Universidad Autónoma de Madrid
Centro de Biología Molecular "Severo Ochoa"

Molecular Dynamics and Quantum Mechanics / Molecular Mechanics Analysis of Hydrolysis Reactions in NTPase Proteins

Tesis doctoral

Fernando Martín García

Madrid, 2013

Instituto Universitario de Ciencia de Materiales Nicolás Cabrera

Facultad de Ciencias

Universidad Autónoma de Madrid

Molecular Dynamics and Quantum Mechanics / Molecular Mechanics Analysis of Hydrolysis Reactions in NTPases Proteins

Memoria para optar al título de Doctor en Biofísica,
presentada por:

Fernando Martín García
Licenciado en Biotecnología

Directores de tesis

Dr. Paulino Gómez Puertas **Dr. Jesús Mendieta Gómez**

Centro de Biología Molecular "Severo Ochoa".
Universidad Autónoma de Madrid.
Campus de Cantoblanco.

A mis padres

Agradecimientos

Agradecer a tantas personas con las que me he cruzado a lo largo de estos cinco años es complicado. Pero sobre todo, plasmar con palabras lo mucho que algunas personas han hecho por mí en estos años. Porque aunque una tesis no deja de ser el trabajo y esfuerzo personal de uno mismo, en mi caso no podría haberla terminado sin toda la gente que he tenido a mi alrededor, tanto en el ámbito académico como en el personal. Seguro que me dejo mucha gente, así que gracias a todos de antemano.

Creo que es menester agradecer en primer lugar a las instituciones. A Biomol Informatics S.L. por la financiación facilitada para la realización de la tesis. Así mismo agradecer al Centro de Biología Molecular "Severo Ochoa" en su colaboración con Biomol Informatics S.L., permitiéndome realizar la tesis en sus instalaciones. Por último, al Centro de Computación Científica de la Universidad Autónoma de Madrid, por el servicio de computación prestado en estos años.

Centrándome, ahora sí, en las personas, quería agradecer primeramente a mis dos directores de tesis, los doctores Paulino Gómez Puertas y Jesús Mendieta Gómez, la oportunidad que me han brindado al poder realizar esta tesis bajo su tutela, pero sobretodo por el compromiso que han mostrado desde el primer día que empecé aquí. A Paulino porque fue la primera persona que me abrió los ojos al campo de la biología computacional y me mostró que se podía hacer biología que no fuera solo con pipetas, placas de cultivo o Westerns. A Jesús porque ha sido la persona que me ha enseñado todo lo que sé de este campo, además de ser capaz de contagiarme el ánimo e interés que él muestra en su día a día en la investigación. Muchas gracias a ambos.

Muchas gracias a la gente del Grupo de Modelado Molecular y de Biomol: a los que ya llevan tiempo (Jesús, Iñigo, David, Jan-Jaap y Silvia), como a las nuevas incorporaciones (Daniel y Rurika). En todos estos años he aprendido mucho de nuestras conversaciones (sobretodo los viernes), y discusiones. Esta tesis no hubiera sido igual. Al doctor Eduardo López Viñas, por su ayuda en mis primeros momentos en el "laboratorio" y por todos sus consejos y conversaciones que siguieron a continuación.

Thanks to the people of Copenhagen University: Professor Kresten Lindorff-larsen, Wouter Boomsma and Peng-Fei. Thanks to permit me to collaborate with you. Special thanks to Dra. Elena Papaleo, for all her scientific support in this last year and for the helping during my stay in Copenhagen. Thanks a lot.

No quiero dejar de dar las gracias a mucha de la gente que he conocido en estos años en el centro y en la universidad: a Helena y Manu, por todos los cafés, tés, charlas, risas y buenos momentos, dentro y fuera del CBM; a María Luisa, Sergio, Andrea, Julián, Eugenia, Ana, Josemi, por todos los buenos ratos. A Carlos, Alex, Patri y Eddy. Tras diez años seguís ahí para lo que necesite.

A la Dra. Belmar López (no me podía resistir). A pesar de la distancia (ahora más que nunca), siempre has estado ahí para escucharme y para dar siempre buenos consejos. ¡Eso y para recordarme mis recomendaciones!

A Andrés, Bea, Carol y Joaquín. Tenía que elegir un orden y el alfabético es el más propio. No hacen falta las palabras y los motivos. Habéis estado en los buenos momentos, pero sobretodo, me habéis ayudado en los malos. Me habéis ayudado a que esta tesis esté escrita.

A mis hermanos, Victor y Fede, a mi primo Ángel, y a mi abuela, Juliana, por preocuparos por mí, no solo ahora, si no siempre.

A mis padres, Ángel y Milagros. A vosotros os dedico esta tesis. Porque la habéis hecho posible con vuestro esfuerzo y sacrificio diario.

Fer

Index

Summary

The main aim of macromolecular structure analysis is to study the connection between structure and function; to gain an understanding of the mechanisms via which macromolecular systems perform their functions at the molecular level; and to identify the role of every residue that is involved in the function of these macromolecules.

Molecular biology and biochemistry techniques have offered and continue to offer a great number of techniques that allow us to obtain reliable results in the study of such processes. In recent years, we have seen how new computational methods have been incorporated into the range of analytical tools available. These new methods are oriented towards the analysis of biological systems and they have evolved to offer us highly accurate calculations that reproduce results obtained in the laboratory. Thus, they provide an atomic resolution tool that can provide additional information to that obtained with *in vivo* and *in vitro* techniques.

Focusing on the approximations applied in the studies performed as part of this thesis, molecular dynamics (MD) allow us to simulate and analyze how both biological macromolecules (proteins, DNA, membranes, etc) and small molecules (ligands, cofactors, peptides, ions, etc) evolve over time in simulations that are close to experimental conditions and that are governed by the different ensembles and force fields. The development and constant evolution of computational systems now permits the study of larger and more complex systems, including not only the system to be studied but also its possible interaction with additional systems, thereby providing new and more accurate information.

In addition to the analysis, that MD can offer (protein – protein / ligand interaction, studies of solvation and conformational changes, etc), quantum mechanics (QM) offers us the opportunity to study different chemical reactions that occur in these systems. Although the application of QM to complete macromolecular systems would be unaffordable from a computational point of view, the combination with molecular mechanics (MM) and MD, in what has been called QM /MM /MD or QM / MD, is perfectly affordable. In the present case, this interface allows us to describe the specific role of certain elements (such as the amino acids that constitute the catalytic sites of enzymes), while considering the protein environment, and permits us to generate and corroborate different hypothesis that would be

extremely difficult to formulate and examine in experimental studies.

In accordance with this idea, in this thesis we have developed a new approximation based on QM /MD interface, combined with biased molecular dynamics (BMD) techniques, such us steered molecular dynamics (SMD) and umbrella sampling (US). This combination permits the study of complete macromoleculare systems (also called all atoms systems) via MD; and include the atoms implicated in the reaction in QM calculations. It involves the sampling of the conformational space defined by reaction coordinates in order to clarify how a reaction takes place and what the role of the surrounding protein environment is.

Resumen

El principal objetivo del análisis estructural de macromoléculas consiste en el estudio de la relación estructura-función de las mismas, comprender los mecanismos gracias a los cuales desarrollan su función a nivel molecular y así como el papel que desempeñan cada uno de los residuos que intervienen en la función de dichas macromoléculas.

Las técnicas de biología molecular han ofrecido y siguen ofreciendo un gran número de técnicas que ofrecen resultados fiables en el estudio de tales procesos. Durante los últimos años, hemos visto como las nuevas aproximaciones computacionales orientadas al análisis de sistemas biológicos se han incorporado dentro del rango de herramientas analíticas. Estos nuevos métodos están orientados al análisis de sistemas biológicos y han evolucionado hasta proporcionarnos cálculos muy precisos que reproducen los resultados obtenidos en el laboratorio. Nos proporcionan por tanto una resolución atómica que puede generar información adicional a los obtenidos por técnicas *in vivo* e *in vitro*.

Centrándonos en las aproximaciones aplicadas en los estudios realizados en esta tesis, la dinámica molecular (MD) permite simular y analizar tanto los sistemas biológicos macromoleculares (proteínas, DNA, membranas, etc.) como las pequeñas moléculas (ligandos, cofactores, péptidos pequeños, iones, etc.) observando su evolución con el tiempo bajo condiciones cercanas a las experimentales y que vienen dadas por los ensambles y campos de fuerzas. El desarrollo y la constante evolución de los sistemas de computación, nos permiten el estudio de sistemas cada vez más grandes y complejos, incluyendo no solo el sistema de estudio si no también las posibles interacciones con otros sistemas, proporcionando nueva y más precisa información.

Adicionalmente al análisis que la MD ofrece (interacciones proteína – proteína / ligando, estudios de solvatación y de cambios estructurales, etc.), la mecánica cuántica (QM) nos ofrece la capacidad de estudiar diferentes reacciones químicas llevadas a cabo por los sistemas biológicos de nuestro interés. Aunque su aplicación al estudio de sistemas macromoleculares completos es inabordable en estos momentos desde el punto de vista de la computación, su combinación con MD, en lo que se denomina interfaz QM/MM/MD o

QM/MD, permite afrontar dicho problema. En nuestro caso, dicha interfaz nos permite describir el papel específico de determinados elementos (tales como los aminoácidos que componen los centros catalíticos de los enzimas), considerando el ambiente proteico, permitiendo así generar y corroborar las diferentes hipótesis que resultan difíciles de investigar por medio de técnicas experimentales.

De acuerdo con esta idea, en esta tesis hemos desarrollado una nueva aproximación basada en la interfaz QM/MD, combinada con técnicas de dinámica molecular dirigida (BMD), tales como el *"steered molecular dynamics (SMD)"* y *"umbrella sampling (US)"*. Esta combinación permitirá el estudio de sistemas macromoleculares completos gracias a la parte de MD, incluyendo en los cálculos de QM aquellos átomos implicados en la reacción, analizando el espacio conformacional definido por las coordenadas de reacción con el fin de aclarar como una reacción es llevada a cabo y cuál es el papel de ambiente proteico en la misma.

List of figures

List of videos

Video 1: Study of the solvation of the different catalytic sites in the pentamer structure of FtsZ.
http://www.cbm.uam.es/bioweb/articles/FEBSLettters_2012_sup_video.avi

Video 2: Enzymatic metadynamic reaction of Hras-RasGAP protein.
http://www.cbm.uam.es/bioweb/articles/biophysj102_2012_supp_video_HD.avi

Video 3: Enzymatic metadynamic reaction of F_oF_1-ATPase protein.
http://www.cbm.uam.es/bioweb/articles/Biochemistry_2013_sup_video.avi

List of tables

List of abbreviations

ATP: Adenosine triphophate

ADP: Adenosin diphsophate

BMD: Biased Molecular Dynamics

FtsZ: Filamenting temperature-sensitive mutant Z

GTP: Guanosine triphosphate

GDP: Guanosine diphosphate

GAP: GTPase activating proteins

GEF: Guanosine exchange factors

MD: Molecular Dynamics

MM: Molecular Mechanics

PBC: Periodic boundary conditions

QM/MD: Quantum Mechanics/Molecular Dynamics (also called QM/MM/MD)

SMD: Steered Molecular Dynamics

US: Umbrella Sampling

1. General introduction

In this section we will introduce the theoretical concepts concerning chemical reactions, in order to understand the molecular basis applied to the description of the hydrolysis reaction. We will present a brief introduction to the four biological systems studied in this thesis (FtsZ, HRas-p21, F_oF_1 ATP synthase and myosin) that perform hydrolysis reaction of nucleotides. This introduction will permit us to understand how they work, explaining possible mechanisms via which they perform the hydrolysis reaction, as we will explain in detail in the Results section below.

In the second part of the chapter, we will introduce how macromolecular systems are studied using MD, explaining how these calculations are performed. This part establishes the basis for a better understanding of the theoretical concepts that will be developed below in the Materials and methods section.

To conclude the chapter, we will introduce the idea of energy surface maps and what they represent. Additionally, we will analyze what is the information we can gain by studying them, in order to clarify and explain the information and conclusions we could extract from them.

1.1. Enzymatic reactions

Chemical reactions

Cells are systems whose functions are made at essentially constant temperature (isothermal systems) and constant pressure (isobaric system). Because heat can only do work when it passes to a zone or object at a lower temperature, heat flow could not be a source of energy in living cells. The energy must be obtained from free energy, described by the Gibbs free-energy function, *G*. It is the value of this free energy what allows us to predict the direction of the chemical reactions that occur in the cell (exergonic or endergonic processes), their equilibrium position and the amount of work they can perform at constant temperature and pressure.

According to the different kind of cells, heterotrophics acquire free energy from nutrient molecules while photosynthetic ones acquire it from absorbed solar radiation. All this free energy is transformed in ATP/GTP and another energy-rich molecules that provide energy for the different cell functions. In the case of ATP, it donates some of its chemical energy to endergonic processes such as the synthesis of the metabolic intermediates and the macromolecules from smaller precursors, the transport of substances across the membranes against concentration gradients or the mechanical motion of proteins. The donation of the energy involves the covalent participation of ATP, which implies that ATP will be converted in its products, ADP and P_i, or even in AMP and 2 P_i.

Chemical reactions, consisting of a mixture of reactants and products, tend to constantly change between the former and the latter until equilibrium is reached. At equilibrium, the rate between the both directions of the reaction is equal. However, when the reaction is in a non-equilibrium state, the tendency to move toward equilibrium represents a driving force, the magnitude of which can be expressed as the free-energy change for the reaction, ΔG^o. If we consider standard conditions (pH 7, constant temperature and pressure), the free-energy change in the case of the hydrolysis of ATP is -30.5 kJ/mol (-7.3 kcal/mol). However, this value can vary if we take into account the living cell, were the concentrations of the substrates and products can vary between cells and also in the individual cell the metabolic

conditions in that right moment. One example of this is what happens in resting muscle cells, where the concentration of free ATP oscillate between 1 and 37 µM. We can use the following expression:

$$\Delta G_p = \Delta G'^{o} + RT \ln \frac{[ADP][P_i]}{[ATP]} \qquad (1.1.1)$$

where ΔG_p is the free energy of hydrolysis of ATP under intracellular conditions (also called the phosphorylation potential), and $\Delta G'^{o}$ is the free energy under standard conditions. The value of the free energy in such cells will be -64 kJ/mol. This reveals to us that the free energy released *in vivo* is greater than the standard free energy, $\Delta G'^{o}$.

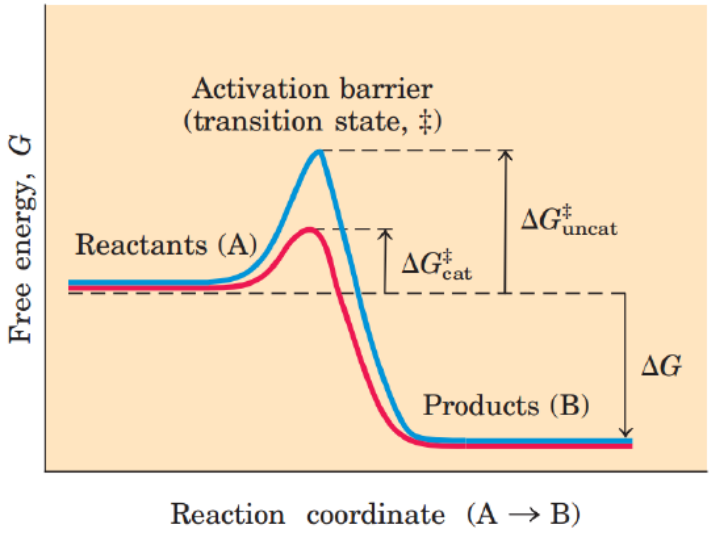

Figure 1.1: **Catalyzed (red) and not catalyzed (blue) reactions.** As we can, ΔG would be the same in both reactions. However, the effect of the protein will vary $\Delta G^{\ddagger}$ corresponding to the activation barrier. Figure adapted from Principles of Biochemistry (Lehninger). Fifth Edition.

In almost all cases, the reactions that occur in a cell are accelerated with respect to the non-catalyzed reaction. NTPases are not an exception. Enzymes accelerate reactions by reducing what is called the activation barrier: the energy barrier that it is necessary to for

reactants to overcome to products, thereby producing the reaction. In the case of enzymes, this activation barrier can be reduced by the stabilization of the chemical reaction thanks to the macromolecular environment (see figure 1.1), which stabilizes the charges and structural geometry during the reaction, thereby increasing the reaction rate.

Most of the time, during enzyme catalysis, the protein undergoes some conformational changes in its secondary or tertiary structure; these changes facilitate the catalysis of the substrate and the liberation of the products of the reaction. These conformational changes are also associated with the process of the binding and unbinding of the substrate. In the present case, although ATP / GTP hydrolysis only involves the release of heat, because of the energy stored in the protein structure, which is spent during synthesis, this binding and releasing of the products is enough to favor protein function from the structural point of view. One example we can find is the case of the myosin protein, in which, the processes of binding, the hydrolysis of ATP and the release of ADP+P_i provoke a whole cycle of attachment to and detachment from actin filaments in muscle contraction (see figure 1.6). Another example can be found in F_oF_1-ATP synthase protein, which, thanks to the energy employed in its synthesis allows the chemical equilibrium to be displaced towards ATP synthesis only in the presence of ADP and P_i (see figure 1.5).

The hydrolysis of these nucleotides is favored by the protein environment, which allows, among other factors, the necessary geometrical configuration of a water molecule, which is for the reaction; and which at the end of the reaction, will form part of the released Pi, forming a covalent bond between the oxygen atom of the water molecule and γ-phosphate (specific reactions are described in Results section below). This correct geometry of the catalytic water molecule is adopted thanks to the residues and ions that constitute the catalytic site of the protein.

Generally, hydrolysis is a well known process, where two kinds of reactions can be produced, depending on the proteic environment, ions and cofactors at the catalytic site. They are the S_N1 and S_N2 reactions. S_N2 (also named the associative reaction) consist of a reaction in which a nucleophilic substitution is produced. A nucleophile particle, the electron donor, attacks an electrophilic particle with a deficient number of electrons. This

provokes the unbinding of a leaving group that was previously bound to the electrophilic particle. In the case we study, the ATP / GTP molecule, the γ-phosphate is the electrophilic group, which is arranged in a tetrahedral conformation (sp3 hybridization) with respect to its three oxygen atoms and the oxygen atom that bonds between γ and β-phosphate, at an angle of 120º with respect to the oxygen atoms. The nucleophile particle is the catalytic water molecule. Transition to the products of the reaction is produced through an activated complex (as we will see in the Transition state theory section below), whose geometry is arranged as a trigonal bipyramid (sp3d hybridation). This situation leads to a weakening of the bond between γ-phosphate atom and the oxygen atom that bonds with the β-phosphate.

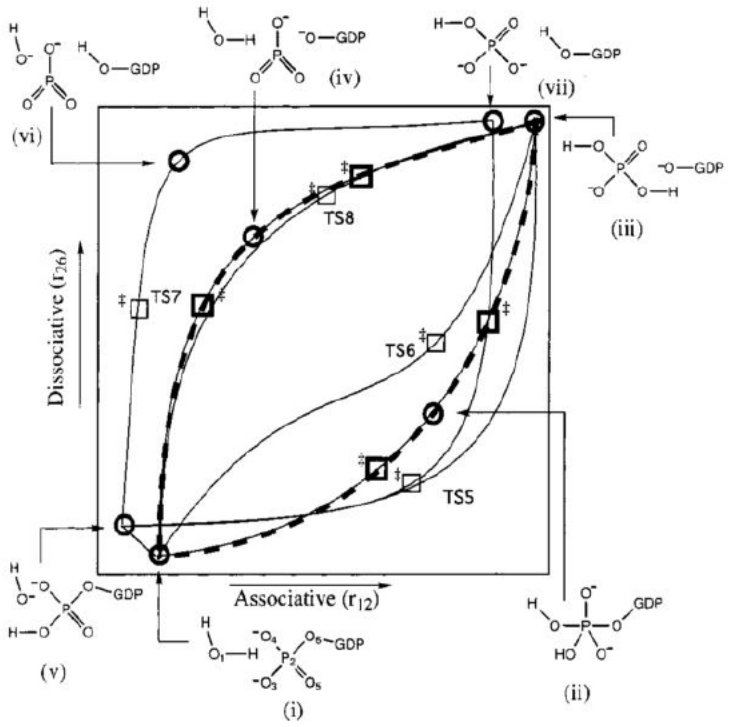

Figure 1.2: Reaction surface for the associative and dissociative paths in a hydrolysis reaction of GTP. (i) represent initial structure before hydrolysis. In this scheme, (ii) and (iv) would represent ideal transition states of S_N2 and S_N1, respectively. Final state (iii) would be very close to final state observed with protein environment. Figure adapted from Warshel et al (2000) (Biochemistry)

In the case of S_N1 (dissociative, figure 1.2) reactions, nucleophilic attack can not be produced because of steric impediments, so the passage from reactants to products is produced in two stages. The first step is the weakening of the bond that the living group forms with the carbocation (in the case of nucleotides, the rupture of the bond between P_i and the nucleotide diphosphate). This situation forces PO_3^{-3} molecule to adopt a planar disposition (sp2 hybridation). This disposition of the ligand is stable thanks to the disposition of the protein residues, and outcome of the first step has been called the intermediary of the reaction. This intermediary is different from the activated complex since in the case of the intermediary, it is delimited by the activation barriers of the two steps of the S_N1 reaction while activated complex is the highly energy structure of that activation barrier. The second step is the nucleophilic attack of the catalytic water molecule over this planar molecule, which generates an $H_2PO_4^-$ molecule.

Focusing on S_N2 reactions, another important point to the hydrolysis of ATP/GTP molecule is that the catalytic water molecule needs to be activated. Activation of the water molecule implies that one of the protons of the water will be accepted by one of the elements of the catalytic site of the protein or substrate (which acts as the base in the reaction), thereby generating an OH^- molecule, which can then react with the positively-charged γ-phosphate atom.

However, sometimes, identification of the acceptor and therefore the base in the reaction, is not as straightforward as we might assume. Mutational analysis helps us to identify which residues are important for the hydrolysis of the nucleotide and so, for the activity of the protein. Results from that analysis, together with structural studies obtained from the resolved NMR and crystal structures, have led to propose some hypotheses with regard to the possible acceptor of the proton and therefore what is responsible for the activation of the water molecule in the case of NTPases. However, because of the short lifetime of the activated complex or intermediaries, evaluation of the residues role of the residues is complicated. At this point, computational analysis has proved to be a very powerful tool, and moreover, in recent decades. Improvements in the force fields, a set of data that defines atomic interactions based on experimental data, and the combination of different MD

techniques with QM have allowed us not only to obtain a detailed computational description of biological systems, but also to simulate reactions using simulation times of the order of femtoseconds, in which laboratory conditions are reproduced. In this way, we can control the reaction coordinates of a chemical reaction and, as explained above, discriminate whether the hydrolysis of a nucleotide in a protein is achieved through an S_N1 or S_N2 reaction.

The hydrolysis of nucleotides is critical in the function of a large number of proteins. In the living cell, we find a lot of chemical processes in which nucleotides such us ATP and GTP play an important role. In glycolysis, for example, glucose is transformed into glucose 6-phosphate by a hexokinase making use of an ATP molecule, where a phosphate group (PO_4^{-2}) is donated from the ATP to one of the oxygen atoms of glucose. A total of four ATP molecules are involved in the whole process, in which a protein is involved in the reaction, hydrolyzing the ATP molecule in ADP and P_i.

A third element necessary for hydrolysis, together with the protein and the nucleotide, is the water environment, where most biological reactions are take place. Although some proteins present closed catalytic sites, during nucleotide exchange, they are exposed to water environment (because of conformational changes or protein-protein separation, like Hras-GAP system). This event permits the access to water molecules to them, which will be enclosed the catalytic site close again. On the other hand, other proteins present catalytic sites that are exposed to water environment, like FtsZ. This event provoke that a single water molecule, helped by the protein environment (or even assisted by the own water environment in what has been called the Grotthuss mechanism[1]), is able to perform a nucleophilic attack on the γ-phosphate, hydrolyzing the nucleotide.

1.1.1. FtsZ: GTP and its role in bacterial division

FtsZ[2,3] is a bacteria GTPase division protein, homolog to tubulin which, in the same way, is able to polymerize in long filaments, in a head-tail disposition (figure 1.3a), and which depends on GTP binding. In *in vitro* experiments, it self-associates forming linear or curved[4] filaments. *In vivo*, it forms a structure known as the division ring or Z-ring, a

polymeric structure that attaches to bacterial membranes, acting as a scaffold for the division machinery, an ensemble of different proteins that receive the name of the divisome. It can be found in most eubacteria and archea and also in chloroplast and mitochondria[5–7]. This protein, which is sensitive to the environmental pH and the presence of different ions (according to the bacterial strain), needs two monomers to constitute the minimal catalytic unit. An active site will be formed by two monomers of FtsZ in a head-tail disposition. In the middle of the interface we find a GTP molecule, which together with a magnesium ion and additional ions such as K+ or Na+[8], is capable of forming polymers. Not much is known about how the hydrolysis of GTP place. A closer look at the catalytic site reveals the presence of aspartic acid residues, which together with the ions in the environment, could act to catalyze the reaction.

A more extensively exists regarding how the Z-ring is organized[9,10]. In this way, single (in an open or close disposition) or multiple filament structures, with lateral association (figure 1.3c, have been proposed. Another debate regards whether the Z-ring itself structure and therefore, the different filament associations, could constitute the force mechanism necessary to begin bacterial division, by provoking the curvature of the membrane[11]. At this point, however, a molecular motor that produces such a force has not yet been identified, in contrast with other mechanisms such as actin/myosin or tubulin/kinesin systems, among others. GTP hydrolysis to GDP would provoke the aperture of the catalytic site or provoke depolymerization.A central location of the Z-ring is necessary for a proper division of the cell. Two principal mechanisms are known. The first is called the *Min system:* a system composed of three proteins (C, D and E), that oscillate from one pole of the cell to the other. This oscillation blocks the Z-ring from locating at the poles of the cells. Mutational studies of these proteins have shown viable cells but which divide at the cell pole, generate small cells without DNA[12,13]. Basically, MinD is located in the peripheral region, at the poles, and only at one of them at any given moment, oscillating between the halves with a period of between 10 and 30 s. In the same way and associated with MinD, we have MinC, which oscillate in the same way as MinD, destabilizing the polymerization of FtsZ ring in the poles. In the absence of MinE, the MinC/D system would be located at one of the poles,

without observe oscillations. MinE migrates as a ring from the center of the cell to the poles, destabilizing the MinC/D system, at one of the poles and preventing from locating in the middle of the cell, where MinE forms again the ring.

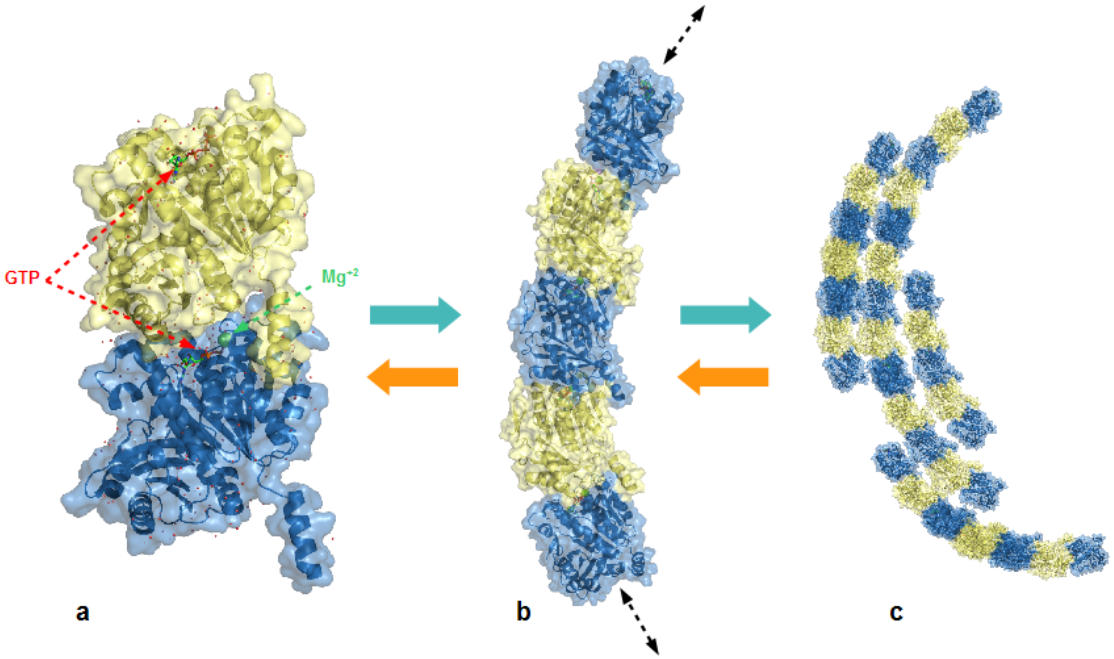

Figure 1.3: Crystal structure of FtsZ dimer (PDB code: 1W5A). Minimal catalytic unit is formed by two monomers of FtsZ (a). The nucleotide binds to the upper site of a FtsZ monomers (blue), while other monomer (yellow) close the catalytic site, in a head-tail disposition. Coupling of different monomers permits to form different length polymers (b), which interact between them forming, Z-ring structure These kind of interactions will consist on head-tail and lateral interactions between polymers (c). Blue arrow indicates polymerization. Orange arrow indicates depolymerization.

The other system is what has been called the nucleoid occlusion. This system, studied by Mulder and Woldringh[14], prevents the formation of the Z-ring anywhere in the cell other than the center of the bacteria; that is, the presence of the nucleoid (chromosome) alters the assembly of the Z-ring. A mutation in the mechanism that governs the distance from the nucleoid to the poles (such as an alteration of the Min system or mutations in cell division proteins) or in chromosome segregation proteins leads to the formation of multiple Z-rings

in free-DNA areas of the bacteria. This effect is not observed in wild type cells, where the segregation mechanism seems to be centered, which creates a small gap where the FtsZ ring polymerize. One of the proteins that take part in the separation and segregation of the chromosome is FtsK[15], a large multispannig protein involved in cell division together with the rest of the divisome proteins and which requires the presence of FtsZ, FtsA and ZipA to localize at the midcell division site[16–18].

As we describe above, the FtsZ ring acts as a scaffold for several proteins, constituting the divisome. The second most conserved protein after FtsZ is FtsA. This protein, recruited in the first steps of cell division, and one of the actin homologous proteins in prokaryotes, seems to play an important role in Z ring stabilization and the recruitment of the rest of the divisome proteins. Although it is capable of binding ATP, it is not known to have any hydrolase activity (with the exception of *B.subtilis* FtsA). FtsA interacts with FtsZ through the c-terminal of the latter and it is capable of anchoring to the cell membrane through a conserved amphipathic c-terminal region; this has led to belief that it could anchor FtsZ to the membrane. Such a role would be shared with ZipA, which is only found in γ-proteobacteria, and which also possess a c-terminal helix that can anchor to the membrane, inserting it into it.

The rest of the proteins that have been identified as being involved in the divisome formation to date (FtsQ, FtsB and FtsL, which form a periplasmic connector; FtsW and FtsI, involved in the peptidoglycan synthesis; and FtsN, invloved in the connection with the peptidoglycan) assemble after early assembly of FtsZ, FtsA, ZipA and FtsK[19].

Based on previous results regarding the independence of the active sites in the filaments of FtsZ[20], we performed a MD study of several filaments, in order to study the behavior of the catalytic sites along the polymer.

1.1.2. Ras: GTP in signal transmission

G proteins (guanine nucleotide binding proteins) constitute an extended family of proteins that are involved in several functions. They are presented as signal transducers attached to the cell surface plasma membrane. This family of proteins act as intermediate in the signal

transduction function between receptors and effectors within the cell. Among some of these receptors that interact with G proteins we can find receptors of hormones, neurotransmitters, chemokines, and autocrine and paracrine factors[21]. These G proteins, after receptor interaction, interact with several proteins thereby transmitting the signal that leads to the cellular response. Some of this effectors are ion channels, metabolic enzymes and the transcription machinery that produce the expression of specific proteins according to the response.

In this large family of proteins, we find the Ras oncogene protein superfamily, a group of low molecular weight proteins, the major representative of which is the protein Ras-p21. Ras proteins are targeted at the plasma membrane by posttransductional modifications in the carboxyl-terminal region, a situation which is critical for their function. Some pathways that Ras proteins take part in have been studied in detail. This is the case of growth factors such as the epidermal growth factor[22]. In that pathway, the Ras protein is activated when the Sos protein, which acts as guanosine exchange factor, is recruited from plasma to membrane by the receptor tyrosine kinase and adapters such as Grb2.

As with the rest of the G proteins family, these proteins are able to bind GTP and hydrolyze it to GDP + P_i. They act as molecular switches, that is, we can find them in a GTP-bound state (activated state) that through a hydrolysis reaction leads to a GDP-bound state (inactivated state). This process reverts by the substitution of GDP for a GTP molecule, thereby the activated state (figure 1.4). Small G proteins show a higher affinity for GDP than for GTP. This means that in a situation where no stimuli activates the signal cascade, the small G protein will remain in the inactivated state. The presence of a stimulus will provoke the replacement of GDP for GTP, mediated by GEF, which in turn will provoke some conformational changes in the structure of the protein that are ligated to the interaction with receptors[23]. The hydrolysis rate of GTP in these proteins is very slow, which is very closely related to their function as molecular switches. In opposition to the general idea that enzymes evolve towards optimization of the rate k_{cat} / K_M, G proteins have evolved with a highly flexible and mobile active site, which provides this switching behavior[24]. The presence or the absence of the γ-phosphate of GTP provoke conformational

changes. Two local regions accumulate most of these changes, permitting interaction with GAP and GEF proteins[25–27].

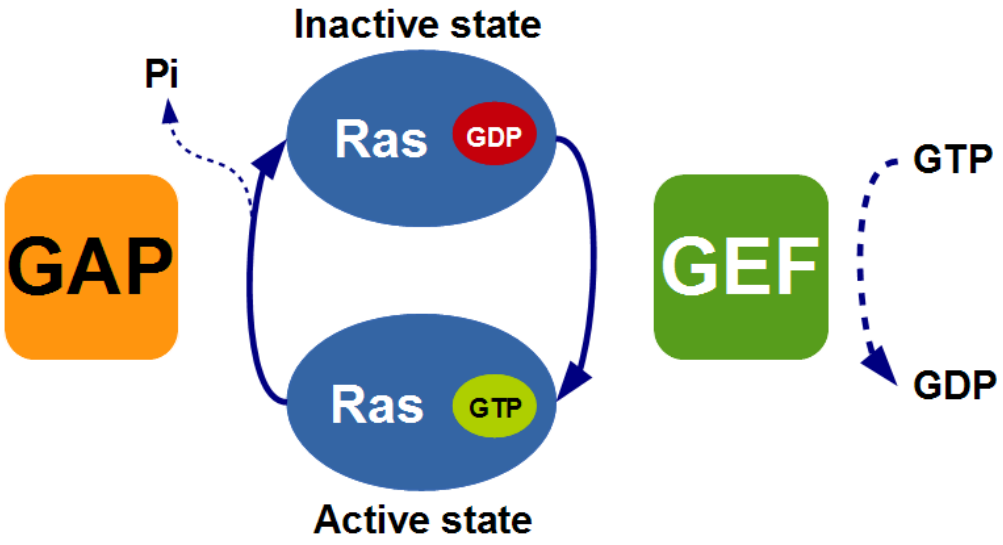

Figure 1.4: Schematic representation of Ras protein as molecular switch. Ras protein alternates from an inactive state (GDP-bound) to an active state (GTP-bound). This transition is helped by external proteins GAP (GTP activation protein) and GEF (Guanosine exchange factor).

Transition from the active to the inactive conformation is governed by the presence of external proteins. When the Ras protein is bound to GTP, hydrolysis is regulated by GTPase activating proteins or GAP, whose function consists of closing the catalytic site of the Ras protein, adding to the Ras catalytic site an additional element called "the arginine finger", an arginine residue, identified by Wittinghofer et al.[28], and which is common in G proteins. The arginine finger binds to GTP between the γ-phosphate and β-phosphate, thereby favoring the weakening of the bond formed by the bridging oxygen atom between γ and β phosphate atoms and the own γ-phosphate itself when a catalytic water molecule attacks the latter. The effect of the attachment of the GAP to Ras-p21 can be seen via the rate constants, where the difference is of several orders of magnitude or the activation barrier, where a difference of more than 10 kcal·mol^{-1} can be observed. Mutational analysis of this

residue[29] has demonstrated the direct influence it has on the reaction, reducing the rate constant ~2000 times, which is related to an increase in the activation barrier of around 4.5 kcal·mol^{-1}. Once the GTP has been hydrolyzed, the Ras protein binds to another protein, the guanine nucleotide exchange factor or GEF. The role of that protein is to catalyze the release of the GDP ligand, which is tightly bound to Ras, and order to replace it with a GTP molecule[30].

The involvement of additional proteins in the hydrolysis of nucleotides is not only confined to G proteins or GTP. We find several cases in which external proteins help in this function. In this way, for example, Actin is necessary for the release of ATP (or concretely, ADP) in order to close the power-stroke cycle produced in muscle contraction. In this way, actin acts an ATP activation protein (AAP). Another example involving ATP could be the effect of DnaJ on the bacterial chaperone DnaK.

Although a lot is known about the function and structure of that protein, some points are still controversial. This is the case of how the hydrolysis reactions is produced. Some studies, based on structural data (reference), have established that Gln 61 is responsible for the activation of the water molecule and therefore is base in the reaction. However, we can find in the literature other studies that establish that this residue plays a structural role and that the substrate acts as the base; a mechanism known as substrate assisted catalysis. In the Result section below we will discuss the role of this residue in accordance with the theory employed.

1.1.3. F_oF_1-ATP synthase: An example of a molecular motor

When we talk about molecular motors or nanomachines, one of the most representative and widely studied is F_oF_1-ATP synthase. This enzyme is responsible for the synthesis of ATP using ADP + P_i as substrates of the reaction. The energy for ATP synthesis is provide by an electrochemical proton gradient that exist between two spaces separated by a membrane. In the case of mitochondria, this gradient is created between the intermembrane space and the matrix, which are separated by the inner membrane. In accordance with this, mitochondrial F_oF_1-ATP synthase is located in the inner membrane. As a general description of its

structure, the protein is formed of two portions (see figure 1.5): F_0, which is the part anchored to the inner membrane in a transmembrane disposition, and whose principal role is the proton translocation; and F_1, which is the portion with ATPase activity and which points to the mitochondrial matrix, and is 90-100 Å in diameter. The two structures can be separated under low ionic strength conditions, in order to gain a better understanding of their functions.

F_0 is formed of three different kinds of subunits, *a*, *b* and *c*. The synthesis of ATP initializes by the proton gradient passing along the *c* subunit of F_0: a cylindrical structure composed of a variable number of monomers depending on the organism (in the case of mitochondria, this value oscillate between 9 and 12). Every monomer is formed of two hydrophobic helices bound by a loop of 18 residues, pointing towards F_1 portion. The *c* subunit is in a transmembrane disposition. The principal characteristic of the structure is that it presents free rotational movement with respect to the membrane, and it is be the base for the function of this protein. Attached to this structure we find the *a* subunit, a monomeric protein that is positionally anchored to the membrane. The transference of the proton gradient is directly related to the disposition of these two structures, *a* and *c*, whose interaction surface consists of two regions (see figure 1.5) that allow the capture and release of protons from one side of the membrane to the other, thereby causing *c* to rotate with respect to *a*.

In the case of this structure only the c subunit has been completely resolved (PDB entry: 1C17) while only a low resolution topology of the *a* protein in membrane is has been achieved to the date[31–33]. The *c* subunit is bound via a polar loop to the γ subunit, a coiled coil structure, that directly maintains together the F_1 and F_0 portions and which with the ε subunit, constitutes what has been called the rotor stalk, a narrow structure of approximately 45 Å in length. When the ε subunit is not attached to the γ-subunit, it present a β-sandwich domain, with a c-terminal region composed of two antiparallel helices, in a helix-turn-helix conformation. When ε is attached to γ-subunit, the c-terminal helices are wrapped around the γ-subunit. It has been suggested that this conformational disposition inhibits ATP hydrolysis by regulating stalk rotation[34]. As mentioned above, the stalk rotates

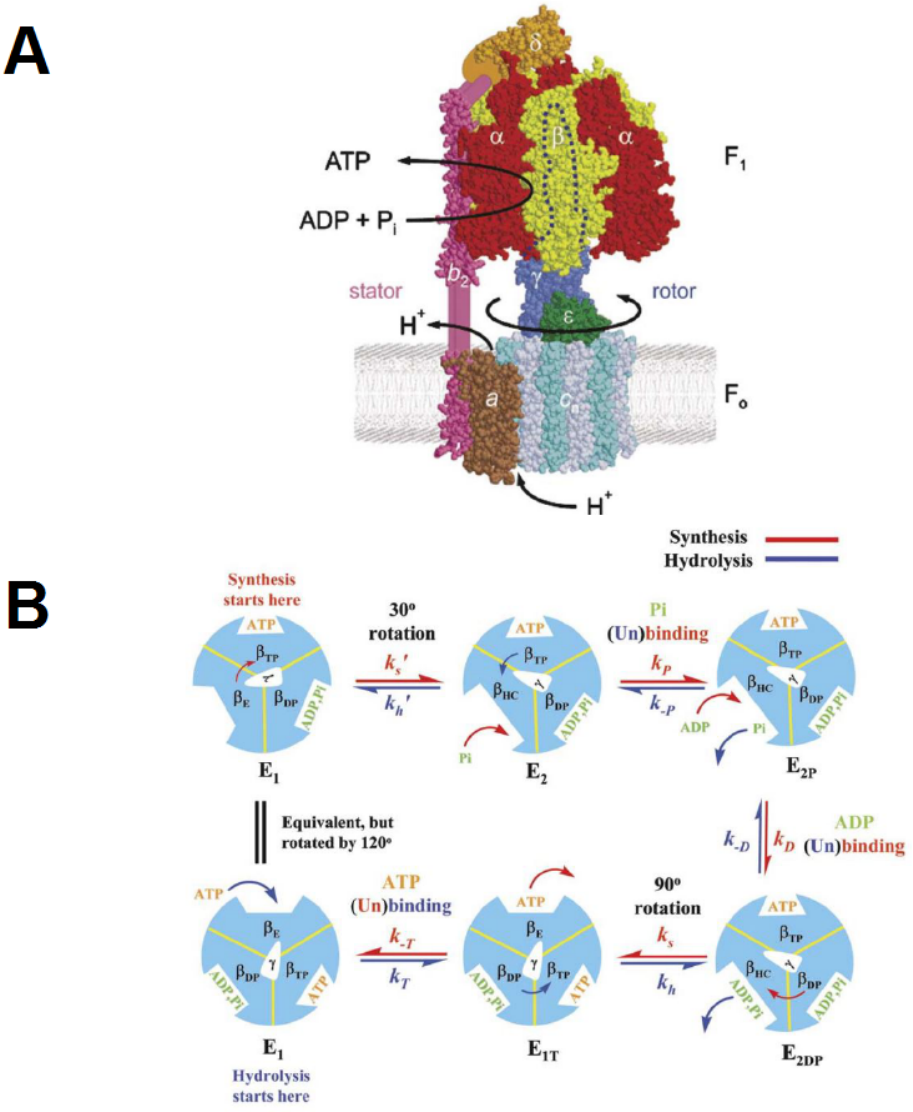

Figure 1.5: (A) Schematic representation of the whole system F_0F_1 ATP synthase. Picture shows how both sections of the protein are disposed respect to the membrane. F_0 would be disposed in a transmembrane disposition, formed by *a* protein and the *c* complex. Stator and rotor stalk would bind F_0 with F_1 region, composed by α and β proteins in an hexameric disposition. Proton gradient to the matrix of mitochondria would provoke the rotation of *c* complex respect to *a* protein, and therefore, rotation of the rotor stalk, formed by γ and ε. Because F_1 is bound to *a* throughout *b* and δ, this will provoke the rotation of the rotor respect to F_1, producing ATP from ADP + Pi. Figure adapted from Senior et al. (2002) (BBA). **(B) Representation of the synthesis / hydrolysis reactions of ATP in F_0F_1 ATP-synthase.** In the picture it shown synthesis (red) and hydrolysis (blue) routes and the disposition of the three β monomers according to the disposition or the rotor stalk (γ protein). The three states of β subunit is indicated: β_{DP}, β_{TP} and β_E. The scheme would show a rotation of 120°. ATP is indicated in orange and ADP/Pi is indicated in green. Adapted from Gao el al. (2005) (Cell)

due to the proton gradient as it passes along F_0. This subunit internally located within the globular structure of F_1-ATPase. F_1-ATPase is formed of 5 types of subunits: α_3, β_3, δ and the already mentioned γ and ε, whose atomic masses are approximately 55, 50, 19, 31 and 14 kDa[35], respectively. The globular part of this structure is composed of six subunits, three subunits and three beta subunits, which possess one fifth of sequence identity, and share a high similarity in structure. When α and β subunits are isolated, they conserve their strong tendency to bind nucleotides to their binding sites, but, they are not capable of hydrolyzing ATP. The presence of both structures, α and β, is required to constitute the catalytic site. Furthermore, F_1-ATPase can not synthesize ATP in the absence of the F_0 subunit and the stalks. Because of its hexameric structure, F_1-ATPase has six catalytic sites, but only three of them are considered catalytically active, those where ATP is bound to the β subunits[35]. The three sites are asymmetric and show very different binding affinities in Mg-nucleotide studies[36,37]. The different affinities are caused by the geometry of the γ-subunit, which interacts with the hexameric region, and presents three different faces. As we have seen before, the proton gradient provokes the rotation of F_0 and the rotation of the rotor stalk.

A consequence of this rotation of the stalk, that internally locates in the globular region of F_1, is a transition between the three different catalytic sites structures: β_{TP}, β_{DP} and β_E which are considered to be the sites to which the ATP and the ADP (and no nucleotide), are bound respectively. The sites were defined in the original crystal structure[38] (a fourth site defined as a half-closed conformation, β_{HC}, was also described by Menz et al. In 2001[39]). The transition is caused by the asymmetric structure of the stalk in this region, which provokes a conformational change in the β subunits for every 120º of rotation that is related to the synthesis of an ATP molecule. MD studies revealed that after the opening motion of the β subunits due of the rotation of the γ subunit, a closing motion appears to be spontaneous after ligand binding. These data very closely reflect with the conformational changes described by normal mode analysis, establishing that these motions requires a low energy cost[40]. In the case of hydrolysis, different studies establish that the different affinities of the catalytic sites for the different ligands (ATP, ADP+Pi, or ATP+H_2O) in accordance with the crystal structures[41,42], provoke the transition from β_E to β_{TP} and to β_{DP} and, as a consequence,

the rotation of the rotor stalk.

During the transition from β_{TP} to β_{DP} the hydrolysis reaction takes place. In the crystal structure, Abrahams et al[38] concluded that β-Glu 188 was the prime candidate for the hydrolysis of ATP because it was hydrogen bonded to the catalytic water molecule that should attack to the γ-phosphate. However, this residue requires a very hydrophobic environment because its low pKa. Taking into account the active site of this protein, conditions are more favorable for a substrate assisted catalysis reaction. In the Result section below we will the possible mechanisms depending on the activation water molecule.

1.1.4. Myosin in muscle contraction

Myosin constitute a superfamily of molecular motors involved in muscle contraction thanks to their interactions with actin filaments in all eukaryotic cells. This superfamily is also involved in a wide range of functions that involve displacement, including cell movement, cytokinesis, vesicle transport, Golgi organization and sensory transduction[43]. The structure of the protein can vary depending on the myosin type, but all respond to a general structure, with a head region, which is the part that attaches to actin filaments and is where the hydrolysis reaction of ATP is occurs. The head domain is completed by a lever arm, which, according to ATP hydrolysis cycle, causes the head to move. That region is bound to a coiled-coil region. To the date, more than 35 classes of myosins have been identified, being 13 found in humans. Single or double headed myosins can be found, but all of them with the capacity to bind to a second protein.

The hydrolysis of ATP in this protein is linked to a series of conformational changes that govern its attaching to actin filaments. The hydrolysis can be defined by four steps: the "rigor state", in which myosin and actin are bound together in the absence of ATP; the "post-rigor state", in which ATP is bound to myosin which no longer interacts with actin; the "post-recovery state", in which a conformational change in the structure has occurred in

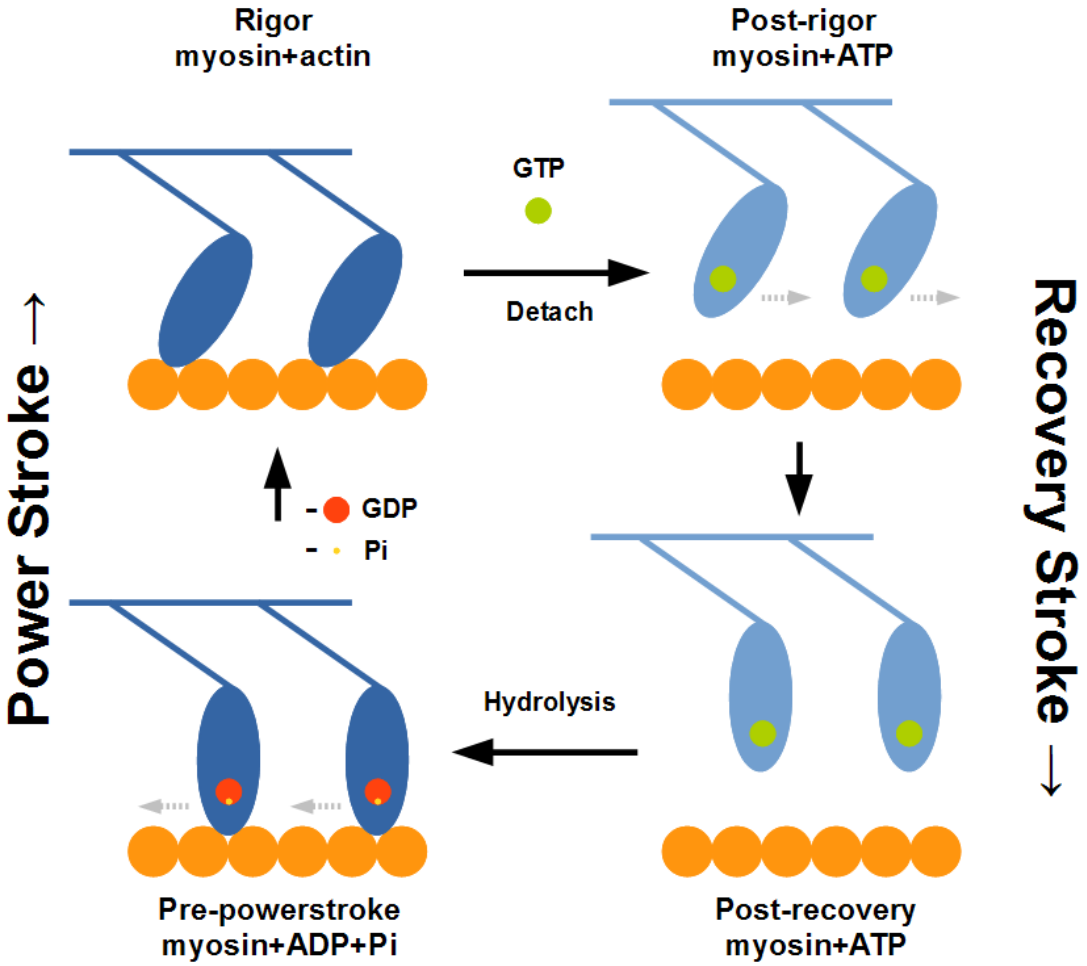

Figure 1.6: Myosin cycle of attaching – detaching of actin. Blue structures represent myosin protein. Orange structures represent actin filaments. The binding of ATP (green) provoke the detaching of myosin. This provoke a conformational change (post-recovery) what leads to the final hydrolysis of ATP to ADP (red) and Pi (yellow). In this step (pre-powerstroke), the myosin attach to actin filament again. The liberation of hydrolysis products provoke the stroke of myosin, displacing actin filament, completing the cycle.

the motor domain with respect to the lever arm; and the "pre-power stroke state", in which ATP hydrolysis occurs. The release of reaction products leads the transition from the pre-power stroke to the rigor state, during which the myosin II binds to the actin filament again and recovers the initial disposition of the head with respect to the lever arm[44,45]. The hydrolysis of ATP is governed in accordance with to the disposition of the different elements that form the catalytic site. These are the switch I (part of U50), P-loop and switch II (that belongs to L50). The states of both switches can be defined as closed or open. The closed state of switch I can be defined as when Ser 237 is forming part of the coordination sphere of the magnesium ion. In the case of switch II, the closed state is defined as when Gly 457 form a hydrogen bond between its mainchain and the O3γ of the γ-phosphate of ATP. We would talk about of an open conformation when the protein lose these interactions.

1.2. Computational simulations and their application in structural biology

MD is a set of computational techniques that allows us to study the energetic and structural properties of molecular systems with atomic-level resolution. The main aim of MD is to describe and explain macroscopic events, based on a microscopic description, using single molecule systems as the starting point, and providing the basic elements to solve the equations of motion that govern these molecular systems, and that are applied to single atoms. From the analysis of the different data obtained, we can obtain relevant information about the different biological functions in which the systems are involved. In this way, we can gain information about their main roles and interactions with macromolecular systems such as nucleotides or polysaccharides, or even small molecules such as ligands, cofactors, substrates and inhibitors.

We have to consider that this kind of computational approximations takes into account macromolecules as physical systems, and that atoms are the minimal units that form them. So, in this way, the behavior of the macromolecules will be governed by the interactions of

their atoms. Every single atom is considered as a punctual mass whose movement is governed by the forces that are projected onto it by the other atoms that conform the system. Because of the large number of atoms that constitute these systems (N), the analytical resolution of the equations of motion for the potential energy function is very complex, so a numerical solution is applied. Although there are some analytical solutions, their convergence is very slow to be considered.

For a particular atom, if we know the exactly position at a specific time, that is, x(t), the new position after a short time period, Δt, will be given by a standard Taylor series (equation 1.2.1):

$$\sum_{n=0}^{\infty} \frac{f^n(a)}{n!}(x-a)^n \qquad (1.2.1)$$

Taking into account position, velocity and acceleration, and also calculating the contribution of the following derivatives in the series, the new position of a single atom in Δt can be calculated with a reasonable precision.

Owing to the fact that the number of series that can be calculated for x(t+Δt) may be infinite, it is necessary to truncate the series at a specific point. Different approximations exists that allow to integrate the equations of motion and we can find different algorithms depending how positions and velocities are calculated, such as the Verlet[46] or "leap frog" algorithms.

MDs is based on the Newton's second law or equation of motion (equation 1.2.2):

$$\frac{\delta V}{\delta r_i} = F = m_i a_i = m_i \frac{\delta^2 r_i}{\delta t^2} \qquad (1.2.2)$$

where F is the force applied to every single atom *(i)*, m is the mass for every atom and a is the acceleration for every one of them. The integration of equation 1.2.2 is considered the starting point to obtain the positions, velocities and accelerations. As a consequence, this is considered a deterministic method, taking the positions and velocities from a former step

and obtaining new positions for the following steps. In this way this is considered as an iterative process that will need starting coordinates. These positions will be taken from experimental data obtained by x-ray crystallography or nuclear magnetic resonance (NMR).

Another important aspect of MD is that we can represent the energy of a protein as a function of the atomic coordinates, and in so doing, we find that the expected states in thermodynamical equilibrium correspond to minimal energy values of the energy function.

The gradient function, that is, the variation of the energy with respect to the positional vector (and therefore, the variation of force as the energy derivative), will orient the force of the single individual atoms of the system. As a result, these energy functions are called force fields.

MD is a very useful tool since we can obtain important biological information from the results. However, the dependence of MD on force fields represents a limitation. The force fields, which contain an accurate parametrization of bonds, angles and dihedrals between the different kinds of atoms, are also a limitation from the point of view of bond breaking and formation. These events, which are present in all chemical reactions, can not be simulated by MD, since all bond lengths are represented by a harmonic potential that maintains the ideal value of the bond with a constant force set to tens of kilocalories per mol and Angstroms. This limitation bring about the necessity for a different approximation, that is, the QM approximation, where bonds are not handled by harmonic potential restrictions, but the electronic cloud (and therefore the specific configuration of the molecular orbitals) of every single atom is considered. This permits the interaction of atomic orbitals between the atoms, creating and breaking bonds in accordance with the atomic environment. Because NTPases are very large systems (and larger if we consider the aqueous environment and the hydrogen bond network between the water molecules), a QM approximation of the whole system would be unaffordable from the computational point of view. This is because every single interaction of the orbitals of the surrounding atoms would have to be calculated (within a specified cut-off), which can not be calculated in a reasonable time with the current computational systems. However, if we want to study an enzymatic reaction, we must consider the protein environment. The most reasonable

approximation (from the computational point of view and that of the reliability of the results) is what has been called the QM / MM method, which was implemented by Warshel and Levitt in 1976[47], and standardized by Karplus et al in 1990[48]. The method is based on considering one part of the system through the QM calculations (the catalytic site) and the rest of the system with MM calculations.

This approximation allows us to study the reaction in the catalytic zone, taking into account the effect of the protein environment. In such a situation, a good definition of the catalytic site is needed, because the lack of one of the elements that take part int the reaction could generate undesirable unreliable results.

1.3. The energy surfaces information

The study of chemical reactions in proteins using the QM/MM approximation is performed using what has been called the energy surface maps. These surface maps are kinetic approximations of what can be seen in topology maps. As in topology maps, we observe hills and valleys of energy (QM energy or potential energy), where the valleys represent chemical structures with lower energy, and therefore, strucuturally stables. What we find in these valleys are the reactants and the products of the reaction, specifically if we consider the lowest minimums. In contrast, we will also observe hills and peaks in these maps, that represent the chemical structures with the highest energy (less stables considering the structure), compared to the reactants and products. Because the energy necessary to produce these structures is very high, they will seldom, if ever, be achieved. Analogously, these regions represent the tops of mountain, where the energy employed to get there is very high.

A third important element of these maps are the saddle points. These regions represent the barriers that a molecule needs to pass through to get from one valley to another one. Saddle points can be explained by considering energy surfaces as variations in the energy system, δE, with respect to the position, δx. If we take into account the first derivative of the energy, that is, the force, and we consider that a stationary state will be one which no force is applied, then we will be at a minimum or a maximum of the function. The second

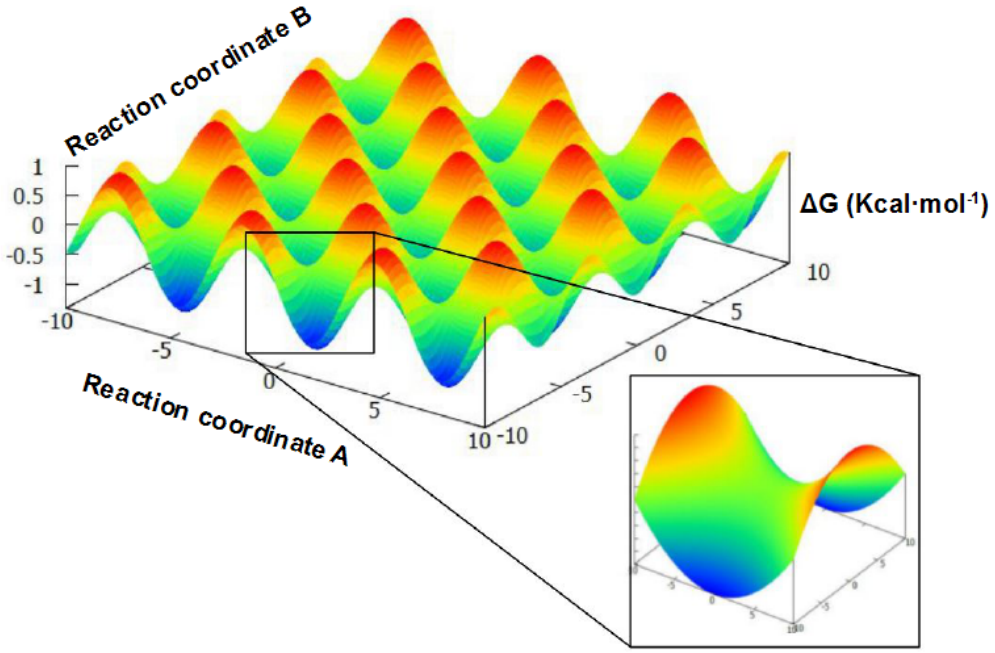

Figure 1.7: Hypothetical energy surface map generated with a sinusoidal function. Map was constructed with a sin(x)*cos(y) function in order to illustrate the different elements that will be found in our studies, detailed in figure 1.8. X and Y axis would represent the reaction coordinates, while Z would be the increment of Gibbs energy. Color code will be the same that we have followed in our studies, shown in the results section. In this way, blue regions corresponds to stable conformations of the system. Red regions would correspond to regions with a high energy, and therefore, the structures would be less stables from the energetic point of view.

derivative will tell us in which we are (depending on whether it is negative of positive, respectively). At a saddle point, it will be a minimum in all directions (the second derivative will be positive) except in one, and that will be the reaction path, where it adopts the behavior of a maximum, and therefore the second derivative will be negative.

The molecular complex associated with saddle point has been called activated complex or transition state. Both terms, activation complex and transition state are commonly used to refer to the same structure. A transition state is considered to be the point on the curve that represents the activated complex. The height of the barrier it is necessary to get the transition state is called the activation barrier and the energy necessary to pass the barrier is the activation energy. We have to consider that a representation of a transition state is a

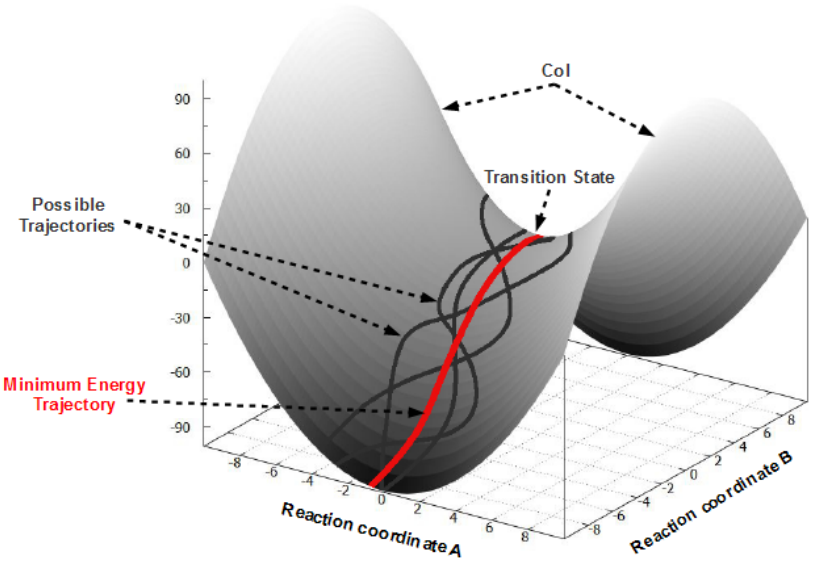

Figure 1.8: Multiple reaction pathways through transition state energy barrier. Every trajectory colored in black represents a slightly different path a reaction may take depending, as we have seen, of the initial coordinates, starting energy and atomic motions. Red line would represent the reaction following the minimum slope in the surface throughout metadynamics. None of the reactions pass through the same point, obtaining an ensemble of reaction paths.

hypothetical structure that can be considered as the "average" structure of the different possible structures that can be found at the saddle point, that contains characteristics of reactants and products, and represent the most unstable structure of the reaction. The main characteristic of the transition state is that its life time is lower that the vibrational move of a bond. This particular vibration of atoms tilts the transition state towards the products. These energy surface maps are plotted against two reaction coordinates, that represent two bonds in the system. However we have to consider that the number of real coordinates of the system is of the order of 3N-6, where N is the number of degrees of freedom, that is, the number of atoms in the system, and would lead to a hypersurface representation, that would be very difficult to imagine and represent. Taking into account all these elements (valleys, peaks, transition states and reaction coordinates), we can obtain an approximation of the possible path that a reaction will follow, by observing the path of minimum slope on the surface.

However, we have to consider that this is not going to be the only pathway along which the reaction passes, but it will be the most probable. Most of the reactions that start at the reactants will pass through the transition state, but, because molecules and atoms are associated to different vibrational modes, and also depending on the initial coordinates, velocities and momenta of the molecules, they will be in different ground state or excited states. This situation will provoke different vibrational states and therefore, different structures, which means that the pathway followed to the products will be slightly different. So, as we have seen above, there will not be a single path to pass from reactants to products, but an ensemble of reaction paths.

An additional element can be found in energy surface maps, known as an intermediary state. It is considered to be stable structure found at a transition state point, so this intermediary state will be surrounded by two transition states, one that separates it from the products and the other between it and the reactants. depending on the height of these barriers, the lifetime of the molecule will be long (if the barriers are high) or short (if they are low). Any chemical structure that remains in a transition state for longer than the vibrational frequency of a bond (between 10^{-13} and 10^{-14} seconds) can be considered as an intermediary.

1.4. Transition state theory

One of the theories used to explain activated complexes studied in our systems is transition state theory. This approach establishes that reactants and activated complex (or transition state complex) are in a quasi-equilibrium condition. Transition state theory analysis assume that the rate (k') by which the product C is produced is directly proportional to the concentration of the activated complex, [AB']:

$$A + B \rightarrow C \tag{1.4.1}$$

$$\frac{d[C]}{dt} = k[A][B] \tag{1.4.2}$$

$$\frac{d[C]}{dt} = k'[AB'] \tag{1.4.3}$$

If we considered a quasi-equilibrium state between reactants and the activated complex, and combining equations 1.4.2 and 1.4.3, we obtain that:

$$[AB'] = K'[A][B] \tag{1.4.4}$$

where K' is the equilibrium constant between the reactants and the activated complex. If we combine equations 1.4.2, 1.4.3 and 1.4.4, we obtain that the global equilibrium constant k from the reactants to the product is:

$$k = k'K' \tag{1.4.5}$$

As we have seen above, activated complexes depends of the vibrational modes of the atoms. This vibration, which governs the transition to the products, is related to the rate constant k' and the frequency of vibration, ν, but does not only depends on vibration modes. The arrangement of the atoms in the activated complex and the rotation of the molecule make that some vibrational modes are not favorable for the pass to the products. This information is given by the transmission coefficient, κ (see equation 1.4.6). Because activated complex should be in equilibrium with reactants, transition state theory considers these vibrational modes and atom rearrangements close to one, being the rate of passage proportional to these two elements:

$$k' = \nu\kappa \tag{1.4.6}$$

With respect to the equilibrium constant K', because the short lifetime of the activated complex, we are unable to have a Boltzmann distribution of states, so it is necessary to perform a statistical mechanics analysis. According to this, K' will be proportional to a new equilibrium constant, $K^{\ddagger}$:

$$K' = \left(\frac{K_B T}{h\nu}\right) K^{\ddagger} \tag{1.4.7}$$

If we return to equations 1.4.5 and 1.4.6, and substitute the terms, we obtain the Eyring equation (equation 1.4.8). This equation only depends on the proper disposition of the atoms in order to pass from activated complex to products (transmission coefficient), the Boltzmann constant, the temperature in the system and Plank's constant.

$$k = \kappa \left(\frac{K_B T}{h} \right) K^{\ddagger} \tag{1.4.8}$$

1.5. Thesis objective

The main objective of this thesis is centered on the implementation of a new theoretical method that permits us to study chemical reactions based on the combination of existing BMD tools and QM theory. To test our approximation, we focus well known systems that are included in NTPases family. Specifically, and because of the limitation of QM theory level, our study has focused on the how catalytic water molecule is activated during nucleotide hydrolysis. Thanks to the flexibility of the method provided by SMD and US, all different hypothesis proposed about which is the base of the reaction can be studied. Energy decomposition provided by QM/MM interface allow us to estimate not only the energy of the reaction (H_{QM}) but also the influence of protein environment ($H_{QM/MM}$). On the other hand, the elevated number of structures generated and the overlapping among them, provide statistical strength in spite of the QM theory employed in the current studies. Another important fact is that the method presented can be combined with different QM theory levels, so new methods can be applied.

A second objective in this thesis consist on the description of the chemical reaction performed in the biological systems used to test the method. Although several experiments reveal the effect of specific residues in the activity of a protein, sometimes, the concrete function cannot be well defined. Here we study, using the method described, the possible functions and effects of the residues that have been proposed as base of the reaction in Ras, F_oF_1 ATPase, and myosin systems, as well as the influence water molecules in catalytic site, which absence is impossible to check current molecular biology techniques.

Finally, a third objective, consist of the study of FtsZ system, in order to obtain a complete model of the polymerization/depolymerization event. We propose a computational model of different length filaments in order to understand the differences of catalytic sites in accordance with their positions in the filament, trying compare the results with experimental ones.

2. Materials and methods

In this chapter, the principal methodological aspects that have been followed during the work for this thesis are explained. The first part is an overview of how force fields define atomic interactions, and which are the simulation conditions in order to reproduce experimental ones. We will see how initial data are obtained, the pipeline followed in the data processing for the protein structures, the conditions we used to perform MD analysis and how we stability is evaluated. Finally, we will introduce the principles of BMD techniques, and how the can be used for the study of chemical reactions.

In a second part of this section, we provide an introduction to QM/MM theory and how energy surface maps, based on it, are produced in order to extract the information explained above and determine reaction paths. In order to validate the method exposed, we will study the ionization of a water molecule case. Finally, we will exposed the possible limitations of the method.

2.1. Molecular dynamics

When we perform an analysis of protein structure using MD, every single data set must be analyzed individually. This includes information regarding the type of molecule analyzed, the experimental conditions that have been used to determine the structure and whether there are missing residues and atoms in it. However, it is also necessary to consider a series of common steps that should be performed for every structure analysis. One of these common steps we include is the analysis of the missing atoms and residues and how they are replaced. Such residues can be modeled based on templates (homologous proteins) or replaced by geometrical restraints, where distances and torsions can maintain the geometry of a missing region. However this latter option should only be adopted when the region does not affect the region of the structure studied.

Another important point is the force field that is going to be used and that describes the interactions between the atoms in the system. A large range of different force fields can be found in the literature, where every one of them presents different properties that cover structure stabilization or charge calculations. Solvation conditions, ion concentration, pH value, the ionization state of the residues or the presence of disulphide bonds are other steps that have to be considered when a biological structure is analyzed. These aspects allow us to perform productive MD that will permit us to obtain a stabilization of the system in accordance with the force field chosen and close to experimental conditions. An evaluation of the stability is necessary in other to perform additional studies of the structure.

All the studies that are presented here were performed using the MD package AMBER (versions 10 and 11)[49], one of the most commonly used MD programs in the scientific literature. The analysis were performed using AmberTools, versions 10 and 11. AMBER has been widely used in our laboratory for many years, proving our team with considerable experience in MD and BMD analysis[8,50–52], and more recently, in QM/MM[53,54]. Different MD packages can be found (CHARMM[55], GROMACS[56] or NAMD[57]), that permit similar analysis of the structures to be performed. We will see how MD stabilization is achieved from obtaining the original data, and how under an established structure, reaction pathways

are studied with energy surface maps using QM/MM theory in combination with BMD techniques.

2.1.1. Determination of the water environment

One aspect that we have to consider when working with biomolecular systems is that, in most cases, the system is going to be solvated with water. MD is able to represent the water environment that enclose the macromolecular systems, in order to reproduce and maintain more accurate cell conditions. There are two principal ways of represent the water environment:

- Implicit water molecules: where the electrostatic embedding is calculated with the Poisson-Boltzmann equation. The aqueous environment is represented as a continuum, without taking the water molecules for the calculation into account, so the system will be accelerated. This is caused because there are not collisions between the system and the water molecules, so no friction exists. This kind of approximation is useful when we consider studying wide conformational changes that cannot be studied using long MD with an explicit solvent. The combination of an implicit solvent with BMD techniques, such as SMD, allows us to study a wider conformational space. A good example of this combination can be found in the study of the F protein of the paramyxovirus performed by our group as complementary work to this thesis[58]. A problem associated with this kind of representation is that the dielectric constant calculation is based on the accessible surface, which implies that hidden surfaces are not taken into account for the calculation, but which may be solvated. These hidden surfaces include the catalytic sites of enzymes, where the presence of water molecules could be essential, as we will see. An additional problem is that for large complex system, the solvation surface is very large, which means that the computational cost would increase.

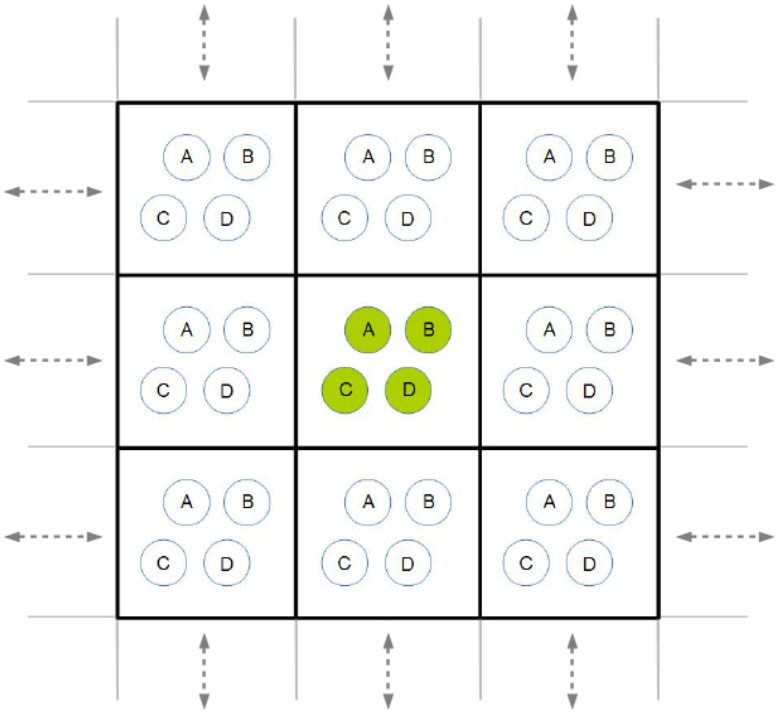

Figure 2.1: Periodic boundary conditions. Green spheres would represent the original system, composed in this case for four monomers, surrounded by a square water box. Rest of the boxes would represent the solvent of the original box, using virtual waters. Original system would be imaged in all the copies that surround the original. It is important to consider the system that is going to be used according to its diffusion. In long simulations, it is common to observe the system out of the original. While globular systems could rotate and translate without contact with their own images, rod-shaped systems (like FtsZ filaments) needs large water boxes in order to avoid that contacts.

– Explicit water molecules: this method uses explicit water molecules as the aqueous environment; that is, the water system is included in the calculation of the MD algorithm. From the computational point of view, this method is more expensive than the implicit one, because the size of the system increases considerably with small variation of water box size. However, this method represent better the conditions of solvation because takes into account the interactions and collisions of the proteic system with water molecules. There are different geometries to solvate the macromolecular system (octahedron, dodecahedron, cubic, triclinic). In order to maintain a continuous environment, it can be established what it has been called

periodic boundary conditions or PBC. A scheme can be found in figure 2.1. This method generates different images of the original system (macromolecule and environment), so the molecules that are placed in the boundaries of the box can sense the molecules of the opposite edge of the box. In this way, the system will be solvated constantly by real or imaginary waters. Another advantage of PBC is that avoids the evaporation of water molecules. Interactions will be calculated between real atoms or between real atoms and imaginary atoms, but never between imaginary ones.

The method chosen here is the explicit one. Several such models can be found in the Isothermal–Isobaric ensemblescientific literature[59,60]. The model used here is the TIP3P model. Although there are several water models that could represent the water environment better from the point of view of density and diffusion, TIP3P is a well-accepted model that is widely used in MD studies and that allows a fast calculation of the solvation environment.

2.1.2. Ensembles

Pressure, volume, temperature and energy system are some of the parameters that have to be controlled by MD and depending on which ones are controlled, we find different kinds of ensembles. In MD an ensemble is considered to be the group of different systems that conforms our phase space, that is, real system and the copies that constitute the lattice in periodic boundary conditions. Principally, we find three different ensembles:

- Microcanonical ensemble (NVE): total energy, volume and number of atoms that comprise the system are considered constant, so no variations are observed in these variables. The system is involved in an adiabatic process where no heat transfer takes place.
- Canonical ensemble (NVT): the ensemble of systems can exchange energy with an external bath, which maintain the temperature of the ensemble. Temperature is maintained using a weak coupling algorithm, that maintain an appropriate total

energy according to the temperature selected. The distribution of atomic collisions tends to ensure an even distribution of the temperature, but only with an explicit solvent, where collisions with the environment avoid acceleration of the protein atoms.

– Isothermal–Isobaric ensemble (NTP): pressure and temperature are maintained constant with variations in the volume and the exchange of particles. This ensemble is the closest to laboratory conditions. This method is not adequate to membrane simulations, where the pressure conditions are not the same along all the axes (anisotropic conditions). All the calculations presented in this thesis were performed using an NTP ensemble.

2.1.3. Force fields

Since several years ago, MD approximations have treated atoms as individual particles with mass, whose movement is resolved by the forces of attraction and repulsion generated by interaction with the other atoms in the system. The energy of the system will depend on the relative positions of every single atom and therefore, it is considered to be potential energy. The total potential energy is the sum of the potential energies of the individual atoms. A simple way to consider this potential energy is as the sum of the energies associated with the covalent and non-covalent bonds between atoms:

$$V_{total} = V_{covalent} + V_{noncovalent} \tag{2.1.1}$$

Covalent interactions are normally defined by equations that represent harmonic potentials. In this way, distances (between contiguous atoms), angles and dihedrals are expressed using the following kinds of expressions:

$$V_{bond} = \sum k_b \left(b - b_0\right)^2 \tag{2.1.2}$$

$$V_{angle} = \sum k_\theta \left(\theta - \theta_0\right)^2 \tag{2.1.3}$$

$$V_{angle} = \sum k_{\phi}\left(1 + cos\left(n\phi\right)\right)^2 \tag{2.1.4}$$

where b_0, θ_0 and Φ_0 are the reference values for distance, angle and dihedral. K_b, K_θ and K_Φ are the force constants that maintain the value of the potential close to the reference value. Strong constants will penalize fluctuations with an increase in the total energy. The fourth types of terms are those that correspond to non-bonded interactions, that is, Coulombic and Van der Waals terms:

$$V_{nonbonded} = \sum\left(\frac{A}{r^{12}}\right) + \sum\left(\frac{C}{r^6}\right) + \sum\left(\frac{\left(q_1 q_2\right)}{\left(\Delta r\right)}\right) \tag{2.1.5}$$

In accordance with all these terms, the total energy of the system will be the result of the sum of all the terms:

$$\begin{aligned} E_{total} = &\sum k_b\left(b - b_0\right)^2 + \sum k_\theta\left(\theta - \theta_0\right)^2 + \sum k_\phi\left(1 + \cos\left(n\phi\right)\right)^2 \\ &+ \sum\left(\frac{A}{r^{12}}\right) + \sum\left(\frac{C}{r^6}\right) + \sum\left(q_1 \frac{q_2}{\left(\Delta r\right)}\right) \end{aligned} \tag{2.1.6}$$

This is the most simple form of expressing the total energy function of a system, which is able to reproduce the fundamental characteristics of protein energy surfaces with atomic resolution. The combination of total energy and the different parameters that have an influence over it, is what has been called the force field. Data sets containing these parameters are obtained from experimental procedures (such as NMR) or high-level quantum mechanical calculations.

Many force fields can be found in the literature, depending not only of the MD implementation used (although they can be used independently of the program) but also on a lot of decisions concerning which data are going to be emphasized in order to obtain a better fit, how areas out of the fit region are handled, and also the computational performance in the calculations[61]. Therefore we will find force fields that are better adapted to specific secondary structures (beta sheets or alpha helices), DNA or polarizable side chains.

In the case of the computational studies performed in this thesis, the force field chosen was ff99SB, which is widely used in the literature[62] and is the basis for more modern force fields that can be found in the latest versions of MD packages, like ff10, ff12, ff99sb-ILDN and ff99SB*-ILDN.

2.1.4. Data sources

The initial point of every single study performed for this thesis was to obtain of the structure on which the study was performed. Mainly, starter data are contained in a PDB file, that is, a structure stored in the Protein Data Base[63] (www.pdb.org), where a collection of thousands of structures that has been solved mainly using x-ray crystallography or Nuclear Magnetic Resonance (NMR) can be found. Every single file contains information on how that structure has been resolved, the resolution of the structure (the confidence in the atoms coordinates), if there are some residues or atoms that could not be resolved (and that will require some type of modeling treatment, as pointed out above), and the coordinates of the atoms that from the system.

Another way to obtain the structure of the system is via molecular modeling of the structure, based on the amino acids sequence of the protein of interest and the structure of a homologous protein.

In most cases, structures obtained from Protein Data Bank have some atoms or missing residues. Two alternatives can be applied at this point. If we talk about missing atoms, they can be added to the pdb file if the residue is known, but always taking into account that there could be some overlap and collisions with the atoms of the nearest residues. In the analyzed structures here, we based on the different rotamers libraries of PyMol when a collision was found after the addition of the missing atoms.

When the missing part is a consecutive sequence of amino acids (mainly loops and highly fluctuating regions), there are two options. First is the use of protections at the ends of each missing sequence. These protections consist of two molecules that form a peptide bond with the end amino acids, adding a methyl group ($-CH_3$) to each end, so the charges are neutralized and not additional ones are added. At this point, and with the objective of

maintaining the molecular structure in that region, additional restrains (distance, angle and dihedral types) should be added. The other option is to model the missing region with either a homologous protein or the result of accessing the different databases and servers where the loop can be model using the amino acid sequence.

In proteins, there are some residues that need to be revised individually. These are histidines and cysteines. In the case of histidines, we need to look at the protonation state of every single one. This will depend on the pH of the environment and the neighborings residues around the histidines. At this point, we find three different protonation states: Nε, Nδ or double protonantion. The decision as to which state is chosen is based on the results given by H^{++} server[64]. For cysteines, it is necessary to check in the structure the position of every cysteine in order to check whether they can disulfide bond, forming a covalent bond between them. Disulfide bond construction is performed using the AmberTools program LeaP. Because structure added to LeaP present all residues with hydrogen atoms according to the pH value, it is necessary to remove the hydrogen atoms of cysteines, thereby obtaining a oxidated state of the residue that permits bond formation. We should change the protonation state of the two residues that comprise the bond by removing the proton bonded to the sulfur atom. If no disulfide-bond is formed, we may experience a change in the tertiary structure of the system.

Proteins are usually resolved with attached cofactors and ligands. In the cases of enzymes, they may be crystallized in the presence of compounds that present an analogous structure to the ligand or the ligand in the transition state form, that is, the highest value of energy in the reaction pathway. These transition states present the peculiarity that they are not stable because at this point the reaction progresses quickly towards the products. Some transition states are stables because of the protein function. In this case they are called intermediaries of the reaction, presenting a longer life than transition state structures.

These analogs are molecules that block the protein in a conformation that approximates that of the transition state. If we focus on NTPase proteins, the analogs of the transition state are usually ligands that reproduce the trigonal bipyramid that the nucleotide form during the transition state. There are several compounds, some of them are AlF_4 or VO_4 derivatives. If

we want to study the reaction with the proper ligand, we must replace the analog atoms for nucleotide ones. It is usually the atom that occupies the position of the γ-phosphate. It could be possible that we can't find these ligands in the force fields, so we may need to parameterize it in order to include them in the calculations.

2.1.5. Topology and parameters

Once the structure has been checked and the proper initial conditions have been obtained, the next step consists of generating the input files for the MD simulation. Although every single MD package has its own file formats, all of them need two initial data: the topology of the molecule and the initial coordinates of the atoms. The term topology refers to the order of the atoms inside the molecule. That is, which is the initial atom, which are the adjacent atoms, whether there are ramifications of the main chain (side chains, chemical modifications, etc) and where the main chain ends. In this way, the bond type between two atoms, or the angle and dihedral values between the atoms are also established. Furthermore, the input file refers to the atom types, which includes the atomic properties (charges, Van der Waals radius, etc). The coordinates of the atoms are the second input parameter; a single file where the coordinates are stored in the same order as it was found in the topology file.

Additional information that can be found in these files will depend on the experimental conditions:

- System solvation: although different solvent can be used, in the case of proteins and concretely in this thesis, the systems were solvated in an explicit water box system (see above). Because periodic boundary conditions are used in simulations, small boxes can be used, but taking into account translations and rotations of the proteins, in order to avoid contacts with the image of the molecule. Information about the box size and position are saved in the coordinate files, with the positions of the water molecules.
- System charge: because the Particle Mesh Ewald method for computing

electrostatics works most reliably when there is no net charge on the system, one of the steps is to neutralize the system. Depending on the total charge, positive or negative ions are used. The normal process consists of substituting some molecules in the solvent with those positive or negatives ions. Additionally, salt concentration can be controlled with total number of ions.

2.1.6. Minimization phase

Once the system has been prepared and experimental conditions have been established, the next step is to minimize the structure. From a computational point of view, the minimization process consists of finding the minimum value for the potential energy function for an ensemble of independent variables. In the case of proteins, these variables correspond to the number of atoms (N) in a given space (3N). Because of the high number of atoms in a protein (a small system can be composed of more than 10-20 atoms), it is very difficult to obtain the minimum energy value of the function. In the case of macromolecules, the accessible conformational space explored by minimization will be very small. As a consequence, only small rearrangements will be produced. In this way, small collisions and the worst steric problems between the side chains of amino acids or bad geometrical values because a ligand substitution (and also in the protein structure) will be solved by minimization.

The idea is that minimization takes the initial coordinates of the atoms that form the protein and scans the space defined with small increments in each coordinate, so that for every increment, the potential energy can be estimated and evaluated in accordance with the previous step. One of the limitations of the method is that, because the conformational space explored is small, many different accessible minimums will never be reached. In such cases, only MD can explore other minimums. During the experiments performed for this thesis, minimization was performed using a combination of two methods:

- *Steepest descent method*: We consider any continuous and differentiable function that can be expanded as a Taylor series. The first derivative is treated as the slope of the potential surface, that is, the force. In the steepest descent method, the direction

of the shift will be parallel to the net force of the potential surface. In other words, the algorithm will search for the direction in which the minimum slope is encountered. The method takes the initial coordinates of the atoms and selects a direction with a chosen increment (see 2.1.7 formula). For each step, if the calculated potential value is lower to the previous step, another iteration is performed. The step size will decrease in the moment the potential value will be higher than in the previous step. One problem with this method is that if two consecutive steps have the same value, the algorithm stops without reaching the minimum of the potential. This method searches for a local minimum, without taking into account small minimums, what provoke that the structure reached will not differ greatly from the initial structure.

$$x_k = x_{k-1} + \lambda_k s_k \tag{2.1.7}$$

- *Conjugate gradient*: This method takes the gradient information of former steps. This information will be applied in the gradient calculation of the following steps. That is, while the steepest descent method always selects the minimum slope, the conjugate gradient method varies the direction considering the previous steps.

In the case of our minimization protocol, all the structures are minimized over 10000 conjugate gradient steps, being the first four cycles of steepest descent at the start of the minimization run (obtaining a fastest minimization protocol) and thereafter every non-bonded pairlist update, following the AMBER protocol.

2.1.7. Equilibration, stabilization phase and productive MD

Once bond distances and large steric problems have been solved by the minimization steps, next phase consists of adapting the system to the experimental conditions in our case, that is, to the canonical ensemble. An adaptation to the selected force field is done too. During the minimization phase, the working temperature is close to 0 K because only potential values have been calculated, so the temperature must then be established at room temperature (298 K), in order to obtain the thermal agitation in the system. This

temperature increase is achieved by increasing the temperature with a δt of 10 K per picosecond. During this phase the Cα trace is maintained with a potential restriction on the dihedral formed by the atoms, taking as reference values those obtained after the minimization phase. This avoids any possible changes in the tertiary structure during the heating process. Additionally, and depending on the system, there are some cases where some distances must be maintained during the heating phase. These distances are usually located in the active site of the protein, where the nucleotides are bound in the NTPases case. These distances involve distances between the nucleotide and waters or side chains bounded to the ligand. This is done in order to maintain the geometrical structure resolved in the x-ray or NMR structures. In the remaining rest of the protein, no such restrictions are applied to the side chains of the amino acids, ions or solvent. In this phase (and if the canonical ensemble has been chosen), the pressure will also be established as atmosphere conditions.

The stabilization phase consists of adapting the structure to the ensemble without any restrictions. For 15 picoseconds, all the restrictions are removed gradually decreasing the constant force of the potential, so, that eventually, no constant force is applied. As mentioned above, depending on the system, some restrictions are maintained. This is the case of transition states, where the structure of the ligand must be maintained. Without these restrictions, the ligand would revert to the products.

Once the restrictions have been removed, what is called productive dynamic starts: a long dynamic in experimental conditions (native conditions), during which the stability of the system is monitored. The simulation time varies depending on what is studied and it is also limited by the size of the system and the computational power avalilable. It takes several weeks to obtain simulations in the range of nanoseconds for systems of hundred of thousand atoms that include explicit waters (although nowadays, GPU computing is solving this limitation, with a high performance in nanoseconds per day, allowing to reach microsecond simulations, which is on the order of conformational changes in real proteins).

There are some parameters that allow us to decide whether the protein is stable or not:

- ***Total Energy*:** This is a good indicator of the stability of the system consist in the monitoring of the total energy of the system (kinetic and potential), observing the lack of fluctuations along the trajectory and the stabilization around a proper value. Energy fluctuations can often be attributed to structural changes.
- ***Root Mean Square Deviation (RMSD)*:** This measure provide the structural difference between the selected atoms of two proteins or states. It respond to the following expression:

$$RMS = \langle (r_i^a - r_i^b)^2 \rangle^{\frac{1}{2}} = \sqrt{\frac{1}{N_i} \sum (r_i^a - r_i^b)^2} \qquad (2.1.8)$$

  A good RMSD value will be under crystal resolution, reducing the increment and fluctuation in time.
- ***Root Mean Square Fluctuation (RMSF)***: Indicates us the mean fluctuations of the atoms selected respect to a reference structures, that corresponds to an average structure of the trajectory.
- ***Radius of Gyration:*** It indicates the variation of the distance between the atoms of the system an its center of mass.
- ***Principal Component Analysis and Cosine Content:*** Data obtained from the covariance matrix of a protein, usually calculated over the Cα trace. Describes if the conformational changes along a trajectory in a protein is because of random diffusion of the structure (value close to one) or because of inner changes in the protein structure (close to zero).

The normal simulation time step is established as 1 fs. When a fastest step is performed (around 2 fs / computational step), we would find a problem associated to the energy of the protons. This is because the potential energies that define the distances with hydrogen atoms are very narrowly defined, so variations in the positions with a bigger step provokes higher energy values, which causes a readjustment in the atoms of the system. In order to avoid this problem, the SHAKE algorithm is applied. In general terms, this algorithm

doesn't take the hydrogen atoms into account in the calculation of the new positions of atoms and therefore, no velocity is applied to them. After every step, hydrogen position is established according to the position of the linked atom.

2.1.8. Steered molecular dynamics (SMD) and umbrella sampling (US)

MD usually works on a nanosecond time scale, reaching microseconds only when small systems are studied[65]. This limitation on the simulation time means that some structural changes that involve a reorganization in the secondary or tertiary structure are impossible to observe during a productive MD, because they are usually produced in a microsecond or millisecond range, accesing to a larger conformational space.

SMD is a computational approximation that permits us to study the inaccessible conformational space of a productive MD. A comparison of this behavior can be found in biotin-avidin experiments. Biotin is so strongly bound to avidin that spontaneous unbinding via thermal events only happens at timescales of weeks while with atomic force microscopy, the event can be studied on a timescale of microseconds to seconds. The SMD technique is included in the BMD techniques. The principles of SMD[66,67] are very similar to those applied in the atomic force microscopy[68], optical tweezers[69] or biomembrane force probe[70].

Two SMD modes can be used. The first consist of increasing the distance with respect to an initial coordinate, x_0, using a constant velocity. The displacement of the molecule, atom, or spatial coordinate we are using to pull will be constrained by an external potential that will present the expression in equation 2.1.9:

$$U = K\frac{(x-x_0)^2}{2} \qquad (2.1.9)$$

where K is the stiffness of the harmonic potential used to pull. Therefore, the force used will correspond to the following expression:

$$F = K\left(x_0 + vt - x\right) \quad (2.1.10)$$

The second SMD working mode consists of applying a constant force during the pulling. In this case, the stiffness of the potential will vary proportionally to the time (K = αt). In this case, the force will correspond to the following expression:

$$F = \alpha t\left(x_0 - x\right) \quad (2.1.11)$$

These principles, that as we have seen are aimed at the study of protein unfolding, ligand unbinding from a protein or protein-proteins/membranes interactions, among others, can also be applied when we study chemical reactions, and they reveal the chemical processes that take part in an enzymatic reaction as we will see in the following sections.

It may be that in the ensemble of structures represented in a MD trajectory, some of the structures present a low density. In the same way, during a SMD trajectory, some structures can be skipped because we are using a high velocity or because some energy transitions can be ignored. In order to increment the density of our sample and study the transitions defined by our coordinate, we use US. This method permits us to obtain a good sampling of our trajectory using small variations in our reaction coordinates. The small increment in distance will create overlapping between the sample windows (in our case, reaction trajectories), permitting us to obtain a good sampling of our defined conformational space. Additionally, the combination of SMD with US permits us to reduce the possible bias effect that we could generate with the former alone.

2.2. Energy surface landscapes using the QM/MM interface

The study of chemical reactions requires the formation and the rupture of chemical bonds. Because MD establishes bond distances between atoms using harmonic potentials in order to maintain the values obtained in experimental results, it is impossible to break bonds. There are a kind of potentials, the Morse potentials, that can define the energy necessary

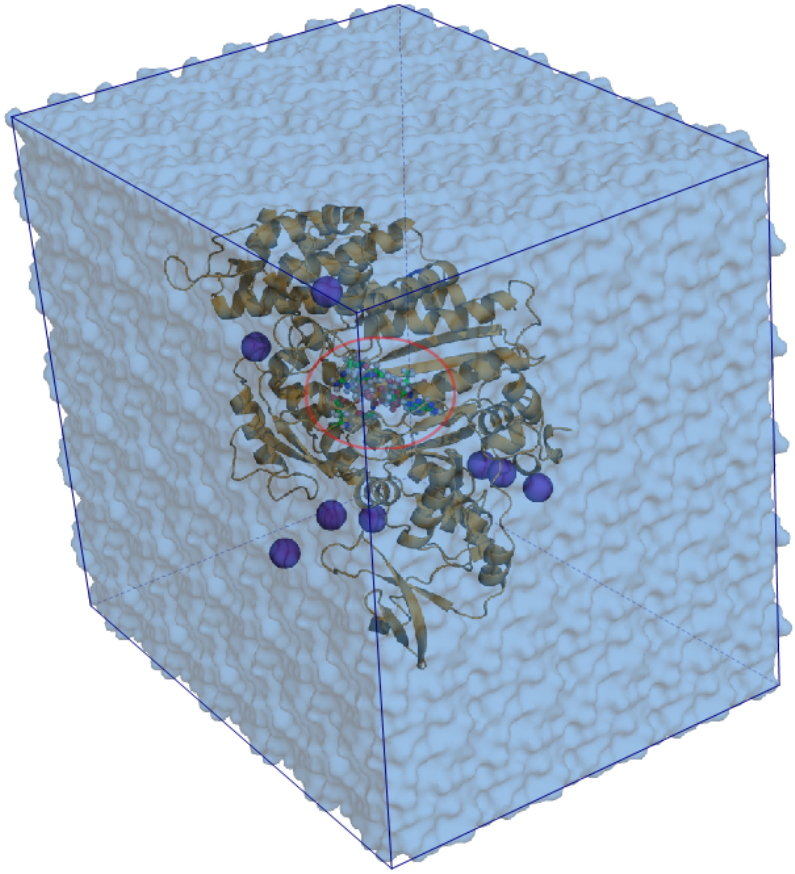

Figure 2.2: Representation of complete solvated protein system. Myosin protein (PDB: 1VOM; orange cartoon) is solvated by a cubic water box (blue box) that presents a size of 12 Å from the protein, presenting PBC conditions (limits are indicated with continuous and dashed blue lines). Potassium ions are represented in purple spheres. During productive MD, all the system is handle by MD equations, calculating the total energy of the system. Once stabilization is reached, it is possible to perform a QM/MD stabilization, where all those atoms that take part in the reaction will be included in H_{QM} term. In this case, QM atoms are represented in ball and sticks. The QM/MD interface would be created between these atoms and the rest of the system, here indicated by the red circle.

for the formation and breaking of a bond, since an activation barrier can be defined in a potential equation. However, this approximation presents several limitations, considering that no reorganization of charges or atomic orbitals are performed when using this kind of potential. Additionally, the assumption of an activation barrier for every bond distance would prevent the study of the activation barrier in the reaction since it would be defined *a priori*.

Therefore, the most suitable approximation to this problem is through the use of QM approximations that allows us to study the catalytic processes inside the protein. The whole system cannot be included in the QM calculations, since every orbital reorganization should be calculated between all the atoms in the system, which would be a severe limitation from the computational point of view. A reasonable number of atoms that performs well in terms of calculation time is close to 150 atoms, compared to the thousands of atoms that forms a solvated system. To resolve this problem, we use QM/MM, or specifically, in our case, the QM/MD implementation. This approximation permits us to define our system in two delimited regions. The first is what has been called QM part, which includes all those atoms that play a direct role in or have direct influence in the catalysis of the reaction. Focusing on the study of nucleotide hydrolysis, this includes the atoms close to the phosphate part of the nucleotide and those residues close to the catalytic water. The second part is formed by the rest of the system and correspond to the MD calculation and that is governed by the classical movement equations as we have seen in the section 2.1.3.

This approximations means that an additional term must be incorporated into the total energy calculation, which, as we have previously seen, is composed of the bonding and non-bonding terms of classical force fields. The distribution of the system into QM and MD parts generates a QM/MD interface, where the interactions between atomic nuclei and electron clouds in the QM part and point charges of the MD atoms in the MD part should be taken into account as an additional term of the total energy in the system, as we see in equation 2.2.1:

$$E_{eff} = \langle \Psi | H_{QM} + H_{QM/MM} | \Psi \rangle + E_{MM} \quad (2.2.1)$$

This $\boldsymbol{H_{QM/MD}}$ energy term can be decomposed in three terms:

$$H_{QM/MM} = \sum_q \sum_M [q_m h_{electron}(x_e, x_{MM}) + z_q q_m h_{core}(x_{QM}, x_{MM}) + (\frac{A}{r_{qm}^{12}} - \frac{B}{r_{qm}^{6}})] \quad (2.2.2)$$

where $\boldsymbol{q_m h_{electron}(x_e, x_{MM})}$ corresponds to the interaction between the electron clouds of the QM atoms and the point charges of the MM part; $\boldsymbol{z_q q_m h_{core}(x_{QM}, x_{MM})}$ is the term that represents the electrostatic interaction between the point charges of the MM atoms and the cores of the QM atoms; and the last term is the Lennard-Jones interactions between the QM and MM atoms.

The definition of our system considering both QM and MD parts, permits us to simulate enzymatic reactions, taking the whole protein environment into account, as governed by MD equations. These are different from other approximations, where the protein environment is minimized with respect to specific structures[71–73]. These kind of approximations permits the use of *ab initio* methods in the calculation of the QM energy, which makes them more accurate in the energy values calculations. This calculation of the energy and structural reaction pathway is defined by local minimizations of initial and final state, defining the activation barrier by the minimization of the considered transition state, what could lead to an error since an ensemble of different reaction pathways pass through transition state. The use of MD would permit to obtain a wider study of atomic rearrangements in the MD part in accordance with the QM transitions. In this way, and as we will see, in our study, the transition state will be compound by an ensemble of structures. The energy value of the transition state will be given by the values of these individual structures.

In this way, we have developed a theoretical approximation that combines the QM/MD interface with the MD techniques of SMD and US. The idea is to study the different reaction pathways that are accessible considering the two reaction coordinates chosen. As a result, we obtain an energy surface landscape that reveals the most favorable pathways

according to the selected coordinates. As we will see, this kind of maps provides information regarding the routes that a chemical reaction could follow according to the reaction coordinates and additionally, they permit us to compare activation barriers in order to compare different routes, evaluating the effect of the environment.

2.2.1. Energy surface maps

For the analysis of a reaction using the QM/MD approximation we need a starting structure that comes from a MD stabilization. For the analysis, we use the structure based on the data collected for the mean distance between the atoms that are to be included in the QM calculation. Another option would be consider an average structure of the last part of the stabilization, minimizing this structure.

Prior to the reaction pathway analysis, the structure obtained must be established under the QM/MD conditions. In this phase, the QM region is defined. In accordance with the nucleotide hydrolysis, the principal atoms selected belong to the ligand (in this case the nucleotide), the waters that could have an influence on the hydrolysis, and all those residues that are close to the ligand. This last step is usually based on experiments in the literature. During QM stabilization, bond distances in the QM region are maintained for around 200 picoseconds by restraints, which are removed them when quantum energy is observed to stabilize, by reducing the force constant during the next 50 picoseconds. The system will be established without any restraint during 50 additional picoseconds.

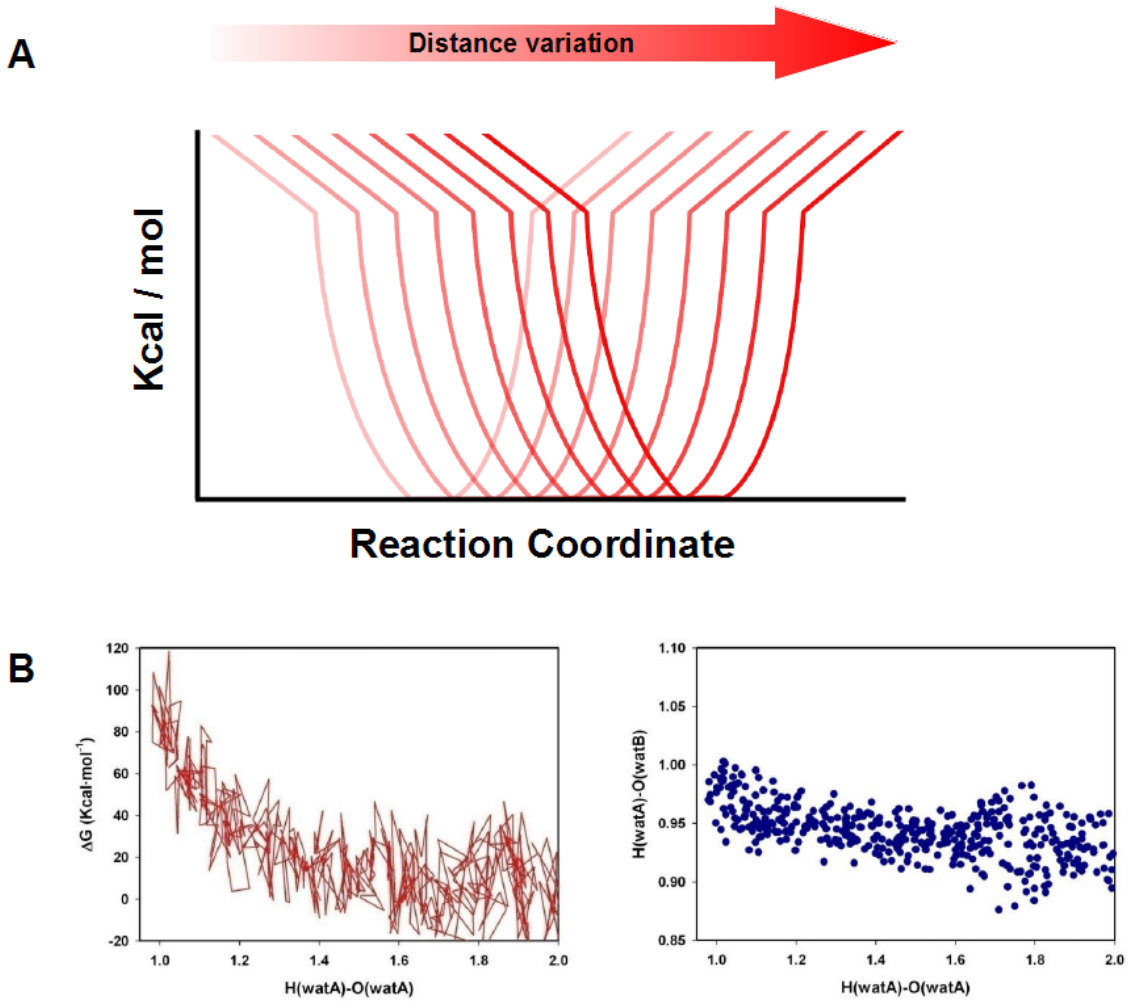

Figure 2.3: (A) Scheme of a reaction coordinate energy potential. In opposition to MD potential energy function that define bonding terms, reaction coordinates are defined with flat-bottomed potentials. These kind of potentials permit a higher sampling along the trajectory and between them. **(B) Example of a single trajectory.** Left panel represents the variation of energy according to one of the reaction coordinates. Checking of the energy (MD or QM energy) can be tested according to the reaction coordinates in order to find a correlation between energy fluctuations and changes in the structure. Right panel shows a plot of a single trajectory according to the two reaction coordinates. In this case, distance H(watA)-O(watB) was maintained fixed with a value of 0.975 Å. Distance H(watA)-O(watA) was incremented 1.050 Å, from 0.950 Å to 2.0 Å. Every single point corresponds to a different structure. A total of 10 structures are stored every 0.025 Å. In this case, a total of 420 different structures are plotted. Direct relation exists with energy graph, where every energy value of the structures is represented. As in the same case of right panel, left panel contain 420 different energy values, associated with a single structure.

Once it has been checked that the structure is stable (the criteria used are the same than in MD, but adding the QM energy of the QM part to the analysis), the next step consists of creating a reference trajectory. The reference trajectory contains an ensemble of structures that will be used as starting structures for the trajectories that will form the surface. This trajectory is created by maintaining one of the chosen coordinates fixed (the value will be taken from the stabilization) and varying the other using the SMD method. We name the coordinate that is held fixed A and the one that is varied B, in order to simplify the terminology. The variation in distance chosen is $\delta r = 0.025$ Å per picosecond. This value permits a good sampling of the structures and energies along the coordinate. Additionally, the small difference in distance allows an overlapping between the adjacent trajectories, applying the US procedure. Larger values could lead to empty regions in the phase space defined by the reaction coordinates. The distance is governed by a flat-bottom harmonic potential, which permits us to obtain this overlapping not only inside the trajectory but also

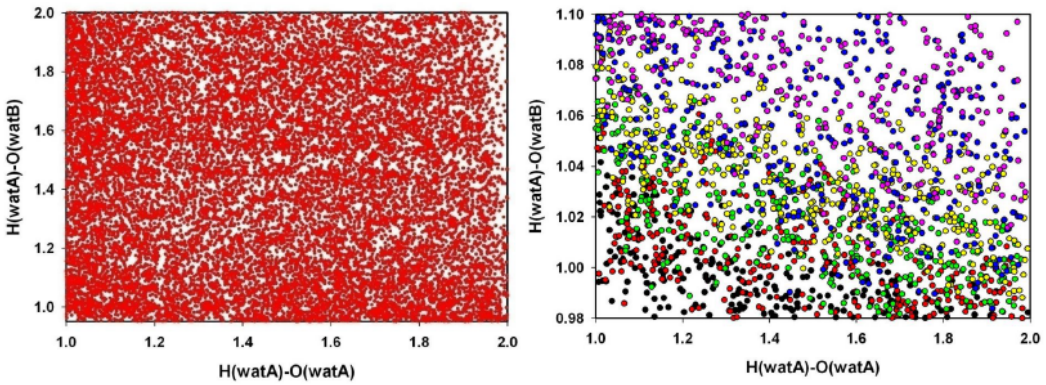

Figure 2.4: Representation of energy surface umbrella sampling. Left panel represent a complete set of trajectories with a distance range from 0.95 Å to 2 Å for both reaction coordinates. As in previous figure, every trajectory contains 10 structures that has been stored every 0.025 Å. In this case, a total of 176400 different structures are plotted in this map. Conformational space defined by the two reaction coordinates (in this case, a water ionization reaction) has been completely mapped. A clear area is observed on the top-right corner, corresponding to an inaccessible energy area as we can see in the figure 2.5. Right panel corresponds to a close view of left panel, ranging the area from 0.98 Å to 1.10 Å of the distance H(watA)-O(watB) and the whole range of the second reaction coordinate. Trajectories has been represented in different colors to differentiate them. As we can see, it exists a high overlap between trajectories (umbrella sampling).

in the adjacent trajectories. Once the starting structures have been generated, we generate the trajectories that define our conformational space. The number of trajectories is established by the increment chosen in the reference trajectory. For their construction, we now fix coordinate A with the value of the starting structure, varying coordinate B, using the same δr, governed by the same flat-bottom harmonic potential and the same velocity. The result is a grid of points that represent the different structures obtained by the variation of the coordinates (see figure 2.4). In this way, every single dot corresponds to an individual structure with an associated energy. Only the H_{QM} and $H_{QM/MD}$ energy terms will be considered, since these terms provide the energy given by the reaction atoms and the influence of the MD environment. The total energy of the MD part would mask the reaction energy since the effect of proteins is to accelerate the reaction rate, and therefore, a downhill surface is obtained.

The surface obtained cannot be analyzed directly because of the large fluctuation in the energy for the individual states. To obtain useful and meaningful information, it is necessary to smooth out the data. We use the LOESS method for this case. A sample result can be seen in figure 2.5. Using this method, we can compare different reaction pathways using different reaction coordinates.

2.2.2. Validation of the method: the activation of a water molecule.

In order to validate this method and check the accuracy of the values obtained, we study the simplest and well known mechanism: the activation of a water molecule using another one, that is, the ionization of two water molecules that results in a molecule of OH^- plus a molecule of H_3O^+. The result of this simulation show us the efficacy of the method.As we have explained, the system must be established in MD. The system consisted on a two water molecules that were solvated with a 12 Å water box. The system was established in MD conditions during 10 ns. Once the system was stable, the 6 atoms of the two water molecules were included in QM part, maintaining the rest of the system in MD calculations. QM stabilization was performed during 300 ps. The reaction coordinates used were the distance between the proton of one of the water molecules, called wat A and the oxygen

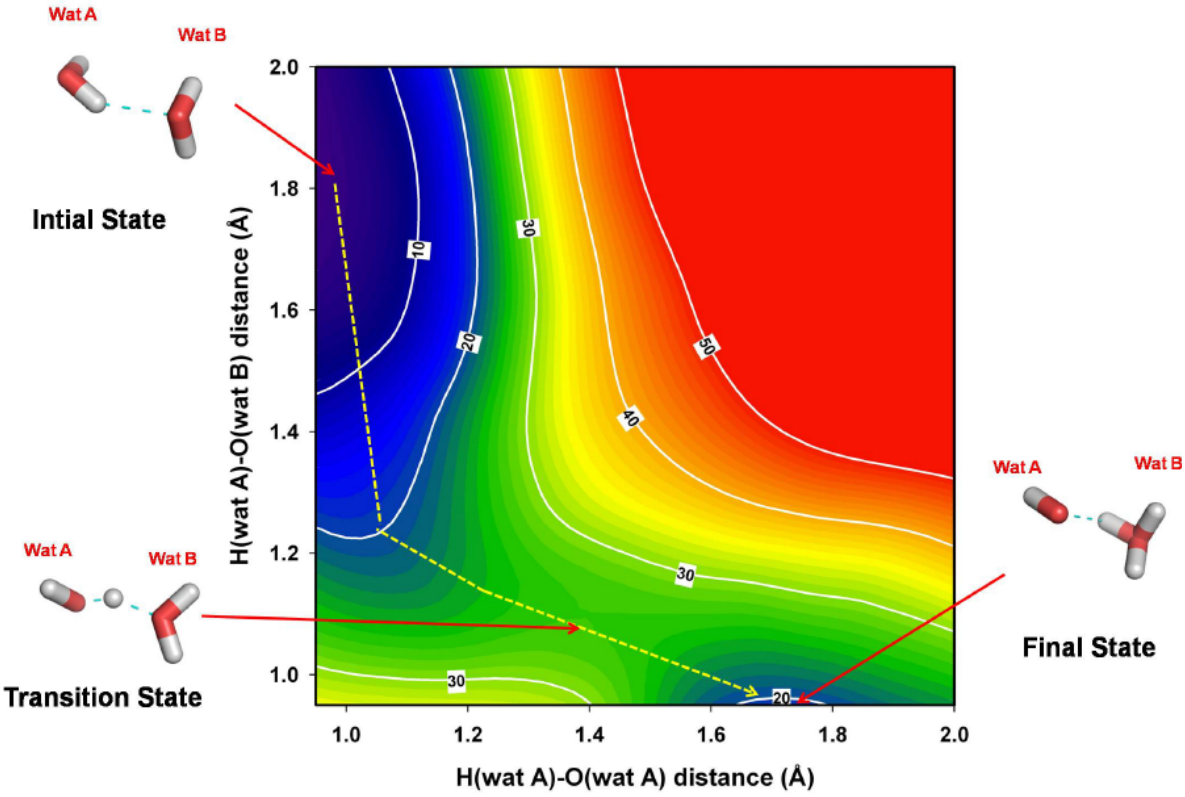

Figure 2.5: Free energy landscape for the ionization of water. G° values of the system are shown for the different states corresponding to the following distances: x–axis, from proton to oxygen atom in the attacking water molecule (wat A); y–axis, from the same proton to the oxygen atom in the second water molecule (wat B). The structures of the two water molecules in the initial state (ΔG° value adjusted to 0), the transition state (ΔG° ~ 28 kcal·mol^{-1}) and final state containing OH plus H_3O^+ molecules (ΔG° =19 kcal·mol^{-1}) are represented.

atom of the other water, referred as wat B molecule (bonding distance); and the distance between the proton involved of wat A and the oxygen of wat A (breaking distance).

The reference trajectory was made maintaining the breaking distance fixed and varying the bonding with a δr of 0.025 Å, from 1.8 Å to 0.95 Å. If we had not maintained the breaking distance, reaction would have produced. A total of 42 initial structures were obtained.They were used to construct 42 adjacent trajectories which have bonding distance restrained. The total number of different structures was 176400 with their own QM energy associated. The density of the conformational space studied can be seen in figure 2.4. The smoothed surface was calculated with LOESS. The free energy landscape obtained gave a ΔG^{o} of around 19 $kcal \cdot mol^{-1}$ between the initial and the final states, in complete agreement with the bulk water value ($\Delta G = -RT \ln K = -19.1$ $kcal \cdot mol^{-1}$ at 298.15 K and pH=7). This result allow us the check that the method was valid to the study of enzymatic reactions in proteins, as we will see in the section of the Results section below.

2.2.3. Limitations of the method

One of the critical steps found during the energy surface analysis was the election of level of theory applied in the QM calculations. *Ab initio* methods, as we explain above, have demonstrated to reach more accurate results since they can converge to the exact solution when solutions are small and the basis defined for the atom interactions can explain the global set of interactions. However these methods show a high demand of computer time and memory usage. This point is critical when we want to study enzymatic systems, where the number of atoms that take part in the reaction is around 100, as in this particular systems. The elevated number of trajectories and their length, makes unaffordable the study using the theory levels. An alternative is found in semi-empirical Hamiltonians, where atom interactions have been precalculated in a set of basis for specific systems, what accelerates the calculations. Because atomic interactions are precalculated, it is necessary the election of a method that include all the atoms in the hydrolysis reaction (C, H, O, N and P). The methods that were available at that point of the work for this thesis and that included those atoms were PM3 and AM1. The better geometrical optimization led us to the choice of PM3.

However, the election of these semi-empirical methods have associated one limitation: there is no information regarding the *d* orbital in the atoms. This point is critical for the global reaction, since there is a reorganization of this orbital respect to the phosphorus atom during transition state, where this atom present a pentacovalent coordination and in general, during the whole reaction. At this point we had to assume a transition state disposition of the ligand (the pentacovalent state with a trigonal bipyramid shape) in order to study the first step of the reaction, which includes the activation of the catalytic water molecule. In spite of this limitation, we could perform a study about the possible bases of the hydrolysis reaction. However, we can not make comparisons with global reactions.

This approximation with PM3 have proved to be highly accurate since the recent incorporation of a new DFT theory level into our methodology is currently reproducing the results in the study with the global reaction. The results obtained reveals us that the combination of QM/MD with SMD and US can provide a lot of information regarding how

an enzymatic reaction occurs. The development of new theory levels, based on DFT, will permit us to obtain a very detailed information regarding the global reaction, observing the particularities of every system.

3. Results

In this section we present the results obtained in the elaboration of this thesis, including MD dynamics studies and QM/MM analysis using the method described in the previous section.

A MD study about the division bacteria protein FtsZ is presented. Based on previous MD studies about the behavior of the catalytic interface in the dimer, i.e, the minimum catalytic unit, and on experimental data that propose a different behavior in polymer interfaces of FtsZ, a MD study of different length polymers were performed in order to characterize the behavior of active site according to its position in the polymer.

Three studies are exposed in order to validate the method explained above, including in the system the influence of a protein environment in a reaction. Concretely, the three systems studied were the HRas, ATPase and myosin proteins. All of them, as we have seen, are well known systems which mechanisms of hydrolysis has been studied for several years. Here we propose a study of the possible bases that permit that hydrolysis was performed in the catalytic center taking into account the protein environment using the QM/MM interface described above.

3.1. Molecular dynamics simulation of GTPase activity in polymers of the cell division protein FtsZ.

3.1.1. Abstract

FtsZ, the prokaryotic ortholog of tubulin, assembles into polymers in the bacterial division ring. The interfaces between monomers contain a GTP molecule, but the relationship between polymerization and GTPase activity is not unequivocally proven. A set of short FtsZ polymers were modelled and the formation of active GTPase structures was monitored using molecular dynamics. Only the interfaces nearest the polymer ends exhibited an adequate geometry for GTP hydrolysis. Simulated conversion of interfaces from close-to-end to internal position and viceversa resulted in their spontaneous rearrangement between active and inactive conformations. This predicted behaviour of FtsZ polymer ends was supported by *in vitro* experiments.

3.1.2. Introduction

FtsZ is the core protein in the process of bacterial division[11,74], being a main target for inhibitor molecules with promising use as future antimicrobials[75,76]. FtsZ assembles in the mid-cell into a large structure called the division-ring, which is formed of, among other proteins, polymers of FtsZ[19]. The structure of the monomer–monomer contacts in the FtsZ dimer, including the GTP site in the interface, was predicted (using 3D computer models) to be equivalent to its eukaryotic homologous structure, the tubulin α/β dimer[3]. This hypothesis was later confirmed by crystallography[77]. Several hypotheses have been posited to explain the role of GTP hydrolysis in the functionality of FtsZ[78–81]. A previous work using molecular dynamics (MD) simulations led to an explanation of the role of K^+ in the

active centre, as well as the prediction, experimentally corroborated, that GTPase activity is dissociated from polymer formation[82]. Recently, analyzing substrate kinetics of the GTPase activity of *E coli* FtsZ, it has been concluded that GTPase active sites in FtsZ are independent of each other[8]. In this work, using MD simulations of modelled short polymers of FtsZ, we have studied the active/inactive state of monomer–monomer interfaces. *In silico* results were then tested by measuring the FtsZ GTPase-specific activity *in vitro*.

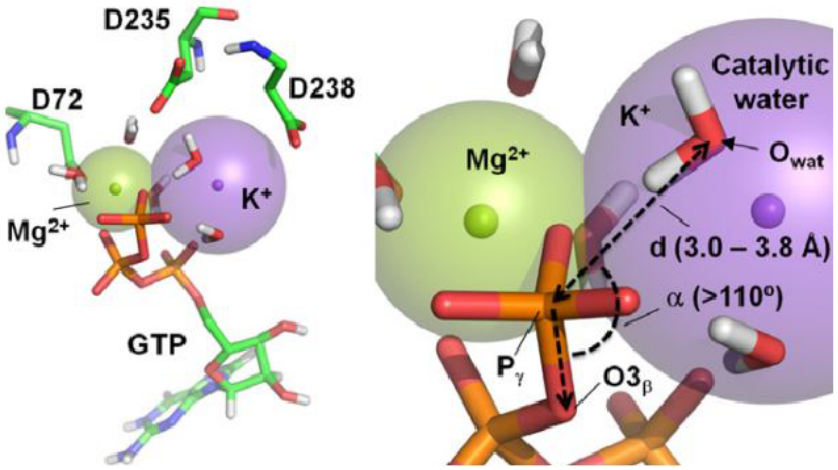

Figure 3.1: Schematic representation of an active FtsZ dimer interface. GTP molecule (sticks) is associated to a Mg^{++} atom (green sphere) coordinating three molecules of water. Two additional water molecules are located in the coordination sphere of the K^{+} atom (purple), including the catalytic water. Adequate "d" distance between O_{wat} and P_{γ} (3.0 Å < d < 3.8 Å) and minimum "α" angle among O_{wat}, P_{γ} and $O3_{\beta}$ ($\alpha > 110°$) necessaries for a correct catalytic activity are indicated.

3.1.3. Materials and methods

3D modelling and molecular dynamics

Three-dimensional models of FtsZ polymers were constructed by successive structural alignment of the two FtsZ subunits of the previously published MD-equilibrated dimer[82]. MD simulations were performed using the PMEMD module of the AMBER10 package[83]. Each modelled FtsZ polymer was surrounded by a rectilinear solvent box with a minimum

distance of 10 Å from the edge of the box to the closest atom of the solute, and with periodic boundary conditions, using LEAP. Adaptation to the AMBER force field was performed by 10,000 steps of energy minimization using a cut-off of 12 Å and a δt of 0.002 ps. During the initial heating phase (200 ps), the temperature was raised from 0 to 300 K, restraining the position of the Cα atoms with a force constant of 20 kcal·mol^{-1}, reducing the force constant in a stepwise fashion in the subsequent phase. After equilibration, unrestrained MD was performed for 8 ns in each case, relocating the hydrogen atoms using the SHAKE algorithm. The coordinates were saved for analysis of atom positions every 20 ps. Continuous tracking of the position of water molecules in the neighborhood of the interfaces was performed using *Ptraj-watershell*.

Measurement of GTPase activity

E. coli FtsZ was purified by the Ca2+-induced precipitation method[84]. Prior to use, protein was dialyzed in reaction buffer (5mM $MgCl_2$, 250mMKCl, 50mMTris pH 7.5) and then incubated at different concentrations with 1mM GTP at 22ºC. Reaction was stopped by dilution in 65mM EDTA in the same buffer, and phosphate concentration was determined by a colorimetric assay using the green malachite reagent[85,86]. Activity values were calculated by measuring the slope of the linear part of the activity curves.Results and discussion

3.1.4. Results and discussion

GTPase interfaces in FtsZ polymers.

To study the behavior of GTPase interfaces in FtsZ polymers, a set of three-dimensional models of short FtsZ polymers (trimer, tetramer and pentamer) was generated using the equilibrated structure through MD procedures of the FtsZ dimer described previously[8] as the initial template. We used this model as starting point because it locates the side chain of some critical residues (i.e. the side chain of Asp residues 235, 238 and 72) in a position suit- able for catalysis, solving some of the structural ambiguities exhibited by the original FtsZ dimer structure[78,87]. A scheme of an active interface is shown in figure 3.1. In order to evaluate the behavior of each active GTP centre, all systems were subjected to 8 ns of

unrestrained MD. Sufficient sampling was evaluated as indicated in Supplementary data. Throughout the MD, the position of the water molecules that came within 3.8 Å of the GTP molecules at the interfaces was continuously tracked (see video 1). These included the two water molecules in the coordination sphere of the K+ as well as all the other water molecules that eventually reached the active centre during the MD. Fixed water molecules in the coordination sphere of Mg^{2+} were excluded. For each water molecule traced, two different measurements were recorded (figure 3.1): first the distance ‘‘d’’ from the oxygen atom of the water molecule (O_{wat}) to the phosphorus atom of the gamma phosphate (P_γ) of GTP and also the angle ‘‘α'' between the axis defined by the same O_{wat} and P_γ atoms, and the axis defined by P_γ and the oxygen 3 atom of the beta phosphate (O3β) of GTP. Only water molecules located at a distance of 3.0– 3.8 Å from P_γ and at an angle greater than 110º with respect to the P_γ–O3β axis, can hydrolyze the GTP molecule[72,88,89]. Those FtsZ polymer interfaces containing water molecules that met these two conditions in a stable way were considered as ‘‘active interfaces’’ in terms of GTPase activity.

Trajectory analysis of GTPase interfaces during molecular dynamics of simulated FtsZ polymerization.

Figure 3.2 (A and B) summarize the analysis of GTPase active and inactive interfaces in simulated FtsZ trimer, tetramer and pentamer (a complete continuous measurement of distance ‘‘d’’ and angle ‘‘α’’ of the water molecules at all GTP interfaces during MD are shown in C and D of the figure 3.2).

In the case of the FtsZ trimer (lane III), both interfaces contained water molecules positioned at a suitable distance and angle, which remained constant through almost the whole simulation, as previously described for the FtsZ dimer[8].

In the FtsZ tetramer, only the interfaces located near the ends (IV_A and IV_B), but not central interface (IV_C), had a water molecule positioned at the correct angle and distance for catalysis (i.e. dark green line in plot IV_B of figure 3.2 C). In the pentamer, only interfaces V_A and V_B (nearest to the ends of the polymer) are active, whereas the central interfaces (V_C and V_D) do not maintain a statistically favorable active position for catalysis. In the interface V_D (equivalent to interface IV_B in the tetramer), five molecules of water circulate

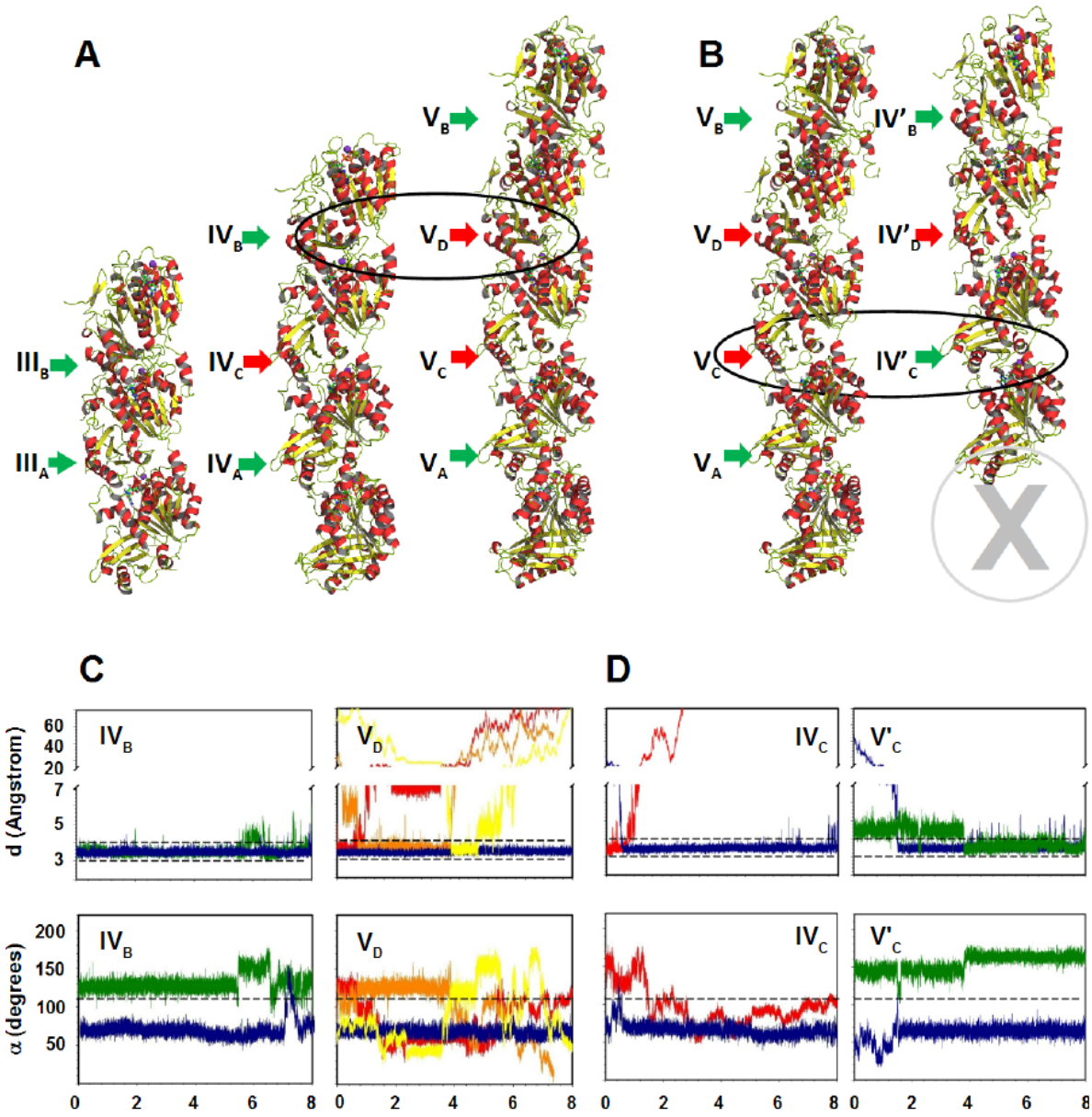

Figure 3.2: Behaviour of FtsZ interfaces in different filaments. (A) Representation of modeled FtsZ trimer (III), tetramer (IV) and pentamer (V). Arrows indicating the position of the interfaces are colored according to their activity state (green: active; red: inactive). Transition from active to inactive state in the case of interfaces IV_B to V_D is highlighted. (B) Representation of modeled FtsZ pentamer (V) and tetramer (IV_0) after subtracting of one FtsZ monomer. Transition from inactive to active state in the case of interfaces V_C to IV'_C is highlighted. (C) Continuous measurement of "d" distance and "α" angle for all water molecules present in the interfaces IV_B and V_D over 8 ns of unrestricted MD. The range 3.0–3.8 Å, which corresponds to the optimal distance, and the position of minimum angle (110°) is indicated in each plot (dotted line). (D) Continuous measurement of "d" distance and "α" angle for all water molecules present in the interfaces V_C and IV'_C over 8 ns of unrestricted MD.

in the neighborhood of P_γ in an unstable way (yellow to red lines in plot V_D of figure 3.2 D), none of them remaining at the end of the trajectory in a stable position to lead to a catalytic event.

In summary, the simulations indicated that only those interfaces located next to the ends of the FtsZ polymers would be active, meaning that there are two active sites per filament. A combination of factors could be the cause of the different behavior exhibited by the interfaces. The relative positions of all the residues located at the interfaces, their contact distances (including the ones used in Mendieta et al.[8]), as well as residues proposed to be related to monomer flexibility[81] were traced during the trajectory. None of them exhibited significant differences among interfaces during the MD procedure, so no individual parameter can be used as a marker for the active/inactive state of each. Alternatively, the putative movements associated with each protein segment in the polymers were modeled using normal mode analysis. The results suggested that the cause appeared to be more associated with the capacity of each monomer to adapt its movements to the adjacent monomers through the corresponding interface (figure 3.6), maintaining the correct geometry of the active site.

GTPase specific activity of FtsZ in vitro.

Consequently with the MD model, if the polymer length increases with the protein concentration, the specific GTPase activity of FtsZ should eventually decrease in parallel, because it is calculated dividing the activity by the total amount of protein, not by the number of active sites. To test this hypothesis, we measured the GTPase specific activity of FtsZ in vitro at different protein concentrations: from 1 to 40 μM. As shown in figure 3.3, once reached the critical concentration for polymerization (1–2 μM), there is an initial phase of rapid increase in GTPase specific activity at concentrations from 1 to 10 μM, which we interpret as resulting from the formation of FtsZ short polymers exhibiting the maximal specific activity (higher ratio of active vs. inactive interfaces). After this, and according to the model, a substantial decrease in specific activity was observed, lowering to 50% of the maximal value at a protein concentration of 40 μM. In previous work by Sossong et al.[80], GTPase activity values were also measured over a similar range of FtsZ

concentration (figure 3.3, insert). The two series of values exhibit a remarkably similar relative variation.

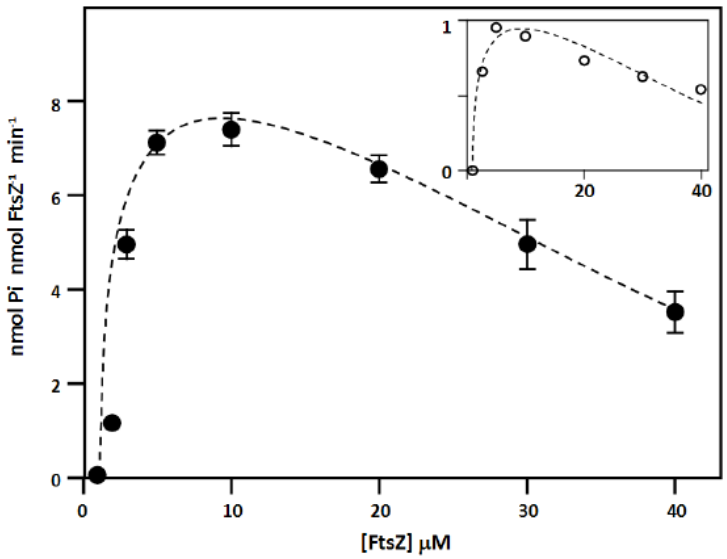

Figure 3.3: Specific GTPase activity at increasing concentrations of FtsZ. Activity data(nmol Pi x nmol $FtsZ^{-1}$ x min^{-1}) were calculated by measuring the slope of the linear part of the activity curves in the presence of GTP. Values are the mean ± standard deviation of four independent experiments. Inset: specific GTPase values re-calculated from previous work by Sossong et al. GTPase activity values shown in their Fig. 1 were converted into specific activity values, by dividing them by the FtsZ concentration, and adapted to a relative scale.

Molecular dynamics of simulated FtsZ depolymerization.

To provide more specific data regarding the relationship between the position of the interface and its active/inactive state, an experiment was designed in which, after the 8 ns MD trajectory, the FtsZ pentamer was shortened to tetramer by removing the lower monomer (figure 3.2 B) and then subjected to an additional 8 ns of unrestricted MD. In the new structure, the interface V_A was no longer present and the interfaces V_B and V_D (now IV'_B and IV'_D) remained in their respective positions and also exhibited the same behavior as in the pentamer. In the interface IV'_B, the initial catalytic water molecule leaves the

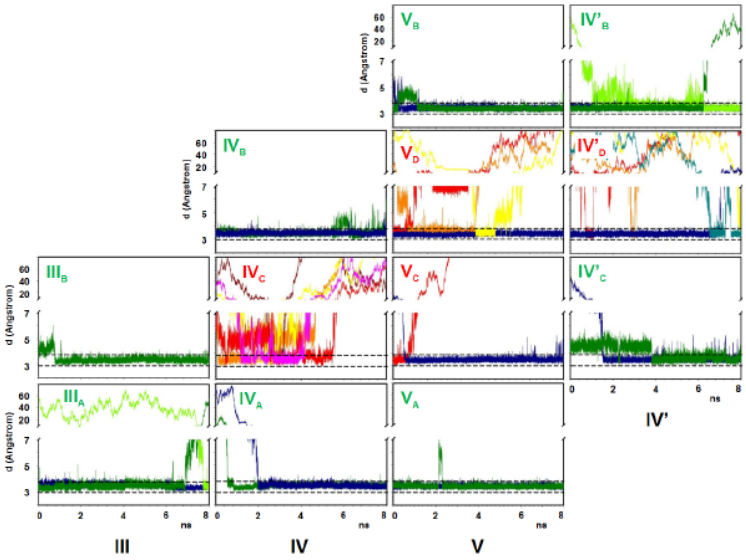

Figure 3.4: Continuous measurement of "d" distance (O_{wat}-P_γ) for all water molecules at the interfaces of the polymers modelled over 8 ns of unrestricted MD. Distances are in angstroms (Å). The range 3.0-3.8 Å, which corresponds to the optimal distance, is indicated (dotted lines). The colour coding of the distance lines in the plots is as follows. Dark green and light green: catalytic water molecules located at less than 3.8 Å and at an angle (see Figure 3.4) of more than 110º; blue: water molecules at less than 3.8 Å with an inadequate angle; yellow to red: unstable water molecules located in the proximities of the active centre. Interfaces are labelled as in Figure 3.2.

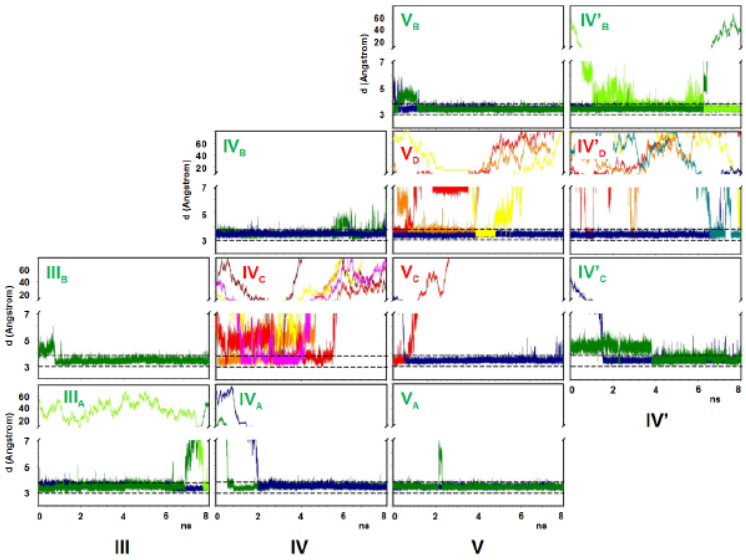

Figure 3.5: **Continuous measurement of "α" angle (O_{wat} - P_γ - $O3_\beta$)** for all water molecules present in the interfaces of modelled polymers over 8 ns of unrestricted MD. The position of minimum angle (110º) is indicated in each plot (dotted line). Colour codes and labels are as in Figure 3.4.

active centre after 6 ns being rapidly replaced by a second water molecule that adopts the correct distance and angle, thus maintaining the active conformation. Results are summarized in table 3.1 and detailed in figures 3.4 and 3.5 A more dramatic effect was observed at interface V_C (now IV'_C) which experimented a change in its position from centred to close- to-end: starting from an inactive state (figure 3.2 B, interface V_C), a water molecule localized in the proximity of P_γ adopted the correct distance and angle within the first 4 ns of the additional trajectory (dark green line in figure 3.2 D, interface IV'_C), maintaining its new stable position until the end of the MD. Thus, the internal structure of an inactive interface switched to active simply due to the change in its relative position from a central to a close-to-end location. In conclusion, our analysis of the different ways in which interfaces in FtsZ polymers behave indicates that GTPase activity may be favoured at the end of the polymers, supporting one putative hypothesis of FtsZ filaments with independent GTP sites recently proposed based on kinetic measurement of GTPase activity[20]. These results raise some questions on the in vivo function of FtsZ, in particular on the connection between the hydrolytic reaction, the filament dynamics and the mechanical properties of the polymer. If the GTPase activity is restricted to filament ends, then polymer bending caused by a putative GTP hydrolysis along the entire polymer cannot be the force generation mechanism. Another important extrapolation of this model is the reduction of the control point to only the two ends. Two discrete points per filament seem easier to manage than a controlling mechanism extending along all the filament subunits.

Trimer (III)	Tetramer (IV)	Pentamer (V)	Tetramer (IV')
-	-	**V_B: 86%**	**IV'_B: 79%**
-	**IV_B: 88%**	V_D: 36%*	IV'_D: 16%
III_B: 92%	IV_C: 29%*	V_C: 09%	**IV'_C: 53%****
III_A: 86%	**IV_A: 92%**	**V_A: 98%**	-

Table 3.1: Maximum time (in percentage of total simulation time) of a water molecule located in a position compatible with GTP hydrolysis in each oligomer interface. In bold, the interfaces predicted to be active. * Active state was maintained only at the beginning of the MD, exhibiting gradual disordering along the trajectory. ** Active state was stabilized during the second half of the trajectory, after gradual re-ordering (see figure 3.2, 3.4 and 3.5)

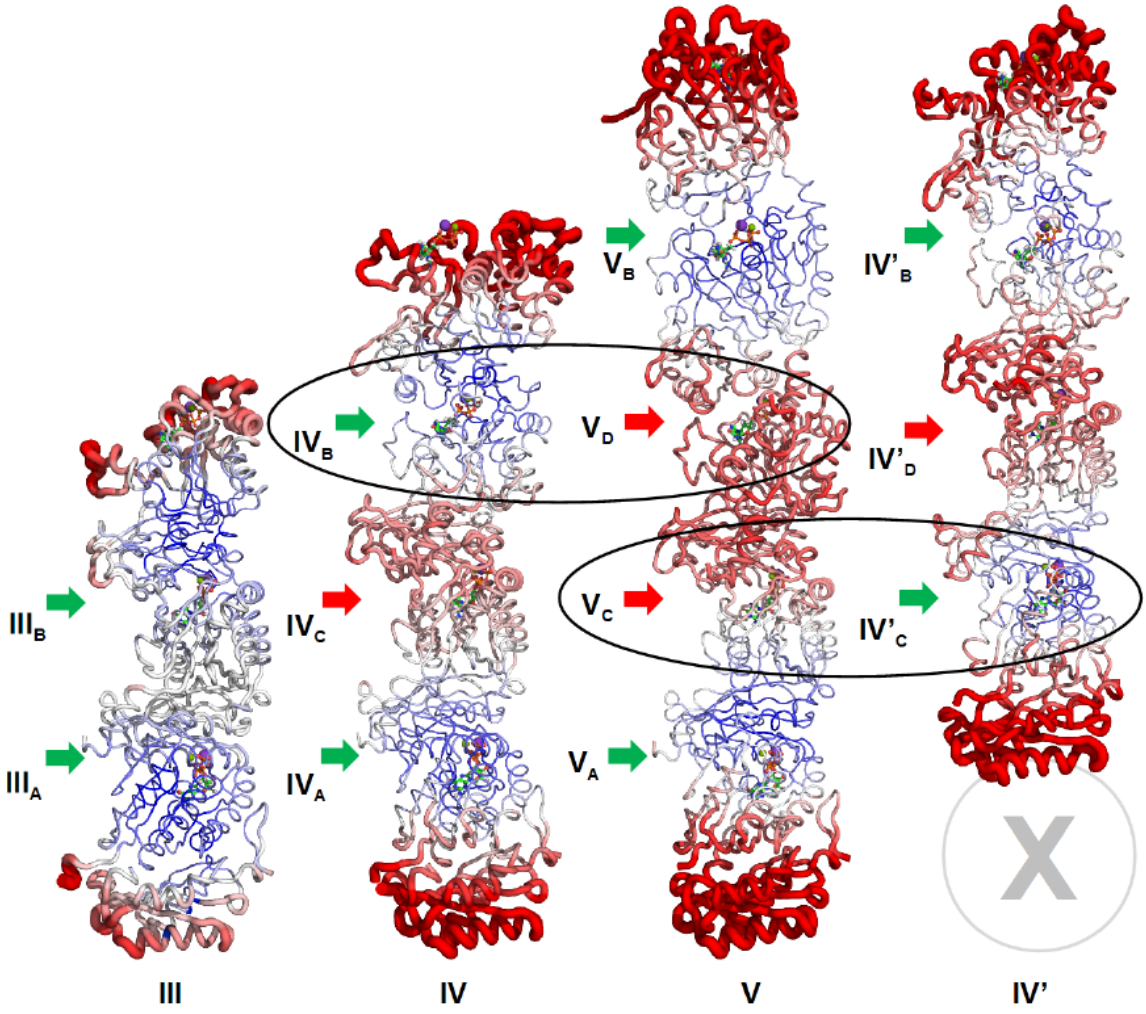

Figure 3.6: Normal mode analysis of modelled FtsZ polymers. The protein trace is coloured according to the normalized mobility values of each residue (blue: limited; red: greater), indicating the putative movements associated with each protein segment in the polymers. Interface labels are coloured according to their predicted behaviour after the MD results: red, inactive; green, active. Normal mode analysis of FtsZ trimer predicted that both interfaces (III_A and III_B) would exhibit limited mobility (green arrows in lane III). In the case of the tetramers and the pentamer, interfaces IV_A, IV_B, V_A, V_B, IV'_B and IV'_C are also predicted to have limited mobility (green arrows in lanes IV, V and IV'). This is probably because the end monomers can adapt their positions relative to the monomers that are in contact with them, thus minimizing the movement of the interface. In contrast, the central interfaces (IV_C, V_C, V_D, and IV'_D) are predicted to experience more movement (red arrows in lane IV and V). Transitions from an active interface (IV_B in lane IV) to an inactive interface (V_D in lane V) and from an inactive interface (V_C in lane V) to an active interface (IV'_C in lane IV') are indicated.

3.2. The role of Gln 61 in HRas GTP hydrolysis: a Quantum Mechanics / Molecular Mechanics (QM/MM) study

3.2.1. Abstract

Activation of the water molecule involved in GTP hydrolysis within the HRas–RasGAP system is analyzed using a tailored approach based on hybrid QM/MM simulation. A new path emerges: transfer of a proton from the attacking water molecule to a second water molecule; then a different proton is transferred from this second water molecule to the GTP. Gln 61 will stabilize the transient OH^- and H_3O^+ molecules thus generated. This newly proposed mechanism was generated by using for the first time the entire HRas-RasGAP protein complex in a QM/MM simulation context. It also offers a rationale explanation for previous experimental results as the decrease of GTPase rate found in the HRas Q61A mutant and the increase exhibited by the HRas Q61E mutant.

3.2.2. Introduction

HRas protein is the most representative member of the Ras subfamily of small GTPases, a form of soluble G-proteins involved in signal transduction[21]. G-proteins usually behave as molecular switches; they shift between an active conformation, bound to GTP, and an inactive GDP-bound conformation. In keeping with their function, these proteins present very low intrinsic GTPase activity and low catalytic efficiency (Kcat/Km ratio); these characteristics lengthen the lifetime of the active state[24]. In the case of HRas, GTP hydrolysis is modulated by the presence of a second protein, the GTPase activating protein RasGAP. RasGAP stabilizes the HRas conformation and supplies an additional arginine

residue that helps to place a water molecule in a position that facilitates catalysis[68], thus leading to GTP hydrolysis and subsequent HRas inactivation.

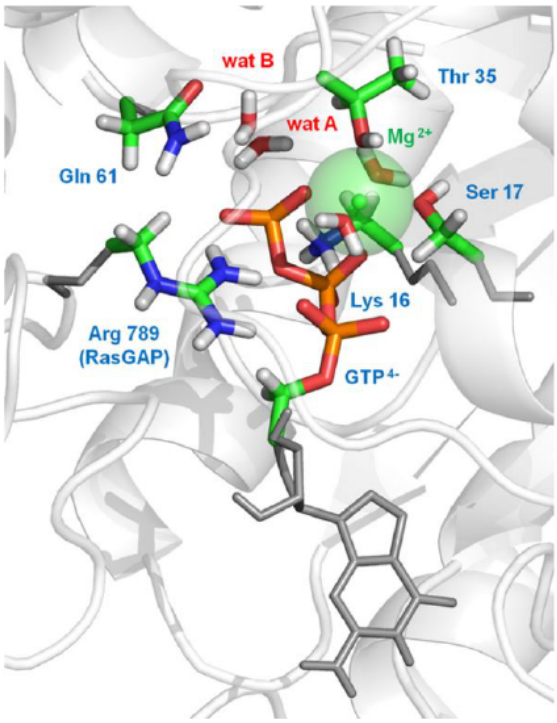

Figure 3.7: Atoms in the QM/MM system. Atoms in the QM region are represented as thick sticks surrounded by the rest of the molecules in the active center included in the MM region (thin sticks) and part of the protein system also in the MM region (ribbons)

Despite all the existing data on the structure and functionality of the HRas–RasGAP complex, the catalytic mechanism of GTP activation and hydrolysis is still controversial[91]. Different mechanisms have been proposed. First, one based on the initial crystal structure[92] suggested that the water molecule was activated by the Gln 61 residue, which acts as the general base proton acceptor. The position of Gln 61 is conserved in the family of guanine nucleotide binding proteins[93] and it is known that mutation of Gln 61 by Ala reduces the rate of GTP hydrolysis[94], whereas substitution of Gln 61 by Glu, which is considered a better proton acceptor, increases the rate of GTP hydrolysis[95]. Some computational studies have simulated GTP hydrolysis on this basis[72]. The conclusions, obtained using an ab initio

6-31 basis set but including only 28 residues in the MM region of the simulation instead of the whole protein, involved an interaction between the amide group of Gln 61 and the carboxylate group of Glu 63, which increases the proton withdrawal potential of Gln 61. This was supported by experiments using an E63K mutant protein[96]. The role of Gln 61 as a proton acceptor has, however, been questioned by other studies[97] that suggest that the GTP molecule acts as the general base for its own hydrolysis[98,99]. In such a substrate-assisted catalysis mechanism, Gln 61 would play a role in the stabilization of the transition state[100]. Finally, more recent reports of the pre-hydrolysis state of the protein obtained using a cryo-technique suggested that a second water molecule is involved in the catalytic mechanism[101], adding an element to the process that, in addition to the introduction of the whole protein structure in the simulations, can be now used to explore alternative ways.

In this work, using a newly developed approach based on the Quantum Mechanics/Molecular Mechanics method, and applying it to the whole HRas-RasGAP protein complex (figure 3.7), we have modeled the activation of the attacking water molecule, a process leading to GTP protonation and subsequent GTP hydrolysis.

3.2.3. Materials and methods

Molecular dynamics (MD) simulations.

The system used in our simulations was based on the X-ray structure of the HRas–RasGAP complex (Protein Data Bank code 1WQ1[28]). It includes the whole HRas protein and residues 714 to 1047 of the catalytic domain of the GTPase activating protein RasGAP. Aluminium fluoride and GDP molecules contained in the crystal structure to mimic the geometry of the transition state of the active center were replaced by a GTP molecule, maintaining all common atoms in the same position as in the crystal structure. The distance between the oxygen of the catalytic water molecule (HOH numbered as “230” in PDB structure 1WQ1) and P_γ of GTP was kept at 2-3 Å by an imposed restraint. A second water molecule was placed in the active site according to the crystal structure of the pre-hydrolysis state obtained by Scheidig et al[101]. K+ ions were added as counterions to neutralize the negative charge of the system and placed in a shell around the system using a

coulombic potential in a grid. The neutralized complexes were then immersed in a rectangular parallelepiped solvent box and a distance of 12 Å was maintained between the wall of the box and the closest atom of the solute. The counterions and the solvent molecules were added using the LEaP module of AMBER[83]. Before the QM/MM simulation itself, 10 ns of unrestricted MD were applied to the system to obtain an equilibrated initial structure, using the PMEMD algorithm of AMBER10 and the parm99 parameter set[83]. Initial relaxation of the system was achieved by performing 10,000 steps of energy minimization using a cut-off of 10.0 Å. Subsequently, and to start the MD simulations, the temperature was raised from 0 to 298 K in a 500-ps heating phase, and velocities were reassigned at each new temperature according to a Maxwell-Boltzmann distribution. During this period, the dihedral of the Cα trace were restrained with a force constant of 500 kcal mol-1 rad-2. During the last 200 ps of the equilibration phase of the MD, the force constant was reduced stepwise to 0. The SHAKE algorithm was used throughout to constrain all the hydrogen bonds to their equilibrium values so that an integration time step of 2 fs could be employed. The list of non-bonded pairs was updated every 25 steps, and coordinates were saved every 2 ps. Periodic boundary conditions were applied and electrostatic interactions were represented using the smooth particle mesh Ewald method with a grid spacing of about 1 Å.

Quantum Mechanics/Molecular Mechanics (QM/MM) simulations

After initial relaxation using MD procedures, as described above, QM/MM simulations were performed using the sander module of AMBER10[95]. The hybrid QM/MM approach employed has been described above, in the mateerial and methods chapter. In our system, the QM region included the two water molecules involved in catalysis and the GTP atoms from the gamma-phosphate group up to the C5'-C4' bond. It also included the Mg++ ion and all the oxygen atoms involved in its coordination sphere, including the hydroxyl groups of the Ser 17 and Thr 35 residues and two coordinating water molecules. Side chains of the Lys 16 and Gln 61 residues of HRas and the chain of Arg 789 belonging to the RasGAP protein were also included (atoms in the QM region are represented in figure 3.7). The QM region contained 75 atoms including 6 link H-atoms used to maintain the integrity of the

covalent bonds sectioned by the QM/MM boundary[83]. In the case of the experiment illustrated in figure 3.8, carboxylate atoms of Glu63 were also included in QM region. The conformation obtained after MD procedures (see above) was equilibrated again for 200 ps using this QM/MM approach. During this equilibration, constraints corresponding to all the covalent bonds between the atoms of the QM region were maintained. All the constraints, except those corresponding to the parameters of the reaction and to the maximum allowable distance position of the catalytic water, were removed progressively over the next 100 ps. In all cases, position of the catalytic water molecule was stable under QM/MM simulation without position restrains. SHAKE was not used for either the MM region or the QM region. Due to the presence of a peptide bond in the side chain of the Gln 61 residue, an MM correction to the peptide linkages was used. A cutoff of 8 Å was used to calculate the QM/MM electrostatic interactions. The extra Gaussian terms that were used in the PM3 Hamiltonian to improve the core-core repulsion term in QM-QM interactions were also included for the QM-MM interactions.

Energy surface calculations.

For the calculation of the energy in the reaction, equation 2.2.1 was used, taking into account only H_{QM} and $H_{QM/MM}$ terms. The reaction coordinates used in the calculations were the bond-breaking distance from the proton to the oxygen atom in the attacking water molecule and the bond-forming distance from this same proton to the oxygen in the ε1 position of Gln 61 (figure 3.8) or to the oxygen of the second water molecule (figure 3.9 and figure 3.10). For each trajectory, the value of one coordinate (x–axis) was increased throughout the simulation time, as in steered MD[102], from 0.95 to 1.8 Å in steps of 0.025 Å; whereas the size of the flat-bottomed part of the harmonic potential of the other coordinate (y–axis) was kept constant within a narrow range. The next trajectory was generated by adding an increment of 0.025 Å to the constant value of the y–axis. A certain degree of overlapping of points was observed from one trajectory to the next, as occurs in umbrella sampling procedures[103], ensuring a better coverage of the whole surface. The generation of a large number of trajectories made it possible to explore in detail the conformational space defined by the reaction coordinates. We sampled the conformational space of the reaction

with around 12,000 homogeneously distributed points on a surface of 0.85 x 0.85 Å. 3D smoothing of the data was applied using the local smoothing technique with tricube weighting and polynomial regression (LOESS). Final ΔG^{o} values in the energy maps were normalized by subtracting from all of them the lowest value obtained in each experiment. Minimum values that were adjusted to 0 in each map were those corresponding to the initial state (figure 3.8) or to the final state (figure 3.9 and figure 3.10).

3.2.4. Results and discussion

In order to evaluate the different mechanisms proposed for the activation of the water molecule involved in catalysis, as well as to identify the proton acceptor and the role of Gln 61 in each pathway, a specific simulation method was devised based on the hybrid QM/MM potential[104] implemented in the AMBER10 package[49,105]. The approach, which shares some characteristics with steered molecular dynamics (MD) and also with umbrella sampling procedures, uses the adaptively biased MD[106] method to obtain a free energy surface in the conformational space defined by the reaction coordinates. Analysis of the surface characteristics provided us with variations in the free energy which helped to describe suitable reaction mechanisms.

Gln61 as the general base

Using the same approach, and including in the QM/MM system the whole crystallized structure of the HRas–RasGAP complex[90], two alternative hypotheses for the activation of the attacking water molecule were tested. As indicated under Materials and Methods, the complete simulation system comprises not only the atoms in the active center but, for the first time to our knowledge, the whole protein complex, all included in a solvent box (total number of atoms > 53,000). The first mechanism analyzed included the assumption that Gln 61 acts as a proton acceptor assisted by Glu 63. The free energy landscape obtained for the activation of the catalytic water molecule (figure 3.8) gave a ΔG^{o} of around 25 $kcal \cdot mol^{-1}$ between the initial and the final states; even higher than for the ion product of water in a polar environment (see above). This result ruled out the possibility of Gln 61 acting as a general base, in agreement with previous equilibrium studies of proton transfer

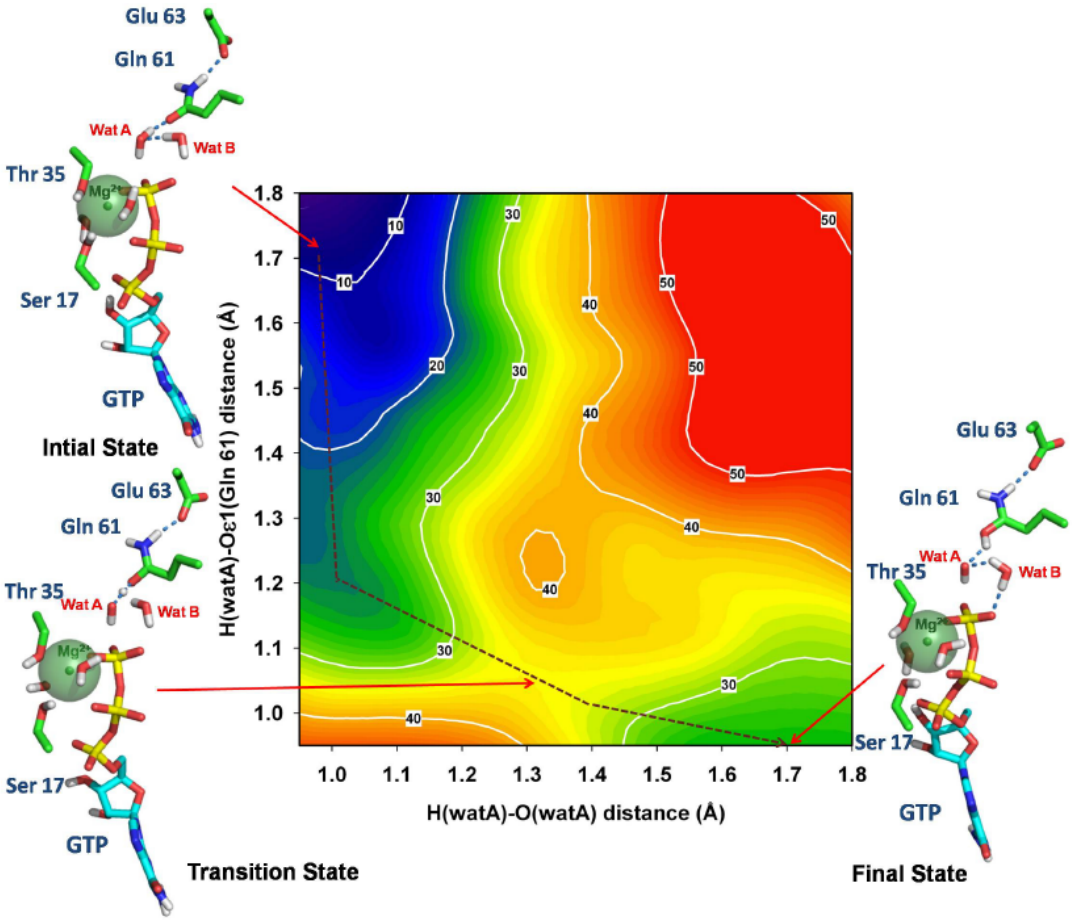

Figure 3.8: Free energy landscape for the activation of the catalytic water molecule supposing Gln61 as proton acceptor assisted by Glu63. ΔG° values obtained for the whole system are shown for the different states corresponding to the distances: x axis, from proton to oxygen atom in the attacking water molecule (wat A); y axis, from the same proton to the oxygen in ε1 position of Gln^{61}. The structure of the active center in the initial state (minimun $\Delta G^{\circ}_{GTP}{}^{-4}$ value adjusted to 0), transition state ($\Delta G^{\circ}_{GTP}{}^{-4}$ ~ 35 kcal·mol^{-1}), and final state ($\Delta G^{\circ}_{GTP}{}^{-4}$ ~ 25 kcal·mol^{-1}) is depicted, indicating the relative positions of Ser^{17}, Thr^{53}, Gln^{61}, Glu^{63}, GTP, and the two water molecules (wat A and wat B).

from the catalytic water molecule to this residue[97]. In order to provide further support to this statement, an unconstrained simulation of the reaction once the final state was reached, was performed (figure 3.11). As expected, in absence of restrictions, the system reverted spontaneously from these situation to the initial reaction substrates (unprotonated Gln61 plus two water molecules), indicating that the products were not stable in the simulated conditions.

ATP as the general base

Once Gln 61 was ruled out as the proton acceptor, an alternative mechanism was tested. According to the arrangement of the side chain of residues and water molecules at the catalytic site of the pre-hydrolysis state of HRas[101], the proton transfer may occur between an attacking water molecule (wat A) and a second water molecule (wat B) acting as the proton acceptor. The free energy landscape obtained in these conditions (figure 3.9) indicated that ΔG^{o} between the initial and the final states is around -3 kcal·mol^{-1}. Detailed analysis of the structure that corresponds to the final state (figure 3.9, bottom right) showed that the GTP molecule is protonated, suggesting that $GTPH^{3-}$ is more stable than GTP^{4-} in the environment of the catalytic site. In addition, and unexpectedly, the results indicated that the proton bound to the GTP molecule is initially not part of wat A but of wat B, which acted as the initial proton acceptor. In order to analyze the behavior of the protons in detail throughout the process, the reaction along the minimum energy path between the initial and the final states was simulated (dashed line in figure 3.9). Tracing the proton movements (a video of the entire process is provided as movie 2) revealed that a proton from wat A is initially transferred to wat B thus forming a hydronium ion. Then, towards the end of the trajectory, a different proton from wat B is transferred to GTP. In addition, at this precise moment, the free OH^{-} molecule approaches the GTP γ-phosphate forcing it to adopt the trigonal bipyramidal geometry characteristic of the pentacovalent state, suggesting a S_N2 reaction mechanism[107]. Although, in our conditions, a S_N2 mechanism appeared as favorable, further studies are still needed to fully unravel the complete mechanism of the GTP hydrolysis as we cannot completely rule out alternative mechanisms, i.e. S_N1 as it has been proposed for Hsc70 ATPase[108].

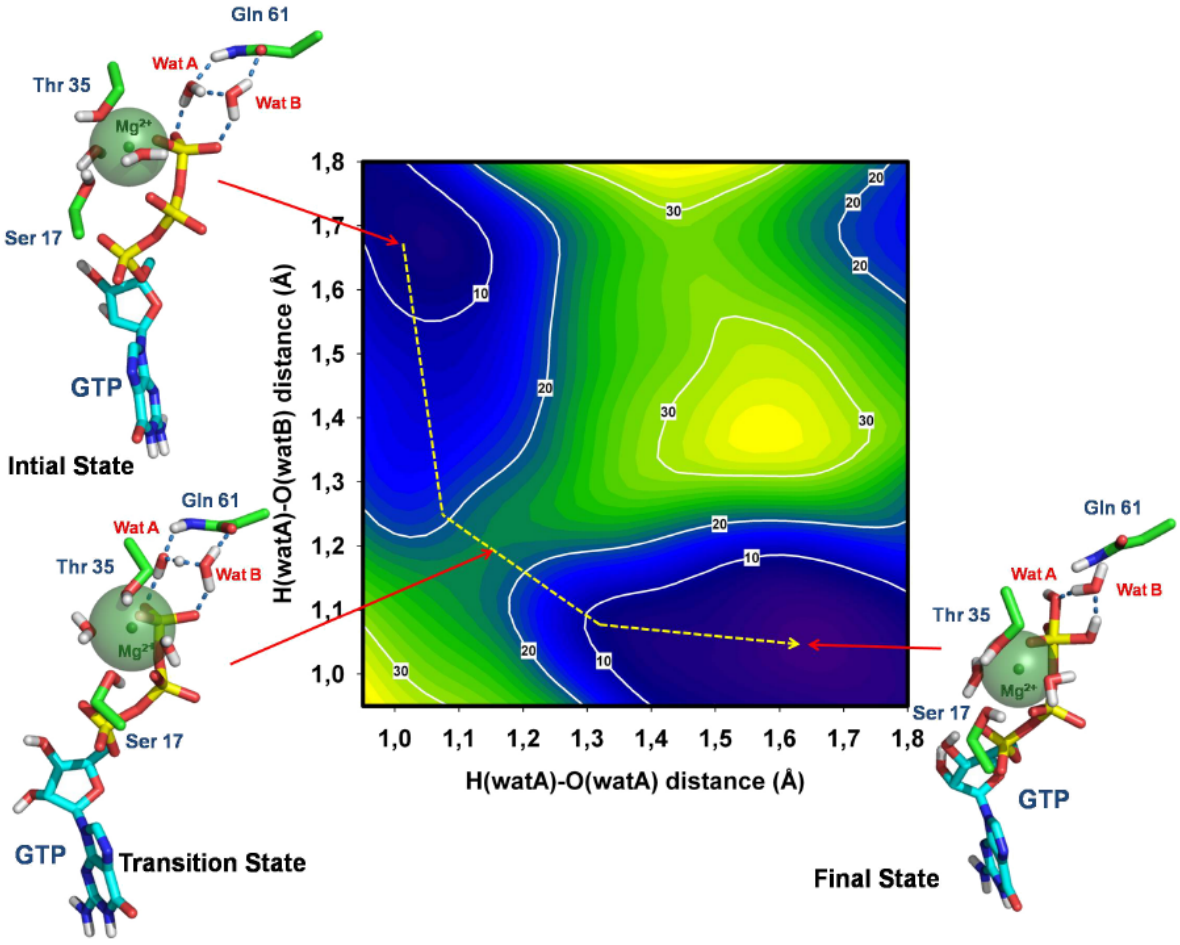

Figure 3.9: Free energy landscape for the proton transfer between the attacking water molecule (wat A) and the second water molecule (wat B). The axes represent the following distances: x axis, from proton to oxygen atom in the attacking water molecule (wat A); y axis, from the same proton to the oxygen atom in the second water molecule (wat B). The structure of the active center in the initial state ($\Delta G^{o}_{GTP}{}^{-4}$ ~ 3 kcal·mol^{-1}), transition state ($\Delta G^{o}_{GTP}{}^{-4}$ ~ 22 kcal·mol^{-1}), and final state ($\Delta G^{o}_{GTP}{}^{-4}$ value adjusted to 0) is depicted, including the same residues as in Fig. 1. GTP molecule with pentacovalent P_γ in the final state is shown by GTP*. Note the position of Gln^{61} in the transition state, where it stabilizes the position of the water molecules during proton transfer.

In our model, the atomic rearrangement at the end of the trajectory suggests that, in the absence of the simulation restraints used to obtain the energy surface, the hydrolysis of GTP can eventually occur spontaneously from the pentacovalent structural state. Additionally, it suggests that the rate-limiting step of the reaction catalyzed by HRas in the presence of RasGAP would be the initial activation of wat A. Under this second hypothesis, the value of the energy barrier located between the initial and final states and corresponding to the transition state, is around 22 kcal·mol^{-1}, lower than the 28 kcal·mol^{-1} of the barrier for the ionization of water in a polar environment (figure 2.5). Analysis of the structure of the transition state at the saddle point (figure 3.9, bottom left) showed a strong interaction between the amide group of Gln 61 and the OH^- formed from wat A. An interaction between a carbonyl atom in the side chain of Gln 61 and the hydronium ion formed by wat B also contributes to the stabilization of the transition state. These results agree with the previously proposed mechanism of substrate-assisted catalysis[98,99], although for the first time they indicate the essential role of wat B in the process: to mediate the transfer of the proton from wat A to GTP. The role of Gln 61 in this new scenario consists of stabilizing the OH^- and H_3O^+ molecules that are transiently generated during proton transfer.In the same free energy landscape, a second local minimum can be observed (figure 3.9, top right-hand corner). The structure of the active center in this region corresponds to protonated GTP. Nevertheless, the origin of the proton bound to GTP in this case is not wat B but wat A, without the mediation of wat B. Although this alternative pathway cannot be completely ruled out, its higher energy barrier (28 kcal·mol^{-1} vs. 22 kcal·mol^{-1}) make it less favorable.

Q61A muntant

In order to obtain additional results to support the proposed role of Gln 61, a second simulation using the same QM/MM approach was performed. In this case, the Gln 61 residue was replaced by Ala to mimic experimental conditions in that reduced GTPase activity of HRas was previously measured *in vitro*[94]. The free energy landscape obtained for the same proton transfer path is shown in figure 3.10. The value of the new energy barrier is clearly higher (around 38 kcal·mol^{-1}). In fact, in the absence of the Gln 61 side chain that stabilizes the transient OH^- and H_3O^+ molecules, the alternative path for direct protonation

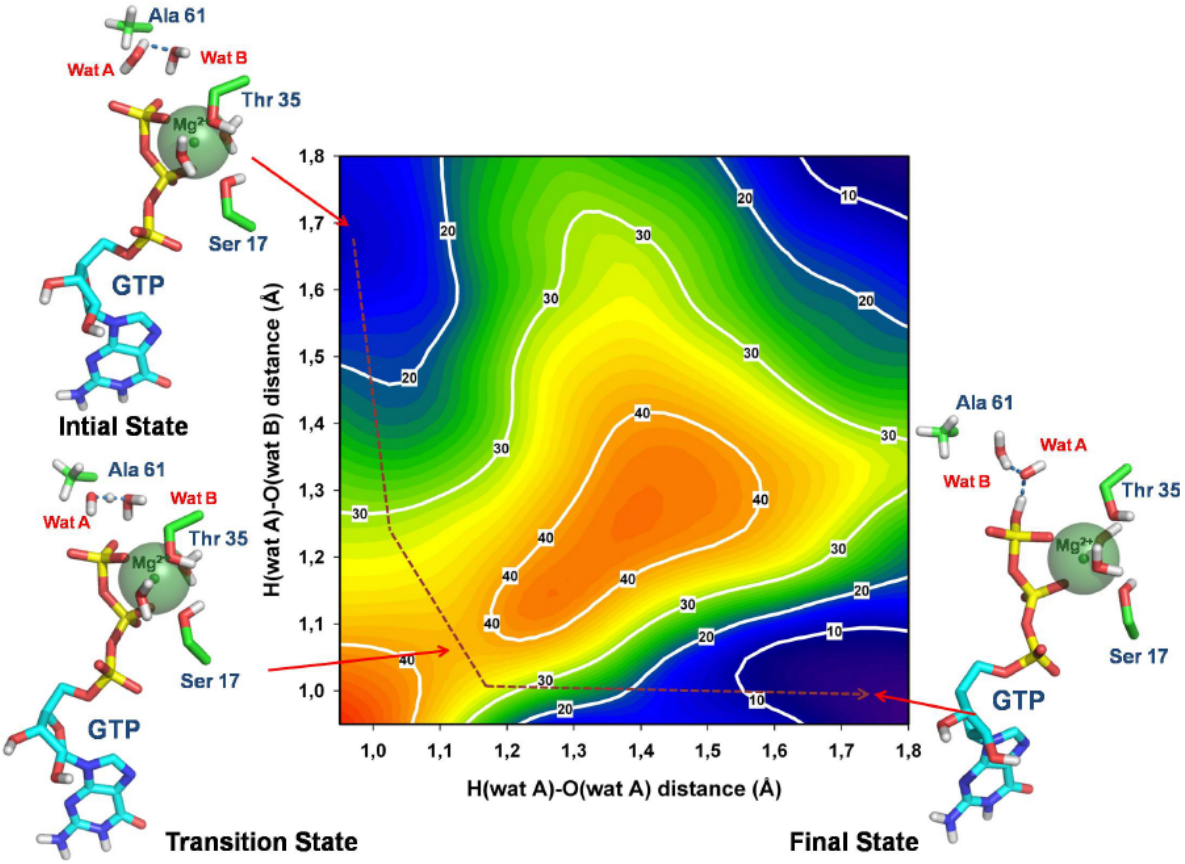

Figure 3.10: Free energy landscape for the Q61A mutant. Axes are defined as in Fig. 3.9. The structure of the active center in the initial state ($\Delta G°_{GTP}{}^{-4}$ ~ 11 kcal·mol^{-1}), transition state ($\Delta G°_{GTP}{}^{-4}$ ~ 38 kcal·mol^{-1}), and final state ($\Delta G°_{GTPH}{}^{-3}$ value adjusted to 0) is depicted, including the side chain of Ala61 residue instead of Gln61.

of GTP appeared to be more favorable, in direct opposition to the situation observed for the wild type protein. These results support the role of Gln 61 in the stabilization of the proton transfer path from wat A to wat B and then to GTP. In addition, they also explain the decrease of the GTPase rate found in vitro for the Q61A mutant[94]. Moreover, theproposed mechanism could also offer a rational explanation for the increased rate of GTP hydrolysis found in the Q61E mutant[95], as it is conceivable that a negatively charged glutamic residue would stabilize better than glutamine the transient hydronium ion.

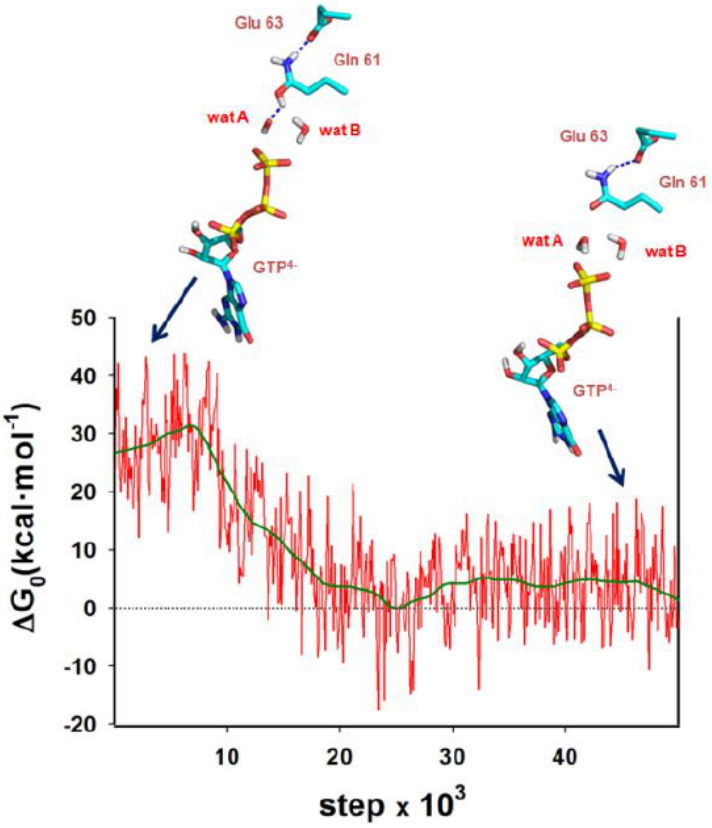

Figure 3.11: Unconstrained simulation of protonated Gln61. Final state of the experiment shown in Figure 2 was subjected to 5.000 steps of unconstrained simulation and ΔG° variation (in kcal•mol^{-1}) was continuously measured. The structure of the active center at the beginning (protonated Gln 61) and after the simulation process in absence of constrains (unprotonated Gln61 plus two water molecules, equivalent to the initial reaction substrates), are represented.

3.3. Simulation of catalytic water activation in mitochondrial F_1-ATPase using a hybrid quantum mechanics/molecular mechanics approach: An alternative role for β-Glu 188

3.3.1. Abstract

The use of quantum mechanics/molecular mechanics (QM/MM) simulations to study the free energy landscape of the water activation at the catalytic site of mitochondrial F_1-ATPase affords us insight into the generation of the nucleophile OH^- prior to the ATP hydrolysis. As a result, the ATP molecule was found to be the final proton acceptor. In the simulated pathway, proton transfer to the nucleotide was not direct, but occurred via a second water molecule in a similar way to the Grotthuss mechanism proposed for proton diffusion. Residue β-Glu 188, previously described as the putative catalytic base, was found to be involved in the stabilization of a transient hydronium ion during the water activation. Simulations in the absence of the carboxylate moiety of β-Glu 188 support this role.

3.3.2. Introduction

Mitochondrial F_oF_1-ATP synthase is the enzyme responsible for the synthesis of ATP from ADP and P_i ($H_2PO_4^-$)[35,109]. Unlike the majority of the enzymes, which work increasing the rate of the reactions that they catalyze, F_oF_1-ATP synthase can force the reaction far from equilibrium by harnessing the proton gradient[110], working as a nanomachine that operates as a mechanical/chemical energy transducer[111]. Therefore, although the hydrolysis of ATP to ADP and P_i is an exergonic process, due to the activity of F_oF_1-ATP synthase, the

concentration ratio ATP:ADP/P_i is close to 1:1 in mitochondria[112,113]. The enzyme is able to generate this gradient even under conditions that favor the hydrolysis reaction by a factor of 2×10^5 [40].

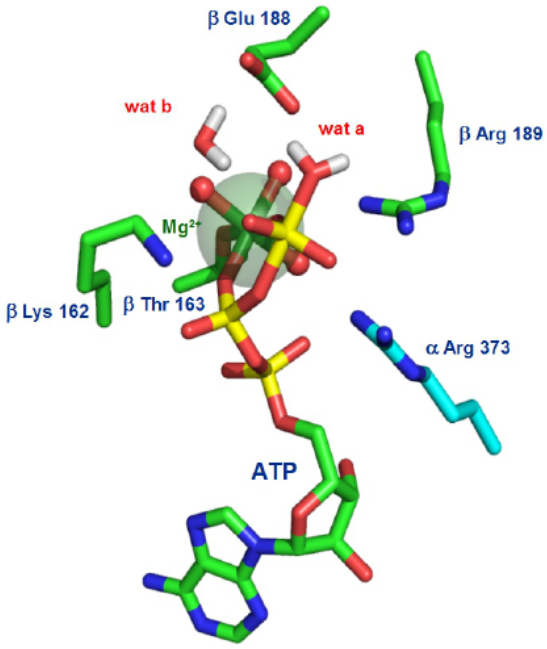

Figure 3.12: Schematic representation of the active center of F1-ATPase. The ATP molecule was generated by superimposing the corresponding atoms of the F1-ATPase structure on the original ADP and aluminium fluoride molecules (PDB entry 1H8E). ATP is associated with a Mg^{2+} atom (green sphere), which includes three water molecules in its coordination sphere (red spheres indicate the positions of the oxygen atoms of the water molecules). The positions of the catalytic water (wat A) and an additional water molecule (wat B) are indicated. Residues of the β subunit (Glu 188, Lys 162, and Arg 189) and the α subunit (Arg 373) located in the active center are also labelled The inset shows a complete MD system, including whole α (white) and β (blue) subunits and showing the position of the atoms in the active center (spheres) at the interface between the two subunits after unconstrained MD simulation.

The three-dimensional structure of F_oF_1-ATP synthase[38,39,114] presents two different sub-structures: a globular catalytic moiety (F_1) and a transmembrane portion (F_o) whose rotation is induced by the proton gradient. F_1-ATP synthase has three pairs of α and β subunits located around the γ subunit, which presents a globular head domain and an extended

coiled-coil tail[38]. Both the α and β subunits are able to bind to nucleotides, but only β subunits are considered catalytically active. Most of the residues that stabilize ATP binding are located in the β subunit, with the catalytic site located at the interface between the α and β subunits. An α subunit closes the catalytic site on the triphosphate moiety. At least one residue from the α subunit (α-Arg 373) is hydrogen-bonded to an oxygen of the nucleotide.

It has been demonstrated that ATP synthesis depends on an external torque on the γ subunit, causing a clockwise rotation and successive structural modifications of the pairs of α and β subunits that results in the generation of ATP molecules from ADP + P_i[115]. During ATP hydrolysis, in contrast, the enzyme rotor turns counterclockwise also in a complex mechanism involving different conformations of the β subunits and consecutive steps of liberation of P_i, Mg^{++} and ADP after the hydrolysis event[116].

As a general mechanism, the first steps in the hydrolysis of triphosphate nucleotides require the activation of a water molecule. The resulting OH^- group attacks the γ-phosphate atom, weakens the bond between this atom and the bridging oxygen and forces the atoms to adopt the trigonal bipyramidal geometry characteristic of the pentacovalent transition state[107]. When the attacking nucleophile has a relatively high pKa, e.g. the case of water, it is commonly assumed that the biological phosphoryl transfer reaction has to be catalyzed by a general base that initially accepts the proton from the nucleophile[117]. The identity of the general base in the different reactions has often been a question of interest and debate.

In the case of F_1-ATPase, the first ATP hydrolysis mechanism proposed was based on the crystal structure of the protein, solved in the presence of ADP and P_i by Abrahams *et al.*[38]. In that structure, the residue β-Glu 188 is hydrogen-bonded to the catalytic water in a position that makes it a strong candidate to act as the catalytic base. In the reverse reaction, during ATP synthesis, β-Glu 188 could provide the proton for the abstraction of the water molecule[118]. The ability of glutamic acid to act as the general base in ATP hydrolysis depends on its pKa. In general, glutamic residues require a very hydrophobic environment to operate as efficient proton acceptors. Although a study predicted that the pKa of a modified residue in the equivalent position in thermophilic F_1-ATPase could be particularly high[119], in the polar environment of the active site of F_1-ATPase it is supposed that the pKa

of β-Glu 188 is low. In these conditions, the pKa of the ATP molecule makes it a much more favorable candidate general base[120]. This hypothesis, in which the nucleotide is involved in its own hydrolysis, has previously been proposed for ras p21[99], and named substrate-assisted catalysis[98].

A favorable reaction path with ATP as the final proton acceptor has previously been proposed for F_1-ATPase using a theoretical approach[121,122]. The proton transfer to the ATP along that path involves two water molecules. However, the system used in the study was only a part of the total β subunits and more than 50% of the atoms included in the study were fixed to preserve the overall shape of the system. More recently, protonation of β-Glu 188 involving three water molecules has been described assuming a different reaction path in a "dissociative" manner[123]. Experimental observations were provided in the same work showing that the rate-determining step of the reaction is the proton transfer from the catalytic water molecule. In similar systems, indirect proton transfer has also been proposed for myosin-catalyzed ATP hydrolysis[81].

We have recently developed a quantum mechanics/molecular mechanics (QM/MM) approach to generate the free energy surface of the conformational space defined by the reaction coordinates for the water activation in the GAP/ras p21 complex[53]. The results suggest that GTP can act as the catalytic base, with the proton being transferred from the attacking water molecule to the nucleotide along an indirect path involving a second water molecule; the most favorable pathway. In this work we use a similar approach to study an F_1-ATPase QM/MM system that includes in the MM region the whole α and β subunits in a fully unconstrained simulation, in contrast to the previously published analysis, that includes in the MM only a small portion of the proteins around the active site[121–123]. The study of the energy surface could afford us insight into the residues involved in the activation of the catalytic water molecule in F_1-ATPase.

3.3.3. Material and methods

Molecular dynamics (MD) simulations

The system used in our simulations was based on the X-ray structure determined by Menz et al.[39] (PDB access code 1H8E). The system includes the F and B chains corresponding to the α and β subunits. Aluminum fluoride and ADP molecules included in the crystal structure so that it resembles the structure of the transition state were replaced by an ATP residue with transition state geometry. The distance between the oxygen of the catalytic water molecule (HOH numbered as F2128 in PDB structure 1H8E) and the Pγ of the ATP was held at 1.9 Å by an imposed restraint. The distance between the Pγ of the ATP and the bridging oxygen interacting with the Pγ was also restrained to 1.9 Å. A second water molecule (HOH F2130) was also included in the simulated system.

The system was immersed in a rectangular parallelepiped solvent box and a distance of 12 Å was maintained between the wall of the box and the closest atom of the solute. K+ ions were added to neutralize the negative charge, with the counterions placed in a shell around the protein moiety using a Coulomb potential in a grid. The counterions and the solvent were added using the LEaP module of AMBER[83]. All the MD simulations were performed using the AMBER11 PMEMD program[83,124] and the parm99 parameter set[83]. Initial relaxation of the system was achieved by performing 10,000 energy minimization steps using a cut-off of 10.0 Å. Subsequently, and to start the molecular dynamics (MD) simulations, the temperature was raised from 0 to 298 K in a 500-ps heating phase, and velocities were reassigned at each new temperature according to a Maxwell-Boltzmann distribution. During this period, the dihedrals of the Cα trace were restrained with a force constant of 500 kcal mol-1rad-2. For the last 200 ps of the equilibration phase of the MD, the force constant was reduced stepwise to 0. The SHAKE algorithm was used throughout to constrain all hydrogen bonds to their equilibrium values so that an integration time step of 2 fs could be employed. The list of non-bonded pairs was updated every 25 steps, and coordinates were saved every 2 ps. Periodic boundary conditions were applied and electrostatic interactions were represented using the smooth particle mesh Ewald method

with a grid spacing of ~ 1 Å. The length of the trajectories was 10 ns in all cases.

Quantum Mechanics/Molecular Mechanics (QM/MM) simulations

The hybrid QM/MM approach is a suitable simulation method to study processes in which chemical bonds are formed and broken, such as enzymatic reactions. The QM/MM simulations were performed using the sander module of AMBER11[124]. The method requires the partitioning of the system into two regions: QM and MM. Calculations involving the atoms belonging to the QM region were performed using the PM3 semi-empirical Hamiltonian. The atoms in the system that are not part of the QM region (those in the MM region) were treated in a classical MM way. In our system, the QM region includes two water molecules involved in the catalysis, and the ATP atoms from the gamma-phosphate group up to the C5'-C4' bond. It also includes the Mg^{++} ion and all the oxygen atoms in its coordination sphere, including the hydroxyl group of Thr 163 from the β subunit and three coordinating water molecules. Side chains of the Glu 188, Lys 162 and Arg 189 residues of the β subunit and the side chain of the Arg 373 belonging to the α subunit are also included. The QM region contains 83 atoms including 6 link H atoms used to complete the covalent bonds cut by the QM/MM boundary[83]. The MM region includes all the other atoms described in the MD simulations (see above) including the whole α and β ATPase subunits. The conformation obtained after classical MD was then equilibrated again for 200 ps using this QM/MM approach. During the equilibration, the constraints corresponding to all the covalent bonds between the atoms in the QM region were maintained. All the restraints, except those corresponding to the parameters of the reaction and the position of the catalytic water, were progressively removed over the next 100 ps. SHAKE was not used for either the MM region or the QM region. A cutoff of 8 Å was used to calculate the QM/MM electrostatic interactions. The extra Gaussian terms that are used in the PM3 Hamiltonian to improve the core–core repulsion term in QM-QM interactions were also included for the QM-MM interactions.

Energy surface calculations

In order to explore the most important part of the conformational space defined by the

reaction coordinates, a new approach was developed. The approach is based on adaptively biased MD[106] and presents some characteristics of steered MD[102] as well as an umbrella sampling[103] procedure. The O-H distance in the catalytic water (breaking distance) and the distance between the H of this catalytic water and the O of the second water molecule (bond distance) were used as the reaction coordinates in all the calculations. The QM/MM trajectories were performed by restraining both reaction coordinates using harmonic potentials with a flat bottom and parabolic sides. For each trajectory, the value of one coordinate was increased over the simulation time, as in steered MD[102], from 0.95 to 2.0 Å; while the size of the flat bottom part of the harmonic potential of the other coordinate was kept constant within a narrow range. Overlapping of points between neighboring trajectories was observed, as in umbrella sampling procedures[103], ensuring good coverage of the whole surface. The generation of a large number of trajectories makes it possible to explore the conformational space defined by the reaction coordinates in detail.

For the representation of the energy in surface maps, only H_{QM} and $H_{QM/MM}$ terms were considered. We sampled the conformational space of the reaction at around 12,000 homogeneously distributed points on a surface of 1.05 x 1.05 Å. 3D smoothing of the data was applied using the local smoothing technique with tricube weighting and polynomial regression (LOESS).

3.3.4. Results and discussion

The starting configuration of the system used in this work includes the whole α and β subunits of the mitochondrial F1-ATP synthase corresponding to the x-ray crystallographic structure determined by Menz et al.[39] (PDB access code 1H8E). The crystal structure contains ADP and aluminum fluoride so as to emulate the hydrolysis transition state. In our system the ADP/aluminum fluoride pair has been replaced by an ATP residue with transition state geometry. The residue was created by overlapping the corresponding portion with the ADP and superimposing the γ-phosphate on the Al atom. The minimum constraints necessary to maintain the geometry were applied to the distances and angles (see Experimental Procedures above). figure 3.12 shows the active site containing five water

molecules: three of them are included in the coordination sphere of the Mg^{++} ion. The other two water molecules are the catalytic one (wat A), which is involved in the generation of the transition state, and an additional water molecule (wat B) that is also located at the active site. After minimization and equilibration, a productive MD of 10 ns was performed. Figure 3.17A shows the root mean square deviation (RMSD) corresponding to the trace of the C-α atoms. During the simulation, the RMSD was almost constant at around 2 Å. This value, close to the resolution of the crystallized structure, shows that no significant conformational changes occurred during the adaptation of the system to the AMBER force field.

The behavior of the water molecules wat A and wat B present in the active center was also monitored throughout the trajectory. Both of them are hydrogen bonded to β-Glu 188. Wat B also interacts through hydrogen bonding with an oxygen in the γ-phosphate. Figure 3.17B shows that the distances corresponding to these interactions remained constant during the simulation. The stability of the interaction between these water molecules and the residues of the active center suggests that both of them could be involved in the first steps of the reaction. Wat A and wat B were included in the QM region of the QM/MM system.

In order to study the nature of the catalytic base involved in ATP hydrolysis in F1-ATPase, we used the hybrid QM/MM potential implemented in the AMBER11 package[124]. This approach presents some characteristics of steered MD and others of umbrella sampling procedures, and it is based on the adaptively biased MD method[106] to obtain a free energy surface of the conformational space defined by the reaction coordinates, as described previously[53].

β-Glu 188 as the general base

Figure 3.13 shows the free energy landscape obtained for the activation of the catalytic water molecule assuming β-Glu 188 is the proton acceptor. The reaction parameters used in this case were the distance between the proton and the oxygen of the catalytic water and the distance between the proton and oxygen Oδ1 of the carboxylic group of β-Glu 188. The

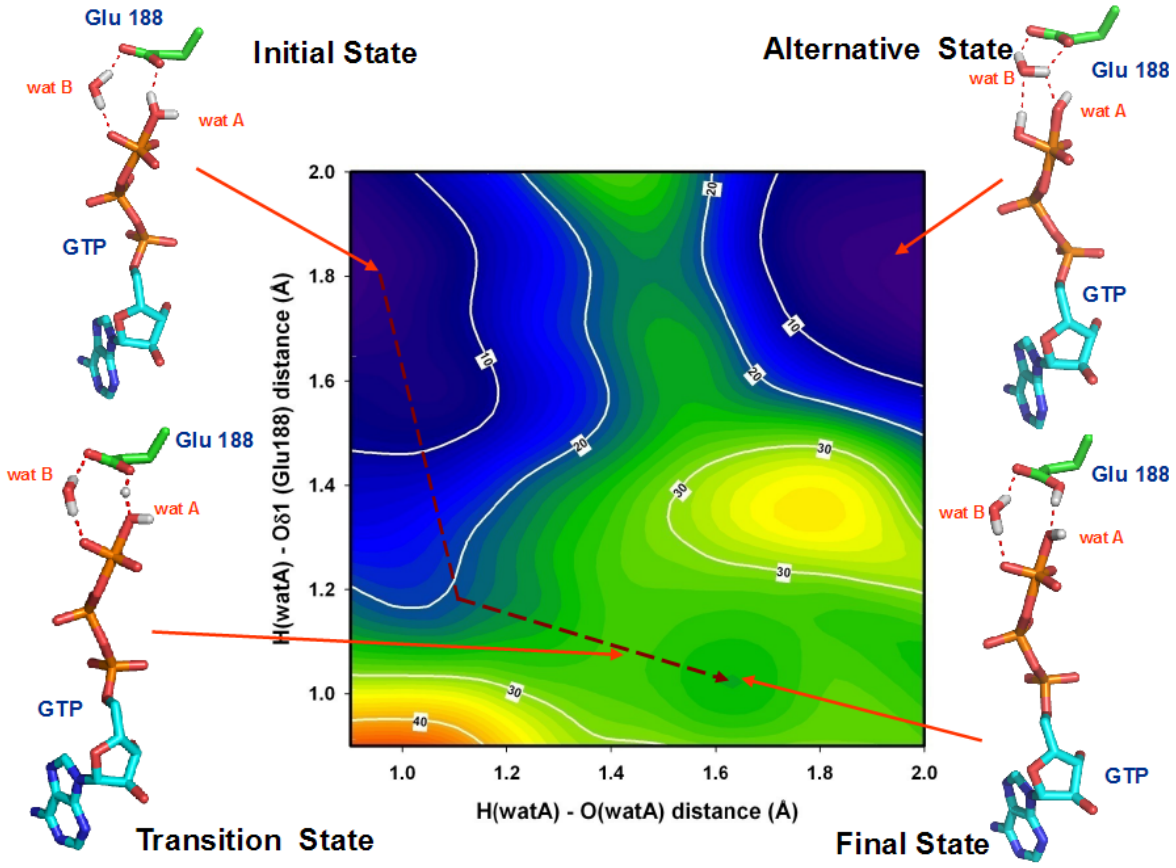

Figure 3.13:. Free energy landscape for the activation of the catalytic water molecule assuming β-Glu 188 is the proton acceptor. ΔG° values obtained for the whole system are shown for the different states corresponding to the distances: x axis, from the proton to the oxygen atom in the active water molecule (wat A); y axis, from the same proton to Oδ1 of β-Glu 188. The structure of the active center is depicted in the initial, transition, and two alternative final states. In all cases, the γ-phosphate group of ATP exhibited a pentacovalent conformation.

reaction proceeded through a high energy barrier (around 28 kcal·mol^{-1}) with the ΔG_0 value obtained for the protonation of the glutamate residue being around 25 kcal·mol^{-1}. This value is compatible with a low pKa characteristic of this acid residue in a polar environment. Both the high energy barrier and high ΔG_0, suggest that β-Glu 188 is not a good candidate to act as the general base. In general, glutamic acid residues require a very hydrophobic environment in order to work as proton acceptors, as in such conditions the electrostatic interactions are stronger than in polar environments and therefore the pKa become higher. The catalytic site of F1-ATPase contains five negative charges (ATP4- and β-Glu 188), five positive charges (β-Lys162, β-Arg189, α-Arg373 and the Mg^{++} ion) and five water molecules: wat A, wat B and the three in the coordination sphere of the Mg^{++} ion, which also includes the β-Thr163. In such a polar environment, β-Glu 188 must exhibit a low pKa which makes it difficult for this residue to act as a catalytic base. In order to provide further support to this statement, an additional simulation of proton transference between a Glu residue and an ATP molecule through an hydronium intermediate was performed in solution, simulating a polar environment (figure 3.16).

As expected, the difference in ΔG_0 values of the system between the protonated states of the two molecules (4.5 kcal•mol^{-1}), which corresponds to 3.3 pKa units, is in the range of the experimentally measured difference of 3.0 - 3.3 pKa units between the two groups in solution. This result indicated that our simulation approach is accurate enough to capture the relative proton affinities of these two groups.

In the same free energy landscape (figure 3.13) a second local minimum is observed at the top right-hand corner. The structure of the active center in this region shows a deprotonated β-Glu 188 residue and the pentacovalent transition state in a protonated form, suggesting that the final proton acceptor during the water activation could be the nucleotide. Interestingly, comparison of the structures in the initial and final states indicates that the original location of the proton bound to ATP in the final state was not the catalytic water molecule (wat A) but the second water molecule (wat B) present at the active site. This suggests a two-water mechanism, similar to that previously described for Human Ras p21[53].

ATP as the general base.

Figure 3.14 shows the free energy landscape obtained for the activation of the catalytic water molecule assuming ATP is the final proton acceptor instead of β-Glu 188. In this case the reaction parameters were the distance between the proton and the oxygen of the catalytic water (wat A) and the distance between the proton and the oxygen of the second water molecule (wat B) present at the active site. Just as indicated above, the analysis of this trajectory indicated that the proton transfer from the catalytic water, wat A, to the nucleotide is not direct, but occurs via the second water, in a similar way to the Grotthuss mechanism proposed for proton diffusion[1], as described for Human Ras p21[53].

The reaction proceeds through an energy barrier of around 22 kcal·mol^{-1}, lower than the 28 kcal·mol^{-1} that corresponds to the barrier when β-Glu 188 was considered to be the proton acceptor (figure 3.13). Significantly, the ΔG_0 value obtained for the final state of this reaction is close to 0 kcal·mol^{-1}. The final structure shows a protonated pentacovalent transition state which is consistent with the high pKa (around 6.8) of ATP in polar environments[120]. This result suggests that the protonation of the ATP and the formation of the nucleophile OH^- ion from the catalytic water ismore favorable than β-Glu 188 acting as a general base in the polar environment of the catalytic site.

To analyze the behavior of the protons in detail throughout the process, the reaction along the minimum energy path between the initial and the final states was simulated (dashed line in figure 3.14). Monitoring the proton movements (a video of the entire process is provided as video 3) revealed that a proton from the catalytic water (wat A) is initially transferred to the additional water (wat B), thus forming a hydronium ion (H_3O^+). Then, a different proton from wat B is transferred to the ATP in the pentacovalent transition state. This behavior confirms a proton transfer path similar to the Grotthuss mechanism[1]

The conformation of the catalytic site at the saddle point, corresponding to the transition state of the water activation (figure 3.14), shows a strong interaction between the carboxylic group of β-Glu 188 and the hydronium ion formed when the additional water molecule (wat B) accepts the proton from the catalytic water molecule (wat A). This interaction stabilizes the transition state of the water activation, reducing the value of the

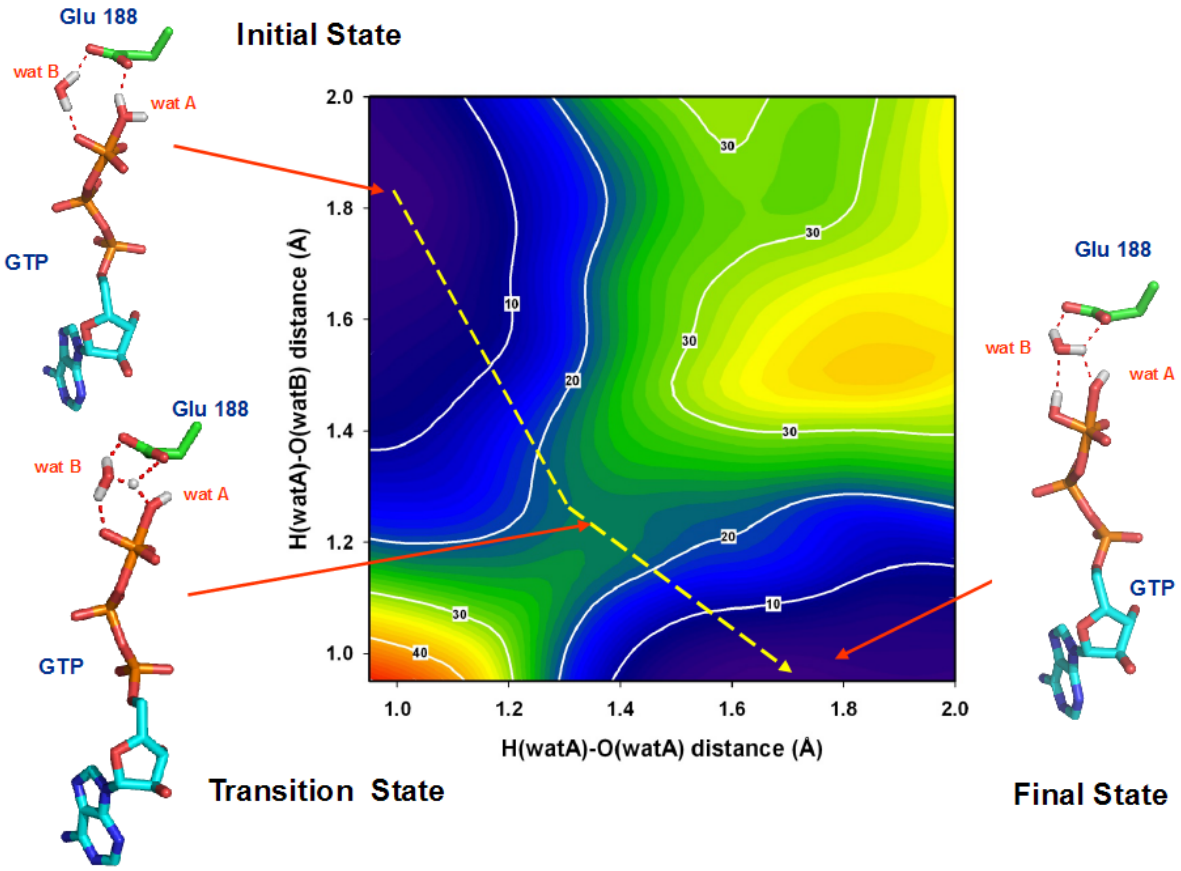

Figure 3.14: Free energy landscape for the proton transfer between the active water molecule (wat A) and the second water molecule (wat B). The axes represent the distances: x axis, from the proton to the oxygen atom in wat A; y axis, from the same proton to the oxygen atom in the second water molecule (wat B). β-Glu 188 stabilizes the position of the water molecules during proton transfer. The structure of the active center is depicted in the initial, transition, and final states. In the final state, a proton from wat B has been transferred to the ATP molecule.

energy barrier.

Non-carboxylate moiety in β-Glu 188 position

The results obtained in the simulations described above suggest a new role for β-Glu 188 which is the stabilization of the hydronium ion generated transiently during the proton transfer from the catalytic water molecule to ATP. This important role could explain the lack of activity produced by mutations of this residue[125–128], although it does not play a direct role as a catalytic base. It has been demonstrated that the presence of a carboxylate group at this position is definitely required for catalysis and its spatial positioning needs to be very precise: only when β-Glu 188 is replaced by residues containing a carboxylate group (Asp or S-carboxymethylcysteine) is a small but detectable degree of ATPase activity retained by the enzyme[127]. Nevertheless, and in order to get additional support for this hypothesis, we generated the free energy landscape for the same proton transfer path but simulating an inactive configuration in which the carboxylate moiety has been removed and the original Glu residue substituted by a neutral one (Ala). Figure 3.15 shows that the water activation in the absence of the carboxylic group of β-Glu 188 proceeds through a very high energy barrier of around 42 kcal·mol^{-1}, with the ΔG_0 value for the final state being close to 0 kcal·mol^{-1} as in the case of the wild-type protein. This result confirms that the neutralization of the transient hydronium ion by the negative charge of β-Glu 188 facilitates the formation of the nucleophile OH^- from the catalytic water, without altering the ΔG_0 of the reaction. A second local minimum observed in the top right-hand corner of Figure 3.15 suggests that, in the absence of the β-Glu 188 residue, a direct proton transfer from the catalytic water molecule to the nucleotide appears to be the most favorable pathway. However, this alternative proton transfer path must proceed through an energy barrier around 32 kcal·mol^{-1}, notably higher than the 22 kcal·mol^{-1} obtained for the sequential transfer of the proton involving the two water molecules in the wild-type protein.

Recently, experimental data based on single-molecule observations[123] have provided evidences that the rate-determining step of the reaction is the proton transference from the catalytic water, a statement that completely agrees with our results. In the same work, additional experimental results were interpreted as favorable to a "dissociative" reaction

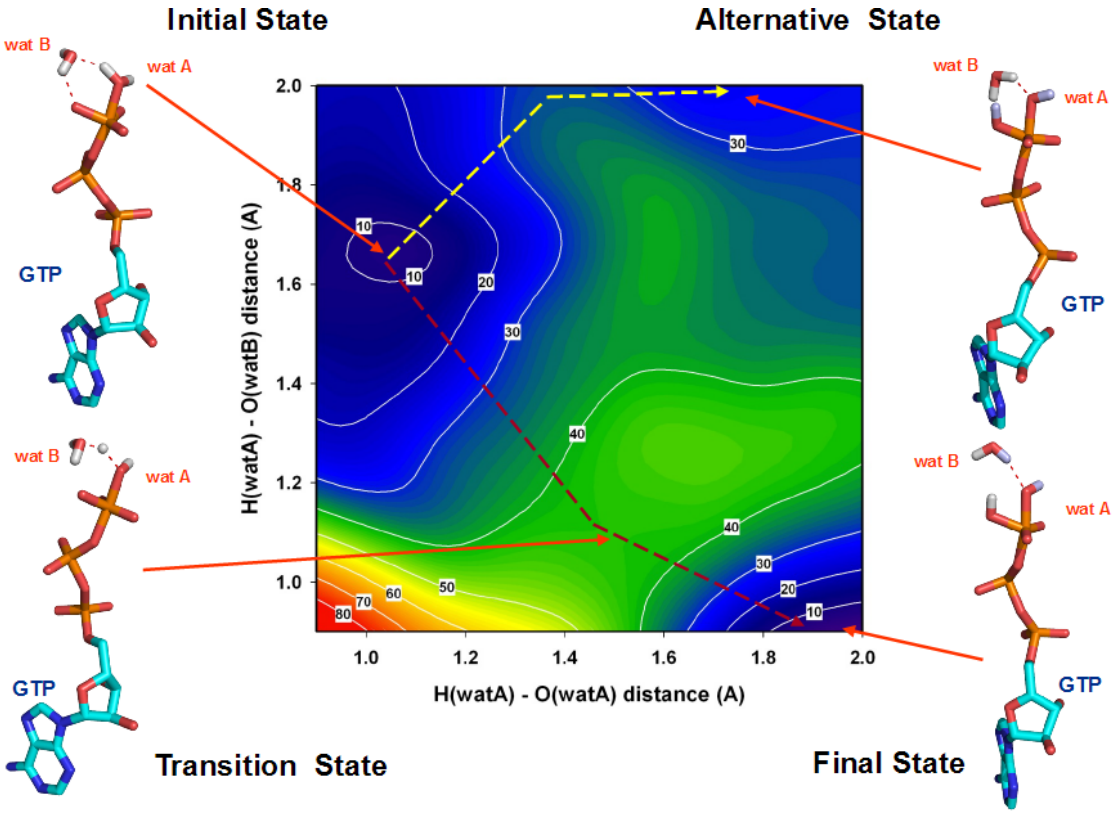

Figure 3.15: Free energy landscape in the absence of the carboxylate moiety in the β-Glu 188 position. Axes are defined as in Figure 4. The structure of the active center is depicted in the initial, transition, and final states. An alternative final state, located in an additional ΔG^0 minimum (top right), is also represented.

path where the P_γ-O_β bond of ATP was weakened before proton transfer event in a "non-conventional" S_N1 process, although the authors do not discard that both processes, the rate-determining step of proton transference and the weakening of the P_γ-O_β bond can occur in a concerted form.

In summary, our results are compatible with a mechanism for catalytic water activation of F1-ATPase involving two water molecules and a proton transfer from the catalytic water (wat A) to a second water (wat B) and then to the ATP molecule which acts as the proton acceptor. The important role of β-Glu 188 is the stabilization of the transient water structures during the proton transfer. This mechanism has been described using a new approach for energy surface calculations of QM/MM trajectories, allowing a detailed analysis of the conformational space defined by the reaction coordinates. Previous results obtained using the same methodology for Human Ras GTP hydrolysis[53] showed a similar mechanism involving two water molecules in the activation. Additional studies are in progress in order to analyze whether this mechanism could be a common solution for several GTPase and ATPase enzymes.

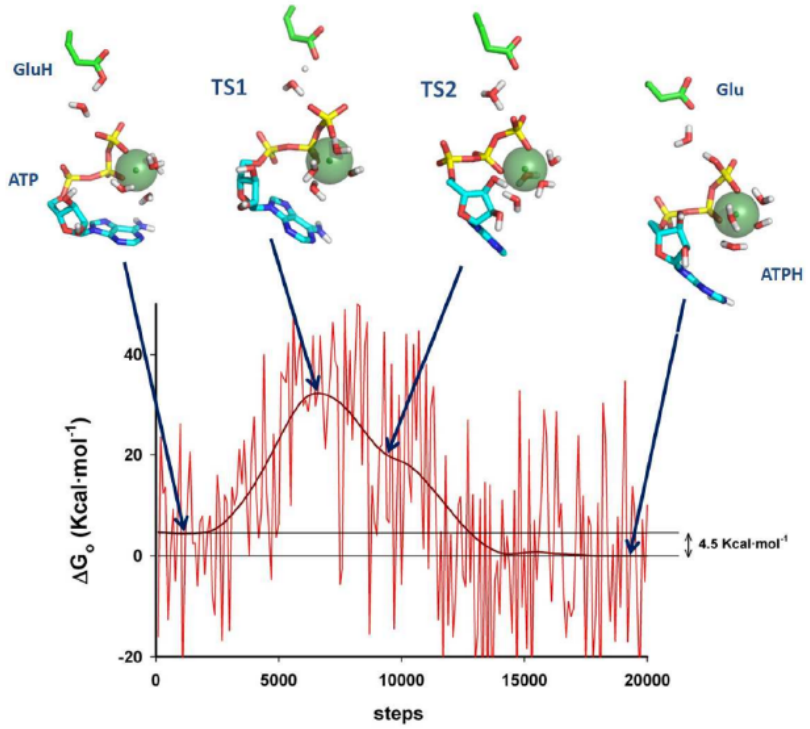

Figure 3.16: Simulation of proton transference between a Glu residue and an ATP molecule through a hydronium intermediate in solution. A minimal system was simulated in a water box containing a protonated Glu residue, the catalytic water and an ATP molecule. From the initial state (left), protonwas forced to move away from the initial GluH state until an energy maximum was reached (transition state TS1). From that position, the system moved spontaneously to a structure where the gamma phosphate of ATP is protonated (ATPH). The difference of ΔG^{o} values between the initial and the final states is 4.5 kcal•mol^{-1}, which corresponds to 3.3 pKa units, in the range of the experimentally measured difference of 3.0- 3.3 pKa units between the two groups in solution. In the same trajectory, the position of the system when the hydronium group is completely formed (TS2 in the figure) suggested that, from this structure, the system will move spontaneously (downhill) to the protonated state of ATP. On the contrary, from this same position, it is predicted that the system would not move easily to the protonated state of Glu residue due to the presence of the ΔG^{o} barrier of TS1.

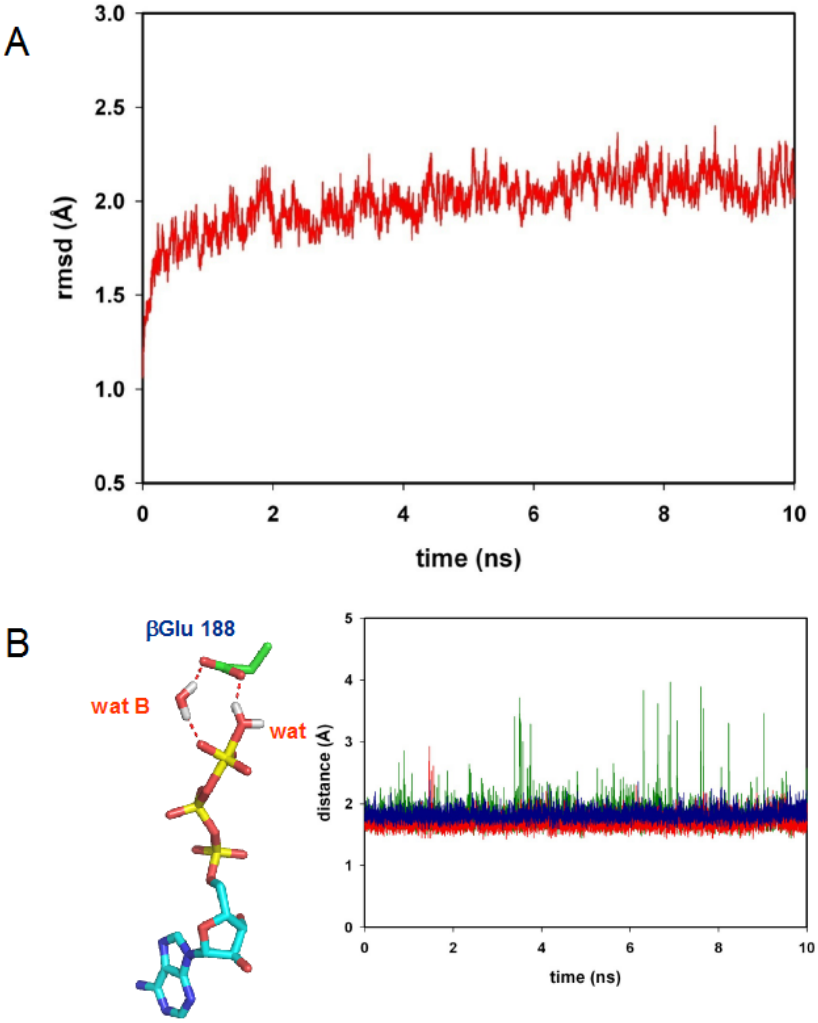

Figure 3.17: (A) Temporal profile of the Cα root-mean square deviation (rmsd) of the whole system during unconstrained MD simulation. (B) Continuous measurement of the hydrogen bond distances between the protons of the wat A and wat B and the oxygen atoms of the β-Glu 188 (red and blue lines, respectively) as well as of the protons of the wat B and the O3γ atom of the nucleotide (green line) during unconstrained MD simulation. Measured distances are graphically represented in the scheme of the left (dashed lines)

3.4. QM/MM study of catalytic water molecule activation in myosin

3.4.1. Abstract

In spite of myosin being a widely studied system and its activity being based on a well-known mechanism of ATP hydrolysis, there still exists a controversy regarding the residues involved in the catalytic water activation immediately prior to ATP hydrolysis. Different candidate residues have previously been proposed. Here, we investigate the reaction pathway of the activation of the catalytic water molecule using a previously described method to study the free energy landscape of the process using a quantum mechanics/molecular mechanics (QM/MM) approach**.** Our results suggest that the nucleophilic OH^- group is generated by the direct transfer of a proton from the catalytic water molecule to the nucleotide, and that the process is facilitated, to a certain extent, by a second water molecule.

3.4.2. Introduction

Myosin is, together with actin, one of the principal components of the system responsible for muscle contraction[44,45]. It is also involved in functions such as cell movement, cytokinesis, vesicle transport, Golgi organization and sensory transduction. The most common type of myosin is myosin II, which is present in muscle fibers. The molecular structure of myosin II can be subdivided into three domains: the N-terminal motor region or "head" (containing the actin binding site and the ATP binding site), a lever arm, and a coiled-coil tail[44]. In addition to myosin II, other types of myosin are found in cells other than muscle cells[129–133].

The motor domain region presents two main subdomains: U50 (the upper 50-kDa domain); and L50 (the lower 50-kDa domain). While outer regions of these two subdomains are involved in actin binding, inner residues are involved in γ-phosphate recognition of ATP. Switch I (U50), switch II (L50) and also the P-loop are the regions that are mostly responsible for this phosphate region recognition of ATP. Binding to ATP stabilizes the closed structure of the U50/L50 cleft. Conformational changes in myosin are dependent on the hydrolysis of ATP, which results in the formation of ADP and Pi as the end products of the reaction. As for all hydrolysis reactions, the presence of a catalytic water molecule, which needs to be activated thus generating an OH^- group, is necessary in the active center to drive the catalytic event. In associative-type (S_N2) mechanisms of hydrolysis, the OH^- group attacks the γ-phosphate of the nucleotide, generating the trigonal bipyramid geometry characteristic of the pentacovalent transition state. Due to the high pKa values of water, the activation of a catalytic water molecule must also be catalyzed by an additional charged group: the catalytic base. The nature of this group is, in the case of myosin, still controversial. One of the residues that has been proposed as the general base is E459[71,73], which forms a salt bridge with R238. Through mutational analysis[134], these residues have been shown to be essential for ATP binding and hydrolysis. E459 is hydrogen-bonded with a second water molecule[135] located near the catalytic water molecule. According to this hypothesis, deprotonation of the catalytic water molecule is performed through this secondary water molecule, generating an H_3O^+ molecule. The excess of charge leads to the transfer of a proton to E459. This reaction pathway generates a highly reactant OH^- molecule that will attack the γ-phosphate of ATP. The product of that reaction would be an HPO_4^{-2} molecule. The final step in the reaction would be the transfer of a proton from E459 to HPO_4^{-2}, generating an $H_2PO_4^-$ molecule, which, together with ADP, is the final product of the hydrolysis.

The second candidate proposed, using DFT approximations[136], as the catalytic base in the activation of the catalytic water molecule is the S236 residue. The hydroxyl group in the side chain of this residue, which is located in the switch I region, establishes a hydrogen bond with ATP through the O2γ atom. In this proposed mechanism, the formation of the

nucleophilic group takes place through indirect proton transfer from the water molecule to the nucleotide using the hydroxyl group of S236 as a relay group.

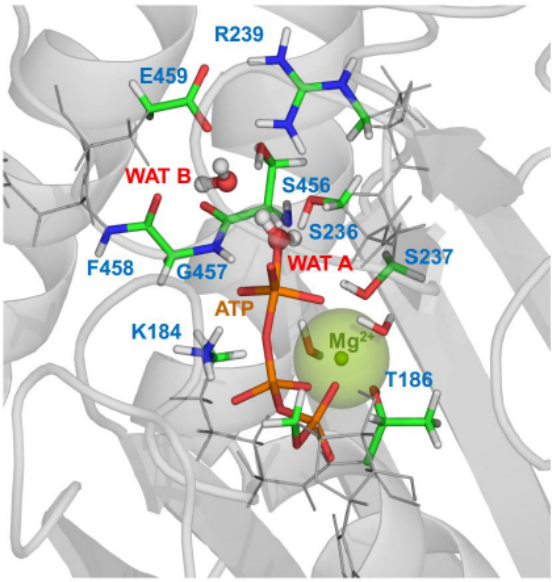

Figure 3.18: Myosin catalytic site. The atoms involved in the QM region are represented as colored sticks. The system is based on the myosin post-rigor structure (PDB: 1VOM). The residues included were: K185, T186, S236, S237, R238, S456, G457, F458 and E459. The phosphate region and the magnesium ion (transparent green sphere) were also included in the QM part. The catalytic (wat A) and secondary (wat B) water molecules are shown in ball-and-stick representation and they were also included in the QM part.

Computer simulations using a QM/MM interface offer the possibility of including reaction atoms in the quantum mechanics (QM) part, at the same time as the whole protein system is simulated in the molecular mechanics (MM) part which have an important role in the behavior of the QM part. QM/MM combined with an approach described previously that is based on adaptively biased MD[53,54] has been shown to be useful in obtaining a reliable free energy surface in the conformational space defined by the reaction coordinates for the water activation. In the present work, using that same methodology, we initially studied the two previous hypotheses explained above for the myosin system, as well as a third hypothesis. The additional hypothesis shares a scheme similar to that of the one which

considers E459 as the catalytic base, involving the second water molecule. The difference is that the final proton acceptor in the system would be the substrate itself, the $H_2PO_4^-$ molecule, without an intermediary HPO_4^{-2}. The results led us to test a fourth additional hypothesis involving direct protonation with ATP as the general base.

3.4.3. Material & methods

Molecular dynamics (MD) simulations

The coordinates of the atoms were obtained from the crystal structure with PDB code 1VOM[135]. As this initial structure was incomplete between residues Q204 and G209, which corresponds to a short loop far from the active center, the position of these residues was determined using the information present in the crystal structure with PDB code: 2JHR[137]. The protonation state of all the histidines in the molecule were examined. As the initial 1VOM crystal structure was resolved using magnesium(II)·ADP·vanadate[135], the vanadium atom was replaced by a phosphorous atom in order to obtain a complete ATP molecule with the original coordinates. One of the oxygen atoms of the vanadate molecule was replaced by the oxygen in the catalytic water molecule (wat A).

The protein was solvated with a 12 Å water box (water model: TIP3[138]) from the edge of the box to the closest atom in the protein. The water box was built using the LEaP program from the AMBER11 package[83]. Periodic boundary conditions were established in order to represent a continuous solvent environment. The system was neutralized with K^+ ions using a grid of 1 Å to determine the charges in the system to be neutralized.

The whole system was exposed to the AMBER ff99SB force field[62] over the minimization period. The minimization was performed over 10,000 steps, with a combination of steepest-descent and conjugate-gradient methods. During this period, the transition state of the ATP was modified from the initial distances established in the ADP·vanadate. Attack (O_{watA} - Pγ) and rupture (Pγ - O3β) distances were reduced from 2.2 and 2.1 Å (respectively) to 1.9 Å. Once the system was minimized, an equilibration period was executed, increasing the temperature from 0K to 300K over 10,000 computational steps with a δt of 2 fs, in order to simulate physiological conditions. To avoid denaturing the system, the Cα trace was

constrained by maintaining the dihedral values of the trace. An additional restriction was used in order to maintain the structure of the catalytic center. These restrictions were applied to the transition state and the salt bridge formed by E459 and R238. The distance between the catalytic water molecule (wat A) and secondary water molecule (W1696 in 1VOM[135]; wat B herein) was maintained at the crystal values throughout the heating phase.

After the equilibration phase, the system was exposed to the ff99SB force field without any restraints. This action was performed during the stabilization phase, during which the force constant of the Cα trace was reduced to 0. Additional constrains were removed, with only the transition state constraint being maintained, in order to analyze whether the system was stable in this conformation. Productive MD was executed for of 18 ns with only this restriction. The SHAKE algorithm was applied to the Particle Mesh Ewald MD calculation. Analysis of the geometry and energy of the system were performed using the ptraj program[83].

Quantum Mechanics/Molecular Mechanics (QM/MM) simulations

The QM region was defined to include the ATP atoms, magnesium, the two magnesium coordination sphere water molecules, the two catalytic water molecules, and the residues K185, T186, S236, S237, R238, S456, G457, F458 and E459. The residues were cut through their non-polar bonds. The atoms that formed those non-polar bonds were replaced by hydrogen atoms. The number of atoms thus included in the QM part was 101 (figure 3.18, colored atoms). The rest of the system was included in the MM calculation. A 300 ps simulation was carried out to adapt the QM part to PM3 Hamiltonian conditions. The sander program was used in this case. The SHAKE algorithm was not applied to the MM calculation and the δt used was 1 fs. During the adaptation phase, active site distance restraints were applied for 250 ps and removed for the last 50 ps, with only the transition state and ion bridge between E459 and R238 being maintained. For the QM/MM electrostatic interactions, a cut-off of 8 Å was used.

As previously described[53,54], this approach combines the steered MD[102] and umbrella sampling[103] methods. Energy surface maps were constructed with trajectories performed by

fixing one of the reaction coordinates (the bond-forming distance), and with a narrow flat-bottomed harmonic potential. The other reaction coordinate (the bond-breaking distance, r) was varied with a δr of 0.025 Å, with an overlap between adjacent trajectories (umbrella sampling). This method allowed us to explore the whole conformational space defined by the reaction coordinates chosen. In order to highlight the effect of the residues on the reaction, only the H_{QM} and $H_{QM/MM}$ energies were taken into account in the construction of the energy maps. 3D smoothing of the data was applied using the local smoothing technique with tricube weighting and polynomial regression (LOESS). The ground state energy level was set at the energy level of the reaction initial state for the calculation of ΔG_0 and ΔG_{TS} for all the maps.

3.4.4. Results & discussion

Using the approach described previously to generate free energy surface maps based on QM/MM simulations[53,54], different hypotheses concerning the mechanism of catalytic water activation in myosin were tested. As an initial step, a detailed analysis of the structure and energy of the whole system was performed during preliminary productive MD and QM/MM stabilization phases. Table 3.2 shows the results for the most important distances measured. In brief, a stable RMSD value (under the resolution of the original crystal structure: 1.9 Å) and invariable total energy were found to hold throughout the 18 ns of the MD stabilization.

After productive MD, the system was studied under QM/MM conditions (see methods and figure 3.18). To test the system, the same distances as those in table 3.2 were measured. Our results indicate that there were no significant variations in the values.

Three different hypotheses for catalytic water activation were then initially tested: water-assisted activation (mainly the same mechanism as previously described for HRas and F1-ATPase[53,54]); and activation with either residue E459[134,139,140] or residue S236[136] as the proton acceptor. A fourth hypothesis, proposed after the initial results, which consisted of direct protonation with ATP as the general base, was also tested.

Residue atom pair	Distance value (Å)
O_{watA} - $P\gamma_{ATP}$	1.92 ± 0.02
$O3\beta_{ATP}$ - $P\gamma_{ATP}$	1.93 ± 0.02
O_{watA} – O_{watB}	3.18 ± 0.17
O_{waB} - $C\delta_{E459}$	3.86 ± 0.25
O_{watA} - O_{G457}	5.14 ± 0.29
$H\gamma_{S236}$ - $O2\gamma_{ATP}$	2.78 ± 0.24
$O\varepsilon1_{E458}$ - $NH1_{R238}$	2.76 ± 0.16
$O\varepsilon2_{E458}$ - $NH2_{R238}$	2.81 ± 0.11
N_{G457} - $O3\gamma_{ATP}$	2.86 ± 0.12
NZ_{K185} - $O3\gamma_{ATP}$	2.68 ± 0.07
NZ_{K185} - $O3\beta_{ATP}$	2.71 ± 0.08

Table 3.2: Distances ± SD measured over 18 ns of productive MD. Low values of the standard deviation indicate that the system remained stable over the whole simulation time.

Water-assisted activation

An initial study was performed based on the idea that a second water molecule (wat B) acts as the acceptor of the proton from wat A, as previously described for HRas and F1-ATPase[53,54]. Distance measurements showed that the wat B molecule was stable during the initial 18 ns of MD and QM stabilization and that it formed two hydrogen bonds: one with the Oε1 of the E459 acetic group and the other with the oxygen in the main chain of G457. Both residues are located in the switch II region of myosin. The reaction coordinates used to build the potential surface map (figure 3.19), were the distances between O_{watA} and H_{watA} (the bond-breaking distance), and between H_{watA} and O_{watB} (the bond-forming distance). No restrictions were applied to the protons that took part in the reaction, so they could be freely transferred to the reaction base, taking into account that H_3O^+ cannot be the final base in the reaction.

Figure 3.19 shows the potential energy of the conformational space corresponding to the reaction coordinates. The hypothetical reaction pathway (dashed red line) shows it taking place via wat B.

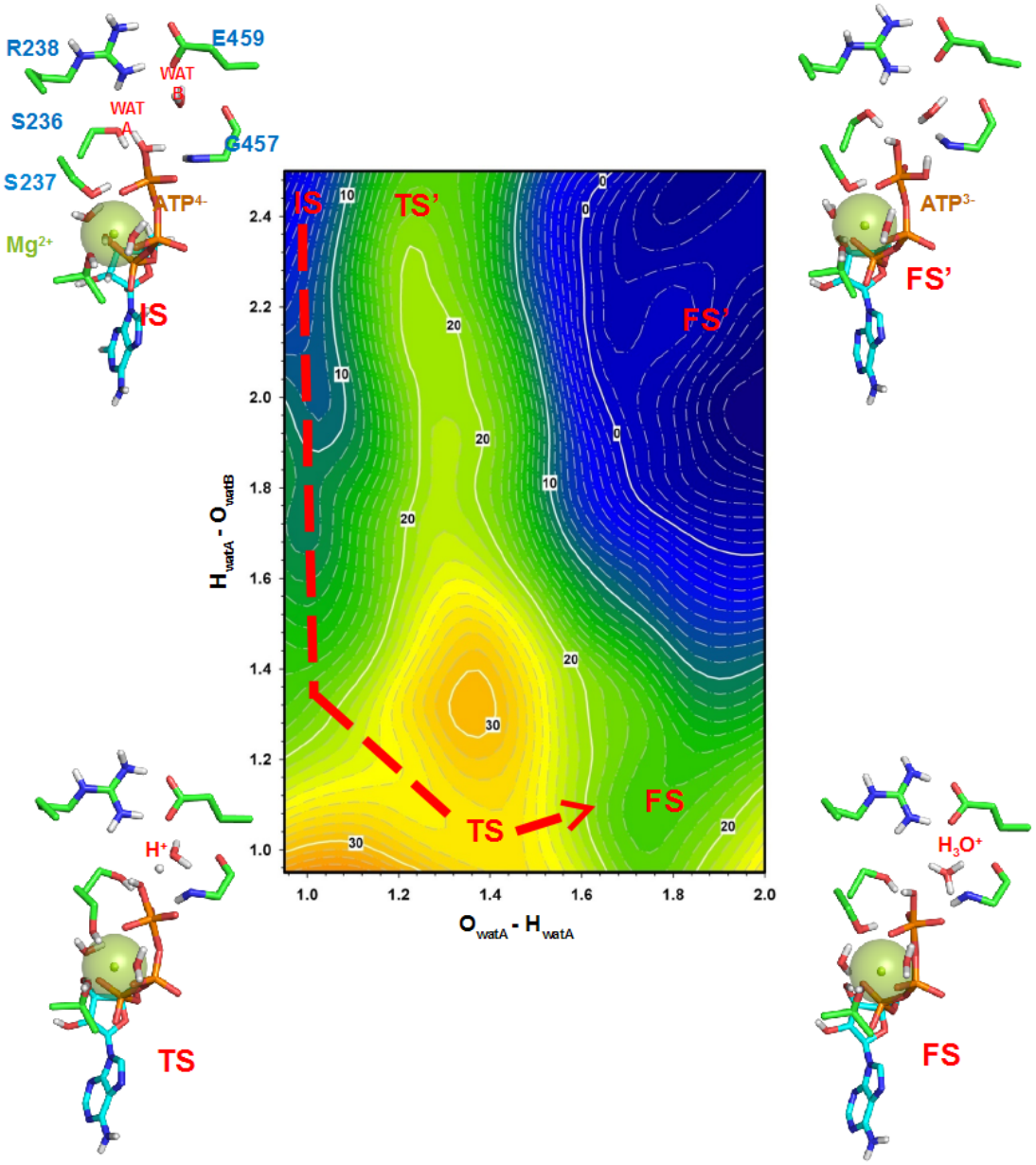

Figure 3.19: Schematic representation of catalytic water molecule activation through wat B. The y-axis represents the distance between the proton of wat A and the oxygen atom of wat B (the bond-forming distance). The x-axis represents the distance between the wat A proton and the oxygen atom of wat A (the bond-breaking distance). The ground state is established in IS. Blue regions represent more stable structures on the map. The color scheme for the surface maps in Figures 3.19, 3.20, 3.21 and 3.22 is the same. The free energy between IS and FS is: $\Delta G_0 \approx 16$ kcal·mol^{-1}. The reaction path (dashed red line) energy barrier is around 26 kcal·mol^{-1}.

The energy surface reveals that the reaction passes through an energy barrier (TS) of around 26 kcal·mol^{-1}, with the final reaction energy, ΔG_0, being ≈ 16 kcal·mol^{-1}. Inspection of the structure of the final state (FS) indicated that a nucleophilic OH^- group had been formed by the transfer of a proton from the catalytic water molecule (wat A) to the wat B molecule, forming an H_3O^+ ion, which remained stable. None of the trajectories showed proton transfer from the H_3O^+ to any subsequent acceptor (E459, S236 or ATP), all of which remained in their unprotonated states. A closer look at wat B revealed the presence of two hydrogen bonds: one with the Oε1 of the E459 acetic group and the other with the oxygen in the main chain of G457; as had been observed during stabilization. The observed O_{watB} – O3γ FS distance was too long for proton transfer from wat B to ATP, indicating that this situation is not a suitable initial step for ATP protonation.

Surprisingly, the same map revealed an alternative path (figure 3.19, labeled as TS' and FS'), with the transition state (TS') located at 2.5 Å for the $H_{watA} - O_{watB}$ coordinate. The value of the TS' energy barrier is 17 kcal·mol^{-1}, lower than the value of TS in the reaction path initially considered. This alternative path shows ΔG_0 ≈ -8 kcal·mol^{-1} at the deepest point of the top-right region, which indicates a more stable FS' structure than FS. Structural analysis of FS' (figure 3.19, upper right) showed a protonated ATP molecule. Trajectory analysis of the alternative path revealed that the origin of that proton was wat A, which appeared in the OH^- state. As this path would indicate direct proton transfer, with ATP acting as the general base, this new hypothesis was also tested

E459 as the general base

A range of studies have shown that E459 plays a critical role in myosin activity[134,139,140]. The residue, located in switch II, forms a salt bridge with R238, in switch I, thereby closing the cleft between the U50 and L50 subunits[43] and maintaining the catalytic center unexposed to the water solvent. Recent studies[71,73] performed using DFT (B3LYP and BB1K) revealed that E459 acts as the base for the reaction. Protonation of E459 occurs via wat B, which accepts a wat A proton. At this point, wat B transfers one of its protons to one of the oxygen atoms of the E459 acetic group. Taking into account the results indicated above (figure 3.19), this proton transfer appeared unfavorable due to the high energy barrier.

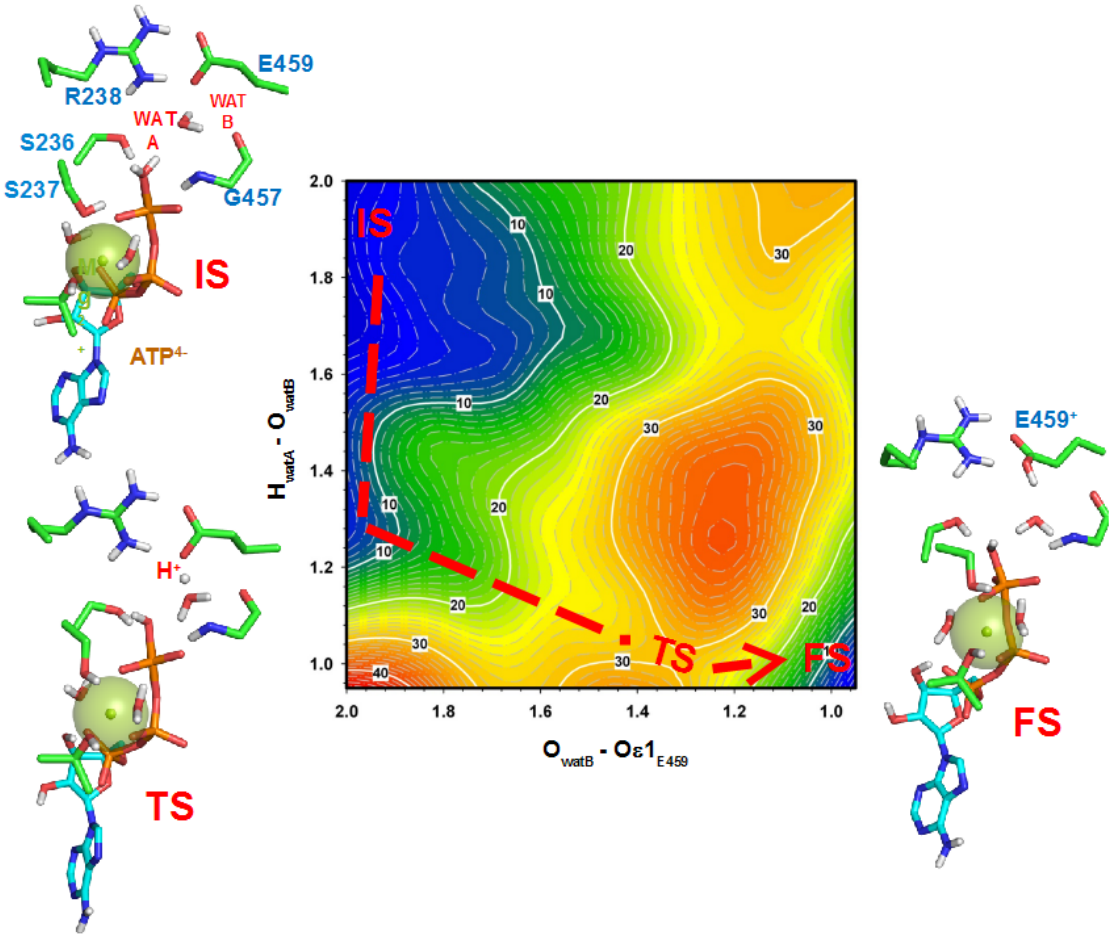

Figure 3.20: Potential energy map of E459 as the general base in the activation of the catalytic water molecule. On the x-axis, the bond-forming distance between the wat B proton and Oε1 of E459 is represented. The second bond-forming distance between the proton of wat A and the oxygen atom of wat B is represented on the y-axis. The dashed red line represents the estimated reaction path. The value of the energy barrier is around 29 kcal·mol^{-1} and the free energy is around: $\Delta G_0 \approx 0$ kcal·mol^{-1}.

Nevertheless, in order to check whether the coordinates selected previously (figure 3.19) were unable to capture this final state, we studied the reaction using $H_{watA} - O_{watB}$ as the new bond-forming coordinate and $H_{watB} - O\varepsilon1_{E459}$ as a second new bond-forming coordinate (figure 3.20). These coordinates fix the reaction path to establish E459 as the final acceptor of the proton.

The top left of figure 3.20 shows a local minimum corresponding to the initial structure of the system, with a bonding distance of 2 Å. The dashed red line shows the reaction path. During this reaction, wat A transfers its proton to wat B and E459 accepts the wat B proton. Analysis of FS showed E459 protonated at Oε1. Via this path, the reaction needs to overcome a barrier of around 29 kcal·mol^{-1}. The free energy of the reaction is around 0 kcal·mol^{-1}. This result shows an equipotential reaction path, which implies a stable structure at the end of the reaction. However, this value is higher than the TS' value in figure 3.19 (around 12 kcal·mol^{-1}) which makes this hypothesis less favorable. In accordance with our results, proton transfer though an intermediate water molecule (wat B) seems to be unfavorable, whether the final proton acceptor is ATP or E459.

In addition, a prediction of the pKa of E459 was made using the PROPKA server[141]. The resulting value, pKa=4.89, makes it difficult to consider E459 as the proton acceptor and it is consistent with pKa values of aspartic acid in highly polarized environments such as the active center of myosin: four negative charges on ATP, one positive on R238, two positives on magnesium, one positive on K185 and the presence of several water molecules.

Direct protonation with ATP as the general base

According to the results obtained in figure 3.19, a direct protonation reaction to ATP was cosidered once E456 was ruled out as base of the reaction. As described above, figure 3.19 shows an alternative reaction path with $\Delta G_{TS} \approx 17$ kcal·mol^{-1} and $\Delta G_0 \approx -8$ kcal·mol^{-1}. Analysis of the putative trajectories in that region showed the transfer of H_{watA} to the O3γ of ATP; that is, a direct substrate-assisted catalysis reaction. In order to elucidate this alternative route, we studied the potential energy surface of the reaction. The bond-breaking coordinate was defined as the $H_{watA} - O_{watA}$ bond distance and the bond-forming coordinate

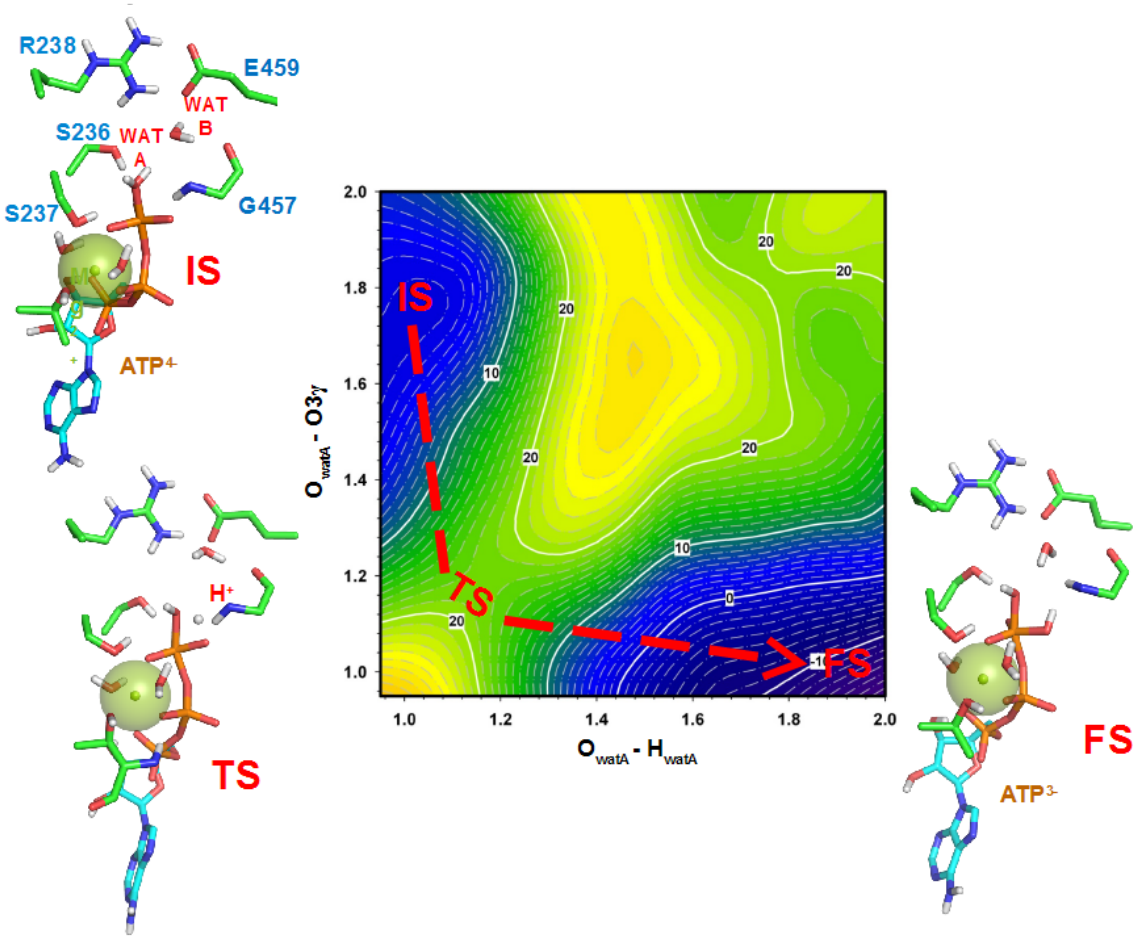

Figure 3.21: Potential energy surface of direct protonation of ATP with wat A proton. The distance between the oxygen of wat A and its own proton involved in the transfer (the bond-breaking distance) is represented on the horizontal axis while the bond-forming distance between the wat A proton and the O3γ of ATP is represented on the vertical axis. The value of the energy barrier is around 18 kcal·mol^{-1} and the free energy is: $\Delta G_0 \approx$ -17 kcal·mol^{-1}. The dashed red line represents the most probable reaction path.

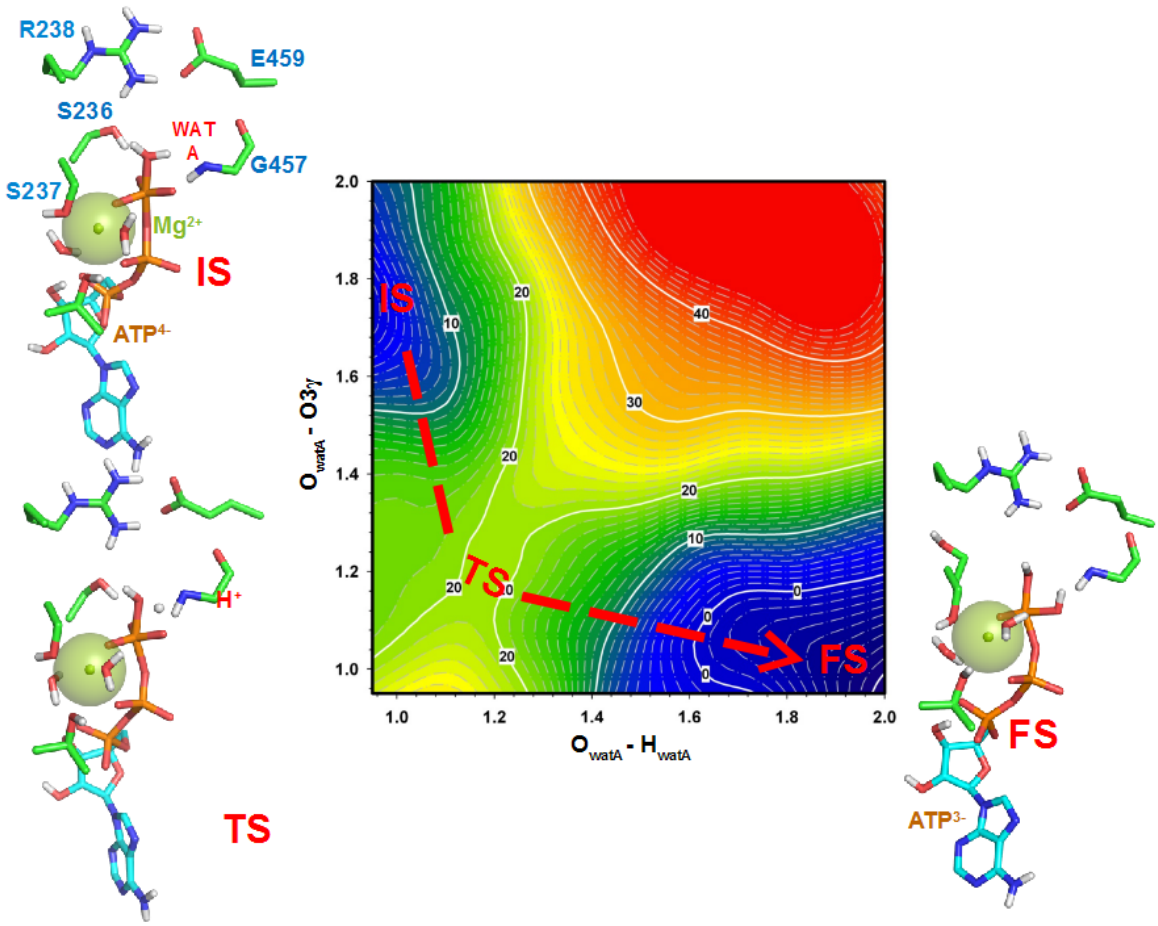

Figure 3.22: Energy map of direct reaction in absence of wat B. The coordinates are the same as in A. Comparison of the two maps reveals a slight effect of wat B on the energy barrier of the transition state (TS). The dashed red line represents the estimated reaction path.

was established as H_{watA} – O3γ.

Figure 3.21 shows the potential energy surface of the direct reaction. On this map, a single path is observed. Analysis of the structures along this reaction path shows the transfer of H_{watA} to the O3γ of ATP through an energy barrier of around 18 kcal·mol^{-1}, a very similar value to that obtained in TS' of figure 3.19 (17 kcal·mol^{-1}). Moreover, ΔG_0 is around -17 kcal·mol^{-1}, which implies a stable final structure (FS; also very similar to FS' in figure 3.19) which corresponds to the protonated ATP molecule. These results lead us to conclude that activation of the catalytic water molecule by direct protonation of the substrate must be considered as the most plausible hypothesis in myosin ATP hydrolysis.

The question regarding the putative role of wat B in the process remained unanswered. In the transition state (TS) of figure 3.21, the distance between H_{watA} and O_{watB} is around 2.5 Å, which also agrees with the values obtained in TS' (Figure 3.19). This data could suggest that wat B is not directly involved in the activation of wat A, but in proton stabilization during the transfer. To evaluate this putative role, we performed a direct reaction in the absence of wat B, using the same reaction coordinates as in Figure 3.21. The resulting potential surface map is shown in Figure 3.22. A comparison of the TS in the two reactions shows a small increment of 2 kcal·mol^{-1} in the absence of wat B, which is consistent with a putative structural role of wat B in the geometry of the catalytic center, similar to that we proposed for S236.

S236 as the relay group

S236 has also been proposed as the proton acceptor[136] in the activation of the catalytic water molecule (wat A). In this reaction path, the S236 side chain would act as a relay group for the proton transfer from the catalytic water molecule to the O2γ of ATP (substrate-assisted catalysis). In order to check whether S236 could act as the proton acceptor, we analyzed the system using H_{watA} – O_{watA} as the bond -breaking coordinate and H_{watA} – O_{S236} as the bond-forming coordinate.

Figure 3.23 shows the potential energy surface using these coordinates. The dashed red line shows the reaction path through S236. The energy barrier (TS) on this path is around 27

kcal·mol^{-1}, similar to the values in the two hypotheses tested above, although in this case ΔG_0 is around -11 kcal·mol^{-1}. The FS structure (figure 3.23, bottom right) reveals that the proton transfer was directed to the O2γ of ATP. Analysis of the trajectory also indicated that proton transfer occurred through S236. As H_{watA} nears O_{S236} of the S236 hydroxyl group, the bond formed in the hydroxyl group weakens thus causing H_{S236} (which forms a stable hydrogen bond) to be transferred to the O2γ of ATP.

In spite of the favorable ΔG_0 value, the high TS energy barrier would rule out S236 as a putative acceptor of the catalytic water proton. Additionally, experimental mutation of S236 to alanine revealed that the ATPase activity is maintained in the mutant S236A protein[142,143].

An alternative path is observed in this map. This alternative route pass through a lower energy barrier of around 9 kcal·mol^{-1}, the lowest seen in all the maps in this study. A possible reaction pathway is difficult to define. However, considering a hypothetical route, the end of the reaction would be the direct protonation of the ATP in O2γ. (top right of figure 3.23). Considering the charge of O2γ and O3γ, the former present a higher negative charge, what could explain that the barrier be lower respect to the protonation of the later. However, taking into account that transition state is maintained, this reaction pathway would present some problems from the structural point of view.

At this point, a new method based on DFT, is being used in our lab. This new approximation permits us to study the whole reaction, so reaction coordinates in this study would change. In this way, bonding coordinate would be the attack of catalytic water molecule to γ-P, while breaking distance would be the bond to the O3β.

Preliminary results reveal a water-assisted catalysis combined with a direct protonation to O2γ. This reaction would be facilitated by the rupture of the breaking coordinate, what would approximate O2γ to second water molecule, favoring the transference. This result was impossible to see with PM3 since transition state should be maintained. Additional experiments are being performed in order to solve the reaction.

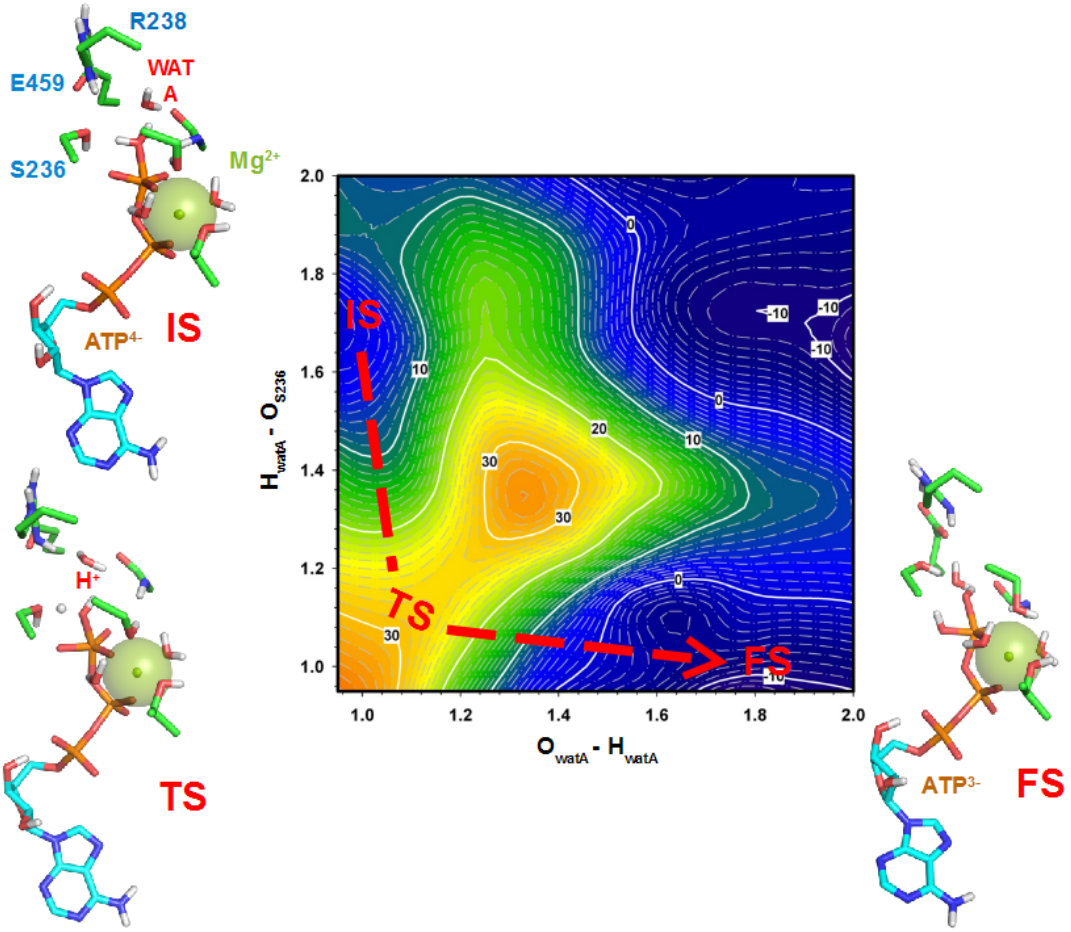

Figure 3.23: Potential energy surface with S236 as intermediary of activation of catalytic water molecule. The distance between the oxygen of wat A and the reactive proton is represented on the horizontal axis. The vertical axis represents the bonding distance between the wat A proton and the oxygen atom of the S236 hydroxyl group. The dashed red line represents the estimated reaction path, as in the rest of the maps.

4. Discussion and conclusions

One of the principal aims of structural biology is the study in the correlation between structure and function. A good description of the molecular mechanism in biomolecular systems provides very useful information in order to discover new alternative treatments for diseases. The combination of multiple methods of study has been essential for the last decades since a single approximation cannot afford the complexities of the biological systems by itself.

To this end, computational methods are incorporating atomic resolution in the biology structural field. In this way, MD, docking, principal component analysis, etc., offer a battery of analytic tools in this way. New computational methods based on QM offer the possibility of study chemical reactions in order to characterize enzymatic mechanisms.

Considering these ideas and the results obtained during the work of this thesis, we can conclude that:

1. We have demonstrated that combination of QM/MD interface with BMD techniques such as SMD and US, provides a broad description of reaction mechanisms, showing the different reaction pathways defined by the reaction coordinates. The high density of structures are sampled thanks to the combination of SMD and US, providing strength to the method. Combination of QM with MD instead of MM permits to the system to access to a higher number of conformations during reaction. For the evaluation of the method, we performed the ionization reaction of a water molecule. The results obtained fitted with the experimental results.
2. The method presented here has permitted us to study the role of Gln61 in the HRas-p21 protein. We have demonstrated the stabilizer role of this residue of the transition state during the activation of the catalytic water molecule, discarding the

possible role attributed as base of the reaction. In order to support this result, we created the mutant Q61A. The result obtained was an increase of the activation barrier respect to the wild type protein. Additionally, a Grotthuss like mechanism has been proposed, where a second water molecule would take part in the activation of the catalytic water molecule.

3. A study about the role of Glu188 in F_oF_1 ATP synthase was performed with our method. As in Hras-p21, Glu188 presents a structural role of the transition state of the reaction, decreasing the activation barrier. The rise of the activation barrier in the mutant E188A confirmed the hypothesis. A Grotthuss like mechanism was described, where a second water molecule plays an important role in the activation of the catalytic water molecule.

4. An evaluation of the different reaction mechanisms proposed in literature was performed in the myosin protein. An evaluation of the different hypothesis revealed a direct protonation of ATP, assisted by the presence of a second water molecule in the catalytic site, as the most probable mechanism. E456 and S236 were discarded as bases of the reaction during the activation of the catalytic water molecule. A closer analysis reveals a structural role in the stabilization of the catalytic site. New studies, where the method is combined with a deeper theory level based on DFT, are supporting these hypothesis.

5. We have suggested that the interfaces of FtsZ polymers present a different behavior depending on their position in the filament. The study of the catalytic sites in different polymers in polymerization and depolymerization conditions has suggested this hypothesis about FtsZ filament behavior. A higher probability of GTPase activity is proposed which depends on the position of the catalytic site in the filament, being higher in the extremes of the filaments. This model agrees with *in vitro* and AFM results.

6. In accordance with the new results obtained with the new theory level based on DFT, an application of the method on FtsZ is necessary to obtain a complete model

of the protein. A good description of the catalytic site and how the protein hydrolyzes the nucleotide would help to the understanding of the mechanism of this protein in bacterial division and obtain a more detailed model.

5. Discusión y conclusiones

Uno de los los principales objetivos de la biología estructural es el estudio de la correlación entre estructura y función. Una buena descripción de los mecanismos moleculares en sistemas biológicos proporciona información útil para el descubrimiento de nuevas alternativas en la cura de enfermedades. La combinación de múltiples métodos de estudio ha sido esencial en las últimas décadas, ya que dada la complejidad de los sistemas biológicos, estos no pueden ser abordados desde un punto de vista único.

De acuerdo con esto, los métodos computacionales están incorporando una resolución atómica en el campo de la biología estructural. La dinámica molecular, las técnicas de docking, el análisis de componentes principales, etc., ofrecen una batería de análisis en este sentido. Nuevos métodos computacionales basados en mecánica cuántica ofrecen la posibilidad de estudiar reacciones químicas con el fin de caracterizar los mecanismos enzimáticos.

Considerando estos aspectos y los resultados obtenidos en esta tesis, podemos concluir que:

1. Hemos demostrado que la combinación de la interfaz de QM/MD con las técnicas de BMD, tales como SMD y US, proporciona una amplia descripción de los mecanismos de reacción, mostrando distintas rutas de reacción definidas por las coordenadas elegidas. La alta densidad de estructuras conseguida gracias a la combinación de SMD y US proporciona una alta robustez al método. La combinación de la mecánica cuántica con la dinámica molecular (QM/MD), en vez del uso de la mecánica molecular (QM/MM), permite a los sistemas el acceso a un número mayor de conformaciones durante el proceso de reacción. La evaluación del valor de energía en la ionización de la molécula de agua confirmó la precisión del método.
2. El método aquí presentado nos ha permitido estudiar el papel de la Gln61 en la proteína HRas-p21. Hemos demostrado el papel estabilizador de este residuo del

estado de transición donde se genera una molécula con carga negativa (hidroxilo) y otra de carga positiva (hidrogenión), contrarrestadas por grupo amida de su cadena lateral, descartando el posible papel atribuido como base de la reacción. Con el fin de respaldar este resultado, creamos el mutante Q61A. El resultado obtenido fue un incremento de la barrera de activación respecto a la proteína nativa. Además, un mecanismo del tipo Grotthuss fue propuesto, donde la segunda molecula de agua juega un papel importante en la activación de la molécula de agua catalítica.

3. Un estudio acerca del papel del Glu188 de F_oF_1 ATP sintetasa se llevó a cabo por medio del método descrito. Como en el caso de HRas-p21, Glu188 presenta un papel estabilizador del estado de transición de la reacción, disminuyendo la barrera de activación. El incremento de la barrera de activación en el mutante E188A confirmó la hipótesis. Un mecanismo del tipo Grotthuss fue descrito, donde una segunda molécula de agua jugó un importante papel en la activación de la molécula de agua catalítica.

4. Se llevó a cabo una evaluación de los diferentes mecanismos de reacción en la proteína miosina. La evaluación de las diferentes hipótesis mostró una protonación directa a la molecular de ATP, asistida por la presencia de una segunda molécula de agua, situada en el sitio catalítico, como el mecanismo más probable. E456 y S236 fueron descartados como bases de la reacción durante la activación de la molécula de agua catalítica. Un análisis más detallado mostró una función estructural en la estabilización del centro catalítico. Nuevos estudios, donde el método se combina con nuevos niveles de teoría basados en DFT, están respaldando este resultado.

5. Hemos propuesto que las distintas interfaces en los polímeros de FtsZ presentan un comportamiento distinto en función de su posición dentro del filamento. Para ello se ha llevado a cabo un estudio de polimerización/despolimerización, observando cambios de estabilidad en función de la posición del centro catalítico. A raíz de estos resultados, se concluye que los centros activos situados en los extremos presentan una mayor probabilidad en cuanto al proceso de hidrólisis se refiere. Esta hipótesis estaría de acuerdo con los resultados obtenidos *in vitro* en el estudio de

actividad GTPasa.

6. De acuerdo con estos resultados expuestos, y considerando el método explicado basado en el nuevo nivel de teoría, se hace patente la necesidad de un nuevo modelo de FtsZ en el que se pueda realizar un estudio de la reacción, pudiendo evaluar el papel de los residuos de ácido aspártico situados en el monómero que constituye el centro catalítico, así como el efecto de los distintos iones durante el proceso de reacción.

6. Published articles

- **Martín-García Fernando**, Papaleo Elena, Gómez Puertas Paulino, Boosma Wouter Lindorff-Larsen Kresten. Comparing molecular dynamics force-fields in the essential subspace. In preparation.
- Pablo Gonzalez de Prado Salas, Ines Hörger, **Fernando Martín-García**, Jesus Mendieta, Paulino Gómez-Puertas, Mario Encinar, Alvaro Alonso, Pedro Tarazona, Marisela Velez. Torsion and curvature of FtsZ filaments. In preparation.
- **Martín-García, F**., Mendieta-Moreno, J. I., Marcos-Alcalde, I., Gómez-Puertas, P. & Mendieta, J. Quantum Mechanics/Molecular Mechanics study of catalytic water molecule activation in myosin. In preparation.
- **Martín-García, F**., Mendieta-Moreno, J. I., Marcos-Alcalde, I., Gómez-Puertas, P. & Mendieta, J. Simulation of Catalytic Water Activation in Mitochondrial F(1)-ATPase Using a Hybrid Quantum Mechanics/Molecular Mechanics Approach: An Alternative Role for β-Glu 188. *Biochemistry* **52**, 959–66 (2013).
- **Martín-García, F**., Mendieta-Moreno, J. I., López-Viñas, E., Gómez-Puertas, P. & Mendieta, J. The Role of Gln61 in HRas GTP hydrolysis: a quantum mechanics/molecular mechanics study. *Biophysical journal* **102**, 152–7 (2012).
- **Martín-García F**, Mendieta-Moreno JI, Mendieta J, Gómez-Puertas P. Molecular dynamics analysis of conformational change of paramyxovirus F proteinvduring the initial steps of membrane fusion. Biochem Biophys Res Commun. 30;420(1):42-7 (2012)
- **Martín-García F**, Salvarelli E, Mendieta-Moreno JI, Vicente M, Mingorance J, Mendieta J, Gómez-Puertas P. Molecular dynamics simulation of GTPase activity in polymers of the cell division protein FtsZ. FEBS Lett. 586(8), 1236-9. (2012).

7. References

1. Cukierman, S. Et tu, Grotthuss! and other unfinished stories. *Biochimica et biophysica acta* **1757**, 876–85 (2006).

2. Bi, E. F. & Lutkenhaus, J. FtsZ ring structure associated with division in Escherichia coli. *Nature* **354**, 161–4 (1991).

3. Löwe, J. & Amos, L. a Crystal structure of the bacterial cell-division protein FtsZ. *Nature* **391**, 203–6 (1998).

4. Encinar, M. *et al.* Polymorphism of FtsZ Filaments on Lipid Surfaces: Role of Monomer Orientation. *Langmuir : the ACS journal of surfaces and colloids* **29**, 9436–46 (2013).

5. Rothfield, L., Justice, S. & García-Lara, J. Bacterial cell division. *Annual review of genetics* **33**, 423–48 (1999).

6. Beech, P. L. & Gilson, P. R. FtsZ and organelle division in Protists. *Protist* **151**, 11–6 (2000).

7. Erickson, H. P. Dynamin and FtsZ: Missing Links in Mitochondrial and Bacterial Division. *The Journal of Cell Biology* **148**, 1103–1106 (2000).

8. Mendieta, J. *et al.* Structural and functional model for ionic (K(+)/Na(+)) and pH dependence of GTPase activity and polymerization of FtsZ, the prokaryotic ortholog of tubulin. *Journal of molecular biology* **390**, 17–25 (2009).

9. Scheffers, D.-J. & Driessen, A. J. M. The polymerization mechanism of the bacterial cell division protein FtsZ. *FEBS letters* **506**, 6–10 (2001).

10. Carballido-López, R. & Errington, J. A dynamic bacterial cytoskeleton. *Trends in Cell Biology* **13**, 577–583 (2003).

11. Erickson, H. P., Anderson, D. E. & Osawa, M. FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one. *Microbiology and molecular biology reviews : MMBR* **74**, 504–28 (2010).

12. Adler, H. I., Fisher, W. D., Cohen, A. & Hardigree, A. A. MINIATURE escherichia coli CELLS DEFICIENT IN DNA. *Proceedings of the National Academy of Sciences of the United States of America* **57**, 321–6 (1967).

13. Reeve, J. N., Mendelson, N. H., Coyne, S. I., Hallock, L. L. & Cole, R. M. Minicells of Bacillus subtilis. *Journal of bacteriology* **114**, 860–73 (1973).

14. Woldringh, C. L., Mulder, E., Huls, P. G. & Vischer, N. Toporegulation of bacterial division according to the nucleoid occlusion model. *Research in microbiology* **142**, 309–20

15. Margolin, W. Themes and variations in prokaryotic cell division. *FEMS microbiology reviews* **24**, 531–48 (2000).

16. Wang, L. & Lutkenhaus, J. FtsK is an essential cell division protein that is localized to the septum and induced as part of the SOS response. *Molecular microbiology* **29**, 731–40 (1998).

17. Wang, S., Arends, S. J. R., Weiss, D. S. & Newman, E. B. A deficiency in S-adenosylmethionine synthetase interrupts assembly of the septal ring in Escherichia coli K-12. *Molecular microbiology* **58**, 791–9 (2005).

18. Aarsman, M. E. G. *et al.* Maturation of the Escherichia coli divisome occurs in two steps. *Molecular microbiology* **55**, 1631–45 (2005).

19. Vicente, M. & Rico, A. I. The order of the ring: assembly of Escherichia coli cell division components. *Mol. Microbiol.* **61**, 5–8 (2006).

20. Salvarelli, E., Krupka, M., Rivas, G., Vicente, M. & Mingorance, J. Independence between GTPase active sites in the Escherichia coli cell division protein FtsZ. *FEBS letters* **585**, 3880–3 (2011).

21. Neves, S. R., Ram, P. T. & Iyengar, R. G protein pathways. *Science (New York, N.Y.)* **296**, 1636–9 (2002).

22. Morrison, D. K., Kaplan, D. R., Rapp, U. & Roberts, T. M. Signal transduction from membrane to cytoplasm: growth factors and membrane-bound oncogene products increase Raf-1 phosphorylation and associated protein kinase activity. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 8855–9 (1988).

23. Hernández-Alcoceba, R., Del Peso, L. & Lacal, J. C. The Ras family of GTPases in cancer cell invasion. *Cellular and molecular life sciences : CMLS* **57**, 65–76 (2000).

24. Kosloff, M. & Selinger, Z. Substrate assisted catalysis -- application to G proteins. *Trends in biochemical sciences* **26**, 161–6 (2001).

25. Sprang, S. R. G proteins, effectors and GAPs: structure and mechanism. *Current opinion in structural biology* **7**, 849–56 (1997).

26. Vetter, I. R. & Wittinghofer, a The guanine nucleotide-binding switch in three dimensions. *Science (New York, N.Y.)* **294**, 1299–304 (2001).

27. Grant, B. J., McCammon, J. A. & Gorfe, A. A. Conformational selection in G-proteins: lessons from Ras and Rho. *Biophysical journal* **99**, L87–9 (2010).

28. Scheffzek, K. The Ras-RasGAP Complex: Structural Basis for GTPase Activation and Its Loss in Oncogenic Ras Mutants. *Science* **277**, 333–338 (1997).

29. Scheffzek, K. *et al.* The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science (New York, N.Y.)* **277**, 333–8 (1997).

30. Fidyk, N. J. & Cerione, R. a Understanding the catalytic mechanism of GTPase-

activating proteins: demonstration of the importance of switch domain stabilization in the stimulation of GTP hydrolysis. *Biochemistry* **41**, 15644–53 (2002).

31. Valiyaveetil, F. I. & Fillingame, R. H. Transmembrane topography of subunit a in the Escherichia coli F1F0 ATP synthase. *The Journal of biological chemistry* **273**, 16241–7 (1998).

32. Vik, S. B., Long, J. C., Wada, T. & Zhang, D. A model for the structure of subunit a of the Escherichia coli ATP synthase and its role in proton translocation. *Biochimica et biophysica acta* **1458**, 457–66 (2000).

33. Stalz, W.-D., Greie, J.-C., Deckers-Hebestreit, G. & Altendorf, K. Direct interaction of subunits a and b of the F0 complex of Escherichia coli ATP synthase by forming an ab2 subcomplex. *The Journal of biological chemistry* **278**, 27068–71 (2003).

34. Tsunoda, S. P. *et al.* Large conformational changes of the epsilon subunit in the bacterial F1F0 ATP synthase provide a ratchet action to regulate this rotary motor enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6560–4 (2001).

35. Boyer, P. D. The ATP synthase--a splendid molecular machine. *Annual review of biochemistry* **66**, 717–49 (1997).

36. Weber, J., Wilke-Mounts, S., Lee, R. S., Grell, E. & Senior, A. E. Specific placement of tryptophan in the catalytic sites of Escherichia coli F1-ATPase provides a direct probe of nucleotide binding: maximal ATP hydrolysis occurs with three sites occupied. *The Journal of biological chemistry* **268**, 20126–33 (1993).

37. Weber, J., Wilke-Mounts, S. & Senior, A. E. Cooperativity and stoichiometry of substrate binding to the catalytic sites of Escherichia coli F1-ATPase. Effects of magnesium, inhibitors, and mutation. *The Journal of biological chemistry* **269**, 20462–7 (1994).

38. Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621–8 (1994).

39. Menz, R. I., Walker, J. E. & Leslie, A. G. W. Structure of Bovine Mitochondrial F1-ATPase with Nucleotide Bound to All Three Catalytic Sites. *Cell* **106**, 331–341 (2001).

40. Gao, Y. Q., Yang, W. & Karplus, M. A structure-based model for the synthesis and hydrolysis of ATP by F1-ATPase. *Cell* **123**, 195–205 (2005).

41. Yang, W., Gao, Y. Q., Cui, Q., Ma, J. & Karplus, M. The missing link between thermodynamics and structure in F1-ATPase. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 874–9 (2003).

42. Gao, Y. Q., Yang, W., Marcus, R. A. & Karplus, M. A model for the cooperative free energy transduction and kinetics of ATP hydrolysis by F1-ATPase. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11339–44 (2003).

43. Sweeney, H. L. & Houdusse, A. Structural and functional insights into the Myosin motor mechanism. *Annual review of biophysics* **39**, 539–57 (2010).

44. Koubassova, N. A. & Tsaturyan, A. K. Molecular mechanism of actin-myosin motor in muscle. *Biochemistry. Biokhimiia* **76**, 1484–506 (2011).

45. Oosawa, F. The unit event of sliding of the chemo-mechanical enzyme composed of myosin and actin with regulatory proteins. *Biochemical and biophysical research communications* **369**, 144–8 (2008).

46. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98–103 (1967).

47. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **103**, 227–249 (1976).

48. Field, M. J., Bash, P. A. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry* **11**, 700–733 (1990).

49. Cheatham, T. E., Woods, R. J. & Case, D. A. The AMBER biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).

50. Mendieta, J. *et al.* Role of histidine-85 in the catalytic mechanism of thymidine phosphorylase as assessed by targeted molecular dynamics simulations and quantum mechanical calculations. *Biochemistry* **43**, 405–14 (2004).

51. Mendieta, J., Gago, F. & Ramírez, G. Binding of 5'-GMP to the GluR2 AMPA receptor: insight from targeted molecular dynamics simulations. *Biochemistry* **44**, 14470–6 (2005).

52. Mendieta, J. *et al.* Phosphorylation modulates the alpha-helical structure and polymerization of a peptide from the third tau microtubule-binding repeat. *Biochimica et biophysica acta* **1721**, 16–26 (2005).

53. Martín-García, F., Mendieta-Moreno, J. I., López-Viñas, E., Gómez-Puertas, P. & Mendieta, J. The Role of Gln61 in HRas GTP hydrolysis: a quantum mechanics/molecular mechanics study. *Biophysical journal* **102**, 152–7 (2012).

54. Martín-García, F., Mendieta-Moreno, J. I., Marcos-Alcalde, I., Gómez-Puertas, P. & Mendieta, J. Simulation of Catalytic Water Activation in Mitochondrial F(1)-ATPase Using a Hybrid Quantum Mechanics/Molecular Mechanics Approach: An Alternative Role for β-Glu 188. *Biochemistry* **52**, 959–66 (2013).

55. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *Journal of computational chemistry* **30**, 1545–614 (2009).

56. Hess, B., Kutzner, C., Van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4**, 435–447 (2008).

57. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781–802 (2005).

58. Martín-García, F., Mendieta-Moreno, J. I., Mendieta, J. & Gómez-Puertas, P. Molecular dynamics analysis of conformational change of paramyxovirus F protein during the initial steps of membrane fusion. *Biochemical and biophysical research communications* **420**, 42–7 (2012).

59. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **112**, 8910 (2000).

60. Wu, Y., Tepper, H. L. & Voth, G. A. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of chemical physics* **124**, 024503 (2006).

61. Ponder, J. W. & Case, D. A. Force Fields for Protein Simulations. *Advances in Protein Chemistry* **66**, 27–87 (2003).

62. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–25 (2006).

63. Berman, H. M. The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242 (2000).

64. Gordon, J. C. *et al.* H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic acids research* **33**, W368–71 (2005).

65. Shaw, D. E. *et al.* Atomic-level characterization of the structural dynamics of proteins. *Science (New York, N.Y.)* **330**, 341–6 (2010).

66. Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* **11**, 224–230 (2001).

67. Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D. & Schulten, K. Steered molecular dynamics investigations of protein function. *Journal of molecular graphics & modelling* **19**, 13–25 (2001).

68. Binnig, G. & Quate, C. F. Atomic Force Microscope. *Physical Review Letters* **56**, 930–933 (1986).

69. Ashkin, A. & Dziedzic, J. Optical trapping and manipulation of viruses and bacteria. *Science* **235**, 1517–1520 (1987).

70. Evans, E., Ritchie, K. & Merkel, R. Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces. *Biophysical journal* **68**, 2580–7 (1995).

71. Grigorenko, B. L. *et al.* Mechanism of the myosin catalyzed hydrolysis of ATP as rationalized by molecular modeling. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7057–61 (2007).

72. Grigorenko, B. L., Nemukhin, A. V, Topol, I. a, Cachau, R. E. & Burt, S. K. QM/MM

modeling the Ras-GAP catalyzed hydrolysis of guanosine triphosphate. *Proteins* **60**, 495–503 (2005).

73. Grigorenko, B. L., Kaliman, I. a & Nemukhin, A. V Minimum energy reaction profiles for ATP hydrolysis in myosin. *Journal of molecular graphics & modelling* **31**, 1–4 (2011).

74. Mingorance, J., Rivas, G., Vélez, M., Gómez-Puertas, P. & Vicente, M. Strong FtsZ is with the force: mechanisms to constrict bacteria. *Trends in microbiology* **18**, 348–56 (2010).

75. Singh, P. & Panda, D. FtsZ inhibition: a promising approach for antistaphylococcal therapy. *Drug news & perspectives* **23**, 295–304 (2010).

76. Awasthi, D., Kumar, K. & Ojima, I. Therapeutic potential of FtsZ inhibition: a patent perspective. *Expert opinion on therapeutic patents* **21**, 657–79 (2011).

77. Mingorance, J., Rueda, S., Gómez-Puertas, P., Valencia, a & Vicente, M. Escherichia coli FtsZ polymers contain mostly GTP and have a high nucleotide turnover. *Molecular microbiology* **41**, 83–91 (2001).

78. Oliva, M. a, Cordell, S. C. & Löwe, J. Structural insights into FtsZ protofilament formation. *Nature structural & molecular biology* **11**, 1243–50 (2004).

79. González, J. M. *et al.* Cooperative behavior of Escherichia coli cell-division protein FtsZ assembly involves the preferential cyclization of long single-stranded fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1895–900 (2005).

80. Sossong, T. M., Brigham-Burke, M. R., Hensley, P. & Pearce, K. H. Self-activation of guanosine triphosphatase activity by oligomerization of the bacterial cell division protein FtsZ. *Biochemistry* **38**, 14843–50 (1999).

81. Martín-Galiano, A. J., Buey, R. M., Cabezas, M. & Andreu, J. M. Mapping flexibility and the assembly switch of cell division protein FtsZ by computational and mutational approaches. *The Journal of biological chemistry* **285**, 22554–65 (2010).

82. Ghosh, B. & Sain, A. Force generation in bacteria without nucleotide-dependent bending of cytoskeletal filaments. *Physical review. E, Statistical, nonlinear, and soft matter physics* **83**, 051924 (2011).

83. Case, D. A. *et al.* The Amber biomolecular simulation programs. *Journal of computational chemistry* **26**, 1668–88 (2005).

84. Rivas, G. *et al.* Magnesium-induced linear self-association of the FtsZ bacterial cell division protein monomer. The primary steps for FtsZ assembly. *The Journal of biological chemistry* **275**, 11740–9 (2000).

85. Hoenig, M., Lee, R. J. & Ferguson, D. C. A microtiter plate assay for inorganic phosphate. *Journal of biochemical and biophysical methods* **19**, 249–51

86. Lanzetta, P. A., Alvarez, L. J., Reinach, P. S. & Candia, O. A. An improved assay for

nanomole amounts of inorganic phosphate. *Analytical biochemistry* **100**, 95–7 (1979).

87. Aylett, C. H. S., Wang, Q., Michie, K. a, Amos, L. a & Löwe, J. Filament structure of bacterial tubulin homologue TubZ. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19766–71 (2010).

88. Grigorenko, B. L., Nemukhin, A. V, Shadrina, M. S., Topol, I. A. & Burt, S. K. Mechanisms of guanosine triphosphate hydrolysis by Ras and Ras-GAP proteins as rationalized by ab initio QM/MM simulations. *Proteins* **66**, 456–66 (2007).

89. Parke, C. L., Wojcik, E. J., Kim, S. & Worthylake, D. K. ATP hydrolysis in Eg5 kinesin involves a catalytic two-water mechanism. *The Journal of biological chemistry* **285**, 5859–67 (2010).

90. Scheffzek, K., Ahmadian, M. R. & Wittinghofer, A. GTPase-activating proteins: helping hands to complement an active site. *Trends in biochemical sciences* **23**, 257–62 (1998).

91. Wittinghofer, A. Phosphoryl transfer in Ras proteins, conclusive or elusive? *Trends in biochemical sciences* **31**, 20–3 (2006).

92. Pai, E. F. *et al.* Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *The EMBO journal* **9**, 2351–9 (1990).

93. Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. & Noda, M. A ras-related gene with transformation suppressor activity. *Cell* **56**, 77–84 (1989).

94. Der, C. J., Finkel, T. & Cooper, G. M. Biological and biochemical properties of human rasH genes mutated at codon 61. *Cell* **44**, 167–76 (1986).

95. Frech, M. *et al.* Role of glutamine-61 in the hydrolysis of GTP by p21H-ras: an experimental and theoretical study. *Biochemistry* **33**, 3237–44 (1994).

96. Fasano, O. *et al.* Analysis of the transforming potential of the human H-ras gene by random mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 4008–12 (1984).

97. Maegley, K. a, Admiraal, S. J. & Herschlag, D. Ras-catalyzed hydrolysis of GTP: a new perspective from model studies. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8160–6 (1996).

98. Schweins, T. *et al.* Substrate-assisted catalysis as a mechanism for GTP hydrolysis of p21ras and other GTP-binding proteins. *Nature structural biology* **2**, 36–44 (1995).

99. Schweins, T., Langen, R. & Warshel, A. Why have mutagenesis studies not located the general base in ras p21. *Nature structural biology* **1**, 476–84 (1994).

100. Privé, G. G. *et al.* X-ray crystal structures of transforming p21 ras mutants suggest a transition-state stabilization mechanism for GTP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 3649–53 (1992).

101. Scheidig, a J., Burmester, C. & Goody, R. S. The pre-hydrolysis state of p21(ras) in complex with GTP: new insights into the role of water molecules in the GTP hydrolysis reaction of ras-like proteins. *Structure (London, England : 1993)* **7**, 1311–24 (1999).

102. Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics* **119**, 3559 (2003).

103. Kästner, J. & Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *The Journal of chemical physics* **123**, 144104 (2005).

104. Kamerlin, S. C. L., Haranczyk, M. & Warshel, A. Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *The journal of physical chemistry. B* **113**, 1253–72 (2009).

105. Walker, R. C., Crowley, M. F. & Case, D. A. The implementation of a fast and accurate QM/MM potential method in Amber. *Journal of computational chemistry* **29**, 1019–31 (2008).

106. Babin, V., Roland, C. & Sagui, C. Adaptively biased molecular dynamics for free energy calculations. *The Journal of chemical physics* **128**, 134101 (2008).

107. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D. & Allen, K. N. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science (New York, N.Y.)* **299**, 2067–71 (2003).

108. Boero, M., Ikeda, T., Ito, E. & Terakura, K. Hsc70 ATPase: an insight into water dissociation and joint catalytic role of K+ and Mg2+ metal cations in the hydrolysis reaction. *Journal of the American Chemical Society* **128**, 16798–807 (2006).

109. Senior, A. E., Nadanaciva, S. & Weber, J. The molecular mechanism of ATP synthesis by F1F0-ATP synthase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1553**, 188–211 (2002).

110. Alberts, B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **92**, 291–294 (1998).

111. Karplus, M. & Pu, J. *How Biomolecular Motors Work: Synergy Between Single Molecule Experiments and Single Molecule Simulations*. **96**, 3–22 (Springer Berlin Heidelberg: Berlin, Heidelberg, 2010).

112. Karplus, M. & Gao, Y. Q. Biomolecular motors: the F1-ATPase paradigm. *Current opinion in structural biology* **14**, 250–9 (2004).

113. Nakamoto, R. K., Ketchum, C. J., Kuo, P. H., Peskova, Y. B. & Al-Shawi, M. K. Molecular mechanisms of rotational catalysis in the F0F1 ATP synthase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1458**, 289–299 (2000).

114. Kagawa, R., Montgomery, M. G., Braig, K., Leslie, A. G. W. & Walker, J. E. The structure of bovine F1-ATPase inhibited by ADP and beryllium fluoride. *The EMBO journal* **23**, 2734–44 (2004).

115. Itoh, H. *et al.* Mechanically driven ATP synthesis by F1-ATPase. *Nature* **427**, 465–8 (2004).

116. Rees, D. M., Montgomery, M. G., Leslie, A. G. W. & Walker, J. E. Structural evidence of a new catalytic intermediate in the pathway of ATP hydrolysis by F1-ATPase from bovine heart mitochondria. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 11139–43 (2012).

117. Cleland, W. W. & Hengge, A. C. Enzymatic mechanisms of phosphate and sulfate transfer. *Chemical reviews* **106**, 3252–78 (2006).

118. Bianchet, M. A., Hullihen, J., Pedersen, P. L. & Amzel, L. M. The 2.8-A structure of rat liver F1-ATPase: Configuration of a critical intermediate in ATP synthesis/hydrolysis. *Proceedings of the National Academy of Sciences* **95**, 11065–11070 (1998).

119. Tozawa, K. *et al.* Unusual pKa of the carboxylate at the putative catalytic position of the thermophilic F1-ATPase β subunit determined by 13C-NMR. *FEBS Letters* **397**, 122–126 (1996).

120. Kaczmarek, P., Szczepanik, W. & Jezowska-Bojczuk, M. Acid-base, coordination and oxidative properties of systems containing ATP, L-histidine and Ni(II) ions. *Dalton transactions (Cambridge, England : 2003)* 3653–7 (2005).doi:10.1039/b508962j

121. Dittrich, M., Hayashi, S. & Schulten, K. On the mechanism of ATP hydrolysis in F1-ATPase. *Biophysical journal* **85**, 2253–66 (2003).

122. Dittrich, M., Hayashi, S. & Schulten, K. ATP hydrolysis in the betaTP and betaDP catalytic sites of F1-ATPase. *Biophysical journal* **87**, 2954–67 (2004).

123. Hayashi, S. *et al.* Molecular mechanism of ATP hydrolysis in F1-ATPase revealed by molecular simulations and single-molecule observations. *Journal of the American Chemical Society* **134**, 8447–54 (2012).

124. Walker, R. C., Crowley, M. F. & Case, D. A. The implementation of a fast and accurate QM/MM potential method in Amber. *J Comput Chem* **29**, 1019–1031 (2008).

125. Nathanson, L. Mutagenesis of beta -Glu-195 of the Rhodospirillum rubrum F1-ATPase and Its Role in Divalent Cation-dependent Catalysis. *Journal of Biological Chemistry* **273**, 10933–10938 (1998).

126. Enoki, S., Watanabe, R., Iino, R. & Noji, H. Single-molecule study on the temperature-sensitive reaction of F1-ATPase with a hybrid F1 carrying a single beta(E190D). *The Journal of biological chemistry* **284**, 23169–76 (2009).

127. Amano, T., Tozawa, K., Yoshida, M. & Murakami, H. Spatial precision of a catalytic carboxylate of F1-ATPase β subunit probed by introducing different carboxylate-

containing side chains. *FEBS Letters* **348**, 93–98 (1994).

128. Nathanson, L. Mutations in the beta -Subunit Thr159 and Glu184 of the Rhodospirillum rubrum F0F1 ATP Synthase Reveal Differences in Ligands for the Coupled Mg2+- and Decoupled Ca2+-dependent F0F1 Activities. *Journal of Biological Chemistry* **275**, 901–905 (2000).

129. Cecchini, M., Houdusse, A. & Karplus, M. Allosteric communication in myosin V: from small conformational changes to large directed movements. *PLoS computational biology* **4**, e1000129 (2008).

130. Chibalina, M. V, Puri, C., Kendrick-Jones, J. & Buss, F. Potential roles of myosin VI in cell motility. *Biochemical Society transactions* **37**, 966–70 (2009).

131. Loubéry, S. & Coudrier, E. Myosins in the secretory pathway: tethers or transporters? *Cellular and molecular life sciences : CMLS* **65**, 2790–800 (2008).

132. Ross, J. L., Ali, M. Y. & Warshaw, D. M. Cargo transport: molecular motors navigate a complex cytoskeleton. *Current opinion in cell biology* **20**, 41–7 (2008).

133. Woolner, S. & Bement, W. M. Unconventional myosins acting unconventionally. *Trends in cell biology* **19**, 245–52 (2009).

134. Onishi, H., Ohki, T., Mochizuki, N. & Morales, M. F. Early stages of energy transduction by myosin: roles of Arg in switch I, of Glu in switch II, and of the salt-bridge between them. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 15339–44 (2002).

135. Smith, C. a & Rayment, I. X-ray structure of the magnesium(II).ADP.vanadate complex of the Dictyostelium discoideum myosin motor domain to 1.9 A resolution. *Biochemistry* **35**, 5404–17 (1996).

136. Yang, Y., Yu, H. & Cui, Q. Extensive conformational transitions are required to turn on ATP hydrolysis in myosin. *Journal of molecular biology* **381**, 1407–20 (2008).

137. Fedorov, R. *et al.* The mechanism of pentabromopseudilin inhibition of myosin motor activity. *Nature structural & molecular biology* **16**, 80–8 (2009).

138. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. chem. Phys.* **79**, 926–935 (1983).

139. Onishi, H. *et al.* Functional transitions in myosin: formation of a critical salt-bridge and transmission of effect to the sensitive tryptophan. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6653–8 (1998).

140. Sasaki, N., Shimada, T. & Sutoh, K. Mutational analysis of the switch II loop of Dictyostelium myosin II. *The Journal of biological chemistry* **273**, 20334–40 (1998).

141. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61**, 704–21 (2005).

142. Frye, J. J., Klenchin, V. a, Bagshaw, C. R. & Rayment, I. Insights into the importance of hydrogen bonding in the gamma-phosphate binding pocket of myosin: structural and functional studies of serine 236. *Biochemistry* **49**, 4897–907 (2010).

143. Shimada, T., Sasaki, N., Ohkura, R. & Sutoh, K. Alanine scanning mutagenesis of the switch I region in the ATPase site of Dictyostelium discoideum myosin II. *Biochemistry* **36**, 14037–43 (1997).