Cryogenic Spectroscopy and Quantum Molecular Dynamics
Determine the Structure of Cyclic Intermediates Involved in Peptide Sequence Scrambling

Oleg Aseev,† Marta A. S. Perez,‡ Ursula Rothlisberger,‡ and Thomas R. Rizzo*†

1Laboratoire de Chimie Physique Moléculaire, École Polytechnique Fédérale de Lausanne, EPFL SB ISIC LCPM, Station 6, CH-1015 Lausanne, Switzerland
‡Laboratory of Computational Chemistry and Biochemistry, École Polytechnique Fédérale de Lausanne, EPFL SB ISIC LCBC, Station 6, CH-1015 Lausanne, Switzerland

Supporting Information

ABSTRACT: Collision-induced dissociation (CID) is a key technique used in mass spectrometry-based peptide sequencing. Collisionally activated peptides undergo statistical dissociation, forming a series of backbone fragment ions that reflect their amino acid (AA) sequence. Some of these fragments may experience a “head-to-tail” cyclization, which after proton migration, can lead to the cyclic structure opening in a different place than the initially formed bond. This process leads to AA sequence scrambling that may hinder sequencing of the initial peptide. Here we combine cryogenic ion spectroscopy and ab initio molecular dynamics simulations to isolate and characterize the precise structures of key intermediates in the scrambling process. The most stable peptide fragments show intriguing symmetric cyclic structures in which the proton is situated on a C2 symmetry axis and forms exceptionally short H-bonds (1.20 Å) with two backbone oxygens. Other nonsymmetric cyclic structures also exist, one of which is protonated on the amide nitrogen, where ring opening is likely to occur.

While tandem mass spectrometry (MS/MS) combined with collision-induced dissociation (CID) is a key method for determining the primary structure of peptides and proteins,1,2 scrambling that can occur during the fragmentation process can lead to errors in the determined sequence. Harrison and co-workers first proposed that peptide sequence scrambling occurs via cyclization of nascent b_n fragment ions followed by opening of the cycle at places different from where it was originally formed.3,4 Such a mechanism relies upon the mobility of the proton upon ion activation, a subject that is now well established.5-10 A number of different experimental and theoretical methods have been used to address the question of peptide sequence scrambling caused by proton mobility: multistage mass spectrometry (MS),11-19 hydrogen-deuterium exchange (HDX),20-24 ion mobility spectrometry (IMS),25-28 infrared spectroscopy (IR),16,20,21,29-41 and density functional theory calculations (DFT).36,37,42-46 Of these, IR spectroscopy is one of the most powerful methods, as it can provide detailed vibrational signatures of key intermediates, yielding insight into their geometry and protonation site. Most spectroscopic studies of b_n ions have been performed using infrared multiple photon dissociation (IRMPD) with free electron lasers. While this approach can provide some structural information and is able to identify cyclic b_n ions, it has important limitations. Because IRMPD experiments have been done at room temperature, the bands are broad, and spectral features of individual conformational isomers cannot be resolved. Moreover, the nonlinear nature of the IRMPD process distorts spectral intensities. Together, these complicating factors make it difficult to analyze the spectra and assign them with confidence to computed structures.

On the theoretical side, the computations necessary to interpret the spectroscopic data typically include classical molecular dynamics (MD) for conformational searching. This has the limitation that one must guess the protonation sites a priori, which provides no insight into the mobility of the excess proton.

Here we report our work combining IR-UV double-resonance spectroscopy in a cryogenic ion trap with ab initio molecular dynamics (AIMD) simulations based on density functional theory to overcome these limitations and reveal detailed structural information on b_n ions produced by collision-induced dissociation. Cooling in a low temperature ion trap eliminates spectral congestion, and combined with IR-UV double resonance this allows the measurement of linear, high resolution, infrared spectra of individual conformers of the fragments produced by CID. The cooling process can trap intermediates that might otherwise be不稳定 at room temperature, helping to elucidate the scrambling mechanism. Moreover, rather than making assumptions about the protonation site, AIMD can help determine the most stable protonation sites concurrently with conformational exploration.

Received: May 25, 2015
Accepted: June 15, 2015
Published: June 15, 2015
We have applied these state-of-the-art techniques to investigate b<sub>6</sub> ions produced from the peptide [FAGFAGPG + H]+. The choice of this molecule was motivated largely by the work of Tirado and Pfeifer, who studied b<sub>6</sub> ions from the analogous species [YAGYAGPG + H]+ using IRMPD. While they clearly identified the presence of a cyclic species, the IRMPD technique they employed is not conformation selective, and as a result it is difficult to determine the structure of the b<sub>6</sub> fragment. We report here high-resolution, conformation-selective IR spectra, which, in combination with AIMD simulations, reveal intriguing totally symmetric cyclic structures for the most stable conformers of this peptide fragment, which are key intermediates in the sequence scrambling process. In addition to the symmetric cyclic species, we have also identified the higher energy, N-protonated, cyclic species that has been proposed to lead to ring opening. Finally, using <sup>15</sup>N substitution, we demonstrate the mobility of the proton between equivalent sites of the N-protonated species.

All spectroscopic experiments described here were conducted using our home-built tandem photofragment mass spectrometer described in detail elsewhere. The parent peptide, [FAGFAGPG + H]+ is dissolved in a solution containing 50/50/0.1 water/methanol/acetic acid to a concentration of 50–100 μM. A nanoelectrospray source is used to produce the gas-phase protonated peptide, which is subsequently introduced into the instrument through a metal capillary. The ions coming from the capillary are collected and focused by an RF ion funnel. We use in-source dissociation through collisions with the ambient gas to fragment the parent ions by applying an acceleration voltage on the two last electrodes of the ion funnel. The CID fragment ions are then guided together with the parent ions through a hexapole and directed to the first quadrupole mass filter, where the b<sub>6</sub> fragments of interest are selected by their mass-to-charge ratio. The preselected fragment ions are deflected 90° by an electrostatic quadrupole bender and enter an octupole ion guide. A second quadrupole bender located downstream from the octupole diverts the ions 90° and introduces them into a cryogenic linear multipole RF ion trap, which is maintained at a temperature of ∼4 K by a closed cycle helium cryostat.

The cold, trapped ions are interrogated by lasers using IR-UV double-resonance photofragment spectroscopy. We first record an ultraviolet photofragment excitation spectrum by monitoring a particular photofragment ion signal while scanning the wavelength of UV laser, which is the frequency-doubled output of a Nd:YAG pumped dye laser. An IR spectrum is then recorded by monitoring the UV-induced photofragment signal at a particular m/z while scanning an IR pulse from a Nd:YAG-pumped OPO, which is introduced 200 ns before the UV laser pulse. When the ions absorb an IR photon they are removed from the initial state involved in the electronic transition and we observe a dip in the UV-induced photofragment signal. Since different conformers have slightly different UV spectra, it allows recording conformation-selective IR spectra by tuning the UV laser to individual, resolved UV bands.

Calculated three-dimensional structures of the conformers were determined from multiple simulated annealing (SA) runs using ab initio molecular dynamics (AIMD) based on density functional theory (DFT). The molecules are heated to high temperatures (~1400K) to explore a large conformational space and then slowly cooled. These SA-AIMD runs are performed using two different software packages: (1) Terachem<sup>49–51</sup> with the functional B3LYP<sup>52</sup> and a 6-31G basis set<sup>53</sup> employing Grimme’s D3 dispersion corrections<sup>54</sup> and (2) CPMD<sup>55</sup> using the functional BLYP<sup>56</sup> and dispersion-corrected atom-centered potentials<sup>57</sup> (DCACPs).

Optimized geometries and harmonic vibrational frequencies of the resulting structures are calculated at the DFT M11/6-31G(d,p)<sup>58</sup> level using Gaussian09 v D.01<sup>59</sup> with tight convergence criteria and an ultrafine grid. The total energies and zero point energy (ZPE) corrections are calculated at the same level. Before comparing the calculated spectra with the experimental results, the frequencies are scaled by factors of 0.946 and 0.963 in the amide NH stretch (3200–3500 cm<sup>−1</sup>) and fingerprint (1400–1800 cm<sup>−1</sup>) regions, respectively. This procedure allowed us to find the structures of most conformers. However, finding the structure of conformer A required additional simulations in which we used the spectral information indicating that the molecule is protonated at one of the phenylalanine nitrogens. SA-AIMD runs were performed for the cyclic structure with N-protonation imposed as a structural constraint in Terachem<sup>51</sup> and a restraint in CPMD.<sup>55</sup> More than 200 different candidate structures were explored in this way.

Five unique conformers have been identified for the b<sub>6</sub> - [FAGFAGPG + H]+ fragment ion under our experimental conditions using IR–UV double-resonance spectroscopy. Figure 1 shows IR spectra in the X–H (X = N, C) stretching region that are representative of the two major classes of conformers. Conformer A has five narrow bands in the higher-energy amide NH stretch region and a broad band around 3050 cm<sup>−1</sup>. The total number of bands in this case corresponds to the number of amino acid residues in the molecule—one for each amide NH. Conformer C, however, shows only three narrow bands in the presented region. To understand the apparent missing bands, we use <sup>15</sup>N isotopic substitution at specific individual residues of the molecule. Isotopic labeling not only helps locate missing spectral bands, but it allows us to assign them to individual amino acid residues, which is necessary for validating calculated spectra and structures.

Figure 2 presents IR spectra of conformer C with natural <sup>14</sup>N isotopic abundance (solid black curves) together with those of selectively <sup>15</sup>N-labeled species (blue-filled spectra). In each case where two matching amino acids (two Phe, two Ala, or two Gly) are labeled at the same time (Figure 2a,c,d), a single
spectral band exhibits a redshift of $7 - 9\ cm^{-1}$. Figure 2b shows a spectrum in which one Phe residue is labeled, and Figure 2e one in which one Ala and one Gly residue are both labeled. The bands that shift upon substitution of two matching residues (Figures 2a,c,d) now split into two. This indicates that each of the three bands of conformer C is degenerate, and this degeneracy is lifted upon isotopic substitution of matching residues. For conformer C to contain three pairs of equivalent NH groups it must be symmetric, and this is only possible if the molecule adopts a cyclic structure. Moreover, the symmetry of the molecule necessarily implies that the added proton is located at the center of symmetry.

Figure 3 demonstrates that conformer A does not exhibit such symmetry, since all bands in its spectrum occur at different positions. As before, we assign these bands using double isotopic substitution of matching residues (Figure 3a,c,d). In order to distinguish between these residues, we also record spectra for the species where only one of each matching pair is labeled (Figure 3b,e). Somewhat to our surprise, the bands in the latter case split into two instead of simply shifting, resulting in an increase in the total number of bands beyond the number of NH oscillators. This must imply that our sample contains a mixture of species with the same geometry but with the isotopic label on each of the matching pair of residues. Since we know on which residue the isotopic label is initially incorporated, and since it is impossible for the $^{15}\text{N}$ to migrate, the only possible explanation for the spectra of Figure 3b,e is that the proton has transferred between matching AA residues on opposite sides of the molecule. This represents direct spectroscopic evidence for a mobile proton. Moreover, as shown in the Supporting Information, this proton mobility is also nicely captured by our AIMD simulations (Movie S1 in the Supporting Information).

The broad bands in the region of 3050 $cm^{-1}$ (Figure 1a) are characteristic of a protonated amide, and isotopic labeling experiments indicate that these bands correspond to the Phe residue (Figure S2 in the Supporting Information). This indicates that conformer A is N-protonated on phenylalanine.

To find the exact three-dimensional structures of the conformers, we performed multiple simulated annealing runs combined with ab initio molecular dynamics. The calculated spectra of the lowest energy structures of the two classes of conformers found in this way are shown in Figure 4 in comparison with the experimental spectra. The assignments of the NH stretch bands in the computed spectra fully correspond to the assignments determined experimentally with $^{15}\text{N}$ isotopic substitution on individual residues. The good agreement between experiment and theory provides a high degree of confidence in the determined structures.

These calculated structures and their ZPE-corrected energies are shown in Figure 5. Based on the calculations, the most favored structure corresponds to conformer C, whereas conformer A is higher in energy by 14.9 kcal/mol. It is important to notice that the structure of conformer A corresponds to the lowest energy N-protonated structure we found. All noncyclic structures observed during SA-AIMD runs lie at least 30 kcal/mol higher in energy relative to C and correspond to N-terminal protonated oxazolone structures, which is consistent with the fact that we do not observe any
bands in the 1800−2000 cm⁻¹ region where the oxazolone CO stretches are expected.  

The spectral band around 1600 cm⁻¹ that appears for conformer C is attributed to a shared proton motion coupled with various amide II vibrations. With regard to the projections on the right side in Figure 5, in this motion the proton oscillates roughly perpendicularly to the axis connecting two oxygens and lying in the plane of the page. The symmetric structure that we determine for conformer C is similar to that suggested by Hernandez et al. in their IRMPD study of the all alanine b₆ ions produced by CID from AAAAAMA. However, the low resolution of their spectra, the lack of conformer selectivity, and the relatively poor agreement between experiment and theory makes their assignment of a symmetric cyclic structure tenuous. The high-resolution, conformer specific spectra measured in this present work provides a solid basis for our assignment of a symmetric, cyclic species.

The combination of IR−UV double-resonance spectroscopy in a cryogenic ion trap and ab initio molecular dynamics simulations allows the characterization of individual conformers of CID fragments as demonstrated by our application to the b₆ ion produced from the peptide [FAGFAGPG + H]⁺. The lowest energy conformer of this fragment ion (C) consists of a highly symmetric cyclic structure in which the proton sits on a C₂ axis and is equally shared between two carbonyl oxygens with ultrashort H-bond distances (1.20 Å). Moreover, although it is generally accepted that cyclic b₆ fragments are more likely protonated at a carbonyl oxygen atom, we have isolated and identified a higher energy cyclic conformer (A) with the protonation situated on the amide nitrogen of phenylalanine. This structure is believed to be involved in a cycle reopening reaction, giving rise to sequence scrambling. Indeed our AIMD simulations performed at high temperature (see Movie S2 in the Supporting Information) show precisely such a ring opening reaction to form an oxazoline structure.

Nitrogen-15 isotopic substitution studies of the non-symmetric conformer shows clear spectroscopic evidence that the proton moves between matched pairs of residues on opposite sides of the cyclic molecule. Given the stability of the symmetric conformer, it may well serve as an intermediate in the proton transfer process between matching residues.

**REFERENCES**

Infrared Multiple-Photon Dissociation Spectroscopy and Gas-Phase Peptide Fragment Size on the Propensity of Cyclization in Collision-Spectrom.


Ions Generated from Doubly Protonated Tryptic Peptides?

The Journal of Physical Chemistry Letters

Structure in Anionic b-Type Peptide Fragments.

Spectrom.


Harrison, A. G.

Why Are B Ions Stable Species in Peptide Spectra?

The Journal of Physical Chemistry Letters


Combined Ion Mobility/Mass Spectrometry.

J. Am. Soc. Mass

and Integrated Ion Mobility/Tandem Mass Spectrometry.

Peptide a- and b-Type Fragment Ions Using Stable Isotope Labeling an

dent Reactions of


