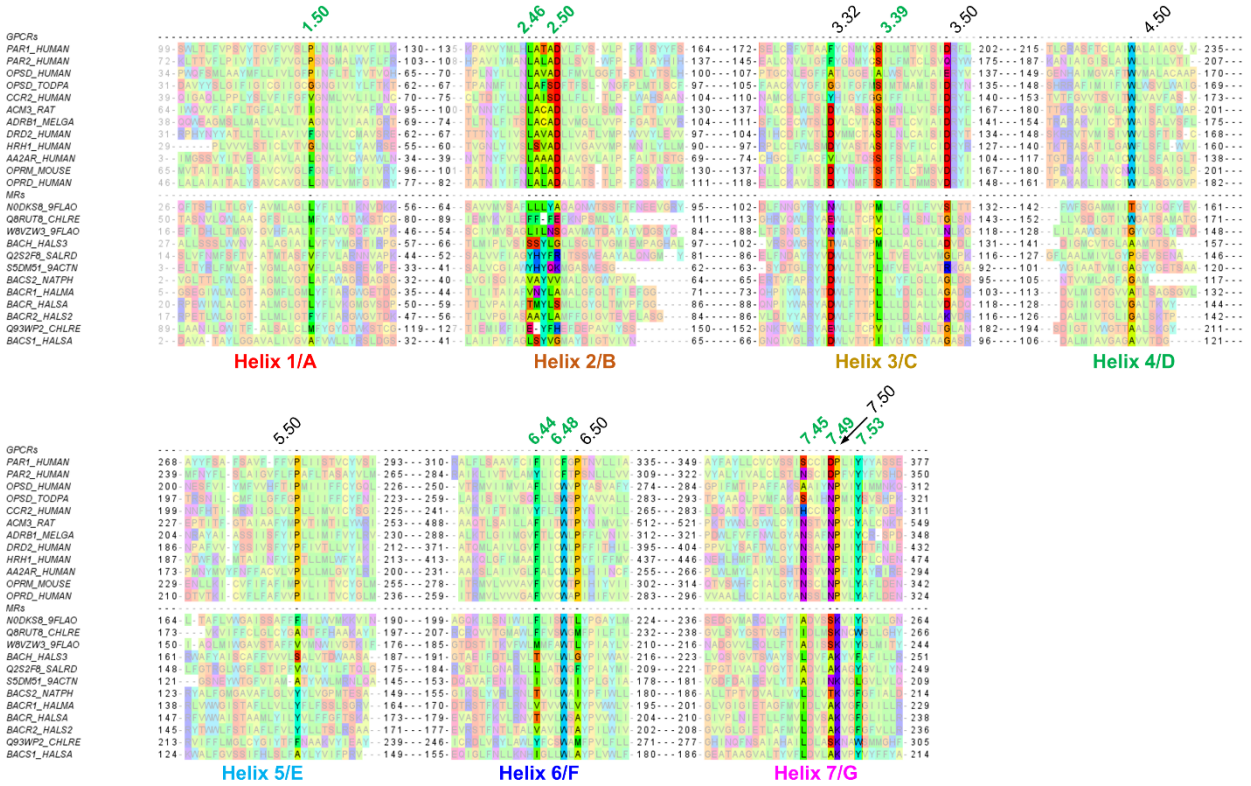


**Figure 4.2.3.** Multiple sequence alignment of MRs and GPCRs constructed based on the structure superposition. Sequences used in the alignment are listed under their UniProt IDs. The most conserved residue (.50 according to Ballesteros–Weinstein nomenclature [339, 340]) in each helix is labeled in black, residues, involved in the coordination of $Na^+$ ion are labeled in green.

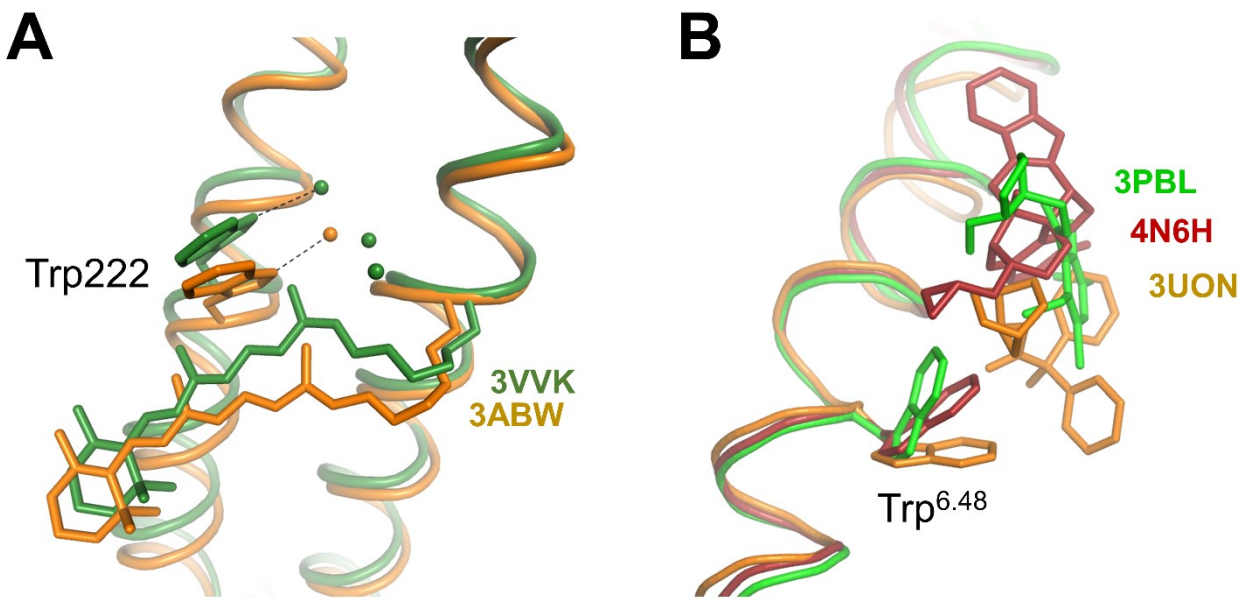


**Figure 4.2.4.** Conserved Trp movement in MRs and GPCRs. A. Coupling of the retinal conformation with the orientation of Trp222 and tilting of the helix F in MRs. Structures of azide-bound halorhodopsin from *Natronomonas pharaonic* are shown, unphotolyzed state (PDB ID 3ABW) is shown in orange, the reactive state (PDB ID 3VVK) is colored green. Water molecules are shown as spheres. B. Sidechain positions of Trp6.48 in GPCRs with different signaling molecules bound. Superposition of three receptors is shown: dopamine $D_3$ receptor (PDB ID 3PBL, green), the δ-opioid 7TM receptor (PDB ID 4N6H, red), and the $M_2$ muscarinic receptor (PDB ID 3UON, orange)

To explore the similarity of MRs with other classes of GPCRs the superposition of the class C human G protein-coupled metabotropic glutamate receptor 1 (GluR, PDB ID 4OR2) with δ-OR and KR2 was constructed. The PDBeFold server was used for the similarity search between the GluR and the whole PDB archive. Among the best matches were turkey β-1 adrenergic receptor (PDB ID 2VT4, RMSD 2.82 Å on 195 residues) and bacteriorhodopsin (PDB ID 1X0K, RMSD 3.87 Å on 189 residues). Those matches were used to build the presented superposition of GluR with δ-OR and KR2.

Currently available structures of GPCRs of Class C [365, 366], show structural similarity to other GPCRs, but do not contain bound $Na^+$ ions. Structural superposition of KR2, δ-OR (class A GPCR), and metabotropic glutamate receptor 1 (class C GPCR, PDB ID 4OR2) in Figure 4.2.5, reveals the conservation of a Trp residue in the helix 6/F. However, the orientation of this Trp residue sidechain in KR2 appears to be intermediate between two different orientations, present in the representative class A and class C GPCR structures (Figure 4.2.5). In δ-OR (class A) $Trp^{6.48}$ coordinates the $Na^+$ ion via a water molecule, while in the class C receptor, $Trp^{6.48}$ is turned away and stabilizes helix 5, which, is positioned closer to the center of the 7TM bundle, as compared to other GPCRs [365, 366].

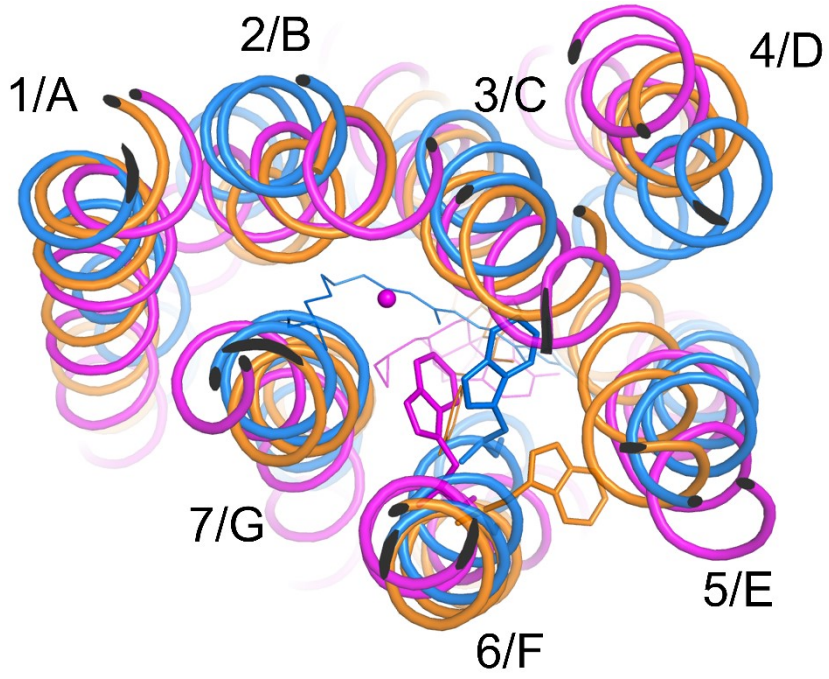


**Figure 4.2.5.** Human glutamate receptor (class C GPCR) superposed with δ-OR (Class A GPCR) and KR2. Glutamate receptor 1 (PDB ID 4OR2) is shown in orange, δ-OR (PDB ID 4N6H) is shown in magenta, KR2 (PDB ID 4XTL) is shown in blue.

Aravind and co-workers [361] have reported comparative genome analysis of the GPCR signaling system. According to their data, some of the early 7TM proteins, which could be already present in the LECA (Last Eukaryotic Common Ancestor), resemble certain bacterial 7TM receptors and show a relationship with the class C GPCRs "suggesting that the latter type could have emerged secondarily from a precursor of the former type" (quoted from [361]). This conclusion supports the notion that class C GPCRs originated directly from a eukaryotic root, independently from other GPCRs [321, 362, 363]. The structural alignment in Figure 4.2.2 and corresponding sequence alignment support the common origin of all GPCRs from MRs and affirm the importance of the Trp residue in the helix 6/F, which is conserved in the glutamate receptors as well. However, the intermediate position of the KR2 structure and sequence between class A and class C GPCRs supports the notion that class C GPCRs might have evolved independently from GPCR of other classes [321, 361-363].

Thus, the evidence from comparative genomics studies [360, 361], phylogenetic analyses [321, 362, 363], and reported here comparative structure analysis indicate that the LECA could already contain 7TM receptors of two types. One of them could be an independent 7TM receptor which likely developed from an ancient KR2 with retained $Na^+$-binding site, that would spawn classes A, B, D, and F of GPCRs. Another type of 7TM receptors in LECA could contain an RGS domain and emerge from bacterial 7TM receptors (which are, in their turn, traceable to bacterial rhodopsins [360]) and later develop into the glutamate receptors (class C GPCRs).

### *4.2.2. Common Origin of G-protein Coupled Receptors and Microbial Rhodopsins*

Proposed here origin of a large group of transmembrane proteins from an ancient sodium pump evokes at least one other case of a (super)family of membrane proteins, that likely originated from an ancient system of Na/K homeostasis. Namely, bacterial and archaeal $Na^+$-translocating ATP synthases have remarkably similar $Na^+$-binding sites. From their structural superposition, the common ancestor of diverse rotary ATP synthases was revealed to be an ATP-driven $Na^+$-pump [416-419]. Already the primordial cells should have possessed systems that expelled $Na^+$, since, as discussed above in Section 3.1.2, several ancient cellular systems, attributed to the Last Universal Cellular Ancestor (LUCA) require $K^+$ ions and are inhibited by $Na^+$ ions [7, 8, 416]. With these results in mind, a similar approach was used here in a comparative analysis of the $Na^+$-binding sites in KR2 (PDB ID 4XTL [369]) and various GPCRs.

Earlier attempts to establish evolutionary relatedness or detect significant similarity between MRs and GPCRs were hindered by the lack of a common function, so that one could not identify a common structure-function relationship or a mechanism, uniting the two groups of proteins, despite obvious overall similarities between the protein families. In this work, the common ability to bind sodium ions was used a such a common function. Availability of crystal structures of GPCRs with bound $Na^+$ ions and structures of novel Na-pumping rhodopsins with surmised $Na^+$-coordinating residues was used not only to establish the structural similarity between the two protein (super)families but also to reveal other common structure-function relationships. Namely, conservation of a Trp residue in helix 6/F, the involvement of this residue in the coupling of the signal (retinal isomerization or agonist binding) with the conformational changes in the protein, and the large movement of the helix 6/F upon the activation are all common for MRs and GPCRs. Multiple sequence alignment of MRs and GPCRs, build on the pairwise alignment of δ-OR and KR2 (Figure 4.2.3), shows that KR2 acts as an intermediate between other MRs and GPCRs of class A. This relation might indicate its proximity to the common ancestor of the two (super)families, which, contained a Na-binding site and likely was a light-driven $Na^+$ export pump (KR2).

Uncovered here conservation of particular structural elements in MRs and GPCRs was used to constructs the most likely evolutionary scenario, describing the evolution of 7TM receptors and the emergence of different GPCRs and MRs from an ancient light-driven sodium export pump (Figure 4.2.6). Pumping $Na^+$ ions necessitates a path through the membrane laced with negatively charged or polar groups, which could compensate the positive charge of the ions [420]. To ensure

one-way pumping, the system also requires a switching mechanism [421]. In the modern ion-pumping MRs, such switching relies on the conserved Trp residue of the helix 6/F; as the light-induced movement of this helix opens a passage for the ions [407-414]. Since both MRs and GPCRs utilize a turn and/or tilt the helix 6/F starting from the conserved Trp, the ancestral protein likely also already possessed this mechanism. Forced tilting of the helix that starts in the middle of a transmembrane segment, additionally accompanied by other significant conformational changes [407-414, 422], is not a common feature, and likely only emerged once, shaped by the tight-triggered isomerization of the retinal.

The ancestral light-driven sodium pump could have evolved further into heptahelical proteins with different functions. As shown in Figure 4.2.6, the relationships between these 7TM proteins can be explained as a series of gains and losses of the $Na^+$ and/or retinal binding residues. For instance, proton- and $Cl^-$-pumping rhodopsins, along with various sensory rhodopsins, have fewer ion-binding residues than KR2 [358-361, 423-426]. Thus, these proteins could have evolved from the $Na^+$-pump after a loss of some of the $Na^+$ ligands left the protein unable to transport $Na^+$, but still able to transport small or negatively charged molecules ($H^+$ or $Cl^-$), or at least still undergo conformational changes in response to the light-induced retinal isomerization. Conversely, the emergence of additional charged or polar groups within the transmembrane path of the translocated ion could increase the conductivity, thus leading to the emergence of channelrhodopsins, which act as light-gated channels for various ions [425, 427].

In parallel with the gains and losses of ion-coordinating residues in MRs, the loss and gain of the retinal-binding residue likely shaped some of the diversity of GPCRs. Such precedent exists, as the loss of the retinal-binding Lys residue was described in some MR lineages [428, 429]. In the ancestral Na-pumping rhodopsin, the loss of the retinal moiety would have left an empty pocket in its place, thus structurally compromising the protein. However, this protein could be stabilized by forming a full-fledged permanent $Na^+$-binding site out of the residues, that coordinated the passing through ions. Newly emerged $Na^+$-coordinating residues in GPCRs, e.g. $Asp^{2.50}$, could thus functionally replace the retinal (see Figure 4.2.1). Without retinal, the Trp residue of the helix 6/F would be able to reach the $Na^+$ ion in the new site and form an additional coordination bond via a water molecule. The remaining cavity could be filled by small organic molecules, which would push down in the Trp6.48, assuming the role of the retinal and triggering the helix 6/F movement. These changes lead to the emergence of $Na^+$-binding heptahelical receptors for small molecules, which are likely to be the ancestors of many, if not all, modern GPCRs (see Figure 4.2.6).

Finally, the ability to bind retinal and utilize its isomerization to detect light could have been re-acquired by the animal rhodopsins. This is evident from the overwhelming differences between the animal visual rhodopsins (class A GPCRs) and MRs. First, the lysine residues, binding the retinal moiety are located in the helix 7/G in both types of rhodopsins, but they are located at different depths within the membrane and their orientation relative to the membrane plane is also different. Second, the light cycles and the isomers of the retinal, observed at each stage also differ between animal and microbial rhodopsins. Specifically, in animal rhodopsins, 11-cis retinal after absorbing light isomerizes into an all-trans form, which has to leave the protein and be regenerated into 11-cis retinal in the cytoplasm [430, 431]. In microbial rhodopsins, the light is absorbed by an all-trans retinal molecule, which isomerizes into 13-cis form, which, in turn, relaxes into the original all-trans form without leaving the protein [432, 433]. Finally, the overall sequence and structure similarity between MRs and animal visual rhodopsins is much lower, than the similarity between MRs and other class A GPCRs (Figure 4.2.3). This divergence can be explained by the protein, first, losing the ancestral, bacterial type retinal, and then acquiring a new, animal type retinal, thus being forced to re-adapt its structure to accommodate for the new type of retinal binding and isomerization. Thus, while insisting on the common origin of GPCRs and MRs, this study still supports the emergence of retinal-based light sensors in the two (super)families of 7TM proteins as a result of the convergent evolution.

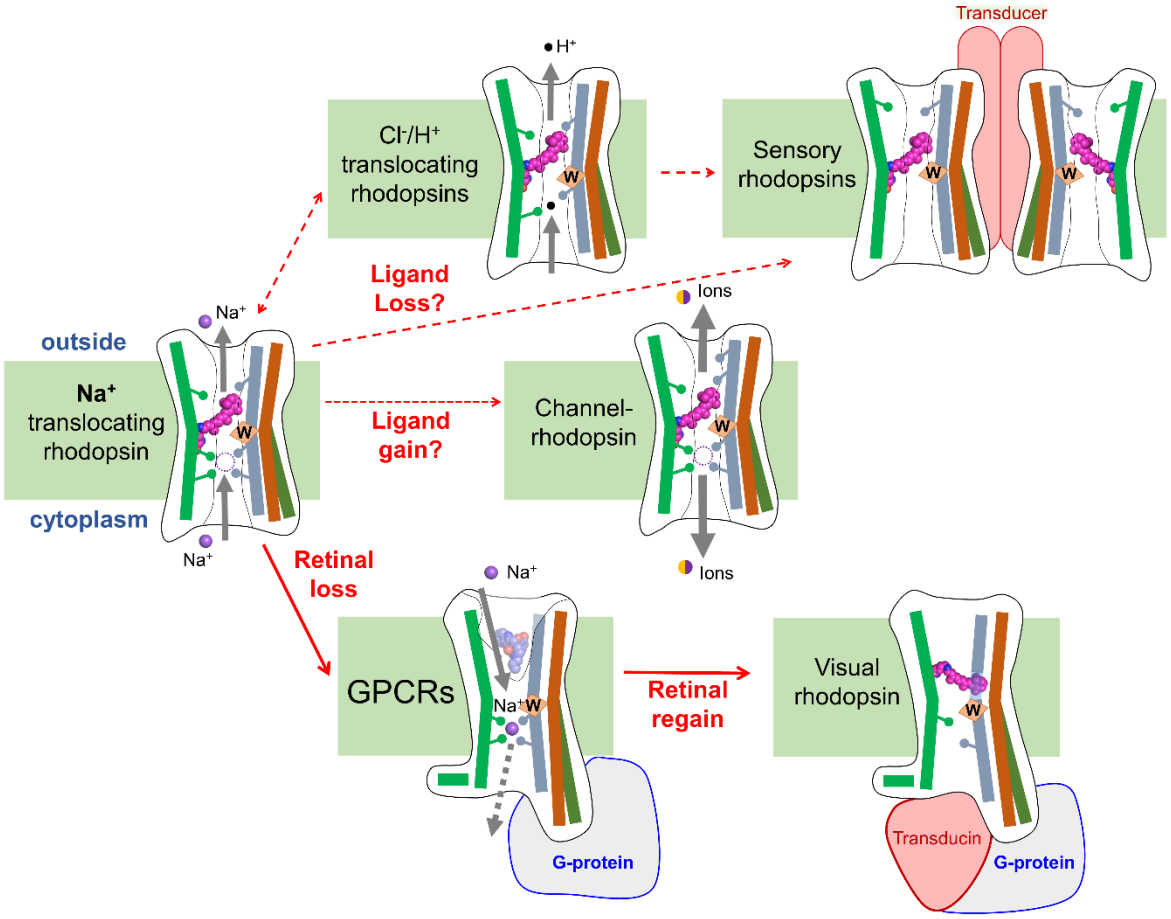


**Figure 4.2.6.** Suggested scheme of MRs and GPCRs evolution. Only three helices of the proteins are shown. Helix 3/C is shown in blue, helix 6/F is shown in brown ("closed" conformation) and orange ("open" conformation), helix 7/G is shown in green. The scheme depicts the proposed order of $Na^+$-ligands and retinal binding losses and gains leading to the emergence of particular functions of the proteins in the course of evolution.

### *4.2.3. Modeling of Class A GPCR activation*

A model of GPCR activation was developed in a manner similar to earlier approaches to GPCRs modeling [381, 434-436]. Additionally, the model includes the possibility of $Na^+$ translocation coupled to the transmembrane potential during receptor activation, similar to the processes occurring in membrane energy-converting proteins [437].

Here, the solution of the model of GPCR activation, as obtained in collaboration with Dr. D.A. Cherepanov of the Russian Academy of Sciences, is presented (Figure 4.2.7A, B). The model describes three possible binary transitions of the receptor: (i) between inactive and active states, (ii) between $Na^+$-bound and $Na^+$-free states, and (iii)between agonist-bound and agonist-free states. These transitions are defined within the model by their equilibrium constants: L - receptor activation constant, M - $Na^+$ association constant, N - agonist association constant. Together all three transitions can be displayed as a cubic graph, where each vertex corresponds to one of the eight states (Figure 4.2.7B, Table 4.3). Additionally, the allosteric coefficients α, β, γ, and δ were used to describe the effect of the coupling between the transitions, i.e. to account for their interdependence. The model was built in two versions to reflect the two possible operation modes of the receptor: carrier-on mode, in which the $Na^+$ ion is translocated into the cytoplasm, and carrier-off mode, in which $Na^+$ ion is returned into the extracellular space.

$Na^+$ ion translocation is described with the probabilistic Boltzmann-like electrostatic terms with $\Phi_1$ and $\Phi_2$ , which are equal to

$$\Phi_1=\exp(\theta \cdot e \cdot \Delta\psi/k_BT) \quad (4.1)$$

and

$$\Phi_2=\exp((\theta-1) \cdot e \cdot \Delta\psi/k_BT) \quad (4.2)$$

where $\theta$ is the coefficient reflecting the relative depth of the $Na^+$ ion in the membrane, with $\theta = 0$ when $Na^+$ ion sits on the extracellular surface of the membrane and $\theta = 1$ when $Na^+$ ion sits on the cytoplasmic surface membrane; $e$ is the elementary charge; $\Delta\psi$ is the transmembrane potential, $k_B$ is the Boltzmann constant, and $T$ is the temperature.

In total, the model describes 12 transitions, each corresponding to an edge of the cube shown in Figure 4.2.7B. The slowest transitions in the system are assumed to be the transitions of the receptor between R (inactive) and R* (active) states.

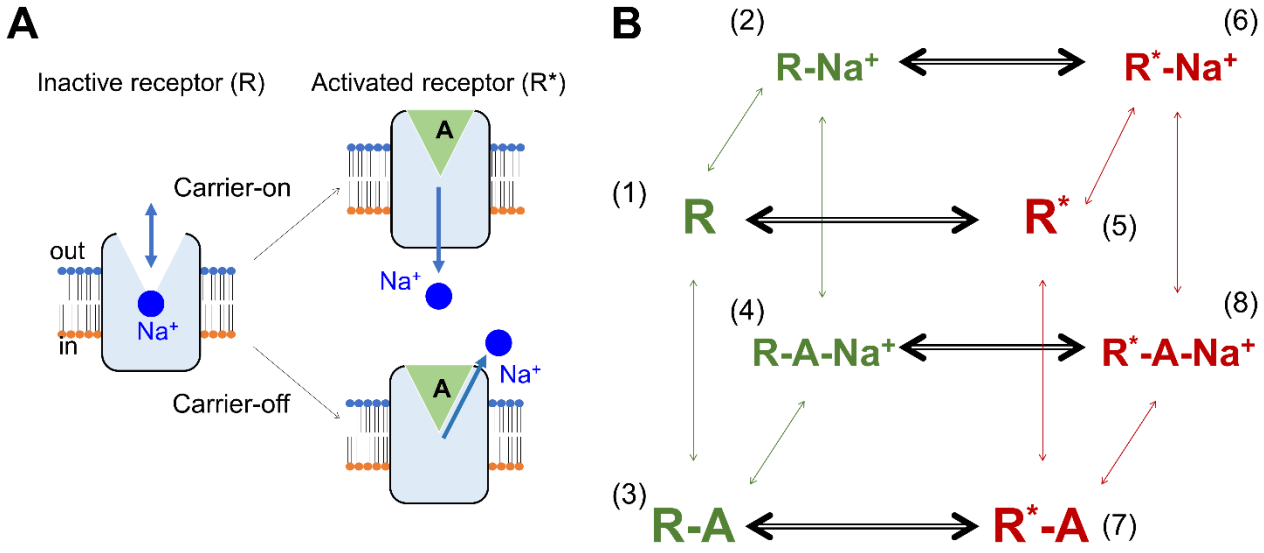


**Figure 4.2.7.** Model describing the $Na^+$ behavior upon GPCR activation.
A. The scheme of the two modes of GPCR activation. In both cases, the $Na^+$ ion binds to the GPCR in its inactive state, but not in its active state. When agonist binding triggers the activation of the receptor, the $Na^+$ ion must leave its binding site. In the carrier-on mode, the $Na^+$ ion is expelled into the cytoplasm, thus traversing the membrane. In the carrier-off mode, the $Na^+$ ion relocates to the extracellular space. B. Model of GPCR activation as a facultative $Na^+$-transporter. Inactive states (R) are shown in green, active states (R*) are shown in red, transitions between R and R* states are shown as thick black arrows. Transition probabilities are defined in Table 4.3.

In the carrier-off mode, the receptor is presumed to be in thermodynamic equilibrium, meaning that all transitions are equal so that the principle of detailed balance can be applied to all active and inactive receptor states. Probabilities $P_1$, …, $P_8$ are determined by the corresponding equilibrium constants (see Table 4.3, right column) and thus can be calculated directly using the normalization requirement:

$$\sum_{i=1..8} P_i = 1$$

In the carrier-on mode of operation, the principle of detailed balance can only be applied separately to inactive (green arrows in Figure 4.2.7B) and active (red arrows in Figure 4.2.7B) states of the receptor, while the transitions from active to inactive states and back (black arrows in Figure 4.2.7B) must be treated as nonequilibrium transitions. Thus, only eight rate constants have been considered explicitly. These include $k_{nm}$ - the rate constants of the four forward transitions from inactive states $n$ to active states $m$; and $k_{mn}$ - the rate constants of the respective backward transitions. All the other transitions were presumed to remain in thermodynamic equilibrium and thus are determined by corresponding equilibrium constants (see Table 4.3, left column). However, because the activation of the receptor is much slower than any other transitions in the model, the detailed balance principle can be applied separately to all inactive and all active receptor states (the left-hand and right-hand sides of the cubic diagram in Figure 4.2.7B, respectively).

The probabilities $P_1$, …, $P_8$ were determined as follows. The cumulative R→R* and R*→R transition probabilities match each other:

$$k_{15}P_1 + k_{26}P_2 + k_{37}P_3 + k_{48}P_4 = k_{51}P_5 + k_{62}P_6 + k_{73}P_7 + k_{84}P_8, \qquad (4.3)$$

where $P_{1\text{-}4}$ are the probabilities of the inactive states, $P_{5\text{-}8}$ – of the active states, and $k_{nm}$ and $k_{mn}$ are corresponding transition rate constants. While the transition rate constants are not fully independent, but they do satisfy the thermodynamic equation $k_{forward}/k_{back} = e^{-\Delta G/RT}$, so Eq. 4.3 can be rewritten thusly:

$$k_{15}(P_1 + \alpha^{1/2}P_2 + \beta^{1/2}P_3 + (\alpha\beta\delta)^{1/2}P_4) = k_{51}(P_5 + \alpha^{-1/2}P_6 + \beta^{-1/2}P_7 + (\alpha\beta\delta)^{-1/2}P_8) \qquad (4.4)$$

The membrane potential affects the probability of certain transitions in both carrier-on and carrier-off modes. Specifically, the cation translocation is affected as the equilibrium constants of sodium binding are changed in both inactive and active states ($\Phi_1$ and $\Phi_2$, respectively). This leads to the following equation:

$$P_5 = \xi \cdot L \cdot P_1,$$

where

$$L = k_{15}/k_{51}$$

and

$$\xi = \frac{1 + \alpha^{1/2}\Phi_1 X_{out} + \beta^{1/2}Y + (\alpha\beta\delta)^{1/2}\Phi_1\gamma X_{out}Y}{1 + \alpha^{1/2}\Phi_2 X_{in} + \beta^{1/2}Y + (\alpha\beta\delta)^{1/2}\Phi_2\gamma X_{in}Y}, \qquad (4.5)$$

where $X_{in}$=M[$Na^+$]$_{in}$, $X_{out}$=M[$Na^+$]$_{out}$, and Y=N[agonist], respectively. Finally, the probabilities normalization requirement is be added to these equations:

$$\sum_{i=1..8} P_i = 1.$$

The resulting set of equations provided a solution to the model.

**Table 4.3.** Probability coefficients for the model of GPCR activation

| Receptor state | N. | Relative probabilities | |
|---|---|---|---|
| | | carrier-on mode | carrier-off mode |
| Inactive | 1 | 1 | 1 |
| Inactive with a $Na^+$ ion | 2 | $\Phi_1 \cdot M[Na^+]^{out}$ | $\Phi_1 \cdot M[Na^+]^{out}$ |
| Inactive with an agonist | 3 | $N \cdot [A]$ | $N \cdot [A]$ |
| Inactive with an agonist and a $Na^+$ ion | 4 | $\gamma \cdot N[A] \cdot \Phi_1 \cdot M[Na^+]^{out}$ | $\gamma \cdot N[A] \cdot \Phi_1 \cdot M[Na^+]^{out}$ |
| Active | 5 | L | L |
| Active with a $Na^+$ ion | 6 | $\alpha \cdot L \cdot \Phi_2 \cdot M[Na^+]^{in}$ | $\alpha \cdot L \cdot \Phi_1 \cdot M[Na^+]^{out}$ |
| Active with an agonist | 7 | $\beta \cdot L \cdot N[A]$ | $\beta \cdot L \cdot N[A]$ |
| Active with an agonist and a $Na^+$ ion | 8 | $\delta(\gamma(L \cdot \beta N[A] \cdot \alpha\Phi_2 M[Na^+]^{in})) = \alpha\beta\gamma\delta \cdot L \cdot N[A] \cdot \Phi_2 M[Na^+]^{in}$ | $\delta(\gamma(L \cdot \beta N[A] \cdot \alpha\Phi_1 M[Na^+]^{out})) = \alpha\beta\gamma\delta \cdot L \cdot N[A] \cdot \Phi_1 M[Na^+]^{out}$ |

Each receptor molecule exists in a steady-state balance between the active (R*) and inactive (R) states; an agonist can shift the distribution in favor of the active states, while $Na^+$ stabilizes the inactive state (Figure 4.2.7A). Both active and inactive states are known to exist as an ensemble of quickly exchanging conformational sub-states [319, 438-440]. For simplicity, these sub-states were not included in the model. Additionally, the interaction of the receptor with other components, such as G-protein or arrestin, was not included in the model as well.

Without any agonists and $Na^+$ ions present, the proportion of the receptors in the active state is determined only by the activation constant L. In principle, the value of L should be about 1; if not, the receptor would be stuck in one of the two states. However, the exact value of L is unknown; its experimental determination would require estimating the proportion of receptors in the two states in the absence of agonists and $Na^+$ ions, such conditions are difficult to create. In this study, the value of L was set to one.

The initial values of the association constants for the agonist and sodium ions were set to $N = 1 \times 10^7\ M^{-1}$ and $M = 1 \times 10^3\ M^{-1}$, respectively, to conform to the physiologically relevant concentrations of agonist and $Na^+$ ions (see [341, 382, 384] and references therein).

The allosteric coefficients α, β, γ, and δ model the coupling between the transitions and their interdependence (see Table 4.4). The coefficient α reflects the effect of sodium binding on the receptor activation. Based on experimental data, $Na^+$ ions inhibit receptor activation [43, 342-347], and thus this coefficient should be $\ll 1$. This way the receptors would exhibit very little activation in the presence of $Na^+$ ions, but still display some activity in their absence (since L=1),

and this reflects the behavior reported in experimental systems (see [43, 319, 341, 343, 345-347, 438].

The coefficient β reflects the effect of agonist binding on the receptor activation. Naturally, as agonist binding triggers the activation of the receptor, this value must be >>1. The coefficient γ reflects the effect of agonist binding on the $Na^+$ ion binding. Available crystal structures of GPCR with $Na^+$ ion in its binding site usually do not contain any agonist molecules, while corresponding structures with agonists bound do not have the space for the $Na^+$ ion [43]; thus, the value of γ should be <<1. Finally, the triple allosteric interaction coefficient δ reflects the coupling between all three transitions. Two more parameters affecting the probabilities of individual states are the concentrations of sodium in and outside of the cell. Here the physiological values of $[Na^+]_{out}$ = 140 mM and $[Na^+]_{in}$ = 10 mM were used (Table 4.4).

Both carrier-on and carrier-off models were implemented separately as functions in MATLAB R2017a [60].

Multiple studies have reported experimental data on voltage sensitivity in different GPCRs (see Table 4.1 and [374, 375] for reviews). The membrane voltage with negatively charged cytoplasmic side activated GPCRs in most cases. Although in few cases, e.g. in the muscarinic receptor $M_1$, the membrane voltage instead decreased the receptor activity [376, 378, 382]. In most studies, GPCR activation at different voltages was evaluated by following the downstream reactions, e.g., by tracking the conductance of the ion channels, regulated by the GPCR under investigation (see Table 4.1, [374, 441] and references therein). It is important to note, however, that such GPCR-regulated channels are themselves voltage-sensitive, which muddles the interpretation of such data. Thus, here the scope of experimental data on the voltage dependence of GPCR activation is limited to only the data obtained by using FRET-based biosensors [349, 382, 384]. Such sensors react to the movement of the helix 6, which is directly linked to the receptor activation [319, 323, 332-338]. Using this method, Rinne and colleagues reported that membrane voltage increases the sensitivity of $\alpha_{2A}$ adrenoreceptor to agonist norepinephrine [384]. Also, membrane voltage was shown to increase the sensitivity of muscarinic receptors $M_3$ and $M_5$ to acetylcholine and carbachol (full agonists), but in the muscarinic receptor $M_1$, the sensitivity to those agonists was decreased [382].

**Table 4.4.** Initial parameters of the model

| Parameter | Value | Description[a] |
|---|---|---|
| M | $10^3$ $M^{-1}$ | Association constant of the sodium ion (initial value) |
| N | $10^7$ $M^{-1}$ | Association constant of the agonist (initial value) |
| L | 1 | Receptor activation constant |
| α | $10^{-3}$ | Intrinsic efficacy of sodium: ratio of affinity of sodium for R* and R |
| β | $10^3$ | Intrinsic efficacy of the agonist: ratio of affinity of the agonist for R* and R |
| γ | $10^{-2}$ | Binding cooperativity between the sodium ion and the agonist: ratio of affinity of A for R-Na and R, or Na for R-A and R |
| δ | $10^2$ | Activation cooperativity between the sodium ion and the agonist: ratio of affinity of A for R*-Na and R-Na, or of Na for R*-A and R-A |
| $[Na^+]_{out}$ | 140 mM | $[Na^+]$ in the extracellular medium |
| $[Na^+]_{in}$ | 10 mM | $[Na^+]$ in the cell cytoplasm |

[a] The designations of the parameters are taken from ref. [435].

The reported here model was tested to see if it can describe the data reported by Rinne and colleagues [382, 384]. To perform the fitting, experimental data points were extracted from the respective publications. In each case, fits were performed separately for the carrier-on and carrier-off modes. For each of the two operation modes, two sets of data points (as obtained at two voltage values, e.g. -90 and +60 mV) were fitted. The two fit curves were calculated by using the same sets of parameters (see Table 4.4), with the only difference being the voltage values that were taken from the respective publications. Both curves were fitted at the same time by adjusting only two parameters, specifically, the $Na^+$ binding constant M and agonist binding constant N. Thus the M and N parameters were kept the same for both curves at different voltages. All other parameters were kept constant at values listed in Table 4.4. In addition, separate fits with $\beta = 100$ were performed (Table 4.5).

The fit curves are shown in Figure 4.2.8 and the fit parameters are listed in Table 4.5. In each case, one of the two operation modes produced a good fit to the experimental data, while the other mode did not. In the cases where the receptor sensitivity was increased by the membrane voltage, the carrier-on model fitted the experimental data. For the data in Figure 4.2.8C, where the

membrane voltage decreased receptor sensitivity, the carrier-off mode provided a good fit. Notably, varying the θ value did not lead to noticeable fit improvements. It is important to note, that the fitted agonist association constants N, shown in Table 4.5, do not correspond to the observable association constants ($K_{obs}$), but rather the intrinsic association constants ($K_a$), which can be much smaller [381, 442].

**Table 4.5.** Sodium ion and agonist binding constants as obtained from fitting experimental data

| Receptor – agonist pairing | Mode | β=1000 | | β=100 | |
|---|---|---|---|---|---|
| | | $Na^+$ association constant, $M^{-1}$ | Agonist association constant, $M^{-1}$ | $Na^+$ association constant, $M^{-1}$ | Agonist association constant, $M^{-1}$ |
| $\alpha_{2A}$-AR with norepinephrine | 1, carrier-on | $2.4 \times 10^4$ | $2.6 \times 10^5$ | $8.7 \times 10^3$ | $1.1 \times 10^6$ |
| $M_3$ muscarinic receptor with carbachol | 1, carrier-on | $3.5 \times 10^5$ | 8.4 | $3.1 \times 10^5$ | 87.9 |
| $M_1$ muscarinic receptor with carbachol | 2, carrier-off | $2.1 \times 10^7$ | $6.4 \times 10^5$ | $2.2 \times 10^5$ | $2.6 \times 10^6$ |

Thus, reported here GPCR activation model was able to quantitatively fit experimental data on the membrane voltage effect on the activation of GPCRs while using the same set of parameters (Table 4.4) and only varying N and M (agonist and $Na^+$ binding constants) (Table 4.1). An example of such a fit is presented in Figure 4.2.9.

Results of the experimental data fitting do not indicate obligatory translocation of $Na^+$ ions into the cell upon GPCR activation. It appears, that while in some cases the translocation does occur, in other cases the $Na^+$ ion is expelled into the extracellular medium upon receptor activation. Still, the carrier-on mode model fits better to the activation curves of those receptor-agonist pairs, where membrane voltage increases the sensitivity. This type of membrane dependence seems to be prevalent, according to the experimental studies (Table 4.1).

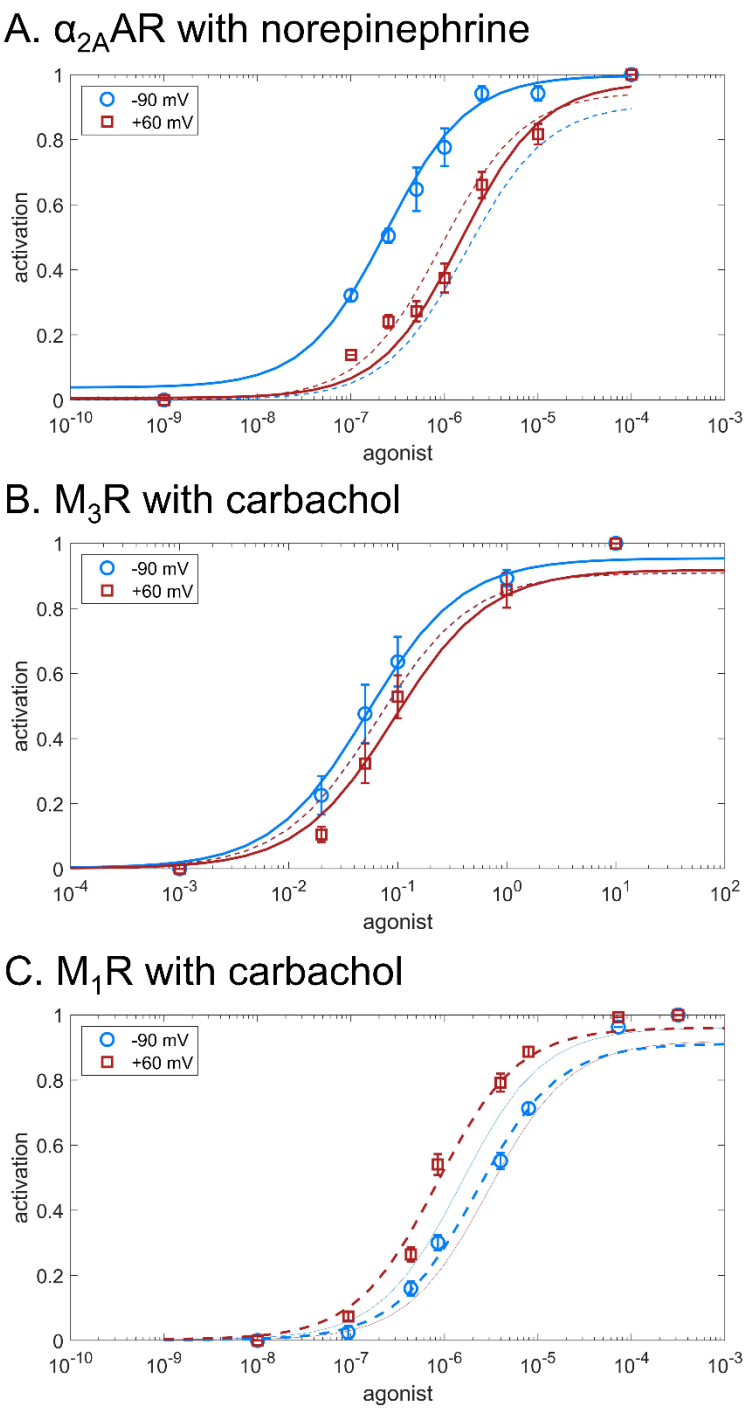


**Figure 4.2.8.** Experimental data on voltage sensitivity of GPCRs fitted with carrier-on and carrier-off models. Fits for the carrier-on model are shown as solid lines, fits for the carrier-off model are shown as dashed lines. A. Data on $\alpha_{2A}$ adrenoreceptor ($\alpha_{2A}$AR) activation by norepinephrine (full endogenous agonist) from [384]. B. Data on muscarinic $M_3$ receptor activation by carbachol (full agonist) from [382]. C. Data on muscarinic $M_1$ receptor activation by carbachol (full agonist) from [382].

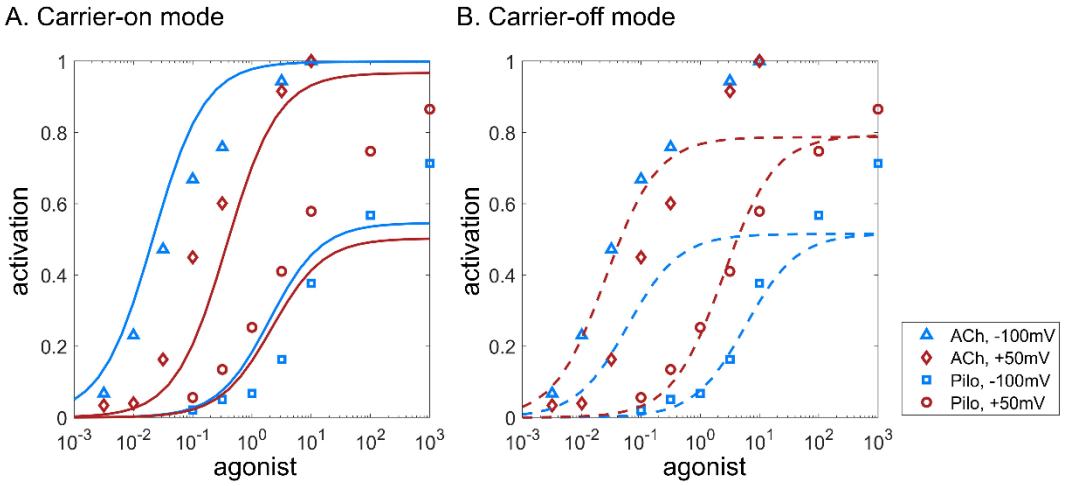


**Figure 4.2.9.** Experimental data on voltage sensitivity of muscarinic acetylcholine receptor $M_2$. Experimental data on M2 receptor activation by acetylcholine (Ach, full endogenous agonist) and Pilocarpine (Pilo, partial synthetic agonist) from [379]. A. Concentration-response curves fitted with the model in carrier-on mode. B. Concentration-response curves fitted with the model in carrier-off mode.

### *4.2.4. Suggested Mechanism of $Na^+$ Translocation Upon GPCR Activation*

Here, $Na^+$-dependent GPCRs are considered as putative electrogenic $Na^+$ carriers. Their ability to translocate a $Na^+$ ion is supported by the following data: (i) the results of MD simulations [336, 346, 347, 372, 373]; (ii) structural similarity and the evolutionary related ness of GPCRs and $Na^+$-pumping MRs (see chapters 4.2 and 4.3, and [443]); and (iii) the data on the "gating currents" or "sensory currents", exhibited in some GPCRs in response to an imposed membrane voltage (see [378, 444] and references therein), which may indicate the movement of a GPCR-bound $Na^+$ ion into the cytoplasm [441].

The primary role of the $Na^+$ ion in the GPCR seems to be the stabilization of the inactive state, thus helping to counteract these intrinsically unstable and noisy receptors, preventing signaling in the absence of agonists. However, such stabilization also imposes rigidity on the receptor, decreasing its sensitivity. To trigger the activation of such $Na^+$-stabilized receptor significantly higher concentrations of agonists would be required. This disadvantage can be negated by recruiting some source of free energy which would have to be coupled to the receptor activation. Since some GPCRs exhibit higher sensitivity under the effect of the transmembrane electric field (Figure 4.2.8A, B, and Figure 4.2.9), they evidently use the energy of this field. However, this energy can be only be used if the receptor activation is mechanically coupled to the translocation of the ion.

There are even more benefits from the activation-coupled translocation of the $Na^+$ ion into the cell. Because the equilibrium constants between the active and inactive states of GPCRs are low (Table 4.4), there could be multiple mechanisms that shift this equilibrium in response to the binding of signaling molecules to different sites within the agonist-binding pocket. Translocation of sodium ion upon activation can only be used for the signal amplification, if the agonist, when bound to the receptor, blocks the pathway for the sodium ion into the extracellular space, thus forcing it to escape into the cytoplasm [43]. In the inactive state, the sodium binding site is located at the bottom of the large agonist-binding pocket, which is open to the extracellular side (Figure 4.1.1A, B). The retreat of sodium ions into the extracellular space through this opening would be much easier, than traversing the receptor in the opposite past layers of hydrophobic residues. However, if the agonist binds in a way that blocks the escape of the sodium ion, the latter is forced to find its way into the cell. It appears, that such specific agonist binding does occur, albeit not in all receptor-agonist pairings, as voltage dependence patterns, indicating the coupling of the receptor activation with translocation of $Na^+$, are not universal among GPCRs (Table 4.1). Specifically, certain receptors show voltage-amplified signaling with some agonists, but others (Table 4.1). Thus, the signal of some agonist molecules, presumably the ones that effectively shield the $Na^+$-binding site from the extracellular space, is amplified by the electric field, as the receptor functions in the carrier-on mode. The signal of other molecules, e.g. partial agonists would be inhibited by the electric field, while the receptor functions in the carrier-off mode. Such a mechanism of discrimination between different agonists would dramatically increase receptor selectivity.

Experimental data suggest that endogenous agonists, which naturally occur as the primary signaling molecules for particular GPCR, induce the carrier-on mode of signaling and fully benefit from the membrane voltage. Specifically, in the muscarinic receptor $M_2$, the signal from acetylcholine (full endogenous agonist) is amplified by voltage, while the signal from pilocarpine (partial synthetic agonist) is decreased by voltage (see Figure 4.2.9, [379, 380], and Table 4.1). Notably, the voltage dependence of the $M_2$ receptor is affected by mutations of the orthosteric ligand-binding site residues, supporting a direct connection between the mechanics of the agonist binding and the voltage dependence [379].

Somewhat similarly, the voltage dependence of muscarinic acetylcholine receptor $M_3$ is exhibited differently with different agonists. Specifically, signaling triggered by acetylcholine (full endogenous agonist) and carbachol (full synthetic agonist) is increased by the voltage, while signaling trigged by choline or pilocarpine is decreased (see Table 4.1, Figure 4.2.8B and [382]). Rinne and coworkers suggested that voltage dependence is determined by the differences in the

binding of specific agonists. In support of this suggestion, they have shown that the mutation of Asn6.52 to Gln reversed the effect of the membrane potential on the carbachol-triggered signaling, but had no effect on the acetylcholine signaling [382]. Rinne and coworkers concluded that this data reinforces the important role of the helix 6 in the activation mechanics of acetylcholine receptors.

Another example is the dopamine $D_{2S}$ receptor. Its sensitivity to dopamine (endogenous full agonist) is increased by the voltage, but the sensitivity to agonists β-phenethylamine and *p*- and *m*-tyramine (synthetic) is decreased (see [387, 388] and Table 4.1 for further references). However, sensitivity to endogenous agonists being increased by voltage is not universal. For example, signaling of the muscarinic receptor $M_1$ is decreased by the membrane voltage both for acetylcholine (endogenous agonist) and carbachol (synthetic agonist) (see Figure 4.2.8C and Table 4.1). Such behavior can be described by the carrier-off mode model, where the $Na^+$ ion is expelled into the extracellular space upon receptor activation. It would appear, that in the $M_1$ receptor $Na^+$ is not translocated, independent of which agonist triggers the receptor activation.

Finally, several receptors have been reported to be insensitive to the membrane voltage (see Table 4.1). Within the suggested model of GPCR activation, this can be explained by the lack of coupling between $Na^+$ binding/exit from the receptor and the receptor activation ($\delta \ll 1.0$). This can occur, for example, if sodium ion leaves the receptor before agonist binding triggers any major conformational changes in the protein.

The effect of the membrane voltage on GPCR activation is contingent on the coupling between the protein activation and electrogenic translocation of $Na^+$ ion across the membrane. Maintaining a sodium ion in the thick of the membrane demands a lot of energy due to the high desolvation penalty for a cation [445]. Thus, every residue in such membrane $Na^+$ binding site is crucial. Particularly, in $Na^+$-dependent ATP synthases, $Na^+$ binding is granted by six amino acids [417, 446]; after losing even a single one of them the protein is no longer able to translocate $Na^+$ and is only capable of proton transport [417, 446, 447]. Since Trp6.48 is involved both in $Na^+$ binding and activation-associated conformational changes in GPCR, this residue is the prime suspect for coupling activation with translocation of sodium. This residue is conserved in most GPCRs, particularly of class A (Table 4.6) and changes its sidechain conformation upon agonist binding (see Figure 4.1.1C and [319, 323, 332-338]). The coordinational bond between $Na^+$ ion and a water molecule, that forms an H-bond with Trp6.48 is present in structures of several GPCRs, e.g. δ-opioid receptor (PDB ID 4N6H), $A_{2A}$ adenosine receptor (PDB ID 4EIY), and $\beta_1$-adrenoceptor (PDB ID 5A8E). Most likely, Trp6.48 is similarly involved in this interaction network in other GPCRs of class A as well. As the bond network around $Na^+$ ion is destabilized

upon agonist binding and Trp6.48 movement, the further stay of the ion in the think of the membrane becomes increasingly unlikely, leading to the $Na^+$ ion leaving the protein and thus coupling its movement with the receptor activation. Thus, whether GPCR activation is accompanied by translocation of $Na^+$ across the membrane, or by the expulsion of the sodium ion into the extracellular space, this process would be affected by the membrane voltage, albeit in opposite directions.

To identify other elements of the structural mechanism of $Na^+$ translocation and its path through the receptor, structures of the muscarinic receptor $M_2$ in its inactive and active states (PDB ID 3UON [328] and PDB ID 4MQT [350], respectively) were compared (Figure 4.2.10). These structures were chosen as they clearly show the differences between the two states [350, 448] and the voltage dependence of the $M_2$ receptor suggests that it can operate in the carrier-on mode (see Figure 4.2.8, Table 4.1 and [378, 379, 448, 449]). Due to the high sequence and structure similarity among class A GPCRs, particularly in the $Na^+$-binding region ([43], Table 4.6), the results of this analysis should apply to other receptors as well. Specifically, several motifs are universally conserved in class A GPCRs: LxxxD in helix 2, DRY in helix 3, WxP in helix 6, and NPxxY in helix 7 [318, 450-452]. Relative positions of this motif in the structure are shown in Figure 4.2.10B, and the extent of their conservation is reported in Table 4.6.

Structures of $M_2$ receptor in active and inactive states show the difference in the helix 6 positions starting at the Trp400$^{6.48}$ residue. A notable difference can also be observed in the position of the Leu65$^{2.46}$ of the LxxxD motif and the NPxxY motif residues (Figure 4.2.10). In the inactive state structure, a large agonist-binding cavity above the $Na^+$-binding site is open to the extracellular space, while the cavity on the cytoplasmic site reaches only up to the NPxxY motif. Two layers of hydrophobic residues block the passage between these two cavities (see yellow and orange residues on Figure 4.2.10B and [336, 347, 372, 373, 453, 454]. In Figure 4.2.10B, the first hydrophobic layer contains Leu65$^{2.46}$, Leu114$^{3.43}$, Val111$^{3.40}$, Ile117$^{3.46}$, Ile392$^{6.40}$, Leu393$^{6.41}$, and Phe396$^{6.44}$ (Figure 4.2.10). These residues are primarily located in helices 2, 3, and 6 and were previously suggested to be involved in the conformational changes upon receptor activation [336, 347, 372, 373, 453, 454]. Furthermore, the majority of these residues are very conserved among class A GPCRs (Table 4.6). This first hydrophobic barrier is structurally supported by residues of the second hydrophobic shell: Val44$^{1.53}$, Ile62$^{2.43}$, Trp148$^{4.50}$, Pro198$^{5.50}$, Met202$^{5.54}$, Tyr206$^{5.58}$, and Ile389$^{6.37}$. While these second shell residues are not as conserved, as the residues from the first layer, their positions are still taken mostly by hydrophobic amino acids in the class A GPCRs (Table 4.6).

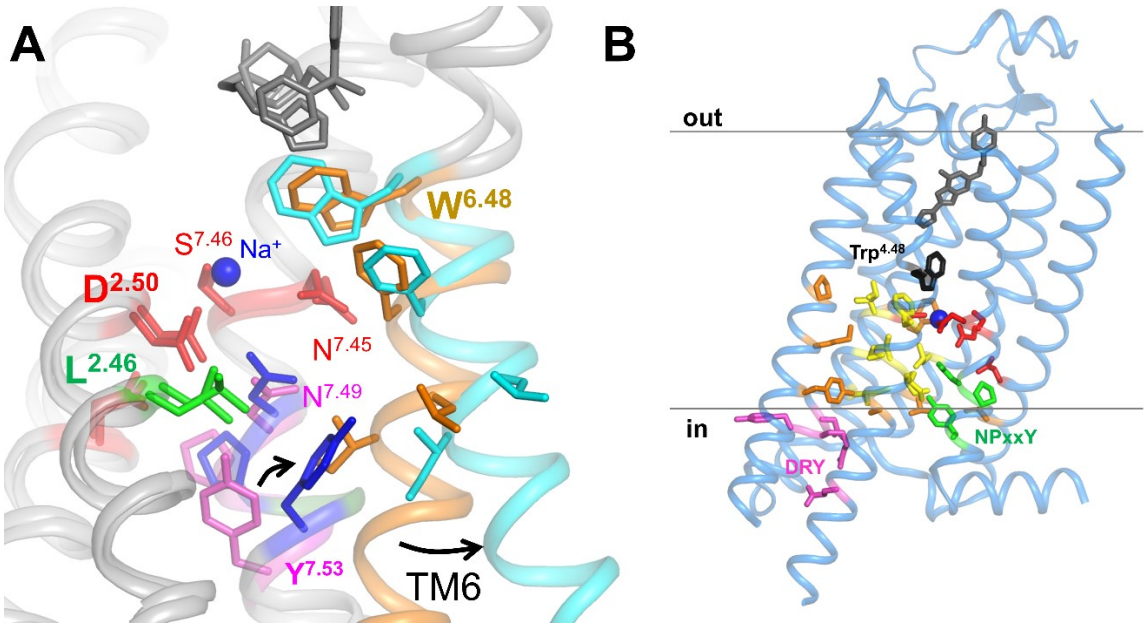


**Figure 4.2.10.** Conserved motifs in active and inactive state structures of GPCRs. A. Structure superposition of $M_2$ receptor, the structures shown are PDB ID 3UON (inactive state [328]) and PDB ID 4MQT (active state [350]). Helix 6 is shown in orange in the inactive state and light blue in the active state; the NPxxY motif is shown in purple in the inactive state and blue in the active state; residues, involved in the coordination of $Na^+$ ion are shown in red; agonist LY2119620 and antagonist 3-quinuclidinyl-benzilate molecules are shown in grey. B. Location of hydrophobic residue clusters and conserved motifs of class A GPCRs. The structure of adenosine $A_{2A}$ receptor is shown (PDB ID 5IU4 [455]). Residue Trp6.48 is shown in black; hydrophobic residue clusters are shown in yellow (first layer) and orange (second layer); $Na^+$-binding residues are shown in red; NPxxY motif – in green; DRY motif - in pink.

The first hydrophobic layer is centered on Leu65$^{2.46}$ of the LxxxD motif (Figure 4.2.10B). Notably, the sidechain of this residue attains different rotamers in active and inactive states (Figure 4.2.10A). A minor rotameric transition of this strictly conserved residue can potentially open a passage for the $Na^+$ ion into the cell (Figure 4.2.11). The LxxxD motif is highly conserved and likely plays a central role in the $Na^+$ translocation. While Asp2.50 directly coordinates $Na^+$ ion, Leu2.46 contributes to the interface between helices 2 and 3.

Since the ability to bind $Na^+$ ions was likely inherited by GPCRs from an ancient $Na^+$-translocating microbial rhodopsin (see Chapter 4.2.2 and [443]), the mechanism and pathway of cation translocation may still hold some resemblance between GPCRs and MRs. In the $Na^+$-translocating rhodopsin, the pathway for the $Na^+$ ion is formed by helices C/3, F/6, and G/7 [369, 370, 443, 456]. Corresponding helices have been implicated in $Na^+$ binding and possibly translocation in GPCRs as well [43, 336, 347, 443].

Recently, a structure of Na-translocating rhodopsin KR2 with $Na^+$ ion trapped within the protein was resolved (PDB ID 6XYT [457]). In this structure the $Na^+$ ion is trapped between Asn110 and Asp114 residues of the characteristic NDQ motif, apparently in one of the transient

binding sites of the ion. The Asn110 and Asp114 residues of KR2 correspond to Asp3.32 and Asn3.35 of GPCRs (Figure 4.2.1). The structure of δ-opioid receptor, used in this work as the most similar to MRs, depicts a water molecule bound between these two residues, in the position similar to that of the $Na^+$ ion trapped in the novel KR2 structure (Figure 4.2.1). Unfortunately, Asp3.32 and Asn3.35 bridged by a water molecule face the extracellular cavity that precedes the $Na^+$-binding site of GPCRs, so that these new structural data do not help in determining the pathway of $Na^+$ translocation from the $Na^+$-binding site into the cytoplasm by GPCRs.

Based on the structure analysis of the active and inactive states of muscarinic $M_2$ receptor (Figure 4.2.10) and the previously discussed similarities between GPCRs and $Na^+$-translocating microbial rhodopsin KR2 (Chapters 4.2.1 and 4.2.2, [443]) the following pathway of $Na^+$ translocation by class A GPCRs can be suggested. As an agonist binds and blocks the retreat of the sodium ion into the extracellular space, the cation remains in its binding pocket, albeit in an unstable state due to the displacement of one of its ligands – Trp6.48. Electric field pushes the $Na^+$ ion towards the cytoplasm, forcing the rotameric transition of Leu2.46. This transition is accompanied by the breakage of the coordination bond between Asp2.50 and the $Na^+$ ion. As the $Na^+$ ion escapes, residues of the NPxxY motif undergo conformational changes, blocking the $Na^+$ ion from returning to its binding site. Specifically, Tyr7.53 residue forms a new Hbond via a water molecule with Tyr5.58, while the Asn7.49 residue turns towards Asp2.50 (Figure 4.2.10A, [335-337]). These new hydrogen bonds stabilize the active state of the receptor, which is achieved additionally by displacement of several helices, the collapse of the $Na^+$-binding pocket, and closing of the cytoplasmic $Na^+$ passageway (Fig. 1B). The suggested mechanical coupling between residues Leu2.46, Asp2.50, and Tyr7.53 and their involvement in the $Na^+$ ion translocation is supported by the results of MD simulations reported by Filizola and colleagues [347].

The $Na^+$ ion could escape from the protein either by reaching the water phase directly from the cytoplasmic side, passing straight between the helices, or even earlier, by sliding between helices 2 and 3 near the DRY motif (Figure 4.2.11). In the latter scenario, the $Na^+$ ion would be released into the phosphate group layer of the membrane lipids. Negatively charged phosphates can serve as potent $Na^+$ ligands, thus attracting the $Na^+$ ion and helping it pass between the protein helices. This is where that path of the $Na^+$ ion could safely end. MD simulations of membranes have shown that $Na^+$ ions commonly reside among the phosphate groups of membrane phospholipids and compensate for their negative charges [458].

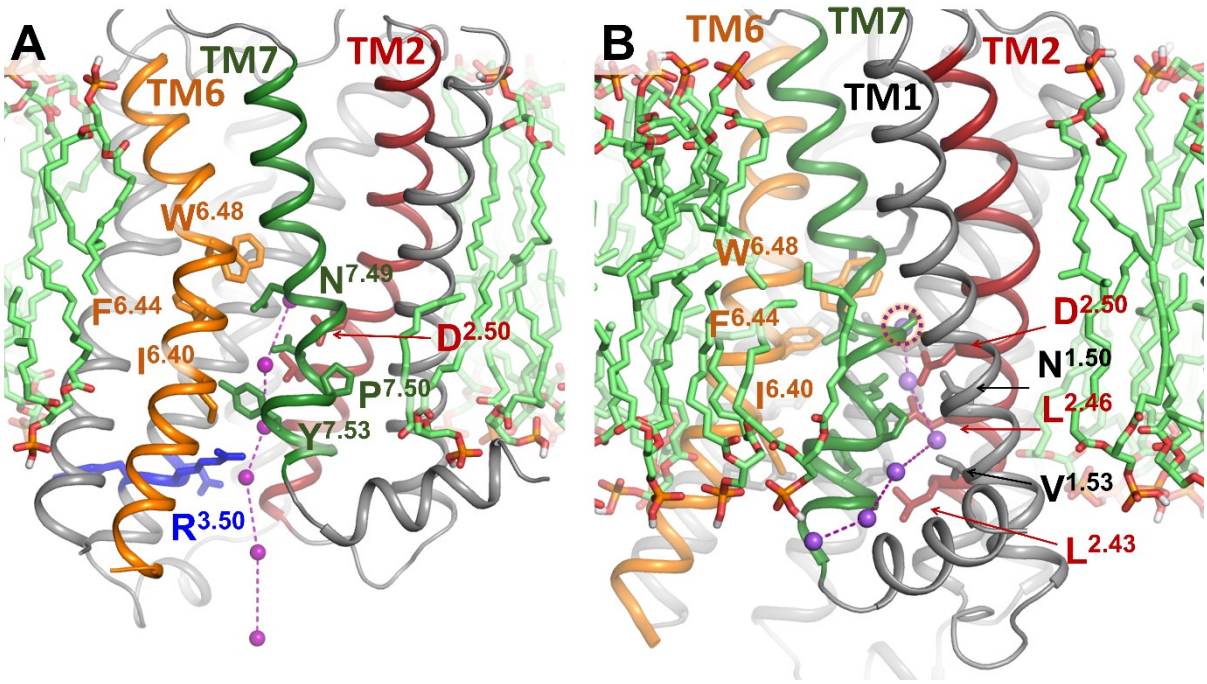


**Figure 4.2.11.** Two proposed pathways of $Na^+$ translocation to the cytoplasm upon activation of class A GPCRs, shown in the example of the $M_2$ receptor (PDB ID 4MQT). A. $Na^+$ ion exits into the cytoplasm through the center for the heptahelical bundle. B. $Na^+$ ion exits through the opening between helices 1, 2, and 7. Helices are colored as follows: 1 – gray, 2 - red, 6 - orange, 7 – green. Residues DRY of the ionic lock are shown in blue. Lipid molecules of the membrane around the protein were constructed with CHARMM-GUI software [77].

Suggested here mechanisms of facultative $Na^+$ translocation by class A GPCRs open a new aspect of structure-function relationships in this undoubtedly important protein (super)family. Reported here activation model was able to describe experimental data on voltage-dependent activation in receptors that exhibit increased, as well as decreased signaling at higher membrane voltages. This model can be useful in the future in interpreting the new experimental data and could be instrumental in identifying the mode of operation for GPCR activation triggered by novel agonists. Structure analysis reported here suggests, that the specific pattern of agonist binding, which blocks the access for the $Na^+$-binding site to the extracellular space, is crucial in determining the carrier-on mode of operation, allowing the receptor to increase its sensitivity by utilizing the energy of the membrane potential. This mechanism can be exploited in the design of novel agonists, which can be made specifically to trigger carrier-on or carrier-off mode of receptor activation. Finally, provided here evidence of the ability of the GPCRs to function as $Na^+$ transporters further supports their common origin with MRs from ancient $Na^+$ pumping rhodopsin. With multiple new structures of 7TM proteins being released every year, more common mechanisms, uniting these two groups of proteins, could be discovered.

**Table 4.6.** Conservation of functionally important residues in Class A GPCRs

| Residue in $M_2$ receptor (PDB: 3UON) | Generic number | Most common residue | | Second most common residue | | Residue type, % |
|---|---|---|---|---|---|---|
| | | AA | % | AA | % | |
| | $Na^+$ coordination | | | | | |
| Asn41 | 1.50 | N | 98 | S | 1 | Polar, 100 |
| Asp69 | 2.50 | D | 92 | N | 3 | Polar, 98 |
| Ser110 | 3.39 | S | 72 | T | 8 | Polar, 83 |
| Trp400 | 6.48 | W | 68 | F | 16 | Aromatic, 87 |
| Asn432 | 7.45 | N | 67 | S | 11 | Polar, 93 |
| Ser433 | 7.46 | S | 64 | C | 13 | Polar, 72 |
| | (C)WxP motif | | | | | |
| Thr399 | 6.47 | C | 71 | S | 10 | Small, 86 |
| Trp400 | 6.48 | W | 68 | F | 16 | Aromatic, 87 |
| Pro402 | 6.50 | P | 99 | N/A | N/A | Helix kink, 99 |
| | The hydrophobic shell around the $Na^+$ pocket | | | | | |
| Leu65 | 2.46 | L | 90 | M | 4 | Hydrophobic, 99 |
| Val111 | 3.40 | I | 40 | V | 24 | Hydrophobic, 88 |
| Leu114 | 3.43 | L | 73 | I | 10 | Hydrophobic, 98 |
| Ile117 | 3.46 | I | 56 | L | 16 | Hydrophobic, 99 |
| Ile392 | 6.40 | V | 37 | I | 28 | Hydrophobic, 93 |
| Leu393 | 6.41 | V | 41 | L | 20 | Hydrophobic, 91 |
| Phe396 | 6.44 | F | 75 | V | 4 | Hydrophobic, 92 |
| | The second hydrophobic shell | | | | | |
| Val44 | 1.53 | V | 65 | A | 14 | Hydrophobic, 92 |
| Ile62 | 2.43 | L | 36 | I | 35 | Hydrophobic, 97 |
| Trp148 | 4.50 | W | 96 | F | 1 | Hydrophobic, 99 |
| Pro198 | 5.50 | P | 79 | V | 5 | Hydrophobic, 95 |
| Met202 | 5.54 | I | 33 | M | 30 | Hydrophobic, 90 |
| Tyr206 | 5.58 | Y | 72 | S | 5 | Hydrophobic, 86 |
| Ile389 | 6.37 | L | 38 | V | 21 | Hydrophobic, 91 |
| | NPxxY motif | | | | | |
| Asn436 | 7.49 | N | 72 | D | 20 | Polar, 98 |
| Pro437 | 7.50 | P | 94 | A | 2 | Hydrophobic, 98 |
| Tyr440 | 7.53 | Y | 89 | F | 4 | Aromatic,93 |
| | DRY motif (ionic lock) | | | | | |
| Asp120 | 3.49 | D | 64 | E | 21 | Polar, 97 |
| Arg121 | 3.50 | R | 95 | H | 1 | Polar, 98 |
| Tyr122 | 3.51 | Y | 66 | F | 10 | Hydrophobic, 87 |
| Glu382 | 6.30 | E | 26 | K | 15 | Polar, 79 |

Residue conservation is presented as obtained from the GPCRdb database [459] alignment of human Class A (Rhodopsin-like) GPCRs. Relative positions of the residues in the structure are shown in Figure 4.1.1.

# 5. Conclusions

1. Molecular dynamic simulations of ATP and GTP molecules in complex with $Mg^{2+}$ revealed the effect different monovalent cations have on the phosphate chain shape. The binding sites on the phosphate chain were shown to be similar for all studied cations. Furthermore, both binding sites are routinely occupied by positively charged moieties in experimental structures of P-loop NTPases. Specifically, the binding site between β- and γ-phosphates harbors the universally conserved Lys of the P-loop/Walker A motif. The binding site between α- and γ-phosphates harbors activating cations in the $M^+$-dependent NTPases or positively charged protein residues in non-$M^+$-dependent NTPases. Analysis of available experimental structures also revealed that the phosphate chain shape of P-loop bound NTP-like molecules is similar to the Mg-NTP conformations sampled from MD simulations with $K^+$ or $NH_4^+$ ions, longer than the triphosphate chain conformations sampled from the simulations with smaller $Na^+$ ions, and shorter than in the simulations in the absence of $K^+$/$Na^+$/$NH_4^+$ ions.

2. Molecular dynamic simulations of $K^+$-dependent tRNA modification GTPase MnmE showed that the activating $K^+$ ion links the oxygen atoms of α- and γ-phosphates, leading to the turn of the γ-phosphate. This brings the phosphate chain of the GTP to a catalytically productive near-eclipsed conformation. The γ-phosphate turn also results in a new hydrogen bond with the backbone nitrogen atom of the $Asn226^{K-3}$ residue, which coordinates $K^+$ ion.

3. MD simulations of translation factor EF-Tu in the presence of $K^+$ and $Mg^{2+}$ ions showed that the $K^+$ atom enters the catalytic site and occupies the same place as $K^+$ ions in $K^+$-dependent GTPases. The binding of a $K^+$ ion between α- and γ-phosphates lead to the rotation of γ-phosphate, the emergence of a new H-bond with the backbone amino group of $Asp21^{K-3,}$ and rearrangement of the H-bond network. These data support the suggestion that the activation of EF-Tu and other translation factors on the ribosome is $K^+$-dependent.

4. Building on the results of MD simulations, comparative analysis of representative structures of P-loop NTPases revealed the following steps of the common activation mechanism of these proteins:

i. Mg-NTP binding to the P-loop motif drives the triphosphate chain into an extended conformation, with β- and γ-phosphates in a near-eclipsed conformation;
ii. Cationic moiety (monovalent cation, Arg/Lys/Asn residue or backbone nitrogens of the characteristic motif) is inserted into the catalytic site and interacts with the γ-phosphate, which is rotated towards an almost eclipsed conformation;
iii. The new position of γ-phosphate is stabilized by an additional H-bond between the $O^{2G}$ oxygen atom and the backbone amino group of the K-3 amino acid of the P-loop motif.

5. Comparison of high-resolution structures of the sodium-translocating bacterial rhodopsin and various $Na^+$-binding GPCRs revealed striking similarities of their sodium-binding sites. This similarity was used to construct a structure-guided sequence alignment for the two (super)families, which highlighted their evolutionary relatedness. These results support a common origin and a shared underlying molecular mechanism for both families that involves a highly conserved aromatic residue playing a pivotal role in the rotation of the 6th transmembrane helix.

6. Modeling and structure analysis of Class A GPCRs suggests that $Na^+$-binding GPCRs translocate $Na^+$ ions upon their activation and exploit the energy of the transmembrane sodium potential to increase their sensitivity and selectivity.

# 6. Outlook

The cytoplasm of all living cells contains much more potassium ions than sodium ions. Modern cells maintain this gradient of ions using sophisticated pumps, energy-dependent proteins that are embedded in advanced ion-tight membranes. The first cells could not possess such structures, and could not sustain the difference in concentrations of small molecules and ions between protocell contents and their environment. Yet, many ubiquitous and, by extension, ancient proteins require potassium ions in high concentrations. Sodium ions cannot replace potassium in such systems, and sometimes even inhibit their activity. Thus, the distinct ionic composition of the cytoplasm and specific requirements of the ancient proteins have been attributed to the origin of the first cells in a potassium-rich environment, supposedly of anoxic geothermal fields [7].

Many of the ubiquitous proteins, attributed to the Last Universal Cellular Ancestor (LUCA) belong to the P-loop NTPase fold. Furthermore, the P-loop structural motif itself was suggested to be one of the peptide ancestors that preceded the modern diversity of domain folds [102]. Reported here MD simulations of ATP and GTP in water revealed that the phosphate chain shape in the presence of $K^+$ ions (or similar in size $NH_4^+$ ions) resembles the phosphate chain shape of the P-loop bound nucleotide. This consonance supports the emergence of the P-loop structural motif in a $K^+$-rich environment as well.

Since the primary function of the P-loop motif is merely binding of the phosphate chain in an extended conformation, it is easy to imagine how NTP hydrolysis at the early stages of cellular evolution was catalyzed by the free $K^+$ ions from the solution, while the short P-loop motif-containing peptides were responsible for binding and transportation of the NTP molecules. The next step in the development of modern NTPases was the emergence of $K^+$-binding sites, which allowed coupling of $K^+$ binding and NTP hydrolysis with specific interactions, ensuring that NTP hydrolysis occurs in a controlled manner. This mechanism is still employed in many modern NTPases that are involved in protein synthesis. Alongside tRNA modification GTPase MnmE and translation factor EF-Tu, covered in this work, this group of proteins includes most other translation factors and a vast variety of ribosome-associated NTPases.

The next step was the recruitment of Arg/Lys fingers and other activating moieties to replace the cations in the active site. The ample variety of activating moieties in the TRAFAC class even allows reconstruction of their evolution, revealing the independent emergence of different types of Arg fingers in different families [8]. Other classes of P-loop fold NTPases have more homogeneous activation mechanisms within each class, suggesting that the emergence of their corresponding activation mechanisms occurred concurrently with the acquisition of new

domains and expansion of the P-loop domain core. Still, even the most sophisticated of the P-loop machines employ the same basic activation mechanism, that is employed by small ancient $K^+$-dependent NTPases and was uncovered in this work.

Since the cytoplasm already in the very first protocells must have contained high concentrations of potassium, the invasion of the living organisms into more common habitats, such as marine or fresh waters, would require adaptations that would allow cells to maintain the concentration gradient on the membrane. Such adaptations were ion-tight membranes composed of complex lipids and transmembrane proteins, pumping ions across the membranes. Many families of proton pumps include subfamilies of sodium-pumping proteins. Some modern organisms have entirely sodium-based energetics, utilizing membrane proteins transporting sodium ions instead of protons.

In 1986, Vladimir Skulachev and his colleagues analyzed a variety of such organisms and proposed the term "Sodium World" to describe their common features [460-462]. Later, the Sodium World expanded with the discoveries of Na-pumping proteins in different families involved in membrane bioenergetics. In this work comparative analysis of modern $Na^+$-pumping microbial rhodopsins and eukaryotic G-protein-coupled receptors (GPCRs) reveals their common origin from an ancient light-driven sodium export pump. Such an ancient pump would be of great use to the first cells, not only in maintaining the vital concentration of $K^+$ but also in creating an ion concentration gradient on the membrane as a depot of consumable energy.

Reported here study of class A GPCRs reveals that these proteins retained not only the Na-binding site in the center of the transmembrane heptahelical bundle but also the ability to translocate this ion. Harvesting the energy of this translocation GPCRs can increase both their sensitivity and selectivity. These features underlie the effectiveness of GPCR as sensors, which explains the abundance and diversity of these receptors. The human genome alone contains 700 representatives of class A GPCRs and hundreds of other GPCR and GPCR-like coding genes. Thus, GPCRs, the largest protein family coded by the human genome, stem from the Sodium World, which encourages exploration of other Na-dependent enzymes of eukaryotes [9].

# 7. Summary

In this work the evolutionary biophysics approach is applied to the two of largest protein superfamilies present in human genomes, namely P-loop NTPases and G-protein coupled receptors (GPCRs). This approach combines comparative analysis of protein structures and sequences with molecular modeling techniques, in order to reveal not only the conservation of particular residues among proteins within each superfamily but also their role in the fundamental mechanisms underlying common functions.

The study of the hydrolysis activation mechanism in P-loop NTPases started with the molecular dynamics simulations of Mg-NTP complexes (Mg-ATP and Mg-GTP) in the presence of $K^+$, $NH_4^+$, and $Na^+$ ions. These simulations have shown that in the presence of large cations ($K^+$ and $NH_4^+$) the phosphate chain of ATP and GTP attains longer conformations, with large distances between α- and γ-phosphates. These conformations resemble the shape of ATP and GTP molecules and their analogs, as seen in the crystal structures of various P-loop NTPases. Comparison with the structures of cation-dependent P-loop NTPases revealed that cation binding sites, observed in MD simulations in water coincide with the location of positively charged moieties in the protein NTP-binding sites.

To clarify the role of monovalent cations in P-loop proteins, MD simulations were conducted for two cation-dependent GTPases: tRNA modification GTPase MnmE and translation factor EF-Tu. MD simulations of Mg-GTP/EF-Tu complex bound to the tRNA and ribosome fragment in the presence of $K^+$ ions have shown consistent binding of cations from the solution to the AG site, similar to the cation binding in MnmE and other cation-dependent P-loop GTPases. In both proteins binding of $K^+$ ion in the AG site lead to the rotation of γ-phosphate, which brought this group closer to the eclipsed state with α-phosphate. The new rotated position of the γ-phosphate was stabilized by a novel H-bond between the $O^{2G}$ oxygen atom of γ-phosphate and the backbone nitrogen of the K-3 residue (relative to the ubiquitously conserved Lys) of the P-loop motif.

The activation mechanism, observed in MD simulations of MnmE and EF-Tu could be envisioned as basic for P-loop NTPases, as these cation-dependent proteins are among the most ancient members of the P-loop superfamily. This mechanism was used as a basis for extensive comparative analysis of representative proteins from all major classes of P-loop NTPases. Based on the established conservation and presence of the key features in active sites of P-loop NTPases, the chain of events leading to the nucleophilic attack and γ-phosphate cleavage has been proposed as the basic universal activation mechanism of NTP hydrolysis in P-loop NTPases.

The first stage of the proposed mechanism involves binding of the NTP molecule to the phosphate chain and attaining an extended, hydrolysis-prone conformation. This conformation is secured by interactions with the signature Lys residue of the P-loop/Walker A motif, backbone atoms of the P-loop, and $Mg^{2+}$ ion, which is in its turn secured by residues of Walker A and B motifs. At the second stage, a stimulating moiety is introduced to the active site by a specific partner – another domain, adjacent subunit in a dimer or oligomer, or an entirely separate protein. P-loop NTPases from different classes and families exhibit a wide range of activating partners and interactions, as well as stimulating moieties. Still, a cationic stimulating moiety that interacts with either both α- and γ-phosphates, or only the γ-phosphate was identified in all protein families with available structures of active state complexes. Triggered by the introduction of the stimulating moiety the γ-phosphate is rotated towards an eclipsed conformation with α-phosphate and forms a novel hydrogen bond with the backbone nitrogen atom of K-3 residue of the P-loop motif. Finally, the catalytic water molecule is stabilized in its attacking position by auxiliary residues.

The second part of this work explores the activation of GPCRs as sodium-translocating receptors. Crystal structures of the novel Na-pumping microbial rhodopsin along with the recent avalanche of GPCR structures provided the basis for comparative structure analysis, focused on investigating the similarities in the Na-binding sites of the two superfamilies. Structure superposition of GPCRs and MRs based on comparison of their Na-binding sites was used to produce structure-guided sequence alignments of the two superfamilies. Non-opsin GPRCs, particularly those that have sodium binding sites, were shown to be more similar to MRs, than visual rhodopsin. The only residue, universally conserved between the two superfamilies was Trp in the helix 6/F (Trp6.48 in GPCRs). In both families, the signaling mechanism directly involves this residue, which is likely to be an ancient feature, inherited from the common ancestor of MRs and GPCRs –Na-pumping light-activated rhodopsin.

The similarity of GPCRs with light-activated sodium pumps endorses the suggestion, that GPCRs also function as $Na^+$ ion translocators. A model of GPCR activation accompanied by translocation of $Na^+$ was constructed to demonstrate how this mechanism can explain the voltage sensitivity of certain Class A GPCRs. Two modes of activation were modeled – one where $Na^+$ ion is transported into the cytoplasm, and the one where $Na^+$ ion is expelled to the intracellular space. The two modes fit well with experimental data on voltage-activated and voltage-suppressed receptors, respectively. Finally, further structure scrutiny and rotamer analysis provided a plausible pathway of $Na^+$ transmembrane translocation through the helical bundle of GPCRs.

## *Publications*


1. Shalaeva DN, Dibrova DV, Galperin MY, Mulkidjanian AY.
**Modeling of interaction between cytochrome c and the WD domains of Apaf-1: bifurcated salt bridges underlying apoptosome assembly.**
Biol Direct. 2015 May 27;10:29. doi: 10.1186/s13062-015-0059-4.

2. Shalaeva DN, Galperin MY, Mulkidjanian AY.
**Eukaryotic G protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins.**
Biol Direct. 2015 Oct 15;10:63. doi: 10.1186/s13062-015-0091-4.

3. Rieger B, Shalaeva DN, Söhnel AC, Kohl W, Duwe P, Mulkidjanian AY, Busch KB.
**Lifetime imaging of GFP at CoxVIIIa reports respiratory supercomplex assembly in live cells.**
Scientific reports. 2017 Apr 6;7(1):1-9. doi: 10.1038/srep46055

4. Shalaeva DN, Cherepanov DA, Galperin MY, Golovin AV, Mulkidjanian AY.
**Evolution of cation binding in the active sites of P-loop nucleoside triphosphatases in relation to the basic catalytic mechanism.**
Elife. 2018 Dec 11;7:e37373. doi: 10.7554/eLife.37373.

5. Shalaeva DN, Cherepanov DA, Galperin MY, Vriend G, Mulkidjanian AY.
**G protein-coupled receptors of class A harness the energy of membrane potential to increase their sensitivity and selectivity.**
Biochim Biophys Acta Biomembr. 2019 Dec 1;1861(12):183051.
doi: 10.1016/j.bbamem.2019.183051.

6. Kozlova MI, Bushmakin IM, Belyaeva JD, Shalaeva DN, Dibrova DV, Cherepanov DA, Mulkidjanian AY.
**Expansion of the "Sodium World" through Evolutionary Time and Taxonomic Space.**
Biochemistry (Mosc). 2020 Dec;85(12):1518-1542.
doi: 10.1134/S0006297920120056.

# *Abbreviations*

| | |
|---|---|
| 2D | Two-dimentional |
| 3D | Three-dimentional |
| 7TM | Heptahelical transmembrane (protein) |
| aa-tRNA | aminoacyl-tRNA |
| ACP | Adenosine 5′-[*β*,*γ*-methylene]triphosphate |
| ADP | Adenosine-5'-diphosphate |
| AGS | Adenosine 5′-[γ-thio]triphosphate |
| amu | Atomic mass unit |
| ANP | Adenosine 5′-[β,γ-imido]triphosphate |
| ATP | Adenosine-5'-triphosphate |
| cAMP | Cyclic adenosine monophosphate |
| ChR2 | Channelrhodopsin 2 |
| cryo-EM | Cryogenic electron microscopy |
| DNA | Deoxyribonucleic acid |
| δ-OR | δ-opioid receptor |
| FRET | Fluorescence resonance energy transfer |
| GAP | GTPase-activating protein |
| GCP | Guanosine 5′-[*β*,*γ*-methylene]triphosphate |
| GDP | Guanosine-5'-diphosphate |
| GIRK | G protein-gated inwardly rectifying potassium ($K^+$) |
| GluR | Metabotropic glutamate receptor 1 |
| GNP | Guanosine 5′-[β,γ-imido]triphosphate |
| GPCR | G-protein coipled receptor |
| GSP | Guanosine 5′-[γ-thio]triphosphate |
| GTP | Guanosine-5'-triphosphate |
| H-bond | Hydrogen bond |
| HMM | Hidden Markov model |
| KR2 | $Na^+$-transporting rhodopsin from the bacterium *Krokinobacter eikastus* |
| LECA | Last eukaryotic common ancestor |
| LSU | Large subunit (ribosomal) |
| LUCA | Last universal cellular ancestor |
| MD | Molecular dynamics |
| MR | Microbial rhodopsin |
| mRNA | Messenger RNA |
| MSA | Multiple sequence alignment |
| NDP | Nucleoside diphosphate |
| NMR | Nuclear magnetic resonance |
| NTP | Nucleoside triphosphate |
| PBC | Periodic boundary conditions |
| PDB | Protein Data Bank |
| RDF | Radial distribution function |
| RGS | Regulators of G-protein signaling |
| RMSD | Root mean square deviation |

| | |
|---|---|
| RMSF | Root mean square fluctuation |
| RNA | Ribonucleic acid |
| rRNA | Ribosomal RNA |
| SR | Signal recognition particle receptor |
| SRL | Sarcin-ricin loop |
| SRP | Signal recognition particle |
| SSU | Small subunit (ribosomal) |
| tRNA | Transfer RNA |
| WB | Walker B (motif) |

# *Supplementary Information*

**Table S1.** Lifetimes of the βγ-conformation of $Mg^{-}$ATP complex in MD simulations

| **Cation** | **$K^+$** | **$Na^+$** | **$NH_4^+$** | **no $M^+$** |
|---|---|---|---|---|
| Average lifetime (ns) | 9.49 | 10.59 | 11.04 | 9.45 |
| Standard deviation | 6.52 | 8.28 | 7.82 | 7.85 |
| Lifetime (ns) for each MD run (total run time, 20 ns) | 7.68 | 13.18 | 0.88 | 0.91 |
| | 16.16 | 0.15 | 19.75 | 3.43 |
| | 16.8 | 1.18 | 19.55 | 0.61 |
| | 7.93 | 11.8 | 8.48 | 14.73 |
| | 4.93 | 4.03 | 2.71 | **20** |
| | 0.28 | **20** | 2.21 | 12.38 |
| | 6.06 | **20** | 1.58 | **20** |
| | 2.75 | 7.01 | 12.8 | 19.23 |
| | 11.76 | 3.86 | 10.71 | 0.26 |
| | 6.33 | 20 | 2.65 | 0.93 |
| | 13.43 | 2.36 | **20** | **20** |
| | 2.65 | 20 | 16.98 | 6.2 |
| | 8.11 | **20** | 17.66 | 0.21 |
| | 11.21 | 20 | 9.16 | 10.58 |
| | 4.9 | **20** | **20** | 3.03 |
| | 8.03 | 16.41 | 1.18 | 13.38 |
| | 0.7 | 3.15 | 1.21 | 1.03 |
| | 14.68 | 2.25 | 1.06 | 5.31 |
| | **20** | 10.93 | 8.36 | 0.36 |
| | **20** | 0.18 | **20** | 9.11 |
| | 0.38 | 6.83 | 5.63 | **20** |
| | 4.68 | **20** | **20** | 6.2 |
| | 8.01 | 0.66 | 20 | 8.48 |
| | **19.75** | 0.83 | 13.38 | 20 |
| | **20** | 20 | 20 | 20 |

For each system, 25 independent 20-ns MD simulation runs were conducted, each starting with the Mg-ATP complex in the βγ conformation. The stability of the βγ conformation was tracked by measuring the distance from the $Mg^{2+}$ ion to the nearest oxygen atom of α-phosphate, and the time periods during which the βγ conformation was retained were compared between different systems. The one-way ANOVA analysis did not reveal any significant dependence of the stability of the βγ-coordination on the monovalent cation present. For each monovalent cation, the βγ-coordination was retained during the whole 20 ns in at least four cases (shown by bold numbers). These simulations were used to characterize the shape of the phosphate chain of ATP with βγ-coordination of the $Mg^{2+}$ ion.

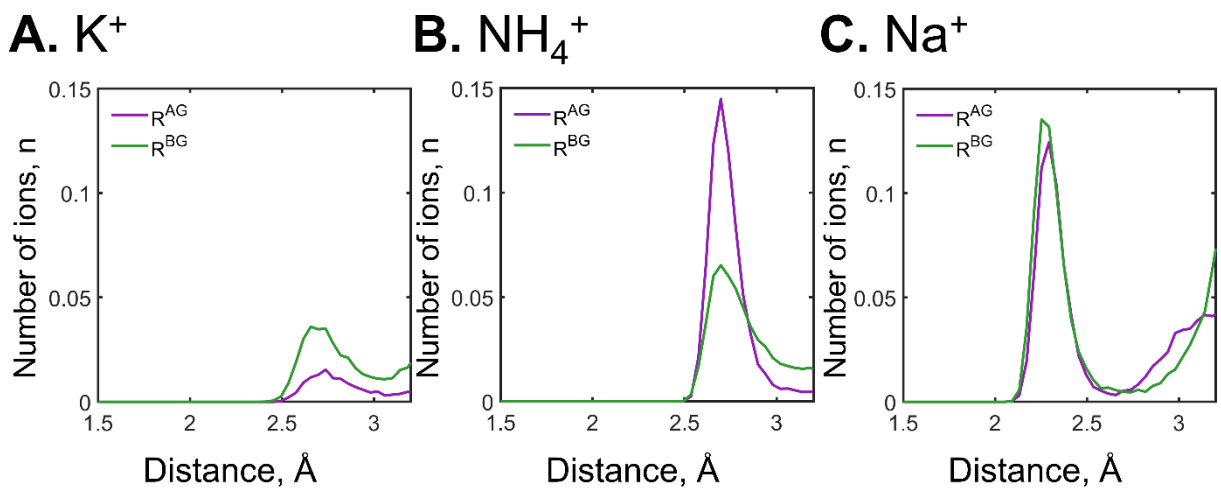


**Figure S1**. $M^+$ ions binding to Mg-GTP as observed in MD simulations. Cation distances to the binding sites ($R^{AG}$ and $R^{BG}$) were calculated as shown on Figure 3.2.4). A-C. Distributions of distances for $K^+$, $NH_4^+$, and $Na^+$ ions, respectively.

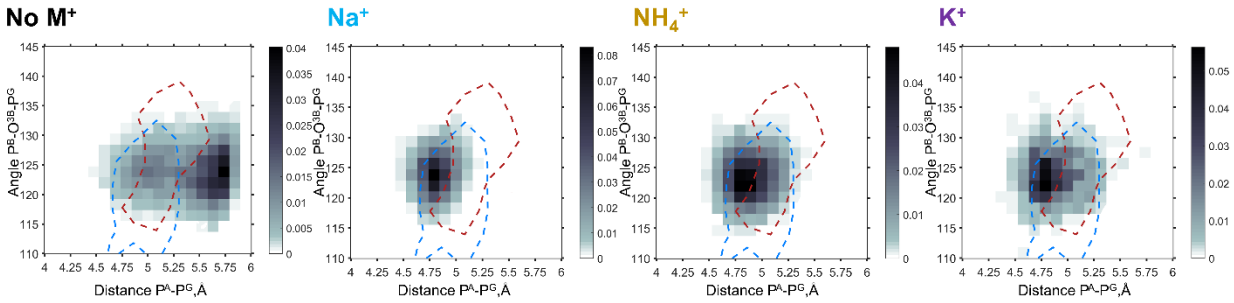


**Figure S2.** Heat maps of the phosphate chain conformation distributions of the Mg-GTP complex as oserved in MD simulations. Conformations were sorted based on distances $P^A$-$P^G$ and angles $P^B$-$O^{3B}$-$P^G$ and plotted as heatmaps. Heatmaps for systems with $M^+$ ions show only those conformations of Mg-GTP complexes, that has at least one $M^+$ ion present within 4 Å radius of the phosphate chain. The colors reflect the probability (estimated as normalized frequency) of the corresponding conformation. Dashed lines outline areas of the conformational space of NTP-mimiking molecules observed in X-ray structures of P-loop proteins: blue for the non-hydrolasable analogs, red for the transition state analogs (see Figure 3.2.9). Data from simulations no. 9–12 in Table 2.1.

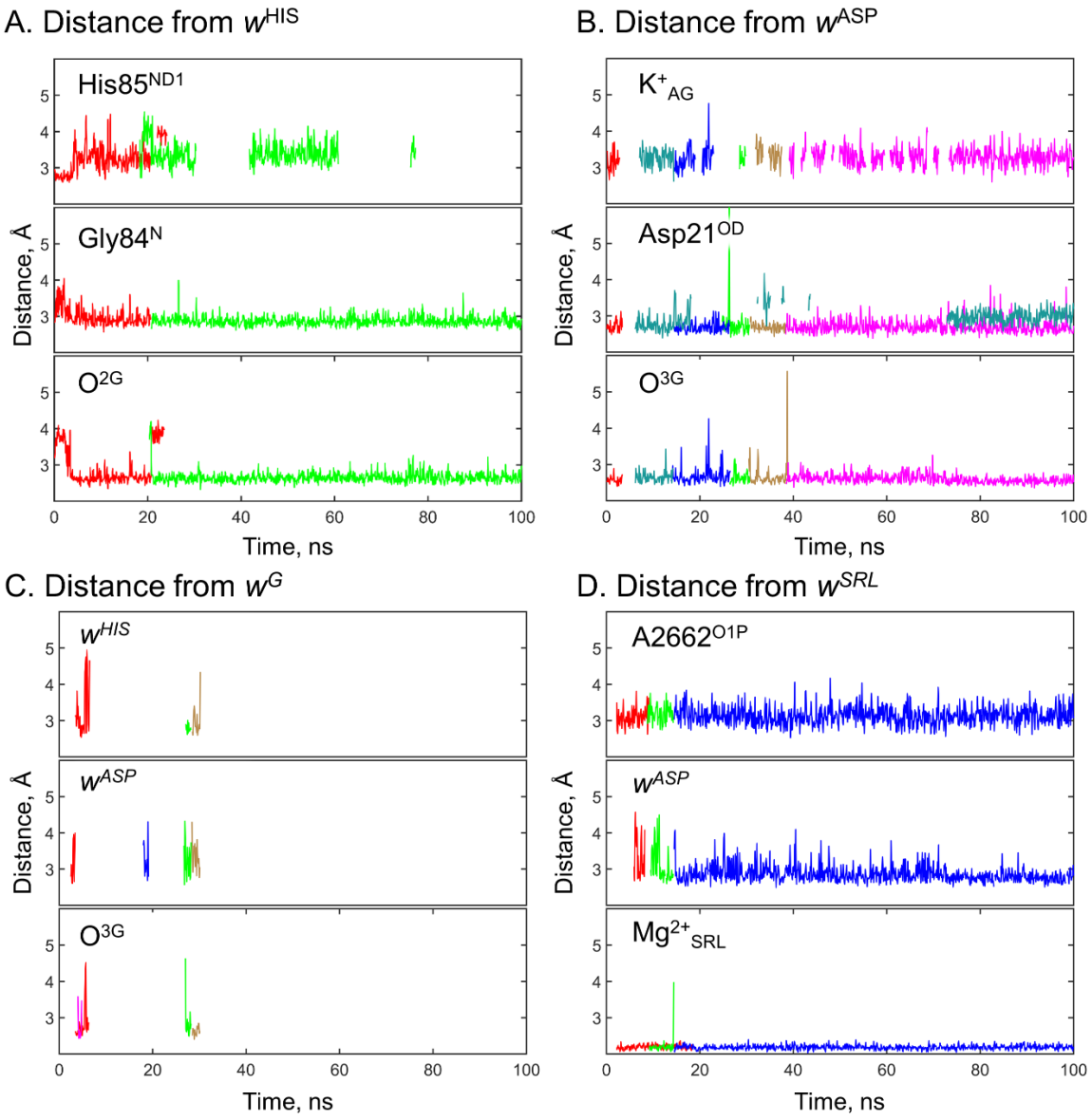


**Figure S3**. Stability of water molecules in the active site of EF-Tu as seen in MD simulation #1. Water molecules nomenclature as shown in Figure 3.2.27. Stability of water molecules in the active site of EF-Tu as seen in MD simulations. A. Distances measured from oxygen atom of water molecules in position $w^{HIS}$ to $N^{D1}$ atom of $His^{D+4}$, backbone nitrogen of $Gly^{D+3}$, and the $O^{2G}$ atom of GTP. B. Distances measured from the oxygen atom of water molecules in position $w^{ASP}$ to the $K^+$ ion in the AG site, the nearest $O^{D1}/O^{D2}$ oxygen atom of $Asp^{K-3}$ sidechain, and the $O^{3G}$ atom of GTP. C. Distances measured from the oxygen atom of the $w^G$ water molecule to oxygen atoms of water molecules $w^{HIS}$ and $w^{ASP}$, and the $O^{3G}$ atom of GTP. D. Distances measured from the oxygen atom of $w^G$ water molecule to the $O^{1P}$ atom of A2662, the oxygen atom of the $w^{ASP}$ water molecule and the $Mg^{2+}{}_{SRL}$ ion. All distances were measured to oxygen atoms of water molecules; different colors correspond to different individual water molecules, within each panel same colors correspond to the same molecules. Data are shown only for the cases when a short distance (<5Å) was maintained for at least 1 ns.

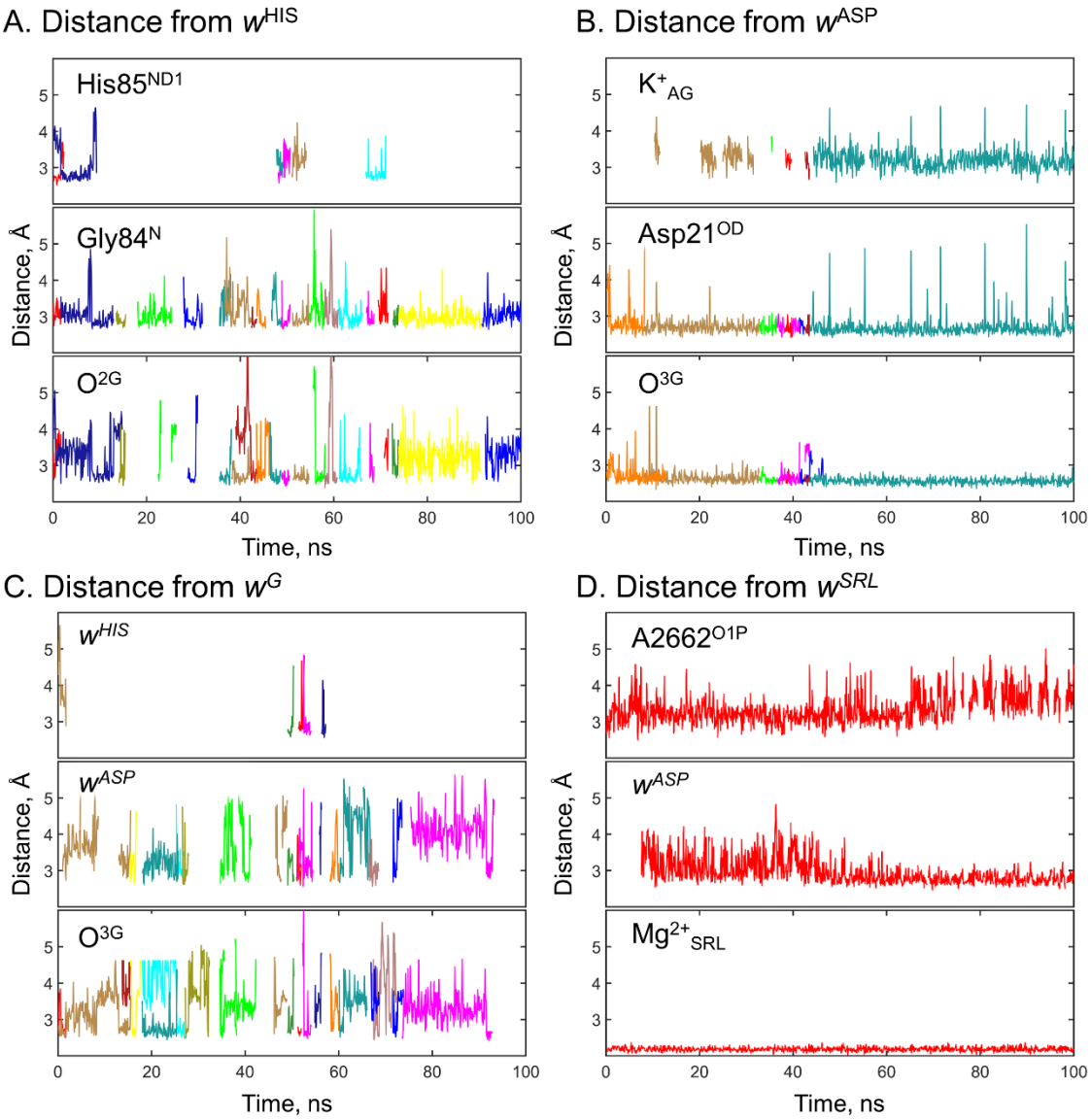


**Figure S4**. Stability of water molecules in the active site of EF-Tu as seen in MD simulation #2. Water molecules nomenclature as shown in Figure 3.2.27. A. Distances measured from oxygen atom of water molecules in position $w^{HIS}$ to $N^{D1}$ atom of $His^{D+4}$, backbone nitrogen of $Gly^{D+3}$, and the $O^{2G}$ atom of GTP. B. Distances measured from the oxygen atom of water molecules in position $w^{ASP}$ to the $K^{+}$ ion in the AG site, the nearest $O^{D1}/O^{D2}$ oxygen atom of $Asp^{K-3}$ sidechain, and the $O^{3G}$ atom of GTP. C. Distances measured from the oxygen atom of the $w^{G}$ water molecule to oxygen atoms of water molecules $w^{HIS}$ and $w^{ASP}$, and the $O^{3G}$ atom of GTP. D. Distances measured from the oxygen atom of $w^{G}$ water molecule to the $O^{1P}$ atom of A2662, the oxygen atom of the $w^{ASP}$ water molecule and the $Mg^{2+}{}_{SRL}$ ion. All distances were measured to oxygen atoms of water molecules; different colors correspond to different individual water molecules, within each panel same colors correspond to the same molecules. Data are shown only for the cases when a short distance (<5Å) was maintained for at least 1 ns.

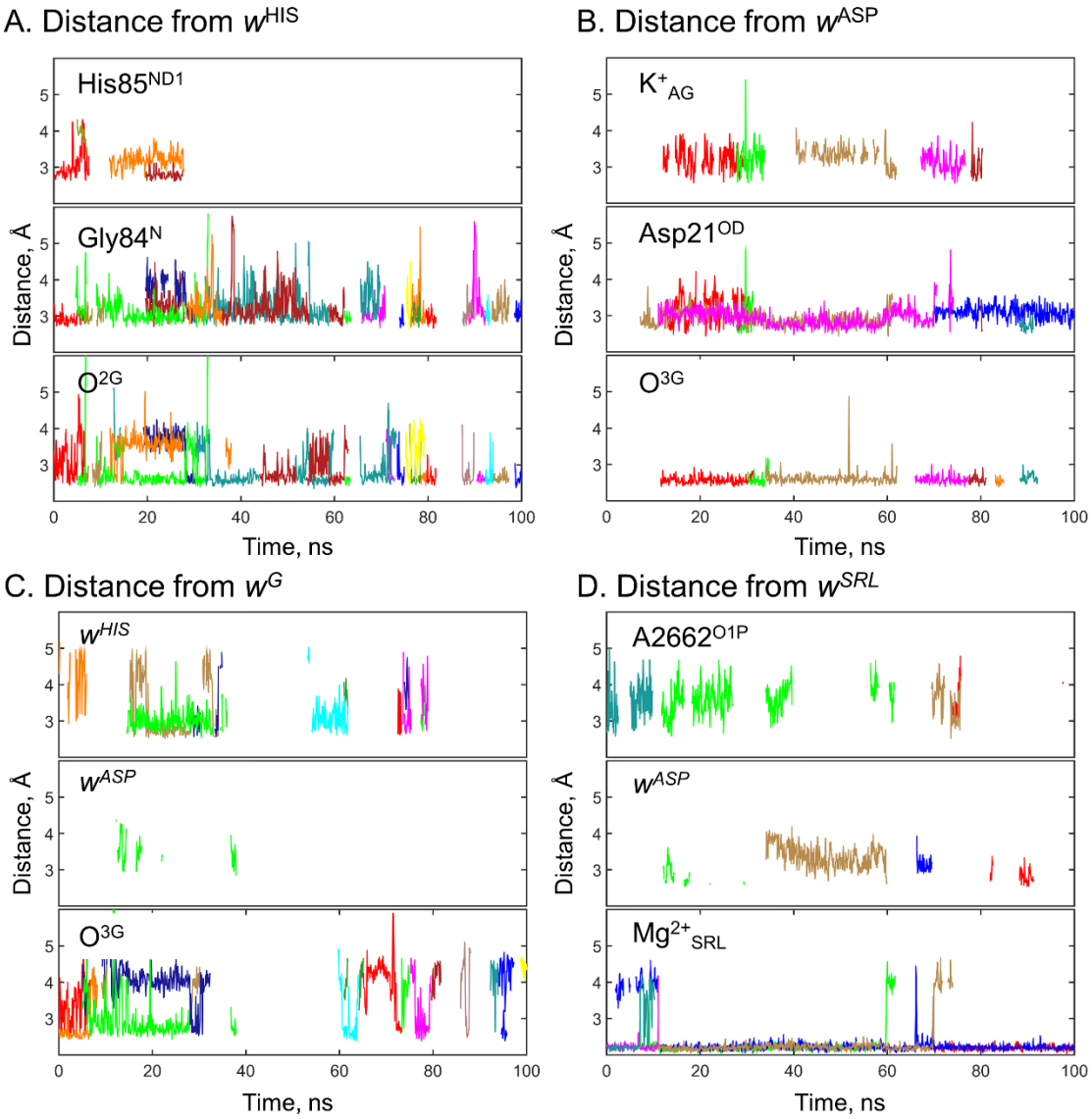


**Figure S5.** Stability of water molecules in the active site of EF-Tu as seen in MD simulation #4. Water molecules nomenclature as shown in Figure 3.2.27. A. Distances measured from oxygen atom of water molecules in position $w^{HIS}$ to $N^{D1}$ atom of $His^{D+4}$, backbone nitrogen of $Gly^{D+3}$, and the $O^{2G}$ atom of GTP. B. Distances measured from the oxygen atom of water molecules in position $w^{ASP}$ to the $K^{+}$ ion in the AG site, the nearest $O^{D1}/O^{D2}$ oxygen atom of $Asp^{K-3}$ sidechain, and the $O^{3G}$ atom of GTP. C. Distances measured from the oxygen atom of the $w^{G}$ water molecule to oxygen atoms of water molecules $w^{HIS}$ and $w^{ASP}$, and the $O^{3G}$ atom of GTP. D. Distances measured from the oxygen atom of $w^{G}$ water molecule to the $O^{1P}$ atom of A2662, the oxygen atom of the $w^{ASP}$ water molecule and the $Mg^{2+}{}_{SRL}$ ion. All distances were measured to oxygen atoms of water molecules; different colors correspond to different individual water molecules, within each panel same colors correspond to the same molecules. Data are shown only for the cases when a short distance (<5Å) was maintained for at least 1 ns.

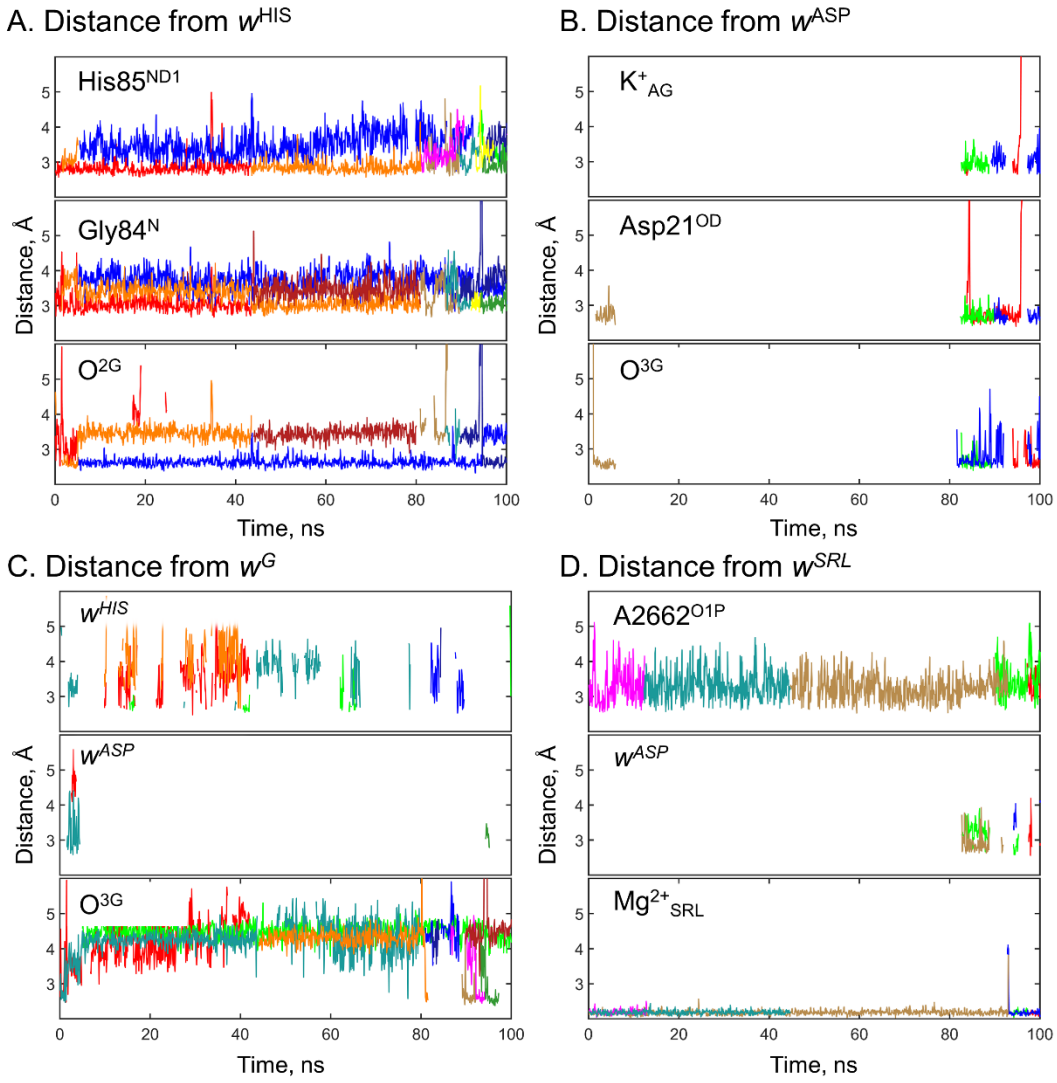


**Figure S6.** Stability of water molecules in the active site of EF-Tu as seen in MD simulation #5. Water molecules nomenclature as shown in Figure 3.2.27. A. Distances measured from oxygen atom of water molecules in position $w^{HIS}$ to $N^{D1}$ atom of $His^{D+4}$, backbone nitrogen of $Gly^{D+3}$ and the $O^{2G}$ atom of GTP. B. Distances measured from the oxygen atom of water molecules in position $w^{ASP}$ to the $K^{+}$ ion in the AG site, the nearest of $O^{D1}/O^{D2}$ oxygen atoms of $Asp^{K-3}$ sidechain, and the $O^{3G}$ atom of GTP. C. Distances measured from the oxygen atom of the $w^{G}$ water molecule to oxygen atoms of water molecules $w^{HIS}$ and $w^{ASP}$, and the $O^{3G}$ atom of GTP. D. Distances measured from the oxygen atom of $w^{G}$ water molecule to the $O^{1P}$ atom of A2662, the oxygen atom of the $w^{ASP}$ water molecule and the $Mg^{2+}{}_{SRL}$ ion. All distances were measured to oxygen atoms of water molecules; different colors correspond to different individual water molecules, within each panel same colors correspond to the same molecules. Data are shown only for the cases when a short distance (<5Å) was maintained for at least 1 ns.

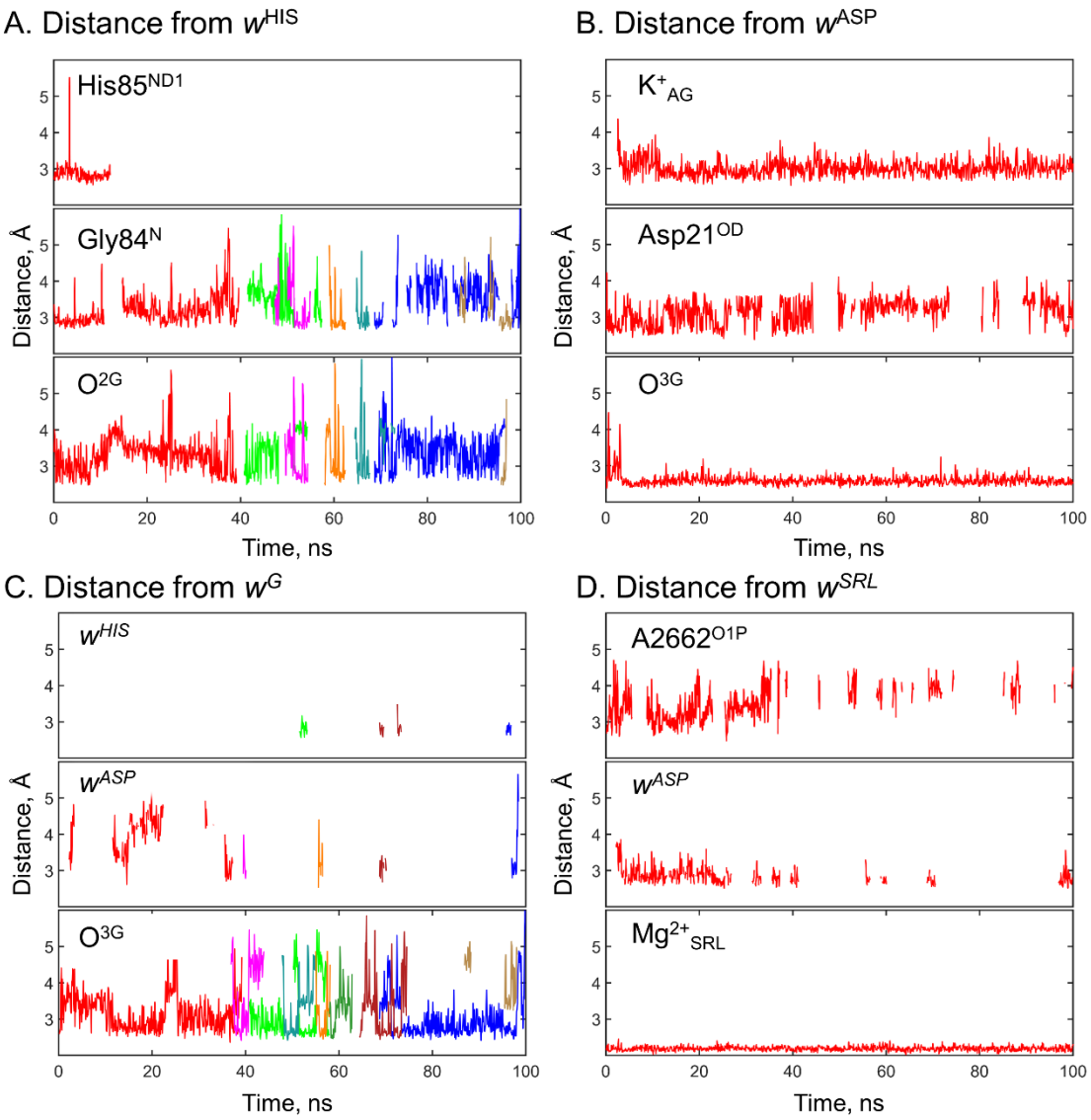


**Figure S7.** Stability of water molecules in the active site of EF-Tu as seen in MD simulation #6. Water molecules nomenclature as shown in Figure 3.2.27. A. Distances measured from oxygen atom of water molecules in position $w^{HIS}$ to $N^{D1}$ atom of $His^{D+4}$, backbone nitrogen of $Gly^{D+3}$, and the $O^{2G}$ atom of GTP. B. Distances measured from the oxygen atom of water molecules in position $w^{ASP}$ to the $K^{+}$ ion in the AG site, the nearest $O^{D1}/O^{D2}$ oxygen atom of $Asp^{K-3}$ sidechain, and the $O^{3G}$ atom of GTP. C. Distances measured from the oxygen atom of the $w^{G}$ water molecule to oxygen atoms of water molecules $w^{HIS}$ and $w^{ASP}$, and the $O^{3G}$ atom of GTP. D. Distances measured from the oxygen atom of $w^{G}$ water molecule to the $O^{1P}$ atom of A2662, the oxygen atom of the $w^{ASP}$ water molecule and the $Mg^{2+}{}_{SRL}$ ion. All distances were measured to oxygen atoms of water molecules; different colors correspond to different individual water molecules, within each panel same colors correspond to the same molecules. Data are shown only for the cases when a short distance (<5Å) was maintained for at least 1 ns.

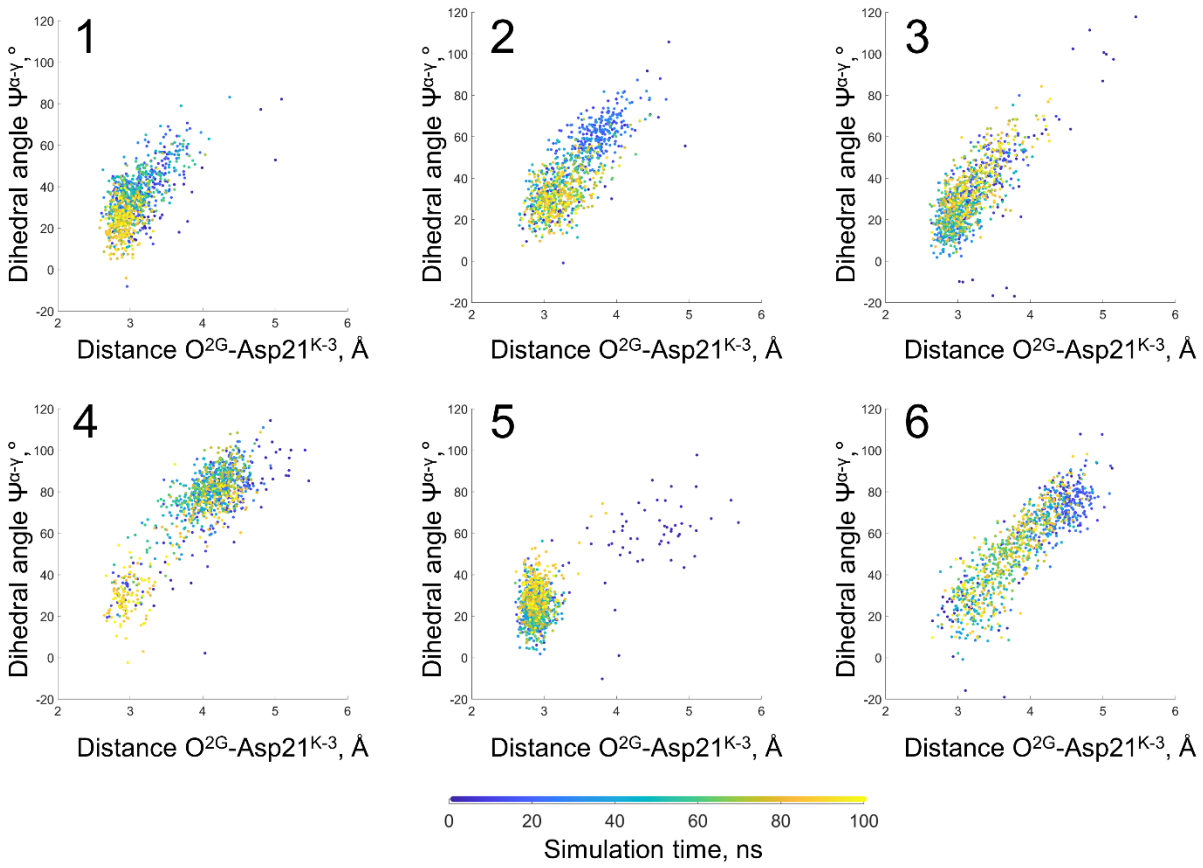


**Figure S8.** $K^+$-induced rotation of γ-phosphate and formation of an H-bond between the $O^{2G}$ oxygen atom and the backbone nitrogen atom of $Asp21^{K-3}$. Conformational space of GTP during MD simulation of EF-Tu is shown as a scatter plot, where Y-axis shows the $\Psi^{\alpha-\gamma}$ dihedral angle and the X-axis shows the length of the H-bond between the $O^{2G}$ oxygen atom and the backbone nitrogen atom of $Asp21^{K-3}$; the color of each dots corresponds to the point of time during the simulation. Data from simulations as listed in Table 2.3.

# References


1. Macallum AB: **The Palaeochemistry of the Ocean in Relation to Animal and Vegetable Protoplasm, by AB Macallum**: University library; 1904.
2. Macallum AB: **The paleochemistry of the body fluids and tissues**. *Physiol Rev* 1926, **6**(2):316-357.
3. Lehninger AL, Nelson DL, Cox MM: **Lehninger principles of biochemistry**: Macmillan; 2005.
4. Westheimer FH: **Why Nature Chose Phosphates**. *Science* 1987, **235**(4793):1173-1178.
5. Sekowska A, Kung HF, Danchin A: **Sulfur metabolism in Escherichia coli and related bacteria: facts and fiction**. *J Mol Microbiol Biotechnol* 2000, **2**(2):145-177.
6. Danchin A, Nikel PI: **Why Nature Chose Potassium**. *J Mol Evol* 2019, **87**(9-10):271-288.
7. Mulkidjanian AY, Bychkov AY, Dibrova DV, Galperin MY, Koonin EV: **Origin of first cells at terrestrial, anoxic geothermal fields**. *Proc Natl Acad Sci U S A* 2012, **109**(14):E821-830.
8. Dibrova DV, Galperin MY, Koonin EV, Mulkidjanian AY: **Ancient systems of sodium/potassium homeostasis as predecessors of membrane bioenergetics**. *Biochemistry-Moscow* 2015, **80**(5):495-516.
9. Kozlova M, Bushmakin I, Belyaeva J, Shalaeva D, Dibrova D, Cherepanov D, Mulkidjanian A: **Expansion of the "Sodium World" through Evolutionary Time and Taxonomic Space**. *Biochemistry (Moscow)* 2020, **85**(12):1518-1542.
10. Skulachev VP: **Membrane-linked energy buffering as the biological function of Na+/K+ gradient**. *FEBS Lett* 1978, **87**(2):171-179.
11. Gruber R, Horovitz A: **Allosteric Mechanisms in Chaperonin Machines**. *Chem Rev* 2016, **116**(11):6588-6606.
12. Wilbanks SM, McKay DB: **How potassium affects the activity of the molecular chaperone Hsc70. II. Potassium binds specifically in the ATPase active site**. *J Biol Chem* 1995, **270**(5):2251-2257.
13. Partono S, Lewin AS: **The rate and specificity of a group I ribozyme are inversely affected by choice of monovalent salt**. *Nucleic Acids Res* 1991, **19**(3):605-609.
14. Rozov A, Khusainov I, El Omari K, Duman R, Mykhaylyk V, Yusupov M, Westhof E, Wagner A, Yusupova G: **Importance of potassium ions for ribosome structure and function revealed by long-wavelength X-ray diffraction**. *Nat Commun* 2019, **10**(1):2519.
15. Lambert D, Leipply D, Shiman R, Draper DE: **The influence of monovalent cation size on the stability of RNA tertiary structures**. *J Mol Biol* 2009, **390**(4):791-804.
16. Meyer S, Bohme S, Kruger A, Steinhoff H-J, Klare JP, Wittinghofer A: **Kissing G domains of MnmE monitored by X-ray crystallography and pulse electron paramagnetic resonance spectroscopy**. *PLoS biology* 2009, **7**(10):e1000212.
17. Anand B, Majumdar S, Prakash B: **Structural basis unifying diverse GTP hydrolysis mechanisms**. *Biochemistry* 2013, **52**(6):1122-1130.
18. Fasano O, De Vendittis E, Parmeggiani A: **Hydrolysis of GTP by elongation factor Tu can be induced by monovalent cations in the absence of other effectors**. *J Biol Chem* 1982, **257**(6):3145-3150.
19. Chinali G, Parmeggiani A: **The coupling with polypeptide synthesis of the GTPase activity dependent on elongation factor G**. *J Biol Chem* 1980, **255**(15):7455-7459.
20. Dubnoff JS, Maitra U: **Characterization of the ribosome-dependent guanosine triphosphatase activity of polypeptide chain initiation factor IF 2**. *J Biol Chem* 1972, **247**(9):2876-2883.
21. Kuhle B, Ficner R: **A monovalent cation acts as structural and catalytic cofactor in translational GTPases**. *EMBO J* 2014, **33**(21):2547-2563.
22. Koonin EV: **Comparative genomics, minimal gene-sets and the last universal common ancestor**. *Nature Reviews Microbiology* 2003, **1**(2):127-136.
23. Verstraeten N, Fauvart M, Versees W, Michiels J: **The universally conserved prokaryotic GTPases**. *Microbiol Mol Biol Rev* 2011, **75**(3):507-542, second and third pages of table of contents.
24. Leipe DD, Wolf YI, Koonin EV, Aravind L: **Classification and evolution of P-loop GTPases and related ATPases**. *J Mol Biol* 2002, **317**(1):41-72.

25. Koonin EV, Wolf YI, Aravind L: **Protein fold recognition using sequence profiles and its application in structural genomics**. *Advances in Protein Chemistry, Vol 54* 2000, **54**:245-275.
26. Leipe DD, Koonin EV, Aravind L: **Evolution and classification of P-loop kinases and related proteins**. *J Mol Biol* 2003, **333**(4):781-815.
27. Bos JL, Rehmann H, Wittinghofer A: **GEFs and GAPs: Critical elements in the control of small G proteins**. *Cell* 2007, **129**(5):865-877.
28. Scrima A, Wittinghofer A: **Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element**. *EMBO J* 2006, **25**(12):2940-2951.
29. Ash M-R, Maher MJ, Mitchell Guss J, Jormakka M: **The cation-dependent G-proteins: in a class of their own**. *FEBS letters* 2012, **586**(16):2218-2224.
30. Chappie JS, Acharya S, Leonard M, Schmid SL, Dyda F: **G domain dimerization controls dynamin's assembly-stimulated GTPase activity**. *Nature* 2010, **465**(7297):435-440.
31. Manikas RG, Thomson E, Thoms M, Hurt E: **The K(+)-dependent GTPase Nug1 is implicated in the association of the helicase Dbp10 to the immature peptidyl transferase centre during ribosome maturation**. *Nucleic Acids Research* 2016, **44**(4):1800-1812.
32. Achila D, Gulati M, Jain N, Britton RA: **Biochemical characterization of ribosome assembly GTPase RbgA in Bacillus subtilis**. *J Biol Chem* 2012, **287**(11):8417-8423.
33. Ebel C, Guinet F, Langowski J, Urbanke C, Gagnon J, Zaccai G: **Solution studies of elongation factor Tu from the extreme halophile *Halobacterium marismortui***. *J Mol Biol* 1992, **223**(1):361-371.
34. Ash MR, Guilfoyle A, Clarke RJ, Guss JM, Maher MJ, Jormakka M: **Potassium-activated GTPase reaction in the G Protein-coupled ferrous iron transporter B**. *J Biol Chem* 2010, **285**(19):14594-14602.
35. Tomar SK, Kumar P, Prakash B: **Deciphering the catalytic machinery in a universally conserved ribosome binding ATPase YchF**. *Biochem Biophys Res Commun* 2011, **408**(3):459-464.
36. Rafay A, Majumdar S, Prakash B: **Exploring potassium-dependent GTP hydrolysis in TEES family GTPases**. *FEBS Open Bio* 2012, **2**:173-177.
37. Foucher AE, Reiser JB, Ebel C, Housset D, Jault JM: **Potassium Acts as a GTPase-Activating Element on Each Nucleotide-Binding Domain of the Essential Bacillus subtilis EngA**. *PLoS One* 2012, **7**(10).
38. Hwang J, Inouye M: **An essential GTPase, Der, containing double GTP-binding domains from Escherichia coli and Thermotoga maritima**. *Journal of Biological Chemistry* 2001, **276**(33):31415-31421.
39. Anand B, Surana P, Prakash B: **Deciphering the catalytic machinery in 30S ribosome assembly GTPase YqeH**. *PLoS One* 2010, **5**(4):e9944.
40. Shim KS, Schmutte C, Yoder K, Fishel R: **Defining the salt effect on human RAD51 activities**. *DNA Repair (Amst)* 2006, **5**(6):718-730.
41. Lowenstein JM: **The stimulation of transphosphorylation by alkali-metal ions**. *Biochem J* 1960, **75**:269-274.
42. Saier MH, Jr.: **A functional-phylogenetic classification system for transmembrane solute transporters**. *Microbiol Mol Biol Rev* 2000, **64**(2):354-411.
43. Katritch V, Fenalti G, Abola EE, Roth BL, Cherezov V, Stevens RC: **Allosteric sodium in class A GPCR signaling**. *Trends Biochem Sci* 2014, **39**(5):233-244.
44. Alder BJ, Wainwright TE: **Phase Transition for a Hard Sphere System**. *Journal of Chemical Physics* 1957, **27**(5):1208-1209.
45. Alder BJ, Wainwright TE: **Studies in Molecular Dynamics .1. General Method**. *Journal of Chemical Physics* 1959, **31**(2):459-466.
46. Rahman A: **Correlations in the Motion of Atoms in Liquid Argon**. *Phys Rev* 1964, **136**(2A):A405-A411.
47. Stilling.Fh, Rahman A: **Improved Simulation of Liquid Water by Molecular-Dynamics**. *Journal of Chemical Physics* 1974, **60**(4):1545-1557.
48. Mccammon JA, Gelin BR, Karplus M: **Dynamics of Folded Proteins**. *Nature* 1977, **267**(5612):585-590.

49. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I *et al*: **CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields**. *Journal of Computational Chemistry* 2010, **31**(4):671-690.
50. Christen M, Hunenberger PH, Bakowies D, Baron R, Burgi R, Geerke DP, Heinz TN, Kastenholz MA, Krautler V, Oostenbrink C *et al*: **The GROMOS software for biomolecular simulation: GROMOS05**. *Journal of Computational Chemistry* 2005, **26**(16):1719-1751.
51. Humphrey W, Dalke A, Schulten K: **VMD: Visual molecular dynamics**. *Journal of Molecular Graphics & Modelling* 1996, **14**(1):33-38.
52. **The PyMOL Molecular Graphics System, Version 1.7.2.1**. In.: Schrödinger, LLC.; 2010.
53. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML: **Comparison of Simple Potential Functions for Simulating Liquid Water**. *Journal of Chemical Physics* 1983, **79**(2):926-935.
54. Joung IS, Cheatham TE: **Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations**. *Journal of Physical Chemistry B* 2008, **112**(30):9020-9041.
55. Callahan KM, Casillas-Ituarte NN, Roeselova M, Allen HC, Tobias DJ: **Solvation of Magnesium Dication: Molecular Dynamics Simulation and Vibrational Spectroscopic Study of Magnesium Chloride in Aqueous Solutions**. *Journal of Physical Chemistry A* 2010, **114**(15):5141-5148.
56. Ryckaert J-P, Ciccotti G, Berendsen HJC: **Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes**. *Journal of Computational Physics* 1977, **23**(3):327-341.
57. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR: **Molecular-Dynamics with Coupling to an External Bath**. *Journal of Chemical Physics* 1984, **81**(8):3684-3690.
58. Feller SE, Zhang YH, Pastor RW, Brooks BR: **Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method**. *Journal of Chemical Physics* 1995, **103**(11):4613-4621.
59. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D *et al*: **GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit**. *Bioinformatics* 2013, **29**(7):845-854.
60. **MATLAB and Statistics Toolbox Release 2017a**. In., 2017a edn: The MathWorks, Inc., Natick, MA; 2017.
61. Rieger B, Shalaeva DN, Sohnel AC, Kohl W, Duwe P, Mulkidjanian AY, Busch KB: **Lifetime imaging of GFP at CoxVIIIa reports respiratory supercomplex assembly in live cells**. *Sci Rep* 2017, **7**:46055.
62. Webb B, Sali A: **Comparative Protein Structure Modeling Using MODELLER**. *Curr Protoc Protein Sci* 2016, **86**:2 9 1-2 9 37.
63. Verlet L: **Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules**. *Phys Rev* 1967, **159**(1):98-+.
64. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM: **LINCS: A linear constraint solver for molecular simulations**. *J Comput Chem* 1997, **18**(12):1463-1472.
65. Orengo CA, Thornton JM: **Protein families and their evolution-a structural perspective**. *Annu Rev Biochem* 2005, **74**:867-900.
66. Zuckerkandl E, Pauling L: **Evolutionary divergence and convergence in proteins**. In: *Evolving genes and proteins.* Elsevier; 1965: 97-166.
67. Russell RB, Saqi MA, Sayle RA, Bates PA, Sternberg MJ: **Recognition of analogous and homologous protein folds: analysis of sequence and structure conservation**. *J Mol Biol* 1997, **269**(3):423-439.
68. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.
69. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**(17):3389-3402.

70. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment**. *J Mol Biol* 2000, **302**(1):205-217.
71. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor**. *Bioinformatics* 2004, **20**(3):426-427.
72. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J *et al*: **Pfam: the protein families database**. *Nucleic Acids Res* 2014, **42**(Database issue):D222-230.
73. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank**. *Nucleic Acids Res* 2000, **28**(1):235-242.
74. Berman H, Henrick K, Nakamura H: **Announcing the worldwide Protein Data Bank**. *Nature Structural Biology* 2003, **10**(12):980.
75. Kleywegt GJ, Harris MR, Zou JY, Taylor TC, Wahlby A, Jones TA: **The Uppsala Electron-Density Server**. *Acta Crystallogr D Biol Crystallogr* 2004, **60**(Pt 12 Pt 1):2240-2249.
76. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI: **OPM: orientations of proteins in membranes database**. *Bioinformatics* 2006, **22**(5):623-625.
77. Jo S, Kim T, Iyer VG, Im W: **CHARMM-GUI: a web-based graphical user interface for CHARMM**. *J Comput Chem* 2008, **29**(11):1859-1865.
78. Horn F, Weare J, Beukers MW, Horsch S, Bairoch A, Chen W, Edvardsen O, Campagne F, Vriend G: **GPCRDB: an information system for G protein-coupled receptors**. *Nucleic Acids Res* 1998, **26**(1):275-279.
79. Pandy-Szekeres G, Munk C, Tsonkov TM, Mordalski S, Harpsoe K, Hauser AS, Bojarski AJ, Gloriam DE: **GPCRdb in 2018: adding GPCR structure models and ligands**. *Nucleic Acids Res* 2018, **46**(D1):D440-D446.
80. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, Green RK, Goodsell DS, Prlic A, Quesada M *et al*: **The RCSB Protein Data Bank: new resources for research and education**. *Nucleic Acids Res* 2013, **41**(Database issue):D475-482.
81. Ye Y, Godzik A: **Flexible structure alignment by chaining aligned fragment pairs allowing twists**. *Bioinformatics* 2003, **19 Suppl 2**:ii246-255.
82. Krissinel E, Henrick K: **Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions**. *Acta Crystallogr D Biol Crystallogr* 2004, **60**(Pt 12 Pt 1):2256-2268.
83. Prlic A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, Bourne PE: **Pre-calculated protein structure alignments at the RCSB PDB website**. *Bioinformatics* 2010, **26**(23):2983-2985.
84. Stryer L: **Biochemistry**. 2013.
85. Lipmann F: **Metabolic generation and utilization of phosphate bond energy**. *Adv Enzymol Relat Areas Mol Biol* 1941, **1**:99-162.
86. Spessard GO: **ACD Labs/LogP dB 3.5 and ChemSketch 3.5**. *Journal of chemical information and computer sciences* 1998, **38**(6):1250-1253.
87. Niefind K, Putter M, Guerra B, Issinger OG, Schomburg D: **GTP plus water mimic ATP in the active site of protein kinase CK2**. *Nature Structural Biology* 1999, **6**(12):1100-1103.
88. Weller M, Haaga M, Laing W: **GTP, as well as ATP, can act as a substrate for the intrinsic protein kinase activity of synaptic plasma membranes**. *Molecular and cellular biochemistry* 1981, **40**(2):75-85.
89. Fiske CH, Subbarow Y: **Phosphorus Compounds of Muscle and Liver**. *Science* 1929, **70**(1816):381-382.
90. Astumian RD, Bier M: **Mechanochemical coupling of the motion of molecular motors to ATP hydrolysis**. *Biophysical Journal* 1996, **70**(2):637-653.
91. Vetter IR, Wittinghofer A: **Signal transduction - The guanine nucleotide-binding switch in three dimensions**. *Science* 2001, **294**(5545):1299-1304.
92. Sprang SR: **G proteins, effectors and GAPs: structure and mechanism**. *Curr Opin Struc Biol* 1997, **7**(6):849-856.
93. Grant BJ, Gorfe AA, McCammon JA: **Large conformational changes in proteins: signaling and other functions**. *Curr Opin Struc Biol* 2010, **20**(2):142-147.
94. Lupas AN, Martin J: **AAA proteins**. *Curr Opin Struc Biol* 2002, **12**(6):746-753.

95. Hanson PI, Whiteheart SW: **AAA+ proteins: Have engine, will work**. *Nature reviews Molecular cell biology* 2005, **6**(7):519-529.
96. Sahlin K, Harris RC: **The creatine kinase reaction: a simple reaction with functional complexity**. *Amino Acids* 2011, **40**(5):1363-1367.
97. Casino P, Rubio V, Marina A: **The mechanism of signal transduction by two-component systems**. *Curr Opin Struct Biol* 2010, **20**(6):763-771.
98. Lupas AN, Ponting CP, Russell RB: **On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world?** *Journal of Structural Biology* 2001, **134**(2-3):191-203.
99. Ponting CP, Russell RR: **The natural history of protein domains**. *Annu Rev Biophys Biomol Struct* 2002, **31**:45-71.
100. Söding J, Lupas AN: **More than the sum of their parts: on the evolution of proteins from peptides**. *BioEssays* 2003, **25**(9):837-846.
101. Ranea JA, Sillero A, Thornton JM, Orengo CA: **Protein superfamily evolution and the last universal common ancestor (LUCA)**. *J Mol Evol* 2006, **63**(4):513-525.
102. Alva V, Soding J, Lupas AN: **A vocabulary of ancient peptides at the origin of folded proteins**. *eLife* 2015, **4**:e09410.
103. la Cour TF, Nyborg J, Thirup S, Clark BF: **Structural details of the binding of guanosine diphosphate to elongation factor Tu from E. coli as studied by X-ray crystallography**. *EMBO J* 1985, **4**(9):2385-2388.
104. Jurnak F: **Induction of elongation factor Tu-GDP crystal polymorphism by polyethylene glycol contaminants**. *J Mol Biol* 1985, **185**(1):215-217.
105. Hingorani VN, Ho YK: **A structural model for the alpha-subunit of transducin. Implications of its role as a molecular switch in the visual signal transduction mechanism**. *FEBS Lett* 1987, **220**(1):15-22.
106. Pai EF, Krengel U, Petsko GA, Goody RS, Kabsch W, Wittinghofer A: **Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis**. *EMBO J* 1990, **9**(8):2351-2359.
107. Walker JE, Saraste M, Runswick MJ, Gay NJ: **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**(8):945-951.
108. Saraste M, Sibbald PR, Wittinghofer A: **The P-loop--a common motif in ATP- and GTP-binding proteins**. *Trends Biochem Sci* 1990, **15**(11):430-434.
109. Bourne HR, Sanders DA, McCormick F: **The GTPase superfamily: conserved structure and molecular mechanism**. *Nature* 1991, **349**(6305):117-127.
110. Blackburn GM, Cherfils J, Moss GP, Richards NGJ, Waltho JP, Williams NH, Wittinghofer A: **How to name atoms in phosphates, polyphosphates, their derivatives and mimics, and transition state analogues for enzyme-catalysed phosphoryl transfer reactions (IUPAC Recommendations 2016)**. *Pure Appl Chem* 2017, **89**(5):653-675.
111. Milburn MV, Tong L, deVos AM, Brunger A, Yamaizumi Z, Nishimura S, Kim SH: **Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins**. *Science* 1990, **247**(4945):939-945.
112. John J, Sohmen R, Feuerstein J, Linke R, Wittinghofer A, Goody RS: **Kinetics of interaction of nucleotides with nucleotide-free H-ras p21**. *Biochemistry* 1990, **29**(25):6058-6065.
113. Vetter IR, Wittinghofer A: **Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer**. *Q Rev Biophys* 1999, **32**(1):1-56.
114. Allin C, Gerwert K: **Ras catalyzes GTP hydrolysis by shifting negative charges from gamma- to beta-phosphate as revealed by time-resolved FTIR difference spectroscopy**. *Biochemistry* 2001, **40**(10):3037-3046.
115. Shalaeva DN, Cherepanov DA, Galperin MY, Golovin AV, Mulkidjanian AY: **Evolution of cation binding in the active sites of P-loop nucleoside triphosphatases in relation to the basic catalytic mechanism**. *Elife* 2018, **7**.
116. Gerwert K, Mann D, Kotting C: **Common mechanisms of catalysis in small and heterotrimeric GTPases and their respective GAPs**. *Biol Chem* 2017, **398**(5-6):523-533.

117. Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, Schmitz F, Wittinghofer A: **The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants**. *Science* 1997, **277**(5324):333-338.
118. Scheffzek K, Ahmadian MR, Wittinghofer A: **GTPase-activating proteins: helping hands to complement an active site**. *Trends Biochem Sci* 1998, **23**(7):257-262.
119. Ogura T, Whiteheart SW, Wilkinson AJ: **Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases**. *J Structural Biology* 2004, **146**(1-2):106-112.
120. Wendler P, Ciniawsky S, Kock M, Kube S: **Structure and function of the AAA+ nucleotide binding pocket**. *Biochim Biophys Acta* 2012, **1823**(1):2-14.
121. Wittinghofer A, Vetter IR: **Structure-function relationships of the G domain, a canonical switch motif**. *Annu Rev Biochem* 2011, **80**:943-971.
122. Kamerlin SC, Sharma PK, Prasad RB, Warshel A: **Why nature really chose phosphate**. *Quarterly Reviews of Biophysics* 2013, **46**(1):1-132.
123. Jin Y, Molt RW, Jr., Blackburn GM: **Metal fluorides: Tools for structural and computational analysis of phosphoryl transfer enzymes**. *Top Curr Chem (Cham)* 2017, **375**(2):36.
124. Nam K, Pu J, Karplus M: **Trapping the ATP binding state leads to a detailed understanding of the F1-ATPase mechanism**. *Proc Natl Acad Sci U S A* 2014, **111**(50):17851-17856.
125. Graham DL, Lowe PN, Grime GW, Marsh M, Rittinger K, Smerdon SJ, Gamblin SJ, Eccleston JF: **MgF(3)(-) as a transition state analog of phosphoryl transfer**. *Chem Biol* 2002, **9**(3):375-381.
126. Bourne HR, Sanders DA, McCormick F: **The GTPase superfamily: a conserved switch for diverse cell functions**. *Nature* 1990, **348**(6297):125-132.
127. Voorhees RM, Schmeing TM, Kelley AC, Ramakrishnan V: **The mechanism for activation of GTP hydrolysis on the ribosome**. *Science* 2010, **330**(6005):835-838.
128. Calixto AR, Moreira C, Kamerlin SCL: **Recent Advances in Understanding Biological GTP Hydrolysis through Molecular Simulation**. *ACS Omega* 2020, **5**(9):4380-4385.
129. Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG, Sprang SR: **Structures of active conformations of Gi alpha 1 and the mechanism of GTP hydrolysis**. *Science* 1994, **265**(5177):1405-1412.
130. Sondek J, Lambright DG, Noel JP, Hamm HE, Sigler PB: **GTPase mechanism of Gproteins from the 1.7-A crystal structure of transducin alpha-GDP-AIF-4**. *Nature* 1994, **372**(6503):276-279.
131. Wittinghofer A: **Signaling mechanistics: aluminum fluoride for molecule of the year**. *Curr Biol* 1997, **7**(11):R682-685.
132. Menz RI, Walker JE, Leslie AG: **Structure of bovine mitochondrial $F_1$-ATPase with nucleotide bound to all three catalytic sites: implications for the mechanism of rotary catalysis**. *Cell* 2001, **106**(3):331-341.
133. Davies DR, Hol WG: **The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes**. *FEBS Lett* 2004, **577**(3):315-321.
134. Jin Y, Richards NG, Waltho JP, Blackburn GM: **Metal fluorides as analogues for studies on phosphoryl transfer enzymes**. *Angew Chem Int Ed Engl* 2017, **56**(15):4110-4128.
135. Menetrey J, Llinas P, Mukherjea M, Sweeney HL, Houdusse A: **The structural basis for the large powerstroke of myosin VI**. *Cell* 2007, **131**(2):300-308.
136. Knowles JR: **Enzyme-catalyzed phosphoryl transfer reactions**. *Annu Rev Biochem* 1980, **49**:877-919.
137. Shabarova ZA, Bogdanov AA: **Advanced organic chemistry of nucleic acids**: John Wiley & Sons; 2008.
138. Bowler MW, Cliff MJ, Waltho JP, Blackburn GM: **Why did Nature select phosphate for its dominant roles in biology?** *New Journal of Chemistry* 2010, **34**(5):784-794.
139. Lassila JK, Zalatan JG, Herschlag D: **Biological phosphoryl-transfer reactions: understanding mechanism and catalysis**. *Annual Review of Biochemistry* 2011, **80**:669-702.
140. Kiani FA, Fischer S: **Comparing the catalytic strategy of ATP hydrolysis in biomolecular motors**. *Phys Chem Chem Phys* 2016, **18**(30):20219-20233.
141. Daigle DM, Brown ED: **Studies of the interaction of *Escherichia coli* YjeQ with the ribosome in vitro**. *J Bacteriol* 2004, **186**(5):1381-1387.

142. Perez-Arellano I, Spinola-Amilibia M, Bravo J: **Human Drg1 is a potassium-dependent GTPase enhanced by Lerepo4**. *FEBS J* 2013, **280**(15):3647-3657.
143. Conway TW: **On the role of ammonium or potassium ion in amino acid polymerization**. *Proc Natl Acad Sci USA* 1964, **51**:1216-1220.
144. Conway TW, Lipmann F: **Characterization of a Ribosome-Linked Guanosine Triphosphatase in Escherichia Coli Extracts**. *Proc Natl Acad Sci U S A* 1964, **52**:1462-1469.
145. Lubin M, Ennis HL: **On the Role of Intracellular Potassium in Protein Synthesis**. *Biochim Biophys Acta* 1964, **80**:614-631.
146. Ash MR, Maher MJ, Guss JM, Jormakka M: **The initiation of GTP hydrolysis by the G-domain of FeoB: insights from a transition-state complex structure**. *PLoS One* 2011, **6**(8):e23355.
147. Mishra R, Gara SK, Mishra S, Prakash B: **Analysis of GTPases carrying hydrophobic amino acid substitutions in lieu of the catalytic glutamine: implications for GTP hydrolysis**. *Proteins* 2005, **59**(2):332-338.
148. Yamanaka K, Hwang J, Inouye M: **Characterization of GTPase activity of TrmE, a member of a novel GTPase superfamily, from Thermotoga maritima**. *J Bacteriol* 2000, **182**(24):7078-7082.
149. Moreau M, Lee GI, Wang Y, Crane BR, Klessig DF: **AtNOS/AtNOA1 is a functional Arabidopsis thaliana cGTPase and not a nitric-oxide synthase**. *J Biol Chem* 2008, **283**(47):32957-32967.
150. Bohme S, Meyer S, Kruger A, Steinhoff HJ, Wittinghofer A, Klare JP: **Stabilization of G domain conformations in the tRNA-modifying MnmE-GidA complex observed with double electron electron resonance spectroscopy**. *J Biol Chem* 2010, **285**(22):16991-17000.
151. Skulachev VP: **Membrane-linked energy buffering as the biological function of Na+/K+ gradient**. *FEBS letters* 1978, **87**(2):171-179.
152. Drever JI, Marion G: **The geochemistry of natural waters: surface and groundwater environments**. *Journal of Environmental Quality* 1998, **27**(1):245-245.
153. Oren A: **Thermodynamic limits to microbial life at high salt concentrations**. *Environ Microbiol* 2011, **13**(8):1908-1923.
154. Yan L, Ma Y, Sun Y, Gao J, Chen X, Liu J, Wang C, Rao Z, Lou Z: **Structural basis for mechanochemical role of Arabidopsis thaliana dynamin-related protein in membrane fission**. *J Mol Cell Biol* 2011, **3**(6):378-381.
155. Villarroya M, Prado S, Esteve JM, Soriano MA, Aguado C, Perez-Martinez D, Martinez-Ferrandis JI, Yim L, Victor VM, Cebolla E *et al*: **Characterization of human GTPBP3, a GTP-binding protein involved in mitochondrial tRNA modification**. *Mol Cell Biol* 2008, **28**(24):7514-7531.
156. Sehorn MG, Sigurdsson S, Bussen W, Unger VM, Sung P: **Human meiotic recombinase Dmc1 promotes ATP-dependent homologous DNA strand exchange**. *Nature* 2004, **429**(6990):433-437.
157. Liu Y, Stasiak AZ, Masson JY, McIlwraith MJ, Stasiak A, West SC: **Conformational changes modulate the activity of human RAD51 protein**. *J Mol Biol* 2004, **337**(4):817-827.
158. Rice KP, Eggler AL, Sung P, Cox MM: **DNA pairing and strand exchange by the *Escherichia coli* RecA and yeast Rad51 proteins without ATP hydrolysis: on the importance of not getting stuck**. *J Biol Chem* 2001, **276**(42):38570-38581.
159. Amunugama R, He Y, Willcox S, Forties RA, Shim KS, Bundschuh R, Luo Y, Griffith J, Fishel R: **RAD51 protein ATP cap regulates nucleoprotein filament stability**. *J Biol Chem* 2012, **287**(12):8724-8736.
160. Li Y, He Y, Luo Y: **Conservation of a conformational switch in RadA recombinase from Methanococcus maripaludis**. *Acta Crystallogr D Biol Crystallogr* 2009, **65**(Pt 6):602-610.
161. Voorhees RM, Ramakrishnan V: **Structural basis of the translational elongation cycle**. *Annual Review of Biochemistry* 2013, **82**:203-236.
162. Spirin A: **Ribosomes**: Springer Science & Business Media; 2013.
163. Lin J, Gagnon MG, Bulkley D, Steitz TA: **Conformational changes of elongation factor G on the ribosome during tRNA translocation**. *Cell* 2015, **160**(1-2):219-227.
164. Gagnon MG, Lin J, Bulkley D, Steitz TA: **Crystal structure of elongation factor 4 bound to a clockwise ratcheted ribosome**. *Science* 2014, **345**(6197):684-687.

165. Ero R, Kumar V, Chen Y, Gao YG: **Similarity and diversity of translational GTPase factors EF-G, EF4, and BipA: From structure to function**. *RNA Biol* 2016, **13**(12):1258-1273.
166. Maracci C, Rodnina MV: **Review: Translational GTPases**. *Biopolymers* 2016, **105**(8):463-475.
167. Fischer N, Neumann P, Bock LV, Maracci C, Wang Z, Paleskava A, Konevega AL, Schroder GF, Grubmuller H, Ficner R *et al*: **The pathway to GTPase activation of elongation factor SelB on the ribosome**. *Nature* 2016, **540**(7631):80-85.
168. Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H: **Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM**. *Nature* 2015, **520**(7548):567-570.
169. Frank J: **The translation elongation cycle-capturing multiple states by cryo-electron microscopy**. *Philos Trans R Soc Lond B Biol Sci* 2017, **372**(1716).
170. Rodnina MV, Fischer N, Maracci C, Stark H: **Ribosome dynamics during decoding**. *Philos Trans R Soc Lond B Biol Sci* 2017, **372**(1716).
171. Burroughs AM, Aravind L: **The Origin and Evolution of Release Factors: Implications for Translation Termination, Ribosome Rescue, and Quality Control Pathways**. *Int J Mol Sci* 2019, **20**(8).
172. Itoh Y, Sekine S, Yokoyama S: **Crystal structure of the full-length bacterial selenocysteine-specific elongation factor SelB**. *Nucleic Acids Res* 2015, **43**(18):9028-9038.
173. Dubey A, Copeland PR: **The Selenocysteine-Specific Elongation Factor Contains Unique Sequences That Are Required for Both Nuclear Export and Selenocysteine Incorporation**. *PLoS One* 2016, **11**(11):e0165642.
174. Zinoviev A, Kuroha K, Pestova TV, Hellen CUT: **Two classes of EF1-family translational GTPases encoded by giant viruses**. *Nucleic Acids Res* 2019, **47**(11):5761-5776.
175. Koonin EV: **How many genes can make a cell: the minimal-gene-set concept**. *Annu Rev Genomics Hum Genet* 2000, **1**:99-116.
176. Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T: **Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes**. *Proc Natl Acad Sci USA* 1989, **86**(23):9355-9359.
177. Baldauf SL, Palmer JD, Doolittle WF: **The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny**. *Proc Natl Acad Sci USA* 1996, **93**(15):7749-7754.
178. Mohr D, Wintermeyer W, Rodnina MV: **Arginines 29 and 59 of elongation factor G are important for GTP hydrolysis or translocation on the ribosome**. *EMBO J* 2000, **19**(13):3458-3464.
179. Ban N, Nissen P, Hansen J, Capel M, Moore PB, Steitz TA: **Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit**. *Nature* 1999, **400**(6747):841-847.
180. Gutell RR, Schnare MN, Gray MW: **A compilation of large subunit (23S- and 23S-like) ribosomal RNA structures**. *Nucleic Acids Research* 1992, **20 Suppl**:2095-2109.
181. Loveland AB, Demo G, Grigorieff N, Korostelev AA: **Ensemble cryo-EM elucidates the mechanism of translation fidelity**. *Nature* 2017, **546**(7656):113-117.
182. Loveland AB, Demo G, Korostelev AA: **Cryo-EM of elongating ribosome with EF-Tu*GTP elucidates tRNA proofreading**. *Nature* 2020.
183. Tourigny DS, Fernandez IS, Kelley AC, Ramakrishnan V: **Elongation factor G bound to the ribosome in an intermediate state of translocation**. *Science* 2013, **340**(6140):1235490.
184. Adamczyk AJ, Warshel A: **Converting structural information into an allosteric-energy-based picture for elongation factor Tu activation by the ribosome**. *P Natl Acad Sci USA* 2011, **108**(24):9827-9832.
185. Beaudry P, Sander G, Grunberg-Manago M, Douzou P: **Cation-induced regulatory mechanism of GTPase activity dependent on polypeptide initiation factor 2**. *Biochemistry* 1979, **18**(1):202-207.
186. Ivell R, Fasano O, Crechet JB, Parmeggiani A: **Characterization of a kirromycin-resistant elongation factor Tu from Escherichia coli**. *Biochemistry* 1981, **20**(5):1355-1361.
187. Parmeggiani A, Sander G: **Properties and regulation of the GTPase activities of elongation factors Tu and G, and of initiation factor 2**. *Mol Cell Biochem* 1981, **35**(3):129-158.

188. Kuhle B: **Structural and functional studies on GTPases involved in eukaryal translation initiation**. Göttingen: Georg-August University; 2014.
189. Maracci C, Peske F, Dannies E, Pohl C, Rodnina MV: **Ribosome-induced tuning of GTP hydrolysis by a translational GTPase**. *Proc Natl Acad Sci U S A* 2014, **111**(40):14418-14423.
190. Mondal D, Warshel A: **EF-Tu and EF-G are activated by allosteric effects**. *Proc Natl Acad Sci U S A* 2018, **115**(13):3386-3391.
191. Jin Y, Molt RW, Jr., Waltho JP, Richards NG, Blackburn GM: **$^{19}$F NMR and DFT analysis reveal structural and electronic transition state features for RhoA-catalyzed GTP hydrolysis**. *Angew Chem Int Ed Engl* 2016, **55**(10):3318-3322.
192. Molt RW, Jr., Pellegrini E, Jin Y: **A GAP-GTPase-GDP-Pi Intermediate Crystal Structure Analyzed by DFT Shows GTP Hydrolysis Involves Serial Proton Transfers**. *Chemistry* 2019, **25**(36):8484-8488.
193. Prasad BR, Plotnikov NV, Lameira J, Warshel A: **Quantitative exploration of the molecular origin of the activation of GTPase**. *P Natl Acad Sci USA* 2013, **110**(51):20509-20514.
194. Rudack T, Xia F, Schlitter J, Kotting C, Gerwert K: **Ras and GTPase-activating protein (GAP) drive GTP into a precatalytic state as revealed by combining FTIR and biomolecular simulations**. *Proc Natl Acad Sci USA* 2012, **109**(38):15295-15300.
195. Mann D, Teuber C, Tennigkeit SA, Schroter G, Gerwert K, Kotting C: **Mechanism of the intrinsic arginine finger in heterotrimeric G proteins**. *Proc Natl Acad Sci USA* 2016, **113**(50):E8041-E8050.
196. Gremer L, Gilsbach B, Ahmadian MR, Wittinghofer A: **Fluoride complexes of oncogenic Ras mutants to study the Ras-RasGap interaction**. *Biol Chem* 2008, **389**(9):1163-1171.
197. Cohn M, Hughes TR, Jr.: **Nuclear magnetic resonance spectra of adenosine di- and triphosphate. II. Effect of complexing with divalent metal ions**. *J Biol Chem* 1962, **237**:176-181.
198. Jiang L, Mao XA: **Conformation of adenosine-5 '-triphosphate in the presence of Mg2+ at different pH**. *Polyhedron* 2002, **21**(4):435-438.
199. Huang SL, Tsai M-D: **Does the magnesium(II) ion interact with the α-phosphate of adenosine-triphosphate? An investigation by oxygen-17 nuclear magnetic resonance**. *Biochemistry* 1982, **21**(5):951-959.
200. Cowan JA: **Metallobiochemistry of Magnesium - Coordination-Complexes with Biological Substrates - Site Specificity, Kinetics and Thermodynamics of Binding, and Implications for Activity**. *Inorganic Chemistry* 1991, **30**(13):2740-2747.
201. Abrahams JP, Leslie AG, Lutter R, Walker JE: **Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria**. *Nature* 1994, **370**(6491):621-628.
202. Schweins T, Wittinghofer A: **GTP-binding proteins. Structures, interactions and relationships**. *Curr Biol* 1994, **4**(6):547-550.
203. Harding MM: **Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium**. *Acta Crystallogr D Biol Crystallogr* 2002, **58**(Pt 5):872-874.
204. Harding MM: **The architecture of metal coordination groups in proteins**. *Acta Crystallogr D Biol Crystallogr* 2004, **60**(Pt 5):849-859.
205. Sigel A, Sigel H, Sigel RKO (eds.): **The Alkali Metal Ions: Their Role for Life**: Springer; 2016.
206. Smith R, Martell A, Chen Y: **Critical evaluation of stability constants for nucleotide complexes with protons and metal ions and the accompanying enthalpy changes**. *Pure Appl Chem* 1991, **63**(7):1015-1080.
207. De Stefano C, Milea D, Pettignano A, Sammartano S: **Modeling ATP protonation and activity coefficients in NaCl aq and KCl aq by SIT and Pitzer equations**. *Biophysical chemistry* 2006, **121**(2):121-130.
208. Stellwagen E, Stellwagen NC: **Quantitative analysis of cation binding to the adenosine nucleotides using the variable ionic strength method: Validation of the Debye–Hückel–Onsager theory of electrophoresis in the absence of counterion binding**. *Electrophoresis* 2007, **28**(7):1053-1062.
209. Warshel A, Prasad BR: **Perspective on Computer Modelling of Enzymatic Reactions**. In: *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis*. Edited by Tunon I, Moliner V. London: Royal Society of Chemistry; 2017: 1-30.

210. Akola J, Jones RO: **ATP hydrolysis in water - A density functional study**. *Journal of Physical Chemistry B* 2003, **107**(42):11774-11783.
211. Grigorenko BL, Rogov AV, Nemukhin AV: **Mechanism of triphosphate hydrolysis in aqueous solution: QM/MM simulations in water clusters**. *J Phys Chem B* 2006, **110**(9):4407-4412.
212. Harrison CB, Schulten K: **Quantum and classical dynamics simulations of ATP hydrolysis in solution**. *J Chem Theory Comput* 2012, **8**(7):2328-2335.
213. Mildvan AS: **Role of magnesium and other divalent cations in ATP-utilizing enzymes**. *J Am Coll Nutr* 1987, **6**(1):28-33.
214. Chaudhry C, Farr GW, Todd MJ, Rye HS, Brunger AT, Adams PD, Horwich AL, Sigler PB: **Role of the gamma-phosphate of ATP in triggering protein folding by GroEL-GroES: function, structure and energetics**. *EMBO J* 2003, **22**(19):4877-4887.
215. Wang J, Boisvert DC: **Structural basis for GroEL-assisted protein folding from the crystal structure of (GroEL-KMgATP)14 at 2.0A resolution**. *J Mol Biol* 2003, **327**(4):843-855.
216. Kotting C, Gerwert K: **Time-resolved FTIR studies provide activation free energy, activation enthalpy and activation entropy for GTPase reactions**. *Chem Phys* 2004, **307**(2-3):227-232.
217. Shutes A, Der CJ: **Real-time in vitro measurement of intrinsic and Ras GAP-mediated GTP hydrolysis**. *Methods Enzymol* 2006, **407**:9-22.
218. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztanyi Z, El-Gebali S, Fraser M *et al*: **InterPro in 2017-beyond protein family and domain annotations**. *Nucleic Acids Res* 2017, **45**(D1):D190-D199.
219. Golicnik M: **Metallic fluoride complexes as phosphate analogues for structural and mechanistic studies of phosphoryl group transfer enzymes**. *Acta Chim Slov* 2010, **57**(2):272-287.
220. Koenig P, Oreb M, Hofle A, Kaltofen S, Rippe K, Sinning I, Schleiff E, Tews I: **The GTPase cycle of the chloroplast import receptors Toc33/Toc34: Implications from monomeric and dimeric structures**. *Structure* 2008, **16**(4):585-596.
221. Gasper R, Meyer S, Gotthardt K, Sirajuddin M, Wittinghofer A: **It takes two to tango: regulation of G proteins by dimerization**. *Nature reviews Molecular cell biology* 2009, **10**(6):423-429.
222. Cherfils J, Zeghouf M: **Regulation of small GTPases by GEFs, GAPs, and GDIs**. *Physiol Rev* 2013, **93**(1):269-309.
223. Chen Z, Yang H, Pavletich NP: **Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures**. *Nature* 2008, **453**(7194):489-484.
224. Walker J: **ATP synthesis by rotary catalysis**. *Angew Chem Int Ed Engl* 1998, **37**:2309-2319.
225. Senior AE, Nadanaciva S, Weber J: **The molecular mechanism of ATP synthesis by $F_1F_0$-ATP synthase**. *Biochim Biophys Acta* 2002, **1553**(3):188-211.
226. Skordalakes E, Berger JM: **Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor**. *Cell* 2006, **127**(3):553-564.
227. Komoriya Y, Ariga T, Iino R, Imamura H, Okuno D, Noji H: **Principal role of the arginine finger in rotary catalysis of $F_1$-ATPase**. *J Biol Chem* 2012, **287**(18):15134-15142.
228. Goitre L, Trapani E, Trabalzini L, Retta SF: **The Ras superfamily of small GTPases: the unlocked secrets**. *Methods Mol Biol* 2014, **1120**:1-18.
229. Meier TI, Peery RB, McAllister KA, Zhao G: **Era GTPase of *Escherichia coli*: binding to 16S rRNA and modulation of GTPase activity by RNA and carbohydrates**. *Microbiology* 2000, **146**(Pt 5):1071-1083.
230. Tu C, Zhou X, Tarasov SG, Tropea JE, Austin BP, Waugh DS, Court DL, Ji X: **The Era GTPase recognizes the GAUCACCUCC sequence and binds helix 45 near the 3' end of 16S rRNA**. *Proc Natl Acad Sci USA* 2011, **108**(25):10156-10161.
231. Klare JP: **Site-directed spin labeling EPR spectroscopy in protein research**. *Biol Chem* 2013, **394**(10):1281-1300.
232. Tu C, Zhou X, Tropea JE, Austin BP, Waugh DS, Court DL, Ji X: **Structure of ERA in complex with the 3' end of 16S rRNA: implications for ribosome biogenesis**. *Proc Natl Acad Sci USA* 2009, **106**(35):14843-14848.
233. Qian X, He Y, Wu Y, Luo Y: **Asp302 determines potassium dependence of a RadA recombinase from *Methanococcus voltae***. *J Mol Biol* 2006, **360**(3):537-547.
234. Auffinger P, Grover N, Westhof E: **Metal ion binding to RNA**. *Met Ions Life Sci* 2011, **9**:1-35.

235. Leonarski F, D'Ascenzo L, Auffinger P: **Nucleobase carbonyl groups are poor Mg(2+) inner-sphere binders but excellent monovalent ion binders-a critical PDB survey**. *RNA* 2019, **25**(2):173-192.
236. Parmeggiani A, Krab IM, Okamura S, Nielsen RC, Nyborg J, Nissen P: **Structural basis of the action of pulvomycin and GE2270 A on elongation factor Tu**. *Biochemistry* 2006, **45**(22):6846-6857.
237. Kummer E, Leibundgut M, Rackham O, Lee RG, Boehringer D, Filipovska A, Ban N: **Unique features of mammalian mitochondrial translation initiation revealed by cryo-EM**. *Nature* 2018, **560**(7717):263-267.
238. Auffinger P, D'Ascenzo L, Ennifar E: **Sodium and Potassium Interactions with Nucleic Acids**. *Met Ions Life Sci* 2016, **16**:167-201.
239. Chen Y, Feng S, Kumar V, Ero R, Gao YG: **Structure of EF-G-ribosome complex in a pretranslocation state**. *Nature Structural & Molecular Biology* 2013, **20**(9):1077-1084.
240. Rodnina MV, Peske F, Peng BZ, Belardinelli R, Wintermeyer W: **Converting GTP hydrolysis into motion: versatile translational elongation factor G**. *Biol Chem* 2019, **401**(1):131-142.
241. Piepenburg O, Pape T, Pleiss JA, Wintermeyer W, Uhlenbeck OC, Rodnina MV: **Intact aminoacyl-tRNA is required to trigger GTP hydrolysis by elongation factor Tu on the ribosome**. *Biochemistry* 2000, **39**(7):1734-1738.
242. Cochella L, Green R: **An active role for tRNA in decoding beyond codon:anticodon pairing**. *Science* 2005, **308**(5725):1178-1180.
243. Schmeing TM, Voorhees RM, Kelley AC, Ramakrishnan V: **How mutations in tRNA distant from the anticodon affect the fidelity of decoding**. *Nature Structural & Molecular Biology* 2011, **18**(4):432-436.
244. Agmon I, Bashan A, Zarivach R, Yonath A: **Symmetry at the active site of the ribosome: structural and functional implications**. *Biol Chem* 2005, **386**(9):833-844.
245. Ramakrishnan U, Hadly EA: **Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies**. *Molecular Ecology* 2009, **18**(7):1310-1330.
246. Bokov K, Steinberg SV: **A hierarchical model for evolution of 23S ribosomal RNA**. *Nature* 2009, **457**(7232):977-980.
247. Fox GE: **Origin and evolution of the ribosome**. *Cold Spring Harb Perspect Biol* 2010, **2**(9):a003483.
248. Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, Fox GE, Harvey SC, Wartell RM, Hud NV *et al*: **History of the ribosome and the origin of translation**. *Proc Natl Acad Sci U S A* 2015, **112**(50):15396-15401.
249. Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG, Harvey SC, Fox GE, Wartell RM: **Evolution of the ribosome at atomic resolution**. *Proceedings of the National Academy of Sciences* 2014, **111**(28):10251-10256.
250. Mirkin BG, Fenner TI, Galperin MY, Koonin EV: **Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes**. *BMC Evol Biol* 2003, **3**:2.
251. Murzin AG, Brenner SE, Hubbard T, Chothia C: **SCOP: a structural classification of proteins database for the investigation of sequences and structures**. *J Mol Biol* 1995, **247**(4):536-540.
252. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG *et al*: **CATH: comprehensive structural and functional annotations for genome sequences**. *Nucleic Acids Research* 2015, **43**(Database issue):D376-381.
253. Schaeffer RD, Liao Y, Cheng H, Grishin NV: **ECOD: new developments in the evolutionary classification of domains**. *Nucleic Acids Research* 2017, **45**(D1):D296-D302.
254. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: **The Pfam protein families database: towards a more sustainable future**. *Nucleic Acids Res* 2016, **44**(D1):D279-D285.
255. Delbaere LT, Sudom AM, Prasad L, Leduc Y, Goldie H: **Structure/function studies of phosphoryl transfer by phosphoenolpyruvate carboxykinase**. *Biochim Biophys Acta* 2004, **1697**(1-2):271-278.

256. Matte A, Tari LW, Delbaere LT: **How do kinases transfer phosphoryl groups?** *Structure* 1998, **6**(4):413-419.
257. Maegley KA, Admiraal SJ, Herschlag D: **Ras-catalyzed hydrolysis of GTP: a new perspective from model studies**. *Proc Natl Acad Sci U S A* 1996, **93**(16):8160-8166.
258. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: **Jalview Version 2--a multiple sequence alignment editor and analysis workbench**. *Bioinformatics* 2009, **25**(9):1189-1191.
259. Kanade M, Chakraborty S, Shelke SS, Gayathri P: **A Distinct Motif in a Prokaryotic Small Ras-Like GTPase Highlights Unifying Features of Walker B Motifs in P-Loop NTPases**. *J Mol Biol* 2020, **432**(20):5544-5564.
260. Gasper R, Wittinghofer F: **The Ras switch in structural and historical perspective**. *Biol Chem* 2019, **401**(1):143-163.
261. Sirajuddin M, Farkasovsky M, Zent E, Wittinghofer A: **GTP-induced conformational changes in septins and implications for function**. *Proc Natl Acad Sci USA* 2009, **106**(39):16592-16597.
262. Byrnes LJ, Singh A, Szeto K, Benvin NM, O'Donnell JP, Zipfel WR, Sondermann H: **Structural basis for conformational switching and GTP loading of the large G protein atlastin**. *EMBO J* 2013, **32**(3):369-384.
263. Wittinghofer A: **GTP and ATP hydrolysis in biology**. *Biopolymers* 2016, **105**(8):419-421.
264. Ligeti E, Welti S, Scheffzek K: **Inhibition and termination of physiological responses by GTPase activating proteins**. *Physiol Rev* 2012, **92**(1):237-272.
265. Fisher AJ, Smith CA, Thoden JB, Smith R, Sutoh K, Holden HM, Rayment I: **X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.BeFx and MgADP.AlF4**. *Biochemistry* 1995, **34**(28):8960-8972.
266. Geeves MA: **Review: The ATPase mechanism of myosin and actomyosin**. *Biopolymers* 2016, **105**(8):483-491.
267. Cross RA: **Review: Mechanochemistry of the kinesin-1 ATPase**. *Biopolymers* 2016, **105**(8):476-482.
268. Bange G, Sinning I: **SIMIBI twins in protein targeting and localization**. *Nat Struct Mol Biol* 2013, **20**(7):776-780.
269. Lutkenhaus J: **The ParA/MinD family puts things in their place**. *Trends in Microbiology* 2012, **20**(9):411-418.
270. Ataide SF, Schmitz N, Shen K, Ke A, Shan SO, Doudna JA, Ban N: **The crystal structure of the signal recognition particle in complex with its receptor**. *Science* 2011, **331**(6019):881-886.
271. Cheek S, Ginalski K, Zhang H, Grishin NV: **A comprehensive update of the sequence and structure classification of kinases**. *Bmc Struct Biol* 2005, **5**.
272. Kenyon CP, Roth RL, van der Westhuyzen CW, Parkinson CJ: **Conserved phosphoryl transfer mechanisms within kinase families and the role of the C8 proton of ATP in the activation of phosphoryl transfer**. *BMC research notes* 2012, **5**:131.
273. Lansdon EB, Segel IH, Fisher AJ: **Ligand-induced structural changes in adenosine 5'-phosphosulfate kinase from Penicillium chrysogenum**. *Biochemistry* 2002, **41**(46):13672-13680.
274. Kerns SJ, Agafonov RV, Cho YJ, Pontiggia F, Otten R, Pachov DV, Kutter S, Phung LA, Murphy PN, Thai V *et al*: **The energy landscape of adenylate kinase during catalysis**. *Nat Struct Mol Biol* 2015, **22**(2):124-131.
275. Moser J, Lange C, Krausze J, Rebelein J, Schubert WD, Ribbe MW, Heinz DW, Jahn D: **Structure of ADP-aluminium fluoride-stabilized protochlorophyllide oxidoreductase complex**. *Proc Natl Acad Sci U S A* 2013, **110**(6):2094-2098.
276. Voigts-Hoffmann F, Schmitz N, Shen K, Shan SO, Ataide SF, Ban N: **The structural basis of FtsY recruitment and GTPase activation by SRP RNA**. *Mol Cell* 2013, **52**(5):643-654.
277. Poyraz O, Brunner K, Lohkamp B, Axelsson H, Hammarstrom LG, Schnell R, Schneider G: **Crystal structures of the kinase domain of the sulfate-activating complex in Mycobacterium tuberculosis**. *PLoS One* 2015, **10**(3):e0121494.
278. Thomsen ND, Berger JM: **Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases**. *Mol Microbiol* 2008, **69**(5):1071-1090.

279. Krishnan A, Burroughs AM, Iyer LM, Aravind L: **Comprehensive classification of ABC ATPases and their functional radiation in nucleoprotein dynamics and biological conflict systems**. *Nucleic Acids Research* 2020, **48**(18):10045-10075.
280. Longo LM, Jablonska J, Vyas P, Kanade M, Kolodny R, Ben-Tal N, Tawfik DS: **On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment**. *Elife* 2020, **9**.
281. Iyer LM, Leipe DD, Koonin EV, Aravind L: **Evolutionary history and higher order classification of AAA+ ATPases**. *J Struct Biol* 2004, **146**(1-2):11-31.
282. Iyer LM, Makarova KS, Koonin EV, Aravind L: **Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging**. *Nucleic Acids Research* 2004, **32**(17):5260-5279.
283. Aravind L, Iyer LM, Leipe DD, Koonin EV: **A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain**. *Genome Biol* 2004, **5**(5):R30.
284. Ammelburg M, Frickey T, Lupas AN: **Classification of AAA+ proteins**. *Journal of Structural Biology* 2006, **156**(1):2-11.
285. Karata K, Inagawa T, Wilkinson AJ, Tatsuta T, Ogura T: **Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. Site-directed mutagenesis of the ATP-dependent protease FtsH**. *J Biol Chem* 1999, **274**(37):26225-26232.
286. Leipe DD, Koonin EV, Aravind L: **STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer**. *J Mol Biol* 2004, **343**(1):1-28.
287. Cheng TC, Hong C, Akey IV, Yuan S, Akey CW: **A near atomic structure of the active human apoptosome**. *Elife* 2016, **5**.
288. Li P, Nijhawan D, Budihardjo I, Srinivasula SM, Ahmad M, Alnemri ES, Wang X: **Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade**. *Cell* 1997, **91**(4):479-489.
289. Shalaeva DN, Dibrova DV, Galperin MY, Mulkidjanian AY: **Modeling of interaction between cytochrome c and the WD domains of Apaf-1: bifurcated salt bridges underlying apoptosome assembly**. *Biol Direct* 2015, **10**:29.
290. Zhou M, Li Y, Hu Q, Bai XC, Huang W, Yan C, Scheres SH, Shi Y: **Atomic structure of the apoptosome: mechanism of cytochrome c- and dATP-mediated activation of Apaf-1**. *Genes Dev* 2015, **29**(22):2349-2361.
291. Singleton MR, Dillingham MS, Wigley DB: **Structure and mechanism of helicases and nucleic acid translocases**. *Annual Review of Biochemistry* 2007, **76**:23-50.
292. Sengoku T, Nureki O, Nakamura A, Satoru KI, Yokoyama S: **Structural basis for RNA unwinding by the DEAD-box protein Drosophila vasa**. *Cell* 2006, **125**(2):287-300.
293. Rees DC, Johnson E, Lewinson O: **ABC transporters: the power to change**. *Nat Rev Mol Cell Biol* 2009, **10**(3):218-227.
294. Dean M, Rzhetsky A, Allikmets R: **The human ATP-binding cassette (ABC) transporter superfamily**. *Genome Res* 2001, **11**(7):1156-1166.
295. Kerr ID: **Sequence analysis of twin ATP binding cassette proteins involved in translational control, antibiotic resistance, and ribonuclease L inhibition**. *Biochem Biophys Res Commun* 2004, **315**(1):166-173.
296. Decottignies A, Goffeau A: **Complete inventory of the yeast ABC proteins**. *Nat Genet* 1997, **15**(2):137-145.
297. Hopfner KP, Karcher A, Shin DS, Craig L, Arthur LM, Carney JP, Tainer JA: **Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily**. *Cell* 2000, **101**(7):789-800.
298. Eisen JA: **A phylogenomic study of the MutS family of proteins**. *Nucleic Acids Research* 1998, **26**(18):4291-4300.
299. Oldham ML, Chen J: **Snapshots of the maltose transporter during ATP hydrolysis**. *Proc Natl Acad Sci U S A* 2011, **108**(37):15152-15156.
300. Rees DM, Leslie AG, Walker JE: **The structure of the membrane extrinsic region of bovine ATP synthase**. *Proc Natl Acad Sci USA* 2009, **106**(51):21597-21601.

301. Savvides SN, Yeo HJ, Beck MR, Blaesing F, Lurz R, Lanka E, Buhrdorf R, Fischer W, Haas R, Waksman G: **VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion**. *EMBO J* 2003, **22**(9):1969-1980.
302. Satyshur KA, Worzalla GA, Meyer LS, Heiniger EK, Aukema KG, Misic AM, Forest KT: **Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility**. *Structure* 2007, **15**(3):363-376.
303. Mancl JM, Black WP, Robinson H, Yang Z, Schubot FD: **Crystal Structure of a Type IV Pilus Assembly ATPase: Insights into the Molecular Mechanism of PilB from Thermus thermophilus**. *Structure* 2016, **24**(11):1886-1897.
304. McCallum M, Benlekbir S, Nguyen S, Tammam S, Rubinstein JL, Burrows LL, Howell PL: **Multiple conformations facilitate PilT function in the type IV pilus**. *Nat Commun* 2019, **10**(1):5198.
305. Jean NL, Rutherford TJ, Lowe J: **FtsK in motion reveals its mechanism for double-stranded DNA translocation**. *Proc Natl Acad Sci U S A* 2020, **117**(25):14202-14208.
306. Dong Y, Zhang S, Wu Z, Li X, Wang WL, Zhu Y, Stoilova-McPhie S, Lu Y, Finley D, Mao Y: **Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome**. *Nature* 2019, **565**(7737):49-55.
307. Gai D, Zhao R, Li D, Finkielstein CV, Chen XS: **Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen**. *Cell* 2004, **119**(1):47-60.
308. Lu K-Y, Chen W-F, Rety S, Liu N-N, Wu W-Q, Dai Y-X, Li D, Ma H-Y, Dou S-X, Xi X-G: **Insights into the structural and mechanistic basis of multifunctional S. cerevisiae Pif1p helicase**. *Nucleic Acids Research* 2018, **46**(3):1486-1500.
309. Lawson MR, Ma W, Bellecourt MJ, Artsimovitch I, Martin A, Landick R, Schulten K, Berger JM: **Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein**. *Mol Cell* 2018, **71**(6):911-922 e914.
310. Itsathitphaisarn O, Wing RA, Eliason WK, Wang J, Steitz TA: **The hexameric helicase DnaB adopts a nonplanar conformation during translocation**. *Cell* 2012, **151**(2):267-277.
311. Singleton MR, Sawaya MR, Ellenberger T, Wigley DB: **Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides**. *Cell* 2000, **101**(6):589-600.
312. Yamaichi Y, Niki H: **Active segregation by the Bacillus subtilis partitioning system in Escherichia coli**. *Proc Natl Acad Sci USA* 2000, **97**(26):14656-14661.
313. Abe J, Hiyama TB, Mukaiyama A, Son S, Mori T, Saito S, Osako M, Wolanin J, Yamashita E, Kondo T *et al*: **Circadian rhythms. Atomic-scale origins of slowness in the cyanobacterial circadian clock**. *Science* 2015, **349**(6245):312-316.
314. Pasqualato S, Cherfils J: **Crystallographic evidence for substrate-assisted GTP hydrolysis by a small GTP binding protein**. *Structure* 2005, **13**(4):533-540.
315. Muneyuki E, Noji H, Amano T, Masaike T, Yoshida M: **F0F1-ATP synthase: general structural features of 'ATP-engine' and a problem on free energy transduction**. *Bba-Bioenergetics* 2000, **1458**(2-3):467-481.
316. Majumdar S, Acharya A, Prakash B: **Structural plasticity mediates distinct GAP-dependent GTP hydrolysis mechanisms in Rab33 and Rab5**. *Febs Journal* 2017, **284**(24):4358-4375.
317. Kotting C, Kallenbach A, Suveyzdis Y, Wittinghofer A, Gerwert K: **The GAP arginine finger movement into the catalytic site of Ras increases the activation entropy**. *Proc Natl Acad Sci U S A* 2008, **105**(17):6260-6265.
318. Katritch V, Cherezov V, Stevens RC: **Structure-function of the G protein-coupled receptor superfamily**. *Annu Rev Pharmacol Toxicol* 2013, **53**:531-556.
319. Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, Hilger D, Lerch MT, Kobilka TS, Thian FS, Hubbell WL *et al*: **Structural Insights into the Dynamic Process of beta2-Adrenergic Receptor Signaling**. *Cell* 2015, **161**(5):1101-1111.
320. Ishchenko A, Gati C, Cherezov V: **Structural biology of G protein-coupled receptors: new opportunities from XFELs and cryoEM**. *Curr Opin Struct Biol* 2018, **51**:44-52.
321. Nordstrom KJV, Almen MS, Edstam MM, Fredriksson R, Schioth HB: **Independent HHsearch, Needleman-Wunsch-Based, and Motif Analyses Reveal the Overall Hierarchy for Most of the G Protein-Coupled Receptor Families**. *Molecular biology and evolution* 2011, **28**(9):2471-2480.

322. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE *et al*: **Crystal structure of rhodopsin: A G protein-coupled receptor**. *Science* 2000, **289**(5480):739-745.
323. Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP: **Crystal structure of the ligand-free G-protein-coupled receptor opsin**. *Nature* 2008, **454**(7201):183-187.
324. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP, Ernst OP: **Crystal structure of opsin in its G-protein-interacting conformation**. *Nature* 2008, **455**(7212):497-502.
325. Kobilka B, Schertler GF: **New G-protein-coupled receptor crystal structures: insights and limitations**. *Trends Pharmacol Sci* 2008, **29**(2):79-83.
326. Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, Han GW, Roth CB, Heitman LH, IJzerman AP *et al*: **Structural basis for allosteric regulation of GPCRs by sodium ions**. *Science* 2012, **337**(6091):232-236.
327. Kruse AC, Li J, Hu J, Kobilka BK, Wess J: **Novel insights into M3 muscarinic acetylcholine receptor physiology and structure**. *J Mol Neurosci* 2014, **53**(3):316-323.
328. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T *et al*: **Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist**. *Nature* 2012, **482**(7386):547-U147.
329. Fenalti G, Giguere PM, Katritch V, Huang XP, Thompson AA, Cherezov V, Roth BL, Stevens RC: **Molecular control of delta-opioid receptor signalling**. *Nature* 2014, **506**(7487):191-196.
330. Miller-Gallacher JL, Nehme R, Warne T, Edwards PC, Schertler GF, Leslie AG, Tate CG: **The 2.1 A resolution structure of cyanopindolol-bound beta1-adrenoceptor identifies an intramembrane Na+ ion that stabilises the ligand-free receptor**. *PLoS One* 2014, **9**(3):e92727.
331. Wacker D, Stevens RC, Roth BL: **How ligands illuminate GPCR molecular pharmacology**. *Cell* 2017, **170**(3):414-427.
332. Knierim B, Hofmann KP, Gartner W, Hubbell WL, Ernst OP: **Rhodopsin and 9-demethyl-retinal analog: effect of a partial agonist on displacement of transmembrane helix 6 in class A G protein-coupled receptors**. *J Biol Chem* 2008, **283**(8):4967-4974.
333. Knierim B, Hofmann KP, Ernst OP, Hubbell WL: **Sequence of late molecular events in the activation of rhodopsin**. *Proc Natl Acad Sci U S A* 2007, **104**(51):20290-20295.
334. Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros JA, Javitch JA: **Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch**. *J Biol Chem* 2002, **277**(43):40989-40996.
335. Eddy MT, Lee MY, Gao ZG, White KL, Didenko T, Horst R, Audet M, Stanczak P, McClary KM, Han GW *et al*: **Allosteric Coupling of Drug Binding and Intracellular Signaling in the A2A Adenosine Receptor**. *Cell* 2018, **172**(1-2):68-80 e12.
336. Yuan S, Hu Z, Filipek S, Vogel H: **W246(6.48) opens a gate for a continuous intrinsic water pathway during activation of the adenosine A2A receptor**. *Angew Chem Int Ed Engl* 2015, **54**(2):556-559.
337. Yuan S, Palczewski K, Peng Q, Kolinski M, Vogel H, Filipek S: **The mechanism of ligand-induced activation or inhibition of mu- and kappa-opioid receptors**. *Angew Chem Int Ed Engl* 2015, **54**(26):7560-7563.
338. Yuan S, Vogel H, Filipek S: **The role of water and sodium ions in the activation of the mu-opioid receptor**. *Angew Chem Int Ed Engl* 2013, **52**(38):10112-10115.
339. Ballesteros JA, Weinstein H: **Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors**. *Methods in neurosciences* 1995, **25**:366-428.
340. Isberg V, de Graaf C, Bortolato A, Cherezov V, Katritch V, Marshal FH, Mordalski S, Pin JP, Stevens RC, Vriend G *et al*: **Generic GPCR residue numbers - aligning topology maps while minding the gaps**. *Trends in Pharmacological Sciences* 2015, **36**(1):22-31.
341. Massink A, Gutierrez-de-Teran H, Lenselink EB, Ortiz Zacarias NV, Xia L, Heitman LH, Katritch V, Stevens RC, IJzerman AP: **Sodium ion binding pocket mutations and adenosine A2A receptor function**. *Mol Pharmacol* 2015, **87**(2):305-313.

342. White KL, Eddy MT, Gao ZG, Han GW, Lian T, Deary A, Patel N, Jacobson KA, Katritch V, Stevens RC: **Structural Connection between Activation Microswitch and Allosteric Sodium Site in GPCR Signaling**. *Structure* 2018, **26**(2):259-269 e255.
343. Pert CB, Pasternak G, Snyder SH: **Opiate agonists and antagonists discriminated by receptor binding in brain**. *Science* 1973, **182**(4119):1359-1361.
344. Horstman DA, Brandon S, Wilson AL, Guyer CA, Cragoe EJ, Jr., Limbird LE: **An aspartate conserved among G-protein receptors confers allosteric regulation of alpha 2-adrenergic receptors by sodium**. *J Biol Chem* 1990, **265**(35):21590-21595.
345. Livingston KE, Traynor JR: **Disruption of the Na+ ion binding site as a mechanism for positive allosteric modulation of the mu-opioid receptor**. *Proc Natl Acad Sci U S A* 2014, **111**(51):18369-18374.
346. Shang Y, LeRouzic V, Schneider S, Bisignano P, Pasternak GW, Filizola M: **Mechanistic insights into the allosteric modulation of opioid receptors by sodium ions**. *Biochemistry* 2014, **53**(31):5140-5149.
347. Hu X, Wang Y, Hunkele A, Provasi D, Pasternak GW, Filizola M: **Kinetic and thermodynamic insights into sodium ion translocation through the μ-opioid receptor from molecular dynamics and machine learning analysis**. *Plos Computational Biology* 2019, **15**(1):e1006689.
348. Ye L, Neale C, Sljoka A, Lyda B, Pichugin D, Tsuchimura N, Larda ST, Pomes R, Garcia AE, Ernst OP *et al*: **Mechanistic insights into allosteric regulation of the A2A adenosine G protein-coupled receptor by physiological cations**. *Nat Commun* 2018, **9**(1):1372.
349. Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS *et al*: **Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor**. *Nature* 2011, **469**(7329):175-180.
350. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hubner H, Pardon E, Valant C, Sexton PM *et al*: **Activation and allosteric modulation of a muscarinic acetylcholine receptor**. *Nature* 2013, **504**(7478):101-106.
351. Huang W, Manglik A, Venkatakrishnan AJ, Laeremans T, Feinberg EN, Sanborn AL, Kato HE, Livingston KE, Thorsen TS, Kling RC *et al*: **Structural insights into micro-opioid receptor activation**. *Nature* 2015, **524**(7565):315-321.
352. Ovchinnikov YA: **Rhodopsin and bacteriorhodopsin: structure-function relationships**. *FEBS Lett* 1982, **148**(2):179-191.
353. Ernst OP, Lodowski DT, Elstner M, Hegemann P, Brown LS, Kandori H: **Microbial and animal rhodopsins: structures, functions, and molecular mechanisms**. *Chem Rev* 2014, **114**(1):126-163.
354. Isom DG, Dohlman HG: **Buried ionizable networks are an ancient hallmark of G protein-coupled receptor activation**. *Proc Natl Acad Sci U S A* 2015, **112**(18):5702-5707.
355. Oesterhelt D, Stoeckenius W: **Rhodopsin-like protein from the purple membrane of Halobacterium halobium**. *Nat New Biol* 1971, **233**(39):149-152.
356. Zhai Y, Heijne WH, Smith DW, Saier MH, Jr.: **Homologues of archaeal rhodopsins in plants, animals and fungi: structural and functional predications for a putative fungal chaperone protein**. *Biochim Biophys Acta* 2001, **1511**(2):206-223.
357. Spudich JL, Yang CS, Jung KH, Spudich EN: **Retinylidene proteins: structures and functions from archaea to humans**. *Annu Rev Cell Dev Biol* 2000, **16**:365-392.
358. Brown LS: **Eubacterial rhodopsins - unique photosensors and diverse ion pumps**. *Biochim Biophys Acta* 2014, **1837**(5):553-561.
359. Inoue K, Tsukamoto T, Sudo Y: **Molecular and evolutionary aspects of microbial sensory rhodopsins**. *Biochim Biophys Acta* 2014, **1837**(5):562-577.
360. Anantharaman V, Aravind L: **Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria**. *BMC genomics* 2003, **4**(1):34.
361. Anantharaman V, Abhiman S, de Souza RF, Aravind L: **Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system**. *Gene* 2011, **475**(2):63-78.

362. Krishnan A, Almen MS, Fredriksson R, Schioth HB: **The origin of GPCRs: identification of mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in fungi**. *PLoS One* 2012, **7**(1):e29817.
363. de Mendoza A, Sebe-Pedros A, Ruiz-Trillo I: **The evolution of the GPCR signaling system in eukaryotes: modularity, conservation, and the transition to metazoan multicellularity**. *Genome Biol Evol* 2014, **6**(3):606-619.
364. Taddese B, Upton GJ, Bailey GR, Jordan SR, Abdulla NY, Reeves PJ, Reynolds CA: **Do plants contain g protein-coupled receptors?** *Plant physiology* 2014, **164**(1):287-307.
365. Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, Niswender CM, Katritch V, Meiler J, Cherezov V *et al*: **Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator**. *Science* 2014, **344**(6179):58-64.
366. Dore AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, Errey JC, Jazayeri A, Khan S, Tehan B *et al*: **Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain**. *Nature* 2014, **511**(7511):557-562.
367. Inoue K, Kato Y, Kandori H: **Light-driven ion-translocating rhodopsins in marine bacteria**. *Trends Microbiol* 2015, **23**(2):91-98.
368. Inoue K, Ono H, Abe-Yoshizumi R, Yoshizawa S, Ito H, Kogure K, Kandori H: **A light-driven sodium ion pump in marine bacteria**. *Nat Commun* 2013, **4**:1678.
369. Gushchin I, Shevchenko V, Polovinkin V, Kovalev K, Alekseev A, Round E, Borshchevskiy V, Balandin T, Popov A, Gensch T *et al*: **Crystal structure of a light-driven sodium pump**. *Nat Struct Mol Biol* 2015, **22**(5):390-395.
370. Kato HE, Inoue K, Abe-Yoshizumi R, Kato Y, Ono H, Konno M, Hososhima S, Ishizuka T, Hoque MR, Kunitomo H *et al*: **Structural basis for Na(+) transport mechanism by a light-driven Na(+) pump**. *Nature* 2015, **521**(7550):48-53.
371. Zhang C, Srinivasan Y, Arlow DH, Fung JJ, Palmer D, Zheng Y, Green HF, Pandey A, Dror RO, Shaw DE *et al*: **High-resolution crystal structure of human protease-activated receptor 1**. *Nature* 2012, **492**(7429):387-392.
372. Yuan S, Filipek S, Palczewski K, Vogel H: **Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway**. *Nat Commun* 2014, **5**:4733.
373. Vickery ON, Carvalheda CA, Zaidi SA, Pisliakov AV, Katritch V, Zachariae U: **Intracellular Transfer of Na(+) in an Active-State G-Protein-Coupled Receptor**. *Structure* 2018, **26**(1):171-180 e172.
374. Mahaut-Smith MP, Martinez-Pinna J, Gurung IS: **A role for membrane potential in regulating GPCRs?** *Trends Pharmacol Sci* 2008, **29**(8):421-429.
375. Vickery ON, Machtens JP, Tamburrino G, Seeliger D, Zachariae U: **Structural Mechanisms of Voltage Sensing in G Protein-Coupled Receptors**. *Structure* 2016, **24**(6):997-1007.
376. Ben-Chaim Y, Tour O, Dascal N, Parnas I, Parnas H: **The M-2 muscarinic G-protein-coupled receptor is voltage-sensitive**. *Journal of Biological Chemistry* 2003, **278**(25):22482-22491.
377. Richards MH, Vangiersbergen PLM: **Human Muscarinic Receptors Expressed in A9l and Cho Cells - Activation by Full and Partial Agonists**. *Brit J Pharmacol* 1995, **114**(6):1241-1249.
378. Ben-Chaim Y, Chanda B, Dascal N, Bezanilla F, Parnas I, Parnas H: **Movement of 'gating charge' is coupled to ligand binding in a G-protein-coupled receptor**. *Nature* 2006, **444**(7115):106-109.
379. Navarro-Polanco RA, Galindo EGM, Ferrer-Villada T, Arias M, Rigby JR, Sanchez-Chapula JA, Tristani-Firouzi M: **Conformational changes in the M2 muscarinic receptor induced by membrane voltage and agonist binding**. *J Physiol-London* 2011, **589**(7):1741-1753.
380. Moreno-Galindo EG, Alamilla J, Sanchez-Chapula JA, Tristani-Firouzi M, Navarro-Polanco RA: **The agonist-specific voltage dependence of M2 muscarinic receptors modulates the deactivation of the acetylcholine-gated K(+) current (I KACh)**. *Pflugers Arch* 2016, **468**(7):1207-1214.
381. Copeland R, Keseru G: **Thermodynamics and Kinetics of Drug Binding**. 2015.
382. Rinne A, Mobarec JC, Mahaut-Smith M, Kolb P, Bunemann M: **The mode of agonist binding to a G protein-coupled receptor switches the effect that voltage changes have on signaling**. *Sci Signal* 2015, **8**(401).

383. Ohana L, Barchad O, Parnas I, Parnas H: **The metabotropic glutamate G-protein-coupled receptors mGluR3 and mGluR1a are voltage-sensitive**. *J Biol Chem* 2006, **281**(34):24204-24215.
384. Rinne A, Birk A, Bunemann M: **Voltage regulates adrenergic receptor function**. *Proc Natl Acad Sci U S A* 2013, **110**(4):1536-1541.
385. Birk A, Rinne A, Bunemann M: **Membrane Potential Controls the Efficacy of Catecholamine-induced beta1-Adrenoceptor Activity**. *J Biol Chem* 2015, **290**(45):27311-27320.
386. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG: **Agonist-bound adenosine A(2A) receptor structures reveal common features of GPCR activation**. *Nature* 2011, **474**(7352):521-U154.
387. Sahlholm K, Marcellino D, Nilsson J, Fuxe K, Arhem P: **Voltage-sensitivity at the human dopamine D-2S receptor is agonist-specific**. *Biochem Bioph Res Co* 2008, **377**(4):1216-1221.
388. Sahlholm K, Barchad-Avitzur O, Marcellino D, Gomez-Soler M, Fuxe K, Ciruela F, Arhem P: **Agonist-specific voltage sensitivity at the dopamine D2S receptor--molecular determinants and relevance to therapeutic ligands**. *Neuropharmacology* 2011, **61**(5-6):937-949.
389. Copeland RA: **Conformational adaptation in drug-target interactions and residence time**. *Future Med Chem* 2011, **3**(12):1491-1501.
390. Balashov SP, Imasheva ES, Dioumaev AK, Wang JM, Jung KH, Lanyi JK: **Light-driven $Na^+$ pump from *Gillisia limnaea*: a high-affinity $Na^+$ binding site is formed transiently in the photocycle**. *Biochemistry* 2014, **53**(48):7549-7561.
391. Kanada S, Takeguchi Y, Murakami M, Ihara K, Kouyama T: **Crystal Structures of an O-Like Blue Form and an Anion-Free Yellow Form of pharaonis Halorhodopsin**. *Journal of Molecular Biology* 2011, **413**(1):162-176.
392. Gushchin I, Reshetnyak A, Borshchevskiy V, Ishchenko A, Round E, Grudinin S, Engelhard M, Buldt G, Gordeliy V: **Active State of Sensory Rhodopsin II: Structural Determinants for Signal Transfer and Proton Pumping**. *Journal of Molecular Biology* 2011, **412**(4):591-600.
393. Volkov O, Kovalev K, Polovinkin V, Borshchevskiy V, Bamann C, Astashkin R, Marin E, Popov A, Balandin T, Willbold D *et al*: **Structural insights into ion conduction by channelrhodopsin 2**. *Science* 2017, **358**(6366).
394. Gmelin W, Zeth K, Efremov R, Heberle J, Tittor J, Oesterhelt D: **The crystal structure of the L1 intermediate of halorhodopsin at 1.9 angstrom resolution**. *Photochem Photobiol* 2007, **83**(2):369-377.
395. Ran TT, Ozorowski G, Gao YY, Sineshchekov OA, Wang WW, Spudich JL, Luecke H: **Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin complexes**. *Acta Crystallogr D* 2013, **69**:1965-1980.
396. Gushchin I, Chervakov P, Kuzmichev P, Popov AN, Round E, Borshchevskiy V, Ishchenko A, Petrovskaya L, Chupin V, Dolgikh DA *et al*: **Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria**. *P Natl Acad Sci USA* 2013, **110**(31):12631-12636.
397. Luecke H, Schobert B, Stagno J, Imasheva ES, Wang JM, Balashov SP, Lanyi JK: **Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore**. *P Natl Acad Sci USA* 2008, **105**(43):16561-16565.
398. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S: **Crystal structure of the μ-opioid receptor bound to a morphinan antagonist**. *Nature* 2012, **485**(7398):321-326.
399. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK *et al*: **High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor**. *Science* 2007, **318**(5854):1258-1265.
400. Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V *et al*: **Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist**. *Science* 2010, **330**(6007):1091-1095.
401. Egloff P, Hillenbrand M, Klenk C, Batyuk A, Heine P, Balada S, Schlinkmann KM, Scott DJ, Schutz M, Pluckthun A: **Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli**. *Proc Natl Acad Sci U S A* 2014, **111**(6):E655-662.

402. Wang C, Jiang Y, Ma JM, Wu HX, Wacker D, Katritch V, Han GW, Liu W, Huang XP, Vardy E *et al*: **Structural Basis for Molecular Recognition at Serotonin Receptors**. *Science* 2013, **340**(6132):610-614.
403. Hanson MA, Roth CB, Jo EJ, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC *et al*: **Crystal Structure of a Lipid G Protein-Coupled Receptor**. *Science* 2012, **335**(6070):851-855.
404. Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Zhang W *et al*: **Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex**. *Science* 2013, **341**(6152):1387-1390.
405. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V: **The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure**. *J Mol Biol* 2004, **342**(2):571-583.
406. Jekely G: **Evolution of phototaxis**. *Philos Trans R Soc Lond B Biol Sci* 2009, **364**(1531):2795-2808.
407. Subramaniam S, Gerstein M, Oesterhelt D, Henderson R: **Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin**. *EMBO J* 1993, **12**(1):1-8.
408. Vonck J: **A three-dimensional difference map of the N intermediate in the bacteriorhodopsin photocycle: part of the F helix tilts in the M to N transition**. *Biochemistry* 1996, **35**(18):5870-5878.
409. Heberle J, Fitter J, Sass HJ, Buldt G: **Bacteriorhodopsin: the functional details of a molecular machine are being resolved**. *Biophys Chem* 2000, **85**(2-3):229-248.
410. Radzwill N, Gerwert K, Steinhoff HJ: **Time-resolved detection of transient movement of helices F and G in doubly spin-labeled bacteriorhodopsin**. *Biophysical Journal* 2001, **80**(6):2856-2866.
411. Nakanishi T, Kanada S, Murakami M, Ihara K, Kouyama T: **Large deformation of helix F during the photoreaction cycle of Pharaonis halorhodopsin in complex with azide**. *Biophysical Journal* 2013, **104**(2):377-385.
412. da Silva GF, Goblirsch BR, Tsai AL, Spudich JL: **Cation-Specific Conformations in a Dual-Function Ion-Pumping Microbial Rhodopsin**. *Biochemistry* 2015.
413. Sattig T, Rickert C, Bamberg E, Steinhoff HJ, Bamann C: **Light-induced movement of the transmembrane helix B in channelrhodopsin-2**. *Angew Chem Int Ed Engl* 2013, **52**(37):9705-9708.
414. Krause N, Engelhard C, Heberle J, Schlesinger R, Bittl R: **Structural differences between the closed and open states of channelrhodopsin-2 as observed by EPR spectroscopy**. *FEBS Lett* 2013, **587**(20):3309-3313.
415. Klare JP, Bordignon E, Engelhard M, Steinhoff HJ: **Sensory rhodopsin II and bacteriorhodopsin: light activated helix F movement**. *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* 2004, **3**(6):543-547.
416. Mulkidjanian AY, Galperin MY, Koonin EV: **Co-evolution of primordial membranes and membrane proteins**. *Trends Biochem Sci* 2009, **34**(4):206-215.
417. Mulkidjanian AY, Galperin MY, Makarova KS, Wolf YI, Koonin EV: **Evolutionary primacy of sodium bioenergetics**. *Biol Direct* 2008, **3**:13.
418. Mulkidjanian AY, Dibrov P, Galperin MY: **The past and present of sodium energetics: may the sodium-motive force be with you**. *Biochim Biophys Acta* 2008, **1777**(7-8):985-992.
419. Mulkidjanian AY, Makarova KS, Galperin MY, Koonin EV: **Inventing the dynamo machine: the evolution of the F-type and V-type ATPases**. *Nat Rev Microbiol* 2007, **5**(11):892-899.
420. Burykin A, Kato M, Warshel A: **Exploring the origin of the ion selectivity of the KcsA potassium channel**. *Proteins-Structure Function and Bioinformatics* 2003, **52**(3):412-426.
421. Oesterhelt D, Tittor J, Bamberg E: **A unifying concept for ion translocation by retinal proteins**. *J Bioenerg Biomembr* 1992, **24**(2):181-191.
422. Kaulen AD: **Electrogenic processes and protein conformational changes accompanying the bacteriorhodopsin photocycle**. *Biochim Biophys Acta* 2000, **1460**(1):204-219.
423. Blanck A, Oesterhelt D, Ferrando E, Schegk ES, Lottspeich F: **Primary structure of sensory rhodopsin I, a prokaryotic photoreceptor**. *EMBO J* 1989, **8**(13):3963-3971.

424. Skulachev VP: **Interrelations of bioenergetic and sensory functions of the retinal proteins**. *Q Rev Biophys* 1993, **26**(2):177-199.
425. Spudich JL, Sineshchekov OA, Govorunova EG: **Mechanism divergence in microbial rhodopsins**. *Biochim Biophys Acta* 2014, **1837**(5):546-552.
426. Beja O, Lanyi JK: **Nature's toolkit for microbial rhodopsin ion pumps**. *Proc Natl Acad Sci USA* 2014, **111**(18):6538-6539.
427. Nagel G, Szellas T, Kateriya S, Adeishvili N, Hegemann P, Bamberg E: **Channelrhodopsins: directly light-gated cation channels**. *Biochem Soc Trans* 2005, **33**(Pt 4):863-866.
428. McCarren J, DeLong EF: **Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla**. *Environ Microbiol* 2007, **9**(4):846-858.
429. Brown LS: **Fungal rhodopsins and opsin-related proteins: eukaryotic homologues of bacteriorhodopsin with unknown functions**. *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* 2004, **3**(6):555-565.
430. Rando RR: **The biochemistry of the visual cycle**. *Chem Rev* 2001, **101**(7):1881-1896.
431. Hubbard R, Wald G: **The mechanism of rhodopsin synthesis**. *Proc Natl Acad Sci U S A* 1951, **37**(2):69-79.
432. Kawanabe A, Kandori H: **Photoreactions and structural changes of anabaena sensory rhodopsin**. *Sensors (Basel)* 2009, **9**(12):9741-9804.
433. Klare JP, Chizhov I, Engelhard M: **Microbial rhodopsins: scaffolds for ion pumps, channels, and sensors**. *Results Probl Cell Differ* 2008, **45**:73-122.
434. Leff P: **The two-state model of receptor activation**. *Trends Pharmacol Sci* 1995, **16**(3):89-97.
435. Hall DA: **Modeling the functional effects of allosteric modulators at pharmacological receptors: an extension of the two-state model of receptor activation**. *Mol Pharmacol* 2000, **58**(6):1412-1423.
436. Copeland RA: **The drug-target residence time model: a 10-year retrospective**. *Nat Rev Drug Discov* 2016, **15**(2):87-95.
437. Wikstrom M, Krab K, Saraste M: **Proton-translocating cytochrome complexes**. *Annu Rev Biochem* 1981, **50**:623-655.
438. Lamichhane R, Liu JJ, Pljevaljcic G, White KL, van der Schans E, Katritch V, Stevens RC, Wuthrich K, Millar DP: **Single-molecule view of basal activity and activation mechanisms of the G protein-coupled receptor beta2AR**. *Proc Natl Acad Sci U S A* 2015, **112**(46):14254-14259.
439. Ye L, Van Eps N, Zimmer M, Ernst OP, Prosser RS: **Activation of the A2A adenosine G-protein-coupled receptor by conformational selection**. *Nature* 2016, **533**(7602):265-268.
440. Susac L, Eddy MT, Didenko T, Stevens RC, Wuthrich K: **A2A adenosine receptor functional states characterized by (19)F-NMR**. *Proc Natl Acad Sci U S A* 2018, **115**(50):12733-12738.
441. Vickery ON, Machtens JP, Zachariae U: **Membrane potentials regulating GPCRs: insights from experiments and molecular dynamics simulations**. *Curr Opin Pharmacol* 2016, **30**:44-50.
442. Strange PG: **Agonist binding, agonist affinity and agonist efficacy at G protein-coupled receptors**. *Br J Pharmacol* 2008, **153**(7):1353-1363.
443. Shalaeva DN, Galperin MY, Mulkidjanian AY: **Eukaryotic G protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins**. *Biol Direct* 2015, **10**:63.
444. Barchad-Avitzur O, Priest MF, Dekel N, Bezanilla F, Parnas H, Ben-Chaim Y: **A Novel Voltage Sensor in the Orthosteric Binding Site of the M2 Muscarinic Receptor**. *Biophys J* 2016, **111**(7):1396-1408.
445. Warshel A, Russell ST: **Calculations of electrostatic interactions in biological systems and in solutions**. *Q Rev Biophys* 1984, **17**(3):283-422.
446. Meier T, Krah A, Bond PJ, Pogoryelov D, Diederichs K, Faraldo-Gomez JD: **Complete ion-coordination structure in the rotor ring of Na+-dependent F-ATP synthases**. *J Mol Biol* 2009, **391**(2):498-507.
447. Kaim G, Wehrle F, Gerike U, Dimroth P: **Molecular basis for the coupling ion selectivity of F1F0 ATP synthases: probing the liganding groups for Na+ and Li+ in the c subunit of the ATP synthase from Propionigenium modestum**. *Biochemistry* 1997, **36**(30):9185-9194.

448. Kruse AC, Hu J, Kobilka BK, Wess J: **Muscarinic acetylcholine receptor X-ray structures: potential implications for drug development**. *Curr Opin Pharmacol* 2014, **16**:24-30.
449. Ben-Chaim Y, Tour O, Dascal N, Parnas I, Parnas H: **The M2 muscarinic G-protein-coupled receptor is voltage-sensitive**. *J Biol Chem* 2003, **278**(25):22482-22491.
450. Oliveira L, Paiva ACdM, Vriend G: **A common motif in G-protein-coupled seven transmembrane helix receptors**. *Journal of Computer-Aided Molecular Design* 1993, **7**(6):649-658.
451. Nygaard R, Frimurer TM, Holst B, Rosenkilde MM, Schwartz TW: **Ligand binding and micro-switches in 7TM receptor structures**. *Trends Pharmacol Sci* 2009, **30**(5):249-259.
452. Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A, Filipek S: **Action of molecular switches in GPCRs--theoretical and experimental studies**. *Curr Med Chem* 2012, **19**(8):1090-1109.
453. Okada T, Ernst OP, Palczewski K, Hofmann KP: **Activation of rhodopsin: new insights from structural and biochemical studies**. *Trends Biochem Sci* 2001, **26**(5):318-324.
454. Tehan BG, Bortolato A, Blaney FE, Weir MP, Mason JS: **Unifying family A GPCR theories of activation**. *Pharmacol Ther* 2014, **143**(1):51-60.
455. Segala E, Guo D, Cheng RK, Bortolato A, Deflorian F, Dore AS, Errey JC, Heitman LH, AP IJ, Marshall FH *et al*: **Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength**. *J Med Chem* 2016, **59**(13):6470-6479.
456. Kovalev K, Polovinkin V, Gushchin I, Alekseev A, Shevchenko V, Borshchevskiy V, Astashkin R, Balandin T, Bratanov D, Vaganova S *et al*: **Structure and mechanisms of sodium-pumping KR2 rhodopsin**. *Sci Adv* 2019, **5**(4):eaav2671.
457. Kovalev K, Astashkin R, Gushchin I, Orekhov P, Volkov D, Zinovev E, Marin E, Rulev M, Alekseev A, Royant A *et al*: **Molecular mechanism of light-driven sodium pumping**. *Nat Commun* 2020, **11**(1):2137.
458. Dickey A, Faller R: **Examining the contributions of lipid shape and headgroup charge on bilayer behavior**. *Biophys J* 2008, **95**(6):2636-2646.
459. Isberg V, Vroling B, van der Kant R, Li K, Vriend G, Gloriam D: **GPCRDB: an information system for G protein-coupled receptors**. *Nucleic Acids Res* 2014, **42**(Database issue):D422-D425.
460. Bakeeva LE, Chumakov KM, Drachev AL, Metlina AL, Skulachev VP: **The sodium cycle. III. Vibrio alginolyticus resembles Vibrio cholerae and some other vibriones by flagellar motor and ribosomal 5S-RNA structures**. *Biochim Biophys Acta* 1986, **850**(3):466-472.
461. Dibrov PA, Kostryko VA, Lazarova RL, Skulachev VP, Smirnova IA: **The sodium cycle. I. Na+-dependent motility and modes of membrane energization in the marine alkalotolerant vibrio Alginolyticus**. *Biochim Biophys Acta* 1986, **850**(3):449-457.
462. Dibrov PA, Lazarova RL, Skulachev VP, Verkhovskaya ML: **The sodium cycle. II. Na+-coupled oxidative phosphorylation in Vibrio alginolyticus cells**. *Biochim Biophys Acta* 1986, **850**(3):458-465.

# *Acknowledgments*

First of all, I would like to express my gratitude to my supervisor, PD Dr. Armen Y. Mulkidjanian for the opportunity to work on many interesting projects. I am grateful for all the effort and time, for understanding and patience, and for invaluable lessons in the scientific method of thinking and writing.

I would like to thank Prof. Dr. Heinz-Jürgen Steinhoff for his continuous interest in my work and the possibility to use the infrastructure of his department.

I would like to thank Prof. Dr. Karin Busch for the opportunity to work with her on a fascinating project. I am also very grateful to Prof. Dr. Karin Busch, Prof. Dr. Philipp Maass, and Dr. Gabriele Deckers-Hebestreit for their kind agreement to participate in the examination committee.

I am grateful to Dr. Dmitry Cherepanov from the A.N. Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, Moscow for his guidance and lessons in molecular modeling and molecular dynamics simulations, and for his design of the GPCR activation model used in this work.

I would like to express my gratitude to Dr. Michael Y. Galperin from National Center of Biotechnology Information, NLM, NIH, and Prof. Dr. Gert Vriend from Radboud University Nijmegen Medical Centre for their guidance and their input in the study of the role of sodium in GPCRs.

I would like to acknowledge the work of Dr. Daria Dibrova, who has always been a role model for me and who started the investigation of the role of monovalent cations in P-loop NTPases in the context of early evolution and origin of life many years ago in her dissertation.

I am very thankful to all the members of our department, past and present: Dr. Johan Klare, Katharina Rudi, Alex Kolbasevici, Maria Kozlova, Dr. Daniel Klose, Dr. Philipp Orekhov, Dr. Marine Bozdaganyan, and all the others for creating an accepting and supportive environment, that was a pleasure to be a part of.

I am grateful to Dr. Michael Sippach, Dr. Magdalena Schumacher, Dr. Ioan Orban-Glaß, Melanie Appelfeller, and Marius Krause for all the adventures we have been on, I will cherish those memories forever.

Finally, I would like to deeply thank my parents, Nikolai Shalaev and Vera Shalaeva, for everything they have done for me in my life. I am especially grateful to my elder brother Sergey Shalaev and his family for their ongoing support.

## *Declaration of scientific integrity*

I hereby declare that this thesis is the result of my own research and neither any part of this work nor the whole thesis in the original or similar form has been submitted for a degree in any other university or academic institution. I further declare, to the best of my knowledge and belief, that any materials from the work of other people, previously published or not, included in my thesis are appropriately referenced or acknowledged.

## *Erklärung über die Eigenständigkeit*

Ich habe die Dissertation selbständig angefertigt und mich außer der angegebenen keiner weiteren Hilfsmittel bedient. Alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen wurden, sind als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen. Ich erkläre, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

…….……  …….……

Osnabrück, Germany  Daria Shalaeva