Classe di Scienze

Ph.D. in Chemistry

Spectroscopic Studies of Molecular Systems Relevant in Astrobiology

Candidate
Teresa Fornaro

Advisor
Prof. Vincenzo Barone

Co-advisor
Dr. Malgorzata Biczysko

Academic Year 2014-2015

Abstract

In the Astrobiology context, the study of the physico-chemical interactions involving "building blocks of life" in plausible prebiotic and space-like conditions is fundamental to shed light on the processes that led to emergence of life on Earth as well as to molecular chemical evolution in space.
In this PhD Thesis, such issues have been addressed both experimentally and computationally by employing vibrational spectroscopy, which has shown to be an effective tool to investigate the variety of intermolecular interactions that play a key role in self-assembling mechanisms of nucleic acid components and their binding to mineral surfaces. In particular, in order to dissect the contributions of the different interactions to the overall spectroscopic signals and shed light on the intricate experimental data, feasible computational protocols have been developed for the characterization of the spectroscopic properties of such complex systems. This study has been carried out through a multi-step strategy, starting the investigation from the spectroscopic properties of the isolated nucleobases, then studying the perturbation induced by the interaction with another molecule (molecular dimers), towards condensed phases like the molecular solid, up to the case of nucleic acid components adsorbed on minerals.
A proper modeling of these weakly bound molecular systems has required, firstly, a validation of dispersion-corrected Density Functional Theory methods for simulating anharmonic vibrational properties. The isolated nucleobases and some of their dimers have been used as benchmark set for identifying a general, reliable and effective computational procedure based on fully anharmonic quantum mechanical computations of the vibrational wavenumbers and infrared intensities within the generalized second order vibrational perturbation theory (GVPT2) approach, combined with the cost-effective dispersion-corrected density functional B3LYP-D3, in conjunction with basis sets of double-ζ quality such as N07D and SNSD.
Such a protocol has been then applied to the dimers of nucleobases in order to study the perturbation on the vibrational frequencies and infrared intensities induced by the intermolecular hydrogen-bonding interactions. Efforts have been made to solve the problems of simulating strongly anharmonic vibrations within hydrogen-bonded bridges, focusing on the requirement of a very accurate description of the underlying potential energy surface. Improved results for such vibrations have been achieved by means of hybrid models, where the harmonic part of the force-field is computed

at a higher level of theory like B2PLYP, or by application of the less demanding ONIOM B2PLYP:B3LYP scheme, which is a focused model where only the part of the molecular system forming the hydrogen bonds is treated at the B2PLYP level of theory. Moreover, for improving the vibrational frequencies of modes like the stretching of C=O and N-H functional groups, which are particularly sensitive to hydrogen-bonding, correction parameters for the B3LYP-D3/N07D frequencies have been determined.

Afterwards, the treatment of the vibrational properties of nucleobases in condensed phases has been faced, focusing on uracil in the solid state. In particular, a heptamer cluster of uracil molecules has been considered as model to represent the properties in the solid state. The relative vibrational frequencies have been computed at anharmonic level within the VPT2 framework, combining two cost-effective approaches, namely the hybrid B3LYP-D3/N07D:DFTBA model, where the harmonic frequencies are computed with B3LYP-D3/N07D method and the anharmonic corrections are evaluated with the less expensive DFTBA method, and the reduced dimensionality VPT2 (RD-VPT2) approach, in which only selected vibrational modes are calculated anharmonically (including the couplings with the other modes) while the remaining modes are treated at the harmonic level, using the B3LYP-D3/N07D method only. The reliability of such theoretical results has been validated with respect to experiments, by performing infrared measurements of uracil in the solid state through the Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) technique. The good performance in predicting the experimental shifts of the vibrational frequencies of uracil due to the intermolecular hydrogen bonds in the solid state with respect to uracil isolated in Argon matrix, has allowed also to provide some new assignments of the experimental spectrum of uracil in the solid state.

Finally, the study of molecule-mineral interactions has been addressed, investigating experimentally the thermodynamics of the adsorption process of nucleic acid components on brucite, a serpentinite-hosted hydrothermal mineral, through determination of the equilibrium adsorption isotherms. Additionally, surface complexation studies have been carried out to get the stoichiometry of surface reactions and the associated electrical work. Such surface complexation modeling has provided reasonable inferences for the possible surface complexes, determining the number of inner/outer-sphere linkages for the adsorbates and the number of surface sites involved in the reaction stoichiometry. However, to distinguish the specific functional groups which constitute the points of attachment to the surface, further quantum mechanical simulations on the energetics of these complexes and spectroscopic characterizations are in progress.

Contents

1 Introduction **1**
1.1 Addressing The Scientific Question about the Origin of Life 1
1.2 *In Silico* Modeling: Computational Spectroscopy 4
1.3 Summary of the Thesis 8

2 Validation of Dispersion-corrected DFT Methods for Anharmonic Vibrational Spectra: Nucleobases and their Dimers **17**
2.1 Nucleobase Monomers 19
2.1.1 DFT and Hybrid CC/DFT Methods: the case of Uracil 19
2.1.2 Dispersion-Corrected B3LYP-based Approaches 22
2.2 Hydrogen-Bonded and Stacked Dimers of Adenine and Uracil 25
2.3 Concluding Remarks 30

3 Improvements of Vibrational Properties for Functional Groups involved in Hydrogen Bonds: Uracil-Water Complexes and Nucleic Acid Base Dimers **31**
3.1 Validation of the B2PLYP-D3 Method: the 2-Thiouracil and the Dimers of Acetamide and Formamide 34
3.2 Hybrid and ONIOM models 39
3.2.1 Validation of the Computational Method: the Uracil-Water Complexes 39
3.2.2 Effects of Hydrogen-Bonding Interactions on IR Spectra: the Hydrogen-Bonded Uracil Dimers 44
3.3 Cost-Effective Correction for ν(C=O) and ν(N-H) within Hydrogen-Bonded Bridges 49
3.3.1 Determination and Validation of Correction Parameters for B3LYP-D3/N07D frequencies of ν(C=O) and ν(N-H) involved in Hydrogen Bonds: the Nucleobase Dimers 50
3.4 Concluding Remarks 54

4 Development of Feasible Protocols for Vibrational Properties of Large Systems: Uracil in the Solid State **55**
4.1 Selection of the Cluster Model for Uracil in the Solid State 57

4.2 Validation of Hybrid and Reduced Dimensionality VPT2 Approaches 62
4.2.1 Hybrid B3LYP-D3/DFTBA Method 62
4.2.2 RD-VPT2 Method 65
4.3 Computations *vs.* Experiment 67
4.4 Concluding Remarks 72

5 Study of Molecule-Mineral Interactions: Binding of Nucleic Acids Components to Brucite 73
5.1 Synthesis and Characterization of Brucite Mineral 75
5.1.1 Brucite Synthesis and Hydrothermal Treatment 75
5.1.2 Characterization of Synthetic Brucite 75
5.2 Thermodynamics of Adsorption of Nucleic Acids Components on Brucite 79
5.2.1 Batch Adsorption Experiments 79
5.2.2 Surface Complexation Modeling 80
5.3 Results and Discussion 84
5.4 Concluding Remarks and Further Investigations 98

6 Conclusions 101

A Theoretical background 105
A.0.1 Vibrational energies 106
A.0.2 Band intensities 109

Bibliography 113

List of Articles 141

Academic Education and Scientific Activities 143

Acknowledgements 147

Chapter 1

Introduction

1.1 Addressing The Scientific Question about the Origin of Life

A fundamental question in Astrobiology concerns the transition from the geochemical world to the biochemical one. The so-called abiogenesis, the set of processes that led to emergence of life from inorganic matter on early Earth and potentially on other planets, is still matter of debate within the scientific community. To face such a complex issue, several aspects need to be addressed. First, it is necessary to investigate the plausible conditions and the possible chemical pathways for the synthesis of the monomeric components of the biomolecules ("building blocks of life"). The subsequent steps involve the study of the interactions among such abiotically-formed organics in plausible environmental conditions, investigating the transformations induced by astrobiologically relevant external agents, towards the study of the co-evolution of compartmentalization, metabolism and informational polymers, up to the emergence of the first simple living systems.[1–8]

A variety of chemical routes, both endogenous and exogenous, have been found for the abiotic formation of key organic molecules and more complex biomolecules in many plausible environmental conditions. The presence of organics in meteorites and comets is an evidence of the occurrence of abiotic processes in the interstellar medium and molecular clouds, that have been widely simulated in laboratory.[9–18] In the interstellar medium (ISM) about 200 complex organic molecules (COMs) have been found so far, i.e. saturated or unsaturated molecules containing carbon and more than five atoms.[19] Examples of organic interstellar molecules are acetone, dimethyl ether, ethylene glycol, formamide, acetamide, acetaldehyde, simple sugars like glycolaldehyde, other species like cyanopolyynes, cumulenes, benzene.[19,20] Even though the presence of the simplest amino acid, glycine, in the ISM is still matter of controversy,[21,22] it has been detected in several carbonaceous meteorites and in cometary dust from the Stardust mission.[23] Several other amino acids like alanine, valine, isovaline, proline, leucine, isoleucine, threonine, serine, aspartic acid,

glutamic acid, have been found in meteorites, some of them showing enantiomeric excess. There is also evidence for extraterrestrial hydrocarbons, carboxylic acids, urea, ketones, alcohols, aldehydes, glycolaldehyde, sugars, purines like xanthine and adenine, pyrimidines like uracil.[11,24–27]
A significant input of such abiotically-formed organics was likely delivered on early Earth and the other terrestrial planets as a consequence of impacts by interplanetary dust particles, asteroids and comets during the Late Heavy Bombardment period occurred approximately 4.1 to 3.8 billion years ago.[28]
On the other side, endogenous generation of organic molecules likely occurred through many physico-chemical processes promoted by the energy sources available on the prebiotic Earth (such as electric discharges, high energy radiations, hydrothermal fluxes) and by mineral catalysts.[1,2,20,29–39] Indeed, several plausible reactive pathways have been found for the synthesis of the building blocks of the biological macromolecules, occurring in the Earth's Archean atmosphere,[40,41] or in the mid-ocean hydrothermal systems,[42–48] in surface and shallow subsurface environments.[49–52]
Abiotic routes have been discovered also to account for homochirality, which is one of the most distinctive biochemical signatures of life and a prerequisite for life. Examples are deracemization induced by UV circular polarized light,[53–58] partial sublimation and homochiral self-organization,[59,60] enantioselective adsorption on mineral surfaces,[61–63] asymmetric auto-catalysis.[64–74] In support of the prebiotic origin of homochirality, as already mentioned, enantiomeric excesses in chiral organic compounds have been detected in meteorites.[25,75–79]

Then, given the prebiotic availability of the biomolecule precursors, the crucial transition from the inanimate matter to the biological systems probably occurred through the selection, concentration and organization of the organic precursors, yielding to the essential macromolecules of life.[62,80]
Mineral surfaces might have played a very important role in such processes, due to their ability to selectively adsorb and concentrate organic molecules on a local scale, removing key organics from the aqueous environments where the hydrolysis prevails on the polymerization, promoting self-organization and catalyzing important chemical reactions.[13,62,81–96] In some cases, mineral surfaces feature an enantioselective potential, which might have been responsible of the emergence of chirality in prebiotic conditions.[61,63,68,97–100] Moreover, the physico-chemical interactions between molecules and minerals might affect the lifetime of molecules under space conditions, e.g. photocatalytic minerals may promote a rapid degradation of the adsorbed molecules or catalyze reactions towards more complex species.[12,13,17,93,101–103] On the other hand, it has been observed that some mineral matrices are able to protect molecules against photodegradation.[104–108] This suggests a pivotal role of minerals in the prebiotic evolution of complex chemical systems.

However, the huge amount of possible reactive pathways, the complexity of the involved heterogeneous systems, the extreme environmental conditions which need to be taken into account, make the study of the abiogenesis an overwhelming task. In this perspective, spectroscopic methods can be employed to investigate a variety

of properties and processes occurring in the prebiotic systems, giving access to detailed information about the intermolecular interactions occurring in the gas-phase, the conformational stability of flexible molecules, chirality, solvent effects, interactions with solid substrates and photochemical processes.[109–112]

Accurate spectroscopic data can be obtained by laboratory experimental measurements, used for analyzing astrophysical observations gained by remote sensing which provide high-resolution data of unknown samples. The analysis and interpretation of such experimental data can be rather intricate, due to several different effects usually shuffled into such molecular systems. Even in the case of small molecules in the gas-phase,[113–117] a variety of intramolecular interactions can occur, strongly affecting the spectroscopic features. Analogously, the spectra of flexible molecules are of difficult interpretation, since separate contributions originating from the different conformers are usually mixed up in the spectra.[116–118]

Additionally, an important source of complexity originates from the environmental effects, such as the molecule-solvent[119,120] and molecule-mineral interactions.[94,103] Indeed, the formation of supra-molecular complexes significantly alters the spectroscopic features with respect to the gas-phase scenario, due to the introduction of new intermolecular interactions as well as the perturbation of the intramolecular ones, causing changes in the spectroscopic selection rules.[121] Especially in the case of hydrogen-bonding interactions, quite remarkable effects on the spectra are usually observed experimentally.[122–124] For instance the red shift and intensity enhancement in the X-H stretching frequencies involved in hydrogen bonds are included among the criteria based on spectroscopy in recent IUPAC definition for hydrogen bond formation.[125] In such cases the overall spectroscopic signals ensue from the contribution of all the different kinds of intra- and inter-molecular interactions which take place within the system or between the system and the environment, hampering the interpretation of the experimental outcomes. This has important implications for life detection in space mission context, casting doubts on the interpretation of the astronomical data, because assignments of the spectroscopic features based on laboratory gas-phase data may be misleading.[103]

In such circumstances, *in silico* modeling represents a useful tool in order to assist the interpretation of the experimental/observational measurements.[119,126–142]

Generally speaking, computer simulations have found in the last years a broad application to figure out phenomena of interest in Astrobiology,[143,144] ranging from astrophysical problems such as the chemistry of simple molecules in the early universe[18,145–148] or the evolution of planetary atmospheres,[149–152] up to the evolution of biomolecules,[153–160] passing through issues that properly lie in the realm of chemistry such as the characterization of the properties and reactivity of key prebiotic molecules in various plausible environmental conditions.[138]

The latter issues were mainly inspected in this work, focusing on the development of general, reliable and effective computational protocols for the characterization of the spectroscopic properties of building blocks of life such as the nucleobases and their complexes in plausible prebiotic environments. The main results on this sub-

ject have been summarized in the present Thesis.

1.2 *In Silico* Modeling: Computational Spectroscopy

The quantum mechanical (QM) simulation of molecular spectra requires the solution of the complete (electronic and nuclear) Schrödinger equation for polyatomic molecules, and the treatment of the light-matter interaction in an effective way, in order to determine the position and intensity of the spectroscopic peaks.[111,161–166] Recent improvements in computational methods by the DREAMSlab (Dedicated Research Environment for Advanced Modeling and Simulations), research group of Prof. Vincenzo Barone at the Scuola Normale Superiore, have led to the development of a powerful tool to interpret and predict experimental spectroscopic data by means of a multi-frequency spectrometer, which covers rotational, vibrational (e.g. infrared, Raman, vibrational circular dichroism), electronic (e.g. absorption, emission, electronic circular dichroism), as well as magnetic resonance (e.g. electron paramagnetic resonance, nuclear magnetic resonance) spectral ranges, for simulating different kinds of spectra that can be directly compared with experimental outcomes through graphical facilities.[132,167] Such computational spectroscopy techniques have shown to be quite effective to unravel the intricate experimental data when dealing with molecular systems of increasing size and complexity.[111,132,168–173]

Regarding vibrational spectroscopy, the resolution of the nuclear Schrödinger equation within the harmonic approximation leads to strong limitations both qualitatively, e.g. due to "missing transitions" corresponding to overtones and combination bands which are forbidden by the selection rules but can be significant features in the experimental spectra, and quantitatively because of the inaccuracy in the estimate of the position of the experimental peaks. Therefore, the harmonic model is not able to reproduce the complex overall band pattern of experimental spectra and, as a consequence, it does not allow to correctly analyze experimental outcomes, for example distinguishing low-intensity features related to non-fundamental transitions of the most populated species present in experimental mixtures from fundamental transitions of the less abundant ones.[134,174–178] An accurate reproduction of the experimental features requires a more complex model which takes into account more realistic description of the potential energy surface (PES), i.e. the treatment of the vibrational molecular motion at the anharmonic level.
The common approach to correct frequencies for anharmonicity and improve their agreement with the experimental findings is obtained by using simple scaling factors[179–183] or more sophisticated scaling methods.[184–186] Mode specific scaling improves the agreement between computed and experimental vibrational frequencies, but the uncertainty of the optimized scaling factors cannot be lower than 0.02[181] and the problem of transferability is not trivial. Furthermore, due to the differ-

ent role played by anharmonicity in vibrational frequencies, zero-point vibrational energies (ZPVE) and partition functions, different scaling factors must be used. As a consequence, the definition of a consistent procedure becomes rather cumbersome.[181,187,188]

In contrast, consistent procedures can be derived from QM computations of vibrational properties beyond the harmonic approximation. However, with the inclusion of anharmonicity, the Schrödinger equation for the nuclear motion cannot be exactly solved and approximate methods must be used, for instance within perturbative[189–207] or self-consistent field[208–212] frameworks. In particular, effective schemes have been developed by the DREAMSlab based on the second order vibrational perturbation theory (the VPT2 approach[199,200,213]), which employs a fourth order Taylor expansion of the energy along the normal modes as definition of the intramolecular potential. Recently, the method has been extended to compute vibrational energies and transition intensities from the vibrational ground state to fundamentals, overtones and combination bands.[199,200,214–216] In order to calculate anharmonic frequencies and IR intensities, the underlying semi-diagonal quartic potential energy and cubic electric dipole moment surfaces are derived through numerical differentiations on geometries displaced from equilibrium along the normal modes.

One of the problems affecting the VPT2 theory concerns the so-called Fermi resonances[162] that may occur, especially in polyatomic molecules, between two vibrations which have the same energy, when some symmetry conditions are fulfilled. In these cases, a mutual perturbation takes place, resulting in the repulsion of the two levels associated with these vibrations. The wavefunctions describing the perturbed states become combinations of the original wavefunctions, causing a redistribution of intensities between the transitions from the ground state to the two perturbed levels. Various effective schemes for the treatment of the Fermi resonances have been developed in the last years.[199,201,213,217] One of the well-tested models, which has been used throughout this work, is based on the variational treatment of the resonant dyads, and is usually referred to as the Generalized VPT2 (GVPT2[199,201,213]) scheme. In the GVPT2 model, nearly-resonant contributions are firstly identified by means of *ad hoc* tests, such as the empirical test proposed by Martin *et al.*,[201] then removed from the perturbative treatment (leading to the so-called deperturbed model, DVPT2) and, finally, treated through a proper reduced dimensionality variational computation at a later stage.[194,199,201] With respect to the transition intensities, similarly to vibrational frequencies, it is necessary to face problems of singularities leading to excessive contributions from anharmonic terms, like the above mentioned Fermi resonances and the 1-1 resonances occurring among fundamental transitions. Analogously to the treatment of energy resonances, *ad hoc* tests[201,215] are used to remove from the calculation terms identified as resonant, but no variational treatment is performed *a posteriori*.

More details about the GVPT2 model are reported in Appendix A.

Using the GVPT2 model, as implemented in the GAUSSIAN package,[218] the com-

putation of anharmonic vibrational spectra, as well as the vibrational contributions to the thermodynamic functions, or to rotational parameters and magnetic properties, is feasible and provides a high accuracy.[132,134,219–223]

For the electronic part, methods based on the Density Functional Theory (DFT) are very practical for the calculation of the molecular potential, since they combine a good accuracy with the feasibility of calculations for large systems. Among these, hybrid (e.g. B3LYP[224]) or double-hybrid (e.g. B2PYLP[225–227]) functionals combined with double- or triple-ζ basis sets, respectively, are usually found to provide very good results.[118,227,228] For amino acids and nucleobases, which are among the most important building blocks of the biological macromolecules, such computational approaches have been extensively applied, and have been found to provide very accurate results at a feasible computational burden.[118,119,128,131,133,170,229,230] However, the presence of complex interactions among the biomolecules (i), such as hydrogen bonds or van der Waals interactions, or with the environment, such as the solvent (ii) and/or the mineral surfaces (iii), introduces new challenges from a computational point of view.

Concerning the treatment of the dispersion interactions (i), unfortunately standard density functionals fail in describing such non-local and non-classical electronic interactions,[231] and many approaches[232–239] have been proposed so far in order to include the relevant corrections at the DFT level, when weakly bound molecular complexes are under study. However, in many cases[128,134,175,240] some of the most successful last-generation functionals (e.g. the M06-2X[236] and the ωB97X[237,238]) do not predict vibrational wavenumbers with the same accuracy usually found using more standard DFT functionals like B3LYP. On the other side, a particularly high accuracy has been found[128,230,240] when simple semi-empirical dispersion corrections are just added to the B3LYP or B2PLYP functionals such as in the case of B3LYP-D.[232,233] In DFT-D,[233,241,242] dispersion interactions are modeled by adding to the density functional a semi-empirical dispersion term (DFT-D), that is a long-range attractive pair-potential (inversely proportional to the sixth power of the intermolecular distance) multiplied by a damping function which determines the range of the dispersion correction in order to avoid near singularities for very small distances and double-counting of correlation effects at intermediate distances.
A very good performance in predicting structural parameters and binding energies of chemical systems involving mainly dispersion forces has been observed using the improved version of Grimme's correction, B3LYP-D3[234,243] coupled to double-ζ basis sets such as the N07D[244,245] and SNSD.[134,228] In DFT-D3, atom-pairwise specific dispersion coefficients and cutoff radii computed from first principles have been introduced together with dispersion coefficients dependent on coordination numbers (geometry). These modifications lead to an improved accuracy, less empiricism and a wide range of applicability, since it has been parameterized for 94 elements of the periodic table. Tripeptide-folding, metallic systems, graphene, benzene on the Ag(111) surface and other molecular complexes have been successfully simulated

using such method.[234]
Another promising approach is the one based on the DiLabio's pesudopotentials,[246–250] the recently developed B3LYP-DCP method with the recommended companion 6-31+G(2d,2p) basis set. This corrects B3LYP by using atom-centered effective core potentials (dispersion-correcting potentials DCPs), composed of Gaussian-type functions, able to model dispersion by modifying the environment in which the valence electrons move.[246–248,250] Effective core potentials are normally used in simulations involving metals and other heavy atoms, and in this case they are employed to account for dispersion interaction. Such method has shown to model satisfactorily π-stacking, steric repulsion noncovalent interactions and also hydrogen-bonding.[249,251,252]

Solvent effects (ii) can be implicitly accounted for, i.e. considering the solvent as a continuum interacting with the solute charge distribution in a pure electrostatic picture,[253–258] or alternatively an explicit solvent model can be used, where e.g. a QM method is employed for the description of the solute in conjunction with a less accurate and low-cost method applied to the solvent molecules. An intermediate approach in which the solute and few explicit solvent molecules are embedded in a continuum polarizable solvent model, has also been shown to achieve highly accurate results by including the polarization effects of the bulk of the solvent media in the most cost-effective way (see, for example, Improta *et al.*[259]).

Regarding the treatment of the mineral surfaces (iii), periodic slab models or non-periodic cluster models can be employed, although the former are usually less practical when spectroscopic properties at the anharmonic level are computed, due to frustrated translational and/or rotational degrees of freedom in the representation of the internal potential function.[260] Using such models for the mineral surface, computational studies have been carried out for several amino acids adsorbed on different substrates, for example cometary dust analogs such as forsterite in the presence of amorphous water ice,[261,262] other silicates like clays,[263] or silicon,[131] silica[264–272] and quartz,[273–276] or metal and metal oxide surfaces,[277,277–291] carbonates like calcite,[292,293] phosphates like hydroxyapatite,[294–298] sulfides like pyrite.[299–301] Also the adsorption of nucleobases has been computationally studied, for instance on the surfaces of clays,[302–307] graphite[88,308,309] and graphene,[310–314] metals.[315–317] Such computational studies allow a factual determination of the adsorption geometry by investigating the energetics of the possible surface complexes or the dynamics of the adsorption process under various conditions. Moreover, the computation of the binding energies is particularly helpful in distinguishing between physisorption and chemisorption processes, providing more easily a better estimate of the energies involved in the adsorption process with respect to experimental determinations of the adsorption enthalpy by means of the van't Hoff equation.[318]
On the whole, it is nowadays possible to derive realistic molecular models, considering all necessary inter- and intra-molecular interactions or environmental effects. Therefore, computational simulations represent a powerful tool to exhaustively understand the real adsorption mechanism, unraveling the phenomena occurring be-

tween organics and inorganic matrices in prebiotic environments. In the Astrobiology context, such studies are fundamental to fill the gap concerning the transition from the geochemical world to the biochemical one.

1.3 Summary of the Thesis

The present research project has been realized in the Theoretical and Computational research group led by Prof. Vincenzo Barone, during a three-years Ph.D. program in Chemistry at the Scuola Normale Superiore in Pisa.
The main activity pertains to the development of comprehensive yet feasible computational protocols for the characterization of the spectroscopic properties of "building blocks of life" such as nucleic acid components in plausible prebiotic environments. These computational procedures provide a tool to investigate the intermolecular interactions playing a key role in the self-assembling mechanisms of such molecules and their binding to mineral surfaces, which are of great interest to unravel the physical and chemical phenomena that led to the emergence of life on Earth as well as the molecular chemical evolution in space.
Additionally, complementary experimental studies, concerning infrared measurements and adsorption experiments, have been carried out at the National Institute for Astrophysics (INAF)-Astrophysical Observatory of Arcetri in Florence, under the supervison of Dr. John Robert Brucato, and at the Geophysical Laboratory of the Carnegie Institution of Washington (CIW), DC, USA, under the supervision of Prof. Robert M. Hazen and Prof. Dimitri Sverjensky of the Johns Hopkins University.

Nucleobases are particularly relevant in the prebiotic context because they are coding components of nucleic acids. Due to the possibility to form intermolecular hydrogen bonds and stacking interactions, nucleobases can spontaneously self-assemble forming dimers, complex aggregates, up to mono- and multilayers adsorbed on minerals in plausible prebiotic conditions.[87,319] Such capability of self-organization has very likely been one of the driving forces that led to the origin of life.[87] For instance, processes driven by hydrogen-bonding interactions have been proposed as plausible evolutionary pathways toward contemporary nucleic acids based on supramolecular noncovalent assembly of nucleobases or some plausible building blocks of pre-RNAs.[4,320]
The investigation on the nature, formation and evolution of such supra-molecular structures requires a detailed analysis of the physico-chemical interactions taking place among the molecules and between the molecules and the environment, which can be performed by spectroscopic techniques.[109,321–325] However, due to the multiple interactions which may be involved, especially the weak inter- and intra-molecular interactions such as dispersive interactions and hydrogen bonds, the modeling of these systems turns out to be quite challenging. The non-local and non-

classical dispersive interactions cannot be described using standard functionals.[231] The vibrational treatment of hydrogen-bonded systems is even more demanding, due to enhanced anharmonicity, which implies both limitations of the perturbative treatment and the requirement of a very accurate description of the underlying potential energy surface (PES).[326] Nevertheless, reliable theoretical description of the relative spectroscopic features would be highly desirable to facilitate the analysis of the experimental outcomes, since remarkable effects on vibrational frequencies and IR intensities are usually observed when functional groups are involved in hydrogen bonds.[122–124,170–173,327]

In this context, to define a suitable molecular model, we have adopted a multi-step strategy starting the investigation from the spectroscopic properties of the isolated molecules, then studying the perturbation induced by the interaction with another molecule (molecular dimers), towards condensed phases like the molecular solid, up to monolayers of biomolecules on solids.
The structures and numbering schemes of the main systems investigated in this work are summarized in:

- Figure 1.1: monomers of nucleobases, namely the purines adenine (A), hypoxanthine (Hy), and the pyrimidines cytosine (C), uracil (U) and thymine (T);

- Figure 1.2: hydrogen-bonded dimeric complexes, namely four 1:1 uracil-water complexes (U-Wn, n=1,2,3,4), six different hydrogen-bonded uracil dimers (U-Un, n=1,2,3,4,5,6), one of the most stable hydrogen-bonded adenine homodimer (A-A), the Watson-Crick nucleic acid base pairs adenine-uracil (A-U), adenine-thymine (A-T), guanine-cytosine (G-C) and the adenine-4-thiouracil heterodimer (A-ThioU);

- Figure 1.3: stacked dimeric complexes, namely the adenine homodimer (A$\cdots$A), the adenine-naphthalene heterodimer (A$\cdots$Naph) and the uracil homodimer (U$\cdots$U);

- Figure 1.4: uracil clusters, namely the tetramer (U4), the hexamer (U6), the heptamer (U7), the two decamers (U10a) and (U10b), and the cluster composed by fourteen uracil molecules (U14);

- Figure 1.5: other nucleic acid components for which the adsorption on the mineral brucite has been studied, namely uridine (UD), uridine 5'-monophosphate (UMP) and adenosine 5'-monophosphate (AMP).

The protocols that have been developed and the main results on this subject[170–173] obtained during the Ph.D. program have been summarized in the present Thesis as follows.
In Chapter 2, it has been reported the study of the IR spectroscopic properties of

the isolated nucleobases adenine, hypoxanthine, uracil, thymine and cytosine, shown in Figure 1.1, and some of the most stable hydrogen-bonded and stacked dimers of adenine and uracil, namely A-A, U-U3 shown in Figure 1.2 and A$\cdots$A, A$\cdots$Naph, U$\cdots$U shown in Figure 1.3. This has allowed to identify a general, reliable and effective computational procedure based on fully anharmonic quantum mechanical computations of the vibrational wavenumbers and IR intensities through the generalized second order vibrational perturbation theory (GVPT2) approach.
Since the modeling of the vibrational spectra of weakly bound molecular complexes requires to take into proper account the dispersion interactions, the first issue addressed in this work has been the validation of dispersion-corrected approaches for the anharmonic (GVPT2) computation of the IR spectra of nucleobases and their dimers.[170]
Then, in Chapter 3, the problem of improving the prediction of the vibrational properties of functional groups within hydrogen-bonded bridges has been faced, considering as model systems a set of four uracil-water complexes and six different hydrogen-bonded uracil dimers, U-Wn (n=1,2,3,4) and U-Un (n=1,2,3,4,5,6) in Figure 1.2, respectively.[173] In particular, the computational protocol described in Chapter 2 has been extended toward hybrid models[134,221,222,227,228,328] where the harmonic part of the force-field is computed at a higher level of theory.
With the aim of developing protocols to study larger and more complex systems, within hybrid schemes, another approach has also been considered to improve the description of harmonic frequencies for modes involved in hydrogen-bonding interactions through less demanding ONIOM (Our own N-layered Integrated Molecular Orbital) computations, which represent focused models where only the part of the molecular system forming the hydrogen bonds is treated at higher level of theory while the remaining part is treated at lower level.
Nevertheless, all these approaches become less and less feasible as the size of the system increases and cost-effective methods are required. In particular, to address the problem of predicting the vibrational frequencies of modes which are highly sensitive to hydrogen-bonding, like the stretching of C=O and even more N-H functional groups featuring significant red-shifts for the stretching vibrational modes and analogous blue-shifts for the out-of-plane vibrational modes,[329,330] a more general method has been developed in this work, defining and validating correction parameters for the B3LYP-D3/N07D frequencies of ν(C=O) and ν(N-H) within hydrogen-bonded bridges. This has been done through a linear regression based on the more accurate B2PLYP-D3/maug-cc-pVTZ frequencies, computed for a training set composed by the six different hydrogen-bonded uracil homodimers (U-Un, n=1,2,3,4,5,6, in Figure 1.2) and a test set of other five nucleobases homo- and heterodimers shown in Figure 1.2, namely one of the most stable hydrogen-bonded adenine homodimer (A-A), the nucleic acid base pairs adenine-uracil (A-U), adenine-thymine (A-T), guanine-cytosine (G-C) and the adenine-4-thiouracil heterodimer (A-ThioU).
Then, in Chapter 4, the computational protocol developed for the nucleic acid bases dimers has been extended to a larger system, namely the uracil heptamer (U7 in

Figure 1.4), chosen to represent the properties in the solid state. Within the VPT2 framework, with the aim of reducing the computational cost of anharmonic computations, two approaches have been considered: the hybrid model,[134] in which the harmonic frequencies are computed at a higher level of theory with anharmonic corrections obtained by less expensive methods, and the reduced dimensionality VPT2 (RD-VPT2) computations,[131,134,331] in which only selected vibrational modes are calculated anharmonically (including the couplings with the other modes), while the remaining modes are treated at the harmonic level.
The reliability of the theoretical results has allowed to extend the interpretation of the experimental findings, providing some new assignments of the IR spectrum recorded for uracil in the solid state.
In Chapter 5, the study of molecule-mineral interactions has been addressed, considering an astrobiologically relevant environment like the serpentinite-hosted hydrothermal vents. Specifically, we have studied the Lost City hydrothermal field, which has been suggested as plausible environment for the origin of life because the disequilibria and strong redox gradient established by mixing the hot and highly reducing hydrothermal fluids with the overlying cold seawater may act as driving force for the formation of prebiotic molecules.[47,48,332] In this context, we focused on the adsorption of nucleic acid components like uracil (U), uridine (UD), uridine 5'-monophosphate (UMP) and adenosine 5'-monophosphate (AMP), shown in Figure 1.5, onto magnesium hydroxide, the so called brucite ($Mg(OH)_2$), which is a product of the serpentinization process.[333] We aimed at evaluating the potential role that serpentinite-hosted hydrothermal minerals play in concentrating and selecting fundamental prebiotic molecules from dilute aqueous environments, and potentially catalyze the formation of more complex species in plausible prebiotic conditions.[334]
This study has started from the experimental investigation of the thermodynamics of the adsorption process, determining the equilibrium adsorption isotherms, which describe the distribution of molecular solute species between the fluid and solid phases. Then, the adsorption on the mineral surfaces has been modeled through surface complexation studies performed with the GEOSURF program,[335] obtaining plausible inferences for the surface complexes, discriminating between inner- and outer-sphere interactions for the adsorbates and identifying the number of surface sites involved in the reaction stoichiometry. However, to distinguish the specific molecular functional groups involved in the molecule-surface interactions, further quantum mechanical simulations are foreseen.

For more details, a list of the related published articles has been provided at the end of the Thesis.

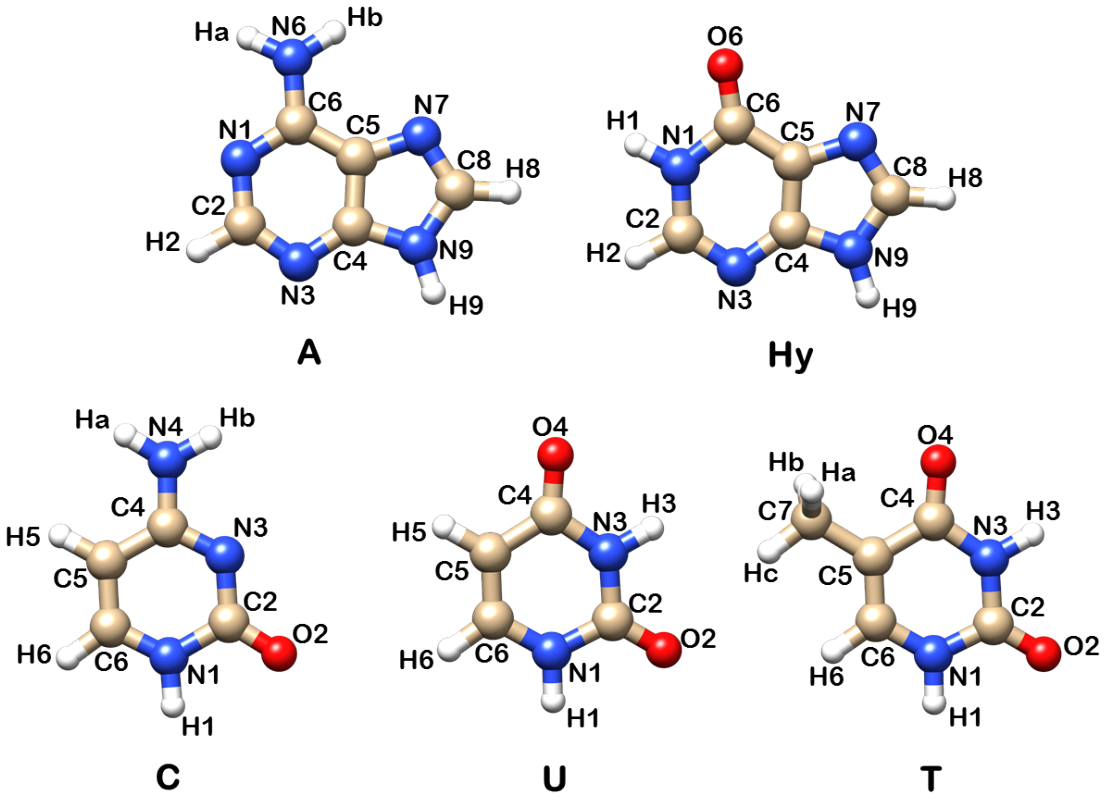

Figure 1.1: Structures and numbering schemes of the nucleobases monomers investigated in this work, namely the purines adenine (A), hypoxanthine (Hy), and the pyrimidines cytosine (C), uracil (U) and thymine (T).

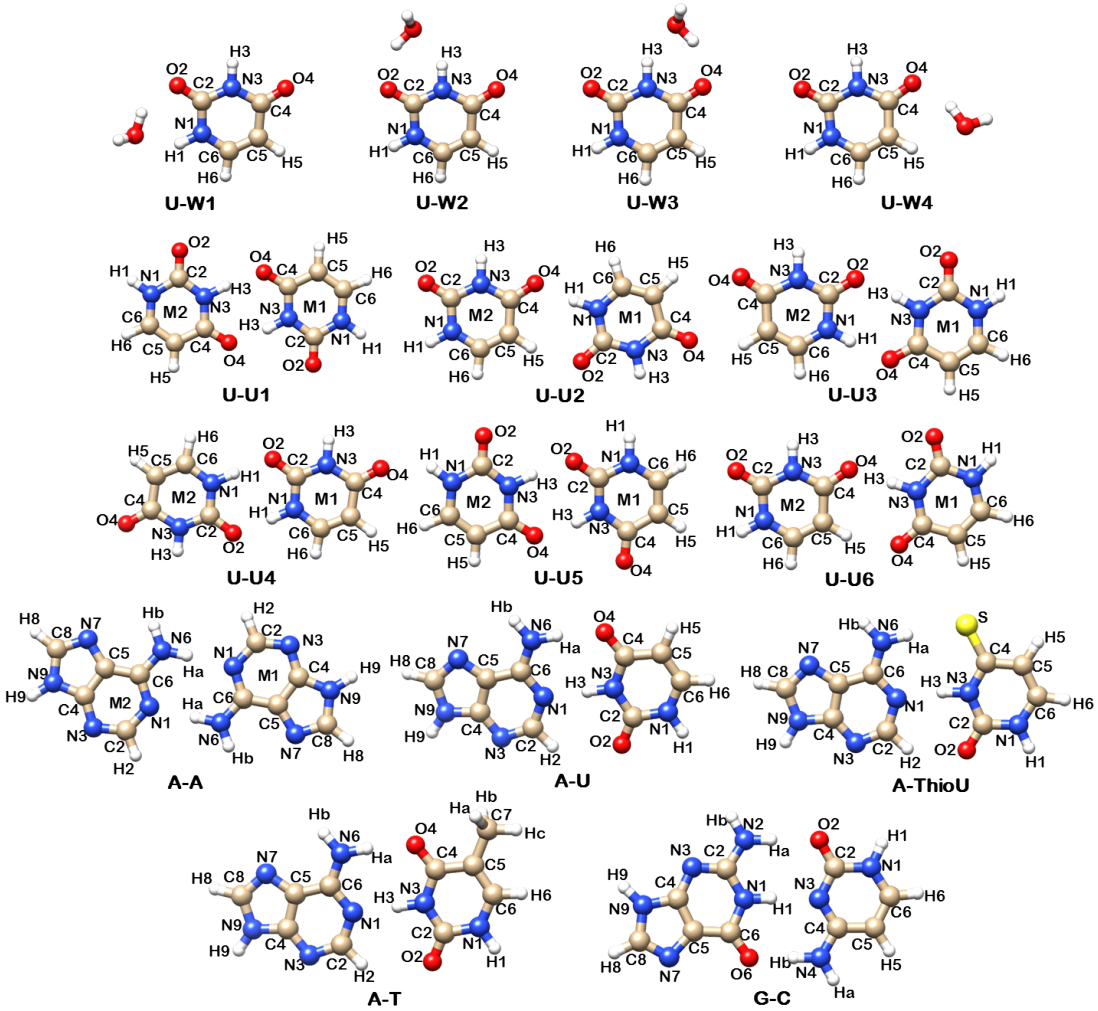

Figure 1.2: Structures and numbering schemes of all the hydrogen-bonded dimeric complexes investigated in this work, namely four uracil-water complexes (U-Wn, n=1,2,3,4), six different hydrogen-bonded uracil dimers (U-Un, n=1,2,3,4,5,6), one of the most stable hydrogen-bonded adenine homodimer (A-A), the nucleic acid base pairs adenine-uracil (A-U), adenine-thymine (A-T), guanine-cytosine (G-C) and the adenine-4-thiouracil heterodimer (A-ThioU).

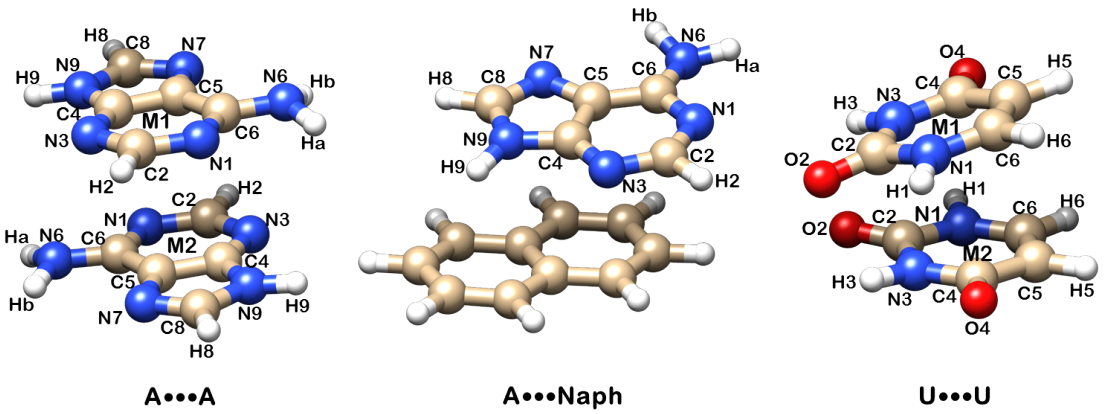

Figure 1.3: Structures and numbering schemes of the stacked dimeric complexes investigated in this work, namely the adenine homodimer (A···A), the adenine-naphthalene heterodimer (A···Naph) and the uracil homodimer (U···U).

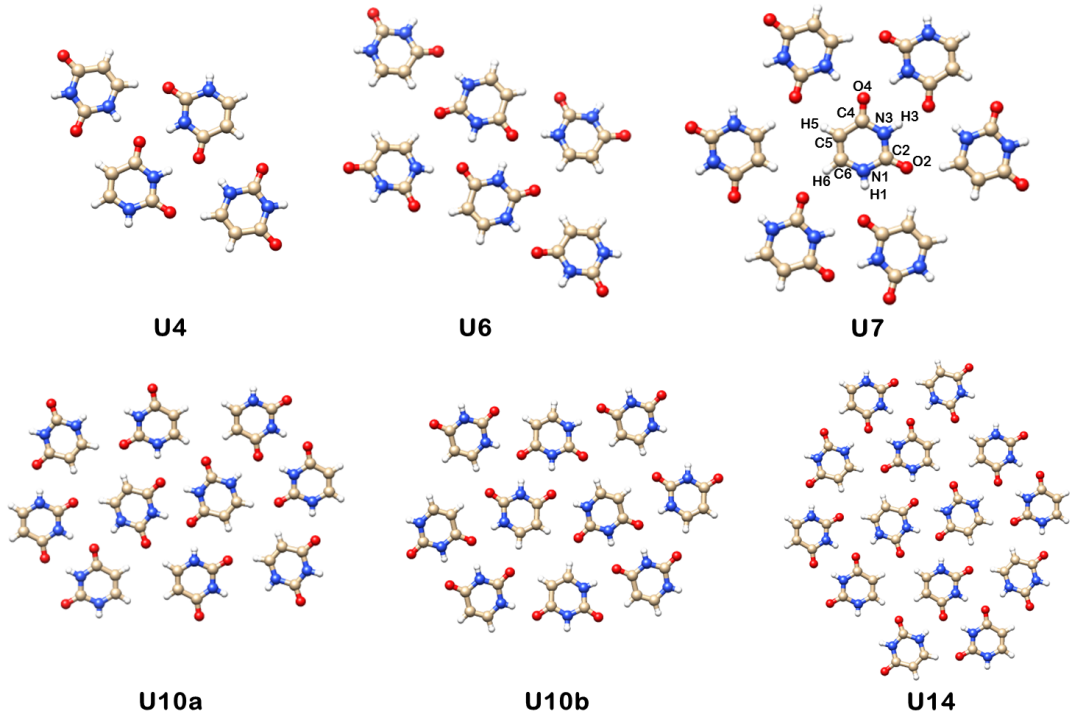

Figure 1.4: Structures of the uracil clusters investigated in this work, namely the tetramer (U4), the hexamer (U6), the heptamer (U7), the two decamers (U10a) and (U10b), and the cluster composed by fourteen uracil molecules (U14).

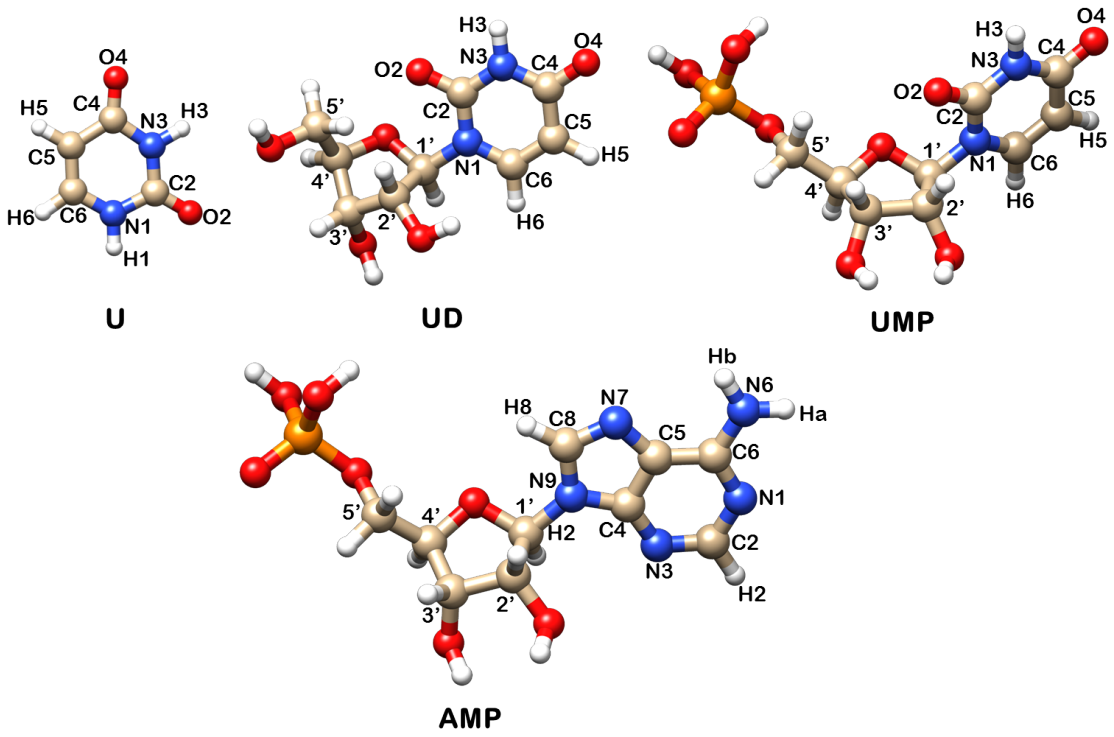

Figure 1.5: Structures of uracil (U), uridine (UD), uridine 5'-monophosphate (UMP) and adenosine 5'-monophosphate (AMP).

Chapter 2

Validation of Dispersion-corrected DFT Methods for Anharmonic Vibrational Spectra: Nucleobases and their Dimers

London dispersion interactions can be empirically defined as the attractive part of the van der Waals (vdW)-type interaction potential between atoms and molecules that are not directly bonded to each other.[336,337] The origin of such interactions is the instantaneous correlation in the motion of electrons, which induces the development of reciprocally attractive dipole moments when two electron densities get close each other.

In modeling weakly bound molecular complexes and molecules adsorbed on solid substrates, dispersion interactions must be taken into account because they could play a major role in determining the stability of the systems. Especially for the chemistry and physics of large systems, the so-called chemical accuracy can be obtained only through the inclusion of dispersion in theoretical simulations.[243] Furthermore, dispersion interactions also influence the accuracy of theoretical (reaction) thermodynamics, since they are due to electron correlation.

Therefore, it is indispensable to use theoretical methods which take into account the presence of dispersion forces. However, these kinds of interactions are non-local and non-classical and require a proper treatment of dynamic electronic correlation, which can be obtained with post-Hartree-Fock methods.[338–341] Such methods are characterized by very good chemical accuracy once coupled with extended basis sets, at the expense of the high computational cost which makes them applicable only for small molecular systems. On the other hand, cost-effective density functional theory methods do not describe correctly the long-range dispersion interactions because the density functionals are semilocal.[231,342–344]

For modeling large systems, it is extremely convenient to include an effective description of the dispersive forces into the density functional theory (DFT) approaches.

Recently, many efforts have been done in this direction.[243,246–248,250,345–358] It has been observed that this type of interactions can be modeled at a relatively low computational cost by using dispersion-corrected DFT methods and new functionals.[232–239] However, some of the most successful last-generation functionals (M06-2X[236] and ωB97X[237,238]) do not predict vibrational wavenumbers with an accuracy sufficient for spectroscopic studies.[128,134,175,240] On the other hand, addition of semi-empirical dispersion corrections to B3LYP (leading to B3LYP-D[232,233]) showed better promises for accurate computation of vibrational properties for larger weakly bound molecular systems where dispersion/stacking interactions cannot be neglected.[128,230,240]

In this work we have validated two recently developed dispersion-corrected methodologies, which had already shown good performance in predicting structural parameters and binding energies of non-covalent adducts, namely the improved version of Grimme's correction (D3[234,243]) and the DiLabio's dispersion-correcting pseudopotentials (DCP[246–250]), both in conjunction with the B3LYP functional, for the calculation of anharmonic vibrational properties of nucleobases and their dimers through the GVPT2 approach.[199,201,213] In order to identify the most reliable DFT-based dispersion-corrected model for simulating vibrational spectra, anharmonic frequencies have been compared with both B3LYP data and experimental findings, while the computed interaction energies and structures of complexes have been compared to the best available theoretical estimates.

2.1 Nucleobase Monomers

Geometries and atom numbering of the investigated nucleobases adenine, hypoxanthine, uracil, thymine and cytosine, are shown in Figure 1.1.
For validating the DFT method to be used for simulation of anharmonic vibrational spectra within the GVPT2 scheme, we have started our analysis from the uracil monomer, for which highly accurate harmonic frequencies computed at Coupled Cluster level are available.[119] These can be used to improve the estimate of the overall anharmonic frequencies within hybrid schemes,[134,222,328] where the harmonic part of the force-field, which represents the larger and less demanding one, is computed at a higher level of theory using for example the double-hybrid B2PLYP functional[225,227] or the Coupled Cluster method,[338,359] while the more computationally expensive anharmonic corrections are obtained with cost-effective approaches such as B3LYP.[224] In particular, for the uracil monomer we have tested the performance of several methods, namely B3LYP/SNSD, B3LYP-D3/SNSD, B3LYP-DCP/6-31+G(2d,2p), M06-2X/SNSD, ωB97XD/SNSD, and the hybrid B2PLYP/B3LYP and CCSD(T)/B3LYP schemes. Then, for the other nucleobases, we mainly compared the performance of the B3LYP-D3/SNSD and B3LYP-DCP/6-31+G(2d,2p) approaches.

2.1.1 DFT and Hybrid CC/DFT Methods: the case of Uracil

Anharmonic vibrational frequencies have been computed for uracil at B3LYP/SNSD, B3LYP-D3/SNSD, B3LYP-DCP/6-31+G(2d,2p), M06-2X/SNSD, ωB97XD/SNSD, and hybrid B2PLYP/B3LYP and CCSD(T)/B3LYP levels of theory, and compared with experimental frequencies in Table 2.1. The corresponding harmonic frequencies computed at the DFT level have been compared with the best theoretical estimates obtained by means of a composite scheme to account for electron correlation and basis-set effects, based on Coupled Cluster singles and doubles approximation augmented by a perturbative treatment of triple excitations [CCSD(T)][360] and correlation-consistent basis sets (aug)-cc-p(C)VnZ (n = T,Q).[361–363] The harmonic frequencies have been extrapolated to the complete basis set (CBS) limit, evaluating also corrections due to core-valence (CV) correlation and effects due to diffuse functions (aug) in the basis set at the second-order Møller-Plesset perturbation theory (MP2)[339] level. In this way, harmonic frequencies of CCSD(T)/CBS(T,Q+aug)+CV quality have been obtained.[119]
In line with previous studies, B3LYP/SNSD has provided fairly accurate anharmonic frequencies, with a Mean Absolute Error (MAE) of about 12 cm^{-1} and maximum discrepancies lower than 35 cm^{-1}. The overall result is due to the good quality of both harmonic contributions and anharmonic corrections, and not to a fortuitous cancellation of errors.
The overall accuracy has been improved by hybrid computations, with the harmonic part computed at higher levels of theory, here B2PLYP/aug-cc-pVTZ and

CCSD(T)/CBS, which in both cases lead to MAEs within 10 cm^{-1}. Moreover, the best theoretical estimates at the CC/DFT level have lowered maximum discrepancies to 25 cm^{-1}.

It is noteworthy that, although harmonic frequency computations at the CCSD(T)/CBS level are currently limited to systems with up to 10-15 atoms, the B2PLYP/B3LYP model can also be applied for larger systems.[134,229]

Concerning dispersion-corrected DFT approaches, B3LYP-D3/SNSD has provided results essentially equal to B3LYP/SNSD, for both harmonic frequencies and anharmonic corrections, so the same improvements can be obtained by hybrid models. All the other DFT models, B3LYP-DCP/6-31+G(2d,2p), M06-2X/SNSD and ωB97XD/SNSD have yielded less accurate harmonic frequencies, with MAEs of about 20 cm^{-1} and maximum discrepancies larger than 50 cm^{-1}. Moreover, for M06-2X/SNSD and ωB97XD/SNSD the quality of results has worsened for anharmonic corrections. Although for uracil the MAEs shown by M06-2X/SNSD and ωB97XD/SNSD are slightly lower than those previously reported for other systems,[128,134,175,240] both functionals have shown again non-systematic maximum deviations (MAX) larger than 100 cm^{-1}, accompanied by large errors (50-100 cm^{-1}) for several important spectral features (e.g. N-H or C=O stretching vibrations). On the other hand, the large MAE of anharmonic B3LYP-DCP/6-31+G(2d,2p) frequencies can be essentially attributed to the errors in the harmonic part.

Thus, for remaining molecular systems we have focused on B3LYP/SNSD, B3LYP-D3/SNSD and B3LYP-DCP/6-31+G(2d,2p) results, considering B3LYP/SNSD as a reference for both harmonic frequencies and anharmonic corrections.

Comparing these three DFT models with experimental data of uracil, it can be observed that B3LYP/SNSD and B3LYP-D3/SNSD approaches give better predictions than B3LYP-DCP/6-31+G(2d,2p), with MAE and MAX of 11.8 cm^{-1} and 32 cm^{-1}, 12.3 cm^{-1} and 34 cm^{-1}, and 18.7 cm^{-1} and 56 cm^{-1}, respectively. Moreover, B3LYP/SNSD and B3LYP-D3/SNSD anharmonic frequencies all agree with respect to experiments within a 20 cm^{-1} range, while larger discrepancies (over 30 cm^{-1}) have been observed only for the vibrational mode γ(N1H) at 562 cm^{-1}. For B3LYP-DCP/6-31+G(2d,2p), errors larger than 30 cm^{-1} have been observed for several vibrations in a higher frequency range (ν(C-C), ν(C=O) and ν(C5H)) and for ν(ring), δ(NH), δ(CH) at $\sim$ 1240 cm^{-1}.

Table 2.1: Harmonic and anharmonic vibrational frequencies (cm^{-1}) for uracil monomer computed at B3LYP/SNSD, B3LYP-D3/SNSD, B3LYP-DCP/6-31+G(2d,2p), M06-2X/SNSD, ωB97XD/SNSD, and hybrid B2PLYP/B3LYP and CCSD(T)/B3LYP levels of theory. Maximum absolute errors (MAE), maximum positive (MAX) and negative (MIN) deviations are reported with respect to CCSD(T) for harmonic frequencies, and experimental data[364–366] for anharmonic GVPT2 values (in bold).

Calculated	**Experimental**														
Assignment[a]	Ar matrix[b]	B3LYP		B3LYP-D3		B3LYP-DCP		M06-2X		ωB97XD		B2PLYP/B3LYP		CCSD(T)/B3LYP[c]	
		harm	GVPT2	harm	GVPT2	harm	GVPT2	harm	GVPT2	harm	GVPT2	harm	GVPT2	harm	GVPT2
ν(N1H)	3482	3639	3473	3639	3473	3643	3475	3671	3511	3691	3589	3650	3494	3653	3472
ν(N3H)	3433	3594	3427	3595	3430	3598	3429	3623	3437	3647	3545	3602	3442	3602	3430
ν(C5H)	3130	3246	3111	3246	3109	3293	3162	3280	3120	3270	3064	3266	3134	3253	3113
ν(C6H)		3205	3062	3203	3060	3252	3118	3233	3197	3229	3060	3225	3090	3218	3062
ν(C2O)	1762	1798	1771	1799	1770	1832	1801	1855	1814	1845	1822	1789	1752	1790	1771
ν(C4O)	1733	1764	1749	1765	1749	1801	1773	1828	1796	1814	1802	1754	1734	1762	1747
ν(C5C6)	1644	1673	1640	1672	1638	1728	1694	1714	1674	1710	1679	1673	1643	1678	1640
ν(ring), δ(N1H)	1473	1499	1463	1500	1463	1535	1492	1523	1491	1527	1484	1503	1464	1505	1455
δ(N3H) + δ(CH)	1401	1417	1386	1417	1386	1430	1396	1433	1409	1437	1409	1423	1406	1427	1390
ν(ring), δ(N1H), δ(N3H)	1389	1403	1372	1405	1371	1450	1418	1420	1385	1427	1405	1411	1384	1414	1367
ν(ring), δ(N3H), δ(CH)	1359	1381	1349	1383	1350	1401	1362	1394	1365	1402	1378	1388	1365	1394	1349
ν(ring), δ(NH), δ(CH)	1219	1227	1202	1230	1204	1225	1203	1238	1210	1250	1217	1235	1229	1248	1204
ν(ring), δ(NH), δ(CH)	1186	1195	1167	1199	1165	1277	1242	1208	1191	1220	1208	1204	1184	1205	1167
ν(ring), δ(CH), δ(N1H)	1076	1085	1064	1087	1062	1109	1091	1095	1090	1103	1080	1091	1065	1084	1064
δ(ring)	987	992	977	991	976	1005	979	997	982	1002	990	996	984	995	978
ν(ring), δ(N3H), δ(CH)	963	964	942	968	946	994	973	982	962	983	968	984	951	968	947
γ(CH)	958	973	949	972	948	1005	981	994	973	991	978	969	959	973	946
γ(C4O), γ(CH)	806	822	803	822	803	845	830	833	827	832	818	826	796	814	803
γ(C2O)	757	769	746	769	745	792	778	784	768	782	766	771	750	773	753
ring breathing	759	770	753	769	752	794	780	783	769	782	765	768	753	765	748
γ(CH)	718	731	713	731	712	745	732	739	741	740	726	736	716	728	714
γ(N3H)	662	689	643	687	641	699	679	685	766	689	658	688	666	670	659
γ(N1H)	562	571	530	569	528	581	577	567	597	570	550	570	556	559	558
δ(ring)	551	559	550	561	560	572	563	567	559	567	568	559	550	545	549
δ(CO)	537	542	535	542	535	555	548	547	542	551	543	541	531	541	534
δ(ring)	516	521	515	523	516	531	525	523	515	526	523	520	510	517	515
δ(CO), δ(ring)	391	387	387	389	387	396	395	399	403	402	398	401	397	388	395
τ	411	401	391	400	390	411	407	395	395	394	395	386	374	387	387
τ	185	170	163	170	162	170	168	168	163	168	161	167	154	159	165
τ		155	147	154	146	153	152	154	151	154	151	149	128	140	147
MAE		8	**12**	8	**12**	25	**19**	17	**19**	18	**24**	7	**8**		**10**
MIN		-21	**-32**	-18	**-34**	-24	**-17**	-10	**-22**	2	**-66**	-13	**-37**		**-24**
MAX		18	**16**	16	**16**	72	**56**	67	**104**	55	**112**	18	**12**		**14**

[a]Abbreviations: ν = stretching; δ = in-plane bending; γ = out-of-plane bending; τ = torsional; sciss = scissoring; rock = rocking; wagg = wagging; asym = asymmetric; sym = symmetric. [b]From Ref.[364–366] [c]Best estimated harmonic harmonic frequencies evaluated by the composite scheme at CCSD(T)/CBS(T,Q + aug) + CV level from Ref.[119]

2.1.2 Dispersion-Corrected B3LYP-based Approaches

A statistical analysis of the deviations of computed vibrational frequencies for all nucleobases with respect to experimental data and B3LYP results has been presented in Table 2.2. Inspection of Table 2.2 shows that B3LYP/SNSD and B3LYP-D3/SNSD results agree fairly well with experiments, with MAEs of about 11 cm^{-1} for the whole set of molecules and maximum positive (MAX) and negative (MIN) discrepancies not exceeding 30 cm^{-1} and 41 cm^{-1}, respectively. It should be also noted that larger discrepancies observed for N-H stretching vibrations can be attributed to red-shifts of about 20 cm^{-1} due to the Argon matrix environment. The relative errors for B3LYP-DCP/6-31+G(2d,2p) are significantly larger, with MAE, MAX and MIN of 23 cm^{-1}, 72 cm^{-1} and 80 cm^{-1}, respectively. The accuracy of B3LYP results confirms that it can stand as a reference for comparison between the two dispersive methods B3LYP-DCP and B3LYP-D3.
Direct comparison between the theoretical models has allowed us to analyze different contributions to each overall anharmonic frequency, namely the harmonic part and the anharmonic correction. The accuracy of the latter is of particular interest as the harmonic part can be corrected by more expensive computations within hybrid schemes, as shown above for the case of uracil.
It emerges that harmonic and anharmonic frequencies obtained using B3LYP/SNSD and B3LYP-D3/SNSD methods are nearly equivalent. On the other hand, major deviations appear in the case of B3LYP-DCP/6-31+G(2d,2p) frequencies, with weighted mean absolute errors of 22 cm^{-1} and 26 cm^{-1}, respectively, for the harmonic and anharmonic frequencies of all the nucleobases; a small deviation (6 cm^{-1}) of the anharmonic shifts is, instead, observed. On the whole, B3LYP/SNSD and B3LYP-D3/SNSD approaches provide more accurate vibrational wavenumbers with respect to B3LYP-DCP/6-31+G(2d,2p). In particular, the high accuracy of B3LYP, fully sufficient for the analysis of experimental spectra, is retained by computations with B3LYP-D3. However, even though less accurate than B3LYP-D3, B3LYP-DCP computations perform significantly better than most of the other dispersion-corrected DFT models[128,134] and, most importantly, provide reliable anharmonic corrections.

As an example, the comparison between simulated and experimental[127] infrared spectra of the adenine molecule has been shown in Figure 2.1. The difference between the full experimental spectrum (EXP) and the one resulting from the fundamental transitions only (EXP FUNDAM) has also been highlighted.
The experimental spectroscopic features are reproduced quite well by the B3LYP/SNSD and B3LYP-D3/SNSD anharmonic calculations, taking into account also the presence of relatively intense nonfundamental transitions, which could not be obtained by scaling the harmonic frequencies. B3LYP-DCP/6-31+G(2d,2p) also provides a rather complex spectrum but the band positions show larger deviations from experimental values and different band patterns.
For more details, the interested reader is referred to the original article.[170]

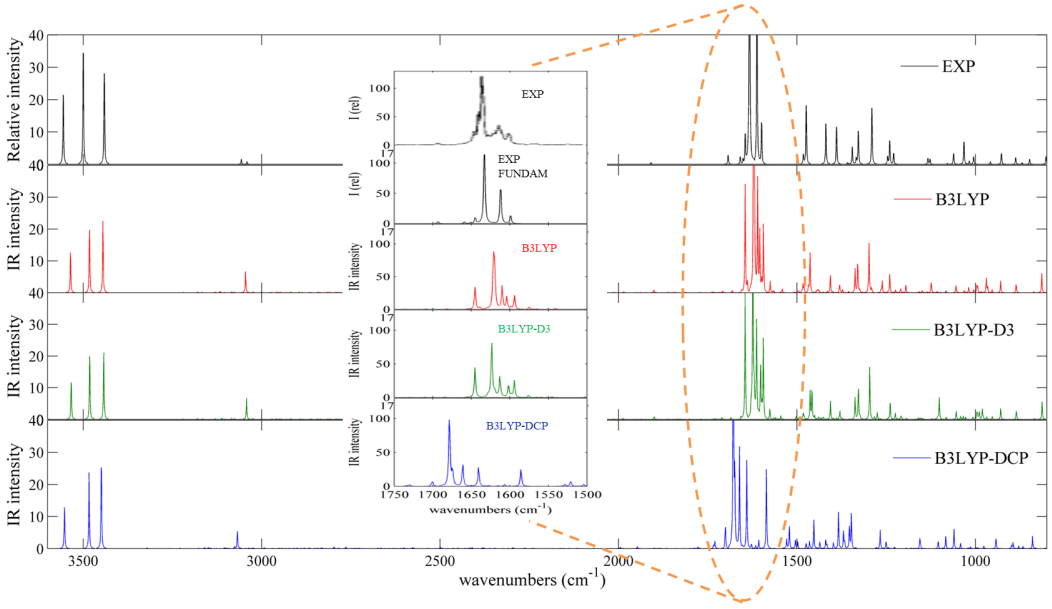

Figure 2.1: Computed anharmonic and experimental infrared spectra of isolated adenine molecule in the 800-3600 cm^{-1} spectral range. Experimental IR spectrum recorded in the low-temperature Ar matrix has been generated using the data of Table 3 of ref.[127] IR spectra line-shapes (both theoretical and experimental) have been convoluted using Lorentzian functions with a half-width at a half-maximum (HWHM) of 1 cm^{-1}. The inset shows the 1500-1700 cm^{-1} spectral range, with both the experimental spectrum obtained by convolution of fundamental vibrational modes (EXP FUNDAM) and the experimental spectrum shown in Fig. 3 of ref.[127](EXP).

Table 2.2: Weighted maximum absolute errors (MAE), maximum positive (MAX) and negative (MIN) deviations of computed vibrational frequencies (cm^{-1}) for the whole set of monomeric nucleobases investigated in this work, namely adenine, hypoxanthine, uracil, thymine and cytosine, with respect to experimental data[a] and B3LYP results.

	Calculated vs Experimental[a]	**Calculated vs B3LYP**		
	Δanh[b]	Δharm[c]	Δanh[b]	ΔGVPT2[d]
	MAE			
B3LYP	12.3			
B3LYP-D3	12.5	1.4	1.7	1.5
B3LYP-DCP	23.6	21.6	26.4	6.1
	MIN			
B3LYP	-46			
B3LYP-D3	-46	-14	-17	-19
B3LYP-DCP	-80	-103	-102	-23
	MAX			
B3LYP	27			
B3LYP-D3	28	15	18	20
B3LYP-DCP	72	82	75	52

[a]From Ref.[119,127,128,364–370] [b]Deviation between anharmonic frequencies. [c]Deviation between harmonic frequencies. [d]Deviation of anharmonic shift for the selected method respect to B3LYP method.

2.2 Hydrogen-Bonded and Stacked Dimers of Adenine and Uracil

Geometries and atom numbering of the dimers of adenine and uracil investigated in this work are labeled as A-A, U-U3 in Figure 1.2 and A$\cdots$A, A$\cdots$Naph, U$\cdots$U in Figure 1.3.
The structures and energies of all dimers considered in the present study have been previously evaluated by highly accurate computational strategies, mainly by combination of MP2 and Coupled Cluster (CCSD(T)) approaches along with extrapolation to the complete basis set limit (CBS),[371–374] allowing the comparison and assessment of the performance of less computationally demanding B3LYP-based models.
In particular, counterpoise-corrected binding energies (ΔE_{bind}) of hydrogen-bonded and stacked dimer structures have been calculated and compared to reference values[371–374] in Figure 2.2. Structural parameters have been compared as well, and the percentage MAEs of rotational constants of dimer structures are shown in Figure 2.2. Results show that B3LYP calculations fail for stacked dimers, as expected for systems that mainly interact through dispersion forces. Instead, both B3LYP-D3 and B3LYP-DCP models predict reliable binding energies and structural parameters with similar good accuracy. As shown in Figure 2.2, even in the case of hydrogen-bonded structures, B3LYP shows the largest deviations for both binding energies and structural parameters with respect to the best theoretical estimates,[371–374] while B3LYP-D3 and B3LYP-DCP show very good performances and provide rather similar values, making both approaches of essentially equal accuracy in predicting structural parameters and binding energies.

Regarding vibrational properties, unfortunately, reliable reference data are not available in the literature. To the best of our knowledge, anharmonic computations at the Coupled Cluster level for the systems considered in this work (except the uracil molecule) have not been performed yet. Additionally, experimental spectroscopic studies of weakly bound nucleobase complexes or similar systems are still rather scarce[375,376] and, most importantly, results might be biased by interpretations, thus being not fully adequate for benchmark purposes. For a proper benchmark reference it is necessary to combine structure identification of the complexes (by means, e.g., of rotational spectra) with an analysis of the vibrational transitions. To the best of our knowledge such experimental studies for the nucleobase complexes are not available and the present work could be useful for supporting future investigations. Moreover, the few experimental data available for dimers are largely related to the X-H stretching vibrations, at high wavenumbers, which are more difficult to describe at the DFT level. Thus, benchmark studies based on the high-frequency spectral zone would unbalance the accuracy of the lower-wavenumber regions, which are relevant as well for the current project. For these reasons, the validation of the performance of these dispersion-corrected models in predicting vibrational properties has been carried out by comparison with well-established data for monomers,

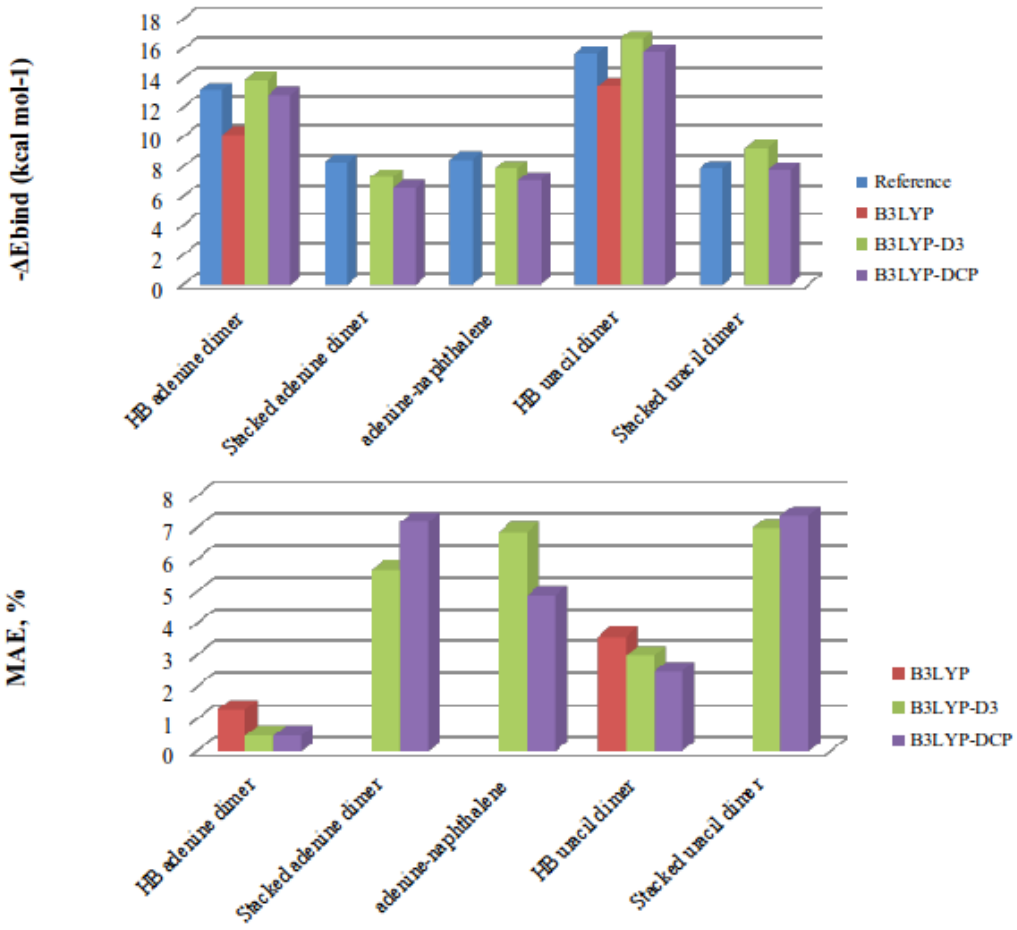

Figure 2.2: Counterpoise-corrected binding energies (ΔE_{bind})of hydrogen-bonded and stacked dimer structures in kcal mol^{-1} and percentage mean absolute errors of rotational constants (MAE, %) of optimized dimer structures computed with B3LYP/SNSD, B3LYP-D3/SNSD and B3LYP-DCP/6-31+G(2d,2p) methods, compared to reference values.[371–374]

and only a critical comparison between both models has been performed.
A detailed statistical analysis of the deviations of the harmonic and anharmonic frequencies computed with B3LYP-D3 and B3LYP-DCP methods for all the investigated dimers is presented in Table 2.3. The mean absolute difference (MAD) has been computed considering all normal modes, while MIN and MAX have been evaluated by excluding 10 of the 341 vibrational modes exhibiting exceptionally large discrepancies. Comparison of harmonic and anharmonic frequencies obtained with both dispersive methods (Table 2.3) indicates a MAD of about 20 cm^{-1} for the deviations between the harmonic frequencies, about 30 cm^{-1} for the deviations between the anharmonic frequencies and about 10 cm^{-1} for the deviations of anharmonic shifts. It can be noted that the relative performance of these two methods in calculating vibrational frequencies in the case of the dimers is the same observed for monomers (with respective MADs again of about 20 cm^{-1}, 30 cm^{-1} and 10 cm^{-1}). Absolute maximum deviations of 65 cm^{-1}, 67 cm^{-1} and 51 cm^{-1} have been obtained for harmonic frequencies, anharmonic frequencies and the anharmonic shifts, respectively, excluding 10 of the 341 vibrational modes which exhibit exceptionally large discrepancies. The vibrations excluded from the statistical analysis are torsional vibrational modes of the amino group and out-of-plane bendings of NH and CH groups, which are quite hard to model due to both PES requirements and limitations of perturbative treatment. However, as it can be inferred from the MADs evaluated taking into account all the vibrational modes, the role of such modes is negligible in the weighted average, justifying their exclusion in the evaluation of MAX and MIN.

Table 2.3: Maximum absolute difference (MAD), maximum positive (MAX) and negative (MIN) deviations of harmonic and anharmonic vibrational frequencies (cm^{-1}) for uracil and adenine dimers computed with B3LYP-DCP/6-31+G(2d,2p) respect to the B3LYP-D3/SNSD method.

	MAD			MIN[d]			MAX[d]		
	harm[a]	anh[b]	GVPT2[c]	harm[a]	anh[b]	GVPT2[c]	harm[a]	anh[b]	GVPT2[c]
U-U	20.3	31.7	14.9	-6	-2	-22	60	65	51
U···U	20.6	28.1	13.5	-15	-16	-20	55	59	14
A-A	21.4	29.2	13.2	-63	-63	-37	47	63	46
A···A	22.1	22.2	3.3	-1	-3	-16	53	55	14
A···Naph	24.2	25.4	5.9	-65	-65	-23	52	67	33
all	21.5	27.3	10.3	-65	-65	-37	60	67	51

[a]Deviation between harmonic frequencies. [b]Deviation between anharmonic frequencies. [c]Deviation of anharmonic shift. [d]Evaluated by excluding 10 of 341 vibrational modes.

Given the better performance of B3LYP-D3 in calculating vibrational frequencies for the monomeric nucleobases, this method has been chosen to compare IR spectra of the monomers and the dimers in order to investigate the effect of the intermolecular interactions on the vibrational frequencies of nucleobases.
In particular, as case study, the anharmonic infrared spectra of adenine monomer and dimers, computed at the B3LYP-D3 level, have been compared in Figure 2.3.

ΔHB and ΔSTACK denote the largest anharmonic frequency shifts of the hydrogen-bonded and stacked dimers with respect to the isolated monomer, respectively. Detailed anharmonic frequency results for adenine and uracil dimers can be found in the original work.[170]

Figure 2.3 points out that the effects of the intermolecular interactions on IR spectra concern both frequency shifts and intensity changes.

In the hydrogen-bonded adenine dimer, the most significant shifts of anharmonic frequencies concern some vibrational modes of the amino group, which is indeed involved in the hydrogen-bonding interaction. These are: $\nu_{sym}(NH_2)$ at 2965 cm^{-1} with a considerable shift of $\sim$-500 cm^{-1} and a large intensity increase giving the most intense band in the spectrum, $\delta_{sciss}(NH_2)$, ν(C6N6) at 1495 cm^{-1} with a shift of -70 cm^{-1}, $\gamma_{wagg}(NH_2)$, and τ(Rr) at $\sim$152 cm^{-1} with a shift of $\sim$-53 cm^{-1}. It is noteworthy that all these shifts are negative indicating that hydrogen bonds weaken the force constants of the amino group for such vibrations. Other bands which gather intensity include the $\nu_{asym}(NH_2)$ vibrational mode at 3502 cm^{-1}, the ν(C8H) vibrational mode at 3101 cm^{-1}, the ν(C2H) vibrational mode at 3030 cm^{-1}, the $\delta_{sciss}(NH_2)$, ν(C5C6), ν(C6N6) vibrational mode at 1647 cm^{-1} and 1592 cm^{-1}, the $\delta_{sciss}(NH_2)$, ν(C4C5), ν(C5C6) vibrational mode at 1583 cm^{-1}, and the ν(C2N3), ν(N1C2) vibrational mode at 1311 cm^{-1}.

In the stacked adenine dimer, the main anharmonic frequency shifts involve out-of-plane vibrational modes, such as $\tau(NH_2)$, γ(N9H) at 425 cm^{-1} with a shift of -66 cm^{-1} and $\gamma_{wagg}(NH_2)$ at $\sim$280 cm^{-1} with a shift of $\sim$77 cm^{-1}. The most intense band in the spectrum of the dimer corresponds to the δ(C8H), ν(N7C8), δ(N9H) vibrational mode at 1242 cm^{-1}, which acquires intensity with respect to the case of the isolated adenine molecule.

Then, the stacked adenine-naphthalene heterodimer may be considered as a first step in the investigation of the interactions between nucleobases and solid substrates. Therefore, we have examined how the naphthalene support affects the vibrational spectroscopic features of the adenine molecule. In particular, the most relevant anharmonic frequency shift concerns the $\gamma_{wagg}(NH_2$, τ(Rr) vibrational mode of the adenine molecule at 330 cm^{-1} with a shift of 125 cm^{-1}, which also shows an intensity increase. Other shifts are observed for the $\nu_{asym}(NH_2)$ vibrational mode at 3520 cm^{-1} (-14 cm^{-1}), ν(N9H) vibrational mode at 3463 cm^{-1} (-19 cm^{-1}), $\nu_{sym}(NH_2)$ vibrational mode at 3425 cm^{-1} (-18 cm^{-1}), γ(C8H) vibrational mode at 824 cm^{-1} (+11 cm^{-1}), γ(C8H), τ(R), τ(r), γ(C6N6) vibrational mode at 810 cm^{-1} (+30 cm^{-1}), and γ(N9H), γ(C2H), τ(R), τ(r) vibrational mode at 567 cm^{-1} (+23 cm^{-1}) which has vanishing intensity in the spectrum of the isolated nucleobase, but gather intensity when adenine interacts with naphthalene, δ(R), γ(N9H) vibrational mode at 523 cm^{-1} (+18 cm^{-1}) which leads to a remarkable intensity increase giving rise to the most intense band in the spectrum of the adenine-naphthalene heterodimer, $\tau(NH_2)$, γ(N9H) vibrational mode at 518 cm^{-1} (+20 cm^{-1}) which is one of the most intense bands in the spectrum of the isolated nucleobase, but decreases its intensity when adenine interacts with naphthalene, γ(N9H) vibrational mode at 522 cm^{-1}

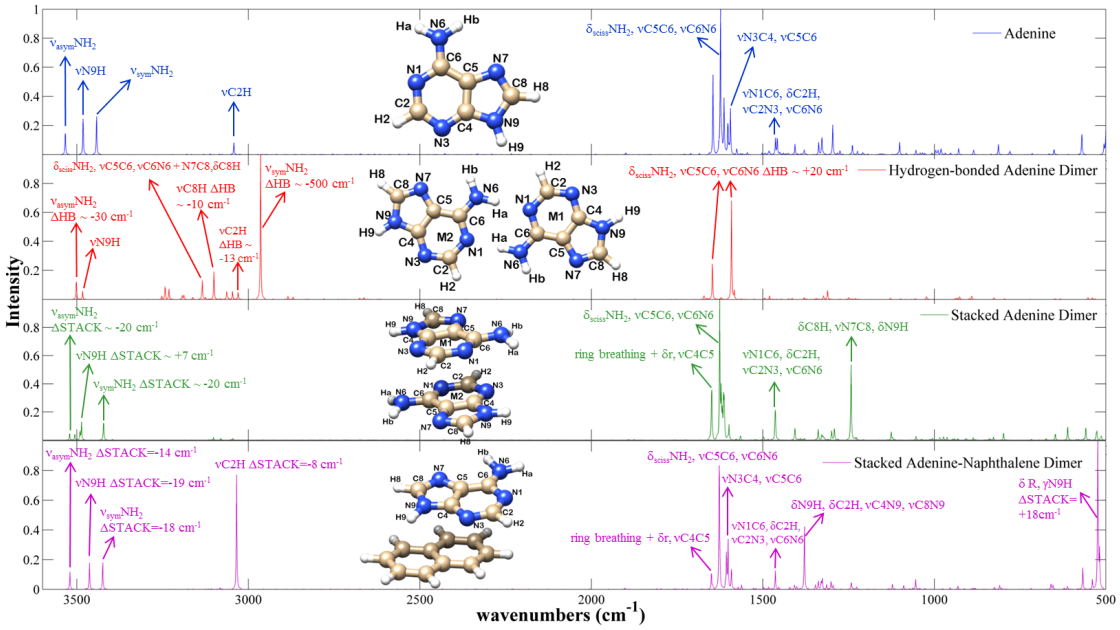

Figure 2.3: Anharmonic IR spectra of adenine dimers computed with B3LYP-D3/SNSD method, compared with IR spectrum of isolated adenine molecule in the 500-3600 cm^{-1} spectral range, along with the assignment of the most pronounced bands. Theoretical spectra line-shapes have been convoluted with Lorentzian functions with a half-width at half-maximum (HWHM) of 1 cm^{-1}.

(+31 cm^{-1}) which shows also an intensity increase, $\tau(NH_2)$, δ(R) vibrational mode at 461 cm^{-1} (-25 cm^{-1}) which decreases in intensity, τ(Rr), $\gamma_{wagg}(NH_2)$ vibrational mode at 218 cm^{-1} (-21 cm^{-1}) which is subjected to a significant decrease, and $\gamma_{wagg}(NH_2)$, τ(Rr) vibrational mode at 168 cm^{-1} (-37 cm^{-1}) which is subjected to a significant decrease. Other bands which gather intensity when adenine interacts with naphthalene include the ν(C2H) vibrational mode at 3034 cm^{-1} which becomes the second most intense band in the spectrum, the ν(N1C6), δ(C2H), ν(C2N3), ν(C6N6) vibrational mode at 1464 cm^{-1} and the δ(N9H), δ(C2H), ν(C4N9), ν(C8N9) vibrational mode at 1379 cm^{-1} which becomes particularly intense.
As expected, comparison of the IR spectra of the monomer and the dimers shows important shifts of the vibrational frequencies and/or IR intensity variations for specific functional groups. These correspond to proton donor and acceptor moieties which interconnect the molecules through intermolecular hydrogen bonds. Even in the stacked structures and in the heterodimer, significant changes in the spectral features have been noticed for some out-of-plane vibrational modes, which are mostly influenced by the stacked configuration.

2.3 Concluding Remarks

Our investigation indicates that spectra simulated using dispersion-corrected B3LYP approaches may be used confidently to interpret experimental data for nucleobase complexes. In particular, the two dispersion-corrected methods, B3LYP-D3/SNSD and B3LYP-DCP/6-31+G(2d,2p), are capable of predicting reliable values for binding energies and structural parameters, for both hydrogen-bonded and stacked structures. For vibrational frequencies, B3LYP-D3/SNSD yields more accurate results, yet B3LYP-DCP/6-31+G(2d,2p) outperforms several other dispersion-corrected DFT approaches and provides reliable anharmonic corrections. However, considering also its applicability for essentially all elements of the periodic table, B3LYP-D3/SNSD can be suggested as the method of first choice.

Chapter 3

Improvements of Vibrational Properties for Functional Groups involved in Hydrogen Bonds: Uracil-Water Complexes and Nucleic Acid Base Dimers

Hydrogen bonds represent key inter- and intramolecular interactions in biology, responsible for stabilizing biological macromolecules and for driving many fundamental processes like protein folding and misfolding, DNA replication, transcription and translation, where the base pairing through hydrogen bonds provides the coding mechanism.[377–380]

Hydrogen-bonding interactions might have been among the driving forces of prebiotic self-organization and compartmentalization.[87,94,103,171,319,381,382] Several examples of self-assembly processes have been reported in the literature like the spontaneous formation of ordered bilayer membranes from amphiphilic complexes of complementary hydrogen bond pairs, which in turn may spontaneously arrange into closed membranous vesicles.[381–383] Such self-assembled structures might have been the primordial membranes of the first cellular lifeforms, produced from amphiphilic compounds on prebiotic Earth.[384,385]

Processes driven by hydrogen-bonding interactions have been proposed also as plausible evolutionary pathways toward contemporary nucleic acids, based on supramolecular noncovalent assembly of nucleobases or some plausible building blocks of pre-RNAs.[4,320] Indeed, thanks to the formation of hydrogen bonds, nucleobases feature a great capability of self-assembling, forming dimers, clusters, up to monolayers adsorbed on solid substrates or more complex three-dimensional structures.[87,319]

Infrared spectroscopy techniques are particularly suitable for studying hydrogen-bonding interactions, since remarkable effects on vibrational frequencies and IR intensities have been observed when functional groups are involved in hydrogen

bonds.[122–124,170–173,327] Reliable theoretical description of the spectroscopic features of hydrogen-bonded systems would be highly desirable to facilitate the analysis of experimental outcomes which, in many cases, is hampered by huge changes in the IR spectra, like shifts of the order of magnitude of hundreds of cm^{-1} and significant increases in the intensity of IR bands of vibrational modes for functional groups directly involved in the hydrogen-bonded bridges.[170,172,173]

In the last years computational spectroscopy has become an effective tool to analyze and assign IR spectra of molecular systems of increasing size and complexity.[109,111,132,134,168,169,321–325,386] Nevertheless, the accurate prediction of vibrational frequencies of functional groups within strong hydrogen-bonded bridges is still an important challenge because of the high computational demand of the vibrational treatment, due to both enhanced anharmonicity and accuracy requirements of the underlying potential energy surface (PES).[326] For these reasons, our group put effort into the development of comprehensive yet feasible computational protocols for accurately predicting IR spectroscopic features of hydrogen-bonded systems. Specifically, we have resorted to the computational procedure previously validated for predicting structural parameters, binding energies and IR spectroscopic properties of isolated nucleobases and some of the most stable hydrogen-bonded and stacked dimers of adenine and uracil, which is based on fully anharmonic quantum mechanical computations within the GVPT2 approach, combined with the B3LYP-D3 method, in conjunction with basis sets of double-ζ quality.[170] In order to improve the prediction of the vibrational properties of functional groups within hydrogen-bonded bridges, such computational procedure has been used to study the hydrogen-bonding effects on the IR spectra of uracil-water complexes and hydrogen-bonded uracil dimers.[173] From the computational point of view, these systems are small enough to allow the use of accurate quantum mechanical methods for investigating vibrational properties beyond the harmonic approximation.

In a first step, two routes have been followed to achieve improved description of the overall anharmonic frequencies, : (i) a hybrid scheme[134,222,328] in which the harmonic part of the force-field is computed at a higher level of theory, namely with the double-hybrid B2PLYP functional[225,227] in conjunction with maug-cc-pVTZ basis set;[387] (ii) a ONIOM B2PLYP:B3LYP focus model, where only the part of the molecular system forming the hydrogen bonds is treated at the B2PLYP level of theory.

The use of hybrid schemes is legitimated by the observation that anharmonic shifts computed at different levels of theory, like CC and DFT, are usually very close.[134] In addition, anharmonic shifts are small corrections with respect to the harmonic part of the whole vibrational frequency and, at the same time, represent the computationally most demanding part of the calculations, requiring the evaluation of the numerical third and fourth derivatives of the PES and the treatment of the Fermi resonances. Therefore, a possible route to decrease the computational cost retaining a high accuracy is to use two different methods for the evaluation of the harmonic frequencies and the anharmonic shifts. This can be performed, in the first approxi-

mation, through a very simple and user friendly procedure, not requiring any code implementation, running two separate calculations, one for the harmonic part using larger basis sets or more accurate methods, and one for the anharmonic corrections using a more approximate method, and then combining the two parts *a posteriori* for each frequency of interest. A more refined hybrid approach can be also employed, including the harmonic frequencies computed at higher level of theory directly into the VPT2 treatment at low level, obtaining a hybrid anharmonic force field, which allows a more consistent treatment of the Fermi resonances, higher accuracy and the automatic application of the hybrid model to all the frequencies of overtones and combination bands. Nevertheless, such an approach can be too cumbersome in the case of large molecular systems with many vibrational modes since it requires the assignment of all vibrational modes, while the simple *a posteriori* combination of harmonic frequencies and anharmonic shifts computed at different levels of theory can be preferred when interested in few vibrational modes of the overall system.
The ONIOM approach is particular useful for localized spectroscopic phenomena. ONIOM is the acronym of Our own N-layered Integrated Molecular Orbital, which indicates an approach to describe complex molecular systems by resorting to multi-scale or "focused" models, where the system is separated in multiple layers and the target structure is treated at a higher level of theory while the rest of the system is described with less accurate but computationally less demanding approaches.[388–395] In this work such computational methodologies have been validated for describing the effect of hydrogen-bonding interactions on the IR spectroscopic features, considering as reference the experimental frequencies available for the set of four uracil-water complexes. Then, fully anharmonic infrared spectra of six hydrogen-bonded uracil dimers have been simulated, highlighting the influence of different bonding patterns on the overall IR spectra line-shapes and the fingerprint features of the specific hydrogen-bonding interactions.
However, it is worth noting that all these approaches become less and less feasible as the size of the system increases. Therefore, a third route has been explored in order to develop more general cost-effective methods. In particular, to address the problem of the toughness in predicting the vibrational frequencies of modes which are particularly sensitive to hydrogen-bonding, like the stretching of C=O and especially N-H functional groups,[329,330] we have defined and validated correction parameters for the B3LYP-D3/N07D frequencies of ν(C=O) and ν(N-H) within hydrogen-bonded bridges. This has been done through a linear regression based on the more accurate B2PLYP-D3/maug-cc-pVTZ frequencies, computed for a training set composed by the six different hydrogen-bonded uracil homodimers (U-Un, n=1,2,3,4,5,6, in Figure 1.2) and a test set of other five nucleobases homo- and heterodimers shown in Figure 1.2, namely one of the most stable hydrogen-bonded adenine homodimer (A-A), the nucleic acid base pairs adenine-uracil (A-U), adenine-thymine (A-T), guanine-cytosine (G-C) and the adenine-4-thiouracil heterodimer (A-ThioU).

3.1 Validation of the B2PLYP-D3 Method: the 2-Thiouracil and the Dimers of Acetamide and Formamide

The performance of the double hybrid density functional B2PLYP in predicting harmonic frequencies and anharmonic corrections has been validated for a wide set of small- to medium-size molecules in comparison with Coupled Cluster calculations and experiments.[227] For instance, a recent study about the anharmonic vibrational properties of molecules containing C≡N moieties,[141] for which a comparison with high-resolution experimental data is available, demonstrates that this functional is more robust than B3LYP to get accurate predictions for such puzzling vibrational modes without any ad hoc scaling factor. From these studies it emerges that the double hybrid B2PLYP functional, in conjunction with basis sets of double- to quadruple-ζ quality, is able to provide very accurate harmonic frequencies, while the anharmonic corrections usually feature the same accuracy of B3LYP/N07D at significantly larger computational cost. Therefore, a good compromise between accuracy and computational cost can be obtained by hybrid approaches, where the harmonic part of vibrational frequencies computed at B2PLYP level of theory is combined with anharmonic corrections evaluated by means of less demanding methods like B3LYP.[227]
Concerning the choice of the basis set, the maug-cc-pVTZ one,[387] in which d functions on hydrogens have been removed, appears to be the most suitable for B2PLYP computations, based on previous experience about increased basis-set requirements.[227–229,396,397]

We have validated the B2PLYP-D3/maug-cc-pVTZ method as the reference approach for computing harmonic vibrational frequencies, using smaller size systems as models for nucleobases complexes, namely 2-thiouracil, acetamide, formamide, and systems featuring hydrogen-bonding interactions like the dimers of acetamide and formamide, for which experimental data are available.[126,213,398–400] The structures of these molecular systems are shown in Figure 3.1.
For the 2-thiouracil molecule, geometry optimization followed by harmonic and anharmonic vibrational calculations has been performed at B2PLYP-D3/maug-cc-pVTZ[225,227] and B3LYP-D3/N07D[234,243–245] levels of theory and the results have been compared with experimental data reported in the literature.[126,213]
Table 3.1 collects the harmonic and anharmonic vibrational frequencies and IR intensities of 2-thiouracil computed with B2PLYP-D3/maug-cc-pVTZ and B3LYP-D3/N07D, and the hybrid anharmonic frequencies obtained using the B2PLYP-D3/maug-cc-pVTZ level for the harmonic part and the B3LYP-D3/N07D for the anharmonic corrections, in comparison with experiments. Data relative to ν(N-H) and ν(C=O) are highlighted in bold. From such results it emerges that B2PLYP-D3/maug-cc-pVTZ is able to improve the harmonic part of the vibrational frequency for ν(C=O) yet slightly improving the ν(N-H), while providing the same accuracy of B3LYP-D3/N07D for anharmonic shifts.

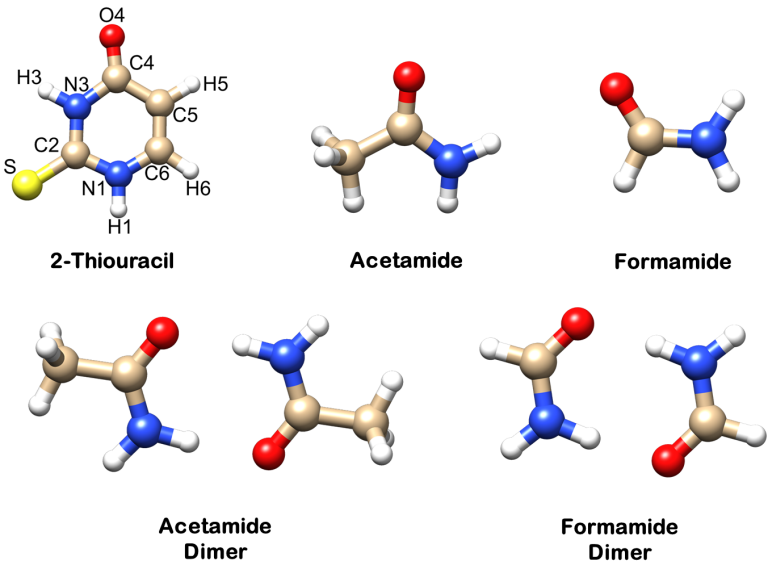

Figure 3.1: Structures of 2-thiouracil, acetamide, formamide, and the dimers of acetamide and formamide.

Since the anharmonic correction represents the computationally most demanding part of the calculations, it is more advantageous to adopt hybrid schemes where this correction is computed at a lower level of theory like B3LYP-D3/N07D. With such a hybrid approach, it is possible to obtain very good predictions both for ν(C=O) and ν(N-H), with mean absolute errors with respect to experiment of about 8 cm^{-1} and 2 cm^{-1}, respectively, instead of about 14 cm^{-1} and 13 cm^{-1} in the case of full anharmonic B3LYP-D3/N07D computations. This further confirms the accuracy of the anharmonic shifts computed with the B3LYP-D3/N07D method and points out the importance to use more accurate harmonic frequencies, like those B2PLYP-D3/maug-cc-pVTZ, for a quantitative comparison with experiments.
On the whole, the MAEs obtained for the anharmonic frequencies computed at B2PLYP-D3/maug-cc-pVTZ and hybrid B2PLYP-D3/B3LYP-D3 levels of theory are very small, i.e. 6.4 cm^{-1} and 5.5 cm^{-1}, respectively, slightly better than the one predicted by B3LYP-D3/N07D that is 7.5 cm^{-1}. Moreover, the intensities computed with B2PLYP-D3/maug-cc-pVTZ reproduce pretty well the trend of IR intensities observed in the experimental spectrum. Therefore, B2PLYP-D3/maug-cc-pVTZ can be generally considered a reliable method for the prediction of vibrational properties and, additionally, shows a better performance with respect to B3LYP-D3/N07D for vibrations like the stretching of C=O and N-H groups which are mainly influenced by hydrogen-bonding interactions.

Furthermore, we have validated the B2PLYP-D3/maug-cc-pVTZ method also for calculating the vibrational properties of systems featuring hydrogen-bonding in-

Table 3.1: Harmonic and anharmonic vibrational frequencies and IR intensities of 2-thiouracil, computed with B2PLYP-D3/maug-cc-pVTZ and B3LYP-D3/N07D methods, and the hybrid B2PLYP-D3/B3LYP-D3 anharmonic frequencies, in comparison with experiments.[126]

Mode	Exp[b]		B2PLYP-D3/maug-cc-pVTZ		B3LYP-D3/N07D		B2PLYP-D3/maug-cc-pVTZ		B3LYP-D3/N07D		HYBRID B2PLYP-D3/B3LYP-D3
Assignment[a]	freq	I relative	harm	I(harm)	harm	I(harm)	anh	I(anh)	anh	I(anh)	anh
ν(N1H)	**3457**	**151**	3631	94	3634	97	**3461**	**80**	**3461**	**80**	**3458**
ν(N3H)	**3415**	**79**	3588	64	3599	70	**3419**	**56**	**3428**	**59**	**3417**
ν(C5H)			3266	2	3261	3	3136	1	3125	1	3129
ν(C6H)			3226	3	3222	3	3104	2	3089	3	3093
ν(C4=O)	**1738**	**543**	1761	632	1783	668	**1732**	**335**	**1752**	**325**	**1730**
ν(C5=C6)	1634	29	1665	38	1668	63	1627	23	1633	38	1631
δ(N1H), δ(N3H), ν(C6N1)	1534	623	1571	699	1573	678	1527	453	1527	522	1526
δ(N3H) + δ(CH)	1430	4	1457	23	1456	28	1419	8	1419	9	1421
δ(C6H)	1389	14	1402	49	1397	72	1366	50	1362	77	1366
δ(NH)	1376	38	1396	33	1395	10	1360	24	1356	3	1357
δ(C5H), ν(C6N1), δ(C6H), ν(N3C4)	1223	33	1243	18	1237	6	1217	4	1212	10	1219
ν(C2N3), δ(C6H), δ(N1H), ν(N3C4)	1191	150	1216	230	1212	255	1185	199	1179	189	1184
ν(C=S), δ(ring), δ(N3H)	1148	192	1169	154	1160	133	1140	86	1132	77	1142
ν(C6N1), δ(C5H), ν(C5=C6)	1060	4	1085	4	1082	8	1062	4	1059	7	1062
δ(ring)	986	15	1003	13	1001	11	979	16	980	12	982
γ(CH)			981	0	967	0	966	0	947	0	962
ν(N3C4), ν(C4C5), ν(C=S)	907	4	916	6	911	5	899	6	894	5	900
γ(C5H), γ(C4O), γ(C6H)	806	58	825	63	816	69	816	59	804	67	813
γ(N3H), γ(C4O), γ(C5H)	727	4	745	8	740	17	735	4	727	10	731
δ(ring), ν(C4C5), ν(C=S), ν(N1C2)	710	5	720	4	717	4	708	3	706	3	709
τ(ring), γ(N3H)	694	67	716	64	709	64	699	65	693	53	700
γ(C=S), γ(N1H)	643	16	649	9	642	28	645	11	637	23	643
γ(N1H)	604	31	618	32	620	26	610	27	608	29	606
δ(ring)	530	13	536	9	538	9	529	6	531	7	529
δ(C4=O)	491	12	497	11	499	12	492	10	494	11	492
δ(ring), ν(C=S)	451	15	453	13	449	14	448	13	444	14	448
τ(ring)	395	11	404	14	402	13	402	7	398	10	400
δ(C=S)	269	10	272	9	271	9	268	7	269	5	271
τ(ring)			162	1	165	1	160	1	162	1	159
τ(ring)			128	1	131	1	129	1	130	1	127
MAE			29.2		28.2		6.4		7.5		5.5
MIN			2		-2		-23		-27		-23
MAX			174		184		10		14		7

[a] Abbreviations: ν = stretching; δ = in-plane bending; γ = out-of-plane bending; τ = torsional. [b] Ref.[126]

teractions like the dimers of acetamide and formamide, focusing in particular on the ν(C=O) and ν(N-H).
In Table 3.2 and Table 3.3 all the harmonic and anharmonic stretching frequencies of the N-H and C=O functional groups, respectively, computed with the B3LYP-D3/N07D, B2PLYP-D3/maug-cc-pVTZ and hybrid B2PLYP-D3/B3LYP-D3 methods, have been compared with experiments for the molecular systems 2-thiouracil, acetamide, formamide, and the dimers of acetamide and formamide. From these results it is clearly evident that B2PLYP-D3/maug-cc-pVTZ overall is able to predict much better the experimental frequencies for ν(C=O) and ν(N-H), even for those involved in hydrogen bonds, giving MAEs for the anharmonic vibrational frequencies of 3.2 cm^{-1} and 7.2 cm^{-1}, respectively, instead of 13.0 cm^{-1} and 17.5 cm^{-1} obtained with B3LYP-D3/N07D. Accordingly, the absolute maximum errors obtained with B2PLYP-D3/maug-cc-pVTZ for ν(C=O) and ν(N-H) are 6 cm^{-1} and 7 cm^{-1}, respectively, instead of 23 cm^{-1} and 59 cm^{-1} obtained with B3LYP-D3/N07D. Especially for ν(C=O), such improvements are due to the harmonic part, while the anharmonic corrections evaluated by means of B3LYP-D3/N07D have almost the same accuracy of the B2PLYP-D3/maug-cc-pVTZ ones, as can be inferred by the MAEs of 5.0 cm^{-1} for ν(C=O) and 6.9 cm^{-1} for ν(N-H) obtained using the hybrid B2PLYP-D3/B3LYP-D3 approach. Such results confirm the reliability of the B2PLYP-D3/maug-cc-pVTZ as reference method to calculate vibrational frequencies of systems like the nucleobases complexes for which experimental data are not available for comparison.

Table 3.2: Harmonic and anharmonic stretching vibrational frequencies of the N-H functional groups for the molecular systems 2-thiouracil, acetamide, formamide, and the dimers of acetamide and formamide, computed with the B3LYP-D3/N07D, B2PLYP-D3/maug-cc-pVTZ and hybrid B2PLYP-D3/B3LYP-D3 methods, in comparison with experiments.[126,213,398–400]

Mode	Exp[b]	B3LYP-D3/N07D		B2PLYP-D3/maug-cc-pVTZ		HYBRID B2PLYP-D3/B3LYP-D3
Assignment[a]	freq	harm	anh	harm	anh	anh
2-thiouracil						
ν(N1H)	3457	3634	3461	3631	3461	3458
ν(N3H)	3415	3599	3428	3588	3419	3417
formamide						
ν_{asym}(NH)	3547	3729	3536	3738	3544	3545
ν_{sym}(NH)	3427	3587	3414	3597	3425	3425
formamide dimer						
$\nu_{asym-asym}$(NH)	3515	3678	3503	3690	3518	3515
$\nu_{asym-sym}$(NH)	3131	3314	3097	3334	3122	3117
acetamide						
ν_{asym}(NH)	3557	3739	3544	3747	3550	3552
ν_{sym}(NH)	3436	3598	3426	3607	3435	3435
acetamide dimer						
$\nu_{asym-asym}$(NH)	3515	3687	3509	3698	3522	3519
$\nu_{asym-sym}$(NH)	3140	3301	3081	3322	3109	3102
MAE		172.4	17.5	181.2	7.2	6.9
MIN		160	-59	170	-31	-38
MAX		184	13	203	7	4

[a]Abbreviations: ν = stretching; sym = symmetric; asym = asymmetric. [b]Ref.[126,213,398–400]

Table 3.3: Harmonic and anharmonic stretching vibrational frequencies of the C=O functional groups for the molecular systems 2-thiouracil, acetamide, formamide, and the dimers of acetamide and formamide, computed with the B3LYP-D3/N07D, B2PLYP-D3/maug-cc-pVTZ and hybrid B2PLYP-D3/B3LYP-D3 methods, in comparison with experiments.[126,213,398–400]

Mode	Exp[b]	B3LYP-D3/N07D		B2PLYP-D3/maug-cc-pVTZ		HYBRID B2PLYP-D3/B3LYP-D3
Assignment[a]	freq	harm	anh	harm	anh	anh
2-thiouracil						
ν(C4=O)	1738	1783	1752	1761	1732	1730
formamide						
ν(C=O)	1739	1795	1762	1778	1744	1744
formamide dimer						
ν(C=O)	1728	1779	1742	1768	1729	1731
acetamide						
ν(C=O)	1728	1769	1739	1755	1727	1724
acetamide dimer						
ν(C=O)	1694	1746	1698	1737	1697	1689
MAE		49.2	13.0	34.4	3.2	5.0
MIN		41	4	23	-6	-8
MAX		56	23	43	5	5

[a]Abbreviations: ν = stretching. [b]Ref.[126,213,398–400]

3.2 Hybrid and ONIOM models

3.2.1 Validation of the Computational Method: the Uracil-Water Complexes

The availability of the carefully assigned experimental frequencies[401] for the set of four uracil-water complexes, U-Wn (n=1,2,3,4) shown in Figure 1.2, has allowed to assess the performance of the different computational methods in predicting anharmonic vibrational frequencies of molecular systems where strong hydrogen bonds between proton donor and acceptor moieties take place.
Firstly, the anharmonic vibrational frequencies of the most stable uracil-water complex (U-W1) have been computed with the B3LYP, B3LYP-D3, B3LYP-DCP and M06-2X methods, in conjunction with the less expensive N07D basis set,[244,245] with the exception of B3LYP-DCP for which the companion 6-31+G(2d,2p) basis set has been recommended.[246–250] The harmonic vibrational frequencies have been computed also with the B2PLYP/maug-cc-pVTZ method and adopting a two-layer ONIOM[402,403] B2PLYP:B3LYP scheme in which the model system (high level, B2PLYP/maug-cc-pVTZ) corresponds to the part of molecular system directly involved in the hydrogen-bonding interaction, while the whole complex is taken as the real system (lower level, B3LYP/N07D). Partitioning within the uracil monomer, leading to the definition of model system, has been always performed along a formally single bond. In total four different partitioning schemes (1-2, 2-3, 3-4 and 4-6) have been defined (see Figure 3.2). This allowed to compute also anharmonic frequencies using the hybrid schemes B2PLYP/B3LYP-D3, B2PLYP/B3LYP-DCP and B2PLYP:B3LYP/B3LYP-D3.
Table 3.4 reports harmonic and anharmonic OH and NH stretching vibrational frequencies for the most stable uracil-water complex (U-W1), which are mainly influenced by the hydrogen-bonding interactions, in comparison with their experimental counterparts. It turns out that the best theoretical description of the effect of hydrogen-bonding on the vibrational frequencies is obtained with the hybrid B2PLYP/B3LYP-D3 and B2PLYP:B3LYP/B3LYP-D3 approaches, providing MAEs with respect to experiment of about 10 cm^{-1}, and maximum discrepancies well below 20 cm^{-1}, further confirming the accuracy of the anharmonic shifts computed with the B3LYP-D3/N07D method. Application of B3LYP-DCP/6-31+G(2d,2p) anharmonic corrections within B2PLYP/B3LYP-DCP model leads to slightly worse results, with MAE of 16 cm^{-1} and maximum discrepancy exceeding 30 cm^{-1}. Among the less expensive methods used in this work, M06-2X/N07D yields the worst results, with a MAE of 67 cm^{-1}, while the B3LYP-based approaches provides better predictions, with MAEs of 31, 25 and 44 cm^{-1} in the case of B3LYP/N07D, B3LYP-D3/N07D and B3LYP-DCP/6-31+G(2d,2p) method, respectively. In particular, B3LYP-D3/N07D outperforms B3LYP-DCP/6-31+G(2d,2p) in calculating the harmonic part of the overall anharmonic frequencies, whereas both methods are roughly equally reliable in the prediction of the anharmonic shifts, as already shown

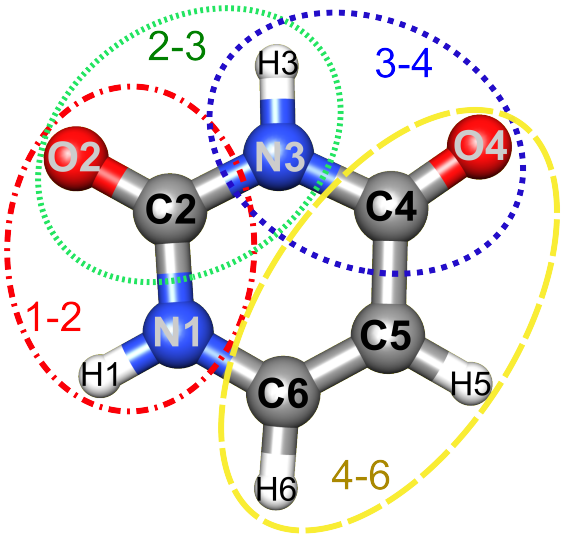

Figure 3.2: Partitioning schemes of the uracil monomer within ONIOM B2PLYP:B3LYP computations, parts considered as high level/model systems (computed at the B2PLYP/maug-cc-pVTZ level) are marked by circles (1-2, 2-3, 3-4 or 4-6).

in Chapter 2.[170] Thus, we have focused only on B3LYP-D3/N07D and B3LYP-DCP/6-31+G(2d,2p) methods for computing anharmonic vibrational frequencies of the other less stable uracil-water complexes. Such results have been reported in Table 3.5, considering as references for anharmonic computations the hybrid B2PLYP/B3LYP-D3, B2PLYP/B3LYP-DCP and B2PLYP:B3LYP/B3LYP-D3 approaches, given their high accuracy compared to the experiment. Similarly to the anharmonic computations for the most stable uracil-water complex, B3LYP-D3/N07D provides more accurate results with respect to B3LYP-DCP/6-31+G(2d,2p), which can be essentially attributed to improvements in the harmonic part. Indeed, the quite similar MAEs obtained using the hybrid B2PLYP/B3LYP-D3 and B2PLYP/B3LYP-DCP methods point out the comparable performance of B3LYP-D3/N07D and B3LYP-DCP/6-31+G(2d,2p) in predicting the anharmonic corrections. Moreover, it is apparent that the hybrid B2PLYP:B3LYP/B3LYP-D3 model, in which the most expensive B2PLYP computations are restricted to the part of molecular system directly involved in hydrogen-bonding interaction, yields very accurate results, with MAEs of 8-12 cm^{-1} and low maximum absolute deviations with respect to experiments within a range of about 13-21 cm^{-1}. This is due not only to the reliability of the anharmonic shifts computed with the B3LYP-D3/N07D method, but also to the increased accuracy of the harmonic part computed through the ONIOM B2PLYP:B3LYP scheme, which is comparable to the accuracy of the full B2PLYP computations, at strongly reduced computational cost.

In summary, results of the anharmonic vibrational computations for the uracil-

water complexes indicate that dispersion-corrected B3LYP-based methods are reliable enough for a quantitative comparison with experiments, reproducing well also vibrations involved in hydrogen-bonded bridges that are characterized by more anaharmonic PES with respect to those of the isolated molecules.[123,125,326] In particular, B3LYP-D3/N07D yields more accurate results with average uncertainties within 22 cm^{-1} in the spectral region of hydrogen-bonding interactions. Although B3LYP-DCP/6-31+G(2d,2p) gives higher errors with an average MAE of 38 cm^{-1}, it outperforms several other dispersion-corrected DFT approaches and provides reliable anharmonic corrections. However, the need of a reliable yet not expensive computational method for predicting both the harmonic part and the anharmonic correction to vibrational frequencies suggests B3LYP-D3/N07D as the method of choice.

Furthermore, a more accurate description of the overall anharmonic frequencies can be obtained by hybrid computations, where the harmonic part of the overall vibrational frequencies is computed at a higher level of theory like B2PLYP/maug-cc-pVTZ, or using the less demanding ONIOM B2PLYP:B3LYP scheme, where only the part of the molecular system forming the hydrogen bonds is treated at B2PLYP/maug-cc-pVTZ level of theory while the remaining part is treated with B3LYP/N07D, obtaining average uncertainties of about 10 cm^{-1} in the spectral region of hydrogen-bonding interactions.

Table 3.4: Harmonic and anharmonic vibrational frequencies in the spectral region of NH and OH stretching vibrational modes for the most stable uracil-water complex (U-W1) compared with experimental data[a]. Mean absolute errors (MAEs) and maximum absolute deviations ($|MAX|$) with respect to the experimental data are also reported.

		Harmonic					
mode[b]	**Exp.**[a]	B3LYP	B3LYP-D3	B3LYP-DCP	M06-2X	B2PLYP	B2PLYP:B3LYP
OH	3727	3885	3886	3882	3958	3899	3901
OH (HB)	3468	3600	3604	3591	3710	3626	3629
N3H	3443	3605	3607	3585	3631	3599	3606
N1H (HB)	3317	3428	3423	3405	3480	3451	3454
MAE		141	141	127	206	155	159
$\|MAX\|$		162	164	155	243	172	174

		Anharmonic						
mode[b]	**Exp.**[a]	B3LYP	B3LYP-D3	B3LYP-DCP	M06-2X	B2PLYP/B3LYP-D3	B2PLYP/B3LYP-DCP	B2PLYP:B3LYP
OH	3727	3703	3707	3702	3801	3720	3719	3722
OH (HB)	3468	3427	3430	3417	3558	3452	3452	3455
N3H	3443	3439	3440	3419	3497	3432	3433	3439
N1H (HB)	3317	3260	3277	3239	3365	3305	3285	3308
MAE		31	25	44	67	11	16	8
$\|MAX\|$		58	40	78	91	16	32	13

[a] Experimental data from IR spectra recorded in Helium nanodroplets.[401]
[b] Modes involved in hydrogen-bonding interactions are labeled as (HB).

Table 3.5: Harmonic and anharmonic vibrational frequencies in the spectral region of NH and OH stretching vibrational modes for uracil-water complexes U-W2, U-W3 and U-W4, compared with experimental data[a]. Mean absolute errors (MAEs) and maximum absolute deviations ($|MAX|$) with respect to the experimental data are also reported.

		Harmonic			
mode[b]	**Exp.**[a]	B3LYP-D3	B3LYP-DCP	B2PLYP	B2PLYP:B3LYP
U-W2					
OH	3728	3889	3885	3903	3904
OH (HB)	3501	3647	3635	3667	3671
N1H	3492	3648	3631	3648	3649
N3H (HB)	3256	3420	3393	3430	3439
MAE		157	142	168	172
$\|MAX\|$		164	158	175	183
U-W3					
OH	3722	3887	3883	3900	3902
OH (HB)	3468	3600	3628	3626	3627
N1H	3492	3645	3586	3646	3648
N3H (HB)	3271	3402	3376	3415	3424
MAE		145	130	159	162
$\|MAX\|$		164	161	178	180
U-W4					
OH	3723	3892	3890	3903	3903
OH (HB)	3508	3620	3607	3647	3648
N1H	3492	3644	3628	3644	3645
N3H	3443	3605	3585	3598	3605
MAE		149	136	156	159
$\|MAX\|$		169	167	180	180

		Anharmonic				
mode[b]	**Exp.**[a]	B3LYP-D3	B3LYP-DCP	B2PLYP/B3LYP-D3	B2PLYP/B3LYP-DCP	B2PLYP:B3LYP
U-W2						
OH	3728	3707	3703	3721	3721	3722
OH (HB)	3501	3479	3467	3499	3499	3503
N1H	3492	3480	3464	3480	3481	3481
N3H (HB)	3256	3252	3222	3262	3259	3271
MAE		15	30	7	6	9
$\|MAX\|$		22	34	12	11	15
U-W3						
OH	3722	3705	3701	3718	3718	3720
OH (HB)	3468	3422	3460	3448	3458	3449
N1H	3492	3478	3412	3479	3472	3481
N3H (HB)	3271	3233	3206	3246	3245	3255
MAE		29	43	16	15	12
$\|MAX\|$		46	80	25	26	19
U-W4						
OH	3723	3710	3707	3721	3720	3721
OH (HB)	3508	3459	3444	3486	3484	3487
N1H	3492	3481	3463	3481	3479	3482
N3H	3443	3438	3416	3431	3429	3438
MAE		19	34	12	14	10
$\|MAX\|$		49	64	22	24	21

[a] Experimental data from IR spectra recorded in Helium nanodroplets.[401]
[b] Modes involved in hydrogen-bonding interactions are labeled as (HB).

3.2.2 Effects of Hydrogen-Bonding Interactions on IR Spectra: the Hydrogen-Bonded Uracil Dimers

Reliable computational methods are necessary to reproduce the remarkable effects of hydrogen-bonding interactions on the vibrational features with the required accuracy for a quantitative comparison with experiments.
For the uracil monomer and the hydrogen-bonded uracil dimers, it is feasible to carry out a full VPT2 anharmonic treatment at B3LYP-D3/N07D level and to use a ONIOM B2PLYP:B3LYP/B3LYP-D3 scheme, which allows to improve the accuracy of computed vibrational frequencies.
Firstly, we have analyzed the effects of hydrogen-bonding interactions on the IR spectra of the hydrogen-bonded uracil dimers U-Un (n=1,2,3,4,5,6) shown in Figure 1.2, using the B3LYP-D3/N07D method. Then, more accurate results have been obtained, with only a slight increase of the computational burden, computing the hybrid anharmonic frequencies using the ONIOM B2PLYP:B3LYP scheme for the harmonic part of the whole vibrational frequency and the B3LYP-D3/N07D method for the anharmonic correction.

In the absence of hydrogen bonds, i.e. in the case of the isolated uracil molecule, the accuracy of fully anharmonic VPT2 calculations using the B3LYP-D3/N07D method is particularly high, with MAE of about 10 cm^{-1}, like in the case of B3LYP-D3/SNSD as observed in Chapter 2. Such an accuracy is on pair with the much more computationally demanding approaches with the harmonic frequencies evaluated by a composite scheme at the Coupled Cluster (CC) level (CCSD(T)/CBS(T,Q+aug)+CV).[119] Figure 3.3 shows the fully anharmonic IR spectrum of the uracil monomer computed at the B3LYP-D3/N07D level of theory, compared with the experimental spectrum of uracil isolated in Argon matrix.[364] A very good agreement between theory and experiment has been achieved not only for the fundamental transitions but for the whole IR spectra. In fact, while anharmonic frequencies improve the accuracy of their harmonic counterparts, anharmonic intensities (including both mechanical and electric anharmonic terms) are often needed to obtain even qualitatively correct spectra shapes since overtones and combination bands have vanishing intensities at the harmonic level. The experimental features in the 1600-1800 cm^{-1} spectral range, shown in the inset, are well reproduced by anharmonic calculations, in particular taking into account also the presence of relatively intense non-fundamental transitions. Anharmonic computations show, in addition to the two most intense bands corresponding to the C=O stretching fundamentals and one weaker transition related to the ν(C5=C6) mode, also five non-fundamental transitions. In line with experimental findings two of them are nearly as intense as the ν(C4=O) fundamental. These results further confirm that GVPT2/DVPT2 anharmonic computations employing DFT potential energy surfaces provide qualitatively correct IR intensities of all transitions, largely sufficient to simulate realistic band patterns.

In order to investigate the effect of the intermolecular interactions on the vibrational frequencies of nucleobases, the fully anharmonic infrared spectra of six dif-

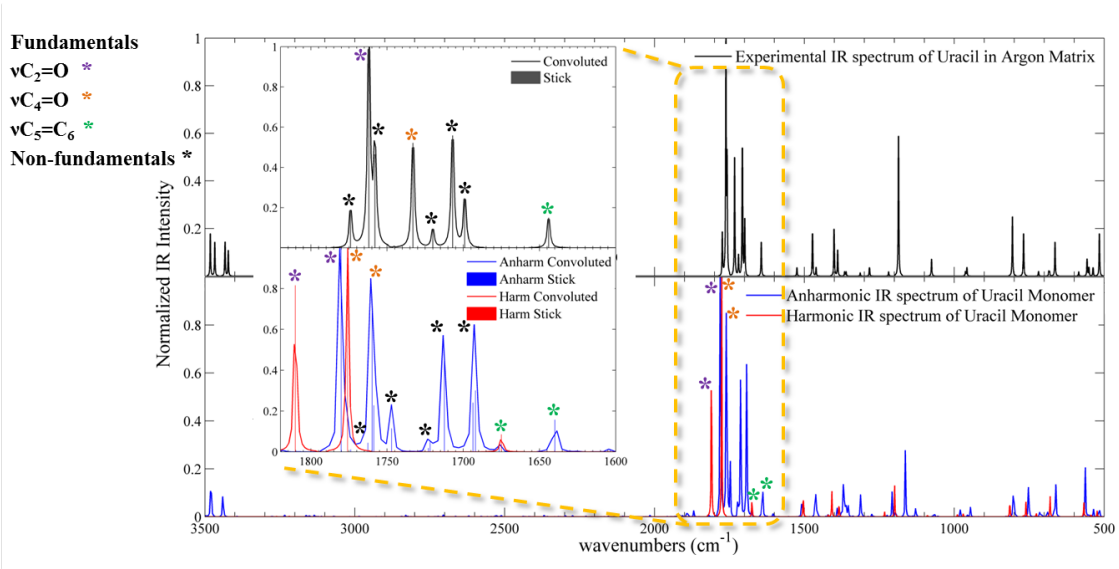

Figure 3.3: The harmonic and anharmonic infrared spectrum of isolated uracil molecule computed with the B3LYP-D3/N07D method, compared with the IR experimental spectrum of uracil isolated in Argon matrix,[364] in the 500-3500 cm^{-1} spectral range, along with the assignment of some IR bands. Experimental IR spectrum recorded in the low-temperature Ar matrix has been generated using the data reported in ref.[364] IR spectra line-shapes (both theoretical and experimental) have been convoluted using Lorentzian functions with a half-width at a half-maximum (HWHM) of 1 cm^{-1}. The inset shows the 1600-1800 cm^{-1} spectral range, highlighting both stick and convoluted spectra.

ferent hydrogen-bonded uracil dimers have been computed at the B3LYP-D3/N07D level of theory and compared to the IR spectrum of the isolated uracil monomer. Figure 3.4 shows the anharmonic IR spectra of all uracil dimers, along with the assignment of the NH and CO stretching vibrational modes, pointing out the different shifts of the corresponding vibrational frequencies with respect to uracil monomer due to the formation of the hydrogen-bonding interactions. An important observation is that, depending on the specific hydrogen bonds, the infrared spectral features are significantly different, and such reliable theoretical simulations can help to distinguish among the different dimers in experimental mixtures.[404] As expected, the most important shifts of the vibrational frequencies and/or IR intensity variations concern the functional groups that are directly involved in the hydrogen-bonding interactions and the magnitude of such vibrational shifts can be considered as an indication of the strength of the intermolecular interactions taking place in the complexes, which in turn reflects their relative stability. In particular, the NH modes are extremely sensitive to H-bonding, featuring significant red-shifts for the stretching vibrations and analogous blue-shifts for the out-of-plane bendings, as already observed in other systems,[329] and used to establish empirical correlations between these spectral parameters and the H-bond properties.[330]

For a more detailed discussion about the harmonic and anharmonic vibrational frequencies and IR intensities of these dimers, the interested reader is referred to the original work.[173] Here we mainly focus on the region of hydrogen-bonding interactions (N-H and C=O stretchings and N-H bendings), reporting in Table 3.6 the relative harmonic and anharmonic vibrational frequencies computed with the B3LYP-D3/N07D method, along with the more accurate results obtained through the hybrid B2PLYP:B3LYP/B3LYP-D3 scheme. The mean absolute deviation of the anharmonic B3LYP-D3/N07D frequencies for all the dimers with respect to the hybrid anharmonic frequencies computed at the B2PLYP:B3LYP/B3LYP-D3 level of theory is of about 28 cm^{-1} in the specific zone of the hydrogen-bonded functional groups, mainly due to the variation of N-H stretching and bending vibrations, by 49 cm^{-1} and 22 cm^{-1}, on average, respectively. At variance, for ν(NH), ν(C=O) and γ(NH) modes not involved in hydrogen-bonding the B3LYP-D3/N07D and B2PLYP:B3LYP/B3LYP-D3 frequencies agree within 5 cm^{-1}. Thus the ONIOM B2PLYP:B3LYP scheme can be suggested as inexpensive method to correct the harmonic part of the vibrational frequencies of functional groups involved in strong hydrogen bonds, using the B3LYP-D3/N07D method for computing the anharmonic corrections to these modes together with all other vibrational frequencies and IR intensities.

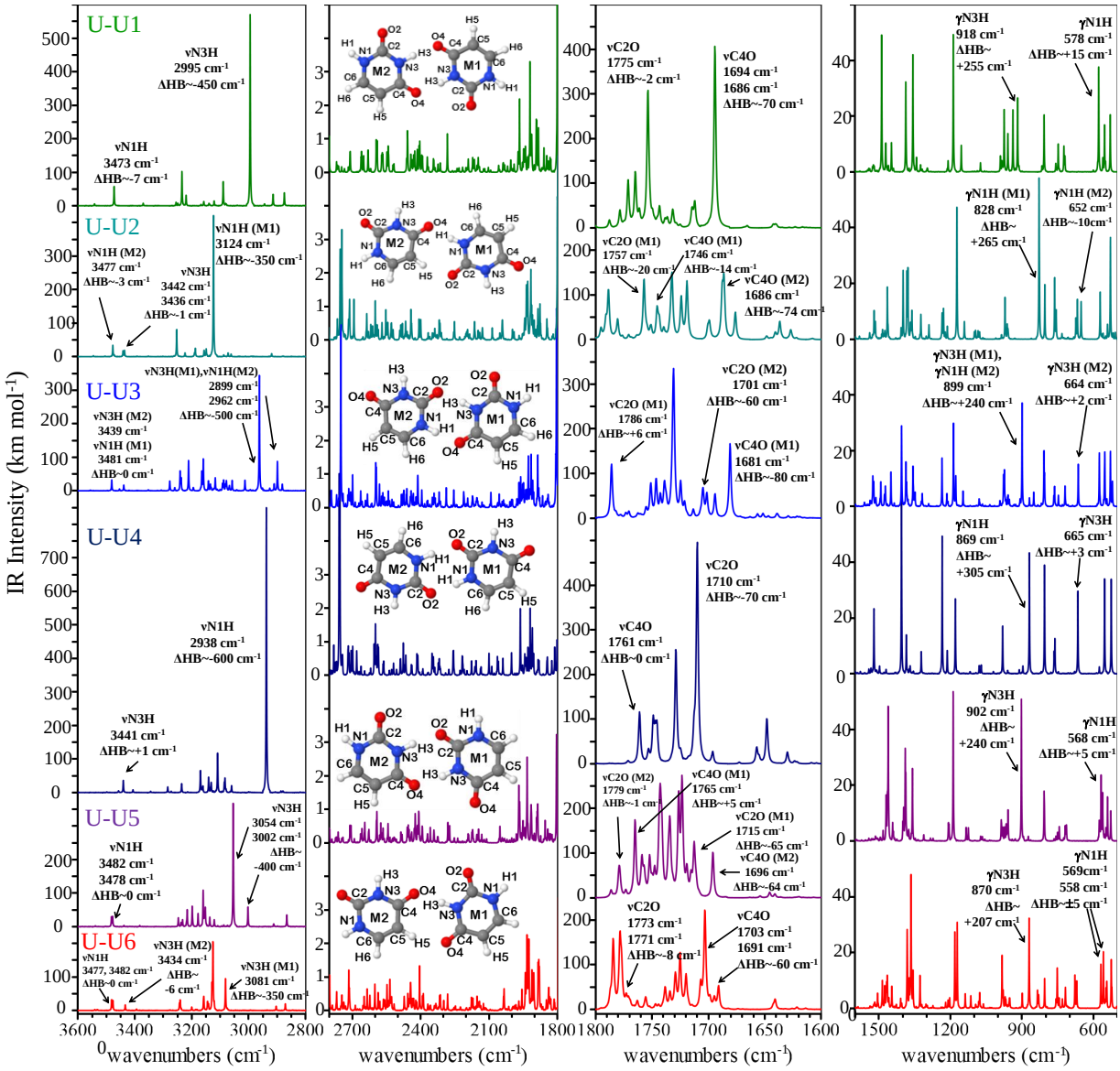

Figure 3.4: Anharmonic infrared spectra of the six different hydrogen-bonded uracil dimers computed with the B3LYP-D3/N07D method, along with the assignment of the NH and CO stretchings and NH bendings, and the corresponding shifts of the vibrational frequencies with respect to uracil monomer due to the formation of the hydrogen-bonding interactions (ΔHB).

Table 3.6: Harmonic and anharmonic vibrational frequencies for the hydrogen-bonded uracil dimers, selected in the spectral region of hydrogen-bonding interactions, computed with the B3LYP-D3/N07D method and hybrid B2PLYP:B3LYP/B3LYP-D3 scheme.

	B3LYP-D3			**B2PLYP:B3LYP**	
mode[a]	harm	anh	Δanh	harm	anh
U-U1					
ν(N1H) (M1,M2)	3645	3473	-172	3648	3476
ν(N1H) (M1,M2)	3645	3473	-172	3648	3476
ν(N3H)-HB (M1,M2)	3248	2995	-253	3303	3050
ν(N3H)-HB (M1,M2)	3202	2956	-246	3265	3019
ν(C2O) (M1,M2)	1814	1782	-33	1809	1776
ν(C2O) (M1,M2)	1812	1775	-37	1807	1770
ν(C4O)-HB (M1,M2)	1743	1694	-49	1729	1681
ν(C4O)-HB (M1,M2)	1734	1686	-47	1724	1676
γ(N3H)-HB (M1,M2)	942	918	-24	909	885
γ(N3H)-HB (M1,M2)	918	892	-25	885	860
γ(N1H) (M1,M2)	582	578	-4	575	571
γ(N1H) (M1,M2)	582	578	-4	575	571
U-U2					
ν(N1H) (M2)	3641	3477	-164	3643	3479
ν(N1H) (M1)	3610	3441	-168	3610	3441
ν(N3H) (M1)	3604	3436	-168	3604	3436
ν(N3H)-HB (M2)	3292	3123	-169	3326	3157
ν(C2O) (M2)	1817	1780	-36	1814	1778
ν(C2O) (M1)	1787	1757	-30	1760	1730
ν(C4O) (M1)	1767	1746	-21	1776	1754
ν(C4O)-HB (M2)	1727	1686	-41	1709	1668
γ(N1H)-HB (M1)	871	828	-43	863	819
γ(N1H)-HB (M1), γ(CH) (M2)	849	826	-23	847	824
γ(N3H) (M1)	678	652	-27	673	646
γ(N1H) (M2)	590	572	-19	587	568
U-U3					
ν(N1H) (M1)	3642	3481	-161	3644	3484
ν(N3H) (M2)	3606	3439	-167	3606	3439
ν(N3H)-HB (M1), ν(N1H)-HB (M2)	3237	2962	-275	3283	3009
ν(N3H)-HB (M1), ν(N1H)-HB (M2)	3186	2899	-287	3240	2953
ν(C2O) (M1)	1820	1786	-34	1814	1780
ν(C2O) (M2)	1782	1755	-27	1777	1750
ν(C4O)-HB (M2)	1764	1701	-62	1754	1691
ν(C4O)-HB (M1)	1728	1681	-48	1717	1669
ν(N3H)-HB (M1), ν(N1H)-HB (M2)	932	899	-33	905	872
ν(N1H)-HB (M2), ν(N3H)-HB (M1)	890	850	-40	872	832
γ(N3H) (M2)	682	664	-19	677	658
γ(N1H) (M1)	591	575	-16	584	569
U-U4					
ν(N3H) (M1,M2)	3606	3441	-165	3606	3441
ν(N3H) (M1,M2)	3606	3441	-165	3606	3440
ν(N1H)-HB (M1,M2)	3235	2938	-297	3272	2975
ν(N1H)-HB (M1,M2)	3181	2871	-310	3223	2913
ν(C4O) (M1,M2)	1788	1761	-27	1783	1756
ν(C4O) (M1,M2)	1783	1760	-23	1781	1758
ν(C2O)-HB (M1,M2)	1759	1710	-49	1747	1698
ν(C2O)-HB (M1,M2)	1755	1711	-44	1743	1699
γ(N1H)-HB (M1,M2)	907	869	-38	893	854
γ(N1H)-HB (M1,M2)	873	832	-40	861	821
γ(N3H) (M1,M2)	682	665	-17	677	660
γ(N3H) (M1,M2)	682	664	-17	677	660
U-U5					
ν(N1H) (M1)	3647	3482	-165	3648	3483
ν(N1H) (M2)	3645	3478	-168	3648	3481
ν(N3H)-HB (M1,M2)	3288	3054	-233	3337	3104
ν(N3H)-HB (M1,M2)	3242	3002	-240	3302	3062
ν(C2O) (M2)	1813	1779	-34	1807	1773
ν(C2O) (M1)	1782	1765	-17	1776	1759
ν(C4O)-HB (M1)	1768	1715	-52	1755	1702
ν(C4O)-HB (M2)	1740	1696	-44	1728	1685
γ(N3H)-HB (M1,M2)	927	902	-25	894	870
γ(N3H)-HB (M1,M2)	901	879	-22	869	848
γ(N1H) (M2)	581	568	-9	574	566
γ(N1H) (M1)	571	561	-10	566	556
U-U6					
ν(N1H) (M1)	3647	3482	-166	3650	3484
ν(N1H) (M2)	3645	3477	-168	3647	3479
ν(N3H) (M2)	3604	3434	-170	3604	3433
ν(N3H)-HB (M1)	3289	3081	-208	3332	3125
ν(C2O) (M1,M2)	1812	1773	-39	1809	1770
ν(C2O) (M1,M2)	1810	1771	-39	1805	1766
ν(C4O)-HB (M1,M2)	1755	1703	-51	1737	1685
ν(C4O) (M1,M2)	1738	1691	-47	1721	1674
γ(N3H)-HB (M1)	910	870	-40	877	837
γ(N3H) (M2)	687	670	-17	683	666
γ(N1H) (M2)	580	569	-11	575	564
γ(N1H) (M1)	575	558	-17	569	552

[a] Modes involved in hydrogen-bonding interactions are labeled as (HB). Modes localized on one of the uracil monomers are marked as (M1) and (M2), while modes delocalized over the dimer are labeled as (M1,M2), respectively. Abbreviations: ν = stretching; γ = out-of-plane bending.

3.3 Cost-Effective Correction for ν(C=O) and ν(N-H) within Hydrogen-Bonded Bridges

The description of vibrational properties for functional groups like C=O and N-H involved in hydrogen bond networks is a big challenge for computational spectroscopy, due to the strong anharmonicity of the corresponding stretching vibrational modes which implies both limitations of the perturbative treatment and the requirement of a very accurate description of the underlying PES. Even though this problem may be fixed by the hybrid computations with harmonic frequencies corrected at higher-level of theory, like Coupled Cluster or B2PLYP,[134] for studying very big and complex systems the use of less demanding yet accurate methods is necessary.
As shown in this Chapter, B3LYP-D3/N07D performs very well in predicting vibrational properties of medium- to large-size systems at relatively low computational cost, reproducing with high accuracy also the vibrations involved in hydrogen bonds.[173] Therefore, the vibrational properties computed with this method constitute a good starting point for applying further refinements.
Here we have developed a more general, inexpensive method to correct specifically the stretching vibrational modes of C=O and N-H functional groups involved in hydrogen bonds beyond the B3LYP-D3/N07D, which can be applied to a broad range of hydrogen-bonded molecular systems. In particular, we have firstly realized a database of very accurate harmonic frequencies computed at B2PLYP-D3/maug-cc-pVTZ level of theory, which can be used as benchmark for future validation of other computational methods and also for predicting anharmonic frequencies within hybrid schemes, for a training set of six different hydrogen-bonded uracil homodimers (U-Un, n=1,2,3,4,5,6, in Figure 1.2) and a test set of other five nucleobases homo- and heterodimers shown in Figure 1.2, namely one of the most stable hydrogen-bonded adenine homodimer (A-A), the nucleic acid base pairs adenine-uracil (A-U), adenine-thymine (A-T), guanine-cytosine (G-C) and the adenine-4-thiouracil heterodimer (A-ThioU). Afterward, the results relative to the ν(C=O) and ν(N-H) of all hydrogen-bonded uracil dimers constituting the training set have been used to define correction parameters for the B3LYP-D3/N07D frequencies of ν(C=O) and ν(N-H) involved in hydrogen bonds.

3.3.1 Determination and Validation of Correction Parameters for B3LYP-D3/N07D frequencies of ν(C=O) and ν(N-H) involved in Hydrogen Bonds: the Nucleobase Dimers

Correction parameters for ν(C=O) and ν(N-H) beyond B3LYP-D3/N07D have been determined using the training set of six different hydrogen-bonded uracil dimers, named U-Un (n=1,2,3,4,5,6) in Figure 1.2, to realize linear regressions as follows:

$$\omega_{\text{B2PLYP-D3}} = k * \omega_{\text{B3LYP-D3}} + \Delta$$

In this way, it has been possible to define the correction parameters, k and Δ, for the B3LYP-D3/N07D harmonic frequencies of ν(C=O) and ν(N-H), based on the more accurate B2PLYP-D3/maug-cc-pVTZ ones.
The linear fits of the B2PLYP-D3/maug-cc-pVTZ harmonic frequencies *vs.* the B3LYP-D3/N07D ones have been shown in Figure 3.5. The estimates of k and Δ turn out to be, respectively, 0.91 ± 0.04 and $(1.2 \pm 0.6) * 10^2$ for ν(C=O), and 0.935 ± 0.007 and $(5 \pm 2) * 10^1$ for ν(N-H). These parameters have been applied to correct the B3LYP-D3/N07D harmonic frequencies for ν(C=O) and ν(N-H) of the test set of other homo- and heterodimers of nucleobases, and the corrected B3LYP-D3/N07D frequencies have been compared to the B2PLYP-D3/maug-cc-pVTZ ones in Table 3.7. The results indicate that the correction parameters of B3LYP-D3/N07D harmonic frequencies defined for ν(C=O) and ν(N-H) using as training set the hydrogen-bonded uracil dimers, are able to improve both ν(C=O) and ν(N-H) of the other nucleobases dimers of the test set, even though they are characterized by different kinds of hydrogen bonds and quite different vibrational frequencies. Indeed, the correction parameters have been derived from hydrogen-bonded systems where only N-H$\cdots$C=O hydrogen bond type is present. The test set, instead, features much different hydrogen bond types, specifically $NH_2\cdots$N, N-H$\cdots$N and $NH_2\cdots$C=O. This demonstrates that the applicability of such parameters is pretty broad and may be extended to any hydrogen-bonded system involving C=O and N-H functional groups, independently of the partner in the hydrogen-bonded bridge. The only exception has been observed for the $\nu_{asym}(NH_2)$, which is well described with B3LYP-D3/N07D and does not need correction.
In more detail, in the case of A-A, which presents two $NH_2\cdots$N hydrogen bond type, for the $\nu_{asym}(NH_2)$ the discrepancies of the B3LYP-D3/N07D harmonic frequencies with respect to the B2PLYP-D3/maug-cc-pVTZ ones increase of about 6 cm^{-1} as a consequence of the correction, while a significant improvement of 22-25 cm^{-1} is observed for $\nu_{sym}(NH_2)$.
In the case of A-U, featuring one $NH_2\cdots$C=O and one N-H$\cdots$N hydrogen bonds, the correction causes a worsening of about 7 cm^{-1} for the harmonic frequency of the adenine $\nu_{asym}(NH_2)$, but the harmonic frequency of $\nu_{sym}(NH_2)$ is improved of about

13 cm^{-1} and, mostly, the uracil ν(N3H) undergoes a great improvement of about 46 cm^{-1}. Regarding the carbonyl stretching vibrational modes of uracil, both involved in the hydrogen bonds within the dimer, the correction allows to improve both the ν(C2O) and ν(C4O) with corresponding discrepancies, compared to B2PLYP-D3/maug-cc-pVTZ, of -2.7 cm^{-1} and 0.7 cm^{-1}, respectively, instead of 14.9 cm^{-1} and 12.6 cm^{-1} for non-corrected B3LYP-D3/N07D.
For A-ThioU and A-T, the situation is analogous to A-U, with a worsening of about 7 cm^{-1} for the harmonic frequency of the adenine $\nu_{asym}(NH_2)$, but the harmonic frequency of $\nu_{sym}(NH_2)$ is improved of about 12-13 cm^{-1} and, mostly, the thiouracil and thymine ν(N3H) undergo a great improvement of about 42-45 cm^{-1}, reaching a value very close to the corresponding B2PLYP-D3/maug-cc-pVTZ one with a discrepancy of only -2.7 cm^{-1} in the case of the thiouracil. The ν(C2O) of thiouracil is improved as well, with discrepancy of 0.2 cm^{-1} compared to B2PLYP-D3/maug-cc-pVTZ, instead of 18.0 cm^{-1} for non-corrected B3LYP-D3/N07D. Improvements in the description of ν(C2O) and ν(C4O) for thymine lead to corresponding discrepancies, compared to B2PLYP-D3/maug-cc-pVTZ, of -2.7 cm^{-1} and 1.3 cm^{-1}, respectively, instead of 14.7 cm^{-1} and 12.0 cm^{-1} for non-corrected B3LYP-D3/N07D.
In the case of G-C, characterized by two $NH_2 \cdots C{=}O$ and one N-H$\cdots$N hydrogen bonds, the worsening of the $\nu_{asym}(NH_2)$ is about 3 cm^{-1} for guanine and 6 cm^{-1} for cytosine, while the harmonic frequency of $\nu_{sym}(NH_2)$ for guanine is improved of about 14 cm^{-1}. Other noteworthy improvements have been observed for intermolecular modes within the hydrogen-bonded bridges, i.e. the simultaneous stretching of N1H, NH_2 of guanine and NH_2 of cytosine, whose discrepancy decreases of 27 cm^{-1}, the simultaneous stretching of NH_2 of cytosine and N1H of guanine, whose discrepancy decreases of 32 cm^{-1}, and the simultaneous stretching of C6O of guanine and C2O of cytosine, with discrepancy of -3.0 cm^{-1} compared to B2PLYP-D3/maug-cc-pVTZ, instead of 10.6 cm^{-1} for non-corrected B3LYP-D3/N07D.
The generalizability of such parameters is very promising for correcting B3LYP-D3/N07D frequencies of ν(C=O) and ν(N-H) involved in any kind of hydrogen bonds, and allows to apply them also to isolated molecules including such functional groups.

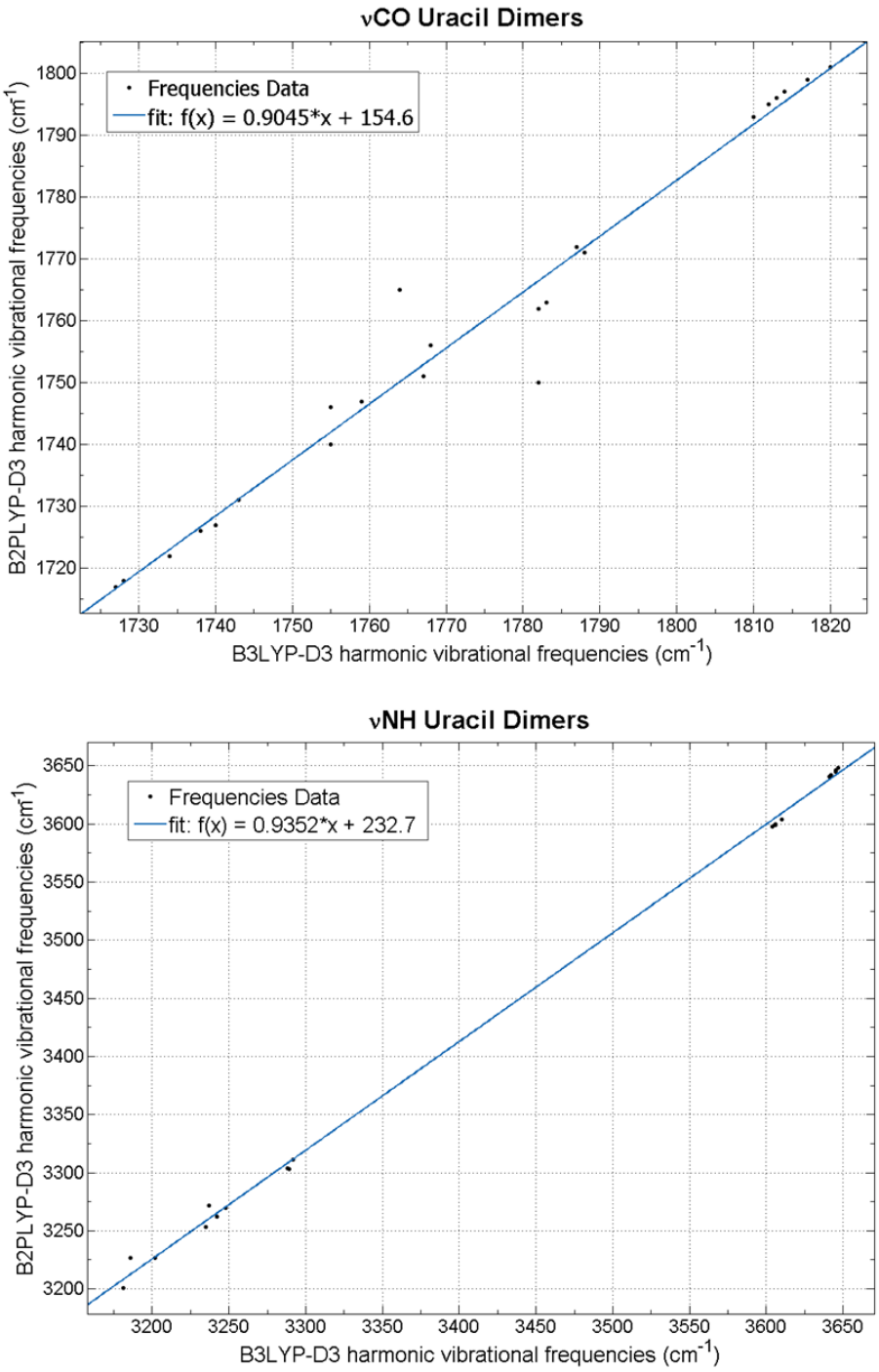

Figure 3.5: Linear fits of B2PLYP-D3/maug-cc-pVTZ harmonic frequencies *vs.* B3LYP-D3/N07D ones, for ν(C=O) (top) and ν(N-H) (bottom).

Table 3.7: Comparison between B2PLYP-D3/maug-cc-pVTZ harmonic frequencies and B3LYP-D3/N07D ones corrected using the linear regression parameters $k = 0.91 \pm 0.04$ and $\Delta = (1.2 \pm 0.6) * 10^2$ for ν(C=O), and $k = 0.935 \pm 0.007$ and $\Delta = (5 \pm 2) * 10^1$ for ν(N-H).

Assignment[a]	B2PLYP-D3/maug-cc-pVTZ	B3LYP-D3/N07D		corrected B3LYP-D3/N07D	
Mode	harm	harm	Δharm	harm	Δharm
A-A					
$\nu_{asym}(NH_2)$ (M1,M2)	3680	3679	-1	3673	-7
$\nu_{asym}(NH_2)$ (M1,M2)	3680	3679	-1	3673	-7
$\nu_{sym}(NH_2)$ (M1,M2)	3295	3254	-41	3276	-19
$\nu_{sym}(NH_2)$ (M1,M2)	3255	3210	-46	3235	-21
A-U					
$\nu_{asym}(NH_2)$ (A)	3694	3693	-2	3686	-8
$\nu_{sym}(NH_2)$ (A)	3408	3387	-21	3400	-8
ν(N3H) (U)	2940	2880	-61	2926	-15
ν(C2O) (U)	1788	1803	15	1785	-3
ν(C4O), ν(ring), δ(N3H) (U)	1731	1744	13	1732	1
A-ThioU					
$\nu_{asym}(NH_2)$ (A)	3696	3693	-3	3687	-9
$\nu_{sym}(NH_2)$ (A)	3436	3409	-27	3421	-15
ν(N3H) (ThioU)	2982	2937	-45	2979	-3
ν(C2O) (ThioU)	1787	1805	18	1787	0
A-T					
$\nu_{asym}(NH_2)$ (A)	3695	3693	-1	3687	-8
$\nu_{sym}(NH_2)$ (A)	3406	3385	-22	3398	-8
ν(N3H) (T)	2954	2895	-59	2940	-14
ν(C2O) (T)	1786	1800	15	1783	-3
ν(C4O), δ(N3H) (T)	1719	1731	12	1720	1
G-C					
$\nu_{asym}(NH_2)$ (G)	3704	3706	2	3699	-5
$\nu_{asym}(NH_2)$ (C)	3689	3686	-4	3679	-10
$\nu_{sym}(NH_2)$ (G)	3398	3378	-20	3392	-6
ν(N1H), $\nu(NH_2)$ (G), $\nu(NH_2)$ (C)	3206	3172	-34	3199	-7
$\nu(NH_2)$ (C), ν(N1H) (G)	3133	3091	-42	3123	-10
δ(N1H) , ν(C6O) (G), $\delta_{sciss}(NH_2)$, ν(C2O) (C)	1751	1761	11	1748	-3

[a] Abbreviations: ν = stretching; δ = in-plane bending; γ = out-of-plane bending; τ = torsional; asym = asymmetric; sym = symmetric; M1 = Monomer 1; M2= Monomer 2.

3.4 Concluding Remarks

The possibility of accurately predicting the effect of hydrogen-bonding interactions on the infrared spectral features is of fundamental importance in order to correctly interpret the experimental outcomes for hydrogen-bonded systems.
The problems of simulating strongly anharmonic vibrations within hydrogen-bonded bridges are related to both possible limitations of the VPT2 treatment and the requirement of a very accurate description of the underlying PES.
This work shows that the GVPT2 model, combined with a semi-diagonal fourth-order polynomial representation of the anharmonic force field in terms of normal modes, evaluated by means of Density Functional Theory using the B3LYP-D3 functional in conjunction with the N07D basis set, is reliable enough to describe the vibrational properties of hydrogen-bonded systems like uracil-water complexes, with average uncertainties within 22 cm^{-1} in the spectral region of hydrogen-bonding interactions.
An improved description of the harmonic part of the vibrational frequencies of functional groups involved in the hydrogen-bonding interactions has been achieved by application of the ONIOM B2PLYP:B3LYP scheme which is a focused model where only the part of the molecular system forming the hydrogen bonds is treated at B2PLYP level of theory. It leads to a definition of inexpensive hybrid approach, with anharmonic corrections computed at B3LYP-D3 level, allowing to obtain average uncertainties of about 10 cm^{-1} in the spectral region of hydrogen-bonding interactions.
Furthermore, we have addressed the problem of improving the description of the vibrational frequencies of modes which are particularly sensitive to hydrogen-bonding, like the stretching of C=O and N-H functional groups, defining and validating correction parameters for the B3LYP-D3/N07D frequencies of ν(C=O) and ν(N-H) within hydrogen-bonded bridges, through a linear regression based on the more accurate B2PLYP-D3/maug-cc-pVTZ frequencies. Our results point out a wide applicability of such parameters for correcting B3LYP-D3/N07D frequencies of ν(C=O) and ν(N-H) involved in any kind of hydrogen bonds.

Chapter 4

Development of Feasible Protocols for Vibrational Properties of Large Systems: Uracil in the Solid State

In this Chapter we have addressed the problem of defining a computational protocol for simulation of the vibrational properties at anharmonic level for large-size molecular systems, like nucleobases in condensed phases. From a theoretical point of view, the main challenge of simulating the spectroscopic properties of supra-molecular systems consists in the development of a comprehensive yet feasible computational protocol that would be a compromise between accuracy and computational cost.
A quantitative comparison with the experimental measurements can be obtained only treating the vibrational problem beyond the harmonic approximation, but a fully-dimensional anharmonic computation of vibrational frequencies (with either perturbative[134] or variational approaches[211,405,406]) would be too cumbersome with large molecular systems, due to the increasing number of underlying electronic structure computations (energies, gradients or Hessians) necessary to describe multi-dimensional potential energy surfaces (PES).
It is worth noting that the general VPT2 framework developed by Barone and co-workers[199,200,214–216] to compute thermodynamic properties, vibrational energies and transition intensities is not limited by the number of nuclear degrees of freedom, as far as the necessary force-fields are provided. However, such computations are not feasible at relatively high levels of theory with the full VPT2 approach, so that less demanding approaches are very sought after.
Currently, theoretical anharmonic simulations of vibrational spectra of very large molecular systems are still far from standard, but few pioneering applications to the systems like chlorophyll[132,331] or polypeptides,[407,408] glycine[131,134,331] or ammonia[409] adsorbed on inorganic surfaces, graphene[408] or water clusters,[407,408] have shown promising results.

Here we have extended the computational GVPT2-based protocol developed for the dimers of nucleobases, described in the previous Chapters, to a larger system,

namely the uracil heptamer, chosen to represent the properties in the solid state. To study this system within the VPT2 framework, with the aim of reducing the computational cost of anharmonic computations, we have considered and combined two approaches applicable to the medium- to large-size molecular systems, namely: the hybrid model,[134] in which the harmonic frequencies are computed at a higher level of theory with anharmonic corrections obtained by less expensive methods, and the reduced dimensionality VPT2 (RD-VPT2) computations,[131,134,331] in which only selected vibrational modes are calculated anharmonically (including the couplings with the other modes), while the remaining modes are treated at the harmonic level. In both cases, the B3LYP-D3/N07D has been considered as the highest level of theory, whereas in the case of the hybrid approach VPT2 computations have been performed using the analytical formulation[410] of the Self-Consistent-Charge Density Functional Tight Binding (SCC-DFTB, or simply DFTB for conciseness) method,[411,412] which has already shown promising results for computation of anharmonic corrections.[413] Among the semi-empirical methods, the DFTB approach turns out to be suitable for studying the vibrational properties of large systems,[414–416] especially when the parameters are fitted in order to reproduce accurate vibrational frequencies,[414,415,417] or when analytical functions (DFTBA[410]) with continuous first and second derivatives are used to obtain fully analytical second energy derivatives. In this work we have specifically employed the latter version of the method.

4.1 Selection of the Cluster Model for Uracil in the Solid State

The molecular crystal of uracil is a monoclinic system, belonging to the $P21/a$ space group[418] with cell parameters a = 11.938 ± 0.001, b = 12.376 ± 0.0009, c = 3.6552 ± 0.0003 Å, $\beta = 120°54' \pm 0.4'$. The uracil molecules in the crystal interact through N-H$\cdots$O hydrogen bonds, forming parallel layers along the (001) crystal plane, and such layers are held together by stacking interactions, with a regular spacing of 3.136 Å.[418]
In order to study the spectroscopic properties of the uracil crystal,[94,419] the structural parameters obtained by crystallographic measurements of uracil in the solid state[418] have been used to design different cluster models of increasing size to take gradually into account all the inter-molecular interactions influencing the vibrational frequencies, starting from the monomeric unit of uracil (U in Figure 1.1), the two different types of uracil dimers appearing in the crystal (U-U1 and U-U2 in Figure 1.2), which feature two N3-H$\cdots$O4 and one N1-H$\cdots$O4 hydrogen bonds, respectively, up to larger cluster models like those represented in Figure 1.4, composed by four (U4), six (U6), seven (U7), ten (U10a and U10b, based on U-U1 and U-U2, respectively), and fourteen uracil molecules (U14). Furthermore, for reasons related to the evaluation of the binding energies (*vide infra*), four ring-shaped cluster models (r6, r8a, r8b, r10) have been also employed, obtained by removing the one, two and four central uracil units from U7, U10a, U10b and U14, respectively. In all these cluster models the influence of the stacked layers on energies, geometries and vibrational frequencies has been neglected.
The selection of the model has been done studying the convergence of the energy for such clusters of increasing size. The total binding energy of a cluster model composed by $n > 1$ uracil units can be evaluated as the difference between the total energy of the cluster (E_{Un}) and the total energy of n isolated uracil units at the solid state geometry (E_{U1}). By dividing such an energy by n or by the total number of the hydrogen bonds (n^{HB}) occurring in the cluster, the average binding energies per uracil unit ($\langle E_n \rangle^{U}$) or per hydrogen bond ($\langle E_n \rangle^{HB}$) are obtained as:

$$\langle E_n \rangle^{U} = \frac{E_{\mathrm{Un}} - nE_{\mathrm{U1}}}{n} \tag{4.1}$$

and

$$\langle E_n \rangle^{HB} = \frac{E_{\mathrm{Un}} - nE_{\mathrm{U1}}}{n^{HB}}. \tag{4.2}$$

However, both these approaches tend to treat all the uracil units in the same fashion, while the clusters are intended to be focused models, providing a more realistic description for the central uracil unit, considering the outer ones as an environment. For this reason, in the case of the largest cluster models – U7, U10a, U10b and U14 – a more suitable definition of the average hydrogen bond energy can be formulated by calculating the total binding energy as the difference between the total energy of

the cluster and the total energy of the non-interacting central units (E_c) – U, U-U1, U-U2 and U4, respectively – and rings (E_r) – r6, r8a, r8b and r10, respectively –

$$\langle E_{\mathrm{c+r}} \rangle^{HB} = \frac{E_{\mathrm{Un}} - (E_c + E_r)}{n^{HB}_{\mathrm{c+r}}}, \tag{4.3}$$

where E_{Un} is related to U7, U10a, U10b and U14. In this case the normalization is performed by dividing the energy by the number of the hydrogen bonds connecting the central units and the rings ($n^{HB}_{\mathrm{c+r}}$).
It has been also verified that the latter approach is less affected by the basis set superposition error (BSSE).[420]
All the total energies have been obtained by single point calculations at the B3LYP-D3 and DFTBA levels, using the experimental crystallographic geometries, without performing any optimization.
Figures 4.1 and 4.2 report the resulting $\langle E_n \rangle^{U}$, $\langle E_n \rangle^{HB}$ and $\langle E_{\mathrm{c+r}} \rangle^{HB}$ as a function of the cluster size. The convergence of normalized binding energies is estimated by corresponding fit functions ($f(n)$), which are shown on the graph along with coefficients of determination (R^2) and the error bars computed as mean absolute errors (MAEs) between binding energies and the $f(n)$. Despite different absolute values, the trends obtained with the B3LYP-D3 and the DFTBA methods are quite similar and the energy convergence is reached at the heptamer level. Furthermore, for both B3LYP-D3 and the DFTBA methods, the value of $\langle E_{\mathrm{c+r}} \rangle^{HB}$ remains almost constant in the different cluster models, with the more accurate B3LYP-D3 result of about 6 kcal mol^{-1} (average of 5.7 kcal mol^{-1} and MAE of 0.2 kcal mol^{-1}) despite the size of the system.
Then, geometry optimizations with tight cutoffs on forces and stepsize convergence have been performed for U, U-U1, U-U2 and U7 models, at the B3LYP-D3/N07D and the DFTBA levels. Comparison of the structural parameters of the optimized geometry with the available experimental crystallographic data for uracil in the solid state,[418] shows that the optimized geometries are still a good description of the real crystalline system, with average errors of about 4% with the B3LYP-D3/N07D method and 8% at the DFTBA level for all the three cluster models.
Further confirmation for the choice of the model is obtained comparing the full VPT2 anharmonic frequencies for U, U-U1 and U-U2, previously computed at the optimized geometries using the B3LYP-D3/N07D method, with the experimental data of the uracil crystal.
The experimental characterization of the IR spectroscopic features of poly-crystalline uracil has been performed at the INAF-Astrophysical Observatory of Arcetri with a Vertex 70v (Bruker) vacuum FT-IR spectrometer, equipped with a Praying Mantis Diffuse Reflection Accessory (Harrick DRIFT), in order to use the Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) technique, which is particularly suitable for the study of solid powder samples (*vide infra* for experimental frequencies).
Comparing the theoretical results with experiment, we have observed that the

monomeric and dimeric models provide only a partial description of the effect on vibrational frequencies due to the intermolecular interactions present in the solid state, giving large errors for several modes, especially the ones related to the moieties involved in the hydrogen bonds. For example, the errors of both the N1-H1 and N3-H3 stretching frequencies of U are larger than 350 cm^{-1}, as well as the errors related to the frequency of the C=O stretching mode (about 100 cm^{-1}). Moreover, each of the dimers accounts only for specific hydrogen-bonding interactions (N3-H$\cdots$O4 or N1-H$\cdots$O4), whereas errors of the same magnitude are found for the frequencies of the other N-H stretching modes not involved in the hydrogen bonds. This suggests that the correct inclusion of all the hydrogen bonds of the crystal is required in order to assign and predict the experimental infrared spectrum in the solid state. In this sense, the uracil heptamer can be viewed as a focused model in which the hydrogen bonds between the central monomer and its surroundings (N1-H1-O4-H3-N3 motif) are included consistently with the periodic crystal, leading to a model mimicking the chemical interactions occurring in the solid state uracil.
However, such a system involves 84 atoms, resulting in 246 normal modes, and the complete VPT2 treatment requires 493 Hessian computations in order to obtain the full semi-diagonal quartic force field via numerical differentiation of the analytical second derivatives. Therefore, the full VPT2 treatment is not feasible at the B3LYP-D3/N07D level for this system, and less expensive methods such as the hybrid B3LYP-D3/DFTBA scheme and the RD-VPT2 approach have been applied. More details can be found in the original paper.[172]

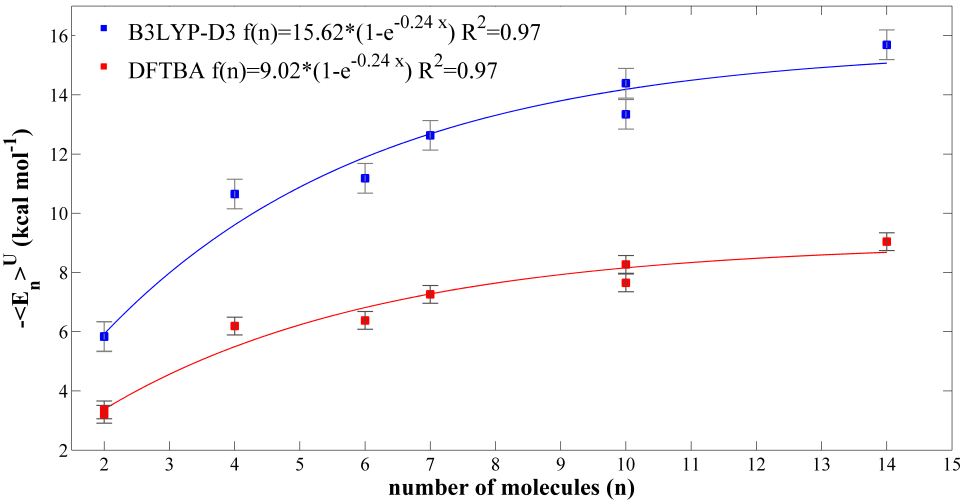

Figure 4.1: Normalized binding energies for the different uracil cluster models, $\langle E_n \rangle^U$ (Eq. 4.1), with each energy term obtained through single point calculations on the experimental crystallographic structure[418] performed at the B3LYP-D3 (blue squares) and DFTBA (red squares) levels. Computed values are fitted to the functional form shown on the graph along with coefficients of determination (R^2). The error bars correspond to the mean absolute errors (MAEs) between computed binding energies and the ($f(n)$).

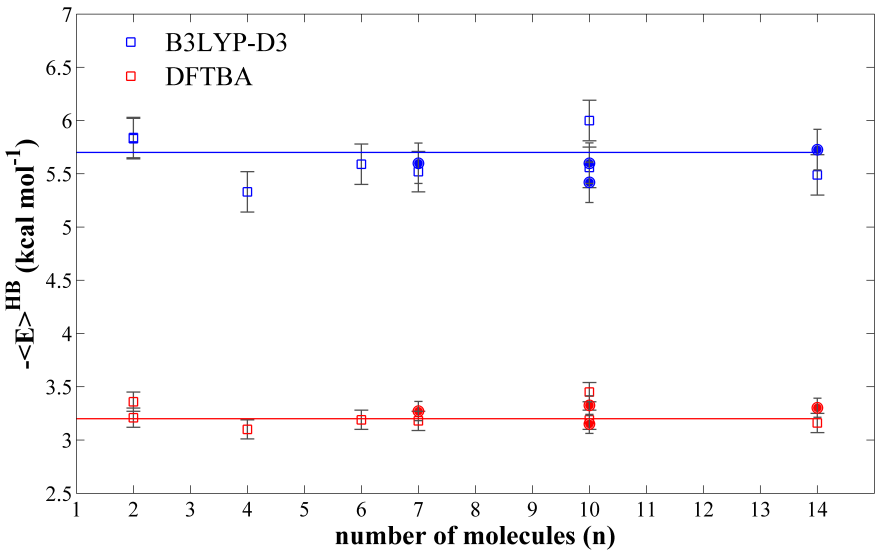

Figure 4.2: Mean hydrogen bond energies for the cluster models $\langle E \rangle^{HB}$: $\langle E_n \rangle^{HB}$ (Eq. 4.2, squares) and $\langle E_{\mathrm{c+r}} \rangle^{HB}$ (Eq. 4.3, circles), evaluated at the B3LYP-D3 (blue) and DFTBA (red) levels. The average $\langle E \rangle^{HB}$ over all cluster models are shown as a horizontal lines. The error bars correspond to the mean absolute errors (MAEs) between computed binding energies and the average $\langle E \rangle^{HB}$.

4.2 Validation of Hybrid and Reduced Dimensionality VPT2 Approaches

Validation for the hybrid and RD-VPT2 approaches has been carried out by comparison with the B3LYP-D3/N07D full-dimensional VPT2 results for uracil dimers. Then, such approaches have been used for computations of anharmonic vibrational frequencies for the heptamer model of uracil in the solid state.

4.2.1 Hybrid B3LYP-D3/DFTBA Method

The Hybrid approach[134,221,227,228,240,328,413] is particularly suitable for simulating large systems, in order to reduce the computational burden associated with the calculation of the anharmonic frequencies. Within a hybrid scheme the harmonic part is computed at a higher level of theory and/or with larger basis sets, while the anharmonic corrections are evaluated by computationally less demanding models (see Ref.[134] and references therein). In this work the harmonic part has been evaluated at the B3LYP-D3/N07D level and the anharmonic corrections have been included in an approximate fashion with the DFTBA method.
Proper use of the hybrid approach requires that the vibrational modes computed with the two different methods have a biunivocal correspondence, otherwise the respective anharmonic force fields should be transformed accordingly, which would be unfeasible for large molecular systems. The similarity between two sets of normal modes can be inferred by analyzing a "Duschinsky-like matrix"[421] ($\boldsymbol{J}^{HL}$). Given $\boldsymbol{L}^{H}$ and $\boldsymbol{L}^{L}$ the rectangular matrices which define the mass weighted cartesian - mass weighted normal coordinates transformation of the high and low level methods, respectively, the $\boldsymbol{J}^{HL}$ matrix is defined as the product matrix $\boldsymbol{L}^{H}\boldsymbol{L}^{L}$, and it deviates from the identity matrix as far as the normal mode definition is different between the two methods. With such an approach, $\boldsymbol{J}^{HL}$ describes the projection of the normal coordinate basis vectors computed at the DFTBA level on those computed at the B3LYP-D3/N07D level. The latter, are used for the assignments of molecular vibrations, based on inspection of the atomic displacements along normal modes, while normal modes computed at the DFTBA level are assigned as linear combinations of the B3LYP-D3/N07 ones, as derived from the analysis of the $\boldsymbol{J}^{HL}$ matrix coefficients. For example, highlighted DFTBA vibrations (in columns, see Figure 4.3) correspond to the combinations of normal modes assigned as γ(CH), γ(C4O)(B3LYP-D3/N07D, in rows).
Figure 4.3 reports $\boldsymbol{J}^{HL}$ related to the B3LYP-D3/N07D and the DFTBA modes of the uracil heptamer, using a color scale for the magnitude of the $\boldsymbol{J}^{HL}$ matrix elements (the darker the more it is close to unity). In this figure sub-sets of normal modes which are consistently defined by the two methods can be identified, suggesting that the hybrid approach can be applied.

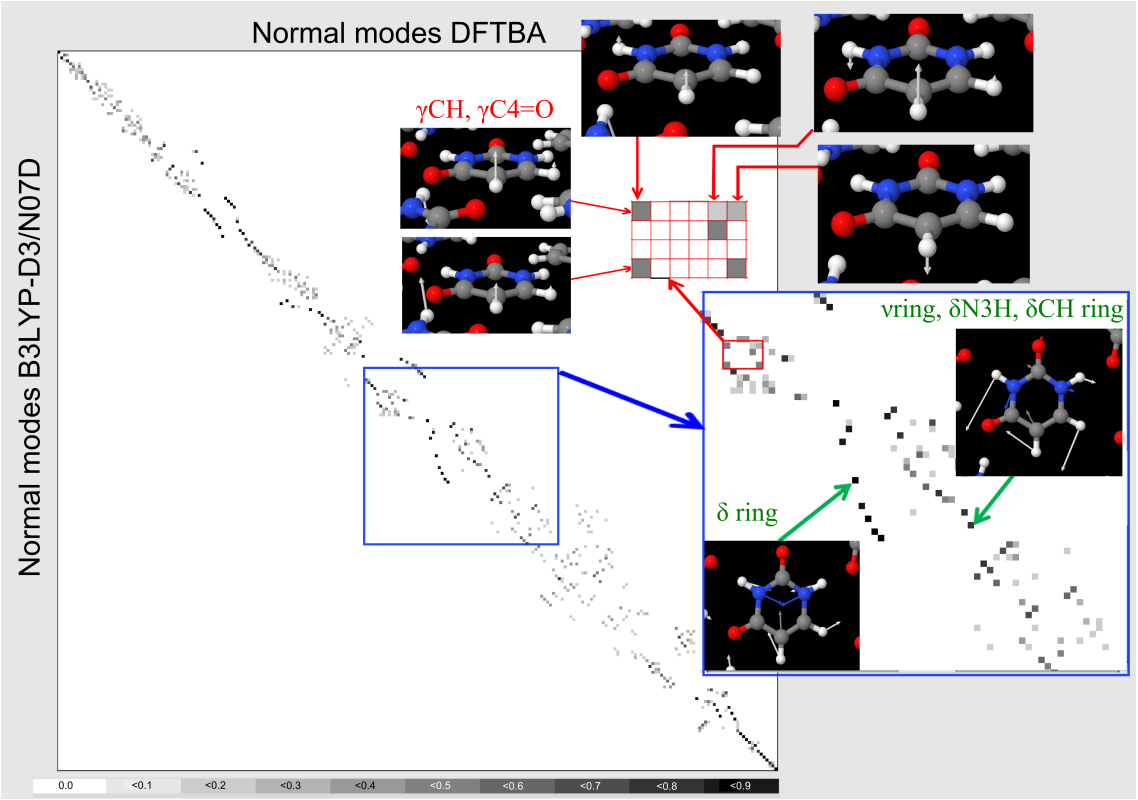

Figure 4.3: Graphical representation of the correspondences between normal modes computed with B3LYP-D3/N07D and DFTBA methods for the uracil heptamer. Assignments were performed based on the atomic displacements along normal modes. $\boldsymbol{J}^{HL}$ matrix and normal modes are visualized using the VMS-Draw tool,[167] a gray scale is used to represent the magnitude of each element of the $\boldsymbol{J}^{HL}$ matrix, with a white square for a near-null value and a black square for 1 (equivalent normal modes). Selected modes of the central uracil unit with large and negligible mode mixing are marked by red and green arrows, respectively, and shown in the insets.

Regarding the frequencies, a statistical analysis of the deviations between the ones computed with the DFTBA method and the ones at the B3LYP-D3/N07D level for the uracil monomer and the dimers U-U1 and U-U2 is presented in Table 4.1. Although the absolute values of the harmonic and anharmonic DFTBA vibrational frequencies show large differences with respect to B3LYP-D3/N07D for both the dimers and the monomer, the anharmonic corrections computed with DFTBA method are sufficiently close to the ones computed at the B3LYP-D3/N07D level, with MAEs of 16.7 cm^{-1} for the uracil monomer, 43.6 cm^{-1} for U-U1, and 27.9 cm^{-1} for U-U2. Moreover, the most significant deviations of the anharmonic shifts for the dimers are related to the stretching modes of the NH groups involved in the hydrogen bonds, namely both the N3H groups of U-U1, and the N1H group of the monomer M1 in U-U2. However, it is worth noting that such strongly anharmonic vibrations usually require accurate description of the underlying PES,[326] which is difficult to achieve especially when approximate approaches such as the DFTBA method are employed. Excluding these modes, the MAEs are lower than about 22-24 cm^{-1} for both dimers, with maximum deviations of about 40 cm^{-1}. These results indicate that the DFTBA method can be used for the computation of the anharmonic corrections within the hybrid scheme, accordingly to Ref.[413]

Table 4.1: Deviations of harmonic and anharmonic vibrational frequencies (cm^{-1}) computed with the DFTBA method respect to the B3LYP-D3/N07D results for the uracil monomer and the dimers U-U1 and U-U2. In brackets the mean absolute errors (MAEs), the minimum negative deviations (MIN) and the maximum positive deviations (MAX) obtained by exclusion of the N3H stretching vibrational modes of both the monomeric units in the dimer U-U1, and the N1H stretching vibrational mode of the monomer 1 in U-U2, respectively.

DFTBA vs B3LYP-D3			
	Δharm[a]	Δanh[b]	Δanh shift[c]
	MAE		
u	65.3	55.3	16.7
u2a	141.9	155.7	43.6 (22.1)
u2b	133.3	126.3	27.9 (24.0)
	MIN		
u	-233	-198	-40
u2a	-270	-580	-310 (-40)
u2b	-471	-456	-102 (-40)
	MAX		
u	180	149	39
u2a	178	146	38 (38)
u2b	166	136	43 (43)

[a] Deviation between harmonic frequencies. [b] Deviation between anharmonic frequencies. [c] Deviation of anharmonic shifts.

4.2.2 RD-VPT2 Method

The Reduced Dimensionality VPT2 (RD-VPT2) approach[131,134,331] has been employed as an alternative for the evaluation of approximate anharmonic frequencies. Within the RD-VPT2 approach, the most demanding computation, namely the numerical differentiation of analytical Hessian, is performed only along a restricted set of "active modes". Such selected modes are included into the VPT2 treatment along with the couplings with the other modes which, instead, are treated at the harmonic level. A detailed description of such an approach is reported in Appendix A and has been given in Refs.[134,331] Here we just note that the reliability of the RD-VPT2 approach is ensured when the set of active modes is large enough to include all the couplings between the modes contributing to the frequencies of interest. The correct choice of the active modes can be based on the nature of the vibrations and the related energy range, allowing to take into account all the vibrations of specific functional groups, which have similar frequencies and are likely to be coupled.[331] Alternatively, lower-level electronic structure computations, resonance conditions, or normal mode similarity (See Ref.[422] and references therein) can be also applied to define the reduced dimensionality schemes.

An important drawback of this approach is that the computational cost is strongly dependent on the specific chemical system under study. In fact, the more the normal modes are coupled with each other, the larger is the number of modes which need to be included in the computations. However, for systems like organic molecules adsorbed on inorganic surfaces, usually the couplings between the molecular vibrations and the surface phonons are very weak, and the RD-VPT2 approach allows to sensibly decrease the computational demand of the calculations.

In this work the validation of the active modes has been performed on the basis of the full dimensionality B3LYP-D3/N07D anharmonic computations available in the case of the dimers, comparing those full dimensional vibrational frequencies with the anharmonic frequencies obtained through the RD-VPT2 treatment.

As reported in Table 4.2, choosing as active modes those corresponding to the high-frequency region, namely ν(N1H), ν(N3H), ν(C5H), ν(C6H), ν(C2O), ν(C4O), ν(C5C6), δ(N1H) and ν(ring), it is possible to obtain anharmonic frequencies which are in good agreement to the full dimensional VPT2 ones. It can be noted that the overall deviations between the full dimensional and RD-VPT2 frequencies are very small, with MAEs of 6.2 cm^{-1} and 4.9 cm^{-1} for U-U1 and U-U2, respectively, and maximum deviations below 35 cm^{-1}. Therefore we have assumed that such vibrational modes may constitute a minimum "active set" to obtain a reliable prediction of the higher vibrational frequencies for the uracil molecule in the center of the heptamer cluster.

Table 4.2: Anharmonic vibrational frequencies (cm^{-1}) for the hydrogen-bonded uracil dimers U-U1 and U-U2, computed at B3LYP-D3/N07D level through a fully anharmonic and reduced dimensionality VPT2 (RD-VPT2) treatment. Mean absolute errors (MAEs), minimum negative deviations (MIN) and maximum positive deviations (MAX) with respect to the fully anharmonic VPT2 treatment are also reported.

U-U1			
	FULL-VPT2	RD-VPT2	
mode	anh	anh	Assignment[a]
66	3476	3473	ν(N1H) (M1,M2)
65	3476	3473	ν(N1H) (M1,M2)
62	2998	2977	ν(N3H) (M1,M2)
59	2960	2926	ν(N3H) (M1,M2)
64	3124	3121	ν(C5H) (M1,M2)
63	3128	3126	ν(C5H) (M1,M2)
61	3089	3088	ν(C6H) (M1,M2)
60	3080	3078	ν(C6H) (M1,M2)
58	1788	1782	ν(C2O) (M1,M2)
57	1783	1776	ν(C2O) (M1,M2)
56	1696	1692	ν(C4O), ν(ring), δ(N3H) (M1,M2)
55	1687	1684	ν(C4O), ν(ring), δ(N3H) (M1,M2)
54	1630	1633	ν(C5C6) (M1,M2)
53	1643	1633	ν(C5C6) (M1,M2)
52	1473	1478	δ(NH) + δ(CH) (M1,M2)
51	1474	1479	δ(NH) + δ(CH) (M1,M2)
50	1447	1453	δ(N3H) (M1,M2)
49	1444	1448	δ(N3H) (M1,M2)
48	1390	1389	ν(ring), δ(N1H) (M1,M2)
47	1387	1386	ν(ring), δ(N1H) (M1,M2)
MAE		6.2	
MIN		-34	
MAX		6	
U-U2			
	FULL-VPT2	RD-VPT2	
mode	anh	anh	Assignment[a]
66	3472	3477	ν(N1H) (M2)
63	3121	3117	ν(N1H) (M1)
65	3440	3442	ν(N3H) (M1)
64	3435	3436	ν(N3H) (M2)
62	3124	3123	ν(C5H) (M1)
59	3074	3068	ν(C5H) (M2)
61	3088	3086	ν(C6H) (M2)
60	3081	3082	ν(C6H) (M1)
58	1779	1780	ν(C2O) (M2)
57	1761	1755	ν(C2O) (M1)
56	1747	1723	ν(C4O) (M1)
55	1694	1678	ν(C4O) (M2)
54	1639	1633	ν(C5C6) (M1)
53	1641	1628	ν(C5C6) (M2)
52	1522	1524	ν(ring), δ(N1H) (M1)
51	1468	1467	ν(ring), δ(N1H) (M2)
50	1399	1404	δ(NH) + δ(CH) (M1)
49	1391	1390	δ(NH) + δ(CH) (M2)
48	1386	1387	ν(ring), δ(NH) (M2)
47	1379	1380	ν(ring), δ(N1H) (M1)
MAE		4.9	
MIN		-24	
MAX		6	

[a] Abbreviations: ν = stretching; δ = in-plane bending; γ = out-of-plane bending; M1 = Monomer 1 in the uracil dimers; M2 = Monomer 2 in the uracil dimers.

4.3 Computations *vs.* Experiment

The hybrid B3LYP-D3/DFTBA approach has been applied to the heptamer model, providing anharmonic frequencies whose harmonic part has been computed at the B3LYP-D3/N07D level while the anharmonic correction has been evaluated by means of the DFTBA method.
In a first step, we have selected, among the 246 normal modes, the vibrational frequencies relative to the central uracil molecule, which has been considered as the target for predicting the vibrational frequencies of uracil in the solid state. Due to the unavoidable mixing of the contributions of the various atoms to the normal modes occurring in very large systems like the heptamer cluster, the assignment of the normal modes of the central uracil molecule has been performed not only by visual inspection but also examining in detail the values of the atomic displacements along the normal modes. It can be noted that such procedure can be facilitated by localized mode approaches[407,408,423] or use of other (i.e. mixed Cartesian-internal) coordinate systems. Then, we have analyzed the $\boldsymbol{J}^{HL}$ matrix of the heptamer to verify the correspondences between the modes of the central uracil molecule computed with the B3LYP-D3/N07D method and the ones computed at DFTBA level. In particular, we have found four modes for which the correspondence is not unique, namely the δ(CH), ν(ring), δ(NH) mode at 1424 cm^{-1}, the γ(CH), γ(C4O) modes at 873 cm^{-1} and 851 cm^{-1}, the δ(ring) mode at 592 cm^{-1} (all frequency values and assignments refer to the B3LYP-D3/N07D harmonic computations). Thus, such modes have been excluded from the hybrid calculation and afterwards treated with a RD-VPT2 approach.
Table 4.3 reports the anharmonic frequencies of the central uracil unit in the heptamer model computed with the hybrid B3LYP-D3/DFTBA approach. As shown, such an approach provides anharmonic vibrational frequencies that are in rather good agreement with the experimental ones, with particular exception for those corresponding to the stretching of the N-H groups and, in some extent the carbonyl groups, which are involved in the hydrogen bonds. As discussed above for uracil dimers, vibrations involved in hydrogen-bonded bridges are characterized by an underlying strongly anaharmonic PES which is poorly described by the less accurate DFTBA method. Indeed, despite the MAE is 32.0 cm^{-1} taking into account all the vibrational modes, it lowers to 11.8 cm^{-1} excluding the modes in the high-frequency region (reported in brackets in Table 4.3), similarly the minimum negative and maximum positive deviations decrease from -202 cm^{-1} to -7 cm^{-1} and from 42 cm^{-1} to 34 cm^{-1}, respectively.

The hybrid B3LYP-D3/DFTBA results for the heptamer have then been improved by application of the RD-VPT2 treatment, which has been used to compute the anharmonic frequencies of the 4 modes excluded from the hybrid calculation due to their non-univocal correspondence between the B3LYP-D3/N07D and DFTBA methods (as discussed above) along with the highest vibrational frequencies, largely influenced by the hydrogen bonds, namely the 11 vibrational modes of the central

uracil molecule in the heptamer corresponding to the ν(N1H), ν(N3H), ν(C5H), ν(C6H), ν(C2O), ν(C4O), ν(C5C6), δ(N1H) and ν(ring) vibrational modes, previously validated as preferential active set.
The RD-VPT2 frequencies have been reported in Table 4.3 (highlighted in bold) and combined with the hybrid B3LYP-D3/DFTBA ones within the final hybrid+RD-VPT2 set, which has been compared with the experimental data for uracil in the solid state.[103,419] With this approach a better agreement with the highest experimental frequencies of the solid uracil has been achieved and the overall MAE, obtained mixing the RD-VPT2 frequencies with the hybrid ones, turns out to be of 17.3 cm^{-1}. Such an error can be considered sufficiently accurate, in particular taking into account the size of the system and some uncertainties of the experimental frequencies related to the N-H stretching vibrational modes.[103,419,424]

Furthermore, it is also possible to extend the computational model (based on additivity assumptions) applying corrections to the harmonic frequencies of the central unit, based on the best theoretical estimates for the uracil monomer, along with the anharmonic and complexation effects at the hybrid+RD-VPT2 level as described above. In the present case we have used the harmonic frequencies of the uracil monomer evaluated by a composite scheme at Coupled Cluster (CC) level (CCSD(T)/CBS(T,Q+aug)+CV),[119] and the results are reported in the last column of the Table 4.3. It can be noted that such correction leads to some improvement for the highest frequency modes, related to the N-H stretching vibrations, but on the whole it does not improve over the B3LYP-D3/DFTBA results, highlighting the good accuracy of the latter.

Finally, we have compared the anharmonic results for the uracil heptamer with the harmonic and anharmonic frequencies for the uracil monomer and the harmonic B3LYP-D3/N07D values for uracil heptamer, also listed in Table 4.3. It is clear that neither harmonic values or best anharmonic estimates for monomer are adequate for the assignment of experimental data, showing discrepancies as large as several hundreds of wavenumbers for modes involved in hydrogen-bonding interactions in solid uracil. The situation is improved once the complexation effects are accounted for in the uracil heptamer, but clearly only including both hydrogen-bonding in molecular cluster and anharmonic effects it is possible to obtain sufficiently accurate results.

Moreover, to verify the reliability and accuracy of our computational model, Table 4.4 compares the shifts of the experimental frequencies of uracil in the solid state[103,419] with respect to uracil isolated in Argon matrix[364–366] and the corresponding shifts obtained from the simulations for the uracil dimers and the heptamer, i.e. the shifts of the fully anharmonic frequencies of the dimers and the RD-VPT2 anharmonic frequencies of uracil in the heptamer with respect to the fully anharmonic frequencies for the uracil monomer, computed at the B3LYP-D3/N07D level.[170] The results show that the heptamer model is more suitable than the dimers for assigning the IR bands of the solid uracil. Indeed, both models based on the dimers provide only a partial description of the multiple interactions involving the uracil molecules

Table 4.3: Vibrational frequencies (cm^{-1}) for the uracil monomer and central uracil unit in the heptamer model, obtained at different levels of theory including the hybrid B3LYP-D3/DFTBA and the reduced dimensionality VPT2 (RD-VPT2) approaches, with corrections based on the best estimated harmonic frequencies available for the uracil monomer evaluated by composite scheme at CCSD(T)/CBS(T,Q+aug)+CV level.[a] Mean absolute errors (MAEs), minimum negative deviations (MIN) and maximum positive deviations (MAX) of computed vibrational frequencies with respect to the experimental data for uracil in the solid state[b,c] are also reported.

Uracil Monomer			**Uracil Heptamer**				**Solid Uracil**	
B3LYP-D3 /N07D	**CCSD(T) /CBS+CV[a]**	**CCSD(T) /B3LYP[a]**	**B3LYP-D3 /N07D**	**Hybrid B3LYP-D3/DFTBA**	**RD-VPT2 +Hybrid**	**CC+RD-VPT2 +Hybrid**	**Experimental**	
harm	harm	anh	harm	anh	anh	anh		Assignment[d]
3647	3653	3472	3298	2999	**3074**	**3079**	3106 sh [b]	ν(N1H)
			3243	2904	**3076**	**3081**		
3607	3602	3430	3266	2935	**3036**	**3031**	3000 m [b]	ν(N3H)
3261	3253	3113	3231	3109	**3069**	**3060**	3088 s [b]	ν(C5H), ν(N1H)
3219	3218	3062	3208	3005	**3053**	**3051**	3040 m [c]	ν(C6H)
1810	1790	1771	1792	1803	**1746**	**1725**	1761 vs [c]	ν(C2O)
			1773	1781	**1725**	**1705**		
1776	1762	1747	1691	1634	**1651**	**1636**	1652 m [c]	ν(C4O)
1675	1678	1640	1652	1624	**1615**	**1618**	1616 w [c]	ν(C5C6)
1504	1505	1455	1570	1520	**1537**	**1538**	1509 s [c]	δ(N1H)
1409	1414	1367	1505	1481	**1471**	**1477**	1456 vs [c]	ν(ring), δ(N1H), δ(N3H)
1421	1427	1390	1458	1436	1436	1443	1421 vs [c]	δ(N3H) + δ(CH)
1385	1394	1349	1424		**1395**	**1405**	1391 s [c]	δ(CH), ν(ring), δ(NH)
1232	1248	1204	1285	1276	1276	1292	1242 vs [c]	ν(ring), δ(NH), δ(CH)
1089	1084	1064	1133	1115	1115	1110	1099 w [c]	ν(ring), δ(CH), δ(N1H)
970	968	947	1033	1019	1019	1016	1006 vs [c]	ν(ring), δ(N3H), δ(CH)
990	995	978	1000	995	995	1001	994 s [c]	δ(ring)
968	973	946	873		**781**	**786**	828 s [c]	γ(CH), γ(C4O)
815	814	803	851		**771**	**770**	807 m [c]	γ(CH), γ(C4O)
770	765	748	801	796	796	791	782 w [c]	ring breathing
761	773	753	772	762	762	774	762 s [c]	γ(C2O)
562	545	549	592		**585**	**569**	585 s [c]	δ(ring)
544	541	534	586	586	586	583	569 vs [c]	δ(CO), δ(ring)
523	517	515	552	547	547	541	554 m [c]	δ(ring)
			542	554	554	548		
95.1	95.0	73.2	58.9	32.0 (11.8)	17.3	21.7		MAE
-47	-42	-89	-12	-202 (-7)	-47	-56		MIN
607	602	430	266	42 (34)	36	50		MAX

[a] From Ref.[119] [b] From Ref.[419] [c] From Ref.[103] [d] Abbreviations: ν = stretching; δ = in-plane bending; γ = out-of-plane bending.

in close contact in the solid phase, while in the heptamer model the central molecule is completely surrounded by other uracil molecules simulating more realistically the actual environment of each uracil molecule in the crystal structure. Experimentally, the remarkable effect of the intermolecular interactions taking place through strong hydrogen bonds between the proton donor and acceptor moieties is rather evident, with the most significant red-shifts with respect to isolated uracil of about 376 cm^{-1} for ν(N1H1), 433 cm^{-1} for ν(N3H1), 42 cm^{-1} for ν(C5H5), 81 cm^{-1} for ν(C4O4). These findings are all well reproduced by the anharmonic computations, using the heptamer as molecular model. Considering the accuracy of fully anharmonic frequencies of the uracil monomer of about 10 cm^{-1}, these results indicate that the B3LYP-D3/N07D RD-VPT2 treatment provides a good prediction of the vibrational frequencies for the heptamer.

Table 4.4: Shifts of the experimental frequencies (cm^{-1}) of uracil in the solid state[b] with respect to uracil isolated in Argon matrix[c], compared with the shifts (cm^{-1}) obtained by B3LYP-D3/N07D computations for the uracil dimers U-U1 and U-U2 (fully anharmonic frequencies) and the heptamer (RD-VPT2 frequencies) with respect to the fully anharmonic frequencies of the uracil monomer.

	$\Delta\nu$ Experimental	**$\Delta\nu$ Calculated**		
Assignment[a]		U-U1	U-U2	U7
ν(N1H)	**-376**	3 (M1,M2)	-1 (M2)	**-399**
		3 (M1,M2)	**-351 (M1)**	
ν(N3H)	**-433**	**-432 (M1,M2)**	11 (M1)	**-393**
		-470 (M1,M2)	5 (M2)	
ν(C5H)	-42	15 (M1,M2)	15 (M1)	-40
		20 (M1,M2)	-35 (M2)	
ν(C2O)	-1	18 (M1,M2)	9 (M2)	-24
		13 (M1,M2)	-10 (M1)	
ν(C4O)	**-81**	-53 (M1,M2)	-2 (M1)	**-98**
		-62 (M1,M2)	-55 (M2)	
ν(C5C6)	-28	-9 (M1,M2)	0.3 (M1)	-23
		5 (M1,M2)	2 (M2)	

[a] Abbreviations: ν = stretching; δ = in-plane bending; γ = out-of-plane bending; M1 = Monomer 1 in the uracil dimers; M2 = Monomer 2 in the uracil dimers. [b] From Ref.[103,419] [c] From Ref.[364–366]

Afterward, we have attempted to assign some of the experimental bands of polycrystalline uracil, not assigned previously due to the lack of reliable computational data for the comparison.[103] Figure 4.4 shows the experimental spectrum of solid uracil, measured using the Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) technique, providing the most complete assignment up to date for this molecular system.

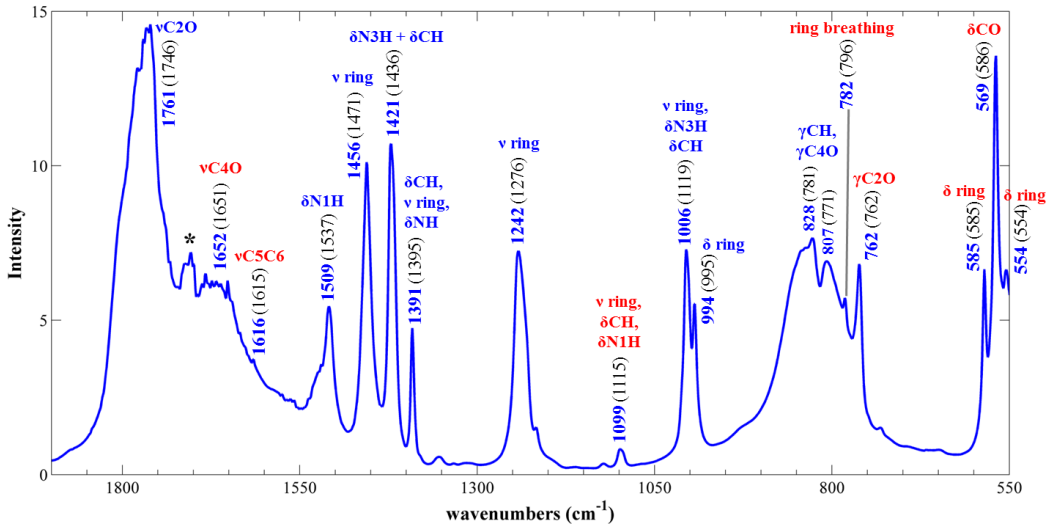

Figure 4.4: Experimental IR spectrum of solid uracil in the 550-1900 cm^{-1} range[103] along with experimental (blue) and theoretical (black, in parenthesis) wavenumbers. Asterisk corresponds to tentative assignment as non-fundamental transition. Band assignment proposed in this work is marked in red.

4.4 Concluding Remarks

In this Chapter we have presented a reliable and feasible computational protocol for simulating the IR spectra of quite large systems at anharmonic level, combining the hybrid B3LYP-D3/DFTBA approach with the RD-VPT2 one. In this way, we have obtained an overall accurate description of the vibrational frequencies of uracil in the solid state with a MAE of 17.3 cm^{-1}. These theoretical results reproduce very well the significant experimental shifts of the vibrational frequencies of uracil due to the intermolecular hydrogen bonds in the solid state with respect to uracil isolated in Argon matrix,[364–366] allowing us also to provide some new assignments of the experimental spectrum of uracil in the solid state.[103]
Such a protocol is particularly promising to analyze and assign the IR spectroscopic features and, hence, to investigate the properties and processes of large systems, ranging from molecular clusters, biopolymers, up to macromolecular systems in complex environments like biomolecules adsorbed on solid supports, which are particularly relevant in the prebiotic context.

Chapter 5

Study of Molecule-Mineral Interactions: Binding of Nucleic Acids Components to Brucite

A fundamental question in Astrobiology concerns the transition from the geochemical world to the biochemical one. This transition probably occurred as a consequence of the combination of multiple complex phenomena between organic and inorganic systems. In this context, studying how organic molecules interact with minerals may be a step forward in resolving unsolved questions about the origin of life.
Investigations on the binding mechanism of "building blocks of life" such as nucleic acid components to mineral surfaces, the stability and reactivity of such molecule-mineral complexes under plausible prebiotic conditions, would give key inputs to figure out the physico-chemical phenomena that led to the emergence of life on Earth.

In this work, we have considered astrobiologically relevant environments like submarine serpentinite-hosted hydrothermal vents and, specifically, the Lost City hydrothermal field, that is a kind of off-axis vent located several kilometers away from the volcanic spreading zones. In such locations, through the serpentinization process, fluid-rock reactions produce methane- and hydrogen-rich fluids, that are highly alkaline (pH 9 to 11), with temperatures ranging from <40 °C to 90 °C.
Quite different conditions are found in the case of the Black Smokers, which are located directly above magma chambers and emit hot chemically modified water (up to 405 °C), with acidic effluents rich in dissolved transition metals, e.g. Fe(II) and Mn(II), plus CO_2, H_2S and CH_4 since they are fueled by volcanoes.
The serpentinization process occurring in the Lost City field is a geological low-temperature metamorphic alteration involving the hydration of magnesium-rich olivine forsterite, in which Fe^{2+} in the rocks reduces H_2O to produce Fe^{3+}, H_2, CH_4 and hydrocarbons. The resulting minerals are magnetite (Fe_3O_4), brucite ($Mg(OH)_2$) and a hydroxylated magnesium-iron silicate called serpentine ($Mg_{2.85}Fe_{0.15}Si_2O_5(OH)_4$).[333]

The Lost City hydrothermal field has been suggested as a plausible environment for the origin of life because the disequilibria and strong redox gradients established by mixing the hot and highly reducing hydrothermal fluids with the overlying cold seawater may provide the energy necessary for the synthesis of key prebiotic molecules.[47,48,332] Additionally, lower temperatures typical of the Lost City hydrothemal fluids seem to favor biosynthesis.[332,425] Thus, it has been speculated that these locations might have been incubators for the emergence of the first lifeforms. The observation that present-day hydrothermal vent microorganisms harbour physiological characteristics that resemble the earliest microbial ecosystems on Earth is an evidence in support of this hypothesis.[47]
In this geochemical context, we focused on the adsorption properties of magnesium hydroxide, named brucite ($Mg(OH)_2$), which is a product of the serpentinization process,[333] with the aim of evaluating potential roles that serpentinite-hosted hydrothermal minerals play in concentrating and selecting fundamental prebiotic molecules such as nucleic acids components from dilute aqueous environments, and potentially in catalyzing the formation of more complex species in plausible prebiotic conditions.[334]
Mg-containing minerals are supposed to have played critical roles in prebiotic geochemistry.[426] Indeed, magnesium is the eighth most abundant element on Earth, the third element in the marine water, a common element also on the other Earth-like planets and, noteworthy, magnesium cations are involved in many fundamental processes of the modern biochemistry, being essential in the function of ribozymes, having a preserved position in ribosomes, a role in the cellular metabolism, a six-oxygen coordination ability responsible for stabilizing di- and tri-phosphates groups of nucleotides, assisting the folding of RNA and promoting the formation of nucleobases and carbohydrates like ribose. Therefore, the study of the adsorption properties of Mg-containing minerals may shed light on the possible involvement of magnesium in prebiotic geochemical processes.

This study has been performed during a short-term internship at the Geophysical Laboratory of the Carnegie Institution of Washington, DC, USA.

5.1 Synthesis and Characterization of Brucite Mineral

5.1.1 Brucite Synthesis and Hydrothermal Treatment

Microcrystalline brucite powder was synthesized using a precipitation method reported in the literature,[334,427,428] adding a 2.0 M $MgCl_2$ aqueous solution (99.99% Alfa Aesar dry ampoules under Argon gas) dropwise, at a rate of approximately 2.5 mL min^{-1} using a graduated buret, to a 2.0 M NH_4OH aqueous solution (Sigma-Aldrich, 25 wt.% solution) kept at a temperature of 45 ± 3 °C under magnetic stirring. After precipitation of brucite, the suspension was cooled down for 3 days at room temperature and the soluble reaction byproducts were washed from the precipitate using 18.2 MΩ·cm Milli-Q water (Millipore). Finally, the precipitate was dried at 135 °C using a hotplate.
Then, the crystallinity and morphology of the synthetic brucite sample was improved through hydrothermal aging treatment, performed by sealing an aqueous suspension of the synthetic brucite within a PTFE-lined stainless steel reaction vessel (Col-Int-Tech) and heating it in a furnace at 150 °C for 3 days. Afterwards, the brucite-water suspension was cooled to room temperature and dried at 135 °C using a hotplate. Eventually, the hydrothermally-aged brucite powder was stored in a glass vial under an Argon gas atmosphere at room temperature.

5.1.2 Characterization of Synthetic Brucite

Crystalline magnesium hydroxide, named brucite, belongs to the bivalent metal hydroxides group, whose crystal structure is a layered CdI_2-type arrangement, with trigonal crystal system, hexagonal scalenohedral class, P3m1 space group. Specifically, successive hexagonal magnesium cations (Mg^{2+}) layers and hydroxyl (OH^-) ions layers are stacked one upon another, the magnesium cations are six-fold coordinated by hydroxyl groups, thus forming $Mg(OH)_6$ octahedra. Such a layered crystal structure is an advantage for platelet-shaped crystallization of the compound.
Characterization of the synthetic brucite powder has been performed through Scanning Electron Microscopy (SEM) imaging (JEOL 8500F). As shown in Figure 5.1, the untreated synthetic brucite is characterized by the presence of a mixture of rod-like morphological structures and rosettes of rough, circular and oblong platelets, while only clusters of lamellar, hexagonal platelets are revealed after the hydrothermal treatment.
Additionally, the powder X-ray Diffraction (XRD) pattern (Bruker D2 Phaser) has shown an increase in the intensity of the peaks after the hydrothermal treatment (Figure 5.2).

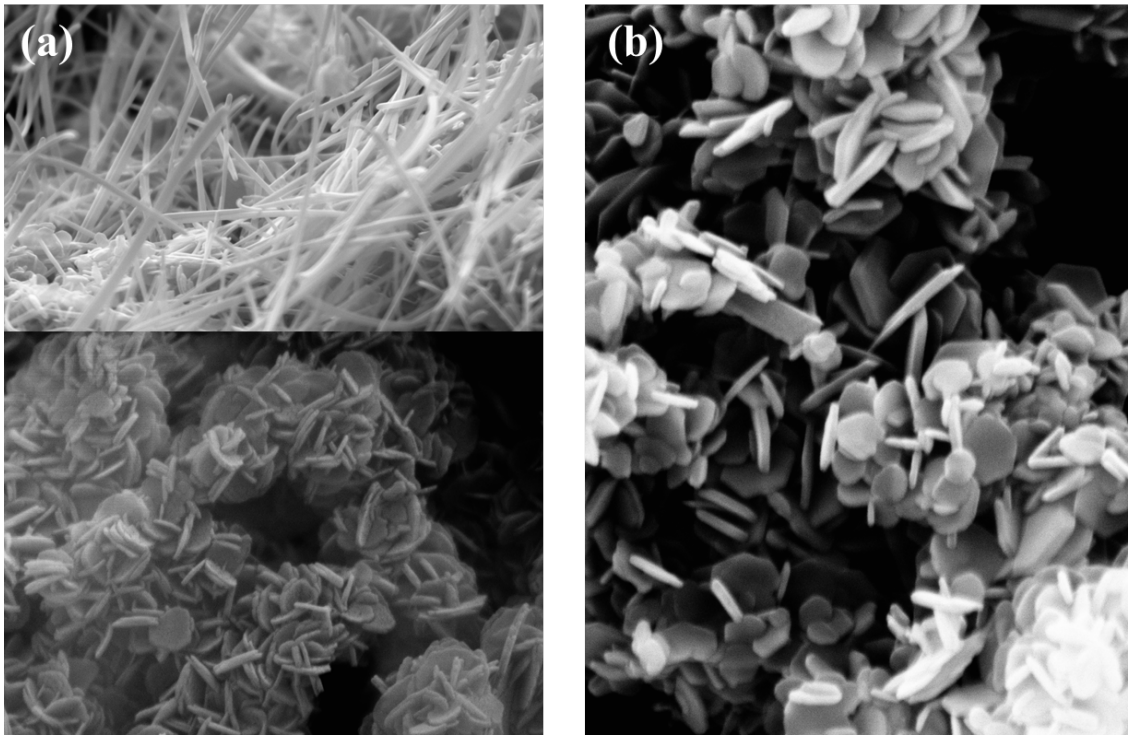

Figure 5.1: SEM image of synthetic brucite (a) before hydrothermal treatment and (b) after hydrothermal treatment at 150°C for 3 days.

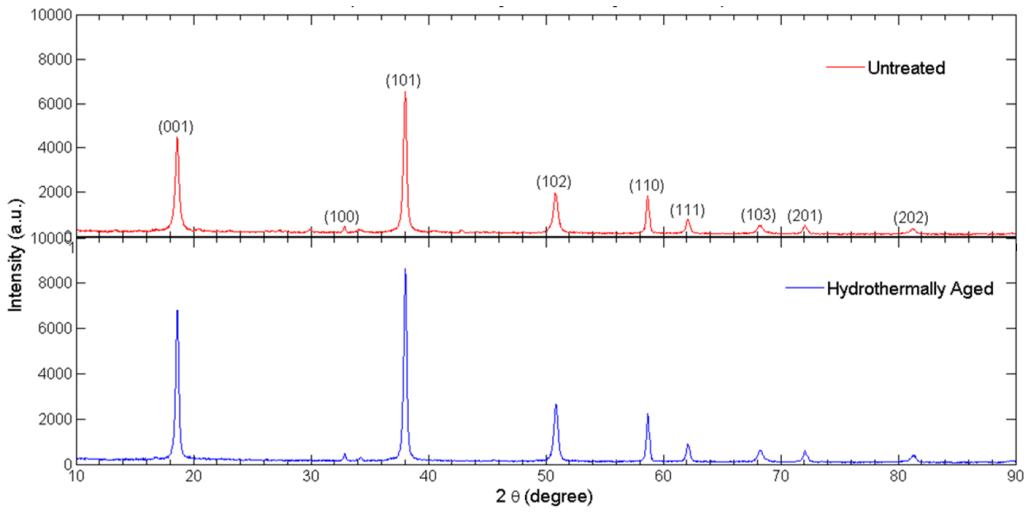

Figure 5.2: Synthetic brucite powder XRD patterns for untreated (red) and hydrothermally-treated (blue) mineral. (hkl) reflections are also indicated.

Further measurements have been carried out to characterize the lateral surfaces of the synthetic brucite. Indeed, several studies indicate that for minerals with superficial hydroxyl functional groups only lateral surfaces are active in the protonation and deprotonation reactions leading to ligand adsorption.[334,429–431] Previous studies have already been carried out on brucite samples, synthesized and hydrothermally-treated using the same experimental procedure adopted in this work, to determine the distribution between the basal and lateral surfaces of this mineral.[334] Specifically, through low-pressure Argon gas adsorption measurements,[334] it was determined that about 89.5% of the total surface area for such samples is represented by basal brucite surfaces, whereas 10.5% is represented by lateral brucite surfaces. As measured by multi-point BET N_2 adsorption (Micromeritics, Norcross, GA, USA), our synthetic brucite has a total specific surface area of 22.7 ± 0.1 m^2g^{-1}. Thus, we have considered an edge-specific surface area (ESA) of 2.4 m^2g^{-1} as the best estimate for the more reactive lateral surfaces.

Moreover, to characterize lateral surfaces of our synthetic brucite samples, transmission electron microscopy (TEM) imaging was performed using the High Resolution Analytical Electron Microbeam (HRAEM) facility at the Johns Hopkins University (courtesy of Dr. Kenneth Livi). The Selected Area Electron Diffraction (SAED) crystallographic experimental technique was used for TEM measurements. The programs CrystalMaker and SingleCrystal assisted in determining the orientation of the SAED patterns, relating that to the crystal faces.

The analysis of the TEM images, shown in Figure 5.3, allowed to derive the particle size ranges (not statistically representative), that are 90-730 nm for the longest dimension in the basal plane, and 20-90 nm for the width of the crystals along (hk0). Additionally, the analysis of face orientations revealed the presence of many vicinal faces that can be broken down into lower index faces, but the (110) face appears to be the most exposed lateral surface. Therefore, in the first approximation we have confidently considered the adsorption of nucleic acids components on the (110) face, which is characterized by all equivalent geminal surface sites.

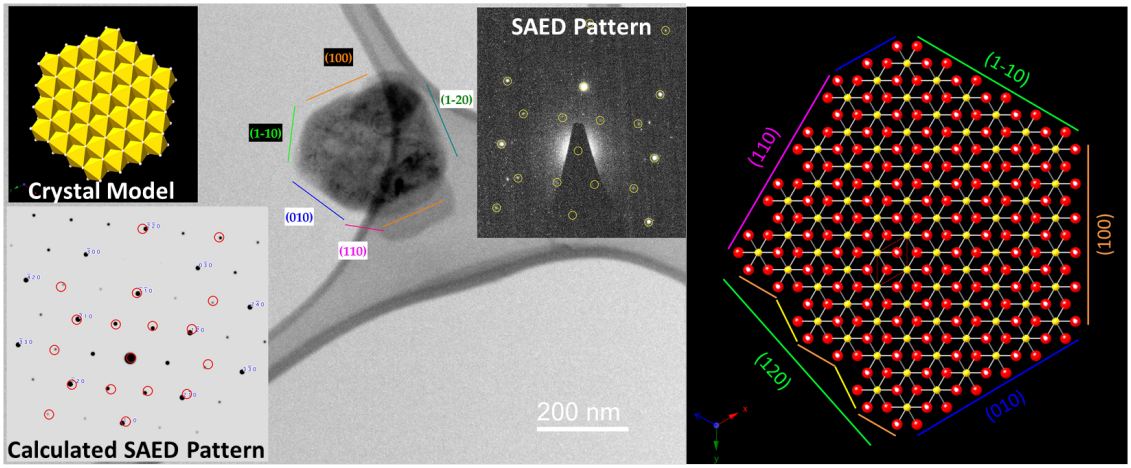

Figure 5.3: Synthetic brucite powder TEM micrograph, analysis of SAED crystallographic pattern and determination of crystal face orientations.

5.2 Thermodynamics of Adsorption of Nucleic Acids Components on Brucite

In a first step, we have studied the thermodynamics of adsorption of nucleic acids components on brucite experimentally, determining the equilibrium adsorption isotherms at room temperature. Then, a rigorous and quantitative thermodynamic characterization of the adsorption data has been provided using the computer code GEOSURF,[335] based on the extended triple-layer model (ETLM) for predicting surface speciation as a function of environmental conditions, as well as the stoichiometry and equilibrium constants.

5.2.1 Batch Adsorption Experiments

Batch adsorption experiments were performed in water (18.2 MΩ·cm Milli-Q water) for the nucleic acid components uracil (U), uridine (UD), uridine 5'-monophosphate (UMP) and adenosine 5'-monophosphate (AMP) shown in Figure 1.5. In specific cases, to better constrain the adsorption mechanism, different ionic strength conditions were considered, namely 0.01 and 0.1 M NaCl.
Uracil ≥99%, uridine ≥99%, uridine 5'-monophosphate disodium salt and adenosine 5'-monophosphate sodium salt ≥99% (Sigma-Aldrich) were used to prepare aqueous stock solutions at initial concentrations ranging from 10 μM to 300 μM.
Batch adsorption samples were prepared by mixing 50 mg of hydrothermally-aged brucite mineral with 5 mL of aqueous stock solutions in 15 mL sterile Falcon tubes (mineral concentration of 10 ± 0.5 g L^{-1}). Then, samples were homogenized by vortex and put on a test tube rotator at room temperature and pressure (25 ± 1 °C, 1 bar) for about 24 h; this time is largely sufficient to reach steady state adsorption as established by preliminary kinetics experiments. Afterwards, equilibrated suspensions were centrifuged for 10 min with a relative centrifugal force of 1073 g, to pellet suspended mineral particles and separate the supernatants.
The analytical concentration of stock solutions and supernatants was measured through ion chromatography (IC), using a Dionex ICS-5000 DP dual pump system, equipped with an AminoPac10 column and an integrated pulsed amperometry (IPAD) electrochemical detector, and Chromeleon 6.8 software (Dionex Corporation, Sunnyvale, CA, USA). In order to more accurately estimate the concentrations, the samples were analyzed in duplicate, determining the mean concentrations within ±1 standard error (σm).
The amount of nucleic acid component adsorbed in equilibrium with the mineral was calculated by subtracting the amount of molecules in the supernatant from the total amount of molecules initially added. The surface adsorption, i.e. the moles of molecules adsorbed on the surface normalized on the specific surface area of the mineral (μmol m^{-2}), was determined as:

$$\frac{n_{\text{ads}}}{A_{\text{s ESA}}} = \frac{(C_0 - C_{\text{eq}})V}{m_{\text{mineral}} A_{\text{s ESA}}} \tag{5.1}$$

where C_0 is the initial concentration of the stock solutions of nucleic acid components (in μM), C_{eq} is the concentration of nucleic acid components remaining in the liquid supernatants at equilibrium (in μM), V is the volume of stock solution initially added to brucite (in L), $m_{mineral}$ is the mass of the mineral used to prepare the samples (in g), and A_{sESA} is the edge-specific surface area (ESA) of our synthetic brucite mineral (2.4 m^2g^{-1}, as previously evaluated).
The pH of the supernatant solutions after equilibrium adsorption was measured with a combination glass electrode (Thermo-Electron Orion 8103BNUWP), calibrated with NIST-standardized buffers (Fisher Scientific). In all cases, thanks to the strong basicity of the brucite powder, the pH at equilibrium was buffered at 10.15 $\pm$ 0.07.
Finally, the pellets were dried in an oven at 50 °C and stored for further analysis.

5.2.2 Surface Complexation Modeling

The experimental batch adsorption data were modeled using the computer code GEOSURF,[335] based on a mechanistic model of equilibrium adsorption which describes the specific surface-molecule interactions and treats quantitatively the energies involved. In more detail, it is based on the extended triple-layer model (ETLM) for predicting surface speciation as a function of environmental conditions such as pH, ionic strength, ligand-to-solid ratios.[432–439]
According to the triple-layer model, there are three planes in the interfacial region (schematically represented in Figure 5.4): the mineral surface or 0-plane, where protons and hydroxide ions adsorb; the inner Helmholz or β-plane, where electrolyte ions adsorb, which is defined by the centers of specifically adsorbed anions and cations; and the outer Helmholz or d-plane, corresponding to the closest distance of approach of the diffuse swarm of counterions induced near the surface to counter the local charge density due to the adsorption of the protons and electrolyte ions, which is responsible for the formation of a net electrical charge at the surface. The three planes of charge are associated with three planes of potential and treated as a series of pairs of parallel-plate capacitors with capacitances C_1 and C_2 in Figure 5.4. A detailed description of this model is provided in Refs.[335,440] and references therein. Noteworthy, the ETLM not only accounts for specific electrolyte adsorption, i.e. the equilibrium constant for the adsorption of a particular electrolyte ion depends on the specific identity of the ion, but also specifically accounts for the electrical work associated with desorption of chemisorbed water molecules during inner-sphere surface complexation,[441] given by $\Delta\Psi$ = -n_{H_2O} (Ψ_0 - Ψ_β), where n_{H_2O} is the number of water molecules released from the surface when inner-sphere bonds are formed,

Ψ_0 is the electric potential at the 0-plane, i.e. the mineral surface where the surface sites may be protonated or deprotonated, and Ψ_β is the electric potential at the inner Helmholz or β-plane, i.e. the plane of electrolyte adsorption. As a consequence, it indicates the number of inner-sphere linkages for the adsorbate, as well as the number of surface sites involved in the reaction stoichiometry, significantly constraining the likely mode of surface attachment. This approach has been validated with respect to spectroscopic data and quantum chemical computations,[289,437,442] providing reliable predictions for the attachment of aqueous organic compounds to the surface of oxide minerals.

To apply the ETLM to our experimental adsorption data for brucite in NaCl aqueous solution, we used surface protonation and electrolyte adsorption parameters previously evaluated by Estrada *et al.* (2015)[334] from a regression of surface titration data measured by Pokrovsky and Schott (2004),[444] which are reported in Table 5.1. Regarding the brucite properties, we used a lateral surface area (A_s) of 2.4 m^2g^{-1}, as previously measured, constant capacitances (C_1 and C_2) of 190 μF cm^{-2}, site density (N_s) of 38 sites nm^{-2} and point of zero charge (pH_{PZC}) of 10.5, as estimated by Estrada *et al.* (2015)[334] for brucite samples synthesized with the same procedure used in this work. Protonation constants of the nucleobase uracil (U), the nucleotide uridine (UD), and the ribonucleotides uridine 5'-monophosphate (UMP) and adenosine 5'-monophosphate (AMP), were derived from literature.[445–448] Our calculations take into account also the complexation of sodium ions with nucleotides in solution, whose equilibrium constants were derived from Smith *et al.* (1991).[449] Table 5.1 lists all the parameters for brucite mineral, the protonation and electrolyte adsorption constants for brucite surface equilibria, the protonation constants for the aqueous species equilibria, and the surface complexation equations proposed in this study with the corresponding equilibrium constants.
In order to establish the surface reaction stoichiometries that better predict the experimental outcomes, the ETLM calculations were iteratively carried out, obtaining fits of the experimental adsorption data over a wide range of surface loading and different ionic strengths.

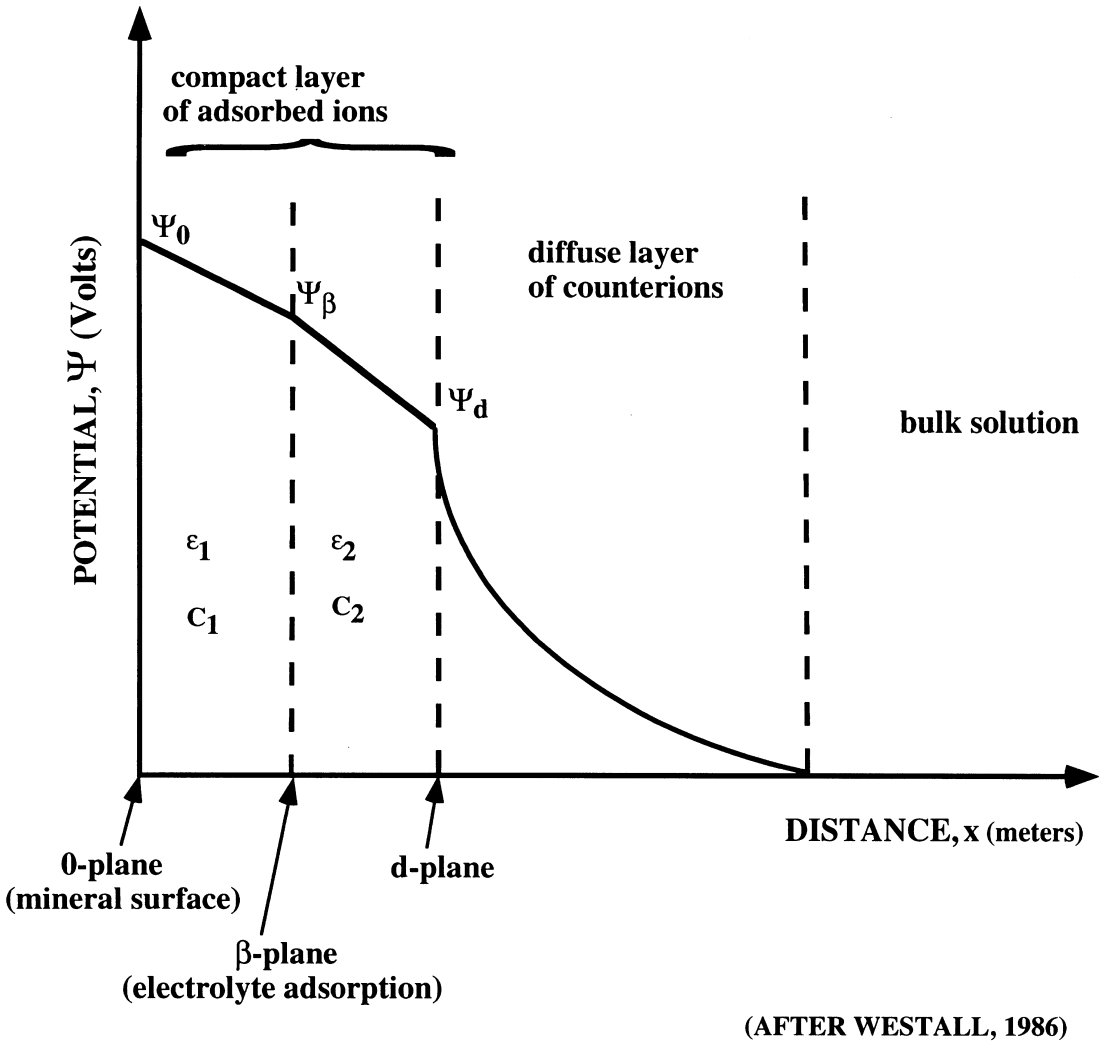

Figure 5.4: Schematic representation of potential-distance relationships assumed in triple-layer model.[443]

Table 5.1: Aqueous nucleic acid components properties, brucite characteristics and ETLM parameters for proton, electrolyte and nucleic acid components adsorption onto brucite.

Reaction type	**Reaction**	**log K**
Aqueous uracil equilibrium	$U^- + H^+ = HU$	9.5[a]
Aqueous uridine equilibrium	$UD^- + H^+ = HUD$	9.2[a]
Aqueous UMP equilibria	$UMP^{3-} + H^+ = HUMP^{2-}$	9.45[b]
	$HUMP^{2-} + H^+ = H_2UMP^-$	6.15[b]
	$H_2UMP^- + H^+ = H_3UMP$	0.7[b]
	$Na^+ + H_3UMP = NaHUMP^- + 2H^+$	-5.97[b,c]
Aqueous AMP equilibria	$AMP^{2-} + H^+ = HAMP^-$	6.21[d]
	$HAMP^- + H^+ = H_2AMP$	3.84[d]
	$H_2AMP + H^+ = H_3AMP^+$	0.4[d]
	$Na^+ + H_2AMP = NaAMP^- + 2H^+$	-9.17[c]
Surface equilibria	Hypothetical 1.0 m standard state	
log K_1^0	$>SOH + H^+ = >SOH_2^+$	9.60[e]
log K_2^0	$>SO^- + H^+ = >SOH$	11.40[e]
log $K^0_{Na^+}$	$>SOH + Na^+ = >SO^-\_Na^+ + H^+$	-9.00[e]
log $K^0_{Cl^-}$	$>SOH + Cl^- + H^+ = >SOH_2^+\_Cl^-$	11.80[e]
Surface complexes formation	$>SOH + HU = >SOH_2^+\_U^-$	2.8
	$>SOH + HUD = >SOH_2^+\_UD^-$	2.75
	$2>SOH + HUD + H^+ = (>SOH_2^+)_2\_UD^-$	16.2
	$3>SOH + H_3UMP = (>SOH_2^+)\_>S_2UMP^- + 2H_2O$	23.75
	$2>SOH + H_3UMP + Na^+ = (>SOH_2^+)_2\_NaHUMP^-$	23.45
	$3>SOH + H_2AMP + H^+ = (>SOH_2^+)\_>S_2AMP + 2H_2O$	29.4
	$>SOH + H_2AMP + Na^+ = (>SOH_2^+)\_NaAMP^- + H^+$	4.95

[a]Dawson *et al.* (1986).[447] [b]Martin (1985), Danyluk (1968).[445,446] [c]Smith (1991).[449] [d]Sigel (1994).[448] [e]Estrada *et al.* (2015).[334]

5.3 Results and Discussion

In order to study the effect of the different molecular components of nucleic acids on the surface interactions, we started our study about the binding mechanisms of nucleic acid components on the surface of brucite, comparing the adsorption of uracil with the corresponding nucleoside and nucleotide, uridine and uridine 5'-monophosphate, respectively.
Experimental adsorption isotherms for uracil (U), uridine (UD), and uridine 5'-monophosphate (UMP) in water are reported in Figure 5.5, as a function of the concentration of the nucleic acid components in the aqueous phase at equilibrium. It is clear that the surface adsorption increases as a function of the equilibrium concentration in the aqueous phase and, consistently, the percentage of adsorption decreases with increasing surface loading.
For uracil, surface adsorption at a pH of 10.0 ± 0.1 ranges from $(3\pm2)\times10^{-2}$ μmol m^{-2} to $(6.5\pm0.3)\times10^{-1}$ μmol m^{-2}, which corresponds to a percentage of adsorption with respect to the amount of nucleobase initially added of about 6.4-5.4%.
For uridine, surface adsorption at a pH of 10.1 ± 0.1 ranges from $(1.77\pm0.08)\times10^{-2}$ μmol m^{-2} to $(3.66\pm0.03)\times10^{-1}$ μmol m^{-2}, which corresponds to a percentage of adsorption with respect to the amount of nucleoside initially added of about 4.6-3.4%.
For UMP, surface adsorption at a pH of 10.2 ± 0.1 is much higher, ranging from $(3.213\pm0.006)\times10^{-1}$ μmol m^{-2} to 3.046 ± 0.005 μmol m^{-2}, which corresponds to a percentage of adsorption with respect to the amount of nucleotide initially added of about 82.2-30.5%.
As inferred from the experimental data, the similar adsorption profile of uracil and uridine indicates that the ribose moiety should not be involved in the adsorption process. The very low adsorption observed for uracil and uridine should be indicative of weak outer-sphere interactions with the mineral surface. On the other hand, the great increase in adsorption observed in the case of UMP points out the involvement of the phosphate group in the surface-molecule interactions, which is able to enhance by 10 times the surface attachment, consistently with other studies reported in the literature.[450,451]

In agreement with the previous finding, applying iteratively the ETLM to fit our experimental data for the adsorption of uracil on brucite in water, we observed that the surface reaction which better reproduces the experimental outcomes pertains to the formation of only outer-sphere interactions. Figure 5.6 shows a very good agreement between experimental data (green triangles) and modeling results (dashed black line). Specifically, we proposed the following overall reaction stoichiometry:

$$>\mathrm{SOH} + \mathrm{HU} = >\mathrm{SOH_2^+\_U^-}$$

for which the overall electrical work is given by $\Delta\Psi_r = +\Psi_0 - \Psi_\beta$, >SOH represents

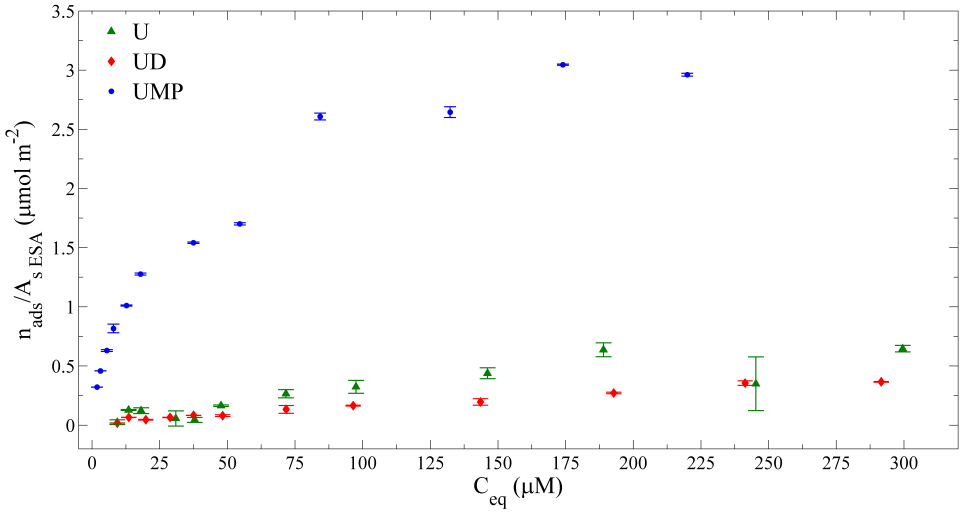

Figure 5.5: Experimental adsorption isotherms for uracil (U), uridine (UD) and uridine 5'-monophosphate (UMP) onto brucite in water at 25 ± 1 °C and 1 bar, using a mineral concentration of 10 g L^{-1}. The plotted experimental data are the average of a duplicate run and error bars are ±1 standard error from the average.

a neutral generic surface site (the symbol ">" implies that the surface site is bonded to the underlying bulk mineral), HU is the neutral uracil molecule, U^- is the deprotonated uracil molecules, and the symbol "_" indicates outer-sphere interactions. Such reaction derives from the sum of the following non-mechanistic steps, involving the protonation of the surface sites (i), the deprotonation of the uracil molecules (ii) and the complexation of anionic uracil molecules on the surface (iii):

$$>SOH + H^+ = >SOH_2^+ \text{ (i)}$$
$$HU = U^- + H^+ \text{ (ii)}$$

for which the electrical work, associated to the attachment of a proton (positive charge) to the surface (0-plane), is given by $\Delta\Psi = +\Psi_0$, and

$$>SOH_2^+ + U^- = >SOH_2^+\_U^- \text{ (iii)}$$

for which the electrical work, due to the approach of an anionic uracil molecule (negative charge) to the β-plane during an outer-sphere complexation, is given by $\Delta\Psi = -\Psi_\beta$.

Such theoretical results are consistent with the experimental conditions, since at the equilibrium solution pH≈10, which is below the point of zero charge estimated for brucite,[334] the hydroxyl groups on the surface of brucite should be in the acidic form, determining a positively-charged surface, while uracil molecules are expected to undergo deprotonation at the N3 site,[447] i.e. molecules are in the anionic form.

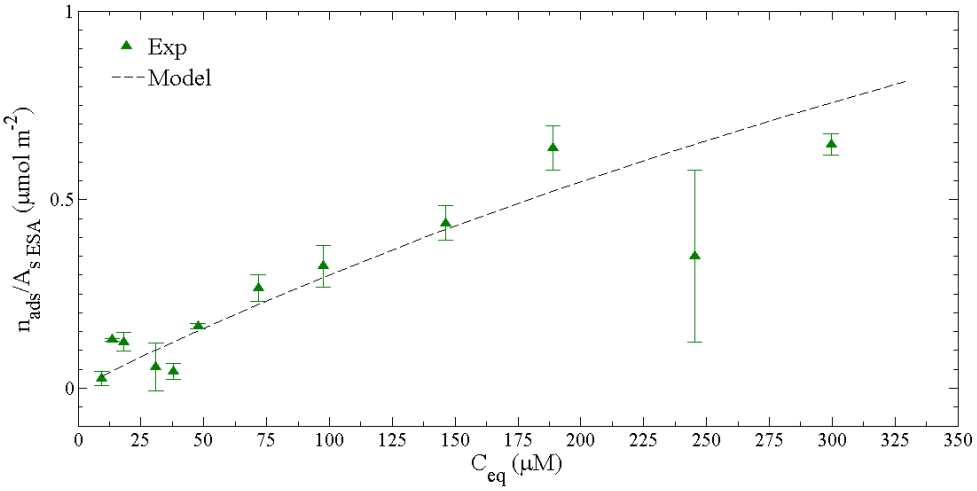

Figure 5.6: Adsorption of uracil onto brucite as function of equilibrium concentration of uracil in the aqueous phase. The green triangles represent experimental data that are the average of a duplicate run and error bars are ±1 standard error from the average. The dashed black curve represents the modeled fit to the experimental data, calculated with the parameters reported in Table 5.1, according to the reaction: >SOH + HU = >SOH_2^+_U^-.

Therefore, the overall complexation reaction can be interpreted as the formation of monodentate outer-sphere surface complexes on positively-charged surface sites, i.e. the anionic molecules do not covalently bind to the surface of the mineral but they approach the positively-charged surface sites through hydrogen-bonding interactions, involving mainly the negative charge located on the C4O group of the uracil molecule. In this kind of approach to the β-plane, the outer-sphere complexes do not keep their hydration shell, but they give specific (hydrogen-bond) interactions, corresponding to a chemisorption process.
A possible representation of the uracil-brucite complex predicted by surface complexation calculations is shown in Figure 5.7.

In the case of uridine on brucite, applying the ETLM using the parameters reported in Table 5.1, the best fit to our experimental data is given by two different models. These are displayed as dashed and solid black curves in Figure 5.8. One of the reaction stoichiometry (dashed black curve in Figure 5.8) is:

$$>SOH + HUD = >SOH_2^+\_UD^-$$

for which the overall electrical work is given by $\Delta\Psi_r = +\Psi_0 - \Psi_\beta$.
In analogy with to the case of uracil, such reaction derives from the sum of non-mechanistic steps, involving the protonation of the surface sites (i), the deprotonation of the uridine molecules (ii) and the complexation of anionic uridine molecules

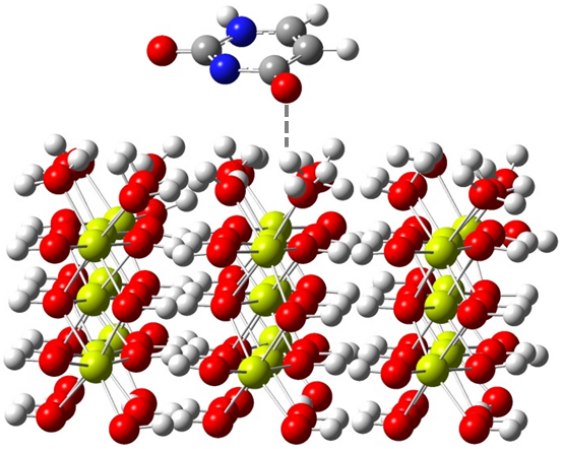

Figure 5.7: Surface species proposed for the adsorption of uracil onto (110) brucite crystal face, corresponding to a monodentate outer-sphere surface complex of anionic uracil on protonated surface sites.

on the surface (iii):

$$>\mathrm{SOH} + \mathrm{H^+} = >\mathrm{SOH_2^+} \text{ (i)}$$
$$\mathrm{HUD} = \mathrm{UD^-} + \mathrm{H^+} \text{ (ii)}$$

for which the electrical work, associated to the attachment of a proton to the surface (0-plane), is given by $\Delta\Psi = +\Psi_0$, and

$$>\mathrm{SOH_2^+} + \mathrm{UD^-} = >\mathrm{SOH_2^+\_UD^-} \text{ (iii)}$$

for which the electrical work, due to the approach of an anionic uridine molecule to the β-plane during an outer-sphere complexation, is given by $\Delta\Psi = -\Psi_\beta$.
The model representing such a reaction is a monodentate outer-sphere surface complex (depicted in panel **a** of Figure 5.9), which is plausibly assumed to interact with the protonated surface sites through the C4O group of the uracil moiety that exhibits a negative charge as a consequence of the deprotonation at the N3 site above pH 9.2.[447]
The other reaction (solid black curve in Figure 5.8) involves two surface sites:

$$2>\mathrm{SOH} + \mathrm{HUD} + \mathrm{H^+} = (>\mathrm{SOH_2^+})_2\_\mathrm{UD^-}$$

for which the overall electrical work is given by $\Delta\Psi_r = +2\Psi_0 - \Psi_\beta$.
This involves the following reaction steps:

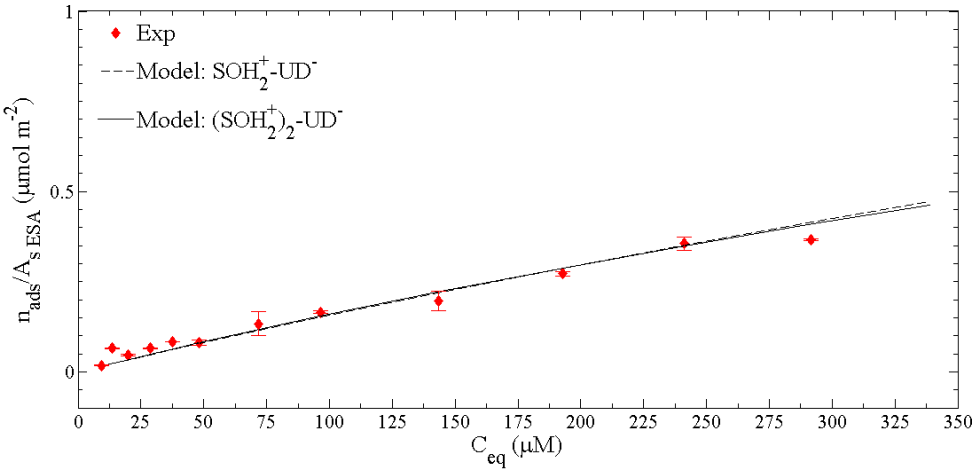

Figure 5.8: Adsorption of uridine onto brucite as function of equilibrium concentration of uridine in the aqueous phase. The red diamonds represent experimental data that are the average of a duplicate run and error bars are ±1 standard error from the average. The dashed and solid black curves represent the modeled fits to the experimental data, calculated with the parameters reported in Table 5.1, according to the reactions: >SOH + HUD = $>SOH_2^+$_UD^-, and 2>SOH + HUD + H^+ = $(>SOH_2^+)_2$_UD^-, respectively.

$$2{>}SOH + 2H^+ = 2{>}SOH_2^+ \text{ (i)}$$
$$HUD = UD^- + H^+ \text{ (ii)}$$

for which the electrical work, associated to the attachment of two protons to the surface (0-plane), is given by $\Delta\Psi = +2\Psi_0$, and

$$2{>}SOH_2^+ + UD^- = (>SOH_2^+)_2\text{_}UD^- \text{ (iii)}$$

for which the electrical work, due to the approach of an anionic uridine molecule to the β-plane during an outer-sphere complexation, is given by $\Delta\Psi$ = -Ψ_β.
This is interpreted as a bidentate outer-sphere surface complex on positively-charged surface sites, whose most probable interactions concern the C4O^- and C2=O functional groups of the uracil moiety, as shown in panel **b** of Figure 5.9.
The comparable adsorption of uridine and uracil suggests that the ribose moiety does not play any significant role in enhancing adsorption. Thus, interactions through the hydroxyl groups in 5' or 2' positions of the sugar moiety is considered less probable. Ongoing work concerning quantum mechanical calculations on the energetics of these complexes will allow to clarify which is the preferential adsorption configuration.

As inferred from Figure 5.5, the adsorption of UMP on brucite is much more favored. The great affinity of UMP for the surface of this mineral may be explained considering that, at the equilibrium solution pH≈10, these molecules feature three

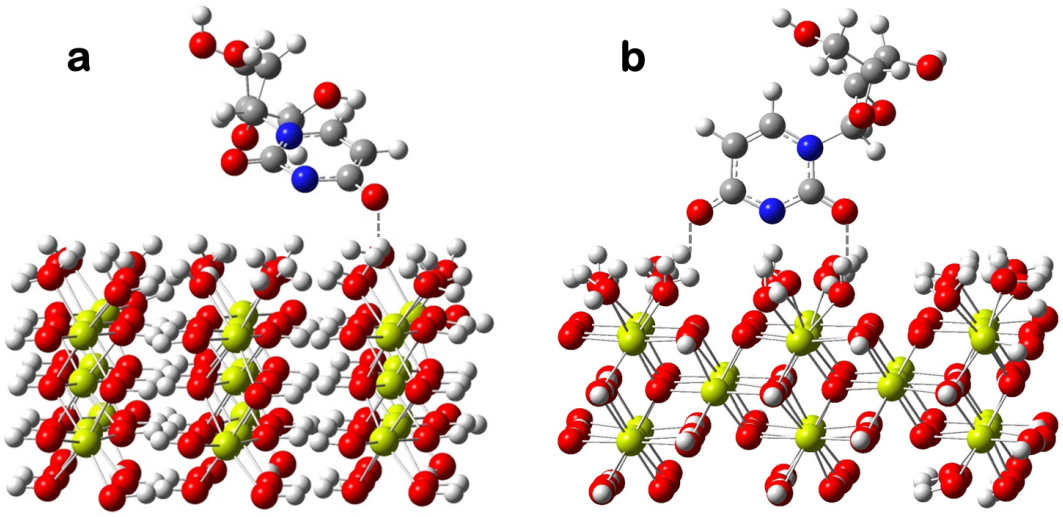

Figure 5.9: Surface species proposed for the adsorption of uridine onto (110) brucite crystal face, corresponding to a monodentate outer-sphere surface complex interacting with a protonated surface site via the C4O$^-$ group of the uracil moiety (panel **a**) and a bidentate outer-sphere surface complex interacting with positively-charged surface sites via C4O$^-$ and C2=O functional groups of the uracil moiety (panel **b**).

negative charges as a consequence of deprotonation of the phosphate group and the N3 site of the uracil moiety,[445,446] which can be strongly attracted by the positively-charged surface. This leads to an adsorption trend typical of Langmuir/high affinity isotherms, up to equilibrium concentrations of UMP in the aqueous phase of about 60 μM, where an inflection point denotes the coverage for a complete monolayer and, then, the adsorption increases again and reaches a new saturation. In fact, comparison between the total exposed edge-specific surface area of the mineral used in our samples with the total area occupied by the estimated total amount of molecules adsorbed on those lateral surfaces around the first and the second plateau of the isotherm, confirms that just one monolayer is formed at the first plateau with a close packing of molecules inclined on the surface, and two monolayers at the second one. It is plausible that at high concentrations of UMP molecules, stacking interactions through the base moieties occur between free UMP molecules and UMP molecules already adsorbed, giving rise to multilayers. This is a fascinating hypothesis in the prebiotic context which requires further investigations in order to identify the possible macromolecular structures which form at high concentrations.

In order to better constrain the adsorption mechanism for UMP on brucite, we performed the batch adsorption experiments in two different ionic strength conditions: water and 0.1 M NaCl. As we can see in Figure 5.10, the presence of sodium ions, Na^+, significantly suppresses the adsorption of UMP, which is indicative of a

possible involvement of Na^+ in the adsorption process. In fact, as widely observed in the literature, sodium ions have a great ability to trap oxygen atoms, resulting in strong complexation with phosphates and/or coordination of water molecules, inducing the formation of complex structures in the aqueous phase which can inhibit, for instance, processes like duplication of DNA, ribosomal peptide synthesis and so on.[452,453] This explains the very low concentration of Na^+ in the cell cytoplasm of all "modern" living cells, which is approximately 0.01 mol L^{-1}, and suggests that the first protocells could not have emerged in NaCl solutions.[452]
In particular, to fit our experimental data of surface adsorption of UMP on brucite in these two different ionic strength conditions, it was necessary to combine two surface reactions, with one involving the adsorption of a Na^+-UMP complex. The first reaction is:

$$3{>}SOH + H_3UMP = (>SOH_2^+)\_{>}S_2UMP^- + 2H_2O$$

for which the overall electrical work is given by $\Delta\Psi_r = +\Psi_0 - \Psi_\beta$.
This involves the following reaction steps:

$$3{>}SOH + 3H^+ = 3{>}SOH_2^+ \text{ (i)}$$
$$H_3UMP = UMP^{3-} + 3H^+ \text{ (ii)}$$

for which the electrical work, associated to the attachment of three protons to the surface (0-plane), is given by $\Delta\Psi = +3\Psi_0$, and

$$3{>}SOH_2^+ + UMP^{3-} = (>SOH_2^+)\_{>}S_2UMP^- + 2H_2O \text{ (iii)}$$

for which the overall electrical work is given by $\Delta\Psi = -\Psi_\beta - 2\Psi_0$, due to the sum of the electrical work for the adsorption of UMP^{3-} on the β-plane during an inner-sphere complexation involving two sites ($-3\Psi_\beta$) and the electrical work associated with the desorption of two chemisorbed water molecules ($-2(\Psi_0-\Psi_\beta)$).
In this case three surface sites are involved in the adsorption process. The molecule employs two points of attachment to form inner-sphere linkages and a third point of attachment for an outer-sphere interaction.
The second reaction includes the complexation of Na^+ with UMP and can be represented as follows:

$$2{>}SOH + H_3UMP + Na^+ = (>SOH_2^+)_2\_NaHUMP^-$$

for which the overall electrical work is given by $\Delta\Psi_r = +2\Psi_0 - \Psi_\beta$.
This involves the following reaction steps:

$$2{>}SOH + 2H^+ = 2{>}SOH_2^+ \text{ (i)}$$
$$H_3UMP + Na^+ = NaHUMP^- + 2H^+ \text{ (ii)}$$

for which the electrical work, associated to the attachment of two protons to the surface (0-plane), is given by $\Delta\Psi = +2\Psi_0$. The reaction (ii) corresponds to the complexation of sodium ion with UMP. In this reaction Na^+ would mainly interact with the negatively-charged phosphate group of the UMP molecule which preferentially undergoes deprotonation.
Then, the third step corresponds to the adsorption of such a $NaHUMP^-$ complex onto the surface:

$$2{>}SOH_2^+ + NaHUMP^- = (>SOH_2^+)_2\text{-}NaHUMP^- \text{ (iii)}$$

for which the electrical work due to the approach of an anionic $NaHUMP^-$ complex to the β-plane during an outer-sphere complexation, is given by $\Delta\Psi = -\Psi_\beta$.
The speciation predicted combining these two surface reactions using the log K reported in Table 5.1, indicates that the tridentate surface complex dominates in water, while such bidentate outer-sphere surface complex with Na^+ is predominant in 0.1 M NaCl. The dashed black curves in Figure 5.10 represent the modeling results obtained combining these two surface reactions.
Assuming that both complexes interact with the surface through the phosphate group, we interpreted the first reaction as the formation of two covalent bonds between the oxygens of the phosphate group and two surface sites, with concomitant release of two water molecules, and one hydrogen bond very likely involving the $C4O^-$ group of the uracil moiety which features a negative charge. A possible representation of such a tridentate complex is shown in Figure 5.11. Other configurations involving hydrogen bonds via the hydroxyl groups of the ribose moiety should not be significant.
The second reaction can represent the formation of a bidentate outer-sphere surface complex of $NaHUMP^-$, where the complexation with Na^+ hampers the formation of covalent bonds with the surface sites, but the phosphate group should be involved in the adsorption process as well through weaker interactions. Indeed, despite the suppression in adsorption observed in the presence of sodium ions due to the complexation with the phosphate group of UMP molecules, the adsorption of UMP in 0.1 M NaCl is much higher than in the case of uridine in water. This indicates that, on the one hand, the ribose moiety should not be involved in the adsorption process and, on the other hand, the phosphate still plays a fundamental role in the attachment.

Then, we investigated the binding mechanism of another ribonucleotide, namely adenosine 5'-monophosphate (AMP). In this case, to constrain the adsorption mechanism, we performed the batch adsorption experiments in three different ionic strength conditions: water, 0.01 M and 0.1 M NaCl. Experimental adsorption isotherms are reported in Figure 5.12, as a function of the concentration of AMP in the aqueous

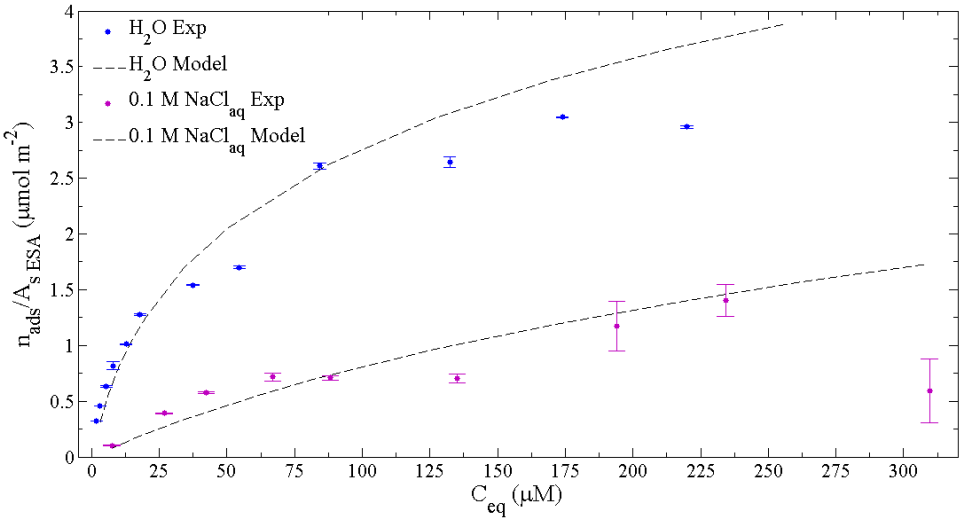

Figure 5.10: Adsorption of UMP onto brucite as function of equilibrium concentration of UMP in the aqueous phase, in two different ionic strength conditions: H_2O and 0.1 M NaCl. The blue and pink dots represent experimental data in H_2O and 0.1 M NaCl, respectively, that are the average of a duplicate run and error bars are ± 1 standard error from the average. The dashed black curves represent the modeled fits to the experimental data, calculated with the parameters reported in Table 5.1, combining two reactions: $3{>}SOH + H_3UMP = (>SOH_2^+)\_>S_2UMP^- + 2H_2O$, and $2{>}SOH + H_3UMP + Na^+ = (>SOH_2^+)_2\_NaHUMP^-$.

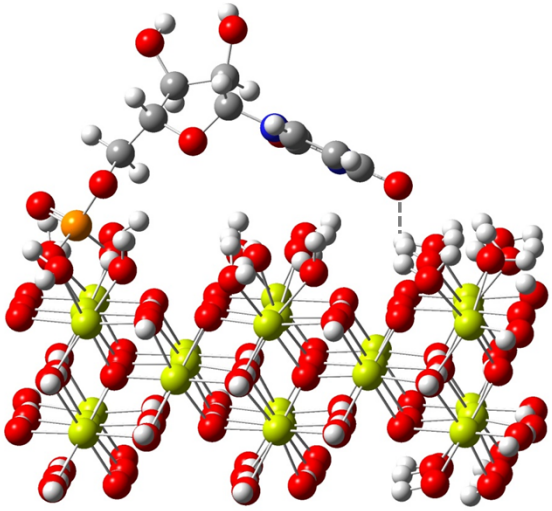

Figure 5.11: Surface species proposed for the adsorption of UMP onto (110) brucite crystal face, corresponding to a tridentate surface complex on protonated surface sites, with two inner-sphere linkage via the phosphate group and one hydrogen bond involving the $C4O^-$ group of the uracil moiety.

phase at equilibrium. For AMP, similarly to UMP, high surface adsorption was observed in water at a pH of 10.1 ± 0.1, ranging from (1.4±0.2)×10^{-1} μmol m^{-2} to 2.9±0.1 μmol m^{-2}, which corresponds to a percentage of adsorption with respect to the amount of nucleotide initially added of about 36.4-25.8%. Lower adsorptions with percentages of about 23.2-22.2% and 22.0-18.8% were observed in 0.01 M and 0.1 M NaCl at pHs of 10.2 ± 0.1 and 10.3 ± 0.1, respectively.
As shown in Figure 5.12, at low concentrations AMP adsorption data resemble S-class isotherms, according to the Giles classification,[454] which is a typical trend of unfavorable cooperative adsorption, where the molecules already adsorbed promote the adsorption process. This behavior also highlights the importance of intermolecular interactions in the adsorption mechanism. The second part of the curves, then, presents an inflection point reached at increasing concentrations as the ionic strength increases, thereafter giving a plateau.
Moreover, the effect of ionic strength on the adsorption of AMP on brucite is clearly evident from Figure 5.12, where increasing the concentration of sodium ions, a remarkable decrease of adsorption occurs. As already observed for UMP, the suppression in adsorption occurring in NaCl aqueous solutions may be due to the complexation of AMP molecules with Na^+, which occupies the possible points of attachment reducing the adsorption capabilities of the molecules.

To compare the adsorption mechanisms of the two ribonucleotides, UMP and AMP, we applied a normalization method where adsorption data are normalized with respect to a reference phase. Specifically, saturated solutions of these nucleotides were used as reference phase, and the adsorption data for UMP and AMP on brucite in water and 0.1 M NaCl were plotted in Figure 5.13 against the ratio between the equilibrium concentration of the nucleotide in the aqueous phase and its solubility.[451] As shown in Figure 5.13, the normalized adsorption isotherms in water (panel **a**) and in 0.1 M NaCl (panel **b**) do not superimpose, indicating that a different adsorption mechanism should occur.
Analysis of the data with the ETLM suggests that the best model for predicting adsorption of AMP on brucite results from the combination of two surface reactions, with one involving the adsorption of a Na^+-AMP complex. Specifically, the first reaction is:

$$3{>}SOH + H_2AMP + H^+ = (>SOH_2^+)_{-}{>}S_2AMP + 2H_2O$$

for which the overall electrical work is given by $\Delta\Psi_r = +\Psi_0$.
This involves the following reaction steps:

$$3{>}SOH + 3H^+ = 3{>}SOH_2^+ \text{ (i)}$$
$$H_2AMP = AMP^{2-} + 2H^+ \text{ (ii)}$$

for which the electrical work, associated to the attachment of three protons to the

surface (0-plane), is given by $\Delta\Psi = +3\Psi_0$, and

$$3{>}SOH_2^+ + AMP^{2-} = (>SOH_2^+)\_{>}S_2AMP + 2H_2O \text{ (iii)}$$

for which the overall electrical work is given by $\Delta\Psi$ = -$2\Psi_0$, due to the sum of the electrical work for the adsorption of AMP^{2-} on the β-plane during an inner-sphere complexation involving two sites (-$2\Psi_\beta$) and the electrical work associated with the desorption of two chemisorbed water molecules (-$2(\Psi_0$-$\Psi_\beta)$).
In this case three surface sites are involved in the adsorption process. The molecule employs two points of attachment to form inner-sphere linkages and a third point of attachment for an outer-sphere interaction with a protonated surface site.
The second reaction includes the complexation of Na^+ with AMP and can be represented as follows:

$$>SOH + H_2AMP + Na^+ = (>SOH_2^+)\_NaAMP^- + H^+$$

for which the overall electrical work is given by $\Delta\Psi_r = +\Psi_0$ -Ψ_β.
This involves the following reaction steps:

$$>SOH + H^+ = >SOH_2^+ \text{ (i)}$$
$$H_2AMP + Na^+ = NaAMP^- + 2H^+ \text{ (ii)}$$

for which the electrical work, associated to the attachment of a protons to the surface (0-plane), is given by $\Delta\Psi = +\Psi_0$. The reaction (ii) corresponds to the complexation of sodium ion with AMP. In this reaction Na^+ mainly interacts with the phosphate group of the AMP molecule which undergoes deprotonation.
Then, the third step corresponds to the adsorption of such a $NaAMP^-$ complex onto the positively-charged surface of brucite:

$$>SOH_2^+ + NaAMP^- = (>SOH_2^+)\_NaAMP^- \text{ (iii)}$$

for which the electrical work due to the approach of an anionic $NaAMP^-$ complex to the β-plane during an outer-sphere complexation, is given by $\Delta\Psi$ = -Ψ_β.
The speciation predicted combining these two surface reactions using the log K reported in Table 5.1, indicates that the tridentate surface complex dominates in water, while the monodentate outer-sphere surface complex with Na^+ is more and more significant as the concentration of Na^+ increases. The dashed black curves in Figure 5.12 represent the modeling results obtained combining these two surface reactions.
We interpreted the first reaction as the formation of two covalent bonds between the oxygens of the phosphate group and two surface sites, with concomitant release

of two water molecules, and one hydrogen bond very likely involving the nucleophile amino group of the adenine moiety, which can be strongly attracted by the positively-charged surface of brucite at the equilibrium solution pH≈10. A possible representation of such a tridentate complex is shown in Figure 5.14. Other configurations involving hydrogen bonds via the hydroxyl groups of the ribose moiety are less plausible. The second reaction represents the formation of a monodentate outer-sphere surface complex of $NaAMP^{-}$ interacting with the protonated surface via the phosphate group which holds a negative charge. Both complexes are different from those observed for UMP on brucite. The substantial dissimilarity for the tridentate complexes of UMP and AMP can be traced back to the charge of the complex: the tridentate complex with UMP presents a negative charge on the C4O group of the base, while the one with AMP is neutral. This could explain why UMP has a great affinity for the surface of brucite, giving a Langmuir/high affinity-type isotherm, while AMP shows an unfavorable cooperative adsorption with a S-shape trend. The cooperativity in the case of adsorption of AMP on brucite can easily occur through intermolecular interactions between neutral AMP molecules adsorbed on the surface of brucite and AMP^{2-} molecules in solution, giving rise to AMP-AMP^{2-}-type pairs. The presence of a negative charge for the complex of UMP with brucite, instead, should prevent a cooperative self-association with other UMP^{3-} molecules in the solution due to charge repulsion. Additionally, it is well known that the adenine residues have the greatest self-stacking tendency and, in general, purines associate much better than pyrimidines due to stronger stacking forces. Especially for adenine derivatives, "head-to-tail" stacking has been observed, with the five-membered and six-membered rings alternating in the stack.[455] Moreover, coordination of metal ions reduces the repulsive effect of the negatively-charged phosphate groups on the self-association and facilitates self-stacking of nucleotides.[455]

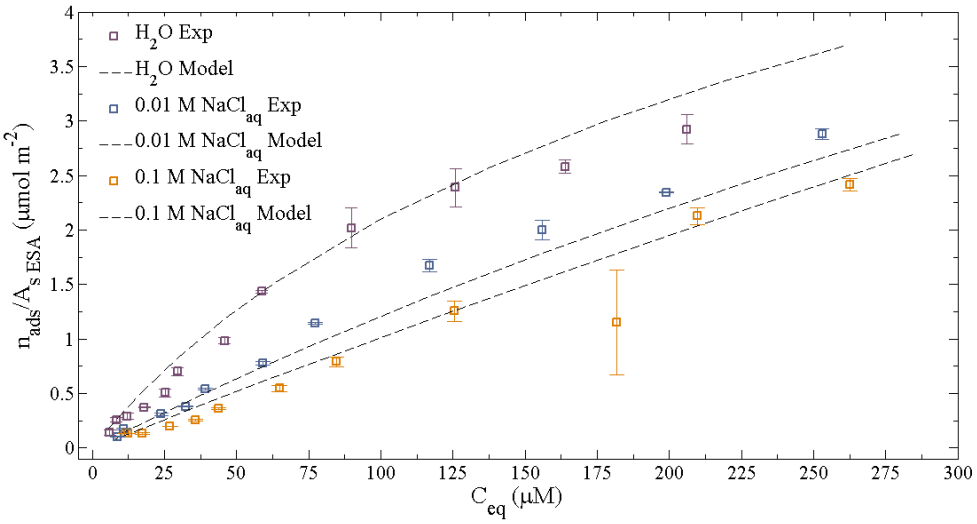

Figure 5.12: Adsorption of AMP onto brucite as function of equilibrium concentration of AMP in the aqueous phase, in three different ionic strength conditions: H_2O, 0.01 M and 0.1 M NaCl. The violet, blue and orange squares represent experimental data in H_2O, 0.01 M and 0.1 M NaCl, respectively, that are the average of a duplicate run and error bars are ±1 standard error from the average. The dashed black curves represent the modeled fits to the experimental data, calculated with the parameters reported in Table 5.1, combining two reactions: 3>SOH + H_2AMP + H^+ = (>SOH_2^+)_>S_2AMP + 2H_2O, and >SOH + H_2AMP + Na^+ = (>SOH_2^+)_NaAMP$^-$ + H^+.

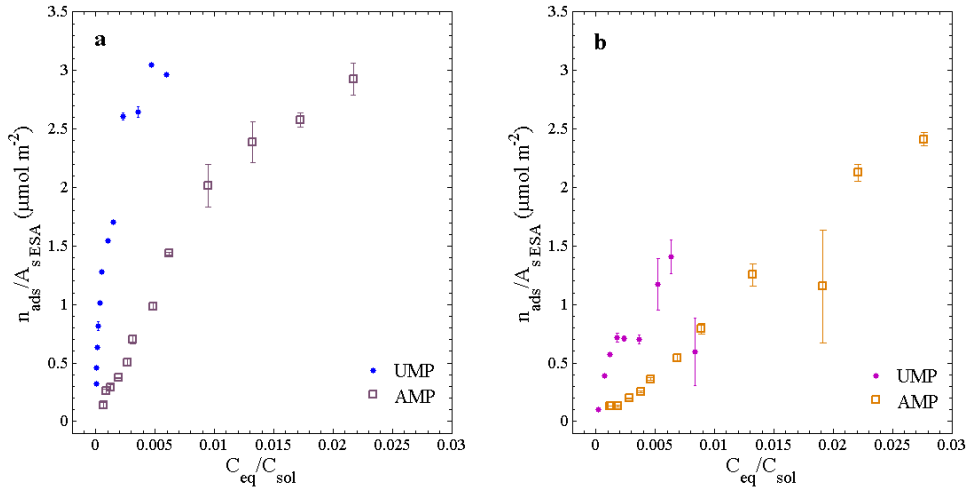

Figure 5.13: Adsorption isotherms of UMP and AMP onto brucite normalized to saturated solutions of the nucleotides, in H_2O (panel **a**) and in 0.1 M NaCl (panel **b**). The nucleotide solubilities used as reference (C_{sol}) were calculated from the Virtual Computational Chemistry Laboratory ALOGPS.[451,456,457]

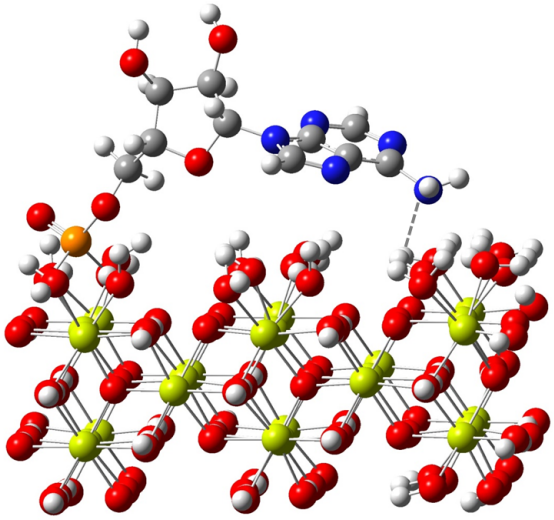

Figure 5.14: Surface species proposed for the adsorption of AMP onto (110) brucite crystal face, corresponding to a tridentate surface complex on protonated surface sites, with two inner-sphere linkage via the phosphate group and one hydrogen bond involving the amino group of the adenine moiety.

5.4 Concluding Remarks and Further Investigations

The results on the adsorption of nucleic acid components on brucite indicate that this serpentinite-hosted hydrothermal mineral is a good adsorbent for organic compounds that are "building blocks of life" and, in particular, selectively adsorbs nucleic acid components from dilute aqueous environments, suggesting a role in concentrating biomolecules in prebiotic conditions, potentially protecting such molecules against hydrolysis in dilute solutions and potentially favoring further prebiotic synthesis even from low yielding reactions.
In more details, brucite favors the adsorption of nucleotides respect to nucleosides and nucleobases, thanks to the strong interactions occurring via the phosphate group. Indeed, according to our interpretation of the surface complexation modeling results, inner-sphere complexation occurs only in the presence of phosphates functional groups, which are able to covalently bind to the hydroxyl groups on the surface of brucite leading to desorption of water molecules. More labile interactions take place for nucleosides and nucleobases thanks to the electrostatic attraction between the anionic adsorbates and the positively-charged substrate.
Specifically, the chemisorption of both ribonucleotides, UMP and AMP, occurs through the formation of tridentate surface complexes with two inner-sphere linkages via the phosphate group and one hydrogen bond, but different adsorption mechanisms are involved. The adsorption of UMP on brucite resembles a Langmuir/high affinity-type isotherm, while AMP presents a S-shape unfavorable cooperative trend. The high affinity of UMP for brucite is easily explained considering that, at the equilibrium pH$\approx$10, UMP molecules feature three negative charges while the surface of brucite is positively-charged. The tridentate complexes formed by UMP present a negative charge on the C4O group of the base, which hampers subsequent cooperative self-association. The tridentate complexes formed by AMP, instead, are neutral favoring self-association with other AMP^{2-} molecules in solution, giving rise to AMP-AMP^{2-}-type pairs.
For uridine, chemisorption occurs through hydrogen-bonds with the surface of brucite, but it is not yet clear whether a mono- or a bi-dentate surface complex forms.
Then, uracil gives monodentate outer-sphere surface complexes where the negatively-charged C4O group very likely interacts with the positively-charged surface sites.

The surface complexation modeling has provided reasonable inferences for the possible surface complexes, discriminating between inner-sphere and outer-sphere interactions, identifying the number of inner-sphere linkages for the adsorbate, as well as the number of surface sites involved in the reaction stoichiometry. However, further quantum mechanical simulations on the energetics of the possible complexes are ongoing, in order to confirm our predictions and establish more confidently which are the chemical groups involved in the molecule-mineral interactions, acting as points of attachment on the surface.

Such investigations on the capability of mineral surfaces to induce well-defined orientations of the molecules through specific molecule-mineral interactions are of primary importance to gain deeper insights into the possible role played by minerals in assisting prebiotic self-organization, possibly promoting chemical reactions towards increasing molecular complexity, by properly arranging the reactive chemical groups of neighboring molecules adsorbed on the surface.
In this regard, our experimental findings of AMP and UMP adsorbed on brucite reveal the formation of supramolecular aggregates at high concentrations, whose identification requires further analyses, such as molecular dynamics simulations and extraction experiments, in order to evaluate their potential relevance in the emergence of complex biochemistry and the geochemical origin of life.

Chapter 6

Conclusions

In the Astrobiology context, the study of the physico-chemical interactions involving "building blocks of life" in plausible prebiotic and space-like conditions is fundamental to shed light on the processes that led to the emergence of life on Earth as well as the chemical evolution in space.
In this PhD Thesis, vibrational spectroscopy has been used as an effective tool to investigate the intermolecular interactions playing a key role in the self-assembling mechanisms of nucleic acid components such as nucleobases and their binding to mineral surfaces. Such issues have been addressed through both computational and experimental studies.
From a theoretical point of view, comprehensive yet feasible computational protocols have been developed for the characterization of the spectroscopic properties of such complex systems, progressing step by step in the study of the variety of intermolecular interactions, which is typical of these systems, in order to dissect their contributions to the overall spectroscopic signals and shed light on the intricate experimental data.
First of all, the spectroscopic properties of the isolated nucleobases have been investigated, identifying a general, reliable and effective computational protocol based on fully anharmonic quantum mechanical computations of the vibrational wavenumbers and IR intensities within the generalized second order vibrational perturbation theory (GVPT2) approach, combined with the cost-effective dispersion-corrected density functional B3LYP-D3, in conjunction with basis sets of double-ζ quality such as N07D and SNSD.
Then, the dimers of nucleobases have been considered in order to study the perturbation on the vibrational frequencies and infrared intensities induced by the intermolecular interactions.
Interestingly, vibrational spectroscopy is diagnostic for the characterization of hydrogen-bonded systems, since remarkable changes of the spectroscopic features are observed for functional groups involved in hydrogen bonds. At the same time, the computational simulation of strongly anharmonic vibrations within hydrogen-bonded bridges is a big challenge due to both possible limitations of the VPT2 treatment and the

requirement of a very accurate description of the underlying PES. In this work, improved results for such vibrations have been achieved extending the previously validated computational protocol toward hybrid models, where the harmonic part of the force-field is computed at a higher level of theory like B2PLYP or by application of the less demanding ONIOM B2PLYP:B3LYP scheme, which is a focused model where only the part of the molecular system forming the hydrogen bonds is treated at B2PLYP level of theory. Moreover, for improving the vibrational frequencies of modes like the stretching of C=O and N-H functional groups, which are particularly sensitive to hydrogen-bonding, correction parameters for the B3LYP-D3/N07D frequencies of such modes have been determined.
After monomers and dimers, the treatment of the spectroscopic properties of nucleobases in condensed phases has been faced. Specifically, a heptamer cluster of uracil molecules has been considered as model to represent the properties in the solid state and the relative vibrational frequencies have been computed at anharmonic level within the VPT2 framework, combining two cost-effective approaches, namely the hybrid B3LYP-D3/N07D:DFTBA model, where the harmonic frequencies are computed with B3LYP-D3/N07D method and the anharmonic corrections are evaluated with the less expensive DFTBA method, and the reduced dimensionality VPT2 (RD-VPT2) approach, in which only selected vibrational modes are calculated anharmonically (including the couplings with the other modes) while the remaining modes are treated at the harmonic level, using the B3LYP-D3/N07D method only. The reliability of such theoretical results has been validated with respect to experiments, by performing infrared measurements of uracil in the solid state through the Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) technique. The good performance in predicting the significant experimental shifts of the vibrational frequencies of uracil due to the intermolecular hydrogen bonds in the solid state with respect to uracil isolated in Argon matrix, has allowed also to suggest some new assignments of the experimental spectrum of uracil in the solid state.
Such a protocol has paved the route for further studies on biomolecules adsorbed on minerals.
Finally, the problem of molecule-mineral interactions has been addressed, investigating experimentally the thermodynamics of the adsorption process of nucleic acid components on brucite, a serpentinite-hosted hydrothermal mineral, through determination of the equilibrium adsorption isotherms. Additionally, surface complexation studies have been carried out to get the stoichiometry of surface reactions and the associated electrical work. Such surface complexation modeling has provided reasonable inferences for the possible surface complexes, determining the number of inner/outer-sphere linkages for the adsorbates and the number of surface sites involved in the reaction stoichiometry. However, to distinguish the specific functional groups which constitute the points of attachment to the surface, further quantum mechanical simulations on the energetics of these complexes and spectroscopic characterizations are required.
Therefore, future investigations will concern mainly quantum mechanical calcula-

tions on the energetics of the possible complexes that nucleic acid components form on the surface of brucite, in order to better determine the adsorption geometry, providing insights into the adsorption mechanism. Then, the computational protocols developed in this work will be used for the spectroscopic characterization of the most stable complexes.
Moreover, further studies will deal with the characterization of the possible macromolecular structures assembling at high concentrations in the case of adsorption of uridine 5'-monophosphate and adenosine 5'-monophosphate onto brucite, which may be interesting in the origin of life research for exploring possible pathways towards the emergence of molecular complexity.

Appendix A

Theoretical background

As mentioned in the introduction, second-order vibrational perturbation theory (VPT2) is particularly appealing to treat medium-to-large systems. Among the available VPT2 models,[163,164,191,193,194,199,200,206,214] Barone and coworkers[199,200,214–216] has developed a general framework to compute thermodynamic properties, vibrational energies and transition intensities from the vibrational ground state to fundamentals, overtones and combination bands. One of the advantages of the VPT2 approach comes from its cost efficiency to compute accurate vibrational properties, at least for medium-sized semi-rigid systems, in particular when applied to a fourth-order normal mode representation of the anharmonic force field. Moreover, difficulties related to the electronic structure computations necessary to explore the multidimensional anharmonic potential energy surface can be overcome with purposely tailored reduced dimensionality VPT2 models.[331] The combination of the anharmonic thermodynamic properties, vibrational energies and intensities allows to simulate very accurate vibrational spectra of single molecules or mixtures of several species or conformers, which can be directly compared with experimental data. For simplicity, the theoretical background presented in the following is intended for Abelian symmetry groups, i.e. without degeneracy. Then, for reasons of compatibility with previous works, we have preferred to adopt units commonly used in the spectroscopic literature (instead of SI units), namely cm^{-1} for vibrational wavenumbers and km mol^{-1} for infrared intensities.

A.0.1 Vibrational energies

For a system with N internal degrees of freedom, the vibrational energies of the states of interest are given by,[163,194,199]

$$\text{Fundamentals: } \nu_{1_i} = \omega_i + 2\chi_{ii} + \frac{1}{2}\sum_{\substack{j=1\\ j\neq i}}^{N}\chi_{ij} \tag{A.1}$$

$$\text{Overtones: } \nu_{2_i} = 2\omega_i + 6\chi_{ii} + \sum_{\substack{j=1\\ j\neq i}}^{N}\chi_{ij} \tag{A.2}$$

$$\begin{aligned}\text{Combinations: } \nu_{1_i1_j} &= \omega_i + \omega_j + 2\chi_{ii} + 2\chi_{jj} + 2\chi_{ij} \\ &\quad + \frac{1}{2}\sum_{\substack{k=1\\ k\neq i,j}}^{N}\left[\chi_{ik} + \chi_{jk}\right] \\ &= \nu_{1_i} + \nu_{1_j} + \chi_{ij}\end{aligned} \tag{A.3}$$

where the anharmonic χ matrix has the form,[163,194,199,200,216]

$$16\chi_{ii} = k_{iiii} - \frac{5{k_{iii}}^2}{3\omega_i} - \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{\left(8{\omega_i}^2 - 3{\omega_j}^2\right){k_{iij}}^2}{\omega_j\left(4{\omega_i}^2 - {\omega_j}^2\right)} \tag{A.4}$$

$$\begin{aligned}4\chi_{ij} = k_{iijj} &- \frac{2\omega_i{k_{iij}}^2}{4{\omega_i}^2 - {\omega_j}^2} - \frac{2\omega_j{k_{ijj}}^2}{4{\omega_j}^2 - {\omega_i}^2} \\ &- \frac{k_{iii}k_{ijj}}{\omega_i} - \frac{k_{jjj}k_{iij}}{\omega_j} \\ &+ \sum_{\substack{k=1\\ k\neq i,j}}^{N}\left[\frac{2\omega_k\left({\omega_i}^2 + {\omega_j}^2 - {\omega_k}^2\right){k_{ijk}}^2}{\Delta_{ijk}} - \frac{k_{iik}k_{jjk}}{\omega_k}\right] \\ &+ \frac{4\left({\omega_i}^2 + {\omega_j}^2\right)}{\omega_i\omega_j}\sum_{\tau=x,y,z}B_e^{\tau}\left(\zeta_{ij}^{\tau}\right)^2\end{aligned} \tag{A.5}$$

$$\Delta_{ijk} = {\omega_i}^4 + {\omega_j}^4 + {\omega_k}^4 - 2\left({\omega_i}^2{\omega_j}^2 + {\omega_i}^2{\omega_k}^2 + {\omega_j}^2{\omega_k}^2\right)$$

ω_i is the harmonic wavenumber (in cm^{-1}) associated to mode i, B_e^{τ} is the diagonal inertial tensor of the molecule at equilibrium geometry and ζ_{ij}^{τ} the Coriolis constant coupling mode i and j along the rotation axis I_τ. Finally, k_{ijk} and k_{ijkl} are the third and fourth derivatives of the potential energy V with respect to the dimensionless normal coordinates $\boldsymbol{q}$, also referred to as cubic and quartic force constants,

$$k_{ijk} = \frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} \quad \text{and} \quad k_{ijkl} = \frac{\partial^4 V}{\partial q_i \partial q_j \partial q_k \partial q_l}$$

It should be noted that a less general-purpose notation with respect to Refs.[199,200,216] has been adopted here since we consider only energy minima.

One of the problem of the VPT2 treatment concerns resonances which result in vanishing or nearly-vanishing denominators in the formulae used to compute the quantities of interest. Vibrational energies, for instance, suffer from the presence of so-called Fermi resonances, occurring when the wavenumber of a mode is twice the wavenumber of another mode ($\omega_i \approx 2\omega_j$, type I) or equal to the sum of the wavenumbers of two other modes ($\omega_i \approx \omega_j + \omega_k$, type II), which results in an improper account of the anharmonic correction. Workarounds have been proposed in the literature to overcome such a problematic situation.[194,201,217,458] A commonly adopted scheme is based on a two-step procedure.[194,199] First, resonant terms are identified by mean of a *ad hoc* test, such as the one proposed by Martin *et al.*,[201] and successively removed, giving the deperturbed VPT2 (DVPT2). In the second step, the discarded terms are reintroduced through a variational treatment. We refer to the whole procedure as the generalized VPT2 (GVPT2). This approach has shown to provide very accurate results.[119,195,219–221,227,240,459,460]
The Martin test[201] used in our GVPT2 implementation relies on a double control on the difference in wavenumber ($|\omega_i - (\omega_j + \omega_k)| \leq \Delta_\omega$ where $j = k$ for type I) and the magnitude of the possibly resonant terms,[199]

$$\text{For type I, } \Delta^1_{ij} = \frac{{k_{ijj}}^4}{256(2\omega_j - \omega_i)^3} \geq \mathcal{K}^I$$

$$\text{For type II, } \Delta^1_{ijk} = \frac{{k_{ijk}}^4}{64(\omega_j + \omega_k - \omega_i)^3} \geq \mathcal{K}^{II}$$

where $\mathcal{K}^I$ and $\mathcal{K}^{II}$ are empirically defined thresholds. Default, reliable values are provided internally for these parameters but can be easily modified by the user. Moreover, it is possible to redefine the resonant terms, which can be convenient to allow a direct comparison with the results from other theoretical approaches, as well as to test the influence of any specific interaction on the overall results.[119] Terms defined as resonant are removed and subsequently treated variationally.
The quality of vibrational wavenumbers computed at the GVPT2 level,[163,194] as implemented in the GAUSSIAN package,[199,200,216] has already been well documented.[119,128,213,227,240,460,461] Here we will only discuss the computational requirements for the force field evaluation. First, equilibrium structures have to be optimized using tight convergence criteria (maximum forces and displacements lower than 1.5×10^{-5} Hartree/Bohr and 6×10^{-5} Å, respectively). The semi-diagonal quartic force field is obtained by numerical differentiation of the analytical second derivatives along each normal coordinate (with the standard 0.01 Å step). In order to get accurate results, all computations are done with at least an ultrafine integration grid (99 radial shells and 590 angular points per shell) for the numerical integrations to obtain the two-electron integrals and their derivatives. Concerning the nearly-resonant contributions, the test proposed by Martin *et al.*[201] is applied with the

following settings $\Delta_\omega = 200$ cm^{-1}, $\mathcal{K}^I = 1$ cm^{-1}, $\mathcal{K}^{II} = 1$ cm^{-1} used by default.

Reduced dimensionality vibrational perturbative approach

In reduced dimensionality VPT2 computations, a set of M "active" normal modes is defined, for which anharmonic wavenumbers are evaluated. The choice of active modes is related to the system under study, for instance the most intense transitions in the IR or Raman spectrum, a range of energy or a molecular fragment, such as a molecular probe. In practice, numerical differentiation is performed only along the normal coordinates of active modes, so that for any non-linear N-atomic system $2M + 1$ Hessians need to be computed, instead of $6N - 11$. This approach leads to a reduced number of force constants available for the perturbative treatment. Assuming that index i corresponds to an active mode while j and k correspond to inactive modes, all the cubic force constants where index i is present at least once (k_{ijk}, k_{ijj}, k_{iik}, k_{iii}) and all k_{iijj} and k_{iiij} quartic force constants are computed. However, a limited number of cubic force constants (terms including only j and k indices) is not evaluated. In practice, this omission has an impact with respect to the full-dimensionality treatment only when a subset of the strongly coupled vibrations is included in the set of active modes. Indeed, there are only two terms in Eq. A.5 which would not be computed, namely:

$$\Delta_j = \frac{k_{jjj}k_{iij}}{4\omega_j} \quad \text{and} \quad \Delta_{jk} = \frac{k_{iik}k_{jjk}}{4\omega_k} \tag{A.6}$$

The first term, Δ_j, should not be omitted when both k_{jjj} and k_{iij} are large. Since $k_{iij}/4\omega_j$ is known, a qualitative estimate of the impact of the whole term in the calculation of χ_{ij} is possible. If $k_{iij}/4\omega_j$ is very small, it is likely that Δ_j gives a small contribution and, therefore, can be safely neglected. A larger value of k_{iij} reflects a significant coupling between modes i and j, meaning that also mode j must be included in the set of active modes. Chemical intuition can also help to assess more precisely the magnitude of $k_{jjj}k_{iij}$ by noting that k_{jjj} is important only when mode j is strongly anharmonic. In such a case, mode j must be included in the set of active modes to be treated anharmonically. If the numerical differentiation along j is too cumbersome and is not relevant *per se* (e.g. mode j is not in the wavenumber range under study), an alternative approach would be to include only the missing term k_{jjj} through numerical evaluation of one-dimensional energy third derivatives. This task is much less computationally demanding than the complete computation.

The second term, Δ_{jk}, reflects the coupling of mode k with both modes i and j. As before, the known k_{iik} can be used to guess the possible influence of the whole term, which does not contribute to Eq. A.5 if $k_{iik}/4\omega_k$ is negligible, i.e. when i and k are uncoupled. On the other hand, for large k_{iik} constants, the whole Δ_{jk} term is not negligible if mode j is also strongly coupled to mode k. A more detailed analysis has been given in Ref.[331] Here we just summarize that the reduced dimensionality

approach can be safely applied if all modes which are strongly coupled are included simultaneously in the VPT2 computations, so that the necessary force constants are evaluated. The choice of normal modes can be based on "chemical intuition", for example the geometrical confinement of vibrations and the spanned wavenumber range, since the vibrations localized in different regions of the molecular systems and with significantly different wavenumbers are unlikely to be strongly coupled. Then, such a choice must be followed by proper validation.[331] Otherwise, the choice of active modes can be derived from lower-level electronic structure computations, resonance conditions, or normal mode similarity (See Ref.[422] and references therein).

A.0.2 Band intensities

The need to account for both the anharmonicity of the potential energy surface (PES) and the property of interest makes particularly difficult the derivation of equations for the transition moments at the anharmonic level. Due to the complexity of such a treatment, various approximations have been employed, in particular by considering independently the wave function and the property, so that different levels of theory can be applied to each term and only one of them is treated beyond the harmonic approximation.[462,463] Following the first complete derivation by Handy and co-workers,[464] we adopted the alternative approach presented by Vázquez and Stanton[206] and proposed a general formulation for any property function of the normal coordinates or their associated momenta, which can be expanded in the form of a polynomial truncated at the third order. In this work, we will only consider the molar absorption coefficient $\epsilon(\bar{\nu}_0)$,[215]

$$\epsilon(\bar{\nu}_0) = \frac{8\pi^3 \mathcal{N}_A \bar{\nu}_0}{3000 \ln(10) hc(4\pi\epsilon_0)} \sum_o \sum_e D_{o,e} \delta(\varepsilon_e - \varepsilon_o - \bar{\nu}_0) \tag{A.7}$$

$$\tag{A.8}$$

where $\bar{\nu}_0$ is the wavenumber of the incident light and δ is the Dirac function. The summation runs over all initial states o and final states e with wavenumbers ε_o and ε_e, respectively.

The static dipole strength $D_{o,e}$ is given by,

$$D_{o,e} = |\langle \boldsymbol{\mu} \rangle_{o,e}|^2 \tag{A.9}$$

$$\tag{A.10}$$

with $\boldsymbol{\mu}$ the electric dipole.

The general formulation to compute the transition moments of the electric dipole vector $\langle \boldsymbol{\mu} \rangle_{o,e}$ and polarizability tensor $\langle \boldsymbol{\alpha} \rangle_{o,e}$ has been given in Ref.[215] Since the properties of interest here depend on the normal coordinates and not on their

conjugate momenta, we can use a more compact formula for fundamentals than the one used in Ref.,[215]

$$
\begin{aligned}
\langle\, P^a \,\rangle_{0,1_i} &= \frac{P^a_i}{\sqrt{2}} + \frac{1}{4\sqrt{2}} \sum_j P^a_{jji} \\
&- \frac{1}{8\sqrt{2}} \sum_{jk} k_{ijkk} P^a_j \left[\frac{1}{\omega_i + \omega_j} - \frac{1-\delta_{ij}}{\omega_i - \omega_j} \right] \\
&- \frac{1}{8\sqrt{2}} \sum_{jk} \Bigg\{ k_{ijk} P^a_{jk} \left(\frac{1}{\omega_i + \omega_j + \omega_k} - \frac{1}{\omega_i - \omega_j - \omega_k} \right) \\
&\quad + \frac{2k_{jkk}}{\omega_j} P^a_{ji} \Bigg\} \\
&+ \frac{1}{2\sqrt{2}} \sum_{jk} \left(\sum_\tau B^\tau_e \zeta^\tau_{ik} \zeta^\tau_{jk} \right) P^a_j \Bigg\{ \frac{\sqrt{\omega_i \omega_j}}{\omega_k} \left(\frac{1}{\omega_i + \omega_j} + \frac{1-\delta_{ij}}{\omega_i - \omega_j} \right) \\
&\quad - \frac{\omega_k}{\sqrt{\omega_i \omega_j}} \left(\frac{1}{\omega_i + \omega_j} - \frac{1-\delta_{ij}}{\omega_i - \omega_j} \right) \Bigg\} \\
&+ \frac{1}{16\sqrt{2}} \sum_{jkl} k_{ikl} k_{jkl} P^a_j \Bigg\{ \frac{4\omega_j (\omega_k + \omega_l)(1-\delta_{ij})(1-\delta_{ik})(1-\delta_{il})}{({\omega_j}^2 - {\omega_i}^2)\left[(\omega_k + \omega_l)^2 - {\omega_i}^2\right]} \\
&\quad + \frac{(\omega_k + \omega_l)\left[(\omega_k + \omega_l)^2 - 3{\omega_i}^2\right] \delta_{ij}(1+\delta_{ik})(1-\delta_{il})}{\omega_i \left[(\omega_k + \omega_l)^2 - {\omega_i}^2\right]^2} \\
&\quad + \frac{4\omega_j (3\omega_k + 4\omega_i)(1-\delta_{ij})(1-\delta_{ik})\delta_{il}}{\omega_k ({\omega_j}^2 - {\omega_i}^2)(\omega_k + 2\omega_i)} \Bigg\} \\
&+ k_{ijk} k_{llk} P^a_j \Bigg\{ \frac{\delta_{ij}}{\omega_i \omega_k} \left(1 + \frac{2\delta_{ik}\delta_{il}}{9} \right) \\
&\quad + \frac{4\omega_j (1-\delta_{ij})(1-\delta_{ik})(1-\delta_{il})}{\omega_k \left({\omega_j}^2 - {\omega_i}^2\right)} \\
&\quad + \frac{4\omega_j \delta_{ik}(1-\delta_{ij})}{\omega_i ({\omega_j}^2 - {\omega_i}^2)} \left(1 + \frac{2\delta_{ij}}{3} \right) \Bigg\}
\end{aligned}
\tag{A.11}
$$

where P^a is a cartesian component of the electric dipole and δ_{ij} is the Kronecker delta. $\boldsymbol{P}_i$, $\boldsymbol{P}_{ij}$ and $\boldsymbol{P}_{ijk}$ are respectively the first, second and third derivatives with respect to the dimensionless normal coordinates,

$$
\boldsymbol{P}_i = \frac{\partial \boldsymbol{P}}{\partial q_i} \; ; \; \boldsymbol{P}_{ij} = \frac{\partial^2 \boldsymbol{P}}{\partial q_i \partial q_j} \; ; \; \boldsymbol{P}_{ijk} = \frac{\partial^3 \boldsymbol{P}}{\partial q_i \partial q_j \partial q_k}
$$

For overtones and combination bands, simpler formulae are obtained, respectively,

$$\begin{aligned}\langle P^a \rangle_{0,2_i} &= \frac{1}{2\sqrt{2}} P^a_{ij} \\ &+ \frac{1}{4\sqrt{2}} \sum_k k_{ijk} P^a_k \left(\frac{1}{\omega_i + \omega_j - \omega_k} - \frac{1}{\omega_i + \omega_j + \omega_k} \right)\end{aligned} \tag{A.12}$$

$$\begin{aligned}\langle P^a \rangle_{0,1_i 1_j} &= \frac{1}{2} P^a_{ij} \\ &+ \frac{1}{4} \sum_k k_{ijk} P^a_k \left(\frac{1}{\omega_i + \omega_j - \omega_k} - \frac{1}{\omega_i + \omega_j + \omega_k} \right)\end{aligned} \tag{A.13}$$

Similarly to vibrational energies, Eqs. A.11, A.12 and A.13 are plagued by the existence of potential singularities, which lead to excessive contributions from the anharmonic terms. In addition to the Fermi resonances described above, 1-1 resonances ($\omega_i \approx \omega_j$) are present for fundamental transitions. The protocol used to avoid unphysical contributions in the anharmonic correction is similar to the one used for energies. Terms identified as resonant through *ad hoc* tests are removed from the calculations. Contrary to the GVPT2 model used for wavenumbers, this step is not followed by any variational treatment of the resonant terms. In practice, the same Martin test is adopted to find Fermi resonances. For 1-1 resonances, the test described in Ref.[215] is used here. It relies on two complementary conditions. The first one is the difference in wavenumbers,

$$|\omega_i - \omega_j| \leq \Delta_\omega^{1-1}$$

The second one is based the magnitude of the numerator and uses two thresholds depending on the term, $\mathcal{K}_1^{1-1}$ for cubic terms and $\mathcal{K}_2^{1-1}$ for quartic and Coriolis terms,

$$\begin{aligned}&|k_{ijk} k_{lmn}| \geq \mathcal{K}_1^{1-1} \\ &|k_{ijkl}| \geq \mathcal{K}_2^{1-1} \; ; \; \left| \sum_\tau B_e^\tau \zeta_{ik}^\tau \zeta_{jk}^\tau \right| \geq \mathcal{K}_2^{1-1}\end{aligned}$$

with the default values as following: $\Delta_\omega^{1-1} = 2\ \mathrm{cm}^{-1}$, $\mathcal{K}_1^{1-1} = 1\ \mathrm{cm}^{-2}$, $\mathcal{K}_2^{1-1} = 1\ \mathrm{cm}^{-1}$.

Bibliography

[1] Miller, S. L. *Science* **1953**, *117*, 528–529.

[2] Wächtershäuser, G. *Proceedings of the National Academy of Sciences* **1990**, *87*(1), 200–204.

[3] Saladino, R.; Crestini, C.; Pino, S.; Costanzo, G.; Di Mauro, E. *Physics of life reviews* **2012**, *9*(1), 84–104.

[4] Cafferty, B. J.; Hud, N. V. *Current opinion in chemical biology* **2014**, *22*, 146–157.

[5] Di Mauro, E.; Saladino, R.; Trifonov, E. N. *Journal of Biomolecular Structure and Dynamics* **2014**, *32*(4), 512–522.

[6] Gupta, N. S.; Steele, A.; Fogel, M.; Griffin, P.; Adams, M.; Summons, R. E.; Yang, H.; Cody, G. D. *Organic Geochemistry* **2014**, *67*, 35–40.

[7] Pino, S.; Sponer, J. E.; Costanzo, G.; Saladino, R.; Mauro, E. D. *Life* **2015**, *5*(1), 372–384.

[8] Sponer, J. E.; Sponer, J.; Giorgi, A.; Di Mauro, E.; Pino, S.; Costanzo, G. *The Journal of Physical Chemistry B* **2015**, *119*(7), 2979–2989.

[9] Bernstein, M. P.; Sandford, S. A.; Allamandola, L. J.; Gillette, J. S.; Clemett, S. J.; Zare, R. N. *Science* **1999**, *283*(5405), 1135–1138.

[10] Ehrenfreund, P.; Charnley, S. B. *Annual Review of Astronomy and Astrophysics* **2000**, *38*(1), 427–483.

[11] Dworkin, J. P.; Deamer, D. W.; Sandford, S. A.; Allamandola, L. J. *Proceedings of the National Academy of Sciences* **2001**, *98*(3), 815–819.

[12] Emeline, A.; Otroshchenko, V.; Ryabchuk, V.; Serpone, N. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **2003**, *3*(3), 203–224.

[13] Brucato, J.; Strazzulla, G.; Baratta, G.; Rotundi, A.; Colangeli, L. *Origins of Life and Evolution of Biospheres* **2006**, *36*(5-6), 451–457.

[14] Saladino, R.; Botta, G.; Delfino, M.; Di Mauro, E. *Chemistry-A European Journal* **2013**, *19*(50), 16916–16922.

[15] Woods, P. M.; Slater, B.; Raza, Z.; Viti, S.; Brown, W. A.; Burke, D. J. *The Astrophysical Journal* **2013**, *777*(2), 90.

[16] Calcutt, H.; Viti, S.; Codella, C.; Beltrán, M. T.; Fontani, F.; Woods, P. M. *Monthly Notices of the Royal Astronomical Society* **2014**, *443*(4), 3157–3173.

[17] Saladino, R.; Carota, E.; Botta, G.; Kapralov, M.; Timoshenko, G. N.; Rozanov, A. Y.; Krasavin, E.; Di Mauro, E. *Proceedings of the National Academy of Sciences* **2015**, *112*(21), E2746–E2755.

[18] Barone, V.; Latouche, C.; Skouteris, D.; Vazart, F.; Balucani, N.; Ceccarelli, C.; Lefloch, B. *Monthly Notices of the Royal Astronomical Society: Letters* **2015**, *453*(1), L31–L35.

[19] Herbst, E.; Van Dishoeck, E. F. *Annual Review of Astronomy and Astrophysics* **2009**, *47*, 427–480.

[20] Botta, O. In *Astrobiology: Future Perspectives*; Ehrenfreund, P., Irvine, W., Owen, T., Becker, L., Blank, J., Brucato, J., Colangeli, L., Derenne, S., Dutrey, A., Despois, D., Lazcano, A., Robert, F., Eds., Vol. 305 of *Astrophysics and Space Science Library;* Springer Netherlands, 2005; pages 359–391.

[21] Garrod, R. T. *The Astrophysical Journal* **2013**, *765*(1), 60.

[22] Caselli, P.; Viti, S.; others. *The Astrophysical Journal Letters* **2014**, *787*(2), L33.

[23] Elsila, J. E.; Glavin, D. P.; Dworkin, J. P. *Meteoritics & Planetary Science* **2009**, *44*(9), 1323–1330.

[24] Martins, Z.; Watson, J.; Sephton, M.; Botta, O.; Ehrenfreund, P.; Gilmour, I. *Meteoritics & Planetary Science* **2006**, *41*(7), 1073–1080.

[25] Pizzarello, S.; Huang, Y.; Alexandre, M. R. *Proceedings of the National Academy of Sciences* **2008**, *105*(10), 3700–3704.

[26] Pizzarello, S. *Journal of the Mexican Chemical Society* **2009**, *53*(4), 253–260.

[27] Callahan, M. P.; Smith, K. E.; Cleaves, H. J.; Ruzicka, J.; Stern, J. C.; Glavin, D. P.; House, C. H.; Dworkin, J. P. *Proceedings of the National Academy of Sciences* **2011**, *108*(34), 13995–13998.

[28] Gomes, R.; Levison, H. F.; Tsiganis, K.; Morbidelli, A. *Nature* **2005**, *435*(7041), 466–469.

[29] Miller, S. L. *Journal of the American Chemical Society* **1955**, *77*(9), 2351–2361.

[30] Miller, S. L. *Biochimica et Biophysica Acta* **1957**, *23*(0), 480 – 489.

[31] Kvenvolden, K.; Lawless, J.; Pering, K.; Peterson, E.; Flores, J.; Ponnamperuma, C.; Kaplan, I. R.; Moore, C. *Nature* **1970**, *228*, 923–926.

[32] Stoks, P. G.; Schwartz, A. W. *Nature* **1979**, *282*, 709–710.

[33] Schlesinger, G.; Miller, S. *Journal of Molecular Evolution* **1983**, *19*(5), 376–382.

[34] Miller, S.; Schlesinger, G. *Origins of life* **1984**, *14*(1-4), 83–90.

[35] Wächtershäuser, G. *Origins of life and evolution of the biosphere* **1990**, *20*(2), 173–176.

[36] Chyba, C.; Sagan, C. *Nature* **1992**, *355*, 125–132.

[37] Botta, O.; Bada, J. *Surveys in Geophysics* **2002**, *23*(5), 411–467.

[38] Ehrenfreund, P.; Irvine, W.; Becker, L.; Blank, J.; Brucato, J. R.; Colangeli, L.; Derenne, S.; Despois, D.; Dutrey, A.; Fraaije, H.; Lazcano, A.; Owen, T.; Robert, F.; an International Space Science Institute ISSI-Team. *Reports on Progress in Physics* **2002**, *65*(10), 1427.

[39] Hazen, R. *Genesis: The Scientific Quest for Lifes Origin;* Joseph Henry Press: Washington, D.C., 2005.

[40] Dobson, C. M.; Ellison, G. B.; Tuck, A. F.; Vaida, V. *Proceedings of the National Academy of Sciences* **2000**, *97*(22), 11864–11868.

[41] Tuck, A. *Surveys in Geophysics* **2002**, *23*(5), 379–409.

[42] Hennet, R.-C.; Holm, N.; Engel, M. *Naturwissenschaften* **1992**, *79*(8), 361–365.

[43] Marshall, W. L. *Geochimica et Cosmochimica Acta* **1994**, *58*(9), 2099 – 2106.

[44] Brandes, J. A.; Boctor, N. Z.; Cody, G. D.; Cooper, B. A.; Hazen, R. M.; Yoder, H. S. *Nature* **1998**, *395*, 365–367.

[45] Cody, G. D.; Boctor, N. Z.; Filley, T. R.; Hazen, R. M.; Scott, J. H.; Sharma, A.; Yoder, H. S. *Science* **2000**, *289*(5483), 1337–1340.

[46] Russell, M. *Acta Biotheoretica* **2007**, *55*(2), 133–179.

[47] Martin, W.; Baross, J.; Kelley, D.; Russell, M. J. *Nature Reviews Microbiology* **2008**, *6*, 805–814.

[48] Braakman, R. *Proceedings of the National Academy of Sciences* **2013**, *110*(33), 13236–13237.

[49] Miller, S. L.; Urey, H. C. *Science* **1959**, *130*(3389), 1622–1624.

[50] Oró, J. *Nature* **1961**, *191*, 1193–1194.

[51] Shapiro, R. *Origins of life and evolution of the biosphere* **1995**, *25*(1-3), 83–98.

[52] Zimmer, C. *Eos, Transactions American Geophysical Union* **2002**, *83*(47), 550–550.

[53] Bonner, W. A.; Rubenstein, E. *Biosystems* **1987**, *20*(1), 99 – 111.

[54] Meinert, C.; De Marcellus, P.; Le Sergeant dHendecourt, L.; Nahon, L.; Jones, N. C.; Hoffmann, S. V.; Bredehöft, J. H.; Meierhenrich, U. J. *Physics of life reviews* **2011**, *8*(3), 307–330.

[55] Boyd, R.; Kajino, T.; Onaka, T. In *Journal of Physics: Conference Series*, Vol. 403, page 012032. IOP Publishing, 2012.

[56] Famiano, M.; Boyd, R.; Kajino, T.; Onaka, T.; Koehler, K.; Hulbert, S. *Symmetry* **2014**, *6*(4), 909–925.

[57] Meinert, C.; Hoffmann, S. V.; Cassam-Chenaï, P.; Evans, A. C.; Giri, C.; Nahon, L.; Meierhenrich, U. J. *Angewandte Chemie* **2014**, *126*(1), 214–218.

[58] Meinert, C.; Cassam-Chenaï, P.; Jones, N. C.; Nahon, L.; Hoffmann, S. V.; Meierhenrich, U. J. *Origins of Life and Evolution of Biospheres* **2015**, pages 1–13.

[59] Tarasevych, A. V.; Sorochinsky, A. E.; Kukhar, V. P.; Guillemin, J.-C. *Origins of Life and Evolution of Biospheres* **2013**, *43*(2), 129–135.

[60] Tarasevych, A. V.; Sorochinsky, A. E.; Kukhar, V. P.; Guillemin, J.-C. *Chemical Communications* **2015**, *51*(32), 7054–7057.

[61] Hazen, R. M.; Filley, T. R.; Goodfriend, G. A. *Proceedings of the National Academy of Sciences* **2001**, *98*(10), 5487–5490.

[62] Hazen, R. M. *American Mineralogist* **2006**, *91*(11-12), 1715–1729.

[63] Yun, Y.; Gellman, A. J. *Nature chemistry* **2015**, *7*, 520–525.

[64] Bonner, W. *Origins of life and evolution of the biosphere* **1991**, *21*(2), 59–111.

[65] Soai, K.; Shibata, T.; Morioka, H.; Choji, K. *Nature* **1995**, *378*, 767–768.

[66] Feringa, B. L.; van Delden, R. A. *Angewandte Chemie International Edition* **1999**, *38*(23), 3418–3438.

[67] Kondepudi, D. K.; Asakura, K. *Accounts of Chemical Research* **2001**, *34*(12), 946–954.

[68] Podlech, J. *Cellular and Molecular Life Sciences CMLS* **2001**, *58*(1), 44–60.

[69] Sato, I.; Urabe, H.; Ishiguro, S.; Shibata, T.; Soai, K. *Angewandte Chemie* **2003**, *115*(3), 329–331.

[70] Soai, K.; Sato, I.; Shibata, T.; Komiya, S.; Hayashi, M.; Matsueda, Y.; Imamura, H.; Hayase, T.; Morioka, H.; Tabira, H.; Yamamoto, J.; Kowata, Y. *Tetrahedron: Asymmetry* **2003**, *14*(2), 185 – 188.

[71] Blackmond, D. G. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*(16), 5732–5736.

[72] Pizzarello, S.; Weber, A. L. *Science* **2004**, *303*(5661), 1151.

[73] Blanco, C.; Ribó, J. M.; Crusats, J.; El-Hachemi, Z.; Moyano, A.; Hochberg, D. *Physical Chemistry Chemical Physics* **2013**, *15*(5), 1546–1556.

[74] Ribó, J. M.; Crusats, J.; El-Hachemi, Z.; Moyano, A.; Blanco, C.; Hochberg, D. *Astrobiology* **2013**, *13*(2), 132–142.

[75] Cronin, J. R.; Pizzarello, S. *Science* **1997**, *275*(5302), 951–955.

[76] Engel, M. H.; Macko, S. A. *Nature* **1997**, *389*, 265–268.

[77] Pizzarello, S.; Cronin, J. *Geochimica et Cosmochimica Acta* **2000**, *64*(2), 329 – 338.

[78] Pizzarello, S.; Zolensky, M.; Turk, K. A. *Geochimica et Cosmochimica Acta* **2003**, *67*(8), 1589 – 1595.

[79] Glavin, D. P.; Dworkin, J. P. *Proceedings of the National Academy of Sciences* **2009**, *106*(14), 5487–5492.

[80] James Cleaves II, H.; Michalkova Scott, A.; Hill, F. C.; Leszczynski, J.; Sahai, N.; Hazen, R. *Chem. Soc. Rev.* **2012**, *41*, 5502–5525.

[81] Bernal, J. *The physical basis of life;* Routledge and Paul: London, 1951.

[82] Lahav, N.; Chang, S. *Journal of Molecular Evolution* **1976**, *8*(4), 357–380.

[83] Gibbs, D.; Lohrmann, R.; Orgel, L. *Journal of Molecular Evolution* **1980**, *15*(4), 347–354.

[84] Sowerby, S.; Heckl, W.; Petersen, G. *Journal of Molecular Evolution* **1996**, *43*(5), 419–424.

[85] Zamaraev, K.; Romannikov, V.; Salganik, R.; Wlassoff, W.; Khramtsov, V. *Origins of life and evolution of the biosphere* **1997**, *27*(4), 325–337.

[86] Parsons, I.; Lee, M. R.; Smith, J. V. *Proceedings of the National Academy of Sciences* **1998**, *95*(26), 15173–15176.

[87] Sowerby, S.; Heckl, W. *Origins of life and evolution of the biosphere* **1998**, *28*(3), 283–310.

[88] Sowerby, S. J.; Edelwirth, M.; Heckl, W. M. *The Journal of Physical Chemistry B* **1998**, *102*(30), 5914–5922.

[89] Sowerby, S.; Petersen, G.; Holm, N. *Origins of life and evolution of the biosphere* **2002**, *32*(1), 35–46.

[90] Ertem, G. *Origins of life and evolution of the biosphere* **2004**, *34*(6), 549–570.

[91] Schoonen, M.; Smirnov, A.; Cohn, C. *AMBIO: A Journal of the Human Environment* **2004**, *33*(8), 539–551.

[92] Hazen, R. M.; Sverjensky, D. A. *Cold Spring Harbor Perspectives in Biology* **2010**, *2*(5), 815–824.

[93] Saladino, R.; Brucato, J. R.; De Sio, A.; Botta, G.; Pace, E.; Gambicorti, L. *Astrobiology* **2011**, *11*(8), 815–824.

[94] Fornaro, T.; Brucato, J. R.; Branciamore, S.; Pucci, A. *International Journal of Astrobiology* **2013**, *12*, 78–86.

[95] Fuchida, S.; Masuda, H.; Shinoda, K. *Origins of Life and Evolution of Biospheres* **2014**, pages 1–16.

[96] Burcar, B. T.; Barge, L. M.; Trail, D.; Watson, E. B.; Russell, M. J.; McGown, L. B. *Astrobiology* **2015**, *15*(7), 509–522.

[97] Lahav, N. *Biogenesis: Theories of Lifes Origin;* Oxford University Press: New York, 1999.

[98] Hazen, R. M.; Sholl, D. S. *Nature Materials* **2003**, *2*(6), 367–374.

[99] Downs, R. T.; Hazen, R. M. *Journal of Molecular Catalysis A: Chemical* **2004**, *216*(2), 273 – 285.

[100] Sholl, D. S.; Gellman, A. J. *AIChE Journal* **2009**, *55*(10), 2484–2490.

[101] Senanayake, S.; Idriss, H. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*(5), 1194–1198.

[102] Shkrob, I. A.; Marin, T. M.; Adhikary, A.; Sevilla, M. D. *The Journal of Physical Chemistry C* **2011**, *115*(8), 3393–3403.

[103] Fornaro, T.; Brucato, J. R.; Pace, E.; Guidi, M. C.; Branciamore, S.; Pucci, A. *Icarus* **2013**, *226*(1), 1068 – 1085.

[104] Franchi, M.; Gallori, E. *Origins of life and evolution of the biosphere* **2004**, *34*(1-2), 133–141.

[105] Gallori, E.; Biondi, E.; Franchi, M. In *Life in the Universe*; Seckbach, J., Chela-Flores, J., Owen, T., Raulin, F., Eds., Vol. 7 of *Cellular Origin and Life in Extreme Habitats and Astrobiology;* Springer Netherlands, 2004; pages 145–148.

[106] Scappini, F.; Casadei, F.; Zamboni, R.; Franchi, M.; Gallori, E.; Monti, S. *International Journal of Astrobiology* **2004**, *3*, 17–19.

[107] Biondi, E.; Branciamore, S.; Maurel, M.-C.; Gallori, E. *BMC Evolutionary Biology* **2007**, *7*(Suppl 2), S2.

[108] Poch, O.; Jaber, M.; Stalport, F.; Nowak, S.; Georgelin, T.; Lambert, J.-F.; Szopa, C.; Coll, P. *Astrobiology* **2015**, *15*(3), 221–237.

[109] Laane, J., Ed. *Frontiers of Molecular Spectroscopy;* Elsevier: Amsterdam, 2009.

[110] Quack, M., Merkt, F., Eds. *Handbook of High-resolution Spectroscopy;* John Wiley & Sons, Inc., 2011.

[111] Barone, V., Ed. *Computational Strategies for Spectroscopy, from Small Molecules to Nano Systems.;* John Wiley & Sons, Inc. Hoboken: New Jersey, 2011.

[112] Grunenberg, J., Ed. *Computational Spectroscopy: Methods, Experiments and Applications;* John Wiley & Sons, Inc., 2010.

[113] Tasinato, N.; Pietropolli Charmet, A.; Stoppa, P.; Giorgianni, S.; Buffa, G. *J. Chem. Phys.* **2010**, *132*, 044315.

[114] Stoppa, P.; Pietropolli Charmet, A.; Tasinato, N.; Giorgianni, S.; Gambi, A. *J. Phys. Chem. A* **2009**, *113*, 1497.

[115] Pietropolli Charmet, A.; Stoppa, P.; Tasinato, N.; Baldan, A.; Giorgianni, S.; Gambi, A. *J. Chem. Phys.* **2010**, *133*, 044310.

[116] Espinoza, C.; Szczepanski, J.; Vala, M.; Polfer, N. *J. Phys. Chem. A* **2010**, *114*, 5919.

[117] Stepanian, S.; Reva, I.; Radchenko, E.; Rosado, M.; Duarte, M.; Fausto, R.; Adamowicz, L. *J. Phys. Chem. A* **1998**, *102*, 1041.

[118] Biczysko, M.; Bloino, J.; Carnimeo, I.; Panek, P.; Barone, V. *Journal of Molecular Structure* **2012**, *1009*, 74–82.

[119] Puzzarini, C.; Biczysko, M.; Barone, V. *J. Chem. Theory Comput.* **2011**, *7*(11), 3702–3710.

[120] Aamouche, A.; Ghomi, M.; Coulombeau, C.; Jobic, H.; Grajcar, L.; Baron, M.; Baumruk, V.; Turpin, P.; Henriet, C.; Berthier, G. *J. Phys. Chem.* **1996**, *100*, 5224.

[121] Tolstoy, V. P.; Chernyshova, I. V.; Skryshevsky, V. A. In *Handbook of Infrared Spectroscopy of Ultrathin Films;* John Wiley & Sons, Inc., 2003; pages 691–710.

[122] Dušan, H., Ed. *Theoretical Treatments of Hydrogen Bonding.;* John Wiley & Sons: Chichester, UK, 1997.

[123] Del Bene, J. E.; Jordan, M. J. T. *Int. Rev. Phys. Chem.* **1999**, *18*(1), 119–162.

[124] Buckingham, A. D.; Del Bene, J.; McDowell, S. A. C. *Chem. Phys. Lett.* **2008**, *463*(1-3), 1 – 10.

[125] Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P.; Kjaergaard, H. G.; Legon, A. C.; Mennucci, B.; Nesbitt, D. J. *Pure and Applied Chemistry* **2011**, *83*, 1619–1636.

[126] Lapinski, L.; Rostkowska, H.; Nowak, M. J.; Kwiatkowski, J. S.; Leszczyski, J. *Vibrational Spectroscopy* **1996**, *13*(1), 23 – 40.

[127] Nowak, M. J.; Lapinski, L.; Kwiatkowski, J. S.; Leszczynski, J. *J. Phys. Chem.* **1996**, *100*(9), 3527–3534.

[128] Biczysko, M.; Panek, P.; Barone, V. *Chem. Phys. Lett.* **2009**, *475*, 105–110.

[129] Santoro, F.; Barone, V.; Improta, R. *Journal of the American Chemical Society* **2009**, *131*(42), 15232–15245.

[130] Santoro, F.; Barone, V.; Lami, A.; Improta, R. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4934–4948.

[131] Carnimeo, I.; Biczysko, M.; Bloino, J.; Barone, V. *Phys. Chem. Chem. Phys.* **2011**, *13*(37), 16713–16727.

[132] Barone, V.; Baiardi, A.; Biczysko, M.; Bloino, J.; Cappelli, C.; Lipparini, F. *Phys. Chem. Chem. Phys.* **2012**, *14*(36), 12404–12422.

[133] Puzzarini, C.; Biczysko, M.; Barone, V.; Pena, I.; Cabezas, C.; Alonso, J. L. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16965–16975.

[134] Barone, V.; Biczysko, M.; Bloino, J. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759–1787.

[135] Penocchio, E.; Piccardo, M.; Barone, V. *J. Chem. Theory Comput.* **2015**, *0*(ja), DOI:10.1021/acs.jctc.5b00622.

[136] Piccardo, M.; Penocchio, E.; Puzzarini, C.; Biczysko, M.; Barone, V. *J. Phys. Chem. A* **2015**, *119*, 2058–2082.

[137] Piccardo, M.; Bloino, J.; Barone, V. *Int. J. Quant. Chem.* **2015**, *115*, 948–982.

[138] Barone, V.; Biczysko, M.; Puzzarini, C. *Acc. Chem. Res.* **2015**, *48*, 1413–1422.

[139] Puzzarini, C.; Ali, A.; Biczysko, M.; Barone, V. *The Astrophysical Journal* **2014**, *792*(2), 118–123.

[140] Puzzarini, C.; Biczysko, M.; Bloino, J.; Barone, V. *The Astrophysical Journal* **2014**, *785*(2), 107.

[141] Vazart, F.; Latouche, C.; Cimino, P.; Barone, V. *J. Chem. Theory Comput.* **2015**, *11*, DOI:10.1021/acs.jctc.5b00638.

[142] Vazart, F.; Calderini, D.; Skouteris, D.; Latouche, C.; Barone, V. *J. Chem. Theory Comput.* **2015**, *11*(3), 1165–1171.

[143] Shenhav, B.; Lancet, D. *Origins of life and evolution of the biosphere* **2004**, *34*(1-2), 181–194.

[144] Rimola, A.; Sodupe, M.; Ugliengo, P. In *Genesis - In The Beginning*; Seckbach, J., Ed., Vol. 22 of *Cellular Origin, Life in Extreme Habitats and Astrobiology;* Springer Netherlands, 2012; pages 345–362.

[145] Galli, D.; Palla, F. *Astron. Astrophys.* **1998**, *335*, 403–420.

[146] Bodo, E.; Gianturco, F.; Martinazzo, R. *Physics Reports* **2003**, *384*(3), 85 – 119.

[147] Bovino, S.; urk, R.; Galli, D.; Tacconi, M.; Gianturco, F. A. *The Astrophysical Journal* **2012**, *752*(1), 19.

[148] Simakov, A.; Miller, G. B. S.; Bunkan, A. J. C.; Hoffmann, M. R.; Uggerud, E. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16615–16625.

[149] Smith, M. L.; Claire, M. W.; Catling, D. C.; Zahnle, K. J. *Icarus* **2014**, *231*(0), 51 – 64.

[150] Simoncini, E.; Virgo, N.; Kleidon, A. *Earth System Dynamics* **2013**, *4*(2), 317–331.

[151] Marion, G.; Kargel, J.; Catling, D.; Lunine, J. *Icarus* **2014**, *236*(0), 1 – 8.

[152] Stahl, F.; Schleyer, P.; III, H. S.; Kaiser, R. *Planetary and Space Science* **2002**, *50*(78), 685 – 692.

[153] Segré, D.; Pilpel, Y.; Lancet, D. *Physica A: Statistical Mechanics and its Applications* **1998**, *249*(14), 558 – 564.

[154] Segré, D.; Ben-Eli, D.; Lancet, D. *Proceedings of the National Academy of Sciences* **2000**, *97*(8), 4112–4117.

[155] Branciamore, S.; Gallori, E.; Szathmry, E.; Czrn, T. *Journal of Molecular Evolution* **2009**, *69*(5), 458–469.

[156] Gallori, E.; Branciamore, S. In *Genesis - In The Beginning*; Seckbach, J., Ed., Vol. 22 of *Cellular Origin, Life in Extreme Habitats and Astrobiology;* Springer Netherlands, 2012; pages 55–66.

[157] Briones, C.; Stich, M.; Manrubia, S. C. *RNA* **2009**, *15*(5), 743–749.

[158] Gevertz, J.; Gan, H. H.; Schlick, T. *RNA* **2005**, *11*(6), 853–863.

[159] Pasquali, S.; Gan, H. H.; Schlick, T. *Nucleic Acids Research* **2005**, *33*(4), 1384–1398.

[160] Ono, N.; Ikegami, T. In *Advances in Artificial Life*; Kelemen, J., Sosk, P., Eds., Vol. 2159 of *Lecture Notes in Computer Science;* Springer Berlin Heidelberg, 2001; pages 186–195.

[161] E. Bright Wilson, J., Decius, J. C., Cross, P. C., Eds. *MOLECULAR VIBRATION. The Theory of Infrared and Raman Vibrational Spectra;* McGRAW-HILL BOOK COMPANY, INC.: New York, USA, 1955.

[162] Califano, S., Ed. *Vibrational States;* John Wiley & Sons: J. W. Arrowsmith Ltd., Winterstoke Road, Bristol BS2 2NT, 1976.

[163] Mills, I. M. In Rao, K. N., Mathews, C. W., Eds., *Molecular Spectroscopy: Modern Research, Volume 1*, page 115, 1972.

[164] Nielsen, H. H. *Reviews of Modern Physics* **1951**, *23*(2), 90–136.

[165] Aliev, M.; Watson, J. In *Molecular Spectroscopy: Modern Research*; Rao, K., Ed.; Academic Press, 1985; pages 1 – 67.

[166] Papoušek, D. D.; Aliev, M. R. M. R. *Molecular vibrational-rotational spectra : theory and applications of high resolution infrared, microwave, and Raman spectroscopy of polyatomic molecules;* Amsterdam ; New York : Elsevier Scientific Pub. Co. ; New York : Distribution for the U.S.A. and Canada, Elsevier /North-Holland, 1982.

[167] Licari, D.; Baiardi, A.; Biczysko, M.; Egidi, F.; Latouche, C.; Barone, V. *J. Comp. Chem.* **2015**, *36*, 321–334.

[168] Jensen, P.; Bunker, P. R. *Computational Molecular Spectroscopy;* John Wiley and Sons Ltd, Chichester, UK, 2000.

[169] Grunenberg, J., Ed. *Computational Spectroscopy: Methods, Experiments and Applications;* Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2010.

[170] Fornaro, T.; Biczysko, M.; Monti, S.; Barone, V. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10112–10128.

[171] Fornaro, T.; Carnimeo, I. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering;* Elsevier, 2014.

[172] Fornaro, T.; Carnimeo, I.; Biczysko, M. *The Journal of Physical Chemistry A* **2015**, *119*(21), 5313–5326.

[173] Fornaro, T.; Burini, D.; Biczysko, M.; Barone, V. *The Journal of Physical Chemistry A* **2015**, *119*(18), 4224–4236.

[174] Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. *Phys. Chem. Chem. Phys.* **2013**, *15*(5), 1358–1363.

[175] Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. *J. Chem. Theory Comput.* **2013**, *9*(3), 1533–1547.

[176] Olbert-Majkut, A.; Lundell, J.; Wierzejewska, M. *J. Phys. Chem. A* **2014**, *118*(2), 350–357.

[177] Latouche, C.; Barone, V. *J. Chem. Theory Comput.* **2014**, *10*(12), 5586–5592.

[178] Reva, I.; M. Nunes, C.; Biczysko, M.; Fausto, R. *J. Phys. Chem. A* **2015**, *119*, 2614–2627.

[179] Sinha, P.; Boesch, S. E.; Gu, C.; Wheeler, R. A.; Wilson, A. K. *The Journal of Physical Chemistry A* **2004**, *108*(42), 9213–9217.

[180] Andersson, M. P.; Uvdal, P. *The Journal of Physical Chemistry A* **2005**, *109*(12), 2937–2941.

[181] Teixeira, F.; Melo, A.; Cordeiro, M. N. D. *The Journal of chemical physics* **2010**, *133*(11), 114109.

[182] Alecu, I.; Zheng, J.; Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2010**, *6*(9), 2872–2887.

[183] Katsyuba, S. A.; Zvereva, E. E.; Burganov, T. I. *The Journal of Physical Chemistry A* **2013**, *117*(30), 6664–6670.

[184] Rauhut, G.; Pulay, P. *The Journal of Physical Chemistry* **1995**, *99*(10), 3093–3100.

[185] Baker, J.; Jarzecki, A. A.; Pulay, P. *The Journal of Physical Chemistry A* **1998**, *102*(8), 1412–1424.

[186] Fábri, C.; Szidarovszky, T.; Magyarfalvi, G.; Tarczay, G. *The Journal of Physical Chemistry A* **2011**, *115*(18), 4640–4649.

[187] Irikura, K. K.; Johnson III, R. D.; Kacker, R. N.; Kessel, R. *The Journal of chemical physics* **2009**, *130*(11), 114102.

[188] Pernot, P.; Cailliez, F. *The Journal of chemical physics* **2011**, *134*(16), 167101.

[189] Mills, I. M.; Academic Press, New York, 1972; chapter Vibration-Rotation Structure in Asymmetric- and Symmetric-Top Molecules, pages 115–140.

[190] Truhlar, D. G.; Olson, R. W.; Jeannotte, A. C.; Overend, J. *J. Am. Chem. Soc.* **1976**, *98*(9), 2373–2379.

[191] Isaacson, A. D.; Truhlar, D. G.; Scanlon, K.; Overend, J. *J. Chem. Phys.* **1981**, *75*(6), 3017–3024.

[192] Clabo Jr., D. A.; Allen, W. D.; Remington, R. B.; Yamaguchi, Y.; Schaefer III, H. F. *Chem. Phys.* **1988**, *123*(2), 187–239.

[193] Allen, W. D.; Yamaguchi, Y.; Császár, A. G.; Clabo Jr., D. A.; Remington, R. B.; Schaefer III, H. F. *Chem. Phys.* **1990**, *145*(3), 427–466.

[194] Amos, R. D.; Handy, N. C.; Green, W. H.; Jayatilaka, D.; Willetts, A.; Palmieri, P. *J. Chem. Phys.* **1991**, *95*(11), 8323–8336.

[195] Maslen, P. E.; Handy, N. C.; Amos, R. D.; Jayatilaka, D. *J. Chem. Phys.* **1992**, *97*(6), 4233–4254.

[196] Gaw, F.; Willetts, A.; Handy, N.; Green, W. In *Advances in Molecular Vibrations and Collision Dynamics*; Bowman, J. M., Ed.; JAI Press, 1991; Vol. 1B; pages 169–185.

[197] Zhang, Q.; Day, P. N.; Truhlar, D. G. *J. Chem. Phys.* **1993**, *98*, 4948–4958.

[198] Barone, V. *J. Chem. Phys.* **1994**, *101*, 10666–10676.

[199] Barone, V. *J. Chem. Phys.* **2005**, *122*(1), 014108.

[200] Barone, V. *J. Chem. Phys.* **2004**, *120*(7), 3059–3065.

[201] Martin, J. M. L.; Lee, T. J.; Taylor, P. M.; François, J.-P. *J. Chem. Phys.* **1995**, *103*(7), 2589–2602.

[202] Stanton, J. F.; Gauss, J. *J. Chem. Phys.* **1998**, *108*, 9218–9220.

[203] Ruud, K.; Åstrand, P.-O.; Taylor, P. R. *J. Chem. Phys.* **2000**, *112*(6), 2668–2683.

[204] Stanton, J. F.; Gauss, J. *Int. Rev. Phys. Chem.* **2000**, *19*, 61–95.

[205] Neugebauer, J.; Hess, B. A. *J. Chem. Phys.* **2003**, *118*(16), 7215–7225.

[206] Vázquez, J.; Stanton, J. F. *Mol. Phys.* **2006**, *104*(3), 377–388.

[207] Vázquez, J.; Stanton, J. F. *Mol. Phys.* **2007**, *105*(1), 101–109.

[208] Christiansen, O. *J. Chem. Phys.* **2003**, *119*, 5773.

[209] Bowman, J. M. *Science* **2000**, *290*, 724.

[210] Carter, S.; Sharma, A. R.; Bowman, J. M.; Rosmus, P.; Tarroni, R. *J. Chem. Phys.* **2009**, *131*, 224106.

[211] Roy, T. K.; Gerber, R. B. *Phys. Chem. Chem. Phys.* **2013**, *15*, 9468–9492.

[212] Hermes, M. R.; Hirata, S. *International Reviews in Physical Chemistry* **2015**, *34*(1), 71–97.

[213] Barone, V.; Festa, G.; Grandi, A.; Rega, N.; Sanna, N. *Chemical Physics Letters* **2004**, *388*(46), 279 – 283.

[214] Barone, V.; Bloino, J.; Guido, C. A.; Lipparini, F. *Chem. Phys. Lett.* **2010**, *496*(1-3), 157–161.

[215] Bloino, J.; Barone, V. *J. Chem. Phys.* **2012**, *136*(12), 124108.

[216] Bloino, J.; Biczysko, M.; Barone, V. *J. Chem. Theory Comput.* **2012**, *8*(3), 1015–1036.

[217] Kuhler, K. M.; Truhlar, D. G.; Isaacson, A. D. *J. Chem. Phys.* **1996**, *104*, 4664–4671.

[218] Gaussian 09 Revision D.01. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; *et al.*. **2009**.

[219] Burcl, R.; Handy, N. C.; Carter, S. *Spectrochim. Acta A* **2003**, *59*(8), 1881 – 1893.

[220] Burcl, R.; Carter, S.; Handy, N. C. *Phys. Chem. Chem. Phys.* **2004**, *6*, 340–343.

[221] Boese, A. D.; Martin, J. *J. Phys. Chem. A* **2004**, *108*, 3085–3096.

[222] Carbonniere, P.; Lucca, T.; Pouchan, C.; Rega, N.; Barone, V. *J. Comp. Chem.* **2005**, *26*(4), 384–388.

[223] Cane, E.; Miani, A.; Trombetti, A. *J. Phys. Chem. A* **2007**, *111*(33), 8218–8222.

[224] Becke, A. D. *The Journal of chemical physics* **1993**, *98*(7), 5648–5652.

[225] Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.

[226] Neese, F.; Schwabe, T.; Grimme, S. *J. Chem. Phys.* **2007**, *126*, 124115.

[227] Biczysko, M.; Panek, P.; Scalmani, G.; Bloino, J.; Barone, V. *J. Chem. Theory Comput.* **2010**, *6*(7), 2115–2125.

[228] Carnimeo, I.; Puzzarini, C.; Tasinato, N.; Stoppa, P.; Charmet, A. P.; Biczysko, M.; Cappelli, C.; Barone, V. *J. Chem. Phys.* **2013**, *139*(7), 074310.

[229] Biczysko, M.; Bloino, J.; Brancato, G.; Cacelli, I.; Cappelli, C.; Ferretti, A.; Lami, A.; Monti, S.; Pedone, A.; Prampolini, G.; others. *Theoretical Chemistry Accounts* **2012**, *131*(4), 1–19.

[230] Thicoipe, S.; Carbonniere, P.; Pouchan, C. *Phys. Chem. Chem. Phys.* **2013**, *15*, 11646–11652.

[231] Kristyán, S.; Pulay, P. *Chemical physics letters* **1994**, *229*(3), 175–180.

[232] Grimme, S. *J. Comp. Chem.* **2004**, *25*(12), 1463–1473.

[233] Grimme, S. *J. Comp. Chem.* **2006**, *27*, 1787–1799.

[234] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*(15), 154104.

[235] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*(15), 154108.

[236] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[237] Chai, J.-D.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106/1–15.

[238] Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

[239] Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*(3), 669–676.

[240] Puzzarini, C.; Biczysko, M.; Barone, V. *J. Chem. Theory Comput.* **2010**, *6*, 828.

[241] Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.

[242] Wu, Q.; Yang, W. *The Journal of chemical physics* **2002**, *116*(2), 515–524.

[243] Grimme, S. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*(2), 211–228.

[244] Barone, V.; Cimino, P.; Stendardo, E. *J. Chem. Theory Comput.* **2008**, *4*(5), 751–764.

[245] Barone, V.; Cimino, P. *Chem. Phys. Lett.* **2008**, *454*(1-3), 139–143.

[246] DiLabio, G. A. *Chem. Phys. Lett.* **2008**, *455*(46), 348 – 353.

[247] Mackie, I. D.; DiLabio, G. A. *J. Phys. Chem. A* **2008**, *112*(43), 10968–10976.

[248] Mackie, I. D.; DiLabio, G. A. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6092–6098.

[249] Torres, E.; DiLabio, G. A. *Journal of Physical Chemistry Letters* **2012**, *3*(13), 1738–1744.

[250] DiLabio, G. A.; Koleini, M.; Torres, E. *Theoretical Chemistry Accounts* **2013**, *132*(10), 1–13.

[251] Torres, E.; DiLabio, G. A. *Journal of Chemical Theory and Computation* **2013**, *9*(8), 3342–3349.

[252] DiLabio, G. A.; Johnson, E. R.; Otero-de-la Roza, A. *Physical Chemistry Chemical Physics* **2013**, *15*(31), 12821–12828.

[253] Barone, V.; Cossi, M. *The Journal of Physical Chemistry A* **1998**, *102*(11), 1995–2001.

[254] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *Journal of computational chemistry* **2003**, *24*(6), 669–681.

[255] Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chemical Physics Letters* **1996**, *255*(4), 327–335.

[256] Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*(8), 3032–3041.

[257] Cancès, E.; Mennucci, B. *Journal of mathematical chemistry* **1998**, *23*(3-4), 309–326.

[258] Mennucci, B.; Cancès, E.; Tomasi, J. *The Journal of Physical Chemistry B* **1997**, *101*(49), 10506–10517.

[259] Improta, R.; Barone, V. *Journal of the American Chemical Society* **2004**, *126*(44), 14320–14321.

[260] Manzhos, S.; Carrington, T.; Yamashita, K. *Surface Science* **2011**, *605*(5), 616–622.

[261] Escamilla-Roa, E.; Moreno, F. *Planetary and Space Science* **2012**, *70*(1), 1 – 9.

[262] Escamilla-Roa, E.; Moreno, F. *Planetary and Space Science* **2013**, *75*(0), 1 – 10.

[263] Newman, S. P.; Di Cristina, T.; Coveney, P. V.; Jones, W. *Langmuir* **2002**, *18*(7), 2933–2939.

[264] Costa, D.; Lomenech, C.; Meng, M.; Stievano, L.; Lambert, J.-F. *Journal of Molecular Structure: THEOCHEM* **2007**, *806*(13), 253 – 259.

[265] Lomenech, C.; Bery, G.; Costa, D.; Stievano, L.; Lambert, J. F. *ChemPhysChem* **2005**, *6*(6), 1061–1070.

[266] Stievano, L.; Piao, L. Y.; Lopes, I.; Meng, M.; Costa, D.; Lambert, J.-F. *European Journal of Mineralogy* **2007**, *19*(3), 321–331.

[267] Rimola, A.; Sodupe, M.; Tosoni, S.; Civalleri, B.; Ugliengo, P. *Langmuir* **2006**, *22*(15), 6593–6604.

[268] Rimola, A.; Sodupe, M.; Ugliengo, P. *The Journal of Physical Chemistry C* **2009**, *113*(14), 5741–5750.

[269] Nonella, M.; Seeger, S. *ChemPhysChem* **2008**, *9*(3), 414–421.

[270] Zhao, Y. L.; Koppen, S.; Frauenheim, T. *The Journal of Physical Chemistry C* **2011**, *115*(19), 9615–9621.

[271] Costa, D.; Tougerti, A.; Tielens, F.; Gervais, C.; Stievano, L.; Lambert, J. F. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6360–6368.

[272] Rimola, A.; Civalleri, B.; Ugliengo, P. *Langmuir* **2008**, *24*(24), 14027–14034.

[273] Gambino, G. L.; Lombardo, G. M.; Grassi, A.; Marletta, G. *The Journal of Physical Chemistry B* **2004**, *108*(8), 2600–2607.

[274] Gambino, G. L.; Grassi, A.; Marletta, G. *The Journal of Physical Chemistry B* **2006**, *110*(10), 4836–4845.

[275] Han, J. W.; Sholl, D. S. *Langmuir* **2009**, *25*(18), 10737–10745.

[276] Wright, L. B.; Walsh, T. R. *The Journal of Physical Chemistry C* **2012**, *116*(4), 2933–2945.

[277] Langel, W.; Menken, L. *Surface Science* **2003**, *538*(12), 1 – 9.

[278] Schravendijk, P.; Ghiringhelli, L. M.; Site, L. D.; van der Vegt. *The Journal of Physical Chemistry C* **2007**, *111*(6), 2631–2642.

[279] Giacomelli, C. E.; Avena, M. J.; De Pauli, C. P. *Langmuir* **1995**, *11*(9), 3483–3490.

[280] Rankin, R. B.; Sholl, D. S. *Langmuir* **2006**, *22*(19), 8096–8103.

[281] Kühnle, A.; Molina, L. M.; Linderoth, T. R.; Hammer, B.; Besenbacher, F. *Phys. Rev. Lett.* **2004**, *93*, 086101.

[282] Jones, G.; Jones, L.; Thibault-Starzyk, F.; Seddon, E.; Raval, R.; Jenkins, S.; Held, G. *Surface Science* **2006**, *600*(9), 1924 – 1935.

[283] Rankin, R. B.; Sholl, D. S. *Surface Science* **2005**, *574*(1), L1 – L8.

[284] Rankin, R. B.; Sholl, D. S. *The Journal of Physical Chemistry B* **2005**, *109*(35), 16764–16773.

[285] Chen, Q.; Frankel, D.; Richardson, N. *Surface Science* **2002**, *497*(13), 37 – 46.

[286] Toomes, R.; Kang, J.-H.; Woodruff, D.; Polcik, M.; Kittel, M.; Hoeft, J.-T. *Surface Science* **2003**, *522*(13), L9 – L14.

[287] Efstathiou, V.; Woodruff, D. *Surface Science* **2003**, *531*(3), 304 – 318.

[288] Tonner, R. *ChemPhysChem* **2010**, *11*(5), 1053–1061.

[289] Parikh, S. J.; Kubicki, J. D.; Jonsson, C. M.; Jonsson, C. L.; Hazen, R. M.; Sverjensky, D. A.; Sparks, D. L. *Langmuir* **2011**, *27*(5), 1778–1787.

[290] Monti, S.; Walsh, T. R. *The Journal of Physical Chemistry C* **2010**, *114*(50), 22197–22206.

[291] Monti, S.; van Duin, A. C. T.; Kim, S.-Y.; Barone, V. *The Journal of Physical Chemistry C* **2012**, *116*(8), 5141–5150.

[292] Orme, C.; Noy, A.; Wierzbicki, A.; McBride, M. T.; Grantham, M.; Teng, H.; Dove, P.; DeYoreo, J. *Nature* **2001**, *411*, 775–779.

[293] Asthagiri, A.; Hazen, R. M. *Molecular Simulation* **2007**, *33*(4-5), 343–351.

[294] Pan, H.; Tao, J.; Xu, X.; Tang, R. *Langmuir* **2007**, *23*(17), 8972–8981.

[295] Rimola, A.; Corno, M.; Zicovich-Wilson, C. M.; Ugliengo, P. *Journal of the American Chemical Society* **2008**, *130*(48), 16181–16183.

[296] Jimenez-Izal, E.; Chiatti, F.; Corno, M.; Rimola, A.; Ugliengo, P. *The Journal of Physical Chemistry C* **2012**, *116*(27), 14561–14567.

[297] Almora-Barrios, N.; Austen, K. F.; de Leeuw, N. H. *Langmuir* **2009**, *25*(9), 5018–5025.

[298] Corno, M.; Rimola, A.; Bolis, V.; Ugliengo, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6309–6329.

[299] Boehme, C.; Marx, D. *Journal of the American Chemical Society* **2003**, *125*(44), 13362–13363.

[300] Nair, N. N.; Schreiner, E.; Marx, D. *Journal of the American Chemical Society* **2006**, *128*(42), 13815–13826.

[301] Pollet, R.; Boehme, C.; Marx, D. *Origins of Life and Evolution of Biospheres* **2006**, *36*(4), 363–379.

[302] Mignon, P.; Sodupe, M. *Phys. Chem. Chem. Phys.* **2012**, *14*, 945–954.

[303] Mignon, P.; Ugliengo, P.; Sodupe, M. *The Journal of Physical Chemistry C* **2009**, *113*(31), 13741–13749.

[304] Michalkova, A.; Robinson, T. L.; Leszczynski, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7862–7881.

[305] Michalkova, A.; Leszczynski, J. In *Practical Aspects of Computational Chemistry I*; Leszczynski, J., Shukla, M. K., Eds.; Springer Netherlands, 2012; pages 645–672.

[306] Robinson, T.; Michalkova, A.; Gorb, L.; Leszczynski, J. *Journal of Molecular Structure* **2007**, *844845*(0), 48 – 58.

[307] Mignon, P.; Sodupe, M. *The Journal of Physical Chemistry C* **2013**, *117*(49), 26179–26189.

[308] Edelwirth, M.; Freund, J.; Sowerby, S.; Heckl, W. *Surface Science* **1998**, *417*(23), 201 – 209.

[309] Ortmann, F.; Schmidt, W. G.; Bechstedt, F. *Phys. Rev. Lett.* **2005**, *95*, 186101.

[310] Le, D.; Kara, A.; Schrder, E.; Hyldgaard, P.; Rahman, T. S. *Journal of Physics: Condensed Matter* **2012**, *24*(42), 424210.

[311] Lee, J.-H.; Choi, Y.-K.; Kim, H.-J.; Scheicher, R. H.; Cho, J.-H. *J. Phys. Chem. C* **2013**, *117*(26), 13435–13441.

[312] Panigrahi, S.; Bhattacharya, A.; Banerjee, S.; Bhattacharyya, D. *J. Phys. Chem. C* **2012**, *116*(7), 4374–4379.

[313] Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2722–2729.

[314] Spiwok, V.; Hobza, P.; Řezáč, J. *The Journal of Physical Chemistry C* **2011**, *115*(40), 19455–19462.

[315] Chen, Q.; Frankel, D. J.; Richardson, N. V. *Langmuir* **2002**, *18*(8), 3219–3225.

[316] Preuss, M.; Schmidt, W. G.; Bechstedt, F. *Phys. Rev. Lett.* **2005**, *94*, 236102.

[317] Kong, H.; Sun, Q.; Wang, L.; Tan, Q.; Zhang, C.; Sheng, K.; Xu, W. *ACS Nano* **2014**, *8*(2), 1804–1808.

[318] Sowerby, S. J.; Mörth, C.-M.; Holm, N. G. *Astrobiology* **2001**, *1*(4), 481–487.

[319] Sivakova, S.; Rowan, S. J. *Chem. Soc. Rev.* **2005**, *34*, 9–21.

[320] Hud, N. V.; Cafferty, B. J.; Krishnamurthy, R.; Williams, L. D. *Chemistry & biology* **2013**, *20*(4), 466–474.

[321] Siebert, F., Hildebrandt, P., Eds. *Vibrational Spectroscopy in Life Science;* Wiley-VCH Verlag GmbH and Co. KGaA: Weinheim, Germany, 2008.

[322] Berova, N., Polavarapu, P. L., Nakanishi, K., Woody, R. W., Eds. *Comprehensive Chiroptical Spectroscopy: Applications in Stereochemical Analysis of Synthetic Compounds, Natural Products, and Biomolecules, Volume 2;* John Wiley & Sons, Inc. Hoboken: New Jersey, 2012.

[323] Watson, M. D.; Fechtenkötter, A.; Müllen, K. *Chem. Rev.* **2001**, *101*(5), 1267–1300.

[324] Elemans, J. A. A. W.; Lei, S.; De Feyter, S. *Angew. Chem. Int. Ed.* **2009**, *48*(40), 7298–7332.

[325] Iwaura, R.; Iizawa, T.; Minamikawa, H.; Ohnishi-Kameyama, M.; Shimizu, T. *Small* **2010**, *6*(10), 1131–1139.

[326] Biczysko, M.; Latajka, Z. *J. Phys. Chem. A* **2002**, *106*(13), 3197–3201.

[327] Wormer, P. E. S.; van der Avoird, A. *Chem. Rev.* **2000**, *100*(11), 4109–4144.

[328] Begue, D.; Carbonniere, P.; Pouchan, C. *J. Phys. Chem. A* **2005**, *109*(20), 4611–4616.

[329] Jarmelo, S.; Reva, I.; Carey, P.; Fausto, R. *Vibrational spectroscopy* **2007**, *43*(2), 395–404.

[330] Rozenberg, M.; Shoham, G.; Reva, I.; Fausto, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2376–2383.

[331] Barone, V.; Biczysko, M.; Bloino, J.; Borkowska-Panek, M.; Carnimeo, I.; Panek, P. *Int. J. Quant. Chem.* **2012**, *112*, 2185.

[332] Shock, E.; Canovas, P. *Geofluids* **2010**, *10*(1-2), 161–192.

[333] Bach, W.; Paulick, H.; Garrido, C. J.; Ildefonse, B.; Meurer, W. P.; Humphris, S. E. *Geophysical research letters* **2006**, *33*(13).

[334] Estrada, C. F.; Sverjensky, D. A.; Pelletier, M.; Razafitianamaharavo, A.; Hazen, R. M. *Geochimica et Cosmochimica Acta* **2015**, *155*, 172–186.

[335] Sahai, N.; Sverjensky, D. A. *Computers & Geosciences* **1998**, *24*(9), 853–873.

[336] Kaplan, I. G. *Intermolecular interactions: physical picture, computational methods and model potentials;* John Wiley & Sons, 2006.

[337] Stone, A. *The theory of intermolecular forces;* Oxford University Press, 2013.

[338] Bartlett, R. J. *Annual Review of Physical Chemistry* **1981**, *32*(1), 359–401.

[339] Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

[340] Hirao, K. *Chemical Physics Letters* **1992**, *190*(34), 374 – 380.

[341] Cramer, C. J. *Essentials of computational chemistry: theories and models;* John Wiley & Sons, 2013.

[342] Hobza, P.; Reschel, T.; others. *Journal of computational chemistry* **1995**, *16*(11), 1315–1325.

[343] Pérez-Jordá, J.; Becke, A. D. *Chemical physics letters* **1995**, *233*(1), 134–137.

[344] Pérez-Jordá, J. M.; San-Fabián, E.; Pérez-Jiménez, A. J. *The Journal of chemical physics* **1999**, *110*(4), 1916–1920.

[345] Zhao, Y.; Truhlar, D. G. *The Journal of chemical physics* **2006**, *125*(19), 194101.

[346] Grimme, S.; Antony, J.; Schwabe, T.; Mück-Lichtenfeld, C. *Organic & Biomolecular Chemistry* **2007**, *5*(5), 741–758.

[347] Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. *Journal of Chemical Theory and Computation* **2008**, *4*(12), 1996–2000.

[348] Johnson, E. R.; Becke, A. D. *The Journal of chemical physics* **2008**, *128*(12), 124105.

[349] Pavone, M.; Rega, N.; Barone, V. *Chemical Physics Letters* **2008**, *452*(4), 333–339.

[350] Gräfenstein, J.; Cremer, D. *The Journal of chemical physics* **2009**, *130*(12), 124105.

[351] Johnson, E. R.; Mackie, I. D.; DiLabio, G. A. *Journal of Physical Organic Chemistry* **2009**, *22*(12), 1127–1135.

[352] Marom, N.; Tkatchenko, A.; Scheffler, M.; Kronik, L. *Journal of Chemical Theory and Computation* **2009**, *6*(1), 81–90.

[353] Bretherick, N. H.; van Mourik, T. *Journal of Chemical Theory and Computation* **2010**, *6*(9), 2687–2700.

[354] Sato, T.; Nakai, H. *The Journal of chemical physics* **2010**, *133*(19), 194101.

[355] Vazquez-Mayagoitia, A.; Sherrill, C. D.; Apra, E.; Sumpter, B. G. *Journal of Chemical Theory and Computation* **2010**, *6*(3), 727–734.

[356] Austin, A.; Petersson, G. A.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K. *Journal of Chemical Theory and Computation* **2012**, *8*(12), 4989–5007.

[357] Vydrov, O. A.; Van Voorhis, T. *Journal of Chemical Theory and Computation* **2012**, *8*(6), 1929–1934.

[358] Grimme, S.; Steinmetz, M. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031–16042.

[359] Gauss, J. In *Encyclopedia of Computational Chemistry;* John Wiley Sons, Ltd, 2002.

[360] Raghavachari, K.; Trucks, G.; Pople, J.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

[361] T.H. Dunning, J. *J. Chem. Phys.* **1989**, *90*, 1007.

[362] Kendall, R.; T.H. Dunning, J.; Harrison, R. *J. Chem. Phys.* **1992**, *96*, 6796.

[363] Woon, D.; T.H. Dunning, J. *J. Chem. Phys.* **1995**, *103*, 4572.

[364] Szczesniak, M.; Nowak, M. J.; Rostkowska, H.; Szczepaniak, K.; Person, W. B.; Shugar, D. *J. Am. Chem. Soc.* **1983**, *105*, 5969–5976.

[365] Graindourze, M.; Smets, J.; Zeegers-Huyskens, T.; Maes, G. *J. Mol. Struct.* **1990**, *222*(3-4), 345–364.

[366] Chin, S.; Scott, I.; Szczepaniak, K.; Person, W. B. *J. Am. Chem. Soc.* **1984**, *106*, 3415–3422.

[367] Lapinski, L.; Reva, I.; Nowak, M. J.; Fausto, R. *Physical chemistry chemical physics* **2011**, *13*(20), 9676–9684.

[368] Ramaekers, R.; Maes, G.; Adamowicz, L.; Dkhissi, A. *Journal of Molecular Structure* **2001**, *560*(1), 205–221.

[369] Szczepaniak, K.; Szczesniak, M. M.; Person, W. B. *The Journal of Physical Chemistry A* **2000**, *104*(16), 3852–3863.

[370] Kwiatkowski, J. S.; Leszczynski, J. *The Journal of Physical Chemistry* **1996**, *100*(3), 941–953.

[371] Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*(32), 10142–10151.

[372] Morgado, C. A.; Jurečka, P.; Svozil, D.; Hobza, P.; Šponer, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3522–3534.

[373] Rezác, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*(8), 2427–2438.

[374] Cho, Y.; Min, S. K.; Yun, J.; Kim, W. Y.; Tkatchenko, A.; Kim, K. S. *J. Chem. Theory Comput.* **2013**, *9*(4), 2090–2096.

[375] Plutzer, C.; Hunig, I.; Kleinermanns, K. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1158–1163.

[376] Ottiger, P.; Frey, J. A.; Frey, H.-M.; Leutwyler, S. *The Journal of Physical Chemistry A* **2009**, *113*(18), 5280–5288.

[377] Saenger, W.; Jeffrey, G. *Hydrogen Bonding in Biological Structures;* Springer-Verlag: Berlin Heidelberg, Germany, 1991.

[378] Robertson, E. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1–18.

[379] de Vries, M. S.; Hobza, P. *Annu. Rev. Phys. Chem.* **2007**, *58*, 585–612.

[380] Garand, E.; Kamrath, M. Z.; Jordan, P. A.; Wolk, A. B.; Leavitt, C. M.; McCoy, A. B.; Miller, S. J.; Johnson, M. A. *Science* **2012**, *335*(6069), 694–698.

[381] Kimizuka, N.; Kawasaki, T.; Kunitake, T. *Journal of the American Chemical Society* **1993**, *115*(10), 4387–4388.

[382] Deamer, D.; Dworkin, J. P.; Sandford, S. A.; Bernstein, M. P.; Allamandola, L. J. *Astrobiology* **2002**, *2*(4), 371–381.

[383] Kunitake, T. *Angewandte Chemie International Edition in English* **1992**, *31*(6), 709–726.

[384] Segré, D.; Ben-Eli, D.; Deamer, D. W.; Lancet, D. *Origins of Life and Evolution of the Biosphere* **2001**, *31*(1-2), 119–145.

[385] Deamer, D.; Dworkin, J. In *Prebiotic Chemistry*; Walde, P., Ed., Vol. 259 of *Topics in Current Chemistry;* Springer Berlin Heidelberg, 2005; pages 1–27.

[386] Barone, V.; Biczysko, M.; Borkowska-Panek, M.; Bloino, J. *ChemPhysChem* **2014**, *15*, 3355–3364.

[387] Papajak, E.; Leverentz, H. R.; Zheng, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*(5), 1197–1202.

[388] Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298–305.

[389] Barone, V.; Improta, R.; Rega, N. *Acc. Chem. Res.* **2008**, *41*(5), 605–616.

[390] Kamerlin, S. C. L.; Warshel, A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10401–10411.

[391] Labat, F.; Le Bahers, T.; Ciofini, I.; Adamo, C. *Acc. Chem. Res.* **2012**, *45*(8), 1268–1277.

[392] Mennucci, B. *Phys. Chem. Chem. Phys.* **2013**, *15*(18), 6583–6594.

[393] Lipparini, F.; Egidi, F.; Cappelli, C.; Barone, V. *J. Chem. Theory Comput.* **2013**, *9*(4), 1880–1884.

[394] Pruitt, S. R.; Bertoni, C.; Brorsen, K. R.; Gordon, M. S. *Acc. Chem. Res.* **2014**, *47*(9), 2786–2794.

[395] Jacob, C. R.; Neugebauer, J. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*(4), 325–362.

[396] Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. *J. Chem. Phys.* **2014**, *141*(3), 034107/1–17.

[397] Puzzarini, C.; Biczysko, M. *J. Phys. Chem. A* **2015**, *119*, 5386–5395.

[398] King, S. *Spectrochimica Acta Part A: Molecular Spectroscopy* **1972**, *28*(1), 165–175.

[399] Mardyukov, A.; Sánchez-García, E.; Rodziewicz, P.; Doltsinis, N. L.; Sander, W. *The Journal of Physical Chemistry A* **2007**, *111*(42), 10552–10561.

[400] Albrecht, M.; Rice, C. A.; Suhm, M. A. *The Journal of Physical Chemistry A* **2008**, *112*(33), 7530–7542.

[401] Choi, M. Y.; Miller, R. E. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3565–3573.

[402] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *Journal of Molecular Structure: THEOCHEM* **1999**, *461*, 1–21.

[403] Vreven, T.; Morokuma, K. *J. Comp. Chem.* **2000**, *21*(16), 1419–1432.

[404] Casaes, R. N.; Paul, J. B.; McLaughlin, R. P.; Saykally, R. J.; van Mourik, T. *J. Phys. Chem. A* **2004**, *108*(50), 10989–10996.

[405] Császár, A. G.; Fabri, C.; Szidarovszky, T.; Matyus, E.; Furtenbacher, T.; Czako, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1085–1106.

[406] Christiansen, O. *Phys. Chem. Chem. Phys.* **2012**, *14*, 6672–6687.

[407] Panek, P. T.; Jacob, C. R. *ChemPhysChem* **2014**.

[408] Cheng, X.; Steele, R. P. *J. Chem. Phys.* **2014**, *141*(10), 104105.

[409] Kurten, T.; Biczysko, M.; Rajamaki, T.; Laasonen, K.; Halonen, L. *J. Phys. Chem. B* **2005**, *109*(18), 8954–8960.

[410] Trani, F.; Barone, V. *Journal of Chemical Theory and Computation* **2011**, *7*(3), 713–719.

[411] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.

[412] Zheng, G.; Lundberg, M.; Jakowski, J.; Vreven, T.; Frisch, M. J.; Morokuma, K. *Int. J. Quant. Chem.* **2009**, *109*, 1841.

[413] Barone, V.; Carnimeo, I.; Scalmani, G. *J. Chem. Theory Comput.* **2013**, *9*, 2052.

[414] Witek, H.; Irle, S.; Morokuma, K. *J. Chem. Phys.* **2004**, *121*, 5163.

[415] Witek, H.; Morokuma, K. *J. Comp. Chem.* **2004**, *25*, 1858.

[416] Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. A* **2006**, *110*, 13551.

[417] Malolepsza, E.; Witek, H.; Morokuma, K. *Chem. Phys. Lett.* **2005**, *412*, 237.

[418] Stewart, R. F.; Jensen, L. H. *Acta Crystallographica* **1967**, *23*(6), 1102–1105.

[419] Rozenberg, M.; Shoham, G.; Reva, I.; Fausto, R. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2004**, *60*(10), 2323 – 2336.

[420] Boys, S.; Bernardi, F. *Mol. Phys.* **1970**, *19*(4), 553–566.

[421] Duschinsky, F. *Acta Physicochimica URSS* **1937**, *7*, 551.

[422] Hirata, S.; Yagi, K. *Chem. Phys. Lett.* **2008**, *464*, 123–134.

[423] Jacob, C. R.; Reiher, M. *J. Chem. Phys.* **2009**, *130*(8), 084106.

[424] Singh, J. *Journal of Molecular Structure* **2008**, *876*(1), 127–133.

[425] Shock, E. L.; Schulte, M. D. *Journal of Geophysical Research: Planets (1991–2012)* **1998**, *103*(E12), 28513–28527.

[426] Holm, N. G. *Geobiology* **2012**, *10*(4), 269–279.

[427] Henrist, C.; Mathieu, J.-P.; Vogels, C.; Rulmont, A.; Cloots, R. *Journal of Crystal Growth* **2003**, *249*(1), 321–330.

[428] Lv, J.; Qiu, L.; Qu, B. *Journal of Crystal Growth* **2004**, *267*(3), 676–684.

[429] Prélot, B.; Villiéras, F.; Pelletier, M.; Gérard, G.; Gaboriaud, F.; Ehrhardt, J.-J.; Perrone, J.; Fedoroff, M.; Jeanjean, J.; Lefèvre, G.; others. *Journal of colloid and interface science* **2003**, *261*(2), 244–254.

[430] Hassan, M. S.; Villieras, F.; Gaboriaud, F.; Razafitianamaharavo, A. *Journal of colloid and interface science* **2006**, *296*(2), 614–623.

[431] Liu, X.; Cheng, J.; Sprik, M.; Lu, X.; Wang, R. *Geochimica et Cosmochimica Acta* **2013**, *120*, 487–495.

[432] Sverjensky, D. A.; Sahai, N. *Geochimica et Cosmochimica Acta* **1996**, *60*(20), 3773–3797.

[433] Sverjensky, D. A. *Geochimica et Cosmochimica Acta* **2003**, *67*(1), 17–28.

[434] Sverjensky, D. A. *Geochimica et Cosmochimica Acta* **2005**, *69*(2), 225–257.

[435] Sverjensky, D. A.; Fukushi, K. *Geochimica et cosmochimica acta* **2006**, *70*(15), 3778–3802.

[436] Fukushi, K.; Sverjensky, D. A. *Geochimica et cosmochimica acta* **2007**, *71*(15), 3717–3745.

[437] Jonsson, C. M.; Jonsson, C. L.; Sverjensky, D. A.; Cleaves, H. J.; Hazen, R. M. *Langmuir* **2009**, *25*(20), 12127–12135.

[438] Bahri, S.; Jonsson, C. M.; Jonsson, C. L.; Azzolini, D.; Sverjensky, D. A.; Hazen, R. M. *Environmental science & technology* **2011**, *45*(9), 3959–3966.

[439] Lee, N.; Sverjensky, D. A.; Hazen, R. M. *Environmental science & technology* **2014**, *48*(16), 9358–9365.

[440] Brown, G. E.; Henrich, V. E.; Casey, W. H.; Clark, D. L.; Eggleston, C.; Felmy, A.; Goodman, D. W.; Grätzel, M.; Maciel, G.; McCarthy, M. I.; others. *Chemical Reviews* **1999**, *99*(1), 77–174.

[441] Sverjensky, D. A.; Fukushi, K. *Environmental science & technology* **2006**, *40*(1), 263–271.

[442] Lee, N.; Hummer, D. R.; Sverjensky, D. A.; Rajh, T.; Hazen, R. M.; Steele, A.; Cody, G. D. *Langmuir* **2012**, *28*(50), 17322–17330.

[443] Westall, J. C. In *Geochemical processes at mineral surfaces*, Vol. 323; American Chemical Society Washington, DC, 1986; pages 54–78.

[444] Pokrovsky, O. S.; Schott, J. *Geochimica et Cosmochimica Acta* **2004**, *68*(1), 31–45.

[445] Danyluk, S. S.; Hruska, F. E. *Biochemistry* **1968**, *7*(3), 1038–1043.

[446] Martin, R. B. *Accounts of Chemical Research* **1985**, *18*(2), 32–38.

[447] Dawson, R.; Elliott, D.; Elliott, W.; Jones, K. In *Data for Biochemical Research;* Oxford University Press, Oxford, 1986.

[448] Sigel, H.; Massoud, S. S.; Corfu, N. A. *Journal of the American Chemical Society* **1994**, *116*(7), 2958–2971.

[449] Smith, R.; Martell, A.; Chen, Y. *Pure and Applied Chemistry* **1991**, *63*(7), 1015–1080.

[450] Cleaves, H. J.; Jonsson, C. M.; Jonsson, C. L.; Sverjensky, D. A.; Hazen, R. M. *Astrobiology* **2010**, *10*(3), 311–323.

[451] Feuillie, C.; Sverjensky, D. A.; Hazen, R. M. *Langmuir* **2014**, *31*(1), 240–248.

[452] Natochin, Y. V. *Herald of the Russian Academy of Sciences* **2007**, *77*(6), 581–591.

[453] Dubrovskii, V.; Sibirev, N.; Eliseev, I.; Vyazmin, S. Y.; Boitsov, V.; Natochin, Y. V.; Dubina, M. *The Journal of chemical physics* **2013**, *138*(24), 244906.

[454] Giles, C. H.; Smith, D.; Huitson, A. *Journal of Colloid and Interface Science* **1974**, *47*(3), 755–765.

[455] Yamauchi, O.; Odani, A.; Masuda, H.; Sigel, H. *Metal ions in biological systems* **1996**, *32*, 207–207.

[456] Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; others. *Journal of computer-aided molecular design* **2005**, *19*(6), 453–463.

[457] Tetko, I. V. *Drug Discovery Today* **2005**, *10*(22), 1497–1500.

[458] Krasnoshchekov, S. V.; Isayeva, E. V.; Stepanov, N. F. *J. Phys. Chem. A* **2012**, *116*(14), 3691–3709.

[459] Barone, V. *J. Phys. Chem. A* **2004**, *108*, 4146–4150.

[460] Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. *Phys. Chem. Chem. Phys.* **2013**, *15*, 10094.

[461] Barone, V.; Bloino, J.; M. Biczysko, M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1092–1101.

[462] Bouř, P. *J. Phys. Chem.* **1994**, *98*(36), 8862–8865.

[463] Faulkner, T. R.; Marcott, C.; Moscowitz, A.; Overend, J. *J. Am. Chem. Soc.* **1977**, *99*(25), 8160–8168.

[464] Miller, W. H.; Hernandez, R.; Handy, N. C.; Jayatilaka, D.; Willets, A. *Chem. Phys. Lett.* **1990**, *172*(1), 62–68.

List of Articles

- **"Adsorption of nucleic acid bases on magnesium oxide (MgO)"**
  T. Fornaro, J. R. Brucato, S. Branciamore and A. Pucci
  International Journal of Astrobiology, 12, 78-86 (2013)

- **"Infrared spectral investigations of UV irradiated nucleobases adsorbed on mineral surfaces"**
  T. Fornaro, J. R. Brucato, E. Pace, M. Cestelli Guidi, S. Branciamore and A. Pucci
  Icarus, 226 (1), 1068-1085 (2013)

- **"Development of extraction protocols for life detection biosensor-based instruments"**
  T. Fornaro, J. R. Brucato, A. Pucci and S. Branciamore
  Planetary and Space Science, 86, 75-79 (2013)

- **"Dispersion corrected DFT approaches for anharmonic vibrational frequency calculations: nucleobases and their dimers"**
  T. Fornaro, M. Biczysko, S. Monti and V. Barone
  Phys. Chem. Chem. Phys., 16, 10112-10128 (2014)

- **"Computer Simulations of Prebiotic Systems"**
  T. Fornaro and I. Carnimeo
  Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, DOI: 10.1016/B978-0-12-409547-2.11025-X (2014)

- **"Toward Feasible and Comprehensive Computational Protocol for Simulation of the Spectroscopic Properties of Large Molecular Systems: The Anharmonic Infrared Spectrum of Uracil in the Solid State by the Reduced Dimensionality/Hybrid VPT2 Approach"**
  T. Fornaro, I. Carnimeo and M. Biczysko
  The Journal of Physical Chemistry A, 119 (21), 5313-5326 (2015)

- **"Hydrogen-Bonding Effects on Infrared Spectra from Anharmonic Computations: Uracil-Water Complexes and Uracil Dimers"**
  T. Fornaro, D. Burini, M. Biczysko and V. Barone
  The Journal of Physical Chemistry A, 119 (18), 4224-4236 (2015)

- **"Toward the Design of Alkynylimidazole Fluorophores: Computational and Experimental Characterization of Spectroscopic Features in Solution and in Poly(methylmethacrylate)"**
  V. Barone, F. Bellina, M. Biczysko, J. Bloino, T. Fornaro, C. Latouche, M. Lessi, G. Marianetti, P. Minei, A. Panattoni, A. Pucci
  Phys. Chem. Chem. Phys., 17, 26710-26723 (2015)

Papers in preparation:

- **"Toward Improved Vibrational Frequencies of ν(C=O) and ν(N-H): the Nucleic Acid Bases Dimers"**
  T. Fornaro, M. Biczysko and V. Barone

- **"Inspecting the Role of Serpentinite-hosted Hydrothermal Minerals in Prebiotic Processes: Binding of Nucleic Acids Components to Brucite"**
  T. Fornaro, C. Feuillie, D. A. Sverjensky, R. M. Hazen, J. R. Brucato, M. Biczysko and V. Barone

Academic Education and Scientific Activities

Classes attended during the three-years Ph.D. program in Chemistry at the Scuola Normale Superiore of Pisa:

- "Molecular Spectroscopy"
- "Astrobiology"
- "Computational Modeling of Bio- and Nano-systems" integrated with the "Laboratory of numerical simulation for physico-chemical processes"
- "Astrochemistry and Molecular Astrophysics"

Internships:

- **National Institute for Astrophysics (INAF)-Astrophysical Observatory of Arcetri, Florence, Italy**
  Experimental laboratory activities in support of the PhD Thesis concerning IR spectroscopy measurements using the Diffuse Reflectance Infrared Fourier Transform Spectroscopy technique (DRIFTS).
  Supervisor: Dr. John Robert Brucato.

- **Carnegie Institution of Washington (CIW)-Geophysical Laboratory, Washington DC, USA**
  Laboratory activities in support of the PhD Thesis concerning studies of molecular binding mechanisms, stability and reactivity of nucleobases, nucleosides and nucleotides on brucite, under prebiotic conditions.
  Supervisors: Prof. Robert M. Hazen (Carnegie Institution of Washington) and Prof. Dimitri A. Sverjensky (Johns Hopkins University).

Conferences and Schools:

- V Workshop of the Italian Astrobiology Society Life in a Cosmic Context, Trieste (Italy), September 2015.
  Oral presentation: "Nucleic Acid Components adsorbed on mineral surfaces: A test bed for searching signs of life on Mars", T. Fornaro, J.R. Brucato, M. Biczysko, V. Barone.

- Astrobiology Science Conference 2015 Habitability, Habitable Worlds, and Life, Chicago, Illinois (USA), June 2015.
  Oral presentation: "Studies of nucleobases binding mechanisms to minerals, their stability and reactivity under prebiotic and space-like conditions", T. Fornaro, J. R. Brucato, M. Biczysko, V. Barone, C. Feuillie, D. Sverjensky and R. M. Hazen.

- Seminar at NASA Goddard Space Flight Center in Washington DC (USA), March 3, 2015, entitled: "Infrared spectral investigations of nucleobases adsorbed on mineral surfaces under space-like conditions".

- Winter Modeling 2014, Pisa (Italy), December 2014.
  Oral presentation: "The Anharmonic Infrared Spectrum of Uracil in the Solid State by Reduced Dimensionality/Hybrid VPT2 approach", T. Fornaro, I. Carnimeo.

- Astrobiology Graduate Conference 2014 (NASA Astrobiology Institute), Troy, New York (USA), July 2014.
  Oral presentation: "Infrared spectral investigations of UV irradiated nucleobases adsorbed on mineral surfaces", T. Fornaro, J.R. Brucato, M. Biczysko, V. Barone.

- NASA Astrobiology Research Focus Group Workshop 2014, Bolton's Landing, New York, USA.
  Development of a proposal, as co-PI with a group of other early-career astrobiologists, concerning the stability and reactivity of tholins under environmental conditions found on Titan's atmosphere.

- Seminar at the Geophysical Laboratory of the Carnegie Institution of Washington, Washington DC (USA), July 24, 2014, entitled: "Infrared spectral investigations of UV irradiated nucleobases adsorbed on mineral surfaces".

- International Symposium on Molecular Spectroscopy 2014, University of Illinois, Champaign-Urbana, Illinois (USA), June 2014.
  Oral presentation: "Anharmonic IR Spectra of Biomolecules: Nucleobases and their Oligomers", V. Barone, M. Biczysko, J. Bloino, I. Carnimeo, T. Fornaro.

- Joint Meeting of the Nordic Network of Astrobiology and the centre of Geobiology - Biosignatures across space and time, Bergen (Norway), May 2014.
  Poster presentation: "Infrared spectral investigations of UV irradiated nucleobases adsorbed on mineral surfaces", T. Fornaro, J.R. Brucato, M. Biczysko, V. Barone.

- AstroBigné seminar at the INAF - Arcetri Astrophysical Observatory, Florence (Italy), December 10, 2013, entitled: "Infrared spectral investigations of UV irradiated nucleobases adsorbed on mineral surfaces".

- Avogadro Colloquia II edition (Italian Chemical Society), Scuola Normale Superiore, Pisa (Italy), September 2013.
  Oral presentation: "Computational spectroscopy studies of nucleobases and their complexes", T. Fornaro, M. Biczysko, S. Monti, V. Barone.

- CODECS Summer School 2013 Theoretical Spectroscopy (COST), Geneva (Switzerland), August 2013.
  Poster presentation: "Computational spectroscopy studies of nucleobases and their complexes", T. Fornaro, M. Biczysko, S. Monti, V. Barone.

- XI Congresso Nazionale di Scienze Planetarie, Bormio (Italy), February 2013.
  Oral presentation: "Adsorption of Nucleic Acid Bases on Magnesium Oxide: hints of Life Detection on Mars", T. Fornaro, J.R. Brucato, S. Branciamore, A. Pucci.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. ERC-2012-AdG-320951-DREAMS and from Italian MIUR (under the project PON01-01078/8).
The author thanks the members of the DREAMSlab who have supported this work at the Scuola Normale Superiore and the high performance computer facilities of the DREAMS center (http://dreamshpc.sns.it) for providing computer resources.
The support of the COST CMTS-Action CM1002 "COnvergent Distributed Environment for Computational Spectroscopy (CODECS)" is also acknowledged.
Moreover, the author thanks Dr. John Robert Brucato of the INAF-Astrophysical Observatory of Arcetri, for mentoring the experimental part of the PhD program and, in particular, for providing the equipment for IR spectroscopic measurements.
The experimental studies performed at the INAF-Astrophysical Observatory of Arcetri received funding from the Italian Space Agency (ASI) under grant agreement No. I/060/10/0-ExoMars Science.
Furthermore, the Carnegie Institution of Washington, DC, USA, is acknowledged for sponsoring a short-term position as predoctoral associate at the Geophysical Laboratory. In particular, the author thanks Prof. Robert M. Hazen of the Carnegie Institution of Washington and Prof. Dimitri Sverjensky of the Johns Hopkins University for their supervision during this internship and for making available equipment of the Mineral Surfaces Laboratory for adsorption experiments.
The author acknowledges also Dr. Cecile Feuillie and Jihua Hao for supporting and giving key inputs to the work carried out in the Mineral Surfaces Laboratory, the acting director of the Geophysical Laboratory Dr. George D. Cody for fruitful discussions and for making available his laboratories, Dr. Dionysis Foustoukos, Dr. Timothy Strobel, John Armstrong and Paul Goldey, for their technical assistance in the hydrothermal treatments, XRD and SEM measurements, and Dr. Kenneth Livi of the Johns Hopkins University for the TEM measurements.