Interaction of amino acids and peptides with TiO2 surfaces: adsorption and reactivity

Stefano Pantaleone

Universitat Autònoma de Barcelona

FACULTAT DE CIÈNCES
Departament de Química

Interaction of amino acids and peptides with TiO_2 surfaces: adsorption and reactivity

Ph.D. thesis

Candidate:
Stefano Pantaleone

Thesis Supervisors:
Prof. Mariona Sodupe Roure
Dr. Albert Rimola Gibert

2018

Memòria presentada per aspirar al Grau de Doctor per:

Stefano Pantaleone

Vist i plau,

Mariona Sodupe Roure **Albert Rimola Gibert**

Bellaterra, ...

Contents

Preface

How life began is one of the most intriguing and unsolved problems of science. The events cascade that led to the formation of complex systems such as polymeric macromolecules like proteins, DNA, RNA, sugars, etc., starting from small and simple molecules like H_2O, CH_4, CO, CO_2, NH_3, H_2, is already an object of intense studies. One of the crucial steps of the origin of life is the condensation reaction between amino acids to form peptides. The peptide bond formation reaction is thermodynamically and kinetically unfavourable in gas phase, but also in water. Considering the reaction between two glycine molecules:

$$Gly + Gly \rightarrow Digly + H_2O \tag{1}$$

the equilibrium constant is around 10^{-5}, and it is clear that in a water environment, for the Le Chatelier's principle, the reaction goes back to reagents, since water is one of the products. According to the seminal hypothesis proposed by Bernal, this crucial step for life formation could have seen the light at the interface with mineral surfaces.[1] Therefore, in this thesis we will focus on the interaction of amino acids with TiO_2 surfaces elucidating the peptide bond for-

mation mechanism occurring on these surfaces. Then we analyse the adsorption of secondary structures (α-helices and β-sheets), and the role of the mineral surface in the peptide folding.

Chapter 1

Introduction

In this Chapter we will discuss the concept of "chemical evolution" in the prebiotic era, *i.e.* those processes which led to the chemical complexity of the present world, starting from the relative simple prebiotic Universe. Then we will describe the main features of amino acids and peptides. In Section 1.4 the importance of mineral surfaces in prebiotic era and in particular the TiO_2 will be discussed. Finally, the state of the art of the most interesting works on amino acids and peptides interacting with mineral surfaces will be presented.

1.1 Chemical evolution in prebiotic era

Considering the Big Bang as starting point of the Universe we know nowadays, chemistry was born around 300000 years after the Big Bang, when the universe cooled enough to allow for the formation of atoms and molecules.[2] As depicted in Figure 1.1, taken from Ref. [2], chemistry may be considered as the bridge between the funda-

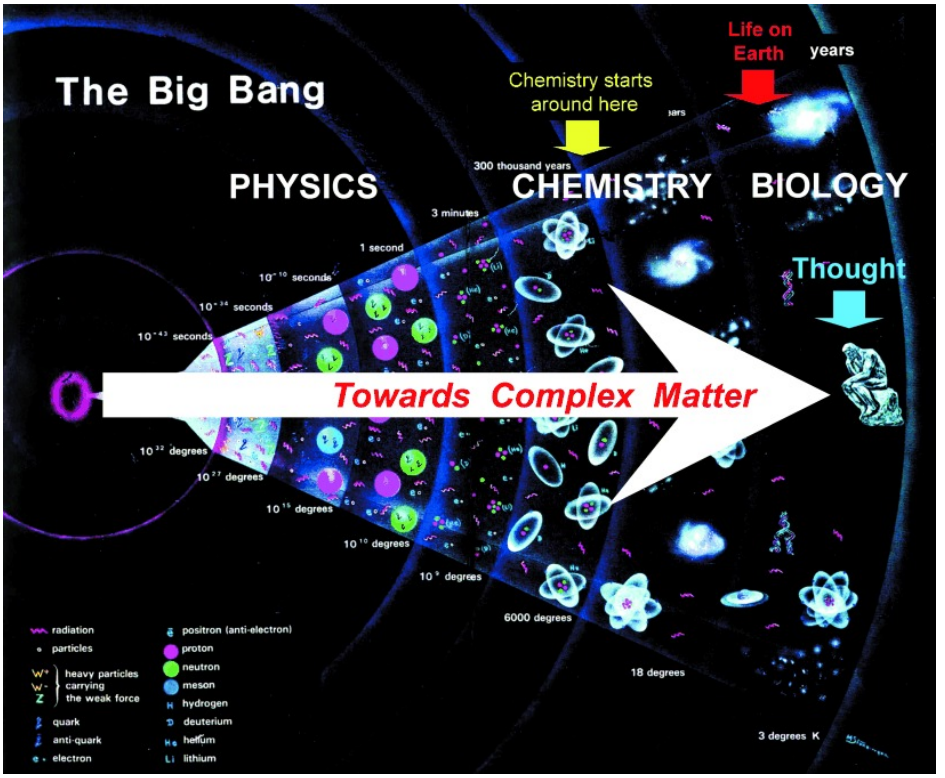


Figure 1.1: Evolution of the universe towards increasingly complex forms of matter: from particles to atoms, to molecules, to life, and to the thinking organism. This figure has been taken from Ref. [2]

mental laws of physics and the most complex forms of matter, *i.e.* living and thinking organisms. Therefore, chemistry accounts for those physico-chemical reactions which occur in the universe, changing the matter towards more and more stages of complexity. The driving force which governs this evolution process is self-organization,[3, 4] *i.e.* the generation of organized molecular and supramolecular architectures by self-assembly in certain environmental conditions. This opens the field to the so-called adaptive chemistry, which studies the response of a system to an external physico-chemical stimulus. In particular, in the present thesis, we will discuss the influence of mineral surfaces and the role they play in amino acid self-assembly. Within this broad field, one of the crucial steps of chemical evolution is the amino acids polymerization, which led to the first small peptides (oligomers) and finally to proteins, one of the macromolecules of life.

Nowadays none knows with absolute certainty the cascade of events and conditions that led to the emergence of life.[5] Many hypothesis have been proposed, from a possible extraterrestrial origin of life, as suggested by Helmholtz and Kelvin, who consider the possibility that comets and meteors transported very simple life systems from the deep space (exogenous theories),[6, 7] to the primordial soup theory,[8–10] an accumulation of organic compounds that, as a result of many reactions, may have led to the formation of more and more complex organic polymers until the emergence of life (endogenous theories).

The first accepted theories about an endogenous growth of life were born at the beginning of the XX century with Oparin and Haldane,[11] who were the pioneers of a heterotrophic hypothesis of

the origin of life. Within the same theory, they proposed two different scenarios. Haldane proposed an atmosphere rich of CO_2 and NH_3 and Oparin one rich in CH_4 and NH_3, the former being less efficient in producing organic compounds.[12] However, since nowadays there are lots of disputes on the real composition of the prebiotic atmosphere and accordingly both interpretations can be accepted. Oparin was the first who proposed a primordial reducing atmosphere and a consequent translation from heterotrophy to autotrophy, leading the atmosphere from an anoxic to an O_2 rich environment. Despite the absence of a geological proof on the existence of the primordial soup, Oparin's theories were supported over years by several experiments, such as those of Miller and Urey[13–15] on the formation of amino acids from simple molecules (H_2O, NH_3, CH_4, CO, N_2) by electric discharges, Oró's ones[16] on the formation of adenine from HCN, and pyrimidines from cyanoacetylene,[17] or the prebiotic synthesis of sugars developed by Butlerow.[18]

Despite the great importance of these milestone experiments, life emergence requires at least four steps, as pointed out by Robert Hazen:[19]

- the emergence of biomolecules
- the emergence of macromolecular systems
- the emergence of self-replicating cycles of molecules
- the emergence of molecular evolution by natural selection

The first step is the simplest one, and it finds positive feedback by the scientific community thanks to the experiments cited above. The

remaining points are object of many controversies. In this context, Hazen and co-workers highlight the importance of considering the role of mineral surfaces in the emergence of life,[19–21] which was already postulated in the 50's by the British physicist J. D. Bernal.[1]

One of the advantages of considering mineral surfaces is that they can protect the adsorbed molecules from the adverse environmental conditions of the prebiotic earth. Moreover, they can strongly interact with biomolecules, concentrate them, thereby activating them toward certain reactions both from geometrically and statistically viewpoints.[22, 23] Finally, the interest for mineral surfaces is also due to the fact that they are important sources of energy for the reactions necessary for chemical evolution and so, the scientific community proceeded over the years with continuous demonstrations and validations of Bernal's hypothesis. Hence, the aim of this work is to study the interactions of biomolecules with an inorganic material which, in the last decades, the scientific community increasingly pays attention for, the TiO_2; this interest was inspired by the experiments of Martra et al. who demonstrate the efficiency of this material with respect to others in catalysing the peptide bond formation among glycine molecules.[24, 25]

1.2 The amino acids

1.2.1 General features

The main feature of amino acids is the presence of carboxyl (–COOH) and amine ($–NH_2$) groups. They can be classified according to the position of the amine with respect to the carboxyl group (α, β, γ, δ),

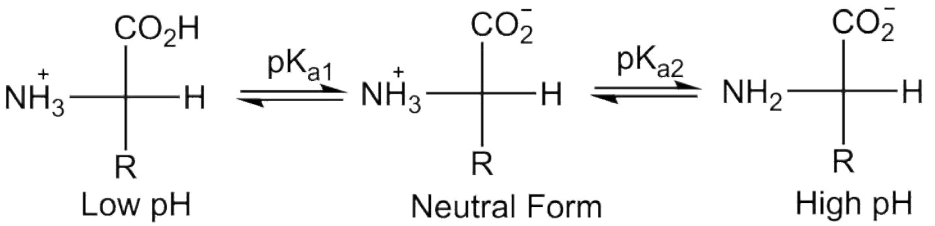
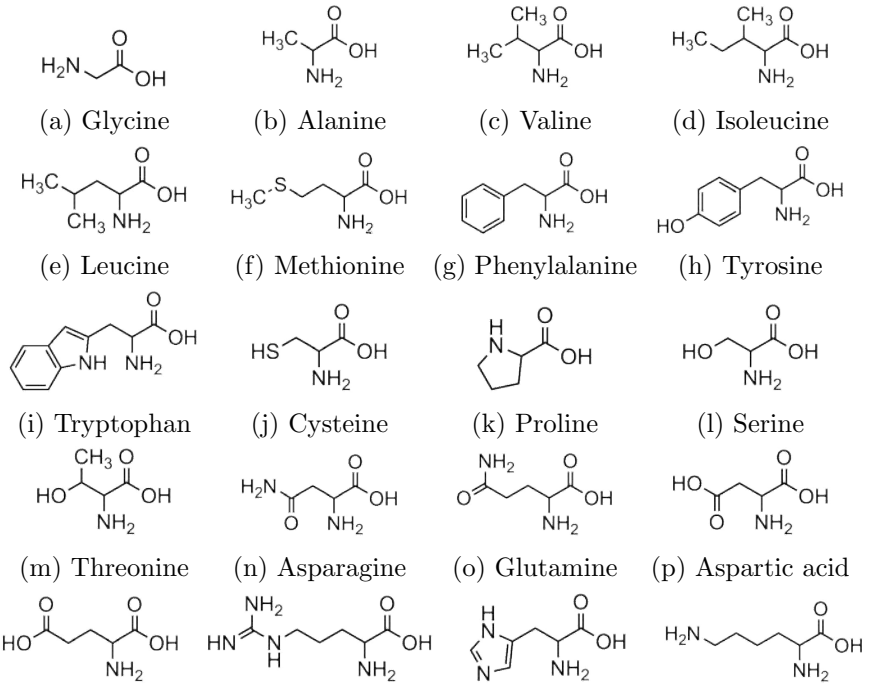


Figure 1.2: pH dependent equilibrium structures for a generic amino acid.

their polarity and the nature of the side chain (aliphatic, aromatic, ...). The general formula of natural amino acids (α-amino acids) is $NH_2CHRCOOH$, where R is the side chain. The amino acid structure in gas phase is the canonical one, *i.e.* the neutral one without any charge separation. In contrast, in water environment, the structure depends on the pH; one of their most evident characteristic is the pK_{a1} value of the COOH group (around 2), higher than generic organic carboxylic acids (around 4.5), due to inductive effects of the amino group in α position with respect to the carboxyl group. At the isoelectronic point (pI) the structure is globally neutral (see Figure 1.2), but with a strong charge separation, because the amino acid is in its zwitterionic form, *i.e.* it presents a formal negative charge on the carboxylic group (deprotonated) and a formal positive charge on the amine group (protonated). At basic pH the carboxylic group deprotonates ($–COO^-$) while the amino group remains neutral ($–NH_2$); while at acidic pH the carboxylic group remains neutral (–COOH) and the amino group is protonated ($–NH_3^+$). Therefore amino acids present at least 2 pK_a: pK_{a1} which guides the protonation/deprotonation of the carboxylic and pK_{a2} the protonation/deprotonation of the amino



(a) Glycine (b) Alanine (c) Valine (d) Isoleucine

(e) Leucine (f) Methionine (g) Phenylalanine (h) Tyrosine

(i) Tryptophan (j) Cysteine (k) Proline (l) Serine

(m) Threonine (n) Asparagine (o) Glutamine (p) Aspartic acid

(q) Glutamic acid (r) Arginine (s) Histidine (t) Lysine

Figure 1.3: Structures of the 20 proteogenic amino acids in their canonical form, *i.e.* with no charged functional groups.

group.

More than 500 amino acids exist in nature, but only 20 of them are present in living organisms (see Figure 1.3), plus other two which are the result of special translation mechanisms; the L-Selenocysteine, *i.e.* a modification of the L-Cysteine with a Se atom instead of the S atom, and the L-Pyrrolisine (not present in humans).

All α-amino acids but glycine (which is not chiral) may exist in two enantiomeric forms, the L and D, depending on the absolute configuration of the α-carbon; however only L-amino acids can be found

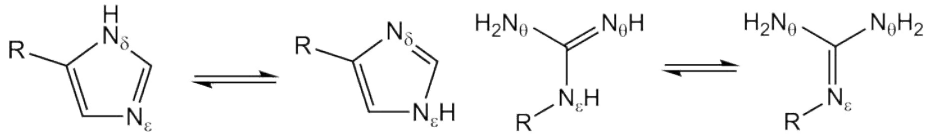
in proteins. Therefore, in human body there are only 19 amino acids, and only in their L enantiomeric form (plus Gly which is not chiral); the origin of this strong selectivity nowadays is still unknown. Despite the small number of these fundamentals building blocks, since now around 18 000 proteins have been detected,[26, 27] but the Human Proteome Project (HPP), an international project organized by the Human Proteome Organization (HUPO), has predicted more than 500 000 undetected proteins.

Amino acids are not only important for their function in proteins, but also as precursor of neurotransmitter and other molecules: for example, Tryptophan is a precursor of serotonin, Tyrosine of dopamine, epinephrine and norepinephrine, Phenylalanine of Tyrosine, Glycine of porphyrins like heme, etc., so their importance is very wide.

1.2.2 Side chains

One of the parameters to classify amino acids is the polarity, which is defined by their lateral chains. In particular, considering the 20 natural proteinogenic α-amino acids, we can distinguish among non-polar amino acids (Glycine, Alanine, Valine, Leucine, Isoleucine, Proline, Methionine, Phenilalanine), polar amino acids (Serine, Cysteine, Threonine, Tyrosine, Tryptophan), polar/acidic (Aspartic Acid, Glutamic acid), polar/amidic (Asparagine, Glutamine) and polar/basic (Lysine, Histidine, Arginine).

Obviously different side chains lead to different properties of the amino acid. One of the main differences among the amino acids is that some of them can have more than two pK_a, depending on the lateral chain: in particular, polar/acidic and polar/basic amino



(a) Histidine (b) Arginine

Figure 1.4: Histidine (left) and Arginine (right) tautomeric forms. Here, the functionality of the side chain is only shown, R being the rest of the amino acid.

acids have three pK_a, because also the lateral chain can be protonated/deprotonated as a function of the environmental pH. The lateral chain has also strong effects on the main pK_a values: Histidine (pK_{a1}=1.82) and Tryptophan (pK_{a1}=2.83) differ in one order of magnitude in their pK_{a1}, while Cysteine (pK_{a2}=8.18) and Proline (pK_{a2}=10.60) differ by more than two orders of magnitude of difference in their pK_{a2}. At physiological pH, the polar/acidic amino acids exhibit a global negative charge, because also the lateral chain is deprotonated, while polar/basic amino acids have a global positive charge, because the lateral chain is protonated, with the only exception of Histidine, whose lateral chain is not protonated at pH=7.4 (pK_{a3}^{HIS}=6.00); this is due to the fact that the lone pair of the N of the imidazole is engaged in the resonance forms, and it is not so available to accept protons. Among polar/basic lateral chains, Histidine and Arginine present two slightly different structures because of the tautomeric forms;[28, 29] Histidine may be protonated at N_δ or at N_ϵ, and Arginine can have the imino group on the N_ϵ or on the N_θ, as shown in Figure 1.4.

One of the most peculiar reactions that may occur among the several lateral chains is the disulphide bridge between two Cysteines to

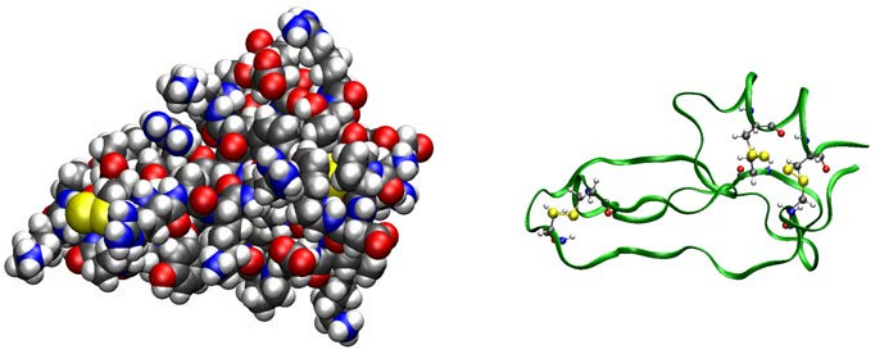


Figure 1.5: Van der Waals (left) and ribbon with the Cysteine residues highlighted in ball and stick (right) representations of the BPTI protein.

form a S–S bond after the oxidation of the –SH groups. This bond is the strongest that a protein can possess (60 kcal/mol) and is responsible for stabilizing and holding proteins, and so for their structure.[30] One of the most famous examples of this phenomenon is the BPTI (bovine pancreatic trypsin-inhibitor, see Figure 1.5); the degradation temperature of the protein drops down drastically with the reduction of the number of disulphide bridges.[31] It is clear that the formation of S–S bonds depends on the environmental conditions: in the internal part of the cells the environment is reducing and the sulphidric groups tend to remain in the –SH form, while in the external part, in an oxidizing environment the S–S bridges are stabilized.

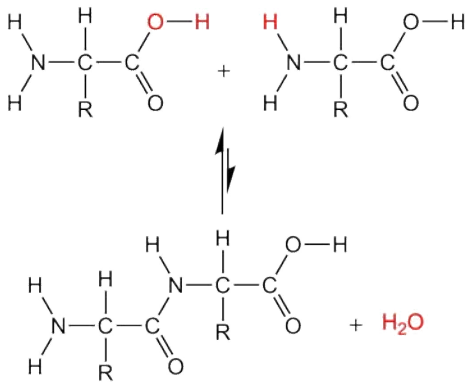


Figure 1.6: Peptide bond formation reaction.

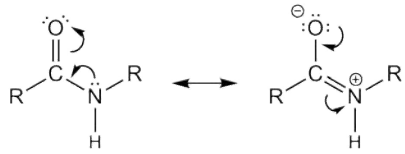


Figure 1.7: Peptide bond resonance forms.

1.3 Peptides

1.3.1 The peptide bond

The peptide bond is formed by a condensation reaction between two amino acids, *i.e.* the formation of a new C–N bond and the elimination of a water molecule, as depicted in Figure 1.6. The peptide bond is almost planar because of the presence of a very important resonance form shown in Figure 1.7.

In gas phase the condensation reaction is kinetically unfavourable because of its high energy barrier to overcome. In water environment the kinetic barrier drops down by some amount thanks to the assistance of the other water molecules, but the reaction is thermodynamically unfavourable (the ΔG of the reaction is positive) and, as one of the reaction products is H_2O, for the Le Chatelier's principle the reaction goes back to the reactants. However, the hydrolysis of the peptide bond (the reverse reaction) is very slow in absence of an appropriate catalyst, because the peptide bond is very stable, and its cleavage is not so easy.[32] The sequence of amino acids involved in a polypeptide chain is called primary structure.

1.3.2 The secondary structure

When a polypeptide chain folds, different parts of the primary structure interact to each other to form secondary structures. The main forces responsible of these interactions are H-bonds between the H atoms of —NH groups and O atoms of the –CO groups, and van der Waals dispersive forces. This kind of interactions, although individually are weak, are amplified in a protein of thousands of atoms, so

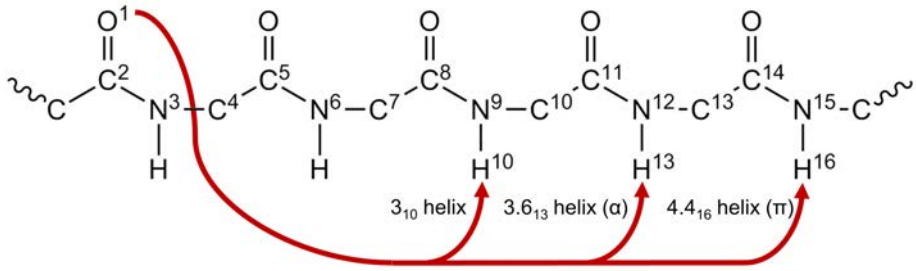


Figure 1.8: H-bond formation scheme for the different helices.

that their global contribution is crucial for the folding of the protein itself. As the rotation around the peptide bond is hindered because of its double bond character, all the possibilities of secondary structures are focused on the rotations of the two dihedral angles around the C_α, *i.e.* C(O)—C_α—N—C(O) (ϕ) and N—C_α—C(O)—N (ψ).

The α-helix

Pauling and coworkers, in the first half of the XX century, proposed the most probable values of ϕ and ψ compatible with stable secondary structures. Values around ϕ = -57° and ψ = -47° lead to the α-helix, one of the most frequent secondary structures in proteins. By convention, a helix is defined by two numbers n_N, where n is the number of residues per turn and N identify the number of atoms of the backbone in a pocket closed by a H-bond (counting the O and H atoms of the H-bond). For the α-helix these two values are indicated as 3.6_{13}; it means that each carbonyl O forms a H-bond with the amidic H of the fourth residue (see Figure 1.8). There are other possible helix structures, but not so common as the α-helix:

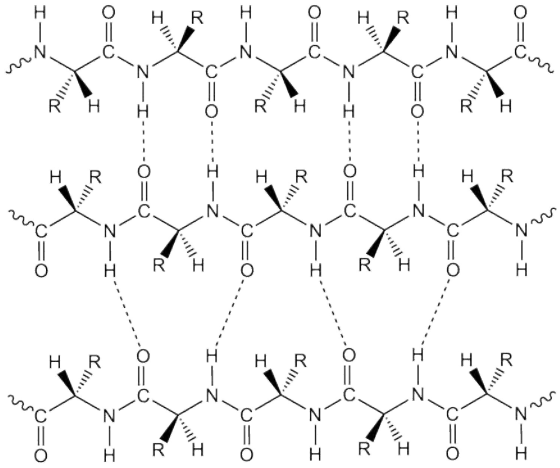


Figure 1.9: Antiparallel (top two β-strands) and parallel (bottom two β-strands) β-sheet.

the 3_{10} helix (*i.e.* it possesses 10 residues per turn, and H-bonds every three residues) and the π-helix (4.4_{16}, *i.e.* H-bond every five residues). The values of ϕ and ψ for the 3_{10} helix are -49° and -26°, respectively, while for the π-helix they are -57° and -70°, respectively. From a steric point of view, they may exist, but they are not so stable as the α-helix, and the π-helix has never been observed in any protein yet.

The β-sheets

Another very common structure that can be found in proteins is the β-sheet; it is formed by at least two polypeptide chains (β-strands), which can be with either the same N/C orientation (parallel β-sheet) or opposite N/C orientation (antiparallel β-sheet), as shown in Figure 1.9. The two (or more) chains are linked by H-bonds between the

carbonyl O and the amidic H.

The values of ϕ and ψ are different from the helices; that is, -139° and +135°, respectively, for the antiparallel β-sheet and -119° and +113°, respectively, for the parallel β-sheet. Antiparallel β-sheets are more stable than parallel β-sheets because of the different strength of the H-bond pattern; as one can see in Figure 1.9, in the antiparallel conformation the H-bonds are more directional and shorter ($\sim$2.10 Å) than those in the parallel conformation ($\sim$2.62 Å).

Structural motifs

The combination of different simple secondary structures lead to more or less complex motifs. For example helices may wrap to each other to form complex and beautiful structures like collagen, α-keratin, and, going to larger scale assemblies, also myoglobin and hemoglobin proteins. β-sheets may form simple structures like β-hairpins, greek key motifs, β-α-β motifs, *i.e.* a parallel β-sheets interplayed by an α-helix, β-meander motifs. *i.e.* two or more consecutive antiparallel β-strands, psi-loop motifs, *i.e.* two antiparallel β-strands interplayed by another β-strand.

1.4 Mineral surfaces

Surface science is the study of those phenomena that occur at the interface between a solid phase surface and the vacuum (a bare surface), the solid surface and one or more gas phase molecules, or the solid surface and a liquid. This field has been more and more studied during the last years, because of its large application in science, and

in the particular field of chemistry, in catalysis.[33] From a theoretical point of view bare surfaces are basically bulk materials where one of the three periodic direction has been removed; the new system has a 2D periodicity and one finite dimension (the slab thickness). The most important feature of surfaces is their reactivity; this is due to the fact that the outermost atoms have lost part of their chemical environment, and, accordingly, they will have new properties and a different reactivity with respect to the bulk atoms. There are many examples of the surfaces used in industrial catalysed processes; the synthesis of NH_3 on iron-doped surfaces,[34, 35] the oxidation of SO_2 to SO_3 for the production of H_2SO_4 catalysed by noble metals like Pt, Ag or Au,[36] or the conversion of NH_3 to NO for the synthesis of HNO_3 on Pt surfaces.[37] Surface science has been applied since many years to heterogeneous catalysis, as one can see from the reactions mentioned above. In the last decades many studies have also focused on the interaction of biological molecules with mineral surfaces due to their application in many scientific fields; a clear example is the nanomedicine[38] which allowed the development of biocompatible materials such as titanium bridges and bioglasses for dental and orthopedic implants.[39, 40]

1.4.1 Mineral surfaces in prebiotic era

As previously said in Section 1.1 mineral surfaces are important in chemical evolution events. Talking about the prebiotic era, one may think that we are considering only Earth. However, the prebiotic era started with the formation of the Universe. In this context, the field of astrochemistry studies the reactions which take place in the

interstellar medium. In the interstellar medium low temperatures and the lack of matter, 1 up to 10^6 atoms per cubic centimeter (a cubic centimeter in the room has about 10^{19} atoms), hinder molecular reactions from a kinetic point of view.[41] Hence the importance of a solid scaffold; at their interface many reactions, kinetically or thermodynamically hindered in gas phase, occur. For example the formation in the interstellar medium of small molecules like H_2 on Fe^{2+} doped olivine,[42–44] and NH_3 on pyrite[45] have been studied.

For the purposes of the present thesis, we will discuss the minerals likely present in a prebiotic Earth and their evolution. Talking about evolution, our mind instinctively focuses on living systems, *i.e.* their changes through increasing complexity; however, evolution is a process that involves all the components of the world (living and non-living).[46] As a result, Hazen et al. summarise the mineral evolution into three broad eras.[20] The first era spans from more than 4.56 to 4.55 Ga ago, *i.e.* the period of the earth formation, from the mineralogical simplicity of the pre-stellar dense molecular clouds (few tens of mineral species), to the formation of the young earth (few hundreds of mineral species). In the second era, from 4.55 Ga to 2.55 Ga, the number of minerals grew up to more than one thousand, due to different physical and chemical processes, such as crystallization and melting, vertical tectonic, subduction and collisional zones, hydrothermal activity, etc. The third era, from 2.55 Ga to current days, is associated with the biological activity, *i.e.* a co-evolution of the geo- and the biospheres occurred (and still now occurs), leading to more than 4300 known species of minerals. The different stages of the mineral evolution, with some examples of the most important minerals are summarized in Table 1.1.

Table 1.1: Ten stages of mineral evolution of terrestrial planets, with possible timing on Earth, examples of minerals, and estimates of the cumulative number of different mineral species. This Table has been taken from ref.[20]

Stage	Age (Ga)	Examples of minerals	∼Cumulative n° species
		The era of planetary accretion (>4.55 Ga)	
1. Primary chondrite minerals	>4.56 Ga	Mg-olivine/pyroxene,Fe-Ni metal, FeS, CAls	60
2. Planetesimal alteration/differentiation	>4.56 to 4.55 Ga		250
a) aqueous alteration		phyllosilicates,hydroxides,sulfates,carbonates,halite	
b) thermal alteration		albite, feldspathoids, biopyriboles	
c) shock phases		ringwoodite, majorite, akimotoite, wadsleyite	
d) achondrites		quartz, K-feldspar, titanite, zircon	
e) iron meteorites		many transition metal sulfides and phosphates	
		The era of crust and mantle reworking (4.55 to 2.5 Ga)	
3. Igneous rock evolution	4.55 to 4.0 Ga		350 to 500
a) fractionation		felspathoids, biopyriboles (volatile-poor planets)	350
b) volcanism, outgassing, surface hydration		hydroxides,clay minerals (volatile-rich planets)	500
4. Granite formation	4.0 to 3.5 Ga		1000
a) granitoids		quartz, alkali feldspar (perthite), hornblende, micas, zircon	
b) pegmatite		beryl, tourmaline, spodumene, pollucite, many others	
5. Plate tectonics	>> 3.0 Ga		1500
a) hydrothermal ores		sulfides, selenides, arsenides, antimonides, tellurides, sulfosalts	
b) metamorphic minerals		kyanite, sillimanite, cordierite, chloritoid, jadeite, staurolite	
6. Anoxic biological words	3.9 to 2.5 Ga		1500
a) metal precipitates		banded iron formation (Fe and Mn)	
b) carbonates		ferroan carbonates, dolostones, limestones	
c) sulfates		barite, gypsum	
d) evaporites		helides, borates	
e) carbonate skarns		diopside, tremolite, grossularite, wollastonite, scapolite	
		The era of bio-mediated mineralogy (>2.5 Ga to present)	
7. Paleoproterozoic atmospheric changes surface oxidation	2.5 to 1.9 Ga	>2000 new oxide/hydroxide species, especially ore minerals	>4000
8. Intermediate ocean	1.9 to 1.0 Ga	minimal mineralogic innovation	>4000
9. Neoproterozoic biogeochemical changes	1.0 to 0.542 Ga		>4000
a) glaciation		extensive ice deposition, but few new minerals	
b) post-glacial oxidation		extensive oxidation weathering of all surface rocks	
10. Phanerozoic Era	0.542 Ga to present		4300+
a) biomineralization		extensive skeletal biomineralization calcite, aragonite, dolomite, hydroxylapatite, and opal	
b) bio-weathering		increased production of clay minerals, soils	

The first stage of Table 1.1 represents the minerals present in the so called "dense molecular clouds" in the interstellar medium. They are nano and micrometer-sized nanoparticles formed several minerals such as Graphite (C), some silicates such as Forsterite (Mg_2SiO_4) Perovskite ($MgSiO_3$) but also orthopyroxene, augite and anorthite, which include in their structures other cations (Fe^{2+}, Ca^{2+}, Al^{3+}). Moreover there are nitrites (TiN, Si_3N_4), sulfides (FeS, CaS), and pure metal alloys (Fe-Ni). Notably, also TiO_2, the material of interest for the present work, was present, in particular in the form of the rutile polymorph. The aggregation of these nanoparticles, driven by gravitational forces, led to the formation of stars and planets. During the second stage, when planets have already formed, many new minerals appeared, as a direct consequence of the adverse environmental conditions of the early Earth. Among many, talc ($Mg_3Si_4O_10(OH)_2$), halite (NaCl), quartz (SiO_2), potassium feldspar ($KAlSi_3O_8$), zircon ($ZrSiO_4$) have been formed in this period. Moreover, meteor bombardments further enrich the number of mineralogic species in this era, providing Earth with a source of reactive phosphorous $(Fe,Ni)_2P$ and $(Fe,Ni)_3P$, another element mandatory for life generation.

During the second era, when life probably began, there was an exponential increasing of mineral species, due to the crust and mantle reworking, caused by volcanism, outgassing phenomena, and plate tectonics. These phenomena led to sudden changes of temperature and pressure which are responsible of stabilizing new phases from existing minerals, but also of the emergence of new species by melting or pressuring already existing minerals. At last, the final stage of this era is driven by the generation of new minerals due to the interface with the first life-forms. Microorganisms are prob-

ably responsible of many sedimentary rocks (Banded Iron Formation (BIF) precipitation), because of biological oxidation processes of Fe^{2+} and Mn^{2+}. Among BIFs there are many known oxides such as hematite (Fe_2O_3), magnetite (Fe_3O_4), sulfides such as pyrite (FeS) and chalcopyrite ($CuFeS_2$), and carbonate such as siderite ($FeCO_3$), Fe-dolomite [Ca(Mg,Fe)CO_3], and calcite ($CaCO_3$).

Finally, during the last era a further increasing (more than in the second one) of mineral species occurred (more than 4000), mostly due to atmospheric changes, and in particular oxidation phenomena (the so called "Great Oxidation Event"). The increase of atmospheric oxygen was caused by photosynthesis of cyanobacteria, and is responsible of more than 2000 new oxides and hydroxides. The mineral evolution still occurs, and it is estimated that around 50 types of new minerals are identified each year.

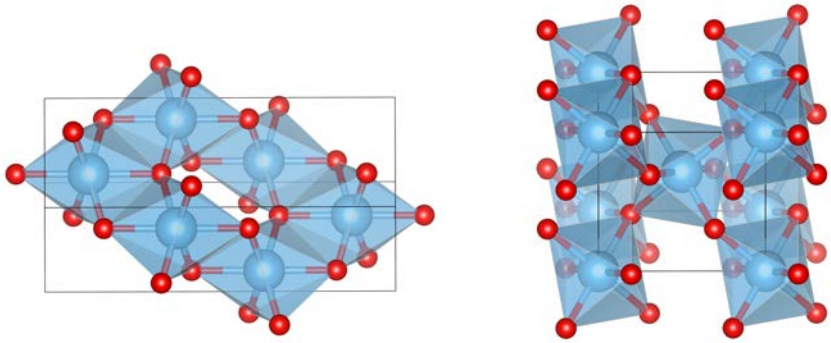
This work is focused on the context of the interface between amino acids and mineral surfaces, and the role they have played in the amino acid polymerization, one of the key step for life emergence. One of the pioneers of this field was John Desmond Bernal, who advocated the key role of clays in life's origin.[1] Among the several ingredients necessary to build up a living systems, only lipids can self-organize in a water environment.[47–51] The other biological macromolecules, such as DNA, RNA, carbohydrates, proteins, tend to break down in water environment.[52] By contrast, mineral surfaces can concentrate those monomers useful to create the previously mentioned macromolecules,[53–57] protect them from adverse environmental conditions (lightning storms and ultraviolet radiation may break down biopolymers), and also activate them for those reactions forbidden in gas phase. This is possible thanks to the surface electro-

static charges, ubiquitously present in each mineral, which allow the adsorption of organic molecules. Many experimental evidences along the years supported Bernal's hypothesis.[58–65]

1.4.2 The TiO_2 material

It is well known that the most abundant component in the outermost part of the earth is SiO_2 (over 50%) followed by Al_2O_3 (more than 15%) and Ca, Mg and Fe oxides (less than 10% each one).[66] Although TiO_2 is not the major part of the earth crust (around 1%),[66] its abundance does not have to be very high to be effective in prebiotic processes, acting as catalyst. As discussed in Section 1.4.1, TiO_2 was present in the Archean Era, and in particular the anatase polymorph, which has been detected in several meteorites,[67, 68] and also on the earth, in zone of ancient asteroid impacts.[69]

Titanium dioxide crystallizes in several polymorphs: rutile, anatase, brookite, columbite, baddeleyite, cotunnite, pyrite, fluorite, and tridymite.[70] Most of these phases are stable in particular conditions of high pressure, but only rutile, anatase and brookite can be commonly found in nature. TiO_2 has been extensively studied in the last years because of its broad application in many fields of science. One of the applications of this material is as photocatalyst[71] for water splitting (into H_2 and O_2);[71–73] despite its low quantum yield in the conversion of solar energy,[74] the addition of dye molecules improves the efficiency of energy production.[75] It is also used for photoassisted degradation of organic molecules, the applications of which span from purification of water,[76] disinfection,[77] self-cleaning coating of can windshields,[78] and protecting coating of



(a) Anatase (b) Rutile

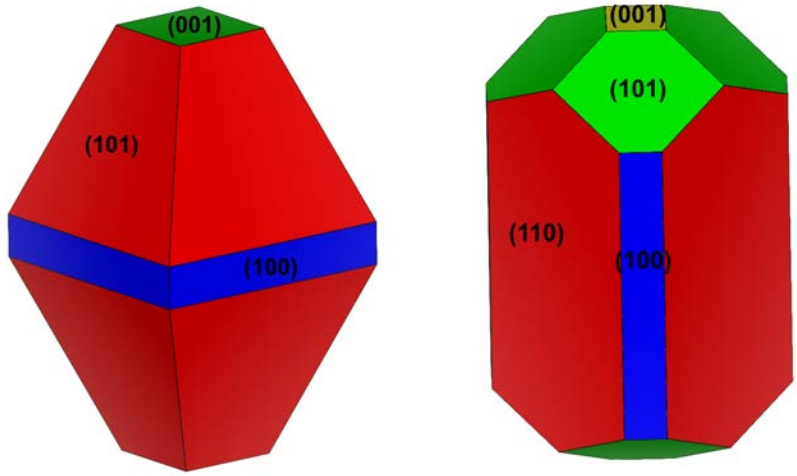
Figure 1.10: Crystallographic bulk structures of the two polymorphs considered in the present thesis.

marble.[79] Moreover, it is a biocompatible material; thus it is used in medicine for dental bridges and bone implants.[80–82]

In particular, in the present thesis only the anatase and rutile polymorphs have been considered (see Figure 1.10), because they are the only used for technological applications. The TiO_2 polymorphs are formed by hexacoordinated Ti atoms (distorted octahedron) and tricoordinated O atoms; the difference among the several polymorphs is in the linkage among octahedrals and in the distortion of the octahedrals from the perfect O_h symmetry, which leads to different space groups for each polymorph.

The surface science of titanium oxide has been abundantly reviewed by Diebold;[83] in this section we will only highlight the most interesting properties for our purposes.

A surface is obtained by cutting the bulk structure along a specific direction. Thus, the outermost atoms will be in a situation of chemically "instability" because some of their bonds have been



(a) Anatase (b) Rutile

Figure 1.11: Wulff polyhedra for the TiO_2 anatase and rutile polymorphs.

cut out to build the surface. As a result a surface presents electrostatic charges, which are responsible of their capability to adsorb organic molecules. In particular, the TiO_2 surfaces analysed in the present thesis have pentacoordinated Ti atoms which are Lewis acid centres, *i.e.* they can accept electrons, and electron-rich bivalent O atoms which are Brønsted basic centres, able to deprotonate acidic molecules, for example. However, theoretically there are several possible cuts for each material, which lead to different surfaces with different features. Accordingly, the properties described above about the reactivity of Lewis and Brønsted centres will be influenced by the surface morphology.

It is possible to cut several surfaces from the same bulk structure,

thus generating a molecule with a completely loss of periodicity, the so called Wulff polyhedron. George Wulff stated that the length of a vector drawn normal to a crystal face will be proportional to the surface energy. The surface energy is an important property which represents the energetic cost to cut the surface from the bulk; in other words the less the surface energy of a certain plane, the less the energetic cost to cut that surface from the bulk, and the major its stability. Hence, it is clear that the abundance in nature of a certain surface depends on its stability. Getting back to the Wulff polyhedron, the major contribution to its shape will be given by the most stable surfaces of the considered material. In Figure 1.11 the Wulff construction for the two most studied polymorphs of TiO_2 (anatase and rutile) are shown. As one can see, the shapes of the two polyhedra are indeed different.

Recently, the group of Prof. G. Martra of the University of Torino synthesized small oligopeptides from the vapour deposition of Gly on TiO_2 anatase nanoparticles. In particular, what they have noticed is that the condensation occurs on the most stable facet of the anatase polymorph, *i.e.* the (101) (red part in Figure 1.11a), which is also the most abundant one, and thus, the one expected to largely contribute to the peptide bond formation.[84] Facets with higher surface energy area seems not to have an important contribution in the peptide bond formation, despite their higher reactivity.[85] Thus, the present work was inspired by this experiment, and we will give insights by means of periodic DFT simulations of glycine polymerization on (101) anatase from an atomistic point of view.

1.5 Previous computational works on the interaction of amino acids and peptides with mineral surfaces

Characterizing with high level of accuracy of what happens, for example, on a catalyst surface when a certain adsorption and/or reaction occurs is one of the hardest challenges on this field. A powerful technique like X-ray diffraction, the most accurate instrument to determine with high resolution the spatial position of the atoms in a crystalline structure (the errors are in the magnitude order of the milliÅ), cannot give a detailed atomistic vision of the interaction of molecules with a surface since the transfer from a 3D to a 2D system leads to a drastic drop of resolution. Some techniques like solid state NMR (Nuclear Magnetic Resonance),[86] FTIR (Fourier-Transform Infrared Spectroscopy), AFM (Atomic Force Microscopy),[87] XPS (X-Ray Photoelectron Spectroscopy),[87] and STM (Scanning Tunneling Microscope)[88–90] can provide useful information about the interaction of molecules with surfaces, but they are still far from a detailed atomic resolution.[91, 92] Because of that, computational techniques are for these cases very useful to complement the experimental information at the atomistic level.

Several computational works have studied the adsorption of amino acids with mineral surfaces; in particular many of them have been concentrated on glycine, the most simple amino acid, in its interaction with alumina,[93] zinc oxide,[94] silica,[95] and hydroxyapatite.[96] Other studies considered several different amino acids, in order to account for the effect of lateral chains, in their interaction with metallic

surfaces (Au and Pd),[97–99] and with surfaces where adsorption occurs *via* weak forces, being them H-bond interactions (silica)[100] or dispersion forces (graphene).[101]

More complex systems than the simple adsorption of a single amino acid have been analysed by means of molecular dynamics simulations, by using classical force fields. Short peptides have been studied in their interaction with rutile,[102–104] muscovite and HOPG (highly oriented pyrolytic graphite), showing, in accordance with experimental results, that the adsorption on the first two surfaces occurs through direct contacts, whereas on graphite adsorption takes place *via* weak interaction with the π system.[105] Larger peptides interacting with SiO_2 and TiO_2 surfaces,[106] and with layered double hydroxides (LDH), such as brucite like clays with positive layered charges, have also been studied. Rimola et al. have provided important insights about mineral induced peptide folding by means of accurate full-QM calculations, and revealed that the α-helix conformation is preferred instead of the random coil conformation, when a small peptide is adsorbed on the hydroxyapatite.[107] The peptide folding induced by naturally-occurring surfaces and the subsequent activation of a potentially "hidden" bioactivity might have triggered the first biocatalytic reactions in a primordial Earth (in absence of life), hence, giving rise to the emergence of the metabolic cycles, a crucial aspect for the origin of life.[108, 109]

Costa and coworkers studied the interaction of Glycine with a stepped (101) Boehmite AlOOH surface mediated by water, by means of AIMD simulations; they have clearly seen that Gly prefers a direct contact (*i.e.* through the COO^- groups, forming an Al-OCO adduct) with the surface than the interplay through a water

layer.[110] Similarly, the interaction of a glycine molecule has been simulated in the interlayer of bulk montmorillonite both in dry and wet environments,[111] but also with hydroxyapatite, simulating a microsolvation by the interplay of some water molecules between the surface and the glycine molecule[112, 113]

Ugliengo and co-workers studied the peptide bond formation catalysed by molecular models of Lewis and Brønsted sites, represented by a single molecule of AlF_3 and HF, respectively.[114] Despite the *naif* nature of these models, one can observe the drastic decrease of the activation barrier respect to the non-catalysed reaction (55 kcal/mol) when one of the two catalyst is used (30 kcal/mol), and when the two catalysts cooperate (15 kcal/mol). Similarly, Aquino et al. studied the amide bond formation between acetic acid and methylammine catalysed by several molecules and cation complexes.[115] More complex simulations have been proposed, cutting a mineral surface in order to build up realistic cluster models of clay,[116, 117] by means of the composite ONIOM method. By means of the cluster approach the efficiency of Brønsted/Lewis sites of the feldspar surface between two glycines has been demonstrated, showing a sensitive reduction of the reaction barrier from 50 kcal/mol to 26 kcal/mol.[116, 117] Using a similar approach Phuakkong et al. simulate the peptide bond formation between two glycine molecules with cluster models of faujasite.[118] Also the FeS surface shows strong interaction energies with amino acids, and the charge transfer from the adsorbate to the surface may indicate the activation of the amino acid to the peptide bond formation, because of the easier nucleophilic attack of another amino acid to the carbonyl C.[119] Greenwell and co-workers provides very important insights on the peptide bond for-

mation, demonstrating the existence of a driving force for the condensation reaction among amino acids due to the formation of a water molecule which in turn hydrate the interlayer, with a sensible gain in energy.[120]

1.5.1 Simulations on TiO_2 surfaces

TiO_2 is one of the most studied surfaces in material science, because of its broad applications in many field of science.

In relation to the adsorption of amino acids and peptides on TiO_2 surfaces many studies are available in literature by means of experimental and computational techniques. Computational works focus on the several adsorption modes of the amino acids on these surfaces, showing a general preference for bidentate adsorptions, both (N,O) and (O,O). This is intuitive because TiO_2 surface exposes in their outermost atomic positions undercoordinated Ti atoms, which are strong Lewis sites that are stabilized by the interaction with the lone pairs of O and N atoms (dative bonds). TiO_2 surface also presents Brønsted basic sites, *i.e.* the most exposed O atoms, which play an important role on these surfaces, because they are important anchoring points as H-bond acceptors. An important point of debate concerns the spontaneous (or not) deprotonation of the amino acids on these sites. On the (110) surface, the most stable one of the rutile polymorph, and probably the most intensively studied TiO_2 surface, amino acids adsorb through a dissociative process and in a bidentate (O,O) manner[121–127] in accordance to IR,[127] photoelectron diffraction,[122] NEXAFS (near edge X-ray absorption fine structure) spectroscopy,[126] and scanning tunneling microscopy data.[123] An

other surface of interest is the (101) of the anatase polymorph. It is of particular interest because of the large synthesis of symmetric Wulff like nanoparticles, exposing the (101) facet as the predominant surface.[128] This is because of its very low surface energy, less than a half than that of the (110) rutile surface and the other anatase facets. Szieberth et al. showed that the preferred adsorption mode for glycine on the (101) anatase surface is the bidentate (N,O) one but in its canonical structure (*i.e.* the gas phase non-deprotonated structure). Overall, the deprotonation seems to depend on the polymorph; in particular the main features of the surface which drive the protonation/deprotonation process are its energy (surfaces with higher energy are more reactive and, accordingly more prone to deprotonate amino acids) and its morphology (the presence of an O atoms close to the adsorption site which can accept the proton).

Monti et al. studied large systems through long MD simulations by means of reactive classical force fields (ReaxFF) parametrized by DFT calculations on cluster models,[129] also considering the presence of water. However there is a lack of systematic studies on the evaluation of a complete set of amino acids in order to study the effect of the several lateral chains; moreover no studies have been carried out on the mechanism of peptide bond formation on TiO_2 surfaces.

1.6 Objectives

As already mentioned in previous sections, literature misses a systematic study on the adsorption of different and most representative groups of lateral chains among the several amino acids, and there are

no studies on the peptide bond formation mechanism on the surface of TiO_2 material.

In the present work, we will provide accurate full QM calculations by means of DFT approach using the VASP periodic code. In particular the main objectives are:

- To study the pure TiO_2 bulk systems in its anatase and rutile poylmorphs, and their most stable surfaces, *i.e.* the (101) for the anatase and the (110) for the rutile.
- To study the adsorption of representative amino acids in order to cover a complete range of the lateral chain features, both at the (101) anatase and (110) rutile surfaces.
- To study the peptide bond formation mechanism on the (101) surface, simulating the conditions of the experiments carried out by Martra et al.
- To study the relative stability of secondary structures (α-helices and β-sheets) on the (101) anatase surface by means of full periodic calculations, *i.e.* 1-D for α-helix and 2-D for β-sheets.

Chapter 2

Methodology

Theoretical and computational chemistry are disciplines devoted to study physico-chemical problems at an atomistic level. In particular, theoretical chemistry deals with the development of new methodologies which allow the treatment of the many-electron problem, while computational chemistry applies those instruments developed by theoretical chemists to study chemical processes of interest through the use of several computational codes. Two classes of code-types are available for the scientific community: one devoted to the visualization and the editing of molecular systems, and another one devoted to solve the equations provided by theoretical chemists throughout efficient algorithms, to calculate all the properties of interest of the system under study.

In this Chapter we will briefly introduce the birth of quantum mechanics and some milestones of quantum physics which allowed the evolution of the modern theories implemented in current codes. Then we will discuss how to model a realistic system, to finally go

in deeper details about the methods and codes used in the present thesis.

2.1 Dawn of quantum mechanics

At the end of the XIX century, many experimental observations demonstrated that classical physics cannot describe properly the properties related to atomic and sub-atomic particles. Among them it is worth mentioning the photoelectric effect, observed for the first time in 1887 by Heinrich Rudolf Hertz,[130] and the black body radiation, described separately by Balfour Stewart[131] and Gustav Kirchhoff[132] (the term "black body" was coined by Kirchhoff himself) in 1858 and 1859, respectively. The former is the emission of electrons from a surface when the electromagnetic radiation hits the material; the latter is the radiation within or surrounding a black body, *i.e.* an idealized physical body that absorbs all incident electromagnetic radiation. Both these phenomena could not be explained by the Maxwell's equations on the magnetic radiation; it actually generated real paradoxes. The solution came from Max Planck in 1900, who, for the first time, proposed that atoms may emit and absorb electromagnetic radiation only in discrete packets,[133, 134] called quanta. Because of this theory Max Planck can be considered the father of the quantum physics. Few years later, in 1905, Einstein, by applying the Planck's quantum theory, gave a reasonable explanation for the photoelectric effect, hypothesizing that the magnetic radiation should consist of energy packets,[135] then called photons. Starting from these experiment and theories, the first half

of the XX century classical physics experienced a real revolution. In 1913 Niels Bohr applied the quantum theory also to the mechanical energy, proposing an empirical model on the hydrogen atom where the electron can move only in a discrete ensemble of orbits;[136–140] this was confirmed in 1914 by Franck and Hertz. Another important evidence of a pure quantum phenomenon is the wave-particle duality: this was already clarified for the light nature, but in 1923 Louis Victor de Broglie hypothesized that also sub-atomic particles might have the same behaviour.[141] His theory was validated by the Compton scattering, observed by Arthur Compton in 1922[142] and, few years later (1927), by the Davisson-Germer experiment,[143] in which they sent slow electrons against crystalline Nickel; what they found was the same diffraction pattern obtained by William Bragg using X-ray diffraction.

All these (and many others) discoveries led to the formulation of a partial differential equation by Erwin Schrödinger,[144] who describes sub-atomic particles as waves, following the de Broglie's hypothesis. Due to its complexity this equation can not be solved analytically but for the H atom, H_2^+, and hydrogenoid systems, *i.e.* systems with only one electron. Still nowadays the many-electron problem has not been solved yet. Starting from this equation, in the following years many implementations have been applied to it in order to both simplify its complexity (the number of variables depends to the number of coordinates of electrons and nuclei) or to further improve its description of the matter. Among different approaches, it is worth mentioning the Born-Oppenheimer approximation (1927),[145] which decouples the electrons and nuclei motion, the orbital approximation, where each electron is considered to belong to

a hydrogen-like atom and the nuclear charge is corrected by using the effective nuclear charge, and the WKB (Wentzel-Kramers-Brillouin, 1926) approximation,[146–149] where the wave function is recast as exponential function, expanded in powers of the Plank constant. A fundamental improvement was given by Dirac (1928);[150, 151] as the Schrödinger equation is based on a classical Hamiltonian, it can not describe pure quantum behaviours, such as particles with zero mass (photons), also introducing the spin formalism (relativistic Schrödinger equation). In the same years the detachment of classical and quantum mechanics has been formally demonstrated by one of the most important theories of quantum mechanics, the uncertainty principle, formulated by Werner Heisenberg in 1927.[152]

Further improvements and approximations of the Schrödinger equation have been given with the aim of extend the method to more complex systems. One of the most important was the Hartree-Fock method (1930-1935),[153, 154] that became the father of computational methods. However, it was not until 1950s that the first *ab-initio* calculations were possible, thanks to the development by Roothaan[155] and Hall[156] who revisited the Hartree-Fock method, and to the development of the first digital computers. Over years, many methods were born to solve the many electron problem, only partially treated in the Hartree-Fock formalism; among many, one is particularly remarkable, the DFT (Density Functional Theory), developed by Kohn and Sham in 1965[157] starting from the consideration of the Thomas-Fermi electron gas model[158, 159] (1927) and the Hohenberg and Kohn theorems.[160] This approach faces the problem by a completely different point of view with respect to the HF method, substituting the complex wave function (many variables,

depending by the number of electrons and nuclei) with the electron density (a physical observable depending by only three variables). Still nowadays new DFT functionals arise every year and DFT is the most used approach for *ab-initio* calculations.

2.2 The Schrödinger equation

The Schrödinger equation is a partial differential, linear, complex, non-relativistic, eigenvalue equation, and, in its most compact form, it is written as:

$$\widehat{H}\Psi = E\Psi \tag{2.1}$$

Let us now analyse in deeper details its characteristics:

- Partial differential: *i.e.* an equation which contains multivariable functions and their partial derivatives,
- Linear: *i.e.* the maximum exponent of its unknowns is equal to 1,
- Complex: *i.e.* it contains both real and imaginary parts,
- Non-relativistic: *i.e.* a classical Hamiltonian-based equation, so it cannot treat particles with zero mass or with spin (Dirac equation),
- Eigenvalue: *i.e.* by applying the Hamiltonian operator to the unknown function Ψ, the result is the same function Ψ times a real value E (the energy of the systems). Therefore Ψ is an eigenfunction of H and E its corresponding eigenvalue.

Despite the "simple" form of Equation 2.1, the Hamiltonian H can be expressed as the sum of the kinetic energy of nucleus (T_n) and electron (T_e) and as the sum of the potential energy for the interactions between nucleus-nucleus (V_{nn}), nucleus-electron (V_{ne}) and electron-electron (V_{ee}). The Hamiltonian for a time-independent non-relativistic Schrödinger equation can be written as:

$$\begin{aligned}\widehat{H} = &- \frac{h^2}{8\pi^2}\sum_{\alpha}\frac{1}{M_\alpha}\nabla_\alpha^2 - \frac{h^2}{8\pi^2 m_e}\sum_{i}\nabla_i^2 + \frac{e^2}{4\pi\epsilon_0}\sum_{\alpha}\sum_{\beta>\alpha}\frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} \\ &- \frac{e^2}{4\pi\epsilon_0}\sum_{\alpha}\sum_{\beta>\alpha}\frac{Z_\alpha}{r_{\alpha i}} + \frac{e^2}{4\pi\epsilon_0}\sum_{i}\sum_{j>i}\frac{1}{r_{ij}}\end{aligned} \tag{2.2}$$

where h is the Planck constant, M_α is the mass of nucleus α, m_e is the mass of the electron, Z_α and Z_β are the atomic number for nuclei α and β, respectively, ϵ_0 is the vacuum dielectric constant, e is the electron charge, and ∇^2 is the Laplace operator, *i.e.* the sum of the second partial derivatives with respect to the coordinates, expressed as:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.3}$$

In atomic units Equation 2.2 can be written in a more compact form:

$$\begin{aligned}\widehat{H} = &- \frac{1}{2}\sum_{\alpha}\frac{1}{M_\alpha}\nabla_\alpha^2 - \frac{1}{2}\sum_{i}\nabla_i^2 + \sum_{\alpha}\sum_{\beta>\alpha}\frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} \\ &- \sum_{\alpha}\sum_{\beta>\alpha}\frac{Z_\alpha}{r_{\alpha i}} + \sum_{i}\sum_{j>i}\frac{1}{r_{ij}}\end{aligned} \tag{2.4}$$

$$\widehat{H} = \widehat{T}_n + \widehat{T}_e + \widehat{V}_{nn} + \widehat{V}_{ne} + \widehat{V}_{ee} \tag{2.5}$$

Now we can introduce the Born-Oppenheimer approximation (mentioned in the section above), *i.e.* the decoupling of nuclei and electrons motion:[145]

$$\Psi(\vec{R}, \vec{r}) = \Psi_n(\vec{R})\Psi_e(\vec{r}; \vec{R})) \tag{2.6}$$

where the subscripts n and e stand for the nuclear and the electronic parts, respectively. In $\Psi_e(\vec{r}; \vec{R})$ the electronic wave function explicitly depends by electronic coordinates r, but only parametrically by nuclear coordinates. This approximation finds its foundation by the fact that nuclei and electrons have different velocities due to their different masses (the proton mass is 1840 times heavier than the electron one); hence the electrons relax instantaneously for each nuclei configuration. Thank to this approximation the first term of Equation 2.5 (T_n) can be neglected and the third term (V_{nn}) can be considered as a constant. So one can solve the electronic Schrödinger equation, thus obtaining the potential energy of the system:

$$\widehat{H_e}\Psi_i(\vec{r}; \vec{R}) = (\widehat{T}_e + \widehat{V}_{ne} + \widehat{V}_{ee} + \widehat{V}_{nn})\Psi_i(\vec{r}; \vec{R}) = U_i(\vec{R})\Psi_i(\vec{r}; \vec{R}) \tag{2.7}$$

The main problem of the Schrödinger equation is that the wave function is unknown. To solve this problem the approach known as variational principle was developed, which systematically drives the solution to the best approximate wave function for a determined electronic state, starting with a trial function ϕ. For a given system with

its proper Hamiltonian, the variational principle states that:

$$\epsilon[\phi] = \frac{\langle\phi|\widehat{H}|\phi\rangle}{\langle\phi|\phi\rangle} \tag{2.8}$$

Within this approximation $\epsilon[\phi] > E_0$, where E_0 is the solution for the real wave function Ψ, $\epsilon[\phi] = E_0$ if, and only if, the trial function correspond to the real wave function of the system. Despite now we have a reasonably well approximated wave function, one of the ingredients of the Hamiltonian is still nowadays elusive, the electron correlation V_{ee}. Many theories and methods were born in order to find a good approximation of this crucial contribution to the energy, since the Hartree-Fock approximation, until the most recent developments on the Density Functional Theory.

2.3 Wave function-based methods

The consequence of the problem discussed above, *i.e.* the electron-electron interaction (V_{ee}), is that the Schrödinger equation can be solved only for those systems with non-interacting electrons. Obviously this is a huge limitation for interesting applications. Therefore, many strategies have been developed in order to solve this problem and to apply this powerful equation to systems of physico-chemical interest. The first solution was proposed by Hartree and Fock in 1930, with the consequent birth of what we can consider the father of the *ab-initio* methods.

2.3.1 The Hartree-Fock method

The foundations of the Hartree-Fock method were raised by Hartree in 1928 with the so called Hartree product,[161–163] where the wave function of a system with N electron is approximated with the product of N-orthonormal one-electron wave functions:

$$\Psi(x_1, x_2, ..., x_N) = \chi_1(x_1)\chi_2(x_2)...\chi_N(x_N) \tag{2.9}$$

However, as pointed out by Slater in the same year, this approximation does not describe adequately the multi-electron wave function because it is not antisymmetric with respect to the exchange of the coordinates of any pair of electrons. Thus, he proposed a similar product but organized in the Slater determinant:[154]

$$\Psi(x_1, x_2, ..., x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \dots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \dots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \dots & \chi_N(x_N) \end{vmatrix} \tag{2.10}$$

where the spin orbitals $\chi_i(x_i)$ are defined as:

$$\chi_i(x_i) = \phi_i(\vec{r})\phi_i(s) \tag{2.11}$$

where $\phi_i(\vec{r})$ is a spatial function and $\phi_i(s)$ is a spin function, which can be either α or β. In this case, the expectation value of the electronic Hamiltonian (H_{el}) is the Hartree-Fock energy:[153]

$$E_{HF} = \sum_{i}^{N} \langle \chi_i | \widehat{h}_i | \chi_i \rangle + \frac{1}{2} \sum_{ij}^{N} (\langle \chi_j | \widehat{J}_i | \chi_j \rangle - \langle \chi_j | \widehat{K}_i | \chi_j \rangle) \qquad (2.12)$$

where $\widehat{h}_i$ is the mono-electron operator that gives the contribution of the kinetic energy and the potential energy for the electron-nucleus interaction ($T_e + V_{ne}$), while $\widehat{J}_i$ and $\widehat{K}_i$ represent the Hartree-Fock approximation of the potential energy for the electron-electron interaction (V_{ee}), *i.e.* they are bi-electronic operators. In particular $\widehat{J}_i$ is the Coulomb operator and describes the classical repulsion between charges of the same sign, and $\widehat{K}_i$ is the Exchange operator, which arises by the antisymmetric character of the Slater determinant, so it has a purely quantum meaning: it describes the Fermi hole, or, in other words, the Pauli exclusion principle. As the Hartree-Fock belongs to those methods which exploit the variational principle, E_{HF} is minimised with respect to the spin orbitals, with the constraint that the spin orbitals remain orthonormal. The minimization of the energy leads to a new set of spin orbitals which, in turn, can be used to further minimise the energy, repeating the procedure until a pre-set convergence criterion. Because of this, the Hartree-Fock is a SCF (self-consistent field)-type method. After a proper transformation of the spin orbitals (χ_j) to spatial orbital (ψ_j), which finally are the molecular orbitals of interest, we face with another problem, *i.e.* the mathematical expression of molecular orbitals is unknown. A further improvement of this methodology was given by Roothaan and Hall, who introduced the LCAO (linear combination of molecular orbitals) approximation,[155, 156] which in turn represents the

foundation of the basis set concept. Within this approximation the molecular orbitals can be expressed as:

$$\psi_i = \sum_{\mu=1}^{Z} C_{\mu i}\phi_\mu \qquad Z = 1, 2, \ldots, N \tag{2.13}$$

where ϕ_μ is a set of basis functions and $C_{\mu i}$ the expansion coefficients. For a complete set of ϕ_μ the expansion would be exact; however for computational reasons we have to use a finite set of K basis functions. The Hartree-Fock approximation has an important limitation: the bi-electronic operator $\widehat{K}_i$, as it describes the formalism of the Pauli principle exclusion, only takes into account electron correlation between electrons with the same spin $\alpha\alpha$ or $\beta\beta$. Consequently, it describes electrons moving in an averaged field generated by the others electrons. Therefore, HF method has a lack of electron correlation.

2.3.2 Post Hartree-Fock methods

The electron correlation energy may be defined as:

$$E_{CORR} = E_{EXACT} - E_{HF} \tag{2.14}$$

One can distinguish between two types of electron correlation:

- Dynamic correlation: the electrons, in the Hartree-Fock approximation, do not interact with each other, but each electron interact with an average field created by the other electrons
- Static correlation: the Hartree-Fock approximation is represented by a single Slater determinant, thereby not accounting for degenerate electronic configurations

To recover this small (but very important!) amount of energy the post Hartree-Fock methods, based on expanding the wave function, can be used, the most important of which are: the configuration interaction (CI) method, the many-body perturbation theory (MBPT), and the coupled cluster (CC) method.

Methods based on density functional theory can also account for electron correlation

2.4 Density functional methods

Density Functional Theory (DFT) is the most popular method to include the effects of the electron correlation with a relatively cheap computational cost. The power of DFT is that, first of all, it is based on the electron density ($\rho(\vec{r})$) which, in contrast to the wave function, is a physical observable experimentally measured (X-Ray diffraction), while the wave function has no direct physical meaning and its form is unknown. The electron density depends on three variables (x,y,z), while the wave function variables depends on the number of atoms (nucleus and electron coordinates). The other advantage of using a physical variable is that $\rho(\vec{r})$ is a non-negative real function (it has no imaginary parts as the wave function) which vanishes at infinity ($\lim_{\vec{r}\to\infty} \rho(\vec{r}) = 0$) and, integrated over the space gives the number of electrons ($\int \rho(\vec{r}) d\vec{r} = N$).

2.4.1 Foundations of DFT: from Thomas-Fermi model to Hohenberg-Kohn theorems

Shortly after the formulation of the Schrödinger equation, Llewellyn Thomas and Enrico Fermi proposed in the same year (1927) a quantum mechanical theory for the electronic structure based on the electron density.[158, 159] Despite the *naif* nature of the uniform electron gas described by the Thomas-Fermi model, it opened up the foundation of what, thirty years later, was named as DFT.

The most important improvement which established the birth of DFT was the Hohenberg and Kohn theorems;[160] the first states that "[The external potential] $V_{ext}(\vec{r})$ is (to within a constant) a unique functional of $\rho(\vec{r})$; since, in turn, $V_{ext}(\vec{r})$ fixes $\widehat{H}$ we see that the full many-particle ground state is a unique functional of $\rho(\vec{r})$". Through a *reduction ad absurdum* they demonstrated a one to one relationship among the electron density and the Hamiltonian, the wave function and the energy, and, accordingly, all the related properties. The second Hohenber-Kohn theorem is the variational principle which, as for wave function-based variational methods, provides the exact energy (E_0) of the ground state through the minimization of $E[\rho(\vec{r})]$:

$$E_0 = E[\rho_0(\vec{r})] \leq E[\rho(\vec{r})] \tag{2.15}$$

where $\rho_0(\vec{r})$ is the real ground-state density.

2.4.2 The Kohn-Sham method

Following the Hartree-Fock approximation, in 1965 Kohn and Sham developed a method to study a system of interacting electrons start-

ing from a fictitious system of non-interacting electrons that generate the same density as that of interacting electrons. In an analogous way as the Schrödinger wave function within the Born-Oppenheimer approximation (see Equation 2.7), we can write the Hamiltonian of the system as:

$$\widehat{H} = \widehat{T} + \widehat{V} + \widehat{J} + \widehat{V}_{XC} \tag{2.16}$$

where $\widehat{T}$ (also called $T_s[\rho]$) is the kinetic energy of a system of non-interacting electrons:

$$T_s[\rho] = \frac{1}{2}\int \nabla^2 \rho(\vec{r}) d\vec{r} \tag{2.17}$$

$\widehat{V}$ (also known as $V_{ext}[\rho]$) is the potential energy from the external field due to positively charged nuclei:

$$V_{ext}[\rho] = \int v(\vec{r})\rho(\vec{r}) d\vec{r} \tag{2.18}$$

J ($J[\rho]$) is the classical Coulomb energy of electron-electron repulsion:

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|r_1 - r_2|} d\vec{r}_1 d\vec{r}_2 \tag{2.19}$$

and the V_{XC} operator which take into account the exchange and correlation behaviour of interacting electrons, as well as the difference between the kinetic energy of the interacting and non-interacting electrons ($T[\rho] - T_s[\rho]$), whose exact form is unknown. As one can see, all these operator are functionals of the electron density. Putting this operator in the one-particle Schrödinger equation and recovering the Hartree-Fock formalism we can write:

$$\widehat{h}_{KS}\chi_i(\vec{r}) = \epsilon_i(\vec{r})\chi_i(\vec{r}) \tag{2.20}$$

where $\widehat{h}_{KS}$ is the Kohn-Sham operator before called generically $\widehat{H}$. Solving the Schrödinger equation iteratively (also DFT is self - consistent!) by starting with a trial $\rho(\vec{r})$, one will get the converged $\rho_0(\vec{r})$, always remembering that $\rho(\vec{r})$ must satisfy the following requirement:

$$\rho(\vec{r}) = \sum_{i=1}^{N} |\chi_i(\vec{r})|^2 \tag{2.21}$$

V_{XC} is the critical part of all the formalism; it is the functional derivative of the exchange-correlation energy:

$$V_{XC} = \frac{dE_{XC}[\rho(\vec{r})]}{d\rho(\vec{r})} \tag{2.22}$$

$E_{XC}[\rho]$ can be divided into two parts: the exchange term ($E_X[\rho]$) and the correlation term ($E_C[\rho]$):

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \tag{2.23}$$

which represent the interaction between electrons with the same and opposite spins, respectively. Density functional methods differ on the treatment of these terms, which will be discussed in the next section.

2.4.3 Exchange-Correlation functionals

The simplest approach is the Local Density Approximation (LDA), which depends only on $\rho(\vec{r})$. The first LD approximation was proposed by Dirac in 1930,[164] within the approximation of the Thomas-

Fermi model:

$$E_X^{LDA} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho^{\frac{4}{3}}(\vec{r}) d(\vec{r}) \tag{2.24}$$

The analytical part of the correlation energy is not known, but Monte Carlo simulations allowed the construction of the very popular functional by Vosko, Wilk and Nusair, referred as VWN.[165] It is very well known that LDA functionals tend to delocalize too much the electrons, and, because of that, they underestimate bond distances and, accordingly, overestimate binding energies, especially for those systems with localized electron. However, not surprisingly, they provide good results in the description of solid metals, whose electrons are delocalized over all the material, and, accordingly, well described by the uniform electron gas model. A more generic approximation is the Local Spin Density Approximation (LSDA) with the electron density is split in its two spin components (α and β), in order to describe open-shell systems.

A more accurate description of the electron density is given by Generalized Gradient Approximation methods (GGAs), which introduce the first derivative of the density (the gradient). One of the first and most common exchange functional has been proposed by Becke (B88) as a correction of the LDA formula:[166]

$$E_X^{B88} = E_X^{LDA} - \beta \sum_\sigma \int \rho_\sigma^{\frac{4}{3}} \frac{x_\sigma^2}{1 + 6\beta x_\sigma sinh^{-1} x_\sigma} d\vec{r} \qquad x = \frac{|\nabla \rho|}{\rho^{\frac{4}{3}}} \tag{2.25}$$

where $\beta = 0.0042$, which is a parameter that was determined by fitting Hartree-Fock exchange energies of several atoms. A popular

correlation functional is the LYP (Lee-Yang-Parr) which has been built with the correlation formula developed by Colli and Salvetti. Within the GGA methods it is worth remember the PBE (Perdew-Burke-Ernzerhof) exchange-correlation functional,[167] the one which has been mostly used in this work.

More accurate description of the $E_{XC}[\rho]$ is given by meta-GGA functionals, where the Laplacian (*i.e.* the curvature) of the electron density has been introduced. The TPSS (Tao-Perdew-Staroverov-Scuseria) is an example of this class of functionals; this is a non-empirical functional and it may be considered as an improvement over the PBE.[168] Other functionals that deserve to be mentioned are the Minnesota one (M06L) developed by Trhular and coworkers.

Hybrid density functionals (also known as hyper-GGA methods) combine a percentage of the exact Hartree-Fock exchange (depending on the functional) with the exchange-correlation of a GGA method. This class includes the most popular density functional methods; one of them is the B3LYP, which combine the B88 exchange functional and the LYP correlation functional with a 20% of exact Hartree-Fock exchange and 3 empirical parameters:[169, 170]

$$\begin{aligned} E_{XC}^{B3LYP} = (1-a_0)E_X^{LSDA} + a_0 E_X^{HF} + a_x \Delta E_X^{B88} \\ + a_c E_C^{LYP} + (1-a_c)E_C^{VMW} \end{aligned} \tag{2.26}$$

where a_0, a_x, and a_c corresponding to 0.2, 0.7 and 0.8 values, respectively. The other very popular hybrid functional is the PBE0, *i.e.* the PBE as it is with a 25% of exact exchange.[171]

The hyper-meta GGA methods, as the name suggests, combine

meta GGA methods with a percentage of exact exchange. Among them the already cited TPSS has been augmented with a 10% of Hartree-Fock exchange. The Minnesota are a very huge class of functionals with different percentages of exact exchange, since M05,[172, 173] M06[174–176] and M08[177] classes, which consist of both meta and hyper-meta GGA functional, until the newest M11,[178, 179] M12[180, 181] and M15,[182, 183] which introduce the range separation, which is a partitioning of electron-electron interactions between short range (local DFT exchange) and long range (Hartree-Fock exchange) in order to represent correctly the interaction potential decay.

2.4.4 Some difficult cases for DFT

DFT methods are the most popular choice to perform *ab-initio* calculations, because many of the different functionals are implemented in several codes, which allow the scientific community to address any kind of chemical problem. In general they provide good geometries and, accordingly also good energies and all their related properties, such as dipole moments, frequency calculations, and excited states. However, there are some cases where pure DFT methods cannot reach the desired accuracy. This is the reason of the development, along the years, of so many functionals, to include corrections over the pure DFT formalism. The most frequent cases where DFT fails are: weak interactions, charge transfer processes, and open-shell systems.

Weak interactions

Weak interactions are basically the so-called van der Waals forces. Let consider graphite, for example: this system is formed by graphene layers, *i.e.* sheets of sp2 carbons of one-C-atom thickness, which interact to each other with the electron cloud of the π-system, as depicted in Figure 2.1. Thus, there are not chemical bonds between graphene layers, nor dative interactions, but only dispersive forces. The experimental distance between the graphene layer is 3.35 Å, and if pure DFT describes quite bad this distance (left part of Figure 2.1) the DFT including dispersion (D) corrections (right part of figure 2.1) is much more accurate in the description of these weak forces, providing an interlayer distance close to the experimental value. DFT-D methods have been implemented by Stefan Grimme, and they are fundamentally *a-posteriori* corrections of energies and gradients based on experimental data and atom-pairwise correction based on atomic ionization potentials. The general formulation of this correction is:

$$E_{DFT-D} = E_{DFT} - S_6 \sum_{i \neq j} \frac{C_6^{ij}}{R_{ij}^6} f_{damp}(R_{ij}) \tag{2.27}$$

where R_{ij} is the distance between atoms i and j, C_6^{ij} is a dispersion coefficient based on atomic polarizabilities and ionization potentials, f_{damp} is a dumping function in order to avoid non-physical behaviours for small distances, and S_6 is an empirical parameter different for each functional. DFT-D2 correction considers all pair of atoms,[184, 185] DFT-D3 also triplets of atoms in order to take into account three-body effects,[186, 186] while the more recent DFT-D4 include atomic partial charges computed at semi-empirical quantum level to scale

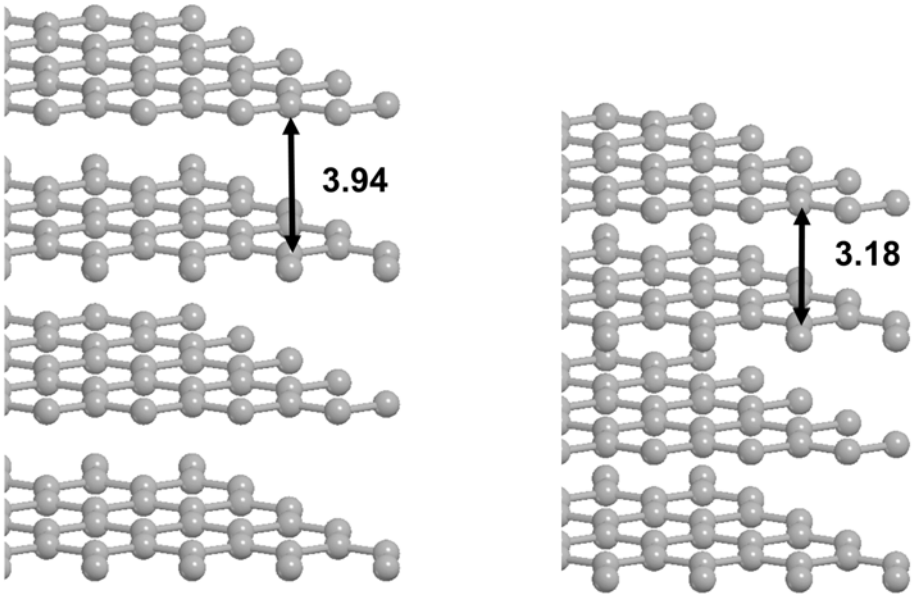


Figure 2.1: PBE (left) and PBE-D2 (right) optimised geometries for the graphite bulk structure.

polarizabilities.[187]

Charge transfer interactions

This kind of interactions occur in metal transition complexes when a fraction of the electronic charge is transferred (metal-to-ligand or ligand-to-metal charge transfer). Pure DFT functionals tend to delocalize the electrons and, accordingly, overestimate the energy of these interactions. However, by adding a proper percentage of exact Hartree-Fock exchange (hybrid functionals) the problem usually vanishes.

Open-shell systems

This is the case of many transition metals (particularly first-row), or radical systems. In this case the problem of pure DFT results from an over-delocalization of the spin density due to the bad cancellation of the self-interaction part of the exchange functional; also in this case DFT predicts an over-stabilization of these systems, and the solution, as for charge transfer interactions, is the inclusion of the exact exchange. The majority of hybrid functionals does not have high percentages of exact exchange because the Hartree-Fock tends to be inaccurate in the opposite direction with respect to DFT, *i.e.* lack of electron correlation. Because of this, the hybrid functionals have been balanced with a proper amount of Hartree-Fock exchange.

2.5 Molecular mechanics based methods

An approach completely different from quantum methods is to use equations deriving by classical physics to get the energy and all related properties of the system of interest. The main difference with respect to quantum methods is that molecular mechanics (MM) based methods do not account for electrons; the basic idea is that atoms are treated as classical particles, and they are bound to each other by springs of different forces (depending on atoms type and hybridation).[188, 189] These methods are also called Force Field methods.

2.5.1 The Force Field energy

The Force Field energy is written as a sum of several terms:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \tag{2.28}$$

where E_{str} is the energy required for stretching a bond between two atoms, E_{bend} for bending an angle, E_{tors} for rotating a bond (torsional angle), E_{vdw} and E_{el} account for interactions between non-bonded atoms, and E_{cross} describes couplings between the first three terms.

E_{str} is described in its simplest form as a second order truncated Taylor expansion around the equilibrium bond length R_0:

$$\begin{aligned} E_{str}(R^{AB} - R_0^{AB}) &= E(0) + \frac{dE}{DR}(R^{AB} - R_0^{AB}) \\ &\quad + \frac{1}{2}\frac{d^2R}{dR^2}(R^{AB} - R_0^{AB})^2 \\ &= k^{AB}(\Delta R^{AB})^2 \end{aligned} \tag{2.29}$$

where k^{AB} is the force constant of the A–B bond, and the equation represents the parabolic behaviour of the potential with respect to an infinitesimal displacement of the atoms from the equilibrium geometry. This is the simplest expression of E_{str} because it requires the only parameter k^{AB}; the E_{str} can be improved by adding more terms in the Taylor expansion, thus requiring more parameters. However, for the cubic (k_3^{AB}) and quartic (k_4^{AB}) terms the energy tends to -∞ and +∞, respectively, so they lead to non-physical results for large deviations with respect to the equilibrium geometry. A simple

function which describes properly the energy for large stretchings is the Morse potential:[190]

$$E_{morse}(\Delta R) = D(1 - \exp^{-\alpha \Delta R})^2 \qquad \alpha = \sqrt{\frac{k}{2D}} \tag{2.30}$$

where D is the dissociation energy.

E_{bend} has an expression analogous to E_{str}:

$$E_{bend}(\theta^{ABC} - \theta_0^{ABC}) = k^{ABC}(\theta^{ABC} - \theta_0^{ABC})^2 \tag{2.31}$$

in its harmonic approximation, which can also be expanded to third, fourth, or higher orders.

E_{tors} has a different form because of both the periodic conditions this parameter must present, and also because of the lower energy required to distort a torsion angle, in comparison to stretching and bending, *i.e.* large deviations from the minimum structures may occur. Accordingly, a truncated Taylor expansion in this case leads to non-physical constraints. Usually E_{tors} is written as a Fourier series:

$$E_{tors}(\omega) = \sum_{n=1} V_n cos(n\omega) \tag{2.32}$$

where n accounts for the periodicity of the torsion angle, and V_n is the barrier for the rotation around the dihedral angle.

E_{vdw} term describes the attraction and repulsions between atoms not directly bound. This energy should decrease to zero when the distance between the atoms increases, while increases very fast for short

distances because of the repulsion between electrons (Pauli exclusion principle). At intermediate distances it describes the weak attraction due to instantaneous dipole-dipole interactions between the electron clouds of the considered atoms (electron correlation).[191] The most popular function which describes correctly the above mentioned behaviours is the Lennard-Jones potential:[192]

$$E_{vdw}(R) = \epsilon \left[\left(\frac{R_0}{R} \right)^{12} - 2 \left(\frac{R_0}{R} \right)^{6} \right] \tag{2.33}$$

E_{el} describes the interaction with permanent dipoles and charges, and the expression of the energy is the Coulomb potential, in an analogous form with respect to that described in the Schödinger equation:

$$E_{el}(R^{AB}) = \frac{Q^A Q^B}{\epsilon R^{AB}} \tag{2.34}$$

E_{tors}, as said before, takes into account the coupling between the first three terms (stretching, bending, torsion), which are usually written as first-order Taylor expansion. The most important coupling is due to stretching and bending:

$$E_{str/bend} = k^{ABC}(\theta^{ABC} - \theta_0^{ABC})[(R^{AB} - R_0^{AB}) - (R^{BC} - R_0^{BC})] \tag{2.35}$$

but, depending on the force field used, also the other possible permutations among stretching, bending and torsions may be present.

2.5.2 Advantages and limitations of Force Field methods

Obviously, the main advantage of MM based methods is the high speed of the calculation also for very large systems. While with QM methods many cores and many days/weeks are necessary to optimise systems with few hundreds of atoms, with MM methods it is possible to treat systems with thousands of atoms with a personal computer in few minutes/hours. One of the limitations of using MM based methods is the drop of accuracy with respect to QM methods; moreover for each new systems studied it has to be previously parametrized. To parametrize a force field one has to use accurate QM calculations or experimental data. Moreover it is impossible to study the reaction profile of a certain reaction by searching for transition state structures; due to the harmonic approximation of the chemical bond, MM methods cannot describe its cleavage/formation. However, this problem can be overcome thanks to the development of a new Force Field named ReaxFF.[193]

2.6 Modelling solids and surfaces

A perfect crystal can be considered as a periodic structure defined by a unit cell repeated in the three directions of space. There are two strategies to model a solid: the cluster and the periodic approaches.

2.6.1 Cluster approach

As we can consider the crystal as a large molecule, one of the simplest approaches to model a solid is to cut from the crystal a small portion of atoms, with the approximation that this portion maintains the properties of the whole crystal. In principle, for different cuts of several sizes, the properties of the macromolecule should converge to the crystal ones as the cluster size increases. An advantage of this approach is that any molecular code may treat a finite system, with the unique limitations being those arising from the electronic structure methods. However, cluster models may suffer from edge effects, leading to too flexible systems, if the size is not large enough.While increasing the size of the cluster is the obvious strategy, this may lead to systems unaffordable from a computational points of view. A solution may be basically blocking those atoms at the edge of the cluster during geometry optimizations, in order to simulate the bulk rigidity. Another approach, which is able to treat very large systems, would be the ONIOM method,[194–196] which consists in dividing the system in two parts, treated at different levels of theory (for example DFT/MM).

2.6.2 Periodic approach

The previous approach can not take into account neither the long range order nor the periodicity of the crystal, and in many cases it fails in the description of the macroscopic properties of the crystal itself. By contrary, the periodic approach can lead to the desired properties of the studied system with a relatively small amount of atoms, thanks to their repetition in the three directions. Hence the

key point of this approach is the translational symmetry of the unit cell in three directions (3D periodic systems, *i.e.* bulk materials), in two directions (2D periodic systems, *i.e.* surfaces), or in one direction (1D periodic systems, *i.e.* polymers). While in molecular systems symmetry is useful to save computational time and also for the interpretation of some results (such as the assignation of normal modes in IR spectra, electron transition in UV-Vis spectra), but not mandatory, in periodic systems the translational symmetry plays a key role, without which it would not be possible the formal treatment and the solution of the Schrödinger equation. In fact, extending the LCAO scheme for a solid, one would have to consider an infinite number of basis functions, with consequently infinite matrices. However, taking into account the translational symmetry and Bloch functions (see the following section) the problem can be factorized into n problems of m dimension, where n is the number of irreducible representation (IR) of the translational group and m the number of basis function.[197]

The Bloch theorem

In order to understand the solution of the periodic Schrödinger equation it is useful to introduce some basic concepts. We have talked about translational symmetry; so the direct lattice is a collection of points repeated in the space at intervals of $\vec{a_1}$, $\vec{a_2}$, $\vec{a_3}$ (the basis vectors). Hence the lattice vector $\vec{g}$ can be written as:[198]

$$\vec{g} = n_1\vec{a_1} + n_2\vec{a_2} + n_3\vec{a_3} \tag{2.36}$$

Where n_1, n_2 and n_3 are integers and, accordingly, the $\vec{g}$ lattice vector defines the n_1, n_2, n_3 cell with respect to the cell origin (0-

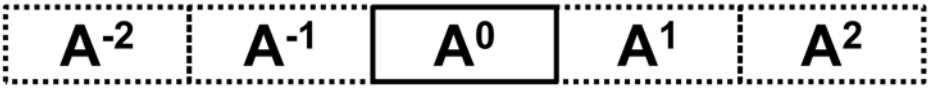


Figure 2.2: Schematic representation of a 1D-periodic system, with one atom per unit cell.

cell). For the sake of simplicity let's consider the polymer depicted in Figure 2.2 (periodic in only one dimension, hence $a_2 = a_3 = 0$); in this case the 0-cell which generates by translational symmetry the whole system is the one with the atom A^0 (continuous line). The atom A^1 will be defined by the lattice vector $\vec{g} = 1\vec{a_1}$.

Each direct lattice admits the construction of its reciprocal lattice of basis vectors (b_1, b_2, b_3) which obey to the following important orthogonality relationship:

$$\vec{a_i}\vec{b_j} = 2\pi\delta_{ij} \tag{2.37}$$

Hence, because of the translation symmetry of the crystal, the Schrödinger equation becomes:

$$\widehat{H}(r - \vec{g})\Psi(r - \vec{g}) = E\Psi(r - \vec{g}) \tag{2.38}$$

in which the $\vec{g}$ vector is introduced. With the assumption of a periodic nature of the potential $V(r)$, the eigenfunctions of the Schrödinger equation will obey the Bloch theorem:[199]

$$\Phi^{\mathbf{k}}(r + \vec{g}) = e^{i\mathbf{k}\vec{g}}\Phi^{\mathbf{k}}(r) \tag{2.39}$$

where $\mathbf{k}$ is a point in the reciprocal space, with an analogous definition as g in the direct space, and $e^{i\mathbf{k}\vec{g}}$ is a phase factor which has the same

periodicity of the crystal. In the schematic representation of Figure 2.2 this means that the functions centred on the atom A^0 will be the same as on A^1, and the other A atoms replicated by translational symmetry. By imposing the Born-Von Karman periodic boundary conditions on the Bloch functions on our model system of Figure 2.2, Equation 2.39 becomes:

$$\Phi^{\mathbf{k}}(r + Na) = e^{i\mathbf{k}Na}\Phi^{\mathbf{k}}(r) = \Phi^{\mathbf{k}}(r) \qquad j = 1, 2, 3 \tag{2.40}$$

where N is the number of cells in the crystal. Equation 2.40 implies that:

$$e^{i\mathbf{k}Na} = 1 \tag{2.41}$$

Introducing the property of Equation 2.37 we have:

$$e^{2\pi i\mathbf{k}N\kappa} = 1 \tag{2.42}$$

which is satisfied for:

$$\kappa = \frac{n}{N} \tag{2.43}$$

where n is integer. Hence for a number of cells which tends to infinity, also the number of **k**-points tend to infinity, thus covering completely the space of the reciprocal lattice, so that k can be considered a continuous variable, represented finally by the bands of the considered material. Each **k**-point will be the IR of the translation group of the considered system, and because of a general theorem of quantum mechanics which states that the product of functions belonging to

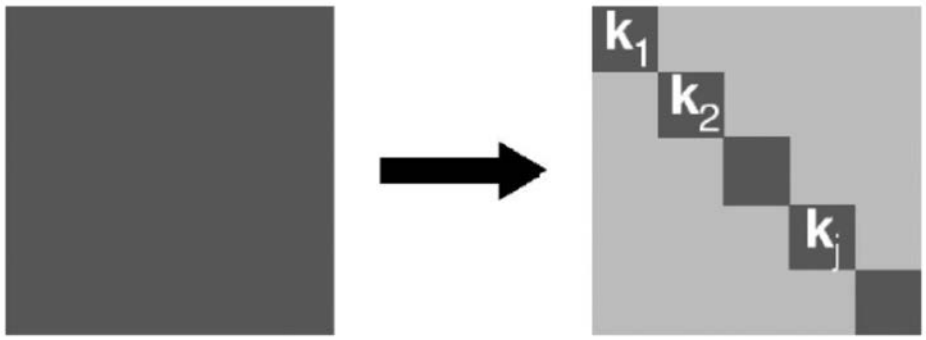


Figure 2.3: Transformation of the infinite Hamiltonian matrix when expressed in the basis of Bloch functions. The figure has been taken from ref. [198]

different IRs is zero, we have a transformation of the infinite Hamiltonian matrix in a block-diagonal matrix, where each block has the symmetry of the j^{th} **k**-point (see Figure 2.3).

Despite the elegance and the real improvement of the Bloch theorem in solving the periodic Schrödinger equation, we have basically reduced a problem of infinite size into infinite problems of finite size. But this is actually the solution of our problem, because the eigenvalues smoothly change with the **k**-points, and, accordingly, a finite number of **k**-points may be used for the solution of the periodic Schrödinger equation; in practical words one will see the convergence of the energy (and so the related properties) for a limited number of **k**-points, depending on the system of study. In Figure 2.4 two opposite systems are shown: an insulator (urea) and a conductor (Be). As one can see, the bands of insulating systems change smoothly with the **k**-points, *i.e.* the energy does not change so much, while in conducting systems the energy of each **k**-point is different from one

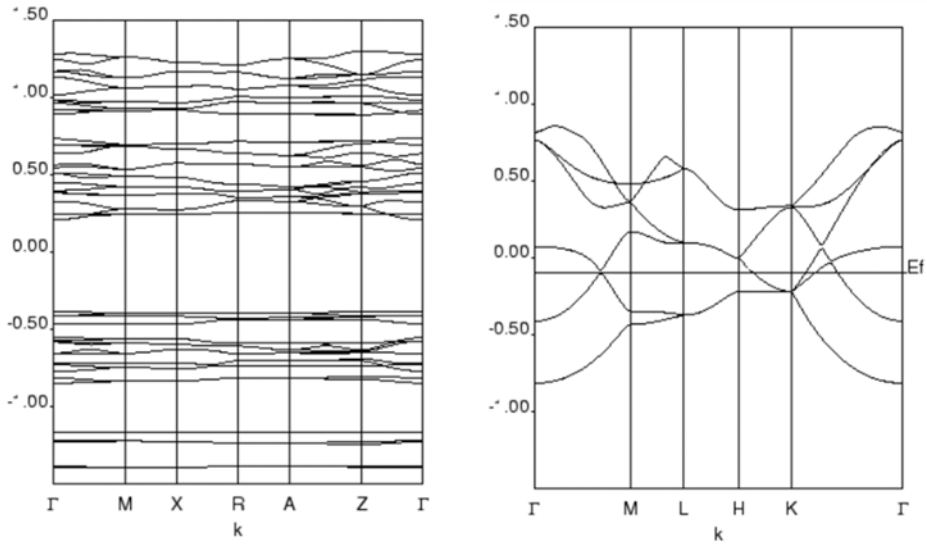


Figure 2.4: Band structure for the Urea bulk (left) and Be bulk (right). The figure has been taken from the CRYSTAL tutorials web site: `http://tutorials.crystalsolutions.eu/`

another, until a crossing with non-occupied band; in this case a much larger **k**-point sampling is required, to lead to converged energy.

The most used methodology for the **k**-points sampling has been developed by Monkhorst and Pack;[200] in particular the **k**-points grid over which the integrals are evaluated is built up not in the whole reciprocal cell, but in a small portion of it, called the Brillouin Zone (see Figure 2.5). **k**-points with special positions, depending on the symmetry space group of a particular crystal, are defined by capital letters, as shown in Figures 2.4 and 2.5; the Γ-point has the coordinates (0,0,0).

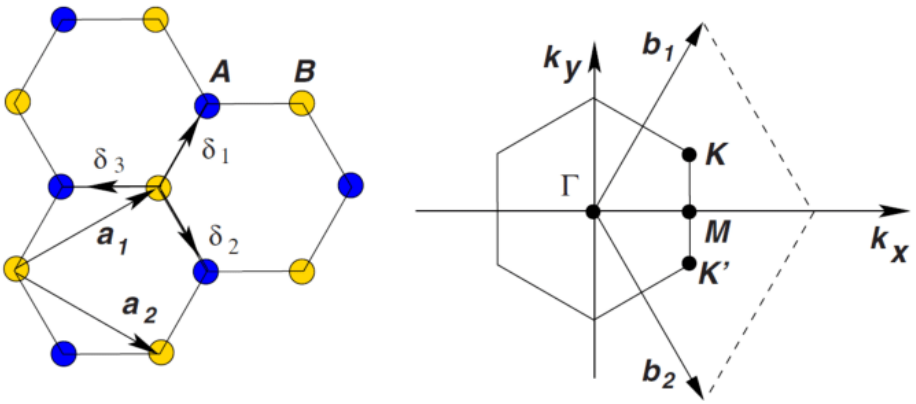


Figure 2.5: Direct (left) and reciprocal (right) space for the graphene. The first Brillouin Zone (BZ) is the small triangle defined by ΓMK.

Surface models

A surface is a periodic system where the bulk material looses one of its three periodic directions, *i.e.* a plane parallel to a given Miller index (hkl) is exposed to the vacuum. From an experimental point of view the outermost atoms of the surface are connected to the bulk atoms, *i.e.* the periodicity of the bulk for inner atoms is conserved (see Figure 2.6).

However, from a computational viewpoint it is not possible to reproduce the structure depicted in left part of Figure 2.6; we are forced to do a double cut of the bulk structure, building the so called thin film (or slab) model, periodic along, for example, x and y axes, and with a finite thickness along z. The thickness of the slab cannot be random, but one has to choose it carefully, with the constraint that the inner atoms of the slab represent with a good approximation the bulk structure; in other words the inner atoms must not be influenced

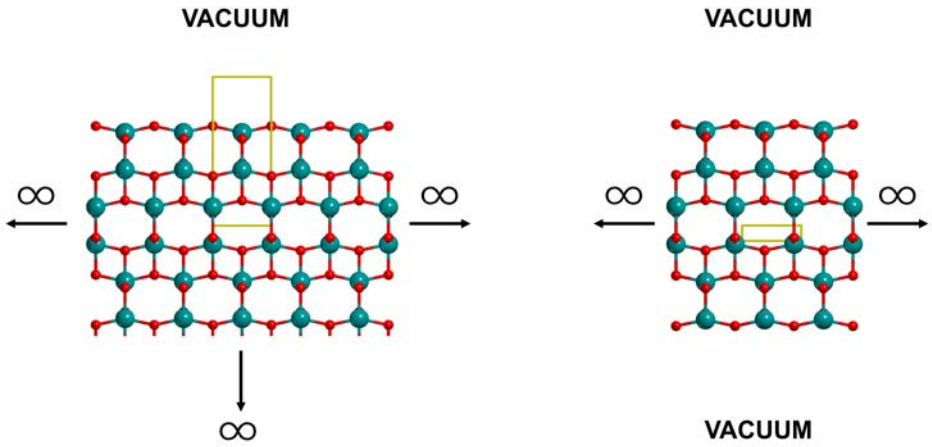


Figure 2.6: Real (left) and model (right) surface.

by edge effects of the surface. To ensure the leading to converged properties of the surface, one can check the surface formation energy, *i.e.* the energetic cost to cut a surface from the bulk, which can be computed as:

$$E_S = \frac{E(n)_{SLAB} - nE_{BULK}}{2A} \tag{2.44}$$

where $E(n)_{SLAB}$ is the energy of the slab containing n-layers, E_{BULK} is the bulk energy and A is the surface area of the slab.[201] E_S should converge for a limited, but increasing, number of layers. Another important requirement is that the surface should be neutral, *i.e.* the stoichiometry of the bulk must be respected. The reason of this is quite obvious; if the unit cell of the surface is charged, that charge will be repeated by translational symmetry, leading to an infinite charged structure, which cannot be stable. The last prerequisite is that the net dipole moment along the non-periodic direction is almost 0. If

the thinnest slab, *i.e.* the 1-layer one, has a dipole moment it will be repeated n-times with the increasing number of layers; large dipole moments may cause instabilities in the electronic structure of the slab leading to a conducting system (electron transfer from valence bands to conducting bands), even if the surface should not be a conductor.

Basis sets

Many methods have been developed by theoretical chemists in order to mathematically treat the atoms of a system. We can distinguish between two classes of basis functions: localized basis sets, which, in turn, can be represented by Slater-type functions, Gaussian functions, or numerical atomic orbitals, and plane waves, a combination of sine and cosine functions which are not centred on nuclei but uniformly fill the space of a given cell. All the basis sets present either *pros* and *cons*, and one has to carefully take them into account depending on the system under study. Generally, Gaussian functions are mostly used to study molecular systems, while plane waves are preferred for periodic systems.

Also Gaussian functions may be used in periodic codes (such as CP2K[202–208] and CRYSTAL[209, 210]), but they are expressed as Gaussian-type Bloch functions, in order to take into account the periodicity of the system (see previous sections). Each Gaussian basis function is a linear combination of Gaussian primitives with different exponent, with the aim of describing with a good accuracy both core electrons (high exponents) and valence electron (low exponents) which are far from the nuclei.

Because of their nature, the number of plane waves does not de-

pend directly by the number of atoms of the system, but by the size of the cell. Because they fill all the space of a cell, using plane waves intrinsically leads to 3D-periodic calculations, also for molecules; this is due to the fact that plane waves cannot be exposed to the vacuum. This means that also calculations of small molecules can be relatively expensive. By contrary, using localized basis sets, the calculation of small molecules is very fast, because the number of basis functions depends on the number of atoms of the system.

An incredible advantage of plane waves is that they constitute an universal basis sets, because they neither depend on the position of the atoms in the cell nor by their nature, and one may choose a more accurate (and more expensive) or a less accurate (and less expensive) treatment of the system only by changing one parameter: the cutoff energy value. In contrast, Gaussian basis sets have to be calibrated and optimised for each system. Moreover, one may play with the accuracy and the cost of the basis set by adding or removing core functions (which describe core electrons), valence functions (which describe the outermost electrons), but also polarized and diffuse functions (which, by adding virtual orbitals, give a better description of the chemical bond and of electrons far from the nucleus, respectively).

Another important aspect of plane waves is that they are BSSE (basis set superposition error)-free. BSSE is a spurious stabilizing effect due to the overlap of the basis function centred on different molecules when they approach to each other (intermolecular BSSE) but also of different part of the same molecule (intramolecular BSSE); this error derives from the incompleteness of Gaussian basis sets, while plane waves, which are intrinsically infinite, form a complete

set by definition.

A disadvantage of using plane waves is that, despite their efficiency in many parts of the SCF procedure for pure GGA methods, the inclusion of exact Hartree-Fock exchange is very expensive, in comparison to localized functions; to give an idea, while with Gaussian basis sets the cost of a hybrid method is 3/4 times more expensive than a pure GGA, with plane waves this cost rise to 2 orders of magnitude.

Finally plane waves do not describe properly core electrons, which are strictly localized close to the nuclei, but this can be solved by using pseudopotentials for the description of ionic cores and plane waves for valence electrons.

2.7 Computational algorithms

The plethora of available computational packages allow the scientific community to simulated a very wide variety of chemical processes, from the most simple ones, such as energy calculations and geometry optimizations, to more complex and realistic situations, like MD simulations of large systems (proteins, interface of proteins and mineral surfaces), elastic and piezoelectric properties of solids, simulation of IR, UV-Vis, NMR, and many other spectra.

In the following, we will discuss the main algorithms used in this work.

2.7.1 The SCF procedure

As said in the previous Section, the Hartree-Fock is a self-consistent method; this means that it starts with a trial function, which will be a rough approximation of the real one, to finally land, through an iterative procedure, to a wave function which gives an energy as similar as possible to the real one. A schematic representation of the SCF procedure for a periodic calculation is shown in Figure 2.7. The set of equations to be solved are the Roothaan-Hall equations:

$$F^{\mathbf{k}}C^{\mathbf{k}} = S^{\mathbf{k}}C^{\mathbf{k}}\epsilon^{\mathbf{k}} \tag{2.45}$$

where F is the Fock matrix, S the overlap matrix, C the coefficient matrix, ϵ a diagonal matrix representing the energy of each molecular orbitals, and $\mathbf{k}$ represents the $\mathbf{k}$-point set.

During the years several algorithms have been developed in order to accelerate the SCF convergence.[211] Among many it is worth citing the damping technique where elements of the matrix at the cycle n are mixed with elements of the matrix at the cycle $n-1$ in order to prevent energy oscillations. Another technique is the level shifting where occupied MO and virtual MO (or CO) are separated by an amount of energy, in order to avoid electronic state changes during the SCF procedure due to the presence of near degenerate orbitals.[212] Finally, perhaps the most popular method is the DIIS (direct inversion in the iterative subspace),[213] which is based on an extrapolation procedure. In particular, it stores the storyboard of previous Fock matrices in order to extrapolate the trend of those matrices. Despite it may fail for bad geometries, for reasonable structures it ensures a very fast convergence.

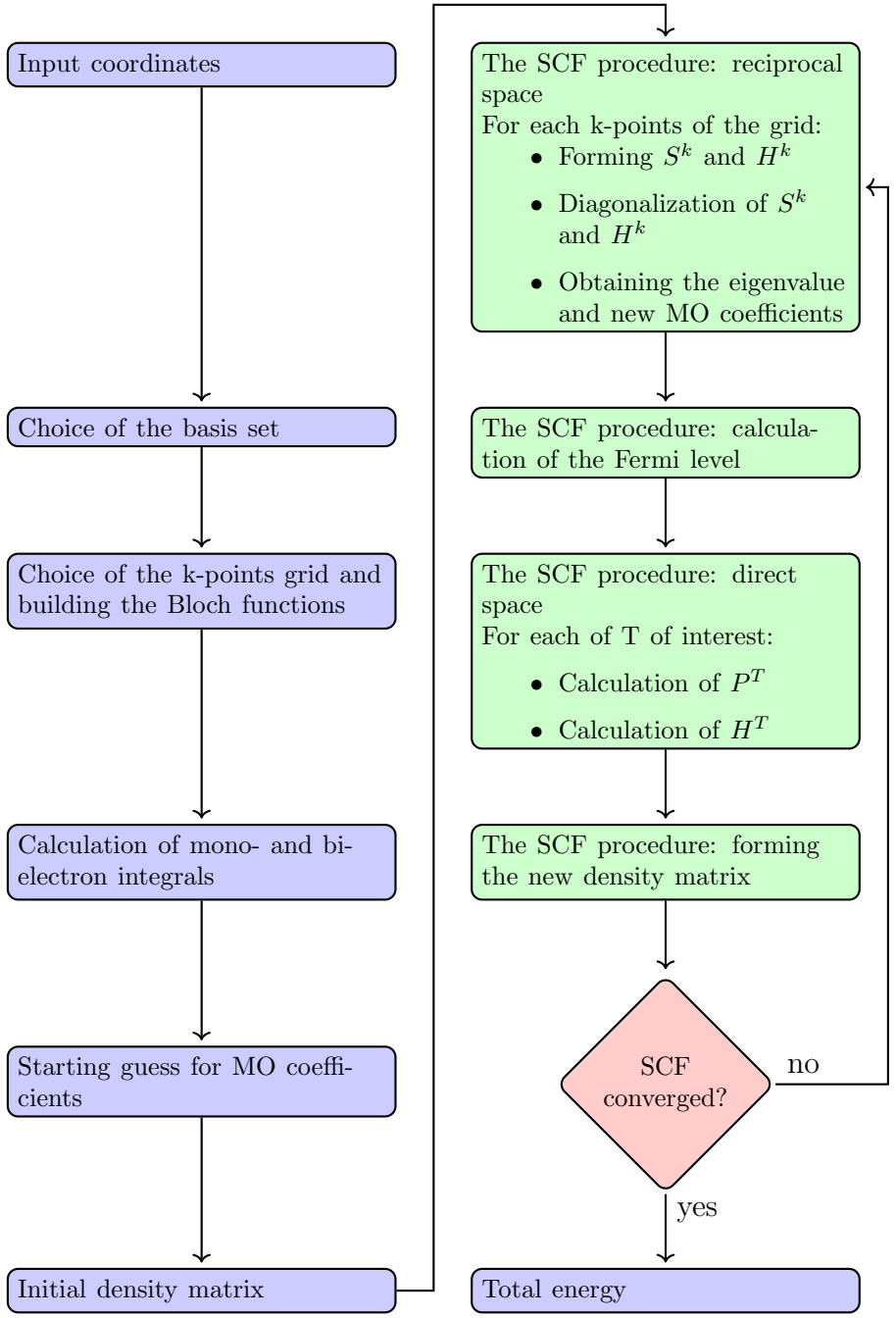


Figure 2.7: Schematic representation of the SCF steps in periodic calculations.

2.7.2 Exploring potential energy surfaces

The geometry optimization is also an iterative procedure which moves the atoms of a system along the PES (potential energy surface) in the sense of minimizing the energy of the system itself, until the forces acting among the atoms are zero. Thus, one can imagine the PES as a hyperspace of at least $3N$ dimensions, where N is the number of atoms of the system, or, more generically, the degrees of freedom of the system, in which each point corresponds to a determined geometry of the system. During the optimization process the system leads to a different structure with respect to the previous ones and, because of this, a SCF calculation is required for each optimization step. The mathematical tool which mainly guides the geometry optimization procedure is the gradient calculation; once the gradients tend to zero the optimization stops and the stationary point is found. As for the SCF, also in this case each computational code includes many different optimization algorithms, which can be more or less fast and robust in reaching the stationary point.

The most simple and intuitive method is the Steepest Descendent (SD).[214] The gradient vector g points in the direction where PES increases most, thus the search direction is $d = -g$. Along this direction, a series of functions are evaluated and, where the function starts to increase an approximate minimum may be found by interpolation between the calculated points. In the new point a new gradient will be calculated and the procedure will be repeated. The advantage is that it is a very simple method, and it only requires the storage of a gradient vector. However, it can only be used to find minima (and not other interesting stationary points), and when it approaches to

a minimum it goes in troubles. Moreover it is problematic for PES with long narrow valleys.

An improvement over the SD method is given by the Conjugate Gradient (CG) method;[215] in this case the line search direction is not calculated only on the point "n", but is mixed with the previous line searches (obviously the first step is given by a pure SD method). This method is quite reliable and very robust to find local minima.

Another class of optimization algorithms is given by the Newton-Raphson method. It expands the function $f(x)$ to the second order around the current point x_0:

$$f(x) \approx f(x_0) + g(x - x_0) + \frac{1}{2}(x - x_0)H(x - x_0) \tag{2.46}$$

and requiring the gradient to be zero, it produces the step:

$$(x - x_0) = -H^{-1}g \tag{2.47}$$

where H is the Hessian matrix, *i.e.* the second derivatives of the energy with respect to the atomic position, which is quite demanding from a computational point of view. The class of methods which use a Newton-Raphson approach, usually starts with an approximation of the Hessian (it can be a unit matrix), which is updated during the optimization process thanks to the gradient calculation. The several updating schemes lead to the different so called pseudo-Newton-Raphson methods, the most popular being the BFGS one, where the Hessian update can be based on the full story of the previous gradient,[216–218] or only the last m gradients, leading to the lim-

ited memory version of the BFGS (L-BFGS), particularly useful from large systems.[219–222] It is possible to combine this approach with extrapolation methods; among many, the most reliable is the GDIIS (Geometry direct inversion in the iterative subspace)[223] which uses the same idea of the DIIS for the SCF, *i.e.* it stores and interpolates gradient calculations of previous steps.

The situation is slightly different when one is interested in exploring reaction paths, because in this case one wants to search for not only minima, but also transition state structures (*i.e.* stationary points represented by first order saddle points) which connect the minima. In this case what the optimization algorithm does is the minimization of all the coordinates but the one which connects reactants and products, which has to be maximized; the PES will be a minimum in all the directions, and a maximum in only one direction. The transition state search is quite difficult from an algorithmic viewpoint, and usually a previous frequency calculation is required, in order to have the real Hessian which will give to the optimizator the correct eigenvector to follow, along which the energy that has to be maximized.

2.7.3 Thermodynamic properties

From the calculated energy on the microscopic system, it is possible to extrapolate macroscopic thermodynamic quantities, such as the Enthalpy and the Gibbs energy. This is possible thanks to the statistical thermodynamic, which is the bridge connecting the microscopic world with the macroscopic one, throughout the calculation of the partition functions Q within the canonical ensemble approximation:

$$Q(N,V,T) = \sum_i e^{-\frac{E_i(N,V)}{k_B T}} \tag{2.48}$$

where i runs over the possible energy states of the system and k_B is the Boltzmann constant. The thermodynamic quantities necessary to reach the Gibbs energy are:

$$U = k_B T^2 \left(\frac{\partial ln Q}{\partial T} \right)_{N,V} \tag{2.49}$$

$$H = U + PV \tag{2.50}$$

$$G = H - TS \tag{2.51}$$

$$S = k_B ln Q + k_B T \left(\frac{\partial ln Q}{\partial T} \right)_{N,V} \tag{2.52}$$

Where U is the internal energy of the system, H is the Enthalpy, G the Gibbs (or free) energy and S the entropy. The partition function Q is factorized in its electronic, vibrational, rotational and translational parts:

$$Q = Q_{el} + Q_{vib} + Q_{rot} + Q_{trans} \tag{2.53}$$

Q_{el} is given by a normal energy calculation, while for Q_{vib} a frequency calculation within the harmonic oscillator approximation is requested, which means to compute the Hessian, *i.e.* the second derivatives of the energy with respect to the atomic coordinates, which, in turn, can be diagonalized to finally obtain the vibrational

normal modes of the system. Q_{rot} and Q_{trans} are evaluated within the rigid-rotor approximation. The last two terms of Equation 2.53 are neglected in solid systems. The calculation of the Hessian is also a useful proof to be sure the structure is really a stationary point, either a minimum (all the elements positive) or a transition state (all the elements positive but one negative).

2.7.4 AIMD: *Ab-Initio* Molecular Dynamics

Since now we talked about frozen nuclei; their motion during the optimization process does not mean that we are following the nuclei motion in time, but they are only following gradients in order to reach a stationary point. Once the stationary point has been found the nuclei are frozen in their position; in the types of calculation described before there is neither dependence of temperature nor time.

Molecular dynamic simulations allow to study the dynamical properties of a system following the evolution of the system in time.[224–227] This is possible within the approximation that nuclei are heavy particles, so that they can be treated as classical particles, thus solving the Newton second law:

$$-\frac{dV}{dr} = m\frac{d^2r}{dt^2} \tag{2.54}$$

where V is the potential energy, r the nuclei position and t the time. One of the milestones of molecular dynamics derives by the studies of Car and Parrinello, with the technique known as *ab-initio* molecular dynamics.[228] Despite the nuclei motion is classically treated, for each nuclei position energy and gradient of the system are computed by QM methods; then, following the gradient, the nuclei position

and velocities are accordingly updated. For a small time step Δt the positions are given by a Taylor expansion:

$$r_{i+1} = r_i + \frac{\partial r}{\partial t}(\Delta t) + \frac{1}{2}\frac{\partial^2 r}{\partial t^2}(\Delta t^2) + \dots \tag{2.55}$$

Analogously, the previous time step will be given by:

$$r_{i-1} = r_i - \frac{\partial r}{\partial t}(\Delta t) + \frac{1}{2}\frac{\partial^2 r}{\partial t^2}(\Delta t^2) + \dots \tag{2.56}$$

Adding Equation 2.55 to Equation 2.56 we have:

$$r_{i+1} = (2r_i - r_{i-1} + \frac{1}{2}\frac{\partial^2 r}{\partial t^2}(\Delta t^2) + \dots \tag{2.57}$$

$$\frac{\partial^2 r}{\partial t^2} = a_i = -\frac{1}{m_i}\frac{dV}{dr_i} \tag{2.58}$$

This is the Verlet algorithm[229] which means that, at each time step the acceleration is evaluated by the forces acting on each particle of the system. The velocities at the initial step are evaluated randomly according to a Maxwell-Boltzmann distribution at the initial temperature or by a previous frequency calculation. The time step has to be chosen carefully, depending on what we are interested in; typically, molecular motions (*i.e.* vibrations and rotations) occur with frequencies of about $10^{11} - 10^{14} s^{-1}$, thus the time step should in the order of femtoseconds ($10^{-15} s$). The smaller the time step, the better the numerical stability of the simulation.

Ensembles in Molecular Dynamics

In Molecular Dynamic simulations (both classical and *ab-initio*) it is possible to keep constant several thermodynamic quantities of the system, depending on what one wants to study; the different choice of the thermodynamic variables leads to different thermodynamic ensembles. The most popular ones are:

- NVE (also called micro-canonical) ensemble, where N (the number of particles), V (the volume of a given cell), E (the total energy of the system) are kept constant,
- NVT (canonical) ensemble, where N, V, T (the temperature) are kept constant,
- NPT (isothermal-isobaric) ensemble, where N, P (the pressure), T are kept constant,
- μVT (grand canonical) ensemble, where μ (the chemical potential),V , T are kept constant.

The disadvantage of the NVE ensemble is that it does not represent a real chemical process, because from an experimental point of view the energy is not conserved in a box of given dimensions, but may exchange with the "external world"; however it is useful for particular simulations. For example, if one wants to simulate a reaction in the interstellar medium, when energy exchanges are prevented by the lack of matter. The canonical ensemble and its derivatives (NPT and μVT) better represent real systems where the temperature is controlled by a thermal bath. Usually, the protocol for an AIMD

simulation is split in two parts: i) the equilibration, a relatively short period, normally within the NVT ensemble, used to reach the target temperature, and ii) the production, an indefinitely long simulation period in the desired ensemble to simulate the process one is interested in.

2.8 Software used

The majority of the calculations have been carried out with the Vienna *Ab-initio* Simulation Package (VASP),[230–233] both for static (geometry optimization, frequency calculation) and dynamic simulations. In this package crystalline orbitals are represented as linear combination of plane waves, evaluated over a **k**-points grid in the reciprocal space. As said before, plane waves cannot describe properly core electrons; thus, we used the PAW (projector-augmented wave) pseudopotentials to describe inner electron shells and plane waves for valence electrons. The choice of this particular basis set ensures an efficient and robust convergence of the SCF iterative procedure, and, at the same time, a good scaling over many hundreds of cores, as we tested on HPC (High Performance Computing) services.

For the calculations of anatase nanoparticles the CRYSTAL[209, 210] code has also been used. Many of the features of the CRYSTAL code are similar to the VASP ones; however, instead of a plane wave basis set, it uses either full electron GTOs (Gaussian Type Orbitals) or combined pseudopotentials (core electron) and GTOs (valence electrons)

Moreover, in order to analyse very big nanoparticles, we used the

GULP code,[234–236] which is completely different from the two previously mentioned ones, as it is a MM (molecular mechanics) based code. It allows the user to perform simulations (optimizations, transition state search, MM-MD simulations) on large systems, with a strong emphasis on condensed phase; thus, it is possible to perform calculations with periodic boundary conditions. Ones validated the accuracy of the GULP code, by comparing its results with the VASP and CRYSTAL ones, it allowed us to perform calculation on nanoparticles with thousands of formula units.

Chapter 3

Titanium dioxide bare structures

3.1 Introduction

In this chapter we will discuss the bulk and surface bare structures for the two most studied polymorphs of TiO_2, *i.e* anatase and rutile. In particular, we will focus on the starting setup of our work and the comparison with experimental results, in order to calibrate and choose our computational tools. Moreover, in the last part of the Chapter we will discuss the project carried out in the theoretical chemistry group of Torino, *i.e.* the building of anatase nanoparticles starting from the bulk structure, and their analysis with different methods.

3.2 Computational details

All the calculations of bulk and surface structures were carried out using the Perdew-Burke-Ernzerhof (PBE)[167] functional to obtain equilibrium geometries and energies. In order to choose the optimal computational parameters, which ensure a good convergence of the energy and a good agreement with experimental results, we carried out several tests on the bulk structure. Therefore, we tested the convergence of the energy ranging the k-point mesh from 4 to 16, as well as the kinetic energy cutoff for plane waves from 300 to 800 eV. Results indicated that the best compromise between the accuracy and the computational cost is to set the k-point mesh to 8 in each direction, and the cutoff for plane waves to 500 eV. To obtain a good match with experimental results (cell parameters and IR measurements) we used both the pure PBE and PBE-D2* (with *a posteriori* Grimme D2* correction, *i.e.* the D2 correction[184, 237] modified for solid systems.[238]) The SCF iterative procedure has been converged to $\Delta E = 10^{-6}$ eV while the tolerance on gradients for geometry optimizations has been set to 0.01 eV/Å for each atom in each direction, both for bulks and surfaces.[239–248] As VASP uses plane waves, the surfaces are artificially replicated in the direction perpendicular to the slab and, accordingly, the number of **k**-points along that direction has been set equal to 1. The c direction, *i.e.* the non-periodic one, has been set to 40 Å in order to avoid fictitious interactions between surface replicas. Phonon calculations have been carried out on the bulk structures by means of the supercell approach on the full Hessian to obtain accurate results in order to match experimental IR frequencies.

For the calculation of nanoparticles we use the VASP code with the computational parameters discussed above (the k-point mesh has been set to 1 as nanoparticles have no periodicity) using the pure PBE functional, and with a vacuum space of almost 10 Å among replicas.

To make a comparison with VASP results, also the CRYSTAL[209, 210] program has been used, with a full electron basis set both for Ti atoms (86-411(d31)G) and for O atoms (6-31G(d)). As CRYSTAL uses GTOs, hybrid functionals are available at reasonable computational cost and, accordingly, we performed calculation both at PBE and PBE0 levels. The DFT integration grid has been set to the default (XLGRID). Tolerances for SCF and geometry optimization convergence has been also set to the default. Integral tolerances have been set to 7 7 7 7 14.

Finally, also the GULP[234–236] program has been used, in order to optimised large nanoparticles, once the results have been demonstrated to be accurate, by comparing them with the VASP and CRYSTAL ones. The force field has been parametrized using a Buckingham potential[249] developed by Matsui and Akaogi,[250] but revisited for the anatase polymorph.[251]

To build the nanoparticles analysed in this work we used a home made program, which cuts stochiometric Wulff nanoparticles starting from the geometrical data of the bulk material and the surface energy of the most stable facets of the material of interest.

3.3 TiO_2 bulk structure

3.3.1 Anatase

The anatase bulk belongs to the I4/amd tetragonal space group ($a = b \neq c$, $\alpha = \beta = \gamma = 90°$), number 141 in the standard listing (see Figure 3.1). Ti atoms are hexacoordinated by O atoms of two different types: apical with Ti–O1 distances of 1.979 Å, and equatorial with Ti–O2 distances of 1.932 Å.[252] In all cases equatorial Ti–O2 bonds are described better than the apical Ti–O1 ones, as one can appreciate by the percentage errors in Table 3.1. Similarly, also the two shorter cell parameters (a and b) are closer to the experimental ones than the c parameter. Dispersive forces are not so important in this case, because TiO_2 has a highly ionic behaviour where electrostatic forces are dominant. However, as one can observe in Table 3.1 and 3.2, the D2* correction give better results with respect to the pure PBE. We have also carried out frequency calculations (see Table 3.2) on the bulk structure in order to further confirm our choice of the D2* correction; in particular we have performed phonon calculations, *i.e.* not only in Γ point, but taking into account the interaction between neighbouring cells by means of the supercell approach. Again, pure PBE is not so accurate in describing the IR vibrations as one can see by the percentage deviation errors in Table 3.2.

Table 3.1: Experimental, PBE, and PBE-D2* optimised structures for the anatase bulk, with the percentage error deviation with respect to the measures carried out at 15 K. The number in parenthesis are the temperatures (in K) at which neutron powder patterns are registered. Bond distances are in Å; volumes are in $Å^3$.

	Exp (295)	Exp (15)	PBE	Dev %	PBE-D2*	Dev %
Ti–O1 (apical)	1.9799	1.9788	2.0077	1.45	2.0093	1.53
Ti–O2 (equatorial)	1.9338	1.9322	1.9506	0.96	1.9406	0.45
a	3.7848	3.7822	3.8098	0.73	3.7878	0.15
b	3.7848	3.7822	3.8098	0.73	3.7878	0.15
c	9.5124	9.5023	9.7102	2.19	9.7292	2.39
Volume	136.26	135.93	140.94	3.68	139.59	2.69

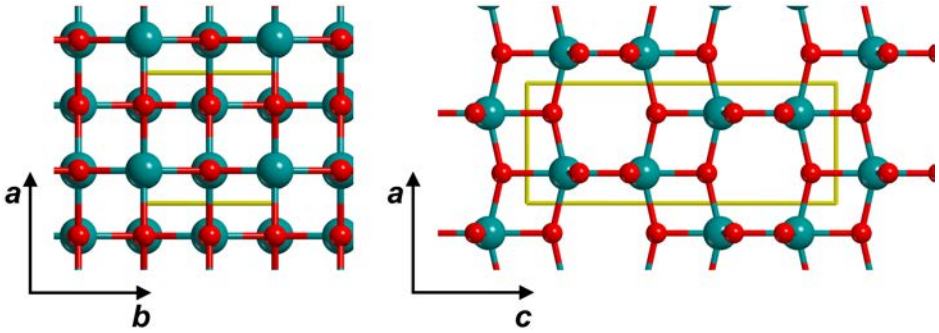


Figure 3.1: PBE-D2* optimised structure for the anatase bulk. *ab* plane on the left and *ac* plane on the right.

Table 3.2: PBE Frequency calculations with different level of dispersion correction for the anatase bulk, with the percentage error deviation with respect to experimental results.[253] Vibrations are in cm^{-1}.

Symmetry	Exp	PBE	Dev %	PBE-D2*	Dev %
A2u	367	442	20.44	384	4.63
Eu	262	223	14.88	240	8.40
Eu	435	389	10.57	406	6.67

3.3.2 Rutile

The rutile bulk belongs to the P42/mnm tetragonal space group ($a = b \neq c$, $\alpha = \beta = \gamma = 90°$), number 136 in the standard listing (see Figure 3.2). Ti atoms are hexacoordinated by O atoms of two different types: apical with Ti–O1 distances of 1.976 Å, and equatorial with Ti–O2 distances of 1.946 Å.[252] In Table 3.3 the experimental and computed cell parameters and bond distances are reported; as one can see also in this case D2* correction gives better results, both for geometries and frequency calculations (see Table 3.4), in comparison with experimental data. Thus, in consistency with the anatase system, we also chose PBE-D2*.

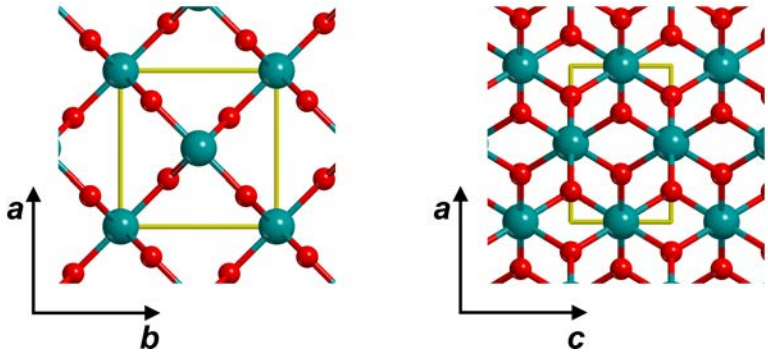


Figure 3.2: PBE-D2* optimised structure for the rutile bulk. *ab* plane on the left and *ac* plane on the right.

Table 3.3: Experimental, PBE, and PBE-D2* optimised structures for the rutile bulk, with the percentage error deviation with respect to the measures carried out at 15 K. The number in parenthesis are the temperatures (in K) at which neutron powder patterns are registered. Bond distances are in Å; volumes are in Å^3.

	Exp (295)	Exp (15)	PBE	Dev %	PBE-D2*	Dev %
Ti–O1	1.9796	1.9764	2.0040	1.42	1.9843	0.42
Ti–O2	1.9486	1.9459	1.9629	0.87	1.9609	0.77
a	4.5931	4.5867	4.6480	1.34	4.6137	0.59
b	4.5931	4.5867	4.6480	1.34	4.6137	0.59
c	2.9589	2.9541	2.9717	0.60	2.9745	0.69
Volume	62.42	62.15	64.20	3.30	63.32	1.88

Table 3.4: PBE Frequency calculations with different level of dispersion correction for the rutile bulk, with the percentage error deviation with respect to experimental results.[254–256] Vibrations are in cm^{-1}.

Symmetry	Exp	PBE	Dev %	PBE-D2*	Dev %
A2u	142	155	9.15	157	10.56
Eu	189	237	25.40	203	7.41
Eu	381	352	7.61	372	2.36
Eu	508	517	1.77	521	2.56

3.4 TiO_2 surface structure

3.4.1 Anatase

Starting from the crystal bulk structure of the TiO_2 anatase polymorph, we built crystalline periodic slab models for the nonpolar (101) surface with thicknesses ranging from ∼3 Å to ∼15 Å, which correspond to slab containing 1-5 TiO_2 layers. Computed surface energies showed that the 3-layer slab model is the best compromise between the accuracy and the computational cost (see Figure 3.3). The (101) surface presents a corrugated morphology (right part of Figure 3.4) with pentacoordinated Ti atoms and bivalent O atoms in the outermost positions of the slab, which correspond to Lewis acidic and Brønsted basic sites, respectively. In the electrostatic potential maps (left part of Figure 3.4) red zones correspond to the bivalent O rows, while blue zones correspond to the pentavalent Ti atoms. As a result of the surface reconstruction, the slab does not present apical and equatorial Ti–O bonds, despite that the intermediate layers maintain the bulk structure.

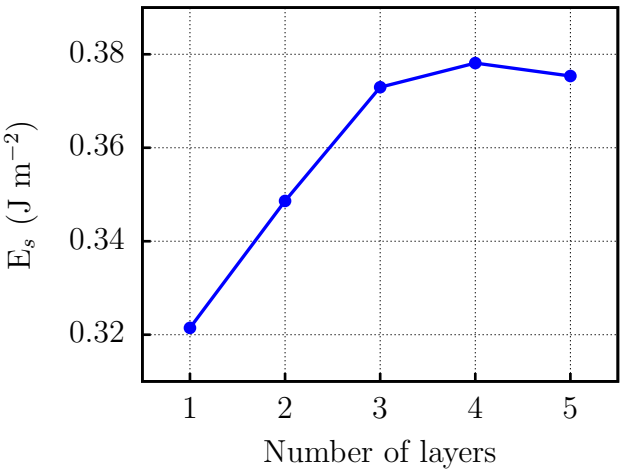


Figure 3.3: Surface formation energy (J m^{-1}) for the (101) surface of the anatase polymorph as a function of the number of TiO_2 layers.

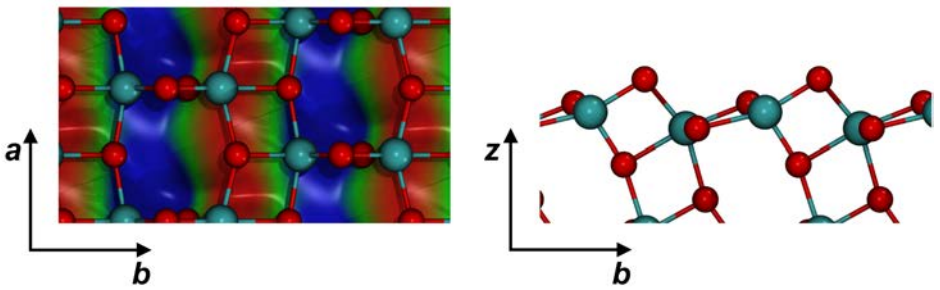


Figure 3.4: Electrostatic potential maps (left) and side view (right) of the (101) anatase surface. Negative values (-0.006 a.u.) are in red, positive values (0.006 a.u.) are in blue.

3.4.2 Rutile

Similarly to TiO_2 anatase polymorph, we started from the crystal bulk structure, and then we built crystalline periodic slab models for the nonpolar (110) surface with thickness ranging from ∼3 Å to ∼21 Å, which correspond to slab containing 1-7 TiO_2 layers. Computed surface energies showed that the 5-layer slab model is the best compromise between accuracy and computational cost (see Figure 3.5). The (110) surface presents a flat morphology (right part of Figure 3.6) with pentacoordinated Ti atoms and bivalent O atoms in the outermost positions of the slab which, as for the anatase case, correspond to Lewis acidic and Brønsted basic sites, respectively. In the electrostatic potential map (left part of Figure 3.6) red zones correspond to O atoms, while as previously mentioned for the anatase, the outermost atoms do not present apical and equatorial Ti–O bonds, as the internal part of the slab do.

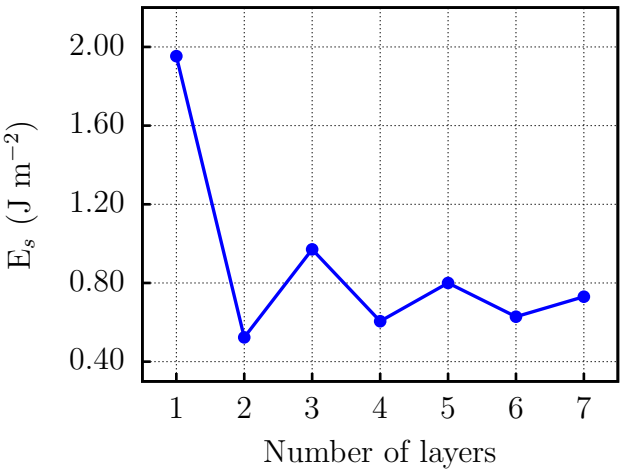


Figure 3.5: Surface formation energy (J m^{-1}) for the (101) surface of the rutile polymorph as a function of the number of TiO_2 layers.

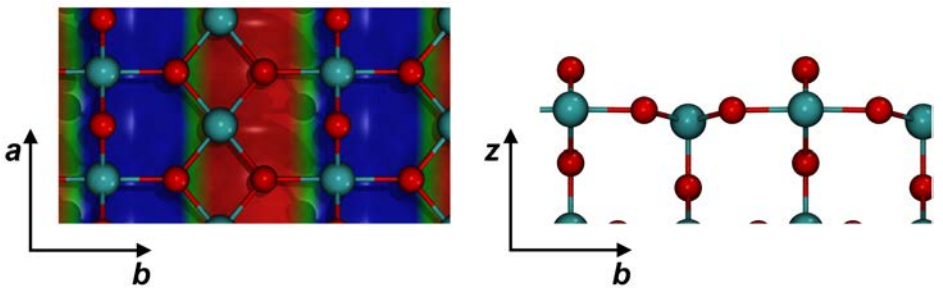


Figure 3.6: Electrostatic potential maps (left) and side view (right) of the (110) rutile surface. Negative values (-0.006 a.u.) are in red, positive values (0.006 a.u.) are in blue.

3.5 Anatase nanoparticles

The analysed Wulff anatase nanoparticles expose only the (101) facet of the TiO_2 anatase polymorph, *i.e.* the most stable one. We optimised nanoparticles of several sizes, starting from the smallest one, which has only 10 formula units, then going to increasing sizes, *i.e.* 35 and 84 formula units.

As one can see from Figure 3.7, FF energies match very well with the CRYSTAL ones. In contrast, VASP tends to overestimate the stability of the studied nanoparticles. This is probably due to the fact that VASP calculations has intrinsically 3D periodicity, because plane waves, by definition, fill uniformly all the space of a given cell. Accordingly, also in a 0D periodic system, there are basis functions all around the molecule (which is not the case of CRYSTAL calculations, where basis functions are localized only on atoms), thus inferring a fictitious stabilization to the system.

In Figure 3.8 and in Table 3.5 the aligned optimised structures and the RMSD among the several optimised structures, respectively, are reported. As expected, between CRYSTAL and VASP the differences are negligible, as also between PBE and PBE0 geometries. The GULP force field describes very well bulk atoms (*i.e.* the innermost atoms of the nanoparticles) and also pentacoordinated Ti atoms and bivalent O atoms on the nanoparticle surface; however, as shown in Figure 3.8, the atoms on the tip of the nanoparticle are quite far from CRYSTAL and VASP geometries, due to highly undercoordinated atoms (tetracoordinated Ti atoms and dangling O atoms) for which the force field is not correctly parametrized. A possible strategy to overcome this problem is to fix the coordinates of those atoms

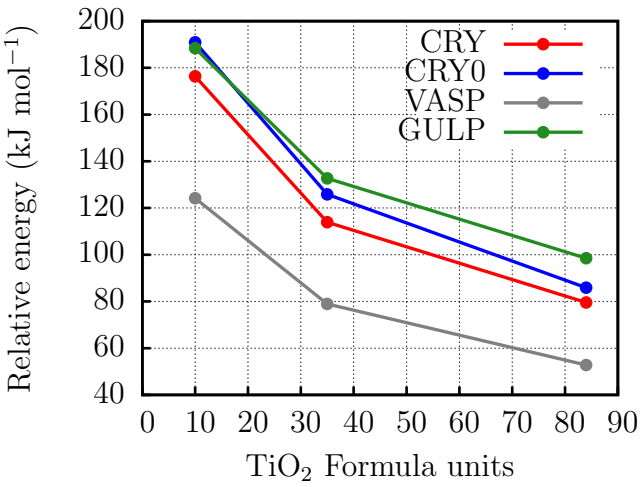


Figure 3.7: Relative energies of optimised nanoparticles with respect to the bulk structure normalized per TiO_2 unit *versus* the number of TiO_2 units in the nanoparticles studied.

which are not well described with the force field, or to use a more complex and accurate force field.

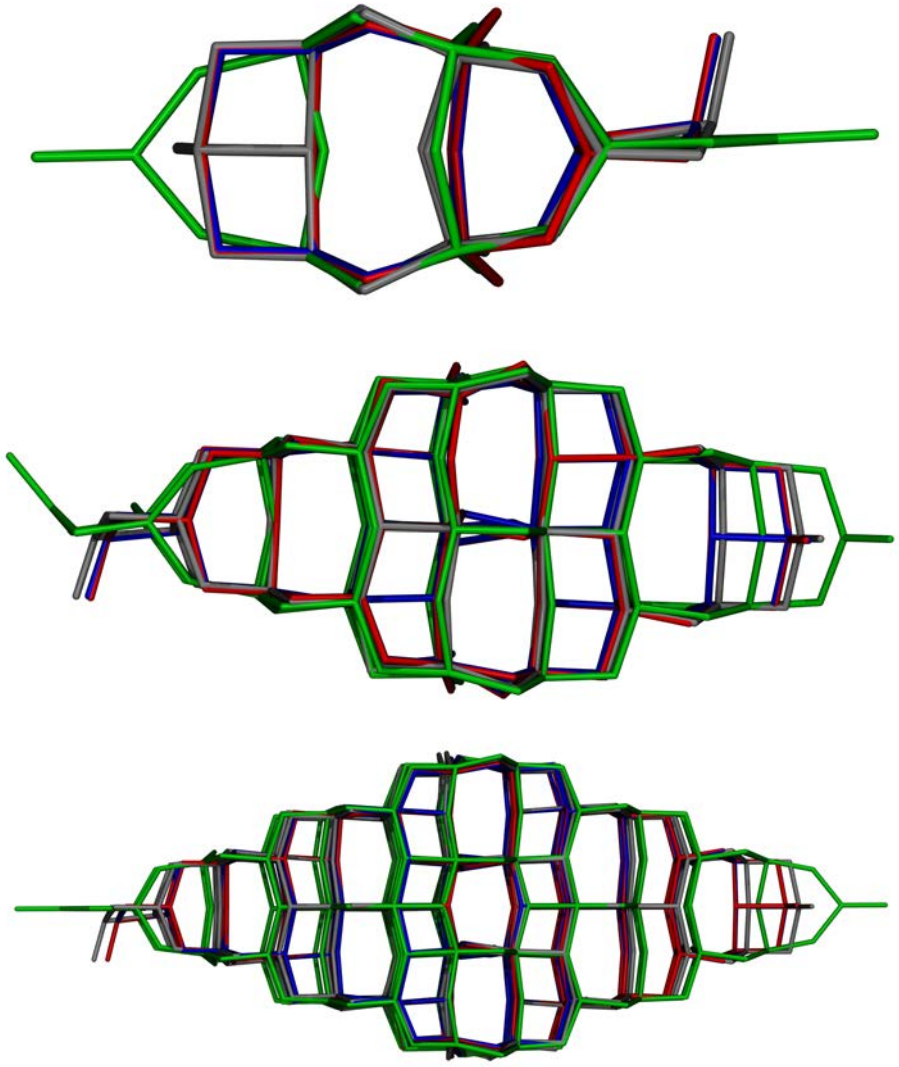


Figure 3.8: Aligned structures for the 10 (top), 35 (middle), 84 (bottom) units TiO_2 nanoparticles. Red is the CRYSTAL optimised structure at PBE level, blue is CRYSTAL optimised structure at PBE0 level, grey is VASP optimised structure at PBE level, green is the GULP optimised structure.

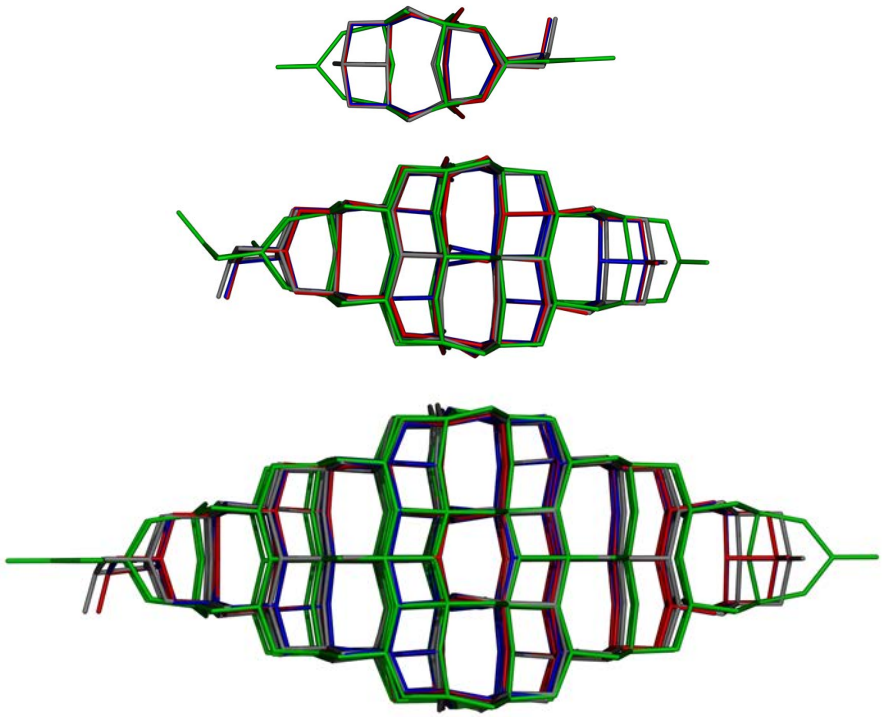


Figure 3.9: Aligned structures for the 10 (top), 35 (middel), 84 (bottom) units TiO_2 nanoparticles. Red is the CRYSTAL optimised structure at PBE level, blue is CRYSTAL optimised structure at PBE0 level, grey is VASP optimised structure at PBE level, green is the GULP optimised structure.

Table 3.5: RMSD among the optimised structures with the different codes and methods. CRY and CRY0 correspont to CRYSTAL calculations at PBE and PBE0 levels, respectively.

10 TiO_2 units	
System	RMSD
CRY-CRY0	0.012
CRY-VASP	0.197
CRY0-VASP	0.196
CRY-GULP	0.685
CRY0-GULP	0.686
VASP-GULP	0.573

35 TiO_2 units	
System	RMSD
CRY-CRY0	0.025
CRY-VASP	0.147
CRY0-VASP	0.156
CRY-GULP	0.581
CRY0-GULP	0.588
VASP-GULP	0.448

84 TiO_2 units	
System	RMSD
CRY-CRY0	0.002
CRY-VASP	0.236
CRY0-VASP	0.235
CRY-GULP	0.651
CRY0-GULP	0.651
VASP-GULP	0.465

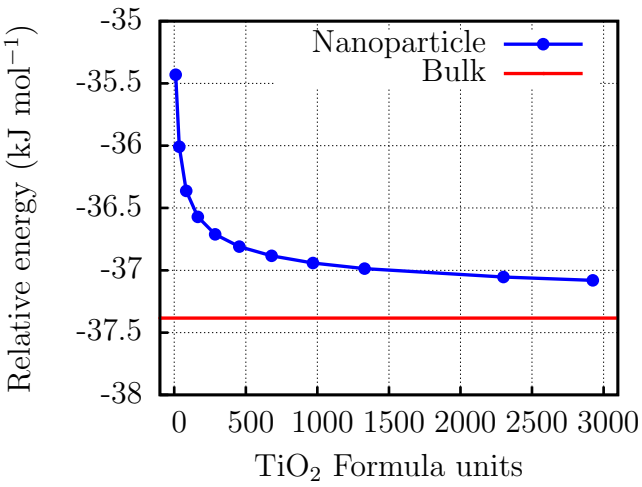


Figure 3.10: Relative energies of GULP optimised nanoparticles with respect to the bulk structure normalized per TiO_2 unit *versus* the number of TiO_2 units in the nanoparticles studied.

Finally, as we have demonstrated the good accuracy of GULP in calculating the relative energy of the nanoparticles with respect to the bulk one, we have carried out optimization with increasing nanoparticles sizes up to 8775 atoms. As one can see from Figure 3.10 the energy of the nanoparticles converges to the bulk by increasing the size of the cluster, as expected. Notably, in Figure 3.10 one may observe a similar behaviour to that of surface energy (see Figure 3.3 and 3.5); *i.e.* by increasing the size of the nanoparticles or the number of layers of a surface, the bulk structure and properties of inner atoms are correctly reproduced. For nanoparticles the convergence is achieved with almost 1000 atoms, as they present stronger edge effects than surfaces.

3.6 Final remarks

In this Chapter the bare TiO_2 material has been analysed both in its bulk and surface structures for the anatase and rutile polymorphs. The surfaces which we have taken into account are the most stable ones for each polymorph, namely (101) and (110) for anatase and rutile, respectively. The comparison between experimental and calculated data is presented, is order to justify the choice of the methodology adopted along all this work.

Finally, Wulff like anatase nanoparticles has been built with a home-made program developed in our research group, and they have been optimized both with QM and MM methods. As MM methods gave results comparable with QM ones, we have optimized nanoparticles of increasing sizes, up to 8775 atoms, showing a convergence of the energy of the clusters with respect to the bulk energy, as expected.

Chapter 4

Amino acids adsorption on TiO_2 anatase and rutile surfaces

In this Chapter we will present the results on the interaction of 11 amino acids with the TiO_2 (101) anatase and (110) rutile surfaces (*i.e.* the most stable surfaces for the two polymorphs), by means of periodic QM simulations. Several adsorption states, with the amino acids in their canonical, zwitterionic, or deprotonated forms, were considered. The effect of the temperature has been introduced by performing AIMD simulations in the NVT (T = 298 K) ensemble.

4.1 Introduction

The interaction between biomolecules and mineral surfaces is of great relevance for science, because it has important technologi-

cal applications in many fields[257–260] such as biomedicine,[261] nanotechnology,[262–264] and water treatment,[265, 266] among others.

The interaction of amino acids and peptides with mineral surfaces is also of particular interest in the field of prebiotic chemistry.[22] According to a seminal hypothesis proposed by Bernal,[1] mineral surfaces could have played a significant role in this step since amino acids can establish strong interactions with mineral surfaces, which in turn can protect and concentrate them and even activate them to carry out condensation reactions that would finally lead to the formation of the first oligopeptides.

Previous experimental and computational studies have analysed the adsorption of amino acids and other carboxylic acids on rutile (110) and anatase (101) TiO_2 surfaces.[24, 121–127, 267–269] The strongest interactions correspond to dative bonds between the N and O electron pairs of the amino acids and the Ti atoms of the surface. Results show that the polymorphs of the surface determine the preferred adsorption mode. That is, for glycine, the most stable configuration on anatase corresponds to the deprotonated (N,O) binding mode, whereas on rutile (110) surface, the deprotonated (O,O) binding mode was identified as the preferred one. The adsorption of the remaining amino acids is driven by a delicate trade-off among dative bond of the lateral chain (Arg, Glu, Gln, Lys, His), attractive dispersion forces (Phe, Leu, Met), H-bond interactions (Ser, Cys), and their sterical hindrance with the surface.

4.2 Computational details

4.2.1 Methods

All the calculations were carried out using the PBE-D2*[167, 184, 237, 238] method for energies and geometry optimizations. The SCF iterative procedure has been converged to $\Delta E = 10^{-6}$ eV while the tolerance on energy for geometry optimizations has been set to $\Delta E = 10^{-4}$ eV. The energy cutoff of the plane waves basis set has been set to 500 eV. The k-points mesh was set to (3,3,1) for the adsorption of the smallest amino acids (glycine, serine, cysteine, and leucine) and to (2,2,1) for the larger ones (methionine, lysine, glutamic acid, glutamine, arginine, phenylalanine, and histidine). The Monkhorst-Pack sampling of the Brillouin zone was used for the k-points mesh. Vibrational frequencies of low lying adsorbed glycine structures were computed, at the Γ point, by numerical differentiation of the analytical first derivatives, using the central difference formula (*i.e.*, two displacement for each atom in each direction), to confirm that optimised structures are minima and to estimate thermal effects. Results showed that located structures are indeed minima and that the relative stability does not change upon including thermal effects.

Ab-initio molecular dynamics (AIMD) simulations were carried out on glycine and serine low lying structures. The value of the energy cutoff for the plane wave basis set was set to 400 eV, while the self-consistent field (SCF) iterative procedure was converged to a tolerance in total energy of $\Delta E = 10^{-6}$ eV. The simulations were carried out considering an equilibration period of 1 ps (1000 steps of 1 fs), at fixed temperature (T = 300 K), followed by 10 ps of

production within the canonical (NVT) ensemble. During both the equilibration and production periods, the inner atomic layers of the TiO_2 slabs were maintained at fixed positions; *i.e.*, only the external atoms of the surface and those of the amino acids were allowed to move according to the motion's equations. We chose this option to avoid unrealistic deformations of the internal structure of the slabs, since real TiO_2 surfaces are linked to a macroscopic bulk.

4.2.2 Surface models

As discussed in Chapter 3, for the anatase the 3-layer slab model (thickness of ∼9 Å) has been used for the adsorption of all the amino acids studied, while for the rutile surface the convergence is reached with the 5-layer slab (thickness of ∼15 Å).

In order to avoid lateral interactions between adsorbates of adjacent cells, we have chosen different supercell models depending on the amino acid considered. The reference unit cell (*i.e.* the 2×2 supercell) used for the smallest amino acids (glycine, serine, cysteine, and leucine) has periodic parameters $a = 7.569$ Å and $b = 10.23935$ Å. For amino acids of intermediate size (methionine, lysine, glutamic acid, glutamine, and arginine), we used a 2×4 supercell ($a = 7.569$ Å, $b = 20.4787$ Å), while for the largest amino acids (phenylalanine and histidine) we used a 4×4 supercell ($a = 15.138$ Å, $b = 20.4787$ Å). Regarding the rutile surface, for the smallest amino acids (glycine, serine, cysteine) we used a 3×2 supercell with parameters $a = 8.8761$ Å and $b = 12.9929$ Å. For the other amino acids studied (leucine, methionine, phenylalanine, histidine, lysine, glutamic acid, glutamine and arginine) we used a 4×2 supercell ($a = 11.8348$ Å, $b = 12.9929$

Å). The interlayer distance, regulated by the c value, was defined considering an empty space of 10 Å between the most external atom of the amino acid and the upper layer, which is enough to avoid mutual fictitious interactions between the periodically repeated slabs. Accordingly, the c value ranges between 25 and 30 Å, depending on the size of the amino acids for the anatase surface, while it has been set to 30 Å for the rutile surface.

Adsorption energies have been calculated as follows:

$$\Delta E_{ADS} = E_{CPLX} - (E_{AA} + E_{SURF}) \tag{4.1}$$

where E_{CPLX} is the energy of the complex "amino acid/TiO_2 surface", E_{AA} is the energy of the amino acid in its most stable gas phase conformation, and E_{SURF} is the energy of the bare surface, each one optimised separately. Moreover, we added another term in Equation 4.1 in order to simulate very low coverage regimes:

$$\Delta E_L = E_{AA(a,b,c)} - E_{AA(30,30,30)} \tag{4.2}$$

where $E_{AA(a,b,c)}$ is the energy of the amino acid alone in its geometry, orientation, and cell parameters as adsorbed on the surface, and $E_{AA(30,30,30)}$ is the same of $E_{AA(a,b,c)}$ but using a cubic cell of $30 \times 30 \times 30$ Å^3. In this way we can evaluate lateral interactions between adsorbate replicas. If ΔE_L is negative the lateral interactions are attractive and overstabilise the complex, and *vice versa*.[270]

4.3 Amino acids adsorpion on the (101) TiO_2 anatase surface

It is well-known that the most stable form of amino acids in gas phase is the canonical one, whereas that in water at pH = 7 is the zwitterionic one. Thus, one may expect that adsorption from gas phase occurs in the canonical form. However, as already mentioned, surface Ti atoms can act as Lewis acid sites and interact with the lone pairs of electron donor atoms of the amino acids, the most relevant ones in the context of this work being O, N and S. Furthermore, surface O atoms are Brønsted basic sites that may act as H-bond acceptors. Because of that, the surface is capable of stabilizing, in addition to the canonical (C) form, the zwitterionic (Z) and also the deprotonated (D) (the amino acid transfers its proton to the surface) forms, which are unstable in gas phase. Thus, adsorption of amino acids adopting these three different forms has been explored in all cases.

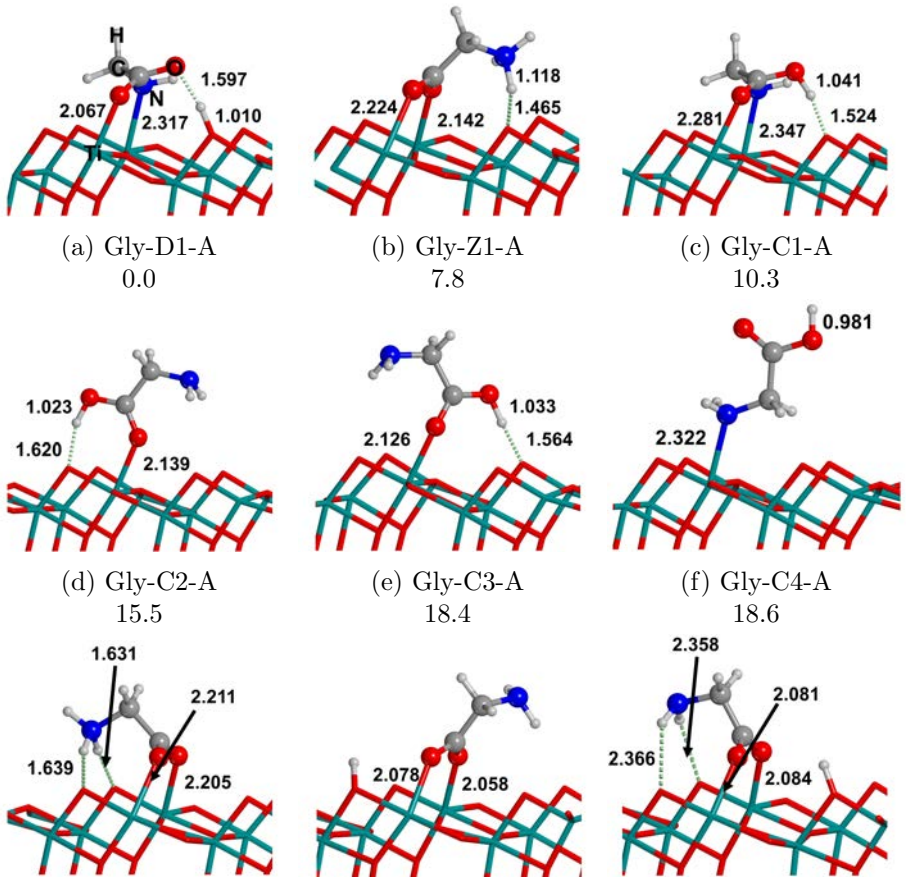
4.3.1 Glycine

Several glycine adsorption states, using as benchmark those structures reported in the work of Szieberth et al.[267] were investigated. C, D and Z forms were considered adopting different configurations upon adsorption. Figure 4.1 shows the different complexes found, and Table 4.1 reports the computed adsorption energies corrected for lateral interactions (ΔE^C_{ADS}) as well as the relative energies. The most stable adduct corresponds to a deprotonated state (Gly-D1-A). This complex is followed by the corresponding most stable zwitte-

rionic (Gly-Z1-A) and canonical (Gly-C1-A) forms. For these three cases, two atoms of the amino acid, (N,O) or (O,O), interact with two Ti surface atoms, in such a way that coordination becomes pseudo-octahedral (as in the bulk), thus inferring an additional stability. Moreover, in these three cases, H-bonds are also established.

For Gly-D1-A, because of the deprotonation of glycine, the surface acts as proton donor and the carboxylate as proton acceptor, while in Gly-Z1-A and Gly-C1-A the proton donor is glycine. Gly-D1-A and Gly-Z1-A are more stable than Gly-C1-A because the two former structures present a COO- group, which establishes more efficient electrostatic interactions with the surface than the –COOH group. This is reflected by the Ti–O distances: 2.067 Å, 2.142 Å, and 2.281 Åfor Gly-D1-A, Gly-Z1-A, and Gly-C1-A, respectively. Gly-D1-A is more stable than Gly-Z1-A because the Ti–N bond is stronger than the Ti–O one. This is validated by the calculated adsorption energy of NH_3 and H_2O on the (101) anatase surface (-112.9 and -86.9 kJ mol^{-1}, respectively).

The remaining structures are less stable (relative energies are larger than 15 kJ mol^{-1}) because they either present less adsorption anchoring points, do not present H-bonds, or the interactions with the surface are weaker. The preference for an adsorption through a (N,O) binding with two Ti atoms on anatase (101) surface has already been described by Szieberth et al.[267] However, this coordination was only explored for glycine in its canonical form, and not for the deprotonated one, which according to our results is the most stable adduct on this surface. Note that the (N,O) binding of the deprotonated form (Gly-A-D1) is 10.3 kJ mol^{-1} more stable than the (N,O) binding of the canonical one (Gly-C1-A).



(a) Gly-D1-A 0.0 (b) Gly-Z1-A 7.8 (c) Gly-C1-A 10.3

(d) Gly-C2-A 15.5 (e) Gly-C3-A 18.4 (f) Gly-C4-A 18.6

(g) Gly-Z2-A 38.0 (h) Gly-D2-A 38.6 (i) Gly-D3-A 51.3

Figure 4.1: PBE-D2* optimised geometries for glycine (Gly) adsorbed on the TiO_2 (101) anatase surface. "C" refers to Gly in its canonical form, "D" in its deprotonated form, and "Z" in its zwitterionic form. A refers to the anatase polymorph. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

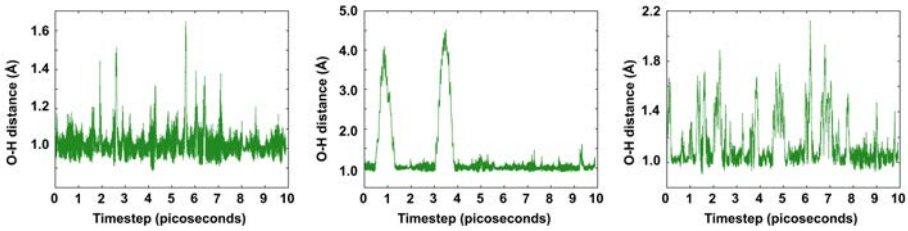
Table 4.1: Calculated energetic and geometric data for all the amino acids analysed.

Complex	ΔE^C_{ADS}	ΔE_{rel}	Ti–N	Ti–O1	Ti–O2
Gly-D1-A	-108.2	0.0	2.32	2.07	–
Gly-Z1-A	-100.4	7.8	–	2.22	2.14
Gly-C1-A	-97.9	10.3	2.35	2.28	–
Gly-C2-A	-92.7	15.5	–	2.14	–
Gly-C3-A	-89.8	18.4	–	2.13	–
Gly-C4-A	-89.6	18.6	2.32	–	–
Gly-Z2-A	-79.6	38.0	–	2.2	2.21
Gly-D2-A	-70.2	38.6	–	2.08	2.06
Gly-D3-A	-56.9	51.3	–	2.08	2.08
Leu-Z-A	-101.9	0.0	–	2.32	2.11
Leu-C-A	-77.3	24.6	2.45	2.29	–
Leu-D-A	-72.1	29.8	2.42	2.06	–
Met-Z-A	-109.1	0.0	–	2.19	2.12
Met-D-A	-85.1	23.9	2.41	2.05	–
Met-C-A	-80.3	28.8	2.48	2.3	–
Phe-Z-A	-93.8	0.0	–	2.19	2.12
Phe-C-A	-76.9	16.5	2.45	2.27	–
Phe-D-A	-76.1	17.4	2.42	2.02	–
Ser-Z-A	-116.8	0.0	–	2.25	2.14
Ser-D1-A	-100.8	15.9	2.39	2.01	–
Ser-D2-A	-96.3	20.5	2.34	2.03	–
Ser-C1-A	-95.7	21.1	2.44	2.23	–
Ser-C2-A	-93.8	23.0	2.41	2.2	–
Cys-Z-A	-114.3	0.0	–	2.24	2.14
Cys-D1-A	-102.9	11.4	2.37	2.02	–
Cys-D2-A	-98.2	16.1	2.35	2.02	–
Cys-C1-A	-95.6	18.7	2.43	2.26	–
Cys-C2-A	-94.1	20.2	2.43	2.24	–
Glu-D1-A	-112.6	0.0	2.48	2.16	–
Glu-C1-A	-109.0	3.5	2.43	2.33	–
Glu-Z-A	-103.5	9.0	–	2.2	2.13
Glu-D2-A	-85.2	27.4	2.4	2.03	–
Glu-C2-A	-78.4	34.2	2.45	2.28	–
Gln-D-A	-127.5	0.0	2.49	2.14	–
Gln-C-A	-124.7	2.8	2.43	2.34	–
Gln-Z-A	-110.8	16.7	–	2.21	2.14
Arg-Z-A	-106.1	0.0	–	2.22	2.14
Arg-D1-A	-103.6	2.6	2.56	2.23	–
Arg-C1-A	-100.9	5.2	2.71	2.26	–
Arg-D2-A	-64.4	41.7	2.54	2.05	–
Arg-C2-A	-63.8	42.3	2.62	2.28	–
Lys-D1-A	-139.2	0.0	2.38	2.11	–
Lys-C1-A	-128.0	10.2	2.41	2.33	–
Lys-Z1-A	-114.8	24.5	–	2.2	2.14
Lys-C2-A	-85.8	53.5	2.45	2.23	–
Lys-D2-A	-83.0	55.2	2.41	2.04	–
Lys-Z2-A	-31.6	107.6	–	2.16	2.15
His-D-A	-116.2	0.0	2.38	2.02	–
His-Z-A	-110.4	5.8	–	2.19	2.13
His-C-A	-110.2	6.0	2.43	2.27	–

ΔE^C_{ADS} is the adsorption energy of the complex calculated with respect to isolated amino acid and surface, and corrected for the lateral interaction; ΔE_{rel} is the relative energy respect to the most stable structure of each amino acid. In the last three columns the Ti–N and Ti–O1 bond lengths of the backbone are reported, while the Ti–O2 represents the interaction of the lateral chain with the surface. The energy values are in kJ mol^{-1} and the bond distances are in Å.

This preference for a (N,O) coordination is different to that observed for the adsorption of glycine on rutile (110) surface, for which the deprotonated (O,O) binding mode was determined to be the preferred one. In the present case, the adduct resulted from the deprotonated (O,O) binding, which corresponds to the Gly-D2-A structure, lies significantly higher in energy (38.6 kJ mol^{-1}). This different behaviour is attributed to the surface morphology between the surface polymorph. Indeed, the Ti–Ti distance in the (101) anatase surface is 3.81 Å, which is close to the distance between the C atom of the carboxylic group and the N atom of glycine (3.72 Å), thus allowing for a very efficient (N,O) interaction with the Ti atoms. However, the Ti–Ti distance in the (110) rutile surface is significantly smaller, and hence that the (O,O) interaction is more efficient.

In view of the low relative energies of Gly-D1-A and Gly-C1-A (10.3 kJ mol^{-1}) and to analyse temperature effects and validate that the preferred adsorption mode is the deprotonated structure, we carried out 10 ps (NVT) molecular dynamics simulations starting from each of these two structures. Additionally, we also carried out molecular dynamics simulations for the zwitterionic Gly-Z1-A structure. Figure 4.2 shows the O(surface)-H(acid) bond length for Gly-D1-A (Figure 4.2a), Gly-C1-A (Figure 4.2c), and Gly-Z1-A (Figure 4.2b). Results from the simulations indicate that, in the very first 100-150 steps of the equilibration, the canonical structure converts to the deprotonated structure, which remains as such for the majority of the simulation (Figure 4.2c). Gly-D1-A AIMD simulation confirms the results obtained for Gly-C1-A: the H atom remains on the surface during almost all the simulation (Figure 4.2a), except very few femtoseconds in which the Gly-D1-A structure converts to the Gly-C1-A



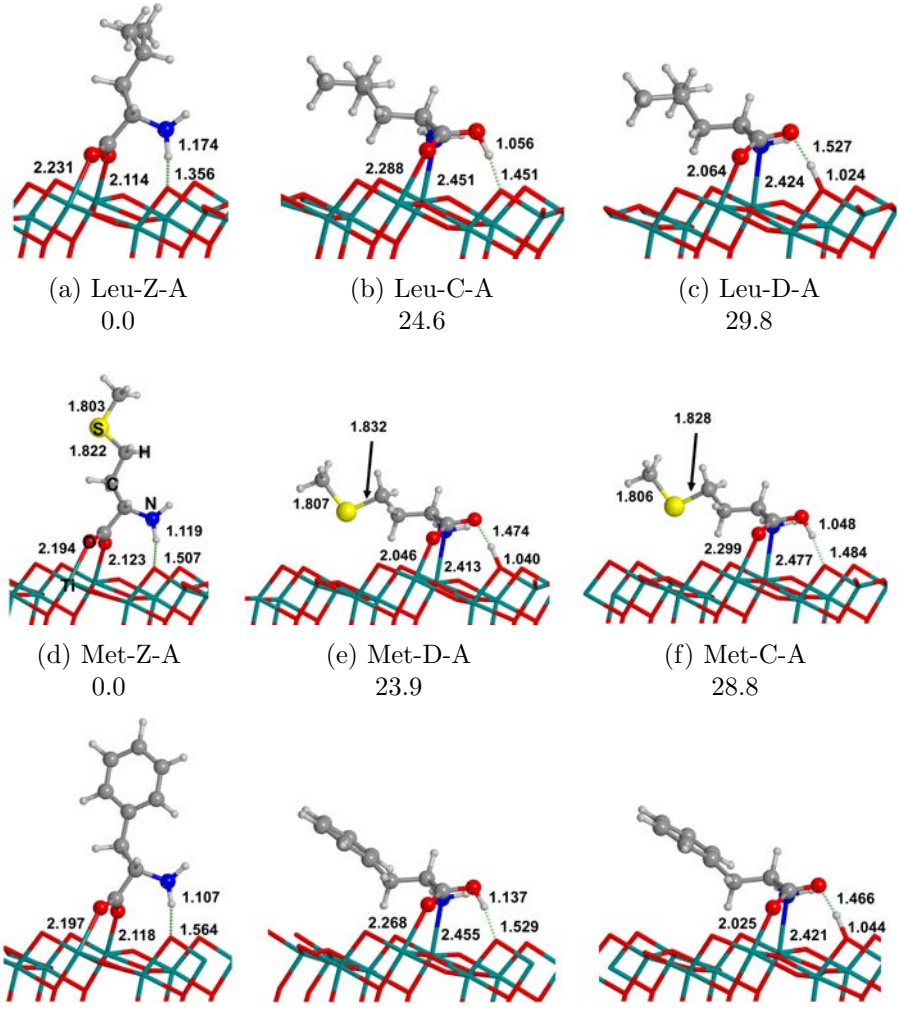
(a) Gly-D1-A (b) Gly-C1-A (c) Gly-Z1-A

Figure 4.2: Evolution of the O(surface)-H(glycine) bond length for Gly-D1-A, Gly-C1-A, Gly-Z1-A complexes in (NVT)-AIMD simulations at 300 K.

structure. For Gly-Z1-A, Figure 4.2b reveals the presence of another deprotonated structure, where the proton of the $-NH_3^+$ group is transferred to the surface. Accordingly, we have optimised a snapshot of this AIMD simulation with the proton on the surface. This deprotonated structure is found to be 4.5 kJ mol^{-1} less stable than Gly-Z1-A. However, NVT simulations indicate that the statistical distribution favours the deprotonated structure, thus highlighting the importance of dynamic (*i.e.* thermal and entropic) effects to perform an appropriate conformational exploration of the adsorption of amino acids on TiO_2 surfaces. In view of these results, the low-lying Gly-D1-A, Gly-C1-A, and Gly-Z1-A complexes were chosen as the starting guess structures to study all the other amino acid adsorption. That is, the H atom of the glycine was replaced by the corresponding amino acidic side chains, taking special caution that the resulting amino acids are in the proper enantiomeric form of the proteinaceous amino acids.

4.3.2 Nonpolar Amino Acids: Leucine, Methionine, and Phenylalanine

Figure 4.3 shows the complexes formed by the adsorption of the nonpolar amino acids on the (101) anatase surface, while Table 4.1 reports the calculated adsorption and relative energies. For these amino acids we expect a similar behaviour to that shown by glycine, because the hydrophobic lateral chains do not allow establishing new H-bonds or interactions with Ti, the interaction of the lateral chains being essentially through dispersive forces. In the case of methionine, we can consider that the S lone pairs can interact with Ti; however, S is a very weak Lewis base, and results do indeed indicate that it does not interact with Ti. In all cases, the zwitterionic structure is found to be the most stable one and thus, compared to the glycine case, there is an inversion of stability between the Z and the D forms. This can be attributed to the sterical hindrance induced by the side chain in the deprotonated form, which leads to larger Ti–N distances than those computed for glycine (2.42 Å for Leu, 2.41 Å for Met, 2.42 Å for Phe and 2.34 Å for Gly), thereby inducing a destabilization of the D state. This sterical hindrance is not present in the Z structures because the side chains are far from the surface and accordingly, for these systems, the Z states become more stable. This Ti–N enlargement is also observed in the C states. The relative stability of the D and C states varies depending on the system; *i.e.*, the D form is more stable than the C one for Met, whereas the reverse trend is observed for Phe and Leu. Results indicate that relative stability between these two structures result from a delicate balance between electron donation, steric hindrance, and dispersion interactions.



(a) Leu-Z-A 0.0 (b) Leu-C-A 24.6 (c) Leu-D-A 29.8

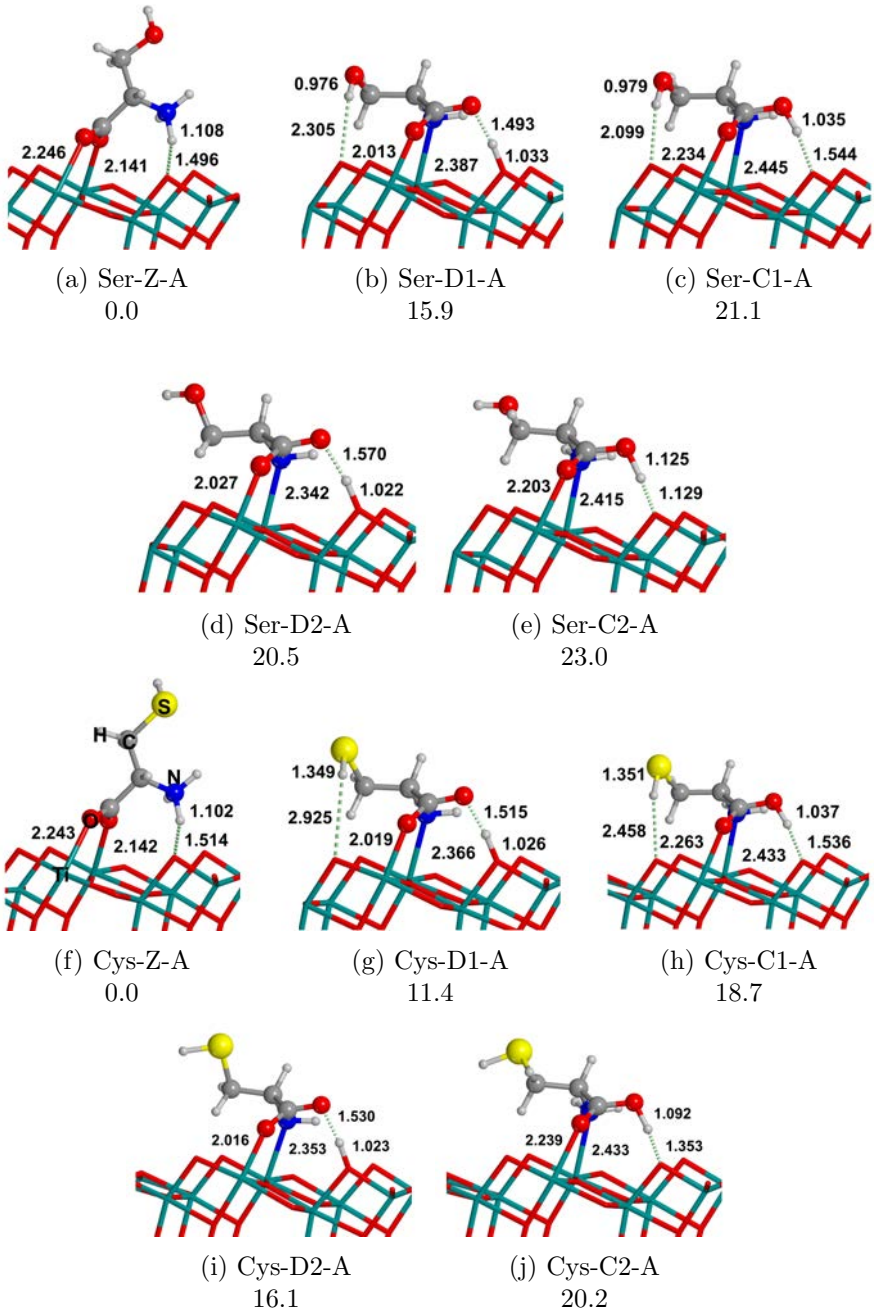
(d) Met-Z-A 0.0 (e) Met-D-A 23.9 (f) Met-C-A 28.8

(g) Phe-Z-A 0.0 (h) Phe-C-A 16.5 (i) Phe-D-A 17.4

Figure 4.3: PBE-D2* optimised geometries for leucine (Leu), methionine (Met), and phenylalanine (Phe) adsorbed on the TiO_2 (101) anatase surface. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

4.3.3 Polar Amino Acids: Serine and Cysteine

These systems exhibit a larger number of adsorption modes because of the additional interaction of the side chains with the surface; *i.e.*, serine and cysteine can lead to the formation of H-bonds between the –OH and –SH groups and the O atom of the surface. Indeed, as one can see in Figure 4.4 and Table 4.1, these H-bonds (present in the "C1" and "D1" complexes) infer an extra stabilization compared to those in which the side chains do not interact with the surface ("C2" and "D2" complexes). Irrespective of that, the most stable structures present the amino acids in the zwitterionic form. As occurred for the non-polar amino acids, the Ti–N bond, in the C and D structures (Table 4.1), increases compared to that in Gly, whereas the Ti–O distances in the Z forms are quite similar. Accordingly, the Z forms are the most stable complexes. Here, the H-bonds arising from the side chains are not strong enough to balance the destabilization caused by the elongation of the Ti–N bonds given in the C1 and D1 forms. As a representative case of those amino acids in which the adsorbed zwitterionic structure is preferred with respect to the deprotonated or canonical, we have also carried out AIMD simulations for serine starting from both the zwitterionic and deprotonated structures. Results, shown in Figure 4.5, indicate that the $–NH_3^+$ group is not so prone to dissociate, as in the Gly-Z1-A case. Indeed, the optimization of a snapshot of the AIMD simulation with the proton of the $–NH_3^+$ group transferred to the surface evolves to the Z structure; *i.e.*, the deprotonated zwitterionic structure is not a minimum of the potential energy surface. By contrast, simulations for Ser-D1-A reveal that this structure interconverts to Ser-C1-A only



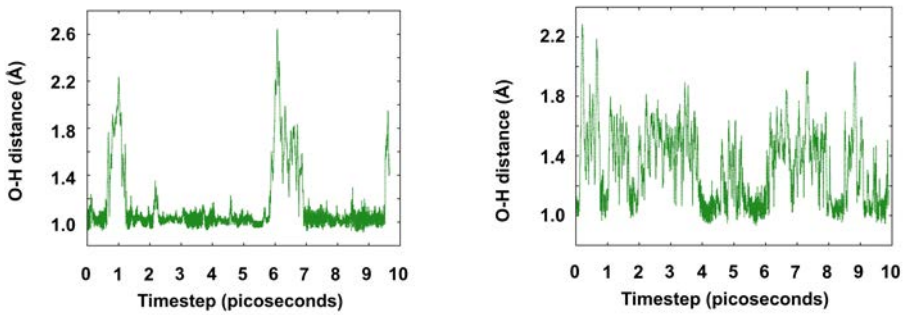
(a) Ser-Z-A 0.0 (b) Ser-D1-A 15.9 (c) Ser-C1-A 21.1

(d) Ser-D2-A 20.5 (e) Ser-C2-A 23.0

(f) Cys-Z-A 0.0 (g) Cys-D1-A 11.4 (h) Cys-C1-A 18.7

(i) Cys-D2-A 16.1 (j) Cys-C2-A 20.2

Figure 4.4: PBE-D2* optimised geometries for serine (Ser) and cysteine (Cys) adsorbed on the TiO_2 (101) anatase surface with the lateral chains interacting with the surface. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.



(a) Ser-D1-A (b) Ser-C1-A

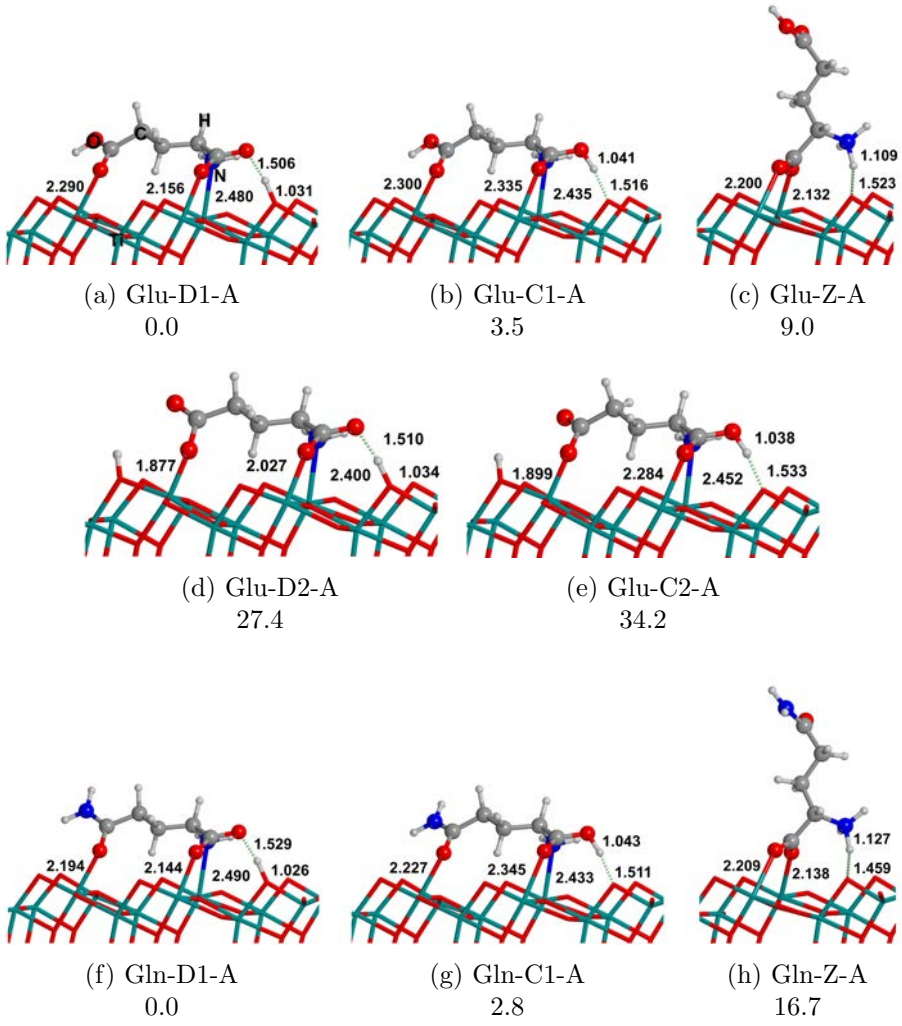
Figure 4.5: Evolution of the O(surface)-H(serine) bond length for Ser-D1-A and Ser-C1-A complexes in (NVT)-AIMD simulations at 300 K.

for few femtoseconds along the entire dynamic.

4.3.4 Polar/Acidic and Polar/Amidic Amino Acids: Glutamic Acid and Glutamine

The most important characteristic of these amino acids is the presence of carboxyl (glutamic acid) and amide (glutamine) functionalities in the side chain; so we expect a stronger interaction between the lateral chains and the surface than those given by serine and cysteine. Figure 4.6 shows the optimised structures, and Table 4.1 shows the adsorption and relative energies. The most stable complexes have the amino acids in the D form, as in the glycine case. As for the previous polar amino acids, deprotonated forms exhibit a larger Ti–N bond distance than that of the Gly-analogue complex, which would render the D forms more unstable. However, the interaction of the side chains is very favourable and strong enough to overcome the

destabilization associated with the weakening of the Ti–N bond. Interestingly, we have also located other complexes similar to Glu-D1-A and Glu-C1-A but with the lateral chain deprotonated due to a H transfer to the surface (Glu-D2-A and Glu-C2-A). However, these two complexes are significantly less stable than Glu-D1-A and Glu-C1-A. This major instability is explained by the lower acidity of the COOH lateral chain group (pKa = 4.25) compared to that of the backbone (pKa = 2.19).



(a) Glu-D1-A
0.0

(b) Glu-C1-A
3.5

(c) Glu-Z-A
9.0

(d) Glu-D2-A
27.4

(e) Glu-C2-A
34.2

(f) Gln-D1-A
0.0

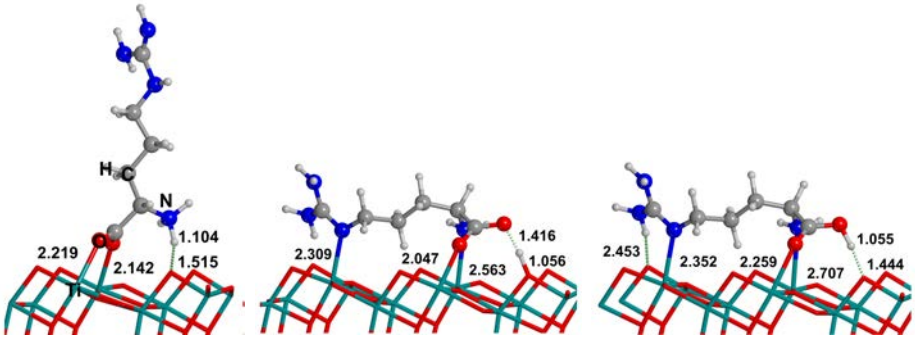
(g) Gln-C1-A
2.8

(h) Gln-Z-A
16.7

Figure 4.6: PBE-D2* optimised geometries for glutamic acid (Glu) and glutamine (Gln) adsorbed on the TiO_2 (101) anatase surface. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

4.3.5 Polar/Basic Amino Acids: Lysine, Arginine, and Histidine

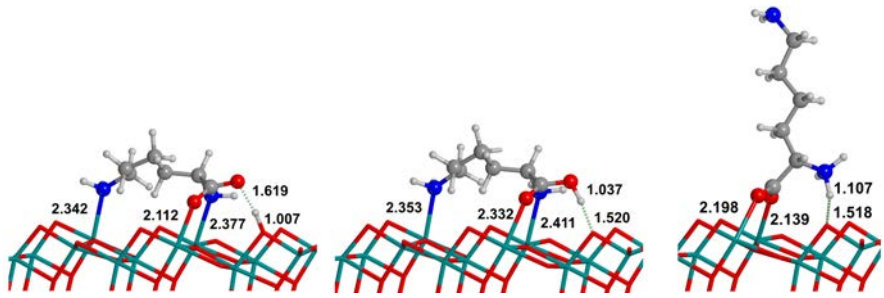
Finally, the adsorption of three amino acids with basic lateral chains; *i.e.*, with the guanidinium, amino, and imidazole groups, has been studied (Figure 4.7 and 4.8). These groups have lone pairs on the N atoms that can strongly interact with Ti atoms of the surface. In the particular case of arginine, the side chain exhibits two tautomeric forms: the -$(CH_2)_3$-NH-C(=NH)NH_2 (T1) and the -$(CH_2)_3$-N=C$(NH_2)_2$ (T2). In the gas phase, the T2 tautomer is more stable, which moreover is more relevant for the present work because the central N is more basic and, accordingly, the interaction with Ti is expectedly stronger. From the calculated ΔE^C_{ADS} and ΔE_{rel} values (Table 4.1) one can observe how the adsorption of these tautomeric forms present different stabilities. Indeed, the C1 and D1 forms (involving the T2 tautomer) present Ti–N distances significantly shorter than C2 and D2 (involving the T1 tautomer) and hence the former structures are about 40 kJ mol^{-1} more stable than the latter ones. For lysine, structures similar to those found for the serine and cysteine cases were investigated, *i.e.*, structures in which the lateral chain interacts (D1, C1, Z2) or not (D2, C2, Z1) with a Ti atom. For D1 and C1, there is a very high stabilization due to this interaction. This is not the case for Z2, as the deformation of the amino acid hinders the system from having a net gain in energy. For histidine, only structures in which Nδ is deprotonated were taken into account, because Nϵ deprotonated complexes do not present interactions with Ti atoms. As mentioned above, the interactions of Ti with N and/or O are fundamental toward the stabilization of the complexes. Remark-



(a) Arg-Z-A
0.0

(b) Arg-D1-A
2.6

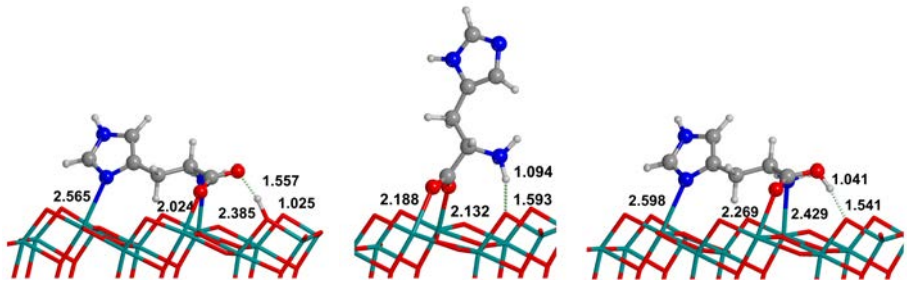
(c) Arg-C1-A
5.2



(d) Lys-D1-A
0.0

(e) Lys-C1-A
10.2

(f) Lys-Z1-A
24.6

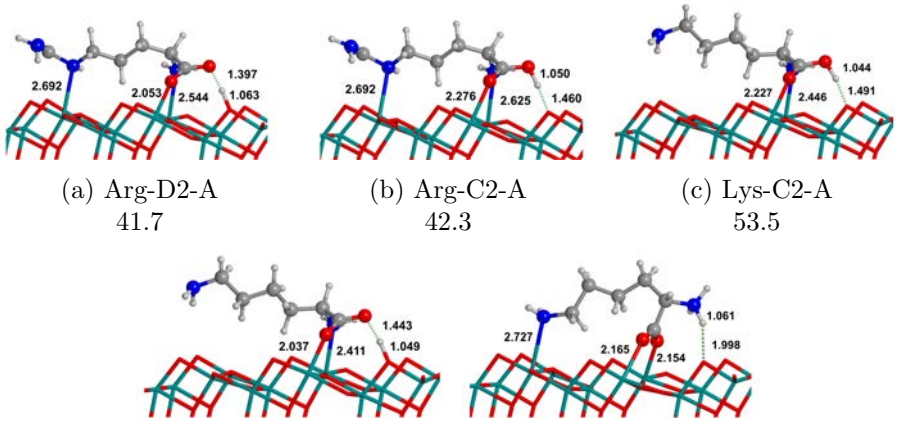


(g) His-D-A
0.0

(h) His-Z-A
5.8

(i) His-C-A
6.0

Figure 4.7: PBE-D2* optimised geometries for arginine (Arg), lysine (Lys), and histidine (His) adsorbed on the TiO_2 (101) anatase surface. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.



(a) Arg-D2-A 41.7 (b) Arg-C2-A 42.3 (c) Lys-C2-A 53.5

(d) Lys-D2-A 55.2 (e) Lys-Z2-A 107.6

Figure 4.8: PBE-D2* optimised geometries for less stable arginine (Arg) and lysine (Lys) adsorbed on the TiO_2 (101) anatase surface. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

ably, for the basic amino acids considered, the preferred adsorption mode involves the deprotonated form of the amino acid, except for the Arg case, where the zwitterionic state is the most stable adsorption structure, plus a Ti–N binding interaction with the side chain.

4.3.6 Trends

It is interesting to correlate the calculated adsorption energies of the different amino acids with their properties. For comparison and to analyse the role of the side chain, the Gly-D1-A and Gly-Z1-A complexes (ΔE^{C}_{ADS} = -108.2 and -100.4 kJ mol^{-1}, respectively) are taken as reference systems. Figure 4.9 shows the adsorption energies of all amino acids considering the most stable form.

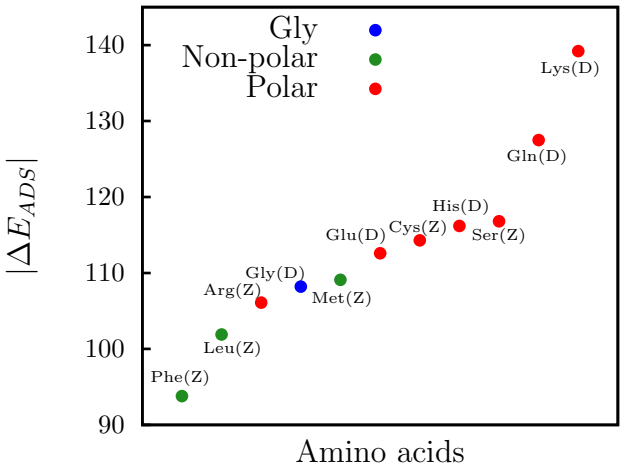


Figure 4.9: Calculated adsorption energies in absolute values, $\left|\Delta E^{C}_{ADS}\right|$ (in kJ mol^{-1}) for the most stable complexes. On the x-axis amino acids are specified in order of increasing $\left|\Delta E^{C}_{ADS}\right|$.

First of all, one can observe that adsorption energies range from -93.8 kJ mol^{-1} (Phe-Z-A) to -139.2 kJ mol^{-1} (Lys-D1). The most favourable values correspond to polar amino acids with strong side chain interactions, whereas the less favourable ones correspond to non-polar amino acids. Lys is the one that exhibits the largest adsorption energy, in absolute value, because of the additional very favourable Ti–N covalent interaction established by the lateral chain with the surface. Non-polar amino acids, plus Ser and Cys, with weak side chain interactions with the surface, and Arg, with a strong side chain interaction with the surface, present the zwitterionic form as the most stable adsorbed complex. This contrasts to Gly and all other amino acids, for which the preferred adsorption mode is the (N,O) deprotonated configuration. This is due to the fact that ster-

ical hindrances introduced by the side chain induce an enlargement of the Ti–NH_2 and Ti–CO_2^- bond distances in the D form, thereby inducing a destabilization of this state. This destabilization does not occur in the Z structures because coordination with the two Ti atoms through the –CO_2^- group forces the side chain to remain far from the surface, so that it becomes the most stable configuration. Arg is the amino acid that exhibits a larger Ti–NH_2 distance and thus destabilization is not compensated by the strong interaction observed between the side chain and the surface.

For those amino acids in which the most stable adsorbed complex has the amino acid in its zwitterionic form, no clear correlation between the calculated adsorption energies with coordination bond distances was found to explain the more/less favourable adsorption energies compared to Gly-Z1-A. With respect to the relative stability between the canonical and deprotonated forms, all amino acids except Leu and Phe seem to prefer the deprotonated structure, the relative energy ranging from 3 to 10 kJ mol^{-1}. Phe and Leu exhibit a different behavior, which results from a delicate balance between side chain and backbone interactions. Note, however, that relative energies between these two structures are very small (1-5 kJ mol^{-1}). Finally, it is worth mentioning that, in contrast to the adsorption on silica where adsorption energies correlate reasonably well with hydropathy index and hydration potential,[100] no clear correlation has been found for the adsorption onto TiO_2. This is due to the fact that adsorption on silica occurs through the interaction with silanol groups, which somewhat mimic the behaviour of water, whereas, in the present case adsorption mainly occurs through dative interactions.

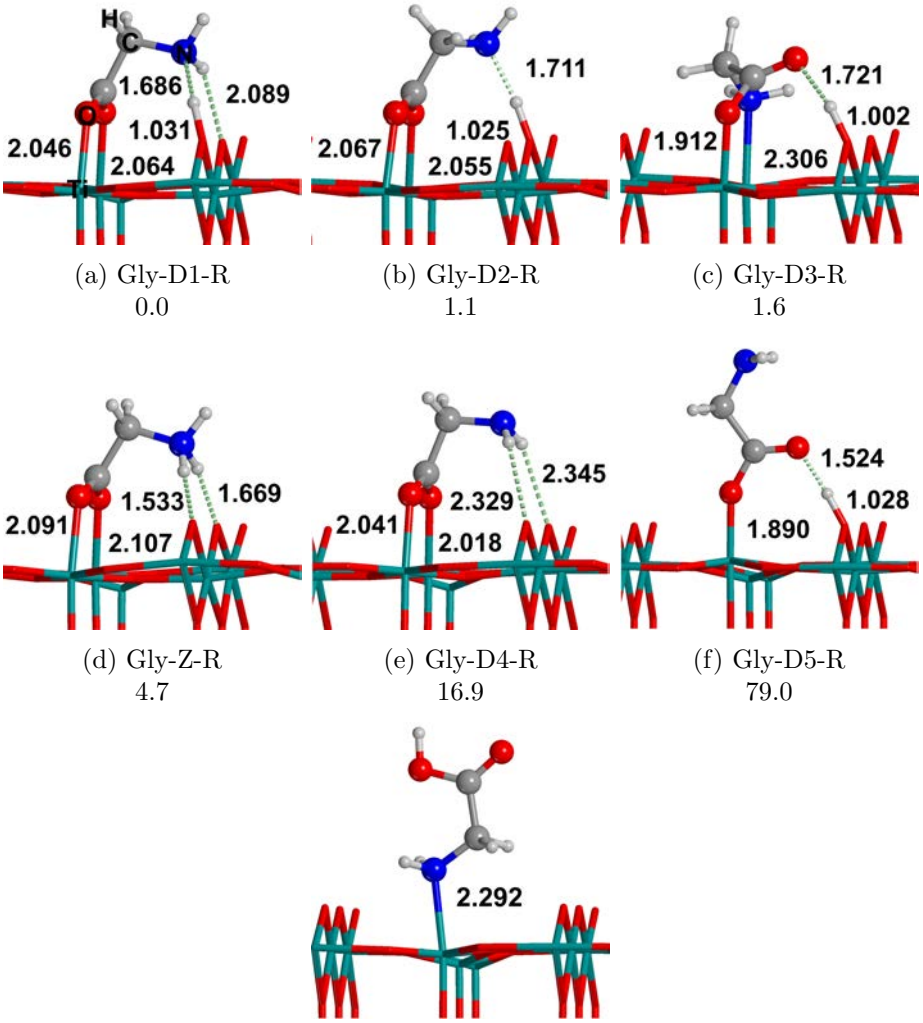
4.4 Amino acids adsorpion on the (110) TiO_2 rutile surface

As for the anatase surface, here we have studied the adsorption of the same amino acids on the TiO_2 (110) rutile surface, in their canonical (C), deprotonated (D) and zwitterionic (Z) forms.

4.4.1 Glycine

In Figure 4.10 different complexes for the Gly/TiO_2 adduct are shown and in Table 4.2 the relative and adsorption energies are reported. The most stable structure is a deprotonated glycine (Gly-D1-R), followed by other two deprotonated structures, Gly-D2-R and Gly-D3-R. All the most stable structures (Gly-D1-R, Gly-D2-R, Gly-D3-R, Gly-Z-R, and Gly-D4-R) present two strong anchoring points, the two O atoms or one O and the N atom, and then some H-bonds. In contrast, Gly-D5-R and Gly-C-R have only one anchoring point and only one or none H-bonds and, accordingly, they are less stable (about 80-100 kJ mol^{-1}).It should be noticed that Gly-D3-R and Gly-D5-R complexes, were obtained from the corresponding "C" forms. During the optimization process the proton jumps to the surface, which is indicative that amino acids are very prone to dissociate on the rutile surface. The only canonical structure obtained is the Gly-C-R complex which is much more unstable than the other ones, because it has only one anchoring point. Another important feature of the (110) rutile surface is the high adsorption energy for the analysed complexes, as reported in Table 4.2 (ΔE^C_{ADS} = -238.3 kJ mol-1 for Gly-D1-R and ΔE^C_{ADS} = -354.2 kJ mol^{-1} for Arg-D-R).

Frequency calculations performed on the low lying structures, *i.e.* Gly-D1-R, Gly-D2-R, Gly-D3-R and Gly-Z-R, reveal an inversion of stability due to thermal correction: Gly-D2-R became more stable than Gly-D1-R by about -0.5 kJ mol^{-1}. However the relative energies are below the chemical accuracy of quantum mechanical calculations (1 kcal mol^{-1}), so we cannot say with certainty which of them is the most stable structure. Furthermore, we have carried out a MD simulation on the Gly-D1-R in order to study its equilibrium between the deprotonated and the zwitterionic forms: Gly-D1-R and Gly-Z-R are basically the same structure with the only difference of one displaced proton. In Figure 4.11 the MD simulation on the Gly-D1-R complex following the two marked H-bonds is shown. As one can see, since the first femtoseconds in the equilibration period, the H atom gets away from the surface. Between 7 and 9 picoseconds both distances drop down to ∼1.0 Å, *i.e.* the surface protonates the NH_2 group and then the formed NH_3 group deprotonates on another O site of the surface in a structure similar to Gly-D2-R. The relative short period of the MD simulation is sufficient to explore three different structures confirming that these complexes are very accessible because of their close relative energies. To analyse the other amino acids we have chosen Gly-D1-R as starting guess, because it is present for the majority of the MD simulation (in the final part of the MD Gly-D2-R complex converts to Gly-D1-R one more time). Gly-Z-R and Gly-C-R have been used as starting guess structures to build up the zwitterionic and canonical forms, respectively, for the other amino acids.



(a) Gly-D1-R
0.0

(b) Gly-D2-R
1.1

(c) Gly-D3-R
1.6

(d) Gly-Z-R
4.7

(e) Gly-D4-R
16.9

(f) Gly-D5-R
79.0

(g) Gly-C-R
106.8

Figure 4.10: PBE-D2* optimised geometries for glycine (Gly) adsorbed on the TiO_2 (110) rutile surface. "C" refers to Gly in its canonical form, "D" in its deprotonated form, and "Z" in its zwitterionic form. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

Table 4.2: Calculated energetic and geometric data for all the amino acids analysed.

Complex	ΔE^C_{ADS}	ΔE_{rel}	Ti–O	Ti–O^a/ Ti–N^b	Ti–X^{a,b,c}
Gly-D1-R	-238.3	0.0	2.05	2.06^a	–
Gly-D2-R	-237.2	1.1	2.07	2.05^a	–
Gly-D3-R	-236.7	1.6	1.91	2.31^b	–
Gly-Z-R	-233.6	4.7	2.09	2.11^a	–
Gly-D4-R	-221.5	16.9	2.04	2.02^a	–
Gly-D5-R	-159.4	79.0	1.89	–	–
Gly-C-R	-127.8	106.8	–	2.29^b	–
Leu-Z-R	-236.9	0.0	2.09	2.07^a	–
Leu-D-R	-231.1	5.8	2.03	2.03^a	–
Leu-C-R	-154.2	83.6	–	2.28^b	–
Met-D-R	-267.7	0.0	2.04	2.00^a	2.87^c
Met-Z-R	-267.4	0.3	2.18	2.03^a	2.89^c
Met-C-R	-190.4	80.7	–	2.31^b	2.85^c
Phe-Z-R	-250.0	0.0	2.04	2.04^a	–
Phe-D-R	-247.7	2.3	2.08	2.09^a	–
Phe-C-R	-148.1	104.1	–	2.30^b	–
Ser-D-R	-235.9	0.0	2.05	2.09^a	–
Ser-Z-R	-234.8	1.2	2.10	2.12^a	–
Ser-C-R	-129.4	105.2	–	2.27^b	–
Cys-D-R	-230.6	0.0	2.03	2.07^a	–
Cys-Z-R	-228.1	2.5	2.08	2.11^a	–
Cys-C-R	-155.2	73.2	–	2.30^b	2.76^c
Glu-Z-R	-258.6	0.0	2.04	2.03^a	–
Glu-D-R	-256.1	2.6	2.10	2.05^a	–
Glu-C-R	-180.6	76.9	–	2.34^b	2.10^a
Gln-D-R	-267.2	0.0	1.96	2.13^a	2.11^a
Gln-Z-R	-256.2	11.0	2.00	2.25^a	2.06^a
Gln-C-R	-195.2	69.8	–	2.33^b	2.06^a
Arg-D-R	-354.2	0.0	2.08	2.13^a	2.07^b
Arg-Z-R	-353.0	1.2	2.13	2.15^a	2.08^b
Arg-C-R	-208.7	142.6	–	2.34^b	2.46^b
Lys-Z-R	-305.1	0.0	2.12	2.07^a	2.26^b
Lys-D-R	-302.8	2.3	2.07	2.05^a	2.33^b
Lys-C-R	-191.7	115.7	–	2.35^b	2.33^b
His-D-R	-284.8	0.0	2.15	1.98^a	2.29^b
His-Z-R	-279.0	5.8	2.17	2.05^a	2.31^b
His-C-R	-224.4	56.5	–	2.29^b	2.32^b

ΔE^C_{ADS} is the adsorption energy of the complex calculated with respect to isolated amino acid and surface, and corrected for the lateral interaction; ΔE_{rel} is the relative energy respect to the most stable structure of each amino acid. In the forth and fifth columns the Ti–O and Ti–O/Ti–N bond lengths of the backbone are reported, respectively. Ti–X represents the interaction of the side chain with the surface; Ti–X^a and Ti–X^b are Ti–O and Ti–N, respectively, while Ti–X^c is Ti–S bond. The energy values are in kJ mol^{-1} and the bond distances are in Å.

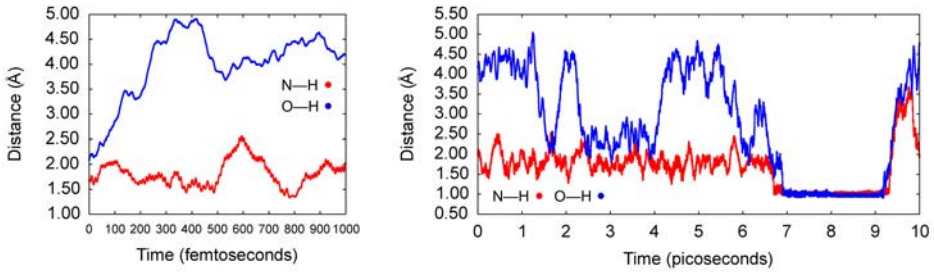
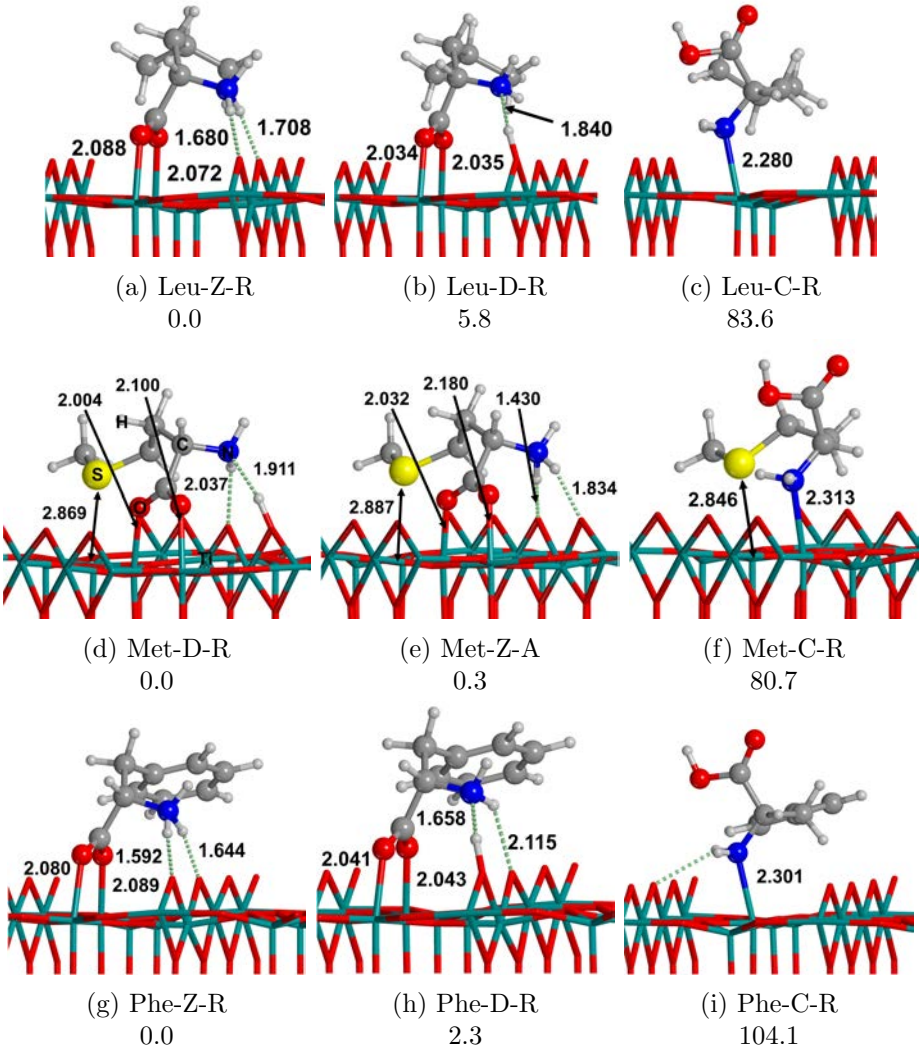


Figure 4.11: Evolution of N–H (1.686 Å) and O–H (2.089 Å) H-bonds for the Gly-D1-R complex. Equilibration period (left) in the NVE ensamble at fixed temperature (300 K) and production period (right) in the NVT ensamble (300 K).

4.4.2 Non-polar amino acids: Leucine, Methionine and Phenylalanine

In Figure 4.12 the complexes of non-polar amino acidic residues are shown. The interaction of the lateral chains with the surface for these amino acids occurs essentially through dispersive forces, particularly for Leu and Phe. In the case of Met, S atom is a weak Lewis base and so its interaction with Ti atoms is much weaker than that of the amino or carboxylic groups; however, as one can see, despite the long Ti–S distance, one of the lone pars of the S atom points towards a Ti atom of the surface for all the structures (with Ti–S distances of about 2.8-2.9 Å). As a result the Met presents the highest adsorption energies among the non-polar amino acids (see Table 4.2). The equilibrium between the Z and D forms is delicate and it is very difficult to identify which of them is more stable (the difference in energy is below the chemical accuracy). However, the difference between the two forms in Leu is slightly larger than the

other two amino acids. This could be explained by the higher pKa value of its NH_3^+ group, which is 0.4/0.5 larger than that of Met and Phe. The canonical structures are disfavoured because of their less contact points with the surface compared to the Z and D forms. The difference in the relative energies between Met and Phe are due to the interaction of the S atom with the surface. The major stability of the Leu-C-R with respect to the Phe-C-R is due to a delicate balance between dispersion forces and the sterical hindrance of the lateral chains; despite the major dispersion interactions of an aromatic ring with respect to a iso-butil group, the major size of the ring leads to a sterical repulsion with bridge O atoms of the surface.



(a) Leu-Z-R 0.0 (b) Leu-D-R 5.8 (c) Leu-C-R 83.6

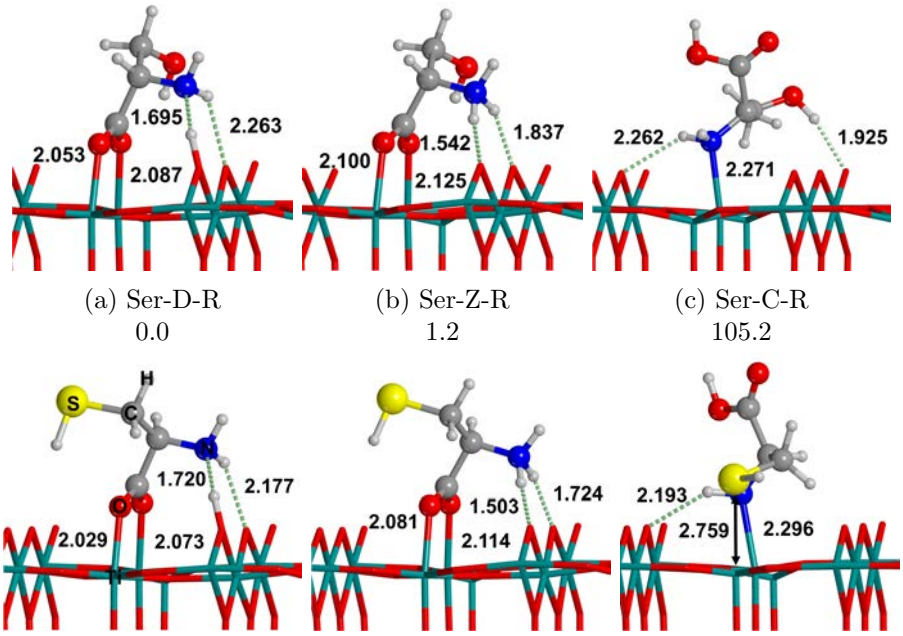
(d) Met-D-R 0.0 (e) Met-Z-A 0.3 (f) Met-C-R 80.7

(g) Phe-Z-R 0.0 (h) Phe-D-R 2.3 (i) Phe-C-R 104.1

Figure 4.12: PBE-D2* optimised geometries for leucine (Leu), methionine (Met), and phenylalanine (Phe) adsorbed on the TiO_2 (110) rutile surface. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

4.4.3 Polar amino acids: Serine and Cysteine

In this case both Ser and Cys present the most stable complex in the deprotonated form (see Figure 4.13). The most stable forms, *i.e.* the deprotonated and the zwitterionic, do not present any additional favourable interaction caused by the H-bonds between the side chains and the surface, because the two partners are very far from each other. Indeed, their adsorption energies are smaller than the previous amino acids studied, including Gly (see Table 4.2). By contrary, Ser-C-R presents a H-bond between the side chain and the surface, while Cys-C-R is anchored to the surface via a weak (but significant) Ti–S dative bond (0.1 Å shorter than Met).



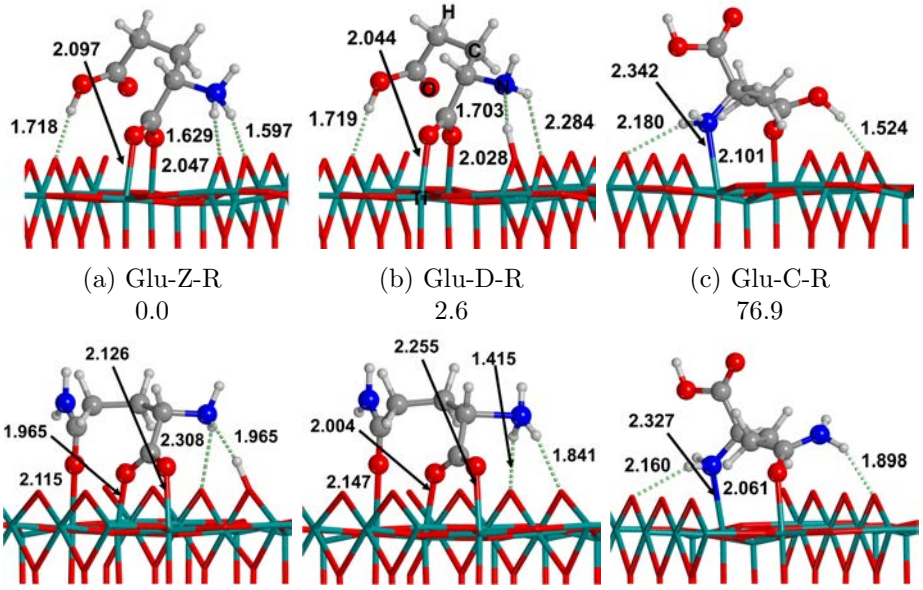
(a) Ser-D-R 0.0 (b) Ser-Z-R 1.2 (c) Ser-C-R 105.2

(d) Cys-D-R 0.0 (e) Cys-Z-R 2.5 (f) Cys-C-R 73.2

Figure 4.13: PBE-D2* optimised geometries for serine (Ser) and cysteine (Cys) adsorbed on the TiO_2 (110) rutile. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

4.4.4 Polar/acidic and polar/amidic amino acids: Glutamic Acid and Glutamine

These amino acids present acidic (Glu) and amidic (Gln) lateral chains, with carbonyl O atoms that can strongly interact with the surface with dative bonds, as it happens for the backbone. As one can see by the structures reported in Figure 4.14, Glu-Z-R and Glu-D-R do not present strong interactions between the lateral chain and the surface, but only one H-bond between the lateral chain proton and one O atom of the surface. Structures with the O of the lateral chain bound to a Ti atom exist, but they are less stable than the ones reported; this is possibly due to the relative short lateral chain of the Glu (just two CH_2 groups) which inhibits a proper orientation for a stronger interaction. The amidic oxygen of Gln is more prone to interact with a Ti atom because of the presence of a resonance form (responsible of the planar structure of the amide bond) with a formal negative charge on the O atom, which overcome the sterical hindrance and the strain of the lateral chain. This is confirmed by the canonical structures, where one can observe a short decrease of the Ti–O bond distance of Gln-C-R with respect to Glu-C-R, and a more stable relative energy. Because of this extra-stabilization we have found canonical structures not very high in energy with respect to the deprotonated or zwitterionic ones. Also in this case we can explain the inversion of stability of the zwitterionic form with the deprotonated one with the smaller pKa values of the Glu compared to the Gln: 9.67 and 9.13, respectively.



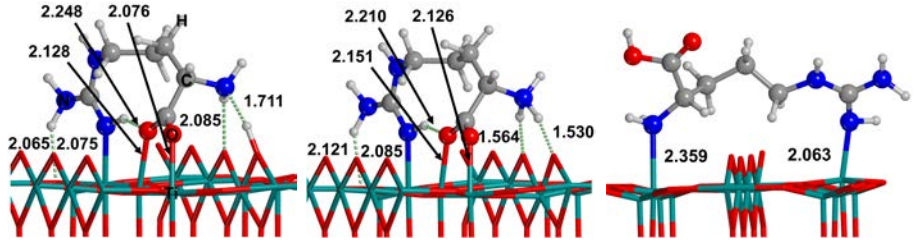
(a) Glu-Z-R 0.0 (b) Glu-D-R 2.6 (c) Glu-C-R 76.9

(d) Gln-D-R 0.0 (e) Gln-Z-R 11.0 (f) Gln-C-R 69.8

Figure 4.14: PBE-D2* optimised geometries for glutamic acid (Glu) and glutamine (Gln) adsorbed on the TiO_2 (110) rutile surface. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

4.4.5 Polar/basic amino acids: Lysine, Arginine and Histidine

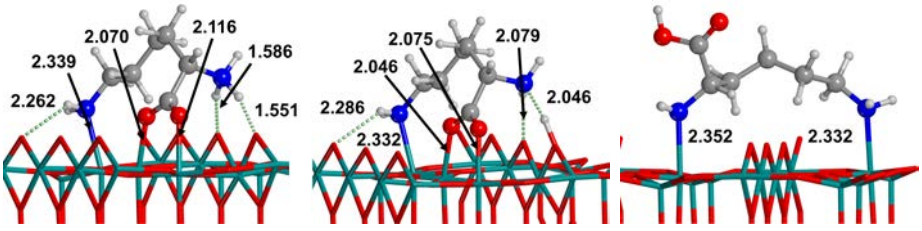
Finally we have studied those amino acids with basic functionalities in the lateral chain, like Arg, Lys and His (see Figure 4.15). These amino acids are expected to interact with the surface even more strongly than Glu or Gln; indeed, as one can see from Table 4.2, very high adsorption energies are reported, in particular for Arg (ΔE^C_{ADS} = -354.2 kJ mol^{-1}). The higher adsorption energies respect to Glu and Gln are due to the fact that nitrogen is more basic (or a stronger Lewis base) than oxygen, and, accordingly, more prone to donate its electron pair to a Lewis acid (Ti atoms of the surface). However, in organic chemistry nitrogen presents several functionalities: here we have chosen to study amino acids with different basic functionalities, *i.e.* the guanidinium group (Arg), the amino group (Lys), and the imidazole group (His). The pKa of the conjugated acids of these groups are reported to be ∼12.5, 10.5, and 6.0 for Arg, Lys, and His, respectively. The strong basicity of the guanidinium group is due to the presence of an imidic group, in addition to a primary and a secondary amine groups. It is well known that imines are more basic than amines; this reflects on Ti–N distances, ∼2.0/2.1 for the lateral chain of the arginine, and ∼2.3 for the lateral chain of Lys and for the backbone. His has a different feature, because the N atom attached on a Ti atom of the surface is formally part of an imidic group, but the aromaticity of the imidazole spreads on the whole ring the electron density, lowering the basicity of the lateral chain of His. Previous consideration are confirmed by the adsorption energies of these complexes, ΔE^C_{ADS} = -354.2 kJ mol^{-1}, ΔE^C_{ADS} = -305.1 kJ



(a) Arg-D-R
0.0

(b) Arg-Z-R
1.2

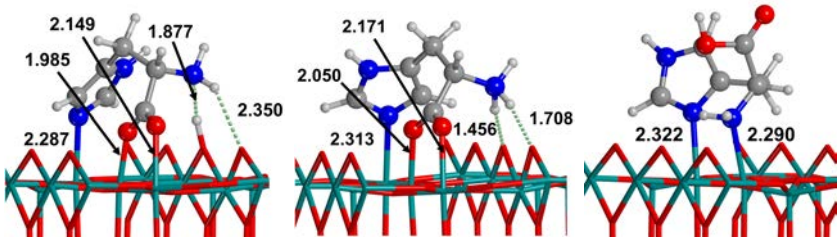
(c) Arg-C-R
142.6



(d) Lys-Z-R
0.0

(e) Lys-D-R
2.3

(f) Lys-C-R
115.7



(g) His-D-R
0.0

(h) His-Z-R
5.8

(i) His-C-R
56.5

Figure 4.15: PBE-D2* optimised geometries for arginine (Arg), lysine (Lys), and histidine (His) adsorbed on the TiO_2 (110) rutile surface. Bond distances are in Å. Relative energy values are in kJ mol^{-1}.

mol^{-1}, ΔE^{C}_{ADS} and ΔE^{C}_{ADS} = -284.8 kJ mol^{-1} for Arg, Lys, and His, respectively.

4.5 Final remarks

Comparison between the (101) anatase surfaces and the (110) rutile

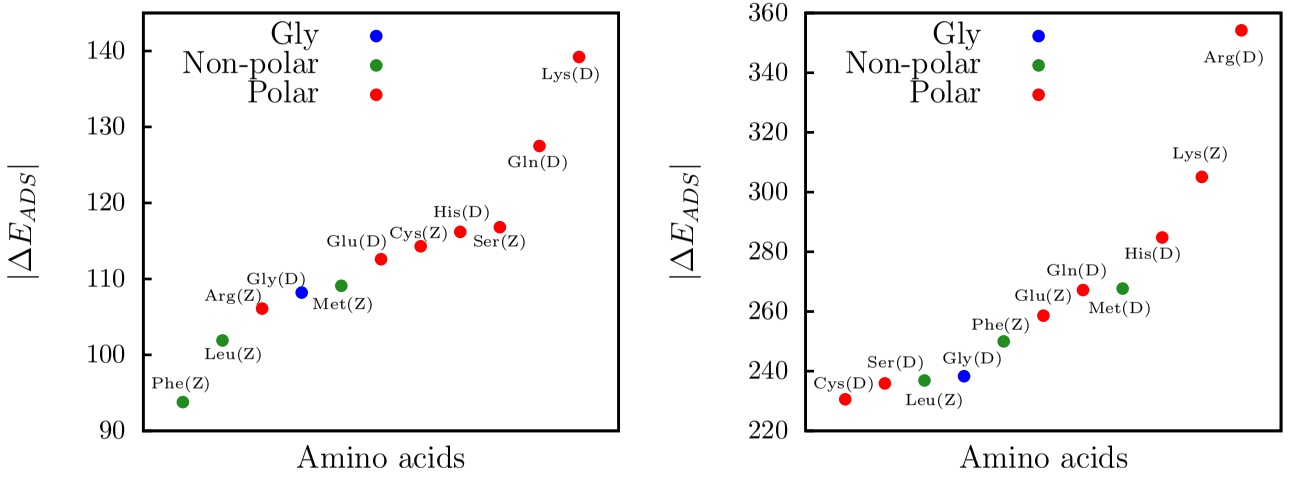
From a morphologic point of view the (110) rutile surface is flatter than the more corrugated shape of the (101) anatase surface. Moreover, considering the pure surfaces, the (110) rutile surface presents a slower convergence of the surface formation energy with the slab thickness in comparison to the anatase one; this is possibly due to the higher surface formation energy of the (110) rutile (E_s = 0.80 J m^{-2} for the 5-layer) with respect to the (101) anatase (E_s = 0.37 J m^{-2} for the 3-layer).

The first difference of the AA/Surface systems between the (110) rutile and the (101) anatase TiO_2 surfaces is the higher tendency of the rutile in deprotonating the amino acids. As one can see from Figure 4.10, only one canonical structure has been found. As mentioned, the corresponding canonical structure of Gly-D3-R and Gly-D5-R were attempted but during the optimization process, the amino acid deprotonates forming Gly-D3-R and Gly-D5-R. The only way to obtain a canonical structure is the adsorption via only the amino group, the acidic group being far from the surface. Obviously this structure presents a lower adsorption energies for all the amino acids analysed (80-100 kJ mol^{-1} less than the corresponding deprotonated and zwitterionic forms) due to the single N-anchoring point of the

backbone. This is a direct consequence of the higher surface formation energy for the rutile surface, which, accordingly, is more reactive. Another consequence of the major reactivity of the (110) rutile with respect to the (101) anatase is its higher adsorption energy for all the analysed complexes. From Table 4.2 one can see that the adsorption energies are almost twice and, in some cases, thrice (e.g. Arg). This property correlates reasonably well with the E_s of the two surfaces, for rutile being twice with respect to anatase.

Another important difference, probably due to the surface morphology, is the preference, in the anatase case, for the (N,O) binding mode. On rutile surface the most stable adsorption is through both the O atoms of the carboxylic group (O,O binding). The Ti–Ti distance on the anatase surface is ∼3.8 Å, and the N...O distance of a glycine molecule is ∼2.9 Å; on rutile the Ti–Ti distance is ∼3.0 and the O...O distance in glycine is ∼2.3 Å. Accordingly, the (N,O) and (O,O) binding modes requires a modest distortion of the amino acid when it is adsorbed on the anatase and rutile surface, respectively.

Finally, as one can observe in Figure 4.16 polar amino acids generally show a higher adsorption energy with respect to the non-polar ones, in particular those which can establish additional dative bonds between the lateral chain (Arg, Lys) and the surface. In some cases this trend is not observed due to either the morphology of the surface (no Ti atoms close to the lateral chain) or the length of the lateral chain (see Glu on rutile, Figure 4.14, and Ser and Cys on rutile, Figure 4.13)



(a) Anatase

(b) Rutile

Figure 4.16: Calculated adsorption energies in absolute values, $\left|\Delta E^{C}_{ADS}\right|$ (in kJ mol^{-1}) for the most stable complexes both adsorbed on (101) anatase and on (110) rutile surfaces. On the x-axis amino acids are specified in order of increasing $\left|\Delta E^{C}_{ADS}\right|$.

Chapter 5

Peptide bond formation mechanism

In the present Chapter we will present the results on the mechanism of the peptide bond formation between two glycine (Gly) molecules by means of periodic simulations on the TiO_2 (101) anatase surface. The efficiency of the surface catalytic sites is demonstrated by comparing the reactions in gas phase and on the surface. Moreover, the effect of molecules capable to act as proton transfer assistants (*i.e.*, H_2O and Gly as well) has also been considered, showing a significant energy barrier decrease. The reaction on the surface is also favourable from a thermodynamic standpoint, providing very large and negative reaction energies. This is due to the fact that the anatase surface also acts as a dehydration agent since the outermost coordinatively unsaturated Ti atoms strongly anchor the released water molecules.

5.1 Introduction

Since the first experiments on prebiotic chemistry, during the fifties/sixties of the last century, there are still several open questions that have not been unambiguously answered. The experiments carried out by Miller,[13, 14] Oró[16] and Butlerov[18] demonstrated the prebiotic formation of important building blocks such as amino acids, nucleic acids, and sugars, respectively.

The next step after the synthesis of the biomolecular building blocks is the formation of the corresponding biopolymers, which are essential macromolecules for life. This is the case, for instance, the synthesis of peptides by polymerization/condensation of amino acids through the peptide bond formation, or the linkage of ribose/deoxyribose with nitrogenous bases and phosphates, through N-glycosidic and phosphoester bonds, respectively, to form a nucleotide monomer, which in turn polymerize via phosphodiester bonds to form a nucleotide strand.[271–274]

In addition to this prebiotic interest, the reaction of the peptide (and amide) bond formation is also of interest for industrial purposes. Indeed, formation of amide bonds by condensation of non-activated amines and carboxylic acids catalysed by nanostructured mineral oxide surfaces holds potentialities as a sustainable route for the industrial production of amides.[275, 276] Nowadays, the amide formation reaction is a key process for pharmaceutical companies, in which in the current synthetic routes the adoption of powerful activating agents is routinely used. Unfortunately, these routes are expensive and environmentally unfriendly, with toxic/corrosive by-products and large quantities of waste. Thus, developing clean and low cost syn-

thetic strategies with good atom economy is highly pursued.[277]

The present Chapter focuses on the problem of the amino acid polymerization, and in particular deals with the condensation between two glycine molecules, which is accompanied by water elimination, as the simplest test case for the peptide bond formation reaction. As the reaction is kinetically hampered in gas phase because of the high activation energy barrier about (45 - 55 kcal mol^{-1}) and thermodynamically unfavourable in water solution since it releases water, the effect of a mineral surface as catalyst has been considered. In particular, we will provide a comprehensive atomistic interpretation of the experiment carried out by Prof. G. Matra and co-workers, where they observed the formation of glycine oligomers by successive feeding of monomer vapours onto several mineral surfaces (hydroxyapatite, silica, and titania).[25] In Figure 5.1 the IR and mass spectra are reported. As one can see both the spectra are similar for SiO_2 and TiO_2, where the polymerization occurs, while they are different from that of HA, where only the monomer is present. These are clear proofs of the efficiency of SiO_2 and particularly TiO_2 in Gly polymerization.

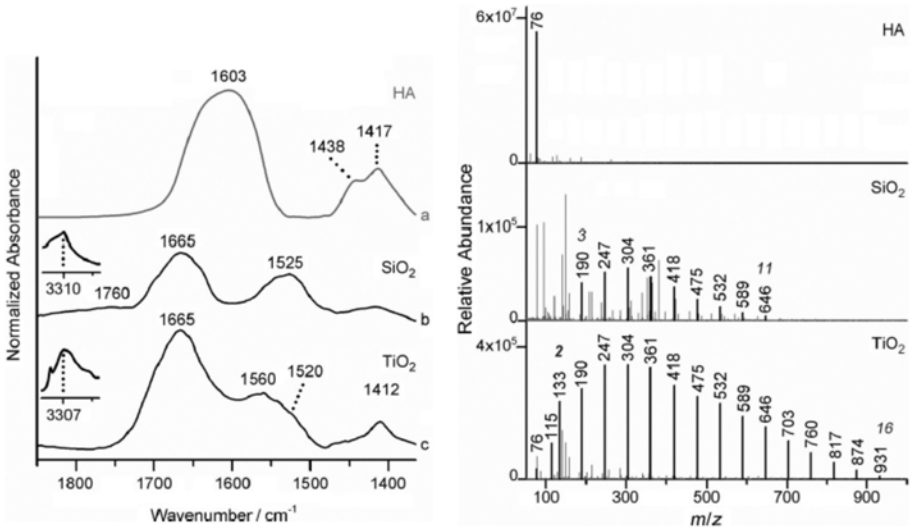


Figure 5.1: IR (left) and mass (right) spectra of the adsorption of Gly monomers onto Hdroxiapatite (HA), silica (SiO_2) and titania (TiO_2). This Figure has been taken from ref. [25]

5.2 Computational details

5.2.1 Methods

Geometry optimizations and frequency calculations were performed with the PBE-D2* method; however, the reaction energetics were refined by performing single-point energy calculations at PBE0-D2* theory level on the optimised stationary points, as it is well known that hybrid methods provide more accurate results than pure GGA methods, particularly for the calculation of activation barriers.[278–280] The SCF iterative procedure has been converged to $\Delta E = 10^{-6}$ eV while the tolerance on gradients for geometry optimizations has been set to 0.01 eV/Å for each atom in each direction. The k-points mesh was set to (3,3,1) for all the reaction studied.

Vibrational frequencies of a reduced Hessian matrix (*i.e.*, considering only the displacements of the atoms corresponding to the reactant molecules and to the first layer of the surface) were computed, at the Γ point, by numerical differentiation of the analytical first derivatives, using the central difference formula (*i.e.* two displacements for each atom in each direction). Frequency calculations were useful to characterize reactants, products and intermediates as minima of the potential energy surface (no imaginary frequencies) and that transition state structures are saddle points (one imaginary frequency associated with the reaction coordinate). From the calculated frequencies, moreover, we computed free energy values at T = 298.15 K using a home-made script that allows us to calculate quasi-harmonic thermochemical corrections. The Grimme correction[281] was used with a frequency cut-off to 100 cm^{-1}; that is, as for low

vibrational normal modes the vibrational entropy is underestimated by the rigid rotor harmonic oscillator approximation, the free-rotor approximation has been applied to the vibrational entropies below this defined cut-off, and a damping function is used to interpolate the values of the two approaches close to the cut-off frequency, in order to avoid a discontinuity. Free energy values at PBE0-D2* were obtained by introducing the PBE-thermal corrections to the PBE0 electronic energies.

For all the reactant structures, both in gas phase and on the surface, we carried out a Bader charge analysis in order to examine whether the surface infers any electronic structure reorganization to the reactants. In particular, we followed the charge on the C atom where the nucleophilic attack occurs. Calculations were performed with the Bader charge analysis code developed by the Henkelman group.[282–285]

5.2.2 Surface models

The standard 1×1 unit cell for the TiO_2 (101) anatase surface is the same we used for the amino acids adsorption in the previous Chapter , *i.e.* with lattice parameters of $a = 3.7845$ Å, $b = 5.119675$ Å, and an angle of 90 degrees. However, to simulate the reaction without steric clashes but maintaining at the same time a reasonable computational cost, we used a 2×2 supercell for all the condensation reactions except for the reaction in the presence of $2H_2O$ molecules, in which a 3×2 supercell was used. The distance between slab images, regulated by the c value, was defined ensuring an empty space of 10 Å between the outermost atom of the amino acid and the upper layer. This condition

is satisfied by setting the c value to 25 Å, which is enough to avoid mutual fictitious interactions between the periodically repeated slabs.

5.3 Results

The gas-phase —NH–C(=O)– peptide bond formation through the condensation of two glycine (Gly) molecules can be both a concerted or a stepwise reaction; however as the barrier of the two mechanisms are similar,[286] we used as reference for the gas phase reaction the concerted mechanism (more probable, as it presents only one barrier). It involves a nucleophilic attack of the NH_2 group of one Gly to the C atom of the carbonyl CO group of the other Gly. This step is simultaneous to a proton transfer from the attacking NH_2 group to the OH group of the second glycine, thus yielding the elimination of one water molecule (see Figure 5.2).

The reaction has already been studied theoretically in gas phase,[114, 116, 117, 287, 288] adopting either the reaction of NH_3 + HCOOH → NH_2CHO + H_2O as the simplest model reaction, or the reaction between more complex forms such as NH_3 + CH_3COOH NH_2COCH_3) or the actual Gly + Gly → $NH_2CH_2CONHCH_2COOH$ (this latter species hereafter referred to as GlyGly) condensation reaction. These calculations indicated that the reactions present free energy barriers at T = 298 K ($\Delta G^{\ddagger}_{298}$) ranging from 40 to 50 kcal mol^{-1}, depending on the type of reaction and the method employed. The calculated reaction free energies at T = 298 K ($\Delta_r G_{298}$) indicated that these processes are either isoergonic or slightly exoergonic. In this work, the uncatalysed gas-phase reaction has an intrinsic $\Delta G^{\ddagger}_{298}$

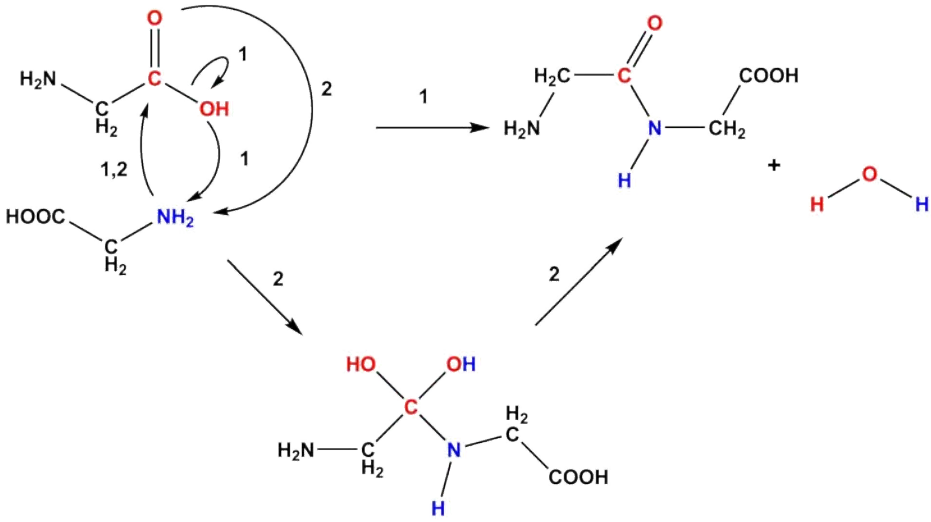


Figure 5.2: Reaction of the peptide bond formation between two glycine molecules adopting a concerted (1) or a stepwise (2) mechanism. Atoms in red and blue are those actually involved in the condensation reaction.

= 41.8 kcal mol^{-1} and $\Delta_r G_{298}$ = -7.3 kcal mol^{-1} (see Figure 5.3). These values will be our reference to assess the catalytic role of the (101) anatase TiO_2 surface.

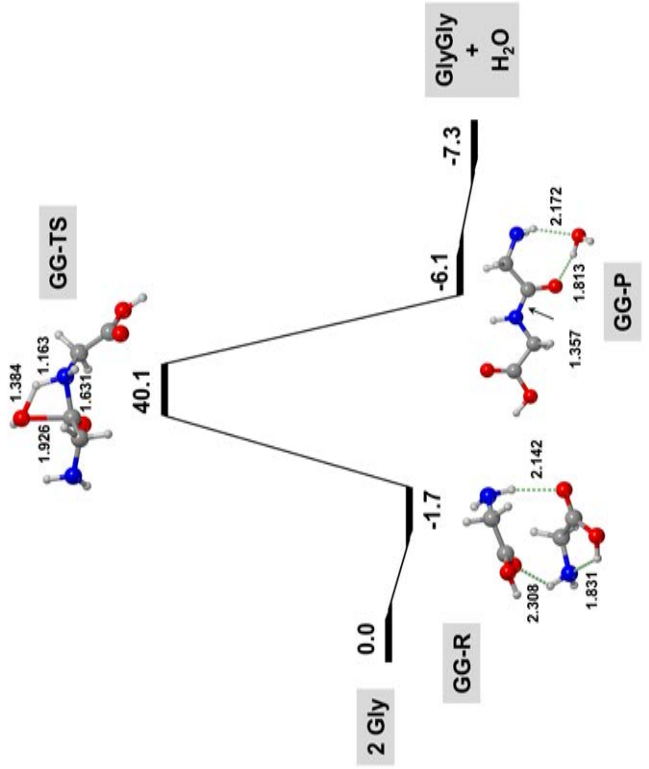


Figure 5.3: Relative free energy profile (T = 298 K) in kcal mol^{-1} at PBE0-D2*//PBED2* theory level for the uncatalyzed gas-phase Gly + Gly condensation reaction. Bond distances are in Å.

5.3.1 Gly + Gly → GlyGly + H_2O

Chapter 4 showed that the most stable complex of Gly interacting with the (101) anatase surface is that in which Gly is deprotonated due to a proton transfer from the –COOH group to the surface. This structure will be referred to as Gly/TiO_2 system along the work. The Gly/TiO_2 system was chosen as the pre-reactant adduct through which the peptide bond formation occurs by the reaction with an incoming Gly molecule.

Figure 5.4 shows the free energy profile at T = 298 K of the reaction, taking as the 0^{th} energy reference the asymptote Gly/TiO_2 + Gly, here named SGG-AS. This asymptote is defined as Gly/TiO_2 and an isolated Gly molecule infinitely separated. At variance with the process in gas-phase, the reaction on the surface adopts a stepwise mechanism. The first step involves the simultaneous nucleophilic attack and the proton transfer. In this case, however, since the adsorbed Gly is in its deprotonated state, the release of H_2O is not possible. The dehydration takes place in the second step, in which the surface proton transfers to the OH group to form water. This water, once released, can be attached to the surface either through H-bond with surface O atoms or by interaction with a coordinatively unsaturated Ti atom, the latter being a much more favourable situation. The calculated free energy profile is overall more favourable than the uncatalysed gas-phase process. The first step shows the highest energy barrier ($\Delta G^{\ddagger}_{298}$ = 35.6 kcal mol^{-1}), while the energy barrier of the dehydration step (with respect to the SGG-AS asymptote) is $\Delta G^{\ddagger}_{298}$ = 17.5 kcal mol^{-1}. The first step is about 6 kcal mol^{-1} lower than the analogous process in gas phase, thus showing

that the interaction with the surface does indeed exert a catalytic effect. With the aim to have deeper insights onto this aspect, we have performed a Bader charge analysis of the reactant structure, in particular of the C atom of the carboxylic group. The interaction with the surface induces an increase of the positive charge of the C atom of the COO group. Indeed, for gas-phase Gly, Bader charge is 1.51 e, while on the surface (*i.e.*, the Gly/TiO_2 complex) it is 1.58 e. This means that the C atom becomes more electrophilic upon adsorption and accordingly more prone to be attacked by the N atom of the incoming Gly molecule. Such charge differences are also reflected on the values of the C–N distance in the reactant and transition state structures of the uncatalysed reaction and on the surface. That is, on the surface the C-N distances are significantly shorter (2.866 and 1.604 Åfor SGG-R and SGG-TS1, respectively, see Figure 5.4) than in the gas phase (3.077 and 1.631 Å, respectively, see Figure 5.3). Despite this catalytic effect, the barrier is still significantly high due to the fourth-membered ring present in the transition state structure (see structure SGG-TS1 of Figure 5.4). The thermodynamics of the reaction is, however, only favourable when the released water interacts with the surface through a surface Ti atom ($\Delta_r G_{298}$ = -23.6 kcal mol^{-1}, see structure SGG-P2 of Figure 5.4). This fact is very important because it points out that the surface can act as a dehydrating agent capturing the formed water, thereby favouring the energetics of the reaction.

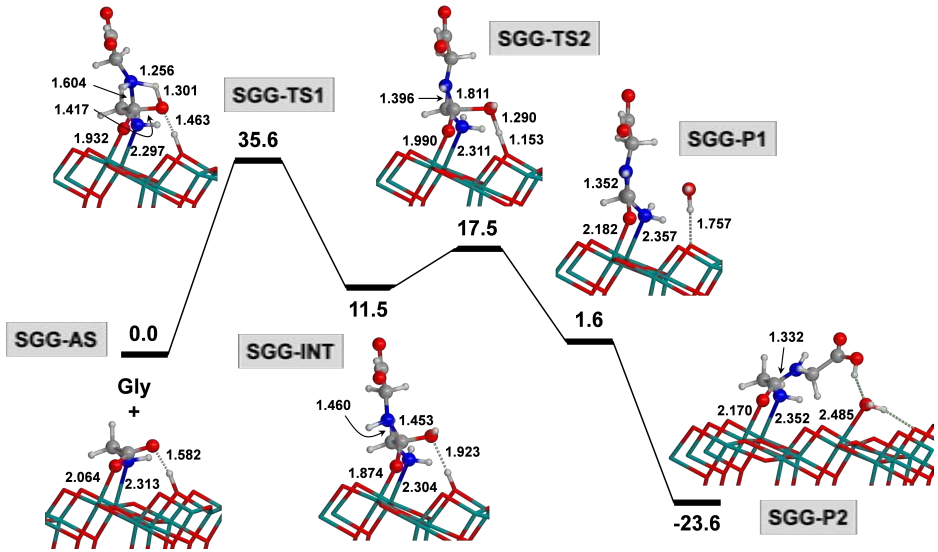


Figure 5.4: Relative free energy profile (T = 298 K) in kcal mol^{-1} at PBE0-D2*//PBE-D2* theory level for the peptide bond formation between two Gly molecules on the (101) anatase surface. The asymptote 0^{th} energy reference is Gly/TiO_2 + Gly (SGG-AS structure). Bond distances are in Å.

5.3.2 Gly + Gly + nH$_2$O → GlyGly + (n+1)H$_2$O (n=1,2)

Some catalytic role of the (101) anatase surface was demonstrated in the previous section. However, the energy barrier is still significantly high to be surmountable at normal conditions. As mentioned above, the fourth-membered ring transition state (SGG-TS1) structure is geometrically highly strained. Thus, a reasonable way to decrease the energy barrier is by reducing the geometrical strain of the transition state. It has long been recognized that water is an efficient catalyst for this purpose because it can act as a proton transfer assistant; *i.e.*, it can help the occurrence of proton transfers by accepting/donating protons simultaneously, reducing the geometrical strain. Accordingly, we have studied the peptide bond formation on the (101) anatase surface in the presence of 1 and 2 water molecules assisting the proton transfer of the first step, resembling moderately dried conditions, as in the experiment.

The calculated free energy profile at 298 K for the reaction assisted by one water molecule is shown in Figure 5.5. The 0^{th} energy reference is the asymptote SGGF-AS, which involves the Gly/TiO_2 system and the isolated Gly and H_2O molecules, all of them infinitely separated. The energy barrier of the first step lowers up to $\Delta G^{\ddagger}_{298} = 20.7$ kcal mol^{-1}, which means a reduction of about 15 kcal mol^{-1} compared to the non-water-assisted mechanism. This lowering is due to the less strained ring present in the transition state structure (SGGW-TS1). That is, with one water acting as a proton transfer assistant, the transition state exhibits a sixth-membered ring. In relation to the second step, calculated $\Delta G^{\ddagger}_{298}$ with respect to the asymptote is 22.9

kcal mol^{-1} (see SGGW-TS2), which is 5.4 kcal mol^{-1} higher than the analogue step in the non-water-assisted reaction. This increase is due to the destabilization of the SGGW-TS2 transition state, which is reflected by the $H_{surface}$- - -O_{Gly} forming bond distance; *i.e.*, while in SGG-TS2 (non-water-assisted) this distance is 1.153 Å, in SGGW-TS2 (1-water-assisted) is 1.465 Å, meaning that in this second case the surface OH is almost completely broken, yielding a more energetic transition state. In the same line of the non-water-assisted mechanism, the thermodynamics of the reaction is favourable ($\Delta_r G_{298}$ = -29.3 kcal mol^{-1}, see SGGW-P2) only when the two water molecules present in the final product (namely, that assisting the proton transfer and that formed in the process) are interacting with Ti surface atoms through dative covalent bonds.

The calculated free energy profile at 298 K for the reaction assisted by two water molecules is shown in Figure 5.6. Here, the asymptote corresponds to Gly/TiO_2 plus a Gly molecule and two water molecules per separated. Considering only the potential energy surface, we identified a first stationary point (SGG2W-R) in which the N–C bond forms without the simultaneous H transfer. Here, a HOOC–CH_2–NH_2^+-COO^{2-}–CH_2–NH_2 zwitterionic species is formed. This is at variance of the two previous processes in which the N-C bond formation takes place simultaneously to the H transfer. Such a zwitterion is stable (in terms of potential energies) because its charges are stabilized by the two water molecules. Interestingly, formation of this stationary point is barrierless, as the energy decreases continuously by decreasing the N–C distance (*i.e.*, by performing a scan calculation using the N-C distance as the distinguished coordinate). However, when considering the free energies, SGG2W-R is

found to be unstable (*i.e.*, it is 6.4 kcal mol^{-1} more energetic than the asymptote) mainly due to entropic effects so that it is a physically unsounded structure of the free energy profile and can be neglected.

This peptide bond reaction is followed by the proton transfer from the NH_2 to the COO groups. Here, the 2 waters directly participate assisting the proton transfer. The transition state structure (see SGG2W-TS1 of Figure 5.6) exhibits an even less strained ring (with eight members) than the 1-water-assisted-analogue. The intrinsic free energy barrier of this step is actually low ($\Delta G^{\ddagger}_{298}$ = 8.0 kcal mol^{-1}) due to the geometry relaxation of this transition state. The next and final step is the formation and release of water, which is performed by the proton transfer from the surface to the OH group. This dehydration has a $\Delta G^{\ddagger}_{298}$ = 18.3 kcal mol^{-1}, which is relatively lower than the 1-water-assisted analogue step. The final product is the peptide attached on the surface and three water molecules, which in case to be attached to interact with surface Ti Lewis sites, the reaction is highly favourable ($\Delta_r G_{298}$ = -45.0 kcal mol^{-1}).

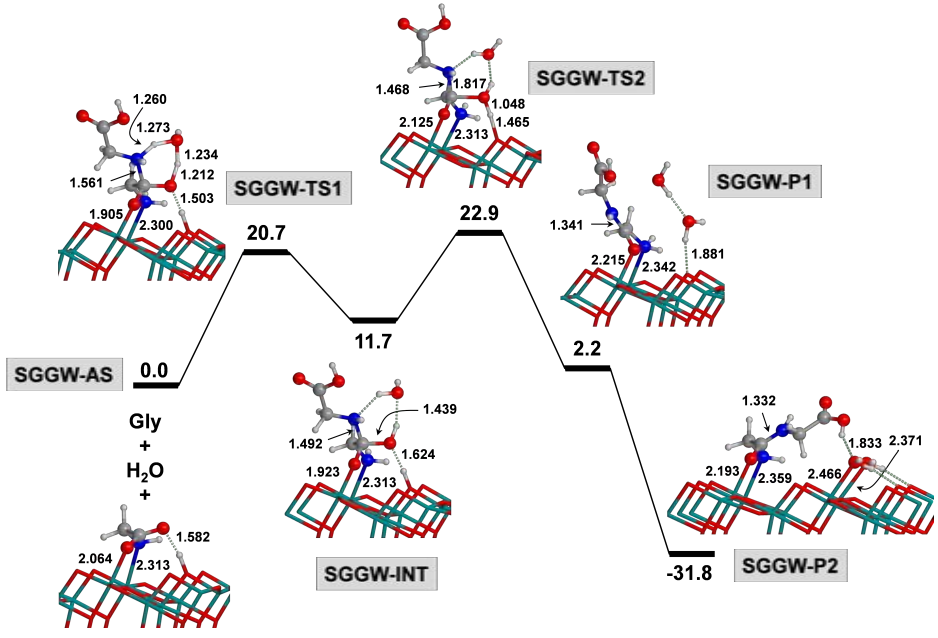


Figure 5.5: Relative free energy profile (T = 298 K) in kcal mol^{-1} at PBE0-D2*//PBE-D2* theory level for peptide bond formation between two Gly molecules on the (101) anatase surface in the presence of 1 water molecule assisting the proton transfer. The asymptote 0^{th} energy reference is Gly/TiO_2 + Gly + H_2O (SGGW-AS structure). Bond distances are in Å.

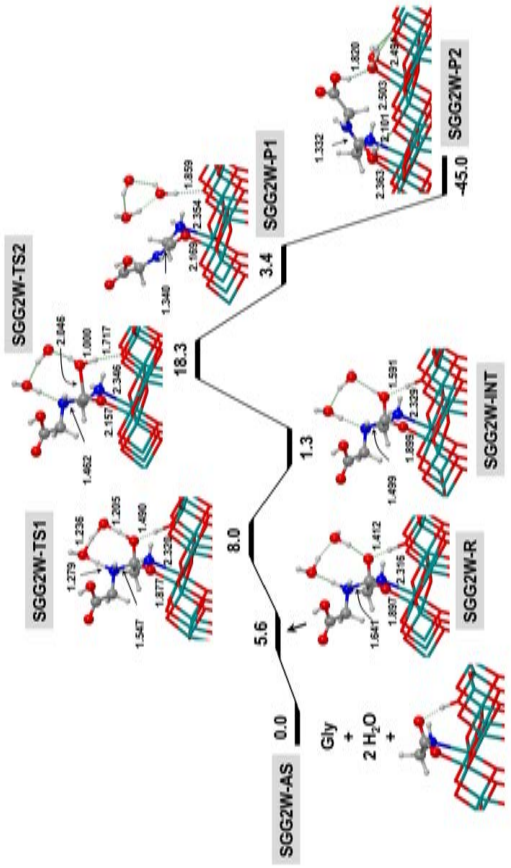


Figure 5.6: Relative free energy profile (T = 298 K) in kcal mol^{-1} at PBE0-D2*//PBE-D2* theory level for the peptide bond formation between two Gly molecules on the (101) anatase surface in the presence of 2 water molecule assisting the proton transfer. The asymptote 0^{th} energy reference is $Gly/TiO_2 + Gly + 2H_2O$ (SGG2W-AS structure). Bond distances are in Å.

5.3.3 Gly + Gly + Gly → GlyGly + Gly

The two previous sections have demonstrated that the energy barriers involved in the peptide bond formation can be significantly reduced by the catalytic activity exerted, on one side, by the interaction of Gly with surface Ti Lewis sites, which makes the C atom more electrophilic, and on the other side, by discrete water molecules assisting the proton transfers. The proposed mechanisms shown above evidence that condensation between two Gly molecules (and by extension between amino acids) is feasible in a prebiotic scenario under the premise of two conditions: i) the reaction has to occur on anatase surfaces, and ii) fluctuating wetting/drying cycles have to operate, since the reaction occurs under moderately dried conditions. These results, however, do not actually explain the experimental findings obtained by Martra et al.[25] in which Gly-based oligopeptides were formed on anatase TiO_2 surfaces by condensation of vapours of monomers. Indeed, in their experiments, the presence of initial water molecules in the experimental set up are rule out since the mineral samples were outgassed at high temperatures to remove residual water and the Gly monomers were introduced by successive feedings from the vapour phase. The aim of this last section is, thus, to provide an atomistic interpretation of these experimental results. It is anticipated that the proposed mechanism advocates the role of a third Gly molecule participating in the reaction as a proton transfer assistant through its COOH group.

Figure 5.7 shows the free energy profile at T = 298 K of the calculated mechanism. The 0^{th} energy reference asymptote (SGGG-AS) is Gly/TiO_2 plus two Gly molecules infinitely separated. As

observed for the mechanism with 2 water molecules, a stationary point in which the N–C bond is formed was also identified. In this case, charges of the organic moiety are stabilized by its interaction with the third Gly molecule. At variance with the $2H_2O$-assisted mechanism, however, in terms of free energies, this stationary point is stable with respect to the asymptote (5.8 kcal mol^{-1} more stable) and accordingly it can be understood as an actual intermediate species of the reaction. The next step, involving the proton transfer from the NH_2 group, is assisted by the COOH group of the third Gly. In this proton-assisted mechanism the C=O group acts as a proton acceptor and the OH as a proton donor (see structure SGGG-TS1 of Figure 5.7). This transition state structure has an eight-membered ring (as occurred in the $2H_2O$-assisted mechanism), which is a low strained structure. Accordingly, the intrinsic $\Delta G^{\ddagger}_{298}$ is actually low (2.0 kcal mol^{-1}, with respect to SGGG-INT1). It is worth mentioning that the immediate intermediate structure after this proton transfer (SGGG-INT2) is 0.3 kcal mol^{-1} higher in energy than the TS. This is due to the thermal corrections included to calculate the free energies. The final step leads to the water elimination and has an intrinsic $\Delta G^{\ddagger}_{298}$ value (*i.e.*, with respect to SGGG-INT1) of 17.7 kcal mol^{-1} (see SGGG-TS2). This transition state is a product-like species, the emerging water molecule being almost formed with a large C–OH distance (2.234 Å, practically broken). The most stable structure for the final product is SGGG-P2 ($\Delta_r G_{298}$ = -26.7 kcal mol^{-1}), in which the released water is attached to the surface on a Ti atom, and the third Gly is interacting with the formed peptide. It is worth mentioning that the same process (*i.e.*, the peptide bond formation assisted by a third Gly molecule) has been calculated in absence of

the anatase surface. The reaction was found to be concerted with a calculated $\Delta G^{\ddagger}_{298}$ of 23.4 kcal mol^{-1} (see Figure 5.8), which implies a decrease to almost half of the barrier computed for the uncatalysed gas-phase process. Despite this reduction, the energy barrier on the surface is dramatically lower, indicating that the anatase surface is of paramount importance. Furthermore, formation of the trimer from three glycine molecules in gas-phase is disfavoured by entropic reasons and by the relative low vapour pressure, which render the process unlikely. On the contrary, it is easy to imagine that during the drying process the glycine solution is concentrated, ending up with clustering of the glycine molecules on the TiO_2 surface.

Interestingly, the highest calculated free energy barrier of the proposed mechanism is that involving the H_2O formation (SGGG-TS2 of Figure 5.7). The reaction in the experiments of Martra et al.[25] occurred at the IR beam temperature; *i.e.*, about 50° C (323 K). Accordingly, we have calculated the free energy barrier at this temperature ($\Delta G^{\ddagger}_{298}$ = 16.5 kcal mol^{-1}), as well as the rate constant for this rate determining step using the classical Eyring equation (considering a first-order reaction since it starts from SGGG-INT2). The obtained results are k $\sim$ 46 s^{-1} and a $t_{1/2} \sim 1.5 \times 10^{-2}$ s, thus indicating that the reaction is actually fast and, therefore, the occurrence of this mechanism is feasible under the experimental conditions.

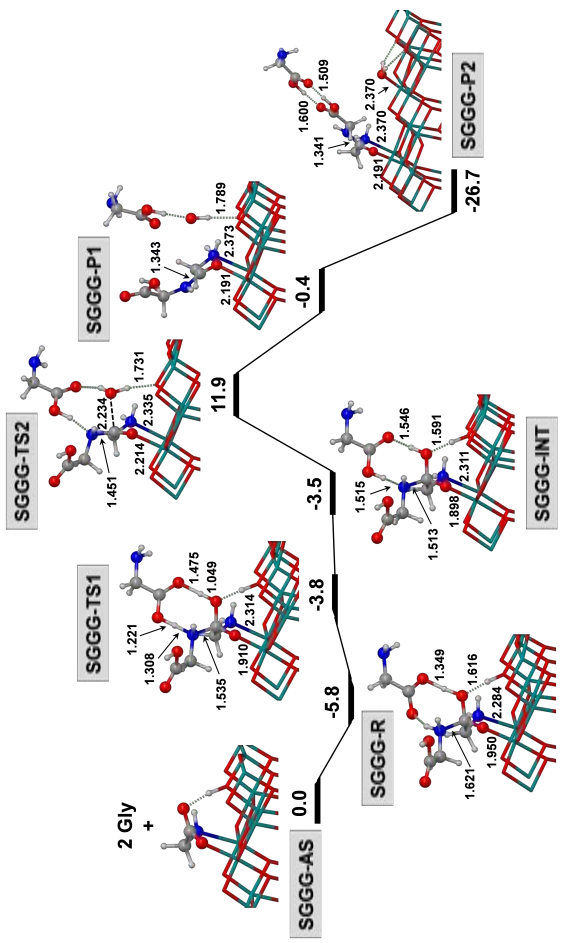


Figure 5.7: Relative free energy profile (T = 298 K) in kcal mol^{-1} at PBE0-D2*//PBE-D2* theory level for the peptide bond formation between two Gly molecules on the (101) anatase surface in the presence of third Gly molecule assisting the proton transfer. The asymptote 0^{th} energy reference is Gly/TiO_2 + 2Gly (SGGG-AS structure). Bond distances are in Å.

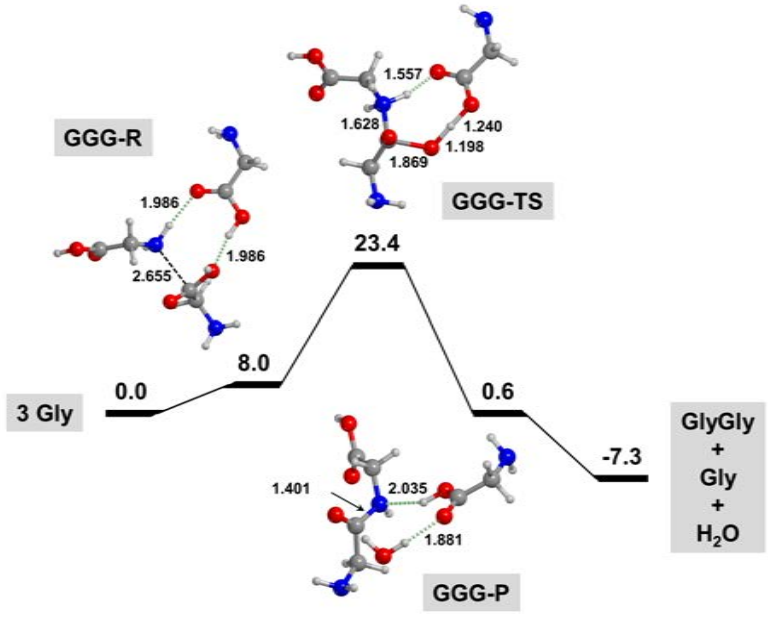


Figure 5.8: Relative free energy profile (T = 298 K) at PBE0-D2*//PBE-D2* theory level for the Gly + Gly condensation reaction in the presence of a third Gly molecule assisting the proton transfer. Bond distances are in Å.

5.4 Final remarks

In the sequence of organizational events leading to the emergence of life, many competing scenarios have been proposed to explain the appearance of the first biopolymers such as proteins and RNA. Formation of an amide bond between two amino acids to give a peptide (*i.e.*, the peptide bond formation reaction) is a critical and poorly understood step in this organizational ladder. A major concern is the so-called "water paradox", since the reaction in highly diluted water solutions is thermodynamically disfavoured because amino acid condensation is followed by water elimination. The reaction also presents low kinetics with reaction half-times of the order of several centuries. An old but still fashionable proposal by Bernal,[1] in which mineral surfaces play a key role in the reaction by concentrating the monomers and catalysing their polymerization by action of surface active sites. This polymerization on the rocks, in combination with fluctuating wetting/drying cycles,[289] in which the condensation becomes possible during the drying cycle, in a primitive Earth is one of the hypothesis to explain the occurrence of the reaction under prebiotic conditions. Despite a great deal of successful experimental work using several minerals (see Introduction), few attempts have been provided from the computational viewpoint to give a mechanistic interpretation of the catalytic role of the surfaces. The present study aims to fill in this gap for the particular case of the condensation reaction between two glycine (Gly) molecules in the presence of the TiO_2 (101) anatase surface containing both coordinatively unsaturated Ti atoms and O atoms acting as Lewis and Brønsted sites, respectively, from an atomistic point of view.

For all the studied reactions, the most stable adduct of Gly on the anatase surface was taken as the pre-reactant complex, in which Gly is in its deprotonated form due to a spontaneous proton transfer to the surface.[290] The main conclusions can be drawn by comparing the energetics for the uncatalysed gas-phase reaction with those resulting on the TiO_2 surface. The gas-phase process is concerted and envisages a nucleophilic attack of the NH_2 group of one Gly to the C atom of the COOH group of the other Gly forming a N–C bond, followed by a proton transfer from the NH_2 group to the OH group to release water. As mentioned, in gas phase the reaction in concerted with a free energy barrier (T = 298 K) of 41.8 kcal mol-1 and reaction free energy of -7.3 kcal mol^{-1}. In constrast, all the studied condensations on the (101) anatase surface adopt a stepwise mechanism, in which the nucleophilic attack and the release of water splits in two steps because of the deprotonated state of Gly in the Gly/TiO_2 pre-reactant complex. The Gly + Gly/TiO_2 reaction presents a free energy barrier decrease for the N–C nucleophilic attack step of about 6 kcal mol^{-1}. This activation energy reduction is caused by the Gly/TiO_2 surface interaction, which renders the C atom more electrophilic than in the gas phase, as indicated a Bader charge analysis. The final dehydration step takes place by proton transfer from the surface to the OH of the organic moiety releasing water, with an intrinsic free energy barrier of 17.5 kcal mol^{-1}. The thermodynamics of the reaction is also more favourable than the gas-phase process because the released water can interact favourably with the (101) anatase surface through covalent dative bonds between the O atom of the water and surface Ti Lewis sites. These largely favourable free reaction energies are kept for the other studied reactions, indicating

that the surface clearly helps the global reaction energies.

The reaction has also been studied in the presence of 1 and 2 water molecules acting as proton transfer assistants. Their presence dramatically reduces the free energy barriers for the nucleophilic attack to 20.7 and 8.0 kcal mol^{-1}, respectively. This reduction is due to the low-strained rings present in respective transition state structures in the proton transfer (6th- and 8th-membered rings, respectively) compared to the highly strained ring in the non-water assisted reactions (4th-membered ring). In both cases, due to this energy barrier decrease, the final dehydration becomes the more energetic step, with energy barriers of 22.9 and 18.3 kcal mol^{-1}, respectively.

Finally, the Gly + Gly/TiO_2 condensation reaction has also been studied in the presence of a third Gly molecule acting as proton transfer helper. This process corresponds to the aggregation of glycine molecules during the drying process on the TiO_2 surface. Here, the mechanism involves three steps. The first is the N–C bond formation, which occurs in a barrierless fashion, forming a stable intermediate in which a HOOC–CH_2-NH_2^+–COO^{2-}–CH_2–NH_2 organic moiety is attached on the surface. The second is the proton transfer from the surface to the organic moiety. This proton transfer is assisted by the COOH group of the third Gly, in which the C=O group acts as proton acceptor and the OH group as proton donor forming a low strained 8th-membered ring. The calculated free energy barrier is of 2 kcal mol^{-1}. The final step is the dehydration, which is found to be the most energetic one with a free energy barrier of about 11.9 kcal mol^{-1}. This mechanistic proposal is consistent and provides an atomistic interpretation of the experimental findings obtained by Martra et al.[25] in which catalytic polymerization of Gly monomers

on TiO_2 surfaces was successfully achieved by successive feedings of the monomers from strict gas-phase conditions.

Chapter 6

Secondary structures: α-helices and β-sheets

The biological activity of a given protein is partly due to its secondary structure and conformational state. Peptide chains are rather flexible so that the interest in finding ways that force the protein folding in a well-defined state is of great importance. Among the different constraint techniques, the interaction of proteins with inorganic surfaces is a fruitful strategy to stabilize selected folded states. Surface-induced peptide folding can have potential applications in different biomedicine areas but it can also be of fundamental interest due to its implications in prebiotic chemistry if a peptide turns-on its biological activity when folded in a given state. In this work, full periodic quantum mechanics (QM) static optimizations and *ab-initio* molecular dynamics (AIMD) simulations at the PBE-D2* level have been carried out to study the adsorption and the relative stability of secondary polyglycine structures (*i.e.*, linear, α-helix and β-sheet) ad-

sorbed on the (101) TiO_2 anatase surface. In this study, to reduce the computational cost, periodic boundary conditions (PBC) have been applied both to the surface and to the biological part, thus obtaining a periodic polypeptide system on the TiO_2 surface. Static QM results indicate that the preferred conformation on the (101) anatase surface is the α-helix over the β-sheet, with calculated adsorption energies per glycine unit of -33.7 and -30.2 kJ mol^{-1}, respectively. AIMD simulations have been very useful to identify the most stable complex, in which thermal and entropic effects are shown to be essential. This is particularly true for the α-helix system, in which on the (101) anatase surface the helical structure becomes partially denatured because the peptide C=O groups close to the surface form Ti–O dative bonds rather than keeping the α-helix H-bond pattern. This structure was identified as the most stable one on the TiO_2 surfaces in comparison to the adsorption of β-sheet structures.

6.1 Introduction

6.2 Computational details

6.2.1 Methods

All the calculations were carried out using the Perdew-Burke-Ernzerhof (PBE)[167] functional to obtain equilibrium geometries end energies with the *a posteriori* Grimme D2* correction.[184, 237, 238] The SCF iterative procedure has been converged to $\Delta E = 10^{-6}$ eV while the tolerance on gradients for geometry optimizations has been set to 0.01 eV/Å for each atom in each direction. The k-points

mesh has been set to (3,3,1) for the smallest systems (*i.e.*, adsorption of primary structure and β-sheets) and to (2,2,1) for the α-helix/TiO_2 system.

Vibrational frequencies of a reduced Hessian (following only CO displacements) on the optimised structures have been computed at the Γ point by numerical differentiation of the analytical first derivatives, using the central difference formula (i.e. two displacements for each atom in each direction). The SCF iterative procedure for the frequency calculations was converged to a tolerance in total energy of $\Delta E = 10^{-6}$ eV.

Ab-initio molecular dynamics (AIMD) simulations have been carried out on all the secondary structures analysed. The value of the energy cutoff for the plane wave basis set has been set to 400 eV, while the self-consistent field (SCF) iterative procedure was converged to a tolerance in total energy of $\Delta E = 10^{-5}$ eV. AIMD simulations have been carried out considering an equilibration period of 1 ps (1000 steps of 1 fs) in the NVE ensemble at fixed temperature (T = 300K) with the velocities scaled at each step of the calculation, followed by 10 ps of production within the canonical ensemble (NVT), with the frequency of the temperature oscillation controlled by the Nosé mass. During both the equilibration and production periods the last atomic layer of the TiO_2 slab have been maintained at fixed positions; *i.e.*, only the upper atoms of the surface and those of the polyglicine systems have been allowed to move according to the motion's equations. We have chosen this option to avoid unrealistic deformations of the internal structure of the slabs, since real TiO_2 surfaces are linked to a macroscopic bulk.

6.2.2 Surface and polyclycine models

From the crystal bulk structure of the TiO_2 anatase polymorph, we have built a crystalline periodic slab model for the non-polar (101) surface with a thickness of 10 Å (3 layers of TiO_2), as in the previous Chapters. For the adsorption of the polyglycine systems on the (101) anatase surface the internal atomic positions were only optimised, while the cell parameters remained fix to the experimental values ($a = 7.569$ Å and $b = 10.23935$ Å). The interlayer distance, regulated by the c value, was defined considering an empty space of at least 15 Å between the outermost atoms of the polyglycine systems and the upper layer, which is enough to avoid mutual fictitious interactions between the periodically repeated slabs. Accordingly, the c value was set to 30 Å.

The polyglycine/TiO_2 systems have been modelled adopting a full periodic approach, in which the periodicity used to model the slab is also used to model the peptides. For the β-sheets systems, periodicity along the a direction has served to model the infinite β-strands and along the b direction to account for the lateral interactions between strands of adjacent unit cells. For the primary peptide structure and the α-helix, periodicity along one direction has been used to model the infinite linear and helical structures (note that for these cases, no interactions between adjacent peptides take place). Figures 6.1, 6.2, 6.3 and 6.4 show the isolated periodic polyglycine systems. This strategy has been useful to reduce the computational cost of the calculations: the largest polyglycine systems is the α-helix, which is modelled by repeating 7 glycine residues (193 atoms per unit cell), while for a reasonable non-periodic α-helix system we should use a

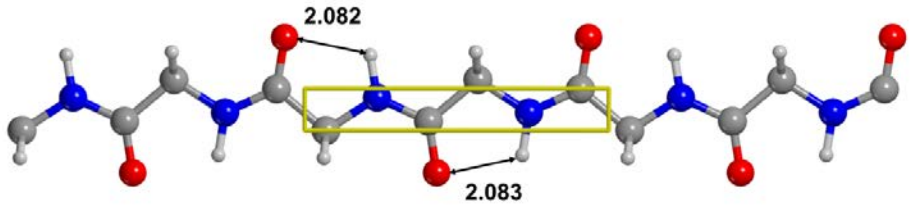


Figure 6.1: PBE-D2* optimised geometry for the isolated periodic linear polyglycine polymer. The unit cell is represented in yellow. Bond distances are in Å.

minimum of 12 glycine (more than 500 atoms). This strategy is possible due to the very good match between the cell dimensions of the periodic polyglycine systems and the (101) anatase surface. The linear polymer has a unit cell of $a = 7.2358$ Å (very close to the a value of the surface); for the β-sheets in their parallel and antiparallel conformations the unit cell parameters are $a = 7.2638$ Å and $b = 10.1084$ Å, and $a = 7.2782$ Å and $b = 9.6926$ Å, respectively (very close to the a and b values of the surface), and for the α-helix $a = 10.4244$ Å (very close to the b value of the surface). For the adsorption of the linear peptide and the β-sheet systems we have used a 2×2 supercell ($a = 7.569$ Å and $b = 10.23935$ Å), while for the α-helix we have used a 4×2 supercell ($a = 15.138$ Å and $b = 10.23935$ Å) to avoid lateral interactions along the a direction.

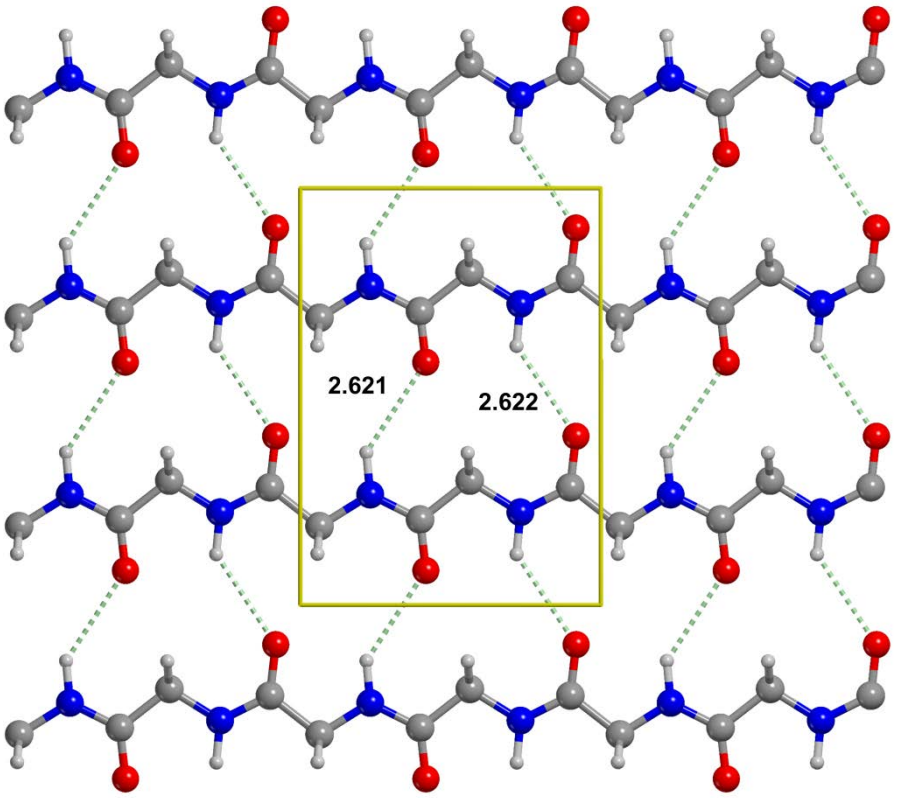


Figure 6.2: PBE-D2* optimised geometry for the isolated periodic parallel β-sheet. The unit cell is represented in yellow. Bond distances are in Å.

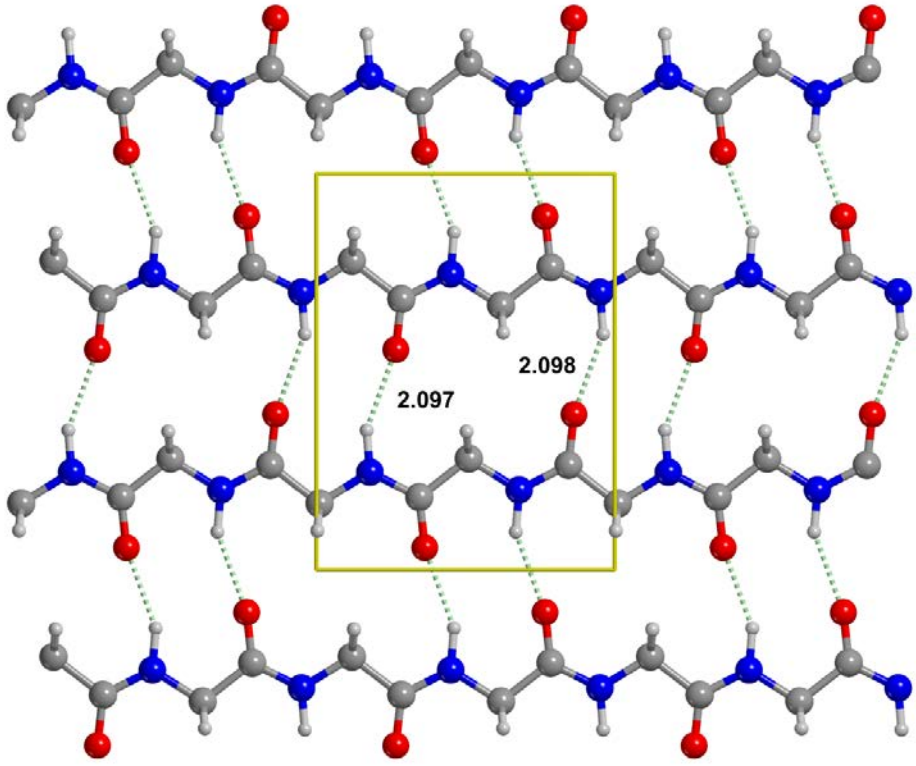


Figure 6.3: PBE-D2* optimised geometry for the isolated periodic antiparallel β-sheet. The unit cell is represented in yellow. Bond distances are in Å.

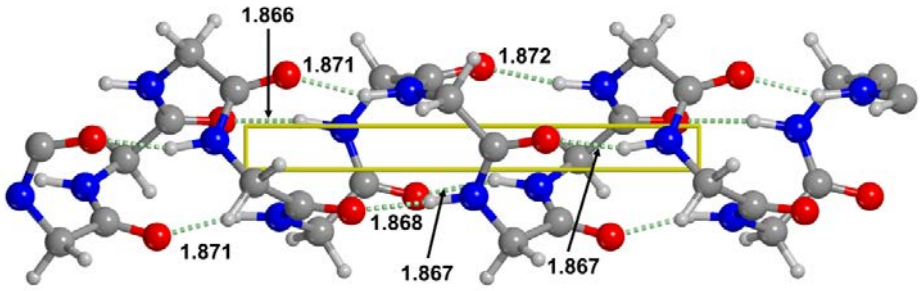


Figure 6.4: PBE-D2* optimised geometry for the isolated periodic α-helix. The unit cell is represented in yellow. Bond distances are in Å.

6.2.3 Adsorption energies

Adsorption energies have been calculated as:

$$\Delta E_{ADS} = E_{CPLX} - E_{PGLY} - E_{SURF} \tag{6.1}$$

where E_{CPLX} is the energy of the complex (polyglycine/surface), E_{PGLY} is the energy of the isolated polyglycine system, and E_{SURF} is the energy of the isolated surface. To compute E_{PGLY} we have optimised the polyglycine systems using an empty space of 20 Å in the non-periodic directions, relaxing both atomic positions and cell parameters in order to include the cost of the cell deformation when the polyglycine interacts with the inorganic surface.

In the case of the β-sheet complexes it is interesting to evaluate all the most important contributions to the total adsorption energy, *i.e.* not only the adsorption energy of the whole β-sheet, but also the adsorption energy of each β-strand with the surface and the interaction energy of the H-bond interactions between the β-strands.

Figure 6.5 represents schematically the β-sheet complex, formed by two β-strands (P1 and P2) per unit cell interacting with the surface (S). This is a common three-body problem, not resolvable analytically. However, in order to estimate each energy contribution to the total energy with a good accuracy, we have proceeded as follows. From a given optimised β-sheet/TiO_2 complex, we removed one or two of the three bodies, and performed single-point energy calculations. This way, we can decompose the total interaction energy as: i) E_{Hb}, the interaction energy between P1 and P2 without S (the H-bond contribution), ii) $\Delta E1$, the interaction energy of P1 with S without P2, and iii) $\Delta E2$, the interaction energy of P2 with S without P1. Considering that ES, E_{P1} and E_{P2} are the single-point absolute energies of isolated S, P1 and P2, respectively, calculation of each contribution (E_{Hb}, $E1$ and $E2$) is done as:

$$\Delta E_{Hb} = E_{\beta} - (E_{P1} - E_{P2}) \tag{6.2}$$

where E_{β} is the single-point absolute energy of the β-sheet without S;

$$\Delta E1 = E_{P1S} - (E_S - E_{P1}) \tag{6.3}$$

where E_{P1S} is the single-point absolute energy of the P1/S complex without P2;

$$\Delta E2 = E_{P2S} - (E_S - E_{P2}) \tag{6.4}$$

where E_{P2S} is the single-point absolute energy of the P2/S complex without P1.

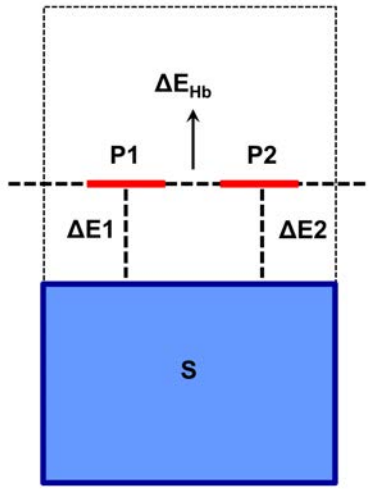


Figure 6.5: Schematic representation of the β-sheet/TiO_2 complexes and the interactions involved.

6.3 Results

6.3.1 Linear polyglycine

The TiO_2 (101) anatase surface presents Lewis and Brønsted sites, represented by the outermost Ti and O atoms, respectively; these sites are capable to strongly interact with amino acidic residues. Secondary structures present particular H-bond patterns between different residues: H-bond donors are N–H groups, while H-bond acceptors are C=O groups. The surface is capable to break these interactions: when a C=O group is close to a Ti atom of the surface it could prefer to interact with the Ti instead through H-bonding, since the Ti–O dative bond is largely stronger.

The simplest polyglycine system is its linear polymer forming a

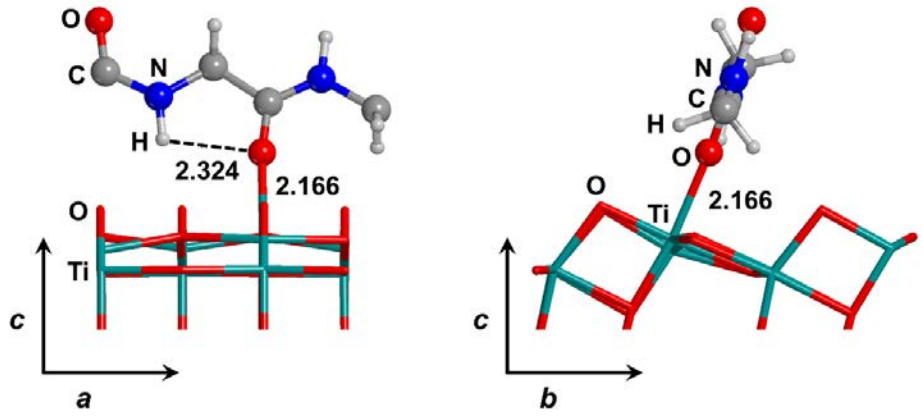


Figure 6.6: Lateral views of the optimised structure at PBE-D2* level of the linear/TiO_2 complex. Bond distances are in Å.

primary structure (see Figure 6.6). The complex of this linear peptide structure in interaction with the (101) anatase surface will be named along the Chapter as linear/TiO_2 complex. Here, we modelled the primary structure within a periodic framework by repeating 2 amino acidic units along the b direction of the unit cell. As one can see, the polymer interacts with the surface through a Ti–O dative bond, with ΔE_{ADS} = -62.0 kJ mol^{-1} and ΔE_{ADS}/AA unit = -31.0 kJ mol^{-1} (see Table 6.2). This very simple system is enough to observe that interaction with the surface induces structural changes on the peptide structure. Indeed, in absence of the surface, the distance between two adjacent C=O/N–H groups is 2.08 Å (see Figure 6.1). However, on the surface, the distance increases to 2.32 Å. This interaction between the C=O and N–H groups as a weak H-bond may be related to the fact that the interaction with the surface weakens the H-bond between C=O and N–H groups.

6.3.2 β-sheet

β-sheet both in its parallel and antiparallel conformations (hereafter referred to as β-sheet_P and β-sheet_AP, respectively) have been considered. In β-sheet_AP, the two β-strands run in opposite directions, so that the N-terminus of one β-strand is adjacent to the C-terminus of the other. In contrast, in β-sheet_P, the two β-strands run in the same direction and accordingly the N-terminus of one β-strand is opposite the N-terminus of the other β-strand (see Figures 6.2 and 6.3). Due to these dispositions of the N- and C- terminus, in β-sheet_AP the H-bonds are aligned directly opposite each other, while in β-sheet_P the geometry of the individual AA residue forces the H-bond to occur at an angle. Thus, H-bonds in β-sheet_AP are stronger and more stable than in β-sheet_P.

In both β-sheet_AP/TiO_2 and β-sheet_P/TiO_2 complexes, the β-sheets are formed by two strands per unit cell. The periodicity along the two cell directions defining the surface allows us to model extended β-sheets. Figures 6.7 and 6.8 show the optimised structures of the β-sheet/TiO_2 complexes and their evolution during the MD simulation. In both systems, one β-strand interacts with the surface via a Ti–O dative bond, while the other through dispersion forces, since neither Ti-O nor H-bond interactions are established.

By comparing the structures of the isolated β-sheets with those on the surface, for both β-sheet conformations, adsorption leads to an increase of the H-bond distances (2.621 Å vs 2.668 – 2.802 Å and 2.097 – 2.098 Å vs 2.297 – 2.587 Å for β-sheet_P and β-sheet_AP , respectively), for the H-bonds involving C=O groups, particularly those ones forming Ti–O dative bonds. Therefore, upon adsorption,

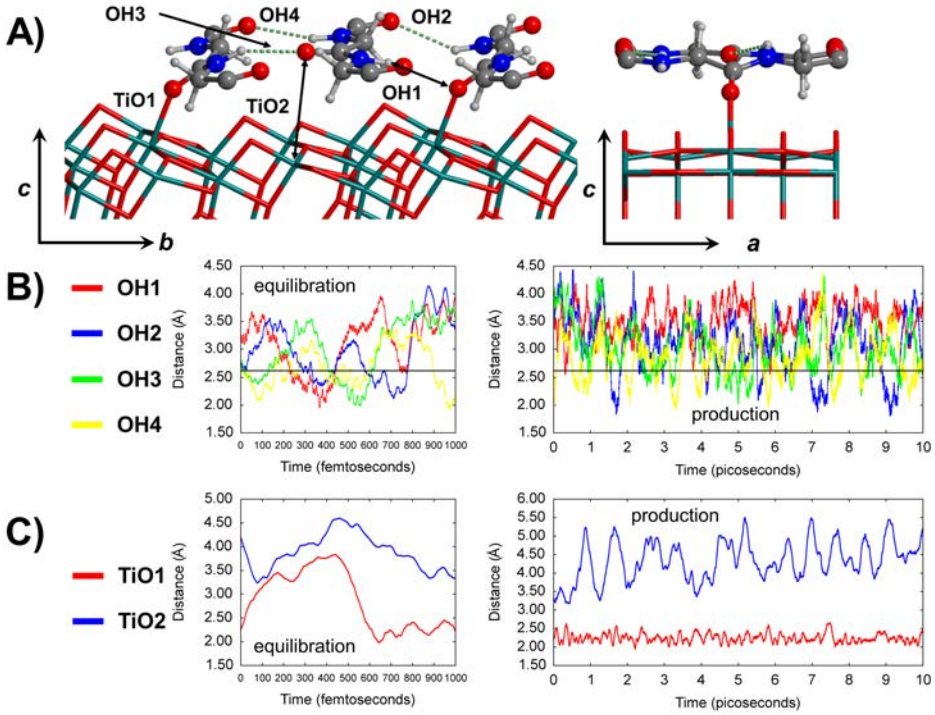


Figure 6.7: PBE-D2* optimised structure of the β-sheet_P/TiO_2 complex (A), and evolution of the H-bond distances (B) and of the Ti-O distances (C) along the AIMD simulations. The equilibration and production periods are separated. Bond distances are in Å.

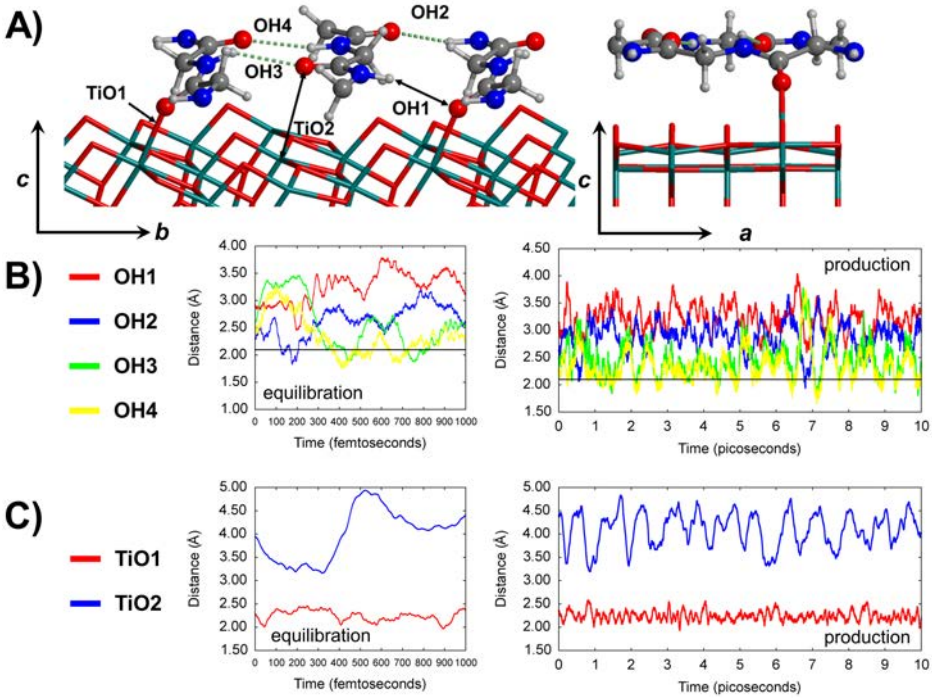


Figure 6.8: PBE-D2* optimised structure of the β-sheet_AP/TiO_2 complex (A), and evolution of the H-bond distances (B) and of the Ti-O distances (C) along the AIMD simulations. The equilibration and production periods are separated. Bond distances are in Å.

the H-bond pattern in β-sheet_AP/TiO_2 is stronger and more stable than in β-sheet_P/TiO_2. Accordingly, the former complex presents a higher adsorption energy than the latter one (ΔE_{ADS} = -120.8 and -115.7 kJ mol^{-1} and ΔE_{ADS}/AA unit = -30.2 and -28.9 kJ mol^{-1}, respectively). Data are presented in Table 6.2

As previously mentioned, by performing single-point energy calculations on the β-sheet/TiO_2 complexes, we have decomposed the adsorption energy into its single energy contributions. We name P1 as the β-strand bound to the surface via Ti–O dative bond and P2 as the β-strand bound via dispersion forces. Accordingly, we have calculated $\Delta E1$, *i.e.* the interaction of P1 with the surface, $\Delta E2$, *i.e.* the interaction of P2 with the surface, and ΔE_{Hb}, *i.e.* the energy of the H-bond pattern. All data are presented in Table 6.1. As expected ΔE_{Hb} is larger for the antiparallel β-sheet than for the parallel one, and $\Delta E1$ is larger than $\Delta E2$. Moreover, as ΔE_{int} is higher ΔE_{calc} it is clear that there is a negative cooperativeness, *i.e.* competitiveness, among the three body.

In view of the large ΔE_{ADS}/AA unit shown by the linear/TiO_2 complex due to the Ti-O dative bonds (see above), we have carried out MD simulations on both β-sheet/TiO_2 complexes to check if transformation from the β-sheets to the linear structure takes place by rotation of the two strands with respect to the normal surface, enabling a new Ti–O dative bond in the P2 strand. However, in both β-sheet_AP/TiO-2 and β-sheet_P/TiO_2, the systems prefer to maintain the H-bond patterns rather than establishing additional Ti–O bonds. In view of that, from the β-sheet_AP/TiO_2 complex, we have manually built a structure in which the peptide adopts a primary-like conformation by rotating the two β-strands (see Figure 6.9). The op-

Table 6.1: Energy terms for the interaction of β-sheet structures with the (101) anatase surface. $\Delta E1$ is the interaction between P1 and S (the strand where Ti–O bonds are present), $\Delta E2$ is the interaction between P2 and S (the strand with disperions forces), ΔE_{Hb} is the interaction energy of the H-bond pattern. ΔE_{int} is the interaction energy calculated as $\Delta E_{int} = E_{CPLX} - E_{P1}//CPLX - E_{P2}//CPLX - E_S//CPLX$. ΔE_{calc} is the sum of all the contribution to the interaction energy, *i.e.* $\Delta E1 + \Delta E2 + \Delta E_{Hb}$ (see the Computational Details section for more details). $|\Delta E|$ is the difference in energy between ΔE_{calc} and ΔE_{int}. Units are in kJ mol^{-1}.

System	$\Delta E1$	$\Delta E2$	ΔE_{Hb}	ΔE_{int}	ΔE_{calc}	$\lvert\Delta E\rvert$
β-sheet_AP/TiO_2	-111.1	-35.1	-79.6	-201.6	-225.8	24.2
β-sheet_P/TiO_2	-113.7	-33.2	-73.1	-197.0	-220.0	23.1

timised resulting structure, hereafter named as β-sheet_BRK/TiO_2, is slightly more favourable with respect to β-sheet_AP/TiO_2 and β-sheet_P/TiO_2 (ΔE_{ADS} = -123.2 kJ mol^{-1} and ΔE_{ADS}/AA unit = -30.8 kJ mol^{-1}). From this β-sheet_BRK/TiO_2 structure, we have also executed AIMD simulations to assess the evolution of this structure with time (see Figure 5). Following the evolution of the O–H and Ti–O, since the first step of the simulation, the two strands rotate to restore β-sheet_AP conformation. Therefore, we can conclude that, despite that potential energy values indicate the β-sheet_AP/TiO_2 complex less stable than the β-sheet_BRK/TiO_2 one, when thermal and entropic effects are accounted for, the antiparallel β-sheet structure adsorbed on the surface is found to be the most stable complex. The inclusion of dynamic effects are shown here to be crucial to elucidate this phenomenon.

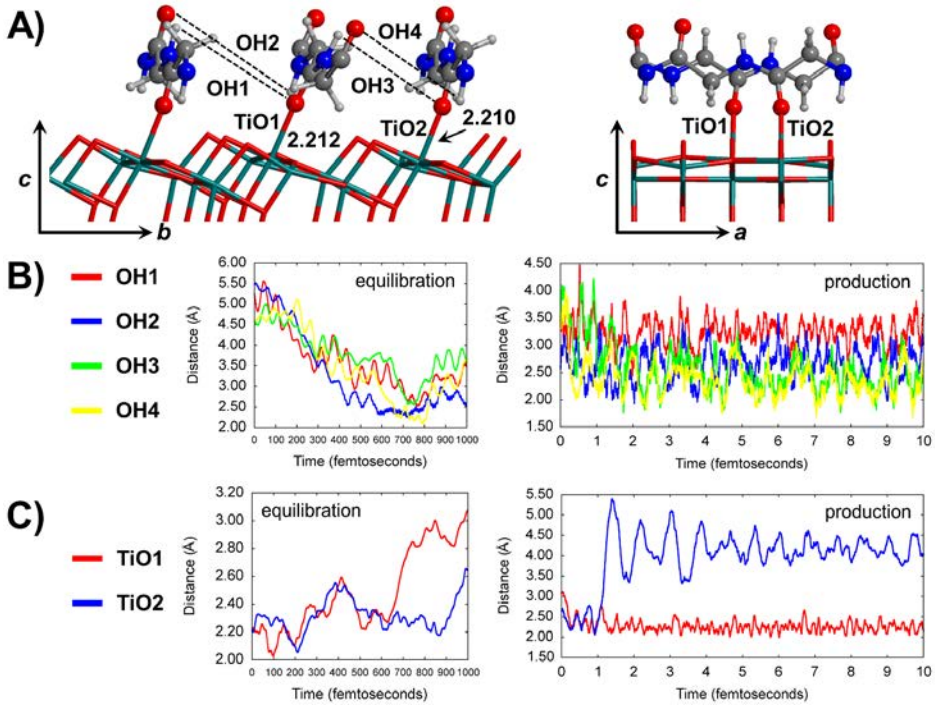


Figure 6.9: PBE-D2* optimised structure of the β-sheet_AP/TiO_2 complex (A), and evolution of the H-bond distances (B) and of the Ti-O distances (C) along the AIMD simulations. The equilibration and production periods are separated. Bond distances are in Å.

6.3.3 α-helix

The other very common secondary structure found in proteins is the α-helix (see Figure 6.4). The α-helix structure presents typically $i + 4 \rightarrow i$ H-bonds; that is, the N–H group of an amino acid forms a H-bond with the C=O group of the amino acid four residues earlier. Figure 6.10 shows the optimised structure of the α-helix on the (101) anatase surface (hereafter referred to as α-helix_1/TiO_2). As one can see in this complex, the periodicity along the a direction of the surface unit cell has been used to model an infinite α-helix. By comparing the free α-helix with that adsorbed on the surface the main structural change is the breaking of one H-bond due to the formation of a Ti–O bond. The rest of H-bonds are not significantly altered due to the surface interaction, being around 1.86 – 1.87 Å. The total calculated adsorption energy is ΔE_{ADS} = -220.1 kJ mol^{-1}, which when normalized per AA unit it is ΔE_{ADS}/AA unit = -31.5 kJ mol^{-1}, very similar to the values obtained for the linear/TiO_2 and β-sheet_AP/TiO_2 complexes.

AIMD simulations have been carried out onto the optimised α-helix_1/TiO_2 complex to assess dynamic and temperature effects on the structures. During the simulations one can identify some denaturation of the peptide; *i.e.*, the α-helix structure is partly lost. Indeed, new Ti–O dative bonds are formed, at expenses to break the $i + 4 \rightarrow i$ H-bonds close to the surface. Due to that, the NH groups involved in the broken $i + 4 \rightarrow i$ H-bonds lead to form a weak $i + 3 \rightarrow i$ H-bonds; that is, the NH establishes a H-bond with a C=O group three residues earlier, with a H-bond distance of about 2.0 Å. The evolution of these and other structural parameters are shown in

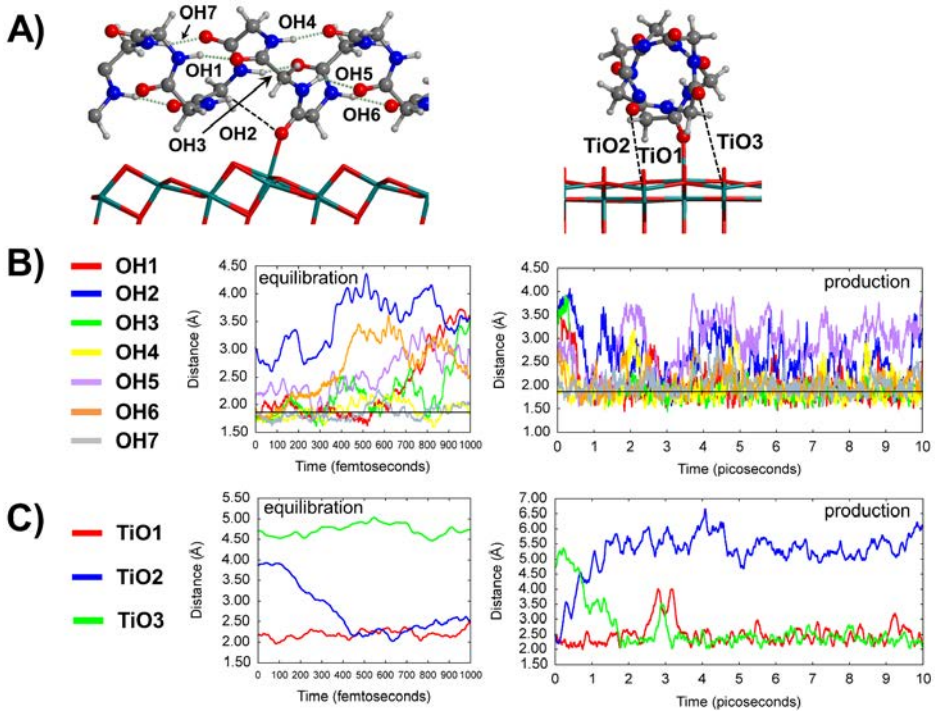


Figure 6.10: PBE-D2* optimised structure of the α-helix_1/TiO_2 complex (A), and evolution of the H-bond distances (B) and of the Ti-O distances (C) along the AIMD simulations. The equilibration and production periods are separated. Bond distances are in Å.

Figure 6.10. Because of the formation of this new Ti–O bond, we have chosen the last snapshot of the MD simulation as the initial structure for a new geometry optimization. The optimised structure is shown in Figure 6.11 (α-helix_2/TiO_2), which indeed presents the different Ti–O bonds and both $i + 4 \rightarrow i$ and $i + 3 \rightarrow i$ H bond patterns. The calculated adsorption energies of α-helix_2/TiO_2 are larger than α-helix_1/TiO_2 (ΔE_{ADS} = -236.1 kJ mol^{-1}, ΔE_{ADS}/AA unit = -33.7 kJ mol^{-1}) (see Table 6.2), indicating that this structure is more stable on the surface, although the helix structure is more distorted. Moreover, this system is the one having the most favourable ΔE_{ADS}/AA unit in comparison to the primary and β-sheet peptides, and accordingly, we can conclude that a polyglycine system adsorbed on the (101) anatase surface would prefer the helix conformation rather than the β-sheet folding.

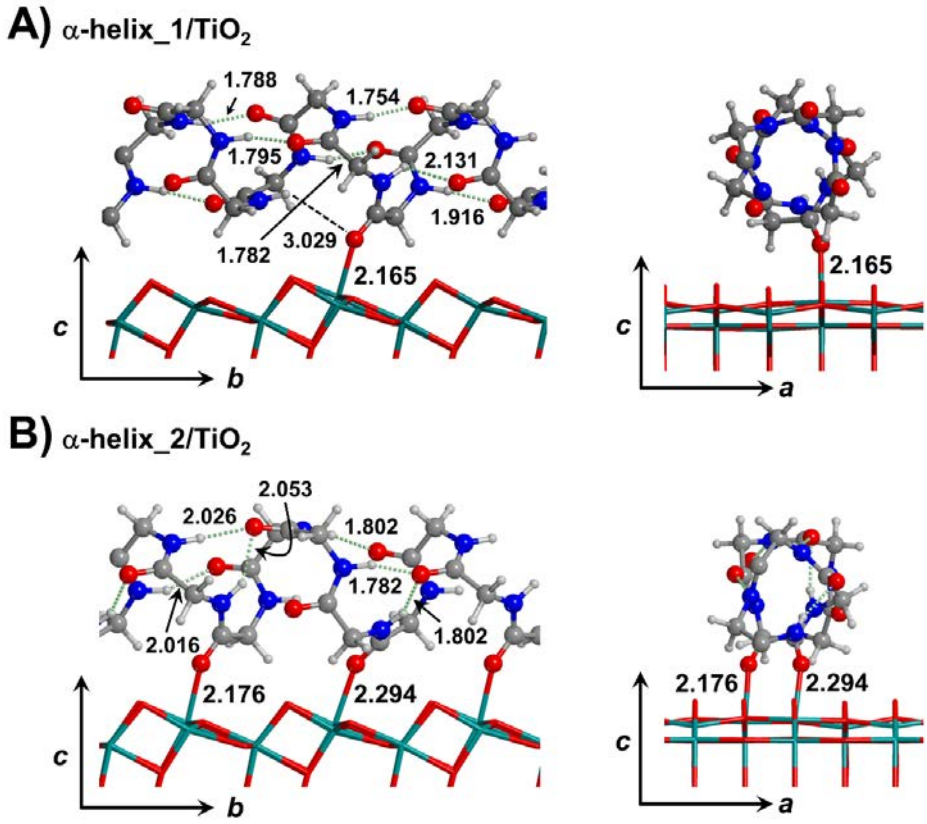


Figure 6.11: PBE-D2* optimized structures for the α-helix_1/TiO_2 (A) and α-helix_2/TiO_2 (B) complexes. Bond distances are in Å.

Table 6.2: Calculated total adsorption energies (ΔE_{ADS}) and normalized per amino acid unit (ΔE_{ADS}/AA unit) for each studied complex. ΔE_h/AA unit is the energetic gain due to the isolated secondary structure formation with respect to the isolated linear polymer (primary structure). Units are in kJ mol^{-1}

System	AA units	ΔE_{ADS}	ΔE_{ADS}/ AA unit	ΔE_h/ AA unit
linear/TiO_2	2	-62.0	-31.0	0.0
β-sheet_P/TiO_2	4	-115.6	-28.9	-18.1[a]
β-sheet_AP/TiO_2	4	-120.8	-30.2	-22.2[a]
β-sheet_BRK/TiO_2	4	-123.2	-30.8	0.0
α-helix_1/TiO_2	7	-220.5	-31.5	-21.5[b]
α-helix_2/TiO_2	7	-235.9	-33.7	-21.5[b]

[a] $\Delta E_h = (E_{\beta-sheet} - 2E_{linear})/4$

[b] $\Delta E_h = (E_{\alpha-helix} - 7/2E_{linear})/7$

6.3.4 CO frequency calculations

An interesting property to analyse, which can also be useful from an experimental point of view, is the stretching vibration of the peptide C=O bonds, ν(CO). Free peptide C=O groups present an intense stretching vibrational band at about 1600-1700 cm^{-1}. This region of the IR spectrum is usually clean so that ν(CO) IR bands are not overlapped by other vibrational frequencies. Because of that, this band is used as a reference mode frequency in C=O-containing molecules, and its blue/red shifts are used to elucidate the interactions with other species. In our systems, two main features can perturb the ν(CO) values: i) the H-bond interactions with N–H groups forming the secondary structures, and ii) the dative interactions with the pentacoordinated surface Ti atoms. It is worth mentioning that the PBE functional underestimates vibrational stretching frequencies, so that our values are shifted by -50/-60 cm^{-1} with respect to the experimental values. Table 6.3 reports the calculated ν(CO) values for all the polyglycine systems studied in this work, both in its isolated state and on the (101) anatase surface, as well as the frequency shifts ($\Delta\nu$) due to the interaction with the surface.

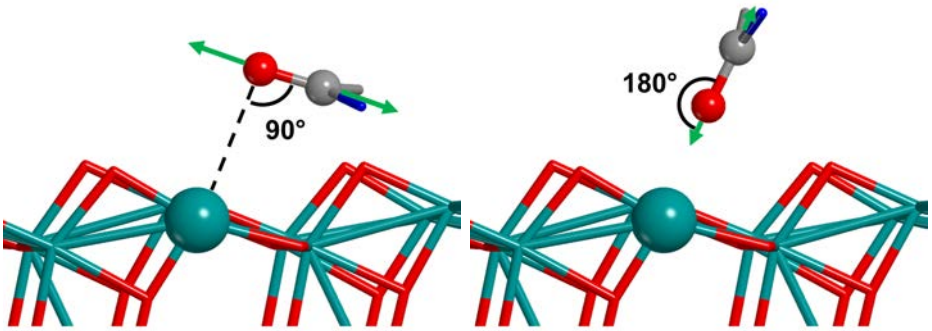
In the isolated linear polyglyicne structure, calculated ν(CO) are 1655 and 1622 cm^{-1}. This linear peptide has two independent C=O groups (one per Gly unit in the unit cell) but the ν(CO) motions are coupled since the first and second ν(CO) values correspond to the antisymmetric and symmetric stretching modes, respectively. As expected, on the TiO_2 surface, these values are red-shifted to 1599 and 1619 cm^{-1} due to the Ti–O bonds; as in the isolated polyglycine structure, the two values are coupled, giving a symmetric (1599 cm^{-1})

and an antisymmetric (1619 cm^{-1}) frequency.

In the isolated β-sheet structures, ν(CO) values are red-shifted with respect to the values of the linear primary structure due to the H-bond pattern between the β-strands. In the antiparallel conformation the ν(CO) range values is 1656 – 1578 cm^{-1}, while in the parallel one 1650 – 1592 cm^{-1}. The red shift is more accentuated in the antiparallel conformation due to its stronger H-bond interactions between strands. Experimentally β-sheet systems present two bands at 1685 and 1629 cm^{-1}, in reasonable agreement with the computed values. When the β-sheets are adsorbed on the surface, a red-shift of all the ν(CO) modes is observed. The red shift is particularly important in those C=O groups which stablish Ti–O bonds with the surface; *i.e.*, in β-sheet_AP/TiO_2, the frequency at 1578 cm^{-1} moves to 1547 cm^{-1} ($\Delta\nu = 31$ cm^{-1}) and in β-sheet_AP/TiO_2 the frequency at 1592 cm^{-1} moves to 1544 cm^{-1} ($\Delta\nu = 48$ cm^{-1}). Also for the β-sheet_BRK/TiO_2 complex a red-shift is observed with respect to the linear polyglycine system. In this complex, two band groups can be established: i) the bands at 1587 and 1583 cm^{-1}, attributed to the C=O groups non interacting with the surface, and ii) the bands at 1616 and 1609 cm^{-1}, attributed to the C=O groups interacting with the Ti surfaces. These latter values indicate an inversion with respect to what is observed in the previous β-sheet/TiO_2 systems. The reason is the relative orientation of the polyglycine systems with respect to the surface. In β-sheet_AP/TiO_2 and β-sheet_P/TiO_2 structures, the carbonyl groups interact with Ti atoms with C-O-Ti angles of about 147 degrees (33 degrees with respect to the normal of the surface). In β-sheet_BRK/TiO_2, the angles are about 170 - 173º (10 - 7 degrees with respect to the surface normal). That is, the β-sheet_BRK

is almost perpendicularly adsorbed on the surface, while the other two structures are adsorbed in a more parallel fashion. In both cases there is a σ-donation from the HOMO (highest occupied molecular orbital) localized on the O atom of the C=O bond to d-orbitals of the Ti atoms of the surface, thus producing a weakening of the C=O bond itself. However in the former case O atoms point directly to Ti atoms, and the wall effect (the repulsive potential due to the vibration against a rigid surface) becomes dominant, overcoming the σ-donation, thus causing a blue-shift of the C=O stretching. Notably, in the case of linear/TiO_2 this phenomenon is not observed, probably due to the coupling of the frequency of the two C=O groups. The hypothesis is that the presence of a second linear polyglycine (β-sheet_BRK/TiO_2) decouples the frequencies of the C=O because of the interactions between the two polyglycine strands. A schematic representation of the wall effect is given in Figure 6.12.

Finally, in the isolated α-helix system, an overall red-shift of the ν(CO) values compared to the isolated primary structure is observed, due to the H-bond pattern defining the helical structure. In α-helix_1/TiO_2, the C=O group involved in the Ti-O bonds does indeed decrease its ν(CO) value (from 1611 to 1558 cm^{-1}, $\Delta\nu$ = -53 cm^{-1}). However, for the rest of the ν(CO) vibrations one can observe both red-shifts and blue shifts. We attribute this behaviour to the breaking of the homogeneous α-helical content upon adsorption. In the isolated α-helix, the H-bond distances range from 1.866 to 1.872 Å, whereas on the surface the range increase from 1.754 to 1.916 Å. Therefore, some H-bonds become stronger while others weaker, and hence the red- and blueshifts. For instance, the highest mode ν(CO) = 1679 cm^{-1} does indeed correspond to the longest H-bond (1.916



(a) Negligible wall effect (b) Appreciable wall effect

Figure 6.12: Schematic representation of the wall effect phenomenon. The angle O-C-Ti is marked. The green arrows represent the vibration of the C=O group.

Å). In the α-helix_2/TiO_2 system, the ν(CO) values of the C=O involved in the Ti-O bonds are similar to those C=O groups involved in simultaneous i + 4 $\rightarrow$ i and i + 3 $\rightarrow$ i H-bond patterns (1599 and 1601 cm^{-1}, respectively). Therefore, the electronic effect on the peptide C=O bond when it interacts with a surface Ti atoms is similar as when it is involved in the two H-bond patterns. The highest ν(CO) normal mode (1662 cm^{-1}) regards the only C=O not involved in Ti–O bonds nor H-bond with N–H groups (see Figure 6.11B).

Table 6.3: PBE-D2* vibrational C=O stretching for the systems studied. Units are in cm^{-1}.

System	Isolated system	on TiO_2
linear	1622	1599
	1655	1619
β-sheet_AP	1578	1547
	1595	1574
	1621	1613
	1656	1623
β-sheet_P	1592	1544
	1602	1573
	1631	1608
	1650	1626
β-sheet_BRK	–	1583
	–	1587
	–	1609
	–	1616
α-helix_1	1611	1558
	1612	1608
	1621	1615
	1624	1624
	1625	1640
	1646	1651
	1647	1679
α-helix_2	1611	1599
	1612	1601
	1621	1616
	1624	1621
	1625	1625
	1646	1648
	1647	1662

6.4 Final remarks

In this Chapter, the adsorption and stability of secondary structures (linear polymer, α-helix and β-sheet) adsorbed on the (101) TiO_2 anatase surface has been theoretically studied with PBE-D2* periodic simulations. For a proper comparison between the adsorption of different polypeptides, the computed adsorption energies have been normalized per AA unit (*i.e.* ΔE_{ADS}/AA unit, the total adsorption energy for each structure has been divided for the number of amino acidic residues in the unit cell). The academic case of the adsorption of the linear polymer indicates that the peptide C=O groups are prone to form covalent dative bonds with coordinatively unsaturated Ti atoms placed at the outermost positions of the surface. For the secondary structures the adsorption presents a delicate trade-off between keeping the H-bond pattern of the secondary structures and formation of Ti–O bonds.

Among the different considered complexes, the α-helix_2/TiO_2 structure has been found as the most stable one (ΔE_{ADS}/AA unit = -33.7 kJ mol^{-1}). This structure presents the peptide C=O groups close to the surface forming Ti–O bonds, thus rendering the helical content partly distorted. This structure has been identified by running AIMD simulations, in which the formation of the Ti–O bonds is observed during the simulation evolution. For the adsorption of the β-sheet structures, considering only the potential energy values, the β-sheet_BRK/TiO_2 complex is the most stable (ΔE_{ADS}/AA unit = -30.8 kJ mol^{-1}), closely followed by β-sheet_AP/TiO_2 (ΔE_{ADS}/AA unit = -30.2 kJ mol^{-1}). The difference between these two complexes is that the former presents two Ti–O bonds at expenses to break

several H-bond interactions, breaking in part the β-sheet structure, while in the latter only one Ti–O bonds is present, thus keeping the β-sheet structure. Irrespective of the tiny energetic difference between these two complexes, AIMD simulations taking β-sheet_BRK/TiO_2 as initial guess structure show clearly the conversion from this complex to β-sheet_AP/TiO_2, thus indicating that β-sheet_BRK/TiO_2 is not stable at 300 K in favour of the β-sheet_AP/TiO_2 complex.

Finally, the peptide C=O vibrational stretching modes have been calculated for both the primary and secondary structures both in their isolated and adsorbed states. The red-/blue-shifts undergone by the ν(CO) as a consequence of the adsorption have been analysed and interpreted from an atomistic perspective. These vibrational data is of relevance from an experimental point of view as they can be useful to elucidate the TiO_2-induced peptide folding.

Chapter 7

Conclusions

In the present thesis the adsorption of amino acids and secondary structures, and the mechanism of peptide bond formation on TiO_2 surfaces has been studied by means of QM periodic simulations. The adsorption of several amino acids, in order to explore all the types of side chains, has been analysed both on (101) anatase and on (110) rutile. Moreover, the mechanism of peptide bond formation between two glycine molecules has been studied in dry conditions (Gly + Gly), in moderately wet conditions (Gly + Gly + $n H_2O$, $n = 1, 2$), and with a high concentration of Gly (Gly + Gly + Gly, where the third Gly assists the proton trasnfer) on the (101) anatase surface, in order to evaluate the effect of TiO_2 in catalysing this process with respect to the gas phase. Finally the static and dynamic properties of secondary structures (α-helix and β-sheets) has been simulated on (101) anatase.

Amino acids adsorption

Both (101) anatase and (110) rutile surfaces present high adsorption energies for all the amino acids analysed, because all the amino acids present al least one, and, in the majority of the cases two, strong dative bonds between the backbone and the surface. Lateral chains may exhibit either dispersion interactions (Leu, Phe, Met), or H-bonds (Ser, Cys), or additional dative bonds (Lys, Arg, His) with the surface. An important difference between the two surfaces studied is their stability with respect to the bulk structures, *i.e.* their reactivity, which has direct consequences on the adsorbed amino acids. The first one is the higher tendency of the (110) rutile in deprotonating the amino acids, being higher its surface energy with respect to (101) anatase. The second consequence of the higher reactivity of the (110) rutile is the higher adsorption energy of the adsorbed amino acids.

An other important difference between the two surfaces is the morphology, which also drives the adsorption mode of the amino acids: on anatase the preferred adsorption mode of the backbone is (N,O), while on rutile is (O,O). This is mainly due to different Ti–Ti distances of the outermost atoms of the surface.

Peptide bond formation mechanism

By comparing the uncatalysed reactions in gas phase and on the surface, in all the cases a sensible reduction of the activation barrier has been observed. In gas phase the reaction may exhibit both a concerted and a stepwise mechanism; in contrast, with the presence of the surface the reaction proceeds through a stepwise mechanism, where the nucleophilic attack and the release of water occur in two

separate steps. By performing a Bader charge analysis, it is clear that the surface modify the electronic structure of the adsorbed amino acids; in particular an increase of the electrofilicity of the carboxyl C is observed and, accordingly, the nucleophilic attack of the second amino acids is more efficient, and the barrier drops down by $\sim$6 kcal mol^{-1} for the non assisted reaction. When the assistance of the proton transfer by small molecules is taken into account the reaction barrier is further reduced, in particular considering the presence of a third glycine as a proton transfer assistant. These simulations give a reasonable explanation, from an atomistic point of view, of the Martra's experiment.[25]

Adsorption of secondary structures

Finally, the adsorption of secondary structure on the anatase surface has been studied, starting from the academic case of a linear polyglycine systems, and then analysing the most common secondary structures (α-helix and β-sheet). The α-helix is the preferred conformation with respect to the linear polymer and the β-sheets, thus confirming that mineral surfaces may induce the protein folding.

It has also been demonstrated that MD simulation are mandatory for these systems, particularly in the case of the β-sheet_BRK/TiO_2, *i.e.* full coverage of the anatase surface with linear glycine polymers. Despite the potential energy is lower for the linear conformation when it is adsorbed on the surface, the introduction of thermal effects and a sufficiently long MD simulation cause the rotation of the polymers (with the consequent cleavage of strong interactions with the surface), in order to restore the β-sheet conformation. Moreover the

MD simulation on the α-helix_1/TiO_2 system allows us to find a more stable structure, *i.e.* the α-helix_2/TiO_2 complex, which is slightly "denaturated" with respect the isolated α-helix.

Future perspectives

The work done in this thesis has opened some interesting new projects, in particular regarding more complex secondary structures. The idea is to analyse not only polyglycine systems but, as we have done for the amino acids, also others polyamino acidic structures to evaluate the effect of several side chains; the work in progress regards polylysine, polyglutamic acid and polyalanine systems, isolated and adsorbed on the anatase surface. Moreover, we are also studying β-sheets with a finite number of strands, *i.e.* with 1D periodicity, with the aim of studying the β-sheet growing on the anatase surface, in order to give some insights from an atomistic point of view on the β-amyloid fibrillation, which is a theme of great important in medicine, as the aggregation of many β-amyloids produces the formation of plaques, which are one of the main causes of the Alzheimer's disease.[291, 292]

Bibliography

[1] J. D. Bernal. The Physical Basis of Life. *Proc. Phys. Soc. A*, 62(9):537–558, 1949.

[2] J.-M. Lehn. Perspectives in Chemistry - Steps Towards Complex Matter. *Angew. Chem. Int. Ed.*, 52:2836–2850, 2013.

[3] J.-M. Lehn. Toward Self-Organization and Complex Matter. *Science*, 295:2400–2403, 2002.

[4] M. Eigen. Selforganization of Matter and the Evolution of Biological Macromolecules. *Naturwissenschaften*, 58:465–523, 1971.

[5] K. R-Mirazo, C. Briones, and A. de la Escosura. Prebiotic Systems Chemistry: New Perspectives for the Origin of Life. *Chem. Rev.*, 114:285–366, 2014.

[6] E. Parrilli, F. Sannino, G. Marino, and M. L. Tutino. Life in Icy Habitats: New Insights Supporting Panspermia Theory. *Rend. Fis. Acc. Lincei*, 22:375–383, 2011.

[7] E. Parrilli, F. Sannino, G. Marino, and M. L. Tutino. Life in Icy Habitats: New Insights Supporting Panspermia Theory. *Rend. Fis. Acc. Lincei*, 22:375–383, 2011.

[8] S. Tirard. J. B. S. Haldane and the Origin of Life. *J. Genet.*, 69:735–739, 2017.

[9] A. Lazcano. The Origin and Early Evolution of Life: Where, When and How? *Evo. Edu. Outreach*, 5:334–336, 2012.

[10] J. Peretó, J. L. Bada, and A. Lazcano. Charles Darwin and the Origin of Life. *Orig. Life Evol. Biosph.*, 39:395–406, 2009.

[11] J. D. Bernal. *The Origin of Life. Translation by Oparin by Ann Synge.* Weidenfeld & Nicolson, 1967.

[12] S. L. Miller, J. W. Schopf, and A. Lazcano. Oparin's "Origin of Life": Sixty Years Later. *J. Mol. Evol.*, 44:351–353, 1997.

[13] S L. Miller. A Production of Amino Acids under Possible Primitive Earth Conditions. *Science*, 117:528–529, 1952.

[14] S. L. Miller. Production of Some Organic Compounds under Possible Primitive Earth Conditions. *J. Am. Chem. Soc.*, 77:2351–2361, 1955.

[15] S. L. Miller and H. C. Urey. Organic Compound Synthesis on the Primitive Earth. *Science*, 130:245–251, 1959.

[16] J. Orò. Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature*, 191:1193–1194, 1961.

[17] J. P. Ferris, R. A. Sanchez, and L. E. Orgel. Studies n Prebiotic Synthesis III. Synthesis of Pyrimidines from Cyanoacetylene and Cyanate. *J. Mol. Biol.*, 33:693–704, 1968.

[18] A. Butlerow. Formation synthétique d'une substance sucrée. *C. R. Acad. Sci.*, 53:145–147, 1861.

[19] R. M. Hazen. Mineral Surfaces and the Prebiotic Selection and Organization of Biomolecules. *Am. Mineral.*, 91:1715–1729, 2006.

[20] R. M. Hazen, D. Papineau, W. Bleeker, R. T. Downs, J. M. Ferry, T. J. McCoy, D. A. Sverjensky, and H. Yang. Mineral Evolution. *Am. Mineral.*, 93:1693–1720, 2008.

[21] R. M. Hazen and D. A. Sverjensky. Mineral Surfaces, Geochemical Complexities, and the Origins of Life. *Cold Spring Harb. Perspect. Biol.*, 2:a002162, 2010.

[22] J. F. Lambert. Adsorption and Polymerization of Amino Acids on Mineral Surfaces: A Review. *Orig. Life Evol. Biosph.*, 38:211–242, 2008.

[23] D. A. M. Zaia. A Review of Adsorption of Amino Acids on Minerals: Was it Important for the Origin of Life? *Amino Acids*, 27:113–118, 2004.

[24] C. Deiana, Y. Sakhno, M. Fabbiani, M. Pazzi, M. Vincenti, and G. Martra. Direct Synthesis of Amides from Carboxylic Acids and Amines by Using Heterogeneous Catalysts: Evidence of Surface Carboxylates as Activated Species. *ChemCatChem*, 5:2832–2834, 2013.

[25] G. Martra, C. Deiana, Y. Sakhno, I. Barberis, M. Fabbiani, M. Pazzi, and M. Vincenti. The Formation and Self-Assembly of Long Prebiotic Oligomers Produced by the Condensation of Unactivated Amino Acids on Oxide Surfaces. *Angew. Chem. Int. Ed.*, 53:4671–4674, 2014.

[26] M. Wilhelm, J. Schlegl, H. Hahne, A. M. Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Burtzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J. H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber, and B. Kuster. Mass-Spectrometry-Based Draft of the Human Proteome. *Nature*, 509:582–587, 2014.

[27] M. S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madungundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. N. Selvan, A. H. Patil, V. Nanjappa, A. Radjakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel,

A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T. C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. K. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda, and A. Pandey. A Draft Map of the Human Proteome. *Nature*, 509:575–581, 2014.

[28] P-Hudáky and A. Perczel. Comformation Dependence of pK_a: Ab Initio and DFT Investigation of Histidine. *J. Phys. Chem. A*, 108:6195–6205, 2004.

[29] J. Rak, P. Skurski, J. Simons, and M. Gutowski. Low-Energy Tautomers and Conformers of Neutral and Protonated Arginine. *J. Am. Chem. Soc.*, 123:11695–11707, 2001.

[30] D. Fass and C. Thorpe. Chemistry and Enzymology of Disulfide Cross-Linking in Proteins. *Chem. Rev.*, 118:1169–1198, 2018.

[31] J. P. Vincent, R. Chicheportiche, and M. Lazdunski. The Conformational Properties of the Basic Pancreatic Trypsin-Inhibitor. *Eur. J. Biochem.*, 23:401–411, 1971.

[32] A. Radzicka and R. Wolfenden. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *J. Am. Chem. Soc.*, 118:6105–6109, 1996.

[33] J. H. Sinfelt. Role of Surface Science in Catalysis. *Surf. Sci.*, 500:923–946, 2002.

[34] R. Schlögl. Catalytic Synthesis of Amonia - A "Never-Ending Story". *Angew. Chem. Int. Ed.*, 42:2004–2008, 2003.

[35] G. Ertl, M Weiss, and S. B. Lee. Reprint of: The Role of Potassium in the Catalytic Synthesis of Amonia. *Chem. Phys. Lett.*, 589:18–20, 2013.

[36] C. G. Vayenas and H. M. Saltsburg. The SO_2 Oxidation on Noble Metals. *J. Catal.*, 57:296–314, 1979.

[37] L. B. Hunt. The Ammonia Oxidation Process for Nitric Acid Manufacture. *Platin. Met. Rev.*, 2 (4):129–134, 1958.

[38] D. G. Castner. Biomedical Surface Science: Foundations to Frontiers. *Surf. Sci.*, 500:28–60, 2002.

[39] C. C. Perry, S. V. Patwardhan, and O. Deschaume. From Biominerals to Biomaterials: the Role of Biomolecule-Mineral Interactions. *Biochem. Soc. Trans.*, 37:687–691, 2009.

[40] L. L. Hench, R. J. Splinter, W. C. Allen, and T. K. Greenlee. Bonding Mechanisms at the Interface of Ceramic Prosthetic Materials. *J. Biomed. Mater. Res.*, 2(1):117–141, 1971.

[41] Javier Navarro Ruiz. *Theoretical Investigation of H_2 Formation on Interstellar Silicates Surfaces.* Phd thesis, Universitat Autònoma de Barcelona, 2015.

[42] J. Navarro-Ruiz, P. Ugliengo, M. Sodupe, and A. Does Fe^{2+} in Olivine-Based Interstellar Grains Play Any Role in the Formation of H_2? Atomistic Insights from DFT Periodic Simulations. *Chem. Commun.*, 52:6873–6876, 2016.

[43] J. Navarro-Ruiz, J. A. Martínez-González, M. Sodupe, P. Ugliengo, and A. Rimola. Relevance of Silicate Surface Morphology in Interstellar H_2 Formation. Insights from Quantum Chemical Calculations. *Mon. Not. R. Astron. Soc.*, 453:914.924, 2015.

[44] J. Navarro-Ruiz, M. Sodupe, P. Ugliengo, and A. Rimola. Interstellar H Adsorption and H_2 Formarion on the Crystalline (010) Forsterite Surface: a B3LYP-D2* Periodic Study. *Phys. Chem. Chem. Phys.*, 16:17447–17457, 2014.

[45] A. Stirling, T. Rozgonyi, M. Krack, and M. Bernasconi. Prebiotic NH_3 Formation: Insights from Simulations. *Inorg. Chem.*, 55:1934–1939, 2016.

[46] Robert M. Hazen. *Genesis - The scientific quest for life's origin.* Joseph Henry Press, 2005.

[47] C. Tanford. The Hydrophobic Effect and the Organization of Living Matter. *Science*, 200:1012–1018, 1978.

[48] D. W. Deamer and R. M. Pashley. Amphiphilic Components of the Murchison Carbonaceous Chondrite: Surface Properties and Membrane Formation. *Origins Life Evol. B*, 19:21–38, 1989.

[49] D. Segré, D. Ben-Eli, D. W. Deamer, and D. Lancet. The Lipid Water. *Origins Life Evol. B*, 31:119–145, 2001.

[50] P. L. Luisi and F. J. Varela. Self-Replicating Micelles - A Chemical Version of a Minimal Autopoietic System. *Origins Life Evol. B*, 19:633–643, 1989.

[51] P. A. Bachmann, P. L. Luisi, and J. Lang. Autocatalytic Self-Replicating Micelles as Models for Prebiotic Structures. *Nature*, 357:57–59, 1992.

[52] R. M. Hazen. Mineral Surfaces and the Prebiotic Selection and Organization of Biomolecules. *Am. Mineral.*, 91:1715–1729, 2006.

[53] S. J. Sowerby, W. H. Heckl, and G. B. Petersen. Chiral symmetry breaking during the self-assembly of monolayers from achiral purine molecules. *J. Mol. Evol.*, 43:419–424, 1996.

[54] S. J. Sowerby, M. Edelwirth, and W. M. Heckl. Self-Assembly at the Prebiotic Solid-Liquid Interface: Structures of Self-Assembled Monolayers of Adenine and Guanine Bases Formed on Inorganic Surfaces. *J. Phys. Chem.*, 102:5914–5922, 1998.

[55] S. J. Sowerby, M. Edelwirth, M. Reiter, and W. M. Heckl. Scanning Tunneling Microscopy Image Contrast as a Function of Scan Angle in Hydrogen-Bonded Self-Assembled Monolayers. *Langmuir*, 14:5195–5202, 1998.

[56] S. J. Sowerby, G. B. Petersen, and N. G. Holm. Primordial Coding of Amino Acids by Adsorbed Purine Bases. *Orig. Life Evol. Biosph.*, 32:35–46, 2002.

[57] T. Uchihashi, T. Okada, Y. Sugawara, K. Yokoyama, and S. Morita. Self-Assembled Monolayer of Adenine Base on Graphite Studied by Noncontact Atomic Force Microscopy. *Phys. Rev. B*, 60:8309–8313, 1999.

[58] N. Lahav, D. White, and S. Chang. Peptide formation in the prebiotic era: thermal condensation of glycine in fluctuating clay environments. *Science*, 201:67–69, 1978.

[59] G. Ertem and J. P. Ferris. Template-Directed Synthesis Using the Heterogeneous Templates Produced by Montmorillonite Catalysis. A Possible Bridge Between the Prebiotic and RNA Worlds. *J. Am. Chem. Soc.*, 119:7197–7201, 1997.

[60] M. M. Hanczyc, S. M. Fujikawa, and K. W. Szostak. Experimental Models of Primitive Cellular Compartments: Encapsulation, Growth, and Division. *Science*, 302:618–622, 2003.

[61] J. P. Ferris. Mineral Catalysis and Prebiotic Synthesis: Montmorillonite-Catalyzed Formation of RNA. *Elements*, 1:145–149, 2005.

[62] J. P. Ferris. Oligomerization of ribonucleotides on montmorillonite: reaction of the 5'-phosphorimidazolide of adenosine. *Science*, 257:1387–1389, 1992.

[63] J. P. Ferris. Montmorillonite catalysis of RNA oligomer formation in aqueous solution. A model for the prebiotic formation of RNA. *J. Am. Chem. Soc.*, 115:12279–12275, 1993.

[64] J. P. Ferris, A. R. Hill Jr, R. Liu, and L. E. Orgel. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381:59–61, 1996.

[65] G. Ertem and J. P. Ferris. Synthesis of RNA oligomers on heterogeneous templates. *Nature*, 379:238–240, 1996.

[66] A. A. Yaroshevsky. Abundances of Chemical Elements in the Earth's Crust. *Geochem. Int.*, 44:48–55, 2006.

[67] J. Han and A. J. Brearley. Formation of TiO_2 Nanoparticles in a CAI-Like Object from an AOA in the ALHA 77307 CO3.0 Carbonaceous Chondrite. *Meteorit. Planet. Sci.*, 46:5190, 2011.

[68] A. Wang, K. Kuebler, B. Jolliff, and L. A. Haskin. Mineralogy of a Martian meteorite as determined by Raman spectroscopy. *J. Raman Spectrosc.*, 35:504–514, 2004.

[69] L. Zücher and D. A. Kring. Hydrothermal alteration in the core of the Yaxcopoil-1 borehole, Chicxulub impact structure, Mexico. *Meteorit. Planet. Sci.*, 115:1199–1221, 11156-11162.

[70] T. Zhu and S.P. Gao. The Stability, Electronic Structure, and Optical Property of TiO_2 Polymorphs. *J. Phys. Chem. C*, 14:11385–11396, 2014.

[71] A. Fujishima and K. Honda. Electrochemical Photolysis of Water at a Semiconductor Electrode. *Nature*, 238:37–38, 1972.

[72] V. E. Heinrich, G. Dresselhaus, and H. J. Zeiger. Observation of Two-Dimensional Phases Associated with Defect States on the Surface of TiO_2. *Phys. Rev. Lett.*, 36:1335–1339, 1976.

[73] W. J. Lo, Y. W. Chung, and G. A. Somorjai. Electron spectroscopy studies of the chemisorption of O_2, H_2 and H_2O on the $TiO_2(100)$ surfaces with varied stoichiometry: Evidence for the photogeneration of Ti^{+3} and for its importance in chemisorption. *Surf. Sci.*, 71:199–219, 1978.

[74] B. O'Regan and M. Grätzel. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films. *Nature*, 353:737–740, 1991.

[75] K. Rajeshwar. Materials aspects of photoelectrochemical energy conversion. *J. Appl. Electrochem.*, 15:1–22, 1985.

[76] A. Mills, R. H. Davies, and D. Worsley. Water Purification by Semiconductor Photocatalysis. *Chem. Soc. Rev.*, 22:417–425, 1993.

[77] P.-C. Maness, S. Smolinski, D. M. Blake, Z. Huang, E. J. Wolfrum, and W. A. Jacoby. Bactericidal Activity of Photocatalytic TiO_2 Reaction: toward an Understanding of Its Killing Mechanism. *Appl. Environ. Microb.*, 65:4094–4098, 1999.

[78] Y. Paz, Z. Luo, L Rabenberg, and A. Heller. Photooxidative self-cleaning transparent titanium dioxide films on glass. *J. Mater. Res.*, 10:2842–2848, 1995.

[79] I. Poulios, P. Spathis, A. Grigoriadou, K. Delidou, and P. Tsoumparis. Protection of marbles against corrosion and microbial corrosion with TiO_2 coatings. *J. Environ. Sci Health*, 34:1455–1471, 1999.

[80] K. S. Brammer adn C. J. Frandsen and S. Jin. TiO_2 nanotubes for bone regeneration. *Trends Biotechnol.*, 30:315–322, 2012.

[81] L. M. Bjursten, L. Rasmusson, S. Oh, G. C. Smith, and K. S. Brammer. Titanium dioxide nanotubes enhance bone bonding in vivo. *J. Biomed. Mater. Res.*, 92A:1218–1224, 2010.

[82] N. K. Awad, S. L. Edwards, and Y. S. Morsi. A review of TiO_2 NTs on Ti metal: Electrochemical synthesis, functionalization and potential use as bone implants. *Mat. Sci. Enf. C-Mater.*, 76:1401–1412, 2017.

[83] U. Diebold. The Surface Science of Titanium Oxide. *Surf. Sci. Rep.*, 48:53–229, 2003.

[84] Marco Fabbiani. *Physical and Chemical Aspects of the Interactions of Molecules with External Surface and Structural Cavities of Nanomaterials*. Phd thesis, University of Insubria, 2017.

[85] M. Fabbiani, M. Pazzi, M. Vincenti, G. Tabacchi, E. Fois, and G. Martra. Does the Abiotic Formation of Oligopeptides on TiO_2 Require Special

Catalytic Sites? Apparently Not. *J. Nanosci. Nanotechno.*, 18:5854–5857, 2018.

[86] K. S. K. Lin, Y. H. Tseng, Y. Mou, Y. C. Hsu, C. M. Yang, and J. C. C. Chan. Mechanistic Study of Apatite Formation on Bioactive Glass Surface Using ^{31}P Solid-State NMR Spectroscopy. *Chem. Mater.*, 17:4493–4501, 2005.

[87] I. V. de Keere, R. Willaert, A. Hubin, and J. Vereecken. Interaction of Human Plasma Fibrinogen with Commercially Pure Titanium as Studied with Atomic Force Microscopy and X-ray Photoelectron Spectroscopy. *Langmuir*, 24:1844–1852, 2008.

[88] S. Mizuno, H. Tochihara, Y. Matsumoto, and K. I. Tanaka. STM Observation of Restructured Cu (001) Surfaces Induced by Li Deposition. *Surf. Sci.*, 393:L69–L76, 1997.

[89] S. Griessl, M. Lackinger, M. Edelwirth, M. Hietschold, and W. M. Heckl. Self-Assembled Two-Dimensional Molecular Host-Guest Atchitectures From Trimesic Acid. *Single Mol.*, 3:25–31, 2002.

[90] L. Kampschulte, S. Griessl, W. M. Heckl, and M. Lackinger. Mediated Coadsorption at the Liquid-Solid Interface: Stabilization through Hydrogen Bonds. *J. Phys. Chem. B*, 109:14074–14078, 2005.

[91] G. A. Somorjai and J. Y. Park. Concepts, Instruments, and Model Systems that Enabled the Rapid Evolution of Surface Science. *Surf. Sci.*, 603:1293–1300, 2009.

[92] M. A. Van Hove. Atomic-Scale Structure: From Surfaces to Nanomaterials. *Surf. Sci.*, 603:1301–1305, 2009.

[93] C. Arrouvel, B. Diawara, D. Costa, and P. Marcus. DFT Periodic Study of the Adsorption of Glycine on the Anhydrous and Hydroxylated (0001) Surfaces of α-Alumina. *J. Phys. Chem. C*, 111:18164–18173, 2007.

[94] S. Irrera, D. Costa, and P. Marcus. DFT periodic study of adsorption of glycine on the (0 0 0 1) surface of zinc terminated ZnO. *J. Mol. Struc-Theochem*, 903:49–58, 2009.

[95] A. Rimola, M. Sodupe, S. Tosoni, B. Civalleri, and P. Ugliengo. Interaction of Glycine with Isolated Hydroxyl Groups at the Silica Surface: First Principles B3LYP Periodic Simulation. *Langmuir*, 22:6593–6604, 2006.

[96] A. Rimola, Y. Sakhno, L. Bertinetti, M. Lelli, G. Martra, and P. Ugliengo. Toward a Surface Science Model for Biology: Glycine Adsorption on Nanohydroxyapatite with Well-Defined Surfaces. *J. Phys. Chem. Lett.*, 2:1390–1394, 2011.

[97] G. Hong, H. Heinz, R. R. Naik, B. L. Farmer, and R. Pachter. Toward Understanding Amino Acid Adsorption at Metallic Interfaces: A Density Functional Theory Study. *ACS Appl. Mater. Inter.*, 1:388–392, 2009.

[98] R. Di Felice, A. Selloni, and E. Molinari. DFT Study of Cysteine Adsorption on Au(111). *J. Phys. Chem. B*, 107:1151–1156, 2003.

[99] M. Hoefling, F. Iori, S Corni, and K-E Gottschalk. Interaction of Amino Acids with the Au (111) Surface: Adsorption Free Energies from Molecular Dynamics Simulations. *Langmuir*, 26:8347–8351, 2010.

[100] A. Rimola, M. Sodupe, and P. Ugliengo. Affinity Scale for the Interaction of Amino Acids with Silica Surface. *J. Phys. Chem. C*, 113:5741–5750, 2009.

[101] H. Tavassoli Larijani, M. Darvish Ganji, and M. Jahanshahi. Trends of amino acid adsorption onto graphene and graphene oxide surfaces: a dispersion corrected DFT study. *RSC Adv.*, 5:92843–92857, 2015.

[102] S. Monti, V. Carravetta, W. Zhang, and J. Yang. Effects Due to Interadsorbate Interactions on the Dipeptide/TiO_2 Surface Binding Mechanism Investigated by Molecular Dynamics Simulations. *J. Phys. Chem. C*, 111:7765–7771, 2007.

[103] S. Monti, V. Carravetta, C. Battocchio, G. Iucci, and G. Polzonetti. Peptide/TiO_2 Surface Interactions: A Theoretical and Experimental Study on the Structure of Adsorbed ALA-GLU and ALA-LYS. *Langmuir*, 24:3205–3214, 2008.

[104] C. Wu, M. Chen, A. A. Skelton, P. T. Cummings, and T. Zheng. Adosrption of Arginine-Glycine-Aspartate Tripeptide onto Negatively Changed Rutile (110) Mediated by Cations: The Effect of Surface Hydroxylation. *ACS Appl. Mater. Interfaces*, 5:2567–2579, 2013.

[105] S. Monti, M. Alderighi, C. Duce, R. Solaro, and M. R. Tiné. Adsorption of Ionic Peptides on Inorganic Supports. *J. Phys. Chem. C*, 113:2433–2442, 2009.

[106] H. Chen, X. Su, K. G. Neoh, and W. S. Choe. Context-Dependent Adsorption Behavior of Cyclic and Linear Peptides on Metal Oxide Surfaces. *Langmuir*, 25:1588–1593, 2009.

[107] A. Rimola, M. Aschi, R. Orlando, and P. Ugliengo. Does Adsorption at Hydroxyapatite Surfaces Induce Peptide Folding? Insights from Large-Scale B3LYP Calculations. *J. Am. Chem. Soc.*, 134:10899–10910, 2012.

[108] G Wächterschäuser. Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci. USA*, 87:200–204, 1990.

[109] G. D. Cody, N. Z. Boctor, T. R. Filley, R. M. Hazen, J. H. Scott, A. Sharma, and H. S. Yoder Jr. Primordial Carbonylated Iron-Sulfur Compounds and the Synthesis of Pyruvate. *Science*, 289:1337–1340, 2000.

[110] A. Motta, M. P. Gaigeot, and D. Costa. AIMD Evidence of Inner Sphere Adsorption of Glycine on a Stepped (101) Boehmite AlOOH Surface. *J. Phys Chem. C*, 116:23418–23427, 2012.

[111] E. Escamilla-Roa, F. J. Huertas, A. Hernández-Laguna, and C. I. Sainz-Díaz. A DFT Study of the Adsorption of Glycine in the Interlayer Space of Montmorillonite. *Phys. Chem. Chem. Phys.*, 19:14961–14971, 2017.

[112] A. Rimola, M. Corno, C. M. Zicovich-Wilson, and P. Ugliengo. Ab Initio Modeling of Protein/Biomaterial Interactions: Glycine Adsorption at Hydroxyapatite Surfaces. *J. Am. Chem. Soc.*, 130:16181–16183, 2008.

[113] A. Rimola, M. Corno, C. M. Zicovich-Wilson, and P. Ugliengo. Ab initio modeling of protein/biomaterial interactions: competitive adsorption between glycine and water onto hydroxyapatite surfaces. *Phys. Chem. Chem. Phys.*, 11:9005–9007, 2009.

[114] A. Rimola, S. Tosoni, M. Sodupe, and P. Ugliengo. Peptide Bond Formation Activated by the Interplay of Lewis and Brønsted Catalysts. *Chem. Phys. Lett.*, 408:295–301, 2005.

[115] A. J. A. Aquino, D. Tunega, M. H. Gerzabek, and H. Lischka. Modeling Catalytic Effects of Clay Mineral Surfaces on Peptide Bond Formation. *J. Phys. Chem. B*, 108:10120–10130, 2004.

[116] A. Rimola, M. Sodupe, and P. Ugliengo. Aluminosilicate Surfaces as Promoters for Peptide Bond Formation: An Assesment of Bernal's Hypothesis by ab Initio Methods. *J. Am. Chem. Soc.*, 129:8333–8344, 2007.

[117] A. Rimola, P. Ugliengo, and M. Sodupe. Formation versus Hydrolysis of the Peptide Bond froma Quantum-Mechanical Viewpoint: The Role of Mineral Surfaces and Implications for the Origin of Life. *Int. J. Mol. Sci.*, 10:746–760, 2009.

[118] O. Phuakkong, K. Bobuatong, P. Pantu, B. Boekfa, M. Probst, and J. Limtrakul. Glycine Peptide Bond Formation Catalysed by Faujasite. *ChemPhysChem*, 12:2160–2168, 2011.

[119] N. Y. Dzade, A. Roldan, and N. H. de Leeuw. Surface and Shape Modification of Mackinawite (FeS) Nanocrystals by Cysteine Adsorption: A First-Principles DFT-D2 Study. *Phys. Chem. Chem. Phys.*, 18:32007–32020, 2016.

[120] V. Erastova, M. T. Degiacomi, D. G. Fraser, and H. C. Greenwell. Mineral Surface Chemistry Control for Origin of Prebiotic Peptides. *Nat. Commun.*, 8(2033):1–9, 2017.

[121] J. M. R. Muir and H. Idriss. Computational Study of Cysteine Interaction with the Rutile TiO_2 (110) Surface. *Surf. Sci.*, 617:60–67, 2013.

[122] G. J. Fleming, K. Adib, J. A. Rodriguez, M. A. Barteau, and H. Idriss. Proline Adsorption on TiO_2 (110) Single Crystal Surface: A Study by High Resolution Photoelectron Spectroscopy. *Surf. Sci.*, 601:5726–5731, 5731.

[123] T. Qiu and M. A. Barteau. STM Study of Glycine on TiO_2 (110) Single Crystal Surfaces. *J. Colloid Interf. Sci.*, 303:229–235, 2006.

[124] G. J. Fleming, K. Adib, J. A. Rodriguez, M. A. Barteau, J. W. White, and H. Idriss. The Adsorption and Reactions of the Amino Acid Proline on Rutile TiO_2 (110) Surfaces. *Surf. Sci.*, 602:2029–2038, 2008.

[125] R. Tonner. Adsorption of Proline and Glycine on the TiO_2 (110) Surface: A Density Functional Theory Study. *ChemPhysChem*, 11:1053–1061, 2010.

[126] A. G. Thomas, W. R. Flawell, C. P. Chatwin, A. R. Kumarasinghe, S. M. Rayner, P. F. Kirkham, D. Tsoutsou, T. K. Johal, and S. Patel. Adsorption of Phenylalanine on Single Crystal Rutile TiO_2 (110) Surface. *Surf. Sci.*, 601:3828–3832, 2007.

[127] L. Ojamäe, C. Aulin, H. Pedersen, and P. O. Käll. IR and Quantum-Chemical Studies of Carboxylic Acid and Glycine Adsorption on Rutile TiO_2 Nanoparticles. *J. Colloid Interf. Sci.*, 296:71–78, 2006.

[128] H. Zhang and J. F. Banfield. Thermodynamic Analysis of Phase Stability of Nanocrystalline Titania. *J. Mater. Chem.*, 8(9):2073–2076, 1998.

[129] S. Monti, A. C. T. van Duin, S. Y. Kim, and V. Barone. Exploration of the Conformational and Reactive Dynamics of Glycine and Diglycine on TiO_2: Computational Investigation in the Gas Phase and in Solution. *J. Phys. Chem. C*, 116:5141–5150, 2012.

[130] H. R. Hertz. Ueber sehr schnelle electrische Schwingungen. *Ann. Phys.*, 267(7):421–448, 1887.

[131] B. Stewart. An account of some experiments on radiant heat. *Trans. R. Soc. Edinburgh*, 2:1–20, 1858.

[132] G. Kirchhoff. Ueber das Verhältniss zwischen dem Emissionsvermögen und dem Absorptionsvermögen der Körper für Wärme and Licht. *Ann. Phys.*, 109(2):275–301, 1860.

[133] M. Planck. Über eine Verbesserung der Wienschen Spektralgleichung. *Verh. Dtsch. Phys. Ges.*, 2:202–204, 1900.

[134] M. Planck. Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum. *Verh. Dtsch. Phys. Ges.*, 2:237–245, 1900.

[135] A. Einstein. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Ann. Phys.*, 322(6):132–148, 1905.

[136] N. Bohr. On the Constitution of Atoms and Molecules, Part I. *Philos. Mag.*, 26(151):1–24, 1913.

[137] N. Bohr. On the Constitution of Atoms and Molecules, Part II: Systems Containing Only a Single Nucleus. *Philos. Mag.*, 26(153):476–502, 1913.

[138] N. Bohr. On the Constitution of Atoms and Molecules, Part III: Systems Containing Several Nuclei. *Philos. Mag.*, 26(155):857–875, 1913.

[139] N. Bohr. The Spectra of Helium and Hydrogen. *Nature*, 92(2295):231–232, 1914.

[140] N. Bohr. Atomic Structure. *Nature*, 107(2682):104–107, 1921.

[141] L. V. de Broglie. *Recherches sue la Thèorie des Quanta*. Masson & C^{ie}, 1925.

[142] A. H. Compton. A Quantum Theory of the Scattering of X-Rays by Light Elements. *Phys. Rev.*, 21(5):483–502, 1923.

[143] C. Davisson and L. H. Germer. The Scattering of Electrons by a Single Crystal of Nickel. *Nature*, 119:558–560, 1927.

[144] E. Schrödinger. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.*, 28(6):1049–1070, 1926.

[145] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Phys.*, 84:457–484, 1927.

[146] H. Jeffreys. On Certain Approximate Solutions of Linear Differential Equations of the Second Order. *P. Lond. Math. Soc.*, 23:428.436, 1924.

[147] G. Wentzel. Eine Verallgemeinerung der Quantenbedingungen für die Zwecke der Wallenmechanik. *Z. Phys.*, 38(6-7):518–529, 1926.

[148] H. A. Kramers. Wallenmechanik und Halbzahlige Quantisierung. *Z. Phys.*, 39(10-11):828–840, 1926.

[149] L. Brillouin. La mécanique ondulatoire de Schrödinger: une méthode générale de resolution par approximations successives. *Comptes Rendus Acad. Sci.*, 183(1):24–26, 1926.

[150] P. A. M. Dirac. The Quantum Theory of the Electron. *P. R. Soc. Lond. A-Conta.*, 117(778):610–624, 1928.

[151] P. A. M. Dirac. A Theory of Electrons and Protons. *Proc. R. Soc. Lond. A*, 126:360–365, 1930.

[152] W. Heisenberg. Über den anschaulichen Inhalt der Quantentheoretischen Kinematik und Mechanik. *Z. Phys.*, 43(3-4):172–198, 1927.

[153] V. Fock. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys.*, 61(1-2):126–148, 1930.

[154] J. C. Slater. Note on Hartree'e Method. *Phys. Rev.*, 35:210–211, 1930.

[155] C. C. J. Roothaan. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.*, 23(2):69–89, 1951.

[156] G. G. Hall. The Molecular Orbital Theory og Chemical Valency VIII. A Method of Calculating Ionization Potentials. *Proc. R. Soc. Lond. A*, 205(1083):541–552, 1951.

[157] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140(4A):A1133–A1138, 1965.

[158] L. H. Thomas. The Calculation of Atomic Fields. *Proc. Camb. Phil. SOc.*, 23:542–548, 1927.

[159] E. Fermi. Statistical Methods to Determine sone Properties of Atoms. *Rend. Accad. Naz. Lincei*, 6:602–607, 1927.

[160] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136(3B):B864–B871, 1964.

[161] D. R. Hartree. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Math. Proc. Cambridge*, 24 (1):89–110, 1928.

[162] D. R. Hartree. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Math. Proc. Cambridge*, 24 (1):111–132, 1928.

[163] D. R. Hartree. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part III. Term Values and Intensities in Series in Optical Spectra. *Math. Proc. Cambridge*, 24 (3):426–437, 1928.

[164] P. A. M. Dirac. Note on Exchange Phenomena in the Thomas Atom. *Math. Proc. Cambridge*, 26 (3):376–385, 1930.

[165] S. H. Vosko, L. Wilk, and M. Nusair. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: a Critical Analysis. *Can. J. Phys.*, 58 (8):1200–1211, 1980.

[166] A. D. Becke. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behaviour. *Phys. Rev. A*, 38:3098–3100, 1988.

[167] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996.

[168] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.*, 91 (14):146401, 2003.

[169] A. D. Becke. Density-Functional Thermochemistry. III. The Role of the Exact Exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[170] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab-Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.*, 98:11623–11627, 1994.

[171] C. Adamo and V. Barone. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.*, 110 (13):6158–6170, 1999.

[172] Y. Zhao, N. E. Schultz, and D. G. Truhlar. Exchange-Correlation Functional with Broad Accuracy or Metallic and Nonmetallic Coumpounds, Kinetics, and Noncovalent Interactions. *J. Chem. Phys.*, 123 (16):161103, 2005.

[173] Y. Zhao, N. E. Schultz, and D. G. Truhlar. Design of Density Functionals by Combining the Method od Constraint Satisfaction with Parametrization fot Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.*, 2 (2):364–382, 2006.

[174] Y. Zhao and D. G. Truhlar. A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Phys. Chem.*, 125 (19):194101, 2006.

[175] Y Zhao and D. G. Truhlar. The M06 Suite of Density Funcional for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, abd Transition Elements: Two New Functionals and

Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.*, 120:215–241, 2006.

[176] Y. Zhao and D. G. Truhlar. Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States. *J. Phys. Chem. A*, 110 (49):131126–131130, 2006.

[177] Y. Zhao and D. G. Truhlar. Exploring the Limit Accuracy of the Global Hybrid Meta Density Functional fot Main-Group Thermochemistry, Kinetics and Noncovalent Interactions. *J. Chem. Theory Comput.*, 4 (11):1849–1868, 2008.

[178] R. Peverati and D. G. Truhlar. M11-L: A Local Density Functional that Provides Improved Accuracy for Electronic Structure Calculations in Chemistry and Physics. *J. Phys. Chem. Lett.*, 3 (1):117–124, 2012.

[179] R. Peverati and D. G. Truhlar. Improving the Accuracy of Hybrid Meta-GGA Density Functionals by Range Separation. *J. Phys. Chem. Lett.*, 2 (21):2810–2817, 2011.

[180] R. Peverati and D. G. Truhlar. An Improved and Broadly Accurate Local Approximation to the Exchange-Correlation Density Functional: The MN12-L Functional fot Electronic Structure Calculations in Chemistry and Physics. *Phys. Chem. Chem. Phys.*, 14 (38):12171–13174, 2012.

[181] R. Peverati and D. G. Truhlar. Screebed-Exchange Density Functional with Broad Accuracy for Chemistry and Solid-State Physics. *Phys. Chem. Chem. Phys.*, 14 (47):16187–16191, 2012.

[182] H. S. Yu, X. He, S. Li, and D. G. Truhlar. MN15: a Kohn-Sham Global-Hybrid Exchange-Correlation Density Functional with Broad Accuracy fot Multi-Reference and Single-Reference Systems and Noncovalent Interactions. *Chem. Sci.*, 7:5032–5051, 2016.

[183] H. S. Yu, X. He, and D. G. Truhlar. MN15-L: A New Local Exchange-Correlation Functional fot Kohn-Sham Density Functional Theory with Broad Accuracy fot Atoms, Molecules, and Solids. *J. Chem. Theory Comput.*, 12 (3):1280–1293, 2016.

[184] S. Grimme. Accurate Description of Van der Walls Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.*, 25:1463–1473, 2004.

[185] S. Grimme, S. Ehrlich, and L. Goerigk. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.*, 32:1456–1465, 2011.

[186] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A Consistent and Accurate Ab-Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H- Pu. *J. Chem. Phys.*, 132:154104, 2010.

[187] E. Caldeweyher, C. Bannwarth, and S. Grimme. Extension of the D3 Dispersion Coefficient Model. *J. Chem. Phys.*, 147:034112, 2017.

[188] U. Dinur and A. T. Hagler. New approaches to empirical force fields. In Kenny B. Lipkowitz and Donald B. Boyd, editors, *Reviews in Computational Chemistry, Volume 2*, chapter 4, page 99. Wiley-VCH, Weinheim, 2007.

[189] F. Jensen. Force field methods. In Frank Jensen, editor, *Introduction to Computational Chemistry*, chapter 2, page 22. John Wiley & Sons Ltd, Chichester, 2007.

[190] P. M. Morse. Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Phys. Rev.*, 34:57–64, 1929.

[191] F. London. Zur Theorie und Systematik der Molekularkräfte. *Z. Physik*, 63:245–279, 1930.

[192] J. E. Lennard Jones. On the determination of molecular fields. II. From the equation of state of gas. *Proc. R. Soc. London Ser. A*, 106:463–477, 1924.

[193] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A*, 105:9396–9409, 2001.

[194] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alber Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.*, 100 (50):19357–19363, 1996.

[195] T. Vreven and K. Morokuma. On the Application of the IMOMO (Integrated Molecular Orbital + Molecular Orbital) Method. *J. Comput. Chem.*, 21 (16):1419–1432, 2000.

[196] F. Maseras and K. Morokuma. IMOMM: A New Integrated Ab-Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.*, 16 (9):1170–1179, 1995.

[197] Cesare Pisani. *Quantum Mechanical Ab-Initio Calculation of the Properties of Crystalline Materials*. Springer Berling Heidelberg, 1996.

[198] R. Dovesi, B. Civalleri, C. Roetti, V. R. Saunders, and R. Orlando. Ab-initio quantum simulation in solid state chemistry. In Kenny B. Lipkowitz, Raima Larter, and Thomas R. Cundari, editors, *Reviews in Computational Chemistry*, chapter 1, pages 1–126. John Wiley & Sons, New York, 2005.

[199] F. Bloch. Über die Quantenmechanik der Elektronen in Kristallgittern. *Zeitschrift für Phys.*, 52 (7-8:555–600, 1929.

[200] H. J. Monkhorst and J. D. Pack. Special Points for Brillouin-Zone Investigations. *Phys. Rev. B*, 13 (12):5188–5192, 1976.

[201] V. Fiorentini and M. Methfessel. Extracting Convergent Surface Energies from Slab Calculations. *J. Phys. Condens. Matter*, 8:6525–6529, 1996.

[202] O. Schütt, P. Messmer, J. Hutter, and J. VandeVondele. Gpu-accelerated sparse matrix–matrix multiplication for linear scaling density functional theory. In R. C. Walker and A. W. Götz, editors, *Reviews in Computational Chemistry*, chapter 8, pages 173–190. John Wiley & Sons, New York, 2016.

[203] U. Borstnik, J. VandeVondele, V. Weber, and J. Hutter. Sparse matrix multiplication: The distributed block-compressed sparse. *Parallel Comput.*, 40 (5-6):47–58, 2014.

[204] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele. CP2K: atomistic simulations of condensed matter systems. *WIREs Comput. Mol. Sci.*, 4 (1):15–25, 2014.

[205] J. VandeVondele and J. Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.*, 127 (11):114105, 2007.

[206] M. Krack. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.*, 114 (1-3):145–152, 2005.

[207] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.*, 167 (2):103–128, 2005.

[208] J. VandeVondele and J. Hutter. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.*, 118 (10):4365–4369, 2003.

[209] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rèrat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman. Quantum-mechanical condensed matter simulations with CRYSTAL. *WIREs Comput. Mol. Sci.*, 8 (31360):1–36, 2018.

[210] R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, P. D'Arco, M. LLunell, M. Causà, Y. Noël, L. Maschio, A. Erba, M. Rerat, and S. Casassa. *CRYSTAL17 User's Manual*, 2017.

[211] C. Kollmar. Convergence Optimization of Restricted Open-Shell Self-Consistent Field Calculations. *Int. J. Quant. Chem.*, 62:617–637, 1997.

[212] M. F. Guest and V. R. Saunders. On methods for converging open-shell Hartree-Fock wave-functions. *Mol. Phys.*, 28:819–828, 1974.

[213] P. Pulay. Improved SCF convergence acceleration . *Comp. Chem.*, 3:556–560, 1982.

[214] P. Debye. Näerungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index. *Mathematische Annalen*, 67:535–558, 1909.

[215] M. R. Hestenes and E. Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *J. Res. Nat. Bur. Stand.*, 49:409–436, 1952.

[216] C. G. Broyden. The Convergence of a Class of Double-Rank Minimization Algorithm. I: General Considerations. *J. Inst. Math. its Appl.*, 6:76–90, 1970.

[217] R. Fletcher. A New Approach to Variable Metric Algorithms. *Comput. J.*, 13:317–322, 1970.

[218] D. Goldfarb. A Family of Variable-Metric Methods Derived by Variational Means. *Math. Comput.*, 24:23–26, 1970.

[219] J. Nocedal. Updating Quasi-Newton Matrices with Limited Storage. *Math. Comput.*, 35:773–782, 1980.

[220] D. C. Liu and J. Nocedal. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Program.*, 45:503–528, 1989.

[221] R. H. Byrd, J. Nocedal, and R. B. Schnabel. Representation of Quasi-Newton Matrices and their Use in Limited Memory Methods. *Math. Program.*, 63:129–156, 1994.

[222] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.*, 16:1190–1208, 1995.

[223] P. Csaszar and P. Pulay. Geometry Optimization by Direct Inversion in the Iterative Subspace. *J. Mol. Struc.*, 114:31–34, 1984.

[224] E. Fermi, J. Pasta, and S. Ulam. Studies of Nonlinear Problems. *Los Alamos Report*, pages 1–22, 1955.

[225] B. J. Alder and T. E. Wainwright. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.*, 31 (2):459–466, 1959.

[226] A. Rahman. Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.*, 136 (2A):A405–A411, 1964.

[227] T. Schlick. Pursuing laplace's vision on modern computers. In J. P. Mesirov, K. Schulten, and D. W. Sumners, editors, *Mathematical Applications to Biomolecular Structure and Dynamics. The IMA Volumes in Mathematics and its Applications*, pages 218–247. Springer, New York, 1996.

[228] R. Car and M. Parrinello. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, 55:2471–2474, 1985.

[229] L. Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159 (1):98–103, 1967.

[230] G. Kresse and J. Hafner. Ab-initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B*, 47:558–562, 1993.

[231] G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B*, 49:14251–14269, 1994.

[232] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, 1996.

[233] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.*, 6:15–50, 1996.

[234] J. D. Gale. GULP: A computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc. Faraday Trans.*, 93:629–637, 1997.

[235] J. D. Gale and A. L. Rohl. The General Utility Lattice Program (GULP). *Mol. Simulat.*, 29:291–341, 2003.

[236] J. D. Gale. GULP: Capabilities and prospects. *Z. Krist.*, 220:552–554, 2005.

[237] S. Grimme. Semiempirical GGa-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.*, 27:1787–1799, 2006.

[238] B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano, and P. Ugliengo. B3LYP augmented with an empirical dispersion term (B3LYP-D*) as applied to molecular crystals. *CrystEngComm*, 10:405–410, 2008.

[239] G. Henkelmann and H. Jónsson. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.*, 111:7010–7022, 1999.

[240] A. Heyden, A. T. Bell, and F. J. Keil. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *J. Chem. Phys.*, 123:224101, 2005.

[241] J. Kästner and P. Sherwood. Superlinearly converging dimer method for transition state search. *J. Chem. Phys.*, 128:014106, 2008.

[242] P. Xiao, D. Sheppard, J. Rogal, and G. Henkelmann. Solid-state dimer method for calculating solid-solid phase transitions. *J. Chem. Phys.*, 140:174104, 2014.

[243] H. Jónsson, G. Mills, and K. W. Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. In B. J. Berne, G. Ciccotti, and D. F. Coker, editors, *Classical and Quantum Dynamics in Condensed Phase Simulations*, pages 385–404. World Scientific, Singapore, 1998.

[244] G. Henkelmann and H. Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, 113:9978–9985, 2000.

[245] G. Henkelmann, B. P. Uberuaga, and H. Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.*, 113:9901–9904, 2000.

[246] D. Sheppard, R. Terrel, and G. Henkelmann. Optimization methods for finding minimum energy paths. *J. Chem. Phys.*, 128:134106, 2008.

[247] D. Sheppard, P. Xiao, W. Chemelewski, D. D. Johnson, and G. Henkelmann. A generalized solid-state nudged elastic band method. *J. Chem. Phys.*, 136:074103, 2012.

[248] D. Sheppard and G. Henkelmann. Paths to which the nudged elastic band converges. *J. Comput. Chem.*, 32:1769–1771, 2011.

[249] T. H. Hill. Steric Effects. I. Van der Waals Potential Energy Curves . *J. Chem. Phys.*, 16:399–404, 1948.

[250] M. Matsui and M. Akaogi. Molecular Dynamics Simulation of the Structural and Physical Properties of the Four Polymorphs of TiO_2. *Mol. Simul.*, 6:239–244, 1991.

[251] O. Lamiel-Garcia, A. Cuko, M. Calatayud, F. Illas, and S. Bromley. Predicting size-dependent emergence of crystallinity in nanomaterials: titania nanoclusters versus nanocrystals. *Nanoscale*, 9:1049–1058, 2017.

[252] J. K. Burdett, T. Hughbanks, G. J. Miller, J. W. Richardson Jr., and J. V. Smith. Structural-Electronic Relationship in Inorganic Solids: Powder Neutron Diffraction Studies of the Rutile and Anatase Polymorphs of Titanium Dioxide at 15 and 295 K. *J. Am. Chem. Soc.*, 109:3639–3646, 1987.

[253] R. J. Gonzalez and R. Zallen. Infrared Reflectivity and Lattice Fundamentals in Anatase TiO_2. *Phys. Rev. B*, 55(11):7014–7017, 1997.

[254] J. T. Luxon and R. Summit. Interpretation of the Infrared Absorption Spectra of Stannic Oxide and Titanium Dioxide (Rutile) Powders. *J. Chem. Phys.*, 50(3):1366–1370, 1969.

[255] F. Gervais and B. Piriou. Anharmonicity in Several-Polar-Mode Crystals: Adjusting Phonon Self-Energy in LO and TO Modes in Al_2O_3 and TiO_2 to Fit Infrared Reflectivity. *J. Phys. C*, 7:2374–2386, 1974.

[256] F. Gervais and B. Piriou. Temperature Dependence of Transverse- and Longitudinal-Optic Modes in TiO_2 (Rutile). *Phys. Rev. B*, 10(4):1642–1654, 1974.

[257] S. V. Patwardhan, G. Patwardhan, and C. C. Perry. Interactions of Biomolecules with Inorganic Materials: Principles, Applications and Future Prospects. *J. Mater. Chem-.*, 17:2875–2884, 2007.

[258] D. Costa, L. Savio, and C. M. Pradier. Adsorption of Amino Acids and Peptides on Metal and Oxide Surfaces in Water Environment: A Synthetic and Prospective Review. *J. Phys. Chem. B*, 120:7039–7052, 2016.

[259] D. Costa, P. A. Garrain, and M. Baaden. Understanding Small Biomolecule-Biomaterial Interactions: A Review of Fundamental Theoretical and Experimental Approaches for Biomolecule Interactions with Inorganic Surfaces. *J. Biomed. Mater. Res. Part A*, 101:1210–1222, 2013.

[260] A. Rimola, D. Costa, M. Sodupe, J. F. Lambert, and P. Ugliengo. Silica Surface Features and their Role in the Adsorption of Biomolecules: Computational Modeling and Experiments. *Chem. Rev.*, 113:4216–4313, 2013.

[261] M. C. Roco. Convergence with Modern Biology and Medicine. *Curr. Opin. Biotechnol.*, 14:337–346, 2003.

[262] M. Ghadiri, W. Chrzanowski, and R. Rohanizadeh. Biomedical Applications of Cationic Clay Minerals. *RSC Adv.*, 5:29467–29481, 2015.

[263] E. Mahon andA. Salvati, F. Baldelli Bombelli, I. Lynch, and K. A. Dawson. Designing the Nanoparticle-Biomolecule Interface for Targeting and Therapeutic Delivery. *J. Controlled Release*, 161:164–174, 2012.

[264] C. Tamerler and M. Sarikaya. Biomimetics: Nanotechnology and Bionanotechnology Using Genetically Engineered Peptides. *Philos. Trans. R. Soc. A*, 267:1705–1726, 2009.

[265] E. I. Unuabonah, C. Günter, J. Weber, S. Lubahn, and A. Taubert. Hybrid Clay: A New Highly Efficient Adsorbent for Water Treatment. *ACS Sustainable Chem. Eng.*, 1:966–973, 2013.

[266] L. Liu, L. Q. Yang, H. W. Liang, H. P. Cong, J. Jiang, and S. H. Yu. Bioinspired Fabrication of Hierarchical FeOOH Nanostructure Array Films at the Air-Water Interface, their Hydrophobicity and Application for Water Treatment. *ACS Nano*, 7:1368–1378, 2013.

[267] D. Szieberth, A. M. Ferrari, and X. Dong. Adsorption of Glycine on the Anatase (101) Surface: An ab Initio Study. *Phys. Chem. Chem. Phys*, 12:11033–11040, 2010.

[268] W. Li, K. Kotsis, and S. Manzhos and. Comparative Density Functional Theory and Density Functional Tight Binding Study of Arginine and Arginine-Rich Cell Penetrating Peptide TAT Adsorption on Anatase TiO_2. *Phys. Chem. Chem. Phys.*, 18:19902–19917, 2016.

[269] W. Friedrichs, S. Köppen, and W. Langel. Titanium Binding Dodecapeptides and the Impact of Water Structure. *Surf. Sci.*, 617:42–52, 2013.

[270] P. Mignon, P. Ugliengo, and M. Sodupe. Theoretical Study of the Adsorption of RNA/DNA Bases on the External Surfaces of Na^+-Montmorillonite. *J. Phys. Chem. C*, 113 (31):13741–13749, 2009.

[271] N. Nitadai and S. Maruyama. Origins of building blocks of life: A review. *Geosci. Front.*, 9:1117–1153, 2018.

[272] J. E. Šponer, J. Šponer, and E. Di Mauro. New evolutionary insights into the non-enzymatic origin of RNA oligomers. *Wiley Interdiscip. Rev. RNA*, 8:e1400, 2016.

[273] J. E. Šponer, J. Šponer, A. Giorgi, E. Di Mauro, S. Pino, and G. Costanzo. Untemplated nonenzymatic polymerization of 3', 5' cGMP: a plausible route to 3', 5'-linked oligonucleotides in primordia. *J. Phys. Chem. B*, 119:2979–2989, 2015.

[274] J. E. Šponer, J. Šponer, O. Nováková, V. Nrabec, O. Šedo, Z. Zdráhal, G. Costanzo, S. Pino, R. Saladino, and E. Di Mauro. Emergence of the first catalytic oligonucleotides in a formamide-based origin scenario. *Chem. Eur. J.*, 22:3572–3586, 2016.

[275] M. Rodriguez-Garcia, A. J. Surman, G. J. T. Cooper, I. Suárez-Marina, Z. Hosni, M. P. Lee, and L. Cronin. Formation of oligopeptides in high yield under simple programmable conditions. *Nature*, 6:8385, 2015.

[276] V. R. Pattabiraman and J. W. Bode. Rethinking amide bond synthesis. *Nature*, 480:471–479, 2011.

[277] D. J. C. Constable, P. J. Dunn, J. D. Hayler, G. R. Humphrey, J. J. L. Leazer, R. J. Lindermann, K. Lorenz, J. Manley, B. A. Pearlmann, A. Wells, A. Zaks, and T. Y. Shang. Key green chemistry research areas—a perspective from pharmaceutical manufacturers. *Green Chem.*, 9:411–420, 2007.

[278] S. F. Sousa, P. A. Fernandes, and M. J. Ramos. General Performance of Density Functional. *J. Phys. Chem. A*, 111:10439–10452, 2007.

[279] G.-J. Kroes. Toward a Database of Chemically Accurate Barrier Heights for Reaction of Molecules with Metal Surfaces. *J. Phys. Chem. Lett.*, 6:4106–4114, 2015.

[280] S. Andersson and M. Griining. Performance of Density Functional for Calculating Barrier Heights of Chemical Reactions Relevant to Astrophysics. *J. Phys. Chem. A*, 108:7621–7636, 2004.

[281] S. Grimme. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem. Eur. J.*, 18:9955–9964, 2012.

[282] W. Tang, E. Sanville, and G. Henkelman. A Grid-Based Bader Analysis Algorithm without Lattice Bias. *J. Phys. Condens. Matter*, 21:084204, 2009.

[283] E. Sanville, S. D. Kenny, R. Smith, and G. Henkelman. An Improved Grid-Based Algorithm for Bader Charge Allocation. *J. Comp. Chem.*, 28:899.908, 2007.

[284] G. Henkelman, A. Arnaldsson, and H. Jonsson. A Fast and Robust Algorithm for Bader Decomposition of Charge Density. *Comput. Mater. Sci.*, 36:354–360, 2006.

[285] M. Yu and D. R. Trinkle. Accurate and Efficient Algorithm for Bader Charge Integration. *J. Chem. Phys.*, 134:064111, 211.

[286] J. H. Jensen, K. K. Baldridge, and M. S. Gordon. Uncatalyzed Peptide Bond Formation in the Gas Phase. *J. Phys. Chem.*, 96:8340–8351, 1992.

[287] A. Rimola, M. Sodupe, and P. Ugliengo. Amide and Peptide Bond Formation: Interplay between Strained Ring Defects and Silanol Groups at Amorphous Silica Surfaces. *J. Phys. Chem. C*, 120:24817–24826, 2016.

[288] A. Rimola, S. Tosoni, M. Sodupe, and P. Ugliengo. . *ChemPhysChem*, 7:157–163, 2006.

[289] N. Lahav, D. White, and S. Chang. . *Science*, 201:67–69, 1978.

[290] S. Pantaleone, A. Rimola, and M. Sodupe. Canonical, Deprotonated, or Zwitterionic? A Computational Study on Amino Acid Interaction with the TiO_2 (101) Anatase Surface. *J. Phys. Chem. C*, 121:14156–14165, 2017.

[291] B. V. Slokovic. Neurovascular mechanisms of Alzheimer's neurodegeneration. *Trends Neurosci.*, 28:202–208, 2005.

[292] R. B. Maccioni ad G. Farías, I. Morales, and L. Navarrete. The revitalized tau hypothesis on Alzheimer's disease. *Arch. Med. Res.*, 41:226–231, 2010.