



HAL
open science

Influence of nanohydration on the structure and radiation-induced fragmentation of gas-phase biomolecules

Juliette Leroux

► **To cite this version:**

Juliette Leroux. Influence of nanohydration on the structure and radiation-induced fragmentation of gas-phase biomolecules. Atomic and Molecular Clusters [physics.atm-clus]. Normandie Université; Universiteitsmuseum (Groningue, Pays-Bas), 2024. English. NNT : 2024NORMC224 . tel-04822556

HAL Id: tel-04822556

<https://theses.hal.science/tel-04822556v1>

Submitted on 6 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité **PHYSIQUE**

Préparée au sein de l'**Université de Caen Normandie**

En cotutelle internationale avec l'**UNIVERSITE GRONINGEN PAYS-BAS, PAYS BAS**

Influence of nanohydration on the structure and radiation-induced fragmentation of gas-phase biomolecules

Présentée et soutenue par
JULIETTE LEROUX

Thèse soutenue le 19/11/2024
devant le jury composé de :

M. JEAN-YVES CHESNEL	Professeur des universités - Université de Caen Normandie (UCN)	Directeur de thèse
MME SADIA BARI	Professeur - GRONINGEN - UNIVERSITY OF GRONINGEN	Co-directeur de thèse
MME DEBORA SCUDERI	Professeur des universités - Université Paris Saclay	Président du jury
MME JULIE DOUADY	Maître de conférences - Université de Caen Normandie (UCN)	Membre du jury
M. RONNIE HOEKSTRA	Professeur - GRONINGEN - UNIVERSITY OF GRONINGEN	Membre du jury
M. MARC ROUSSEAU	Professeur des universités - Laboratoire de Physique corpusculaire de Caen	Membre du jury
M. CASPAR VAN DER WAL	Professeur - GRONINGEN - UNIVERSITY OF GRONINGEN	Membre du jury
MME MELANIE SCHNELL	Professeur - Christian-Albrechts-Universität zu Kiel	Rapporteur du jury

Thèse dirigée par **JEAN-YVES CHESNEL** (Centre de recherche sur les ions, les matériaux et la photonique (Caen)) et **SADIA BARI** (GRONINGEN - UNIVERSITY OF GRONINGEN)

**Influence of nanohydration on the
structure and radiation-induced
fragmentation of gas-phase
biomolecules**

Juliette Leroux

Zernike Institute for Advanced Materials PhD thesis number: 2024-023

ISSN: 1570-1530

Printed by: Imprimerie moderne de Bayeux

Cover by: Marine Leroux

This work was conducted at the Deutsches Elektronen-Synchrotron, DESY in Germany, and the CIMAP laboratory in France as part of a co-agreement between the REGION NORMANDIE and the Center for Molecular Water Science (CMWS) at DESY. The X-ray absorption experiments were performed at the PO4 beamline of the PETRA III synchrotron at DESY and the UE52_PGM Ion Trap beamline of the BESSY II synchrotron in Berlin, Germany.



university of
 groningen

faculty of science
 and engineering

zernike institute for
 advanced materials



UNIVERSITÉ
 CAEN
 NORMANDIE



CMWS Centre for Molecular
 Water Science



university of
 groningen



UNIVERSITÉ
 CAEN
 NORMANDIE

Influence of nanohydration on the structure and radiation-induced fragmentation of gas-phase biomolecules

PhD thesis

to obtain the degree of PhD at the
 University of Groningen
 on the authority of the
 Rector Magnificus of the University of Groningen
 Prof. J.M.A. Scherpen
 and in accordance with
 the decision by the College of Deans

and

to obtain the degree of PhD at the
 University of Caen Normandy
 on the authority of the
 Prof. L. Adoui
 and in accordance with
 the decision by the doctoral school ED 591 PSIME.

Double PhD degree

This thesis will be defended in public on

Tuesday 19 November 2024 at 14:30 hours

by

Juliette Sarah Pauline Leroux

born on 18 November 1997
 in Caen, France

Supervisors

Prof. S. Bari

Prof. JY. Chesnel

Assessment Committee

Prof. M. Schnell

Prof. D. Scuderi

Prof. M. Rousseau

Prof. R.A. Hoekstra

Contents

Introduction	1
1 Context	5
1.1 Biomolecules	5
1.1.1 Amino acids and peptides	6
1.1.2 Nucleic acids	8
1.2 Photoabsorption processes in biomolecules	9
1.2.1 Interaction between biomolecules and IR light	10
1.2.2 Electronic excitations of biomolecules by UV photons	13
1.2.3 Interactions between biomolecules and VUV or soft X-ray photons	15
1.3 Near-edge X-ray absorption spectroscopy	17
1.4 Influence of the hydration on biomolecular structure and fragmentation	19
2 Computational methods	23
2.1 Born-Oppenheimer approximation	23
2.2 Electronic Structure Methods	24
2.2.1 Quantum-chemical methods	25
2.2.2 Electron density-based methods	25
2.3 Density Functional Theory	26
2.4 Computations of ground state geometries and X-ray absorption spectra	29
2.4.1 Ground state: geometry and IR spectrum	29
2.4.2 Excited states: X-ray spectrum	31
3 Experimental methods	35
3.1 The Spektrolatius setup	35
3.1.1 Electrospray ionization	37
3.1.2 Mass filter	39
3.1.3 RF trap	41
3.1.4 Time-of-flight mass spectrometer	43

3.1.5	Acquisition cycle	44
3.2	The Ion Trap end-station	45
3.3	Synchrotron facilities	47
3.3.1	PETRA III	47
3.3.2	BESSY II	47
3.4	A new electrospray source dedicated to the production of hydrated biomolecules	48
3.4.1	Description of the hydration ESI source developed during this thesis	50
4	Influence of the number of water molecules on the conformation and oxygen K-edge absorption spectra of a deprotonated nucleotide	53
4.1	Introduction	53
4.2	Methods	55
4.3	Structures characterization	55
4.4	Calculated IR spectra of hydrated AMP ⁻	63
4.5	Calculated absorption spectra in the soft X-ray regime	67
4.6	Conclusion	71
5	Influence of a single water molecule on the X-ray absorption spectra of a phosphorylated amino acid	73
5.1	Introduction	73
5.2	Methods	75
5.2.1	Experimental method	75
5.2.2	Theoretical calculations	76
5.3	Results and discussion	78
5.3.1	Mass spectra	78
5.3.2	X-ray absorption spectra	83
5.3.3	Discussion of the calculated conformers	92
5.3.3.1	Singly hydrated conformers of protonated phosphotyrosine	92
5.3.3.2	Comparison of the calculated absorption spectra with the NEXAMS spectra	96
5.4	Conclusion	102
6	Electronic signatures of the protonation sites of model peptides	103
6.1	Abstract	103
6.2	Introduction	103

6.3	Methods	106
6.3.1	Experimental method	106
6.3.2	Theoretical calculations	108
6.4	Results and Discussion	111
6.4.1	Photodissociation mass spectra analysis	111
6.4.2	Different protonation sites on amino groups in peptides	117
6.4.2.1	[GGGGG+H] ⁺	117
6.4.2.2	[PGGGG+H] ⁺	121
6.4.2.3	[GGGGH+H] ⁺	123
6.4.2.4	[GGGGR+H] ⁺	126
6.4.2.5	[GGGGK+H] ⁺	130
6.4.2.6	Conclusion	132
6.4.3	Protonation at the amide oxygen: The case of triglycine	135
6.4.3.1	Comparison of the experimental spectra of G ₅ and G ₃	137
6.4.3.2	Theory/experiment comparison at the N K-edge	140
6.4.3.3	Theory/experiment comparison at the O K-edge	147
6.4.3.4	Conclusion	150
6.5	Conclusions	151
7	Summary	153
8	Samenvatting	159
9	Résumé en français	165
9.1	Introduction	165
9.2	Contexte	167
9.2.1	Molécules d'intérêt biologique	167
9.2.2	Spectroscopie d'absorption induite par rayons X	169
9.3	Méthodes	171
9.3.1	Calculs de Chimie Quantique	171
9.3.2	Spectrométrie de masse	176
9.3.3	Une nouvelle source electrospray pour la production de biomolécules hydratées	177
9.3.3.1	Description de la source electrospray développée du- rant cette thèse	179
9.4	Résultats	182

9.4.1	Influence du nombre de molécules d'eau sur les propriétés thermodynamiques et spectroscopiques d'un nucléotide déprotoné: Une étude théorique	182
9.4.2	Influence d'une seule molécule d'eau sur le spectre d'absorption de rayon X d'un acide aminé protoné	185
9.4.3	Signatures électroniques des sites de protonation de peptides modèles	188
9.5	Conclusion et perspectives	190
	Bibliography	192
	Acknowledgements	219

Introduction

Water is essential to life. The presence of water on Earth provides a suitable environment for the initial form of life [1], [2]. Prebiotic molecules are hypothesized to have formed under physiochemical conditions through the interaction of water with carbon atoms. As a ubiquitous solvent, water possesses the essential properties to dissolve ions and establish a versatile environment for fundamental chemical reactions, such as the formation of amino acids and peptides [3].

Water is an integral part of biological matter, composing 70 % of a cell, and its presence in the molecular environment of the proteins impacts their assembly and dynamics. This is particularly important as proteins' functions are dictated by both their structure and the interplay with their surrounding environment [4]–[7]. For example, a certain threshold level of hydration is necessary to fully activate the dynamics and functionality of globular proteins, on average 40% of water [4], [8]. Moreover, water interacts through hydrogen bonds with the proteins or other biomolecules, which affects their three-dimensional structure and biological functionalities [9]. The water in the biological medium also affects protein-protein binding and the stability of local interactions such as salt bridges [10]. However, the mechanisms behind the solvation of biomolecules, such as the influence of the first solvation shell on the 3D structure or the dynamics of solvation, are not yet fully understood. This problem extends to the photophysical and photochemical properties of absorbing chromophores within proteins, which exhibit a high sensitivity to the local hydration environment [5].

Furthermore, it is not yet understood whether the solvation environment provides additional relaxation pathways to prevent radiation-induced damage to biomolecules or if it could be responsible for further fragmentation besides the creation of radical oxygen species. Therefore, the role of solvation effects on the fragmentation of biomolecules after ionization or excitation has been studied in bulk [11]–[13] and microsolvation environments [14]–[20]. However, it is still under discussion whether water enhances or reduces radiation damage in biomolecules.

Studying the interaction of water with biomolecules on a molecular level remains challenging in the liquid phase because of the complexity involved in disentangling the

solvent's effect from the intrinsic properties of the biomolecules. To investigate the first steps of solvation, the method of choice is to carry out experiments in the gas phase, as it allows removing any interaction between the biomolecule and the solvent. From this starting point, a bottom-up approach can be adopted by first studying the isolated system and then progressively adding water molecules in a stepwise manner until eventually approaching the bulk properties.

The development of the electrospray ionization source to introduce thermally fragile molecules into the gas phase by John Fenn [21] made it possible to design experimental studies at the molecular level focused on biomolecules such as proteins and DNA strands. In combination with mass spectrometry, the technique serves as a valuable tool for investigating the stepwise solvation of biomolecules. This approach enables the initial study of isolated systems, followed by the step-by-step investigation of the solvation process, thereby elucidating the solvent's influence on the conformation and intrinsic properties of biomolecules. To obtain a better understanding of the effect of water on the conformation, reactivity, and photo-processes of biomolecules, continuous research has been devoted to studying stepwise solvation theoretically and experimentally over the last decades [7], [15], [17], [22]–[25]. Theoretical methods allow the exploration of the conformational space of hydrated molecules for various sizes and hydration levels [7], [26]. Therefore, theory can be employed to provide physical interpretation to experiments, while experiments can serve as benchmarks for testing how appropriate the theoretical treatments of hydrated systems are. Zhan and Fenn developed an electrospray source dedicated to producing hydrated clusters [15], [27], where they showed the possibilities of introducing hydrated biomolecules with up to 40 or more water molecules, allowing to study stepwise solvation of biomolecules in the gas phase.

Furthermore, combining mass spectrometry with advanced light sources, such as tabletop lasers, synchrotron, or free-electron lasers, allows us to probe the radiation process and dynamics in the biomolecular target and, thus, understand the different properties of isolated or solvated biomolecules. Electromagnetic radiation serves as a versatile probe for elucidating the structure and radiation damage in molecular systems. The energy of the photons absorbed dictates the induced transitions in a molecular system, ranging from rotational and vibrational to electronic excitations in the molecule. Consequently, diverse relaxation pathways become observable. Sufficient energy can lead to molecular ionization and/or dissociation. While infrared (IR) light induces vibrational transitions within the molecule, resonant photoexcitation with ultraviolet (UV) light triggers electronic transitions, promoting valence electrons to unoccupied molecular orbitals. Additionally, the photo-absorption process in the soft X-ray range

allows us to selectively probe specific atoms within the biomolecule or its environment by exciting the electrons of the atomic core level.

In this context, measuring and analyzing the molecule's absorption spectra at different energies will provide extensive information regarding not only its geometric structure but also its electronic structure. However, for gaseous samples, direct absorption is challenged by the low density of biomolecules, which complicates direct measurements of the absorption of light in the sample. Therefore, one approach to address this issue is to perform so-called action spectroscopy, which involves measuring how the absorbed photon affects the molecule, often by monitoring fragmentation patterns using mass spectrometry. This method has been implemented in many studies for investigating isolated but also hydrated systems combining electrospray source and mass spectrometry with light sources [25], [28]–[30].

The work of this thesis is inscribed in this context and focuses on combining theoretical and experimental approaches to investigate the influence of the aqueous environment on the structure and fragmentation of gas-phase biomolecules. To investigate solvated biomolecules in the gas phase, we developed a dedicated hydration source based on the source of Prof. Nonose [24], which is based on the same concept as the one developed by Zhan and Fenn. Theoretical calculations have been performed on a deprotonated nucleotide to understand how the stepwise solvation with up to twenty water molecules affects the three-dimensional structure and absorption spectra. Moreover, combining both calculations and action-spectroscopy experiments in the soft X-ray range allowed us to decipher the 3D structure and electronic structure and elucidate the influence of a single water molecule or a proton on the properties of a protonated amino acid and protonated peptides.

The manuscript is divided into seven chapters.

In Chapter 1, we detail the context of this PhD thesis. The photoabsorption processes that occur in biomolecules upon irradiation with various photon energies, ranging from infrared (IR) to soft X-rays, are discussed. Additionally, it explores the current understanding of how hydration affects the properties of these biomolecules.

Chapter 2 explains the theoretical approach used to understand biomolecules' vibrations and electronic properties. It introduces the Density Functional Theory (DFT) computational method and explains how it is used to optimize molecular structures and to calculate properties of excited states.

In Chapter 3, we describe the experimental method. The two experimental setups and synchrotron beamlines used for near-edge X-ray absorption spectroscopy are de-

scribed. A focus is made on the tandem mass spectrometer Spektrolatius at DESY. The final section of this chapter provides a detailed description of the novel electrospray source developed during this thesis, which is specifically designed to generate hydrated biomolecules.

In Chapter 4, we investigate theoretically the influence of the number of water molecules on the structure of a deprotonated nucleotide using density functional theory calculations. We first investigate the influence of stepwise solvation on the 3D structure of the nucleotide. The second part of this study explores how these varying hydration levels affect the vibrational and electronic properties of the deprotonated nucleotide. Chapter 5 presents the effect of a single water molecule on the fragmentation and (electronic) structure of a phosphorylated amino acid. The experimental X-ray spectra at the carbon and oxygen K-edge of a mixture of isolated and hydrated protonated phosphotyrosine were recorded and compared with DFT calculations of potential structures for both systems.

In Chapter 6, we discuss the influence of the protonation sites on the electronic transitions of protonated peptides at the nitrogen K-edge. First, the electronic transitions of different protonation sites at nitrogen atoms of model pentapeptides are explained. The second part of this chapter compares the X-ray absorption spectra of triglycine when protonated at the N-terminus or on one amide oxygen. This demonstrates the sensitivity of the experimental technique to the specific atoms involved in protonation. Chapter 7 summarizes the experimental and theoretical results and provides an outlook of this thesis.

1 | Context

1.1 Biomolecules

Biomolecules are fundamental components of every living organism, essential for various biological processes such as cell division or development. They regulate all the chemical reactions necessary for an organism's growth and survival. The four main types of biomolecules are carbohydrates, lipids, nucleic acids, and proteins. Carbohydrates serve as essential energy sources and are composed of four types of sugar units (monosaccharides, disaccharides, polysaccharides, and oligosaccharides). Lipids are vital for forming cell membranes, which separate the cells from the environment from one another, ensuring cellular integrity and functionality. Nucleic acids, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), play a crucial role in storing and transmitting genetic information. They encode the sequence of nucleotides that dictates the arrangement of amino acids in proteins, which are essential for life. Proteins have their structure and functions defined by the amino-acid sequence. For example, proteins such as elastin or keratin act as structural components of cells and complex organisms, and insulin acts as a hormone. Proteins are typically defined as chains of 50 or more amino acids, while shorter chains are referred to as peptides. Considering the size of the systems studied in this work (from a single amino acid to pentapeptides), the term peptide will be employed in the following. This section will provide a more detailed description of the peptides' and nucleic acids' nomenclatures and properties, which are relevant to this work.

1.1.1 Amino acids and peptides

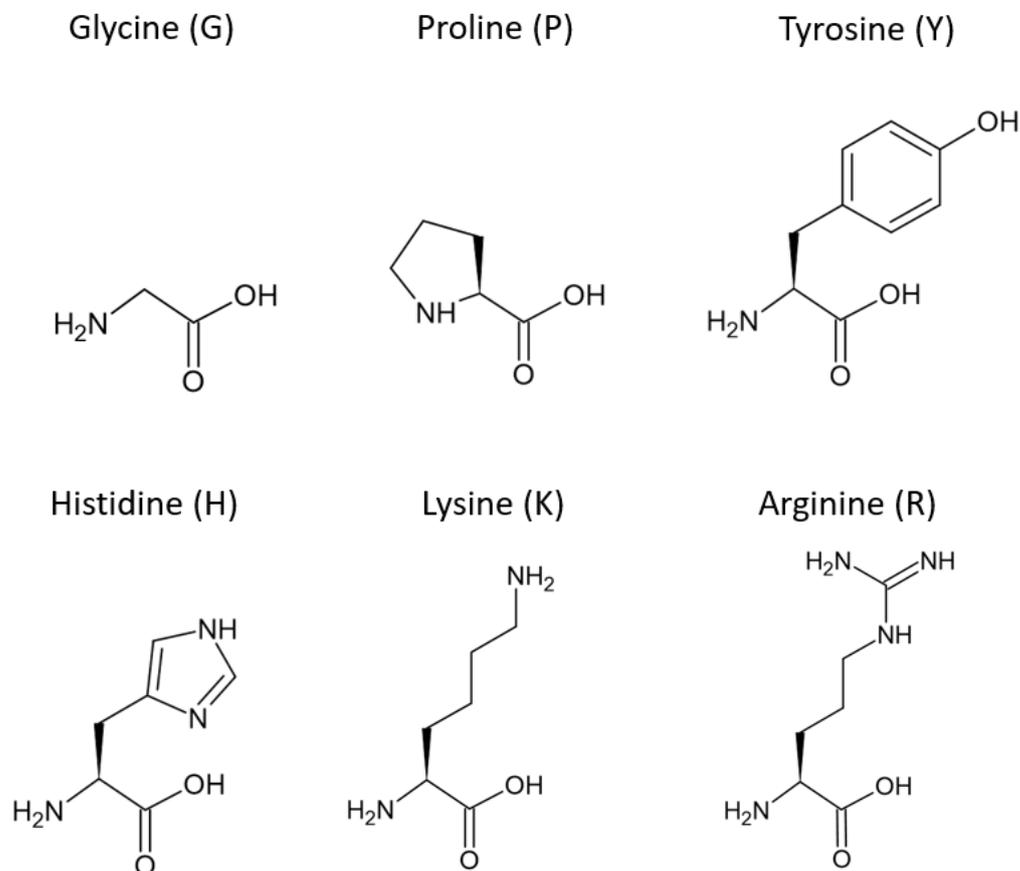


Figure 1.1.1: The chemical structure of the amino acids used in this PhD thesis. Glycine, proline, histidine, lysine, and arginine were used for the protonation site study in Chapter 6. Tyrosine was used in Chapter 5 to study the influence of a single water molecule on an amino acid.

Amino acids are the building blocks of peptides. They consist of an amino group (NH_2), a carboxyl group (COOH), and a side chain, all connected to the central α -carbon (C_α), Figure 1.1.1 shows the amino acids which are parts of the sequence of the peptides studied in the course of this thesis. The side chain is what differs from one amino acid to another. For example, for tyrosine, the side chain is a phenol group of formula $\text{C}_6\text{H}_5\text{O}$, whereas in the simplest amino acid, glycine, the side chain consists of only a hydrogen atom.

During the synthesis of peptides, the amino group of one amino acid reacts with the

carboxyl group of another amino acid. It forms a peptide bond under the release of H_2O , as shown in Figure 1.1.2. A peptide is formed when more α -amino acids react this way.

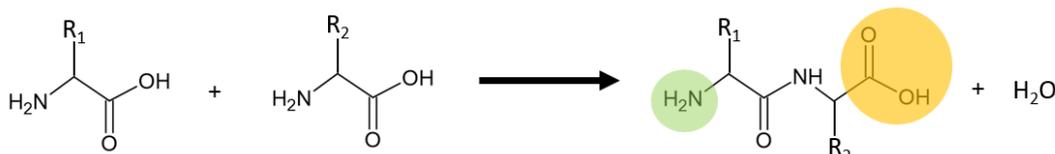


Figure 1.1.2: Peptide bond formation between two amino acids with a side chain R_1 and R_2 respectively. The N-terminus of the peptide is represented in the green circle, and the C-terminus is highlighted in the yellow circle.

Peptides are a diverse group of biomolecules that play crucial roles within living organisms. We can categorize them based on their origin and function. Neuropeptides, for example, are a specific type that influences learning, memory, and motivation [31]. Moreover, structural proteins, like mitochondrial cytochrome c oxidase, catalyze electron transfer and proton translocation reactions across membranes in living beings. Cytochrome c oxidase uses electrons to reduce molecular oxygen and thus produce water [32]. Therefore, it is essential to understand the fundamental properties of the building blocks of such proteins. Peptides can be formed from the standard 22 amino acids and modified versions. These modifications, known as post-translational modifications (PTMs), occur after a protein is initially formed and can significantly alter the peptide's function. PTMs involve covalent chemical changes, such as adding acetyl, phosphoryl, or methyl groups to specific amino acids. These modifications regulate various aspects of a protein's function, including its stability, activity, localization within the cell, and interactions with other molecules [33]. More than 400 PTMs have been described, with the most common being phosphorylation, acetylation, ubiquitination, and methylation [34]. Notably, phosphorylation can affect a protein's biological activities and stabilize its structure by inducing new covalent bonds within the proteins or disrupting protein-protein interactions [35].

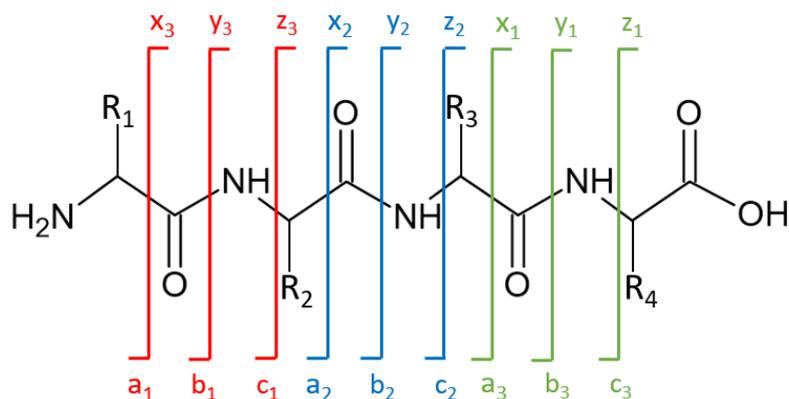


Figure 1.1.3: Nomenclature of the fragments of a peptide with four amino acids.

To interpret the results of experiments in which fragmentation of peptides occurs, a nomenclature system is used to name and identify the fragments originating from the different cleavage sites in the peptide backbone. This system was first introduced by Roepstorff and Fohlman [36] and then modified by Biemann [37]. Peptide fragment ions are indicated by a, b, or c if the charge is retained on the N-terminus side and by x, y, or z if the charge is located on a C-terminus side as shown in Figure 1.1.3. The subscript indicates where the cleavage is in the chain. Double backbone scissions generate internal fragments. Internal fragments of formula $\text{RCH}=\text{NH}_2^+$, originating from a single side chain by a combination of a- and y-type cleavages, are called immonium ions.

1.1.2 Nucleic acids

Nucleotides are the monomeric units of nucleic acids, including ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). A nucleotide comprises a phosphate group, a sugar group, and a nitrogen-containing base. The bases found in DNA are adenine (A), cytosine (C), guanine (G), and thymine (T). In RNA, uracil (U) replaces thymine. The chemical structure of the five bases of DNA and RNA and the structure of a nucleotide are shown in Figure 1.1.4. The main difference between DNA and RNA molecules is the order of their building blocks (nucleotides) and structural differences. RNA is composed of ribose sugar groups with A-U pairs, while DNA is composed of deoxyribose sugar groups with A-T pairs. RNA is composed of a single strand of nucleotides, while DNA is composed of two strands, forming a helical structure. Nucleotide sequences are important to life as they are involved in the cell division processes and carry the encoding of all biological molecules of a living organism [38].

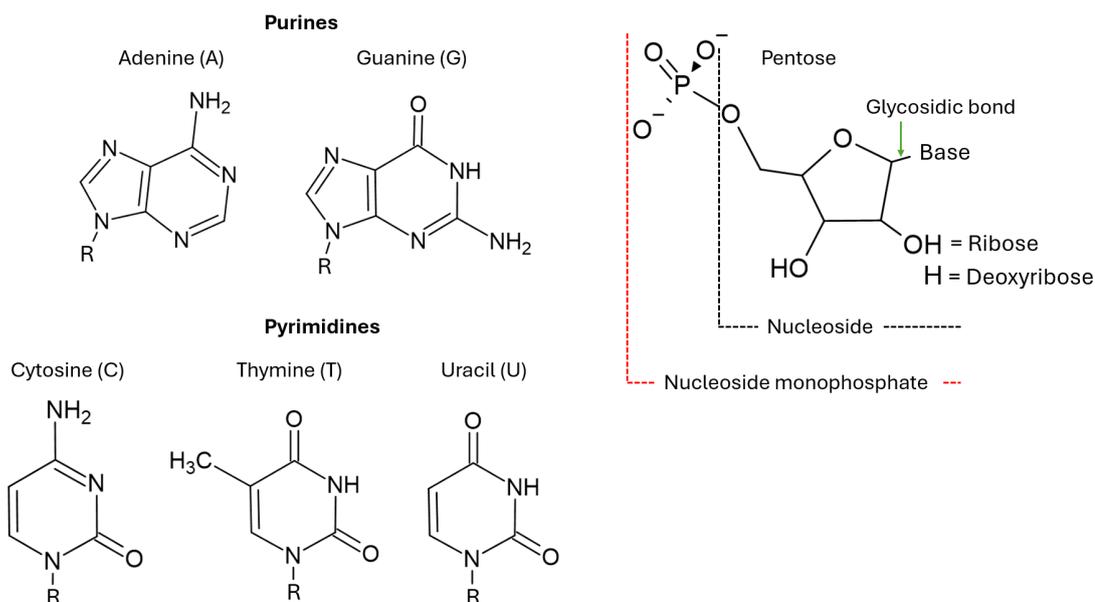


Figure 1.1.4: Representation of the five nucleobases, R represents the sugar in the nucleoside or nucleosides monophosphates.

Moreover, not only DNA and RNA have biological activity, but simpler units such as adenosine monophosphate (AMP), adenosine diphosphate (ADP), or adenosine triphosphate (ATP) have roles in the biological medium. ATP is an energy metabolite, and ADP and ATP also mediate the action of hormones [39], [40]. Jauker et al. showed that a liquid mixture of AMP with amino acid leads to the spontaneous formation of RNA strands and peptidyl RNA [41]. Chapter 4 studies the effect of stepwise solvation of deprotonated AMP on its structure and spectroscopic properties.

1.2 Photoabsorption processes in biomolecules

The electromagnetic spectrum is divided into radio waves, microwaves, infrared (IR) light, visible light, ultraviolet (UV) light, vacuum-UV (VUV), soft X-rays, hard X-rays, and gamma rays (see Table 1.1). Depending on the photon that is absorbed by the molecule, different kinds of transitions can be probed, ranging from rotational and vibrational to electronic transitions, and thus, different relaxation channels can be observed. If enough energy is given to the system, ionization of the molecule occurs. Measuring and analyzing the absorption spectra of a molecule at different energies will provide extensive information regarding, e.g., its geometric and electronic structures. One can determine the fraction of incident radiation absorbed at specific photon energies by measuring the transmitted radiation through a sample, whether solid, liquid,

or gaseous. However, for gaseous samples, direct absorption is challenged by the low density of ions ($< 10^8 \text{ cm}^{-3}$), which complicates direct absorption measurements [42]. One approach to address this issue and measure the "absorption" spectrum of diluted samples is to perform so-called action spectroscopy. This method involves measuring the action of the absorbed photon on the molecule, typically its fragmentation, but also electron or photoemission. As an indirect method, this technique allows the study of low-density samples.

Table 1.1: The different spectra domains of the electromagnetic spectrum according to [43].

Spectra domain	Photon energy	Wavelength
Radio waves	$< 12.4 \text{ meV}$	$> 1 \text{ cm}$
Microwaves	$0.08 - 1.24 \text{ meV}$	$1 \text{ mm} - 1.5 \text{ cm}$
IR	$1.24 \text{ meV} - 1.6 \text{ eV}$	$760 \text{ nm} - 1 \text{ mm}$
Visible	$1.6 - 3.3 \text{ eV}$	$380 - 760 \text{ nm}$
UV	$3.1 - 12.4 \text{ eV}$	$100 - 400 \text{ nm}$
VUV	$6.2 - 124 \text{ eV}$	$10 - 200 \text{ nm}$
Soft X-rays	$124 \text{ eV} - 12.4 \text{ keV}$	$0.1-10 \text{ nm}$
Hard X-rays	$12.4 \text{ keV} - 1.24 \text{ MeV}$	$0.001 - 0.1 \text{ nm}$
Gamma rays	$> 1.24 \text{ MeV}$	$< 0.001 \text{ nm}$

This section explores the photoabsorption and subsequent fragmentation processes observed in peptides and nucleic acids upon irradiation with photons from the infrared (IR) to the soft X-ray regime.

1.2.1 Interaction between biomolecules and IR light

When a molecule absorbs an IR photon, its vibrational states are excited, causing the molecule to heat up. Consequently, the molecule may dissociate, dissipate the energy into the molecular environment, or undergo radiative relaxation and emit IR photons, whose energies depend on the excited vibrational modes. IR spectroscopy operates on the principle of resonant excitation of characteristic vibrational modes of chemical bonds through photoabsorption.

As a fundamental technique for molecular structure characterization, IR spectroscopy relies on the vibrations of a molecule's atoms. Three methods are commonly used

to measure the IR action spectrum of gas-phase samples: Infrared multiple photon dissociation (IRMPD), IR dip-ultraviolet photodissociation, and the "tagging" (or "messenger") technique. IRMPD requires the absorption of multiple photons by the molecule until the molecule possesses enough vibrational energy to fragment, as represented in Figure 1.2.1. In brief, the energy deposited upon absorption of a photon at the fundamental vibrational frequency is dissipated into the bath of vibrational degrees of freedom by intramolecular vibrational redistribution (IVR). This leads to an increase in the molecular internal energy and the possibility of re-absorbing at the fundamental frequency. Once the dissociation threshold of the molecule is exceeded, it fragments. The fragmentation yield is then recorded as a function of the photon energy [42].

The IR dip-UV photodissociation has been developed by Stearns et al. [44]. In the IR dip-UV technique, a UV laser is first set on a particular feature in the electronic spectrum while scanning the wavenumber with an IR laser preceding the UV absorption. IR absorption removes the population from the ground state; this is detected as a depletion in the UV-induced photo fragmentation. With this technique, conformer-specific IR spectra can be obtained.

Last, the "tagging" method lowers the photodissociation threshold to the absorption of a single IR photon [45] by attaching to the target molecule a weakly bonded messenger atom or molecule, such as H₂ [45], He [46], or Argon [47]. The loss of the messenger is then recorded as a function of the photon energy, providing the IR spectrum of the biomolecule.

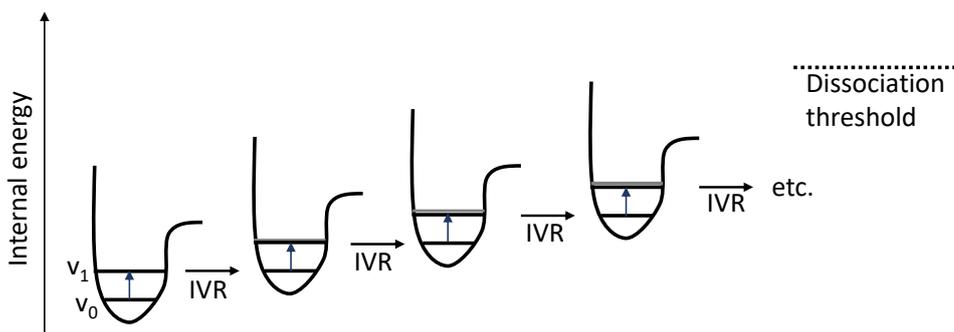


Figure 1.2.1: Schematic representation of the mechanism of IRMPD. Photon absorption at a fundamental vibrational mode v_0 is dissipated by IVR, leading to an increase in the internal energy of the ion and the possibility of re-absorbing. At higher internal energies, the density of states of the ion is also increased (indicated by a broadening of the v_1 linewidth). Once the dissociation threshold of the ion is exceeded, the molecule fragments.

IR spectroscopy can also be used to solve the three-dimensional structure of peptides by comparing the experimental results to calculated spectra of optimized geometries obtained, for example, by molecular dynamics or density functional theory calculations [48]–[53]. Lucas et al. investigated the IR signature of the protonation site of a dialanine peptide [49]. They found that the favored protonation site is the terminal amino group with the adjacent amide bond in the regular trans conformation. The same group also investigated the protonation site of N-acetyl-alanine and alanyl-histidine via IRMPD. They discovered that for N-acetyl-alanine, the acetyl oxygen acts as the protonation site, whereas the side chain imidazole nitrogen is protonated in the case of alanyl-histidine [54].

IRMPD spectroscopy can also be used to map modifications in the amino-acid sequence; Correia et al. investigated phosphorylated amino acids and identified the vibrational signature of a phosphorylated amino acid [51]. They claimed that IRMPD is a valuable tool for identifying phosphorylation sites (replacement of an alcohol group with a phosphate group in an amino acid) in bigger systems as there is no spectral congestion in the $900 - 1300 \text{ cm}^{-1}$ region where the P-O and P=O bands appear [51]. Moreover, IR spectroscopy measurements have been performed to decipher the environment of biomolecules [55]–[58]; Demireva et al. studied diprotonated 1,7-diaminoheptane hydrated with up to thirty water molecules [55]. They demonstrated that the 3D structure of the biomolecules becomes folded when more than fourteen water

molecules are added, and they tracked the folding process through the OH stretch region of the IR spectrum (2600 - 4000 cm^{-1}) [55].

In nucleic acids, IR-UV spectroscopy has been used to record the mid-IR spectra of guanine, guanosine, and deoxyguanine [52]. It was shown that the keto hydrogen-bonded structure is the lowest energy conformer in the gas phase for guanosine and deoxyguanine, whereas the keto form is the most stable structure of 9-ethyl guanine. Moreover, Salpin et al. recorded the IRMPD spectra of uracil, thymine, and cytosine in the mid-IR region [53]. They observed that the lowest energy structure is the enol form for all three structures, with a small fraction of the oxo tautomer in their spectra. They suggested a facile tautomerization, from keto to enol, of the neutral or protonated nucleobases in a protic solvent [53].

Thus, IR spectroscopy is a valuable tool for investigating three-dimensional structures of biomolecules in the absence of solvent. IR spectroscopy can also provide information on the secondary structure of proteins, mainly thanks to the C=O stretching vibration, which is sensitive to hydrogen bonding [52], [57]. However, no information on the electronically excited states of a molecule can be probed using IR light. This can be studied using UV radiation.

1.2.2 Electronic excitations of biomolecules by UV photons

While interaction with IR light causes molecules to undergo vibrational transitions (as explained in the previous section), resonant photoexcitation with UV light (200 - 400 nm) causes the molecules to undergo electronic transitions, leading to the promotion of valence electrons to empty molecular orbitals. This brings the molecule from its ground state to an excited state, as represented in Figure 1.2.2. In the UV range (> 200 nm), transitions typically involve the bonding orbital π and the nonbonding n. Specifically, transitions occur from π to π^* and from n to π^* . The σ to σ^* electronic transitions require more energetic photons (< 120 nm) and can thus be probed in the VUV regime. In contrast to IR photons, a single UV photon can provide enough energy to dissociate a molecule, thus allowing direct measurement of its fragments.

For nucleotides and nucleobases, the electronic transitions from the ground state to $\pi\pi^*$ states are responsible for the characteristic absorption band around 260 nm. In oligonucleotides, there is an ultrafast decay of the $\pi\pi^*$ state due to the existence of a conical intersection allowing a rapid return to the ground state [59], [60]. Moreover, it was shown that the protonation state of the oligonucleotide adenosine monophosphate AMP influences the deexcitation process upon absorption of 266 nm photons [61]. Deprotonated AMP relaxes to its ground state through internal conversion (IC) and

intramolecular vibrational energy redistribution (IVR). In contrast, protonated AMP dissociates before IVR via a nonstatistical process [61]. In amino acids, only the ones possessing an aromatic side chain absorb UV light at around 260 nm [62]. For example, tyrosine can absorb at 275 nm through electronic transitions from the ground state to a $\pi\pi^*$ state. However, in peptides, UV light can be absorbed in two ways: excitation through localized chromophore (tyrosine, tryptophan, and phenylalanine) at 260 nm or excitation through the peptide bond chromophores throughout the molecules with absorption bands in the far-UV region at $\lambda < 210$ nm [63]. Excitation of a peptide with UV light generally results in its fragmentation. At 263 nm, the fragmentation of the peptides can be triggered by a hydrogen transfer in the $\pi\pi^*$ excited state followed by C-C or C-N bond cleavage; other dissociation channels typically result from vibrationally induced fragmentation in the ground state reached via internal conversion [64].

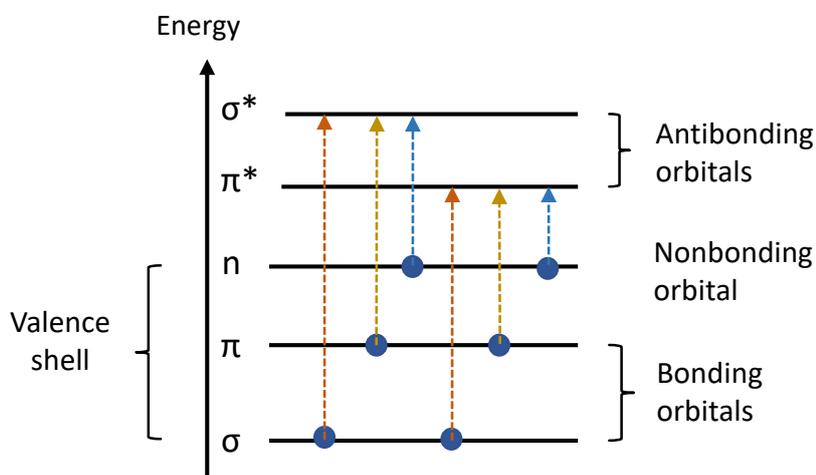


Figure 1.2.2: Schematic representation of the molecular transitions probed by the absorption of a UV photon.

1.2.3 Interactions between biomolecules and VUV or soft X-ray photons

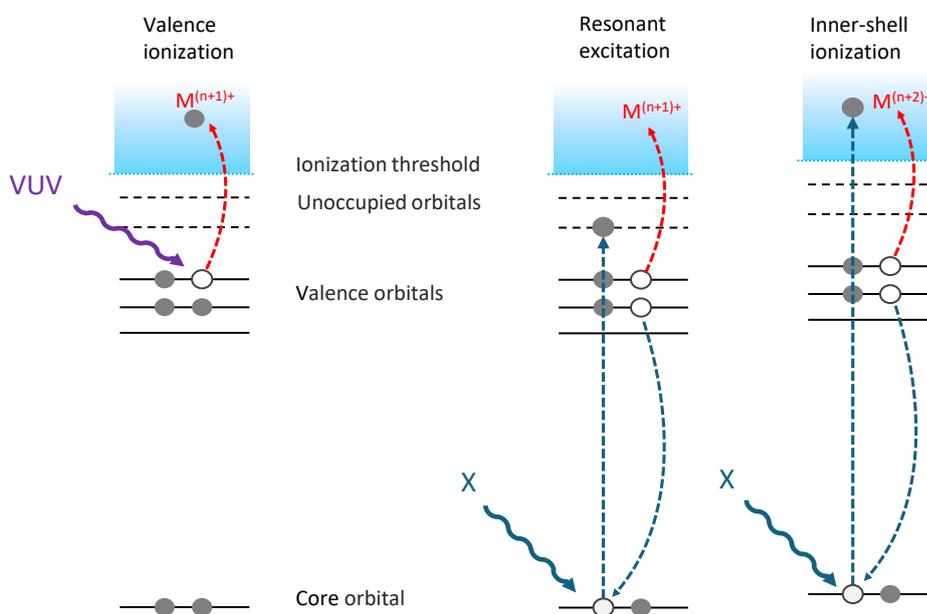


Figure 1.2.3: Schematic representation of the different processes after irradiation with VUV in the valence shell (left) or soft X-ray photons in the core shell (right). Core-shell excitation/ionization can be followed by Auger-Meitner decay.

If the photon energy increases beyond roughly 10 eV, photoionization becomes dominant. In the vacuum ultraviolet (VUV) region, this photoionization primarily ejects electrons from the outermost molecular orbitals (valence orbitals) of the biomolecule [65], whereas soft X-ray photons possess enough energy to excite or ionize the core electrons [28] (see Figure 1.2.3). In the VUV, the ejection of a valence electron leads to an unpaired valence electron or valence hole. Biomolecules with unpaired valence electrons, known as radicals, are highly reactive due to their unstable electronic configuration. This instability causes them to fragment through radical-induced dissociation pathways [66].

Schwell et al. investigated the ionization energies of small nucleic acid bases and amino acids [67]. They have found that their vertical ionization energy lies between 8.2 eV for Adenine and 9.6 eV for α -Amino-isobutyric acid. For nucleic acid bases, the produced cations do not fragment even when they have been produced with up to 5.35 eV of

excess energy, whereas the amino acid cations are unstable and rapidly fragment via a C-C(OOH) bond cleavage [67]. Moreover, intra-cluster peptide bond formation following VUV excitation (above 10 eV) has been observed in protonated serine dimers. [68]. This demonstrates that excited state dynamics in the VUV regime can lead to non-statistical pathways.

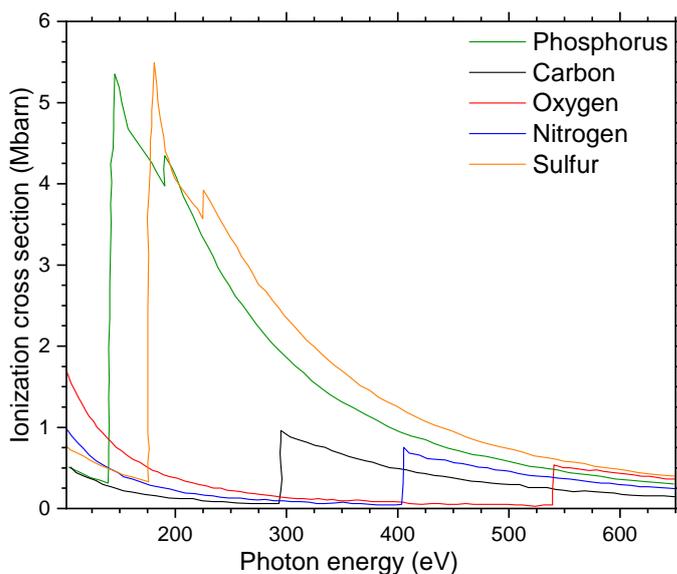


Figure 1.2.4: Atomic photoionization cross section for selected elements. Adapted from [69].

In the soft X-ray domain, the photon energy is tuned across an ionization edge to be element-specific, with photoabsorption being the dominant process. Figure 1.2.4 shows the ionization cross-section for some selected elements relevant to biomolecules. The photoabsorption creates a core hole that is filled either radiatively, emitting a fluorescent photon, or non-radiatively, emitting an Auger-Meitner electron as shown in the schematic Figure 1.2.3. The branching ratio between these two processes primarily depends on the atomic number of the probed element, with the lighter elements predominantly decaying via the Auger-Meitner effect. In the pre-edge region, resonant core excitation is the main process and is followed by Auger-Meitner decay, as depicted in Figure 1.2.3. As for IR and UV spectroscopy, this region is rich in structural information and is the focus of X-ray absorption spectroscopy studies (see next section). Above the edge, the absorption mechanism leaves the molecule in a core-hole

state with a departing photo-electron. Following this, the hole is filled, emitting an Auger-Meitner electron, thus leaving the system with two additional positive charges. It is also possible that more than one Auger-Meitner electron is emitted after photoionization [70]–[72].

Müller et al. showed that the interaction of a single photon at an energy of 288 eV with a ground-state C^{1+} ion results in the production of a C^{4+} ion following triple Auger-Meitner decay [72]. After Auger-Meitner decay, the biomolecular ion can still have enough internal energy to fragment further. For instance, a study by Egorov et al. revealed how the size of a peptide influences its response to soft X-ray radiation [73]. Larger peptides, around 10 kDa, primarily experience non-dissociative ionization or lose neutral atoms from their original structure.

In contrast, smaller peptides below 1 kDa rarely undergo non-dissociative ionization. Instead, their interaction with soft X-rays leads to fragmentation into sequence ions. Moreover, in the case of a protonated tri-oligonucleotide composed of fluorouracil, adenine, and guanine, Wang et al. observed soft X-ray-induced intramolecular hydrogen transfer [74]. They demonstrated that X-ray absorption unspecifically facilitates hydrogen transfer to the adenine moiety, but hydrogen transfer to the guanine is only observed when the photoabsorption does not occur in the guanine [74].

1.3 Near-edge X-ray absorption spectroscopy

X-ray absorption spectroscopy (XAS) has been widely used over the last decades to study the electronic and structural properties of biomolecular systems. XAS was first used in the 1920s for structural investigations of matter, where fine structures near the absorption edges were observed. Near-edge X-ray absorption fine structure (NEXAFS) investigates how X-ray photons are absorbed by exciting core electrons into unoccupied bound or continuum states. The NEXAFS presents prominent and unique characteristics in the energy range slightly below and extending up to approximately 50 eV beyond the absorption threshold [75]. The primary experimental techniques for recording NEXAFS spectra of solid-state samples involve electron yield and fluorescence yield detection. Both methods directly measure the formation of a core hole created by X-ray absorption, thereby quantifying the absorption cross-section [76]. Additionally, ion yield detection can be employed as an alternative to electron yield. While an indirect measurement of X-ray absorption, recording the total ion yield as a function of photon energy provides an approximation of the total absorption spectrum [75].

Pioneering work over two decades ago employed NEXAFS spectroscopy to analyze thin

films of amino acids and small peptides in the solid phase [77]. This technique was further utilized to measure the carbon, nitrogen, and oxygen K-edge NEXAFS spectra of the 20 common proteinogenic amino acids, along with simple glycine oligopeptides and various homopolypeptides [78]–[80]. The studies revealed that strong signals in amino acid spectra originate from electron transitions involving the C-terminal carboxyl group’s π^* orbitals, aromatic ring systems in aromatic side chains, and the carboxyl groups of acidic amino acid side chains. Interestingly, in the case of peptides, characteristic features emerge at the carbon, nitrogen, and oxygen K-edges due to electron transitions from 1s core orbitals to the π^* orbitals of the peptide bond [81], [82]. For nucleic acid bases and nucleotides, the $1s \rightarrow \pi^*$ transitions are particularly sensitive to the chemical environment of the nitrogen with oxygen substitution on ring carbon atoms resulting in a blue shift of the lowest $1s \rightarrow \pi^*$ transitions [83].

Studying biomolecules in the gas phase offers the advantage of investigating their intrinsic properties absent of environmental influences. By combining NEXAFS with mass spectrometry, one can study charged biomolecules in the gas phase and record dissociation yield and ionization yield as a function of the photon energy. The technique is also referred to as near-edge X-ray absorption mass spectrometry (NEXAMS) [84]–[86]. NEXAMS was first employed to elucidate dissociation processes in biomolecules, including non-dissociative ionization, backbone scission, or site-selective dissociation [85], [87]–[89], or to understand the structure of charged molecules and the influence of the environment on their structure [86], [90]–[93].

Egorov et al. studied the different reaction pathways of protonated melittin in the gas phase when interacting with soft X-ray radiation at the carbon, nitrogen, and oxygen K-edges [87]. They observed three main reaction channels for protonated melittin: non-dissociative ionization, loss of small neutral molecules from the protein, and formation of backbone scission ions. This study emphasized the capability of NEXAMS to assign specific X-ray absorption spectral features to individual ions observed in the mass spectrum. Moreover, in an earlier study, the same authors have shown that the dissociation processes after absorption of a soft X-ray photon depend on the size of the molecules [73]. Furthermore, Schwob et al. have shown that irradiating biomolecules in the gas phase with soft X-rays can induce site-selective dissociation, as evidenced by their investigation of protonated methionine enkephalin at the sulfur L-edge [85]. Their work revealed that two competing processes could occur following the excitation of a sulfur 2p electron: site-selective dissociation and statistical fragmentation mediated by intramolecular vibrational energy redistribution.

However, NEXAMS can also be used to probe the structure of biomolecules and to study how factors like protons or their environment influence their electronic prop-

erties [86], [91], [93]. Buckley and Besley utilized theoretical calculations to explore the sensitivity of NEXAFS to protein secondary structures (α -helix, β -sheet, turn) [91]. They studied a protein called Agitoxin 2 and computed its NEXAFS spectra for each secondary structure at the carbon, nitrogen, and oxygen K-edges. While minor variations were observed at the carbon and oxygen edges, the nitrogen K-edge showed significant differences between the α -helix and the β -sheet, particularly in the $\pi^*(\text{C}=\text{O})$ excitations. This highlights the sensitivity of NEXAMS to specific structural elements within proteins.

Another example comes from the work of Mattioli et al., who investigated the influence of protonation on the molecular structure of Adenosine triphosphate (ATP) [93]. They combined NEXAMS (focusing on carbon and nitrogen K-edges) with DFT calculations. They detected subtle changes in the electronic structure of ATP upon protonation, compared to its neutral form, using inner-shell C and N photoexcitation. Moreover, they demonstrated that the folded structure of ATP observed in solution is maintained even in gas-phase clusters, stabilized by a few water molecules.

In conclusion, X-ray photoabsorption is a powerful tool for studying molecular structure, fragmentation, and dynamics. Combining near-edge X-ray absorption fine structure with mass spectrometry can be employed to elucidate the structure and dissociation processes in gas-phase charged biomolecules. Additionally, being element-specific, this technique can be used to investigate the influence of specific atoms or surrounding environments on the molecular structure.

1.4 Influence of the hydration on biomolecular structure and fragmentation

The aqueous environment of the biomolecules plays an important role in their structure and functionality. As explained by Helms in his review [4], a certain threshold level of hydration is necessary to fully activate the dynamics and functionality of globular proteins, i.e., 0.40 grams of water per gram of protein. Water interacts through hydrogen bonds with the proteins or other biomolecules, which determine their three-dimensional structure and biological functionalities [9]. Moreover, it also affects protein-protein binding as well as the stability of local interactions such as salt bridges [10]. Using UV radiation from a synchrotron, Milosavljević et al. demonstrated that three water molecules stabilize a protonated dimer of a peptide by around 1.5 eV [19]. As shown by Ahmad et al., the interfacial water within the gap between

two hydrophilic proteins forms a cohesive hydrogen bond network [94]. This network acts to stabilize early folding intermediates by facilitating interactions before the formation of native contacts. This collective hydrogen bond network observed at the protein-water interface has also been implicated in the folding of antifreeze proteins [95]. However, studying the interaction of water with biomolecules in the liquid phase remains challenging on a molecular level because of the infinite number of degrees of freedom of the solvent.

To obtain a better understanding of the effect of water on the conformation and reactivity of biomolecules, continuous research has been devoted to studying stepwise solvation theoretically and experimentally over the last decades [7], [15], [16], [19], [22], [23], [25], [55], [58], [96]–[99]. Density functional theory (DFT) and Møller-Plesset perturbation theory (MP2) are the preferred theoretical methods for small systems such as single amino acids. These two methods give results that agree with the experiments, and their relatively cheap computational cost allows us to perform systematic conformational studies for systems such as amino acids surrounded by up to 10 water molecules. For instance, Aikens et al. studied the hydration of neutral glycine - a system extensively studied theoretically over the last ten years - using MP2 and DFT methods [97]. They found several global minima for the different cluster sizes and observed changes in the structure of the amino acids after adding three to five water molecules, depending on the level of theory used to compute the structures. However, these methods require manually adding the water molecules to the system, which includes human bias, and it becomes more difficult to sample all the different possibilities as the number of water molecules in the system increases, even in systems as small as single amino acids. Kim et al. showed that conventional approaches with manual positioning of water molecules are limited as the number of possible hydrogen bond (HB) sites increases [100]. DFT and MP2 limitations rely on molecular energetics to find local minima. Thus, they overlook the structural characteristics of the given system. To differentiate the patterns of HB networks and go over the limitations of the two other methods, Kim et al. developed a method based on the Monte-Carlo Basin hopping method combined with a molecular graph method [100]. With this method, they found new global minima for glycine hydrated with ten water molecules in both neutral and zwitterionic states and the minimum number of water to stabilize to the zwitterionic form of tyrosine (eleven water molecules).

Although the details we can obtain for bigger systems are limited with conventional methods, they allow us to investigate systems with several dozen water molecules with good accuracy and determine different properties of the interaction of water with biomolecules, such as the preferred positions of the water molecules or thermodynam-

ical properties [7], [10], [101]. Chutia et al. investigated the influence of water on the structure of Ac-Ala_{5,8}-LysH⁺ peptides depending on whether the peptide is in its helical form or not [7]. Using molecular dynamics (MD) and DFT methods, they determined the preferred sites of hydration of these peptides. They discovered that water molecules cannot be inserted within the helical structure, indicating that a helical peptide is stable against solvent/water disruption. Additionally, they observed a transition from the protonated to the deprotonated state of the carboxyl group when the peptide was surrounded by 152 water molecules.

With the development of the electrospray ionization source to introduce fragile molecules into the gas phase by John Fenn [21], it has been possible to design experimental studies to investigate stepwise solvation and gain an understanding of the role of water on biomolecular conformation and properties. Zhan and Fenn developed an electrospray source dedicated to producing hydrated clusters where they demonstrated the possibilities of introducing hydrated biomolecules with up to 40 or more water molecules [15]. This paved the way for thermodynamics studies combining electrospray ionization with a drift tube and a mass spectrometer, which Gao et al. used to study the hydration effect on the thermodynamic properties of protonated amino acids [102], [103]. They found that the ammonium and the carboxyl group are the preferred binding sites of amino acids, with the first water molecule binding more strongly than the next three. Liu et al. investigated various sizes of peptides, from protonated dialanine to neurotensin (13 amino acids), with up to twenty water molecules [99]. They observed that the water binding energies are in the order of 7 to 15 kcal.mol⁻¹ and that the hydration level depends on the peptide's charge state, yielding a ratio of five to seven water molecules per unit charge. However, this finding contradicts the study by Nagornova et al., who recorded the infrared (IR) spectra of diprotonated gramicidin S with up to 50 water molecules [25]. Furthermore, Silveira et al. performed ion mobility combined with mass spectrometry (IMS-MS), where they studied protonated water clusters and two different peptides [104]. With their IMS-MS study, they could reproduce the magic number of protonated water clusters (cluster which presents a more stable structure than the other cluster from the series) [105], [106]. They also observed magic numbers in diprotonated gramicidin S (8, 11, and 14 water molecules) but could still observe higher hydration levels. However, they did not observe any magic number in protonated bradykinin. This illustrates that hydration is not yet fully understood, and the existence of a magic number probably depends on the peptides and their structures.

In order to understand how or where the water binds to the biomolecules, mass spectroscopy has been widely used, especially when combined with UV and IR lasers,

to this end, specific experimental instruments have been developed to perform spectroscopy [25], [29], [56], [58], [107]. Infrared spectroscopy of hydrated clusters can follow the changes in structure during stepwise solvation. Fischer et al. showed that the lack of free NH or OH stretches for glycine and alanine hydrated with four water molecules is characteristic of a completed first solvation shell [16]. Using collision-assisted stripping with IR spectroscopy, Hirata et al. showed that adding three water molecules to protonated benzocaine changes the protonation site from the carboxyl group to the nitrogen atom [58].

Furthermore, water can influence the structure of a biomolecule and its dissociation processes. For instance, Liu et al. studied the dissociation of a deprotonated hydrated nucleotide, adenosine monophosphate (AMP), after collision with neutral atoms [14]. They demonstrated that for less than ten water molecules attached to AMP, the dominant fragmentation channels are the loss of adenine and the loss of the phosphate group. However, the loss of water becomes the dominating fragmentation process, and the AMP anion is fully protected from fragmentation when more than thirteen water molecules are added to the biomolecule. They concluded that water protects the deprotonated nucleotide from fragmentation with neutral atoms [14].

However, Milosavljević et al. pointed out a different behavior after hydration of a diprotonated peptide interacting with soft X-ray radiation at the oxygen K-edge [17]. They studied substance P ions, bare and solvated with four and eleven water molecules, and observed that the hydration of substance P enhances its fragmentation efficiency in the range of 534 to 538 eV [17].

Water is undeniably critical for protein function, influencing their conformation, reactivity, and dissociation pathways. Despite significant theoretical and experimental efforts, a complete understanding of how water modulates protein structure and dissociation/deexcitation processes remains elusive. To bridge this knowledge gap and explore the interplay between water and biomolecules at synchrotron and FEL facilities, we have designed and built a novel electrospray source. This innovative source enables the production of hydrated (de)protonated biomolecules, providing a tool to investigate the influence of water on protein structure and fragmentation dynamics.

2 | Computational methods

Theoretical calculations can provide information on a system's molecular geometric and electronic structures and physicochemical properties. For instance, global and local energy minima of a given molecular structure can be investigated using molecular dynamics or electronic structure calculations such as Density Function Theory (DFT). The latter methods are based on solving the Schrödinger equation and the use of the Born-Oppenheimer approximation. Once candidate geometries have been found, their vibrational spectra can be calculated within the harmonic oscillation model. Furthermore, their electronically excited states can be described using time-dependent DFT or by combining DFT with restricted open-shell configuration interaction with singles (ROCIS).

This chapter introduces the Born-Oppenheimer approximation and the concept of DFT.

2.1 Born-Oppenheimer approximation

In molecular physics, the Born-Oppenheimer (BO) approximation is most commonly used to solve the Schrödinger equation. This approximation is based on the large mass difference between the nuclei and the electrons and, correspondingly, the time scale of their motion. We can, then, consider that the electrons move in the field of fixed nuclei and adapt instantaneously and adiabatically to the displacement of the nuclei while remaining in their electronic ground state. In this context, the total wave function for the electronic ground state, $\Psi_{tot}(\vec{r}_i, \vec{R}_a)$, reads as a product:

$$\Psi_{tot}(\vec{r}_i, \vec{R}_a) = \Psi_e(\vec{r}_i, \vec{R}_a)\chi_n(\vec{R}_a), \quad (2.1.1)$$

where $\Psi_e(\vec{r}_i, \vec{R}_a)$ is the electronic wave function of the ground state, which depends parametrically on the electron and nuclear coordinates (\vec{r}_i and \vec{R}_a , respectively), and $\chi_n(\vec{R}_a)$ is the nuclear wave function.

With the BO approximation, the total Hamiltonian is divided into nuclear and

electronic terms:

$$\hat{H}_{tot} = \hat{T}_n + \hat{H}_e, \quad (2.1.2)$$

where \hat{T}_n is the kinetic energy operator of the nuclei and \hat{H}_e is the electronic Hamiltonian without considering the nuclear kinetic energy, defined as the sum of three operators:

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne}, \quad (2.1.3)$$

\hat{T}_e being the kinetic energy operator, \hat{V}_{ee} the electron-electron repulsion energy operator and $\hat{V}_{ne} = \sum_{i=1}^N v_{ext}(\vec{r}_i)$ the electron-nuclei attraction energy operator. In atomic units and for a system composed of N electrons and M nuclei, these operators are defined as

$$\hat{T}_e = - \sum_{i=1}^N \frac{\nabla_i^2}{2}; \hat{V}_{ee} = \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|\vec{r}_i - \vec{r}_j|}; \hat{V}_{ne} = - \sum_{a=1}^M \sum_{i=1}^N \frac{Z_a}{|\vec{R}_a - \vec{r}_i|},$$

where Z_a and \vec{R}_a are the atomic number and position vector of the nucleus a and \vec{r}_i is the position vector of the electron i . ∇_i^2 is the Laplace operator and is expressed using the cartesian coordinates as $\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$.

In this context, the major computational work is solving the electronic time-independent Schrödinger equation for a given set of nuclear coordinates or geometry:

$$\hat{H}_e |\Psi_e\rangle = E_e |\Psi_e\rangle, \quad (2.1.4)$$

E_e being the electronic ground state energy, $|\Psi_e\rangle$ the ket associated to the electronic wave function $\Psi_e(\vec{r}_i, \vec{R}_a)$. It is assumed that the nuclei are moving within the potential energy surface (PES), which can be obtained by solving the previous equation.

2.2 Electronic Structure Methods

From the above section, we have shown that the electronic ground state energy eigenvalue E_e is based on obtaining the electronic wave function and resolving the integro-differential problem, as defined in Equation 2.1.4. Depending on the approximations and the type of properties investigated, different approaches based on electronic methods exist.

2.2.1 Quantum-chemical methods

Wave function-based methods comprise the Hartree-Fock (HF) and Post-HF methods. The HF method reduces a multi-particle problem into a one-particle by a single determinant. In this representation, one particle (electron in this case) is interacting with the rest of the system as an averaged (mean) field. The *a posteriori* based-on HF methods aim to recover electronic correlation using the Slater determinant of the resulting lowest HF energy as an initial guess determinant. Those are the post-HF methods, and they recover electronic correlation employing different approaches such as Møller-Plesset (MP) methods [108], Configuration Interaction (CI) methods [109], and Coupled Cluster (CC) methods [110], [111].

In MP methods, an additional term is added over the original Hamiltonian, which is an external perturbation over the potential contribution in a power series term. This expansion yields orders of perturbation over the initial Hamiltonian, the most relevant one expressing doubly excited Slater determinants being the second-order Møller-Plesset (MP2) method.

From the methods based on the CI theory, the wave function is described as a linear combination of Slater determinants, whereby considering all the possible configuration state functions of a system (full CI) to be the exact solution for the Schrödinger equation [112]. Last, all methods based on the CC theory use an exponential Ansatz operator to build a multi-electron determinant from the obtained one from the HF solution as an initial guess.

In CC methods, cluster operators are truncated to get different configurations, including single, double, up to n -th excitations, n being the total number of electrons involved in the system. In terms of accuracy, the CC with full treatment of excitations singles, doubles, and perturbative triples (CCSD(T)) is considered the gold standard. However, truncation over the determinants to be evaluated is required (such as in CC theory) for computational feasibility purposes. In the case of CI theory, the common methods include single excitations (CIS) and double excitations (CISD).

In both theories (CC and CI), configurations of several determinants imply the inclusion of a further representation of electron occupation in different spin-orbitals space, which is fundamental for understanding the vertical photoabsorption process.

2.2.2 Electron density-based methods

Electron density-based methods (Density Functional Theory (DFT)) are based on calculating the electron density rather than the electronic wave function. These methods

transform the problem of electrons interacting and evolving in a nuclear potential to a problem of independent electrons evolving in an effective potential, which assumes that the energy could be written as a functional of the electronic density.

The choice of one or another computational method is a compromise between the system size, the computational costs, and the level of accuracy needed for a property target. The computational cost of solving, e.g., CCSD equations, has been shown to grow up to the power of six with respect to the molecular size. This implies that a combination of other alternatives is required for the size of the systems studied in this thesis. This is the case of the electron density-based methods that recover correlation energy, and the computational cost is orders of magnitude lower than post-HF methods. Therefore, the density functional theory (DFT) method is an interesting alternative and is the most commonly used method for quantum mechanical calculations of the electronic structure of medium-sized systems [113]–[118]. For these reasons, we have chosen to compute the energy ground state at the DFT level. In the next section, the basics of DFT will be explained.

2.3 Density Functional Theory

The DFT approach uses functionals of the spatial-dependent electron density, which describes the ground state energy of the system. This makes the computation of the many-electron systems computationally possible. The electron density can be written as

$$\rho(\vec{r}) = N \int \cdots \int \Psi^*(\vec{r}, \vec{r}_2, \dots, \vec{r}_N) \Psi(\vec{r}, \vec{r}_2, \dots, \vec{r}_N) d\vec{r}_2 \dots d\vec{r}_N. \quad (2.3.1)$$

For an electronic system described by the Hamiltonian 2.1.3 above, the energy and the ground state wave function are determined by minimizing the functional energy $E[\Psi]$. DFT is based on the two Hohenberg-Kohn (HK) theorems [119]. The first HK theorem shows that for a many-electron system in its ground state, the energy is uniquely determined by the electron density $\rho(\vec{r})$. From this theorem, it follows that $\rho(\vec{r})$ determines the external potential (i.e., the potential created by the nuclei in the systems) and the number N of electrons. As a consequence, $\rho(\vec{r})$ determines the energy and all other ground-state electronic properties of the system. The second HK theorem follows the variational principle, enabling ground-state electron density to minimize the energy functional with respect to the corresponding density, which is exactly the ground state electronic density $\rho_0(\vec{r})$.

We can, thus, write the energy of the system as

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{ne}[\rho] = F_{HK}[\rho] + V_{ne}[\rho], \quad (2.3.2)$$

$F_{HK}[\rho]$ being a universal functional independent of the external potential created by the nuclei. Then, the variational principle with respect to $\rho(\vec{r})$ is applied to obtain the ground state energy from this equation:

$$E[\rho_{trial}(\vec{r})] \geq E[\rho_0(\vec{r})] = E_0, \quad (2.3.3)$$

where $\rho_{trial}(\vec{r})$ can be a trial density, which minimizes the energy. The exchange term (implicitly involved in the HF construction) and the electronic correlation are contained in the universal functional.

By solving Equation 2.3.3, the form of the universal functional $F_{HK}[\rho]$ should be known, as $V_{ne}[\rho]$ is then also known. However, it is impossible to know a mathematical expression of $T[\rho]$ and $V_{ee}[\rho]$ for a system of interacting particles. To overcome this problem, Kohn and Sham proposed to replace the system of interacting particles subject to the nuclear potential $v_{ext}(\vec{r})$ with a non-interacting system subject to a particular external potential v_{KS} having the exact density of the system [120]. With this approach, the electronic energy of the fictitious system will be calculated from an auxiliary set of single-particle orbitals Φ^{KS} called the Kohn-Sham (KS) orbitals. In the KS formulation, the energy is given by

$$E^{KS}[\rho] = T_{KS} + U_{Hartree}[\rho] + E_{xc}[\rho] + V_{ext}[\rho], \quad (2.3.4)$$

with

$$\rho(\vec{r}) = \sum_{i=1}^N \Phi_i^{KS*}(\vec{r}) \Phi_i^{KS}(\vec{r}). \quad (2.3.5)$$

The first term of Equation 2.3.4 is given by the KS orbitals, $U_{Hartree}[\rho]$ is the Hartree energy, and $V_{ext}[\rho]$ is a functional of the density depending of the external potential created by the nuclei. All these terms are defined as:

$$T_{KS} = -\frac{1}{2} \sum_{i=1}^N \int \Phi_i^{KS*}(\vec{r}) \nabla^2 \Phi_i^{KS}(\vec{r}) d\vec{r},$$

$$U_{Hartree}[\rho] = \frac{1}{2} \int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}'|},$$

$$V_{ext}[\rho] = \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r}.$$

$E_{xc}[\rho]$ includes the exchange and correlation energies containing the differences between the real system and the non-interacting one:

$$E_{xc}[\rho] = (T[\rho] - T_{KS}) + (V_{ee}[\rho] - U_{Hartree}[\rho]).$$

DFT is aimed to obtain the exact energy if the exchange-correlation functional used is exact. Thus, the accuracy of the DFT depends on the accuracy of the description of the exchange-correlation functional.

There are many different approximate functionals that are divided into categories. The most simple approximation is the local density approximation (LDA). These functionals are based on the approximation of the exchange-correlation potential for a uniform electron gas. This is a model system with a consistent density of electrons. Therefore, it is inaccurate for molecular systems such as peptides and proteins. A second category of functionals is the gradient corrected functionals (GGA). These functionals depend on $\vec{\nabla}\rho$ as well as on ρ , which helps to consider a deviation from uniformity in molecular systems. This brings results that are closer to the experiments in terms of binding energies and H-bonds, but this is still not enough for systems like peptides.

An additional degree of accuracy can be reached by combining the exchange and correlation obtained using the GGA method with some percentage of the exchange potential as described by the Hartree-Fock theory. These functionals are called hybrid functionals and are widely used for calculating the properties of peptides and proteins. The B3LYP functional is one of the most popular functional as it yields good structural and thermochemical properties with a reduction of self-interaction errors due to the semi-local approximation [121]. The B3LYP functional can be written as:

$$E_{xc}^{B3LYP} = (1 - a_0)E_x^{LDA} + a_0E_x^{HF} + a_xE_x^{B88} + a_cE_c^{LYP} + (1 - a_c)E_c^{VWN}, \quad (2.3.6)$$

where E_x^{LDA} is the exchange energy in the LDA approximation [120], E_x^{HF} is the Hartree Fock exchange energy [122], E_x^{B88} is defined by the gradient correction of Becke [123]. E_c^{LYP} is the correlation energy given by Lee, Yang, and Parr [124], and E_c^{VWN} is the local correlation functional VWN(V) from Vosko, Wilk, and Nusair [125]. The values a_0 , a_x and a_c are a set of parameters that were proposed by Becke after fitting a test data set with a similar hybrid functional B3PW91 [126]. The fit resulted in the following values:

$$a_0 = 0.2, a_x = 0.72, a_c = 0.81$$

All calculations for the hydrated species such as deprotonated adenosine monophosphate (Chapter 4) or protonated phosphotyrosine (Chapter 5), and part of the calculations done for the protonation sites in peptides project (Chapter 6) were done with the B3LYP functional as implemented in the ORCA software [127].

2.4 Computations of ground state geometries and X-ray absorption spectra

2.4.1 Ground state: geometry and IR spectrum

Insights on the physicochemical properties, molecular structures, chemical reactivity, or vibrational spectra can be understood by first exploring the change of the potential energy when nuclear coordinates change. The potential energy surface (PES) is thus a function of a system's potential energy with respect to one or more nuclear coordinates. For small molecules, a PES tends to have a global minimum, which is the lowest energy that a molecule with these nuclear coordinates can reach. For medium and large molecules, a PES tends to have local minima, implying that there is no single nuclear coordinate representation at the lowest energy. As outlined on the PES, stable molecular structures are found in the deep well, and the barrier energy to overcome this deep well is the energy necessary to dissociate the interaction of those nuclear coordinates. Representing the PES as a harmonic or an anharmonic oscillator, the shape of the valley at the deep well determines the vibrational spectrum. Then, the exploration of the PES and the identification of its stationary points (minimum, saddle points, ...) are crucial.

Geometry Optimizations

A set of approaches based on estimating forces between atoms or clusters of atoms are force field computational models. These force field models are not grounded on *ab initio* methods but follow a morse potential behavior in which their components include the interatomic interactions as the covalent and the intermolecular interactions as the non-covalent ones. In this context, the energy is the sum of parameterized potentials (covalent bonding and non-covalent bonding terms) in which the electronic structure is not described explicitly, as each atom is treated as a point charge. The force field parameters are usually valid for a particular electronic ground state at a given covalent structure. Moreover, force field methods significantly lower the computational cost while reducing the precision of the system description. This reduction enables the modeling of systems with several thousand atoms, which is typically unfeasible with other methods due to their computational cost.

Usually, the most searched stationary points in the PES exploration are the local minima corresponding to a first-order energy gradient equal to zero and positive second-

order energy derivatives as :

$$\frac{\partial E}{\partial R_a^\alpha} = 0 \text{ and } \frac{\partial^2 E}{\partial R_a^\alpha \partial R_b^\beta} > 0, \quad (2.4.1)$$

where α and β specify the direction X, Y, or Z of the nuclear coordinates R.

These PES sites give energetically stable molecular structures, and the search for these structures is done by geometry optimizations. In this work, the geometry optimizations are performed at the DFT level as DFT has a very favorable cost-accuracy ratio and because analytical gradients are available for most functionals like the B3LYP. Moreover, Hybrid DFT optimizations can be sped up by the use of the RIJCOSX approximation [128] implemented in the ORCA software [127]. In this thesis, the geometry optimizations are done at the DFT-B3LYP level with the RIJCOSX approximation and the TZVP basis set.

IR Spectrum

Once a stable structure is found, we perform the calculation of its vibrational spectrum. The molecular normal modes are, in general, calculated within the harmonic oscillation model. If the molecule is in its equilibrium conformation, it sits at the lowest point (at least locally) on the PES. The cross-section of the PES profile close to this point can then be assumed to be approximately parabolic, such that the second derivative of the energy with respect to a nuclear coordinate can be interpreted as a force constant for the harmonic oscillation of an atom along this coordinate. Since molecular vibrations in polyatomic molecules involve the simultaneous displacement of multiple atoms, this harmonic oscillator model can be generalized to multiple nuclear coordinates. The normal modes and their frequencies then become eigenvectors and eigenvalues of a force constant matrix. Then, to obtain the vibrational spectrum of a molecular system, we first calculate the Hessian matrix, i.e. the second derivative of the energy with respect to all coordinates defined as

$$H_{ab}^{\alpha\beta} = \frac{\partial^2 E}{\partial R_a^\alpha \partial R_b^\beta}. \quad (2.4.2)$$

Upon its diagonalization, we get the frequencies of the normal modes. However, frequencies calculated by *ab initio* methods often do not match the experimental spectra if a scaling factor (between 0.8 and 1) is not applied to the calculated frequency. This scaling compensates for the fact that the PES is not harmonic but follows a Morse potential and that the electronic structure calculation is approximate. In this thesis, the frequencies and the associated intensities are calculated at the DFT-B3LYP level with the TZVP basis set, and a scaling factor of 0.965 is applied.

2.4.2 Excited states: X-ray spectrum

Different methods, such as the previously mentioned CIS and CCSD, have been implemented to describe the electronically excited states of large molecules. However, in most cases, these methods become computationally expensive with the increasing size of the system. Another possible method is to use time-dependent DFT (TDDFT) as an affordable method for calculating transition energies and moments on the basis of the Kohn-Sham determinant. However, with the classically used functionals, TDDFT can be highly approximate and not give the right results.

As DFT is a single determinantal method (built from N lowest-energy KS orbitals), it cannot properly deal with multiplet problems, as is the case for X-ray absorption calculations at the L-edge. Therefore, Roemelt et al. have developed a new method for the calculation of transition metal L-edge X-ray absorption spectra [129], using the ORCA software [127], combining DFT with restricted open-shell configuration interaction with singles (ROCIS). A complete description of the DFT/ROCIS method as implemented in the ORCA software is explained in references [129]–[131]. To determine the absorption spectra, the DFT/ROCIS works with the framework of the minimal coupling scheme [132]. The absorption intensity for a transition state between $|I\rangle$ and $|F\rangle$ is proportional to the absorption cross section σ

$$\sigma = \frac{2\pi\omega}{I}(f_{ed} + f_{md} + f_{eq})\delta(E_{IF} - \hbar\omega), \quad (2.4.3)$$

ω and I being the angular frequency and time-averaged intensity of the light beam, respectively, E_{IF} is the energy difference between state $|I\rangle$ and $|F\rangle$ and f_{ed} , f_{md} and f_{eq} are the contribution from the electric dipole, the magnetic dipole, and the electric quadrupole terms respectively, as defined in reference [133].

Additionally, the calculated transition energies with their respective intensities enable any spectra simulation programs to generate a simulated absorption spectrum by convoluting the transitions with Gaussian-type or pseudo-Voigt-type functions with predefined full-width-at-half-maximum (FWHM). Indeed, it is necessary to fit the absorption line as the energy bandwidth of the light source, the core-hole lifetime, and the electronic broadening of the calculated transitions is not computed in theoretical calculations. Two types of functions are mainly used: Gaussian-type and pseudo-Voigt functions.

Gaussian-type functions average all types of experimental broadenings, such as the spectra bandwidth of the light source, the core-hole lifetime, and the intrinsic broadening of the electronic excitation, into one homogeneous set of functions with fixed FWHM value. With Gaussian-type functions, the FWHM is set between 0.1 eV and

3 eV depending on the atomic-edge probed, and the type of experiment performed [91], [134]–[137].

A pseudo-Voigt function can be approximated by a linear combination of Gaussian and Lorentzian terms. The Gaussian term accounts for the photon-energy broadening, whereas the Lorentzian term accounts for the core-hole lifetime broadening [138], [139]. The FWHM of the Lorentzian term can be calculated according to the work of Thompson et al. [140], which gives 0.6, 0.7, and 0.75 eV at the carbon, nitrogen, and oxygen K-edges, respectively.

The methods presented in this chapter were used to obtain the results of Chapter 4, Chapter 5, and Chapter 6. Figure 2.4.1 provides a flowchart summarizing the full procedure used during this thesis to calculate the different thermodynamic properties and vibrational or electronic spectra. To summarize, first a structural search is performed to compute a large set of possible geometries of the system of interest. The structures found are then optimized at the DFT level. To reduce the number of possible geometries, a selection using clustering methods or a selection through their electronic energies and geometry parameters is done. After reducing the number of conformers to a few relevant ones, we can perform frequency calculations to obtain the thermodynamic properties and vibrational spectra, or DFT/ROCIS calculations to compute their X-ray absorption spectra at the different atomic edges.

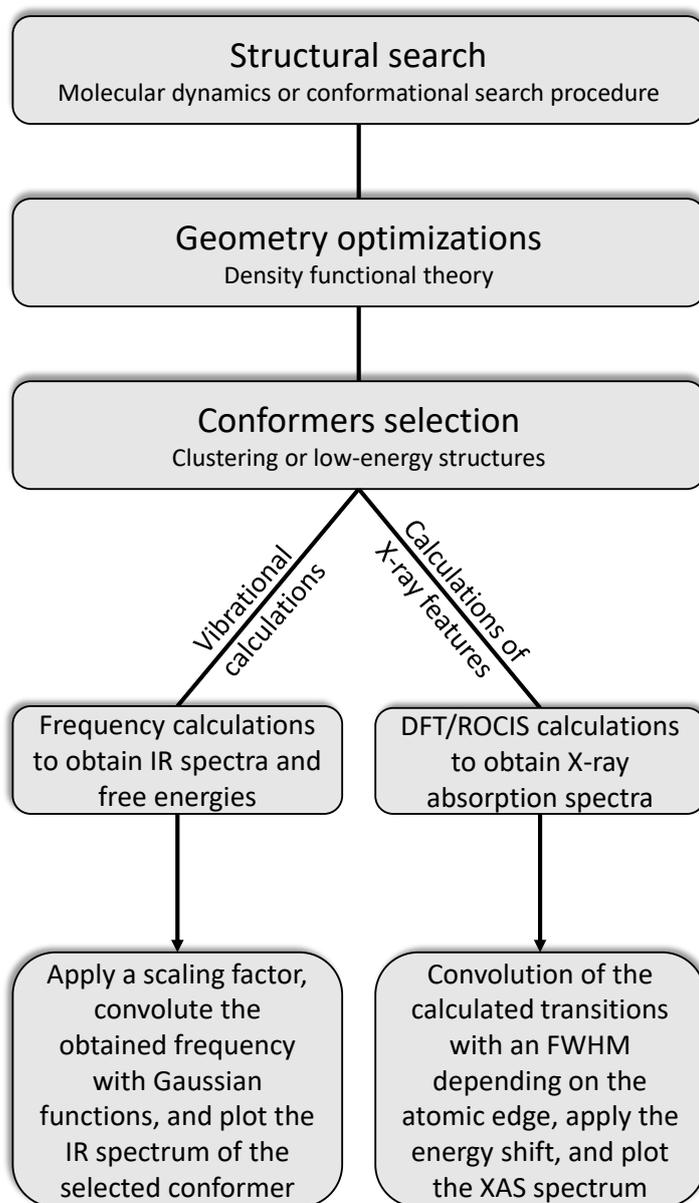


Figure 2.4.1: Flow chart highlighting the main steps of the quantum-chemical calculations performed for this work.

3 | Experimental methods

This chapter presents the experimental technique used in this thesis. Mass spectrometry is a technique used to analyze charged particles in the gas phase. Two setups were used in the course of this PhD thesis: the tandem mass spectrometer of the FS-BIG group located at the Deutsches Elektronen-Synchrotron, DESY in Hamburg, Germany, and the Ion Trap end-station of the BESSY II synchrotron in Berlin, Germany. The experiments were performed at the PETRA III synchrotron, Hamburg, Germany, and at BESSY II, Berlin, Germany. A brief description of the two setups, the two synchrotrons, and the beamlines used for the experiments detailed in Chapter 5 and Chapter 6 will be given.

3.1 The Spektrolatius setup

This experimental setup was used to obtain the data presented in Chapter 5 and part of the results in Chapter 6.

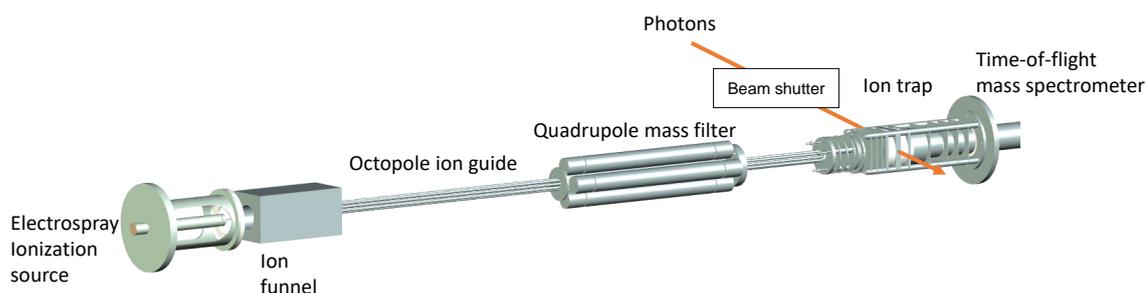


Figure 3.1.1: Schematic layout of the tandem mass spectrometer.

The tandem mass spectrometer used in this thesis was built in the FS-BIG group at DESY. A schematic of the setup is shown in Figure 3.1.1, and a picture of it is shown in Figure 3.1.2. The setup is composed of five different chambers, all acting

as differential pumping stages, from atmospheric pressure to 10^{-9} mbar. Note that the 10^{-9} mbar is base pressure in the trap and time-of-flight chambers when no buffer gas is introduced in the ion trap. Such low pressure is necessary to avoid, e.g., space charge repulsion in the ion trap by ionizing the background molecules while performing experiments using ionizing radiations (VUV, soft X-ray). The pressure values given below are those while in operation. This setup relies on a high-fluence electrospray ionization (ESI) source to introduce the ions into the gas phase. The electrosprayed ions are transported, from atmospheric pressure, by a heated capillary into the first pumping stage, which hosts an ion funnel (pressure of around 10^{-1} mbar) that collects and focuses the ions. They continue through a radio frequency (RF) octupole ion guide (pressure of around 10^{-3} mbar) to a quadrupole mass filter (QMS) where the ions of interest are selected according to their mass-to-charge (m/z) ratio. The quadrupole mass filter is operated at a pressure of 10^{-6} mbar. Then, a second octupole and a set of Einzel lenses guide and focus the ions into an RF 3D Paul trap, where they are accumulated prior to irradiation with photons. To allow efficient trapping of the ions, their kinetic energy is cooled down to room temperature by collisions with a helium buffer gas. Therefore, the operating pressure of the Paul trap is set around 10^{-4} mbar, measured in the trap chamber. All the cationic products are eventually extracted from the trap and analyzed in a reflectron time-of-flight (Re-TOF) mass spectrometer ($m/\Delta m = 1800$), which operates at low 10^{-7} mbar.

A beam shutter can be implemented between the trap chamber and the synchrotron beamline to control the duration of light exposure. The whole operation of the setup, including ion injection in the trap, irradiation, and extraction in the Re-TOF, is cyclic, with a typical period ranging from 100 ms to several seconds depending on the experiment parameters (e.g., photon flux, absorption cross-section). Additionally, the setup can be operated in so-called *duty cycle* mode as described in subsection 3.1.5.

The functioning of the different parts of the instruments will be detailed in the following sections.

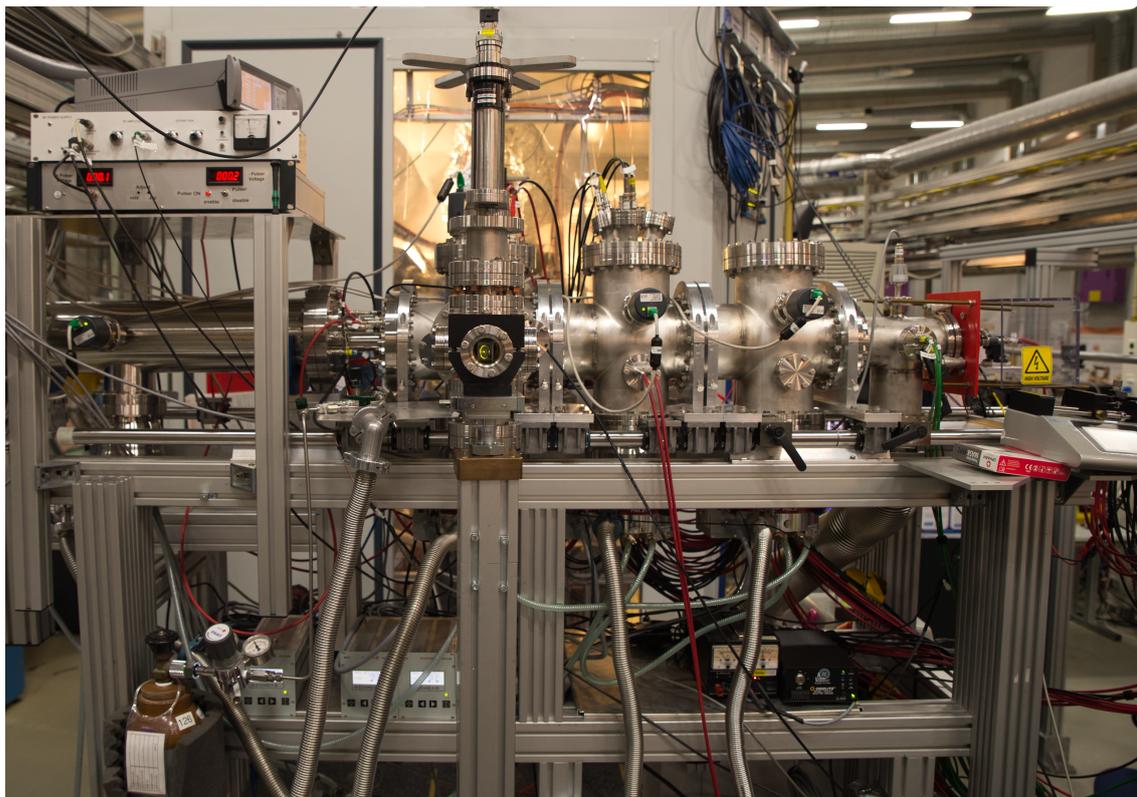


Figure 3.1.2: Picture of the Spektrolatius setup when coupled to the P04 beamline at PETRA III. From left to right: the TOF chamber, the trap chamber, the QMS chamber, the octupole chamber, and the source.

3.1.1 Electrospray ionization

Electrospray ionization is a soft ionization technique developed by John Fenn [21], used to transfer ions from a solution into the gas phase. In the case of biomolecules such as peptides, the ion beam can be formed of multiple charged states, protonated $[M + nH]^{n+}$ or deprotonated $[M - nH]^{n-}$, of the initial molecular compound M . It is a soft ionization technique which means that little residual energy is left after ionization and generally no fragmentation occurs upon ionization. In 2002, the technique obtained a Nobel Prize; John Fenn, the inventor of the method, shared a Nobel Prize in mass spectrometry with Koichi Tanaka and Kurt Wüthrich. As depicted in Figure 3.1.3 for positive ion mode, a dilute (around $30 \mu\text{M}$) analyte solution consisting of 1% of formic acid (to allow for protonation of the analyte) together with water and methanol (1:1) is injected by a syringe pump through a metal needle that is placed a few mm in front of a heated capillary which is the entrance of the vacuum chambers. The typical infusion speed used is 0.1 mL/hour . The capillary is heated to speed

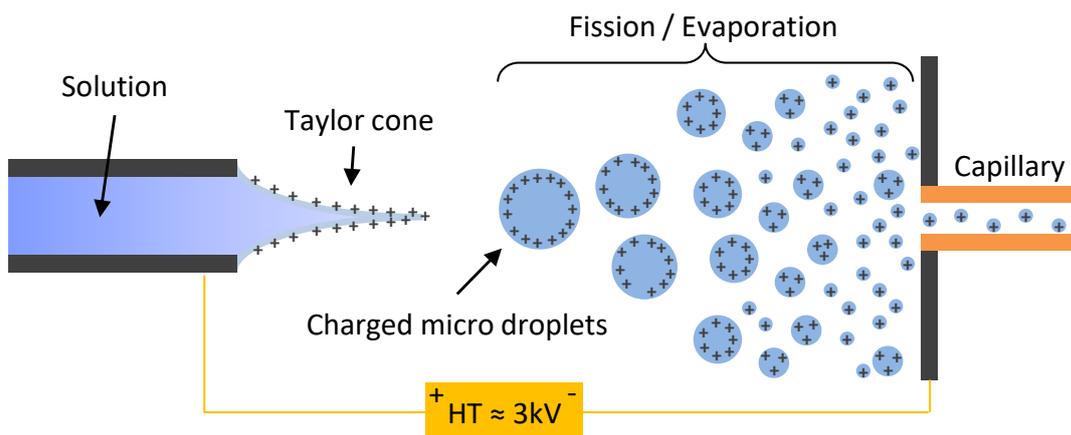


Figure 3.1.3: Principle of an electro spray ionization source working in positive mode.

up the desolvation process and prevent clustering due to the expansion cooling into vacuum. A high voltage is applied to the needle (3 - 4 kV), whereas a low voltage is applied to the capillary. Because of this strong electric field, charges accumulate at the tip of the needle, and the solvent is shaped into a Taylor cone. The ESI source can also be used in the negative mode by changing the solvent and swapping the voltage of the ESI needle and capillary. The charged droplets are emitted from the cone and the solvent evaporates until the surface tension cannot compensate the Coulomb repulsion anymore. This process continues until the Rayleigh limit is reached and fission into smaller droplets occurs. Two principle mechanisms have been proposed to explain the last step of the formation of gas-phase analyte ions starting from these highly charged droplets: the charge residual model [141], [142] or the ion evaporation model [143]. The charge residual model assumes that the gas-phase ion comes from a very small droplet that contains only one analyte ion, which undergoes evaporation of the remaining solvent. The ion evaporation model assumes that after repeated solvent evaporation and fission, the radii of the charged droplets decrease to a given size when the electric field is strong enough to cause direct emission of the solvated ions. Generally, large molecules like proteins follow the charge residual model, whereas small organic and inorganic ions follow the ion evaporation model [144].

This technique produces a very low ion current in the pA to nA range. Typically, we measured around 300 pA of ion current in our ion funnel for a protonated pentapeptide (methionine enkephalin). However, as it produces ionic species, it is possible to employ electric fields to achieve pure and dense targets by using ion guides, ion filters, and a radio frequency (RF) ion trap.

3.1.2 Mass filter

Since many species are introduced by the electrospray ionization source, e.g. analyte clusters, analytes of multiple charge states, impurities in the solution, or fragments of the analyte ions produced by multiple collisions with the residual gas molecules (for example, in the ion funnel) in the first stages of the experimental setup, it becomes necessary to select the ions of interest before accumulation in the Paul trap. A quadrupole mass spectrometer (QMS) can be used to select and guide ions according to their mass-to-charge m/z ratio. It consists of four parallel, cylindrical rods on which a combination of DC and RF electric fields is applied. A schematic representation is shown in Figure 3.1.4

The electric potential Φ applied to each pair of opposite rods is composed of a DC component U and an oscillating component $V\cos(\Omega t)$, with Ω being the angular frequency of the applied voltage. Opposite rods have the sign of the electric potential Φ , and, in contrast, adjacent rods have the opposite sign. The ion trajectory can be determined using Mathieu's equations. They imply that depending on the U/V ratio and Ω , ions with a certain m/z ratio have either a stable trajectory or an unstable trajectory. Ions with a stable trajectory go through the QMS, and the others collide with the rods [145]. This allows us to select the ions according to their m/z ratio.

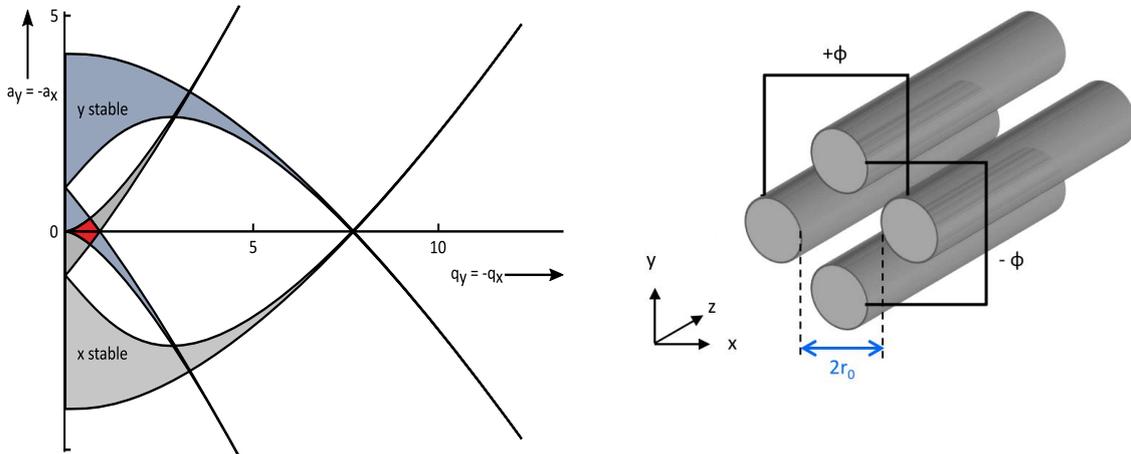


Figure 3.1.4: Left: The stability diagram for the two-dimensional quadrupole field. A QMS works in the region colored in red. (Figure adapted from [146]) Right: Schematic view of the electrodes of the QMS.

For a sinusoidally operated hyperbolic mass filter, the electric potential Φ is given

by

$$\Phi(x, y, t) = [U + V \cos(\Omega t)] \frac{x^2 - y^2}{2r_0^2}, \quad (3.1.1)$$

x and y are the distances along the given coordinate axes orthogonal to the QMS axis, and r_0 is the distance from the center axis to the inner surface of the rods (see Figure 3.1.4). Thus, the magnitude of the electric field along the x and y directions can be obtained:

$$E_x = -\frac{\partial \Phi}{\partial x} = -[U + V \cos(\Omega t)] \frac{x}{r_0^2}, \quad (3.1.2)$$

$$E_y = -\frac{\partial \Phi}{\partial y} = [U + V \cos(\Omega t)] \frac{y}{r_0^2}, \quad (3.1.3)$$

$$E_z = -\frac{\partial \Phi}{\partial z} = 0. \quad (3.1.4)$$

By defining the dimensionless parameter a and q such that

$$a = \frac{4zeU}{\Omega^2 r_0^2 m}, \quad (3.1.5)$$

$$q = \frac{2zeV}{\Omega^2 r_0^2 m}, \quad (3.1.6)$$

where e denotes the elementary charge [147]. The equation of motion can be defined by the Mathieu's differential equation

$$\frac{d^2 u}{d\tau^2} + (a_u + 2q_u \cos(2\tau))u = 0, \quad (3.1.7)$$

$$\tau = \frac{\Omega t}{2}, \quad (3.1.8)$$

u being x or y .

The Mathieu equation has two types of solution: stable motion when the particles oscillate in the $x - y$ plane with limited amplitude or unstable motion. The stability of the motion only depends on the parameters a and q . Therefore, in a $a - q$ map (represented in Figure 3.1.4), there are regions of stability and instability; in our case, only the overlapping region where x and y are stable is of interest. For fixed values of r_0 , Ω , U , and V , all the ions with the same m/z are stable on the same region of the $a - q$ map. Since $a/q = 2U/V$ and does not depend on m , all masses lie along the line $a/q = \text{const}$. Changing U and V simultaneously, such that a/q remains constant, brings the ions of various mass successively into the stability region, which allows us to choose which ions have a stable motion. A pure ion guiding mode can be obtained when only the RF field is applied to the rods and $U = 0$. This is the so-called RF-only mode of the QMS.

3.1.3 RF trap

As the ESI source produces low ion flux, it is necessary to increase the ion density before irradiation with photons. Increasing the ion density improves the signal-to-noise ratio and tackles the low X-ray absorption cross-section. Therefore, a quadrupole ion trap is implemented between the QMS and the time-of-flight mass spectrometer to accumulate the ions into a dense cloud. The Paul trap is an RF ion trap developed by Wolfgang Paul, who shared a Nobel prize with Hans Georg Dehmelt for the development of the ion trap technique [146]. The Paul trap consists of three hyperbolic electrodes: one central ring, to which an RF field is applied, and two cap electrodes placed at the ends, called injection and extraction electrodes. A representation of the Paul trap and a picture of the one implemented in the Spektrolatius setup are shown in Figure 3.1.5.

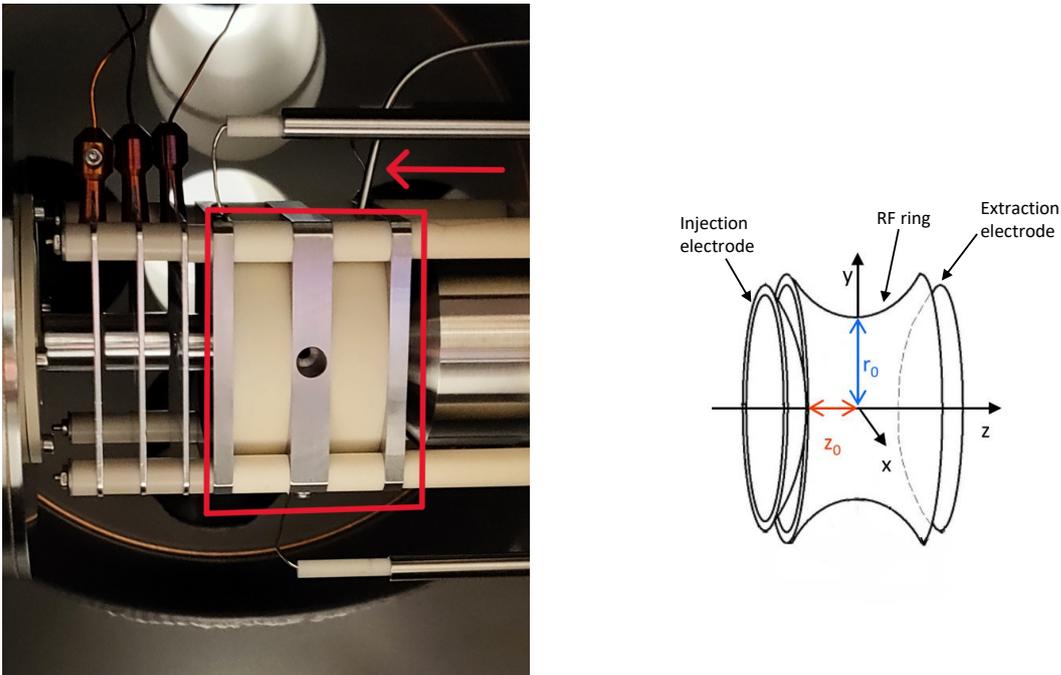


Figure 3.1.5: Left: photography of the trap. The red square shows where the trap is located, and the left arrow shows the He buffer gas injection line. On the left of the photograph, one sees the three electrodes of the Einzel lens that focus the ions inside the trap. On the right side of the trap, the first electrodes of the Reflectron-TOF mass spectrometer are located. Right: representation of the ion trap.

The ratio between the radius r_0 of the central ring and the distance z_0 between the center of the trap and one of the cap electrodes should respect the relation $r_0^2 = 2z_0^2$ to

obtain a quadrupolar field everywhere in the trap. Applying a potential of the form $U_0 + V_0 \cos(\Omega t)$ between the ring and the two end electrodes allows the creation of a quadrupolar field expressed in cylindrical coordinates by:

$$\Phi(r, z, t) = \frac{U_0 + V_0 \cos(\Omega t)}{2r_0^2} (2z^2 - r^2), \quad (3.1.9)$$

with $r^2 = x^2 + y^2$.

The ions' motion in the Paul trap can be as well described by Mathieu's equations, see Equation 3.1.7 [145].

By comparison with equation 3.1.9 and equation 3.1.7, the relevant a and q parameters can be obtained:

$$a_z = -2a_r = -\frac{8zeU_0}{m\Omega^2 r_0^2}; \quad q_z = -2q_r = \frac{4zeV_0}{m\Omega^2 r_0^2}. \quad (3.1.10)$$

For the trap, there are different values for a and q for which the solutions are stable simultaneously for the axial and radial directions. If this is fulfilled, the ions are confined within the trap. Our Paul trap operated in RF-only mode; thus, $U_0 = 0$, i.e., $a_z = 0$. Therefore, the ions have a stable trajectory when $q_z \leq 0.908$, defined as the lowest mass-to-charge ratio (also named low-mass cutoff) [147]. This means that at fixed Ω and V_0 , we can trap ions with a m/q of

$$\frac{m}{q} \geq \frac{4V_0}{0.908\Omega^2 r_0^2}, \quad (3.1.11)$$

where $q = z.e$ is the charge of the ions. For example, for the Paul trap of Spektrolatius, we have $r_0 = 10$ mm, $\Omega = 1$ MHz; therefore, for $V_0 = 500$ V, this gives a low-mass cutoff at m/z 54.

Here, we use the trap to accumulate the ions in a relatively dense cloud (10^6 ions.cm⁻³ [148]) and irradiate this cloud using photon beams. To this end, first, the ions are injected into the trap, and then their kinetic energy is dissipated by using a helium buffer gas. The motion of ions is focused on the center, and the trapping efficiency is increased. The last step is to extract the ions into the time-of-flight mass spectrometer. To do so, the RF voltage is switched off, and an electric field is applied between the end caps.

3.1.4 Time-of-flight mass spectrometer

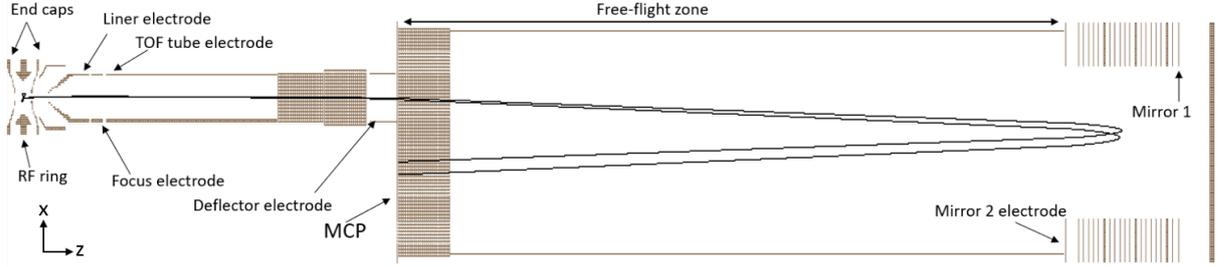


Figure 3.1.6: SIMION [149] representation of the Re-TOF mass spectrometer. In black, the trajectory of two ions, with different starting positions in the trap, inside the Re-TOF.

The analysis of the fragments produced by the photodissociation of the parent ion is performed in a reflectron time-of-flight mass spectrometer. As suggested by the name, this type of mass spectrometer utilizes the measurement of the flight time Δt of charged fragments to determine their m/z ratios. The ions in the Paul trap are extracted from it at the time t_i by imposing an electric field created by a potential difference between the trap's end caps. The charged particles are re-accelerated and focused using a set of three electrodes, indicated as Liner, Focus, and TOF tube electrodes in the sketch of Figure 3.1.6. Then, they enter the free-flight zone (without a field), are reflected in the reflectron's mirror, and are finally detected at the time t_f on a microchannel plate (MCP) detector. The time-of-flight of the ions is then $\Delta t = t_f - t_i$, and the intensity of the signal as a function of Δt is called the time-of-flight spectrum.

The time of flight of a particle depends on its m/z ratio. Therefore, one can obtain a mass spectrum using a time-of-flight spectrum. Indeed, when ions are accelerated along the z -axis by a uniform electric field of norm E , Newton's second law gives:

$$ma = qE, \quad (3.1.12)$$

with m being the mass, a its acceleration and $q = z \cdot e$ the charge of the ions. Thus, in the acceleration zone, the position of the ion at all times is given by

$$z(t) = \frac{qE}{2m}t^2. \quad (3.1.13)$$

Therefore, the measurement of the time of flight Δt of an ion at a given position z makes it possible to obtain its m/z ratio by

$$m/z \propto \Delta t^2. \quad (3.1.14)$$

Using a calibration, Equation 3.1.14 can be used to obtain a mass spectrum from the time-of-flight spectrum that has been recorded.

The reflectron time-of-flight spectrometer (Re-TOF, see Figure 3.1.6) is used to increase the flight distance and to improve the mass separation by reducing the spread of flight times due to the ions' different kinetic energies. Its principle is to reverse the trajectory of the ions by imposing a force opposite to their velocity vector due to a homogeneous electric field generated by electric potentials applied to parallel electrodes. Thus, for the same m/z ratio, the more energetic ions will penetrate deeper into the reflectron and spend more time in it than the less energetic ones, which therefore makes it possible to compensate for the initial dispersion in the kinetic energy of the ions. Consequently, time focusing of ions with the same m/z ratio is achieved on the detector, and the resolution is improved compared to a linear TOF.

3.1.5 Acquisition cycle

The experiments performed with this tandem mass spectrometer require a very high signal-to-noise ratio to distinguish the results of irradiation from background products. Despite the low background pressure and the optimization of the ESI source, contamination from the ESI source or from the residual gas can contribute to the mass spectra recorded. The residual gas contains the expected molecules N_2 , O_2 , CO_2 or H_2O present in the atmosphere. However, these small molecules do not interfere with our molecule's fragments as they are below the low-mass cutoff used during our experimental campaigns. Our residual gas is also composed of molecules with a higher mass-to-charge ratio ($> m/z$ 45), mainly neutral molecules from the ESI source, and presumably also degassing material from our pulse valve and the Teflon-coated shutter blades.

Therefore, the experiments are done in duty cycle mode to obtain the mass spectrum resulting only from the fragmentation of the target molecule with the photon beam. This means successive cycles of three mass scans to subtract all possible sources of background noise. A schematic of the duty cycle mode is shown in Figure 3.1.7.

In each cycle, first, a spectrum with only the parent molecule is recorded (ESI-only). This is used to measure the precursor ion intensity and fragments coming from fragmentation with the buffer gas. Second, the mass spectrum resulting from the irradiation of the trapped molecules is recorded. For the third spectrum, the ESI is switched off, and the mass spectrum resulting from the irradiation of the residual gas is recorded. To obtain the final spectrum for further analysis, the mass spectrum resulting from

ESI-only and the one from the irradiation of the residual gas are subtracted from the mass spectrum obtained by irradiating the trapped molecules. A series of a hundred cycles is accumulated to average over long-term fluctuations of the source.

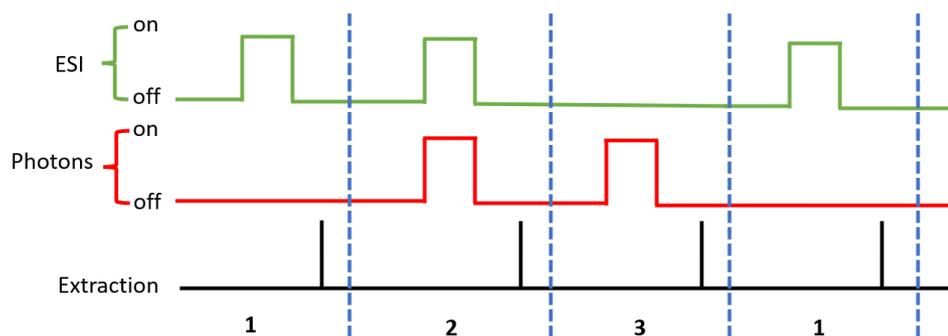


Figure 3.1.7: Schematic of the duty cycle acquisition principle. Extraction **1** represents the acquisition of the ESI-only spectrum, **2** the acquisition of the irradiation of the parent molecule, and **3** the acquisition of the photon-only spectrum.

3.2 The Ion Trap end-station

Part of the photon-induced fragmentation study shown in chapter 6 was conducted at the UE52_PGM Ion Trap beamline of the third-generation synchrotron radiation source BESSY II at the Helmholtz-Zentrum Berlin (Germany) (see subsection 3.3.2 for more details) [90]. A scheme of the apparatus is shown in Figure 3.2.1.

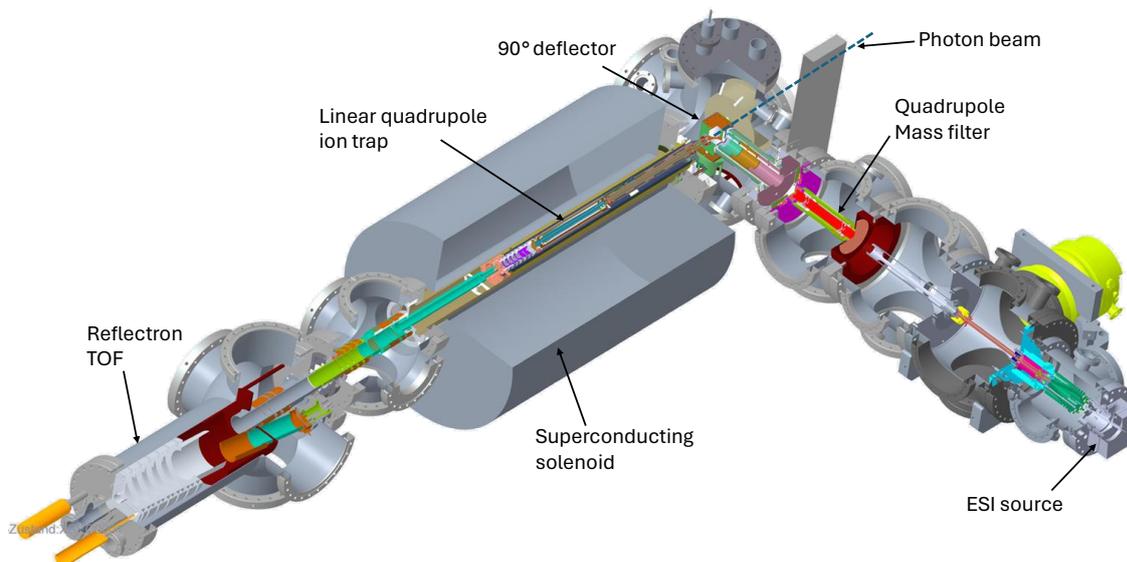


Figure 3.2.1: Schematic layout of the Ion Trap endstation at the UE52_PGM beamline at BESSY II (courtesy of Laura Berneaga and Max Flach).

The Ion Trap setup follows the same concept as *Spektrolatius*, described in section 3.1, but has a linear quadrupole trap that can be cooled down to 3 K with a superconducting solenoid magnet (0 - 5 T). The setup was originally designed for X-ray magnetic circular dichroism spectroscopy but can also be used for tandem mass spectrometry on biomolecular ions. For example, recent studies on the peptides melittin, methionine-enkephalin, and leucine-enkephalin have shown that the ion trap is suited for photodissociation studies of electrosprayed biomolecules [85]–[87].

The Ion Trap consists of a commercial electrospray source that introduces the ions into the gas phase. The electrosprayed ions are focused in an RF ion funnel and guided by a hexapole to an RF quadrupole mass filter, where they are filtered according to the m/z . The mass resolution of the mass filter is about $m/\Delta m = 1300$. The mass-selected ions pass through a 90° bender to enter the linear quadrupole ion trap, in axis with the photon beam path, where they are cooled down to 10 - 20 K with a helium buffer gas at cryogenic temperature. 2-D trapping is done by applying an RF field to the quadrupole rods when a static potential is applied to the trap entry and exit electrodes to trap the ions in the third dimension. The trap is continuously filled to reach a target density up to the theoretical value of $5 \times 10^8 \text{ cm}^{-3}$ [90]. Synchronously to the filling of the trap, the ions are irradiated with photons. The ions are extracted into a reflectron time-of-flight mass spectrometer by applying pulsed voltages to the exit lenses of the trap. Due to the design of the linear trap, only a specific mass range can be detected.

For instance, the ions are extracted from the ion trap and dispersed with respect to their mass-to-charge ratio while traversing the distance from the ion trap to the mass spectrometer. Therefore, ions arrive at the extraction region of the time-of-flight mass spectrometer with a mass-dependent delay. The mass range analyzed can be selected by tuning the time delay between the pulsed extraction from the ion trap and the extraction pulse of the first acceleration region of the mass spectrometer.

3.3 Synchrotron facilities

3.3.1 PETRA III

PETRA III is one of the most brilliant storage ring-based X-ray sources for high-energy photons worldwide. Located in Hamburg, Germany, it operates at 6 GeV particle energy and at photon energies between 150 eV and 200 keV, depending on the beamline. PETRA III runs at 120 mA electron beam current in continuous mode and 100 mA in timing mode. PETRA III hosts 25 operational beamlines for user access.

The results presented in Chapter 5 and Chapter 6 have been obtained by interfacing the Spektrolatius setup (see section 3.1 for details) with the variable XUV beamline P04 of the PETRA III synchrotron [150]. The P04 beamline provides photons over the range 250 - 3000 eV with a high photon flux of up to 10^{15} photons per second at 1 keV. The energy resolution is superior to 30000 at 335 eV. At the focus point, a spot size of $10 \times 10 \mu\text{m}^2$ can be achieved. For better overlap with the ion cloud in the Spektrolatius setup, the setup was placed 20 cm away from the focus point to reach a spot size of $100 \times 150 \mu\text{m}^2$.

3.3.2 BESSY II

BESSY II is a 3rd generation synchrotron radiation source that produces soft X-ray light in the range of 6 meV to 150 keV. BESSY II can accelerate the electrons to an energy of up to 1.7 GeV with a nominal beam current of 300 mA. Around 40 experimental stations are accessible to users.

Part of the NEXAMS spectra shown in Chapter 6 was conducted at the UE52_PGM Ion Trap beamline. This is an undulator beamline with a plane-grating monochromator delivering soft X-ray radiation with variable polarization in the energy range 120 - 1600 eV with a photon flux of 10^{12} photons per second. This includes the K-absorption

edges of carbon, nitrogen, and oxygen [28], [90]. The energy resolution is superior to 10000 at 400 eV, which is more than sufficient to record X-ray absorption spectra at the C, N, and O K-edges. Furthermore, the UE52_PGM Ion trap beamline is placed downstream from the focus point to create ideal conditions for the irradiation of the ion cloud in the linear quadrupole ion trap of the Ion Trap end-station.

3.4 A new electrospray source dedicated to the production of hydrated biomolecules

Over the last thirty years, there have been massive efforts to develop experimental techniques to study hydrated species in the gas phase in a controlled way in combination with an electrospray ionization (ESI) source. Different experimental methods have been tested, such as in-source solvation or partial desolvation and solvation in cold trapping devices. This section reviews briefly some examples of such techniques. Nguyen et al. were the first to produce hydrated biomolecules [151], using partial desolvation of the ions in the ESI source. To do so, the ESI needle was aimed directly at the entrance orifice of the capillary, and the nitrogen curtain gas flow, usually used to help desolvation, was reduced from 0.6 L/min to 0.1 L/min. With this technique, they could produce hydrated amine with up to 30 water molecules. Partial desolvation in the ESI source has been the method of choice for most studies on hydrated biomolecules [104], [152], [153]. Silveira et al. used cryogenic ion mobility mass spectrometry (IM-MS) combined with partial desolvation to probe the low-field dissociative limit for cluster ions held by hydrogen bonds [104]. They have shown that lowering the capillary temperature from 386 K to 340 K increases the maximum number of water molecules that can be retained in the diprotonated bradykinin (from 0 to 73 water molecules) but also increases the relative intensity of the hydrated clusters.

However, partial desolvation is not the only technique to produce solvated ions from an electrospray ionization source; solvation into a drift cell or trap allows the production of hydrated species. Liu et al. used a dedicated setup to produce hydrated clusters after their ESI source [99]. Their setup was originally designed to study the thermodynamical properties of biomolecules. It consists of an ESI source, a drift cell that can be heated up to 300 K, and a quadrupole mass spectrometer. This setup allows rehydration of the biomolecules in the drift cell, which can be filled with water vapor at pressure up to 3 mbar and thus works as a pick-up cell of water molecules. They obtained maximum water pickup at 260 K and 1.7 mbar of water vapor. In these experimental conditions, they could attach 20 water molecules to triply protonated

Neurotensin, a 13-residues peptide.

Marsh et al., using a similar concept as Liu et al., developed a dual cryogenic ion trap spectrometer for the formation and characterization of solvated species [154]. This setup consists of an electrospray source, a reaction ion trap in which they produce solvated species by introducing bursts of helium buffer gas seeded with solvent vapor. After the reaction trap, there are two radio-frequency ion guides and a second ion trap in which they can do an irradiation experiment. By cooling the reaction trap to 80 K, they were able to produce hydrated ions with up to 30 water molecules. Using this setup, they characterized protonated diglycine with infrared photodissociation (IRPD) spectroscopy with up to four water molecules.

The recipient of the Nobel prize for developing the electrospray technique has developed a dedicated electrospray source to produce solvated species, where they, first, completely desolvate the ions and hydrate them again in a second chamber by introducing water vapor [15], [27]. Such a source has been further improved by Nonose et al. introducing electrodes to guide the ions through the two chambers [24]. With this technique, as the hydrated clusters are produced in the source, it is then possible to further select the hydrated clusters of interest using the quadrupole mass filter. This allows for further study of the hydrated fragments using activation methods or spectroscopic techniques. Partial desolvation in an ESI source is solution-dependent and requires tuning the parameters of the source for each new sample, whereas the dedicated solvation ESI source is designed to hydrate any kind of biomolecule. With this, Zhan and Fenn produced hydrated peptides with up to 30 water molecules [27] and Nonose et al. produced protonated tryptophan with up to 40 water molecules attached [24].

3.4.1 Description of the hydration ESI source developed during this thesis

The hydration ESI source developed during this thesis is inspired by the design of Nonose et al. [24]. Prof. Nonose kindly provided the scheme, which was instrumental in designing the two chambers of the source. A scheme of the concept of the source is given in Figure 3.4.1.

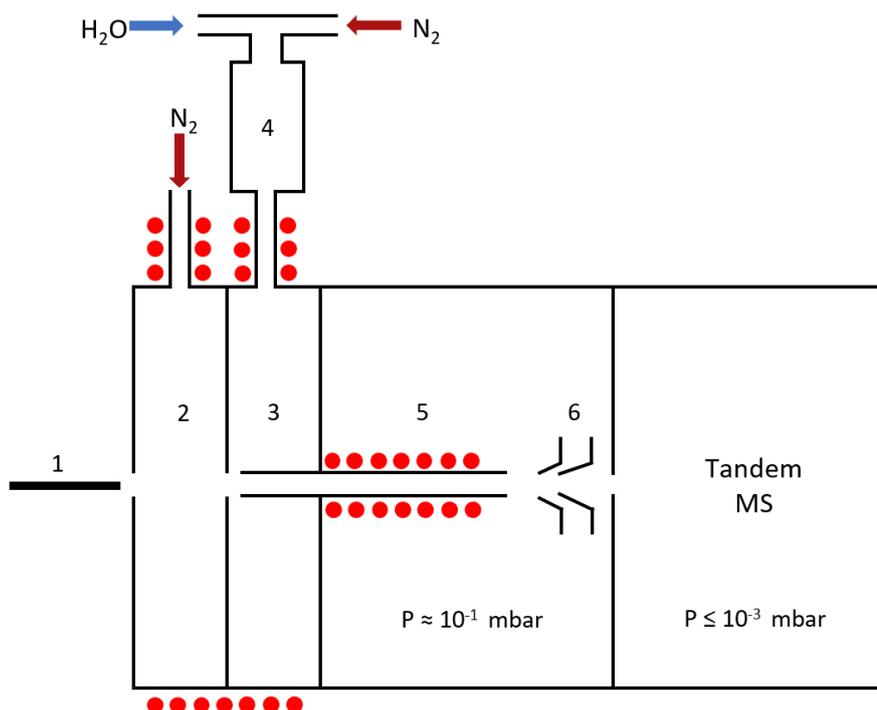


Figure 3.4.1: Schematic of the hydration source and the first stage of the tandem mass spectrometer. (1) electro spray needle, (2) drying chamber, (3) solvation chamber, (4) vaporizing chamber, (5) heated capillary, (6) skimmers. The red dots highlight the parts of the source that can be heated. The pressure in the drying chamber and solvation chamber is kept to atmospheric pressure, and the pressures in the skimmer chamber and the mass spectrometer are indicated in the schematic.

This source consists of an electro spray ionization source, but what changes is the addition of a solvation chamber between the ESI needle and the entrance of the capillary. The ions first enter a drying chamber to remove every molecule of solvent that could still be around the charged biomolecules. In the drying chamber, heated nitrogen gas is injected in a controlled way to remove the remaining solvent molecules. Once they are completely free of solvent, they enter the solvation chamber that is filled with

3.4. *A new electrospray source dedicated to the production of hydrated biomolecules*

a carrier gas, in our case nitrogen, seeded with water vapor. Two electrodes, spaced by 3 mm and with an inner diameter of 3 mm, are placed between the needle and the capillary. The needle is spaced by a few millimeters from the first electrode, and the capillary is spaced by less than a mm from the exit of the second electrode. These two electrodes play two roles: making the separation between the drying chamber and the solvation chamber through the small orifice (3 mm) and helping focus the ions, thus increasing the ion transmission by application of an electric field. Typically, in positive ion mode, a potential of 300 to 500 V is applied on the first electrode, and a potential of 50 to 150 V is applied on the second electrode. 3 mm after the exit of the capillary, a set of two skimmers, spaced by 2 mm, at the entrance of the mass spectrometer has been used as a differential pumping orifice and as well to focus the ions before being guided in the octupole. The first skimmer after the capillary has an orifice of 2 mm diameter and an inner angle of 30°, and the second skimmer has a smaller orifice of 1 mm diameter with an inner diameter of 20°. A 3D schematic of the actual source is given in Figure 3.4.2.

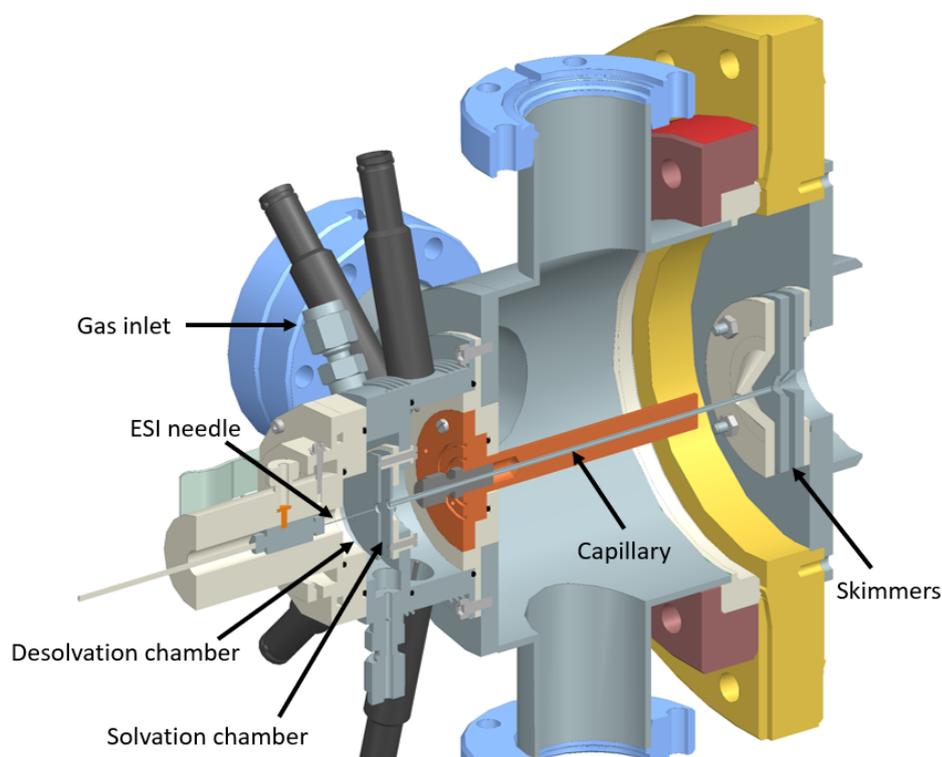


Figure 3.4.2: Design of the hydration source developed during this thesis.

A commercial flow meter from Bronkhorst, Ruurlo, the Netherlands (EL-FLOW select) controls the nitrogen gas flow. The mixture of nitrogen gas and water vapor is done using the controlled evaporator mixer (CEM W-102A-220-K) from Bronkhorst, in combination with a mass flow controller for the nitrogen gas and a flow controller for liquid (mini CORI-FLOW, Bronkhorst) to control the water flow. The mini CORI-FLOW has to be used in combination with a pump to provide constant liquid pressure. With this setup, we can control the flow of liquid and gas, and according to the liquid flow, the temperature of the CEM is set to ensure a constant and stable flow of water vapor. Typically, to obtain a water vapor at a water flow of 4 g/h and a nitrogen flow of 1.5 l_n/min, the temperature of the CEM needs to be set at 48°C. The temperature of the CEM can be defined by using the FLUIDAT online tool from Bronkhorst. This tool calculates the temperature that has to be set to the CEM to produce stable vapor depending on the desired liquid flow and gas flow.

In conclusion, we have built a new electrospray source dedicated to the production of hydrated (de)protonated biomolecules. During this thesis, the source has been designed and built. We were able to produce a stable signal of isolated biomolecules from m/z 137 (4-Amino-benzoic acid) to m/z 891 (RWMG₇) to perform NEXAMS experiments. We could also spray multiply charged cytochrome c (16+, m/z 770), a heme protein of mass 12 kDa. The results of Chapter 5 have been obtained with this ESI source. However, the capacity of the source to produce hydrated biomolecules still needs to be tested and characterized. Several factors require further investigation to fully characterize the source and use it for further experiments. These include:

- The impact of the water vapor concentration within the nitrogen gas in the solvation chamber.
- The influence of the nitrogen flow rates.
- The effect of the distance between the ESI needle and the first electrode.
- The influence of the temperatures of the capillary and the water vapor.

4 | Influence of the number of water molecules on the conformation and oxygen K-edge absorption spectra of a deprotonated nucleotide

4.1 Introduction

Adenosine monophosphate (AMP) is a nucleotide, a building block of DNA and RNA, that consists of the nucleobase adenine, the sugar ribose, and a phosphate group where deprotonation can take place (see Figure 4.1.1). AMP plays a role in metabolic processes, such as forming DNA strands or being part of cellular signaling. Jauker et al. showed that a liquid mixture of AMP with amino acid leads to the spontaneous formation of RNA strands and peptidyl RNA [41]. For these reasons, AMP has been extensively studied in the past decades and is still today an important subject [14], [20], [152], [155], [156].

Hydration of deprotonated AMP (AMP^-) has been previously studied at the molecular level by several groups. Liu et al. investigated the dissociation of the hydrated deprotonated AMP nucleotide induced by energetic collisions with neutral atoms [14] and by electron capture [20]. They showed that AMP^- was fully protected from fragmentation when hydrated by more than 14 water molecules in the case of collisions with neutral atoms, whereas in the case of electron capture, the authors concluded that the fragmentation was enhanced. Moreover, Milosavljević et al. studied the role of one water molecule on the radiation-induced fragmentation of protonated AMP when exposed to UV/VUV photons [152]. Using action spectroscopy, they demonstrated that the presence of a single water molecule inhibits dissociation in the UV region. This inhibition is attributed to intrinsic structural changes induced by the water molecule [152].

Calvo and Douady conducted a theoretical study on the stepwise hydration and evaporation of deprotonated adenosine monophosphate anions (AMP^-) [101]. In their paper, the authors explored the potential energy surface (PES) of AMP^- complexes containing up to 20 water molecules using replica-exchange molecular dynamics (REMD) with the explicit polarizable Amber ff99 force field combined with the TIP3P flexible potential for describing the water molecules. In the REMD simulations, 40 replica temperatures were used in the 80 - 1000 K range to cover the whole PES. The lowest-energy structures found with REMD were subsequently optimized using DFT calculations. The standard B3LYP hybrid exchange-correlation functional with the Gaussian-type basis set were used. Using these methods, they discovered that hydrated AMP^- undergoes significant solvent rearrangements within the 150 - 200 K range. For AMP^- hydrated with twenty water molecules, a few water molecules are lost after 1 μs , and it loses up to half of its water molecules after 1 s if thermalized at 350 K.

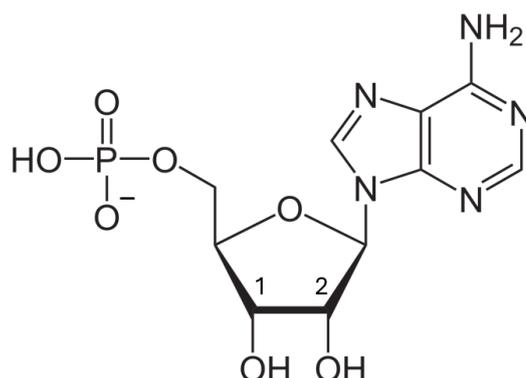


Figure 4.1.1: Chemical structure of deprotonated AMP (which is the more common form when AMP is electrosprayed). The numbers 1 and 2 indicate the two carbon atoms where the hydroxyl group can be replaced by a hydrogen atom.

This chapter extends the theoretical work from Calvo and Douady using their kindly given stable structures. The aim is to understand how the structure of AMP^- evolves with stepwise hydration by geometry optimization. First, the different hydrated structures and their binding energies are presented. Then, the spectroscopic properties of the different hydrated clusters are discussed for the calculated infrared spectra and the X-ray absorption spectra at the oxygen K-edge.

4.2 Methods

All calculations were performed using the ORCA electronic structure package [127]. As this work is an extension of the theoretical work of Calvo and Douady [101], the initial structures from this work were their, kindly given, most stable configurations for each cluster. All structures were, then, fully reoptimized employing the TZVP Ahlrichs basis set [157] in combination with the B3LYP exchange-correlation functional [123]. More details about the B3LYP functional can be found in Chapter 2. In all calculations done in this chapter, relativistic effects have been considered using the zeroth-order regular approximation (ZORA) [158].

In Figure 4.3.3, the number of hydrogen bonds was calculated as a function of the number of water molecules, as described later. A hydrogen bond was assumed to be formed if the distance between two water molecules or between a water molecule and an oxygen or hydrogen atom of AMP⁻ was shorter than 3 Å [8]. Analytical frequency calculations were performed at the same level of theory to simulate the IR spectra of the different hydrated AMP⁻. A scaling factor of 0.965, the theoretical scaling factor for B3LYP in combination with the TZVP basis set, was used for all computed IR spectra with a Gaussian-broadening of 50 cm⁻¹.

In the last section of this chapter, the calculated X-ray absorption spectrum at the oxygen K-edge of each cluster will be presented. The calculations were done using the density functional theory (DFT) combined with the restricted-open-shell configuration interaction with singles (ROCIS) [130]. The final absorption spectra were then broadened, using a pseudo-Voigt profile with a full width at half-maximum (FWHM) of 0.6 eV [91] and shifted by 14 eV according to the theoretical shift given by the B3LYP functional at the oxygen K-edge [159].

4.3 Structures characterization

Figure 4.3.1 and Figure 4.3.2 show the lowest-energy conformations of deprotonated AMP, isolated and hydrated, AMP⁻(H₂O)_m with m=0,1,2,3,5,10,15 and 20. Isolated AMP⁻ was found to be linear with an intra-hydrogen bond between an oxygen atom of the phosphate group and the -OH group of the sugar ribose (Figure 4.3.1). Solvation of AMP⁻ with one water molecule already induces structural changes. In this case, the hydrated nucleotide is folded with the water molecule between the adenine group and the phosphate group, as shown in Figure 4.3.1. The water molecule is bound to the phosphate group through an inter-hydrogen bond. The intra-hydrogen bond between the oxygen of the phosphate group and the -OH group of the sugar observed

in bare AMP⁻ is conserved when adding the water molecule. It is worth noting that this intra-hydrogen bond in AMP⁻ is conserved through stepwise hydration.

When a second water molecule is added to the system (see Figure 4.3.1), the folded structure of AMP⁻ is conserved, and the second water molecule forms an extra inter-hydrogen bond with an oxygen atom from the sugar group. Interestingly, for AMP⁻(H₂O)₃ (Figure 4.3.1), the third water molecule does not bind to the third oxygen atom of the phosphorus group (-POH) but to one of the water molecules also forming a second hydrogen bond with the oxygen atom at the deprotonation site of the phosphor group.

For AMP⁻ solvated with five water molecules (see Figure 4.3.1), the structure of the nucleotide is more folded than in the case of the nucleotide solvated by three water molecules, probably due to attraction between the water droplet and the adenine group. For instance, we observed three inter-hydrogen bonds between the water droplet and the adenine group of AMP⁻. Within the water droplet, two water molecules are bound to the deprotonation site of AMP⁻ (as for lower hydration level), and a third one creates an inter-hydrogen bond with the oxygen atom of the phosphate group, which already has an intra-hydrogen bond with the -OH group of the sugar. The two last water molecules form a hydrogen bond network with the water molecule bound to the nucleotide.

Similar behavior can be observed for AMP⁻(H₂O)₁₀ (Figure 4.3.2) where three water molecules bind to the oxygen atoms of the phosphate group. Moreover, a water molecule does two extra inter-hydrogen bonds with the nitrogen atoms of the adenine moieties of the nucleotide. The six other water molecules form a water droplet close to the phosphate group but bind to the adenine group through this one water molecule. This changes the structure of the nucleotide; the calculated structure shows a more linear structure than for the other hydration states shown in this section, probably due to Coulomb repulsion between the water droplet and the nucleotide, as already demonstrated in the work of Demireva et al. [55]. For solvation states higher than AMP⁻(H₂O)₁₀, we have also observed up to three inter-HBs between the water molecules and the adenine group where the other water molecules bind to each other. However, in AMP⁻(H₂O)₂₀, an additional HB is observed between a water molecule and the other -OH group of the ribose in the opposite plane. This could be explained by a first solvation shell being full after twenty water molecules. The number of intra- and inter-HBs observed in the cluster is shown in Figure 4.3.3.

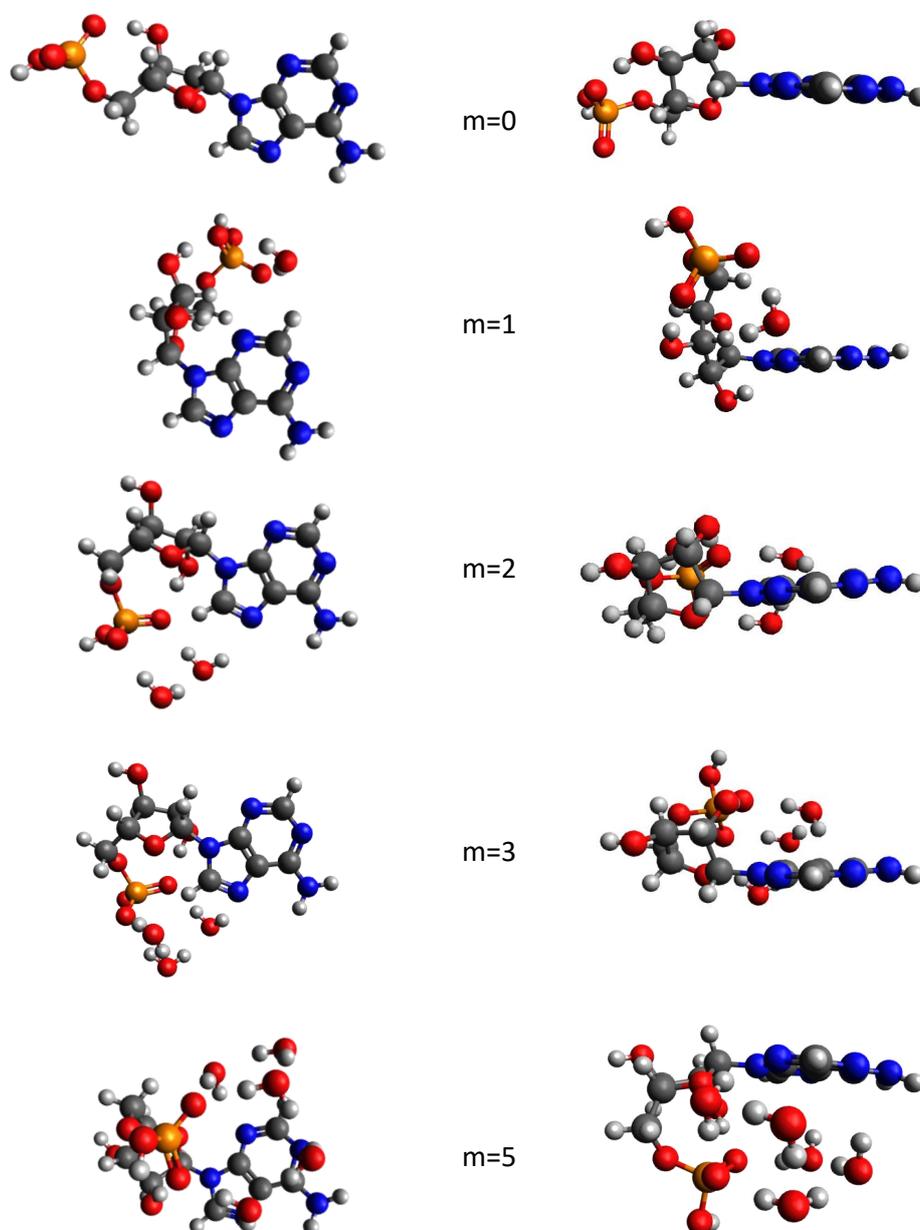


Figure 4.3.1: Structures of AMP⁻(H₂O)_m, m=0,1,2,3 and 5, after optimization at the B3LYP/TZVP level.

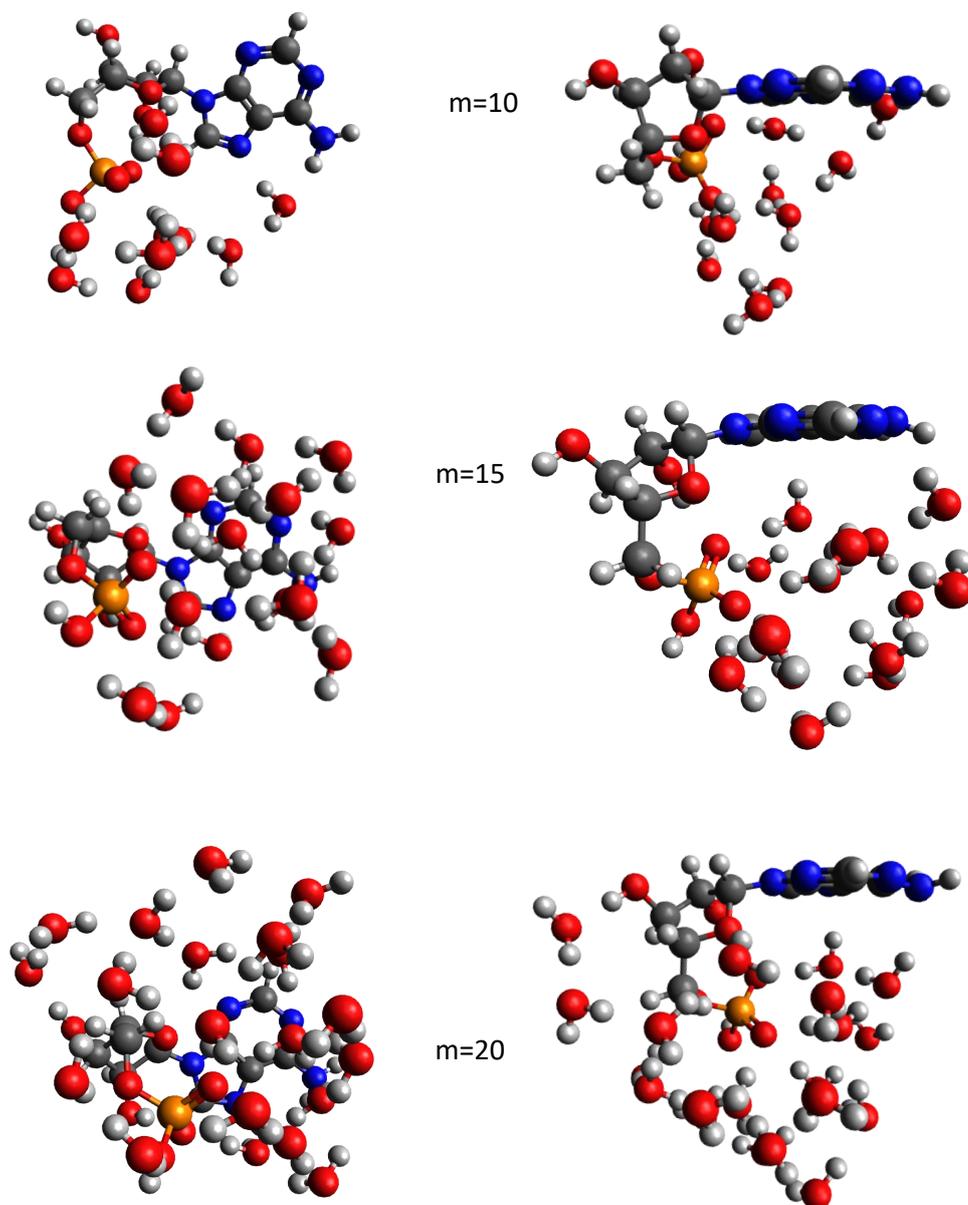


Figure 4.3.2: Structures of AMP⁻(H₂O)_m, m=10,15 and 20, after optimization at the B3LYP/TZVP level.

In conclusion, from the HBs network and the computed structures for hydrated AMP⁻, we showed that when AMP⁻ is isolated, its structure is preferentially linear, and already one water molecule is enough to fold the system and bring the phosphate group closer to the adenine moiety. Moreover, with increasing hydration levels, the

additional water molecules preferentially bind to one another to form a water drop of increasing size. For up to approximately nine water molecules added to AMP⁻, there is competition between the inter- and intra-hydrogen bonds. We observed a convergence of the number of inter-HBs after eight water molecules attached to AMP⁻, whereas the number of intra-HBs in the water droplet keeps increasing linearly with the hydration (see Figure 4.3.3). After ten water molecules, the Coulomb repulsion between the water droplet and the nucleotide becomes more important, and AMP⁻ unfolds itself. These results are in agreement with the study of Calvo and Douady as well as the early study of Liu et al. [101], [155].

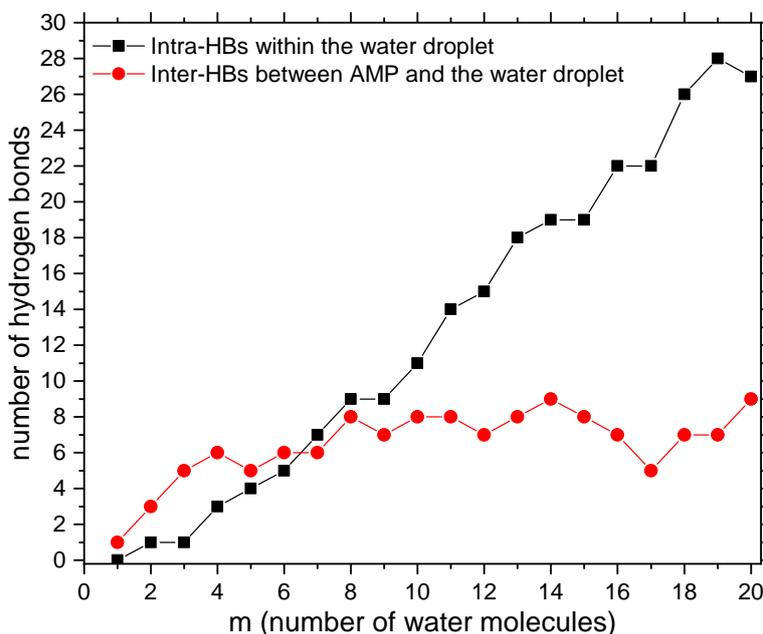


Figure 4.3.3: Number of hydrogen bonds (HBs) between the water drop and the nucleotide (inter) and within the water drop (intra) as a function of the number of water molecules.

To obtain more insight into the influence of solvation on the thermochemistry of AMP⁻, one can also study the binding energy of the system, $E_{bind}(m)$, as a function of the number of water molecules m . The binding energy can be computed from the energies of the full system (AMP⁻(H₂O)_m) minus the sum of the two isolated subsystems (AMP⁻ and (H₂O)_m) defined as

$$E_{bind}(m) = E_{AMP^-(H_2O)_m} - E_{AMP^-} - E_{(H_2O)_m}. \quad (4.3.1)$$

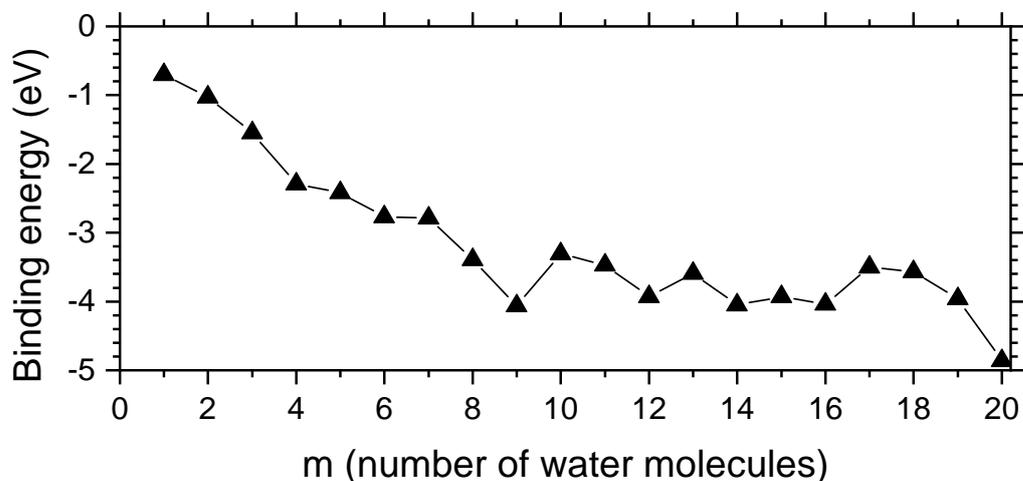


Figure 4.3.4: Binding energy of the different clusters as a function of their original size, calculated upon the removal of all water molecules (without consideration of molecular rearrangement).

The binding energy, as shown in Figure 4.3.4, is defined by the minimum amount of energy required to remove all water molecules from the cluster to obtain the isolated AMP⁻ (without considering molecular rearrangement). As expected, the binding energy decreases with the size of the cluster due to the different hydrogen bonds. The presence of the solvent allows the creation of inter-HBs with the nucleotide and HBs between water molecules, thus promoting the stability of the cluster. This decrease of the binding energy is very pronounced for clusters containing fewer than ten water molecules and is correlated with the increase in the number of inter-HBs for small systems (see Figure 4.3.3). Moreover, the binding energy stabilizes around -3.8 eV after eight to nine water molecules are added to the isolated AMP⁻. From AMP⁻(H₂O)₈ to AMP⁻(H₂O)₁₈, the binding energy remains approximately constant before decreasing again for AMP⁻(H₂O)₂₀. This can be explained by the number of intra-HBs between the nucleotide and the water drop, shown in Figure 4.3.3. Indeed, in this figure, the number of inter-hydrogen bonds between the water molecules and the nucleotides follows the same trend (inversed) as the binding energies over the full range of solvated clusters, whereas the number of intra-HB within the water droplet increases further with the hydration state. This suggests that after a certain hydration level, the added water molecules do not participate in the stabilization of the cluster. However, the

binding energy (Figure 4.3.4) shows a slight decrease from $\text{AMP}^-(\text{H}_2\text{O})_{17}$ and our highest solvation state, which is similarly reproduced in the inter-HBs network. This could be understood as the appearance of a second solvation shell, which would stabilize the cluster, or just fluctuations of the structure and binding energy within the calculation. Therefore, further REMD simulations combined with DFT calculations are needed to conclude whether the trend between $\text{AMP}^-(\text{H}_2\text{O})_{17}$ and $\text{AMP}^-(\text{H}_2\text{O})_{20}$ continue for higher level of hydration.

To better understand whether the calculated structures can be considered as possible structures for hydrated AMP^- , the energetics of stepwise hydration can be characterized for all the structures obtained. The hydration energies also provide information regarding the presence of stabilization by intramolecular hydrogen bonding in the nucleotide. The dissociation energy is computed from the energies of the full system ($\text{AMP}^-(\text{H}_2\text{O})_m$) minus the sum of the two systems $\text{AMP}^-(\text{H}_2\text{O})_{m-1}$ and H_2O defined as

$$E_{diss}(m) = |E_{\text{AMP}^-(\text{H}_2\text{O})_m} - E_{\text{AMP}^-(\text{H}_2\text{O})_{m-1}} - E_{\text{H}_2\text{O}}|. \quad (4.3.2)$$

This definition is the formula used by Chutia et al. for their study on solvated peptides and corresponds to the energy required to remove one water molecule from the cluster [7]. The evolution of the calculated dissociation energies as a function of the number of water molecules is given in Figure 4.3.5.

The dissociation energy oscillates between 0.23 and 0.75 eV, whereby it decreases on average with the number of water molecules. This trend has already been observed with semi-empirical PM6 or DFT methods in the study by Douady and Calvo [101]. The values obtained for the hydration energies are within the range of accepted values from thermochemical measurements of similar systems (0.3 to 0.6 eV / 6 to 14 kcal.mol⁻¹) [123], [155], [156], [160].

Moreover, the binding energies of the first four water molecules were determined experimentally by Liu et al. for similar nucleotides and especially deoxy- AMP^- (d AMP^-) which differs from AMP^- only by the replacement of a hydroxyl group by a hydrogen atom at the second position of the ribose moiety [155]. The sugar group of AMP^- possesses two carbon atoms that can be either bound to a hydroxyl group or a hydrogen atom (denoted 1 and 2 in Figure 4.1.1); thus, the second position is the carbon atom that is the closest to the adenine moiety. They showed that the experimental hydration energies for the first four water molecules that bind d AMP^- stem between 10.3 and 8.6 kcal.mol⁻¹, which is in good agreement with the results obtained in this study for the small hydrated clusters.

The maxima and minima of these energies should correspond, respectively, to so-called

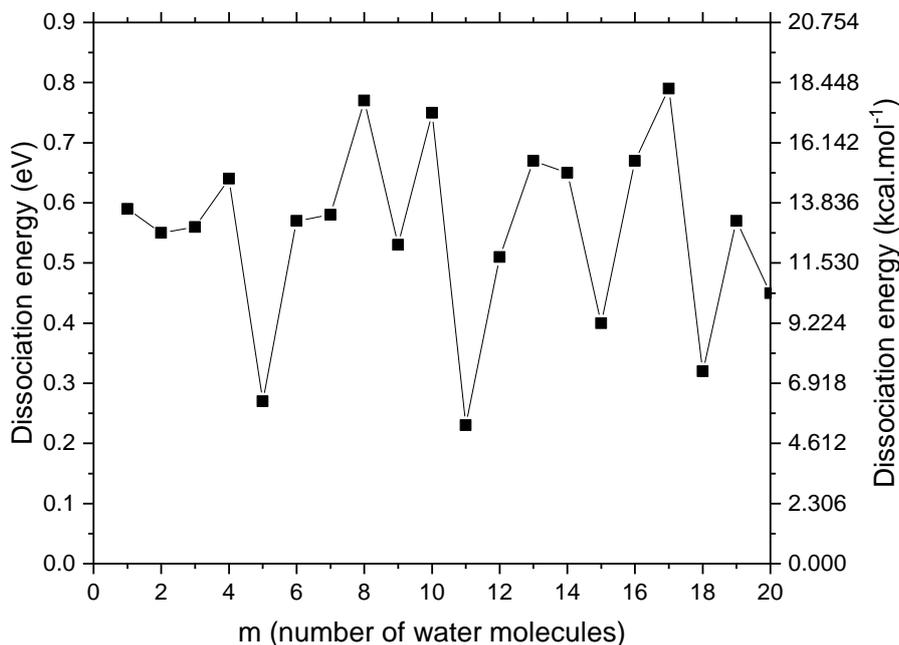


Figure 4.3.5: Dissociation energies of $\text{AMP}^-(\text{H}_2\text{O})_m$ $m=1,\dots,20$ calculated for the structures shown in Figure 4.3.1 and Figure 4.3.2.

magic and antimagic clusters. In this study, the DFT/B3LYP/ZORA-TZVP calculations predict magic numbers at sizes $m = 8, 10,$ and 17 as well as antimagic numbers at $m = 5, 11, 15$ and 18 . These results are partially in good agreement with those of Douady and Calvo [101]. However, the differences in dissociation energies are more likely to be imputable to the methods themselves rather than to different minima. All dissociation energies reported here are not corrected by the counterbalance correction to account for basis set superposition errors (BSSE). This effect could explain the differences with the DFT/B3LYP/6-31+G* calculations of Douady and Calvo. Therefore, experiments such as mass spectrometry, which record the mass spectrum of the full range of hydrated clusters, as presented in the work of Silveira et al. for hydrated gramicidin S [104], are necessary to conclusively determine the existence of magic and antimagic numbers for AMP^- .

Adding water molecules around the phosphate group of the nucleotide and making hydrogen bonds with the adenine group should also influence the spectroscopic properties of the system, and the changes should be observed in the electronic and vibrational transitions of the molecules. Indeed, IR spectroscopy is a fundamental tool

for investigating the structure of a given molecule or complex through the excitation of its vibrational states. X-ray radiation can be used as a probe of the local environment of a given system; this makes soft X-rays at the oxygen an interesting tool to learn more about the influence of the water molecules on the electronic structure of AMP⁻.

4.4 Calculated IR spectra of hydrated AMP⁻

This section describes the calculated IR spectra of isolated and solvated deprotonated AMP with up to ten water molecules. IR spectroscopy is based on the principle of resonant excitation of characteristic vibrational modes of chemical bonds through photoabsorption. It is a fundamental technique in molecular structure characterization that can be used to solve the three-dimensional structure of isolated and hydrated biomolecules.

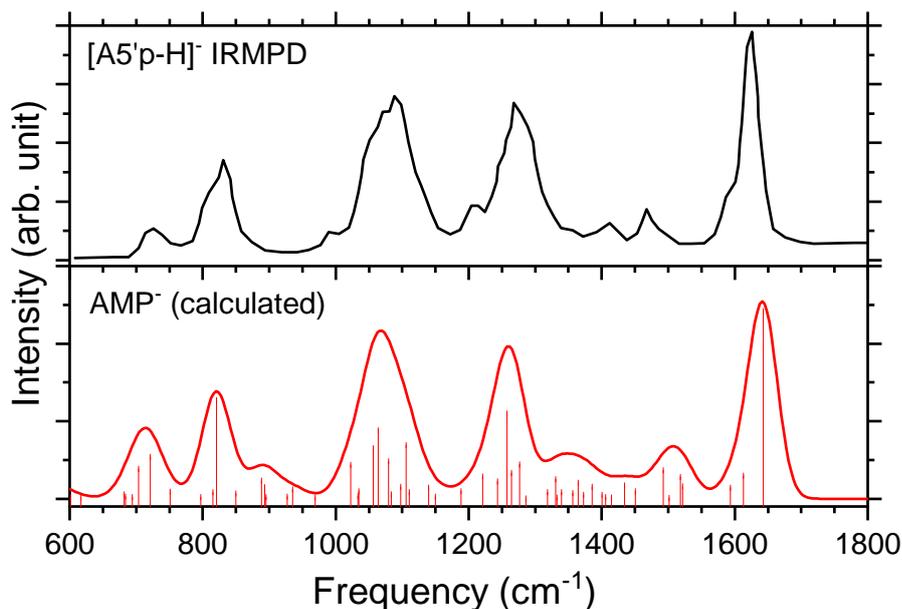


Figure 4.4.1: Top: experimental IRMPD spectrum of bare AMP⁻ in the region 600 to 1800 cm⁻¹ (from Nei et al. [50]). Bottom: calculated IR spectrum of isolated AMP⁻ structure from this study (scaling factor 0.965, FWHM=50 cm⁻¹).

Figure 4.4.1 shows the comparison between the experimental IRMPD spectrum of isolated deprotonated AMP measured by Nei et al. [50] and the calculated IR

spectrum from the stable structure of bare AMP⁻ in the fingerprint region, i.e. 600 - 1800 cm⁻¹. The calculated IR spectrum of the nucleotide is in excellent agreement with the experimental IRMPD spectrum, with the different resonances and their relative intensities well reproduced. Due to the theoretical scaling factor, a redshift of 50 cm⁻¹ is observed between the experimental and calculated spectra and accounted for.

Based on the work of Nei et al. [50] and on the analysis of the vibrational bands, the bands in the fingerprint region can all be attributed. They concluded that the features in the region between 650 and 1400 cm⁻¹ are vibrational signatures of the phosphate and ribose moieties with minor contributions arising from the nucleobase while bands in the region 1400 and 1800 cm⁻¹ arise from the IR active modes of the nucleobase. We can assign the different bands by analyzing the results of the DFT calculation for isolated AMP⁻. The following frequencies were all shifted by +50 cm⁻¹ in order to be compared to the experimental spectrum. The first main band at 1050 cm⁻¹ can be attributed to CO asymmetrical stretching and CO bending. A P=O stretching, as well as a CH bending, are responsible for the intense band stemming at 1250 cm⁻¹. Both results of the DFT calculation are in good agreement with the study of Nei et al. [50]. The low-intensity band at 1500 cm⁻¹ can be attributed to vibrations within the adenine group of the nucleotide via C-C stretching and CH bending mode. The intense band at 1650 cm⁻¹ is due to an NH bending from the adenine group combined with a C-C stretching of the adenine.

Furthermore, Nei et al. also compared their experimental spectrum with theoretical calculations. They performed their structural search at the MP2 level, and showed six possible structures of deprotonated AMP, all lying within 140 meV. For all their computed structures, the two hydroxyl groups of the ribose are in the same plane, whereas in our calculated structure, they appear to be in opposite planes. In their case, however, the phosphate group also binds to the ribose through a hydrogen bond between the deprotonation site and the hydroxyl group 1 (at position 1 as defined in Figure 4.1.1), and the global structure of AMP⁻ is linear as well. Although there are some discrepancies in the structure, our structure is in good agreement with their computed structures. Moreover, the analysis of the vibrational bands aligns with the findings of Nei et al., confirming that the calculated structure for AMP⁻ could be a viable candidate. This underlines the suitability of the theoretical method used here to simulate the structure and properties of the nucleotide. Therefore, we have extended the study to hydrated clusters of AMP⁻ using the same theoretical method.

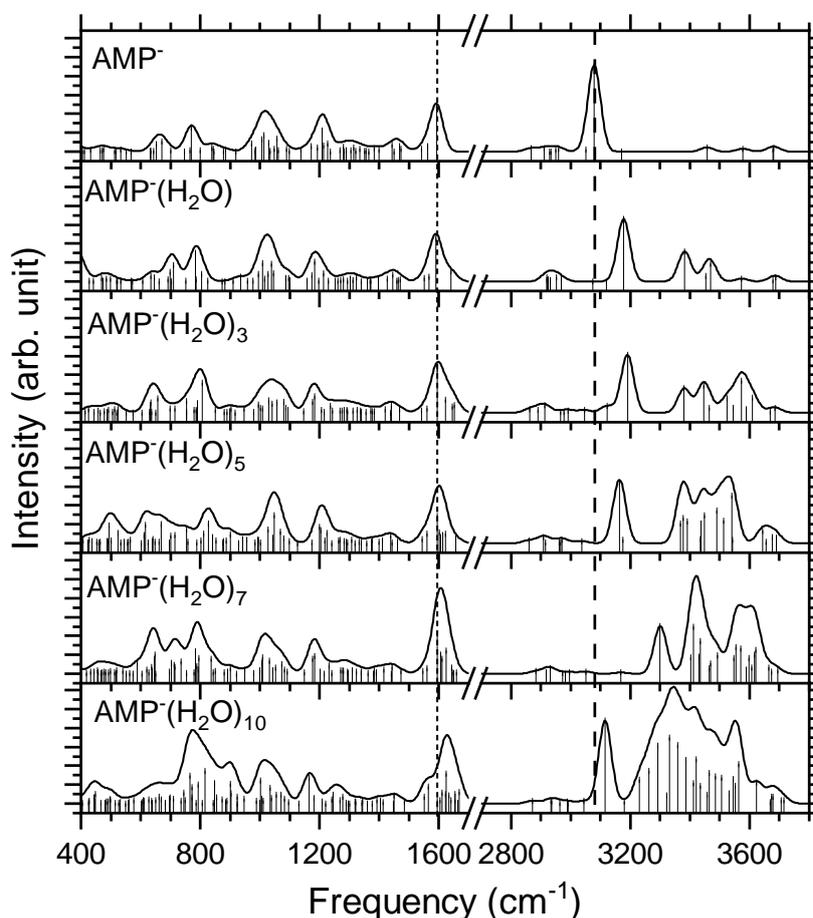


Figure 4.4.2: Calculated IR spectra of AMP⁻(H₂O)_m m=0,1,3,5,7, and 10 at the B3LYP/ZORA-TZVP level of theory, in the region 400 to 1800 cm⁻¹ and 2800 to 3800 cm⁻¹ (scaling factor 0.965, FWHM = 50 cm⁻¹). Dashed lines have been added at 1600 cm⁻¹ and 3080 cm⁻¹ to facilitate reading. The scale on the intensity axis is the same for all computed IR spectra.

Figure 4.4.2 shows the calculated IR spectra of AMP⁻(H₂O)_m, m=0,1,2,3,7, and 10, in the region 400 – 3800 cm⁻¹. In this section, no shift was applied to the computed values for better comparison between the spectra. The assignment of the different vibrational modes participating in the calculated IR spectra and the calculated bands'

centroid are summarized in Table 4.1.

In the fingerprint region, 400 to 1800 cm^{-1} , the spectra are mostly similar, with the bands at 660 cm^{-1} and 790 cm^{-1} varying with the number of water molecules. The band at 660 cm^{-1} is blueshifted by 40 cm^{-1} for $\text{AMP}^-(\text{H}_2\text{O})$ and redshifted back to 640 cm^{-1} for higher solvation state. The band at 790 cm^{-1} is blueshifted by 30 cm^{-1} for all hydration states up to $\text{AMP}^-(\text{H}_2\text{O})_7$. In the case of $\text{AMP}^-(\text{H}_2\text{O})_{10}$, this band is broader and more intense and appears at the same frequency as the isolated structure, i.e. 790 cm^{-1} . The band stems from vibrations of a combination of OH wagging and twisting, mostly from the water molecules but also from an OH vibration from the sugar group. This explains why the band at 790 cm^{-1} is more intense for $\text{AMP}^-(\text{H}_2\text{O})_{10}$, since $\text{AMP}^-(\text{H}_2\text{O})_{10}$ has the most water molecules. For all clusters, the bands at 1000 cm^{-1} and 1200 cm^{-1} stem from CO bending modes and CH rocking modes, respectively (see Table 4.1).

For the hydrated clusters with more than two water molecules, the band observed at 1600 cm^{-1} can be attributed mainly to an OH bending mode from the different water molecules with a small contribution from an NH bending mode in the case of $\text{AMP}^-(\text{H}_2\text{O})_{10}$. This band is red-shifted in isolated AMP^- and $\text{AMP}^-(\text{H}_2\text{O})$ and comes from the NH bending modes of the adenine group. For hydrated clusters, the NH bending modes of the adenine group are also present at 1600 cm^{-1} , but the intensity of this frequency is negligible compared to the intensity of the OH bending modes.

More differences are observed between the clusters in the region 2800 – 3800 cm^{-1} . For bare AMP^- , only one band arises at 3080 cm^{-1} , which can be attributed to the OH stretch of the OH group of the ribose. This band appears to be blue-shifted for small hydrated clusters (five water molecules) by less than 100 wavenumbers. However, for $\text{AMP}^-(\text{H}_2\text{O})_7$, this band is blue-shifted by 220 wavenumbers, appearing at 3300 cm^{-1} with lower intensity. For $\text{AMP}^-(\text{H}_2\text{O})_{10}$, this band is not observed. At higher frequencies, a broad feature arises between 3380 and 3660 cm^{-1} for all hydrated species. This feature becomes broader with an increasing number of water molecules and corresponds to the OH stretchings of the various water molecules in the cluster. The behavior of $\text{AMP}^-(\text{H}_2\text{O})_{10}$ shows a distinct characteristic in its vibrational spectrum. The observed band at 3115 cm^{-1} does not originate from a blue-shifted OH stretch of the ribose, as identified for the other clusters. Instead, this 3115 cm^{-1} band arises from the symmetric OH stretching of water molecules within the water cluster. At higher frequencies, the broad feature between 3250 cm^{-1} and 3600 cm^{-1} corresponds to OH stretchings from the different water molecules involving inter-hydrogen bonds with AMP^- .

Table 4.1: Calculated absorption bands (in cm^{-1}) for the different clusters of $\text{AMP}^-(\text{H}_2\text{O})_m$ at DFT/B3LYP/ZORA-TZVP level (scaling factor of 0.965).

m	CO bending	CH rocking	OH stretching / AMP^-	OH stretching / water
0	1000 - 1010	1210	3080	
1	1010 - 1015	1170 - 1185	3177	3380
3	1000 - 1010	1180	3190	3380 - 3520
5	1050 - 1070	1200 - 1220	3162	3368 - 3643
7	1000 - 1030	1180 - 1215	3300	3420 - 3600
10		1180		3250 - 3650

In conclusion, in the fingerprint region, the general shape of the IR spectrum for each structure is similar, with the same bands appearing. However, frequency shifts can be observed, which allows us to differentiate the different hydrated clusters. Moreover, for $\text{AMP}^-(\text{H}_2\text{O})_{10}$, a new band arises at 790 cm^{-1} corresponding mostly to OH vibrations from the water molecules not interacting with the nucleotide. More differences can be observed in the -OH stretch region between 3000 cm^{-1} and 3600 cm^{-1} . In the case of isolated AMP^- , only one main band is observed, corresponding to -OH stretches of the hydroxyl groups of the ribose. This band is blueshifted for all hydrated clusters up to $\text{AMP}^-(\text{H}_2\text{O})_7$. For $\text{AMP}^-(\text{H}_2\text{O})_{10}$ this band is not present. Another band arises at 3115 cm^{-1} for the symmetric OH stretching of the different water molecules within the water cluster. A broad feature appears after already one water molecule is added to the system between 3200 and 3600 cm^{-1} . This corresponds to OH stretching within the water cluster. However, for $\text{AMP}^-(\text{H}_2\text{O})_{10}$, this corresponds to OH stretching from only the water molecules having inter-HB with AMP^- whereas for all other hydrated clusters, this corresponds to OH stretching from any water molecules. Having studied the influence of water molecules on the 3D structure in vibrational spectra of AMP^- , we will now further investigate the effects of water on the local electronic structure of deprotonated AMP by calculating the X-ray absorption spectra at the oxygen K-edge.

4.5 Calculated absorption spectra in the soft X-ray regime

In this section, the calculated oxygen K-edge X-ray absorption spectra (XAS) of AMP^- , isolated and hydrated with up to 10 water molecules, are presented and com-

pared. The X-ray absorption spectra were calculated using the ORCA program at the DFT/ROCIS level using the B3LYP functional with the TZVP basis set; relativistic effects were considered using the zeroth-order regular approximation [158] (more details can be found in Chapter 2). The analysis of the different XAS spectra was done by considering the main transition lines for the attribution of the main resonances, as well as the low-intensity transitions, in order to disentangle the influence of the water molecules.

Figure 4.5.1 shows the calculated absorption spectra of $\text{AMP}^-(\text{H}_2\text{O})_m$ $m=0,1,3,5,7$ and 10 at the oxygen K-edge. To obtain a deeper insight into the contribution of the different excitations of the oxygen atoms composing the $\text{AMP}^-(\text{H}_2\text{O})_m$ clusters, the transition lines were divided into three groups according to the atomic orbitals from which the core electrons are excited. The blue spectra represent the contribution of the oxygen orbitals of the water molecules. The contribution of the oxygen atoms of the ribose and the phosphate groups are shown in pink and orange, respectively. All spectra were shifted by +14 eV to account for the B3LYP functional energy shift and convoluted with Gaussian functions with a FWHM of 0.6 eV, as commonly performed [136]. The energy shift is the same for all spectra to be able to compare possible shifts in energy for the different transitions depending on the number of water molecules added to isolated AMP^- .

When looking at the total XAS of the isolated nucleotide (top window in Figure 4.5.1) at first glance, its shape is mainly due to oxygen excitations from the phosphate group. In particular, the contribution of the first resonance located around 533 eV is shared between the three chemical groups of AMP^- . It corresponds to the $\text{O } 1s \rightarrow \sigma^*(\text{OH})$ transitions in the phosphate and ribose groups and to an $\text{O } 1s \rightarrow \pi^*(\text{CN})$ transition in the adenine moiety. The $\text{O } 1s \rightarrow \pi^*$ transition was also observed for a protonated oligonucleotide $^{\text{F}}\text{UAG}$ at the oxygen K-edge [74]. The broad feature between 535 and 539 eV is mainly due to $\sigma^*(\text{CH})$ and $\sigma^*(\text{NH})$ in the adenine group coming from excitations of the $\text{O } 1s$ orbitals from the ribose and phosphate groups. Table 4.2 summarizes the transition energies of the $\text{O } 1s \rightarrow \pi^*(\text{CN})$ and $\text{O } 1s \rightarrow \sigma^*(\text{OH})$ from the phosphate group, sugar group, and water molecules for all calculated structures discussed in this section.

When the solvent molecules are added to AMP^- , the overall shape of the total spectra remains similar to the isolated spectrum (see the black curves in Figure 4.5.1). The first resonance is observed around 533 eV. Between 534 and 537 eV, there is a broad feature that increases with the number of water molecules. However, both the 533 eV resonance and the broad feature are changing differently depending on the number of

water molecules added.

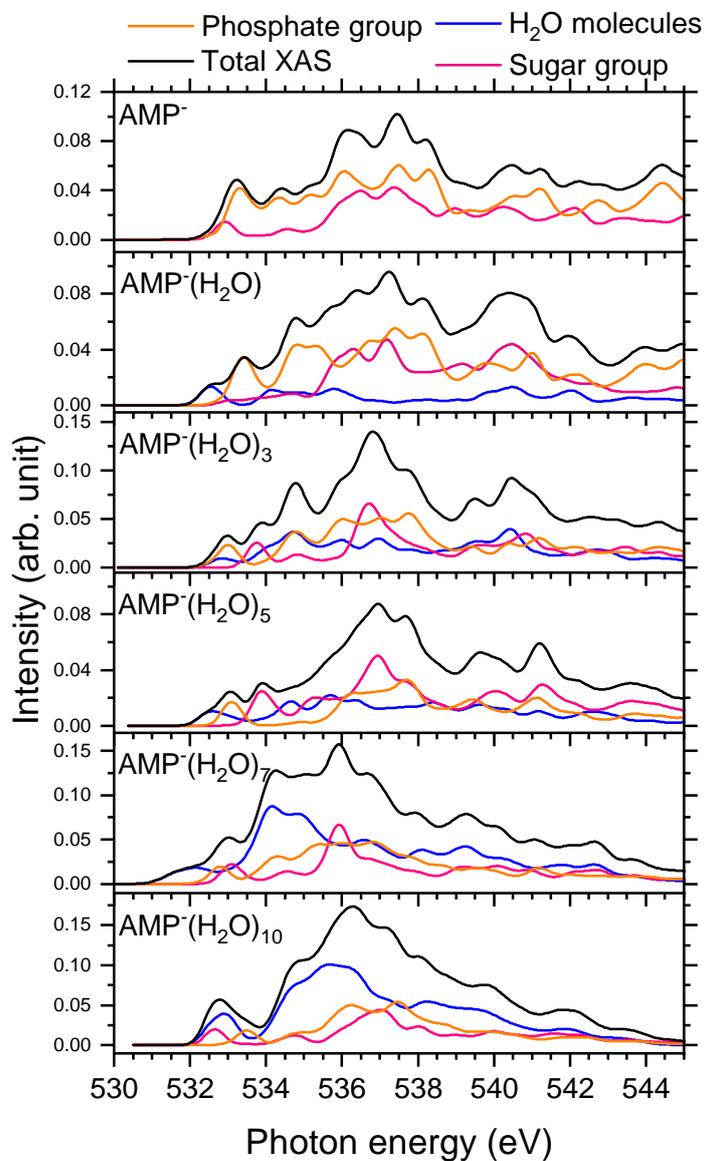


Figure 4.5.1: DFT/ROCIS absorption spectra of $\text{AMP}^-(\text{H}_2\text{O})_m$ $m=0,1,3,5,7$ and 10 at the oxygen K-edge. The absorption lines were broadened using a Gaussian profile with a 0.6 eV FWHM. The total calculated XAS is shown in black. The different contributions from the water molecules, sugar, and phosphate groups are represented in blue, pink, and orange, respectively. A shift of $+14$ eV was applied to consider the energy shift due to the B3LYP functional.

Up to seven water molecules, a splitting of the resonance at 533 eV is observed. For $\text{AMP}^-(\text{H}_2\text{O})_1$ and $\text{AMP}^-(\text{H}_2\text{O})_7$, the splitting is due to the additional transitions of $\text{O } 1s \rightarrow \sigma^*(\text{OH})$ from the water molecules at 532.5 eV (see the blue curves in Figure 4.5.1). However, for three and five water molecules added to AMP^- , the splitting does not come from the addition of the $\sigma^*(\text{OH})$ transition of the water molecules but from the splitting of the electronic contributions from the ribose and the phosphate (see the pink and orange curves in Figure 4.5.1). For instance, for $\text{AMP}^-(\text{H}_2\text{O})_3$ the transition from an O 1s orbital to the $\sigma^*(\text{OH})$ orbital of the phosphate is redshifted by 0.3 eV whereas the $\sigma^*(\text{OH})$ transition in the ribose group is blue-shifted by 0.7 eV compared to isolated AMP^- , similarly to $\text{AMP}^-(\text{H}_2\text{O})_5$ in which the same redshift is observed for the $\sigma^*(\text{OH})$ transitions of the phosphate group and a blueshift of 0.8 eV is observed from the $\sigma^*(\text{OH})$ resonance of the ribose.

Moreover, we could not observe a discernible contribution of the water molecules to the X-ray spectra at higher photon energies until at least seven water molecules are added to AMP^- .

As shown in Figure 4.5.1, the contribution of the water molecules consists of a number of low-intensity transitions with negligible participation to the total XAS spectra, as no intense transitions stemming from the water molecules were found for small clusters (below seven water molecules) during the analysis of the different transitions. However, in the case of $\text{AMP}^-(\text{H}_2\text{O})_7$ and $\text{AMP}^-(\text{H}_2\text{O})_{10}$, the contribution of water become higher in intensity, enough to create new features in the total absorption spectra, such as the band at 535 eV in the XAS of $\text{AMP}^-(\text{H}_2\text{O})_7$. This is due to the fact that all low-intensity transitions from the different oxygen atoms of the ten water molecules start to compete with the more intense transitions of the sugar and phosphate moieties. It is worth noting that, as observed in Figure 4.3.3, the threshold in which the number of intra-HBs becomes higher than the number of inter-HBs is also seven water molecules. The change of the XAS spectra for a level of hydration higher than seven water molecules could be explained by the creation of a water droplet after seven water molecules. This would then be responsible for the stronger water-related bands observed between 534 and 536 eV that would appear in the XAS spectrum only for a water droplet of sufficient size (here above seven water molecules).

Interestingly, the total calculated spectrum of $\text{AMP}^-(\text{H}_2\text{O})_{10}$ is similar to the experimental XAS spectrum of a solid-state DNA sample [161], [162].

We can, thus, conclude that when more than six water molecules are added to AMP^- , the low-intensity transitions cannot be neglected when studying the total calculated XAS. The question remains if this can be probed experimentally or if it is an artifact of the calculations. However, the shift of 0.7 eV in the resonance at 533 eV,

Table 4.2: Calculated energy transition for the different clusters of $\text{AMP}^-(\text{H}_2\text{O})_m$ at the DFT/B3LYP/ZORA-TZVP level.

m	$\pi^*(\text{CN})$	$\sigma^*(\text{OH})$ phosphate	$\sigma^*(\text{OH})$ ribose	$\sigma^*(\text{OH})$ water
0	533 eV	533.3 eV	533 eV	
1	534.7 eV	533.5 eV	535.7 eV	532.5 eV
3		533 eV	533.7 eV	533 eV (very weak)
5		533 eV	533.8 eV	532.5 eV (very weak)
7	531.6 eV	532.8 eV	533.1 eV	532 eV and 535 eV (very weak)
10	533 eV		532.6 eV	532.5 eV

as well as the difference in broadness and intensity of the 534 - 537 eV feature, can be probed and is already a clear signature of the water adducts.

4.6 Conclusion

In conclusion, this chapter has dealt with the influence of stepwise hydration of a deprotonated nucleotide, adenosine monophosphate, employing density functional theory at the B3LYP/TZVP level. The geometric structures of hydrated clusters with up to twenty water molecules have been investigated, and their dissociation energies determined. We showed that the isolated AMP^- has a more linear structure, whereas the interaction of AMP^- with water molecules folds the structure of the nucleotide. We could define two types of hydrogen bonds: intramolecular HBs, i.e., HB within the water droplet, and inter-HBs, which are the HBs between the water droplet and AMP^- . The dissociation energy of the full water droplet follows the inverse trend of the number of intra-HBs. However, after eight water molecules attached to AMP^- , we found that water molecules do not bind to AMP^- anymore but to the other water molecules forming a water droplet of increasing size.

From the vibrational spectra, we revealed that the fingerprint region is composed of the same bands for all clusters with shifts in energy of the bands at 660 cm^{-1} , 790 cm^{-1} and 1600 cm^{-1} . However, the OH stretch region, between 2800 cm^{-1} and 3800 cm^{-1} , possesses more differences between the clusters. We showed that the band corresponding to the OH stretches of the sugar group is blueshifted for all hydrated clusters compared to the isolated AMP^- and is not present in the case of $\text{AMP}^-(\text{H}_2\text{O})_{10}$. A band of increasing size and intensity arose between 3200 and 3600 cm^{-1} corresponding to the OH stretching within the water cluster. Additionally to the information on the 3D

structure and vibrational spectra of hydrated AMP⁻, the X-ray absorption spectra at the oxygen K-edge of chosen hydrated structures were calculated to understand possible energy transfer between the water molecules and the deprotonated nucleotide. We determined that below seven water molecules, the presence of water molecules is not reflected at high photon energy but only in the first resonance appearing at 533 eV for isolated AMP⁻. A resonance due to O 1s → $\sigma^*(\text{OH})$ transitions from the water molecules is responsible for a low-intensity resonance at around 532.5 eV that is not present in the isolated AMP⁻ spectrum. At a higher solvation level, the number of O 1s → $\sigma^*(\text{OH})$ transitions due to water molecules increased significantly compared to the lower solvation states. The increased number of low-intensity transitions is responsible for the broad feature observed in the total calculated spectra of AMP⁻(H₂O)₇ and AMP⁻(H₂O)₁₀ between 534 eV and 536 eV.

With this study, we demonstrated that the differences in the structure of AMP⁻ induced by water can be extracted from the vibrational spectra and the XAS spectra at the oxygen K-edge. Future experimental data must show whether the structural influence of the water molecules on the nucleotide can actually be determined, for example, whether the low-intense transitions of the water droplet can be resolved.

5 | Influence of a single water molecule on the X-ray absorption spectra of a phosphorylated amino acid

5.1 Introduction

Understanding solvation effects represents a fundamental aspect of chemistry, as the structure and reactivity of molecules in a solvent environment differ significantly from those in vacuum. A single water molecule can already alter the protonation site of a tripeptide [163]. Interactions with solvent molecules can modify the relative stability. It has been shown for a diprotonated organic molecule fully linear in the gas phase that the most stable structure is folded when it interacts with water molecules [55]. The effect of a few water molecules on the structure and the reaction dynamics of amino acids have been studied theoretically [23], [96], [97], as well as experimentally using mass spectrometry combined with infrared (IR) and ultra-violet (UV) spectroscopy [16], [56], [164].

Additionally, soft X-ray light can be employed to understand the influence of solvents on the properties of biomolecules. Over the last decades, soft X-ray spectroscopy has become a valuable tool to study the electronic structures of biomolecular systems in the condensed [165], [166] and gas phases [28], [86]. Soft X-ray excitation is element-specific, allowing site-selective excitation of a molecule to specific unoccupied molecular orbitals by tuning the photon energy to a resonance. [76]. Recently, Milosavljević et al. performed oxygen K-shell spectroscopy of an isolated protonated peptide (substance P) solvated with four and eleven water molecules, formed using electrospray ionization [17]. They showed that following resonant absorption of X-rays in the core level of the oxygen atoms of the water molecules, the peptide backbone fragmentation is increased. This suggests a charge or energy transfer, such as in intermolecular Coulombic decay (ICD), from the solvent to the peptide. However, Milosavljević et al. also studied

the effect of three water molecules on the stabilization of a peptide dimer upon VUV irradiation and demonstrated that three water molecules already stabilize the dimer by about 1.5 eV and, thereby, prevent damage by the VUV photons [19]. This shows that the mechanisms behind the nanosolvation effect on photostability are not yet well understood.

The phosphorylation of the alcohol group on the side chain of tyrosine, serine, or threonine is one of the most frequent post-translational modifications (PTM) of proteins [34]. The replacement of the -OH group by a phosphate group H_3PO_4 results generally in the modification of the protein's biological functions. While Milosavljević et al. focused on peptides consisting of natural amino acids [17], [19], it is also important to understand the influence of the solvent on the structure and fragmentation process of phosphorylated amino acids and proteins. To understand the effect of a single solvent molecule, we studied protonated phosphotyrosine (hereafter written as pTyr) in the gas phase as a model for phosphorylation. Phosphotyrosine is an amino acid where the -OH group of the benzene ring of tyrosine is replaced by a phosphate group H_3PO_4 . Its chemical structure is shown in Figure 5.1.1. The simple presence of the phosphate group is likely to direct water attachment toward it, as shown in Chapter 4 for deprotonated AMP. pTyr has already been studied by infrared multiple photon dissociation (IRMPD) spectroscopy by Scuderi et al. [107]. They have shown that adding a water molecule to pTyr redshifts the ammonium NH stretches and the phosphate P=O stretch, which they could link to the formation of a strong hydrogen bond bridge between the phosphate group and the ammonium through the water molecule. In this manuscript, we extend this work by studying the influence of a single water molecule on the X-ray fragmentation and electronic structure of pTyr by analysis of the soft X-ray spectra at the carbon and oxygen K-edges.

This chapter is divided as follows: first, the experimental and theoretical methods are explained. Then, the experimental mass spectra and X-ray absorption spectra at the carbon and oxygen K-edges are described. In the last section, a discussion of the potential structural candidates is done in light of our experimental results.

5.2 Methods

5.2.1 Experimental method

Phosphotyrosine was purchased from Sigma-Aldrich at $\geq 95\%$ purity and used without further purification. The sample solution was prepared at 1.5 mM concentration in 1:1 water/methanol with 1 % of formic acid to ensure the protonation.

The results were obtained during an experimental campaign by coupling the Spektrolatius setup (described in detail in section 3.1) to the P04 soft X-ray beamline (described in subsection 3.3.1) at the PETRA III synchrotron, Hamburg, Germany. Briefly, the setup relies on a high-fluence electrospray ionization (ESI) source to introduce the ions into the gas phase. The electrosprayed ions are transported by a capillary into the first pumping stage, which hosts two skimmers that focus the ions. The second skimmer plays the role of conductance limit for differential pumping between the ESI source at atmospheric pressure and the next stage at 10^{-3} mbar. The ions continue through a radio frequency (RF) octupole ion guide to a quadrupole mass filter where the ions of interest are selected according to their mass-to-charge (m/z) ratio. Then, a second RF octupole and a set of Einzel lenses guide and focus the ions into an RF 3D ion trap, where they are accumulated before irradiation with photons. To allow efficient trapping of the ions, their kinetic energy is cooled down to room temperature by collisions with a helium buffer gas. After irradiation with X-ray photons for 2 seconds, all cationic products are eventually extracted from the trap and analyzed in a reflectron time-of-flight (Re-TOF) mass spectrometer ($m/\Delta m = 1800$). At the carbon (C) K-edge, two sets of data were recorded: one over the range from

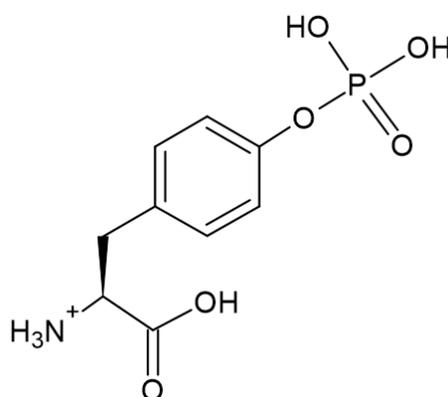


Figure 5.1.1: Chemical structure of protonated phosphotyrosine.

284 eV to 291 eV with 0.2 eV steps and an energy bandwidth of 250 meV and a second from 284.55 eV to 286.75 eV with 0.05 eV steps and an 81 meV bandwidth to improve the spectral resolution around the first resonance at 285 eV. At the oxygen (O) K-edge, the energy resolution was set to 80 meV, and the photon energy scan was performed from 528 eV to 545 eV with steps of 0.1 eV. A photodiode placed downstream of the ion trap was used to record the photon flux.

The NEXAMS spectra obtained were normalized to the photon flux and to the precursor ion intensity. To account for all sources of background ions and average over long-term fluctuations of the ESI source, the data acquisition was divided into repeated cycles of two mass scans. First, the mass spectrum resulting from the irradiation of the trapped molecules was recorded. For the second cycle, the ESI was switched off, and the mass spectrum resulting from the irradiation of the residual gas was taken and subtracted from the mass spectrum resulting from the irradiation of the trapped molecules. The final mass spectrum is a series of a hundred accumulated cycles.

5.2.2 Theoretical calculations

To assign the different resonances of the NEXAMS spectra at the C K-edge and O K-edge, theoretical calculations were performed to compute the X-ray absorption spectra (XAS) of structural candidates for pTyr and hydrated pTyr.

First, possible geometries of isolated pTyr were calculated by means of the DFT method. Twenty starting structures were generated from the PubChem database [167] and fully optimized employing the TZVP Ahlrichs basis set [157] in combination with the B3LYP exchange-correlation functional [123]. More details about the B3LYP functional can be found in Chapter 2. All *ab initio* calculations have been performed using the ORCA electronic structure package [127]. All optimized conformers were then ranked according to their root mean square deviation (RMSD) values. The RMSD values with respect to each conformer have been calculated using the VMD software after alignment of every conformer on their carbon atoms to ensure proper calculation of the RMSD value [168], [169]. It simply measures the average distance between the atoms of superimposed molecules. High RMSD values mean a large distance between the atoms, i.e., structures different from one another, whereas a low RMSD value means a small distance between the atoms and, thus, similar structures.

Finally, five structures were chosen for the isolated pTyr according to their final (lowest) DFT energies and RMSD values (see Figure 5.2.1). The lowest energy conformer (pTyr_1) has been chosen as a reference for the calculations of the RMSD values. The four other structures (pTyr_2 to pTyr_5) were chosen to be at a relative electronic en-

ergies of maximum +0.4 eV with different conformations according to the calculated RMSD values. The X-ray absorption spectra (XAS) at the C and O K-edges were calculated for these five structures and later compared with our experimental spectra.

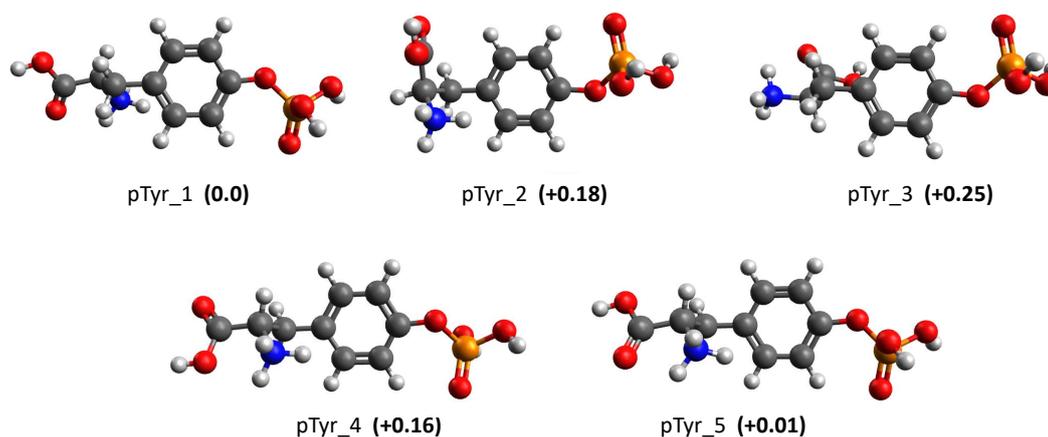


Figure 5.2.1: Representation of the five isolated structures obtained at the B3LYP/TZVP level of theory, their relative electronic energies are shown in parentheses in eV.

To obtain hydrated structures, initially, a solvation shell was created around the phosphate group of the five selected isolated structures. Subsequently, the water shell was narrowed down to five possible hydrated pTyr configurations for each isolated structure. For this, an 8.0 Å shell was first created around the whole molecule using the CHIMERA software [170]. The full solvation shell was further reduced to a 6 Å shell around the phosphate group, as will be explained later. The positions of the water molecules in the five initial systems were optimized at the PM7 level of theory using the MOPAC software (the pTyr molecule was frozen during this calculation) [171]. The five structures with the optimized solvation shell are shown in Figure 5.2.2.

From these five solvated structures, 24 singly hydrated structures were chosen and optimized before calculating their XAS at the carbon and oxygen K-edges. The different singly hydrated structures obtained will be discussed in subsection 5.3.3. To compare with the experimental results, the final X-ray absorption spectra were broadened using a Gaussian profile with a full width at half-maximum (FWHM) of

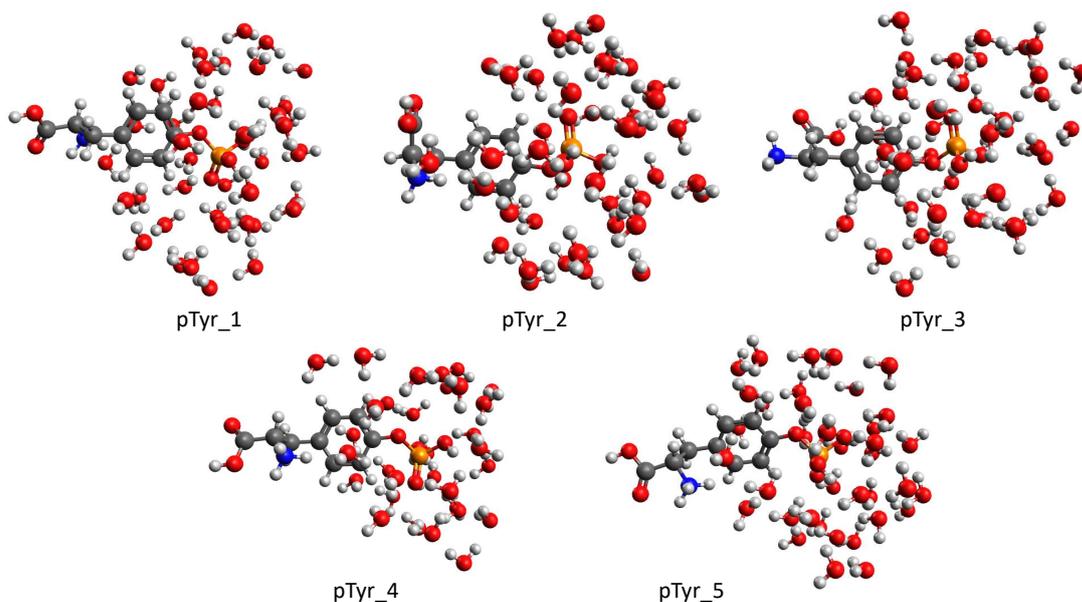


Figure 5.2.2: Representation of the solvation shell around the phosphate group for the five structures after optimization of the position of the water molecules with MOPAC.

0.6 eV unless otherwise stated [91]. The energy shift of the theoretical spectra has been defined by matching the first resonance of the theoretical spectra to the corresponding one in the experimental spectra, representing $\pi^*_{C=C}$ in the case of the C K-edge calculated spectra and $\pi^*_{C=O}$ in the case of the O K-edge calculated spectra, as will be explained later [172]. This gives a shift of around 11.5 eV in the case of the C K-edge and 14 eV in the case of the O K-edge calculations. This shift is due to the B3LYP functional underestimation of the core excitation energies [159], [173].

5.3 Results and discussion

5.3.1 Mass spectra

In this section, the ESI-only spectrum (no photons) of $[pTyr+H+H_2O]^+$ as well as the photodissociation spectrum of $[pTyr+H+H_2O]^+$ at the O K-edge will be discussed.

The ESI-only spectrum, recorded without photons, is the result of the collisional activation of the molecule with the helium buffer gas. In the ESI-only mass spectrum reported in Figure 5.3.1, we observed multiple neutral losses from hydrated pTyr. The main neutral loss observed is the loss of H_2O from the selected parent ion, leading

to isolated pTyr. Although only the hydrated phosphotyrosine was mass selected in the QMS, the intensity of isolated pTyr was almost equal to that of hydrated pTyr due to collision with the helium buffer gas in the trap. Moreover, bare pTyr could further fragment in the trap before photon irradiation. A corresponding fragment from isolated pTyr at m/z 216 results from the loss of H_2O and CO as already observed by Scuderi et al. [107]. Interestingly, this sequential loss of H_2O and CO from the carboxyl group is not observed in the case of hydrated pTyr (it would appear at m/z 234), indicating that the addition of a water molecule non-covalently bound to the amino acid affects the possible fragmentation channels.

From hydrated pTyr, we observed the loss of HPO_3 at m/z 200 and the subsequent loss of H_2O at m/z 182. The loss of HPO_3 is a characteristic neutral loss in phosphorylated amino acids and peptides [51], [174]. We suggest that m/z 182 does not come from the loss of HPO_3 for the bare pTyr. If that was the case, then the intensity of fragment m/z 182 would have been the same as m/z 200. Since the intensities of bare pTyr and hydrated pTyr were nearly equal, we could have expected the intensity of HPO_3 loss to be the same for both molecules. However, we could not rule it out completely from the ESI-only spectrum as it could also be that hydration of phosphotyrosine enhances the loss of HPO_3 .

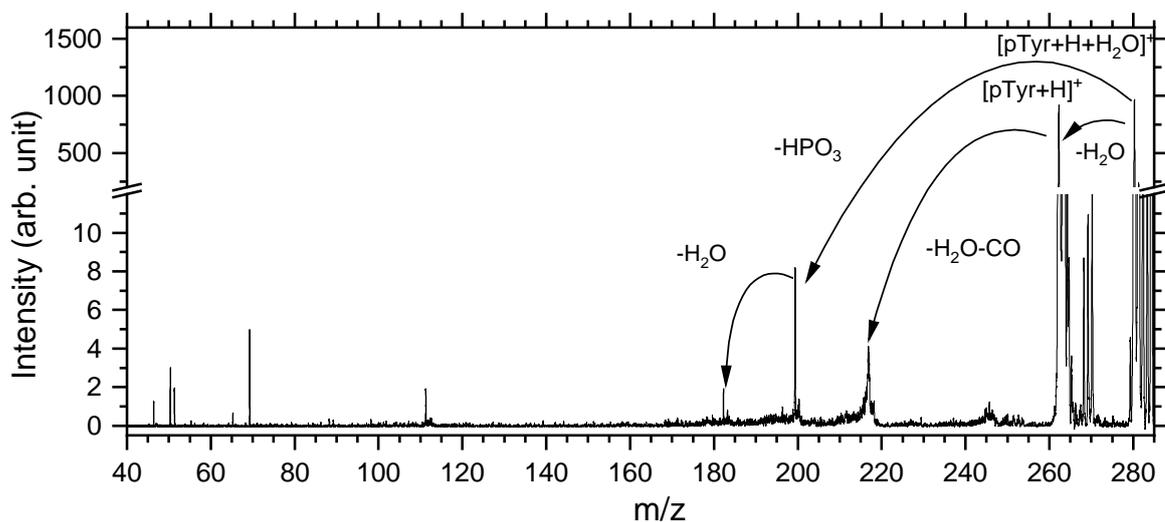


Figure 5.3.1: ESI-only mass spectrum of hydrated phosphotyrosine. The precursor ion $[\text{pTyr}+\text{H}+\text{H}_2\text{O}]^+$ is observed at m/z 280.

5. Influence of a single water molecule on the X-ray absorption spectra of a phosphorylated amino acid

The mass spectrum revealed peaks in the low mass region (m/z 45 to m/z 115) that could not be assigned to fragments of pTyr or its hydrated forms. These peaks likely originated from contaminants in the ESI solution that may have further broken down upon entering the instrument and/or ion trap. While these contaminants could not be eliminated during the experiments, they could be distinguished from pTyr because they did not match the expected fragmentation pattern of pTyr [175], [176]. We typically expect benzene fragments from the tyrosine side-chain [175] and phosphate losses [176]. However, subtraction of the ESI-only spectrum from the photodissociation spectrum (see Figure 5.3.2) revealed that the only fragments present were those characteristic of pTyr and hydrated pTyr irradiated at 536.3 eV, clearly indicating $[\text{pTyr}+\text{H}+\text{H}_2\text{O}]^+$ and $[\text{pTyr}+\text{H}]^+$ as the parent ions.

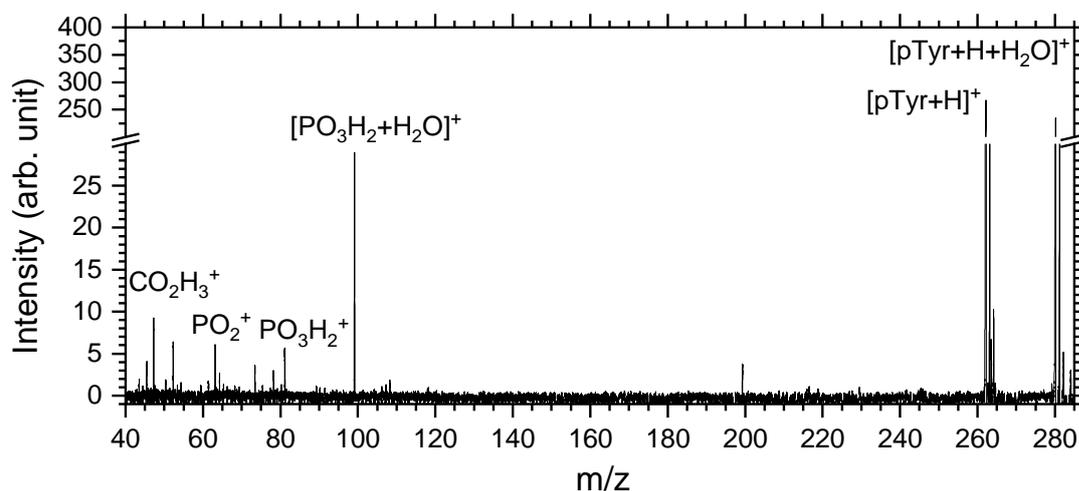


Figure 5.3.2: Photodissociation mass spectrum of the mixture of hydrated and bare pTyr measured at 536.3 eV.

We observed 13 fragments in total from isolated and hydrated pTyr. Table 5.1 summarizes the attribution of the different fragments observed after photon irradiation. Most fragments could be attributed to tyrosine side-chain fragments such as m/z 52, 65, 78, 80, 107, and 108. Fragments at m/z 63, 81, and 99 are attributed to fragments of the phosphate group, PO_2^+ , PO_3H_2^+ and $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ while, m/z 47 and 62 are backbone fragments, respectively CO_2H_3^+ and $\text{C}_2\text{O}_2\text{H}_6^+$. Based solely on this mass spectrum, it is not possible to conclude whether PO_3H_2^+ is produced by the evaporation of the water molecule from $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ or directly from fragmen-

tation of the isolated $[\text{pTyr}+\text{H}]^+$. This will be further discussed in the context of the partial yield spectra at the O K-edge (see next section). Note that the observation of the hydrated fragment $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ reinforces the conclusion that the water molecule binds to the phosphate group in phosphotyrosine. In the photodissociation mass spectrum, we could identify this fragment unambiguously as originating from the hydrated pTyr, which means that despite the large internal energy (> 20 eV) after soft X-ray photoabsorption [73], [87], the non-covalent bond between pTyr and the water molecule is conserved. This is different from collision-induced fragmentation in an ion trap where for the deprotonated nucleotide AMP, it has been shown that under collisions with neutrals, the hydrated molecule first undergoes loss of the water molecule and then further fragments [14]. However, hydrated fragments have been observed after irradiation at the oxygen K-edge of a hydrated peptide with four water molecules [17].

It is also worth mentioning that the complementary fragment of $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ is the protonated tyrosine amino acid minus one hydrogen atom, that should appear at m/z 181. This fragment was absent from our mass spectra, suggesting that the amino acid undergoes further dissociation after the loss of the $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ fragment. All fragments mentioned in this section have been used to obtain the total ion yield spectra shown in the next section.

Table 5.1: Attribution of the different fragments produced by the irradiation of $[\text{pTyr}+\text{H}+\text{H}_2\text{O}]^+$ and $[\text{pTyr}+\text{H}]^+$ with photons at 536.3 eV.

m/z	Attribution
47	CO_2H_3^+
50	PH_3O^+
51	C_4H_3^+
52	C_4H_4^+
62	$\text{C}_2\text{O}_2\text{H}_6^+$
63	PO_2^+
65	C_5H_5^+
78	C_6H_6^+
80	C_6H_8^+
81	PO_3H_2^+
99	$[\text{PO}_3\text{H}_2 + \text{H}_2\text{O}]^+$
107	$\text{C}_7\text{H}_7\text{O}^+$
108	$\text{C}_7\text{H}_8\text{O}^+$

We can already draw some conclusions based on the ESI-only mass spectrum and the photodissociation mass spectrum. The fragmentation channels between the isolated and the hydrated pTyr appeared different. From hydrated pTyr, we could observe the loss of HPO_3 with subsequent loss of H_2O . The loss of HPO_3 from bare pTyr was negligible in our experimental conditions. Furthermore, the sequential losses of H_2O and CO from the carboxyl group have not been observed in the case of hydrated pTyr but in the case of bare pTyr. This shows that adding a single water molecule to pTyr readily affects the fragmentation channels.

Moreover, the fact that the only hydrated fragment is $[\text{PO}_3\text{H}_2 + \text{H}_2\text{O}]^+$ indicated that the water molecule binds to the phosphate group. Furthermore, observing hydrated fragments after irradiation with soft X-ray photons suggests a fast, non-statistical fragmentation process. This fragmentation might occur prior to the intramolecular vibrational energy redistribution (IVR) of the deposited energy, as IVR would lead first to the cleavage of weak non-covalent bonds. However, the loss of HPO_3 from hydrated pTyr can be observed without the loss of the non-covalently bound water molecule, suggesting the water molecule is not only bound to the phosphate group. These two statements suggest that the water molecule forms a bridge between the phosphate group and another chemical group of the amino acid. The next section

presents and analyzes the experimental X-ray absorption spectra, considering the results of the DFT calculations. Additionally, it discusses the potential conformers of hydrated pTyr.

5.3.2 X-ray absorption spectra

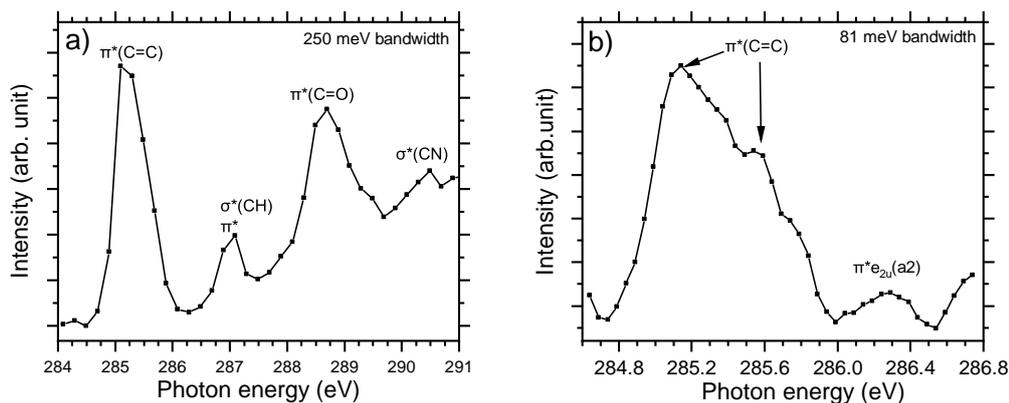


Figure 5.3.3: Total ion yield (TIY) spectra at the C K-edge of the mixture of hydrated and bare pTyr. a) The full range measurement; b) the high-resolution TIY spectrum is shown.

The X-ray absorption spectra at the C K-edge of the mixture of bare and hydrated pTyr are shown in Figure 5.3.3. At the C K-edge, we observed three main resonances, as can be seen in Figure 5.3.3.a: the first one is at 285 eV and originates from a C $1s \rightarrow \pi^*(C=C)$ transition from the benzene ring. We observed in the high-resolution scan around 285 eV that the $\pi^*(C=C)$ resonance is split into two resonances separated by 0.4 eV (see Figure 5.3.3.b). This is explained by the fact that the molecular environment is not identical for the different carbon atoms of the ring, as has been described in the X-ray absorption spectrum of neutral benzene and gas-phase neutral tyrosine [177], [178]. A second, weaker resonance can be observed in Figure 5.3.3.b, located at 286.2 eV. Theoretical calculation has attributed this to a dipole-forbidden C $1s \rightarrow \pi^*e_{2u}(a2)$ resonance, which is vibronically allowed due to symmetry breaking of the benzene ring [179]. This has also been observed experimentally by Püttner et al. using photo-electron spectroscopy [178] and well described with high-resolution spectroscopy by Rennie et al. [180].

The second resonance present in Figure 5.3.3.a, at 287.3 eV, is a combination of

two different transitions located at 287.1 eV and 287.4 eV due to $C_{\text{ring}} 1s \rightarrow \pi^*$ and $C_{\text{ring}} 1s \rightarrow \sigma^*(\text{CH})$ respectively. This resonance has been observed for solid-state tyrosine and tyrosine-containing peptides [77], [86], [181]. The last main resonance at 288.7 eV has been attributed to $C 1s \rightarrow \pi^*$ transition from the benzene ring [177], [178], more precisely to $C 1s \rightarrow \pi^*(\text{C}=\text{O})$ transitions in small peptides in solid phase [80], [82] and gas-phase [86], [89]. Accordingly, our calculations indicated that the resonance at 288.7 eV predominantly originates from $C 1s \rightarrow \pi^*(\text{C}=\text{O})$ transitions from the carboxyl group (see in the next section Figure 5.3.12).

A very weak resonance is observed in the TIY (Figure 5.3.3) and for most partial ion yields (PIYs) (Figure 5.3.4) around 290.5 eV. This resonance has already been observed for phenylalanine and tyrosine and is attributed to $C 1s \rightarrow \sigma^*(\text{CC})$ and $C 1s \rightarrow \sigma^*(\text{CN})$ [177]. All transitions in the TIY at the C K-edge can be attributed to transitions already observed in gas-phase neutral tyrosine or protonated peptides. This could be expected as phosphorylation of the side chain of tyrosine does not change the direct neighboring atom of the different carbon atoms.

Comparing the partial ion yield spectra of the different fragments in Figure 5.3.4, it appeared that the water molecule has little influence at the carbon K-edge. Even the PIY of the hydrated fragment $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ that undoubtedly originates from $[\text{pTyr}+\text{H}+\text{H}_2\text{O}]^+$, does not show any clear difference (new peaks or shifts) that could identify an effect of the solvation.

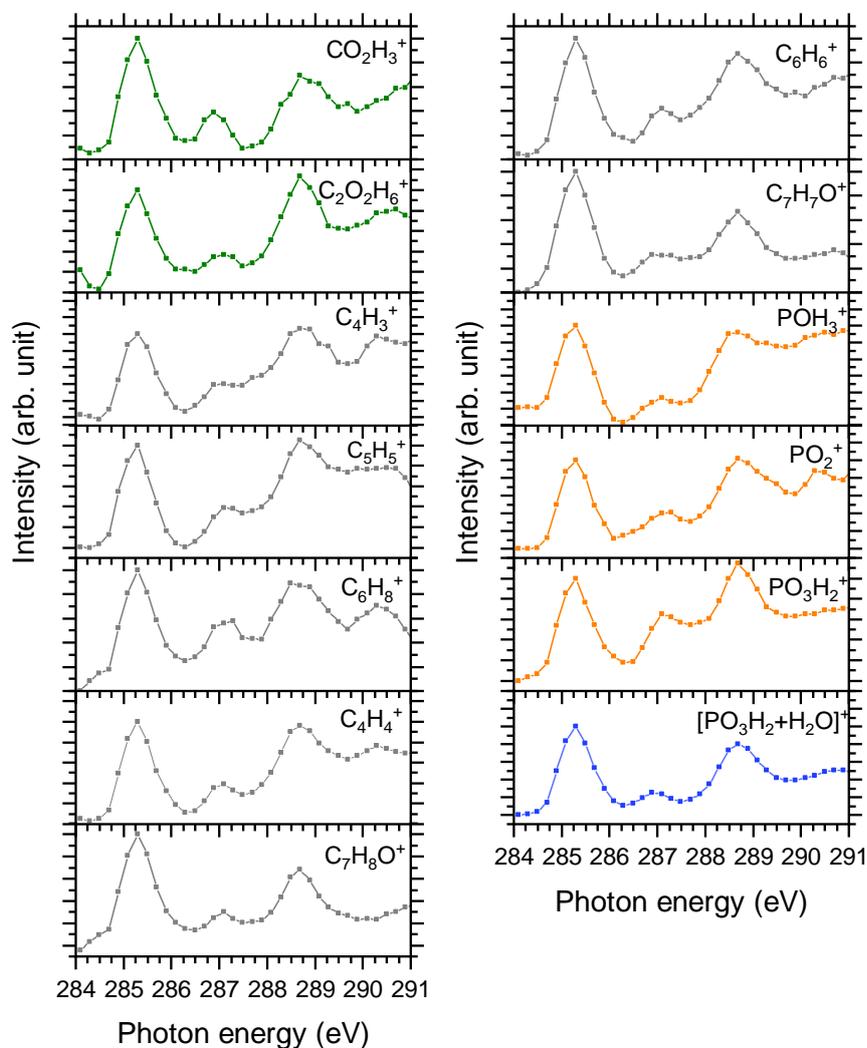


Figure 5.3.4: PIYs of the different fragments originating from the mixture of $[\text{pTyr}+\text{H}+\text{H}_2\text{O}]^+$ and $[\text{pTyr}+\text{H}]^+$ at the C K-edge. All PIYs have been normalized on the resonances at 285.2 eV. The PIYs of the carboxyl-containing fragment are shown in green, and the tyrosine side-chain fragments are shown in grey. The PIYs of the phosphorus-containing fragments are shown in orange, and the PIY of the hydrated fragment is shown in blue.

In contrast to the carbon K-edge, the phosphorylation and the water molecule are anticipated to have significant signatures at the oxygen K-edge, given their elemental composition. The oxygen K-edge offers direct probing of the oxygen atoms within the amino acid, i.e., the phosphate group and carboxyl group, as well as the water molecule. In this energy range, the main resonances of the amino acid are expected at 532.1 eV, 535.2 eV, and 541 eV according to the work of Zhang et al. on neutral tyrosine [177], whereas for the water molecule, the two main resonances are expected to stem at 534.2 eV and 536.2 eV according to Wilson et al. [182]. From the study of Wang et al., we could expect a weak resonance around 534.5 eV due to transitions to the phosphate group [74]. However, the study of Ward et al. on phosphate-containing molecules suggested that the fingerprint of PO_4 appears as a broad and high-intense resonance at 536.7 eV [183].

Therefore, the oxygen K-edge allows us to distinguish between the resonances of the different oxygen sites of hydrated pTyr, as we do not expect any overlap between them.

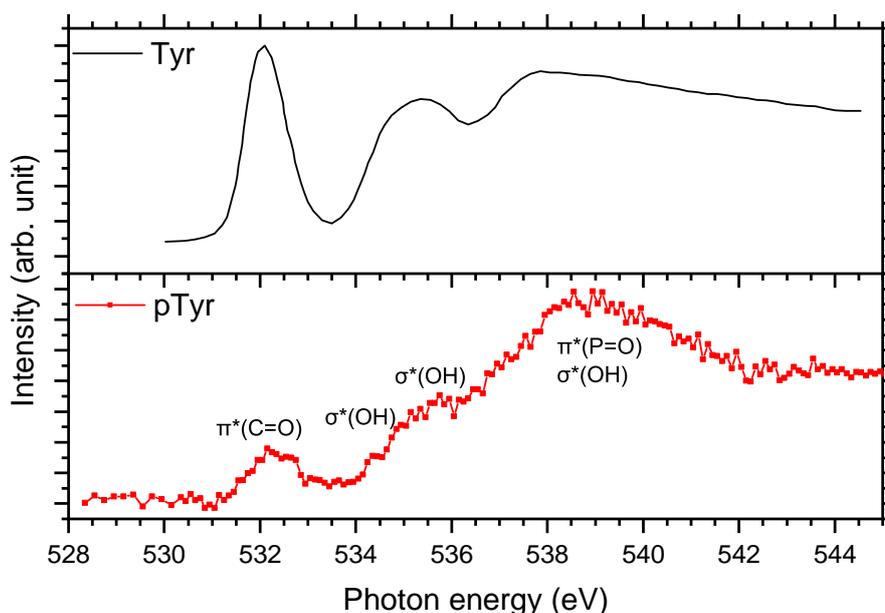


Figure 5.3.5: Top: experimental XAS spectrum for gas-phase neutral tyrosine, adapted from [177]. Bottom: experimental absorption spectrum of a mixture of $[\text{pTyr}+\text{H}+\text{H}_2\text{O}]^+$ and $[\text{pTyr}+\text{H}]^+$ at the O K-edge.

In Figure 5.3.5, the NEXAMS spectrum of the mixture of $[\text{pTyr}+\text{H}+\text{H}_2\text{O}]^+$ and $[\text{pTyr}+\text{H}]^+$ is shown. We observed one main resonance at 532.2 eV, which was also

observed in the X-ray absorption spectrum of neutral tyrosine [177]. The authors attributed this resonance to an $O\ 1s \rightarrow \pi^*(C=O)$ transition from the carboxyl group. Our calculations at the O K-edge of isolated pTyr and hydrated pTyr confirmed that this attribution is also valid for pTyr (see next section). The weak resonance at 534.5 eV, appearing as a shoulder on the broad feature, can be attributed to a $\sigma^*(OH)$ transition from the water molecule, as discussed in the following. A resonance near 536 eV stems from transitions from $O\ 1s \rightarrow \sigma^*(OH)$ mostly from the carboxyl group of the molecule and with a small contribution from the water molecule when it forms a bridge between the phosphate group and the carboxyl group (see the OP_bridge structure in the following). The last broad resonance at 539 eV is due to many electronic transitions, mainly $\pi^*(P=O)$ and $\sigma^*(OH)$ transitions from the phosphate group. (see the next section for the discussion of the different calculated conformers and specific attributions of the resonances).

As the NEXAMS spectrum shown in Figure 5.3.5 is the result of the irradiation of both pTyr and hydrated pTyr, a more precise conclusion cannot be made based only on the TIY spectrum. However, additional information can be extracted from the PIYs at the oxygen K-edge, especially from the PIY of the hydrated fragment $[PO_3H_2+H_2O]^+$.

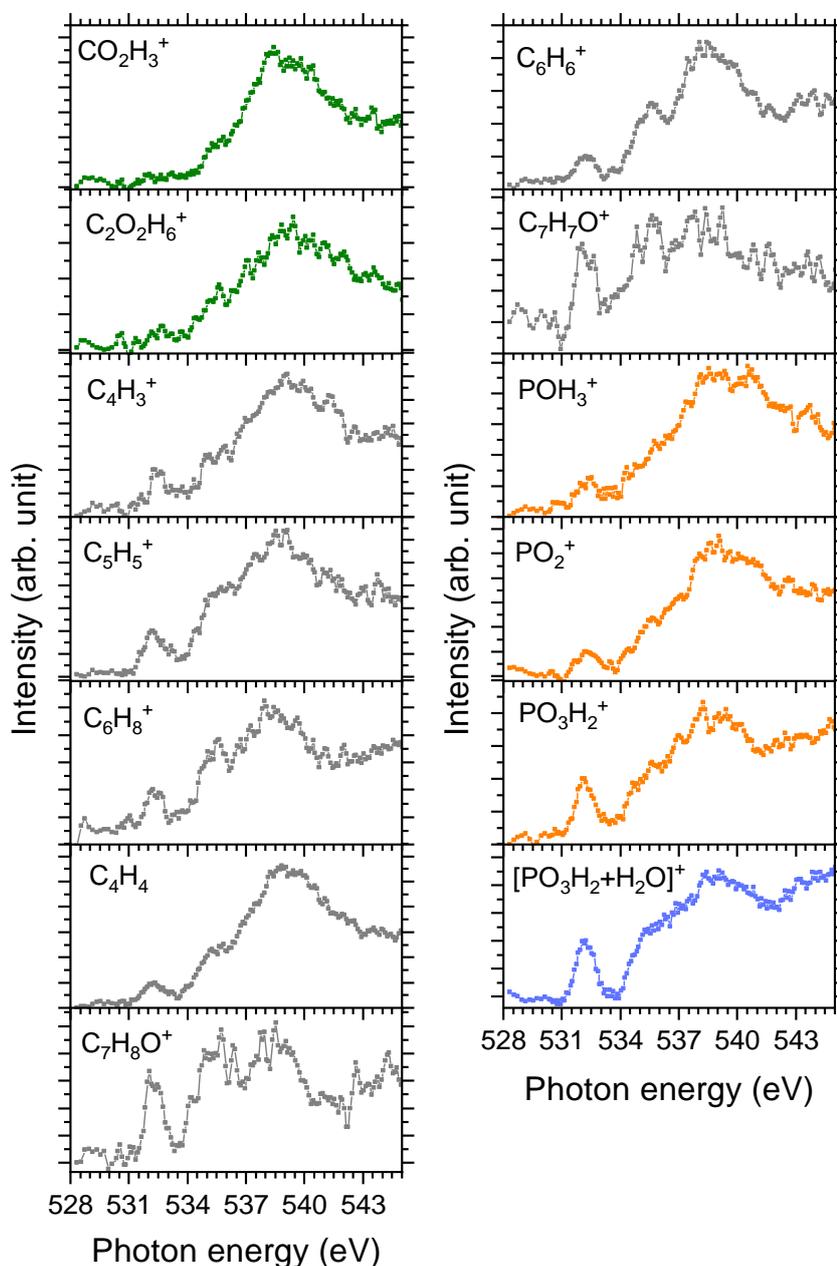


Figure 5.3.6: PIYs of the fragments originating from a mixture of $[\text{pTyr}+\text{H}+\text{H}_2\text{O}]^+$ and $[\text{pTyr}+\text{H}]^+$ at the oxygen K-edge. All PIYs have been normalized on the resonances at 532.15 eV. The PIYs of the carboxyl-containing fragment are shown in green, and the PIYs of tyrosine side-chain fragments are shown in grey. The PIYs of phosphorus-containing fragments are shown in orange, and the PIY of the hydrated fragment is shown in blue.

In Figure 5.3.6, the PIYs of every fragment at the oxygen K-edge are shown. Some differences in the PIYs of the fragments can be observed. Interestingly, the two backbone fragments CO_2H_3^+ and $\text{C}_2\text{O}_2\text{H}_6^+$ do not exhibit the first $\text{O } 1s \rightarrow \pi^*(\text{C}=\text{O})$ resonance at 532.15 eV. These two fragments contain an intact carboxyl group. The absence of the resonance could then be understood as the fact that exciting the $\pi^*(\text{C}=\text{O})$ orbitals will first and foremost induce cleavage of $\text{C}=\text{O}$ and/or $\text{C}-\text{OH}$ bonds in the carboxyl group. Thus, fragments containing an intact COOH group cannot be produced through this resonant excitation.

We observed that for all tyrosine side-chain fragments, plotted in grey in Figure 5.3.6, as well as for phosphate-containing fragments PH_3O^+ and PO_2^+ , the relative intensity between the resonance at 532.15 eV and the resonances stemming at higher photon energy is similar. The fragment $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ and its bare form PO_3H_2^+ present a different ratio between the resonances at 532.15 eV and at 535 eV and might therefore originate from different processes as the tyrosine side-chain fragments. This will be discussed further at the end of this section. Note that for $\text{C}_7\text{H}_7\text{O}^+$ and $\text{C}_7\text{H}_8\text{O}^+$ the relative intensity of each resonance is also not the same as the other tyrosine side-chain fragments. However, no conclusion could be made, as the corresponding spectra are of very low intensity and are affected by statistical noise.

We observed differences between a tyrosine side-chain fragment (C_6H_6^+) coming from the isolated pTyr and the hydrated fragment $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$. Specifically looking at the differences in intensities of the first resonance, we observed for C_6H_6^+ a dip at 536.5 eV whereas for $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ the resonances at 535.5 eV and the one centered around 539 eV are not well resolved (see Figure 5.3.7). This indicates that for $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$ an additional resonance appears at around 536.5 eV. We explain this new resonance by the presence of transitions from the water molecule around 536 eV, as shown in Figure 5.3.7. Wilson et al. measured the total ion-yield spectrum of gas-phase water at the oxygen K-edge; they observed three main peaks in the spectrum, at 534.2, 536.2, and 537.5 eV [182] assigned to transitions from the ground state to a mixture of empty molecular orbitals and Rydberg states [184].

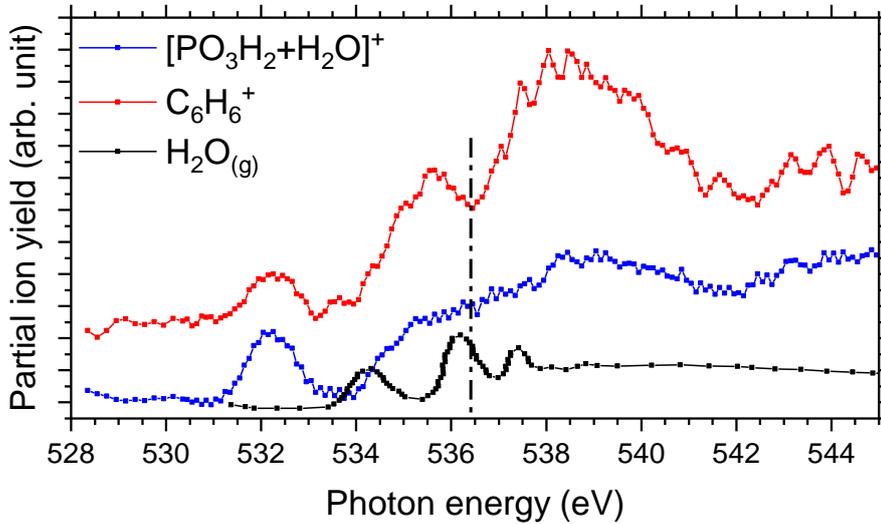


Figure 5.3.7: Comparison of the ion yield spectra of $C_6H_6^+$ and $[PO_3H_2+H_2O]^+$ with the experimental XAS of gaseous water, adapted from [182]. A vertical dashed line has been added at 536.4 eV to guide the eyes.

As mentioned in subsection 5.3.1, it was not possible to conclude that $PO_3H_2^+$ is produced by the evaporation of the water molecule from $[PO_3H_2+H_2O]^+$ on the basis of only the photodissociation mass spectrum. However, it is now possible to draw some conclusions by comparing the PIYs of both fragments at the oxygen K-edge. As illustrated in Figure 5.3.7 highlighting the differences between the $[PO_3H_2+H_2O]^+$ and $C_6H_6^+$ fragments, the PIY spectra of isolated pTyr and hydrated pTyr fragments exhibit category-dependent variations. Notably, $PO_3H_2^+$ has a different PIY spectrum than PO_2^+ , as can be observed in Figure 5.3.6. The resonance at 532.15 eV is more intense in the case of $PO_3H_2^+$ and the feature between 535 and 538 eV is lower in intensity for PO_2^+ compared to $PO_3H_2^+$. Therefore, we can suggest that both fragments have their origin in different photodissociation processes. If $PO_3H_2^+$ comes from the evaporation of the water molecule from $[PO_3H_2+H_2O]^+$ after irradiation with photons, the hydrated fragment $[PO_3H_2+H_2O]^+$ and its bare form $PO_3H_2^+$ should have a similar PIY spectra. Figure 5.3.8 shows that their PIY spectra are identical once normalized on the first resonance at 532.15 eV, and this indicates that they should come from the same process. For instance, after being produced, the hydrated fragment loses its water molecule due to further fragmentation in the trap.

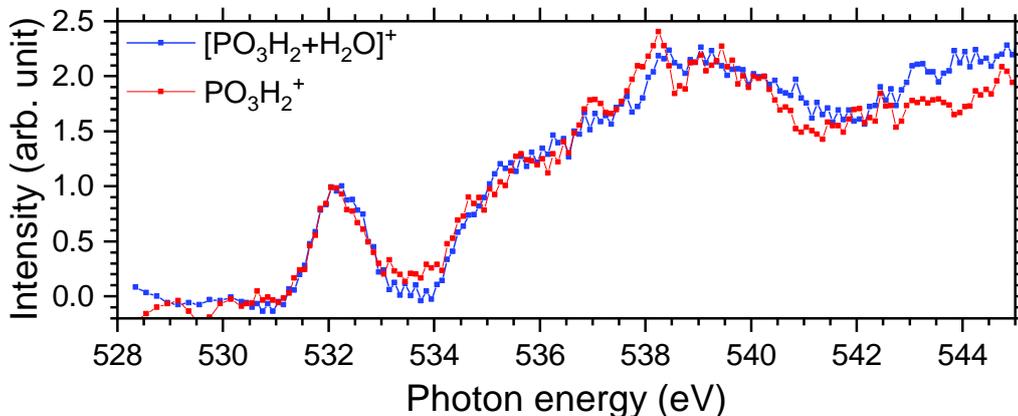


Figure 5.3.8: Ion yield spectra of hydrated and isolated PO_3H_2 fragment. Both spectra have been baseline-corrected at 529 eV and normalized on the $\pi^*(\text{C}=\text{O})$ resonance at 532.15 eV.

To summarize, the NEXAMS spectra gave us more insight into the electronic structure of pTyr and hydrated pTyr. At the carbon K-edge, the total ion-yield spectrum is similar to the one of gas-phase tyrosine measured by Zhang et al. [177]. No clear differences have been observed in the PIY spectra of the hydrated fragment compared to the isolated fragments. However, the water molecule is directly probed at the oxygen K-edge since the NEXAMS technique is element-specific. Although the NEXAMS spectrum at the oxygen K-edge resembles gas-phase tyrosine [177], the relative intensity of the different resonances varies. We showed that the PIYs of a fragment coming from the isolated pTyr have an electronic signature different from the hydrated fragment $[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$, which shows an additional transition at 536.4 eV from the water molecule. We also demonstrated that PO_3H_2^+ originated from the water molecule loss from the hydrated phosphorus-containing fragment.

To obtain more insight into the different electronic transitions and the possible structure of hydrated pTyr, DFT calculations have been performed and compared with the X-ray absorption spectra at the C and O K-edges. This will be discussed in the next section.

5.3.3 Discussion of the calculated conformers

5.3.3.1 Singly hydrated conformers of protonated phosphotyrosine

The analysis of the photodissociation mass spectra drove all calculations for the hydrated conformers.

The spectra revealed that the main hydrated fragment, at m/z 99 ($[\text{PO}_3\text{H}_2+\text{H}_2\text{O}]^+$), contains the phosphate group, thereby directing the structural analysis towards conformations where the water molecule binds to the phosphate group. Moreover, the loss of HPO_3 from hydrated pTyr was observed without the loss of H_2O , indicating that the water molecule probably binds not only to the phosphate group.

To reduce the number of possibilities for the water molecule positions from the full solvation shell of the five possible structures (shown in Figure 5.2.2), we selected singly hydrated structures where the water molecule is either hydrogen-bonded to the phosphate group or bound to the NH_3^+ , starting from the 6 Å shell around the phosphate group, or close to the benzene ring (on top of it). This gave four to six possible water molecule positions for each selected conformer of pTyr. These possible structures were further optimized at the B3LYP/TZVP level of theory. The following categories were identified within the 24 hydrated conformers: 7 conformers have the water molecule making a bridge between the NH_3^+ group and the phosphate group (called NP_bridge in the following), and 15 of them have the water molecule making hydrogen bonds (HB) with the phosphate group only, called POH later in the text. The three lowest energy conformers of both categories were chosen as representatives, and their X-ray absorption spectra were calculated. Moreover, two hydrated conformers showed different structures than the 21 others. One had the water molecule making a single hydrogen bond with the NH_3^+ of the phosphotyrosine (Nterm in the following), and the second one displayed a water bridge between the phosphate group and the carboxyl group of the phosphotyrosine, which is thus named the OP_bridge. The structures of the eight hydrated conformers with their relative electronic energies that were considered can be seen in Figure 5.3.9, and their X-ray absorption spectra were calculated. The efficiency of the B3LYP functional for computing geometries and *ab initio* X-ray absorption spectra of phosphotyrosine were tested. Tests using gradient-corrected (GGA), meta-GGA, and double-hybrid functionals were performed. The results of the tests revealed that the structures computed with all different functionals lead to very similar results with very close RMSD values for all conformers (≤ 0.02 , which means that their structures are roughly the same) and single point energy ranging in less than an eV range which evidences that B3LYP is robust for performing calculations of phosphorylated molecules. This has already been shown for some phosphorus com-

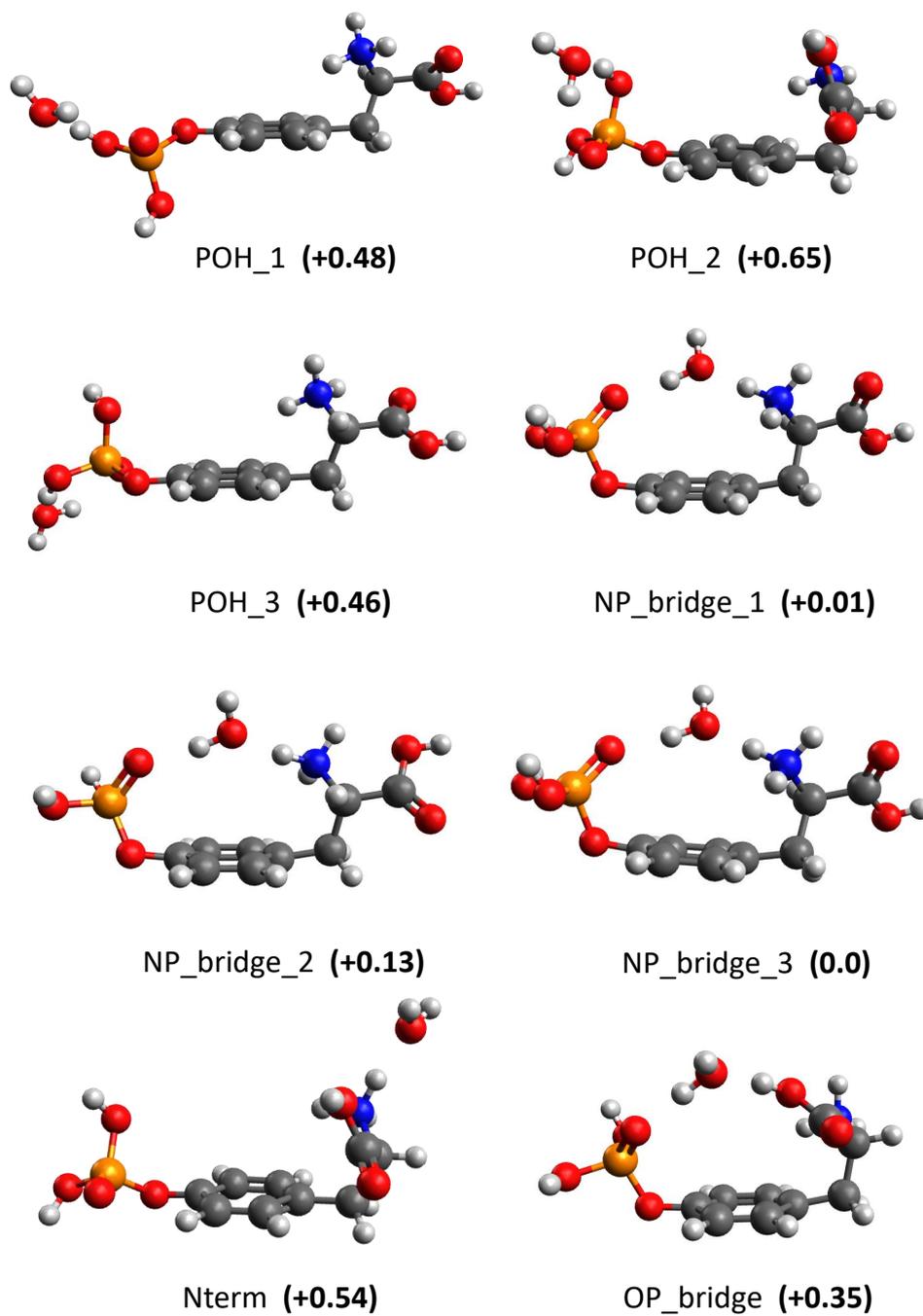


Figure 5.3.9: Structures of the eight singly hydrated conformers of pTyr with their relative electronic energies in eV.

pounds by Nguyen et al. [185] and for phosphorylated proteins by Rudbeck et al. [98]. B3LYP has already been employed to simulate phosphorylated amino acid structures and IRMPD spectra [51], [107].

Finally, to reduce the number of hydrated conformers, a first choice of the isolated conformers and hydrated conformers was made by comparing the calculated XAS of the five isolated possible conformers and the nine hydrated conformers with the experimental carbon K-edge total ion yield spectrum, as it should not be influenced by the water molecule (see subsection 5.3.2).

The relative intensity of each resonance and the transition positions were considered to compare the calculated spectra with the experimental spectrum at the C K-edge. We employed the Pendry reliability factor R_p [186] for an unbiased criterion for this comparison. The equation used to calculate the R_p factor can be written as follows:

$$R_p = \frac{\sum_{i=0}^n (Y_1 - Y_2)^2}{\sum_{i=0}^n (Y_1^2 + Y_2^2)}, \quad (5.3.1)$$

with Y_i , $i=1,2$ being the renormalized logarithmic derivatives defined as

$$Y_i = \frac{L_i^{-1}}{(L_i^{-2} + V_0^2)},$$

with

$$L_i = \frac{\partial I_i}{\partial x}$$

being the derivative of the intensity as a function of the photon energy over the intensity, and V_0 is the approximate half width of the peaks (here taken to be 0.6 eV). A perfect agreement between the theoretical spectrum and the experimental ones is represented by $R_p = 0$, while $R_p = 1$ means no agreement.

Figure 5.3.10 shows the calculated spectra at the carbon K-edge for the eight chosen singly hydrated conformers and their calculated R_p factor. Structures from which their calculated spectra at the carbon K-edge were not in good agreement with the experimental NEXAMS were discarded, i.e. POH_1, POH_3, NP_bridge_1 and NP_bridge_3. NP_bridge_2 has an R_p factor a bit higher than the other conformer. However, this structure was kept, being the lowest energy conformer of the three NP_bridge structures, as we could not rule out the possibility that the water molecule forms a bridge between the N-terminus and the phosphate group, as already shown by Scuderi et al. [107]. Moreover, as discussed previously, we established that the water molecule binds to the phosphate group. This ruled out the Nterm conformer from being further analyzed. There is also the possibility that the water molecule binds only to the phosphate group. Therefore, POH_2, of which the XAS at the

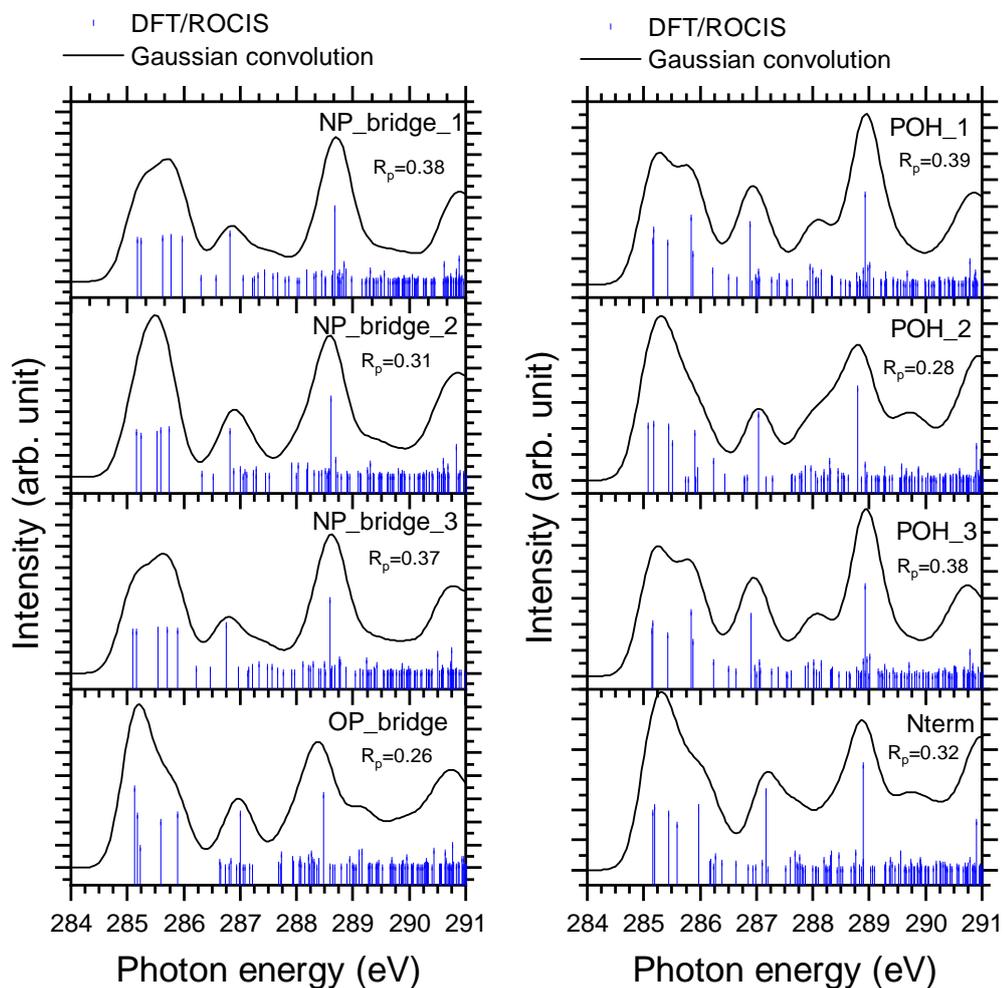


Figure 5.3.10: Simulated spectra at the C K-edge and their Pendry factor for all calculated hydrated conformers of pTyr. The absorption lines were shifted by approximately 11.5 eV.

C K-edge fits the experimental spectrum, was considered as a possible structure of hydrated pTyr. Thus, in the next section, only three (NP_bridge_2, POH_2, and OP_bridge) of the nine hydrated conformers were considered. It is to be noted that OP_bridge and POH_2 structures were derived from the isolated structure pTyr_2, while NP_bridge_2 comes from the hydration of pTyr_1 (see Figure 5.2.1). In the following, the three conformers will be discussed in light of the experimental spectra

at the carbon and oxygen K-edges.

5.3.3.2 Comparison of the calculated absorption spectra with the NEXAMS spectra

In this section, the calculated structures and their agreement with the experimental spectra will be discussed.

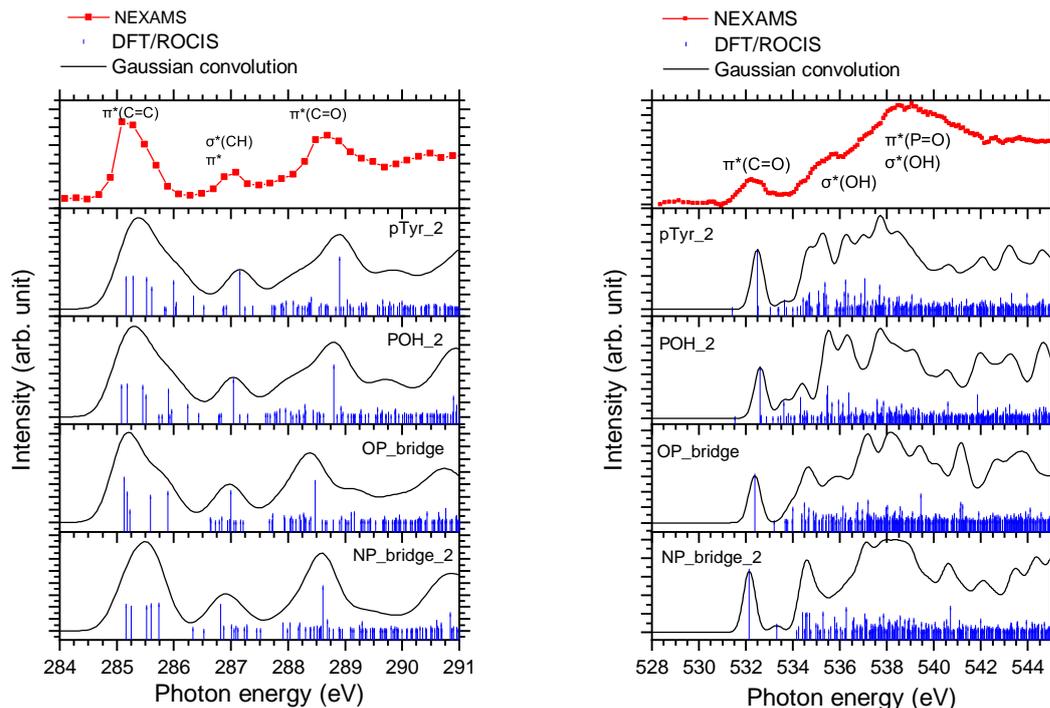


Figure 5.3.11: Comparison between the calculated X-ray absorption spectra of the isolated pTyr (pTyr_2) and the three selected hydrated pTyr conformers (POH_2, OP_bridge and NP_bridge_2) and the experimental total ion yield spectra (red) at the C K-edge (left) and at the O K-edge (right).

Figure 5.3.11 shows the comparison of the experimental total ion yield spectra at the carbon and oxygen K-edges and the four selected conformers. pTyr_2 is the isolated conformer that was chosen to be the final representative for isolated pTyr after a comparison of its calculated spectrum with the NEXAMS at the carbon K-edge. At the carbon K-edge, some differences were found between the different structures. The analysis of the NEXAMS spectrum based on the Gaussian convolution of the calculated transition lines indicates a poorer match for NP_bridge_2 compared to the other

three structures. According to the calculated Pendry factor (see Figure 5.3.10, from the chosen conformers, NP_bridge_2 has the highest R_p factor. For this structure (NP_bridge_2), the relative intensity of the C 1s $\rightarrow\pi^*(C=C)$ resonance at 285 eV compared to C 1s $\rightarrow\pi^*(C=O)$ resonance at 288.7 eV does not reproduce what is observed in the NEXAMS spectrum, which shows the inverse trend. However, all calculated transitions of NP_bridge_2 are in agreement with the attribution of the different resonances in the TIY spectrum of pTyr. Furthermore, pTyr_2 and POH_2 have similar spectra with a resonance at 288 eV appearing as a shoulder on the C 1s $\rightarrow\pi^*(C=O)$ resonance stemming at 288.9 eV (blue-shifted by 0.2 eV compared to the experimental result). This shoulder is not present in the spectrum of OP_bridge. However, the C 1s $\rightarrow\pi^*(C=O)$ resonance is red-shifted by 0.3 eV for OP_bridge compared to the NEXAMS spectrum (see Figure 5.3.11).

To achieve a clearer visualization of the different atomic contributions, we decomposed the calculated spectrum into the contributions from the absorbing atoms in the various groups of the molecule. An example is given in Figure 5.3.12, where the calculated spectrum of pTyr_2 has been divided into two partial contributions: the contribution from the benzene carbon atoms from the tyrosine side-chain and the contribution of the other carbon atoms i.e. α -, β - and carboxylic carbon atoms. This allowed us to conclude that transitions towards the $\pi^*(C=C)$ and $\pi^*(C=O)$ orbitals are only due to excitations of the carbon atoms of the benzene ring, whereas the weak resonance at 288 eV (not resolved in the experimental spectrum) is due to transitions from the α -, β -, and carboxylic carbon atoms and from the benzene group, thus the contribution to this resonance is shared between both groups. Moreover, the last resonance at 288.7 eV (the $\pi^*(C=O)$) has mainly contributions from the α -, β -, and carboxylic carbon atoms. The same is true for POH_2.

From the comparison of the NEXAMS and the calculated spectra at the C K-edge, we infer that although NP_bridge_2 exhibits a convoluted spectrum with differences in shape compared to the NEXAMS, all four conformers are in agreement and can be possible fitting structures for bare and hydrated pTyr in our experimental setup. However, to confirm that our pTyr_2 conformer and the three hydrated conformers can be good structural candidates for pTyr, their calculated spectra should also be in agreement with the experimental spectrum at the oxygen K-edge.

Figure 5.3.13 shows the comparison of the NEXAMS with the decomposed calculated spectra at the oxygen K-edge for the three singly hydrated conformers, POH_2, OP_bridge, and NP_bridge_2. At first glance, we observed that the XAS of OP_bridge

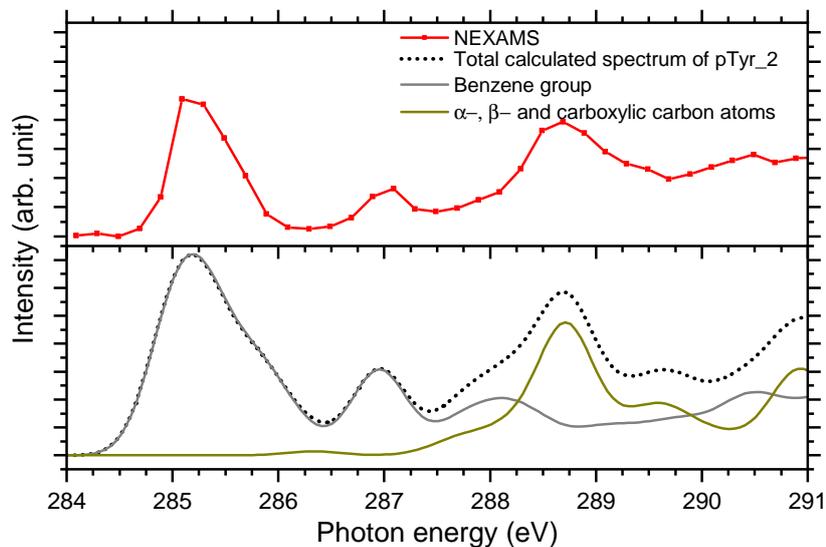


Figure 5.3.12: Comparison of the experimental spectrum at the C K-edge (top) with the theoretical spectrum of pTyr_2 (bottom). The contribution of the carbon atoms of the benzene group is represented in the grey curve, and the contribution of the α -, β -, and carboxylic carbon atoms are represented in dark yellow. Here, the convolution of the theoretical transitions has been done with a sum of Gaussian functions of FWHM=0.6 eV.

and NP_bridge_2 are similar in shape. The first resonance at 532.15 eV is entirely due to $O1s \rightarrow \pi^*(C=O)$ from the carboxyl group, with no transition from other oxygen atoms in the molecules, and the broad feature between 535 and 540 eV comes mostly from transitions from an oxygen atom of the phosphate group exciting σ^* antibonding orbitals. For both conformers bridging the phosphate group and either the N-terminus or the carboxyl group, the water molecule does not exhibit strong transitions and seems to act more like a spectator in the calculation, except for a weak resonance making a shoulder at 534 eV on the resonance at 534.7 eV. However, this shoulder is more pronounced for the OP_bridge conformer compared to NP_bridge_2. This weak resonance, at 534 eV, is due to transitions from the oxygen atom of the water molecule into a $\sigma^*(OH)$ orbital (LUMO+7, Figure 5.3.14) where most of the density is located on the water molecule (see Figure 5.3.14). This could explain the low-intensity resonance at 534.2 eV in our experimental spectrum, which would also be consistent with the TTY spectrum of gas-phase water of Wilson et al. [182]. This transition is also

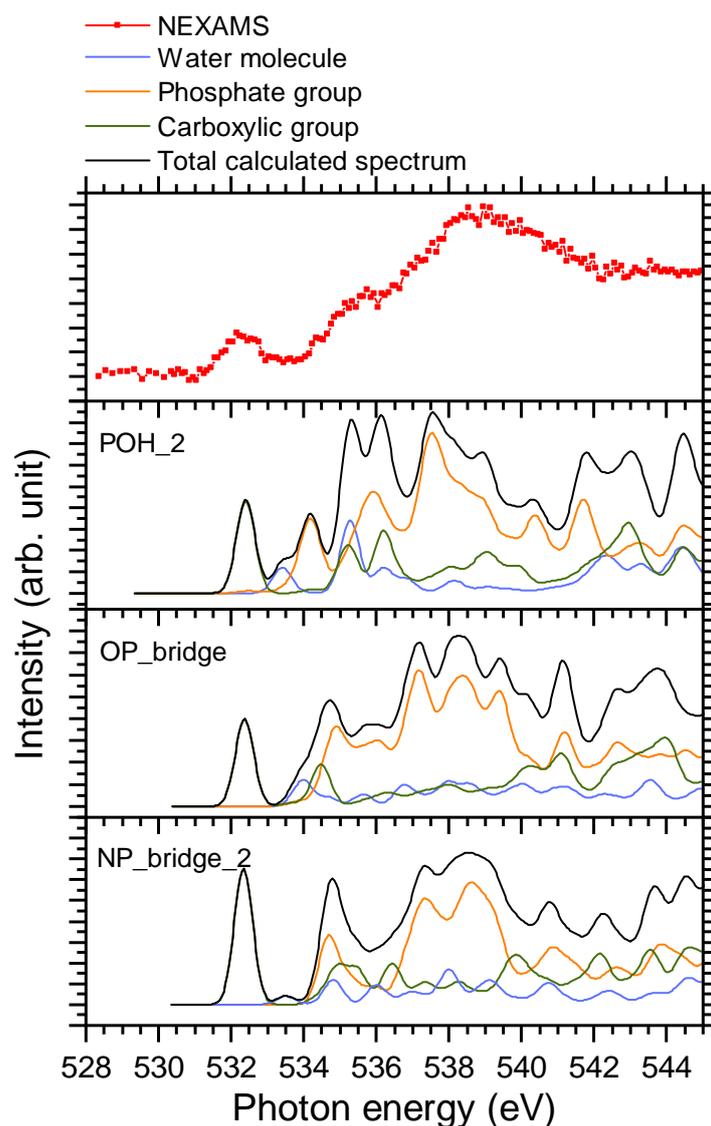


Figure 5.3.13: Comparison of the experimental spectrum at the O K-edge with the decomposition of the calculated XAS for the three singly hydrated conformers. The orange curve represents the contribution of the oxygen atoms of the phosphate group to the calculated XAS. The contribution of the oxygen atoms of the carboxyl group is represented in brown, and the oxygen atom of the water molecule is represented in blue. The convolution of the theoretical transitions has been done with a sum of Gaussian functions of FWHM=0.6 eV.

observed in the XAS of NP_bridge_2 at 533.5 eV (LUMO+3 Figure 5.3.14), but the intensity of this transition is even lower. Thus, this transition does not contribute to the total calculated spectrum. Moreover, for NP_bridge_2, the resonance at 534.7 eV is shared equally by contributions of the carboxyl group and the phosphate group through $O\ 1s \rightarrow \pi^*(C=O)$ and $O\ 1s \rightarrow \sigma^*(OH)$ transitions (LUMO and LUMO+6 Figure 5.3.14). For the OP_bridge conformer, starting from 535 eV, the intense transitions are all due to the phosphate group (LUMO+14 Figure 5.3.14).

Furthermore, the POH_2 conformer exhibits an XAS spectrum rather different from the bridge-conformers, where the water molecule is responsible for a strong transition at 533.6 eV (LUMO+6 Figure 5.3.14) and partly contributing to the resonance at 535.5 eV which is shared with transitions from oxygen atoms of the carboxyl group to the $\sigma^*(OH)$ orbital. From 536 eV to 540 eV, the oxygen atoms of the phosphate group are excited to $\pi^*(C=O)$ and $\sigma^*(OH)$ orbitals (LUMO+1 and LUMO+4 Figure 5.3.14).

The differences in the calculated spectra's lower energy range can be explained by the differences in structure between POH_2 and the two conformers forming a bridge. The resonance around 534 eV (at 533.5 eV for POH_2) is due to transitions from the water molecule into the $\sigma^*(OH)$ orbital, mostly where most of the electronic density is located around the water molecule (LUMO+6, Figure 5.3.14). These transitions show that the water molecule interacts with itself and partially with the -OH group of the phosphate group. As POH_2 makes only one hydrogen bond with the water molecule, the water molecule interacts less with pTyr compared to the bridge structures where the water molecule is more strongly bound to the biomolecule, and the electronic density is shared between the water molecule and pTyr through the phosphate group and either the N-terminus or the carboxyl group. That is why for OP_bridge and NP_bridge_2, the intensity of these water-related transitions is very low, and most of the electronic transitions in the calculated spectra of both conformers come from electronic excitation from the carboxylic and phosphate groups. Therefore, the calculated oxygen K-edge spectra for both bridge conformers resemble more the calculated spectrum of pTyr_2 (see Figure 5.3.11).

Moreover, from the NEXAMS spectra, it is clear that no intense transitions are present between the $O\ 1s \rightarrow \pi^*(C=O)$ transitions at 532.15 eV and the $O\ 1s \rightarrow \sigma^*(OH)$ transitions at 535.5 eV, even though the possibility that some transitions are present cannot be ruled out (the TIY intensity being non-zero between the two resonances). Thus, we could conclude that POH_2 can be present in our experimental setup but is not the major conformer of the target ion cloud. However, from the ESI-only and the photodissociation mass spectra, we concluded that the water molecule not only binds

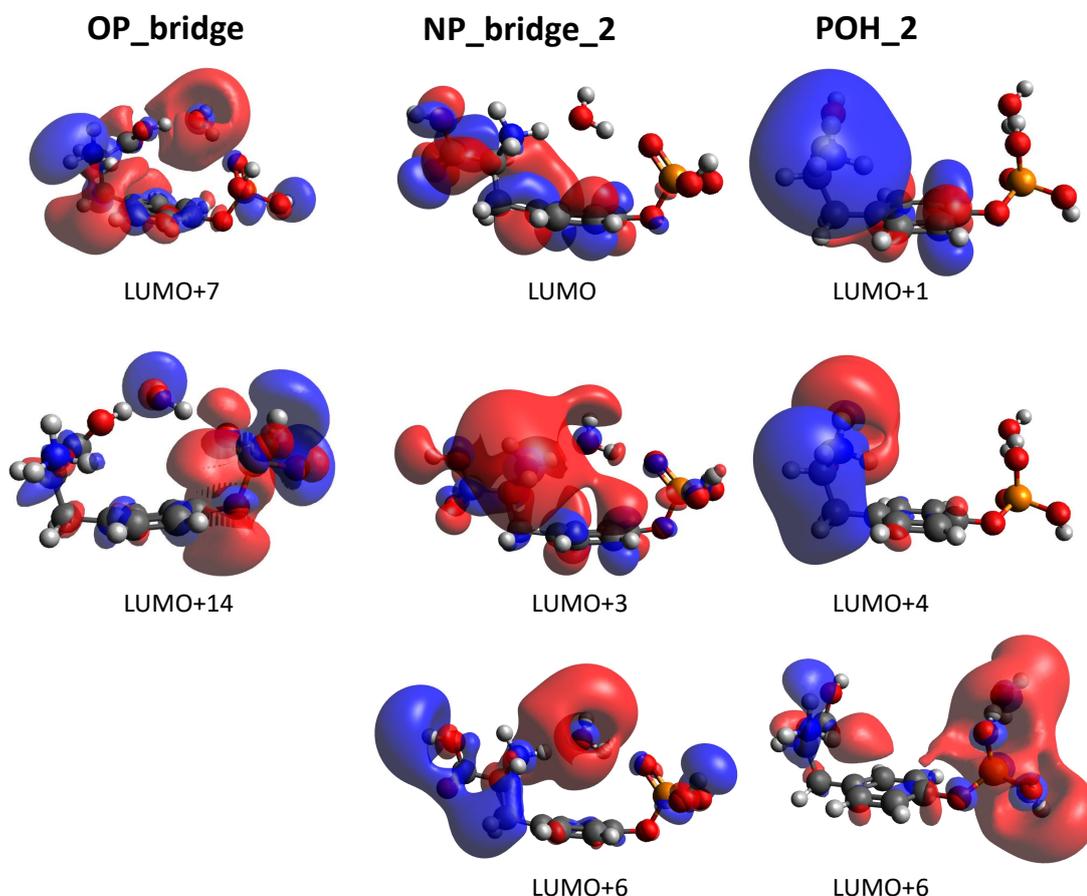


Figure 5.3.14: Molecular orbitals responsible for the main transitions at the oxygen K-edge for the OP_bridge, NP_bridge_2, and POH_2 conformers.

to the phosphate group but should bind to another group to allow the loss of HPO_3 without the loss of the water molecule. Therefore, the two bridge conformers are good structural candidates. Of both conformers, OP_bridge shows a stronger water transition at 534 eV, which might be the explanation for our experimental resonance at 534.3 eV.

For singly-hydrated protonated phosphotyrosine, Scuderi et al. showed that under their experimental conditions, the most stable conformers have the water molecule bridging the phosphate and the ammonium, where the water molecule is an acceptor from the N-H and a donor to O=P [107]. This is similar to what we obtained for the NP_bridge_2 and OP_bridge structures. For the OP_bridge structure, H_2O is an acceptor from the -OH of the carboxyl group and a donor to O=P. Scuderi et al. also found a similar structure to OP_bridge during their structural search, but they did

not consider it further as it was a higher energy conformer [107]. We concluded that, even if we cannot rule out the possibility of POH_2 or NP_bridge_2 and similar structures, OP_bridge is in best agreement with the NEXAMS at the carbon and oxygen K-edges and its structure shows a bridge between the carboxyl group and the phosphate group through the water molecule. The OP_bridge structure was therefore chosen as the structural representative of hydrated pTyr in our experimental condition.

5.4 Conclusion

In conclusion, we have measured the experimental X-ray spectra at the carbon and oxygen K-edges of a mixture of isolated and singly hydrated protonated phosphotyrosine to investigate the effect of a single water molecule on the fragmentation and spectroscopic fingerprint of the biomolecule in the soft X-ray range. We have shown that the combination of DFT/ROCIS calculations and soft X-ray spectroscopy allows the determination of the location of the water molecule. After a first structural search, three possible conformers were calculated: one where the water molecule binds to the phosphate group only and two where the water molecule forms a bridge between the phosphate group and either the carboxyl group or the N-terminus. The comparison of the NEXAMS spectra with the calculated spectra of these three conformers at the oxygen K-edge allowed us to rule out the first conformer. In the light of the photodissociation mass spectrum and the comparison of the calculated spectra and NEXAMS, we concluded that the most representative conformer present under our experimental condition was a structure where the water forms a bridge between the phosphate and the carboxyl groups. Finally, our experimental technique needs to be improved to probe the bare and hydrated molecule separately, for example, using resonant extraction in the trap, which would allow selective removal of one of the species from the trap, or a dedicated electrospray source that would enhance the yield of hydrated or desolvated species, such as the one developed during this thesis and described in section 3.4. This work opens the way for further extended studies of the influence of additional water molecules on the structure and photo-fragmentation of protonated phosphotyrosine or biomolecules in general.

6 | Electronic signatures of the protonation sites of model peptides

This chapter has been adapted from our recently published article [187].

6.1 Abstract

Near-edge X-ray absorption mass spectrometry (NEXAMS) around the nitrogen and oxygen K-edges was employed on gas-phase peptides to probe the electronic transitions related to their protonation sites, namely at basic side chains, N-terminal, and amide oxygen. The experimental results are supported by Replica Exchange Molecular Dynamics and density-functional theory, as well as restricted open-shell configuration interaction with singles calculations to attribute the transitions responsible for the experimentally observed resonances. We studied five tailored-made glycine-based pentapeptides, where we identified the signature of the protonation site of N-terminal proline, histidine, lysine, and arginine, at 406 eV, corresponding to $N\ 1s \rightarrow \sigma^*(NH_x^+)$ ($x = 2$ or 3) transitions, depending on the peptides. We compared the spectra of pentaglycine and triglycine to evaluate the sensitivity of NEXAMS to protomers. Separate resonances have been identified to distinguish two protomers in triglycine, the protonation site at the N-terminus at 406 eV and the protonation site at the amide oxygen, characterized by a transition at 403.1 eV.

6.2 Introduction

In the biological medium, the charge states of molecules, such as peptides, proteins, and nucleic acids, play a significant role as they influence conformation and reactivity. Within the physiological pH range, an important mechanism altering the charge state is the protonation and deprotonation of specific chemical groups. Improving our understanding of biological processes and the structure-function relationship requires us

to localize the (de)protonation sites in biomolecules. Experimentally, protonation sites in peptides have been studied by different techniques over the last decades, mainly by gas-phase basicity studies [188]–[191] to evaluate proton affinities of chemical groups. According to the proton affinities of the different sites, protonation in amino acids and small peptides occurs preferably at the nitrogen site of the most basic side chains (arginine, histidine, and lysine with proton affinities of 250.2, 231.8, and 237.1 kcal.mol⁻¹, respectively [192]) and at the N-terminus primary amine -NH₂ [188]. However, in certain cases, for example, protonation can also occur on the peptide backbone when the number of protons exceeds the number of basic residues. Action spectroscopy techniques, and in particular infrared (IR) spectroscopy, are powerful tools for the determination of molecular structures in the gas phase that can be used to assess the protonation site of molecules as big as small peptides. For example, in the study of Grégoire et al., they first investigated the IR signature of the protonation sites of di-alanine peptide combining resonant IRMPD with different theoretical approaches [49]. The authors further used the same experimental and theoretical methods to investigate the protonation site of the small protonated peptides GGG, RGD, and RGDFV [48].

Protonated triglycine has been studied theoretically and experimentally by many groups, mostly to understand its structure and sites of protonation that remain controversial. In her theoretical study on glycine oligomers, Chung-Phillips concluded that the most stable conformer of triglycine is protonated at an amide oxygen in the backbone [193]. Wu and Mc Mahon employed infrared multiphoton dissociation (IRMPD) to determine the structure and protonation site of triglycine. They showed that, under their experimental conditions, two distinct protomers of triglycine with different protonation sites were responsible for the recorded spectrum [194], one protonated at the N-terminus of the tripeptide and the other protonated on the first amine oxygen. The same conclusion was drawn by Grégoire et al. in a study in which they combined resonant IRMPD of protonated triglycine and Car-Parrinello molecular dynamics calculations [48]. However, in a recent study, Li et al. calculated that the protonation site of triglycine is the N-terminal nitrogen rather than the amine oxygen, which they found to be at least 0.14 kcal.mol⁻¹ more stable [195]. The authors also calculated the X-ray absorption spectra at the carbon, nitrogen, and oxygen K-edges of triglycine protomers with different conformations. They observed clear differences between the linear and folded conformers of triglycine, with resonances unique to some conformers and with the strongest effects being observed at the nitrogen and oxygen edges. This showed theoretically that near-edge X-ray absorption fine structure (NEXAFS) spectroscopy can be sensitive to the different protonation sites of a peptide. Furthermore,

this method has already been applied to determine the position of hydrogen atoms and protonation state for biomolecules in liquid phase and crystals. For example, Edwards et al. utilized the combination of NEXAFS and time-dependent DFT (TDDFT) to determine the structure of organic crystals and the positions of hydrogen atoms [196] while Eckert et al. combined NEXAFS in aqueous solution with TDDFT to establish the structure of 2-Mercaptopyridine and its (de)protonation sites [197].

Over the last decades, soft X-ray spectroscopy has become an important technique for studying the electronic structure of biomolecular systems, both in the condensed [80], [165] and gas phases [85], [86], [89]. In fact, soft X-ray excitation is element-specific, so tuning the photon energy to a resonance allows site-selective excitation of a molecule to specific unoccupied molecular orbitals. Very recently, Wang et al. demonstrated, using near-edge X-ray absorption mass spectrometry (NEXAMS), that the electronic structure of gas-phase oligonucleotides is influenced by the protonation site [198]. In a previous study, Otero and Urquhart observed that peptides in the solid phase show differences in their N 1s NEXAFS spectra depending on the protonation state of the amine group [166]. For amine-protonated species, they observed a feature at high energy around 406 eV, well above the resonance around 402 eV, characterizing an unprotonated amine group.

To go further and map the different electronic transitions responsible for the protonation site in peptides, we studied tailor-made glycine-based pentapeptides. Glycine is the simplest amino acid with a single hydrogen as its side chain and has been studied extensively [79], [81], [82], [189], [199]–[202]. Therefore, studying peptides made from glycines and one basic amino acid at the C-terminus or a proline at the N-terminus allowed us to avoid probing effects from side chains' atoms that are not at the actual protonation site while keeping the system simple for comparison with calculations. To probe the electronic structures of the different protonation sites in peptides, we studied five different glycine-based protonated pentapeptides by soft X-ray action spectroscopy. As protonation occurs preferentially at nitrogen sites, we present here a study at the nitrogen K-edge of model peptides with known protonation site: $[\text{GGGGG}+\text{H}]^+$, $[\text{PGGGG}+\text{H}]^+$, $[\text{GGGGH}+\text{H}]^+$, $[\text{GGGGR}+\text{H}]^+$, and $[\text{GGGGK}+\text{H}]^+$ (noted G_5 , PG_4 , G_4H , G_4R , and G_4K , respectively, in the following), where G stands for glycine, P for proline, H for histidine, R for arginine, and K for lysine, as well as protonated triglycine $[\text{GGG}+\text{H}]^+$, noted G_3 in the following. The chemical structures of G_5 and of the amino acids that make up the other pentapeptide sequences are shown in Figure 6.2.1. G_5 and PG_4 are both N-terminal-protonated but with primary and secondary amine, respectively. Histidine, arginine, and lysine are basic residues, and the protonation occurs on their side chains. The study of triglycine allows us to test

the theory of Li et al. [195].

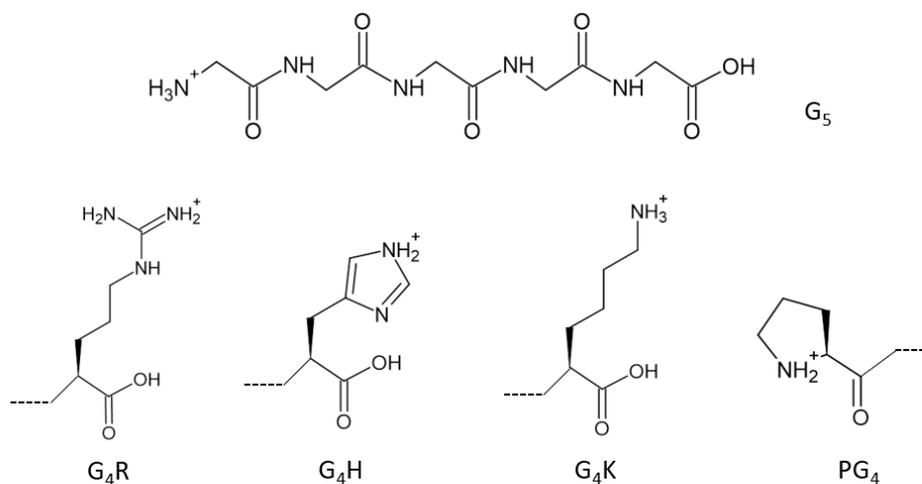


Figure 6.2.1: Chemical structure of protonated pentaglycine (top), the chemical structures of the last amino acid of G_4R , G_4H , G_4K and the first amino acid of the sequence PG_4 , from left to right (bottom).

6.3 Methods

6.3.1 Experimental method

All custom peptides were purchased from ProteoGenix (Schiltigheim, France), at more than 95% purity and used without further purification. The sample solutions were prepared at 30 μ M concentration in 1:1 water/methanol with 1% of formic acid to ensure the protonation of the peptides. The experiments have been performed during two experimental campaigns, one at the BESSY II synchrotron (HZB, Berlin, Germany) and one at the PETRA III synchrotron (Deutsches Elektronen-Synchrotron DESY, Germany) on two different apparatus with a similar principle. The results for protonated G_4H , G_4R , and PG_4 have been obtained by coupling the home-built tandem mass spectrometer Spektrolatius (described in detail in Chapter 3) to the P04 soft X-ray beamline of PETRA III. This setup relies on a high-fluence electrospray ionization (ESI) source to introduce the ions into the gas phase. The electrosprayed ions are transported by a heated capillary into the first pumping stage, which hosts an ion

funnel that collects and focuses the ions. They continue through a radio frequency (RF) octupole ion guide to a quadrupole mass filter where the ions of interest are selected according to their mass-to-charge (m/z) ratio. Then, a second octupole and a set of Einzel lenses guide and focus the ions into an RF 3D Paul ion trap, where they are accumulated prior to irradiation with photons. To allow efficient trapping of the ions, their kinetic energy is cooled down to room temperature by collisions with helium buffer gas. All the cationic products are eventually extracted from the trap and analyzed in a reflectron time-of-flight (Re-TOF) mass spectrometer ($m/\Delta m = 1800$). The duration of the synchrotron light exposure, typically 2 seconds, is controlled with an optomechanical shutter. At the N K-edge, the energy bandwidth was set to 250 meV. The photon energy scans were performed by steps of 200 meV across the N K-edge (397 - 412 eV). A photodiode placed downstream of the ion trap was used to record the photon flux. The NEXAMS spectra obtained during this beamtime were normalized to the photon flux and to the precursor ion intensity to remove the contribution of the ESI source fluctuations. To account for all sources of background ions, the data acquisition was divided into repeated cycles of three mass scans. First, a mass spectrum of the ESI-only was recorded to measure the precursor ion intensity and fragments originating from energetic collisions with the buffer gas. Second, the mass spectrum resulting from the irradiation of the trapped molecules was recorded. For the third spectrum, the ESI is switched off and the mass spectrum resulting from the irradiation of the residual gas was recorded. The mass spectrum resulting from the ESI and the one resulting from the irradiation of the residual gas were subtracted from the mass spectrum resulting from the irradiation of the trapped molecules. To obtain the final mass spectrum, a series of a hundred cycles was accumulated to average over long-term fluctuations of the source. Therefore, in the final subtracted mass spectrum, the precursor ion peak of each molecule appears negative due to this procedure. The results on protonated G_3 , G_5 , and G_4K have been obtained at the UE52_PGM Ion Trap beamline of the BESSY II synchrotron, described elsewhere [28], [90]. The Ion Trap setup from BESSY II has been described in Chapter 3. Briefly, the instrument is equipped, similarly to the setup described above, with an ESI source, a quadrupole mass filter, a cryogenically cooled linear ion trap, and a Re-TOF mass spectrometer. For G_4K , the temperature of the trap was set at 290 K. Contrary to side-chain protonation, backbone protonation is more sensitive to a peptide's conformation. In an attempt to reduce the number of conformers in the trap and limit them to the lowest energy conformers, the temperature of the trap was set to 8 K for G_3 and G_5 . The N K-edge spectra were recorded by steps of 100 meV and 210 meV bandwidth across the range 397 - 410 eV (only to 408 eV for G_4K). Measurements at the O K-edge for

G₅ and G₃ were additionally performed to assess the effect of potential amide oxygen protonation. Here, the photon-energy bandwidth was 200 meV, and the photon energy was ramped up in 100 meV steps between 528 and 545 eV. A photodiode located behind the interaction region was used to record the photon flux for data normalization. Note that the design of the Ion Trap instrument limits the detection of the irradiation products to a small m/z window. The extraction parameters were set to observe fragments over a mass range of about m/z 50-150 for G₄K and over a mass range of m/z 25-70 for the measurements of G₅ and G₃. The NEXAMS spectra, or total ion yield (TIY) spectra, can be assumed to be good approximations of the X-ray absorption cross-section and, thus, comparable to the calculated DFT/ROCIS oscillator strengths. Indeed, the partial ion yields for individual fragments are the results of the convolution of the absorption cross-section of the molecule with the fragments' branching ratio. By summing up the contribution of all fragments into a "total ion yield" spectrum, we approach the result of an absorption spectrum [203]. Nevertheless, disagreement in spectral intensity could arise in the case of detecting multiple ions from the same photoabsorption event or due to discrepancies in ions' overall detection efficiencies. NEXAMS, or TIY NEXAFS, has been extensively used and compared with X-ray absorption calculation [113], [137], [204]–[207].

6.3.2 Theoretical calculations

Structural candidates for G₃ and the different pentapeptides were obtained with different structure search techniques. For G₃, a random structure search was performed, where all dihedral angles of the molecule were allowed to rotate with random angles. The search was conducted using the GenSec package [208], [209]. We sampled 600 structures starting from the three different protonation states labeled GGGH1, GGGH2, and GGGH3 (see Figure 6.4.28). These structures were fully optimized with the PBE0 exchange-correlation functional [210] including the pairwise TS-vdW corrections [211] with the FHI-aims program package [212]. The final optimized structures were ranked by their total energy, and we selected conformers lying on the lowest 1 eV range for each protonation state. In total, we obtained 10 GGGH1, 10 GGGH2, and 5 GGGH3 conformers. This search strategy was targeted at finding low-temperature, enthalpically stabilized conformers. We subsequently calculated in-silico X-ray absorption spectra using the density functional theory (DFT) combined with the restricted-open-shell configuration interaction with singles excitations (ROCIS) for the lowest energy conformer of each protonation state [130]. The calculated X-ray absorption

spectra of GGGH3 do not reproduce the experiment at all, results for this conformer are not further discussed in this paper. Apart from the fact that only a few conformations were found in the *ab initio*-based GenSec structure search performed in this work, there is evidence in the literature that only a few conformers contribute to the spectroscopic signals of small glycine peptides, as shown in the study of Wu and Mc Mahon [194]. In addition, the extensive conformer search performed by Hudgins et al. found essentially only one stable conformation in the case of glycine pentapeptides and smaller glycines by combining ion mobility measurements with molecular dynamics simulations using the CHARMM force field [213]. A different strategy was adopted for the pentapeptides, which was targeted at finding higher temperature stable conformers. We conducted molecular dynamics simulations and DFT/ROCIS calculations for the pentapeptides using the GROMACS program (version 2018.8) and the ORCA electronic structure package [127], respectively. To efficiently sample the potential energy surface (PES), exponentially distributed Replica Exchange Molecular Dynamics (REMD) simulations with different temperatures ranging from 300 K to 648 K were performed based on the Amber ff14SB force field in GROMACS [214]. The conformers generated by the REMD simulation were represented by the values of their dihedral angles for the backbone and side chains. Calculating these descriptors for each snapshot of the REMD simulation resulted in a feature dataset of the REMD trajectory. These features were used to cluster conformers with similar structures using the Probabilistic Analysis of Molecular Motifs (PAMM)[215], [216] such that we could identify recurring molecular conformations of different metastable states sampled during the REMD. PAMM is specifically designed to work in relatively high-dimensional spaces and can partition the underlying probability distribution function into modes corresponding to different conformers. The lowest REMD energy conformation of each identified cluster was chosen as the representative conformer of that cluster and considered for the calculation of the absorption spectra, as discussed by Kotobi et al. [217]. A Boltzmann weighting factor based on the REMD energy differences between the representative conformers and normalized by the cluster population was employed to estimate the statistical relevance of the various metastable states and to obtain weighted average spectra considering a canonical ensemble of the conformers. Conformers from REMD, having a weighting factor lower than 2% of contribution, were assumed to be negligible. All theoretical absorption spectra presented in this paper were obtained using the DFT/ROCIS approach, employing the TZVP Ahlrichs basis set [157] in combination with the B3LYP exchange-correlation functional [123]. The optimized conformers were used to calculate transition energies and the dipole moments for carbon, nitrogen, and oxygen K-edges. We calculated the averaged absorption spectra by weighting the

contribution of different conformers using the previously calculated weighting factors. The final absorption spectra were then broadened, using a pseudo-Voigt profile with a full width at half-maximum (FWHM) of 0.6 eV [91]. The shift in energy of the theoretical spectra is done by matching the first resonance of the theoretical spectra to the corresponding one in the experimental spectra, representing π^*_{CONH} electronic transitions in the peptide backbone, as explained later [172]. This shift is due to the B3LYP functional underestimation of the core excitation energies [159], [173]. Such method has already been successfully used to compare DFT calculations with soft X-ray absorption spectra [81], [85], [86], [218].

6.4 Results and Discussion

6.4.1 Photodissociation mass spectra analysis

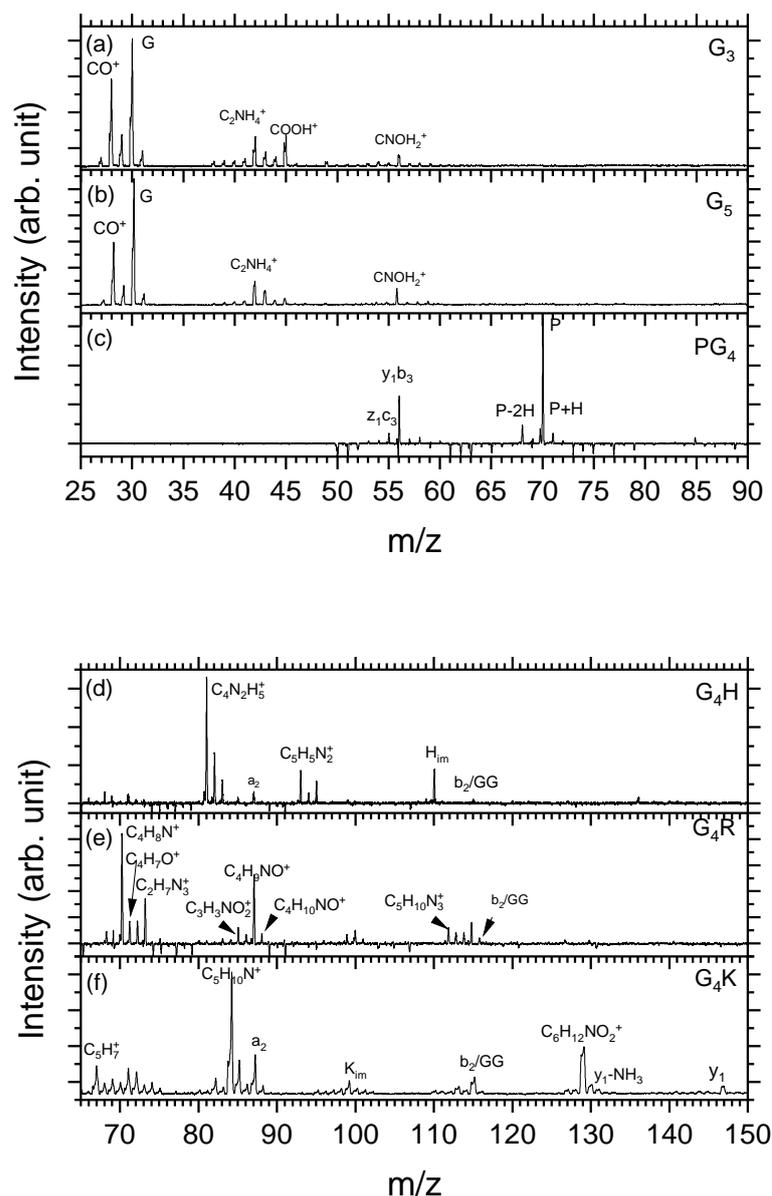


Figure 6.4.1: Low mass region of the photodissociation mass spectra of all investigated protonated peptides measured at 401.6 eV photon energy. The mass range changes between the two sets of mass spectra to show the main fragments produced by photodissociation. Peaks of negative intensity are due to the background subtraction.

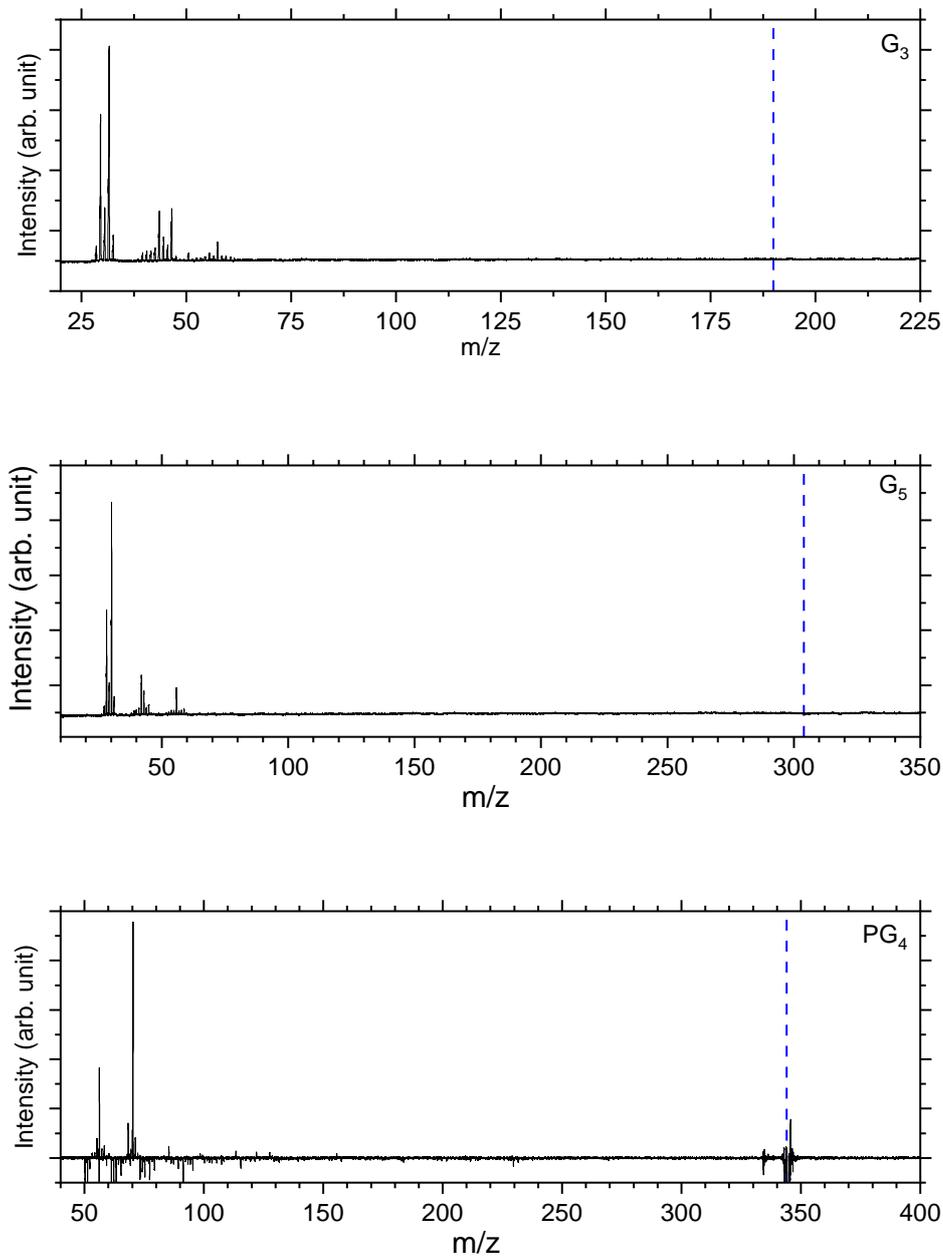


Figure 6.4.2: Photodissociation mass spectrum over the full mass range of G_3 , G_5 measured at 401.6 eV (BESSY II), and PG_4 at 401.6 eV (PETRA III). The mass window was set to see the fragments from m/z 25 to m/z 70 for the measurement of G_3 and G_5 . The negative peaks in the spectrum of PG_4 come from the background subtraction. The dashed blue line represents the position of the precursor ion.

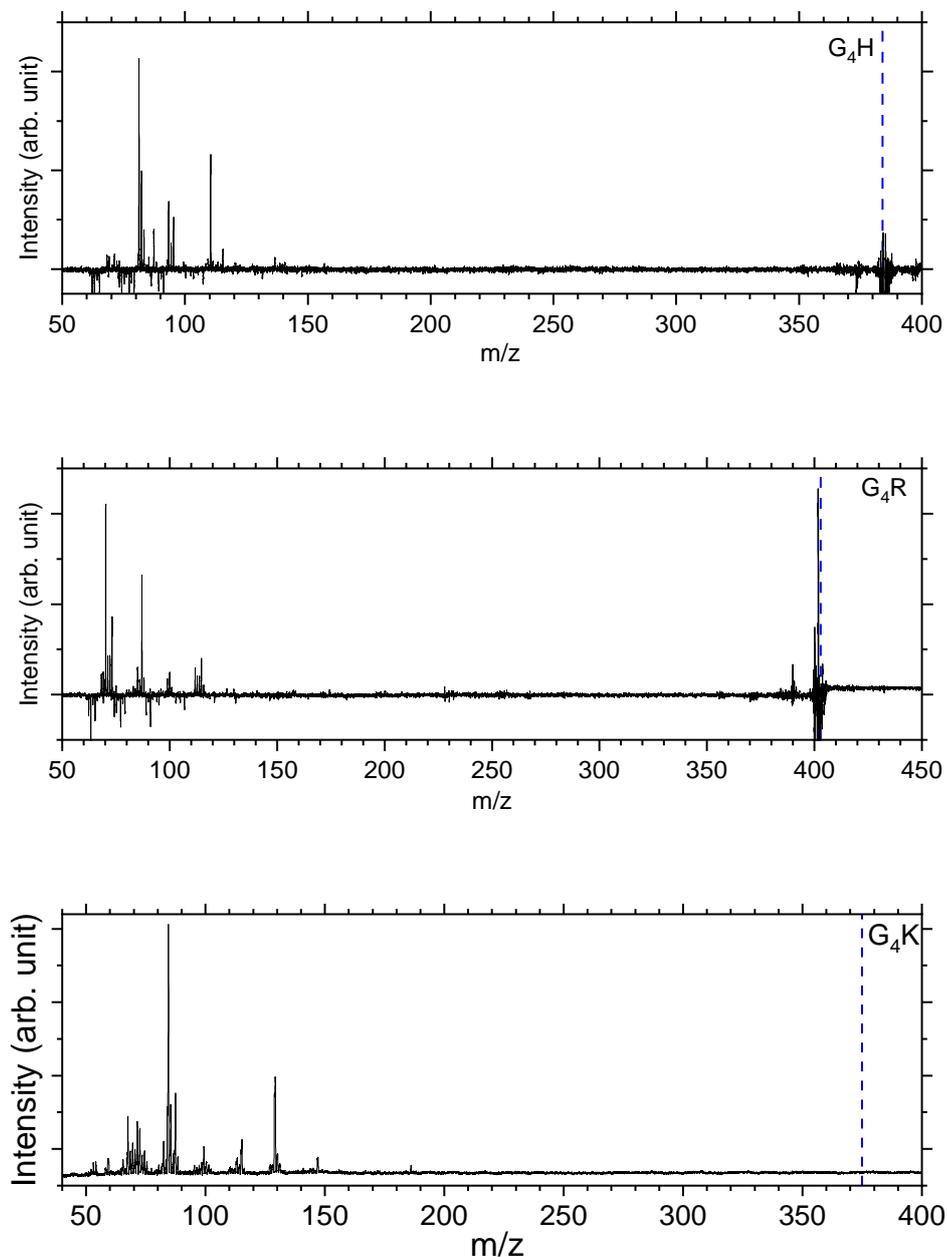


Figure 6.4.3: Subtracted photodissociation mass spectrum over the full mass range of G₄H and G₄R measured at 401.6 eV (PETRA III), and G₄K at 401.6 eV (BESSY II). For the G₄K measurement, the mass window was set to see the fragment from m/z 50 to m/z 160. The negative peaks come from the background subtraction. The dashed blue line represents the position of the precursor ion.

Figure 6.4.1(a) to Figure 6.4.1(f) show the mass spectra measured for the peptides G₃ (m/z 190), G₅ (m/z 304), PG₄ (m/z 344), G₄H (m/z 384), G₄R (m/z 403), and

G₄K (m/z 375) at the photon energy of 401.6 eV ($\pi^*(\text{CONH})$ resonance, as discussed later). The photodissociation mass spectra over the full mass range for the six peptides are shown in Figure 6.4.2 and Figure 6.4.3. As can be seen in Figure 6.4.2 and Figure 6.4.3, no intact photoionized precursor ions have been observed for all peptides. Fragments are labeled according to the peptide fragment nomenclature established by Roepstorff, Fohlman, and Biemann [36], [37]. A table summarizing the different fragments is shown in Table 6.1. Triglycine and pentaglycine are composed of the same residues, which is reflected in their similar mass spectra (Figure 6.4.1(a) and Figure 6.4.1(b)) with fragments stemming from the backbone. m/z 28 corresponds to a fragment that can be attributed to C_2H_4^+ or CH_2N^+ , m/z 29 is attributed to the CHO^+ fragment. The immonium ion of glycine is observed at m/z 30. At m/z 31 is the fragment CH_3O^+ . m/z 42 corresponds to $\text{C}_2\text{H}_4\text{N}^+$, m/z 43 to CHNO^+ , m/z 45 to COOH^+ , and m/z 56 to C_2NOH_2^+ . These fragments have also been reported for neutral glycine under VUV ionization at 21.21 eV by Lago et al. [219] and by Majer et al. using electron-impact ionization [220].

In the photodissociation mass spectrum of PG₄ (Figure 6.4.1(c)), fewer fragments are observed. The most prominent one is the fragment at m/z 70, which is the immonium ion of the proline amino-acid residue, that is also produced by collision-induced dissociation (CID) as observed by Zhang et al. [175]. The two other fragments at m/z 68 and m/z 71 are related to proline as well. The two fragments at m/z 56 and m/z 55 are attributed to internal backbone fragments, C_2NOH_2^+ and C_2NOH^+ , respectively. In the spectrum of G₄H, shown in Figure 6.4.1(d), most of the ions observed result from fragmentation taking place at the side chain of histidine. The spectrum shows at m/z 110 the immonium ion of histidine ($\text{C}_5\text{H}_8\text{N}_3^+$), as observed in CID experiment [221], accompanied by an ammonia loss ($-\text{NH}_3$) at m/z 93 and by the additional loss of HCN at m/z 83. The fragments at m/z 82 and 81 result from further hydrogen losses, the latter being the dominant fragment. The peak at m/z 87 is attributed to the a_2 fragment. The fragment at m/z 115 can be attributed to the internal backbone fragment GG or the b_2 fragment, which has the same elemental composition.

For the mass spectrum of G₄R, presented in Figure 6.4.1(e), some fragments remain unassigned, and the others are mainly due to fragmentation at the protonated arginine side chain. We do not observe the immonium ion of arginine at m/z 129, but we do observe the related ions at m/z 112, 87, 70. Other fragments at m/z 71, 72, 73, 88, and 85 are attributed to the arginine residue, the fragment's elemental compositions are given in Table 6.1. Only the fragments at m/z 70, 71, 72, 88, and 112 have been reported in CID experiments. Note that the intense peak at m/z 87 most likely also has a contribution from the a_2 ion. GG/ b_2 fragments are also observed at m/z 115.

In the spectrum of G₄K (Figure 6.4.1(f)), we were not able to clearly identify either the fragments around m/z 70 or those around m/z 100. These fragments are not from the background and were, therefore, included in the analysis. We observed the y_1 fragment at m/z 147 and an -NH₂ loss from the y_1 fragment at m/z 131. m/z 67 and m/z 84, attributed to C₅H₇⁺ and C₅H₁₀N⁺ ions from the lysine side chain have also been observed in CID [175]. Other lysine-related fragments are observed at m/z 129, m/z 112, and m/z 101, the latter being the immonium ion of lysine. m/z 87 is the a_2 fragment from the peptide. The fragment at m/z 115 can be attributed to the internal backbone fragment GG or the b_2 fragment as for G₄H and G₄R. As the amino-acid sequences of G₄K and G₄H are the same for the first four residues, it is expected to observe a_2 and b_2 for all peptides under study except PG₄. The internal fragment GG should also be a common fragment of all peptides. The fact that it is not observed in PG₄ (see Figure 6.4.2) suggest the fragment m/z is most likely due to a b_2 fragment rather than GG in G₄H, G₄R and G₄K.

To summarize, in the mass spectra of G₄K, G₄H, and G₄R, the majority of the ions are attributed to fragments from the side chain of the basic residues in the peptides, lysine, histidine, and arginine, respectively. PG₄, G₅, and G₃ do not contain a basic side chain; thus, the fragments observed arise only from backbone dissociation. Most of the observed fragments were also reported in collision-induced dissociation (CID) experiments [175], suggesting that, in our case, internal conversion followed by intramolecular vibrational redistribution of the internal energy is the main process following photoexcitation/photoionization and Auger decay, eventually leading to statistical dissociation of the peptides.

Table 6.1: Attribution of the fragments observed during our study for each peptide. The letters G and P stand here for the immonium ion forms of the amino acid glycine and proline, respectively.

Mass	G ₅ /G ₃	PG ₄	G ₄ H	G ₄ R	G ₄ K
28	C ₂ H ₄ ⁺ /CH ₂ N ⁺				
29	CHO ⁺				
30	G				
31	CH ₃ O ⁺				
42	C ₂ H ₄ N ⁺				
43	CHNO ⁺				
45	COOH ⁺				
55	C ₂ NOH ⁺	C ₂ NOH ⁺			
56	C ₂ NOH ₂ ⁺	C ₂ NOH ₂ ⁺			
67					C ₅ H ₇ ⁺
68		[P-2H] ⁺			
70		P		C ₄ H ₈ N ⁺	
71		[P+H] ⁺		C ₄ H ₇ O ⁺	
72				C ₄ H ₁₀ N ⁺	
81			C ₄ H ₅ N ₂ ⁺		
82			C ₄ H ₆ N ₂ ⁺		
83			C ₄ H ₇ N ₂ ⁺		
84					C ₅ H ₁₀ N ⁺
85				C ₃ H ₃ NO ₂ ⁺	
87			a ₂	C ₄ H ₉ NO ⁺ /a ₂	a ₂
88				C ₄ H ₁₀ NO ⁺	
93			C ₅ H ₅ N ₂ ⁺		
110			C ₅ H ₈ N ₃ ⁺		
112				C ₅ H ₁₀ N ₃ ⁺	
115			GG/b ₂	GG/b ₂	GG/b ₂
131					y ₁ -NH ₃
147					y ₁

6.4.2 Different protonation sites on amino groups in peptides

The total ion yields (TIY) of each molecule were calculated by summing the areas under the peaks of all detected fragments in the photo-induced mass spectra for each photon energy over the scanned energy range. No specific fragmentation channels could be identified over the scanned photon energies for the different studied peptides, as shown in the different partial ion yield spectra in Figure 6.4.8, Figure 6.4.11, Figure 6.4.13, Figure 6.4.16, Figure 6.4.20, and Figure 6.4.27.

6.4.2.1 $[\text{GGGGG}+\text{H}]^+$

Figure 6.4.4(a) presents the experimental TIY and calculated absorption spectra of G_5 at the N K-edge. The final DFT/ROCIS spectrum is the convolution of the two most populated conformers according to the REMD simulations that have been performed, with a weighting factor of 0.52 for the first conformer (conformer 1), shown in Figure 6.4.5, and a weighting factor of 0.47 for the second conformer (conformer 2), shown in Figure 6.4.5. Conformers 1 and 2 are very similar, conformer 1 being more compact, with the protonated N-terminus making hydrogen bonds with the oxygen atoms of the backbone and of the carboxyl group at the C-terminus of the peptides. Our calculated structures and DFT/ROCIS spectrum are in agreement with those obtained by Li et al. [222]. Figure 6.4.6 and Figure 6.4.7 show the respective calculated DFT spectra of conformers 1 and 2 with the different molecular orbitals responsible for the main transitions.

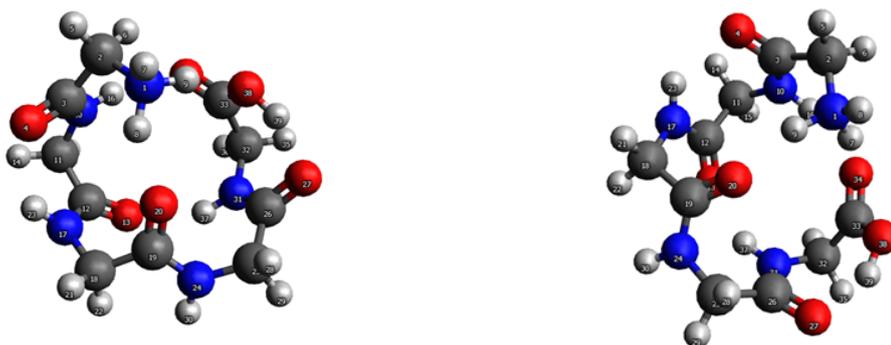


Figure 6.4.5: DFT-optimized geometry of conformer 1 (left) and conformer 2 (right) of G_5 (B3LYP/ZORA-TZVP).

The first resonance A at 401.6 eV is attributed to $\text{N } 1s \rightarrow \pi^*(\text{CONH})$ transitions within peptide bonds according to the DFT/ROCIS transition lines. This has already

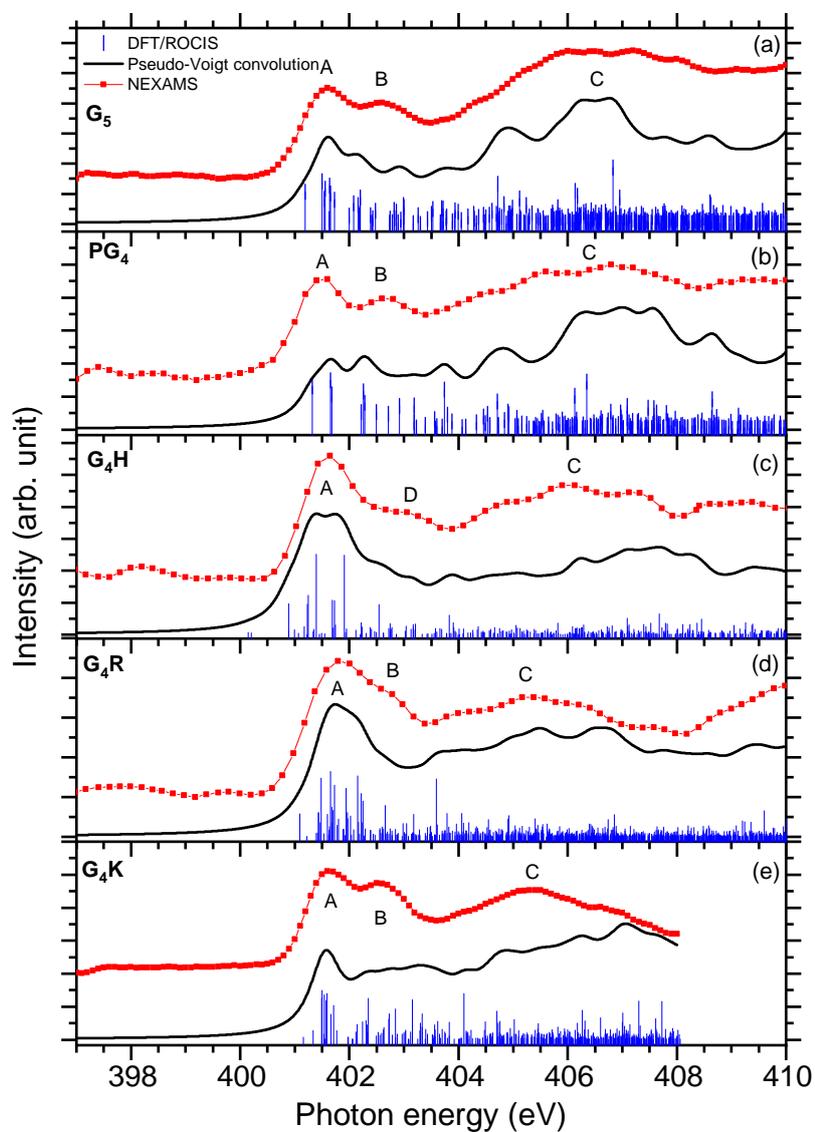


Figure 6.4.4: The experimental nitrogen K-edge total ion yield for G_5 , PG_4 , G_4H , G_4R and G_4K , are shown with red dotted lines. In blue, the DFT/ROCIS X-ray absorption lines are displayed for each molecule. The absorption lines were shifted by ≈ 13 eV and were broadened using a pseudo-Voigt profile with a 0.6 eV FWHM represented in black.

been reported for peptides in previous papers [79], [81], [86], [223]. The origin of resonance B at 402.6 eV is controversial. For gas-phase diglycine, Feyer et al. attributed this feature to a $N\ 1s \rightarrow \pi^*(N-C\alpha)$ transition [202], as well as Dörner et al. for protonated methionine-enkephalin [86]. Zubavichus et al. proposed that this feature is an $N\ 1s \rightarrow \pi^*(CONH)$ transition, energetically shifted because of potential different conformations existing in their studied homopolypeptides [223]. Using time-dependent density functional theory (TDDFT) to calculate NEXAFS spectra of α -helix and β -turn models based on glycines, Buckley and Besley attributed the first two narrow resonances at the N K-edge to $N\ 1s \rightarrow \pi^*(CONH)$, with the second band associated with π^* orbitals in a neighboring residue [91]. Their computed spectra suggest as well that this second feature is highly susceptible to the peptide's conformation, being more pronounced in an α -helix than in a β -turn. In the energy range of resonance B, the analysis of the lowest unoccupied molecular orbitals (LUMOs) of the conformers of G_5 (LUMO+1 of conformer 1 and LUMO+3 of conformer 2) leads to transitions from the N 1s level in one peptide bond to unoccupied $\pi^*(CONH)$ orbitals in neighboring peptide bonds. We finally attribute, according to our calculation and in agreement with the work of Buckley and Besley [91], the resonance B to $N_i\ 1s \rightarrow \pi^*(CON_jH)$ where N_i and N_j are different nitrogen atoms in the peptide backbone. The DFT/ROCIS calculations show two major transitions in the energy region of 404 - 408 eV, one resonance at 404.8 eV and one at 406.6 eV, that are contributing in the experimental spectrum to one broad spectral feature, noted C. The study of the molecular orbitals of the two main conformers of G_5 (LUMO+1 and LUMO+2 of conformer 1, LUMO of conformer 2) assigned the first resonance to a $N\ 1s \rightarrow \pi^*(C=O)$ transition in the carboxyl group of the C-terminus, where the DFT/ROCIS calculations show that the transition is due to excitation from the nitrogen atoms of the N-terminus to the carboxyl group. Such transition exists thanks to the folded conformation of G_5 and the spatial proximity of the N and C-termini. The peak at 406.6 eV is assigned to transitions located at the protonated N-terminus, $N\ 1s \rightarrow \sigma^*(NH_3^+)$ with the density of the molecular orbitals on the N-terminus (LUMO+5 of conformer 1 and conformer 2). Otero and Urquhart have previously reported similar transitions at 406.8 and 406.5 eV in solid-state NEXAFS spectra of zwitterionic glycine and glycine hydrochloride, respectively, and assigned them to both $N\ 1s \rightarrow \sigma^*(NH)$ and $N\ 1s \rightarrow \sigma^*(CN)$ transitions [166].

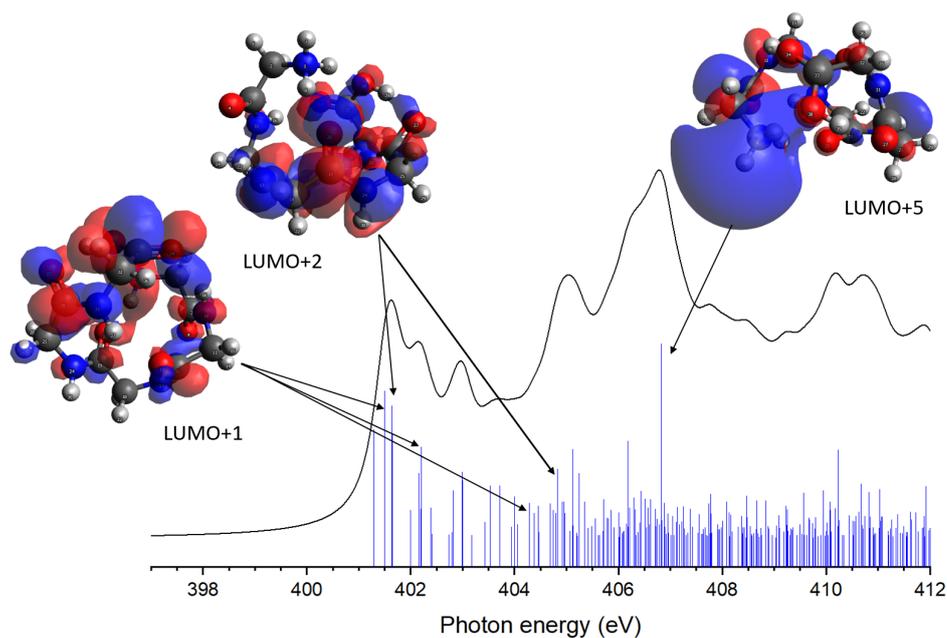


Figure 6.4.6: Calculated spectrum of conformer 1 of G₅ with the respective LUMOs responsible for the main transitions.

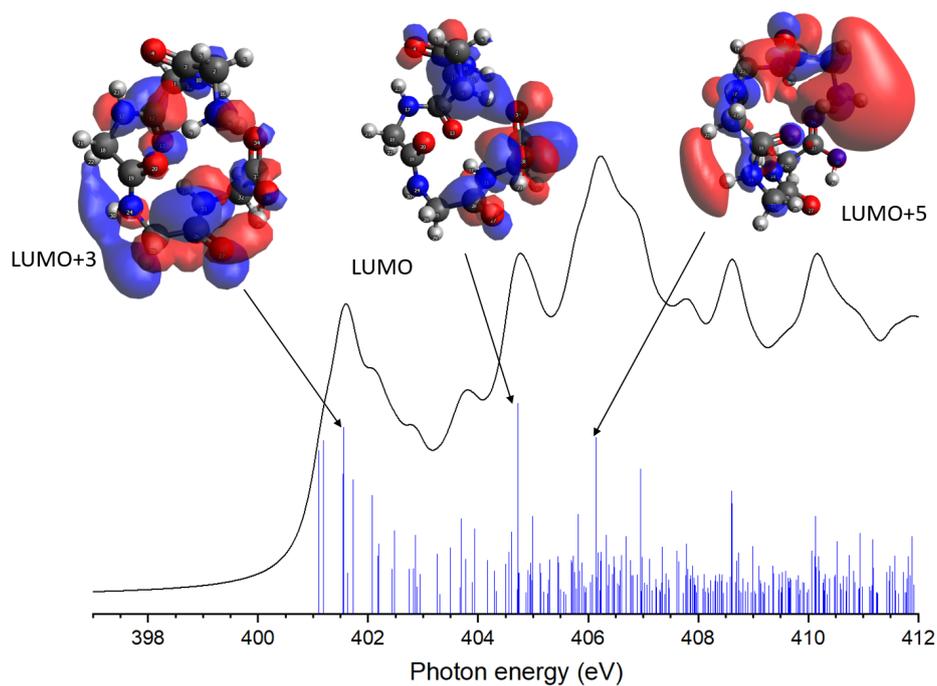


Figure 6.4.7: Calculated spectrum of conformer 2 of G₅ with the respective LUMOs responsible for the main transitions.

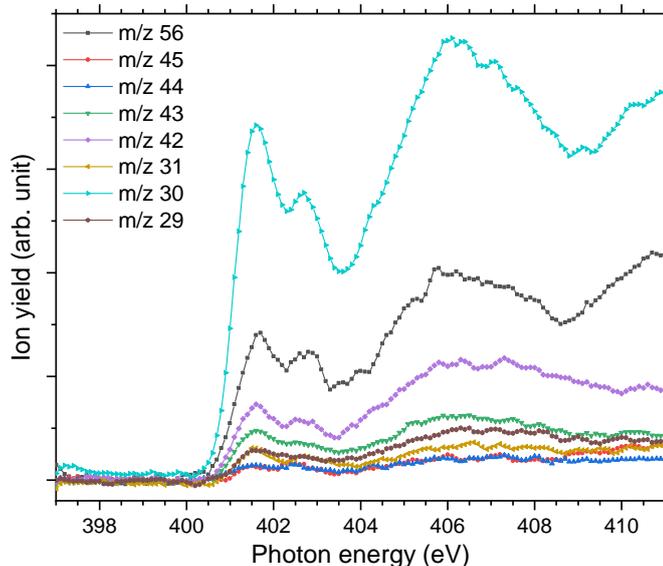


Figure 6.4.8: Experimental nitrogen K-edge partial ion yield spectra of G_5 .

6.4.2.2 [PGGGG+H]⁺

The NEXAMS spectrum of protonated PG_4 , presented in Figure 6.4.4(b), is similar to the spectrum of protonated G_5 . The DFT-calculated structure of the main conformer (weighting factor of 0.97), shown in Figure 6.4.9, is folded with the N-terminal nitrogen making hydrogen bonds to the oxygen atoms of the C-terminus and the backbone. The proline group stands out from the folded system. Figure 6.4.10 shows the respective calculated DFT spectrum of PG_4 with the different molecular orbitals responsible for the main transitions.

The first resonance A at 401.6 eV is attributed to $N\ 1s \rightarrow \pi^*(CONH)$ transitions in a peptide bond according to the DFT/ROCIS transition lines (LUMO). In the energy range of resonance B, the DFT/ROCIS calculations show a resonance at 402.1 eV and one at 403.7 eV, involving the molecular orbitals LUMO+1 and LUMO+3, which can both be attributed to electronic transitions from one peptide bond to a neighboring peptide bond. However, experimentally, only a resonance at 402.6 eV has been observed. We therefore attribute the resonance B to $N_i\ 1s \rightarrow \pi^*(CON_jH)$ transitions to neighboring peptide bonds, as discussed for G_5 and consistently with our observations for the different protonated pentapeptides of this study. Feature C is composed of different electronic transitions explaining the peak's broadness. The DFT/ROCIS

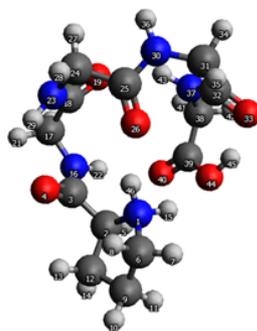


Figure 6.4.9: DFT-optimized geometry of PG₄ (B3LYP/ZORA-TZVP).

result gives a resonance at 404.7 eV mainly due to an N 1s $\rightarrow \sigma^*(\text{NH})$ transition from the backbone's nitrogen atoms, whereas at 406.5 eV, the calculated molecular orbital (LUMO+7) shows most of the density around the proline residue and describes an N 1s $\rightarrow \sigma^*(\text{CH})$ electronic transition due to the excitation of the protonated secondary amine nitrogen. Note that, for the immonium ion (m/z 70), shown in Figure 6.4.11, the intensity of feature C is lower than for the other fragments contributing to the total ion yield spectrum.

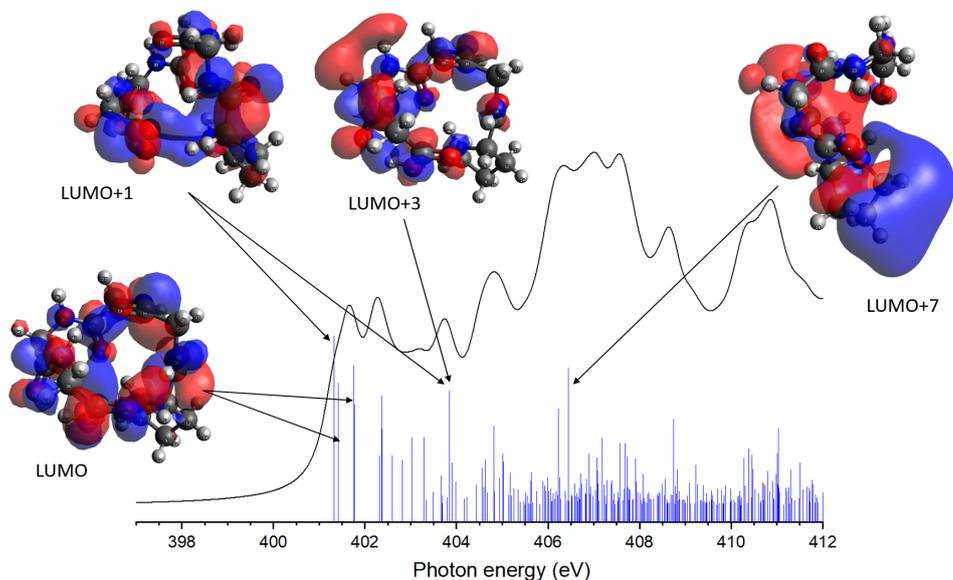


Figure 6.4.10: Calculated spectrum of PG₄ with the respective LUMOs responsible for the main transitions.

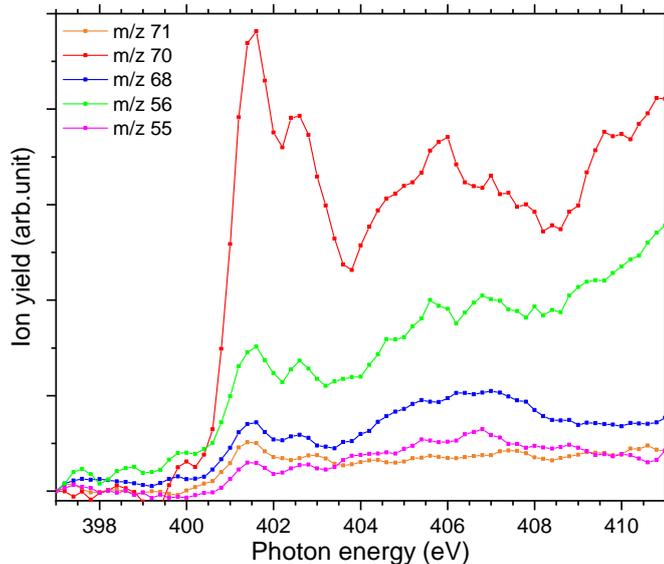


Figure 6.4.11: Experimental nitrogen K-edge partial ion yield spectra of PG_4 .

6.4.2.3 $[\text{GGGGH}+\text{H}]^+$

The experimental NEXAMS spectrum of G_4H , shown in Figure 6.4.4(c), also presents three main features. For G_4H , only the conformer with a weighting factor of 0.953 was studied. The other calculated conformers have a weighting factor below 0.01 and thus were not considered in the final analysis. The DFT-optimized geometry of G_4H is shown in Figure 6.4.12. The structure is mostly folded around the backbone of the peptide. The protonated nitrogen of the histidine side chain makes hydrogen bonds with the oxygen atoms of the backbone. Contrary to G_5 , the N-terminus does not make any hydrogen bonds with the backbone, which leads to G_4H being less compact than G_5 . Figure 6.4.14 shows the respective calculated DFT spectrum of G_4H with the different molecular orbitals responsible for the main transitions.

The first resonance A is located at 401.6 eV and can be partly attributed to $\text{N } 1s \rightarrow \pi^*(\text{CONH})$ transitions in peptide according to the analysis of the DFT/ROCIS transition lines at 401.2 eV. Additionally, the main resonance peak overlaps energetically with $\text{N } 1s \rightarrow \pi^*(\text{C}=\text{N})$ electronic transitions in the imidazole ring of the histidine side chain. The calculations give the transition lines at 400.8 eV, 401.4 eV, and 401.9 eV due to the double bond $\text{C}=\text{N}$ of the histidine's imidazole ring (LUMO) and the transition lines at 401.2 eV, involving the LUMO+1 and LUMO+2, stemming from

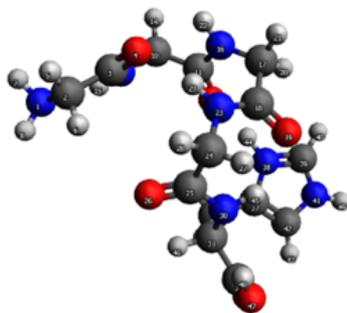


Figure 6.4.12: DFT-optimized geometry of G_4H (B3LYP/ZORA-TZVP).

the $N\ 1s \rightarrow \pi^*(CONH)$ peptide bond transitions. These additional histidine-related transitions contribute to the increased intensity and broadening of resonance A in comparison with the other peptides. In contrast to the other studied pentapeptides, neither experimentally nor theoretically, does the absorption spectrum of protonated G_4H show a resonance B at 402.6 eV, attributed to $N_i\ 1s \rightarrow \pi^*(CON_jH)$ transition in neighboring peptide bonds. Only in the partial ion yield spectrum of the immonium ion ($m/z\ 110$) (Figure 6.4.13) is resonance B at 402.6 eV distinguishable. That could be attributed to the fact that absorption in the histidine side chain in resonance A is not leading to the production of the immonium ion (similarly to what was discussed for the proline immonium ion in PG_4). Thus, resonance A is relatively smaller for this fragment, and resonance B can be observed. It is worth noting here that the PIY of G_4H is dominated by the ion yield of the fragment $m/z\ 81$ that involves H_3CN loss from the histidine side chain (Figure 6.4.13). This fragment shows a very high yield at the resonance A, which most probably correlates with the $N\ 1s \rightarrow \pi^*(C=N)$ electronic transitions within the histidine side chain.

Resonance D at 403.1 eV is found, experimentally, at significantly higher energy than resonance B at 402.6 eV for the other protonated pentapeptides. The DFT/ROCIS calculation gives, at 402.5 eV, a resonance due to $N\ 1s \rightarrow \sigma^*(CH)$ and $N\ 1s \rightarrow \sigma^*(NH)$ electronic transitions (LUMO+6) in the histidine side chain where the protonation site is located. For G_4H , the resonance at 403.1 eV is the signature of the protonation site, which explains the low intensity of the broad feature C compared to the intensity of feature C in G_5 . The broad spectral feature located in C between 404 and 408 eV remains unclear as it is due to different electronic transitions from the backbone of the protonated peptides. The DFT/ROCIS transitions are all of low intensity and cannot be attributed to one resonance because the molecular orbitals are highly delocalized. This feature has a contribution from LUMO+13 and LUMO+19 for the transition

lines at 407.1 eV and from LUMO+34 for the transition lines around 407.9 eV. The analysis of the LUMO+13 and LUMO+19 orbitals shows that feature C includes contributions from σ^* transitions from the backbone and the side chain.

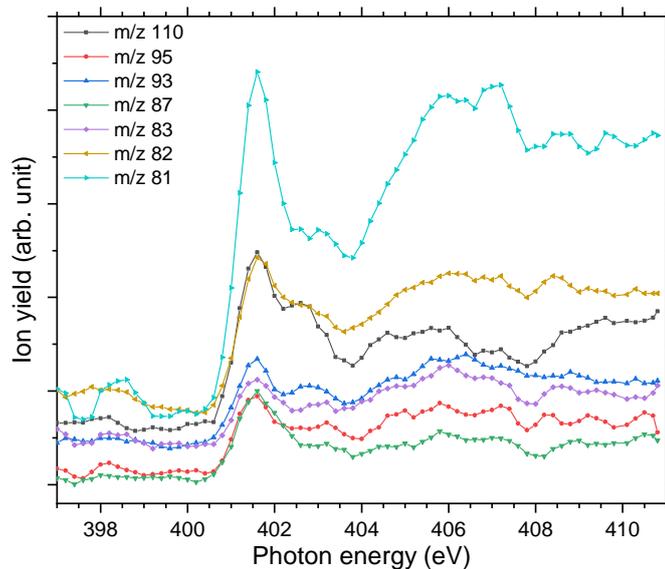


Figure 6.4.13: Experimental nitrogen K-edge partial ion yield spectra of G_4H .

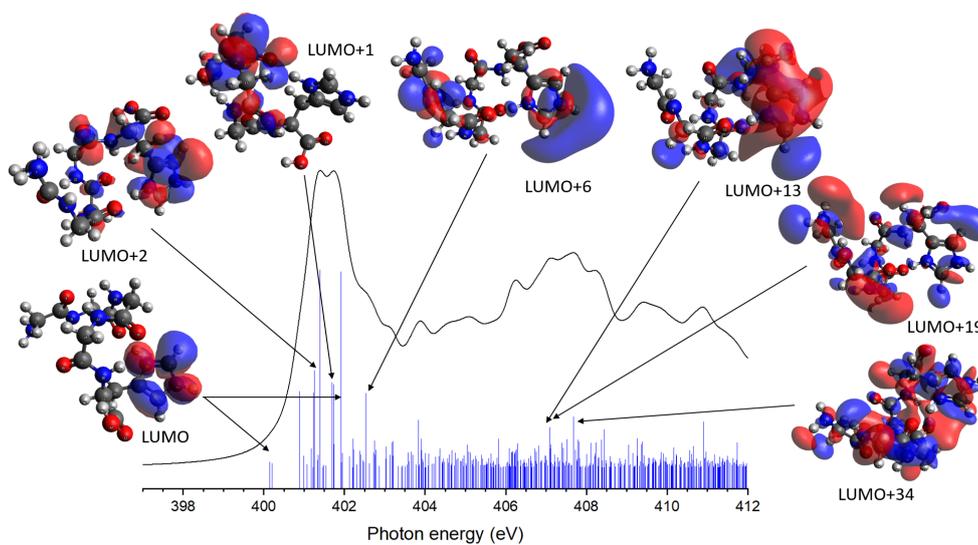


Figure 6.4.14: Calculated spectrum of G_4H with the respective LUMOs responsible for the main transitions.

6.4.2.4 [GGGGR+H]⁺

In Figure 6.4.4(d), the experimental NEXAMS spectrum of G₄R at the nitrogen K-edge is compared with the result of the DFT/ROCIS simulations. The final DFT/ROCIS spectrum is the convolution of the two most populated conformers according to the REMD simulations that have been performed, with a weighting factor of 0.63 for the first conformer (conformer 1, shown in Figure 6.4.15) and a weighting factor of 0.28 for the second conformer (conformer 2, shown in Figure 6.4.15). The calculated structures are both folded around the side chain and the C-terminus. The N-terminus of both conformers does not form hydrogen bonds with the C-terminus nor with the backbone or the side chain. Instead, the guanidinium group, where the proton is located, establishes a hydrogen bond network with the carboxyl group and the backbone. For conformer 1, the nitrogen atom with the protonation site forms only one hydrogen bond with an oxygen atom of the backbone, whereas the one from conformer 2 forms a hydrogen bond with a nitrogen atom of the first glycine as well as one with an oxygen atom. Conformer 2 establishes more hydrogen bonds than conformer 1, which leads to a more folded structure. Figure 6.4.17 and Figure 6.4.18 show the respective calculated DFT spectrum of conformers 1 and 2 with the different molecular orbitals responsible for the main transitions.

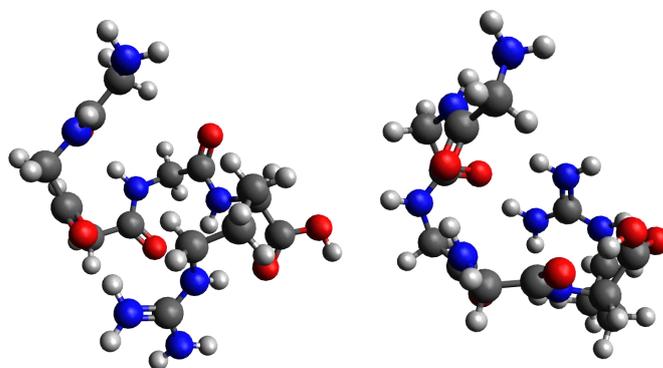


Figure 6.4.15: DFT-optimized geometry of conformer 1 (left) and conformer 2 (right) of G₄R (B3LYP/ZORA-TZVP).

In the experimental spectrum (Figure 6.4.4(d)), the first two resonances are not well resolved compared to the other studied peptides. The first resonance A at 401.6 eV has been attributed to the peptide bond resonance $N\ 1s \rightarrow \pi^*(CONH)$ coming from transitions to the first LUMOs of both conformers (LUMO to LUMO+6 for conformer 1 and from LUMO to LUMO+2 for conformer 2). According to our DFT/ROCIS

calculations, two resonances are responsible for the second feature B at 402.6 eV: $N_i 1s \rightarrow \pi^*(CON_jH)$ (LUMO of conformer 1) to a neighboring peptide bond and $N 1s \rightarrow \pi^*(C=N)$ (LUMO+4 of conformer 2) in the arginine side chain. For the case of zwitterionic arginine, Zubavichus et al. observed a π^* resonance located around 401.4 eV that they attributed to the guanidine group of arginine, whereas Leinweber et al. observed the π^* resonance of arginine at 401.7 eV [80], [165]. Moreover, Stewart-Ornstein et al. attributed the resonance at 402.9 eV of a Sub6 solid sample, an arginine-containing antimicrobial peptide, to a $\pi^*(C=N)$ transition in the arginine residue [224]. In the case of protonated G_4R , our calculations show that this resonance is shifted higher in energy to lie between the two peptide bond resonances at 402 eV. The overlap with this second transition explains why the resonance here is not clearly separated by one eV from the first peptide bond resonance, as some transition lines from the DFT/ROCIS calculations are related to the $\pi^*(C=N)$ resonance. Interestingly, the resonance B is more apparent in the PIY of the fragment m/z 73 ($C_2H_7N_3^+$) that is formed of the intact guanidine group of the arginine side chain. It can therefore be understood that the $N 1s \rightarrow \pi^*(C=N)$ transition in the arginine would likely result in a C=N bond cleavage, explaining why these transitions are not observed in the PIY of m/z 73 (Figure 6.4.16) and thus better resolving the separation of resonances A and B.

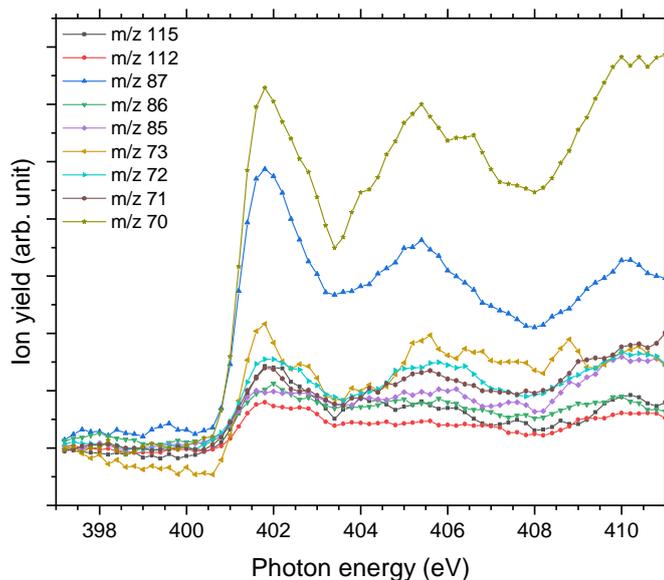


Figure 6.4.16: Experimental nitrogen K-edge partial ion yield spectra of G_4R .

The DFT/ROCIS calculations show a transition at 403.6 eV that can possibly be responsible for the small feature before the broad one between 404 and 408 eV. This transition is attributed to $N\ 1s \rightarrow \sigma^*(NH_2^+)$ transitions from the arginine side chain due to the additional proton for conformer 1 (LUMO+9). For conformer 2, it is more ambiguous as there are two main transition lines at 403.96 eV and 404.1 eV that are responsible for this small feature. The transition at 403.96 eV is due to the molecular orbital (LUMO+2) and is attributed to $N\ 1s \rightarrow \pi^*(C=O)$ transitions, whereas at 404.1 eV, the molecular orbital (LUMO+13) shows most of the density around the arginine side chain, with some of the density in the backbone as well, and is attributed to a $N\ 1s \rightarrow \sigma^*(NH)$ transition due to an electronic excitation in the protonated side chain. Therefore, we concluded that the fingerprint of the protonation site is the first resonance of feature C, at 404 eV. We observed a broad feature, noted C between 404 eV and 408 eV, which is less intense than feature C of the other protonated peptides of this study relative to feature A. As the spectra are normalized on the intensity of peak A, this effect could be explained by a higher intensity of peak A in the case of G₄R because of the additional $\pi^*(C=N)$ transition from the arginine side chain that can be found between resonances A and B. The DFT/ROCIS calculations show different electronic excitations responsible for feature C with very low amplitude, coming mainly from highly delocalized molecular orbitals. Comparing with the NEXAMS spectra of the other pentapeptides of this study, we conclude that the feature is mainly due to σ^* transitions from the backbone.

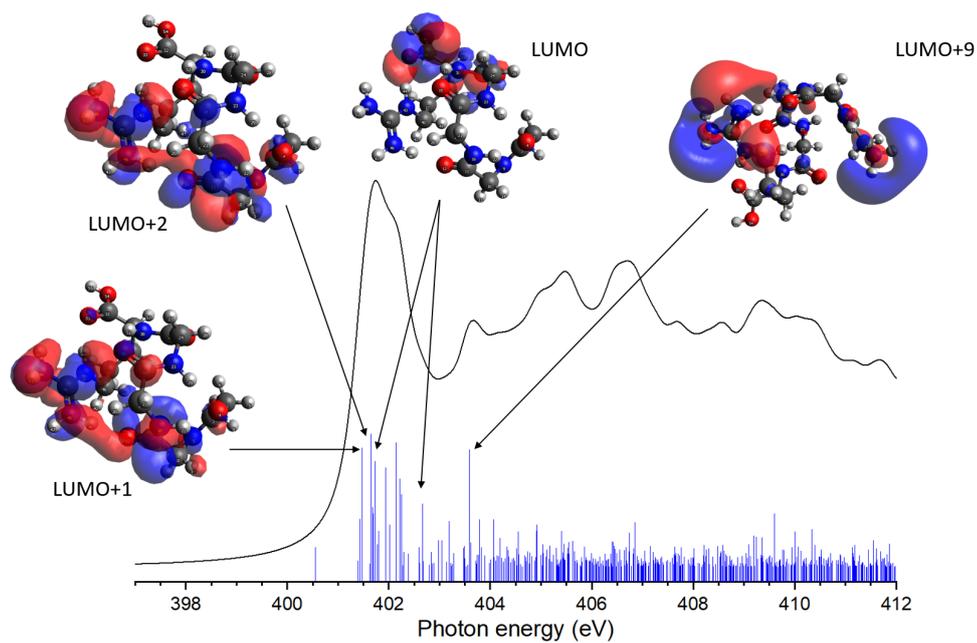


Figure 6.4.17: Calculated spectrum of conformer 1 of G₄R with the respective LUMOs responsible for the main transitions.

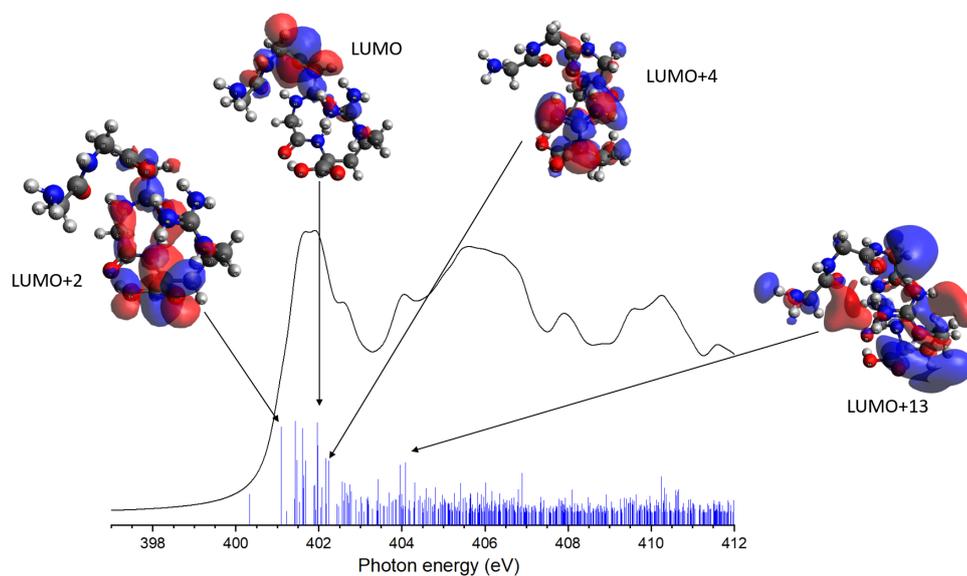


Figure 6.4.18: Calculated spectrum of conformer 2 of G₄R with the respective LUMOs responsible for the main transitions.

6.4.2.5 [GGGGK+H]⁺

The REMD calculation of the conformers of G₄K gave two main structures responsible for the experimental spectrum, which we refer to as conformer 1 and conformer 2 in the following, with a weighting factor of 0.45 each. The structures, shown in Figure 6.4.19, are mostly folded, with the protonated amino group of the lysine side chain establishing hydrogen bonds with oxygen atoms along the peptide backbone and at the C-terminus. Figure 6.4.21 and Figure 6.4.22 show the respective calculated DFT spectra of conformer 1 and 2 with the different molecular orbitals responsible for the main transitions.

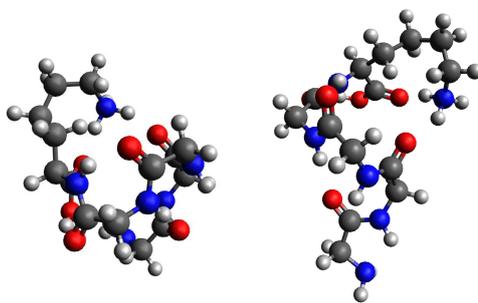


Figure 6.4.19: DFT-optimized geometry of conformer 1 (left) and conformer 2 (right) G₄K (B3LYP/ZORA-TZVP).

The experimental spectrum of G₄K presented in Figure 6.4.4(e) has the same three main features as the other peptides studied. The first peak A is the main peptide bond resonance $N\ 1s \rightarrow \pi^*(CONH)$ at 401.6 eV (LUMO+1 of conformer 1 and LUMO+2 of conformer 2). This is consistent with the results we obtained for the other peptides. As for protonated G₄R, the second feature at 402.6 eV is due to different contributions. The DFT/ROCIS calculation of conformer 1 shows that the peak at 402.4 eV only originates from $N_i\ 1s \rightarrow \pi^*(CON_jH)$ transitions to neighboring peptide bonds (LUMO+5). On the other hand, the results of the DFT/ROCIS simulation for conformer 2 show that the resonance at 402.6 eV has major contributions from two different electronic transitions. The first one at 402.3 eV is attributed to $N\ 1s \rightarrow \pi^*(COOH)$ (LUMO+5) in the C-terminus but with the excitation coming from the nitrogen of the fourth glycine rather than the nitrogen of the C-terminus connected to the lysine residue. The second electronic transition is $N\ 1s \rightarrow \sigma^*(NH)$ in the protonated lysine side chain at 403.1 eV (LUMO+3). The third feature is due to $N\ 1s \rightarrow \sigma^*(CH)$ electronic transitions from the backbone (LUMO+8 of conformer 1) around 404.5 eV, and $N\ 1s \rightarrow \sigma^*(NH)$ from the backbone and protonation site (LUMO+3 of conformer 1, LUMO+4 and LUMO+3,

of conformer 2) at 407 eV. Relative to the yield of the $N\ 1s \rightarrow \pi^*(CONH)$ resonance, the signature of the protonation site in G_4K is less intense than for G_5 .

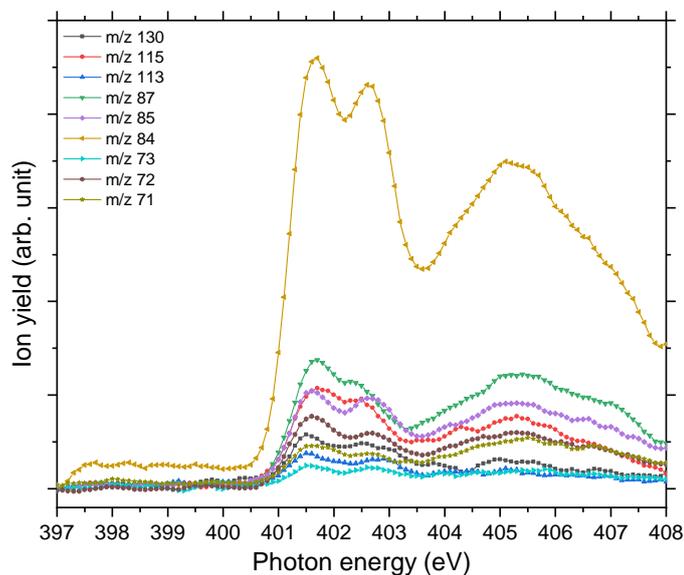


Figure 6.4.20: Experimental nitrogen K-edge partial ion yield spectra of G_4K .

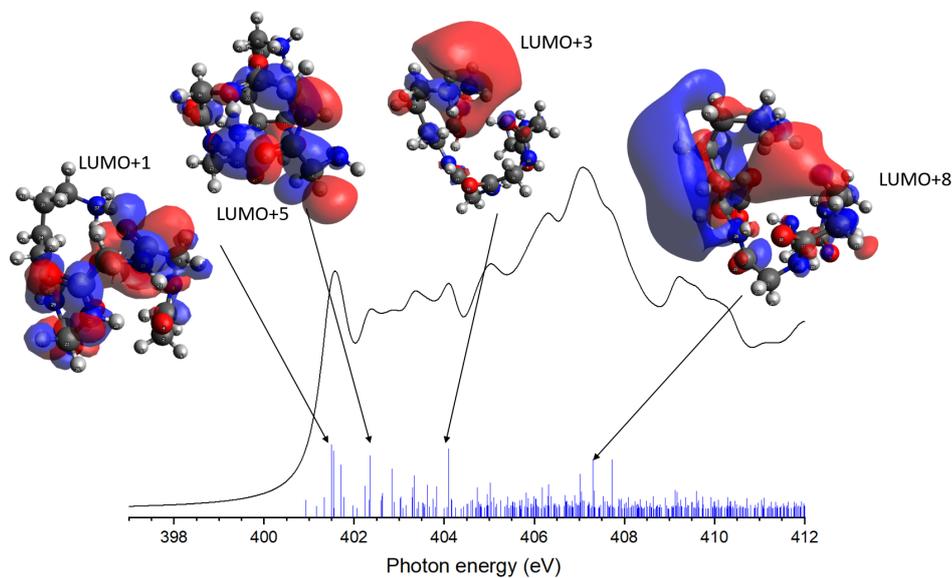


Figure 6.4.21: Calculated spectrum of conformer 1 of G_4K with the respective LUMOs responsible for the main transitions.

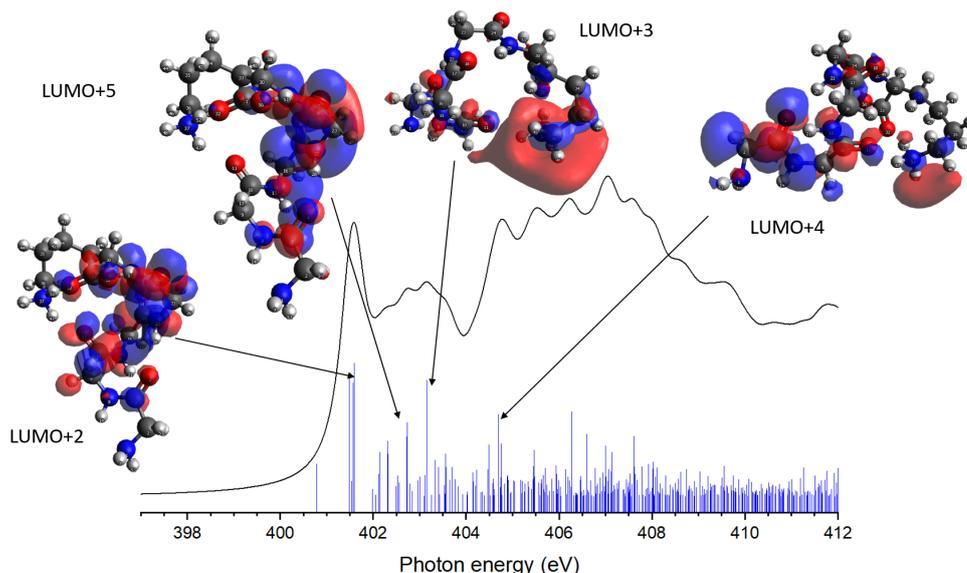


Figure 6.4.22: Calculated spectrum of conformer 2 of G_4K with the respective LUMOs responsible for the main transitions.

6.4.2.6 Conclusion

The study of these five protonated model pentapeptides at the N K-edge revealed some features that are common to any protonated peptide. The main results of the study are summarized in Table 6.2. As was reported previously, all peptides show a strong resonance related to electronic transitions in the peptide bonds, always located at 401.6 eV [80], [81], [202]. In the NEXAMS spectra, all peptides show a resonance at 402.6 eV that we attribute to transitions from a nitrogen 1s level in one peptide bond to $\pi^*(CONH)$ antibonding orbitals in a neighboring peptide bond, except for protonated G_4H , which has a feature centered at 403.1 eV due to $N\ 1s \rightarrow \sigma^*(CH)$ transitions in the histidine side chain where the protonation site is.

For protonated G_4R , the $\pi^*(C=N)$ transition from the arginine side chain contributes to the resonance at 401.6 eV and the one at 402.6 eV, which results in the resonance appearing broader than for the other peptides. Similarly, G_4H shows $\pi^*(C=N)$ transitions in the histidine's imidazole ring, which contribute to the resonance at 401.6 eV. The NEXAMS spectrum of every peptide shows a broad feature between 404 and 408 eV with different intensities relative to the $N\ 1s \rightarrow \pi^*(CONH)$ resonance. The comparison of the experimental spectra with DFT/ROCIS calculations shows that the feature comes from σ^* resonances from the peptide's backbone with major contribution from $\sigma^*(NH)$ resonances located at the protonated nitrogen site of these peptides, both from the basic side chain and the N-terminus. In PG_4 and G_4H , we observed that the

immonium ion exhibits a lower intensity of feature C, which can be explained by the resonant excitation in the peptide bond leading to more immonium ion production compared to the absorption of the photon in the side chain, around 406 eV. Interestingly, we did not observe any resonance from the unprotonated N-terminus in the experimental total ion yield spectra in the cases of G₄H, G₄R, and G₄K. Furthermore, for peptides with a protonated side chain, we observed that resonance A is more intense than resonance C as compared to the peptides with a protonation site on the N-terminus. The additional transitions in the side chains energetically close to the resonances A and B lead to resonance A being more intense than for G₅ or PG₄. We could observe differences between G₄H, G₄R, and G₄K NEXAMS spectra in the shape and the intensity of the resonances A and B. The total ion yield spectrum of G₄H does not exhibit any resonance at 402.6 eV but one at 403.1 eV. The resonance B appears as a shoulder on resonance A for G₄R, while for G₄K resonance B is almost as intense as resonance A. We discussed that these transitions are better resolved, or absent, in the PIY of certain fragments, especially when comparing e.g. immonium ion and fragments involving C=N bond break. These differences distinguish them from each other. As shown in the previous section, the high intensity of resonance C is due to the protonation site located on the N-terminal. For neutral diglycine in the gas phase, the feature observed at 406 eV is weaker and shows about the same intensity as the resonance at 401.6 eV [82], which agrees with our conclusion.

Table 6.2: Experimental energy position (exp) and proposed assignment of the spectral feature observed in the total nitrogen K-edge NEXAMS spectra of the different peptides as well as the energies of the DFT/ROCIS absorption lines (calc).

Peak	Exp (eV)	Calc (eV)	Assignment
G₅			
A	401.6	401.6	N 1s \rightarrow π^* (CONH)
B	402.6	402.1	N _i 1s \rightarrow π^* (CON _j H)
C	404 - 408	404.8	N 1s \rightarrow π^* (C=O)
		406.6	N 1s \rightarrow σ^* (NH ₃ ⁺)
PG₄			
A	401.6	401.6	N 1s \rightarrow π^* (CONH)
B	402.6	402.1	N _i 1s \rightarrow π^* (CON _j H)
		403.7	
C	404 - 408	404.7	N 1s \rightarrow σ^* (NH)
		406.5	N 1s \rightarrow σ^* (CH)
G₄R			
A	401.6	401.7	N 1s \rightarrow π^* (CONH)
B	402.6	402.6	N _i 1s \rightarrow π^* (CON _j H)
			N 1s \rightarrow π^* (C=N)
C	404 - 408	403.6	N 1s \rightarrow σ^* (NH ₂ ⁺)
		403.96	N 1s \rightarrow π^* (C=O)
		404.1	N 1s \rightarrow σ^* (NH)
G₄H			
A	401.6	400.8	N 1s \rightarrow π^* (CONH)
		401.4	N 1s \rightarrow π^* (C=N)
D	403.1	402.5	N 1s \rightarrow σ^* (NH)
C	404 - 408	407.1	N 1s \rightarrow σ^*
		407.9	
G₄K			
A	401.6	401.7	N 1s \rightarrow π^* (CONH)
B	402.6	402.4	N _i 1s \rightarrow π^* (CON _j H)
		402.3	N 1s \rightarrow π^* (COOH)
		403.1	N 1s \rightarrow σ^* (NH)
C	404 - 408		N 1s \rightarrow σ^* (NH)
			N 1s \rightarrow σ^* (CH)

6.4.3 Protonation at the amide oxygen: The case of triglycine

Rodriquez et al. investigated the proton migration in protonated triglycine using density functional theory [225]. In this study, they found that the protonation of the N-terminal nitrogen atom is higher in free energy than the different isomers resulting from protonation at the carbonyl oxygen of the N-terminal residue. However, in a recent study, Li et al. [195] calculated that the protonation site of triglycine is the N-terminal nitrogen rather than the amine oxygen, which they found to be at least $0.61 \text{ kcal.mol}^{-1}$ more stable than the other protonation sites possible for triglycine. They suggested that experimental studies using soft X-ray spectroscopy could provide a way to distinguish the various G_3 isomers since NEXAFS spectra should show unique peaks at the nitrogen and oxygen K-edges for the different protomers. Using protonated pentaglycine, we were able to establish the N K-edge fingerprint of an N-terminal protonated peptide. Thus, comparing the NEXAMS spectra of G_5 and G_3 should show whether only conformers of G_3 with the protonation site at the N-terminal nitrogen are present under our experimental conditions. As protonation of an amine oxygen is expected to influence the NEXAMS spectra at the oxygen K-edge, we present, in the following, results for both, the N and O K-edges. The carbon K-edge spectra of G_5 and G_3 have also been recorded, but as they exhibit only little differences, with one main resonance at 288.3 eV for both protonated peptides, which comes from $\text{C } 1s \rightarrow \pi^*(\text{CONH})$ transitions from the backbone, they are only shown in Figure 6.4.23 and will not be further discussed in this manuscript. The partial ion yield spectra at the oxygen K-edge are shown in Figure 6.4.24 and in Figure 6.4.25, and the partial ion yield spectra of G_3 at the nitrogen K-edge are shown in Figure 6.4.27.

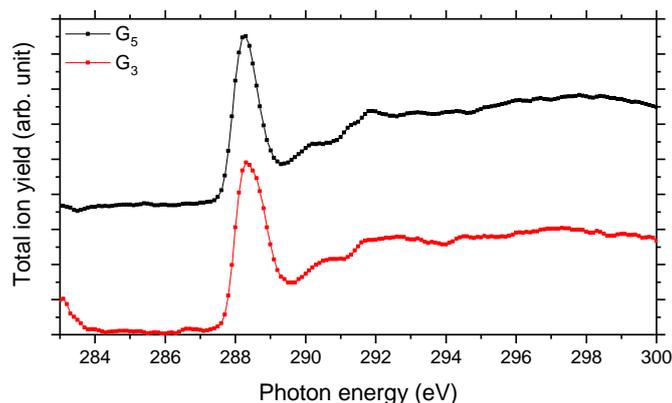


Figure 6.4.23: Experimental carbon K-edge total ion yield spectra of G_5 in black, and in red G_3 .

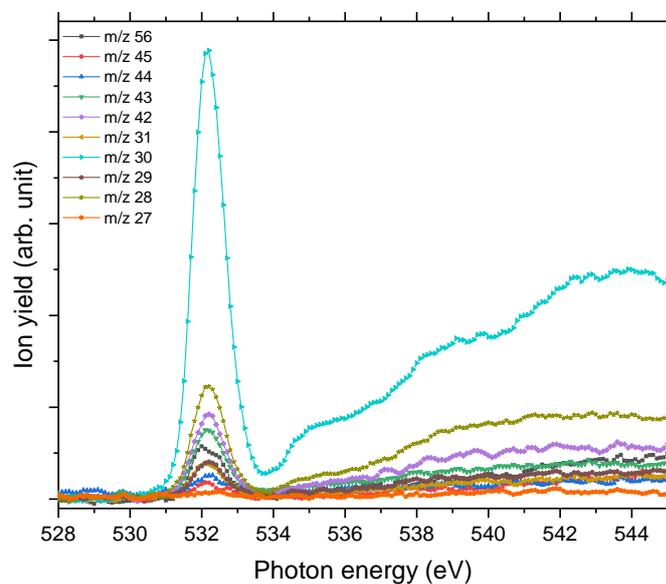


Figure 6.4.24: Experimental partial ion yield spectra at the nitrogen K-edge of G_5 .

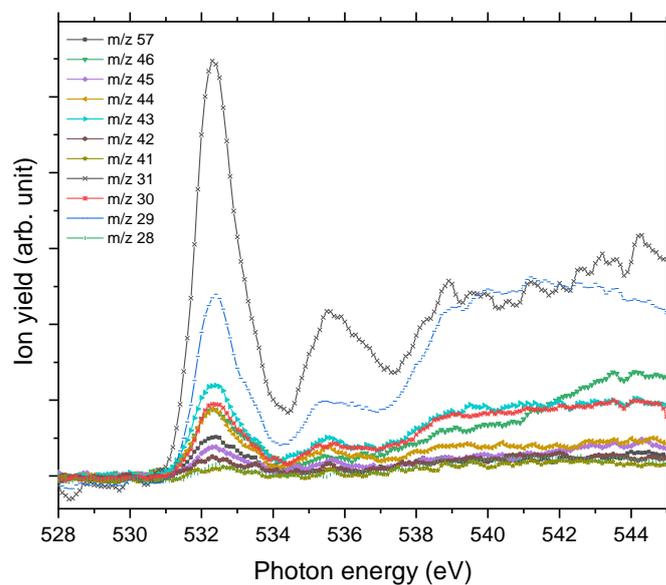


Figure 6.4.25: Experimental partial ion yield spectra at the nitrogen K-edge of G_3 .

6.4.3.1 Comparison of the experimental spectra of G_5 and G_3

In this section, we first compare the experimental NEXAMS spectra of protonated pentaglycine and triglycine. We performed DFT/ROCIS calculations on different isomers and possible protonation sites of triglycine to interpret the results of the experimental spectra.

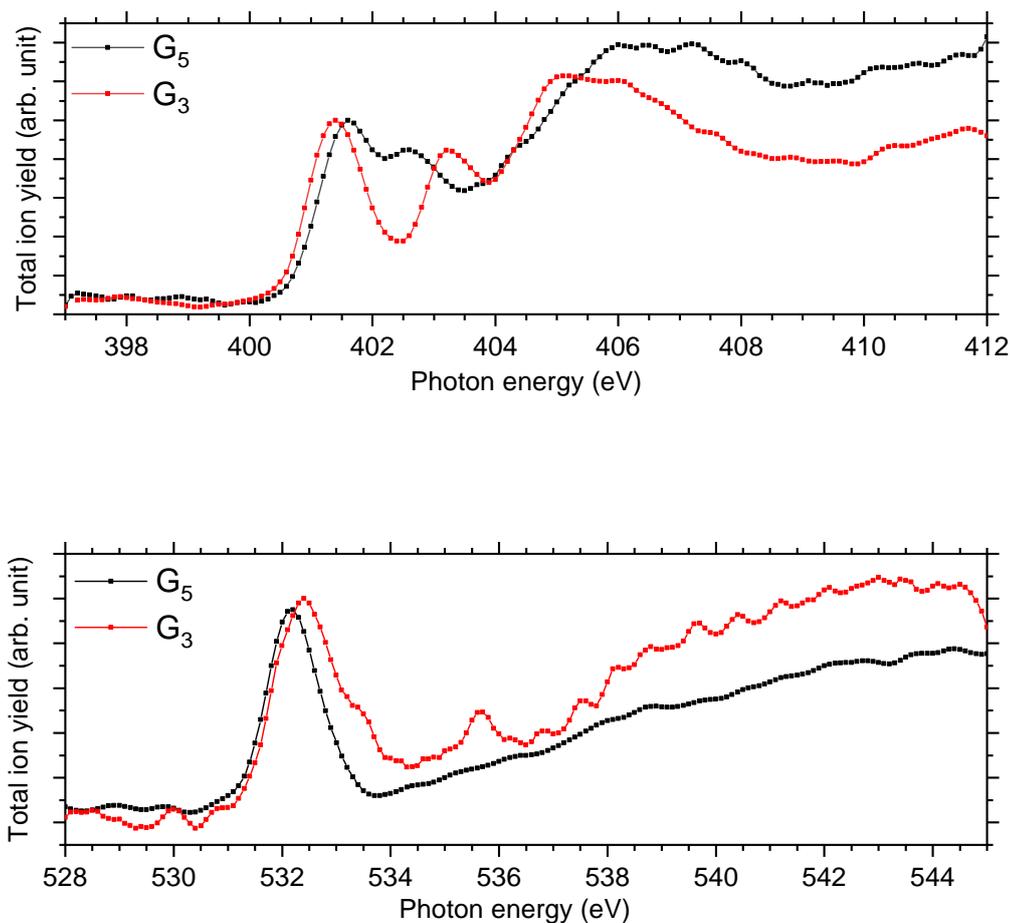


Figure 6.4.26: Experimental NEXAMS spectra at the N K-edge (top) and at the O K-edge (bottom) of G_5 in black and G_3 in red.

As shown in the previous section and in previous theoretical and experimental studies [190], [194], the protonation site of protonated pentaglycine is located at the N-terminus, i.e. on the amide nitrogen of the peptide. We have shown in the previous section that the electronic transition due to the protonation site of pentaglycine is

located in the broad feature centered at 406 eV. The protonation site is responsible for the relatively high intensity of this feature as compared to the main $N\ 1s \rightarrow \pi^*(\text{CONH})$ resonance. Thus, we can use this resonance to map the difference between the two glycine-based peptides.

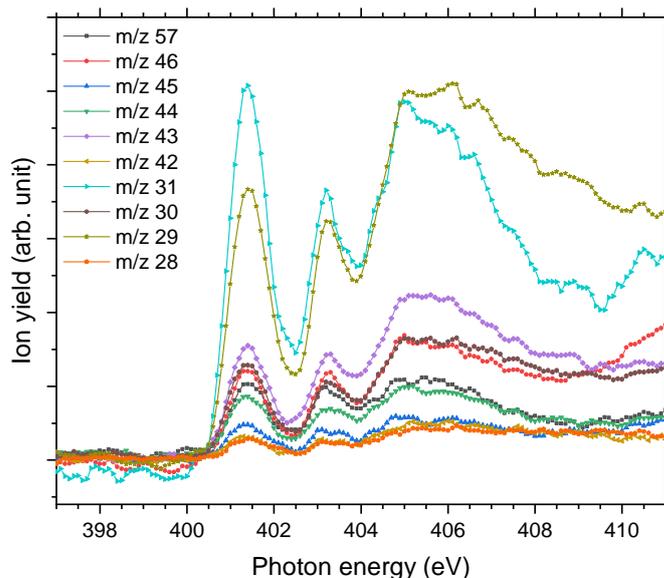


Figure 6.4.27: Experimental nitrogen K-edge partial ion yield spectra of G3.

At the N K-edge, see Figure 6.4.26 (top), we observed clear differences and similarities. For both protonated peptides, a resonance is located at 401.6 eV and is due to a $N\ 1s \rightarrow \pi^*(\text{CONH})$ electronic transition in the peptide bonds blue-shifted by 0.1 eV for G_3 compared to G_5 . The resonance at 402.6 eV is not observed for protonated triglycine, but we observed a feature at 403.1 eV that has not been observed for protonated pentaglycine. For G_4H , a resonance is observed at 403.1 eV for $N\ 1s \rightarrow \sigma^*(\text{NH})$. Note that in the NEXAMS spectrum of solid triglycine, no resonance between the one at 401.5 eV and the broad peak centered at 406 eV has been observed [82]. In the work of Gordon et al. [82] as well as of Zubavichus et al. [79] on solid-state triglycine, they attributed the feature at 406 eV to $\sigma^*(\text{NH})$ transitions from the amide group that is protonated in the zwitterionic form, which is consistent with our conclusion on the protonation site of peptides. This suggests that the peak we observe at 403.1 eV in the N K-edge spectrum of G_3 could be due to an electronic transition from a different protonation site than the one of pentaglycine. This being said, both protonated peptides

exhibit this broad feature between 404 and 408 eV that we attributed previously to excitations to the $\sigma^*(\text{NH})$ and $\sigma^*(\text{CH})$ orbitals, with an intense contribution from an $-\text{NH}_3^+$ protonated amino group. This suggests, on the contrary, that triglycine would also be protonated on its N-terminus. Thus, from the experimental spectra only, we can highlight differences but cannot conclude whether triglycine has one or two protonation sites, as the experimental NEXAMS spectrum could suggest two different ones. A comparison with DFT/ROCIS calculations is shown in the next section for a better assignment of the different resonances. At the O K-edge, see Figure 6.4.26 (bottom), both spectra show an intense resonance centered at 532.2 eV, blue-shifted by 0.2 eV for G_5 compared to G_3 . This feature has previously been observed in protonated peptides [17], [206], [223] and proteins [87], [226] as well as for gas-phase diglycine by Feyer et al. [202]. This resonance is attributed to an $\text{O } 1s \rightarrow \sigma^*(\text{CONH})$ transition in the peptide bonds. In their study, Gordon et al. also attributed the resonance at 532.2 eV of solid diglycine and triglycine to an $\text{O } 1s \rightarrow \pi^*(\text{CONH})$ resonance with a minor contribution of the $\text{O } 1s \rightarrow \pi^*(\text{COO}^-)$ electronic transition [82]. G_3 exhibits a shoulder at 533.4 eV that is not observed in the case of G_5 and was not observed for gas-phase diglycine. Dörner et al. attributed this transition to an $\text{O } 1s \rightarrow \pi^*(\text{COOH})$ transition in the carboxyl group, which was observed at 532.5 eV for leucine-enkephalin [86]. In the case of protonated triglycine, a second resonance is observed at 535.6 eV that is not observed in the experimental spectrum of G_5 . This resonance has been attributed to an $\text{O } 1s \rightarrow \pi^*(\text{C=O})$ transition in the carboxyl group by Feyer et al. for gas-phase glycine and diglycine [202]. The spectrum of protonated triglycine at the oxygen K-edge does not exhibit any feature that at first glance would be associated to the protonation site.

To more deeply analyze the NEXAMS spectra for triglycine and better understand our experimental observation and the differences with pentaglycine, we compared the result of the NEXAMS experiment of protonated triglycine with DFT/ROCIS calculations.

6.4.3.2 Theory/experiment comparison at the N K-edge

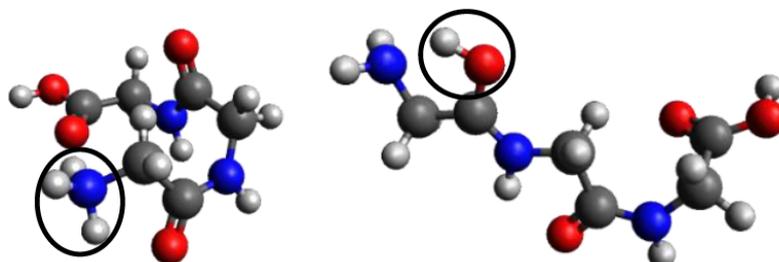


Figure 6.4.28: Calculated structures for GGGH1 (left) and GGGH2 (right) (B3LYP/ZORA-DEF2-TZVP). The dark circles highlight the position of the protonation site.

In the case of protonated triglycine, a random structure search was performed to get different structures of this protonated peptide. Theoretically, the structure search has been performed at 0 K to probe the lowest-energy conformers of triglycine that should compare with our experimental condition, where the ions in the trap were cryogenically cooled down to 8 K. The result of the conformer search and geometry optimization gives two main protomers, noted GGGH1 and GGGH2. The calculated structures of GGGH1 and GGGH2 are shown in Figure 6.4.28. The first protomer, GGGH1, has a more folded structure, with hydrogen bonds between the protonated N-terminus and the carboxyl oxygen atoms. In contrast, the GGGH2 structure is more linear, with the proton bound to the first amide oxygen, making a hydrogen bond with the terminal nitrogen. In the theoretical study of Li et al., they found out that the most stable conformer of the traditional N-terminal protonation site of triglycine is $2.53 \text{ kcal.mol}^{-1}$ more stable than the most stable one of the O-protonated conformers at the highly accurate coupled cluster singles and doubles (CCSD) level of theory [195]. Thus, it is expected that the experimental spectrum will be reproduced by mainly the calculated spectrum of GGGH1.

Figure 6.4.29 presents the results of the DFT/ROCIS calculations for the absorption spectra of both isomers at the nitrogen K-edge. Figure 6.4.30 and Figure 6.4.31 show the respective calculated DFT spectra of GGGH1 and GGGH2 at the N K-edge with the different molecular orbitals responsible for the main transitions.

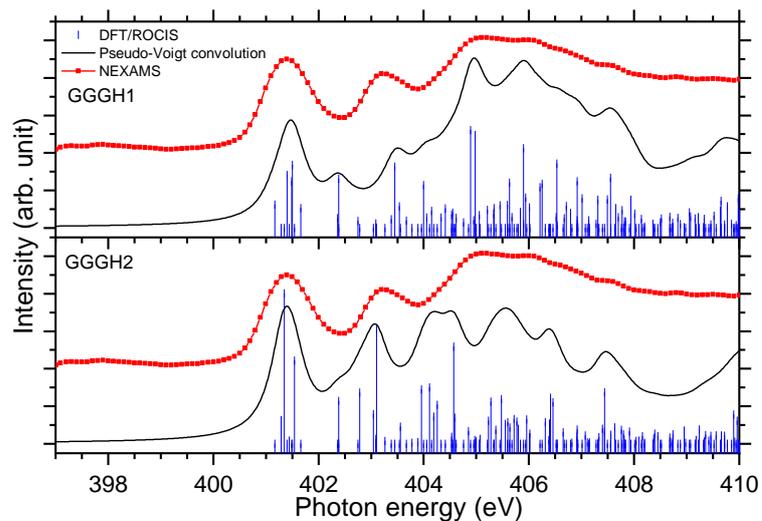


Figure 6.4.29: Experimental nitrogen K-edge ion yield spectrum in red of G_3 as well as DFT/ROCIS spectrum for the calculated structure GGGH1 (shifted +12.07 eV) and for the calculated structure GGGH2 (shifted +12.56 eV). The absorption lines were broadened using a pseudo-Voigt profile with a 0.6 eV FWHM represented in black.

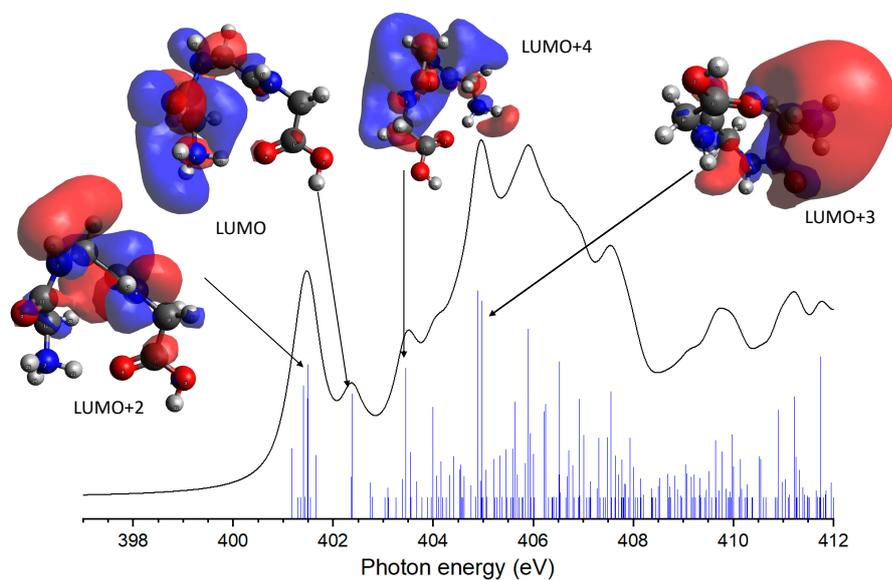


Figure 6.4.30: Calculated spectrum of GGGH1 at the N K-edge with the respective LUMOs responsible for the main transitions.

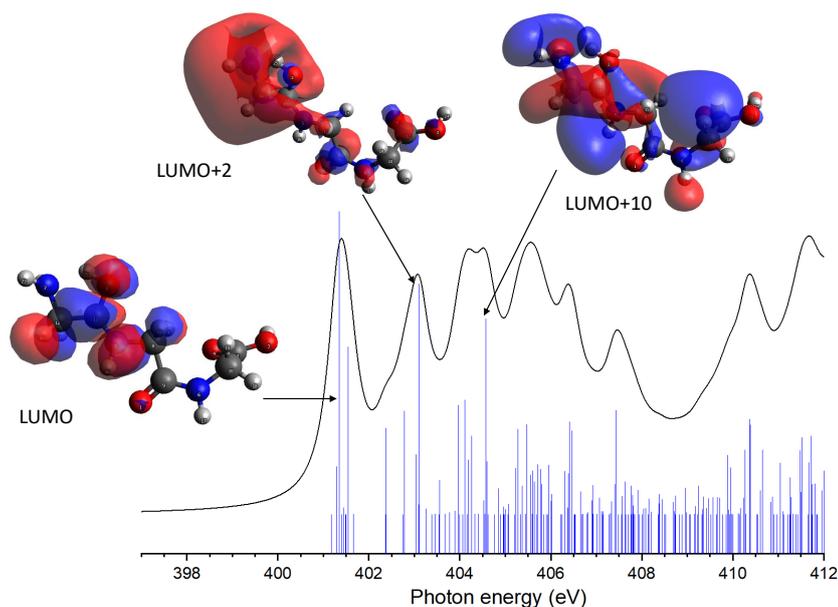


Figure 6.4.31: Calculated spectrum of GGGH2 at the N K-edge with the respective LUMOs responsible for the main transitions.

At first glance, the two calculated spectra show strong differences, especially in the region from 402 eV to 408 eV, and neither spectrum reproduces the experimental spectrum well overall. As for the other peptides, the DFT/ROCIS transition lines of the first resonance at 401.6 eV stem from $N\ 1s \rightarrow \pi^*(\text{CONH})$ transitions in the peptide bonds for both protomers. According to our calculations, in the case of GGGH1, the resonance at 403.4 eV is attributed to a transition from $N\ 1s$ to $\sigma^*(\text{CH})$ orbitals (LUMO+4) in the backbone of triglycine, whereas in the case of GGGH2, it is a $N\ 1s \rightarrow \sigma^*(\text{NH})$ transition in the unprotonated terminal- NH_2 (LUMO+2). Interestingly, the resonance calculated for the GGGH2 conformer is more intense than it is in the case of GGGH1 and reproduces better the peak observed at this energy in the experimental spectrum. The broad feature centered at 406 eV is due to many transitions, as described previously for the protonated pentapeptides. For GGGH1, the DFT/ROCIS calculations reproduce well the broad feature, and the analysis of the different electronic transitions shows that, for this isomer, the feature is mainly due to $N\ 1s \rightarrow \sigma^*(\text{CH})$ transitions from the backbone and $N\ 1s \rightarrow \sigma^*(\text{NH})$ transitions from the protonated terminal- NH_3^+ . In the case of GGGH2, the calculated spectrum does not reproduce the high intensity of the feature between 404 and 408 eV. The main transition calculated at 404.8 eV (LUMO+10) is highly delocalized and can be due to $N\ 1s \rightarrow \sigma^*(\text{CH})$ or $N\ 1s \rightarrow \sigma^*(\text{NH})$ transitions from the backbone. The main resonances observed for G_3 at the nitrogen and oxygen K-edges are summarized in Ta-

ble 6.3. Ultimately, the experimental spectrum cannot be reproduced with sufficient confidence by either of the two calculated isomers since only the spectrum of GGGH2 can explain the resonance at 403.1 eV, while only the spectrum of GGGH1 reproduces the broader feature centered at 405 eV. As in the work of Wu and Mc Mahon [194], the comparison of the experimental spectrum with the calculated ones suggests that both isomers might be present under our experimental conditions and contribute to the measured NEXAMS spectrum. In this experiment conducted at the Ion Trap end-station, only the linear ion trap was cooled down to 8 K. The other parts of the setup were at room temperature, such that we can expect to have different stable isomers of protonated triglycine transferred and cooled down in the trap, as explained by Li et al. [195] and already shown for biomolecular ions by Warnke et al. using ion mobility spectrometry (IMS) combined with a cryogenic ion trap [46]. In the study of Li et al. [195], the authors showed that at a temperature below 150 K, all isomer conversions are forbidden; thus, if many conformers are present in the experimental setup, they can only come from the way the ions were first generated, which shows that if many conformers are already in the solution, we should observe them when doing spectroscopy. In their study, Warnke et al. investigated the conformation of biomolecular ions using their experimental setup, where they selected a certain conformer by means of their IMS before performing IR spectroscopy at cryogenic temperature [46]. By comparing the IR spectra of conformer-selected ions with the hot conformer-selected one, they demonstrated that the latter can be reproduced by summing both conformer-selected spectra with a 4:1 ratio, both conformers differing in size by less than 1 %. Together, these results suggest that the room-temperature structures of a biomolecular ion are still conserved during the cooling down process in the cryogenic ion trap.

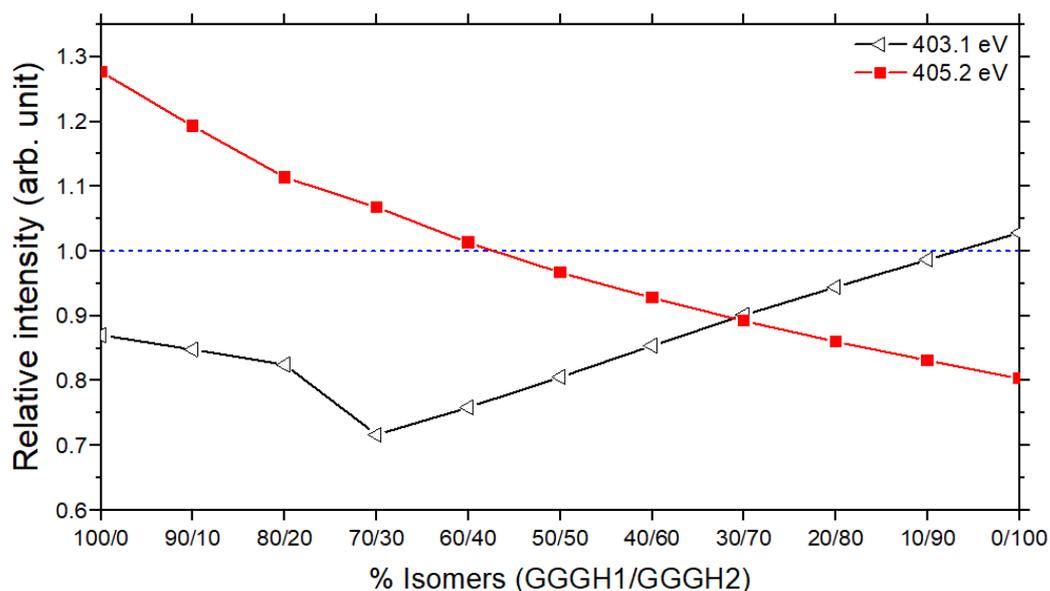


Figure 6.4.32: Relative intensities of the calculated resonances at 403.1 eV (black, $\sigma^*(\text{NH})$ of the unprotonated $-\text{NH}_2$ group) and at 405.2 eV (red, $\sigma^*(\text{CH})$ from the backbone), compared to the experimental ones, as a function of the isomer mixing ratio GGGH1 and GGGH2. The dashed blue line represents a relative intensity of 1.

We have therefore attempted to estimate the relative populations of the two isomers present in our experiment by mixing the contribution of both isomers into a new calculated spectrum and by comparing the peak intensities with the experimental spectrum. The two resonances we considered as markers for the comparison are the one at 403.1 eV to obtain the contribution of GGGH2 to the experimental spectrum and the broad feature centered at 405.2 eV to obtain the contribution of GGGH1, which are both related to the protonation state of the N-terminus. To do this, we normalized the experimental spectrum and the calculated spectra to the first resonance, which is the peptide bond resonance, which is similar in both calculated spectra. After normalization, we compared, for different isomer mixing ratios, the relative intensity between the calculated and experimental peak height for both resonances. The results are shown in Figure 6.4.32.

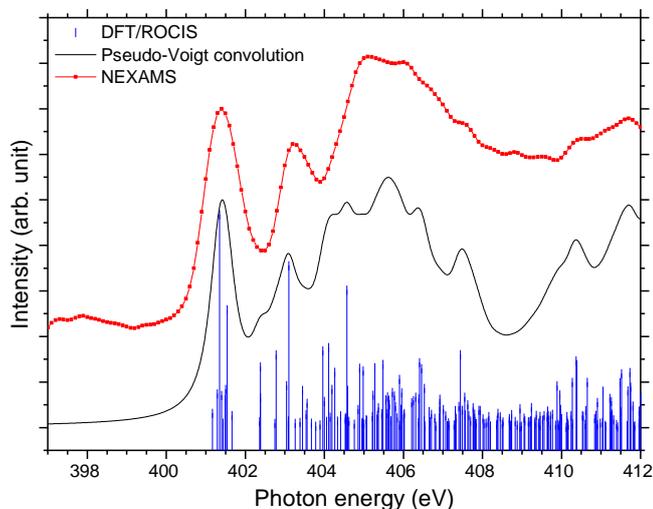


Figure 6.4.33: Comparison of, in red, the experimental NEXAMS spectra of protonated triglycine at the N K-edge and, in black, the result of the convolution of the two calculated spectra GGGH1 and GGGH2 with 30% of GGGH1 and 70% of GGGH2.

On the basis of the spectral comparison, we concluded that our experimental spectrum arises from a relative population of N-protonated triglycine of 30% with 70% of O-protonated triglycine. This ratio is the closest one to the experimental spectrum with 90% of relative intensity for both resonances. The result of the isomer mixing on the calculated absorption spectra is shown in Figure 6.4.33. This is in very good agreement with the results of Voss et al., who performed IR-IR double resonances of protonated triglycine in an ion trap at 10 K and concluded with a relative population of 35% N-protonated and 65% O-protonated [163]. To obtain this result, the authors combined IR-IR cryogenic ion vibrational predissociation spectroscopy (IR-IR CIVS spectroscopy), which allows for isomer-specific spectroscopy of mass-selected ions combined with the harmonic calculation of the chosen structure of protonated triglycine. They used weakly-bound D_2 as a messenger tag to ensure low internal energy in the triglycine ions and to reduce spectral congestion. The main isomer found under their experimental condition is the same amide protonated structure (O-protonated) identified by Wu and Mc Mahon [194]. This structure differs from our O-protonated structure by having a trans bond between the N-terminus and the amide when our calculated structure has a cis-bond between the two groups, but their ratio is similar to what we observed under our experimental condition. Moreover, they showed that their experimental spectrum could only be explained if 35% of N-protonated triglycine

with an H-bond between the N-terminus and the carboxyl group is also present in their trap. Overall, this is in agreement with what was found when comparing the NEXAMS spectrum of protonated triglycine with the DFT calculations done for the present study. It should also be pointed out that this population is more representative of the isomers present at 300 K originating from the ESI source rather than the 0 K population distribution.

Table 6.3: Experimental energy position (exp) and proposed assignment of the spectral feature observed in the total nitrogen K-edge and oxygen K-edge NEXAMS spectra of G_3 as well as the energies of the DFT/ROCIS absorption lines (calc).

Exp (eV)	Calc (eV)	Assignment
G_3		
401.6	401.5	N 1s \rightarrow $\pi^*(\text{CONH})$
403.2	403.1	N 1s \rightarrow $\sigma^*(\text{NH}_2)$
404 - 408	404.8	N 1s \rightarrow $\sigma^*(\text{CH})$
		N 1s \rightarrow $\sigma^*(\text{NH}_3^+)$
532.3	532.4	O 1s \rightarrow $\pi^*(\text{CONH})$
535.5	534.7	O 1s \rightarrow $\pi^*(\text{C=O})$
	535.6	O 1s \rightarrow $\sigma^*(\text{OH})$

6.4.3.3 Theory/experiment comparison at the O K-edge

Figure 6.4.34 shows the NEXAMS spectrum obtained at the oxygen K-edge and its comparison with DFT/ROCIS calculations for the two main isomers of protonated triglycine. Figure 6.4.35 and Figure 6.4.36 show the respective calculated DFT spectra of GGGH1 and GGGH2 at the O K-edge with the different molecular orbitals responsible for the main transitions.

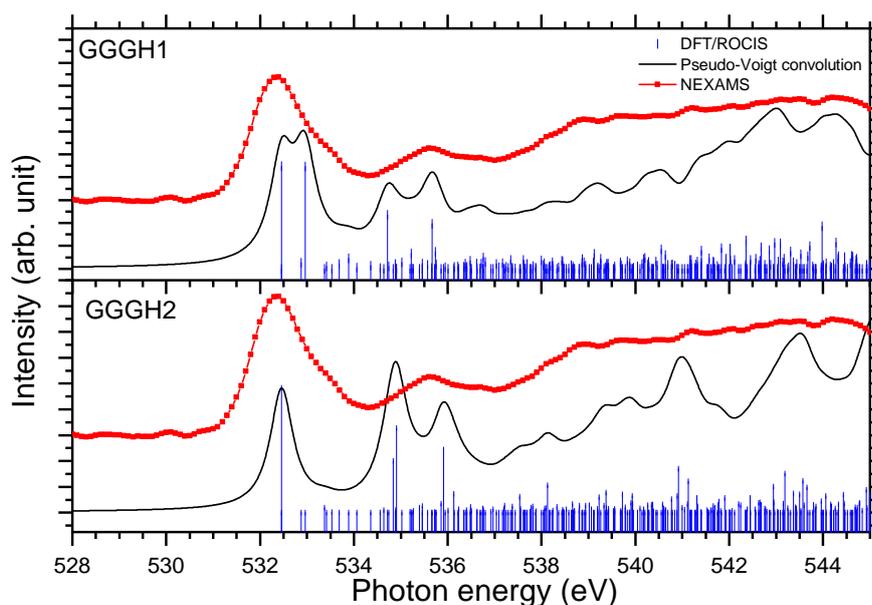


Figure 6.4.34: Experimental oxygen K-edge ion yield spectrum in red of G_3 as well as DFT/ROCIS spectrum (shifted +14.15 eV) for the calculated structure GGGH1 and for the calculated structure GGGH2 (shifted +14.76 eV). The absorption lines were broadened using a pseudo-Voigt profile with a 0.6 eV FWHM represented in black.

The different resonances are summarized in Table 6.3. For GGGH1, the first resonance at 532.4 eV is attributed to $O\ 1s \rightarrow \pi^*(CONH)$ transitions (LUMO and LUMO+2) in the peptide bond. The other main transition line at 533 eV, which is shown in the calculated spectrum, can be attributed to $O\ 1s \rightarrow \pi^*(C=O)$ (LUMO+1) in the carboxyl group starting from the oxygen atom of the $-C=O$ group. In contrast, in the case of GGGH2, we observed only the first resonance related to the $O\ 1s \rightarrow \pi^*(CONH)$ transition. For GGGH1, the pseudo-Voigt convolution of the DFT/ROCIS spectrum results in a much broader peak closer to the experimental result.

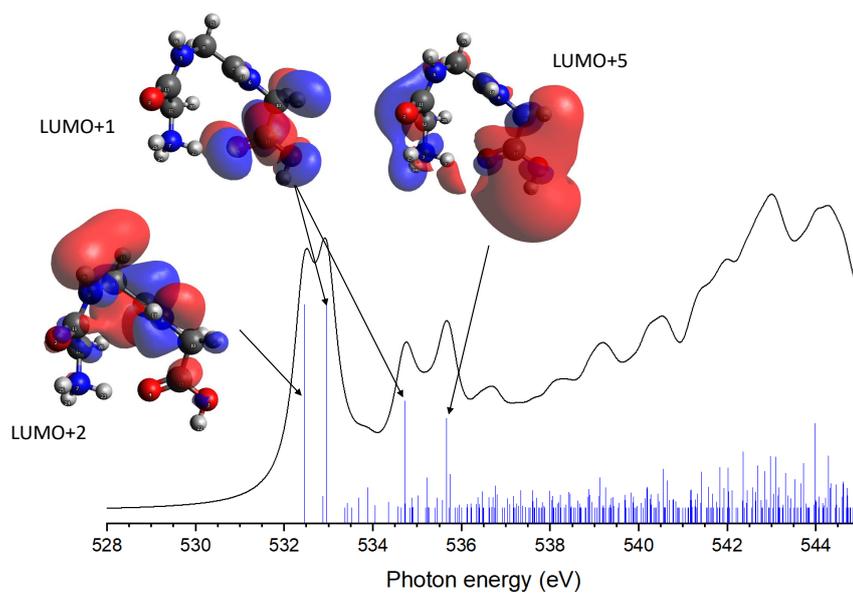


Figure 6.4.35: Calculated spectrum of GGGH1 at the O K-edge with the respective LUMOs responsible for the main transitions.

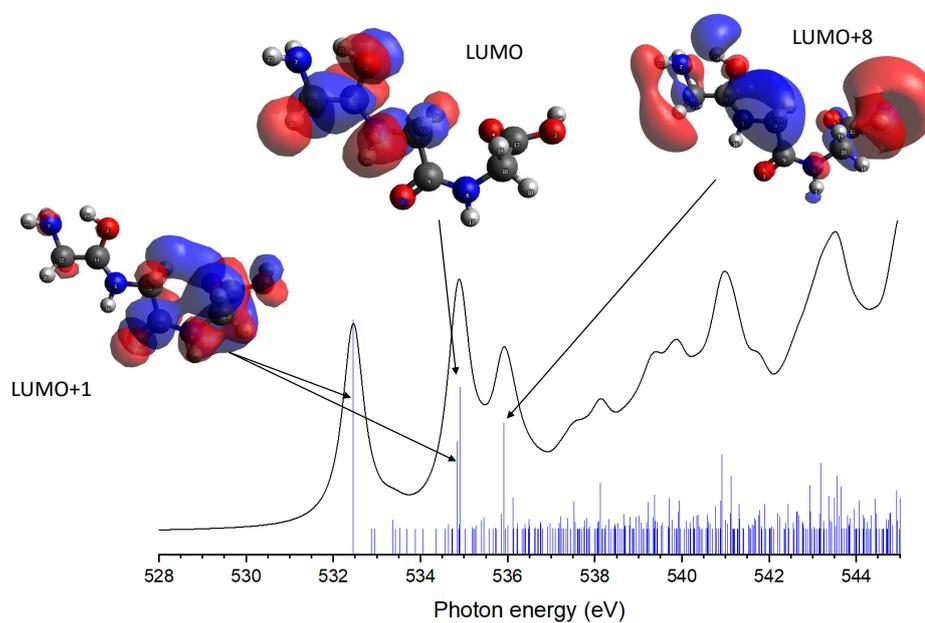


Figure 6.4.36: Calculated spectrum of GGGH2 at the O K-edge with the respective LUMOs responsible for the main transitions.

For both calculated spectra, the second feature includes two resonances, at 534.7 eV

and 535.7 eV for GGGH1 (534.9 eV and 535.9 eV for GGGH2), but only one was observed experimentally, centered at 535.6 eV. In the case of GGGH1, the first resonance at 534.7 eV is due to an $O\ 1s \rightarrow \pi^*(C=O)$ transition to the LUMO+1 in the carboxyl group but starting from the oxygen atom of the -OH group, and the second is attributed to $O\ 1s \rightarrow \sigma^*(OH)$ transitions (LUMO+5) in the carboxyl group of protonated triglycine. For GGGH2, the first resonance can be attributed to the peptide bond resonance $O\ 1s \rightarrow \pi^*(CONH)$ in the molecular orbitals LUMO and LUMO+1, whereas the resonance at 535.9 eV is due to $O\ 1s \rightarrow \sigma^*(OH)$ transitions (LUMO+8) at the C-terminus of the peptide, similarly to GGGH1. At higher energy, the spectra are mainly constituted of excitations to σ^* orbitals. These results are in good agreement with the previous study of Feyer et al. for neutral gas-phase diglycine [202].

To confirm the results obtained at the nitrogen K-edge, we combined the two calculated spectra at the oxygen K-edge with the same factor, 30/70, for GGGH1 and GGGH2 as in the previous section. The resulting graph is shown in Figure 6.4.37. Here, we achieved a good agreement with the experimental spectrum. We reproduced the different experimental resonances with the calculations. Despite the relative intensity of the resonance at 535 eV being higher for the calculated spectrum than the experimental spectrum, the attributed electronic transitions agree well with our results and the previous studies [81], [82], [202].

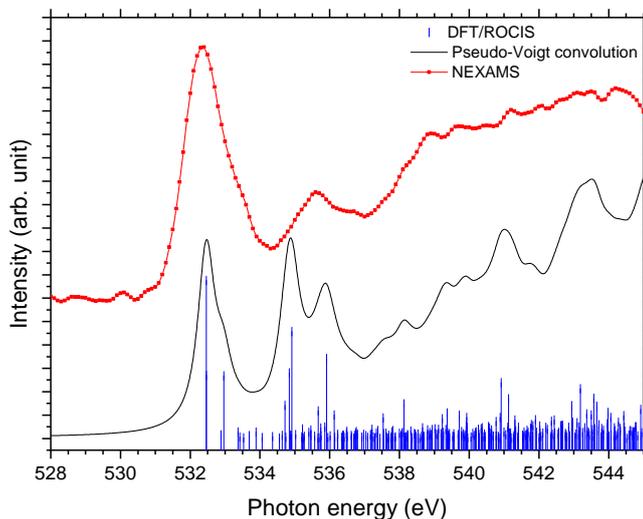


Figure 6.4.37: Comparison of, in red, the experimental NEXAMS spectra of protonated triglycine at the O K-edge and, in black, the result of the convolution of the two calculated spectra GGGH1 and GGGH2 with 30% of GGGH1 and 70% of GGGH2.

6.4.3.4 Conclusion

In this section, we have shown that there are clear differences between a protonated amide nitrogen and a protonated amine oxygen and that the difference is clearly detectable in the N K-edge NEXAMS spectra. When the protonation site is located at the oxygen, we observed a resonance at 403.1 eV that is not observed for the case of a peptide protonated at a nitrogen atom; this resonance stems from an N 1s \rightarrow $\sigma^*(\text{NH})$ transition in the unprotonated $-\text{NH}_2$ N-terminus. However, our experimental spectrum still contains the signature of a protonated N-terminus with a broad and intense feature centered around 405 eV, which can be explained by our experimental conditions. In fact, during this experimental campaign, only the ion trap was held at 8 K, whereas the other parts of the setup were at room temperature, which allowed different conformers to coexist in the trap. Moreover, as shown in the study of Wu and Mc Mahon [194], smaller peptides of glycine are more stable in a linear structure, whereas peptides containing more than three glycines are more stable in a folded structure. It can be assumed that as triglycine is of an intermediate size, the folded structure and the linear structure can coexist. According to the simulations, protonated triglycine is more stable in a folded structure when the proton is on the N-terminal, and the linear structure is the most stable with the proton on the first amide oxygen. By comparing the DFT/ROCIS calculations with the N K-edge experimental spectrum, we could confirm that the latter is representative of a mixing of protomers with a relative population of N-protonated triglycine of 30% with 70% of O-protonated triglycine. At the oxygen K-edge, no striking difference appears between the two protomers on the calculated spectra, which indicates that the signature of the protonation site is best observed at the nitrogen K-edge. However, our measurement did not show most of the resonances expected by the theoretical work of Li et al. Their simulated NEXAFS spectra at the nitrogen K-edge did not show any feature around the energy of the peptide bond excitation, and we did not observe any π^* transition originating from the NH_2 terminal expected by Li et al. [195] but a σ^* transition lying higher in energy than the expected transitions. They predicted many transitions in the region of 404 - 406 eV that we were not able to resolve experimentally. At the oxygen K-edge, we also observed striking differences between our experimental spectrum and their simulated one: we did not observe any peak at 531.3 eV or at 533.6 eV, which they attributed to an O 1s \rightarrow $\pi^*(\text{C}=\text{O})$ transition only due to cyclic isomers or to linear isomers, respectively. One reason that could explain the differences between our results and their spectra can be the temperature used for the simulation; they calculated their spectra at a temperature of 498 K, while our ion trap was held at 8 K, which could

result in different conformations of the protomers in the trap which would affect the absorption spectra. Overall, the NEXAMS technique has proven to be sensitive to the protonation site of peptides, and we have shown that we can distinguish the different protomers present in the experimental setup and give an approximate ratio.

6.5 Conclusions

In conclusion, we have performed a NEXAMS study on five glycine-based protonated pentapeptides to investigate the effect of the protonation site on the electronic structure and to record its spectroscopic fingerprint in the soft X-ray range. We have shown that the combination of DFT/ROCIS calculations and soft X-ray spectroscopy allows us to map the electronic transitions of the protonation site. The influence of a protonation site on a nitrogen atom has only a moderate effect on the entire spectrum of protonated pentapeptides. Experimentally, the NEXAMS technique does not allow for straightforward identification of the resonances imparted to the protonation site of a peptide, essentially owing to the numerous overlapping transitions to σ^* orbitals in these molecular systems and the intrinsic band broadening imparted to vibrational excitations in the core-hole excited state (Franck-Condon principle). However, we could observe that the high intensity of the feature centered at 406 eV was due to the protonation site, and we discussed how certain changes in the spectral features, especially the shape and position of the resonances, normally attributed to $N\ 1s \rightarrow \pi^*(\text{CONH})$ and $N_i\ 1s \rightarrow \pi^*(\text{CON}_j\text{H})$, can be attributed to the different basic residues in the peptides, e.g. by additional transitions to $\pi^*(\text{C}=\text{N})$ and $\sigma^*(\text{NH})$ in the side chain of histidine and arginine. We could also identify that the production of immonium ions of a residue is quenched when the absorption takes place on its side. This coincides with the increased yield of side-chain-related fragments in the same condition, which suggests that fast bond cleavage in the side-chain is induced after localized photon absorption prior to energy redistribution in the system. In addition, we are able to answer an open question by Dörner et al. [86] concerning the presence of a peak at 399.2 eV in protonated methionine enkephalin $[\text{YGGFM}+\text{H}]^+$ but not in leucine enkephalin $[\text{YGGFL}+\text{H}]^+$. The authors proposed, based on literature and calculation, that this peak could be attributed to either an $N\ 1s \rightarrow \sigma^*(\text{NH})$ transition in an unprotonated terminal NH_2 group or an $N\ 1s \rightarrow \pi^*(\text{C}=\text{N})$ transition in a peptide bond in an enol form. In our set of experiments, peptides G_4H , G_4R , and G_4K have an unprotonated terminal NH_2 group but do not show a prominent peak at 399.2 eV. The same is true for the calculations of O-protonated triglycine, which do not show a transition at this energy. The assignment of this band to $N\ 1s \rightarrow \sigma^*(\text{NH})$ can thus be excluded.

Moreover, we have demonstrated that the NEXAMS technique is sensitive to backbone protonation, as demonstrated by studying protonated triglycine at the nitrogen K-edge and the oxygen K-edge. We have shown that a protonation site located at an amide oxygen leads to a resonance around 403 eV that is not present in the case of a peptide protonated at a nitrogen atom. However, the measurements did not reveal as striking differences between the protomers as expected from the work by Li et al. [195]. We have shown that the combination of DFT/ROCIS calculations and NEXAMS can help estimate the mixing ratio between the protomers observed during the experiment. This work should also be extended to the search of deprotonation site signatures in peptides and in other organic molecules where the information on the (de)protonation site(s) could be revealed using NEXAMS. Additionally, it could be tested experimentally if signatures of conformational structures such as α -helix and β -turn can be resolved in the $N_i 1s \rightarrow \pi^*(CON_jH)$ transitions as proposed theoretically by Buckley and Besley [91]. With this work, we can foresee that NEXAMS is perfectly suited to follow ultrafast proton transfer/migration in biomolecules, for example, in pump-probe studies at X-ray free-electron laser facilities, where a pump pulse would trigger the proton transfer process and a probe X-ray pulse would interrogate the position of that proton on the molecule. In addition, NEXAMS will allow us to follow the protonation/hydrogenation change in nanohydrated peptides and locate the water molecules on the biomolecules. Finally, due to the ionizing properties of soft X-ray radiation, NEXAMS experiments do not suffer from the limitations of dissociation-based spectroscopy techniques with respect to “stable” molecules, M^{n+} , that would not break but produce *a minima* intact photoionized precursor ions, $M^{(n+1)+\bullet}$, from which an absorption spectrum can be retrieved [73]. In these cases, NEXAMS is a valuable complementary technique to e.g. IRMPD.

7 | Summary

In summary, the first chapters described the different photoabsorption processes relevant to biomolecules over the electromagnetic spectrum ranging from IR to soft X-rays. We also introduced the computational approach (density functional theory), experimental mass spectrometers, and synchrotron beamlines used during this PhD thesis. A novel electrospray source dedicated to the production of hydrated biomolecules based on the design of Prof. Nonose has been presented [24]. During this thesis, this source was developed and implemented in our tandem mass spectrometer. The source can produce a very stable signal of biomolecular ions thanks to the use of nitrogen curtain gas. However, further tests are needed to produce hydrated biomolecules using this source. Moreover, we studied three biomolecular systems theoretically and experimentally to investigate the influence of the molecular environment on the electronic and three-dimensional structures, as well as on the fragmentation. In particular, the effects of a proton and nanosolvation with up to 20 water molecules were examined.

In Chapter 4, we have studied a deprotonated nucleotide, adenosine monophosphate (AMP^-), theoretically, using density functional theory. We have shown that isolated AMP^- has a linear structure and that already a single water molecule is sufficient to induce folding of AMP^- . We observed that only a few water molecules make hydrogen bonds with the nucleotide while most of the water molecules bind to one another to create a water droplet of increasing size with the hydration level. Adding water molecules around the phosphate group of the nucleotide and making hydrogen bonds with the adenine group also influence the spectroscopic properties of the system, inducing observable changes in the electronic and vibrational transitions of the molecules. These changes have been identified for hydrated AMP^- in the computed IR spectra and X-ray absorption spectra (XAS) at the oxygen K-edge. The calculated IR spectrum of isolated AMP^- showed excellent agreement with the experimental IRMPD spectrum of Nei et al. [50]. Frequency shifts in the fingerprint region can be observed between the different levels of hydration of AMP^- . In the 2800 - 3600 cm^{-1} region, the band corresponding to the -OH stretches of the hydroxyl groups of the ribose is

blueshifted in hydrated AMP⁻ compared to isolated AMP⁻, whereas the band arising from the OH stretches within the water droplet becomes larger with the hydration level. In addition, the water molecules make a clear contribution to the oxygen K-edge X-ray spectra of the different hydration levels of AMP⁻. When less than seven water molecules are attached, a splitting of the first resonance at 533 eV (due to transitions to $\sigma^*(\text{OH})$ from the phosphate and ribose groups) is observed, along with a low-intensity resonance at 532.5 eV from $\sigma^*(\text{OH})$ transitions in the water molecules. At higher solvation levels, the increased number of low-intensity transitions from the water molecules is responsible for a broad feature centered at 535 eV.

The differences in the vibrational and electronic spectra that have been observed theoretically now need to be confirmed by further experimental investigations. Such experiments could also serve as a benchmark for testing our computational method. As the water droplet is involved in a hydrogen bond network not only with the phosphate group but also with the adenine group, we could complete our study with the investigation of the effect of stepwise hydration on the X-ray spectra at the phosphorus and nitrogen K-edges. Moreover, this work could be extended to larger systems such as adenosine triphosphate, which possesses three phosphate groups and is, therefore, likely to allow more HBs between the biomolecule and water molecules. Studying the hydration of other nucleotides, such as thymidine monophosphate, cytidine monophosphate, or guanosine monophosphate, would allow us to understand the influence of the nucleobase on the hydration of nucleotides.

In Chapter 5, we combined experiments and theoretical calculations to investigate the influence of a single water molecule on the structure and fragmentation of a phosphorylated amino acid, phosphotyrosine (pTyr). We obtained the X-ray absorption spectra at the carbon and oxygen K-edges of a mixture of isolated and singly hydrated protonated phosphotyrosine. From the mass spectra, we could already highlight that the water molecule binds to the phosphate group, but not only. At the carbon K-edge, we have shown that the total ion yield spectrum resemble the XAS spectrum of neutral Tyr [177]. While no spectral fingerprint of the water molecule was observed at the C K-edge, at the O K-edge, an additional resonance at 536.5 eV appeared in the partial ion yield of the hydrated fragment $[\text{PO}_3\text{H}_2 + \text{H}_2\text{O}]^+$ compared to a fragment coming from bare pTyr. We attributed this resonance to a transition from the water molecule. The comparison of the NEXAMS spectrum at the carbon K-edge with the structural search of singly hydrated pTyr at the DFT level showed three possible structural candidates for singly hydrated pTyr: the water molecule binds to the phosphate group only, or the water molecule forms a bridge between the phosphate group and either

the NH_3^+ or the carboxyl group. The comparison between the O K-edge NEXAMS spectrum and the calculated XAS for the three structural candidates suggests that the water molecule binds to the carboxylic and the phosphate group. Thus, we were able to demonstrate that the combination of DFT/ROCIS calculation and soft X-ray spectroscopy allows for the determination of the location of the water molecule.

To go further, we could first optimize our experimental conditions to measure separately the X-ray absorption spectra of isolated pTyr and singly hydrated pTyr. This would enable us to better disentangle the contribution of both species to the NEXAMS spectra and obtain more information on the influence of the water molecule on the X-ray spectrum at the oxygen K-edge. Measuring the X-ray spectrum of singly hydrated pTyr at the nitrogen K-edge could tell us more unambiguously if the structure having the water bound to the protonated amino group can be completely discarded from the possible structures present in our experimental setup, as we could expect the water to influence the electronic transitions at this energy range. Photoelectron photoion coincidence (PEPICO) or PEPIPICO measurements would be required to gain more insight into the fragmentation processes of singly hydrated phosphotyrosine. This would allow us to investigate specifically the processes leading to hydrated fragments, explore the energy transfers after irradiation at the oxygen K-edge, and possibly demonstrate intermolecular Coulombic decay (ICD) in hydrated biomolecules as suggested by Milosavljević et al. [17]. Furthermore, we could investigate a higher solvation state to observe whether and how the fragmentation pattern and electronic structure change with an increasing number of water molecules. This would allow us to observe experimentally what we observed in Chapter 4, that the water transitions become dominant for high solvation states. Last, studying the deprotonated species would allow us to investigate the effect of the (de)protonation site on the structure and X-ray spectra of singly hydrated pTyr.

In Chapter 6, the influence of the proton localization on the electronic structure of model protonated peptides has been investigated. We have combined theoretical calculations, using replica exchange molecular dynamics and DFT, with NEXAMS at the nitrogen K-edge to investigate the structures of protonated model peptides and the electronic fingerprint of the protonated site. Our study initially focused on the influence of the proton when located at nitrogen atoms by analyzing the N K-edge NEXAMS spectrum of glycine-based pentapeptides. Subsequently, we examined the electronic spectra of protonated triglycine, which has been shown to be protonated either on the N-terminal nitrogen or on a backbone amide oxygen atom [194], [195]. The protonated pentapeptides all present a strong resonance at 401.6 eV related to

electronic transitions in the peptide bonds. Moreover, the signature of the protonation site for all peptides except G₄H is located in the broad feature between 404 and 408 eV. This feature comes essentially from σ^* resonances from the peptide's backbone with a major contribution from the $\sigma^*(\text{NH})$ resonances located at the protonated nitrogen site either from the N-terminus or the side chain of lysine and arginine. For the histidine-containing peptide G₄H, the signature of the protonation site is located at 403.1 eV through a N 1s $\rightarrow \sigma^*(\text{NH})$ transition.

For protonated triglycine, we have demonstrated that two protomers coexist in our experimental conditions: a linear structure with the protonation at the amide oxygen and a folded structure with the protonation at the N-terminus. We were able to distinguish the signature of both structures in the X-ray absorption spectrum at the nitrogen K-edge through the resonance at 403.1 eV (N 1s $\rightarrow \sigma^*(\text{NH})$ transition from an unprotonated N-terminus), which is only present in the O-protonated conformer, and the broad resonance centered around 405.5 eV for the N 1s $\rightarrow \sigma^*(\text{NH})$ transition from a protonated N-terminus. In combination with DFT/ROCIS calculations, we have shown that the population in our experimental conditions is represented by a relative population of 30% of N-protonated triglycine with 70% of O-protonated triglycine.

A direct follow-up to this work is the study of the influence of deprotonation sites in peptides on the electronic structure. Mapping the protonation and deprotonation sites of model peptides would allow easier experimental attribution of the different resonances in bigger systems, such as small proteins. Bigger systems are often multiply charged. Thus, we could investigate multiply charged peptides with fixed and mobile protons using NEXAMS. Moreover, we could also study the influence of solvation on the protonation site of these model peptides.

In summary, this work showed that NEXAMS is a technique suited to study the structure and the fragmentation processes of hydrated biomolecules and can be used to locate the protonation site in systems at least as big as pentapeptides. We have shown that one water molecule is already enough to change the structure of a phosphorylated amino acid and of a small nucleotide. In the future, we will use the newly built electrospray source to systematically explore the effect of the first water shell around biomolecules on their structure and fragmentation using NEXAMS. We could also implement a cooling system for our Paul trap to reduce the conformer population and minimize the evaporation of water molecules as the hydrated sample enters the trap. Incorporating an ion mobility spectrometer (IMS) before the Paul trap would also allow for selection of specific ion conformations. We also foresee that this method can be used to follow ultra-fast proton transfer/migration in biomolecules, for example,

in pump-probe studies at X-ray free-electron laser facilities. In such experiments, a pump pulse in the IR or UV range would trigger the proton transfer process, while a probe X-ray pulse would interrogate the position of that proton on the molecule.

Amino acids and the sugars in DNA and RNA are chiral molecules, with the amino acids being left-handed and the sugars being right-handed. Circular dichroism (CD) studies investigate the differential absorption of circularly polarized radiation by a chiral material, which can be used to reveal the structure of amino acids. However, the CD signal from circularly polarized UV photons is quite low (less than 4.5%); thus, attribution of CD signal from electrospray ions is challenging. In the soft X-ray regime, a chiral asymmetry of 20% has been detected in the C 1s core level photoemission of Fenchone enantiomers [227]. By combining mass spectrometry with X-ray circularly polarized light from synchrotron radiation, we could extend the possibility of structural determination of NEXAMS.

8 | Samenvatting

Nanohydratatie van biomoleculen in de gas fase: De invloed op de structuur en fragmentatie na fotoabsorptie

In de eerste hoofdstukken zijn de verschillende fotoabsorptieprocessen beschreven die relevant zijn voor biomoleculen, waarin we een ruim bereik in het elektromagnetische spectrum behandelen, van infrarood tot röntgenstraling. We hebben ook de simulatie methode geïntroduceerd (dichtheidsfunctionaaltheorie, DFT), alsmede de massaspectrometers en synchrotronbundellijnen die tot de resultaten in de vorm van de NEXAMS spectra hebben geleid. Er is een nieuwe electrospray ionisatiebron gepresenteerd voor de productie van gehydrateerde biomoleculen, gebaseerd op het ontwerp van Prof. Nonose [24]. Voor dit proefschrift werd deze bron ontwikkeld en succesvol geïmplementeerd in onze tandem massaspectrometer. De bron kan een zeer stabiel signaal van biomoleculaire ionen produceren, dankzij het gebruik van stikstof als schermgas. Verdere tests zijn echter nodig om gehydrateerde biomoleculen te produceren met deze bron. Naast de ontwikkeling en bouw van de ionenbron hebben we drie biomoleculaire systemen theoretisch en experimenteel bestudeerd om de invloed van de moleculaire omgeving op zowel de fragmentatie als de elektronische en driedimensionale structuren te onderzoeken. In het bijzonder hebben we de effecten van een proton's positie en nanosolvatie met maximaal 20 watermoleculen onderzocht.

In hoofdstuk 4 hebben we een gedeprotoneerde nucleotide, adenosinemonofosfaat (AMP^-), theoretisch bestudeerd met behulp van DFT. We hebben getoond dat het geïsoleerde AMP^- een lineaire structuur heeft en dat al één watermolecuul voldoende is om te zorgen voor een gevouwen structuur van AMP^- . We hebben waargenomen dat slechts enkele watermoleculen een waterstofbrug vormen met de nucleotide, terwijl de meeste watermoleculen zich aan elkaar binden om een waterdruppel te vormen die steeds groter wordt naarmate het hydratatieniveau toeneemt. Het toevoegen van watermoleculen rond de fosfaatgroep van de nucleotide en het maken van waterstofbruggen met de adeninegroep beïnvloeden ook de spectroscopische eigenschappen van

het systeem, waardoor waarneembare veranderingen optreden in de elektronische en vibrationele transitie van de moleculen. Deze veranderingen zijn geïdentificeerd voor gehydrateerd AMP⁻ in de berekende IR-spectra en röntgenabsorptiespectra rond de 1s ionisatie en excitatie van zuurstof. Het berekende IR-spectrum van geïsoleerd AMP⁻ toont een uitstekende overeenkomst met het experimentele IRMPD-spectrum van Nei et al.. Frequentieverschuivingen in de 'fingerprint'-regio kunnen worden waargenomen tussen de verschillende hydratationiveaus van AMP⁻. De spectrale band in het gebied van 2800 - 3600 cm⁻¹ vindt zijn oorsprong in de -OH-strekkingen van de hydroxylgroepen van de ribose. In gehydrateerd AMP⁻ is deze band naar een hogere energie verschoven (blue-shifted), vergeleken met geïsoleerd AMP⁻, terwijl de band die voortkomt uit OH-strekkingen binnen de waterdruppel groter wordt met het hydratationiveau. Bovendien leveren de watermoleculen een duidelijke bijdrage aan de zuurstof K-kant bij gehydrateerde AMP⁻. Wanneer er minder dan zeven watermoleculen zijn verbonden, wordt een splitsing van de eerste resonantie bij 533 eV geobserveerd (door transitie naar $\sigma^*(\text{OH})$ orbitalen van de fosfaat- en ribosegroepen), samen met een lage-intensiteit resonantie bij 532,5 eV van $\sigma^*(\text{OH})$ overgangen in de watermoleculen. Bij hogere solvatationiveaus is het toegenomen aantal overgangen met lage intensiteit van de watermoleculen verantwoordelijk voor een brede eigenschap gecentreerd bij 535 eV. De verschillen in de theoretisch voorspelde vibrationele- en elektronische transitie moeten nu bevestigd worden door verder experimenteel onderzoek. Zulke experimenten kunnen ook dienen als maatstaf voor het testen van onze rekenmethode. Aangezien de waterdruppel niet alleen betrokken is bij een waterstofbrugnetwerk met de fosfaatgroep, maar ook met de adeninegroep, zouden we onze studie kunnen voltooien met het onderzoeken van het effect van stapsgewijze hydratatie op de röntgenspectra aan de fosfor- en stikstof K-kant. Bovendien zou dit onderzoek kunnen worden uitgebreid naar grotere systemen zoals adenosinetrifosfaat, dat drie fosfaatgroepen bezit en daarom waarschijnlijk meer waterstofbruggen tussen het biomolecuul en de watermoleculen toestaat. Het bestuderen van de hydratatie van andere nucleotiden, zoals thymidinemonofosfaat, cytidinemonofosfaat of guanosinemonofosfaat, zou ons in staat stellen om de invloed van de nucleobase op de hydratatie van nucleotiden te begrijpen.

In hoofdstuk 5 hebben we experimenten en theoretische berekeningen gecombineerd om de invloed van een enkel watermolecuul op de structuur en fragmentatie van een gefosforyleerd aminozuur, fosfotyrosine (pTyr), te onderzoeken. We verkregen de röntgenabsorptiespectra aan de koolstof- en zuurstof K-kant van een mengsel van geïsoleerd en enkelvoudig gehydrateerd geprotoneerd fosfotyrosine. Uit de massaspectra konden we al opmaken dat het watermolecuul zich niet alleen bindt aan

de fosfaatgroep, maar ook aan andere functionele groepen in pTyr. Met behulp van de totale ionenopbrengst van NEXAMS spectra aan de K-kant van koolstof hebben we aangetoond dat de spectra veel overeenkomsten tonen met het XAS-spectrum van neutraal pTyr [177]. Hoewel er geen spectrale vingerafdruk van het watermolecuul werd waargenomen bij de C K-kant, verscheen er bij de O K-kant wel een extra resonantie bij 536,5 eV, geobserveerd in de ionenopbrengst van het gehydrateerde fragment $[\text{PO}_3\text{H}_2 + \text{H}_2\text{O}]^+$. We schreven deze resonantie toe aan een elektronische transitie van het watermolecuul. De vergelijking van het NEXAMS spectrum op de koolstof K-kant met het structuuronderzoek van enkelvoudig gehydrateerd pTyr op DFT niveau toonde drie mogelijke structurele kandidaten voor enkelvoudig gehydrateerd pTyr: (i) het watermolecuul bindt zich alleen aan de fosfaatgroep, of (ii) het watermolecuul vormt een brug tussen de fosfaatgroep en de NH_3^+ , of (iii) het vormt een brug tussen de fosfaatgroep en de carboxylgroep. De vergelijking tussen het O K-edge NEXAMS spectrum en de berekende XAS voor de drie structurele kandidaten suggereert dat het watermolecuul bindt aan de carboxylgroep en de fosfaatgroep. We hebben dus kunnen aantonen dat de combinatie van DFT/ROCIS berekening en röntgenspectroscopie het mogelijk maakt om de locatie van het watermolecuul te bepalen. Als volgende stap zouden we onze experimentele condities kunnen optimaliseren om de röntgenabsorptiespectra van geïsoleerd pTyr en gehydrateerd pTyr afzonderlijk te kunnen meten. Dit zou ons in staat stellen om de bijdrage van beide samples aan de NEXAMS-spectra aan de zuurstof K-kant beter te ontrafelen en meer informatie te verkrijgen over de invloed van het watermolecuul hierin. Het meten van het röntgenspectrum van enkelvoudig gehydrateerd pTyr bij de stikstof K-kant zou kunnen vaststellen of de structuur met het water gebonden aan de geprotoneerde aminogroep volledig kan worden uitgesloten van de mogelijke structuren die aanwezig zijn in onze experimentele opstelling, omdat we kunnen verwachten dat het water de elektronische overgangen in dit energiegebied beïnvloedt. Verder zouden we een hogere hydratatioestand kunnen onderzoeken om te observeren of en hoe het fragmentatiepatroon en de elektronische structuur veranderen bij een toenemend aantal watermoleculen. Dit zou ons in staat stellen om experimenteel te observeren wat we in hoofdstuk 4 hebben geobserveerd, namelijk dat de waterovergangen dominant worden voor hoge hydratatioestanden. Ten slotte zouden we het effect van de (de)protonatie locatie op de structuur en röntgenspectra van enkelvoudig gehydrateerd pTyr kunnen onderzoeken door de gedeprotoneerde samples te bestuderen.

In hoofdstuk 6 is de invloed van de protonlokalisatie op de elektronische structuur van geprotoneerde modelpeptiden onderzocht. We hebben theoretische berekeningen,

gebruikmakend van replica exchange moleculaire dynamica en DFT, gecombineerd met NEXAMS aan de stikstof K-kant om de structuren van geprotoneerde modelpeptiden en de elektronische vingerafdruk van de proton locatie te onderzoeken. Onze studie richtte zich in eerste instantie op de invloed van het proton, wanneer het zich op stikstofatomen bevindt, door NEXAMS spectra aan de N K-kant van glycine-gebaseerde pentapeptiden te analyseren. Vervolgens onderzochten we de elektronische spectra van geprotoneerd triglycine, waarvan is aangetoond dat het ofwel geprotoneerd is op de stikstof bij het amino-uiteinde, of op een amide-zuurstofatoom in de peptidebindingen[194], [195].

De geprotoneerde pentapeptiden vertonen allemaal een sterke resonantie bij 401,6 eV die gerelateerd is aan elektronische overgangen in de peptidebindingen. Bovendien is de vingerafdruk van de protoneringslocatie voor alle peptiden behalve G₄H te vinden in de brede resonantie tussen 404 en 408 eV. Deze eigenschap is voornamelijk afkomstig van σ^* resonanties van de bindingsketen van het peptide met een belangrijke bijdrage van de $\sigma^*(\text{NH})$ resonanties gelokaliseerd op de geprotoneerde stikstof van de N-terminus of de zijketen van lysine en arginine. In het histidine-bevattende peptide G₄H bevindt zich de signatuur van de protoneringslocatie zich bij 403,1 eV door een N 1s \rightarrow $\sigma^*(\text{NH})$ transitie.

In geprotoneerd triglycine hebben we aangetoond dat er twee protomeren naast elkaar bestaan onder onze experimentele condities: een lineaire structuur met de proton gelokaliseerd aan de amide zuurstof, en een gevouwen structuur met de protonering aan de N-terminus. We konden beide structuren onderscheiden in het röntgenabsorptiespectrum bij de stikstof K-kant dankzij de resonantie bij 403,1 eV (N 1s \rightarrow $\sigma^*(\text{NH})$ transitie van een ongeprotoneerde N-terminus), die alleen aanwezig is in de O-geprotoneerde conformatie, en de brede resonantie gecentreerd rond 405,5 eV voor de N 1s \rightarrow $\sigma^*(\text{NH})$ overgang van een geprotoneerde N-terminus. In combinatie met DFT/ROCIS-berekeningen hebben we aangetoond dat de populatie in onze experimentele omstandigheden wordt vertegenwoordigd door een relatieve populatie van 30% N-geprotoneerd triglycine en 70% O-geprotoneerd triglycine. Een direct vervolg op dit onderzoek is de studie van de invloed van deprotoneringslocaties in peptiden op de elektronische structuur. Het in kaart brengen van de protonatie- en deprotonatielocaties van modelpeptiden zou de experimentele toewijzing van de verschillende resonanties in grotere systemen, zoals kleine eiwitten, vergemakkelijken. Grotere systemen zijn vaak meervoudig geladen. Zo zouden we met NEXAMS meervoudig geladen peptiden met vaste en mobiele protonen kunnen onderzoeken. Bovendien zouden we de invloed van solvatie op de protonatieplaats van deze modelpeptiden kunnen bestuderen.

Samenvattend toonde dit werk aan dat NEXAMS een techniek is die geschikt is om de structuur en de fragmentatieprocessen van gehydrateerde biomoleculen te bestuderen en gebruikt kan worden om de proton te lokaliseren in systemen die minstens zo groot zijn als pentapeptiden. We hebben aangetoond dat één watermolecuul al genoeg is om de structuur van een gefosforyleerd aminozuur en van een kleine nucleotide te veranderen. In de toekomst zullen we de nieuw gebouwde electrospray bron gebruiken om systematisch het effect van de eerste waterschil rond biomoleculen op hun structuur en fragmentatie te onderzoeken met behulp van NEXAMS. We kunnen ook een koelstelsel implementeren voor onze Paul-val om de conformatie populatie te reduceren en de verdamping van de watermoleculen te beperken wanneer het gehydrateerde sample de ionenval binnenkomt. Een ionenmobiliteitsspectrometer (IMS) zou ook kunnen worden geïmplementeerd vóór de Paul-val om specifieke ionconformaties te selecteren. We verwachten ook dat deze methode kan worden gebruikt om ultrasnelle protonoverdracht/migratie in biomoleculen te volgen, bijvoorbeeld in pump-probe studies bij röntgen-vrij-elektronenlaserfaciliteiten. In dergelijke experimenten zou een pomppuls in het IR- of UV-bereik het protonoverdrachtsproces in gang zetten, waarna een röntgenstralingpuls de positie van dat proton op het molecuul zou kunnen onthullen. Aminozuren en suikers in DNA en RNA zijn chirale moleculen, waarbij de aminozuren linkshandig zijn en de suikers rechtshandig. Circulair dichroïsme (CD) onderzoekt de differentiële absorptie van circulair gepolariseerde straling door een chiraal materiaal, wat gebruikt kan worden om de structuur van aminozuren te onthullen. Het CD-sigitaal van circulair gepolariseerde UV-fotonen is echter vrij laag (minder dan 4,5%); het is dus een uitdaging om CD-signalen van electrospray-ionen toe te schrijven. In het röntgenbereik is een chirale asymmetrie van 20% gedetecteerd in de foton emissie door middel van C 1s ionisatie van Fenchone enantiomeren [227]. Door massaspectrometrie te combineren met circulair gepolariseerd röntgenlicht van synchrotronstraling zouden we de mogelijkheden voor structuurbepaling van NEXAMS kunnen uitbreiden.

9 | Résumé en français

Influence de la nanohydratation sur la structure et la fragmentation radio-induite de molécules d'intérêt biologique en phase gazeuse

9.1 Introduction

L'eau est essentielle à la vie. La présence d'eau sur Terre a fourni un environnement approprié pour la forme initiale de la vie [1], [2]. Il est admis que les molécules prébiotiques se sont formées dans des conditions physico-chimiques impliquant l'interaction de l'eau avec des atomes de carbone. En tant que solvant omniprésent, l'eau possède les propriétés essentielles pour solubiliser les ions et établir un environnement polyvalent pour les réactions chimiques fondamentales, telles que la formation d'acides aminés et de peptides [3]. Le rôle des effets de solvatation sur la fragmentation des biomolécules après ionisation ou excitation a été étudié dans des milieux denses [11]–[13] et de microsolvatation [14]–[20]. Cependant, la question de savoir si l'eau augmente ou réduit les dommages causés par les radiations sur les biomolécules est toujours en cours de débat. L'étude de l'interaction de l'eau avec les biomolécules au niveau moléculaire reste difficile en phase liquide en raison de la difficulté à identifier l'effet du solvant des propriétés intrinsèques des biomolécules. Pour étudier les premières étapes de la solvatation, la méthode de choix consiste à réaliser des expériences en phase gazeuse, car elle permet d'éliminer toute interaction entre la biomolécule et le solvant. À partir de ce point de départ, une approche ascendante peut être adoptée en étudiant d'abord le système isolé, puis en ajoutant progressivement des molécules d'eau de manière progressive jusqu'à ce que l'on s'approche finalement des propriétés de la phase liquide. Le développement par John Fenn [21] de la source d'ionisation par électrospray (électrospray) pour introduire des molécules thermiquement fragiles dans la phase gazeuse a permis de concevoir des études expérimentales au niveau moléculaire portant sur des biomolécules telles que les protéines et les brins d'ADN.

Combinée à la spectrométrie de masse, cette technique constitue un outil de choix pour étudier la solvation progressive des biomolécules. Cette approche permet l'étude initiale de systèmes isolés, suivie de l'étude progressive du processus de solvation, ce qui permet d'élucider l'influence du solvant sur la conformation et les propriétés intrinsèques des biomolécules. Afin de mieux comprendre l'effet de l'eau sur la conformation, la réactivité et les processus photosensibles des biomolécules, des recherches approfondies ont été consacrées à l'étude théorique et expérimentale de la solvation par étapes au cours des dernières décennies [7], [15], [17], [22]–[25]. Les méthodes théoriques permettent d'explorer l'espace conformationnel des molécules hydratées pour différentes tailles et niveaux d'hydratation [7], [26]. Par conséquent, la méthode théorique peut être utilisée pour fournir une interprétation physique des observations expérimentales, tandis que les expériences peuvent servir de référence pour tester la pertinence des traitements théoriques des systèmes hydratés. Zhan et Fenn ont développé une source d'électrospray dédiée à la production d'agrégats hydratés [15], [27] et ont montré la possibilité de produire des biomolécules hydratées avec jusqu'à 40 molécules d'eau ou plus, ce qui permet d'étudier la solvation progressive des biomolécules en phase gazeuse.

De plus, la combinaison de la spectrométrie de masse avec des sources de lumière avancées, telles que les lasers de table, les synchrotrons ou les lasers à électrons libres, nous permet de sonder l'interaction du rayonnement électromagnétique avec une cible biomoléculaire et, par conséquent, de comprendre les différentes propriétés des biomolécules isolées ou solvatées. Le rayonnement électromagnétique est une sonde polyvalente qui permet d'élucider la structure et les dommages causés par les rayonnements dans les systèmes moléculaires. L'énergie des photons absorbés détermine les transitions induites dans un système moléculaire, allant des excitations rotationnelles et vibrationnelles aux excitations électroniques. Par conséquent, diverses voies de relaxation deviennent observables. Une énergie suffisante peut conduire à l'ionisation et/ou à la dissociation moléculaire. Alors que la lumière infrarouge (IR) induit des transitions vibrationnelles au sein de la molécule, la photoexcitation résonante avec la lumière ultraviolette (UV) entraîne des transitions électroniques, promouvant les électrons de valence vers des orbitales moléculaires inoccupées. En outre, le processus de photo-absorption dans la gamme des rayons X mous nous permet de sonder sélectivement des atomes précis au sein de la biomolécule ou de son environnement en excitant les électrons de cœur.

Dans ce contexte, la mesure et l'analyse des spectres d'absorption de la molécule à différentes énergies de photons fournissent de nombreuses informations concernant non seulement sa structure géométrique, mais aussi sa structure électronique. Toute-

fois, pour les échantillons gazeux, l'absorption directe est rendue difficile par la faible densité des biomolécules, ce qui complique les mesures directes de l'absorption de la lumière dans l'échantillon. Une approche pour résoudre ce problème consiste donc à effectuer une spectroscopie dite d'action, qui consiste à mesurer comment le photon absorbé affecte la molécule, souvent en explorant les voies de fragmentation à l'aide de la spectrométrie de masse. Cette méthode a été mise en œuvre dans de nombreux travaux pour étudier les systèmes isolés mais aussi hydratés, en combinant la source électrospray et la spectrométrie de masse avec des sources de photons [25], [28]–[30].

Le travail de cette thèse s'inscrit dans ce contexte et se concentre sur la combinaison d'approches théoriques et expérimentales pour étudier l'influence de l'environnement aqueux sur la structure et la fragmentation des biomolécules en phase gazeuse. Pour étudier les biomolécules solvatées en phase gazeuse, nous avons développé une source d'hydratation dédiée, basée sur la source du Prof. Nonose [24] et qui est conçue sur le même principe que celle développée par Zhan et Fenn [15]. Des calculs théoriques ont été effectués sur un nucléotide déprotoné pour comprendre comment la solvation progressive impliquant jusqu'à vingt molécules d'eau affecte la structure tridimensionnelle et les spectres d'absorption. De plus, la combinaison des calculs et des expériences de spectroscopie d'action dans le domaine des rayons X mous nous a permis de révéler la structure 3D et la structure électronique, ainsi que d'élucider l'influence d'une seule molécule d'eau ou d'un proton sur les propriétés d'un acide aminé protoné et de peptides protonés.

9.2 Contexte

9.2.1 Molécules d'intérêt biologique

Les biomolécules sont des composants fondamentaux de tout organisme vivant, essentiels à divers processus biologiques tels que la division cellulaire ou le développement biologique. Elles régulent toutes les réactions chimiques nécessaires à la croissance et à la survie d'un organisme. Les acides nucléiques, tels que l'acide désoxyribonucléique (ADN) et l'acide ribonucléique (ARN), jouent un rôle crucial dans le stockage et la transmission de l'information génétique. Ils codent la séquence de nucléotides qui dicte l'arrangement des acides aminés dans les protéines essentielles à la vie. La structure et les fonctions des protéines sont définies par la séquence des acides aminés. Par exemple, des protéines telles que l'élastine ou la kératine agissent comme des com-

posants structurels des cellules et des organismes complexes, et l'insuline agit comme une hormone. Les protéines sont généralement définies comme des chaînes de 50 acides aminés ou plus, tandis que les chaînes plus courtes sont appelées peptides. Les acides aminés sont les éléments constitutifs des peptides. Ils se composent d'un groupe amine (NH_2), d'un groupe carboxyle (COOH) et d'une chaîne latérale, tous reliés au carbone α central ($\text{C}\alpha$). La Figure 9.2.1 montre les acides aminés qui composent la séquence des peptides ainsi que des deux autres molécules étudiées dans le cadre de cette thèse. La chaîne latérale est ce qui diffère d'un acide aminé à l'autre. Par exemple, pour la tyrosine, la chaîne latérale est un groupe phénol de formule C_6H_5 , alors que pour l'acide aminé le plus simple, la glycine, la chaîne latérale n'est constituée que d'un atome d'hydrogène.

Lors de la synthèse des peptides, le groupe amine d'un acide aminé réagit avec le groupe carboxyle d'un autre acide aminé. Il se crée une liaison peptidique sous l'effet du dégagement de H_2O . Un peptide est formé lorsque plusieurs acides α -aminés réagissent de cette façon. Pour interpréter les résultats des expériences de fragmentation des peptides, un système de nomenclature est utilisé pour nommer et identifier les fragments provenant des différents sites de fragmentation du squelette peptidique. Ce système a été introduit pour la première fois par Roepstorff et Fohlman [36], puis modifié par Biemann [37]. Les fragments d'ions peptidiques sont indiqués par a, b ou c si la charge est retenue du côté du N-terminal et par x, y ou z si la charge est située du côté du C-terminal. Un indice indique où se situe la rupture de liaison dans la chaîne. Les doubles scissions du squelette génèrent des fragments internes. Les fragments internes de formule $\text{RCH}=\text{NH}_2^+$, provenant d'une seule chaîne latérale par une combinaison de clivages de types a et y, sont appelés ions immonium.

Les nucléotides sont les unités monomériques des acides nucléiques, y compris l'acide ribonucléique (ARN) et l'acide désoxyribonucléique (ADN). Un nucléotide comprend un groupe phosphate, un groupe sucre et une base azotée. Les bases présentes dans l'ADN sont l'adénine (A), la cytosine (C), la guanine (G) et la thymine (T). Dans l'ARN, l'uracile (U) remplace la thymine. La principale différence entre les molécules d'ADN et d'ARN réside dans l'ordre de leurs éléments constitutifs (nucléotides) et dans leurs différences structurelles. L'ARN est composé de groupes de sucre ribose avec des paires A-U, tandis que l'ADN est composé de groupes de sucre désoxyribose avec des paires A-T. L'ARN est composé d'un seul brin de nucléotides, tandis que l'ADN est composé de deux brins, formant ainsi une structure hélicoïdale.

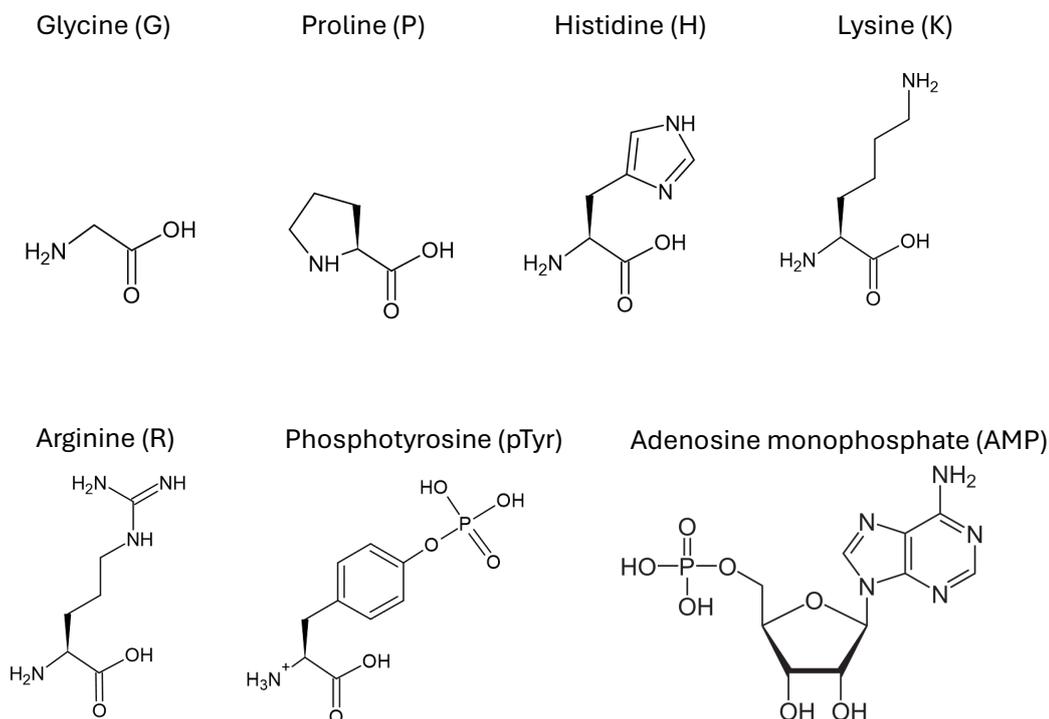


Figure 9.2.1: Structures des acides aminés composant les pentapeptides, ainsi que la structure de la phosphotyrosine protonée étudiée dans le chapitre 6 et de l'adenosine monophosphate étudiée dans le chapitre 4.

9.2.2 Spectroscopie d'absorption induite par rayons X

La spectroscopie d'absorption des rayons X (XAS) a été largement utilisée au cours des dernières décennies pour étudier les propriétés électroniques et structurales des systèmes biomoléculaires. La spectroscopie d'absorption des rayons X a été utilisée pour la première fois dans les années 1920 pour l'étude structurale de la matière, où de fines structures proches des seuils d'absorption ont été observées. La structure fine d'absorption des rayons X (NEXAFS) étudie la manière dont les photons de rayons X sont absorbés en excitant les électrons de cœur dans des états liés ou continus inoccupés. La spectroscopie NEXAFS présente des caractéristiques importantes et uniques dans la gamme d'énergie légèrement inférieure à l'énergie d'ionisation et s'étendant jusqu'à environ 50 eV au dessus de ce seuil [75]. Les principales techniques expérimentales pour la réalisation de spectres NEXAFS d'échantillons à l'état solide impliquent la mesure du taux d'électrons et du taux de fluorescence. Ces deux mesures

renseignent directement sur la formation d'une lacune en couche profonde du cortège électronique après absorption des rayons X, ce qui conduit à la détermination de la section efficace d'absorption [76]. De plus, la détection du taux d'ions peut être utilisée comme alternative au taux d'électrons. Bien qu'il s'agisse d'une mesure indirecte de l'absorption des rayons X, l'enregistrement du taux total des ions en fonction de l'énergie des photons permet d'obtenir une approximation du spectre d'absorption total [75].

Il y a plus de vingt ans, des travaux pionniers ont utilisé la spectroscopie NEXAFS pour analyser des films minces d'acides aminés et de petits peptides en phase solide [77]. Cette technique a ensuite été utilisée pour mesurer les spectres NEXAFS au seuil K du carbone, de l'azote et de l'oxygène des 20 acides aminés protéinogéniques courants, ainsi que des oligopeptides simples de glycine et de divers homopolypeptides [78]–[80]. Les études ont révélé que les signaux forts dans les spectres d'acides aminés proviennent de transitions électroniques impliquant les orbitales π^* du groupe carboxyle du C-terminal, les systèmes de cycles aromatiques dans les chaînes latérales aromatiques et les groupes carboxyles des chaînes latérales d'acides aminés. Il est intéressant de noter que dans le cas des peptides, des caractéristiques spectrales apparaissent aux seuils K du carbone, de l'azote et de l'oxygène en raison des transitions électroniques des orbitales centrales 1s vers les orbitales π^* de la liaison peptidique [81], [82]. Pour les bases d'acides nucléiques et les nucléotides, les transitions $1s \rightarrow \pi^*$ sont particulièrement sensibles à l'environnement chimique de l'azote, la substitution des atomes de carbone par des atomes d'oxygène du cycle entraînant un décalage vers le bleu des transitions $1s \rightarrow \pi^*$ les plus basses en énergie [83].

L'étude des biomolécules en phase gazeuse offre l'avantage d'étudier leurs propriétés intrinsèques en l'absence de toute influence d'un environnement chimique. En combinant la NEXAFS et la spectrométrie de masse, il est possible d'étudier des biomolécules chargées en phase gazeuse et d'enregistrer le taux de dissociation et le taux d'ionisation en fonction de l'énergie des photons. Cette technique est également appelée spectrométrie de masse par absorption de rayons X (NEXAMS) [84]–[86]. Le NEXAMS a d'abord été utilisé pour élucider les processus de dissociation dans les biomolécules, y compris l'ionisation non dissociative, la scission du squelette ou la dissociation sélective en site [85], [87]–[89], ou pour comprendre la structure des molécules chargées et l'influence de l'environnement sur leur structure [86], [90]–[93].

9.3 Méthodes

9.3.1 Calculs de Chimie Quantique

Motivations

Les calculs de Chimie Quantique peuvent fournir des informations sur les structures géométriques et électroniques d'un système moléculaire en lien avec ses propriétés physicochimiques. Ils peuvent être combinés à des expériences (par exemple de mesure de spectre ou de propriétés thermodynamiques), ce qui permet d'interpréter le résultat de ces expériences et d'obtenir des informations complémentaires sur le système étudié. Les propriétés physico-chimiques, les structures moléculaires, la réactivité chimique ou les spectres vibrationnels peuvent être mieux appréhendés en explorant les modifications de la surface d'énergie potentielle (PES) du système suite à des variations de ses coordonnées nucléaires. Or, selon le système (taille, composition ...), la PES peut être plus ou moins complexe. Pour les petites molécules, la PES présente généralement un minimum global, correspondant à l'énergie potentielle la plus basse associée à la géométrie d'équilibre. Pour les systèmes de taille moyenne et de grande taille, la PES tend à comporter plusieurs minima locaux, ce qui implique l'existence de plusieurs géométries relativement proches en énergie. En représentant la PES comme l'énergie potentielle d'un oscillateur harmonique ou anharmonique, la forme de la vallée au niveau du puits détermine le spectre vibratoire. L'exploration de la PES et l'identification de ses points stationnaires (minimum, points de contact, ...) sont alors cruciales.

D'un point de vue théorique, les minima énergétiques globaux et locaux d'un système moléculaire donné peuvent être obtenus à l'aide de la dynamique moléculaire ou de calculs de structure électronique. Ces dernières méthodes dites *ab initio* sont basées sur la résolution de l'équation de Schrödinger dans l'approximation de Born-Oppenheimer. Selon le type de propriétés étudiées et la taille du système, il existe différentes approches : celles basées sur le calcul explicite de la fonction d'onde comme les méthodes Hartree-Fock (HF), Møller-Plesset (appelées MPn et basées sur la théorie de la perturbation à l'ordre n), les méthodes de clusters couplés (CC) et celles basées sur le calcul de la densité électronique comme la théorie de la fonctionnelle de la densité (DFT). Dans cette thèse, les géométries des systèmes pris dans leur état fondamental ont été obtenues à l'aide de calculs DFT avec la fonctionnelle hybride B3LYP.

Une fois que les géométries candidates ont été trouvées, on peut calculer leurs spectres vibrationnels (IR) et/ou leurs spectres d'absorption en rayon X. Dans cette thèse les spectres infrarouge (IR) ont été obtenus au niveau DFT-B3LYP et dans le cadre

du modèle de l'oscillation harmonique. Pour un système donné, la modélisation de son spectre X nécessite la prise en compte de son état fondamental mais également de ses états excités. Dans la littérature plusieurs approches théoriques sont envisageables comme la DFT dépendante du temps (TDDFT) ou une méthode hybride combinant la DFT avec l'interaction de configuration ouverte restreinte aux simples excitations (ROCIS) [130]. C'est cette dernière approche qui a été utilisée pour générer les spectres X des systèmes envisagés.

Optimisation de la géométrie d'un système

D'un point de vue numérique, l'optimisation de la géométrie d'un système nécessite de calculer les forces appliquées sur chacun des atomes qui le composent. Cette étape nécessite donc de calculer les gradients de l'énergie potentielle du système vis à vis de toutes les coordonnées nucléaires. Selon la description de l'énergie potentielle, cette étape peut être fastidieuse. Par conséquent, l'obtention de la géométrie d'équilibre d'un système complexe comme ceux envisagés dans le cadre de cette thèse s'effectue en deux étapes. La première consiste à explorer la PES du système à l'aide de calculs dits de "champ de force" afin de générer des structures "stables". La seconde étape consiste à réoptimiser ces structures à l'aide de calculs *ab initio* comme la DFT.

Les modèles de champ de force supposent que l'énergie potentielle est la somme de potentiels paramétrés (termes de liaison covalente et non covalente) dans lesquels la structure électronique n'est pas décrite explicitement puisque chaque atome est traité comme une charge ponctuelle. Ainsi, les méthodes de champ de force réduisent considérablement le coût de calcul tout en limitant la précision de la description du système. Cette réduction permet de modéliser des systèmes comportant plusieurs milliers d'atomes, ce qui est typiquement irréalisable avec des méthodes *ab initio* en raison de leur coût de calcul.

Les minima énergétiques de la surface de potentiel décrite en champ de force (ou fournissent) des structures moléculaires énergétiquement stables, qui seront ensuite réoptimisées en tenant compte explicitement de la structure électronique du système. Dans cette thèse, l'approche DFT-B3LYP a été retenue car la DFT permet un rapport coût/précision très favorable et parce que les gradients analytiques sont disponibles pour la plupart des fonctionnelles comme la B3LYP [123]. De plus, les optimisations hybrides de la DFT peuvent être accélérées par l'utilisation de l'approximation RIJCOSX [128] implémentée dans le logiciel ORCA [127]. Dans cette thèse, les optimisations de la géométrie sont effectuées au niveau DFT-B3LYP dans l'approximation RIJCOSX et à l'aide de la base TZVP [157].

Spectre vibrationnel infrarouge

Une fois qu'une structure stable est trouvée, nous pouvons calculer son spectre vibrationnel. Les modes normaux moléculaires sont, en général, calculés dans le cadre du modèle de l'oscillateur harmonique. Si la molécule est dans sa conformation d'équilibre, elle correspond au point le plus bas (au moins localement) de la PES. La section transversale du profil PES près de ce point peut alors être supposée comme étant approximativement parabolique, de sorte que la dérivée seconde de l'énergie par rapport à une coordonnée nucléaire peut être interprétée comme une constante de force pour l'oscillation harmonique d'un atome le long de cette coordonnée. Étant donné que les vibrations moléculaires dans les molécules polyatomiques impliquent le déplacement simultané de plusieurs atomes, ce modèle d'oscillateur harmonique peut être généralisé à plusieurs coordonnées nucléaires. Ainsi les modes normaux et les fréquences associées deviennent alors des vecteurs propres et des valeurs propres d'une matrice de constante de force.

Par conséquent, pour obtenir le spectre vibrationnel d'un système moléculaire, nous commençons d'abord par calculer la matrice hessienne, c'est-à-dire la dérivée seconde de l'énergie par rapport à toutes les coordonnées nucléaires. Après diagonalisation de cette dernière, on obtient les fréquences des modes normaux. Cependant, les spectres IR calculés par les méthodes *ab initio* ne correspondent pas toujours pas aux spectres expérimentaux. Dans ce cas, il est usuel d'appliquer un facteur d'échelle (entre 0,8 et 1) à l'ensemble des fréquences calculées. Cette mise à l'échelle compense le fait que la PES n'est pas purement harmonique mais suit un potentiel de Morse et que le calcul de la structure électronique est approximatif. Dans cette thèse, les fréquences et les intensités associées sont calculées au niveau DFT-B3LYP avec la base TZVP, avec un facteur d'échelle de 0,965.

Spectres électroniques d'absorption des rayons X

La DFT étant une méthode à déterminant unique (construite à partir de N orbitales Kohn-Sham de plus faible énergie), elle ne peut pas traiter correctement les problèmes de multiplets, comme c'est le cas pour les calculs d'absorption des rayons X au seuil de la couche L. Pour y remédier Roemelt et al. ont développé une méthode hybride combinant la DFT avec l'interaction de configuration ouverte restreinte aux simples excitations (ROCIS). Cette méthode hybride DFT-ROCIS implémentée dans le logiciel ORCA [127] permet de calculer de façon satisfaisante les spectres d'absorption des rayons X au seuil de la couche L des métaux de transition [129] et peut être appliquée aux excitations au seuil de la couche K des atomes légers [85], [86]. Une description

complète de cette méthode telle qu'elle est mise en œuvre dans le logiciel ORCA est expliquée dans les références [129]–[131]. Dans cette thèse, les spectres d'absorption calculés ont été réalisés en utilisant la méthode de DFT-ROCIS avec la fonctionnelle B3LYP et la base TZVP. Un décalage en énergie d'environ 12 eV (adapté en fonction de la molécule et de l'atome excité) a été appliqué à l'ensemble du spectre.

Par ailleurs, les énergies de transition calculées avec leurs intensités respectives permettent à tout programme de simulation de spectres de générer un spectre d'absorption simulé en convoluant les transitions avec des fonctions de type Gaussienne ou de type pseudo-Voigt avec une largeur maximale à demi-hauteur (FWHM) prédéfinie. En effet, il est nécessaire d'ajuster la raie d'absorption car la largeur de bande énergétique de la source lumineuse, la durée de vie de la lacune et l'élargissement électronique des transitions calculées ne sont pas pris en compte dans les calculs théoriques. Deux types de fonctions sont principalement utilisés : les fonctions de type Gaussienne qui font la moyenne de tous les types d'élargissements expérimentaux [91], [134]–[137] et les fonctions type pseudo-Voigt qui séparent l'élargissement de l'énergie des photons et la durée de vie des lacunes électroniques [138], [139]. La Figure 9.3.1 résume les différentes étapes nécessaires au calcul d'un spectre d'absorption pour une molécule donnée.

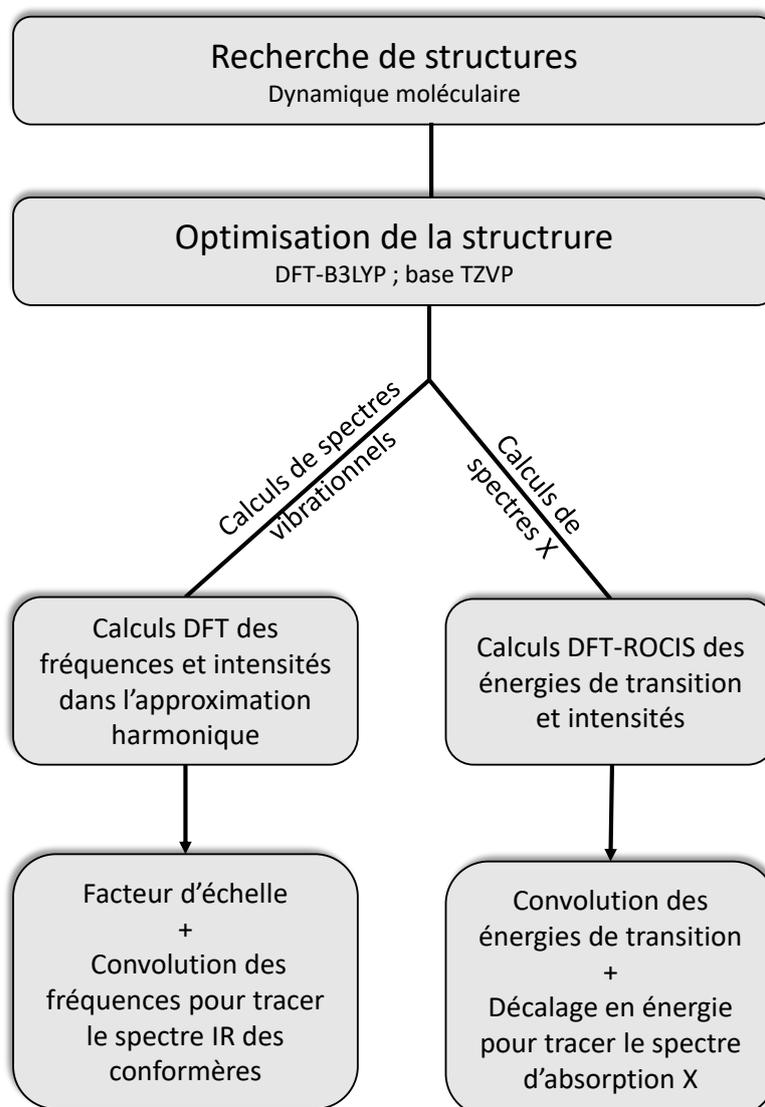


Figure 9.3.1: Schéma représentant les différentes étapes d'un calcul de spectre d'absorption infrarouge ou X.

9.3.2 Spectrométrie de masse

La spectrométrie de masse est une technique utilisée pour analyser les particules chargées en phase gazeuse. Deux installations ont été utilisées dans le cadre de cette thèse de doctorat : le spectromètre de masse en tandem du groupe FS-BIG situé au Deutsches Elektronen-Synchrotron, DESY à Hambourg, Allemagne, et la station Ion Trap du synchrotron BESSY II à Berlin, Allemagne. Les expériences ont été réalisées au synchrotron PETRA III à Hambourg et à BESSY II à Berlin. Ces dispositifs expérimentaux ont été utilisés pour obtenir les données présentées aux chapitre 5 et 6 du manuscrit principal discutés brièvement dans la suite de ce chapitre.

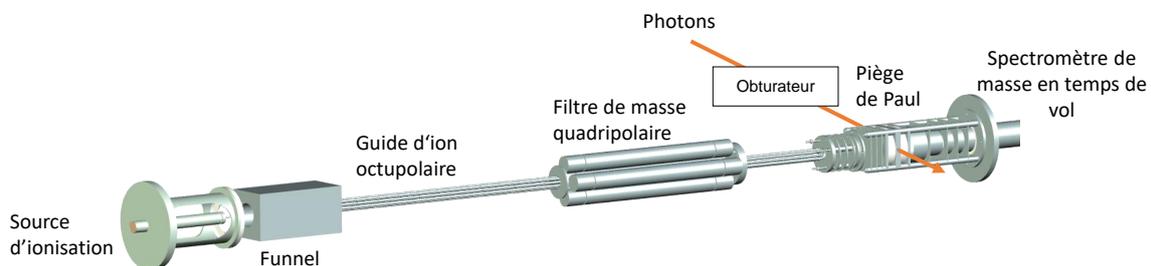


Figure 9.3.2: Schéma du spectromètre de masse Spektrolatius présent à Hambourg, Allemagne.

Le spectromètre de masse en tandem utilisé dans cette thèse a été construit dans le groupe de Prof. Sadia BARI à DESY. Un schéma de l'installation est présenté dans la figure 9.3.2. L'installation est composée de cinq chambres différentes, qui agissent toutes comme des étages de pompage différentiel, allant de la pression atmosphérique à 10^{-9} mbar. Il convient de noter que la pression de 10^{-9} mbar est la pression de base dans le piège et le spectromètre de masse en temps de vol lorsqu'aucun gaz tampon n'est introduit dans le piège à ions. Une pression aussi basse est nécessaire pour éviter, par exemple, la répulsion de charge d'espace dans le piège à ions due à l'ionisation des molécules parasites lors d'expériences utilisant des radiations ionisantes. Les valeurs de pression indiquées ci-après sont celles en fonctionnement. Cette installation repose sur une source d'ionisation par électronébulisation pour introduire les ions dans la phase gazeuse. Les ions sont transportés, à partir de la pression atmosphérique, par un capillaire chauffé dans le premier niveau de pompage, qui contient un entonnoir à ions (funnel) dont le rôle est de collecter et de concentrer les ions. Ces ions traversent ensuite un guide d'ions octupolaire à radiofréquence (RF) (pression

d'environ 10^{-3} mbar) jusqu'à un filtre de masse quadripolaire où les ions d'intérêt sont sélectionnés en fonction de leur rapport masse sur charge (m/z). Le filtre de masse quadripolaire fonctionne à une pression de 10^{-6} mbar. Ensuite, un deuxième octupole et un ensemble de lentilles d'Einzel guide et focalise les ions dans un piège de Paul, où ils sont accumulés avant d'être irradiés par des photons. Pour permettre un piégeage efficace des ions, leur énergie cinétique est réduite en refroidissant le nuage ionique jusqu'à la température ambiante par collision avec un gaz tampon d'hélium.

Par conséquent, la pression de fonctionnement du piège de Paul est fixée à environ 10^{-4} mbar, telle que mesurée dans la chambre du piège. Tous les produits cationiques sont finalement extraits du piège au moyen d'un champ électrique et analysés dans un spectromètre de masse à temps de vol par réflectron (Re-TOF) ($m/\Delta m = 1800$), qui fonctionne à une faible pression de 10^{-7} mbar. Un obturateur de faisceau peut être installé entre la chambre de piégeage et la ligne de faisceau synchrotron pour contrôler la durée de l'exposition à la lumière. L'ensemble du fonctionnement de l'installation, y compris l'injection d'ions dans le piège, l'irradiation et l'extraction dans le Re-TOF, est cyclique, avec une période typique allant de 100 ms à plusieurs secondes en fonction des paramètres de l'expérience (par exemple, le flux de photons, la section efficace d'absorption).

9.3.3 Une nouvelle source electrospray pour la production de biomolécules hydratées

Au cours des trente dernières années, des efforts considérables ont été déployés pour mettre au point des techniques expérimentales associées à une source d'ionisation par électronébulisation (ESI) qui permettent d'étudier de manière contrôlée les espèces hydratées en phase gazeuse. Différentes méthodes expérimentales ont été testées, telles que la solvatation dans la source ou la désolvatation partielle ainsi que la solvatation dans des dispositifs de piégeage à froid. Cette section passe brièvement en revue quelques exemples de ces techniques. Nguyen et al. ont été les premiers à produire des biomolécules hydratées [151], en utilisant la désolvatation partielle des ions dans la source ESI. Pour ce faire, l'aiguille ESI a été dirigée directement vers l'orifice d'entrée du capillaire, et le flux de gaz d'azote, habituellement utilisé pour faciliter la désolvatation, a été réduit de 0.6 L/min à 0.1 L/min. Cette technique a permis de produire des amines hydratées comportant jusqu'à 30 molécules d'eau. La désolvatation partielle dans la source ESI a été la méthode de choix pour la plupart des études sur les biomolécules hydratées [104], [152], [153]. Silveira et al. ont utilisé la spectrométrie de masse à mobilité ionique cryogénique (IM-MS) associée à une désolvatation par-

tielle pour sonder la limite de dissociation en champ faible pour les agrégats ionisés maintenus par des liaisons hydrogène. [104]. Ils ont montré que l'abaissement de la température du capillaire de 386 K à 340 K augmente le nombre maximum de molécules d'eau qui peuvent être maintenues autour de la bradykinine doublement protonée (de 0 à 73 molécules d'eau) mais augmente également l'intensité relative des agrégats hydratés. Cependant, la désolvatation partielle n'est pas la seule technique permettant de produire des ions solvatés à partir d'une source d'ionisation par électronébulisation ; la solvatation dans une cellule de drift ou un piège permet la production d'espèces hydratées. Liu et al. ont utilisé un dispositif spécial associé à une source ESI afin de produire des agrégats hydratés [99]. Ce dispositif était à l'origine conçu pour étudier les propriétés thermodynamiques des biomolécules. Il se compose d'une source ESI, d'une cellule de drift qui peut être chauffée jusqu'à 300 K et d'un spectromètre de masse quadripolaire. Ce dispositif permet de réhydrater les biomolécules dans la cellule de drift qui peut être remplie de vapeur d'eau à une pression pouvant atteindre jusqu'à 3 mbar et qui fonctionne donc comme une cellule de capture des molécules d'eau. Les auteurs ont ainsi obtenu une captation maximale de l'eau à 260 K et à 1,7 mbar de vapeur d'eau. Dans ces conditions expérimentales, ils ont pu attacher 20 molécules d'eau à la neurotensine triplement protonée, un peptide de 13 résidus. Marsh et al, utilisant un concept similaire à celui de Liu et al, ont mis au point un spectromètre cryogénique à piège à ions pour la formation et la caractérisation d'espèces solvatées [154]. Ce dispositif se compose d'une source à électrospray et d'un piège à ions réactionnel dans lequel ils produisent des espèces solvatées en introduisant des impulsions de gaz tampon d'hélium chargé de vapeur de solvant. En refroidissant le piège de réaction à 80 K, ils ont pu produire des ions hydratés contenant jusqu'à 30 molécules d'eau. Grâce à cette installation, ils ont caractérisé la diglycine protonée par spectroscopie de photodissociation infrarouge (IRPD) avec jusqu'à quatre molécules d'eau. Le lauréat du prix Nobel pour le développement de la technique d'électrospray a mis au point une source à électrospray dédiée à la production d'espèces solvatées, où ils désolvent d'abord complètement les ions et les hydratent à nouveau dans une deuxième chambre en introduisant de la vapeur d'eau [15], [27]. Une telle source a été améliorée par Nonose et al. en ajoutant des électrodes pour guider les ions à travers les deux chambres [24]. Avec cette technique, lorsque les agrégats hydratés sont produits dans la source, il est possible de sélectionner les agrégats hydratés d'intérêt à l'aide du filtre de masse quadripolaire. Cela permet de poursuivre l'étude des fragments hydratés à l'aide de méthodes d'activation ou de techniques spectroscopiques. La désolvatation partielle dans une source ESI dépend de la solution et nécessite de régler les paramètres de la source pour chaque nouvel échantillon, alors que la source ESI à

solvatation dédiée est conçue pour hydrater n'importe quel type de biomolécule. Zhan et Fenn ont ainsi produit des peptides hydratés contenant jusqu'à 30 molécules d'eau [27] et Nonose et al. ont produit du tryptophane protoné avec jusqu'à 40 molécules d'eau attachées [24].

9.3.3.1 Description de la source electrospray développée durant cette thèse

La source ESI d'hydratation développée au cours de cette thèse s'inspire de la source de Nonose et al [24]. Nonose nous a gracieusement fourni le schéma, qui a été déterminant pour la conception des deux chambres composant la source. Un schéma de principe de cette source est donné dans la Figure 9.3.3.

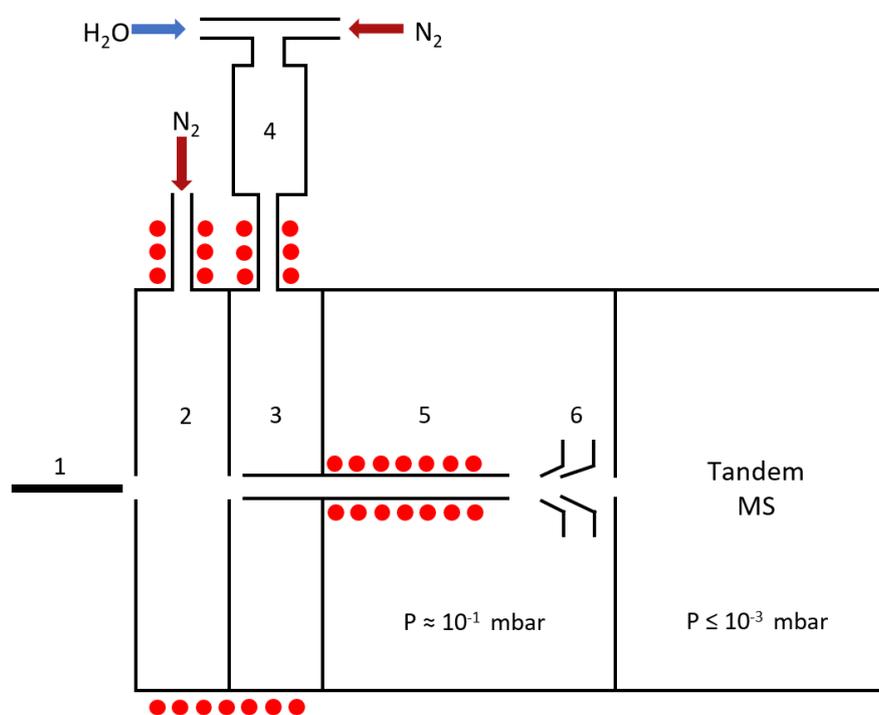


Figure 9.3.3: Schéma de la source d'hydratation et de la première chambre du spectromètre de masse Spektrolatius : (1) aiguille d'électrospray, (2) chambre de désolvatation, (3) chambre de solvatation, (4) chambre de vaporisation, (5) capillaire chauffé, (6) skimmers. Les points rouges indiquent les parties de la source qui peuvent être chauffées. La chambre de désolvatation et la chambre de solvatation sont maintenues à la pression atmosphérique. Les pressions dans la chambre de pompage différentiel contenant les deux skimmers et le spectromètre de masse sont indiquées dans le schéma.

Cette source consiste en une source d'ionisation par électrospray, mais ce qui change par rapport aux sources ESI traditionnelles, c'est l'ajout d'une chambre de solvation entre l'aiguille ESI et l'entrée du capillaire. Les ions entrent d'abord dans une chambre de séchage afin d'éliminer toutes les molécules de solvant qui pourraient encore se trouver autour des biomolécules chargées. Dans la chambre de séchage, de l'azote gazeux chauffé est injecté de manière maîtrisée pour éliminer les molécules de solvant restantes. Une fois qu'elles sont complètement débarrassées du solvant, les biomolécules pénètrent dans la chambre de solvation qui est remplie d'un gaz porteur (dans notre cas de l'azote) chargé de vapeur d'eau. Deux électrodes, espacées de 3 mm et d'un diamètre intérieur de 3 mm, sont placées entre l'aiguille et le capillaire. L'aiguille est distante de quelques millimètres de la première électrode, et le capillaire est distant de moins d'un millimètre de la sortie de la seconde électrode. Ces deux électrodes jouent deux rôles : faire la séparation entre la chambre de séchage et la chambre de solvation à travers le petit orifice (3 mm) et aider à focaliser les ions, augmentant ainsi la transmission des ions par l'application d'un champ électrique. Typiquement, en mode ions positifs, un potentiel de 300 à 500 V est appliqué sur la première électrode, et un potentiel de 50 à 150 V est appliqué sur la seconde électrode. A une distance de 3 mm après la sortie du capillaire, un ensemble de deux skimmers, espacés de 2 mm à l'entrée du spectromètre de masse, a été utilisé comme orifice de pompage différentiel et aussi pour focaliser les ions avant d'être guidés dans l'octupole. Le premier skimmer après le capillaire a un orifice de 2 mm de diamètre et un angle intérieur de 30°, tandis que le second skimmer a un orifice plus petit de 1 mm de diamètre avec un diamètre intérieur de 20°. Un schéma en 3D de la source réelle est présenté à la Figure 9.3.4.

Un débitmètre commercial de Bronkhorst, Ruurlo, Pays-Bas (EL-FLOW select) contrôle le flux d'azote gazeux. Le mélange d'azote gazeux et de vapeur d'eau est réalisé à l'aide d'un mélangeur évaporateur régulé (CEM W-102A-220-K) de Bronkhorst, couplé à un débitmètre massique pour l'azote gazeux et un régulateur de débit pour le liquide (mini CORI-FLOW, Bronkhorst) afin de contrôler le flux d'eau. Le mini CORI-FLOW doit être couplé à une pompe pour fournir une pression de liquide constante. Avec cette configuration, nous pouvons contrôler le flux de liquide et de gaz, et en fonction du flux de liquide, la température du CEM est réglée pour assurer un flux constant et stable de vapeur d'eau. Typiquement, pour obtenir une vapeur d'eau à un débit d'eau de 4 g/h et un flux d'azote de 1,5 ln/min, la température de la CEM doit être réglée à 48°C. La température du CEM peut être définie en utilisant l'outil en ligne FLUIDAT de Bronkhorst. Cet outil calcule la température qui doit être réglée

sur le CEM pour produire une vapeur stable en fonction du débit de liquide et du flux de gaz désirés.

Pour résumer, nous avons construit une nouvelle source électrospray dédiée à la production de biomolécules hydratées (dé)protonées. Au cours de cette thèse, la source a été conçue et construite. Nous avons pu produire un signal stable de biomolécules isolées de m/z 137 (acide 4-Amino-benzoïque) à m/z 891 (RWMG7) pour réaliser des expériences NEXAMS. Nous avons également pu faire un spray de cytochrome c (16+, m/z 770), une hémoprotéine d'une masse de 12 kDa. Les résultats du chapitre 5 ont été obtenus avec cette source ESI. Cependant, la capacité de la source à produire des biomolécules hydratées doit encore être testée et caractérisée. Plusieurs facteurs doivent être étudiés de manière plus approfondie afin de caractériser complètement la source et de l'utiliser pour d'autres expériences.

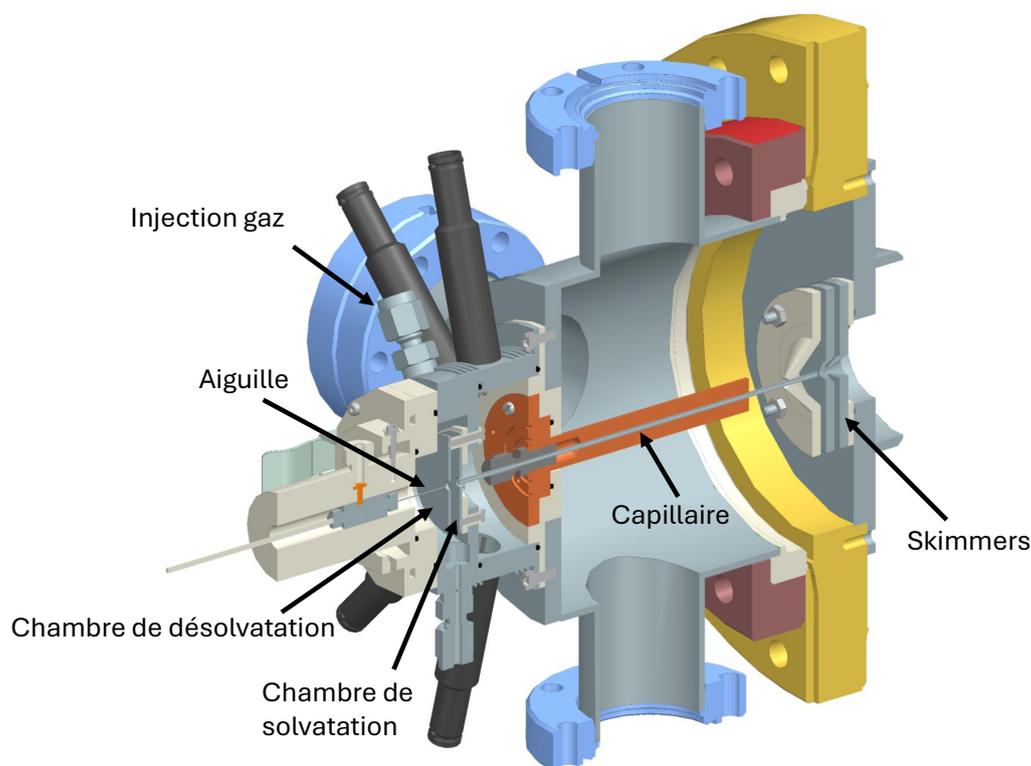


Figure 9.3.4: Vue 3D de la source d'hydratation développée pendant cette thèse.

9.4 Résultats

9.4.1 Influence du nombre de molécules d'eau sur les propriétés thermodynamiques et spectroscopiques d'un nucléotide déprotoné: Une étude théorique

L'adénosine monophosphate (AMP) est un nucléotide, élément constitutif de l'ADN et de l'ARN, qui se compose de la nucléobase adénine, du sucre ribose et d'un groupe phosphate sur lequel la déprotonation peut avoir lieu. Calvo et Douady ont mené une étude théorique sur l'hydratation et l'évaporation par étapes des anions adénosine monophosphate déprotonés (AMP^-) [101]. Dans leur article, les auteurs ont exploré la surface d'énergie potentielle (PES) des complexes AMP^- comportant jusqu'à 20 molécules d'eau en utilisant la dynamique moléculaire par échange de répliques (REMD) avec le champ de force polarisable explicite Amber ff99 associé au potentiel flexible TIP3P pour décrire les molécules d'eau. Dans les simulations REMD, 40 températures de répliques ont été utilisées dans la gamme 80 - 1000 K pour couvrir l'ensemble de la PES. Les structures les moins énergétiques trouvées avec REMD ont ensuite été optimisées à l'aide de calculs de DFT. La fonctionnelle d'échange-corrélation hybride standard B3LYP avec l'ensemble de base de type gaussien a été utilisée. En utilisant ces méthodes, ils ont constaté que l' AMP^- hydratée connaît d'importants réarrangements de solvant dans la zone de température comprise entre 150 et 200 K. Pour l' AMP^- hydratée avec vingt molécules d'eau, quelques molécules d'eau sont perdues après 1 μs , et l'agrégat perd jusqu'à la moitié de ses molécules d'eau après 1 seconde s'il est thermalisé à 350 K. Dans cette étude, nous avons poursuivi l'étude de Calvo et Douady en utilisant leurs structures optimisées pour les différents agrégats hydratés. C'est pourquoi, nous avons étudié l' AMP^- hydratée de manière théorique, en utilisant la théorie de la fonctionnelle de la densité, à partir de leur structures. Nous avons d'abord étudié l'influence des molécules d'eau sur la structure du nucléotide et l'influence du réseau de liaisons hydrogène (HB), créé par les molécules d'eau, sur les énergies de liaison de l'agrégat. À partir du réseau d'HBs et des structures calculées pour l' AMP^- hydraté, nous avons montré que lorsque l' AMP^- est isolée, sa structure est plutôt linéaire et qu'une seule molécule d'eau suffit à replier le système et à approcher le groupe phosphate de la partie adénine. De plus, avec l'augmentation des niveaux d'hydratation, les molécules d'eau supplémentaires se lient de préférence les unes aux autres pour former une goutte d'eau de taille croissante. Jusqu'à environ neuf molécules d'eau ajoutées à l' AMP^- , il y a compétition entre les liaisons inter- et intra-hydrogène, donc entre les liaisons entre la goutte et l'AMP, d'une part, et

les HBs entre les molécules d'eau, d'autre part. Nous avons observé une convergence du nombre de liaisons inter-hydrogène après huit molécules d'eau attachées à l'AMP⁻, alors que le nombre de liaisons intra-hydrogène dans la gouttelette d'eau continue d'augmenter de façon linéaire avec l'hydratation. Après dix molécules d'eau, la répulsion coulombienne entre la gouttelette d'eau et le nucléotide devient plus importante, de sorte que l'AMP⁻ se déplie. L'énergie de liaison quant à elle décroît de façon importante quand le nombre de liaisons inter-hydrogène augmente et finit par se stabiliser autour de -3,5 eV quand plus de 8 molécules d'eau forment la goutte, comme montré sur la Figure 9.4.1. Ces résultats sont en accord avec l'étude de Calvo et Douady ainsi qu'avec les premières études de Liu et al [101], [155].

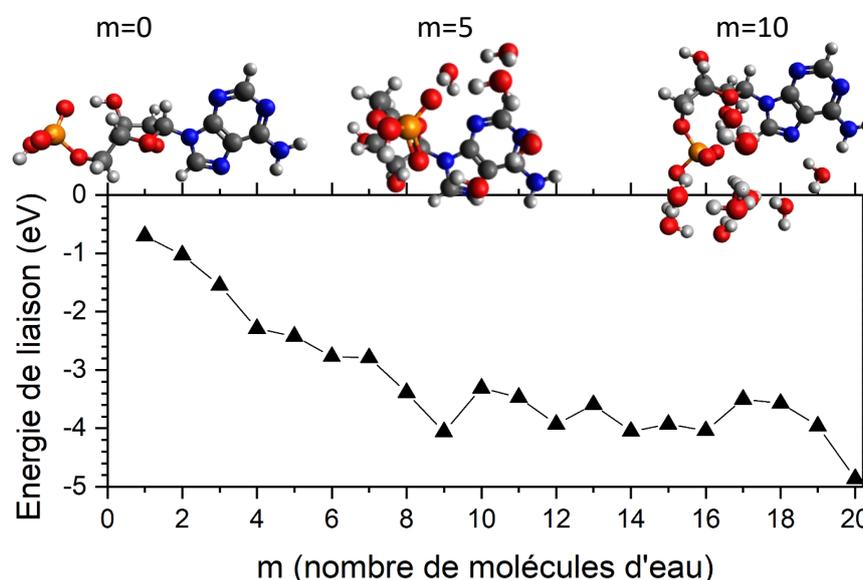


Figure 9.4.1: Énergie de liaison entre la goutte d'eau de taille croissante et le nucléotide déprotoné. Les structures optimisées de l'AMP isolée et hydratée avec cinq et dix molécules sont également représentées.

L'ajout de molécules d'eau autour du groupe phosphate du nucléotide et l'établissement de liaisons hydrogène avec le groupe adénine influencent également les propriétés spectroscopiques du système, induisant des changements observables dans les transitions électroniques et vibrationnelles des molécules. Ces changements ont été identifiés pour l'AMP⁻ hydratée dans les spectres IR calculés et les spectres d'absorption des rayons X (XAS) au seuil K de l'oxygène. Le spectre IR calculé de l'AMP⁻ isolée a montré une excellente correspondance avec le spectre IRMPD expérimental de Nei et al [50] (voir Figure 4.4.1 du Chapitre 4 dans le manuscrit en anglais). Des décalages

de fréquence dans la région de nombres d'onde 400 - 1600 cm^{-1} peuvent être observés entre les différents niveaux d'hydratation de l'AMP⁻. Dans la région 2800 - 3600 cm^{-1} , la bande correspondant aux étirements -OH des groupes hydroxyles du ribose est décalée vers le bleu dans l'AMP⁻ hydratée par rapport à l'AMP⁻ isolée, tandis que la bande provenant des étirements OH à l'intérieur de la gouttelette d'eau devient plus importante avec le niveau d'hydratation. Par ailleurs, les molécules d'eau contribuent clairement aux spectres de rayons X au seuil K de l'oxygène des différents niveaux d'hydratation de l'AMP⁻. Lorsque moins de sept molécules d'eau sont attachées, on observe un dédoublement de la première résonance à 533 eV (due aux transitions vers $\sigma^*(\text{OH})$ des groupes phosphate et ribose), ainsi qu'une résonance de faible intensité à 532,5 eV provenant des transitions $\sigma^*(\text{OH})$ dans les molécules d'eau. À des niveaux de solvation plus élevés, le nombre accru de transitions de faible intensité provenant des molécules d'eau est responsable d'une large bande centrée à 535 eV (voir Figure 4.5.1 du chapitre 4 dans le manuscrit en anglais). Les différences dans les spectres vibrationnels et électroniques qui ont été observées théoriquement doivent maintenant être confirmées par d'autres études expérimentales. Ces expériences pourraient également servir de référence pour tester notre méthode de calcul. Comme la goutte d'eau est impliquée dans un réseau de liaisons hydrogène non seulement avec le groupe phosphate mais aussi avec le groupe adénine, nous pourrions compléter notre étude en examinant l'effet de l'hydratation par étapes sur les spectres de rayons X aux seuils K du phosphore et de l'azote. De plus, ce travail pourrait être étendu à des systèmes de plus grande taille tels que l'adénosine triphosphate, qui possède trois groupes phosphates et qui est donc susceptible de permettre davantage de liaisons hydrogène entre la biomolécule et les molécules d'eau. L'étude de l'hydratation d'autres nucléotides, tels que la thymidine monophosphate, la cytidine monophosphate ou la guanosine monophosphate, nous permettrait de comprendre l'influence de la nucléobase sur l'hydratation des nucléotides.

9.4.2 Influence d'une seule molécule d'eau sur le spectre d'absorption de rayon X d'un acide aminé protoné

Pour comprendre l'effet de la présence d'une seule molécule de solvant, nous avons étudié la phosphotyrosine protonée (ci-après appelée pTyr) en phase gazeuse comme modèle de phosphorylation. La phosphotyrosine est un acide aminé dans lequel le groupe OH du groupe benzène de la tyrosine est remplacé par un groupe phosphate H_3PO_4 . La simple présence du groupe phosphate est susceptible de diriger l'attachement de l'eau vers lui, comme nous l'avons montré pour l'AMP déprotoné dans la section précédente. La pTyr a déjà été étudiée par spectroscopie de dissociation de photons multiples dans l'infrarouge (IRMPD) par Scuderi et al [107]. Ils ont montré que l'ajout d'une molécule d'eau à la pTyr décalait les elongations NH de l'ammonium et l'elongation P=O du phosphate, ce qu'ils ont pu associer à la formation d'un pont de liaisons hydrogène fortes par la molécule d'eau entre le groupe phosphate et l'ammonium.

Dans ce document, nous étendons ces travaux en étudiant l'influence d'une seule molécule d'eau sur la fragmentation par rayons X mous et sur la structure électronique de la pTyr par l'analyse des spectres de rayons X mous aux seuils K du carbone et de l'oxygène. Dans cette étude, nous avons associé expériences et calculs théoriques pour étudier l'influence d'une seule molécule d'eau sur la structure et la fragmentation de cet acide aminé phosphorylé. Nous avons obtenu les spectres d'absorption de rayons X aux seuils K du carbone et de l'oxygène d'un mélange de phosphotyrosine protonée isolée et une fois hydratée.

Les spectres de masse montrent déjà que la molécule d'eau se lie au moins au groupe phosphate. Au seuil K du carbone, nous avons montré que le spectre d'absorption de la pTyr ressemble au spectre XAS de la tyrosine neutre [177]. Alors qu'aucune empreinte spectrale de la molécule d'eau n'a été observée au seuil K du carbone, au seuil K de l'oxygène, une résonance supplémentaire à 536,5 eV est observée dans le spectre d'absorption du fragment hydraté $[PO_3H_2 + H_2O]^+$ par rapport au spectre d'absorption d'un fragment provenant de la pTyr isolée. Nous avons attribué cette résonance à une transition de la molécule d'eau vers elle-même.

La comparaison du spectre NEXAMS au seuil K du carbone avec la recherche structurale de la pTyr hydratée au niveau DFT a montré trois candidats structurels possibles pour la pTyr hydratée : une structure où la molécule d'eau se lie uniquement au groupe phosphate, une où la molécule d'eau forme un pont entre soit le groupe phosphate et le groupe NH_3^+ ou entre le groupe phosphate et le groupe carboxyle. La comparaison entre le spectre NEXAMS au seuil K de l'oxygène et le spectre d'absorption

calculé pour les trois candidats structurels suggère que la molécule d'eau se lie préférentiellement au groupe carboxyle et au groupe phosphate (présentée Figure 9.4.2). Ainsi, nous avons pu démontrer que la combinaison des calculs DFT/ROCIS et de la spectroscopie à rayons X mous permet de localiser la molécule d'eau. La Figure 9.4.2, résume les résultats obtenus à la couche K de l'oxygène et montre les deux structures préférentielles de la pTyr isolée et hydratée lors de notre campagne expérimentale.

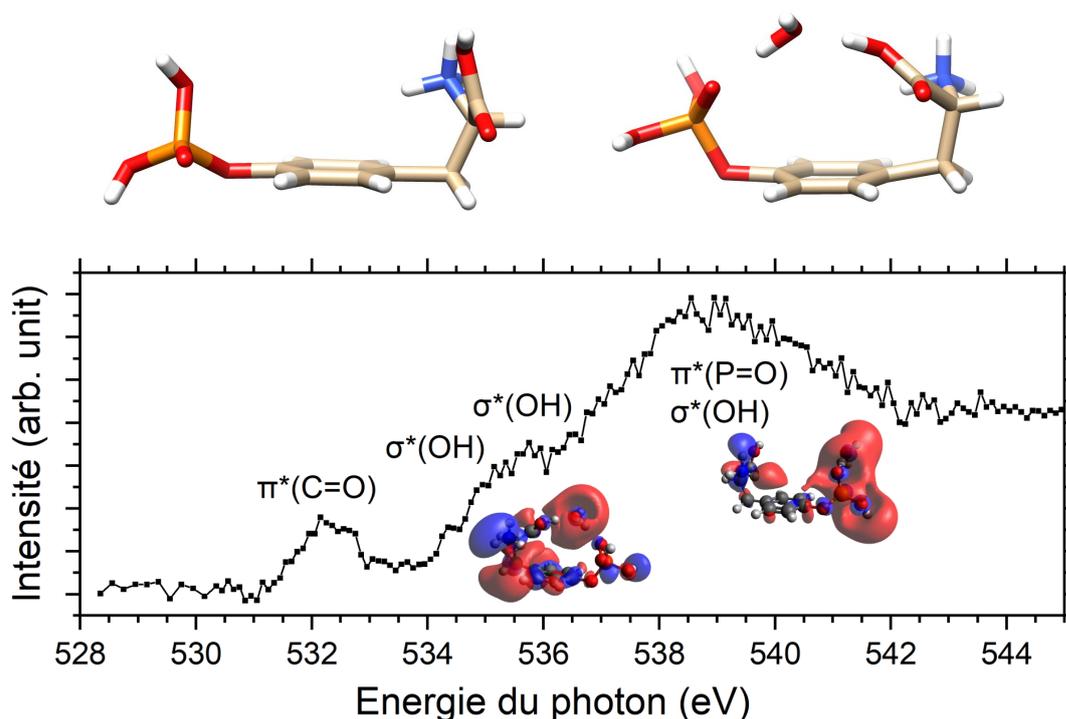


Figure 9.4.2: Spectre expérimental au seuil K de l'oxygène obtenu pour le mélange de pTyr isolée et hydratée. Les attributions des différentes résonances sont notées au dessus de celles-ci. Au dessus du spectre se trouve la représentation de la pTyr isolée, ainsi qu'une possible conformation de la pTyr hydratée, après optimisation en DFT.

Pour aller plus loin, nous pourrions commencer par optimiser nos conditions expérimentales afin de mesurer séparément les spectres d'absorption des rayons X de la pTyr isolée et de la pTyr hydratée par une molécule d'eau. Cela nous permettrait de mieux démêler la contribution des deux espèces aux spectres NEXAMS et d'obtenir plus d'informations sur l'influence de la molécule d'eau sur le spectre des rayons X au seuil K de l'oxygène. La mesure du spectre de rayons X de la pTyr hydratée au

seuil K de l'azote pourrait nous indiquer plus clairement si la structure dans laquelle l'eau est liée au groupe amino protoné peut être complètement écartée des structures présentes dans notre dispositif expérimental, car nous nous attendons à ce que l'eau influence les transitions électroniques dans cette gamme d'énergie. Des mesures de coïncidence de photoélectrons et de photoions (PEPICO) ou PEPIPICO permettraient de mieux comprendre les processus de fragmentation de la phosphotyrosine hydratée. Cela nous permettrait d'étudier spécifiquement les processus menant aux fragments hydratés, d'explorer les transferts d'énergie après irradiation au seuil K de l'oxygène, et éventuellement de démontrer la décroissance coulombienne intermoléculaire (ICD) dans les biomolécules hydratées, comme suggéré par Milosavljević et al. [17]. De plus, nous pourrions étudier un état de solvatation plus élevé pour observer si et comment le schéma de fragmentation et la structure électronique changent avec un nombre croissant de molécules d'eau. Cela nous permettrait d'observer expérimentalement ce que nous avons observé au chapitre précédent, à savoir que les transitions de l'eau vers l'eau deviennent dominantes pour les états de solvatation élevés. Enfin, l'étude d'espèces déprotonées nous permettrait d'étudier l'effet du site de déprotonation sur la structure et les spectres d'absorption X de la pTyr hydratée.

9.4.3 Signatures électroniques des sites de protonation de peptides modèles

Dans le milieu biologique, l'état de charge des molécules telles que les peptides, les protéines et les acides nucléiques, joue un rôle important car il influence la conformation géométrique et la réactivité chimique. Dans la gamme de pH physiologique, un mécanisme important modifiant l'état de charge est la protonation et la déprotonation de groupes chimiques spécifiques. Pour améliorer notre compréhension des processus biologiques et de la relation structure-fonction des molécules, une analyse des sites de (dé)protonation dans les biomolécules est nécessaire. Expérimentalement, les sites de protonation dans les peptides ont été étudiés par différentes techniques au cours des dernières décennies, principalement par des études de la basicité en phase gazeuse [188]–[191] pour déterminer les affinités protoniques des groupes chimiques. Toutefois, dans certains cas, par exemple, la protonation peut également avoir lieu sur le squelette du peptide lorsque le nombre de protons dépasse le nombre de résidus basiques présents dans le peptide. Les techniques de spectroscopie d'action, et en particulier la spectroscopie infrarouge (IR), sont des outils puissants pour la détermination des structures moléculaires en phase gazeuse qui permettent d'évaluer le site de protonation de molécules de taille comparable à celles de petits peptides [48], [49]. Au cours des dernières décennies, la spectroscopie de rayons X mous est devenue une technique majeure pour l'étude de la structure électronique des systèmes biomoléculaires, tant en phase condensée [80], [165] qu'en phase gazeuse [85], [86], [89]. En effet, l'excitation par les rayons X mous est spécifique à chaque élément, de sorte que la sélection de l'énergie du photon à une résonance particulière permet l'excitation sélective d'une molécule vers des orbitales moléculaires inoccupées spécifiques.

Pour aller plus loin et établir la correspondance entre les différentes transitions électroniques responsables des sites de protonation dans les peptides, nous avons étudié des pentapeptides à base de glycine conçus spécialement pour cette étude. La glycine est l'acide aminé le plus simple dont la chaîne latérale est constituée d'un seul hydrogène et elle a été largement étudiée [79], [81], [82], [189], [199]–[202]. Par conséquent, en étudiant des peptides composés de glycines et d'un acide aminé basique à l'extrémité C-terminal ou d'une proline à l'extrémité N-terminal, nous pouvons nous affranchir de l'étude des effets des chaînes latérales qui ne se trouvent pas sur le site de protonation, et ainsi le système demeure suffisamment simple pour une comparaison avec des calculs de chimie quantique. Pour étudier les structures électroniques des différents sites de protonation dans les peptides, nous avons analysé cinq pentapeptides pro-

tonés différents à base de glycine par spectroscopie d'action à rayons X mous. La protonation se produisant préférentiellement sur les sites azotés, nous présentons ici une étude au seuil de la couche K de l'azote de peptides modèles dont le site de protonation est connu : $[\text{GGGGG}+\text{H}]^+$, $[\text{PGGGG}+\text{H}]^+$, $[\text{GGGGH}+\text{H}]^+$, $[\text{GGGGR}+\text{H}]^+$ et $[\text{GGGGK}+\text{H}]^+$, où G représente la glycine, P pour la proline, H pour l'histidine, R pour l'arginine et K pour la lysine, ainsi que la triglycine protonée $[\text{GGG}+\text{H}]^+$. Les spectres expérimentaux obtenus sont montrés sur la figure 9.4.3.

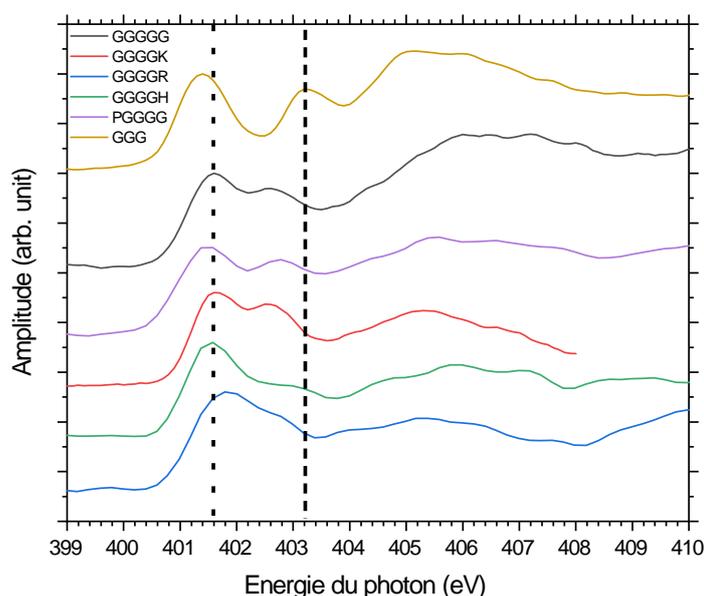


Figure 9.4.3: Spectres d'absorption expérimentaux au seuil K de l'azote des cinq pentapeptides protonés et de la triglycine protonée. Deux lignes noires en pointillées ont été introduites à 401,6 eV et à 403,1 eV pour faciliter la lecture.

Les pentapeptides protonés présentent tous une forte résonance à 401,6 eV liée à des transitions électroniques dans les liaisons peptidiques. De plus, la signature du site de protonation pour tous les peptides sauf $[\text{GGGGH}+\text{H}]^+$ est située dans la bande de résonance large entre 404 et 408 eV. Cette structure provient essentiellement des résonances σ^* du squelette du peptide, avec une contribution majeure des résonances $\sigma^*(\text{NH})$ situées sur l'atome d'azote protoné, donc soit sur le N-terminal, soit sur la chaîne latérale de la lysine et de l'arginine. Pour le peptide $[\text{GGGGH}+\text{H}]^+$, la signature du site de protonation est située à 403,1 eV que l'on peut attribuer à une transition $\text{N } 1s \rightarrow \sigma^*(\text{NH})$ de l'histidine.

Pour la triglycine protonée, nous avons démontré que deux protomères coexistent dans nos conditions expérimentales : une structure linéaire avec le site de protonation au niveau de l'oxygène amide et une structure repliée avec un site de protonation au niveau du N-terminal. Nous avons pu différencier la signature des deux structures dans le spectre d'absorption des rayons X au seuil K de l'azote grâce à la résonance à 403,1 eV (transition $N\ 1s \rightarrow \sigma^*(NH)$ d'un N-terminal non protoné), qui n'est présente que dans le conformère O-protoné, et à la large résonance centrée autour de 405,5 eV pour la transition $N\ 1s \rightarrow \sigma^*(NH)$ d'un N-terminal protoné. En comparant ces résultats expérimentaux avec des calculs DFT/ROCIS, nous avons montré que la population dans nos conditions expérimentales est représentée par une population relative de 30% de triglycine N-protonée avec 70% de triglycine O-protonée. Une suite directe à ce travail est l'étude de l'influence des sites de déprotonation dans les peptides sur la structure électronique. La description des sites de protonation et de déprotonation des peptides modèles faciliterait l'attribution expérimentale des différentes résonances dans des systèmes plus grands, tels que les petites protéines. Les systèmes plus grands sont souvent multichargés. Ainsi, nous pourrions étudier des peptides multichargés avec des protons fixes et mobiles à l'aide de NEXAMS. En outre, nous pourrions également étudier l'influence de la solvatation sur le site de protonation de ces peptides modèles qui, comme montré par Fischer et al pour la triglycine protonée, peut jouer le rôle de catalyseur pour le transfert de protons [57].

9.5 Conclusion et perspectives

Pour conclure, ce travail a montré que NEXAMS est une technique adaptée à l'étude de la géométrie et des processus de fragmentation des biomolécules hydratées et qu'elle peut être utilisée pour déterminer le site de protonation dans des systèmes au moins aussi grands que les pentapeptides. De plus, nous avons montré qu'une molécule d'eau suffit à modifier la structure d'un acide aminé phosphorylé et d'un petit nucléotide. À l'avenir, nous utiliserons la nouvelle source à électrospray pour explorer systématiquement l'effet de la première enveloppe de molécule d'eau autour des biomolécules sur leur structure et leur fragmentation à l'aide de la méthode NEXAMS. Nous pourrions également installer un système de refroidissement pour notre piège de Paul afin de réduire la population de conformères et de minimiser l'évaporation des molécules d'eau lorsque l'échantillon hydraté est introduit dans le piège. L'intégration d'un spectromètre de mobilité ionique (IMS) avant le piège de Paul permettrait également de sélectionner des conformations ioniques spécifiques.

Nous envisageons également d'utiliser cette méthode pour suivre les transferts ultra-

rapides des protons dans les biomolécules, par exemple dans le cadre d'études pompe-sonde dans les centres de laser à électrons libres à rayons X. Dans de telles expériences, une impulsion de pompe dans la gamme IR ou UV déclencherait le processus de transfert de protons, tandis qu'une impulsion de sonde dans les rayons X interrogerait la position de ce proton dans la molécule.

Les acides aminés et les sucres de l'ADN et de l'ARN sont des molécules chirales, les acides aminés étant des énantiomères gauches et les sucres des énantiomères droits. Les études de dichroïsme circulaire (CD) traitent de l'absorption différentielle d'un rayonnement polarisé circulairement par un matériau chiral, ce qui peut être utilisé pour déterminer la structure des acides aminés. Cependant, le signal CD des photons UV polarisés circulairement est assez faible (moins de 4,5 %) ; l'attribution du signal CD des ions issus d'une source électrospray est donc difficile. Dans le régime des rayons X mous, une asymétrie chirale de 20 % a été détectée dans la photoémission à partir de la couche C 1s des énantiomères de fenchone, ce qu'il serait possible d'observer avec notre système expérimental [227]. Ainsi, en combinant la spectrométrie de masse avec l'irradiation par les rayons X polarisés circulairement du rayonnement synchrotron, nous pourrions étendre les possibilités de détermination structurale de la technique NEXAMS.

Bibliography

- [1] L. J. Rothschild and R. L. Mancinelli, “Life in extreme environments,” *Nature*, vol. 409, no. 6823, pp. 1092–1101, Feb. 2001, ISSN: 0028-0836. DOI: 10.1038/35059215.
- [2] P. Ball, “Water is an active matrix of life for cell and molecular biology,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 51, pp. 13327–13335, Dec. 2017, ISSN: 0027-8424. DOI: 10.1073/pnas.1703781114.
- [3] F. Westall and A. Brack, “The Importance of Water for Life,” *Space Science Reviews*, vol. 214, no. 2, p. 50, Mar. 2018, ISSN: 0038-6308. DOI: 10.1007/s11214-018-0476-7.
- [4] V. Helms, “Protein Dynamics Tightly Connected to the Dynamics of Surrounding and Internal Water Molecules,” *ChemPhysChem*, vol. 8, no. 1, pp. 23–33, Jan. 2007, ISSN: 1439-4235. DOI: 10.1002/cphc.200600298.
- [5] S. K. Pal, J. Peon, and A. H. Zewail, “Biological water at the protein surface: Dynamical solvation probed directly with femtosecond resolution,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 4, pp. 1763–1768, Feb. 2002, ISSN: 0027-8424. DOI: 10.1073/pnas.042697899.
- [6] S. K. Pattanayak and S. Chowdhuri, “Effect of Water on Solvation Structure and Dynamics of Ions in the Peptide Bond Environment: Importance of Hydrogen Bonding and Dynamics of the Solvents,” *The Journal of Physical Chemistry B*, vol. 115, no. 45, pp. 13241–13252, Nov. 2011, ISSN: 1520-6106. DOI: 10.1021/jp206027e.
- [7] S. Chutia, M. Rossi, and V. Blum, “Water adsorption at two unsolvated peptides with a protonated lysine residue: From self-solvation to solvation,” *Journal of Physical Chemistry B*, vol. 116, no. 51, pp. 14788–14804, 2012, ISSN: 15205207. DOI: 10.1021/jp3098268.
- [8] M. R. N. Murthy, “Protein Hydration,” *Current Science*, vol. 120, no. 1, p. 186, Jan. 2021, ISSN: 0011-3891. DOI: 10.18520/cs/v120/i1/186-192.

- [9] A. R. Fersht, J.-p. Shi, J. Knill-Jones, *et al.*, “Hydrogen bonding and biological specificity analysed by protein engineering,” *Nature*, vol. 314, no. 6008, pp. 235–238, Mar. 1985, ISSN: 0028-0836. DOI: 10.1038/314235a0.
- [10] M. C. Bellissent-Funel, A. Hassanali, M. Havenith, *et al.*, “Water Determines the Structure and Dynamics of Proteins,” *Chemical Reviews*, vol. 116, no. 13, pp. 7673–7697, 2016, ISSN: 15206890. DOI: 10.1021/acs.chemrev.5b00664.
- [11] J. Gu, J. Leszczynski, and H. F. Schaefer, “Interactions of Electrons with Bare and Hydrated Biomolecules: From Nucleic Acid Bases to DNA Segments,” *Chemical Reviews*, vol. 112, no. 11, pp. 5603–5640, Nov. 2012, ISSN: 0009-2665. DOI: 10.1021/cr3000219.
- [12] E. Wang, X. Ren, W. Baek, H. Rabus, T. Pfeifer, and A. Dorn, “Water acting as a catalyst for electron-driven molecular break-up of tetrahydrofuran,” *Nature Communications*, vol. 11, no. 1, p. 2194, May 2020, ISSN: 2041-1723. DOI: 10.1038/s41467-020-15958-7.
- [13] M. Smyth, J. Kohanoff, and I. I. Fabrikant, “Electron-induced hydrogen loss in uracil in a water cluster environment,” *The Journal of Chemical Physics*, vol. 140, no. 18, May 2014, ISSN: 0021-9606. DOI: 10.1063/1.4874841.
- [14] B. Liu, S. B. Nielsen, P. Hvelplund, *et al.*, “Collision-induced dissociation of hydrated adenosine monophosphate nucleotide ions: Protection of the ion in water nanoclusters,” *Physical Review Letters*, vol. 97, no. 13, p. 133401, Sep. 2006, ISSN: 00319007. DOI: 10.1103/PhysRevLett.97.133401.
- [15] D. Zhan, J. Rosell, and J. B. Fenn, “Solvation studies of electrospray ions - Method and early results,” *Journal of the American Society for Mass Spectrometry*, vol. 9, no. 12, pp. 1241–1247, 1998, ISSN: 10440305. DOI: 10.1016/S1044-0305(98)00107-X.
- [16] K. C. Fischer, S. L. Sherman, J. M. Voss, J. Zhou, and E. Garand, “Microsolvation Structures of Protonated Glycine and l -Alanine,” *Journal of Physical Chemistry A*, vol. 123, no. 15, pp. 3355–3366, 2019, ISSN: 15205215. DOI: 10.1021/acs.jpca.9b01578.
- [17] A. R. Milosavljević, K. Jänkälä, M. L. Ranković, *et al.*, “Oxygen K-shell spectroscopy of isolated progressively solvated peptide,” *Physical Chemistry Chemical Physics*, vol. 22, no. 23, pp. 12909–12917, 2020, ISSN: 1463-9076. DOI: 10.1039/D0CP00994F.

- [18] J. Kočišek, A. Pysanenko, M. Fárník, and J. Fedor, “Microhydration Prevents Fragmentation of Uracil and Thymine by Low-Energy Electrons,” *Journal of Physical Chemistry Letters*, vol. 7, no. 17, pp. 3401–3405, 2016, ISSN: 19487185. DOI: 10.1021/acs.jpcllett.6b01601.
- [19] A. R. Milosavljevic, V. Z. Cerovski, F. Canon, L. Nahon, and A. Giuliani, “Nanosolvation-induced stabilization of a protonated peptide dimer isolated in the Gas phase,” *Angewandte Chemie - International Edition*, vol. 52, no. 28, pp. 7286–7290, 2013, ISSN: 14337851. DOI: 10.1002/anie.201301667.
- [20] B. Liu, N. Haag, H. Johansson, *et al.*, “Electron capture induced dissociation of nucleotide anions in water nanodroplets,” *Journal of Chemical Physics*, vol. 128, no. 7, Feb. 2008, ISSN: 00219606. DOI: 10.1063/1.2839597.
- [21] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, *Electrospray ionization for mass spectrometry of large biomolecules*, 1989. DOI: 10.1126/science.2675315.
- [22] M. Kohtani and M. F. Jarrold, “The initial steps in the hydration of unsolvated peptides: Water molecule adsorption on alanine-based helices and globules,” *Journal of the American Chemical Society*, vol. 124, no. 37, pp. 11 148–11 158, Sep. 2002, ISSN: 00027863. DOI: 10.1021/ja012755r.
- [23] L. C. Snoek, R. T. Kroemer, and J. P. Simons, “A spectroscopic and computational exploration of tryptophan-water cluster structures in the gas phase,” in *Physical Chemistry Chemical Physics*, vol. 4, 2002, pp. 2130–2139. DOI: 10.1039/b200059h.
- [24] S. Nonose, S. Iwaoka, K. Mori, Y. Shibata, and K. Fuke, “Structures and reactions of hydrated biomolecular cluster ions,” *European Physical Journal D*, vol. 34, no. 1-3, pp. 315–319, Jul. 2005, ISSN: 14346060. DOI: 10.1140/epjd/e2005-00166-9.
- [25] N. S. Nagornova, T. R. Rizzo, and O. V. Boyarkin, “Interplay of intra- and intermolecular H-bonding in a progressively solvated macrocyclic peptide,” *Science*, vol. 336, no. 6079, pp. 320–323, 2012, ISSN: 10959203. DOI: 10.1126/science.1218709.
- [26] C. Baldauf and M. Rossi, “Going clean: Structure and dynamics of peptides in the gas phase and paths to solvation,” *Journal of Physics Condensed Matter*, vol. 27, no. 49, 2015, ISSN: 1361648X. DOI: 10.1088/0953-8984/27/49/493002.

- [27] D. Zhan and J. B. Fenn, “Gas phase hydration of electrospray ions from small peptides,” *International Journal of Mass Spectrometry*, vol. 219, no. 1, pp. 1–10, 2002, ISSN: 13873806. DOI: 10.1016/S1387-3806(02)00545-6.
- [28] S. Bari, D. Egorov, T. L. Jansen, *et al.*, “Soft X-ray Spectroscopy as a Probe for Gas-Phase Protein Structure: Electron Impact Ionization from Within,” *Chemistry - A European Journal*, vol. 24, no. 30, pp. 7631–7636, May 2018, ISSN: 15213765. DOI: 10.1002/chem.201801440.
- [29] E. Garand, “Spectroscopy of Reactive Complexes and Solvated Clusters: A Bottom-Up Approach Using Cryogenic Ion Traps,” *The Journal of Physical Chemistry A*, vol. 122, no. 32, pp. 6479–6490, Aug. 2018, ISSN: 1089-5639. DOI: 10.1021/acs.jpca.8b05712.
- [30] A. R. Milosavljević, F. Canon, C. Nicolas, C. Miron, L. Nahon, and A. Giuliani, “Gas-phase protein inner-shell spectroscopy by coupling an ion trap with a soft X-ray beamline,” *Journal of Physical Chemistry Letters*, vol. 3, no. 9, pp. 1191–1196, 2012, ISSN: 19487185. DOI: 10.1021/jz300324z.
- [31] D. de Wied, “Peptides and behavior,” *Life Sciences*, vol. 20, no. 2, pp. 195–204, Jan. 1977, ISSN: 00243205. DOI: 10.1016/0024-3205(77)90313-7.
- [32] H. Michel, J. Behr, A. Harrenga, and A. Kannt, “CYTOCHROME C OXIDASE: Structure and Spectroscopy,” *Annual Review of Biophysics and Biomolecular Structure*, vol. 27, no. 1, pp. 329–356, Jun. 1998, ISSN: 1056-8700. DOI: 10.1146/annurev.biophys.27.1.329.
- [33] J. M. Lee, H. M. Hammarén, M. M. Savitski, and S. H. Baek, “Control of protein stability by post-translational modifications,” *Nature Communications*, vol. 14, no. 1, pp. 1–16, 2023, ISSN: 20411723. DOI: 10.1038/s41467-023-35795-8.
- [34] S. Ramazi and J. Zahiri, “Post-translational modifications in proteins: Resources, tools and prediction methods,” *Database*, vol. 2021, no. 7, pp. 1–20, 2021, ISSN: 17580463. DOI: 10.1093/database/baab012.
- [35] P. Cohen, “The origins of protein phosphorylation,” *Nature Cell Biology*, vol. 4, no. 5, E127–E130, May 2002, ISSN: 1465-7392. DOI: 10.1038/ncb0502-e127.
- [36] P. Roepstorff and J. Fohlman, *Letter to the editors*, 1984. DOI: 10.1002/bms.1200111109.
- [37] K. Biemann, “Contributions of mass spectrometry to peptide and protein structure,” *Biological Mass Spectrometry*, vol. 16, no. 1-12, pp. 99–111, 1988, ISSN: 10969888. DOI: 10.1002/bms.1200160119.

- [38] L. Pauling and R. B. Corey, "A Proposed Structure For The Nucleic Acids," *Proceedings of the National Academy of Sciences*, vol. 39, no. 2, pp. 84–97, Feb. 1953, ISSN: 0027-8424. DOI: 10.1073/pnas.39.2.84.
- [39] J. RIBEIRO and J. WALKER, "THE EFFECTS OF ADENOSINE TRIPHOSPHATE AND ADENOSINE DIPHOSPHATE ON TRANSMISSION AT THE RAT AND FROG NEUROMUSCULAR JUNCTIONS," *British Journal of Pharmacology*, vol. 54, no. 2, pp. 213–218, Jun. 1975, ISSN: 0007-1188. DOI: 10.1111/j.1476-5381.1975.tb06931.x.
- [40] M. Rajendran, E. Dane, J. Conley, and M. Tantama, "Imaging Adenosine Triphosphate (ATP)," *The Biological Bulletin*, vol. 231, no. 1, pp. 73–84, Aug. 2016, ISSN: 0006-3185. DOI: 10.1086/689592.
- [41] M. Jauker, H. Griesser, and C. Richert, "Spontaneous Formation of RNA Strands, Peptidyl RNA, and Cofactors," *Angewandte Chemie - International Edition*, vol. 54, no. 48, pp. 14 564–14 569, 2015, ISSN: 15213773. DOI: 10.1002/anie.201506593.
- [42] N. C. Polfer, "Infrared multiple photon dissociation spectroscopy of trapped ions," *Chemical Society Reviews*, vol. 40, no. 5, pp. 2211–2221, 2011, ISSN: 14604744. DOI: 10.1039/c0cs00171f.
- [43] ISO-21348:2007, "Space environment (natural and artificial)-Process for determining solar irradiances," Tech. Rep., 2007.
- [44] J. A. Stearns, S. Mercier, C. Seaiby, M. Guidi, O. V. Boyarkin, and T. R. Rizzo, "Conformation-Specific Spectroscopy and Photodissociation of Cold, Protonated Tyrosine and Phenylalanine," *Journal of the American Chemical Society*, vol. 129, no. 38, pp. 11 814–11 820, Sep. 2007, ISSN: 0002-7863. DOI: 10.1021/ja0736010.
- [45] M. Okumura, L. I. Yeh, J. D. Myers, and Y. T. Lee, "Infrared spectra of the cluster ions H_7O+3H_2 and H_9O+4H_2 ," *The Journal of Chemical Physics*, vol. 85, no. 4, pp. 2328–2329, Aug. 1986, ISSN: 0021-9606. DOI: 10.1063/1.451079.
- [46] S. Warnke, A. Ben Faleh, R. P. Pellegrinelli, N. Yalovenko, and T. R. Rizzo, "Combining ultra-high resolution ion mobility spectrometry with cryogenic IR spectroscopy for the study of biomolecular ions," *Faraday Discussions*, vol. 217, pp. 114–125, 2019, ISSN: 13645498. DOI: 10.1039/c8fd00180d.
- [47] J. M. Headrick, E. G. Diken, R. S. Walters, *et al.*, "Spectral Signatures of Hydrated Proton Vibrations in Water Clusters," *Science*, vol. 308, no. 5729, pp. 1765–1769, Jun. 2005, ISSN: 0036-8075. DOI: 10.1126/science.1113094.

- [48] G. Grégoire, M. P. Gageot, D. C. Marinica, J. Lemaire, J. P. Schermann, and C. Desfrancois, “Resonant infrared multiphoton dissociation spectroscopy of gas-phase protonated peptides. Experiments and Car-Parrinello dynamics at 300 K,” *Physical Chemistry Chemical Physics*, vol. 9, no. 24, pp. 3082–3097, 2007, ISSN: 14639076. DOI: 10.1039/b618094a.
- [49] B. Lucas, G. Grégoire, J. Lemaire, *et al.*, “Investigation of the protonation site in the dialanine peptide by infrared multiphoton dissociation spectroscopy,” *Physical Chemistry Chemical Physics*, vol. 6, no. 10, pp. 2659–2663, 2004, ISSN: 14639076. DOI: 10.1039/b316053j.
- [50] Y. W. Nei, K. T. Crampton, G. Berden, J. Oomens, and M. T. Rodgers, “Infrared multiple photon dissociation action spectroscopy of deprotonated RNA mononucleotides: Gas-phase conformations and energetics,” *Journal of Physical Chemistry A*, vol. 117, no. 41, pp. 10 634–10 649, 2013, ISSN: 10895639. DOI: 10.1021/jp4039495.
- [51] C. F. Correia, P. O. Balaj, D. Scuderi, P. Maitre, and G. Ohanessian, “Vibrational signatures of protonated, phosphorylated amino acids in the gas phase,” *Journal of the American Chemical Society*, vol. 130, no. 11, pp. 3359–3370, 2008, ISSN: 00027863. DOI: 10.1021/ja073868z.
- [52] A. Abo-Riziq, B. O. Crews, I. Compagnon, *et al.*, “The mid-IR spectra of 9-ethyl guanine, guanosine, and 2-deoxyguanosine,” *Journal of Physical Chemistry A*, vol. 111, no. 31, pp. 7529–7536, Aug. 2007, ISSN: 10895639. DOI: 10.1021/jp072183i.
- [53] J.-Y. Salpin, S. Guillaumont, J. Tortajada, L. MacAleese, J. Lemaire, and P. Maitre, “Infrared Spectra of Protonated Uracil, Thymine and Cytosine,” *ChemPhysChem*, vol. 8, no. 15, pp. 2235–2244, Oct. 2007, ISSN: 1439-4235. DOI: 10.1002/cphc.200700284.
- [54] B. Lucas, G. Grégoire, J. Lemaire, *et al.*, “Infrared multiphoton dissociation spectroscopy of protonated N-acetyl-alanine and alanyl-histidine,” *International Journal of Mass Spectrometry*, vol. 243, no. 2, pp. 105–113, May 2005, ISSN: 13873806. DOI: 10.1016/j.ijms.2005.01.002.
- [55] M. Demireva, J. T. O’Brien, and E. R. Williams, “Water-induced folding of 1,7-diammoniumheptane,” *Journal of the American Chemical Society*, vol. 134, no. 27, pp. 11 216–11 224, Jul. 2012, ISSN: 00027863. DOI: 10.1021/ja303313p.

- [56] P. Çarçabal, R. T. Kroemer, L. C. Snoek, *et al.*, “Hydrated complexes of tryptophan: Ion dip infrared spectroscopy in the ‘molecular fingerprint’ region, 100–2000 cm^{-1} ,” *Physical Chemistry Chemical Physics*, vol. 6, no. 19, pp. 4546–4552, Oct. 2004, ISSN: 14639076. DOI: 10.1039/b411757c.
- [57] K. C. Fischer, J. M. Voss, J. Zhou, and E. Garand, “Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of $\text{Gly}_3\text{H}^+(\text{H}_2\text{O})$,” *Journal of Physical Chemistry A*, vol. 122, no. 41, pp. 8213–8221, 2018, ISSN: 15205215. DOI: 10.1021/acs.jpca.8b07546.
- [58] K. Hirata, F. Haddad, O. Dopfer, S. I. Ishiuchi, and M. Fujii, “Collision-assisted stripping for determination of microsolvation-dependent protonation sites in hydrated clusters by cryogenic ion trap infrared spectroscopy: the case of $\text{benzocaineH}^+(\text{H}_2\text{O})_n$,” *Physical Chemistry Chemical Physics*, vol. 24, no. 10, pp. 5774–5779, 2022, ISSN: 14639076. DOI: 10.1039/d1cp05762f.
- [59] H. Kang, K. T. Lee, B. Jung, Y. J. Ko, and S. K. Kim, “Intrinsic Lifetimes of the Excited State of DNA and RNA Bases,” *Journal of the American Chemical Society*, vol. 124, no. 44, pp. 12 958–12 959, Nov. 2002, ISSN: 0002-7863. DOI: 10.1021/ja027627x.
- [60] M. Barbatti, A. J. A. Aquino, J. J. Szymczak, D. Nachtigallová, P. Hobza, and H. Lischka, “Relaxation mechanisms of UV-photoexcited DNA and RNA nucleobases,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 50, pp. 21 453–21 458, Dec. 2010, ISSN: 0027-8424. DOI: 10.1073/pnas.1014982107.
- [61] S. B. Nielsen, J. U. Andersen, J. S. Forster, *et al.*, “Photodestruction of Adenosine 5'-Monophosphate (AMP) Nucleotide Ions in vacuo : Statistical versus Nonstatistical Processes,” *Physical Review Letters*, vol. 91, no. 4, p. 048 302, Jul. 2003, ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.91.048302.
- [62] D. Creed, “THE PHOTOPHYSICS AND PHOTOCHEMISTRY OF THE NEAR-UV ABSORBING AMINO ACIDS—II. TYROSINE AND ITS SIMPLE DERIVATIVES,” *Photochemistry and Photobiology*, vol. 39, no. 4, pp. 563–575, Apr. 1984, ISSN: 0031-8655. DOI: 10.1111/j.1751-1097.1984.tb03891.x.
- [63] J. P. Reilly, “Ultraviolet photofragmentation of biomolecular ions,” *Mass Spectrometry Reviews*, vol. 28, no. 3, pp. 425–447, May 2009, ISSN: 0277-7037. DOI: 10.1002/mas.20214.

- [64] M. Pérot, B. Lucas, M. Barat, J. A. Fayeton, and C. Jouvet, “Mechanisms of UV Photodissociation of Small Protonated Peptides,” *The Journal of Physical Chemistry A*, vol. 114, no. 9, pp. 3147–3156, Mar. 2010, ISSN: 1089-5639. DOI: 10.1021/jp908937s.
- [65] F. Canon, A. R. Milosavljevic, G. Van Der Rest, *et al.*, “Photodissociation and dissociative photoionization mass spectrometry of proteins and noncovalent protein-ligand complexes,” *Angewandte Chemie - International Edition*, vol. 52, no. 32, pp. 8377–8381, Aug. 2013, ISSN: 14337851. DOI: 10.1002/anie.201304046.
- [66] C. Brunet, R. Antoine, P. Dugourd, F. Canon, A. Giuliani, and L. Nahon, “Formation and fragmentation of radical peptide anions: Insights from vacuum ultraviolet spectroscopy,” *Journal of the American Society for Mass Spectrometry*, vol. 23, no. 2, pp. 274–281, 2012, ISSN: 10440305. DOI: 10.1007/s13361-011-0285-7.
- [67] M. Schwell, H.-W. Jochims, H. Baumgärtel, F. Dulieu, and S. Leach, “VUV photochemistry of small biomolecules,” *Planetary and Space Science*, vol. 54, no. 11, pp. 1073–1085, Sep. 2006, ISSN: 00320633. DOI: 10.1016/j.pss.2006.05.017.
- [68] O. Licht, D. Barreiro-Lage, P. Rousseau, *et al.*, “Peptide Bond Formation in the Protonated Serine Dimer Following Vacuum UV Photon-Induced Excitation,” *Angewandte Chemie*, vol. 135, no. 15, Apr. 2023, ISSN: 0044-8249. DOI: 10.1002/ange.202218770.
- [69] J. Yeh and I. Lindau, “Atomic subshell photoionization cross sections and asymmetry parameters: $1 < Z < 103$,” *Atomic Data and Nuclear Data Tables*, vol. 32, no. 1, pp. 1–155, Jan. 1985, ISSN: 0092640X. DOI: 10.1016/0092-640X(85)90016-6.
- [70] J. Eland, P. Linusson, L. Hedin, E. Andersson, J. Rubensson, and R. Feifel, “Triple ionisation of methane by double Auger and related pathways,” *Chemical Physics Letters*, vol. 485, no. 1-3, pp. 21–25, 2010, ISSN: 00092614. DOI: 10.1016/j.cpllett.2009.11.072.
- [71] T. A. Carlson and M. O. Krause, “Experimental Evidence for Double Electron Emission in an Auger Process,” *Physical Review Letters*, vol. 14, no. 11, pp. 390–392, Mar. 1965, ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.14.390.

- [72] A. Müller, A. Borovik, T. Buhr, *et al.*, “Observation of a Four-Electron Auger Process in Near- K -Edge Photoionization of Singly Charged Carbon Ions,” *Physical Review Letters*, vol. 114, no. 1, p. 013 002, Jan. 2015, ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.114.013002.
- [73] D. Egorov, L. Schwob, M. Lalande, R. Hoekstra, and T. Schlathölder, “Near edge X-ray absorption mass spectrometry of gas phase proteins: The influence of protein size,” *Physical Chemistry Chemical Physics*, vol. 18, no. 37, pp. 26 213–26 223, 2016, ISSN: 14639076. DOI: 10.1039/c6cp05254a.
- [74] X. Wang, S. Rathnachalam, V. Zamudio-Bayer, *et al.*, “Intramolecular hydrogen transfer in DNA induced by site-selective resonant core excitation,” *Physical Chemistry Chemical Physics*, vol. 24, no. 13, pp. 7815–7825, 2022, ISSN: 14639076. DOI: 10.1039/d1cp05741c.
- [75] J. Stöhr, *NEXAFS Spectroscopy* (Springer Series in Surface Sciences). Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, vol. 25, ISBN: 978-3-642-08113-2. DOI: 10.1007/978-3-662-02853-7.
- [76] G. Hähner, “Near edge X-ray absorption fine structure spectroscopy as a tool to probe electronic and structural properties of thin organic films and liquids,” *Chem. Soc. Rev.*, vol. 35, no. 12, pp. 1244–1255, 2006, ISSN: 0306-0012. DOI: 10.1039/B509853J.
- [77] J. Boese, A. Osanna, C. Jacobsen, and J. Kirz, “Carbon edge XANES spectroscopy of amino acids and peptides,” *Journal of Electron Spectroscopy and Related Phenomena*, vol. 85, no. 1-2, pp. 9–15, Jul. 1997, ISSN: 03682048. DOI: 10.1016/S0368-2048(97)00032-7.
- [78] Y. Zubavichus, A. Shaporenko, V. Korolkov, M. Grunze, and M. Zharnikov, “X-ray Absorption Spectroscopy of the Nucleotide Bases at the Carbon, Nitrogen, and Oxygen K-Edges,” *The Journal of Physical Chemistry B*, vol. 112, no. 44, pp. 13 711–13 716, Nov. 2008, ISSN: 1520-6106. DOI: 10.1021/jp802453u.
- [79] Y. Zubavichus, M. Zharnikov, A. Schaporenko, and M. Grunze, “NEXAFS study of glycine and glycine-based oligopeptides,” *Journal of Electron Spectroscopy and Related Phenomena*, vol. 134, no. 1, pp. 25–33, Jan. 2004, ISSN: 03682048. DOI: 10.1016/j.elspec.2003.09.006.
- [80] Y. Zubavichus, A. Shaporenko, M. Grunze, and M. Zharnikov, “Innershell Absorption Spectroscopy of Amino Acids at All Relevant Absorption Edges,” *The Journal of Physical Chemistry A*, vol. 109, no. 32, pp. 6998–7000, Aug. 2005, ISSN: 1089-5639. DOI: 10.1021/jp0535846.

- [81] G. Cooper, M. Gordon, D. Tulumello, C. Turci, K. Kaznatcheev, and A. P. Hitchcock, "Inner shell excitation of glycine, glycyglycine, alanine and phenylalanine," in *Journal of Electron Spectroscopy and Related Phenomena*, vol. 137-140, Jul. 2004, pp. 795-799. DOI: 10.1016/j.elspec.2004.02.102.
- [82] M. L. Gordon, G. Cooper, C. Morin, *et al.*, "Inner-shell excitation spectroscopy of the peptide bond: Comparison of the C 1s, N 1s, and O 1s spectra of glycine, glycyglycine, and glycyglycyglycine," *Journal of Physical Chemistry A*, vol. 107, no. 32, pp. 6144-6159, Aug. 2003, ISSN: 10895639. DOI: 10.1021/jp0344390.
- [83] S. M. Kirtley, O. C. Mullins, J. Chen, *et al.*, "Nitrogen chemical structure in DNA and related molecules by X-ray absorption spectroscopy," *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, vol. 1132, no. 3, pp. 249-254, Oct. 1992, ISSN: 01674781. DOI: 10.1016/0167-4781(92)90157-U.
- [84] O. González-Magaña, G. Reitsma, M. Tiemens, L. Boschman, R. Hoekstra, and T. Schlathölter, "Near-Edge X-ray Absorption Mass Spectrometry of a Gas-Phase Peptide," *The Journal of Physical Chemistry A*, vol. 116, no. 44, pp. 10745-10751, Nov. 2012, ISSN: 1089-5639. DOI: 10.1021/jp307527b.
- [85] L. Schwob, S. Dörner, K. Atak, *et al.*, "Site-Selective Dissociation upon Sulfur L-Edge X-ray Absorption in a Gas-Phase Protonated Peptide," *Journal of Physical Chemistry Letters*, vol. 11, no. 4, pp. 1215-1221, Feb. 2020, ISSN: 19487185. DOI: 10.1021/acs.jpcllett.0c00041.
- [86] S. Dörner, L. Schwob, K. Atak, *et al.*, "Probing Structural Information of Gas-Phase Peptides by Near-Edge X-ray Absorption Mass Spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 32, no. 3, pp. 670-684, Mar. 2021, ISSN: 18791123. DOI: 10.1021/jasms.0c00390.
- [87] D. Egorov, S. Bari, R. Boll, *et al.*, "Near-Edge Soft X-ray Absorption Mass Spectrometry of Protonated Melittin," *Journal of the American Society for Mass Spectrometry*, vol. 29, no. 11, pp. 2138-2151, Nov. 2018, ISSN: 18791123. DOI: 10.1007/s13361-018-2035-6.
- [88] L. Schwob, M. Lalande, D. Egorov, *et al.*, "Radical-driven processes within a peptidic sequence of type I collagen upon single-photon ionisation in the gas phase," *Physical Chemistry Chemical Physics*, vol. 19, no. 34, pp. 22865-22904, 2017, ISSN: 14639076. DOI: 10.1039/c7cp03376a.

-
- [89] A. R. Milosavljević, C. Nicolas, M. L. Ranković, F. Canon, C. Miron, and A. Giuliani, “K-Shell Excitation and Ionization of a Gas-Phase Protein: Interplay between Electronic Structure and Protein Folding,” *Journal of Physical Chemistry Letters*, vol. 6, no. 16, pp. 3132–3138, Aug. 2015, ISSN: 19487185. DOI: 10.1021/acs.jpcllett.5b01288.
- [90] K. Hirsch, J. T. Lau, P. Klar, *et al.*, “X-ray spectroscopy on size-selected clusters in an ion trap: from the molecular limit to bulk properties,” *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 42, no. 15, p. 154029, Aug. 2009, ISSN: 0953-4075. DOI: 10.1088/0953-4075/42/15/154029.
- [91] M. W. Buckley and N. A. Besley, “A theoretical study of the near edge X-ray absorption fine structure of amino acids and proteins,” *Chemical Physics Letters*, vol. 501, no. 4-6, pp. 540–546, Jan. 2011, ISSN: 00092614. DOI: 10.1016/j.cplett.2010.12.004.
- [92] H. Shimada, T. Fukao, H. Minami, *et al.*, “Structural changes of nucleic acid base in aqueous solution as observed in X-ray absorption near edge structure (XANES),” *Chemical Physics Letters*, vol. 591, pp. 137–141, Jan. 2014, ISSN: 00092614. DOI: 10.1016/j.cplett.2013.11.026.
- [93] G. Mattioli, R. Schürmann, C. Nicolafrancesco, A. Giuliani, and A. R. Milosavljević, “Effect of Protonation on the Molecular Structure of Adenosine 5-Triphosphate: A Combined Theoretical and Near Edge X-ray Absorption Fine Structure Study,” *Journal of Physical Chemistry Letters*, vol. 14, no. 45, pp. 10173–10180, 2023, ISSN: 19487185. DOI: 10.1021/acs.jpcllett.3c01666.
- [94] M. Ahmad, W. Gu, T. Geyer, and V. Helms, “Adhesive water networks facilitate binding of protein interfaces,” *Nature Communications*, vol. 2, no. 1, 2011, ISSN: 20411723. DOI: 10.1038/ncomms1258.
- [95] T. Sun, F.-h. Lin, R. L. Campbell, J. S. Allingham, and P. L. Davies, “An Antifreeze Protein Folds with an Interior Network of More Than 400 Semi-Clathrate Waters,” *Science*, vol. 343, no. 6172, pp. 795–798, Feb. 2014, ISSN: 0036-8075. DOI: 10.1126/science.1247407.
- [96] D. S. Ahn, S. W. Park, I. S. Jeon, *et al.*, “Effects of Microsolvation on the Structures and Reactions of Neutral and Zwitterion Alanine: Computational Study,” *Journal of Physical Chemistry B*, vol. 107, no. 50, pp. 14109–14118, 2003, ISSN: 15206106. DOI: 10.1021/jp031041v.

- [97] C. M. Aikens and M. S. Gordon, "Incremental solvation of nonionized and zwitterionic glycine," *Journal of the American Chemical Society*, vol. 128, no. 39, pp. 12 835–12 850, 2006, ISSN: 00027863. DOI: 10.1021/ja062842p.
- [98] M. E. Rudbeck, S. O. Nilsson Lill, and A. Barth, "Influence of the molecular environment on phosphorylated amino acid models: A density functional theory study," *Journal of Physical Chemistry B*, vol. 116, no. 9, pp. 2751–2757, 2012, ISSN: 15205207. DOI: 10.1021/jp206414d.
- [99] D. Liu, T. Wyttenbach, P. E. Barran, and M. T. Bowers, "Sequential hydration of small protonated peptides," *Journal of the American Chemical Society*, vol. 125, no. 28, pp. 8458–8464, Jul. 2003, ISSN: 00027863. DOI: 10.1021/ja034638x.
- [100] Y. Kim, J. Kim, K. Y. Baek, and W. Y. Kim, "Efficient structural elucidation of microhydrated biomolecules through the interrogation of hydrogen bond networks," *Physical Chemistry Chemical Physics*, vol. 20, no. 12, pp. 8185–8191, 2018, ISSN: 14639076. DOI: 10.1039/c7cp08372f.
- [101] F. Calvo and J. Douady, "Stepwise hydration and evaporation of adenosine monophosphate nucleotide anions: a multiscale theoretical study," *Physical Chemistry Chemical Physics*, vol. 12, no. 14, p. 3404, 2010, ISSN: 1463-9076. DOI: 10.1039/b923972c.
- [102] B. Gao, T. Wyttenbach, and M. T. Bowers, "Hydration of protonated aromatic amino acids: Phenylalanine, tryptophan, and tyrosine," *Journal of the American Chemical Society*, vol. 131, no. 13, pp. 4695–4701, Apr. 2009, ISSN: 00027863. DOI: 10.1021/ja8085017.
- [103] B. Gao, T. Wyttenbach, and M. T. Bowers, "Protonated arginine and protonated lysine: Hydration and its effect on the stability of salt-bridge structures," *Journal of Physical Chemistry B*, vol. 113, no. 29, pp. 9995–10 000, Jul. 2009, ISSN: 15206106. DOI: 10.1021/jp903307h.
- [104] J. A. Silveira, K. A. Servage, C. M. Gamage, and D. H. Russell, "Cryogenic ion mobility-mass spectrometry captures hydrated ions produced during electrospray ionization," *Journal of Physical Chemistry A*, vol. 117, no. 5, pp. 953–961, Feb. 2013, ISSN: 10895639. DOI: 10.1021/jp311278a.
- [105] J. Q. Searcy and J. B. Fenn, "Clustering of water on hydrated protons in a supersonic free jet expansion," *The Journal of Chemical Physics*, vol. 61, no. 12, pp. 5282–5288, Dec. 1974, ISSN: 0021-9606. DOI: 10.1063/1.1681876.

-
- [106] T. F. Magnera, D. E. David, and J. Michl, “The first twenty-eight gas-phase proton hydration energies,” *Chemical Physics Letters*, vol. 182, no. 3-4, pp. 363–370, Aug. 1991, ISSN: 00092614. DOI: 10.1016/0009-2614(91)80230-U.
- [107] D. Scuderi, J. M. Bakker, S. Durand, *et al.*, “Structure of singly hydrated, protonated phospho-tyrosine,” *International Journal of Mass Spectrometry*, vol. 308, no. 2-3, pp. 338–347, 2011, ISSN: 13873806. DOI: 10.1016/j.ijms.2011.08.031.
- [108] C. Møller and M. S. Plesset, “Note on an Approximation Treatment for Many-Electron Systems,” *Physical Review*, vol. 46, no. 7, pp. 618–622, Oct. 1934, ISSN: 0031-899X. DOI: 10.1103/PhysRev.46.618.
- [109] P.-O. Löwdin, “Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction,” *Physical Review*, vol. 97, no. 6, pp. 1474–1489, Mar. 1955, ISSN: 0031-899X. DOI: 10.1103/PhysRev.97.1474.
- [110] F. Coester and H. Kümmel, “Short-range correlations in nuclear wave functions,” *Nuclear Physics*, vol. 17, no. C, pp. 477–485, Jun. 1960, ISSN: 00295582. DOI: 10.1016/0029-5582(60)90140-1.
- [111] J. Čížek, “On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods,” *The Journal of Chemical Physics*, vol. 45, no. 11, pp. 4256–4266, Dec. 1966, ISSN: 0021-9606. DOI: 10.1063/1.1727484.
- [112] R. J. Bartlett and G. D. Purvis, “Many-body perturbation theory, coupled-pair many-electron theory, and the importance of quadruple excitations for the correlation problem,” *International Journal of Quantum Chemistry*, vol. 14, no. 5, pp. 561–581, Nov. 1978, ISSN: 0020-7608. DOI: 10.1002/qua.560140504.
- [113] G. Gopakumar, P. H. Svensson, O. Grånäs, *et al.*, “X-ray Induced Fragmentation of Protonated Cystine,” *Journal of Physical Chemistry A*, vol. 126, no. 9, pp. 1496–1503, 2022, ISSN: 15205215. DOI: 10.1021/acs.jpca.1c10158.
- [114] M. N. Blom, I. Compagnon, N. C. Polfer, *et al.*, “Stepwise solvation of an amino acid: The appearance of zwitterionic structures,” *Journal of Physical Chemistry A*, vol. 111, no. 31, pp. 7309–7316, 2007, ISSN: 10895639. DOI: 10.1021/jp070211r.

- [115] S. L. Sherman, K. C. Fischer, and E. Garand, "Effects of Methyl Side Chains on the Microsolvation Structure of Protonated Tripeptides," *The Journal of Physical Chemistry A*, vol. 127, no. 30, pp. 6275–6281, Aug. 2023, ISSN: 1089-5639. DOI: 10.1021/acs.jpca.3c03711.
- [116] D. J. Miller and J. M. Lisy, "Hydration of ion-biomolecule complexes: Ab initio calculations and gas-phase vibrational spectroscopy of $K^+ + (\text{indole})_m (\text{H}_2\text{O})_n$," *Journal of Chemical Physics*, vol. 124, no. 18, 2006, ISSN: 00219606. DOI: 10.1063/1.2191047.
- [117] J. Houmøller, M. Wanko, A. Rubio, and S. B. Nielsen, "Effect of a Single Water Molecule on the Electronic Absorption by o- and p-Nitrophenolate: A Shift to the Red or to the Blue?" *Journal of Physical Chemistry A*, vol. 119, no. 47, pp. 11498–11503, 2015, ISSN: 15205215. DOI: 10.1021/acs.jpca.5b08634.
- [118] J. Boschmans, S. Jacobs, J. P. Williams, *et al.*, "Combining density functional theory (DFT) and collision cross-section (CCS) calculations to analyze the gas-phase behaviour of small molecules and their protonation site isomers," *Analyst*, vol. 141, no. 13, pp. 4044–4054, 2016, ISSN: 13645528. DOI: 10.1039/c5an02456k.
- [119] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Physical Review*, vol. 136, no. 3B, B864–B871, Nov. 1964, ISSN: 0031-899X. DOI: 10.1103/PhysRev.136.B864.
- [120] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Physical Review*, vol. 140, no. 4A, A1133–A1138, Nov. 1965, ISSN: 0031-899X. DOI: 10.1103/PhysRev.140.A1133.
- [121] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields," *The Journal of Physical Chemistry*, vol. 98, no. 45, pp. 11623–11627, Nov. 1994, ISSN: 0022-3654. DOI: 10.1021/j100096a001.
- [122] A. D. Becke, "Hartree–Fock exchange energy of an inhomogeneous electron gas," *International Journal of Quantum Chemistry*, vol. 23, no. 6, pp. 1915–1922, 1983, ISSN: 1097461X. DOI: 10.1002/qua.560230605.
- [123] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A*, vol. 38, no. 6, pp. 3098–3100, Sep. 1988, ISSN: 0556-2791. DOI: 10.1103/PhysRevA.38.3098.

- [124] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, no. 2, pp. 785–789, Jan. 1988, ISSN: 0163-1829. DOI: 10.1103/PhysRevB.37.785.
- [125] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *Canadian Journal of Physics*, vol. 58, no. 8, pp. 1200–1211, 1980, ISSN: 0008-4204. DOI: 10.1139/p80-159.
- [126] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, Apr. 1993, ISSN: 0021-9606. DOI: 10.1063/1.464913.
- [127] F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, "The ORCA quantum chemistry program package," *The Journal of Chemical Physics*, vol. 152, no. 22, p. 224 108, Jun. 2020, ISSN: 0021-9606. DOI: 10.1063/5.0004608.
- [128] F. Neese, F. Wennmohs, A. Hansen, and U. Becker, "Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange," *Chemical Physics*, vol. 356, no. 1-3, pp. 98–109, 2009, ISSN: 03010104. DOI: 10.1016/j.chemphys.2008.10.036.
- [129] D. Maganas, S. DeBeer, and F. Neese, "A Restricted Open Configuration Interaction with Singles Method to Calculate Valence-to-Core Resonant X-ray Emission Spectra: A Case Study," *Inorganic Chemistry*, vol. 56, no. 19, pp. 11 819–11 836, 2017, ISSN: 1520510X. DOI: 10.1021/acs.inorgchem.7b01810.
- [130] M. Roemelt, D. Maganas, S. Debeer, and F. Neese, "A combined DFT and restricted open-shell configuration interaction method including spin-orbit coupling: Application to transition metal L-edge X-ray absorption spectroscopy," *Journal of Chemical Physics*, vol. 138, no. 20, 2013, ISSN: 00219606. DOI: 10.1063/1.4804607.
- [131] M. Roemelt and F. Neese, "Excited states of large open-shell molecules: An efficient, general, and spin-adapted approach based on a restricted open-shell ground state wave function," *Journal of Physical Chemistry A*, vol. 117, no. 14, pp. 3069–3083, 2013, ISSN: 10895639. DOI: 10.1021/jp3126126.
- [132] D. P. Craig and T. Thirunamachandran, *Molecular Quantum Electrodynamics* (Dover Books on Chemistry). Dover Publications, 2012, ISBN: 9780486135632.

- [133] S. DeBeer George, T. Petrenko, and F. Neese, "Time-dependent density functional calculations of ligand K-edge X-ray absorption spectra," *Inorganica Chimica Acta*, vol. 361, no. 4, pp. 965–972, 2008, ISSN: 00201693. DOI: 10.1016/j.ica.2007.05.046.
- [134] S. Bari, L. Inhester, K. Schubert, *et al.*, "Inner-shell X-ray absorption spectra of the cationic series $NH^y +$ ($y = 0-3$)," *Physical Chemistry Chemical Physics*, vol. 21, no. 30, pp. 16 505–16 514, 2019, ISSN: 1463-9076. DOI: 10.1039/C9CP02864A.
- [135] L. I. Vogt, J. J. H. Cotelesage, N. V. Dolgova, *et al.*, "Sulfur X-ray Absorption and Emission Spectroscopy of Organic Sulfones," *The Journal of Physical Chemistry A*, vol. 127, no. 16, pp. 3692–3704, Apr. 2023, ISSN: 1089-5639. DOI: 10.1021/acs.jpca.2c08647.
- [136] N. A. Besley, "Density Functional Theory Based Methods for the Calculation of X-ray Spectroscopy," *Accounts of Chemical Research*, vol. 53, no. 7, pp. 1306–1315, 2020, ISSN: 15204898. DOI: 10.1021/acs.accounts.0c00171.
- [137] Y. J. Chiang, W. C. Huang, C. K. Ni, C. L. Liu, C. C. Tsai, and W. P. Hu, "NEXAFS spectra and specific dissociation of oligo-peptide model molecules," *AIP Advances*, vol. 9, no. 8, Aug. 2019, ISSN: 21583226. DOI: 10.1063/1.5112151.
- [138] G. K. Wertheim, M. A. Butler, K. W. West, and D. N. Buchanan, "Determination of the Gaussian and Lorentzian content of experimental line shapes," *Review of Scientific Instruments*, vol. 45, no. 11, pp. 1369–1371, 1974, ISSN: 00346748. DOI: 10.1063/1.1686503.
- [139] C. Nicolas and C. Miron, "Lifetime broadening of core-excited and -ionized states," *Journal of Electron Spectroscopy and Related Phenomena*, vol. 185, no. 8-9, pp. 267–272, 2012, ISSN: 03682048. DOI: 10.1016/j.elspec.2012.05.008.
- [140] P. Thompson, D. E. Cox, and J. B. Hastings, "Rietveld refinement of Debye-Scherrer synchrotron X-ray data from Al_2O_3 ," *Journal of Applied Crystallography*, vol. 20, no. 2, pp. 79–83, Apr. 1987, ISSN: 0021-8898. DOI: 10.1107/S0021889887087090.
- [141] G. Schmelzeisen-Redeker, L. Bütfering, and F. W. Röllgen, "Desolvation of ions and molecules in thermospray mass spectrometry," *International Journal of Mass Spectrometry and Ion Processes*, vol. 90, no. 2, pp. 139–150, 1989, ISSN: 01681176. DOI: 10.1016/0168-1176(89)85004-9.

-
- [142] S. Nguyen and J. B. Fenn, “Gas-phase ions of solute species from charged droplets of solutions,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 4, pp. 1111–1117, 2007, ISSN: 00278424. DOI: 10.1073/pnas.0609969104.
- [143] J. V. Iribarne and B. A. Thomson, “On the evaporation of small ions from charged droplets,” *The Journal of Chemical Physics*, vol. 64, no. 6, pp. 2287–2294, 1976, ISSN: 00219606. DOI: 10.1063/1.432536.
- [144] S. Banerjee and S. Mazumdar, “Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte,” *International Journal of Analytical Chemistry*, vol. 2012, pp. 1–40, 2012, ISSN: 1687-8760. DOI: 10.1155/2012/282574.
- [145] P. E. Miller and M. B. Denton, “The quadrupole mass filter: Basic operating concepts,” *Journal of Chemical Education*, vol. 63, no. 7, pp. 617–622, 1986, ISSN: 00219584. DOI: 10.1021/ed063p617.
- [146] W. Paul, “Electromagnetic Traps for Charged and Neutral Particles (Nobel Lecture),” *Angewandte Chemie International Edition in English*, vol. 29, no. 7, pp. 739–748, Jul. 1990, ISSN: 0570-0833. DOI: 10.1002/anie.199007391.
- [147] R. E. March, “An Introduction to Quadrupole Ion Trap Mass Spectrometry,” *Journal of Mass Spectrometry*, vol. 32, no. 4, pp. 351–369, Apr. 1997, ISSN: 1076-5174. DOI: 10.1002/(SICI)1096-9888(199704)32:4<351::AID-JMS512>3.0.CO;2-Y.
- [148] G. Z. Li, S. Guan, and A. G. Marshall, “Comparison of equilibrium ion density distribution and trapping force in Penning, Paul, and combined ion traps,” *Journal of the American Society for Mass Spectrometry*, vol. 9, no. 5, pp. 473–481, 1998, ISSN: 10440305. DOI: 10.1016/S1044-0305(98)00005-1.
- [149] D. A. Dahl, “simion for the personal computer in reflection,” *International Journal of Mass Spectrometry*, vol. 200, no. 1-3, pp. 3–25, Dec. 2000, ISSN: 13873806. DOI: 10.1016/S1387-3806(00)00305-5.
- [150] J. Viefhaus, F. Scholz, S. Deinert, *et al.*, “The Variable Polarization XUV Beamline P04 at PETRA III: Optics, mechanics and their performance,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 710, pp. 151–154, May 2013, ISSN: 01689002. DOI: 10.1016/j.nima.2012.10.110.

- [151] V. Q. Nguyen, X. G. Chen, and A. L. Yergey, "Observations of magic numbers in gas phase hydrates of alkylammonium ions," *Journal of the American Society for Mass Spectrometry*, vol. 8, no. 11, pp. 1175–1179, 1997, ISSN: 10440305. DOI: 10.1016/S1044-0305(97)00188-8.
- [152] A. R. Milosavljević, V. Z. Cerovski, F. Canon, *et al.*, "Energy-dependent UV photodissociation of gas-phase adenosine monophosphate nucleotide ions: The role of a single solvent molecule," *Journal of Physical Chemistry Letters*, vol. 5, no. 11, pp. 1994–1999, Jun. 2014, ISSN: 19487185. DOI: 10.1021/jz500696b.
- [153] A. Zviagin, V. Kopysov, and O. V. Boyarkin, "Gentle nano-electrospray ion source for reliable and efficient generation of microsolvated ions," *Review of Scientific Instruments*, vol. 93, no. 11, pp. 1–6, 2022, ISSN: 10897623. DOI: 10.1063/5.0119580.
- [154] B. M. Marsh, J. M. Voss, and E. Garand, "A dual cryogenic ion trap spectrometer for the formation and characterization of solvated ionic clusters," *Journal of Chemical Physics*, vol. 143, no. 20, 2015, ISSN: 10897690. DOI: 10.1063/1.4936360.
- [155] D. Liu, T. Wyttenbach, and M. T. Bowers, "Hydration of mononucleotides," *Journal of the American Chemical Society*, vol. 128, no. 47, pp. 15 155–15 163, Nov. 2006, ISSN: 00027863. DOI: 10.1021/ja062418o.
- [156] A. T. Blades, Y. Ho, and P. Kebarle, "Hydration in the gas phase of the orthophosphate anion, (HO)₂PO₂⁻, and the conversion of the orthophosphate to the metaphosphate, PO₃⁻, ion," *Journal of the American Chemical Society*, vol. 118, no. 1, pp. 196–201, 1996, ISSN: 00027863. DOI: 10.1021/ja952169w.
- [157] M. F. Peintinger, D. V. Oliveira, and T. Bredow, "Consistent Gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations," *Journal of Computational Chemistry*, vol. 34, no. 6, pp. 451–459, 2013, ISSN: 01928651. DOI: 10.1002/jcc.23153.
- [158] E. Van Lenthe, J. G. Snijders, and E. J. Baerends, "The zero-order regular approximation for relativistic effects: The effect of spin-orbit coupling in closed shell molecules," *Journal of Chemical Physics*, vol. 105, no. 15, pp. 6505–6516, 1996, ISSN: 00219606. DOI: 10.1063/1.472460.
- [159] P. J. Lestrange, P. D. Nguyen, and X. Li, "Calibration of Energy-Specific TDDFT for Modeling K-edge XAS Spectra of Light Elements," *Journal of Chemical Theory and Computation*, vol. 11, no. 7, pp. 2994–2999, 2015, ISSN: 15499626. DOI: 10.1021/acs.jctc.5b00169.

- [160] H. Wincel, "Microhydration of Protonated Nucleic Acid Bases and Protonated Nucleosides in the Gas Phase," *Journal of the American Society for Mass Spectrometry*, vol. 20, no. 10, pp. 1900–1905, 2009, ISSN: 10440305. DOI: 10.1016/j.jasms.2009.06.017.
- [161] K. Akamatsu and A. Yokoya, "X-ray absorption near edge structures of DNA or its components around the oxygen K-shell edge," *Radiation Research*, vol. 155, no. 3, pp. 449–452, 2001, ISSN: 00337587. DOI: 10.1667/0033-7587(2001)155[0449:XRANES]2.0.CO;2.
- [162] K. Fujii, K. Akamatsu, Y. Muramatsu, and A. Yokoya, "X-ray absorption near edge structure of DNA bases around oxygen and nitrogen K-edge," *Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms*, vol. 199, pp. 249–254, 2003, ISSN: 0168583X. DOI: 10.1016/S0168-583X(02)01538-0.
- [163] J. M. Voss, K. C. Fischer, and E. Garand, "Revealing the structure of isolated peptides: IR-IR predissociation spectroscopy of protonated triglycine isomers," *Journal of Molecular Spectroscopy*, vol. 347, pp. 28–34, May 2018, ISSN: 1096083X. DOI: 10.1016/j.jms.2018.03.006.
- [164] K. C. Fischer, S. L. Sherman, and E. Garand, "Competition between Solvation and Intramolecular Hydrogen-Bonding in Microsolvated Protonated Glycine and β -Alanine," *The Journal of Physical Chemistry A*, vol. 124, no. 8, pp. 1593–1602, Feb. 2020, ISSN: 1089-5639. DOI: 10.1021/acs.jpca.9b11977.
- [165] P. Leinweber, J. Kruse, F. L. Walley, *et al.*, "Nitrogen K-edge XANES - An overview of reference compounds used to identify 'unknown' organic nitrogen in environmental samples," *Journal of Synchrotron Radiation*, vol. 14, no. 6, pp. 500–511, Oct. 2007, ISSN: 09090495. DOI: 10.1107/S0909049507042513.
- [166] E. Otero and S. G. Urquhart, "Nitrogen 1s near-edge X-ray absorption fine structure spectroscopy of amino acids: Resolving zwitterionic effects," *Journal of Physical Chemistry A*, vol. 110, no. 44, pp. 12 121–12 128, 2006, ISSN: 10895639. DOI: 10.1021/jp064082a.
- [167] S. Kim, E. E. Bolton, and S. H. Bryant, "PubChem3D: conformer ensemble accuracy," *Journal of Cheminformatics*, vol. 5, no. 1, p. 1, Dec. 2013, ISSN: 1758-2946. DOI: 10.1186/1758-2946-5-1.
- [168] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, Feb. 1996, ISSN: 02637855. DOI: 10.1016/0263-7855(96)00018-5.

- [169] J. Eargle, D. Wright, and Z. Luthey-Schulten, “Multiple Alignment of protein structures and sequences for VMD,” *Bioinformatics*, vol. 22, no. 4, pp. 504–506, Feb. 2006, ISSN: 1367-4811. DOI: 10.1093/bioinformatics/bti825.
- [170] E. F. Pettersen, T. D. Goddard, C. C. Huang, *et al.*, “UCSF Chimera - A visualization system for exploratory research and analysis,” *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004, ISSN: 01928651. DOI: 10.1002/jcc.20084.
- [171] J. J. Stewart, “MOPAC: A semiempirical molecular orbital program,” *Journal of Computer-Aided Molecular Design*, vol. 4, no. 1, pp. 1–103, 1990, ISSN: 0920654X. DOI: 10.1007/BF00128336.
- [172] Y. Imamura and H. Nakai, “Time-dependent density functional theory (TDDFT) calculations for core-excited states: Assessment of an exchange functional combining the Becke88 and van Leeuwen-Baerends-type functionals,” *Chemical Physics Letters*, vol. 419, no. 1-3, pp. 297–303, 2006, ISSN: 00092614. DOI: 10.1016/j.cplett.2005.11.084.
- [173] A. Chantzis, J. K. Kowalska, D. Maganas, S. Debeer, and F. Neese, “Ab Initio Wave Function-Based Determination of Element Specific Shifts for the Efficient Calculation of X-ray Absorption Spectra of Main Group Elements and First Row Transition Metals,” *Journal of Chemical Theory and Computation*, vol. 14, no. 7, pp. 3686–3702, 2018, ISSN: 15499626. DOI: 10.1021/acs.jctc.8b00249.
- [174] P. J. Boersema, S. Mohammed, and A. J. Heck, “Phosphopeptide fragmentation and analysis by mass spectrometry,” *Journal of Mass Spectrometry*, vol. 44, no. 6, pp. 861–878, 2009, ISSN: 10765174. DOI: 10.1002/jms.1599.
- [175] P. Zhang, W. Chan, I. L. Ang, *et al.*, “Revisiting Fragmentation Reactions of Protonated α -Amino Acids by High-Resolution Electrospray Ionization Tandem Mass Spectrometry with Collision-Induced Dissociation,” *Scientific Reports*, vol. 9, no. 1, p. 6453, Apr. 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-42777-8.
- [176] J. Li and X. Zhan, “Mass spectrometry analysis of phosphotyrosine-containing proteins,” *Mass Spectrometry Reviews*, no. January 2022, pp. 1–31, 2023, ISSN: 10982787. DOI: 10.1002/mas.21836.
- [177] W. Zhang, V. Carravetta, O. Plekan, *et al.*, “Electronic structure of aromatic amino acids studied by soft x-ray spectroscopy,” *The Journal of Chemical Physics*, vol. 131, no. 3, Jul. 2009, ISSN: 0021-9606. DOI: 10.1063/1.3168393.

- [178] R. Püttner, C. Kolczewski, M. Martins, *et al.*, “The C 1s NEXAFS spectrum of benzene below threshold: Rydberg or valence character of the unoccupied r-type orbitals,” *Chemical Physics Letters*, vol. 393, pp. 361–366, 2004. DOI: <https://doi.org/10.1016/j.cplett.2004.06.053>.
- [179] W. H. Schwarz, T. C. Chang, U. Seeger, and K. H. Hwang, “Core excitations of symmetrical aromatic molecules. Specific correlations in the valence shell and localization in the core shells,” *Chemical Physics*, vol. 117, no. 1, pp. 73–89, 1987, ISSN: 03010104. DOI: 10.1016/0301-0104(87)80098-8.
- [180] E. E. Rennie, B. Kempgens, H. M. Köppe, *et al.*, “A comprehensive photoabsorption, photoionization, and shake-up excitation study of the C 1s cross section of benzene,” *The Journal of Chemical Physics*, vol. 113, no. 17, pp. 7362–7375, 2000, ISSN: 0021-9606. DOI: 10.1063/1.1290029.
- [181] R. M. Petoral and K. Uvdal, “XPS and NEXAFS study of tyrosine-terminated propanethiol assembled on gold,” *Journal of Electron Spectroscopy and Related Phenomena*, vol. 128, no. 2-3, pp. 159–164, Feb. 2003, ISSN: 03682048. DOI: 10.1016/S0368-2048(02)00271-2.
- [182] K. R. Wilson, B. S. Rude, T. Catalano, *et al.*, “X-ray Spectroscopy of Liquid Water Microjets,” *The Journal of Physical Chemistry B*, vol. 105, no. 17, pp. 3346–3349, May 2001, ISSN: 1520-6106. DOI: 10.1021/jp010132u.
- [183] J. D. Ward, M. Bowden, C. T. Resch, *et al.*, “Identification of Uranyl Minerals Using Oxygen K-Edge X-Ray Absorption Spectroscopy,” *Geostandards and Geoanalytical Research*, vol. 40, no. 1, pp. 135–148, Mar. 2016, ISSN: 1639-4488. DOI: 10.1111/j.1751-908X.2015.00337.x.
- [184] M. N. Piancastelli, A. Hempelmann, F. Heiser, O. Gessner, A. Rüdél, and U. Becker, “Resonant photofragmentation of water at the oxygen K edge by high-resolution ion-yield spectroscopy,” *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 59, no. 1, pp. 300–306, 1999, ISSN: 10941622. DOI: 10.1103/PhysRevA.59.300.
- [185] M. T. Nguyen, S. Creve, and L. G. Vanquickenborne, “Properties of phosphorus compounds by density functional theory: CH₃P species as a test case,” *Journal of Chemical Physics*, vol. 105, no. 5, pp. 1922–1932, 1996, ISSN: 00219606. DOI: 10.1063/1.472062.
- [186] J. B. Pendry, “Reliability factors for LEED calculations,” *Journal of Physics C: Solid State Physics*, vol. 13, no. 5, pp. 937–944, Feb. 1980, ISSN: 0022-3719. DOI: 10.1088/0022-3719/13/5/024.

- [187] J. Leroux, A. Kotobi, K. Hirsch, *et al.*, “Mapping the electronic transitions of protonation sites in peptides using soft X-ray action spectroscopy,” *Physical Chemistry Chemical Physics*, vol. 25, no. 37, pp. 25 603–25 618, 2023, ISSN: 1463-9076. DOI: 10.1039/D3CP02524A.
- [188] K. Zhang, C. J. Cassady, and A. Chung-Phillips, “Ab Initio Studies of Neutral and Protonated Triglycines: Comparison of Calculated and Experimental Gas-Phase Basicity,” *Journal of the American Chemical Society*, vol. 116, no. 25, pp. 11 512–11 521, Dec. 1994, ISSN: 0002-7863. DOI: 10.1021/ja00104a035.
- [189] K. Zhang, D. M. Zimmerman, A. Chung-Phillips, and C. J. Cassady, “Experimental and ab initio studies of the gas-phase basicities of polyglycines,” *Journal of the American Chemical Society*, vol. 115, no. 23, pp. 10 812–10 822, Nov. 1993, ISSN: 0002-7863. DOI: 10.1021/ja00076a044.
- [190] J. Wu and C. B. Lebrilla, “Gas-Phase Basicities and Sites of Protonation of Glycine Oligomers (GLY_n; n = 1–5),” *Journal of the American Chemical Society*, vol. 115, no. 8, pp. 3270–3275, 1993, ISSN: 15205126. DOI: 10.1021/ja00061a027.
- [191] Z. Wu and C. Fenselau, “Proton affinity of arginine measured by the kinetic approach,” *Rapid Communications in Mass Spectrometry*, vol. 6, no. 6, pp. 403–405, 1992, ISSN: 10970231. DOI: 10.1002/rcm.1290060610.
- [192] C. Bleiholder, S. Suhai, and B. Paizs, “Revising the Proton Affinity Scale of the Naturally Occurring α -Amino Acids,” *Journal of the American Society for Mass Spectrometry*, vol. 17, no. 9, pp. 1275–1281, 2006, ISSN: 10440305. DOI: 10.1016/j.jasms.2006.05.010.
- [193] A. Chung-Phillips, “Polyglycine conformational analysis: Calculated vs experimental gas-phase basicities and proton affinities,” *Journal of Physical Chemistry A*, vol. 109, no. 26, pp. 5917–5932, 2005, ISSN: 10895639. DOI: 10.1021/jp040758v.
- [194] R. Wu and T. B. McMahon, “Protonation sites and conformations of peptides of glycine (Gly 1-5H⁺) by IRMPD spectroscopy,” *Journal of Physical Chemistry B*, vol. 113, no. 25, pp. 8767–8775, Jun. 2009, ISSN: 15206106. DOI: 10.1021/jp811468q.
- [195] H. Li, J. Jiang, and Y. Luo, “Identification of the protonation site of gaseous triglycine: the cis-peptide bond conformation as the global minimum,” *Physical Chemistry Chemical Physics*, vol. 19, no. 23, pp. 15 030–15 038, 2017, ISSN: 1463-9076. DOI: 10.1039/C7CP01997A.

- [196] P. T. Edwards, L. K. Saunders, D. C. Grinter, *et al.*, “Determination of H-Atom Positions in Organic Crystal Structures by NEXAFS Combined with Density Functional Theory: a Study of Two-Component Systems Containing Isonicotinamide,” *The Journal of Physical Chemistry A*, vol. 126, no. 19, pp. 2889–2898, May 2022, ISSN: 1089-5639. DOI: 10.1021/acs.jpca.2c00439.
- [197] S. Eckert, P. S. Miedema, W. Quevedo, *et al.*, “Molecular structures and protonation state of 2-Mercaptopyridine in aqueous solution,” *Chemical Physics Letters*, vol. 647, pp. 103–106, 2016, ISSN: 00092614. DOI: 10.1016/j.cplett.2016.01.050.
- [198] X. Wang, S. Rathnachalam, K. Bijlsma, *et al.*, “Site-selective soft X-ray absorption as a tool to study protonation and electronic structure of gas-phase DNA,” *Physical Chemistry Chemical Physics*, vol. 23, no. 20, pp. 11 900–11 906, May 2021, ISSN: 14639076. DOI: 10.1039/d1cp01014j.
- [199] B. M. Messer, C. D. Cappa, J. D. Smith, *et al.*, “pH dependence of the electronic structure of glycine,” *Journal of Physical Chemistry B*, vol. 109, no. 11, pp. 5375–5382, Mar. 2005, ISSN: 15206106. DOI: 10.1021/jp0457592.
- [200] O. Plekan, V. Feyer, R. Richter, *et al.*, “Investigation of the amino acids glycine, proline, and methionine by photoemission spectroscopy,” *Journal of Physical Chemistry A*, vol. 111, no. 43, pp. 10 998–11 005, Nov. 2007, ISSN: 10895639. DOI: 10.1021/jp075384v.
- [201] R. Wu and T. B. McMahon, “Infrared multiple-photon dissociation mechanisms of peptides of glycine,” *Chemistry - A European Journal*, vol. 14, no. 26, pp. 7765–7770, 2008, ISSN: 09476539. DOI: 10.1002/chem.200800594.
- [202] V. Feyer, O. Plekan, R. Richtery, M. Coreno, K. C. Prince, and V. Carravetta, “Photoemission and photoabsorption spectroscopy of glycyl-glycine in the gas phase,” *Journal of Physical Chemistry A*, vol. 113, no. 40, pp. 10 726–10 733, Oct. 2009, ISSN: 10895639. DOI: 10.1021/jp906843j.
- [203] A. R. Milosavljević, A. Giuliani, and C. Nicolas, “Gas-Phase Near-Edge X-Ray Absorption Fine Structure (NEXAFS) Spectroscopy of Nanoparticles, Biopolymers, and Ionic Species,” in *X-ray and Neutron Techniques for Nanomaterials Characterization*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 451–505, ISBN: 9783662486061. DOI: 10.1007/978-3-662-48606-1_8.

- [204] K. Schubert, M. Guo, K. Atak, *et al.*, “The electronic structure and deexcitation pathways of an isolated metalloporphyrin ion resolved by metal L-edge spectroscopy,” *Chemical Science*, vol. 12, no. 11, pp. 3966–3976, Mar. 2021, ISSN: 20416539. DOI: 10.1039/d0sc06591a.
- [205] S. Dörner, L. Schwob, K. Schubert, *et al.*, “The influence of the methionine residue on the dissociation mechanisms of photoionized methionine-enkephalin probed by VUV action spectroscopy,” *European Physical Journal D*, vol. 75, no. 4, Apr. 2021, ISSN: 14346079. DOI: 10.1140/epjd/s10053-021-00147-y.
- [206] Y. S. Lin, C. C. Tsai, H. R. Lin, *et al.*, “Highly Selective Dissociation of a Peptide Bond Following Excitation of Core Electrons,” *Journal of Physical Chemistry A*, vol. 119, no. 24, pp. 6195–6202, 2015, ISSN: 15205215. DOI: 10.1021/acs.jpca.5b04415.
- [207] I. Ljubić, A. Kivimäki, and M. Coreno, “An experimental NEXAFS and computational TDDFT and Δ dFT study of the gas-phase core excitation spectra of nitroxide free radical TEMPO and its analogues,” *Physical Chemistry Chemical Physics*, vol. 18, no. 15, pp. 10 207–10 217, 2016, ISSN: 14639076. DOI: 10.1039/c6cp00490c.
- [208] D. Maksimov and M. Rossi, *GenSec: Generate and Search*. Nov. 2021. DOI: 10.5281/ZENODO.5651513.
- [209] D. Maksimov, “Characterization and prediction of peptide structures on inorganic surfaces,” Ph.D. dissertation, EPFL, 2022. DOI: 10.5075/epfl-thesis-9740.
- [210] C. Adamo and V. Barone, “Toward reliable density functional methods without adjustable parameters: The PBE0 model,” *Journal of Chemical Physics*, vol. 110, no. 13, pp. 6158–6170, 1999, ISSN: 00219606. DOI: 10.1063/1.478522.
- [211] A. Tkatchenko and M. Scheffler, “Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data,” *Physical Review Letters*, vol. 102, no. 7, pp. 6–9, 2009, ISSN: 00319007. DOI: 10.1103/PhysRevLett.102.073005.
- [212] V. Blum, R. Gehrke, F. Hanke, *et al.*, “Ab initio molecular simulations with numeric atom-centered orbitals,” *Computer Physics Communications*, vol. 180, no. 11, pp. 2175–2196, 2009, ISSN: 00104655. DOI: 10.1016/j.cpc.2009.06.022.

- [213] R. R. Hudgins, Y. Mao, M. A. Ratner, and M. F. Jarrold, "Conformations of Gly(n)H⁺ and Ala(n)H⁺ peptides in the gas phase," *Biophysical Journal*, vol. 76, no. 3, pp. 1591–1597, 1999, ISSN: 00063495. DOI: 10.1016/S0006-3495(99)77318-2.
- [214] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *Journal of Chemical Theory and Computation*, vol. 11, no. 8, pp. 3696–3713, 2015, ISSN: 15499626. DOI: 10.1021/acs.jctc.5b00255.
- [215] P. Gasparotto, R. H. Meißner, and M. Ceriotti, "Recognizing Local and Global Structural Motifs at the Atomic Scale," *Journal of Chemical Theory and Computation*, vol. 14, no. 2, pp. 486–498, 2018, ISSN: 15499626. DOI: 10.1021/acs.jctc.7b00993.
- [216] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, "Unsupervised Learning Methods for Molecular Simulation Data," *Chemical Reviews*, vol. 121, no. 16, pp. 9722–9758, 2021, ISSN: 15206890. DOI: 10.1021/acs.chemrev.0c01195.
- [217] A. Kotobi, L. Schwob, G. B. Vonbun-Feldbauer, *et al.*, "Reconstructing the infrared spectrum of a peptide from representative conformers of the full canonical ensemble," *Communications Chemistry*, vol. 6, no. 1, p. 46, Mar. 2023, ISSN: 2399-3669. DOI: 10.1038/s42004-023-00835-3.
- [218] R. A. Ingle, A. Banerjee, C. Bacellar, *et al.*, "Carbon K-edge x-ray emission spectroscopy of gas phase ethylenic molecules," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 55, no. 4, 2022, ISSN: 13616455. DOI: 10.1088/1361-6455/ac4e66.
- [219] A. F. Lago, L. H. Coutinho, R. R. Marinho, A. Naves De Brito, and G. G. De Souza, "Ionic dissociation of glycine, alanine, valine and proline as induced by VUV (21.21 eV) photons," *Chemical Physics*, vol. 307, no. 1, pp. 9–14, 2004, ISSN: 03010104. DOI: 10.1016/j.chemphys.2004.06.052.
- [220] J. R. Majer, B. I. Al-Ali, and A. S. Azzouz, "The mass spectrum of glycine," *Organic Mass Spectrometry*, vol. 16, no. 4, pp. 147–152, 1981, ISSN: 10969888. DOI: 10.1002/oms.1210160402.
- [221] I. A. Papayannopoulos, *The interpretation of collision-induced dissociation tandem mass spectra of peptides*, Jan. 1995. DOI: 10.1002/mas.1280140104.

- [222] R. Xiong, L. Xu, Y. Tang, M. Cao, and H. Li, “Identifying the protonation site and the scope of non-proline cis-peptide bond conformations: a first-principles study on protonated oligopeptides,” *Physical Chemistry Chemical Physics*, vol. 25, no. 20, pp. 13 989–13 998, 2023, ISSN: 14639076. DOI: 10.1039/d3cp00690e.
- [223] Y. Zubavichus, A. Shaporenko, M. Grunze, and M. Zharnikov, “NEXAFS Spectroscopy of Homopolypeptides at All Relevant Absorption Edges: Polyisoleucine, Polytyrosine, and Polyhistidine,” *The Journal of Physical Chemistry B*, vol. 111, no. 33, pp. 9803–9807, Aug. 2007, ISSN: 1520-6106. DOI: 10.1021/jp073922y.
- [224] J. Stewart-Ornstein, A. P. Hitchcock, D. H. Cruz, *et al.*, “Using intrinsic X-ray absorption spectral differences to identify and map peptides and proteins,” *Journal of Physical Chemistry B*, vol. 111, no. 26, pp. 7691–7699, Jul. 2007, ISSN: 15206106. DOI: 10.1021/jp0720993.
- [225] C. F. Rodriguez, A. Cunje, T. Shoeib, I. K. Chu, A. C. Hopkinson, and K. W. Siu, “Proton migration and tautomerism in protonated triglycine,” *Journal of the American Chemical Society*, vol. 123, no. 13, pp. 3006–3012, 2001, ISSN: 00027863. DOI: 10.1021/ja0015904.
- [226] A. Giuliani, A. R. Milosavljević, F. Canon, and L. Nahon, “Contribution of synchrotron radiation to photoactivation studies of biomolecular ions in the gas phase,” *Mass Spectrometry Reviews*, vol. 33, no. 6, pp. 424–441, Nov. 2014, ISSN: 10982787. DOI: 10.1002/mas.21398.
- [227] V. Ulrich, S. Barth, S. Joshi, *et al.*, “Giant Chiral Asymmetry in the C 1 s Core Level Photoemission from Randomly Oriented Fenchone Enantiomers,” *The Journal of Physical Chemistry A*, vol. 112, no. 16, pp. 3544–3549, Apr. 2008, ISSN: 1089-5639. DOI: 10.1021/jp709761u.

Acknowledgements

It still feels like yesterday when I arrived in Hamburg for my Master's thesis three and a half years ago. Now, I have reached the end of my student journey, and I have much (or not) to say about it. Especially about the past 3 years as a PhD student between Caen in France and Hamburg in Germany, graduating in Groningen! If you ask me how or why I ended up in such a program, my answer would still be "Don't ask" but I pretty much enjoyed it. This PhD thesis would have never been written, let alone finished, without the help and support of many people.

First, I would like to thank Melanie Schnell, Marc Rousseau, Ronnie Hoekstra, and Debora Scuderi for agreeing to be part of my Assessment committee and reviewing my thesis. Thank you, Debora, for accepting this role so late, and also, thank you for the phosphotyrosine idea; it is a nasty but exciting molecule.

I would like to thank my supervisors.

Sadia, thank you for all the support and guidance. These three years have not always been smooth but your positivity made things much easier. You never gave up on all the bureaucratic deadends that came with establishing this double PhD program with Caen. Most importantly, thank you for always being reachable, even on weekends when I needed advice.

Jean-Yves, first, thank you for taking over the supervision in France and navigating the bureaucracy of the agreement between the University of Caen and Groningen with such calm, especially when the different discussions stopped making sense. Thank you for letting me take over a few of your classes with the L1 and L2 so I could get experience in teaching during my PhD. Your feedback on my work and presentations was always constructive, and I thank you for traveling to Hamburg for our beamtime at PETRA and attending the defense in Groningen!

I would have never become the experimentalist I am without you, Lucas. From the time you explained to me how Spektrolatius and NEXAMS experiments worked when I joined the group as a master student to the night shifts at P04, these times have been memorable. I learned a lot from all our discussions. Thank you for sharing your passion for comic books with me. It's too bad Sadia never accepted our idea of

building a LEGO wall in the lab; it would have been amazing!

I never worked alone, but the hydration source and the experimental campaigns were a success thanks to the help of so many of you. I would like to thank Prof. Nonose for kindly answering my email about the functioning of the source and providing details on how to build it, as well as Bart and Torben for drawing and helping with the design of the final source. I also thank the teams at the Ion Trap Endstation at BESSY II and the P04 beamline of PETRA III for their help during the beamtimes. Additionally, I am grateful to Florian Trinter for his help during the P04 beamtimes and for reviewing the drafts of the papers that resulted from these experiments.

From the CIMAP lab in Caen, I would like to thank the PhD students for the nice coffee breaks and beer evenings when I was around. I want to thank the SIMUL team for accepting an experimentalist on their team; I enjoyed my time with you. A special thanks to Julie, without whom I would have probably never started doing DFT calculations. Thank you for always believing in me, even when I didn't believe in myself.

Doing a PhD would not have been the same without the FS-BIG team in Hamburg. Thank you all for your help during experiments and for the many coffee breaks or beers after work. I enjoyed our work together over the last few years, the beamtimes, the countless hours trying to optimize signals or fix broken parts of Spektrolatius, or, better, the water leaks! As Bart said, not without humor, I was much better at hydrating the setup than the biomolecules. Spektrolatius has never been so clean! Carlos, our discussions were never short but always passionate. Thanks for your patience in teaching me how to do calculations and for reviewing the theoretical chapter of my thesis. For all the things that made this experience about more than just work: late-night Mario Kart races during beamtimes, Corona shifts, cakes, coffee, escape games together, and everything I'm forgetting (did I mention coffee?).

I would also like to thank all my friends who made these three years unforgettable. Tristan, PE, Chloé, Florian, Sophie, and Thibault, studying would have been far less enjoyable without "les jeudis chez Tristan" or the long video calls as we scattered across Germany and France. Christian, thank you for being a great friend and an excellent German teacher—best of luck with your PhD in Grenoble! I also want to thank my friends from Hamburg—Thea, Sarah, Marine, George, and many others. Our climbing sessions and evenings together were always refreshing and much-needed breaks from work. I hope for many more of those great moments. A special thanks to Marine for reading part of my thesis and offering valuable advice.

I would like to thank my parents, Céline and Vincent, my siblings, Marine and Thomas, and my grandmothers, Monique and Geneviève, for their love and support. Merci à

tous pour votre soutien et d'avoir fait le déplacement jusqu'à Groningen pour ma soutenance. Un grand merci à Marine pour avoir réalisé la couverture de ma thèse! Robert, I never imagined I would be so lucky and share this whole adventure with you. Thanks for moving to Hamburg to be with me, and thank you for always being there.

Influence of nanohydration on the structure and radiation-induced fragmentation of gas-phase biomolecules

Abstract:

In the biological medium, water conditions the structure and function of biomolecules, with the first solvation shell playing a fundamental role in these properties. To understand this particular role of water, a bottom-up approach can be utilized, such as investigating the stepwise hydration of biomolecules in the gas phase. This thesis presents an experimental and theoretical study on the influence of nanohydration on the structure and physical processes of gas-phase biomolecules by irradiation with soft X-ray photons.

First, we described the experimental setup and the electrospray ionization source dedicated to producing nanohydrated biomolecules. In the second part, we studied the solvation of two model systems: deprotonated adenosine monophosphate (AMP) and protonated phosphorylated amino acid phosphotyrosine. Using quantum chemical calculations, we investigated the influence of stepwise solvation on the three-dimensional structure and spectroscopic characteristics of AMP. We demonstrated that a single water molecule already affects its geometry. Furthermore, we investigated the influence of a single water molecule on the structure and X-ray absorption spectra of phosphotyrosine. By combining soft X-ray spectroscopy with quantum chemical calculations, we could determine the location of the water molecule and demonstrate its features in the X-ray absorption spectrum at the oxygen K-edge. Lastly, we examined the electronic fingerprint of the protonation sites in model peptides using their X-ray spectra at the nitrogen K-edge. We mapped the differences in the electronic transitions caused by different proton locations at the nitrogen and oxygen sites present within the peptide backbone and amino acid side chains.

Key-words

Solvation, biomolecules, density-functional theory, soft X-ray spectroscopy, time-of-flight spectrometry, photoabsorption, electronic structure.

Influence de la nanohydratation sur la structure et la fragmentation radio-induite de biomolécules en phase gazeuse

Résumé

Dans le milieu biologique, l'eau conditionne la structure et la fonction des biomolécules, la première couche de solvation jouant un rôle fondamental dans ces propriétés. Pour comprendre ce rôle particulier de l'eau, une approche bottom-up peut être utilisée, comme l'étude de l'hydratation progressive des biomolécules en phase gazeuse. Cette thèse présente une étude expérimentale et théorique de l'influence de la nanohydratation sur la structure et les processus physiques des biomolécules en phase gazeuse par irradiation avec des photons X.

Dans un premier temps, nous avons décrit le dispositif expérimental et la source d'ionisation par électrospray dédiée à la production de biomolécules nanohydratées. Dans la deuxième partie, nous avons étudié la solvation de deux systèmes modèles : l'adénosine monophosphate (AMP) déprotonée et la phosphotyrosine protonée. À l'aide de calculs de chimie quantique, nous avons étudié l'influence de la solvation progressive sur la structure tridimensionnelle et les caractéristiques spectroscopiques de l'AMP. Nous avons démontré qu'une seule molécule d'eau affecte déjà sa géométrie. En outre, nous avons étudié l'influence d'une unique molécule d'eau sur la structure et les spectres d'absorption des rayons X de la phosphotyrosine. En combinant la spectroscopie à rayons X avec des calculs de chimie quantique, nous avons pu déterminer l'emplacement de la molécule d'eau et démontrer ses caractéristiques dans le spectre d'absorption des rayons X au seuil K de l'oxygène. Enfin, nous avons examiné l'empreinte électronique des sites de protonation dans des peptides modèles dans leurs spectres de rayons X au seuil K de l'azote.

Mots-clefs

Solvation, biomolécules, théorie de la fonctionnelle de la densité, spectroscopie à rayons X mous, spectrométrie de temps de vol, photoabsorption, structure électronique.
