In silico screening of NRas protein oncogenic mutations : new structural and physico-chemical insights into the catalytic activity

Ruth Helena Tichauer

To cite this version:

Ruth Helena Tichauer. In silico screening of NRas protein oncogenic mutations : new structural and physico-chemical insights into the catalytic activity. Catalysis. Université Paul Sabatier - Toulouse III, 2019. English. NNT : 2019TOU30028 . tel-02090107v2

HAL Id: tel-02090107

https://theses.hal.science/tel-02090107v2

Submitted on 31 Aug 2020

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Ruth Helena TICHAUER

Le 1er Février 2019

In Silico Screening of NRas Protein Oncogenic Mutations: New Structural and Physico-Chemical Insights into the Catalytic Activity

Ecole doctorale: **SDM - SCIENCES DE LA MATIERE - Toulouse**

Spécialité: **Physique de la Matière**

Unité de recherche:
LAAS - Laboratoire d'Analyse et d'Architecture des Systèmes

Thèse dirigée par
Marie BRUT

Jury

Carsten KÖTTING, PD Dr, Ruhr University	Rapporteur
Eric HENON, Pr, Université de Reims	Rapporteur
Fabienne ALARY, Pr, Université Toulouse 3	Examinatrice
Georges CZAPLICKI, Pr, Université Toulouse 3	Examinateur
A. Marco SAITTA, Pr, Sorbonne Université	Examinateur
Gilles FAVRE, PU, PH, IUCT Toulouse	Invité
Marie BRUT, Dr, Université Toulouse 3	Directrice de Thèse

"I would feel more optimistic about a bright future for man if he spent less time proving that he can outwit Nature and more time tasting her sweetness and respecting her seniority."

E.B. White, "Coon Tree", 1977

UNIVERSITÉ TOULOUSE III - PAUL SABATIER

Abstract

Doctor of Philosophy

In Silico Screening of NRas Protein Oncogenic Mutations: New Structural and Physico-Chemical Insights into the Catalytic Activity

by Ruth Helena TICHAUER

Ras subfamily of small GTPase proteins holds a key position in cell proliferation pathways. Indeed, the transmission of cell growth signals is controlled by proteins belonging to it. In their GTP-bound conformation, these proteins interact and activate downstream effectors of cell replication and differentiation. The hydrolysis reaction that takes place in their center, terminates these interactions, thereby leading to the GDP-bound inactive state.

Point mutations of key residues lead to a hydrolysis rate drop that keeps Ras in a GTP-bound active state. Now, high concentrations of active Ras have been associated to abnormal cell proliferation, emblematic of cancerous tissues dissemination. With this into consideration, the elucidation of Ras mechanisms for accelerating GTP cleavage appears as a major step in the development of cancer targeted therapies that would consist in restoring the hydrolyzing capabilities within oncogenic Ras to a wild-type rate.

In an attempt to gain insight into Ras catalyzing properties at the atomic level, unconstrained Molecular Dynamics (MD) simulations describing the nucleotide binding domain, commonly named G domain, at different levels of theory (Molecular Mechanics (MM), Semi-empirical and Density Functional Theory (DFT)) were carried out for NRas member in its wild-type and Gln 61 mutated forms. These simulations were coupled to biomechanic characterizations of the complexes under inspection employing the Static Modes (SM) approach. The latter method, allows the identification, through certain physicochemical criteria, of hot spots *i.e.* responsive residues of the biomolecule, that have a mechanical influence on the GTPase function of the protein. Hence, they could serve as suitable sites to host drug-like molecules containing specific chemical groups that would facilitate GTP hydrolysis.

The obtained results show that water molecules positioning is crucial for efficiently catalyzing the reaction that takes place in NRas center. Indeed, the precise positioning observed within the wild-type is lost within the mutant proteins studied here. Furthermore, the active site structural modifications undergone upon Gln 61 substitutions, together with solvent distribution in it, impact directly GTP electronic density. The latter is accommodated to a GDP-like state within the wild-type only, as experimentally determined in previous investigations. Thus, oncogenic Gln 61 mutations impair this major catalyzing effect. Finally, among three engineered NRas proteins of the Q61R mutated form proposed during this thesis, one enables the recovery of water distribution as within the wild-type.

Acknowledgements

The fruit of the work carried during these three memorable years, would not have been harvest without the direct/indirect implication of several persons I would like to honor in this paragraph.

First, I would like to thank my PhD supervisor, Dr. Marie Brut, for giving me the opportunity to integrate and work on this project. Very soon, I realized that Ras proteins constituted a subject dear to her heart. Yet, she handed it over to me, letting me carry out the research on it following the directions I selected despite I had no experience whatsoever in computational structural biology. Her trust let my academic background in theoretical physics express itself for studying Ras proteins, thereby leading to the study of quantities that had not been envisioned. Thank you Marie for granting me freedom in developing this subject as I desired.

Subsequently, I would like to thank the permanent members of M^3 team, Anne and Georges. Thank you for permanently providing scientific, technical and moral support. It has been delightful sharing this time with you. I can only hope that our research paths will keep crossing over one another.

Then, I would like to thank the PhD students of the group, in particular Miha, for permanent support and infinite patience notably during the writing of this manuscript. Thank you!

Not every day in our lab but always bringing about relevant questions, thank you Antoine (a postdoc within our team) for your enthusiasm. It is so frustrating having to share you with another institution..

I would like to address a special thanks to all my friends, in particular Ilias and Lorena, for support all along the PhD experience. Thanks to you this phase has been marked with periods of laughter and joy.

Last but not least, I would like to thank both my parents who taught me from my youngest years that knowledge is the greatest wealth one can ever acquire. Thank you for educating me in a manner that I intend to become as wealthy as my mind allows for it. I can only wish to become as exemplary as you are to me.

Finally, I would like to thank Walter, Fernanda and Hilda, dear to my heart, for ever lasting phone conversations and keeping me in touch with my roots.

Contents

List of Figures

List of Tables

List of Abbreviations

AO	**A**tomic **O**rbitals
DFT	**D**ensity **F**unctional **T**heory
FTIR	**F**ourier **T**ransformed **I**nfra**R**ed
GAP	**G**TPase **A**ctivating **Protein**
GDP	**G**uanosine **D**i **P**hosphate
GEF	**G**uanine nucleotide **E**xchange **F**actor
GGA	**G**eneralized **G**radient **A**pproximation
GTP	**G**uanosine **T**ri **P**hosphate
HOMO	**H**ighest **O**ccupied **M**olecular **O**rbital
IR	**I**nfra**R**ed
KS	**K**ohn **S**ham
LCAO	**L**inear **C**ombination of **A**tomic **O**rbitals
LDA	**L**ocal **D**ensity **A**pproximation
LUMO	**L**owest **U**noccupied **M**olecular **O**rbital
MD	**M**olecular **D**ynamics
MGGA	**M**eta **G**eneralized **G**radient **A**pproximation
MM	**M**olecular **M**echanics
MO	**M**olecular **O**rbitals
PES	**P**otential **E**nergy **S**urface
QM/MM	**Q**uantum **M**echanics / **M**olecular **M**echanics
RDF	**R**adial **D**istribution **F**unction
RTK	**R**eceptor **T**yrosine **K**inases
RMSD	**R**oot **M**ean **S**quare **D**eviation
SCF	**S**elf **C**onsistent **F**ield

To my Pachi…

Introduction

Biochemistry [49], in the sense of explaining living processes through chemical reactions finds its beginnings in the 19th century when, to the detriment of vitalism, Friedrich Wöhler[1] synthesized urea from ammonium cyanate. By this achievement, he proved that organic compounds could be artificially created, hence discarding the hypothesis that only living beings could produce the molecules of life.

The emergence of this discipline has brought into light the organizational complexity of the numerous lifeless molecules living organisms rely on. Through essential compounds, namely biomolecules, living organisms extract, transform, channel and use energy from their surroundings, adapt to the latter and replicate, thereby ensuring survival. At the microscopic level, this is achieved through a colossal chemical complexity: each entity has a defined function, and regulated interactions among them, enable sensing and responding mechanisms for adapting to the constantly changing environment they are immersed in.

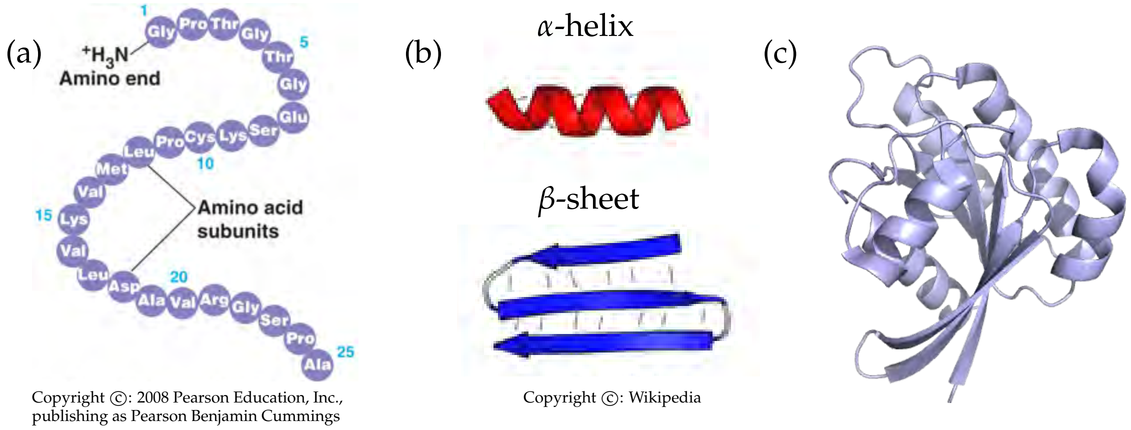
The vast diversity of both biomolecules (lipids, carbohydrates, proteins, nucleic acids) and the mechanisms they are involved in, have lead to the division of biochemistry in three main fields: *i)* metabolism, which deals with the set of transformations that occur within living cells, *ii)* molecular genetics, which studies the structure and function of genes at the molecular level and *iii)* protein science, which investigates proteins' structure and function at the molecular level as well.

The work presented in this manuscript belongs to the latter.

Proteins are the most abundant biological macromolecules in cells. They are built by the association of single units called amino acids that are covalently bound through their *backbone*. This is achieved by eliminating atoms corresponding to the elements of water. As such, these monomeric subunits are usually referred to as *amino acid residues* or simply *residues* within the polymer they form through their union. Different combinations of the finite number of naturally existing amino acids (20 in total plus some variants) give rise to thousands of different proteins. They range in size from

---

[1] German chemist, 1800-1882



Copyright ©: 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings

Copyright ©: Wikipedia

FIGURE 1: A protein primary (a), secondary (b) and tertiary (c) structures.

small peptides to macromolecules and have diverse biological functions (enzymes, hormones, antibodies, transporters, muscle fibers..). The linear amino acid sequence through which a protein is built is called its *primary structure* (see figure 1(a)).

The primary structure is assembled from RNA coding of another linear sequence: DNA. Hence, proteins are the molecular expression of genetic information. As the effective storage, expression and reproduction of the latter defines individual species and assures their lasting over successive generations, proteins can be viewed as the most important final products of information paths, because their correct functioning ensures the organism survival. Indeed, numerous genetic diseases have been traced to the production of defective proteins that loose their function. This is due to unrepaired mistakes that occur during the DNA replication process that result in point changes within the nucleotide and, thus, within the protein amino acid sequence. These point changes are commonly called *mutations* and may *i)* enhance the protein function, *ii)* have no impact whatsoever on it, or *iii)* destroy it. As such, proteins perpetrating a certain function usually originate from diverse amino acid sequences. The latter contain *i)* dissimilar regions with little or no effect on protein function and *ii)* identical regions that are crucial to it and have been conserved over evolutionary time. Both diversity and similarity in primary structures has lead to the construction of separate protein families. The constituent members usually share common amino acid residues essential to perpetrate the specific function associated to the family. Because the primary structure of a protein encodes its functioning / malfunctioning, it appears that the information carried by the sequence also determines protein folding into a unique three dimensional (3D) structure. Indeed, the latter is closely related to its function within the cell.

Between the primary and 3D structures of a protein, lies an intermediate folding of neighboring amino acid residues called the *secondary structure* (see figure 1(b)). This intermediate 3D arrangement is remarkable because a finite number of stable recurring structural patterns arise like α helices, β strands, γ turns, to name but a few. Each type of secondary structure tends to be constituted by the same type of residues. Indeed specific amino acids better take part in the major non-bonded interactions that govern each type of intermediate folding.

The overall 3D folding of a protein is called the *tertiary structure* (see figure 1(c)). Amino acid residues that are far apart in the sequence and that might be part of different secondary structures, are often brought to interact within the completely folded protein. Differently from other molecules that can adopt an unlimited number of conformations through free rotations about their bonds, proteins exist in a small number of 3D structures called *native* conformations. The latter are thermodynamically the most stable, *i.e.* they have the lowest Gibbs free energy (G) under the given set of environmental conditions (pH, ion strength, metal ion concentrations..):

$$G = H - TS \quad \text{where} \quad \begin{array}{l} H \text{ is the enthalpy} \\ T \text{ is the temperature} \\ S \text{ is the entropy} \end{array} \tag{1}$$

This energy is minimized, *i.e.* the protein native conformation is stabilized by:

- Individual covalent bonds such as disulfide bonds that link distant parts of the primary structure
- Ionic interactions between charged groups
- Hydrophobic weak interactions
- Hydrogen bonds

The cumulative effect of the many *weak* non-covalent bonds contribute the most: the protein native conformation is generally the one with the maximum number of weak interactions[2].

Although protein folding, hence protein function, is dictated by the primary structure that holds the genetic information, the environment also has a major impact on it as it influences the strength and number of weak non-bonded interactions. Indeed, the pH of the surrounding medium determines the protonation state (*i.e.* fully protonated, dipolar ion or fully unprotonated) of the protein constituent residues. In this manner,

---

[2] Insoluble proteins and proteins within a membrane follow different rules but weak interactions are still critical structural elements

the ionic groups that are effectively present are defined, and so are possible ionic interactions. Moreover, the polarity of the solvent determines the strength of hydrophobic interactions[3]. In water, they result from the burial of non-polar residues that aims at reducing the solvation layer surrounding them. In this manner, the number of highly ordered water molecules that arise in the vicinity of hydrophobic surfaces is minimized, reducing thereby the high entropic cost such water structuring represents.

Tertiary structures are not static. The numerous non-covalent interactions that govern the native conformation confer flexibility to proteins, which is necessary for their biological function. Indeed, interconversions between the finite number of 3D structural forms proteins can adopt, enable the achievement of the task they have specialized for over evolutionary time, for example: binding other molecules, catalyzing reactions or transporting small molecules.

According to the tertiary structure, proteins are classified in two major groups: fibrous proteins and globular proteins. The latter only is briefly described in what follows as the protein studied in this thesis belongs to this category.

Globular proteins, as inferred from their name, fold into a spherical, globular shape. They contain several types of secondary structures that fold back on each other, generating a compact core comparable in compactness to crystals.

The most varied and specialized globular proteins are enzymes. They catalyze specific chemical reactions. Both their diversity and complexity might arise from the fact that living cells require, simultaneously, precise amounts of thousands of different biomolecules and thus, organisms provided with mechanisms that enhanced the precise synthesis of the needed compounds, were advantaged over evolutionary time.

By providing a surface that complements the transition state of the reaction in terms of stereochemistry, polarity and charge, enzymes accelerate reactions without affecting their equilibrium: reaction rates are greatly increased by a reduction of activation energies only (see figure 2).

Enzymes are organized within cells such that the product of one reaction becomes the substrate in the next one. The sequence of consecutive enzyme catalyzed reactions is referred to as a *metabolic pathway*. The small intermediate molecules in these pathways are called *metabolites*. They have various functions, among which regulating the activity of enzymes from pathways they are involved in. Regulation first ensures that the cell requirements are fulfilled. As metabolites can be part of numerous pathways,

---

[3] Hydrophobic interactions among lipids and between lipids and proteins play a major role in determining their structure within biological membranes

Image source: http://www.chem.uiuc.edu/rogers/text13/tx133/tx133fr.html

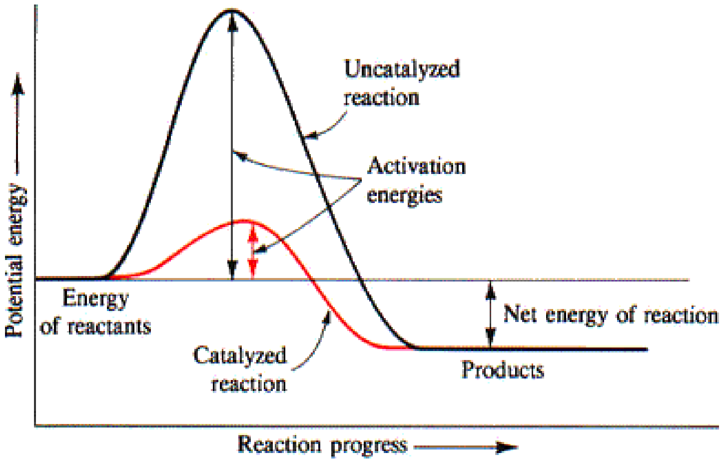


FIGURE 2: Enzymes effect on activation energies.

regulation also ensures that, the change in concentration of a specific metabolite due to a specific pathway activity, does not impact the activity of other pathways. Indeed, undesired influence between distinct cellular processes must be avoided.

The multiplicity of enzymes, their specificity and their sensitivity to regulation, bestow cells the capacity to lower activation barriers selectively and, thereby, finely tune the mechanisms that are necessary for cell survival and replication. Enzymes are thus the sculptors that shape living organisms the way we know them.

GTPases [9] are the enzymes central to this work. As inferred by their name, they bind and catalyze the hydrolysis reaction of guanosine triphosphate (GTP) into guanosine diphosphate (GDP). Through this chemical reaction they adopt three main states: *i)* GTP bound before hydrolysis, *ii)* GDP bound after the reaction, and *iii)* a transient empty state during which the products are released (GDP + inorganic phosphate P_i) and a new reactant molecule (GTP) is loaded. As in cytoplasm the concentration of GTP is higher than that of GDP, the former is more likely to enter the active site, initiating thereby a new cycle.

The bound nucleotide defines two main closely related conformations of the protein that, despite their resemblance, allow interactions with different biomolecules. Because of the irreversibility of bond breakage during hydrolysis that renders the reaction cycle unidirectional, GTPases are considered to function as molecular switches between two states, commonly named, "ON" or *active* and "OFF" or *inactive* (see figure 3). This peculiar switching mechanism has been conserved over evolutionary time but has also evolved, bestowing thereby with functional diversity the GTPase superfamily [8].

This manuscript, which carries the result of the work accomplished during three years,

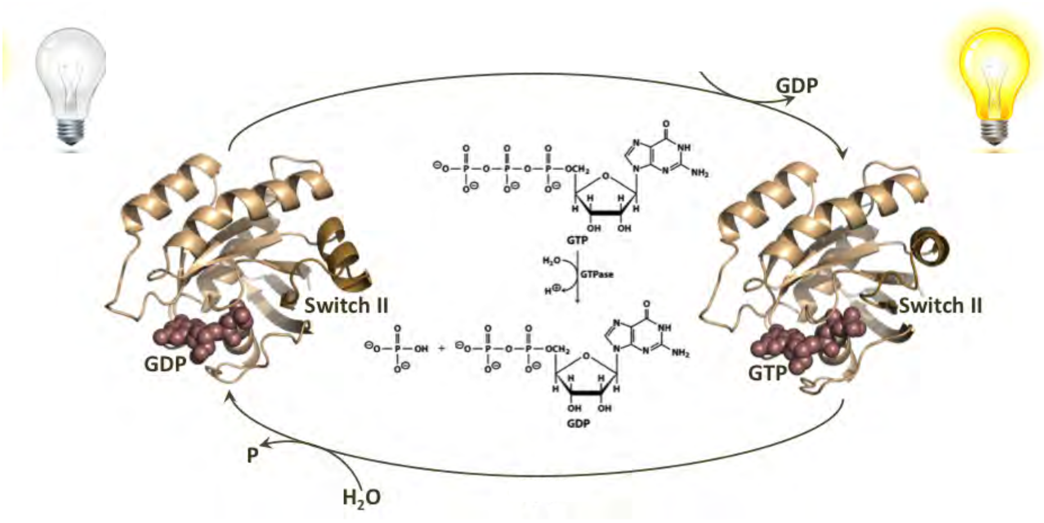


FIGURE 3: GTPases switching mechanism.

is devoted to the study of GTPases from the Ras family from a chemical physics perspective. More specifically, the highly conserved G domain of NRas isoform is investigated within the field of computational structural biology [39]. Indeed, the mechanisms employed by members of the Ras family to catalyze GTP hydrolysis remain to be elucidated. This is of major importance since point mutations of Ras specific residues are found in more than 30% of human tumours [56]. Moreover, these mutant proteins display a reduced hydrolysis rate [45, 73] indicating that they remain locked in the GTP bound state. Hence it appears that high concentrations of this active conformation [15] play a major role in bestowing cells with capabilities characteristic of the dissemination of cancerous tissues. As no suitable treatment has yet been found for overcoming the proliferation of such malignancies, and as restoring hydrolysis within the mutated forms could be an alternative, the detailed study of the active site appears to be crucial. This is done in an attempt to determine:

How NRas Gln 61 mutants impair GTP hydrolysis in the protein active site?

To answer this question, in a first part, an overview of Ras proteins is given. Their role within the cell is presented. Key structural features that are essential to Ras function are introduced, as well as the suggestions that have emerged concerning its catalyzing effects. In a second place, the methods used to study the active site of GTP bound NRas are described. Each of them enables to gain new insights into different effects of the protein through different levels of accuracy employed. In the third and last part, the obtained results are presented for wild-type and mutant NRas. The structural characteristics of the active site are assessed. The associated water distribution is determined. The ligand (GTP) intrinsic properties are evaluated. To finish, the characterization of three engineered double mutant proteins is presented.

Chapter 1

Ras proteins: key role, conserved structure

Genes encoding Ras proteins were first identified by Edward M. Scolnick and colleagues in a study of cancer related viruses discovered within rats' sarcomas by Jennifer Harvey and Werner Kirsten in the 1960's. These genes, namely HRAS and KRAS, were then found within human cancerous tissues. They encode 21 kDa proteins (p21) of the same name *i.e.* $p21^{H-ras}$ and $p21^{K-ras}$ that we will refer to as HRas and KRas, respectively. These proteins, along with NRas, belong to the Ras family, itself part of the Ras superfamily of small GTPase proteins.

1.1 Ras position in the cell machinery

Ras proteins are taken as the representatives of small GTPase proteins. They are monomeric subunits constituted of 189 amino acids that fold into a tertiary structure containing a single G domain (see figure 1.1). The overall 3D structure can be divided in two main domains.

The first, the catalytic domain, consists in residues $1-165$ of the protein sequence. As inferred from its denomination, residues forming Ras active site, where GTP hydrolysis takes place, are all found within this portion of the sequence. This domain can be further divided in two regions: the effector lobe (residues $1-86$) and the allosteric lobe (residues $87-165$). In the former, residues that take part in interactions with other biomolecules are found. They are 100% conserved in HRas, KRas and NRas. The latter, the allosteric lobe, contains regions that bind specific membrane components. It displays 90% of residue identity within Ras family members.

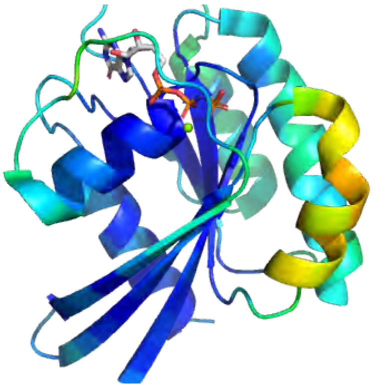


FIGURE 1.1: GTP bound Ras 3D structure (PDB entry: 1QRA).

The second domain is designated as the hyper-variable region. It consists in residues $166 - 189$ and, as inferred from its name, contains a small number of identical residues within the family members. Nevertheless, an important ending *motif* is conserved: CAAx. In this four amino-acid sequence, C represent a cysteine, A an aliphatic amino acid (alanine, isoleucine, proline or valine) and x is either a methionine or a serine. It is recognized by farnesyltransferases that catalyze the prenylation of a farnesyl group to the conserved cysteine residue. The addition of this isoprenoid is critical for anchoring Ras proteins to various cellular membranes.

As briefly described in the introduction, Ras proteins, like other GTPases, function as molecular switches. They cycle between an active GTP bound state and an inactive GDP bound state. The conformational changes Ras undergoes in this process are mainly triggered by the hydrolysis reaction (see equation 1.1) that takes place within its active site and by the nucleotide exchange upon it.

$$GTP + H_2O \longrightarrow GDP + P_i \tag{1.1}$$

Mutations of Ras key residues, namely Gly 12, Gly 13 and Gln 61, are found in more than 30% of human tumors [56]. These mutant proteins, display a drop in the intrinsic hydrolysis rate [45, 73]. However, from the very beginning of the investigation of Ras implication in tumors formation, *in vitro* experiments showed that the sole impaired intrinsic hydrolysis within mutant proteins could not be directly correlated to the transforming potential that cells are bestowed with *in vivo* [15]. In fact, for a long time, Ras cellular function remained mysterious [15] even though its ability to bound GDP/GTP lead to the speculation that it was a signal transducer, because of similarities with α subunits of heterotrimeric G proteins [1].

---

[1] Well known proteins that belong to the extended family of GTPases and that transduce sensory and hormonal stimuli across the plasma membrane

In 1987, Meg Trahey and Frank McCormick [75], in an experimental study of wild-type NRas and two mutated proteins at site 12 (G12N and G12V), showed that Ras biological activity depends on the bound nucleotide and is regulated by other enzymes. Indeed, both mutant proteins were predominantly associated to GTP and induced maturation of *Xenopus* [2] oocytes, whereas the non-mutated form remained inactive. More importantly, they found that the wild-type intrinsic hydrolysis leading to its GDP bound state was stimulated by more than a 200 fold by a cytoplasmic protein they named GTPase Activating Protein (GAP). It appeared that Ras function in cellular proliferation was regulated through interactions with another enzyme. Oncogenic mutations appeared to escape the regulatory process leading to the accumulation of active GTP bound complexes. This finding triggered the search for analogous biomolecules that would activate normal Ras.

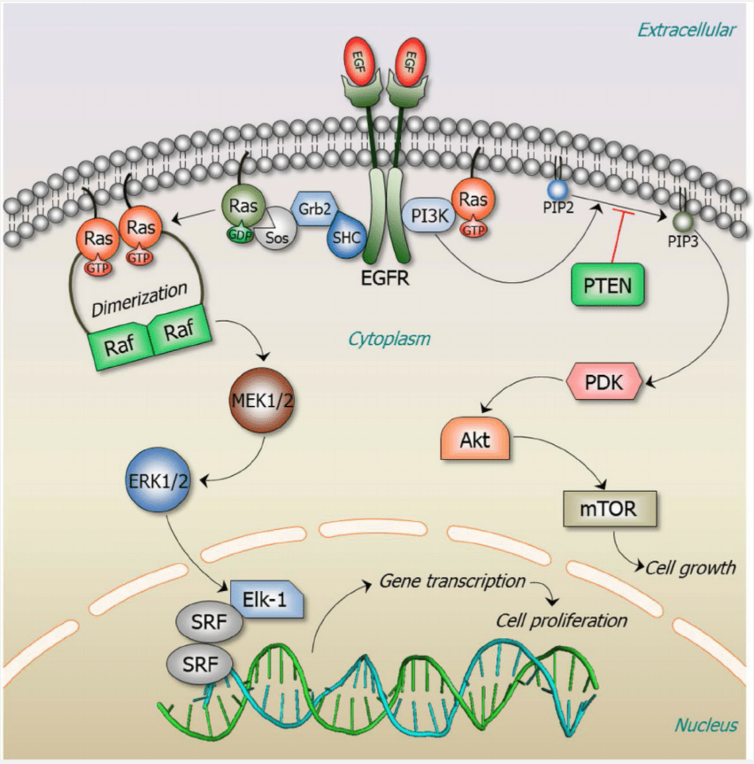
Ever since, a myriad of biomolecules have been identified as upstream and downstream regulators of members of the Ras superfamily [6, 7, 12]. More importantly, it is now clear that Ras cellular function is determined by its affinity with different effector molecules [12, 44, 79]. The exceptional transforming properties that Ras single mutations bestow cells is explained by the numerous enzymes it interacts with and that accomplish a large panel of cellular functions.

Concerning self-sufficiency proliferation, typical of cancerous tissues, Ras proteins take part by transmitting growth signals to effector molecules that control cell replication, differentiation and survival. This signal transmission is illustrated in what follows, taking as an example only one effector pathway [12, 79] that is implicated in these critical tasks within the cell. Nevertheless, it is important to bear in mind that no less than eleven biomolecules have been identified as Ras effectors. Moreover, these numerous signaling pathways are interconnected [12] and might influence each others activity in manners that remain to be revealed.

From the extracellular medium, growth signals are transmitted inside the cell by the binding of growth factors to transmembrane proteins commonly named receptor tyrosine kinases (RTK). The binding at the protein extracellular domain, provokes reactions of dimerization and autophosphorylation of tyrosine residues at the protein inner cytoplasmic domain. These reactions result in the creation of specific binding sites at the inner surface of the plasma membrane. Adapter proteins Grb2, bound to guanine nucleotide exchange factors (GEFs) in the plasma, recognize these binding sites. Through the binding to one of these sites, a Grb2 can translocate a GEF (like SOS for exemple) to this part of the cell. The latter, as inferred by its name, catalyzes the exchange of GDP

---

[2] Aquatic frogs commonly studied as model organisms



Copyright ©: R. Nusinov et al, Oncotarget (2014)

FIGURE 1.2: Ras position in the cellular machinery. EGFR is one of many RTK.

by GTP. Once located in the plasma membrane, GEFs can interact with Ras, activating it by promoting the release of GDP and the binding of GTP. In the active state, Ras is in turn able to interact with its effectors. For inducing cell replication, differentiation and survival for example, Ras interacts with a kinase named Raf, that in turn activates a subsequent kinase named MEK, which activates another kinase named ERK. The latter is translocated to the nucleus of the cell where it activates either other kinases as Rsk or transcriptor factors as Elk. The signal is terminated by GTP hydrolysis that returns Ras to its inactive GDP bound state, unable to interact with any of its effector molecules.

Ras activity is hence determined by *i)* the rate of its activation, determined by GEFs, *ii)* the engaged interactions with diverse effector molecules, determined by its location within the cell, and *iii)* the rate of its inactivation, determined by both its intrinsic hydrolysis rate and the enhancement of this reaction by GAPs. Ras oncogenic mutations hinder GTP hydrolysis, locking thereby the protein in a GTP bound active state that is hence able to indefinitely interact with its numerous effectors. In this manner, cells are bestowed with capabilities such as limitless replicative potential in the absence of external growth signals. Moreover, concerning the Raf signalization pathway given as an example, it has been observed that within the wild-type, GAP is unable to enhance the hydrolysis rate while Ras remains bound to Raf [79]. Thus, if restoring sensitivity towards GAP action within the mutant proteins appears as a major requirement for diminishing the concentration of GTP bound active Ras within the cell, recovering the intrinsic catalyzing power is also essential for inactivating certain pathways.

In what follows, Ras structural components that play a major role in its intrinsic GTPase capability as well as in its affinity for different effector and regulator molecules are presented.

1.2 Ras key regions

A major step in understanding both Ras intrinsic and GAP enhanced catalysis of GTP hydrolysis was the determination of Ras [30, 50, 55, 63] and Ras-GAP [62] crystal structures. Prior to this achievement, several parts of the protein sequence had nevertheless been identified as playing a major role in its function, either by directed mutagenesis experiments, or by homology studies of the protein within the more extended family of GTPases. For assessing dynamic properties of Ras in solution, FTIR [3, 4, 14, 18, 34, 35], NMR [21, 36, 71] and Raman [77] spectroscopy methods have also been employed. Although the structural features are inferred in an indirect manner, the advantage of the three latter methods over X-ray spectroscopy arises from the circumvent of constraining the protein by crystal packing forces. Harvesting information from all these analysis, key Ras sequence - structure - function relationships are presented in this section in a non exhaustive manner.

The first region of Ras proteins that is essential to their GTPase function is the *P-loop*, also referred to as loop L1. It extends from residue 10 to 17 (see figure 1.3). These residues are implicated in binding the phosphate moiety of the nucleotide. Indeed, this negatively charged part of the ligand is stabilized by a positively charged potential created by all the backbone amide groups forming the loop, the Mg^{2+} ion and the last amide group of Lys 16 residue side chain. They all take part in hydrogen bonds with GDP/GTP α, β and γ phosphate groups. These binding pattern has been preserved over evolutionary time as the the motif GxxxxGK(S/T), contained in the *P-loop*, is highly conserved among GTPases.

Complementary to this region, in the sense of binding both GDP and GTP, are sequences 112-119 and 143-147 that contain residues that stabilize the binding of the guanidinium moiety of the nucleotide by hydrophobic interactions and hydrogen bonds.

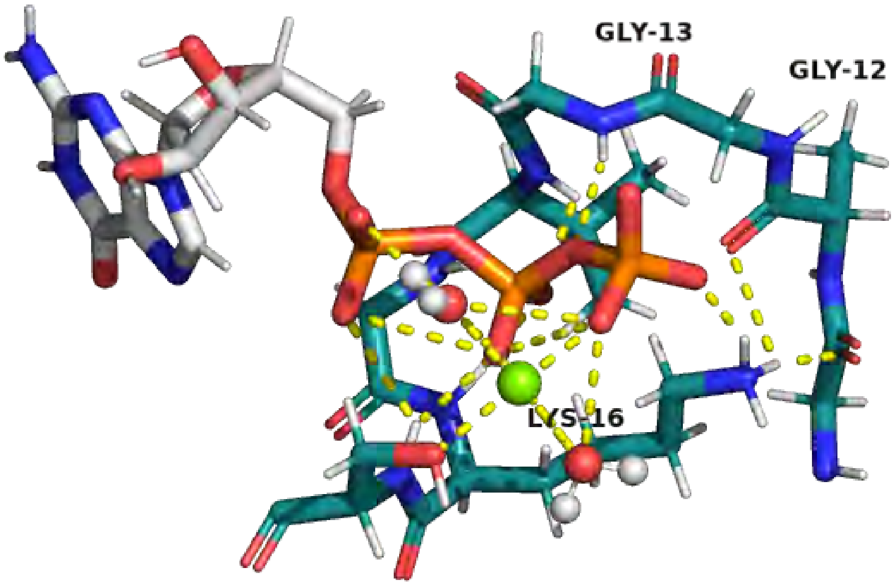


FIGURE 1.3: View of Ras *P-loop* (residues 10-17) while GTP bound.

A second region of major importance comprises residues 30 to 40 of the protein sequence and is called the *Switch I* (see figure 1.4). It is also referred to as *effector region* and contains loop L2. It has been shown to be crucial in the establishment of interactions between Ras and its effector molecules, thus in Ras function. Indeed, mutations in this region, for example D38A, introduced in oncogenic Ras, bring to an end the protein transforming potential [26]. Actually, the *Switch I* region is highly flexible providing the protein structure with dynamic properties which are fundamental for the formation of macromolecular complexes. Not only Ras crystallographic structures present a poor resolution of this region [30, 50, 55, 62, 63] but also, Kraulis et al NMR spectroscopy investigation [36] has confirmed that this region has significant rapid internal motions on the nanosecond time scale. Moreover, Geyer et al [21] using NMR spectroscopy as well, have shown the existence of two main states of the *Switch I* region that determine Ras affinity for its effector molecules. Tyr 32 residue conformations characterize the state assumed. In state 1, Tyr 32 side chain lays far from the bound nucleotide while in state 2 the hydroxyl group hydrogen bonds with GTP γ phosphate group. Effector molecules preferentially interact with Ras when the *Switch I* is in state 2. The only conserved residue of this region in all GTPases, Thr 35, has been pointed out as the one triggering the conformational change that leads to state 2. Spoerner et al [71] have precisely shown that the methyl group of Thr 35 triggers this conformational change. State 2 is then stabilized by the coordination of Thr 35 side-chain hydroxyl group to the cation. Although the *Switch I* is involved in major interactions of Ras with its effector molecules, residues laying outside this region also contribute to binding and affinity

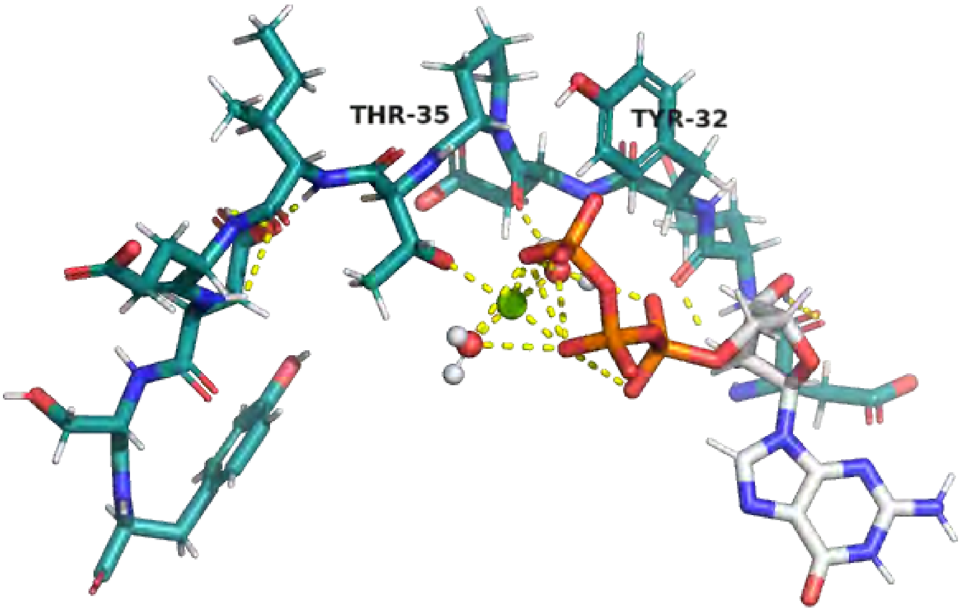


FIGURE 1.4: View of Ras *Switch I* region (residues 30-40) while GTP bound.

towards these molecules. For instance, mutations of residues 57 and 59 in HRas inhibit binding to full length Raf.

The third and last region of Ras proteins playing a fundamental role in its function, is the *Switch II* (see figure 1.5) that contains loop L4. It is formed by residues 57 to 67 which are also bestowed with genuine intrinsic mobility according to Kraulis et al [36]. This region contains the DxxG motif, conserved in all GTPases. The common aspartic acid at position 57 within Ras, coordinates the Mg^{2+} ion through a water molecule [36]. Gln 61, the residue that is the most subject to debate in this region and in the entire protein, has been identified as playing a major role in both intrinsic and GAP enhanced GTP hydrolysis. Indeed, Der et al [17] site directed mutagenesis experiments showed that the replacement of the glutamine at position 61 by seventeen other amino acid residues has no impact on nucleotide binding affinities but result in a reduced GTP hydrolysis rate. The various scenarios this residue has been consequently implicated in because it locks Ras in an active GTP bound state, are discussed in the next section.

Both GEF and GAP proteins that regulate Ras activation and inactivation, respectively, make contacts with the *P-loop*, the *Switch I* and *Switch II* regions. As neither GEFs, nor GAPs, have conserved structures [7], they approach Ras proteins in diverse manners. Nevertheless, they manage to recognize Ras conserved regions in order to perpetrate their respective functions.

GEFs [7] catalyze the nucleotide exchange by decreasing the affinity of Ras for GDP. This is achieved by deforming the active site in a manner that the Mg^{2+} ion is occluded

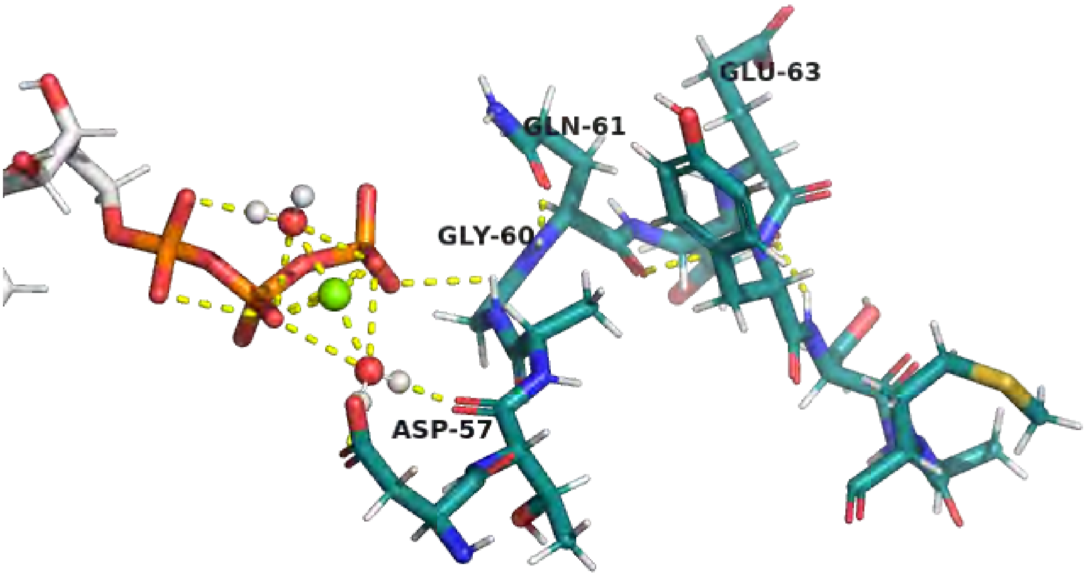


FIGURE 1.5: View of Ras *Switch II* region (residues 57-67) while GTP bound.

by either GEF residues or by the alanine side chain of the conserved DxxG motif from *Switch II* region. Glu 62 residue, conserved in almost all small GTPases, facilitates the nucleotide release by electrostatic repulsions with GDP terminal phosphate group. Subsequently, the higher concentration of GTP within the cell, lead to its binding that in turn decreases the affinity between GEFs and Ras, promoting thereby the dissociation of the protein complex.

GAPs action has been open to debate, alike the role of Gln 61 residue, certainly because the enzymatic GTP hydrolysis remains undetermined. Some of the propositions that have emerged will be presented in the next section. Nevertheless, a general consensus has been reached through experimental studies. Indeed, the first crystallographic structure of the Ras-GAP complex [62] has brought substantial information concerning the interactions between the two proteins. It appears that the overall structures of both proteins in the complex are similar to the isolated ones, discarding a major allosteric effect that could be induced by GAP. Nevertheless, GAP appears to stabilize the *Switch II* region of Ras as it is well defined in contrast with structures of isolated Ras. Moreover, GAP provides an arginine residue, Arg 789, inserted in the active site that interacts with the γ phosphate group. Mutational studies [2] of this *arginine finger*, have determined that this GAP residue is the one that increases the reaction rate by up to 5 orders of magnitude. It participates in Ras-GAP binding but is not fundamental for protein recognition.

The implication of these structural elements in catalyzing GTP hydrolysis is presented in the next section within the various scenarios that have emerged.

1.3 Mechanisms for GTP enzymatic hydrolysis

Although the intrinsic hydrolysis activity of Ras proteins is low, making them poor GTPases, the further diminution of the reaction rate by oncogenic mutations is dramatic. Moreover, as demonstrated by Meg Trahey and Frank McCormick [75], these mutations also hinder the hydrolysis rate enhancement provided by GAPs. The unambiguous implication in tumors formation of high concentrations of active GTP bound Ras within the cell, has placed Ras proteins at the centre of numerous experimental and theoretical studies. The need of Gln 61 residue demonstrated by Der et al [17] site directed mutagenesis experiments, has lead to several propositions concerning Ras and Ras-GAP catalyzing mechanisms in general, and Gln 61 residue implication in particular. This section is devoted to the description of the dominant alternative mechanisms that have emerged. An initial part presents Ras and Ras-GAP effects on GTP upon its binding. These effects contribute to catalyzing GTP hydrolysis by the solely formation of the protein-ligand complex.

1.3.1 Ras and Ras-GAP effects on GTP properties

As protein function is intrinsically related to the dynamic properties of its structure, numerous experimental studies have resorted to methods that allow to get insight on Ras and Ras-GAP catalyzing mechanisms without perturbing protein structure.

One such method is time-resolved Fourrier Transformed Infra Red spectroscopy (FTIR). This method also makes it possible to follow GTP hydrolysis as it occurs. Using it, numerous studies have gain insight on Ras [3, 14, 34] and Ras-GAP [4, 35] catalyzing effects. Indeed, the infrared (IR) spectra of both GDP and GTP present major differences in the environment provided by the protein active site compared to their respective spectra in solution. Not surprisingly, the first difference observed is the sharpening of the bands associated to the nucleotide, indicating that its degrees of freedom are reduced upon binding to the protein.

Concerning Ras effects [3, 14], its active site induces a downshifted of the vibration associated to GTP β phosphate group. This indicates a decrease in $P-O$ bond order within the group. It is due to strong interactions of O_β atoms with the protein binding

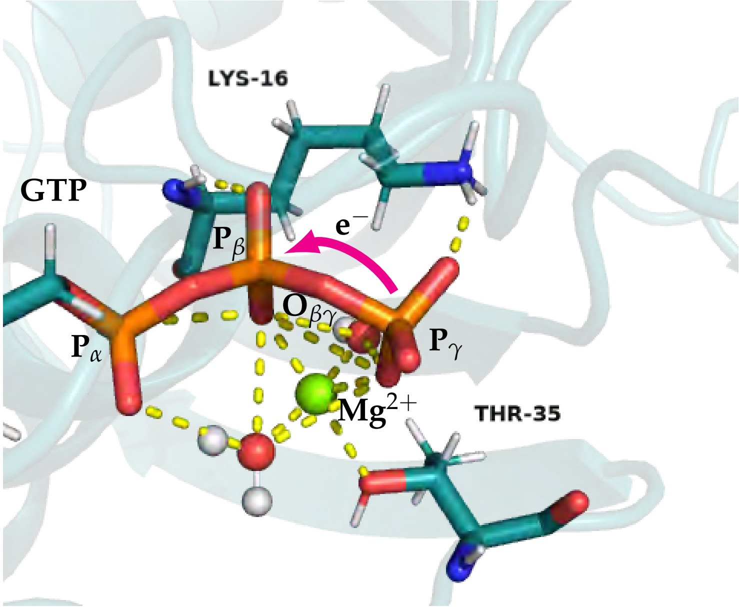


FIGURE 1.6: Ras induced charge transfer from the γ to the β phosphate group of GTP.

site that provides a positively charged environment, notably through the backbone nitrogen atoms of the *P-loop* described above. As a consequence, the β phosphorus atom becomes electron deficient. It retrieves then electrons from the $O_{\beta\gamma}$ bridging atom, weakening thereby the bond between the β and γ phosphate groups. Moreover, the vibration associated to the γ phosphate group is upshifted, indicating an increase in $P-O$ bond order within the group. This reflects that the γ phosphate group binding to Ras active site is weaker than that of the β phosphate group. It appears that compared to GTP in solution, within the protein active site, there is a decrease in negative charge at the γ oxygens and an increase at the β oxygens. Hence, Ras induces a charge shift from the γ to the β phosphate group (see figure 1.6), stabilizing thereby a GDP like state. This charge shift constitutes a major catalyzing effect. It occurs independently of the associative or dissociative nature of the enzymatically catalyzed GTP hydrolysis reaction.

The spectra associated to GDP reveals moreover that the β oxygen atoms remain strongly coordinated to the protein upon hydrolysis. Hence, the binding site strongly stabilizes this group. A final reaction detail brought to light by this spectroscopic study, is the observation of two bands characteristic of singly protonated inorganic phosphate. The products of the reaction then probably involve: protonated GDP + HPO_4^{2-}.

A subsequent study of Ras-GAP [4] employing the same techniques, showed that GAP further promotes the charge shift leading to a GTP charge distribution that resembles that of the GDP product. Indeed, the nucleotide β phosphate band is downshifted

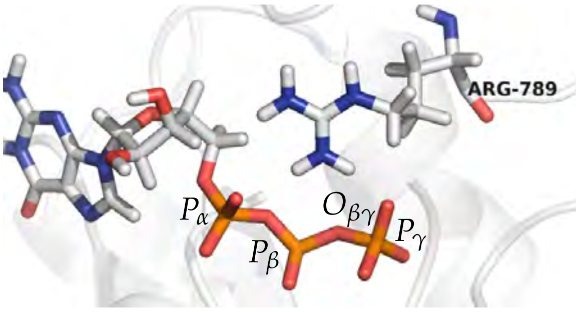


FIGURE 1.7: GAP *arginine finger* positioning within Ras active site (PDB entry: 1WQ1).

by $79cm^{-1}$ in the Ras-GAP complex while it was downshifted by $28cm^{-1}$ only when bound to HRas alone. The *arginine finger* could be at the origin of such enhancement through interactions with the β oxygen atoms (see its positioning in figure 1.7). Lys 16 residue could also play this role through a conformational change induced by GAP. In this study it is moreover determined that bond breakage by the lytic water molecule accompanied by the inorganic phosphate formation is not the rate limiting step of the reaction. HPO_4^{2-} release is. Indeed, its IR bands indicate that it is tightly bound to the protein complex since they are phosphorylation like.

To gain insight into GTP geometric characteristics that lead to its charge distribution upon binding to Ras, QM/MM calculations were performed within the same research group [32]. The obtained results were validated by comparing the theoretical IR spectra to the experimental one previously introduced [14]. It appears that upon binding to HRas, $P_\beta - O_\beta$ bond lengths are increased, reflecting the experimentally determined decrease in bond order within the β phosphate group. More importantly, $P_\beta - O_{\beta\gamma}$ and $P_\gamma - O_{\beta\gamma}$ bond lengths are also increased. The latter is accompanied by a decrease in the $P_\gamma \widehat{O_{\beta\gamma}} P_\beta$ angle and charge shifts from P_γ phosphorous atom to $O_{\beta\gamma}$ oxygen bridging atom. The charge accumulated in the β phosphate group is provided by both the α and γ phosphates but also by the CH_2 group modelling the guanidium moiety. Lys 16 residue and the Mg^{2+} ion appear to help these charge transfers.

Another method enabling to get details on how an enzymatic reaction proceeds without disrupting protein structure, is Raman spectroscopy. Wang et al [77] have resorted to this method to study GTP hydrolysis within HRas. They reach the same conclusion as FTIR studies, that is: Ras tightly binds both GDP and GTP β phosphate groups further stabilizing a GDP-like state of GTP prior hydrolysis. Indeed, the vibrational frequency associated to it in the Raman spectra is also downshifted. The study is coupled to *ab initio* calculations at the HF/3-21g* level to gain insight on this observation. It appears that the Mg^{2+} ion positioning is at the origin of the downshift of the β phosphate group vibration.

Copyright ©: Rudack et al, *Biophysical Journal* (2012)

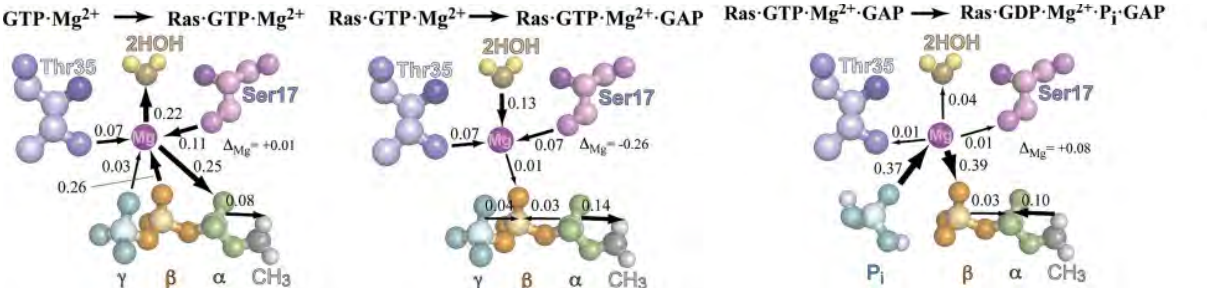


FIGURE 1.8: The role of the Mg^{2+} ion in the enzymatic GTP hydrolysis.

To further investigate the role of the cation, Rudack et al [60] study in detail the charge distribution of the bound nucleotide and the Mg^{2+} ion within the protein active site through QM/MM simulations. The results are validated by comparing the calculated infrared spectra to the experimental ones [3, 4]. It appears that already in solution, the cation drags and stores electrons from GTP γ phosphate group rendering it more positive. Binding to Ras and Ras-GAP further enhances this flow, rendering the terminal phosphate even more positive and reducing thereby the energy barrier of the reaction (see figure 1.8). The β phosphate group is rendered more negative by the intruding arginine finger of GAP. After $P_\gamma - O_{\beta\gamma}$ bond breakage, the electric charge stored in the Mg^{2+} ion is returned to the β phosphate group.

The geometric details associated to these charge shifts are assessed in a subsequent study [59]. It appears that from a tridentated coordination in solution, the Mg^{2+} ion becomes bidentated upon binding to Ras. This induces a rotation of both γ and β phosphate groups such that they adopt en eclipsed conformation relative to each other. The increased electronic repulsion between non-bridging oxygen atoms in such conformation provokes the elongation of the $P_\gamma - O_{\beta\gamma}$ bond. GAP binding to Ras does not change the coordination of the cation but Arg 789 residue induces similarly a rotation of the α phosphate group in a manner that the α and β phosphates also become eclipsed. Hence, in the Ras-GAP complex all three phosphates are in an eclipsed position, further elongating the $P_\gamma - O_{\beta\gamma}$ bond. It is postulated that this energetically unfavorable eclipsed conformation contributes to catalysis because it destabilizes the reactant state of GTP thereby pushing it to adopt the conformation of the transition state.

In a later FTIR study, Kötting and Gerwert [34] determine that the HRas induced diminution of the activation energy in GTP hydrolysis is mainly enthalpic. Indeed, of the main catalytic effect $\Delta\Delta G^* = 5.8$ kcal/mol, 5.2 come from the enthalpy difference

Copyright ©: Kötting et al, *PNAS* (2008)

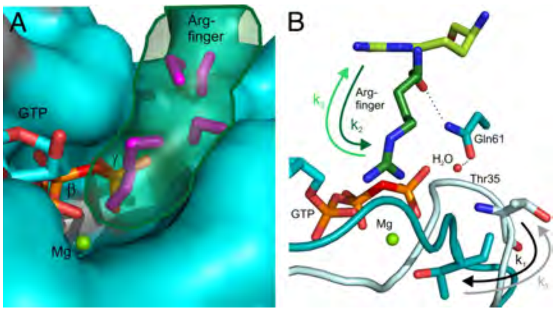


FIGURE 1.9: The *arginine finger* movement in the catalytic pocket.

($\Delta\Delta H^*$) while the entropic contribution is of $\Delta\Delta S^* = 0.6$ kcal/mol only. As their previous time resolved FTIR studies of the reaction within HRas [14] and HRas-GAP [4] showed that GAP binding promotes a greater charge shift towards the β phosphate group, they propose that the *arginine finger* achieves this by a movement towards this group of the nucleotide.

To assess if such a movement takes place, they subsequently study HRas-GAP complex [35] using the same experimental method. Three consecutive events are identified, depicted in figure 1.9. The first event, constitutes a conformation change of Ras *Switch I* region from an incompetent to a competent GTPase state. These two states certainly correspond to the ones identified by Geyer et al [21] NMR spectroscopy study. GAP binding has no influence on the rate of this step. The second event is associated to the introduction of GAP Arg 789 residue into the binding pocket together with $P_\gamma - O_{\beta\gamma}$ bond breakage. Because the method is unable to kinetically separate both events, it is assumed that hydrolysis occurs immediately after the *arginine finger* is positioned. Finally, the third event, corresponds to inorganic phosphate release, return of *Switch I* to the incompetent state (state 1 according to Geyer et al [21]) and exit of GAP Arg 789 from Ras active site. This last step constitutes the rate limiting step of the reaction. As the activation enthalpy associated to GTP hydrolysis within HRas-GAP complex is similar to that determined within HRas alone [34], it is concluded that GAP must dramatically enhance the reaction rate (up to 10^5 times) by an entropic contribution. The *arginine finger* is expected to make such a contribution by moving highly structured water molecules from the protein active site to the bulk solvent. Water removal would moreover change the dielectric constant of the binding pocket, thereby increasing the charge shift towards GTP β phosphate group depicted in figure 1.6.

As the catalyzing effect that consists in promoting a charge shift from the γ to the β phosphate group is accompanied by an increase in the bond length of the bond to be broken ($P_\gamma - O_{\beta\gamma}$), this mechanism is consistent with a more dissociative reaction

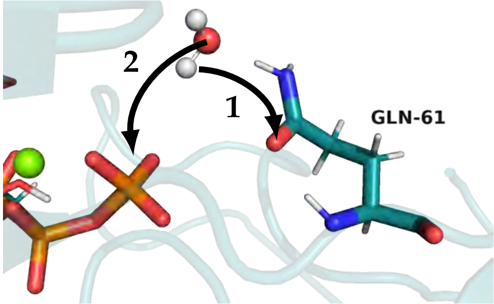


FIGURE 1.10: Subsequent steps in reaction scheme: Gln 61 is the general base of the reaction.

scheme for the enzymatic GTP hydrolysis. Nevertheless, the negative charge accumulation on the β phosphate group that leads to a less charged γ phosphate group would also facilitate a nucleophilic attack on the latter in a more associative scheme. Thus, Ras and Ras-GAP might use features from both associative and dissociative reaction schemes although the associative/dissociative nature of the reaction has been subject to a heated debate. The major propositions that have arisen are presented in what follows.

1.3.2 Gln 61 as the general base of the reaction

From the first high resolution crystallographic structure of HRas [50], resolved at 1.35 Å, emerged the primary role proposed for Gln 61 residue: that of acting as the general base of the reaction. Indeed, in this structure of HRas bound to a slowly hydrolyzing GTP analogue GppNp, Gln 61 appears to contact one of the two water molecules (numbered 175 and 189) present in the active site and that could chemically participate in the reaction. As water molecule 175 is placed opposite to the $\beta - \gamma$ bond, it is the one among the two identified as the lytic water molecule. In the proposed mechanism, Gln 61 would *activate* water 175 by abstracting one of its protons. A hydroxyl ion would thus be formed, and would subsequently attack GTP γ phosphorous atom (see figure 1.10). The proton abstraction by Gln 61 residue would be assisted by Thr 35 and Glu 63 residues. The former could polarize the water molecule through hydrogen bonds while the latter could orient Gln 61 and/or increase its proton withdrawing potential. The rate limiting step of GTP hydrolysis would be a conformational change of the very mobile *Switch II* region.

In support of this scenario, Krengel et al [37] propose that within HRas oncogenic mutants the enzymatic reaction is impaired by:

i) Steric hindrance for Gly 12 mutants,

ii) The chemical nature of the substituent residue for Gln 61 mutants accompanied by a loss of flexibility of the *Switch II* region.

Indeed, the crystallographic structures of oncogenic HRas, resolved at 2 - 2.6 Å, reveal that the overall structural changes are very small. For G12R and G12V, the only differences observed in the *P-loop* are the longer side chains of the substituent residues that push away water 175 and Gln 61 side chain. For Q61H and Q61L, the replacing histidine and leucine have no proton abstraction potential and the *Switch II* region appears to be less flexible.

To evaluate the feasibility of this reaction scheme inferred from crystallographic data, Langen et al [38] in a computer simulation study, evaluate the associated energy. Their results exclude that Gln 61 could play such a role. Indeed, solely the proton transfer step yields a higher energy barrier than that of the overall reaction (30 vs 23 kcal/mol, respectively). Moreover, it appears that Glu 63 stabilizing effect of the protonated Gln 61 through electrostatic interactions is shielded by water molecules.

This last study triggered the use of computational methods for studying GTP hydrolysis within the active site of Ras and Ras-GAP complex. As a result, diverse reaction schemes emerged, further opening the starting debate concerning Gln 61 role in the enzymatic reaction.

Indeed, in a subsequent theoretical study of wild-type and Q61E mutant HRas [20], the side chain groups of both Gln and Glu at position 61 interact with the presumed lytic water molecule placed near residue Thr 35. Such arrangement supports the "Gln 61 as the general base of the reaction" mechanism. Moreover, these results were coupled to experimental characterizations of Q61E mutant hydrolyzing capability. Indeed, if residue 61 takes part in the enzymatic hydrolysis by activating the lytic water molecule, a glutamic acid at this position should enhance the rate of the reaction as its carboxylate side chain is a better proton abstractor (*i.e.* has a higher pK_a) than the carboxamide side chain of wild-type glutamine. The intrinsic activity of Q61E is reported to be indeed higher than that of the wild-type protein. Precisely, GTP hydrolysis is reported to proceed 20 times faster within Q61E than within wild-type HRas in this study.

A similar mechanism in which Gln 61 takes action in the enzymatic GTP hydrolysis by abstracting a proton from a water molecule is presented in the next section.

Copyright ©: Grigorenko et al, *PROTEINS: Structure, Function and Bioinformatics* (2005)

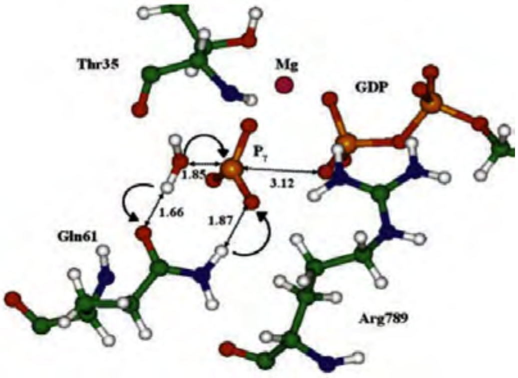


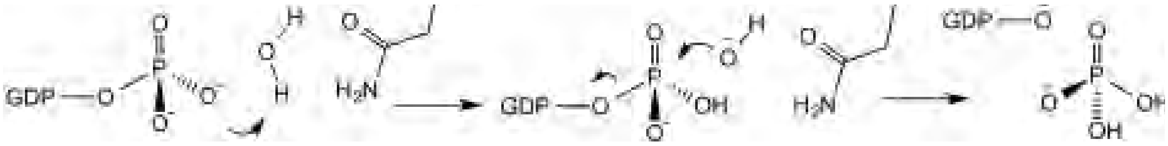
FIGURE 1.11: The proton shuttle mechanism.

1.3.3 Gln 61 in concerted proton transfers

Through QM/MM molecular dynamics simulations, Grigorenko et al [23, 24] present a new mechanism in which Gln 61 residue also participates chemically in the reaction. These investigations of the protein complex propose that GTP $P_\gamma - O_{\beta\gamma}$ bond is cleaved first. Subsequently, Gln 61 residue simultaneously abstracts a proton from the lytic water molecule while conceding a proton from its terminal amide group to an oxygen atom of GTP γ phosphate group (see figure 1.11). These concerted proton transfers lead to a product state consisting of GDP, $H_2PO_4^-$ and the imide form of Gln 61 residue. The latter energetically unfavorable species is postulated to become energetically favorable over the amide form in the presence of the doubly protonated inorganic phosphate. As in this scenario the protonation of Gln 61 residue is transitory, it is considered to participate in the enzymatic hydrolysis of the reaction by serving as a *proton shuttle*.

Assuming that GTP hydrolysis proceeds through the same reaction pathway in both Ras [23] and Ras-GAP [24], they suggest that the arginine finger enhances the reaction rate by limiting the degrees of freedom of the reactants such that the hydroxyl ion can only proceed to attack GTP γ phosphate group in the Ras-GAP complex.

Finally, in a later study carried out by the same group [31], the complete reaction pathway is presented. That is, a few steps are added enabling to restore the amide form of Gln 61 within Ras active site. The rate limiting step then appears to be the tautomeric reaction Gln 61 is involved in and that enables the active site to recover its pre-hydrolysis form. Using transition state theory, the rate constants governing GTP enzymatic hydrolysis are inferred from the energy barriers associated to the various transition states identified in this reaction pathway. They correlate remarkably well with experimental results. In this manner, the Gln 61 as a proton shuttle mechanism



Copyright ©: Carvalho et al, *Archives of Biochemistry and Biophysics* (2015)

FIGURE 1.12: Substrate assisted catalysis: Gln 61 stabilizes the transition state.

has managed to survive despite having being strongly criticized due to its complexity.

In subsequent scenarios, Gln 61 residue does not participate chemically in the reaction *i.e.* none of its bond are broken, nor new are formed.

1.3.4 Gln 61 stabilizes the transition state

As the first role speculated for Gln 61 residue, another role of this residue has been proposed from crystallographic data. Indeed, Privé et al [55] suggest that as the nucleotide binding pocket presents a localized positively charged region formed by the Mg^{2+}, Lys 16 residue and the main chain amide groups of residues 13-17, 35 and 60, catalysis of GTP hydrolysis by Ras proteins is achieved by lowering the energy barrier to reach the transition state of the reaction. Gln 61 role would be to stabilize the pentavalent phosphate intermediate. It is excluded that Gln 61 activates the lytic water molecule located near Thr 35 residue as mutants with higher proton affinities in this position impair hydrolysis to the same extent as poor proton abstractors. They suggest that if the lytic water molecule was to be activated, GTP could accomplish this task as it is a much more efficient general base.

Shortly after this suggestion, through computer simulations, Schweins et al [65] validate the reaction scheme in which GTP is the general base during its own hydrolysis. Indeed, the energy barrier associated to it according to their calculations (24,5 kcal/-mol) is close to the experimental value (23 kcal/mol) inferred from the measured reaction rate. They further support this mechanism by arguing that GTP γ phosphate group has the highest pK_a in Ras active site hence a proton transfer to it should be less energy demanding than the protonation of any other active site residue. Moreover, the protonated state would reduce the electrostatic repulsion between it and the attacking hydroxyl ion HO^-, thereby facilitating the formation of the pentacoordinated intermediate of the reaction.

In a subsequent study, through NMR-spectroscopy measurements at different pH values, this same group [68] monitors the titration of GTP in the environment provided by the protein to determine the pK_a value of its terminal phosphate group. They find that this value is different within the active site of wild-type and mutant HRas. Moreover, it appears that the pK_a of GTP γ phosphate group has a linear relationship with the logarithm of the rate of GTP hydrolysis, at neutral pH. Hence, the higher the pK_a of this group within the environment provided by the considered active site, the faster the reaction rate at neutral pH. These results strongly support that GTP itself is the general base of the reaction in a *substrate assisted catalysis* scheme (see figure 1.12). It appears that oncogenic mutations change the electrostatic properties of the protein active site. This in turn results in a diminished pK_a of GTP terminal phosphate group that leads to a reduced hydrolysis rate. In this scenario, the role of Gln 61 is to stabilize the transition state of the reaction through hydrogen bonds.

As some mutant proteins do not follow the same linear relationship between the pK_a and the logarithm of the reaction rate, in a subsequent publication [67], Schweins et al. argue that mutants that present the same slope correlating $ln(k)$ and pK_a affect the reaction in the same manner that is, by stabilizing or destabilizing the same resonant structure. Those with a different slope, stabilize or destabilize a different intermediate. Moreover, as the same linear relationship $ln(k) \propto pK_a$ is observed within the Ras-GAP complex, the substrate assisted catalysis is also assumed. GAP is proposed to enhance the reaction rate by increasing the pK_a of GTP γ phosphate group.

The reaction pathway associated to this reaction scheme is proposed [66] to proceed through four states linked by three events: *i)* proton transfer from the lytic water molecule to the γ phosphate, *ii)* formation of a pentacoordinnated intermediate by the attack of the hydroxyl HO^- ion on the terminal phosphate group and *iii)* bond breakage leading to the products GDP and doubly protonated inorganic phosphate $H_2PO_4^-$. The proton transfer is considered to be the thermodynamic bottleneck of the reaction while the nucleophilic step the kinetic bottleneck *i.e.* rate limiting step as it passes through the lowest probability configuration. It is suggested that even if the proton transfer does not involve the highest energy barrier, it contributes nevertheless to the overall rate of the reaction. This would explain why a change in the pK_a of the γ phosphate leads to a change in the reaction rate.

This reaction scheme *i.e.* substrate assisted catalysis is also supported by the crystallographic structure of Ras-GAP complex [62]. In this structure, previously discussed in section 1.2 describing Ras structural characteristics, the provided arginine residue, Arg 789, interacts with GTP γ phosphate group. It is then suggested that it enhances the reaction rate by stabilizing the charge developed on this group. Gln 61 residue takes part in catalyzing the reaction by stabilizing the transition state.

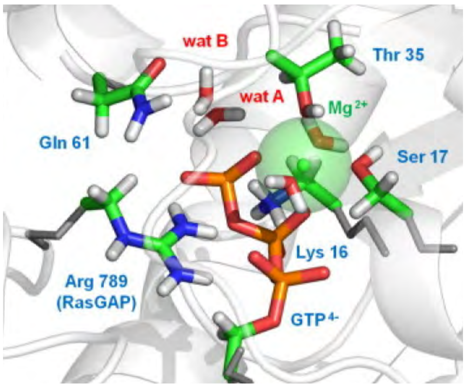


FIGURE 1.13: Two water molecules scheme. Gln 61 positions the lytic water molecule.

Nevertheless, a subsequent crystallographic structure of Ras has lead to further propositions concerning the role of Gln 61 residue.

1.3.5 Gln 61 positions the lytic water molecule

A fourth, yet not last, role for Gln 61 residue has emerged, again from crystallographic structure evidence. Indeed, Scheidig et al [63], in the crystallographic structure of GTP bound HRas presented, identified an additional water molecule (besides water 175 from Pai et al structure [50]) positioned in the vicinity of the γ phosphate group (see figure 1.13). Hence, they propose that two water molecules participate in the enzymatic hydrolysis. In the reaction scheme they present, a γ oxygen atom first gets protonated by this second water molecule, numbered 189. A hydroxyl ion HO^- is thereby formed. Changes of the electrostatic environment produced by this event, enable water 175 to get closer to the P_γ phosphorus atom. In this latter positioning, water 175 protonates the hydroxyl ion, thereby recovering water 189 while generating another hydroxyl ion. In a final step, the latter attacks GTP terminal phosphate group. Gln 61 in this scheme takes action by positioning the lytic water molecule, *i.e.* water 175, along with Thr 35 residue.

This mechanism is supported in a review [42]. The authors propose that Gly 13 residue, conserved in all GTPases, takes part in catalyzing the reaction by forming hydrogen bonds through its backbone amide group with the $O_{\beta\gamma}$ atom. In this scheme, Gln 61 residue, along with Thr 35, participates by correctly positioning the lytic water molecule.

Classical molecular dynamics of HRas [70] also suggests that Gln 61 residue orientates and polarizes these water molecules so that one of them can more easily transfer a proton to the GTP. Indeed, during the simulations, a conformational change triggered by the highly conserved Gly 60 residue moves Gln 61 away from the active site. Its interactions with the γ phosphate group, present in HRas crystallographic structure [50], are in this manner disturbed and bring it to the new positioning in which it interacts with the two water molecules.

Using QM/MM molecular dynamics, Martín-García et al [43] also support a scenario in which two water molecules are involved. However, they suggest that Gln 61 residue contributes to catalysis by stabilizing the transition state. Indeed, the latter comprises hydroxyl OH^- and hydronium H_3O^+ ions. They result from the protonation of one water molecule by the second one, leading to the formation of both ions. Subsequently, the hydronium ion protonates the GTP while the hydroxyl ion attacks the γ phosphate.

In a subsequent study using quantum based methods, Topol et al [74], propose a slightly different scheme for GTP hydrolysis within Ras-GAP complex but that nevertheless implicates Gln 61 in water positioning. Their study suggests that, first, the $P_\gamma - O_{\beta\gamma}$ bond is cleaved. Subsequently, the lytic water molecule reacts with GTP terminal phosphate group leading to the formation of a doubly protonated inorganic phosphate $H_2PO_4^-$. Thr 35 and Gln 61 residues take part by correctly positioning this lytic water molecule. GAP Arg 789 residue orients Gln 61 in a suitable position to fulfill its function.

A subsequent theoretical investigation of HRas-GAP complex [57], also supports a role for Gln 61 in positioning the lytic water molecule, considering one water molecule only, in a new reaction scheme. Indeed, in the course of the molecular dynamics, the lytic water molecule appears to interact with both Gln 61 and the backbone carbonyl oxygen of GAP Arg 789 residue. It is hence proposed that while Gln 61 positions this water molecule, the *arginine finger* activates it. It is moreover suggested that the side chain guanidine group neutralizes the charge accumulated on the nucleotide as it interacts with GTP α and γ phosphate groups. Finally, it is also speculated that GAP has an allosteric effect on Ras that promotes the conformation of the active site in which Gln 61 is near GTP γ phosphate group .

A final, more subtle, role of Gln 61 residue has been suggested. It is presented in the next section.

1.3.6 Gln 61 has an indirect role

From the first investigation of Gln 61 residue implication in tumors formation, Der et al [17] suggested that this residue has no direct chemical implication in GTP hydrolysis. Indeed, the seventeen mutant proteins studied, despite having an equivalent reduced GTP hydrolysis rate, are associated to different transforming capabilities. It appears that hydrolysis rate cannot be quantitatively correlated to transforming potential. Hence, it is proposed that Gln 61 takes part in accelerating the reaction by ensuring an optimal catalytic conformation of Ras active site. In such a scenario, different substitutions of this residue result in different overall protein conformations. The various mutant proteins, because the catalytically optimal conformation of their active site has been lost, present an equally reduced hydrolysis rate. Yet, they have different affinities for regulator and effector molecules, thereby bestowing different transforming potencies on different Ras mutants.

In support to an indirect action of Gln 61 residue, NMR spectroscopy investigations [36] concluded that the intrinsic low [5] catalytic activity of Ras proteins is due to a small proportion of time the *Switch I* and *Switch II* regions spend in a catalytically optimal conformation. Moreover, Gln 61 substitutions influence *Switch II* mobility affecting it even more in adopting the catalytically active conformation. In this case, the action of GAP is suggested to be that of stabilizing the catalytically active conformation by limiting the mobility of these region, which is the case according to the first Ras-GAP crystallographic structure [62].

Furthermore, in a theoretical study of Ras-GAP complex [69] it is also deduced that Gln 61 stabilizes a pre-catalytic configuration of the active site. This conformation, further stabilizes the transition state than the ground state of the reaction. Because Gln 61 occupies a pivotal position in Ras active site, it is suggested that mutations of this residue destroy the pre-catalytic configuration such that the ground state of GTP hydrolysis is then more stabilized than the transition state.

Although crucial, Gln 61 implication in the enzymatic hydrolysis of GTP remains controversial. Even more, the associative/dissociative nature of the reaction remains undetermined. Indeed, to the date, even if some insight on Ras and Ras-GAP catalyzing effects has been gained, experimental methods are unable to assess Gln 61 precise role, nor to distinguish the associative/dissociative character of GTP hydrolysis reaction. As probably noticed all along this section, computational methods have become unavoidable for getting detailed insight at the molecular and atomic level. Indeed, experimental observations are systematically accompanied by theoretical investigations. For

X-ray spectroscopy, simulations allow to restore the dynamic properties of the protein complex that are lost during crystallization as a single conformation is captured. For FTIR and Raman spectroscopy, atomic details are gained by reproducing the structures that lead to similar spectra to the experimentally determined. For kinetic measurements, the transition states are determined and validated by correlating the activation energies to the reaction constants. Using this combined approach, diverse reaction pathways have been proposed, evaluated and either discarded or validated.

Besides the ongoing debate on both Gln 61 role and the associative/dissociative nature of the reaction, it appears that the number of water molecules implicated in the enzymatic hydrolysis of GTP also remains to be elucidated. One theoretical study has been devoted to answer this question [52]. Although it is concluded that the enzymatic reaction most probably involves two water molecules such that one assists the other in donating a proton to the bound nucleotide, it is stressed out that this proton shuttle does not contribute to catalysis, thus leaving the one water model valid.

It appears that despite the numerous studies that have been carried out to the date, a further characterization of the protein active site is still required. This work is hence devoted to the characterization, at the atomic level, of NRas active site. Indeed, mutations of this protein are present in one out of four malignant melanomas [56]. Moreover, more than 60% of these mutations concern Gln 61 residue that is replaced by, in decreasing occurrence, an Arginine (45.5%), a Lysine (35.8%), a Leucine (10.4%), a Histidine (7.1%), a Proline (0.8%) or a Glutamic Acid (0.4%) [19] (*i.e.* Q61R > Q61K > Q61L > Q61H > Q61P > Q61E). NRas proteins hence appear as attractive targets for the pharmaceutical industry. Because the role of GAP in reducing the amount of active GTP bound Ras within the cell is fundamental, its arginine binding loop is also included in the simulations that have been carried out. The employed computational methods are presented in the next chapter (chapter 2). The corresponding results, in the forthcoming one (chapter 3).

Chapter 2

Methods

Nowadays, understanding the complexity of living organisms through chemical reactions and physical processes is beyond conducting chemical experiments. Computational chemistry makes it possible to study, at different levels, biomolecules and the chemical processes they are involved in, within a variety of environments. In this chapter, the foundations of molecular modeling methods [16, 29, 40, 41], along with the scope of their application, employed to study NRas protein, are described.

2.1 Molecular Modeling Methods

2.1.1 Relating a molecule's shape to its stability: the PES

The definition of chemistry as being the scientific discipline that studies the structure and properties of matter together with the set of reactions that transform it, gave rise to the emergence of numerous representations that aim at describing it at the molecular level. Indeed, geometry is a key feature and molecular modeling methods aim at identifying structures of interest by relating geometry to relative stability. This is achieved by evaluating, for each of the compound's conformation, its internal energy, thus leading to the construction of a potential energy surface (see figure 2.1).

The potential energy surface (PES) can be regarded as a function that assigns to a relative arrangement of atoms the energy resulting from this particular arrangement:

$$PES : (x_1, y_1, z_1; ...; x_n, y_n, z_n) \mapsto E(x_1, y_1, z_1; ...; x_n, y_n, z_n) \tag{2.1}$$

This definition 2.1 holds for a molecule containing n atoms, each with coordinates (x_i, y_i, z_i)

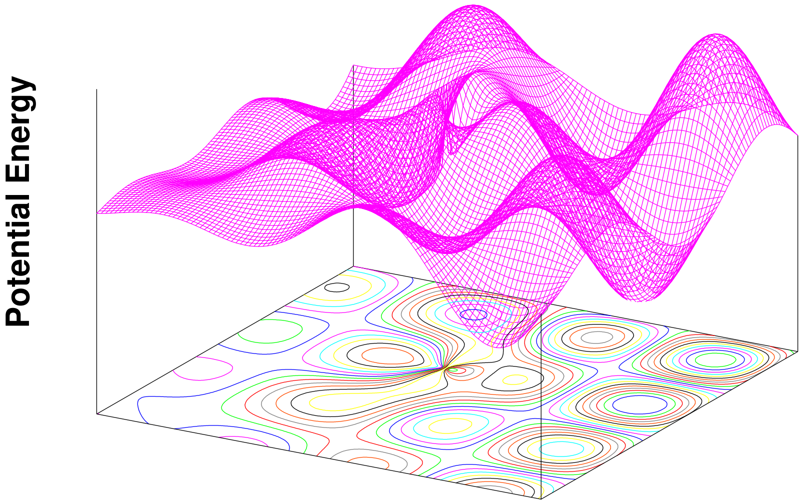


FIGURE 2.1: Potential Energy Surface

The construction of the PES is made within the Born[1]-Oppenheimer[2] approximation that enables the separation of nuclear and electronic motion. Indeed, because of the nucleon/electron mass ratio, nuclear kinetic energy is insignificant compared to that of electrons. Therefore, the atoms' nucleus positions can be considered as remaining *fixed*. The interactions between these nuclei establish a potential energy $V_{Nuclear}$. The addition of the zero point energy V_{0K} (described below) and the electrons total energy $V_{electronic}$ to it, lead to the molecule overall internal energy:

$$E = V_{Nuclear} + V_{electronic} + V_{0K} \tag{2.2}$$

Molecular modeling methods intend to explore the PES and identify its stationary points. The latter are characterized by the first derivative of the energy with respect to atomic coordinates that vanishes in all directions *i.e* all first partial derivatives equal zero:

$$\frac{\partial E}{\partial q_i} = 0 \, \forall \, i \;\text{ with }\; i = 1, ..., 3n \tag{2.3}$$

The most stable conformations of a molecule, that have a finite lifetime, and the transition states, that only exist for an instant during a reaction, are in this manner determined. The former correspond to local or global energy minima in the PES and the latter to saddle points connecting two minima. They are distinguished by the values of

---

[1] Max Born, German Physicist, 1913-1970

[2] Robert Oppenheimer, American Physicist, 1904-1967

the second derivative of the energy with respect to the atoms coordinates. This derivative is positive in all directions for a minimum *i.e* all second partial derivatives are over zero, denoting that a minimum is a minimum in all directions:

$$\frac{\partial^2 E}{\partial q_i^2} > 0 \, \forall \, i \tag{2.4}$$

While, for a transition state, following the direction defined by the lowest energy pathway between the two minima it connects (called the intrinsic reaction coordinate IRC), the second partial derivative in this direction is negative, denoting that a transition state is a maximum following the direction of the IRC:

$$\frac{\partial^2 E}{\partial q_i^2} < 0 \quad \text{for} \; i = IRC \tag{2.5}$$

This distinction between minima and transition states, *i.e.* the sign of the second derivative of the energy with respect to atomic coordinates, leads to the characterization of stationary points found on the PES by calculating the *normal modes* associated to the corresponding geometries. The normal modes are the simplest vibrations of a molecule. They are characterized by the *in-phase* movement of the constituent atoms: they reach their maximum displacement and their equilibrium positions all at the same time. The set of normal modes corresponds to the experimental infrared (IR) spectrum. The combination of this elementary vibrations leads to the more complex vibrations a molecule experiences.

A non linear molecule made of n atoms has $3n - 6$ normal modes while a linear molecule has $3n - 5$. The normal mode frequency involving atoms A and B of masses m_A and m_B, respectively, is given by:

$$\tilde{\nu} = \frac{1}{2\pi c}\left(\frac{k}{\mu}\right)^{1/2} \quad \text{with} \;\; \mu = \frac{m_A m_B}{m_A + m_B} \;\; \text{reduced mass} \tag{2.6}$$

where $\tilde{\nu}$ is the frequency in cm^{-1}, c the celerity of light and k the force constant of the vibration.

The force constants k_i in equation 2.6 ($i = 7, ..., 3n$ as the first 6 correspond to translational and rotational motion of the overall molecule) are a measure of the stiffness of the molecule towards the corresponding vibration. They are obtained from the diagonalization of the Hessian[3] matrix $\mathbf{H}$ that contains the second derivatives of the energy

---

[3] Ludwig Otto Hesse, German Mathematician, 1811-1874

with respect to atomic coordinates.

$$\underset{\mathbf{H}}{\begin{pmatrix} \frac{\delta^2 E}{\delta q_1 q_1} & \cdots & \frac{\delta^2 E}{\delta q_1 q_{3n}} \\ \vdots & \ddots & \\ \frac{\delta^2 E}{\delta q_{3n} q_1} & \cdots & \frac{\delta^2 E}{\delta q_{3n} q_{3n}} \end{pmatrix}} = \underset{\mathbf{P}}{\begin{pmatrix} q_{11} & \cdots & q_{1(3n)} \\ \vdots & \ddots & \\ q_{(3n)1} & \cdots & q_{(3n)(3n)} \end{pmatrix}} \underset{\mathbf{k}}{\begin{pmatrix} k_1 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & k_{3n} \end{pmatrix}} \mathbf{P}^{-1} \tag{2.7}$$

Upon diagonalization 2.7, the eigenvalues k_i of the Hessian matrix constitute the magnitude of the displacements associated to the corresponding normal mode i. The eigenvectors (column i of **P**) depict the directions of motion of atoms involved in normal mode i.

Minima on the PES correspond to all the normal mode force constants k_i being positive (because of 2.4) which leads to real IR frequencies only. Stationary points corresponding to a transition state, have one negative force constant (because of 2.5) leading to one imaginary IR frequency in its spectra. Therefore, the calculation of vibrational frequencies, enables the distinction between stationary points on the PES. Beyond, it is used to obtain the zero point energy V_{0K} of the molecule. This energy corresponds to the vibrational energy of the molecule at $T = 0K$. Indeed, even at this temperature the molecule vibrates. Otherwise, its momentum and position would be exactly determined thereby transgressing the Heisenberg[4] uncertainty principle. The zero point energy needs to be accounted for in order to accurately evaluate the total internal energy E (see equation 2.2).

Molecular modeling methods described in the following sections, differ by the representation of atoms therefore, by the equations that are used to determine the internal energy E associated to a specific geometry of the system under study. The stationary points of the PES are hence characterized differently according to the chosen approach. These methods are introduced following a higher level of accuracy which is associated to an increased computer cost.

2.1.2 Molecular Mechanics

Molecular mechanics (MM) is a method from computational chemistry that describes atoms as simple spheres of van der Waals[5] radii, making no distinction between the nuclei and electrons forming it. Within a molecule, the spheres interact with their bonded neighbors through springs that hold them together as depicted in figure 2.2.

---

[4] Werner Heisenberg, German Physicist, 1901-1976

[5] Johannes Diderik van der Waals, Dutch Physicist, 1837-1923

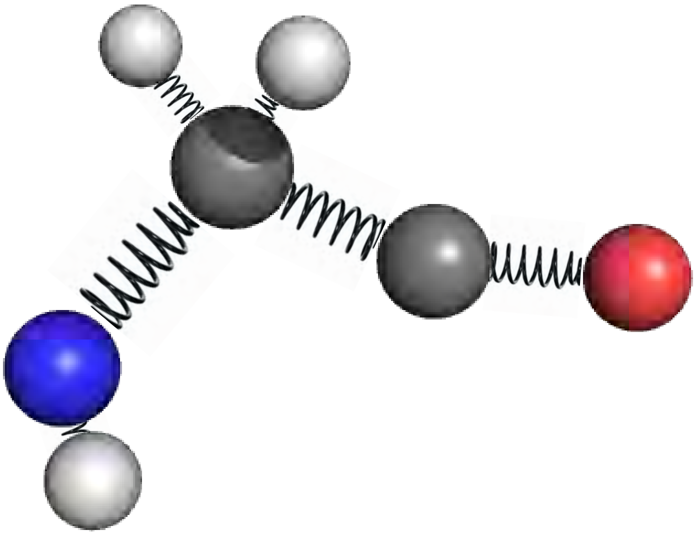


FIGURE 2.2: Mechanical model of molecules: van der Waals spheres & springs

This mechanical description leads to a *mechanical* potential energy of bonded atoms which is the sum of energy terms accounting for:

- the stretching and compressing of the spring between two consecutive spheres:

$$V_{stretch_{ij}} = k_{stretch_{ij}}(l_{ij} - l_{eq_{ij}})^2$$

- the angle bending between two springs connecting three consecutive spheres:

$$V_{bend_{ij}} = k_{bend_{ijk}}(\theta_{ijk} - \theta_{eq_{ijk}})^2$$

- the torsional angle between the edge springs connecting four consecutive spheres:

$$V_{torsion_{ijkl}} = k_{torsion_{ijkl}}(1 + cos(l\phi_{ijk}))$$

The accuracy of each of these *mechanical* terms can be improved by the addition of higher order anharmonic terms. Indeed, the above corresponds to the harmonic approximation where only up to the second-order terms of a Taylor expansion around the minimum of a potential energy function $V(r)$ are accounted for.

Pauli repulsive and van der Waals attractive interactions between two neutral non-bonded atoms, *i* and *j*, are accounted for using a Lennard-Jones potential of the form:

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.8}$$

where ϵ_{ij} is the Lennard-Jones[6] well depth, σ_{ij} is the interatomic distance at which repulsive and attractive forces exactly compensate and r_{ij} is the distance between the two spheres.

For charged atoms, the electrostatic interactions between two spheres, *i* and *j*, holding partial charges q_i and q_j, respectively, is evaluated through:

$$V_{elec}(r_{ij}) = \frac{q_i q_j}{\varepsilon_{ij} r_{ij}} \tag{2.9}$$

The total potential energy of a molecule is therefore:

$$V = \sum_{bonds} V_{stretch} + \sum_{angles} V_{bend} + \sum_{dihedrals} V_{torsion} + \sum_{pairs} \left(V_{LJ}(r_{ij}) + V_{elec}(r_{ij}) \right) \tag{2.10}$$

Given a spatial configuration of van der Waals atoms forming a molecule, its total energy is readily calculated if the parameters corresponding to the equilibrium bond lengths, angles and dihedrals, along with their corresponding stretch, bend and torsion force constants, as well as the Lennard-Jones well depth and zero force interatomic distances, are known. These parameters are usually determined from experimental data and are assembled in what is commonly named a *force field*.

The strength of molecular mechanics, besides laying on the simple yet accurate [7] van der Waals sphere model, is also due to the transferability of force fields. Indeed, the latter are build for specific molecules (organic/inorganic compounds, proteins, lipids, nucleotides...) in a manner that they are fitted to reproduce experimental measurements of the ones forming the *training set*. A given force field can then be used to study a plethora of molecules of the same nature as the ones forming the training set.

Because of the simplicity of the equations, molecular mechanics is well-suited for studying the PES associated to macromolecules within reasonable computer time.

However, its weakness also comes from both the simplicity of the model it is based on and the force fields. Indeed, because electrons are neglected, molecular mechanics cannot be used to describe electronic properties. Moreover, a given force field, despite being well-suited for interpolation, is not designed for extrapolation *i.e.* exotic molecules will hardly be accurately described by an *inappropriate* force field.

---

[6] John Lennard-Jones, British Mathematician, 1894-1954

[7] when used in the boundaries of its scope

2.1.3 Semi-empirical methods

Semi-empirical methods, as inferred by their name, partly lay on parameters fitted to experimental data. The family of methods described in this section concern the ones that were developed in an attempt to provide fast yet accurate alternatives to *ab initio* approaches. Therefore, as within the latter, the representation of atoms is not approximated: both nuclei and electrons are considered when calculating the energy of a molecule. Moreover, the atom's constituent entities are treated in the framework of quantum mechanics leading to consider their dual wave-particle behavior.

As in *ab initio* methods, the total energy E of a molecule is obtained by solving the time-independent non-relativistic Schrödinger equation of the N atomic nuclei and n electrons forming it:

$$\hat{H}^{tot}\Psi^{tot}(1,...,N,1,...,n) = E\Psi^{tot}(1,..,N,1,...,n) \tag{2.11}$$

This is an eigenvalue equation where the set of energies E and wave functions $\Psi^{tot}(1,..,N,1,...,n)$ solving it are the eigenvalues and the eigenfunctions, respectively, of the total energy operator $\hat{H}^{tot}$. This operator is defined by:

$$\hat{H}^{tot} = -\sum_{\mu}\frac{\hbar^2}{2M_\mu}\nabla^2_\mu + \sum_{\mu<\eta}\frac{Z_\mu Z_\eta e^2}{4\pi\varepsilon_0 r_{\mu\eta}} - \sum_i \frac{\hbar^2}{2m_i}\nabla^2_i + \sum_{i<j}\frac{e^2}{4\pi\varepsilon_0 r_{ij}} - \sum_\mu\sum_i \frac{Z_\mu e^2}{4\pi\varepsilon_0 r_{\mu i}} \tag{2.12}$$

where:

$-\frac{\hbar^2}{2M_\mu}\nabla^2_\mu$ is the kinetic energy of the μ^{th} atomic nucleus of mass M_μ

$\frac{Z_\mu Z_\eta e^2}{4\pi\varepsilon_0 r_{\mu\eta}}$ is the nuclear repulsion between the μ^{th} and η^{th} nuclei

$-\frac{\hbar^2}{2m_i}\nabla^2_i$ is the kinetic energy of the i^{th} electron of mass m_i

$\frac{e^2}{4\pi\varepsilon_0 r_{ij}}$ is the electronic repulsion between the i^{th} and j^{th} electrons

$\frac{Z_\mu e^2}{4\pi\varepsilon_0 r_{\mu i}}$ is the electrostatic interaction between the μ^{th} nucleus and the i^{th} electron

As in molecular mechanics, the Born-Oppenheimer approximation is assumed leading to the atoms' nuclei kinetic energy being neglected and their positions considered fixed. Because of the latter, $\hat{H}^{tot}$ can be separated in nuclear and electronic energy operators $\hat{V}_{NN}$ and $\hat{H}^{el}$, respectively.

Within this approximation, the nuclear energy operator consists in the nucleus-nucleus repulsion term only:

$$\hat{V}_{NN} = \sum_{\mu<\eta} \frac{Z_\mu Z_\eta e^2}{4\pi\varepsilon_0 r_{\mu\eta}} \tag{2.13}$$

For a geometry of interest, the potential resulting from inter-nuclei interactions is hence readily calculated from the atoms considered positions.

In this scenario, electrons move in a stationary field generated by the atomic nuclei. Their contribution to the total molecular energy is obtained by solving the time independent non-relativistic Schrödinger[8] equation for the electron energy operator:

$$\hat{H}^{el} = \sum_i -\frac{\hbar^2}{2m_i}\nabla_i^2 + \sum_{i<j}\frac{e^2}{4\pi\varepsilon_0 r_{ij}} - \sum_\mu\sum_i \frac{Z_\mu e^2}{4\pi\varepsilon_0 r_{\mu i}} \tag{2.14}$$

Equation 2.11 for $\hat{H}^{el}$ leads to the electronic energy being:

$$\varepsilon = \frac{\int \Psi_{el}^*(1,...,n)\hat{H}^{el}\Psi_{el}(1,...,n)d\tau}{\int \Psi_{el}(1,...,n)^*\Psi_{el}(1,...,n)d\tau} \quad \text{with } d\tau = dx\,dy\,dz\,d\xi \tag{2.15}$$

where $\Psi_{el}(1,...,n)$ is a purely electronic wavefunction that describes the n electrons motion in the field generated by the molecule's nuclei. Ψ_{el}^2 is regarded as the probability density function that describes molecular electron density. The integration is carried out with respect to the three spatial coordinates x, y, z and the spin variable ξ of each electron, leading thereby to an integration over $4n$ variables.

The difficulty in determining the electronic energy ε lies in the electron-electron repulsive interaction term from 2.14. Because of it, the Schrödinger equation for $\hat{H}^{el}$ cannot be separated in n one-electron equations that could be independently solved, exactly, and then assembled to yield the total electronic energy ε of the molecule. Indeed, the Schrödinger equation can be solved analytically for the hydrogen atom only and leads to one-electron wave functions named orbitals (or atomic orbitals AOs).

To overcome this difficulty, a subsequent approximation consists in considering the total poly-electron wave function $\Psi_{el}(1,...,n)$ of the molecule as a Slater[9] determinant of one-electron wave functions:

$$\Psi_{el}(1,...,n) = \frac{1}{\sqrt{n!}}\begin{vmatrix} \psi_1(1)\alpha(1) & \psi_1(1)\beta(1) & \dots & \psi_{n/2}(1)\alpha(1) & \psi_{n/2}(1)\beta(1) \\ \vdots & & & & \\ \psi_1(n)\alpha(n) & \psi_1(n)\beta(n) & \dots & \psi_{n/2}(n)\alpha(n) & \psi_{n/2}(n)\beta(n) \end{vmatrix} \tag{2.16}$$

---

[8] Erwin Schrödinger, Austrian Physicist, 1887-1961

[9] John Clarke Slater, American Physicist, 1900-1976

$\psi_i(j)$ is the i^{th} molecular spatial wave function occupied by the j^{th} electron. $\alpha(j)$ and $\beta(j)$ are spin functions that determine the spin, $+\frac{1}{2}$ and $-\frac{1}{2}$ respectively, of the j^{th} electron occupying the i^{th} spatial wave function. These spin functions are orthogonal *i.e.* $\alpha(j)\beta(j) = \delta_{\alpha\beta}$. Because the Pauli[10] exclusion principle forbids two electrons bearing identical quantum numbers from occupying the same quantum state, $\alpha(j)$ and $\beta(j)$ spin functions are associated to each spatial wave function ψ_i to ensure that each of these is occupied by maximum two electrons of opposed spin. $\psi_i(j)\alpha(j)$ and $\psi_i(j)\beta(j)$ are thus two one-electron molecular spin-orbitals holding two electrons of opposite spin in the same i^{th} spatial wave function. Note that the Slater determinant allows all electrons to roam among all molecular one-electron spin-orbitals. For closed shell systems which consist in an even number n of electrons, each paired with an electron of opposite spin, the Slater determinant is constructed from $\frac{n}{2}$ spatial wave functions.

It is important to keep in mind that orbitals constitute the solutions of the Schrödinger equation for the hydrogen atom and thus, assigning one-electron orbitals to atoms or molecules (atomic orbitals AOs or molecular orbitals MOs, respectively) containing more than one electron constitutes an approximation.

The mathematical properties of a determinant make it possible to apply arithmetic operations to the one-electron molecular spin-orbitals without changing the total poly-electron molecular wave function $\Psi_{el}(1, ..., n)$ constructed from these as in 2.16. Hence, the constituent one-electron molecular orbitals ψ_i can be transformed leaving the properties of the molecular poly-electron orbital $\Psi_{el}(1, ..., n)$ unchanged. In this manner, if the one-electron MOs ψ_i are manipulated making them orthonormal, the total poly-electron energy of the molecule is determined from:

$$\varepsilon = \int \Psi^*_{el}(1, ..., n)\hat{H}^{el}\Psi_{el}(1, ..., n)d\tau \text{ as } \int \psi_i(p)\psi_j(p)d\tau = \delta_{ij}\forall p \in [1, n] \tag{2.17}$$

Separating the electronic hamiltonian in two contributions

$$\hat{H}^{el} = \sum_{electrons} \hat{H}_{core} + \hat{H} \text{ with } \hat{H}_{core} = -\frac{\hbar^2}{2m_i}\nabla_i^2 - \sum_{\mu} \frac{Z_\mu e^2}{4\pi\varepsilon_0 r_{\mu i}}$$
$$\hat{H} = \sum_{i<j} \frac{e^2}{4\pi\varepsilon_0 r_{ij}}$$

and adopting the Dirac[11] notation for equation 2.17 *i.e.* $\varepsilon = \langle\Psi_{el}|\hat{H}^{el}|\Psi_{el}\rangle$, its integration while replacing the molecular poly-electron orbital $\Psi_{el}(1, ..., n)$ by its expression

---

[10] Wolfgang Pauli, Austrian Physicist, 1900-1958

[11] Paul Dirac, English Physicist, 1902-1984

as a Slater determinant 2.16, yields the following terms:

$$H_{ii} = \langle \psi_i(p) | \hat{H}_{core} | \psi_i(p) \rangle \tag{2.18}$$

the expectation value of the one-electron core hamiltonian $\hat{H}_{core}$ that represents the energy of one electron, here electron p, subject to the potential generated by all the fixed nuclei μ in the spatial orbital ψ_i, all other electrons being removed. This energy is the same for each of the n electrons as they are identical.

$$J_{ij} = \langle \psi_i(p)\psi_i(p) | \hat{H} | \psi_j(q)\psi_j(q) \rangle \tag{2.19}$$

the expectation value of the two electrons hamiltonian $\hat{H}$ when electron p is in the spatial orbital ψ_i and electron q in ψ_j. J is named the Coulomb[12] integral and represents the electrostatic repulsion between the charge cloud of spatial orbitals ψ_i and ψ_j. This term also arises in the form $J_{ij} = \langle \psi_i(p)\psi_j(q) | \hat{H} | \psi_i(p)\psi_j(q) \rangle$ and hence has to be counted twice.

$$K_{ij} = \langle \psi_i(p)\psi_j(q) | \hat{H} | \psi_i(q)\psi_j(p) \rangle \tag{2.20}$$

the expectation value of the two electrons hamiltonian $\hat{H}$ when electrons p and q can be assigned to the same spatial orbital. K is named the Exchange integral and is considered to be a correction to the electrostatic repulsion. Indeed, following the approximation of using one electron spin-orbitals results in the overestimation of the coulomb repulsion as each electron is then repelled by all the others when, in fact, electrons of same spin *intrinsically* avoid each other, reducing in this manner the coulomb repulsion that would arise between them if they had no spin.

The poly-electron energy, defined in equation 2.17 can therefore be written:

$$\varepsilon = 2\sum_{i}^{n/2} H_{ii} + \sum_{i}^{n/2}\sum_{j}^{n/2}(2J_{ij} - K_{ij}) \tag{2.21}$$

As each spatial orbital can hold two electrons of opposite spin, the energy of two electrons has to be accounted for each spatial orbital when evaluating the core hamiltonian, hence the "2" premultiplying term for the sum over H_{ii}. Because we consider a closed shell system consisting of a molecule with n electrons, $n/2$ spatial orbitals are used to construct the Slater determinant of Ψ_{el}. The sums are thus considered from spatial orbital ψ_1 to spatial orbital $\psi_{n/2}$.

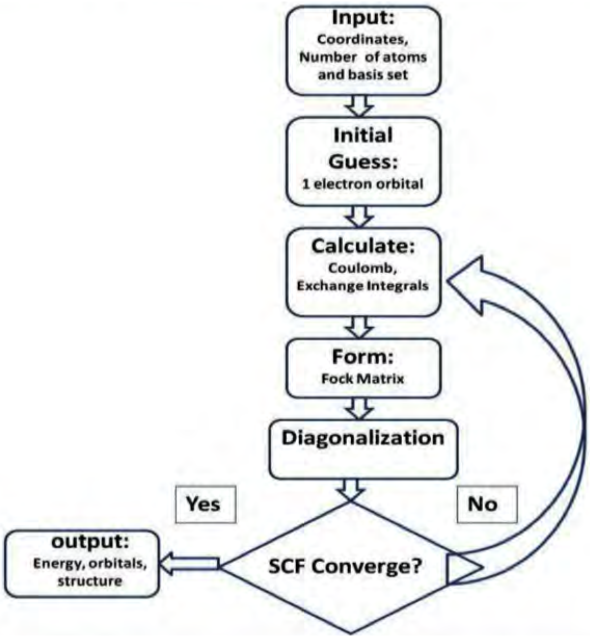
---

[12] Charles-Augustin de Coulomb, French military engineer and Physicist, 1736-1806

In order to calculate the molecular poly-electronic energy according to equation 2.21, the one-electron MOs ψ_i need to be known. Indeed, even when discarding the fact that ε is the expectation value of the electron hamiltonian $\hat{H}^{el}$, the Coulomb J_{ij} and Exchange K_{ij} integrals depend *directly* on them as the operators associated to these quantities are defined as:

$$\hat{J}_j(p) = \int \psi_j^*(q)\hat{H}\psi_j(q)\, d\tau_q \tag{2.22}$$

$$\hat{K}_j(p)\psi_i(p) = \left[\int \psi_j^*(q)\hat{H}\psi_i(q)\, d\tau_q\right]\psi_j(p) \tag{2.23}$$



Copyright ©: P.C. Deshmukh, *DST-SERC-School publication (2011)*

FIGURE 2.3: Self Consistent Field procedure

To determine $\hat{J}_j$ and $\hat{K}_j$ and overcome the fact that the obtention of the eigenvalues ε of $\hat{H}^{el}$ requires the knowledge of its eigenfunctions Ψ_{el} when, usually, the latter are determined once the former have been calculated, the Hartree method is employed. This method consists in *i)* using an *initial guess* for the one-electron MOs ψ_j, *ii)* calculating the associated energies using the expression in 2.21, *iii)* determining the associated eigenvectors that arise and *iv)* comparing them to the initial guess. If the last step, the comparison, leads to similar MOs *i.e.* there is no significant change according to certain criteria between the initial guess MOs and the resulting MOs, the electrons field is consistent with itself, *i.e. self consistent*, and thus the obtained eigenvalues and eigenfunctions of $\hat{H}^{el}$ are considered to be the *real* ones. If not, the resulting MOs are taken to repeat the cycle until self consistency is achieved. Hence, in semi-empirical methods, as in wave function-based *ab initio* approaches, a Self Consistent Field (SCF) procedure is adopted to determine the electronic energy of a molecule.

A final approximation consists hence in building the initial guess of the one electron MOs ψ_j through a Linear Combination of Atomic Orbitals (LCAO approach):

$$\psi_j(x,y,z) = \sum_r c_{rj}\phi_r(x,y,z) \tag{2.24}$$

These AOs constitute the basis functions. Each basis function $\phi_r(x,y,z)$ contributes to each one-electron MO $\psi_j(x,y,z)$ according to the c_{rj} weighting coefficients. To ensure that the one-electron MOs remain orthogonal, the AOs must satisfy:

$$\sum_{rs} c^*_{ri}c_{sj}S_{rs} = \delta_{ij} \text{ with } S_{rs} = \int \phi_r(p)\phi_s(p)\,d\tau_p\, \forall\, p \in [1,n] \tag{2.25}$$

The set of basis functions constitutes the *basis set*. Because the linear combination of n AOs $\phi_r(x,y,z)$ leads to the construction of n MOs $\psi_j(x,y,z)$, for a closed shell system of n electrons, a basis set made of at least $\frac{n}{2}$ AOs is required. Nevertheless, the basis set can be expanded as desired to reach higher levels of accuracy. However, such calculations are accompanied by an increase in the need of computer resources. Whatever the size of the basis set chosen, the Slater determinant of the molecular poly-electronic wave function Ψ_{el}, is constructed from the occupied MOs ψ_j.

Replacing the one-electron MOs ψ_j by their expansion in LCAO ϕ_r leads the terms in expression 2.21 of the poly-electron energy to become :

$$H_{ii} = \sum_{rs} c^*_{ri}c_{si}H_{rs} \text{ with } H_{rs} = \int \phi^*_r(p)\hat{H}^{core}\phi_s(p)\,dv_p \tag{2.26}$$

$$J_{ij} = \sum_{rstu} c^*_{ri}c_{si}c^*_{tj}c_{uj} \int \phi^*_r(p)\phi_s(p)\hat{H}\phi^*_t(q)\phi_u(q)\,dv_p\,dv_q \tag{2.27}$$

$$K_{ij} = \sum_{rstu} c^*_{ri}c_{ui}c^*_{tj}c_{sj} \int \phi^*_r(p)\phi_u(q)\hat{H}\phi^*_t(p)\phi_s(q)\,dv_p\,dv_q \tag{2.28}$$

Adopting the more compact notation $\int \phi^*_r(p)\phi^*_s(p)\hat{H}\phi^*_t(q)\phi_u(q)\,dv_p\,dv_q = (rs|tu)$ that denotes that electron p occupies AOs r and s and electron q AOs t and u, the poly-electron energy ε in the LCAO approach is expressed as:

$$\varepsilon = 2\sum_{i}^{occupied}\sum_{rs} c^*_{ri}c_{si}H_{rs} + \sum_{ij}^{occupied}\sum_{rstu} c^*_{ri}c_{si}c^*_{tj}c_{uj}[2(rs|tu)-(ru|ts)] \tag{2.29}$$

As this quantity is the expectation value of the electron hamiltonian operator $\hat{H}^{el}$ for the poly-electronic wave function $\Psi_{el}(1,...,n)$ (see equation 2.15), the poly-electron energy evaluated through equations 2.17 and following is only an approximation of the *true* value. Indeed, in these equations (2.17 and 2.21), $\Psi_{el}(1,...,n)$ is approximated by a Slater determinant of one-electron MOs $\psi_j(x,y,z)$ which are initially constructed by a LCAO $\phi_r(x,y,z)$, in the expression 2.29 we use. The variational principle ensures that the value of the energy ε obtained, using the approximated Ψ_{el}, is an upper-bound to the *real* value. Hence, the lower the value, the better the energy (in the sense of closer to the *true* value of ε) and the more accurate the wave functions describing electron density obtained through the SCF procedure.

The electronic ground state of the molecule considered geometry corresponds to the lowest energy MOs being filled. The energy associated to this state is the minimum electronic energy. The ground state eigenvalues and eigenfunctions of $\hat{H}^{el}$ are therefore obtained by minimizing ε, defined in equation 2.29, with respect to the one-electron MOs ψ_j which initial *guess* is defined in equation 2.24. Indeed, if the ground state energy is reached, a small variation of the associated electronic ground wave functions $\delta\psi_j = \sum\limits_r \delta c_{rj}\phi_r$ should not affect significantly its value since it is a minimum.

The minimization of ε with respect to the MOs ψ_j is carried out using the method of the undetermined Lagrangian[13] multipliers ε_{ij} with the constraint that the MOs ψ_j remain orthonormal. Indeed, this property was enforced in the Slater determinant (see equation 2.17 and 2.25) that subsequently led to the expression of ε in equations 2.21 and 2.29. Thus the set of MOs ψ_j that minimizes this energy (ε) is the one leaving the quantity $L = \varepsilon - 2\varepsilon_{ij}c^*_{ri}c_{sj}S_{rs}$ constant. Therefore:

$$\begin{aligned}\delta L = &\\ & 2\sum_i^{occ}\sum_{rs}\delta c^*_{ri}c_{si}H_{rs} + \sum_{ij}^{occ}\sum_{rstu}(\delta c^*_{ri}c_{si}c^*_{tj}c_{uj} + c^*_{ri}\delta c_{si}c^*_{tj}c_{uj})[2(rs|tu) - (ru|ts)] \\ & - 2\sum_{ij}\sum_{rs}\varepsilon_{ij}\delta c^*_{ri}c_{sj}S_{rs} + \ \mathit{complex}\ \ \mathit{conjugate}\ = 0\end{aligned}$$

that leads to:

$$\begin{aligned} & \sum_s\left(c_{si}H_{rs} + \sum_j^{occ}\sum_{stu}c_{si}c^*_{tj}c_{uj}[2(rs|tu) - (ru|ts)]\right) = \sum_j\varepsilon_{ij}\sum_s c_{sj}S_{rs} \\ \iff & \sum_i\sum_s\left(H_{rs} + \sum_{tu}c^*_{tj}c_{uj}[2(rs|tu) - (ru|ts)]\right)c_{si} = \sum_i\varepsilon_{ij}\sum_s S_{rs}c_{si}\end{aligned}$$

---

[13] Giuseppe Luigi Lagrangia, Italian Mathematician and Astronomer, 1736-1813

The Lagrangian multipliers ε_{ij} can be transformed in order to obtain $\varepsilon_{ij} = \begin{cases} \varepsilon_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$.

In this manner, for each molecular orbital i, the coefficients c_{si} that minimise the energy ε satisfy:

$$\sum_s \left(H_{rs} + \sum_{tu} c^*_{tj} c_{uj} [2(rs|tu) - (ru|ts)] \right) c_{si} = \varepsilon_i \sum_s S_{rs} c_{si} \tag{2.30}$$

where $F_{rs} = H_{rs} + \sum\limits_{tu} c^*_{tj} c_{uj} [2(rs|tu) - (ru|ts)]$ is the expectation value of an effective one-electron operator called Fock[14] operator and defined by:

$$\hat{F} = \hat{H}^{core} + \sum_j (2\hat{J}_j - \hat{K}_j) \tag{2.31}$$

H_{rs} is a one-electron energy term (the core energy of an electron occupying MO ψ_i) and $G_{rs} = \sum\limits_{tu} c^*_{tj} c_{uj} [2(rs|tu) - (ru|ts)]$ is a two-electron energy term (the coulomb and exchange energies of the i^{th} electron in interaction with all others). The set of equations 2.30 is the Roothaan[15]-Hall[16] version of the Hartree[17]-Fock equations.

Replacing F_{rs} into this set of equations 2.30 leads to write it as a matrix equation:

$$\underbrace{\begin{pmatrix} F_{11} & \cdots & F_{1s} & \cdots & F_{1n} \\ \vdots & \ddots & & & \\ F_{r1} & & F_{rs} & & F_{rn} \\ \vdots & & & \ddots & \\ F_{n1} & & F_{ns} & & F_{nn} \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} c_{11} & \cdots & c_{1i} & \cdots & c_{1n} \\ \vdots & \ddots & & & \\ c_{s1} & & c_{si} & & c_{sn} \\ \vdots & & & \ddots & \\ c_{n1} & & c_{ni} & & c_{nn} \end{pmatrix}}_{\mathbf{C}} = \underbrace{\begin{pmatrix} \varepsilon_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & & & \\ 0 & \cdots & \varepsilon_i & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & \cdots & 0 & \cdots & \varepsilon_n \end{pmatrix}}_{\mathbf{E}} \underbrace{\begin{pmatrix} S_{11} & \cdots & S_{1s} & \cdots & S_{1n} \\ \vdots & \ddots & & & \\ S_{r1} & & S_{rs} & & S_{rn} \\ \vdots & & & \ddots & \\ S_{n1} & & S_{ns} & & S_{nn} \end{pmatrix}}_{\mathbf{S}} \underbrace{\begin{pmatrix} c_{11} & \cdots & c_{1i} & \cdots & c_{1n} \\ \vdots & \ddots & & & \\ c_{s1} & & c_{si} & & c_{sn} \\ \vdots & & & \ddots & \\ c_{n1} & & c_{ni} & & c_{nn} \end{pmatrix}}_{\mathbf{C}}$$

$$\mathbf{F\,C} \;= \mathbf{E\,S\,C} \tag{2.32}$$

F is the energy matrix, commonly named Fock matrix. **C** is the matrix containing the weighting factors c_{si} that define the contribution of basis function ϕ_s to MO ψ_i. **S** is the overlap matrix which elements S_{rs} describe how well pairs of basis functions ϕ_r and ϕ_s overlap. It appears that the Lagrangian multipliers ε_i, contained in the diagonal matrix **E**, are the energy levels of the Fock matrix.

For computational purposes, it is suitable that the overlap matrix **S** be a diagonal matrix *i.e.* $S_{rs} = \delta_{rs}$. This corresponds to an orthonormal basis set which is not the case for usual basis sets built from usual AOs (the ones obtained for the hydrogen atom

---

[14] Vladimir Aleksandrovich Fock, Soviet Physicist, 1898-1974

[15] Clemens Roothaan, Dutch Physicist born 1918

[16] George Garfield Hall, Northern Irish Mathematician, 1925-2018

[17] Douglas Hartree, English Mathematician and Physicist, 1897-1958

when solving analytically the Schrödinger equation). Nonetheless, this can be easily achieved through a Löwdin[18] orthogonalisation that is systematically employed in computational chemistry.

When applying this procedure to equation 2.32, it becomes:

$$\mathbf{F}^{\tau}\mathbf{C}^{\tau} = \mathbf{E}\,\mathbf{C}^{\tau} \tag{2.33}$$

and so the coefficients c_{si}^{τ} are determined from the set of linear equations:

$$\sum_{s}(F_{rs}^{\tau} - \varepsilon_i \delta_{rs})c_{si}^{\tau} = 0 \tag{2.34}$$

Finally, the coefficients c_{si} of the one-electron MOs ψ_i that minimize the poly-electronic energy ε, are obtained by applying to the c_{si}^{τ} coefficients the inverse procedure of the one that enabled the orthogonalization of the overlap matrix **S**.

Semi-empirical methods are much less demanding in computer resources than *ab initio* approaches because of several quantities that, instead of being numerically calculated, are fitted to reproduce experimental values such as heats of formation, ionization energies, dipole moments and geometries. The main differences between these quantum-based approaches are:

1. Semi-empirical methods consider valence electrons only. Core electrons are described as part of the atomic nucleus, which produce a different potential than that defined in equation 2.13. As a direct consequence, the number of MOs ψ_j needed for constructing the Slater determinant is half the number of valence electrons instead of half the number of total electrons.

2. The basis set in semi-empirical methods is limited to the number of valence AOs while in *ab initio* calculations, it can be as large as desired.

3. The overlap matrix **S** in semi-empirical calculations is considered as being unity from the beginning, hence the orthogonalization procedure of this matrix do not need to be carried out.

4. In semi-empirical methods, the different integrals that emerge in the expression of the Fock matrix elements (see equations 2.30) are not systematically calculated while they are all evaluated numerically in *ab initio* methods. Indeed, semi-empirical methods differ between them in the manner these integrals are handled. For many, the one-electron energy terms H_{rs} in 2.30 are obtained through

---

[18] Per-Olov Löwdin, Swedish Physicist, 1916-2000

fitting to reproduce experimental data. The two-electrons energy terms from G_{rs} in 2.30 are calculated when centered in one AO, approximated when centered in two AOs and completely neglected when centred in three or four AOs.

The gain in speed when employing semi-empirical methods is accompanied by a small loss in accuracy when comparing results to *ab initio* calculations. Nevertheless, the empirical parameters make the use of semi-empirical methods unreliable for molecules that are not similar to the ones used in the training set employed for obtaining these parameters.

2.1.4 Density Functional Theory

An alternative to wave function-based quantum approaches, as *ab initio* and semi-empirical methods, is Density Functional Theory (DFT). For modeling multi-electron systems at the atomic level, DFT methods are based on the assumption that the ground state properties of the studied system depend on its spatial, ground state, electronic density function $\rho_0(x, y, z)$ only. In this manner, for a system holding n electrons, the mathematical problem of $4n$ variables in wave function-based methods is reduced to 3 variables in DFT.

Within this approach, electrons are treated as forming a negatively charged cloud, the shape of it being described by the spatial electron density function $\rho(x, y, z)$. In this *continuum* description, as opposite to a *discrete one*, the sum of point charges held by electrons in the terms of equation 2.14 become:

$$\sum_{i=1}^{n} e_i \mapsto \int \rho(\vec{r}) d\vec{r} \tag{2.35}$$

The theoretical foundations of DFT approaches lie on the first and second Hohenberg[19]-Kohn[20] theorems [27], which state that:

1. The external potential is a unique functional of electron density.
2. The functional that delivers the ground state energy of the system gives the lowest energy if, and only if, the input density is the true ground state density.

---

[19] Pierre Hohenberg, French-American Physicist, 1934-2017

[20] Walter Kohn, Austrian Physicist, 1923-2016

In the first theorem, the *external potential* refers to the potential generated by the nuclei positions *i.e.* by the system geometry. The properties associated to it can be determined directly from the corresponding ground state electron density $\rho_0(\vec{r})$, according to the second theorem. Therefore, according to these two theorems, the PES is shaped by a functional of the ground state electron density. Indeed, as previously introduced and within the Born-Oppenheimer approximation, the PES relates to the geometry of a given system its total energy (see 2.1). In DFT, this energy, instead of being calculated from the nuclei positions, can be obtained directly from the ground state electron density of the system. Hence, molecular energy is a functional of $\rho_0(\vec{r})$:

$$E[\rho_0(\vec{r})] \tag{2.36}$$

If molecular properties can be determined through a functional of the ground state electron density, the second theorem ensures that the *true* value is approached by an upper bound. The ground state electronic energy determined through this functional will only be equal to the *true* value if the electron density is the *true* ground state electron density, it will be greater otherwise:

$$E[\rho(\vec{r})] \geq E_0[\rho_0(\vec{r})] \tag{2.37}$$

Determining the functional that enables the calculation of molecular properties from electron density, is the major challende in DFT. Because approximate functionals are used, it is not granted that $E_0[\rho_0]$ is approached by an upper bound. The Kohn-Sham[21] approach [33] reduces the impact of the lack of knowledge of this functional by evaluating the difference between the electronic energy of the considered system and that of a fictitious reference system of non interacting electrons, with a density identical to that of the studied system $\rho_{ref}(\vec{r}) = \rho(\vec{r})$.

In this manner the electronic energy of the system, $E[\rho(\vec{r})]$ defined by:

$$E[\rho] = \langle V_{Ne}[\rho] \rangle + \langle V_{ee}[\rho] \rangle + \langle T[\rho] \rangle \tag{2.38}$$

and the electronic energy of the fictitious reference system $E^{ref}[\rho(\vec{r})]$ defined by:

$$E^{ref}[\rho] = \langle V_{Ne}^{ref}[\rho] \rangle + \langle V_{ee}^{ref}[\rho] \rangle + \langle T^{ref}[\rho] \rangle$$

lead to the following expression for the electronic energy of the system of interest:

---

[21] Lu Jeu Sham, Chinese Physicist born in 1938

$$E[\rho] = -\sum_{\mu} Z_{\mu} \int \frac{\rho(\vec{r})}{4\pi\varepsilon_0\vec{r}} d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{4\pi\varepsilon_0\vec{r}_{12}} d\vec{r}_1 d\vec{r}_2 - \frac{1}{2m} \sum_{i=1}^{n} \langle\psi^{KS}|\nabla_i^2|\psi^{KS}\rangle + E_{XC}[\rho] \quad (2.39)$$

where:

$$-\sum_{\mu} Z_{\mu} \int \frac{\rho(\vec{r})}{4\pi\varepsilon_0\vec{r}} d\vec{r}$$

is the nucleus-electron attraction potential energy $\langle V_{Ne}[\rho]\rangle$. It is obtained from the same energy term in equation 2.12 when passing from the discrete to the continuum description of electrons defined in 2.35.

$$\frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{4\pi\varepsilon_0\vec{r}_{12}} d\vec{r}_1 d\vec{r}_2$$

is the electron-electron repulsion potential energy $\langle V_{ee}^{ref}[\rho]\rangle$ between electronic clouds 1 and 2 of an infinitesimal volume within the fictitious reference system. Indeed, when substracting $E^{ref}[\rho(\vec{r})]$ to $E[\rho(\vec{r})]$, the electron-electron repulsion energy difference is $\Delta\langle V_{ee}[\rho]\rangle = \langle V_{ee}[\rho]\rangle - \langle V_{ee}^{ref}[\rho]\rangle$ hence $\langle V_{ee}[\rho]\rangle = \langle V_{ee}^{ref}[\rho]\rangle + \Delta\langle V_{ee}[\rho]\rangle$. It has to be emphasized that considering electrons as a continuum negatively charged cloud, rather than considering them as point particles, leads to the electron-electron repulsion term being over-estimated because of electron self-interaction that arises.

$$-\frac{1}{2m} \sum_{i=1}^{n} \langle\psi^{KS}|\nabla_i^2|\psi^{KS}\rangle$$

is the electron kinetic energy of the reference fictitious system $\langle T^{ref}[\rho]\rangle$. Indeed, as for the electron-electron repulsion term, the substraction between the kinetic energies yields $\Delta\langle T[\rho]\rangle = \langle T[\rho]\rangle - \langle T^{ref}[\rho]\rangle$ hence $\langle T[\rho]\rangle = \langle T^{ref}[\rho]\rangle + \Delta\langle T[\rho]\rangle$. As the fictitious system consists in n non interacting electrons, their overall wavefunction Ψ^{KS} can be expressed exactly as a Slater determinant of one-electron Kohn-Sham spin-orbitals ψ^{KS} as defined in 2.16.

$$E_{XC}[\rho] = \Delta\langle V_{ee}[\rho]\rangle + \Delta\langle T[\rho]\rangle$$

is the exchange-correlation energy term that holds the differences between the real and the fictitious systems of n electrons. The exchange and correlation energies are accounted for in this term. The former is analogous to the exchange term that arises in wavefunction based methods (see equation 2.20) and that is considered to be a correction to the coulomb repulsion energy due to electrons of same spin that avoid each other. The latter is a further correction that accounts for correlation between electrons. Indeed, just as spin results in a *Fermi*[22] *hole*, *i.e* a reduced probability of finding electrons of same spin in a region of space, electrons treated as point particles also avoid each other more efficiently than what is estimated when treating them through negatively charged clouds. The correlation energy hence accounts for what is commonly named the *Coulomb hole* and that is only accounted for in post Hartree-Fock wave function methods.

$E_{XC}[\rho]$ is the only term that needs the knowledge of the functional. As it involves the deviations only in kinetic and repulsive potential energies between the two systems, an approximate knowledge of the functional has a small impact on the overall electronic energy.

Although DFT is not a wave function based method, for computational purposes an overall electronic Kohn Sham mathematical wave function Ψ^{KS} is used.

The electron density function of a poly-electronic system, central within this quantum approach, is then defined by:

$$\rho(x,y,z) = \sum_j n_j |\psi_j^{KS}|^2 \text{ with } n_j = 0,1,2 \tag{2.40}$$

where ψ_j^{KS} is the j^{th} Khon-Sham spatial MO of a set of orthogonal Kohn-Sham MOs, holding each n_j electrons that are used to build the poly-electronic Kohn Sham wave function Ψ^{KS} as in 2.16.

As in *ab initio* and semi-empirical methods, the spatial MOs ψ_j^{KS} are initially build from a LCAO (see equation 2.24) in order to have an initial *guess* of the electron density $\rho(\vec{r})$ that is subsequently refined in a self-consistent procedure. In fact, the same basis sets are employed in both wave function based and DFT approaches.

In an analogous manner to that described in the previous section, the energy $E[\rho]$ defined in equation 2.39 is minimized with respect to the ψ_j^{KS} MOs using the method of the Lagrange multipliers, with the constraint that the set of MOs remain orthonormal. This procedure leads to the Kohn-Sham equations:

$$\left[-\frac{1}{2m}\nabla_i^2 - \sum_\mu \frac{Z\mu}{4\pi\varepsilon_0} + \int \frac{\rho(\vec{r}_2)}{4\pi\varepsilon_0 r_{12}} + v_{XC}(p)\right] \psi_i^{KS}(p) = \varepsilon_i^{KS}\psi_i^{KS}(p) \tag{2.41}$$

---

[22] Enrico Fermi, Italian Physicist, 1901-1954

where

$$\hat{h} = -\frac{1}{2m}\nabla_i^2 - \sum_{\mu}\frac{Z_\mu}{4\pi\varepsilon_0} + \int\frac{\rho(\vec{r}_2)}{4\pi\varepsilon_0 r_{12}} + v_{XC}(p) \text{ is the Kohn-Sham operator}$$

$v_{XC}(p)$ is the exchange correlation potential

The determination of the exchange correlation potential gives rise to different levels of DFT implementation. Indeed, the accuracy depends on the exchange correlation energy $E_{XC}(\rho)$ and thus on the form of the functional to evaluate it. To name but a few, in order of increasing sophistication, the Local Density Approximation (LDA) functionals consider $E_{XC}(\rho)$ equal to the energy of an electron gaz of same density ρ. For the Generalized Gradient Approximation (GGA) functionals, the value of $E_{XC}(\rho)$ is approached in the same manner than by LDA functionals but considering that this value also depends in the gradient of ρ. Further sophistications, as in meta GGA functionals (MGGA), consist in also accounting for the second derivative of ρ and so on.

After selecting the functional for determining the exchange correlation potential, the solution of the set of $n/2$ equations (for a closed shell system of n electrons) is obtained following the previously adopted strategy, that is: *i)* orthogonalizing, *ii)* diagonalizing and *iii)* transforming $\mathbf{C}^\tau$ to get $\mathbf{C}$.

DFT methods demand far more computer resources than semi-empirical methods but, within the same computer time, they lead more accurate results than Hartree-Fock *ab initio* approaches because the latter cannot account for electron correlation.

> *"While solutions to the Hartree-Fock equations may be viewed as exact solutions to an approximate description, the Kohn-Sham equations are approximations to an exact description"*
>
> **Merrill GN, Gronert S, Kass SR** (1997)

A gain in accuracy can be achieved through larger basis sets and more accurate exchange-correlation energy functionals for Hartree-Fock and DFT methods, respectively.

2.2 Computer algorithms

2.2.1 Geometry Optimization

Geometry optimization is the procedure that enables to locate and characterize a stationary point on the PES. The initial structure is changed by the selected algorithm until a stationary point is found, *i.e.* the first derivative of the energy with respect to the geometric parameters cancels out (see equation 2.3). Subsequently, its nature is determined by the curvature at the considered stationary point, *i.e.* by evaluating the second derivative of the energy with respect to the geometric parameters.

Geometry optimizations, which consist in finding a minimum on the PES, are usually referred to as energy minimizations. Several algorithms perform this procedure. They differ in the employed strategy to assess which direction allows to approach the nearest minimum, and which size step choose when moving in this direction. The three main categories are:

- Derivative free methods: simulated annealing, simplex minimization and genetic algorithms,
- First derivative methods: steepest descent and multiple conjugate gradient minimization algorithms,
- Second derivative methods: Newton[23]-Raphson[24] and quasi-Newton algorithms.

The least sophisticated the method, the least time it spends at each point of the PES but the more steps have to be done before reaching the minimum.

2.2.2 Static Modes approach

The Static Modes Method [10, 11], developed at LAAS-CNRS, allows the evaluation of molecular deformations induced by external forces applied on a system of atoms. It was initially implemented to introduce full protein flexibility in docking algorithms. Indeed, flexibility plays a major role in the function of biomolecules, notably in the formation of macromolecular complexes. Yet, it is often discarded in the study of such processes because of the difficulty and associated computational cost in evaluating it.

To repair this flaw, the static modes approach harnesses information from the Hessian matrix, previously introduced (see section 2.1.1). The latter contains the force constants of the system, derived from a chosen energy model (empirical, semi-empirical, *ab initio*...), through which its constituent atoms are coupled. Hence, straining one of them prompts a response from those that are linked to it through the force constants stored in this matrix.

In the static modes approach, by associating force with displacement, the intrinsic deformations of an atomic system are evaluated *separately*. This is achieved by straining

---

[23] Isaac Newton, English scientist, 1643-1727

[24] Joseph Raphson, English Mathematician, 1648-1715

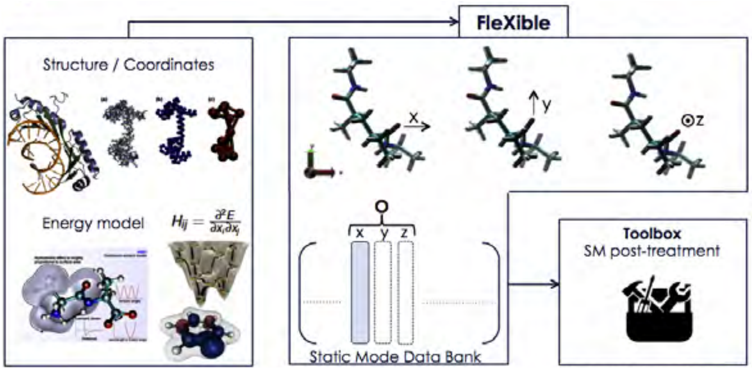


FIGURE 2.4: The Static Modes approach

each atom, in a specific direction of the conformational space and storing the displacement of all the other atoms in this manner induced. Thus, each *mode* is the elementary molecular deformation resulting from the excitation of one atom, in one direction x, y, z of the cartesian space.

To a molecular system constituted of n atoms, $3n - 6$ modes are associated. Indeed, each atom can be strained in each direction of 3D space such that: 3 solicitations are associated to translations of the entire system along each axis and 3 others to rotations of the entire system about each axis. These modes are stored to be subsequently used and, in this manner, efficiently calculate structural changes arising from external stimuli that mimic single or multi-atom contact. Indeed, the availability of the static modes, makes it possible to determine the overall response of the system to a custom designed constraint through combinations of the elementary modes.

The calculations consequently rely on a two-step procedure: *i)* a first calculation to obtain all the modes of a molecule, each one being the elementary deformation resulting from the forced displacement of a specific atom, *ii)* a subsequent combination of these modes to obtain the molecule response to the solicitation the operator wishes to mimic.

A variety of tools are available, allowing the user to design specific actions applied on single or multi-atom sites. In this manner, the impact on the molecule structure and mechanical properties of elaborated multi-direction constraints, involving several atoms, can be anticipated through an immediate and low-cost calculation. Reciprocally, a desired molecular shape can be imposed to obtain the constraint that would achieve such deformation, leading to the determination of the atoms that should be strained and the force that should be applied on them, specifying both the amplitude and direction of this force.

The static mode approach is used here to perform efficient, predictive and adaptable *in silico* experiments. The latter, simulate enzyme (in)activation mechanisms and screen

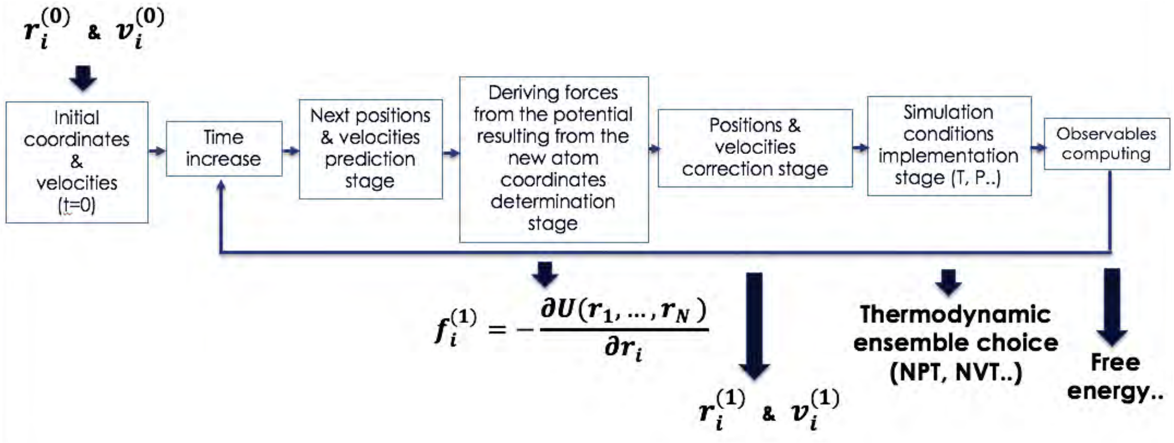


FIGURE 2.5: Molecular dynamics algorithm

mutations effects to propose a new roadmap for target discovery. The algorithms implemented demand insignificant computing time, specially if the resources used for obtaining the Hessian matrix are not accounted as part of the static mode procedure.

2.2.3 Molecular Dynamics

Molecular dynamics (MD), as inferred from its name, is a computer simulation method that aims at studying the dynamic, time-dependent behavior (as opposed to stationary) of a collection of elementary entities interacting with each other. The elementary entities can be atoms or bigger portions of a molecule as in coarse grained approaches. This method is widely spread as it enables the exploration of the PES, the temperature playing the role of an energy reservoir that makes it possible to overcome energy barriers between different energy wells.

In what follows, atoms are considered as the elementary entities.

The algorithm is based on the numerical integration of Newton's equations of motion applied to each atom forming the system:

$$\vec{F} = m\vec{a}(t) \tag{2.42}$$

From the initial coordinates $\vec{r}(t=0)$ and velocities $\vec{v}(t=0)$ of atoms forming the system under study (molecule, surface, condensed phase material among others), the trajectory of each atom is built in the course of the simulation, step by step, from the interactions it is involved in. These interactions are usually obtained from molecular mechanics force fields, interatomic potentials or single-point time-demanding quantum calculations.

Indeed, as forces between two atoms derive from their potential energy $\vec{F} = -\vec{\nabla} V(\vec{r}_{ij})$ and as the acceleration of the i^{th} atom is defined to be:

$$\vec{a}_i(t) = \frac{d^2\vec{r}_i(t)}{dt^2} \text{ with } \vec{r}_i(t) \text{ instant position of atom } i \tag{2.43}$$

the subsequent position and velocity of atom i are determined by:

$$\vec{r}_i(t{=}\Delta t) = \vec{r}_i(t{=}0) + \Delta t \cdot \vec{v}_i(t{=}0) + \frac{\Delta t^2}{2} \cdot \vec{a}_i(t{=}0) + O(\Delta t^4) \tag{2.44}$$

$$\vec{v}_i(t{=}\Delta t) = \vec{v}_i(t{=}0) + \Delta t \cdot \frac{\vec{a}_i(t{=}0) + \vec{a}_i(t{=}\Delta t)}{2} + O(\Delta t^4) \tag{2.45}$$

where Δt is the chosen time step and the acceleration terms are inferred from the employed force field/potential/quantum calculation.

In equation 2.45, the velocity of atom i is expressed adopting the Verlet algorithm. This algorithm is widely used in MD because of its stability, its smaller cumulative error compared to same order numerical integration schemes built from Taylor expansions, its time reversibility and the fact the it preserves phase space.

When evaluating the force applied on a specific atom i, an additional constraint can be applied to force the evolution of the system towards a specific state *i.e.* to a particular stationary point of the PES. One MD implementation of this kind is termed Steered Molecular Dynamics (SMD) as the exploration of these particular regions of the PES is accelerated compared to the amount of simulation time it would take to reach them with *unconstrained* MD.

Another Nobel-awarded implementation of MD is when the forces of a portion of the system are evaluated from *ab initio*, semi-empirical or DFT approaches, while the rest of the system, usually the larger part together with the solvent molecules (if the simulation is carried in explicit solvent) is treated with forces deriving from mechanical force fields or interatomic potentials. This hybrid implementation is called Quantum Mechanics / Molecular Mechanics (QM/MM) approach [78] and has been widely used in particular within biological macromolecules in order to study active regions at the quantum level. This approach enables chemical reactions to take place under the influence of the rest of the biomolecule, which usually acts through key steric effects.

Copyright ©: G. Groenhof, *Met. Mol. Biol. (2013)*

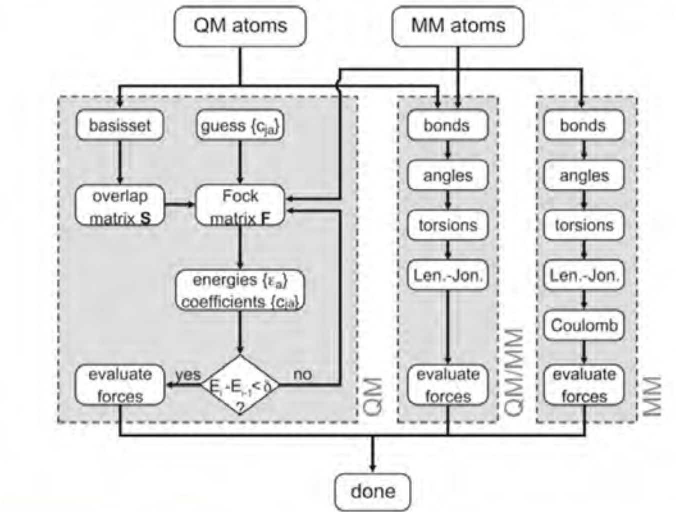


FIGURE 2.6: QM/MM molecular dynamics algorithm.

2.3 Simulation analysis tools

The results obtained from molecular modeling calculations are usually subject to diverse post-processing analysis in order to yield system properties of interest. For each method previously introduced, a few of the corresponding analysis tools are described in detail in the following sections.

2.3.1 Analyzing Molecular Dynamics trajectories

As molecular dynamics (MD) consist in constructing the trajectory of each atom during the simulation, dynamic properties of the system can be calculated. Among several analysis that are usually carried out, a few are presented in the following sections.

System properties

To assess the quality and stability of trajectories obtained from MD simulations, system properties as the potential, kinetic and total energies but also temperature, volume, density and pressure are usually plotted as a function of time. Common flaws can be inferred from this plots, as:

- Instabilities by discontinuities in the potential or total energy,
- System explosion by a sudden increase in volume,

- Constituent atoms overlap by spikes in temperature,
- Insufficient solvation layer, when running simulations in explicit solvent, by an unexpected density value.

Root Mean Square Deviation

The Root Mean Square Deviation (RMSD) from the starting positions of atoms can be calculated and plotted against the simulation elapsed time. In simulations involving proteins, this analysis is usually carried out for the backbone atoms in order to identify the *equilibration* from the *production* phase. The equilibration phase consists in the initial simulated time during which the observables defining the chosen ensemble (NPT or NVE usually) adjust to their selected value. The variations in temperature, pressure or simply the relaxation from the constraints induced from crystalizing the protein (as usually starting structures come from X-ray crystallographic data) are accompanied by a steady increase in the RMSD. The production phase starts when equilibrium is considered to be reached. In this phase, the observables defining the chosen thermodynamic ensemble fluctuate around their selected value. The RMSD of the overall protein backbone is expected to stop increasing. The exploration of the PES starts during this phase.

The RMSD of individual residue side-chains can also be evaluated to assess their intrinsic mobility. Their plot is useful to identify the conformational changes they undergo during the simulation.

Native contacts

The *native/non-native contacts* analysis consists in identifying, for a set of atoms, possible interacting partners according to a distance criterion. Atoms that satisfy this criterion in the initial structure are considered to form *native contacts* with the set of atoms under analysis. Atoms that get within the selected threshold distance in the course of the simulation, form *non native contacts*. The strongest *native/non-native contacts* can be considered to be *i)* the ones that remain the longest, *ii)* the ones that involve several atoms from the same residue.

Solvent presence

Radial Distribution Function To assess the presence of solvent molecules around a particular atom, molecule or region in general, the Radial Distribution Function (RDF) appears as a useful analysis tool to identify hydration shells (assuming water as the solvent). This analysis consists in determining the number of water molecules within a certain *radius* of the atom, molecule or region of interest. Usually, the results from this analysis are plotted against the distance from the selected atom, molecule or region in order to identify at which distance from the considered entity lays the highest probability of finding a solvent molecule. The plot depicts hence solvent density as a function of distance from the chosen center. The first peak is often interpreted as the first hydration shell, the second as the second hydration shell and so on, in simulations run in explicit solvent.

2D RDF As RDF evaluates solvent presence according to a single parameter *i.e.* distance, water molecules density appears as a 1D plot. In order to map water molecules in 2D, an algorithm analogous to usual RDF was implemented but such that water molecules are projected on a plane of interest. This representation makes it possible to identify the location and extent of hydration sites rather than just the distance that separates them from the selected center. To define the plane of projection, an orthonormal basis is built from the coordinates of three points corresponding to three atoms, molecules or regions centers. These coordinates can be chosen a single time at the beginning of the simulation run or at each step. The latter allows to account for the changes of the selected coordinates when these are defined from the centers of atoms or molecules.

2.3.2 Population analysis

Ab initio, semi-empirical and DFT calculations, which consider electrons explicitly, enable to assign partial charges to the constituent atoms. This is achieved by partitioning the overall poly-electronic wave function Ψ_{el} (or Ψ_{el}^{KS} in DFT) into atomic contributions according to the weighting coefficients c_{rj} of each AO ϕ_r from the basis set.

Adopting the definition of electron density 2.40 and using one-electron orthonormal spin-orbitals ψ_j built from a LCAO as in 2.24, leads to the total number N of electrons being determined from:

$$N = \sum_j \left(\sum_r c_{rj}^2 + \sum_{r \neq s} c_{rj} c_{sj} S_{rs} \right) \text{ as } N = \int \rho(\vec{r}) d\vec{r} \tag{2.46}$$

where $S_{rs} = \int \phi_r(\vec{r}_j)\phi_s(\vec{r}_j)d\vec{r}_j$ is the overlap of AOs $\phi_r(\vec{r}_j)$ and $\phi_s(\vec{r}_j)$. It appears hence that each MO ψ_j holds contributions provided by AO ϕ_r only, and by the overlap between AOs ϕ_r and ϕ_s.

When trying to assess how these N electrons are distributed, for each AO ϕ_r of the μ^{th} atom, the charged it holds centered on this atom comes from the c_{rj}^2 contributions it makes to every one-electron spin-orbital ψ_j. Hence, the charge centered on atom μ held by its AO ϕ_r is:

$$n_r = \sum_j c_{rj}^2 \tag{2.47}$$

It is called Mulliken net population as Robert Mulliken[25] was the first to propose this scheme for attributing partial charges to atoms in molecules.

The amount of electrons shared between atom μ and others within the molecule, due to overlap between AOs ϕ_r and ϕ_s, is obtained by the contribution this overlap S_{rs} makes to each one-electron spin-orbital ψ_j. Therefore the Mulliken overlap population is:

$$n_{r/s} = \sum_j 2 c_{rj} c_{sj} S_{rs} \tag{2.48}$$

The electron population on atom μ due to its AO ϕ_r, named Mulliken gross population, is the sum of its Mulliken net population *i.e.* electrons centered on this atom by AO ϕ_r, plus half the Mulliken overlap population *i.e.* electrons shared by this atom through AO ϕ_r:

$$N_r = n_r + \frac{1}{2} n_{r/s} \tag{2.49}$$

In this manner, Mulliken decided to attribute an equal share of electrons to each atom μ forming a bond with another atom through the overlap of one of its AO ϕ_r. A different partitioning could be adopted accounting for electronegativity of the species involved.

Finally, the overall electronic charge on atom μ within the molecule is given by:

$$N_\mu = \sum_{r \in \mu} N_r \tag{2.50}$$

---

[25] American Physicist and Chemist, 1896-1986

This quantity is named Mulliken reduced partial charge as the actual *partial charge* δq_μ of atom μ within a molecule is considered to be the sum of its nuclei charge and its overall electronic charge *i.e* $\delta q_\mu = Z_\mu - N_\mu$.

Charges obtained through Mulliken population analysis do not converge to a unique value when using larger basis sets to obtain more accurate overall poly-electronic wave functions Ψ_{el} (or Ψ_{el}^{KS} in DFT). Moreover, this analysis can result in AOs carrying a number of electrons superior to the quantity allowed by the Pauli exclusion principle or inferior to zero.

To palliate these problems, Löwdin proposed to symmetrically orthogonalize the AOs forming the basis set before carrying out the population analysis. This simple yet ingenious operation leads to far more stable reduced partial atomic charges N_μ when using larger basis sets. Nevertheless, Löwdin reduced partial charges also become unstable when using diffuse functions so, the use of very large basis sets remains unsuitable when carrying this type of analysis.

Chapter 3

Results

Members of the Ras family, which was described in detail in chapter 1, share 95% of overall residue identity. The elucidation of structural features that play a major role in its function hence appears to be fundamental for understanding the regulation of a plethora of cellular processes. In this chapter, the active site of NRas catalytic domain, formed of the previously described *P-loop*, *Switch I* and *Switch II* regions, is studied in detail in its GTP bound conformation (see figure 3.1).

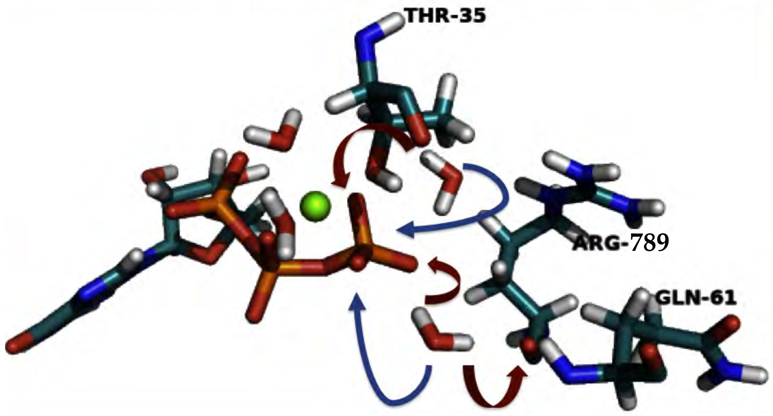


FIGURE 3.1: NRas catalytic domain.

The wild-type protein as well as Gln 61 mutants are analyzed to identify characteristics of the nucleotide binding site that contribute to catalysis. Indeed, substitutions of this residue systematically impair both the intrinsic and GAP enhanced hydrolysis rate, leaving NRas in its GTP bound active state that participates in the development of cancerous tissues. Moreover, not only Gln 61 residue is the most recurrent substitution in NRas (60% [56]) but also, identifying its implication in the enzymatic hydrolysis of GTP, would probably be valid for all Ras family members. Using the methods described in chapter 2, NRas active site structural characteristics are evaluated, the intrinsic properties of the ligand in this environment are disclosed and the impact of Gln 61 mutations assessed. A final part is devoted to the study of engineered double mutant proteins that have been proposed as possible restoring enzymatic function candidates.

3.1 Wild-type NRas active site according to classical MD

A major question, which is also the first to arise when planing to characterize a molecule through computational chemistry methods, is: which level of theory implemented through which algorithm should be used? The choice depends of course on what is aimed to be learned about the molecule, defining thereby the level of accuracy to be employed, but it is also based on the availability of computer resources.

In chapter 2, molecular modeling methods were presented in order of both increasing accuracy and need of computer resources. Following this rating, the 3D structure of wild-type NRas is first characterized at the classical level through Molecular Mechanics (MM) force fields during Molecular Dynamics (MD) simulations.

Indeed, MM MD simulations demand the least computer resources while providing information about the dynamic behavior of the protein structure. The different conformations it can adopt are then sampled to the extent allowed by the chosen force field.

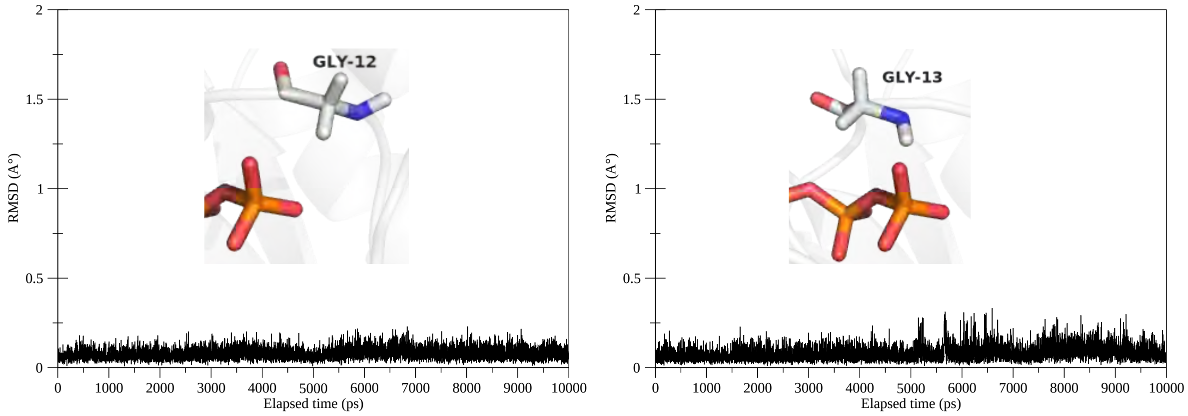
The obtention of the initial 3D structure and the simulation setup are described in appendix A. Results obtained through this approach are presented and discussed in what follows.

3.1.1 Active site conformation

To evaluate the conformational changes undergone by key NRas active site residues, the Root Mean Square Deviation (RMSD) [58] from the initial position of their side chains was plotted against the simulation time *i.e.* RMSD(t). This straightforward analysis is however insufficient to depict the behavior of GAP Arg 789 residue. Instead, for this residue only, the lifetime of its strongest *native/non-native contacts* [58] is represented, enabling thereby to infer the conformational changes this residue undergoes.

Several NRas residues that have been identified to play a key role in the protein function through mutational assays and/or that are conserved within the superfamily, namely Gly 12, Gly 13, Thr 35 and Gly 60, present small fluctuations ($\sim 0.1 - 0.2$ Å) around a unique conformation already adopted in the protein crystallographic structures [48, 62, 63] used for these simulations (see figure 3.2).

(a) Glycine 12 (b) Glycine 13



(c) Threonine 35 (d) Glycine 60

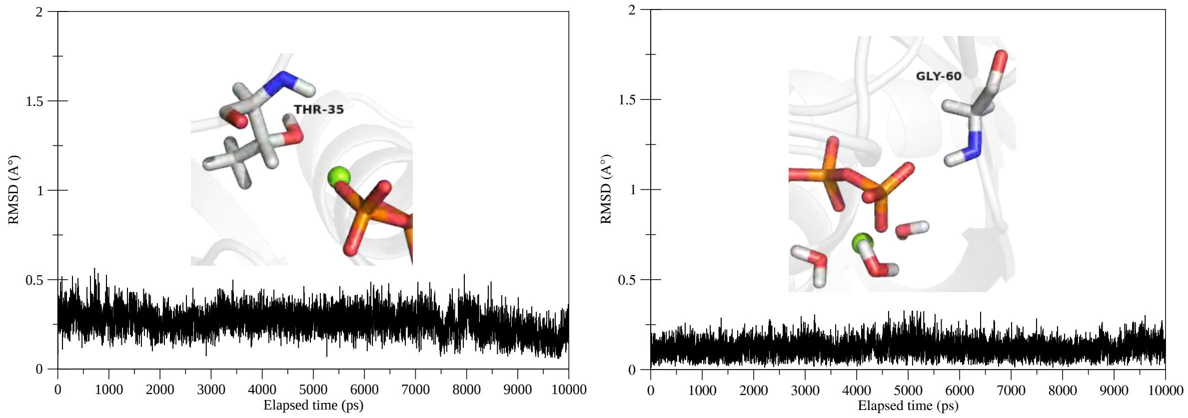
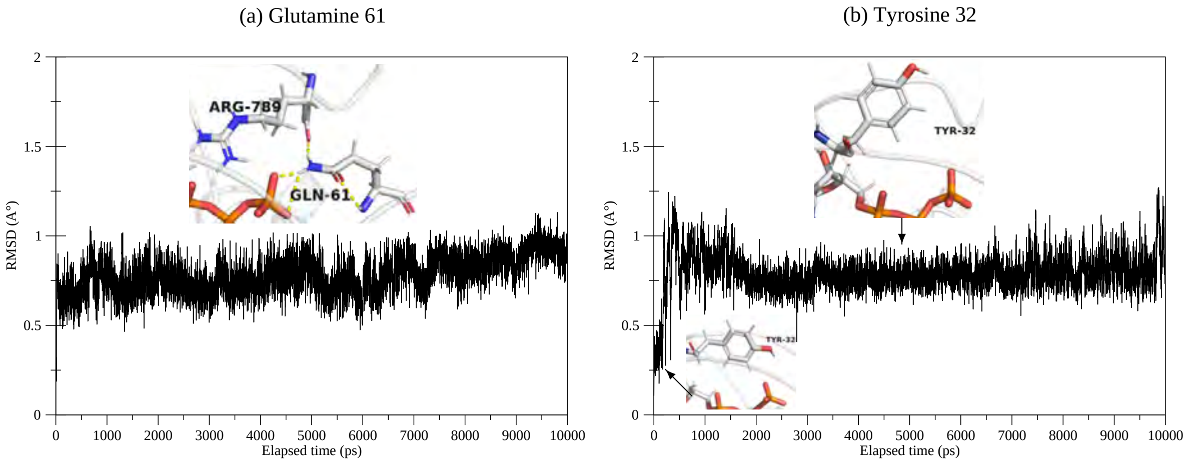


FIGURE 3.2: RMSD(t) plot of Gly 12 (a), Gly 13 (b), Thr 35 (c) and Gly 60 (d) from wild-type NRas during MM molecular dynamics.

Other important active site residues adopt multiple conformations that need to be closely inspected to determine their role in protein function.

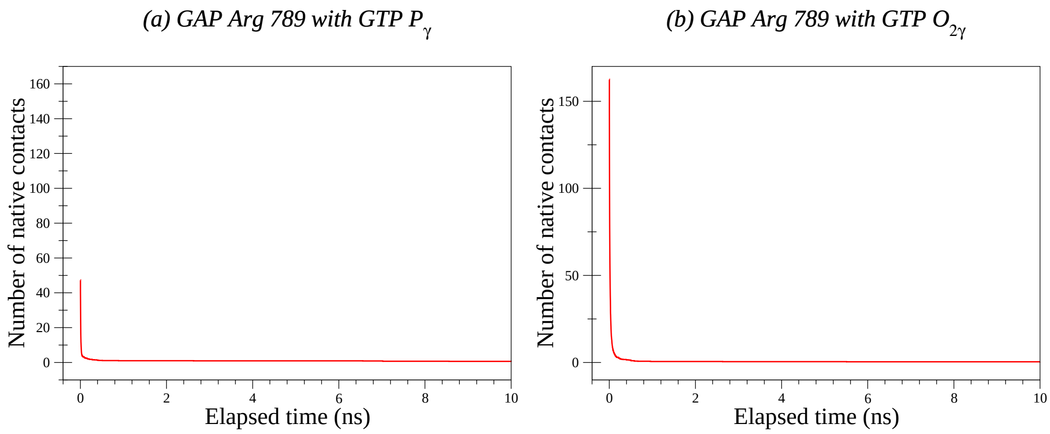
Gln 61 RMSD(t) plot presents an initial jump of 0.7 Å related to the amine group of its side chain. This group, initially bent towards Ala 59 residue, moves towards GTP γ-phosphate group and GAP Arg 789 residue. Stable interactions are in this manner established between Gln 61 amine group and GTP γ-phosphate and Arg 789 backbone oxygen atoms. The conformation thereby adopted remains throughout the simulation with small fluctuations around it $\sim 0.2 - 0.3$ Å (see figure 3.3). According to these results, Gln 61 could play a role in stabilizing the transition state [43, 55, 62, 65, 66, 68] or in positioning the lytic water molecule [57, 63, 70, 74], as previously proposed in the literature. However, a role as the base in the hydrolysis reaction [20, 37, 50] is excluded for Gln 61 as its side chain oxygen atom remains far from the reactants ($\sim$ 4.6 and 4.7 Å away from a presumed lytic water molecule and from GTP P_γ atom, respectively).


(a) Glutamine 61

(b) Tyrosine 32

FIGURE 3.3: RMSD(t) plot of Gln 61 (a) and Tyr 32 (b) from wild-type NRas during MM MD.

Tyr 32 residue is found in two conformations, previously identified in Geyer et al. NMR spectroscopy study [21]. The first one, adopted in the initial structure, corresponds to its side chain hydroxy group engaging in hydrogen bonds with GTP O_γ atoms. The second one, corresponds to this same group shifted away from the GTP γ phosphate group. Both, the *closed* and *open* conformations, respectively, are depicted in figure 3.3 (b). Within the simulated time, the *open* one is mainly adopted. However, it is not excluded that it might switch back to a *closed* state if the simulation is pursued as it has been shown, also through NMR spectroscopy[36], that *Switch* regions present dynamic fluctuations in the nanosecond timescale.


(a) GAP Arg 789 with GTP P_γ

(b) GAP Arg 789 with GTP $O_{2\gamma}$

FIGURE 3.4: Lifetime curves of GAP Arg 789 *native contacts* with GTP P_γ atom (a) and $O_{2\gamma}$ atom (b) within wild-type NRas during MM molecular dynamics.

The lifetime curve of GAP Arg 789 *native contacts* with GTP P_γ atom shows that they are maintained throughout the simulation (see figure 3.4 *(a)*). Strong *native contacts* with $O_{2\gamma}$ atom also remain during the entire simulation (see figure 3.4 *(b)*) while *non-native contacts* are formed mainly with Gln 61 residue.

As water molecules are among the reactants in the hydrolysis reaction of GTP and as the number of solvent molecules implicated during the enzymatic catalysis is subject to debate, we proceed to evaluate water presence in wild-type NRas active site.

3.1.2 Water distribution

The radial distribution function (RDF) of water molecules around GTP P_γ atom was computed so that solvent presence within wild-type NRas active site could be estimated. The corresponding plot is presented in figure 3.5.

A first peak of water density arises at $\sim$ 3.5 Å from GTP P_γ atom, with an amplitude of 0.26. A second peak of water density is observed at $\sim$ 6.1 Å from GTP P_γ atom. Then, at $\sim$ 7 Å from this atom bulk solvent density starts.

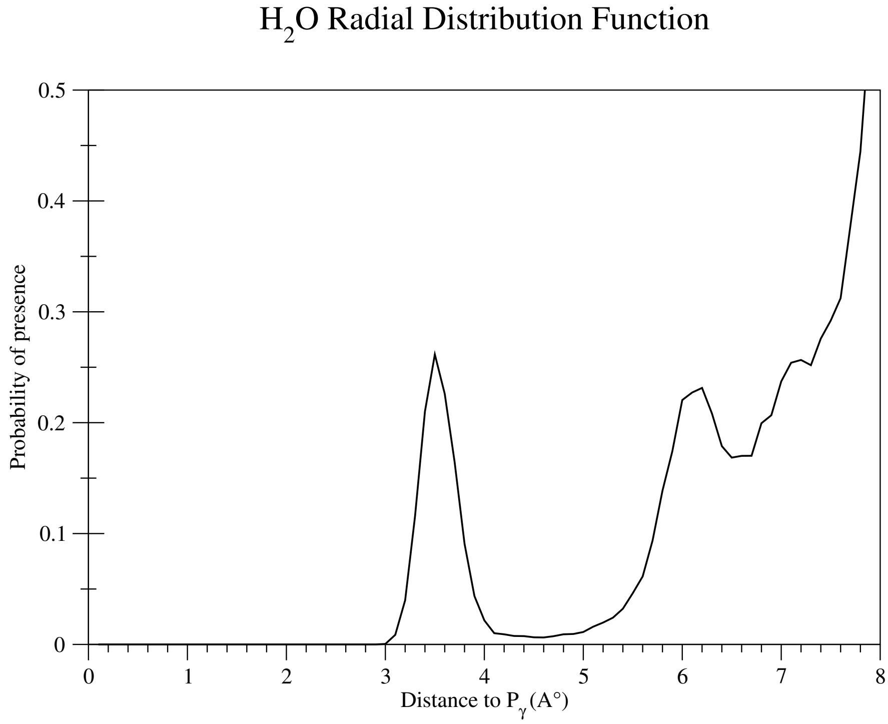


FIGURE 3.5: RDF of water molecules around GTP P_γ atom within wild-type NRas during MM molecular dynamics.

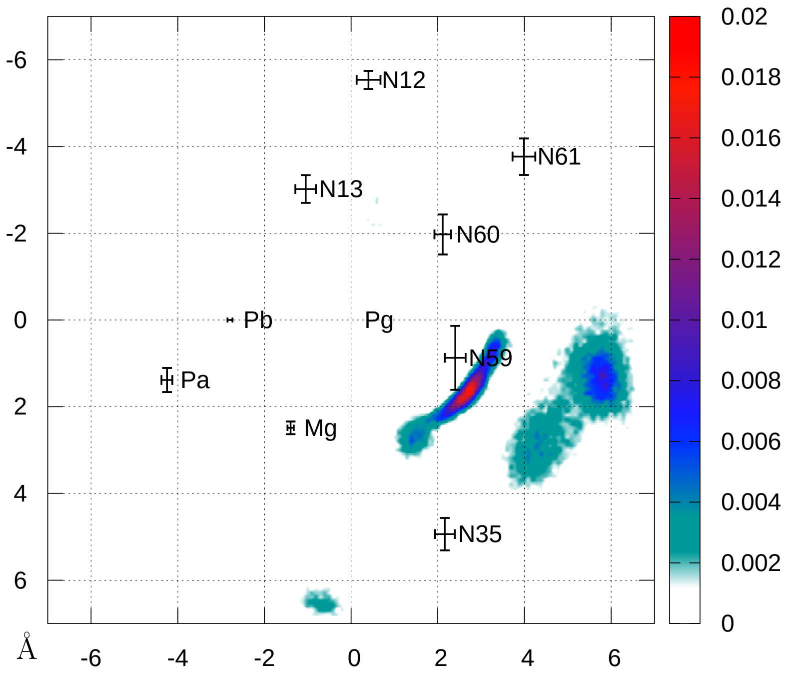


FIGURE 3.6: 2D RDF of water molecules within wild-type NRas active site during MM molecular dynamics. The average coordinates of nitrogen backbone atoms from residues 12, 13, 35, 59, 60 and 61, the phosphorus atoms from GTP and the Mg^{2+} ion are also projected. The corresponding standard deviation values are represented with error bars.

In an attempt to precisely localize the closest water molecules to GTP P_γ atom, which appear to be in a 3.5 Å radius of this atom according to the RDF plot (figure 3.5), an algorithm that maps regions of the active site where water molecules tend to remain was designed. This algorithm is described in detail in appendix B. The 2D water presence probability density function is thereby determined. Here a 7 Å threshold distance from GTP P_γ atom was taken for evaluating the 2D water distribution within wild-type NRas active site.

Figure 3.6 shows that water density is high in the vicinity of residues 35 and 59, the highest density peak being placed next to residue 59. This positioning can be correlated to a presumed lytic water molecule identified by previous studies [20, 50, 55, 62, 63]. Moreover, a single water molecule, besides the two water molecules coordinated to the Mg^{2+} ion, appears to occupy a place in the active site close to GTP P_γ atom.

To validate the obtained wild-type NRas active site conformation together with water molecules positioning in it, their effect on the bound nucleotide electronic density is subsequently assessed. Indeed, as described in detail in chapter 1, previous FTIR [3, 4, 14, 18] and Raman [77] spectroscopy studies, have reported a strong binding of the protein to the GTP β phosphate group, thereby accommodating GTP electronic density

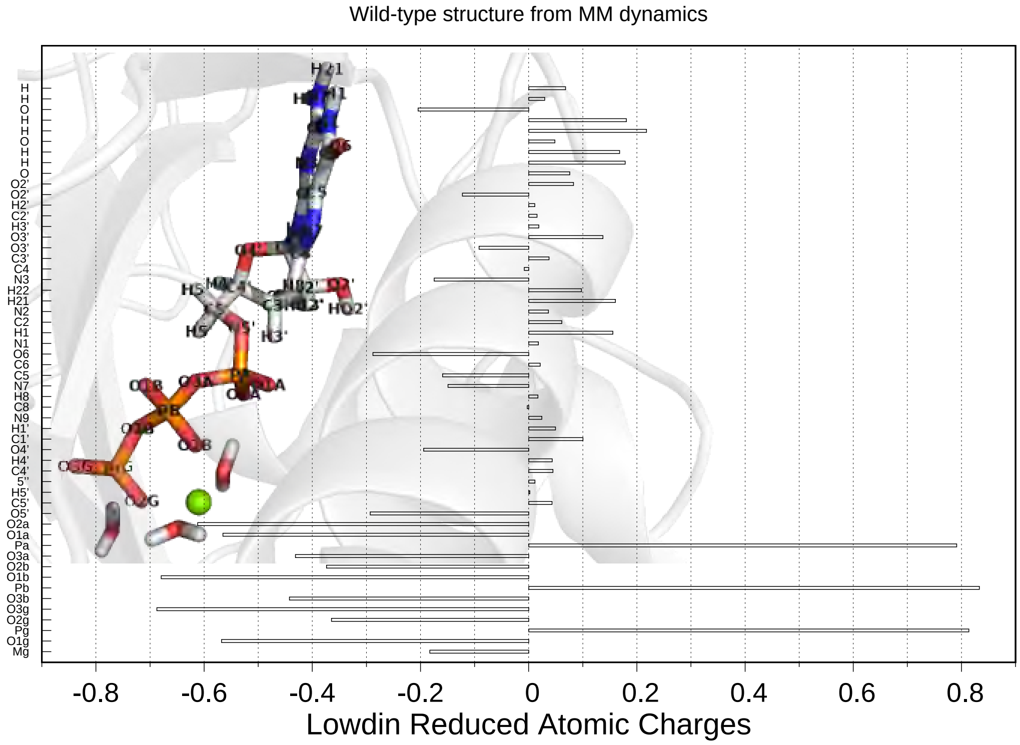


FIGURE 3.7: Löwdin population analysis of GTP, Mg^{2+} and water molecules in wild-type NRas active site conformation according to MM molecular dynamics.

to a more GDP like state. The comparison of the ligand electronic distribution with experimental data allows hence to determine the likelihood of the previously presented theoretical results concerning NRas active site structure.

3.1.3 GTP electronic density

To observe how electronic charges are distributed among GTP atoms, energy minimizations were performed treating the cofactors and water molecules in their vicinity at the DFT level of theory. The effects of the protein together with the bulk solvent surrounding the DFT system were taken into account within a classical framework through MM forcefields. The simulations setup is described in detail in appendix A.

GTP electronic distribution was then assessed through a Löwdin population analysis. Individual partial charges are thereby attributed to each atom forming this molecule (see figure 3.7). Nevertheless, as this analysis is not suitable for attributing *absolute* charges to atoms, but rather for comparing *relative* differences within a system, the overall Löwdin reduced charge held by each GTP phosphate group was calculated to identify the relative differences among these groups.

The sum of reduced charges attributed to atoms belonging to the same phosphate group shows that the overall negative charge held by the γ-phosphate is bigger than that of the β-phosphate group when GTP is bound to wild-type NRas active site (-0.805843 a.u. vs -0.662219 a.u. for GTP γ and β phosphate groups, respectively). This GTP electronic distribution is in disagreement with the numerous experimental investigations [3, 4, 14, 18, 77] that concluded that Ras catalyzes GTP hydrolysis by stabilizing a more GDP like state in which the β phosphate group holds more negative charge than the γ phosphate.

This disagreement could be due to the level of theory employed for assessing NRas active site conformation and water distribution within it. Indeed, during MM dynamics, the neglect of electrons might prevent residues to adopt certain conformations that are dictated by the geometries of the atomic orbitals and that, as a consequence, cannot be attained when these are neglected. Moreover, in the study of a protein surface exposed to solvent, as performed here, this level of theory certainly overlooks the electrostatic coupling between protein residues and water molecules. Since the subsequent DFT/MM minimizations to assess GTP charge distribution, do not treat the active site at the quantum level, its overall conformation remains described by classical force-fields during this procedure.

To ascertain if this possibility is the one leading to the discrepancy between the MM MD simulations results and the experimental ones, we proceed to employ a higher level of theory to investigate wild-type and Gln 61 mutant NRas active site conformations together with water molecules positioning inside it.

3.2 NRas active site according to QM/MM dynamics

The size of proteins, makes their study at a quantum level hardly attainable during MD simulations. Indeed, even when using semi-empirical hamiltonians, such simulations carried on systems containing more than a hundred atoms become extremely expensive both in time and computer resources. A quantum mechanics/molecular mechanics (QM/MM) approach [78] is then adopted. This hybrid scheme enables to treat residues of interest at a quantum level while accounting for the influence of the entire protein, together with the bulk solvent, within a classical, less demanding, framework.

3.2.1 Individual residues intrinsic mobility

The intrinsic motions of key NRas active site residues are studied through QM/MM MD simulations. These calculations were performed on the wild-type and on six Gln 61 substituted proteins (Q61E, Q61P, Q61H, Q61L, Q61K, Q61R) to measure the structural impact of such oncogenic mutations. The initial structures include the *arginine binding loop* of GAP to account for the electrostatic and steric effects of the so called *arginine finger*. The procedure to obtain these structures is described in detail in appendix A.

As water molecules tightly bound to proteins have distinctly different properties from water molecules of the bulk solvent and can be essential to the protein structure and function, water molecules present in the active site (*i.e.* within 5Å of GTP P_γ atom) were also treated quantum mechanically. The electrostatic coupling between these water molecules and the active site residues is in this manner taken into account.

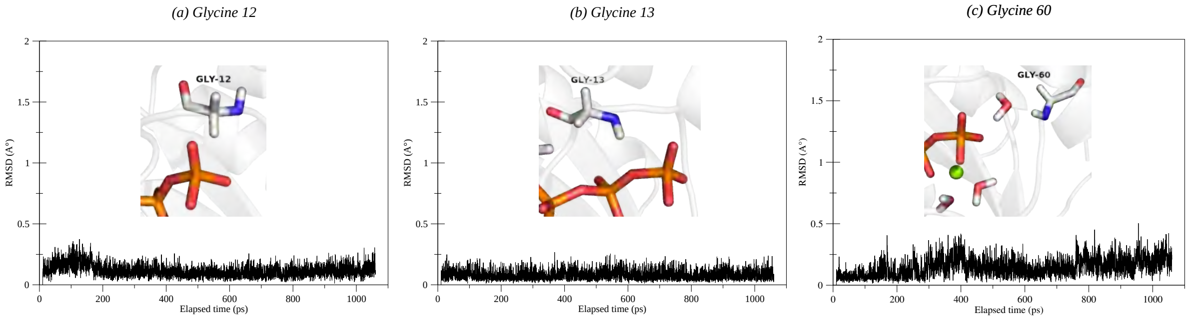
The simulation setup can be found in appendix A. The obtained trajectories were post-processed using analysis methods described in chapter 2. Conformational changes undergone by quantumly treated residues (Gly 12, Gly 13, Tyr 32, Thr 35, Gly 60, 61, Arg 789) were identified, as when analyzing MM MD of the wild-type *i.e.* by plotting the RMSD from the initial conformation of their side-chains against the simulated time, RMSD(t). Because such straightforward analysis poorly depicts the observed [28] behavior of GAP Arg 789 residue, the lifetime of its strongest *native contacts* was instead computed for this residue only. In this manner, only based on a distance criterion, the evolution of the possible interactions the arginine finger can engage in due to different adopted conformations is monitored.

Comparing results for wild-type and mutant NRas sheds light on how these residues intrinsic motions, stability and interacting neighbors are impacted by different Gln 61 substitutions.

Wild-type p21^{N-ras}

The RMSD(t) plot of residues 12, 13 and 60 within the wild-type present very small fluctuations, approximately 0.1 - 0.2 Å, around a unique conformation (see figure 3.8).

It appears that these residues have a reduced mobility in the active site of GTP bound wild-type NRas. For residues 12 and 13, this is in accordance with an NMR study [36] that concluded that the region comprising residues 10 to 18 has no conformational flexibility in solution.


(a) Glycine 12

(b) Glycine 13

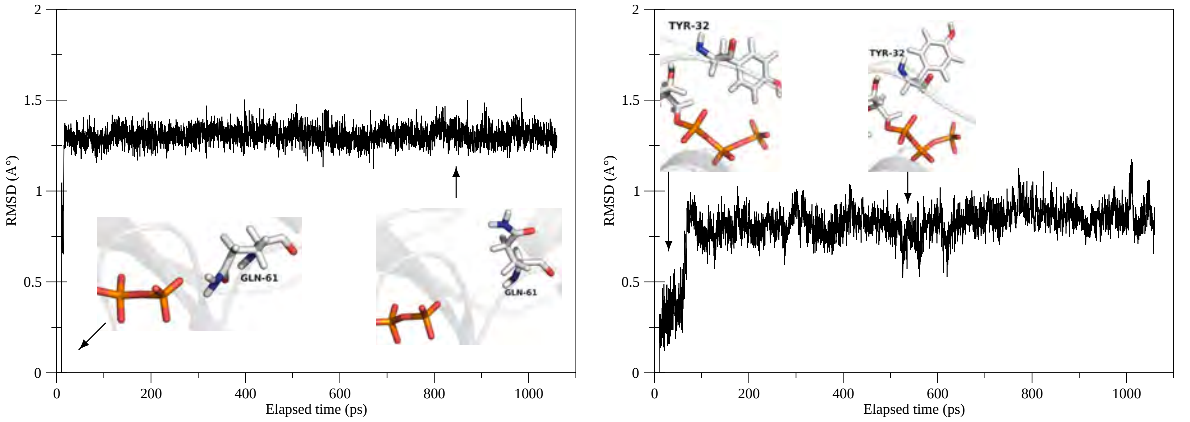
(c) Glycine 60

FIGURE 3.8: RMSD(t) plot of Gly 12 (a), Gly 13 (b), and Gly 60 (c) from wild-type NRas during QM/MM molecular dynamics.

Residues 61, 32, 35 adopt, on the other hand, multiple conformations. This is again in accordance with Kraulis et al. NMR study [36] that attributed to the poor resolution in crystallographic structures of *Switch I* and *Switch II* regions, a genuine mobility of the residues forming them. This urge us to describe in detail their RMSD(t) plots to identify, at the atomic level, the multiple adopted conformations. GAP arginine finger multiple conformations are assessed by the lifetime curve of its strongest *native contacts*.

The fluctuations of the presumed catalytic Gln 61 residue are, as for residues 12, 13 and 60, in the $\sim 0.1 - 0.2$ Å range, although its RMSD(t) plot presents an initial jump of ~ 1.25 Å occurring during the equilibration phase of the simulation (figure 3.9 *(a)*). Gln 61 side chain, initially bent towards the GTP accordingly to crystallographic structures 1WQ1 [62] and 5UHV [30], rapidly unbends, getting out of the active site. It remains in this state for the rest of the simulation time as reported by previous dynamical studies [57, 70].

Tyr 32, which is part of the *Switch I* region, is initially *closed* with its side-chain bent towards the ligand so that its hydroxy group engages in hydrogen bonds with GTP O_γ atoms. It subsequently gets *open* as its side-chain moves out of the active site. Both conformations are stable during the simulation and have both been identified in an earlier NMR study [21]. Indeed, Geyer et al. explain that the ensemble of distant *i.e. open* conformation of Tyr 32, relative to the nucleotide, and the *closed* one, constitute two different chemical environments surrounding the β phosphate group. These two chemical environments induce the splitting they observe of this phosphate group resonance vibration. During the simulated time and within the wild-type, the *open* one remains the longest (see figure 3.9 *(b)*) but it is not excluded that Tyr 32 side chain might get *closed* again as Kraulis el al. [36] showed that *Switch* regions are flexible in solution and have internal motions in the nanosecond time scale.


(a) Glutamine 61

(b) Tyrosine 32

(c) Threonine 35

(d) Native contacts of GAP Arg 789 with GTP Pg atom

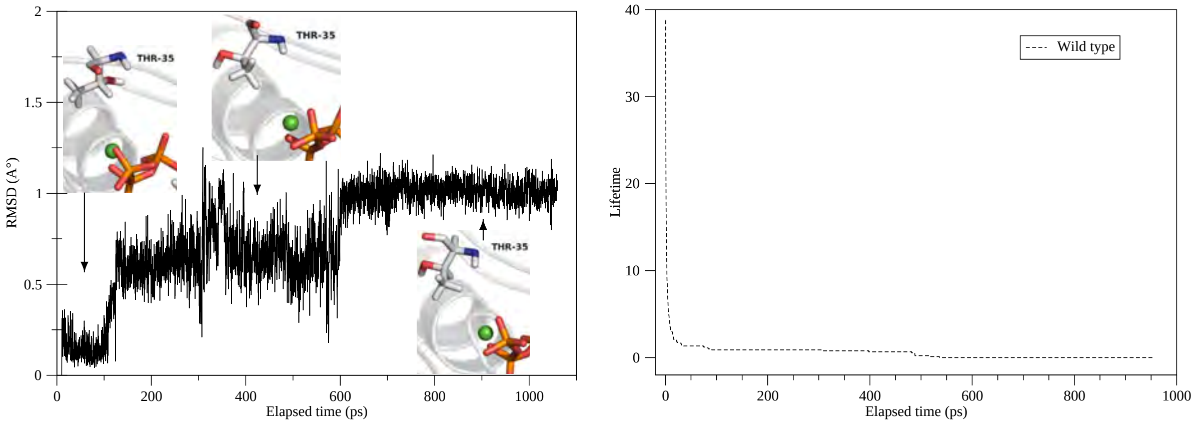


FIGURE 3.9: RMSD(t) plot alongside the associated conformational changes of Gln 61 (a), Tyr 32 (b), Thr 35 (c) from wild-type NRas and lifetime curve of *native contacts* between GTP P_γ atom and GAP Arg 789 residue (d) during QM/MM molecular dynamics.

Thr 35 RMSD(t) plot presents three main stable stages that are associated to three main conformations adopted by this residue (figure 3.9 *(c)*). The transition between the first two steps is associated to the side chain hydroxy and methyl groups that switch places: the hydroxy group, initially towards the inside of the active site, moves outwards and, the methyl group, initially outwards, moves inwards. In the third stage, Thr 35 side-chain stretches further, its methyl group getting even closer to the Mg^{2+} ion and the backbone oxygen atom, initially pointing inwards the active site, moves to point outwards. The three conformations of this residue, that also belongs to the *Switch I* region, are represented in the insets of figure 3.9 *(c)*. The clearly identified and stable conformations adopted by Thr 35 methyl group might be part of the structural inherent role associated to this chemical group by Spoerner et al. [71]. In their study that combines NMR spectroscopy, X-ray crystallography and binding affinity essays, they suggest that Thr 35 methyl group regulates Ras dynamical flexibility, necessary for its molecular switch function and for its interactions with numerous effector molecules. Indeed, their mutational assays that omits it (T35S) results in a reduced affinity for Raf-RBD.

Finally, the conformational changes of GAP Arg 789 side chain are inferred from the lifetime of its strongest *native contacts* [58]. This curve depicts the span of potential interactions the arginine finger could engage in, based on a 6 Å distance criterion and that were present in the starting structure. Within the wild-type, these contacts are made with GTP P_γ atom and, although the lifetime decays during the simulation, at least one Arg 789 atom satisfies the selected criterion during the entire simulation (see figure 3.9 *(d)*). Other strong *native contacts* with atoms from the γ phosphate group also persist ($O_{1\gamma}$, $O_{2\gamma}$) and several *non-native contacts* develop during the simulation, notably with GTP $O_{3\beta}$ and $O_{3\gamma}$ atoms. *Non-native contacts* also arise with atoms from Gln 61 residue.

Gln 61 mutants

As within the active site of wild-type NRas, the RMSD(t) plots of residues 12, 13 and 60 within Gln 61 mutant proteins present small fluctuations ($\sim 0.1 - 0.2$ Å). These residues are not affected by Gln 61 substitutions and remain in their initial conformation.

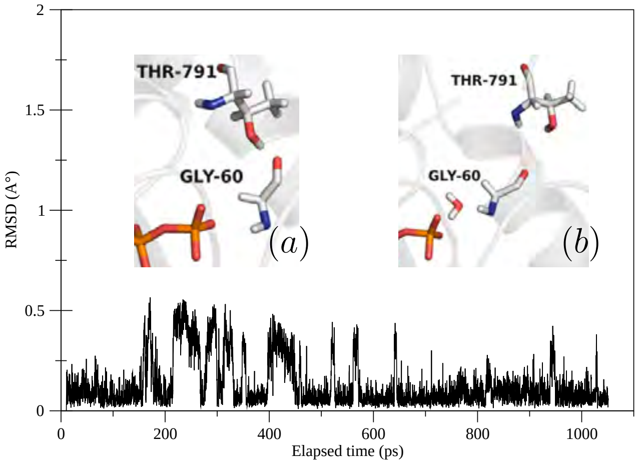


FIGURE 3.10: RMSD(t) plot of Gly 60 within Q61R mutant during QM/MM molecular dynamics. This residue is shown forming hydrogen bonds with: GAP Thr 791 residue within Q61R (a), a transient water molecule within wild-type NRas (b).

Q61R is an exception. Within this mutant, even though the fluctuations in the RMSD(t) plot of Gly 60 residue remain small ($\sim$ 0.2 Å), two different stages can be identified (see figure 3.10). They correspond to two conformations adopted by the backbone oxygen atom. Initially pointing out of the active site, it turns inwards because of remarkably stable hydrogen bonds formed with GAP Thr 791 residue, depicted in figure 3.10 inset *(a)*. This hydrogen bonding cannot occur within the wild-type because of different orientations adopted by both residues that make impossible such interaction as depicted in figure 3.10 inset *(b)*.

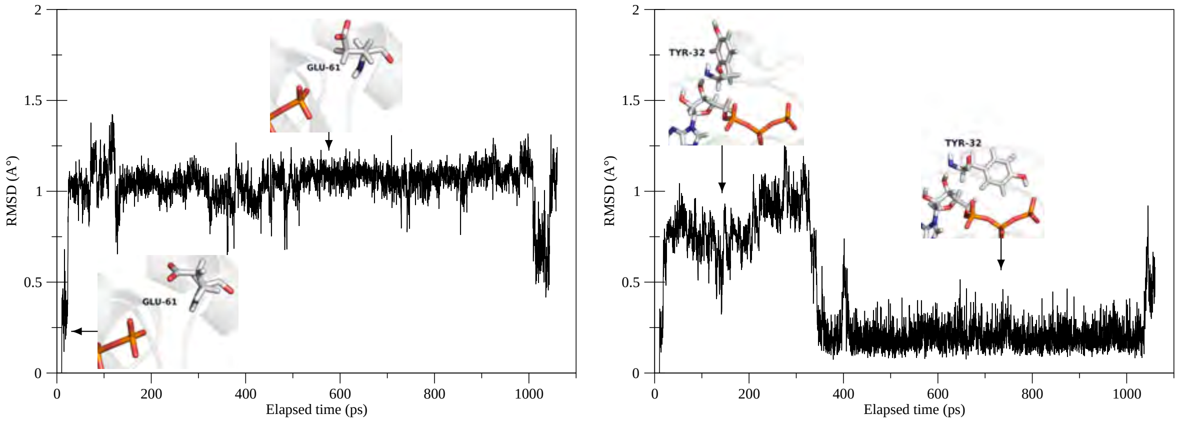
It appears that the substitution of Gln 61 by an Arg impacts directly the interaction network of Gly 60 residue. Indeed, within the wild-type, through its backbone nitrogen atom, Gly 60 mainly engages in hydrogen bonds with transient water molecules (see inset *(b)* of figure 3.10) whereas, within Q61R, hydrogen bonds are formed with GTP O_γ atoms. From a structural point of view, in these calculations that include GAP *arginine binding loop* only, Q61R seems to modify the docking between NRas and GAP, resulting in different interactions engaged between the two proteins and also in different interactions with solvent molecules that get in the active site.

Q61E Within Q61E, the RMSD(t) plot of Glu 61 (figure 3.11 *(a)*) presents slightly greater fluctuations ($\sim$ 0.2 − 0.3 Å) than those observed for Gln 61 ($\sim$ 0.1 − 0.2 Å) within the wild-type (figure 3.9 *(a)*). Glu 61 appears to be more mobile than wild-type Gln 61. The carboxylate group of this residue is able to point towards the γ phosphate group of GTP, a conformation also observed by Frech et al [20] during MD calculations. This state remains stable during the simulation as shown on figure 3.11 *(a)*.

As described above for wild-type NRas, Tyr 32 is found in both *closed* and *open* conformations. Within Q61E the *closed* conformation is the one that remains the longest within the simulation time, as shown on figure 3.11 *(b)*.

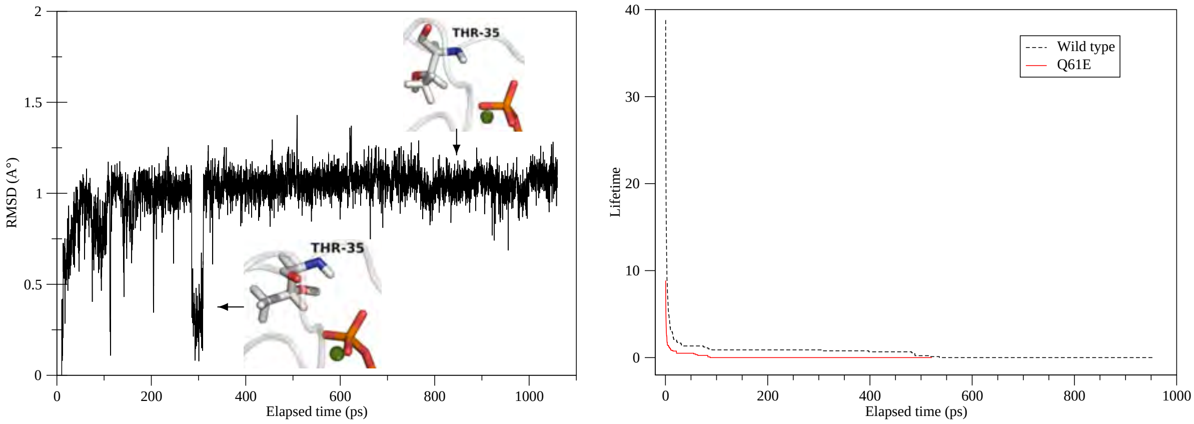
In Q61E NRas active site, Thr 35 is found in the first two conformations identified within the wild-type (see figure 3.11 *(c)*). Within this mutant, this residue mostly remains in the second one, *i.e.* the side chain hydroxy group pointing out of the protein active site and the methyl group pointing inwards.

Arg 789 residue is initially engaged in *native contacts* with GTP P_γ atom, as within the wild-type. This might be due firstly to the adopted procedure to obtain Q61E initial structure (for details report to appendix A). Nevertheless, these contacts are lost during the simulation and none of them remains after 500 ps (figure 3.11 *(d)*). Moreover, from the beginning, fewer *native contacts* are present compared to the wild-type because several disappear already during the previous QM/MM minimization cycles performed to remove eventual bad contacts in the initial structure (see simulation setup details in



(a) Glutamic acid 61

(b) Tyrosine 32



(c) Threonine 35

(d) Native contacts of GAP Arg 789 with GTP Pg atom

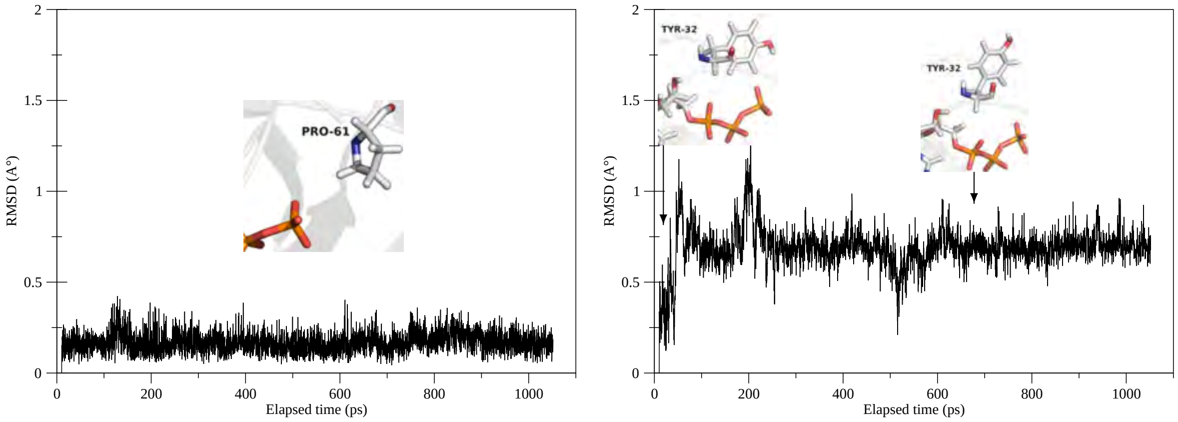
FIGURE 3.11: RMSD(t) plots alongside the associated conformational changes of Glu 61 (a), Tyr 32 (b), Thr 35 (c) from Q61E NRas and lifetime curve of *native contacts* between GTP P_γ atom and GAP Arg 789 (d) during QM/MM molecular dynamics.

appeandix A). *Non-native contacts* develop during the simulation, mainly with Tyr 32 residue.

Q61P In Q61P NRas, Pro 61 residue remains stable with small fluctuations around a unique conformation ($\sim 0.1 - 0.2$ Å) as shown on figure 3.12 *(a)*.

Tyr 32 residue switches from the initial *closed* conformation to the *open* one during the equilibration phase (see figure 3.12 *(b)*). It remains in this state throughout the simulation.

Thr 35 adopts the second and the third conformations identified within the wild-type. In the third one, the backbone oxygen atom does not follow the conformational change observed for the wild-type protein (figure 3.12 *(c)*).



(a) Proline 61 *(b) Tyrosine 32*

(c) Threonine 35 *(d) Native contacts of GAP Arg 789 with GTP Pg atom*

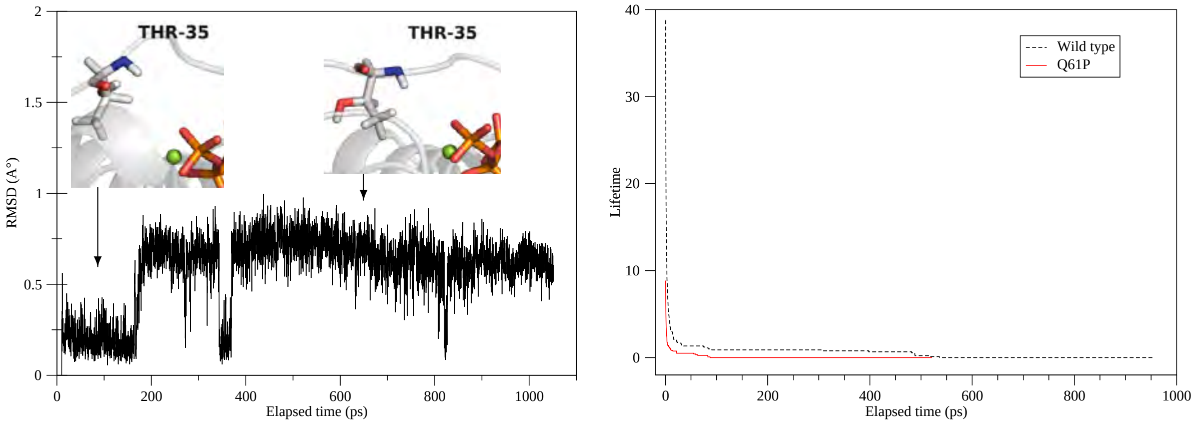


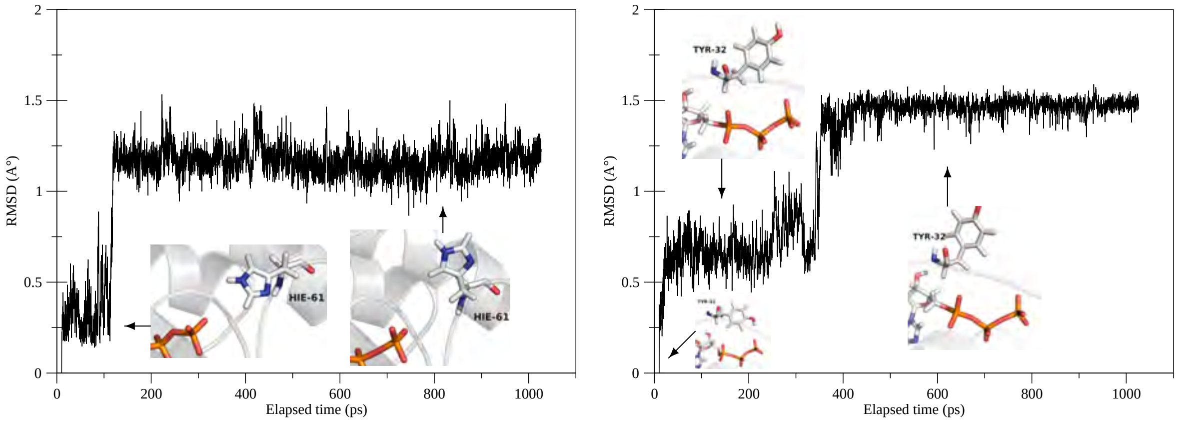
FIGURE 3.12: RMSD(t) plots alongside the associated conformational changes of Pro 61 (a), Tyr 32 (b), Thr 35 (c) from Q61P NRas and lifetime curve of *native contacts* between GTP P_γ atom and GAP Arg 789 (d) during QM/MM molecular dynamics.

Finally, the side chain of GAP Arg 789, initially engaged in *native contacts* with GTP P_γ atom, looses them all after 500 ps (figure 3.12 *(d)*). *Non-native contacts* develop mainly with Tyr 32 residue.

Q61H Two main stages are clearly identified on the RMSD(t) plot of His 61 residue (see figure 3.13 *(a)*). They are associated to two conformations of its side chain, *bent* and *unbent* toward the ligand, as depicted in the insets of this figure. The fluctuations around each conformation are very small, approximately $\sim 0.1 - 0.2$ Å.

Tyr 32 remains in the *open* conformation during the simulation, although it was *closed* in the initial structure, as shown on figure 3.13 *(b)*.

Thr 35 side chain adopts the second and third conformations identified within the


(a) Histidine 61

(b) Tyrosine 32

(c) Threonine 35

(d) Native contacts of GAP Arg 789 with GTP Pg atom

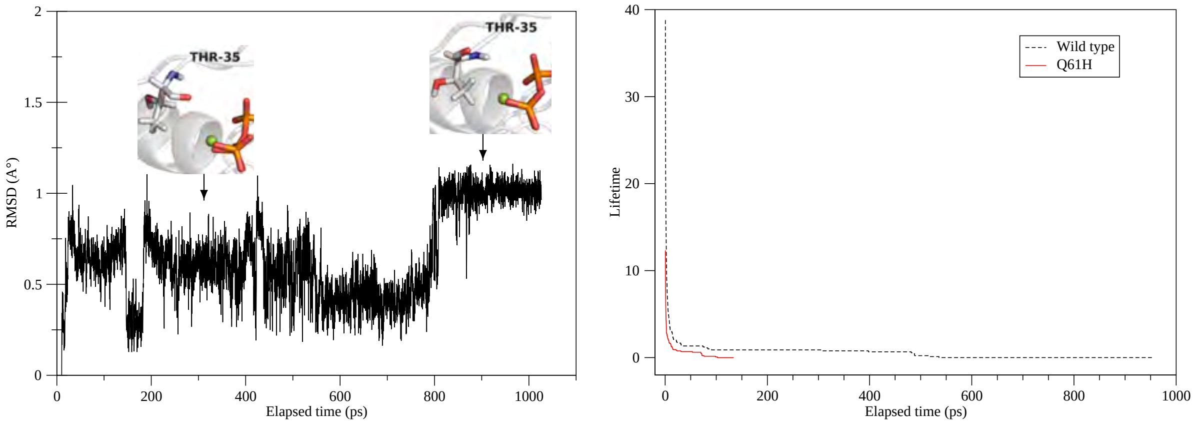
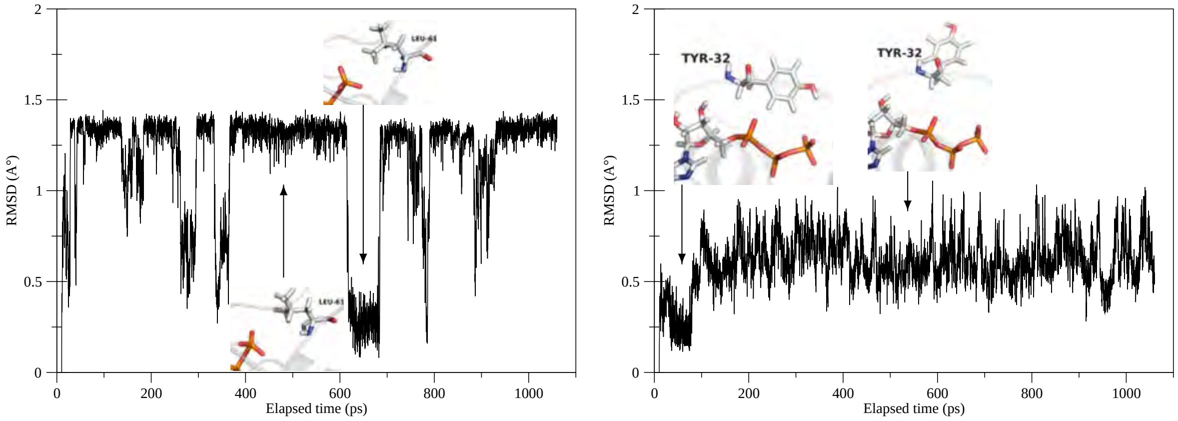


FIGURE 3.13: RMSD(t) plots alongside the associated conformational changes of His 61 (a), Tyr 32 (b), Thr 35 (c) from Q61H NRas and lifetime curve of *native contacts* between GTP P_γ atom and GAP Arg 789 (d) during QM/MM molecular dynamics.

wild-type although, in the third one, the backbone oxygen atom does not follow the conformational change observed within the wild-type (see figure 3.13 *(c)*). Within Q61H NRas, this residue mainly remains in the second conformation (*i.e.* the side chain hydroxy group pointing out of the protein active site and the methyl group pointing inwards).

When bound to Q61H mutant, the initial *native contacts* between Arg 789 and GTP P_γ atom are rapidly lost: none of them remains after 150 ps as shown in figure 3.13 *(d)*. As within Q61E and Q61P, *non-native contacts* appear with Tyr 32 residue.

Q61L The RMSD(t) plot of Leu 61 residue presents small fluctuations ($\sim 0.1 - 0.2$ Å) around two main conformations adopted during the simulation. They are shown in figure 3.14 *(a)*. The conformational change consists in a 90° rotation about the carbon 3 - carbon 4 single bond.


(a) Leucine 61

(b) Tyrosine 32

(c) Threonine 35

(d) Native contacts of GAP Arg 789 with GTP Pg atom

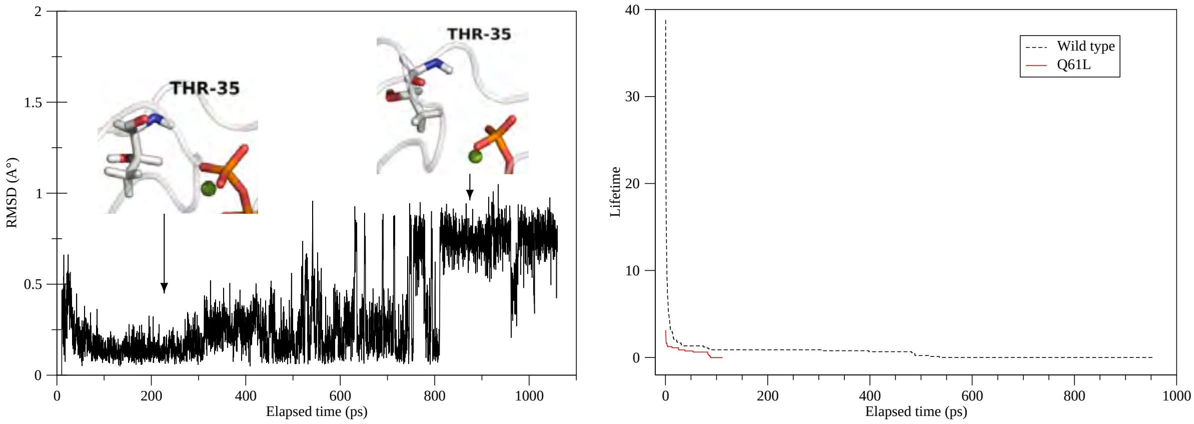
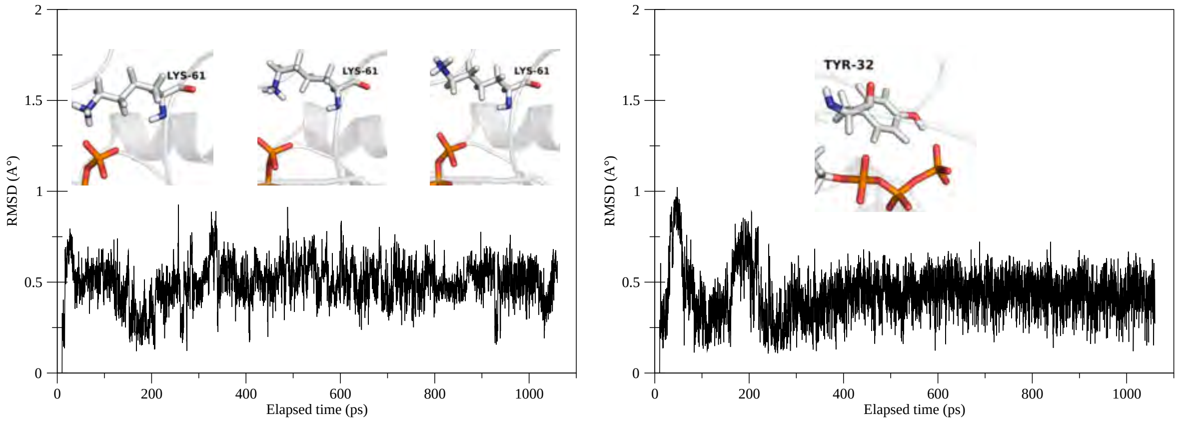


FIGURE 3.14: RMSD(t) plots alongside the associated conformational changes of Leu 61 (a), Tyr 32 (b), Thr 35 (c) from Q61L NRas and lifetime curve of *native contacts* between GTP P_γ atom and GAP Arg 789 (d) during QM/MM molecular dynamics.

During the simulation time, Tyr 32 side chain mainly remains in the *open* configuration as within wild-type NRas.

Thr 35 adopts the second conformation encountered within the wild-type (*i.e.* the side chain hydroxy group pointing out of the protein active site and the methyl group pointing inwards), as well as a new orientation such that both hydroxy and methyl groups point out of the active site (see figure 3.14 *(c)*). This last conformation is the longer lasting within the simulated time.

Finally, the side chain of GAP Arg 789 rapidly looses the few initial *native contacts* with GTP P_γ atom and *non-native contacts* appear with Tyr 32 residue (see figure 3.14 *(d)*).



(a) Lysine 61

(b) Tyrosine 32

(c) Threonine 35

(d) Native contacts of GAP Arg 789 with GTP Pg atom

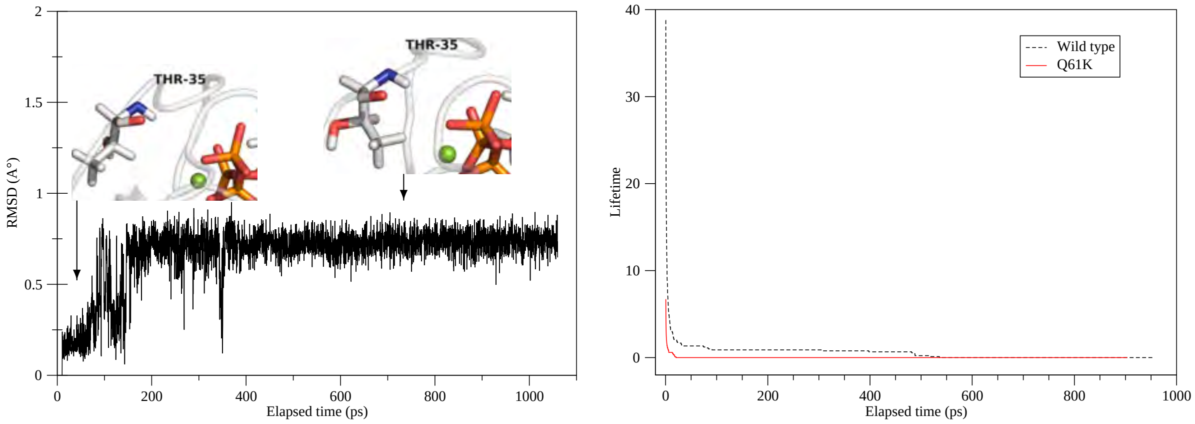


FIGURE 3.15: RMSD(t) plots alongside the associated conformational changes of Lys 61 (a), Tyr 32 (b), Thr 35 (c) from Q61K NRas and lifetime curve of *native contacts* between GTP P_γ atom and GAP Arg 789 (d) during QM/MM molecular dynamics.

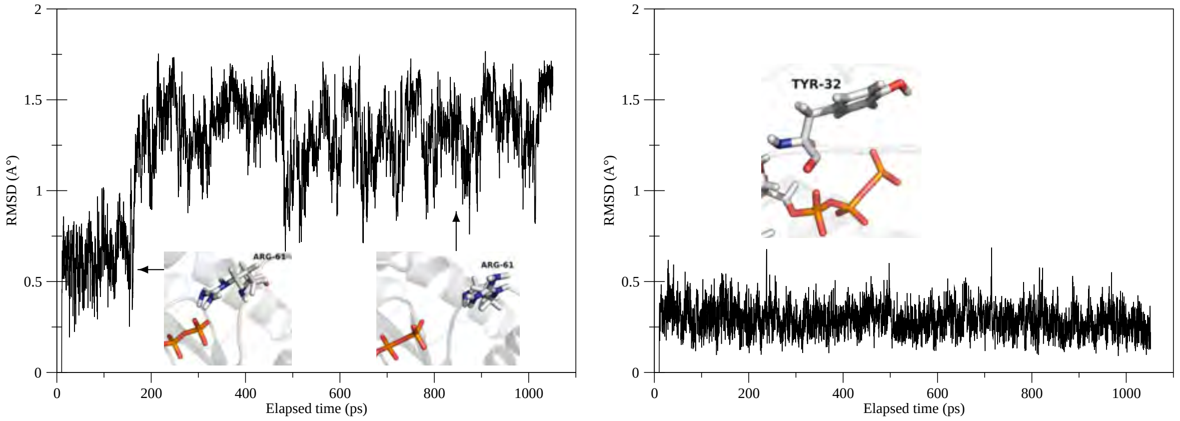
Q61K Within Q61K, the residue at position 61 appears to be more mobile than the corresponding one within wild-type NRas. Indeed, the comparison between figures 3.15 *(a)* and 3.9 *(a)* respectively, shows larger fluctuations in the RMSD(t) plot of Lys 61 ($\sim$ 0.3 – 0.4 Å) than those observed for wild-type Gln 61 ($\sim$ 0.1 – 0.2 Å). This is due to conformational changes undergone by its side chain that nevertheless allow the terminal amine group to form stable hydrogen bonds with GTP O_γ atoms during most of the simulation time. These conformations are depicted in the insets of figure 3.15 *(a)*.

Tyr 32 remains in the *closed* conformation (figure 3.15 *(b)*).

Two main stages are observed in Thr 35 RMSD(t) plot (see figure 3.15 *(c)*). The first one corresponds to the conformation encountered within Q61L mutant in which both the hydroxy and the methyl groups of the side chain point out of the active site. The second one, corresponds to the second conformation observed within the wild-type.

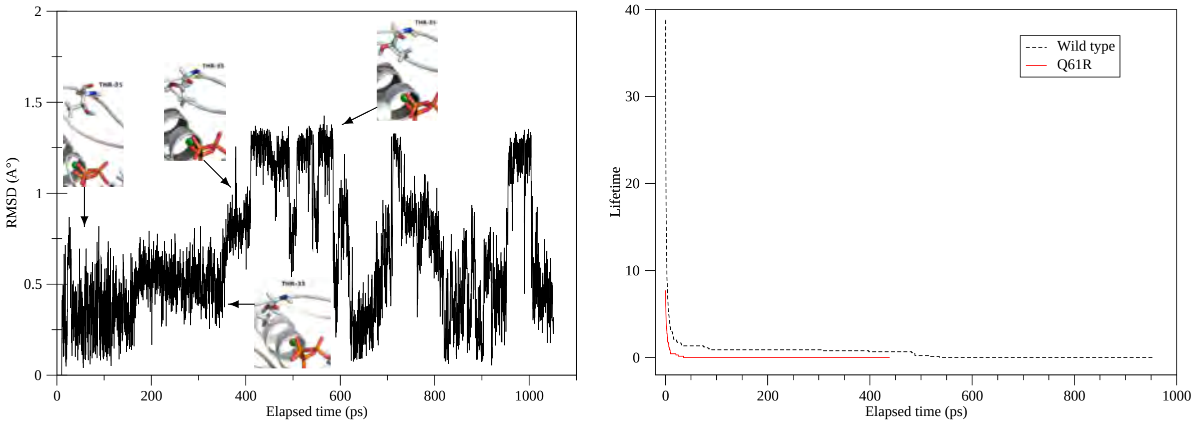
Both are depicted in the insets of figure 3.15 *(c)*.

GAP Arg 789, maintains the initial *native contacts* made with GTP P_γ atom almost as long as when bound to wild-type NRas (figure 3.15 *(d)*). *Non-native contacts* arise with Tyr 32 residue.



(a) Arginine 61

(b) Tyrosine 32



(c) Threonine 35

(d) Native contacts of GAP Arg 789 with GTP Pg atom

FIGURE 3.16: RMSD(t) plots alongside the associated conformational changes of Arg 61 (a), Tyr 32 (b), Thr 35 (c) from Q61R NRas and lifetime curve of *native contacts* between GTP P_γ atom and GAP Arg 789 (d) during QM/MM molecular dynamics.

Q61R The fluctuations observed in the RMSD(t) plot of Arg 61 residue (figure 3.16 *(a)*) are significantly greater ($\sim 0.5 - 0.6$ Å) than those of Gln 61 within the wild-type ($\sim 0.1 - 0.2$ Å). Moreover, Arg 61 is found in two main conformations depicted in the insets of figure 3.16 *(a)*. In the first one, the side chain guanidine engages hydrogen bonds with GTP O_γ atoms. In the second one, Arg 61 side chain is out of the active site. This *open* conformation lasts longer during the simulated time.

From the *closed* conformation in the initial structure, Tyr 32 switches to the *open* one during the first steps of the simulation, although the amplitude of the side chain motion is limited compared to the wild-type case (see figure 3.16 *(b)*).

Q61R appears to be the mutant in which Thr 35 is the least stable. Indeed, it continuously switches between the three main conformations identified within wild-type NRas and a fourth one identified within Q61L mutant (see figure 3.16 *(c)*).

As within other Gln 61 mutants, Arg 789, initially involved in native contacts with GTP P_γ atom, looses them all after 500 ps (figure 3.16*(d)*). *Non-native contacts* are subsequently formed with Glu 31 residue.

Intrinsic residues mobility conclusions

From the previous comparisons between wild-type and Gln 61 mutant NRas proteins, it appears that the substitution of Gln 61 has a structural impact mainly on residues 32 and 35 from NRas, on residue 60 in the case of Q61R only, and on residue 789 from GAP *arginine binding loop*. However, to assess the impact of Gln 61 mutants on NRas-GTP-GAP complex, calculations including the entire GAP protein are needed.

Depending on the mutant, the *open* or the *closed* conformations are mainly adopted by Tyr 32 residue during the simulation time.

Multiple conformations are adopted by Thr 35 residue within wild-type and mutant NRas but with different relative stabilities within each.

For Q61R only, the interactions network of Gly 60 residue is modified.

The *arginine finger* supplied by the portion of GAP considered during these unrestrained QM/MM simulations, rapidly looses the initially present *native contacts* with GTP P_γ atom within all Gln 61 substituted proteins, except Q61K. Moreover, it forms *non-native* contacts with residue 61 when bound to the wild-type only. It appears that, within the active site of the mutated proteins, the conformational changes undergone by GAP Arg 789 residue:

1. pull it appart from GTP γ atoms, reducing thereby the probability of interactions between the two,
2. move it towards the *Switch I*, provoking thereby interactions with residues from this region of the active site rather than with Gln 61 from *Switch II* as within the wild-type protein.

3.2.2 Active site water occupancy

The behaviour differences of key residues, described in the previous section, constitute structural modifications of the protein active site that occur upon Gln 61 substitution. These active site structural rearrangements are accompanied by a modification of solvent presence in it. Indeed, the visualization [28] of the obtained trajectories revealed that water molecules accessibility and remaining time in the protein active site vary whether the wild-type or a Gln 61 mutated form of NRas is considered. A first proof of this was briefly discussed in the previous section when the conformations adopted by Gly 60 residue revealed that a water molecule it interacts with within the wild-type is no longer present in that part of the active site within Q61R NRas.

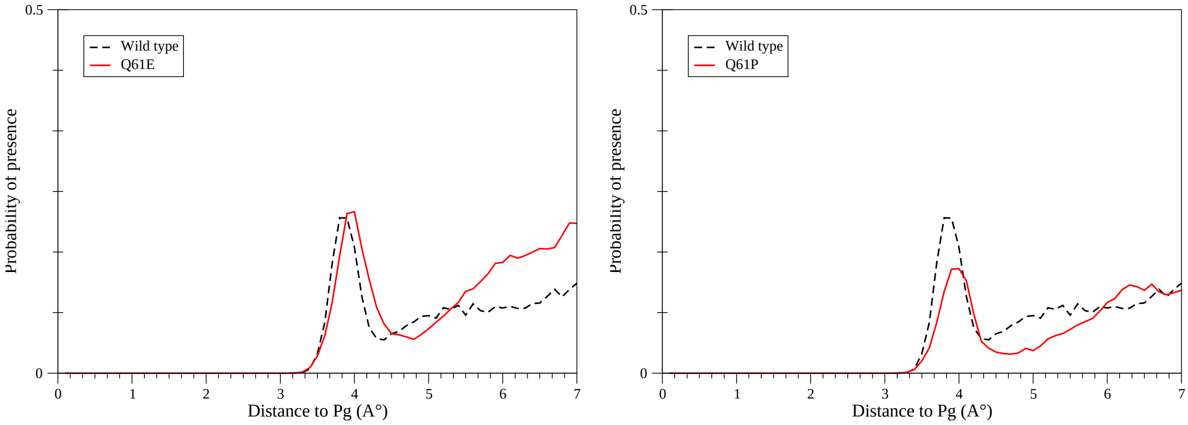
To assess qualitatively and quantitatively water occupancy differences between wild-type and mutant NRas, the radial distribution function (RDF) of water molecules around GTP phosphorus atoms was computed. Water distribution was then further mapped in 2D by projecting, on a chosen plane, water molecules found within 7Å of GTP P_γ atom.

Radial Distribution Function of water molecules

An algorithm aiming to determine the RDF of water molecules within NRas active site was developed. This choice was made with the purpose of freely selecting optimal parameters for the studied systems like the accuracy, but also in order to *i)* account for molecules treated with different energy models due to the hybrid QM/MM approach used, *ii)* discard specific molecules (such as the two water molecules coordinated to the Mg^{2+} ion) and *iii)* process all the trajectories in a single run. This custom designed RDF code was validated against the corresponding analysis tool provided by Visual Molecular Dynamics package [28] when using identical specifications for calculating this function.

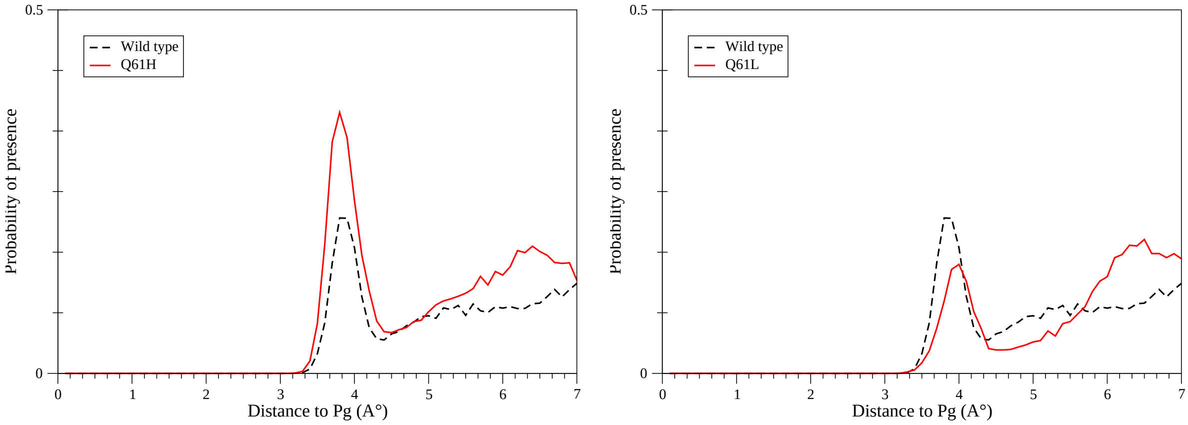
Figure 3.17 represents the RDF of water molecules around GTP P_γ atom for wild-type NRas (black dashed line) and each of the six studied Gln 61 mutants (red solid line).

For the wild-type, a first peak is found centered at ~ 3.8 Å from GTP P_γ atom. This peak was previously reported in a molecular dynamics study of wild-type Hras-GTP complex [47]. Surprisingly, the same peak is observed for the six studied mutant proteins. It is shifted to ~ 4 Å for Q61E, Q61P, Q61L and Q61R, and presents noticeable amplitude variations for Q61P, Q61H and Q61L (0.14, 0.36 and 0.15 respectively vs 0.21 for wild-type NRas).



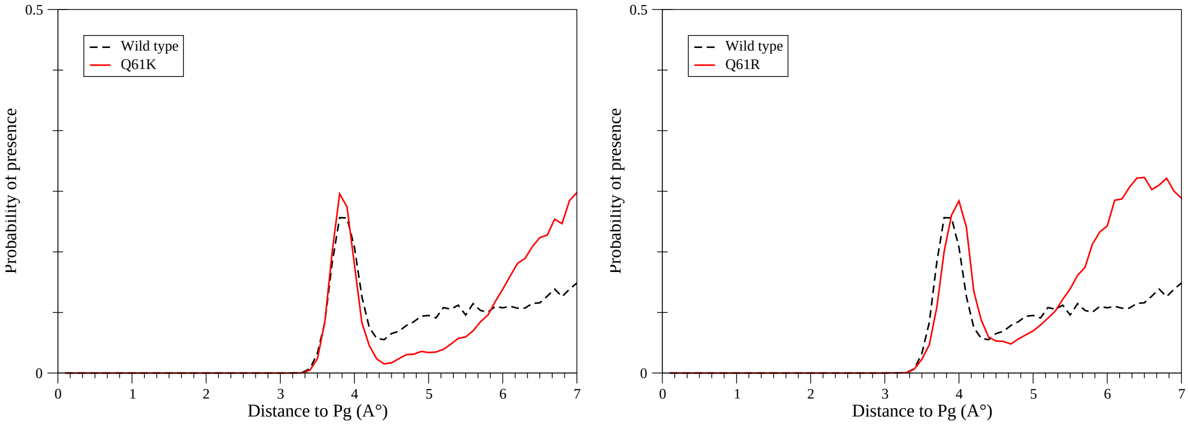
(a) WT vs Q61E

(b) WT vs Q61P



(c) WT vs Q61H

(d) WT vs Q61L



(e) WT vs Q61K

(f) WT vs Q61R

FIGURE 3.17: RDF of water molecules around GTP P_γ atom within wild-type NRas vs (a) Q61E, (b) Q61P, (c) Q61H, (d) Q61L, (e) Q61K and (f) Q61R mutants during QM/MM molecular dynamics simulations.

However, at $\sim 6-7$ Å from GTP P_γ atom, the differences in RDF between wild-type and Gln 61 mutant proteins is noticeable (figure 3.17). The probability to find a water molecule in this area is lower within the wild-type than within any of the Gln 61 mutated forms. Indeed, for wild-type NRas, the RDF curve fluctuates around a density value of 0.08 while it reaches values of 0.20, 0.11, 0.17, 0.16, 0.24 and 0.25 for Q61E, Q61P, Q61H, Q61L, Q61K and Q61R, respectively. Furthermore, the integral of the RDF curve in this region reveals the presence of 1.12 water molecules for the wild-type compared to 1.92, 1.24, 1.71, 1.80, 2.01 and 2.74 water molecules for Q61E, Q61P, Q61H, Q61L, Q61K and Q61R respectively.

In order to trace back where the solvent excess is located, that observed at $\sim 6-7$ Å from GTP P_γ atom within all Gln 61 mutant proteins, the RDF of water molecules around GTP P_α atom was also calculated. A striking difference appears between the wild-type and the mutated NRas forms, in particular for Q61R as depicted in figure 3.18 *(f)*. Indeed, a pick arises at ~ 4 Å from GTP P_α atom indicating that a large amount of water is found in a 4 Å radius of this atom in the active site of Q61R.

A similar peak is observed for Q61E, Q61P, Q61H and Q61L (figures 3.18 *(a)* to *(d)*). This peak is shifted to a distance of ~ 3.9Å from GTP P_α atom for Q61E and Q61P, ~ 3.7Å for Q61H and ~ 3.8Å for Q61L, with significantly smaller amplitude for these four mutants (*i.e.* 0.02, 0.10, 0.07, 0.06 for Q61E, Q61P, Q61H, Q61L respectively, compared to 0.27 for Q61R). As for wild-type NRas, this peak is absent on the RDF plot of Q61K. The integration of this curve, leads to 0.76 water molecules in a 7 Å radius of GTP P_α atom for Q61K and to 0.20 water molecules for wild-type NRas . This indicates that, although both RDF curves have similar shapes, there are more water molecules in the vicinity of GTP α phosphorus atom for Q61K than for wild-type NRas.

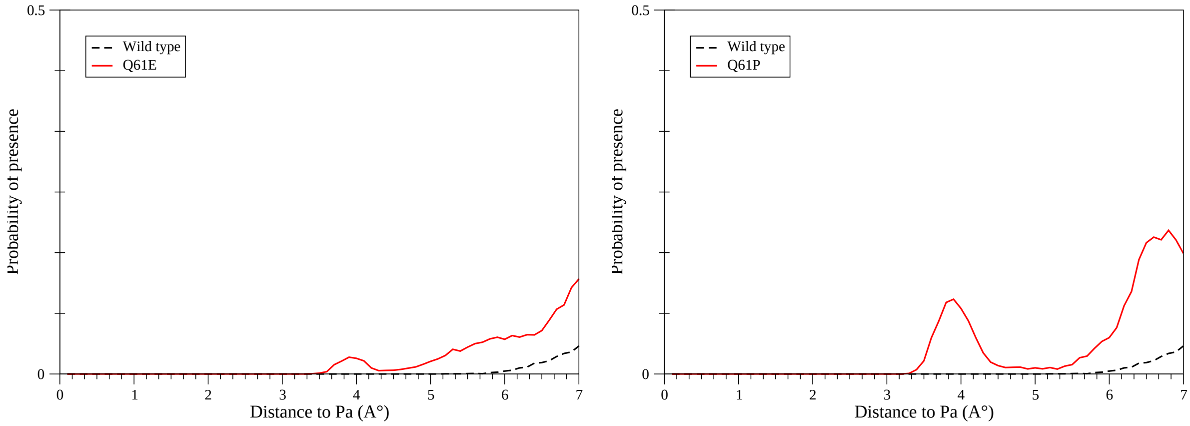
As water molecules are among the reactants in hydrolysis reactions, these results suggest that for catalyzing GTP hydrolysis, NRas active site is more efficient when its structure enables an accurate positioning of water molecules than when a higher solvent presence density is found in the vicinity of the bound nucleotide. This accurate water molecules positioning in NRas active site is precisely determined by computing the 2D water probability density.

2D RDF of water molecules

An algorithm was developed in order to map and visualize the regions of the active site where water molecules tend to remain during the unrestrained QM/MM MD simulations. This algorithm counts water molecules within a given radius of a given atom

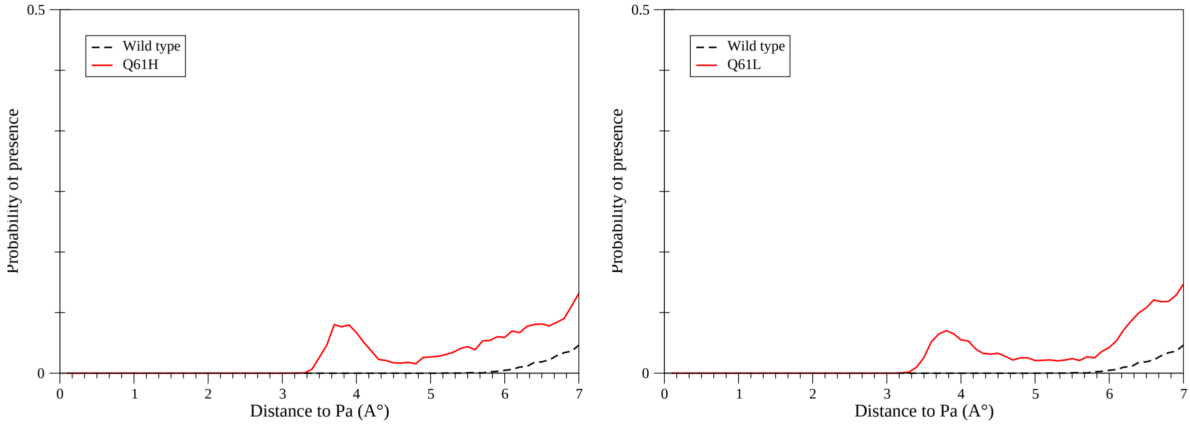
(a) WT vs Q61E

(b) WT vs Q61P



(c) WT vs Q61H

(d) WT vs Q61L



(e) WT vs Q61K

(f) WT vs Q61R

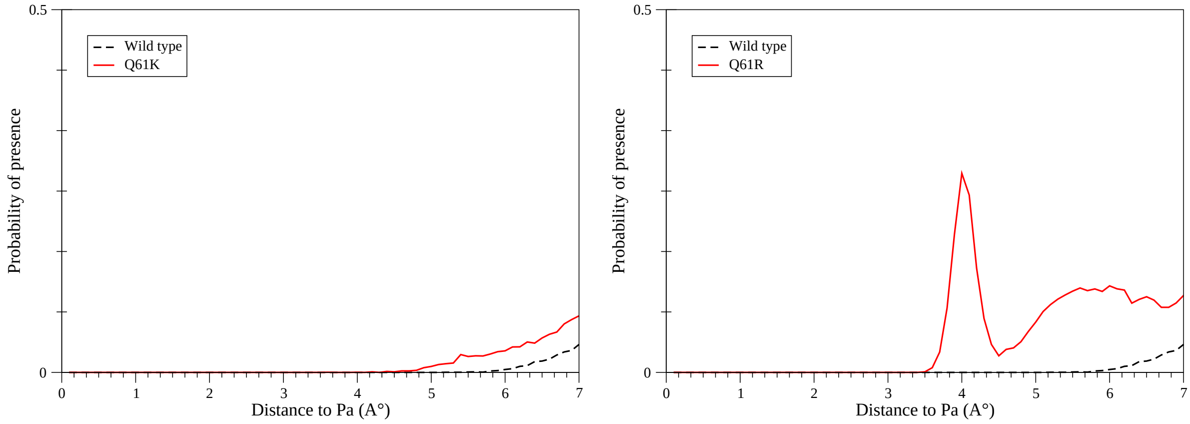


FIGURE 3.18: RDF of water molecules around GTP P_α atom within wild-type NRas vs (a) Q61E, (b) Q61P, (c) Q61H, (d) Q61L, (e) Q61K and (f) Q61R mutants during QM/MM molecular dynamics simulations.

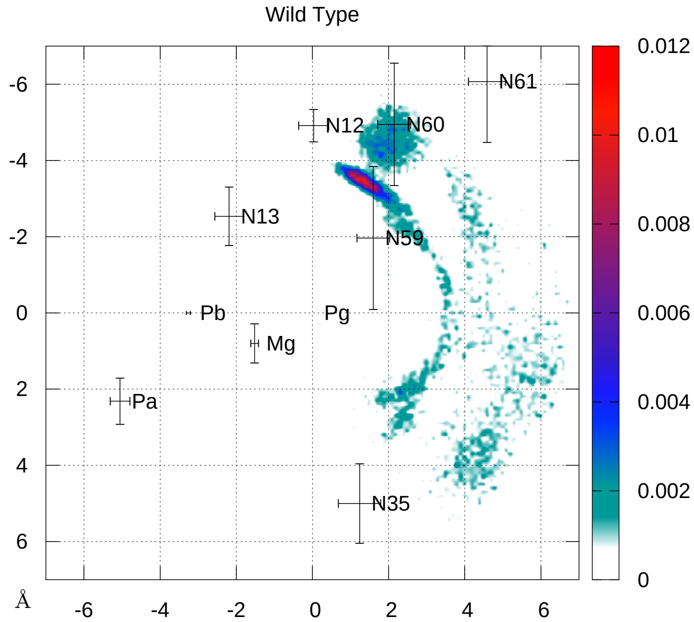


FIGURE 3.19: 2D RDF of water molecules within the active site of wild-type NRas during QM/MM molecular dynamics. The average coordinates of nitrogen backbone atoms from residues 12, 13, 35, 59, 60 and 61 have also been projected as well as GTP phosphorus atoms and the Mg^{2+} ion. The corresponding standard deviation values are represented with error bars. GTP P_γ atom is placed at the center of the plane.

and projects them on a chosen plane. Its detailed description is provided in appendix B.

For studying wild-type and Gln 61 mutant NRas, a plane containing GTP P_β and P_γ atoms, was considered to be the most appropriated *projection* plane. Water molecules found within 7 Å of GTP P_γ atom were projected on it. The Mg^{2+} ion, GTP phosphorus atoms and nitrogen backbone atoms of several NRas residues were also projected in order to see how water molecules accommodate within the active site vis-a-vis of these atoms. Nitrogen backbone atoms were selected because of their hydrogen bond donor/acceptor competence that makes them critical in the establishment of a hydrogen bond network within the protein active site. Results for wild-type NRas and Gln 61 mutants are presented in figures 3.19 and 3.20 respectively.

Wild-type p21^{N-ras} For wild-type NRas, water molecules stay close to GTP P_γ atom, forming an arch of density around the γ phosphate group, extended from residues 12 to 35 (see figure 3.19). The highest water density is localized in a small region between residues 12, 59 and 60. Indeed, although water molecules are unconstrained during the whole QM/MM simulation, they naturally tend to stay in this area because of stable hydrogen bonds formed with the nitrogen backbone atoms of Gly12 and Gly60 residues:

$$O_{H_2O} \text{ (donor) - - - } N \text{ atom of Gly12 (acceptor)}$$
$$N \text{ atom of Gly60 (donor) - - - } O_{H_2O} \text{ (acceptor)}$$

A second region of high water density is found in the vicinity of residue Thr35[1], although the associated amplitude is significantly smaller. This region corresponds to the area of the protein active site where earlier studies of wild-type Ras [20, 50, 55, 62, 63] identified a presumed lytic water molecule. Finally, no water molecules are found in the vicinity of GTP P_α nor P_β atoms.

Gln 61 mutations RDF around GTP P_γ atom of wild-type NRas and Q61E are similar, with an excess of solvent found at $\sim 6-7$ Å from this atom for Q61E (figure 3.17 *(a)*). However, the 2D RDF plotted for this mutant (figure 3.20 *(a)*), shows that water molecules are delocalized within its active site compared to the wild-type. Three water density peaks arise between residues 35 and 59, next to residue 60 and between residues 12 and 61. The amplitude of each of these three peaks is one order of magnitude smaller than for the highest peak found within wild-type NRas. Water density is higher in other regions of the active site, in particular in the vicinity of GTP P_β atom.

According to RDF plots, Q61P and Q61L mutations lead to a lower probability of finding water molecules at ~ 4 Å from GTP P_γ atom than what is observed for the wild-type (figure 3.17 *(b)* and *(d)*). By contrast, a peak of density arises at the same distance from P_α atom (figure 3.18 *(b)* and *(d)*).

In the 2D RDF of Q61P, the arch of water molecules observed for the wild-type can also be observed although it is scattered compared to the non mutated protein (see figure 3.20 *(b)*). The peak found between residues 12 and 60 is now shifted toward residue 59. Another peak of water density arises close to GTP P_γ atom, following the $\overrightarrow{P_\beta P_\gamma}$ direction. Both present similar amplitudes, one order of magnitude smaller than the peak held by residues 12 and 60 within wild-type NRas and, again, water density is higher in other parts of the active site, in particular in the vicinity of the Mg^{2+} ion.

For Q61L mutant (figure 3.20 *(d)*), water positioning in the active site is delocalized compared to that observed for wild-type NRas. The peak of water density held by residues 12 and 60 within the wild-type, does not appear anymore but two others are observed between residues 59 and 35. Their amplitude is one order of magnitude smaller than the highest density peak within the non mutated form. Water density is

---

[1] only its nitrogen backbone atom N35 is represented

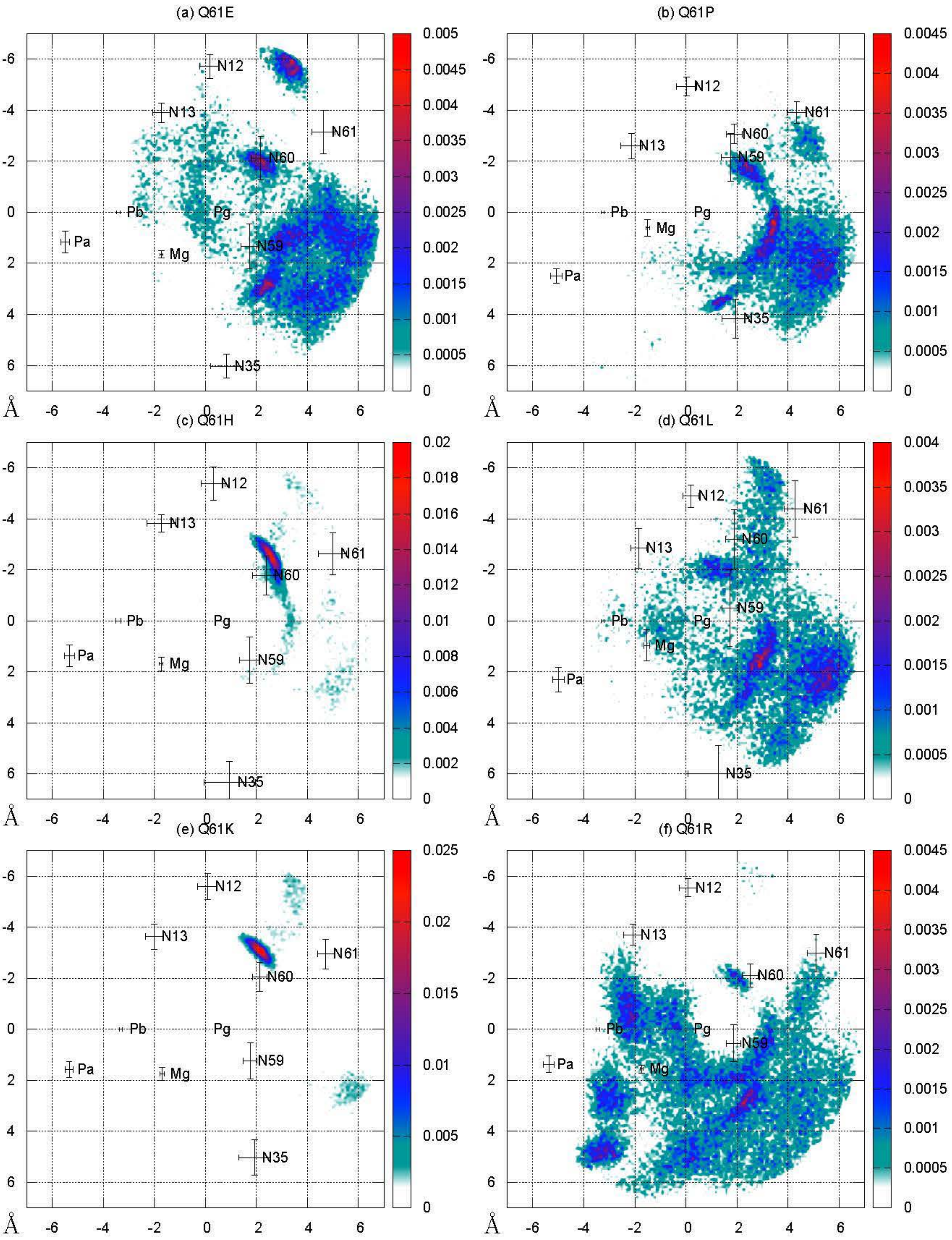


FIGURE 3.20: 2D RDF of water molecules within the active site of (a) Q61E, (b) Q61P, (c) Q61H, (d) Q61L, (e) Q61K and (f) Q61R NRas mutants during QM/MM molecular dynamics. The average coordinates of nitrogen backbone atoms from residues 12, 13, 35, 59, 60 and 61 have also been projected as well as GTP phosphorus atoms and Mg^{2+} ion. The corresponding standard deviation values are represented with error bars.

higher in other regions of the active site, in particular in the vicinity of residue 13 and GTP P_β atom.

For mutations Q61H and Q61K (figure 3.20 *(c)* and *(e)*), water distribution is more localized than within the active site of wild-type NRas.

Indeed, within Q61H (figure 3.20 *(c)*), the peak of density held by residues 12 and 60 presents a higher amplitude than within the wild-type, alike the RDF plot around P_γ atom (figure 3.17 *(c)*). Moreover, the arch of water molecules together with the second highest density peak, placed in the vicinity of residue Thr 35, disappear.

Mutation Q61K also results in a higher amplitude for the density peak held by residues 12 and 60 and in the disappearance of the peak previously found next to residue 35 (figure 3.20 *(e)*).

Q61R is the mutation that leads to the most delocalized water distribution in the protein active site (figure 3.20 *(f)*). The peak held by residues 12 and 60 is not observed anymore but, according to the RDF plots, the probability of finding water molecules at 4 Å from P_γ atom is similar to that of wild-type NRas (figure 3.17 *(f)*). This is due to water molecules placed all around the γ phosphate group (figure 3.20 *(f)*). The second peak of water density, found close to Thr 35, has a higher amplitude within Q61R than within the wild-type. High densities of water molecules are observed around α and β phosphate groups.

Water distribution conclusions

Comparing water distribution within the active site of wild-type and Gln 61 mutant NRas, leads to the hypothesis that, for catalyzing GTP hydrolysis, the protein is more efficient when its structure enables an accurate positioning of water molecules rather than a larger amount of solvent present in the vicinity of the nucleotide. Indeed, according to the RDF plots, there are less water molecules in the vicinity of GTP P_α atom within the wild-type than within any of the Gln 61 mutated forms. The same is true when observing the RDF around GTP P_γ atom at $\sim 6-7$ Å from this atom. Moreover, the 2D RDF figures clearly show an accurate positioning within the wild-type protein *i.e.* the arch of density, which is lost within the mutants either because of *water delocalization* (*i.e.* Q61E, Q61P, Q61L, Q61R) or because of *water over-localization* (*i.e.* Q61H, Q61K).

In regards to *water delocalization* that leads to high amounts of solvent present in the vicinity of α and β phosphate groups within most of oncogenic mutant proteins, water molecules in this region of the active site might disturb the electronic density of the ligand, making the $P_\gamma - O_{\beta\gamma}$ bond breakage more difficult to be achieved. This hypothesis is investigated in the next section.

As for *water over-localization*, although within the wild-type the density peak observed next to residue 35 presents a negligible amplitude compared to the amplitude of the peak held by residues 12 and 60 (figure 3.19), water presence in both regions of the active site appears to be crucial for catalyzing GTP hydrolysis. Indeed, even if the peak maintained by residues 12 and 60 has a higher amplitude within Q61H and Q61K than within the wild-type, a single water density peak is insufficient for efficiently catalyzing the reaction.

Overall, it appears that the water molecule found next to GTP P_γ atom, held by hydrogen bonds formed with residues 12 and 60, has an important role in the enzymatic catalysis of the reaction. Nevertheless, its role remains to be elucidated. Among several possibilities, it could consist in *i)* reacting with the ligand, *ii)* participating in individual/concerted proton transfers or *iii)* structuring the protein active site without being chemically involved in hydrolysis. The unrestrained QM/MM dynamics show however that residues 12 and 60 ensure the positioning of this water molecule by hydrogen bonds and, thereby, participate in catalyzing the reaction. Nevertheless, water presence in the vicinity of these two residues only is not sufficient for catalyzing GTP hydrolysis. This suggests that a second water molecule participates in the enzymatic reaction, in accordance with previous studies [23, 43, 53, 63, 70]. It does not imply that both water molecules participate chemically in GTP hydrolysis. Their precise role in the reaction catalyzed by Ras remains to be determined.

Regardless of the numerous reaction schemes that have been proposed for GTP hydrolysis within wild-type NRas, involving one or more water molecules, it is known that enzymes typically limit conformational flexibility of the bound ligand, generally constraining it to a conformation close to that of the reaction transition state. Hence, evaluating the ligand intrinsic properties within the wild-type and mutant proteins active site appears as a the most convenient study to proceed with.

3.3 GTP properties within NRas active site

Among other attributed catalyzing effects, it has been experimentally shown [14, 32] that GTP binding to Ras active site induces the elongation of the bond to be broken (between P_γ and $O_{\beta\gamma}$ atoms). This is expected to lower the hydrolysis reaction energy barrier by several kcal/mol compared to the same reaction in solution. GTP geometric details, *i.e.* bond lengths and angles, are first investigated within the active site of wild-type and Gln 61 mutant NRas.

3.3.1 GTP bond lengths and angles

Table 3.1 summarizes a few geometric characteristics of the bound ligand. There are no significant differences in bond lengths between wild-type and Gln 61 mutant NRas according to QM/MM molecular dynamics simulations. Indeed, the $P_\gamma - O_{\beta\gamma}$ bond, the one to be broken during hydrolysis, differs by 0.01 to 0.03 Å between wild-type and mutant NRas. About the same differences are observed in the length of $P_\beta - O_{\beta\gamma}$ bond. Moreover, these bonds appear to be equally constrained by the considered protein active sites as they present the same standard deviations from their average values (*i.e.* 0.04 Å).

	Wild-type	Q61E	Q61P	Q61H	Q61L	Q61K	Q61R
$P_\gamma - O_{\beta\gamma}$	1.76 ± 0.04	1.78 ± 0.04	1.77 ± 0.04	1.78 ± 0.04	1.76 ± 0.04	1.75 ± 0.04	1.78 ± 0.04
$P_\beta - O_{\beta\gamma}$	1.66 ± 0.04	1.70 ± 0.04	1.65 ± 0.04	1.69 ± 0.04	1.65 ± 0.04	1.66 ± 0.04	1.70 ± 0.04
$P_\beta \widehat{O_{\beta\gamma}} P_\gamma$	146.6 ± 6.0	158.4 ± 9.0	145.4 ± 5.7	163.5 ± 11.6	145.2 ± 6.0	152.1 ± 7.0	168.0 ± 8.5

TABLE 3.1: GTP remarkable bond lengths (in Å) and angles (in°) when bound to wild-type and six oncogenic mutant NRas proteins.

More noticeable differences are observed in GTP $P_\beta \widehat{O_{\beta\gamma}} P_\gamma$ angle. Although similar values are adopted within wild-type, Q61P and Q61L NRas, a difference of up to 22° is observed with Q61R (12, 17 and 6° with Q61E, Q61H and Q61K respectively).

Hence, the directions in which GTP P_γ atom can move appear to be different within wild-type and some mutant NRas. We proceed hence to determine these directions through the use of the static modes approach.

3.3.2 P_γ intrinsic directional motions

The obtention of wild-type and mutant NRas structures defining the conformations in which GTP P_γ atom intrinsic directional motions are evaluated, is described in appendix A.

In order to perform this mechanistic analysis of the bound nucleotide, the static modes associated to each protein complex were calculated. The information they carry was subsequently harnessed through an algorithm that performs linear combinations of the static modes in order to satisfy a given constraining criterium. The criterium selected here was the maximization of the distance between GTP P_γ and $O_{\beta\gamma}$ atoms as the bond between them is broken during hydrolysis, thus an increase of the distance separating them is expected.

While determining the force directions that maximize $P_\gamma - O_{\beta\gamma}$ bond length, the γ phosphorus atom was monitored to determine the amplitude and direction of its intrinsic motions during $P_\gamma - O_{\beta\gamma}$ opening. The corresponding results, for wild-type and Gln 61 mutant NRas, are presented in figure 3.21. In order to adopt the same viewpoint for observing GTP P_γ atom intrinsic motions, Gly 12 and Gly 13 residues conformation as well as secondary structure elements were taken as a reference. Indeed, these residues belong to the rigid *P-loop* region and overall, secondary structure elements are conserved upon mutation (according to Kraulis et al. NMR spectroscopy study [36] and Krengel et al. X-ray crystallography results [37], respectively).

For wild-type NRas, the largest amplitude motion is allowed in a direction towards the side of the active site were the *Switch I* region is located. More specifically, this arrow points to Asp 33 residue. The same is true for Q61P although the highest amplitude arrow points towards Pro 34 residue. Within this mutant, GTP $P_\beta\widehat{O_{\beta\gamma}}P_\gamma$ angle presented the slightest deviation from the value reported for the wild-type (see table 3.1).

For Q61L, which also presents a $P_\beta\widehat{O_{\beta\gamma}}P_\gamma$ angle value close to that of the wild-type, P_γ most favorable motion appears to be in the direction of Asp 57 residue. It is towards Asp 33 residue within Q61H, as within the wild-type protein, despite GTP $P_\beta\widehat{O_{\beta\gamma}}P_\gamma$ angle having significantly different values within both ($\sim$ 163° vs 146° for Q61H and wild-type NRas, respectively according to table 3.1).

Within Q61K and Q61R, the most widespread NRas Gln 61 substitutions, GTP P_γ highest amplitude directions of motion are the most dissimilar to the one observed for the wild-type protein. Within Q61K, this motion is directed towards the *Switch II*, more

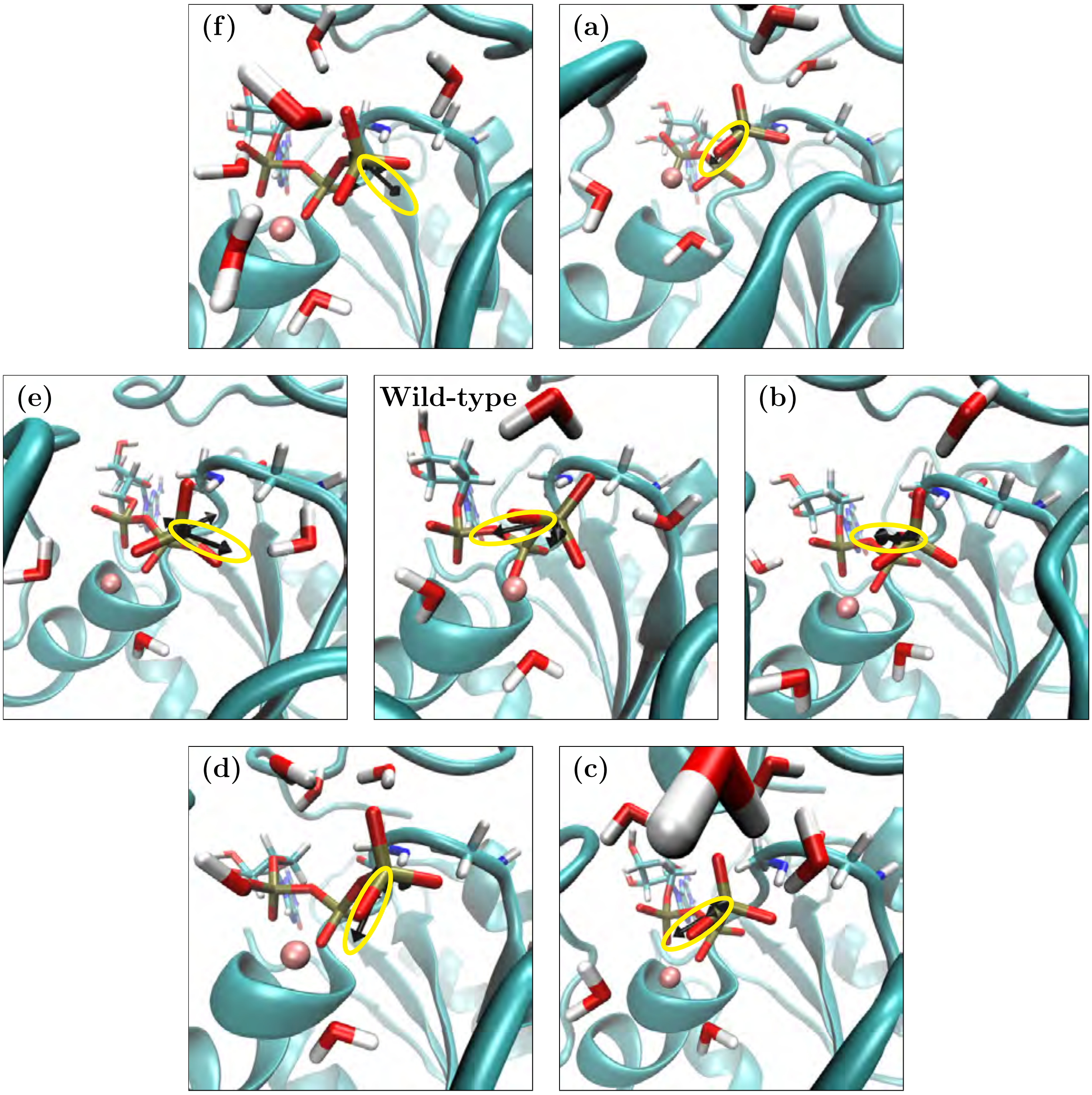


FIGURE 3.21: Intrinsic directional motions of GTP P_γ atom within the active site of wild-type NRas (central panel) vs Q61E (a), Q61P (b), Q61H (c), Q61L (d), Q61K (e) and Q61R (f) mutants. The amplitude of the black arrows indicate the extent of the allowed motion in the considered direction.

specifically towards Gly 60 residue. Within the active site of Q61R, the largest amplitude motion points towards Lys 16 residue.

Beyond revealing the directionality of GTP P_γ atom intrinsic motions, depicting the ligand within the active sites of wild-type and Gln 61 mutant NRas, as done in figure 3.21, discloses the fact the the γ phosphate group is differently oriented. Indeed, oxygen atoms belonging to this group eclipse to a different extend oxygen atoms from the β phosphate group within the various considered active site proteins. For the wild-type protein, an eclipsed conformations between GTP phosphate groups was reported [59]. Nevertheless, the static analysis presented here does not allow to draw conclusions on preferably adopted conformations as it is tied to a single active site configuration of the protein complexes.

In a previous theoretical study of the role of the Mg^{2+} ion [60], Rudack et al. reported that a change in the cation coordination sphere induced changes in GTP charge distribution. Similarly, a change of the number of water molecules surrounding both the Mg^{2+} ion and the nucleotide, observed upon Gln 61 substitutions, could influence their charge distribution. This possibility is investigated in the next section.

3.3.3 GTP electronic density

The nucleotide electronic distribution is first evaluated through a Löwdin population analysis. Subsequently, it is further characterized by depicting the Kohn-Sham Highest Occupied Molecular Orbital (KS HOMO) associated to the ensemble of atoms taken into account for performing these charge distribution analysis. The ensemble of atoms consists in the ligand, the cation and water molecules found within 5Å of GTP P_γ atom. They are treated at the DFT level of theory while the protein and bulk solvent are treated at the MM level. The simulations setup, as well as the procedure followed to obtain the structures used during these calculations, can be found in appendix A.

Löwdin population analysis

The reduced atomic charges attributed to the DFT atoms through a Löwdin population analysis are only examined for comparing the different environments provided by the protein active sites under consideration. In this manner, the relative changes in GTP and Mg^{2+} ion charge distribution, induced by the structural rearrangements of NRas active site along with water molecules positioning, upon Gln 61 substitutions, are evaluated.

The results of this analysis are presented in figure 3.22 for wild-type (black rectangles) and Gln 61 mutant NRas proteins (red rectangles).

Some common features arise for the six studied mutant proteins, namely: *i)* the Mg^{2+} ion is more negatively charged within the wild-type than within any of the mutant proteins, in particular within Q61P, the reduced Löwdin charge attributed to the cation is positive signalling a depletion of electrons, *ii)* GTP P_γ atom is more positively charged within wild-type NRas than within any of the mutant proteins and *iii)* $O_{3\gamma}$ and $O_{1\beta}$ atoms hold less negative charge within the wild-type than within any of the mutated forms (even if the charge difference on $O_{1\beta}$ is small i.e. $\sim$ 0.005 a.u. for Q61R).

For the rest of the nucleotide atoms, specific mutations lead to specific differences.

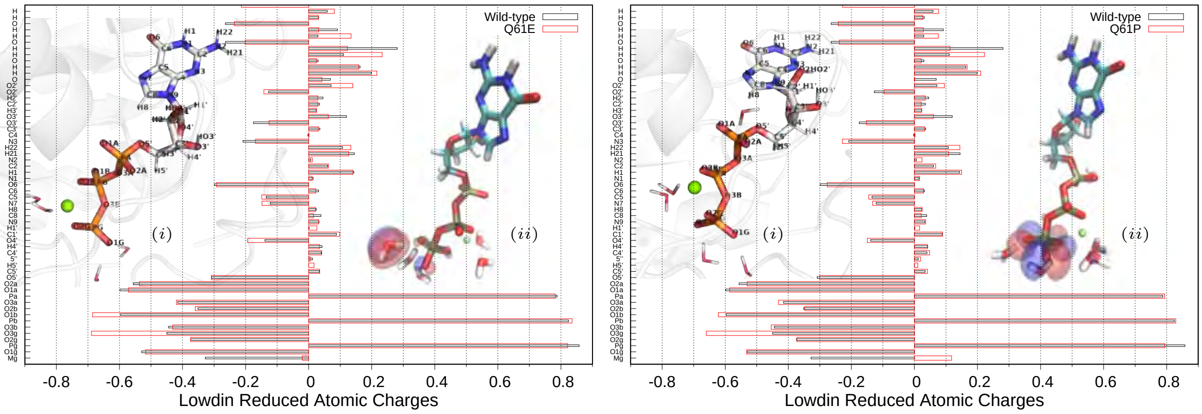
$O_{1\gamma}$ and $O_{2\gamma}$ atoms hold the same negative Löwdin reduced charge within wild-type, Q61E and Q61P mutant NRas. They are less negatively charged within Q61H and Q61L. Within Q61K and Q61R, $O_{1\gamma}$ is more negatively charged while $O_{2\gamma}$ less negatively charged than within the wild-type.

The bridging oxygen atom between the γ and β phosphate groups holds a slightly larger negative charge ($\sim$ 0.015 a.u.) within Q61P, Q61H, Q61L and Q61K mutants than within wild-type NRas. It is slightly less negatively charged ($\sim$ 0.01 a.u.) within Q61E and Q61R than within the non-mutated protein.

The Löwdin reduced charge held by P_β, $O_{2\beta}$ and $O_{\alpha\beta}$ atoms is nearly the same within the wild-type and mutated proteins active site (differences of $\sim \pm 0.01$ a.u.).

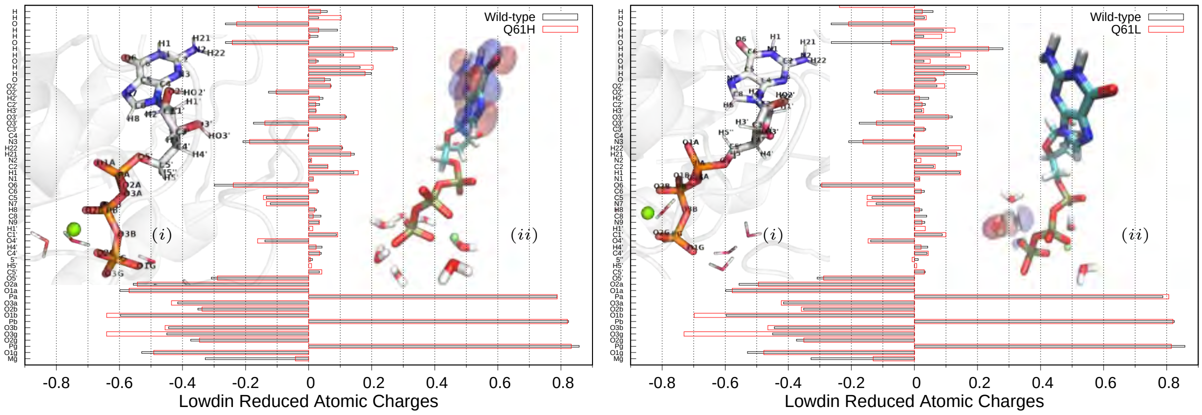
Finally, the α phosphorus atom also holds approximately the same positive reduced charge in the different environments provided. $O_{1\alpha}$ and $O_{2\alpha}$ atoms are less negatively charged within Q61E, Q61P, Q61H, Q61L and Q61R than within the wild-type. Within Q61K, $O_{1\alpha}$ is also less negatively charged while $O_{2\alpha}$ holds 0.047 more negative charge than wild-type NRas.

The overall reduced Löwdin charges attributed to GTP α, β and γ phosphate groups and to the Mg^{2+} ion, bound to the different NRas forms, are summarized in table 3.2. From this table, it can be easily seen that the wild-type active site arrangement, along with the water molecules included in it, is the only structural configuration that leads to a bound ligand with a more negatively charged β phosphate group than its γ phosphate group is. This is in accordance with numerous experimental [3, 4, 14, 18, 34, 35, 77] and theoretical [32, 59, 60] investigations of wild-type Ras proteins that concluded that this charge shift from the γ to the β phosphate group, induced by the protein, constitutes a major catalyzing effect.



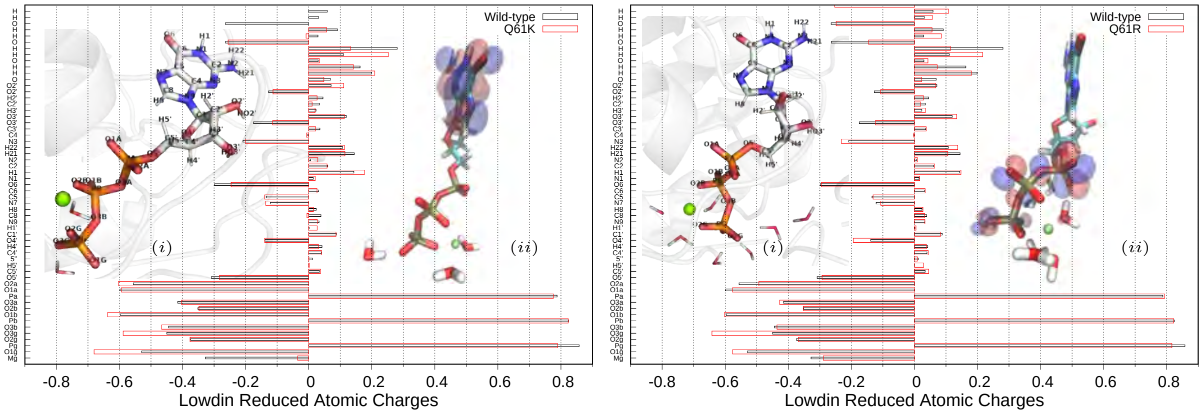
(a) WT vs Q61E

(b) WT vs Q61P



(c) WT vs Q61H

(d) WT vs Q61L



(e) WT vs Q61K

(f) WT vs Q61R

FIGURE 3.22: Löwdin population analysis of GTP, Mg^{2+} and water molecules within wild-type NRas vs Q61E (a), Q61P (b), Q61H (c), Q61L (d), Q61K (e) and Q61R (f). In inset *(ii)* of each panel the KS HOMOs of the DFT system is depicted. The isovalue selected for representing the KS HOMO is $|0.02|e^{-}/a_0^3$. The positive lobes of the KS HOMOs are depicted in blue and the negative ones in red.

	Wild-type	Q61E	Q61P	Q61H	Q61L	Q61K	Q61R
$\gamma\ PO_4^{2-}$	-0.495546	-0.758753	-0.771163	-0.645419	-0.744287	-0.855361	-0.770763
$\beta\ PO_4^{-}$	-0.568316	-0.64049	-0.594811	-0.613388	-0.702032	-0.628504	-0.568822
$\alpha\ PO_4^{-}$	-0.781513	-0.745417	-0.73535	-0.759775	-0.687824	-0.824228	-0.704198
Mg^{2+}	-0.327834	-0.020431	0.117928	-0.042110	-0.130809	-0.035087	-0.289204

TABLE 3.2: Löwdin reduced charges in a.u. for wild-type and Gln 61 mutant NRas.

Furthermore, for five of the six studied Gln 61 mutant proteins (i.e. Q61E, Q61P, Q61L, Q61K and Q61R), not only the γ phosphate group is more negatively charged than the β phosphate group, but also it holds more negative charge than the α phosphate group. Hence, for these five NRas mutants, GTP phosphate groups charge follows the trend:

$$\gamma_{PO_4^{2-}} \text{ electronic density} > \alpha_{PO_4^{-}} \text{ electronic densitty} > \beta_{PO_4^{-}} \text{ electronic density}$$

while for the wild-type it is:

$$\alpha_{PO_4^{-}} > \beta_{PO_4^{-}} > \gamma_{PO_4^{2-}}$$

Q61H lies in between as GTP phosphate group overall charge follows:

$$\alpha_{PO_4^{-}} > \gamma_{PO_4^{2-}} > \beta_{PO_4^{-}}$$

These results suggest that Gln 61 substitutions hinder the catalyzing effect of NRas that consists in modifying GTP electronic density in a manner that further stabilizes a more GDP like transition state than the GTP reactant.

KS HOMO

The previous Löwdin population analysis aimed at attributing partial charges to individual atoms from the overall electron density of the system. In this part, the KS HOMO is studied within the environment provided by each protein complex to assess where the highest energy electrons tend to be located as they are the most likely to react.

Wild-type NRas For wild-type NRas, the KS HOMO electronic density appears prominently on GTP α oxygen atoms, followed by β oxygen atoms and finally γ oxygen

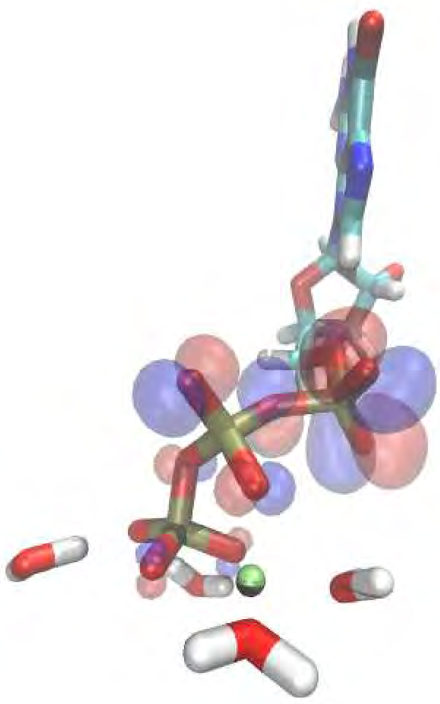


FIGURE 3.23: KS HOMO associated to the DFT system consisting in GTP, Mg^{2+} ion and water molecules within the active site of wild-type NRas. The positive lobes are depicted in blue and their negative counterparts in red.

atoms. According to the chosen cutoff for representing the delimiting surface of the KS HOMO (isovalue of $|0.02|e^- / a_0^3$), the highest energy electrons are distributed following the same trend as observed when attributing partial charges to GTP phosphate groups by a Löwdin population analysis, *i.e.*:

$$\alpha_{PO_4^-} \text{ electronic density} > \beta_{PO_4^-} \text{ electronic density} > \gamma_{PO_4^{2-}} \text{ electronic density}$$

The agreement between both analyses suggests that the KS HOMO makes large contributions when assigning reduced charges to GTP phosphate groups following the Löwdin scheme, although electronic distribution on γ oxygen atoms does not follow the same trend according to both approaches. Indeed, $O_{1\gamma}$ atom holds no electronic density according to the KS HOMO analysis while it is the most negatively charged γ oxygen atom according to the Löwdin scheme. This suggests that the reduced charge on this atom is mainly contributed by lower energy electrons.

Gln 61 mutant NRas Using the same isovalue as for the wild-type ($|0.02|e^- / a_0^3$), striking differences are observed in the KS HOMO associated to the GTP ligand, the

Mg^{2+} ion and the water molecules present in the active site of the Gln 61 mutated proteins.

For Q61E and Q61L, the highest energy electrons are mainly localized, respectively, on one and two of the water molecules present within the corresponding active sites. For Q61E, a part of the highest energy electron density is also located on the three γ oxygen atoms such that the $O_{3\gamma}$ atom holds a larger share than the two others. For Q61L, $O_{1\gamma}$ atom is the only one to also holds a portion of electron density, while significantly smaller than the amount located on the two water molecules. Within the selected cutoff for delimiting the KS HOMO surface, only GTP γ oxygen atoms appear to hold some electron density. For this reason, the γ phosphate group appears to be more negatively charged than the α and β phosphate groups within these two particular mutants. This is in accordance with the Löwdin population analysis. The Q61E GTP γ phosphate group appears to be more negatively charged than the same chemical group within Q61L, again in accordance with the previous reduced charge attribution (see table 3.2). The transient water molecules within the two mutant proteins appear equally reactive towards an electrophile.

For Q61P, the electron density is prominent on GTP γ oxygen atoms. A significantly smaller amount is located on (in decreasing order): *i)* one of the water molecules coordinated to the Mg^{2+} ion, *ii)* one of the transient water molecules from the active site and *iii)* the $O_{\beta\gamma}$ bridging oxygen atom. As within Q61E and Q61L, according to the KS HOMO, GTP γ phosphate group within Q61P is more negatively charged than the α and β phosphate groups, again in accordance with the Löwdin population analysis. The intake of electron density by one of the Mg^{2+} coordinated water molecules could explain why a positive Löwdin reduced charge is attributed to this ion within this mutant protein.

For Q61H and Q61K, the highest energy electron density is located on the guanine moiety. The KS HOMO depicted within the selected cutoff does not allow to assess the relative charge held by GTP phosphate groups. When attributing reduced Löwdin charges to the phosphate moiety of the nucleotide, lower energy occupied MO must contribute more importantly than the KS HOMO.

Finally, within Q61R, the highest energy electron distribution according to the KS HOMO is the most similar to that observed within the wild-type. It appears prominently on GTP α oxygen atoms, followed by $O_{\alpha\beta}$ and $O_{1\beta}$ atoms that display a similar amount, succeeded by $O_{\beta\gamma}$ and $O_{1\gamma}$ atoms after which, $O_{2\gamma}$ atom holds a smaller share. Lastly, one of the water molecules coordinated to the Mg^{2+} ion holds the smallest part, according of the selected isovalue ($|0.02|e^{-}/a_0^3$). In addition to the Mg^{2+} coordinated water molecule, which appears to take a non negligible amount of KS HOMO electron density, two important differences arise within Q61R compared to wild-type NRas: *i)* $O_{2\beta}$ atom holds no high-energy electron density contrary to what is observed within the wild-type and *ii)* $O_{1\gamma}$ atom has the largest share among the γ oxygen atoms, while it has none within the non-mutated form.

GTP electronic density conclusions

The study of GTP electronic distribution through a Löwdin population analysis shows that wild-type NRas active site conformation, obtained through QM/MM MD, is the only one that leads to the GTP β phosphate group being more negatively charged than its γ phosphate group is. This is in agreement with experimental studies [3, 4, 14, 18, 77] that showed that GTP binding to Ras induces a more GDP like electronic repartition of the ligand charge. Moreover, Rudack et al [60] suggested in their theoretical study that this process is partially achieved by the Mg^{2+} ion that temporarily stores electrons from GTP γ phosphate group, rendering it more positive. This mechanism, according to the performed population analysis, is more importantly observed within the wild-type. Indeed, in its active site, the cation holds a larger Löwdin reduced negative charge than in the active site of the mutant proteins. This is associated to the γ phosphorus atom that is also more positively charged within the wild-type than within any of the studied NRas mutant proteins.

Hence, Gln 61 substitutions result in the loss of NRas catalyzing effect, which consists in stabilizing a more GDP like state. Indeed, within the six studied mutant proteins, the γ phosphate group holds more negative charge than the β phosphate group.

Finally, when analyzing GTP electronic density by depicting the KS HOMO, within the non mutated form, the highest energy electrons, *i.e.* the most reactive ones, appear to be distributed in the same order as determined by the Löwdin population analysis *i.e.* $\alpha_{PO_4^-} > \beta_{PO_4^-} > \gamma_{PO_4^{2-}}$. For all Gln 61 mutant proteins, except Q61R, other lower energy electrons contribute when attributing reduced charges to atoms forming GTP phosphate groups.

3.4 Oncogenic Gln 61 mutants impact

In the previous sections, a systematic comparison between wild-type and mutant NRas was made in order to identify the differences induced by Gln 61 substitutions. It is essential to elucidate how these structural rearrangements, described in detail at the beginning of this chapter, influence other active site and ligand properties, also discussed along this chapter, and hence how they impact the protein function. In this section, a final effort is made to relate these structural modifications to the protein active site function.

To facilitate this task, table 3.3 summarizes RMSD, RDF and 2D RDF results for wild-type and Gln 61 mutant NRas.

3.4.1 Qualitative Structure-Activity Relationships

Table 3.3, shows that Q61R mutation destabilizes the most the protein active site. Indeed, within this mutated form, although Thr 35 is able to adopt the conformations encountered within the wild-type, it is *unstable* in Q61R active site as it constantly switches between these conformations. Such a behavior might prevent this residue from interacting in a stable way with a presumed lytic water molecule identified in several crystallographic structures of Ras [50, 55, 62, 63, 70]. Moreover, Spoerner et al. [71] have shown that Thr 35 side chain methyl group regulates the protein dynamical flexibility, essential to its function. The destabilization of this residue within Q61R probably also hinders this role attributed to Thr 35 methyl group.

Another residue presenting remarkable structural differences within Q61R is Gly 60 residue. Indeed, the latter interacts through stable hydrogen bonds with GAP Thr 791 residue within Q61R active site (inset *(a)* of figure 3.10) while it hydrogen bonds with transient water molecules within the wild-type active site (inset *(b)* of figure 3.10). These changes in the interaction network of Gly 60 might be the reason of the lack of water molecules, within Q61R, in the region corresponding to highest water density peak, held by residues 12 and 60, within the wild-type (see Q61R 2D RDF plot in figure 3.20 *(f)*). Hence, Gly 60, a highly conserved residue, might be indirectly involved in the enzyme catalyzed reaction, by maintaining water molecules in this specific region.

This inference is consistent with previous studies [17, 22, 36, 69] that proposed for Gln 61, within wild-type Ras, an indirect steric effect for catalyzing the GTP hydrolysis reaction. Because this residue is coupled to regions of NRas playing a major role in its function, its mutation destroys the pre-organized catalytic configuration adopted by

	Wild-type		**Q61E**		**Q61P**		**Q61H**		**Q61L**		**Q61K**		**Q61R**	
Residue	Conformations	Stability	Conformations	Stability	Conformations	Stability	Conformations	Stability	Conformations	Stability	Conformations	Stability	Conformations	Stability
Gly 12	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s
Gly 13	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s
Tyr 32	open	s	closed	s	open	s	open	s	open	s	closed	s	open	s
Thr 35	HO in, CH_3 out HO out, CH_3 in CH_3 w/ Mg^{2+}	s	HO in, CH_3 out HO out, CH_3 in	s	HO out, CH_3 in CH_3 w/ Mg^{2+}	s	HO in, CH_3 out HO out, CH_3 in CH_3 w/ Mg^{2+}	s	HO in, CH_3 out HO out, CH_3 in HO out, CH_3 out	s	HO out, CH_3 in HO out, CH_3 out	s	HO in, CH_3 out HO out, CH_3 in HO out, CH_3 out	u
Ala 59	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s
Gly 60	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	crystallographic	s	h-bonds w/ Thr 791	u
61	open	s	open bent to GTP	s	crystallographic	s	open bent to GTP	s	1 CH_3 to GTP 2 CH_3 to GTP	u	closed	s	open	u
RDF	Peak Distance Ampl	Integral 5Å 7Å	Peak Distance Ampl	Integral 5Å 7Å	Peak Distance Ampl	Integral 5Å 7Å	Peak Distance Ampl	Integral 5Å 7Å	Peak Distance Ampl	Integral 5Å 7Å	Peak Distance Ampl	Integral 5Å 7Å	Peak Distance Ampl	Integral 5Å 7Å
P_α	—	0 0.20	3.9 Å 0.02	0.18 1.33	3.9 Å 0.10	0.66 2.41	3.7 Å 0.07	0.53 1.62	3.8 Å 0.06	0.55 1.64	—	0.02 0.76	4.0 Å 0.27	1.35 3.43
P_γ	3.8Å 0.21	1.52 3.42	4.0 Å 0.22	1.56 4.46	4.0 Å 0.14	0.99 2.78	3.8 Å 0.36	2.32 5.05	4.0 Å 0.15	1.03 3.54	3.8 Å 0.25	1.17 3.68	4.0 Å 0.24	1.50 5.33
2D RDF	Arch	2.63 3.35	Delocalized	3.33 4.46	Delocalized Arch	2.12 2.77	Localized	4.17 5.05	Delocalized	2.80 3.54	Localized	2.70 3.68	Delocalized	4.34 5.33

TABLE 3.3: Main conformation(s) and general stability of residues forming the active site along RDF and 2D RDF of water molecules within wild-type and Gln 61 mutants NRas. In the stability column, s denotes *stable residue* and u *unstable residue*. The 2D RDF integrals are calculated for P_γ centred squares which borders extend up to 5Å and 7Å away from GTP P_γ atom

Ras upon binding with GAP. Thus, Gln 61 takes action in an indirect manner in the hydrolysis reaction by enabling Gly 60 to correctly position water molecules.

Indeed, as showed in section 3.2.1, for wild-type NRas, Gln 61 side-chain gets out of the active site at the beginning of the simulation. Its carbonyl and amide groups do not interact with any of the ligand atoms and appear to be too far from the reactants to stabilize the transition state. Although Gln 61 interacts with transient water molecules, this interaction is taken over by Gly 60 residue which holds water molecules, along with Gly 12 residue, by hydrogen bonds formed through its nitrogen backbone atom.

In the 2D RDF plots (figure 3.20), the projection of nitrogen backbone atoms clearly shows the rearrangements of the active site residues upon Gln 61 substitution. In the case of Q61R mutation, the one that destabilizes the most the active site *i.e.* the nitrogen backbone atom of Thr 35, does not appear. Indeed, it has moved further than the selected threshold. The corresponding standard deviations can be correlated to the previously reported flexibility loss [37] of the protein backbone due to oncogenic mutations.

As well revealed by the 2D RDF plots, Q61R mutation, results in a highly delocalized water distribution within NRas active site. This is not only due to its different chemical nature, but also to steric effects that displace Gly 60 such that it can no longer participate in maintaining water molecules in a specific side of GTP γ phosphate group, as discussed above.

The substitution of Gln 61 by a histidine (Q61H) or by a lysine (Q61K) leads to a higher amplitude peak of water density, maintained by residues 12 and 60, than that observed within the wild-type (figure 3.20 *(c)* and *(e)*). Yet, the absence of the peak of water density in the vicinity of Thr 35 residue appears to impact the protein enzymatic action. This supports previous studies [23, 38, 43, 53, 63, 70] that considered the presence of a second water molecule in the active site for catalyzing the hydrolysis reaction. Indeed, as summarized in table 3.3, the double integral of wild-type 2D RDF plot leads to a total amount of 1.5 and 3.4 water molecules present in the active site of the protein complex, when taking into account water molecules within 5Å and 7Å of GTP P_γ atom respectively (discarding the two water molecules coordinated to the Mg^{2+} ion). As observed in figure 3.19, these water molecules do not accumulate in a single region of the active site of wild-type NRas, but are precisely positioned. They form an *arch* of water density between GTP γ phosphate group, extending from residues 12 to 35 of the active site. This observation does not imply that all water molecules present in the active site of wild-type NRas chemically participate in the hydrolysis reaction.

Their specific roles remain to be elucidated and might bring about a new proposition concerning the catalyzing mechanisms of GTP hydrolysis within NRas active site.

Concerning the GAP enhanced GTP hydrolysis reaction, how the structural modifications of key NRas active side residues influence the role of the so-called *arginine finger* is investigated.

3.4.2 Ras-GAP binding loop potential interactions

A remarkable structural difference that arises upon Gln 61 substitution pertains to GAP Arg 789 residue in the NRas-GAP arginine binding loop complex. *Native contacts* of this residue with GTP P_γ atom remain stable during the entire simulation within the wild-type protein while they are completely lost after a certain time within all the studied mutants, except for Q61K where a *native contact* remains almost as long as within the wild-type. Moreover, the non-mutated form is the only form where *non-native contacts* appear between its Gln 61 residue and GAP Arg 789 residue.

Evaluating how Gln 61 mutations induce the loss and absence of respectively GAP Arg 789 *native* and *non-native contacts* appears to be crucial even if, during the unrestrained QM/MM molecular dynamics, only a small portion of GAP (the *arginine binding loop*) was taken into account.

For Q61E, Q61P, Q61L and Q61R, GAP Arg 789 *native contacts* with GTP P_γ atom are lost, probably because of the delocalized positioning of water molecules (see figure 3.20 *(a)*, *(b)*, *(d)* and *(f)*). The displaced, quantumly treated, solvent [2], prevents Arg 789 side chain from remaining in the vicinity of this atom because of steric hindrance. The electrostatic interactions between GAP arginine binding loop, NRas and/or the ligand might also be damped by these water molecules.

For Q61H, which presents a more localized 2D water distribution than the wild-type, the higher water density in the vicinity of GTP P_γ atom (as depicted by the higher amplitude peak in both its RDF curve (figure 3.17 *(c)*) and its 2 RDF mapping (figure 3.20 *(c)*)) might be responsible for the vanishing of *native contacts* between Arg 789 and GTP P_γ atom.

Concerning Q61K, water molecules are also more localized than in the wild-type active site (figure 3.20 *(e)*). This mutant presents the most similar RDF plots to those of the wild-type (figures 3.17 *(e)* and 3.18 *(e)*). Hence, *native contacts* between the arginine finger and GTP P_γ atom can persist almost as long as within the wild-type.

---

[2] as water molecules within 5Å of GTP P_γ atom were treated at this level of theory

Moreover, within all Gln 61 mutant proteins, the rapid loss of *native contacts* with GTP P_γ atom and the absence of *non-native contacts* with Gln 61 residue might be due to *i)* the chemical nature of the replacing residue, *ii)* the active site structural reorganization induced by the mutation that could prevent Arg 789 from adopting the position found within the wild-type, *iii)* the change in water molecules distribution so that they also could prevent Arg 789 from adopting an optimal positioning due to steric hindrance and/or damped electrostatic interactions between the two proteins.

Nevertheless, calculations including the entire GAP protein are needed to assess the impact of Gln 61 substitutions on NRas-GTP-GAP interaction network, in particular concerning the interactions the *arginine finger* can potentially engage in. Such calculations might shed light on how, at the molecular level, oncogenic mutations render Ras insensitive to its inactivation by GAPs, as previously showed experimentally [1, 75].

3.4.3 Addressing Gln 61 mutant proteins

As presented in chapter 1, NRas mutations are present in more than a quarter of skin melanomas. Moreover, 60% of them consist in the substitution of Gln 61 residue that is replaced by, in decreasing occurrence, an Arginine (45.5%), a Lysine (35.8%), a Leucine (10.4%), a Histidine (7.1%), a Proline (0.8%) or a Glutamic Acid (0.4%) [19] (*i.e.* Q61R > Q61K > Q61L > Q61H > Q61P > Q61E). NRas proteins hence appear as an attractive target for the pharmaceutical industry. Indeed, restoring the hydrolysis capacity within two of the most widespread mutations (*i.e.* Q61R and Q61K) through a compound would concern more than 80% of skin malignancies due to NRas Gln 61 substitution. Overall, such an accomplishment, would touch more than 12% of all skin cancer patients accounting for the genetic specificity of their disease. Yet, to date, recovering protein function through the binding of small drug molecules remains to be achieved.

With the development of predictive computational methods, the increase in computing power and the availability of computer resources, the discovery of small compounds is accompanied by computer simulations, leading to the field of computer-aided drug design.

As during site-directed mutagenesis experiments, in this section, a second point mutation is inserted in the most widespread NRas mutant, *i.e.* Q61R. Upon previous predictions through the static mode method, that enabled the identification of a suitable site (confidential) inducing the elongation of GTP $P_\gamma - O_{\beta\gamma}$ bond, the structures of three engendered double mutant proteins were generated and will be studied following the same process adopted all along for studying wild-type and Gln 61 mutants NRas. Indeed, the additional substitutions inserted in the identified site, aim to reverse Q61R effect and in this manner recover the protein catalyzing capabilities. Characterizing these three engineered double mutants, makes it possible to assess if the inserted mutations could contribute to recover protein function. Moreover, these theoretical investigations are coupled to experimental characterizations performed by the CRCT-INSERM. Indeed, the location and the chemical nature of the substituting residue could help to propose therapeutic strategies to target NRas.

Water distribution

As water molecules positioning appeared as a key catalyzing mechanism, water distribution is first evaluated within the active site of the three engineered Q61R double mutants NRas. The results of this analysis are presented in figure 3.24.

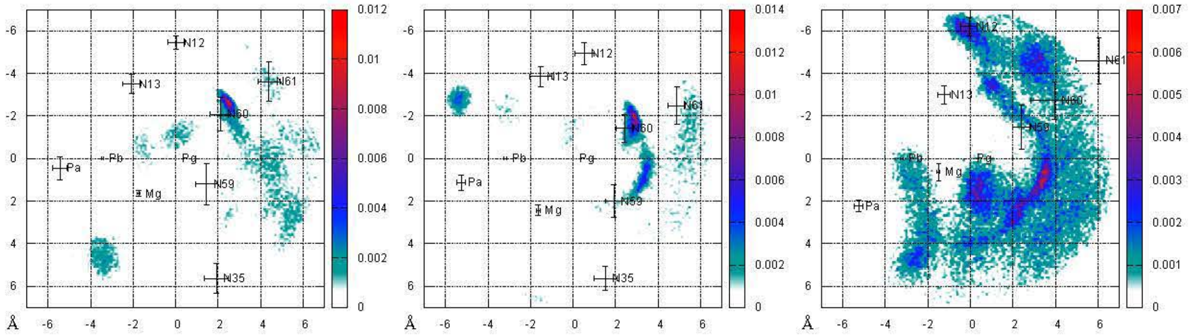


FIGURE 3.24: 2D RDF of water molecules within the active site of the three engineered Q61R double mutants (a), (b) and (c) during QM/MM molecular dynamics. The average coordinates of nitrogen backbone atoms from residues 12, 13, 35, 59, 60 and 61 have also been projected as well as GTP phosphorus atoms and Mg^{2+} ion. The corresponding standard deviation values are represented with error bars.

From this figure, it appears that the three Q61R double mutants impact water delocalization observed within Q61R. The second engineered Q61R double mutant, referred to as $Q61R_{(b)}$, presents the most similar water distribution to that observed within the wild-type protein. The highest amplitude density peak, positioned between residues 12, 59 and 60, which was missing in Q61R, is restored. A smaller arch of water density is also present although it is not as extended as the one observed for the wild-type. The second highest amplitude water density peak is placed next to residue 59 whereas it was positioned between this residue and residue 35 within the non-mutated form. A third peak appears in the side of the active site where GTP P_{α} and P_{β} atoms are located.

Nevertheless, no solvent is present in the immediate vicinity of these atoms, nor next to the Mg^{2+} ion.

Q61R$_{(a)}$ also restores the highest amplitude density peak. An arch of water cannot be discerned within this engineered double mutant although water molecules appear to transit in the corresponding area. Finally, a third water density peak is placed in the side of the active site where the cation is located. A small amount of solvent is present in the vicinity of this ion and near GTP P_β atom.

The third engineered NRas double mutant, Q61R$_{(c)}$, presents the most delocalized water positioning among the three. Nevertheless, water delocalization is lower than within Q61R. Water presence between residues 12, 59 and 60 is restored and the arch can also be distinguished although it is scattered, as within Q61P, and more extended than within the wild-type. Important amounts of solvent are observed next to the Mg^{2+} ion and GTP β phosphate atoms.

The intrinsic properties of the bound nucleotide are crucial for the feasibility of its own hydrolysis within the enzyme. Hence, they are hence subsequently examined within the active sites of Q61R double mutants.

GTP intrinsic properties

As small differences were observed in GTP geometric features between wild-type and mutant NRas proteins, we proceed to directly evaluate GTP P_γ intrinsic mobility that presented more evident differences. The corresponding results are depicted in figure 3.25.

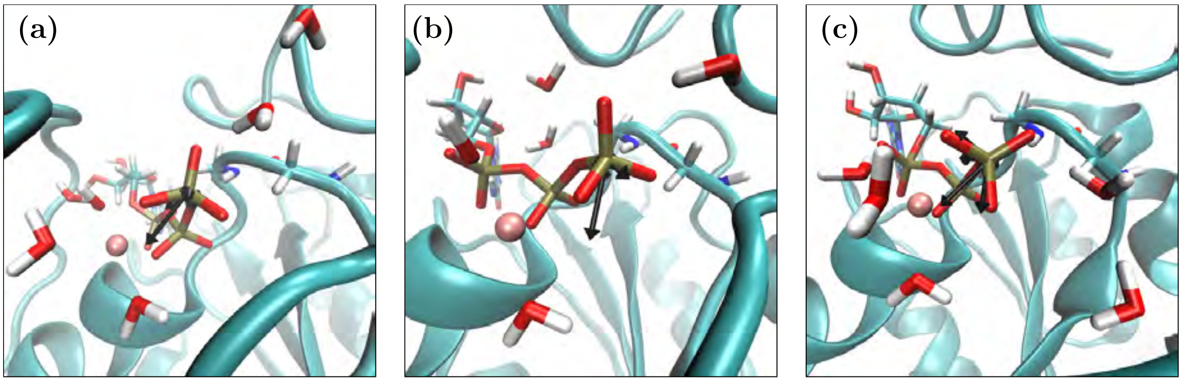


FIGURE 3.25: Intrinsic directional motions of GTP P_γ atom within the active site of the three engineered Q61R double mutants (a), (b) and (c). The amplitude of the black arrows indicate the extent of the allowed motion in the considered direction.

P_γ intrinsic directional motions Within the active site of all the engineered Q61R double mutant proteins, GTP P_γ largest amplitude motion is not directed anymore toward Lys 16 residue, as within Q61R.

For Q61R$_{(a)}$ and Q61R$_{(b)}$ the largest amplitude arrow points towards Asp 57 residue.

For Q61R$_{(c)}$, this motion is directed to the Mg^{2+} ion.

Although motion towards the *Switch I* region, specifically towards Asp 33 residue, is not restored, the largest amplitude motion is again directed to the same side of the active site as within wild-type NRas.

GTP electronic density To end, it appears of major importance to assess if the engineered Q61R double mutant proteins restore NRas major catalyzing effect *i.e.* that of promoting a charge transfer from the γ to the β phosphate group within GTP, which is lost in oncogenic Q61R. As previously, such evaluation is performed first, through a Löwdin population analysis to gain insight on partial charges held by GTP atoms. In a second place, the KS HOMO is studied to see where the highest energy electrons tend to remain.

	Wild-type	Q61R$_{(a)}$	Q61R$_{(b)}$	Q61R$_{(c)}$	Q61R
$\gamma\ PO_4^{2-}$	-0.495546	-0.740905	-0.838226	-0.492496	-0.770763
$\beta\ PO_4^-$	-0.568316	-0.664815	-0.578882	-0.574557	-0.568822
$\alpha\ PO_4^-$	-0.781513	-0.848858	-0.69498	-0.681141	-0.704198
Mg^{2+}	-0.327834	-0.194730	-0.172694	-0.483914	-0.289204

TABLE 3.4: Löwdin reduced charges in a.u. for wild-type, the three engineered Q61R double mutants and Q61R NRas.

The reduced Löwdin charges attributed to the Mg^{2+} ion and to GTP phosphate groups within the three engineered Q61R double mutant proteins are summarized in table 3.4. From this table, it appears that two of these double mutants, Q61R$_{(a)}$ and Q61R$_{(b)}$, leave GTP electronic distribution as it is in Q61R *i.e.* the γ phosphate group is more charged than the β phosphate group. Moreover, the reduced Löwdin charge held by the cation is even smaller than that held within Q61R. This suggests that, upon introduction of the second substitution, the Mg^{2+} ion remains unable to temporarily store electrons during hydrolysis, a mechanism proposed by Rudack et al [60].

For the third engineered Q61R double mutant, Q61R$_{(c)}$, the stabilization of the GDP-like state appears to be restored. Indeed, the charge distribution of GTP follows the same trend as within the wild-type protein *i.e.* $\alpha_{PO_4^-} > \beta_{PO_4^-} > \gamma_{PO_4^{2-}}$. Moreover, the Mg^{2+} holds a bigger Löwdin reduced charge within this double mutant than within the wild-type (-0.483914 vs -0.327834, respectively), suggesting that it can again temporarily store electrons.

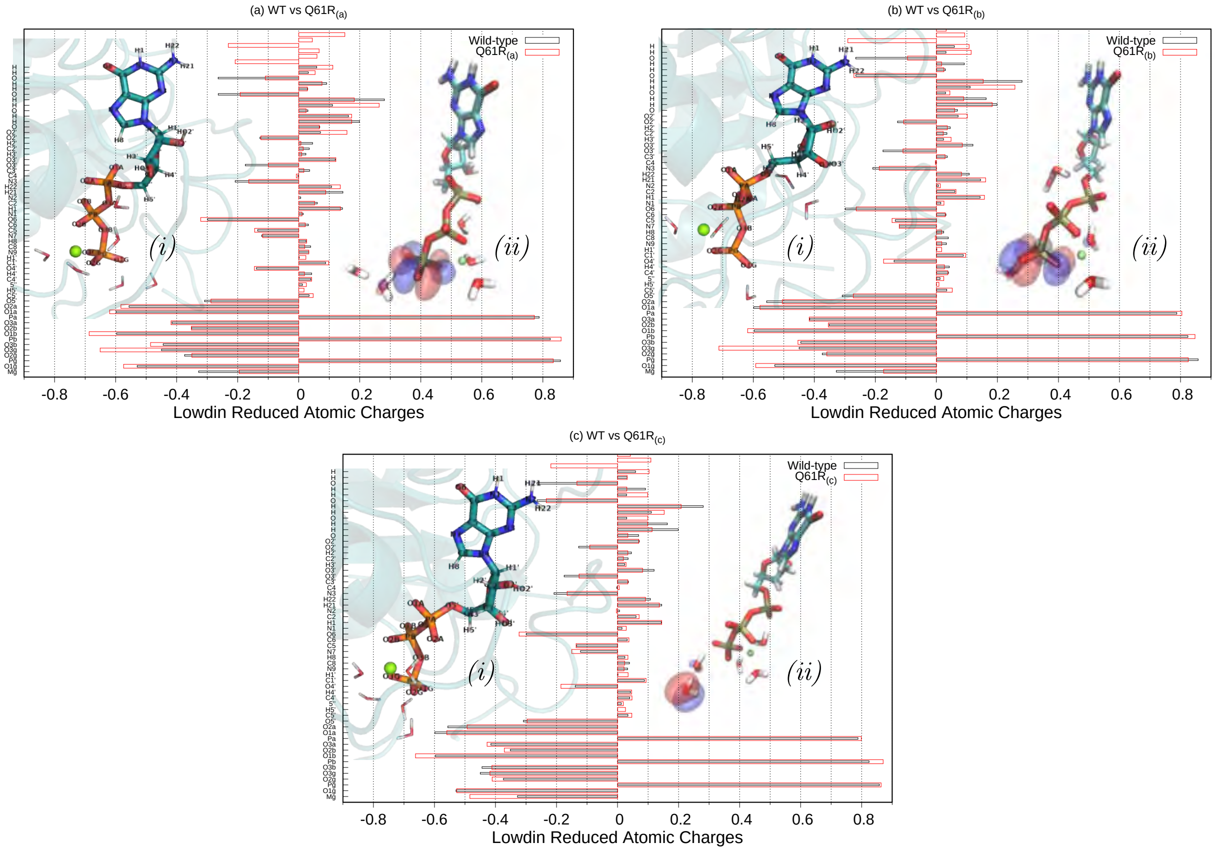


FIGURE 3.26: Löwdin population analysis of GTP, Mg^{2+} and water molecules within the active site of the three engineered Q61R double mutants (a), (b) and (c). In inset *(ii)* of each panel the KS HOMOs of the DFT system is depicted. The isovalue selected for representing the KS HOMO is $|0.02|e^-/a_0^3$. The positive lobes of the KS HOMOs are depicted in blue and the negative ones in red.

The KS HOMO of each engineered Q61R double mutant is depicted in the inset *(ii)* of each panel of figure 3.26.

For Q61R$_{(a)}$ and Q61R$_{(b)}$, the highest energy electrons are localized in the terminal phosphate group of the nucleotide. For Q61R$_{(b)}$ this is in total agreement with the Löwdin population analysis as GTP charge distribution within this double mutant follows the trend $\gamma_{PO_4^{2-}} > \alpha_{PO_4^-} > \beta_{PO_4^-}$ according to both charge distribution analysis. For Q61R$_{(a)}$, the α phosphate group appears as holding the largest amount of negative charge, according to the Löwdin population analysis, while the γ phosphate is the most negatively charged according to the KS HOMO. Hence, the reduced Löwdin charge held by the α phosphate group, within Q61R$_{(a)}$, must be contributed by other lower energy electrons.

For Q61R$_{(c)}$, the highest energy electrons are mainly localized in one of the water molecules from the active site that could act as the lytic water molecule. Hence, charge distribution on the ligand must be due to lower energy electrons. Depicting the lowest unoccupied molecular orbital (LUMO) did not reveal atoms of the phosphate moiety of GTP that could accept the electrons held by this water molecule.

Conclusion and Perspectives

Investigating the active site of wild type NRas and Gln 61 mutant proteins using different methods shed light on the impact such mutations have, and that might, at least in part, explain why the rate of GTP hydrolysis is strongly reduced by substitutions of this residue.

The unrestrained QM/MM molecular dynamics simulations showed that all NRas Gln 61 substitutions induce a structural rearrangement of the active site and modify the distribution of water molecules in it, to a different extent.

Residues Tyr 32, Thr 35, Gly 60 mobility, stability and interaction network are particularly affected depending on the substitution. This is particularly true for Q61R as the flexibility of Tyr 32 decreases, Thr 35 becomes unstable and Gly 60 interacts with GAP Thr 791 residue instead of forming hydrogen bonds with transient water molecules from the active site, as is the case within the wild-type. Residues 35 and 60 being conserved in the more extended family of GTPases, such a change, upon Gln 61 substitution, in the nature and stability of the interactions they are involved in, must contribute to the hydrolysis rate drop observed for all these mutants. Indeed, these residues have been pointed as stabilizing a hydrolysis competent conformation of the active site. They belong to the *Switch I* and *Switch II* regions that are the most flexible parts of Ras proteins and thus adopt multiple conformations. Moreover, Tyr 32 residue, despite not being conserved among all GTPases as Thr 35 and Gly 60, has also been identified as adopting two main conformations that determine the catalyzing competent/incompetent state of the active site.

The chemical participation of water in hydrolysis reactions and the ongoing debate concerning the number of water molecules taking part in Ras and Ras-GAP catalyzed GTP hydrolysis prompt the inspection of water occupancy within NRas active site. For the wild type, water molecules stay close to GTP P_γ atom, forming an arch of density around the γ phosphate group, extended from residues 12 to 35 of the protein. The highest water density is localized between residues 12, 59 and 60. Water molecules are held in this region by hydrogen bonds formed with the nitrogen backbone atoms of Gly 12 and Gly 60 residues. A second peak of density is observed in the vicinity of Thr

35 residue. The presence of two water density peaks, in two distinct regions, supports the presence of an additional water molecule, besides the lytic water molecule, in the protein complex active site for an efficient catalysis of the hydrolysis reaction.

Within all Gln 61 mutant proteins, this precise positioning of water molecules is lost. For Q61E, Q61P, Q61L and Q61R, water molecules distribution is delocalized compared to the wild-type and more solvent appears to be present in the active site. Both delocalization and excess of water might prevent GAP *arginine finger* from interacting with GTP P_γ atom and 61 residue because of steric hindrance. They most probably also disturb GAP Arg 789 crucial electrostatic effect, further damping the interactions of this residue with other Ras residues.

Water distribution within both Q61H and Q61K appears to be more localized than what it observed for the wild-type. In particular, it does not present the second peak of water density placed in the vicinity of Thr 35. This observation further supports the need of the presence of a second water molecule for an efficient catalysis of GTP hydrolysis.

As within the wild-type, Gly 60 residue participates in holding the highest water density peak through hydrogen bonds, this residue is involved in the enzymatic catalysis of the GTP hydrolysis by maintaining water molecules on a specific side of GTP γ phosphate group. This could explain the conservation of Gly 60 residue in all GTPases. The role of Gln 61 then appears to be consistent with an indirect steric effect in stabilizing the pre-organized catalytic configuration of the active site: it enables Gly 60 residue to correctly position water molecules by ensuring a certain conformation of the active site. This conformation is disrupted by Gln 61 substitutions. Both the chemical nature of the replacing residue and the structural rearrangements it induces lead to a different distribution of water molecules within the active site in a manner that NRas loses its catalyzing power.

The participation of Gly 12 residue in maintaining water molecules explains why its substitution by an alanine residue is nevertheless oncogenic. Indeed, it has been demonstrated that mutations of this residue also diminish the rate of GTP enzymatic hydrolysis. The impairment of catalytic power has been attributed to steric hindrance caused by the side chain of the replacing residue that pushes away the presumed lytic water molecule as well as Gln 61 residue. Nevertheless, this would hardly be the case for the small side chain of an alanine residue. Hence, Gly 12 substitutions hinder GTP hydrolysis, not only through steric hindrance, but also because of the role this residue has in water molecules placement.

As the intrinsic properties of GTP are modified by the protein upon binding, the differences of this impact due to the differences in wild-type and mutant NRas active site structuration and water distribution were assessed. No major differences were found in GTP bond lengths and bond angles while in the active site of wild-type and Gln 61 mutant NRas.

Mechanistically, GTP P_γ atom appears to adopt different directions of motion. Within wild-type NRas, the biggest amplitude motion is directed toward the *Switch I* region, specifically toward Asp 33 residue. Within Gln 61 mutants, this movement is directed to other sides of the active site. In particular, for the most widespread mutations, Q61R and Q61K, it points to an opposite direction, toward the *P-loop* and *Switch II*, respectively.

Moreover, it has been experimentally determined that a major catalyzing effect consists in a charge transfer from the γ to the β phosphate group induced by Ras upon GTP binding. This effect is enhanced by GAP Arg 789 residue, further stabilizing the GDP-like charge distribution of GTP. Such charge distribution is observed within wild-type NRas. The DFT calculations show that the reduced atomic charges of the nucleotide result in a charge distribution such that the β phosphate group holds more negative charge than the γ phosphate. This agreement with experiment suggests that the wild-type active site conformation together with water molecules placement in it, reached through QM/MM dynamics, are representative of the protein active site conformation. The same analysis performed for the studied Gln 61 mutant proteins, shows that the charge transfer is hindered by all the considered Gln 61 substitutions. Indeed, GTP charge distribution is such that the γ phosphate group holds more negative charge than the β phosphate. Hence, the reduced GTP hydrolysis rate in oncogenic Gln 61 mutants must also be due to the loss of NRas major catalyzing effect: that of stabilizing a GDP-like state of the reactant nucleotide.

Depicting the KS HOMO of the DFT system, enable the identification of the location of the highest energy electrons, the most likely to react. For wild-type NRas, they appeared to be distributed on the phosphate moiety of the nucleotide in a manner that charge distribution also followed the trend $\alpha_{PO_4^-} > \beta_{PO_4^-} > \gamma_{PO_4^{2-}}$ unveiled by the Löwdin population analysis. The most similar high energy electron distribution was observed for Q61R although, again, the β phosphate group appeared to hold less charge than this same group within the wild type.

Finally, engineered double mutants of the most widespread Gln 61 NRas mutant, Q61R, were investigated. Their water distribution shows that they all enable to recover the highest water density peak, held by Gly 12 and Gly 60 residues, lost in the single mutant. Moreover, they all address water delocalization although the third double mutant, Q61R$_{(c)}$, does it to a lesser extent. The second one, Q61R$_{(b)}$, recovers the arch of water density almost as within the wild-type protein. Concerning the nucleotide intrinsic properties, while the three prevent GTP P_{γ} atom to move toward the *P-loop*, as is the case within Q61R, only the third one, Q61R$_{(c)}$, addressees GTP charge distribution.

Comparing the active site of wild-type and Gln 61 mutant NRas, revealed the impact of such mutations at the molecular level. The latter, shed light on key catalyzing mechanisms and how specifically each substitution hinders them. Nevertheless, the study of the entire NRas-GAP complex shall be performed to get insight on how Gln 61 substitutions hinder the action of GAP in enhancing the rate of GTP hydrolysis. Concerning the ongoing debate of the number of solvent molecules implicated in the enzymatic reaction, the presence of a second water molecule has been revealed. Yet, its precise role in the enzymatic hydrolysis reaction remains to be elucidated. This is being currently investigated and will certainly result in a new tint of how GTP hydrolysis proceeds in NRas active site. Finally, the comparison between wild-type and Gln 61 mutant NRas, also provided hints on how to address each oncogenic mutant effect. The characterization of engineered doubly mutated proteins of the most widespread NRas mutation revealed that some wild-type characteristics could be restored through specific chemical and structural modifications that were introduced. Such mutations have actually been tested *in vitro* by our colleagues from CRCT-INSERM and validate the predictive possibilities of our approach, showing that GTP hydrolysis in Q61R can be restored in this manner. These results pave the way to new therapeutic strategies to target NRas Q61R. Nevertheless, modifications capable of restoring all NRas catalyzing effects, for each Gln 61 mutant, remain to be found. Further investigations need to be carried in particular for understanding the higher occurrence of specific mutations in the development of specific malignancies and thereby target such diseases while accounting for the molecular heterogeneity of cancer.

Appendix A

Initial structures obtainment and simulations setup

A.1 Initial Structures

A.1.1 Wild-type NRas

NRas 3D structure in complex with a slowly hydrolyzable GTP analogue was just recently published [30]. Prior to this structure, none corresponding to the GTP bound NRas was available. Hence, for studying this complex, the initial coordinates were assembled from several X-ray resolved crystallographic structures of different members of the Ras family.

The founder structure of the initial one used during the simulations, corresponds to GDP bound NRas. It was obtained from the Brookhaven Protein Data Bank (PDB) entry 3CON [48]. This structure was carefully aligned to that of HRas in complex with a slowly hydrolyzable GTP analogue (PDB entry: 1QRA [63]) in order to get the nucleotide coordinates within the active site of NRas protein. The coordinates of two water molecules positioned in the vicinity of the γ-phosphate group, the magnesium cation (Mg^{2+}) and two water molecules coordinated to it, were also extracted from 1QRA. Residues 61-71 that are missing in 3CON but are conserved in both NRas and HRas, were taken from 1QRA to complete 3CON structure. To overcome a possible rearrangement of the protein due to the different bound nucleotide (GDP vs GTP), the resulting complex (*i.e.* NRas-GTP) was energy minimized using AMBER package [13]. A 1ns unconstrained molecular mechanics (MM) molecular dynamics (MD) simulation was subsequently carried in implicit solvent using the Born model [25], with a 1 fs time step, at 300K.

In order to account for the effects of the so-called GAP *arginine finger* (residue Arg 789), it is necessary to position GAP with respect to NRas and extract the corresponding GAP coordinates. This was achieved by aligning the NRas-GTP structure, obtained through the previously described procedure, to HRas-GAP complex provided in PDB entry 1WQ1 [62]. The coordinates of the so-called GAP *arginine binding loop* (residues 780 to 795 only) were kept for the initial NRas-GTP-GAP structure used during the simulations.

A.1.2 Gln 61 mutant NRas

Six oncogenic mutant structures were prepared from the wild-type structure, substituting Gln 61 by arginine (Q61R), lysine (Q61K), leucine (Q61L), histidine (Q61H), proline (Q61P) and glutamic acid (Q61E), using PyMOL mutagenesis command [64]. Indeed, a previous X-ray structure analysis [37] concluded that the overall structural changes between the wild-type and oncogenic mutant proteins are very small. The only consistent changes appear in the *Switch I* (residues 30 to 38) and *Switch II* (residues 57 to 67) regions which have been nevertheless reported to be very mobile by several experimental studies [36, 46, 50] . Each mutant structure was first minimized, before adding the solvent, using AMBER package [13].

A.1.3 Static modes structures

Wild-type and Gln 61 mutant NRas structures for which the static modes were calculated, were extracted from the QM/MM trajectories obtained. The most representative configuration of each NRas-GTP complex was identified by a clustering algorithm [58] implemented in the AMBER package. The structures of the protein-ligand complexes together with the magnesium ion, two coordinated water molecules and water molecules present within 5Å of GTP P_γ atom were subsequently energy minimized in implicit solvent, using the generalized Born model [25] until the convergence criterion was reached (root mean square gradient $< 10^{-6} kcal/(mol{\cdot}\text{Å})$). Following this procedure, water molecules occupy positions that correspond to high water density regions identified during the QM/MM study through the 2D RDF algorithm described in appendix B.

A.1.4 DFT initial structures

The protein complexes of wild-type and Gln 61 mutant NRas for performing DFT calculations are the previously energy minimized ones for which the static modes were calculated. These structures were solvated in a periodic box of TIP3P water which boundaries extend up to 14Å from the solute. Na^{+} ions were added to neutralize the obtained systems.

A.2 Simulations setup

A.2.1 MM Molecular Dynamics

All MM molecular dynamics simulations were carried out using AMBER package [13].

The ff99bsc0 force field [51] was employed for describing the protein motions. The simulations were carried out in explicit solvent using the TIP3P water model [54], under periodic boundary conditions, with neutralizing Na^{+} ions. To get rid of possible vacuum bubbles created during the solvation process, 10000 cycles of energy minimization were first performed with a 500 $kcal \cdot mol^{-1} \cdot$ Å^{-2} restraint applied to the solute.

Subsequently, the entire system was gently heated to 300K using the Langevin thermostat with a collision frequency of 2 ps, at constant volume, with weak restraints of 5 $kcal \cdot mol^{-1} \cdot$ Å^{-2} applied to the entire protein complex.

Finally, the production set-up was implemented in the NPT ensemble with 1 atm target pressure, using the Berendsen barostat with a pressure relaxation time of 1 ps. Long-range electrostatic interactions were treated using the particle-mesh Ewald method with a 12.0 Å cutoff. Covalent bonds involving hydrogen atoms were constrained using the shake algorithm [61] so that a 2 fs step could be used.

The system properties of the obtained trajectories (energy, density, temperature, pressure, volume, backbone RMSD) were monitored to assess their quality.

A.2.2 QM/MM Molecular Dynamics

All QM/MM molecular dynamics simulations were carried out using AMBER package [76].

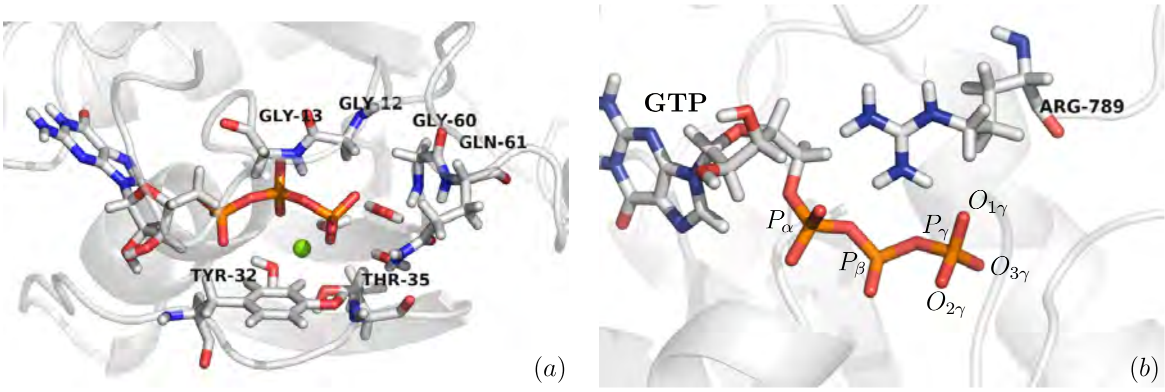


FIGURE A.1: View of the quantum region consisting of: (a) residues 12−13, 32, 35, 60−61 from Ras, GTP and Mg^{2+} ion with two coordinated water molecules, (b) GAP Arg 789 and GTP.

Each molecular system was partitioned into a QM and a MM region, the QM region consisting in critical residues 12−13, 32, 35, 60−61 from NRas, Arg 789 from GAP, GTP and the Mg^{2+} ion with its two coordinated water molecules (see figure A.1). Water molecules found within 5Å of GTP P_γ atom (present in the crystallographic data or introduced by the explicit solvation of the system described below), were also treated quantum mechanically during the simulation. Because of diffusion, their MM/QM identity was updated when they exited/entered the QM region of the protein active site.

The quantum treatment of several residues from the active site leads to the introduction of 10 link atoms, one per QM - MM bond that was sectioned to define the QM region boundaries. These atoms were added following the recommended settings in Amber molecular dynamics package.

The semi-empirical PM3 Hamiltonian [72] and the classical ff99bsc0 force field [51] were used to treat the QM and the MM regions respectively. Na^+ ions were added to neutralize the total charge of the system which was solvated in a periodic box of TIP3P [54] water extending up to minimum cutoff of 8 Å from the protein boundary.

In order to get rid of possible vacuum bubbles created during the solvation of the protein complex, 10000 cycles of minimization were first performed with a 500 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ restraint applied to the solute. QM/MM minimization runs were subsequently carried out in order to remove bad contacts and relax the structure, applying a 5 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ restraint on residues 780 and 795 from GAP only to avoid its undocking. As many runs were performed as needed to reach values of 10^{-2} Å for the RMS. The entire system was then gently heated from 0 to 300 K using the Langevin thermostat with a collision frequency of 2 ps, at constant volume, with weak restraints of 5 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ applied to the entire protein-ligand complex.

Finally, the production set up was implemented in the NPT ensemble under periodic boundary conditions with 1 atm target pressure, and using the Berendsen barostat with a pressure relaxation time of 1 ps. Long-range electrostatic interactions were treated using the particle-mesh Ewald method [76]. Covalent bonds involving hydrogen atoms from the MM region were constrained using the shake algorithm [61] so that a 1 fs step could be used. The first 100 ps of this production phase are considered as the equilibration phase given that the RMSD plots of the backbone atoms do not present a significant drift after this simulation time.

The production was run for 1 ns by portions of 50 ps because of the lack, at the time, of an automated procedure to treat quantum mechanically water molecules within 5 Å of GTP P_γ atom. Indeed, water molecules initially present in the active site are free to move outside but can be replaced by water molecules from the bulk solvent. In this manner, for the first time, transient water molecules were treated quantum mechanically in NRas active site in an adaptative way. Despite being completely unconstrained, both water molecules coordinated to the Mg^{2+} ion remain in their initial positions during the entire simulation in accordance with previous theoretical studies[70] .

For wild-type and Gln 61 mutant NRas proteins, system properties (energy, density, temperature, pressure, volume, back-bone RMSD) were monitored to check the quality of the obtained trajectories.

A.2.3 DFT Calculations

DFT calculations of NRas reaction center were performed using Orca package interfaced to Amber, so that the protein environment together with the solvent effects could be taken into account in a DFT/MM scheme [76].

GTP, water molecules placed within 5Å of its P_γ atom, Mg^{2+} ion and both coordinated water molecules were all treated at the quantum level using the Generalized Gradient Approximation (GGA) functional PBE and the Def2-SVP basis set during consecutive single point energy calculations in the course of a molecular dynamics simulation. This basis set was used despite being a minimal basis for core orbitals because *i)* it is still larger than a minimal basis set due to its double-zeta quality for the valence shell, *ii)* it is suitable for light elements *i.e.* from the three first rows of the periodic table and *iii)* a set of polarization functions is included such that it is also applied to hydrogen atoms. Moreover, as a Löwdin population analysis will be carried out, a larger basis set would not be suitable. Indeed, population analysis based on basis functions, such as the Löwdin population analysis, are more accurate for comparing trends in electron distribution when basis set containing tight functions (as opposed at diffuse functions) are used.

Appendix B

Developed Algorithms

B.1 2D Radial Distribution Function

The 2D RDF algorithm counts the number of water molecules within a certain distance of a given atom and projects them on a chosen plane. To define this plane, an orthonormal basis is built using the coordinates of three arbitrarily selected atoms. This orthonormal basis can be fixed (*i.e.* constructed once at the beginning of the trajectories under analysis) or can be mobile (*i.e.* adapt to the coordinates changes of the three atoms by being re-defined at each step of the analysis run). The latter is preferred to account for stochastic and thermal motion.

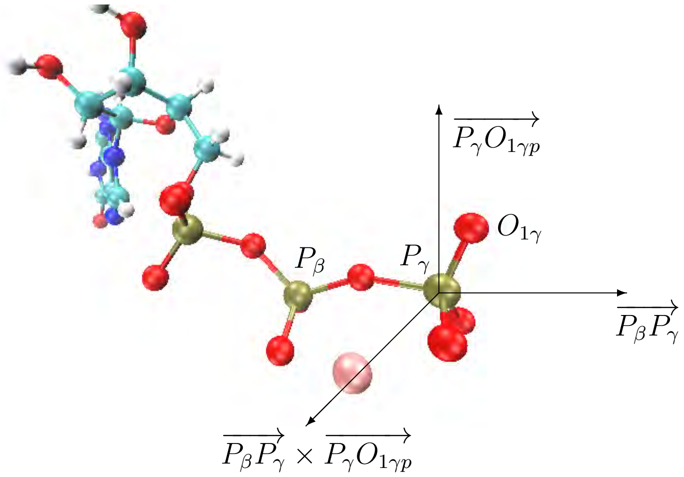


FIGURE B.1: Orthonormal basis defined by GTP P_β, P_γ and $O_{1\gamma}$ atoms

For describing water occupancy within the active site of wild-type and mutant NRas, a plane containing GTP P_β and P_γ atoms was considered to be the most appropriate. The coordinates of three GTP atoms, P_β, P_γ and $O_{1\gamma}$ (see figure B.1), were employed to build the orthonormal basis that in turn was used to define the selected plane.

The first vector of the orthonormal basis is simply normalized $\overrightarrow{P_\beta P_\gamma}$. To build the second vector, $O_{1\gamma}$ coordinates were projected orthogonally on the plane containing P_γ atom and to which $\overrightarrow{P_\beta P_\gamma}$ is a normal vector. In this manner, the second vector is simply normalized $\overrightarrow{P_\gamma O_{1\gamma p}}$, where $O_{1\gamma p}$ is the orthogonal projection of $O_{1\gamma}$. Finally, the third vector is defined by the cross product of the two first *i.e.* $\overrightarrow{P_\beta P_\gamma}$ x $\overrightarrow{P_\gamma O_{1\gamma p}}$.

The plane perpendicular to $\overrightarrow{P_\gamma O_{1\gamma p}}$, containing GTP P_β and P_γ atoms, was selected as the *projection* plane.

Bibliography

[1] H. Adari et al. “Guanosine triphosphatase activating protein (GAP) interacts with the p21 ras effector binding domain”. In: *Science* 240 (1988), pp. 518–521.

[2] M.R. Ahmadian et al. “Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras”. In: *Nature Structural Biology* 4 (1997), pp. 686–689.

[3] C. Allin and K. Gerwert. “Ras catalyzes GTP hydrolysis by shifting negative charges from γ- to β-Phosphate As Revealed by Time-Resolved FTIR Difference Spectroscopy”. In: *Biochemistry* 40 (2001), pp. 3037–3046.

[4] C. Allin et al. “Monitoring the GAP catalyzed H-Ras GTPase reaction at atomic resolution in real time”. In: *PNAS* 98 (2001), pp. 7754–7759.

[5] M. Barbacid. “ras Genes”. In: *Annual Review of Biochemistry* 56 (1987), pp. 779–827.

[6] M.S. Boguski and F. McCormick. “Proteins regulating Ras and its relatives”. In: *Nature* 366 (1993), pp. 643–654.

[7] J.L. Bos, H. Rehmann, and A. Wittinghofer. “GEFs and GAPs: Critical Elements in the Control of Small G proteins”. In: *Cell* 129 (2007), pp. 865–877.

[8] H.R. Bourne, D.A. Sanders, and F. McCormick. “The GTPase superfamily: a conserved switch for diverse cell functions”. In: *Nature* 348 (1990), pp. 125–132.

[9] H.R. Bourne, D.A. Sanders, and F. McCormick. “The GTPase superfamily: conserved structure and molecular mechanism”. In: *Nature* 349 (1991), pp. 117–127.

[10] M. Brut et al. “The Static Modes: An alternative approach for the treatment of macro- and bio-molecular induced-fit flexibility”. In: *The European Physical Journal E* 28 (2009), pp. 17–25.

[11] M. Brut et al. “Toward in Silico Biomolecular Manipulation through Static Modes: Atomic Scale Characterization of HIV-1 Protease Flexibility”. In: *Journal of Physical Chemistry B* 118(11) (2014), pp. 2821–2830.

[12] B.M.T. Burgering and J.L. Bos. “Regulation of Ras-mediated signalling: more than one way to skin a cat”. In: *Trends in Biochemical Sciences* 1995 (1995), pp. 18–22.

[13] D.A. Case et al. *AMBER 2017*. University of California. San Francisco, 2017.

[14] V. Cepus et al. “Time-resolved FTIR studies of the GTPase reaction of H-Ras p21 reveal a key role for the β - phosphate”. In: *Biochemistry* 37 (1998), pp. 10263–10271.

[15] A.D. Cox and C.J. Der. “Ras history: The saga continues”. In: *Small GTPases* 1 (2010), pp. 2–27.

[16] Christopher J. Cramer. *Essentials of Computational Chemistry*. 2nd ed. John Wiley and Sons, Ltd. ISBN: 0-470-09181-9.

[17] C.J. Der, T. Finkel, and G.M. Cooper. “Biological and Biochemical properties of human rasH genes mutated at codon 61”. In: *Cell* 44 (1986), pp. 167–176.

[18] X. Du, H. Frei, and S-H. Kim. “The mechanism of GTP hydrolysis by Ras probed by Fourier Transform Infrared Spectroscopy”. In: *Journal of Biological Chemistry* 275 (2000), pp. 8492–8500.

[19] S.A. Forbes et al. “The Catalogue of Somatic Mutations in Cancer (COSMIC)”. In: *Current protocols in human genetics* (2008).

[20] M. Frech et al. “Role of Glutamine-61 in the hydrolysis of GTP by p21^{H-ras}: an experimental and theoretical study”. In: *Biochemistry* 33 (1994), pp. 3237–3244.

[21] M. Geyer et al. “Conformational transitions in p21ras and in its complexes with the effector protein Raf-RBD and the GTPase activating protein GAP”. In: *Biochemistry* 35 (1996), pp. 10308–10320.

[22] T.M. Glennon, J. Villa, and A. Warshel. “How does GAP catalyze the GTPase reaction of Ras? A computer simulation study”. In: *Biochemistry* 39 (2000), pp. 9641–9651.

[23] B.L. Grigorenko et al. “Mechanisms of Guanosine Triphosphate hydrolysis by Ras and Ras-GAP proteins as rationalized by Ab-initio QM/MM simulations”. In: *PROTEINS: Structure, Function and Bioinformatics* 66 (2007), pp. 456–466.

[24] B.L. Grigorenko et al. “QM/MM Modeling the Ras-GAP catalyzed hydrolysis of Guanosine Triphosphate”. In: *PROTEINS: Structure, Function and Bioinformatics* 60 (2005), pp. 495–503.

[25] G.D. Hawkins, C.J. Cramer, and D.G. Truhlar. “Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium”. In: *Journal of Physical Chemistry* 100 (1996), pp. 19824–19839.

[26] C. Herrmann, G.A. Martin, and A. Wittinghofer. “Quantitative Analysis of the Complex between p21ras and the Ras-binding Domains of the Human Raf-1 Protein Kinase”. In: *Journal of Biological Chemistry* 270 (1995), pp. 2901–2905.

[27] P. Hohenberg and W. Kohn. “Inhomogeneous electron gas”. In: *Physical Review* 136 (1964), pp. 864–871.

[28] W. Humphrey, A. Dalke, and K. Schulten. “VMD – Visual Molecular Dynamics”. In: *Journal of Molecular Graphics* 14 (1996), pp. 33–38.

[29] Frank Jensen. *Introduction to Computational Chemistry*. 2nd ed. John Wiley and Sons, Ltd. ISBN: 978-0-470-01186-7.

[30] C.W. Johnson et al. “The small GTPases K-Ras, N-Ras and H-Ras have distinct biochemical properties determined by allosteric effects”. In: *Journal of Biological Chemistry* 292 (2017), pp. 12981–12993.

[31] M.G. Khrenova et al. “Hydrolysis of Guanosine Triphosphate (GTP) by the Ras-GAP Protein Complex: Reaction Mechanism and Kinetic Scheme”. In: *Journal of Physical Chemistry B* 119 (2015), pp. 12838–12845.

[32] M. Klän, J. Schlitter, and L. Gerwert. “Theoretical IR Spectroscopy Based on QM/MM Calculations Provides Changes in Charge Distribution, Bond Lengths, and Bond Angles of the GTP ligand Induced by the Ras-Protein”. In: *Biophysical journal* 88 (2005), pp. 3829–3844.

[33] W. Kohn and L.J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Physical Review* 140 (1965), pp. 1133–1138.

[34] C. Kötting and K. Gerwert. “Time-resolved FTIR studies provide activation free energy, activation enthalpy and activation entropy for GTPase reactions”. In: *Journal of Chemical Physics* 307 (2004), pp. 227–232.

[35] C. Kötting et al. “The GAP arginine finger movement into the catalytic site of Ras increases the activation entropy”. In: *PNAS* 105 (2008), pp. 6260–6265.

[36] P.J. Kraulis et al. “Solution structure and dynamics of ras p21-GDP determined by heteronuclear 3-dimensional and 4-dimansional NMR-spectroscopy”. In: *Biochemistry* 33 (1994), pp. 3515–3531.

[37] U. Krengel et al. “Three-Dimensional structures of H-Ras p21 mutants: molecular basis for their inability to function as signal switch molecules”. In: *Cell* 62 (1990), pp. 539–548.

[38] R. Langen, T. Schweins, and A. Warshel. “On the mechanism of Guanosine Triphosphate Hydrolysis in ras p21 proteins”. In: *Biochemistry* 31 (1992), pp. 8691–8696.

[39] Michael Levitt. “The birth of computational structural biology”. In: *Nature Structural Biology* 8 (2001), pp. 392–393.

[40] Errol L. Lewars. *Computational Chemistry*. 2nd ed. Springer. ISBN: 978-90-481-3860-9.

[41] Errol L. Lewars. *Computational Chemistry*. 2nd ed. Springer. ISBN: 978-90-481-3860-9.

[42] K.A. Maegley, S.J. Admiraal, and D. Herschlag. “Ras-catalyzed hydrolysis of GTP: A new perspective from model studies”. In: *PNAS* 93 (1996), pp. 8160–8166.

[43] F. Martín-García et al. "The role of Gln 61 in HRas GTP hydrolysis: a Quantum Mechanics/Molecular Mechanics study". In: *Biophysical Journal* 102 (2012), pp. 152–157.

[44] F. McCormick and A. Wittinghofer. "Interactions between Ras proteins and their effectors". In: *Current Opinion in Biotechnology* 7 (1996), pp. 449–456.

[45] J.P. McGrath et al. "Comparative biochemical properties of normal and activated human ras p21 protein". In: *Nature* 310 (1984), pp. 644–649.

[46] M.V. Milburn et al. "Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins". In: *Science* 247 (1990), pp. 939–945.

[47] T. Miyakawa et al. "Analysis of water molecules in the Hras-GTP and GDP complexes with molecular dynamics". In: *Chemistry, Physics and Biology, Progress in Theoretical Chemistry and Physics* 27 (2013), pp. 351–360.

[48] L. Nedyalkova et al. "Crystal structure of the human NRas GTPase bound with GDP". Brookhaven Protein Data Bank entry 3CON. 2008.

[49] David L. Nelson and Michael M. Cox. *Principles of biochemistry*. 4th ed. W.H. Freeman and Company. ISBN: 978-0-7167-4339-2.

[50] E.F. Pai 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". In: *Journal of the European Molecular Biology Organization* 9 (1990), pp. 2351–2359.

[51] A. Perez et al. "Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alpha/gamma Conformers". In: *Biophysical journal* 92 (2007), pp. 3817–3829.

[52] B.R. Prasad, N.V. Plotnikov, and A. Warshel. "Addressing open questions about phosphate hydrolysis pathways by careful free energy mapping". In: *Journal of Physical Chemistry B* 117 (2013), pp. 153–163.

[53] R. Prasad et al. "Quantitative exploration of the molecular origin of the activation of GTPase". In: *PNAS* 110 (2013), pp. 20509–20514.

[54] D. Price and C. Brooks. "A modified TIP3P water potential for simulation with Ewald summation". In: *Journal of Chemical Physics* 121 (2004), pp. 10096–10103.

[55] G.G Privé et al. "X-ray crystal structures of transforming p21 ras mutants suggest a transition-state stabilization mechanism for GTP hydrolysis". In: *Biochemistry* 89 (1992), pp. 3649–3653.

[56] Y. Pylayeva-Gupta, E. Grabocka, and D. Bar-Sagi. "RAS oncogenes: weaving a tumorigenic web". In: *Nature Reviews* 11 (2011), pp. 761–774.

[57] H. Resat et al. "The arginine finger of RasGAP helps Gln-61 align the nucleophilic water in GAP-stimulated hydrolysis of GTP". In: *PNAS* 98 (2001), pp. 6033–6038.

[58] D.R. Roe and T.E. Cheatham. “PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data”. In: *Journal of Chemical Theory and Computation* 9 (2013), pp. 3084–3095.

[59] T. Rudack et al. “Ras and GTPase-activating protein (GAP) drive GTP into a precatalytic state as revealed by combining FTIR and biomoleculat simulations”. In: *PNAS* 109 (2012), pp. 15295–15300.

[60] T. Rudack et al. “The role of magnesium for geometry and charge in GTP hydrolysis, revealed by quantum mechanics/molecular mechanics simulations”. In: *Biophysical Journal* 103 (2012), pp. 293–302.

[61] J.P. Ryckaert, G. Ciccotti, and H. Berendsen. “Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes”. In: *Journal of Chemical Physics* 23 (1977), pp. 327–341.

[62] K. Schefzek et al. “The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants”. In: *Science* 277 (1997), pp. 333–338.

[63] A.J. Scheidig, C. Burmester, and R.S. Goody. “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”. In: *Structure* 7 (1999), pp. 1311–1324.

[64] Schrödinger, LLC. “The PyMOL Molecular Graphics System, Version 1.8”. Nov. 2015.

[65] T. Schweins, R. Langen, and A. Warshel. “Why have mutagenesis studies not located the general base in ras p21”. In: *Nature Structural Biology* 1 (1994), pp. 476–484.

[66] T. Schweins and A. Warshel. “Mechanistic analysis of the observed linear free energy relationships in $p21^{ras}$ and related systems”. In: *Biochemistry* 35 (1996), pp. 14232–14243.

[67] T. Schweins et al. “Linear Free Energy Relationships in the Intrinsic and GTPase Activating Protein-Stimulated Guanosine 5’-Triphosphate Hydrolysis of $p21^{ras}$”. In: *Biochemistry* 35 (1996), pp. 14225–14231.

[68] T. Schweins et al. “Substrate-assisted catalysis as a mechanism for GTP hydrolysis of $p21^{ras}$ and other GTP-binding proteins”. In: *Nature Structural Biology* 2 (1995), pp. 36–44.

[69] A. Shurki and A. Warshel. “Why does the ras switch break by oncogenic mutations?” In: *PROTEINS: Structure, Function and Bioinformatics* 55 (2004), pp. 1–10.

[70] T.A. Soares, J.H. Miller, and T.P. Straatsma. “Revisiting the structural flexibility of the complex p21ras-GTP: The catalytic conformation of the molecular switch II”. In: *PROTEINS: Structure, Function and Genetics* 45 (2001), pp. 297–312.

[71] M. Spoerner et al. "Dynamic properties of the Ras switch I region and its importance for binding to effectors". In: *PNAS* 98 (2001), pp. 4944–4949.

[72] J.P Stewart. "Optimization of parameters for semiempirical methods I. Method". In: *Journal of Computational Chemistry* 10 (1989), pp. 209–220.

[73] R.W. Sweet et al. "The product of ras is a GTPase and the T24 oncogenic mutant is deficient in this activity". In: *Nature* 311 (1984), pp. 273–275.

[74] I.A. Topol et al. "Quantum chemical modeling of the GTP hydrolysis by the RAS-GAP protein complex". In: *Biochimica and Biophysica Acta* 1700 (2004), pp. 125–136.

[75] M. Trahey and F. McCormick. "A cytoplasmic protein stimulates normal N-ras p21 GTPase, but does not affect oncogenic mutants". In: *Science* 238 (1987), pp. 542–545.

[76] R. Walker, M. Crowley, and D. Case. "The implementation of a fast and accurate QM/MM potential method in Amber". In: *Journal of Computational Chemistry* 29 (2008), pp. 1019–1031.

[77] J.H. Wang et al. "Raman Difference Studies of GDP and GTP Binding to c-Harvey ras". In: *Biochemistry* 37 (1998), pp. 11106–11116.

[78] A. Warshel and M. Levitt. "Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme". In: *Journal of Molecular Biology* 103 (1976), pp. 227–249.

[79] A. Wittinghofer and C. Herrmann. "Ras-effector interactions, the problem of specificity". In: *Federation of European Biochemical Societies* 369 (1995), pp. 52–56.