Novel Approaches in Cold Ion Spectroscopy for Structural Characterization of Biomolecules in the Gas Phase

PAR

Vladimir KOPYSOV

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

Résumé

La spectrométrie de masse est une technique analytique puissante, qui est devenue inestimable dans plusieurs domaines scientifiques, tels que les science en «-omique» (p. ex., protéomique), le contrôle antidopage et les sciences judiciaires, le développement de médicaments, l'analyse environnementale, la recherche géologique, etc. Développé il y a plus d'un siècle pour mesurer des masses atomiques des éléments et les abondances naturelles des isotopes, aujourd'hui la spectrométrie de masse est un outil essentiel à caractériser la structure de biomolécules, allant des petits métabolites aux grands assemblages macromoléculaires. Cela inclut l'indentification de composés inconnus, la quantification de composés connus, et l'exploration de la structure moléculaire. Cependant, l'analyse par spectrométrie de masse se fonde sur la mesure du rapport masse sur charge d'ions, une quantité qui n'est pas sensible à leur structure tridimensionnelle (3D). La spectroscopie d'ions refroidis est une autre méthode d'étudier les ions en phase gazeuse, qui a prouvé sa capacité de révéler leurs empreintes spectroscopiques, qui sont caractéristiques de la structure 3D d'ions à un niveau fondamental. Dans cette thèse nous présentons des nouvelles approches à l'identification structurale de molécules, qui sont fondées sur l'intégration de la spectrométrie de masse (MS) de haute résolution avec la spectroscopie haute résolution de photofragmentation d'ions refroidis dans l'ultraviolet (UV).

La première partie de ce travail décrit le couplage d'un spectromètre de masse Orbitrap Exactive à un spectromètre d'ions refroidis et l'utilisation de ce nouvel instrument hybride pour mesurer des spectres bidimensionnels (2D) UV-MS d'espèces ioniques. La synergie des deux techniques complémentaires fait ces spectres les empreintes uniques d'ions biomoléculaire.

La deuxième partie présente une approche qui utilise l'analyse mathématique de ces empreintes très spécifiques pour l'identification de biomolécules isomériques dans des mélanges. En utilisant des bibliothèques d'empreintes provisoirement enregistrées,

1

l'approche d'UV-MS a été appliquée avec succès pour l'identification quantitative des isomères de position d'un peptide phosphorylé, des diastéréoisomères d'un peptide opioïde, et des diastéréoisomères d'un médicament. Nous démontrons également une approche sans bibliothèques, qui permet d'estimer le nombre de composantes présentes dans un mélange et d'obtenir leurs spectres de masse de fragmentation et spectres d'absorption UV directement du spectre 2D UV-MS du mélange.

Dans la troisième partie nous employons l'approche sans bibliothèques et le phénomène de la photofragmentation non-statistique pour identifier conformères d'ions biomoléculaires en phase gazeuse et récupérer leurs spectres d'absorption UV et spectres de masse individuels. Nous validons cette approche sur un dipeptide de référence, Tyr-Ala, et puis l'appliquons à un decapeptide, gramicidin S, et à un petit médicament, homatropine. Une analyse minutieuse des spectres d'absorption UV et de masse dérivés des conformères de gramicidin S et de homatropine fournit quelques renseignements structurels qui peuvent être utilisés pour résoudre leurs structures 3D.

Dans la dernière partie nous démontrons une utilisation de la spectroscopie UV-MS pour estimer les distances entre les chromophores phénylalanine et tyrosine dans quatre peptides opioïdes en phase gazeuse. Nous avons aussi calculé les structures 3D des leurs conformères de moindre énergie. Nous analyserons ces structures dans le cadre des distances interchromophore estimées et des propriétés pharmacologiques des peptides.

Mots-clés: spectroscopie d'ions refroidis, spectroscopie dans l'ultraviolet, spectrométrie de masse de haute résolution, photodissociation, Orbitrap, UV-MS, phase gazeuse, ions biomoléculaires, isomères, conformères, FRET, peptides phosphorylés, peptides opioïdes, éphédrine, gramicidin S, homatropine, leucine enképhaline.

Abstract

Mass spectrometry is a powerful analytical technique, which has become invaluable across a wide range of scientific fields, such as "omics" sciences (e.g., proteomics), doping control and forensic sciences, drug development, environmental analysis, geologic research, etc. Originally developed over a century ago to measure elemental atomic weights and the natural abundances of isotopes, nowadays mass spectrometry is a vital tool for structural characterization of biomolecules, ranging from small metabolites to large macromolecular assemblies. This includes identifying unknown compounds, quantifying known compounds, and exploring molecular structures. However, the mass spectrometric analysis is based on measurements of the mass-to-charge ratio of ions, a quantity that is not sensitive to their three-dimensional (3D) structure. Cold ion spectroscopy is another method of studying ions in the gas phase, that has demonstrated its capability to reveal their spectroscopic fingerprints, which are characteristic of the 3D structure of the ions on a fundamental level. In this thesis we present novel approaches to molecular structure identification, which are based on integration of high-resolution mass spectrometry (MS) with high-resolution ultraviolet (UV) photofragmentation spectroscopy of cold ions.

The first part of this work describes the coupling of an Exactive Orbitrap-based mass spectrometer to a cold ion spectrometer and the use of this novel hybrid instrument for measuring two-dimensional (2D) UV-MS spectra of ionic species. The synergy of the two complementary techniques makes these spectra unique fingerprints of biomolecular ions.

The second part presents an approach that uses mathematical analysis of these highly specific fingerprints for identification of isomeric biomolecules in mixtures. Using preliminarily recorded libraries of fingerprints, the UV-MS approach was successfully employed for quantitative identification of positional isomers of a phosphorylated peptide, diastereomers of an opioid peptide, and diastereomers of a drug molecule. We also demonstrate a library-free approach, which allows estimating the number of components present in a mixture and deriving their UV absorption and fragmentation mass spectra directly from the 2D UV-MS spectrum of the mixture.

In the third part, we employ the library-free approach and the phenomenon of nonstatistical photofragmentation to identify conformers of gas-phase biomolecular ions and to recover their individual UV absorption and fragmentation mass spectra. We validate this approach with a benchmark dipeptide, Tyr-Ala, and then apply it to a decapeptide, gramicidin S, and a small drug molecule, homatropine. A careful analysis of the derived UV and mass spectra of the conformers of gramicidin S and homatropine provides some structural information that can be used for solving their 3D structures.

In the last part, we demonstrate a use of 2D UV-MS spectroscopy for estimating the distances between the phenylalanine and tyrosine chromophores in four gas-phase opioid peptides. We have also calculated the 3D structures of their lowest energy conformers. We discuss these structures in the context of the estimated interchromophore distances and the pharmacological properties of the peptides.

Contents

Chapter 1

Introduction

One of the basic principles of molecular biology is that the functions of biomolecules are intimately related to their three-dimensional (3D) structure, which, in its turn, is determined by their chemical structure. Therefore, over the years a number of experimental and computational techniques have been developed for structural characterization of biomolecules. For instance, X-ray crystallography is one of the most widely used tools for 3D structure determination of biomolecules: most of the structures included in Protein Data Bank (more than 10^5 entries) were determined using this method.[1] Also, X-ray crystallography was the first technique that allowed solving the 3D structures of very large species, such as globular proteins[2,3] and nucleic acids.[4,5] Owing to its remarkable resolving power, X-ray crystallography can, in principle, provide detailed information on every atom in a molecule. However, it has some serious limitations.[6–8] Obtaining large enough, well-ordered crystals of a molecule of interest is a prerequisite for solving its 3D structure using X-ray crystallography, and the accuracy of the determined atomic structure critically depends on the quality of these crystals. For instance, due to the inherent thermal motion of atoms in the mobile regions of biomolecules (e.g., the side chains of a protein), their coordinates cannot be reliably determined. Moreover, flexible regions of proteins (e.g., intrinsically disordered domains) often diffract so weakly that they are invisible in the crystallographic electron density maps. Many important biomolecules cannot be crystallized at all and thus cannot be studied using X-ray crystallography. Finally, the conditions required to crystallize biomolecules may not be the same as those *in vivo*, which may significantly disturb their native 3D structure.

In contrast to X-ray crystallography, NMR spectroscopy enables the determination of the 3D structures of biomolecules directly in solution. Although it has been demonstrated that

this technique can be employed to study macromolecular complexes with masses up to 1 MDa,[9,10] its practical application is limited to small- and medium-sized biomolecules with molecular weight below 50 kDa.[11] Another limitation stems from the fact that NMR spectroscopy does not determine the 3D coordinates of the atoms in a molecule but rather gives a list of restraints, such as internuclear distances and torsion angles, which is then used to build its 3D model. Typically, this model represents an ensemble of structures, all of which are consistent with the experimentally determined restraints. Therefore, as in the case of X-ray crystallography, the 3D structure of flexible regions of biomolecules, for which NMR spectroscopy provides fewer and/or less strict restraints, cannot be accurately determined. Despite recent advances in the field, the 3D structure determination of flexible biomolecules, such as intrinsically disordered proteins, still presents challenges to NMR spectroscopy.[12]

Mass spectrometry is another powerful analytical technique, which is based on measurements of the mass-to-charge ratios of ions and their fragments. Although originally developed to measure elemental atomic weights and the natural abundances of isotopes, nowadays mass spectrometry is a vital tool for identification and structural characterization of biomolecules, ranging from small metabolites to large macromolecular assemblies.[13–18] A great variety of ionization techniques, which have been developed over the last century, allow studying solid, liquid, and gaseous samples. Besides measuring the mass-to-charge ratio of ions of interest, one can also fragment them and obtain their fragmentation mass spectra, which reflect structural features of the molecules, from which they are derived from. Such spectra may serve as fingerprints of the molecular species[19] or can be used for the chemical structure (e.g., primary structure of proteins) determination.[20] Moreover, such fragmentation techniques as electron capture and transfer dissociation, which result in prompt, non-ergodic dissociation, may provide information about the higher order structure of biomolecules (e.g., secondary and tertiary structure of proteins) or even give some hints about dynamical processes, such as protein unfolding.[21] Different types of mass analyzers have their own advantages and disadvantages.[22] For instance, triple quadrupole mass spectrometers have very high sensitivity and hence are perfectly suited for quantification of known molecules, such as disease-related biomarkers, in a sample,[23] while Orbitrap-based mass spectrometers offer very high mass resolution and accuracy, which are required for identification of protein sequences and post-translational modifications.[24] Finally, it is worth noting that mass

spectrometry can be readily coupled with various separation techniques, such as liquid chromatography or ion mobility spectrometry, for analyzing very complex biological samples, which contain thousands of different molecules.

Another general method used for identification and structural characterization of biomolecules is optical spectroscopy. Although this technique is not able to determine the 3D structure of biomolecules directly, a number of experimental approaches that provide information on their structural features have been developed. For instance, infrared (IR) spectroscopy enables identification of the presence of certain functional groups in biomolecules, and similarly to fragmentation mass spectra, IR spectra may serve as spectroscopic fingerprints of molecular species.[25] Also, since IR spectra reflect vibrational motion of atoms in molecules, IR spectroscopy may also reveal some features of their 3D structure, such as hydrogen-bond patterns.[26–29] Ultraviolet-visible (UV-Vis) and fluorescence spectroscopy, although they require the presence of a chromophore/fluorophore in a biomolecule, can be used to probe its local structural features. For example, fluorescence quenching of tryptophan side chain indicates whether it is exposed to solvent or buried in protein hydrophobic core;[30] and the measurements of Förster Resonance Energy Transfer (FRET) allow estimating distances between two or more fluorophores in a biomolecule.[31] Circular dichroism (CD) spectra recorded in the far-UV spectral region (190-250 nm) can be used to determine the fraction of different secondary structure types (α-helix, β-sheet, and random coil) in a protein, while the CD signal in the near-UV spectral region (250-350 nm) is sensitive to protein tertiary structure.[32,33] Although a lot of structural information can be obtained from optical spectra, they are usually very broad because of the interaction with solvent molecules, which results in the inhomogeneous spectral broadening. Obviously, this issue can be overcome by isolating biomolecules in the gas phase.

As virtually all biological molecules possess low vapor pressure, special techniques have to be used to transfer biomolecules into the gas phase. Thermal heating is one of them. However, this method often requires very high temperatures and thus results in significant thermal decomposition. Therefore, it has been used only for a limited number of small biomolecules.[34–37] Alternatively, one can use a pulsed laser to instantly heat molecules deposited on a matrix in order to avoid the thermal decomposition. Moreover, when the laser-desorbed molecules are injected into the collision zone of a supersonic expansion and

9

entrained in a carrier gas pulse, they are cooled by multiple collisions with the carrier gas atoms.[38] This allows for rotational cooling to temperatures smaller than 10 K.[39,40] Although vibrational cooling is less efficient, vibrational temperatures in the range of 15 to 50 K can be achieved.[39,40] Such a combination of laser desorption with supersonic expansion has been widely used to obtain highly resolved UV and IR spectra of intact isolated biomolecules with molecular weights up to 2 kDa in the gas phase.[41–49]

Typically, UV spectroscopy in molecular beams involves a use of so-called action spectroscopy. In contrast to conventional absorption spectroscopy, which detects a change of light intensity due to absorption, action spectroscopy monitors changes that happen with the matter itself upon its interaction with light. For instance, laser-induced fluorescence (LIF) spectroscopy implies measurements of the total (undispersed) fluorescence signal as a function of the excitation wavelength, and resonance-enhanced multiphoton ionization (REMPI) technique employs ionization of molecules by two or more photons to measure the photoionization yield as a function of the excitation wavelength.[50]

The first biomolecule studied in a supersonic jet was neutral amino acid tryptophan. In 1985, Levy and co-workers employed resonant two-photon ionization (R2PI) spectroscopy to measure its electronic spectrum,[51] and the power-dependence studies of this spectrum revealed the presence of six different conformers.[52] The use of LIF spectroscopy not only confirmed the assignment but also allowed them to study photophysics of neutral tryptophan in a solvent-free environment.[53] Later, similar experiments were carried out on the other two UV-absorbing amino acids, tyrosine and phenylalanine,[54] and a number of other small biomolecules.[35,41,43,55,56]

An addition of one more laser to these spectroscopic schemes enabled the measurements of conformer-selective UV and IR spectroscopic fingerprints. The first, so-called pump laser is used to deplete the ground state of one of the conformers, while the second, so-called probe laser probes the effect induced by the first one, that is, a decrease in fluorescence intensity or photoionization yield. For instance, UV-UV spectroscopy allows for the assignment of UV peaks to distinct conformers.[50,57] Depending on which of the two lasers is scanned, it is referred to as either hole-burning (the probe laser is scanned) or ion-dip (the pump laser is scanned) spectroscopy. When using an IR laser instead of a UV laser to deplete the ground state, conformer-selective IR spectra can be obtained. In this case, the UV laser is

fixed at a certain vibronic transition of a single conformer and the IR laser is scanned. Whenever the frequency of the IR laser is resonant with a vibrational transition of this conformer, its ground state is depopulated, which results in a dip in the recorded spectrum.[50,58,59] Because of the very efficient intramolecular vibrational redistribution (IVR) process, the IR-induced depletion signal can exceed 50 % (the theoretical limit for a simple two-level system). Alternatively, when the IR laser is fixed at a certain vibrational transition belonging to a single conformer, one can scan the UV laser to record an IR-UV hole-burning spectrum. The application of IR-UV ion-dip spectroscopy to measurements of conformer-selective IR spectra of a molecule requires its UV spectrum to be sufficiently resolved to be able to set the UV laser on a distinct vibronic transition. However, it was recently demonstrated that even in the case when the UV spectra of different conformers significantly overlap conformer-selective IR spectra can be obtained by means of IR-IR-UV spectroscopy.[60]

A use of vacuum ultraviolet (VUV) single-photon ionization instead of R2PI enables studying molecules that lacks a UV chromophore. For instance, if the energy of a VUV photon is chosen to be a bit lower than the ionization energy of a molecule, absorption of an additional IR photon prior to the VUV photon allows reaching the ionization threshold and hence results in a photoionization signal.[61] Also, the IR absorption may affect photodissociation channels even if the energy of a VUV photon alone is sufficient for ionizing the molecule. Thus, one can acquire an IR spectrum by measuring the photodissociation mass spectrum while scanning the IR laser.[62] For example, this enables studying weakly bound clusters,[63] which are challenging cases for IR-UV spectroscopy because they typically have a very short excited state lifetime and hence cannot be efficiently ionized by nanosecond UV laser pulses. It is worth noting that, alternatively, one can employ femtosecond multiphoton ionization technique to ionize them.[64,65]

A great variety of neutral biomolecules isolated in the gas phase have been studied over the last thirty years. A combination of high-resolution spectroscopy with high-level quantum chemical calculations allowed for the elucidation of the 3D structure of peptides,[46,66–85] carbohydrates,[86–92] nucleobases.[70,93–97] Also, spectroscopic techniques have been successfully applied to study their weakly bound complexes[44,68,70,93,98–106] as well as biomolecules microsolvated by water molecules.[70,87,88,92,95,107–117]

However, since biomolecules *in vivo* are often charged (at least, locally), studying of biomolecular ions is as important as the study of neutral biomolecules. Moreover, there are several advantages. For instance, using matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) enables transferring intact biological molecules, ranging from single amino acids to large macromolecular assemblies, into the gas phase in the form of closed-shell molecular ions. Then, one can select the ions of interest on the basis of their mass-to-charge ratios or collisional cross sections. Finally, the ions can be easily manipulated and analyzed using a variety of mass-spectrometric techniques. Thus, photodissociation of ionic species can be routinely detected, which allows using action spectroscopy to obtain their UV and IR spectra (the very low typical ion density in a mass spectrometer impedes the application of direct absorption spectroscopy).

For instance, photofragmentation of ions can be induced by the absorption of multiple IR photons, which, however, requires an IR laser source to deliver a relatively high photon flux (the rate of photon absorption should exceed the rate of collisional and radiative energy relaxation). With the advent of high pulse-energy optical parametric oscillators and free electron lasers, numerous IR multiple photon dissociation (IRMPD) studies appeared.[118–124]

While this spectroscopic technique works well for small biomolecular ions,[118,125–135] the 3D structure determination of larger species becomes less convincing and reliable.[128,136,137] Indeed, as the size of biomolecules and hence the number of vibrational modes increase, the IR spectra become much more complicated. Also, in contrast to conventional spectroscopic methods based on one-photon absorption, IRMPD spectroscopy involves the sequential absorption of a large number of IR photons, which may distort the resulting IR spectra (vibrational bands are typically broadened and red-shifted).[138,139] Since the 3D structure is determined by comparing experimentally measured and theoretically calculated vibrational frequencies, using vibrational spectroscopy as a structural tool for larger molecules requires reducing all sources of spectral congestion.

One of the sources is thermal broadening, which arises from distribution of thermally populated states. In the case of spectroscopic studies of neutral biomolecules in supersonic jets, thermal broadening is suppressed by cooling their internal degrees of freedom in collisions with a carrier gas. Another source of spectral congestion is conformational heterogeneity. And again, rotational and vibrational cooling of biomolecules in supersonic jet

expansions often makes their UV spectra highly resolved, which allows utilizing IR-UV double-resonance spectroscopy to obtain their conformer-selective IR spectra. Thus, cooling of biomolecular ions is required to facilitate spectroscopic analysis and their 3D structure determination.

One of the widely used approaches to cooling of gas-phase ions is buffer-gas cooling, which uses a cold ion trap to trap and cool the ions. The ion trap is installed inside a box, which is cooled down to a desired temperature and filled with a buffer gas. The buffer gas is either leaked or pulsed into the trap, where it thermalizes in collisions with the surrounding cold box. Thus, the ions entering the cold ion trap lose their kinetic and internal energy in inelastic collision with the cold buffer gas atoms/molecules. For instance, in 2004, Weinkauf and co-workers used a modified commercial ESI mass spectrometer with a 3D Paul ion trap cooled by liquid nitrogen to measure the electronic spectrum of cold protonated tryptophan;[140] and in 2006, three separate groups reported the development of cold ion spectrometers based on a cryogenically cooled 22-pole ion trap,[141–143] introduced by Dieter Gerlich in 1993.[144] A number of other cold ion traps and spectrometers have been developed over the last decade.[145–159]

One of the spectroscopic techniques that can be employed to measure the IR spectra of cold gas-phase ions is a "messenger tagging" approach,[160] which is particularly useful in the absence of an electronic chromophore. In this approach, inert species, such as H_2 and N_2 molecules or rare gas atoms, are condensed onto the ions that are trapped and cooled down to a temperature of about 10 K in a cold ion trap, and IR excitation of the formed adducts results in fast photo-evaporation of these "tags". Thus, by monitoring the photofragmentation yield of the adducts as a function of the excitation laser wavenumber, one can acquire the IR spectrum of the ion (the weakly bound "tags" are thought to have little effect on the vibrational spectrum).[161–163]

Recently, a different approach to producing ultracold biomolecular ions and studying them by means of laser spectroscopy was proposed by von Helden and co-workers.[164] In this approach, gas-phase ions are produced by ESI and trapped in a room temperature ion trap. Ultracold helium droplets, produced by a helium droplet source, traverse the trap and pick up the ions. Once the pick-up takes place, the ions are cooled down to the temperature of the helium droplets (≈ 0.4 K), and then, the doped droplets are intersected by a laser beam. Upon

UV or IR excitation, the ions are ejected from the droplets. Thus, UV and/or IR spectra of ultracold biomolecular ions embedded in helium droplets can be obtained by recording the photo-ejection yield while scanning the excitation laser wavenumber.[157,165,166]

In our Laboratory, buffer-gas cooling in a 22-pole linear ion trap, mounted on a cold head of a closed cycle refrigerator, is used to cool biomolecular ions, produced by ESI. For instance, Boyarkin et al. measured UV spectra of cold protonated tyrosine and tryptophan by recording the photodissociation signal as a function of the UV laser wavenumber.[141] While the obtained UV spectrum of protonated tyrosine with a well-resolved vibronic structure and the 2.7 cm^{-1} linewidth of individual peaks proved the capability of this method to cool biomolecular ions in the gas phase to very low internal temperatures (≈ 10 K), the UV spectrum of protonated tryptophan remained rather broad due to its short excited state lifetime. However, Mercier et al. demonstrated that solvation of protonated tryptophan even by only two water molecules dramatically sharpens its UV spectrum.[167] Moreover, the appearance of sharply resolved features in UV spectra of cold ions allowed for the measurements of their conformer-specific IR spectra using IR-UV double-resonance spectroscopy, when a UV laser is set to a distinct sharp UV peak and an IR laser is scanned. The IR pulse precedes the UV pulse by approximately 100 ns so that, whenever the IR laser frequency is resonant with a ground state vibration of the conformer excited by the UV laser, its ground state is depopulated, which is detected as a depletion in the UV-induced photofragmentation signal. For example, Stearns et al. employed IR-UV spectroscopy to measure the conformer-specific IR spectra of protonated phenylalanine and tyrosine and, by comparing them with the results of high-level quantum chemical calculations, determined the 3D structure of the conformers.[168] Later on, the same approach was used for the 3D structure determination of a number of small protonated peptide containing either phenylalanine or tyrosine as a UV chromophore.[169–175] To date, the largest species, for which its exact 3D structure has been revealed by cold ion spectroscopy, is the most abundant gas-phase conformer of a doubly-protonated decapeptide, gramicidin S.[176] Furthermore, Nagornova et al. demonstrated that cold ion spectroscopy can be applied to track changes in its 3D structure upon addition of 1 to 50 water molecules in a gas-phase cluster.[177]

While cold ion spectroscopy has been successfully employed for structural studies of relatively small biomolecular ions, its application to larger species becomes less

straightforward. First, investigations of larger systems are usually inhibited because of their lower photofragmentation yields. This issue can be, at least, partially solved by infrared multiple-photon excitation (IRMPE) of ions after the UV excitation of their chromophores. For instance, Guidi et al. reported an increase in the photofragmentation yield of up to two orders of magnitude by using a pulsed CO_2 laser for IRMPE.[178] Alternatively, one can use the "messenger tagging" approach described above. Another problem is due to conformational heterogeneity, which can still be a significant source of spectral congestion. One possible solution to this problem is to separate conformers prior to the spectroscopic analysis using ion mobility spectrometry. For example, a combination of high-field asymmetric waveform ion mobility spectrometry (FAIMS) and cold ion spectroscopy enabled to partially separate the conformational families of a doubly- and triply-protonated peptide, bradykinin, and thus to resolve their spectra.[154,179,180] UV spectra of large biomolecular ions may also be unstructured because of the very large number of low-frequency vibrations. Although this inhibits the use of conformer-selective IR-UV spectroscopy, it was demonstrated that it is still possible to record all-conformer IR "gain" spectra of molecules as large as small proteins.[181] Moreover, an addition of one more IR laser to this spectroscopic scheme may, in principle, enable measurements of conformer-specific IR spectra. For instance, IR-IR-UV spectroscopy has been recently employed to obtain individual IR spectra of two conformers of protonated tryptophan,[182] the UV spectrum of which is intrinsically broad. Finally, the practical use of cold ion spectroscopy for the 3D structure determination of biomolecules larger than di- and tri-peptides is usually impeded by unfeasibly time-consuming quantum chemical calculations. Thus, development of novel approaches that do not rely on the high-level calculations would be highly desirable.

While cold ion spectroscopy is inherently based on various mass-spectrometric techniques, the information contained in photofragmentation mass spectra is typically overlooked. At the same time, it was shown that two conformational families of protonated tyrosine display different photofragmentation mass spectra,[168] which, in principle, may allow distinguishing them. Also, we have demonstrated that the enhanced selectivity offered by a combination of cold ion spectroscopy and mass spectrometry may be useful, in particular, for identification of tyrosine-phosphorylated peptides and, in general, for identification of UV chromophores in biomolecular ions.[183]

To sum up, the main objective of this work is to develop novel approaches for structural characterization of cold biomolecular ions in the gas phase that use not only the spectroscopic fingerprints of these species but also the information provided by the fragmentation mass spectra that the electronic excitation produces. First, we describe a coupling of the existing cold ion spectrometer with a commercial high-resolution Orbitrap-based mass spectrometer, which allows recording the entire photofragmentation mass spectra at each excitation wavenumber. Then, we demonstrate three particular applications of this new hybrid instrument, namely: 1) identification of isomeric biomolecules in mixtures; 2) identification of conformers of biomolecular ions and deriving structural information on the basis of their conformer-specific UV and MS spectra; and 3) measurements of interchromophore distances in gas-phase opioid peptides.

Chapter 2

Experimental Setup

2.1. Cold Ion Spectrometer

A schematic diagram of the cold ion spectrometer that was built more than a decade ago and upgraded by adding a cryogenically cooled 22-pole ion trap is shown in Figure 2.1. It consists of a number of differentially pumped vacuum chambers (not shown in Figure 2.1) and an ion optical system, which guides ions from their entrance to the spectrometer from the atmospheric-pressure region, where they are produced by nano-electrospray ionization (nanoESI) technique, to the region, where they are detected by a channeltron detector. The ion optical system includes a transfer capillary, a skimmer, a hexapole ion trap, two quadrupole mass filters, two quadrupole deflectors, an octupole ion guide, electrostatic lenses, and a 22-pole ion trap, mounted on a cold head of a cryocooler. The cold ion spectrometer is not linear, but rather Z-shaped, with two deflectors giving two 90° turns to the ion beam. Apart from reducing the gas load to turbomolecular pumps and reducing the background pressure in the high-vacuum stages, this creates a clear path for laser beams on the axis of the 22-pole ion trap, allowing them to enter and exit the cold ion spectrometer through windows placed at Brewster's angle.

A typical experimental workflow is the following. The sample to be studied is dissolved in water, methanol, or their mixture with addition of acetic acid and sprayed at atmospheric pressure from a metal-coated glass needle towards the entrance of the metal-coated glass capillary. The ions, produced by nanoESI, are transferred with the airflow through the capillary to the first vacuum stage, which is pumped to a pressure of about 2 mbar. Then, the ions pass through the skimmer and are pre-trapped in the 6 cm long linear hexapole ion trap, which spans the second and the third vacuum stages. The pressures in these stages are at the level of $5 \cdot 10^{-2}$ and $5 \cdot 10^{-5}$ mbar, respectively. A pulsed voltage is applied to the exit electrode of the
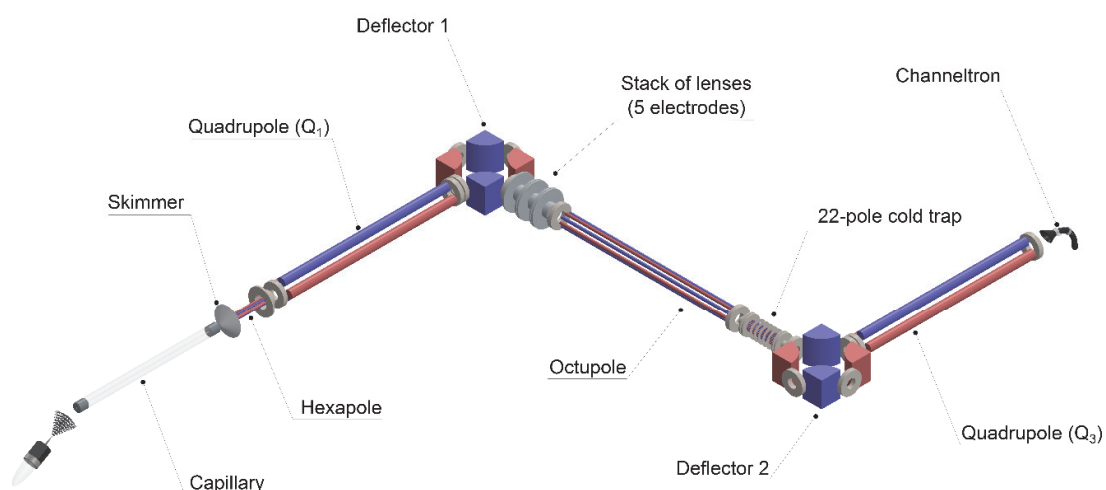
Figure 2.1. Schematic diagram of the cold ion spectrometer.

hexapole, which allows accumulating ions and periodically releasing them as dense ion packets. The hexapole is followed by the first quadrupole mass filter (4th vacuum stage, $p \approx 2 \cdot 10^{-7}$ mbar), which selects ions of a particular mass-to-charge ratio (m/z). The mass-selected ions are turned by 90° by the first quadrupole deflector and transmitted through the linear RF-only octupole ion guide to the 22-pole ion trap, where they get trapped and cooled in collisions with a pulse of helium buffer gas, which is introduced into the trap $\approx$ 0.2 ms before the arrival of the ion packet. The trap is installed on the cold head of the cryocooler (SRDK-408R, Sumitomo Heavy Industries). Thus, helium atoms thermalize in collisions with the walls of the trap and attain a translational temperature of 6 K, which allows cooling the ions to a vibrational temperature of about 10 K.[141] After about 30 ms, when the helium is pumped out, photofragmentation spectroscopy is performed inside the 22-pole ion trap by irradiating the trapped ions with a UV and/or IR laser pulse. Then both parent and fragment ions are released from the trap and turned by 90° by the second quadrupole deflector. The second quadrupole mass filter selects either parent ions or fragment ions of a particular m/z, and the transmitted ions are detected by the channeltron detector. The pressures in the octupole ion guide, the 22-pole ion trap, and the second quadrupole mass filter, which correspond to the 5th, 6th, and 7th vacuum stages, respectively, are in the range of 10^{-9} mbar (ultimate vacuum) to 10^{-4} mbar (peak pressure during a He pulse).

2.2. Laser Setups and Spectroscopic Techniques

UV laser light used for electronic excitation of chromophores is produced by frequency doubling in a BBO (beta barium borate) crystal the visible output of a dye laser (HD-500, Lumonics) pumped by the third harmonic of a Nd:YAG (neodymium-doped yttrium aluminum garnet) laser (GCR-250, Spectra Physics). The resulting UV beam is separated from the remaining visible light via reflections on two dichroic mirrors and sent into the vacuum manifold of the cold ion spectrometer through a BaF_2 window placed at Brewster's angle. This is schematically depicted in Figure 2.2. For all the experiments described in this thesis Coumarin 540A dye was used. The BBO crystal is mounted along with a fused silica compensator in an Autotracker III (Inrad), and both are automatically rotated to the phase-matching angle as the visible wavelength is scanned. The UV pulses have a duration of 5 to 10 ns, and a typical energy is in the range of 1 to 4 mJ at 10 Hz repetition rate of the Nd:YAG laser.
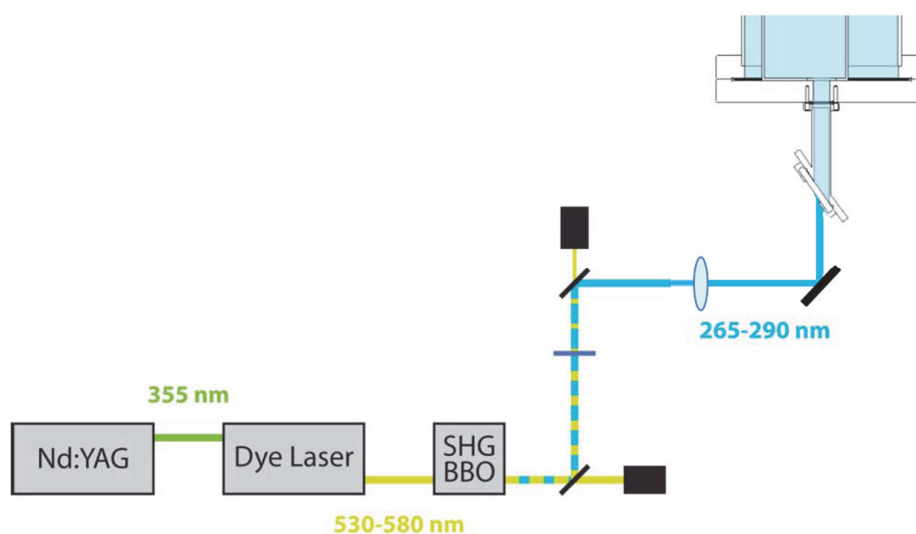
Figure 2.2. Schematic diagram of the UV light generation.

To generate IR laser light in the 3 μm region, we use an optical parametric oscillator/amplifier (OPO/OPA) system (LaserVision), pumped by the 1064 nm fundamental mode of a Nd:YAG laser (SpitLight 600, Innolas).The vertically polarized idler beam from the OPO/OPA is sent into the vacuum manifold of the cold ion spectrometer through another BaF_2 window placed at Brewster's angle (Figure 2.2). The Nd:YAG laser operates at 5 or 10 Hz

repetition rate. The IR pulse energy varies in the range of 1 to 10 mJ depending on the IR wavenumber. To generate IR light in the 6 μm region we remove a Brewster plate polarizer at the exit of the OPA stage and perform a difference frequency mixing of the horizontally polarized signal (≈ 1.5 μm) and vertically polarized idler (≈ 3 μm) outputs of the OPA stage in a AgGaSe$_2$ crystal. Both signal and idler waves are removed by a ZnSe filter, placed right after the nonlinear crystal, while the 6 μm light goes to the cold ion spectrometer via the same optical path (Figure 2.2). The IR pulse energy in the 6 μm region varies in the range of 0.5 to 2 mJ depending on the IR wavenumber.
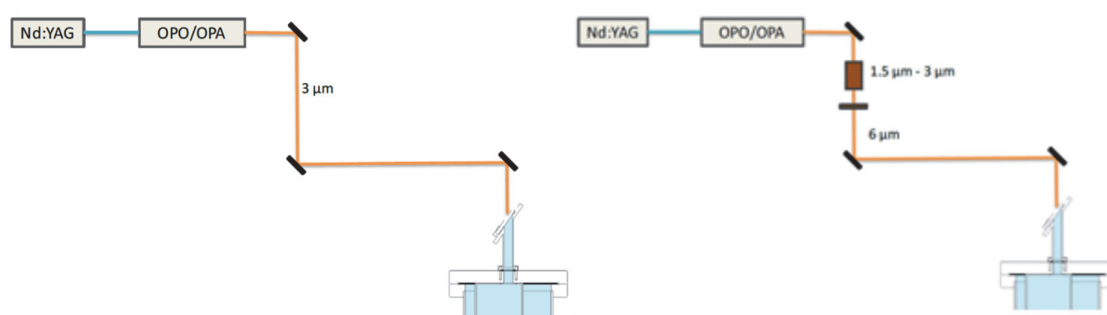
Figure 2.3. Schematic diagram of the generation of IR light in the 3 μm (left panel) and 6 μm (right panel) regions.

Different types of spectra can be obtained using the cold ions spectrometer. First, one can record a photofragmentation mass spectrum by fixing the wavenumber of the UV laser and scanning the Q$_3$ quadrupole mass filter over a certain m/z range. Once the m/z of a photofragment(s) is known, a UV spectrum is recorded by fixing the Q$_3$ quadrupole mass filter at the m/z of a particular fragment and monitoring the intensity of the photofragmentation signal (i.e., the number of photofragments) as a function of the UV laser wavenumber (Figure 2.4(a)). In order to increase the photofragmentation signal, the ions can be additionally activated by a CO$_2$ laser pulse following the UV excitation. CO$_2$ laser light enters the cold ions spectrometer via the same optical path as IR laser light.

To record a conformer-selective IR spectrum, we set the UV laser wavelength on a particular feature (i.e., narrow vibronic transitions) in UV spectrum. In this experiment UV and IR laser pulses are generated at 20Hz and 10 Hz repetition rates, respectively. The IR pulse precedes the UV pulse by about 100 ns. We measure the difference between the number of photofragments in the two alternating cycles, with UV-only and IR-UV excitation, as a function

of the IR laser wavenumber (Figure 2.4(b)). Alternatively, the UV laser wavenumber is tuned to the low-wavenumber, red edge of the UV spectrum. In this case we measure an increase of the photofragmentation signal of IR-preheated ions (Figure 2.4(c)).
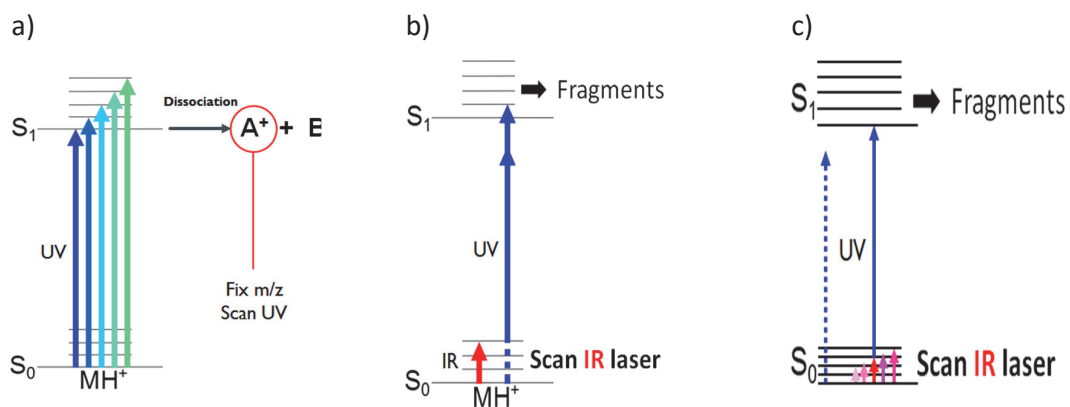
Figure 2.4. Spectroscopic schemes for the measurements of **(a)** UV, **(b)** IR depletion and **(c)** IR gain spectra.

Chapter 3

Coupling an Orbitrap-based Mass Spectrometer to the Existing Cold Ion Spectrometer

This chapter describes coupling of an Exactive Orbitrap-based mass spectrometer to the existing cold ion spectrometer. First, we give a brief description of the Exactive mass spectrometer. Then, we present an interface that was used to couple the two instruments. And finally, we describe a typical experimental workflow and the details of experimental data processing.

3.1. Introduction

The Orbitrap, invented by Alexander Makarov in 1999,[184,185] is a purely electrostatic mass analyzer based on the principle of orbital trapping, introduced by Kingdon in 1923. [186] It consists of an outer barrel-like electrode and a central spindle-like electrode (see the Orbitrap Mass Analyzer in Figure 3.1), which create a quadrologarithmic electrostatic potential:

$$U(r,z) = \frac{k}{2} \cdot \left(z^2 - \frac{r^2}{2} \right) + \frac{k}{2} \cdot R_m^2 \cdot \ln \left(\frac{r}{R_m} \right) + const \qquad (3.1)$$

where r and z are cylindrical coordinates, k is field curvature and R_m is the characteristic radius. Stable ion trajectories involve both orbital motion around the central electrode and oscillations in the z-direction around the plane of symmetry ($z = 0$).[185] Since the electrostatic potential is quadratic with respect to the z coordinate, the ion motion along the z-axis can be

described as simple harmonic oscillations with an angular frequency determined by the mass-to-charge ratio (m/z) of the ions:

$$\omega = \sqrt{\frac{k}{(m/z)}} \qquad (3.2)$$

The outer electrode is split into two halves at $z = 0$ so that the oscillations of the ion cloud induce an image current on the two halves, which is differentially amplified and subsequently converted from the frequency domain into the m/z domain by means of the fast Fourier transform (FFT).

The key features of any mass spectrometer are its mass resolution, mass accuracy and dynamic range. Mass resolution, R, depends on the spectrum acquisition time: a longer acquisition time results in a higher resolution in the frequency domain and hence in a higher mass resolution. The maximum acquisition time is determined by several factors such as collisions with background gas, field imperfections, electric field instability, and space-charge effects, which all result in ion losses and an ion package dephasing.[187] In commercial instruments these effects are markedly reduced by employing highly accurate manufacturing, very stable power supplies, and ultrahigh vacuum (< 10^{-9} mbar) so that a resolving power above 10^6 has been demonstrated for m/z below 300-350.[188] The mass accuracy is affected by such factors as variations in the initial injection positions of ions, space-charge effects, very weak cross-terms (r, z) in the electrostatic potential, low signal-to-noise (S/N) ratio. Currently mass accuracy better than 1 ppm can be achieved.[189] The dynamic range is determined by the maximum total number of ions that can be analyzed with the Orbitrap and its detection limit. The former is limited by the size and configuration of an Orbitrap analyzer and lies in the range of (0.5-1)·10^6 elementary charges,[24] while the latter is limited mainly by the internal noise of the used preamplifier and can be as low as 1-2 ions.[190] For instance, it was shown that a mass accuracy of 5 ppm can be reached at a dynamic range of more than 5000.[191]

Orbitrap-based mass spectrometers have been successfully applied in the fields of proteomics,[192–195] lipidomics,[196–199] metabolomics,[196,200–202] environmental analysis,[203,204] doping control,[205,206] etc. A number of fragmentation techniques have been implemented on Orbitrap mass spectrometers for MS/MS analysis.[189,193,207–210] Particularly, infrared multiple-

photon dissociation (IRMPD) and vacuum ultraviolet (VUV) dissociation were integrated into commercial Orbitrap mass spectrometers,[211–215] however, to our knowledge, there are no examples of using tunable light sources and Orbitrap mass spectrometers for measuring 2D spectra (i.e., high-resolution mass spectra as a function of the excitation wavelength) of ions.

3.2. Description of the Exactive Mass Spectrometer

The Exactive Orbitrap-based mass spectrometer (Thermo Fisher Scientific) available to us is a standalone commercial instrument intended for high-throughput LC-MS applications. Figure 3.1 schematically shows its main components: an atmospheric pressure ionization (API) source, ion optics, a curved linear ion trap (C-Trap), a collision cell, and an Orbitrap mass analyzer.
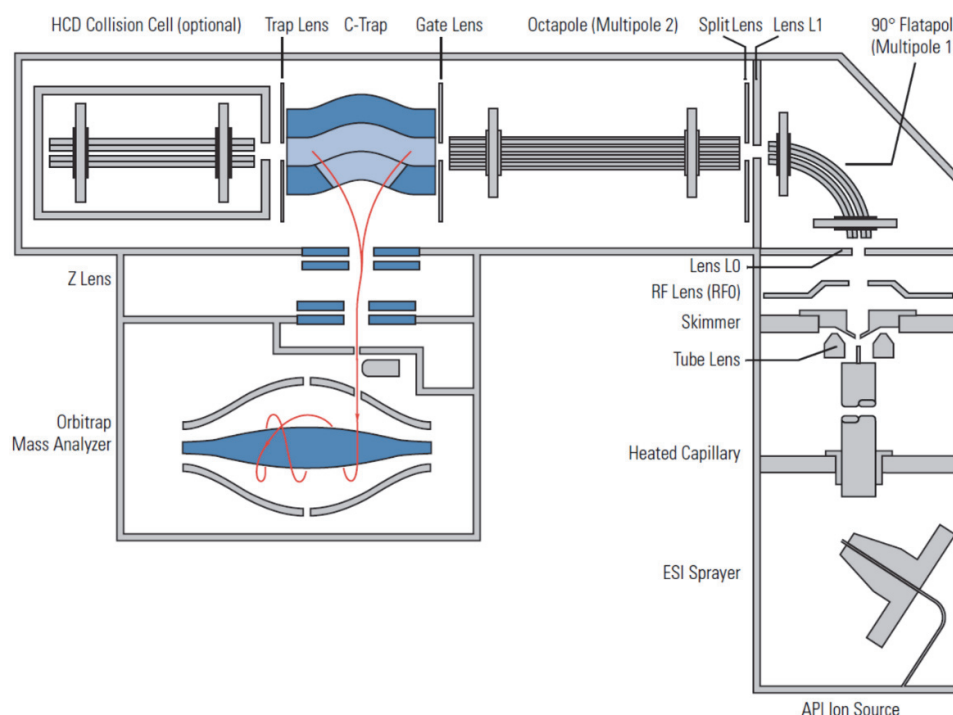

Figure 3.1. Schematic representation of the Exactive Orbitrap-based mass spectrometer.

Samples are introduced into the API source either directly via a syringe pump or after preliminary separation by means of a LC system. Ions generated from a solution using an electrospray ionization (ESI) technique are transferred through the heated capillary and

focused by the tube lens towards the orifice of the skimmer. The skimmer serves as a vacuum baffle between the higher and lower pressure regions of the vacuum manifold. The capillary-skimmer region is pumped out by a rotary-vane pump (Sogevac SV 40 BI, Oerlikon Leybold Vacuum). After passing the skimmer, the ions are focused by the RF lens (RF0) and transmitted to the RF-only flatapole (Multipole 1) through the aperture of the lens L0. The flatapole rods are bent through a 90° arc in order to minimize the transmission of unwanted neutral species. Ions from the flatapole pass through the lens L1 and the Split lens to the RF-only octupole (Multipole 2), which transmits them to the C-Trap. Lenses L0 and L1, both act as vacuum baffles and split the second vacuum chamber into three differential pumping stages, pumped out by a split-flow turbomolecular pump (SplitFlow 310, Pfeiffer Vacuum). The injection of ions into the C-Trap is controlled by applying either a "closed" (± 350 V) or an "open" (0 V) voltage to the Split lens. The C-Trap is a curved linear ion trap intended for accumulating the ions, squeezing, and injecting them into the Orbitrap mass analyzer. Voltages applied to the two end-cap electrodes (gate and trap lenses) of the C-Trap create a potential well along its axis. A constant nitrogen flow ($\approx$ 0.2 mL/min) is directed to the C-Trap so that the trap is kept at a pressure of about $2 \cdot 10^{-5}$ mbar. The ions entering the C-Trap lose their kinetic energy in collisions with the nitrogen buffer gas and are eventually focused close to the trap axis. In order to extract the trapped ions towards the mass analyzer, the C-Trap RF voltage is switched off, while the DC voltages applied to its rods and to the end-cap electrodes squeeze the ions along the curved axis and push them orthogonally to this axis. Then, the ions are transmitted through the Z-lens, electrostatically deflected, and injected into the Orbitrap through a slot in its outer electrode. The use of the Z-lens and the deflector allows for very efficient differential pumping so that the pressure inside the Orbitrap is kept lower than $1 \cdot 10^{-9}$ mbar. The vacuum in the analyzer chamber is maintained by a second split-flow turbomolecular pump (Turbovac TW 290/20/20-UHV, Oerlikon Leybold Vacuum).

Our Exactive mass spectrometer is also equipped with a collision cell, which allows performing higher-energy collision-induced dissociation (HCD) experiments. The collision cell is an additional vacuum chamber with an octupole ion trap inside. The HCD cell is coupled to the C-Trap, so that for HCD experiments the ions leaving the Multipole 2 pass through the C-Trap (both gate and trap lenses are set to transmit ions) and are accelerated at the entrance to the HCD cell, which is filled with a collision gas. In collisions with the buffer gas the ions are

fragmented, and then both parent and fragment ions are returned back to the C-Trap. After being accumulated in the C-Trap, the ions are sent to the Orbitrap for mass analysis.

The Exactive mass spectrometer is controlled by an external PC running the Exactive Tune software. The Exactive contains an internal computer, which communicates with the external PC via an Ethernet interface. The internal computer, therefore, serves as a front end for all the electronics of the Exactive, which comprise a number of printed circuit boards (PCB). The PCBs supply the Exactive components with the DC and RF voltages; they also monitor readbacks, control the vacuum system (turbomolecular pumps, pressure gauges, valves), supply high voltage to the ESI needle, communicate with peripheral devices (syringe pump, LC system), acquire and process the induced current from the Orbitrap mass analyzer, etc. The PCBs extensively communicate with each other to synchronize the operation of all the Exactive components. After all, there are interlocks that ensure safe and reliable operation of the Exactive as an entire instrument.
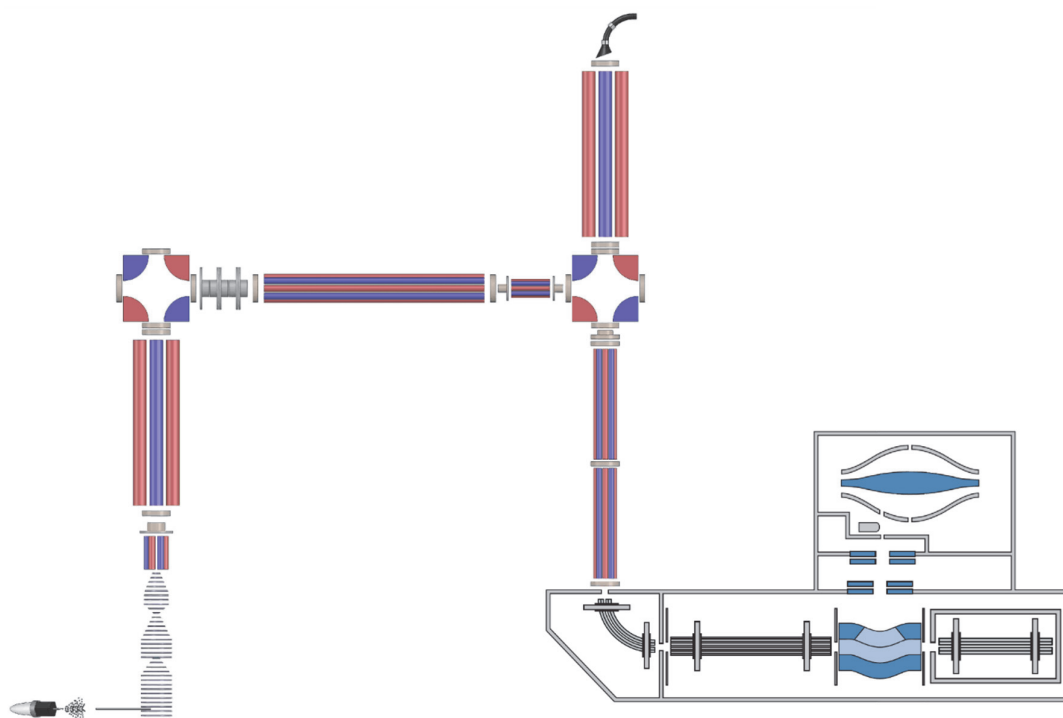

Figure 3.2. Schematic view of the coupled instrument.

3.3. Design of the Interface

As it follows from the above description of the Exactive, it is a complicated instrument, intended for being used as a standalone mass spectrometer or as a part of an LC-MS system. Thus, coupling it to our existing cold ion spectrometer (CIS) and using them together required certain technical issues to be solved. First, in order to increase the number of both parent and photofragment ions as well as to improve the stability of the ion signal, we replaced the skimmer-based ion source by an orthogonal double-ion-funnel ion source (Spectroglyph) and replaced the 22-pole by an 8-pole cold ion trap (for details see ref. [158] and the Supporting Information in ref. [216]). Then, we wanted to employ the Exactive for measuring mass spectra of the ions stored in the cold ion trap but at the same time to keep the possibility of using the quadrupole mass filter (Q3) for our experiments. Therefore, we decided to couple the Exactive to the second quadrupole deflector (D2) of the CIS, which allows turning the ions from the cold trap either to the Q3 or towards the Exactive (Figure 3.2). However, since the Exactive itself does not provide means for coupling its vacuum manifold as well as its ion optics to the CIS, we have replaced the ion source section of the Exactive by a custom piece, which mimics the overall dimensions of the original assembly and has a CF flange on one side.

The Exactive mass spectrometer can operate with a number of API ion sources, such as ESI and nano-ESI (nESI) ionization, atmospheric pressure chemical ionization (APCI) and atmospheric pressure photoionization (APPI) sources. The Exactive monitors whether any of these ion sources are connected to switch on the ion optics and high voltage (8kV) power supplies and attempts to automatically identify the ion source type. Therefore, the only interlock that we had to overcome after the replacement of the ion source section of the Exactive is the ion source interlock. Each API source compatible with the Exactive has an identifier (ID), which is encoded in the source D-Sub plug via resistors soldered to its pins. The Exactive has a corresponding D-Sub socket on its front side so that, when an API source is connected to the Exactive, it probes the resistance between certain pins and determines the ion source type. To emulate the ESI ion source we connected to the D-Sub socket a plug with the following pinout: a 2.2 kΩ resistor is soldered between pins 7 and 8; pins 5 and 12 and pins 6 and 13 are short-circuited in pairs; pins 9, 10 and 11 are soldered together; the remaining pins are left unconnected.

The pressure inside the source chamber of the Exactive is on the order of 10^{-5} mbar, while the pressure in the region of D2 is in the range of 10^{-9} mbar (ultimate vacuum) to 10^{-4} mbar (peak pressure during He pulse). Therefore, to isolate the vacuum manifolds of the Exactive and the CIS from each other, we have coupled them via an interface (Figure 3.3) that consists of two differentially pumped vacuum chambers. Each chamber is evacuated by a 260 L/s turbomolecular pump (HiPace 300, Pfeiffer Vacuum) backed by a 4.8 m^3/h dry pump (XDS 5, Edwards). The pressure inside the chambers is monitored by cold cathode gauges (IKR 261, Pfeiffer Vacuum).
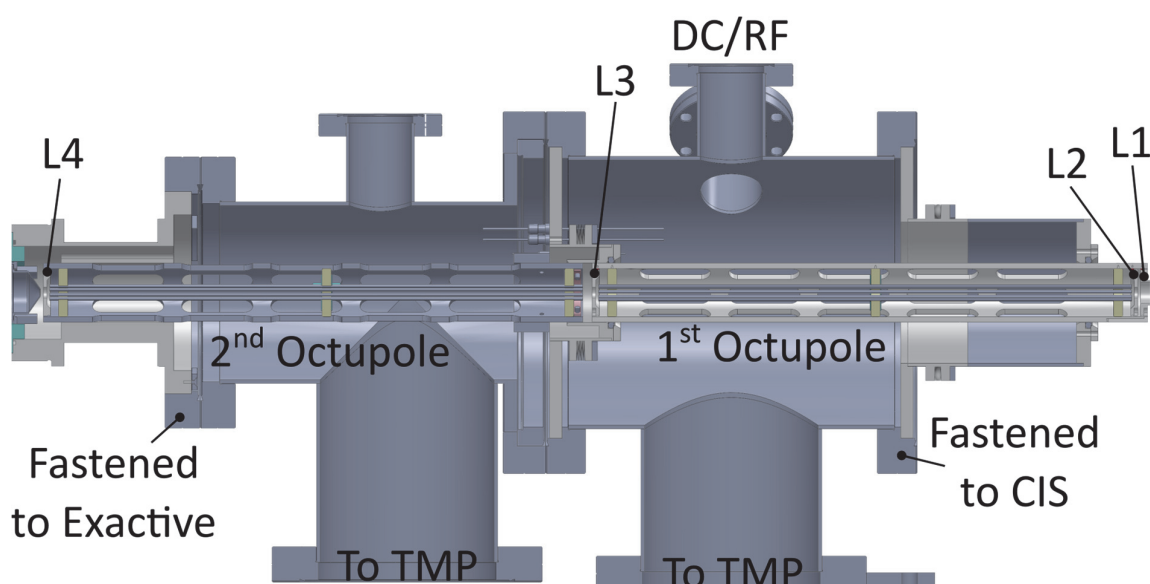
Figure 3.3. Section view of the interface that couples the cold ion spectrometer (on the right side) and the Exactive mass spectrometer (on the left side).

The interface ion optics include two 30 cm long RF-only octupoles (Thermo Fisher Scientific) with 1.85 mm diameter rods on an inscribed radius of 2.875 mm, driven by a 1 MHz RF power supply built in-house by our electronic workshop, and five DC electrodes. The peak-to-peak RF amplitude is determined by an external DC power supply (MCL 160-650, F.u.G. Elektronik) and can be varied from zero to a few hundred volts. A PC-controlled DC power supply (Extrel CMS) is used to supply voltages to the DC electrodes.

3.4. Synchronization of the Duty Cycles

Once the Exactive mass spectrometer had been physically connected to the CIS, the next step was to synchronize their duty cycles. The duty cycle of the CIS is determined by the repetition rate of a Nd:YAG laser. For stable and reliable operation, the laser flashlamps should be triggered with a repetition rate of 20 ± 5 Hz. Therefore, the duration of the duty cycle has to be a multiple of 50 ms and its minimum duration is limited by 50 ms. On the other hand, the Exactive mass spectrometer is much more complicated system with a number of parameters affecting its duty cycle (ion injection time, resolution, various delays, HCD settings, etc.) and a number of PCBs synchronizing over the serial peripheral interface (SPI) bus and coaxial cables. For this reason, we decided not to interfere in the internal logic of the Exactive but rather to trigger the CIS with a certain synchronization pulse from the Exactive.

Figure 3.4. **(Left panel)** Schematic diagram of triggering the cold ion spectrometer by external pulses, taken from the test point TP 17 on the Ion Optic Supply RF board of the Exactive mass spectrometer. **(Right panel)** Snapshot of the oscilloscope display that shows a synchronization between the pulse voltage applied to the Split lens (from 0 to 350 V) and the TTL pulses at the test point TP 17.

The most natural way for doing this is to trigger the master pulse generator of the CIS by the pulsed voltage applied to the Split lens. Then, the settings of the other two pulse generator, which trigger the ion optics of the CIS and the Nd:YAG laser, can be readily adjusted to release ions from the cold ion trap exactly during the period of time when the Split lens is set to transmit ions. Thus, from the Exactive side there is no difference between its normal operation mode, when a continuous flow of ions, generated by the API source, is transmitted to the C-Trap only during the injection time period, and a mode when ion bunches from the cold ion trap are injected to the C-Trap at a certain repetition rate. Because the voltage applied

to the Split lens is rather high (pulsed from 0 to ±350 V), we trigger the master pulse generator by the TTL pulses taken from the test point TP17 on the Ion Optic Supply RF board and synchronized with the Split lens pulsed voltage (Figure 3.4).

3.5. Data Acquisition and Processing

3.5.1. Measurement of a 2D UV-MS spectrum

A typical experimental workflow is the following (Figure 3.2). The ions produced by nESI are transferred through a metal capillary and two ion funnels and pre-trapped in a hexapole ion trap. Ion bunches are periodically released from the hexapole and guided into the cold octupole ion trap, where they get trapped and cooled down to a vibrational temperature of 10 K in collisions with a pulse of helium buffer gas. The Q1 quadrupole mass filter can be either set to an RF-only mode to transmit all the ions stored in the hexapole, or alternatively, it can be set to a certain m/z value to select only the ions of interest (e.g., isomers, which have the same m/z). The ions are fragmented by a UV laser pulse, and both fragment and parent ions are guided into the Orbitrap mass analyzer. At each UV wavenumber a single mass spectrum is obtained by averaging 2-10 microscans (each microscan corresponds to a single laser shot) of the Orbitrap and a computer-controlled data acquisition system, synchronized with the CIS, records the UV laser power, measured by a pyroelectric detector, and the UV laser wavenumber, measured by a wavelength meter (WS/5, HighFinesse/Ångstrom). 2D UV-MS spectra, or UV-MS fingerprints, are measured by continuously recording the photofragmentation mass spectra on the Orbitrap while scanning the UV laser. Regardless of the duty cycle of the Exactive, which is mainly determined by a requested mass resolution, the flashlamps and the Q-switch of the Nd:YAG laser are triggered at 20 and 10 Hz, respectively.

3.5.2. Converting raw data into a matrix

The produced data is a set of mass spectra stored as a Thermo RAW file and a set of pairs of the UV laser wavenumbers and pulse energies stored as a plain text file. First, the recorded mass spectra are converted from the RAW data format into plain text by an in-house developed program, written in C++ with using the MSFileReader library (Thermo Fisher

Scientific). Briefly, the whole m/z range is divided into a set of equal bins of a certain size (typically, 0.1 Da). Each high-resolution mass scan from the RAW file is then sampled over this set, and the resulting array of ion intensities is written as a single space-separated string in a text file. Thus, the output text file contains the recorded 2D spectrum in a matrix form, where the rows correspond to different UV laser wavenumbers and the columns correspond to different m/z values. This matrix as well as the corresponding pairs of the UV laser wavenumbers/pulse-energies are loaded into and subsequently processed by the MATLAB software suite.

3.5.3. Baseline correction

The UV-MS matrix imported into MATLAB is baseline-corrected as follows. Each column of the matrix represents a UV spectrum recorded at a certain m/z value. The spectrum is smoothed using a Savitzky-Golay filter (MATLAB built-in function "smooth" with the "method" argument set to "sgolay").[217] The sum of such smoothed spectra over the entire wavenumber range gives an estimated "signal" value, and the norm of the residual between the smoothed and raw spectra gives an estimated "noise" value. The "signal" values as a function of m/z essentially represent a mass spectrum integrated over the wavenumber scale. Typically, a fragmentation mass spectrum contains only a few peaks of considerable intensity but large peak-free regions. The second derivative of the integrated mass spectrum is used to find these regions. Two types of points in the second-derivative spectra are considered to be peak-related: positive local maxima and those negative local minima, the absolute values of which are below some threshold. All other points are considered to belong to peak-free regions. The choice of the threshold strikes a compromise between a loss of some peaks for a too high threshold and an erroneous assignment of noise spikes as truly mass peaks for a too low threshold. Typically, we use a relative threshold value of 20, which is multiplied by the "noise" level at each m/z to give the m/z-dependent absolute threshold values. Additionally, we manually remove the peaks that systematically appear as artifacts in our mass spectra from the set of peak-related data points. After the identification of the peak-free regions, for each mass spectrum we employ a shape-preserving piecewise cubic interpolation (MATLAB built-in function "interp1" with the "method" argument set to "pchip") taking these regions

as a set of sample data points to calculate a baseline and then subtract it. Finally, we replace all the negative values in the resulting UV-MS matrix by zeros.

3.5.4. Peak picking

After the baseline correction the following algorithm is applied to the UV-MS matrix in order to transform it into a set of UV spectra measured at the most prominent fragment and parent mass peaks. First, the integrated over the wavenumber scale mass spectrum is recalculated for the baseline-corrected UV-MS matrix. Then, starting from the lowest m/z value and going to higher m/z, we search for a local maximum in the integrated mass spectrum that have a S/N ratio greater than 5. Once such a maximum is found, we determine the edges of the corresponding peak by searching in both directions (i.e., going to lower and higher m/z) the nearest local minimum or a point, at which the S/N ratio is below 5. We add a tuple containing the low and high m/z edges and the m/z of the peak maximum to a list of identified peaks and continue searching for a next local maximum with S/N > 5 starting from the high m/z edge of the identified peak. This procedure is repeated until the highest m/z end of the mass spectrum is reached. Then, for each identified peak in the list, the UV spectra in the UV-MS matrix that correspond to the m/z values between the low and high edges of the peak are summed up. The resulting UV spectra are concatenated to yield a new UV-MS matrix, and the m/z values of the peak maxima are stored as a list of fragment and parent m/z. Finally, we may set an absolute or relative threshold to retain in the resulting UV-MS matrix the UV spectra of only those peaks, the integrated (over the wavenumber scale) intensity of which is higher than a certain value.

Chapter 4

Identification of Biomolecules in Mixtures

In this chapter we report on a use of 2D UV-MS spectroscopy for identification of biomolecules in mixtures. The synergy of the two orthogonal techniques, ultraviolet (UV) spectroscopy and mass spectrometry (MS), makes the recorded 2D UV-MS spectra unique fingerprints of molecular ions. We demonstrate how a library of such fingerprints can be utilized for qualitative and quantitative identification of 1) positional isomers of a phosphopeptide, 2) stereoisomers of a peptide, and 3) stereoisomers of a drug molecule in their mixtures, – a library-based approach. Then, we present a library-free approach, which allows estimating the number of components present in a mixture and deriving their UV absorption and fragmentation mass spectra directly from the 2D UV-MS spectrum of the mixture. Finally, we discuss various aspects of practical applicability of the 2D UV-MS approach.

4.1. Introduction

Structural identification of biomolecules remains in the core of life-science research on the fundamental level. Since the introduction of electrospray ionization (ESI)[218,219] and matrix-assisted laser desorption ionization techniques (MALDI),[220,221] mass spectrometry (MS) has become one of the most powerful and versatile tools in this field. Although analysis by state-of-the-art high-resolution MS demonstrates impressively accurate determination of, for instance, peptide sequences, the identification of isobaric molecules, and especially isomers, remains among the challenges of this technique.[222,223] Coupling of MS to liquid

chromatography (LC) or ion mobility (IM) often enables separation of such species and tagging each of them with a retention or arrival time, respectively.[224–226] The separation does not yet ensure their reliable identification, because these tags are not fundamental to molecules and are rather sensitive to experimental conditions, which may not be always reproduced with sufficient accuracy.[227–229] Optical spectroscopy, which is orthogonal to MS as well as to LC and IM separation techniques, is another method of structural identifications. When combined with cryogenic cooling of polyatomic ions, optical spectroscopy possesses the capability to map-up the ionic vibrational energy levels, which are characteristic of the 3D structure of the ions on a fundamental level.[85,141,161,176,177,230]

Here we present an approach to molecular structural identification, which is based on integration of broadband high-resolution MS with ultraviolet (UV) photofragmentation spectroscopy of cold ions. This fusion multiplies the selectivities of the two complementary techniques and produces highly specific two-dimensional (fragmentation yield vs UV wavenumber and m/z) fingerprints of ionic species. For instance, UV photofragmentation involves electronic excitation of ions and therefore may produce fragments specific to their structure, which are absent (or low abundant) in the products of thermal-like dissociation.[183,231] And compared with single-wavelength photofragmentation by VUV-UV light,[232–235] in the UV-MS approach the fragmentation patterns are recorded at several wavelengths within the absorption bands of UV chromophore groups. Vibrational resolution, achievable by cryogenic cooling of ions, makes these patterns sharply wavenumber-dependent, adding a great number of details into ionic fingerprints, which allows distinguishing very similar species, including isomers.

Once stored as libraries, these fingerprints can be used for qualitative and quantitative identification of ions in their mixtures employing numerical algorithms of matrix analysis. Moreover, in certain cases the decomposition of the UV-MS spectrum of a mixture of different molecules allows for recovery of their UV and MS identities without a use of libraries. Here we demonstrate the capability of the approach to identify ions in mixtures of different singly protonated isomers, which are particular difficult cases for MS even when coupled to LC or IM and for which the electron capture/transfer dissociation technique, for instance, is not applicable at all.

4.2. Library-based Approach

4.2.1. Data analysis

The library-based approach consists in measuring the 2D fingerprints of the species to be identified, constructing a library of the corresponding UV-MS matrices, and decomposing the recorded 2D spectra of the samples to be analyzed into a linear combination of the library matrices. This procedure is very similar to expressing a k-dimensional vector as a sum of k basis set vectors multiplied by the coefficients reflecting the coordinates of the vector relative to this basis.

At the first step, the 2D spectra of individual molecules are converted into matrices and subsequently processed using the MATLAB software suite, as described in the previous chapter. At the second step, we define a set of n wavenumbers, at which the 2D spectrum of a mixture will be analyzed, and generate a set of m m/z values of all the peaks detected in the UV-MS spectra of individual components. Each of the matrices is then mapped onto these new wavenumber and m/z scales in the following way: if an initial matrix contains a peak at a certain m/z (from the new set of m/z values), then the UV spectrum at this m/z (i.e., a column vector in the initial UV-MS matrix) is projected onto the new wavenumber scale by means of linear interpolation; otherwise, the respective column vector in the resulting matrix is filled with zeros. Thus, if we denote the number of individual components by k, the new matrices form a basis of a k-dimensional subset of a vector space of n-by-m matrices.

Now, consider a sample containing a mixture of these k species (some of them may be absent in the mixture, which corresponds to zero concentration). Its recorded 2D spectrum is converted into a UV-MS matrix, which is then baseline corrected and normalized to the total number of ions. Since the set of relevant m/z values is already defined, the peak-picking step is omitted and the UV spectra corresponding to these m/z values are directly interpolated on the prescribed wavenumber scale. The resulting n-by-m matrix should be a linear combination of the k library matrices:

$$D_M = \sum_{i=1}^{k} c_i \cdot D_i \qquad (4.1)$$

where D_M is the UV-MS matrix of the mixture, and c_i and D_i are the concentration and the UV-MS matrix of the i-th individual component, respectively.

However, due to noise in the experimental data and possible data processing errors, the system of $n \times m$ equations becomes inconsistent. A general approach to find an approximate solution of an overdetermined system (indeed, $n \times m \gg k$) is to minimize the sum of squared residuals with respect to variables. More specifically, we minimize the squared Frobenius norm of the residual matrix subject to non-negativity constraints:

$$\min_{x \geq 0} \left\| D_M - \sum_{l=1}^{k} x_l \cdot D_l \right\|_F^2 = \min_{x \geq 0} \sum_{i=1}^{n} \sum_{j=1}^{m} \left((D_M)_{ij} - \sum_{l=1}^{k} x_l \cdot (D_l)_{ij} \right)^2 \qquad (4.2)$$

using the MATLAB built-in function "lsqnonneg", and subsequently, to find the relative concentrations of the mixture components, the determined coefficients are divided by their sum:

$$\hat{c}_i = \frac{x_i}{\sum_{j=1}^{k} x_j} \qquad (4.3)$$

Finally, the accuracy of the identification of k individual components in s various mixtures is assessed in terms of the root-mean-square deviation (RMSD):

$$RMSD = \sqrt{\frac{1}{k \cdot s} \cdot \left(\sum_{i=1}^{s} \sum_{j=1}^{k} \left(c_j^{(i)} - \hat{c}_j^{(i)} \right)^2 \right)} \qquad (4.4)$$

which is a measure of discrepancy between the relative concentration $c_j^{(i)}$ of the j-th component in the i-th mixture and its predicted value $\hat{c}_j^{(i)}$ averaged over all the components and all the mixtures.

4.2.2. Positional isomers of phosphorylated TSAAATSY peptide

Protein and peptide phosphorylation is one of the most abundant natural post-translational modification (PTM), which drastically increases the diversity of their biological functions. Abnormal phosphorylation in living organisms is associated with certain cell pathologies, and therefore, the understanding and detection of this modification can be used for human disease diagnostics.[236,237] Methods of phosphorylation analysis include

several biochemical, wet techniques and high-resolution MS. Although tandem MS in combination with collision-induced dissociation (CID), higher-energy collisional dissociation (HCD), electron capture/transfer dissociation (ExD) techniques can accurately identify the presence of phosphorylated peptides,[237–241] there are still many unresolved challenges regarding unambiguous determination of phosphorylation sites.[237]

We have already demonstrated how one can employ the technique of cold ion spectroscopy for unambiguous identification of peptide phosphorylation on tyrosine (pTyr) residues.[183] However, peptides phosphorylated on serine (pSer) and threonine (pThr) do not exhibit any specific spectral features in the UV absorption or fragmentation mass spectra, which could be utilized for distinguishing them. Since 2D UV-MS fingerprints are very specific not only to the amino acid sequence of peptides but even to their 3D structure, one can build a library of such fingerprints of isomeric peptides and subsequently use it for identification of the peptides. Here we present the results of identification of phosphorylated octapeptides that differ only by their single phosphorylation site.

The amino acid sequence of the studied peptides is Thr-Ser-(Ala)$_3$-Thr-Ser-Tyr with a phosphate group located on one of the Thr/Ser/Tyr residues. There are five possible phosphorylation sites and hence five isomeric peptides to be identified. Figure 4.1(a) shows the HCD spectra of the five peptides. As one should expect, the most prominent fragments correspond to the neutral loss of phosphoric acid (H$_3$PO$_4$, 98 Da) from the parent ions for all the protonated peptides, except TSAAATSpY. Indeed, the phosphate group of peptides phosphorylated on Ser or Thr residues is relatively labile, which opens up an additional low-energy fragmentation pathway that competes with backbone fragmentation. The extent of the resulting neutral H$_3$PO$_4$ loss depends on various factors, such as the structure of a peptide, its charge state and the applied activation energy, but in general, the loss of H$_3$PO$_4$ is the dominant fragment in the CID mass spectra of serine- and threonine-phosphorylated peptides.[242] In tyrosine-phosphorylated peptides, the H$_3$PO$_4$ group cannot be eliminated from the Tyr residue, because of the aromatic character of its side chain.[243] Instead, the loss of HPO$_3$ (or even sequential loss of HPO$_3$ and H$_2$O) is observed, provided basic residues (i.e., lysine or arginine) are present in a protonated peptide,[244] which, however, is not the case for the TSAAATSpY peptide.

Figure 4.1. **(a)** HCD fragmentation mass spectra of five singly protonated phosphopeptides: pTSAAATSY (red), TpSAAATSY (blue), TSAAApTSY (green), TSAAATpSY (magenta), and TSAAATSpY (cyan). HCD collision energy was set to 40 eV. The rightmost peaks, labeled [M+H]$^+$, correspond to the parent ions, which all have the same m/z = 851.316 Th (mass accuracy is 2 ppm). Labels W and P denote the neutral losses of water and phosphoric acid, respectively. Vertical dashed lines show the correspondence between fragments of different peptides that have exactly the same m/z. For graphical clarity all the spectra are normalized by the intensities of the most abundant fragments. **(b)** Total ion current (TIC) chromatogram of a mixture of the five peptides (20 % each). LC-MS analysis was performed on an Acquity UPLC system (Waters) coupled to a Xevo G2-S QToF mass spectrometer (Waters). The peptide mixture was fractionated on a 2.1x50 mm, 1.8 µm Zorbax Extend-C18 column (Agilent). The mobile phase consisted of A (water containing 0.1 % formic acid) and B (acetonitrile containing 0.1 % formic acid). The gradient program was as follows: 2 % of B for 0.2 min; 2 % to 95 % of B in 3 min; 95 % of B for 1 min; 95 % to 2 % of B in 0.1 min; and 2 % of B for 1 min. The flow rate was 0.5 mL/min.

Apart from the H_3PO_4 group loss, characteristic of phosphorylated peptides, the HCD spectra display the series of b- and y-ions, typical for CID-like fragmentation. The mass difference of two consecutive b- or y-ions corresponds to an amino-acid residue mass so that the sequences of the peptides could, in principle, be derived from the fragmentation mass spectra. However, the peptides have a number of common fragments with exactly the same masses (e.g., the peptides pTSAAATSY and TpSAAATSY differ only by the masses of b_1 and y_7 ions), and the combination of backbone fragmentation with H_3PO_4 losses makes the HCD

spectra more complicated for *de novo* analysis. Finally, there is another fragmentation channel specific for serine- and threonine-containing peptides, namely, the neutral loss of water (H_2O, 18 Da).[245] The elimination of H_2O from Ser and Thr residues gives rise to dehydroxyalanine and dehydrobutyrine, respectively, which are essentially the same products as for the fragmentation through the neutral loss of the H_3PO_4 group from pSer and pThr residues. Thus, identification of these five isomeric peptides (may be, except for the TSAAATSpY peptide, which has a very distinct fragmentation pattern) only on the basis of their HCD (or CID) mass spectra becomes challenging. It is worth noting that the ExD technique, which preserves phosphorylation and generates sequence-specific c- and z-ions,[242] cannot be employed in this case, because the studied peptides are singly protonated and ExD is feasible only for multiply charged species (even for doubly charged peptides the ExD efficiency is quite low).

For the reasons discussed above and, in general, because of the complexity of biological samples, a MS-based characterization of phosphorylated peptides is usually complemented by a preliminary separation stage, such as high-performance liquid chromatography (HPLC).[237] We performed a standard ultra-performance liquid chromatographic (UPLC) analysis of an equimolar mixture of all five peptides, and the recorded total ion chromatogram is shown in Figure 4.1(b). Only the TSAAATSpY peptide could be reliably separated from the others, while the pairs of the peptides pTSAAATSY/TpSAAATSY and TSAAApTSY/TSAAATpSY eluted almost simultaneously and the peptides within each pair turned to be virtually indistinguishable. Of course, the UPLC analysis was not specific for the given peptides and can be further optimized to separate them but at the price of losing universality (in fact, most of the fractionation techniques have been developed for general separation of analytes and not for the analysis of certain species). Still, it may not ensure their unambiguous identification due to the known issues in reproducibility of retention times, including the phenomenon of peak inversion.[228,229]

Our 2D UV-MS approach to identification of isomeric species does not require the separation stage prior to MS analysis but rather analyzes the entire mixture at once. For this, one has to build a library of the UV-MS fingerprints of individual components first. The UV-MS spectra were recorded individually for each of the studied here phosphorylated peptides, and Figure 4.2(a,b) shows their cumulative UV absorption and photofragmentation mass spectra. The photofragmentation mass spectra display a number of specific fragments that are not

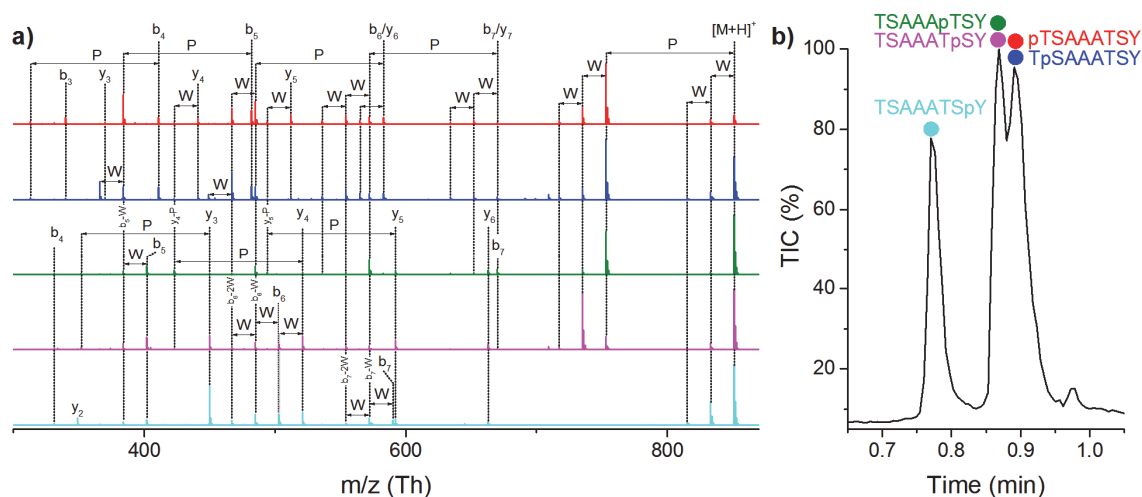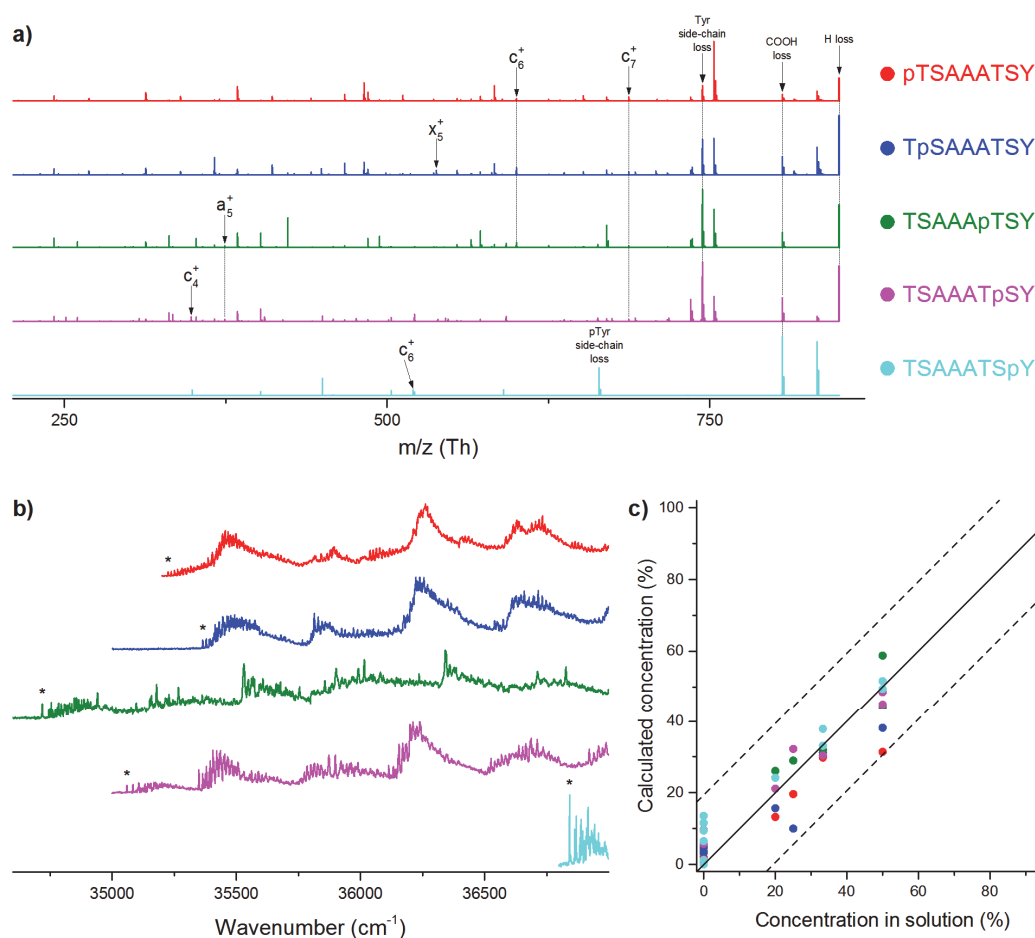**Figure 4.2.** **(a)** Photofragmentation mass spectra and **(b)** UV absorption spectra of five singly protonated phosphopeptides, pTSAAATSY (red), TpSAAATSY (blue), TSAAApTSY (green), TSAAATpS (magenta), and TSAAATSpY (cyan), calculated by integrating their UV-MS spectra either along the UV wavenumber or along the m/z dimension, respectively. **(a)** Labels designate the peaks that are not present in the HCD fragmentation mass spectra of these peptides. All other peaks correspond to b- and y- ions, neutral losses of water and phosphoric acid, and their combinations. Vertical dashed lines show the correspondence between fragments of different peptides that have exactly the same m/z. For graphical clarity the peaks corresponding to parent ions are not shown and all the spectra are normalized by the intensities of the most abundant fragments. **(b)** Asterisks indicate the position of the UV absorption band origins of the peptides. For each spectrum, the spectra are normalized to the maximum signal. **(c)** Calculated relative concentrations of the five phosphopeptides (dots) as a function of their relative concentrations in solution (0–100 %) for 12 different mixtures. The solid line shows the ideal 1:1 dependence, and the area between the two dashed lines corresponds to the 99.7 % confidence band (3 × RMSD) for calculating concentrations.

present in the HCD spectra of these peptides. For instance, the fragment at m/z = 744.267 Th corresponds to the cleavage of the C_α-C_β bond of the Tyr side chain and is a common product of photodissociation of Tyr-containing peptides.[168,170,183,246,247] The corresponding fragment due to the loss of the pTyr side chain of the TSAAATSpY peptide appears at m/z = 664.230 Th. Owing to the high mass resolution provided by the Orbitrap analyzer, another abundant fragment could be detected at m/z = 850.309 Th, which corresponds to the loss of the H atom. Although the H-loss was not directly observed in the case of protonated tyrosine, it was shown to be the dominant photodissociation channel of protonated tryptophan.[248] Interestingly, the H-loss fragment is absent in the photofragmentation spectrum of TSAAATSpY, while it is one of the most prominent photofragments for the other four peptides. Both side-chain loss and H-loss are well-known products of prompt photodissociation, a process that occurs on the electronically excited state directly after the UV excitation of a chromophore. However, the fragmentation mass spectra contain a few more UV-specific fragments that are typically not observed in the UV photodissociation of protonated peptides. They were unambiguously assigned (see the labels in Figure 4.2(a)) on the basis of their accurate masses ($\delta m/m$ < 10 ppm). For example, the most intense peak in the mass spectrum of the TSAAATSpY peptide at m/z = 806.320 Th corresponds to the neutral loss of the COOH group.

Although the presence of UV-specific fragments apparently improves the specificity of UV-MS spectra, it is the UV absorption spectra that make the 2D fingerprints unique. Indeed, the UV spectra of the five peptides, shown in Figure 4.2(b), even visually look different. For instance, the band origin of TpSAAATSY is located at 35365.3 cm^{-1}, which is only slightly red-shifted with respect to that of neutral tyrosine.[54] This suggests that in TpSAAATSY the Tyr side chain is relatively remote from the rest of the ion. On the contrary, the band origin of TSAAApTSY (34719.3 cm^{-1}) is significantly shifted to lower wavenumbers. Such a large spectral shift was previously observed for the protonated Ala-Tyr dipeptide and rationalized in terms of the interaction between the π-electron cloud of the Tyr chromophore and the protonated N-terminus.[170] Thus, one may assume that in the case of TSAAApTSY phosphorylation on the Thr residue alters the hydrogen-bond pattern and hence the 3D structure of the ion in such a way that the protonated N-terminus becomes located in close proximity to the Tyr side chain. In general, regardless of the exact relationship between the 3D structure of an ion and its UV

spectrum, the position of the UV band origin and the vibronic structure of UV spectra are known to be highly sensitive to the local environment of chromophores.[54,168,170,249,250]

Thus, in this case, both photofragmentation mass spectra and UV absorption spectra contribute to the specificity of the recorded UV-MS fingerprints. Then, in order to assess the performance of the library-based approach, a set of 12 different mixtures of the five phosphorylated peptides was prepared with the total peptide concentration of 20 μM and the relative concentration of each peptide ranging from 0 to 100 %. The UV-MS matrix for each mixture was constructed by measuring the UV-MS spectra within two spectral regions, 34600-35800 cm^{-1} and 36800-37000 cm^{-1}, and then merging them together. The merged UV-MS matrices were decomposed in the basis set of the library matrices, and the results of the decomposition are shown in Figure 4.2(c). The library-based analysis correctly identified the peptides in all the mixtures and reproduced the relative concentrations of the peptides with an RMSD of 6.5 %. Although the RMSD value is low enough, the origin of the discrepancy between the calculated concentrations and the concentrations in solution still has to be investigated in more detail to improve the performance of the identification.

4.2.3. Stereoisomers of [Ala^2, Leu^5]-enkephalin

Racemization is another type of PTM. Except for glycine, all amino acids can exist in the form of either of two isomers, called L- and D-forms. While all amino acids found in proteins during the biosynthesis in ribosomes are present in the L-form, D-amino acids may slowly accumulate in biological systems during their aging.[251–253] For instance, racemization of aspartic acid, which occurs much faster than for any other amino acid, can be employed for age estimation in forensic sciences.[252,254] Here we present the results of identification of the Tyr-Ala-Gly-Phe-Leu peptide with all naturally left-handed amino acids ([L-Ala^2, L-Leu^5]-enkephalin) and its stereoisomer with the Ala and Leu residues in the right-handed form ([D-Ala^2, D-Leu^5]-enkephalin). As the peptides differ in the chirality of only two of five stereocenters, they are not mirror images of each other, that is, they are diastereomers. However, it should be challenging to distinguish these isomers by MS only, because they have exactly the same sequence of bonded atoms, and therefore, their fragmentation mass spectra are likely to be very similar.
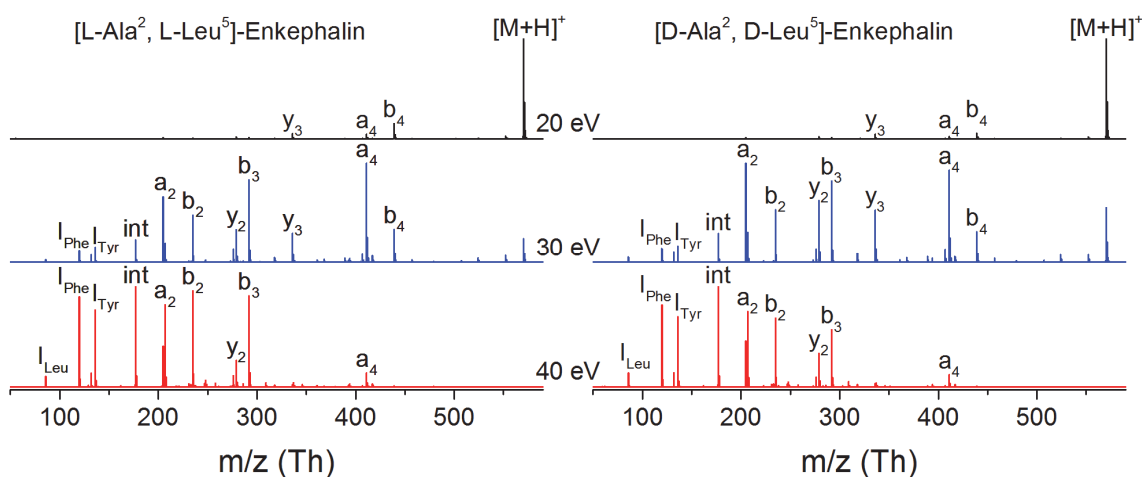
Figure 4.3. HCD fragmentation mass spectra of two singly protonated stereoisomers of [Ala^2, Leu^5]-enkephalin. HCD collision energy was set to 20, 30 and 40 eV. The rightmost peaks, labeled [M+H]$^+$, correspond to the parent ions with m/z = 570.291 Th (mass accuracy is 2 ppm). Labels int denote an internal fragment, and labels I$_{Tyr}$ and I$_{Phe}$ denote the immonium fragments corresponding to the Tyr and Phe residues, respectively. For graphical clarity all the spectra are normalized by the intensities of the most abundant ions.

Figure 4.3 shows the HCD fragmentation spectra recorded at the collision energies of 20, 30 and 40 eV. And indeed, both peptides exhibit the same fragmentation patterns, similar to those previously measured for leucine enkephalin (Tyr-Gly-Gly-Phe-Leu).[255] At a relatively low collision energy the most abundant fragments are the a$_4$ and b$_4$ ions, which correspond to the lowest-barrier dissociation channels. Upon increase of the activation energy, the fragmentation pattern shifts towards lower m/z values due to consecutive fragmentation. Also, the higher-energy activation gives rise to the Tyr and Phe immonium ions. Thus, the two isomers have almost identical fragmentation spectra, which only slightly differ in the relative intensities of the fragment peaks.

Since the peptides are practically indistinguishable by MS, a use of a complementary technique is required to separate them prior to MS analysis. Besides liquid chromatography, which can, in principle, be used for separating stereoisomers,[224] there is another technique, namely, ion mobility spectrometry (IMS) capable of doing this. It was demonstrated that both conventional IMS and differential IMS enable separation of peptides in quite complex mixtures.[225,256] It was also shown that a combination of LC and IMS can be particularly useful
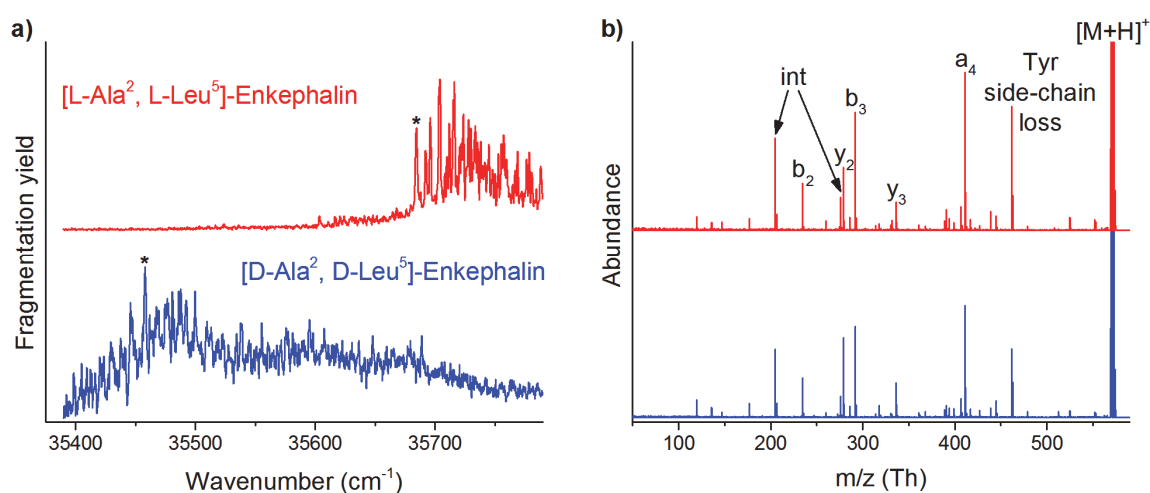
Figure 4.4. **(a)** UV absorption and **(b)** photofragmentation mass spectra of [L-Ala^2, L-Leu^5]-enkephalin (red) and [D-Ala^2, D-Leu^5]-enkephalin (blue). UV spectra are the slices of the respective UV-MS spectra taken at the Tyr side-chain loss fragment (m/z = 462.238 Th, mass accuracy is 7 ppm) and normalized to the maximum signal. Photofragmentation mass spectra were calculated by averaging the respective UV-MS spectra over the UV peaks indicated by the asterisks. Labels designate the parent ions ([M+H]$^+$) and the prominent photofragments. For graphical clarity the [M+H]$^+$ peaks are cut off (their intensity is 200 times higher).

for resolving mixtures of isomeric peptides.[223,257] However, it turned out that the difference in collisional cross sections of the studied here stereoisomers is so small that a conventional IMS with a resolving power of 80 could not attain a baseline-separation and only partially resolve them.[226]

To test the ability of the 2D UV-MS approach to discriminate between [L-Ala^2, L-Leu^5]-enkephalin and [D-Ala^2, D-Leu^5]-enkephalin, we first measured their individual fingerprints. The UV absorption spectra of the two peptides, shown in Figure 4.4(a), are strikingly different. While the spectrum of [L-Ala^2, L-Leu^5]-enkephalin exhibits a well-resolved vibronic structure in the region near its band origin, the spectrum of its isomer looks rather congested and is red-shifted by approximately 200 cm^{-1}. It is likely that such dramatic spectral changes are caused by a substantial alteration of the 3D structure of the peptide upon replacement of the L-amino acids by their D-forms. On the other hand, this results in only a 1.1 % relative difference in collisional cross sections of the two peptides.[226] In general, this
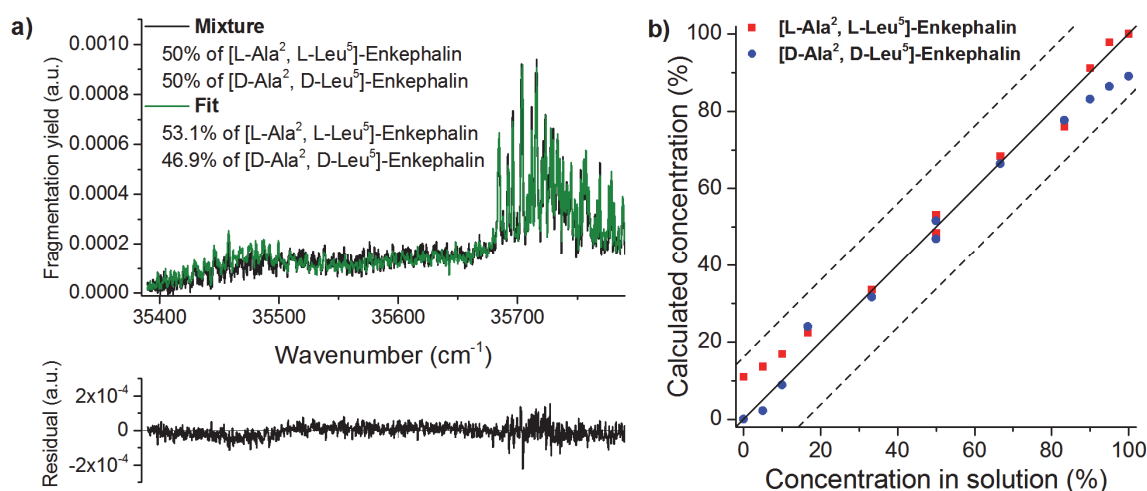
Figure 4.5. **(a)** UV absorption spectrum (black) of an equimolar mixture of two stereoisomers of [Ala^2, Leu^5]-enkephalin, taken as a slice of the experimental UV-MS spectrum at the Tyr side-chain loss fragment, and its fit (green) by a linear combination of the UV spectra of the individual components (see Figure 4.4(a)). The resulting residual plot is shown at the bottom. **(b)** Calculated relative concentrations of [L-Ala^2, L-Leu^5]-enkephalin (red squares) and [D-Ala^2, D-Leu^5]-enkephalin (blue circles) as a function of their relative concentrations in solution (0–100 %) for 12 different mixtures. The solid line shows the ideal 1:1 dependence, and the area between the two dashed lines corresponds to the 99.7 % confidence band (3 × RMSD) for calculating concentrations.

example illustrates the much higher sensitivity of UV spectra compared with collisional cross section values to changes in a molecular structure. The photofragmentation mass spectra (Figure 4.4(b)), in contrast, resemble the HCD fragmentation spectra, specifically those recorded at a collisional energy of 30 eV. Although the UV excitation leads to the appearance of the specific Tyr side-chain loss fragments, the relative abundance of these peaks in the mass spectra is almost the same.

The library-based identification of the two peptides was performed for a set of 12 different mixtures with the total peptide concentration of 10 μM and the relative peptide concentrations ranging from 0 to 100 %. The UV-MS spectrum of each mixture was measured within the 35350-35800 cm^{-1} range and decomposed in the basis of the UV-MS matrices of the individual isomers. The photofragmentation mass spectra do not contribute, at least significantly, to the specificity of the UV-MS fingerprints of the isomers. Therefore, the quality of the matrix-based decomposition in this case can be readily assessed by comparing the UV

spectrum of a mixture with a linear combination of the UV spectra of the isomers taken with coefficients returned by the decomposition algorithm. For example, such a linear combination of the UV spectra of [L-Ala^2, L-Leu^5]-enkephalin and [D-Ala^2, D-Leu^5]-enkephalin shown in Figure 4.5(a) almost perfectly fits the UV spectrum of an equimolar mixture of these peptides, which validates the performed decomposition. The results of the library-based analysis of the 12 mixtures are presented in Figure 4.5(b). The peptides were correctly identified in all the mixtures, and their relative concentrations were reproduced with an RMSD of 5.4 %. Moreover, the data points corresponding to 7 of 12 mixtures are located very close to the ideal identity line, while the other points noticeably deviate from it, and the higher the relative concentration of [D-Ala^2, D-Leu^5]-enkephalin is, the more is the deviation. Such a trend indicates the presence of a systematic error and is probably due to an inaccuracy in the composition of these mixtures. Thus, one may expect that the ultimate accuracy of the isomer identification is much better than 5.4 %.

4.2.4. Stereoisomers of ephedrine

Identification of isomeric compounds is required not only in the field of proteomics but also in pharmacology, doping control, and forensics. Indeed, isomers of many drugs exhibit different pharmacokinetic and/or pharmacodynamics properties.[258,259] Isomers can differ in
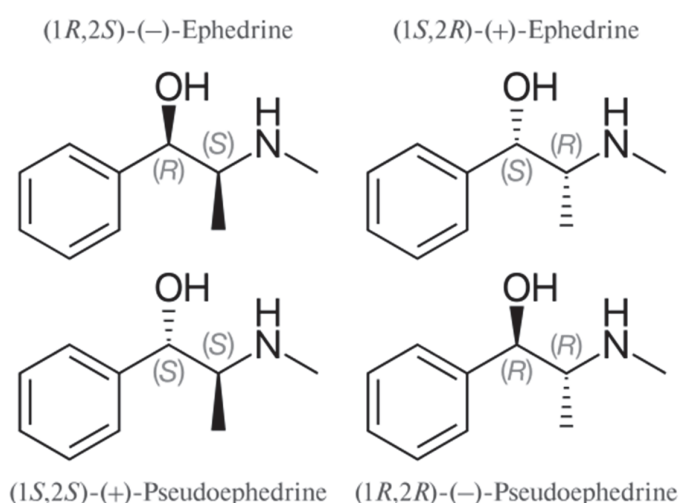
Figure 4.6. Four stereoisomers of ephedrine. By convention, the pair of enantiomers (1R,2S)-ephedrine and (1S,2R)-ephedrine is called ephedrine, while the pair of enantiomers (1R,2R)-ephedrine and (1S,2S)-ephedrine is called pseudoephedrine.

their rates of absorption,[260] volumes of distribution[261], metabolic pathways,[262] and pharmacological activity.[258] For example, ephedrine, which is an alkaloid found in plants of *Ephedra* type and used for the treatment of asthma, nasal congestion, and obesity, has two stereogenic centers and hence exists as four stereoisomers (Figure 4.6). While the mechanism of action of all the isomers of ephedrine involves activation of α- and β-adrenergic receptors,[263] their potency varies by 2-3 orders of magnitude.[264] Being a central nervous system stimulant, ephedrine is prohibited in sport by the World Anti-Doping Agency (WADA),[265] and the inequivalence in the potency of the ephedrine isomers is reflected in the levels at which they should be detected: the decision limit for reporting an Adverse Analytical Finding for ephedrine (the pair of enantiomers (1R,2S)-ephedrine and (1S,2R)-ephedrine) is 11 µg/mL, and for pseudoephedrine (the pair of enantiomers (1R,2R)-ephedrine and (1S,2S)-ephedrine) it is 170 µg/mL.[266] Here we use the library-based analysis not only for qualitative but also for quantitative identification of two diastereomers of ephedrine, namely, (+)-ephedrine and (+)-pseudoephedrine.

Since the isomers studied here are diastereomers, one may expect that, as in the case of the diastereomers of [Ala^2, Leu^5]-enkephalin, it should be challenging to identify them by MS only. Indeed, their HCD spectra, presented in Figure 4.7, have very similar fragmentation
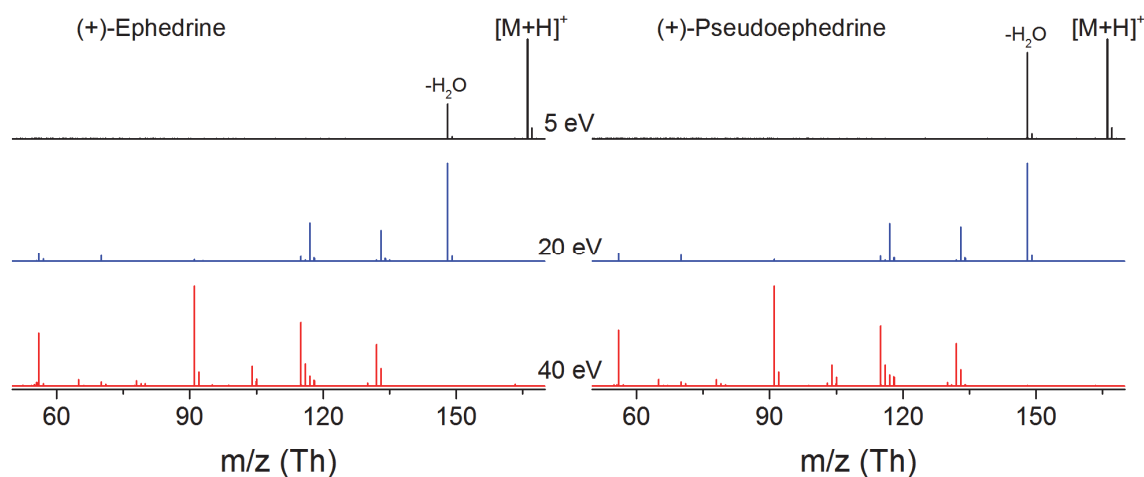
Figure 4.7. HCD fragmentation mass spectra of two singly protonated stereoisomers of ephedrine. HCD collision energy was set to 5, 20 and 40 eV. The rightmost peaks, labeled [M+H]$^+$, correspond to the parent ions with m/z = 166.123 Th (mass accuracy is less than 1 ppm). For graphical clarity all the spectra are normalized by the intensities of the most abundant ions.

patterns. The predominant fragment at low collision energies corresponds to the neutral water loss, which at higher activation energies undergoes a secondary fragmentation, thus yielding low m/z fragments.[267] Because the resulting mass spectra differ only in the relative intensities of the fragment peaks, the isomers have to be fractionated prior to MS analysis. This requires developing methods specific to this isomer pair. For instance, an UPLC analysis employed for the simultaneous detection of 63 stimulants in urine was not able to separate ephedrine stereoisomers.[268] Therefore, several LC-based methods have been developed specifically for the determination of ephedrine and pseudoephedrine in biological samples.[268–272] On the contrary, our approach is based on direct measurements of fundamental characteristics of ions and should be equally applicable both to peptides and to ephedrines.

First, of course, we measured the 2D UV-MS fingerprints of (+)-ephedrine and (+)-pseudoephedrine. Their UV absorption spectra, shown in Figure 4.8(a), display a nicely resolved structure. The UV band origin of (+)-ephedrine, at least, of its major conformer (this is the most intense peak) at 37550.4 cm^{-1} is shifted by only ≈ 2 cm^{-1} with respect to that of the neutral ephedrine, studied in a supersonic jet-cooled molecular beam by resonance two-photon ionization (R2PI) spectroscopy.[273,274] On the other hand, the two lowest-energy peaks of (+)-pseudoephedrine at 37562.0 cm^{-1} and 37575.6 cm^{-1}, which are likely to be the band origins of two different conformers, are both shifted by more than 10 cm^{-1} with respect to the band origin of any of the neutral pseudoephedrine conformers.[273,275] To assess the structural implications of these spectral shifts, one may compare them to the known data for protonated phenylalanine, which is comparable in size to ephedrines and possesses essentially the same chromophore. Two conformers of protonated phenylalanine were previously characterized by a use of IR-UV depletion spectroscopy.[168] It was shown that the two conformers differ in the orientation of the backbone with respect to the side chain, namely, one of them corresponds to *anti* and the other to *gauche* conformations. Such a notable structural difference, at least, for a relatively small molecule results in a rather modest spectral shift of 8.7 cm^{-1}. We may thus conclude that the protonation of (+)-pseudoephedrine substantially alters its 3D structure, while (+)-ephedrine is likely to have a structure similar to that of the major conformer of neutral ephedrine, although measurements of conformer-selective IR spectra and high-level quantum chemical calculations are required for the 3D structure elucidation.
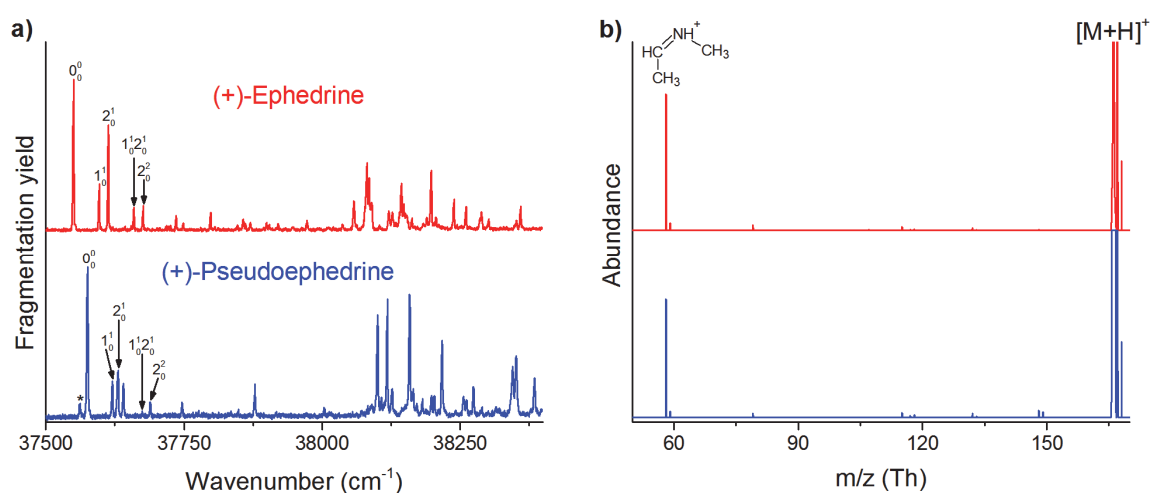
Figure 4.8. UV absorption (left panel) and photofragmentation mass spectra (right panel) of (+)-ephedrine (red) and (+)-pseudoephedrine (blue). UV spectra are the slices of the respective UV-MS spectra taken at the dominant fragment (m/z = 58.066 Th, mass accuracy is 15 ppm) and normalized to the maximum signal. Labels designate the 0-0 transitions as well as the first several vibronic transitions. The 0-0 transition of a minor conformer of (+)-pseudoephedrine is indicated by an asterisk. Photofragmentation mass spectra were calculated by averaging the respective UV-MS spectra over the UV band origins (0_0^0). For graphical clarity the peaks corresponding to the parent ions ([M+H]$^+$) are cut off (their intensity is 100 times higher).

Regardless of any knowledge of the exact structure of ephedrines, clearly the stereochemistry of the α-carbon significantly affects the UV spectra of the isomers. First, the band origin of the major conformer of (+)-pseudoephedrine is blue-shifted by 25.2 cm^{-1} with respect to that of (+)-ephedrine. Second, the low-frequency Franck-Condon active modes of the two isomers have slightly different frequencies: 46.6 and 63.1 cm^{-1} in (+)-ephedrine and 45.2 and 55.6 cm^{-1} in (+)-pseudoephedrine. Finally, since the frequency of the first Franck-Condon active mode in (+)-pseudoephedrine is 45.2 cm^{-1}, the transition at 37562.0 cm^{-1} should be assigned to the band origin of a minor conformer rather than to a hot band. Thus, the difference in the stereochemistry makes (+)-pseudoephedrine more flexible (in terms of conformational diversity) compared with (+)-ephedrine. Similarly, only two conformers of neutral ephedrine were observed experimentally, while four distinct conformers of neutral pseudoephedrine were unambiguously identified.[273] To sum up, it is the high sensitivity of UV spectra to any changes in a 3D structure in combination with a high spectral resolution,

provided by cryogenic cooling, that makes the UV-MS spectra indeed distinct and enables using them as fingerprints for distinguishing the isomers.

On the contrary, as might be expected, the photofragmentation mass spectra of the two isomers are almost identical (Figure 4.8(b)). Interestingly, the mass spectra are dominated by only one fragment, which is not observed in CID-like fragmentation (there is a fragment in the HCD spectra at m/z = 56.051 Th, but the photofragment is at m/z = 58.066 Th). This indicates that there should be a very efficient process that occurs on the excited-state potential energy surfaces and leads to prompt dissociation. The fragment at m/z = 58 Th was previously observed in the electron ionization (EI) mass spectra[276] and upon resonant two-photon ionization of neutral ephedrines.[273–275,277] In both cases the first step is the ionization of neutral ephedrines, followed by a cleavage of the carbon-carbon bond adjacent to the nitrogen.[278] However, it is very unlikely that the UV excitation results in ionization of the ions. The reported ionization energy (IE) of neutral ephedrine is 8.3 eV,[274] and the IE of protonated ephedrines is expected to be higher. For instance, for polypeptides the average IE increases by $\approx$ 1.1 eV per each additional proton.[279] Taking into account that this trend was attributed to Coulombic repulsion, for a small molecule such as ephedrine this increase should be, at least, the same or even more pronounced. Therefore, ionization of protonated ephedrines requires more than 9.4 eV, while the energy of the two UV photons corresponding to the band origin of (+)-ephedrine is 9.3 eV. Moreover, both EI and R2PI produce a number of other fragments that are not present in the photofragmentation spectra in Figure 4.8(b) or, at least, are much less abundant than the 58 Th fragment. Thus, we believe that the photodissociation mechanism of protonated ephedrines differs from that of neutral ephedrines, although the resulting fragment has the same nominal mass (58 Da) and the same elemental composition ($C_3H_8N^+$), which we determined on the basis of the measured accurate m/z value.

A possible explanation for the observed fragmentation can be drawn from a comparison of protonated ephedrines and protonated phenylalanine. Indeed, they both have a protonated amino group and an unsubstituted phenyl ring separated by two carbon atoms. The most abundant photofragment of protonated phenylalanine is due to the neutral loss of its side chain, that is, due to the cleavage of the C_α-C_β bond.[168] The same photofragmentation channel was observed in the case of protonated Phe-containing peptides,[178,179,246,280,281] and it was shown that the mechanism of this specific photodissociation involves triplet states.[246]

Due to structural similarity between ephedrines and phenylalanine, one may expect that photofragmentation of ephedrines, which also leads to the C_α-C_β bond cleavage, is mediated by triplet states, which are not accessible upon collisional activation. In such a case, it might be worth investigating the photofragmentation mechanism of protonated ephedrines in more detail, because it can be a very general mechanism of prompt photodissociation of molecules with aromatic chromophores.

Once the 2D fingerprints of (+)-ephedrine and (+)-pseudoephedrine were recorded, we prepared a set of 30 different mixtures with the total concentrations ranging from 40 ng/mL ($2.4 \cdot 10^{-7}$ M) to 16.8 µg/mL ($1.0 \cdot 10^{-4}$ M) and the relative isomer concentrations ranging from 0 to 100 %. The UV-MS spectra of all the mixtures were measured within 37540-35650 cm^{-1} range because in this region the UV peaks of the isomers do not overlap, which ensures the maximum selectivity. The results of decomposition are presented in Figure 4.9. First, as in the
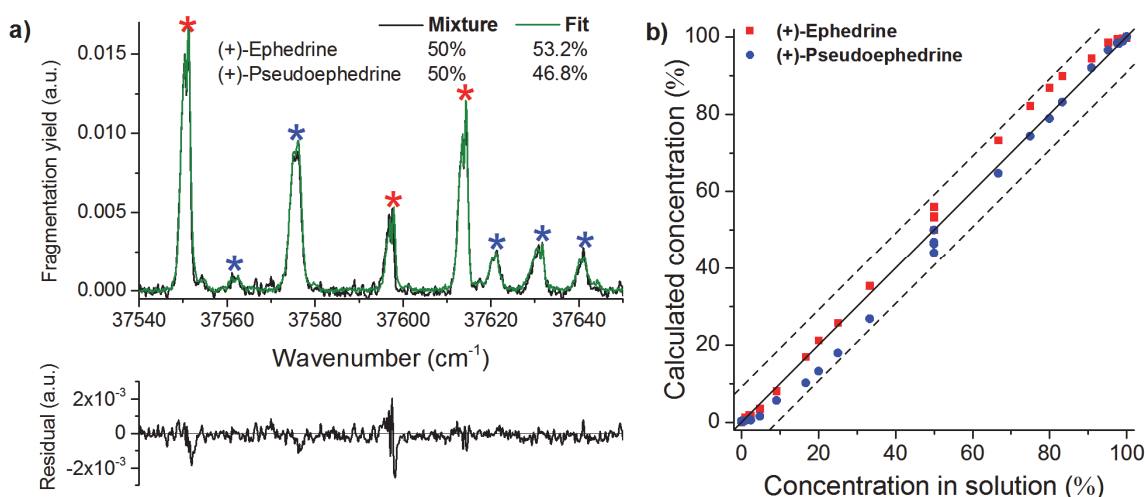
Figure 4.9. **(a)** UV absorption spectrum (black) of an equimolar mixture of two stereoisomers of ephedrine taken as a slice of the experimental UV-MS spectrum at m/z = 58.066 Th, and its fit (green) by a linear combination of the UV spectra of the individual component (see Figure 4.8(a)). The resulting residual plot is shown at the bottom. The peaks corresponding to (+)-ephedrine and (+)-pseudoephedrine are indicated by the red and blue asterisks, respectively. **(b)** Calculated relative concentrations of (+)-ephedrine (red squares) and (+)-pseudoephedrine (blue circles) as a function of their relative concentrations in solution (0–100 %) for 30 different mixtures. The solid line shows the ideal 1:1 dependence, and the area between the two dashed lines corresponds to the 99.7 % confidence band (3 × RMSD) for calculating concentrations.

case of [Ala^2, Leu^5]-enkephalins, the quality of the matrix-based decomposition can be assessed by comparing the UV spectrum of a mixture and its fit by a linear combination of the UV spectra of the isomers. Figure 4.9(a) displays a perfect match obtained for the decomposition of an equimolar mixture. Every single peak is clearly resolved and fitted by the corresponding peak of either (+)-ephedrine or (+)-pseudoephedrine. Unfortunately, the plot of the calculated concentrations of the two isomers versus their relative concentrations (Figure 4.9(b)) in solution reveals the same deficiency as in the case of [Ala^2, Leu^5]-enkephalins: about half of the data points almost perfectly lie on the identity line, while the others noticeably deviate from it. We inspected in detail the performed decompositions and did not find any signs of an imprecision of the numerical analysis itself. Thus, we have to conclude that this discrepancy is due to systematic errors in sample preparation. Nonetheless, the overall RMSD calculated for the entire set of mixtures is 3.1 %, which is a reasonably low value.

However, from a practical point of view, one needs to determine the absolute concentrations of analytes rather than their relative concentrations. Here, we assume that the total concentration of the two isomers can be readily determined using an external or internal standard.[282] Then, the relative concentrations of the isomers can be converted to the respective absolute concentrations by multiplying them by the total concentrations. Figure 4.10(a) demonstrates a rather good linearity of determining the absolute concentration of (+)-ephedrine within the entire range of concentrations, while there is a tendency to underestimate the absolute concentration of (+)-pseudoephedrine when its concentration in solution is low (< 1 µg/mL). Note that logarithmic scales are used for both axes because the range of concentrations spans approximately three orders of magnitude. The resulting RMSD value for determining the absolute concentrations of the isomers is 0.1 µg/mL for the whole set of the mixtures. For comparison, according to WADA requirements, the concentration of ephedrines has to be determined with an absolute uncertainty not exceeding 0.5 µg/mL.[266]

Obviously, the performance of the analysis should be concentration-dependent and cannot be described by only one value, RMSD. The relative error in determining the absolute concentrations of the isomers as a function of their concentrations in solution is shown in Figure 4.10(b). Indeed, the relative error increases with lowering the concentration of either
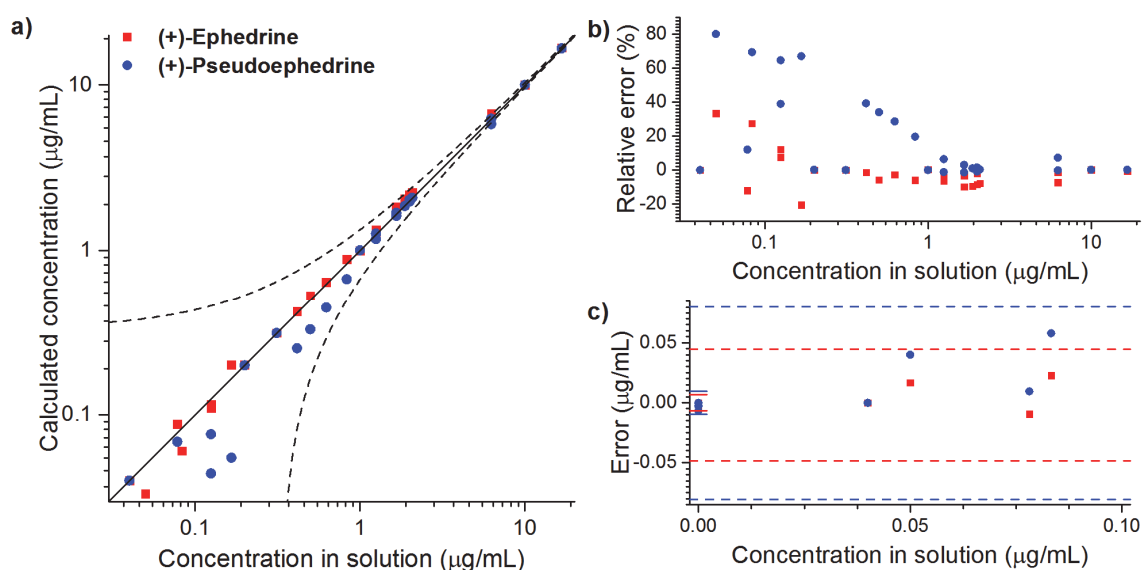
Figure 4.10. **(a)** Calculated absolute concentrations of (+)-ephedrine (red squares) and (+)-pseudoephedrine (blue circles) as a function of their absolute concentrations in solution for 30 different mixtures. The solid line shows the ideal 1:1 dependence, and the area between the two dashed lines corresponds to the 99.7 % confidence band (3 × RMSD) for calculating concentrations. **(b)** The corresponding relative errors of the calculated concentrations with respect to the concentrations in solution. **(c)** The absolute errors of the calculated concentrations in the limit of low concentrations (< 0.1 µg/mL). The solid lines around zero concentration show the confidence bands (3 × RMSD) for determining concentrations in blank samples and the dashed lines show the confidence bands (3 × RMSD) for determining concentrations in the limit of low concentrations.

(+)-ephedrine or (+)-pseudoephedrine. Although at very low concentrations the relative error can be as large as 80 %, it does not exceed 100 %. Qualitatively, this means that, even when the concentration of one of the isomers is 40 ng/mL, we can still detect its presence in a sample. For a quantitative assessment, one has to introduce the limit of blank (LoB) and the limit of detection (LoD).[283]

First, even the analysis of a blank sample (i.e., a sample that does not contain an analyte) may result in its non-zero detected concentration. Assuming a Gaussian error distribution, at the 95 % confidence level the LoB is:[283]

$$LoB = \langle c_{blank} \rangle + 1.645 \cdot \sigma_{blank} \qquad (4.5)$$

Table 4.1. Limit of blank, limit of detection and dynamic range estimated for (+)-ephedrine and (+)-pseudoephedrine.

	(+)-ephedrine	(+)-pseudoephedrine
LoB (in ng/mL)	4.9	7.6
LoD (in ng/mL)	29.3	51.7
Dynamic range	341	193

where $\langle c_{blank} \rangle$ and σ_{blank} are the mean and the standard deviation of the calculated analyte concentrations for a set of blank samples. That is, 95 % of the concentrations measured for blank samples will not exceed the LoB. Therefore, the analyte concentration should be, at least, higher than the LoB to be distinguished from "analytical noise". Then, to take into account the uncertainty associated with the analyte actually present in a sample, one defines the LoD as follows:[283]

$$LoD = LoB + 1.645 \cdot \sigma_{low} \qquad (4.6)$$

where σ_{low} is the standard deviation calculated for a set of samples containing a very low concentration of the analyte. Only if the analyte concentration is higher than LoD, it can be reliably, at the 95 % confidence level, distinguished from "analytical noise". The estimated LoB and LoD values for the studied here ephedrines are summarized in Table 4.1.
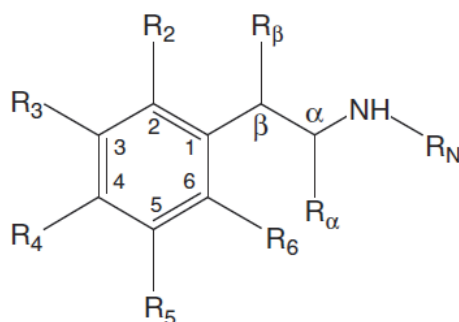
Figure 4.11. General structural formula of substituted β-phenylethylamines. Labels R_x, where x = N, α, β, 2, 3, 4, 5, 6, denote possible substitution sites. In ephedrines $R_N = R_\alpha = CH_3$, $R_\beta = OH$ and $R_2 = R_3 = R_4 = R_5 = R_6 = H$.

It turned out that the library-based analysis is capable of selectively detecting the diastereomers of ephedrine at concentrations much lower than required by WADA. However, ephedrines belong to a more general category of prohibited substances, namely, to the category of stimulants (S6),[265] which are to be routinely detected at a level of 100 ng/mL. Thus, in principle, one can record a comprehensive library of the 2D UV-MS fingerprints of all the stimulants specified by WADA and use it in doping control. Moreover, ephedrine is a representative of the class of substituted β-phenylethylamines (Figure 4.11), many of which are psychoactive drugs. For instance, ephedrine differs from methamphetamine, a strong central nervous system stimulant illegal in many countries,[284] only by the presence of a hydroxyl group. Because of the enormous diversity of substituted β-phenylethylamines (8 positions available for substitution), the high selectivity provided by the 2D UV-MS approach can be particularly useful in forensic chemistry.

4.3. Library-free Approach

4.3.1. Non-negative matrix factorization

An apparent inconvenience of the library-based approach is the need to build a library of the 2D fingerprints of individual compounds prior to their identification in a mixture. However, in certain cases the number of mixture components can be estimated and their UV and mass spectra can be derived directly from the UV-MS spectrum of the mixture.

Consider a mixture and its preprocessed (as described in the previous chapter) UV-MS matrix D. The n-by-m non-negative matrix D can always be factorized into a non-negative n-by-k matrix W of UV spectra and a non-negative k-by-m matrix H of mass spectra, where their common dimension k corresponds to an estimated number of components. Then, the matrix product $W \cdot H$ is a lower-rank approximation to D.[285,286] Since all the matrices, D, W and H, are non-negative, this approach was called non-negative matrix factorization (NMF).[287] One can find the best k-rank approximation by minimizing the Frobenius norm of the residual:

$$\|D - W \cdot H\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} \left(D_{ij} - \sum_{l=1}^{k} W_{il} \cdot H_{lj} \right)^2} \qquad (4.7)$$

For a given k, the MATLAB built-in function *nnmf*, which employs the iterative alternating least-squares algorithm, performs such factorization. The maximum number of iterations for each factorization procedure was 1000 (typically the algorithm converges much faster, after about 100 iterations). In order to find the global solution, the factorization of the UV-MS matrix was repeated 100 times for each rank k. The accuracy of the NMF can be assessed as the root-mean-square deviation (RMSD) between the matrix D and its k-rank approximation:

$$RMSD(k) = \frac{\|D^{n \times m} - W^{n \times k} \cdot H^{k \times m}\|_F}{\sqrt{n \cdot m}} \qquad (4.8)$$

where superscripts denote the size of the matrices. Instead of absolute RMSD values one can also use the relative RMSD, which, apparently, does not depend on the matrix size:

$$RMSD_{rel}(k) = \frac{RMSD(k)}{RMSD(0)} = \frac{\|D^{n \times m} - W^{n \times k} \cdot H^{k \times m}\|_F}{\|D^{n \times m}\|_F} \qquad (4.9)$$

RMSD is a good measure of accuracy of the factorization itself, but choosing an appropriate number of components k on the basis of the RMSD values is often problematic, because larger values of k always result in a better approximation to D and hence to lower RMSD. Therefore, to estimate the number of important components another approach, which is called bi-cross-validation (BCV),[288] was employed.

4.3.2. Bi-cross validation

The general idea behind the BCV is to hold out a submatrix of the matrix D (i.e., a set of elements of D) and use the rest of D to predict the held out submatrix. The quality of the k-term NMF is expressed by a BCV error. For instance, in the case of r-by-s submatrices (r and s are such small integers, that n and m are the multiples of r and s, respectively) the BCV error is evaluated as follows. The matrix D is partitioned into $(u \times v)$ submatrices ($u = {}^n/_r$, $v = {}^m/_s$):

$$D^{n \times m} = \begin{pmatrix} D_{11}^{r \times s} & \cdots & D_{1v}^{r \times s} \\ \vdots & \ddots & \vdots \\ D_{u1}^{r \times s} & \cdots & D_{uv}^{r \times s} \end{pmatrix} \qquad (4.10)$$

Then, for a given $i \in \{1, \dots, u\}$ and $j \in \{1, \dots, v\}$, the matrix D has the form:

$$D^{n \times m} = \begin{pmatrix} X_{11}^{(i-1)r \times (j-1)s} & Z_1^{(i-1)r \times s} & X_{12}^{(i-1)r \times (v-j)s} \\ Y_1^{r \times (j-1)s} & D_{ij}^{r \times s} & Y_2^{r \times (v-j)s} \\ X_{21}^{(u-i)r \times (j-1)s} & Z_2^{(u-i)r \times s} & X_{22}^{(u-i)r \times (v-j)s} \end{pmatrix} \qquad (4.11)$$

NMF is invariant under any permutations of rows and columns of D up to the corresponding permutations of rows of W and columns of H, respectively. Therefore, one can rearrange D in such a way that D_{ij} will be the upper left submatrix:

$$\begin{pmatrix} D_{ij}^{r \times s} & Y_1^{r \times (j-1)s} & Y_2^{r \times (v-j)s} \\ Z_1^{(i-1)r \times s} & X_{11}^{(i-1)r \times (j-1)s} & X_{12}^{(i-1)r \times (v-j)s} \\ Z_2^{(u-i)r \times s} & X_{21}^{(u-i)r \times (j-1)s} & X_{22}^{(u-i)r \times (v-j)s} \end{pmatrix} = \begin{pmatrix} D_{ij}^{r \times s} & Y^{r \times (m-s)} \\ Z^{(n-r) \times s} & X^{(n-r) \times (m-s)} \end{pmatrix} \qquad (4.12)$$

where the submatrices X, Y, and Z are defined as follows:

$$X^{(n-r) \times (m-s)} = \begin{pmatrix} X_{11}^{(i-1)r \times (j-1)s} & X_{12}^{(i-1)r \times (v-j)s} \\ X_{21}^{(u-i)r \times (j-1)s} & X_{22}^{(u-i)r \times (v-j)s} \end{pmatrix}$$

$$Y^{r \times (m-s)} = \begin{pmatrix} Y_1^{r \times (j-1)s} & Y_2^{r \times (v-j)s} \end{pmatrix} \qquad (4.13)$$

$$Z^{(n-r) \times s} = \begin{pmatrix} Z_1^{(i-1)r \times s} \\ Z_2^{(u-i)r \times s} \end{pmatrix}$$

Since both r and s are small, the submatrix X contains almost the same information as the entire matrix D and the k-term NMF of X is likely to be of the same accuracy as that of D. Literally, this means that, if the transformed matrix D, in which D_{ij} is the upper left submatrix, is factorized into W and H, so that:

$$\begin{pmatrix} D_{ij}^{r \times s} & Y^{r \times (m-s)} \\ Z^{(n-r) \times s} & X^{(n-r) \times (m-s)} \end{pmatrix} = D^{n \times m} \approx W^{n \times k} \cdot H^{k \times m}$$

$$= \begin{pmatrix} \widehat{W}_D^{r \times k} \\ \widehat{W}_X^{(n-r) \times k} \end{pmatrix} \cdot \begin{pmatrix} \widehat{H}_D^{k \times s} & \widehat{H}_X^{k \times (m-s)} \end{pmatrix} \qquad (4.14)$$

then the k-term NMF of X itself can be approximated as:

$$X^{(n-r) \times (m-s)} \approx W_X^{(n-r) \times k} \cdot H_X^{k \times (m-s)} \approx \widehat{W}_X^{(n-r) \times k} \cdot \widehat{H}_X^{k \times (m-s)} \qquad (4.15)$$

Within the accuracy of the NMF of the entire matrix D, the approximation to D_{ij} is:

$$D_{ij}^{r \times s} \approx \widehat{D}_{ij}^{r \times s} = \widehat{W}_D^{r \times k} \cdot \widehat{H}_D^{k \times s} \qquad (4.16)$$

and using the following two approximations:

$$\widehat{W}_D^{r \times k} \approx \widehat{W}_D^{r \times k} \cdot \widehat{H}_X^{k \times (m-s)} \cdot \left(\widehat{H}_X^{k \times (m-s)}\right)^+ \approx Y^{r \times (m-s)} \cdot \left(H_X^{k \times (m-s)}\right)^+ \qquad (4.17)$$

$$\widehat{H}_D^{k \times s} \approx \left(\widehat{W}_X^{(n-r) \times k}\right)^+ \cdot \widehat{W}_X^{(n-r) \times k} \cdot \widehat{H}_D^{k \times s} \approx \left(W_X^{(n-r) \times k}\right)^+ \cdot Z^{(n-r) \times s}$$

one can finally get:

$$D_{ij}^{r \times s} \approx \widehat{D}_{ij}^{r \times s} \approx Y^{r \times (m-s)} \cdot \left(H_X^{k \times (m-s)}\right)^+ \cdot \left(W_X^{(n-r) \times k}\right)^+ \cdot Z^{(n-r) \times s} \approx$$

$$\approx Y^{r \times (m-s)} \cdot \left(W_X^{(n-r) \times k} \cdot H_X^{k \times (m-s)}\right)^+ \cdot Z^{(n-r) \times s} \qquad (4.18)$$

$$\approx Y^{r \times (m-s)} \cdot \left(X^{(n-r) \times (m-s)}\right)^+ \cdot Z^{(n-r) \times s}$$

where the superscript " $+$ " denotes matrix pseudoinversion. In fact, this means, that the small submatrix D_{ij} can be reconstructed on the basis of the other three submatrices:

$$D_{ij}^{r \times s} \approx Y^{r \times (m-s)} \cdot \left(X^{(n-r) \times (m-s)}\right)^+ \cdot Z^{(n-r) \times s} \qquad (4.19)$$

Furthermore, this approximation becomes the exact equality, provided the rank of the submatrix X equals the rank of the entire matrix D, that is:

$$D_{ij}^{r \times s} = Y^{r \times (m-s)} \cdot \left(X^{(n-r) \times (m-s)}\right)^+ \cdot Z^{(n-r) \times s} \qquad (4.20)$$

The accuracy of the approximation of D_{ij} is assessed in terms of Frobenius norm of the residual:

$$\left\| D_{ij}^{r \times s} - \widehat{D}_{ij}^{r \times s} \right\|_F \qquad (4.21)$$

Since the procedure, described above, can be performed for any $i \in \{1, \dots, u\}$ and $j \in \{1, \dots, v\}$, the BCV error is defined as:

$$BCV(k) = \sqrt{\frac{\sum_{i=1}^{u} \sum_{j=1}^{v} \left\| D_{ij}^{r \times s} - \widehat{D}_{ij}^{r \times s} \right\|_F^2}{n \cdot m}} \qquad (4.22)$$

As in the case of RMSD, one may also use the relative BCV error:

$$BCV_{rel}(k) = \frac{BCV(k)}{BCV(0)} = \frac{\sqrt{\sum_{i=1}^{u} \sum_{j=1}^{v} \left\| D_{ij}^{r \times s} - \widehat{D}_{ij}^{r \times s} \right\|_F^2}}{\left\| D^{n \times m} \right\|_F} \qquad (4.23)$$

When the value of k is less than the true number of components, the BCV error is mainly determined by the NMF itself (low accuracy of the k-rank approximation), and therefore the BCV error is close to the RMSD error. When the value of k becomes greater than the true number of components, the NMF starts to fit the noise in the experimental data by the

redundant components, which is called overfitting. However, the submatrix D_{ij} is estimated on the basis of NMF of the submatrix X. The overfitting of X does not lead to a better approximation to D_{ij}, because the noise is not correlated in these two submatrices. Thus, the large values of k will not yield a smaller BCV error, while the RMSD error monotonically decreases upon increase of k.

Of course, the exact dependence of the BCV error on the number of components k is determined by the quality of the experimental data matrix D (e.g., signal-to-noise ratio) and the size of submatrices, r and s, used for the BCV analysis. Therefore, after plotting the BCV error as a function of k and determining a value of k_{BCV}, which corresponds to a minimum of the BCV error, one has to manually inspect the UV and MS spectra of individual components derived from the k_{BCV}-term as well from the $(k_{BCV} \pm 1)$-term NMF to validate the estimated number of components.

4.3.3. Contributions of the derived components

The contribution of the derived components to the decomposition of the initial matrix D can be evaluated in the following way. The columns of the matrix W and the rows of the matrix H represent the UV and MS vectors of the components, respectively. The outer product of each pair of vectors, corresponding to a certain component, yields a UV-MS matrix. The matrix D is therefore:

$$D \approx \sum_{l=1}^{k} D_l = \sum_{l=1}^{k} W_{\bullet l} \otimes H_{l \bullet}. \qquad (4.24)$$

The sum of all the elements of the matrix D is a measure of the integral photofragmentation efficiency of all the ionic species over the whole range of wavenumbers and fragments, and the sum of all the elements of the matrix D_l is the integral photofragmentation efficiency of the l-th component. Thus, one can define the following quantity, that characterizes the relative contribution of the l-th component:

$$c_l = \frac{\sum_{i,j}(D_l)_{ij}}{\sum_{p=1}^{k}\sum_{i,j}(D_p)_{ij}} \qquad (4.25)$$

where $\left(D_p\right)_{ij}$ is the element of the matrix D_l from the i-th row and j-th column. The value of c_l reflects the relative concentration of the l-th component in a sample as well as its absorption cross-section and photofragmentation efficiency.

4.3.4. A test case

It is fundamental to the approach that only the components, which differ both in absorption and in fragmentation patterns, can be revealed mathematically. Because of this limitation, the library-free analysis cannot be used to distinguish the stereoisomers of [Ala[2], Leu[5]]-enkephalin or ephedrine (the photofragmentation mass spectra of the stereoisomers turned out to be almost identical). However, the positional isomers of the singly phosphorylated TSAAATSY peptide meet the prerequisite for the feasibility of matrix factorization and hence can be a suitable system to test the performance of the library-free approach.

Figure 4.12 illustrates such "blind" analysis for two binary mixtures of isomers, namely, an equimolar mixture of pTSAAATSY and TSAAApTSY and an equimolar mixture of TpSAAATSY and TSAAATpSY. The dependence of the BCV error on the number of components, used for a non-negative matrix factorization, indeed, suggests the most likely presence of two components in both mixtures. While the RMSD error, as might be expected, monotonically decreases and approaches zero as the number of components increases, the BCV error rapidly converges and reaches an almost constant level at k = 2 (Figure 4.12(a,d)). The derived spectra match well to the UV and MS spectra of individual components, obtained by integrating the respective library fingerprints over wavenumber and m/z, respectively. A slight discrepancy in the UV spectra of Component 2 and TSAAApTSY (Figure 4.12(b)) in the 35400-35500 cm[-1] region is likely to be due to an imperfect normalization by the total ion current or by the UV laser power. The derived photofragmentation mass spectra differ from those of individual isomers only in the relative abundances of peaks, while the overall fragmentation patterns are recovered correctly. This can be explained by a dependence of photodissociation on the UV excitation energy, which makes it impossible to accurately represent the UV-MS matrix of an individual species as a product of one UV vector and one MS vector and requires a use of more than one component for matrix factorization.

Figure 4.12. **(a, d)** Plots of the BCV and RMSD errors as a function of the number of component used for non-negative matrix factorization of the UV-MS spectra of equimolar mixtures of singly protonated **(a)** pTSAAATSY and TSAAApTSY and **(d)** TpSAAATSY and TSAAATpSY peptides. **(b)** UV absorption and **(c)** photofragmentation mass spectra of the two components, derived by factorization of the UV-MS matrix of the mixture of pTSAAATSY and TSAAApTSY peptides. **(e)** UV absorption and **(f)** photofragmentation mass spectra of the two components, derived by factorization of the UV-MS matrix of the mixture of TpSAAATSY and TSAAATpSY peptides. For a comparison, the UV and MS spectra of individual isomers, obtained by integrating their UV-MS fingerprints over wavenumber and m/z, respectively, are shown (in color) below the derived traces. For graphical clarity, all spectra are normalized to their respective integral intensities.

In order to estimate relative concentrations of the components, their relative contributions were normalized to the total integrals of the respective library fingerprints (i.e., to the sum of all elements of a matrix). Such normalization accounts for the total

photofragmentation yields of the components. The estimated relative concentrations of the components in the two mixtures 52:48% and 51:49% are in a very good agreement with their 50:50% relative concentrations in solution.

By comparing the UV spectra of isomers in the two mixtures, one can clearly see that the spectral overlap for the pair of serine-phosphorylated peptides is higher compared to that for the peptides phosphorylated on threonine. This leads to a different behavior of the BCV error as a function of the number of components, k. While the matrix factorization at $k = 1$ accounts for ≈ 80 % of variance in the UV-MS matrices of both mixtures, the addition of one more component to the factorization model reduces the unexplained variance to 13 and 18 % in the case of peptides with pThr and pSer, respectively. In other words, it is less important to add the second component in the latter case, because from numerical point of view the one-component model describes the UV-MS matrix almost equally well as the two-component one. Thus, the degree of diversity of the UV and MS spectra of individual components is a critical factor, determining the number of component, that can be unambiguously revealed by the "blind" analysis. For instance, the BCV analysis of a four-component mixture, shown in Figure 4.13(a), suggests the presence of only three components in the mixture. The derived UV and MS spectra of two of them, Component 1 and Component 3, are in a reasonable agreement with the UV and MS spectra of the pTSAAATSY and TSAAApTSY peptides (Figure 4.13(b,c)). However, the two peptides phosphorylated on serine could not be discriminated, and they both contribute to the spectra of Component 2. To check whether these isomers can be distinguished, provided the correct value of k is used for NMF, we performed a four-component factorization of the UV-MS matrix of the mixture. Figure 4.13(d,e) clearly demonstrates, that the spectra of the TpSAAATSY and TSAAATpSY are still mixed, although now they are represented by two distinct components. Moreover, such overfitting may even deteriorate the obtained results (e.g., the derived mass spectrum of Component 3 apparently exhibits a wrong fragmentation pattern) due to less numerical stability of the algorithm.

In overall, these examples validate the use of the library-free approach for qualitative identifications of unknown components in their mixtures. For instance, one may use their derived mass spectra as an input for searching MS databases and/or *de novo* sequencing algorithms. The derived UV spectra may further narrow the range of candidate structures,

64

because the UV band origins of various aromatic residues are located in different spectral regions.[183]
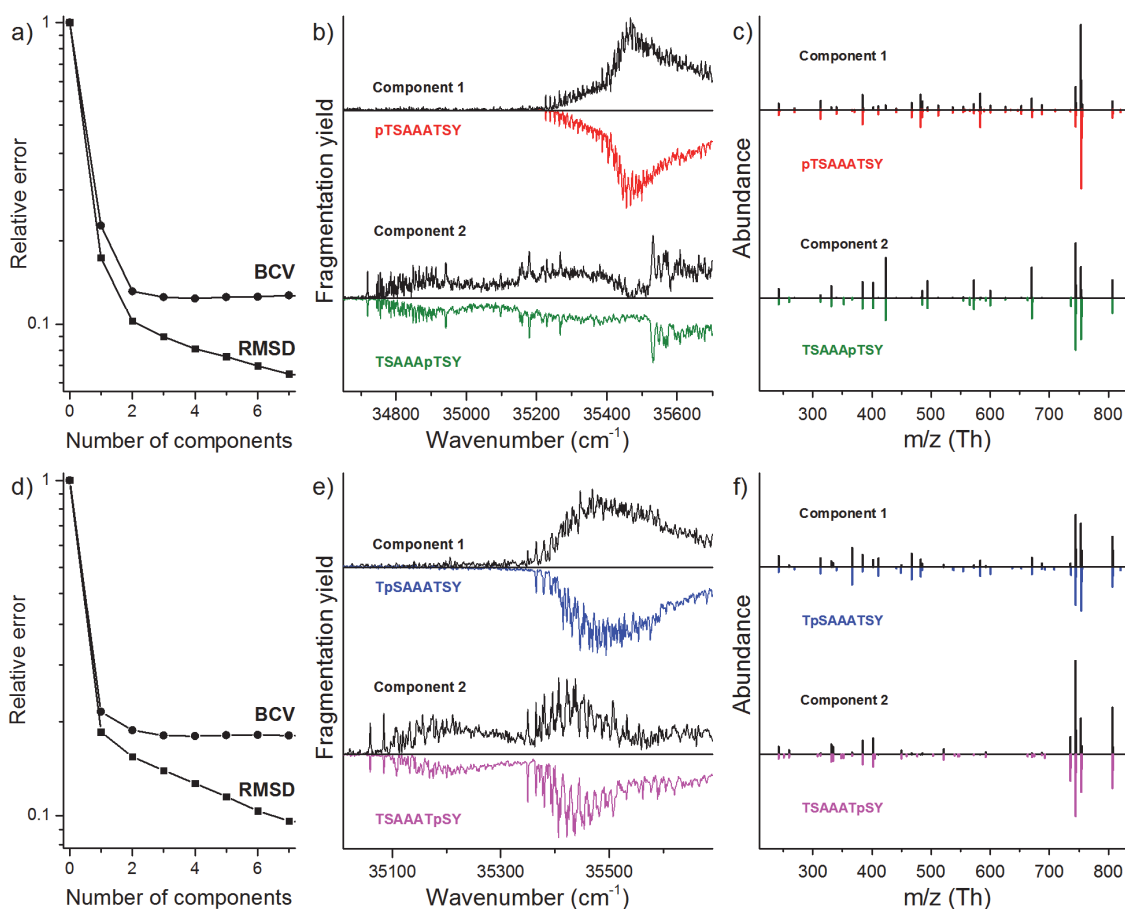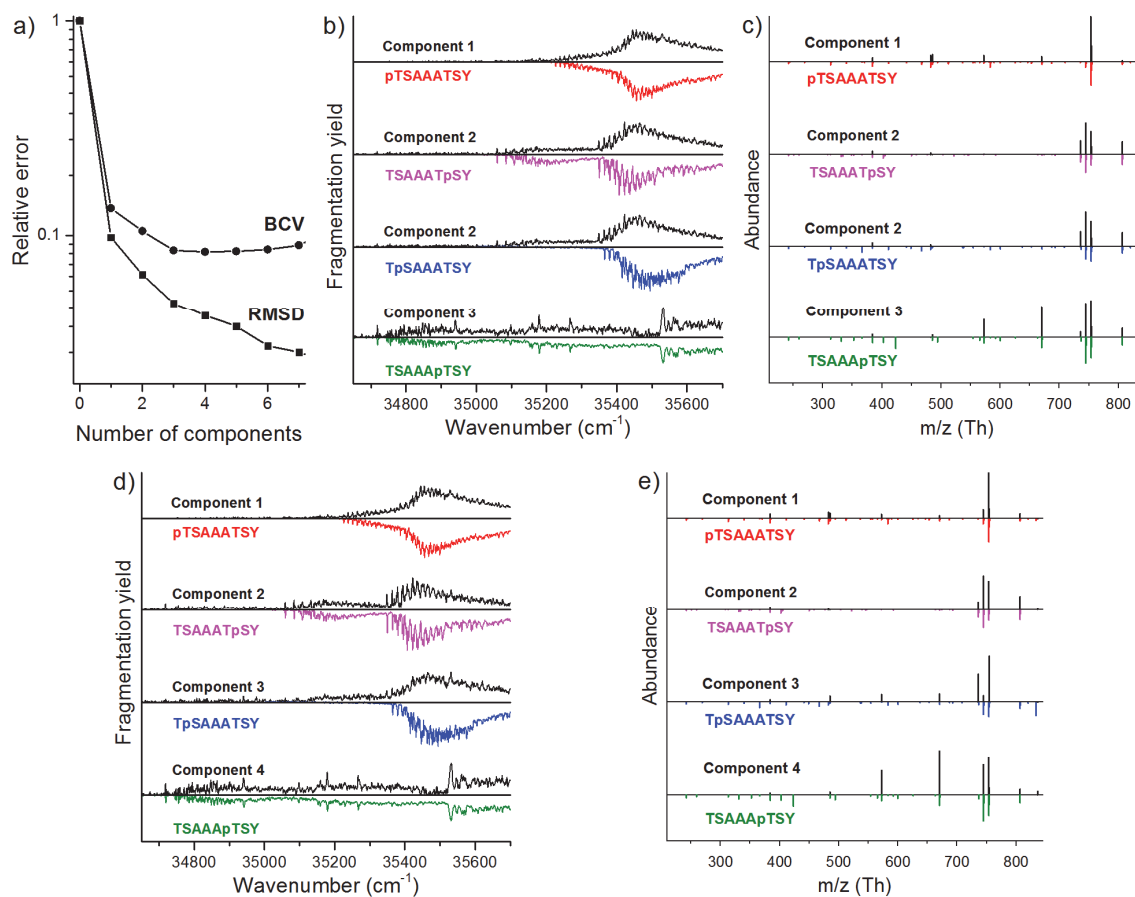

Figure 4.13. **(a)** Plot of the BCV and RMSD errors as a function of the number of component used for non-negative matrix factorization of the UV-MS spectrum of a mixture of singly protonated pTSAAATSY, TpSAAATSY, TSAAApTSY, and TSAAATpSY peptides (25 % each). **(b)** UV absorption and **(c)** photofragmentation mass spectra of the components, derived by three-component factorization of the UV-MS matrix. **(d)** UV absorption and **(e)** photofragmentation mass spectra of the components, derived by four-component factorization of the UV-MS matrix. For a comparison, the UV and MS spectra of individual isomers, obtained by integrating their UV-MS fingerprints over wavenumber and m/z, respectively, are shown (in color) below the derived traces. For graphical clarity, all spectra are normalized to their respective integral intensities.

4.4. Notes on Practical Application

An important issue for the library-based approach is whether it can be used for online identification, when the time for a fingerprint measurement is limited by the width of LC peak to a few to tens of seconds. High-resolution 2D fingerprints, discussed in this chapter, were measured at thousands of UV wavenumbers, and typically it takes several hours to record one 2D spectrum. Thus, a library of 2D fingerprints can be created only offline. However, for an analysis of a mixture instead of recording its full 2D spectrum one can measure a set of photofragmentation mass spectra at a few critical wavenumbers and apply the same decomposition algorithm to such a truncated UV-MS matrix.

For instance, for the library of phosphorylated peptides considered above, the five UV absorption band origins give a minimum full set of the five critical wavenumbers, which can be complemented each time by the wavenumbers of nearby peaks. Using the full 2D UV-MS matrices of the five peptides and their mixtures, we simulated how the performance of the library-based identification depends on the number of wavenumbers, retained in the truncated matrices, and on the spectral linewidth of the excitation laser pulse. The first five critical wavenumbers are the positions of the UV band origins, taken in the order of decreasing energy of the transitions (e.g., the first wavenumber in the set corresponds to the band origin of the TSAAATSpY peptide). The next five wavenumbers (number of wavenumbers from 6 to 10 in Figure 4.14) correspond to the intense peaks, nearest to the respective band origins. These 10 wavenumbers are then complemented by the positions of the next two intense peaks in each of the five fingerprints, giving totally a set of 20 wavenumbers, and so on. Thus, a set of 100 wavenumbers contains the positions of 20 intense peaks (counting from the band origin) in each of the UV spectra of the five peptides. Additionally, we generated 100 sets of randomly chosen wavenumbers. To simulate the effect of the laser linewidth, instead of taking a slice of the 2D fingerprint at a certain wavenumber we calculated a weighted average of the fragmentation mass spectra (slices of 2D fingerprints) in the region around this wavenumber using a Gaussian function with the full width at half maximum of 1 and 10 cm^{-1}. Figure 4.14(a) corresponds to the linewidth of the UV laser, that was used to record the 2D fingerprints, which effectively means that no averaging was applied in this case.
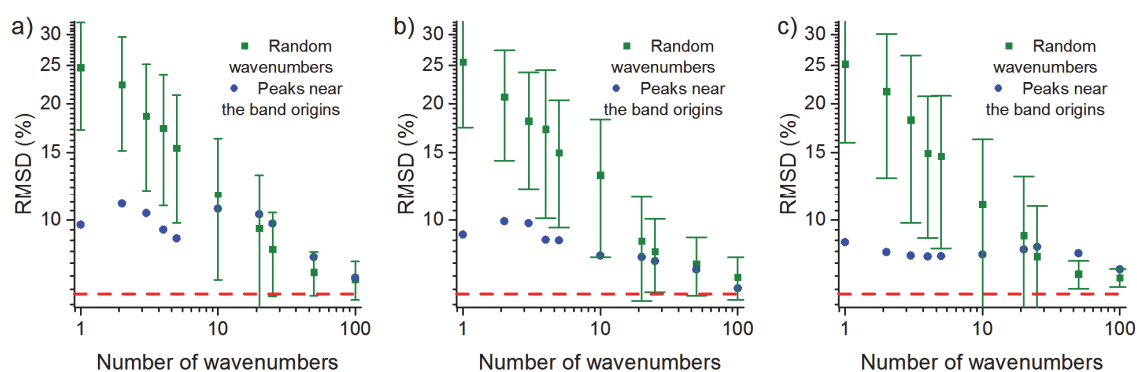
Figure 4.14. RMSD of the library-based analysis of 12 mixtures of the isomeric phosphorylated peptides plotted as a function of the number of wavenumbers, retained for matrix decomposition. Photofragmentation mass spectrum at each wavenumber was taken either **(a)** as a slice of the respective 2D UV-MS matrix at this wavenumber or as a weighted average of the photofragmentation mass spectra, recorded in the region around this wavenumber, using a Gaussian function with the FWHM of **(b)** 1 and **(c)** 10 cm^{-1}. Each green dot represents an average over 100 randomly generated sets of wavenumbers, and the error bar indicates the dispersion of the RMSD values. Blue dots are the RMSDs, calculated for sets of wavenumbers, corresponding to the UV peaks near the band origins of the five phosphorylated peptides.

Figure 4.14 shows that a random choice of a few wavenumbers results in unacceptably high RMSD values, which, however, rapidly decrease and for a set of 100 wavenumbers approach that of the full decomposition. As might be expected, increasing the number of wavenumbers also reduces the dispersion of the RMSD. Interestingly, the RMSD almost does not depend on the width of the weight function, which may indicate that the major contribution to the analysis accuracy results from the randomness of the choice of wavenumbers. However, this is not the case for the sets of critical wavenumbers. For a narrow linewidth (0.1 cm^{-1}) the RMSD strongly depends on the number of wavenumbers in a set, while for a broad linewidth (10 cm^{-1}) it stays almost constant. Apparently, slight imperfections in the data analysis procedures (e.g., baseline correction or normalization) introduce numerical errors in the library fingerprints as well as in the 2D spectra of mixtures, which makes the decomposition algorithm very sensitive to the exact choice of wavenumbers when their number in a set is small. Thus, by increasing the number of retained wavenumber one actually performs a sort of averaging, which improve the accuracy of analysis. On the contrary,

67

averaging with a broad weight function prior to decomposition smooths all the sharp features in UV spectra, and therefore, even for a set of 100 wavenumbers the RMSD remains relatively large.

The linewidth of 1 cm^{-1} seems to be an optimal value. On one hand, for a small number of wavenumbers in a set the decomposition is stable with respect to small errors in the input data. On the other hand, for a set of 100 wavenumbers the RMSD value reaches that of the full analysis. And even a use of 10-20 wavenumbers for the library-based identification results in a reasonable RMSD value of 8 %. An ultrahigh field Orbitrap analyzer, set to a moderate mass resolution of $6 \cdot 10^4$, needs only about 0.15 s to complete a measurement of one mass spectrum.[289] This duration corresponds to 3 s for measuring the 2D spectrum of a mixture at 20 wavenumbers, suggesting that the rate of the UV-MS measurements can be sufficiently fast for online LC-MS identifications, using preliminarily measured libraries.

Another important question that we would like to address here is whether the cryogenic cooling is a prerequisite for a successful use of the UV-MS approach. For example, ephedrines exhibit rather narrow UV peaks and have a relatively sparse pattern of vibronic transition near their UV band origins. Therefore, even at an elevated temperature of the cold ion trap ($\approx$ 64 K) their UV spectra retain a clearly resolvable vibronic structure (Figure 4.15). Moreover, the UV spectra of the two stereoisomers are still distinguishable and only slightly overlap. However, because of broadening of the rotational contours of vibronic transitions and because the population of the ground vibrational level is depleted according to the Boltzmann distribution, the photofragmentation ion signal at a single wavenumber of the UV laser is much lower compared to that at the temperature of 10 K. The noise level is mainly determined by the Orbitrap analyzer and does not depend on the temperature of the cold ion trap. Thus, increasing the temperature results in a deteriorated signal-to-noise (SNR) ratio in recorded spectra.

Figure 4.15(b,c) compares the performed decomposition of the UV-MS matrices of an equimolar mixture of (+)-ephedrine and (+)-pseudoephedrine, obtained at two different temperatures of the cold ion trap, 10 and 64 K. Although the UV spectra of the mixture are fitted equally well by a linear combination of the UV spectra of the stereoisomers and their estimated relative concentrations are very similar, the uncertainty in the calculated concentrations in the case of T ≈ 64 K is likely to be larger. Indeed, since the concentrations
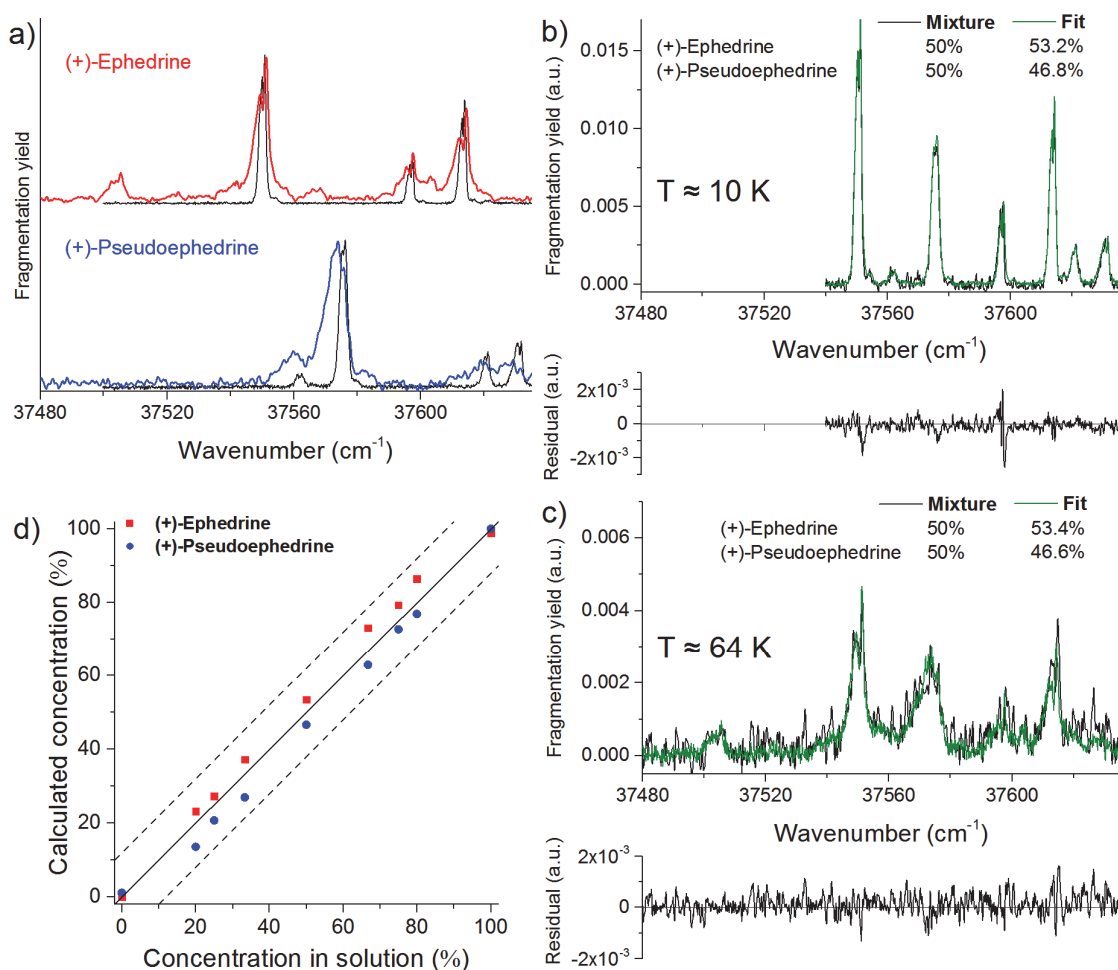
Figure 4.15. **(a)** UV absorption of (+)-ephedrine (red) and (+)-pseudoephedrine (blue) at the temperature of the cold ion trap T ≈ 64 K, taken as slices of the respective UV-MS spectra at m/z = 58.066 Th and normalized to the maximum signal. For comparison, the UV spectra of ultracold ions (T ≈ 10 K) are shown in black. **(b, c)** UV absorption spectrum (black) of an equimolar mixture of two stereoisomers of ephedrine, taken as a slice of the experimental UV-MS spectra at m/z = 58.066 Th, and its fit (green) by a linear combination of the UV spectra of the individual component. The resulting residual plot is shown at the bottom. The temperature of the cold ion trap was set to **(b)** 10 K and **(c)** 64 K. **(d)** Calculated relative concentrations of (+)-ephedrine (red squares) and (+)-pseudoephedrine (blue circles) as a function of their relative concentration in solution (0–100 %) for 9 different mixtures. A solid line shows the ideal 1:1 dependence, and the area between two dashed lines corresponds to the 99.7 % confidence band (3 × RMSD) for calculating concentrations.

are obtained as a solution of a linear least squares problem, the respective relative error is inversely proportional to the square root of SNR, which can be estimated as follows:

$$SNR = \frac{\sigma_{signal}^2}{\sigma_{noise}^2} = \frac{\left\|\sum_{l=1}^{k} x_l \cdot D_l\right\|_F^2}{\left\|D_M - \sum_{l=1}^{k} x_l \cdot D_l\right\|_F^2} \qquad (4.26)$$

For instance, SNR is 49.7 and 4.4 for the considered mixture at T ≈ 10 K and 64 K, respectively, which means, that the relative error in the calculated concentrations is approximately 3 times larger. Apparently, SNR depends on the concentration of the species to be detected and on their photofragmentation yields. Thus, the analysis performed for 9 different mixtures of the isomers of ephedrine at reasonably high concentrations (at T ≈ 64 K) still correctly identified them (Figure 4.15(d)) with RMSD of 4 %. The accuracy of such analysis can be improved by increasing the photofragmentation efficiencies of the isomer.

Finally, it is worth noting that the vibrational resolution in UV spectra of species is affected not only by the laser linewidth and the temperature of the cold ion trap, but also by the size of these species. For instance, the UV spectra of even relatively small proteins (e.g., cytochrome c) can be broaden due to significant conformational heterogeneity and rich vibrational structure of large species. The large size of proteins also reduces their UV fragmentation yield, making MS detection of the fragments difficult. In opposite, there are examples of vibrationally resolved UV spectra of cold protonated peptides containing up to 17 amino acids.[178,249] These data suggest the applicability of the UV-MS technique, in particular, to tryptic peptides, which are on the same size-scale.

4.5. Conclusions

In summary, we have presented a potentially broadly useful approach for structural identification of UV-absorbing molecules. Based on direct measurements of fundamental characteristics of ionic species, namely, the wavenumbers of their vibronic transitions and the m/z of the parent and fragment ions, the UV-MS method is insensitive to moderate changes of the experimental conditions, making the 2D fingerprints highly specific and reproducible.

We have demonstrated its application for a library-based identification of isomers in their mixture. In particular, using the library-based approach we have correctly identified five isomeric peptides, differing only in their phosphorylation site, which could not be fully separated by a standard ultra-performance liquid chromatography, and two diastereomers of a five-amino-acid peptide, which could not be baseline-resolved by ion mobility spectrometry. We have also shown that it can be applied for identification of non-peptidic molecules. Moreover, the estimated accuracy, limits of detection and dynamic range for determining the concentrations of two diastereomers of ephedrine, a natural central nervous system stimulant, meet the requirements of the World Anti-Doping Agency, which suggests, that the UV-MS method may be successfully employed in the fields of doping analysis and forensic science.

Besides the library-based analysis of 2D fingerprints, we have proposed a more sophisticated, library-free approach, which in certain cases is capable of estimating the number of components present in a mixture and deriving their respective UV absorption and photofragmentation mass spectra directly from the UV-MS spectrum of the mixture.

The exact limitations for applicability of UV-MS approach has yet to be assessed. The hardware and operational conditions utilized for the approach are not specific to molecules and should be suitable for fingerprinting of any small to midsize UV-absorbing species compatible with MS analysis, such as peptides, containing aromatic residues, drugs, metabolites, etc. Our analysis demonstrates that the UV-MS fingerprinting also has the potential for coupling it to liquid chromatography for online identifications in a broad range of applications.

Chapter 5

Identification of Conformers of Biomolecules

In this chapter we report on a use of UV-MS spectroscopy, complemented by a numerical analysis of the recorded 2D data arrays, for identification of conformers of biomolecular ions and recovering their individual UV absorption and photofragmentation mass spectra. We first validate this approach with a benchmark dipeptide, Tyr-Ala, and then apply it to a decapeptide, gramicidin S, and to a small drug molecule, homatropine. We also show how a careful analysis of the derived UV and MS spectra allows deducing information on the structural features of the conformers. For instance, the revealed characteristic structural difference between the conformers of gramicidin S identifies one of its previously calculated structures as the most likely geometry of the conformer C, whose structure still remains unknown. And the analysis of the UV-MS spectrum of homatropine provides some structural information, which can be used for solving its 3D structure.

5.1. Introduction

Conformational heterogeneity of biomolecules is a key factor, that enables high diversity of their biological functions. In drug design, conformational heterogeneity is one of the factors that limit pharmacological efficiency of drug molecules.[290,291] Recognition of the conformers of a molecule and, subsequently, determination of the difference in their 3D structures are therefore highly desired for modeling functionality of biomolecules on a fundamental, molecular level. A few approaches have been developed over the last decades to reveal and characterize conformers of biological ions isolated in the gas phase. Ion mobility spectrometry (IMS) is capable of determining conformer-specific cross-sections for collisions of ions with a

buffer gas.[292] These cross-sections can be used for validation of calculated conformational structures, although relying on a single value (cross-section) to characterize complex structures of biomolecules leaves certain ambiguity in these validations and may require a combination of IMS with other techniques.[155,293] IR-UV depletion spectroscopy is another method that has demonstrated its capability to reveal conformers of gas-phase aromatic biomolecules and to characterize each of the conformers by its IR spectrum.[58,60,103,116,176,181] The use of such spectra for a stringent validation of 3D molecular structures, however, requires their accurate calculation, which is currently very challenging already for mid-size biomolecules. Additional structural constraints are therefore highly desired to guide conformational search and to reduce the number of calculated candidate structures prior to computation of their IR spectra.

Our approach to identify conformers of peptides, containing aromatic residues, is based on the fact, that UV electronic excitation of peptides may produce some specific fragments, which do not appear in thermal (e.g. collision-induced or IR multiple photon) dissociation.[168,170,178,183,231,232,248,250,294] The high sensitivity of electronic states to local environment of UV chromophores[168,170,177,249,295] makes the abundance of such fragments, to a certain extent, dependent on the geometry of ionic species. This dependence may allow for distinguishing conformers, in which the chromophore environments are sufficiently different.

We have employed 2D UV-MS spectroscopy, which was demonstrated for identification of isomeric species in the previous chapter, to explore this idea. Similarly to our previous analysis of isomers, a recorded two-dimensional data array (ion abundance vs m/z and UV wavenumber) is analyzed to identify the distinguishable conformers of ions using mathematical algorithms of matrix analysis.[285–288] The data array is approximated by a product of two non-negative matrices of appropriate dimensions, which are varied to minimize the norm of the residual. If the number of the conformers is k, the first matrix of the optimized product contains k UV absorption spectra, while the second one collects k photofragmentation mass spectra of these conformers. The number of essential conformers, contributing to the measured UV-MS matrix, is evaluated using the bi-cross-validation (BCV) approach,[288] in which goodness of k-term matrix factorizations is assessed by a metric, called BCV error. The value of k that minimizes the BCV error is an estimate of the number of

conformers. It is fundamental to the approach that only the conformers, which differ both in absorption and in fragmentation, can be revealed mathematically.

Here we first validate this approach with a benchmark dipeptide, Tyr-Ala, and then apply it to a decapeptide, gramicidin S. Finally, we demonstrate the identification of conformers of a small non-peptidic drug molecule, homatropine. Moreover, in each case we use the derived UV absorption and photofragmentation mass spectra of the conformers to deduce structural information on the local environment of their chromophores.

5.2. Conformers of a Dipeptide, Tyr-Ala

In order to validate our 2D UV-MS approach, we applied it to a singly protonated dipeptide, Tyr-Ala, for which two distinct pairs of conformers had been previously found and characterized by a multiple use of IR-UV double resonance spectroscopy.[170] Since the position of the Tyr absorption band origin is sensitive to 3D structure, all the four conformers differ in their UV spectra. It was also shown that, at least, the conformers A and B have different photofragmentation spectra (Figure 2 in ref. [170]). Thus, matrix factorization should be able to discriminate between these conformers and recover their respective UV and MS spectra.

First, we measured the UV-MS spectrum of [YA + H]$^+$ and performed a BCV analysis (Figure 5.1 (a)) to estimate the number of significant components, contributing to the spectrum. In contrast to the monotonically decreasing RMSD error, the BCV error as a function of the number of components, used for the UV-MS matrix factorization, attains its minimum at k = 3, which suggests the presence of three major components, denoted below as I, II and III. Their UV absorption and photofragmentation mass spectra are shown in Figure 5.1(b,c). The assignment of peaks in the UV spectrum of the dipeptide clearly indicates that the component I essentially corresponds to the conformers B and D, while the component II – to the conformers A and C, although the derived spectrum of the component I contains a very small admixture of the UV spectra of the conformers A and C. The nearly identical conformers within each pair differ only in the orientation (*syn/anti*) of hydroxylic OH group of Tyr with respect to the rest of the ion. Therefore, they are likely to have very similar photofragmentation mass spectra and could not be distinguished. The derived mass spectra
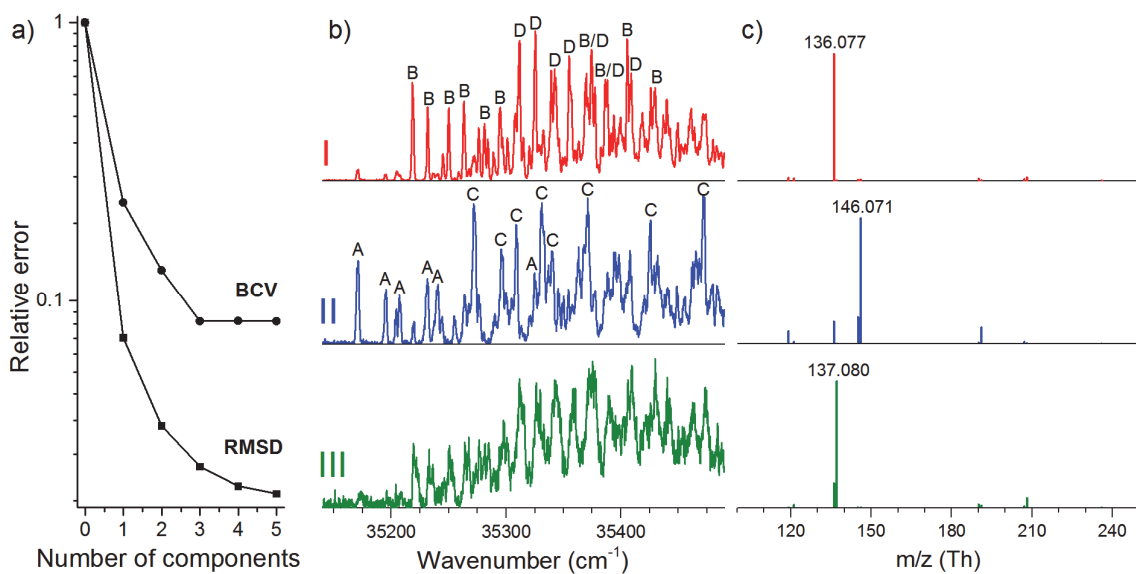
Figure 5.1. **(a)** Plot of the RMSD and BCV errors as a function of the number of components, used for non-negative matrix factorization of the UV-MS spectrum of singly protonated Tyr-Ala dipeptide. **(b)** UV absorption and **(c)** photofragmentation mass spectra, derived by three-component factorization of the UV-MS matrix. UV spectra are normalized to the maximum signal and MS spectra are normalized by the intensity of the most abundant fragments. Peaks in the UV spectra are labeled according to the conformational assignment, performed in ref. [170].

are also in a good agreement with those, obtained by Stearns et al. The spectrum of the component I is dominated by the Tyr immonium ion, similarly to that of the conformer B, recorded at its UV band origin. And the two most abundant fragments in the spectrum of the component II correspond to those observed for the conformer A at its UV band origin, although the relative intensities of the two peaks are somewhat different, because of slight admixing of the conformers A and C to the component I. However, in general, the match of the derived spectra of the components I and II to the previously measured conformer-specific UV transitions and photofragmentation mass spectra validates the performed "blind" analysis.

Notably, the BCV analysis suggested the presence of one additional component. None of the previously found and characterized conformers has the UV absorption and/or fragmentation mass spectra similar to those of the component III. However, a close inspection of its derived spectra reveals three indications, that this component corresponds to an isotopologue of the component I with one ^{13}C in the Tyr immonium fragment. First, the most

abundant peak in its photofragmentation MS is at 137.080 Th, that is, shifted exactly by the ^{12}C-^{13}C mass difference (1.0033 Da) with respect to the monoisotopic peak of the Tyr immonium ion. Second, the determined relative contribution of the component III (8 %) is close to the abundance of the Tyr immonium ion with one ^{13}C (8.9 %). Finally, the spectroscopic evidence of our assignment arises from the shape of the derived UV spectrum. All the UV transitions are slightly shifted and broadened compared with those of the component I. The spectral shift of the UV band origin can be explained by the change of the zero-point vibrational energies (ZPVEs) of the ground (S_0) and the first excited singlet (S_1) states upon the isotopic substitution. The change in vibrational frequencies in the S_1 state also affects the positions of vibronic transitions in the UV spectrum. The observed spectral broadening of UV transitions and particularly of the band origin indicate some inhomogeneity. For instance, the ^{12}C-^{13}C isotopic substitution of various carbon atoms in the immonium fragment may result in slightly different ZPVEs.
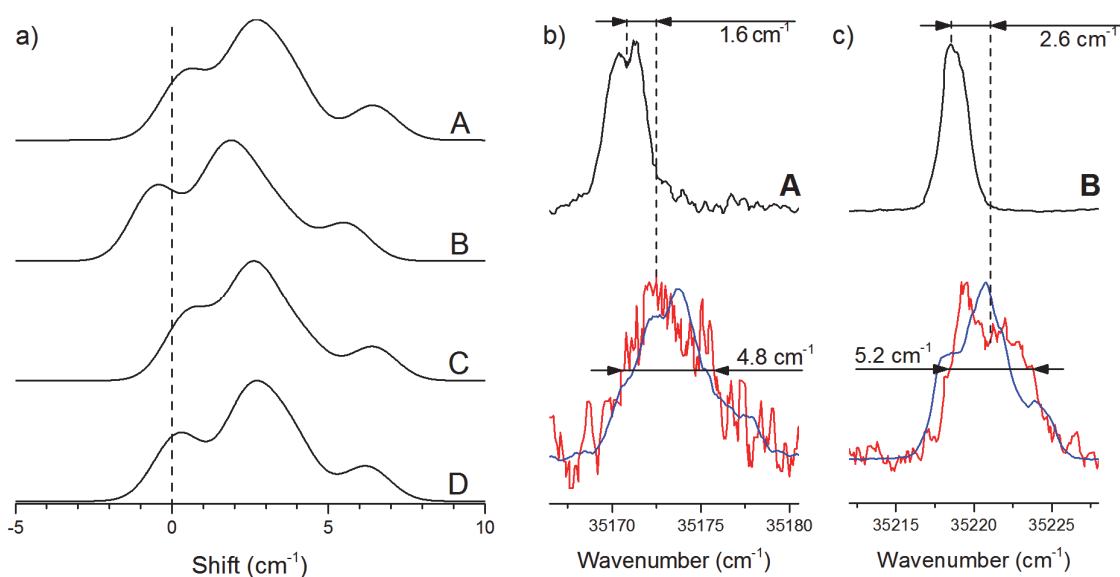
Figure 5.2. **(a)** Simulated broadening of a single UV peak of the isotopologues of the conformers A, B, C, and D (from top to bottom) with one ^{13}C. The spectral shift is calculated relative to the position of the UV peak of the respective monoisotopic species, indicated by the dashed line. **(b, c)** UV spectra around the band origins of the conformers A **(b)** and B **(c)**, taken as slices of the UV-MS spectrum at m/z = 136.077 (black) and 137.080 Th (red), and simulated shapes of the band origins (blue) of their isotopologues with one ^{13}C.

In order to test this hypothesis, we simulated the shapes of the UV band origins of the four conformers of [YA + H]$^+$ and their isotopologues. First, we took the Cartesian coordinates of the four lowest energy structures, calculated at the B3LYP/6-31+G** level of theory.[183] The geometry of these conformers are in a good agreement with those calculated by Stearns et al.[170] Therefore, we adopt the notation A, B, C, and D for the conformers. For our calculations we employ a Coulomb-attenuated version of the hybrid B3LYP functional (CAM-B3LYP),[296] because of the known deficiency of B3LYP in describing excited states with significant charge-transfer character,[297,298] and a double-zeta basis set without diffuse functions (6-31G**), because the geometry optimization and the Hessian calculations of the S_1 state would be too time-consuming. For all the calculations the GAMESS (US) software package[299] was used.

Ground-state geometries of the four conformers were re-optimized and subsequently used as a starting point for the S_1 excited state geometry optimization. Both ground- and excited-state geometry optimizations were performed with a convergence criterion of 10^{-5} a.u. for the atomic forces. The Hessian matrices were calculated semi-numerically for S_0 and S_1 equilibrium geometries of each conformer. Using these matrices and the exact mass of the ^{13}C isotope (13.003355 Da), the harmonic frequencies and the respective ZPVEs for each of the isotopomers (i.e., species with the single ^{12}C-^{13}C isotopic substitution of various carbon atoms) were evaluated. All the S_0 and S_1 calculations were performed at the DFT and TD-DFT levels of theory, respectively. To simulate the shape of the UV band origin of each conformer at the fragment m/z of 136.077 Th and 137.080 Th we divided all the isotopomers into two groups: those, which contain ^{13}C in the immonium fragment, and those, which contain ^{13}C in the rest of the ion. The shapes of the band origins were evaluated as follows:

$$I(\tilde{v}, 136.077\ Th) = (1 - \alpha) \cdot I_0(\tilde{v}) + \alpha \cdot \sum_k I_0(\tilde{v} - \lambda \cdot \Delta\tilde{v}_k)$$

$$I(\tilde{v}, 137.080\ Th) = \alpha \cdot \sum_k I_0(\tilde{v} - \lambda \cdot \Delta\tilde{v}_k)$$

(5.1)

where $\Delta\tilde{v}_k$ is the calculated shift of the difference between the S_1 and S_0 ZPVEs of the k-th isotopomer (from the first and the second group, respectively) relative to that of the monoisotopic species; $\alpha = 0.011$ is the abundance of ^{13}C; $\lambda = 0.976$ is the scaling factor for ZPVE;[300] and $I_0(\tilde{v})$ is the shape of the band origin of monoisotopic species. $I_0(\tilde{v})$ was

approximated either by a Gaussian function with FWHM = 1.77 cm^{-1} (Figure 5.2(a)) or by the exact shape of the band origin of the conformers A (Figure 5.2(b)) and B at m/z = 136.077 Th (Figure 5.2(c)).

The performed calculation of ZPVEs of the two electronic states of the isotopologues confirms that the observed spectral broadening of UV peaks arises from a dispersion of the spectral shifts in various isotopomers and allows us to reproduce (at least, qualitatively) the shape of the UV band origins of the isotopologues of the conformer A (Figure 5.2(b)) and B (Figure 5.2(c)). Moreover, the calculations reveal the origin of such dispersion. Figure 5.3 shows the calculated ground-state structures of the conformers, in which the size of carbon atoms is proportional to the spectral shifts caused by the isotopic substitution of these carbons. In all the four structures the largest shifts arise from the substitution of the carbons in the aromatic ring, that are the nearest and the farthest to the protonated amino group, indicating that the electronic excitation is mainly perturbed by the proton-π interaction. Overall, the detection and unambiguous assignment of the isotopologue of a conformer in this study strengthens our confidence in the ability of the new approach to distinguish conformational families and to derive their UV absorption and fragmentation mass spectra.
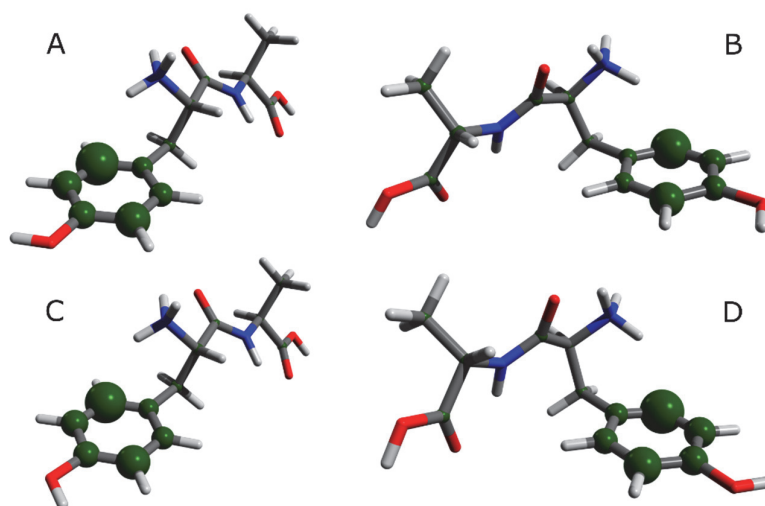

Figure 5.3. Calculated here structures of the conformers A, B, C, and D. The size of the carbon atoms (green spheres) is drawn proportional to the calculated spectral shifts, resulting from ^{12}C-^{13}C isotopic substitution of the respective carbons.

The photofragmentation mass spectra of the components I and II and hence of the conformers B/D and A/C differ drastically. The UV-specific Tyr side-chain loss (m/z = 146.071 Th) is by far the main fragmentation channel for the conformers A/C, while the mass spectrum of the conformers B/D contains mainly thermal fragments (e.g., Tyr immonium ion at m/z = 136.077 Th). Regarding the structures of the two conformational families (Figure 5.3), we attribute this difference to the different degree of spatial isolation of the chromophore from the peptide backbone. The preferential cleavage of the C_α-C_β bond was earlier observed in isolated chromophores;[294,301–303] it was also detected in single amino acids and peptides.[168,178,183,231,246] Different models have been elaborated to explain this UV-specific fragmentation.[250,294,301,304] Recent experimental studies provide evidence that the mechanism of the Phe and Tyr side-chain loss in protonated peptides involves triplet states, which are associated with local excitation of the aromatic ring.[246] In the framework of this model, the intersystem crossing in such chromophores should intrinsically lead to C_α-C_β bond cleavage following UV excitation of amino acids and peptides, provided a non-radiative decay of the electronic excitation through other channels is hindered by high energy barriers.[248] Close proximity of a chromophore to other functional groups of a molecule may reduce these barriers,[250,305] opening, in particular, an internal conversion through conical intersections as a competitive pathway towards thermal fragmentation.[306] Consistently, for the conformers A/C of the dipeptide, in which the chromophore is relatively remote from the rest of the ion, the main fate of the electronic excitation is the C_α-C_β bond cleavage, while the proximity of the chromophore to the backbone in the conformer B/D reroutes UV excitation towards thermal fragmentation.

5.3. Conformers of a Decapeptide, Gramicidin S

The observed correlation between the structure and UV fragmentation of the dipeptide helps us in deriving some structural constraints for a larger peptide, gramicidin S (GS). Four conformers, labeled A, B, C and D, of this doubly protonated peptide with the amino-acid sequence *cyclo*-(-Val-Orn-Leu-D-Phe-Pro-)$_2$ were previously revealed by UV spectroscopy,[249] the existence of three of which was distinctly confirmed by IR-UV spectroscopy.[176,249,307] Moreover, the 3D of the conformers A and B have been calculated and validated by

comparison with the experimental conformer-selective IR spectra.[176,307,308] However, the structure of the less abundant conformer C still remains unknown. Because this conformer apparently has higher energy than both A and B, additional structural constraints will facilitate the search of its 3D structure.

The BCV analysis of the measured UV-MS spectrum of [GS + 2H]$^{2+}$ suggests the presence of two or three distinct components (Figure 5.4(a)). Indeed, although the BCV error mostly stabilizes for $k \geq 2$, it still slightly decreases. Therefore, one has to verify manually, whether this decrease is due to the presence of a minor conformer or not. The derived UV spectra (Figure 5.4(b,d)) of the components I and II contain all the transitions previously assigned to the conformers A/B (the structures of the conformers A and B were shown to be very similar,[307,308] making them indistinguishable by our approach) and C/D, respectively,[249] while the spectrum of the third component (III) does not look like a real UV spectrum, but rather resembles noise. The respective fragmentation spectra (Figure 5.4(c,e)) reveal, that it originates from an interference of the two closely spaced peaks at 549.84 and 549.86 Th (and also of the corresponding isotopic peaks at 550.34 and 550.36 Th). The former is not related to the UV photodissociation and may be an artefact, while the latter is a photoinduced fragment of relatively low abundance, which can be assigned to the neutral loss of isocyanic acid (HN=C=O), based on its accurate monoisotopic mass ($\delta m/m \approx 12$ ppm). Allocating the noise to the third component improves the signal-to-noise ratio of the components I and II. However, the three-component matrix factorization does not indicate any signs of the presence of additional conformers. Thus, for further analysis we focus on the UV and mass spectra derived for the components I and II by three-component factorization.

Unlike in the early study of GS, where the fragments were detected by a low-resolution quadrupole mass spectrometer, the high mass resolution, provided by the Orbitrap analyzer, allows for an unambiguous assignment of the detected photofragments. The most abundant photofragments in the derived spectra (Figure 5.4(e)) of the component I and II correspond to the neutral losses of a CO group (m/z = 557.36 Th) and of the Phe side chain (m/z = 525.84 Th), respectively. The loss of CO is a typical thermal fragment in a thermal-like (e.g., collision-induced or IR multiple photon) dissociation of cyclic peptides, produced right after the peptide ring opening.[245] In contrast, the Phe side-chain loss is a UV-specific fragmentation channel of
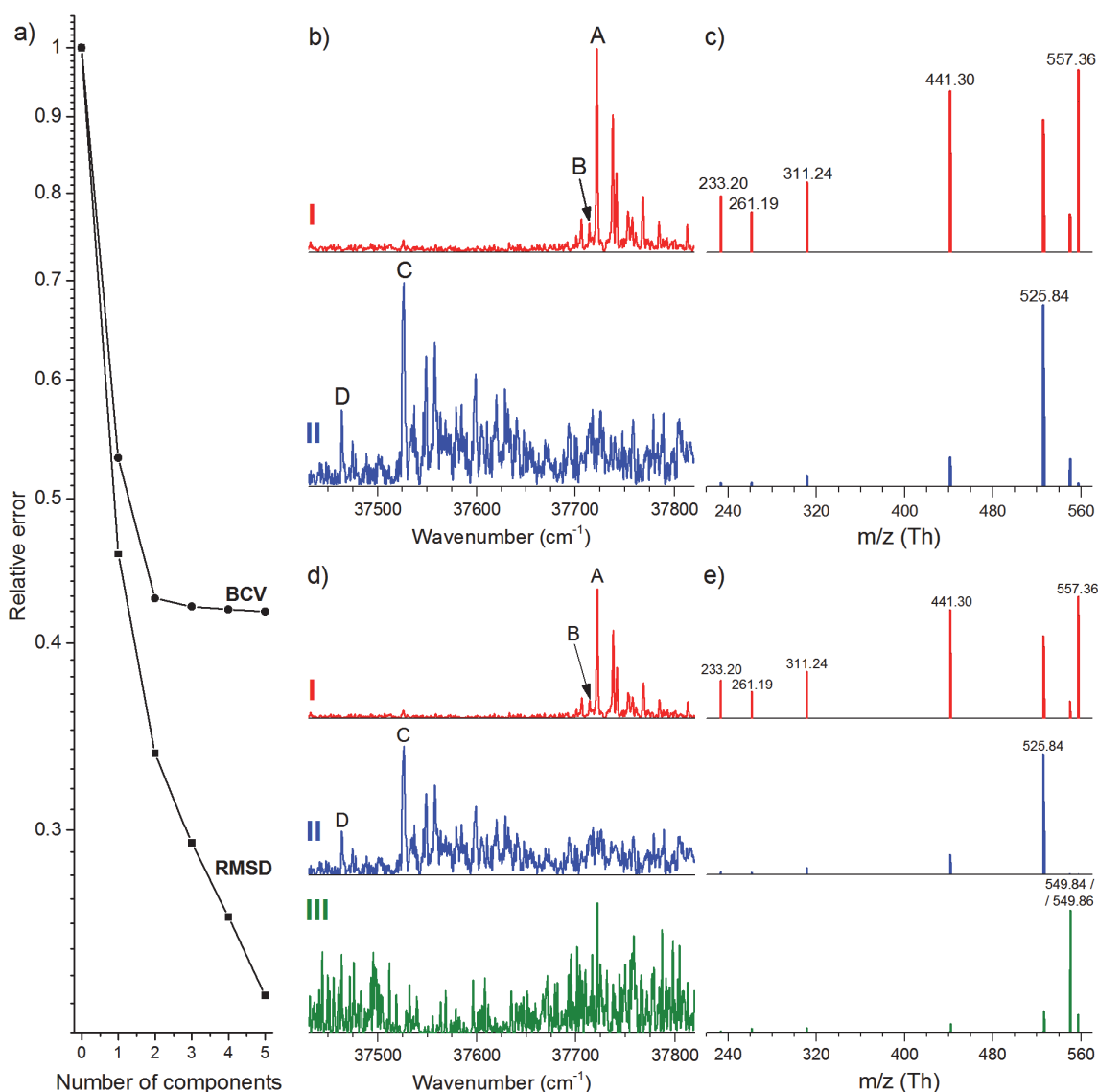
Figure 5.4. **(a)** Plot of the RMSD and BCV errors as a function of the number of components, used for non-negative matrix factorization of the UV-MS spectrum of $[GS + 2H]^{2+}$. **(b)** UV absorption and **(c)** photofragmentation mass spectra, derived by two-component factorization of the UV-MS matrix. **(d)** UV absorption and **(e)** photofragmentation mass spectra, derived by three-component factorization of the UV-MS matrix. UV spectra are normalized to the maximum signal and MS spectra are normalized by the intensity of the most abundant fragments. Peaks in the UV spectra are labeled according to the assignment in ref. [249]. The UV fragmentation was assisted by a subsequent pulse of CO_2 laser.[178]

Phe-containing protonated peptides.[168,178,246,250] Additionally, the fragmentation mass spectrum of the component I contains four peaks at lower m/z. On the basis of their accurate

masses and isotopic patterns we assign the fragments at m/z = 261.19 and 311.24 Th to the singly protonated b/y-ions, containing FL and PVO amino-acid sequences, respectively, and the fragment at m/z = 441.30 Th corresponds to the doubly protonated PVOLFPO sequence. Finally, the neutral loss of CO by the fragment at m/z = 261.19 Th yields the fragment at m/z = 233.20 Th. Thus, among all the identified photofragments only the Phe side-chain loss results from the prompt dissociation in the electronically excited state directly after the UV excitation of the Phe chromophore, while the others are products of the statistical dissociation in the ground electronic state.

It is worth noting that due to a low resolution of the previously recorded photofragmentation mass spectra the peak at m/z = 557.36 Th was erroneously ascribed to the cleavage of the C_γ-C_δ bond in the Orn side chain and this finding was interpreted as an evidence of the energy transfer from the UV-excited Phe chromophore to the nearby protonated amino group of the Orn side chain. [176] The Phe side-chain loss, on the contrary, was not observed, likely, because a large molecule such as GS requires an additional activation (e.g., by means of a CO_2 laser) to overcome a dissociation barrier in the triplet state.

Similar to the case of the dipeptide, the difference in the photofragmentation spectra of the components I and II are related to the structural features of the corresponding conformers. The spectrum of the component II is dominated by the Phe side-chain loss fragment, suggesting, that the phenyl rings in the conformers C/D are likely to be remote from the rest of the ion and are not involved in any (at least, strong) interactions with the backbone or the protonated Orn side chains. Conversely, a substantial interaction between the chromophores and their local environment in the conformers A/B leads to the efficient internal conversion from the electronically excited state and subsequent thermal fragmentation from the ground electronic state. Indeed, the known structures of the conformers A/B exhibit a tight packing of the aromatic rings to the backbone by π-proton interaction of the aromatic rings with NH_3^+ of the nearby Orn residues. [176,307,308] In addition to the observed fragmentation pattern, in A/B conformers this coupling leads to significant blue shifts of the UV band origins relative to the band origin in neutral Phe. [54,67] On the other hand, the lack of such a shift for the conformers C/D implies no coupling between the chromophores and the charged group, which is consistent with the relative isolation of the chromophore in these conformers.
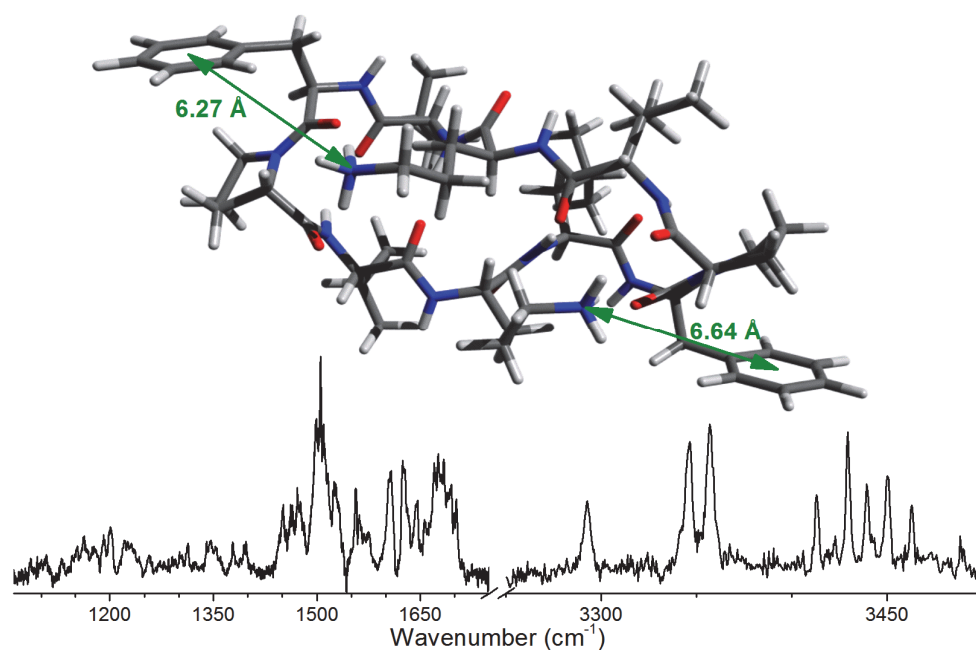
Figure 5.5. Candidate structure and IR-UV depletion spectrum of the conformer C of doubly protonated GS (the UV fragmentation was assisted by a subsequent pulse of CO_2 laser). The 3D structure is generated from the Cartesian coordinates of the structure R06 in ref. [309], for which the NH_3^+-Phe distances is 6.3 and 6.6 Å compared with 3 Å for the most stable conformer. For details of the measurement of the IR depletion spectrum see Supporting Information in ref. [307].

Such structures were previously predicted by means of high-level density functional theory (DFT) calculations as the low-energy geometries of the gas-phase GS, namely, structures R04-R08 in ref. [309]. The structures R04, R05, and R08 belong to the C_2 point symmetry group, and therefore, only four NH stretches are expected to be observed in the IR spectrum of these conformers. Since the experimental IR spectrum of the conformer C exhibits, at least, five distinct NH stretch vibrations, the structures R04, R05, and R08 can be discarded from further consideration. In the structure R07 only one of the Phe rings is remote from the charge, while another Phe ring is rather close to the NH_3^+ group. Since both Phe rings should be isolated in the structure of the conformer C, we believe that the structure R07 can also be disregarded. Thus, only the structure R06 meets all the requirements derived from the analysis of the experimental data. Although the final validation still requires calculation of its accurate IR spectrum and subsequently its comparison with the experimentally measured IR

spectrum (Figure 5.5), we believe that the structure R06 or a similar one, likely, corresponds to the experimentally observed conformer C of gramicidin S.

5.4. Conformers of a Drug Molecule, Homatropine

A quick assessment of conformational heterogeneity might be particularly useful in drug design. Unlike peptides, most of the common drugs have no well-defined structural units, which complicates the structural analysis of such species. In this regard, we applied the UV-MS approach to singly protonated homatropine, a small drug molecule, used in eye drops as an anticholinergic agent. Homatropine belongs to a family of tropane alkaloids, that is, plant-derived organic compounds, containing a tropane ring in their chemical structure. They are among the most economically important pharmaceuticals[310] and are applied as mydriatics, antiemetics, antispasmodics, anesthetics and bronchodilators.[311] Moreover, one of the principal drugs of abuse, cocaine, is also a tropane derivative.

Neither the number of conformers nor their 3D structures were known for gas-phase homatropine. The BCV analysis (Figure 5.6(c)) clearly indicates the presence of two significant components in the UV-MS spectrum of this ion. The derived photofragmentation mass spectra (Figure 5.6(e)) of the components I and II exhibit predominant fragments at m/z = 259.158 and 260.162 Th. The peak at 259.158 Th can be unambiguously (mass accuracy is 5 ppm) assigned to the neutral loss of the OH group. And the peak at 260.162 Th is shifted exactly by the ^{12}C-^{13}C mass difference. This suggests that the component II corresponds to the isotopologue of the species associated with the component I with one ^{13}C isotope. Notably, the OH loss is the major dissociation channel of neutral benzyl alcohol (C_6H_5OH) upon its photoexcitation at 193 nm,[312] which essentially is the chromophore of homatropine. This fragment is absent in the HCD mass spectrum of homatropine (Figure 5.6(b)), in contrast to the fragments at m/z = 124.112 and 142.123 Th, which correspond to the esteric CO bond cleavages.[313] Thus, we conclude that UV excitation of homatropine mainly results in prompt dissociation, which occurs on the electronically excited state.

Figure 5.6(d) shows the derived UV spectra, which display nicely resolved peaks with the full width at half maximum of about 2 cm^{-1}. We assign the leftmost peak, indicated by the
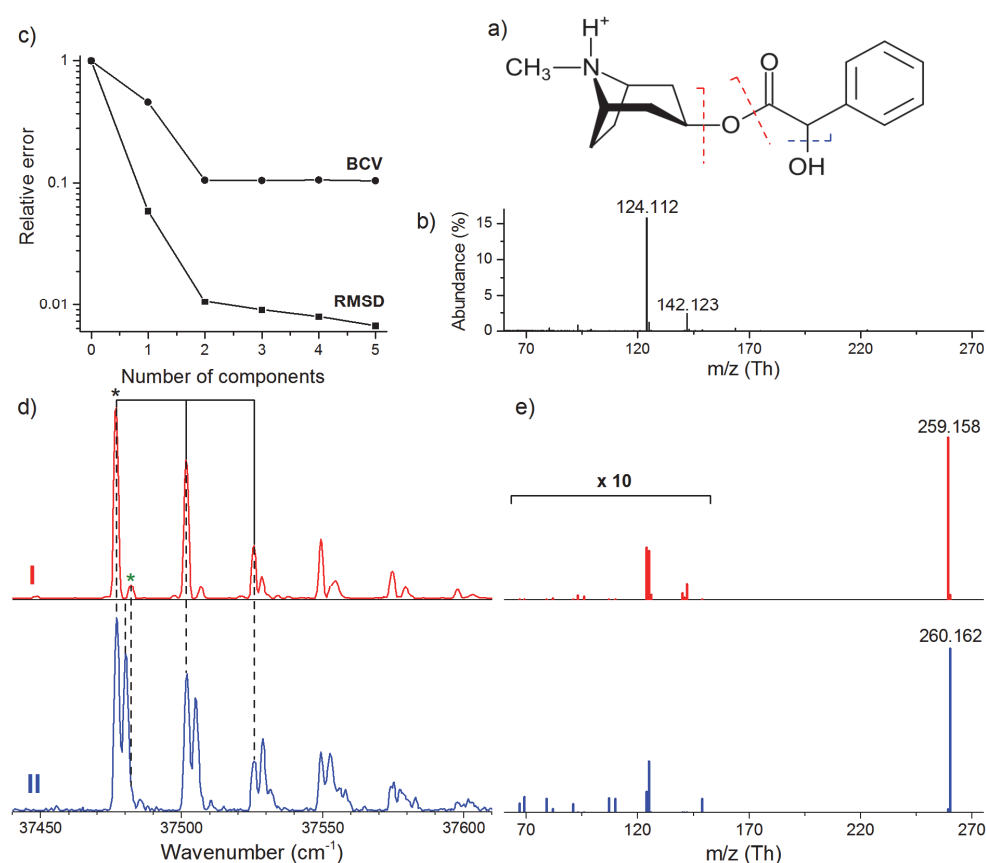
Figure 5.6. **(a)** Chemical structure of singly protonated homatropine and **(b)** its HCD fragmentation mass spectrum, measured at a HCD collision energy of 20 eV. **(c)** Plot of the RMSD and BCV errors as a function of the number of components, used for non-negative matrix factorization of the UV-MS spectrum of singly protonated homatropine. **(d)** UV absorption and **(e)** photofragmentation mass spectra, derived by two-component factorization of the UV-MS matrix. **(a)** The bond cleavages, corresponding to the main thermal and prompt dissociation channels, are indicated by the red and blue dashed lines, respectively. **(b)** For graphical clarity the base peak, corresponding to the parent ions at m/z = 276.159 Th is not shown. **(d)** UV spectra are normalized to the maximum signal. The band origins of the major and minor conformers are designated by the black and green asterisks, respectively. The correspondence between the peaks in the spectra of the two components is indicated by the dashed lines. A series of solid lines points to the first three peaks of the lowest-frequency Franck-Condon progression. **(e)** Mass spectra are normalized to the intensity of the most abundant fragments. For graphical clarity the m/z scale is aligned with that of the HCD fragmentation mass spectrum and the peaks in the 60-150 Th region are enlarged by a factor of 10.

86

black asterisk in Figure 5.6(d), to the absorption band origin of one of the conformers of protonated homatropine. It appears in the UV spectra of both components at 37476.9 cm^{-1}, only 0.5 cm^{-1} red-shifted with respect to the band origin of neutral toluene.[314] Similarly to the considered above cases of dipeptide and gramicidin S, the lack of a significant spectral shift in the UV spectrum and the predominance of the prompt photofragmentation in the mass spectrum allow suggesting that in protonated homatropine the aromatic ring is remote to the rest of the ion and, in particular, does not interact with the protonated tertiary amino group.

In the spectrum of the component II the band origin at 37476.9 cm^{-1} slightly overlaps with a relatively intense peak at 37480.2 cm^{-1}. Since this component corresponds to the isotopologue, the observed splitting of the band origin may result from the change of the ZPVEs of the S_0 and S_1 states upon the isotopic substitution. However, unlike in the case of the dipeptide, the isotopic substitution does not lead to a spectral broadening. Indeed, if the high symmetry of the aromatic ring is not perturbed by its interactions with the local environment, the spectral shift of the band origin upon the isotopic substitution is relatively insensitive to the exact substitution position. For instance, in neutral toluene the substitution of hydrogen by deuterium in the *orto-*, *meta-*, and *para-*positions yields the average spectral shift of 32.8 cm^{-1}, while the dispersion of the spectral shift values is only 1.9 cm^{-1}.[315] A fit of the split band origin with two Gaussian functions (not shown here) yields the 8.5:7.5 ratio of integral intensities for the two peaks (there are 16 carbons in total). Therefore, one may suggest, that the isotopic substitution of any carbon in the aromatic ring and, may be, one or two carbons, that are nearest to it, lead to the spectral shift of 3.3 cm^{-1}, while the substitution of the other carbons almost does not affect the position of the band origin. This finding provides an additional evidence of the free chromophore, non-interacting with its local environment.

In the spectrum of the component I the band origin is accompanied by a small peak at 37482.1 cm^{-1} (designated by the green asterisk in Figure 5.6(d)). Although the photofragmentation mass spectrum of this component exhibits a tiny isotopic peak at 260.162 Th, mainly the component I corresponds to the monoisotopic species of homatropine. Hence, this satellite peak is unlikely to be explained by any isotopic substitutions. Then, the UV spectrum of the component I displays a low-frequency ($\approx$ 25 cm^{-1}) Franck-Condon progression. However, the presence of a very low-energy (5.2 cm^{-1}) vibration, which could explain the satellite peak, is also ruled out for such a relatively small molecule (for instance, the lowest-
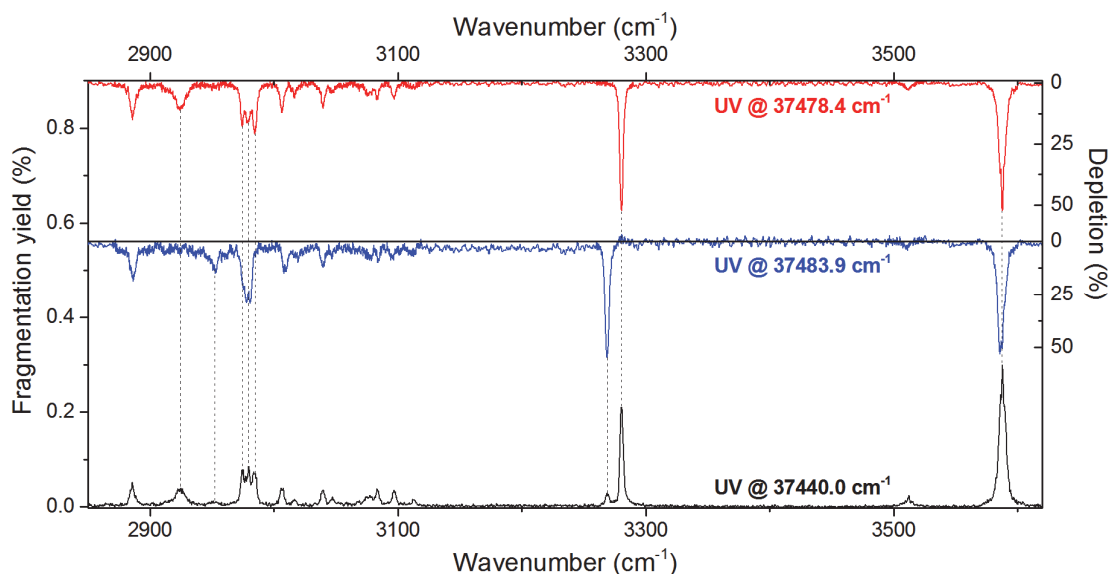
Figure 5.7. Experimental IR-UV gain (black) and depletion (red and blue) spectra, measured by isolating the parent ions at m/z = 276 Th with the first quadrupole (Q1) mass filter, fragmenting them by a UV laser pulse in the cold ion trap and monitoring the fragmentation signal at m/z = 259 Th with the second quadrupole (Q3) mass filter in the two alternating cycles, with and without an IR laser pulse, preceding the UV pulse. The spectra are obtained as a difference between the IR-UV and UV-only signals. The depletion spectra are normalized to the UV-only signal and the gain spectrum to the number of parent ions.

energy Franck-Condon-active vibration in an ion of comparable size, $[YA + H]^+$, has a frequency of 13.3 cm^{-1} (Figure 5.1(b))). We thus conclude that the satellite peak at 37482.1 cm^{-1} must be the band origin of another, minor conformer, which has a photofragmentation mass spectrum similar to that of the major conformer and therefore could not be disentangled by the UV-MS method. The very small difference in the UV band origin positions and the identical fragmentation patters of the conformers of such a small molecule suggests a considerable similarity of their 3D structures.

This conclusion has been further confirmed by IR-UV double resonance spectroscopy. The all-conformer IR spectrum, shown in Figure 5.7, contains two peaks in the region of NH stretch vibrations, one of which is an order of magnitude more intense, than the other. The conformer-selective IR spectra measured with the UV laser wavenumber fixed at 37476.9 and 37482.1 cm^{-1} exhibit only one NH stretch either at 3281.3 or 3268.7 cm^{-1}, respectively. They also reveal differences in the region of CH stretches, while the peak, corresponding to the OH

88

stretch, appears in all the IR spectra at 3587.6 cm^{-1}. Thus, the peak at 37482.1 cm^{-1} is indeed the band origin of a minor conformer, which can, for instance, have an inverted arrangement of the proton and the methyl group around the tetrahedral nitrogen atom.

Overall, the performed analysis of the 2D UV-MS spectrum of singly protonated homatropine allowed us to correctly identify one major and one minor conformer (and even their isotopologue) and predict that in both conformers, which are likely to have a very similar 3D structure, the aromatic ring is not involved in any interactions with the rest of the ion.

5.5. Conclusions

In conclusion, we have demonstrated a use of non-statistical UV fragmentation, detected by high-resolution 2D UV-MS cold ion spectroscopy, for revealing conformers of biomolecular ions. Unlike, for instance, the ion mobility spectrometry, which characterizes each conformer (or conformational family) by a single value, collisional cross-section, the UV-MS approach provides conformer-specific UV absorption and photofragmentation mass spectra, which can be used for a rational analysis of molecular structures.

We have validated our approach using a benchmark singly protonated dipeptide, Tyr-Ala, for which, in addition to two families of the previously known conformers, their isotopologues were unambiguously identified. On the basis of the calculated 3D structures of the conformers we propose a model, which, in particular, explains the difference in their photofragmentation mass spectra and, in general, establishes a correlation between the relative abundance of the side-chain loss fragments, resulting from the prompt dissociation of aromatic amino-acid residues upon their UV excitation, and the extent of interaction of the respective chromophore with its local environment.

Application of the UV-MS analysis to a doubly protonated decapeptide, gramicidin S, reveals, that the Phe aromatic rings in the conformer C, for which the geometry has not yet been validated, should be highly isolated from the peptide backbone. This conclusion constrains the most likely structure of conformer C to one of the previously calculated low-energy geometries of gramicidin S.

Finally, we have demonstrated, that the UV-MS approach can be equally well applied for identification of conformers of non-peptidic biomolecular ions, containing a UV-absorbing chromophore. The conclusions about the structural features of a singly protonated drug molecule, homatropine, drawn on the basis of a careful analysis of its UV-MS spectrum, have been fully confirmed by the experimentally measured conformer-selective IR spectra.

Chapter 6

Interchromophore Distance Measurements in Gas-Phase Opioid Peptides

In this chapter we employ 2D UV-MS spectroscopy for measuring the distances between natural phenylalanine and tyrosine chromophores in four opioid peptides. We first provide evidence of the energy transfer from phenylalanine to tyrosine and demonstrate how the efficiency of such a process can be evaluated on the basis of UV-MS spectra. Then, we propose an approach for deriving the interchromophore distances from these data. Finally, we calculate the precise three-dimensional structures of the lowest energy conformers of the peptides and discuss them in the context of the experimental results.

6.1. Introduction

The biological functionality of peptide ligands is governed by the three-dimensional (3D) geometry they adopt at the binding sites of target proteins. These structures, which are typically determined in the condensed phase by NMR spectroscopy or X-ray crystallography, are often used to predict the pharmacological efficiency of candidate drugs. However, the prediction may become ambiguous when peptides lack rigidity and can undergo significant conformational changes upon interaction with receptors in vivo, particularly because of the hydrophobic environment of their binding pockets.[290,316] Isolation of biomolecules in the gas phase removes intermolecular interactions, creating a water-free environment. Thus, structural features measured in the gas-phase peptides may complement the condensed-phase data for modeling ligand-receptor interactions. Opioid peptides, which bind to

transmembrane opioid receptors, have been used as efficient drugs for pain treatment for decades.[317–319] They all contain two aromatic rings, the spacing between which is considered one of the key structural parameters that determine the pharmacology of these drugs.[290,320,321]

Although resonance energy transfer (RET) in solution is a proven approach that provides an estimate of the average interchromophore distances in biomolecules,[322,323] researchers have only recently demonstrated the use of fluorescence[324–326] and, more recently, photodissociation mass spectrometry[327–329] for detecting RET in gas-phase ions labeled with large chromophores. Such labeling would not be appropriate for opioid peptides, which experience drastic pharmacology changes with only slight structural modifications.[320,330,331]

Instead, the Tyr and Phe aromatic residues naturally contained in these peptides can be involved in RET.[332] Herein, we employ UV fragmentation spectroscopy-mass spectrometry (UV-MS) of ultracold ions to measure conformer-specific efficiencies of resonance energy transfer (RET) in four similar opioid peptides, leucine enkephalin (Tyr-Gly-Gly-Phe-Leu) and three stereoisomers of [Ala^2, Leu^5]-leucine enkephalin (Tyr-Ala-Gly-Phe-Leu), isolated in the gas phase. Recording the entire fragmentation mass spectrum at each UV wavenumber allows for more accurate measurements, and the high spectral resolution offered by cooling makes them conformer selective. We use this data, the known gas-phase geometry of singly protonated leucine enkephalin, and our low-level structural calculations to estimate the Tyr-Phe ring spacing in the stereoisomers and compare these distances with the pharmacological activity of the respective ligands.

6.2. Evidence for Energy Transfer from Phenylalanine to Tyrosine in Gas-Phase Opioid Peptides

The four pentapeptides studied here possess two common chromophores, namely, tyrosine (Tyr) and phenylalanine (Phe). Figure 6.1(a) shows the UV absorption spectrum of singly protonated leucine enkephalin (denoted below as LE) in the region of the Tyr absorption band origin, generated by integrating the measured UV–MS spectrum over the most prominent fragments. The band origin of Tyr in LE appears at 35683.1 cm^{-1}, quite close to that

of the neutral Tyr (various conformers have their band origins in the 35492-35660 cm^{-1} region)[54] and significantly blue-shifted with respect to those of the protonated Tyr (35081-35235 cm^{-1}).[168] This blue shift was previously associated with the size of the respective ions.[247] Indeed, a correlation between the size of the CID fragments of LE and the blue shift of their band origins of Tyr absorption was observed,[172] however, it is more likely that the position of the Tyr band origin in LE reflects the strength of the interaction between Tyr aromatic ring and the charge, localized on the N-terminus. The gas-phase structure of protonated LE, calculated by Burke et al.,[247] reveals that the N-terminal protonated amino group is involved in a relatively strong hydrogen bonding. Namely, it forms a C_{11} hydrogen bond with the CO group of the third residue, glycine, and a C_{17} hydrogen bond with the C-terminal CO group. The third hydrogen of the amino group faces the Tyr phenyl ring and might be involved in a proton-π interaction. However, this interaction is much less strong, therefore, the charge should be localized on the other two hydrogens. This is consistent with the experimental IR spectra of protonated LE and tyrosine. In the protonated tyrosine the proton-π interaction shifts the frequency of the NH stretch, associated with the π-bound hydrogen,
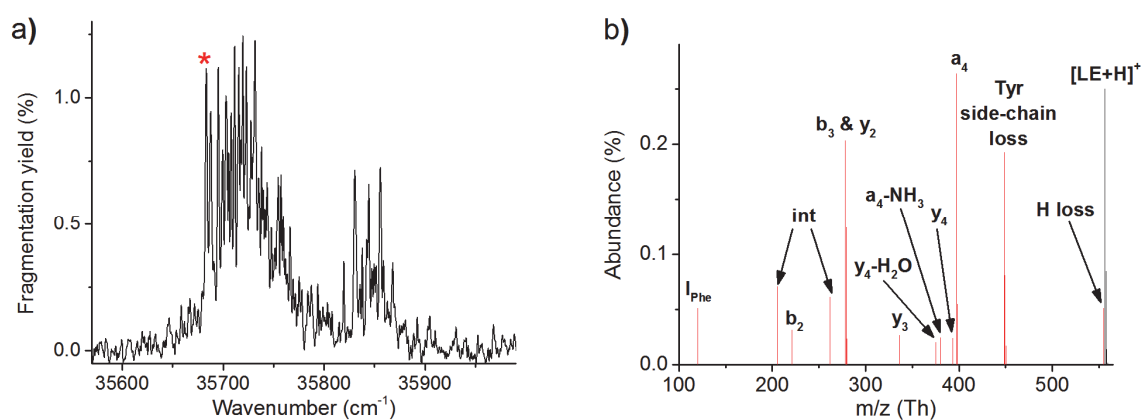
Figure 6.1. (a) UV absorption and **(b)** photofragmentation mass spectra of singly protonated leucine enkephalin. **(a)** UV spectrum is obtained by integrating the UV-MS spectrum over the most prominent fragments. **(b)** Photofragmentation mass spectrum is obtained by averaging the UV-MS spectrum over the UV peak, indicated by the asterisk. The rightmost peak, labeled [LE+H]$^+$, corresponds to the parent ion with m/z = 556.281 Th (mass accuracy is 7 ppm). Labels int and I$_{Phe}$ denote internal fragments and the immonium fragment corresponding to Phe residue, respectively. For graphical clarity the intensity of the [LE+H]$^+$ peak is reduced by a factor of 400.

to the 3000-3200 cm^{-1} region and the free NH stretch appears in the 3320-3350 cm^{-1} region.[168] On the other hand, in the protonated LE the frequencies of the two NH stretches, associated with the NH$_3^+$ group, are shifted down to the 2900-3100 cm^{-1} region, while the NH stretch, corresponding to the hydrogen facing the phenyl ring, appears at around 3300 cm^{-1}.[247] This suggests rather weak proton-π interaction and explains the lack of the spectral shift of the Tyr band origin.

Overall, the UV absorption spectrum of LE near the Tyr band origin, shown in Figure 6.1(a), looks very similar to that recorded by Burke et al.,[247] although distinct vibronic peaks are much narrower (FWHM ≈ 1.8 cm^{-1}) in the spectrum presented here (this may be, for instance, due to a better cooling of ions in our cold trap). Since the UV spectra of cryogenically cooled Tyr- and Phe-containing protonated peptides are known to be sensitive to their 3D structure,[168,170,249] we can assume that the protonated LE adopt the same 3D structure as in the experiments performed by Burke et al. Thus, this allows us to use the previously calculated and validated 3D structure of the protonated LE as a reference for further discussion.

The photofragmentation mass spectrum of LE, generated by averaging the UV-MS spectrum over the Tyr band origin, also resembles the one recorded by Burke et al.[247] It mainly consists of the known CID-like fragments such as b- and y-type ions, internal and immonium fragments, and the a$_4$-ion, specific to fragmentation of the LE peptide.[255] In addition to these statistical fragments, the spectrum exhibits the Tyr side-chain loss fragment, which is known to be due to a prompt dissociation taking place on the electronically excited state directly after the UV excitation of the Tyr chromophore,[168,170,183,231,246–248] and the fragment, corresponding to the neutral loss of the hydrogen atom. It was shown that the H-loss is the dominant prompt photodissociation channel of protonated tryptophan, but it was not directly observed in the photofragmentation mass spectrum of protonated tyrosine.[248] Here, we could readily detect and unambiguously assign this fragment due to the high mass accuracy and resolution of the Orbitrap mass analyzer.

The second chromophore present in LE, namely, phenylalanine starts absorbing the UV light at higher wavenumbers, quite far from the Tyr band origin. Figure 6.2(a) shows the UV absorption spectrum of LE, obtained by integrating the UV–MS spectrum, recorded in a wide range of wavenumbers, over the most prominent fragments. The intensity in this spectrum is mainly due to the strong absorption of Tyr. The spectrum displays a number of vibronic bands,

the most pronounced of which are due to the breathing mode of the Tyr phenyl ring (shown in Figure 6.2(a)).[250,333,334] Although some of the bands retain fine structure (the linewidth of distinct peaks is about a few cm^{-1}) even 1500 cm^{-1} above the band origin, they are getting
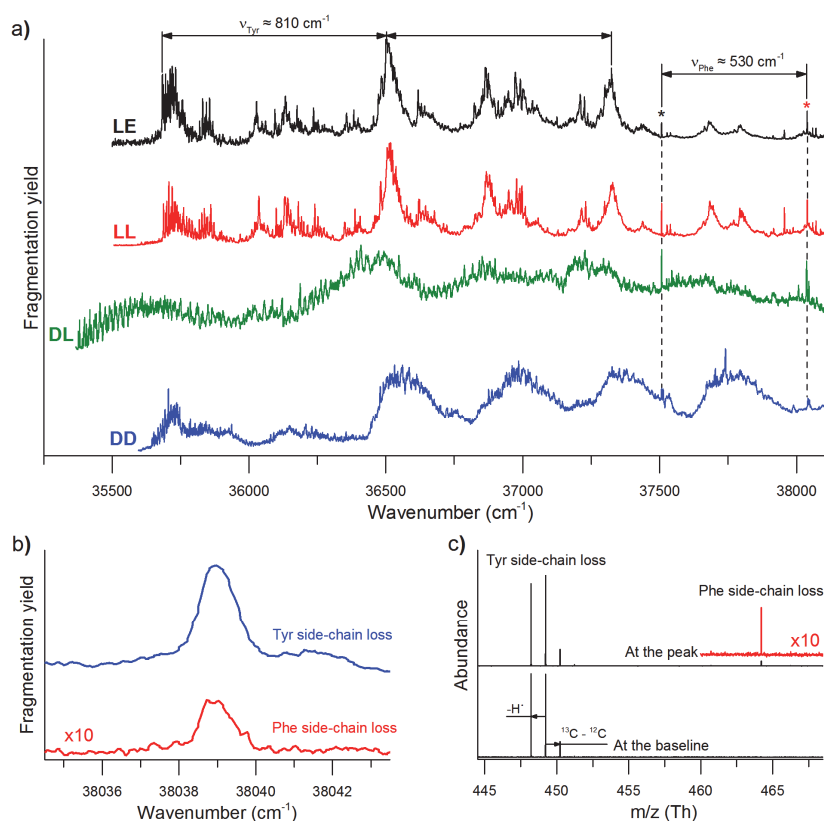
Figure 6.2. (a) UV absorption spectra of singly protonated LE (black), LL (red), DL (green) and DD (blue) enkephalins obtained by integrating the respective UV–MS spectra over the most prominent fragments. For graphical clarity, the spectra are normalized to their maximum fragmentation yields. The spectra of LL, DL and DD are offset by +3.2, +123.4 and +246.2 cm^{-1}, respectively, to align their Phe band origins with that of the LE, indicated by the black asterisk. **(b)** High-resolution UV absorption spectra of the LE in the region around the peak, indicated by the red asterisk. The spectra correspond to the side-chain losses of Tyr (blue) and Phe (red). For graphical clarity, the lower trace is enlarged by a factor of 10. **(c)** Photofragmentation mass spectra of the LE obtained by averaging the UV-MS spectrum over the wavenumber ranges of 38035-38036 cm^{-1} (bottom) and 38038.5-38039.5 cm^{-1} (top). The red trace corresponds to the Phe side-chain loss peak enlarged by a factor of 10.

noticeably broader upon increase of the excitation wavenumber. Therefore, the sharp peak (FWHM ≈ 1.5 cm^{-1}) at 37 507.6 cm^{-1} on top of the comparatively smooth absorption by Tyr, has been assigned to the band origin of Phe. Such narrow peaks in the region of wavenumber around 37500 cm^{-1} are characteristic of Phe-containing cold peptides.[168,171,178,249,280,335] Another sharp peak at 38038.9 cm^{-1} further supports our assignment, because the difference between the wavenumbers of these two peaks, 531.3 cm^{-1}, corresponds well to the 6b non-totally symmetric vibration in n-alkylbenzenes[314] as well as in phenylalanine and Phe-containing peptides.[168,250,295,336] Finally, the photofragmentation mass spectrum at this vibronic peak (Figure 6.2(c)) exhibits the Phe side-chain loss fragment, which is known to arise upon UV excitation of the Phe chromophore in gas-phase peptides.[168,178,246,250]

The UV absorption spectra in the region around the peak at 38038.9 cm^{-1}, taken as slices of the UV-MS spectrum at the Tyr and Phe side-chain losses, are presented in Figure 6.2(b). It turned out, that the UV excitation of Phe results not only in the Phe side-chain loss, but also in the Tyr side-chain loss. Moreover, the apparent fragmentation yield to the Tyr side-chain loss channel is higher by an order of magnitude. This fact suggests the presence of the energy transfer from the Phe to the Tyr chromophore. Indeed, as mentioned above, both side-chain losses are associated with the processes, which occur in the electronically excited states after the UV excitation of a chromophore, and cannot be produced by a statistical dissociation. Therefore, the UV excitation has to be somehow transferred from the S_1 state of Phe to the S_1 state of Tyr. At interchromophore distances not exceeding a few nanometers this process proceeds non-radiatively and is typically described by either Förster or Dexter mechanisms.[31]

In addition to the UV spectrum of LE, Figure 6.2(a) shows the absorption spectra of three stereoisomers, namely, [L-Ala^2, L-Leu^5]-, [D-Ala^2, L-Leu^5]-, and [D-Ala^2, D-Leu^5]-enkephalins (denoted below as LL, DL, and DD, respectively), generated by integrating the measured UV-MS spectra over the most prominent fragments. They differ from LE by the second amino-acid residue (i.e., Ala instead of Gly). However, the UV spectra of LE and LL look quite similar, which suggests a likely structural similarity of these peptides. In contrast, the spectra of DL and DD differ strikingly in the position of the Tyr and Phe band origins and in the shape of the UV bands. This observation implies quite different 3D structures of DL and DD compared with those of LE and LL, which can be associated with the changes in the chirality of Ala and Leu residues. This will be discussed later in the context of our ET measurements and 3D structure

calculations. We only note that we have assigned the Phe band origins in the UV spectra of the three stereoisomers (in Figure 6.2(a) they are aligned with the Phe band origin in LE) and they all exhibit the energy transfer from the Phe chromophore to the Tyr chromophore.

6.3. Evaluation of the Energy Transfer Efficiencies

In principle, the peptides can exist in a number of conformations. We denote this number by k and the number of ions of the i-th conformer, present in the cold trap at T ≈ 10 K, by N_i. However, the position of the Phe band origin is known to be sensitive to the conformational state of a peptide.[54,168,249] Thus, the identified Phe band origins very likely belong to a single conformer, say, to the j-th conformer, and the ET efficiency (φ_{ET}) should be associated only with this conformer. In order to evaluate φ_{ET}, we assume that the UV laser fluence (Φ) is sufficiently low, so that both Tyr and Phe chromophores absorb the UV light linearly with respect to Φ. Then, a general expression for the photofragmentation signal at a certain $^m/_z$ in the region around the Phe band origin of the j-th conformer is:

$$I(^m/_z, \tilde{v}) = \sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)}(\tilde{v}) \Phi(\tilde{v}) \cdot \varphi_{Tyr}^{(i)}(^m/_z, \tilde{v}) + \tag{6.1}$$

$$N_j \cdot \sigma_{Phe}(\tilde{v}) \Phi(\tilde{v}) \cdot \left[(1 - \varphi_{ET}) \cdot \varphi_{Phe}(^m/_z, \tilde{v}) + \varphi_{ET} \cdot \varphi_{Tyr}^{(j)}(^m/_z, \tilde{v})\right]$$

where $\sigma_{Tyr}^{(i)}(\tilde{v})$ and $\varphi_{Tyr}^{(j)}(^m/_z, \tilde{v})$ are the absorption cross-section of the Tyr chromophore of the i-th conformer at wavenumber $\tilde{v}$ and its photodissociation yield to $^m/_z$ fragment upon the UV excitation of Tyr; $\sigma_{Phe}(\tilde{v})$ and $\varphi_{Phe}(^m/_z, \tilde{v})$ are the respective absorption cross-section and the photodissociation yield of the Phe chromophore of the j-th conformer. The first term in this expression corresponds to the Tyr absorption in all the conformers, while the second represents only the Phe absorption in the j-th conformer. Note, that the absorption by the Tyr chromophore in the region around the Phe band origin is relatively smooth and does not change significantly over the width of the peak (see Figure 6.2(b)). The laser fluence, $\Phi(\tilde{v})$, also stays almost constant within this narrow spectral region. Therefore, we disregard their dependencies on the wavenumber and consider them as constants, $\sigma_{Tyr}^{(i)}$ and Φ.

On the contrary, the Phe absorption cross-section is strongly wavenumber-dependent. It reaches a maximum at the center of the peak ($\tilde{v}_0$) and drops to zero at the baseline ($\tilde{v}$). Therefore, we can write down the photofragmentation signal at the baseline and at the peak maximum ($\sigma_{Phe}(\tilde{v}_0) \equiv \sigma_{Phe}$, $\sigma_{Phe}(\tilde{v}) \equiv 0$) as follows:

$$I_{BL}(^m/_z) = I(^m/_z, \tilde{v}) = \sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)} \Phi \cdot \varphi_{Tyr}^{(i)}(^m/_z, \tilde{v})$$

$$I_{peak}(^m/_z) = I(^m/_z, \tilde{v}_0) = \sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)} \Phi \cdot \varphi_{Tyr}^{(i)}(^m/_z, \tilde{v}_0) +$$

$$+ N_j \cdot \sigma_{Phe} \Phi \cdot \left[(1 - \varphi_{ET}) \cdot \varphi_{Phe}(^m/_z, \tilde{v}_0) + \varphi_{ET} \cdot \varphi_{Tyr}^{(j)}(^m/_z, \tilde{v}_0) \right]$$

(6.2)

Generally speaking, the photodissociation yields, associated with the UV excitation of Tyr ($\varphi_{Tyr}^{(i)}$) and Phe (φ_{Phe}) chromophores, depend on the UV photon wavenumber.[250] For instance, the photodissociation yield of thermal, CID-like fragments is obviously determined by the amount of internal energy available for the breaking molecular bonds. But the relative change in the UV photon energy is less than 0.01% on the width of the Phe peak (FWHM ≈ 1 cm^{-1}), which implies, that the photodissociation yield of thermal fragments should be a constant in this wavenumber region. Thus, the photofragmentation signals due to the Tyr absorption (I_{BL}) and due to the Phe absorption ($\Delta I = I_{peak} - I_{BL}$) integrated over all the prominent b- and y-fragments are:

$$I_{BL}(b/y) = \sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)} \Phi \cdot \varphi_{Tyr}^{(i)}(b/y)$$

$$\Delta I(b/y) = N_j \cdot \sigma_{Phe} \Phi \cdot \left[(1 - \varphi_{ET}) \cdot \varphi_{Phe}(b/y) + \varphi_{ET} \cdot \varphi_{Tyr}^{(j)}(b/y) \right]$$

(6.3)

The prompt photodissociation, especially if it requires overcoming of a certain energy barrier, might be more sensitive to the UV photon energy, because the internal energy in an electronically excited state is lower compared with that in the ground state ($E_{int} = hc\tilde{v} - E_{0-0}$). However, the internal energy available in the S_1 state of Tyr upon its direct (by the UV light) or indirect (by the ET from Phe) excitation in the region of Phe absorption is almost 2000 cm^{-1}, which is still much higher than its increase by about 1 cm^{-1}. Thus, the photodissociation yields of the Tyr side-chain loss at the baseline and at the maximum of the Phe peak should not differ, at least, significantly. Taking into account that the UV excitation of Phe cannot

98

directly lead to the C_α-C_β cleavage of Tyr ($\varphi_{Phe}(C_\alpha\text{-}C_\beta) = 0$), the baseline signal ($I_{BL}$) and the peak height ($\Delta I$) at the Tyr side-chain loss fragment is:

$$I_{BL}(C_\alpha\text{-}C_\beta) = \sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)} \Phi \cdot \varphi_{Tyr}^{(i)}(C_\alpha\text{-}C_\beta) \tag{6.4}$$

$$\Delta I(C_\alpha\text{-}C_\beta) = N_j \cdot \sigma_{Phe}\Phi \cdot \varphi_{ET} \cdot \varphi_{Tyr}^{(j)}(C_\alpha\text{-}C_\beta)$$

Then,

$$\frac{I_{BL}(C_\alpha\text{-}C_\beta)}{I_{BL}(b/y)} = \frac{\sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)} \cdot \varphi_{Tyr}^{(i)}(C_\alpha\text{-}C_\beta)}{\sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)} \cdot \varphi_{Tyr}^{(i)}(b/y)} \tag{6.5}$$

and

$$\frac{\Delta I(C_\alpha\text{-}C_\beta)}{\Delta I(b/y)} = \frac{\varphi_{ET} \cdot \varphi_{Tyr}^{(j)}(C_\alpha\text{-}C_\beta)}{(1-\varphi_{ET}) \cdot \varphi_{Phe}(b/y) + \varphi_{ET} \cdot \varphi_{Tyr}^{(j)}(b/y)}$$

$$= \frac{\varphi_{ET} \cdot \dfrac{\varphi_{Tyr}^{(j)}(C_\alpha\text{-}C_\beta)}{\varphi_{Tyr}^{(j)}(b/y)}}{\varphi_{ET} + (1-\varphi_{ET}) \cdot \dfrac{\varphi_{Phe}(b/y)}{\varphi_{Tyr}^{(j)}(b/y)}} \tag{6.6}$$

Because the b- and y-fragments are generated by the thermal dissociation process, their fragmentation yields should be mainly determined by the amount of vibrational energy, that the parent ions have in the ground state. Therefore, for CID-like fragments we assume $\varphi_{Phe} \approx \varphi_{Tyr}$. Finally, the ratio of the photodissociation yields of the Tyr side-chain loss and of the thermal fragments for the j-th conformer is approximated by an average over all conformers:

$$\frac{\varphi_{Tyr}^{(j)}(C_\alpha\text{-}C_\beta)}{\varphi_{Tyr}^{(j)}(b/y)} \approx \frac{\sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)} \cdot \varphi_{Tyr}^{(i)}(C_\alpha\text{-}C_\beta)}{\sum_{i=1}^{k} N_i \cdot \sigma_{Tyr}^{(i)} \cdot \varphi_{Tyr}^{(i)}(b/y)} = \frac{I_{BL}(C_\alpha\text{-}C_\beta)}{I_{BL}(b/y)} \tag{6.7}$$

Under these assumptions the ET efficiency can be evaluated as follows:

$$\varphi_{ET} = \frac{\Delta I(C_\alpha\text{-}C_\beta)}{\Delta I(b/y)} \cdot \frac{I_{BL}(b/y)}{I_{BL}(C_\alpha\text{-}C_\beta)} \tag{6.8}$$

The values, required to calculate the ET efficiency using this formula, can be derived from experimental spectra by fitting the identified Phe band origins with the following Gaussian function:
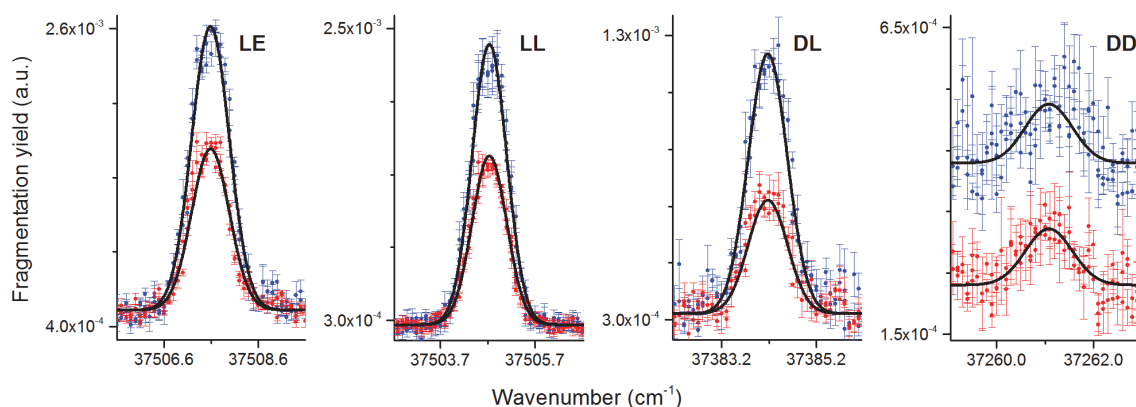
Figure 6.3. High-resolution UV fragmentation spectra of singly protonated LE, LL, DL, and DD (from left to right) peptides in the region of their Phe band origins, corresponding to the prominent thermal, b- and y-fragments (blue dots) and to the Tyr side-chain loss fragment (red dots). The mean photofragmentation signal (dots) and its standard deviation (error bars) obtained by averaging of four measurements. Solid lines represent the best fits with Gaussian functions. For graphical clarity, the UV spectra at the Tyr side-chain loss fragment of LE, LL, DL, and DD are multiplied by 1.369, 0.999, 0.993 and 0.795, respectively. Additionally, the UV spectrum, corresponding to the thermal fragmentation of DD, is offset by $2 \cdot 10^{-4}$ a.u.

Table 6.1. The parameters of the Gaussian functions fitted to the experimental UV fragmentation spectra of singly protonated LE, LL, DL and DD peptides, shown in Figure 6.3, and the respective ET efficiencies. The position ($\tilde{\nu}_0$) and the full width at half maximum (FWHM) of the Phe band origins are given in cm^{-1}.

	LE	LL	DL	DD
$\tilde{\nu}_0$	37507.6	37504.7	37384.2	37261.1
FWHM	0.94	0.86	0.97	1.14
$I_{BL}(b/y)$	$(5.21\pm0.05)\cdot10^{-4}$	$(2.67\pm0.04)\cdot10^{-4}$	$(3.22\pm0.04)\cdot10^{-4}$	$(2.30\pm0.03)\cdot10^{-4}$
$\Delta I(b/y)$	$(2.10\pm0.02)\cdot10^{-3}$	$(2.12\pm0.04)\cdot10^{-3}$	$(0.91\pm0.02)\cdot10^{-3}$	$(0.96\pm0.08)\cdot10^{-4}$
$I_{BL}(C_\alpha\text{-}C_\beta)$	$(3.80\pm0.04)\cdot10^{-4}$	$(2.67\pm0.02)\cdot10^{-4}$	$(3.25\pm0.02)\cdot10^{-4}$	$(2.89\pm0.03)\cdot10^{-4}$
$\Delta I(C_\alpha\text{-}C_\beta)$	$(0.87\pm0.02)\cdot10^{-3}$	$(1.28\pm0.01)\cdot10^{-3}$	$(0.40\pm0.01)\cdot10^{-3}$	$(1.15\pm0.06)\cdot10^{-4}$
φ_{ET}	0.57 ± 0.02	0.60 ± 0.02	0.44 ± 0.02	0.95 ± 0.10

$$I(\tilde{v}) = I_{BL} + \Delta I \cdot e^{-\frac{(\tilde{v}-\tilde{v}_0)^2}{2w^2}}$$ <div align="right">*(6.9)*</div>

Figure 6.3 provides pairs of expanded views of these peaks in the four studied peptides. One spectrum in each pair was generated by integrating the respective high-resolution UV-MS spectrum over the b- and y-fragments, and the other one corresponds to the Tyr side-chain loss. Each spectrum is an average of four measurements. The energy of a UV laser pulse was maintained equal to 1 mJ. For each peptide the two spectra were simultaneously fitted with the Gaussian functions, sharing the $\tilde{v}_0$ and w parameters. The results of the fit as well as the calculated ET efficiency for the four peptides are summarized in Table 6.1.

6.4. Interchromophore Distances

At reasonably large interchromophore distances, the ET process can be described in terms of the ideal dipole approximation, that is, both donor (e.g., Phe) and acceptor (e.g., Tyr) are approximated by point dipoles and hence the electromagnetic interaction between them is governed by a dipole-dipole interaction. Based on this assumption, Theodor Förster developed a theory in the late 1940s, which predicted the ET rate constant to be inversely proportional to the sixth power of the donor-acceptor distance.[322,337,338] In the frame of this theory, named Förster Resonance Energy Transfer (FRET), the ET efficiency can be then expressed as:[31]

$$\varphi_{FRET} = \frac{k_{FRET}}{k_{FRET} + 1/\tau_D} = \frac{1}{1 + \left(R_{DA}/R_0\right)^6}$$ <div align="right">*(6.10)*</div>

where k_{FRET} is the rate constant of the FRET process, τ_D is the lifetime of the donor excited state in the absence of acceptor, R_{DA} is the interchromophore (donor-acceptor) distance and R_0 is the so-called Förster distance, at which $\varphi_{FRET} = 0.5$. R_0 in its turn is determined by the photophysical properties of donor and acceptor and by the refractive index of the medium (n), in which they are embedded:

$$R_0^6 = \frac{9 \cdot (ln\,10) \cdot \kappa^2 \cdot \varphi_D \cdot J}{128 \cdot \pi^5 \cdot n^4 \cdot N_A}$$ <div align="right">*(6.11)*</div>

Here, the orientation factor, κ^2, is the square of the parameter κ, which is defined as:

$$\kappa = \frac{\vec{\mu}_D \cdot \vec{\mu}_A - 3 \cdot (\vec{\mu}_D \cdot \vec{R}_{DA}) \cdot (\vec{\mu}_A \cdot \vec{R}_{DA})}{|\vec{\mu}_D| \cdot |\vec{\mu}_A|} \tag{6.12}$$

where $\vec{\mu}_D$ and $\vec{\mu}_A$ are the transition dipole moments (TDMs) of donor and acceptor, respectively, and $\vec{R}_{DA}$ is a unit vector, pointing from donor to acceptor. Then, φ_D is the quantum yield of the donor fluorescence, that is, $\varphi_D = {}^{\tau_D}/_{\tau_r}$, where τ_r is the donor radiative lifetime. Finally, J is the overlap integral, typically expressed in wavelength (λ) or in wavenumber ($\tilde{v}$) form:

$$J = \int \varepsilon_A(\lambda) \cdot f_D(\lambda) \cdot \lambda^4 d\lambda = \int \frac{\varepsilon_A(\tilde{v}) \cdot f_D(\tilde{v})}{\tilde{v}^4} d\tilde{v} \tag{6.13}$$

between the donor fluorescence spectrum ($f_D(\lambda)$ or $f_D(\tilde{v})$), normalized to unity:

$$\int f_D(\lambda) d\lambda = \int f_D(\tilde{v}) d\tilde{v} = 1 \tag{6.14}$$

and the acceptor absorption spectrum, represented its molar extinction coefficient ($\varepsilon_A(\lambda)$ or $\varepsilon_A(\tilde{v})$).

The relation between the FRET efficiency, determined either from steady-state or time-resolved measurements, and the ratio ${}^{R_{DA}}/_{R_0}$ provides a straightforward way to estimate the interchromophore distance, once the R_0 value is known. In solution the evaluation of the Förster distance is not very challenging, because both the absorption spectrum of the acceptor and the emission spectrum of the donor can be routinely recorded.[339] The fluorescence quantum yield of the donor φ_D can be determined by comparison with standard fluorophores[340] and the refractive index n is often assumed to be near that of a solvent.[339] The largest uncertainty in calculating R_0 often arises from the orientation factor.[341] The κ^2 value can vary between 0 and 4, but typically, one assumes $\kappa^2 = {}^2/_3$, which is a reasonable approximation in the case of fast isotropic motion of both donor and acceptor.[322] However, this is not always the case. For instance, $\kappa^2 = 0.476$ for a very slow rotation of randomly distributed donors and acceptors.[342] And in general, the κ^2 parameter depends strongly on the averaging regime (due to rotational motion) and on the distribution of the TDMs of chromophores in space, so that there is no universal value for κ^2.[341]

In the case of cold, gas-phase ions the situation is even more complicated. Due to the low number density of ions in the cold trap a use of direct absorption and fluorescence

spectroscopy for measuring the absorption and emission spectra of chromophores as well as their fluorescence quantum yields can be rather problematic.[343] Also, because ions are stored in the gas phase and at a very low temperature, the rotational diffusion of chromophores is virtually frozen and the spatial distribution of their TDMs is obviously non-random. In this case, there is no way to approximate κ^2 by a certain value.[342] Here we employed another approach to estimating interchromophore distances. First, note, that Eqs. (6.10) and (6.11) can be rewritten in the following way:

$$\varphi_{FRET} = \frac{1}{1 + \dfrac{R_{DA}^6}{C \cdot \kappa^2}}$$ (6.15)

$$C = \frac{9 \cdot (ln\, 10) \cdot \varphi_D \cdot J}{128 \cdot \pi^5 \cdot n^4 \cdot N_A}$$

In this formulation the FRET efficiency explicitly depends on two parameters, R_{DA} and κ^2, which are determined by the exact 3D structure of ionic species. At the same time, the factor C, which depends only on the photophysical properties of chromophores and on the refractive index n, should not change significantly among peptides of similar size, composition and with the same donor-acceptor pair, so that we can hold it constant for the enkephalins studied here. Then, under this assumption Eq. (6.15) yields:

$$\frac{R_{DA}^6}{\kappa^2} = \frac{R_{ref}^6}{\kappa_{ref}^2} \cdot \frac{\varphi_{ref}}{\varphi_{ET}} \cdot \frac{1 - \varphi_{ET}}{1 - \varphi_{ref}}$$ (6.16)

where φ_{ref}, R_{ref}, κ_{ref}^2 are the ET efficiency, interchromophore distance and orientation factor of a reference peptide. The gas-phase structure of the lowest energy conformer of LE was recently solved and validated by IR-UV spectroscopy.[247] Thus, using the ET efficiency in LE and the R_{DA} and κ^2 values, calculated using its well-defined 3D structure (Figure 6.4), as the reference allows for estimating the $\frac{R_{DA}^6}{\kappa^2}$ ratios in the other three enkephalins.

The interchromophore distance (vector $\vec{R}_{DA}$ in Figure 6.4) is calculated as the distance between the geometric centers of the aromatic rings (the average of the coordinates of six carbon atoms) of Phe and Tyr. The lowest singlet-singlet electronic transition in both Phe and Tyr is of L_b type (Platt's nomenclature[344]), while the higher excited states lie $\approx$ 1 eV above the L_b state[54,250] and are therefore excluded from further consideration. We calculate the κ
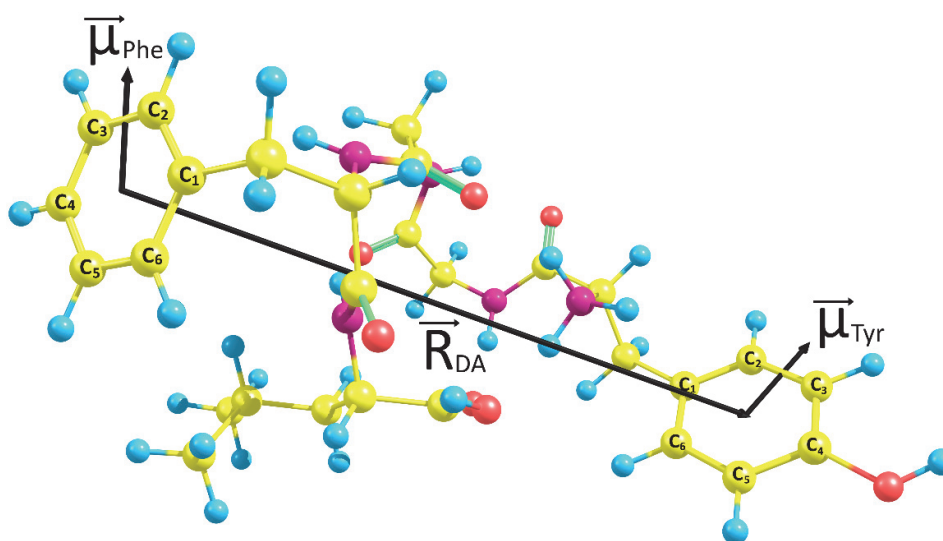
Figure 6.4. 3D structure of singly protonated LE, reconstructed from Cartesian coordinates, kindly provided by N. Burke and Prof. T. S. Zwier.

parameter using Eq. (6.12), where the TDMs of Phe and Tyr are approximated by the vectors ($\vec{\mu}_{Phe}$ and $\vec{\mu}_{Tyr}$ in Figure 6.4) directed from the center of C_5-C_6 bond to the center of C_2-C_3 bond of the respective aromatic rings and the vector is normalized by its length. In the case of LE, shown in Figure 6.4, $R_{DA} = 11.67$ Å and $\kappa^2 = 1.284$ (note that $\kappa^2 \neq {}^2/_3$). Considering these numbers and the determined above ET efficiencies, Eq. (6.16) yields $(1.7 \pm 0.2) \cdot 10^6$, $(3.3 \pm 0.3) \cdot 10^6$, and $(0.0 - 4.1) \cdot 10^5$ Å^6 for $\frac{R_{DA}^6}{\kappa^2}$ in LL, DL, and DD, respectively. We now use these values as constraints for selecting structures of the stereoisomers from the pools of low-energy conformers, calculated at a computationally undemanding molecular mechanics (MM) level of theory.

6.5. Conformational Families of the Stereoisomers of [Ala^2, Leu^5]-enkephalin

A conformational search for the low-energy structures of singly protonated LL, DL, and DD peptide has been performed within the TINKER molecular modeling package[345] using the OPLS-AA/L force field.[346,347] The *scan* program systematically samples the conformational space via a basin hopping method, which combines activation of the located minima along a

number of normal modes with subsequent local geometry optimization. The search starts from the local minimization of an initial structure, which becomes the first local minimum in the search list. The normal mode activation step is performed for each structure from the search list and, whenever a previously unknown minimum is located, it is added to the search list, if its energy falls within a 10 kcal/mol window with respect to the that of the lowest-energy structure. When the program reaches the end of the search list without locating additional minima, it terminates. The convergence criterion for the local geometry minimization was set to 10^{-4} kcal/mol/Å. The chirality of amino-acid residues was maintained during the conformational search.

Figure 6.5 shows the distribution of the interchromophore distances R_{DA} for the resulting pools of the low-energy conformers of LL, DL and DD peptides. The application of the derived constraints for the $\frac{R_{DA}^6}{\kappa^2}$ ratio restricts the entire conformational space of LL (473 structures) and DL (2519 structures) to only 25 and 74 conformers, respectively, thus, removing more than 90 % of the candidate structures. Although only 44 % of the structures have been rejected for DD (1482 of 2630 structures) on the basis of the $\frac{R_{DA}^6}{\kappa^2}$ constraint, this reduces the maximum possible value of R_{DA} from 13.1 to 9.6 Å.
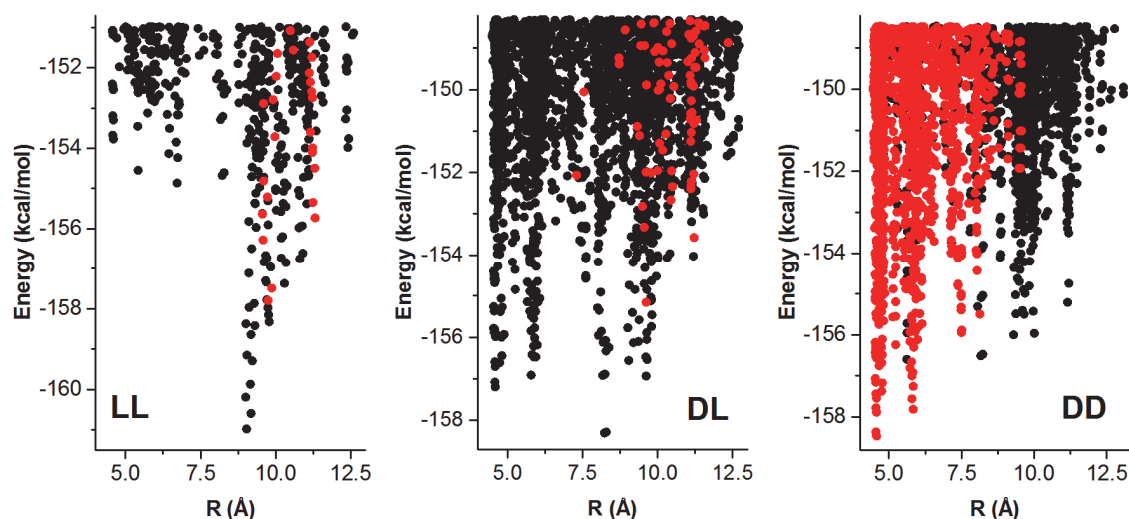

Figure 6.5. Plot of energy versus interchromophore (Phe-Tyr) distance in the conformers of singly protonated LL, DL, and DD (from left to right) peptides, found within a 10 kcal/mol energy window. Each black dot represents a single conformer. Those conformers, for which the ratio $\frac{R_{DA}^6}{\kappa^2}$ satisfies the constraints determined above, are designated by red dots.

The distribution of the remaining LL structures exhibits two families with the average Phe-Tyr distances of 9.8 and 11.2 Å. The distribution of the remaining DL structures is more heterogeneous, however one can clearly distinguish three major groups of conformers with the average Phe-Tyr distances of 9.6, 10.4, and 11.2 Å and two outliers with $R_{DA} \lesssim 7.5$ Å, but with very small orientation factor of about 0.04. Compared with the LL and DL isomers, the DL peptide has a rather diverse conformational distribution, which covers a wide range of interchromophore distances from 4.5 to 9.6 Å. The distribution exhibits a number of low-energy conformational families with R_{DA} of about 4.7, 5.2, 5.8, 7.4 and 8.1 Å. Regardless of the particular number, the interchromophore distance in DD is clearly significantly shorter than in LL and DL. A comparison of the linewidths of the peaks in Figure 6.3 further supports this conclusion. The Phe band origin in DD (FWHM = 1.14 cm^{-1}) is noticeably broader than those of LE (FWHM = 0.94 cm^{-1}), LL (FWHM = 0.86 cm^{-1}) and DL (FHWM = 0.97 cm^{-1}), which may indicate the shorter lifetime of the excited state of Phe in DD. Then, this additional broadening of 0.6-0.7 cm^{-1} corresponds to the S$_1$ lifetime of 7-8 ps. Taking into account that a typical lifetime of Phe chromophore, reported for gas-phase species, is on the order of tens of nanoseconds (or, at least, longer than 1 ns),[55,246,305,348–350] the ET efficiency should not be less than 0.99, which is consistent with the calculated efficiency of 0.95. In general, at such short distances, the Dexter[323] rather than the Förster approximation of ET may become more appropriate for the quantification of R_{DA}.[31,351]

6.6. Validation of the 3D Structure of the Lowest Energy Conformers of the Stereoisomers of [Ala^2, Leu^5]-enkephalin

Although the $\frac{R_{DA}^6}{\kappa^2}$ structural constraint indeed restricts the entire conformational space of the peptides to a few conformational families, the precise 3D structures of their major conformers still require a more stringent validation. To avoid any bias associated with the derived structural constraints, we started the search of the lowest energy conformers from the pools of structures, provided by the MM calculations. Because of the known deficiency of force fields in predicting the energetics of peptides,[352] one cannot rely on the ordering of the calculated structures and choose only a few low-energy structures for a further optimization.

On the other hand, the conformational space of DL and DD peptides is so huge, that the optimization of all the structures directly at the high level of theory would be computationally too expensive. Therefore, we performed quantum chemical calculations in three steps.

First, all the conformers of LL, DL, and DD within a 7.5 kcal/mol energy window (to sample structures from all conformational families) were optimized at the B3LYP/3-21G level with a convergence criterion of 10^{-3} a.u. for the largest component of the nuclear gradient. Thus, after the first step we obtained more or less reliable geometries of 123, 781, and 873 conformers of LL, DL, and DD, respectively, and their relative energies, which were, however, still not accurate enough. At the next step a portion of these structures within a 7.5 kcal/mol energy window was re-optimized at the B3LYP-D/6-31G* level with a convergence criterion of 10^{-4} a.u., which resulted in 70, 44, and 62 final structures of LL, DL, and DD, respectively. This level of theory should be sufficient to adequately describe peptide geometries and predict the relative energies of conformers with an error not exceeding 1 kcal/mol.[353–356] Therefore, only the conformers within a 1 kcal/mol energy window were finally optimized at the B3LYP-D/6-31+G** with a convergence criterion of 10^{-5} a.u. All the quantum chemical calculations were performed using the Firefly software package.[357] For the B3LYP hybrid functional[358,359] the VWN V correlation functional[360] and Grimme's DFT-D version 3 with Becke-Johnson damping dispersion correction[361,362] were used. In the case of the 6-31+G** basis set, the diffuse functions were added only on nitrogen and oxygen atoms. The resulting structures of the lowest energy conformers of LL, DL, and DD peptides are shown in Figure 6.6.

In order to validate the calculated structures, we measured the IR gain spectra[181] of the three stereoisomers. Briefly, the UV laser was tuned to the red edge of the Tyr absorption spectra of the respective peptides, while the IR laser was scanned from 3700 to 2700 cm^{-1} (3 μm region) and from 1800 to 1200 cm^{-1} (6 μm region). The IR pulse preceded the UV pulse by 100-200 ns, so that, if ions do not absorb the IR light at a certain wavenumber, the photofragmentation signal only due to the UV excitation of the cold ions is very low. However, if the ions (any conformer) absorb the IR light at this wavenumber, a significant increase of the photofragmentation signal is observed, because the IR-preheating results in broadening and red-shifting of the UV spectrum. Figure 6.7 shows the comparison of the recorded all-conformer IR spectra of the three peptides with the harmonic frequencies of their lowest-
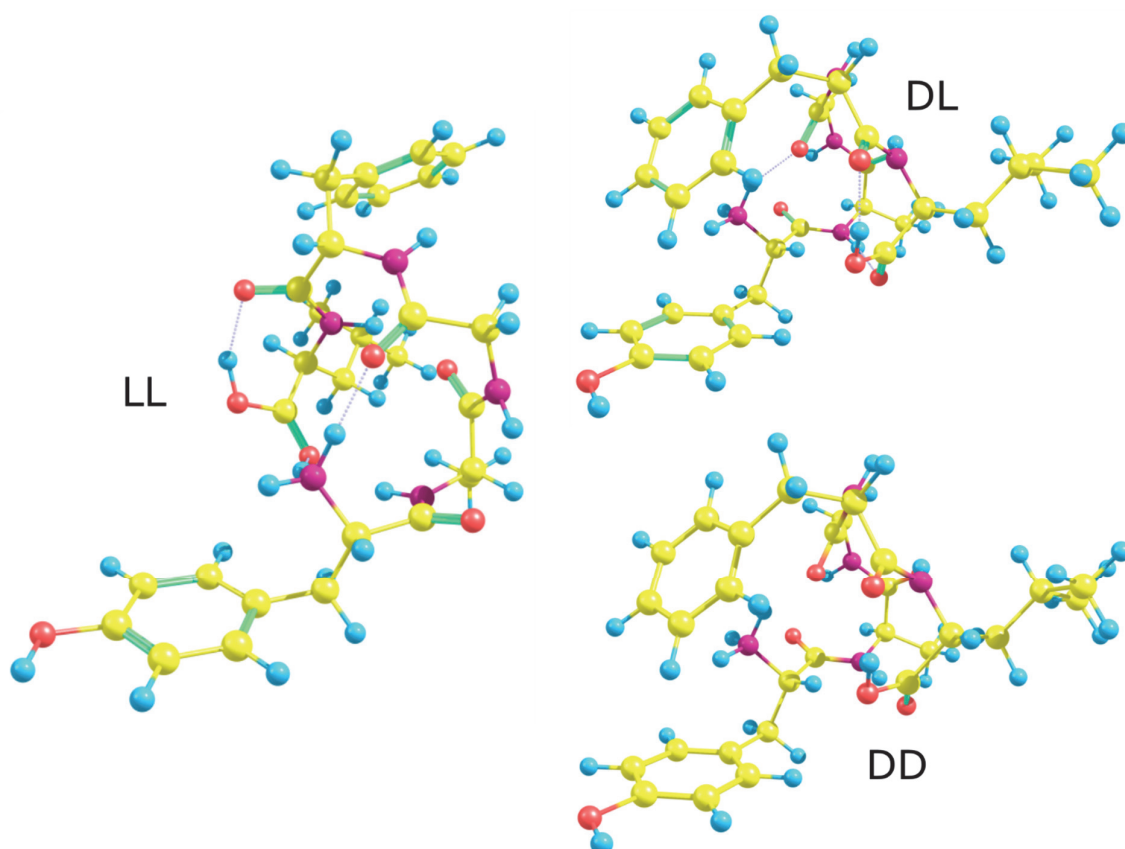
Figure 6.6. 3D structure of the lowest energy conformers of singly protonated LL, DL, and DD peptides.

energy conformers, calculated at the B3LYP-D/6-31+G** level of theory and linearly scaled to account for vibrational anharmonicity.

For the LL peptide the calculated harmonic frequencies match the experimental IR spectrum reasonably well. For instance, in the 3 μm region the peak at ≈ 3648 cm^{-1} can be unambiguously assigned to the free OH stretch of the Tyr side-chain, which is a common feature for all peptides studied here. One sharp peak at 3405 cm^{-1} and two partially resolved peaks at ≈ 3379 cm^{-1} correspond to the weakly H-bonded amide NH stretches of Phe and Ala/Gly residues, respectively, while the amide NH of Leu, involved in a C_{10} hydrogen bond, is shifted to 3321 cm^{-1}. As in the case of LE, the charged NH_3 group at the N-terminus does not seem to be strongly interacting with the Tyr aromatic ring, because the NH stretch, associated with the π-bound hydrogen, appears at 3289 cm^{-1}, only slightly red-shifted with respect to the free NH in protonated Phe and Tyr.[168] In contrast, the remaining two hydrogen of the amino group are strongly H-bonded and the respective NH stretch frequencies are shifted down to

the 2900-3050 cm^{-1} region, where these broad bands overlap with numerous CH stretches. Concerning the OH stretch of the C-terminal carboxylic group, its calculated frequency gets into this region, because of a relatively strong interaction between the hydroxyl group and the amide carbonyl group of Phe residue. On the other hand, by comparison with the experimental and computational data on protonated Phe and Tyr,[168] the peak at 3571 cm^{-1} in the IR spectrum can be assigned to the free carboxylic OH stretch. However, its intensity is too low compared to that of the Tyr OH stretch.[168,170] Therefore, we assign this peak to the free carboxylic OH stretch of a minor conformer. In the major, lowest energy conformer this OH is likely to be H-bonded (as in the case of LE[247]), and hence the respective IR peak is likely to be significantly red-shifted. The assignment in the 6 μm region is less straightforward, because of a considerable vibrational mode mixing, but the calculated frequencies mainly agree with the experimental IR peaks as well. Thus, we may conclude, that only one major conformer of the singly protonated LL peptide is present in the gas phase at the temperature of ≈ 10 K and its 3D structure corresponds to that of the calculated lowest energy conformer.

The IR spectrum of the DD peptide is more complicated. A quick assessment of the number of strong IR transitions suggests the presence of at least two abundant conformers. Because the UV spectrum of this isomer is relatively broad, it was not possible to measure its conformer-selective IR spectra using the IR-UV depletion scheme.[181] Nonetheless, the calculated harmonic frequencies of the lowest energy conformer match some of the IR peaks, which implies that this conformer is likely to be one of those experimentally observed. However, the harmonic frequency analysis for the higher-energy conformers and the measurements of the conformer-selective spectra, for instance, by employing the IR-IR-UV scheme[60] are required for a further conformational assignment. The worst agreement between the experimental and calculated IR transition was obtained in the case of the DL peptide. Although the peptide is likely to have more than one abundant conformer (note a number of small peaks in the 3420-3500 cm^{-1} region), the calculated frequencies match neither the intense peaks, tentatively ascribed to a major conformer, nor the less intense IR transitions of a minor one. Of course, we cannot fully disregard a possible inaccuracy in computing the harmonic frequencies. However, the experimentally observed conformer could also have been erroneously discarded during the multistep refinement of structures from the conformational pool.
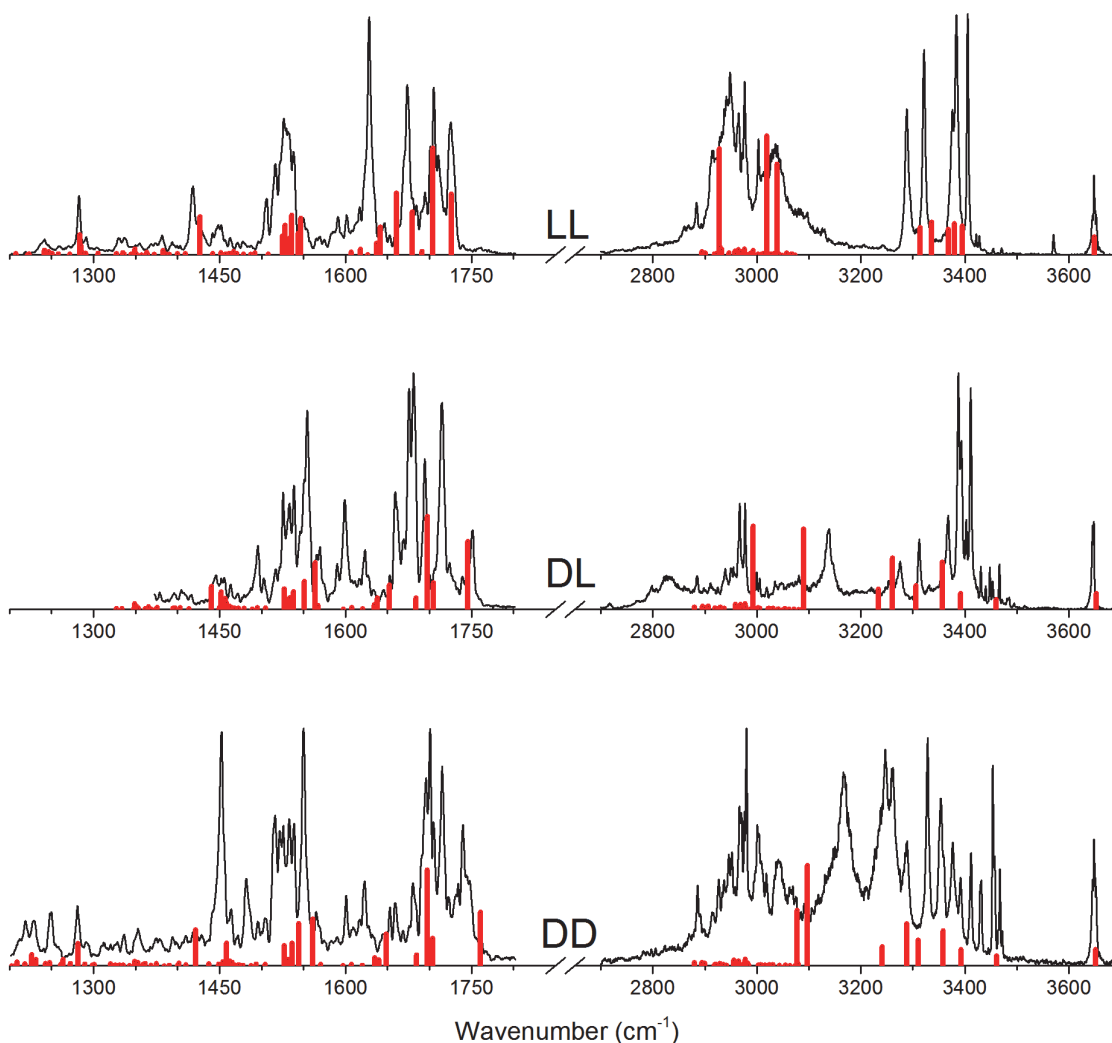
Figure 6.7. Comparison of the experimental all-conformer IR spectra (black lines) of singly protonated LL, DL and DD (from top to bottom) peptides and the scaled harmonic frequencies (red sticks) of their lowest energy conformers, calculated at the B3LYP-D/6-31+G** level of theory. Harmonic frequencies in the 3 and 6 μm regions are scaled by 0.9535 and 0.98, respectively.

6.7. Discussion

To sum up, based on the recorded the 2D UV-MS spectra, we determined the efficiency of ET from Phe to Tyr and estimated the respective interchromophore distances in four opioid peptides. We have also performed an extensive conformational search and, using the derived

constraints for the $\frac{R_{DA}^6}{\kappa^2}$ ratio, reduced the entire conformational space of these peptides to a number of conformational families. Finally, we have calculated the 3D structure of the lowest energy conformers and compared their harmonic frequencies against the experimentally measured all-conformer IR spectra of the peptides.

The most conclusive results were obtained for the LL peptide. The calculated structure of its lowest energy conformer (Figure 6.6) turned out to be very similar to that of the LE peptide (Figure 6.4). This explains the fact that their UV (Figure 6.2(a)) and IR spectra (Figure 6.7 and Figure 4 in ref. [247]) are almost identical and that they demonstrate almost the same ET efficiency (Table 6.1). Furthermore, this conformer is the one with the lowest energy in the conformational family with R_{DA} of 11.2 Å, which was revealed by the constrained conformational search (Figure 6.5).

The DD peptide is represented by several conformers, which have quite a different 3D structure. This follows from a comparison of the UV and IR spectra. For instance, significant red shits (≈ 250 cm^{-1}) of both Tyr and Phe band origins (Figure 6.2(a)) suggest substantial alteration of the local environment of the two chromophores and can be associated with either Tyr-Phe interaction or π-proton interactions, or both. The IR spectra of DD (Figure 6.7), both in 3 and 6 μm regions, provide a complementary hint on its structural features. The whole pattern of NH and CO stretches is shifted to higher wavenumbers compared to that of LL (or LE), which indicates a weakening of the hydrogen bond network. The calculated structure of the lowest energy conformer (Figure 6.6) indeed confirms these hypotheses. While the β-hairpin-like structures of LL, LE and even C-terminally methyl esterified LE are stabilized by two hydrogen bonds, which the protonated N-terminus forms with the carbonyl groups of Gly^3 and C-terminus,[247,363] one of these interactions, namely, the hydrogen bond with the C-terminus disappears in DD. This leads to a distortion of the β-turn, formed around Gly^3-Phe^4, and two C$_7$ hydrogen bonds between NH of Gly^3 and CO of Tyr^1, and between carboxylic OH and CO of Phe^4, respectively. As a result, a global shift of harmonic frequencies to higher wavenumber is observed. Also, because the hydrogen bonds in DD are weakened, the peptide becomes conformationally more flexible, which is clearly reflected in its IR spectrum. Moreover, the broadening of the UV spectra of DD is likely to arise from this conformational heterogeneity, because single vibronic transition in the region of the Tyr band origin are as narrow as in the spectrum of LE and LL (about 1.5 cm^{-1}). Finally, the calculated conformers

belongs to the conformational family with R_{DA} of 4.7 Å, which is fully consistent with our ET efficiency measurements.

Concerning the DL peptide, the results are apparently contradictory. On one hand, the calculated conformational pool (Figure 6.5) and the lowest energy conformer structure (Figure 6.6) resemble those of the DD isomer. On the other hand, all the experimental data (UV and IR spectra, ET efficiency) clearly indicate a difference in the structure of these two isomers, DL and DD. For instance, the Tyr band origin in DL appears at 35253.6 cm^{-1}, about 150 cm^{-1} red-shifted with respect to that in DD, while the Phe band origin in DL is, on the contrary, blue-shifted approximately by 120 cm^{-1} (Figure 6.2(a)). It is likely that in the DL isomer the π-proton interaction of the Tyr side chain is rather strong and the Phe chromophore becomes more remote from the positive charge. This would explain a large interchromophore distance (≈ 10.4 Å), derived from the ET efficiency measurements. Thus, the next step would be to perform the geometry optimization with subsequent harmonic frequency analysis of those conformers of DL, which satisfy the constraint imposed on the $\frac{R_{DA}^6}{\kappa^2}$ ratio, because the final validation requires calculation of their accurate IR spectra.

6.8. Conclusions

Among the four peptides examined herein, the DD isomer, called DADLE, is by far the most potent and selective δ-opioid receptor agonist.[319,330,331] The question remains as to whether the highest efficiency of DADLE correlates with its Phe-Tyr spacing. The available X-ray diffraction data suggest 5 to 13.9 Å (depending on the conformer) and 9.3 Å for this spacing in crystallized leucine enkephalin and DADLE, respectively.[320] These values of R_{DA} exhibit no apparent correlation with the drug efficiencies. Similar uncertainty remains in correlating structures and pharmacological activity for many other opioid peptides.[290,321] In the gas phase, the short Phe-Tyr spacing constrains the structures of DADLE to the double bend family, whereas LL and LE exhibit a similar single bend structural motif. Therefore, there is a qualitative difference between the gas-phase structures of these drugs, which differ significantly in their efficiencies.

Although the interchromophore distance is not the only important pharmacophoric parameter, the peptides in vivo are, likely, zwitterions and the final validation of the calculated structures must be completed by a method such as conformer-selective IR spectroscopy, the question arises as to whether this distance, when determined in the gas phase, is a better pharmacological metric for these receptor ligands than the same distance in crystal structures. Certainly, additional gas-phase data are required before drawing a firm conclusion. Herein we have demonstrated how the Phe-Tyr distance can be estimated in the gas-phase enkephalins using FRET and MM calculations. Most of the opioid peptides contain the Phe-Tyr pair, allowing for the extension of our study to other ligands of this family, as well as to other peptides with these chromophores.

Conclusions

In this thesis, we have described the construction of a novel hybrid instrument that combines a cold ion spectrometer and an Orbitrap-based mass spectrometer. The former allows carrying out high-resolution UV photodissociation spectroscopy of ions cooled to the vibrational temperature of about 10 K, and the latter enables recording the entire high-resolution mass spectrum of both parent and all photofragment ions in a single scan at each UV wavenumber. Vibrational resolution, achievable by cryogenic cooling of ions, makes these fragmentation patterns sharply wavenumber-dependent and characteristic of the 3D structure of the ions on a fundamental level. Thus, this fusion exploits the advantages of the two complementary techniques and produces unique and highly reproducible 2D UV-MS fingerprints (fragmentation yield vs UV wavenumber and mass-to-charge ratio) of ionic species.

One of the obvious applications of such fingerprints is to use them for identification of very similar species that cannot be distinguished even by high-resolution mass spectrometry. Here we have demonstrated that, once stored as libraries, the UV-MS fingerprints can be used in further identifications of isomers in mixtures by employing numerical algorithms of matrix analysis. Particularly, we have successfully applied the library-based UV-MS approach for quantitative identification of positional isomers of a phosphorylated peptide, diastereomers of an opioid peptide, and diastereomers of a drug molecule. These are challenging cases not only for mass spectrometry but also for liquid chromatography and ion mobility spectrometry, which are typically used for separating structurally similar species prior to a mass spectrometric analysis. We have also demonstrated the potential of the UV-MS approach for being complemented by liquid chromatography (or, in principle, ion mobility spectrometry) for online identifications. Besides the library-based analysis of 2D fingerprints, we have proposed a more sophisticated library-free approach, which in certain cases is capable of

estimating the number of components present in a mixture and deriving their respective UV absorption and photofragmentation mass spectra directly from the UV-MS spectrum of the mixture.

Further, we employed 2D UV-MS spectroscopy to identify conformers of gas-phase biomolecular ions and to recover their individual UV absorption and photofragmentation mass spectra. It is the phenomenon of non-statistical fragmentation that makes UV-MS fingerprints not only isomer- but also conformer-specific and hence enables using the library-free approach. We have validated it using a benchmark dipeptide, Tyr-Ala, for which, in addition to two families of the previously known conformers, their isotopologues were unambiguously identified. The application of our approach to a decapeptide, gramicidin S, allowed us to provide a structural constraint that restricts the most likely structure of its conformer C, for which the 3D geometry has not yet been validated, to one of the previously calculated low-energy structures. We have also demonstrated that our approach can be equally well applied for identification of conformers of non-peptidic biomolecular ions containing a UV-absorbing chromophore. A careful analysis of the UV-MS spectrum of a small drug molecule, homatropine, revealed the presence of one major and one minor conformers, which has been fully confirmed by the experimentally measured conformer-selective IR spectra. Finally, we proposed a model that establishes a correlation between the relative abundance of the side-chain loss fragments, resulting from prompt dissociation of aromatic amino-acid residues upon their UV excitation, and the extent of interaction of the respective chromophore with its local environment.

In the last part of this thesis, we have demonstrated a use of 2D UV-MS spectroscopy for estimating the interchromophore (Phe-Tyr) distances in four opioid peptides isolated in the gas phase on the basis of measurements of the conformer-specific efficiencies of resonance energy transfer and computationally inexpensive molecular mechanics calculations. Additionally, we have performed an extensive conformational search and calculated the 3D geometries of the lowest energy conformers of these peptides. An analysis of all the experimental data and the results of high-level quantum chemical calculations clearly indicates that the calculated structure of one of the peptides, namely, [D-Ala[2], L-Leu[5]]-enkephalin is not the one observed experimentally, while the 3D structures of the others are consistent with the estimated interchromophore distances, which validates the

proposed approach. Moreover, in contrast to the condensed-phase structures, the estimated distances appeared to differ substantially in the peptides with different pharmacological efficiencies. This suggests that gas-phase structures might be a better pharmacophoric metric for ligand peptides, which undergo significant conformational changes upon interaction with the hydrophobic binding pockets of receptors.

The exact limitations of the UV-MS approaches described in this thesis report are yet to be assessed. However, the hardware and the operational conditions utilized for the approaches are not specific to molecules and should be suitable for studying any small to midsize UV-absorbing species compatible with mass spectrometric analysis, such as peptides, containing aromatic residues, drugs, and metabolites, in a broad range of applications. Furthermore, other spectroscopic techniques, such as infrared multiple-photon dissociation (IRMPD) or IR-UV double-resonance spectroscopy, can be equally well implemented on the novel instrument.

Bibliography

(1)   PDB Current Holdings Breakdown http://www.rcsb.org/pdb/statistics/holdings.do.

(2)   KENDREW, J. C.; BODO, G.; DINTZIS, H. M.; PARRISH, R. G.; WYCKOFF, H.; PHILLIPS, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666.

(3)   PERUTZ, M. F.; ROSSMANN, M. G.; CULLIS, A. F.; MUIRHEAD, H.; WILL, G.; NORTH, A. C. T. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185* (4711), 416–422.

(4)   FRANKLIN, R. E.; GOSLING, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, *171* (4356), 740–741.

(5)   WATSON, J. D.; CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738.

(6)   Bränd´en, C.-I.; Alwyn Jones, T. Between objectivity and subjectivity. *Nature* **1990**, *343* (6260), 687–689.

(7)   Egli, M. Diffraction Techniques in Structural Biology. In *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010; pp 7.13.1–7.13.35.

(8)   Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **2014**, *83* (1), 553–584.

(9)   Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. NMR analysis of a 900K GroEL GroES complex. *Nature* **2002**, *418* (6894), 207–211.

(10)   Sprangers, R.; Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **2007**, *445* (7128), 618–622.

(11)   Rule, G. S.; Hitchens, T. K. *Fundamentals of Protein NMR Spectroscopy*; Focus on Structural Biology; Springer-Verlag: Berlin/Heidelberg, 2006; Vol. 5.

(12)   Felli, I. C.; Pierattelli, R. *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*; Felli, I. C., Pierattelli, R., Eds.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, 2015; Vol. 870.

(13)   Mehmood, S.; Allison, T. M.; Robinson, C. V. Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annu. Rev. Phys. Chem.* **2015**, *66* (1), 453–474.

(14)   Zaia, J. Mass spectrometry of oligosaccharides. *Mass Spectrom. Rev.* **2004**, *23* (3), 161–227.

(15)   Thevis, M.; Schänzer, W. Mass spectrometry in sports drug testing: Structure characterization and analytical assays. *Mass Spectrom. Rev.* **2007**, *26* (1), 79–107.

(16)   Wang, Y.; Liu, S.; Hu, Y.; Li, P.; Wan, J.-B. Current state of the art of mass spectrometry-based metabolomics studies – a review focusing on wide coverage, high

throughput and easy identification. *RSC Adv.* **2015**, *5* (96), 78728–78737.

(17)   Brügger, B. Lipidomics: Analysis of the Lipid Composition of Cells and Subcellular Organelles by Electrospray Ionization Mass Spectrometry. *Annu. Rev. Biochem.* **2014**, *83* (1), 79–98.

(18)   Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* **2009**, *11* (1), 49–79.

(19)   Shao, W.; Lam, H. Tandem mass spectral libraries of peptides and their roles in proteomics research. *Mass Spectrom. Rev.* **2016**.

(20)   Hughes, C.; Ma, B.; Lajoie, G. A. De Novo Sequencing Methods in Proteomics. In *Methods in molecular biology (Clifton, N.J.)*; 2010; Vol. 604, pp 105–121.

(21)   Kaltashov, I. A.; Eyles, S. J. *Mass spectrometry in structural biology and biophysics: architecture, dynamics, and interaction of biomolecules*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2012.

(22)   Gross, J. H. *Mass Spectrometry. A Textbook*, 2nd ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011.

(23)   Picotti, P.; Aebersold, R. Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **2012**, *9* (6), 555–566.

(24)   Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85* (11), 5288–5296.

(25)   Karoui, R.; Downey, G.; Blecker, C. Mid-Infrared Spectroscopy Coupled with Chemometrics: A Tool for the Analysis of Intact Food Systems and the Exploration of Their Molecular Structure–Quality Relationships – A Review. *Chem. Rev.* **2010**, *110* (10), 6144–6168.

(26)   Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta - Bioenerg.* **2007**, *1767* (9), 1073–1101.

(27)   Barth, A. The infrared absorption of amino acid side chains. *Prog. Biophys. Mol. Biol.* **2000**, *74* (3-5), 141–173.

(28)   Barth, A.; Zscherp, C. What vibrations tell about proteins. *Q. Rev. Biophys.* **2002**, *35* (4), S0033583502003815.

(29)   Remorino, A.; Hochstrasser, R. M. Three-Dimensional Structures by Two-Dimensional Vibrational Spectroscopy. *Acc. Chem. Res.* **2012**, *45* (11), 1896–1905.

(30)   Beechem, J. M.; Brand, L. Time-Resolved Fluorescence of Proteins. *Annu. Rev. Biochem.* **1985**, *54* (1), 43–71.

(31)   *FRET - Förster Resonance Energy Transfer*; Medintz, I., Hildebrandt, N., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2013.

(32)   Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **2007**, *1* (6), 2876–2890.

(33)   Kelly, S. M.; Price, N. C. The use of circular dichroism in the investigation of protein

structure and function. *Curr. Protein Pept. Sci.* **2000**, *1* (4), 349–384.

(34)   Rizzo, T. R.; Park, Y. D.; Levy, D. H. A molecular beam of tryptophan. *J. Am. Chem. Soc.* **1985**, *107* (1), 277–278.

(35)   Park, Y. D.; Rizzo, T. R.; Peteanu, L. A.; Levy, D. H. Electronic spectroscopy of tryptophan analogs in supersonic jets: 3-Indole acetic acid, 3-indole propionic acid, tryptamine, and N-acetyl tryptophan ethyl ester. *J. Chem. Phys.* **1986**, *84* (12), 6539.

(36)   Colarusso, P.; Zhang, K.; Guo, B.; Bernath, P. F. The infrared spectra of uracil, thymine, and adenine in the gas phase. *Chem. Phys. Lett.* **1997**, *269* (1-2), 39–48.

(37)   Lin, J.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L. K.; LeBreton, P. R. Ultraviolet photoelectron studies of the ground-state electronic structure and gas-phase tautomerism of purine and adenine. *J. Am. Chem. Soc.* **1980**, *102* (14), 4627–4631.

(38)   Hutzler, N. R.; Lu, H.-I.; Doyle, J. M. The Buffer Gas Beam: An Intense, Cold, and Slow Source for Atoms and Molecules. *Chem. Rev.* **2012**, *112* (9), 4803–4827.

(39)   Levy, D. H. Laser Spectroscopy of Cold Gas-Phase Molecules. *Annu. Rev. Phys. Chem.* **1980**, *31* (1), 197–225.

(40)   Meijer, G.; de Vries, M. S.; Hunziker, H. E.; Wendt, H. R. Laser desorption jet-cooling of organic molecules. *Appl. Phys. B* **1990**, *51* (6), 395–403.

(41)   Cable, J. R.; Tubergen, M. J.; Levy, D. H. The electronic spectra of small peptides in the gas phase. *Faraday Discuss. Chem. Soc.* **1988**, *86* (0), 143.

(42)   Cable, J. R.; Tubergen, M. J.; Levy, D. H. Laser desorption molecular beam spectroscopy: the electronic spectra of tryptophan peptides in the gas phase. *J. Am. Chem. Soc.* **1987**, *109* (20), 6198–6199.

(43)   Cable, J. R.; Tubergen, M. J.; Levy, D. H. Electronic spectroscopy of small tryptophan peptides in supersonic molecular beams. *J. Am. Chem. Soc.* **1988**, *110* (22), 7349–7355.

(44)   Nir, E.; Kleinermanns, K.; de Vries, M. S. Pairing of isolated nucleic-acid bases in the absence of the DNA backbone. *Nature* **2000**, *408* (6815), 949–951.

(45)   Abo-Riziq, A.; Grace, L.; Nir, E.; Kabelac, M.; Hobza, P.; de Vries, M. S. Photochemical selectivity in guanine-cytosine base-pair structures. *Proc. Natl. Acad. Sci.* **2005**, *102* (1), 20–23.

(46)   Abo-Riziq, A.; Bushnell, J. E.; Crews, B.; Callahan, M.; Grace, L.; de Vries, M. S. Gas phase spectroscopy of the pentapeptide FDASV. *Chem. Phys. Lett.* **2006**, *431* (4-6), 227–230.

(47)   Bakker, J. M.; Plützer, C.; Hünig, I.; Häber, T.; Compagnon, I.; von Helden, G.; Meijer, G.; Kleinermanns, K. Folding Structures of Isolated Peptides as Revealed by Gas-Phase Mid-Infrared Spectroscopy. *ChemPhysChem* **2005**, *6* (1), 120–128.

(48)   Rijs, A. M.; Ohanessian, G.; Oomens, J.; Meijer, G.; von Helden, G.; Compagnon, I. Internal Proton Transfer Leading to Stable Zwitterionic Structures in a Neutral Isolated Peptide. *Angew. Chemie Int. Ed.* **2010**, *49* (13), 2332–2335.

(49) Rijs, A. M.; Kabeláč, M.; Abo-Riziq, A.; Hobza, P.; de Vries, M. S. Isolated Gramicidin Peptides Probed by IR Spectroscopy. *ChemPhysChem* **2011**, *12* (10), 1816–1821.

(50) de Vries, M. S.; Hobza, P. Gas-phase spectroscopy of biomolecular building blocks. *Annu. Rev. Phys. Chem.* **2007**, *58*, 585–612.

(51) Rizzo, T. R.; Park, Y. D.; Peteanu, L.; Levy, D. H. Electronic spectrum of the amino acid tryptophan cooled in a supersonic molecular beam. *J. Chem. Phys.* **1985**, *83* (9), 4819.

(52) Rizzo, T. R.; Park, Y. D.; Peteanu, L. a.; Levy, D. H. The electronic spectrum of the amino acid tryptophan in the gas phase. *J. Chem. Phys.* **1986**, *84* (5), 2534.

(53) Rizzo, T. R.; Park, Y. D.; Levy, D. H. Dispersed fluorescence of jet-cooled tryptophan: Excited state conformers and intramolecular exciplex formation. *J. Chem. Phys.* **1986**, *85* (12), 6945.

(54) Martinez, S. J.; Alfano, J. C.; Levy, D. H. The electronic spectroscopy of the amino acids tyrosine and phenylalanine in a supersonic jet. *J. Mol. Spectrosc.* **1992**, *156* (2), 421–430.

(55) Martinez, S. J.; Alfano, J. C.; Levy, D. H. The Electronic Spectroscopy of Tyrosine and Phenylalanine Analogs in a Supersonic Jet: Basic Analogs. *J. Mol. Spectrosc.* **1993**, *158* (1), 82–92.

(56) Philips, L. A.; Levy, D. H. Rotationally resolved electronic spectroscopy of tryptamine conformers in a supersonic jet. *J. Chem. Phys.* **1988**, *89* (1), 85.

(57) Lipert, R. J.; Colson, S. D. Persistent spectral hole burning of molecular clusters in a supersonic jet. *J. Phys. Chem.* **1989**, *93* (10), 3894–3896.

(58) Page, R. H.; Shen, Y. R.; Lee, Y. T. Infrared–ultraviolet double resonance studies of benzene molecules in a supersonic beam. *J. Chem. Phys.* **1988**, *88* (9), 5362.

(59) Pribble, R. N.; Zwier, T. S. Size-Specific Infrared Spectra of Benzene-(H2O)n Clusters (n = 1 through 7): Evidence for Noncyclic (H2O)n Structures. *Science (80-.).* **1994**, *265* (5168), 75–79.

(60) Shubert, V. A.; Zwier, T. S. IR-IR-UV hole-burning: conformation specific IR spectra in the face of UV spectral overlap. *J. Phys. Chem. A* **2007**, *111* (51), 13283–13286.

(61) Hu, Y.; Guan, J.; Bernstein, E. R. Mass-selected IR-VUV (118 nm) spectroscopic studies of radicals, aliphatic molecules, and their clusters. *Mass Spectrom. Rev.* **2013**, *32* (6), 484–501.

(62) Hu, Y.; Bernstein, E. R. Vibrational and photoionization spectroscopy of biomolecules: Aliphatic amino acid structures. *J. Chem. Phys.* **2008**, *128* (16), 164311.

(63) Hu, Y.; Bernstein, E. R. Vibrational and Photoionization Spectroscopy of Neutral Valine Clusters. *J. Phys. Chem. A* **2009**, *113* (30), 8454–8461.

(64) Nosenko, Y.; Kunitski, M.; Stark, T.; Göbel, M.; Tarakeshwar, P.; Brutschy, B. Vibrational signatures of Watson–Crick base pairing in adenine–thymine mimics. *Phys. Chem. Chem. Phys.* **2013**, *15* (27), 11520.

(65) Nosenko, Y.; Kunitski, M.; Thummel, R. P.; Kyrychenko, A.; Herbich, J.; Waluk, J.; Riehn,

C.; Brutschy, B. Detection and Structural Characterization of Clusters with Ultrashort-Lived Electronically Excited States: IR Absorption Detected by Femtosecond Multiphoton Ionization. *J. Am. Chem. Soc.* **2006**, *128* (31), 10000–10001.

(66) Jaeqx, S.; Oomens, J.; Cimas, A.; Gaigeot, M.-P.; Rijs, A. M. Gas-Phase Peptide Structures Unraveled by Far-IR Spectroscopy: Combining IR-UV Ion-Dip Experiments with Born-Oppenheimer Molecular Dynamics Simulations. *Angew. Chemie Int. Ed.* **2014**, *53* (14), 3663–3666.

(67) Snoek, L. .; Robertson, E. .; Kroemer, R. .; Simons, J. . Conformational landscapes in amino acids: infrared and ultraviolet ion-dip spectroscopy of phenylalanine in the gas phase. *Chem. Phys. Lett.* **2000**, *321* (1-2), 49–56.

(68) Gerhards, M.; Unterberg, C.; Gerlach, A. Structure of a β-sheet model system in the gas phase: Analysis of the C☐O stretching vibrations. *Phys. Chem. Chem. Phys.* **2002**, *4* (22), 5563–5565.

(69) Stamm, A.; Bernhard, D.; Gerhards, M. Structural investigations on a linear isolated depsipeptide: the importance of dispersion interactions. *Phys. Chem. Chem. Phys.* **2016**, *18* (22), 15327–15336.

(70) Robertson, E. G.; Simons, J. P. Getting into shape: Conformational and supramolecular landscapes in small biomolecules and their hydrated clusters. *Phys. Chem. Chem. Phys.* **2001**, *3* (1), 1–18.

(71) Snoek, L. C.; Kroemer, R. T.; Hockridge, M. R.; Simons, J. P. Conformational landscapes of aromatic amino acids in the gas phase: Infrared and ultraviolet ion dip spectroscopy of tryptophan. *Phys. Chem. Chem. Phys.* **2001**, *3* (10), 1819–1826.

(72) Buchanan, E. G.; James, W. H.; Choi, S. H.; Guo, L.; Gellman, S. H.; Müller, C. W.; Zwier, T. S. Single-conformation infrared spectra of model peptides in the amide I and amide II regions: Experiment-based determination of local mode frequencies and inter-mode coupling. *J. Chem. Phys.* **2012**, *137* (9), 094301.

(73) Inokuchi, Y.; Kobayashi, Y.; Ito, T.; Ebata, T. Conformation of <scp>l</scp> -Tyrosine Studied by Fluorescence-Detected UV−UV and IR−UV Double-Resonance Spectroscopy. *J. Phys. Chem. A* **2007**, *111* (17), 3209–3215.

(74) Brenner, V.; Piuzzi, F.; Dimicoli, I.; Tardivel, B.; Mons, M. Chirality-Controlled Formation of β-Turn Secondary Structures in Short Peptide Chains: Gas-Phase Experiment versus Quantum Chemistry. *Angew. Chemie Int. Ed.* **2007**, *46* (14), 2463–2466.

(75) Rijs, A. M.; Ohanessian, G.; Oomens, J.; Meijer, G.; von Helden, G.; Compagnon, I. Internal Proton Transfer Leading to Stable Zwitterionic Structures in a Neutral Isolated Peptide. *Angew. Chemie Int. Ed.* **2010**, *49* (13), 2332–2335.

(76) Chin, W.; Compagnon, I.; Dognon, J.-P.; Canuel, C.; Piuzzi, F.; Dimicoli, I.; von Helden, G.; Meijer, G.; Mons, M. Spectroscopic Evidence for Gas-Phase Formation of Successive β-Turns in a Three-Residue Peptide Chain. *J. Am. Chem. Soc.* **2005**, *127* (5), 1388–1389.

(77) Chin, W.; Piuzzi, F.; Dognon, J.-P.; Dimicoli, I.; Tardivel, B.; Mons, M. Gas Phase

Formation of a 3 10 -Helix in a Three-Residue Peptide Chain: Role of Side Chain-Backbone Interactions as Evidenced by IR–UV Double Resonance Experiments. *J. Am. Chem. Soc.* **2005**, *127* (34), 11900–11901.

(78)   Dean, J. C.; Buchanan, E. G.; Zwier, T. S. Mixed 14/16 Helices in the Gas Phase: Conformation-Specific Spectroscopy of Z-(Gly) n , n = 1, 3, 5. *J. Am. Chem. Soc.* **2012**, *134* (41), 17186–17201.

(79)   James, W. H.; Baquero, E. E.; Shubert, V. A.; Choi, S. H.; Gellman, S. H.; Zwier, T. S. Single-Conformation and Diastereomer Specific Ultraviolet and Infrared Spectroscopy of Model Synthetic Foldamers: α/β-Peptides. *J. Am. Chem. Soc.* **2009**, *131* (18), 6574–6590.

(80)   Vaden, T. D.; Gowers, S. A. N.; de Boer, T. S. J. A.; Steill, J. D.; Oomens, J.; Snoek, L. C. Conformational Preferences of an Amyloidogenic Peptide: IR Spectroscopy of Ac-VQIVYK-NHMe. *J. Am. Chem. Soc.* **2008**, *130* (44), 14640–14650.

(81)   Rijs, A. M.; Kabeláč, M.; Abo-Riziq, A.; Hobza, P.; de Vries, M. S. Isolated Gramicidin Peptides Probed by IR Spectroscopy. *ChemPhysChem* **2011**, *12* (10), 1816–1821.

(82)   Compagnon, I.; Oomens, J.; Bakker, J.; Meijer, G.; von Helden, G. Vibrational spectroscopy of a non-aromatic amino acid-based model peptide: identification of the γ-turn motif of the peptide backbone. *Phys. Chem. Chem. Phys.* **2005**, *7* (1), 13–15.

(83)   Gloaguen, E.; Pagliarulo, F.; Brenner, V.; Chin, W.; Piuzzi, F.; Tardivel, B.; Mons, M. Intramolecular recognition in a jet-cooled short peptide chain: γ-turn helicity probed by a neighbouring residue. *Phys. Chem. Chem. Phys.* **2007**, *9* (32), 4491.

(84)   Fricke, H.; Schäfer, G.; Schrader, T.; Gerhards, M. Secondary structure binding motifs of the jet cooled tetrapeptide model Ac–Leu–Val–Tyr(Me)–NHMe. *Phys. Chem. Chem. Phys.* **2007**, *9* (32), 4592.

(85)   Gloaguen, E.; Loquais, Y.; Thomas, J. A.; Pratt, D. W.; Mons, M. Spontaneous Formation of Hydrophobic Domains in Isolated Peptides. *J. Phys. Chem. B* **2013**, *117* (17), 4945–4955.

(86)   Simons, J. P.; Cristina Stanca-Kaposta, E.; Cocinero, E. J.; Liu, B.; Davis, B. G.; Gamblin, D. P.; Kroemer, R. T. Probing the glycosidic linkage: secondary structures in the gas phase. *Phys. Scr.* **2008**, *78* (5), 058124.

(87)   Barry, C. S.; Cocinero, E. J.; Çarçabal, P.; Gamblin, D. P.; Stanca-Kaposta, E. C.; Remmert, S. M.; Fernández-Alonso, M. C.; Rudić, S.; Simons, J. P.; Davis, B. G. "Naked" and Hydrated Conformers of the Conserved Core Pentasaccharide of N-linked Glycoproteins and Its Building Blocks. *J. Am. Chem. Soc.* **2013**, *135* (45), 16895–16903.

(88)   Cocinero, E. J.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. The Building Blocks of Cellulose: The Intrinsic Conformational Structures of Cellobiose, Its Epimer, Lactose, and Their Singly Hydrated Complexes. *J. Am. Chem. Soc.* **2009**, *131* (31), 11117–11123.

(89)   Jockusch, R. A.; Kroemer, R. T.; Talbot, F. O.; Snoek, L. C.; Çarçabal, P.; Simons, J. P.; Havenith, M.; Bakker, J. M.; Compagnon, I.; Meijer, G.; et al. Probing the Glycosidic Linkage: UV and IR Ion-Dip Spectroscopy of a Lactoside. *J. Am. Chem. Soc.* **2004**, *126* (18), 5709–5714.

(90) Jockusch, R. A.; Talbot, F. O.; Simons, J. P. Sugars in the gas phase Part 2: the spectroscopy and structure of jet-cooled phenyl [small beta]-D-galactopyranoside. *Phys. Chem. Chem. Phys.* **2003**, *5* (8), 1502–1507.

(91) Talbot, F. O.; Simons, J. P. Sugars in the gas phase: the spectroscopy and structure of jet-cooled phenyl β-D-glucopyranoside. *Phys. Chem. Chem. Phys.* **2002**, *4* (15), 3562–3565.

(92) Çarçabal, P.; Jockusch, R. A.; Hünig, I.; Snoek, L. C.; Kroemer, R. T.; Davis, B. G.; Gamblin, D. P.; Compagnon, I.; Oomens, J.; Simons, J. P. Hydrogen Bonding and Cooperativity in Isolated and Hydrated Sugars: Mannose, Galactose, Glucose, and Lactose. *J. Am. Chem. Soc.* **2005**, *127* (32), 11414–11425.

(93) Nir, E.; Hünig, I.; Kleinermanns, K.; de Vries, M. S. The nucleobase cytosine and the cytosine dimer investigated by double resonance laser spectroscopy and ab initio calculations. *Phys. Chem. Chem. Phys.* **2003**, *5* (21), 4780–4785.

(94) Plützer, C.; Kleinermanns, K. Tautomers and electronic states of jet-cooled adenine investigated by double resonance spectroscopy. *Phys. Chem. Chem. Phys.* **2002**, *4* (20), 4877–4882.

(95) Chin, W.; Mons, M.; Piuzzi, F.; Tardivel, B.; Dimicoli, I.; Gorb, L.; Leszczynski, J. Gas Phase Rotamers of the Nucleobase 9-Methylguanine Enol and Its Monohydrate: Optical Spectroscopy and Quantum Mechanical Calculations. *J. Phys. Chem. A* **2004**, *108* (40), 8237–8243.

(96) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. Guanine tautomerism revealed by UV–UV and IR–UV hole burning spectroscopy. *J. Chem. Phys.* **2001**, *115* (10), 4604.

(97) Ligare, M.; Siouri, F.; Bludsky, O.; Nachtigallová, D.; de Vries, M. S. Characterizing the dark state in thymine and uracil by double resonant spectroscopy and quantum computation. *Phys. Chem. Chem. Phys.* **2015**, *17* (37), 24336–24341.

(98) Lee, J. J.; Albrecht, M.; Rice, C. A.; Suhm, M. A.; Stamm, A.; Zimmer, M.; Gerhards, M. Adaptive Aggregation of Peptide Model Systems. *J. Phys. Chem. A* **2013**, *117* (32), 7050–7063.

(99) Gerlach, A.; Unterberg, C.; Fricke, H.; Gerhards *, M. Structures of Ac–Trp–OMe and its dimer (Ac–Trp–OMe) 2 in the gas phase: influence of a polar group in the side-chain. *Mol. Phys.* **2005**, *103* (11-12), 1521–1529.

(100) Stanca-Kaposta, E. C.; Çarçabal, P.; Cocinero, E. J.; Hurtado, P.; Simons, J. P. Carbohydrate–Aromatic Interactions: Vibrational Spectroscopy and Structural Assignment of Isolated Monosaccharide Complexes with p -Hydroxy Toluene and N -Acetyl l -Tyrosine Methylamide. *J. Phys. Chem. B* **2013**, *117* (27), 8135–8142.

(101) Fricke, H.; Gerlach, A.; Gerhards, M. Structure of a β-sheet model system in the gas phase: Analysis of the fingerprint region up to 10 μm. *Phys. Chem. Chem. Phys.* **2006**, *8* (14), 1660.

(102) Gloaguen, E.; Tardivel, B.; Mons, M. Gas phase double-resonance IR/UV spectroscopy of an alanine dipeptide analogue using a non-covalently bound UV-tag: observation of

a folded peptide conformation in the Ac-Ala-NH2–toluene complex. *Struct. Chem.* **2016**, *27* (1), 225–230.

(103) Fricke, H.; Funk, A.; Schrader, T.; Gerhards, M. Investigation of secondary structure elements by IR/UV double resonance spectroscopy: analysis of an isolated beta-sheet model system. *J. Am. Chem. Soc.* **2008**, *130* (14), 4692–4698.

(104) Gerhards, M.; Unterberg, C.; Gerlach, A.; Jansen, A. β-sheet model systems in the gas phase: Structures and vibrations of Ac–Phe–NHMe and its dimer (Ac–Phe–NHMe) 2. *Phys. Chem. Chem. Phys.* **2004**, *6* (10), 2682–2690.

(105) Gerhards, M.; Unterberg, C. Structures of the protected amino acid Ac–Phe–OMe and its dimer: A β-sheet model system in the gas phase. *Phys. Chem. Chem. Phys.* **2002**, *4* (10), 1760–1765.

(106) Vaden, T. D.; Gowers, S. A. N.; Snoek, L. C. Observation of β-Sheet Aggregation in a Gas-Phase Tau-Peptide Dimer. *J. Am. Chem. Soc.* **2009**, *131* (7), 2472–2474.

(107) Jockusch, R. A.; Kroemer, R. T.; Talbot, F. O.; Simons, J. P. Hydrated Sugars in the Gas Phase: Spectroscopy and Conformation of Singly Hydrated Phenyl β- <scp>d</scp> - Glucopyranoside †. *J. Phys. Chem. A* **2003**, *107* (49), 10725–10732.

(108) Fricke, H.; Schwing, K.; Gerlach, A.; Unterberg, C.; Gerhards, M. Investigations of the water clusters of the protected amino acid Ac-Phe-OMe by applying IR/UV double resonance spectroscopy: microsolvation of the backbone. *Phys. Chem. Chem. Phys.* **2010**, *12* (14), 3511.

(109) Zhu, H.; Blom, M.; Compagnon, I.; Rijs, A. M.; Roy, S.; von Helden, G.; Schmidt, B. Conformations and vibrational spectra of a model tripeptide: change of secondary structure upon micro-solvation. *Phys. Chem. Chem. Phys.* **2010**, *12* (14), 3415.

(110) Ebata, T.; Hashimoto, T.; Ito, T.; Inokuchi, Y.; Altunsu, F.; Brutschy, B.; Tarakeshwar, P. Hydration profiles of aromatic amino acids: conformations and vibrations of <scp>l</scp> -phenylalanine–(H $_2$ O) $_n$ clusters. *Phys. Chem. Chem. Phys.* **2006**, *8* (41), 4783–4791.

(111) Lee, K. T.; Sung, J.; Lee, K. J.; Kim, S. K.; Park, Y. D. Resonant two-photon ionization study of jet-cooled amino acid: L-phenylalanine and its monohydrated complex. *J. Chem. Phys.* **2002**, *116* (19), 8251.

(112) Alonso, J. L.; Cocinero, E. J.; Lesarri, A.; Sanz, M. E.; López, J. C. The Glycine–Water Complex. *Angew. Chemie Int. Ed.* **2006**, *45* (21), 3471–3474.

(113) Fricke, H.; Gerlach, A.; Unterberg, C.; Rzepecki, P.; Schrader, T.; Gerhards, M. Structure of the tripeptide model Ac–Val–Tyr(Me)–NHMe and its cluster with water investigated by IR/UV double resonance spectroscopy. *Phys. Chem. Chem. Phys.* **2004**, *6* (19), 4636–4641.

(114) Zwier, T. S. Laser Spectroscopy of Jet-Cooled Biomolecules and Their Water-Containing Clusters: Water Bridges and Molecular Conformation. *J. Phys. Chem. A* **2001**, *105* (39), 8827–8839.

(115) Schwing, K.; Reyheller, C.; Schaly, A.; Kubik, S.; Gerhards, M. Structural Analysis of an

Isolated Cyclic Tetrapeptide and its Monohydrate by Combined IR/UV Spectroscopy. *ChemPhysChem* **2011**, *12* (10), 1981–1988.

(116) Snoek, L. C.; Kroemer, R. T.; Simons, J. P. A spectroscopic and computational exploration of tryptophan–water cluster structures in the gas phase. *Phys. Chem. Chem. Phys.* **2002**, *4* (11), 2130–2139.

(117) Biswal, H. S.; Loquais, Y.; Tardivel, B.; Gloaguen, E.; Mons, M. Isolated Monohydrates of a Model Peptide Chain: Effect of a First Water Molecule on the Secondary Structure of a Capped Phenylalanine. *J. Am. Chem. Soc.* **2011**, *133* (11), 3931–3942.

(118) Bush, M. F.; O'Brien, J. T.; Prell, J. S.; Saykally, R. J.; Williams, E. R. Infrared Spectroscopy of Cationized Arginine in the Gas Phase: Direct Evidence for the Transition from Nonzwitterionic to Zwitterionic Structure. *J. Am. Chem. Soc.* **2007**, *129* (6), 1612–1622.

(119) Fukui, K.; Takada, Y.; Sumiyoshi, T.; Imai, T.; Takahashi, K. Infrared Multiphoton Dissociation Spectroscopic Analysis of Peptides and Oligosaccharides by Using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with a Midinfrared Free-Electron Laser. *J. Phys. Chem. B* **2006**, *110* (32), 16111–16116.

(120) Bakker, J. M.; Besson, T.; Lemaire, J.; Scuderi, D.; Maître, P. Gas-Phase Structure of a π-Allyl–Palladium Complex: Efficient Infrared Spectroscopy in a 7 T Fourier Transform Mass Spectrometer. *J. Phys. Chem. A* **2007**, *111* (51), 13415–13424.

(121) Oomens, J.; van Roij, A. J. A.; Meijer, G.; von Helden, G. Gas-Phase Infrared Photodissociation Spectroscopy of Cationic Polyaromatic Hydrocarbons. *Astrophys. J.* **2000**, *542* (1), 404–410.

(122) Valle, J. J.; Eyler, J. R.; Oomens, J.; Moore, D. T.; van der Meer, A. F. G.; von Helden, G.; Meijer, G.; Hendrickson, C. L.; Marshall, A. G.; Blakney, G. T. Free electron laser-Fourier transform ion cyclotron resonance mass spectrometry facility for obtaining infrared multiphoton dissociation spectra of gaseous ions. *Rev. Sci. Instrum.* **2005**, *76* (2), 023103.

(123) Wu, R.; McMahon, T. B. Infrared Multiple Photon Dissociation Spectra of Proline and Glycine Proton-Bound Homodimers. Evidence for Zwitterionic Structure. *J. Am. Chem. Soc.* **2007**, *129* (16), 4864–4865.

(124) Aleese, L. Mac; Simon, A.; McMahon, T. B.; Ortega, J.-M.; Scuderi, D.; Lemaire, J.; Maître, P. Mid-IR spectroscopy of protonated leucine methyl ester performed with an FTICR or a Paul type ion-trap. *Int. J. Mass Spectrom.* **2006**, *249-250*, 14–20.

(125) Kapota, C.; Lemaire, J.; Maître, P.; Ohanessian, G. Vibrational Signature of Charge Solvation vs Salt Bridge Isomers of Sodiated Amino Acids in the Gas Phase. *J. Am. Chem. Soc.* **2004**, *126* (6), 1836–1842.

(126) Forbes, M. W.; Bush, M. F.; Polfer, N. C.; Oomens, J.; Dunbar, R. C.; Williams, E. R.; Jockusch, R. A. Infrared Spectroscopy of Arginine Cation Complexes: Direct Observation of Gas-Phase Zwitterions. *J. Phys. Chem. A* **2007**, *111* (46), 11759–11770.

(127) Polfer, N. C.; Oomens, J.; Moore, D. T.; von Helden, G.; Meijer, G.; Dunbar, R. C. Infrared Spectroscopy of Phenylalanine Ag(I) and Zn(II) Complexes in the Gas Phase. *J.*

Am. Chem. Soc. **2006**, *128* (2), 517–525.

(128) Polfer, N. C.; Paizs, B.; Snoek, L. C.; Compagnon, I.; Suhai, S.; Meijer, G.; von Helden, G.; Oomens, J. Infrared Fingerprint Spectroscopy and Theoretical Studies of Potassium Ion Tagged Amino Acids and Peptides in the Gas Phase. *J. Am. Chem. Soc.* **2005**, *127* (23), 8571–8579.

(129) Lagutschenkov, A.; Langer, J.; Berden, G.; Oomens, J.; Dopfer, O. Infrared spectra of protonated neurotransmitters: dopamine. *Phys. Chem. Chem. Phys.* **2011**, *13* (7), 2815–2823.

(130) Lagutschenkov, A.; Langer, J.; Berden, G.; Oomens, J.; Dopfer, O. Infrared Spectra of Protonated Neurotransmitters: Serotonin. *J. Phys. Chem. A* **2010**, *114* (50), 13268–13276.

(131) Vaden, T. D.; de Boer, T. S. J. A.; MacLeod, N. A.; Marzluff, E. M.; Simons, J. P.; Snoek, L. C. Infrared spectroscopy and structure of photochemically protonated biomolecules in the gas phase: a noradrenaline analogue, lysine and alanyl alanine. *Phys. Chem. Chem. Phys.* **2007**, *9* (20), 2549.

(132) Polfer, N. C.; Oomens, J. Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom. Rev.* **2009**, *28* (3), 468–494.

(133) Lagutschenkov, A.; Langer, J.; Berden, G.; Oomens, J.; Dopfer, O. Infrared spectra of the protonated neurotransmitter histamine: competition between imidazolium and ammonium isomers in the gas phase. *Phys. Chem. Chem. Phys.* **2011**, *13* (34), 15644.

(134) Correia, C. F.; Clavaguera, C.; Erlekam, U.; Scuderi, D.; Ohanessian, G. IRMPD Spectroscopy of a Protonated, Phosphorylated Dipeptide. *ChemPhysChem* **2008**, *9* (17), 2564–2573.

(135) Correia, C. F.; Balaj, P. O.; Scuderi, D.; Maitre, P.; Ohanessian, G. Vibrational Signatures of Protonated, Phosphorylated Amino Acids in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130* (11), 3359–3370.

(136) Oomens, J.; Polfer, N.; Moore, D. T.; van der Meer, L.; Marshall, A. G.; Eyler, J. R.; Meijer, G.; von Helden, G.; Barth, A.; Zscherp, C.; et al. Charge-state resolved mid-infrared spectroscopy of a gas-phase protein. *Phys. Chem. Chem. Phys.* **2005**, *7* (7), 1345.

(137) Kupser, P.; Pagel, K.; Oomens, J.; Polfer, N. C.; Koksch, B.; Meijer, G.; von Helden, G. Amide-I and -II vibrations of the cyclic beta-sheet model peptide gramicidin S in the gas phase. *J. Am. Chem. Soc.* **2010**, *132* (6), 2085–2093.

(138) Makarov, A. A.; Petrova, I. Y.; Ryabov, E. A.; Letokhov, V. S. Statistical Inhomogeneous Broadening of Infrared and Raman Transitions in Highly Vibrationally Excited XY 6 Molecules. *J. Phys. Chem. A* **1998**, *102* (9), 1438–1449.

(139) Parneix, P.; Basire, M.; Calvo, F. Accurate Modeling of Infrared Multiple Photon Dissociation Spectra: The Dynamical Role of Anharmonicities. *J. Phys. Chem. A* **2013**, *117* (19), 3954–3959.

(140) Nolting, D.; Marian, C.; Weinkauf, R. Protonation effect on the electronic spectrum of

tryptophan in the gas phase. *Phys. Chem. Chem. Phys.* **2004**, *6* (10), 2633.

(141) Boyarkin, O. V.; Mercier, S. R.; Kamariotis, A.; Rizzo, T. R. Electronic Spectroscopy of Cold, Protonated Tryptophan and Tyrosine. *J. Am. Chem. Soc.* **2006**, *128* (9), 2816–2817.

(142) Trippel, S.; Mikosch, J.; Berhane, R.; Otto, R.; Weidemüller, M.; Wester, R. Photodetachment of Cold OH- in a Multipole Ion Trap. *Phys. Rev. Lett.* **2006**, *97* (19), 193003.

(143) Dzhonson, A.; Gerlich, D.; Bieske, E. J.; Maier, J. P. Apparatus for the study of electronic spectra of collisionally cooled cations: para-dichlorobenzene. *J. Mol. Struct.* **2006**, *795* (1-3), 93–97.

(144) Gerlich, D. Experimental investigations of ion–molecule reactions relevant to interstellar chemistry. *J. Chem. Soc., Faraday Trans.* **1993**, *89* (13), 2199–2208.

(145) Asmis, K. R.; Brümmer, M.; Kaposta, C.; Santambrogio, G.; von Helden, G.; Meijer, G.; Rademann, K.; Wöste, L.; Yin, D.; Xu, N.; et al. Mass-selected infrared photodissociation spectroscopy of V4O10+. *Phys. Chem. Chem. Phys.* **2002**, *4* (7), 1101–1104.

(146) Wang, X.-B.; Wang, L.-S. Development of a low-temperature photoelectron spectroscopy instrument using an electrospray ion source and a cryogenically controlled ion trap. *Rev. Sci. Instrum.* **2008**, *79* (7), 073108.

(147) Choi, C. M.; Kim, H. J.; Lee, J. H.; Shin, W. J.; Yoon, T. O.; Kim, N. J.; Heo, J. Ultraviolet Photodepletion Spectroscopy of Dibenzo-18-Crown-6-Ether Complexes with Alkali Metal Cations. *J. Phys. Chem. A* **2009**, *113* (29), 8343–8350.

(148) Choi, C. M.; Choi, D. H.; Heo, J.; Kim, N. J.; Kim, S. K. Ultraviolet-Ultraviolet Hole Burning Spectroscopy in a Quadrupole Ion Trap: Dibenzo[18]crown-6 Complexes with Alkali Metal Cations. *Angew. Chemie Int. Ed.* **2012**, *51* (29), 7297–7300.

(149) Redwine, J. G.; Davis, Z. a.; Burke, N. L.; Oglesbee, R. a.; McLuckey, S. a.; Zwier, T. S. A novel ion trap based tandem mass spectrometer for the spectroscopic study of cold gas phase polyatomic ions. *Int. J. Mass Spectrom.* **2013**, *348*, 9–14.

(150) Goebbert, D. J.; Wende, T.; Bergmann, R.; Meijer, G.; Asmis, K. R. Messenger-Tagging Electrosprayed Ions: Vibrational Spectroscopy of Suberate Dianions. *J. Phys. Chem. A* **2009**, *113* (20), 5874–5880.

(151) Svendsen, A.; Lorenz, U. J.; Boyarkin, O. V.; Rizzo, T. R. A new tandem mass spectrometer for photofragment spectroscopy of cold, gas-phase molecular ions. *Rev. Sci. Instrum.* **2010**, *81* (7), 073107.

(152) Fujihara, A.; Matsumoto, H.; Shibata, Y.; Ishikawa, H.; Fuke, K. Photodissociation and Spectroscopic Study of Cold Protonated Dipeptides. *J. Phys. Chem. A* **2008**, *112* (7), 1457–1463.

(153) Kamrath, M. Z.; Relph, R. A.; Guasco, T. L.; Leavitt, C. M.; Johnson, M. A. Vibrational predissociation spectroscopy of the H2-tagged mono- and dicarboxylate anions of dodecanedioic acid. *Int. J. Mass Spectrom.* **2011**, *300* (2-3), 91–98.

(154) Papadopoulos, G.; Svendsen, A.; Boyarkin, O. V.; Rizzo, T. R. Conformational Distribution of Bradykinin [bk + 2 H](2+) Revealed by Cold Ion Spectroscopy Coupled with FAIMS. *J. Am. Soc. Mass Spectrom.* **2012**, No. February.

(155) Masson, A.; Kamrath, M. Z.; Perez, M. A. S.; Glover, M. S.; Rothlisberger, U.; Clemmer, D. E.; Rizzo, T. R. Infrared Spectroscopy of Mobility-Selected H+-Gly-Pro-Gly-Gly (GPGG). *J. Am. Soc. Mass Spectrom.* **2015**, *26* (9), 1444–1454.

(156) Jašík, J.; Žabka, J.; Roithová, J.; Gerlich, D. Infrared spectroscopy of trapped molecular dications below 4K. *Int. J. Mass Spectrom.* **2013**, *354-355*, 204–210.

(157) Filsinger, F.; Ahn, D.-S.; Meijer, G.; von Helden, G.; Goyal, S.; Schutt, D. L.; Scoles, G.; Fröchtenicht, R.; Toennies, J.; Vilesov, A.; et al. Photoexcitation of mass/charge selected hemin+, caught in helium nanodroplets. *Phys. Chem. Chem. Phys.* **2012**, *14* (38), 13370.

(158) Boyarkin, O. V; Kopysov, V. Cryogenically cooled octupole ion trap for spectroscopy of biomolecular ions. *Rev. Sci. Instrum.* **2014**, *85* (3), 033105.

(159) Gerlich, D. Inhomogeneous RF Fields: A Versatile Tool for the Study of Processes with Slow Ions. In *Advances in Chemical Physics: State-Selected and State-To-State Ion-Molecule Reaction Dynamics, Part 1. Experiment, Volume 82*; Ng, C.-Y., Baer, M., Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons, Inc.: Hoboken, NY, USA, 2007; pp 1–176.

(160) Okumura, M.; Yeh, L. I.; Myers, J. D.; Lee, Y. T. Infrared spectra of the cluster ions H7O+3·H2 and H9O+4·H2. *J. Chem. Phys.* **1986**, *85* (4), 2328.

(161) Garand, E.; Kamrath, M. Z.; Jordan, P. A.; Wolk, A. B.; Leavitt, C. M.; McCoy, A. B.; Miller, S. J.; Johnson, M. A. Determination of Noncovalent Docking by Infrared Spectroscopy of Cold Gas-Phase Complexes. *Science (80-.).* **2012**, *335* (6069), 694–698.

(162) Kamrath, M. Z.; Garand, E.; Jordan, P. A.; Leavitt, C. M.; Wolk, A. B.; Van Stipdonk, M. J.; Miller, S. J.; Johnson, M. A. Vibrational Characterization of Simple Peptides Using Cryogenic Infrared Photodissociation of H 2 -Tagged, Mass-Selected Ions. *J. Am. Chem. Soc.* **2011**, *133* (16), 6440–6448.

(163) Wolk, A. B.; Leavitt, C. M.; Garand, E.; Johnson, M. A. Cryogenic Ion Chemistry and Spectroscopy. *Acc. Chem. Res.* **2014**, *47* (1), 202–210.

(164) Bierau, F.; Kupser, P.; Meijer, G.; von Helden, G. Catching Proteins in Liquid Helium Droplets. *Phys. Rev. Lett.* **2010**, *105* (13), 133402.

(165) González Flórez, A. I.; Ahn, D.-S.; Gewinner, S.; Schöllkopf, W.; von Helden, G. IR spectroscopy of protonated leu-enkephalin and its 18-crown-6 complex embedded in helium droplets. *Phys. Chem. Chem. Phys.* **2015**, *17* (34), 21902–21911.

(166) González Flórez, A. I.; Mucha, E.; Ahn, D.-S.; Gewinner, S.; Schöllkopf, W.; Pagel, K.; von Helden, G. Charge-Induced Unzipping of Isolated Proteins to a Defined Secondary Structure. *Angew. Chemie Int. Ed.* **2016**, *55* (10), 3295–3299.

(167) Mercier, S. R.; Boyarkin, O. V.; Kamariotis, A.; Guglielmi, M.; Tavernelli, I.; Cascella, M.;

Rothlisberger, U.; Rizzo, T. R. Microsolvation effects on the excited-state dynamics of protonated tryptophan. *J. Am. Chem. Soc.* **2006**, *128* (51), 16938–16943.

(168) Stearns, J. A.; Mercier, S.; Seaiby, C.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R. Conformation-specific spectroscopy and photodissociation of cold, protonated tyrosine and phenylalanine. *J. Am. Chem. Soc.* **2007**, *129* (38), 11814–11820.

(169) Stearns, J. A.; Seaiby, C.; Boyarkin, O. V.; Rizzo, T. R. Spectroscopy and conformational preferences of gas-phase helices. *Phys. Chem. Chem. Phys.* **2009**, *11* (1), 125–132.

(170) Stearns, J. A.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R. Conformation-specific infrared and ultraviolet spectroscopy of tyrosine-based protonated dipeptides. *J. Chem. Phys.* **2007**, *127* (15), 154322.

(171) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. Spectroscopic signatures of gas-phase helices: Ac-Phe-(Ala)5-Lys-H+ and Ac-Phe-(Ala)10-Lys-H+. *J. Am. Chem. Soc.* **2007**, *129* (45), 13820–13821.

(172) Wassermann, T. N.; Boyarkin, O. V.; Paizs, B.; Rizzo, T. R. Conformation-Specific Spectroscopy of Peptide Fragment Ions in a Low-Temperature Ion Trap. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (6), 1029–1045.

(173) Zabuga, A. V; Rizzo, T. R. Capping Motif for Peptide Helix Formation. *J. Phys. Chem. Lett.* **2015**, *6* (9), 1504–1508.

(174) Aseev, O.; Perez, M. A. S.; Rothlisberger, U.; Rizzo, T. R. Cryogenic Spectroscopy and Quantum Molecular Dynamics Determine the Structure of Cyclic Intermediates Involved in Peptide Sequence Scrambling. *J. Phys. Chem. Lett.* **2015**, *6* (13), 2524–2529.

(175) Voronina, L.; Masson, A.; Kamrath, M. Z.; Schubert, F.; Clemmer, D. E.; Baldauf, C.; Rizzo, T. R. Conformations of prolyl-peptide bonds in the bradykinin 1-5 fragment in solution and in the gas phase. *J. Am. Chem. Soc.* **2016**, jacs.6b04550.

(176) Nagornova, N. S.; Guglielmi, M.; Doemer, M.; Tavernelli, I.; Rothlisberger, U.; Rizzo, T. R.; Boyarkin, O. V. Cold-ion spectroscopy reveals the intrinsic structure of a decapeptide. *Angew. Chem. Int. Ed. Engl.* **2011**, *50* (23), 5383–5386.

(177) Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V. Interplay of intra- and intermolecular H-bonding in a progressively solvated macrocyclic peptide. *Science* **2012**, *336* (6079), 320–323.

(178) Guidi, M.; Lorenz, U. J.; Papadopoulos, G.; Boyarkin, O. V.; Rizzo, T. R. Spectroscopy of protonated peptides assisted by infrared multiple photon excitation. *J. Phys. Chem. A* **2009**, *113* (5), 797–799.

(179) Papadopoulos, G.; Svendsen, A.; Boyarkin, O. V.; Rizzo, T. R. Spectroscopy of mobility-selected biomolecular ions. *Faraday Discuss.* **2011**, *150*, 243.

(180) Voronina, L.; Rizzo, T. R. Spectroscopic studies of kinetically trapped conformations in the gas phase: the case of triply protonated bradykinin. *Phys. Chem. Chem. Phys.* **2015**, *17* (39), 25828–25836.

(181) Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V. Exploring the Mechanism of IR-UV

Double-Resonance for Quantitative Spectroscopy of Protonated Polypeptides and Proteins. *Angew. Chemie Int. Ed.* **2013**, *52* (23), 6002–6005.

(182) Pereverzev, A. Y.; Cheng, X.; Nagornova, N. S.; Reese, D. L.; Steele, R. P.; Boyarkin, O. V. Vibrational Signatures of Conformer-Specific Intramolecular Interactions in Protonated Tryptophan. *J. Phys. Chem. A* **2016**, acs.jpca.6b05605.

(183) Kopysov, V.; Nagornova, N. S.; Boyarkin, O. V. Identification of Tyrosine-Phosphorylated Peptides Using Cold Ion Spectroscopy. *J. Am. Chem. Soc.* **2014**, *136* (26), 9288–9291.

(184) Makarov, A. Mass spectrometer. US5886346 A, 1999.

(185) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72* (6), 1156–1162.

(186) Kingdon, K. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys. Rev.* **1923**, *21* (4), 408–418.

(187) Makarov, A.; Denisov, E. Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1486–1495.

(188) Denisov, E.; Damoc, E.; Lange, O.; Makarov, A. Orbitrap mass spectrometry with resolving powers above 1,000,000. *Int. J. Mass Spectrom.* **2012**, *325-327* (null), 80–85.

(189) Olsen, J. V; de Godoy, L. M. F.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010–2021.

(190) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* **2012**, *9* (11), 1084–1086.

(191) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (7), 977–982.

(192) Scigelova, M.; Makarov, A. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **2006**, *6 Suppl 2*, 16–21.

(193) Ahlf, D. R.; Compton, P. D.; Tran, J. C.; Early, B. P.; Thomas, P. M.; Kelleher, N. L. Evaluation of the compact high-field orbitrap for top-down proteomics of human cells. *J. Proteome Res.* **2012**, *11* (8), 4308–4314.

(194) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol. Cell. Proteomics* **2010**, *9* (10), 2252–2261.

(195) Boja, E. S.; Phillips, D.; French, S. A.; Harris, R. A.; Balaban, R. S. Quantitative mitochondrial phosphoproteomics using iTRAQ on an LTQ-Orbitrap with high energy collision dissociation. *J. Proteome Res.* **2009**, *8* (10), 4665–4675.

(196) Koulman, A.; Woffendin, G.; Narayana, V. K.; Welchman, H.; Crone, C.; Volmer, D. A. High-resolution extracted ion chromatography, a new tool for metabolomics and

lipidomics using a second-generation orbitrap mass spectrometer. *Rapid Commun. Mass Spectrom.* **2009**, *23* (10), 1411–1418.

(197) Ejsing, C. S.; Sampaio, J. L.; Surendranath, V.; Duchoslav, E.; Ekroos, K.; Klemm, R. W.; Simons, K.; Shevchenko, A. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (7), 2136–2141.

(198) Schuhmann, K.; Herzog, R.; Schwudke, D.; Metelmann-Strupat, W.; Bornstein, S. R.; Shevchenko, A. Bottom-up shotgun lipidomics by higher energy collisional dissociation on LTQ Orbitrap mass spectrometers. *Anal. Chem.* **2011**, *83* (14), 5480–5487.

(199) Almeida, R.; Pauling, J. K.; Sokol, E.; Hannibal-Bach, H. K.; Ejsing, C. S. Comprehensive lipidome analysis by shotgun lipidomics on a hybrid quadrupole-orbitrap-linear ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (1), 133–148.

(200) Breitling, R.; Pitt, A. R.; Barrett, M. P. Precision mapping of the metabolome. *Trends Biotechnol.* **2006**, *24* (12), 543–548.

(201) Werner, E.; Croixmarie, V.; Umbdenstock, T.; Ezan, E.; Chaminade, P.; Tabet, J.-C.; Junot, C. Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution. *Anal. Chem.* **2008**, *80* (13), 4918–4932.

(202) Lim, H.-K.; Chen, J.; Sensenhauser, C.; Cook, K.; Subrahmanyam, V. Metabolite identification by data-dependent accurate mass spectrometric analysis at resolving power of 60,000 in external calibration mode using an LTQ/Orbitrap. *Rapid Commun. Mass Spectrom.* **2007**, *21* (12), 1821–1832.

(203) Hogenboom, A. C.; van Leerdam, J. A.; de Voogt, P. Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography-hybrid linear ion trap Orbitrap mass spectrometry. *J. Chromatogr. A* **2009**, *1216* (3), 510–519.

(204) Blay, P.; Hui, J. P. M.; Chang, J.; Melanson, J. E. Screening for multiple classes of marine biotoxins by liquid chromatography-high-resolution mass spectrometry. *Anal. Bioanal. Chem.* **2011**, *400* (2), 577–585.

(205) Thomas, A.; Geyer, H.; Schänzer, W.; Crone, C.; Kellmann, M.; Moehring, T.; Thevis, M. Sensitive determination of prohibited drugs in dried blood spots (DBS) for doping controls by means of a benchtop quadrupole/Orbitrap mass spectrometer. *Anal. Bioanal. Chem.* **2012**, *403* (5), 1279–1289.

(206) Virus, E. D.; Sobolevsky, T. G.; Rodchenkov, G. M. Introduction of HPLC/orbitrap mass spectrometry as screening method for doping control. *J. Mass Spectrom.* **2008**, *43* (7), 949–957.

(207) Olsen, J. V; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, *4* (9), 709–712.

(208) McAlister, G. C.; Phanstiel, D.; Good, D. M.; Berggren, W. T.; Coon, J. J. Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. *Anal. Chem.* **2007**, *79* (10), 3525–3534.

(209) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668–9673.

(210) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization. *Anal. Chem.* **2015**, *87* (8), 4152–4158.

(211) Vasicek, L. A.; Ledvina, A. R.; Shaw, J.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J.; Brodbelt, J. S. Implementing photodissociation in an Orbitrap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (6), 1105–1108.

(212) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast ultraviolet photodissociation at 193 nm and its applicability to proteomic workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.

(213) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.

(214) Girod, M.; Sanader, Z.; Vojkovic, M.; Antoine, R.; MacAleese, L.; Lemoine, J.; Bonacic-Koutecky, V.; Dugourd, P. UV photodissociation of proline-containing peptide ions: insights from molecular dynamics. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (3), 432–443.

(215) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. Implementation of UV-photodissociation on a benchtop Q Exactive mass spectrometer and its application to phospho-proteomics. *Anal. Chem.* **2016**.

(216) Kopysov, V.; Makarov, A.; Boyarkin, O. V. Colors for Molecular Masses: Fusion of Spectroscopy and Mass Spectrometry for Identification of Biomolecules. *Anal. Chem.* **2015**, *87* (9), 4607–4611.

(217) Orfanidis, S. J. *Introduction to Signal Processing*; Prentice Hall, 1996.

(218) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.* **1985**, *57* (3), 675–679.

(219) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-.).* **1989**, *246* (4926), 64–71.

(220) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion Process.* **1987**, *78*, 53–68.

(221) Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **1988**, *60* (20), 2299–2301.

(222) He, F.; Emmett, M. R.; Håkansson, K.; Hendrickson, C. L.; Marshall, A. G. Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass

Measurement. *J. Proteome Res.* **2004**, *3* (1), 61–67.

(223) Creese, A. J.; Smart, J.; Cooper, H. J. Large-scale analysis of peptide sequence variants: the case for high-field asymmetric waveform ion mobility spectrometry. *Anal. Chem.* **2013**, *85* (10), 4836–4843.

(224) Winter, D.; Pipkorn, R.; Lehmann, W. D. Separation of peptide isomers and conformers by ultra performance liquid chromatography. *J. Sep. Sci.* **2009**, *32* (8), 1111–1119.

(225) Shvartsburg, A. A.; Creese, A. J.; Smith, R. D.; Cooper, H. J. Separation of a set of peptide sequence isomers using differential ion mobility spectrometry. *Anal. Chem.* **2011**, *83* (18), 6918–6923.

(226) Wu, C.; Siems, W. F.; Klasmeier, J.; Hill, H. H. Separation of Isomeric Peptides Using Electrospray Ionization/High-Resolution Ion Mobility Spectrometry. *Anal. Chem.* **2000**, *72* (2), 391–395.

(227) Norbeck, A. D.; Monroe, M. E.; Adkins, J. N.; Anderson, K. K.; Daly, D. S.; Smith, R. D. The utility of accurate mass and LC elution time information in the analysis of complex proteomes. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1239–1249.

(228) Riddle, L. A.; Guiochon, G. Influence of Mobile Phase Gradients on the Retention and Separation of Peptides from a Cytochrome-c Digest by Reversed-Phase Liquid Chromatography. *Chromatographia* **2006**, *64* (3-4), 1–7.

(229) Tarasova, I. A.; Perlova, T. Y.; Pridatchenko, M. L.; Goloborod'ko, A. A.; Levitsky, L. I.; Evreinov, V. V.; Guryca, V.; Masselon, C. D.; Gorshkov, A. V.; Gorshkov, M. V. Inversion of chromatographic elution orders of peptides and its importance for proteomics. *J. Anal. Chem.* **2012**, *67* (13), 1014–1025.

(230) Féraud, G.; Dedonder, C.; Jouvet, C.; Inokuchi, Y.; Haino, T.; Sekiya, R.; Ebata, T. Development of Ultraviolet-Ultraviolet Hole-Burning Spectroscopy for Cold Gas-Phase Ions. *J. Phys. Chem. Lett.* **2014**, *5* (7), 1236–1240.

(231) Tabarin, T.; Antoine, R.; Broyer, M.; Dugourd, P. Specific photodissociation of peptides with multi-stage mass spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19* (20), 2883–2892.

(232) Brodbelt, J. S. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.

(233) Cui, W.; Thompson, M. S.; Reilly, J. P. Pathways of peptide ion fragmentation induced by vacuum ultraviolet light. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.

(234) Joly, L.; Antoine, R.; Broyer, M.; Dugourd, P.; Lemoine, J. Specific UV photodissociation of tyrosyl-containing peptides in multistage mass spectrometry. *J. Mass Spectrom.* **2007**, *42* (6), 818–824.

(235) Yeh, G. K.; Sun, Q.; Meneses, C.; Julian, R. R. Rapid peptide fragmentation without electrons, collisions, infrared radiation, or native chromophores. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 385–393.

(236) Blume-Jensen, P.; Hunter, T. Oncogenic kinase signalling. *Nature* **2001**, *411* (6835),

355–365.

(237) Engholm-Keller, K.; Larsen, M. R. Technologies and challenges in large-scale phosphoproteomics. *Proteomics* **2013**, *13* (6), 910–931.

(238) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (4), 1774–1779.

(239) Palumbo, A. M.; Smith, S. A.; Kalcic, C. L.; Dantus, M.; Stemmer, P. M.; Reid, G. E. Tandem mass spectrometry strategies for phosphoproteome analysis. *Mass Spectrom. Rev. 30* (4), 600–625.

(240) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol. Cell. Proteomics* **2012**, *11* (12), 1758–1767.

(241) Molden, R. C.; Goya, J.; Khan, Z.; Garcia, B. A. Stable isotope labeling of phosphoproteins for large-scale phosphorylation rate determination. *Mol. Cell. Proteomics* **2014**, *13* (4), 1106–1118.

(242) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. Phosphopeptide fragmentation and analysis by mass spectrometry. *J. Mass Spectrom.* **2009**, *44* (6), 861–878.

(243) Tholey, A.; Reed, J.; Lehmann, W. D. Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. *J. Mass Spectrom.* **1999**, *34* (2), 117–123.

(244) Moyer, S. C.; VonSeggern, C. E.; Cotter, R. J. Fragmentation of cationized phosphotyrosine containing peptides by atmospheric pressure MALDI/Ion trap mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (6), 581–592.

(245) Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005**, *24* (4), 508–548.

(246) Zabuga, A. V; Kamrath, M. Z.; Boyarkin, O. V; Rizzo, T. R. Fragmentation mechanism of UV-excited peptides in the gas phase. *J. Chem. Phys.* **2014**, *141* (15), 154309.

(247) Burke, N. L.; Redwine, J. G.; Dean, J. C.; McLuckey, S. A.; Zwier, T. S. UV and IR spectroscopy of cold protonated leucine enkephalin. *Int. J. Mass Spectrom.* **2015**, *378*, 196–205.

(248) Kang, H.; Jouvet, C.; Dedonder-Lardeux, C.; Martrenchard, S.; Grégoire, G.; Desfrançois, C.; Schermann, J. P.; Barat, M.; Fayeton, J. a. Ultrafast deactivation mechanisms of protonated aromatic amino acids following UV excitation. *Phys. Chem. Chem. Phys.* **2005**, *7* (2), 394–398.

(249) Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V. Highly resolved spectra of gas-phase gramicidin s: a benchmark for peptide structure calculations. *J. Am. Chem. Soc.* **2010**, *132* (12), 4040–4041.

(250) Féraud, G.; Broquier, M.; Dedonder, C.; Jouvet, C.; Grégoire, G.; Soorkia, S. Excited State Dynamics of Protonated Phenylalanine and Tyrosine: Photo-Induced Reactions Following Electronic Excitation. *J. Phys. Chem. A* **2015**, *119* (23), 5914–5924.

(251) Bada, J. L. *Posttranslational Modifications Part A*; Methods in Enzymology; Elsevier, 1984; Vol. 106.

(252) Ohtani, S.; Yamada, Y.; Yamamoto, T.; Arany, S.; Gonmori, K.; Yoshioka, N. Comparison of age estimated from degree of racemization of aspartic acid, glutamic acid and alanine in the femur. *J. Forensic Sci.* **2004**, *49* (3), 441–445.

(253) Cloos, P. A. .; Christgau, S. Non-enzymatic covalent modifications of proteins: mechanisms, physiological consequences and clinical applications. *Matrix Biol.* **2002**, *21* (1), 39–52.

(254) Alkass, K.; Buchholz, B. A.; Ohtani, S.; Yamamoto, T.; Druid, H.; Spalding, K. L. Age estimation in forensic sciences: application of combined aspartic acid racemization and radiocarbon analysis. *Mol. Cell. Proteomics* **2010**, *9* (5), 1022–1030.

(255) Sztáray, J.; Memboeuf, A.; Drahos, L.; Vékey, K. Leucine enkephalin--a mass spectrometry standard. *Mass Spectrom. Rev. 30* (2), 298–320.

(256) Srebalus, C. A.; Li, J.; Marshall, W. S.; Clemmer, D. E. Gas-Phase Separations of Electrosprayed Peptide Libraries. *Anal. Chem.* **1999**, *71* (18), 3918–3927.

(257) Srebalus Barnes, C. A.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. Resolving Isomeric Peptide Mixtures: A Combined HPLC/Ion Mobility-TOFMS Analysis of a 4000-Component Combinatorial Library. *Anal. Chem.* **2002**, *74* (1), 26–36.

(258) Chhabra, N.; Aseri, M. L.; Padmanabhan, D. A review of drug isomerism and its significance. *Int. J. Appl. basic Med. Res.* **2013**, *3* (1), 16–18.

(259) Nguyen, L. A.; He, H.; Pham-Huy, C. Chiral drugs: an overview. *Int. J. Biomed. Sci.* **2006**, *2* (2), 85–100.

(260) Rabbaa, L.; Dautrey, S.; Colas-Linhart, N.; Carbon, C.; Farinotti, R. Absorption of ofloxacin isomers in the rat small intestine. *Antimicrob. Agents Chemother.* **1997**, *41* (10), 2274–2277.

(261) Foster, D. J.; Somogyi, A. A.; Dyer, K. R.; White, J. M.; Bochner, F. Steady-state pharmacokinetics of (R)- and (S)-methadone in methadone maintenance patients. *Br. J. Clin. Pharmacol.* **2000**, *50* (5), 427–440.

(262) Lane, S.; Al-Zubiedi, S.; Hatch, E.; Matthews, I.; Jorgensen, A. L.; Deloukas, P.; Daly, A. K.; Park, B. K.; Aarons, L.; Ogungbenro, K.; et al. The population pharmacokinetics of R- and S-warfarin: effect of genetic and clinical factors. *Br. J. Clin. Pharmacol.* **2012**, *73* (1), 66–76.

(263) Solanki, P.; Yadav, P.; Kantharia, N. Ephedrine: direct, indirect or mixed acting sympathomimetic? *Int. J. Basic Clin. Pharmacol.* **2014**, *3* (3), 431.

(264) Vansal, S. S.; Feller, D. R. Direct effects of ephedrine isomers on human β-adrenergic receptor subtypes. *Biochem. Pharmacol.* **1999**, *58* (5), 807–810.

(265) (WADA), W. A.-D. A. The World Anti-Doping Code. The 2015 Prohibited List International Standard https://wada-main-prod.s3.amazonaws.com/resources/files/wada-2015-prohibited-list-en.pdf.

(266) World Anti-Doping Agency (WADA). WADA Technical Document - TD2014DL. Decision Limits for the Confirmatory Quantification of Threshold Substances https://wada-main-prod.s3.amazonaws.com/resources/files/WADA-TD2014DL-v1-Decision-Limits-for-the-Quantification-of-Threshold-Substances-EN.pdf.

(267) Bijlsma, L.; Sancho, J. V; Hernández, F.; Niessen, W. M. A. Fragmentation pathways of drugs of abuse and their metabolites based on QTOF MS/MS and MS(E) accurate-mass spectra. *J. Mass Spectrom.* **2011**, *46* (9), 865–875.

(268) Monfort, N.; Martínez, L.; Bergés, R.; Segura, J.; Ventura, R. Screening method for stimulants in urine by UHPLC-MS/MS: identification of isomeric compounds. *Drug Test. Anal.* **2015**.

(269) Jacob, P.; Haller, C. A.; Duan, M.; Yu, L.; Peng, M.; Benowitz, N. L. Determination of Ephedra Alkaloid and Caffeine Concentrations in Dietary Supplements and Biological Fluids. *J. Anal. Toxicol.* **2004**, *28* (3), 152–159.

(270) Spyridaki, M.-H.; Kiousi, P.; Vonaparti, A.; Valavani, P.; Zonaras, V.; Zahariou, M.; Sianos, E.; Tsoupras, G.; Georgakopoulos, C. Doping control analysis in human urine by liquid chromatography-electrospray ionization ion trap mass spectrometry for the Olympic Games Athens 2004: determination of corticosteroids and quantification of ephedrines, salbutamol and morphine. *Anal. Chim. Acta* **2006**, *573-574*, 242–249.

(271) Deventer, K.; Pozo, O. J.; Van Eenoo, P.; Delbeke, F. T. Development and validation of an LC-MS/MS method for the quantification of ephedrines in urine. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **2009**, *877* (4), 369–374.

(272) Imaz, C.; Carreras, D.; Navajas, R.; Rodriguez, C.; Rodriguez, A. F.; Maynar, J.; Cortes, R. Determination of ephedrines in urine by high-performance liquid chromatography. *J. Chromatogr. A* **1993**, *631* (1-2), 201–205.

(273) Butz, P.; Kroemer, R. T.; Macleod, N. A.; Simons, J. P. Conformational Preferences of Neurotransmitters: Ephedrine and Its Diastereoisomer, Pseudoephedrine. *J. Phys. Chem. A* **2001**, *105* (3), 544–551.

(274) Chervenkov, S.; Wang, P. Q.; Braun, J. E.; Neusser, H. J. Fragmentation and conformation study of ephedrine by low- and high-resolution mass selective UV spectroscopy. *J. Chem. Phys.* **2004**, *121* (15), 7169–7174.

(275) Karaminkov, R.; Chervenkov, S.; Delchev, V.; Neusser, H. J. High-resolution mass-selective UV spectroscopy of pseudoephedrine: evidence for conformer-specific fragmentation. *J. Phys. Chem. A* **2011**, *115* (34), 9704–9713.

(276) Barry, T. L.; Petzinger, G. Structure and fragmentation mechanisms of some ions in the mass spectrum of ephedrine. *Biol. Mass Spectrom.* **1977**, *4* (3), 129–133.

(277) Karaminkov, R.; Chervenkov, S.; Härter, P.; Neusser, H. J. Resonance-enhanced two-photon ionization mass spectroscopy of ephedrine: Indication for a state-selective fragmentation in a flexible molecule. *Chem. Phys. Lett.* **2007**, *442* (4-6), 238–244.

(278) Gohlke, R. S.; McLafferty, F. W. Mass Spectrometric Analysis. Aliphatic Amines. *Anal. Chem.* **1962**, *34* (10), 1281–1287.

(279) Budnik, B. A.; Tsybin, Y. O.; Håkansson, P.; Zubarev, R. A. Ionization energies of multiply protonated polypeptides obtained by tandem ionization in Fourier transform mass spectrometers. *J. Mass Spectrom.* **2002**, *37* (11), 1141–1144.

(280) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. Effects of N-Terminus Substitution on the Structure and Spectroscopy of Gas-Phase Helices. *Chim. Int. J. Chem.* **2008**, *62* (4), 240–243.

(281) Kopysov, V.; Makarov, A.; Boyarkin, O. V. Nonstatistical UV Fragmentation of Gas-Phase Peptides Reveals Conformers and Their Structural Features. *J. Phys. Chem. Lett.* **2016**, 1067–1071.

(282) De Hoffmann, E.; Stroobant, V. *Mass Spectrometry: Principles and Applications*, 3rd ed.; John Wiley & Sons, Ltd: Chichester, UK, 2007.

(283) Armbruster, D. A.; Pry, T. Limit of blank, limit of detection and limit of quantitation. *Clin. Biochem. Rev.* **2008**, *29 Suppl 1*, S49–S52.

(284) International Narcotics Control Board. List of Psychotropic Substances under International Control https://www.incb.org/documents/Psychotropics/greenlist/Green_list_ENG_2015_new.pdf.

(285) Berry, M. W.; Browne, M.; Langville, A. N.; Pauca, V. P.; Plemmons, R. J. Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data Anal.* **2007**, *52* (1), 155–173.

(286) Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5* (2), 111–126.

(287) Lee, D. D.; Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401* (6755), 788–791.

(288) Owen, A. B.; Perry, P. O. Bi-cross-validation of the SVD and the nonnegative matrix factorization. *Ann. Appl. Stat.* **2009**, *3* (2), 564–594.

(289) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell. Proteomics* **2014**, *13* (12), 3698–3708.

(290) Nielsen, B. G.; Jensen, M. Ø.; Bohr, H. G. The probability distribution of side-chain conformations in [Leu] and [Met]enkephalin determines the potency and selectivity to mu and delta opiate receptors. *Biopolymers* **2003**, *71* (5), 577–592.

(291) Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **2003**, *10* (9), 787–797.

(292) Bowers, M. T.; Kemper, P. R.; von Helden, G.; van Koppen, P. A. Gas-phase ion chromatography: transition metal state selection and carbon cluster formation. *Science* **1993**, *260* (5113), 1446–1451.

(293) Warnke, S.; Baldauf, C.; Bowers, M. T.; Pagel, K.; von Helden, G. Photodissociation of

conformer-selected ubiquitin ions reveals site-specific cis/trans isomerization of proline peptide bonds. *J. Am. Chem. Soc.* **2014**, *136* (29), 10308–10314.

(294) Tseng, C.-M.; Lin, M.-F.; Yang, Y. L.; Ho, Y. C.; Ni, C.-K.; Chang, J.-L. Photostability of amino acids: photodissociation dynamics of phenylalanine chromophores. *Phys. Chem. Chem. Phys.* **2010**, *12* (19), 4989–4995.

(295) Loquais, Y.; Gloaguen, E.; Alauddin, M.; Brenner, V.; Tardivel, B.; Mons, M. On the near UV photophysics of a phenylalanine residue: conformation-dependent ππ* state deactivation revealed by laser spectroscopy of isolated neutral dipeptides. *Phys. Chem. Chem. Phys.* **2014**, *16* (40), 22192–22200.

(296) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393* (1-3), 51–57.

(297) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation energies in density functional theory: an evaluation and a diagnostic test. *J. Chem. Phys.* **2008**, *128* (4), 044118.

(298) Richard, R. M.; Herbert, J. M. Time-Dependent Density-Functional Description of the 1 L a State in Polycyclic Aromatic Hydrocarbons: Charge-Transfer Character in Disguise? *J. Chem. Theory Comput.* **2011**, *7* (5), 1296–1306.

(299) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14* (11), 1347–1363.

(300) Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries. *J. Chem. Theory Comput.* **2010**, *6* (9), 2872–2887.

(301) Féraud, G.; Broquier, M.; Dedonder-Lardeux, C.; Grégoire, G.; Soorkia, S.; Jouvet, C. Photofragmentation spectroscopy of cold protonated aromatic amines in the gas phase. *Phys. Chem. Chem. Phys.* **2014**, *16* (11), 5250–5259.

(302) Huang, C.-L.; Jiang, J.-C.; Lee, Y. T.; Ni, C.-K. Photodissociation of ethylbenzene and n-propylbenzene in a molecular beam. *J. Chem. Phys.* **2002**, *117* (15), 7034.

(303) Tseng, C.-M.; Lee, Y. T.; Ni, C.-K.; Chang, J.-L. Photodissociation dynamics of the chromophores of the amino acid tyrosine: p-methylphenol, p-ethylphenol, and p-(2-aminoethyl)phenol. *J. Phys. Chem. A* **2007**, *111* (29), 6674–6678.

(304) Dehon, C.; Soorkia, S.; Pedrazzani, M.; Jouvet, C.; Barat, M.; Fayeton, J. A.; Lucas, B. Photofragmentation at 263 nm of small peptides containing tyrosine: the role of the charge transfer on CO. *Phys. Chem. Chem. Phys.* **2013**, *15* (22), 8779–8788.

(305) Mališ, M.; Loquais, Y.; Gloaguen, E.; Biswal, H. S.; Piuzzi, F.; Tardivel, B.; Brenner, V.; Broquier, M.; Jouvet, C.; Mons, M.; et al. Unraveling the mechanisms of nonradiative deactivation in model peptides following photoexcitation of a phenylalanine residue. *J. Am. Chem. Soc.* **2012**, *134* (50), 20340–20351.

(306) Shemesh, D.; Sobolewski, A. L.; Domcke, W. Efficient excited-state deactivation of the

Gly-Phe-Ala tripeptide via an electron-driven proton-transfer process. *J. Am. Chem. Soc.* **2009**, *131* (4), 1374–1375.

(307) Roy, T. K.; Kopysov, V.; Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V.; Gerber, R. B. Conformational Structures of a Decapeptide Validated by First Principles Calculations and Cold Ion Spectroscopy. *ChemPhysChem* **2015**, *16* (7), 1374–1378.

(308) Joshi, K.; Semrouni, D.; Ohanessian, G.; Clavaguéra, C. Structures and IR spectra of the Gramicidin S peptide: pushing the quest for low-energy conformations. *J. Phys. Chem. B* **2012**, *116* (1), 483–490.

(309) Doemer, M.; Guglielmi, M.; Athri, P.; Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V.; Tavernelli, I.; Rothlisberger, U. Assessing the performance of computational methods for the prediction of the ground state structure of a cyclic decapeptide. *Int. J. Quantum Chem.* **2013**, *113* (6), 808–814.

(310) Raskin, I.; Ribnicky, D. M.; Komarnytsky, S.; Ilic, N.; Poulev, A.; Borisjuk, N.; Brinker, A.; Moreno, D. A.; Ripoll, C.; Yakoby, N.; et al. Plants and human health in the twenty-first century. *Trends Biotechnol.* **2002**, *20* (12), 522–531.

(311) Grynkiewicz, G.; Gadzikowska, M. Tropane alkaloids as medicinally useful natural products and their synthetic derivatives as new drugs. *Pharmacol. Rep. 60* (4), 439–463.

(312) Dyakov, Y. A.; Hsu, W. H.; Ni, C.-K.; Tsai, W.-C.; Hu, W.-P. Photodissociation dynamics of benzyl alcohol at 193 nm. *J. Chem. Phys.* **2012**, *137* (6), 064314.

(313) Wang, P.; Bartlett, M. G. Collision-induced dissociation mass spectra of cocaine, and its metabolites and pyrolysis products. *J. Mass Spectrom.* **1998**, *33* (10), 961–967.

(314) Hopkins, J. B.; Powers, D. E.; Smalley, R. E. Vibrational relaxation in jet-cooled alkyl benzenes. I.Absorption spectra. *J. Chem. Phys.* **1980**, *72* (9), 5039.

(315) Balfour, W. J.; Ram, R. S. Spectroscopic studies on toluene and some deuterated derivatives. II. Electronic spectra. *Can. J. Phys.* **1994**, *72* (11-12), 1225–1235.

(316) Granier, S.; Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Weis, W. I.; Kobilka, B. K. Structure of the δ-opioid receptor bound to naltrindole. *Nature* **2012**, *485* (7398), 400–404.

(317) Lord, J. A. H.; Waterfield, A. A.; Hughes, J.; Kosterlitz, H. W. Endogenous opioid peptides: multiple agonists and receptors. *Nature* **1977**, *267* (5611), 495–499.

(318) Waldhoer, M.; Bartlett, S. E.; Whistler, J. L. Opioid receptors. *Annu. Rev. Biochem.* **2004**, *73*, 953–990.

(319) Janecka, A.; Fichna, J.; Janecki, T. Opioid Receptors and their Ligands. *Curr. Top. Med. Chem.* **2004**, *4* (1), 1–17.

(320) Deschamps, J. R.; George, C.; Flippen-Anderson, J. L. Structural studies of opioid peptides: a review of recent progress in x-ray diffraction studies. *Biopolymers* **1996**, *40*, 121–139.

(321) Deschamps, J. R.; Flippen-Anderson, J. L.; George, C. X-ray studies on ligands.

Biopolymers **2002**, *66* (5), 287–293.

(322) Förster, T. 10th Spiers Memorial Lecture. Transfer mechanisms of electronic excitation. *Discuss. Faraday Soc.* **1959**, *27*, 7–17.

(323) Dexter, D. L.; Schulman, J. H. Theory of Concentration Quenching in Inorganic Phosphors. *J. Chem. Phys.* **1954**, *22* (6), 1063.

(324) Danell, A. S.; Parks, J. H. FRET measurements of trapped oligonucleotide duplexes. *Int. J. Mass Spectrom.* **2003**, *229* (1-2), 35–45.

(325) Talbot, F. O.; Rullo, A.; Yao, H.; Jockusch, R. A. Fluorescence resonance energy transfer in gaseous, mass-selected polyproline peptides. *J. Am. Chem. Soc.* **2010**, *132* (45), 16156–16164.

(326) Frankevich, V.; Chagovets, V.; Widjaja, F.; Barylyuk, K.; Yang, Z.; Zenobi, R. Fluorescence resonance energy transfer of gas-phase ions under ultra high vacuum and ambient conditions. *Phys. Chem. Chem. Phys.* **2014**, *16* (19), 8911–8920.

(327) Daly, S.; Poussigue, F.; Simon, A.-L.; MacAleese, L.; Bertorelle, F.; Chirot, F.; Antoine, R.; Dugourd, P. Action-FRET: Probing the Molecular Conformation of Mass-Selected Gas-Phase Peptides with Förster Resonance Energy Transfer Detected by Acceptor-Specific Fragmentation. *Anal. Chem.* **2014**, *86* (17), 8798–8804.

(328) Hendricks, N. G.; Lareau, N. M.; Stow, S. M.; McLean, J. A.; Julian, R. R. Bond-Specific Dissociation Following Excitation Energy Transfer for Distance Constraint Determination in the Gas Phase. *J. Am. Chem. Soc.* **2014**, *136* (38), 13363–13370.

(329) Kulesza, A.; Daly, S.; MacAleese, L.; Antoine, R.; Dugourd, P. Structural exploration and Förster theory modeling for the interpretation of gas-phase FRET measurements: Chromophore-grafted amyloid-β peptides. *J. Chem. Phys.* **2015**, *143* (2), 025101.

(330) KOSTERLITZ, H. W.; LORD, J. A. H.; PATERSON, S. J.; WATERFIELD, A. A. EFFECTS OF CHANGES IN THE STRUCTURE OF ENKEPHALINS AND OF NARCOTIC ANALGESIC DRUGS ON THEIR INTERACTIONS WITH μ- AND δ-RECEPTORS. *Br. J. Pharmacol.* **1980**, *68* (2), 333–342.

(331) Morley, J. S. Structure-activity relationships of enkephalin-like peptides. *Annu. Rev. Pharmacol. Toxicol.* **1980**, *20*, 81–110.

(332) Scholes, G. D. Long-range resonance energy transfer in molecular systems. *Annu. Rev. Phys. Chem.* **2003**, *54*, 57–87.

(333) Grace, L. I.; Cohen, R.; Dunn, T. .; Lubman, D. M.; de Vries, M. S. The R2PI Spectroscopy of Tyrosine: A Vibronic Analysis. *J. Mol. Spectrosc.* **2002**, *215* (2), 204–219.

(334) Horwitz, J.; Strickland, E. H.; Billups, C. Analysis of the vibrational structure in the near-ultraviolet circular dichroism and absorption spectra of tyrosine derivatives and ribonuclease-A at 77.deg.K. *J. Am. Chem. Soc.* **1970**, *92* (7), 2119–2129.

(335) Zabuga, A. V; Kamrath, M. Z.; Rizzo, T. R. Franck-Condon-like Progressions in Infrared Spectra of Biological Molecules. *J. Phys. Chem. A* **2015**.

(336) Horwitz, J.; Strickland, E. H.; Billups, C. Analysis of vibrational structure in the near-ultraviolet circular dichroism and absorption spectra of phenylalanine and its derivatives. *J. Am. Chem. Soc.* **1969**, *91* (1), 184–190.

(337) Forster, T. Energiewanderung und Fluoreszenz. *Naturwissenschaften* **1946**, *33* (6), 166–175.

(338) Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **1948**, *437* (1-2), 55–75.

(339) *Principles of fluorescence spectroscopy*, 3rd ed.; Lakowicz, J. R., Ed.; Springer US: New York, NY, 2006.

(340) *Standardization and Quality Assurance in Fluorescence Measurements I*; Resch-Genger, U., Ed.; Springer Series on Fluorescence; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; Vol. 5.

(341) van der Meer, B. W. Kappa-squared: from nuisance to new sense. *Rev. Mol. Biotechnol.* **2002**, *82* (3), 181–196.

(342) Steinberg, I. Z. Long-range nonradiative transfer of electronic excitation energy in proteins and polypeptides. *Annu. Rev. Biochem.* **1971**, *40*, 83–114.

(343) *Laser Photodissociation and Spectroscopy of Mass-separated Biomolecular Ions*; Polfer, N. C., Dugourd, P., Eds.; Lecture Notes in Chemistry; Springer International Publishing: Cham, 2013.

(344) Platt, J. R. Classification of Spectra of Cata-Condensed Hydrocarbons. *J. Chem. Phys.* **1949**, *17* (5), 484.

(345) Ponder, J. W. Software Tools for Molecular Design, Version 7.1 http://dasher.wustl.edu/tinker.

(346) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.

(347) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †. *J. Phys. Chem. B* **2001**, *105* (28), 6474–6487.

(348) Hickman, C. G.; Gascooke, J. R.; Lawrance, W. D. The S1–S0(1B2–1A1) transition of jet-cooled toluene: Excitation and dispersed fluorescence spectra, fluorescence lifetimes, and intramolecular vibrational energy redistribution. *J. Chem. Phys.* **1996**, *104* (13), 4887.

(349) Lee, Y.; Jung, J.; Kim, B.; Butz, P.; Snoek, L. C.; Kroemer, R. T.; Simons, J. P. Alanyl Side Chain Folding in Phenylalanine: Conformational Assignments through Ultraviolet Rotational Band Contour Analysis. *J. Phys. Chem. A* **2004**, *108* (1), 69–73.

(350) Hashimoto, T.; Takasu, Y.; Yamada, Y.; Ebata, T. Anomalous conformer dependent S1 lifetime of l-phenylalanine. *Chem. Phys. Lett.* **2006**, *421* (1-3), 227–231.

(351) Braslavsky, S. E.; Fron, E.; Rodríguez, H. B.; Román, E. S.; Scholes, G. D.; Schweitzer, G.; Valeur, B.; Wirz, J. Pitfalls and limitations in the practical use of Förster's theory of resonance energy transfer. *Photochem. Photobiol. Sci.* **2008**, *7* (12), 1444–1448.

(352) Baldauf, C.; Rossi, M. Going clean: structure and dynamics of peptides in the gas phase and paths to solvation. *J. Phys. Condens. Matter* **2015**, *27* (49), 493002.

(353) LEE, H.-J.; PARK, H.-M.; LEE, K.-B. CONFORMATIONAL PREFERENCES OF N-ACETYL–GLYCINE–GLYCINE–N'-METHYLAMIDE: A THEORETICAL STUDY. *J. Theor. Comput. Chem.* **2009**, *08* (05), 799–811.

(354) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. Energetics and structure of glycine and alanine based model peptides: Approximate SCC-DFTB, AM1 and PM3 methods in comparison with DFT, HF and MP2 calculations. *Chem. Phys.* **2001**, *263* (2-3), 203–219.

(355) Jalkanen, K. J. Energetics, structures, vibrational frequencies, vibrational absorption, vibrational circular dichroism and Raman intensities of Leu-enkephalin. *J. Phys. Condens. Matter* **2003**, *15* (18), S1823–S1851.

(356) Jalkanen, K. J.; Elstner, M.; Suhai, S. Amino acids and small peptides as building blocks for proteins: comparative theoretical and spectroscopic studies. *J. Mol. Struct. THEOCHEM* **2004**, *675* (1-3), 61–77.

(357) Granovsky, A. A. Firefly, Version 8.0.1 http://classic.chem.msu.su/gran/firefly/index.html.

(358) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98* (45), 11623–11627.

(359) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648.

(360) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58* (8), 1200–1211.

(361) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.

(362) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465.

(363) Burke, N. L.; DeBlase, A. F.; Redwine, J. G.; Hopkins, J. R.; McLuckey, S. A.; Zwier, T. S. Gas-Phase Folding of a Prototypical Protonated Pentapeptide: Spectroscopic Evidence for Formation of a Charge-Stabilized β-Hairpin. *J. Am. Chem. Soc.* **2016**.

Curriculum Vitae

Vladimir Kopysov

Date of birth: 27.11.1987
Place of birth: Glazov, USSR
Citizenship: Russian Federation
E-mail: kvnlinux@gmail.com
Phone: +41789475602
Address: Avenue du Leman 77, Lausanne
Ch-1005, Switzerland

EDUCATION

2011 – 2016	**PhD Thesis "Novel Approaches in Cold Ion Spectroscopy for Structural Characterization of Biomolecules in the Gas Phase"** École Polytechnique Fédérale de Lausanne, Lasusanne Advisor: Dr. MER Oleg V. Boyarkine
2009 – 2011	**M. S. in Applied Mathematics and Physics** Moscow Institute of Physics and Technology, Moscow Thesis: *Study of Absorption and Emission Spectra of Dibenzoylmethanatoboron Difluoride in Solutions* Advisor: Dr. Vyacheslav A. Sazhnikov
2005 – 2009	**B. S. in Applied Mathematics and Physics** Moscow Institute of Physics and Technology, Moscow Thesis: *Spin Label ESR Study of The Complexation of Para- and ………….Diamagnetic Organic Molecules with Cyclodextrins in ………….Polyacrylamide Hydrogels* Advisor: Prof. Vsevolod A. Livshits

PUBLICATIONS

- Kopysov, V.; Makarov, A.; Boyarkin, O. V. Nonstatistical UV Fragmentation of Gas-Phase Peptides Reveals Conformers and Their Structural Features. *J. Phys. Chem. Lett.* **2016**, 1067–1071.
- Kopysov, V.; Boyarkin, O. V. Resonance Energy Transfer Relates the Gas-Phase Structure and Pharmacological Activity of Opioid Peptides. *Angew. Chemie Int. Ed.* **2016**, *55* (2), 689–692.
- Kopysov, V.; Makarov, A.; Boyarkin, O. V. Colors for Molecular Masses: Fusion of Spectroscopy and Mass Spectrometry for Identification of Biomolecules. *Anal. Chem.* **2015**, *87* (9), 4607–4611.
- Roy, T. K.; Kopysov, V.; Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V.; Gerber, R. B. Conformational Structures of a Decapeptide Validated by First Principles Calculations and Cold Ion Spectroscopy. *ChemPhysChem* **2015**, *16* (7), 1374–1378.

- Kopysov, V.; Nagornova, N. S.; Boyarkin, O. V. Identification of Tyrosine-Phosphorylated Peptides Using Cold Ion Spectroscopy. *J. Am. Chem. Soc.* **2014**, *136* (26), 9288–9291.
- Boyarkin, O. V; Kopysov, V. Cryogenically cooled octupole ion trap for spectroscopy of biomolecular ions. *Rev. Sci. Instrum.* **2014**, *85* (3), 033105.
- Sazhnikov, V. A.; Muzafarov, A. M.; Kopysov, V. N.; Aristarkhov, V. M.; Kononevich, Y. N.; Meshkov, I. B.; Voronina, N. V.; Alfimov, M. V. Silica nanoparticles with covalently attached fluorophore as selective analyte-responsive supramolecular chemoreceptors. *Nanotechnologies Russ.* **2012**, *7* (1-2), 6–14.
- Sazhnikov, V. A.; Kopysov, V. N.; Aristarkhov, V. M.; Shibneva, E. S.; Mironchik, A. G.; Fedorenko, E. V.; Alfimov, M. V. Fluorescence properties and conformation of dibenzoylmethanatoboron difluoride in solutions. *High Energy Chem.* **2011**, *45* (6), 501–504.

CONFERENCES & COURSES

2015, September	**Swiss Chemical Society Fall Meeting**
	Poster
2015, June	**63rd Conference on Mass Spectrometry and Allied Topics.**
	Oral presentation
2013, July	**7th Summer School "Mass Spectrometry in Biotechnology & Medicine"**
	Poster
2011, November	**54th Scientific Conference of MIPT**
	Oral presentation
2011, July	**2nd S3 Summer School**
	Theory of gas sensitivity of nano-structured MOX layers
2011, June	**XXV International Chugaev Conference on Coordination Chemistry**
	Poster
2011, April	**International student, postgraduate and young scientist conference "Lomonosov-2011"**
	Oral presentation
2010, September – December	**Courses of spoken French**
	Moscow Institute of Physics and Technology
2010, August	**Courses of French**
	Université de Perpignan Via Domitia
2008, September – December	**Courses of spoken English**
	Moscow Institute of Physics and Technology

AWARDS AND HONORS

2010	Winner of theoretical round of IV Russian Nanotechnology Olympiad

TEACHING AND WORKING EXPERIENCE

2016, Spring	**Teaching assistant**

	for the course "Mass spectrometry"
2015, Spring	**Teaching assistant**
	for the course "Lasers and applications in chemistry"
2014, Spring	**Teaching assistant**
	for the course "Lasers and applications in chemistry"
2013, Fall	**Teaching assistant**
	for the course "Experimental molecular sciences"
2013, Spring	**Teaching assistant**
	for the course "General chemistrhy" (Laboratory Work)
2008 – 2012	**Research Engineer**
	Photochemistry Center RAS, Moscow
2008, February – August	**PERL Developer**
	AGAVA Ltd, Dolgoprudny
2006 – 2008	**Help Desk Technician**
	AGAVA Ltd, Dolgoprudny
2005 – 2006	**Physics & MathsTutor**
	MIPT Physics and Technology Correspondence School, Dolgoprudny

SKILLS

- Foreign Languages – English (C1 Level), French (A2-B1 Level), German (A1 Level)
- Experimental Techniques – Mass spectrometry, Laser spectroscopy, Electron spin resonance
- Scientific Software – MATLAB, Origin, GAMESS-US, Firefly, TINKER, LabView
- Programming Languages – C/C++, Shell, PHP, Perl, Python

HOBBIES

- Powerlifting
- Cross-country skiing