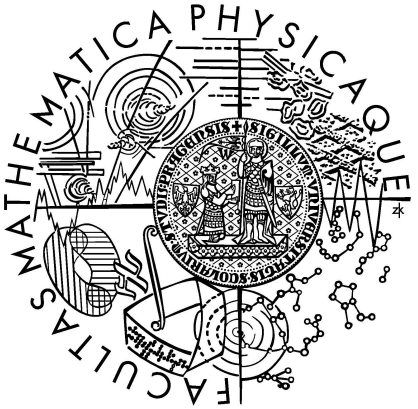
Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Josef Melcr

QM/MM calculations and classical molecular dynamics simulations of biomolecules

Institute of Physics

Supervisor of the master thesis: RNDr. Ivan Barvík, Ph.D.
Study programme: Physics
Specialization: Biophysics and chemical physics

Prague 2013

Many thanks belong to my supervisor, RNDr. Ivan Barvík, Ph.D., who showed remarkable willingness and dedication in the final stages of this work. His helpfulness during the whole span of our collaboration is acknowledged.

I am grateful to my parents, who support my studies and life all the time.

I thank my beloved girlfriend Adéla, who feeded me with strength in times I lacked it.

The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is highly appreciated.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague date ..... signature of the author

Název práce: QM/MM výpočty a klasické molekulárně-dynamické simulace biomolekul

Autor: Josef Melcr

Katedra: Fyzikální ústav Univerzity Karlovy, Matematicko-fyzikální fakulty

Vedoucí diplomové práce: RNDr. Ivan Barvík, Ph.D., Fyzikální ústav Univerzity Karlovy

Abstrakt: Byla provedena literární rešerše v oblastech proteosyntézy, MD simulací, DFT a QM/MM metod. Elongační faktor Tu (EF-Tu) je protein který se účastní proteosyntézy. Do svého aktivního místa váže GTP, jehož hydrolýza hraje důležitou roli při translaci. V nedávných simulacích bylo zaznamenáno spontánní vázání sodných iontů do aktivního místa EF-Tu Z experimentu je přitom známo, že monovalentní ionty stimulují katalytický efekt EF-Tu na GTPázovou reakci. Systematickými QM/MM výpočty byl tudíž v rámci této práce prozkoumán povrch volné energie hydrolytických reakcí methyl-di-fosfátu (ve vakuu a v implicitních vodních prostředích) a GTP v aktivním místě EF-Tu. Bylo použito extrapolační QM/MM schéma ONIOM a DFT. V souladu s experimenty byl zaznamenán slabý katalytický efekt sodíkového iontu na GTPázovou reakci. Jeho přítomnost změnila konformaci GTP interakcemi s jeho záporně nabitými kyslíkovými atomy a vyrovnala také náboje na fosfátových skupinách. To znamenalo především přenos elektronů z γ na β-fosfátovou skupinu, což je charakteristické pro mezistavy při hydrolýze GTP.

Klíčová slova: EF-Tu, GTP, ONIOM, hydrolýza, vlivy iontů

Title: QM/MM calculations and classical molecular dynamics simulations of biomolecules

Author: Josef Melcr

Department: Institute of Physics of Charles University, Faculty of Mathematics and Physics

Supervisor: RNDr. Ivan Barvík, Ph.D., Institute of Physics of Charles University

Abstract: A literature search in the fields of proteosynthesis, MD simulations, DFT and QM/MM methods was done. Elongation factor Tu (EF-Tu) is a GTP-binding protein that plays a crucial role in proteosynthesis. Spontaneous binding of sodium cations into EF-Tu active site was observed in recent simulations It's known from experiment that monovalent ions yield a moderate catalytic effect on the GTPase reaction in EF-Tu. Systematic calculations were performed to uncover the free energy surfaces for hydrolytic reactions of methyl-diphosphate (in vacuum and implicit solvents) and GTP in EF-Tu active site. Density functional theory and ONIOM extrapolative QM/MM scheme were adopted for the assay. In accordance with experiments, the catalytic effect of the sodium cation was mild. It changes the conformation of GTP attracting its negatively charged oxygen atoms. The Na^+ also equilibrates the charges of all phosphate groups of the GTP mostly by transferring electrons from γ to β-phosphate group, which is characteristic for the intermediate states during the hydrolytic reaction.

Keywords: EF-Tu, GTP, ONIOM, hydrolysis, ionic effects

Contents

1. Structure of biomolecules

Different types of biomolecules (nucleic acids, proteins, lipids, carbohydrates) can be found in living organisms. They are surrounded by water molecules and salts [1].

1.1 Nucleic acids

Nucleic acids consist of nucleotides connected together via internucleotide linkages. Nucleic acids play various roles in living cells. They store genetic information, transfer amino acids as carriers and even enzymatically catalyze chemical reactions.

1.1.1 Nucleotides

Nucleotides are the basic building blocks of nucleic acids. They consist of a purine/pyrimidine base, a sugar and a phosphate group. Nucleobases and sugars are connected via glycosidic bonds. Sugars and phosphate groups are joined via esteric bonds (Figure 1.1). A nucleotide moiety without the phosphate group is called *nucleoside*. Nucleoside mono-/di-/tri-phosphates carry 1-3 phosphate groups. So called cyclic nucleotides have a monophosphate group bridging their 3' and 5' carbons. Cyclic nucleotides often serve as signal molecules. [1]

The most common purine-based nucleobases are Adenine and Guanine. The most common pyrimidine-based nucleobases are Cytosine, Thymine and Uracil. Moreover, there are several atypical nucleobases: 5-methyl-cytidine, 5-hydroxymethyl-cytidine, inosine, pseudouridine or 4-thiouridine etc.

The sugar-part of a nucleoside is usually either ribose (RNA) or deoxy-ribose (DNA).

The phosphate group is acidic – it has a deprotonated form in common pH. Especially nucleoside triphosphates are highly charged ($4e^-$). Therefore, their negative charges are usually compensated by divalent magnesium cations Mg^{2+}.

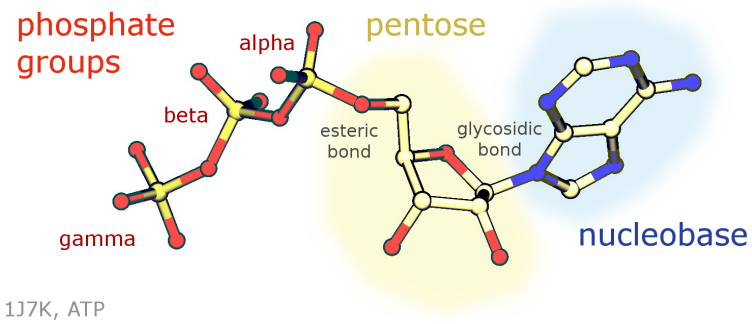


Figure 1.1: Nucleotide structure – the color scheme is as follows: phosphate group (red), sugar (yellow) and nucleobase (blue). Hydrogen atoms are missing.

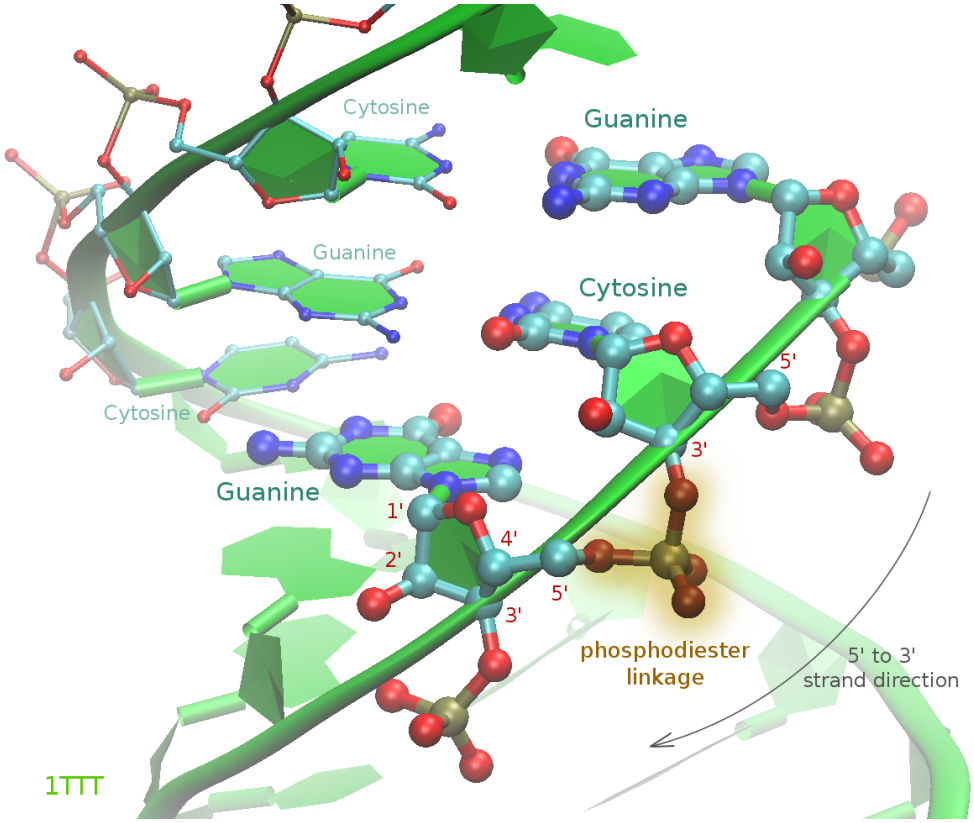


Figure 1.2: Part of a transfer ribo-nucleic acid (tRNA) structure 1TTT (taken from Protein Data Bank). Atoms are colored as follows: carbon (cyan), oxygen (red), nitrogen (blue). Hydrogen atoms are missing in crystal structures. Green ribbons are schematic representations of the tRNA backbone. Nucleotides are bound via phosphodiesteric linkages (highlighted in ochre). The smaller ball-and-stick entities in the back are nucleotides complementary to the larger ones in the front. Nucleotides should be read in the 5'–3' direction (indicated by an arrow). So the oligonucleotide sequence is GCG in the first chain and CGC in the complementary oligonucleotide. The Watson-Crick base pairing is obvious here.

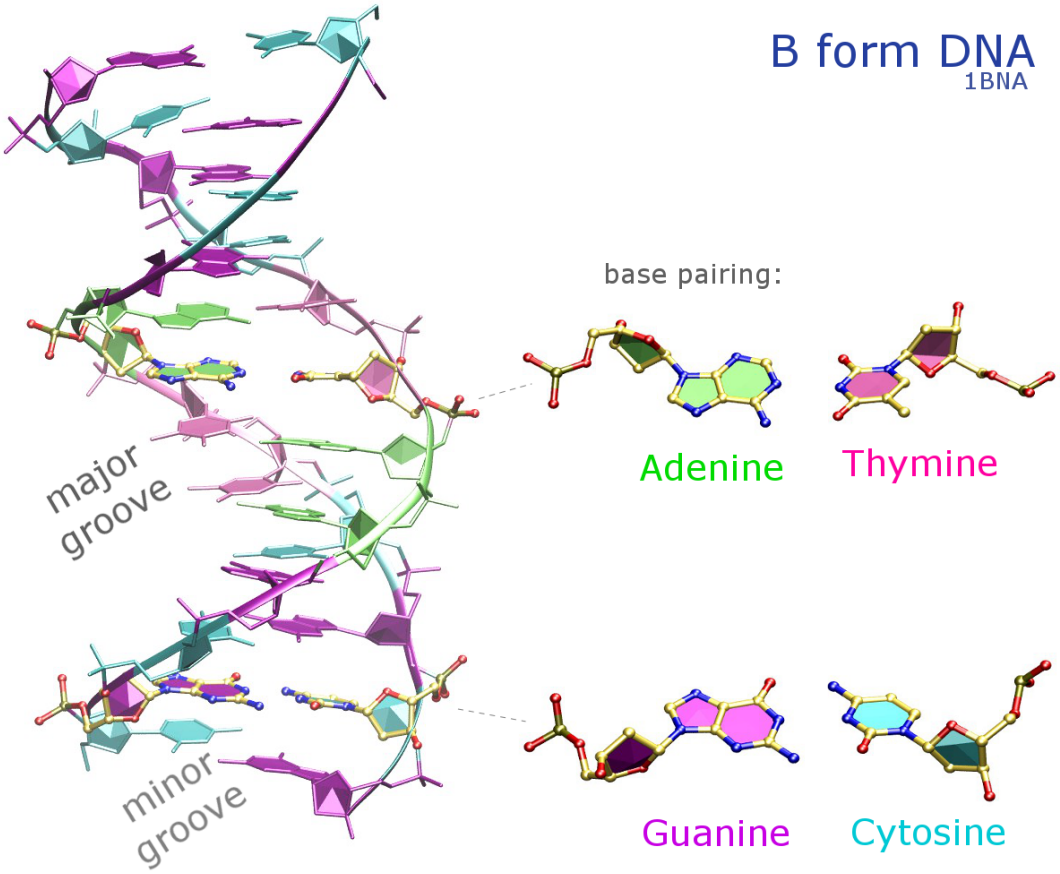


Figure 1.3: B-DNA double helical structure (left). The Watson-Crick pairing of common nucleobases (right). Hydrogen atoms are missing in crystal structures.

1.1.2 Oligonucleotides – DNA, RNA

Oligonucleotides are chains composed of nucleotides (Figure 1.2). The sequence of building blocks in an oligonucleotide is usually written in a specific direction from the 5' to 3' end. In two-stranded complexes of nucleic acids (as in DNA) nucleobases are paired. The most common is the Watson-Crick pairing Guanine–Cytosine and Adenine–Thymine (Figure 1.3). It means that a known sequence of nucleotides in one strand allows to determine nucleotides in the other strand.

Ribo-nucleic acids (RNA)

Ribo-nucleic acids consist of riboses (usually in the C3'-endo/A-form), phosphate internucleotide linkages and nucleobases (Adenine–A, Guanine–G, Cytosine–C and Uracil–U). RNAs form hairpins and take part in gene expression performing many roles. As a consequence, messenger RNA (mRNA), ribosomal RNA (rRNA) or transfer RNA (tRNA, Figure 1.2) are distinguished.

Deoxy-ribonucleic acids (DNA)

Deoxy-ribonucleic acids consist of deoxy-riboses (usually in the C2'-endo/B-form), phosphate internucleotide linkages and nucleobases (Adenine–A, Guanine–G, Cytosine–C and Thymine–T). DNA is the largest macromolecule in cells. It carries the

genetic information. Prokaryotic organisms have a circular DNA that makes a chromosome, whereas eukaryotic organisms possess a linear DNA that can make an array of chromosomes. DNA uses exclusively the Watson-Crick pairing that connects A–T and C–G. DNA folds into a double-helical structure (Figure 1.3). The most common form is the clockwise B-form. It's approximately 20Å wide and one turn is about 36 Å long. The B-form DNA possesses a minor and a major groove in its helical structure.

1.2 Proteins, peptides and amino acids

1.2.1 Amino acids

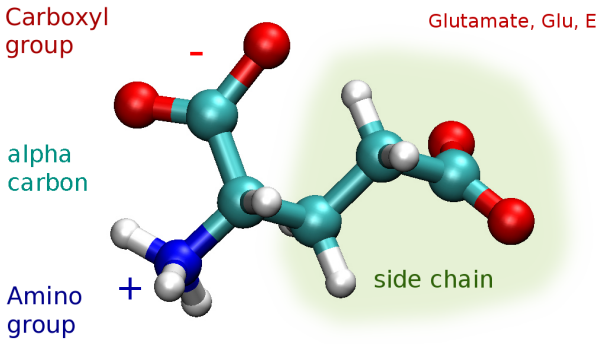


Figure 1.4: Glutamate (Glu or E) as an example of amino acid structure.

Amino acids are basic building blocks of peptides and proteins. They consist of terminal amino and carboxylic acid groups, the α-carbon atom and various side chains (Figure 1.4).

There are only 21 amino acids commonly found in proteins. According to their side chains they can be sorted in 2 groups and 5 sub-groups: *Non-polar* – aliphatic, aromatic; *Polar* – uncharged, positively or negatively charged (Figure 1.5).

The α-carbon atom in each amino acid except Glycine is chiral. Two optical isomers of amino acids can be distinguished – L and D-form. L-amino acids form proteins. D-amino acids are rather rarely found (some antibiotics or bacterial cell walls). [1]

1.2.2 Peptides

Peptides are short chains of amino acids. They are formed by making a peptide bond between a carboxylic acid group of one amino acid and an amino group of the adjacent amino acid. An example of a short peptide structure is shown in Figure 1.6.

The peptide bond has a resonance structure – with about 40% double bond character. This feature prevents free mutual rotation of surrounding groups of atoms. As a consequence peptide bonds remain planar with cis/trans isomerism. Further, peptide bonds are very stable. Enzymes, high temperatures or strong acidic or basic environment are needed for its hydrolysis.

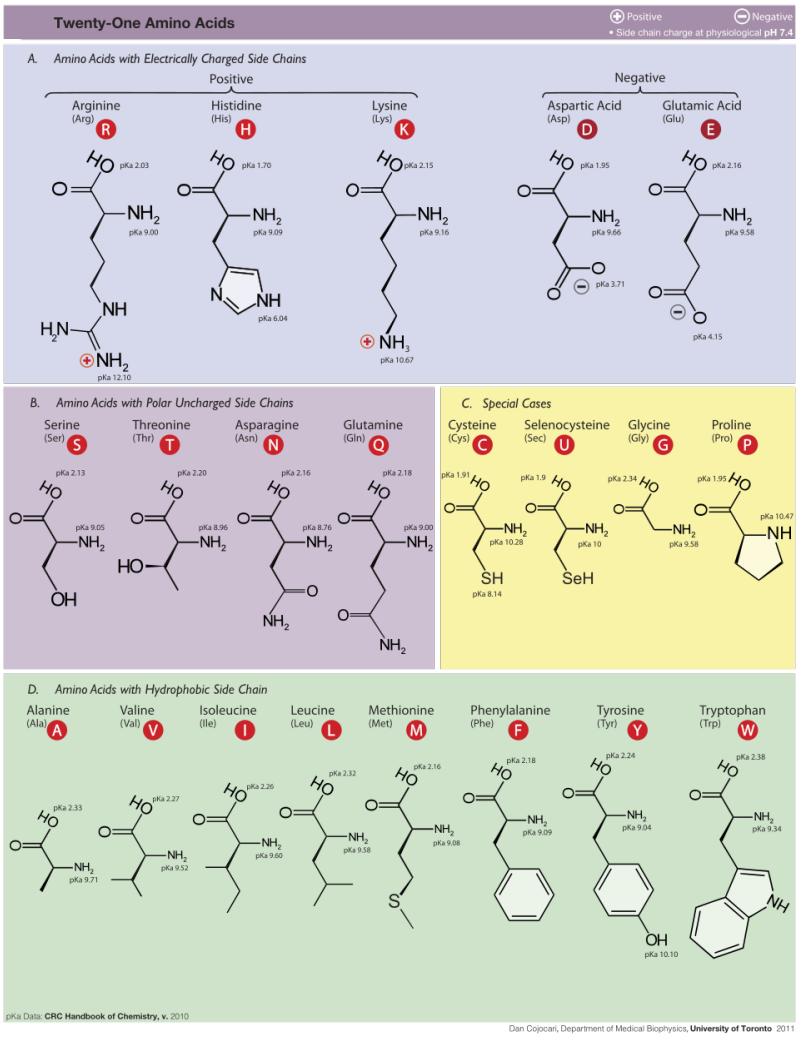


Figure 1.5: 21 basic amino acids divided into groups [2].

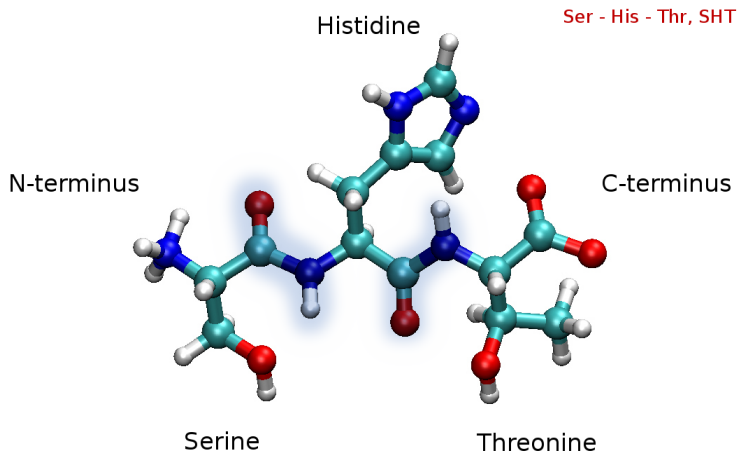


Figure 1.6: Tripeptide consisting of Serine, Histidine and Threonine amino acids. Standard abbreviations used for this molecule are Ser-His-Thr or SHT. Peptide bonds are highlighted using the light-blue clouds.

The process of peptide creation leaves only one amino and one carboxylic acid group free. The concerned residues are called N- and C-terminal (in order). Amino acids numbering always starts from the N-terminus.

1.2.3 Proteins

Proteins are long polypeptides. They are usually products of *translation* in living cells (more in section 2.2). Proteins work as enzymes, transport small molecules or mediate signaling through cell-membranes etc.

Amino acids sequence of a protein dictates its spatial structure – conformation – fold. There are four levels (primary, secondary, tertiary and quarternary) of a protein structure.

The amino acids sequence is a synonym for the primary protein structure.

The secondary protein structure describes conformations of short protein segments mainly stabilized by hydrogen bonds established between backbone amidic and carbonyl groups. An example of a helical secondary structure is depicted in Figure 1.7. As can be seen, it resembles a coil and it is usually represented as a ribbon. Side-chains of amino acids are directed outwards. There are three types of helices: α-helix, π-helix and a 3_{10}-helix. The most common is the α-helix which is distinguishable by a typical pattern: an amino group of a residue n is hydrogen bonded to a carboxyl group of a residue $n + 4$. Another common secondary structural motif is a β-sheet (Figure 1.8). Its usual representation is a broad arrow. The β-sheet moiety is formed by two neighboring protein segments that can have either a parallel or anti-parallel arrangement (N to C orientation).

The tertiary structure describes a spatial arrangement of protein parts in larger domains. The driving force for protein folding at this level is provided by hydrophobicity or hydrophilicity of amino acids. The quarternary structure describes mutual positioning and interactions of protein domains. There's no single common force that is in charge of these conformational preferences. [1]

1.3 Water and salts

Water is essential for all living organisms. The oxygen atom has a higher electronegativity than hydrogen atoms. Therefore, water molecules are polar with a dipole moment of 1.85 D (Figure 1.9). Water is an amphiprotic solvent as it can either donate or accept a proton. Thus, water molecules create hydrogen bonds with its neighbors (usually 3–4 hydrogen bonds per a water molecule).

Furthermore, water molecules are self-ionizable compounds. A water molecule can dissociate according to this chemical equation:

$$\mathrm{H_2O} + \mathrm{H_2O} \longrightarrow H_3O^+ + \mathrm{OH}^- \tag{1.1}$$

thus creating a cation/anion pair. The concentrations of these ions determine overall acidity (or basicity) of solutions (crucial for living organisms). It's usually described by a pH (or pOH) scale which is defined with

$$\mathrm{pH} = -\log\left[\mathrm{H_3O}^+\right] = 14 - \mathrm{pOH} \tag{1.2}$$

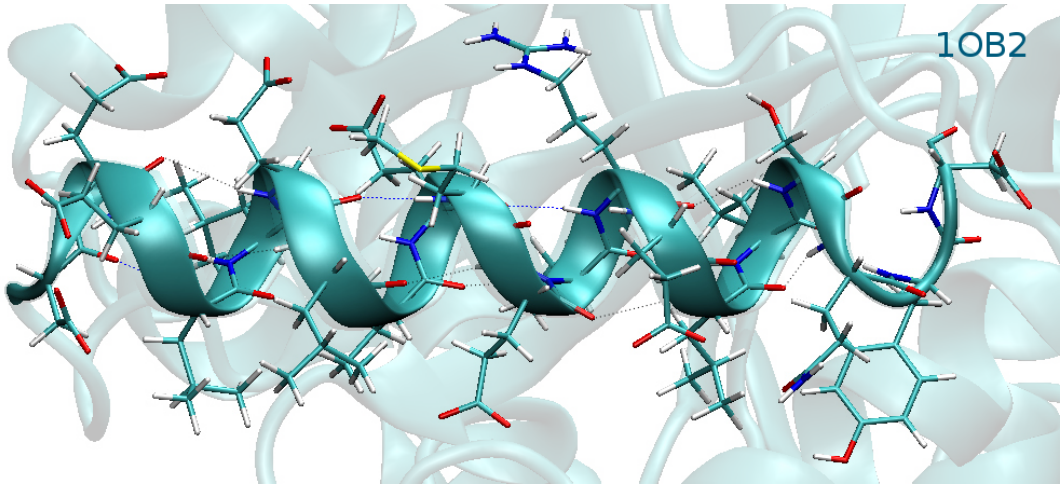


Figure 1.7: α-helical part of the 1OB2 structure represented as a ribbon. Hydrogen bonds are highlighted with dotted lines. Side-chains of amino acids are always directed outwards.

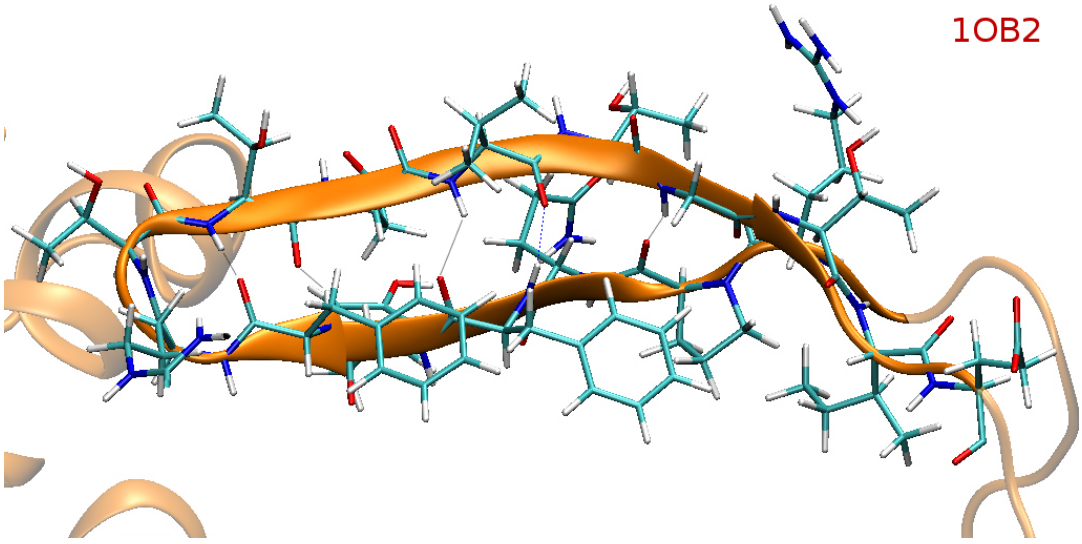


Figure 1.8: An example of an anti-parallel β-sheet structure represented using broad arrows. Hydrogen bonds are highlighted with dotted lines. Based on the 1OB2 crystal structure.

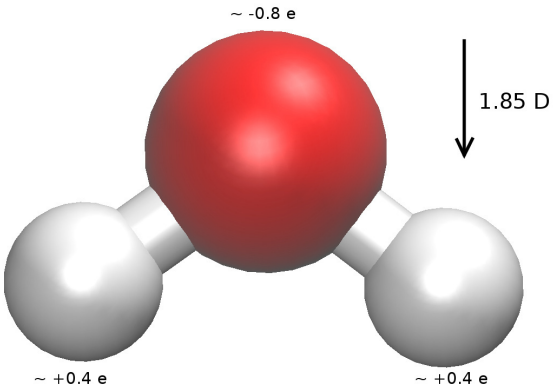


Figure 1.9: Water molecule depicted using a ball-and-stick representation. Dipole moment orientation is indicated by an arrow. Partial charges on atoms are shown as well.

where the square brackets "$[H_3O^+]$" denote the concentration of oxonium cations. In a pure water pH=7. Acidic solutions have lower and basic solutions higher pH than that.

Under physiological conditions, the solvent contains various diluted salts in their ionic form. The most common ions are Na^+, K^+, Cl^-, Ca^{2+} and Mg^{2+}. Their concentrations are essential for living organisms as it creates an osmotic pressure and establishes membrane potential. Most cells maintain their own equilibrium concentrations of different ions (usually about 0.1-0.2 M). [1]

2. Proteosynthesis

Genetic information stored in DNA is copied into mRNA during transcription. The mRNA travels from the cell nucleus into cytoplasm. Triplets of nucleobases in mRNA (codons) are recognized by complementary triplets of nucleobases in tRNA (anticodons). tRNAs carry amino acids which are joined together in peptidyl transferase centres (PTC) of ribosomes (Figures 2.1 and 2.4). Thus, the sequence of amino acids in the synthesized protein is determined by the Watson-Crick pairing between codons and anti-codons [3].

There are four different nucleotides in RNA, which give $4^3 = 64$ possible combinations of nucleotides in codons. Nevertheless, there are only 20 amino acids that are coded by these codons. Therefore, multiple codons can be usually translated into the same amino acid (Figure 2.2). [4]

2.1 Ribosome

Ribosomes catalyze proteosynthesis (translation of genes into proteins). Approximately 22 amino acids per second are incorporated into a new born polypeptide chain. This process is highly accurate. On average, mistakes occur only every 10^3–10^4 step [6].

Ribosomes were discovered in 1955. The atomistic structure was deciphered in 2000 by scientific groups lead by V. Ramakrishnan, T. Steitz and A. Yonath using the X-ray crystallography.

The ribosome of *Thermus Thermophilus* bacteria is depicted in Figure 2.1. Ribosomes consist of a large 50S and a small 30S subunit (S is a Svedberg unit for sedimentation in a centrifuge). Both subunits exist separately in the inactive ribosome state. There are A site which binds incoming aminoacylated tRNAs, P site where peptide transfer occurs and E site where the deacylated tRNA leaves the ribosome [7]. In the vicinity of the A site, there is an important evolutionary conserved sarcin-ricin loop (Figure 2.3), on which elongation factors Tu and G bind ([8] and [9]). It's named after cytotoxines sarcin and ricin that cut a single bond in the sarcin-ricin loop making the ribosome unable to interact properly with the elongation factors and disabling translation [4].

2.2 Translation

Apart from the ribosome, there are many other molecules that participate in various phases (initiation, elongation, termination) of proteosynthesis (Figure 2.5, bacterial translation diagram).

2.2.1 Initiation

The initiation phase of translation starts with ribosomal subunits (30S and 50S) separated [7]. A probable first step is the binding of IF3 to 30S. The Shine-Dalgarno sequence of mRNA recognizes the complementary anti-Shine-Dalgarno sequence of ribosomal RNA (Figure 2.6). The binding of the 30S-IF3-mRNA

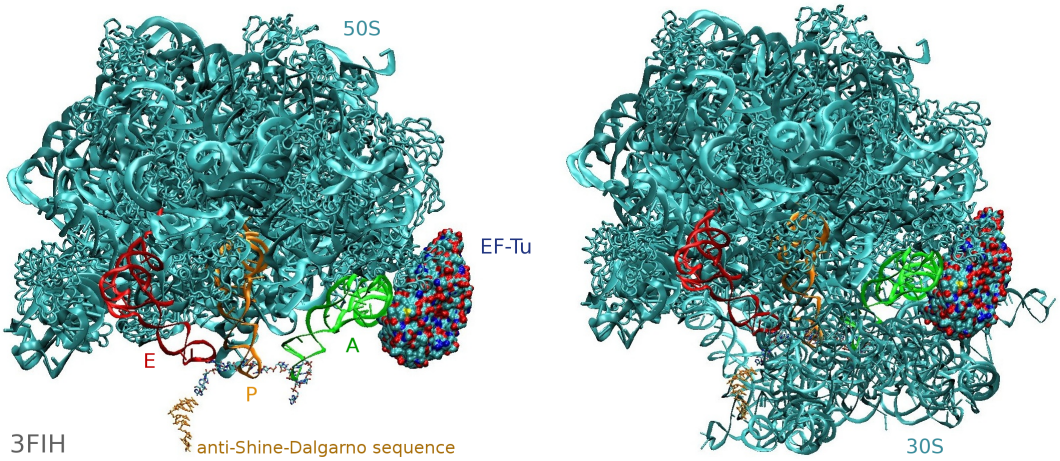


Figure 2.1: The ribosome with tRNAs bound in its A (green), P (orange) and E (red) sites. On the left picture only the large 50S subunit is shown, which renders the short mRNA strand (various colors according to the atomic names) and the anti-Shine-Dalgarno sequence (orange) visible. The structure near A site tRNA is the EF-Tu. The PDB code of this structure is 3FIH.

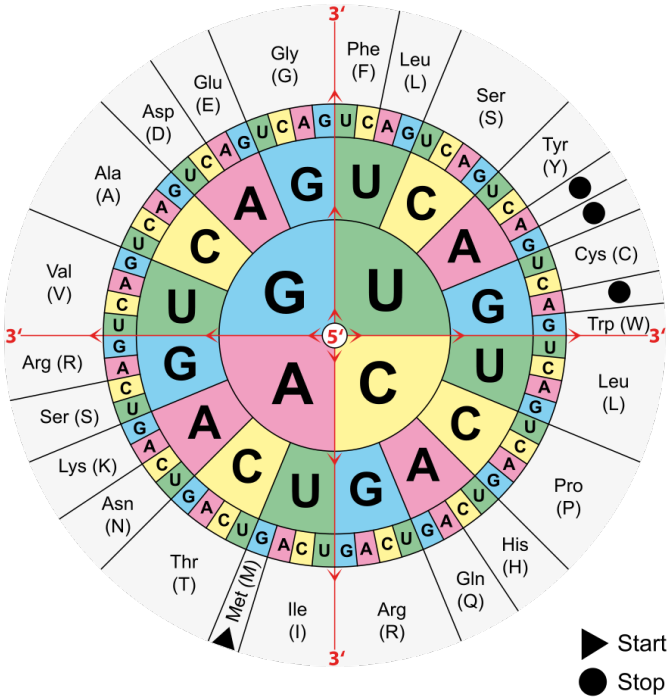


Figure 2.2: Genetic code. Start and stop codons are denoted by corresponding symbols. The codon triplet sequence goes outwards from the center in the direction of red arrows. The third position is obviously the most variable, thus called "wobble position". Image taken from [5].

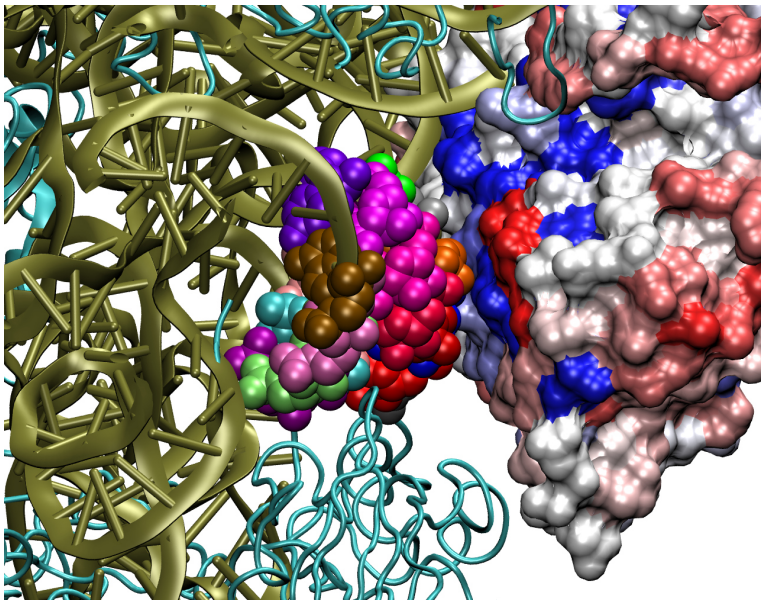


Figure 2.3: EF-Tu (blue-white-red structure to the right) bonded to the ribosome (khaki helices depict nucleic acids, light-blue strands represent proteins). In the middle there is the sarcin-ricin loop highlighted in various colors according to the residue numbers. It binds to EF-Tu closely to its active site. The both of these regions, sarcin-ricin loop and the active site of EF-Tu, are highly conserved among bacteria suggesting their importance [4]. The conserved sites on EF-Tu are blue colored, whereas the divergent are red.

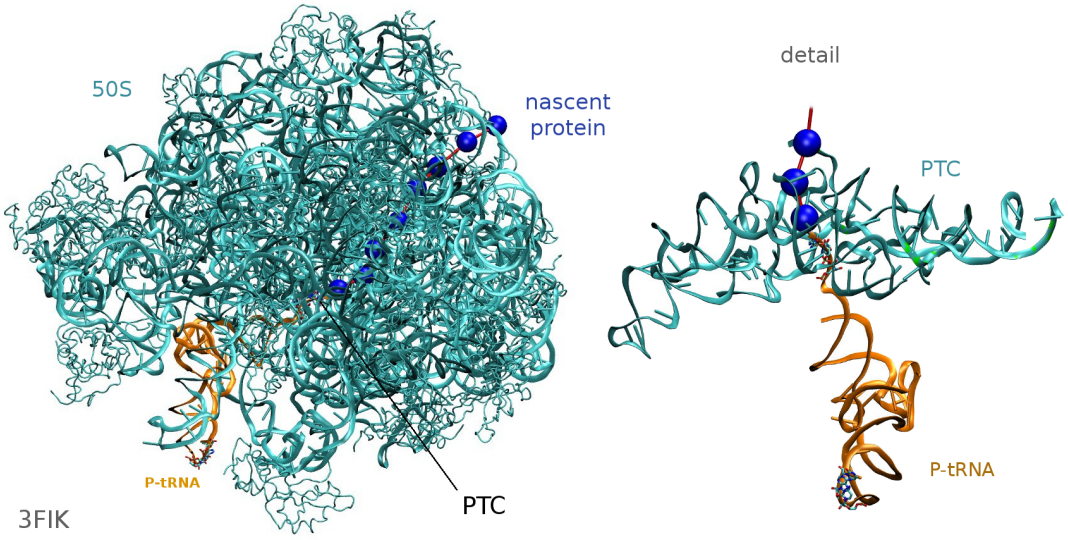


Figure 2.4: The peptidyl transfer center (PTC) of the ribosome. Only the large 50S subunit of the ribosome is shown. The nascent protein is depicted only schematically so that its route through the ribosome could be easily seen. In the right picture a detail of the PTC is shown with aminoacylated P site tRNA. The PDB code for this structure is 3FIK.

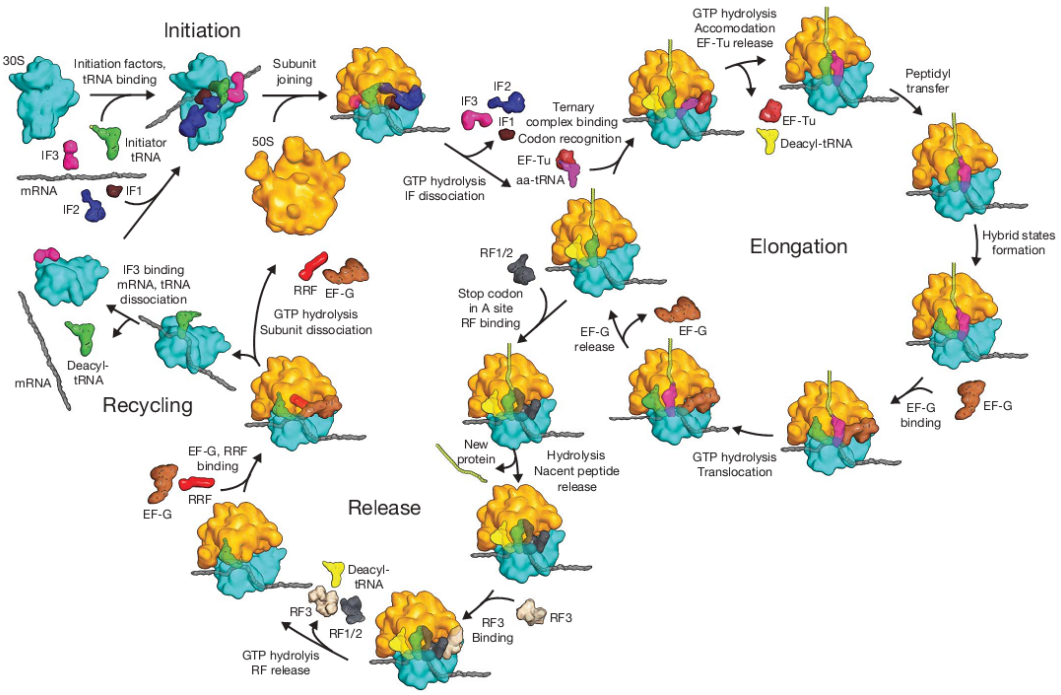


Figure 2.5: Bacterial translation. All intermediate structures are not shown for simplicity. Abbreviations used: 30S and 50S are the small and the large subunit of the ribosome; IF is an initiation factor; aa-tRNA stays for aminoacyl-tRNA; EF is an elongation factor; RF is a release factor; RRF is the ribosome recycling factor. The translation begins with *initiation*, the two ribosomal subunits, 30S and 50S, join together and with the help of IFs the first amino acid is brought to the peptidyl-transferase center of the ribosome. Then the translation proceeds with *elongation*, during which the protein is elongated in presence of aa-tRNAs, EF-Tu and EF-G. When it is a turn of the terminating sequence on mRNA, the stop codon, the ribosome binds release factors that free the nascent protein and the ribosomal subunits dissociate again. Image taken from [7].

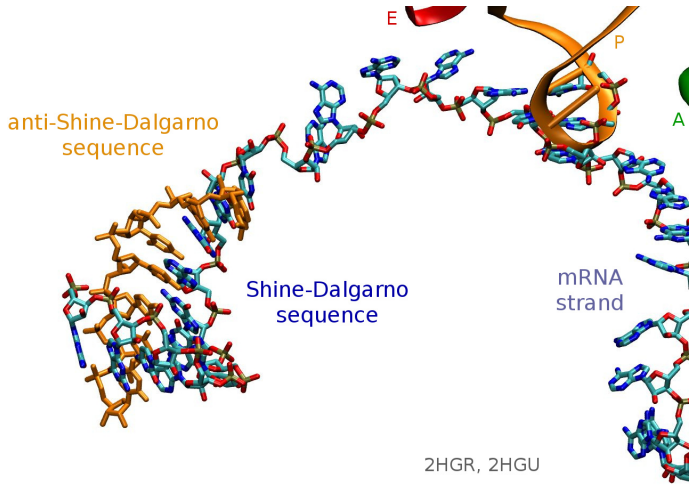


Figure 2.6: Shine-Dalgarno sequence on mRNA and an anti-Shine-Dalgarno sequence on the RNA part of the 30S subunit. The perfect base pairing is apparent. Other parts of the ribosome are not shown for simplicity. The position of the anti-Shine-Dalgarno sequence can be seen in the picture 2.1. Only a short strand of mRNA was included in the crystal structures 2HGR and 2HGU. The green, orange and red ribbons in the top right corner are the cartoon representations of A, P and E site tRNAs respectively.

complex to IF1, IF2 and initiator tRNA results in the 30S initiation complex (30S-IC). The IF2 GTPase promotes joining of ribosomal subunits that form the 70S initiation complex (70S-IC). IF3 is released. IF2 hydrolyzes GTP and releases an inorganic phosphate. The initiating fMet-tRNA$^{\mathrm{fMet}}$ moves into PTC.

2.2.2 Elongation

Amino acids are repeatedly added to a growing polypeptide chain in the elongation phase of translation (Figure 2.5). Initially, the ribosomal P site is occupied by the initiating tRNA that carries the first amino acid. The A site is empty and prepared for binding of ternary complexes composed of EF-Tu, GTP and aminoacyl-tRNA with an anti-codon complementary to the second codon of mRNA. If the codon-anticodon recognition fails, the ternary complex dissociates from the A site. If the codon and anticodon are complementary, tRNA undergoes conformational changes that stimulate EF-Tu to hydrolyze its GTP substrate. With that EF-Tu releases tRNA that moves its aminoacylated tip to PTC (Figure 2.4). The α-amino group of the to-be-added amino acid makes a nucleophilic attack on the ester carbon of the last-in-chain amino acid that is still bound to tRNA in the P site. The exact mechanism of the catalysis is not still fully understood [7].

After that, translocation takes place. It moves all tRNAs bound to a ribosome one step further. Since the movement is rather large, about 50 Å, it requires the help of EF-G [7]. P site tRNA moves to the E site. A site tRNA moves to the P site. A site becomes empty allowing another ternary complex to bind. The elongation cycle repeats again and again until it reaches the stop codon on mRNA.

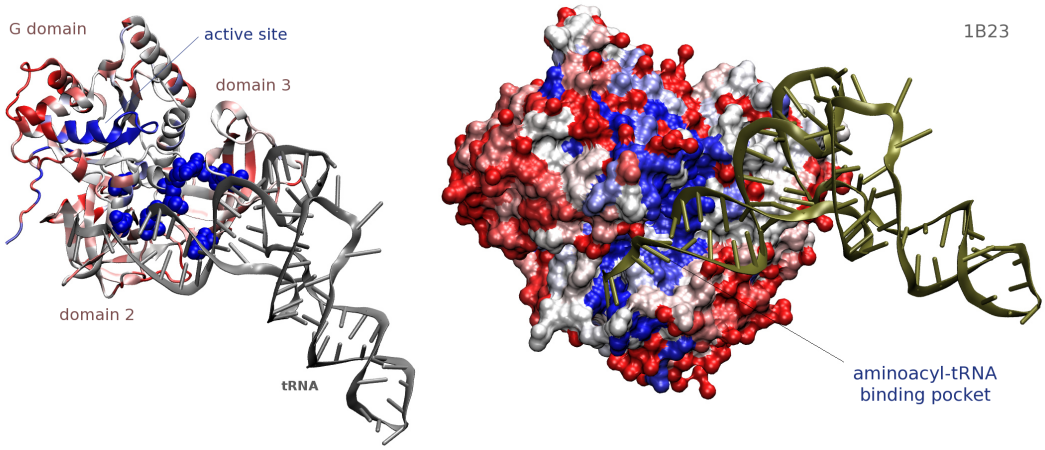


Figure 2.7: The ternary complex of EF-Tu, GTP and aminoacylated tRNA. Sequence identity of EF-Tu of different bacteria is indicated with different colors: blue marks the conserved whereas red the non-conserved residues. There are two notable highly conserved areas: the active site and the aminoacyl-tRNA binding pocket. Both molecules have the same orientation but different representations. The PDB code of this structure is 1B23.

2.2.3 Termination

There are several stop codon triplets: UAA, UAG and UGA (Figure 2.2). Release factors detach P site tRNA from the polypeptide chain which is released to cytosol. Ribosome recycling factor then stimulates separation of ribosomal subunits [7].

2.3 EF-Tu

Elongation factor Tu (EF-Tu) delivers aminoacylated tRNAs (aa-tRNAs) to the ribosome A-site (Figure 2.1) [4]. The so called ternary complex is composed of aa-tRNA, EF-Tu and GTP. EF-Tu shields the ester bond formed between amino acids and tRNAs in the so called aa-tRNA binding pocket. If the mRNA codon and tRNA anticodon triplets are fully complementary and the Watson-Crick base pairing is properly formed, the corresponding tRNA undergoes a conformational change that stimulates EF-Tu to hydrolyze its GTP substrate. With that EF-Tu releases tRNA that moves its aminoacylated tip to PTC (Figure 2.4). In this way EF-Tu helps to secure a remarkable translational fidelity ([6], [3] and [4]). Reactivation of EF-Tu is done by other factor, EF-Ts, that catalyses unbinding of the GDP allowing another GTP to slip in.

EF-Tu consists of 3 domains. The biggest one is called the G domain (Figure 2.7). There are highly conserved parts especially along the active site where either GTP or GDP is usually bound. EF-Tu can be found in the active or inactive form (Figure 2.8) depending on whether GTP or GDP is bound in the active site (Figure 2.7) [3]. There are rather large conformational changes in the G-domain. In contrast, the domains 2 and 3 remain more or less rigid during the conformational transitions between the active–inactive forms of EF-Tu. However, both domains move as rigid bodies with respect to the G domain remarkably.

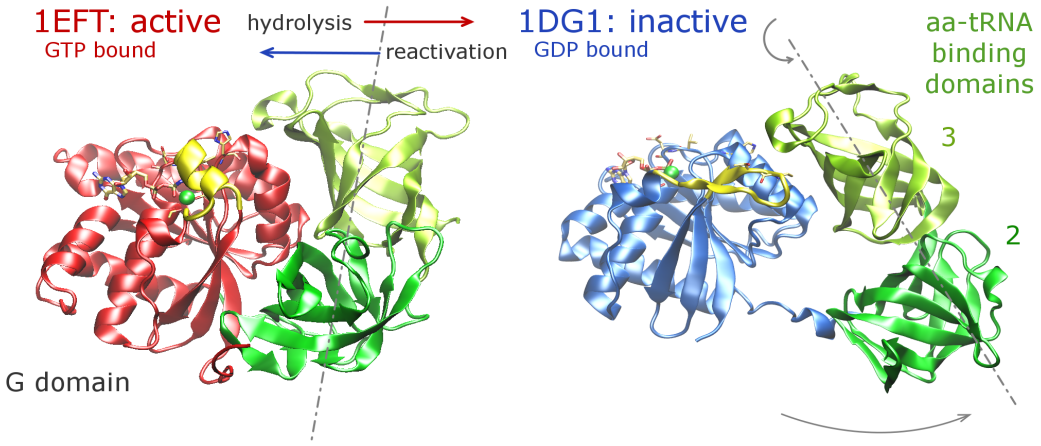


Figure 2.8: The active, GTP bound, and inactive, GDP bound conformations of EF-Tu. The aa-tRNA binding domains undergo conformational changes that unfold them from the G domain and rotate them in the direction shown. Apart from this change the switch I (sw1) α-helix turns into a β-sheet (yellow). The GTP/GDP substrate is shown with several nearby residues.

The architecture of the GTP-binding domain is similar in all translational GTPases from bacteria to higher eukaryotes. The GTP hydrolysis is accompanied by conformational changes in the conserved switch I and II regions. Moreover, translational GTPases bind to the same region on ribosomes in all species. The conserved binding site and structural similarities suggest that there is a common mechanism by which the ribosome activates the GTP hydrolysis. However, despite more than 40 years of research, this mechanism remains elusive. [10]

There are several loops around the EF-Tu active site: P loop (res. 18-23), Switch I (sw1, res. 40-62) and Switch II (sw2, res. 80-100) [8]. Switch II contains the His85 residue that is believed to serve as a general base in the GTPase reaction [10]. Val20 and Ile61 create the so called *Hydrophobic gate*, which was believed to have an effect on the EF-Tu GTPase reaction [8]. But this has been challenged in [10]. MD simulations presented in my bachelor thesis [11] revealed no correlation between the opening of the Hydrophobic gate and penetration of His85 into the active site as well.

The GTP substrate has a charge of $-4e^-$ in common pH. The three phosphate groups are stabilized in the EF-Tu active site by Mg^{2+} and interactions with residues of the P loop (res. 22-25). The γ-phosphate group of GTP interacts with Thr62 from Switch I and through the catalytic water with His85 from Switch II [10].

The role of His85 is a subject of spirited discussions. His85 is undoubtedly a catalytic residue as it has a strong effect on the GTPase activity of EF-Tu [12]. Point mutations of this residue ([13], [14]) support this idea. Another study showed that the EF-Tu GTPase reaction is pH independent [15]. It implies that His85 doesn't act as a general base during the GTPase reaction. Its catalytic effect resides in stabilization of the transition state. A recent study [10] suggests that His85 is pushed into the active site by the sarcin-ricin loop – a highly conserved part of the ribosome (section 2.1 and Figure 2.3).

Attempts to find alternative residues that would serve as a general base in the

EF-Tu GTPase reaction were not successful [16]–[19]. However, a computational study of closely related enzyme Ras p21 [20] suggests that GTP itself could serve as a general base.

2.3.1 Spontaneous binding of sodium ions into the EF-Tu active site

Active sites of GTPases resemble the palm of a hand, where the Thumb stabilizes the catalytic water molecule and the electrophilic Finger modulates redistribution of the charge between separating beta and gamma phosphate groups of GTP.

G-proteins possess some intrinsic GTPase activity. Much more significant GTPase effect is usually reached, if a second helping hand (the so-called G-protein Assisting Protein – GAP) is engaged into the GTP hydrolysis. The Thumb and Finger can be either positioned by GAP to interact optimally with GTP (i.e. endogenous fingers localized on GTPases), or even supplied by GAP into the reaction (i.e. exogenous fingers).

The Thumb and Finger are usually side chains of amino acids, but water molecules and ions can be exploited as well, if the GTPase is able to create an appropriate binding site for them. A number of amino acids can serve as the Thumb stabilizing the catalytic water molecule (Gln [22], [23], [24], His, Asn [25], [26], [27], [28] – Figure 2.9 A), C)). If the Thumb is retracted, the second bridging water molecule can stabilize the catalytic water molecule (Wb+Asp [21], Wb+Glu [29]). The Finger modulating the charge redistribution between the beta and leaving-gamma phosphate groups of GTP is usually a positively charged arginine residue (Arg [22], [23], Figure 2.9 A)). But sometimes a relevant biological activity is observed when the GTPase has in this position just a suitable binding site for positively charged monovalent ions ([30], [31], Figure 2.9 B)).

Elongation factor Tu (EF-Tu) is a translational GTPase that delivers aminoacyl-tRNAs to the ribosomal decoding site. EF-Tu itself has a very low intrinsic GTPase activity, which can be stimulated by positively charged monovalent ions ([32] and [33]) to reach up to activities recorded for the ion stimulated HAS GTPases that are used to explain their biological functions ([21], [29]–[32]).

In my bachelor thesis [11], MD simulations of the 1EFT and 2XQD structures of EF-Tu in water and salt solution were produced. Spontaneous binding of sodium ions into the EF-Tu active site was observed in both cases. Even two ions were settled there in the 2XQD simulation (Figure 2.10).

Generally, sodium ions settled in the position of a water molecule that cross-bridges two residues, Asp21 and Gly60, in many crystal structures of EF-Tu. Those residues are neighbours to Val20 and Ile61 forming the so called Hydrophobic gate mentioned above. Binding of sodium ions into the active site resulted in narrowing of the Hydrophobic gate. Further, the sodium ions stimulated penetration of His85 into the EF-Tu active site, which was independent of the width of the Hydrophobic gate.

A similar observation was reported in a QM/MM study of the Ras-GAP GTP-binding protein. During equilibration MD runs Na^+ ions diffused quickly towards the active site where they were settled [65].

In MD simulations following-up on my bachelor thesis, one of crystallographic waters inside Switch I of EF-Tu was replaced by a monovalent cation (Ion_{sw}) inter-

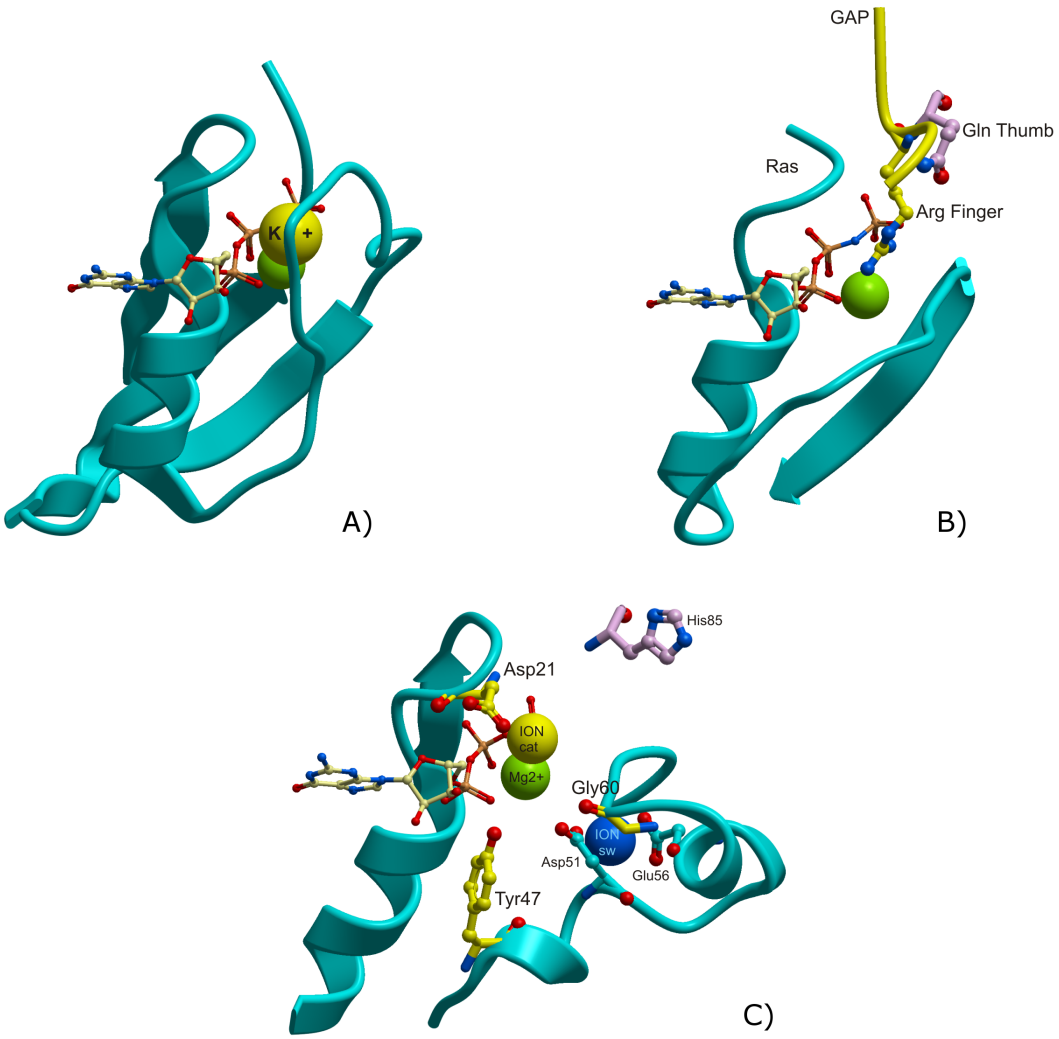


Figure 2.9: Catalytic machinery in active sites of Ras, FeoB and EF-Tu GTPases. A) Ras protein interacts with a GAP, which stimulates hydrolysis by supplying a catalytic residue, an "arginine finger", into the Ras active site ([22] and [23]). This stabilizes the endogenous "glutamine thumb" which in turn stabilizes the water molecule for the in-line attack while the arginine neutralizes a negative charge of β and γ-phosphate oxygens in the transition state. B) Sometimes a GTPase has just a binding site for positively charged monovalent ions instead of an arginine finger. Even this can lead to a relevant biological activity ([30] and [31]). C) In our MD simulations, one of the crystallographic waters inside the Switch I of EF-Tu was replaced by a monovalent cation ($\mathrm{Ion_{sw}}$, depicted as dark blue sphere) interacting with Asp51 and Glu56 side chains. $\mathrm{Ion_{sw}}$ was firmly bound in this position in both Phe47 as well as Tyr47 MD simulations lasting for 500 ns. Interestingly, the stabilization of Switch I achieved using $\mathrm{Ion_{sw}}$ affected positively spontaneous binding of other monovalent ions ($\mathrm{Ion_{cat}}$, depicted as yellow sphere) into the active site of EF-Tu. $\mathrm{Ion_{cat}}$ was bound in the vicinity of the scissile bond between the β and γ-phosphate groups of GTP. $\mathrm{Ion_{cat}}$ interacts with the side chain of Asp21 and main chain carbonyl group of Gly60. Occassional direct interactions with the Tyr47 side chain were found as well. Nevertheless, Tyr47 anchored to the α-phosphate group of GTP provides much more important services for $\mathrm{Ion_{cat}}$ indirectly, through the stabilization of Switch I (including Gly60).

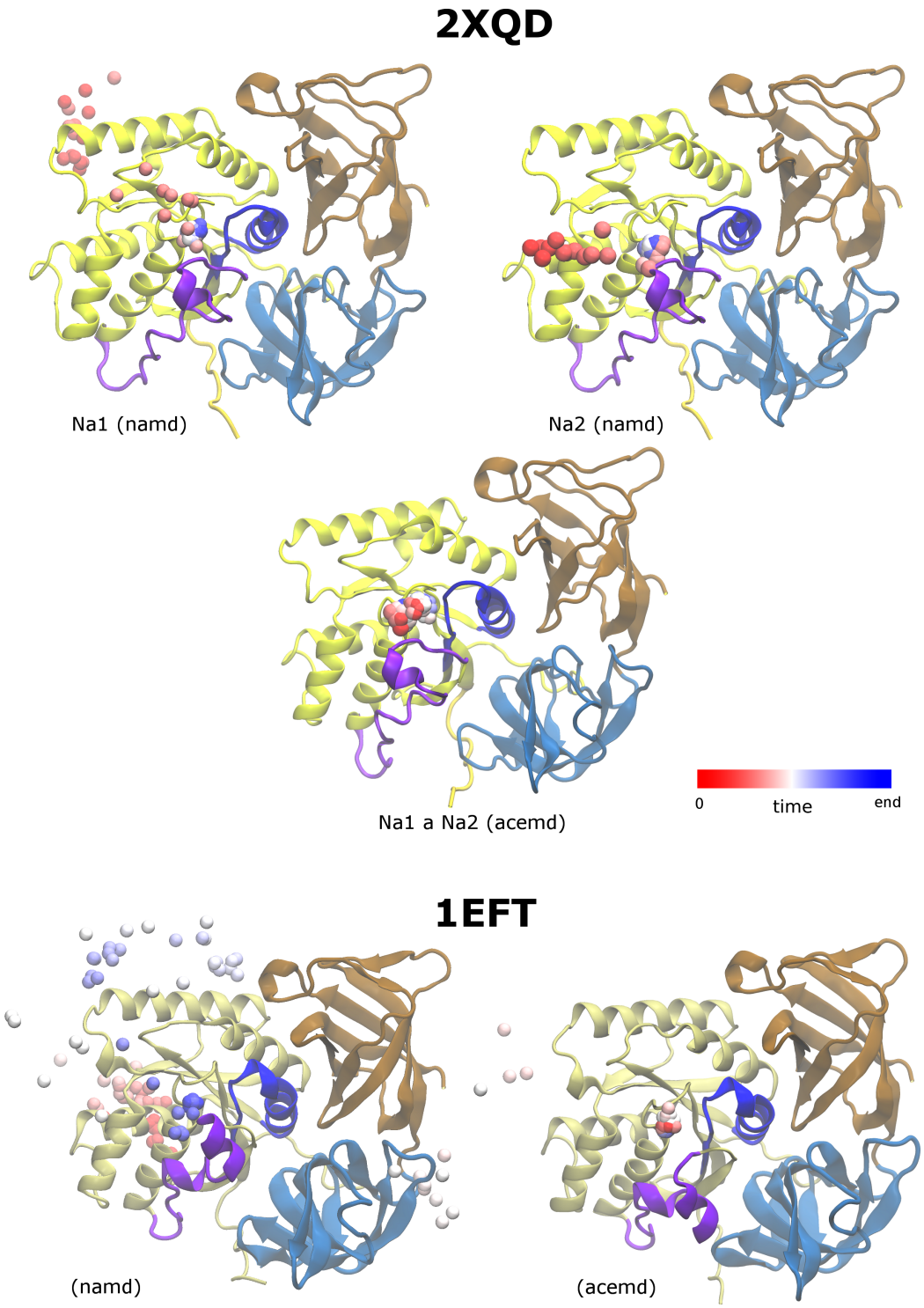


Figure 2.10: Spontaneous binding of sodium ions into the active site of EF-Tu as observed in MD simulations [11]. Even two sodium ions were settled in the active site of 2XQD. Time evolution is indicated by a color change; simulations had different lengths. Words "namd" and "acemd" refer to specific parts of MD simulations. In this particular case, "namd" preceded "acemd". Domains are distinguished by colors, Switches I and II and the P loop are also highlighted.

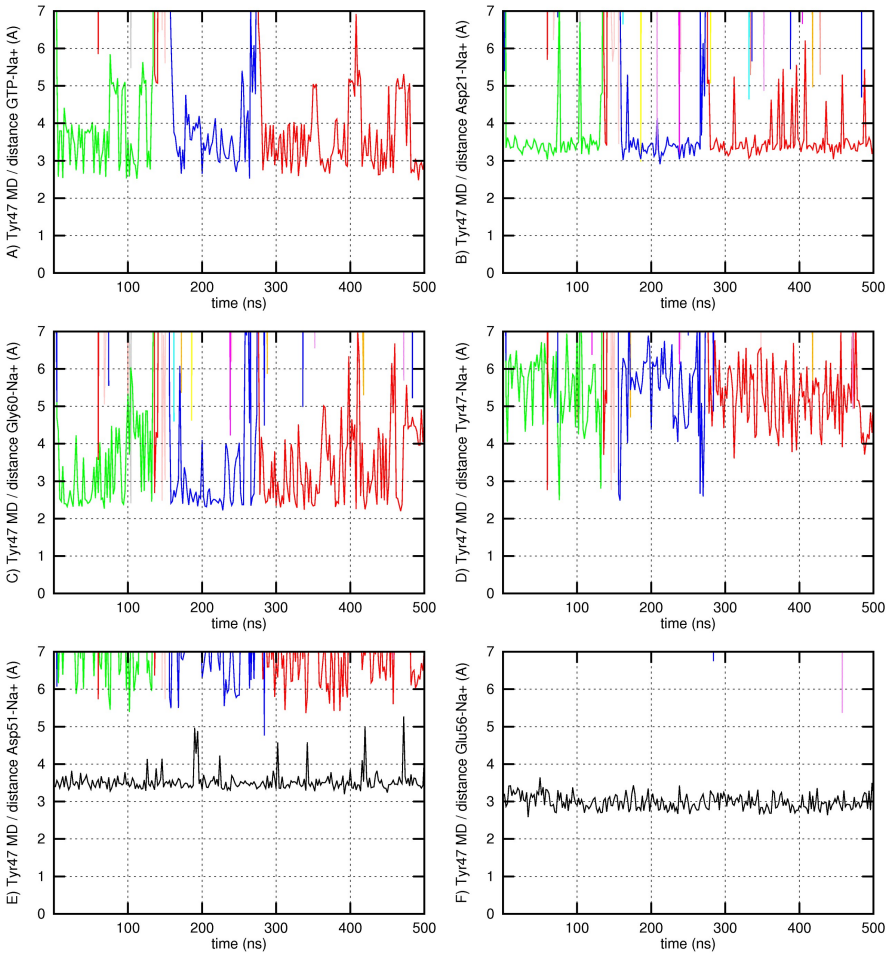


Figure 2.11: Spontaneous binding of Na^+ ions into the EF-Tu active site

acting with Asp51 and Glu56 side chains (Figure 2.9 C)). Ion_{sw} was firmly bound in this position in MD simulations lasting for 500 ns (Figure 2.11, Graph E–F). Interestingly, the Switch I stabilization achieved using Ion_{sw} affected positively repetitive spontaneous binding of other monovalent ions (Ion_{cat}, depicted as yellow sphere in Figure 2.9 C)) into the active site of EF-Tu into the vicinity of the scissile bond between the beta and gamma phosphate groups of GTP (Figure 2.11, Graph A). Ion_{cat} interacts with the side chain of Asp21 (Figure 2.11, Graph B) and main chain carbonyl group of Gly60 (Graph C). Occassional direct interactions with the Tyr47 side chain were found as well (Figure 2.11, Graph D). These spontaneous binding events of Ion_{cat} were considerably long-lasting (there interchanged only 3 ions within 500 ns - Figure 2.11, Graph A).

3. Molecular dynamics simulations

Molecular dynamics simulations (MD) represent a branch of computational methodologies that can be described as theoretical experiments. By running a computer simulation with subsequent application of statistical analysis, distributions and mean values of physical properties can be obtained. These methods were traditionally used in conjunction with experiments to help their interpretation and to test the underlaying theories.

Computer simulations have apparent advantages over experiments: they can be used in cases when the experiment would be hard to accomplish; the resolution in time and space is easily reached. An evident disadvantage and an inherent feature of theirs is, however, that simulations are always only a theoretical model of reality.

3.1 MD approximations, interaction potentials

For an evolution of a system equations of motion must be known. Quantum systems like molecules require the usage of Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi \tag{3.1}$$

where i is the imaginary unit, $\hbar$ is the reduced Planck constant, Ψ is the wave function and $\hat{H}$ is the Hamiltonian operator. The wave function Ψ contains the most complete description of the whole system [36]. In a general case it contains electrons as well as nuclei.

For time-independent potentials, the wave function can be split into a spacial part and a time part. The separation allows the usage of the time-independent Schrödinger equation

$$H\Psi = E\Psi \tag{3.2}$$

which is an eigenproblem for the Hamiltonian operator, where the eigenvalues E have a physical meaning as energies [37].

The Schrödinger equation has a many-body character. It would be much easier to solve if the number of particles that are considered in computation could be lowered. A common approach to achieve that is to use the Born-Oppenheimer approximation, which allows separation of nuclei from electrons splitting the equation in two. It assumes that electrons move in the field of fixed nuclei, which is possible, for nuclei are orders of magnitude heavier and slower than electrons. Next, the nuclei move in an effective field formed by the electrons. Such approximation renders good enough results even for a hydrogen, and it gets better with increasing mass of atoms [38].

The ground state of a given system is studied most often, and the assumptions for the usage of the time-independent Schrödinger equation are also valid in most cases. At this level, it's possible to define an **interaction potential** V for the atoms with the nucleic coordinates as parameters. The interaction potential is

an approximation to the effective potential formed by the electrons from the Born-Oppenheimer approximation. Since the atoms as whole units don't require quantum treatment, classical Newton's equations of motion can be applied. An MD simulation of such system would than represent an approximate solution of the Schrödinger equation 3.1, where the forces F exerted on atoms are calculated as

$$F = -\nabla V \tag{3.3}$$

3.1.1 Types of interaction potentials

The interaction potential V describes forces on atoms. Depending on the way it was obtained, several types are distinguished [39]:

Full PES (potential energy surface) is the most demanding method as it requires an ab-initio solution of V for the whole span of nucleic parameters in advance. The usage of this method is rare and it virtually can be calculated only for small systems.

On-the-fly potential. The interaction potential V is evaluated only for the nucleic coordinates at the current time using ab-initio methods. It is usually used in conjunction with DFT (density functional theory), and depending on the available computational power it can be used for hundreds or even thousands of atoms. It may be also called ab-initio MD.

Force field is an empirical potential. It can be obtained by fitting a particular mathematical formula to an experiment or ab-initio calculations. It is the fastest method yet the most inaccurate.

3.2 Classical force fields

The core of a force field lays in the definition of the empirical formula for the interaction potential. The mathematical form varies among different types of force fields; the harmonic potential is the most common, though, as it is simple and it usually describes the potential in equilibrius position well. By fitting the formula to experimental or ab-initio data, parameters are obtained for the various terms. Force fields for smaller systems (for example salts dissolved in water) can be more precise than those for larger systems, where the formula must be more general due to the high number of types of interactions [37].

Classical force fields like Amber, Charmm, OPLS and UFF are pairwise additive. For example Amber force field uses a simple empirical formula [40]

$$V = \sum_{bonds} K_b(r - r_{eq})^2 + \sum_{angles} K_\Theta(\Theta - \Theta_{eq})^2 + \tag{3.4}$$
$$\sum_{dihedrals} \frac{1}{2} V_n(1 + \cos(n\phi - \delta)) + \sum_{i<j} \left[s_{ij}^{VdW} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + s_{ij}^{q} \frac{q_i q_j}{\epsilon\, r_{ij}} \right]$$

Charmm force field, for example, has a slightly more complex form [41]:

$$
\begin{aligned}
V = & \sum_{bonds} K_b(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{impropers} K_\psi(\psi - \psi_0)^2 \\
& + \sum_{dihedrals} K_\phi(1 + \cos(n\phi - \delta))^2 + \sum_{Urey-Bradley} K_{UB}(r_{1,3,ik} - r_{1,3,0})^2 \\
& + \sum_{VdW} \epsilon_{ij} \left[(\tfrac{R_{min,ij}}{r_{ij}})^1 2 - 2(\tfrac{R_{min,ij}}{r_{ij}})^6 \right] + \sum_{electrostatics} \tfrac{q_i q_j}{r_{ij}}
\end{aligned}
\tag{3.5}
$$

The terms will be explained in the subsequent sections.

Non-bonded interaction are usually evaluated only for atoms separated by more than 3 bonds, as such interactions are already included in the bond stretching and angle bending terms. The number of non-bonded terms scales approximately quadratically with the size of the system taking the most of the time of evaluation of the interaction potential.

3.2.1 Bonding terms

Internal coordinates allow us to use the local symmetry of molecules simplifying the force field formula. Common terms that appear in force fields for biophysical applications are [37]:

Chemical bond stretching is usually described by a harmonic potential $V = K_b(r - r_{eq})^2$. This formula is convenient for equilibrium conditions and doesn't apparently allow bond breaking. For cases where anharmonicity or dissociation plays a role Morse potential $V = D_{eq}(1 - e^{-a(r-r_{eq})})$ can be used.

Valence angle bending is also usually described by a harmonic potential $V = K_\theta(\theta - \theta_{eq})^2$, where θ denotes the valence angle. Another way of modelling the bending is to use triangulation. Instead of using the angle a radial distance is used in the formula, thus making a "virtual chemical bond".

Urey-Bradley potential has usually a harmonic form $V = K_{ik}^{UB}(r_{1,3,ik} - r_{1,3,0})^2$ as well. It improves the description of interaction between the atoms that are separated by 2 bonds by the inclusion of correlation and repulsion of electronic clouds.

Dihedral torsions describe the interaction of atoms separated by 3 bonds. Its general formula is $V = K_\phi(1 + z \cdot \cos(n\phi - \delta))$. The parameter $z = \pm 1$ distinguishes between two common definitions; n sets the periodicity (for example the amino group has $n = 3$), and δ is the equilibrium dihedral angle of the structure.

Improper torsion angle potential is a correction usually for planar systems that would not remain so without it. It also usually has a harmonic form $V_I = K_I(\Phi_{ijkl} - \Phi_0)^2$ and the definition of the angle Φ varies among implementations.

3.2.2 Non-bonding terms

Except the bonding terms that describe the interactions of molecules separated by less than four bonds there also must be such terms that incorporate the strong electrostatic interaction and dispersion.

Electrostatic interaction has a long range impact meaning that its contributions from an infinite distance don't vanish and can't be neglected. A multipole expansion is used for its description: the 1^{st} term is a monopole (charged molecules, ions); the 2^{nd} is a dipole that is usually present in all structures (for ex. water); the 3^{rd} is a quadrupole etc [39].

For the monopole plays the most important part in the multipole expansion, atoms are simplified as partial point charges. If a better precision is needed, dipoles or other terms can be added. The dipole moment can be specified by a simple addition of the term or by representative point charges.

Reaction field method

When isolated systems are studied, the reaction field would be a befitting method for long range electrostatics. The system is surrounded by a dielectric continuum with a defined permitivity.

More implementations have been derived, but it's not in the scope of this work to cover them. The original Onsager's method uses a cavity of a circular shape with a defined radius R [37]. The charges within are summed directly, the rest of them is treated as a dielectric continuum with permitivity ϵ. The dipole moment of the cavity polarizes this dielectric that acts back on the system accordingly to the formula

$$U = \frac{1}{4\pi\epsilon_0} \sum_{j<n} q_j q_n \left[\frac{1}{R_{nj}} + \frac{B_0 r_{nj}^2}{2R^3} \right] \; ; \quad B_0 = \frac{2(\epsilon - 1)}{2\epsilon + 1} \tag{3.6}$$

This method is simpler and faster than the following Ewald summation. In contrast with it, reaction field method allows calculating of long range electrostatics even on charged systems. It works best on isolated systems but it can be used even in a condensed phase. Problems may occur in cases when the system splits or its fragments tend to leave it [39].

Ewald summation

When periodic boundary conditions for bulk compensation (PBC, section 3.3) are used, the long range contribution to the electrostatic interaction cannot be neglected. For neutral systems, it can be treated in a more efficient manner if the summation over distant charges is done in the Fourier space rather than in the real space. The summation would not converge for charged systems.

The procedure uses neutralising charge distributions to make the sum in the real space convergent. Gaussian distribution is usually used as it has good features with the Fourrier transform. The inverse distribution is then summed in reciprocal space, where it converges better. Since the sum also includes interactions of each distribution with itself, they must be subtracted afterwards.

The implementation of the reciprocal summation often uses a particle mesh Ewald method. The system of particles is basically represented as a mesh of

densities on a discrete lattice in space, which allows the usage of the fast Fourrier transform. A careful choice of the lattice constant may provide a great speedup without any significant loss of precision, since the number of particles is usually greater than the number of required grid points [37].

Polarizability and Van der Waals interaction

Induction forces describe the polarizability of atoms. When two atoms draw nearer their electronic structure changes due to electrostatic interaction and electronic correlation. As a consequence, even neutral atoms may attract each other by induced dipoles. The tendency to form them is specified by polarizability, which is a quantity that describes the tendency of atoms to polarize themselves in an outer electric field.

Besides polarizability, there is also a dispersion interaction that describes electronic correlation and repulsion. Introduction of such forces is crucial with the point charge approximation as oppositely charged atoms would otherwise collapse into each other.

The force is not much strong; it's attractive at a distance from the source, where it is approximately proportional to r^{-6}. It is repulsive when atoms are too close and the electronic orbitals overlap a lot, which is approximately proportional to $e^{-\alpha r}$ [38].

There are two forms how to add dispersion to the interaction potential. The *Buckingham potential*, commonly referred to as (exp,6), uses the form

$$U_{VdW-B} = A_{ij} \cdot e^{-\alpha r_{ij}} - \frac{B_{ij}}{r_{ij}^6} \tag{3.7}$$

Its disadvantage is that the interaction becomes attractive once the atoms are too close [35].

Lennard-Jones form doesn't suffer from this artefact. Accordingly to its form it is commonly referred to as (12,6) potential:

$$U_{VdW-LJ} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} = \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \tag{3.8}$$

where ϵ denotes the depth of the minimum and σ is the equilibrious position [39].

It's unusual to define parameters for every pair of atomic types in a force field. The specified parameters often refer to the interaction of two atoms of the same type. The rest of the parameters are obtained by mixing rules, which definition differs among force fields. The most common mixing rules are Lorentz-Berthelot's:

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj})\,; \quad \epsilon_{ij} = \sqrt{\epsilon_{ii}\,\epsilon_{jj}} \tag{3.9}$$

3.3 Various other methods

Periodic boundary conditions (PBC). Most biological systems extend so much that their size is too large for computational studies. Only to perform a simulation, approximations must be undertaken. PBC are best suited for simulations of systems in the condensed phase.

Periodic boundary conditions compensate the bulk of the system by allowing the molecules to diffuse through simulation box boundaries. The simulation box must be generally triclinic; in biological applications and in water solutions it is usually orthorhombic, though. The particle that leaves the box via one of its sides appears on the corresponding facing side preserving its previous direction ([11] and [37]).

Although, the artificial periodicity may create artefacts, its usefulness exceeds its withdrawals. Only effects that are smaller in size than the sizes of the simulation box can be studied. Long range electrostatics need special treatment as the contributions are not negligible (see section 3.2.2).

Rigid bonds. When running a simulation a time-step dt must be chosen accordingly to the fastest movements, which are usually hydrogen vibrations. When such movements are constrained to a fixed length, the time-step dt can be elongated. By doing so a significant speedup can be achieved. With the usual classical way of propagation, such fixation may even improve results as for example C–H vibration has quantum characteristics that wouldn't be described. There are several algorithms for that purpose, for example Shake, Lincs, Settle, Rattle [39].

3.4 Statistical ensembles

Every system subjected to an MD simulation constitutes a statistical ensemble. With the choice of thermodynamic macroscopic conditions (pressure, temperature, etc.) one of the following statistical ensembles is usually used [42]:

Microcanonical ensemble indicates a macrostate with constant energy, volume and number of particles (EVN-ensemble). Internal energy U is thence the corresponding thermodynamic potential that reaches its minimum in equilibrium. Microcanonical ensembles are adiabatic systems. Since any state has the same energy and entropy, the principal of equal probabilities is valid. It's generated straightforward by both MD and Monte Carlo simulations. Because the energy is kept fixed and it consists of potential and kinetic parts $E = E_k + V$, the temperature of the system varies accordingly to kinetic energy $T \approx E_k$.

Canonical ensemble specifies a macrostate with constant temperature, number of particles and volume (TVN-ensemble). Helmholtz free energy F is the corresponding thermodynamic potential (definition in section 4.5). Since the temperature is kept fixed, the total energy $E = E_k + V$ is not constant as in microcanonical ensemble. Furthermore the principal of equal probabilities not valid any more, because the probability P of a state is proportional to total energy $P(E) \approx e^{-E/kT}$. A function that describes the distribution of probabilities of microstates with corresponding energies E_i is called a partition function. Its formula for a canonical ensemble is

$$P_i = \frac{e^{-E_i/kT}}{\sum_j e^{-E_j/kT}} \tag{3.10}$$

Isobaric-isothermic ensemble keeps temperature, pressure and number of particles constant (NPT-ensemble), so the Gibbs free energy G is its corresponding thermodynamic potential. It is an ensemble of choice for biophysical simulations as the conditions stated are the same as in biological systems.

Grandcanonical ensemble indicates a macrostate with constant temperature, volume and chemical potential μ. It may also be called TVμ-ensemble. The number of particles is not constant any more. The usefulness of this ensemble lays for example in simulations of adsorptions.

3.5 Ergodic theorem

The Ergodic theorem is the following tautology [35]:

> In an ergodic system, averageing over ensemble is the same as averageing over trajectory (time). A system is ergodic, if this condition is valid.

It allows evaluation of mean values of physical quantities in molecular dynamics methods. In MD practice the question usually is, how much ergodic a given system is. That mostly depends on the shape of the potential energy surface and the temperature.

3.6 Heat coupling

A direct MD simulation provides, at least approximately, constant total energy and volume. The natural conditions are different, though; many organisms keep their temperature, experiments are usually performed also at a room temperature and atmospheric pressure.

The equipartition theorem is [42]

$$< E_k >= \frac{3}{2}NkT = \left\langle \sum \frac{1}{2}mv_i^2 \right\rangle \tag{3.11}$$

The basic idea is that in equilibrium, the kinetic energy is distributed throughout the system equally among all degrees of freedom. The formula is generally imprecise, however it works in many cases. Assuming its sufficient validity, the relation between temperature and mean value of kinetic energy forbids complete fixation of the temperature. It's possible, though, to keep it constant on average.

There are several methods for heat coupling [39]:

Berendsen thermostat uses direct scaling of particles' velocities with a scaling factor λ

$$\lambda = \left[1 + \frac{\Delta T}{\tau_T}\left(\frac{T_0}{T} - 1\right)\right] \tag{3.12}$$

This method sets the defined temperature T_0 reasonably fast accordingly to the temperature coupling coefficient τ_T. Unfortunately, it may bring artifacts leading to a "fast flying ice cube" in extreme. This thermostat can be used in situations when the temperature is well established and the coupling is weak.

Andersen thermostat works in a stochastic way. It randomly chooses particles to which it assigns velocities from Maxwell-Boltzman distribution. The effect is explained as a collision of the particle with a "heat particle". The strength of the thermostat is set by the rate of the collisions.

Langevin thermosat is also a stochastic thermostat. A random force and a friction is introduced to the system:

$$ma = -tv + f(r) + f_{rand} \tag{3.13}$$

By a correct equilibration of the friction and the random force canonical ensemble is sampled correctly.

Nosè-Hoover thermostat uses an extended Lagrange function – it's added as a new degree of freedom. The Hamilton canonical equations then obtain a form:

$$\begin{aligned} \dot{q}_i &= p_i/m_i \\ \dot{p}_i &= -\frac{\partial}{\partial_{q_i}} U - \zeta p_i \end{aligned} \tag{3.14}$$

And the reservoir has an equation:

$$\dot{\zeta} = \frac{gkT_0}{Q}\left(\frac{T}{T_0} - 1\right) \tag{3.15}$$

The factor ζ is an equivalent to the friction coefficient, Q is the virtual mass of the thermostat. Its disadvantage is that when the temperature is far from the desired value, this thermostat converges slowly and it may oscillate around the value afterwards.

A weak coupling is usually preferred as its influence is not as pronounced. It's also possible to couple more thermostats in a chain and let only the weakest to interact.

Thermostats violate conservation laws for linear and angular momenta. When diffusion or other kinetics is studied, it's better to use the microcanonical ensemble.

3.7 Pressure coupling

The formula for an ideal gas shall illustrate pressure coupling methods:

$$PV = NkT \tag{3.16}$$

If the values of T a P are kept still, it's obvious that the volume V must fluctuate, which is the reason why barostats influence the volume. They are mostly derived with the assumption of periodic boundary conditions (see section 3.3).

Berendsen barostat scales the volume with the following factor

$$\mu = \left[1 - \frac{dt}{\tau_p}(P - P_0)\right] \tag{3.17}$$

so that the volume V becomes $V \rightarrow \mu^3 V$. Anisotropic conditions could also be simulated when a different μ_i (and a different P_0) is set for each dimension [35].

Nosè-Hoover barostat uses, like its heat coupling equivalent, an extended Lagrange function. The equations of motion are rather difficult and may vary among implementations. The variant that is implemented in the NAMD software package is enhanced by Langevin dynamics and is described by the following set of equations:

$$\begin{aligned} \dot{r} &= \frac{p}{m} + \dot{e}r \\ \dot{p} &= F - \dot{e}p - gp + R \\ \dot{V} &= 3\dot{e}V \\ \ddot{e} &= \frac{3V}{W(P - P_0)} - g_e\dot{e} + \frac{R_e}{W} \\ W &= 3N\tau^2 kT \\ < R^2 > &= 2mgkT/h \\ < R_e^2 > &= 2Wg_e kT/h \end{aligned} \tag{3.18}$$

where W is the mass of the piston that equilibrates the pressure, R_e is its noise, R is the noise of the particles [43]. The coordinate r belongs to the piston. The coordinates of particles are scaled after the formulae:

$$r_i = V^{1/3} e_i \qquad v_i = V^{1/3} \dot{e}_i \tag{3.19}$$

where e_i is a vector parallel with one of the box edges with the length set by the previous set of equations 3.18.

3.8 Propagation

An expected course for propagation of a quantum system, which many biophysically interesting systems are, would be to solve the Schrödinger equation 3.1. In classical MD simulations, however, Newton equations of motion are used instead since the studied system are represented classically – atoms are embodied by point masses with point partial charges; the exerting forces are defined by an approximate interaction potential (see section 3.1.1).

3.8.1 Newton's equations of motion

There are three Newton's laws of motion. The first law, law of inertia, states that any object subjected to a zero force is still or moves at a constant speed. With that it establishes the frame of reference.

The second law, the force law, states that an exerted force on an object is proportional to the change of its linear momentum. In systems, where the mass of the objects doesn't change, the force is proportional to the acceleration of the object with the mass as a coefficient.

The third law, the action-reaction law, states that if one object acts on another object with a force, the other object also acts on the first object with the same force but opposite direction.

The force law for systems with constant masses has a formula

$$\vec{F} = m\vec{a} = m\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}t^2} \tag{3.20}$$

where m is the mass of the object, $\vec{r}$ and $\vec{a}$ are its position and acceleration and $\vec{F}$ is the exerting force. It is a differential equation of the second order. To solve it, initial conditions – positions and their derivatives, velocities – must be known.

Such equation can be solved numerically by the finite difference method. In this method, the time is discretized into short intervals, time-steps dt, during which the force is considered constant. The Newton's second law 3.20 is hereby transformed into an equivalent set of differential equations of the first order:

$$\begin{aligned} \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} &= \vec{v} \\ \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} &= \frac{1}{m}\vec{F} \end{aligned} \tag{3.21}$$

3.8.2 Numerical methods

Many differential schemes for solving the Newton's equation of motion 3.21 exist [37], of which several chosen will be explained in the following paragraphs.

The simplest scheme is perhaps the **Euler method**. By rewriting the equations 3.21 using Taylor expansion and neglecting all terms except the first two one obtains

$$\begin{aligned} \vec{r}(t+dt) &= \vec{r}(t) + \vec{v}(t)dt \\ \vec{v}(t+dt) &= \vec{v}(t) + \frac{1}{m}\vec{F}(t)dt \end{aligned} \tag{3.22}$$

This scheme could be used with benefits in systems, where the forces depend on velocities of the particles, because the velocities are known before the evaluation of forces. The drawback is that this method accumulates errors with every step. On an example of a harmonic oscillator the Euler method systematically increases the total energy.

The **Verlet algorithm** can be derived by the addition of Taylor expansions of the second order of particles' positions for times $t + dt$ and $t - dt$ into the Newton's equations of motion 3.21

$$\vec{r}(t+dt) = 2\vec{r}(t) - \vec{r}(t-dt) + \frac{1}{m}\vec{F}(t)(dt)^2 \tag{3.23}$$

Even though the Verlet algorithm is still a simple scheme, it is the most common method in MD applications. It's more precise then Euler method and it doesn't suffer from its drawbacks – it's stable even on long trajectories. This particular variant doesn't include velocities in its formula and so it requires two sets of coordinates as an initial condition, which is quite impractical.

A reformulation that incorporates velocities and yet it generates identical trajectories is the **Velocity Verlet algorithm**. It can be derived from the basic scheme 3.23 by the addition of Taylor expansion of velocities and a few slight adjustments:

$$\vec{r}(t+dt) = \vec{r}(t) + \vec{v}(t)dt + \frac{1}{2m}\vec{F}(t)(dt)^2$$

$$\vec{v}(t+dt) = \vec{v}(t) + \frac{1}{2m}(\vec{F}(t) + \vec{F}(t+dt))dt \tag{3.24}$$

It is usually divided into more particular steps since the "new" velocity $\vec{v}(t+dt)$ depends on forces from two time-steps. This can pose a burden for large systems as it allocates much memory.

The **Leap-frog** variant of the Verlet algorithm is symbolically the same with the Euler equations 3.22. The difference lays in that the velocity is evaluated in the middle of the time-step, that is in the time $t + dt/2$. That gives:

$$\vec{r}(t+dt) = \vec{r}(t) + \vec{v}(t+\frac{1}{2}dt)dt + \frac{1}{2m}\vec{F}(t)(dt)^2$$

$$\vec{v}(t+\frac{3}{2}dt) = \vec{v}(t+\frac{1}{2}dt) + \frac{1}{m}\vec{F}(t)dt \tag{3.25}$$

The main feature is a benefit and a drawback at the same time. It's a stable algorithm unlike Euler's, but the velocities are evaluated in a different time than the coordinates, so that velocity dependant properties can't be evaluated without corrections. The advantage over Verlet's methods is that Leap-Frog could be easily started and only forces from the current time-step are needed.

4. Ab-initio methods

This chapter covers density functional theory and thermochemistry in ab-initio calculations, methods used in subsequent QM and ONIOM studies.

4.1 Density functional theory

Density functional theory (DFT) is in principle an exact theory of electronic structure of atoms and molecules. Instead of using the many-electron wave function $\Psi(r_1, r_2, \ldots)$, it is based on the electron density distribution $n(r)$. In its beginnings it served mainly the solid state physics, but its success there made it also popular among theoretical and computational chemists [44].

DFT is primarily a theory of electronic ground state structure. The traditional wave function based methods are generally preferred for systems with relatively small number of atoms. DFT allows inclusion of more atoms in a single computation, however, the accuracy is lower. It depends on the knowledge of the exchange correlation energy functional $E_{xc}[n(r)]$ that will be introduced in the following text. On the other hand, traditional wave function based methods may reach arbitrary level of accuracy, given a computational resource powerful enough [44].

4.2 Fundamentals of DFT

The fundamentals of DFT will be shown on the simplest class of systems, N non-relativistic, interacting electrons in a non-magnetic state with the following Hamiltonian

$$H = T + V + U \tag{4.1}$$

With the use of atomic units the terms in the Hamiltonian are defined

$$T = -\frac{1}{2}\sum_j \nabla_j^2\,; \quad V = \sum_j v(r_j)\,; \quad U = \frac{1}{2}\sum_{i\neq j}\frac{1}{|r_i - r_j|} \tag{4.2}$$

The arbitrary external potential $v(r)$ extends the definition to a broad class of Hamiltonians, not only the physically relevant Coulomb point charge potentials.

4.2.1 Hohenberg-Kohn theorems

The whole DFT relies on the Hohenberg-Kohn theorems. The first could be stated [44]:

> The specification of the ground state density, $n(r)$, determines the external potential $v(r)$ uniquely (to within an additive constant C)
>
> $$n(r) \longrightarrow v(r) \quad \text{(unique)} \tag{4.3}$$
>
> Since $n(r)$ also determines N by integration, it determines the full Hamiltonian H and thence, implicitly, all properties determined by H. Examples are the full N-particle ground state wave function $\Psi(r_1, ..., r_N)$, the electrical polarizability, the nth excitation energy, vibrational force constants, and potential energy surfaces for chemical reactions.

The minimal principle

The second Hohenberg-Kohn theorem allows a minimal principle for the energy as a functional of $n(r)$ to be derived. The energy $E_{v(r)}[n(r)]$ is defined

$$\begin{aligned} E_{v(r)}[n(r)] &= \int v(r)n(r)\mathrm{d}r \; + \; F[n(r)] \\ F[n(r)] &= (\Psi[n(r)], (T+U)\Psi[n(r)]) \end{aligned} \tag{4.4}$$

Since the wave function Ψ is a functional of $n(r)$, the both of the previous terms are functionals of $n(r)$ too. The minimal principle for a trial density $n(r)$ and the ground state density $n_0(r)$ and energy E could than be stated in the following manner ([44] and [45]):

$$E_{v(r)}[n(r)] \geq E_{v(r)}[n_0(r)] = E \tag{4.5}$$

Thomas-Fermi approximation could be derived by making approximations for $F[n(r)]$. The functional representation for the kinetic energy with good accuracy is impossible to find [46], so the modern DFT uses a different approach of Kohn-Sham. The largest contributions to $F[n(r)]$ are

$$F[n(r)] = T_s[n(r)] + \frac{1}{2}\int \frac{n(r)n(r')}{|r-r'|}\mathrm{d}r\mathrm{d}r' + E_{xc}[n(r)] \tag{4.6}$$

The term $T_s[n(r)]$ represents the kinetic energy of a non-interacting system with density $n(r)$ in the appropriate potential $\tilde{v}(r)$. The middle term is a classical expression for the interaction energy of charge densities. The last term contains the exchange correlation energy. If omitted, the theory would become equivalent to the Hartree approximation [44].

4.3 Kohn-Sham equations

The Euler-Lagrange equation associated with the stationarity of the energy functional $E_{v(r)}[n(r)]$ defined in the equation 4.4 could be rewritten into a set of self-consistent Kohn-Sham equations ([44] and [47])

$$\left(-\frac{1}{2}\nabla^2 + v(r) + \int \frac{n(r')}{|r-r'|}\mathrm{d}r' + v_{xc}(r) - \epsilon_j\right)\phi_j(r) = 0 \tag{4.7}$$

$$n(r) = \sum_{j=1}^{N} |\Phi_j(r)|^2 \tag{4.8}$$

$$v_{xc}(r) = \frac{\delta E_{xc}[n(r)]}{\delta n(r)} \tag{4.9}$$

The terms in the first equation 4.7 are equivalent with the terms in Hartree-Fock theory except the exchange correlation term $v_{xc}(r)$. They are in order: the

kinetic energy, the external potential $v(r)$, the Coulomb point charge potential, the exchange correlation potential $v_{xc}(r)$ and the orbital energy ϵ_j.

These equations must be solved self-consistently, like the Hartree-Fock equations. The exchange correlation potential is calculated in each cycle from the equation 4.9 accordingly to the chosen approximation of $E_{xc}[n(r)]$. It is then substituted to the equation 4.7 for Kohn-Sham orbitals. The ground state energy is given by

$$E = \sum_{1}^{v} \epsilon_j - \frac{1}{2} \int \frac{n(r')}{|r - r'|} \mathrm{d}r\mathrm{d}r' - \int v_{xc}(r) n(r) \mathrm{d}r + E_{xc}[n(r)] \tag{4.10}$$

The additional functional is a minor slowdown to the computation, especially when the equations 4.7–4.9 are in principle exact.

4.4 Exchange correlation energy

The accuracy of the energy obtained via the Kohn-Sham equations 4.7–4.9 depends on the knowledge of the exchange correlation energy functional $E_{xc}[n(r)]$. Several methods and levels of approximation have been derived with various success. The most notable exchange and correlation energy functionals LDA, B88X, LYP and their mixture B3LYP will be presented in this section.

4.4.1 Local density approximation LDA

Local density approximation (LDA) is the name for the simplest approximation of exchange correlation energy functional $E_{xc}[n(r)]$. It has been derived for the uniform interacting electron gas of density n, for which it is exact. It has a form of ([44] and [46]):

$$E_{xc}^{LDA}[n(r)] = \int \epsilon_{xc}(n(r)) n(r) \mathrm{d}r = -C_x \sum_{\sigma} \int n_{\sigma}^{4/3}(r) \mathrm{d}r = E_x[n(r)] \tag{4.11}$$

where $C_x = \frac{3}{4} \left(\frac{6}{\pi}\right)^{1/3}$, and σ denotes the spins α, β.

The LDA is accurate enough in cases where the native length scale over which $n(r)$ varies is large enough. Despite that the condition is rarely met, the LDA gives useful results for most chemical applications [44]. For that reason it often forms the dominant part of many E_{xc} formulations [46].

The spin-dependent generalization of LDA, the local spin density approximation (LSDA), has a clear overbinding tendency as it overestimates binding energies of most species. Still, the properties like bond lengths and angles or vibrational frequencies agree well with experiments. This makes the LSDA a reasonable choice for structural studies [44].

4.4.2 Generalized gradient approximations GGA

The next step in development of a more accurate exchange correlation functional was to include a gradient correction to the formula for LDA, 4.11. The form for

generalized gradient approximation (GGA) is

$$E_{xc}^{GGA} = \int f(n(r), |\nabla n(r)|)\mathrm{d}r \tag{4.12}$$

where $f(n(r), |\nabla n(r)|)$ is a suitably chosen function [44].

Becke exchange functional B88X

In atoms, which have only one centre, there is only dynamic correlation. An example of a GGA functional that was derived for such chemical species is the Becke exchange functional B88X

$$E_x[n] = -\sum_\sigma \left(C_X \int n_\sigma^{4/3}\mathrm{d}r - \int \frac{\beta n_\sigma^{4/3} x_\sigma^2}{1 + 6\beta x_\sigma \sinh^{-1}(x_\sigma)}\mathrm{d}r \right) \tag{4.13}$$

where $\beta = 0.0042$ and $x = |\nabla n| n^{-4/3}$ [46].

Lee-Yang-Parr correlation functional LYP

The Lee-Yang-Parr (LYP) correlation functional was one of the first functionals that turned attention of quantum chemists to DFT. It is based on the work of Colle-Salvetti that derived an approximative formula for correlation in terms of electron density and a Laplacian of the second order HF density matrix. It is:

$$E_c = -0.04918 \int n(R) \frac{1 + 0.173W \exp\frac{-0.58}{\beta}}{1 + \frac{0.89}{\beta}}\mathrm{d}R \tag{4.14}$$

$$\beta = 2.29 n^{1/3}(r)$$

$$W = 0.3814 n^{-8/3} \left[\nabla_r^2 P_{2HF} \left(R + \frac{r}{2}; R - \frac{r}{2} \right) \right]_{r=0}$$

$$r = |r_i - r_j|\,; \quad R = \frac{1}{2}(r_i + r_j)$$

where P_{2HF} denotes the second order HF density matrix [46].

Lee, Yang and Parr reformulated the Colle-Salvetti expression so that it contained only the electron density and the local kinetic energy density. Inserting a gradient expansion for the latter term they obtained [46]:

$$[\nabla_r^2 P_{2HF}(R, r)]_{r=0} = n(R)[t_{HF}(R) - 2t_W(R)] \tag{4.15}$$

$$t_W(R) = \frac{1}{8}\frac{|\nabla n(R)|^2}{n(R)} - \frac{1}{8}\nabla^2 n \tag{4.16}$$

$$t_{HF}(R) = C_F n(R)^{5/3} + \left[\frac{1}{9} t_W(R) + \frac{1}{18}\nabla^2 n \right] \tag{4.17}$$

The LYP correlation functional is acquired by its insertion into the equation 4.14.

Mixed functionals B3PW91 and B3LYP

It was not until 1993, when the paper [48] was released and the DFT became really widespread among practitioners of quantum chemistry. There the DFT methods were systematically assessed on the same set of test molecules as was used by Pople and coworkers in their development of Gaussian-n methods [45]. It was shown, that the Becke correlation functional B88X meant a significant improvement over the non-gradient-corrected methods like LDA; the absolute deviation decreased from 36.2kcal/mol (LDA) to 3.7kcal/mol (B88X).

The role of exact exchange on the precision of density functional methods was also investigated. A semi-empirical generalization for the exchange correlation functional contributions was proposed with the following formula

$$E_{xc} = E_{xc}^{LDA} + a_0(E_x^{exact} - E_x^{LDA}) + a_x \Delta E_x^{B88X} + a_c \Delta E_c^{PW91} \tag{4.18}$$

that was fitted via the coefficients a_0, a_x and a_c to the same set of test molecules as above. The coefficient a_0 determines the extent of electron-gas exchange with exact exchange. The coefficients a_x and a_c determine the influence of B88X as a gradient correction for exchange and PW91 on correlation. The fit was determined with a mean absolute deviation of 2.4kcal/mol, which is very close to the target value of chemical accuracy $\pm$2kcal/mol [48].

The three parameters are denoted in the functional abbreviations by the symbol "3" among the exchange and correlation functionals, so the mixed functional that was obtained by the previous equation is called B3PW91. The very popular B3LYP only uses LYP functional instead of PW91. This was not proposed by Becke's original paper [48], but it was available in the Gaussian 94 programme as an easy-to-use keyword that switched the two correlation functionals using the same semi-empirical parameters. Since the B3LYP functional appeared to perform better than the original B3PW91, it attracted more attention in the chemical community. Eventually, it was refitted with only a slight improvement [49].

4.5 Thermochemistry in ab-initio calculations

The quantities that describe macrostates are called state variables; pressure, temperature, volume, entropy or inner energy are examples of them. A quantity that is a function of state variables is also a state variable. Thermodynamic potentials are special cases of these functions. The motivation to introduce them is that they naturally follow conditions of particular classes of experiments. They also relate directly to the microscopic description of a thermodynamic system [42].

Besides inner energy U, there are three other thermodynamic potentials defined: Free energy F, Enthalpy H and Gibbs free energy G.

$$\begin{aligned} F(T,V) &= U - TS \\ H(S,P) &= U + PV \\ G(T,P) &= U + PV - TS \end{aligned} \tag{4.19}$$

these potentials prefer some variables over other. The knowledge of a thermodynamic potential in its preferred variables provides complete thermodynamic

information. Such functional is called the master function. The preferred variables for the introduced thermodynamic potentials are denoted as dependant variables in parentheses in the previous equations.

The maximal entropy principle implies that any thermodynamic potential reaches its minimum in equilibrium with its preferred variables fixed. That allows principles of thermodynamic potentials minima to be derived [42].

In computational chemistry the microstate of a given system could be obtained. Its knowledge is used to acquire the partition function. Such function contains all information needed to derive equilibrium thermodynamic quantities [42].

For a quantum system, the partition function generally consists of many kinds of contributions: translational, rotational, vibrational and electronic motion. For any component q the following equation for entropy S holds [42]:

$$S = Nk_B + Nk_B \ln \frac{q(V,T)}{N} + Nk_B T \left(\frac{\partial \ln q}{\partial T} \right)_V \tag{4.20}$$

where N is the number of particles, k_B is a Boltzmann factor, T is temperature, V is volume and as a bottom index it denotes that the partial derivative is taken at a constant volume.

The internal energy E can also be obtained from the partition function by differentiating

$$E = Nk_B T^2 \left(\frac{\partial \ln q}{\partial T} \right)_V \tag{4.21}$$

which contains thermal contributions [42]. The both of these equations are temperature dependent.

Ideal gas approximation is the basic approximation taken in most ab-initio programme packages. It assumes that all particles are non-interacting. For the electronic structure, it's also basically assumed that excited states are inaccessible, which is a reasonable assumption for systems without low lying excited states [50]. Thermodynamic potentials, defined by equations 4.19, could then be calculated. With the given approximations the Gibbs free energy would be

$$G = U + PV - TS = U + nRT - TS \tag{4.22}$$

where the ideal gas approximation is used according to the equation 3.16, temperature T is set and the entropy S is evaluated from the formula 4.20. This means that ΔG will be evaluated correctly if the number of moles changes during the reaction [50].

The partition function for contribution from translation movements is given by

$$q_t = \left(\frac{2\pi m k_B T}{h^2} \right)^{3/2} V = \left(\frac{2\pi m k_B T}{h^2} \right)^{3/2} \frac{k_B T}{P} \tag{4.23}$$

where the ideal gas approximation is again used to simplify the evaluation [50]. With the use of Stirling's approximation the entropy contribution from translational movements is

$$S_t = R \left(\ln q_t + 1 + \frac{3}{2} \right) \tag{4.24}$$

The contribution to the internal thermal energy is [50]

$$E_t = \frac{3}{2}RT \tag{4.25}$$

The specific formulae used in Gaussian programme for other contributions from electronic, rotational and vibrational motions are given in the reference [50].

5. QM/MM methods

Computational approaches for studies of biological systems have to deal with their inherent size as most of them contain large number of atoms that need to be included into the assay. A way to do it is to crop the system in "a very Occamish way" so that it contains only essential parts. Depending on the situation, the chosen system and the desired level of accuracy, this can mean taking only a few atoms as well as hundreds.

Proteins (section 1.2.2) are considerably large structures that contain hundreds or thousands of atoms. A particular fold of a protein is important for its function. In particular, if an enzymatic reaction in a protein is under scrutiny, the environment of a substrate must be conserved for the most accurate description. To fulfill this demand, it's usually needed to include at least tens of atoms into the calculation.

Not all atoms need the same level of description, however. In such case a hybrid approach can be used with benefits. By defining layers treated with different levels of theory, a significant speedup can be achieved [51]. Theories that allow an arbitrary number of defined layers exist, though, the most common approach is to dissect the system into two layers.

A layer definition depends on the theory chosen and on a chemical sense of the scientist. In the following text I will refer to these general terms:

Model system is the part that is treated with higher accuracy, usually a QM method. It can be also called as a QM, high or inner layer.

Real system is the whole system – it contains all atoms and is treated with the lower level of theory.

Low layer is the part that remains if the model system is subtracted from the real system. It contains only such atoms that are treated only on the lower level in all subcalculations. It is usually called the MM part of the system as some MM-based method is often used for its description.

5.1 Introduction to generic QM/MM and ONIOM theory

Hybrid methods are a class of computational techniques that combine two or more model chemistries together. A generic QM/MM methods create a large subgroup as they are the most common. ONIOM [1] is more general, however. It can combine any number of molecular orbital methods as well as molecular mechanics methods. In the following text I will rely on an article *Combining quantum mechanics methods with molecular mechanics methods in ONIOM* [51] most often.

[1] ONIOM is an abbreviation of "our Own N-layer Integrated molecular Orbital molecular Mechanics"

5.2 Link atoms

The main distinction among layered hybrid models is the treatment of covalent bonds between regions. A common approach is to use link atoms or frozen orbitals [2]. Both perform well, though the general opinion is that the latter describes the boundary better. Another method is to use pseudopotentials [51].

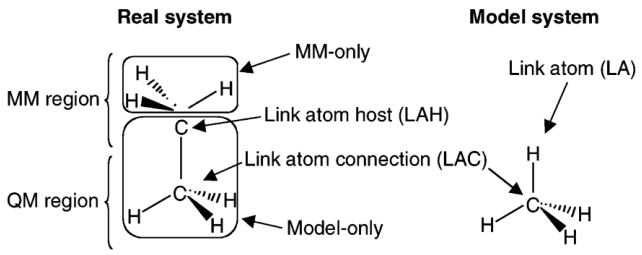


Figure 5.1: An example partitioning of ethane. Taken from [51].

With the link atom method, there are 4 sets of atoms in a 2-layer system:

In Inner layer atoms (both in model and real system)

LA link atoms (in the inner layer only; not in the real system)

Con outer layer atoms covalently connected to In-atoms

Out outer layer atoms – other cases

An example for the the above sets is in the picture 5.1.

In a general case, link atoms are at a specified distance from the connected In-atoms. In ONIOM, the LA position is obtained with a scale factor g.

Link atoms are placed on the line, that connects the In-layer center to which it is connected (LAC – *LA Connection*) with the atom it substitutes (LAH – *LA Host*). LAC–LA distance is obtained by scaling the original LAC-LAH distance with a factor g, which is chosen so that a chemically reasonable LAC-LAH distance yields a reasonable LAC-LA distance.

In order to generate the model system, there must exist a function

$$R_{LA} = f(R_{In}, R_{Con}) \tag{5.1}$$

which can be arbitrarily chosen [52]. In ONIOM a simple bond length scaling is used:

$$\mathbf{x}_{LA} = \mathbf{x}_{LAC} + g(\mathbf{x}_{LAH} - \mathbf{x}_{LAC}) \tag{5.2}$$

Since the LA shall mimic the existing bond, the choice of a bond length scaling as described in the equation 5.2 above is straightforward. Thus, when the bond length changes, the LA's distance changes accordingly to the factor g – a better approach than using fixed lengths that remove one degree of freedom with each Con-atom reducing it by their number. Other ways, for example with g not constant, exist [52].

[2] frozen orbitals are more complicated as they require parametrization of the orbitals

The factor g is chosen so that the resulting LA-LAH bond length is rational – usually in the equilibrium so that for example scaling a C-C bond to the C-H satisfies $|CC| \cdot g = |CH|$. If no g is specified by a user, Gaussian package sets it as

$$g = \frac{d_{In} + d_{LA}}{d_{In} + d_{Con}} \tag{5.3}$$

Scale factors for LA are in the recent implementations of ONIOM different for QM and MM methods. The LA in QM remains the same with its scale factor, but in MM calculation, the scale factor is generally different as the atom type may also be. In the case when the factor is 1, it mimics the generic QM/MM method as the MM terms in the energy equation (equation 5.6 in the following text) cancel identically. This difference improves results of vibrational frequencies (as described in [51]).

Link atoms are usually hydrogen atoms, but any atom is applicable if it mimics the substituted part well. For further improvements a parametrization that changes electronegativity with a shift operator or a parameter may be added.

With the LA definition and an energy definition like

$$E = E(R_{In}, R_{Con}, R_{Out}) \tag{5.4}$$

it is possible to obtain energy gradient ΔE (with a simple Jacobi matrix), Hessian and dipole moment and other electric field derivatives.

5.3 Energy definition, electronic embedding

Another main distinction among layered hybrid models is the treatment of electrostatics. The simplest method is the *classical* or *mechanical embedding*. The interaction between the high and the low layer is described by the low level theory – usually MM. In this particular case, the MM partial charges of atoms from the low layer interact with assigned partial charges of the atoms from the high layer.

A different description of electrostatic interaction between layers is called *electronic embedding*. In this case charge distributions of MM and QM layers interact together by adding the partial MM charges into the QM Hamiltonian. Thus the point charges from the low layer polarize the charge distribution of the high layer. A particular implementation of these methods depend on the QM/MM scheme used [51].

There are two main schemes of energy evaluation. The generic QM/MM uses a summation scheme, which could be in the case of *mechanical embedding* described by the following equation:

$$E_{QM/MM} = E^{MM}_{MM-only} + E^{QM}_{model} + E^{MM}_{MM-only*model-only} \tag{5.5}$$

The terms are: $E^{MM}_{MM-only}$ – MM-energy of the low layer (without LA), $E^{MM}_{MM-only*model-only}$ – interaction between the QM and the MM regions (MM terms with at least one centre in the QM region and one in the MM region). A simplistic explanation of it could be that there are two separate regions described by their methods and the interaction between them is added as a separate term.

ONIOM energy expression is different in that it uses an extrapolation scheme. The MM-energy is evaluated for the whole system as well as for the

model system. In the case of *mechanical embedding*, the interaction between the low and the high layer is then described as a part of the MM-energy of the whole system. The equation for such method is as follows:

$$E_{ONIOM} = E^{MM}_{real} + E^{QM}_{model} - E^{MM}_{model} \tag{5.6}$$

The interactions between partitions are evaluated in the E^{MM}_{real} term of equation 5.6. Only link atoms are a special case as they don't cancel out exactly – the corresponding terms in E^{MM}_{real} and E^{MM}_{model} may contain different values due to different typings and positions.

The above expressions 5.6 and 5.5 become identical, when there are no bonded interactions between the model system and the rest of the system.

As can be seen, ONIOM calculates energies of "complete subsets of the system", whereas generic QM/MM takes "incomplete parts". As a result, only ONIOM can mix two quantum methods together.

5.3.1 Electronic embedding

Electronic embedding (EE) is an extension of the basic ONIOM mechanical embedding scheme, which incorporates point charges from MM-only region into the QM Hamiltonian, thus providing explicit interaction of the electrons with the environment [51].

An equation for generic QM/MM with *electronic embedding* is

$$E_{QM/MM+EE} = E^{MM}_{MM-only} + E^{QM}_{model,v} + E^{MM}_{MM-only*model-only,noQ} \tag{5.7}$$

with corresponding Hamiltonian

$$H^{QM}_{model,v} = H^{QM}_{model} - \sum_{i,N} \frac{s_N q_N}{r_{iN}} + \sum_{J,N} \frac{Z_J s_N q_N}{r_{JN}} \tag{5.8}$$

where N is and index for MM layer atoms, J is and index for QM layer atoms and i is an index for electrons. The symbol s denotes a scaling factor that will be explained in the following text.

The **energy expression for ONIOM with electronic embedding** is

$$E_{ONIOM+EE} = E^{MM}_{real} + E^{QM}_{model,v} - E^{MM}_{model,v} \tag{5.9}$$

$$E^{MM}_{model,v} = E^{MM}_{model} + \sum_{J,N} \frac{q_J s_N q_N}{r_{JN}} \tag{5.10}$$

where N is an index for MM layer atoms and J is an index for QM layer atoms. The quantum term $E^{QM}_{model,v}$ uses the same Hamiltonian 5.8 as above.

A practical result of this formulation is, that when using ONIOM, model system point partial atomic charges must be also specified (not in generic QM/MM).

Scaling factors s, presented in the equation 5.8, are of a great importance as they manage the strong QM–MM[3] electrostatic interaction. These factors are used only with EE scheme and control the polarization of the wave function. They only affect the interaction between the QM and the MM parts [51].

[3]The MM-MM interaction is treated in the force field.

A general consensus is that this interactions shouldn't be counted between atoms that are separated by less then 3 bonds – the corresponding factors $s_N = 0$ as the electrostatic effect involved is already included in the corresponding angle-bending and bond-stretching force field terms. The electrostatic interaction is fully turned on for other atoms by specifying $s_N = 1$. This treatment helps avoiding overpolarization of the wave function caused by strong (and generally imprecise) interaction with near MM point charges.

Scaling factors are also useful for over/under-counting problems. The electrostatic interaction is partially incorporated in the bond stretching, angle bending and torsinal terms. By counting it separately as a non-bond interaction, one actually overcounts it. By leaving it out completely, this energy contribution is on the contrary undercounted – and as a consequence omitted completely from all interactions (with other atoms also). By introducing scaling factors, it is possible to correct this aggravation, that is always present in QM/MM methods due to the atom centred point charges and charge densities incompatibility.

In ONIOM, identical scaling factors s_N are used for both MM and QM model system atoms, as the model system must be identical in both levels of theory.

5.4 Correct partitioning

A careful choice of partitioning of the system must be taken, otherwise cancellation problems may occur when chemical reactivity is studied. Relative energies (for example energetic barriers) have sense only when the system composition does not change – number of atoms, their types and connectivity remain the same. Otherwise it presents discontinuity to the potential energy surface [51].

A safe way to choose the right partitioning is to define the low layer more than 3 bonds away from a bond creation/breaking place. At the border, exactly 3 bond away, all terms exist (bond, angle, dihedral) and may remain same (cancel identically), but there may be a change in the atom type of the link atom (and due to scale factors also in its position), thus providing different non-bonding interactions.

5.5 Microiterative geometry optimizing procedure

When optimizing geometries of QM/MM-described systems, special procedures must be taken. Most of the information provided in this section was taken from the article *Geometry Optimization with QM/MM, ONIOM, and Other Combined Methods. I. Microiterations and Constraints* [53].

With the definition of link atoms, the potential energy surface is well defined and the E gradient in ONIOM scheme is

$$\frac{\partial E}{\partial q} = \frac{\partial E^{high}_{model}}{\partial q} \cdot \mathbf{J} + \frac{\partial E^{low}_{real}}{\partial q} - \frac{\partial E^{low}_{model}}{\partial q} \cdot \mathbf{J} \tag{5.11}$$

where $\mathbf{J}$ is a Jacobian for model $\rightarrow$ real system transformation.

Thanks to the ONIOM energy definition, and the resulting gradient form, it's possible to carry out the optimization of the real-MM system separately (see

equations 5.6 and 5.11). After its full minimization a QM calculation with its minimization step follows. However, the real system naturally incorporates all atoms and so it also depends on the coordinates of the atoms in the high layer. Then the next step is minimizing the real system again with subsequent QM step and so on until convergence is found. This is usually called as the standard microiterative scheme.

When electronic embedding is applied, though, the situation becomes more difficult as the coordinates of otherwise separably optimized systems become entangled; meaning $\frac{\partial E^{high}_{model}}{\partial q^{low}_{real}} \neq 0$ and vice versa. Modified schemes must be used in that case.

The article [53] presents several modifications to the microiterative procedure for electronic embedding:

> Yang et al. approximated the QM charge distribution by electrostatic potential (ESP) charges, which were then used unchanged to describe the Coulombic interaction between the two layers during the optimization of the MM region. After the microiterations, the QM region is fully optimized and the ESP charges are re-evaluated. This sequence is repeated until convergence. However, it is clear that in this scheme the PES used for the geometry optimization, which includes the interaction of the QM ESP charges with the MM partial charges for the electrostatic interaction between the layers, is not the same as that used for the energy calculation, which includes the interaction of the QM nuclei and electrons with the MM partial charges.
>
> Friesner et al. presented a similar scheme, but used the ESP charges for the QM region only to describe the perturbation to the exact gradient during the microiterations. After these microiterations with essentially a frozen wave function, the optimization step in the QM region is taken, and the process is repeated. Because the perturbation goes to zero at convergence, the optimized structure is a true critical point on the surface for the energy calculation.
>
> However, in this approach the MM region is not fully optimized when the QM optimization step is taken, because formally the wave function has to be re-evaluated after the microiterations. It is not clear how this will affect the performance of the QM optimizer, in particular the Hessian update mechanism. Therefore, we propose two modifications to Friesner's scheme.
>
> First, after the microiterations, the wave function should be re-evaluated, followed by a new series of microiterations. Thus, the QM wave function and MM geometry optimization are iteratively repeated until self-consistency. Only then do we take a QM optimization step.
>
> Second, it is not necessary to use ESP (or other) localized charges to represent the QM charge distribution. With fast multipole methods, the electrostatic interaction with the exact QM charge density can be incorporated efficiently

When optimizing the MM region with constraints, all QM-coordinates and the constrained MM-coordinates are kept fixed in space as the procedure is held in

Cartesian coordinates. In that case, the microiterative scheme must be modified so that it permits translations and rotations of the QM-region as a rigid body. Such approach may even enhance the unconstrained optimization as it is no more necessary to reorient the whole MM region to suite the otherwise fixed-in-space QM region.

When optimizing the QM region, a quasi-Newton methods with variable step sizes are usually used. To control the procedure there's usually some "trust radius" fuse to check on the large steps. The test is made only for a large enough step and in that case the programme checks, whether a step back from the new point would really come back. Only if it is successful the microiterations are performed. If not, the microiterations are made more carefully with a smaller step size that slowly increases with the step-number.

The implementation of the microiterative procedure in Gaussian programme can be described in a few points (taken from [53]):

1. Coordinates are separated into two sets, q^m for the QM region and q^l for the MM region so that $\frac{\partial E^{high}_{model}}{\partial q^{low}_{real}} = 0$ and vice versa. Thus, only the energy and gradient of the real system at the low level need to be calculated during the microiterations.

2. Cartesian coordinates are used for q_l to avoid costly coordinate transformations for the large MM region, and redundant internal coordinates are used for the QM region to minimize the number of expensive QM steps needed for the optimization.

3. The MM region is augmented with coordinates to permit the rigid body translation and rotation of the QM region. This is essential if any of the atoms in the MM region are constrained, but also improves the efficiency of unconstrained optimizations.

4. The step size in the QM region needs to be controlled so that the optimization stays in the same valley during the next microiteration.

6. Hydrolysis of methyl diphosphate: QM study

Phosphodiesters are chemical species that involve one or more phosphate groups connected together via ester bonds. DNA, RNA and phosphorylated nucleotides such as GDP/GTP are examples of such molecules. Phosphate groups exist in a deprotonated form in physiological pH around 7. It means that GDP/GTP carry the electrostatic charge of $-3e^-$ or $-4e^-$ depending on the number of phosphate groups. Therefore GDP/GTP attracts salt cations. Specifically GTP is bound to the magnesium divalent cation Mg^{2+} when functioning as a cofactor in GTPases [54].

Substitution is a chemical reaction, where a part of a molecule is replaced by a substituent. In biochemistry, substituents are either *nucleophilic* or *electrophilic*. The electrophilic substituents are likely to attack electron rich sites of molecules like double-bonds. In contrast, nucleophilic substituents favor to attack electron-defficient sites of molecules [1].

Phosphoryl transfer reactions belong to substitution reactions. A general chemical formula for them is

$$\mathrm{R{-}O{-}H_2PO_3} + \mathrm{SA} \longrightarrow \mathrm{R{-}O{-}S} + \mathrm{H_2PO_3{-}A} \tag{6.1}$$

where R is a donor molecule, A is the acceptor of the phosphate group H_2PO_3 and S is the substituent. If the SA is H_2O then we speak about hydrolysis. Hydrolysis is, very simply said, an exothermic reaction that disrupts a covalent bond at a cost of a water molecule. Hydrolysis is widespread in living cells. It is a part of many crucial processes like glycolysis, transcription (RNA synthesis), translation (protein synthesis) and reactions in active sites of GTPases.

There are three possible mechanisms for the phosphate tranfer [55]:

Dissociative mechanism requires a stable metaphosphate ion PO_3^- to be formed. This ion is then attacked by a nucleophile, which could be the rate-limiting step [55]. The metaphosphate anion is stable in the gas phase, but hasn't been observed directly in solution.

Associative mechanism can be split into two steps. The first step is the addition of the nucleophile that creates a phosphorane intermediate. The second step is the elimination of the phosphate with its new bonding partner.

Concerted mechanism is similar to the associative one, but the bond formation to the nucleophile and bond fission of the leaving group occur simultaneously in the transition state [55].

6.1 MDP as a simple model of GTP

To investigate the simplest possible model of polyphosphate hydrolysis, methyl diphosphate (MDP) was chosen. A GTP hydrolysis study [56] justifies the choice of MDP as a model system representing GTP by stating that there's a shift

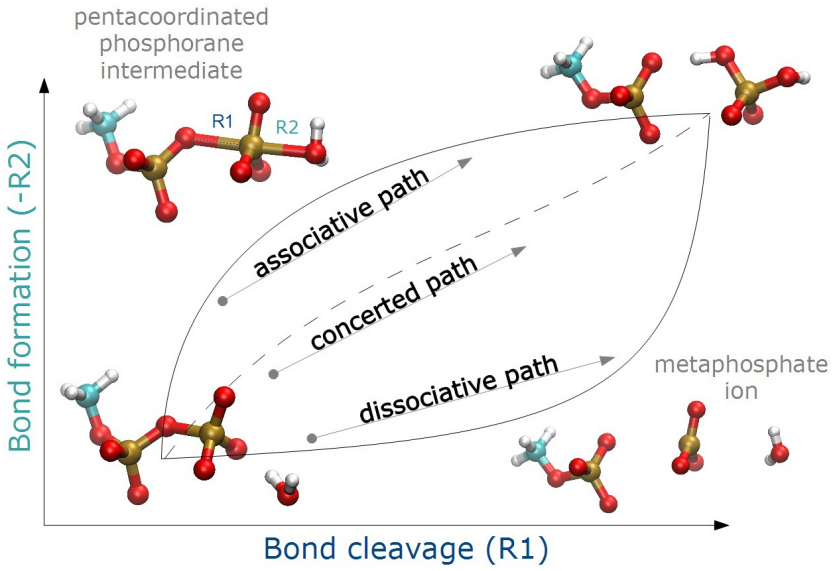


Figure 6.1: A diagramatic representation of the general free energy surface for phosphate transfer in phosphomonoesters. The three possible pathways are depicted accordingly to the reaction coordinates defined in Figure 6.2.

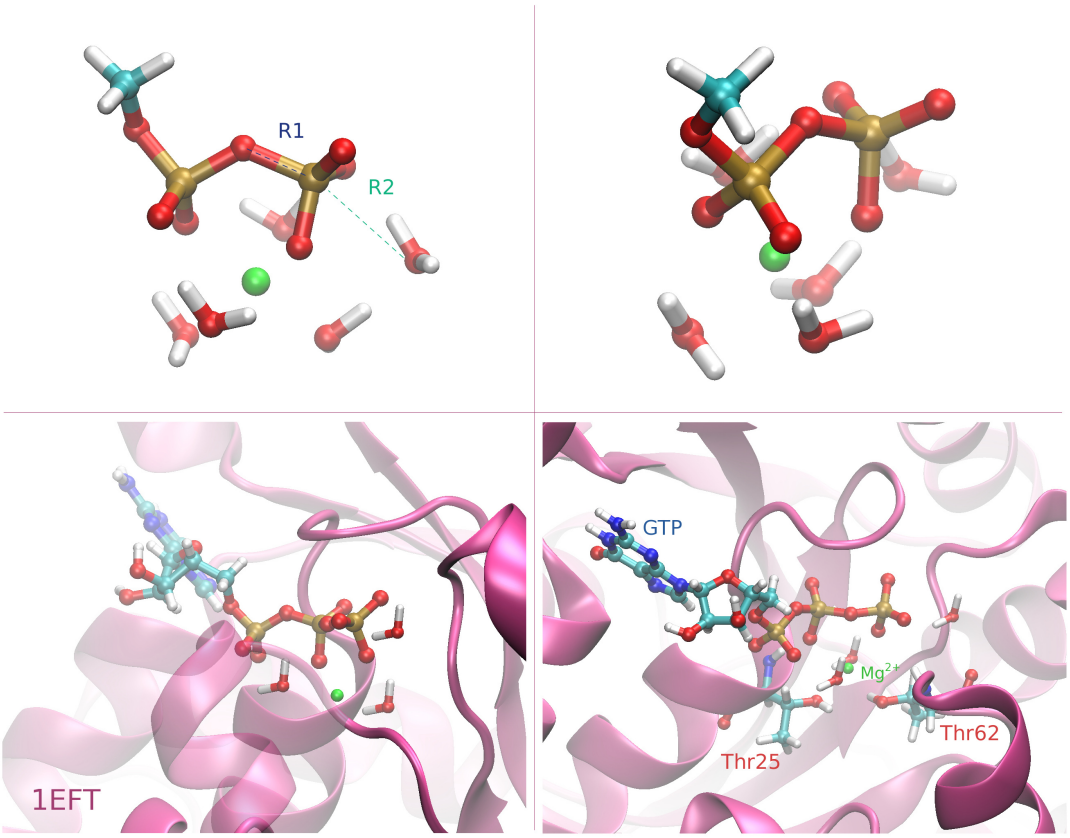


Figure 6.2: Reaction coordinates R1 and R2 used in the QM study of the MDP hydrolysis. Bottom pictures show the 1EFT structure of EF-Tu from which the initial conformer of the MDP model system was derived. Similarity between MDP and GTP is apparent. Two water molecules in the MDP model system replaced hydroxyl groups of Threonine residues in EF-Tu.

of negative charges from the γ-phosphate group towards the β-phosphate group in the context of enzyme Ras. It means that resulting charge distributions are similar for GTP and GDP. As our intention is to study a GTP hydrolysis in the EF-Tu active site, the undertaken simplifications seem to be reasonable.

6.1.1 MDP hydrolysis in the gas phase

The MDP model system is shown in the picture 6.2. It consists of MDP, Mg^{2+}, its first solvation shell and the hydrolytic water molecule. The initial coordinates were chosen so that the MDP model system resembles the equivalent part of the 1EFT crystal structure of EF-Tu.

The reaction coordinates were chosen naturally as the lengths of the scissile and arising bonds (Figure 6.2).

A Free energy surface (FES) was computed using the Gaussian (ver. 2009 rev. A) software package by applying Density Functional Theory (DFT) and B3LYP functional. The temperature was let at its default value 25°C.

The initial geometry optimization of MDP in the gas phase was held in the 6-31+G* basis set using constraints for the reaction coordinates R1 and R2 (Figure 6.2). MDP in selected points on FES was then subjected to both geometry optimization and free energy calculations in the extended basis set 6-311++G**. The same level of theory and basis set was benchmarked against MP2 for similar model systems in [57].

During the free energy calculations, MDP in various (R1,R2) points had to be frequently checked for different conformations and proton transfers to ensure that the resulting configuration is the best one. The resulting MDP conformers in neighboring surface points are consistent. The resulting free energy surface (FES) can be seen in Figure 6.3. The resulting potential energy surface PES is depicted in Figure 6.4. Numerical values of free energies are given in the Table 6.1. The reaction pathway is highlighted in color. The surface points calculated in the extended basis set are bold faced.

Preferences of the MDP system in choosing reaction paths are obvious – the easiest way for the reaction to happen is to flow via a dissociative pathway through the slight depression at (3.4,2.3Å). The activation barrier for the reaction evaluated with the extended basis set is then $\Delta G = 26.3$ kcal/mol with the rate constant $2 \cdot 10^{-6}$ s^{-1}.

6.1.2 MDP hydrolysis in implicit solvents

Further, the MDP model system was solvated using two implicit solvent models: PCM and COSMO ([58] and [59]). This was specified by a keyword `SCRF=PCM` or `SCRF=COSMO` in the Gaussian input file.

Free energy calculations were produced for conformers optimized previously in the gas phase. The optional optimization in implicit solvents wouldn't probably improve results as the reaction field might push the MDP system into artificial minima. A sufficient buffer zone formed by explicit water molecules is needed to avoid this artifact [60] (see Discussion, section 6.2).

The resulting free energy surfaces can be seen in Figures 6.5 and 6.7. The corresponding potential energy surfaces are depicted in Figures 6.6 and 6.8. There

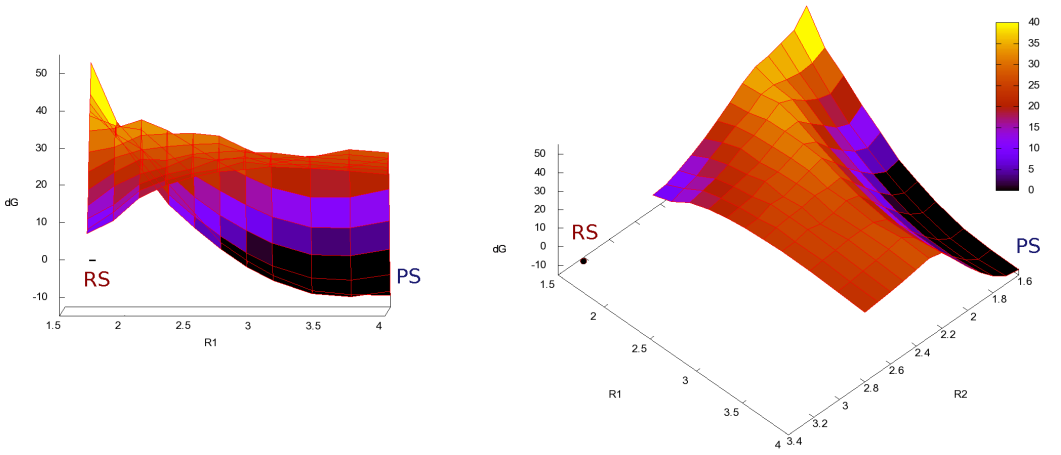


Figure 6.3: FES of the MDP hydrolysis in the gas phase. Energies are given in kcal/mol, distances in Å.

	R1 [Å]										
R2 [Å]	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.4	3.7	4.0
1.6	50.47	34.64	26.85	17.88	10.10	2.38	-2.55	-6.49	-10.48	-11.65	-11.77
1.7	42.11	31.35	21.70	12.67	6.42	0.75	-3.99	-7.68	-11.22	-12.01	-10.58
1.8	39.87	30.71	23.92	15.61	9.66	4.30	-0.14	-3.65	-6.93	-7.48	-5.90
1.9	37.92	31.90	27.70	21.43	15.01	10.09	5.94	2.61	-0.45	-0.80	0.94
2.0	36.90	34.08	33.00	27.04	21.04	16.82	12.99	9.81	6.98	6.78	8.70
2.1	32.90	33.56	35.95	33.05	28.69	24.47	20.88	17.84	15.15	15.22	17.40
2.2	27.31	28.70	31.84	32.31	32.46	31.66	28.30	25.47	**23.07**	**23.49**	25.65
2.3	22.19	26.07	28.01	29.08	29.58	**29.81**	27.34	27.38	**26.14**	**28.18**	27.46
2.4	17.56	22.27	25.81	25.44	27.10	27.71	28.02	25.85	**26.60**	**26.94**	26.12
2.5	13.44	18.72	21.92	23.79	26.61	25.55	26.57	24.49	26.43	25.73	24.77
2.6	10.07	13.88	19.34	21.99	23.45	26.03	26.32	26.21	25.53	24.68	23.90
2.7	7.30	11.90	17.76	20.47	22.32	23.48	25.44	24.70	24.87	23.92	23.15
2.8	6.37	9.71	15.86	19.25	21.45	22.90	23.47	**24.29**	23.88	23.17	22.30

Table 6.1: ΔG matrix in kcal/mol for the reaction in vacuum evaluated at the B3LYP/6-31+G* level

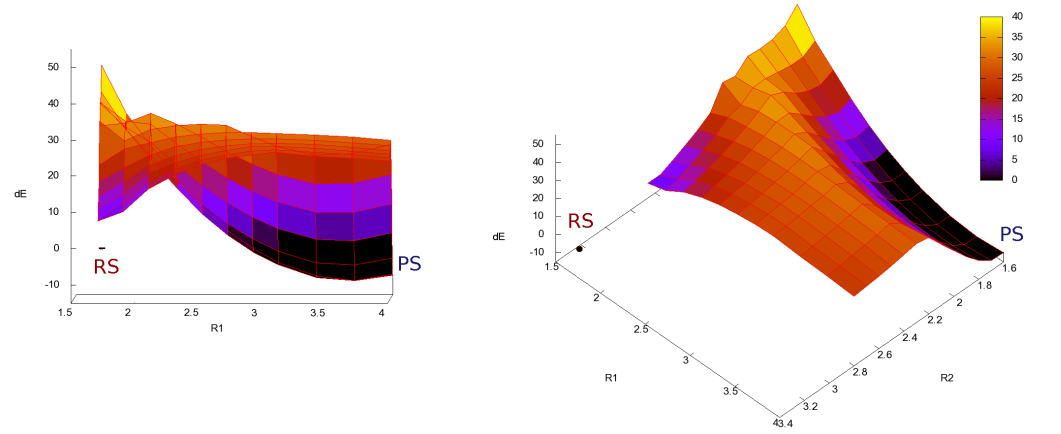


Figure 6.4: PES of the MDP hydrolysis in the gas phase. Energies are given in kcal/mol, distances in Å.

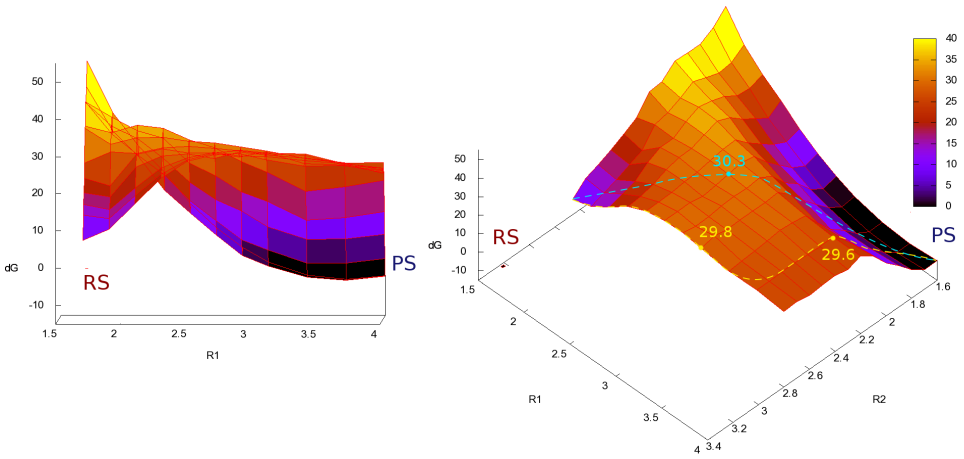


Figure 6.5: FES of the MDP hydrolysis in the PCM implicit solvent. Energies are given in kcal/mol, distances in Å.

	R1 [Å]										
R2 [Å]	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.4	3.7	4.0
1.6	53.20	39.46	30.37	23.39	14.94	8.54	3.22	-0.69	-4.23	-5.26	-4.32
1.7	46.44	34.69	27.79	18.70	12.65	6.60	1.26	-1.55	-4.53	-5.38	-4.24
1.8	42.35	33.87	28.30	22.99	16.08	10.66	5.62	2.90	-0.71	-0.69	0.52
1.9	42.40	36.20	32.84	27.51	21.14	16.59	13.46	9.19	5.82	6.00	7.27
2.0	42.80	38.71	35.15	33.03	26.44	23.01	18.71	16.11	11.09	11.79	13.48
2.1	36.44	34.75	36.67	35.93	32.74	30.69	27.02	24.11	19.03	20.25	21.94
2.2	36.03	31.72	33.54	33.59	32.56	32.09	31.07	29.92	27.92	**26.69**	26.84
2.3	26.88	29.86	29.37	30.94	31.26	**30.80**	**30.11**	**29.35**	**28.88**	**26.34**	24.34
2.4	22.84	20.92	29.30	**30.76**	**29.73**	**29.71**	30.43	29.66	29.49	26.86	25.66
2.5	19.12	18.04	25.99	**29.02**	**29.53**	28.39	30.11	29.35	28.49	27.28	25.24
2.6	16.09	14.76	20.67	26.06	28.62	30.08	**29.16**	29.06	27.97	27.18	24.63
2.7	13.37	12.46	18.66	24.35	28.19	29.93	28.98	30.45	27.33	26.86	26.11
2.8	6.78	9.83	16.49	23.46	27.95	**29.66**	**30.13**	**28.50**	26.78	26.82	25.56

Table 6.2: ΔG matrix in kcal/mol for the reaction with PCM as implicit solvent evaluated at the B3LYP/6-31+G* level

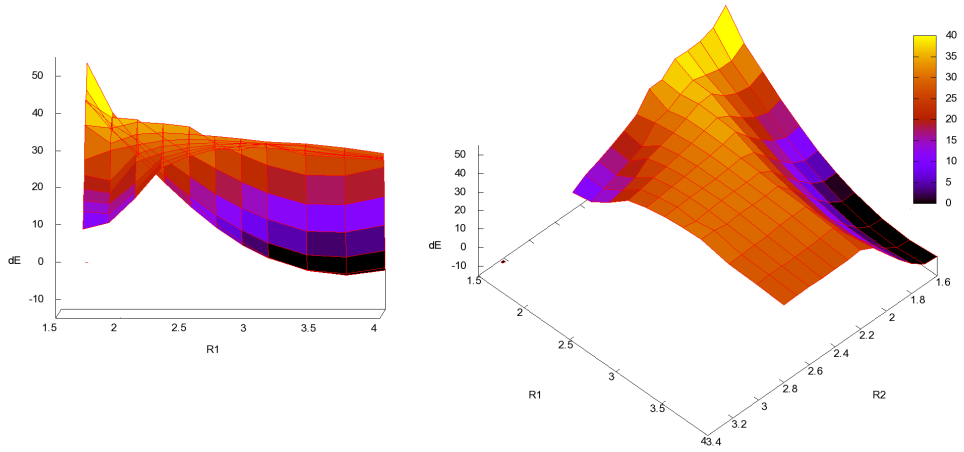


Figure 6.6: PES of the MDP hydrolysis in the PCM implicit solvent. Energies are given in kcal/mol, distances in Å.

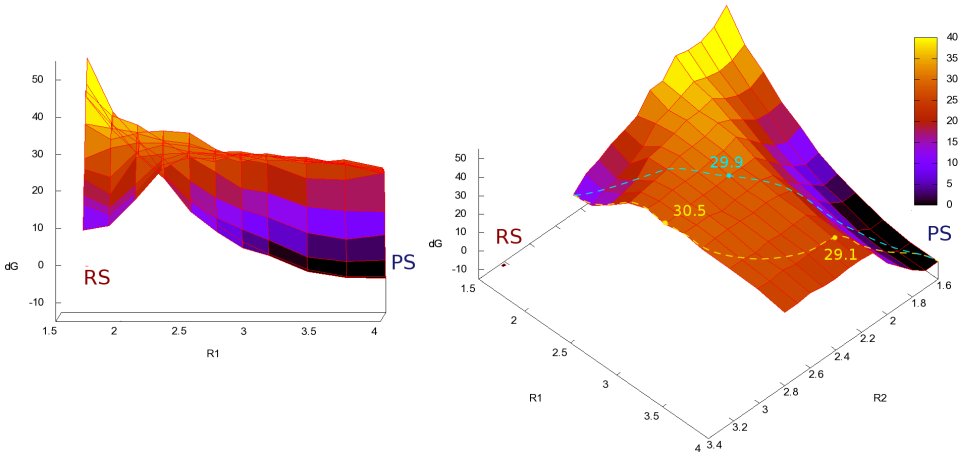


Figure 6.7: FES of the MDP hydrolysis in the COSMO implicit solvent. Energies are given in kcal/mol, distances in Å.

	R1 [Å]										
R2 [Å]	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.4	3.7	4.0
1.6	51.42	37.00	29.24	19.81	13.86	6.67	2.02	-1.16	-5.36	-7.33	-7.26
1.7	44.58	33.56	25.29	18.91	10.28	4.89	0.56	-1.48	-5.77	-7.45	-7.57
1.8	41.28	32.36	26.02	19.96	13.28	8.87	5.71	2.81	-1.72	-2.98	-2.57
1.9	41.40	33.88	29.13	24.59	18.86	14.31	11.26	9.02	4.84	3.15	4.32
2.0	43.20	36.37	34.06	29.56	25.48	21.25	16.98	15.85	12.03	11.06	10.78
2.1	34.46	33.18	32.11	32.53	31.95	27.66	24.97	23.69	**19.77**	19.40	21.36
2.2	34.39	30.36	31.98	30.03	29.33	26.98	27.05	25.84	**25.66**	**23.63**	22.20
2.3	25.30	28.19	29.52	29.52	28.67	**27.41**	26.20	26.76	**25.06**	**24.46**	22.50
2.4	22.67	20.51	27.25	**29.14**	**27.35**	**26.86**	25.94	26.25	24.74	24.13	21.46
2.5	20.12	16.01	26.07	**26.88**	**27.06**	**26.78**	**24.74**	26.25	25.64	23.98	21.90
2.6	16.00	12.74	18.81	25.38	27.36	26.45	**27.11**	26.61	24.41	25.47	23.37
2.7	13.62	10.97	16.78	24.18	26.07	25.91	27.12	26.74	25.22	24.24	22.64
2.8	6.78	7.97	14.90	22.93	26.85	**25.21**	**27.28**	**25.59**	25.44	24.82	22.47

Table 6.3: ΔG matrix in kcal/mol for the reaction with COSMO as implicit solvent evaluated at the B3LYP/6-31+G* level

is also a matrix representation for the ΔG numerical values in Tables 6.2 and 6.3. The highlighting style is the same as in the previous section. A comparison of corresponding FES points is given in the Table 6.4. A graphical representation of the highest surface points for the reaction pathways along with ΔG barriers is depicted in Figure 6.9. A comparison of the obtained free energy surfaces is in Figure 6.10, the optimized configurations along with the two reaction paths are depicted in Figure 6.11.

The FESs of the MDP hydrolysis in the gas phase and implicit solvents are roughly similar. The main difference lays in the flattening effect of the solvent on the part of FES that represents leaving of the phosphate group.

Two reaction pathways were detected: *concerted* that has its TS at (2.7, 2.3Å) and was absent in the gas phase; *dissociative* that has its TS around (3.0, 2.8Å). The ΔG activation barrier for both pathways is basically the same and yields approximately 30 kcal/mol.

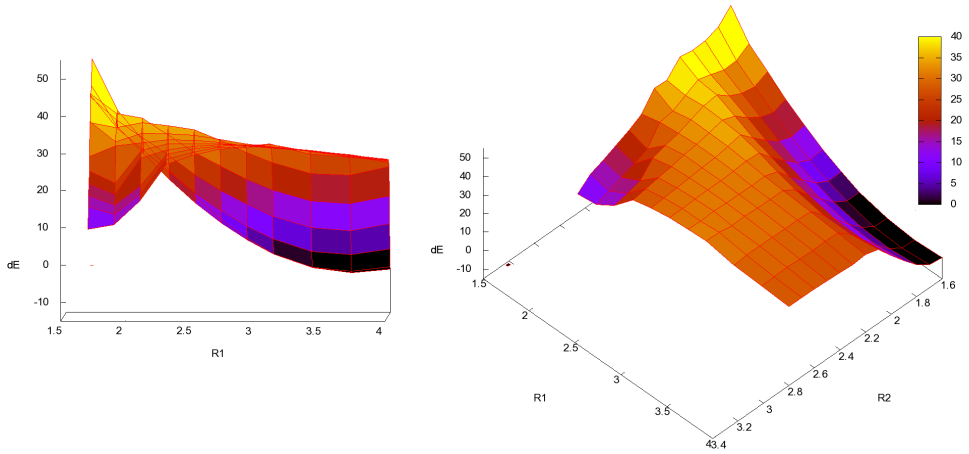


Figure 6.8: PES of the MDP hydrolysis in the COSMO implicit solvent. Energies are given in kcal/mol, distances in Å.

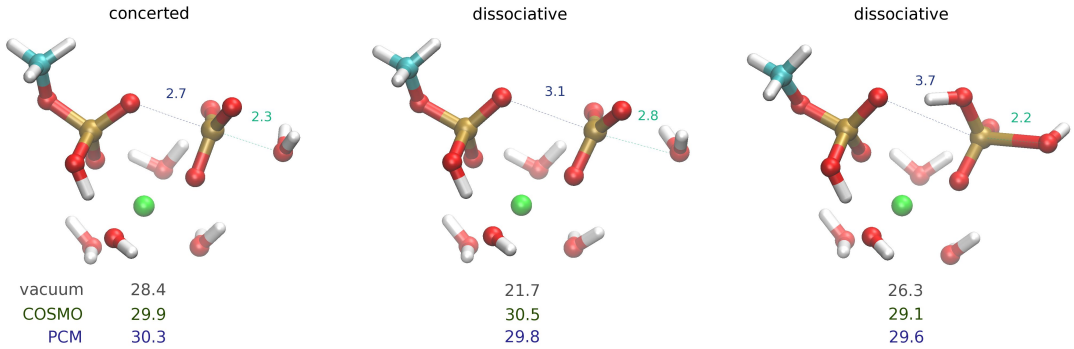


Figure 6.9: Representation of the highest FES points for reaction pathways along with ΔG barriers. The colour scheme is used consistently throughout this chapter.

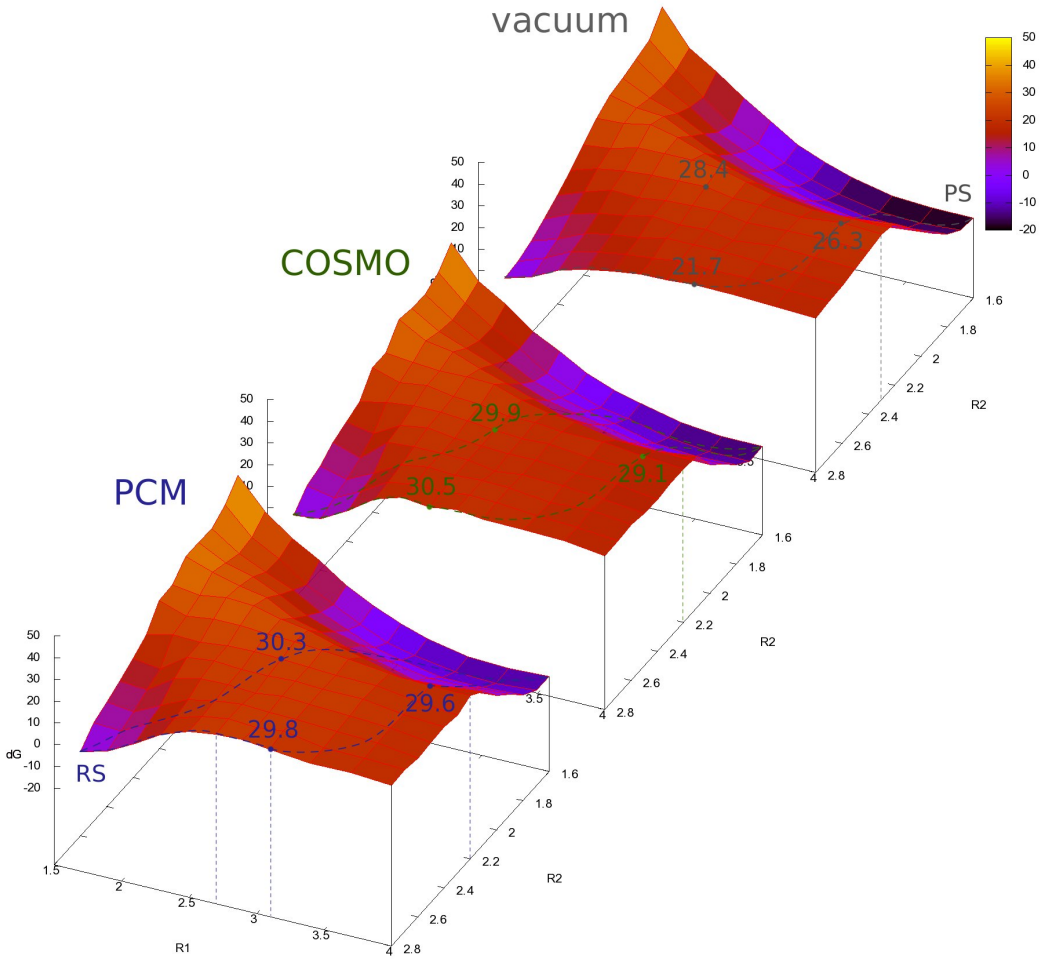


Figure 6.10: Free energy surfaces of the hydrolytic reaction of MDP in different environments. The energy ΔG (dG) is in kcal/mol, the reaction coordinates R1 and R2 are in Å. The surfaces were obtained on the B3LYP/6-31+G* level, whereas the given energies of several points were calculated on a better, B3LYP/6-311++G**, level. Probable reaction paths were highlighted with dashed lines. Both paths share similar activation energy, the concerted path was not located in the gas phase. RS means reactant state, PS product state.

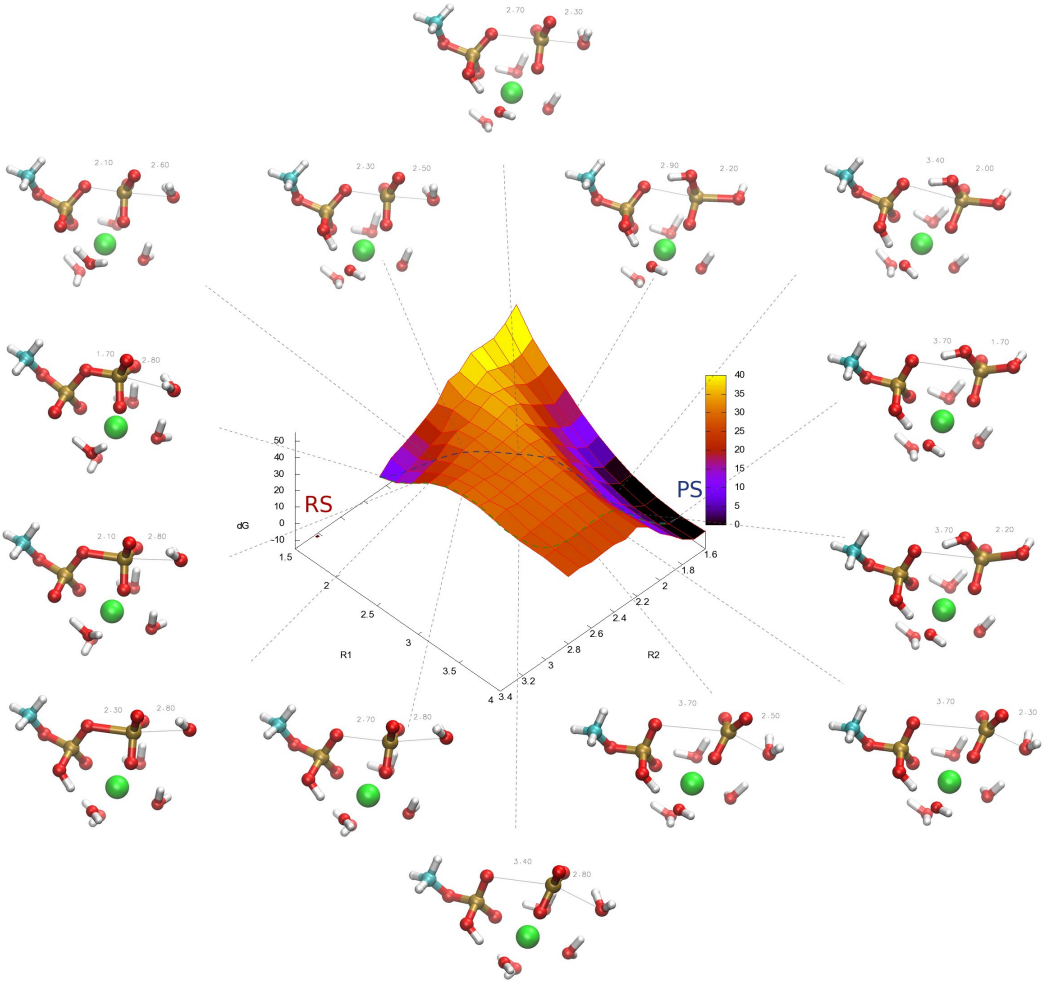


Figure 6.11: A comparison of the optimized configurations along the reaction paths. The surface belongs to the PCM-solvated calculations.

concerted path

environment	R1	R2 [Å]	ΔG [kcal/mol]	k [s^{-1}]
vacuum	2.7	2.3	28.4	$6 \cdot 10^{-8}$
PCM	2.7	2.3	30.3	$3 \cdot 10^{-9}$
COSMO	2.7	2.3	29.9	$5 \cdot 10^{-9}$

dissociative path

environment	R1	R2 [Å]	ΔG [kcal/mol]	k [s^{-1}]
vacuum	3.1	2.8	21.7	
PCM	3.1	2.8	29.8	$6 \cdot 10^{-9}$
COSMO	2.7	2.8	30.5	$2 \cdot 10^{-9}$
vacuum	3.7	2.3	26.3	$2 \cdot 10^{-6}$
PCM	3.7	2.2	29.6	$9 \cdot 10^{-9}$
COSMO	3.7	2.2	29.1	$2 \cdot 10^{-8}$

Table 6.4: Comparison of ΔG barriers and rate constants of various FES points in different environments

6.2 Discussion and conclusions

MDP hydrolysis both in vacuum and in implicit solvents was thoroughly analyzed in this chapter. The FESs were obtained for conformers optimized in the gas phase. For a given point (constraints R1 and R2) the gas phase optimizations directed the MDP model system towards only a few attractors – mostly just one. It means that this partially solvated system proved to be very stable if geometry optimization was applied for different initial conditions. On the other hand, testing geometry optimizations produced using the implicit solvent models often had many attractors – they possessed chaotic behavior as even two close initial conditions led to highly divergent results. Therefore, if further optimizations in the implicit solvent were applied for some of the gas phase optimized structures (as was the approach in [57]), it would be possible to carve the FES almost at will.

In fact, dielectric continuum approaches perform well only in cases, when the electric field exerted by the solute is weak enough (linear response region). As noted in [57] or [60], solvent yields important effects on the behavior of the phosphate groups. In the latter study [60], a hybrid solvent approach, that combines explicit water molecules with implicit dielectric continuum methods, was used. The explicit water molecules functioned as a "buffer zone" for the implicit solvent model. It was found that only 2–3 water molecules are sufficient to stabilize the doubly charged phosphate anion.

In [57], three reaction pathways for MDP hydrolysis with similar activation energies around 32 kcal/mol were found: associative (TS: 2.4,2.1Å), concerted (2.8,2.3Å) and dissociative (3.4,2.8Å). In here presented calculations, the associative pathway was not revealed – it is energetically disfavored. A comparison of corresponding surface points is provided in the Table 6.5. Reaction barriers in each study have similar heights. Nevertheless, here presented reaction pathways have reaction barriers about 2 kcal/mol lower. This seems to be a result of different methodologies chosen for a geometry optimization (as mentioned above).

path	R1	R2 [Å]	ΔG [kcal/mol]
associative	2.4	2.1	≥33
associative[57]	2.4	2.1	32
concerted	2.7	2.3	29.9
concerted[57]	2.8	2.3	32
dissociative	2.7	2.8	30.5
dissociative	3.1	2.8	30.4
dissociative[57]	3.4	2.8	32

Table 6.5: Comparison of ΔG barriers in COSMO with results from [57]

Interestingly, the values gained in this study are closer to the experimentally observed activation barrier of 28 kcal/mol [57].

To save computational time, the FES scan was restricted to surface points R2 $\leq$ 2.8Å. The reaction in a gas phase apparently doesn't suffer from that. The solvated surfaces, however, might have better paths that circumnavigate the currently highest points through the unmapped area. Would this be the case, there still remains a spine that must be crossed and that has approximately the same height as the current highest points. In addition, the attack of the hydrolytic water molecule shall be the rate-determining step in the dissociative pathway according to [55].

The reference [61] presents that Mg^{2+} serves as a temporary storage for electrons in the case of the GTP hydrolysis in solution and in Ras. To assess the field, Mulliken atomic charges were analyzed here. The difference in the Mulliken charge of Mg^{2+} between the reactant state and the intermediate state in the present calculation varies. The gas phase calculations don't suggest any storage during the reaction. The solvated system, however, encounters a charge shift of $-0.07e^-$ in the dissociative pathway and even $-0.15e^-$ in the concerted pathway (becomes less positive). The transferred negative charge is eventually removed from the cation in the product state. The Mulliken atomic charges for the COSMO model reflect very similar changes as for PCM. Although, the absolute values for Mg^{2+} are generally $+0.08e^-$ higher suggesting that the reaction field is probably even stronger. A full negative charge is anticipated to reside on the leaving phosphate group in the loose transition states of monoesters [55]. The leaving phosphate group bore $-0.8e^-$ on average in the implicit solvents and $-0.6e^-$ in the gas phase intermediate states. Taking the effect of Mg^{2+} into account, the partial charges reflect the anticipation well.

6.3 Input file example

A typical Gaussian input file section for a geometry optimization looked as follows:

```
%Nproc=2
# b3lyp/6-31+g* opt=(ModRedundant) freq

R1=17 , R2=28 ; vakuova optimalizace

-1 1
```

```
 P     0.140214     0.292614     0.332889
 O    -0.862002     0.552458     1.681238
 . . . .

B  1 2   1.7  F
B  1 6   2.8  F
```

The first line defines the number of CPU cores required for the run.

The most important parameters are probably on the second line that starts with `#`. The energy evaluation method along with the basis set is simply specified by its common abbreviation. The `opt` keyword sets the program to optimize the geometry of the given system. Internal coordinates are used by default. The `ModRedundant` option allows fine tuning of the internal coordinates used for the optimization. It also allows to freeze them, which was our intention. The `freq` keyword generates frequencies, vibrational modes and thermochemistry.

After that, separated by an empty line, the jobname follows. Then starts the molecule definition. The first line contains specifications of charge and multiplicity, the subsequent lines contain coordinates of atoms. The last required specifications are the modifications to the internal coordinates required by the `ModRedundant` keyword. The example freezes two bonds between atoms 1–2 and 1–6 to the values 1.7Å and 2.8Å.

6.4 Output file example

The output file contains a lot of information by default: Mulliken atomic charges, energy gradients, distance matrices, internal coordinates and much more. The section containing vibrational frequencies, required by `freq` keyword, helpes in judgements, whether the optimized structure is settled in a minimum. The existence of an imaginary vibrational mode would prove that the structure probably occupies a saddle point on the potential surface.

This applies to ground states. In transition states, however, there shall exist exactly one imaginary mode that reflects the lowest pass, a saddle point of the first order, for the reaction [62]. This was checked in the following section:

```
******    1 imaginary frequencies (negative Signs) ******
Diagonal vibrational polarizability:
     194.1090533     173.9615017     375.2847929
Harmonic frequencies (cm**-1), IR intensities (KM/Mole), Raman scattering
activities (A**4/AMU), depolarization ratios for plane and unpolarized
incident light, reduced masses (AMU), force constants (mDyne/A),
and normal coordinates:
                     1                      2                      3
                     A                      A                      A
Frequencies --   -48.2652               39.6250               49.3046
Red. masses --     7.7360                7.7765                4.4276
Frc consts  --     0.0106                0.0072                0.0063
IR Inten    --     6.7543                0.2882                4.1566
 Atom  AN      X      Y      Z        X      Y      Z        X      Y      Z
    1  15    -0.06   0.04   0.00    -0.06  -0.08  -0.05     0.04   0.01   0.02
    2   8    -0.07   0.01   0.00    -0.13  -0.15  -0.07     0.03  -0.02   0.03
    . . . .
```

It is an example of an output file for the FES point (1.9, 2.8Å). The only imaginary frequency belongs to the mode that involves the hydrolytic water approaching the leaving phosphate group as can be seen in the three columns that contain displacements for the vibration (only 2 lines shown).

Further in the output file, the thermochemistry section contains thermodynamic quantities:

```
-------------------
- Thermochemistry -
-------------------
Temperature   298.150 Kelvin.  Pressure   1.00000 Atm.
Atom     1 has atomic number 15 and mass  30.97376
. . . .

Zero-point correction=                           0.197381 (Hartree/Particle)
Thermal correction to Energy=                    0.219919
Thermal correction to Enthalpy=                  0.220863
Thermal correction to Gibbs Free Energy=         0.147682
Sum of electronic and zero-point Energies=          -1831.751489
Sum of electronic and thermal Energies=             -1831.728951
Sum of electronic and thermal Enthalpies=           -1831.728007
Sum of electronic and thermal Free Energies=        -1831.801188
```

The absolute value of Gibbs free energy that was used for ΔG evaluations is on the last line in atomic units Hartree. Since a different energy unit is common in biochemistry, conversion was needed according to the relation

$$1\,\mathrm{Ha} = 627.5095\,\mathrm{kcal/mol} \tag{6.2}$$

7. Hydrolysis of GTP by EF-Tu: ONIOM study

Phosphate compounds play crucial roles in many processes in all living organisms. Guanosine triphosphate (GTP) is a cofactor of GTP-binding proteins that catalyze its hydrolysis. These enzymes usually regulate cellular processes by changing their conformation from GTP- to GDP-bound and vice versa [63]. Elongation factor Tu (EF-Tu) is a GTP-binding protein that is involved in proteosynthesis (see the chapter 2).

A rather complex potential energy surface was reported for the GTPase reaction in a similar GTP-binding protein Ras [64]. Two reaction pathways, associative and dissocitive, were presented. The associative mechanism involved the formation of a pentacoordinated transition state with subsequent elimination of an orthophosphate. A trigonal metaphosphate is involved in the dissociative mechanism. The concerted mechanism, as described in the previous chapter 6, was not considered in [64].

The authors of [55] reviewed catalytic effects of several enzymes on the phosphoryl transfer reaction:

> Comparisons of the structures of enzymes that catalyze phosphoryl transfer show that the only similarity at the active site is the presence of positive charge, in the form either of a binuclear metal center and/or positively charged amino acid side chains. For phosphatases, which catalyze the hydrolysis of phosphomonoesters, it has been suggested that metal ions or cationic side chains might change the normally loose transition state into a more associative process by promoting electron withdrawal from the phosphorus atom, thus promoting a nucleophilic attack.

Metal ions were numerously reported to have an accelerating effect on the hydrolysis. In mechanistic studies, however, the effect varied. Furthermore, either Mg^{2+} or Ca^{2+} do not alter the loose transition state in phosphate transfer [55].

According to [61], the Mg^{2+} ion provides a temporary storage for the electrons taken from the phosphate groups. In addition, a specific coordination of the magnesium ion in the active site of the Ras-GAP protein influences the energetics of the reaction in a major way. Depending on the position of Mg^{2+}, the activation barrier varies from 15 to 32 kcal/mol.

Spontaneous binding of sodium ions into the active site was observed in MD simulations of EF-Tu [11] and Ras [65]. However, the sodium ion was not included in the QM/MM study of Ras "because the position of such an ion is not known and difficult to predict". The choice is justified by a rate constant measurement that suggests only a slight catalytic effect of counterions at room temperature [67].

Charge transfer within the GTPase reaction of Ras-GAP is discussed in [65]. It was found that negative charge accumulates at the β-phosphate group for the dissociative mechanism and at the γ-phosphate group for the associative mechanism.

The importance of the coordination of the attacking water molecule is stressed on an example of myosin, an ATPase. The water molecules around ATP hydrogen bond to protruding oxygens of the charged phosphate groups. Nevertheless, such position is not in favor of a nucleophilic attack required for hydrolysis.

A simplified model of the GTPase reaction in solution was studied in the previous chapter 6. Now GTP in the context of the G domain of EF-Tu will be studied by ONIOM, a hybrid QM/MM scheme described in chapter 5.

7.1 Preparation of the structure

As described in section 2.3, EF-Tu is a complex protein consisting of three domains. Domains 2 and 3 take part in binding of the aminoacylated tRNA. The G domain binds GDP/GTP in its active site. The initial structure for this study was derived from the crystal structure of EF-Tu from bacteria *Thermus Aquaticus* (PDB code 1EFT). It shows EF-Tu in its active conformation. In my bachelor thesis [11], the 1EFT structure was subjected to MD simulations, which revealed spontaneous binding of sodium ions into the active site.

Three different structures were prepared for here presented QM/MM studies. Two of them were based on the 1EFT crystal structure. As the crystal structures contain none or only a few water molecules, the corresponding QM/MM calculations were held in vacuum. The third structure was chosen from conformers observed in the above mentioned MD simulations, which were produced in solution. So a layer of explicit water molecules surrounding EF-Tu was present in QM/MM calculations as well.

It should be mentioned that only parts of EF-Tu relevant for hydrolysis were included (Figure 7.1). The selection comprised almost 3200 atoms: residues 9–211, GTP, Mg^{2+} with its first solvation shell, the hydrolytic water molecule and several other water molecules around the active site. The N-terminal residues of EF-Tu were omitted because they form a tail directed outwards into the space. Residues 212–405 that form domains 2 and 3 and connecting loops were omitted too, because they take part rather in binding of the aminoacylated tRNA. Because of the cuts in the protein chain, the N-terminal and C-terminal residues needed addition of terminal groups.

Layers, subsets of atoms, must be defined for a QM/MM study. Our system was separated into two regions[1] treated with different levels of theory.

The QM layer consists of the tree phosphate groups of GTP, Mg^{2+} along with its first solvation shell, the hydrolytic water molecule and a methyl group that substitutes the guanosine nucleoside. The first solvation shell of Mg^{2+} is formed by two water molecules and two hydroxyl groups of threonine residues. A sodium ion has been included into to the QM layer as well if it was present in the active site.

The methyl group that substitutes the guanosine nucleoside and the two hydroxyl groups of threonines 25 and 62 that form the first solvation shell of Mg^{2+} required link atoms to be defined due to layer boundaries. The link atoms were treated as described in chapter 5. It means that the hydroxyl groups were then represented in the QM layer as two water molecules. A graphical representation

[1] Nomenclature of the chapter 5 is used.

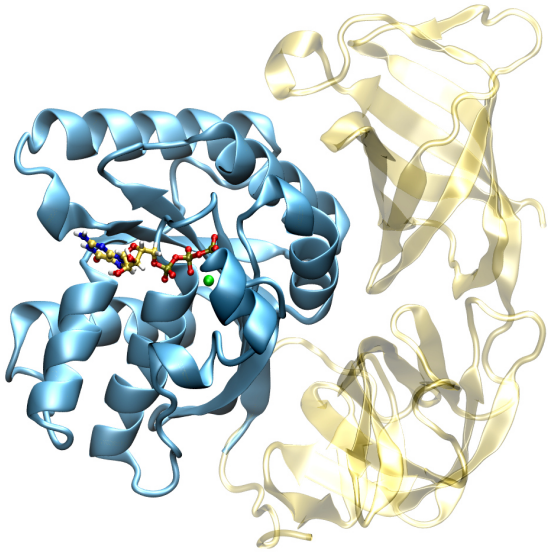


Figure 7.1: Selection of the G-domain that is relevant for the GTPase reaction in EF-Tu (cyan, opaque). Domains 2 and 3 take part in binding of aminoacylated tRNA (yellow, transparent). Eight residues in the N-terminal tail were also excluded. GTP and Mg^{2+} shown in the active site.

of the QM layer definition is depicted in Figure 7.2.

The Low layer naturally contained all atoms but those in the QM layer.

There are well-established force field parameters for monophosphates in the AMBER force field. However, they do not extend well to polyphosphates [66]. The same source also provides better parameters for polyphosphates that are intended for use with the existing AMBER force field. They justified the improvement by fitting the potential energy for both sets of parameters to a potential energy surface calculated at the RHF/6-31+G* level. Since the parameters could also be used with nucleosides, they were adopted here for the description of the methyl-tri-phosphate/GTP.

7.2 Free energy surface

In order to remain consistent with the previous MDP hydrolysis study (see chapter 6), equivalent methods were used. The systems, as described in the previous sections, were subjected to a QM/MM ONIOM study using the software package Gaussian (ver. 2009 rev. A). Density functional theory (DFT) with B3LYP functional was chosen as a model chemistry for the QM layer; the real system and the low layer were described molecular mechanically using the AMBER force field with parameters for polyphosphates taken from [66]. Bonds cut at layer boundaries were treated by link atoms.

The reaction coordinates for the free energy surface (FES) were chosen naturally and consistently with the previous MDP study, as the lengths of the scissile and forming bonds. The definition is illustrated in Figure 7.2.

Each study always started by a proper definition of all atoms for the forcefield, which was done with the help of the GaussView software package. I always

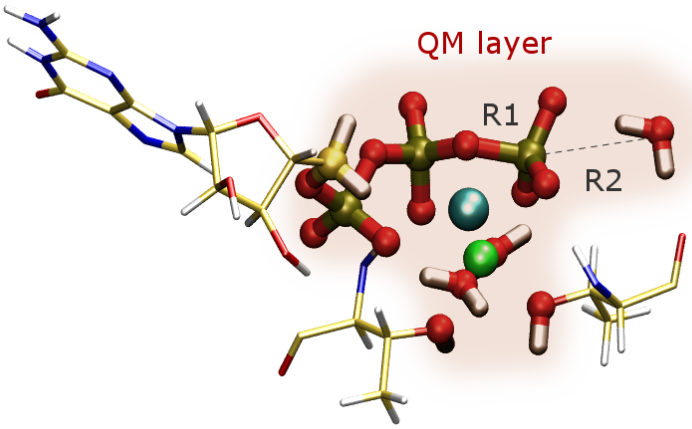


Figure 7.2: The definition of the high layer (the thicker representation in the red cloud). Two threonines 25 and 62 and the guanosine nucleoside are shown to stress the position of cleaved bonds. The missing parts were substituted by a hydrogen atom (link atom) to saturate the dangling bonds. The color code is used consistently throughout this chapter: C yellow, N blue, O red, H white, P golden, Na^+ cyan, Mg^{2+} green.

continued with double-checking the definition by reading an extended output of Gaussian log (required by `IOp(4/33=3)`). If a mistake was made it could be discovered in the output file in the corresponding sections then.

Only when everything was settled and well defined, the system was minimized in the AMBER force field. The minimized structure was then partitioned into layers as specified in the previous section. The ONIOM QM/MM scheme with electronic embedding was used for further study; the geometry optimization procedure used micro-iterations and the quadratic coupling algorithm. The basis set for the QM method was chosen as 6-31G* for the geometry optimization and FES scan as the microiterative procedure usually required many steps to converge. Several interesting surface points were subsequently minimized with a better, 6-311++G**, basis set to obtain free energy differences. This approach is equivalent to the methods used for the previous MDP study in vacuum and solution.

7.2.1 Crystal structure based construct "CSWat"

The first construct that was studied here was based on the 1EFT crystal structure of EF-Tu. For clarity in the following text, it will be referred to as "CSWat". Relevant parts of the protein were extracted and modified as described earlier; the system is depicted in Figure 7.3. Since the crystal structure contained GDPNP, a slowly hydrolysing GTP analogue, the substituted nitrogen atom was switched back to oxygen. The structure was minimized using the Amber force field completed with additional parameters for polyphosphates [66].

The relaxed structure was subjected to minimizations using the ONIOM QM/MM scheme. Several attempts and tweaks to either minimization procedure or structure were needed to find a proper minimum. The initial unsettled structure could easily blow up with too large iterative steps. The "best avail-

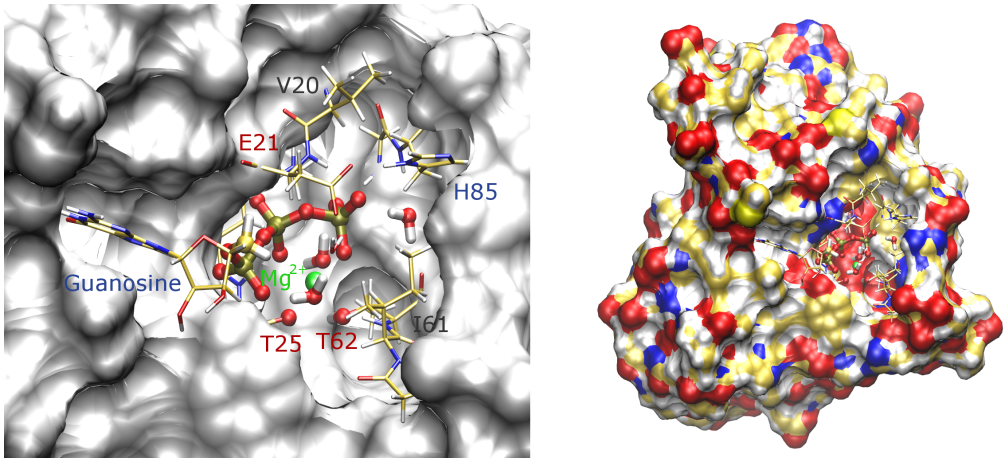


Figure 7.3: The CSWat contruct. The active site with nearby residues and the water molecule substituted by Na^+ in CSNa and MDNa constructs are shown.

able" minimum couldn't be found with too little steps, though. Calculations were repeatedly restarted to find an optimal range for this parameter.

When the energy minimum for the reactants state was found, the FES was uncovered step-by-step using constraints applied to the reaction coordinates. The procedure required several repetitions since the system sometimes changed its conformation and converged to a better energy minimum. This I addressed with a scrupulous scrutiny of resulting structures by means of VMD taking into account RMSD of structures belonging to neighbouring points on FES. In cases where it had happened, all previous surface points had to be recalculated again to preserve consistency. It prolongated the study rather dearly.

The default grid step for the FES scan was set to 0.1Å. In areas, where such detail was not necessary, the step was allowed to be longer. On the contrary, finer grid was scanned around saddle points, which reassured us that the default grid step is sufficient.

The obtained FES can be seen in the Figures 7.4 and 7.5 and in the Table 7.1, which shows two possible reaction pathways. The *concerted* pathway fits nicely with the corresponding MDP pathway. Nevertheless, the other, *associative*, pathway was missing until now. The activation barriers[2] are similar; for the associative pathway it is 28.4 kcal/mol, whereas for the concerted pathway it is 1 kcal/mol less, 27.4 kcal/mol. The barriers along with reaction coordinates and rate constants are also in the Table 7.4.

Analysis of Mulliken atomic charges was made. The magnesium cation was not found to accept electrons as suggested in [61]. It had accepted only $-0.05e^-$ in the intermediate states of associative mechanism and even less in the concerted. Nevertheless, it fed the rest of the system with $0.10e^-$ in the product state.

A significant charge transfer occurred between the β and γ-phosphate groups, though. The donor, γ-phosphate, lost $0.38e^-$ partial electron density in the associative mechanism and even $0.48e^-$ in the concerted. The β-phosphate group was apparently its antagonist as it accepted $0.31e^-$ and $0.43e^-$ partial electron density respectively (became more negative).

Comparison of Mulliken charges in the product state support the idea as the γ-

[2]All presented activation barriers were evaluated with the extended basis 6-311++G**.

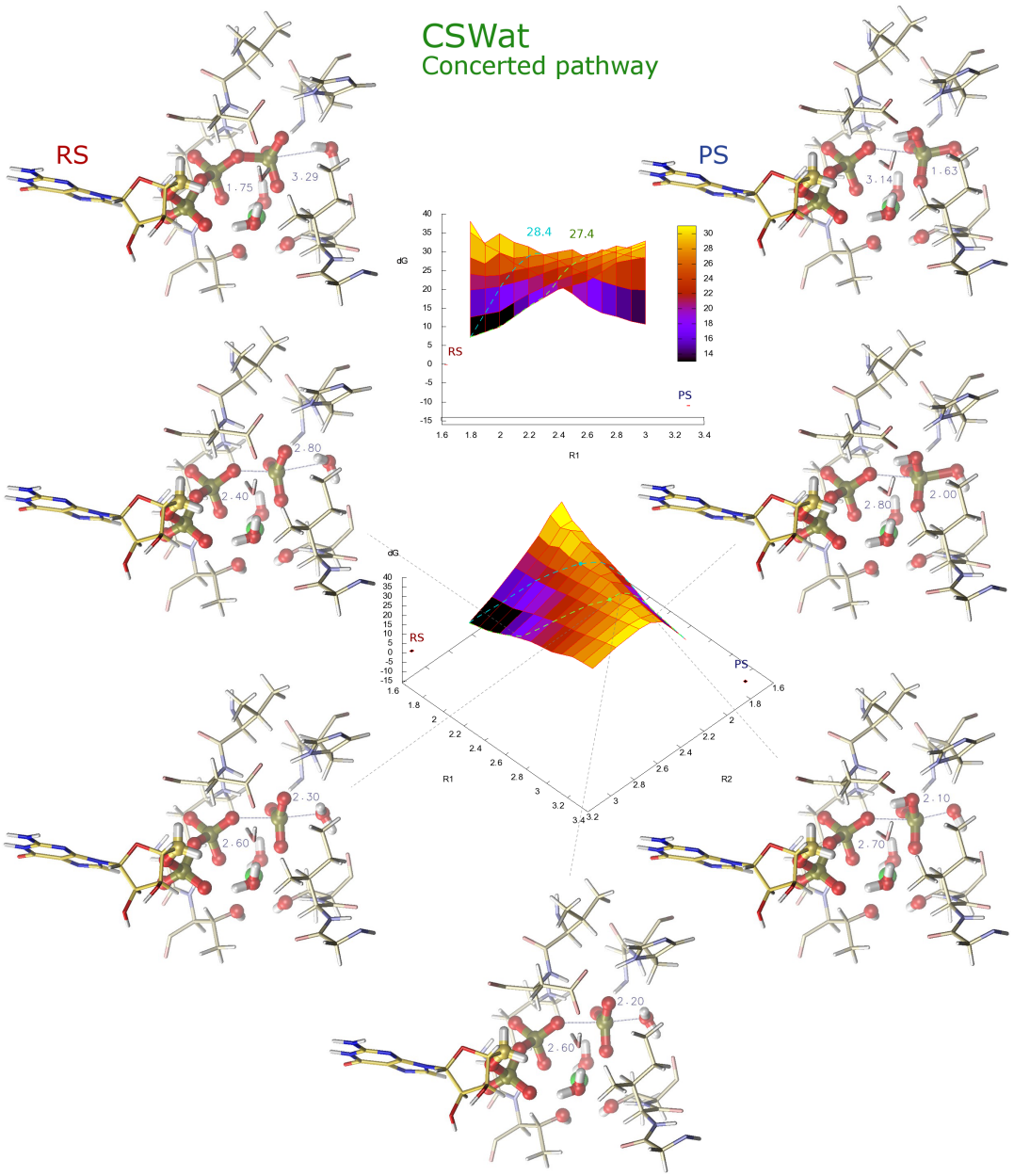


Figure 7.4: The free energy surface for the GTPase reaction in CSWat, concerted pathway. Distances are given in Å, energies in kcal/mol. RS stays for reactant state, PS for product state. Dots denote positions of transition states.

R2 [Å]	R1 [Å]												
	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.00
2.0	36.50	30.33	28.29	25.32	23.84	21.46	19.53	17.31	14.15	12.20	11.24	9.93	9.12
2.1	33.70	30.85	33.31	31.07	**29.90**	**27.88**	26.28	24.18	22.39	20.72	19.47	18.46	18.36
2.2	27.10	25.50	28.43	27.02	**27.38**	**27.82**	28.61	29.25	**27.02**	29.33	28.43	27.46	26.93
2.3	22.91	22.39	22.57	23.77	24.58	25.34	26.47	27.27	**28.12**	28.87	30.43	29.64	30.96
2.4	18.79	19.27	19.74	20.79	21.83	23.23	24.63	25.57	26.51	27.81	29.10	30.46	31.82
2.6	11.8(	12.58	13.36	15.69	18.01	19.88	21.75	23.01	24.27	25.91	27.55	29.13	30.71
2.8	6.79	8.16	9.54	12.25	14.97	17.20	19.43	20.53	21.64	24.16	26.67	27.36	28.05

Table 7.1: ΔG matrix for the crystal structure based study evaluated at the B3LYP/6-31G* level

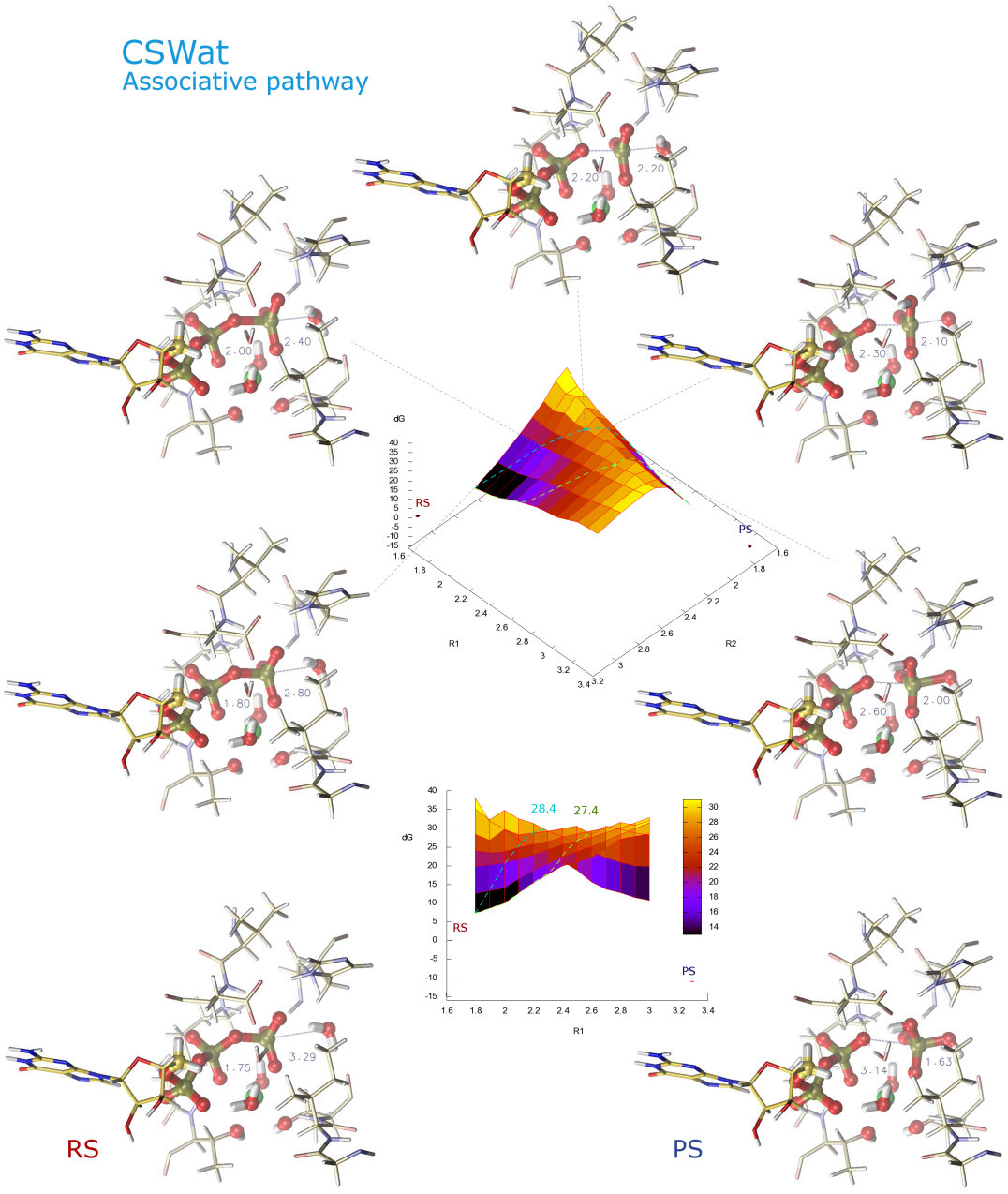


Figure 7.5: The free energy surface for the GTPase reaction in CSWat, associative pathway. Distances are given in Å, energies in kcal/mol. RS stays for reactant state, PS for product state. Dots denote positions of transition states.

phosphate group lost 0.58e$^-$ and the same value was accepted by the β-phosphate group.

7.2.2 Crystal structure based construct with addition of Na^+ "CSNa"

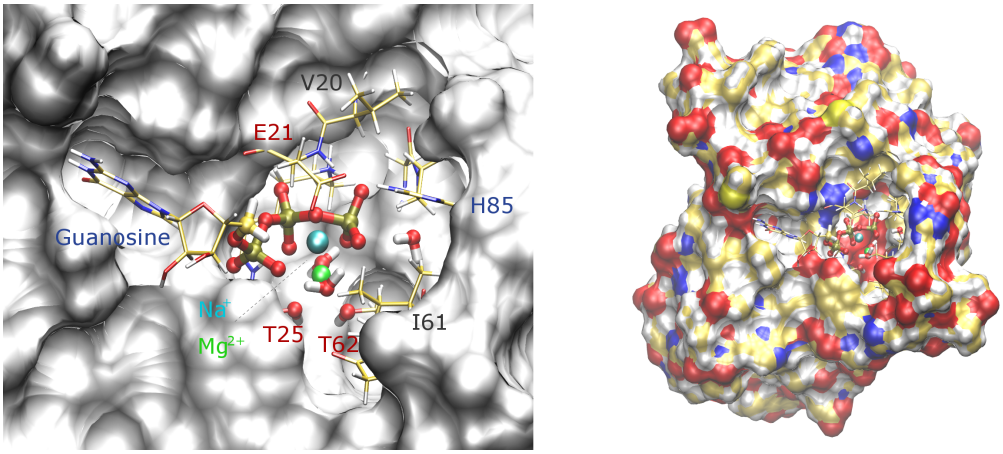


Figure 7.6: The CSNa construct. It differs from CSWat by an exchange of a water molecule with a sodium ion. The active site with nearby residues is shown.

This structure is based on the previous construct CSWat: the reactant state optimized on the ONIOM level was taken and a water molecule from the low layer was exchanged for Na^+ in a place, where it was observed in the MD simulations [11]. The ion was included in the QM layer and no other changes were made. The construct is labelled "CSNa" hence. It can be viewed in the Figure 7.6.

The same procedures as described in the very beginning of this section were undertaken. The obtained FES is in Figure 7.7 and in the Table 7.2. The addition of the ion favors the associative pathway, which activation barrier is 26.8 kcal/mol. The activation barrier for the concerted pathway would be about 1 kcal/mol higher. It is not considered, however, since the FES doesn't contain any evidence of a corresponding saddle point. For comparison of the activation barriers with reaction coordinates and rate constants see also Table 7.4.

The addition of Na^+ also changed the distribution of Mulliken partial charges. It drove electrons from Mg^{2+} in the reactant state making it more positive (from 0.62e$^-$ to 0.91e$^-$). The sodium cation accepted approximately 0.19e$^-$. The both of the cations didn't change their Mulliken charges during the reaction and no significant drift was observed even in the product state.

The addition of Na^+ also made the electron density distribution among the phosphate groups more uniform. The α, β and γ-phosphate groups bore charges of -1.05e$^-$, -1.08e$^-$ and -1.09e$^-$ respectively, whereas without Na^+ the charges were -0.97e$^-$, -0.92e$^-$ and -1.27e$^-$. That suggests that the sodium cation drove approximately 0.17e$^-$ from γ to β-phosphate group providing a little push toward the electron density distribution of the transition state.

The β-phosphate group accepts electrons from the leaving γ-phosphate. The transfer is not so large, though, as a part of the transfer probably occurred in the sodiation. The γ-phosphate looses 0.32e$^-$, whereas the β-phosphate group

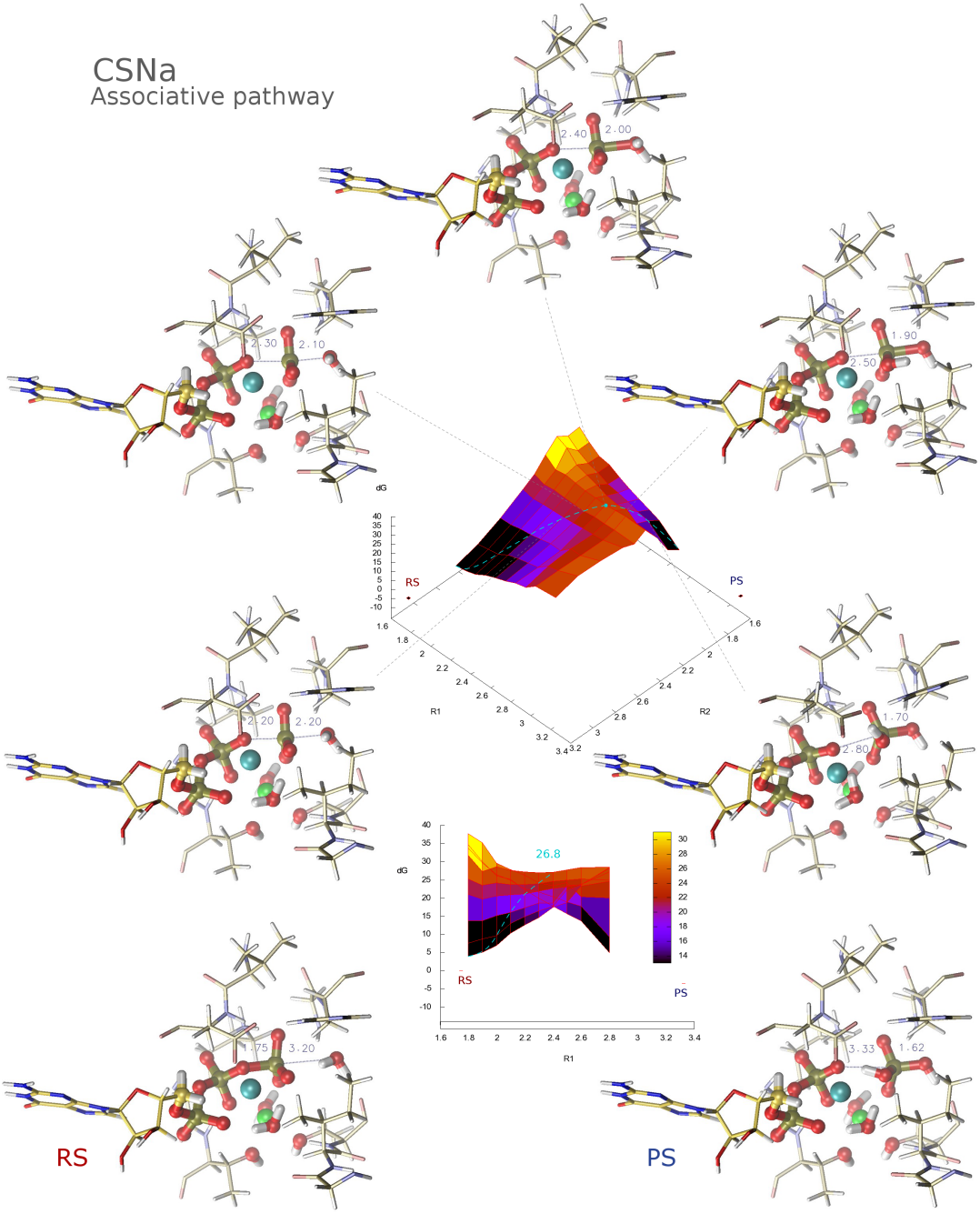


Figure 7.7: The free energy surface for the GTPase reaction in CSNa, which differs from CSWat by an exchange of a water molecule with a sodium ion. The substitution results in stabilization of the pentacoordinated phosphorane intermediate favoring the associative pathway. Distances are given in Å, energies in kcal/mol. RS stays for reactant state, PS for product state. The dot denotes the position of the transition state.

R2 [Å]	R1 [Å] 1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.8
1.7	32.84	28.18	24.28	23.32	20.46	18.08	15.92	13.88	11.97	3.22
1.8	31.97	27.97	24.71	23.27	20.84	19.97	18.72	17.22	16.03	7.64
1.9	36.17	33.68	28.07	26.27	24.72	**22.63**	21.85	22.67	18.80	18.48
2.0	30.35	25.91	27.14	26.20	**25.71**	**25.60**	**25.63**	**26.29**	27.00	27.02
2.1	23.94	22.19	22.85	22.46	22.44	22.61	23.30	23.59	23.95	23.64
2.2	19.87	18.48	19.37	19.68	19.97	20.65	21.49	21.92	22.57	26.00
2.4	12.82	12.79	14.24	15.04	16.38	17.73	17.69	22.46	22.71	27.30
2.6	6.93	8.04	8.91	12.32	14.18	15.75	17.08	21.51	22.93	27.65
2.8	3.58	4.61	6.44	9.85	12.13	14.18	17.02	18.33	22.84	24.14

Table 7.2: ΔG matrix for the crystal structure based study with Na^+ evaluated at the B3LYP/6-31G* level

accepts $-0.29e^-$. If the initial charge transfer was included, the electron loss on the γ-phosphate group would reach $0.49e^-$, which is relatively close to the value without the effect of Na^+, $0.58e^-$.

7.2.3 MD simulation based construct with explicit solvent "MDNa"

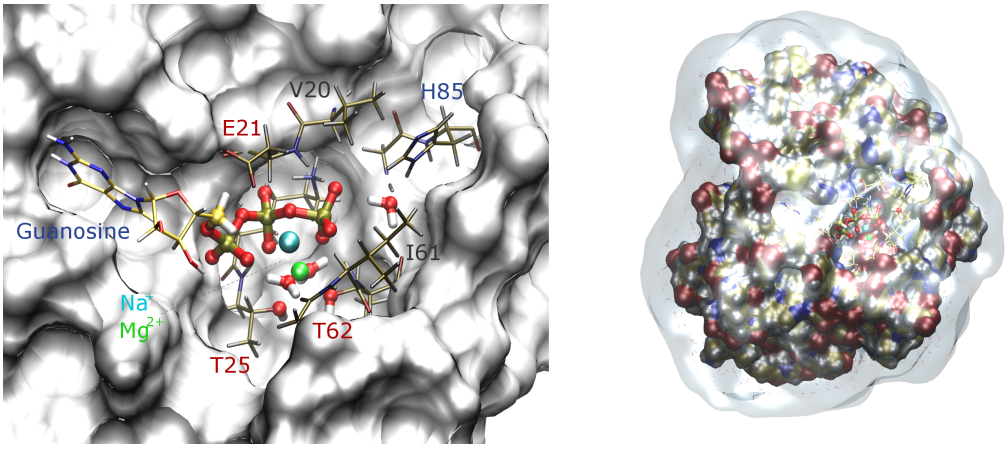


Figure 7.8: The MDNa construct. Apart from the same selection as in the previous constructs CSWat and CSNa, this contains also about 5Å thick layer of explicit solvent. The active site with nearby residues is shown, water molecules are hidden for clarity.

As a part of my bachelor thesis, an MD simulation of EF-Tu was made (section 2.3.1). A frame from the MD trajectory that contained Na^+ bound in the active site of the G domain of EF-Tu was taken and the the same part of the protein as above along with approximately 5Å thick layer of solvent was selected. The solvent contained both water molecules and Na^+. The construct will be referred to as "MDNa", and is depicted in Figure 7.8.

Then it was led through the same preparative and optimizing procedures as described above. Due to the fact that the explicit solvent increased the total number of atoms about twice as much, the study was restricted to a reduced scan

R2 [Å]	R1 [Å] 1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7
2.0	26.87	25.42	24.35	23.55	22.43	21.36	20.56	19.66	17.89
2.1	29.91	29.29	28.86	**28.61**	28.12	27.65	26.82	26.41	25.99
2.2	28.47	28.91	**28.59**	**28.73**	**28.93**	**29.28**	29.60	30.02	30.27
2.3	24.31	25.29	25.46	25.97	26.39	27.01	27.37	28.18	28.99
2.5	18.05	18.96	19.87	21.13	22.38	23.44	24.50	25.78	27.05

Table 7.3: ΔG matrix for the crystal structure based study with Na^+ evaluated at the B3LYP/6-31G* level

of FES so that it was feasible to be done. Since the FES was expected to remain similar in its features to the previous studies, it wasn't hard to guess where the relevant part of the surface would lay.

The free energy surface (Figure 7.9, Table 7.3) suggests that the system chooses the associative pathway, which is favored by about 1 kcal/mol, too. The corresponding activation barrier is then 29.6 kcal/mol. The result follows a similar trend as observed in the MDP hydrolysis study (chapter 6): the solvent elevates the barrier by about 2–3 kcal/mol.

The effect of solvation shifted Mulliken atomic charges comparing with the previous study. The magnesium cation became slightly more positive bearing $0.95e^-$. The water environment affected Na^+ that attracted approximately $0.16e^-$ more electrons (became more negative).

The cations together temporarily stored $-0.08e^-$ in the transition state region. The charge transfer between the γ and β-phosphate groups probably transpired rather directly so. Either way, the β-phosphate accepted $-0.29e^-$ in the product state, a very similar value as in the previous study. The γ-phosphate, however, ended up with $-0.15e^-$ more negative charge – probably a result of both solvation and different protein conformation, since in the transition state it also bore $-0.05e^-$ more negative charge than in the reactant state. That is in contrast with the previous crystal structure based studies, where the γ-phosphate group lost electron density in the transition state regions.

7.3 Discussion and conclusions

The GTPase reaction of Elongation factor Tu was studied in this chapter using the QM/MM ONIOM scheme. Only the G domain of EF-Tu described by the AMBER force field was taken to represent the protein environment. The ligand, GTP, was treated partially with DFT (B3LYP) and partially with the force field. FES for each of the three constructs was scanned and the mechanism of the hydrolytic reaction was analysed. A particular interest was directed towards the effect of Na^+.

The area of FES for scan was chosen accordingly to its shape. As expected, the protein environment restricts the GTPase reaction to a smaller volume than that which applies in solution making the dissociative mechanism unfavourable. For that reason the corresponding area of FES was not scanned. The crystal structure based studies suggested that the system chooses the associative mechanism when Na^+ is present. That had allowed us to reduce the scanned area for the last MD

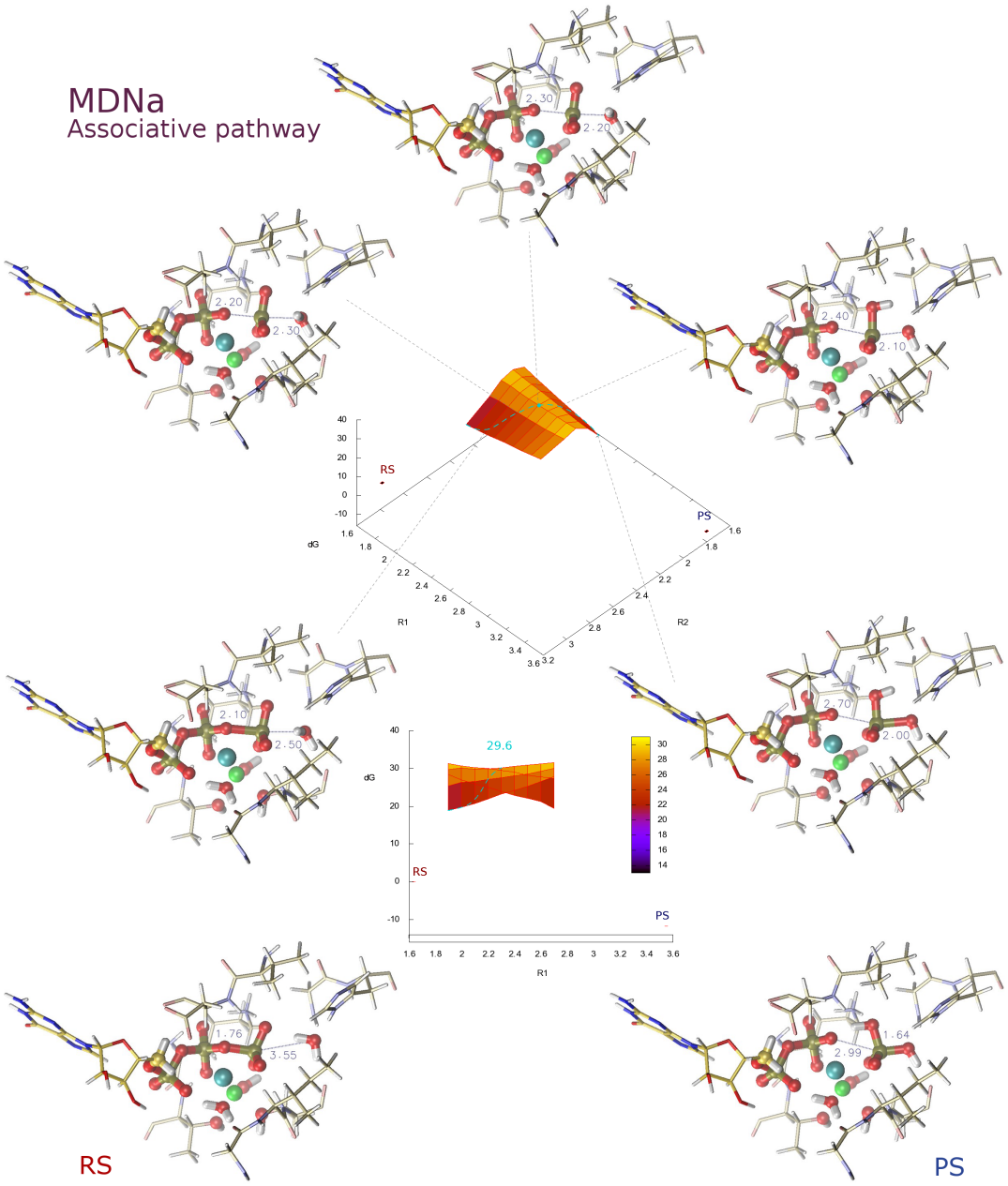


Figure 7.9: The free energy surface for the GTPase reaction in MDNa, which contains about 5Å thick layer of explicit solvent. Distances are given in Å, energies in kcal/mol. RS stays for reactant state, PS for product state. The dot denotes the position of the transition state.

PDB based structure

pathway	R1	R2 [Å]	ΔG [kcal/mol]	k [s^{-1}]
associative	2.2	2.2	28.4	$6 \cdot 10^{-8}$
concerted	2.6	2.3	27.4	$3 \cdot 10^{-7}$

PDB based structure with Na^+

pathway	R1	R2 [Å]	ΔG [kcal/mol]	k [s^{-1}]
associative	2.4	2.0	26.8	$9 \cdot 10^{-7}$

MD simulation based structure with explicit solvent and Na^+

pathway	R1	R2 [Å]	ΔG [kcal/mol]	k [s^{-1}]
associative	2.3	2.2	29.6	$8 \cdot 10^{-9}$

Table 7.4: Activation barriers and rate constants for different structures and reaction pathways

based construct, which generally took more computational time.

The time required for optimization of a single FES point varied. It was significantly longer around the TS regions. On average one such job took about a day to complete on 4 cores of CPU. In the case of the solvated system it took even more. The optimization in the extended basis set was undertaken only for the most interesting FES points, as the required time was even longer. The FESs were composed of tens of points (from 45 to 90), so each FES needed about a month and a half of calculations to be completed.

The activation barriers vary among here presented constructs, but the all of them are reasonably close to the range observed for the GTP hydrolysis in solution: 26–27.5kcal/mol ([67] and [63]). They are still more than 4kcal/mol higher than the observed experimental value of about 22.5kcal/mol of the intrinsic GTPase activity of EF-Tu [16]. In the future it would be valuable to produce EVB MD simulations sampling the entire conformational space between the reactant and product state. Selected frames from EVB MD simulations would then be studied by means of QM/MM methods. This could lead to more realistic results.

The solvation increased the activation barrier by about 2–3kcal/mol, which is a result consistent with the one obtained in the previous chapter 6, where the solvation also led to a similar increase in the activation barrier. Therefore, a deeper understanding of the catalytic effect of the Hydrophobic gate Val20-Ile61 of EF-Tu that apparently de-solvates the gamma phosphate group of GTP would be valuable.

The catalytic effect of Na^+ was mild as suggested in [67]. The QM/MM analysis of the Na^+ effect on the EF-Tu active site is a novel study. Monovalent ions were noted in active sites of various GTPases [65], but none to our knowledge had really examined it. In [65], it was suggested that cation should prevent protonation of GTP making the GTPase reaction more pH independent. This is actually also true for EF-Tu [15].

It should be noted that the three model systems in the current study are hard to compare since EF-Tu adopts different conformers in the crystal- and MD-based constructs. The difference between the corresponding active sites is evident from

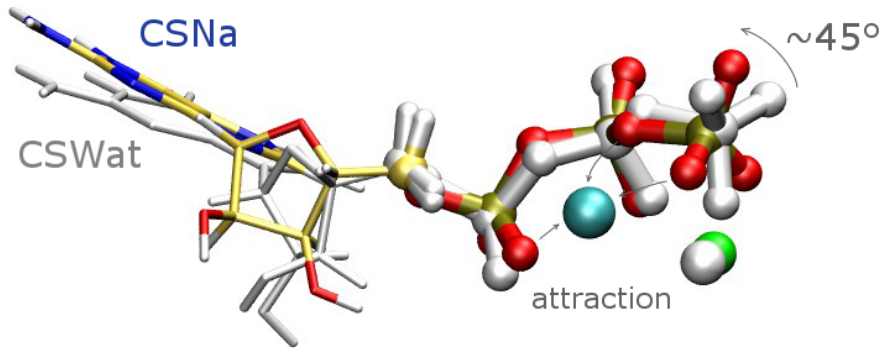


Figure 7.10: Alignment of the CSWat and CSNa constructs. Na^+ attracts negatively charged oxygens of the phosphate groups. The γ-phosphate group rotates by $\approx 45°$ as a consequence.

their representations in Figures 7.7 and 7.9. The active site of MDNa relaxed during the MD simulations so that it forms a longer cavity for the substrate. The His85 residue occupies different positions. In MDNa, it directs towards the hydrolytic water molecule, whereas in CSNa it directs towards an oxygen atom of the gamma phosphate group of GTP. This difference may lead to distinct mechanisms of a proton transfer within reaction events.

The position of Na^+ is equivalent in both models. Nevertheless, the elongated shape of the active site cavity in MDNa allowed Na^+ to settle closer to Mg^{2+} comparing to CSNa. The active site warp also changed positions of Thr25 and Thr62, which slightly shifted the coordination sphere of Mg^{2+} and the arrangement of GTP.

Both classical MD simulations [11] and the CSNa construct agree on that Na^+ rotates the γ-phosphate group of GTP by $\approx 45°$ and that it attracts oxygen atoms from all phosphate groups (Figure 7.10). It was also found that the sodium cation equilibrates charges of all phosphate groups of GTP to be around $-1e^-$. The Na^+ ion presence stabilizes the pentacoordinated transition state favoring the associative reaction path as a consequence. Although the activation barrier was higher in MDNa than in CSNa, this effect was shared by both constructs.

The charge transfer described in the source [61] was not observed here – the Mulliken atomic charges don't show any significant shift that would prove the role of Mg^{2+} as a temporary electron storage. When Na^+ was present in the EF-Tu active site then both cations held their charges almost constant throughout the whole reaction. The source [67] states that the activation barrier for the GTP hydrolysis in solution doesn't change when Mg^{2+} is present or not, which supports the "indifferent" behaviour of cations here. The Mulliken analysis also shows that a negative charge accumulates at the β-phosphate group during and also after the reaction. However, the charge shift from the γ-phosphate group was observed only in those studies that were based on the crystal structure 1EFT. The MD based construct showed opposite behaviour – the γ-phosphate group attracted a small portion of the electron density during the reaction. Either explicit solvent effects or conformational differences may be in charge of this discrepancy. In the future it will be interesting to include more surrounding water molecules as well amino acids side chains into the QM part of the model system.

7.4 Input file example

Before any research could even be started, it was necessary to read through Gaussian input file specifications. Several test jobs were run to find the specific combination of keywords and options that matched our intentions. The most of the options will be introduced in this section.

A general Gaussian input file can be divided into several sections. The input file for the optimization job of the FES point (1.8, 2.0Å) will serve as an example that shall illustrate the options chosen for the studies in the current chapter. The file starts with:

```
%chk=opt-18-20-v3.chk
%nprocshared=4
%mem=8gb
# oniom(b3lyp/6-31g*:amber=SoftFirst)=(MK,embed)
                    opt=(ModRedundant,QuadMacro,MaxStep=20) freq  geom=Check
```

The lines starting with the symbol `%` are the so called *Link 0 commands*. They specify requirements for the job (number of threads and memory) and the filename for the checkpoint file. The checkpoint file contains all information required for the job to be restarted or continued after a crash. It can also be used as a storage of the molecule specification (coordinates, atom types and partial charges, connectivity, link atoms, ...).

The line starting with the symbol `#` is the *Route section*, which has been split in two lines only to fit in the page. It specifies the level of theory, tasks to be done and various options. The QM/MM ONIOM method is required by the `oniom` keyword. The word alone wouldn't make sense, the model chemistries for the high and low layers must also be specified. The example shows the level of theory used for calculation of FES points. The high layer was treated at the DFT level with the B3LYP functional and 6-31G* basis set. The low layer was described by the AMBER force field. The `SoftFirst` parameter allows us to use user-specified *soft* parameters instead of those incorporated in the Gaussian package. Their definition is provided at the end of the file. The `MK` option sets the Van der Waals radii used for RESP charges evaluation [68] to the Merz-Kollman VdW radii instead of the default Universal force field radii. This was taken as a better option since the Merz-Kollman radii were used when deriving the Amber force field [40]. The `embed` requires electronic embedding (section 5.3.1) to be used within the ONIOM scheme.

A geometry optimization procedure is asked by the `opt` keyword. Its options are: `ModRedundant` allows modifications to the internal coordinates (e.g. to freeze them); `QuadMacro` appoints quadratic coupling algorithm for geometry optimizations; `MaxStep=20` sets the maximal optimization step to 20×10^{-2} Bohr or radians. Thermochemistry section is generated by the `freq` keyword. The geometry with atom type definitions, their partial charges, connectivities and link atom definitions are read from the checkpoint file as specified by the last keyword, `geom=Check`.

The next section defines the job name and contains molecule specifications. As the geometry is read from a checkpoint file, the only missing specification is the total charges and multiplicities for the real system, for the high layer treated on the higher (DFT in this example) level and for the high layer treated on the lower level (Amber) in this order.

```
1.8-2.0 opt based on opt-19-20-v3

 -18 1  -2 1  -2 1

B  3161 3162   1.8   F
B  3162 3194   2.0   F
```

The last two lines belong to the `ModRedundant` option of the `opt` keyword. In this example two bonds (`B`) between atoms with numbers 3161–3162 and 3162–3194 are frozen (`F`) at the lengths 1.8Å and 2.0Å.

The last part of the input file belongs to the option `SoftFirst` of the `amber` keyword. It contains missing parameters for triphosphates (taken from [66]), Mg^{2+} (from AMBER force field Parm99 [40]) and the water-like hydroxyl-fragments of threonines 25 and 62 (derived from the internal implementation of the General Amber force field).

```
! from parm99.dat
VDW  MG   0.7926  0.8947
HrmBnd2 OS CT HC    50.0   109.50
!from amber.prm - combined water and CT params - for the water-like fragment of thr
HrmStr1 OH HW     553.0    0.9572
HrmBnd1 HW OH HO    55.0   108.50
!Just substitutions of O2 by O3 as they share parameters (Carlson)
VDW  O3   1.6612  0.2100
HrmStr1 O3 P       525.0  1.480
HrmBnd1 O3 P  OS   100.0   108.23
HrmBnd1 O3 P  O3   140.0   119.90
!Carlson parameters for MTP
HrmBnd2 P  OS P     12.685  150.0
AmbTrs  HC CT OS P    0 0 0 0   0.0 0.0  0.105  0.0      3.0
AmbTrs  CT OS P  OS   0 0 0 0   0.0 0.0 -1.560  0.0      1.0
AmbTrs  CT OS P  O2   0 0 0 0   0.0  -0.812 1.179  0.0      2.0
AmbTrs   P OS P  OS   0 0 0 0   0.897  0.0    0.0    0.0      1.0
AmbTrs   P OS P  O2   0 0 0 0   0.0   -0.709  0.0    0.0      2.0
AmbTrs   P OS P  O3   0 0 0 0   0.0    0.0   -0.255  0.0      3.0
```

7.5 Output file example

The output file contained the same sections as for the MDP study in the previous chapter 6. In addition there was a simple MM-partial-charges summation check:

```
 MM sanity checks:
 All charges sum to:              -18.00000000
 Charges of atoms sum to:         -18.00000000
 Layer 1 (S) has     30 atoms and charge  -2.53300000 sum=  -2.53300000
 Layer 2 (R) adds  3178 atoms and charge -15.46700000 sum= -18.00000000
```

The non-rounded values reflect the fact that there were bonds cut at the boundaries of the layers. The `Layer 1` is more negative due to missing link hydrogen atoms that were omitted from the sum, whereas the `Layer 2` is more positive for a similar reason. The log also informs about the cut bonds:

```
 ONIOM: Cut between C /H   242 and O   244 factor= 0.686800 0.686800
 ONIOM: Cut between C /H   795 and O   797 factor= 0.686800 0.686800
 ONIOM: Cut between C /H  3166 and C  3150 factor= 0.723900 0.723900
```

The contributions from the ONIOM extrapolative energy expression 5.6, chapter 5, are also printed in the output file in atomic units, Hartrees:

```
ONIOM: Microiterations cycle    1 out of a maximum of  25
ONIOM: calculating energy.
ONIOM: gridpoint  1 method:  low   system:  model energy:    -1.434920330886
ONIOM: gridpoint  2 method:  high  system:  model energy: -2398.762948011557
ONIOM: gridpoint  3 method:  low   system:  real  energy:    -9.477123406977
ONIOM: extrapolated energy =   -2406.805151087648
```

It can be checked that the programme really follows the ONIOM energy expression 5.6 by summing the second and the third term together and subtracting the first:

$$(-2398.762948011557) + (-9.477123406977) - (-1.434920330886) = -2406.805151087648$$

7.6 Troubleshooting

Even though the Gaussian software package was intended as a "black-box virtual chemistry lab", preparing and running ONIOM jobs of complex systems is still an arduous task.

The Gaussian manual pages had to be read thoroughly through several times before all options and settings were sufficiently understood.

An example of knowledge that was gained from test studies and is not mentioned in manual pages is that the partial charges appointed to the atoms in the high layer are not used in the optimization procedure. Charges obtained by an electrostatic potential charge fitting procedure are assigned and used for optimization instead.

There was also found an apparent bug that neglects the specified `ModRedundant` options in Gaussian 09 rev. C. The section was fully functional in rev. A that was exclusively used here. The absence of any mention of this neglect in the output file is malicious.

Much time was spent over electronic embedding and its implementation in Gaussian. The manual pages suggest the use of `EmbedCharge` keyword without any alternative. Several test jobs were run on simple systems like a dimer of water, a short peptide or MDP. These model systems had also served for experiments with partitioning, link atom definitions and various settings. With mechanical embedding optimizations evolved reasonably and often converged quickly into configurations similar to those obtained by pure QM (Hartree-Fock) calculations. When electronic embedding was applied by the `EmbedCharge` keyword, however, the optimizations didn't converge in any case. Furthermore, it often led to an explosion of the model system. For example the water dimer got closer in the first few steps and then blew up unexpectedly.

The suspicious behaviour indicated that there was something amiss either in the code or in the model system. Nevertheless, even after several checks no mistake was found in specifications. At last they were compared with an input file of an ONIOM job that successfully used electronic embedding in the past. There the electronic embedding was turned on by a keyword `embed`. After substitution of the previous `EmbedCharge` keyword for this one, the jobs started producing sensible results.

In order to find the difference between the two keywords, several test jobs were run with exactly the same settings but the electronic embedding. No difference was detected among the internal Gaussian options that are printed out in the beginning of every output file.

The outputs were identical and the self-consistent field procedure converged to the same values in both cases. Only the ONIOM high level energy was slightly different; it differed after about tenth decimal place in the test system. The difference was less pronounced when closer MM-partial charges were taken into account (changed by scaling factors defined in chapter 5) and more pronounced otherwise. Our hypothesis is that the `EmbedCharge` keyword does something wrong with the wave function after its evaluation, maybe double-counting something. Since the `embed` keyword matched our intentions and produced sensible results, it was used instead. With awareness the available data were checked for possible errors and for assurance that the procedure does what it is expected to do.

Even when all options were well set, the system blew up occasionally. That led us to consider a restriction for the optimization step value. The default maximal step length is 0.3 Bohr or rad. By experimenting, it was found out that the range 0.16–0.20 works best in our protein-substrate system.

Solvation of the system in an implicit solvent was intended from the very beginning. However, the implicit solvent ONIOMPCM available for ONIOM jobs didn't converge even after a week of permanent computation. For that reason the implicit solvation was abandoned and a layer of explicit solvent was added. That held different problems – a layer of about 8Å thick increased the number of atoms to more than 8 thousand. The problem emerged when thermochemistry was computed. There probably is some internal boundary for the calculation that is not mentioned in the Gaussian manual as such calculations ended up with an error message `NIJ > Max2 in MMCore` without any other explanation. After reduction of the solvent to the level that the system contained "only" 5749 atoms, the calculations went smoothly through.

It was neither our desire nor intention to include almost the entire EF-Tu G-domain in the model system, but it was necessary. In order to describe the active site of EF-Tu properly, at least the surrounding residues needed to be in the places that they naturally occupy. An attempt to include only those parts of the protein that are closer than 12Å from GTP was made. The optimized structure reflected the original protein environment too poorly, however. Since the MM part doesn't slow down ONIOM calculations too much, the entire G-domain was taken into account.

Bibliography

[1] Alberts Bruce, Johnson Alexander, Lewis Julian, Raff Martin, Roberts Keith, Walter Peter, Wilson John, Hunt Tim.
Molecular biology of the cell, fifth edition
ISBN 978-0-8153-4r05, 2008

[2] *Amino acid – Wikipedia, the free encyclopedia* [online]. Cited in March 2013
`http://en.wikipedia.org/wiki/Amino_acids`

[3] Magis Andrew, Chen Ke, Eargle John, Roberts Elijah, Luthey-Shulten Zan. *Evolution of translation, The Ribosome* [online].
Cited in April 2013, released in March 2009.
`http://www.scs.illinois.edu/schulten/tutorials/ribosome/ribosome_tutorial.pdf`

[4] Magis Andrew, Chen Ke, Eargle John, Roberts Elijah, Luthey-Shulten Zan. *Evolution of translation, EF-Tu:tRNA* [online].
Cited in March 2013, released in March 2009
`http://www.scs.illinois.edu/schulten/tutorials/ef-tu/eftu_tutorial.pdf`

[5] *Geneticky kod – Wikipedie* [online]. Cited in March 2013
`http://cs.wikipedia.org/wiki/Geneticky_kod`

[6] Krab Ivo M., Parmeggiani Andrea. *EF-Tu, a GTPase odyssey.*
Biochimica et Biophysica Acta 1443, 1998

[7] Schmeing T. Martin, Ramakrishnan V. *What recent ribosome structures have revealed about the mechanism of translation.*
Nature 461, 2009, p. 1234

[8] Villa Elizabeth, Sengupta Jayati, Trabuco Leonardo G., LeBarron Jamie, Baxter William T., Shaikh Tanvir R., Grassucci Robert A., Nissen Paul, Ehranberg Måns, Shulten Klaus, Frank Joachim. *Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis.*
PNAS 106, 2009, p. 1063

[9] Rodnina Marina V. *Visualizing the protein synthesis machinery: New focus on the translational GTPase elongation factor Tu.*
PNAS 106, 2009, p. 969

[10] Voorhees Rebecca M., Schmeing Martin T., Ramakrishnan V., Kelley Ann C. *The mechanism for activation of GTP hydrolysis on the ribosome.*
Science 330, 2010, p. 835

[11] Melcr Josef.
Bachelor thesis: Molecular dynamics simulations of biomolecular complexes consisting of proteins and nucleic acids.
Physical institute of Charles University in Prague, 2011

[12] Cool Robert H., Parmeggiani Andrea. *Substitution of Histidine-84 and the GTPase mechanism of Elongation factor Tu.*
Biochemistry 30, 1991 , p. 362

[13] Zeidler Waltraud, Egle Christian, Ribeiro Sofia, Wagner Annett, Katunin Vladimir, Kreutzer Roland, Rodnina Marina, Wintermeyer Wolfgang, Sprinzl Mathias. *Site-directed mutagenesis of Thermus thermophilus elongation factor Tu, replacement of His85, Asp81 and Arg300.*
Eur. J. Biochem. 229, 1995, p. 596

[14] Scarano Giuliana, Krab Ivo M., Bocchini Vincenzo, Parmeggiani Andrea. *Relevance of Histidine-84 in the elongation facor Tu GTPase activity and in poly(Phe) synthesis: its substitution by glutamine and alanine.*
FEBS 365, 1995, p. 214

[15] Daviter Tina, Wieden Hans–Joachim, Rodnina Marina V. *Essential Role of Histidine 84 in Elongation Factor Tu for the Chemical Step of GTP Hydrolysis on the Ribosome.*
J. Mol. Biol. 332, 2003, p. 689

[16] Zeidler Waltraud, Schirmer Norbert K., Egle Christian, Ribeiro Sofia, Kreutzer Roland, Sprinzl Mathias. *Limited proteolysis and amino acid replacements in the effector region of Thermus theromophilus elongation factor Tu.*
Eur. J. Biochem. 239, 1996, p. 265

[17] Kraal B., Lippmann C., Kleanthous C. *Translational regulation by Modifications of the Elongation factor Tu.*
Folia Microbiol. 44, 1999, p. 131

[18] Krab Ivo M., Parmeggiani Andrea. *Mutagenesis of three residues, Isoleucine-60, Threonine-61, and Aspartic Acid-80, implicated in the GTPase activity of Escherichia coli Elongation factor Tu.*
Biochemistry 38, 1999, p. 13035

[19] Knudsen Charlotte, Wieden Hans–Joachim, Rodnina Marina V. *The Importance of Structural Transitions of the Switch II Region for the Functions of Elongation Factor Tu on the Ribosome.*
The journal of biological chemistry 276, 2001, p. 22183

[20] Schweins T., Langen R., Warshel A. *Why have mutagenesis studies not located the general base in ras p21.*
Structural biology 1, 1994, p. 476

[21] Baskaran Anand, Parag Surana, Balaji Prakash. *Deciphering the catalytic machinery in 30S ribosome assembly GTPase YqeH.*
PLoS ONE 5(4), april 2010.

[22] Scheffzek K., Ahmadian M. R., Kabsh W., Wiesmuller L., Lautwein A., Schmitz F., Wittinghofer A. *The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras*

mutants.
Science 277, 1997, 333–338

[23] Guangpu Li, Xuejun C. Zhang. *GTP hydrolysis mechanism of Ras-like GTPases.*
J. Mol. Biol. 340, 2004, 921–923

[24] Rittinger K., Walker P. A., Eccleston J. F., Smerdon S. J., Gamblin S. J. *Structure at 1.65 A of RhoA and its GTPase-activating protein in complex with a transition-stateanalogue.*
Nature 389, 1997, 758–762

[25] Daumke O., Weyand M., Chakrabarti P. P., Vetter I. R., Wittinghofer A. *The GTPase-activating protein Rap1GAP uses a catalytic asparagine.*
Nature 429, 2004, 197–201

[26] Menetrey J., Cherfils J. *Structure of the small G protein Rap2 in a non-catalytic complex with GTP.*
Proteins 37, 1999, 465–473

[27] Scrima A., Thomas C., Deaconescu D., Wittinghofer A. *The Rap-RapGAP complex: GTP hydrolysis without catalytic glutamine and arginine residues.*
The EMBO Journal 27, 2008, 1145–1153

[28] Sot B., Kotting C., Deaconescu D., Suveyzdis Y., Gerwert K., Wittinghofer A. *Unravelling the mechanism of dual-specificity GAPs.*
The EMBO Journal 29, 2010, 1205–1214

[29] Scrima A., Wittinghofer A. *Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element.*
The EMBO Journal 25, 2006, 2940–2951

[30] Ash R. M., Maher M. J., Guss M. J., Jormakka M. *The initiation of GTP Hydrolysis by the G-Domain of FeoB: Insights from a Transition-State Complex Structure.*
PLOS ONE 6, 2011

[31] Ash R. M., Guilfoyle A., Clarke R. J., Guss M. J., Maher M. J., Jormakka M. *Potassium-activated GTPase Reaction in the G Protein-coupled Ferrous Iron Transporter B.*
Journal of Biological Chemistry 285, 2010, 14594–14602

[32] Fasano O., Vendittis E., Parmeggiani A. *Hydrolysis of GTP by Elongation Factor Tu Can Be Induced by Monovalent Cations in the Absence of Other Effectors.*
The Journal of Biological Chemistry 257, 1982, 3145–3150

[33] Mistou M. Y., Cool R. H., Parmeggiani A. *Effects of ions on the intrinsic activities of c-H-ras protein p21 – A comparison with elongation factor Tu.*
Eur. J. Biochem. 204, 1992, 179–185

[34] Mishra R., Gara S. K., Mishra S., Prakash B. *Analysis of GTPases Carrying Hydrophobic Amino Acid Substitutions in Lieu of the Catalytic Glutamine: Implications for GTP Hydrolysis.*
PROTEINS: Structure, Function, and Bioinformatics 59, 2005, 332–338

[35] Frenkel D., Smit B. *Understanding molecular simulations.*
Academic Press, New York, 2002

[36] Kapsa Vojtěch, Skála Lubomír, Chen Jian. *From probabilities to mathematical structure of quantum mechanics.*
Physica E: Low-dimensional Systems and Nanostructures 42 (3), January 2010, 293–297

[37] Leach Andrew R. *Molecular modelling - principles and applications.* 2nd release
Essex: Pearson Education Ltd., 2001. ISBN 0-582-38210-6.

[38] Skála Lubomír. *Kvantová teorie molekul.*
Katedra chemické fyziky matematicko-fyzikální fakulty Univerzity Karlovy, Praha, 1944. ISBN 80-7184-007-6

[39] Allen M. P., Tildesley D. J. *Computer simulations of liquids.*
Clarendon Press, Oxford, 1991

[40] Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K. M. Jr., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., Kollman P. A. *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.*
J. Am. Chem. Soc. 117, 1995, 5179–5197

[41] Feig M., MacKerell A. D. Jr., Brooks C. L. *Force field influence on the observation of α-helical protein structures in molecular dynamics simulations.*
Journal of Physical Chemistry B 107, 2003, 2831–2836

[42] Callen H. B.
Thermodynamics and an Introduction to Thermostatistics .
John Wiley & Sons, Inc., New York 1985

[43] Bhandarkar M. et al. *NAMD User's guide* [online]. Cited in March 2013
`http://www.ks.uiuc.edu/Research/namd/2.8b1/ug/`

[44] Kohn W., Becke A. D., Parr Robert G.
Density functional theory of electronic structure.
J. Phys. Chem. 100, 1996, 12974–12980

[45] Parr Robert G., Yang Weitao.
Density-functional theory of atoms and molecules.
Oxford University press, New York, ISBN 0-19-504279-4, 1989

[46] Handy Nicholas C. Becke A. D., Parr Robert G.
The importance of Colle-Salvetti for computational density functional theory.
Theor Chem Acc 123, 2009, 165–169

[47] Pople John A., Gill Peter M. W., Johnson Benny G.
Kohn-Sham density-functional theory wihin a finite basis set.
Chemical physics letters 199 (6), 1992

[48] Becke Axel D.
Density-functional thermochemistry. III. The role of exact exchange.
J. Chem. Phys. 98, 1993, 5648–5652

[49] Raghavachari Krishnan.
Perspective on "Density-functional thermochemistry. III. The role of exact exchange".
Theor Chem Acc 103, 2000, 361–363

[50] Ochterski Joseph W.
Thermochemistry in Gaussian [online]. Cited in April 2013
`http://www.gaussian.com/g_whitepap/thermo/thermo.pdf`

[51] Vreven Tom, Byun K. Suzie, Komáromi István, Dapprich Stefan, Montgomery A. John Jr., Morokuma Keiji, Frish Michael J.
Combining quantum mechanics methods with molecular mechanics methods in ONIOM
JCTC 2006, p. 815-826

[52] Bryun K. Suzie, Komáromi István, Dapprich Stefan, Morokuma Keiji, Frish Michael J.
A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives
Journal of Molecular Structure (Theochem) 1999, p. 461-462

[53] Vreven Tom, Farkas Ödön, Schlegel H. Bernard, Morokuma Keiji, Frish Michael J.
Geometry Optimization with QM/MM, ONIOM, and Other Combined Methods. I. Microiterations and Constraints
J Comput Chem 24, 2003, 760–769,

[54] John Jacob, Frech Matthias, Wittinghofer Alfred. *Biochemical properties of Ha-ras encoded p21 mutants and mechanism of the autophosphorylation reaction.*
The Journal of Biological Chemistry 263, 1988, p. 11792

[55] Cleland Wallace W., Hengge Alvan C. *Enzymatic mechanisms of phosphate and sulphate transfer.*
Chemical reviews 106, 2006, 3252–3278

[56] Allin C., Gerwert K. *Ras Catalyzes GTP Hydrolysis by Shifting Negative Charges from γ- to β-Phosphate As Revealed by Time-Resolved FTIR Difference Spectroscopy.*
Biochemistry 40 (10), 2001, 3037–3046

[57] Kläh Marco, Rosta Ednina, Warshel Arieh.
On the Mechanism of Hydrolysis of Phosphate Monoesters Dianions in Solutions and Proteins
J. Am. Chem. Soc.: 15310-15323, 2006

[58] Thomasi Jacopo, Mennucci Benedetta, Cammi Roberto. *Quantum Mechanical Continuum Solvation Models.*
Chemical reviews 105, 2005, 2999–3093

[59] Klamt Andreas, Schüürmann G. *COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient.*
J. Chem. Soc., Perkin Trans. 2, 1993, 799–805

[60] Pluhařová E., Oncak M., Seidel R., Schroeder C., Schroeder W., Winter B., Bradforth S. E., Jungwirth P., Slavíček P. *Transforming Anion Instability into Stability: Contrasting Photoionization of Three Protonation Forms of the Phosphate Ion upon Moving into Water.*
Journal of Physical Chemistry B, 116 (2012) 13254

[61] Rudack Till, Xia Fei, Kötting Carsten, Shlitter J., Gerwert Klaus. *The Role of Magnesium for Geometry and Charge in GTP Hydrolysis, Revealed by Quantum Mechanics/Molecular Mechanics Simulations.*
Biophysical Journal 103 (2), July 2012, 293–302

[62] Eyring H. *The activated complex in chemical reactions.*
J. Chem. Phys. 3 (107), 1935

[63] Grigorenko Bella L., Rogov Alexander V., Nemukhin Alexander V. *Mechanism of Triphosphate Hydrolysis in Aqueous Solution: QM/MM Simulations in Water Clusters.*
J. Phys. Chem. B 110, 2006, 4407–4412

[64] Glennon Timothy M., Villa Jordi, Warshel Arieh. *How Does GAP Catalyze the GTPase Reaction of Ras?: A Computer Simulation Study.*
Biochemistry 39 (32), 2000, 9641–9651

[65] Kläh Marco, Shlitter J., Gerwert K. *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.*
Biophysical journal 88 (6), 2005

[66] Meagher L. Kristin, Redman T. Luke, Carlson A. Heather.
Development of polyphosphate parameters for use with the AMBER force field
J Comput Chem 24: 1016–1025, 2003

[67] Kotting C., Gerwert K. *Time-resolved FTIR studies provide activation free energy, activation enthalpy and activation entropy for GTPase reactions.*
J. Chem. Phys. 307, 2004, 227–232.

[68] Bayly Christopher I., Cielpak Piotr, Cornell Wendy D., Kollman Peter A. *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model.*
J. Phys. Chem. 97, 1993, 10269–10280

List of abbreviations

RNA	ribo nucleic acid
mRNA	mediator RNA
tRNA	transfer RNA
aa-tRNA	aminoacylated tRNA
EF-Tu	Elongation factor Tu
PTC	peptydyl transferase centre
GTP	guanosine-tri-phosphate
GDP	guanosine-di-phosphate
MDP	methyl-di-phosphate
PDB	protein database `www.pdb.org`
His85	histidine, residue no. 85
E21	aspartic acid, residue no. 21
MD	molecular-dynamics
QM	quantum mechanics
QM/MM	quantum mechanical and molecular mechanical combined method
ONIOM	"Own N-layer Integrated molecular Orbital molecular Mechanics", explained in chapter 5
PCM	polarized continuum model
COSMO	conductor-like solvent model
FES	free energy surface
PES	potential energy surface
CSWat, CSNa and MDNa	specific for chapter 2.3, explained there