Mechanistic Insights into Intramolecular Gas-Phase Crosslinking of Peptides and Carbene or Nitrile Imine Intermediates

Hongyi Zhu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:
František Tureček, Chair
Bo Zhang
Matthew F. Bush

Program Authorized to Offer Degree:
Chemistry

University of Washington

Abstract

Mechanistic Insights into Intramolecular Gas-Phase Crosslinking of Peptides and Carbene or Nitrile Imine Intermediates

Hongyi Zhu

Chair of the Supervisory Committee:
František Tureček
Department of Chemistry

This dissertation presents a comprehensive study on the use of photochemical crosslinking, advanced tandem mass spectrometry, and computational simulations to investigate noncovalent interactions and crosslinking behaviors in gas-phase peptide ions. The work combines experimental and theoretical approaches to analyze how different crosslinking chemistries, protonation sites, amino acid compositions, and scaffold stereochemistries affect peptide structure and reactivity, providing a new framework for understanding the fundamental interactions of intramolecular peptide sequences with various phototags in gas-phase environments, advancing the application of mass spectrometry as a powerful tool for biomolecular structural analysis.

Chapter 1 introduces the principles of mass spectrometry, including ionization, mass analysis, and detection methods, enabling the characterization of biomolecules with minimal sample preparation. This chapter also discusses photochemical crosslinking as a method for probing noncovalent interactions, focusing on the role of reactive intermediates such as nitrenes, carbenes, and nitrile imines, providing insights into three-dimensional structures and molecular interactions, particularly for complex biomolecular assemblies.

In Chapter 2, the efficacy of carbene crosslinkers activated by diazirine under 355 nm is examined using peptide scaffolds (s-LAAG, s-ALAG, and s-AALG). Crosslinking yields were consistent across these scaffolds, demonstrating a robust crosslinking mechanism that is independent of the sequence order of alanine

and leucine residues. Hydrogen-deuterium exchange, carboxyl C-terminus blocking, and analysis of CID-MS^n spectra of reference synthetic products revealed that a significant fraction of crosslinks involved the Gly amide and carboxyl groups. Contact analysis of long Born-Oppenheimer molecular dynamics (BOMD) trajectories was used to count close contacts between the incipient carbene and peptide atoms, and further provided insights into the thermal behavior of peptide ions, validating the s-AALG scaffold as a robust model for carbene crosslinking studies in hydrophobic environments.

Chapter 3 extends the analysis to peptide sequences containing basic residues: proline and histidine, within s-AAPG and s-AAHG scaffolds. Crosslinking yields were significantly reduced in these peptides, attributed to the unique protonation and conformational characteristics of the basic residues. Detailed BOMD simulations revealed that proline forms a stable hydrogen-bond network and affects crosslinking patterns, while histidine's protonation at the imidazole group introduces structural constraints. These findings underscore the role of amino acid basicity and protonation in governing crosslinking efficiency and specificity.

In Chapter 4, nitrile imine intermediates generated by photodissociation of tetrazole-tagged peptide conjugates on cyclohexane scaffolds with distinct stereochemistry (*cis*-1,2- and *trans*-1,4-cyclohexane) is introduced. Despite expected steric hindrance for the trans-1,4 configuration, crosslinking occurs in both scaffold types, challenging conventional stereochemical assumptions. High-resolution cyclic ion mobility mass spectrometry, BOMD, and density functional theory (DFT) calculations allow us to match theoretical and experimental collision cross sections, providing insights into stereochemical effects on crosslinking yields and attachment sites. These findings reveal that the flexibility of the *cis* scaffold facilitates interactions, while the *trans* scaffold imposes conformational constraints that affect crosslinking efficiency.

Chapter 5 focuses on the reactivity of peptides with C-terminal lysine or arginine residues when crosslinked with nitrile imine intermediates. Experiments showed that lysine primarily crosslinks through the carboxyl group, while arginine's guanidine group exhibits unique reactivity, likely due to proton transfer steps that precede C–N bond formation. Computational Gibbs energy calculations reveal this reaction as an endothermic proton transfer followed by an exothermic bond formation step, highlighting a novel mechanism in gas-phase peptide ion chemistry that has not been observed in condensed phases. Ion mobility data allowed for accurate comparisons of experimental and theoretical collision cross sections, offering a deeper understanding of ion structures and reactivity.

Acknowledgements

First, I would like to express my gratitude to my research advisor, Prof. František Tureček, for your invaluable guidance, mentorship, and unwavering support. I am truly grateful that you accepted me into your group in my second year and provided me with the opportunity to work on fascinating research in such an encouraging and inspiring environment. Thank you for taking a chance on me, for providing me with opportunities to work on innovative projects, and for fostering a space where I could grow both academically and personally. I am also grateful for the opportunity you provided to conduct research in the Czech Republic, as well as for the chance to attend conferences, where I was able to connect with other researchers worldwide and gain new perspectives. Thank you for showing me what it means to be an outstanding mentor, both through your academic achievements and your dedication to student development.

I would like to extend my thanks to my undergraduate advisor, Prof. Hui Chao for introducing me to the world of research, and Kangqiang Qiu for patiently guiding me through the fundamentals of research when I was just beginning. Your mentorship and support were instrumental in shaping my academic journey.

I am fortunate to have worked with incredible colleagues and collaborators. Thank you, Martin Sadilek, for your training and for organizing regional meetings; Shu R. Huang, for teaching me solid-phase peptide synthesis and mass spectrometry; Yue Liu, for guiding me in computational analysis; and Václav Zima, for assistance with challenging organic synthesis. I feel so lucky to have Jiahao Wan, Kim Vu, and Emily R. Ding as supportive colleagues and friends in the group. I would like to thank Prof. Karel Lemr for the wonderful collaboration in Olomouc, and Marianna Nytka for her guidance on cyclic ion mobility.

Finally, I would like to thank my committee members, Prof. Matthew F. Bush, Prof. Bo Zhang, and my Graduate School Representative, Prof. Miklos Guttman, for their time, support, and valuable feedback on my research, which has been essential to my growth as a researcher.

DEDICATION

This dissertation is dedicated to my husband, Haoyang Wu for your love, encouragement and support throughout my Ph.D. journey. To my furry family members, Baby Sproutie and Tea Sproutie for your comforting presence and companionship. To my parents and grandparents, thank you for your nurturing, guidance, and boundless love since the day I was born.

Contents

List of Figures

List of Schemes

List of Tables

Chapter 1

Introduction

1.1 Photochemical Crosslinking for Probing Noncovalent Interactions

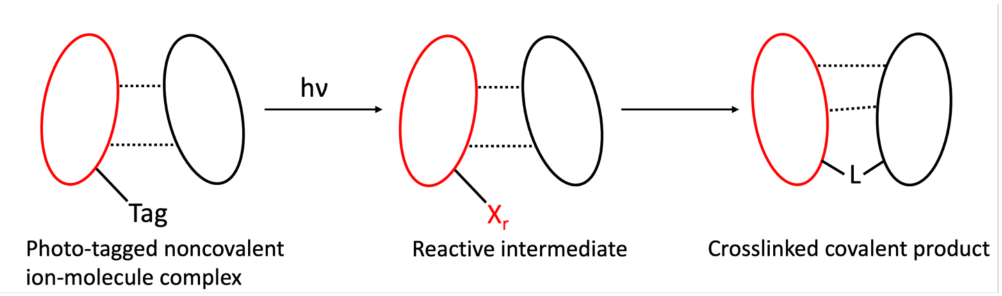
Chemical crosslinking relies on the formation of covalent bonds between components of a noncovalent complex or between different sites in a single large molecule.[1][2] With the introduction of photoactivated[2] and photodissociative[3]-[5] crosslinkers, it became possible to generate highly reactive, short-lived intermediates to identify contacts between reactive groups and sites on the target molecule. A traditional crosslinker contains two reactive groups connected by a spacer arm, which can be adjusted based on the target. When two sites within a noncovalent complex are crosslinked, the distance between them is defined by the length of the spacer arm, providing valuable distance constraints for reconstructing the three-dimensional structure of the complex.

1.1.1 Noncovalent Interactions

Noncovalent interactions are ubiquitous in nature and play a significant role in the cohesion of chemical systems. First identified by J.D. van der Waals, noncovalent interactions are cohesive forces that are distinct from and weaker than covalent bonds that hold atoms or molecules together.[6][7] P.A. Kollman later refined the definition, characterizing them as interactions where electrons remain paired with no net change in bonding.[8] Their strength typically ranges from -0.5 to -50 kcal mol^{-1}[9], arising from electrostatic forces, exchange-repulsion, dispersion, polarization, and charge transfer components.[10]-[12] Although weaker than covalent bonds, noncovalent interactions collectively exert a substantial influence on synthesis, catalysis, and material design. They are essential for maintaining the three-dimensional structures of biomolecules, such as proteins and nucleic acids, and are crucial for their binding activities and biochemical functions. Noncovalent interactions also define the tertiary and quaternary structures of biomolecules like DNA, RNA, proteins, and carbohydrates.[13] Hydrogen bonding and π-π interactions are key contributors to these structures, though their relative strengths are still under investigation. Beyond structural integrity, noncovalent interactions govern molecular recognition processes, including protein-protein interactions, protein-DNA interactions, and ligand binding.[13] These interactions are fundamental to biological processes and also have practical implications for drug design, crystallinity, and self-assembly.

Various methods are employed to study noncovalent interactions. High-resolution techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) reveal static and dynamic structures, though they require high sample purity and are often challenging to perform. Other methods including analytical ultracentrifugation, isothermal titration calorimetry, fluorescence spectroscopy, and surface plasmon resonance can provide insights into binding affinities and subunit interactions.[14] Additionally, affinity purification, chemical crosslinking, and fluorescence resonance energy transfer (FRET) are widely used to probe protein interactions before detailed structural studies are undertaken.

Mass spectrometry has become indispensable for studying noncovalent interactions due to the development of soft ionization techniques, such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). These methods require minimal sample quantities and can be performed without extensive purification. Native mass spectrometry preserves noncovalent interactions in the gas phase, providing stoichiometric information about complex composition. Additional tools, such as ion mobility mass spectrometry (IM-MS), hydrogen-deuterium exchange, and hydroxyl radical labeling, offer further insights into protein folding and dynamics. Crosslinking mass spectrometry specializes in site-specific interaction studies and can even extend to in vivo analyses. While mass spectrometry-based methods do not achieve the atomic resolution of X-ray crystallography or NMR, they offer higher throughput and are particularly effective for weak or transient complexes. Computational modeling complements mass spectrometry by refining experimental data and reconstructing detailed three-dimensional structures of complexes.[15] When experimental data is unavailable, bioinformatic techniques are also employed to predict interaction interfaces.

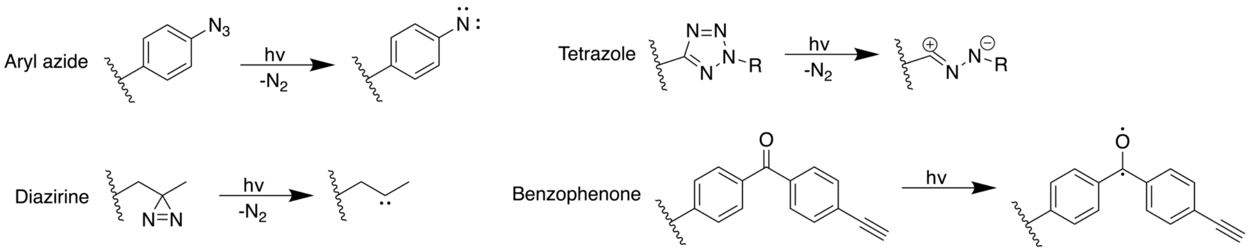


Scheme 1.1: Photochemical crosslinking illustration.

1.1.2 Photochemical Crosslinking and Crosslinking Reagents

Crosslinking has been employed as an indirect method to map noncovalent interactions in protein and peptide complexes. Among the variety of crosslinking strategies,[16] one particular interest here concerns reactive intermediates produced transiently by photodissociation of stable functional groups.

The functional group is introduced as a tag into one of the components where upon photodissociation it produces a transient intermediate (X_r) that reacts spontaneously by forming a covalent bond at a sterically accessible position (Scheme 1.1). The detected bond formation is used to assign point-to-point distance constraints that can provide information on the spatial arrangement of the reacting regions in the biomolecule or noncovalent complex. The reactive intermediates that have been employed frequently include nitrenes[4], carbenes[2][17], and nitrile imines[18] produced by photodissociation, as well as the triplet states formed by photoexcitation in benzophenone-tagged proteins (Scheme 1.2).[19] Among these photocrosslinkers, aryl azides are easy to synthesize but require photon activation at wavelength of < 300 nm which can denature proteins[20]; benzophenones are easy to handle and are resistant to water deactivation but the crosslinker size can sterically alter the non-covalent interactions within the complexes.[21] Stable carbene sources, such as diazirines[22]-[24] and diazoalkanes[25] are better suited for the purpose of our study mainly because of its activation wavelength (330-370 nm) that is outside the absorption by naturally occurring biomolecules, chemical stability and its small steric hindrance.[2] So they have been used extensively in photoaffinity labeling and foot printing studies of various biomolecules.[26]-[28] However, these conventional photocrosslinker have problems like extremely short half-lives, thus resulting in very low target capturing yields and high background.[29] The emerging tetrazole as a photocrosslinker overcomes this issue. Readily available by thermolysis or photolysis of 2,5-diaryl tetrazoles which can be readily introduced as tags into peptides or



Scheme 1.2: Reactive intermediates produced by corresponding photocrosslinkers.

other biomolecules[30], nitrile imines[31] have been suggested for derivatization of peptides and proteins88 via the well known [3 + 2] cycloadditions with dipolarophiles.[32] 2,5-Diphenyltetrazole has a major absorption band with a λ_{max} at 255 nm and $\epsilon = 2 \times 10^4$ L mol^{-1} cm^{-1} that was measured by Blanksby, Trevitt, et al. in gas-phase ions.[33] This makes tetrazole a strong chromophore for photodissociation. The products have been identified by spectroscopic analysis,[34][35] such as EPR spectroscopy,[36] and the reaction kinetics has been studied by laser flash photolysis,[37] analysis of kinetic isotope effects,[38] and quantum mechanical tunneling. The insertion reactions have been found to have very low activation energies.[35] Compared with other methods using chemical reagents, the principal advantage of photochemical crosslinking is that, in contrast to X-ray and NMR, it does not require purified substrates, and can be carried out in a variety of environments such as lipid bilayers and living cells.[39]

1.2 A Brief Introduction to Mass Spectrometry

Mass spectrometry is an analytical technique that can provide both qualitative (structure) and quantitative (molecular mass or concentration) information on analyte molecules after their conversion to ions. The molecules of interest are first introduced into the ionization source of the mass spectrometer, where they are first ionised to acquire positive or negative charges. The ions then travel through the mass analyzer and arrive at different parts of the detector according to their mass/charge (*m/z*) ratio. After the ions make contact with the detector, useable signals are generated and recorded by a computer system. The computer displays the signals graphically as a mass spectrum showing the relative abundance of the signals according to their *m/z* ratio.

1.2.1 Ionization Source

Different ionization techniques have been developed to optimally ionize molecules of different characteristics such as polarity, volatility, thermal lability, stability, and size. Experimental goals also influence the ionization method chosen.

Electron impact (EI) is an ionization method used for samples amenable to gas-phase analysis due to their thermally stable and relatively low molecular weight. For this reason, samples are usually introduced to the ion source for EI after gas chromatography (GC) separation or from a solids probe. An EI source uses a filament set to about 70 eV to create a stream of high-energy electrons that interact with the gas phase sample molecules. Ionization occurs when the collision removes an electron from the sample molecule, creating predominantly singly charged positive ions. Because EI is a high-energy process, it cleaves covalent bonds, producing repeatable fragmentation that can be used to identify compounds using mass spectral libraries. EI is described as a "hard" ionization method since the analytes often undergo fragmentation during the ionization process.

Complementary to EI, chemical ionization (CI) is used to ionize molecules that would fragment excessively by EI, or to ionize molecules without fragmentation to produce a molecular ion that can be used to determine the molecular weights of sample components. In CI, a reagent gas such as methane, isobutene, or ammonia is introduced to the ion source, where it is ionized by the filament. The ionized gas interacts with the sample, which is subsequently ionized by reactions with reagent gas ions, creating singly charged sample components. Because reagent gas is introduced at high concentration relative to the sample, most of the ionization of the sample occurs by CI rather than EI. Depending on the sample molecule and the reagent gas, ionization reactions include proton transfer, proton abstraction, and adduct formation. Compared to EI, CI is a "soft" ionization technique because the reagent gas reactions substantially reduce the energy absorbed by sample molecules, producing substantially less fragmentation or predominantly molecular ions.

Developed by John Fenn in 1984, another "soft" ionization techenique electrospray ionisation mass spectrometry (ESI-MS) has emerged in clinical laboratories.[40] It provides a sensitive, robust, and reliable tool for studying non-volatile and thermally labile bio-molecules that are not amenable to analysis by other conventional techniques. ESI uses electrical energy to assist the transfer of ions from solution into the gaseous phase before they are subjected to mass spectrometric analysis. Ionic species in solution can thus

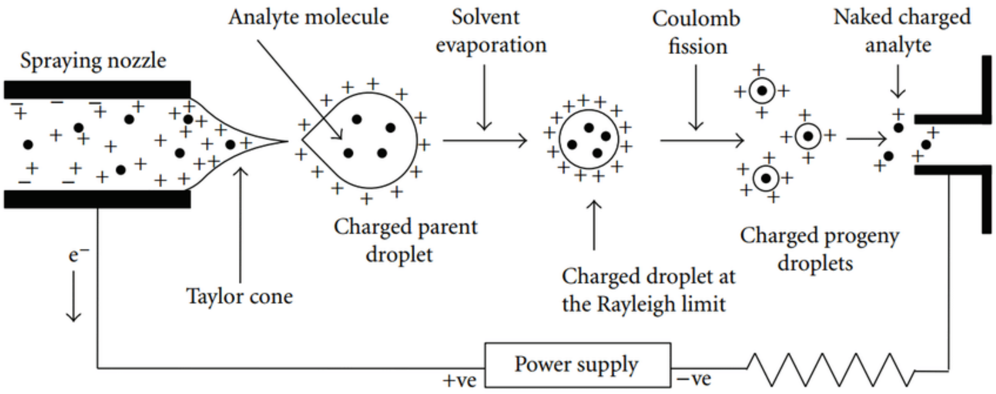


Figure 1.1: Mechanism of electrospray ionization. Image source: Figure 6 from DOI: 10.1155/2012/282574

be analysed by ESI-MS with increased sensitivity. Neutral compounds can also be converted to ionic form in solution or in gaseous phase by protonation or cationisation (e.g. metal cationisation), and hence can be studied by ESI-MS. The transfer of ionic species from solution into the gas phase by ESI involves three steps: (1) dispersal of a fine spray of charge droplets, followed by (2) solvent evaporation and (3) ion ejection from the highly charged droplets (Figure 1.1) which is maintained at a high voltage (2.5 - 6.0 kV) relative to the wall of the surrounding chamber.[41] A mist of highly charged droplets with the same polarity as the capillary voltage is generated. The application of a nebulising gas (e.g. nitrogen), which shears around the eluted sample solution, enhances a higher sample flow rate. The charged droplets, generated at the exit of the electrospray tip, pass down a pressure gradient and potential gradient toward the analyser region of the mass spectrometer. With the aid of an elevated ESI-source temperature and/or another stream of nitrogen drying gas, the charged droplets are continuously reduced in size by evaporation of the solvent, leading to an increase of surface charge density and a decrease of the droplet radius. Finally, the electric field strength within the charged droplet reaches a critical point at which it is kinetically and energetically possible for ions at the surface of the droplets to be ejected into the gaseous phase. The emitted ions are sampled by a sampling skimmer cone and are then accelerated into the mass analyser for subsequent analysis of molecular mass and measurement of ion intensity.

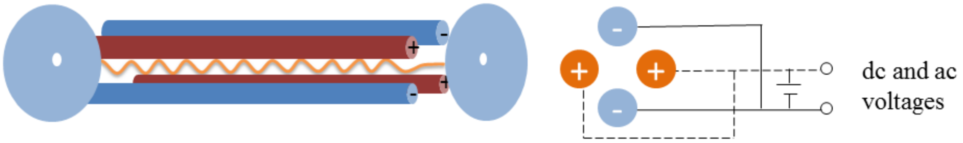


Figure 1.2: Schematic of a quadrupole mass analyzer. Image source: https://www.creative-proteomics.com/blog/index.php/several-types-of-mass-analyzer/

1.2.2 Mass Analyzers and Ion Trap

The mass analyzer is the heart of the mass spectrometer, which takes ionized masses and separates them based on mass to charge ratios. There are many types of mass analyzers that can accomplish the task of distinguishing and separating ions of different mass-to-charge ratios.

1.2.2.1 Quadrupole

The quadrupole mass filter is one of the most commonly used mass analyzers. It consists of four parallel metal rods and each opposing rod pair is connected together electrically (Figure 1.2). One pair of raids is applied with a radio frequency (RF) voltage while another one is applied with a direct current (DC) voltage. At a given DC and RF combination, only the ions of a particular *m/z* show a stable trajectory and can be transmitted to the detector, while other ions with unstable trajectories don't pass the road, because the amplitude of their oscillation becomes infinite. By changing DC and RF in time which usually at a fixed ratio, ions with different *m/z* values can be transmitted to the detector one after another to obtain a mass spectrum.[42]

1.2.2.2 Quadrupole 3-D Ion Trap

Ion traps are modified quadrupole devices with additional functionality of ion trapping. The quadrupole 3-D ion trap is about the size of a small fist and consists of a ring electrode and two hyperbolic end electrodes.[43] It operates at relatively high pressure (10^{-3} torr) with a helium buffer gas that assists the ions to maintain a stable orbital frequency. The buffer gas also serves as the collision gas for collision-induced dissociation (CID) during MS/MS experiments. Ions may be created inside the quadrupole ion trap or, more commonly, externally. An oscillating saddle field inside the trapping volume contains and focuses the ions into the center of the trap. From here the operator can scan the ions out of the trap to create a classic full mass spectral scan of the ions in the trap. Alternatively, a particular ion can be selected (isolated), collisionally fragmented and a scan of all the product ions generated (MS^2 scan). This whole process can be repeated with any one of these fragment ions (MS^3 scan) and as long as there are sufficient ions remaining in the trap to provide an adequate signal-to-noise ratio (S/N), the process can be repeated (Figure 1.3). The number of ions that can be retained in the quadrupole ion trap, or indeed in any trapping-type instrument, is limited by space charging effects. Space charging occurs when the cloud of ions becomes sufficiently dense that coulombic repulsion between the like-charged ions starts to overcome the trapping potential, resulting in degraded mass resolution and accuracy. Limiting the number of ions in the trap at any one time normally controls this effect.

1.2.2.3 Time-of-Flight

Time-of-flight (ToF) mass analyzers operate on a different principle than quadrupole ion traps, determining the *m/z* ratio of ions by measuring their flight time. Ions are accelerated by an electric field of known strength, and their separation occurs based on the time they take to traverse a flight tube of fixed length and reach the detector.[44] The ion trajectory depends on its momentum, kinetic energy, and *m/z* ratio. According to classical physics, ions with lower *m/z* travel faster and arrive at the detector sooner, while those with higher *m/z* arrive later. ToF analyzers are compact and moderately priced, making them a viable alternative to magnetic sector and quadrupole analyzers, especially given their speed and sensitivity. They are known for excellent mass accuracy, straightforward calibration, and the highest practical mass range among mass analyzers. However, the speed of the digitizer can limit the dynamic range of the instrument. With fast acquisition rates, ToF analyzers are ideal for rapid gas chromatography (GC) separations, enabling better resolution of coeluting components compared to slower analyzers like quadrupoles.

1.2.2.4 Ion Mobility Spectrometry

In ion mobility spectrometry (IMS), packets of analyte ions travel through a gas-filled “drift tube” under the influence of a uniform electric field and their arrival time at a detector is recorded. The relationship between the drift velocity of ionic species and the strength of the applied electric field is governed by the

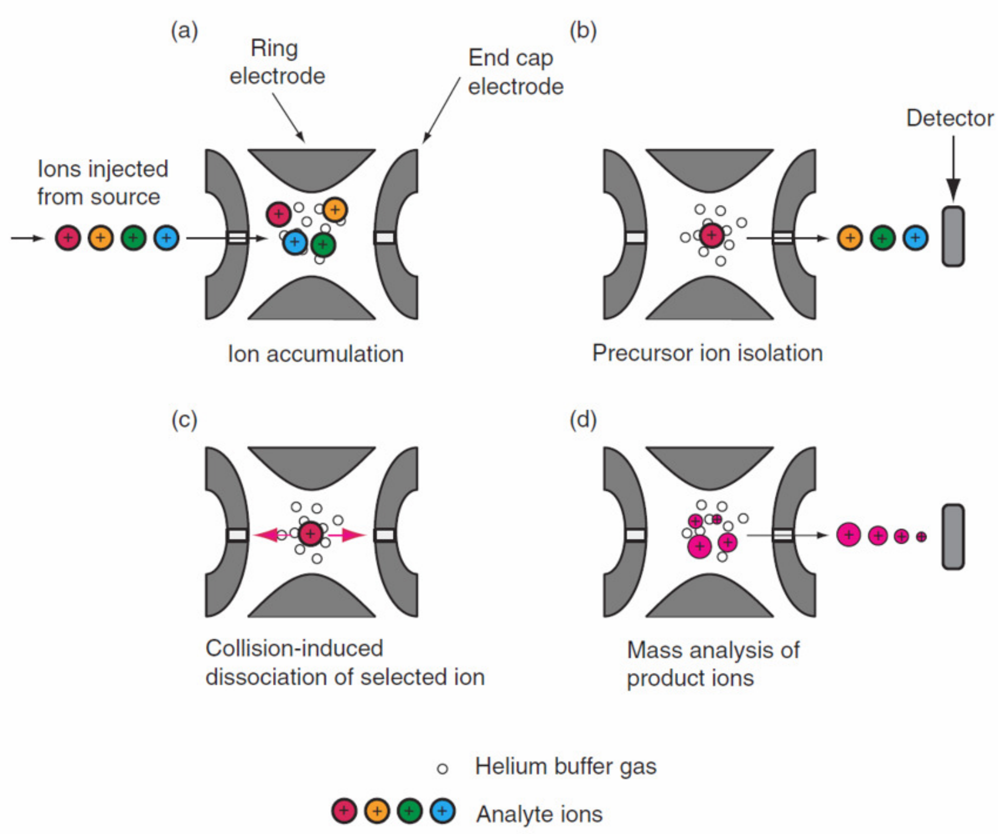


Figure 1.3: Schematic of collision-induced dissociation (CID) in the quadrupole ion trap (MS^2 experiment). In separate events, ions from the source are accumulated and trapped in the space at the center of the electrodes. (a) Ions with a specified*m/z* value are retained in the trap and all others ejected. (b) The specified ions are then collisionally fragmented by axial excitation between the two end caps. (c) The resulting product ions are then sequentially ejected to generate the product ion spectrum. (d) In an MS^3 experiment, one of these product ions may be selectively retained in the trap, excited, and fragmented. Image source: Figure 6 from: DOI: 10.1016/B978-008045382-8.00187-8

mobility (K) of the ion in a buffer or drift gas.[45] K is influenced by the instrument operational parameters of gas temperature and pressure but more importantly for separation, the physicochemical properties of the ion and the gas, including ion charge state, the ion and gas molecule masses, and the rotationally averaged collision cross section (CCS) of the ion and gas molecules.[46] In the traveling wave IM (TWIM),[47] ions are initially accumulated in a trap ion guide and then released as an ion packet into the ion mobility ion guide. Here axial motion through the stack is generated by a repeating sequence of transient DC voltages providing a continuous series of "traveling waves". Ions are then separated as they are driven ahead of these potential hills through the stacked ring ion guides before transfer to the mass analyzer.

TWIM platform has become commercially available that is based on cyclic ion mobility spectrometry (cIMS). The cIMS utilizes a closed-loop TWIM cell (racetrack) for ion mobility separation and enables the separation length to be extended, increasing mobility resolving power.[45] The racetrack is oriented perpendicularly to the rest of the MS system and is situated between the ion two ion guides called the pre- and postarray stores. At the junction where the racetrack meets these ion optics, a multifunctional electrode array generates a traveling wave, enabling the movement of ions between the racetrack and the ion guides by modulating the wave direction on the array. This configuration of ion optics allows researchers the unique ability to isolate ions based on drift time for further analysis. Figure 1.4 illustrates the instrument's design, which is based on the Waters SYNAPT G2-*Si* IM-MS platform, with modifications to the IM separation region to integrate a cIMS device. This device is used for CCS measurements, as discussed in Chapters 4 and 5.

1.2.3 Ion Activation and Fragmentation

The "soft" ionization processes described above typically generate single- or multi-charged molecular ions with little accompanying fragmentation. To obtain structurally informative fragments, these ions must be subject to a second round of mass spectral analysis. This is known as MS/MS or tandem MS. In the first MS stage, an ion is selected or isolated in the mass spectrometer, activated and fragmented, and the product ions mass analyzed in the second MS stage. Depending on the instrument being used, it is possible to perform multistage mass spectrometry (MS^n) and to construct ion fragmentation pathways as part of an exercise in structural elucidation.

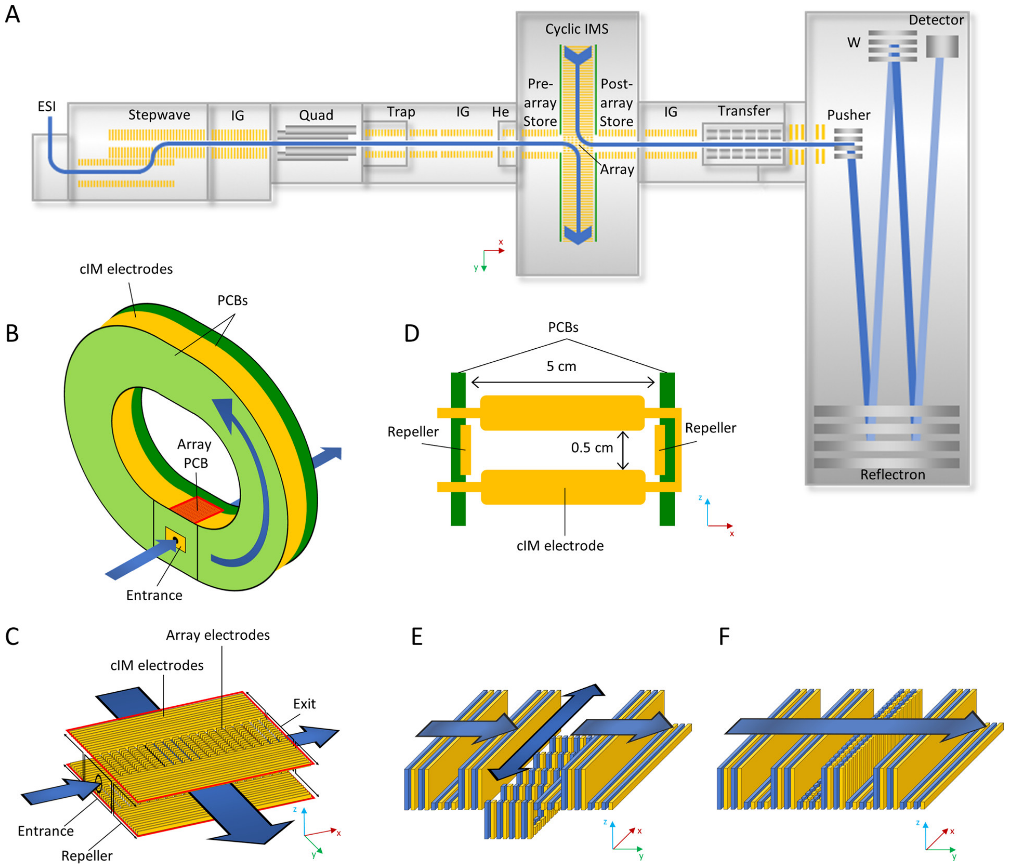


Figure 1.4: Instrument design: (A) instrument overview; (B) cIM device; (C) ion entry/exit region, consisting of array electrodes; (D) structure of cIM electrodes; (E) ion injection/ejection mode, array TWs are applied in the x- (or -x) direction; and (F) separation mode, array TWs are applied in the y-direction. Image source: Figure 1 from DOI: 10.1021/acs.analchem.9b01838

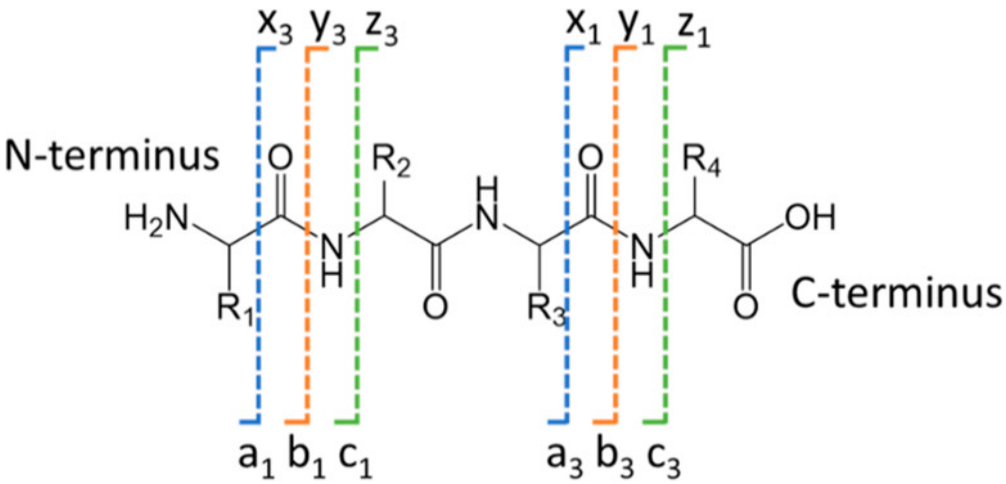


Figure 1.5: Fragmentation nomenclature for peptides and proteins: $\boldsymbol{a}_n$/$\boldsymbol{x}_n$, $\boldsymbol{b}_n$/$\boldsymbol{y}_n$, and $\boldsymbol{c}_n$/$\boldsymbol{z}_n$ ions according to cleavage of backbone bonds. "n" indicates the number of residues contained in the product, as illustrated for a tetrapeptide. Image source: Figure 3 from DOI: 10.1021/acs.chemrev.9b00440

Collision-induced dissociation (CID) is one of the routinely employed techniques that yields fragments based on a two-step process: collision activation and unimolecular dissociation. CID occurs in a collision cell. The cell is differentially pumped and the flow of gas into the cell is carefully controlled. The source of collision gas could be either argon, helium, or nitrogen. In principle, collision energy is transferred to either peptide or small molecule backbone, leading to vibrational excitation. Due to the excitation process, ions tend to undergo gas-phase dissociation, which is visualized as MS^n fragments.[48] In a typical peptide mass spectrum acquired in positive-ion mode, the protonated form of peptides contains protons localized at the basic residues, N-terminus, or on the backbone amide bonds. In theory, during peptide fragmentation, six different fragment ion series would be formed, including the N-terminal $\boldsymbol{a}_n$, $\boldsymbol{b}_n$, and $\boldsymbol{c}_n$ fragments and the C-terminal $\boldsymbol{x}_n$, $\boldsymbol{y}_n$, and $\boldsymbol{z}_n$ fragments (Figure 1.5). In reality, during low-energy collisional activation, only the $\boldsymbol{b}_n$ and $\boldsymbol{y}_n$ ion series are predominantly observed, while the other diagnostic ions series and sequence-specific side chain fragment ions series are unseen.

Ultraviolet/visible photodissociation (UVPD) is a relatively new activation method which utilizes UV and visible light in the electromagnetic spectrum to activate electronic transitions in ions and to induce dissociation. High energy photons are used to trigger bond activation and applied for gas-phase reactions or structure elucidation. Photodissociation can take place by a number of routes, and here are the three

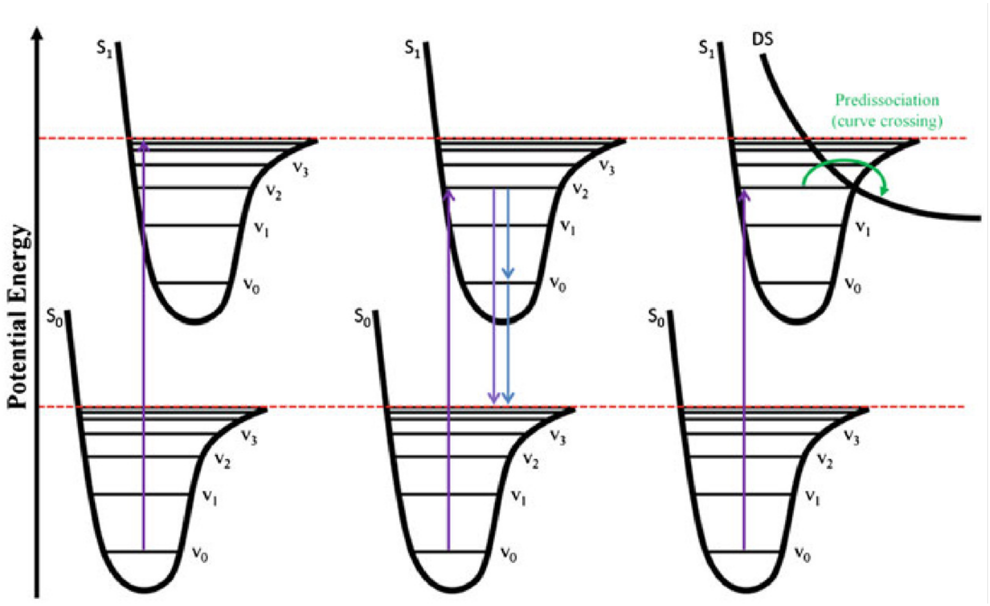


Figure 1.6: Three potential dissociation mechanisms upon absorption of a UV photon. Dissociation thresholds are indicated via red, dashed lines that are available. Image source: Figure 1.11 from DOI: 10.1007/978-3-319-01252-0

most common pathways[49]: in the highest energy mechanism, absorption of a UV photon promotes the molecule from the electronic ground state (S_0) to an excited vibrational level in the electronically excited state (S_1) above the dissociation threshold, resulting in direct dissociation (Figure 1.6 left). The second mechanism differs only in that dissociation proceeds from the ground electronic state, following an internal conversion from the electronically excited state (Figure 1.6 center). The third mechanism proceeds via a potential energy curve crossing to a dissociative state, which has no, or only a shallow potential energy minimum, hence causing dissociation. This results in the dissociation below normal dissociation thresholds (predissociation) (Figure 1.6 right).

Naturally, UVPD experiments require a coupled tunable UV light source and a trapping mass spectrometer. A UVPD spectrum is measured by monitoring the photodissociation yield as a function of incident UV photon frequency. All of the ion-trapping systems are naturally well-suited for UVPD because ions can be accumulated and stored during the photoactivation event, allowing more facile control of energy deposition, fragmentation efficiency, and S/N of the resulting spectra. In our lab, we have modified the Bruker amaZon speed mass spectrometer with optical access into the 3-D ion trap. An Nd-YAG EKSPLA NL301G laser is used to generate a beam of photon pulses. The optical setups allow both the tunable 210-700 nm wavelength and a higher-powered single-wavelength 355 nm beam from different output beam paths (Figure 1.7).[50] We applied the 355 nm wavelength in Chapters 2 and 3 and a 250 nm wavelength in Chapter 4. Additionally, we employed the commercially available Thermo Orbitrap Ascend Tribrid mass spectrometer with an integrated 213 nm laser for UVPD experiments in Chapters 4 and 5.

1.2.4 Ion Detection

The final element of the mass spectrometer is the ion detection system. Once ions are separated, they must be detected and quantified. Detectors range from conversion dynode electron multipliers (EM) to multichannel plates (MCP) and Fourier transform (FT) methods. An EM is a vacuum tube that transforms a primary emission of an analyte ion striking the EM into a secondary emission of electrons from within the tube. This incidental electron current is proportional to the number of analyte ions reaching EM. EM detectors are usually coupled to quadrupole and ion trap mass analyzers. An MCP detector uses similar principles of secondary electron emission and is often used with ToF analyzers. The FT detection methods are used with high mass accuracy analyzers such as the Orbitrap. The principle behind the FT detection is to convert (transform) the unique oscillation orbits of the ions into their corresponding *m/z* values.

1.3 Computational Methods

We apply theoretical calculations to support experimental results, beginning with the creation of initial structures, Born-Oppenheimer Molecular Dynamics (BOMD) calculations by MOPAC under Cyby4 platform is used to generate a trajectory of unique conformers.[51] This process produced thousands of candidate structures, which were clustered by structural similarity using a machine learning model. Then one representative conformer from each cluster is selected for further refinement by enhanced semi-empirical methods such as PM6-D3H4. In the next step, multiple selected complex structures are fully optimized by density functional theory (DFT) methods and ranked by their relative Gibbs energies. The optimized structures and atomic charge densities can be further used for calculations of collision cross sections (CCS), which are structurally relevant parameters obtainable by ion mobility measurements. The experimental localization of crosslinked regions and CCS data along with computational structure and energy analysis represent a powerful approach that allows one to achieve *de novo* structure assignments for ion complexes.

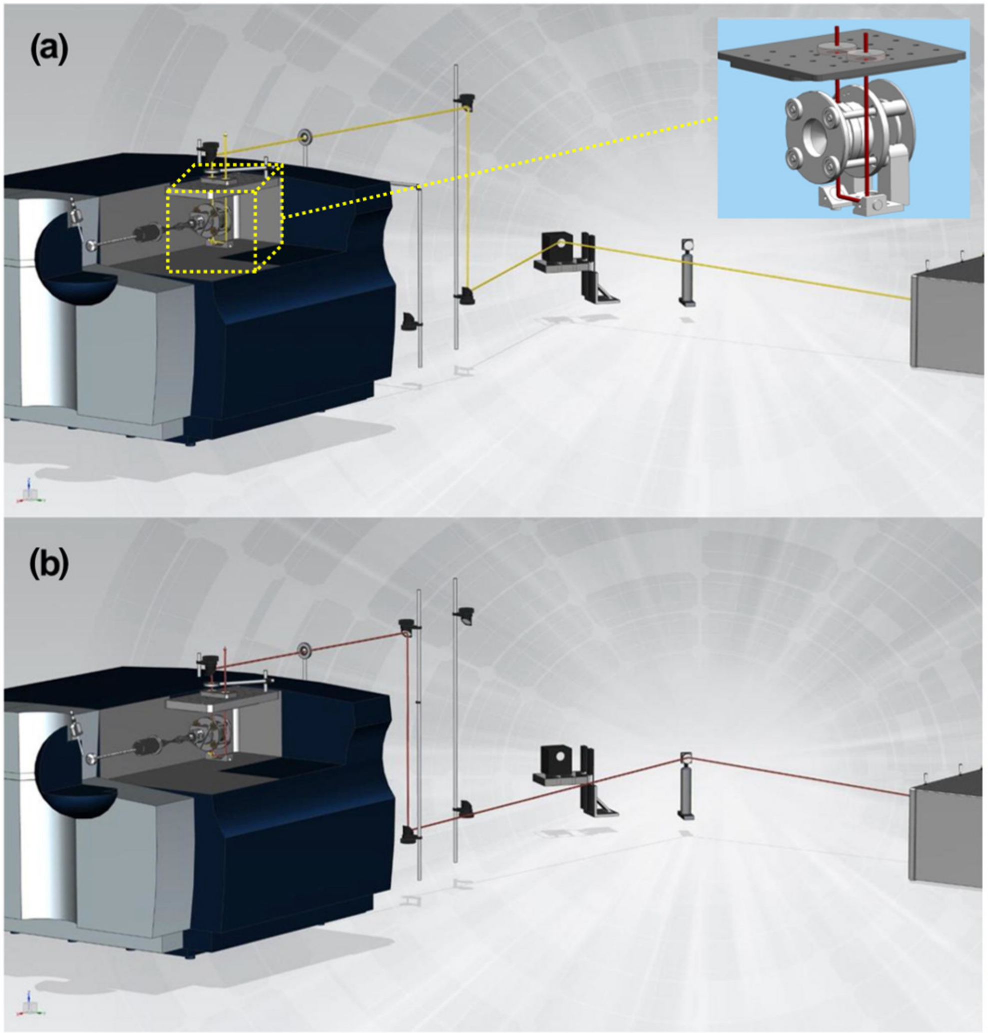


Figure 1.7: Optical setup and modifications of Bruker amaZon mass spectrometer for MS^n-UVPD and automated action spectroscopy, displaying (a) tunable 210-700 nm with modifications for optical access to the trapped ions highlighted in the inset, and (b) single-wavelength 355 nm capabilities using optical flip mount. The (a) tunable setup incorporated a fast steering mirror to improve beam alignment and UVPD reproducibility. Image source: Figure 2 from DOI: 10.1007/s13361-019-02229-z

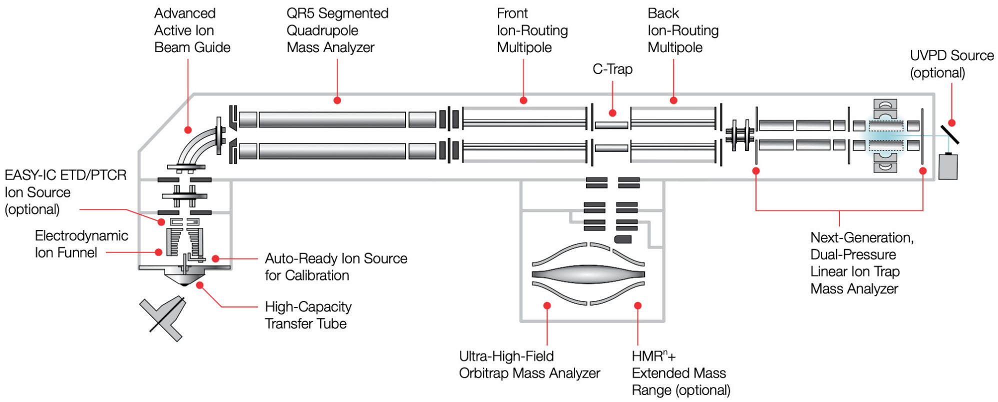


Figure 1.8: Orbitrap Ascend Tribrid mass spectrometer ion path. Image source: https://assets.thermofisher.com/TFS-Assets/CMD/Specification-Sheets/ps-001181-lsms-orbitrap-ascend-tribrid-ms-ps001181-en.pdf

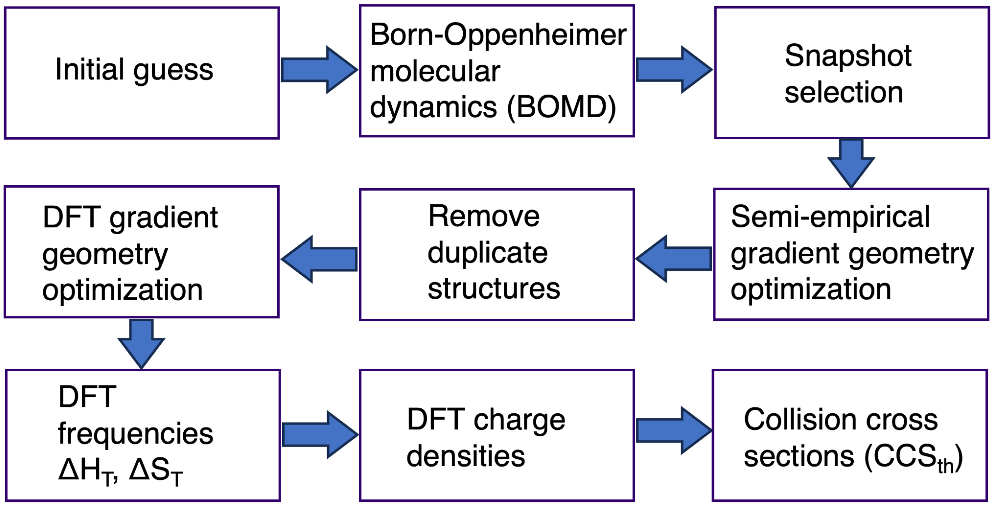


Figure 1.9: Workflow of computational analysis.

1.4 Bibliography

[1] Fancy, D.A.; Kodadek, T. " Chemistry for the Analysis of Protein–Protein Interactions: Rapid and Efficient Cross-Linking Triggered by Long Wavelength Light." *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6020-6024.

[2] Das, J. "Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments." *Chem. Rev.* **2011**, *111*, 4405-4417.

[3] Singh, A.; Thornton, E.R.; Westheimer, F.H. "Photolysis of Diazoacetylchymotrypsin." *J. Biol. Chem.* **1962**, *237*, PC3006-PC3008.

[4] Fleet, G.W.J.; Porter, R.R.; Knowles, J.R. "Affinity Labeling of Antibodies with Aryl Nnitrene as Reactive Group." *Nature (London).* **1969**, *224*, 511-512.

[5] Knowles, J.R. "Photogenerated Reagents for Biological Receptor-Site Labeling." *Acc. Chem. Res.* **1972**, *5*, 155-160.

[6] van der Waals, D. J. "On the Continuity of the Gaseous and Liquid States." Leiden University: Leiden, The Netherlands, **1873**.

[7] Müller-Dethlefs, K.; Hobza, P.; Mu, K.; Hobza, P. "Noncovalent Interactions: A Challenge for Experiment and Theory." *Chem. Rev.* **2000**, *100*, 143-168.

[8] Kollman, P. A. "Noncovalent Interactions." *Acc. Chem. Res.* **1977**, *10*, 365-371.

[9] Adhav, V. A.; Saikrishnan, K. "The Realm of Unconventional Noncovalent Interactions in Proteins: Their Significance in Structure and Function." *ACS Omega.* **2023**, *8*, 22268-22284.

[10] Kollman, P. "Non-Covalent Forces of Importance in Biochemistry." *New Comprehensive Biochemistry.* Vol. 6. Elsevier, **1984**, 55-71.

[11] Morokuma, K. "Molecular Orbital Studies of Hydrogen Bonds. III. C= O··· H–O Hydrogen Bond in H2CO··· H2O and H2CO··· 2H2O." *J. Chem. Phys.* **1971**, *55*, 1236-1244.

[12] Ziegler, T.; Rauk, A. "On the Calculation of Bonding Energies by the Hartree Fock Slater Method." *Theor. Chim. Acta* **1977**, *46*, 1-10.

[13] Chen, F.; Gülbakan, B.; Weidmann, S.; Fagerer, S. R.; Ibáñez, A. J.; Zenobi, R. "Applying Mass Spectrometry to Study Non-Covalent Biomolecule Complexes." *Mass Spectrom. Rev.* **2016**, *35*, 48-70.

[14] Phizicky, E. M.; Fields S. "Protein-Protein Interactions: Methods for Detection and Analysis." *Microbiol. Rev.* **1995**, *59*, 94-123.

[15] Strynadka, N. C.; Eisenstein, M.; Katchalski-Katzir, E.; Shoichet, B. K.; Kuntz, I. D.; Abagyan, R.; Totrov, M.; Janin, J.; Cherfils, J.; Zimmerman, F.; Olson, A.; Duncan, B.; Rao, M.; Jackson, R.; Sternberg, M.; James, M. N. "Molecular Docking Programs Successfully Predict the Binding of a β-Lactamase Inhibitory Protein to TEM-1 β-Lactamase." *Nat. Struct. Biol.* **1996**, *3*, 233-239.

[16] Piersimoni, L.; Kastritis, P. L.; Arlt, C.; Sinz, A. "Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions–A Method for All Seasons." *Chem. Rev.* **2022**, *122*, 7500-7531.

[17] Smith, R. A. G.; Knowles, J. R. "Aryldiazirines. Potential Reagents for Photolabeling of Biological Receptor Sites." *J. Am. Chem. Soc.* **1973**, *95*, 5072.

[18] Wan, J.; Nytka, M.; Qian, H.; Vu, K.; Lemr, K.; Tureček, F. "Nitrile Imines as Peptide and Oligonucleotide Photo-Cross-Linkers in Gas-Phase Ions." *J. Am. Soc. Mass Spectrom.* **2024**, *35*, 344-356.

[19] Galardy, R. E.; Craig, L. C.; Jamieson, J. D.; Printz, M. P. "Photoaffinity Labeling of Peptide Hormone Binding Sites." *J. Biol. Chem.* **1974**, *249*, 3510-3518.

[20] Fleming, S. A. "Chemical Reagents in Photoaffinity Labeling." *Tetrahedron* **1995**, *51*, 12479-12520.

[21] Weber, P. J. A.; Beck-Sickinger, A. G. "Comparison of the Photochemical Behavior of Four Different Photoactivatable Probes." *J. Peptide Res.* **1997**, *49*, 375-383.

[22] Liu, M. T. H. "Chemistry of Diazirines; Vol. I and II." *CRC Press*, Boca Raton, **1987**.

[23] Frey, H. M.; Stevens, I. D. R. "The Photolysis of Dimethyldiazirine." *J. Chem. Soc.* **1963**, 3514-3519.

[24] Barton, D. H. R.; Jaszberenyi, J. C.; Theodorakis, E. A.; Reibenspies, J. H. “The Invention of Radical Reactions. 30. Diazirines as Carbon Radical Traps. Mechanistic Aspects and Synthetic Applications of a Novel and Efficient Amination Process.” *J. Am. Chem. Soc.* **1993**, *115*, 8050-8059.

[25] Korneev, S. M. “Valence Isomerization between Diazo Compounds and Diazirines.” *Eur. J. Org. Chem.* **2011**, *31*, 6153-6175.

[26] Hashimoto, M.; Hatanaka, Y. “Recent Progress in Diazirine-Based Photoaffinity Labeling.” *Eur. J. Org. Chem.* **2008**, 2513-2523.

[27] Dubinsky, L.; Bastiaan, P.; Kroma, B. P.; Meijler, M. M. “Diazirine Based Photoaffinity Labeling.” *Bioorg. Med. Chem.* **2012**, *20*, 554-570.

[28] Kumar, A. B.; Anderson, J. M.; Manetsch, R. “Design, Synthesis and Photoactivation Studies of Fluorous Photolabels.” *Org. Biomol. Chem.* **2011**, *9*, 6284-6292.

[29] Burton, N. R.; Kim, P.; Backus, K. M. “Photoaffinity Labelling Strategies for Mapping the Small Molecule-Protein Interactome.” *Org. Biomol. Chem.* **2021**, *19*, 7792-7809.

[30] Song, W.; Wang, Y.; Qu, J.; Madden, M. M.; Lin, Q. “A Photoinducible 1,3-Dipolar Cycloaddition Reaction for Rapid, Selective Modification of Tetrazole-Containing Proteins.” *Angew. Chem., Int. Ed.* **2008**, *47*, 2832-2835.

[31] Sharp, J. T. “Nitrile Ylides and Nitrile Imines.” in the Chemistry of Heterocyclic Compounds 59, in *Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products*, ed. A. Padwa and W. H. Pearson, John Wiley & Sons, NewYork, **2002**.

[32] Shawali, A. S. “Reactions of Heterocyclic Compounds with Nitrilimines and Their Precursors.” *Chem. Rev.* **1993**, *93*, 2731-2777.

[33] Marshall, D. L.; Menzel, J. P.; McKinnon, B. I.; Blinco, J. P.; Trevitt, A. J.; Barner-Kowolik, C.; Blanksby, S. J. “Laser Photodissociation Action Spectroscopy for the Wavelength-Dependent Evaluation of Photoligation Reactions.” *Anal. Chem.* **2021**, *93*, 8091-8098.

[34] Platz, M. S. “Atom-Transfer Reactions of Aromatic Carbenes.” *Acc. Chem. Res.* **1988**, *21*, 236-242.

[35] Mieusset, J. L.; Brinker, U. H. “The Carbene Reactivity Surface: A Classification.” *J. Org. Chem.* **2008**, *73*, 1553-1558.

[36] Senthilnathan, V. P.; Platz, M. S. “Determination of the Absolute Rates of Decay of Arylcarbenes in Various Low Temperature Matrixes by Electron Spin Resonance Spectroscopy.” *J. Am. Chem. Soc.* **1980**, *102*, 7637-7643.

[37] Hadel, L. M.; Maloney, V. M.; Platz, M. S.; McGimpsey, W. G.; Scaiano, J. C. “The Absolute Kinetics of Several Reactions of Substituted Diphenylcarbenes.” *J. Phys. Chem.* **1986**, *90*, 2488-2491.

[38] Wright, B. B.; Senthilnathan, V. P.; Platz, M. S.; McCurdy Jr, C. W. “Tunneling Parameters for the Hydrogen Atom Abstraction Reactions of Diphenylcarbene in a Low Temperature Toluene Matrix.” *Tetrahedron Lett.* **1982**, *23*, 833-836.

[39] Tureček, F. “Covalent Crosslinking in Gas-Phase Biomolecular Ions. An Account and Perspective.” *Phys. Chem. Chem. Phys.* **2023**, *25*, 32292-32304.

[40] Yamashita, M.; Fenn, J. B. “Electrospray Ion Source. Another Variation on the Free-Jet Theme.” *J. Phys. Chem.* **1984**, *88*, 4451-4459.

[41] Ho, C. S.; Lam, C. W.; Chan, M. H.; Cheung, R. C.; Law, L. K.; Lit, L. C.; Ng, K. F.; Suen, M. W.; Tai, H. L. “Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications.” *Clin. Biochem. Rev.* **2003**, *24*, 3-12.

[42] El-Aneed, A.; Cohen, A.; Banoub, J. “Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers.” *Appl. Spectrosc. Rev.* **2009**, *44*, 210-230.

[43] Makarov A. “Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis.” *Anal. Chem.* **2000**, *72*, 1156-1162.

[44] Greaves, J.; Roboz, J. “Mass Spectrometry for the Novice.” CRC Press, **2013**.

[45] Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson K.; Langridge, D.; Green, M. “A Cyclic Ion Mobility-Mass Spectrometry System.” *Anal. Chem.* **2019**, *91*, 8564-8573.

[46] Mason, E. A.; McDaniel, E. W. "Transport Properties of Ions in Gases." John Wiley & Sons, Ltd., **1988**.

[47] Pringle, S.D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassionos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. "An Investigation of the Mobility Separation of Some Peptide and Protein Ions Using a New Hybrid Quadrupole/Travelling Wave IMS/oa-ToF Instrument." *Int. J. Mass Spectrom.* **2007**, *261*, 1-12.

[48] Wells, J. M.; McLuckey, S. A. "Collision-Induced Dissociation (CID) of Peptides and Proteins." *Methods Enzymol.* **2005**, *402*, 148-185.

[49] Polfer, N. C.; Dugourd, P. "Laser Photodissociation and Spectroscopy of Mass-Separated Biomolecular Ions." Springer, **2013**.

[50] Dang, A.; Korn, J. A.; Gladden, J.; Mozzone, B.; Tureček, F. "UV-Vis Photodissociation Action Spectroscopy on Thermo LTQ-XL ETD and Bruker amaZon Ion Trap Mass Spectrometers: a Practical Guide." *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1558-1564.

[51] Řezáč, J. "Cuby: An Integrative Framework for Computational Chemistry." *J. Comput. Chem.* **2016**, *37*, 1230-1237.

Chapter 2

Carbene Crosslinking with Hydrophobic Peptide Ion Scaffolds in the Gas-Phase

Reproduced in part with permission from Hongyi Zhu, Václav Zima, Emily R. Ding, and František Tureček. Carbene Cross-Linking in Gas-Phase Peptide Ion Scaffolds. *Journal of the American Society for Mass Spectrometry*, **2023**, *34* (4), 763-774.

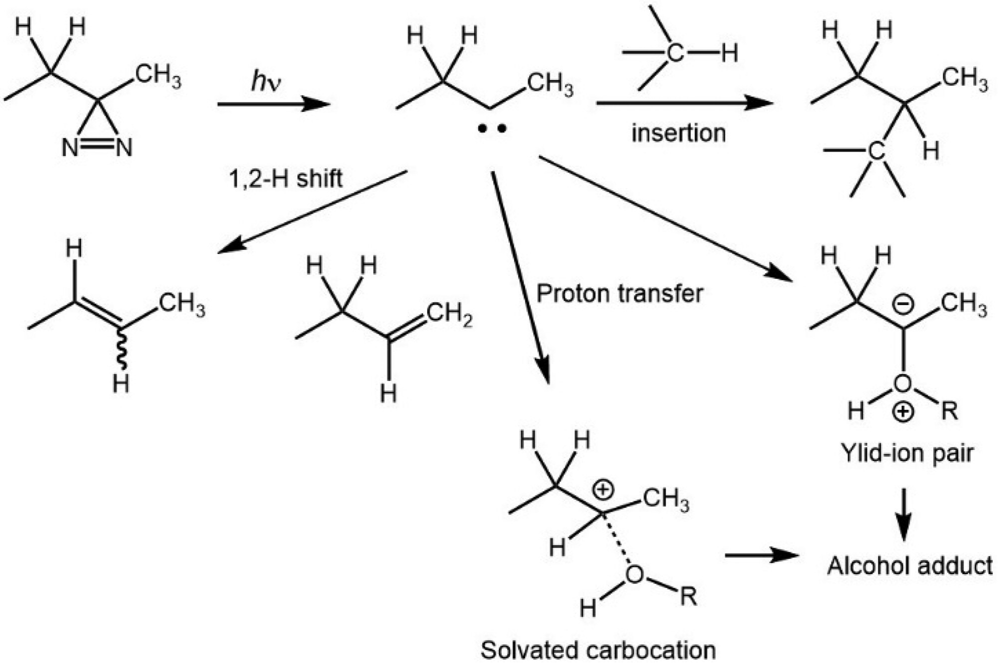
Abstract. Scaffolds consisting of a peptide, a phthalate linker, and a 4,4-azipentyl group were synthesized and used to study intramolecular peptide-carbene crosslinking in gas-phase cations. Carbene intermediates were generated by UV-laser photodissociation at 355 nm of the diazirine ring in mass-selected ions, and the crosslinked products were detected and quantified by collisioninduced dissociation tandem mass spectrometry (CID-MS^n, n = 3-5). Peptide scaffolds containing Ala, and Leu residues with a C-terminal Gly gave 21-26% yields of crosslinked products. Experiments using hydrogen-deuteriumhydrogen exchange, carboxyl group blocking, and analysis of CID-MS^n spectra of reference synthetic products revealed that a significant fraction of crosslinks involved the Gly amide and carboxyl groups. Interpretation of the crosslinking results was aided by Born-Oppenheimer molecular dynamics (BOMD) and density functional theory calculations that allowed us to establish the protonation sites and conformations of the precursor ions. Analysis of long (100 ps) BOMD trajectories was used to count close contacts between the incipient carbene and peptide atoms, and the counting statistics was correlated with the results of gas-phase crosslinking.

2.1 Introduction

Photoaffinity labeling in combination with a chemical probe to covalently bind its target upon UV irradiation has demonstrated considerable promise in drug discovery for identifying new drug targets and binding sites. In particular, carbene-mediated photoaffinity labeling has been widely used in drug target identification owing to its excellent photolabeling efficiency, minimal steric interference and longer excitation wavelength. Carbene-mediated photoaffinity labeling has been increasingly used in target identification owing to the small size of carbenes, high crosslinking efficiency, long excitation wavelength, excellent chemical stability, less disruption of the interaction interface and minimal damage to biological samples. Two major types of carbene precursor diazirine and diazo compounds have occupied a uniquely important place in photoaffinity labeling. In particular, in comparison with the latter, diazirines generate relatively more reactive carbenes[1][2] and have either relatively high thermal and chemical stability[3] or a much shorter lifetime when activated by photoirradiation. Upon exposure to UV light at a certain wavelength (350–380 nm), carbenes generated by diazo compounds or diazirines are capable of spontaneously forming a covalent bond with a neighboring molecule via insertion into bonds such as C–H, O–H, and N–H to stabilize the interaction between a target and a probe[4]-[7].

Our research group have developed methods for gas-phase crosslinking of non-covalent complexes of peptides[8]-[12], dinucleotides[13], and small molecules[14]. A common feature among these peptide-related phototags is the placement of the diazirine ring within an aliphatic chain, flanked by CH_2 and CH_3 groups. This structure allows for rapid, competitive 1,2-hydrogen shifts of the transient singlet carbene, occurring on a high picosecond to low nanosecond timescale, which converts the carbene into an unreactive alkene (Scheme 2.1). We have exploited this side reaction as an internal clock in our molecular dynamics (MD) simulations, run on 100 ps timescales, to identify and quantify the close contacts between the nascent carbene and X–H bonds that may undergo insertion. This approach has enabled us to correlate computationally determined structures with experimental results obtained from photodissociation and sequence analysis by CID-MS^3[8]-[12]. The carbene-alkene isomerization is a highly exergonic reaction which often drives dissociation of the rearranged complex[8]-[14].

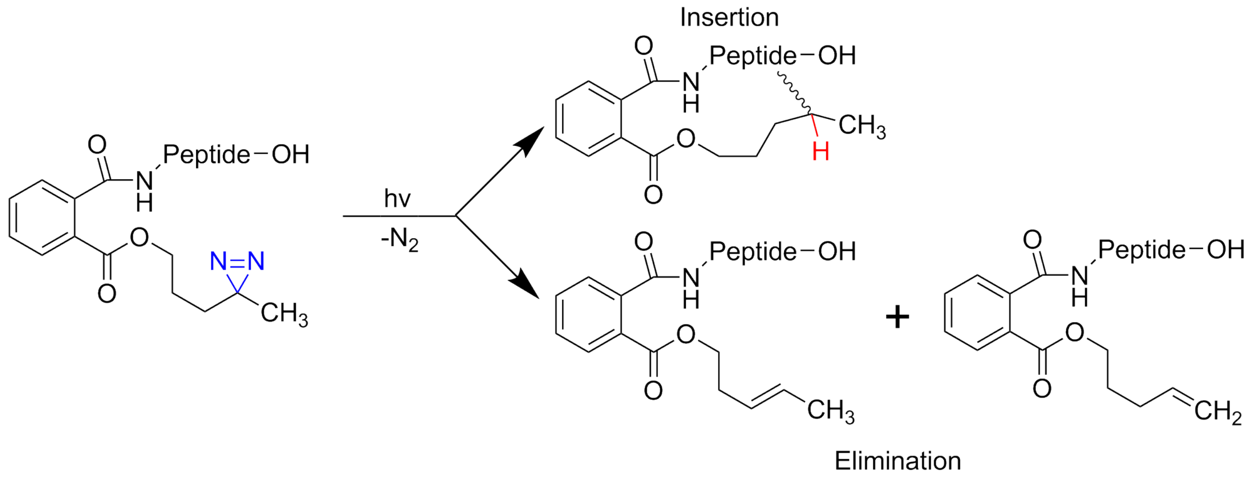
A major limitation in previous studies of peptide crosslinking has been the inability to pinpoint precise insertion sites. This has been due to incomplete sequencing of the crosslinks, which often narrows down



Scheme 2.1: Diazirine photolysis products.

the insertion site to only a few amino acid residues, without identifying the specific atoms of the X–H bond involved. Our research aims to overcome these limitations by advancing carbene crosslinking within model molecular frameworks to achieve a higher degree of spatial resolution regarding insertion sites. To this end, we have engineered a novel model system that combines diazirine-derived carbene precursors with target peptide segments in a single molecular scaffold (Scheme 2.2). This design eliminates issues related to inefficient non-covalent complex formation in protic solvents, observed in previous studies[12]. The scaffold includes peptide sequences composed of hydrophobic amino acids such as alanine, leucine, and glycine, which allow for flexible protonation sites and enhance control over the crosslinking process. Besides, leucine is strategically positioned before, between, and after two alanine residues to examine how different sequence arrangements influence crosslinking behavior, which allows us to explore the potential effects of peptide sequence order on carbene insertion and crosslinking properties.

Moreover, our experimental approach is complemented by computational studies using Born-Oppenheimer Molecular Dynamics (BOMD) and Density Functional Theory (DFT). We select the lowest-energy protomers and conformers for each peptide ion sequence in the scaffolds and determine the dynamics of throughspace contacts of the incipient carbene atom with selected atoms and bonds in the peptide moiety in



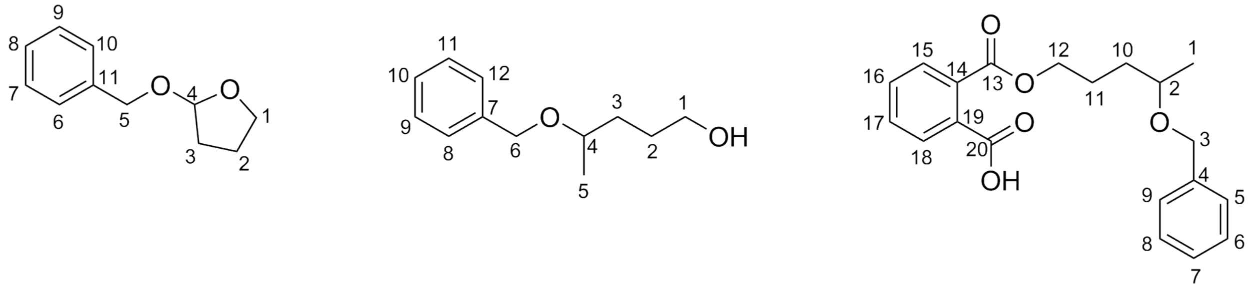
Scheme 2.2: Scaffolds for carbene-peptide crosslinking where Peptide stands for LAAG, ALAG, and AALG sequences, indicating crosslinks and alkene isomerization products. The alternative formation of 4-pentenyl conjugates is not shown.

the course of 100 ps trajectories. The combined experimental and computational results bear on achieving a more detailed understanding of carbene based crosslinking to polar biomolecular targets.

2.2 Experimental Section

2.2.1 Materials and Methods

Unless otherwise noted, all starting materials were purchased from Sigma Aldrich (St. Louis, MO) and were used as received. Acetonitrile was dried using 4A molecular sieves, grade 514 (8-12 mesh). Chromatography was performed on Acros Organics Silica gel, ultrapure, 60-200 micrometer, 60A. TLC was performed on TLC Silica gel 60 F_{254} aluminum sheets. Spots were detected with permanganate stains, bromocresol green stains or under a UV lamp. HPLC was performed on a Shimadzu (Canby, OR) UFLC with a YMC analytical Pack Pro C18 column (5 μm, 250 X 4.6 mm) at a flow rate of 0.5 mL/min over 40 min. Preparative scale purification were performed on a Shimadzu Shim-pack GIS C18 (5 μm, 250 X 10 mm) at a flow rate of 2.3 mL/min over 40 min. The employed mobile phases were water (A) and 65% acetonitrile, 35% 2-propanol (B). A typical gradient for analysis and purification was 10-100% B. UV traces were collected at 254 nm. ^{1}H NMR spectra were recorded on Bruker spectrometers at 300 and 500 MHz. ^{13}C NMR Spectra were recorded at 126 MHz. Coupling constants (J) are given in Hz and chemical



Scheme 2.3: NMR characterization numbering.

shifts are on the δ-scale. NMR characterization was carried out according to the numbering in Scheme 2.3. Mass spectra were measured on a Bruker Daltonik, GmbH (Bremen, Germany), amaZon ion trap mass spectrometer that was equipped with an EKSPLA NL301G Nd-YAG laser (Altos Photonics, Bozeman, MT, USA) working at 20 Hz frequency and 3- to 6-ns pulse width. Ions were produced by electrospray ionization, selected by mass in the ion trap, and irradiated by multiple laser pulses at 355 nm. Multistage tandem mass spectra were obtained by collision-induced dissociation (CID) of mass selected fragment ions. The excitation amplitude was chosen to achieve a high conversion of precursor ions. HRMS spectra were recorded on Thermo-Electron Fisher Orbitrap Velos mass spectrometer. Samples were introduced by flow injection into the ESI source.

2.2.2 Synthetic Procedures

Abbreviations Used

t-Boc: *tert*-Butyloxycarbonyl

Fmoc: fluorenylmethyloxycarbonyl

TFA: trifluoroacetic acid

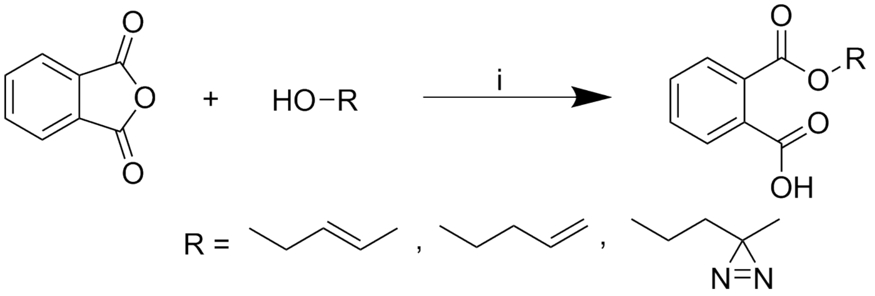
DCM: dichloromethane

DIPEA: di-isopropylethylamine

HBTU: (*o*-(benzotriazol-1-yl)-*N, N, N', N'*-tetramethyluronium hexafluorophosphate

HATU: (1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium-3-oxide hexafluorophosphate

DMF: *N, N*-dimethylformamide



Scheme 2.4: Reagents and conditions (i) Triethylamine, 4-dimethylaminopyridine, ethyl acetate, r.t., overnight.

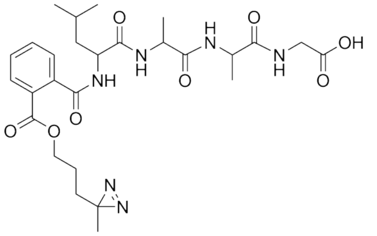
2.2.2.1 General Procedure for Esterification

Phthalic anhydride (1 eq.) was added to the solution of the corresponding alcohol (1 eq.) in ethyl acetate. Triethylamine (1 eq.) and 4-(N,N-dimethylamino)pyridine (0.1 eq.) were added to the reaction mixture and stirred overnight at room temperature. The mixture was washed with 5% HCl for 3 times and brine for once, then dried over $MgSO_4$ and evaporated to dryness under reduced pressure to give the title compound as a pale-yellow oil.

2.2.2.2 General Procedure for Solid Phase Peptide Synthesis

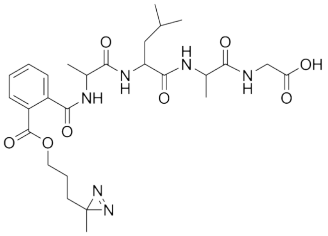
Resin containing an appropriate Fmoc protected C-terminal amino acid (50 mg, 1 equiv.) was weighed into a 3 mL syringe. The beads were swollen with DMF for 20 min. The Fmoc protective group on the beads was removed by 2 mL 20% piperidine in DMF after shaking for 10 min and the beads were washed with DMF 6 times. The coupling cocktail solution for peptide chain elongation was prepared by dissolving a Fmoc protected amino acid (3 equiv.), PyBOP (47 mg, 3 equiv.), HOBt (17 mg, 4.2 equiv.) and 25 μL NMM in 1 mL DMF. The coupling cocktail solution was injected into the syringe and shaken for 20 min for the reaction to complete. Then the beads were washed with DMF 6 times. The Fmoc deprotection, DMF wash, coupling, and DMF wash were repeated in each peptide alongation step until the desired peptide sequence was obtained. After the last amino acid was added and the Fmoc group was deprotected, the corresponding carboxylic acid with or without the diazirine phototag (Scheme 2.4) was coupled to the peptide N-terminus by adding a solution made by dissolving the acid (6 equiv.), PyBOP (94 mg, 6 equiv.), HOBt (34 mg, 8.4 equiv.) and 25 μL NMM in 1 mL DMF, and shaking the reaction mixture overnight. Then, the beads were washed with DMF 6 times, and washed with DCM 6 times to remove DMF. The peptide conjugate was stripped from the resin by injecting a solution mixture of TFA/water/triethylsilane (95:2.5:2.5) twice. The TFA solution containing the peptide product was collected in a small glass vial and the solvent was evaporated with air flow to give 15 mg of the peptide conjugate product.

4,4-Azipent-1-yl-s-LAAG(I)



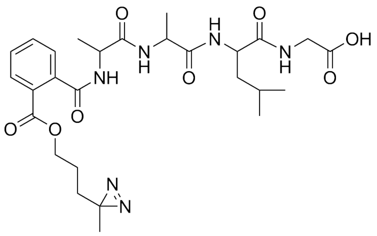
The title conjugate was prepared according to the general procedure for esterification and the general procedure for solid phase peptide synthesis. The product was characterized by MS/MS: *m/z* 575 (M+H)$^+$, 500 (–GlyOH), 429 (–Ala–GlyOH), 358 (–Ala–Ala–GlyOH).

4,4-Azipent-1-yl-s-ALAG(II)



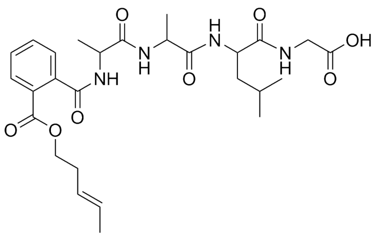
The title conjugate was prepared according to the general procedure for esterification and the general procedure for solid phase peptide synthesis. The product was characterized by MS/MS: *m/z* 575 (M+H)$^+$, 500 (–GlyOH), 429 (–Ala–GlyOH).

4,4-Azipent-1-yl-s-AALG (III)



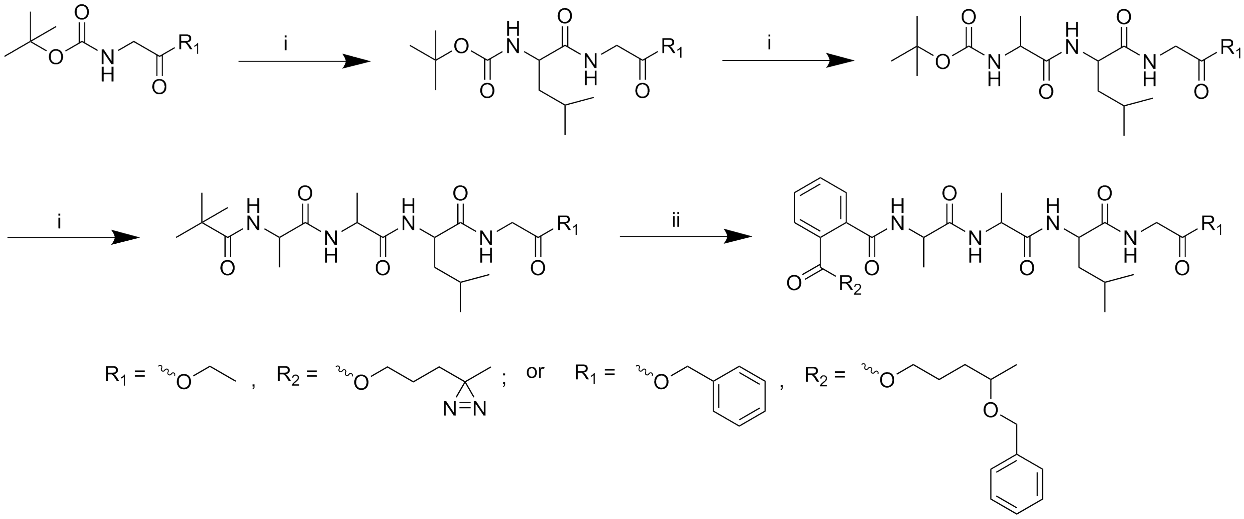
The title conjugate was prepared according to the general procedure for esterification and the general procedure for solid phase peptide synthesis. The product was characterized by MS/MS: *m/z* 575 (M+H)$^+$, 500 (–GlyOH), 387 (–Leu–GlyOH).

3-penten-1-yl-s-AALG (IV)

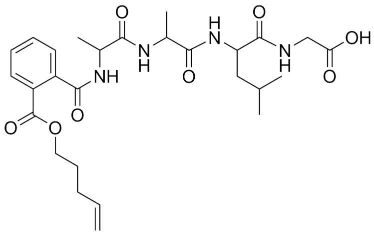


The title conjugate was prepared according to the general procedure for esterification and the general procedure for solid phase peptide synthesis. The product was characterized by MS/MS: *m/z* 547 (M+H)$^+$, 472 (–GlyOH), 359 (–Leu–GlyOH).

4-penten-1-yl-s-AALG (V)



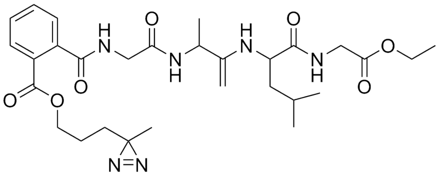
Scheme 2.5: Reagents and conditions (i) TFA/DCM 1:1, DIPEA, HBTU, DMF, r.t., 4h; (ii) TFA/DCM 1:1, DIPEA, HATU, DMF, r.t., 4h.



The title conjugate was prepared according to the general procedure for esterification and the general procedure for solid phase peptide synthesis. The product was characterized by MS/MS: *m/z* 547 $(M+H)^+$, 472 (–GlyOH), 458, 359 (–Leu–GlyOH).

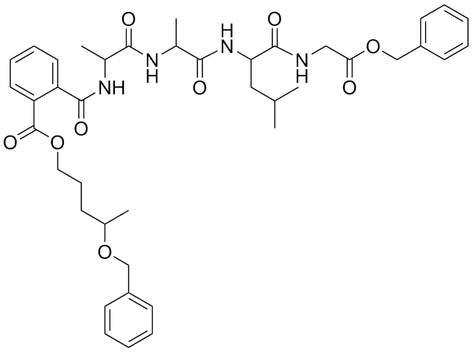
2.2.2.3 General Procedure for In-Solution Peptide Synthesis

t-Boc-protected peptide (1 equiv.) was deprotected using the solution TFA/DCM 1:1 (1 g of peptide/10 mL) for 1 h. The solution was evaporated to dryness and the residue was dissolved in DMF (1M) with DIPEA to ensure basic pH. *t*-Boc-protected amino acid (1.1 equiv.), HBTU (1.1 equiv.) (or HATU for the last synthesis step) and DIPEA (3 equiv.) were stirred in DMF (0.5 M) for 15 min and the solution of the deprotected peptide was added. The solution was stirred for 2.5 h. 2-Aminoethylmorpholine (0.3 equiv.) was added and the reaction mixture was stirred for further 30 min. The reaction mixture was diluted with EtOAc (70 mL) and extracted twice with 5% aqueous solution of HCl, twice with saturated aqueous solution of $NaHCO_3$ and with brine. The organic phase was concentrated to a minimum volume and the peptide was precipitated by tert-butylmethyl ether. 4,4-Azipent-1-yl-s-AALG-OEt(VI)



The title conjugate was prepared according to the general procedure for in solution peptide synthesis. The product was characterized by MS/MS: *m/z* 603 $(M+H)^+$, 500 (–GlyOEt).

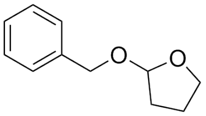
4-Benzyloxypent-1-yl-s-AALG-OBnz(VII)



The title conjugate was prepared according to the general procedure for in solution peptide synthesis. The product was characterized by MS/MS: *m/z* 745 $(M+H)^+$, 580 (–GlyOBnz), 467 (–Leu–GlyOBnz). HRMS (ESI) *m/z* calcd for $C_{41}H_{52}N_4O_9$ $(M+H)^+$ 745.3807; found 745.3856.

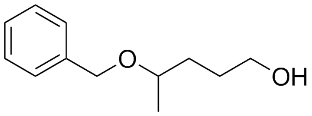
2.2.2.4 Synthetic Protocols

2-Benzyloxyoxolan (VIII)



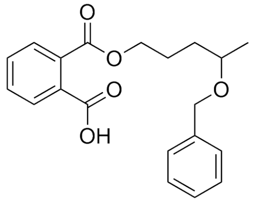
2,3-Dihydrofuran (0.97 g, 13.9 mmol) and pyridinium para-toluensulphonate (0.232 g, 0.93 mmol) were added to benzyl alcohol (1.00 g, 9.25 mmol). The reaction mixture was stirred overnight and purified by column chromatography (SiO_2, hexane/ethyl acetate; 10/1) to give the title compound (1.61 g, 98%). ^1H NMR (500 MHz, $CDCl_3$) δ 7.48–7.19 (m, 5H), 5.25 (s, 1H), 4.82–4.66 (m, 1H), 4.58–4.41 (m, 1H), 3.96 (dd, J = 47.1, 1.6 Hz, 2H), 2.17–1.80 (m, 4H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.4 (C-11), 128.4 (C-7, C-9), 127.9, 127.5 (C-6, C-8, C-10), 103.1 (C-4), 68.8 (C-5), 67.0 (C-1), 32.4 (C-3), 23.5 (C-2).

4-Benzyloxy-1-pentanol (IX)[15]



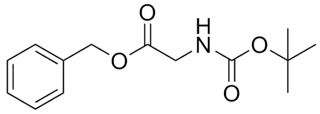
Compound VIII (1.00 g, 5.61 mmol) was dissolved in hexane (30 mL, distilled from CaH_2 and stored over 4A molecular sieves). Trimethylaluminium (8.42 mL, 16.8 mmol) was added to the solution at room temperature. The reaction was stirred and refluxed for 1 h and quenched by slow addition, with vigorous stirring, into cold aqueous saturated NH_4Cl solution. The solids were filtered off and thoroughly washed with ethyl acetate. The aqueous phase was extracted with ethyl acetate (3 × 10 mL), and the combined organic phases were washed with water (3 × 10 mL), dried with $MgSO_4$ and concentrated under reduced pressure. The residue was purified by column chromatography (SiO_2, hexane/ethyl acetate; 4/1) to give the title compound (0.82 g, 75%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41-7.20 (m, 5H), 4.60 (d, J = 11.6 Hz, 1H), 4.45 (d, J = 11.6 Hz, 1H), 3.75-3.44 (m, 3H), 1.95 (s, 1H), 1.77-1.51 (m, 4H), 1.23 (d, J = 6.1 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.8 (C-7), 128.5 (C-9, 11), 127.9, 127.7 (C-8, 10, 12), 74.9, 70.6 (C-4,6), 63.1 (C-1), 33.4 (C-3), 28.9 (C-2), 19.6 (C-5).

4-Benzyloxy-1-pentyl phthalate (X)



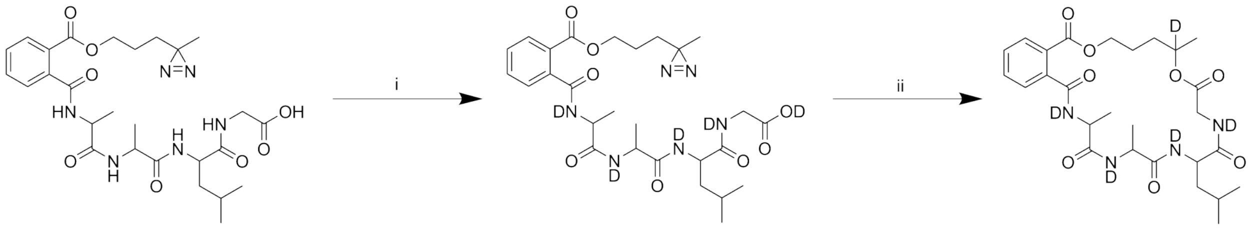
Phthalic anhydride (0.42 g, 2.83 mmol) was added to the solution of compound IX (0.63 g, 2.83 mmol) in ethyl acetate (25 mL). Triethylamine (0.29 g, 2.83 mmol) and 4-dimethylaminopyridine (35 mg, 0.283 mmol) were added to the reaction and stirred overnight. The mixture was washed with 5% HCl (3 × 20 mL) and brine, dried over $MgSO_4$ and evaporated to dryness under reduced pressure. The residue was purified by column chromatography (SiO_2, DCM/MeOH; 15/1) to give the title compound (0.48 g, 50%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.82-7.70 (m, 1H), 7.70-7.58 (m, 1H), 7.48 (td, J = 7.1, 1.6 Hz, 2H), 7.31- 7.12 (m, 5H), 4.54 (d, J = 11.7 Hz, 1H), 4.38 (d, J = 11.7 Hz, 1H), 4.36-4.25 (m, 1H), 4.25-4.16 (m, 1H), 3.60-3.38 (m, 1H), 1.90-1.75 (m, 1H), 1.75-1.61 (m, 2H), 1.61-1.46 (m, 1H), 1.14 (d, J = 6.1 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.1, 168.2 (C-13, 20), 138.4 (C-4), 133.0 (C-14), 131.8 (C-19), 131.0, 129.7, 129.0 (C-15, 18), 128.5, 128.0 (C-16, 17, 6, 8), 127.8 (C-5, 7, 9), 74.7, 70.6 (C-2, 3), 65.9 (C-12), 33.3 (C-10), 24.6 (C-11), 19.6 (C-1).

N-tert-Butyloxycarbonylglycine benzyl ester (XI)[16]

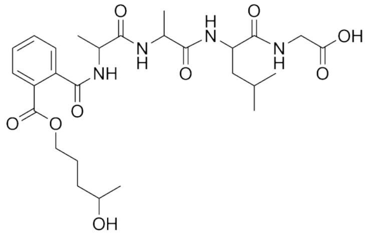


To a stirred solution of N-Boc-glycine (1.00 g, 5.71 mmol) in DCM, was added benzyl alcohol (0.62 g, 5.71 mmol) and 4-dimethylaminopyridine (70 mg, 0.57 mmol) and the mixture was cooled to 0 °C. A 1 M solution of *N,N'*-dicyclohexylcarbodiimide (1.24 g, 6.0 mmol) in DCM (6 mL) was added. The reaction mixture was left stirring overnight. The precipitated urea was filtered off and the filtrate was washed with 5% NaOH and 5% HCl. The organic phase was then dried over $MgSO_4$ and evaporated to dryness under reduced pressure to give the title compound (1.36 g, 90%). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.35 (s, 5H), 5.18 (s, 2H), 4.97 (d, J = 26.0 Hz, 1H), 3.95 (d, J = 5.7 Hz, 2H), 1.44 (s, 9H).

4-Hydroxy-1-pentyl-s-AALG (XII)[17]

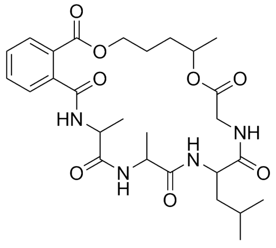


Scheme 2.6: Reagents and conditions (i) MeOD:D_2O 1:1; (ii) UV lamp, overnight.



A mixture of 10% Pd/C and 15% $Pd(OH)_2$ was added to a solution of compound VII (228 mg, 0.31 mmol) in 5 mL MeOH. The inside air was replaced with H_2 (balloon) by three vacuum/H_2 cycles. The reaction mixture was stirred at room temperature for 2 days and was filtered through Celite. The filtrate was concentrated in reduced pressure to give the title conjugate (160 mg, 90%). This product was characterized by MS: *m/z* 565 $(M+H)^+$.

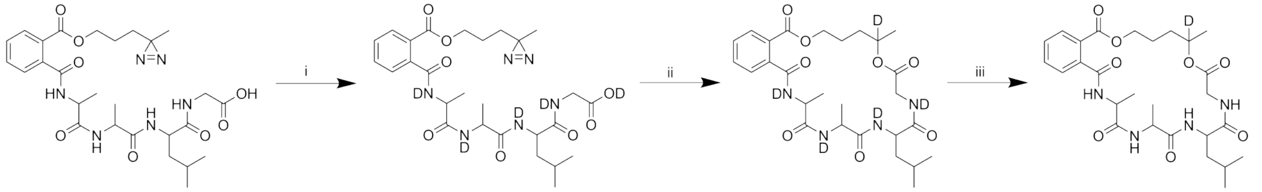
Benzo[c]-6(S), 9(S), 18(R,S)-trimethyl-12(S)-2-methylpropyl-1,17-dioxa-5,8,11,14-tetraazacy-clohenicosan-2, 4, 7, 10, 13, 16-hexaone (XIII)



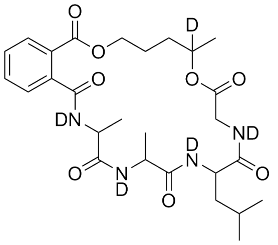
To a stirred solution of compound XII (80 mg, 0.15 mmol) in ACN (60 mL), was added DCC (30 mg, 0.15 mmol) and DMAP (1.8 mg, 0.015 mmol). The reaction mixture was stirred overnight and filtered with cotton. The filtrate was evaporated to dryness under reduced pressure and purified by HPLC (water (A) and 65% ACN, 35% 2-propanol (B)). The product was characterized by MS/MS: *m/z* 547 $(M+H)^+$, 529 (–H_2O), 519 (–CO), 472 (–GlyOH), 461 (–$C_5H_{10}O$), 443 (*m/z* 461–H_2O), 386 (–GlyOH–$C_5H_{10}O$), 359 (–Leu–GlyOH), 288, 275. HRMS (ESI) m/z calcd for $C_{27}H_{38}N_4O_8$ $(M+H)^+$ 547.2762, found 547.2769.

2.2.2.5 Procedure for UVPD in Solution of H/D Exchanged 4,4-Azi-pent-1-yl-s-AALG

4,4-Azipent-1-yl-s-AALG (5 mg) was dissolved in MeOD:D_2O 1:1 (500 μL) and irradiated by a UV lamp at 365 nm overnight. The reaction mixture was diluted with MeOD:D_2O:CD_3COOD in a 1:1:0.01 ratio for electrospray.



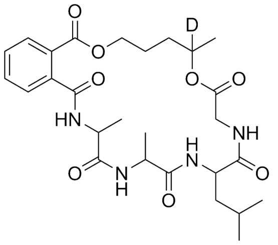
Scheme 2.7: Reagents and conditions (i) MeOD:D_2O 1:1; (ii) UV lamp, overnight, concentrate; (iii) MeOH:H_2O 1:1.



The products including [D_5]-4,4-azipent-1-yl-s-AALG (D_5-XIII) were characterized by MS/MS: *m/z* 553 (D_5-M+D)$^+$, 533 (–D_2O), 525 (–CO), 475 (–[D_2]GlyOD), 466 (–C_5H_9DO), 446 (*m/z* 466–D_2O), 389, 275.

2.2.2.6 Procedure of UVPD and H/D Back Exchange of 4,4-azi-pent-1-yl-s-AALG in Solution

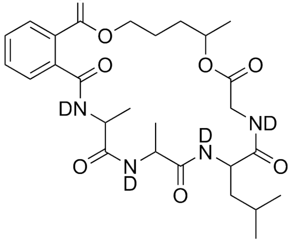
4,4-Azi-pent-1-yl-s-AALG (5 mg) was dissolved in MeOD/D_2O 1:1 (500 μL) and irradiated by a UV lamp (365 nm) overnight. The reaction mixture was concentrated to a minimum volume with a flow of nitrogen and redissolved in MeOH:H_2O:CH_3COOH at a 1:1:0.01 ratio and for electrospray.



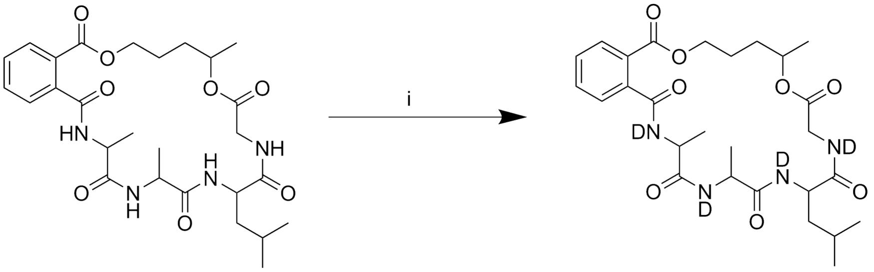
The product of post-photolysis D/H back exchange (D_1-XIII) was characterized by MS/MS: *m/z* 548 (M+H)$^+$, 530 (–H_2O), 473 (–GlyOH), 461 (–C_5H_9DO), 386 (*m/z* 461–GlyOH), 360 (–Leu–GlyOH), 273 (*m/z* 360–C_5H_9DO).

2.2.2.7 Procedure for H/D Exchange of the Cyclized Peptide

The cyclic peptide was dissolved in MeOD:D_2O 1:1 and diluted with MeOD: D_2O: CD_3COOD in a 1:1:0.01 ratio for electrospray.



The product (D_4-XIII) was characterized by MS/MS: m/z 552 (D_4-M+D)$^+$, 532 (–D_2O), 524 (–CO), 494, 475 (–[D_2]GlyOH), 466 (–$C_5H_{10}O$), 446, 388, 361 (–Leu–GlyOH), 289, 275.



Scheme 2.8: Reagents and conditions (i) MeOD:D_2O 1:1.

2.2.3 MS and NMR Characterization of Synthesized Products

2.2.4 Calculations

Born-Oppenheimer molecular dynamics (BOMD) calculations were run as 20 ps trajectories with 1 fs steps at 350-600 K using PM6-D3H4 calculations.[18] The BOMD calculations were run under the high-level Cuby4 platform.[19] The initial ion structures were constructed with different protonation sites that were at the peptide amide carbonyls for the non-basic peptides. In addition to thermal conformational motion, BOMD allows bond changes to proceed along the trajectory, namely, proton migration between different sites in the ion. The 20,000 structures generated by BOMD were sampled at regular intervals and 200 selected snapshots were fully optimized by PM6-D3H4. This augmented semiempirical procedure has been shown to capture hydrogen bonding and dispersion interactions,[20] and we have used it previously for structure analysis of several gas-phase peptide ions.[8]-[14] The PM6-D3H4-optimized structures were sorted out by type and energy, and 15-20 low-energy ions were reoptimized with density functional theory (DFT) calculations, using B3LYP[21] and the 6-31+G(d,p) basis set and including harmonic frequency analysis. Several lowenergy B3LYP structures were reoptimized with M06-2X[22] and the 6-31+G(d,p) basis set. These were used for single-point energy calculations that were carried out with M06-2X/6-311++G(2d,p) to provide relative energies that were combined with B3LYP zero-point vibrational energies, enthalpies, and entropies to obtain relative Gibbs energies for ion conformers and protomers. Ion solvation energies were estimated from self-consistent reaction field polarizable continuum model calculations (PCM)[23] in water dielectric with full gradient optimization by M06-2X/6-31+G(d,p) and single-point M06-2X/6-311++G(2d,p) energies. These standard DFT calculations were carried out with the Gaussian 16 suite of programs (Revision B.01) that was licensed from Gaussian, Inc. (Wallingford, CT). The lowest-energy ion structures from the

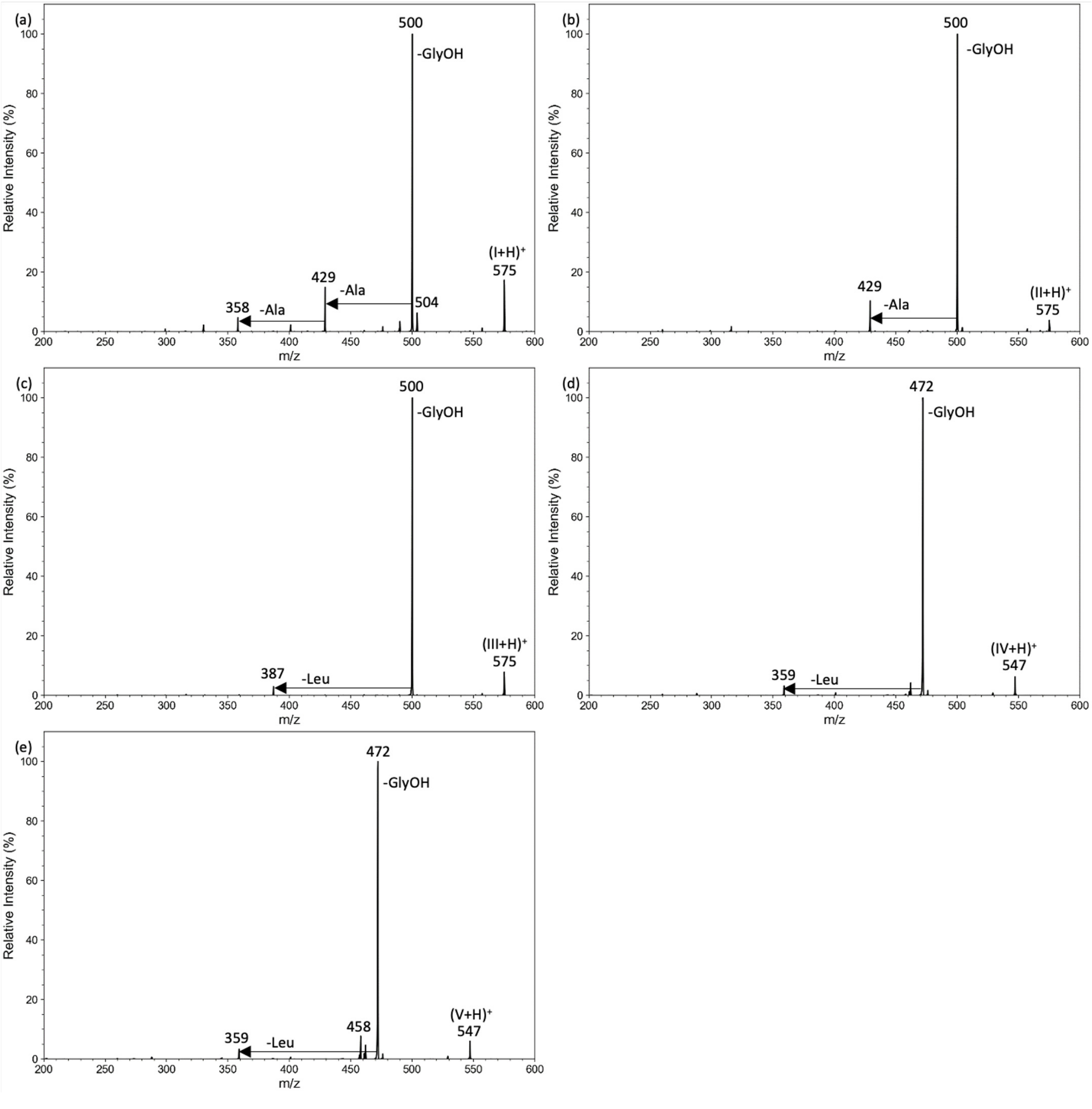


Figure 2.1: CID-MS^2 mass spectra of (a) (I+H)$^+$ ion at *m/z* 575, (b) (II+H)$^+$ ion at *m/z* 575, (c) (III+H)$^+$ ion at *m/z* 575, (d) (IV+H)$^+$ ion at *m/z* 547, (e) (V+H)$^+$ ion at *m/z* 547.

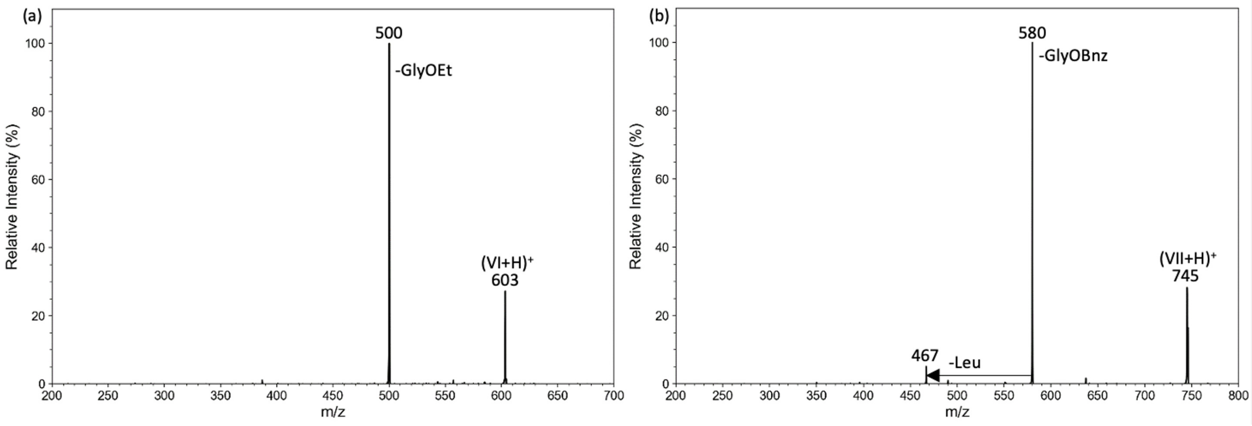


Figure 2.2: CID-MS^2 mass spectra of (a) (VI+H)$^+$ ion at *m/z* 603, and (b) (VII+H)$^+$ ion at *m/z* 745.

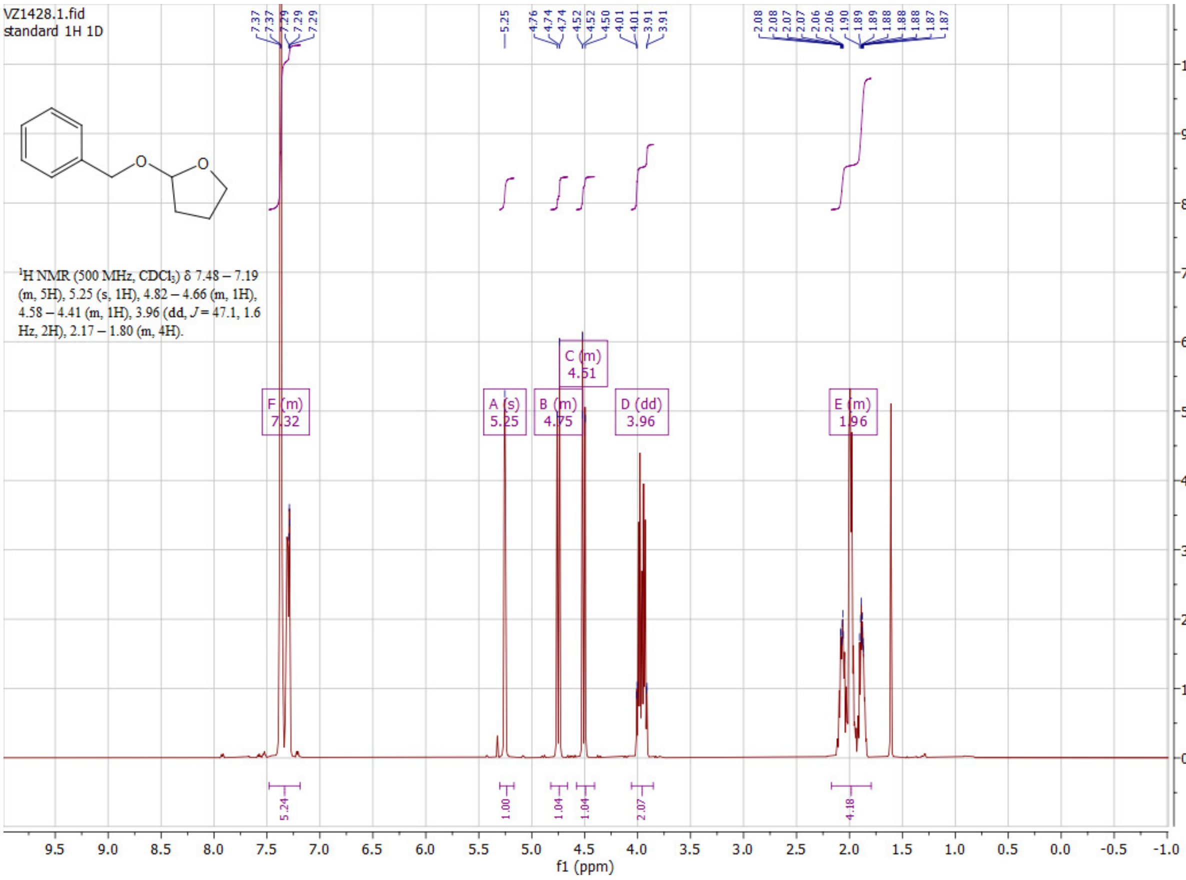

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.48 – 7.19 (m, 5H), 5.25 (s, 1H), 4.82 – 4.66 (m, 1H), 4.58 – 4.41 (m, 1H), 3.96 (dd, J = 47.1, 1.6 Hz, 2H), 2.17 – 1.80 (m, 4H).

Figure 2.3: ^{1}H NMR spectrum of VIII.

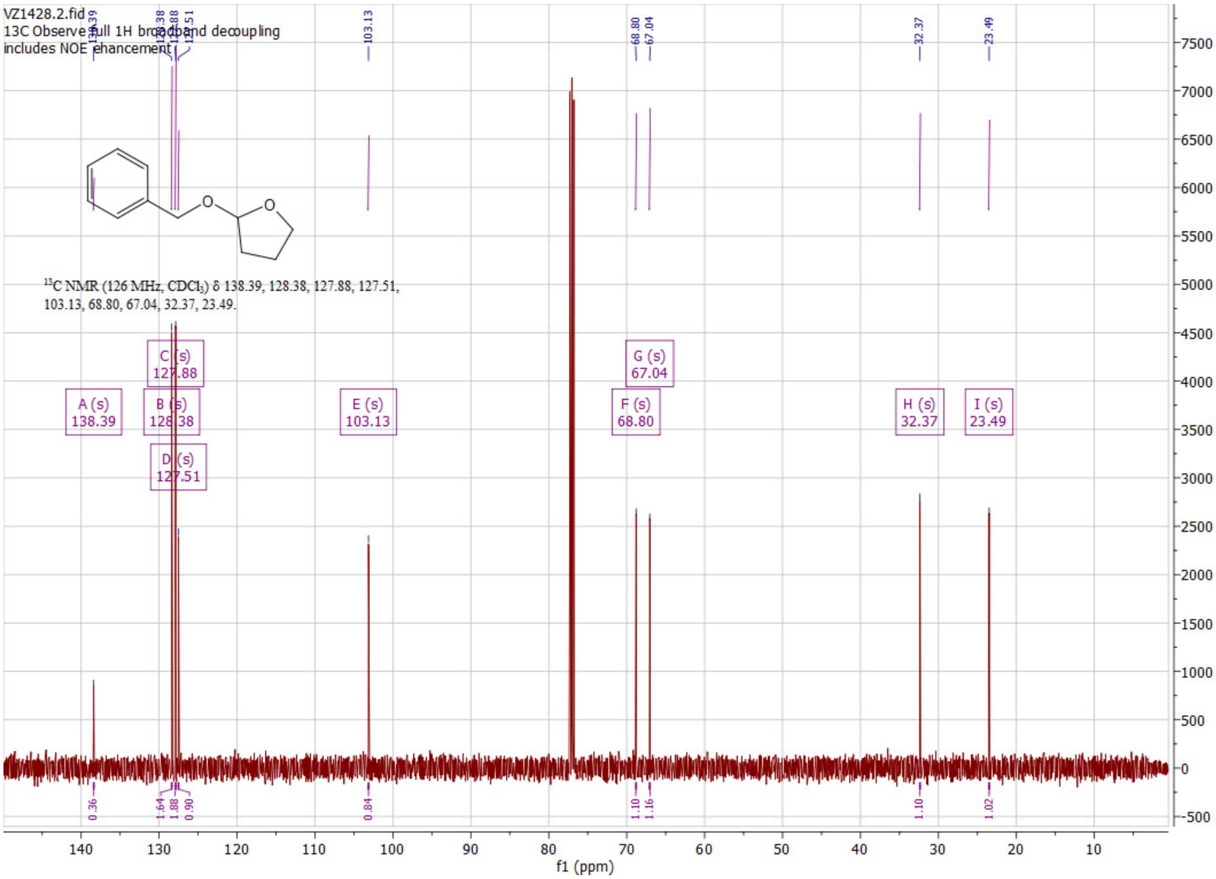
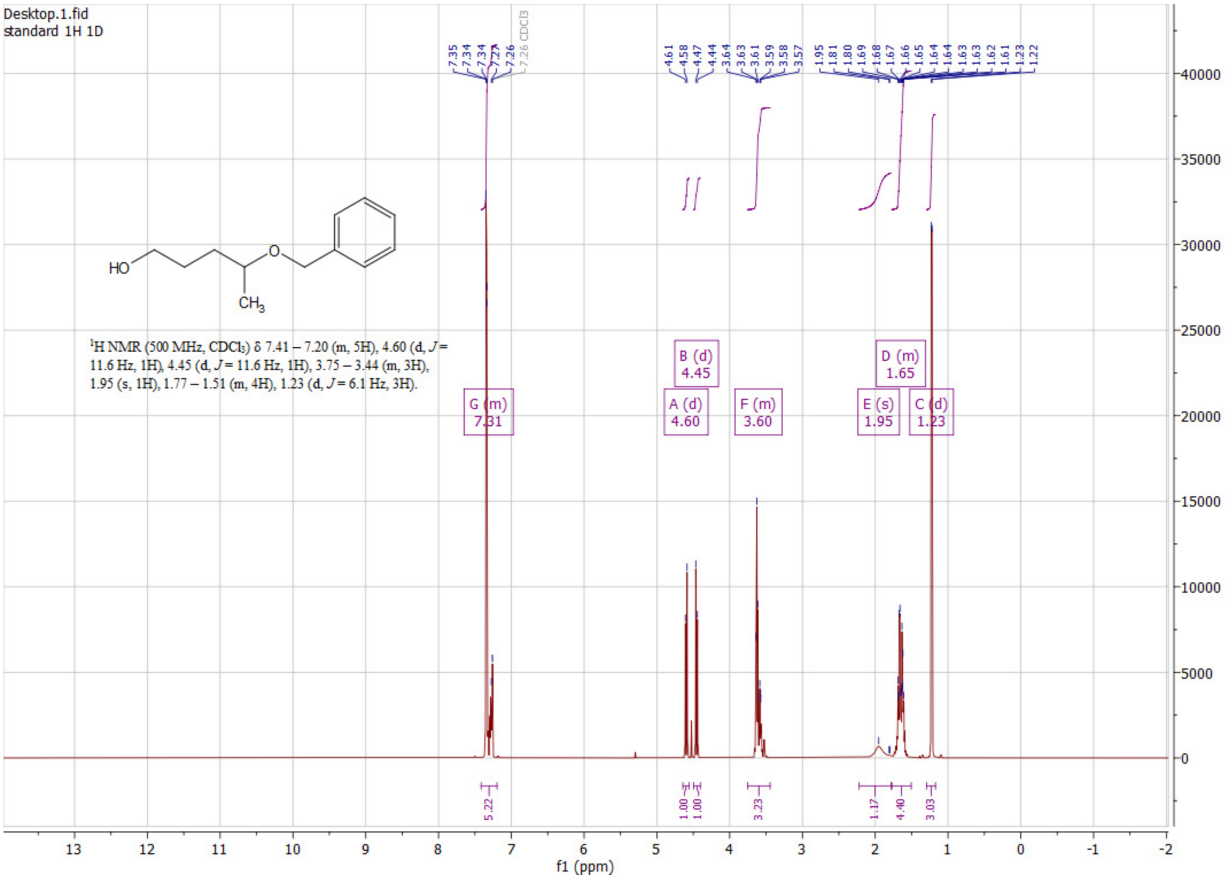


Figure 2.4: ^{13}C NMR spectrum of VIII.



^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 – 7.20 (m, 5H), 4.60 (d, *J* = 11.6 Hz, 1H), 4.45 (d, *J* = 11.6 Hz, 1H), 3.75 – 3.44 (m, 3H), 1.95 (s, 1H), 1.77 – 1.51 (m, 4H), 1.23 (d, *J* = 6.1 Hz, 3H).

Figure 2.5: ^{1}H NMR spectrum of IX.

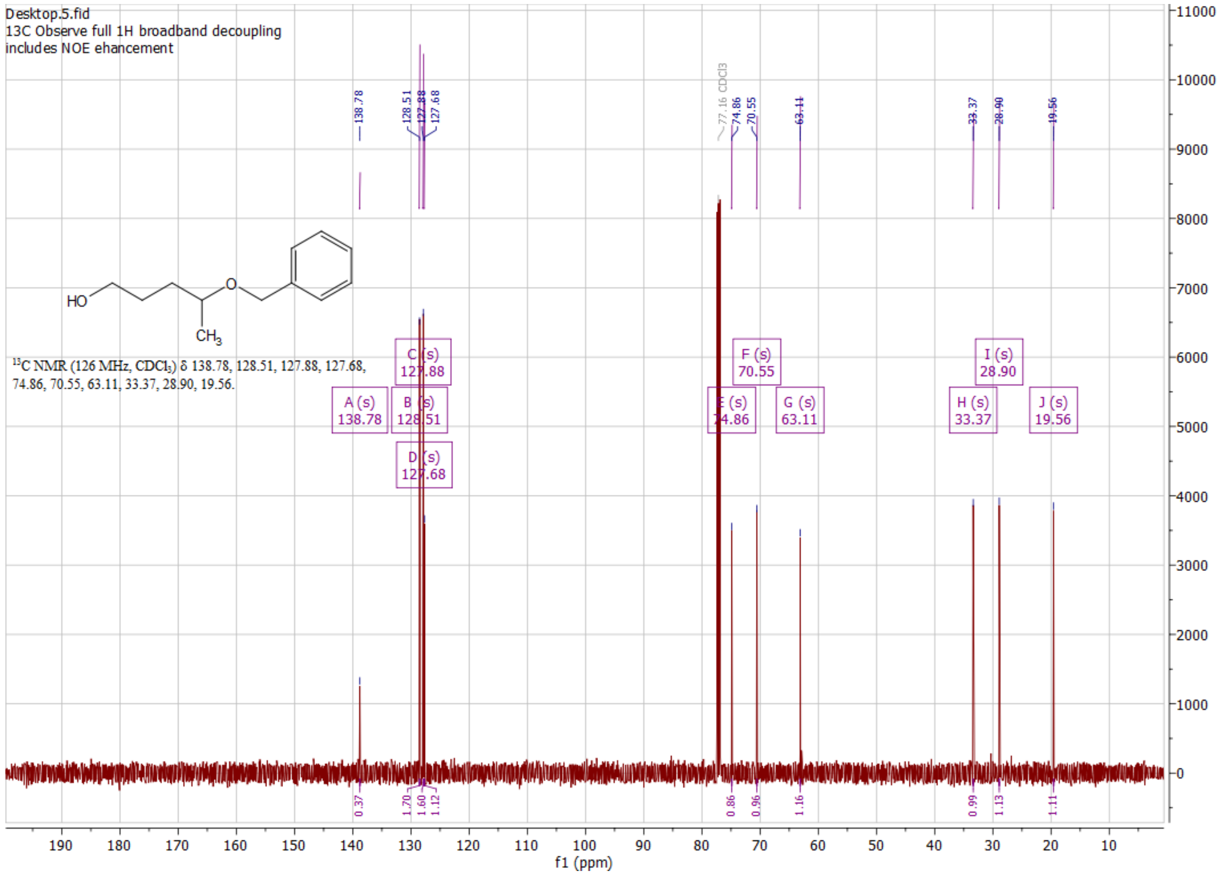
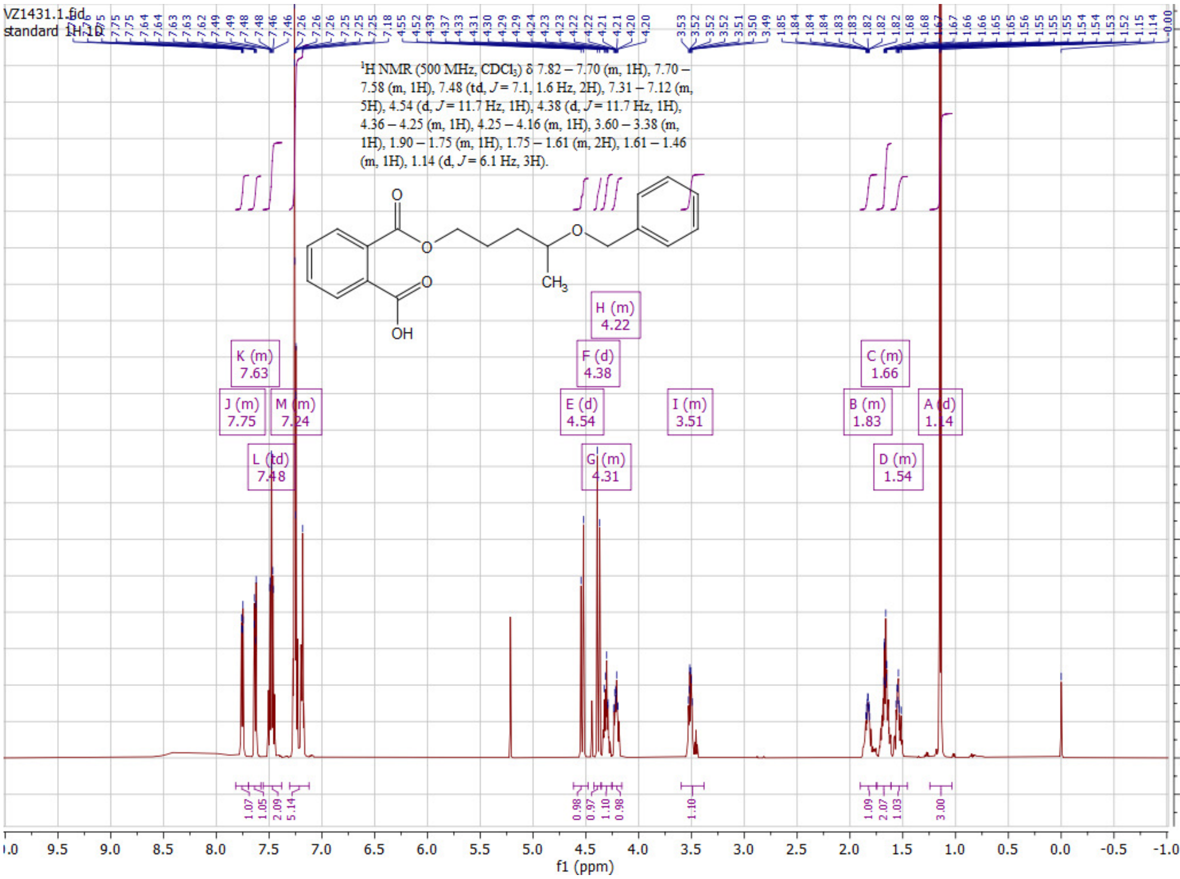
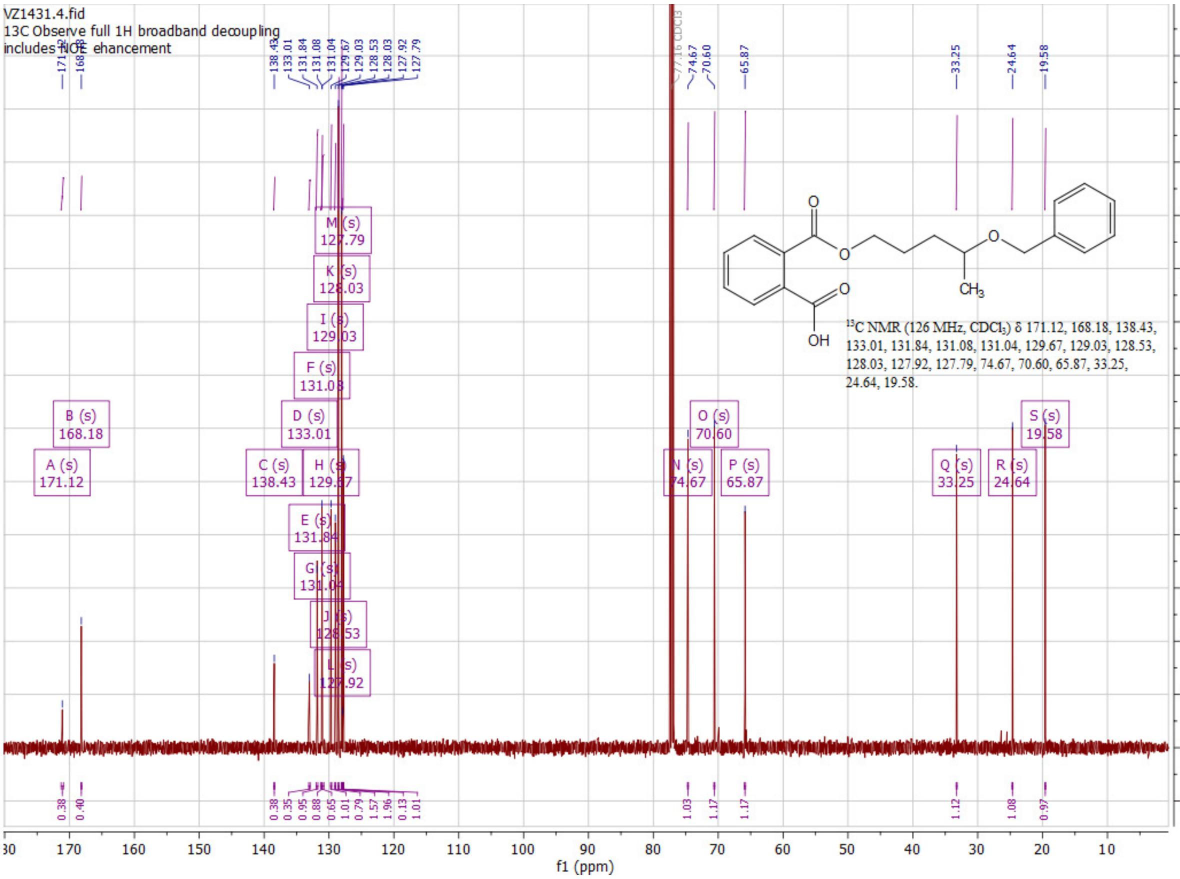


Figure 2.6: ^{13}C NMR spectrum of IX.



^{1}H NMR (500 MHz, CDCl$_3$) δ 7.82 – 7.70 (m, 1H), 7.70 – 7.58 (m, 1H), 7.48 (td, J = 7.1, 1.6 Hz, 2H), 7.31 – 7.12 (m, 5H), 4.54 (d, J = 11.7 Hz, 1H), 4.38 (d, J = 11.7 Hz, 1H), 4.36 – 4.25 (m, 1H), 4.25 – 4.16 (m, 1H), 3.60 – 3.38 (m, 1H), 1.90 – 1.75 (m, 1H), 1.75 – 1.61 (m, 2H), 1.61 – 1.46 (m, 1H), 1.14 (d, J = 6.1 Hz, 3H).

Figure 2.7: ^{1}H NMR spectrum of X.



^{13}C NMR (126 MHz, CDCl$_3$) δ 171.12, 168.18, 138.43, 133.01, 131.84, 131.08, 131.04, 129.67, 129.03, 128.53, 128.03, 127.92, 127.79, 74.67, 70.60, 65.87, 33.25, 24.64, 19.58.

Figure 2.8: ^{13}C NMR spectrum of X.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.35 (s, 5H), 5.18 (s, 2H), 4.97 (d, *J* = 26.0 Hz, 1H), 3.95 (d, *J* = 5.7 Hz, 2H), 1.44 (s, 9H).

Figure 2.9: ^{1}H NMR spectrum of XI.

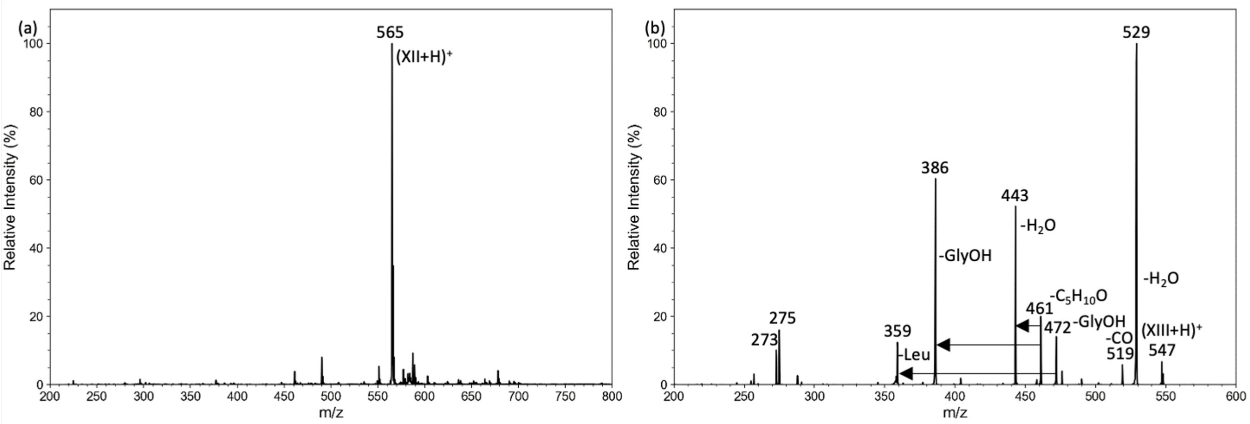


Figure 2.10: (a) Mass spectrum of $(XII+H)^+$ ion, and (b) CID-MS^2 mass spectrum of $(XIII+H)^+$ ion at *m/z* 547.

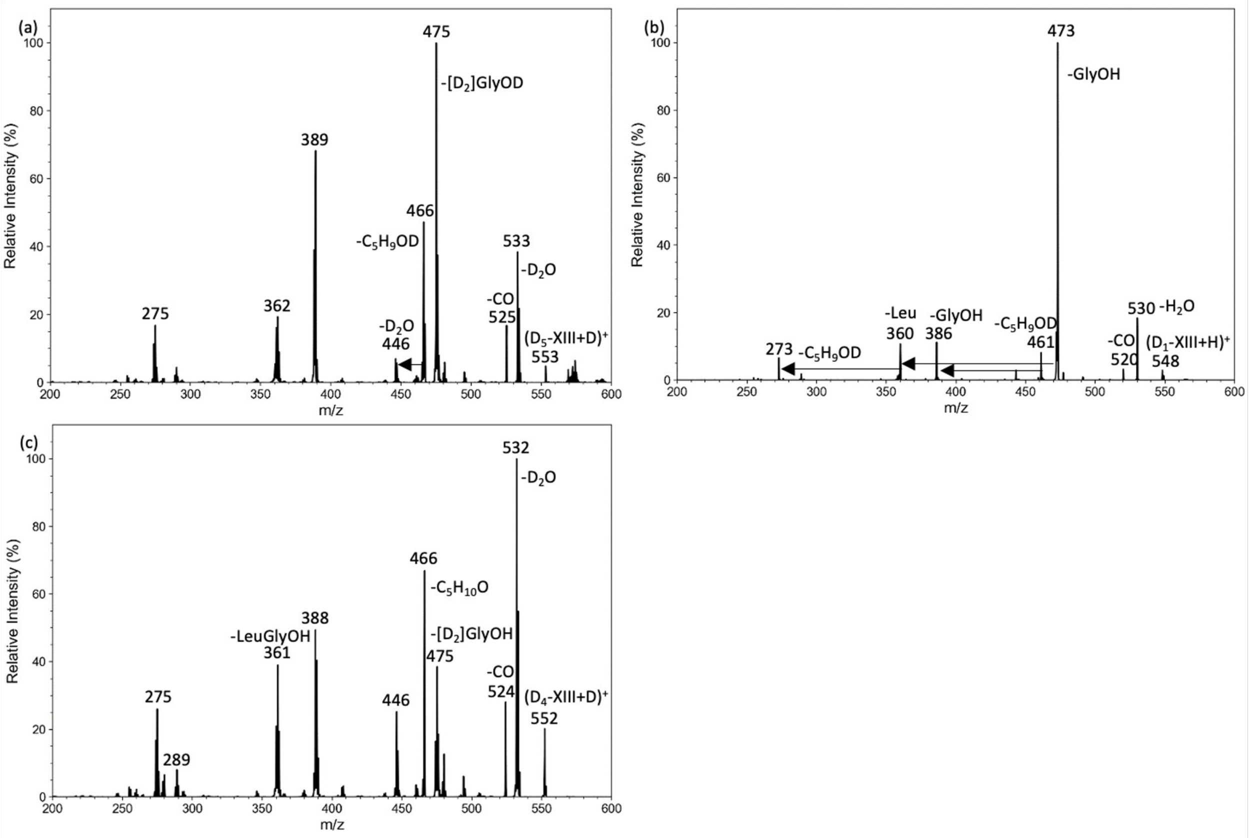


Figure 2.11: CID-MS^2 mass spectra of (a) (D$_5$-XIII+D)$^+$ ion at *m/z* 553, (b) (D$_1$-XIII+H)$^+$ ion at *m/z* 548, and (c) (D$_4$-XIII+D)$^+$ ion at *m/z* 552.

M06-2X calculations were then used as initial geometries for 100 ps BOMD trajectory runs using Cuby4, that each created 100,000 snapshots for atom-atom contact analysis.

2.3 Results and Disscussion

2.3.1 Photodissociation and Crosslinking Yields

Scaffold ions 4,4-Azipent-1-yl-s-LAAG (s-LAAG), 4,4-Azipent-1-yl-s-ALAG (s-ALAG), and 4,4-Azipent-1-yl-s-AALG (s-AALG) (*m/z* 575) were produced by electrospray protonation from aqueous-methanol solutions and selected by mass in the ion trap. Photodissociation at 355 nm resulted in elemination of N_2 from the diazirine ring and formation of product ions (Figure 2.12). Multiple (14) laser pulses were used to achieve conversions that produced photodissociation ion counts allowing further ions analysis by multistage CID-MS^n. Loss of N_2 was accompanies by consecutive dissociations that were driven by the substantial exergonicity of carbene insertion[14] or isomerization[24]. Tetrapeptide scaffolds are the smallest ones to provide fractions of stable $(MH - N_2)^+$ ions for further CID-MS^n studies.

The fragment ions formed through photodissociation were identified using CID-MS^n spectra of the surviving $(MH - N_2)^+$ ions. Additionally, to help with this identification, representative synthetic 3-pentene and 4-pentene derivatives of s-AALG illustrating the byproducts of carbene isomerization to alkenes were analyzed and compared with UVPD spectra (Figure 2.1 (d) and (e)). For s-LAAG, s-ALAG, and s-AALG, CID-MS^n of their pentene esters showed simple peptide backbone cleavages resulting in the sequential loss of the amino acid residues starting with GlyOH and followed by Leu or Ala according to the particular sequence to give $\mathbf{b}_n{}^+$ ions, where n represents the number of residues in the fragment counting from the N-terminus for b-type ions. These fragment ions were used to identify and quantify crosslinks and non-crosslink olefin byproducts from diazirine photodissociation.

Photodissociation spectra of s-LAAG, s-ALAG, and s-AALG gave conversions ranging between 16 and 35% (Table 2.1). The overall conversion depended on the laser power and overlap of the ion cloud with the laser beam in the ion trap that varied in measurements taken on different days. Internal normalization of the photoproducts showed nondissociating (survivor) fractions of 4.5-4.7% for s-LAAG, s-ALAG, and s-AALG (Table 2.1). The orthogonal nature of the crosslink and non-crosslink CID-MS^n spectra allowed us

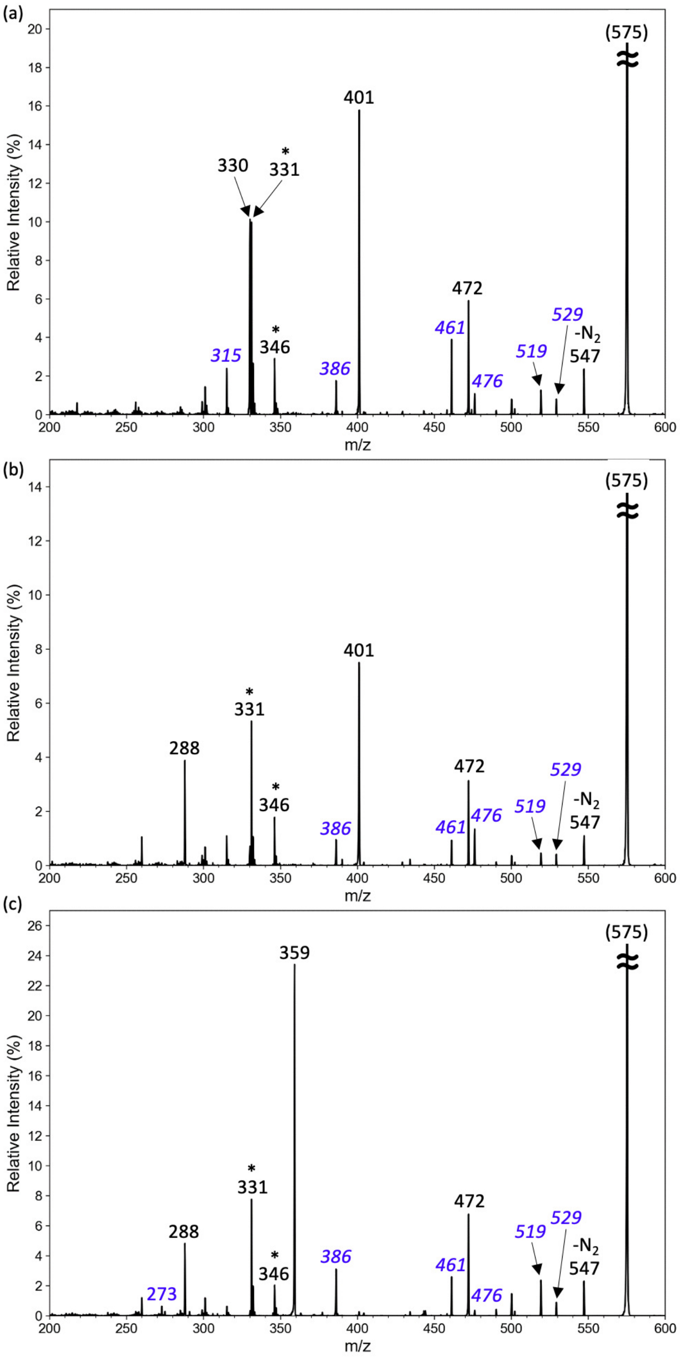


Figure 2.12: UVPD-MS^2 mass spectra (355 nm) of *m/z* 575 from protonated (a) s-LAAG, (b) s-ALAG, and (c) s-AALG. Fragment ions originating from crosslinked photoproducts are annotated by blue italic numerals. Ions from laser desorption-ionization of surface material are denoted by asterisks.

to clearly distinguish these populations. Photodissociation of s-LAAG, s-ALAG, and s-AALG was found to produce 21-26% fractions of crosslinked ions (Table 2.1), showing very similar crosslinking yields. These yields from the peptide scaffolds were in the middle range of yields typically obtained from gas-phase noncovalent peptide-peptide complexes[8]-[12] and orders of magnitude higher than yields of crosslinks for equimolar carbene-substrate ratios in solution[25][26].

Table 2.1: Photodissociation Conversion and Crosslink Yields

Scaffold	Laser Pulses (%)	UVPD Conversion^a (%)	Survivor Fraction^b (%)	Crosslink Fraction^c (%)
s-LAAG	14	29	4.6	26
s-ALAG	14	16	4.5	25
s-AALG	14	35	4.7	21
s-AALG-OEt	36	65	2.5	13

[a] Fraction of total UVPD produced ions.
[b] Fraction of nondissociating $(MH – N_2)^+$ ions relative to the total photoproduct ion intensities.
[c] Fraction of photoproducts identified as originating from crosslinks.

2.3.2 CID-MS^n Analysis of Crosslinked Photoproducts

The $(MH – N_2)^+$ ions from diazirine photodissociation were then investigated by CID-MS^3 (Figure 2.13) and further CID-MS^4 of each prominent fragment from the corresponding MS^3 spectra (Table 2.2-2.4). The $(MH– N_2)^+$ ions at *m/z* 547 from s-LAAG, s-ALAG, and s-AALG gave very similar fragmentation patterns in which the fragment ions were identified on the basis of accurate mass measurements of analogous $(MH– N_2)^+$ ions that were separately generated by CID-MS^2 of the scaffolds. The conspicuous feature of the CID-MS^3 spectra (Figure 2.13) was that they showed a very minor loss of GlyOH (*m/z* 472), which was the dominant dissociation of non-crosslinked scaffold ions. Instead, the spectra showed losses of water, CO, and $C_5H_{10}O$ from the side chain (*m/z* 461) and a loss of an out-of-sequence alanine residue at *m/z* 476 and 458. These features, including sequence scrambling, are typical of cyclic peptides[27][28], affirming that the ions were internal crosslinks in which the peptide and alkyl side chains were connected by a covalent bond. The loss of $C_5H_{10}O$ from the side chain was combined with loss of GlyOH (*m/z* 386) and dipeptide sequence residues LeuGlyOH (*m/z* 273) and AlaGlyOH (*m/z* 315) (Figure 2.13, respectively). This interpretation was corroborated by the CID-MS^4 spectra of m/z 461 ions that showed the pertinent peptide sequence fragment ions as dominant products (Figure 2.14). The facile loss of $C_5H_{10}O$ upon CID indicated that the C–X bond forming the crosslink to the peptide chain was susceptible to proton driven cleavage. This feature is typical of ester C–O and amide C–N bonds but not aliphatic C–C bonds, suggesting that the carbene insertion into the LAAG, ALAG, and AALG sequences involved the peptide amide N–H and/or glycine carboxyl O–H bonds.

Table 2.2: CID-MS^4 Fragments of $(MH - N_2)^+$ *m/z* 547 from s-LAAG Following UVPD

529 –H_2O	519 –CO	476 –Ala	472 –GlyOH	461 –$C_5H_{10}O$	458 –AlaOH	386 461–GlyOH	315 386–Ala
511	502	458	401	443	401	371	287
484	476	448	330	386	390	358	244
454	474	405		358	383	315	170
443	416	401		315	372	274	
416	377	390		287	301		
386	372	371			297		
364	309	333			243		
315		330					
		315					

Table 2.3: CID-MS^4 Fragments of $(MH - N_2)^+$ *m/z* 547 from s-ALAG Following UVPD

529 –H_2O	519 –CO	476 –Ala	472 –GlyOH	461 –$C_5H_{10}O$	434 –Leu	401 472–Ala	386 461–GlyOH	315 386–Ala
520	502	458	401	443	416	288	371	287
511	476	448	288	404	359		358	
507	434	419		386	348		315	
501	309	401		358	288		287	
470	148	390		315	273		201	
458		330		287				
443		315						
386		288						

2.3.3 H/D Exchange in s-AALG Reveals Proton Involvement in Crosslinking Mechanism

The CID-MS^n analysis of the photodissociation products from s-LAAG, s-ALAG, and s-ALAG ions in the gas phase indicated the possibility of the formation of crosslinks with cyclized lactone or amide structures. Despite the different positions of Ala and Leu residues, the fragemntation pattern in both CID and UVPD are the same. Further experiments were carried out to investigate the crosslink's structure for s-AALG in more details. According to the standard mechanism of X–H bond insertion into carbenes (Scheme

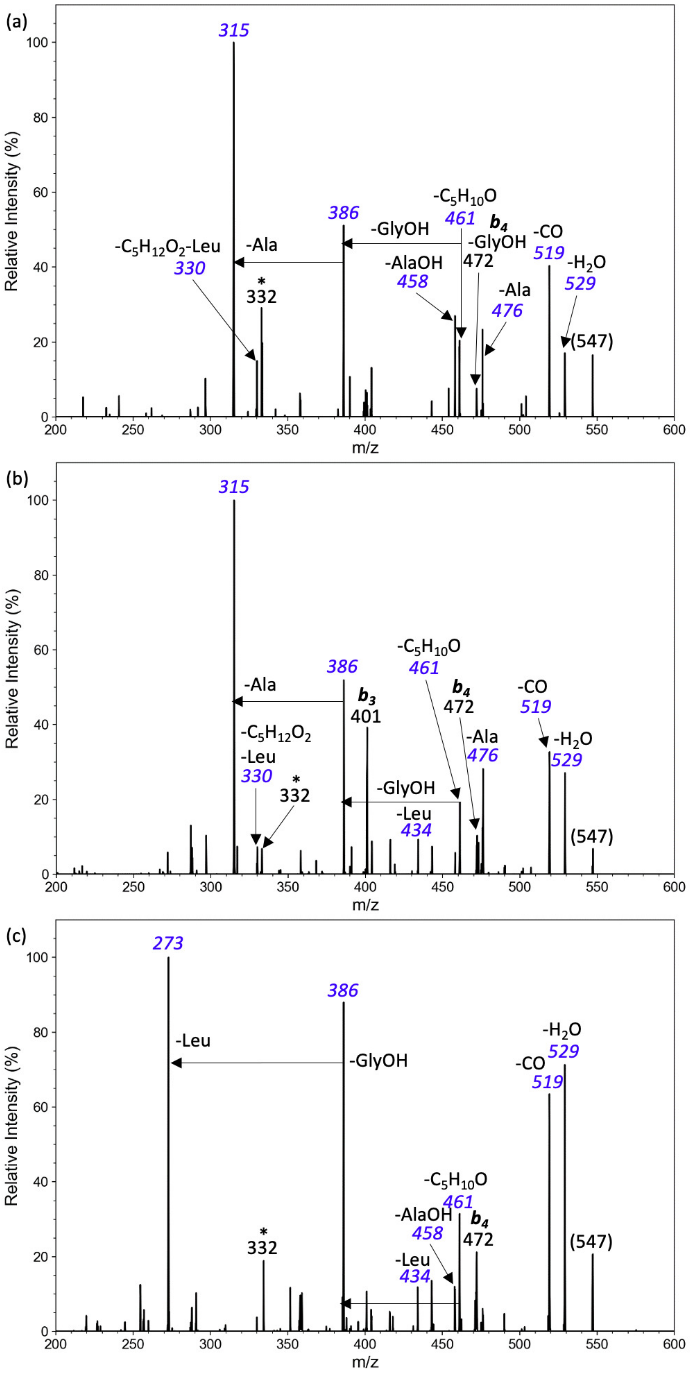


Figure 2.13: CID-MS^3 mass spectra of *m/z* 547 from photodissociation of (a) s-LAAG, (b) s-ALAG, and (c) s-AALG. Fragment ions and neutral losses were assigned from accurate mass measurements. For ion color coding see text in Figure 2.12.

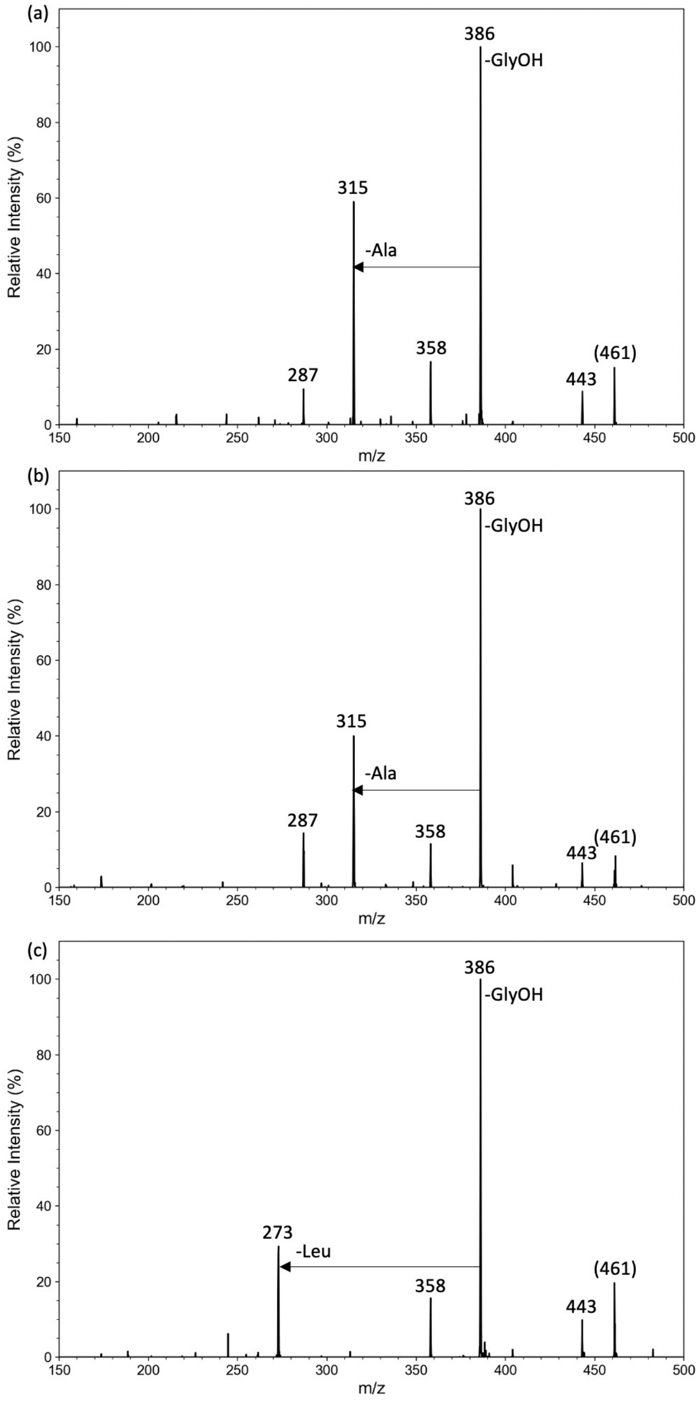


Figure 2.14: CID-MS^4 mass spectra of (MH – $C_5H_{10}O$)$^+$*m/z* 461 from (a) s-LAAG, (b) s-ALAG, and (c) s-AALG following UVPD.

Table 2.4: CID-MS^4 Fragments of $(MH - N_2)^+$ *m/z* 547 from s-AALG Following UVPD

529 $-H_2O$	519 –CO	472 –GlyOH	461 $-C_5H_{10}O$	458 –AlaOH	434 –Leu	386 461–GlyOH	273 386–Leu
511	502	443	443	443	416	371	
501	475	359	386	404	359	358	
459	434		358	386	348	315	
443	431		273	358	288	287	
386	348		245	315	273	201	
372	309			287			
359	288						
293	245						
255							
219							

2.1), the hydrogen atom involved is incorporated in a stable covalent C–H bond. This means that protons from polar O–H or N–H bonds that are prone to hydrogen-deuterium exchange in the reactant become nonpolar if involved in carbene insertion and cannot be further affected by prototropic H/D exchange.

To probe crosslinking in s-AALG, we converted all five exchangeable amide N–H and carboxyl O–H protons to deuterons under mild nonacidic conditions and subjected the sample to UV photolysis in CD_3OD/D_2O solution (Schemes 2.6 and 2.7). The products were then converted back to the protic cycle and analyzed by tandem mass spectrometry. The electrospray mass spectrum showed partial retention of deuterium in the $(MH - N_2)^+$ ion at *m/z* 548, whereas the main fraction was back exchanged (*m/z* 547, Figure 2.15(a), inset). CID-MS^2 of the *m/z* 548 ion indicated that the products of solution photolysis chiefly consisted of alkene isomers, as judged by the loss of glycine and LeuGlyOH residues, giving rise to the respective *m/z* 473 and 386 fragment ions. We note that the *m/z* 548 ion contained ca. 40% of combined ^{13}C and ^{15}N isotopologues of the *m/z* 547 ion that were unresolved from the *m/z* 548 D_1 species and contributed to the formation of the *m/z* 473 and 386 fragment ions. However, the *m/z* 461 fragment ion in the CID-MS^2 spectrum signified a loss of C_5H_9DO that was further corroborated by the combined loss of C_5H_9DO and LeuGly (*m/z* 273, Figure 2.15(a)). The CID-MS^3 spectrum of the *m/z* 461 ion (Figure 2.15(b)) confirmed there was no deuterium present after the loss of C_5H_9DO. This established that an exchangeable proton was incorporated into the transient carbene upon photolysis.

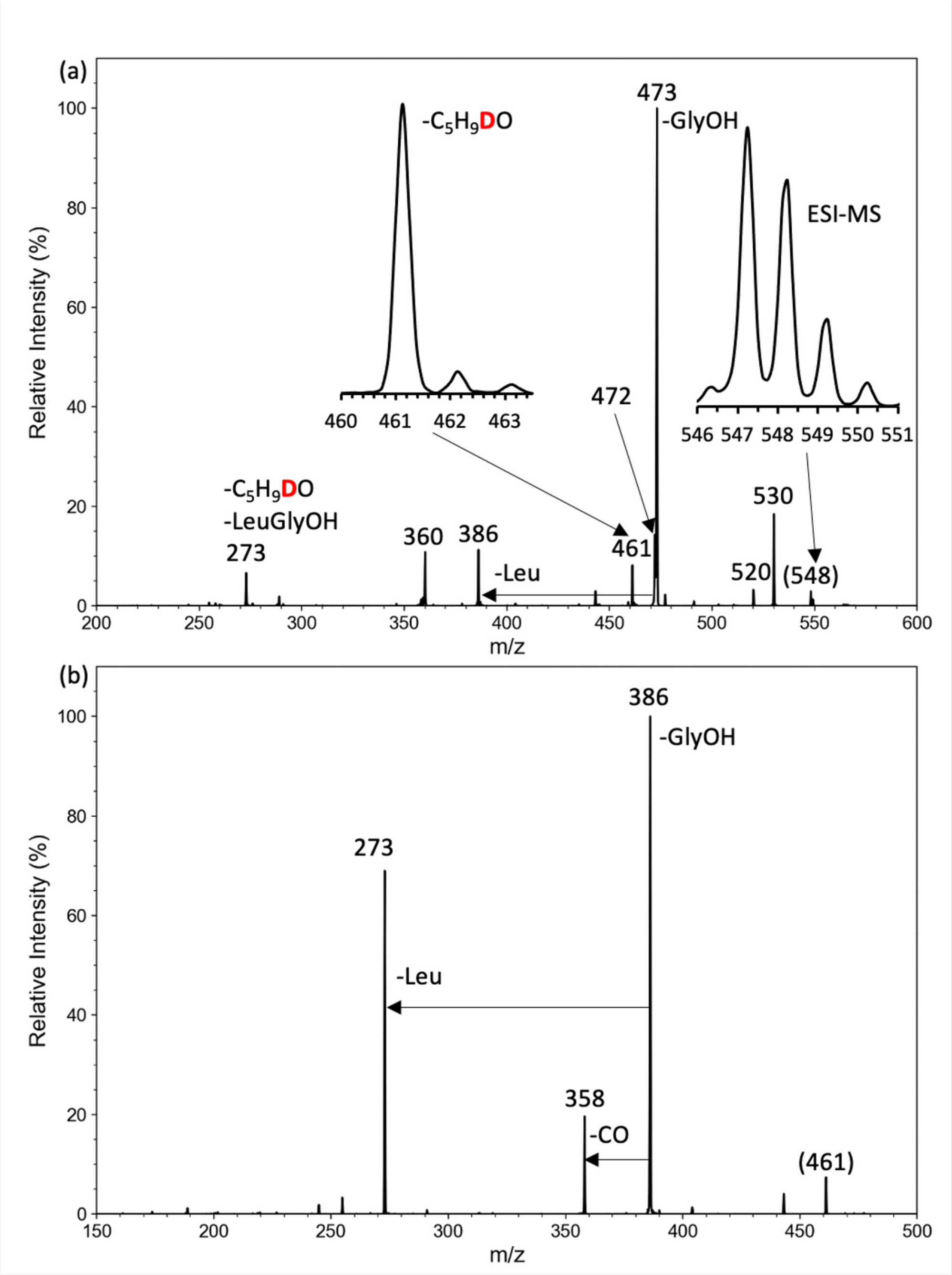


Figure 2.15: (a) CID-MS^2 mass spectra of *m/z* 548 D_1-containing ion from solution photolysis of 4,4-azipent-1-yl-[D_5]-AALG and back D → H exchange. Insets show the precursor ion cluster at *m/z* 547-551 with theoretical isotope distribution, and the peak profile of the *m/z* 461 fragment ion. (b) CID-MS^3 of *m/z* 461.

2.3.4 Dertermination of Crosslinking Sites in s-AALG

To further specify the peptide functional group engaged in crosslinking, we studied gas-phase photodissociation of s-AALG-OEt in which the C-terminal carboxyl group was blocked as an ethyl ester. Table 2.1 data show that photodissociation of s-AALG-OEt proceeded with a smaller survivor $(MH – N_2)^+$ and crosslink fractions when compared to those from s-AALG. This ca. 40% decrease was interpreted as being due to the formation of alkene products caused by the carboxyl blockage. This was corroborated by both the UVPD-MS^2 of s-AALG-OEt at *m/z* 603 and CID-MS^3 of the $(MH – N_2)^+$ ion at *m/z* 575, as shown in Figure 2.16(a) and (b), respectively.

UVPD of s-AALG-OEt showed substantial dissociation of the peptide chain which was typical of the alkene products of N_2 elimination. CID-MS^3 of the $(MH – N_2)^+$ ion showed less abundant *m/z* 472 (loss of GlyOEt) and *m/z* 359 (loss of LeuGlyOEt) sequence ions that would be indicative of an alkene ion where this dissociation predominated (Figure 2.16(c)). Prominent dissociations in the ester group by loss of ethylene and ethanol, *m/z* 547 and 529, respectively, were accompanied by nonsequence loss of Ala (*m/z* 504) and Leu (*m/z* 462) that indicated the presence of crosslinks. Remarkably, loss of $C_5H_{10}O$ from the side chain (*m/z* 386) was observed only following the loss of GlyOEt, so this fragmentation could not be used as a crosslink signature.

To further elucidate the carboxyl participation in carbene quenching, we synthesized an authentic lactone corresponding to the Gly-carboxyl crosslinked product, denoted as cyc-s-AALG. The CID-MS^2 spectrum of protonated cyc-s-AALG (*m/z* 547) showed fragment ions by loss of water, GlyOH, $C_5H_{10}O$, $(C_5H_{10}O + H_2O)$, $(C_5H_{10}O + \text{GlyOH})$, and LeuGlyOH at *m/z* 529, 472, 461, 443, 386, and 359, respectively (Figure 2.17(a)). These fragment ions, which were characteristic of a cyclic structure, also appeared in the CID-MS^3 spectrum of the s-AALG UVPD product, although with different relative intensities (Figure 2.17(b)). In particular, the UVPD product showed a more prominent loss of CO (*m/z* 519) and the combined loss of $C_5H_{10}O$ and LeuGlyOH (*m/z* 273). On the basis of the Gly ethyl ester blocking results (Table 2.1), we estimated the content of the lactone form of cyc-s-AALG at ca. 40%, which allowed us to subtract this contribution from the spectrum of the UVPD product. The resulting renormalized difference spectrum (Figure 2.17 (b)) showed the dominant dissociations by combined losses of $C_5H_{10}O$ and the GlyOH and LeuGlyOH fragments. Along with the results of deuterium labeling, this indicated that the crosslinking

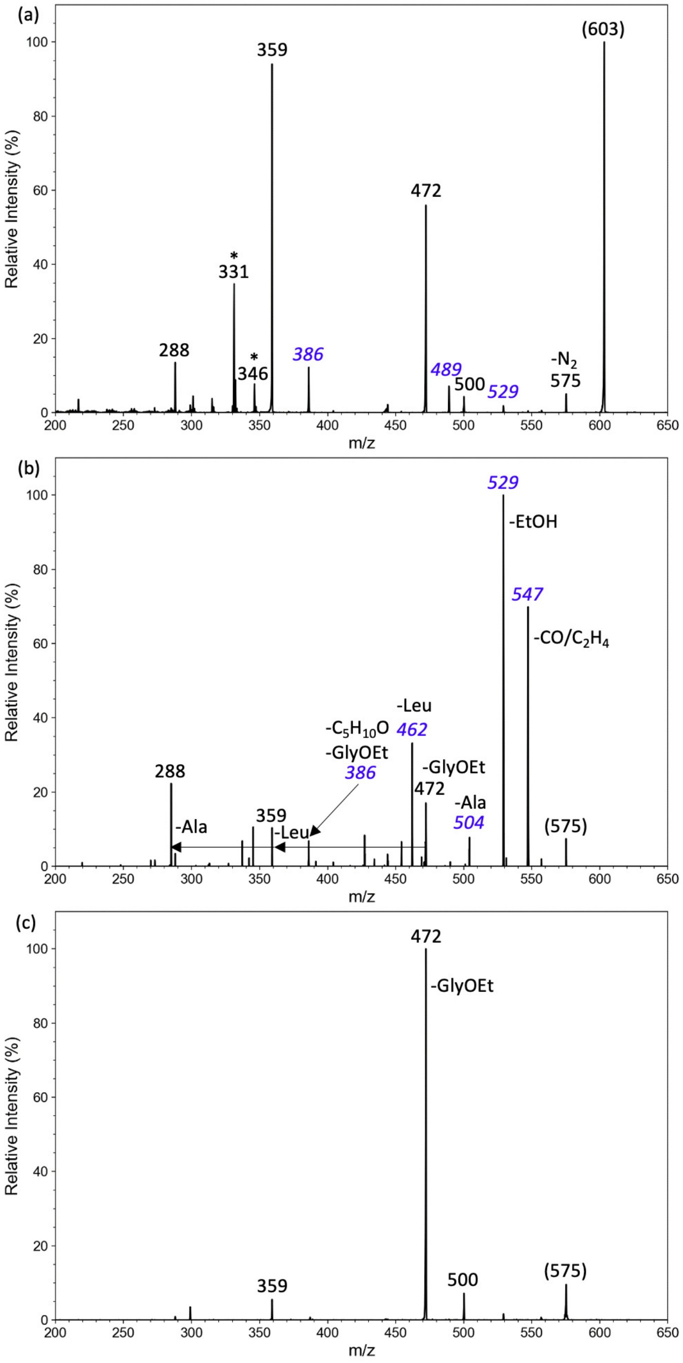


Figure 2.16: (a) CID-MS^2 mass spectrum of *m/z* 603 (M + H)$^+$ ion from 4,4-azipent-1-yl-s-AALG-OEt. (b) CID-MS^3 spectrum of the (MH – N_2)$^+$ ion at *m/z* 575 from UVPD. (c) CID-MS^2 spectrum of the *m/z* 575 (M + H)$^+$ ion from the pentenyl-s-AALG-OEt byproduct.

involved the N–H bond of the Gly amide. However, on the basis of the experimental data we cannot exclude crosslinking to the N–H bond of the Leu amide. It should be noted that increasing the amide group basicity by N-alkylation, as in the amide crosslinks, has been reported to enhance peptide bond cleavage[31], which was consistent with the facile backbone dissociations at the crosslinked residues in the Figure 2.17 (b) spectrum.

The dissociations of the cyc-s-AALG lactone isomer were further investigated with the help of deuterium labeling in the N–D exchanged D_4-derivative (Figure 2.11 (c), Scheme 2.8). The spectrum showed a combined loss of HDO and D_2O, indicating ester elimination and lactone ring opening that involved light hydrogens from the aliphatic side chain. Consistent with this, the losses of GlyOH and LeuGlyOH were comprised of D_1-D_3-containing molecules, indicating the involvement of light hydrogens. In contrast, the loss of $C_5H_{10}O$ did not involve any H/D exchange, which suggested that this neutral molecule was eliminated as a cyclic ether. Note that an elimination of $C_5H_{10}O$ as a pentenol would have necessarily involved a deuteron transfer onto the neutral molecule.

The results from H/D exchange and carboxyl blocking experiments can be summarized as follows. Polar amide and carboxyl groups significantly participate in crosslinking to the transient carbene produced in the scaffold side chain. A fraction, estimated at about 40% of polar crosslinks, is due to participation by the carboxyl OH group forming a lactone product. The other fraction involves one of the amide N–H bonds, specifically that between Leu and Gly in s-AALG, forming a cyclic amide product. Scheme 2.9 depicts the possible pathways from the carbene intermediate to the carboxyl and amide crosslinks. The crosslink creates a chiral center at the former carbene atom which most likely gives rise to racemic products. The efficiency of these pathways, which totals up to 21% according to Table 2.1 data, depends on the kinetics of the first step, which presumably consists of a proton transfer to the electron-rich carbene.[29][30] This step depends on the populations of ion conformations in thermal ions that determine the probability of the proton to appear at a distance to the carbene atom allowing C–H bond formation in the transient carbocation. Conformations in which the proton is remote from the carbene undergo competitive isomerization within the side chain by 1,2-hydrogen migration from the adjacent CH_2 or CH_3 group to form alkene isomers.

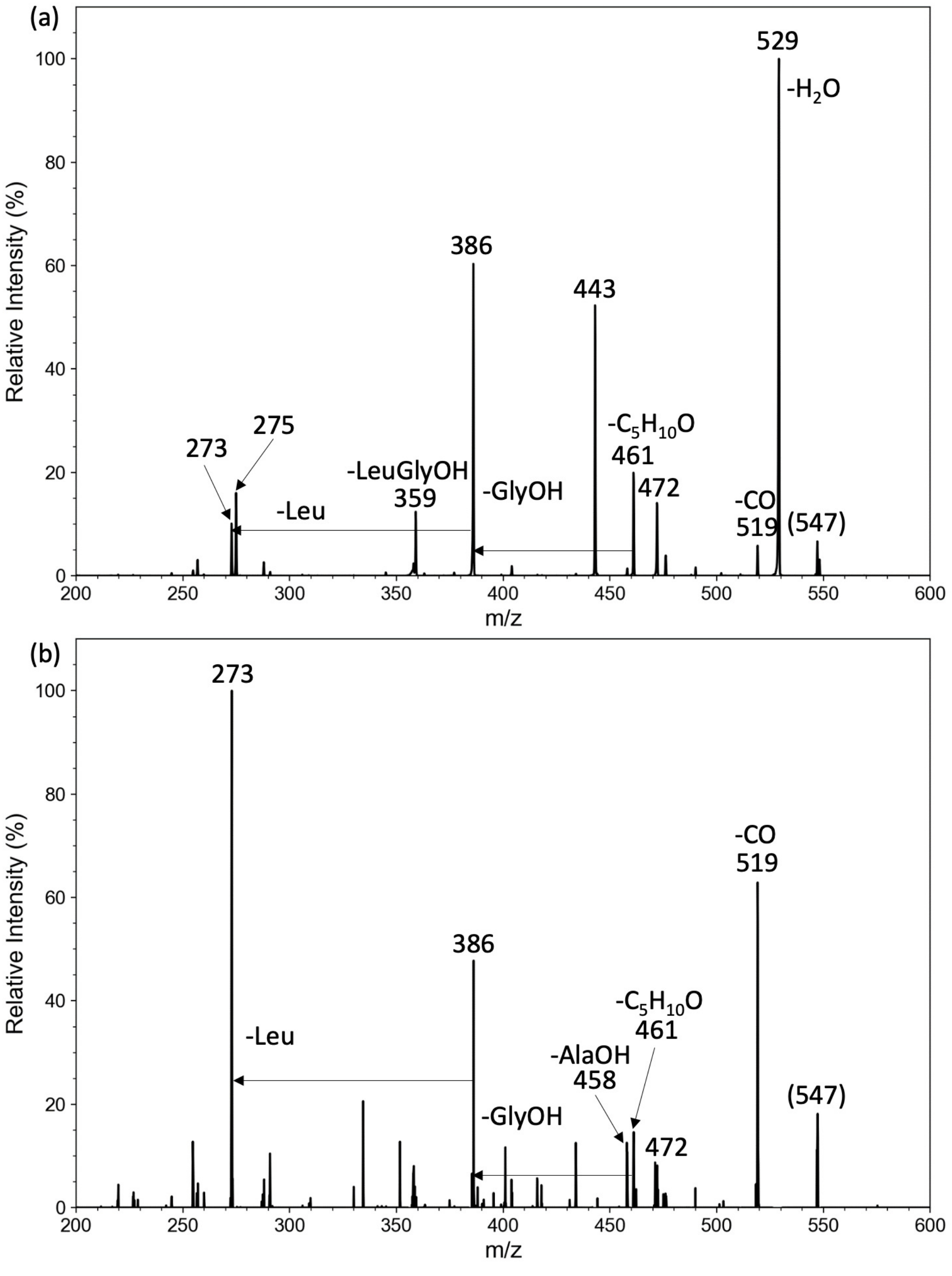
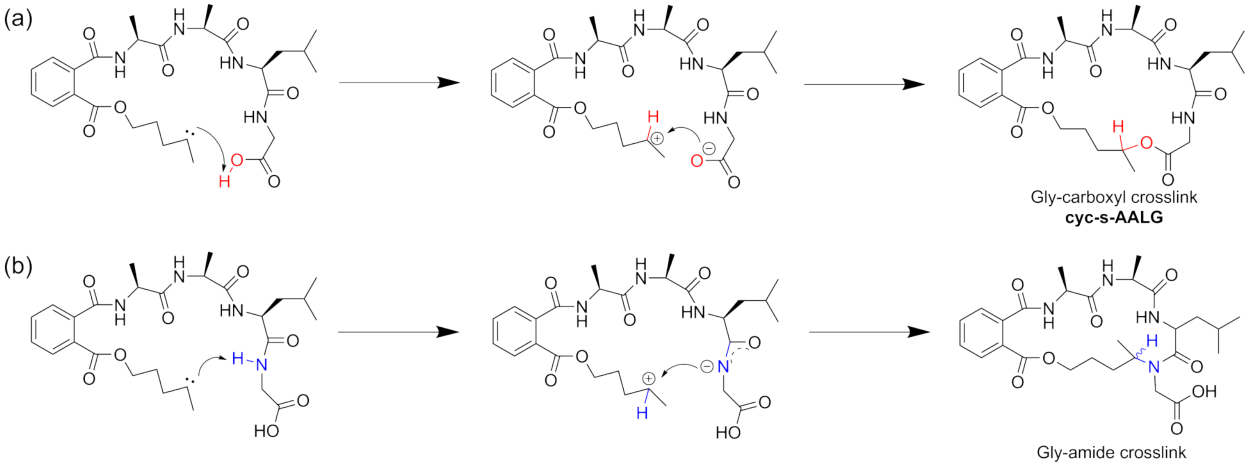


Figure 2.17: (a) CID-MS^2 mass spectrum of *m/z* 547 (M + H)$^+$ ion from cyc-s-AALG. (b) Difference spectrum of *m/z* 547 CID-MS from UVPD of s-AALG and cyc-s-AALG.



Scheme 2.9: Proposed mechanism of carbene interactions in s-AALG peptide scaffolds with (a) OH bonds and (b) NH bonds.

2.3.5 Structure of s-AALG Peptide Ions

To interpret the results of scaffold photodissociation and carbene crosslinking, we conducted extensive calculations of ion structures and dynamics. The primary goals of these calculations were to identify the most energetically favorable protonation sites and ion conformations of the diazirine-containing scaffolds and to analyze the dynamic interactions between the incipient carbene atom and peptide bonds in thermal ions. BOMD calculations of s-AALG ions were performed with initial structures protonated at the phthalate, Ala1, Ala2, and Leu amides. Several low Gibbs energy structures are presented in Figure 2.18. Structures initially protonated at Ala1, Ala2, and Leu collapsed during BOMD to form a single protomer, with the proton positioned at the Ala2 amide. In the lowest energy structures (**A1-A4**), the Ala2 proton was shared by a very short (1.39-1.47 Å) hydrogen bond with the Leu amide, indicating facile proton migration between these two amide groups. This hydrogen bond stabilized the conformation of the Ala2-Leu segment of the peptide chain, causing it to move as a unit during the thermal motion of the ions. Ions **A1-A4** primarily differed in the conformation of the leucine and diazirine-containing side chains. Structures **A5** and **A6** displayed distinct conformations, with the Ala2 proton hydrogen-bonded to the phthalate and Ala1 amide carbonyls at 1.49 and 1.38 Å, respectively. The phthalate-protonated ion **A7** represented another higher energy protomer. Nonetheless, the Gibbs energy differences among the various protomers and conformers were small enough to allow proton migration among the amide groups, ensuring conformational mobility of

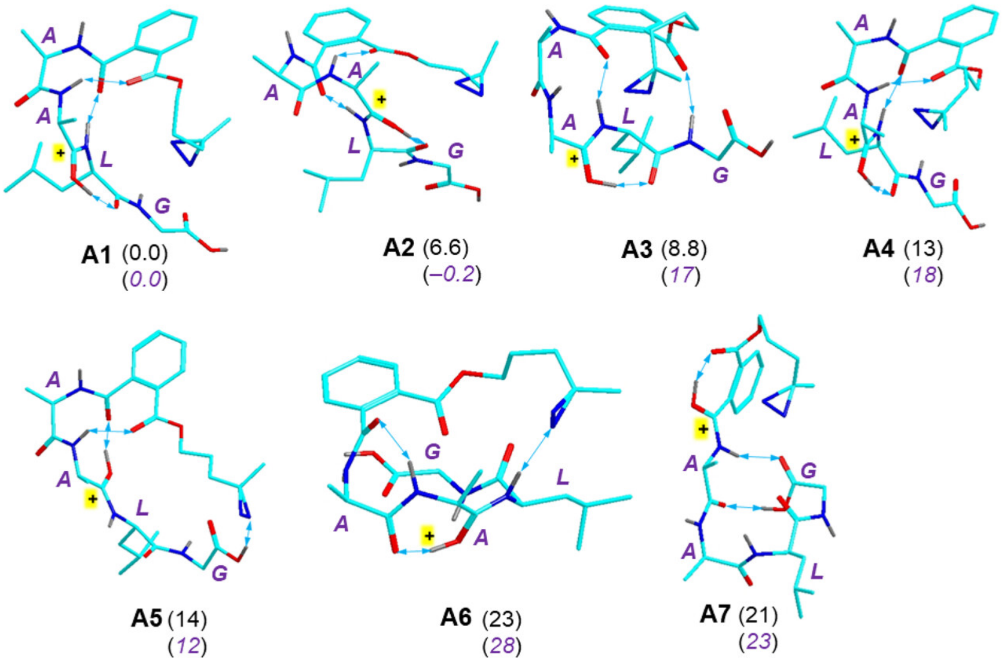


Figure 2.18: M06-2X/6-31+G(d,p)-optimized structures of low-energy s-AALG ions. Atom color coding is as follows: cyan = C, blue = N, red = O, gray = H. Only exchangeable hydrogens are shown to avoid clutter. Hydrogen bonds are indicated by double-headed blue arrows. The protonated amide groups are labeled with plus signs. Relative Gibbs energies in kJ mol^{-1} are from M06-2X/6-311++G(2d,p) single-point energy calculations including B3LYP zero-point energies, enthalpies, and entropies and referring to 310 K. Relative Gibbs energies of water-solvated ions are shown as purple italics.

the peptide C-terminus and side chains in thermal ions.

Solvation by water, as estimated using polarizable continuum model calculations, equalized the relative Gibbs energies of **A1** ($\Delta G_{310,\,\mathrm{aq}}$ = 0.0 kJ mol^{-1}) and **A2** ($\Delta G_{310,\,\mathrm{aq}}$ = -0.2 kJ mol^{-1}), while the energies of **A3-A6** remained above 12 kJ mol^{-1} ($\Delta G_{310,\,\mathrm{aq}} > 12$ kJ mol^{-1}) (Figure 2.18). This indicated that solvent effects that may play a role in ion charging in electrospray droplets were comparable for all s-AALG ion protomers and conformers and were unlikely to significantly change the populations of gas-phase ions.

2.3.6 Atom-Atom Contact Analysis of s-AALG Thermal Ions

The fully optimized geometries of the low Gibbs energy peptide scaffold ions represented static local energy minima, where the incipient carbene atoms could be distant from the carboxyl oxygen and amide nitrogen atoms involved in photodissociative crosslinking. For instance, in structure **A1**, the stationary distances between the incipient carbene atom (C18) and the carboxyl hydroxyl (O41), Gly-amide (N34), and Leu amide (N29) were 5.14, 4.26, and 4.90 Å, respectively. For a complete list of selected interatomic distances, see Table 2.5. Unlike the calculated ion geometries at local energy minima, which correspond to static structures at 0 K, the gas-phase ions studied experimentally were in the buffer gas at the ion trap temperature (310 K). Consequently, the thermal motion of all atoms and the resulting conformational changes had to be considered to capture and map the temporary variations in interatomic distances between the incipient carbene atom (C18) and specific atoms in the peptide chain.

Furthermore, photodissociation led to internal excitation of the nascent carbene ion, which can be expressed by assigning it an effective temperature. According to our calculations, the dissociation of **A1** into N_2 and the carbene intermediate was endergonic by 59 kJ mol^{-1}. When combined with the photon energy (337 kJ mol^{-1} at 355 nm) and the standard rovibrational enthalpy of **A1** (100 kJ mol^{-1} at 310 K), a total of 378 kJ mol^{-1} was distributed between the carbene ion and the departing N_2 molecule. Although the exact kinetic energy of the departing N_2 molecule was unknown, preventing a precise evaluation of the internal energy distribution, we estimated the effective temperature of the carbene ion to be 640 K, as an upper limit corresponding to 378 kJ mol^{-1} of rovibrational energy. The hot nascent carbene ion then underwent cooling collisions with the bath gas, limiting its internal excitation to the range of 310-640 K.

Table 2.5: Interatomic Distances in Peptide Scaffolds of the Incipient Carbene Atom C18 to Selected Atoms[a,b]

Ion	CO**O**H	COO**H**	Gly-**N**H	Gly-N**H**	Leu-**N**H	Leu-N**H**	Ala2-O**H**
A1	5.14	5.57	4.26	4.41	4.90	5.89	5.26
A2	6.86	7.78	5.38	6.27	5.46	6.03	4.15
A3	5.49	5.38	4.63	4.38	4.18	3.73	5.49
A4	12.0	12.7	9.72	10.2	7.53	7.36	7.85
A5	4.1	3.22	6.34	6.57	7.30	7.12	8.50
A6	8.94	9.50	6.61	7.07	4.49	3.51	7.36

[a] Atoms of interest shown with bold letters.
[b] Based on M06-2X/6-31+G(d,p) optimized geometries.

In the contact analysis, we considered that the diazirine ring limits the closest approach of peptide atoms to the incipient carbene atom to a distance based on the sum of the van der Waals radii of the involved atoms, also influenced by the mutual orientation of the interacting groups. A previous estimate for a random approach set this limit at approximately 4 Å[8], but this benchmark distance was still too long to lead to an unambiguous insertion reaction after carbene formation[14]. Therefore, we reexamined it using the results of the 100 ps BOMD trajectories of **A1** and **A2** (Figures 2.19-2.21), counting the encounters of the incipient carbene atom with selected ion atoms over a 2.5-4.5 Å distance. The trajectory contact analysis is shown in Figures 2.22 and 2.23. Despite having very similar Gibbs energies, the conformers followed different trajectories at both 310 and 640 K (Figures 2.19-2.21). At 310 K, ion **A1** experienced 22% of contacts with exchangeable hydrogens within a 3.7 Å distance, corresponding to the experimental crosslinking yield (21%, Table 2.1). Interestingly, the number of these contacts decreased when the ion temperature increased to 640 K, likely due to the larger conformational space accessed at higher temperatures, making specific contacts with exchangeable hydrogens relatively less frequent due to increased interactions with other atoms. The contacts were distributed among the Gly-amide NH (H36), Gly carboxyl OH (H42), and Leu-amide NH (H35). The most acidic proton (H80) showed lower participation at 310 K, which increased at 640 K (green symbols in Figure 2.22). In contrast, conformer **A2** primarily showed contacts with nonexchangeable hydrogen atoms, while contacts with exchangeable protons were minimal at 310 K (Figure 2.23). Most of the 310 K contacts occurred with the hydrogens of the Ala2 methyl group, accounting for 22% of contacts at 3.8 Å. However, while the Ala2 methyl hydrogens were accessible to C18, the methyl carbon atom remained at a distance greater than 4.5 Å throughout the trajectories, likely hindering insertion into the carbene and formation of the C–C bond. The BOMD trajectory of conformer **A2** was more sensitive to the effective temperature, as shown in Figures 2.19 and 2.20. At 640 K, **A2** developed numerous contacts with the Gly-amide NH and the charging proton (Figure 2.21), which could result in proton transfer and crosslinking. Thus, it can be inferred that crosslinking in the nascent, vibrationally hot carbene can involve the Gly amide and Leu OH protons starting from conformer **A2**. Upon collisional cooling, the crosslinking probability of **A2** decreased, while that of conformer **A1** increased. It should be noted that achieving close contact was a necessary, though not sufficient, condition for carbene crosslinking. The accessibility of the other atom of the X–H bond and the energetics of proton transfer were additional factors likely influencing the crosslinking yields and topicity.

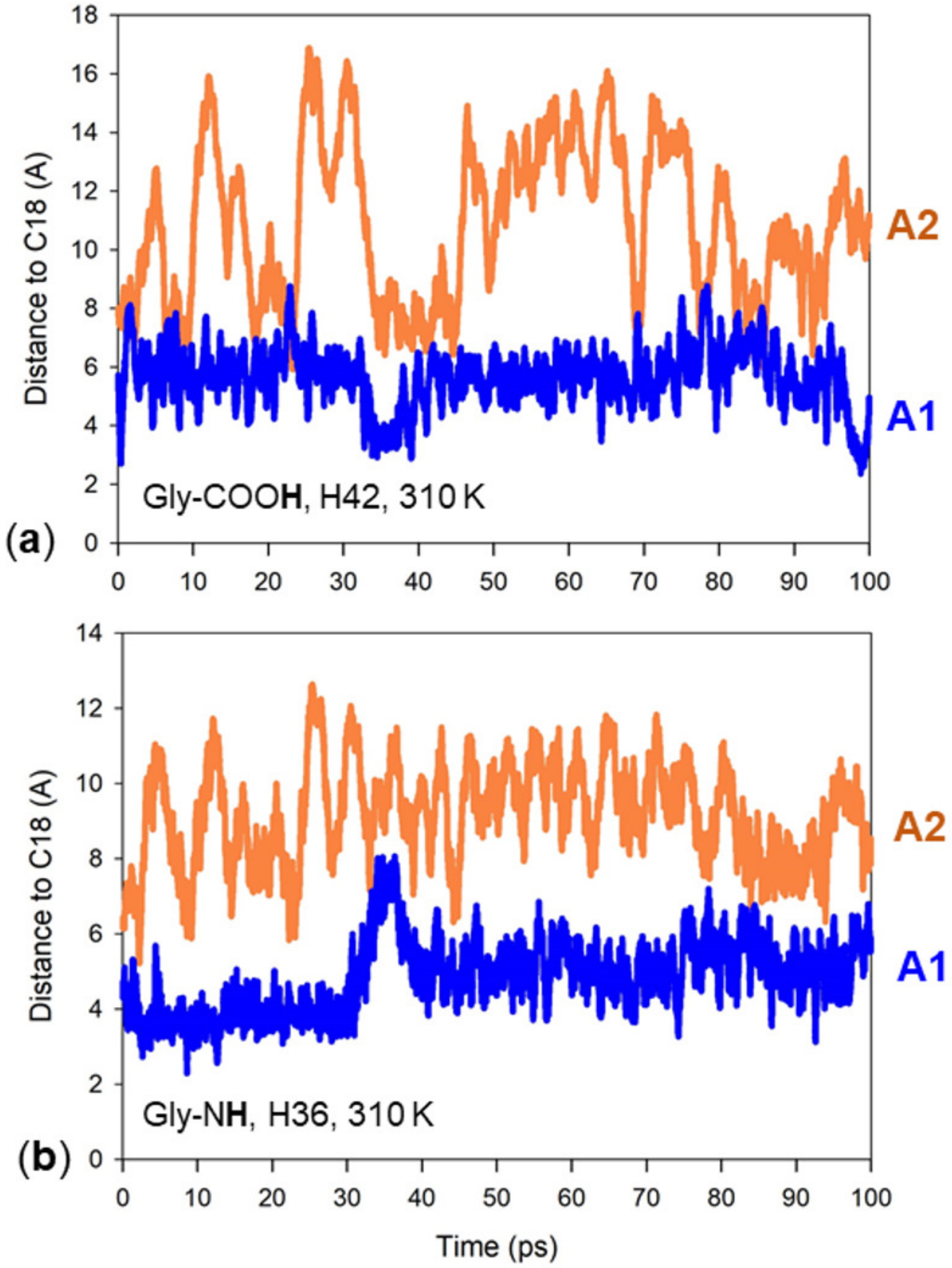


Figure 2.19: Interatomic distances to C18 of (a) Gly-COO**H** (H42) and (b) Gly-N**H** (H36) in s-AALG ions **A1** (blue) and **A2** (brown) along the 100 ps BOMD trajectories at 310 K.

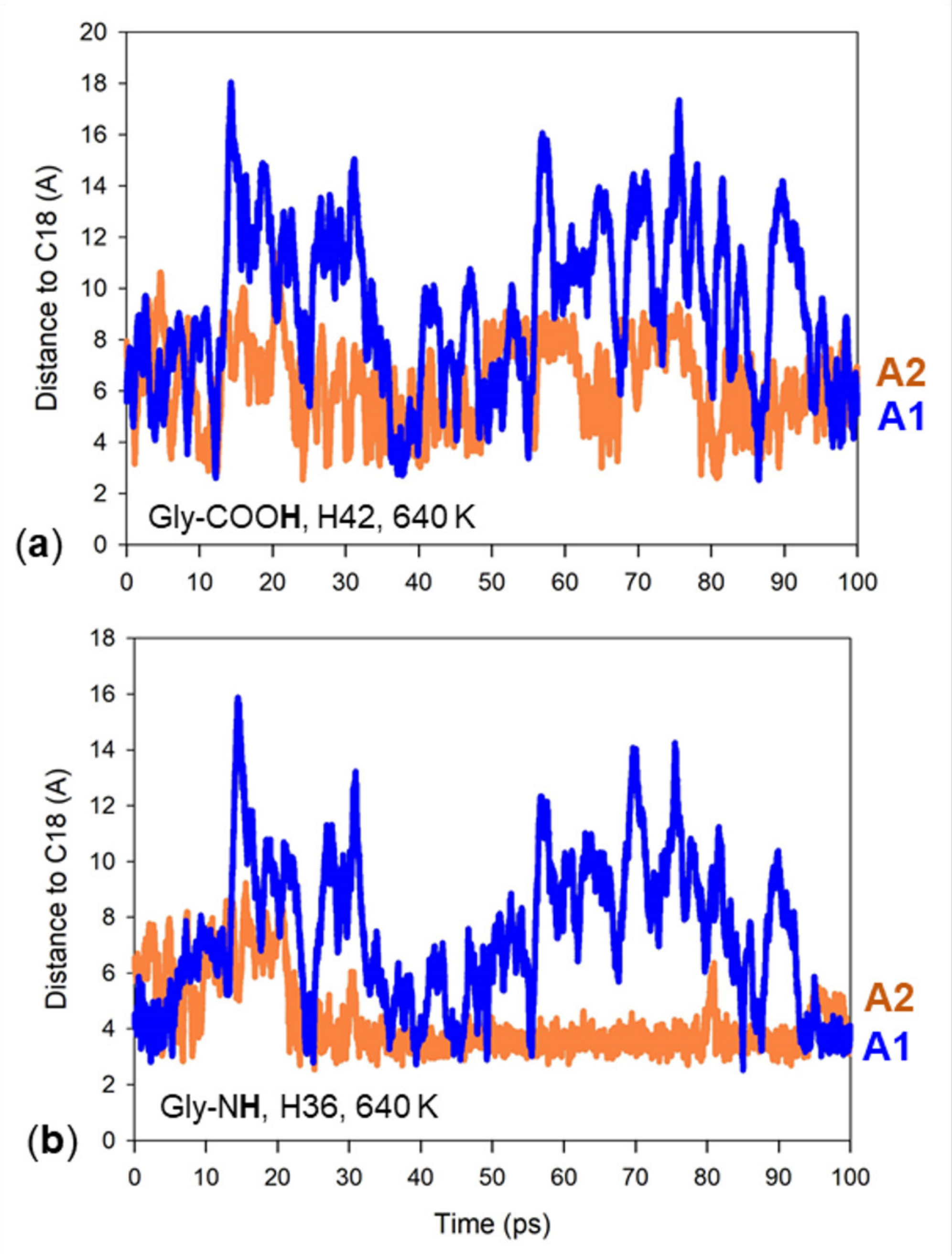


Figure 2.20: Interatomic distances to C18 of (a) Gly-COO**H** (H42) and (b) Gly-N**H** (H36) in s-AALG ions **A1** (blue) and **A2** (brown) along the 100 ps BOMD trajectories at 640 K.

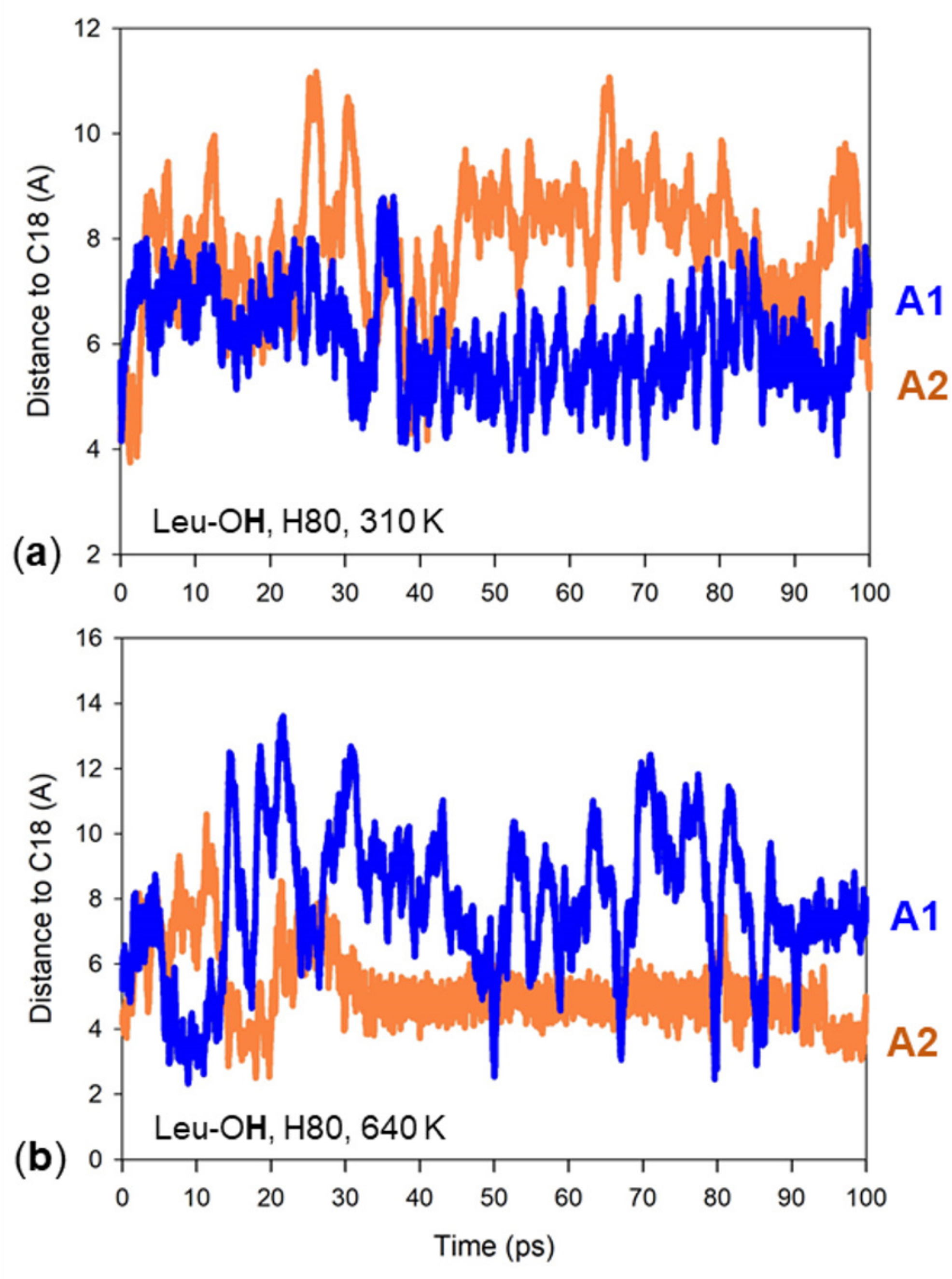


Figure 2.21: Interatomic distances to C18 of Leu-O**H** (H80, top panel) AALG ions **A1** (blue) and**A2** (brown) along the 100 ps BOMD trajectories at (a) 310 K and (b) 640 K.

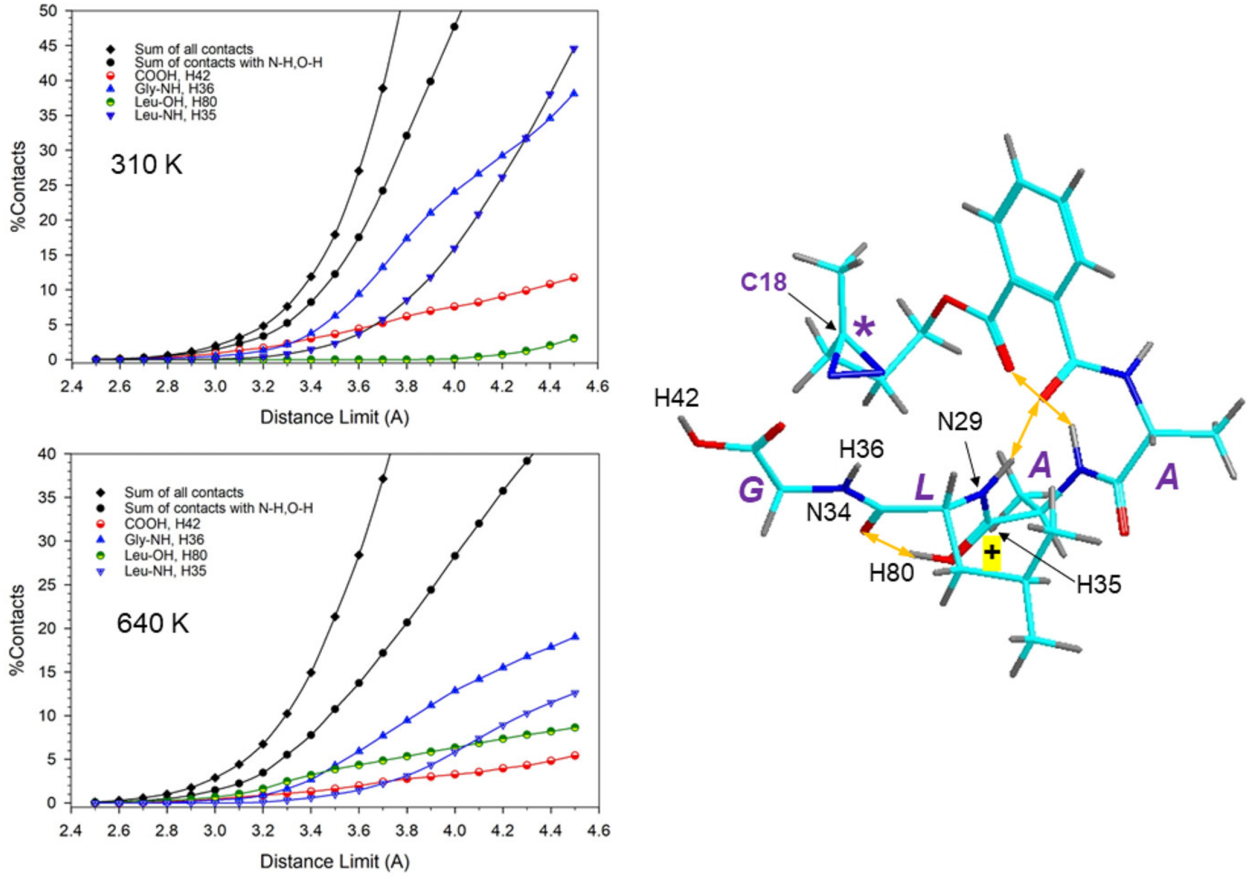


Figure 2.22: Contact analysis of 100 ps BOMD trajectories of ion **A1** at 310 K (top) and 640 K (bottom). For atom number assignment see the structure on the right; the incipient carbene (C18) is annotated with an asterisk. The limits in Å are cutoffs for counting the number of interatomic distances to C18 in the 100 ps trajectories. The total fractions are overcounted due to simultaneous contacts with multiple atoms.

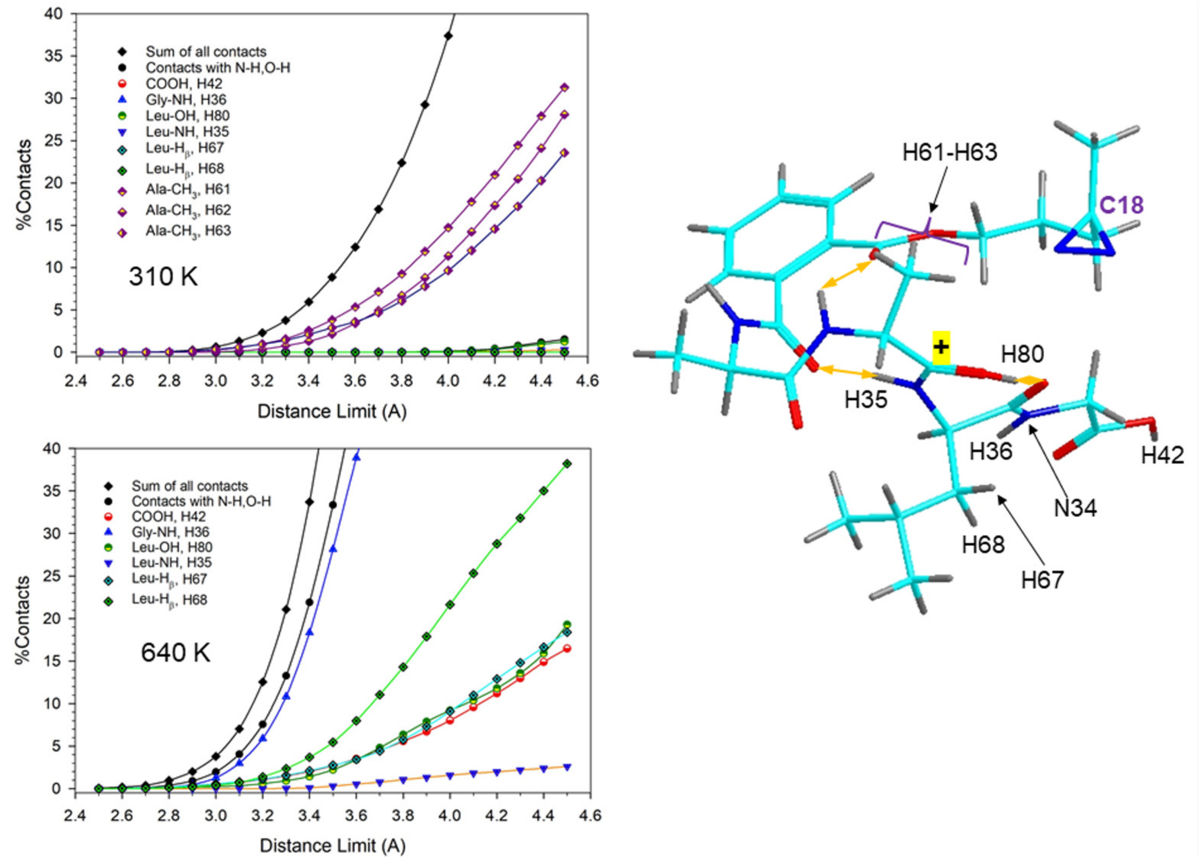


Figure 2.23: Contact analysis of 100 ps BOMD trajectories of ion **A2** at 310 K (top) and 640 K (bottom). For atom number assignment see the structure on the right.

2.4 Conclusions

This study demonstrated the efficacy of utilizing peptide scaffolds with covalently attached diazirine carbene precursors as model systems for examining intramolecular crosslinking behavior in gas-phase peptide ions. Across the sequences analyzed (s-LAAG, s-ALAG, and s-AALG), it was evident that the specific order and positioning of alanine and leucine residues had minimal impact on the overall crosslinking yields, indicating a robust crosslinking mechanism independent of these structural variations. The s-AALG scaffold, due to its consistent crosslinking pattern, served as a focal point for a more in-depth investigation.

Hydrogen-deuterium exchange and CID-MSn spectral analysis of synthetic reference products confirmed that polar groups, particularly the carboxyl OH and amide NH groups of the glycine residue, are major contributors to crosslink formation. This was further substantiated by BOMD simulations, which provided insights into the temperature dependence of these interactions and highlighted the role of thermal ion behavior in governing the proximity and frequency of these critical contacts with the incipient carbene atom. Overall, the comprehensive approach combining experimental techniques and computational simulations solidifies the s-AALG scaffold as a reliable model for understanding carbene crosslinking in hydrophobic peptide environments.

2.5 Bibliography

[1] Blencowe A.; Hayes W. “Development and Application of Diazirines in Biological and Synthetic Macromolecular Systems.” *Soft Matter.* **2005**, *1*, 178-205.

[2] Fleming S. A. “Chemical Reagents in Photoaffinity Labeling.” *etrahedron.* **1995**, *51*, 12479–12520.

[3] Hashimoto M.; Hatanaka Y. “Recent Progress in Diazirine-Based Photoaffinity Labeling.” *Eur. J. Org. Chem.* **2008**, *15*, 2513-2523.

[4] Liu, M. T. H. “Chemistry of Diazirines; Vols. I and II.” *CRC Press: Boca Raton*, **1987**.

[5] Frey, H. M.; Stevens, I. D. R. “The Photolysis of Dimethyldiazirine.” *J. Chem. Soc.* **1963**, 3514-3519.

[6] Barton, D. H. R.; Jaszberenyi, J. C.; Theodorakis, E. A.; Reibenspies, J. H. “The Invention of Radical Reactions. 30. Diazirines as Carbon Radical Traps. Mechanistic Aspects and Synthetic Applications of a Novel and Efficient Amination Process.” *J. Am. Chem. Soc.* **1993**, *115*, 8050-8059.

[7] Korneev, S. M. “Valence Isomerization between Diazo Compounds and Diazirines.” *Eur. J. Org. Chem.* **2011**, *2011*, 6153-6175.

[8] Shaffer, C. J.; Andrikopoulos, P. C.; Řezáč, J.; Rulíšek, L.; Tureček, F. “Efficient Covalent Bond Formation in Gas-Phase Peptide-Peptide Ion Complexes with the Photoleucine Stapler." *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 633-645.

[9] Liu, Y.; Ramey, Z.; Tureček, F. “Non-Covalent Interactions of a Neuroprotective Peptide Revealed by Photodissociative Cross-Linking in the Gas Phase.” *Chem.–Eur. J.* **2018**, *24*, 9259-9263.

[10] Nguyen, H. T. H.; Andrikopoulos, P. C.; Rulíšek, L.; Shaffer, C. J.; Tureček, F. “Photodissociative Cross Linking of Noncovalent Peptide-Peptide Ion Complexes in the Gas Phase.” *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1706-1720.

[11] Nguyen, H. T. H.; Huang, S. R.; Liu, Y.; Liu, Y.; Korn, J. A.; Tureček, F. “Probing Arginine-Phosphopeptide Interactions in Non-Covalent Peptide-Peptide Ion Complexes Using Gas-Phase Cross-Linking and Born-Oppenheimer Molecular Dynamics Calculations.” *Int. J. Mass Spectrom.* **2019**, *435*, 259-271.

[12] Huang, S. R.; Liu, Y.; Tureček, F. “Non-Covalent Complexes of the Amyloid Peptide Fragment Gly-Asn-Asn-Gln-Gln-Asn-Tyr. A Gas-Phase Photodissociative Cross-Linking and Born-Oppenheimer Molecular Dynamics Binding Study.” *Phys. Chem. Chem. Phys.* **2019**, *21*, 2046-2056.

[13] Liu, Y.; Tureček, F. “Photodissociative Crosslinking of Diazirine-Tagged Peptides with DNA Dinucleotides.” *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1992-2006.

[14] Liu, Y.; Liu, Y.; Nytka, M.; Huang, S. R.; Lemr, K.; Tureček, F. “Probing D- and L-Adrenaline Binding to β2-Adrenoreceptor Peptide Motifs by Photodissociation Crosslinking and Ion Mobility Mass Spectrometry.” *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1041-1052.

[15] Olsson, R.; Berg, U.; Frejd T. “Alkylative Cleavage of the Endo-Cyclic C-O Bond in Lactol Ethers Employing Organometallic Reagents." *Acta Chem. Scand.* **1998**, *52*, 172-174.

[16] Veri, E. “Synthesis and Reactivity of α-Trialkylsilyl α-Amino Acids." Ph.D. Dissertation, RWTH Aachen University, Aachen, Germany, **2006**.

[17] Sajiki, H. “Selective Inhibition of Benzyl Ether Hydrogenolysis with Pd/C Due to the presence of Ammonia, Pyridine or Ammonium Acetate." *Tetrahedron Lett.* **1995**, *36*, 3465-3468.

[18] Wan, J.; Brož, B.; Liu, Y.; Huang, S. R.; Marek, A.; Tureček, F. “The DNA Radical Code. Resolution of Identity in Dissociations of Trinucleotide Codon Cation Radicals in the Gas Phase." *J. Am. Soc. Mass Spectrom.* **2023**, *34*, 304-319.

[19] Řezáč, J. “Cuby: An Integrative Framework for Computational Chemistry." *J. Comput. Chem.* **2016**, *37*, 1230-1237.

[20] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. “Semi-Empirical Quantum Chemical PM6 Method Augmented by Dispersion and H Bonding Correction Terms Reliably Describes Various Types of Non-covalent Complexes." *J. Chem. Theory Comput.* **2009**, *5*, 1749-1760.

[21] Becke, A. D. “Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior.” *Phys. Rev. A* **1998**, *38*, 3098-3100.

[22] Zhao, Y.; Truhlar, D. G. “The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals.” *Theor. Chem. Acc.* **2008**, *120*, 215-241.

[23] Tomasi, J.; Mennucci, B.; Cammi, R. “Quantum Mechanical Continuum Solvation Models.” *Chem. Rev.* **2005**, *105*, 2999-3093.

[24] Marek, A.; Tureček, F. “Collision-Induced Dissociation of Diazirine-Labeled Peptide Ions. Evidence for Brønsted-Acid Assisted Elimination of Nitrogen.” *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 778-789.

[25] Ziemianowicz, D. S.; Bomgarden, R.; Etienne, C. W.; Schriemer, D. C. “Amino Acid Insertion Frequencies Arising from Photoproducts Generated Using Aliphatic Diazirines.” *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2011-2021.

[26] Iacobucci, C.; Goetze, M.; Piotrowski, C.; Arlt, C.; Rehkamp, A.; Ihling, C.; Hage, C.; Sinz, A. “Carboxyl-Photo-Reactive MS Cleavable Cross-Linkers: Unveiling a Hidden Aspect of Diazirine-Based Reagents.” *Anal. Chem.* **2018**, *90*, 2805-2809.

[27] Bleiholder, C.; Osbum, S.; Williams, T. D.; Suhai, S.; Van Stipdonk, M.; Harrison, A. G.; Paizs, B. “Sequence-Scrambling Fragmentation Pathways of Protonated Peptides.” *J. Am. Soc. Mass Spectrom.* **2008**, *130*, 17774-17789.

[28] Novák, J.; Lemr, K.; Schug, K. A.; Havlíček, V. “CycloBranch: De Novo Sequencing of Nonribosomal Peptides from Accurate Product Ion Mass Spectra.” *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1780-1786.

[29] Dix, E. J.; Goodman, J. L. “Protonation of Diarylcarbenes by Alcohols: The Importance of Ion Pair Dynamics.” *J. Phys. Chem.* **1994**, *98*, 12609-12612.

[30] Peon, J.; Polshakov, D.; Kohler, B. “Solvent Reorganization Controls the Rate of Proton Transfer from Neat Alcohol Solvents to Singlet Diphenylcarbene.” *J. Am. Chem. Soc.* **2002**, *124*, 6428-6438.

[31] Vaisar, T.; Urban, J. “Probing the Proline Effect in CID of Protonated Peptides.” *J. Mass Spectrom.* **1996**, *31*, 1185-1187.

Chapter 3

Carbene Crosslinking with Basic Peptide Ion Scaffolds in the Gas-Phase

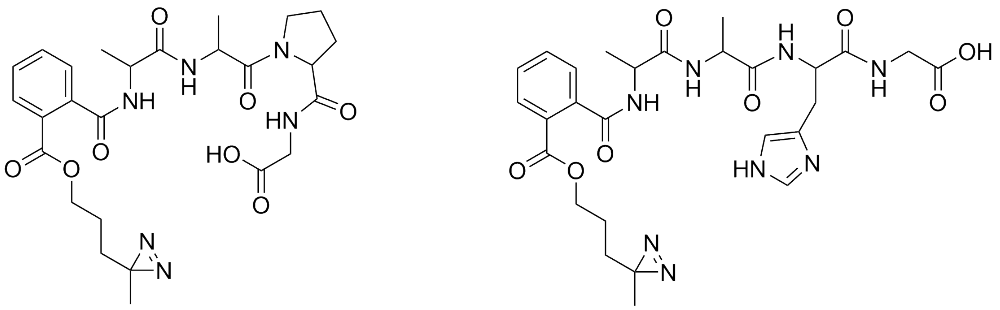
Reproduced in part with permission from Hongyi Zhu, Václav Zima, Emily R. Ding, and František Tureček. Carbene Cross-Linking in Gas-Phase Peptide Ion Scaffolds. *Journal of the American Society for Mass Spectrometry*, **2023**, *34* (4), 763-774.

Abstract. Based on the established intramolecular crosslinking peptide model, we investigated the effects of basic amino acid residues with different protonation sites and conformations on carbene crosslinking. Carbene intermediates were generated by UV-laser photodissociation at 355 nm of the diazirine ring in mass-selected ions, and the crosslinked products were detected and quantified using collision-induced dissociation tandem mass spectrometry (CID-MS^n, n = 3–5). Compared to normal hydrophobic amino acids like alanine, leucine, and glycine, which have only methyl group side chains, the presence of proline and histidine residues significantly decreased crosslinking yields. Interpretation of the crosslinking results was aided by Born-Oppenheimer molecular dynamics (BOMD) simulations and density functional theory (DFT) calculations, which helped establish the protonation sites and conformations of the precursor ions. The analysis of long BOMD trajectories (100 ps) was used to count the close contacts between the incipient carbene and peptide atoms. These statistical counts were then correlated with the gas-phase crosslinking results, providing insights into how different protonation sites and amino acid conformations influence crosslinking efficiency.

3.1 Introduction

The intramolecular crosslinking peptide model we developed demonstrated precise identification of crosslinking sites, down to specific X–H bonds in amino acids, offering high resolution in carbene photodissociation research. The three sequences—s-LAAG, s-ALAG, and s-AALG—were found to exhibit similar crosslinking patterns. In the s-AALG sequence, BOMD calculations revealed that the initial protonation sites, located at the phthalate, Ala1, Ala2, and Leu amides, collapsed upon BOMD simulation to a single protomer, with the proton migrating to the Ala2 amide. To further explore how different protonation sites influence peptide conformation and crosslinking behavior, we used the s-AALG scaffold as a standard and replaced leucine with other amino acid residues containing different funcional groups. This modification also highlights the versatility of our scaffold model, as peptide protonation sites and chain conformations can be easily adjusted.

Certain amino acid residues are known to promote selective cleavages that dominate the spectra and suppress other nonselective cleavages that can aid in peptide identification. Examples of these selective cleavages include the "proline effect," a preference for cleavage N-terminal to proline residues[1]-[3] to give $y_n{}^+$ ions, where n represents the number of residues in the fragment counting from the C-terminus for y-type ions and from the N-terminus for b-type ions[4]. The "proline effect" can in part be rationalized by the increased basicity of the prolyl-amide site. Similarly, histidine is categorized as a basic amino acid due to its strong basicity, with protonation fixed at its imidazole nitrogens. In this study, we aim to leverage our established intramolecular peptide scaffold to investigate sequences with varying basicity, from proline to histidine, examining both basic backbone and side-chain residues (Scheme 4.1). By combining experimental data with BOMD calculations and contact analysis, we aim to map the crosslinking behavior of these different sequences and gain further insights into how peptide structure and protonation influence crosslinking.



Scheme 3.1: Scaffolds design for s-AAPG and s-AAHG.

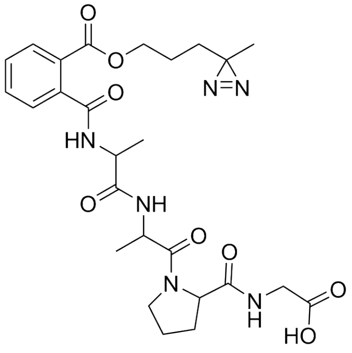
3.2 Experimental Section

3.2.1 Materials and Methods

Unless otherwise noted, all starting materials were purchased from Sigma Aldrich (St. Louis, MO) and were used as received. Acetonitrile was dried using 4A molecular sieves, grade 514 (8-12 mesh). Mass spectra were measured on a Bruker Daltonik, GmbH (Bremen, Germany), amaZon ion trap mass spectrometer that was equipped with an EKSPLA NL301G Nd-YAG laser (Altos Photonics, Bozeman, MT, USA) working at 20 Hz frequency and 3- to 6-ns pulse width. Ions were produced by electrospray ionization, selected by mass in the ion trap, and irradiated by multiple laser pulses at 355 nm. Multistage tandem mass spectra were obtained by collision-induced dissociation (CID) of mass selected fragment ions. The excitation amplitude was chosen to achieve a high conversion of precursor ions.

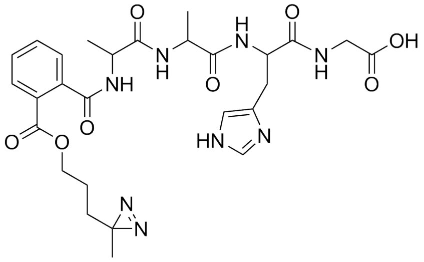
3.2.2 Synthetic Procedures

4,4-Azipent-1-yl-s-AAPG (I)



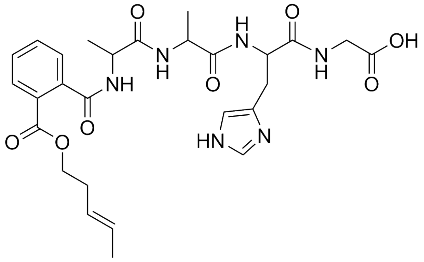
The title conjugate was prepared according to the general procedure for esterification (see Chapter 2.2.2) and the general procedure for solid phase peptide synthesis (see Chapter 2.2.2). The product was characterized by MS/MS: *m/z* 559 (M+H)$^+$, 484 (–GlyOH, b_4), 387 (–Pro-GlyOH, b_3), 316 (–Ala-Pro-GlyOH, b_2), 244 ((Ala-Pro-GlyOH+H)$^+$, y_3), 173 ((Pro-GlyOH+H)$^+$, y_2).

4,4-Azipent-1-yl-s-AAHG (II)



The title conjugate was prepared according to the general procedure for esterification (see Chapter 2.2.2) and the general procedure for solid phase peptide synthesis (see Chapter 2.2.2). The product was characterized by MS/MS: *m/z* 599 $(M+H)^+$, 571 ($–N_2$), 553 ($–N_2–H_2O$), 496 (–GlyOH), 485 ($–N_2–C_5H_{10}O$), 468 (*m/z* 485$–NH_3$), 284 $(\text{Ala-His-GlyOH+H})^+$, 213 $(\text{His-GlyOH+H})^+$.

3-Penten-1-yl-s-AAHG (III)



The title conjugate was prepared according to the general procedure for esterification (see Chapter 2.2.2) and the general procedure for solid phase peptide synthesis (see Chapter 2.2.2). The product was characterized by MS/MS: *m/z* 571 $(M+H)^+$, 553 ($–H_2O$), 496 (–GlyOH), 485 ($–C_5H_{10}O$), 468 (*m/z* 485$–NH_3$), 359 (–His–GlyOH), 284 $(\text{Ala-His-GlyOH+H})^+$, 249, 213 $(\text{His-GlyOH+H})^+$, 195 (*m/z* $–H_2O$).

3.2.3 MS Characterization of Synthesized Products

3.2.4 Calculations

Born-Oppenheimer molecular dynamics (BOMD) calculations were run as 20 ps trajectories with 1 fs steps at 350-600 K using PM6-D3H4 calculations.[5] The BOMD calculations were run under the high-level Cuby4 platform.[6] The initial ion structures were constructed with different protonation sites that were at the peptide amide carbonyls for the non-basic peptides. In addition to thermal conformational motion, BOMD allows bond changes to proceed along the trajectory, namely, proton migration between different sites in the ion. The 20,000 structures generated by BOMD were sampled at regular intervals and 200 selected snapshots were fully optimized by PM6-D3H4. This augmented semiempirical procedure has been shown to capture hydrogen bonding and dispersion interactions,[7] and we have used it previously for structure analysis of several gas-phase peptide ions.[8]-[14] The PM6-D3H4-optimized structures were sorted out by type and energy, and 15-20 low-energy ions were reoptimized with density functional theory (DFT) calculations, using B3LYP[15] and the 6-31+G(d,p) basis set and including harmonic frequency analysis. Several lowenergy B3LYP structures were reoptimized with M06-2X[16] and the 6-31+G(d,p) basis set. These were

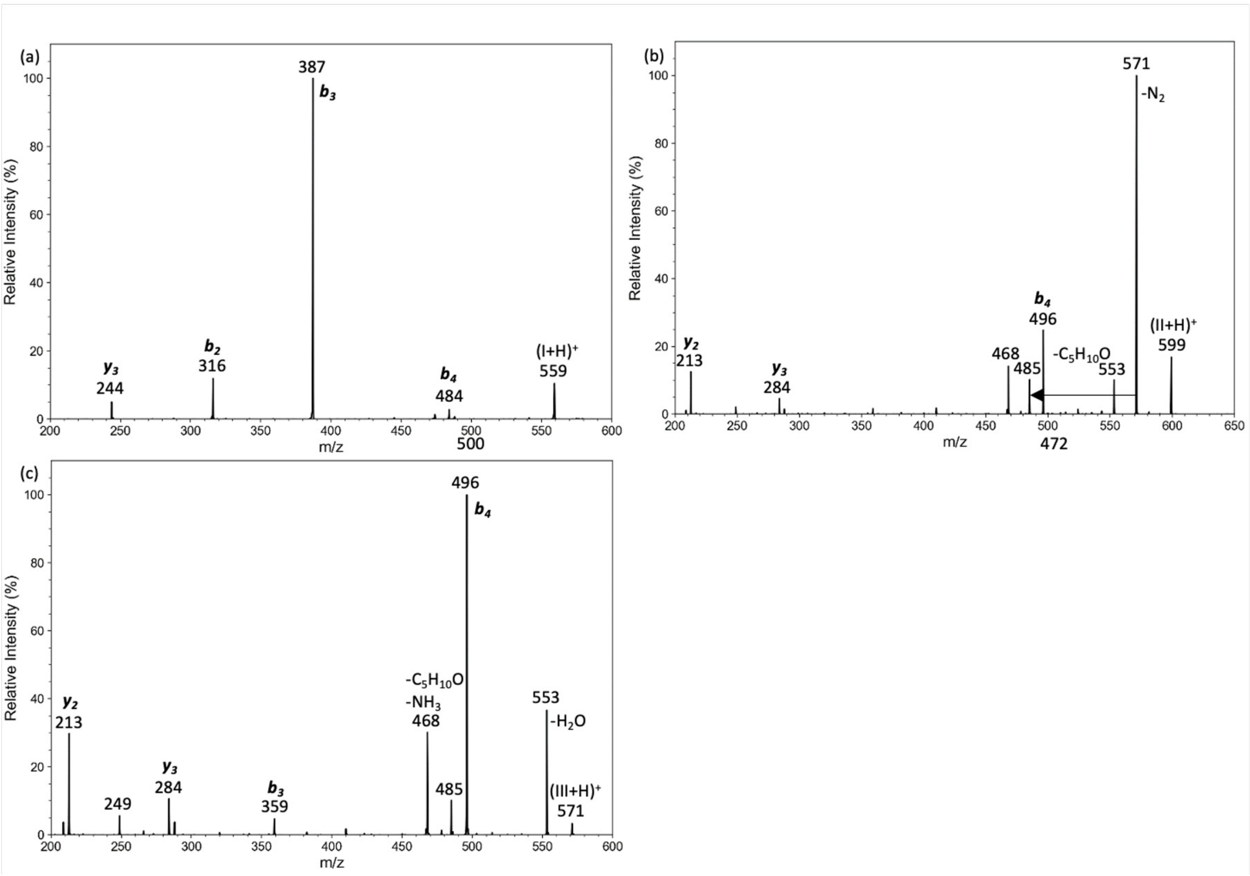


Figure 3.1: CID-MS^2 mass spectra of (a) (I+H)$^+$ ion at *m/z* 559, (b) (II+H)$^+$ ion at *m/z* 599, and (c) (III+H)$^+$ ion at *m/z* 571.

used for single-point energy calculations that were carried out with M06-2X/6-311++G(2d,p) to provide relative energies that were combined with B3LYP zero-point vibrational energies, enthalpies, and entropies to obtain relative Gibbs energies for ion conformers and protomers. Ion solvation energies were estimated from self-consistent reaction field polarizable continuum model calculations (PCM)[17] in water dielectric with full gradient optimization by M06-2X/6-31+G(d,p) and single-point M06-2X/6-311++G(2d,p) energies. These standard DFT calculations were carried out with the Gaussian 16 suite of programs (Revision B.01) that was licensed from Gaussian, Inc. (Wallingford, CT). The lowest-energy ion structures from the M06-2X calculations were then used as initial geometries for 100 ps BOMD trajectory runs using Cuby4, that each created 100,000 snapshots for atom-atom contact analysis.

3.3 Results and Discussion

3.3.1 Photodissociation and Crosslinking Yields

Scaffold ions s-AAPG (*m/z* 559) and s-AAHG (*m/z* 599) were produced by electrospray protonation from 1:1 water:methanol solutions and selected by mass in the ion trap. Photodissociation at 355 nm resulted in elimination of N_2 from the diazirine ring and formation of product ions (Figure 3.2). Photodissociation spectrum of s-AAPG was obtained with 36 laser pulses gave photodissociation conversions at 49%. In contrast, differentiating the UVPD-produced ions in the photodissociation spectrum of s-AAHG was challenging, as all the UVPD-MS^2 induced fragments (Figure 3.2 (b)) were also present in the CID-MS^2 spectrum (Figure 3.1 (b)). To address this, we conducted further investigation through CID-MS^3 analysis of both UVPD and CID fragments. At *m/z* 468, we observed that the fragmentation patterns from UVPD-CID-MS^3 and CID-CID-MS^3 were distinct (Figure 3.3), indicating the presence of UVPD-produced ions. This ion difference was then used to calculate the UVPD conversion rate for s-AAHG. Compared with the scaffolds containing alanine, leucine, and glycine residues in different order, s-AAPG with the basic backbone gave a higher photodissociation yield and lower yields of both the survivor fraction and crosslink products. The last two figures are even more pronounced for s-AAHG with the basic side chain which gave nondissociating photoproducts at 83% but only a 7.8% yield of fragments assigned to crosslinks.

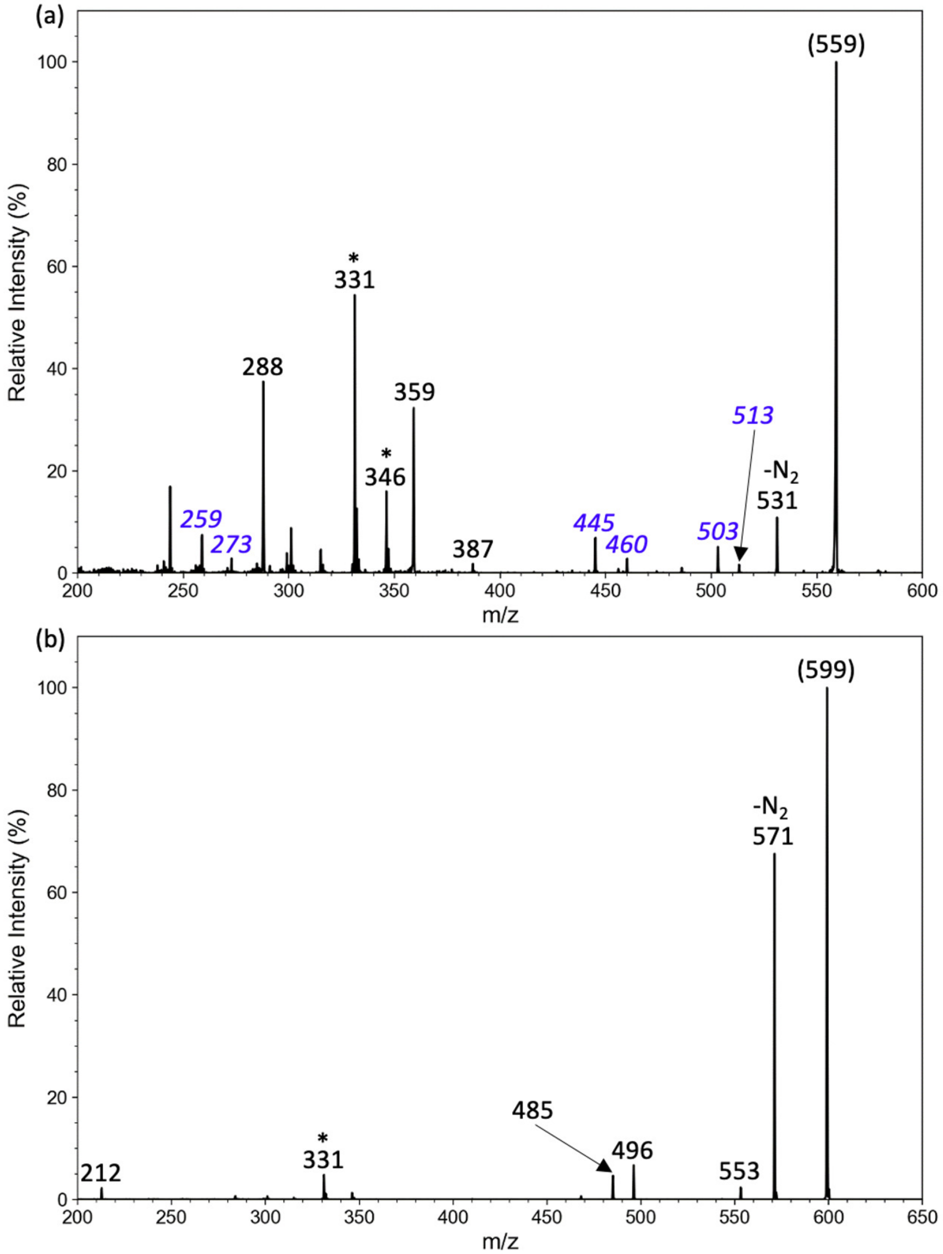


Figure 3.2: UVPD-MS^2 mass spectra (355 nm) of protonated (a) s-AAPG at *m/z* 559 and (b) s-AAHG at *m/z* 599. Fragment ions originating from crosslinked photoproducts are annotated by blue italic numerals. Ions from laser desorption-ionization of surface material are denoted by asterisks.

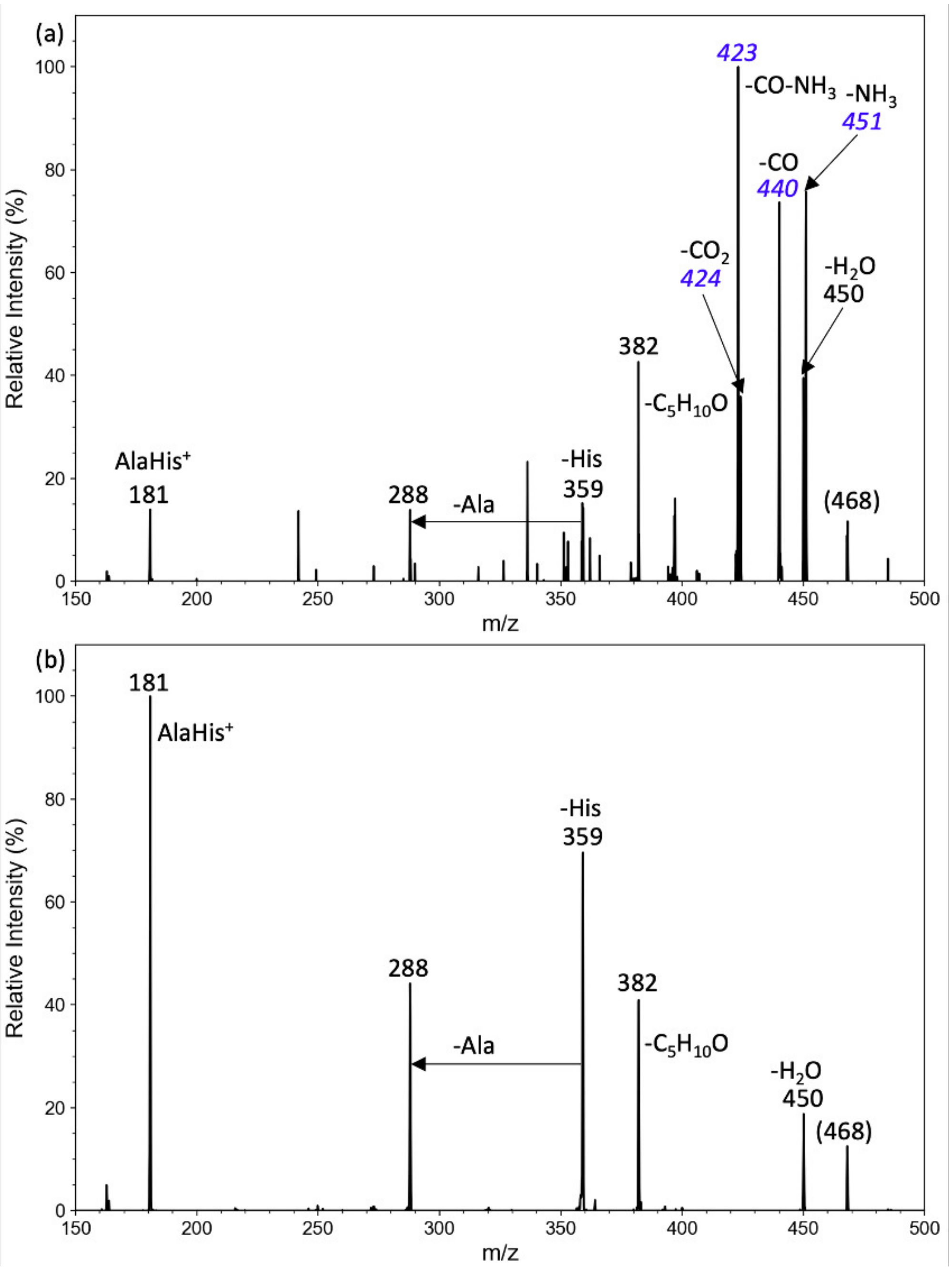


Figure 3.3: (a) CID-MS^4 mass spectrum of (MH – N$_2$ – GlyOH – CO)$^+$ *m/z* 461 of s-AAHG following UVPD; (b) CID-MS^3 mass spectrum of (MH – N$_2$ – GlyOH – CO)$^+$ *m/z* 461 of 3-penten-1-yl-s-AAHG following CID. For ion color coding see text in Figure 3.2.

Table 3.1: Photodissociation Conversion and Crosslink Yields

Scaffold	Laser Pulses (%)	UVPD Conversion[a] (%)	Survivor Fraction[b] (%)	Crosslink Fraction[c] (%)
s-AAPG	36	49	6.4	11
s-AAHG	14	43	83	7.8

[a] Fraction of total UVPD produced ions.
[b] Fraction of nondissociating $(MH - N_2)^+$ ions relative to the total photoproduct ion intensities.
[c] Fraction of photoproducts identified as originating from crosslinks.

3.3.2 CID-MS^n Analysis of Crosslinked Photoproducts

The $(MH - N_2)^+$ ions resulting from diazirine photodissociation were further investigated by CID-MS^n (Figure 3.4, Table 3.2, and Table 3.3). For s-AAPG, fragment ions indicative of a cyclic peptide structure were observed (*m/z* 531, Figure 3.4 (a)). These ions formed through the loss of CO (*m/z* 503), Ala (*m/z* 460), combined with water (*m/z* 442), and $C_5H_{10}O$ (*m/z* 445). Interestingly, the loss of the $C_5H_{10}O$ side chain was not accompanied by the loss of glycine, suggesting that the side chain is conjugated to the proline residue, as evidenced by the $C_5H_{10}O$-ProGly-H_2O^+ (*m/z* 259) and Ala-$C_5H_{10}O$-ProH_2O^+ (*m/z* 273) ions. Additionally, the spectrum displayed fragment ions consistent with non-cyclized structures, such as the *m/z* 359 and *m/z* 288 sequence ions. Notably, these ions also appeared prominently in the CID-MS^2 spectrum of the AAPG-alkene byproduct (*m/z* 531, Figure 3.4 (b)). The internal cleavage ion from the UVPD-CID-MS^3 of s-AAPG (*m/z* 460, corresponding to the loss of Ala) was further analyzed by its CID-MS^4 spectrum (Figure 3.5), which displayed the loss of $C_5H_{10}O$ (*m/z* 374) combined with the loss of glycine (*m/z* 299). As the inverse sequence of losses (glycine followed by $C_5H_{10}O$) was not observed (Figure 3.5), this suggests cross-linking via a glycine ester. Together with the other dissociations, the spectra indicate that the $(MH - N_2)^+$ ions from s-AAPG represent a mixture of linear and cyclic isomers.

The CID-MS^3 spectrum of the abundant survivor $(MH - N_2)^+$ ions from s-AAHG (*m/z* 571, Figure 3.6 (a)) showed features that indicated predominant alkene products. This was inferred from the similarity with the CID-MS^2 spectrum of the authentic AAHG-pentene standard (Figure 3.6 (b)). Thus, the carbene generated in the histidine-containing scaffold chiefly underwent a rearrangement to an alkene or alkenes.

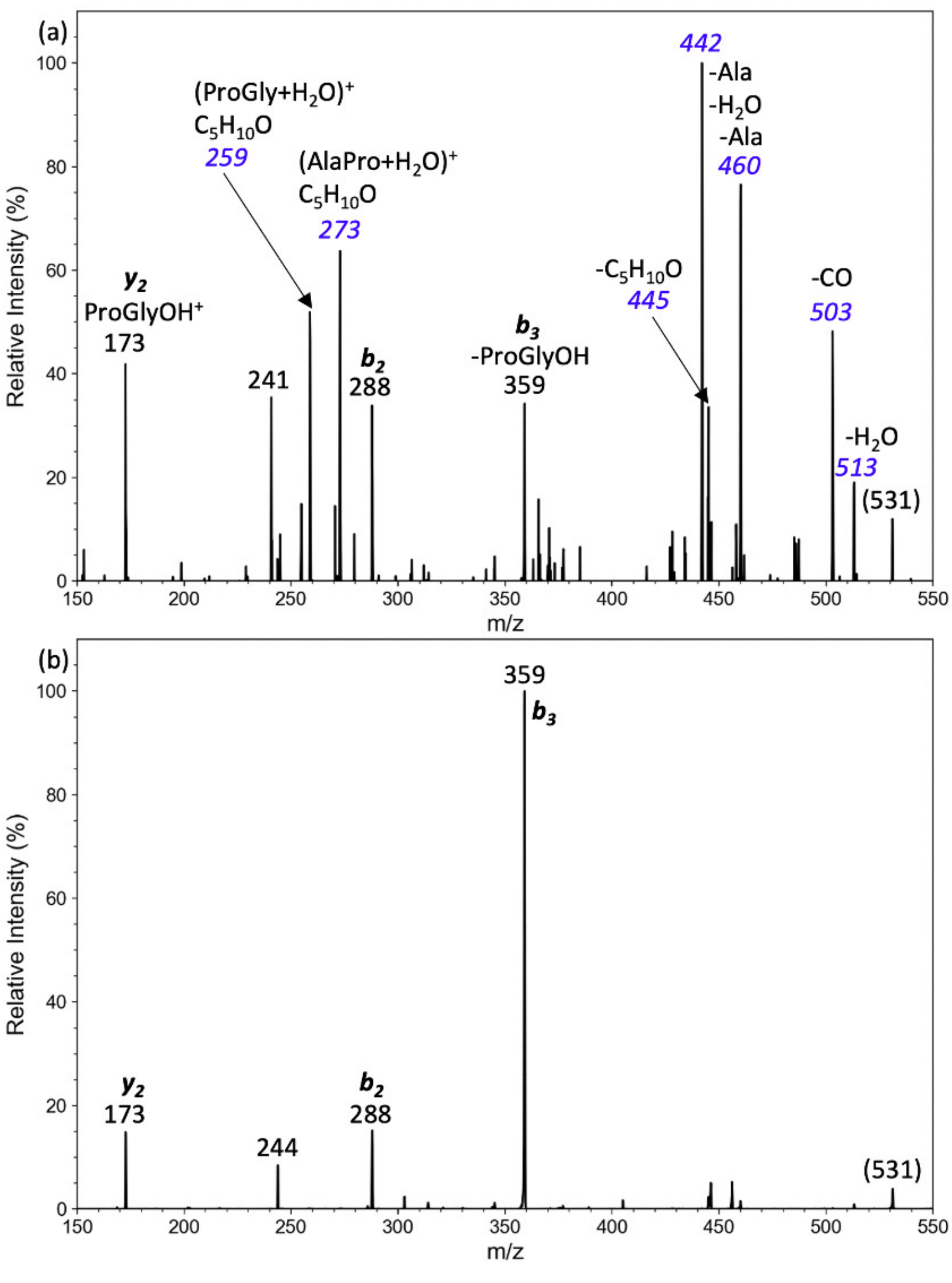


Figure 3.4: (a) CID-MS^3 mass spectrum of ions from photodissociation of s-AAPG at *m/z* 531, and (b) CID-MS^2 mass spectrum of the (M – N_2) byproduct (*m/z* 531) from s-AAPG. For ion color coding see text in Figure 3.2.

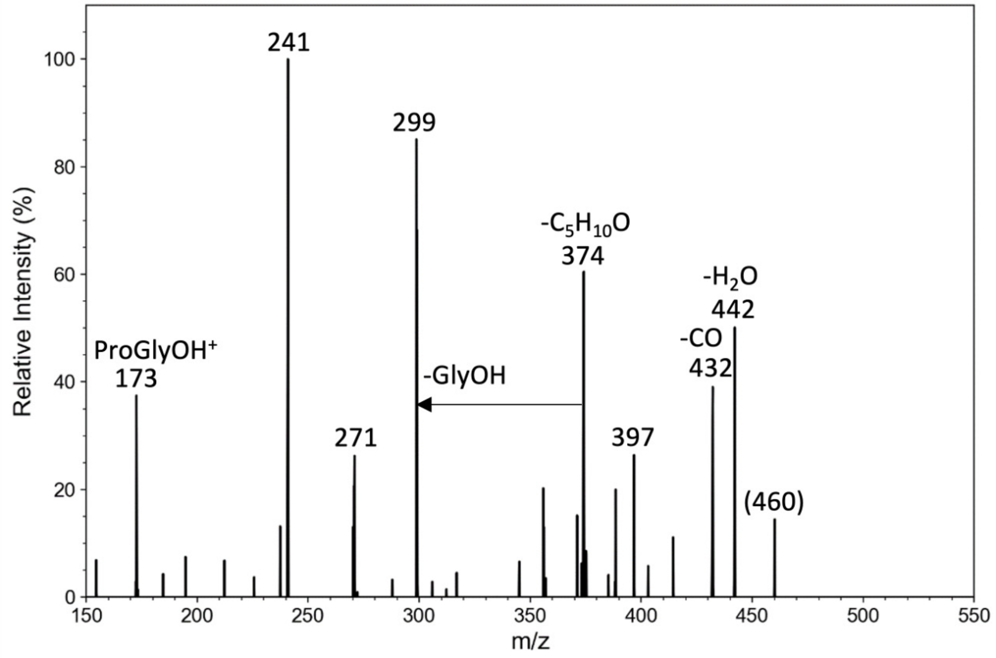


Figure 3.5: CID-MS^4 mass spectrum of *m/z* 460 from UVPD of s-AAPG and loss of Ala.

Table 3.2: CID-MS^4 Fragments of $(MH - N_2)^+$ *m/z* 531 from s-AAPG Following UVPD

513 $-H_2O$	503 –CO	460 –Ala	445 $-C_5H_{10}O$	442 $-Ala-H_2O$	359 –ProGlyOH	288 359-Ala	273 $(AlaPro+H_2O)^+$ $C_5H_{10}O$	259 $(ProGly+H_2O)^+$ $C_5H_{10}O$	241 $ProGly^+$ $C_5H_{10}O$	173 $ProGlyOH^+$
496	486	442	427		288	260				158
456	460	389	371			217				144
373	458	374	273							132
341		319	173							
313		299								
255		289								
		288								
		241								
		173								

3.3.3 Structures of s-AAPG and s-AAHG Peptide Ions

The presence of basic residues in s-AAPG and s-AAHG suggested preferential protonation at the Pro tertiary amide and His imidazole ring, respectively, in the gas-phase ions. Indeed, initial s-AAPG ion structures with the proton placed at different amide groups converged to Pro-protomers upon BOMD. The lowest Gibbs energy structures (**P1**-**P3**) displayed the same peptide pattern with the protonated tertiary amide at Pro, forming a strong hydrogen bond with the phthalate ester carbonyl (Figure 3.7). These nearly isoenergetic structures differed in the conformation of the diazirine-containing side chain, which influenced the distance between the incipient carbene atom (C18) and the peptide atoms. Specifically, structure **P3**

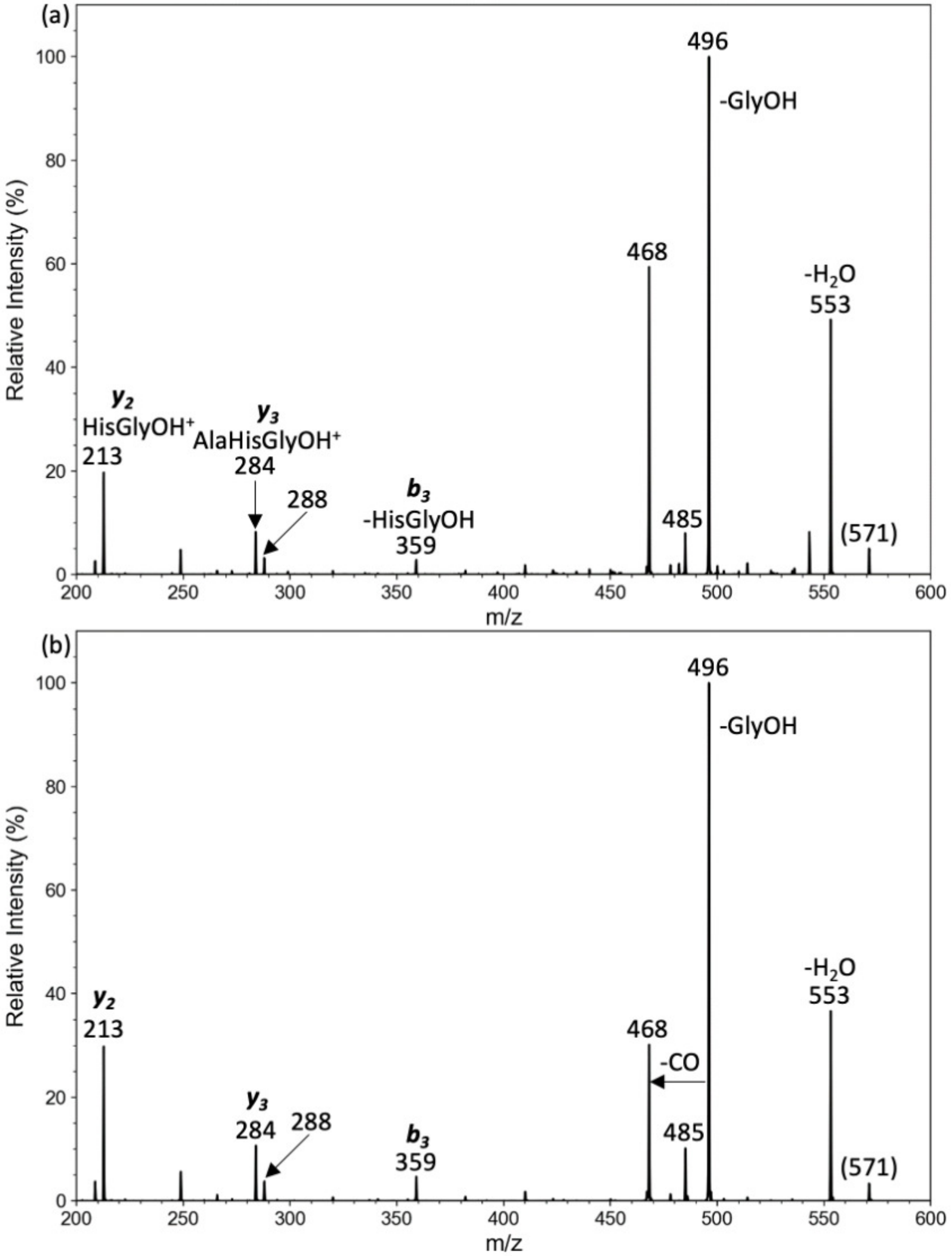


Figure 3.6: (a) CID-MS^3 mass spectrum of ions from photodissociation of s-AAHG at *m/z* 571, and (b) CID-MS^2 mass spectrum of *m/z* 571 $(M + H)^+$ from 3-penten-1-yl-s-AAHG.

Table 3.3: CID-MS^4 Fragments of $(MH – N_2)^+$ *m/z* 571 from s-AAHG Following UVPD

553	543	496	485	468	359	288	284	213
$–H_2O$	–CO	–GlyOH	$–C_5H_{10}O$	496–CO	496-His	359–Ala	AlaHisGlyOH$^+$	HisGlyOH$^+$
536		468	467	451			266	195
535		359	410	450			249	138
525			382	440			209	110
510			273	424			192	
508			213	423				
496			195	397				
478				382				
467				336				
320								
249								

exhibited very short distances between C18 and the carboxyl OH and Gly amide NH, measuring 3.58 and 4.16 Å, respectively (Table 3.4).

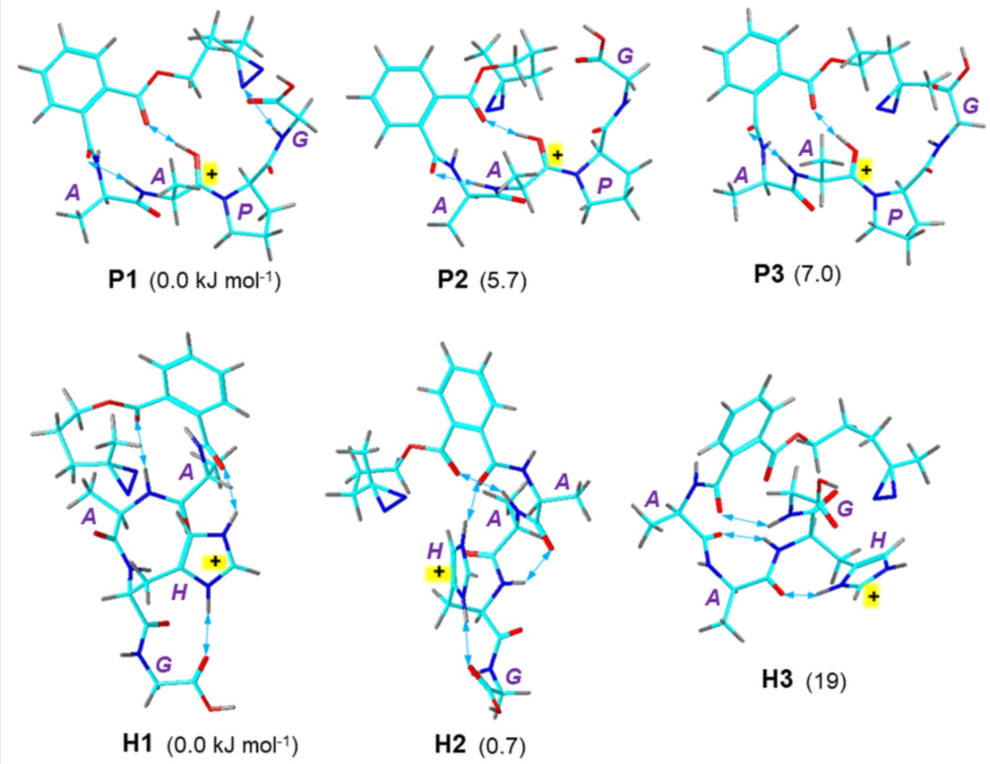
Table 3.4: Interatomic Distances in Peptide Scaffolds of the Incipient Carbene Atom C18 to Selected Atoms^{a,b}

Ion	CO**O**H	COO**H**	Gly-**N**H	Gly-N**H**	Ala2-O**H**
P1	5.38	5.59	4.44	3.67	5.10
P2	7.74	7.54	6.66	6.28	4.06
P3	3.58	3.88	4.16	4.55	4.39
H1	10.9	11.2	8.07	7.83	
H2	10.7	10.5	9.12	9.27	
H3	7.39	7.77	5.83	6.53	

[a] Atoms of interest shown with bold letters.

[b] Based on M06-2X/6-31+G(d,p) optimized geometries.

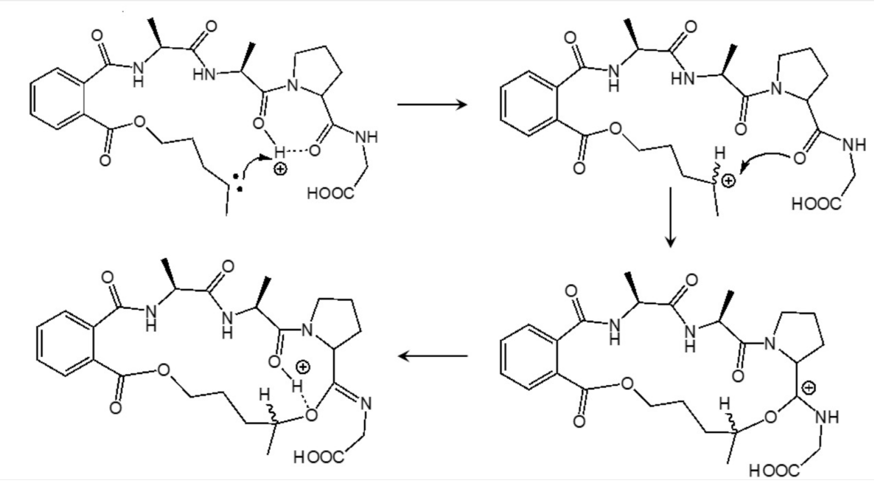
Low-energy s-AAHG ions were all protonated at the His imidazole nitrogens. This protonation allowed the ions to form hydrogen bonds, stabilizing the peptide chain in an extended conformation in the lowest energy structures, **H1** and **H2** (Figure 3.7), while the diazirine-containing side chain remained relatively free to move. In contrast, the higher-energy structure **H3** exhibited a more tightly folded conformation due to multiple hydrogen bonds formed between the neutral amide groups.



P1 (0.0 kJ mol^{-1}) **P2** (5.7) **P3** (7.0)

H1 (0.0 kJ mol^{-1}) **H2** (0.7) **H3** (19)

Figure 3.7: M06-2X/6-31+G(d,p)-optimized structures of low-energy s-AAPG and s-AAHG ions. Atom color coding is as follows: cyan = C, blue = N, red = O, gray = H. Only exchangeable hydrogens are shown to avoid clutter. Hydrogen bonds are indicated by double-headed blue arrows. The protonated amide groups are labeled with plus signs. Relative Gibbs energies in kJ mol^{-1} are from M06-2X/6-311++G(2d,p) single-point energy calculations including B3LYP zero-point energies, enthalpies, and entropies and referring to 310 K. Relative Gibbs energies of water-solvated ions are shown as purple italics.



Scheme 3.2: Proposed Crosslinking Mechanism in s-AAPG Involving the Ala2 Proton.

3.3.4 Atom-Atom Contact Analysis of s-AAPG and s-AAHG Thermal Ions

Contact analysis of s-AAPG trajectories starting from structures **P1** and **P2** revealed key aspects of the conformational dynamics (Figure 3.8). A significant portion of the contacts, when averaged over **P1**-**P3**, involved the Gly-COOH (10%), Gly amide NH (24%), and Pro Hα (15%). However, the Pro-Hα contacts were not associated with contacts involving the Pro Cα (e.g., 0% frequency at the 3.5 Å limit), which would be necessary for carbene insertion into the Pro Cα-H$_\alpha$ bond. A large portion of the contacts occurred with the OH and NH protons at the protonated Ala2 amide (45% and 2.5%, respectively), potentially leading to proton transfer to the basic carbene. The high frequency of these contacts was determined by the equilibrium conformations of **P1**-**P3**, which were largely stabilized by the strong hydrogen bond between the Ala2-OH$^+$ and the phthalate carbonyl (Figure 3.7), and this bond remained intact despite thermal motion at 350-640 K. Protonation of the carbene by Ala2-OH$^+$ can initiate a subsequent nucleophilic attack by the Pro and Gly amide or carboxyl atoms at the transient carbocation, resulting in macrocyclic ring closure (Scheme 3.2). The mechanism outlined in Scheme 3.2 aligns with the results of the crosslinking experiments and the CID-MS^3 spectrum of s-AAPG, which showed a marked reduction in peptide chain dissociation at Pro. According to the ion structures depicted in Scheme 3.2, cleavage of the susceptible tertiary peptide bond between Ala2 and Pro in the crosslinked ion cannot lead to the elimination of Pro-Gly, as it is tethered to the other side chain. Instead, it triggers the elimination of Ala, which is the predominant dissociation observed in the CID-MS^3 spectrum (Figure 3.4(a)).

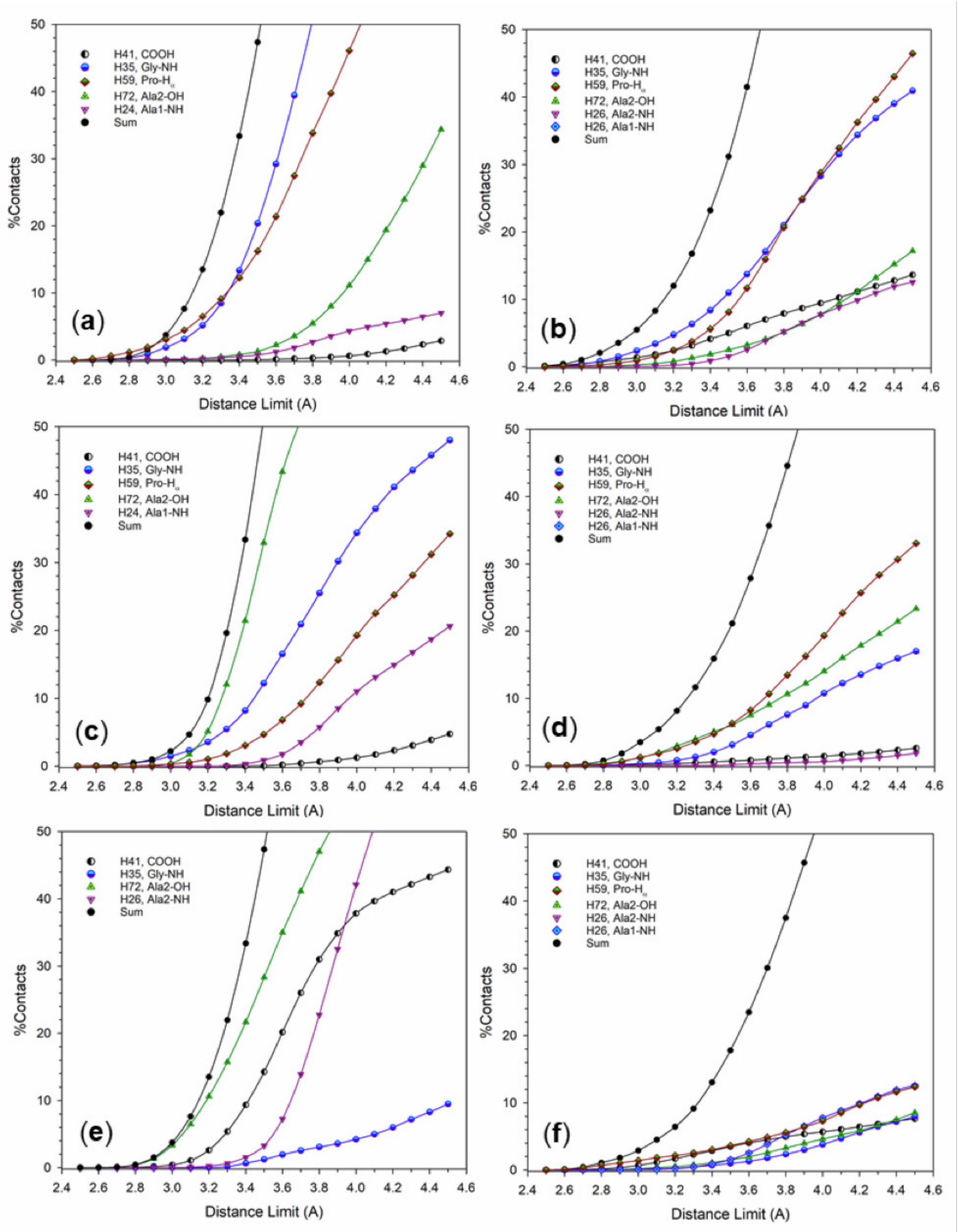


Figure 3.8: Carbene contacts along 100 ps BOMD trajectories for s-AAPG scaffold ions **P1** at (a) 350 K, (b) 640 K; **P2** at (c) 350 K, (d) 640 K; **P3** at (e) 350 K, and (f) 640 K. For equilibrium ion structures refer to Figure 3.7.

Contact analysis of s-AAHG ion trajectories provided results consistent with the crosslinking data, indicating a low efficiency for this peptide sequence. Figure 3.9(a) shows that conformer **H1** had frequent contacts with the His α-hydrogen; however, these were not accompanied by contacts with the His α-carbon atom. These contacts significantly diminished at 640 K (Figure 3.9(b)). The other low-energy conformer, **H2**, exhibited contacts with the Ala1 and Ala2 amide NH at 310 K, and these contacts were maintained even at 640 K (Figures 3.9(b) and 3.9(d)). We note that crosslinking to the Ala1 and Ala2 amides would not prevent these ions from undergoing peptide chain dissociation at the His-Gly sequence, which we used for analyzing the CID-MS^3 spectrum in Figure 3.6(b). Therefore, the contact analysis may suggest a higher yield of crosslinks in s-AAHG than what is inferred from the interpretation of the CID-MS^3 spectrum. It is noteworthy that, in contrast to the solution study[18], crosslinking to His in gas-phase s-AAHG did not stand out when compared with other amino acid residues. This discrepancy can be attributed to the fact that crosslinking yields in gas-phase ions are determined by the ions' static and dynamic conformations, without being influenced by solvent effects, reagent diffusion, or other factors present in the condensed phase.

3.4 Conclusions

This study expanded the investigation into the crosslinking behavior of peptide ion scaffolds containing basic amino acid residues, namely proline and histidine. The results indicated that the introduction of these residues significantly reduced crosslinking yields compared to hydrophobic scaffolds like those with alanine, leucine, and glycine. The s-AAPG and s-AAHG sequences were analyzed in detail to understand how protonation sites and peptide conformations influence carbene crosslinking.

In s-AAPG, the primary protonation site was identified as the proline tertiary amide, which, upon carbene formation, facilitated a nucleophilic attack leading to macrocyclic ring closure. The detailed BOMD simulations and atom-atom contact analysis suggested that the stable hydrogen-bond network around the proline residue governs this behavior. For s-AAHG, the protonation site was fixed at the histidine imidazole group. While computational data indicated a higher crosslinking potential in s-AAHG than observed experimentally, this discrepancy was attributed to structural constraints and dissociation behavior at the His-Gly sequence, which were not fully captured in the experimental setup.

Overall, the findings emphasize the influence of amino acid residues with increasing basicity on crosslink-

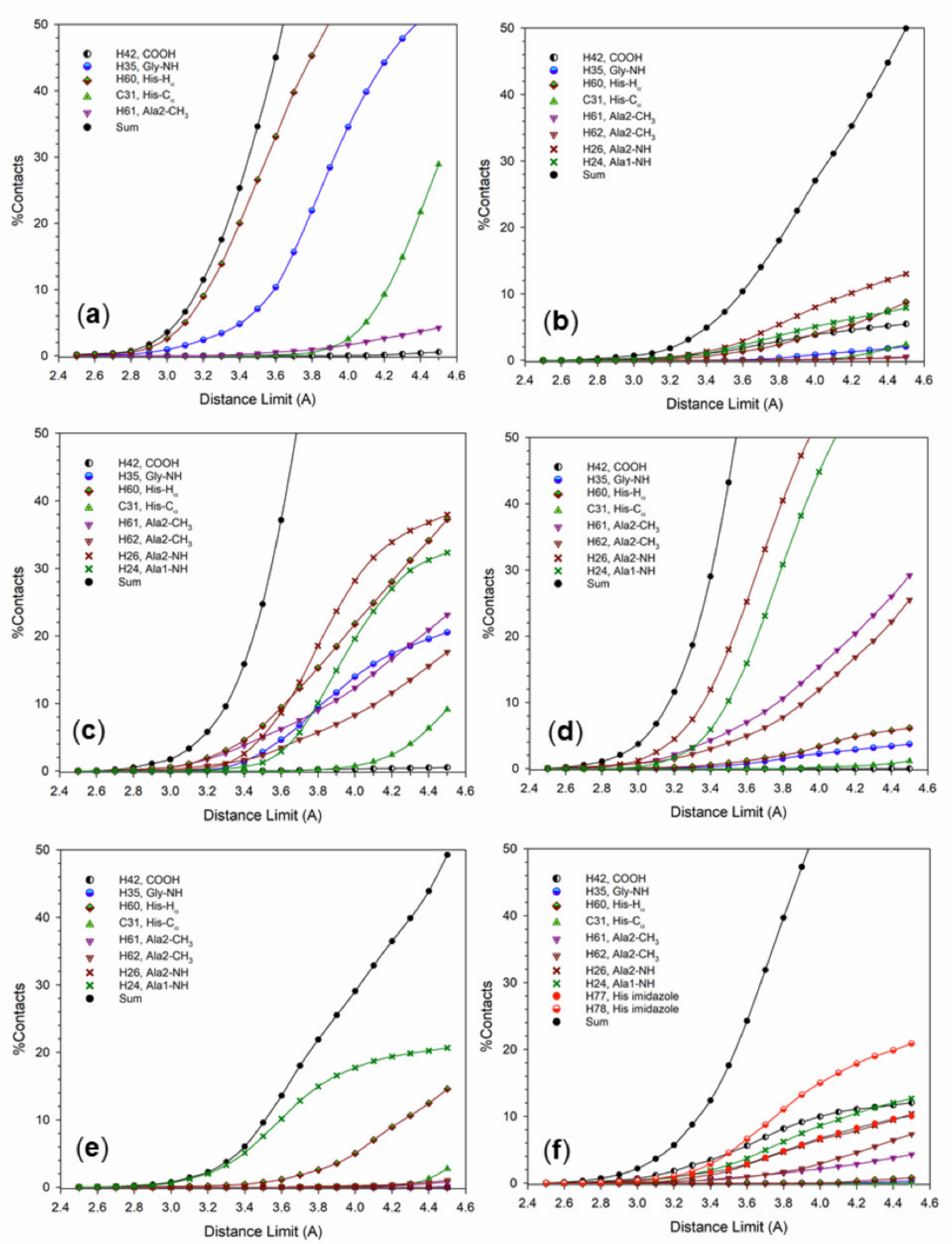


Figure 3.9: Carbene contacts along 100 ps BOMD trajectories for s-AAHG scaffold ions **H1** at (a) 350 K, (b) 640 K; **H2** at (c) 350 K, (d) 640 K, and **H3** at (e) 350 K, and (f) 640 K. For equilibrium ion structures refer to Figure 3.7.

ing efficiency and underscore the importance of computational simulations for predicting peptide behavior in gas-phase environments. The integration of experimental CID-MS^n data with advanced simulations provides a nuanced understanding of how scaffold structure, protonation, and peptide dynamics collectively determine crosslinking outcomes.

3.5 Bibliography

[1] Schwartz, B.L.; Bursey, M.M. “Some Proline Substituent Effects in the Tandem Mass Spectrum of Protonated Pentaalanine.” *Biol. Mass Spectrom.* **1992**, *21*, 92-96.

[2] Vaisar, T.; Urban, J. “Probing the proline effect in CID of protonated peptides.” *J. Mass Spectrom.* **1996**, *31*, 1185-1187.

[3] Raulfs M. D. M.; Breci L.; Bernier M.; Hamdy O. M.; Janiga A.; Wysocki V.; Poutsma J. C. “Investigations of the Mechanism of the “Proline Effect” in Tandem Mass Spectrometry Experiments: The “Pipecolic Acid Effect”.” *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1705-1715.

[4] Roepstorff P. “Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides.” *Biomed. Mass Spectrom.* **1984**, *11*, 601.

[5] Wan, J.; Brož, B.; Liu, Y.; Huang, S. R.; Marek, A.; Tureček, F. “The DNA Radical Code. Resolution of Identity in Dissociations of Trinucleotide Codon Cation Radicals in the Gas Phase." *J. Am. Soc. Mass Spectrom.* **2023**, *34*, 304-319.

[6] Řezáč, J. “Cuby: An Integrative Framework for Computational Chemistry." *J. Comput. Chem.* **2016**, *37*, 1230-1237.

[7] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. “Semi-Empirical Quantum Chemical PM6 Method Augmented by Dispersion and H Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes." *J. Chem. Theory Comput.* **2009**, *5*, 1749-1760.

[8] Shaffer, C. J.; Andrikopoulos, P. C.; Řezáč, J.; Rulíšek, L.; Tureček, F. “Efficient Covalent Bond Formation in Gas-Phase Peptide-Peptide Ion Complexes with the Photoleucine Stapler." *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 633-645.

[9] Liu, Y.; Ramey, Z.; Tureček, F. “Non-Covalent Interactions of a Neuroprotective Peptide Revealed by Photodissociative Cross-Linking in the Gas Phase.” *Chem.–Eur. J.* **2018**, *24*, 9259-9263.

[10] Nguyen, H. T. H.; Andrikopoulos, P. C.; Rulíšek, L.; Shaffer, C. J.; Tureček, F. “Photodissociative Cross Linking of Noncovalent Peptide-Peptide Ion Complexes in the Gas Phase.” *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1706-1720.

[11] Nguyen, H. T. H.; Huang, S. R.; Liu, Y.; Liu, Y.; Korn, J. A.; Tureček, F. “Probing Arginine-Phosphopeptide Interactions in Non-Covalent Peptide-Peptide Ion Complexes Using Gas-Phase Cross-Linking and Born-Oppenheimer Molecular Dynamics Calculations.” *Int. J. Mass Spectrom.* **2019**, *435*, 259-271.

[12] Huang, S. R.; Liu, Y.; Tureček, F. “Non-Covalent Complexes of the Amyloid Peptide Fragment Gly-Asn-Asn-Gln-Gln-Asn-Tyr. A Gas-Phase Photodissociative Cross-Linking and Born-Oppenheimer Molecular Dynamics Binding Study.” *Phys. Chem. Chem. Phys.* **2019**, *21*, 2046-2056.

[13] Liu, Y.; Tureček, F. “Photodissociative Crosslinking of Diazirine-Tagged Peptides with DNA Dinucleotides.” *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1992-2006.

[14] Liu, Y.; Liu, Y.; Nytka, M.; Huang, S. R.; Lemr, K.; Tureček, F. “Probing D- and L-Adrenaline Binding to β2-Adrenoreceptor Peptide Motifs by Photodissociation Crosslinking and Ion Mobility Mass Spectrometry.” *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1041-1052.

[15] Becke, A. D. “Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior.” *Phys. Rev. A* **1998**, *38*, 3098-3100.

[16] Zhao, Y.; Truhlar, D. G. “The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals.” *Theor. Chem. Acc.* **2008**, *120*, 215-241.

[17] Tomasi, J.; Mennucci, B.; Cammi, R. “Quantum Mechanical Continuum Solvation Models.” *Chem. Rev.* **2005**, *105*, 2999-3093.

[18] Ziemianowicz, D. S.; Bomgarden, R.; Etienne, C. W.; Schriemer, D. C. “Amino Acid Insertion Frequencies Arising from Photoproducts Generated Using Aliphatic Diazirines.” *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2011-2021.

Chapter 4

Nitrile Imine Crosslinking in Stereochemically Distinct Scaffolds of Peptides AAAG and AAAHG in the Gas-Phase

Reproduced in part with permission from Hongyi Zhu, Marianna Nytka, Tuan Ngoc Kim Vu, Karel Lemr, and František Tureček. Photochemical and Collision-Induced Cross-Linking in Stereochemically Distinct Scaffolds of Peptides and Nitrile Imines in Gas-Phase Ions. *Journal of the American Society for Mass Spectrometry*, **XXXX**, *XXX*, XXX-XXX.

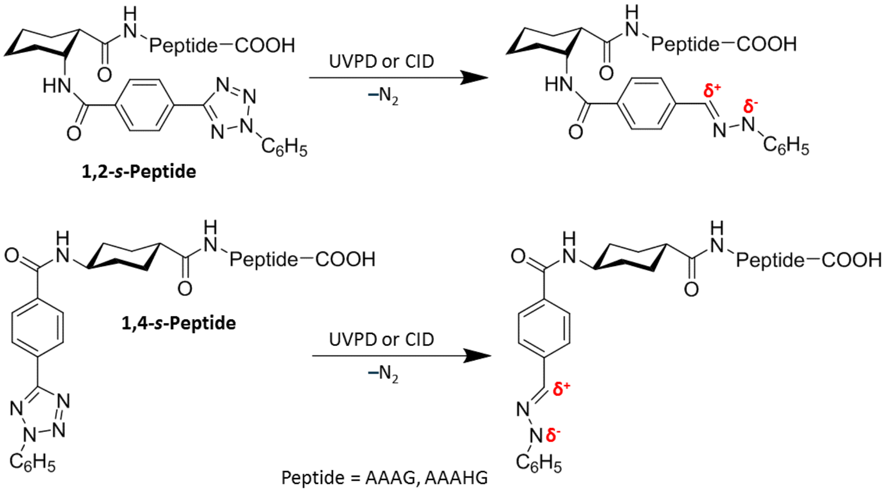
Abstract. Intramolecular crosslinking between peptides and nitrile imine intermediates was studied in stereochemically distinct conjugates, where the reacting components were mounted on *cis*-1,2-cyclohexane and *trans*-1,4-cyclohexane scaffolds, referred to as 1,2-s-peptides and 1,4-s-peptides, respectively. The nitrile imine intermediates were generated through N_2 loss from 2,5-diaryltetrazole tags upon UV-photodissociation at 213 and 250 nm or by collision-induced dissociation, and further analyzed by CID and UVPD-MS^3. Peptide fragment ion series originating from both linear structures and macrocyclic crosslinks were identified and used to quantify the crosslinking yields. The yields in MS^2 varied between 27% for AAAG conjugates and 65% for AAAHG conjugates, depending on the peptide sequence. The CID-MS^3 yields ranged from

57% to 91%, also influenced by the peptide sequence. The structures of 1,2-s-peptide and 1,4-s-peptide ions, along with several nitrile imine intermediates and crosslinked products, were investigated using high-resolution cyclic ion mobility in combination with Born–Oppenheimer molecular dynamics and density functional theory calculations. Matches between experimental and theoretical collision cross sections and ion relative Gibbs energies enabled structural assignment. Peptide conjugates C-terminated with glycine residues underwent crosslinking via the carboxyl group, as confirmed by MS^3 sequencing and supported by carboxyl blocking experiments, which reduced crosslinking yields.

4.1 Introduction

The development of new reagents has been in the forefront of crosslinking studies aimed at the determination of biomolecular structure by mass spectrometry.[1] In addition to photoreactive tags such as diazirines[2]-[4] and benzophenones[5]-[7] that can be introduced into selected amino acid residues, various bifunctional crosslinkers have been introduced[8] and used[9] that targeted lysine residues, and new variations have been proposed that were cleavable by collision-induced dissociation (CID) in tandem mass spectrometry analysis.[10]-[14] Cycloaddition reactions, such as the azide-alkyne click reaction, have expanded the portfolio of crosslinking strategies in solution.[15],[16] The tag approach has the advantage that at least one site of attachment is known so that the detection of crosslinks indicates the residues that are sterically accessible to the tagged ones.

We have been pursuing a different approach in which crosslinking was achieved by photodissociation of gas-phase peptide ions that were site-specifically tagged with diazirine groups.[17] The main aim of these studies has been to elucidate ion structures and determine peptide-peptide[18] and peptide-oligonucleotide[19] noncovalent interactions in gas-phase complexes. Gas-phase crosslinking offers some advantages over classical solution studies. First, the reactive intermediates are distinguished and selected by mass which allows one to quantify the crosslinking yields.[20] Second, the crosslink sites can be elucidated by tandem mass spectrometry *de novo* sequencing.[21] However, similar to crosslinking in solution, gas-phase crosslinking is a low resolution method, the chief limitation being the incomplete and peptide-dependent sequence coverage. Recently, we have reported on a discovery that nitrile imines undergo photoinduced crosslinking to peptides in conjugates that were tagged with a 2,5-diaryltetrazole group.[22] The diaryltetrazole is conve-



Scheme 4.1: Peptide-Nitrile-Imine Conjugates with Stereochemically Distinct *cis*-1,2- and *trans*-1,4-Cyclohexane Scaffolds.

niently attached to a suitable nucleophile in the peptide, such as the lysine ϵ-amine,[23] or cysteine SH.[22] Photodissociation at 250 nm breaks up the tetrazole ring[24] expelling N_2 and creating the reactive nitrile imine. The associated change of mass allows one to track the conversion and quantify the reaction yield.[22] Our previous investigation of nitrile imine crosslinking reactions has revealed that they proceeded as a proton-catalyzed [3+2] addition to the peptide amide group that was followed by ring opening and proton migration.[22] In an effort to further explore nitrile imine crosslinking, we now investigate conjugates in which peptide chains and the 2,5-diaryltetrazole moiety are mounted on cyclohexane scaffolds.[25] This design allows us to exploit the cyclohexane stereochemistry of *cis*-1,2- and *trans*-1,4-substituents to position the reacting moieties in a well defined steric arrangement and study its effect on crosslinking yields and attachment sites (Scheme 4.1).

The flexible nature of the *cis*-1,2-cyclohexane scaffold was expected to provide unimpeded access of the incipient nitrile imine to the peptide backbone amide and side-chain groups. In contrast, the *trans*-1,4-cyclohexane scaffold was presumed to exert steric hindrance to the peptide-nitrile-imine interaction in a fashion analogous to peptides containing 1,4-*cis*- and *trans*-cycloornithine residues.[26] Here, we explore these potential stereochemical effects on crosslinking with peptides containing different functional groups, such as carboxyl, methoxycarbonyl, and imidazole that may compete in their reactions with the nitrile imine group.[27] First the photodissociation and collision-induced dissociation (CID) tandem mass spectra of peptide-tetrazole conjugates with the aim of identifying the fragment ions and establishing crosslinking yields is presented and discussed. This is followed by structure analysis of the conjugates and dissociation products by high-resolution cyclic ion mobility mass spectrometry in combination with Born-Oppenheimer molecular dynamics (BOMD) and density functional theory (DFT) calculations of ion structures to obtain theoretical collision cross sections (CCS_{calc}) for comparison with experimental data.

4.2 Experimental Section

4.2.1 Synthetic Procedures

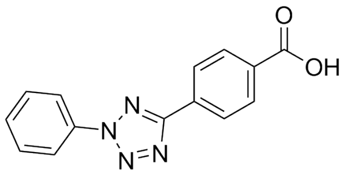
Unless otherwise noted, all starting materials were from (Sigma Aldrich, Fisher Chemical) and were used as received. ^{1}H NMR spectra were recorded on Bruker spectrometers at 300 MHz. Coupling constants (J) are given in Hz and chemical shifts are on the δ-scale. HRMS spectra were recorded on a Thermo Orbitrap Ascend Tribrid Mass Spectrometer. Samples were introduced by flow injection into the ESI source.

4.2.2 Materials and Methods

Abbreviations Used

Fmoc: fluorenylmethyloxycarbonyl
DMF: *N, N*-dimethylformamide
PyBOP: (bensotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
HOBt: 1-hydroxybenzotriazole
NMM: 4-methylmorpholine
DCM: dichloromethane
TFA: trifluoroacetic acid
DCC: dicyclohexylcarbodiimide
DMAP: 4-dimethylaminopyridine

4.2.2.1 Synthetic Procedure for 4-(2-Phenyl-*2H*-tetrazol-5-yl)benzoic acid



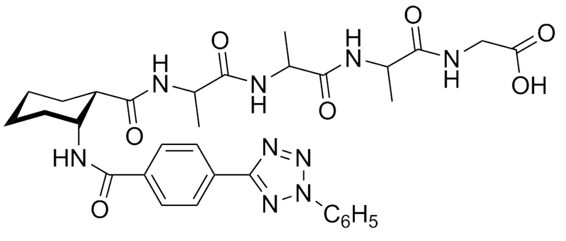
This synthesis was performed following the published literature procedure.[23] 4-Formylbenzoic acid (0.75 g, 5 mmol) was dissolved in 50 mL ethanol, to which 0.86 g (25 mmol) of benzenesulfonylhydrazide was added. The mixture was stirred for 30 min. A white precipitate was formed after adding 100 mL water and was collected in a funnel. This precipitate was then dissolved in 30 ml pyridine to give solution A. In parallel, 0.345 g (5 mmol) of $NaNO_2$ was dissolved in 2 mL of water and added dropwise to a cooled mixture of 0.465 g (5 mmol) of aniline dissolved in 8 mL of a 1:1 water-ethanol mixture and 1.3 mL of concentrated HCl, giving solution B. Solution B was slowly added to solution A, cooled with an ice-salt bath. The reaction mixture was then extracted 3 times with ethyl acetate. Precipitate was formed after adding 250 mL of 3M HCl to the combined organic layers and was collected and dried to give the title compound (0.5 g, 35%): ^{1}H NMR (300 MHz, DMSO-d_6) δ 8.31 (d, J = 6.0 Hz, 2H), 8.21-8.16 (m, 4H), 7.75-7.62 (m, 3H).

4.2.2.2 General Procedure for Solid Phase Peptide Synthesis

Resin containing an appropriate Fmoc protected C-terminal amino acid (50 mg, 1 equiv.) was weighed into a 3 mL syringe. The beads were swollen with DMF for 20 min. The Fmoc protective group on the beads was removed by 2 mL 20% piperidine in DMF after shaking for 10 min and the beads were washed with DMF 6 times. The coupling cocktail solution for peptide chain elongation was prepared by dissolving a Fmoc protected amino acid (3 equiv.), PyBOP (47 mg, 3 equiv.), HOBt (17 mg, 4.2 equiv) and 25 μL NMM in 1 mL DMF. The coupling cocktail solution was injected into the syringe and shaken for 20 min for the reaction to complete. Then the beads were washed with DMF 6 times. The Fmoc deprotection, DMF wash, coupling, and DMF wash were repeated in each peptide alongation step until the desired peptide sequence was obtained. After the last amino acid was added and the Fmoc group was deprotected, 4-(2-phenyl-2H-tetrazol-5-yl)benzoic acid was coupled to the peptide N-terminus by adding a solution made by dissolving 4-(2-phenyl-2H-tetrazol-5-yl)benzoic acid (48 mg, 6 equiv.), PyBOP (94 mg, 6 equiv.), HOBt (34 mg, 8.4 equiv) and 25 μL NMM in 1 mL DMF, and shaking the reaction mixture overnight. Then, the beads were washed with DMF 6 times, and washed with DCM 6 times to remove DMF. The peptide conjugate was stripped from the resin by injecting a solution mixture of TFA:water:triethylsilane (95:2.5:2.5) twice.

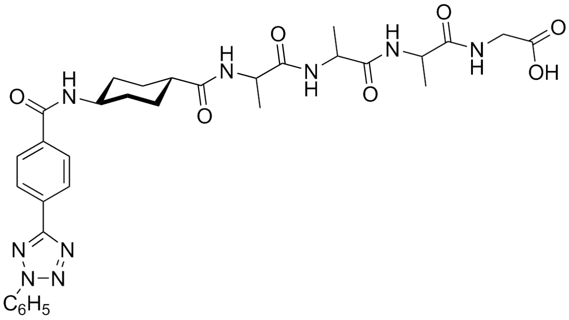
The TFA solution containing the peptide product was collected in a small glass vial and the solvent was evaporated with air flow to give ∼15 mg of the peptide conjugate product.

1,2-*s*-AAAG



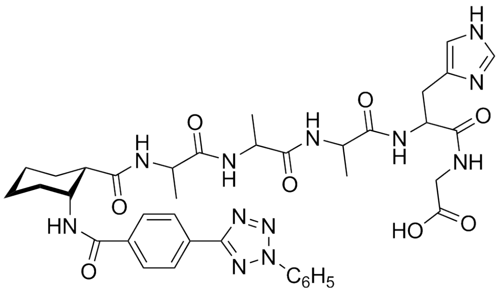
The title conjugate was prepared according to the general procedure for solid phase peptide synthesis. HRMS (ESI) m/z calcd for $C_{32}H_{40}N_9O_7$ $(M+H)^+$ 662.3045, found 662.3041.

1,4-*s*-AAAG



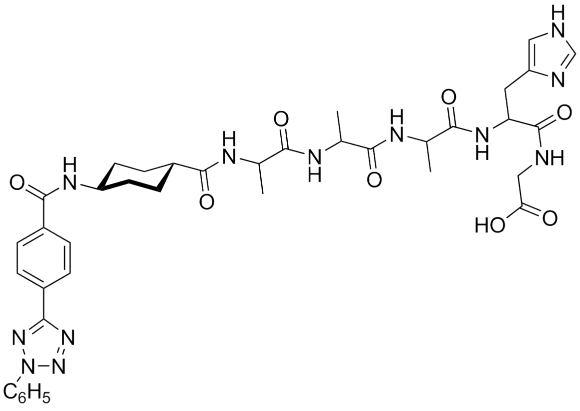
The title conjugate was prepared according to the general procedure for solid phase peptide synthesis. HRMS (ESI) m/z calcd for $C_{32}H_{40}N_9O_7$ $(M+H)^+$ 662.3045, found 662.3037.

1,2-*s*-AAAHG



The title conjugate was prepared according to the general procedure for solid phase peptide synthesis. HRMS (ESI) m/z calcd for $C_{38}H_{47}N_{12}O_8$ $(M+H)^+$ 799.3634, found 799.3620.

1,4-*s*-AAAHG

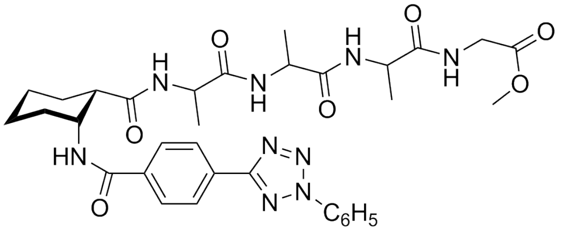


The title conjugate was prepared according to the general procedure for solid phase peptide synthesis. HRMS (ESI) m/z calcd for $C_{38}H_{47}N_{12}O_8$ $(M+H)^+$ 799.3634, found 799.3608.

4.2.2.3 General Procedure for Peptide C-Terminus Methylation

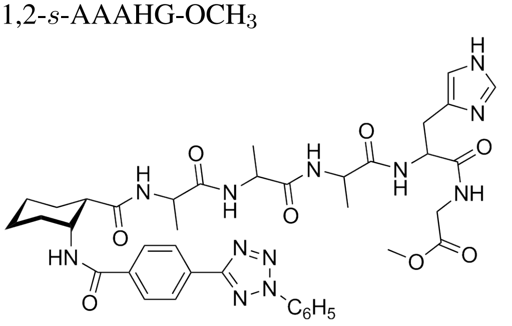
2-3 mg (1 equiv.) of the corresponding peptide conjugate was added to 1 mL methanol. DCC (1 equiv.) and DMAP (0.1 equiv.) were added to the reaction mixture and stirred overnight at room temperature. The precipitated dicyclohexylurea was removed by centrifugation, the solution was evaporated to dryness with a flow of nitrogen and reconstituted in aqueous methanol for electrospray ionization.

1,2-*s*-AAAG-OCH_3



The title conjugate was prepared according to the general procedure for methylation. HRMS (ESI) m/z calcd for $C_{33}H_{42}N_9O_7$ $(M+H)^+$ 676.3202, found 676.3212.

1,2-*s*-AAAHG-OCH_3

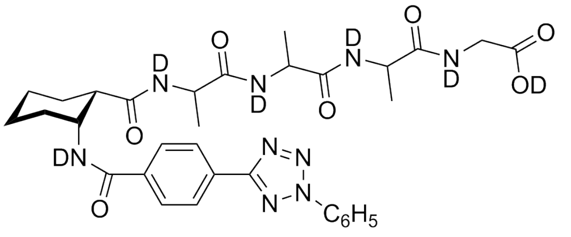


The title conjugate was prepared according to the general procedure for methylation. HRMS (ESI) m/z calcd for $C_{39}H_{49}N_{12}O_8$ $(M+H)^+$ 813.3791, found 813.3825.

4.2.2.4 General Procedure for H/D Exchange

The desired peptide conjugate was dissolved in MeOD:D_2O 1:1 and diluted with MeOD:D_2O: CD_3COOD in a 1:1:0.01 ratio for electrospray to give the (D_n + D) ions.

[D_6]-1,2-s-AAAG



The title conjugate was prepared according to the general procedure for H/D exchange. It was characterized by MS/MS: *m/z* 669 ([D_6]M+D)$^+$, 591 (b_3, –$C_2H_2D_3NO_2$), 519 (b_2, –$C_5H_6D_4N_2O_3$), 447 (b_1, –$C_8H_{10}D_5N_3O_4$).

4.2.3 Cyclic Traveling Wave Ion Mobility-Mass Spectrometry

4.2.3.1 Sample Preparation

LC-MS grade acetonitrile (ACN) and water were purchased from Fisher Chemical (Fisher Scientific, United Kingdom). Acetic acid $\geq$ 99% was obtained from Sigma Aldrich (Prague, Czech Republic). Poly-DL-alanine was purchased from Sigma Aldrich (Prague, Czech Republic).

Solid samples were dissolved in ACN/H_2O (50/50) to prepare stock solutions with the following concentrations: 2 mg/mL for 1,2-s-AAAG; 2 mg/mL for 1,4-s-AAAG; 2 mg/mL for 1,2-s-AAAHG; 2.37 mg/mL for 1,4-s-AAAHG.

The stock solutions were diluted to 1 mg/mL in ACN/H_2O (50:50) and then diluted to a working concentration of 1 μg/mL in ACN/H2O (1:1) with 1% acetic acid. The stock solution of polyalanine (1 mg/ml in ACN/H_2O (50/50)) was diluted to 2 μg/mL in ACN/H_2O (50/50) with 0.1% of formic acid. The calibrant was prepared in duplicate (one data acquisition for each ion).

4.2.3.2 ESI Conditions

The instrument parameters were as follows: capillary voltage 2 kV; cone voltage 10 V; source offset 10 V; source temperature 100 °C; desolvation temperature 280 °C; cone gas flow 30 L/h; desolvation gas flow

600 L/h; trap 6 V; transfer 4 V; stepwave body gradient 20 V; ion guide TW pulse height 0.4 V; trap TW pulse height 4 V; trap entrance 2 V; trap bias 2 V; trap DC -4 V; post trap gradient 3 V; post trap bias 35 V; 4 TOF pushes per bin for $[M+H]^+$ and for $[M+H-N_2]^+$. The precursor ion was isolated by the quadrupole and fragmented in the trap (45V) to knock off N_2.

Table 4.1: cIMS conditions for 1 pass separation experiments for $[M+H]^+$ and for $[M+H-N_2]^+$ in the brackets

Traveling Wave Parameters		**Parameter**	**Sequence Inject**	**Separate**	**Eject and Acquire**
Cyclic TW Velocity (m/s)	375	Time	10	2	Automatic
Array TW Velocity (m/s)	375	Pre Array Gradiant	85	85	85
TW Static Height (V)	18 (15)	Pre Array Bias	70	70	70
TW Start Height (V)	18 (15)	Array Entrance	10	30	50
TW Limit Height (V)	35	Wave Height	2	0	25
TW Ramping Rate (V/ms)	2.5	Array Offset	45	70	45
		Array Mode	Forward	Sideways	Forward Eject
		Array Exit	50	30	2
		Post Array Gradient	35	35	35
		Post Array Bias	10	10	10

4.2.3.3 CCS Calibration

Calibrant arrival times (t_{arr}) as well as analytes corresponding to peak maxima were obtained from one and two pass experiments (98 and 196 cm path length). Arrival time is the value that includes injection time (time needed for ions to arrive on the T-Wave array), drift time (time that ions spend inside the mobility cell), and dead time (time the ions exit the mobility cell to reach the detector).

The calibration and analytes drift times were calculated as t_d = t_{arr}(two passes) – t_{arr} (one pass). The dead time was t_0 = t_{arr}(one pass) — t_d. For determining CCS values, the drift times of analytes (t_d(analyte) = t_{arr}(analyte) – t_0 (an average of 6 measurements)) were divided by the number of passes. Peak maxima were

Table 4.2: Separation Time for Multipass Measurements for Different Analytes

Precursor Ion	**Number of Passes**	**Separation Time (ms)**
1,2-s-AAAG $(M + H)^+$ (*m/z* 662.4)	20	430
1,4-s-AAAG $(M + H)^+$ (*m/z* 662.4)	15	360
1,2-s-AAAHG $(M + H)^+$ (*m/z* 799.4)	4	90
1,2-s-AAAHG $(M + H - N_2)^+$ (*m/z* 771.4)	4	90
1,4-s-AAAHG $(M + H)^+$ (*m/z* 799.4)	15	453
1,4-s-AAAHG $(M + H - N_2)^+$ (*m/z* 771.4)	4	93

detected by DriftScope 2.9. An Excel file, provided by Waters Corp. (© Johanna Hofmann, Kevin Pagel, Fritz-Haber-Institute of the Max-Planck-Society, Berlin, Germany), was used to determine CCS values. A calibration curve was constructed by plotting ln(CCS', Å^2) vs. ln($t_{d'}$) (y = 0.4419X + 5.8235, R^2=0.9999, containing 16 points), where CCS' is the collision cross section corrected for ion charge and reduced mass of ion and drift gas (nitrogen); $t_{d'}$ is a corrected drift time for *m/z* dependent flight time (depends on the Enhanced Duty Cycle (ECD) delay coefficient, which was 1.50 ms for our experiments: $t_{d'} = t_{d\ analyte} -$ 0.001 * EDC delay coefficient * $\sqrt{m/z_{analyte}}$.

Table 4.3: Ions Used for CCS Calibration

m/z	**z**	**Mol. Mass (Da)**	^{DT}**CCS**$_{N_2}$ (**Å^2**)
374.20	1	373.20	180.30
445.24	1	444.20	194.80
516.28	1	515.30	209.70
587.32	1	586.30	226.20
658.35	1	657.40	239.90
729.39	1	728.40	252.50

(Continued)

m/z	z	Mol. Mass (Da)	^{DT}CCS$_{N_2}$ (Å^2)
800.43	1	799.40	265.70
871.46	1	870.50	278.50

4.2.3.4 CCS Calibration for Slicing Experiments

Each calibrant ion was sent to 4 passes, sliced, and then sent to 1 and 2 pass experiments in order to calculate the CCS values of analytes after the slicing experiment the same way as described above.

Table 4.4: cIMS Conditions for Slicing Measurements of Polyalanine Ions[a]

	***m/z* 374.20**	***m/z* 445.24**	***m/z* 516.28**	***m/z* 587.32**	***m/z* 658.35**	***m/z* 729.39**	***m/z* 800.43**	***m/z* 871.46**
Inject	10	10	10	10	10	10	10	10
Separate	32	40	45	40	64	70	80	90
Eject	4.74	4.04	7.92	7.83	8.52	11.06	12.66	14.10
Eject to Pre-Store	3.59	5.13	5.88	4.89	10.81	10.72	9.17	14.80
Hold and Eject	18.74	19.73	31.5	23.82	31.79	24.77	19.98	18.98
Reinject from Pre-Store	3	3	3	3	3	3	3	3
Separate	2(12)	2(16)	2(18)	2(20)	2(30)	2(30)	2(30)	2(30)

[a] The first separate indicates 4 pass experiment and the second indicates 1 (2) pass experiment.

4.2.4 Accurate Mass Measurements and Assignments

Table 4.5: Accurate Mass Measurements of UVPD-MS^2 of (1,2-*s*-AAAG + H)$^+$ Ions

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
662.3041	$C_{32}H_{40}N_9O_7$	-0.4		
634.2981	$C_{32}H_{40}N_7O_7$	-0.3	N_2	
616.2874	$C_{32}H_{38}N_7O_6$	-0.4	H_2N_2O	$-N_2-H_2O$

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
589.2766	$C_{31}H_{37}N_7O_6$	-0.3	CH_3N_3O	–N_2–CO–NH_3
587.2721	$C_{30}H_{35}N_8O_5$	-0.4	C_2H_5NO	b_4
577.2765	$C_{30}H_{37}N_6O_6$	-0.4	$C_2H_3N_3O$	–N_2–[Gly] (β_4)
559.2660	$C_{30}H_{35}N_6O_5$	-0.3	$C_2H_5N_3O_2$	–N_2–GlyOH
543.2558	$C_{26}H_{35}N_6O_7$	-0.4	$C_6H_5N_3$	–N_2–C_6H_5N
516.2351	$C_{27}H_{30}N_7O_4$	-0.3	$C_5H_{10}N_2O_3$	b_3
506.2394	$C_{27}H_{32}N_5O_5$	-0.4	$C_5H_8N_4O_2$	*m/z* 577–[Ala] (β_3)
488.2290	$C_{27}H_{30}N_5O_4$	-0.2	$C_5H_{10}N_4O_3$	*m/z* 559–[Ala]
468.2239	$C_{24}H_{30}N_5O_5$	-0.2	$C_8H_{10}N_4O_2$	*m/z* 543–GlyOH
445.1980	$C_{24}H_{25}N_6O_3$	-0.3	$C_8H_{15}N_3O_4$	b_2
435.2024	$C_{24}H_{27}N_4O_4$	-0.3	$C_8H_{13}N_5O_3$	*m/z* 506–[Ala] (β_2)
417.1920	$C_{24}H_{25}N_4O_3$	-0.1	$C_8H_{15}N_5O_4$	*m/z* 488–[Ala]
414.2345	$C_{18}H_{32}N_5O_6$	-0.2	$C_{14}H_8N_4O$	y_5
397.1868	$C_{21}H_{25}N_4O_4$	-0.2	$C_{11}H_{15}N_5O_3$	*m/z* 468–[Ala]
374.1610	$C_{21}H_{20}N_5O_2$	-0.2	$C_{11}H_{20}N_4O_5$	b_1
364.1654	$C_{21}H_{22}N_3O_3$	-0.2	$C_{11}H_{18}N_6O_4$	*m/z* 435–[Ala] (β_1)
346.1548	$C_{21}H_{20}N_3O_2$	-0.2	$C_{11}H_{20}N_6O_5$	*m/z* 417–[Ala]
326.1497	$C_{18}H_{20}N_3O_3$	-0.2	$C_{14}H_{20}N_6O_4$	*m/z* 417–C_6H_5N
309.1556	$C_{14}H_{21}N_4O_4$	-0.1	$C_{18}H_{19}N_5O_3$	
289.1506	$C_{11}H_{21}N_4O_5$	-0.04	$C_{21}H_{19}N_5O_2$	y_4
271.1400	$C_{11}H_{19}N_4O_4$	-0.1	$C_{21}H_{21}N_5O_3$	*m/z* 289–H_2O
255.1127	$C_{15}H_{15}N_2O_2$	-0.1	$C_{17}H_{25}N_7O_5$	*m/z* 326–[Ala]
234.1236	$C_{12}H_{16}N_3O_2$	-0.1	$C_{20}H_{24}N_6O_5$	
221.0709	$C_{14}H_9N_2O$	-0.04	$C_{18}H_{31}N_7O_6$	*m/z* 346–$C_7H_{11}NO$

193.0647	$C_{14}H_9O$	-0.1		

Table 4.6: Accurate Mass Measurements of CID-MS^2 of (1,2-*s*-AAAG + H)$^+$ Ions

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
662.3041	$C_{32}H_{40}N_9O_7$	-0.4		
587.2718	$C_{30}H_{35}N_8O_5$	-0.7	$C_2H_5NO_2$	b_4
516.2347	$C_{27}H_{30}N_7O_4$	-0.7	$C_5H_{10}N_2O_3$	b_3
445.1979	$C_{24}H_{25}N_6O_3$	-0.4	$CH_8H_{15}N_3O_4$	b_2

Table 4.7: Accurate Mass Measurements of CID-MS^3 of Ion (1,2-*s*-AAAG + H – N$_2$)$^+$ (*m/z* 634) Generated by UVPD of (1,2-*s*-AAAG + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
634.2976	$C_{32}H_{40}N_7O_7$	-0.8		
616.2872	$C_{32}H_{38}N_7O_6$	-0.6	H_2O	
606.3028	$C_{31}H_{40}N_7O_6$	-0.7	CO	
589.2763	$C_{31}H_{37}N_6O_6$	-0.6	CH_3NO	
577.2761	$C_{30}H_{37}N_6O_6$	-0.8	C_2H_3NO	–[Gly] (β_4)
563.2607	$C_{29}H_{35}N_6O_6$	-0.6	C_3H_5NO	–[Ala]
559.2656	$C_{30}H_{35}N_6O_5$	-0.7	$C_2H_5NO_2$	b_4
518.2392	$C_{28}H_{32}N_5O_5$	-0.6	$C_4H_8N_2O_2$	*m/z* 589–[Gly]
509.2136	$C_{25}H_{29}N_6O_6$	-0.7	$C_7H_{11}NO$	
506.2391	$C_{27}H_{32}N_5O_5$	-0.7	$C_5H_8N_2O_2$	–[GlyAla] (β_3)
492.2236	$C_{26}H_{30}N_5O_5$	-0.5	$C_6H_{10}N_2O_2$	*m/z* 563–[Ala]
488.2287	$C_{27}H_{30}N_5O_4$	-0.5	$C_5H_{10}N_2O_3$	b_3
435.2021	$C_{24}H_{27}N_4O_4$	-0.6	$C_8H_{13}N_3O_3$	–[GlyAlaAla] (β_2)

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
417.1918	$C_{24}H_{25}N_4O_3$	-0.3	$C_8H_{15}N_3O_4$	b_2
364.1652	$C_{21}H_{22}N_3O_3$	-0.4	$C_{11}H_{18}N_4O_4$	–[GlyAlaAlaAla] (β_1)
271.1399	$C_{11}H_{19}N_4O_4$	-0.2	$C_{21}H_{21}N_3O_3$	
228.0977	$C_9H_{14}N_3O_4$	-0.2	$C_{23}H_{26}N_4O_3$	

Table 4.8: Accurate Mass Measurements of CID-MS^2 of (1,2-*s*-AAAG-OCH_3 + H)$^+$ Ions

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
676.3227	$C_{33}H_{42}N_9O_7$	2.5		
658.3124	$C_{33}H_{30}N_9O_6$	2.8	H_2O	
587.2741	$C_{30}H_{35}N_8O_5$	1.6	$C_3H_7NO_2$	b_4
516.2371	$C_{27}H_{30}N_7O_4$	1.7	$C_6H_{12}N_2O_3$	b_3
445.1998	$C_{24}H_{25}N_6O_3$	1.5	$C_9H_{17}N_3O_4$	b_2
374.1626	$C_{21}H_{20}N_5O_2$	1.5	$C_{12}H_{22}N_4O_5$	b_1

Table 4.9: Accurate Mass Measurements of UVPD-MS^2 of (1,2-*s*-AAAG-OCH_3 + H)$^+$ Ions

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
676.3212	$C_{33}H_{42}N_9O_7$	1.0		
648.3157	$C_{33}H_{42}N_7O_7$	1.7	N_2	
587.2735	$C_{30}H_{35}N_8O_5$	1.0	$C_3H_7NO_2$	b_4
559.2672	$C_{30}H_{35}N_6O_5$	0.9	$C_3H_7N_3O_2$	–N_2–GlyOCH_3

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
516.2363	$C_{27}H_{30}N_7O_4$	0.9	$C_6H_{12}N_2O_3$	b_3
488.2302	$C_{27}H_{30}N_5O_4$	1.0	$C_6H_{12}N_4O_3$	*m/z* 559–[Ala]
468.2250	$C_{24}H_{30}N_5O_5$	0.9	$C_9H_{12}N_4O_2$	*m/z* 559–C_6H_5N
445.1990	$C_{24}H_{25}N_6O_3$	0.7	$C_9H_{17}N_3O_4$	b_2
428.2511	$C_{19}H_{34}N_5O_6$	0.7	$C_{14}H_8N_4O$	y_5
417.1929	$C_{24}H_{25}N_4O_3$	0.8	$C_9H_{17}N_5O_4$	*m/z* 488–[Ala]
397.1877	$C_{21}H_{25}N_4O_4$	0.7	$C_{12}H_{17}N_5O_3$	*m/z* 468–[Ala]
374.1618	$C_{21}H_{20}N_5O_2$	0.6	$C_{12}H_{22}N_4O_5$	b_1
346.1556	$C_{21}H_{20}N_3O_2$	0.6	$C_{12}H_{22}N_6O_5$	*m/z* 417–[Ala]
339.2033	$C_{16}H_{27}N_4O_4$	0.6	$C_{17}H_{15}N_5O_3$	$C_7H_{11}NO$[AlaAlaAla]
326.1504	$C_{18}H_{20}N_3O_3$	0.5	$C_{15}H_{22}N_6O_4$	*m/z* 417–C_6H_5N
305.1614	$C_{15}H_{21}N_4O_3$	0.6	$C_{18}H_{21}N_5O_4$	
303.1670	$C_{12}H_{23}N_4O_5$	0.7	$C_{21}H_{19}N_5O_2$	y_4
294.1818	$C_{15}H_{24}N_3O_3$	0.6	$C_{18}H_{18}N_6O_4$	*m/z* 339–NH_3–CO
288.1349	$C_{15}H_{18}N_3O_3$	0.6	$C_{18}H_{24}N_6O_4$	*m/z* 305–NH_3
268.1662	$C_{13}H_{22}N_3O_3$	0.6	$C_{20}H_{20}N_6O_4$	$C_7H_{11}NO$[AlaAla]
255.1134	$C_{15}H_{15}N_2O_2$	0.6	$C_{18}H_{27}N_7O_5$	*m/z* 326–[Ala]
234.1242	$C_{12}H_{16}N_3O_2$	0.5	$C_{21}H_{26}N_6O_5$	*m/z* 305–[Ala]
221.0714	$C_{14}H_9N_2O$	0.5	$C_{19}H_{33}N_7O_6$	*m/z* 346–$C_7H_{11}NO$
197.1288	$C_{10}H_{17}N_2O_2$	0.3	$C_{23}H_{25}N_7O_5$	$C_7H_{11}NO$[Ala]
193.0652	$C_{14}H_9O$	0.4	$C_{19}H_{33}N_9O_6$	*m/z* 221–N_2

Table 4.10: Accurate Mass Measurements of CID-MS^3 of (1,2-*s*-AAAG-OCH$_3$ + H – N$_2$)$^+$ Generated by UVPD of (1,2-*s*-AAAG-OCH$_3$ + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
648.3149	$C_{33}H_{42}N_7O_7$	0.9		
630.3041	$C_{33}H_{40}N_7O_6$	0.6	H_2O	
620.3198	$C_{32}H_{42}N_7O_6$	0.7	CO	
604.3249	$C_{32}H_{42}N_7O_5$	0.7	CO_2	
577.2775	$C_{30}H_{37}N_6O_6$	0.6	C_3H_5NO	–[HN=CHCOCH$_3$] (β_4)
559.2669	$C_{30}H_{35}N_6O_5$	0.6	$C_3H_7NO_2$	b_4
539.2619	$C_{27}H_{35}N_6O_6$	0.6	C_6H_7NO	–C_6H_5N–H_2O
506.2402	$C_{27}H_{32}N_5O_5$	0.4	$C_6H_{10}N_2O_2$	–[AlaC$_3$H$_5$NO] (β_3)
488.2297	$C_{27}H_{30}N_5O_4$	0.5	$C_6H_{12}N_2O_3$	b_3
435.2031	$C_{24}H_{27}N_4O_4$	0.4	$C_9H_{15}N_3O_3$	–[AlaAlaC$_3$H$_5$NO] (β_2)
428.2507	$C_{19}H_{34}N_5O_6$	0.3	$C_{14}H_8N_2O$	y_5
417.1927	$C_{24}H_{25}N_4O_3$	0.6	$C_9H_{17}N_3O_4$	b_2
397.1875	$C_{21}H_{25}N_4O_4$	0.5	$C_{12}H_{17}N_3O_3$	b_3–C_6H_5N
364.1660	$C_{21}H_{22}N_3O_3$	0.4	$C_{12}H_{20}N_4O_4$	–[AlaAlaAlaC$_3$H$_5$NO] (β_1)
346.1554	$C_{21}H_{20}N_3O_2$	0.4	$C_{12}H_{22}N_4O_5$	b_1
326.1503	$C_{18}H_{20}N_3O_3$	0.4	$C_{15}H_{22}N_4O_4$	b_2–C_6H_5N
305.1612	$C_{15}H_{21}N_4O_3$	0.4	$C_{18}H_{21}N_3O_4$	
285.1562	$C_{12}H_{21}N_4O_4$	0.5	$C_{21}H_{21}N_3O_3$	
255.1132	$C_{15}H_{15}N_2O_2$	0.4	$C_{18}H_{27}N_5O_5$	b_1–C_6H_5N
242.1139	$C_{10}H_{16}N_3O_4$	0.4	$C_{23}H_{26}N_4O_3$	
234.1241	$C_{12}H_{16}N_3O_2$	0.4	$C_{21}H_{26}N_4O_5$	

Table 4.11: Accurate Mass Measurements of UVPD-MS^2 of (1,4-*s*-AAAG + H)$^+$ Ions

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
662.3014	$C_{32}H_{40}N_9O_7$	-3.1		
634.2973	$C_{32}H_{40}N_7O_7$	-1.1	N_2	
616.2867	$C_{32}H_{38}N_7O_6$	-1.1	H_2N_2O	–N_2–H_2O
587.2715	$C_{30}H_{35}N_8O_5$	-1.0	$C_2H_5NO_2$	b_4
577.2758	$C_{30}H_{37}N_6O_6$	-1.1	$C_2H_3N_3O$	–N_2–[Gly]
559.2653	$C_{30}H_{35}N_6O_5$	-1.0	$C_2H_5N_3O_2$	–N_2–GlyOH
551.3275	$C_{32}H_{39}N_8O$			
533.9087				
516.2345	$C_{27}H_{30}N_7O_4$	-0.9	$C_5H_{10}N_2O_3$	b_3
506.2389	$C_{27}H_{32}N_5O_5$	-0.9	$C_5H_8N_4O_2$	m/z 577–[Ala]
495.2650				
492.2233	$C_{26}H_{30}N_5O_5$	-0.8	$C_6H_{10}N_4O_2$	–N_2–2[Ala]
488.2285	$C_{27}H_{30}N_5O_4$	-0.7	$C_5H_{10}N_4O_3$	*m/z* 559–[Ala]
468.2233	$C_{24}H_{30}N_5O_5$	-0.8	$C_8H_{10}N_4O_2$	*m/z* 559–C_6H_5N
461.2175	$C_{26}H_{29}N_4O_4$	-0.8	$C_6H_{11}N_5O_3$	*m/z* 506–NH_3–CO
445.1975	$C_{24}H_{25}N_6O_3$	-0.8	$C_8H_{15}N_3O_4$	b_2
435.2019	$C_{24}H_{27}N_4O_4$	-0.8	$C_8H_{13}N_5O_3$	*m/z* 506–[Ala]
417.1975	$C_{24}H_{25}N_4O_3$	5.4	$C_8H_{15}N_5O_4$	*m/z* 488–[Ala]
407.2071	$C_{23}H_{27}N_4O_3$	-0.7	$C_9H_{13}N_5O_4$	*m/z* 435–CO
397.1864	$C_{21}H_{25}N_4O_4$	-0.6	$C_{11}H_{15}N_5O_3$	*m/z* 468–[Ala]
390.1806	$C_{23}H_{24}N_3O_3$	-0.6	$C_9H_{16}N_6O_4$	*m/z* 461–[Ala]
383.1401				
364.1650	$C_{21}H_{22}N_3O_3$	-0.6	$C_{11}H_{18}N_6O_4$	*m/z* 435–[Ala]
346.1545	$C_{21}H_{20}N_3O_2$	-0.5	$C_{11}H_{20}N_6O_5$	*m/z* 417–[Ala]
336.1701	$C_{20}H_{22}N_3O_2$	-0.6	$C_{12}H_{18}N_6O_5$	*m/z* 407–[Ala]

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
327.0776				
309.1553	$C_{14}H_{21}N_4O_4$	-0.4	$C_{18}H_{19}N_5O_3$	
251.0464				
239.0812	$C_{14}H_{11}N_2O_2$	-0.3	$C_{18}H_{29}N_7O_5$	
238.1183	$C_{11}H_{16}N_3O_3$	-0.3	$C_{21}H_{24}N_6O_4$	*m/z* 309–[Ala]
234.1234	$C_{12}H_{16}N_3O_2$	-0.3	$C_{20}H_{24}N_6O_5$	
221.0706	$C_{14}H_9N_2O$	-0.3	$C_{18}H_{31}N_7O_6$	*m/z* 417–[Ala]–$C_7H_{11}NO$
218.1132	$C_8H_{16}N_3O_4$	-0.3	$C_{24}H_{24}N_6O_3$	y_3
200.1027	$C_8H_{14}N_3O_3$	-0.3	$C_{24}H_{26}N_6O_4$	*m/z* 218–H_2O

Table 4.12: Accurate Mass Measurements of CID-MS^2 of (1,4-*s*-AAAG + H)$^+$ Ions

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
662.3037	$C_{32}H_{40}N_9O_7$	-0.8		
587.2710	$C_{30}H_{35}N_8O_5$	-1.5	$C_2H_5NO_2$	b_4
516.2345	$C_{27}H_{30}N_7O_4$	-0.9	$C_5H_{10}N_2O_3$	b_3

Table 4.13: Accurate Mass Measurements of CID-MS^3 of Ion (1,4-*s*-AAAG + H – N_2)$^+$ Generated by UVPD of (1,4-*s*-AAAG + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
m/z 634.2968	$C_{32}H_{40}N_7O_7$	-1.6		
m/z 616.2863	$C_{32}H_{38}N_7O_6$	-1.5	H_2O	
m/z 606.3019	$C_{31}H_{40}N_7O_6$	-1.6	CO	
m/z 589.2754	$C_{31}H_{37}N_6O_6$	-1.5	CH_3NO	–NH_3–CO
m/z 577.2755	$C_{30}H_{37}N_6O_6$	-1.4	C_2H_3NO	–[Gly] (β_4)

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
m/z 559.2649	$C_{30}H_{35}N_6O_5$	-1.4	$C_2H_5NO_2$	b_4
m/z 506.2385	$C_{27}H_{32}N_5O_5$	-1.3	$C_5H_8N_2O_2$	–[AlaGly] (β_3)
m/z 492.2229	$C_{26}H_{30}N_5O_5$	-1.2	$C_6H_{10}N_2O_2$	–2[Ala]
m/z 488.2280	$C_{27}H_{30}N_5O_4$	-1.2	$C_5H_{10}N_2O_3$	b_3
m/z 461.2172	$C_{26}H_{29}N_4O_4$	-1.1	$C_6H_{11}N_3O_3$	*m/z* 506–NH_3–CO
m/z 435.2016	$C_{24}H_{27}N_4O_4$	-1.1	$C_8H_{13}N_3O_3$	–[AlaAlaGly] (β_2)
m/z 381.1913	$C_{21}H_{25}N_4O_3$	-0.8	$C_{11}H_{15}N_3O_4$	
m/z 364.1647	$C_{21}H_{22}N_3O_3$	-0.9	$C_{11}H_{18}N_4O_4$	*m/z* 381–NH_3
m/z 336.1698	$C_{20}H_{22}N_3O_2$	-0.9	$C_{12}H_{18}N_4O_5$	*m/z* 461–$C_6H_7NO_2$
m/z 254.1130	$C_{11}H_{16}N_3O_4$	-0.5	$C_{21}H_{24}N_4O_3$	
m/z 228.0974	$C_9H_{14}N_3O_4$	-0.5	$C_{23}H_{26}N_4O_3$	
m/z 200.1025	$C_8H_{14}N_3O_3$	-0.5	$C_{24}H_{26}N_4O_4$	*m/z* 228–CO

Table 4.14: Accurate Mass Measurements of CID-MS^2 of (1,2-*s*-AAAHG + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
799.3619	$C_{38}H_{47}N_{12}O_8$	-1.5		
771.3550	$C_{38}H_{47}N_{10}O_8$	-2.3	N_2	
753.3455	$C_{38}H_{45}N_{10}O_7$	-1.2	H_2N_2O	–N_2–H_2O
743.3616	$C_{37}H_{47}N_{10}O_7$	-0.8	CN_2O	–N_2–CO
724.3313	$C_{36}H_{42}N_{11}O_6$	-0.1	$C_2H_5NO_2$	b_5
714.3352	$C_{36}H_{44}N_9O_7$	-0.6	$C_2H_3N_3O$	–N_2–[Gly]
696.3227	$C_{36}H_{42}N_9O_6$	-2.6	$C_2H_5N_3O_2$	–N_2–GlyOH
666.3113	$C_{32}H_{42}N_8O_8$	-0.7	$C_6H_5N_4$	
605.2824	$C_{30}H_{37}N_8O_6$	-0.7	$C_8H_{10}N_4O_2$	696–C_6H_5N

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
587.2719	$C_{30}H_{35}N_8O_5$	-0.6	$C_8H_{12}N_4O_3$	b_4
577.2759	$C_{30}H_{37}N_6O_6$	-1.0	$C_8H_{10}N_6O_2$	–N_2–[HisGly]
516.2348	$C_{27}H_{30}N_7O_4$	-0.6	$C_{11}H_{17}N_5O_4$	b_3
506.2391	$C_{27}H_{32}N_5O_5$	-0.7	$C_{11}H_{15}N_7O_3$	–N_2–[AlaHisGly]
445.1978	$C_{24}H_{25}N_6O_3$	-0.5	$C_{14}H_{22}N_6O_5$	b_2
435.2022	$C_{24}H_{27}N_4O_4$	-0.5	$C_{14}H_{20}N_8O_4$	–N_2–[AlaAlaHisGly]
426.2091	$C_{17}H_{28}N_7O_6$	-0.5	$C_{21}H_{19}N_5O_2$	y_5
408.1986	$C_{17}H_{26}N_7O_5$	-0.4	$C_{21}H_{21}N_5O_3$	426–H_2O
374.1608	$C_{21}H_{20}N_5O_2$	-0.4	$C_{17}H_{27}N_7O_6$	b_1
364.1652	$C_{21}H_{22}N_3O_3$	-0.4	$C_{17}H_{25}N_9O_5$	435–[Ala]
355.1720	$C_{14}H_{23}N_6O_5$	-0.4	$C_{24}H_{24}N_6O_3$	y_4
337.1616	$C_{14}H_{21}N_6O_4$	-0.3	$C_{24}H_{26}N_6O_4$	355–H_2O
284.1351	$C_{11}H_{18}N_5O_4$	-0.2	$C_{27}H_{29}N_7O_4$	y_3
266.1246	$C_{11}H_{16}N_5O_3$	-0.2	$C_{27}H_{31}N_7O_5$	284–H_2O

Table 4.15: Accurate Mass Measurements of UVPD-MS^2 of (1,2-*s*-AAAHG + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
799.3620	$C_{38}H_{47}N_{12}O_8$	-1.4		
771.3559	$C_{38}H_{47}N_{10}O_8$	-1.4	N_2	
753.3454	$C_{38}H_{45}N_{10}O_7$	-1.3	H_2N_2O	–N_2–H_2O
743.3613	$C_{37}H_{47}N_{10}O_7$	-1.1	CN_2O	–N_2–CO
714.3347	$C_{36}H_{44}N_9O_7$	-1.1	$C_2H_3N_3O$	–N_2–[Gly]
696.3237	$C_{36}H_{42}N_9O_6$	-1.6	$C_2H_5N_3O_2$	–N_2–GlyOH

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
680.3139	$C_{32}H_{42}N_9O_8$	-1.2	$C_6H_5N_3$	–N_2–C_6H_5N
666.3112	$C_{32}H_{42}N_8O_8$	-0.8	$C_6H_5N_4$	
605.2821	$C_{30}H_{37}N_8O_6$	-1.0	$C_8H_{10}N_4O_2$	696–C_6H_5N
577.2757	$C_{30}H_{37}N_6O_6$	-1.2	$C_8H_{10}N_6O_2$	–N_2–[HisGly]
518.2391	$C_{28}H_{32}N_5O_5$	-0.7	$C_{10}H_{15}N_7O_3$	
506.2390	$C_{27}H_{32}N_5O_5$	-0.8	$C_{11}H_{15}N_7O_3$	–N_2–[AlaHisGly]
468.2234	$C_{24}H_{30}N_5O_5$	-0.7	$C_{14}H_{17}N_7O_3$	
446.2139	$C_{20}H_{28}N_7O_5$	-0.7	$C_{18}H_{19}N_5O_3$	
435.2020	$C_{24}H_{27}N_4O_4$	-0.7	$C_{14}H_{20}N_8O_4$	–N_2–[AlaAlaHisGly]
426.2089	$C_{17}H_{28}N_7O_6$	-0.7	$C_{21}H_{19}N_5O_2$	y_5
408.1984	$C_{17}H_{26}N_7O_5$	-0.6	$C_{21}H_{21}N_5O_3$	*m/z* 426–H_2O
397.1865	$C_{21}H_{25}N_4O_4$	-0.5	$C_{17}H_{22}N_8O_4$	*m/z* 468–[Ala]
375.1770	$C_{17}H_{23}N_6O_4$	-0.5	$C_{21}H_{24}N_6O_4$	*m/z* 446–[Ala]
364.1651	$C_{21}H_{22}N_3O_3$	-0.5	$C_{17}H_{25}N_9O_5$	–N_2–[AlaAlaAlaHisGly]
355.1720	$C_{14}H_{23}N_6O_5$	-0.4	$C_{24}H_{24}N_6O_3$	y_4
337.1614	$C_{14}H_{21}N_6O_4$	-0.5	$C_{24}H_{26}N_6O_4$	y_4–H_2O
326.1494	$C_{18}H_{20}N_3O_3$	-0.5	$C_{20}H_{27}N_9O_5$	*m/z* 468–[AlaAla]
304.1400	$C_{14}H_{18}N_5O_3$	-0.4	$C_{24}H_{29}N_7O_5$	*m/z* 446–[AlaAla]
294.1193	$C_{12}H_{16}N_5O_4$	-0.4	$C_{26}H_{31}N_7O_4$	
284.1350	$C_{11}H_{18}N_5O_4$	-0.3	$C_{27}H_{29}N_7O_4$	y_3
266.1244	$C_{11}H_{16}N_5O_3$	-0.4	$C_{27}H_{31}N_7O_5$	*m/z* 284–H_2O
255.1125	$C_{15}H_{15}N_2O_2$	-0.3	$C_{23}H_{32}N_{10}O_6$	*m/z* 468–[AlaAlaAla]
249.0979	$C_{11}H_{13}N_4O_3$	-0.3	$C_{27}H_{34}N_8O_5$	*m/z* 266–NH_3
223.0823	$C_9H_{11}N_4O_3$	-0.3	$C_{29}H_{36}N_8O_5$	*m/z* 294–[Ala]

Table 4.16: Accurate Mass Measurements of CID-MS^3 of ion (1,2-*s*-AAAHG + H – N$_2$)$^+$ Generated by CID of (1,2-*s*-AAAHG + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
771.3551	$C_{38}H_{47}N_{10}O_8$	-2.2		
753.3443	$C_{38}H_{45}N_{10}O_7$	-2.4	H_2O	
743.3606	$C_{37}H_{47}N_{10}O_7$	-1.8	CO	
726.3343	$C_{37}H_{44}N_9O_7$	-1.5	CH_3NO	–CO–NH_3
714.3340	$C_{36}H_{44}N_9O_7$	-1.8	C_2H_3NO	–[Gly] (β_5)
696.3239	$C_{36}H_{42}N_9O_6$	-1.4	$C_2H_5NO_2$	b_5
686.3394	$C_{35}H_{44}N_9O_6$	-1.5	$C_3H_3NO_2$	*m/z* 714–CO
666.3108	$C_{32}H_{42}N_8O_8$	-1.2	$C_6H_5N_2$	
629.2820	$C_{32}H_{37}N_8O_6$	-1.1	$C_6H_{10}N_2O_2$	–[AlaAla]
605.2817	$C_{30}H_{37}N_8O_6$	-1.4	$C_8H_{10}N_2O_2$	*m/z* 696–C_6H_5N
577.2754	$C_{30}H_{37}N_6O_6$	-1.5	$C_8H_{10}N_4O_2$	*m/z* 714–[His]
534.2659	$C_{24}H_{36}N_7O_7$	-1.2	$C_{14}H_{11}N_3O$	
506.2385	$C_{27}H_{32}N_5O_5$	-1.3	$C_{11}H_{15}N_5O_3$	*m/z* 714–[AlaHis]
435.2016	$C_{24}H_{27}N_4O_4$	-1.1	$C_{14}H_{20}N_6O_4$	*m/z* 714–[AlaAlaHis]
408.1981	$C_{17}H_{26}N_7O_5$	-0.9	$C_{21}H_{21}N_3O_3$	y_5–H_2O
364.1648	$C_{21}H_{22}N_3O_3$	-0.8	$C_{17}H_{25}N_7O_5$	*m/z* 714–[AlaAlaAlaHis]
337.1611	$C_{14}H_{21}N_6O_4$	-0.8	$C_{24}H_{26}N_4O_4$	y_4–H_2O
294.1191	$C_{12}H_{16}N_5O_4$	-0.6	$C_{26}H_{31}N_5O_4$	
266.1242	$C_{11}H_{16}N_5O_3$	-0.6	$C_{27}H_{31}N_5O_5$	y_4–H_2O
248.1137	$C_{11}H_{14}N_5O_2$	-0.5	$C_{27}H_{33}N_5O_6$	*m/z* 266–H_2O

Table 4.17: Accurate Mass Measurements of CID-MS^3 of ion (1,2-*s*-AAAHG + H – N_2)$^+$ Generated by UVPD of (1,2-*s*-AAAHG + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
771.3551	$C_{38}H_{47}N_{10}O_8$	-2.2		
753.3444	$C_{38}H_{45}N_{10}O_7$	-2.3	H_2O	
743.3604	$C_{37}H_{47}N_{10}O_7$	-2.0	CO	
724.3184	$C_{37}H_{42}N_9O_7$	-1.8	CH_5NO	
714.3340	$C_{36}H_{44}N_9O_7$	-1.8	C_2H_3NO	–[Gly] (β_5)
696.3236	$C_{36}H_{42}N_9O_6$	-1.7	$C_2H_5NO_2$	b_5
686.3393	$C_{35}H_{44}N_9O_6$	-1.6	$C_3H_3NO_2$	*m/z* 714–CO
666.3105	$C_{32}H_{42}N_8O_8$	-1.5	$C_6H_5N_2$	
605.2817	$C_{30}H_{37}N_8O_6$	-1.4	$C_8H_{10}N_2O_2$	*m/z* 696–C_6H_5N
577.2754	$C_{30}H_{37}N_6O_6$	-1.5	$C_8H_{10}N_4O_2$	–[HisGly] (β_4)
534.2658	$C_{24}H_{36}N_7O_7$	-1.3	$C_{14}H_{11}N_3O$	
506.2385	$C_{27}H_{32}N_5O_5$	-1.3	$C_{11}H_{15}N_5O_3$	–[AlaHisGly] (β_3)
435.2016	$C_{24}H_{27}N_4O_4$	-1.1	$C_{14}H_{20}N_6O_4$	–[AlaAlaHisGly] (β_2)
408.1981	$C_{17}H_{26}N_7O_5$	-0.9	$C_{21}H_{21}N_3O_3$	y_5–H_2O
364.1647	$C_{21}H_{22}N_3O_3$	-0.9	$C_{17}H_{25}N_7O_5$	–[AlaAlaAlaHisGly] (β_1)
337.1611	$C_{14}H_{21}N_6O_4$	-0.8	$C_{24}H_{26}N_4O_4$	y_4–H_2O
294.1191	$C_{12}H_{16}N_5O_4$	-0.6	$C_{26}H_{31}N_5O_4$	
266.1242	$C_{11}H_{16}N_5O_3$	-0.6	$C_{27}H_{31}N_5O_5$	y_3–H_2O
249.0977	$C_{11}H_{13}N_4O_3$	-0.5	$C_{27}H_{34}N_6O_5$	*m/z* 266–NH_3

Table 4.18: Accurate Mass Measurements of UVPD-MS^3 of ion (1,2-*s*-AAAHG + H – N$_2$)$^+$ Generated by CID of (1,2-*s*-AAAHG + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
771.3555	$C_{38}H_{47}N_{10}O_8$	-1.8		
753.3450	$C_{38}H_{45}N_{10}O_7$	-1.7	H_2O	
743.3610	$C_{37}H_{47}N_{10}O_7$	-1.4	CO	
714.3343	$C_{36}H_{44}N_9O_7$	-1.5	C_2H_3NO	–[Gly] (β_5)
696.3240	$C_{36}H_{42}N_9O_6$	-1.3	$C_2H_5NO_2$	b_5
686.3396	$C_{35}H_{44}N_9O_6$	-1.3	$C_3H_3NO_2$	*m/z* 714–CO
682.3084	$C_{35}H_{40}N_9O_6$	-1.2	$C_3H_7NO_2$	–[Ala]–H_2O
668.3292	$C_{35}H_{42}N_9O_5$	-1.1	$C_3H_5NO_3$	*m/z* 696–CO
666.3107	$C_{32}H_{42}N_8O_8$	-1.3	$C_6H_5N_2$	
636.3002	$C_{31}H_{40}N_8O_7$	-1.2	$C_7H_7N_2O$	
605.2820	$C_{30}H_{37}N_8O_6$	-1.1	$C_8H_{10}N_2O_2$	*m/z* 696–C_6H_5N
577.2756	$C_{30}H_{37}N_6O_6$	-1.3	$C_8H_{10}N_4O_2$	–[HisGly] (β_4)
518.2388	$C_{28}H_{32}N_5O_5$	-1.0	$C_{10}H_{15}N_5O_3$	
506.2387	$C_{27}H_{32}N_5O_5$	-1.1	$C_{11}H_{15}N_5O_3$	–[AlaHisGly] (β_3)
488.2282	$C_{27}H_{30}N_5O_4$	-1.0	$C_{11}H_{17}N_5O_4$	b_3
468.2231	$C_{24}H_{30}N_5O_5$	-1.0	$C_{14}H_{17}N_5O_3$	*m/z* 605–[His]
447.2019	$C_{25}H_{27}N_4O_4$	-0.8	$C_{13}H_{20}N_6O_4$	*m/z* 518–[Ala]
435.2017	$C_{24}H_{27}N_4O_4$	-1.0	$C_{14}H_{20}N_6O_4$	–[AlaAlaHisGly] (β_2)
408.1982	$C_{17}H_{26}N_7O_5$	-0.8	$C_{21}H_{21}N_3O_3$	y_5–H_2O
397.1863	$C_{21}H_{25}N_4O_4$	-0.7	$C_{17}H_{22}N_6O_4$	*m/z* 605–[AlaHis]
376.1649	$C_{22}H_{22}N_3O_3$	-0.7	$C_{16}H_{25}N_7O_5$	*m/z* 518–[AlaAla]
364.1649	$C_{21}H_{22}N_3O_3$	-0.7	$C_{17}H_{25}N_7O_5$	–[AlaAlaAlaHisGly] (β_1)
337.1612	$C_{14}H_{21}N_6O_4$	-0.7	$C_{24}H_{26}N_4O_4$	y_3–H_2O
326.1494	$C_{18}H_{20}N_3O_3$	-0.5	$C_{20}H_{27}N_7O_5$	*m/z* 605–[AlaAlaHis]

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
294.1192	$C_{12}H_{16}N_5O_4$	-0.5	$C_{26}H_{31}N_5O_4$	
266.1243	$C_{11}H_{16}N_5O_3$	-0.5	$C_{27}H_{31}N_5O_5$	y_3–H_2O
255.1124	$C_{15}H_{15}N_2O_2$	-0.4	$C_{23}H_{32}N_8O_6$	*m/z* 605–[AlaAlaAlaHis]
249.0978	$C_{11}H_{13}N_4O_3$	-0.4	$C_{27}H_{34}N_6O_5$	*m/z* 266–NH_3
223.0822	$C_9H_{11}N_4O_3$	-0.4	$C_{29}H_{36}N_6O_5$	*m/z* 294–[Ala]

Table 4.19: Accurate Mass Measurements of CID-MS^2 of (1,2-*s*-AAAHG-OCH_3 + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
813.3831	$C_{39}H_{49}N_{12}O_8$	4.0		
785.3761	$C_{39}H_{49}N_{10}O_8$	3.2	N_2	
767.3674	$C_{39}H_{47}N_{10}O_7$	5.0	H_2N_2O	–N_2–H_2O
724.3354	$C_{36}H_{42}N_{11}O_6$	4.0	$C_3H_7NO_2$	b_5
696.3291	$C_{36}H_{42}N_9O_6$	3.8	$C_3H_7N_3O_2$	b_5–N_2
680.3314	$C_{33}H_{44}N_8O_8$	3.7	$C_6H_5N_4$	
668.3347	$C_{35}H_{42}N_9O_5$	4.4	$C_4H_7N_3O_3$	*m/z* 696–CO
652.3391	$C_{35}H_{42}N_9O_4$	3.7	$C_4H_7N_3O_4$	*m/z* 696–CO_2
587.2767	$C_{30}H_{35}N_8O_5$	4.2	$C_9H_{14}N_4O_3$	b_4
559.2704	$C_{30}H_{35}N_6O_5$	4.1	$C_9H_{14}N_6O_3$	b_4–N_2
516.2394	$C_{27}H_{30}N_7O_4$	4.0	$C_{12}H_{19}N_5O_4$	b_3
488.2337	$C_{27}H_{30}N_5O_4$	4.5	$C_{12}H_{19}N_7O_4$	b_3–N_2
440.2291	$C_{18}H_{30}N_7O_6$	3.9	$C_{21}H_{19}N_5O_2$	y_5
417.1958	$C_{24}H_{25}N_4O_3$	3.7	$C_{15}H_{24}N_8O_5$	b_2–N_2
369.1912	$C_{15}H_{25}N_6O_5$	3.1	$C_{24}H_{24}N_6O_3$	y_4

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
346.1581	$C_{21}H_{20}N_3O_2$	3.1	$C_{18}H_{29}N_9O_6$	b_1–N_2
318.1589	$C_{15}H_{20}N_5O_3$	2.8	$C_{24}H_{29}N_7O_5$	
298.1537	$C_{12}H_{20}N_5O_4$	2.7	$C_{27}H_{29}N_7O_4$	y_3

Table 4.20: Accurate Mass Measurements of UVPD-MS^2 of (1,2-*s*-AAAHG-OCH_3 + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
813.3825	$C_{39}H_{49}N_{12}O_8$	3.4		
785.3763	$C_{39}H_{49}N_{10}O_8$	3.4	N_2	
696.3284	$C_{36}H_{42}N_9O_6$	3.1	$C_3H_7N_3O_2$	b_5–N_2
694.3340	$C_{33}H_{44}N_9O_8$	3.3	$C_6H_5N_3$	–N_2–C_6H_5N
680.3309	$C_{33}H_{44}N_8O_8$	3.2	$C_6H_5N_4$	
668.3336	$C_{35}H_{42}N_9O_5$	3.3	$C_4H_7N_3O_3$	b_5–N_2–CO
652.3370	$C_{35}H_{42}N_9O_4$	1.6	$C_4H_7N_3O_4$	b_5–N_2–CO_2
605.2861	$C_{30}H_{37}N_8O_6$	3.0	$C_9H_{12}N_4O_2$	b_5–N_2–C_6H_5N
591.2828	$C_{30}H_{37}N_7O_6$	2.8	$C_9H_{12}N_5O_2$	
587.2762	$C_{30}H_{35}N_8O_5$	3.7	$C_9H_{14}N_4O_3$	*m/z* 605–H_2O
577.2910	$C_{29}H_{37}N_8O_5$	2.9	$C_{10}H_{12}N_4O_3$	*m/z* 605–CO
564.2952	$C_{29}H_{38}N_7O_5$	2.3	$C_{10}H_{11}N_5O_3$	
563.2869	$C_{29}H_{37}N_7O_5$	1.8	$C_{10}H_{12}N_5O_3$	
559.2690	$C_{30}H_{35}N_6O_5$	2.7	$C_9H_{14}N_6O_3$	b_4–N_2
531.2701	$C_{29}H_{35}N_6O_4$	-1.3	$C_{10}H_{14}N_6O_4$	b_4–N_2–CO
516.2377	$C_{27}H_{30}N_7O_4$	2.3	$C_{12}H_{19}N_5O_4$	*m/z* 587–[Ala]
488.2316	$C_{27}H_{30}N_5O_4$	2.4	$C_{12}H_{19}N_7O_4$	b_3–N_2

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
460.2325	$C_{26}H_{30}N_5O_3$	-1.8	$C_{13}H_{19}N_7O_5$	b_3–N_2–CO
417.1945	$C_{24}H_{25}N_4O_3$	2.4	$C_{15}H_{24}N_8O_5$	b_2–N_2
397.1893	$C_{21}H_{25}N_4O_4$	2.3	$C_{18}H_{24}N_8O_4$	b_3–N_2–C_6H_5N
389.1954	$C_{23}H_{25}N_4O_2$	-1.8	$C_{16}H_{24}N_8O_6$	b_2–N_2–CO
369.1902	$C_{15}H_{25}N_6O_5$	2.1	$C_{24}H_{24}N_6O_3$	y_4
346.1570	$C_{21}H_{20}N_3O_2$	2.0	$C_{18}H_{29}N_9O_6$	b_1–N_2
326.1519	$C_{18}H_{20}N_3O_3$	2.0	$C_{21}H_{29}N_9O_5$	*m/z* 397–[Ala]
318.1579	$C_{15}H_{20}N_5O_3$	1.8	$C_{24}H_{29}N_7O_5$	
298.1527	$C_{12}H_{20}N_5O_4$	1.7	$C_{27}H_{29}N_7O_4$	y_3
263.1155	$C_{12}H_{15}N_4O_3$	1.6	$C_{27}H_{34}N_8O_5$	y_3–NH_3–H_2O
255.1143	$C_{15}H_{15}N_2O_2$	1.5	$C_{24}H_{34}N_{10}O_6$	*m/z* 326–[Ala]
227.1151	$C_9H_{15}N_4O_3$	1.2	$C_{30}H_{34}N_8O_5$	y_2

Table 4.21: Accurate Mass Measurements of UVPD-MS^3 of ion (1,2-*s*-AAAHG-OCH_3 + H –N_2)$^+$ Generated by CID of (1,2-*s*-AAAHG-OCH_3 + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
785.3760	$C_{39}H_{49}N_{10}O_8$	3.1		
767.3656	$C_{39}H_{47}N_{10}O_7$	3.2	H_2O	
696.3290	$C_{36}H_{42}N_9O_6$	3.7	$C_3H_7NO_2$	b_5
680.3315	$C_{33}H_{44}N_8O_8$	3.8	$C_6H_5N_2$	
668.3347	$C_{35}H_{42}N_9O_5$	4.4	$C_4H_7NO_3$	b_5–CO
652.3391	$C_{35}H_{42}N_9O_4$	3.7	$C_4H_7NO_4$	b_5–CO_2
559.2699	$C_{30}H_{35}N_6O_5$	3.6	$C_9H_{14}N_4O_3$	b_4

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
531.2709	$C_{29}H_{35}N_6O_4$	-0.5	$C_{10}H_{14}N_4O_4$	b_4–CO
488.2324	$C_{27}H_{30}N_5O_4$	3.2	$C_{12}H_{19}N_5O_4$	b_3
460.2332	$C_{26}H_{30}N_5O_3$	-1.1	$C_{13}H_{19}N_5O_5$	b_3–CO
440.2280	$C_{18}H_{30}N_7O_6$	2.8	$C_{21}H_{19}N_3O_2$	y_5
417.1948	$C_{24}H_{25}N_4O_3$	2.7	$C_{15}H_{24}N_6O_5$	b_2
397.1896	$C_{21}H_{25}N_4O_4$	2.6	$C_{18}H_{24}N_6O_4$	b_3–C_6H_5N
389.1957	$C_{23}H_{25}N_4O_2$	-1.5	$C_{16}H_{24}N_6O_6$	b_2–CO
369.1904	$C_{15}H_{25}N_6O_5$	2.3	$C_{24}H_{24}N_4O_3$	y_4
346.1572	$C_{21}H_{20}N_3O_2$	2.2	$C_{18}H_{29}N_7O_6$	b_1
326.1521	$C_{18}H_{20}N_3O_3$	2.2	$C_{21}H_{29}N_7O_5$	b_2–C_6H_5N
318.1581	$C_{15}H_{20}N_5O_3$	2.0	$C_{24}H_{29}N_5O_5$	
298.1530	$C_{12}H_{20}N_5O_4$	2.0	$C_{27}H_{29}N_5O_4$	y_3

Table 4.22: Accurate Mass Measurements of CID-MS^3 of ion (1,2-*s*-AAAHG-OCH_3 + H –N_2)$^+$ Generated by UVPD of (1,2-*s*-AAAHG-OCH_3 + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
785.3744	$C_{39}H_{49}N_{10}O_8$	1.5		
767.3641	$C_{39}H_{47}N_{10}O_7$	1.7	H_2O	
696.3286	$C_{36}H_{42}N_9O_6$	3.3	$C_3H_7NO_2$	b_5
680.3311	$C_{33}H_{44}N_8O_8$	3.4	$C_6H_5N_2$	
668.3341	$C_{35}H_{42}N_9O_5$	3.8	$C_4H_7NO_3$	b_5–CO
559.2693	$C_{30}H_{35}N_6O_5$	3.0	$C_9H_{14}N_4O_3$	b_4
531.2704	$C_{29}H_{35}N_6O_4$	-1.0	$C_{10}H_{14}N_4O_4$	b_4–CO

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
488.2319	$C_{27}H_{30}N_5O_4$	2.7	$C_{12}H_{19}N_5O_4$	b_3
460.2328	$C_{26}H_{30}N_5O_3$	-1.5	$C_{13}H_{19}N_5O_5$	b_3–CO
440.2275	$C_{18}H_{30}N_7O_6$	2.3	$C_{21}H_{19}N_3O_2$	y_5
417.1944	$C_{24}H_{25}N_4O_3$	2.3	$C_{15}H_{24}N_6O_5$	b_2
397.1892	$C_{21}H_{25}N_4O_4$	2.2	$C_{18}H_{24}N_6O_4$	b_3–C_6H_5N
389.1953	$C_{23}H_{25}N_4O_2$	-1.9	$C_{16}H_{24}N_6O_6$	b_2–CO
369.1900	$C_{15}H_{25}N_6O_5$	1.9	$C_{24}H_{24}N_4O_3$	y_4
346.1569	$C_{21}H_{20}N_3O_2$	1.9	$C_{18}H_{29}N_7O_6$	b_1
326.1518	$C_{18}H_{20}N_3O_3$	1.9	$C_{21}H_{29}N_7O_5$	b_2–C_6H_5N
318.1578	$C_{15}H_{20}N_5O_3$	1.7	$C_{24}H_{29}N_5O_5$	
298.1527	$C_{12}H_{20}N_5O_4$	1.7	$C_{27}H_{29}N_5O_4$	y_3
280.1420	$C_{12}H_{18}N_5O_3$	1.6	$C_{27}H_{31}N_5O_5$	y_3–H_2O
263.1154	$C_{12}H_{15}N_4O_3$	1.5	$C_{27}H_{34}N_6O_5$	y_3–H_2O–NH_3
227.1152	$C_9H_{15}N_4O_3$	1.3	$C_{30}H_{34}N_6O_5$	y_2

Table 4.23: Accurate Mass Measurements of UVPD-MS^2 of (1,4-*s*-AAAHG + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
799.3599	$C_{38}H_{47}N_{12}O_8$	-3.5		
771.3538	$C_{38}H_{47}N_{10}O_8$	-3.5	N_2	
753.3436	$C_{38}H_{45}N_{10}O_7$	-3.1	H_2N_2O	–N_2–H_2O
743.3596	$C_{37}H_{47}N_{10}O_7$	-2.8	CN_2O	–N_2–CO
714.3330	$C_{36}H_{44}N_9O_7$	-2.8	$C_2H_3N_3O$	–N_2–[Gly]
696.3222	$C_{36}H_{42}N_9O_6$	-3.1	$C_2H_5N_3O_2$	–N_2–GlyOH
686.3382	$C_{35}H_{44}N_9O_6$	-2.7	$C_3H_3N_3O_2$	*m/z* 743–[Gly]

(Continued)

m/z	**Ion Formula**	**Error (mu)**	**Neutral Loss**	**Ion Assignment**
680.3126	$C_{32}H_{42}N_9O_8$	-2.5	$C_6H_5N_3$	–N_2–C_6H_5N
605.2805	$C_{30}H_{37}N_8O_6$	-2.6	$C_8H_{10}N_4O_2$	*m/z* 696–C_6H_5N
559.2643	$C_{30}H_{35}N_6O_5$	-2.0	$C_8H_{12}N_6O_3$	*m/z* 696–[His]
506.2378	$C_{27}H_{32}N_5O_5$	-2.0	$C_{11}H_{15}N_7O_3$	*m/z* 714–[AlaHis]
461.2167	$C_{26}H_{29}N_4O_4$	-1.6	$C_{12}H_{18}N_8O_4$	*m/z* 506–NH_3–CO
435.2010	$C_{24}H_{27}N_4O_4$	-1.7	$C_{14}H_{20}N_8O_4$	*m/z* 506–[Ala]
417.1906	$C_{24}H_{25}N_4O_3$	-1.5	$C_{14}H_{22}N_8O_5$	*m/z* 559–2[Ala]
397.1856	$C_{21}H_{25}N_4O_4$	-1.4	$C_{17}H_{22}N_8O_4$	
390.1800	$C_{23}H_{24}N_3O_3$	-1.2	$C_{15}H_{23}N_9O_5$	*m/z* 461–[Ala]
364.1643	$C_{21}H_{22}N_3O_3$	-1.3	$C_{17}H_{25}N_9O_5$	*m/z* 435–[Ala]
355.1712	$C_{14}H_{23}N_6O_5$	-1.2	$C_{24}H_{24}N_6O_3$	y_4
337.1606	$C_{14}H_{21}N_6O_4$	-1.3	$C_{24}H_{26}N_6O_4$	*m/z* 355–H_2O
304.1394	$C_{14}H_{18}N_5O_3$	-1.0	$C_{24}H_{29}N_7O_5$	
294.1187	$C_{12}H_{16}N_5O_4$	-1.0	$C_{26}H_{31}N_7O_4$	
284.1343	$C_{11}H_{18}N_5O_4$	-1.0	$C_{27}H_{29}N_7O_4$	y_3
266.1238	$C_{11}H_{16}N_5O_3$	-1.0	$C_{27}H_{31}N_7O_5$	*m/z* 284–H_2O
262.1290	$C_{12}H_{16}N_5O_2$	-0.9	$C_{26}H_{31}N_7O_6$	
252.1447	$C_{11}H_{18}N_5O_2$	-0.8	$C_{27}H_{29}N_7O_6$	
249.0973	$C_{11}H_{13}N_4O_3$	-0.9	$C_{27}H_{34}N_8O_5$	
223.0818	$C_9H_{11}N_4O_3$	-0.8	$C_{29}H_{36}N_8O_5$	
221.0702	$C_{14}H_9N_2O$	-0.7	$C_{24}H_{38}N_{10}O_7$	*m/z* 417–[Ala]–$C_7H_{11}NO$

Table 4.24: Accurate Mass Measurements of CID-MS^2 of (1,4-*s*-AAAHG + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
799.3608	$C_{38}H_{47}N_{12}O_8$	-2.6		
781.3509	$C_{38}H_{45}N_{12}O_7$	-2.0	H_2O	
771.3538	$C_{38}H_{47}N_{10}O_8$	-3.5	N_2	
753.3449	$C_{38}H_{45}N_{10}O_7$	-1.8	H_2N_2O	$–N_2–H_2O$
743.3611	$C_{37}H_{47}N_{10}O_7$	-1.3	CN_2O	$–N_2–CO$
724.3301	$C_{36}H_{42}N_{11}O_6$	-1.3	$C_2H_5NO_2$	b_5
714.3344	$C_{36}H_{44}N_9O_7$	-1.4	$C_2H_3N_3O$	$–N_2–$[Gly]
696.3229	$C_{36}H_{42}N_9O_6$	-2.4	$C_2H_5N_3O_2$	$–N_2–$GlyOH
666.3109	$C_{32}H_{42}N_8O_8$	-1.1	$C_6H_5N_4$	
605.2819	$C_{30}H_{37}N_8O_6$	-1.2	$C_8H_{10}N_4O_2$	*m/z* 696–C_6H_5N
559.2651	$C_{30}H_{35}N_6O_5$	-1.2	$C_8H_{12}N_6O_3$	*m/z* 696–[His]
506.2387	$C_{27}H_{32}N_5O_5$	-1.1	$C_{11}H_{15}N_7O_3$	*m/z* 714–[AlaHis]
488.2282	$C_{27}H_{30}N_5O_4$	-1.0	$C_{11}H_{17}N_7O_4$	*m/z* 696–[AlaHis]
417.1913	$C_{24}H_{25}N_4O_3$	-0.8	$C_{14}H_{22}N_8O_5$	*m/z* 696–[AlaAlaHis]
355.1717	$C_{14}H_{23}N_6O_5$	-0.7	$C_{24}H_{24}N_6O_3$	y_4
337.1612	$C_{14}H_{21}N_6O_4$	-0.7	$C_{24}H_{26}N_6O_4$	$y_4–H_2O$
284.1348	$C_{11}H_{18}N_5O_4$	-0.5	$C_{27}H_{29}N_7O_4$	y_3

Table 4.25: Accurate Mass Measurements of CID-MS^3 of Ion (1,4-*s*-AAAHG + H – N_2)$^+$ Generated by CID of (1,4-*s*-AAAHG + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
771.3541	$C_{38}H_{47}N_{10}O_8$	-3.2		
753.3431	$C_{38}H_{45}N_{10}O_7$	-3.6	H_2O	
743.3593	$C_{37}H_{47}N_{10}O_7$	-3.1	CO	
726.3332	$C_{37}H_{44}N_9O_7$	-2.6	CH_3NO	*m/z* 743–NH_3

(Continued)

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
725.3492	$C_{37}H_{45}N_{10}O_6$	-2.6	CH_2O_2	*m/z* 743–H_2O
714.3327	$C_{36}H_{44}N_9O_7$	-3.1	C_2H_3NO	–[Gly] (β_5)
696.3226	$C_{36}H_{42}N_9O_6$	-2.7	$C_2H_5NO_2$	b_5
686.3382	$C_{35}H_{44}N_9O_6$	-2.7	$C_3H_3NO_2$	–[Gly]–CO
668.3282	$C_{35}H_{42}N_9O_5$	-2.1	$C_3H_5NO_3$	b_5–CO
666.3096	$C_{32}H_{42}N_8O_8$	-2.4	$C_6H_5N_2$	
629.2810	$C_{32}H_{37}N_8O_6$	-2.1	$C_6H_{10}N_2O_2$	–[AlaAla]
605.2808	$C_{30}H_{37}N_8O_6$	-2.3	$C_8H_{10}N_2O_2$	*m/z* 696–C_6H_5N
577.2747	$C_{30}H_{37}N_6O_6$	-2.2	$C_8H_{10}N_4O_2$	–[HisGly] (β_4)
559.2643	$C_{30}H_{35}N_6O_5$	-2.0	$C_8H_{12}N_4O_3$	b_4
534.2651	$C_{24}H_{36}N_7O_7$	-2.0	$C_{14}H_{11}N_3O$	
506.2378	$C_{27}H_{32}N_5O_5$	-2.0	$C_{11}H_{15}N_5O_3$	–[AlaHisGly] (β_3)
488.2274	$C_{27}H_{30}N_5O_4$	-1.8	$C_{11}H_{17}N_5O_4$	b_3
461.2167	$C_{26}H_{29}N_4O_4$	-1.6	$C_{12}H_{18}N_6O_4$	β_3–NH_3–CO
435.2010	$C_{24}H_{27}N_4O_4$	-1.7	$C_{14}H_{20}N_6O_4$	–[AlaAlaHisGly] (β_2)
417.1907	$C_{24}H_{25}N_4O_3$	-1.4	$C_{14}H_{22}N_6O_5$	b_2
337.1606	$C_{14}H_{21}N_6O_4$	-1.3	$C_{24}H_{26}N_4O_4$	y_4–H_2O
294.1187	$C_{12}H_{16}N_5O_4$	-1.0	$C_{26}H_{31}N_5O_4$	
266.1239	$C_{11}H_{16}N_5O_3$	-0.9	$C_{27}H_{31}N_5O_5$	y_3–H_2O
249.0974	$C_{11}H_{13}N_4O_3$	-0.8	$C_{27}H_{34}N_6O_5$	y_3–H_2O–NH_3

Table 4.26: Accurate Mass Measurements of UVPD-MS^3 of Ion (1,4-*s*-AAAHG + H – N_2)$^+$ Generated by CID of (1,4-*s*-AAAHG + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
771.3533	$C_{38}H_{47}N_{10}O_8$	-4.0		

(Continued)

m/z	**Ion Formula**	**Error (mu)**	**Neutral Loss**	**Ion Assignment**
753.3430	$C_{38}H_{45}N_{10}O_7$	-3.7	H_2O	
743.3589	$C_{37}H_{47}N_{10}O_7$	-3.5	CO	
714.3323	$C_{36}H_{44}N_9O_7$	-3.5	C_2H_3NO	–[Gly] (β_5)
696.3218	$C_{36}H_{42}N_9O_6$	-3.5	$C_2H_5NO_2$	b_5
686.3376	$C_{35}H_{44}N_9O_6$	-3.3	$C_3H_3NO_2$	–[Gly]–CO
668.3272	$C_{35}H_{42}N_9O_5$	-3.1	$C_3H_5NO_3$	b_5–CO
636.2985	$C_{31}H_{40}N_8O_7$	-2.9	$C_7H_7N_2O$	
605.2800	$C_{30}H_{37}N_8O_6$	-3.1	$C_8H_{10}N_2O_2$	b_5–C_6H_5N
577.2857	$C_{29}H_{37}N_8O_5$	-2.4	$C_9H_{10}N_2O_3$	
559.2635	$C_{30}H_{35}N_6O_5$	-2.8	$C_8H_{12}N_4O_3$	b_4
534.2647	$C_{24}H_{36}N_7O_7$	-2.4	$C_{14}H_{11}N_3O$	
518.2373	$C_{28}H_{32}N_5O_5$	-2.5	$C_{10}H_{15}N_5O_3$	
506.2374	$C_{27}H_{32}N_5O_5$	-2.4	$C_{11}H_{15}N_5O_3$	–[AlaHisGly] (β_3)
488.2270	$C_{27}H_{30}N_5O_4$	-2.2	$C_{11}H_{17}N_5O_4$	b_3
461.2161	$C_{26}H_{29}N_4O_4$	-2.2	$C_{12}H_{18}N_6O_4$	–[AlaHisGly]–NH_3–CO
447.2006	$C_{25}H_{27}N_4O_4$	-2.1	$C_{13}H_{20}N_6O_4$	*m/z* 518–[Ala]
435.2007	$C_{24}H_{27}N_4O_4$	-2.0	$C_{14}H_{20}N_6O_4$	*m/z* 506–[Ala]
417.1902	$C_{24}H_{25}N_4O_3$	-1.9	$C_{14}H_{22}N_6O_5$	b_2
397.1852	$C_{21}H_{25}N_4O_4$	-1.8	$C_{17}H_{22}N_6O_4$	b_3–C_6H_5N
390.1795	$C_{23}H_{24}N_3O_3$	-1.7	$C_{15}H_{23}N_7O_5$	*m/z* 447–[Gly]
364.1640	$C_{21}H_{22}N_3O_3$	-1.6	$C_{17}H_{25}N_7O_5$	*m/z* 435–[Ala]
355.1708	$C_{14}H_{23}N_6O_5$	-1.6	$C_{24}H_{24}N_4O_3$	y_4
337.1603	$C_{14}H_{21}N_6O_4$	-1.6	$C_{24}H_{26}N_4O_4$	y_4–H_2O
304.1390	$C_{14}H_{18}N_5O_3$	-1.4	$C_{24}H_{29}N_5O_5$	
294.1184	$C_{12}H_{16}N_5O_4$	-1.3	$C_{26}H_{31}N_5O_4$	
266.1236	$C_{11}H_{16}N_5O_3$	-1.2	$C_{27}H_{31}N_5O_5$	y_3–H_2O

(Continued)

m/z	**Ion Formula**	**Error (mu)**	**Neutral Loss**	**Ion Assignment**
249.0971	$C_{11}H_{13}N_4O_3$	-1.1	$C_{27}H_{34}N_6O_5$	y_3–H_2O–NH_3
223.0816	$C_9H_{11}N_4O_3$	-1.0	$C_{29}H_{36}N_6O_5$	

Table 4.27: Accurate Mass Measurements of CID-MS^3 of Ion (1,4-*s*-AAAHG + H – N_2)$^+$ Generated by UVPD of (1,4-*s*-AAAHG + H)$^+$

m/z	**Ion Formula**	**Error (mu)**	**Neutral Loss**	**Ion Assignment**
771.3534	$C_{38}H_{47}N_{10}O_8$	-3.9		
753.3428	$C_{38}H_{45}N_{10}O_7$	-3.9	H_2O	
743.3590	$C_{37}H_{47}N_{10}O_7$	-3.4	CO	
726.3326	$C_{37}H_{44}N_9O_7$	-3.2	CH_3NO	–NH_3–CO
725.3488	$C_{37}H_{45}N_{10}O_6$	-3.0	CH_2O_2	
714.3323	$C_{36}H_{44}N_9O_7$	-3.5	C_2H_3NO	–[Gly] (β_5)
696.3223	$C_{36}H_{42}N_9O_6$	-3.0	$C_2H_5NO_2$	b_5
686.3378	$C_{35}H_{44}N_9O_6$	-3.1	$C_3H_3NO_2$	–[Gly]–CO
668.3275	$C_{35}H_{42}N_9O_5$	-2.8	$C_3H_5NO_3$	b_5–CO
605.2804	$C_{30}H_{37}N_8O_6$	-2.7	$C_8H_{10}N_2O_2$	b_5–C_6H_5N
577.2741	$C_{30}H_{37}N_6O_6$	-2.8	$C_8H_{10}N_4O_2$	–[HisGly] (β_4)
559.2640	$C_{30}H_{35}N_6O_5$	-2.3	$C_8H_{12}N_4O_3$	b_4
534.2648	$C_{24}H_{36}N_7O_7$	-2.3	$C_{14}H_{11}N_3O$	
506.2375	$C_{27}H_{32}N_5O_5$	-2.3	$C_{11}H_{15}N_5O_3$	–[AlaHisGly] (β_3)
461.2163	$C_{26}H_{29}N_4O_4$	-2.0	$C_{12}H_{18}N_6O_4$	β_3–NH_3–CO
435.2008	$C_{24}H_{27}N_4O_4$	-1.9	$C_{14}H_{20}N_6O_4$	–[AlaAlaHisGly] (β_2)
337.1605	$C_{14}H_{21}N_6O_4$	-1.4	$C_{24}H_{26}N_4O_4$	y_4–H_2O
294.1185	$C_{12}H_{16}N_5O_4$	-1.2	$C_{26}H_{31}N_5O_4$	
266.1237	$C_{11}H_{16}N_5O_3$	-1.1	$C_{27}H_{31}N_5O_5$	y_3–H_2O

(Continued)

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
249.0972	$C_{11}H_{13}N_4O_3$	-1.0	$C_{27}H_{34}N_6O_5$	y_3–H_2O–NH_3
223.0817	$C_9H_{11}N_4O_3$	-0.9	$C_{29}H_{36}N_6O_5$	

4.2.5 Calculations

Structures of the conjugates and their dissociation products were generated by Born-Oppenheimer molecular dynamics (BOMD) calculations that were run as 20 ps trajectories with 1 fs steps, using the Berendsen thermostat[28] at 510, 610, and 810 K. Trajectory analysis confirmed that the systems were reaching a near-stationary state after ca 16000 steps (16 ps), as determined by low root-mean-square deviations of atomic Cartesian coordinates, so no longer trajectory runs were necessary because they generated only duplicate structures. Initial protonation sites were chosen at different amide groups for the non-basic AAAG conjugates to be treated by BOMD. Since BOMD includes both electron and nuclear motion, the different protonation sites coalesced by proton transfer to the most basic position. The initial protonation sites in AAAHG conjugates were at the basic residue His. The trajectory calculations were performed with PM6-D3H4[29] which complements the semiempirical Hamiltonian with dispersion and hydrogen-bonding interactions, using the Cuby4 platform[30] and MOPAC.[31] Two hundred snapshots were selected at regular 100 fs intervals from the BOMD trajectories, and the structures were fully gradient-optimized with PM6-D3H4. Several selected low energy structures were further optimized with B3LYP[32] with the 6-31+G(d,p) basis set. These calculations were augmented by including empirical dispersion corrections (GD3-BJ)[33][34] and were used to provide harmonic frequencies. Several low-energy structures were reoptimized with M06-2X[35] using the 6-31+G(d,p) basis set. The M06-2X/6-31+G(d,p) optimized geometries were used to calculate single-point energies which were carried out with M06-2X and the def2qzvpp[36] basis set (<5000 basis functions). It should be noted that ion thermochemistry obtained at this level of theory is approximate within an estimated 5-10 kJ mol^{-1},[35][37] and thus no efforts were made to derive the ion isomer equilibrium populations from the calculated Gibbs free energies ($\Delta G_{g,310}$). Another set of M06-2X/6-31++G(d,p) single-point calculations were used to obtain charge densities according to Merz, Singh, and Kollman (MK).[38][39] All the standard DFT calculations were run using Gaussian 16 (Revision B.01) that was licensed from Gaussian, Inc. (Wallingford, CT). Collision cross sections in nitrogen were calculated by the modified ion trajectory method (MobCal$_{\text{MPI}}$)[40][41] using the MK charge densities. Standard van der Waals parameters for different atom types were obtained from the MMFF94 data set.[42]

4.3 Results and Discussion

4.3.1 Photodissociation and Crosslinking Yields of 1,2-*s*-AAAG and 1,4-*s*-AAAG

4.3.1.1 Spectra Interpretation of 1,2-*s*-AAAG

Photodissociation of (1,2-*s*-AAAG + H)$^+$ (*m/z* 662) led to backbone cleavages and subsequent loss of N_2, producing a nitrile imine ion (1,2-*s*-AAAG + H – N_2)$^+$ (*m/z* 634) as shown in the UVPD-MS^2 spectrum (Figure 4.1 (a)). The loss of N_2 and the resulting fragment ion assignments were determined through accurate mass measurements, where $\Delta m = 28.0060$ for N_2 (Tables 4.5-4.7). The dissociations upon UVPD showed strong dependence on the photon energy, which was 4.96 ev for the 250 nm and 5.82 eV for the 213 nm laser beams. This was reflected by the more extensive dissociations of the (1,2-*s*-AAAG + H – N_2)$^+$ ion when generated by UVPD at 213 nm (Figure 4.2 (a)). We note that the UV-vis absorption spectrum of 2-(4-carboxyphenyl)-5-phenyl-tetrazole in acetonitrile solution showed strong absorption at both wavelengths due to the 2,5-diaryltetrazole chromophore (Figure 4.3), while absorption at 250 nm in gas-phase ions has also been established by UVPD action spectra.[24]

UVPD produced three series of fragment ions whose relative intensities depended on the photon energy. The first is the peptide backbone cleavage without tetrazole N_2 loss, which produced $\boldsymbol{b}_n$ type ions at *m/z* 587, 516, 445, 374, and [$\boldsymbol{y}_n$ + 2H]$^+$ ions at *m/z* 289, and 271 that were particularly abundant at 213 nm (red-annotated ions in Figure 4.1 (a) and 4.2 (a)). We note that backbone cleavages leading to ***b***-type ions were predominant in the CID-MS^2 spectrum of (1,2-*s*-AAAG + H)$^+$ (Figure 4.4 (a), Table 4.6) that did not show loss of N_2. A second ion series originated from (1,2-*s*-AAAG + H – N_2)$^+$ (*m/z* 634) by loss of C_6H_5N, possibly phenylnitrene or an isomer, giving rise to fragment ions at *m/z* 543 and their sequence ***b***-type ions at *m/z* 468, 397, 326, and 255 (green-annotated ions in Figure 4.1 (a)). This ion series was prominent in the 213 nm UVPD spectrum (Figure 4.2 (a)) while being much less abundant at 250 nm

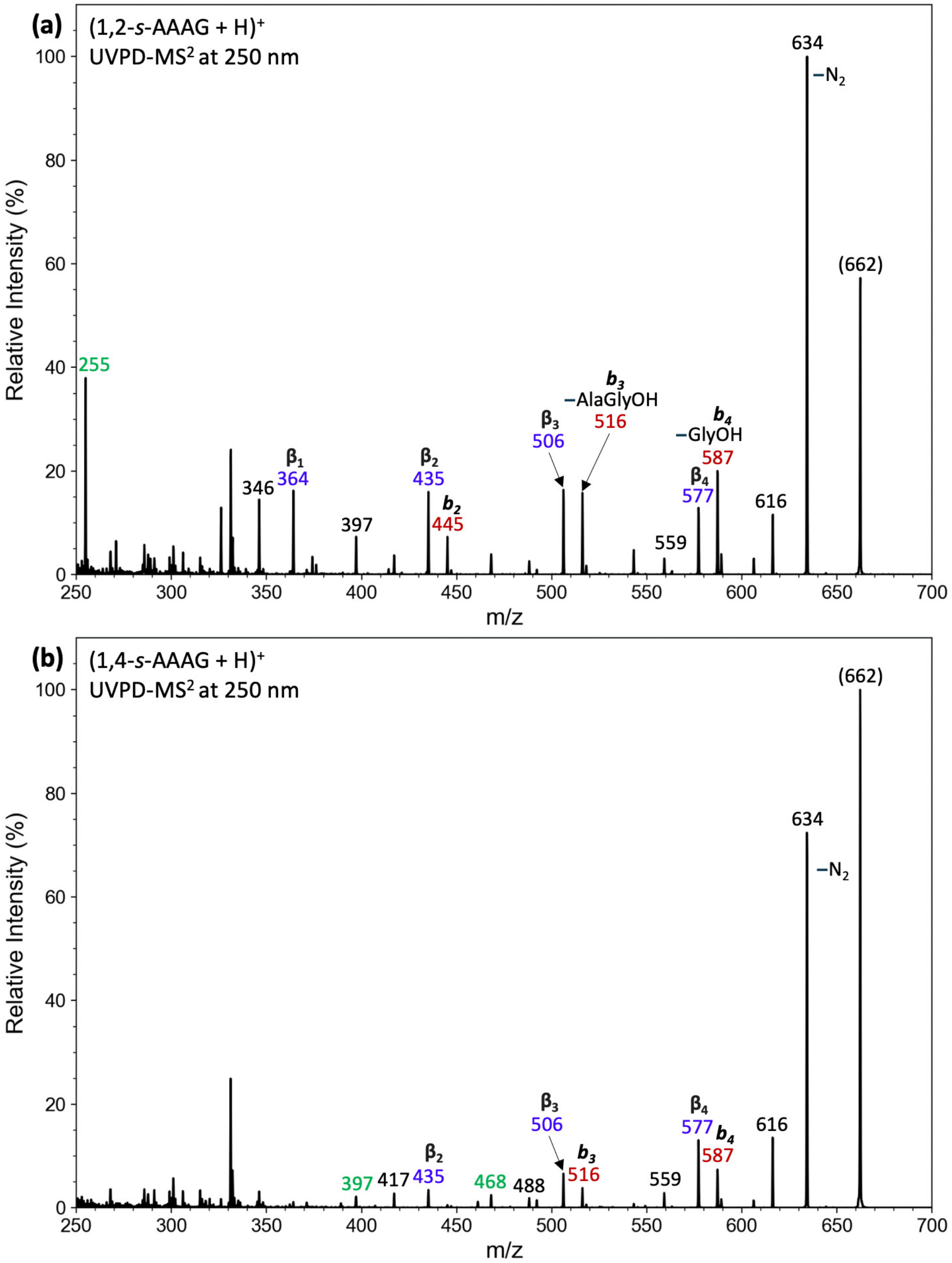


Figure 4.1: UVPD-MS^2 of (a) (1,2-*s*-AAAG + H)$^+$ (*m/z* 662), and (b) (1,4-*s*-AAAG + H)$^+$ (*m/z* 662) obtained at 250 nm. Fragment ion color-coding is as follows: Dark red: $\boldsymbol{b}_n$ and [$\boldsymbol{y}_n$ + 2H]$^+$ from the precursor ion; green: $\boldsymbol{b}_n$ after loss of N_2 and C_6H_5N; blue: β_n ions by loss of truncated residues from crosslinks after loss of N_2. The aminocyclohexane carboxylic acid is counted as the first residue in $\boldsymbol{b}_n$ and β_n ions.

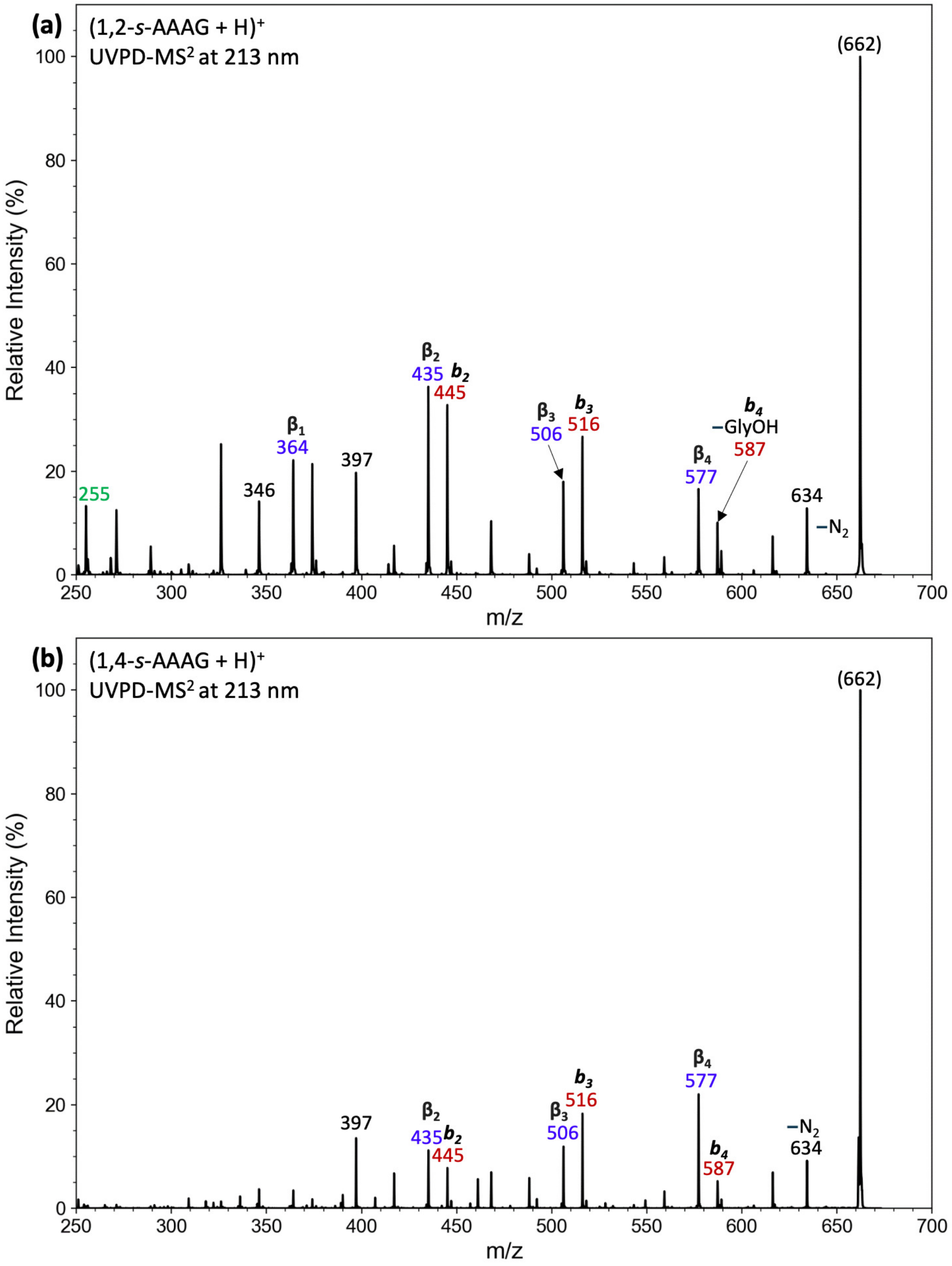


Figure 4.2: UVPD-MS^2 of (a) (1,2-*s*-AAAG + H)$^+$ (*m/z* 662), and (b) (1,4-*s*-AAAG + H)$^+$ (*m/z* 662) obtained at 213 nm. For ion color coding see text in Figure 4.1.

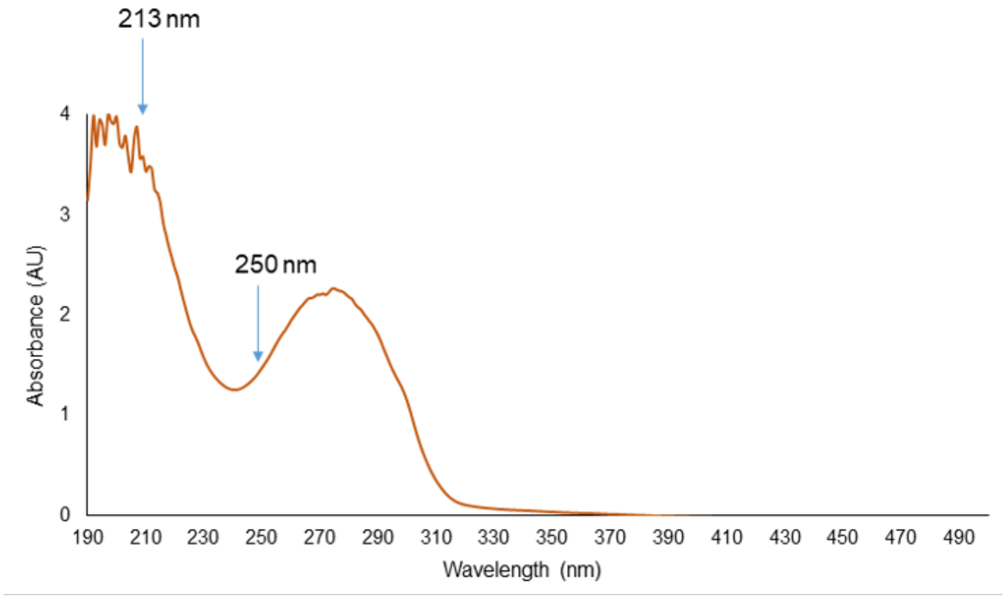


Figure 4.3: UV-vis absorption spectrum of 2-(4-carboxyphenyl)-5-phenyltetrazole in 99:1 acetonitrile:trifluoroacetic acid.

(Figure 4.1 (a)). The third ion series (blue-annotated ions in Figure 4.1 (a) and 4.2 (a)) also originated from (1,2-*s*-AAAG + H – N_2)$^+$ but involved internal residue losses forming *m/z* 577 (loss of [$HNCH_2CO$], denoted as [Gly]), *m/z* 506 (loss of [$HNCH(CH_3)CONHCH_2CO$], denoted as [AlaGly]), *m/z* 435 (loss of [AlaAlaGly]), and *m/z* 364 (loss of [AlaAlaAlaGly]). In this notation, we use square brackets to distinguish the neutral internal losses, [Amino acid residues], from those of standard amino acids and peptides leading to ***b*** and [$\boldsymbol{y}_n$ + 2H]$^+$ fragment ions.[43] Since losses of internal amino acid residues are characteristic of cyclic peptides,[44]-[46] their presence in the UVPD-MS^2 spectra allowed us to assign internal crosslinks between the nitrile imine and peptide groups. The UVPD and CID-MS^2 spectra differed significantly due to distinct ion excitation modes affecting dissociation pathways. Absorption of a 250-nm or 213-nm photon by the 2,5-diaryltetrazole chromophore created excited electronic states that dissociated by N_2 loss before internal energy redistribution, causing vibrational excitation. In contrast, slow heating by collisions drove ground-electronic state dissociations where the peptide bond cleavages prevailed. This was despite that previous energy analysis of tetrazole ring dissociations has indicated a low threshold energy for the N_2 loss.[47][22]

In the next step, we selected ions (1,2-*s*-AAAG + H – N_2)$^+$ (*m/z* 634) and investigated their CID-MS^3 spectra (Figure 4.5 (a) and 4.6 (a), Table 4.7), followed by CID-MS^4 of each fragmentation peak from the

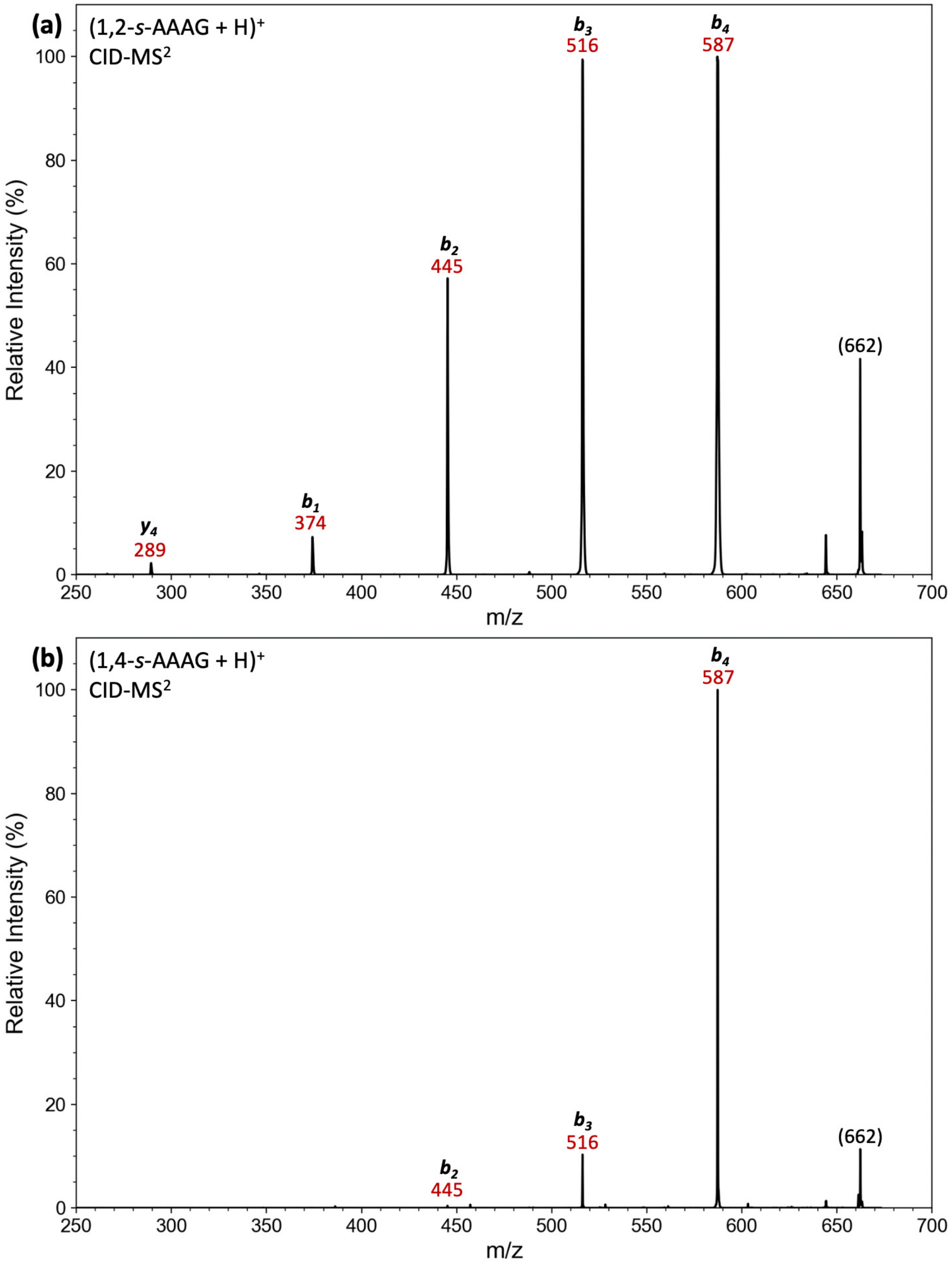


Figure 4.4: CID-MS^2 of (a) (1,2-*s*-AAAG + H)$^+$ (*m/z* 662), and (b) (1,4-*s*-AAAG + H)$^+$ (*m/z* 662). For ion color coding see text in Figure 4.1.

CID-MS^3 spectrum (Table 4.28). The main fragment ion series consisted of internal fragments formed by consecutive losses of [Gly] and [Ala] residues, represented as β_n ions to differentiate them from the standard $\boldsymbol{b}_n$ sequence ions. Among the latter, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$, and $\boldsymbol{b}_4$ were present in the spectrum at low intensities, indicating structures with an open peptide chain in a minor population of (1,2-*s*-AAAG + H – N$_2$)$^+$ ions. The main population, indicated by blue-annotated β_n ions, was attributed to crosslinked structures.

Table 4.28: CID-MS^4 Fragments of (1,2-*s*-AAAG + H – N$_2$)$^+$ *m/z* 634 from 1,2-*s*-AAAG Following UVPD at 213 nm

616	606	589	577	559	518	506	492	488	435	364	271
–H$_2$O	–CO	–CO–NH$_3$	β_4	**b**$_4$	589–[Gly]	β_3	–2[Ala]	**b**$_3$	β_2	β_1	β_0
598	589	518	506	531	447	435	474	470	390	256	243
588	577	501	435	488	376	364	435	460	364	239	228
559	563	447	364	397			421	443	256	200	226
545	535	376		364			364	417	180	109	200
528	518			326				379	152		
474	506			234				328			
403	447			179				309			
397	410							255			
364	384							205			
326								180			
273											
255											

4.3.1.2 Spectra Interpretation of 1,4-*s*-AAAG

The isomeric (1,4-*s*-AAAG + H)$^+$ ion was photolyzed at 250 and 213 nm with the results shown in Figure 4.1 (b) and 4.2 (b). In addition to the loss of N$_2$ from the tetrazole moiety (ion (1,4-*s*-AAAG + H – N$_2$)$^+$, *m/z* 634), UVPD-MS^2 displayed a series of internal β_n fragment ions resulting from sequential dissociations of (1,4-*s*-AAAG + H – N$_2$)$^+$ (blue-annotated ions at *m/z* 577, 506, 435, and 364), similar to those produced by (1,2-*s*-AAAG + H – N$_2$)$^+$. Elemental compositions of these ions are listed in Table 4.11. Red-annotated ions reflected peptide chain cleavages without nitrogen loss, while green-annotated ions originated from nitrile imine fragments. Similar to (1,2-*s*-AAAG + H)$^+$, CID-MS^2 of (1,4-*s*-AAAG + H)$^+$ gave a straight $\boldsymbol{b}_n$ ion series while no loss of N$_2$ was observed (Figure 4.4 (b), Table 4.12). CID-MS^3 of (1,4-*s*-AAAG + H – N$_2$)$^+$ showed a prominent loss of GlyOH (*m/z* 559), suggesting a subset of ions with a linear peptide chain (Figures 4.5 (b) and 4.6 (b)). Additional CID-MS^4 data for each peak in the CID-MS^3 spectrum are recorded in Table 4.29. Most of the CID-MS^3 products from (1,4-*s*-AAAG + H – N$_2$)$^+$ were

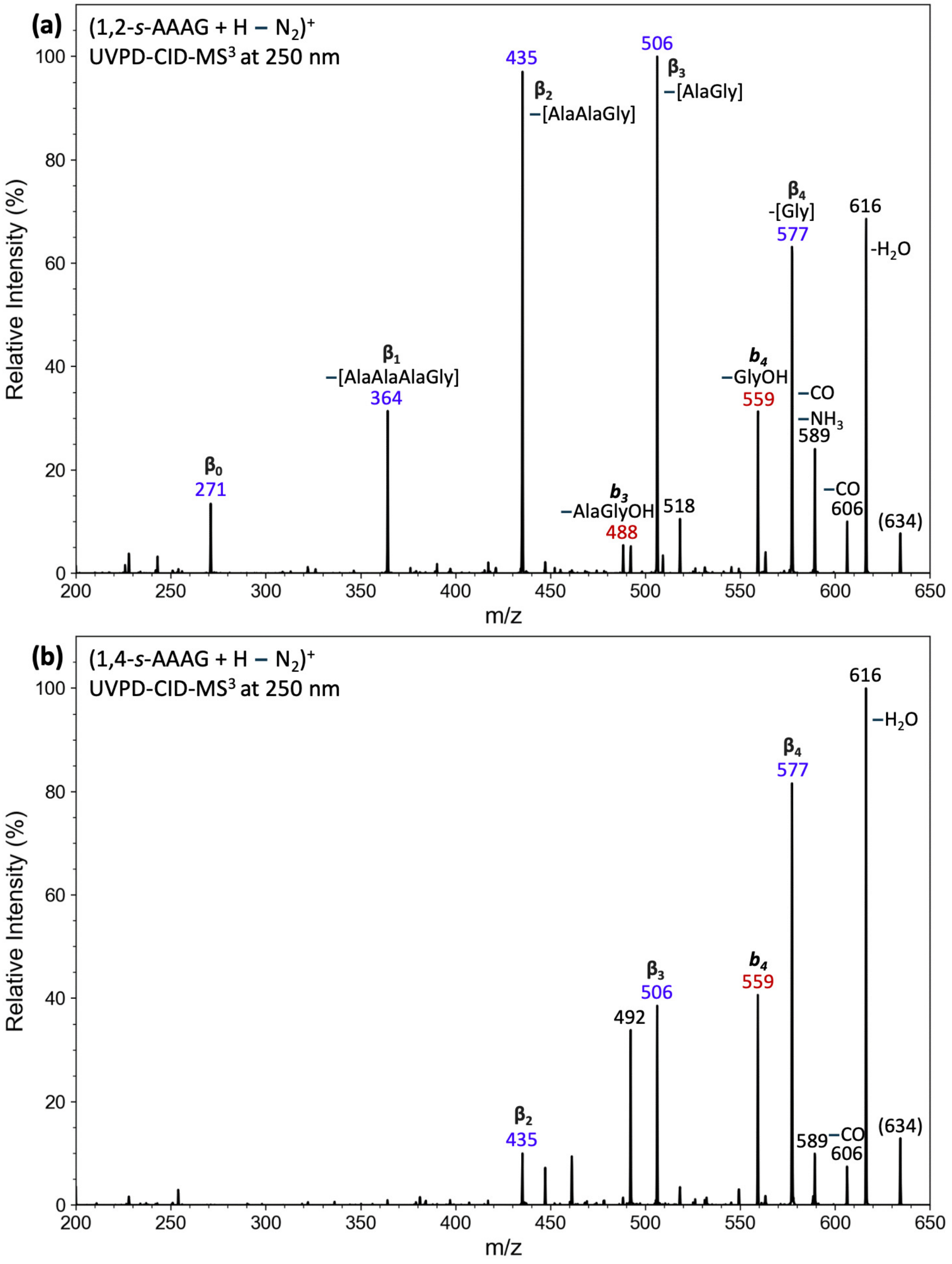


Figure 4.5: CID-MS^3 of (a) (1,2-*s*-AAAG + H – N$_2$)$^+$ (*m/z* 634), and (b) (1,4-*s*-AAAG + H – N$_2$)$^+$ (*m/z* 634) from 250 nm UVPD. For ion color coding see text in Figure 4.1.

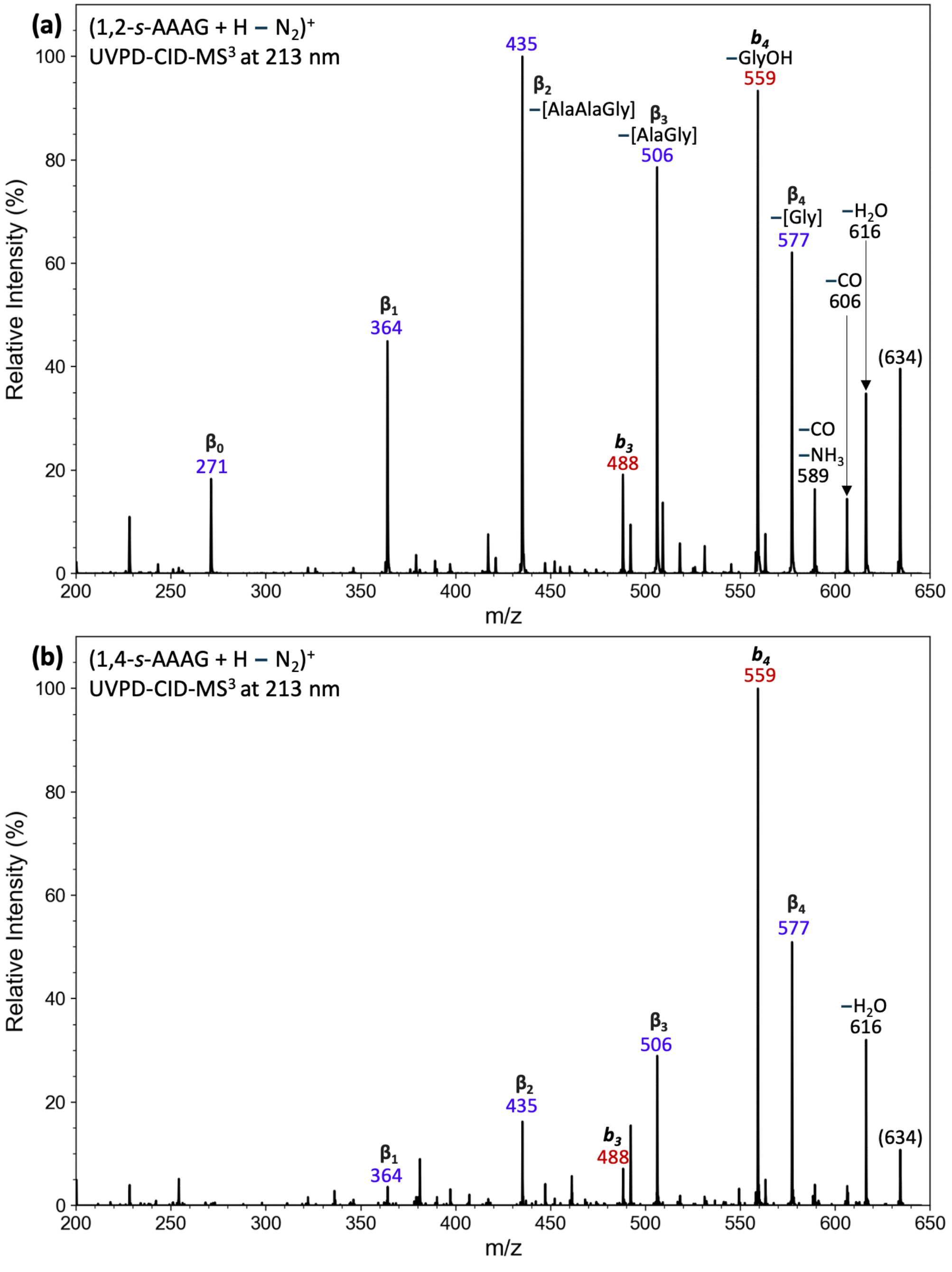


Figure 4.6: CID-MS^3 of (a) (1,2-*s*-AAAG + H – N$_2$)$^+$ (*m/z* 634), and (b) (1,4-*s*-AAAG + H – N$_2$)$^+$ (*m/z* 634) from 213 nm UVPD. For ion color coding see text in Figure 4.1.

β_n internal fragments by loss of [Gly] (*m/z* 577) in combination with [$(\mathrm{Ala})_n$] (*m/z* 506, 435). In addition, loss of [AlaAla] (*m/z* 492) was observed, and cleavage of the Ala_1 N–C_α bond formed the complementary fragment ions at *m/z* 381, 254, and 228. According to their accurate *m/z* (Table 4.13), these ions involved the formation of $\mathrm{HC(CH_3)CO}$-Ala-Ala-$\mathrm{NHCH_2CO}$ neutral or ion fragments that indicated cyclization to the nitrile imine by the Gly carboxyl group.

Table 4.29: CID-MS^4 Fragments of (1,4-*s*-AAAG + H – $\mathrm{N_2}$)$^+$ *m/z* 634 from 1,4-*s*-AAAG Following UVPD at 213 nm

616	606	589	577	559	506	492	461	435
–$\mathrm{H_2O}$	–CO	–CO–$\mathrm{NH_3}$	β_4	**b**$_4$	β_3	–2[Ala]	506–$\mathrm{NH_3}$–$\mathrm{H_2O}$	β_2
598	589	561	549	541	478	474	443	417
588	588	518	532	531	461	464	433	407
573	577		506	516	435	447	390	390
559	561		461	488		435	362	364
545	518			472		384	327	
502	506			407			265	
474	492			381			180	
371	344			364				
202	226			325				
				234				
				179				

4.3.1.3 Difference between β_n and ($\boldsymbol{b}_n$ + $\mathrm{H_2O}$) Ions

The β_n ions discussed above are formally analogous to ($\boldsymbol{b}_n$ + $\mathrm{H_2O}$) fragment ions, which Gaskell and colleagues reported as products of C-terminal carboxyl participation in amide bond cleavage, accompanied by OH group and proton migration.[48]-[50] The two types of ions differ in structure, with β_n ions having the carboxyl oxygen incorporated in the nitrile imine moiety as a result of crosslinking, whereas the ($\boldsymbol{b}_n$ + $\mathrm{H_2O}$) are C-terminally truncated peptide ions. Since the ions are isomeric, their distinction was of concern with regard to peptide-nitrile imine crosslinking. The formation of ($\boldsymbol{b}_n$ + $\mathrm{H_2O}$) fragment ions has been reported to be sequence dependent, and where present in the spectra, they accompanied as minor satellites the more common $\boldsymbol{b}_n$ ions.[49] Perusal of the CID-MS^2 spectra of (1,2-*s*-AAAG + H)$^+$ (Figure 4.4 (a)) and (1,4-*s*-AAAG + H)$^+$ (Figure 4.4 (b)) revealed no ($\boldsymbol{b}_n$ + $\mathrm{H_2O}$) ions accompanying the $\boldsymbol{b}_n$ ion series, indicating that this peptide sequence in the scaffolds did not undergo backbone rearrangements with OH migration.

Additionally, we examined the non-crosslinked ions by performing CID-MS^4 on the signature isomeric

$(MH – N_2 – GlyOH)^+$ ions ($\boldsymbol{b}_4$, *m/z* 559) from (1,2-*s*-AAAG + H)$^+$ and (1,4-*s*-AAAG + H)$^+$, as well as the (1,2-*s*-AAAG-OCH$_3$ + H – N$_2$)$^+$ ions from (1,2-*s*-AAAG-OCH$_3$ + H)$^+$, which was blocked at the C-terminus as a methyl ester and discussed in details later (Figure 4.7). The spectrum of the $\boldsymbol{b}_4$ ion from (1,4-*s*-AAAG + H)$^+$ (Figure 4.7 (c)) was straightforward, showing a dominant loss of CO that is typical of dissociations of ***b***-type ions. The spectrum of the $\boldsymbol{b}_4$ ion from (1,2-*s*-AAAG + H)$^+$ (Figure 4.7 (a)) also showed a loss of CO; however, the spectrum also showed a $\boldsymbol{b}_3$ sequence ion at *m/z* 488 that further eliminated a $C_6H_8N_2O$ neutral fragment. This dissociation clearly indicated a bonding interaction between the peptide chain and the nitrile imine that resulted in oxygen transfer to the latter. The spectrum of the $\boldsymbol{b}_4$ ion from (1,2-*s*-AAAG-OCH$_3$ + H)$^+$ (Figure 4.7 (b)) was still different, showing only a minor loss of CO, whereas the main sequence involved the formation of the $\boldsymbol{b}_3$ sequence ion at *m/z* 488, and its further dissociation by loss of C_6H_5N from the nitrile imine moiety that was followed by loss of [Ala]. The dissociations observed in the Figure 4.7 (a) and (b) spectra are probably driven by the high reactivity of the phenylnitrene moiety that has been known to undergo ring enlargement and nucleophilic addition reactions that may involve the peptide groups in the *m/z* 559 ions.

4.3.1.4 Crosslinking Yields of 1,2-*s*-AAAG and 1,4-*s*-AAAG

We estimated the crosslinking yield by combining the intensities of identified straight-chain and cyclized ions. Cyclized ions were represented by the sum of β_n ion intensities (*m/z* 577, 506, and 435), along with the intensity of the loss of water ion (*m/z* 616) which was absent in the CID spectra of linear peptide ions and may indicate a cyclized structure. Non-cyclized ions were represented by the combined intensities of $\boldsymbol{b}_n$ and ($\boldsymbol{b}_n$ – C_6H_5N) ions. The summarized data are provided in Table 4.30.

Compared to the 250 nm results which the survivor ion retained at 38-60%, UVPD at 213 nm resulted in the more extensive fragmentation, with $(MH – N_2)^+$ ion survivor fractions of 0.9-5%.At 213 nm, crosslinked ion fractions were 27% for 1,2-*s*-AAAG and 38% for 1,4-*s*-AAAG, both lower than the fractions observed at 250 nm, which were 65% and 77%, respectively. The high crosslinking yield from 1,4-*s*-AAAG was unexpected, as the scaffold's trans-1,4-stereochemistry was designed to position the peptide and tetrazole groups on opposite sides of the cyclohexane ring, limiting their steric accessibility for crosslinking. This discrepancy is addressed in the calculations section.

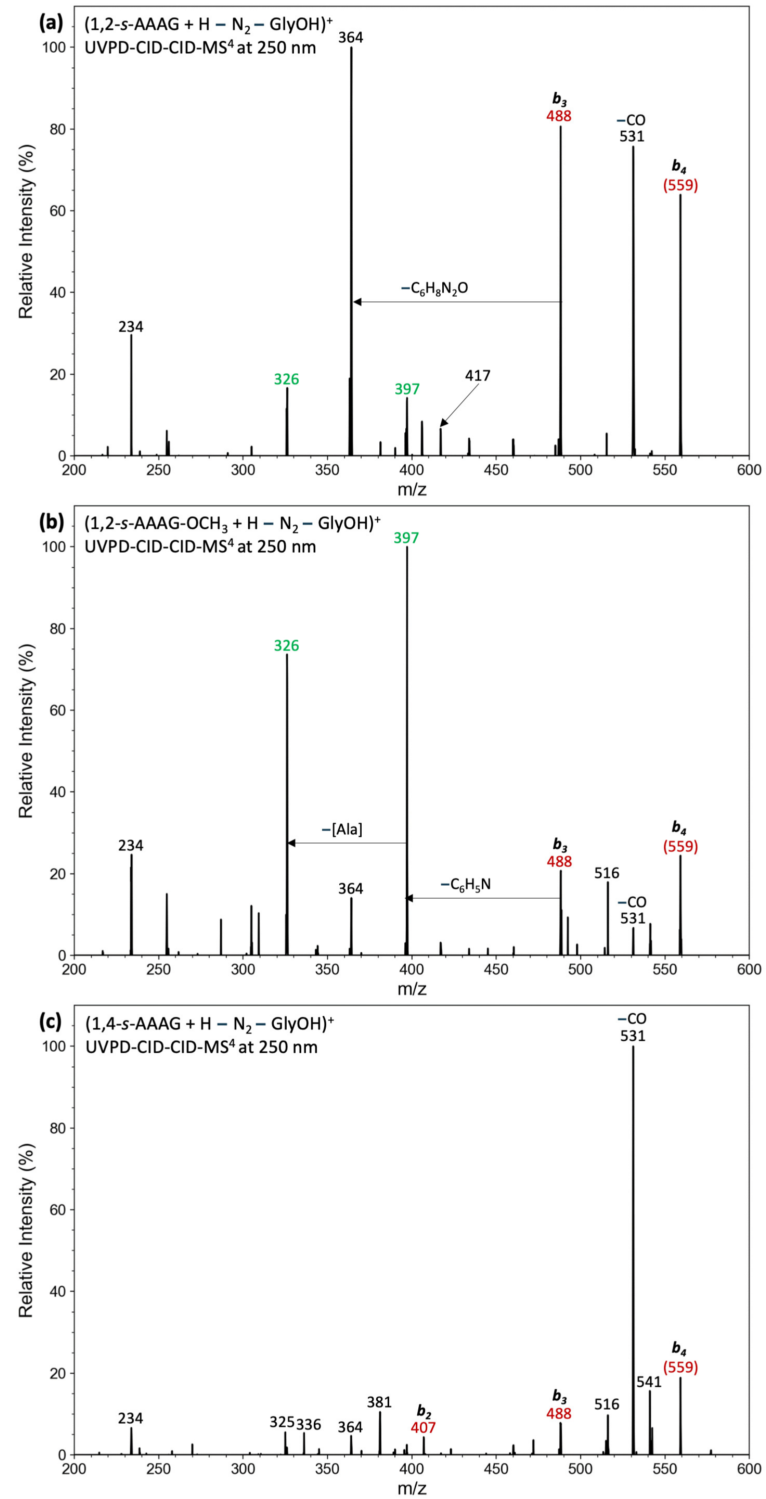


Figure 4.7: CID-MS^4 of isomeric *m/z* 559 sequence ions produced by (a) loss of GlyOH from (1,2-*s*-AAAG + H – N_2)$^+$, (b) loss of $GlyOCH_3$ from (1,2-*s*-AAAG-OCH_3 + H – N_2)$^+$ by UVPD of (1,2-*s*-AAAG-OCH_3 + H)$^+$, and (c) loss of GlyOH from (1,4-*s*-AAAG + H – N_2)$^+$. For ion color coding see text in Figure 4.1.

Crosslinking yields were also determined from CID-MS^3 of nitrile imine ions generated by UVPD at 213 and 250 nm (Table 4.30). The 250 nm data for (1,2-*s*-AAAG + H – N$_2$)$^+$ and (1,4-*s*-AAAG + H – N$_2$)$^+$ showed high crosslink fractions, 89% and 84%, respectively. The increase in crosslink fractions from UVPD-MS^2 to CID-MS^3 likely resulted from the depletion of linear peptide ions during UVPD, due to their tendency toward facile backbone dissociation, leaving a higher proportion of cyclic crosslinks to be probed in the subsequent CID-MS^3 step.

Table 4.30: Crosslinking Yields of Peptide-Nitrile Imine Scaffolds

Compound/Ion	Ion Activation	%(MH–N_2)^a	%Crosslinks^a
UVPD-MS^2			
1,2-*s*-AAAG	250 nm	46	65
1,2-*s*-AAAG	213 nm	0.9	27
1,4-*s*-AAAG	250 nm	60	77
1,4-*s*-AAAG	213 nm	5	38
1,2-*s*-AAAG-OCH_3	250 nm	38	<0.1
1,2-*s*-AAAG-OCH_3	213 nm	2.7	<0.1
1,2-*s*-AAAHG	213 nm	42	65
1,4-*s*-AAAHG	213 nm	47	60
1,2-*s*-AAAHG-OCH_3	213 nm	38	<0.1
CID-MS^3			
1,2-*s*-AAAG–N_2	250 nm		89
1,2-*s*-AAAG–N_2	213 nm		74
1,4-*s*-AAAG–N_2	250 nm		84
1,4-*s*-AAAG–N_2	213 nm		57
1,2-*s*-AAAG-OCH_3–N_2	250 nm		28
1,2-*s*-AAAG-OCH_3–N_2	213 nm		37
1,2-*s*-AAAHG–N_2	250 nm		87
1,2-*s*-AAAHG–N_2	213 nm		91
1,4-*s*-AAAHG–N_2	213 nm		91
1,2-*s*-AAAHG-OCH_3–N_2	213 nm		2.1

[a] Relative to the sum of all identified fragment ion intensities.

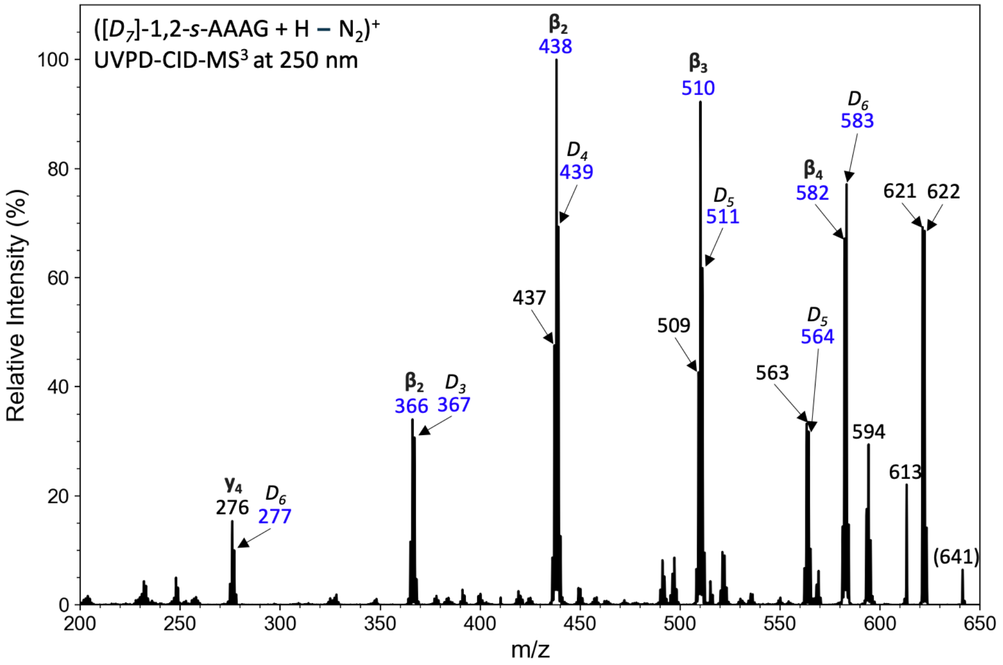
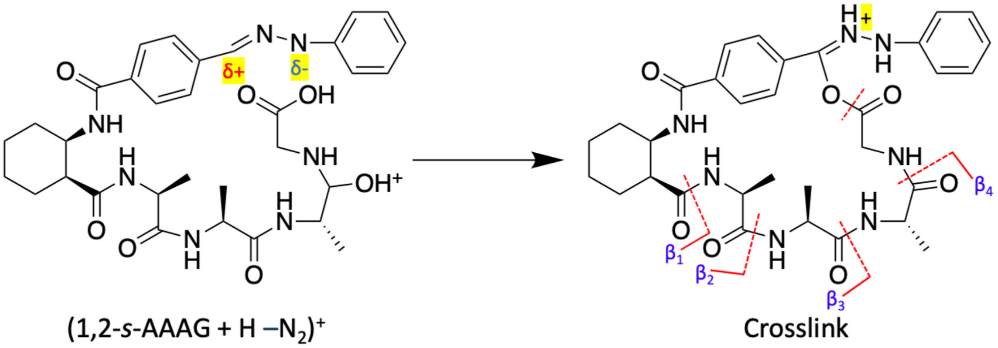


Figure 4.8: CID-MS^3 of ([D_7]-1,2-*s*-AAAG + H – N_2)$^+$ (*m/z* 641) produced by 250 nm UVPD N_2 loss from ([D_7]-1,2-*s*-AAAG + H)$^+$. For ion color coding see text in Figure 4.1.

4.3.2 Dertermination of Crosslinking Sites in 1,2-*s*-AAAG

To track proton migrations in the crosslinks, we used full H/D exchange in (1,2-*s*-AAAG + H)$^+$ to produce the deuterated ion ([D_7]-1,2-*s*-AAAG + H)$^+$ (*m/z* 669) which was then photolyzed at 250 nm, and the ([D_7]-1,2-*s*-AAAG + H – N_2)$^+$ ion formed (*m/z* 641) was investigated by CID-MS^3 (Figure 4.8). The observed deuterium retention in the β_n fragment ions was consistent with the expected neutral losses of [[D_1]Gly] and [[D_1]Gly([D_1]Ala)$_n$] (n = 1-3). It should be noted that the fragment ions underwent back exchange while stored in the ion trap, resulting in a distribution of deuterated species. For example, the abundant β_4 ion of an expected [D_6] content showed a distribution of [D_6], [D_5], and [D_4] species, and likewise for the other β_n fragment ions (Figure 4.8).

The sequential loss of [Gly] and [Ala] from (1,2-*s*-AAAG + H – N_2)$^+$ may be interpreted as resulting from crosslinking to the nitrile imine by the Gly COOH group, which was associated with proton transfer and anchored one of carboxyl oxygen atoms in the nitrile imine moiety. This is shown in Scheme 4.2 that indicates rupture of the crosslinked O–C=O ester bond accompanied by backbone amide dissociations. These two types of dissociations can occur competitively, leading to the same result.



Scheme 4.2: Proposed Formation of β_n Sequence Fragment Ions from Crosslinked ((1,2-*s*-AAAG + H – N_2)$^+$.

This hypothesis was proved by blocking the carboxyl OH by converting it to a methyl ester in (1,2-*s*-AAAG-OCH_3 + H)$^+$ and obtained the CID and UVPD-MS^2 spectra (Figure 4.9 (a) and (b), Table 4.8 and 4.9). CID-MS^3 of the methyl ester ion, (1,2-*s*-AAAG-OCH_3 + H – N_2)$^+$, *m/z* 648, that was generated by UVPD of (1,2-*s*-AAAG-OCH_3 + H)$^+$ (Figure 4.9 (c)), showed a markedly diminished formation of β_n ions while the spectrum was dominated by $\boldsymbol{b}_1$-$\boldsymbol{b}_4$ ions from cleavages of the presumably linear peptide chain (Figure 4.9 (c), Table 4.10). The impact of blocking the Gly carboxyl as a methyl ester in (1,2-*s*-AAAG-OCH_3 + H)$^+$ was substantial, as seen in Table 4.30. UVPD-MS^2 showed nearly complete suppression of crosslinked fragment ions, with the methyl esters exhibiting significantly reduced crosslink formation. This effect was particularly large for the methyl esters where the crosslinks did not dissociate on UVPD and were then revealed by CID-MS^3. Nevertheless, the fraction of crosslinked ions in methyl ester-modified samples was notably lower than that observed for carboxyl-terminated peptide conjugates.

4.3.3 Ion Mobility, Collision Cross Sections, and Ion Structures of 1,2-*s*-AAAG and 1,4-*s*-AAAG

To investigate the composition of precursor ion conformers and crosslinking products, we employed cyclic ion mobility mass spectrometry. Collision cross sections (CCS_{IMS}) were determined and compared with theoretical CCS_{calc} values obtained from fully optimized ion structures and charge distributions. The arrival time distribution (ATD) of (1,2-*s*-AAAG + H)$^+$ obtained after 20 passes (19.6 m path length) showed two peaks with very similar CCS_{IMS} = 255 and 258 Å^2, respectively (Figure 4.10). Compared with that, the ATD of (1,4-*s*-AAAG + H)$^+$ (15 passes) gave one peak of CCSIMS = 271 Å^2 (Figure 4.11) with a minor tail

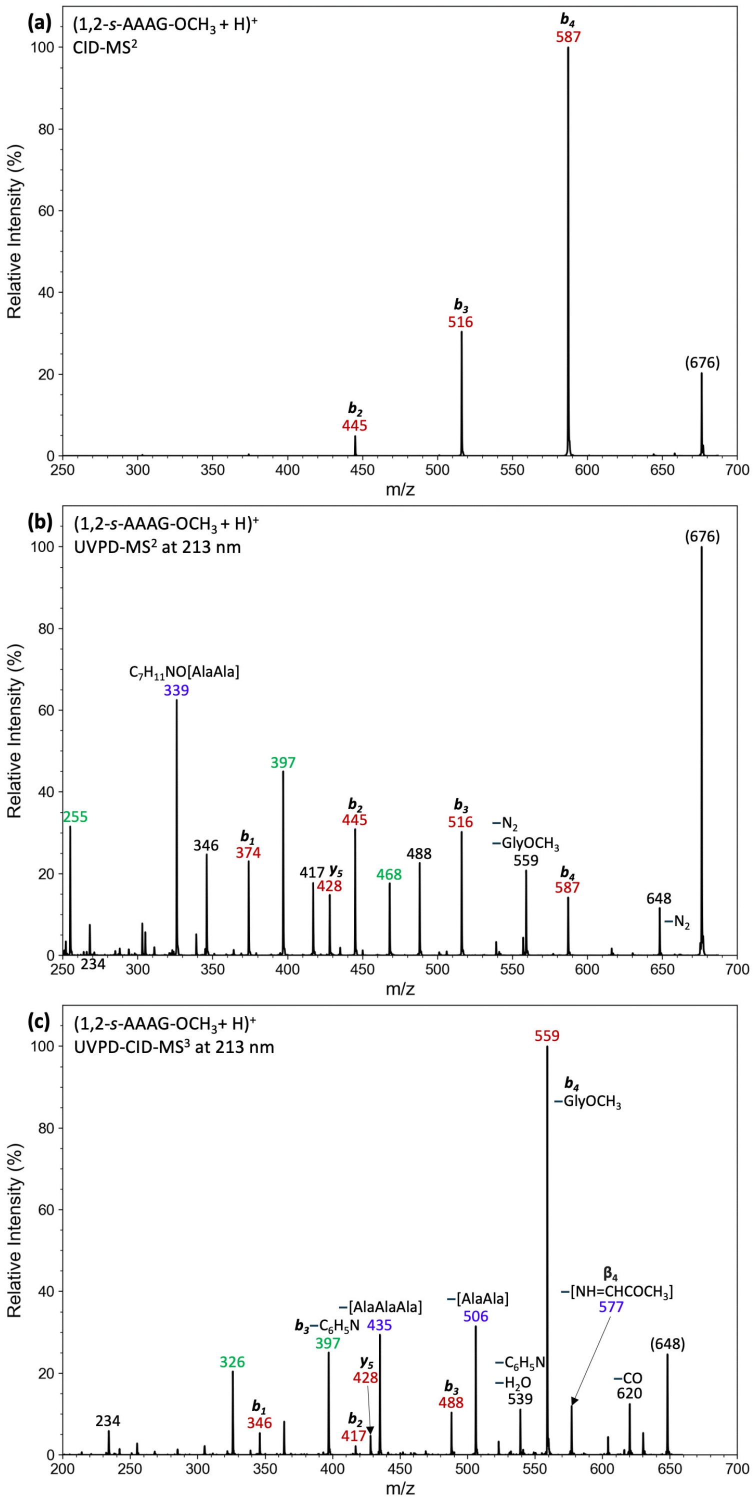


Figure 4.9: (a) CID-MS^2 of (1,2-*s*-AAAG-OCH$_3$ + H)$^+$ (*m/z* 676), (b) UVPD-MS^2 of (1,2-*s*-AAAG-OCH$_3$ + H)$^+$ (*m/z* 676) obtained at 213 nm, and (c) CID-MS^3 of (1,2-*s*-AAAG-OCH$_3$ + H – N$_2$)$^+$ (*m/z* 648) from 213 nm UVPD. For ion color coding see text in Figure 4.1.

toward longer arrival times which indicated that the ion was produced as one component or an unresolved mixture of several components of very close CCS. Using a combination of BOMD of several protonation isomers and M06-2X/6-31+G(d,p) geometry optimization of several low-energy structures, we were able to identify the lowest-energy ions and calculate their $\mathrm{CCS_{calc}}$. Figure 4.10 shows two lowest-energy isomers of (1,2-*s*-AAAG + H)$^+$, **1,2-A1** and **1,2-A2**, that had virtually identical relative energies, and can be considered energetically equivalent at the DFT level of theory. Upon BOMD and gradient optimization of initial structures with several different amide protonation sites, all low-energy structures collapsed to a single protonation isomer being protonated at the diaryltetrazole carboxamide carbonyl. The protonated amides developed strong hydrogen bonds to the Ala amide groups, which were Ala-2 in **1,2-A2** and Ala-3 in **1,2-A1** (Figure 4.10). The structures also differed in the orientation of the *cis*-1,2-diaryltetrazole carboxamide and peptide groups on the cyclohexane ring. The peptide chain was equatorial in **1,2-A1** and axial in **1,2-A2**. The hydrogen bonding patterns determined the compact peptide conformations whereas the rigid 2,5-diaryltetrazole moiety stuck out of the structures. This indicated that the incipient nitrile imine was not in the close proximity of the peptide groups, and crosslinking must have involved peptide conformational change. The $\mathrm{CCS_{calc}}$ of **1,2-A1** and **1,2-A2**, 257 and 260 Å^2, respectively, matched within 1% the experimental $\mathrm{CCS_{IMS}}$ of the peaks detected in the Figure 4.10 ion mobilogram. It should be noted that the MobCal-MPI CCS calculations that were based on averaging 10 iterations gave 0.5-1.1% root-mean-square deviations which should be considered when comparing the calculated and experimental CCS. The $\mathrm{CCS_{calc}}$ of other low-energy conformers were 2-4% larger and provided a less satisfactory fit with the $\mathrm{CCS_{IMS}}$ (Figure 4.10 **1,2-A3** and **1,2-A4**).

Several low-energy optimized structures of (1,4-*s*-AAAG + H)$^+$ were obtained that all showed a uniform protonation site that was on the Ala-2 amide (Figure 4.11). The structures fell into two groups; one with the carboxyl group reaching to the tetrazole ring, as represented by **1,4-A1**, and the other with a folded peptide sequence as represented by **1,4-A2**. The two groups differed in their $\mathrm{CCS_{calc}}$ which ranged between 262-270 Å^2 for the **1,4-A1** types and 282-286 Å^2 for the **1,4-A2** types. The gas-phase structures favored the **1,4-A2** group by Gibbs energy, which was largely mitigated by solvation in water that favored **1,4-A1** (Figure 4.11). The structures of the **1,4-A1** type gave a closer agreement between the $\mathrm{CCS_{IMS}}$ and $\mathrm{CCS_{calc}}$ and thus we considered them representative of the (1,4-*s*-AAAG + H)$^+$ ions produced by electrospray. Loss

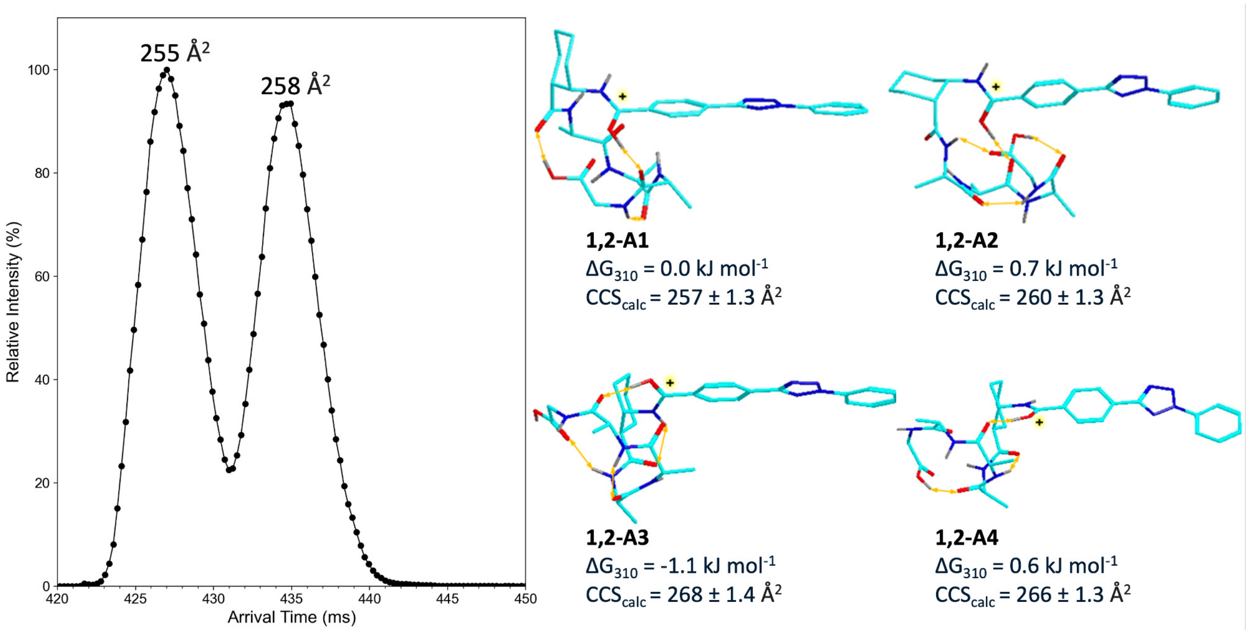


Figure 4.10: Arrival time distributions (ATD) of (1,2-*s*-AAAG + H)$^+$ after 20 passes. M06-2X/6-31+G(d,p) optimized structures, M06-2X/def2qzvpp relative Gibbs energies, and CCS$_{\text{calc}}$ of low energy (1,2-*s*-AAAG + H)$^+$ ions. Atom color coding is as follows: cyan = C, blue = N, red = O, gray = H. Only exchangeable NH, OH hydrogens are shown to avoid clutter. Major hydrogen bonds are indicated by ochre double-headed arrows.

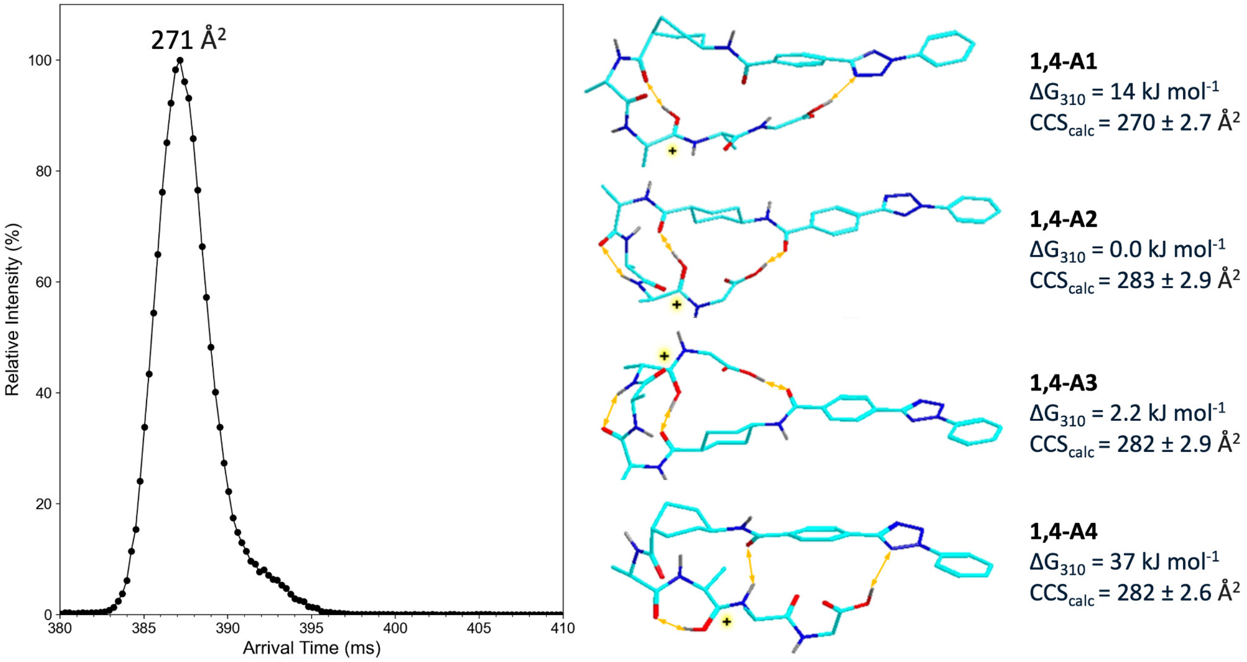
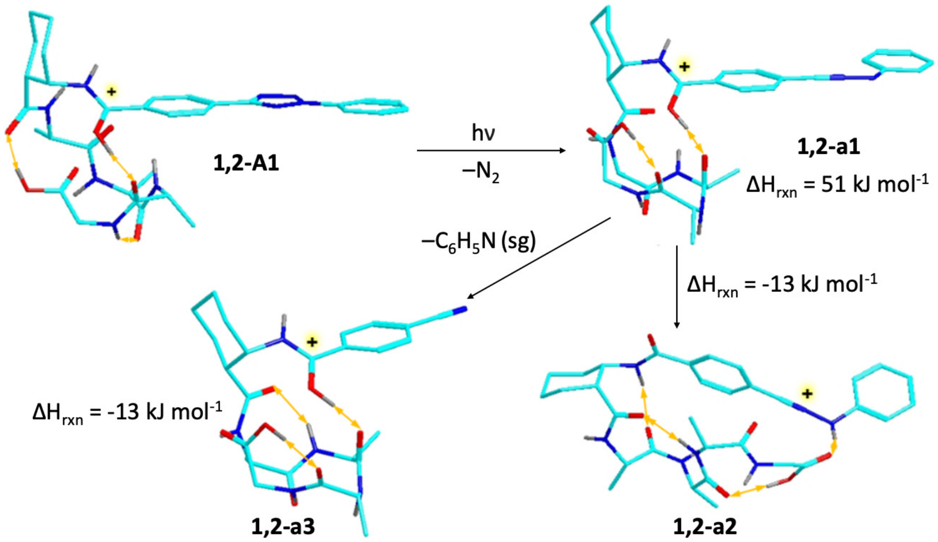


Figure 4.11: Arrival time distributions (ATD) of (1,4-*s*-AAAG + H)$^+$ after 15 passes. M06-2X/6-31+G(d,p) optimized structures, M06-2X/def2qzvpp relative Gibbs energies, and CCS$_{\text{calc}}$ of low energy (1,4-*s*-AAAG + H)$^+$ ions. Structure discription as in Figure 4.10.



Scheme 4.3: Dissociation and Crosslinking Reaction Energies of (1,2-*s*-AAAG + H)$^+$.

of N_2 from (1,2-*s*-AAAG + H)$^+$ could not be studied by CID and therefore we did not have the relevant CCS_{IMS} of the nitrile imine and crosslinked products. Calculations suggested that the loss of N_2 from the tetrazole ring in (1,2-*s*-AAAG + H)$^+$ was only 51 kJ mol^{-1} endothermic leaving a substantial excitation energy in nitrile imine ion **1,2-a1** (Scheme 4.3). From the energy balance one gets E_{exc}(1,2-*s*-AAAG + H – N_2)$^+$ $\leq$ (E_{photon} – 51) = 479 and 562 kJ mol^{-1} at 250 and 213 nm, respectively, where the $\leq$ sign reflects the unknown translational energy of the departing N_2 molecule. The energy excess can drive proton migration onto the nitrile imine group that has been considered essential for facilitating peptide cyclization.[22] Alternatively, the N–N bond in excited **1,2-a1** can dissociate, driving the observed loss of C_6H_5N and subsequent backbone dissociations.

4.3.4 Photodissociation and Crosslinking Yields of 1,2-*s*-AAAHG and 1,4-*s*-AAAHG

To utilize nitrile imine formation by both UVPD and CID, scaffold conjugates that were furnished with the basic [His] residue were investigated. In this way, the peptide backbone fragmentation was suppressed by proton sequestration at the basic site, allowing us to generate nitrile imine intermediates by CID-MS^2 and study their properties by CID and UVPD-MS^3.

4.3.4.1 Spectra Interpretation of 1,2-s-AAAHG

Both CID and UVPD of the (1,2-*s*-AAAHG + H)$^+$ ions gave abundant fractions of the denitrogenated intermediates (1,2-*s*-AAAHG + H – N$_2$)$^+$ (*m/z* 771, Figure 4.12 (a) and 4.14 (a), Table 4.14 and 4.15). Backbone dissociations were chiefly observed for 213 nm UVPD (Figure 4.14 (a)) where they primarily led to two ion series, both originating from (1,2-*s*-AAAG + H – N$_2$)$^+$. In contrast, competitive peptide chain dissociations in tetrazole conjugates, i.e. without N$_2$ loss, were only minor as indicated by the red-annotated ions. The main series of blue-annotated ions in the CID-MS^3 spectra corresponded to β_n fragments by loss of [Gly] (*m/z* 714), and continuing by sequential losses of [His] and [Ala] at *m/z* 577, 506, 435, and 364 (Figure 4.15 (a)). In addition, the presence of the basic His residue steered protonation to complementary β_n ions at *m/z* 266, 337, and 408 that were also indicative of cyclized structures. The β_n ions formally corresponded to [$\boldsymbol{y}_n$ + 2H – H$_2$O]$^+$, which were represented as γ_n.

To investigate further dissociations of the (1,2-*s*-AAAHG + H – N$_2$)$^+$ ions (*m/z* 771), both CID and UVPD in the UVPD-CID-MS^3 and CID-UVPD-MS^3 were applied (Figure 4.15 (a) and 4.16 (a), Tables 4.16 and 4.18). CID-MS^3 of (1,2-*s*-AAAHG + H – N$_2$)$^+$ revealed prominent peaks due to the loss of water and CO at *m/z* 753 and 743, respectively (Figure 4.15 (a)). The absence of these fragments in the (1,2-*s*-AAAHG + H)$^+$ ions suggests they likely originate from cyclized structures. Additionally, the *m/z* 605 ion, resulting from combined loss of GlyOH and C$_6$H$_5$N, indicated the presence of non-cyclized linear structures. Most fragment ions at *m/z* 714, 577, 506, 435, and 364 were of the β_n type. These ions represent cyclized structures and are attributed to the loss of internal residues. CID-MS^4 of the prominent CID-MS^3 peaks of (1,2-*s*-AAAHG + H – N$_2$)$^+$ provided additional evidence supporting these findings (Table 4.31). UVPD of (1,2-*s*-AAAHG + H – N$_2$)$^+$ produced similar fragmentation patterns (Figure 4.16 (a)). However, due to the higher excitation energy in UVPD, the sequence ion series extended to lower-mass ions through more extensive dissociation of light-absorbing intermediate fragments.

4.3.4.2 Spectra Interpretation of 1,4-s-AAAHG

Photodissociation of (1,4-*s*-AAAHG + H)$^+$ (*m/z* 799) at 213 nm gave an intense (1,2-*s*-AAAHG + H – N$_2$)$^+$ ion (*m/z* 771) along with a series of backbone fragment ions (Figure 4.14 (b), Table 4.23). These consisted of the β_n series of sequential losses of [Gly], [His], and [Ala] internal fragments (blue-annotated

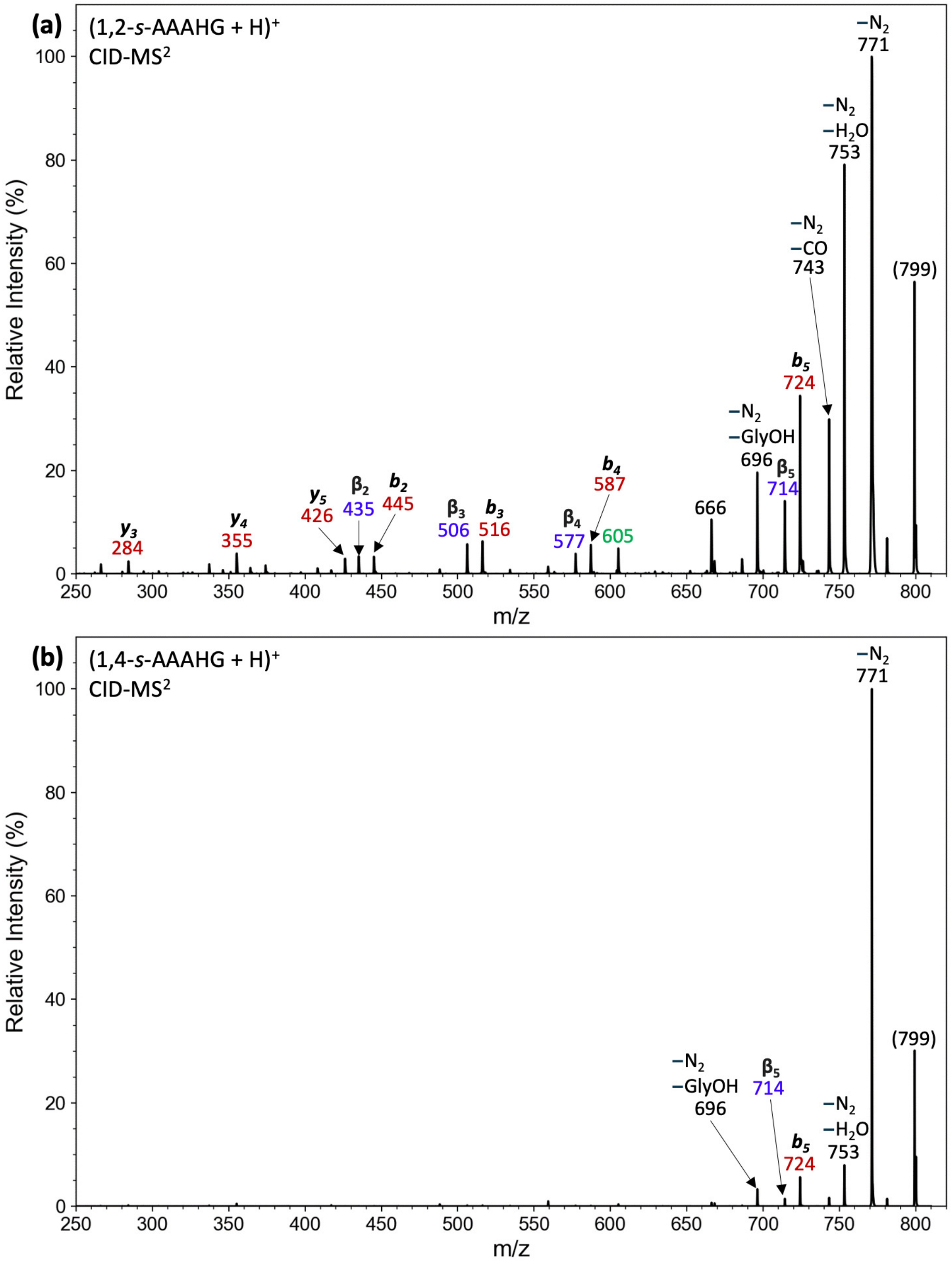


Figure 4.12: (a) CID-MS^2 of (1,2-*s*-AAAHG + H)$^+$ (*m/z* 799), and (b) CID-MS^2 of (1,4-*s*-AAAHG + H)$^+$ (*m/z* 799). For ion color coding see text in Figure 4.1.

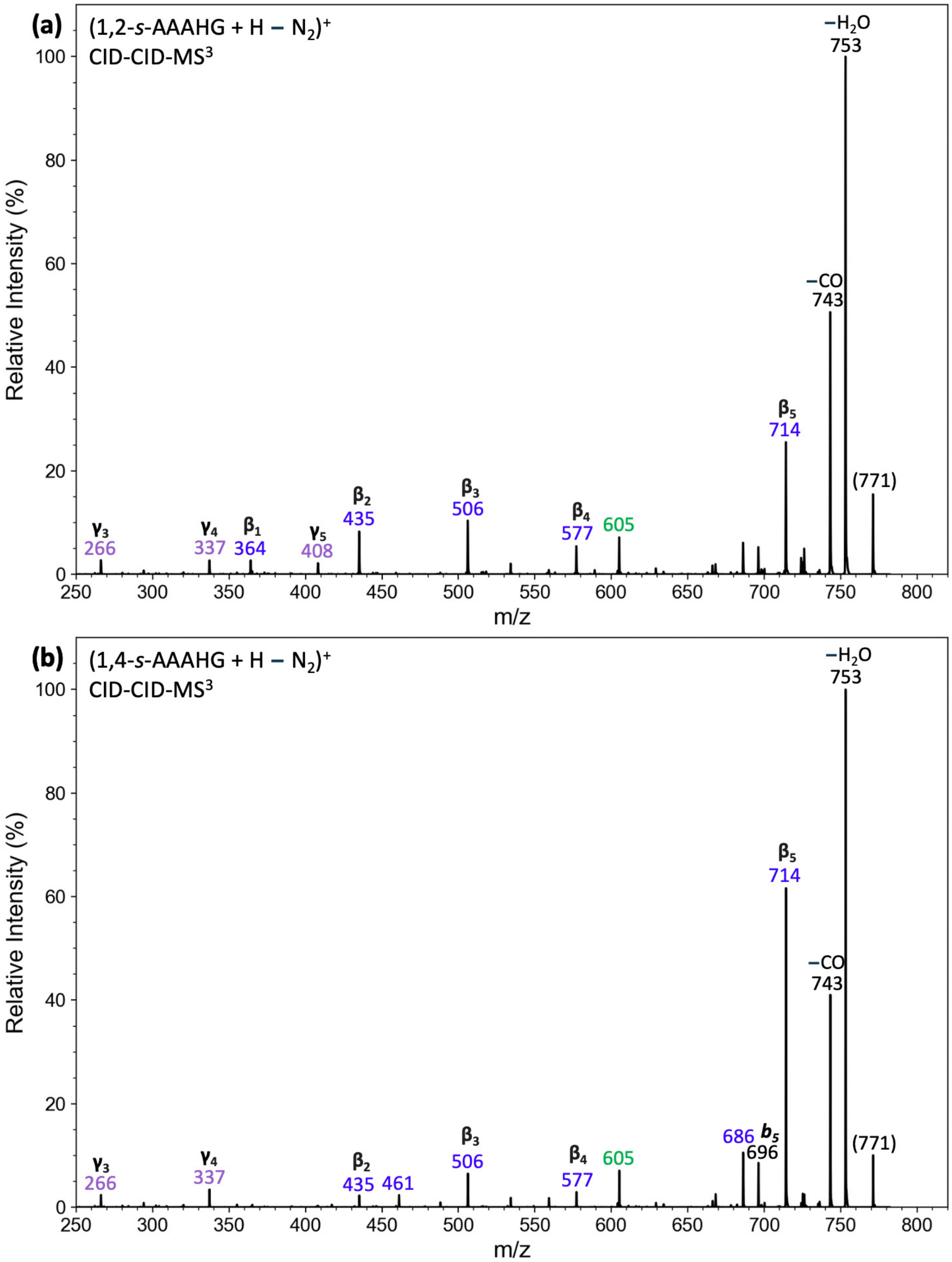


Figure 4.13: (a) CID-MS^3 of (1,2-*s*-AAAHG + H – N$_2$)$^+$ (*m/z* 771) generated by CID of (1,2-*s*-AAAHG + H)$^+$, and (b) CID-MS^3 of (1,4-*s*-AAAHG + H – N$_2$)$^+$ (*m/z* 771) generated by CID of (1,4-*s*-AAAHG + H)$^+$. For ion color coding see text in Figure 4.1.

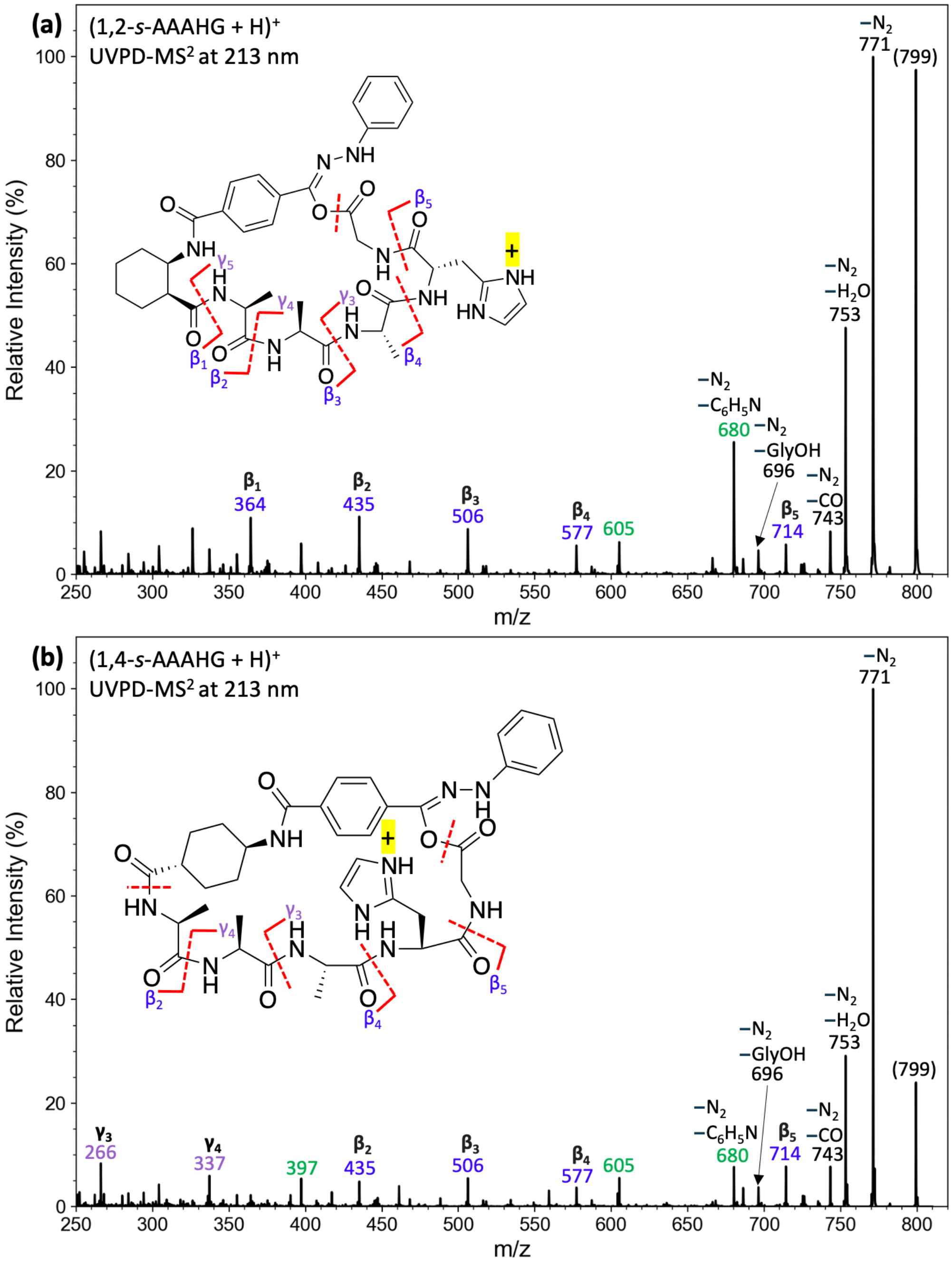


Figure 4.14: (a) UVPD-MS^2 at 213 nm of (1,2-*s*-AAAHG + H)$^+$ (*m/z* 799), and (b) UVPD-MS^2 at 213 nm of (1,4-*s*-AAAHG + H)$^+$ (*m/z* 799). For ion color coding see text in Figure 4.1.

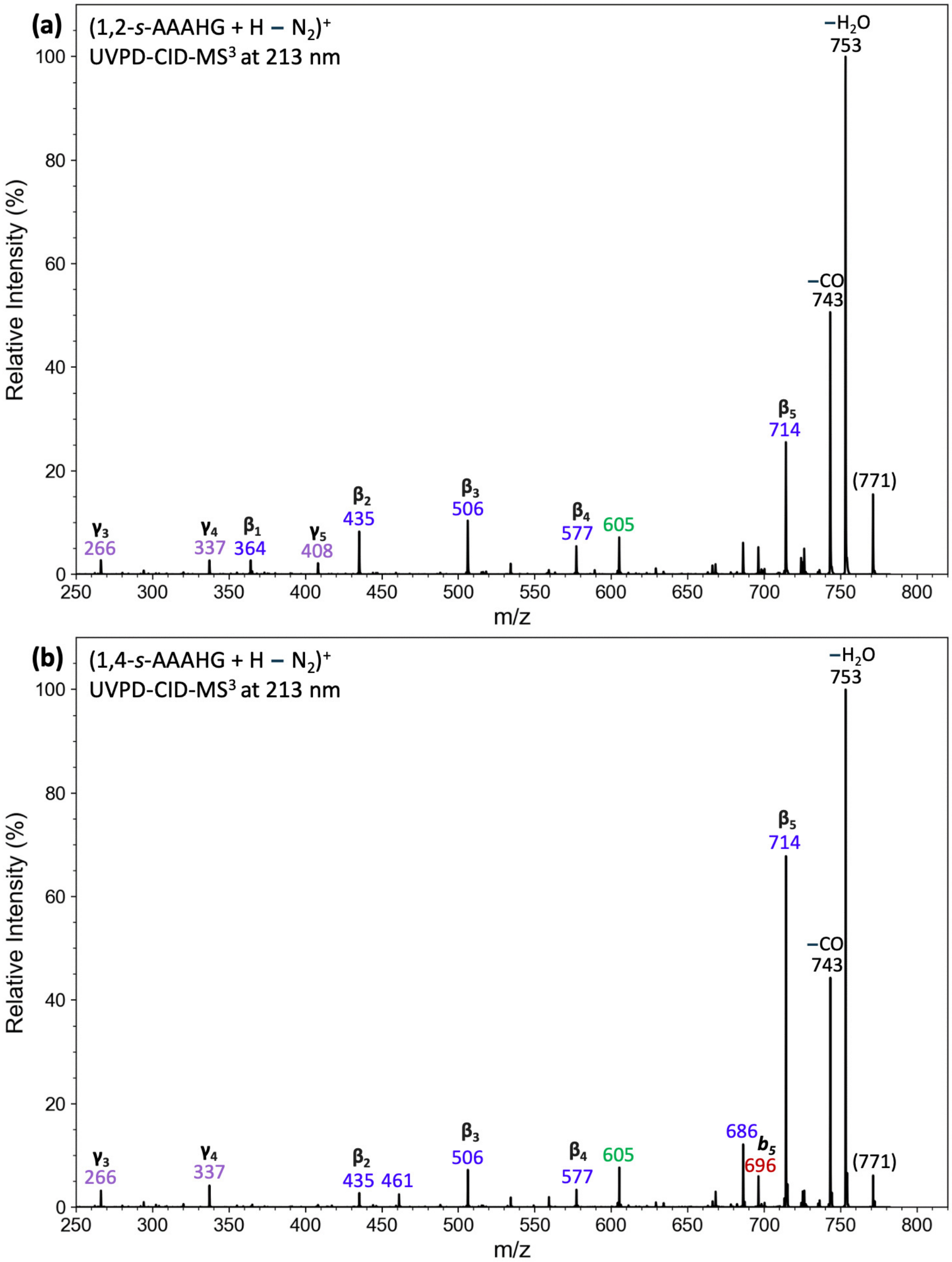


Figure 4.15: (a) CID-MS^3 of (1,2-*s*-AAAHG + H – N_2)$^+$ (*m/z* 771) generated by UVPD of (1,2-*s*-AAAHG + H)$^+$, and (b) CID-MS^3 of (1,4-*s*-AAAHG + H – N_2)$^+$ (*m/z* 771) generated by UVPD of (1,4-*s*-AAAHG + H)$^+$. For ion color coding see text in Figure 4.1.

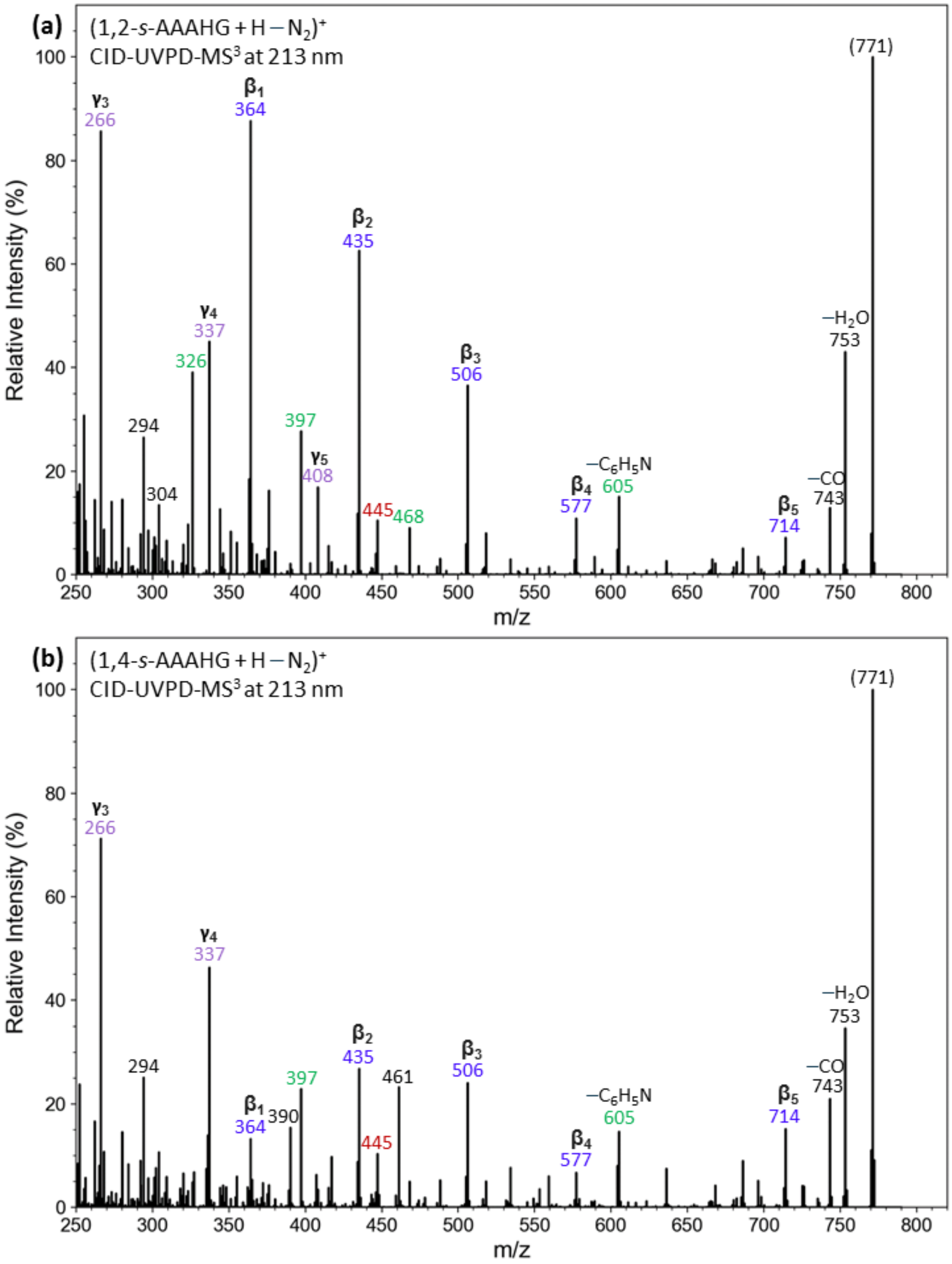


Figure 4.16: (a) UVPD-MS^3 at 213 nm of (1,2-*s*-AAAHG + H – N$_2$)$^+$ (*m/z* 771) generated by CID of (1,2-*s*-AAAHG + H)$^+$, and (b) UVPD-MS^3 at 213 nm of (1,4-*s*-AAAHG + H – N$_2$)$^+$ (*m/z* 771) generated by UVPD of (1,4-*s*-AAAHG + H)$^+$. For ion color coding see text in Figure 4.1.

Table 4.31: CID-MS^4 Fragments of (1,2-*s*-AAAHG + H – N_2)$^+$ *m/z* 771 from 1,2-*s*-AAAHG Following UVPD at 213 nm

753	743	714	605	577	506	435	408
–H_2O	–CO	β_5	$\mathbf{b}_5$–C_6H_5N	β_4	β_3	β_2	γ_5
735	726	686	577	506	435	364	390
725	714	577	468	435	364	180	380
710	698	506	397	364		135	135
682	686	435				120	120
605	604						
577	589						
468	577						
397	518						
326	506						
268							

ions), indicating crosslinking. The other main series was initiated by loss of C_6H_5N from the nitrile imine (*m/z* 680), followed by standard backbone cleavages yielding $\boldsymbol{b}_n$ type fragment ions (green-annotated series). Loss of N_2 was also induced by CID that was followed by backbone cleavages (Table 4.24).

CID-MS^3 (Figure 4.15 (b), Table 4.25) and UVPD-MS^3 of CID-generated (1,4-*s*-AAAHG + H – N_2)$^+$ (*m/z* 771, Table 4.26), as well as CID-MS^3 and further CID-MS^4 of UVPD-generated (1,4-*s*-AAAHG + H – N_2)$^+$ (Table 4.27 and 4.32) showed mainly the β_n ions that were complemented by the γ_3 and γ_4 ions at *m/z* 266 and 337, respectively. Both the sequential fragmentation starting from the C-terminal [Gly], and the absence of direct loss of $[\text{Ala}]_n$ from (1,4-*s*-AAAHG + H – N_2)$^+$ indicated that crosslinking to the nitrile imine was chiefly realized by the [Gly] residue. The linkage and peptide chain dissociations could follow the same mechanism as those shown in Scheme 4.2 for 1,2-*s*-AAAG (Figure 4.14).

4.3.4.3 Crosslinking Yields of 1,2-*s*-AAAHG and 1,4-*s*-AAAHG

The crosslinking yields at 213 nm for (1,2-*s*-AAAHG + H)$^+$ and (1,4-*s*-AAAHG + H)$^+$ were similar at 65 and 60%, respectively (Table 4.30). A significant drop of yield was recorded for (1,2-*s*-AAAHG-OCH_3 + H)$^+$ (< 0.1%), indicating the critical role of the free carboxyl in binding to the nitrile imine. The crosslinking yields upon CID of (1,2-*s*-AAAHG + H – N_2)$^+$ and (1,4-*s*-AAAHG + H – N_2)$^+$ that were

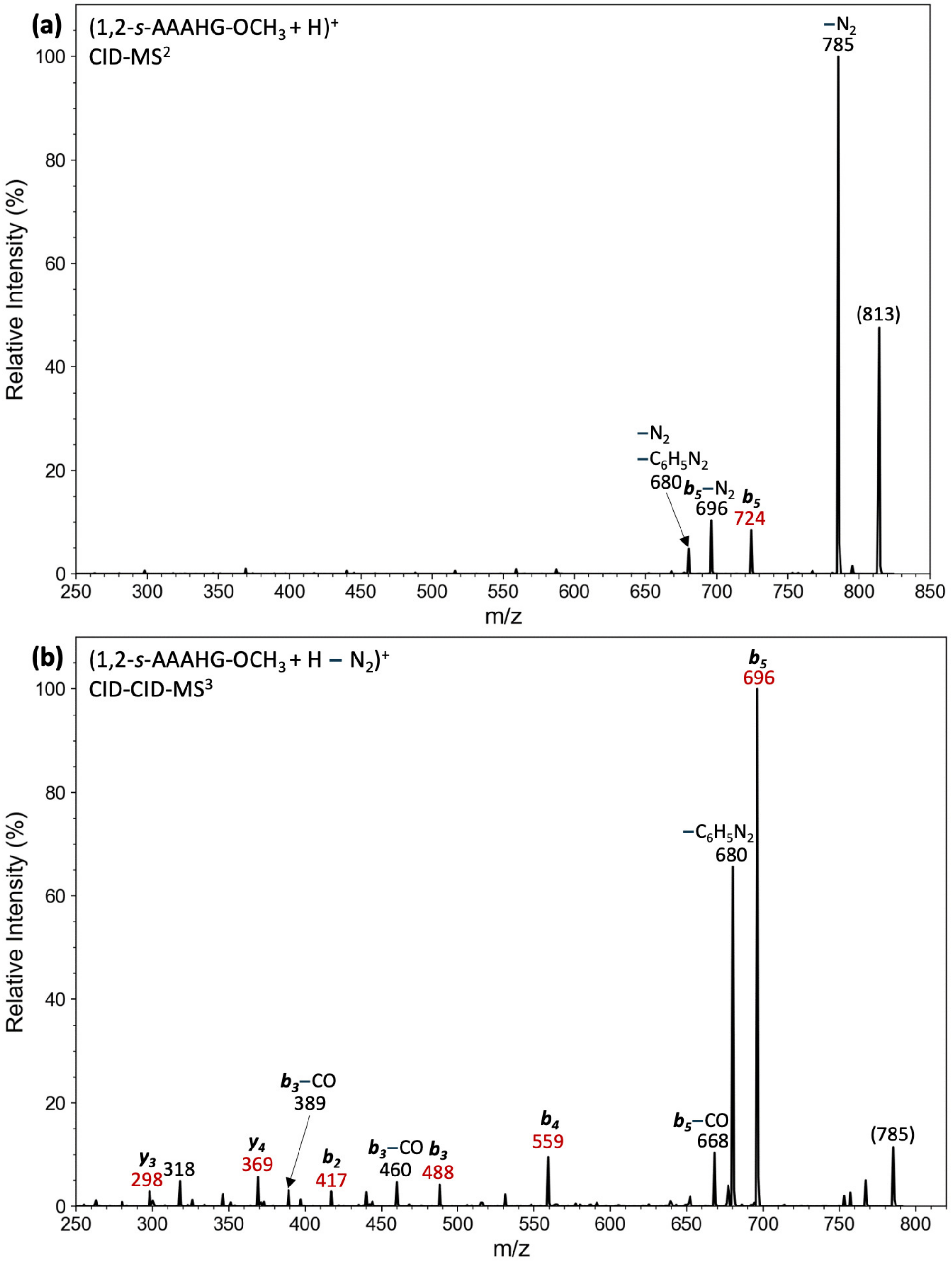


Figure 4.17: (a) CID-MS^2 of (1,2-*s*-AAAHG-OCH$_3$ + H)$^+$ (*m/z* 813), and (b) CID-MS^3 of (1,2-*s*-AAAHG-OCH$_3$ + H – N$_2$)$^+$ (*m/z* 785) generated by CID of (1,2-*s*-AAAHG-OCH$_3$ + H)$^+$. For ion color coding see text in Figure 4.1.

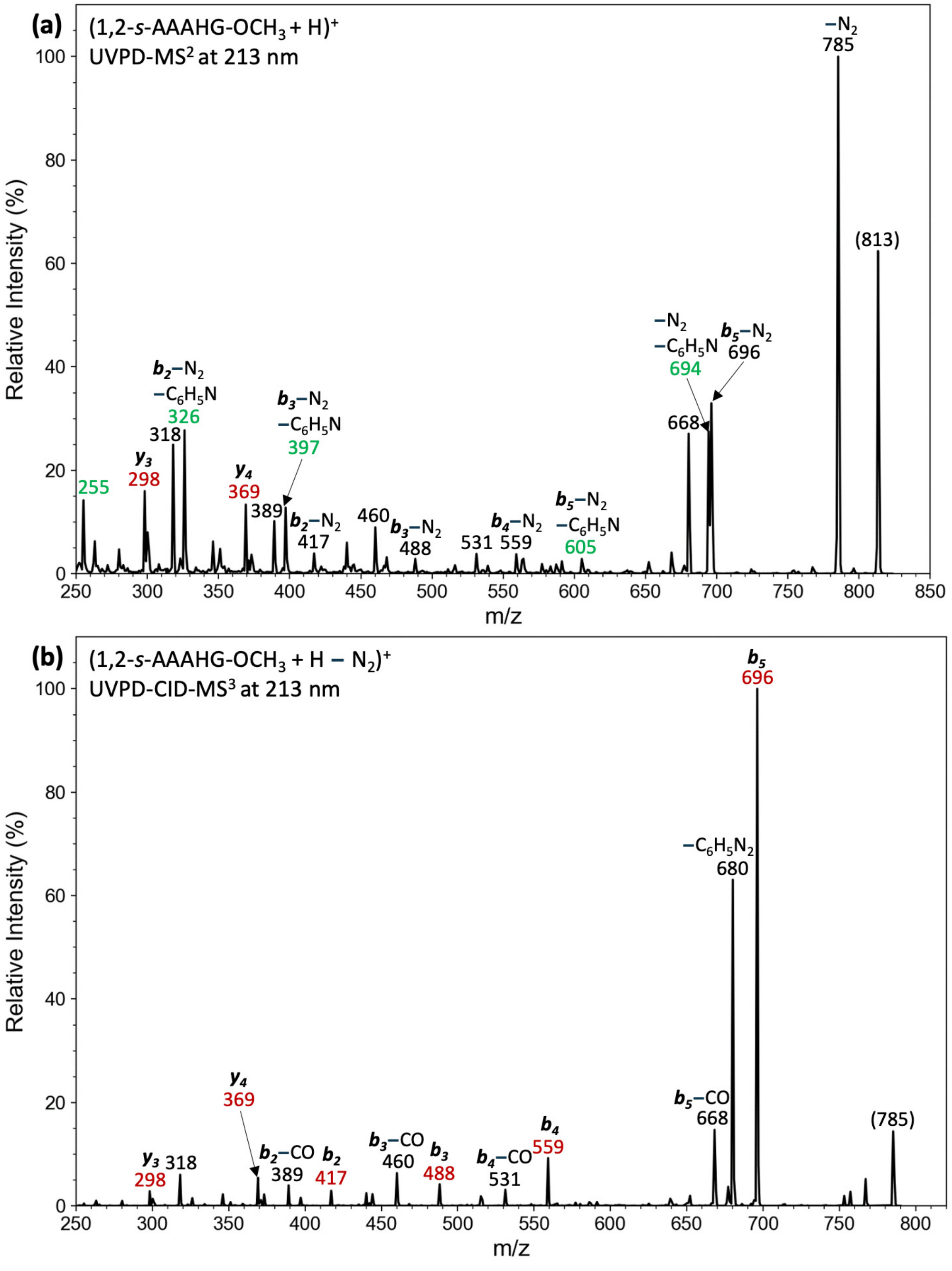


Figure 4.18: (a) UVPD-MS^2 at 213 nm of (1,2-*s*-AAAHG-OCH$_3$ + H)$^+$ (*m/z* 813), and (b) CID-MS^3 of (1,2-*s*-AAAHG-OCH$_3$ + H – N$_2$)$^+$ (*m/z* 785) generated by UVPD of (1,2-*s*-AAAHG-OCH$_3$ + H)$^+$. For ion color coding see text in Figure 4.1.

Table 4.32: CID-MS^4 Fragments of (1,4-*s*-AAAHG + H – N_2)$^+$ *m/z* 771 from 1,4-*s*-AAAHG Following UVPD at 213 nm

753	743	714	696	686	605	577	506	461	435
–H_2O	–CO	β_5	$\mathbf{b}_5$	β_5–CO	$\mathbf{b}_5$–C_6H_5N	β_4	β_3	β_3–NH_3–CO	β_2
735	726	686	668	668	587	549	478	433	417
725	725	670	652	577	577	506	461	390	407
710	714	577	587	506	561	468	435	362	390
696	698	506	577	486	468	397		265	364
682	686		559	461	397				
665	604		515	415					
605	518		488	252					
577	505		468						
397			444						
268			397						
249			245						

generated at 213 nm were comparable at 91% (Table 4.30). Ions (1,2-*s*-AAAHG + H – N_2)$^+$ produced by UVPD at 250 nm gave a similar yield of crosslinks (87%) upon CID. In contrast, carboxyl methylation, as in (1,2-*s*-AAAHG-OCH_3 + H)$^+$, caused a large drop in the crosslinking yield (2.1%, Table 4.30) which was consistent with the yield decrease upon UVPD of (1,2-*s*-AAAHG-OCH_3 + H)$^+$ methyl ester (Figure 4.18. Conversely, this large effect of blocking the carboxyl indicated that the His side-chain imidazole ring did not react with the nitrile imine to affect the crosslinking yield. The fragment ion assignment was corroborated by accurate mass measurements (Table 4.19-4.22).

4.3.5 Ion Mobility, Collision Cross Sections, and Ion Structures of 1,2-*s*-AAAHG and 1,4-*s*-AAAHG

Ion mobility of the (1,2-*s*-AAAHG +H)$^+$ ions (*m/z* 799, 4 passes) showed two resolved peaks with CCS_{IMS} = 277 and 282 Å^2 for the minor (7%) and major (93%) components, respectively (Figure 4.19). These were closely matched by the CCS_{calc} of the two lowest Gibbs energy structures **1,2-H1** and **1,2-H2** of 276 and 284 Å^2 (Figure 4.19). Both these low-energy structures were protonated on His and showed coiled peptide chains with similar hydrogen bonding patterns, reflecting their similar CCS_{calc} that were within 3%. Ion mobility of the (1,4-*s*-AAAHG +H)$^+$ions (*m/z* 799, 15 passes) also showed two peaks of CCS_{IMS} 296 and 301 Å^2 for the major (98%) and minor (2%) components (Figure 4.20). The calculated lowest-energy ion structures fell into two groups. One group, represented by **1,4-H1** and **1,4-H2**2, showed the peptide

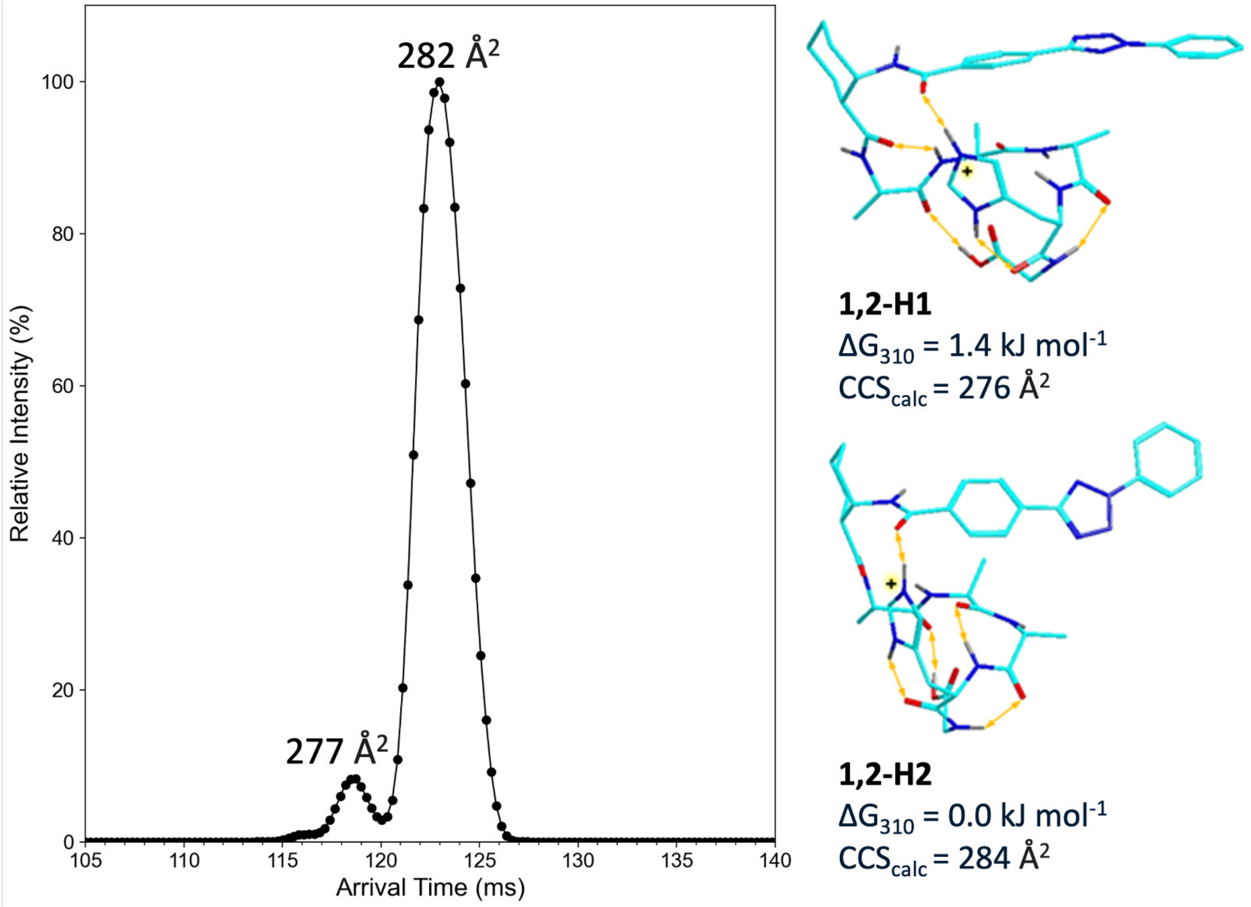


Figure 4.19: Arrival time distributions of (1,2-*s*-AAAHG + H)$^+$ after 4 passes. M06-2X/6-31+G(d,p) optimized structures, M06-2X/def2qzvpp relative Gibbs energies, and CCS$_{\mathrm{calc}}$ of low energy (1,2-*s*-AAAHG + H)$^+$ ions. Structure discription as in Figure 4.10.

hydrogen bonded to the benzoyl C=O which resulted in compact structures. The CCS$_{\mathrm{calc}}$ of **1,4-H1** closely matched the CCS$_{\mathrm{IMS}}$ of the major peak in the Figure 4.20 mobilogram. The other group, represented by **1,4-H3**, had more extended structures, resulting in larger CCS$_{\mathrm{calc}}$.

Ion mobility separation of the CID-denitrogenated ions (1,2-*s*-AAAHG + H – N$_2$)$^+$ and (1,4-*s*-AAAHG + H – N$_2$)$^+$ (*m/z* 771) showed five fully or partially resolved peaks with CCS$_{\mathrm{IMS}}$ ranging between 263-289 Å^2 (Figure 4.21 and 4.22). This made the structure assignment difficult because the CCS$_{\mathrm{calc}}$ of conformers of several different ion structures showed substantial overlap. In addition, the phenyl group conformation in both the nitrile imines and crosslinks had a large effect on the CCS$_{\mathrm{calc}}$, causing overlap between conformers of different structure types. This is illustrated in Figure 4.21 for low energy conformers of the nitrile imine (**1,2-h1**) and carboxyl-crosslinked ions **1,2-h2** and **1,2-h3**. In this case, a match between the CCS$_{\mathrm{calc}}$ and

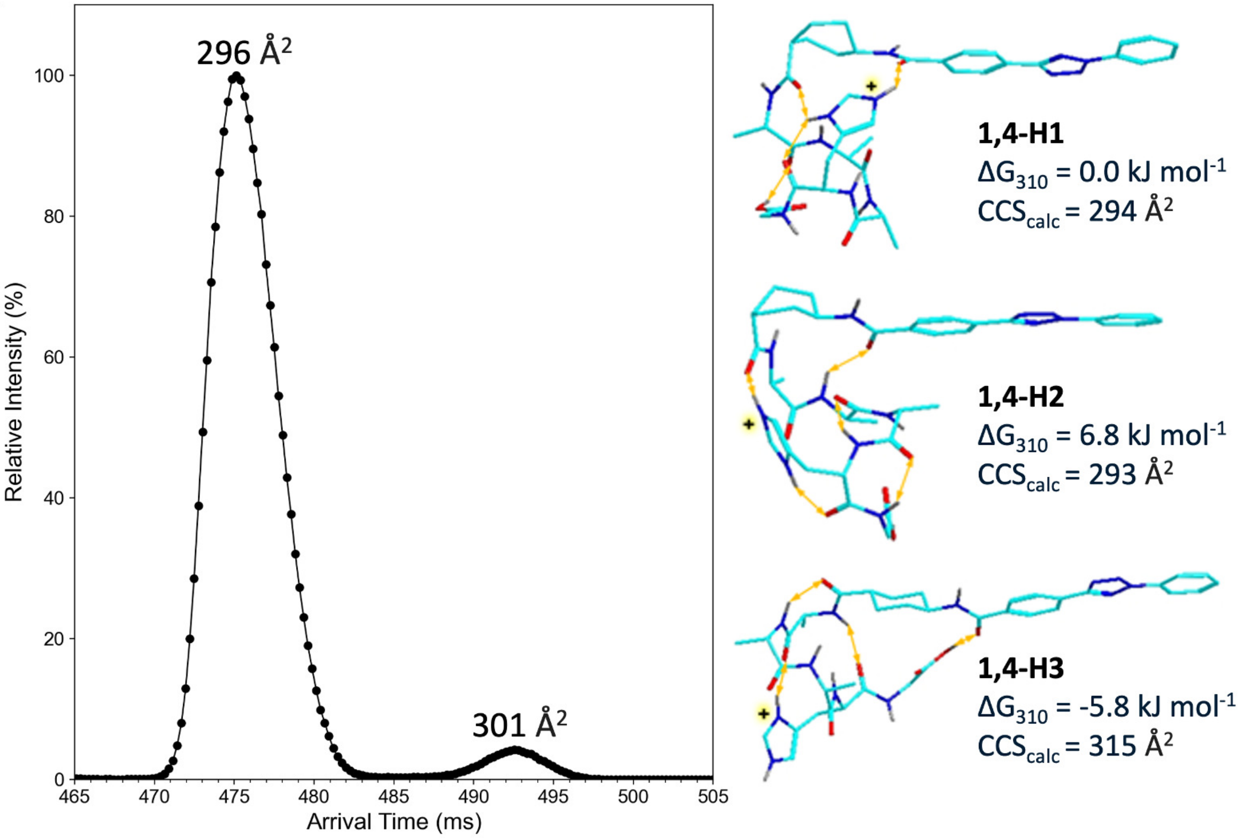


Figure 4.20: Arrival time distributions of (1,4-*s*-AAAHG + H)$^+$ after 15 passes. M06-2X/6-31+G(d,p) optimized structures, M06-2X/def2qzvpp relative Gibbs energies, and CCS$_{\text{calc}}$ of low energy (1,4-*s*-AAAHG + H)$^+$ ions. Structure discription as in Figure 4.10.

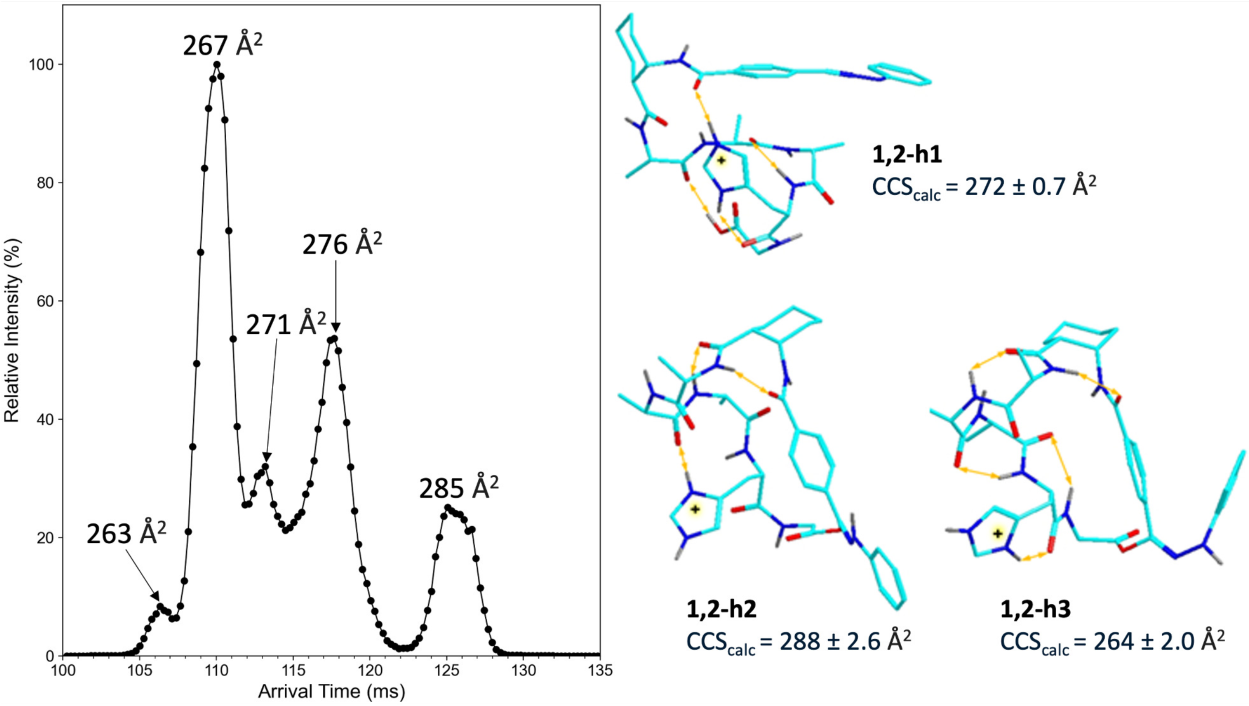


Figure 4.21: Arrival time distributions of $(1{,}2\text{-}s\text{-AAAHG} + \text{H} - \text{N}_2)^+$ after 4 passes. M06-2X/6-31+G(d,p) optimized structures and CCS$_{\text{calc}}$ of low energy **1,2-h1**-**1,2-h3** isomers. Structure discription as in Figure 4.10.

CCS$_{\text{IMS}}$ cannot be taken as an unequivocal structure assignment, although the presence of the linear (nitrile imine) and carboxyl crosslinked structures was strongly indicated by the UVPD-MS^2 and CID-MS^3 spectra (Figure 4.14 (a) and 4.15 (a)).

The CCS$_{\text{calc}}$ for crosslinked structures **1,4-h1** and **1,4-h2** had close matches with peaks in the mobilogram whereas the CCS$_{\text{calc}}$ for the nitrile imines (**1,4-h3**) had not. An interesting feature of the $(1{,}4\text{-}s\text{-AAAHG} + \text{H} - \text{N}_2)^+$ crosslinks was that the carboxyl bond to the nitrile imine was achieved only after flipping the cyclohexane ring to the twist-boat conformation. The **1,4-h3** → **1,4-h2** cyclization was calculated to be exergonic, $\Delta G_{\text{g, 310}} = -59$ kJ mol^{-1}, which compensated for the energy gain due to cyclohexane ring inversion. A further stabilization of the twist-boat conformation could be due to the strong hydrogen bond between the protonated histidine imidazole and the benzoyl carbamide in **1,4-h1** and **1,4-h2** (Figure 4.22).

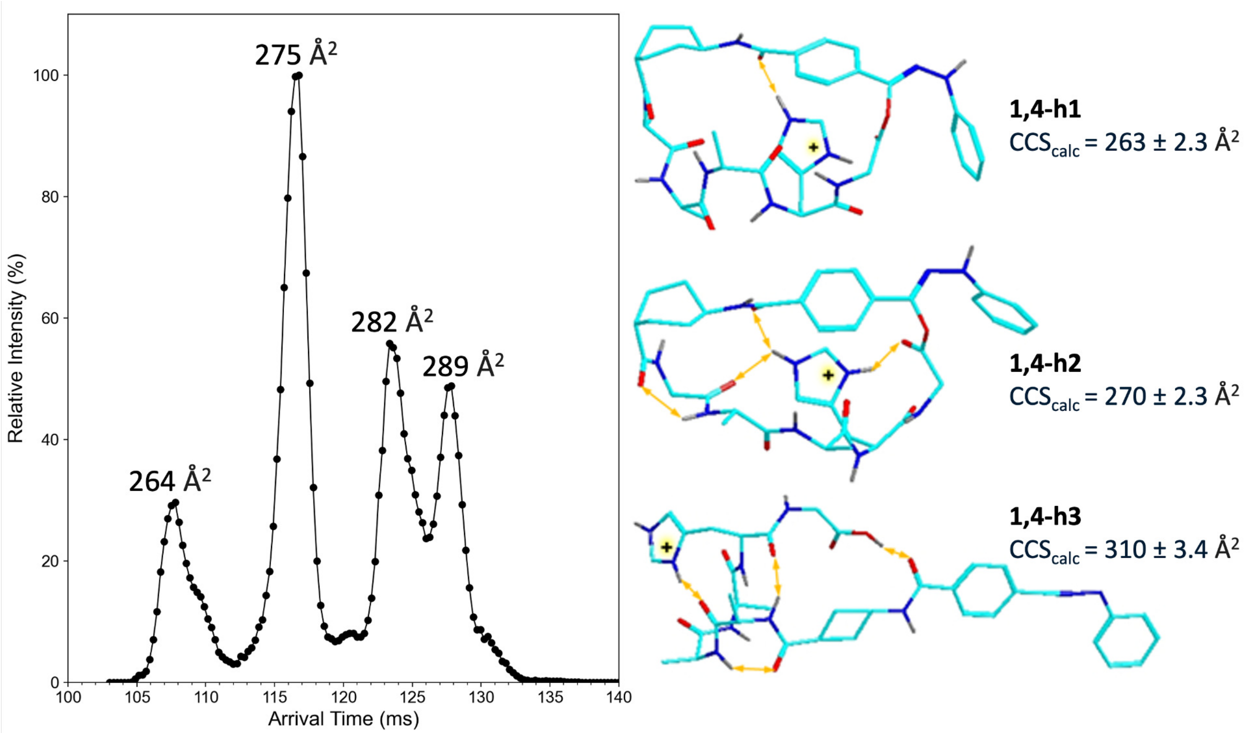


Figure 4.22: Arrival time distributions of (1,4-*s*-AAAHG + H – N_2)$^+$ after 4 passes. M06-2X/6-31+G(d,p) optimized structures and CCS_{calc} of low energy **1,4-h1**-**1,4-h3** isomers. Structure discription as in Figure 4.10.

4.4 Conclusions

Photodissociation of (1,2-s-AAAG + H)$^+$, (1,4-s-AAAG + H)$^+$, (1,2-s-AAAHG + H)$^+$, and (1,4-s-AAAHG + H)$^+$ predominantly led to crosslinking via the carboxyl group, resulting in the formation of diarylhydrazide enol esters (Scheme 4.2 and 4.3). The critical role of the carboxyl group was confirmed by the sharp decrease of crosslinking yields after carboxyl methylation. Similar crosslinking behavior has been reported in solution, where lysine-anchored nitrile imines react with Glu side-chain carboxyl groups in proteins.[51][52]

Experimental IMS data indicated that *cis*-1,2-s-peptide scaffold ions form as mixtures of closely related conformers within the same protonation isomer in the gas phase. In these ions, the cyclohexane ring typically adopts a chair conformation with the peptide chain in an equatorial position. Protonation sites varied with different peptide composition: the benzamide carbonyl was protonated in (1,2-s-AAAG + H)$^+$, whereas basic residues were protonated in (1,2-s-AAAHG + H)$^+$. Additionally, hydrogen bonds frequently formed between the benzamide carbonyl and charged groups within the peptide. In contrast, *trans*-1,4-s-peptide scaffold ions favored a single dominant conformation involving a twist-boat cyclohexane ring, stabilized by strong hydrogen bonding within the peptide and across the ring, thus compensating for the energy required to convert from chair to twist-boat conformations.

Photodissociation and CID of these ions produced a mixture of isomers and conformers, including both crosslinked and linear structures, likely indicating the presence of surviving nitrile imine intermediates. As supported by the absence of crosslinking in sodiated peptide conjugates,[22] crosslinking involves the attack of the nitrile imine group following proton transfer to the imine nitrogen. Notably, the tetrazole nitrogens in precursor ions remained unprotonated, with protonation sites distant from the nitrile imine. This suggests that proton transfer required conformational changes within the peptide, driven by photon absorption or collision energy. This process aligns with the relatively low threshold energies observed for N_2 expulsion, (e.g., 51 kJ mol^{-1} in Scheme 4.3). The requirement for proton transfer and associated conformational adjustments emphasizes that the crosslinking process probes the dynamic nature of peptide ions rather than their static structures. This dynamic behavior was particularly evident in the *trans*-1,4-cyclohexane scaffolds, which likely underwent a chair-to-twist-boat transition during crosslinking.

4.5 Bibliography

[1] Piersimoni, L.; Kastritis, P. L.; Arlt, C.; Sinz, A. “Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions-A Method for All Seasons.” *Chem. Rev.* **2022**, *122*, 7500-7531.

[2] Knowles, J. R. “Photogenerated Reagents for Biological Receptor-Site Labeling.” *Acc. Chem. Res.* **1972**, *5*, 155–160.

[3] Smith, R. A. G.; Knowles, J. R. “Aryldiazirines: Potential Reagents for Photolabeling of Biological Receptor Sites.” *J. Am. Chem. Soc.* **1973**, *95*, 5072–5073.

[4] Das, J. “Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments.” *Chem. Rev.* **2011**, *111*, 4405-4417.

[5] Kauer, J. C.; Erickson-Viitanen, S.; Wolfe, H. R., Jr.; DeGrado, W. W. “P-benzoyl-LPhenylalanine, a New Photoreactive Amino Acid. Photolabeling of Calmodulin with a Synthetic Calmodulin-Binding Peptide.” *J. Biol. Chem.* **1986**, *261*, 10695-10700.

[6] Dorman, G.; Prestwich, G. D. “Benzophenone Photophores in Biochemistry.” *Biochemistry* **1994**, *33*, 5661-5673.

[7] Lee, H. S.; Dimla, R. D.; Schultz, P. G. “Protein-DNA Photo-Crosslinking with a Genetically Encoded Benzophenone-Containing Amino Acid.” *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5222-5224.

[8] Lomant, A. J.; Fairbanks, G. “Chemical Probes of Extended Biological Structures: Synthesis and Properties of the Cleavable Protein Cross-Linking Reagent [35S]Dithiobis(Succinimidyl Propionate).” *J. Mol. Biol.* **1976**, *104*, 243-261.

[9] Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. “Distance Restraints from Crosslinking Mass Spectrometry: Mining a Molecular Dynamics Simulation Database to Evaluate Lysine-Lysine Distances.” *Protein Sci.* **2014**, *23*, 747-759.

[10] Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. "Mass Spectrometry Identifiable Cross- Linking Strategy for Studying Protein-Protein Interactions." *Anal. Chem.* **2005**, *77*, 311-318.

[11] Soderblom, E. J.; Goshe, M. B. "Collision-Induced Dissociative Chemical Cross-Linking Reagents and Methodology: Applications to Protein Structural Characterization Using Tandem Mass Spectrometry Analysis." *Anal. Chem.* **2006**, *78*, 8059-8068.

[12] Kao, A.; Chiu, C.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. "Development of A Novel Cross-Linking Strategy for Fast and Accurate Identification of Cross-Linked Peptides of Protein Complexes." *Mol. Cell. Proteomics* **2011**, *10*, M110.002170.

[13] Petrotchenko, E. V.; Serpa, J. J.; Borchers, C. H. "An Isotopically Coded CID-Cleavable Biotinylated Cross-Linker for Structural Proteomics." *Mol. Cell. Proteomics* **2011**, *10*, M110.001420.

[14] Clifford-Nunn, B.; Showalter, H. D.; Andrews, P. C. "Quaternary Diamines as Mass Spectrometry Cleavable Crosslinkers for Protein Interactions." *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 201-212.

[15] Stadlmeier, M.; Runtsch, L. S.; Streshnev, F.; Wuhr, M.; Carell, T. "A Click-Chemistry-Based Enrichable Crosslinker for Structural and Protein Interaction Analysis by Mass Spectrometry." *ChemBioChem* **2020**, *21*, 103-107.

[16] Nury, C.; Redeker, V.; Dautrey, S.; Romieu, A.; van der Rest, G.; Renard, P. Y.; Melki, R.; Chamot-Rooke, J. "A Novel Bio-Orthogonal Cross-Linker for Improved Protein/Protein Interaction Analysis." *Anal. Chem.* **2015**, *87*, 1853-1860.

[17] Tureček, F. "Covalent Crosslinking in Gas-Phase Biomolecular Ions. An Account and Perspective." *Phys. Chem. Chem. Phys.* **2023**, *25*, 32292-32304.

[18] Liu, Y.; Ramey, Z.; Tureček, F. "Non-Covalent Interactions of a Neuroprotective Peptide Revealed by Photodissociative Cross-Linking in the Gas Phase." *Chem. Eur. J.* **2018**, *24*, 9259- 9263.

[19] Liu, Y.; Tureček, F. "Photodissociative Crosslinking of Diazirine-Tagged Peptides with DNA Dinucleotides in the Gas Phase." *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1992-2006.

[20] Shaffer, C. J.; Andrikopoulos, P. C.; Řezáč, J.; Rulíšek, L.; Tureček, F. "Efficient Covalent Bond Formation in Gas-Phase Peptide-Peptide Ion Complexes with the Photoleucine Stapler." *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 6333-645.

[21] Nguyen, H. T. H.; Andrikopoulos, P. C.; Rulíšek, L.; Shaffer, C. J.; Tureček, F. "Photodissociative Cross Linking of Noncovalent Peptide-Peptide Ion Complexes in the Gas Phase." *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1706-1720.

[22] Wan, J; Nytka, M.; Vu, K.; Qian, H.; Lemr, K.; Tureček, F. "Nitrile Imines as Peptide and Oligonucleotide Photocrosslinkers in Gas-Phase Ions." *J. Am. Soc. Mass Spectrom.* **2024**, *35*, 344-356.

[23] Song, W.; Wang, Y.; Qu, J.; Madden, M. M.; Lin Q. "A Photoinducible 1,3-Dipolar Cycloaddition Reaction for Rapid, Selective Modification of Tetrazole-Containing Proteins." *Angew. Chem.* **2008**, *120*, 2874-2877.

[24] Marshall, D. L.; Menzel, J. P.; McKinon, B. I.; Blinco, J. P.; Trevitt, A. J.; Barner-Kowolik, C.; Blanksby, S. J. "Laser Photo-dissociation Action Spectroscopy for the Wavelength-Dependent Evaluation of Photoligation Reactions." *Anal. Chem.* **2021**, *93*, 8091-8098.

[25] Zhu, H.; Zima, V.; Ding, E.; Tureček, F. "Carbene Crosslinking in Gas-Phase Peptide Ion Scaffolds." *J. Am. Soc. Mass Spectrom.* **2023**, *34*, 763-774.

[26] Marek, A.; Nguyen, H. T. H.; Brož, B.; Tureček, F. "Stereospecific Control of Peptide Gas- Phase Ion Chemistry with cis and trans Cyclo Ornithine Residues." *J. Mass Spectrom.* **2018**, *53*, 124-137.

[27] Sharp, J. T. Nitrile Ylides and Nitrile Imines. In *Chemistry of Heterocyclic Compounds 59: Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products;* Padwa, A., Pearson, W. H., Eds.; John Wiley & Sons: New York, 2002.

[28] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. "Molecular Dynamics with Coupling to an External Bath." *J. Chem. Phys.* **1984**, *81*, 3684-3690.

[29] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. "Semiempirical Quantum Chemical PM6Method Aug-

mented by Dispersion and H Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes." *J. Chem. Theory Comput.* **2009**, *5*, 1749-1760.

[30] Řezáč, J. "Cuby: An Integrative Framework for Computational Chemistry." *J. Comput. Chem.* **2016**, *37*, 1230-1237.

[31] Stewart, J. J. P. *MOPAC 16*; Stewart Computational Chemistry: Colorado Springs, CO, **2016**.

[32] Becke, A. D. "Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior." *Phys. Rev. A* **1988**, *38*, 3098-3100.

[33] Grimme, S.; Ehrlich, S.; Goerigk, L. "Effect of the Damping Function in Dispersion Corrected Density Functional Theory." *J. Comput. Chem.* **2011**, *32*, 1456-1465.

[34] Nickerson, C. J.; Bryenton, K. R.; Price, A. J. A.; Johnson, E. R. "Comparison of Density-Functional Theory Dispersion Corrections for the DES15K Database." *J. Phys. Chem. A* **2023**, *127*, 8712-8722.

[35] Zhao, Y.; Truhlar, D. G. "The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals." *Theor. Chem. Acc.* **2008**, *120*, 215-241.

[36] Weigend, F. "Accurate Coulomb-Fitting Basis Sets for H to Rn." *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057-1065.

[37] Gray, M.; Bowling, P. E.; Herbert, J. M. "Comment on "Benchmarking Basis Sets for Density Functional Theory Thermochemistry Calculations: Why Unpolarized Basis Sets and the Polarized 6-311G Family Should Be Avoided"." *J. Phys. Chem. A* **2024**, *128*, 7739-7745.

[38] Singh, U. C.; Kollman, P. A. "An Approach to Computing Electrostatic Charges for Molecules." *J. Comput. Chem.* **1984**, *5*, 129-145.

[39] Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. "Atomic Charges Derived from Semiempirical Methods." *J. Comput. Chem.* **1990**, *11*, 431-439.

[40] Ieritano, C.; Crouse, J.; Campbell, J. L.; Hopkins, W. S. “A Parallelized Molecular Collision Cross Section Package with Optimized Accuracy and Efficiency.” *Analyst* **2019**, *144*, 1660-1670.

[41] Ieritano, C.; Hopkins, W. S. “Assessing Collision Cross Section Calculations Usi ng MobCal-MPI with a Variety of Commonly Used Computational Methods.” *Mater. Today Commun.* **2021**, *27*, No. 102226.

[42] Halgren, T. A. “Merck Olecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMFF94.” *J. Comput. Chem.* **1996**, *17*, 490-519.

[43] Chu, I. K.; Siu, C.-K.; Lau, J. K.-C.; Tang, W. K.; Mu, X.; Lai, C. K.; Guo, X.; Wang, X.; Li, N.; Yao, Z.; Xia, Y.; Kong, X.; Oh, H.-B.; Ryzhov, V.; Tureček, F.; Hopkinson, A. C.; Siu, K. W. M. “Proposed Nomenclature for Peptide Ion Fragmentation.” *Int. J. Mass Spectrom.* **2015**, *390*, 24-27.

[44] Bleiholder, C.; Osbum, S.; Williams, T. D.; Suhai, S.; Van Stipdonk, M.; Harrison, A. G.; Paizs, B. “Sequence-Scrambling Fragmentation Pathways of Protonated Peptides.” *J. Am. Chem. Soc.* **2008**, *130*, 17774-17789.

[45] Novák, J.; Lemr, K.; Schug, K. A.; Havlíček, V. “CycloBranch: De Novo Sequencing of Nonribosomal Peptides from Accurate Product Ion Mass Spectra.” *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1780-1786.

[46] Townsend, C.; Furukawa, A.; Schwokert, J.; Pye, C. R.; Edmonson, Q.; Lokey, R. S. “CyCLS: Accurate, Whole-Library Sequencing of Cyclic Peptides Using Tandem Mass Spectrometry.” *Bioorg. Med. Chem.* **2018**, *26*, 1232-1238.

[47] Tureček, F. “Covalent Crosslinking in Gas-Phase Biomolecular Ions. An Account and Perspective.” *Phys. Chem. Chem. Phys.* **2023**, *25*, 32292-32304.

[48] Ballard, K. D.; Gaskell, S. J. “Sequential Mass Spectrometry Applied to the Study of the Formation of “Internal” Fragment Ions of Protonated Peptides.” *Int. J. Mass Spectrom. Ion Process.* **1991**, *111*, 173-189.

[49] Thorne, G. C.; Ballard, K. D.; Gaskell, S. J. “Metastable Decomposition of Peptide $[M + H]^+$ Ions via Rearrangement Involving Loss of C-Terminal Amino Acid Residue.” *J. Am. Soc. Mass Spectrom.* **1990**, *1*, 249-257.

[50] Ballard, K. D.; Gaskell, S. J. "Intramolecular Oxygen-18 Isotopic Exchange in the Gas Phase Observed during the Tandem Mass Spectrometric Analysis of Peptides." *J. Am. Chem. Soc.* **1992**, *114*, 64-71.

[51] Herner, A.; Marjanovic, J.; Lewandowski, T. M.; Marin, V.; Patterson, M.; Miesbauer, L.; Ready, D.; Williams, J.; Vasudevan, A.; Lin, Q. "2-Aryl-5-Carboxytetrazole as a New Photoaffinity Label for Drug Target Identification." *J. Am. Chem. Soc.* **2016**, *138*, 14609-14615.

[52] Tian, Y.; Jacinto, M. P.; Zeng, Y.; Yu, Z.; Qu, J.; Liu, W. R.; Lin, Q. "Genetically Encoded 2- Aryl-5-Carboxytetrazoles for Site-Selective Protein Photo-Cross-Linking." *J. Am. Chem. Soc.* **2017**, *139*, 6078-6081.

Chapter 5

Nitrile Imine Crosslinking in Stereochemically Distinct Scaffolds of Peptides with Basic Residues at the C-Terminus in the Gas-Phase

Reproduced in part with permission from Hongyi Zhu, Marianna Nytka, Tuan Ngoc Kim Vu, Karel Lemr, and František Tureček. Photochemical and Collision-Induced Cross-Linking in Stereochemically Distinct Scaffolds of Peptides and Nitrile Imines in Gas-Phase Ions. *Journal of the American Society for Mass Spectrometry*, **XXXX**, *XXX*, XXX-XXX.

Abstract. Intramolecular crosslinking between peptides protonated at C-terminal lysine or arginine residues and nitrile imine intermediates in stereochemically distinct conjugates is investigated. Nitrile imine intermediates, generated by N_2 loss from 2,5-diaryltetrazole tags upon UV photodissociation at 213 nm or by CID, were analyzed by CID and UVPD-MS^3. Crosslinking yields ranged from 28% to 78% in UVPD-MS^2 and increased to 62 to 97% in UVPD-CID-MS^3. Structural analysis of the *cis* and *trans* peptide ions, intermediates of the nitrile imine intermediates, and crosslinks was conducted using cyclic ion mobility with Born-Oppenheimer molecular dynamics and density functional theory calculations. Experimental and theoretical matches in collision cross sections and Gibbs energies were used to confirm structural assign-

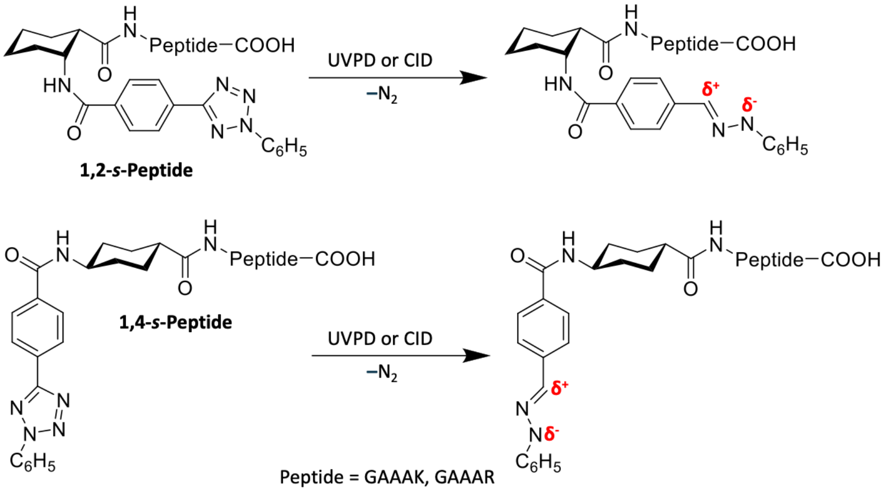
ments. Peptide conjugates with C-terminal lysine underwent crosslinking primarily via the carboxyl group, as confirmed by MS^3 sequencing and supported by carboxyl-blocking experiments that reduced crosslinking yields. Peptide conjugates with C-terminal arginine also showed crosslinking via the guanidine group on the side chain.

5.1 Introduction

In Chapter 4, the results from the AAAG and AAAHG scaffolds demonstrated that crosslinking predominantly involved the C-terminal residue. This conclusion was further supported by C-terminus blocking experiments, which showed a significant decrease in crosslinking yields after converting the Gly carboxyl into a methyl ester, suggesting that the proton at the Gly carboxyl is essential for the crosslinking reaction. To investigate the influence of basic side-chain groups and to better understand the role of proton involvement, we introduced lysine and arginine at the C-terminus to replace glycine. This modification was intended to explore the competition between the proton on the basic residue and the carboxyl group during the crosslinking process.

Lysine and arginine are two positively charged amino acids with high aqueous pKa values (∼10.5 for lysine[1] and ∼13.8 for arginine[2]) that play crucial roles in protein structure and function, primarily through electrostatic interactions and protein solvation. The high pKa of arginine results from the delocalization of the positive charge across the π-bonded system of the side chain's guanidinium ion, which remains protonated even when partially embedded within protein structures or membranes.[3][4] In contrast, the charge on lysine is concentrated on the terminal aliphatic amino group, making its side chain more susceptible to deprotonation within proteins.

We synthesized peptide scaffolds designated as 1,2-s-GAAAK, 1,4-s-GAAAK, 1,2-s-GAAAR, and 1,4-s-GAAAR, where the charge-carrying side chain could donate a proton to catalyze crosslinking and serve as a nucleophile to attack the nitrile imine dipole. The crosslinking behavior of these scaffolds was investigated using UVPD-CID-MS^3 and CID-UVPD-MS^3 at 213 nm, with accurate mass measurements aiding in fragment ion identification. Structural analysis of the conjugates and their dissociation products was performed using high-resolution cyclic ion mobility mass spectrometry, in combination with Born-Oppenheimer molecular dynamics (BOMD) and density functional theory (DFT) calculations to determine



Scheme 5.1: Peptide-Nitrile-Imine Conjugates with Stereochemically Distinct *cis*-1,2- and *trans*-1,4-Cyclohexane Scaffolds.

theoretical collision cross sections (CCS_{calc}), allowing for comparison with experimental data.

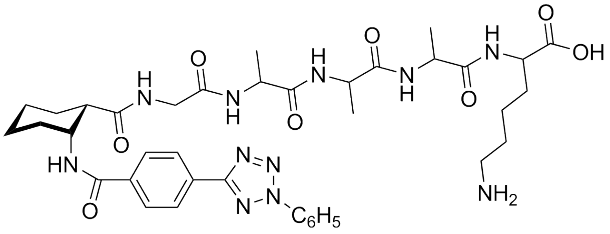
5.2 Experimental Section

5.2.1 Materials and Methods

Unless otherwise noted, all starting materials were from (Sigma Aldrich, Fisher Chemical) and were used as received. ^{1}H NMR spectra were recorded on Bruker spectrometers at 300 MHz. Coupling constants (J) are given in Hz and chemical shifts are on the δ-scale. HRMS spectra were recorded on a Thermo Orbitrap Ascend Tribrid Mass Spectrometer. Samples were introduced by flow injection into the ESI source.

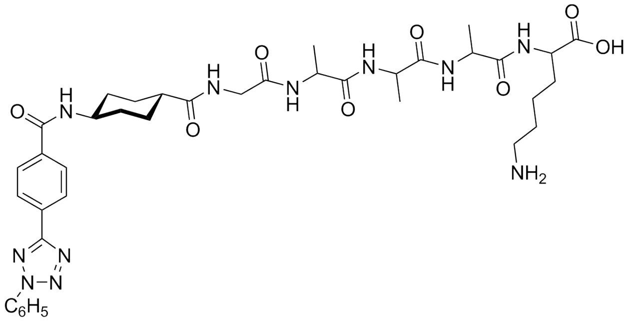
5.2.2 Synthetic Procedures

1,2-*s*-GAAAK



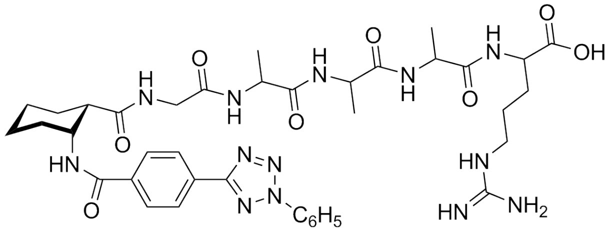
The title conjugate was prepared according to the general procedure for solid phase peptide synthesis (see Chapter 4.2.2). HRMS (ESI) *m/z* calcd for $C_{38}H_{52}N_{11}O_8$ (M+H)$^+$ 790.3995, found 790.3978.

1,4-*s*-GAAAK



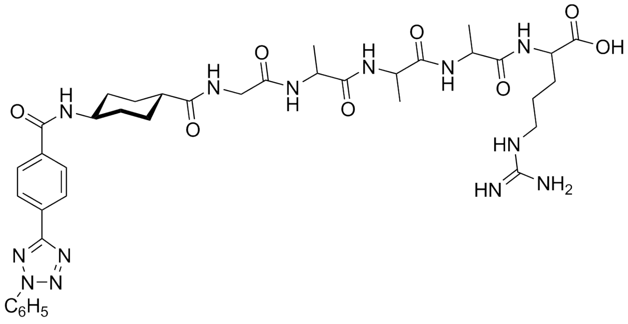
The title conjugate was prepared according to the general procedure for solid phase peptide synthesis (see Chapter 4.2.2). HRMS (ESI) *m/z* calcd for $C_{38}H_{52}N_{11}O_8$ $(M+H)^+$ 790.3995, found 790.3976.

1,2-*s*-GAAAR



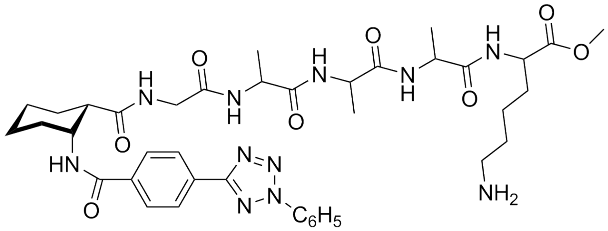
The title conjugate was prepared according to the general procedure for solid phase peptide synthesis (see Chapter 4.2.2). HRMS (ESI) *m/z* calcd for $C_{38}H_{52}N_{13}O_8$ $(M+H)^+$ 818.4056, found 818.4039.

1,4-*s*-GAAAR



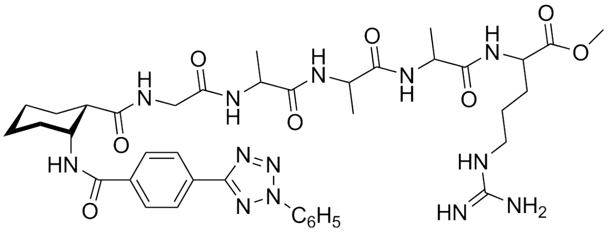
The title conjugate was prepared according to the general procedure for solid phase peptide synthesis (see Chapter 4.2.2). HRMS (ESI) *m/z* calcd for $C_{38}H_{52}N_{13}O_8$ $(M+H)^+$ 818.4056, found 818.4040.

1,2-s-GAAAK-OCH$_3$



The title conjugate was prepared according to the general procedure for methylation (see Chapter 4.2.2). HRMS (ESI) *m/z* calcd for $C_{39}H_{54}N_{11}O_8$ $(M+H)^+$ 804.4151, found 804.4135.

1,2-s-GAAAR-OCH$_3$



The title conjugate was prepared according to the general procedure for methylation (see Chapter 4.2.2). HRMS (ESI) *m/z* calcd for $C_{39}H_{54}N_{13}O_8$ $(M+H)^+$ 832.4128, found 832.4191.

5.2.3 Cyclic Traveling Wave Ion Mobility-Mass Spectrometry

5.2.3.1 Sample Preparation

LC-MS grade acetonitrile (ACN) and water were purchased from Fisher Chemical (Fisher Scientific, United Kingdom). Acetic acid $\geq$ 99% was obtained from Sigma Aldrich (Prague, Czech Republic). Poly-DL-alanine was purchased from Sigma Aldrich (Prague, Czech Republic).

Solid samples were dissolved in ACN/H_2O (50/50) to prepare stock solutions with the following concentrations: 3.06 mg/mL for 1,2-s-GAAAK; 2 mg/mL for 1,4-s-GAAAK; 3.1 mg/mL for 1,2-s-GAAAR; 4 mg/mL for 1,4-s-GAAAR.

The stock solutions were diluted to 1 mg/mL in ACN/H_2O (50:50) and then diluted to a working concentration of 1 μg/mL in ACN/H_2O (1:1) with 1% acetic acid. The stock solution of polyalanine (1 mg/ml in ACN/H_2O (50/50)) was diluted to 2 μg/mL in ACN/H_2O (50/50) with 0.1% of formic acid. The calibrant was prepared in duplicate (one data acquisition for each ion).

5.2.3.2 ESI Conditions

The instrument parameters were the same as Chapter 4.2.3. cIMS conditions for 1 Pass separation experiments of $[M + H]^+$ and $[M + H - N_2]^+$ are listed in Table 4.1. The CCS calibration is listed in Chapter 4.2.3.

5.2.4 Accurate Mass Measurements and Assignments

Table 5.1: Separation Time for Multipass Measurements for Different Analytes

Precursor Ion	**Number of Passes**	**Separation Time (ms)**
1,2-*s*-GAAAK $(M + H)^+$ (*m/z* 790.4)	10	270
1,2-*s*-GAAAK $(M + H - N_2)^+$ (*m/z* 762.4)	20	508
1,4-*s*-GAAAK $(M + H)^+$ (*m/z* 790.4)	10	300
1,4-*s*-GAAAK $(M + H - N_2)^+$ (*m/z* 762.4)	4	93
1,2-*s*-GAAAR $(M + H)^+$ (*m/z* 818.4)	7	180
1,2-*s*-GAAAR $(M + H - N_2)^+$ (*m/z* 790.4)	15	374
1,4-*s*-GAAAR $(M + H)^+$ (*m/z* 818.4)	10	300
1,4-*s*-GAAAR $(M + H - N_2)^+$ (*m/z* 790.4)	4	95

Table 5.2: Accurate Mass Measurements of UVPD-MS^2 of (1,2-*s*-GAAAK + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
790.3957	$C_{38}H_{52}N_{11}O_8$	-3.8		
762.3899	$C_{38}H_{52}N_9O_8$	-3.4	N_2	
744.3795	$C_{38}H_{50}N_9O_7$	-3.3	H_2N_2O	$-N_2-H_2O$
671.3480	$C_{32}H_{47}N_8O_8$	-3.1	$C_6H_5N_3$	$-N_2-C_6H_5N$
634.2956	$C_{32}H_{40}N_7O_7$	-2.8	$C_6H_{12}N_4O$	–N_2–[Lys]
563.2588	$C_{29}H_{35}N_6O_6$	-2.5	$C_9H_{17}N_5O_2$	–N_2–[AlaLys]
540.3117	$C_{24}H_{42}N_7O_7$	-2.3	$C_{14}H_{10}N_4O$	
492.2221	$C_{26}H_{30}N_5O_5$	-2.0	$C_{12}H_{22}N_6O_3$	–N_2–[AlaAlaLys]

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
454.2066	$C_{23}H_{28}N_5O_5$	-1.9	$C_{15}H_{24}N_6O_3$	*m/z* 671–LysOH–[Ala]
421.1853	$C_{23}H_{25}N_4O_4$	-1.7	$C_{15}H_{27}N_7O_4$	–N_2–[AlaAlaAlaLys]
417.2440	$C_{17}H_{33}N_6O_6$	-1.6	$C_{21}H_{19}N_5O_4$	y_5
383.1699	$C_{20}H_{23}N_4O_4$	-1.5	$C_{18}H_{29}N_7O_4$	*m/z* 454–[Ala]
364.1641	$C_{21}H_{22}N_3O_3$	-1.5	$C_{17}H_{30}N_8O_5$	–N_2–[GlyAlaAlaAlaLys]
346.1537	$C_{21}H_{20}N_3O_2$	-1.3	$C_{17}H_{32}N_8O_6$	*m/z* 364–H_2O
312.1330	$C_{17}H_{18}N_3O_3$	-1.3	$C_{21}H_{34}N_8O_5$	*m/z* 454–[AlaAla]
289.1860	$C_{12}H_{25}N_4O_4$	-1.0	$C_{26}H_{27}N_7O_4$	y_3
271.1391	$C_{11}H_{19}N_4O_4$	-1.0	$C_{27}H_{33}N_7O_4$	
256.1071	$C_{14}H_{14}N_3O_2$	-1.0	$C_{24}H_{38}N_8O_6$	
255.1119	$C_{15}H_{15}N_2O_2$	-0.9	$C_{23}H_{37}N_9O_6$	*m/z* 454–[GlyAlaAla]
243.1443	$C_{10}H_{19}N_4O_3$	-0.9	$C_{28}H_{33}N_7O_5$	
239.0806	$C_{14}H_{11}N_2O_2$	-0.9	$C_{24}H_{41}N_9O_6$	
229.1089	$C_{14}H_{15}NO_2$	-0.8	$C_{24}H_{37}N_{10}O_6$	
221.0701	$C_{14}H_9N_2O$	-0.8	$C_{24}H_{43}N_9O_7$	*m/z* 346–$C_7H_{11}NO$

Table 5.3: Accurate Mass Measurements of CID-MS^2 of (1,2-s-GAAAK + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
790.3978	$C_{38}H_{52}N_{11}O_8$	-1.7		
762.3915	$C_{38}H_{52}N_9O_8$	-1.8	N_2	
744.3817	$C_{38}H_{50}N_9O_7$	-1.1	H_2N_2O	–N_2–H_2O
656.3392	$C_{32}H_{46}N_7O_8$	-1.0	$C_6H_6N_4$	–N_2–phenyldiazene
634.2973	$C_{32}H_{40}N_7O_7$	-1.1	$C_6H_{12}N_4O$	–N_2–[Lys]

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
563.2601	$C_{29}H_{35}N_6O_6$	-1.2	$C_9H_{17}N_5O_2$	–N_2–[AlaLys]
492.2232	$C_{26}H_{30}N_5O_5$	-1.0	$C_{12}H_{22}N_6O_3$	–N_2–[AlaAlaLys]
421.1864	$C_{23}H_{25}N_4O_4$	-0.6	$C_{15}H_{27}N_7O_4$	–N_2–[AlaAlaAlaLys]
364.1650	$C_{21}H_{22}N_3O_3$	-0.6	$C_{17}H_{30}N_8O_5$	–N_2–[GlyAlaAlaAlaLys]

Table 5.4: Accurate Mass Measurements of CID-MS^3 of ion (1,2-s-GAAAK + H –N_2)$^+$ Generated by CID of (1,2-s-GAAAK + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
762.3918	$C_{38}H_{52}N_9O_8$	-1.5		
744.3814	$C_{38}H_{50}N_9O_7$	-1.4	H_2O	
734.3970	$C_{37}H_{52}N_9O_7$	-1.4	CO	
717.3705	$C_{37}H_{49}N_8O_7$	-1.4	CH_3NO	–CO–NH_3
691.3550	$C_{35}H_{47}N_8O_7$	-1.2	C_3H_5NO	–[Ala]
671.3499	$C_{32}H_{47}N_8O_8$	-1.2	C_6H_5N	
654.3233	$C_{32}H_{44}N_7O_8$	-1.3	$C_6H_8N_2$	–phenylhydrazine
634.2972	$C_{32}H_{40}N_7O_7$	-1.2	$C_6H_{12}N_2O$	–[Lys] (β_5)
626.3285	$C_{31}H_{44}N_7O_7$	-1.2	$C_7H_8N_2O$	*m/z* 717–C_6H_5N
616.2867	$C_{32}H_{38}N_7O_6$	-1.1	$C_6H_{14}N_2O_2$	b_5
583.2864	$C_{29}H_{39}N_6O_7$	-1.1	$C_9H_{13}N_3O$	–phenylhydrazine–[Ala]
563.2601	$C_{29}H_{35}N_6O_6$	-1.2	$C_9H_{17}N_3O_2$	–[AlaLys] (β_4)
545.2497	$C_{29}H_{33}N_6O_5$	-1.0	$C_9H_{19}N_3O_3$	b_4
538.2650	$C_{28}H_{36}N_5O_6$	-1.0	$C_{10}H_{16}N_4O_2$	*m/z* 583–CO–NH_3
525.2446	$C_{26}H_{33}N_6O_6$	-1.0	$C_{12}H_{19}N_3O_2$	b_5–C_6H_5N

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
492.2232	$C_{26}H_{30}N_5O_5$	-0.9	$C_{12}H_{22}N_4O_3$	–[AlaAlaLys] (β_3)
474.2127	$C_{26}H_{28}N_5O_4$	-0.9	$C_{12}H_{24}N_4O_4$	b_3
467.2281	$C_{25}H_{31}N_4O_5$	-0.8	$C_{13}H_{21}N_5O_3$	*m/z* 538–[Ala]
454.2077	$C_{23}H_{28}N_5O_5$	-0.8	$C_{15}H_{24}N_4O_3$	b_4–C_6H_5N
421.1863	$C_{23}H_{25}N_4O_4$	-0.7	$C_{15}H_{27}N_5O_4$	–[AlaAlaAlaLys] (β_2)
403.1758	$C_{23}H_{23}N_4O_3$	-0.7	$C_{15}H_{29}N_5O_5$	b_2
383.1708	$C_{20}H_{23}N_4O_4$	-0.6	$C_{18}H_{29}N_5O_4$	b_3–C_6H_5N
364.1650	$C_{21}H_{22}N_3O_3$	-0.6	$C_{17}H_{30}N_6O_5$	–[GlyAlaAlaAlaLys] (β_1)
346.1545	$C_{21}H_{20}N_3O_2$	-0.5	$C_{17}H_{32}N_6O_6$	b_1
312.1338	$C_{17}H_{18}N_3O_3$	-0.5	$C_{21}H_{34}N_6O_5$	b_2–C_6H_5N

Table 5.5: Accurate Mass Measurements of CID-MS^3 of ion (1,2-s-GAAAK + H –N_2)$^+$ Generated by UVPD of (1,2-s-GAAAK + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
762.3898	$C_{38}H_{52}N_9O_8$	-3.5		
744.3795	$C_{38}H_{50}N_9O_7$	-3.3	H_2O	
734.3951	$C_{37}H_{52}N_9O_7$	-3.3	CO	
654.3216	$C_{32}H_{44}N_7O_8$	-3.0	$C_6H_8N_2$	–phenylhydrazine
634.2956	$C_{32}H_{40}N_7O_7$	-2.8	$C_6H_{12}N_2O$	–[Lys] (β_5)
563.2586	$C_{29}H_{35}N_6O_6$	-2.7	$C_9H_{17}N_3O_2$	–[AlaLys] (β_4)
492.2220	$C_{26}H_{30}N_5O_5$	-2.1	$C_{12}H_{22}N_4O_3$	–[AlaAlaLys] (β_3)
454.2065	$C_{23}H_{28}N_5O_5$	-2.0	$C_{15}H_{24}N_4O_3$	b_4–C_6H_5N
421.1852	$C_{23}H_{25}N_4O_4$	-1.8	$C_{15}H_{27}N_5O_4$	–[AlaAlaAlaLys] (β_2)

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
383.1698	$C_{20}H_{23}N_4O_4$	-1.6	$C_{18}H_{29}N_5O_4$	b_3–C_6H_5N
364.1640	$C_{21}H_{22}N_3O_3$	-1.6	$C_{17}H_{30}N_6O_5$	–[GlyAlaAlaAlaLys] (β_1)
312.1330	$C_{17}H_{18}N_3O_3$	-1.3	$C_{21}H_{34}N_6O_5$	b_2–C_6H_5N

Table 5.6: Accurate Mass Measurements of CID-MS^2 of (1,2-s-GAAAK-OCH$_3$ + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
804.4135	$C_{39}H_{54}N_{11}O_8$	-1.6		
776.4066	$C_{39}H_{54}N_9O_8$	-2.4	N_2	
670.3552	$C_{33}H_{48}N_7O_8$	-0.7	$C_6H_6N_4$	–N_2–phenyldiazene

Table 5.7: Accurate Mass Measurements of UVPD-MS^2 of (1,2-s-GAAAK-OCH$_3$ + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
804.4128	$C_{39}H_{54}N_{11}O_8$	-2.3		
776.4064	$C_{39}H_{54}N_9O_8$	-2.6	N_2	
685.3649	$C_{33}H_{49}N_8O_8$	-1.9	$C_6H_5N_3$	–N_2–C_6H_5N
670.3543	$C_{33}H_{48}N_7O_8$	-1.6	$C_6H_6N_4$	–N_2–phenyldiazene
668.3388	$C_{33}H_{46}N_7O_8$	-1.4	$C_6H_8N_4$	–N_2–phenylhydrazine
616.2864	$C_{32}H_{38}N_7O_6$	-1.4	$C_7H_{16}N_4O_2$	b_5–N_2
525.2444	$C_{26}H_{33}N_6O_6$	-1.2	$C_{13}H_{21}N_5O_2$	b_5–N_2–C_6H_5N
474.2125	$C_{26}H_{28}N_5O_4$	-1.1	$C_{13}H_{26}N_6O_4$	b_5–N_2–phenylhydrazine
454.2073	$C_{23}H_{28}N_5O_5$	-1.2	$C_{16}H_{26}N_6O_3$	b_4–N_2–C_6H_5N

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
431.2603	$C_{18}H_{35}N_6O_6$	-1.0	$C_{21}H_{19}N_5O_2$	y_5
403.1756	$C_{23}H_{23}N_4O_3$	-0.9	$C_{16}H_{31}N_7O_5$	b_4–N_2–phenylhydrazine
383.1705	$C_{20}H_{23}N_4O_4$	-0.9	$C_{19}H_{31}N_7O_4$	b_3–N_2–C_6H_5N
346.1543	$C_{21}H_{20}N_3O_2$	-0.7	$C_{18}H_{34}N_8O_6$	*m/z* 403–[Gly]
312.1336	$C_{17}H_{18}N_3O_3$	-0.7	$C_{22}H_{36}N_8O_5$	b_2–N_2–C_6H_5N
303.2021	$C_{13}H_{27}N_4O_4$	-0.6	$C_{26}H_{27}N_7O_4$	y_3
271.1396	$C_{11}H_{19}N_4O_4$	-0.5	$C_{28}H_{35}N_7O_4$	
255.1123	$C_{15}H_{15}N_2O_2$	-0.5	$C_{24}H_{39}N_9O_6$	*m/z* 312–[Gly]
232.1651	$C_{10}H_{22}N_3O_3$	-0.5	$C_{29}H_{32}N_8O_5$	y_2

Table 5.8: Accurate Mass Measurements of CID-MS^3 of (1,2-*s*-GAAAK-OCH_3 + H –N_2)$^+$ Generated by CID of (1,2-*s*-GAAAK-OCH_3 + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
776.4074	$C_{39}H_{54}N_9O_8$	-1.6		
758.3972	$C_{39}H_{52}N_9O_7$	-1.2	H_2O	
748.4125	$C_{38}H_{54}N_9O_7$	-1.6	CO	
705.3707	$C_{36}H_{49}N_8O_7$	-1.2	C_3H_5NO	–[Ala]
685.3656	$C_{33}H_{49}N_8O_8$	-1.2	C_6H_5N	
668.3390	$C_{33}H_{46}N_7O_8$	-1.2	$C_6H_8N_2$	–phenylhydrazine
640.3444	$C_{32}H_{46}N_7O_7$	-0.9	$C_7H_8N_2O$	–phenylhydrazine–CO
597.3023	$C_{30}H_{41}N_6O_7$	-0.8	$C_9H_{13}N_3O$	–phenylhydrazine–[Ala]
554.2601	$C_{28}H_{36}N_5O_7$	-0.8	$C_{11}H_{18}N_4O$	
545.2498	$C_{29}H_{33}N_6O_5$	-0.9	$C_{10}H_{21}N_3O_3$	b_4

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
525.2448	$C_{26}H_{33}N_6O_6$	-0.8	$C_{13}H_{21}N_3O_2$	b_5–C_6H_5N
474.2127	$C_{26}H_{28}N_5O_4$	-0.9	$C_{13}H_{26}N_4O_4$	b_3
454.2075	$C_{23}H_{28}N_5O_5$	-1.0	$C_{16}H_{26}N_4O_3$	b_4–C_6H_5N
403.1757	$C_{23}H_{23}N_4O_3$	-0.8	$C_{16}H_{31}N_5O_5$	b_2
383.1706	$C_{20}H_{23}N_4O_4$	-0.8	$C_{19}H_{31}N_5O_4$	b_3–C_6H_5N
346.1544	$C_{21}H_{20}N_3O_2$	-0.6	$C_{18}H_{34}N_6O_6$	b_1
312.1337	$C_{17}H_{18}N_3O_3$	-0.6	$C_{22}H_{36}N_6O_5$	b_2–C_6H_5N
255.1125	$C_{15}H_{15}N_2O_2$	-0.3	$C_{24}H_{39}N_7O_6$	b_1–C_6H_5N

Table 5.9: Accurate Mass Measurements of CID-MS^3 of (1,2-*s*-GAAAK-OCH$_3$ + H – N$_2$)$^+$ Generated by UVPD of (1,2-*s*-GAAAK-OCH$_3$ + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
776.4069	$C_{39}H_{54}N_9O_8$	-2.1		
748.4120	$C_{38}H_{54}N_9O_7$	-2.1	CO	
732.4173	$C_{38}H_{54}N_9O_6$	-1.9	CO_2	
705.3700	$C_{36}H_{49}N_8O_7$	-1.9	C_3H_5NO	–[Ala]
685.3650	$C_{33}H_{49}N_8O_8$	-1.8	C_6H_5N	
683.3494	$C_{33}H_{47}N_8O_8$	-1.7	C_6H_7N	
668.3385	$C_{33}H_{46}N_7O_8$	-1.7	$C_6H_8N_2$	–phenylhydrazine
640.3438	$C_{32}H_{46}N_7O_7$	-1.5	$C_7H_8N_2O$	–phenylhydrazine–CO
597.3017	$C_{30}H_{41}N_6O_7$	-1.4	$C_9H_{13}N_3O$	–phenylhydrazine–[Ala]
545.2495	$C_{29}H_{33}N_6O_5$	-1.2	$C_{10}H_{21}N_3O_3$	b_4
525.2444	$C_{26}H_{33}N_6O_6$	-1.2	$C_{13}H_{21}N_3O_2$	b_5–C_6H_5N

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
474.2125	$C_{26}H_{28}N_5O_4$	-1.1	$C_{13}H_{26}N_4O_4$	b_3
454.2073	$C_{23}H_{28}N_5O_5$	-1.2	$C_{16}H_{26}N_4O_3$	b_4–C_6H_5N
403.1755	$C_{23}H_{23}N_4O_3$	-1.0	$C_{16}H_{31}N_5O_5$	b_2
383.1704	$C_{20}H_{23}N_4O_4$	-1.0	$C_{19}H_{31}N_5O_4$	b_3–C_6H_5N
346.1542	$C_{21}H_{20}N_3O_2$	-0.8	$C_{18}H_{34}N_6O_6$	b_1
312.1336	$C_{17}H_{18}N_3O_3$	-0.7	$C_{22}H_{36}N_6O_5$	b_2–C_6H_5N
255.1123	$C_{15}H_{15}N_2O_2$	-0.5	$C_{24}H_{39}N_7O_6$	b_1–C_6H_5N

Table 5.10: Accurate Mass Measurements of UVPD-MS^2 of (1,4-s-GAAAK +H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
790.3972	$C_{38}H_{52}N_{11}O_8$	-2.3		
762.3908	$C_{38}H_{52}N_9O_8$	-2.5	N_2	
671.3495	$C_{32}H_{47}N_8O_8$	-1.6	$C_6H_5N_3$	–N_2–C_6H_5N
634.2969	$C_{32}H_{40}N_7O_7$	-1.5	$C_6H_{12}N_4O$	–N_2–[Lys]
589.2756	$C_{31}H_{37}N_6O_6$	-1.3	$C_7H_{15}N_5O_2$	*m/z* 634–NH_3–CO
563.2597	$C_{29}H_{35}N_6O_6$	-1.6	$C_9H_{17}N_5O_2$	–N_2–[AlaLys]
535.2651	$C_{28}H_{35}N_6O_5$	-1.2	$C_{10}H_{17}N_5O_3$	*m/z* 563–CO
518.2386	$C_{28}H_{32}N_5O_5$	-1.2	$C_{10}H_{20}N_6O_3$	*m/z* 589–[Ala]
492.2230	$C_{26}H_{30}N_5O_5$	-1.1	$C_{12}H_{22}N_6O_3$	–N_2–[AlaAlaLys]
474.2125	$C_{26}H_{28}N_5O_4$	-1.1	$C_{12}H_{24}N_6O_4$	b_3–N_2
464.2282	$C_{25}H_{30}N_5O_4$	-1.0	$C_{13}H_{22}N_6O_4$	*m/z* 492–CO
454.2075	$C_{23}H_{28}N_5O_5$	-1.0	$C_{15}H_{24}N_6O_3$	b_4–N_2–C_6H_5N
447.2015	$C_{25}H_{27}N_4O_4$	-1.2	$C_{13}H_{25}N_7O_4$	*m/z* 464–NH_3
421.1860	$C_{23}H_{25}N_4O_4$	-1.0	$C_{15}H_{27}N_7O_4$	–N_2–[AlaAlaAlaLys]

(Continued)

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
403.1756	$C_{23}H_{23}N_4O_3$	-0.9	$C_{15}H_{29}N_7O_5$	b_2–N_2
383.1706	$C_{20}H_{23}N_4O_4$	-0.8	$C_{18}H_{29}N_7O_4$	*m/z* 454–[Ala]
364.1648	$C_{21}H_{22}N_3O_3$	-0.8	$C_{17}H_{30}N_8O_5$	–N_2–[GlyAlaAlaAlaLys]
336.1700	$C_{20}H_{22}N_3O_2$	-0.7	$C_{18}H_{30}N_8O_6$	*m/z* 364–CO
313.1177	$C_{17}H_{17}N_2O_4$	-0.6	$C_{21}H_{35}N_9O_4$	
265.0967	$C_{16}H_{13}N_2O_2$	-0.5	$C_{22}H_{39}N_9O_6$	
239.0810	$C_{14}H_{11}N_2O_2$	-0.5	$C_{24}H_{41}N_9O_6$	*m/z* 364–$C_7H_{11}NO$
228.1015	$C_{14}H_{14}NO_2$	-0.4	$C_{24}H_{38}N_{10}O_6$	
221.0705	$C_{14}H_9N_2O$	-0.4	$C_{24}H_{43}N_9O_7$	b_1–N_2–$C_7H_{11}NO$

Table 5.11: Accurate Mass Measurements of CID-MS^2 of (1,4-*s*-GAAAK + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
790.3976	$C_{38}H_{52}N_{11}O_8$	-1.9		
762.3912	$C_{38}H_{52}N_9O_8$	-2.1	N_2	
744.3821	$C_{38}H_{50}N_9O_7$	-0.7	H_2N_2O	–N_2–H_2O
656.3395	$C_{32}H_{46}N_7O_8$	-0.7	$C_6H_6N_4$	–N_2–phenyldiazene
634.2977	$C_{32}H_{40}N_7O_7$	-0.7	$C_6H_{12}N_4O$	–N_2–[Lys]
563.2602	$C_{29}H_{35}N_6O_6$	-1.1	$C_9H_{17}N_5O_2$	–N_2–[AlaLys]
492.2233	$C_{26}H_{30}N_5O_5$	-0.8	$C_{12}H_{22}N_6O_3$	–N_2–[AlaAlaLys]

Table 5.12: Accurate Mass Measurements of CID-MS^3 of (1,4-*s*-GAAAK-OCH_3 + H – N_2)$^+$ Generated by CID of (1,4-*s*-GAAAK + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
762.3912	$C_{38}H_{52}N_9O_8$	-2.1		

(Continued)

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
744.3810	$C_{38}H_{50}N_9O_7$	-1.8	H_2O	
734.3966	$C_{37}H_{52}N_9O_7$	-1.8	CO	
654.3230	$C_{32}H_{44}N_7O_8$	-1.6	$C_6H_8N_2$	–phenylhydrazine
634.2967	$C_{32}H_{40}N_7O_7$	-1.7	$C_6H_{12}N_2O$	–[Lys] (β_5)
616.2865	$C_{32}H_{38}N_7O_6$	-1.3	$C_6H_{14}N_2O_2$	b_5
563.2596	$C_{29}H_{35}N_6O_6$	-1.7	$C_9H_{17}N_3O_2$	–[AlaLys] (β_4)
545.2495	$C_{29}H_{33}N_6O_5$	-1.2	$C_9H_{19}N_3O_3$	b_4
535.2651	$C_{28}H_{35}N_6O_5$	0.1	$C_{10}H_{17}N_3O_3$	*m/z* 563–CO
525.2445	$C_{26}H_{33}N_6O_6$	-1.1	$C_{12}H_{19}N_3O_2$	b_5–C_6H_5N
518.2385	$C_{28}H_{32}N_5O_5$	-1.3	$C_{10}H_{20}N_4O_3$	m/z 535–NH_3
492.2227	$C_{26}H_{30}N_5O_5$	-1.4	$C_{12}H_{22}N_4O_3$	–[AlaAlaLys] (β_3)
474.2124	$C_{26}H_{28}N_5O_4$	-1.2	$C_{12}H_{24}N_4O_4$	b_3
454.2074	$C_{23}H_{28}N_5O_5$	-1.1	$C_{15}H_{24}N_4O_3$	b_4–C_6H_5N
447.2015	$C_{25}H_{27}N_4O_4$	-1.2	$C_{13}H_{25}N_5O_4$	b_3–CO–NH_3
421.1859	$C_{23}H_{25}N_4O_4$	-1.1	$C_{15}H_{27}N_5O_4$	–[AlaAlaAlaLys] (β_2)
403.1754	$C_{23}H_{23}N_4O_3$	-1.1	$C_{15}H_{29}N_5O_5$	b_2
383.1704	$C_{20}H_{23}N_4O_4$	-1.0	$C_{18}H_{29}N_5O_4$	b_3–C_6H_5N

Table 5.13: Accurate Mass Measurements of CID-MS^3 of (1,4-s-GAAAK-OCH$_3$ + H – N$_2$)$^+$ Generated by UVPD of (1,4-s-GAAAK + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
762.3912	$C_{38}H_{52}N_9O_8$	-2.1		
744.3810	$C_{38}H_{50}N_9O_7$	-1.8	H_2O	
634.2966	$C_{32}H_{40}N_7O_7$	-1.8	$C_6H_{12}N_2O$	–[Lys] (β_5)
606.3020	$C_{31}H_{40}N_7O_6$	-1.5	$C_7H_{12}N_2O_2$	β_5–CO

(Continued)

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
588.2914	$C_{31}H_{38}N_7O_5$	-1.5	$C_7H_{14}N_2O_3$	β_5–CO–H_2O
563.2596	$C_{29}H_{35}N_6O_6$	-1.7	$C_9H_{17}N_3O_2$	–[AlaLys] (β_4)
535.2649	$C_{28}H_{35}N_6O_5$	-1.4	$C_{10}H_{17}N_3O_3$	β_4–CO
517.2543	$C_{28}H_{33}N_6O_4$	-1.5	$C_{10}H_{19}N_3O_4$	β_4–CO–H_2O
492.2228	$C_{26}H_{30}N_5O_5$	-1.3	$C_{12}H_{22}N_4O_3$	–[AlaAlaLys] (β_3)
464.2280	$C_{25}H_{30}N_5O_4$	-1.2	$C_{13}H_{22}N_4O_4$	β_3–CO
447.2015	$C_{25}H_{27}N_4O_4$	-1.2	$C_{13}H_{25}N_5O_4$	β_3–CO–NH_3
421.1858	$C_{23}H_{25}N_4O_4$	-1.2	$C_{15}H_{27}N_5O_4$	–[AlaAlaAlaLys] (β_2)

Table 5.14: Accurate Mass Measurements of UVPD-MS^2 of (1,2-s-GAAAR + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
818.4032	$C_{38}H_{52}N_{13}O_8$	-2.4		
790.3970	$C_{38}H_{52}N_{11}O_8$	-2.5	N_2	
762.4025	$C_{37}H_{52}N_{11}O_7$	-2.1	CN_2O	–N_2–CO
699.3554	$C_{32}H_{47}N_{10}O_8$	-1.9	$C_6H_5N_3$	–N_2–C_6H_5N
653.3502	$C_{31}H_{45}N_{10}O_6$	-1.6	$C_7H_7N_3O_2$	–N_2–$C_7H_7NO_2$
563.2598	$C_{29}H_{35}N_6O_6$	-1.5	$C_9H_{17}N_7O_2$	–N_2–[AlaArg]
492.2230	$C_{26}H_{30}N_5O_5$	-1.1	$C_{12}H_{22}N_8O_3$	–N_2–[AlaAlaArg]
454.2075	$C_{23}H_{28}N_5O_5$	-1.0	$C_{15}H_{24}N_8O_3$	b_4–N_2–C_6H_5N
421.1861	$C_{23}H_{25}N_4O_4$	-0.9	$C_{15}H_{27}N_9O_4$	–N_2–[AlaAlaAlaArg]
410.2137	$C_{17}H_{28}N_7O_5$	-0.9	$C_{21}H_{24}N_6O_3$	
383.1706	$C_{20}H_{23}N_4O_4$	-0.8	$C_{18}H_{29}N_9O_4$	b_3–N_2–C_6H_5N
364.1648	$C_{21}H_{22}N_3O_3$	-0.8	$C_{17}H_{30}N_{10}O_5$	–N_2–[GlyAlaAlaAlaArg]
353.1924	$C_{15}H_{25}N_6O_4$	-0.8	$C_{23}H_{27}N_7O_4$	*m/z* 410–[Gly]

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
346.1543	$C_{21}H_{20}N_3O_2$	-0.7	$C_{17}H_{32}N_{10}O_6$	b_1–N_2
312.1336	$C_{17}H_{18}N_3O_3$	-0.7	$C_{21}H_{34}N_{10}O_5$	b_2–N_2–C_6H_5N
300.1660	$C_{12}H_{22}N_5O_4$	-0.6	$C_{26}H_{30}N_8O_4$	
282.1555	$C_{12}H_{20}N_5O_3$	-0.6	$C_{26}H_{32}N_8O_5$	*m/z* 300–H_2O
271.1396	$C_{11}H_{19}N_4O_4$	-0.5	$C_{27}H_{33}N_9O_4$	
256.1075	$C_{14}H_{14}N_3O_2$	-0.6	$C_{24}H_{38}N_{10}O_6$	
254.1607	$C_{11}H_{20}N_5O_2$	-0.5	$C_{27}H_{32}N_8O_6$	
246.1556	$C_9H_{20}N_5O_3$	-0.5	$C_{29}H_{32}N_8O_5$	
243.1447	$C_{10}H_{19}N_4O_3$	-0.5	$C_{28}H_{33}N_9O_5$	
239.0810	$C_{14}H_{11}N_2O_2$	-0.5	$C_{24}H_{41}N_{11}O_6$	
238.0970	$C_{14}H_{12}N_3O$	-0.5	$C_{24}H_{40}N_{10}O_7$	
229.1291	$C_9H_{17}N_4O_3$	-0.4	$C_{29}H_{35}N_9O_5$	

Table 5.15: Accurate Mass Measurements of CID-MS^3 of **4**$^+$ Generated by UVPD of (1,2-*s*-GAAAR + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
790.3977	$C_{38}H_{52}N_{11}O_8$	-1.8		
772.3870	$C_{38}H_{50}N_{11}O_7$	-1.9	H_2O	
762.4029	$C_{37}H_{52}N_{11}O_7$	-1.7	CO	
747.3919	$C_{37}H_{51}N_{10}O_7$	-1.8	CHNO	
730.3655	$C_{37}H_{48}N_9O_7$	-1.6	CH_4N_2O	
698.3479	$C_{32}H_{46}N_{10}O_8$	-1.6	C_6H_6N	
653.3504	$C_{31}H_{45}N_{10}O_6$	-1.4	$C_7H_7NO_2$	

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
634.2971	$C_{32}H_{40}N_7O_7$	-1.3	$C_6H_{12}N_4O$	–[Arg] (β_5)
563.2600	$C_{29}H_{35}N_6O_6$	-1.3	$C_9H_{17}N_5O_2$	–[AlaArg] (β_4)
545.2495	$C_{29}H_{33}N_6O_5$	-1.2	$C_9H_{19}N_5O_3$	b_4
525.2444	$C_{26}H_{33}N_6O_6$	-1.2	$C_{12}H_{19}N_5O_2$	β_5–C_6H_5N–H_2O
492.2231	$C_{26}H_{30}N_5O_5$	-1.0	$C_{12}H_{22}N_6O_3$	–[AlaAlaArg] (β_3)
474.2125	$C_{26}H_{28}N_5O_4$	-1.0	$C_{12}H_{24}N_6O_4$	b_3
454.2075	$C_{23}H_{28}N_5O_5$	-1.0	$C_{15}H_{24}N_6O_3$	β_4–C_6H_5N–H_2O
421.1861	$C_{23}H_{25}N_4O_4$	-0.9	$C_{15}H_{27}N_7O_4$	–[AlaAlaAlaArg] (β_2)
410.2139	$C_{17}H_{28}N_7O_5$	-0.7	$C_{21}H_{24}N_4O_3$	
403.1756	$C_{23}H_{23}N_4O_3$	-0.9	$C_{15}H_{29}N_7O_5$	b_2
388.1761	$C_{22}H_{22}N_5O_2$	-0.7	$C_{16}H_{30}N_6O_6$	
364.1649	$C_{21}H_{22}N_3O_3$	-0.7	$C_{17}H_{30}N_8O_5$	–[GlyAlaAlaAlaArg] (β_1)
353.1925	$C_{15}H_{25}N_6O_4$	-0.7	$C_{23}H_{27}N_5O_4$	*m/z* 410–[Gly]
346.1544	$C_{21}H_{20}N_3O_2$	-0.6	$C_{17}H_{32}N_8O_6$	b_1
282.1556	$C_{12}H_{20}N_5O_3$	-0.5	$C_{26}H_{32}N_6O_5$	*m/z* 353–[Ala]

Table 5.16: Accurate Mass Measurements of CID-MS^2 of (1,2-*s*-GAAAR + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
m/z	**Ion Formula**	**Error (mu)**	**Neutral Loss**	**Ion Assignment**
818.4039	$C_{38}H_{52}N_{13}O_8$	-1.7		
790.3971	$C_{38}H_{52}N_{11}O_8$	-2.4	N_2	

Table 5.17: Accurate Mass Measurements of CID-MS^3 of **4**$^+$ Generated by CID of (1,2-s-GAAAR + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
m/z	**Ion Formula**	**Error (mu)**	**Neutral Loss**	**Ion Assignment**
790.3980	$C_{38}H_{52}N_{11}O_8$	-1.5		
772.3875	$C_{38}H_{50}N_{11}O_7$	-1.4	H_2O	
762.4030	$C_{37}H_{52}N_{11}O_7$	-1.6	CO	
755.3610	$C_{38}H_{47}N_{10}O_7$	-1.4	H_5NO	*m/z* 772–NH_3
747.3926	$C_{37}H_{51}N_{10}O_7$	-1.1	CHNO	
745.3768	$C_{37}H_{49}N_{10}O_7$	-1.2	CH_3NO	*m/z* 762–NH_3
730.3660	$C_{37}H_{48}N_9O_7$	-1.1	CH_4N_2O	*m/z* 747–NH_3
702.3347	$C_{35}H_{44}N_9O_7$	-1.1	$C_3H_8N_2O$	
698.3483	$C_{32}H_{46}N_{10}O_8$	-1.2	C_6H_6N	
653.3505	$C_{31}H_{45}N_{10}O_6$	-1.3	$C_7H_7NO_2$	
634.2973	$C_{32}H_{40}N_7O_7$	-1.1	$C_6H_{12}N_4O$	–[Arg] (β_5)
563.2603	$C_{29}H_{35}N_6O_6$	-1.0	$C_9H_{17}N_5O_2$	–[AlaArg] (β_4)
545.2499	$C_{29}H_{33}N_6O_5$	-0.8	$C_9H_{19}N_5O_3$	b_4
525.2448	$C_{26}H_{33}N_6O_6$	4.3	$C_{12}H_{19}N_5O_2$	b_5–C_6H_5N
492.2232	$C_{26}H_{30}N_5O_5$	-0.9	$C_{12}H_{22}N_6O_3$	–[AlaAlaArg] (β_3)
474.2128	$C_{26}H_{28}N_5O_4$	-0.8	$C_{12}H_{24}N_6O_4$	b_3
454.2078	$C_{23}H_{28}N_5O_5$	-0.7	$C_{15}H_{24}N_6O_3$	b_4–C_6H_5N
421.1864	$C_{23}H_{25}N_4O_4$	-0.6	$C_{15}H_{27}N_7O_4$	–[AlaAlaAlaArg] (β_2)
410.2140	$C_{17}H_{28}N_7O_5$	-0.6	$C_{21}H_{24}N_4O_3$	
403.1759	$C_{23}H_{23}N_4O_3$	-0.6	$C_{15}H_{29}N_7O_5$	b_2
388.1763	$C_{22}H_{22}N_5O_2$	-0.5	$C_{16}H_{30}N_6O_6$	
383.1709	$C_{20}H_{23}N_4O_4$	-0.5	$C_{18}H_{29}N_7O_4$	b_3–C_6H_5N

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
364.1651	$C_{21}H_{22}N_3O_3$	-0.5	$C_{17}H_{30}N_8O_5$	–[GlyAlaAlaAlaArg] (β_1)
353.1927	$C_{15}H_{25}N_6O_4$	-0.5	$C_{23}H_{27}N_5O_4$	*m/z* 410–[Gly]
346.1545	$C_{21}H_{20}N_3O_2$	-0.5	$C_{17}H_{32}N_8O_6$	b_1
282.1557	$C_{12}H_{20}N_5O_3$	-0.4	$C_{26}H_{32}N_6O_5$	*m/z* 353–[Ala]

Table 5.18: Accurate Mass Measurements of CID-MS^2 of (1,2-*s*-GAAAR-OCH$_3$ + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
832.4191	$C_{39}H_{54}N_{13}O_8$	-2.2		
804.4120	$C_{39}H_{54}N_{11}O_8$	-3.1	N_2	

Table 5.19: Accurate Mass Measurements of UVPD-MS^2 of (1,2-*s*-GAAAR-OCH$_3$ + H)$^+$

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
832.4184	$C_{39}H_{54}N_{13}O_8$	-2.3		
804.4117	$C_{39}H_{54}N_{11}O_8$	-3.4	N_2	
787.3858	$C_{39}H_{51}N_{10}O_8$	-2.8	H_3N_3	–N_2–NH_3
713.3704	$C_{33}H_{49}N_{10}O_8$	-2.5	$C_6H_5N_3$	–N_2–C_6H_5N
654.3226	$C_{32}H_{44}N_7O_8$	-2.0	$C_7H_{10}N_4$	
582.3340	$C_{25}H_{46}N_9O_7$	-1.8	$C_{14}H_8N_4O$	y_6
545.2490	$C_{29}H_{33}N_6O_5$	-1.7	$C_{10}H_{21}N_7O_3$	b_4–N_2
491.2386	$C_{26}H_{31}N_6O_4$	-1.5	$C_{13}H_{23}N_7O_4$	–N_2–[ArgOCH$_3$–NH$_3$] –[AlaAla]
474.2121	$C_{26}H_{28}N_5O_4$	-1.5	$C_{13}H_{26}N_8O_8$	b_3–N_2

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
442.2395	$C_{18}H_{32}N_7O_6$	-1.4	$C_{21}H_{22}N_6O_2$	
420.2018	$C_{23}H_{26}N_5O_3$	-1.2	$C_{16}H_{28}N_8O_5$	*m/z* 491–[Ala]
403.1753	$C_{23}H_{23}N_4O_3$	-1.2	$C_{16}H_{31}N_9O_5$	b_2–N_2
385.2182	$C_{16}H_{29}N_6O_5$	-1.2	$C_{23}H_{25}N_7O_3$	*m/z* 442–[Gly]
363.1805	$C_{21}H_{23}N_4O_2$	-1.1	$C_{18}H_{31}N_9O_6$	*m/z* 491–[GlyAla]
346.1540	$C_{21}H_{20}N_3O_2$	-1.0	$C_{18}H_{34}N_{10}O_6$	b_1–N_2
314.1814	$C_{13}H_{24}N_5O_4$	-0.9	$C_{26}H_{30}N_8O_4$	*m/z* 385–[Ala]
271.1394	$C_{11}H_{19}N_4O_4$	-0.7	$C_{28}H_{35}N_9O_4$	
263.0921	$C_{15}H_{11}N_4O$	-0.6	$C_{24}H_{43}N_9O_7$	
243.1445	$C_{10}H_{19}N_4O_3$	-0.7	$C_{29}H_{35}N_9O_5$	*m/z* 314–[Ala]
235.0972	$C_{14}H_{11}N_4$	-0.6	$C_{25}H_{43}N_9O_8$	*m/z* 263–CO

Table 5.20: Accurate Mass Measurements of CID-MS^3 of (1,2-*s*-GAAAR-OCH_3 + H – N_2)$^+$ Generated by UVPD of (1,2-*s*-GAAAR-OCH_3 + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
804.4127	$C_{39}H_{54}N_{11}O_8$	-3.1		
787.3862	$C_{39}H_{51}N_{10}O_8$	-2.4	H_3N	–NH_3
769.3760	$C_{39}H_{49}N_{10}O_7$	-2.0	H_5NO	–NH_3–H_2O
762.3914	$C_{38}H_{52}N_9O_8$	-1.9	CH_2N_2	
759.3917	$C_{38}H_{51}N_{10}O_7$	-2.0	CH_3NO	–NH_3–CO
716.3496	$C_{36}H_{46}N_9O_7$	-1.9	$C_3H_8N_2O$	–NH_3–[Ala]
645.3129	$C_{33}H_{41}N_8O_6$	-1.5	$C_6H_{13}N_3O_2$	–NH_3–[AlaAla]
633.3128	$C_{32}H_{41}N_8O_6$	-1.6	$C_7H_{13}N_3O_2$	–[ArgOCH_3–NH_3]

(Continued)

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
616.2864	$C_{32}H_{38}N_7O_6$	-1.4	$C_7H_{16}N_4O_2$	b_5
588.2914	$C_{31}H_{38}N_7O_5$	-1.5	$C_8H_{16}N_4O_3$	b_5–CO
587.2711	$C_{30}H_{35}N_8O_5$	-1.4		
562.2758	$C_{29}H_{36}N_7O_5$	-1.4	$C_{10}H_{18}N_4O_3$	–[ArgOCH$_3$–NH$_3$]–[Ala]
545.2493	$C_{29}H_{33}N_6O_5$	-1.4	$C_{10}H_{21}N_5O_3$	b_4
517.2545	$C_{28}H_{33}N_6O_4$	-1.3	$C_{11}H_{21}N_5O_4$	b_4–CO
516.2341	$C_{27}H_{30}N_7O_4$	-1.3	$C_{12}H_{24}N_4O_4$	*m/z* 587–[Ala]
491.2389	$C_{26}H_{31}N_6O_4$	-1.2	$C_{13}H_{23}N_5O_4$	–[ArgOCH$_3$–NH$_3$]–[AlaAla]
474.2122	$C_{26}H_{28}N_5O_4$	-1.4	$C_{13}H_{26}N_6O_4$	b_3
442.2397	$C_{18}H_{32}N_7O_6$	-1.2	$C_{21}H_{22}N_5O_2$	
420.2019	$C_{23}H_{26}N_5O_3$	-1.1	$C_{16}H_{28}N_6O_5$	–[ArgOCH$_3$–NH$_3$]–[AlaAlaAla]
403.1753	$C_{23}H_{23}N_4O_3$	-1.2	$C_{16}H_{31}N_7O_5$	b_2
388.1758	$C_{22}H_{22}N_5O_2$	-1.0	$C_{17}H_{32}N_6O_6$	
363.1807	$C_{21}H_{23}N_4O_2$	-0.9	$C_{18}H_{31}N_7O_6$	b_4–CO–$C_7H_7NO_2$–NH$_3$
346.1540	$C_{21}H_{20}N_3O_2$	-1.0	$C_{18}H_{34}N_8O_6$	b_1
314.1816	$C_{13}H_{24}N_5O_4$	-0.7	$C_{26}H_{30}N_6O_4$	
263.0922	$C_{15}H_{11}N_4O$	-0.5	$C_{24}H_{33}N_7O_7$	
238.0970	$C_{14}H_{12}N_3O$	-0.5	$C_{25}H_{32}N_8O_7$	

Table 5.21: Accurate Mass Measurements of CID-MS^3 of (1,2-*s*-GAAAR-OCH$_3$ + H – N$_2$)$^+$ Generated by CID of (1,2-*s*-GAAAR-OCH$_3$ + H)$^+$

m/z	Ion Formula	Error (mmu)	Neutral Loss	Ion Assignment
804.4124	$C_{39}H_{54}N_{11}O_8$	-2.7		

(Continued)

m/z	**Ion Formula**	**Error (mmu)**	**Neutral Loss**	**Ion Assignment**
787.3857	$C_{39}H_{51}N_{10}O_8$	-2.9	H_3N	$–NH_3$
769.3756	$C_{39}H_{49}N_{10}O_7$	-2.4	H_5NO	$–NH_3–H_2O$
762.3910	$C_{38}H_{52}N_9O_8$	-2.3	CH_2N_2	
716.3492	$C_{36}H_{46}N_9O_7$	-2.3	$C_3H_8N_2O$	$–NH_3–H_2O$–[Ala]
654.3226	$C_{32}H_{44}N_7O_8$	-2.0	$C_7H_{10}N_4$	
633.3125	$C_{32}H_{41}N_8O_6$	-1.9	$C_7H_{13}N_3O_2$	–[ArgOCH$_3$–NH$_3$]
616.2860	$C_{32}H_{38}N_7O_6$	-1.8	$C_7H_{16}N_4O_2$	b_5
587.2708	$C_{30}H_{35}N_8O_5$	-1.7	$C_9H_{19}N_3O_3$	
562.2755	$C_{29}H_{36}N_7O_5$	-1.7	$C_{10}H_{18}N_4O_3$	–[ArgOCH$_3$–NH$_3$]–[Ala]
545.2490	$C_{29}H_{33}N_6O_5$	-1.7	$C_{10}H_{21}N_5O_3$	b_4
516.2338	$C_{27}H_{30}N_7O_4$	-1.6	$C_{12}H_{24}N_4O_4$	*m/z* 587–[Ala]
491.2386	$C_{26}H_{31}N_6O_4$	-1.5	$C_{13}H_{23}N_5O_4$	–[ArgOCH$_3$–NH$_3$]–[AlaAla]
474.2120	$C_{26}H_{28}N_5O_4$	-1.6	$C_{13}H_{26}N_6O_4$	b_3
454.2071	$C_{23}H_{28}N_5O_5$	-1.4	$C_{16}H_{26}N_6O_3$	$b_4–C_6H_5N$
445.1969	$C_{24}H_{25}N_6O_3$	-1.4	$C_{15}H_{29}N_5O_5$	
442.2394	$C_{18}H_{32}N_7O_6$	-1.5	$C_{21}H_{22}N_4O_2$	
420.2017	$C_{23}H_{26}N_5O_3$	-1.3	$C_{16}H_{28}N_6O_5$	–[ArgOCH$_3$–NH$_3$]–[AlaAlaAla]
403.1752	$C_{23}H_{23}N_4O_3$	-1.3	$C_{16}H_{31}N_7O_5$	b_2
388.1756	$C_{22}H_{22}N_5O_2$	-1.2	$C_{17}H_{32}N_6O_6$	*m/z* 445–[Gly]
363.1805	$C_{21}H_{23}N_4O_2$	-1.1	$C_{18}H_{31}N_7O_6$	–[ArgOCH$_3$–NH$_3$]–[GlyAlaAlaAla]
346.1539	$C_{21}H_{20}N_3O_2$	-1.1	$C_{18}H_{34}N_8O_6$	b_1
314.1814	$C_{13}H_{24}N_5O_4$	-0.9	$C_{26}H_{30}N_6O_4$	
263.0920	$C_{15}H_{11}N_4O$	-0.7	$C_{24}H_{41}N_7O_7$	
238.0969	$C_{14}H_{12}N_3O$	-0.6	$C_{25}H_{42}N_8O_7$	

Table 5.22: Accurate Mass Measurements of UVPD-MS^2 of (1,4-*s*-GAAAR + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
818.4040	$C_{38}H_{52}N_{13}O_8$	-1.6		
790.3973	$C_{38}H_{52}N_{11}O_8$	-2.2	N_2	
762.4031	$C_{37}H_{52}N_{11}O_7$	-1.5	CN_2O	–N_2–CO
699.3560	$C_{32}H_{47}N_{10}O_8$	-1.3	$C_6H_5N_3$	–N_2–C_6H_5N
653.3506	$C_{31}H_{45}N_{10}O_6$	-1.2	$C_7H_7N_3O_2$	–N_2–$C_7H_7NO_2$
563.2603	$C_{29}H_{35}N_6O_6$	-1.0	$C_9H_{17}N_7O_2$	–N_2–[AlaArg]
545.2498	$C_{29}H_{33}N_6O_5$	-0.9	$C_9H_{19}N_7O_3$	b_4–N_2
492.2233	$C_{26}H_{30}N_5O_5$	-0.8	$C_{12}H_{22}N_8O_3$	–N_2–[AlaAlaArg]
474.2127	$C_{26}H_{28}N_5O_4$	-0.9	$C_{12}H_{24}N_8O_4$	b_3–N_2
464.2284	$C_{25}H_{30}N_5O_4$	-0.8	$C_{13}H_{22}N_8O_4$	*m/z* 492–CO
454.2077	$C_{23}H_{28}N_5O_5$	-0.8	$C_{15}H_{24}N_8O_3$	b_4–N_2–C_6H_5N
447.2020	$C_{25}H_{27}N_4O_4$	-0.7	$C_{13}H_{25}N_9O_4$	*m/z* 492–CO–NH_3
421.1863	$C_{23}H_{25}N_4O_4$	-0.7	$C_{15}H_{27}N_9O_4$	–N_2–[AlaAlaAlaArg]
403.1757	$C_{23}H_{23}N_4O_3$	-0.8	$C_{15}H_{29}N_9O_5$	b_2–N_2
353.1927	$C_{15}H_{25}N_6O_4$	-0.5	$C_{23}H_{27}N_7O_4$	
300.1662	$C_{12}H_{22}N_5O_4$	-0.4	$C_{26}H_{30}N_8O_4$	
282.1557	$C_{12}H_{20}N_5O_3$	-0.4	$C_{26}H_{32}N_8O_5$	*m/z* 353–[Ala]

Table 5.23: Accurate Mass Measurements of CID-MS^2 of (1,4-*s*-GAAAR + H)$^+$

m/z	Ion Formula	Error (mu)	Neutral Loss	Ion Assignment
818.4038	$C_{38}H_{52}N_{13}O_8$	-1.8		
790.3967	$C_{38}H_{52}N_{11}O_8$	-2.8	N_2	

Table 5.24: Accurate Mass Measurements of CID-MS^3 of (1,4-s-GAAAR + H – N_2)$^+$ Generated by CID of (1,4-s-GAAAR + H)$^+$

m/z	**Ion Formula**	**Error (mu)**	**Neutral Loss**	**Ion Assignment**
790.3975	$C_{38}H_{52}N_{11}O_8$	-2.0		
773.3712	$C_{38}H_{49}N_{10}O_8$	-1.7	H_3N	–NH_3
772.3871	$C_{38}H_{50}N_{11}O_7$	-1.8	H_2O	
762.4036	$C_{37}H_{52}N_{11}O_7$	-1.0	CO	
745.3768	$C_{37}H_{49}N_{10}O_7$	-1.2	CH_3NO	–NH_3–CO
730.3657	$C_{37}H_{48}N_9O_7$	-1.4	CH_4N_2O	–NH_3–CHNO
702.3345	$C_{35}H_{44}N_9O_7$	-1.3	$C_3H_8N_2O$	–NH_3–[Ala]
653.3505	$C_{31}H_{45}N_{10}O_6$	-1.3	$C_7H_7NO_2$	
634.2973	$C_{32}H_{40}N_7O_7$	-1.1	$C_6H_{12}N_4O$	–[Arg] (β_5)
616.2866	$C_{32}H_{38}N_7O_6$	-1.2	$C_6H_{14}N_4O_2$	b_5
563.2600	$C_{29}H_{35}N_6O_6$	-1.3	$C_9H_{17}N_5O_2$	–[AlaArg] (β_4)
545.2496	$C_{29}H_{33}N_6O_5$	-1.1	$C_9H_{19}N_5O_3$	b_4
535.2654	$C_{28}H_{35}N_6O_5$	-0.9	$C_{10}H_{17}N_5O_3$	b_4–CO
518.2387	$C_{28}H_{32}N_5O_5$	-1.1	$C_{10}H_{20}N_6O_3$	b_4–CO–NH_3
492.2230	$C_{26}H_{30}N_5O_5$	-1.1	$C_{12}H_{22}N_6O_3$	–[AlaAlaArg] (β_3)
474.2125	$C_{26}H_{28}N_5O_4$	-1.1	$C_{12}H_{24}N_6O_4$	b_3
464.2283	$C_{25}H_{30}N_5O_4$	-0.9	$C_{13}H_{22}N_6O_4$	β_3–CO
447.2016	$C_{25}H_{27}N_4O_4$	-1.1	$C_{13}H_{25}N_7O_4$	β_3–CO–NH_3
421.1862	$C_{23}H_{25}N_4O_4$	-0.8	$C_{15}H_{27}N_7O_4$	–[AlaAlaAlaArg] (β_2)
403.1755	$C_{23}H_{23}N_4O_3$	-1.0	$C_{15}H_{29}N_7O_5$	b_2
375.1696	$C_{23}H_{23}N_2O_3$	-0.7	$C_{15}H_{29}N_9O_5$	
353.1924	$C_{15}H_{25}N_6O_4$	-0.8	$C_{23}H_{27}N_5O_4$	
318.1594	$C_{20}H_{20}N_3O$	-0.7	$C_{18}H_{32}N_8O_7$	
282.1555	$C_{12}H_{20}N_5O_3$	-0.6	$C_{26}H_{32}N_6O_5$	*m/z* 353–[Ala]

Table 5.25: Accurate Mass Measurements of CID-MS^3 of (1,4-s-GAAAR + H – N_2)$^+$ Generated by UVPD of (1,4-s-GAAAR + H)$^+$

m/z	**Ion Formula**	**Error (mu)**	**Neutral Loss**	**Ion Assignment**
790.3976	$C_{38}H_{52}N_{11}O_8$	-1.9		
772.3873	$C_{38}H_{50}N_{11}O_7$	-1.6	H_2O	
762.4040	$C_{37}H_{52}N_{11}O_7$	-0.6	CO	
745.3765	$C_{37}H_{49}N_{10}O_7$	-1.5	CH_3NO	–NH_3–CO
730.3657	$C_{37}H_{48}N_9O_7$	-1.4	CH_4N_2O	–NH_3–CHNO
702.3345	$C_{35}H_{44}N_9O_7$	-1.3	$C_3H_8N_2O$	–NH_3–[Ala]
653.3506	$C_{31}H_{45}N_{10}O_6$	-1.2	$C_7H_7NO_2$	
634.2972	$C_{32}H_{40}N_7O_7$	-1.2	$C_6H_{12}N_4O$	–[Arg] (β_5)
616.2866	$C_{32}H_{38}N_7O_6$	-1.2	$C_6H_{14}N_4O_2$	b_5
563.2600	$C_{29}H_{35}N_6O_6$	-1.3	$C_9H_{17}N_5O_2$	–[AlaArg] (β_4)
545.2497	$C_{29}H_{33}N_6O_5$	-1.0	$C_9H_{19}N_5O_3$	b_4
535.2653	$C_{28}H_{35}N_6O_5$	-1.0	$C_{10}H_{17}N_5O_3$	β_4–CO
518.2387	$C_{28}H_{32}N_5O_5$	-1.1	$C_{10}H_{20}N_6O_3$	β_4–CO–NH_3
492.2230	$C_{26}H_{30}N_5O_5$	-1.1	$C_{12}H_{22}N_6O_3$	–[AlaAlaArg] (β_3)
474.2126	$C_{26}H_{28}N_5O_4$	-1.0	$C_{12}H_{24}N_6O_4$	b_3
464.2283	$C_{25}H_{30}N_5O_4$	-0.9	$C_{13}H_{22}N_6O_4$	β_3–CO
454.2076	$C_{23}H_{28}N_5O_5$	-0.9	$C_{15}H_{24}N_6O_3$	b_4–C_6H_5N
447.2017	$C_{25}H_{27}N_4O_4$	-1.0	$C_{13}H_{25}N_7O_4$	β_3–CO–NH_3
421.1861	$C_{23}H_{25}N_4O_4$	-0.9	$C_{15}H_{27}N_7O_4$	–[AlaAlaAlaArg] (β_2)
403.1756	$C_{23}H_{23}N_4O_3$	-0.9	$C_{15}H_{29}N_7O_5$	b_2
375.1696	$C_{23}H_{23}N_2O_3$	-0.7	$C_{15}H_{29}N_9O_5$	
353.1924	$C_{15}H_{25}N_6O_4$	-0.8	$C_{23}H_{27}N_5O_4$	
300.1660	$C_{12}H_{22}N_5O_4$	-0.6	$C_{26}H_{30}N_6O_4$	
282.1555	$C_{12}H_{20}N_5O_3$	-0.6	$C_{26}H_{32}N_6O_5$	*m/z* 353–[Ala]

5.2.5 Calculations

Structures of the conjugates and their dissociation products were generated by Born-Oppenheimer molecular dynamics (BOMD) calculations that were run as 20 ps trajectories with 1 fs steps, using the Berendsen thermostat[5] at 510, 610, and 810 K. Trajectory analysis confirmed that the systems were reaching a near-stationary state after ca 16000 steps (16 ps), as determined by low root-mean-square deviations of atomic Cartesian coordinates, so no longer trajectory runs were necessary because they generated only duplicate structures. Initial protonation sites were chosen at different amide groups for the non-basic AAAG conjugates to be treated by BOMD. Since BOMD includes both electron and nuclear motion, the different protonation sites coalesced by proton transfer to the most basic position. The initial protonation sites in GAAAK and GAAAR conjugates were at the basic residues. The trajectory calculations were performed with PM6-D3H4[6] which complements the semiempirical Hamiltonian with dispersion and hydrogen-bonding interactions, using the Cuby4 platform[7] and MOPAC.[8] Two hundred snapshots were selected at regular 100 fs intervals from the BOMD trajectories, and the structures were fully gradient-optimized with PM6-D3H4. Several selected low energy structures were further optimized with B3LYP[9] with the 6-31+G(d,p) basis set. These calculations were augmented by including empirical dispersion corrections (GD3-BJ)[10][11] and were used to provide harmonic frequencies. Several low-energy structures were reoptimized with M06-2X[12] using the 6-31+G(d,p) basis set. The M06-2X/6-31+G(d,p) optimized geometries were used to calculate single-point energies which were carried out with M06-2X and the def2qzvpp[13] basis set (<5000 basis functions). It should be noted that ion thermochemistry obtained at this level of theory is approximate within an estimated 5-10 kJ mol^{-1},[12][14] and thus no efforts were made to derive the ion isomer equilibrium populations from the calculated Gibbs free energies ($\Delta G_{g,310}$). Another set of M06-2X/6-31++G(d,p) single-point calculations were used to obtain charge densities according to Merz, Singh, and Kollman (MK).[15][16] All the standard DFT calculations were run using Gaussian 16 (Revision B.01) that was licensed from Gaussian, Inc. (Wallingford, CT). Collision cross sections in nitrogen were calculated by the modified ion trajectory method (MobCal$_{\mathrm{MPI}}$)[17][18] using the MK charge densities. Standard van der Waals parameters for different atom types were obtained from the MMFF94 data set.[19]

5.3 Results and Discussion

5.3.1 Photodissociation and Crosslinking Yields of 1,2-*s*-GAAAK and 1,4-*s*-GAAAK

5.3.1.1 Spectra Interpretation of 1,2-s-GAAAK

UVPD at 213 nm of (1,2-s-GAAAK + H)$^+$ (*m/z* 790) resulted in loss of N_2 followed by backbone dissociations of the (1,2-s-GAAAK + H – N_2)$^+$ ion (*m/z* 762, Figure 5.4 (a), Table 5.2 and 5.26). The dissociations produced two main series of fragment ions. The sequential loss of [Lys], [Ala], and [Gly] gave rise to ions at *m/z* 634, 563, 492, 421, and 364 (blue-annotated ions in Figure 5.4 (a)). These represented the crosslinked fraction of (1,2-s-GAAAK + H – N_2)$^+$ where dissociation started from the C-terminal Lys residue while the carboxyl OH was retained in the β_n ions. These crosslinked fragments were proved further with 1,2-s-GAAAK analogue in which the 2,5-diaryltetrazole moiety was replaced by a benzamide group , annotated as (1,2-b-GAAAK + H)+ to exclude N_2 loss (Figure 5.2). The other main series (green annotated ions) started from (1,2-s-GAAAK + H – N_2)$^+$ by loss of C_6H_5N (*m/z* 671) that was followed by regular backbone cleavages forming the $\boldsymbol{b}_n$ ions at *m/z* 525, 454, 383, and 312. CID-MS^2 of (1,2-s-GAAAK + H)$^+$ gave very similar results (Figure 5.1 (a), Table 5.3). CID-MS^3 of (1,2-s-GAAAK + H – N_2)$^+$ generated by CID (Figure 5.3 (a)) and UVPD (Figure 5.5 (a)) gave very similar spectra that were dominated by the β_n ions. In addition, CID-MS^3 led to a loss of $C_6H_8N_2$ (*m/z* 654) most likely phenylhydrazine, that has been observed for other peptide-nitrile imine conjugates (Table 5.4, 5.5).[20]

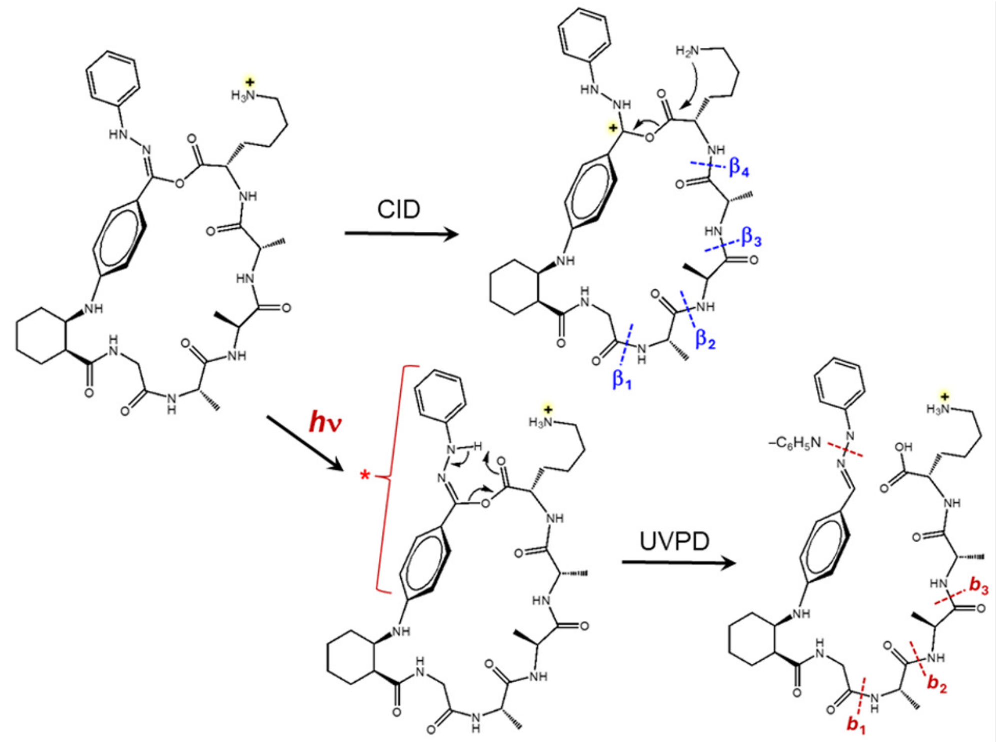
In contrast to the CID spectrum, UVPD of (1,2-s-GAAAK + H – N_2)$^+$ resulted in the loss of C_6H_5N followed by backbone cleavages forming the $\boldsymbol{b}_n$ ions (Figure 5.6 (a)). The Figure 5.5 (a) and 5.6 (a) data apparently gave conflicting results, as the β_n ions represented the crosslinked fraction whereas the $\boldsymbol{b}_n$ ions were deemed to be indicative of the straight chain isomers. The different dissociations on CID and UVPD can be explained by the very different modes of ion excitation. CID under multiple collision conditions follows the slow heating internal energy build up,[21] promoting low activation energy reactions. This is depicted in Scheme 5.2 by proton transfer onto the phenylhydrazone moiety in crosslinked (1,2-s-GAAAK + H – N_2)$^+$, resulting in the weakening and cleavage of the lysine CO–O bond and triggering the formation of the β_n ion series. In contrast, photoexcitation chiefly affects the aromatic ring chromophore, resulting in a photochemical cleavage of the C–OCO bond that restores the open-chain lysine residue. The internal energy supplied by the photon (562 kJ mol^{-1} at 213 nm) then drives standard peptide amide dissociations forming the $\boldsymbol{b}_n$ ions.

Table 5.26: CID-MS^4 Fragments of (1,2-s-GAAAK + H – N_2)$^+$ *m/z* 762 from 1,2-s-GAAAK Following UVPD at 213 nm

744 –H_2O	654 –$C_6H_8N_2$	634 β_5	563 β_4	492 β_3	421 β_2
726	636	606	535	464	364
661	626	563	492	447	256
616	609	492	421	421	166
563	583		364	364	138
545	538				135
525	525				120
492	515				
454	454				
421	383				
383					
312					

5.3.1.2 Spectra Interpretation of 1,4-s-GAAAK

The trans-1,4-cyclohexane conjugate 1,4-s-GAAAK was investigated by a combination of sequential excitations by UVPD and CID (Table 5.10-5.13). The UVPD-MS^2 spectrum of (1,4-s-GAAAK + H)$^+$ (*m/z* 790) showed a series of β_n ions originating from the denitrogenated intermediate (1,4-s-GAAAK + H – N_2)$^+$ (*m/z* 762, Figure 5.4 (b)). CID-MS^3 of (1,4-s-GAAAK + H – N_2)$^+$ (*m/z* 762, Figure 5.5 (b)) and further CID-MS^4 of peaks originating from CID-MS^3 of (1,4-s-GAAAK + H – N_2)$^+$ (Table 5.27) were representative of the ion structure and dissociations. The spectrum showed a series of dominant β_n ions at *m/z* 634, 563, and 492 by sequential loss of [Lys] and [Ala] that resembled the analogous ion series produced from the (1,4-s-GAAAK + H)$^+$ ion (1,4-s-GAAAK + H – N_2)$^+$. A minor C_6H_5N loss on UVPD



Scheme 5.2: Different Dissociations of Crosslinked Intermediate (1,2-*s*-GAAAK + H – $N_2)^+$ upon Collision Activation and 213 nm Photon Absorption.

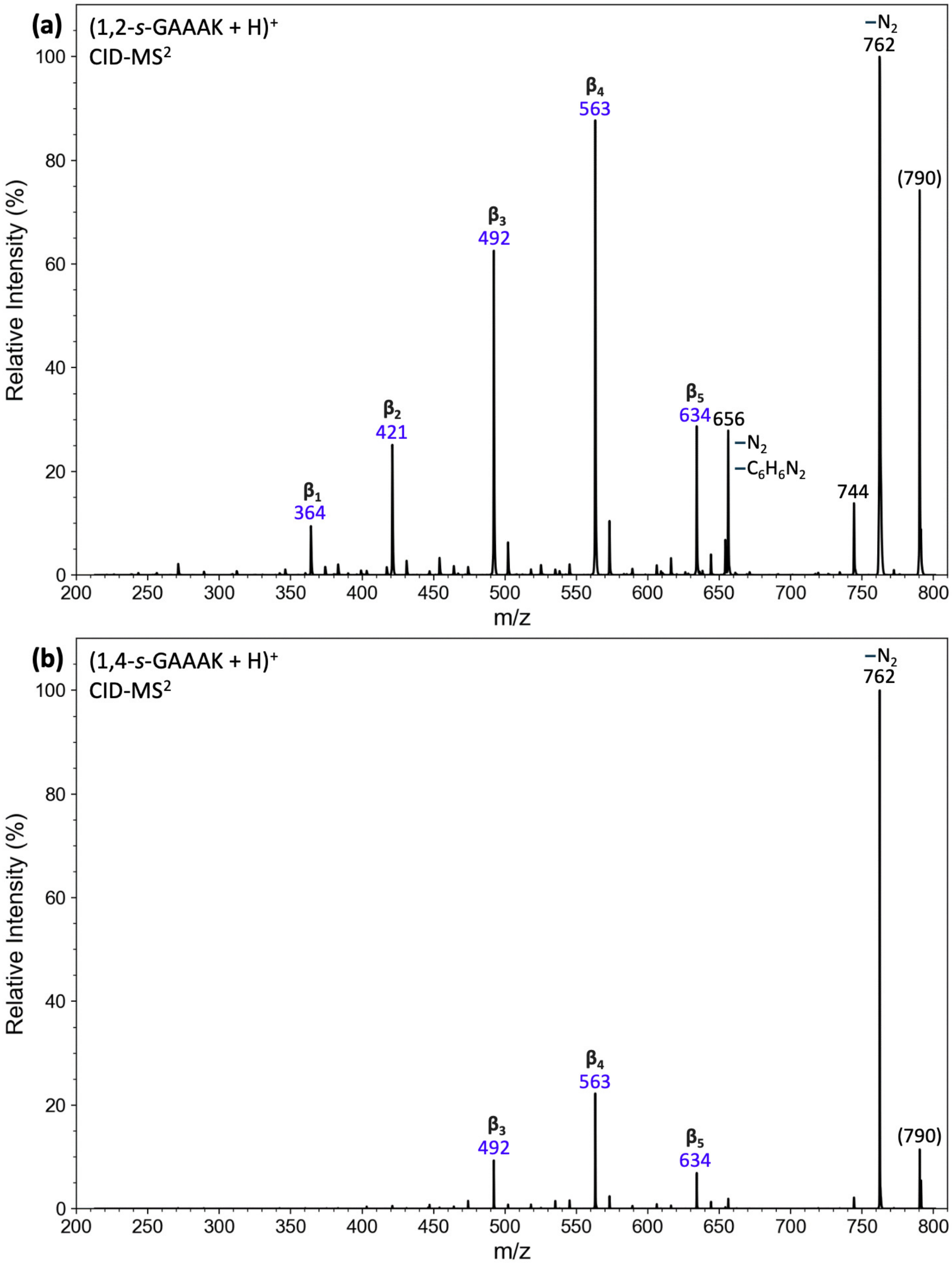


Figure 5.1: CID-MS^2 of (a) (1,2-*s*-GAAAK + H)$^+$ (*m/z* 790), and (b) (1,4-*s*-GAAAK + H)$^+$ (*m/z* 790). Fragment ion color-coding is as follows: Dark red: $\boldsymbol{b}_n$ and [$\boldsymbol{y}_n$ + 2H]$^+$ from the precursor ion; green: $\boldsymbol{b}_n$ after loss of N_2 and C_6H_5N; blue: β_n ions by loss of truncated residues from crosslinks after loss of N_2. The aminocyclohexane carboxylic acid is counted as the first residue in $\boldsymbol{b}_n$ and β_n ions.

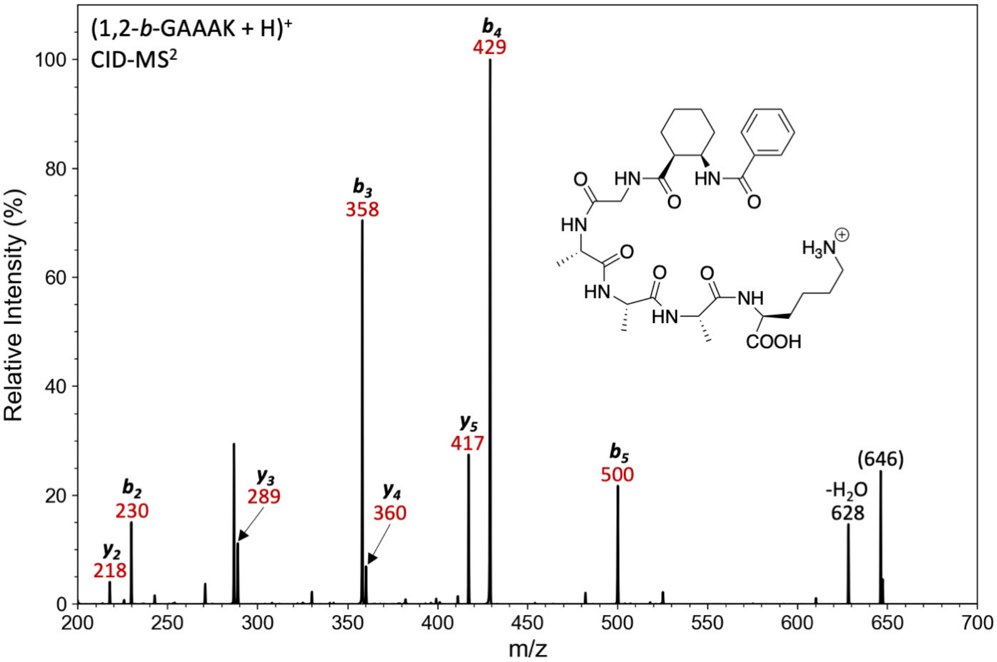


Figure 5.2: CID-MS^2 of (1,2-*b*-GAAAK+ H)$^+$ (*m/z* 646). For ion color coding see text in Figure 5.1.

was observed (*m/z* 671), but its subsequent backbone cleavage to form $\boldsymbol{b}_n$ ions was weak (*m/z* 525, 454, 383, Figure 5.5 (b)). CID-MS^2 of (1,4-s-GAAAK + H)$^+$ was similar to the UVPD spectrum, showing loss of N_2 and β_n ions (Figure 5.1 (b)). Further UVPD at 213 nm of (1,4-s-GAAAK + H – N_2)$^+$ resulted in deep dissociation, forming the $C_{14}H_9N_2O$ and $C_{14}H_{11}N_2O_2$ ions that contained the diarylnitrile imine and its oxygenated form, respectively (Figure 5.6 (b)). Prominent β_n and ($\boldsymbol{b}_n$ – C_6H_5N) ion series were also observed. We interpreted the UVPD and CID-MS^n data as being indicative of prevalent crosslinking of Lys to the transient nitrile imine.

5.3.1.4 Crosslinking Yields of 1,2-s-GAAAK and 1,4-s-GAAAK

Ions (1,2-s-GAAAK + H – N_2)$^+$ generated by UVPD at 250 and 213 nm exhibited high crosslinking yields upon CID analysis, at 94% and 93% respectively (Table 5.28). Both UVPD-MS^2 of (1,4-s-GAAAK + H)$^+$ and CID-MS^3 of (1,4-s-GAAAK + H – N_2)$^+$ yielded substantial crosslinking percentages, at 78% and 97%, respectively (Table 5.28). The crucial role of the carboxyl group in crosslinking was emphasized by comparing yields for (1,2-s-GAAAK-OCH_3 + H)$^+$ and (1,2-s-GAAAK + H)$^+$, where carboxyl methylation

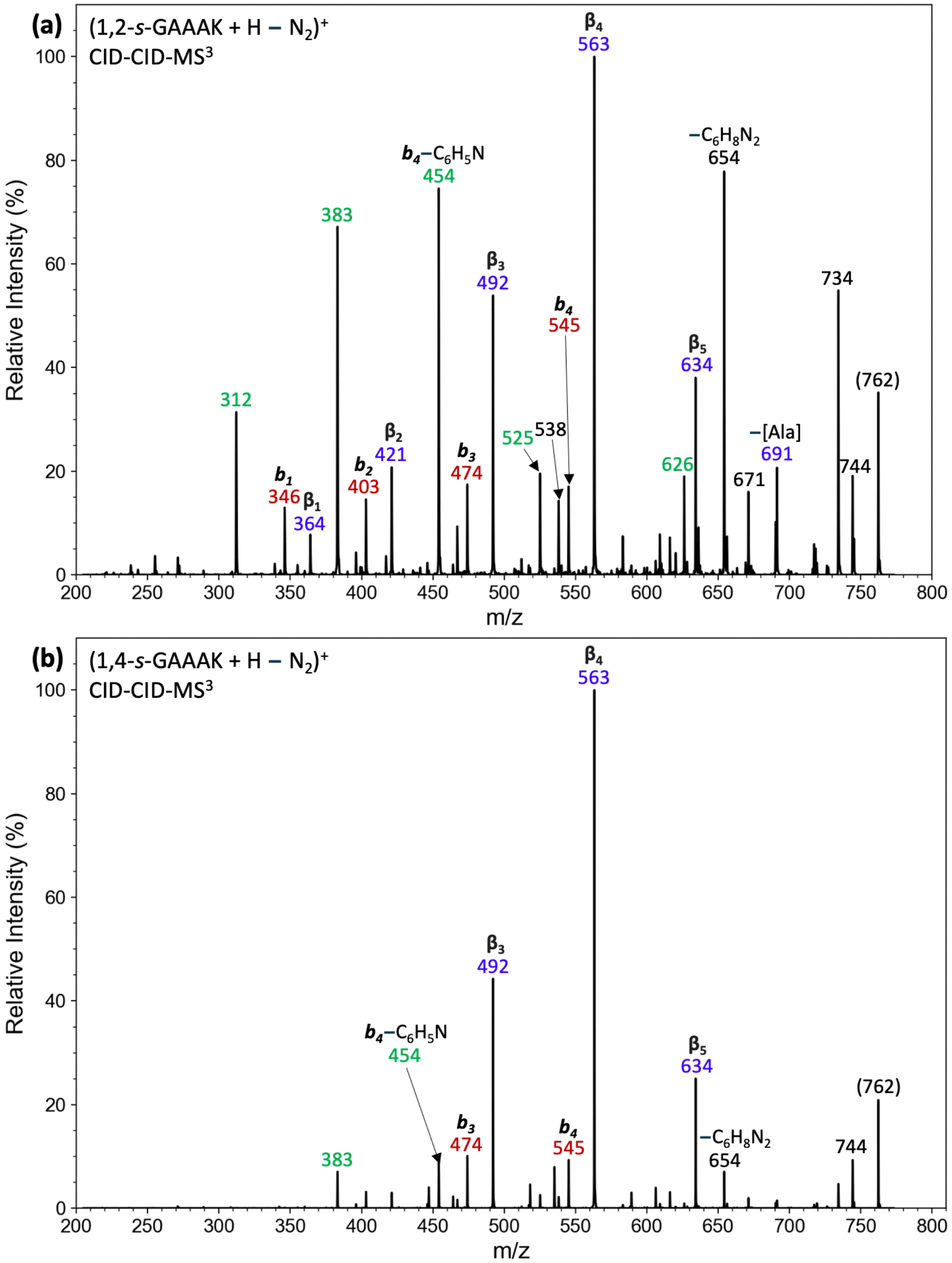


Figure 5.3: CID-MS^3 of (a) (1,2-*s*-GAAAK + H – N$_2$)$^+$ (*m/z* 762) generated by CID of (1,2-*s*-GAAAK + H)$^+$, and (b) (1,4-*s*-GAAAK + H – N$_2$)$^+$ (*m/z* 762) generated by CID of (1,4-*s*-GAAAK + H)$^+$. For ion color coding see text in Figure 5.1.

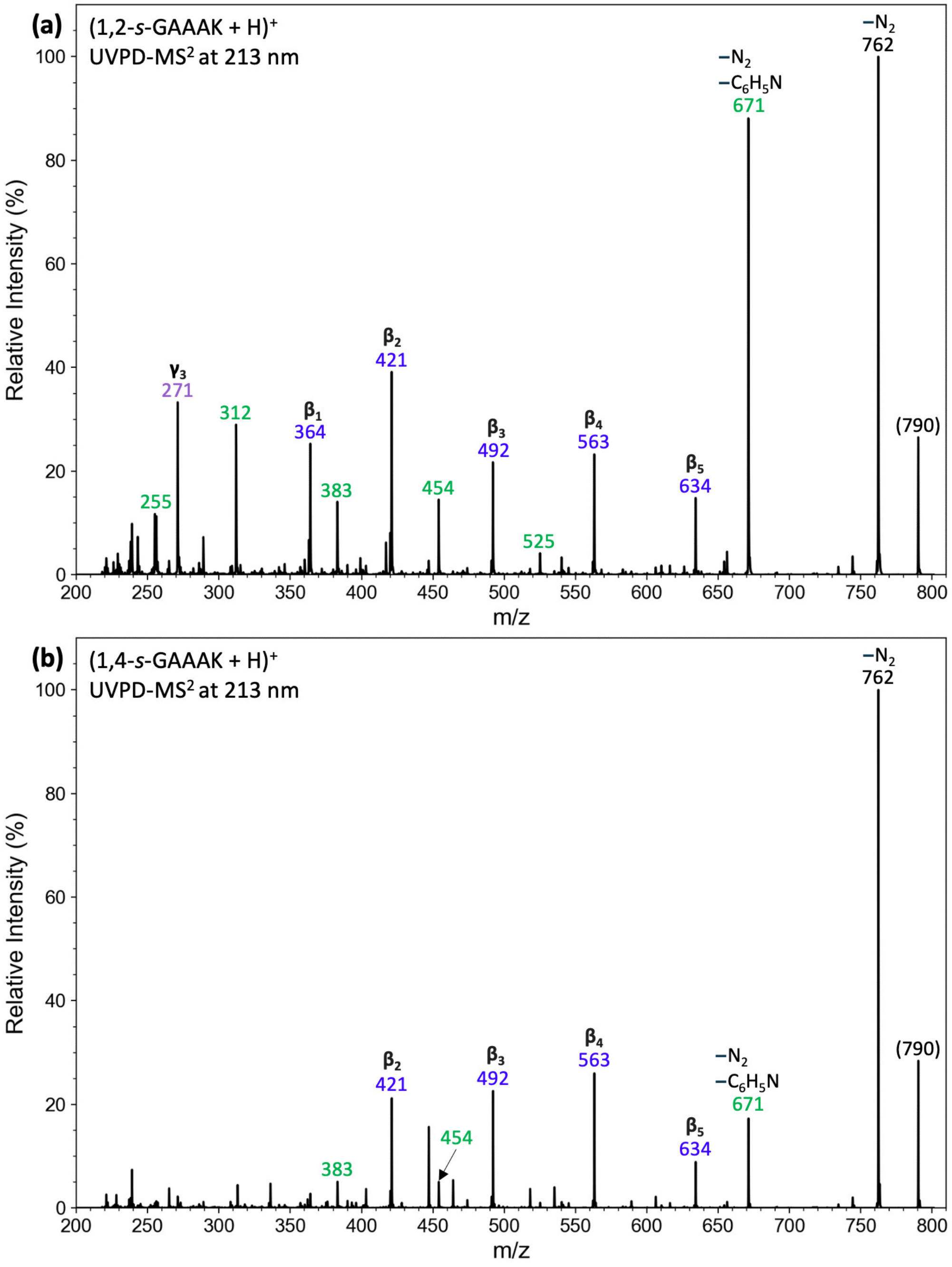


Figure 5.4: UVPD-MS^2 at 213 nm of (a) (1,2-*s*-GAAAK + H)$^+$ (*m/z* 790), and (b) (1,4-*s*-GAAAK + H)$^+$ (*m/z* 790). For ion color coding see text in Figure 5.1.

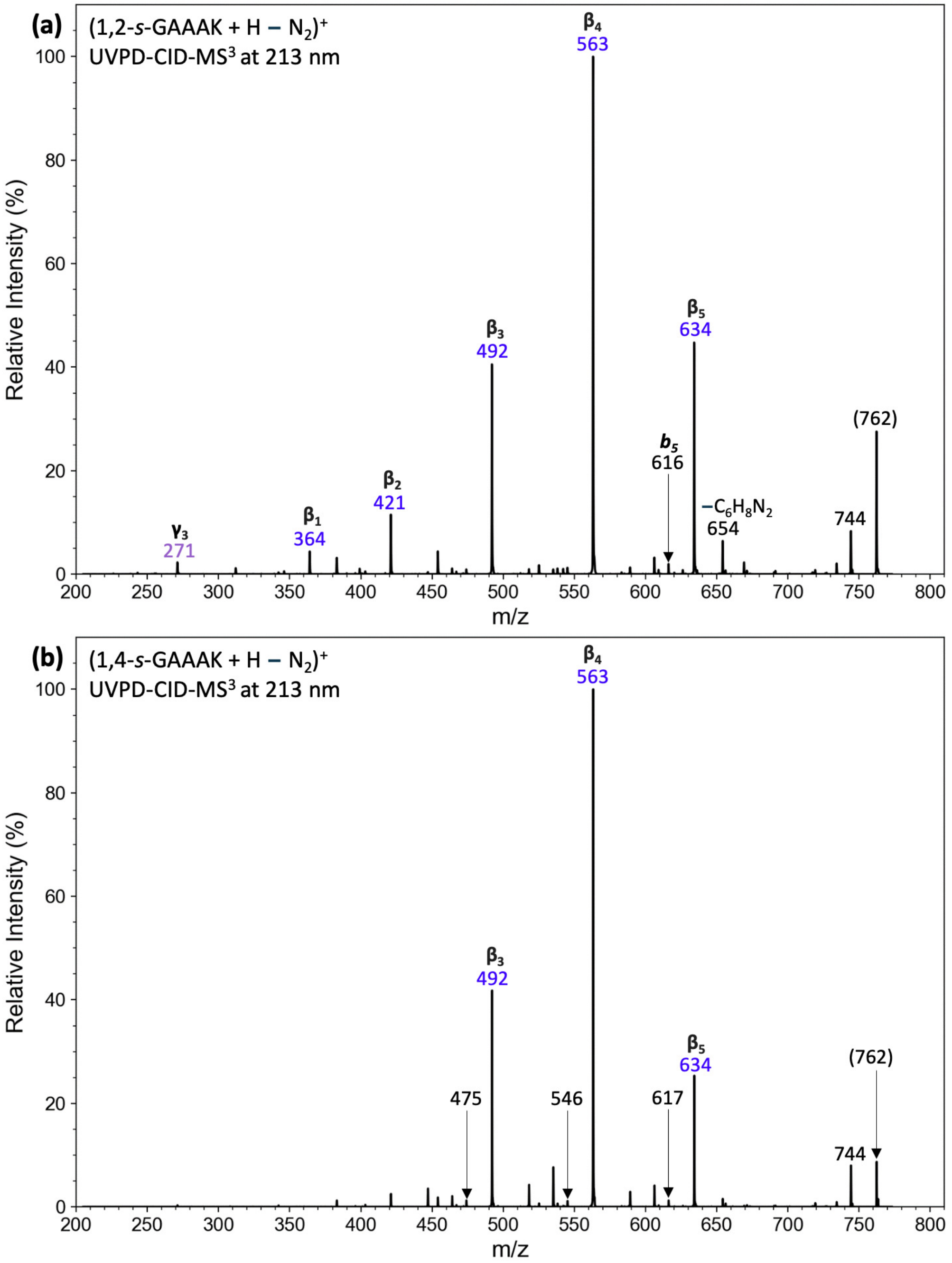


Figure 5.5: CID-MS^3 of (a) (1,2-*s*-GAAAK + H – N$_2$)$^+$ (*m/z* 762) generated by UVPD of (1,2-*s*-GAAAK + H)$^+$ at 213 nm, and (b) (1,4-*s*-GAAAK + H – N$_2$)$^+$ (*m/z* 762) generated by UVPD of (1,4-*s*-GAAAK + H)$^+$ at 213 nm. For ion color coding see text in Figure 5.1.

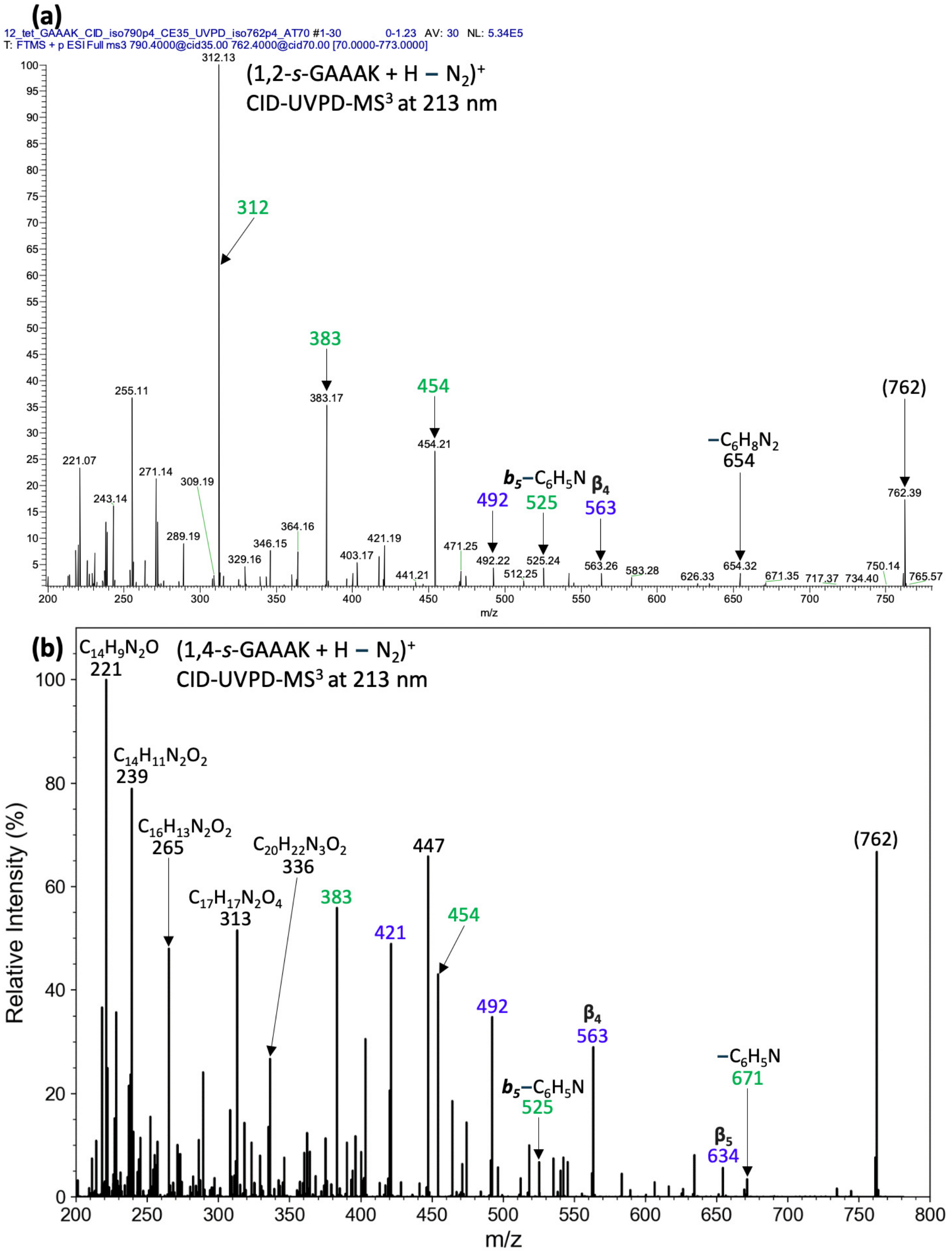


Figure 5.6: UVPD-MS^3 at 213 nm of (a) (1,2-*s*-GAAAK + H – N_2)$^+$ (*m/z* 762) generated by CID of (1,2-*s*-GAAAK + H)$^+$, and (b) (1,4-*s*-GAAAK + H – N_2)$^+$ (*m/z* 762) generated by CID of (1,4-*s*-GAAAK + H)$^+$. For ion color coding see text in Figure 5.1.

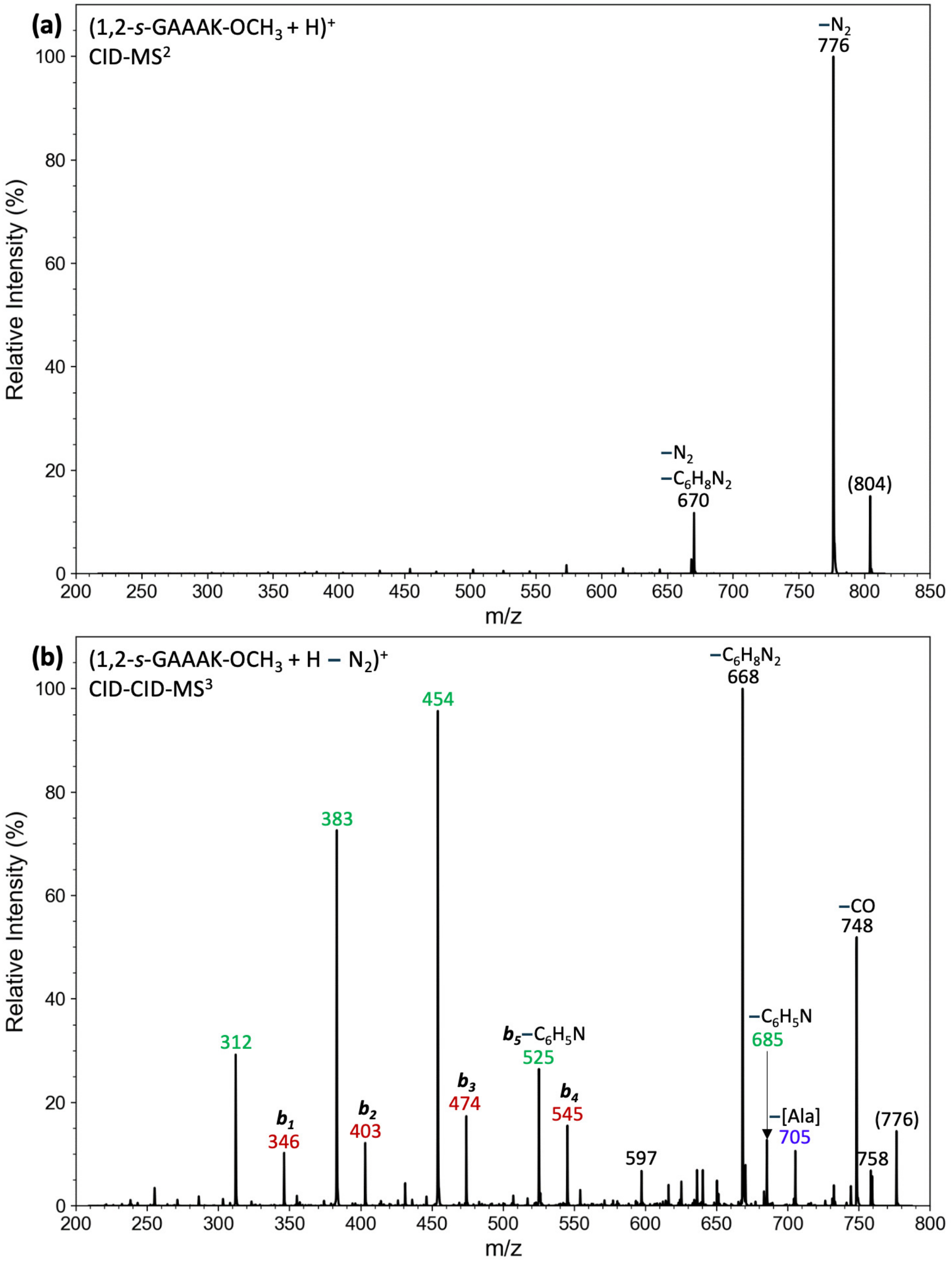


Figure 5.7: CID-MS^2 of (a) (1,2-*s*-GAAAK-OCH_3 + H)$^+$ (*m/z* 804), and (b) CID-MS^3 of (1,2-*s*-GAAAK-OCH_3 + H – N_2)$^+$ (*m/z* 776) generated by CID of (1,2-*s*-GAAAK-OCH_3 + H)$^+$. For ion color coding see text in Figure 5.1.

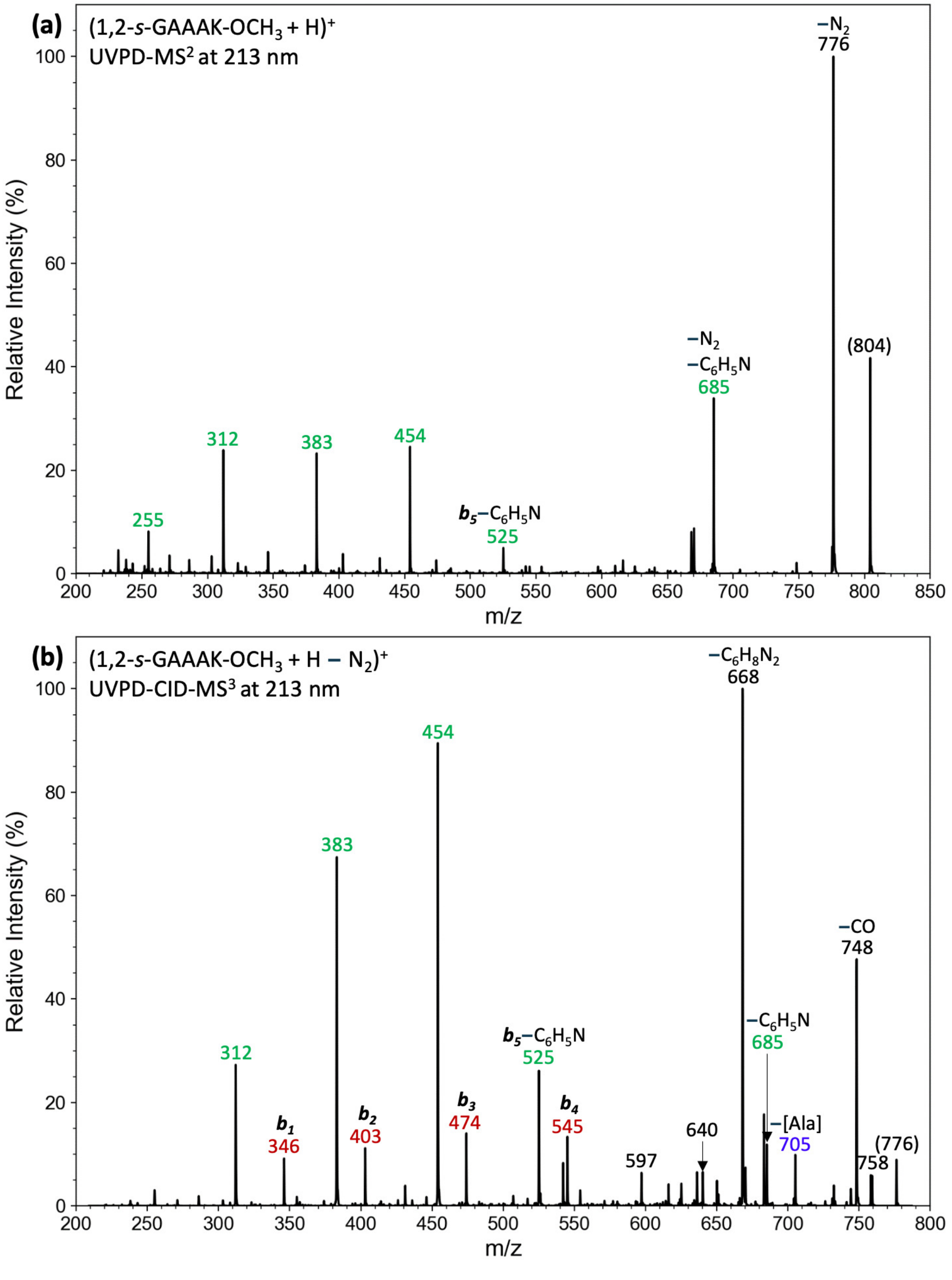


Figure 5.8: UVPD-MS^2 of (a) (1,2-*s*-GAAAK-OCH$_3$ + H)$^+$ (*m/z* 804) at 213 nm, and (b) CID-MS^3 of (1,2-*s*-GAAAK-OCH$_3$ + H – N$_2$)$^+$ (*m/z* 776) generated by UVPD at 213 nm of (1,2-*s*-GAAAK-OCH$_3$ + H)$^+$. For ion color coding see text in Figure 5.1.

notably reduced crosslinking yields in both UVPD-MS^2 and CID-MS^3 for (1,2-s-GAAAK-OCH$_3$ + H – N$_2$)$^+$ (15% and 22%, respectively). Interestingly, the reduction for (1,2-s-GAAAK-OCH$_3$ + H)$^+$ was not as pronounced as that seen for (1,2-s-AAAG-OCH$_3$ + H)$^+$ and (1,2-s-AAAHG-OCH$_3$ + H)$^+$ in Chapter 4, suggesting that a minor fraction of (1,2-s-GAAAK-OCH$_3$ + H)$^+$ may use the lysine amine group in reactions with the nitrile imine.

The significance of the lysine carboxyl in crosslinking was further supported by UVPD results for the methyl ester (1,2-s-GAAAK-OCH$_3$ + H)$^+$. UVPD of this ion (*m/z* 804), as well as UVPD-CID-MS^3 and CID-CID-MS^3 of the denitrogenated intermediate (1,2-s-GAAAK-OCH$_3$ + H – N$_2$)$^+$ (*m/z* 776), primarily produced the $\boldsymbol{b}_n$ ion series after C$_6$H$_5$N loss, with the β_n ions as minor products (Figure 5.8, Tables 5.6-5.9). Loss of C$_6$H$_5$NHNH$_2$ from (1,2-s-GAAAK-OCH$_3$ + H – N$_2$)$^+$ (*m/z* 668, Figure 5.8 (b)) indicated a favorable proton transfer to the nitrile imine group.

Table 5.28: Crosslinking Yields of Peptide-Nitrile Imine Scaffolds

Compound/Ion	Ion Activation	%(MH–N_2)[a]	%Crosslinks[a]
UVPD-MS^2			
1,2-s-GAAAK	250 nm	77	28
1,2-s-GAAAK	213 nm	32	66
1,4-s-GAAAK	213 nm	38	78
1,2-s-GAAAK-OCH_3	213 nm	48	15
1,2-s-GAAAR	213 nm	52	48
1,4-s-GAAAR	213 nm	72	58
1,2-s-GAAAR-OCH_3	213 nm	47	40
CID-MS^3			
1,2-s-GAAAK–N_2	250 nm		94
1,2-s-GAAAK–N_2	213 nm		93
1,4-s-GAAAK–N_2	213 nm		97
1,2-s-GAAAK-OCH_3–N_2	213 nm		22
1,2-s-GAAAR–N_2	213 nm		74
1,2-s-GAAAR–N_2	CID		74
1,4-s-GAAAR–N_2	213 nm		67
1,4-s-GAAAR–N_2	CID		62
1,2-s-GAAAR-OCH_3–N_2	213 nm		74
1,2-s-GAAAR-OCH_3–N_2	CID		78

[a] Relative to the sum of all identified fragment ion intensities.

5.3.2 Ion Mobility, Collision Cross Sections, and Ion Structures of 1,2-*s*-GAAAK and 1,4-*s*-GAAAK

The presence of the Lys residue in these conjugates steered protonation to the basic side-chain ϵ-amino group in all low-energy isomers. Internal solvation of the ϵ-NH_3^+ group then governed the peptide chain folding, as shown for the representative (1,2-*s*-GAAAK + H)$^+$ structures **1,2-K1** and **1,2-K2** (Figure 5.9). Both **1,2-K1** and **1,2-K2** showed peptide hydrogen bonds to the benzamide carbonyl and chiefly differed in the hydrogen bonding of the C-terminal carboxyl and orientation of the phenyltetrazole moiety. These low-energy conformers had CCS_{calc} that closely matched those of the major peaks in the ion mobilogram. We did not find adequate CCS matches for the two minor peaks in the Figure 5.9 mobilogram that corresponded to more compact structures at CCS_{IMS} = 280 and 282 Å^2.

The low-energy conformers of (1,4-*s*-GAAAK + H)$^+$ (**1,4-K1** and **1,4-K2**) had cyclohexane twistboat conformations that allowed the ions to achieve efficient hydrogen bonding between the peptide moiety and benzamide carbonyl, in particular regarding internal solvation of the ϵ-NH_3^+ group by the benzamide in **1,4-K1** (Figure 5.10). We note that the typical Gibbs energy difference between the cyclohexane twist-boat and chair conformations has been reported as 22-24 kJ mol^{-1},[27],[28] which can be readily overcome by the binding energy of a strong intramolecular hydrogen bond. The **1,4-K1** ion of CCS_{calc} = 296 Å^2, which was more compact and had a lower Gibbs energy, also developed a hydrogen bond between the carboxyl and the tetrazole ring. Whether or not that may be conducive to carboxyl crosslinking in the incipient nitrile imine will be discussed later in the paper. The CCS_{calc} of **1,4-K1** and **1,4-K2** showed the best match with the CCS_{IMS} of the two resolved peaks, 296 Å^2 and 304 Å^2, in the Figure 5.10 ion mobilogram.

Ion mobility separation of the CID-denitrogenated ions (1,2-*s*-GAAAK + H – N_2)$^+$ (Figure 5.11) showed two major peaks of CCS_{IMS} = 274 and 276 Å^2, respectively, along with a very minor peak at CCS_{IMS} = 275 Å^2. We considered several types of ion structures, both nitrile imines and crosslinked ones, that were evaluated by their relative energy and CCS_{calc}. The representative lowest-energy conformers in each group are shown in Figure 5.11. The carboxyl-crosslinked isomer **1,2-k1** of CCS_{calc} = 273 Å^2, that was the global energy minimum among the (1,2-*s*-GAAAK + H – N_2)$^+$ ions, showed a close match with the CCS_{IMS} of the first major peak in the mobilogram. A lowenergy isomer crosslinked by the Lys NH_2 (**1,2-k2**) of CCS_{calc} = 274 Å^2 was close to the second major peak in the (1,2-*s*-GAAAK + H – N_2)$^+$ mobilogram.

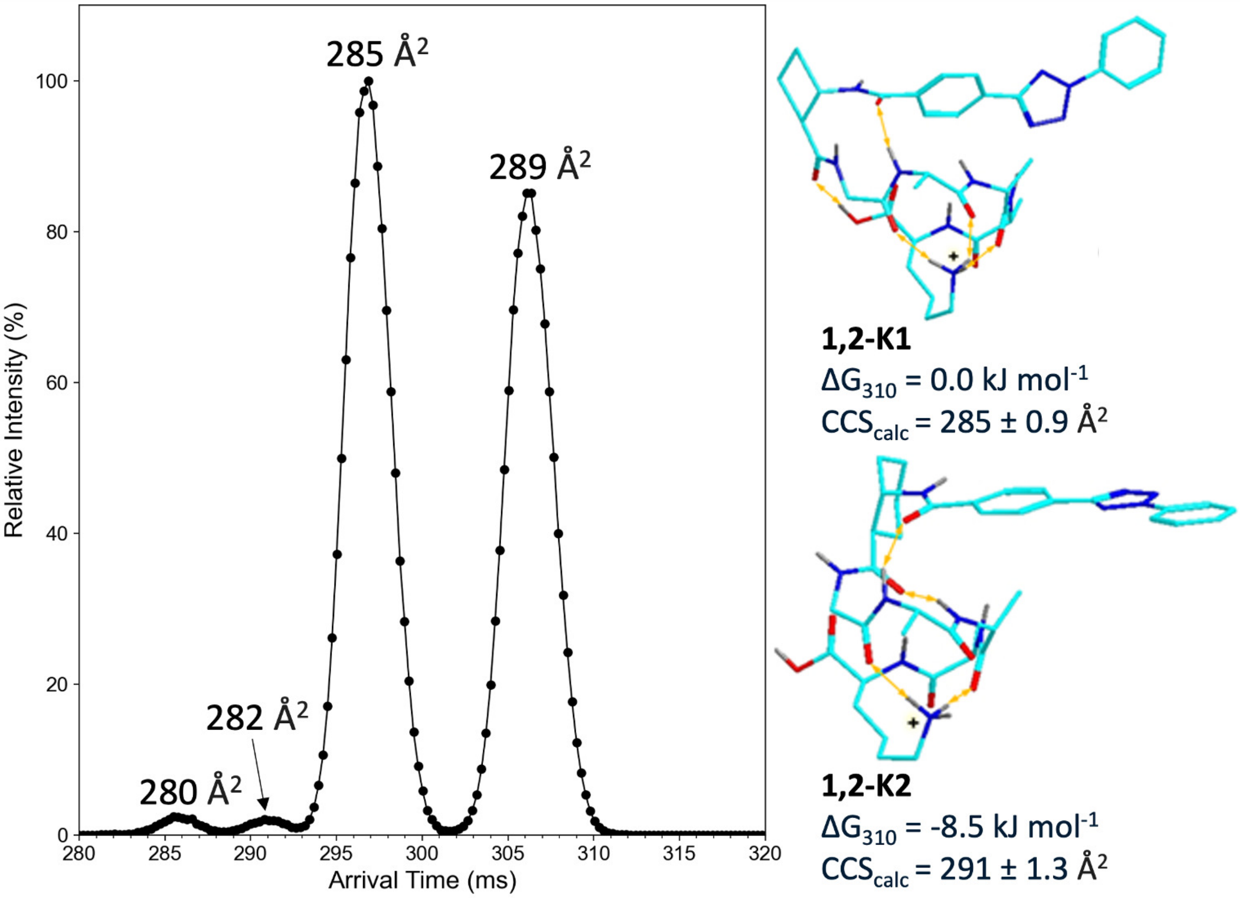


Figure 5.9: Arrival time distributions of (1,2-*s*-GAAAK + H)$^+$ after 10 passes. M06-2X/6-31+G(d,p) optimized structures, M06-2X/def2qzvpp relative Gibbs energies, and CCS$_{\text{calc}}$ of low energy (1,2-*s*-GAAAK + H)$^+$ ions. Atom color coding is as follows: cyan = C, blue = N, red = O, gray = H. Only exchangeable NH, OH hydrogens are shown to avoid clutter. Major hydrogen bonds are indicated by ochre double-headed arrows.

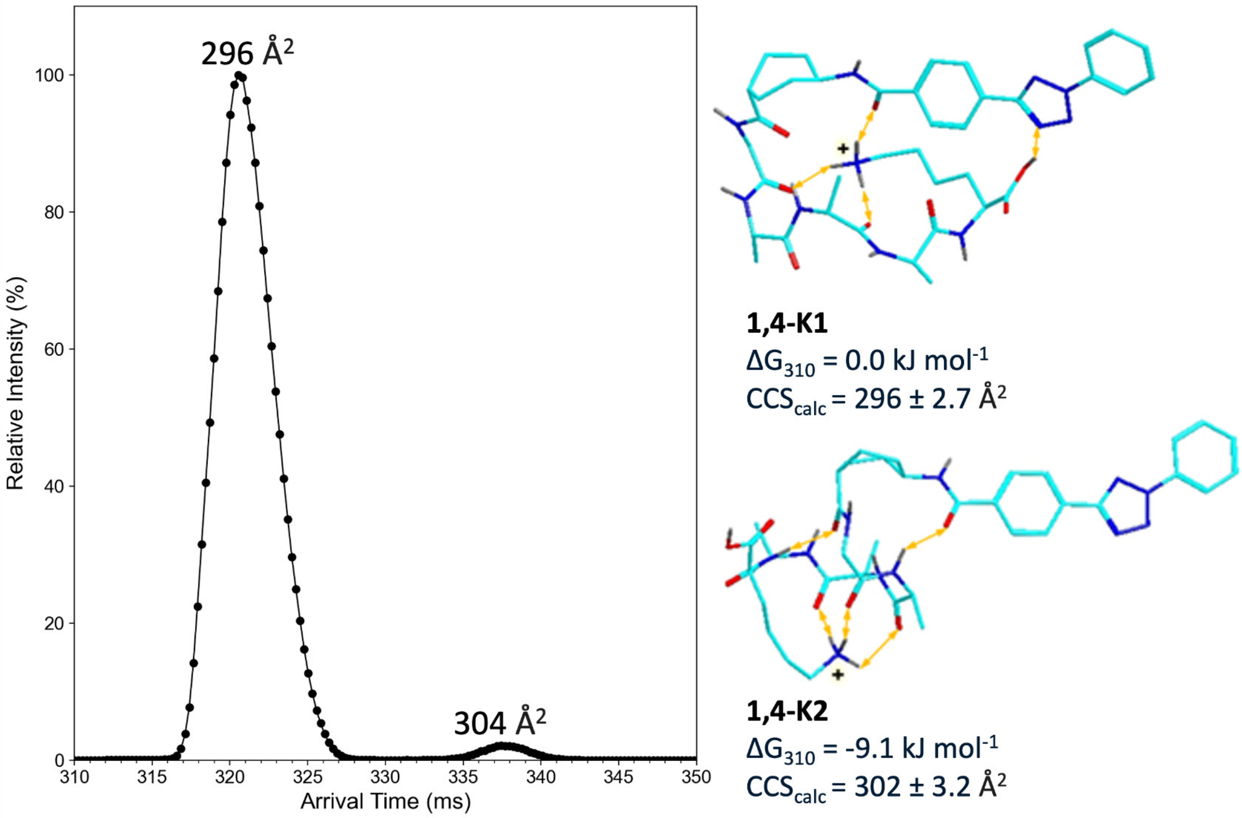
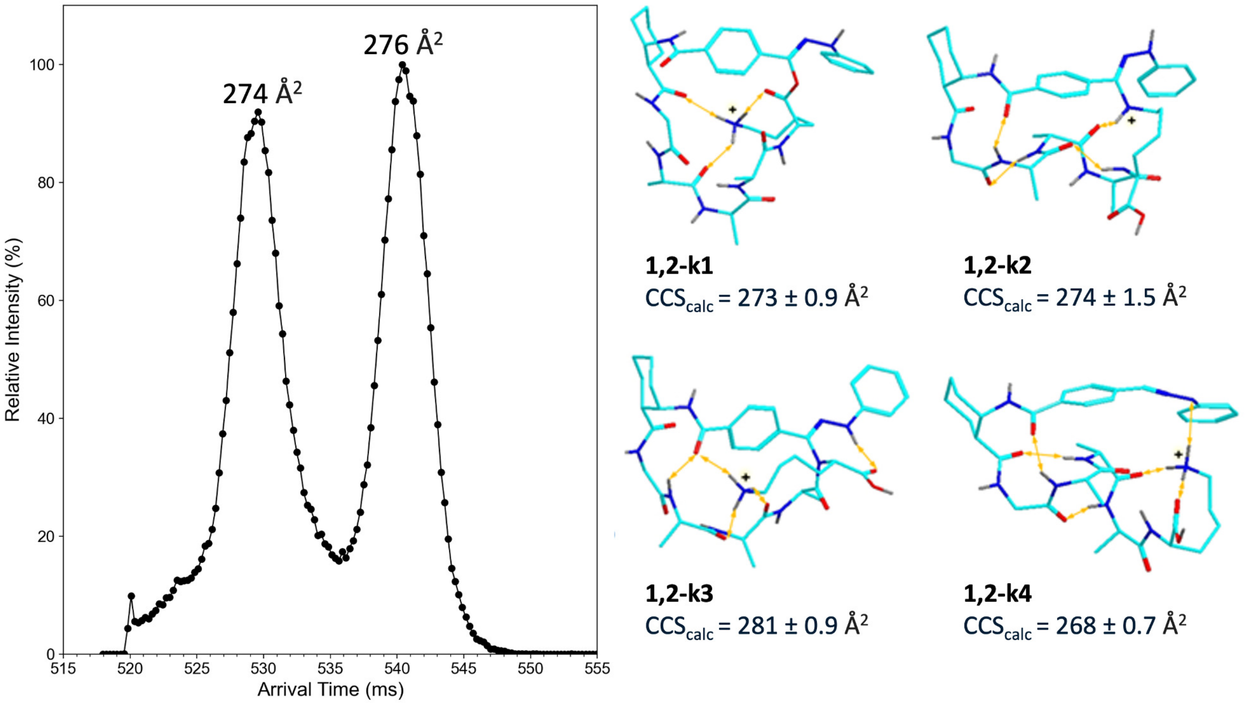


Figure 5.10: Arrival time distributions of (1,4-*s*-GAAAK + H)$^+$ after 10 passes. M06-2X/6-31+G(d,p) optimized structures, M06-2X/def2qzvpp relative Gibbs energies, and CCS$_{\text{calc}}$ of low energy (1,4-*s*-GAAAK + H)$^+$ ions. Structure discription as in Figure 5.9.



1,2-k1
CCS_{calc} = 273 ± 0.9 Å²

1,2-k2
CCS_{calc} = 274 ± 1.5 Å²

1,2-k3
CCS_{calc} = 281 ± 0.9 Å²

1,2-k4
CCS_{calc} = 268 ± 0.7 Å²

Figure 5.11: Arrival time distributions of (1,2-*s*-GAAAK + H – N_2)$^+$ after 20 passes. M06-2X/6-31+G(d,p) optimized structures and CCS_{calc} of low energy isomers **1,2-k1**-**1,2-k4**. Structure discription as in Figure 5.9.

Considering the CID-MS^3 spectrum of (1,2-*s*-GAAAK + H – N_2)$^+$ (Figure 5.5 (a)) that indicated carboxyl crosslinking, the presence of structures of the **1,2-k1** type in the stable ion population was possible and was consistent with the matching CCS. The presence of Lys-NH_2 crosslinked ions was also possible, as indicated by the matching CCS. Interestingly, no close CCS match was found for the nitrile-imine structures, such as **1,2-k4** of CCS_{calc} = 268 Å² (Figure 5.11). The lack of residual nitrile imine structures was consistent with the Figure 5.5 (a) spectrum and the 93-94% crosslinking yield for (1,2-*s*-GAAAK + H – N_2)$^+$ (Table 5.28).

Ion mobility of CID-denitrogenated ions (1,4-*s*-GAAAK + H – N_2)$^+$ (Figure 5.12) showed at least five fully and partially resolved components. The most abundant peak at CCS_{IMS} = 272 Å² found a close match with the carboxyl-crosslinked isomer **1,4-k1** of CCS_{calc} = 272 Å² (Figure 5.12). Another carboxyl-crosslinked conformer (**1,4-k2** of CCS_{calc} = 282 Å²) could fit one of the peaks with CCS_{IMS} = 281 or 288 Å², although this assignment was uncertain. The nitrile imine isomers were more extended and had larger CCS_{calc}, e.g. for **1,4-k3**, that could be accounted for by the weak peak at CCSIMS = 294 Å² in the mobilogram. The assignment of structure **1,4-k1** was consistent with the CID-MS^3 spectrum of (1,4-*s*-

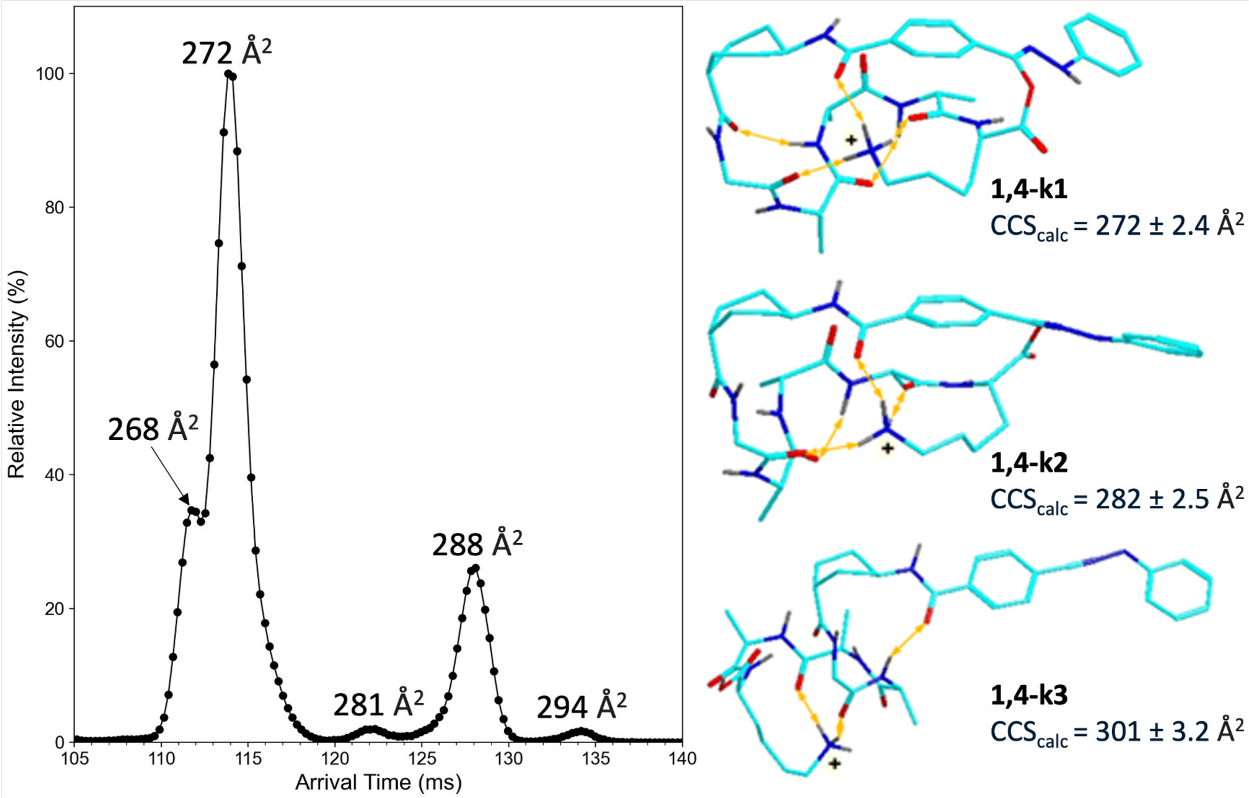


Figure 5.12: Arrival time distributions of (1,4-*s*-GAAAK + H – N_2)$^+$ after 4 passes. M06-2X/6-31+G(d,p) optimized structures and CCS$_{calc}$ of low energy isomers **1,4-k1**-**1,4-k3**. Structure discription as in Figure 5.9.

GAAAK + H – $N_2)^+$ (Figure 5.5 (b)) and the high crosslinking yield for this scaffold (97%, Table 5.28). A salient feature of structures **1,4-k1**–**1,4-k3** was that they all had the cyclohexane ring in a twistboat conformation that allowed for crosslinking in **1,4-k1** and **1,4-k2** but also favored a peptide hydrogen bond to the benzamide carbonyl in **1,4-k3**.

5.3.3 Photodissociation and Crosslinking Yields of 1,2-*s*-GAAAR and 1,4-*s*-GAAAR

5.3.3.1 Spectra Interpretation of 1,2-*s*-GAAAR

UVPD at 213 nm of (1,2-s-GAAAR + H)$^+$ (*m/z* 818) showed some new features that were specific for Arg (Figure 5.15 (a), Table 5.14). In addition to the β_n ion series (*m/z* 563, 492, 421, and 364) and combined loss of N_2 and C_6H_5N (*m/z* 699), there was a prominent ion by loss of $C_7H_7NO_2$ (*m/z* 653). CID-MS^3 of the denitrogenated ion (1,2-s-GAAAR–$N_2)^+$ (*m/z* 790, Table 5.15) showed a similar fragmentation pattern that was dominated by the β_n ion series and the *m/z* 762 (loss of CO) and *m/z* 653 ions (Figure 5.16 (a)). Ion activation by the CID-MS^2 and CID-CID-MS^3 sequences gave similar results (Figure 5.13 (a) and 5.14 (a), Table 5.16 and 5.17). The loss of $C_7H_7NO_2$ pointed to carboxyl crosslinking to the nitrile imine phenyl group that was unprecedented with other peptide sequences. This was further studied by CID-MS^4 of the *m/z* 762 (loss of CO) ion that gave a complete β_n ion series but no *m/z* 653 (Figure 5.19, Table 5.29), indicating that the $C_7H_7NO_2$ neutral fragment was eliminated in a single step as a distinct moiety containing the nitrile imine phenyl linked to the carboxyl group and including two hydrogen atoms transferred from the peptide chain. An interesting feature of the Figure 5.19 spectrum was the elimination of an arginine $C_5H_{12}N_4$ neutral fragment, *m/z* 762.4029 → *m/z* 634.2971, starting the β_4 sequence which clearly indicated crosslinking by the arginine C-terminus. This was further investigated by blocking the free carboxyl as a methyl ester in 1,2-s-GAAAR-OCH_3 (Table 5.18 and 5.19). CID-MS^3 of the denitrogenated ion (1,2-s-GAAAR-OCH_3 – $N_2)^+$, *m/z* 804, Figure 5.17 (b), Table 5.17) displayed three major fragment ion series. The classical $\boldsymbol{b}_n$ ions (*m/z* 346, 403, 474, and 545) indicated a fraction of non-crosslinked nitrile imine isomers that were analogous to those in the spectrum of the 1,2-s-GAAAK derivative (1,2-s-GAAAK-OCH_3–$N_2)^+$ (Figure 5.8 (b), Table 5.9). In contrast to (1,2-s-GAAAK-OCH_3–$N_2)^+$, the Arg methyl ester (1,2-s-GAAAR-OCH_3–$N_2)^+$ showed a major loss of ammonia butvirtually no loss of C_6H_5N, which had a major effect on the Figure 5.18 (b) spectrum. The ammonia loss was followed by internal fragmentations leading to a sequential loss of two [Ala] units. This indicated cyclic peptide structures, although it was not clear if they were present in the (1,2-s-GAAAR-OCH$_3$–N$_2$)$^+$ ion population or produced upon CID by the ammonia loss. Finally, the CID-MS^3 spectrum of (1,2-s-GAAAR-OCH$_3$–N$_2$)$^+$ displayed a series of fragment ions starting with the elimination of $C_7H_{13}N_3O_2$ (*m/z* 633) from the Arg-OCH$_3$ residue and followed by sequential losses of [Ala] and [Gly] units (*m/z* 562, 491, 420, and 363, Figure 5.18 (b)), forming a new β_n ion series. Note that these fragment ions had different formulas (Table 5.20) than those in the Figure 5.16 (a) spectrum, which indicated crosslinking by an NH$_2$ group of the Arg side-chain, followed by N–C bond cleavage leaving the guanidine NH$_2$ group incorporated in the β_4 ions. Guanidine crosslinking was also evident from the small effect that carboxyl blocking had on the crosslinking yields of 1,2-s-GAAAR-OCH$_3$ and ion (1,2-s-GAAAR-OCH$_3$–N$_2$)$^+$.

5.3.3.2 Spectra Interpretation of 1,4-*s*-GAAAR

The GAAAR peptide sequence was investigated as to its propensity for nitrile imine crosslinking when mounted on the trans-1,4-cyclohexane scaffold. Both UVPD and CID-MS^2 of (1,4-s-GAAAR + H)$^+$ (*m/z* 818, Figure 5.15 (b) and 5.13 (b), Table 5.22 and 5.23) resulted in loss of N$_2$, and the resulting ions (1,4-s-GAAAR + H – N$_2$)$^+$ (*m/z* 790) were subjected to further investigations. CID-MS^3 of 1,4-s-GAAAR + H – N$_2$)$^+$ (Figure 5.16 (b) and 5.14 (b), Table 5.24 and 5.25) showed losses of NH$_3$, CO, and $C_7H_7NO_2$ at *m/z* 773, 762, and 653, respectively, that were analogous to the dissociations of (1,2-s-GAAAR + H – N$_2$)$^+$. In addition, we observed a β_4 ion series (*m/z* 634, 563, 492, and 421) that was indicative of crosslinks, and a $\boldsymbol{b}_n$ ion series (*m/z* 545, 474, 403) that were formed from the fraction of linear chain isomers. The loss of $C_7H_7NO_2$ (*m/z* 653) was indicative of a proton and carboxyl transfer to the nitrile imine C_6H_5N moiety and was investigated by CID-MS^4, as detailed in the Supporting Information (Figure 5.20, Table 5.30).

4.3.1.4 Crosslinking Yields of 1,2-s-GAAAR and 1,4-s-GAAAR

The Table 5.28 data showed that the crosslinking yield upon UVPD of 1,2-s-GAAAR dropped was 48% and by only 8% upon carboxyl methylation in 1,2-s-GAAAR-OCH$_3$ and virtually did not change in CID of (1,2-s-GAAAR-OCH$_3$–N$_2$)$^+$, amounting to 74-78% for ions generated by UVPD or CID-MS^2. Compared with the sterically more favorable (1,2-s-GAAAR + H – N$_2$)$^+$ which was 74% under UVPD-CID-MS^3, ion

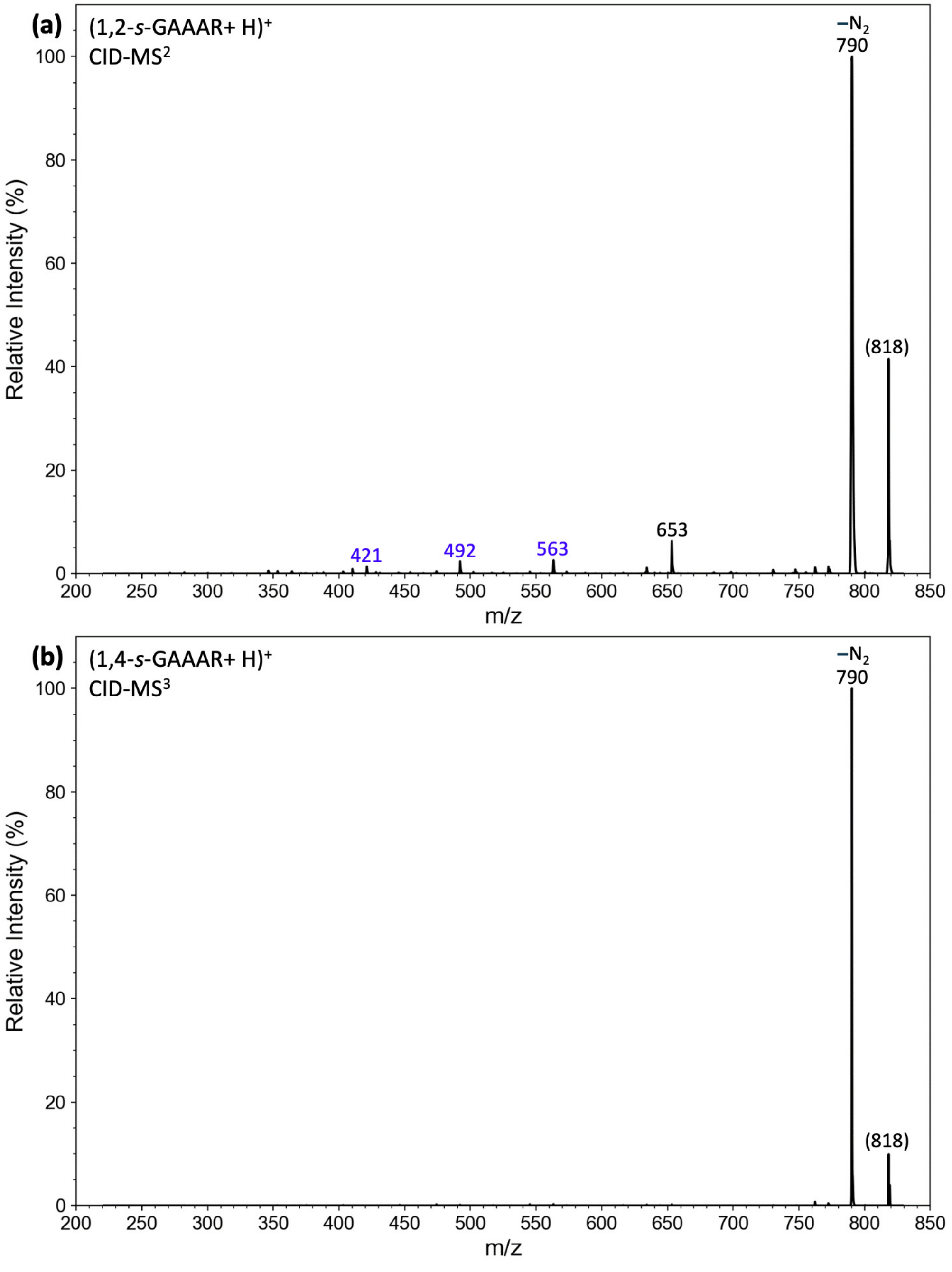


Figure 5.13: CID-MS^2 of (a) (1,2-*s*-GAAAR + H)$^+$ (*m/z* 818), and (b) (1,4-*s*-GAAAR + H)$^+$ (*m/z* 818). For ion color coding see text in Figure 5.1.

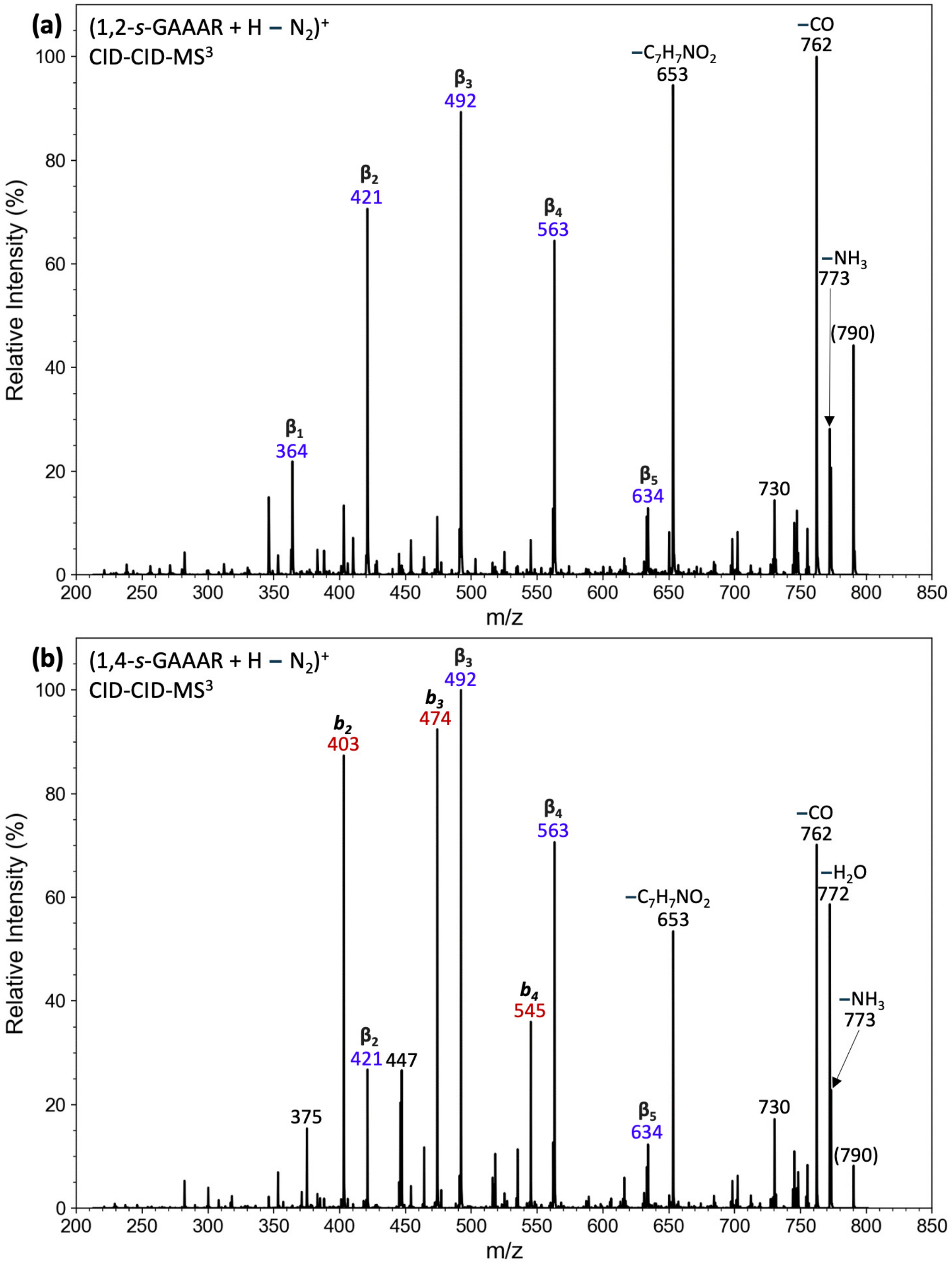


Figure 5.14: CID-MS^3 of (a) (1,2-s-GAAAR + H – N$_2$)$^+$ (m/z 790) generated by CID of (1,2-s-GAAAR + H)$^+$, and (b) (1,4-s-GAAAR + H – N$_2$)$^+$ (m/z 790) generated by CID of (1,4-s-GAAAR + H)$^+$. For ion color coding see text in Figure 5.1.

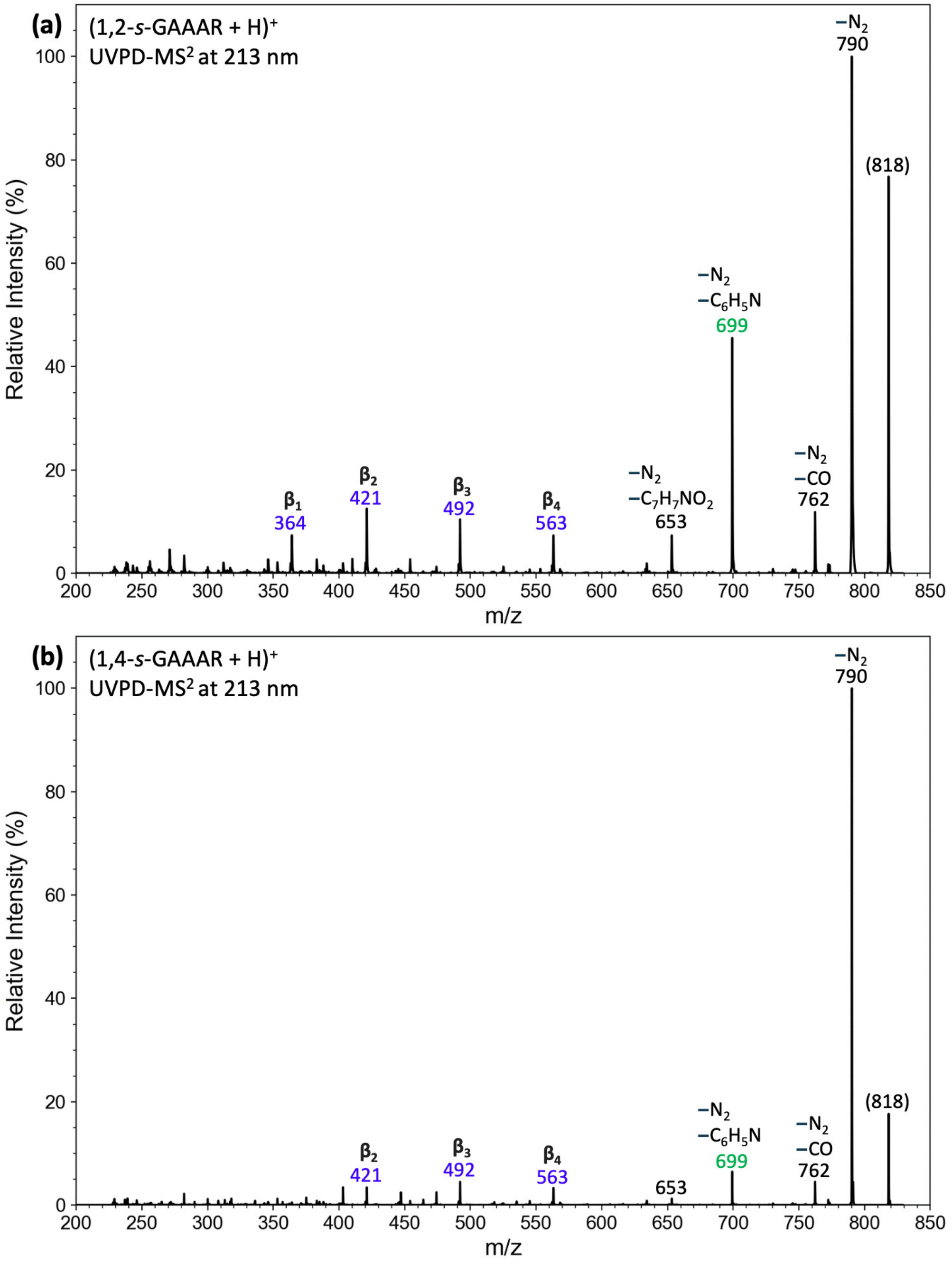


Figure 5.15: UVPD-MS^2 at 213 nm of (a) (1,2-*s*-GAAAR + H)$^+$ (*m/z* 818), and (b) (1,4-*s*-GAAAR + H)$^+$ (*m/z* 818). For ion color coding see text in Figure 5.1.

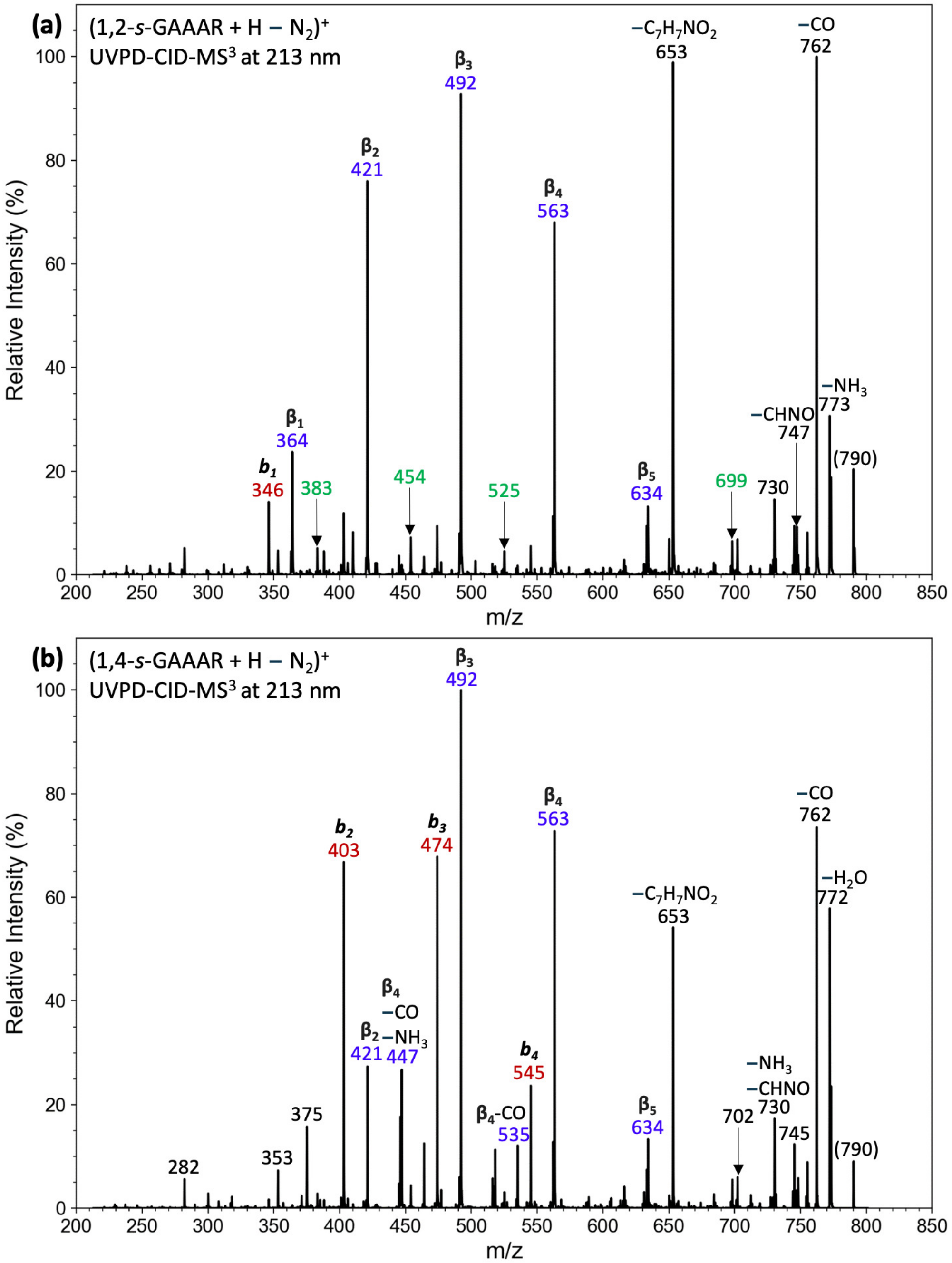


Figure 5.16: CID-MS^3 of (a) (1,2-*s*-GAAAR + H – $N_2)^+$ (*m/z* 790) generated by UVPD at 213 nm of (1,2-*s*-GAAAR + H$)^+$, and (b) (1,4-*s*-GAAAR + H – $N_2)^+$ (*m/z* 790) generated by UVPD at 213 nm of (1,4-*s*-GAAAR + H$)^+$. For ion color coding see text in Figure 5.1.

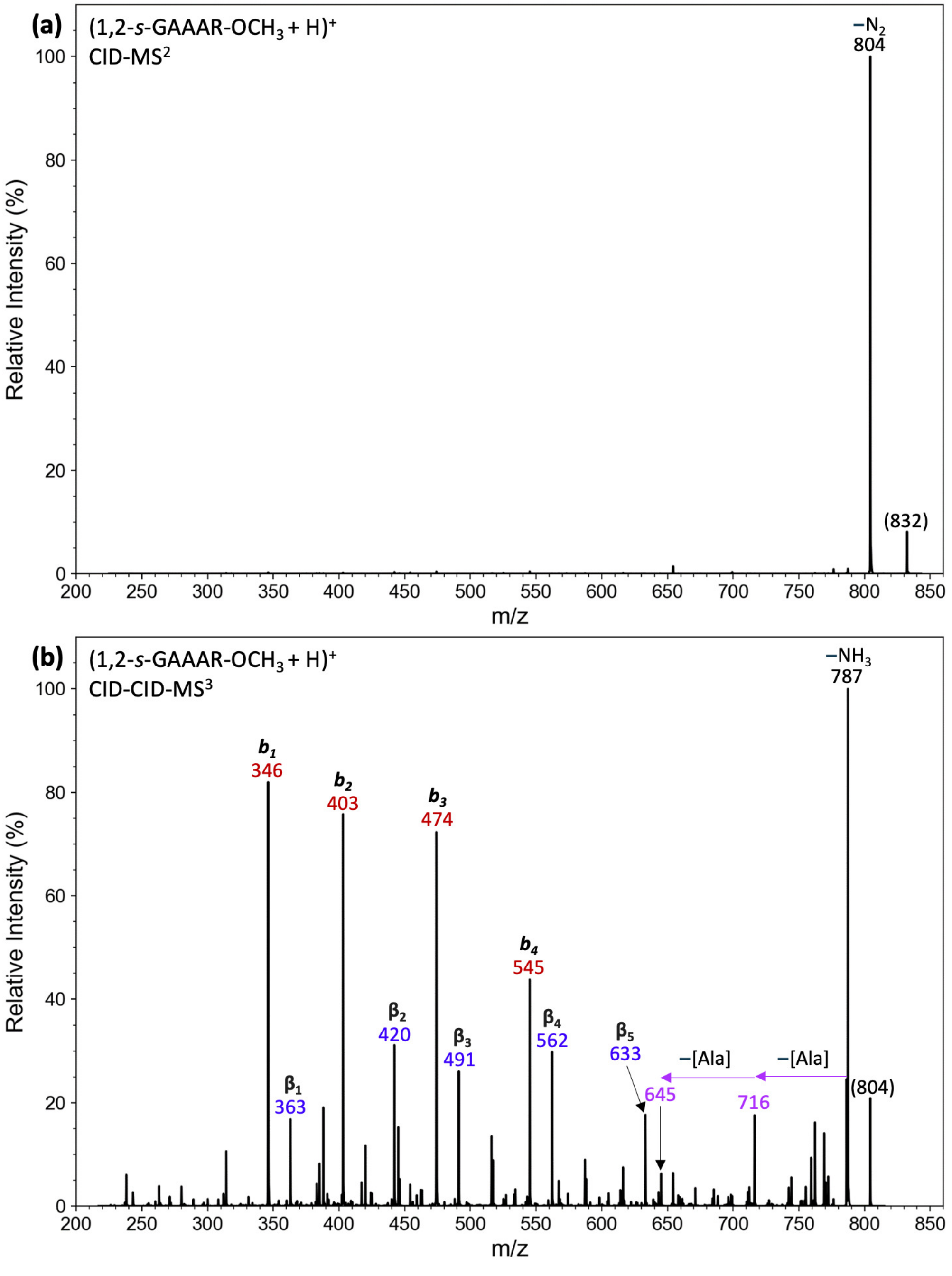


Figure 5.17: CID-MS^2 of (a) (1,2-*s*-GAAAR-OCH$_3$ + H)$^+$ (*m/z* 832), and (b) CID-MS^3 of (1,2-*s*-GAAAR-OCH$_3$ + H – N$_2$)$^+$ (*m/z* 804) generated by CID of (1,2-*s*-GAAAR-OCH$_3$ + H)$^+$. For ion color coding see text in Figure 5.1.

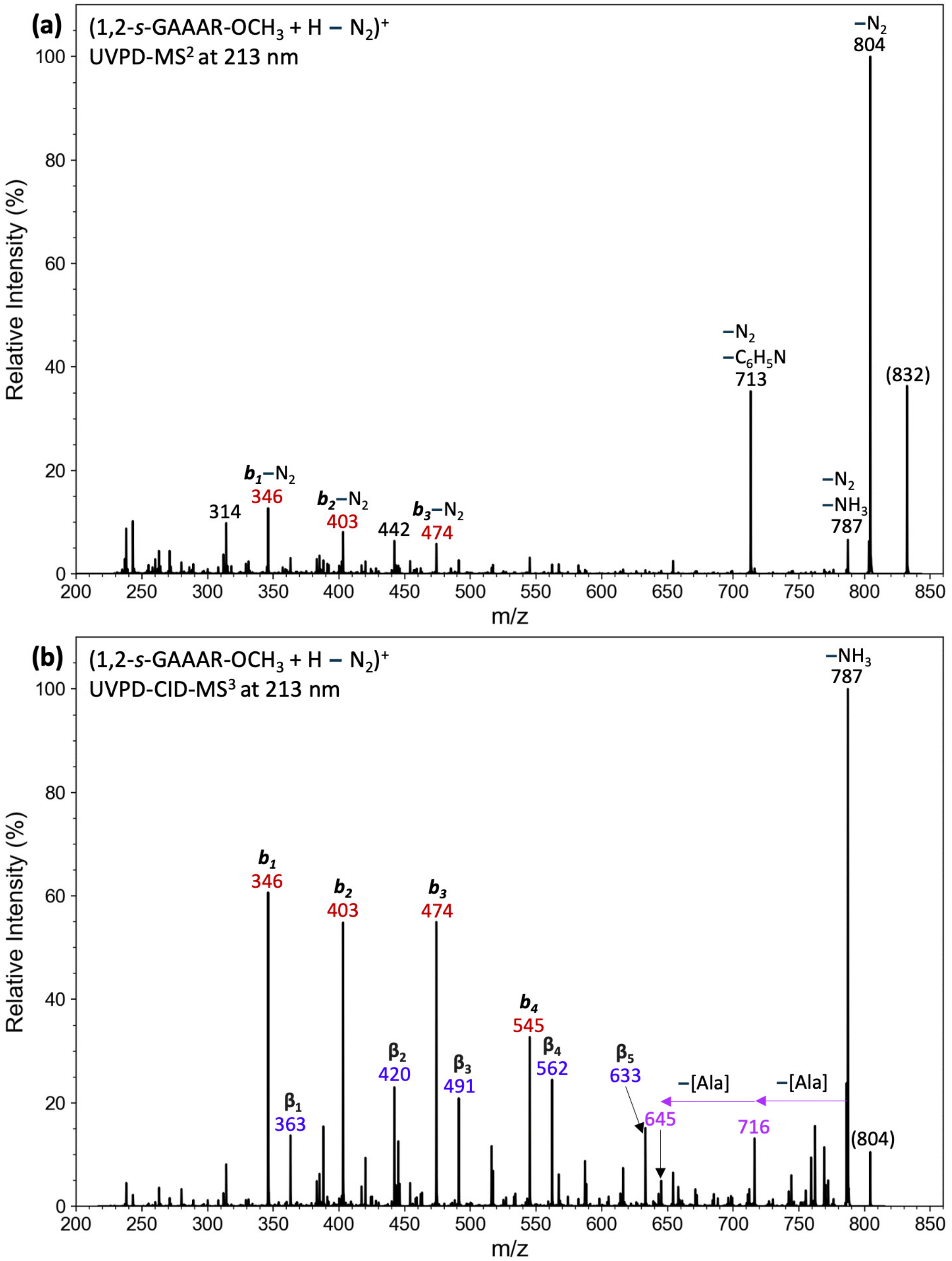


Figure 5.18: UVPD-MS^2 of (a) (1,2-*s*-GAAAR-OCH$_3$ + H)$^+$ (*m/z* 832) at 213 nm, and (b) CID-MS^3 of (1,2-*s*-GAAAR-OCH$_3$ + H – N$_2$)$^+$ (*m/z* 804) generated by UVPD at 213 nm of (1,2-*s*-GAAAR-OCH$_3$ + H)$^+$. For ion color coding see text in Figure 5.1.

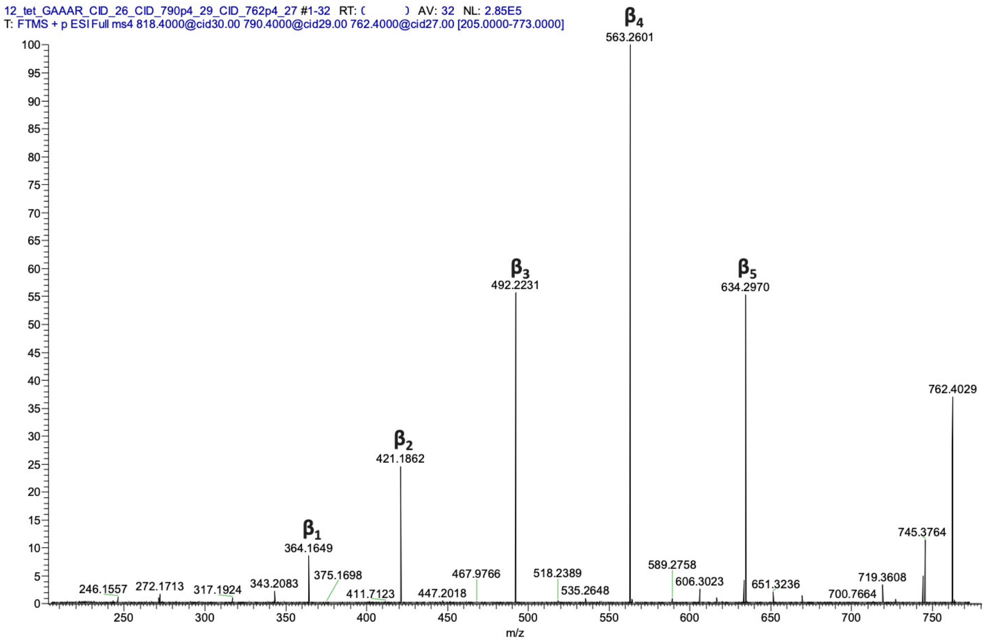


Figure 5.19: High-resolution CID-MS$_4$ of *m/z* 762 generated by CID-UVPD of (1,2-s-GAAAR + H)$^+$.

(1,4-s-GAAAR + H – N$_2$)$^+$ showed a smaller fraction of crosslinks, as expressed by the 67% yield (Table 5.28).

5.3.4 Ion Mobility, Collision Cross Sections, and Ion Structures of 1,2-*s*-GAAAR and 1,4-*s*-GAAAR

IMS of the (1,2-s-GAAAR + H)$^+$ ion was represented by two baseline-resolved peaks in a 36:64 intensity ratio for which we determined CCS$_{\text{IMS}}$ = 280 and 285 Å^2, respectively (Figure 5.21). Several (1,2-s-GAAAR + H)$^+$ ion structures were obtained by BOMD and DFT geometry optimizations. All low-energy structures converged to a common peptide folding pattern shown by **1,2-R1** and **1,2-R2**, in which the charged Arg side-chain group was wrapped by the peptide amides and H-bonded to the benzamide carbonyl (Figure 5.21). A similar H-bonding pattern was found for other low-energy conformers **1,2-R3**-**1,2-R6** (Figure 5.22). Interestingly, the cyclohexane ring in these low-energy conformers appeared in either the chair (**1,2-R1**, **1,2-R3**) or twist-boat (**1,2-R2**, **1,2-R4**-**1,2-R6**) conformation and the ring showed multiple reversible flipping in the course of BOMD trajectories at 610 K. On the basis of their Gibbs energies and

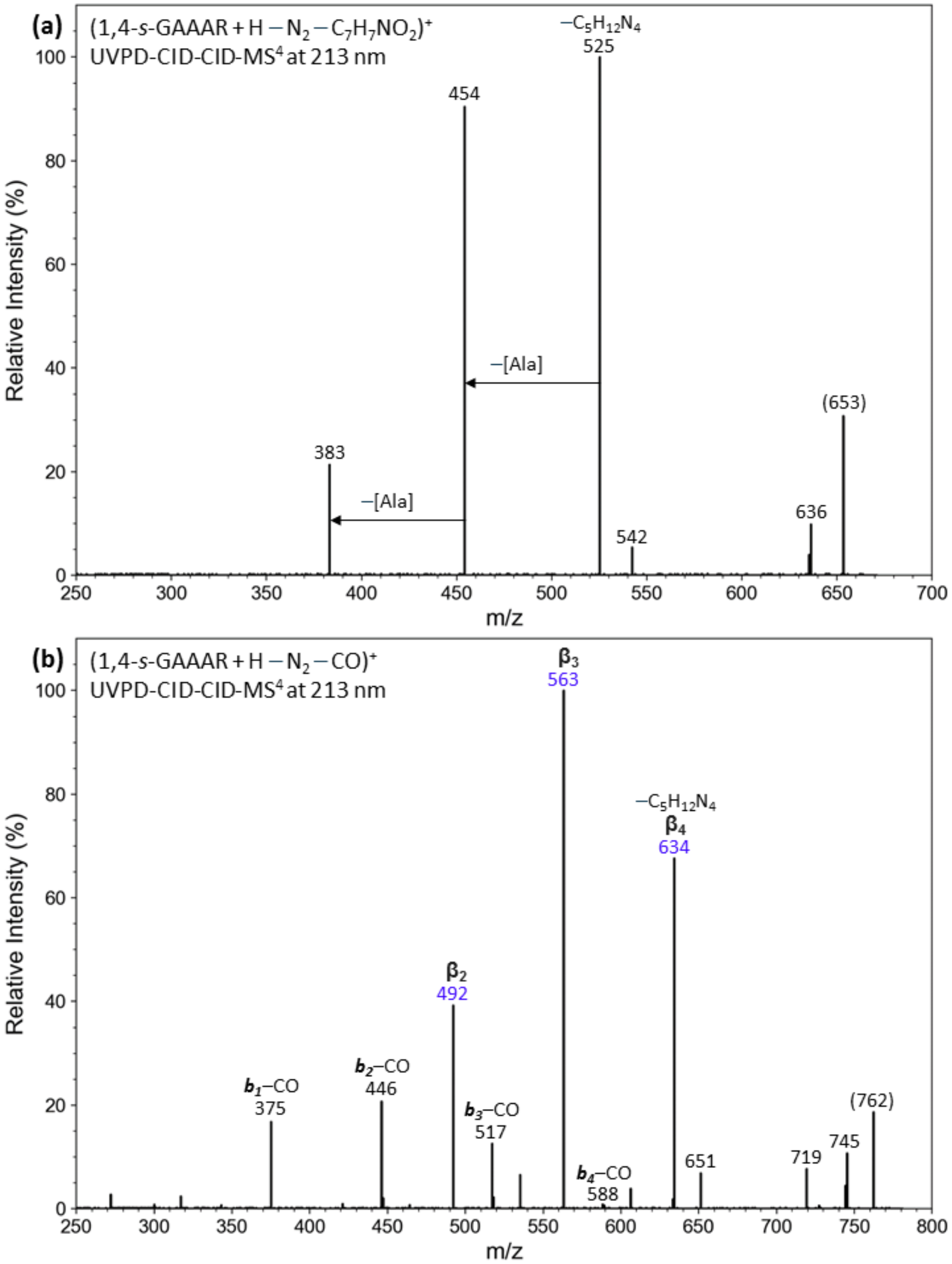


Figure 5.20: (a) CID-MS^4 of the intermediate *m/z* 653 ion by loss of $C_7H_7NO_2$ from the (1,4-*s*-GAAAR + H – N_2)$^+$ ion, and (b) CID-MS^4 of the intermediate *m/z* 762 ion by loss of CO from (1,4-*s*-GAAAR + H – N_2)$^+$.

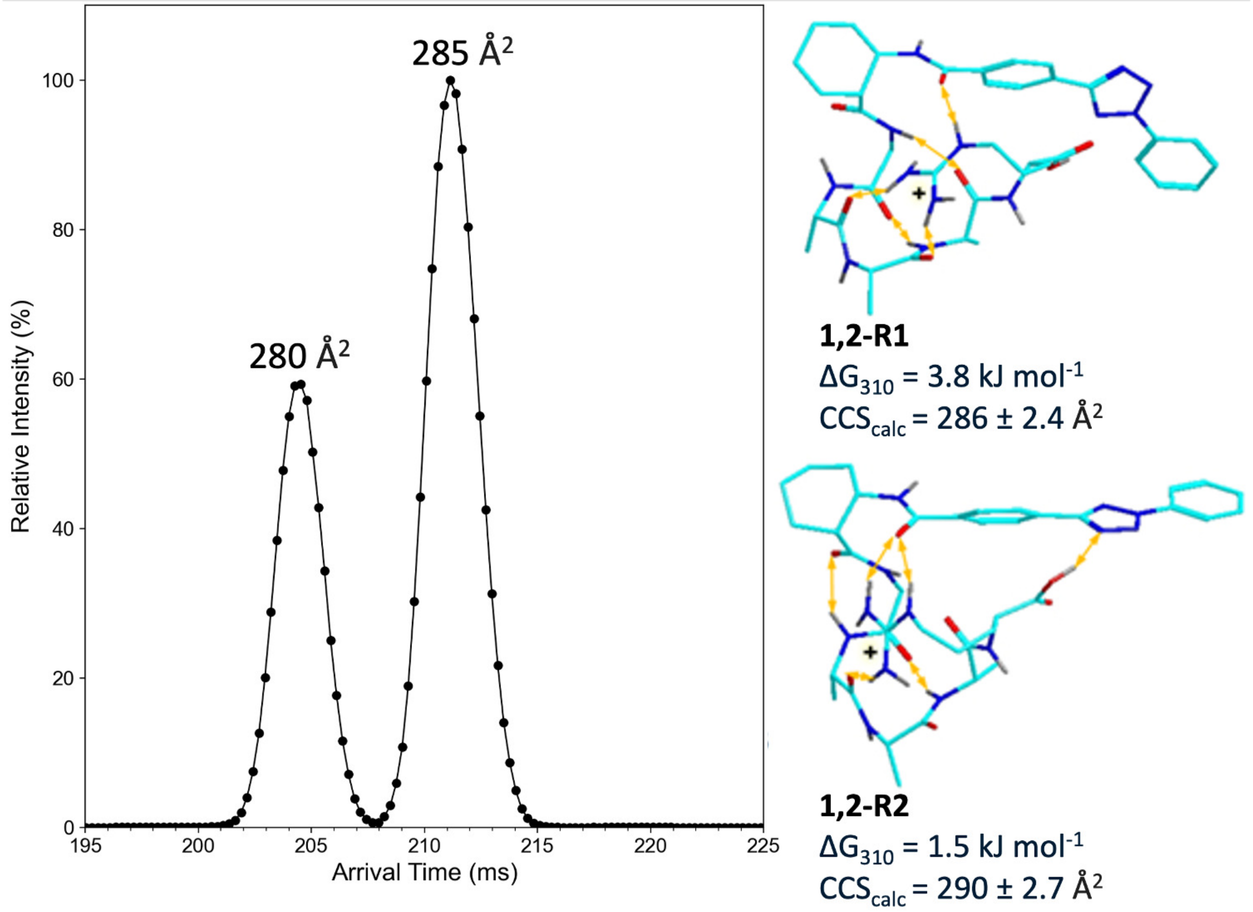


Figure 5.21: Arrival time distributions of (1,2-*s*-GAAAR + H)$^+$ after 7 passes. M06-2X/6-31+G(d,p) optimized structures, M06-2X/def2qzvpp relative Gibbs energies, and CCS$_{\text{calc}}$ of low energy (1,2-*s*-GAAAR + H)$^+$ ions. Structure discription as in Figure 5.9.

CCS$_{\text{calc}}$, structures **1,2-R1** and **1,2-R2**, CCS$_{\text{calc}}$ = 286 and 290 Å^2, respectively, gave the closest fit with the experimental data. The agreement was somewhat inferior to that for the other sequences, although still within 2%. The other low-energy conformers had CCS$_{\text{calc}}$ that were similar or slightly larger than those of **1,2-R1** and **1,2-R2** (Figure 5.22).

IMS of the (1,4-*s*-GAAAR + H)$^+$ ion gave a single peak at CCS$_{\text{IMS}}$ = 301 Å^2 even after 10 passes (Figure 5.23). Geometry optimization pointed to the lowest-energy structures **1,4-R1** and **1,4-R2** with CCS$_{\text{calc}}$ = 302, and 301 Å^2, respectively, providing near perfect match with the CCS$_{\text{IMS}}$. The nearly isoenergetic structures **1,4-R1** and **1,4-R2** were rotamers differing only in the orientation of the phenyltetrazole group whereas their peptide conformations were virtually identical (Figure 5.23). These were characterized by

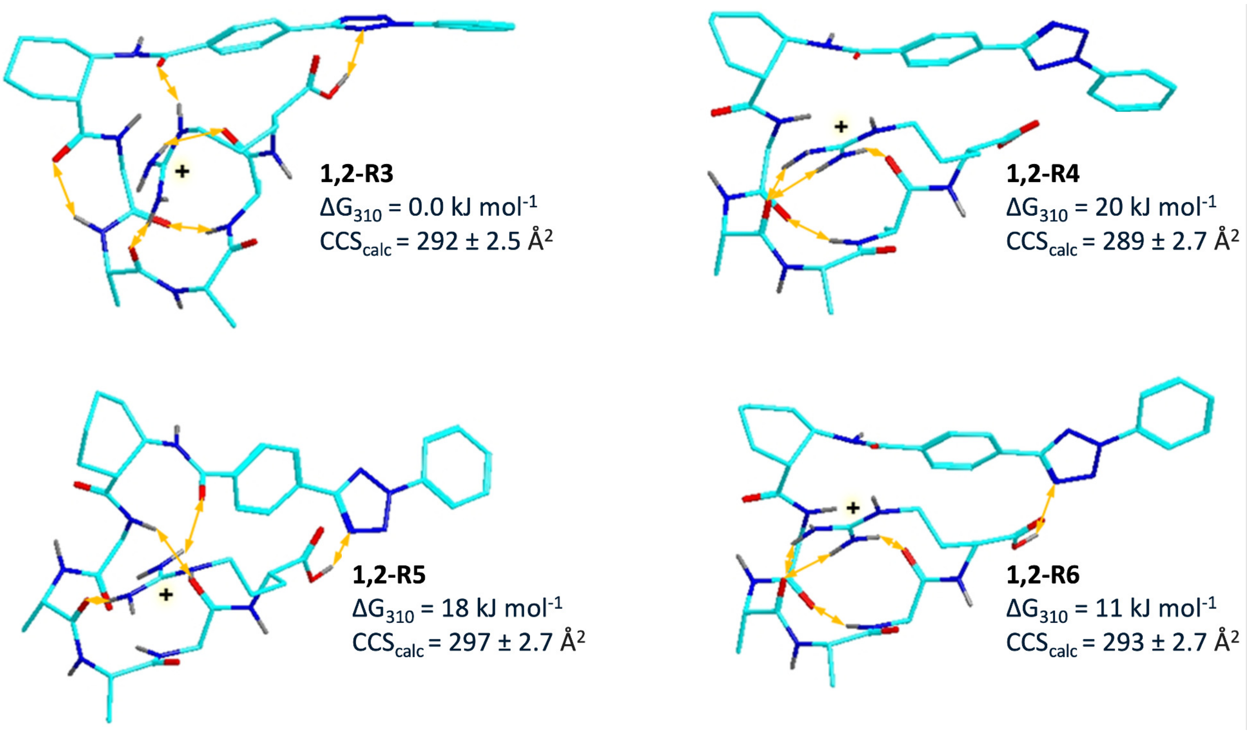


Figure 5.22: Auxiliary M06-2X/6-31+G(d,p) optimized structures of (1,2-*s*-GAAAR + H)$^+$ isomers and CCS$_{\text{calc}}$. The relative Gibbs energies refer to the text Figure 5.21.

a tightly folded peptide chain that was maintained by multiple hydrogen bonds and capped by carboxyl hydrogen bonding to the benzamide carbonyl. The cyclohexane ring was in a twist-chair conformation allowing the compact overall conformation for the peptide conjugates, as in the other low-energy ions with the trans-1,4-cyclohexane scaffold.

IMS of the CID-denitrogenated (1,2-*s*-GAAAR + H – N$_2$)$^+$ ions showed four resolved peaks after 15 passes (Figure 5.24). The major component had CCS$_{\text{IMS}}$ = 273 Å^2 that was accompanied by more compact minor components at CCS$_{\text{IMS}}$ = 269, 270, and 271 Å^2. These CCS$_{\text{IMS}}$ were substantially smaller than the CCS$_{\text{calc}}$ of nitrile imine conformers that were in the 295-296 Å^2 range (**1,2-r3**, **1,2-r4**,Figure 5.25). This result, indicating the absence of non-crosslinked isomers in the (1,2-*s*-GAAAR + H – N$_2$)$^+$ ion population, was qualitatively consistent with the high crosslinking yield for (1,2-*s*-GAAAR + H – N$_2$)$^+$ (Table 5.28). We obtained two types of low-energy crosslinked structures, a carboxyl crosslink **1,2-r1**, and an arginine crosslink **1,2-r2** of CCS$_{\text{calc}}$ = 275 and 277 Å^2, respectively, that were close to the CCS$_{\text{IMS}}$ of the main peak in the mobilogram. The minor components with smaller CCS$_{\text{IMS}}$ could belong to slightly more compact

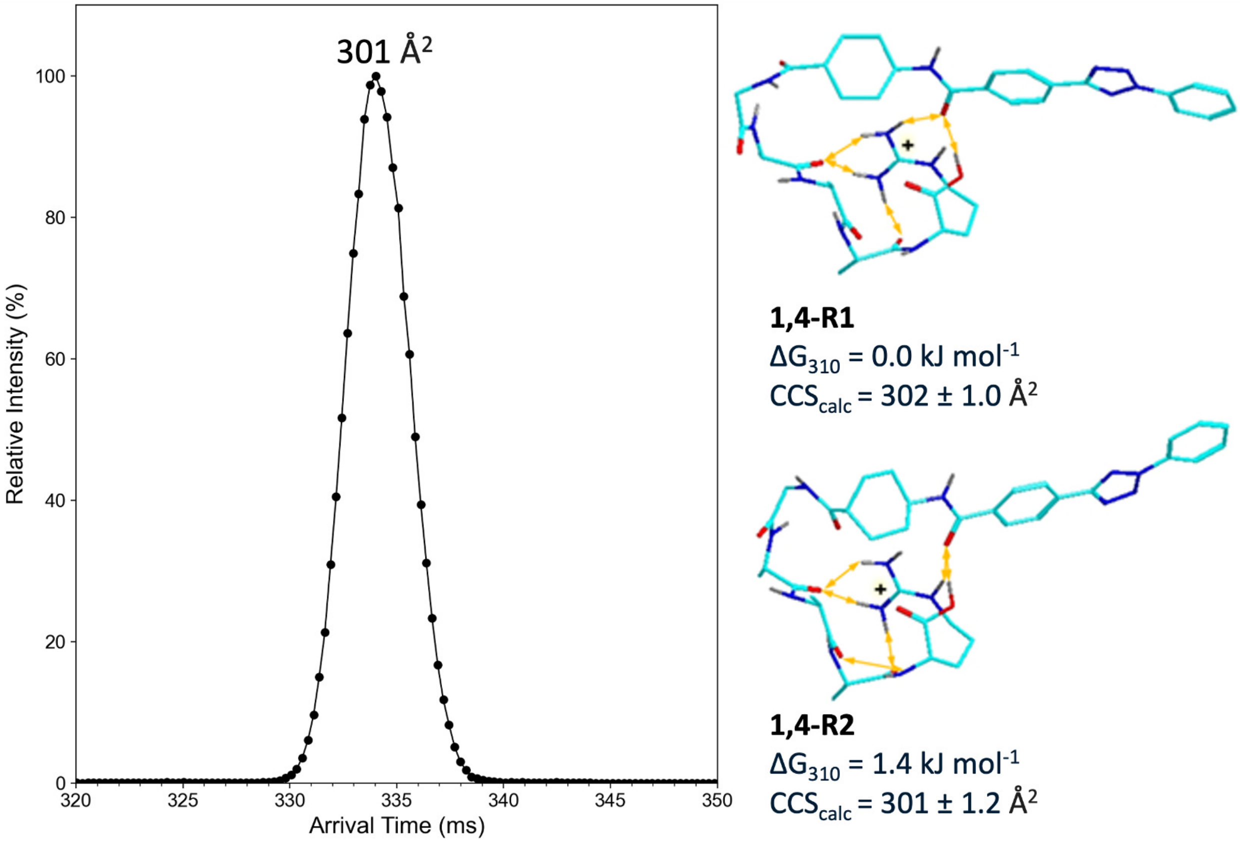


Figure 5.23: Arrival time distributions of (1,4-*s*-GAAAR + H)$^+$ after 10 passes. M06-2X/6-31+G(d,p) optimized structures, M06-2X/def2qzvpp relative Gibbs energies, and CCS$_{\text{calc}}$ of low energy (1,4-*s*-GAAAR + H)$^+$ ions. Structure discription as in Figure 5.9.

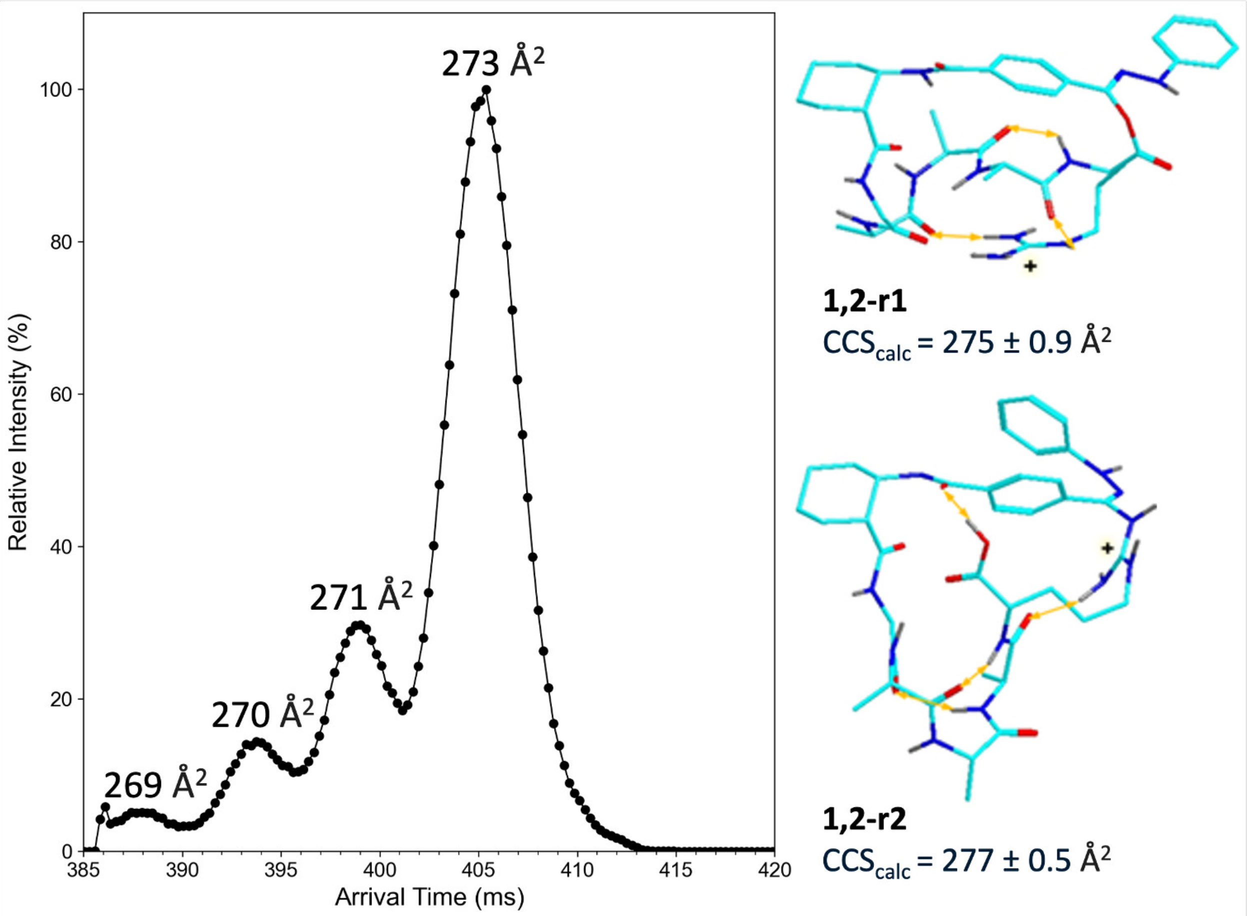


Figure 5.24: Arrival time distributions of (1,2-*s*-GAAAR + H – N_2)$^+$ after 15 passes. M06-2X/6-31+G(d,p) optimized structures and CCS_{calc} of low energy isomers **1,2-r1**-**1,2-r2**. Structure discription as in Figure 5.9.

conformers of **1,2-r1** and **1,2-r2** that, however, did not find matching structures among the other low-energy ions (**1,2-r5**-**1,2-r7**,Figure 5.25) that showed larger CCS_{calc}.

IMS of the CID-denitrogenated (1,4-*s*-GAAAR + H – N_2)$^+$ ions showed two major peaks at CCS_{IMS} = 272 and 291 Å^2 that were accompanied by very minor peaks at CCS_{IMS} = 280 and 285 Å^2 (Figure 5.26). We found the best fit for the major IMS peaks with the Arg-crosslinked structures **1,4-r1** and **1,4-r3** of CCS_{calc} = 274 and 291 Å^2, respectively. A carboxyl-crosslinked structure (**1,4-r2**, CCS_{calc} = 281 Å^2) could account for the minor peaks in the mobilogram. Ions **1,4-r1** and **1,4-r3** differed in their peptide conformations. In **1,4-r1**, the compact structure of the macrocyclic ring was clinched by a hydrogen bond of the carboxyl to the benzamide carbonyl. This feature was absent in **1,4-r2** that had a bulkier macrocyclic ring with fewer

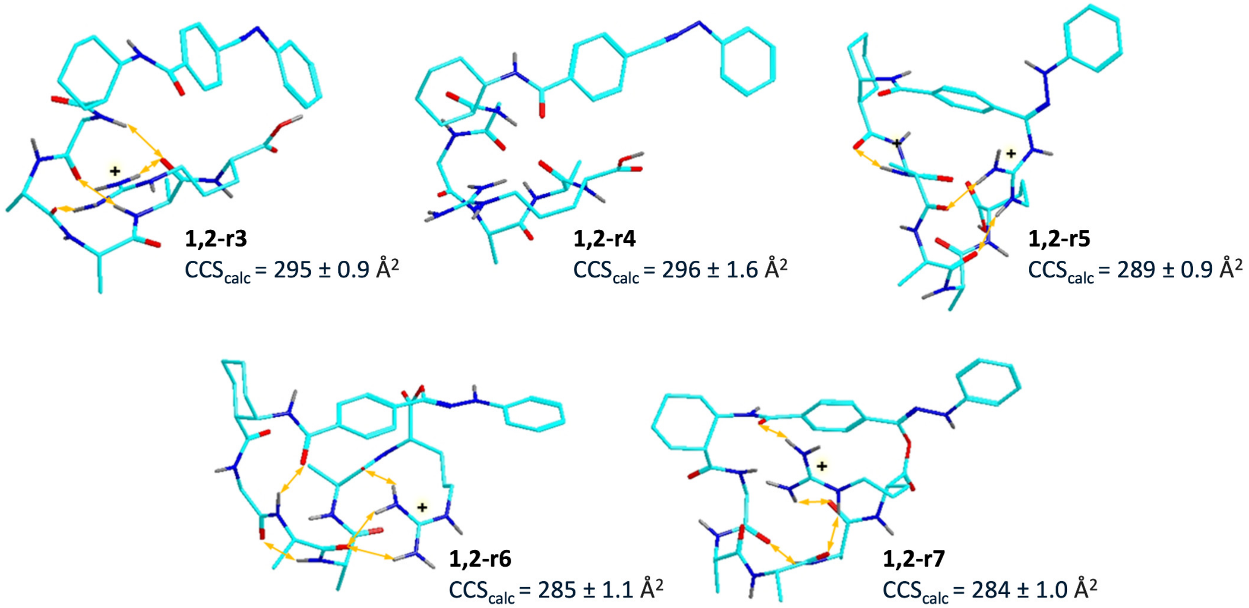


Figure 5.25: Auxiliary M06-2X/6-31+G(d,p) optimized structures of (1,2-*s*-GAAAR + H – N_2)$^+$ isomers and CCS_{calc}. The relative Gibbs energies refer to the main text Figure 5.24.

hydrogen bonds and a free carboxyl group. In addition to **1,4-r1**-**1,4-r3** we obtained several related ions that had CCS_{calc} within the experimental range. These were represented by Arg-crosslinked structures **1,4-r4** and **1,4-r5** of CCS_{calc} = 270 and 274 Å^2, respectively, carboxyl-crosslinked **1,4-r6** and **1,4-r7** with the respective CCS_{calc} = 278 and 281 Å^2, and nitrile imine **1,4-r8** of CCS_{calc} = 275 Å^2 (Figure 5.27). All the crosslinked structures had the cyclohexane ring in a twist-boat conformation while differing in the conformation of the macrocyclic ring. Nitrile imine **1,4-r8** had a standard cyclohexane chair conformation (Figure 5.27). The close similarity of the CCS_{calc} for the Arg-crosslinked ions **1,4-r1**, **1,4-r4** and **1,4-r5** did not allow us to unambiguously assign the 272 Å^2 peak in the mobilogram to a single conformer. It appears, however, that the carboxyl-crosslinked ions **1,4-r2**, **1,4-r6**, and **1,4-r7** were only a small minority among the crosslinks.

5.4 Conclusions

In both *cis*- and *trans*-GAAAK and GAAAR scaffolds, protonation occurs at the basic C-terminal residues. Photodissociation and CID produced mixtures of isomers and conformers, comprising both crosslinked and linear forms. For (1,2-*s*-GAAAK + H)$^+$, photodissociation primarily resulted in crosslinking via the carboxyl group to form diarylhydrazide enol esters, although some crosslinking occurred at the [Lys] residue, as

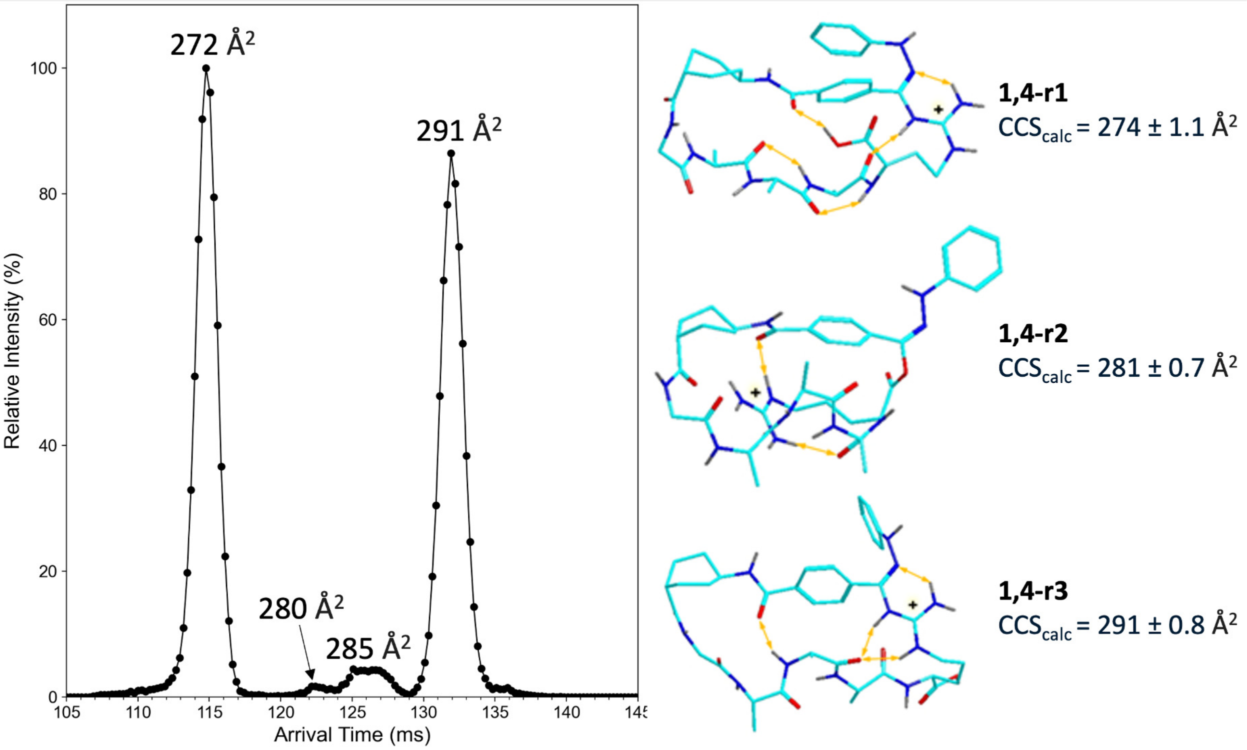
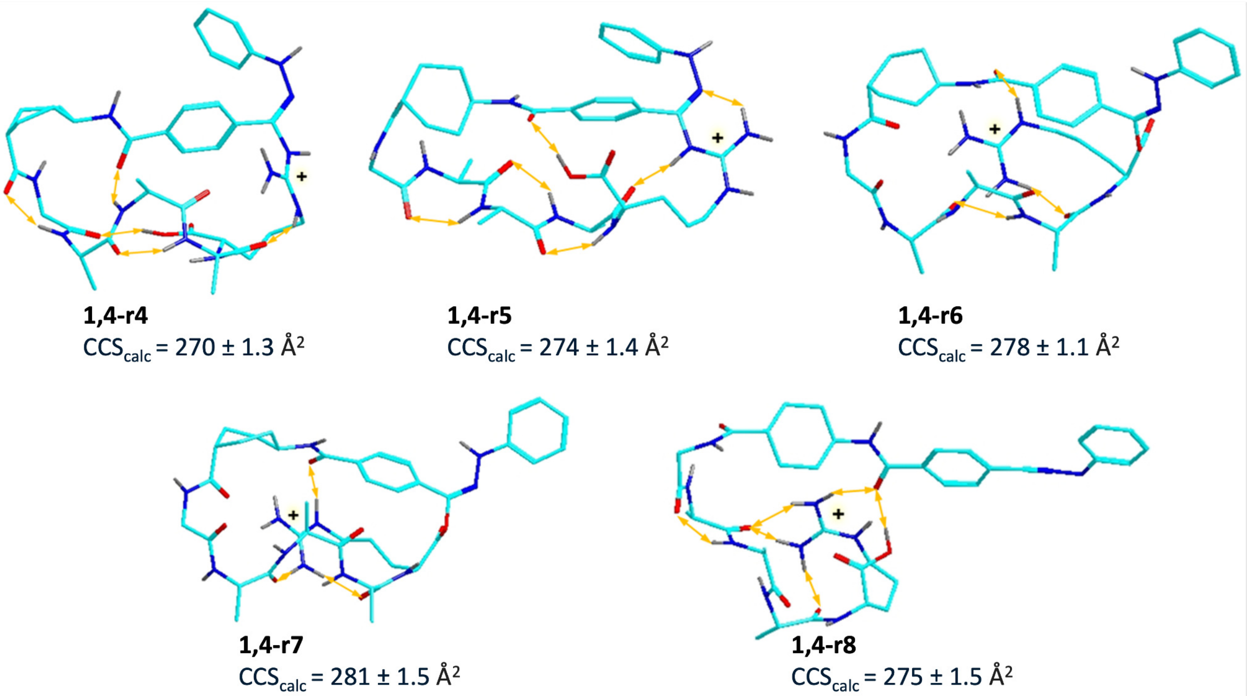


Figure 5.26: Arrival time distributions of (1,4-*s*-GAAAR + H – N_2)$^+$ after 4 passes. M06-2X/6-31+G(d,p) optimized structures and CCS_{calc} of low energy isomers **1,4-r1**-**1,4-r3**. Structure discription as in Figure 5.9.



1,4-r4
CCS_{calc} = 270 ± 1.3 Å^2

1,4-r5
CCS_{calc} = 274 ± 1.4 Å^2

1,4-r6
CCS_{calc} = 278 ± 1.1 Å^2

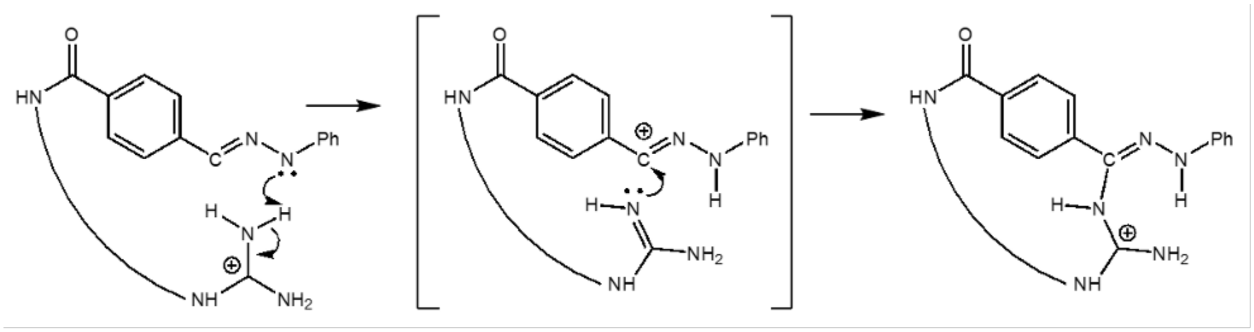
1,4-r7
CCS_{calc} = 281 ± 1.5 Å^2

1,4-r8
CCS_{calc} = 275 ± 1.5 Å^2

Figure 5.27: Auxiliary M06-2X/6-31+G(d,p) optimized structures of (1,4-*s*-GAAAR + H – N_2)$^+$ isomers and CCS_{calc}. The relative Gibbs energies refer to the main text Figure 5.26.

demonstrated by carboxyl methylation blocking experiments. In (1,2-*s*-GAAAR + H)$^+$ and (1,4-*s*-GAAAR + H)$^+$, crosslinking occurred through the [Arg] guanidinium group, highlighting proton transfer steps, while carboxyl crosslinking was minimal, suggesting the carboxyl proton was not involved.

Crosslinking by the guanidinium group can be viewed as a two step process: an endothermic proton transfer from guanidinium to the imine nitrogen, followed by an exothermic C–N bond formation. Calculations at the M06-2X/def2qzvpp level of theory including zero-point corrections indicate Gibbs energy changes of $\Delta G_{310,rxn}$ = -88 kJ mol^{-1} and -72 kJ mol^{-1} for cyclization of the (1,2-*s*-GAAAR + H – N_2)$^+$ and (1,4-*s*-GAAAR + H – N_2)$^+$ intermediates, respectively (Scheme 5.3). This cyclization can be viewed as a [3 + 2] addition of the guanidinium N–H bond across the nitrile imine dipolarophile, proceeding through a proton-transfer intermediate. Similar reactions have not been observed in condensed phases, marking a novel aspect of gas-phase peptide ion chemistry.



Scheme 5.3: Proposed mechanism for arginine crosslinking.

5.5 Bibliography

[1] Harms, M. J.; Schlessman, J. L.; Chimenti, M. S.; Sue, G. R.; Damjanović, A.; García-Moreno, B. "A Buried Lysine that Titrates with a Normal pK(a): Role of Conformational Flexibility at the Protein–Water Interface as a Determinant of pK(a) Values." *Protein Sci.* **2008**, *17*, 833-845.

[2] Fitch C. A.; Platzer G.; Okon M.; García-Moreno, B. "Arginine: Its pKa Value Revisited." *Protein Sci.* **2015**, *24*, 752-761.

[3] Harms, M. J.; Schlessman, J. L.; Sue, G. R.; Garcia-Moreno, B. "Arginine Eesidues at Internal Positions in a Protein are Always Charged." *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 18954-18959.

[4] Li, L. B.; Vorobyov, I.; Allen, T. W. "The Different Interactions of Lysine and Arginine Side Chains with Lipid Membranes." *J. Phys. Chem. B* **2013**, *117*, 11906-11920.

[5] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. "Molecular Dynamics with Coupling to an External Bath." *J. Chem. Phys.* **1984**, *81*, 3684-3690.

[6] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. "Semiempirical Quantum Chemical PM6Method Augmented by Dispersion and H Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes." *J. Chem. Theory Comput.* **2009**, *5*, 1749-1760.

[7] Řezáč, J. "Cuby: An Integrative Framework for Computational Chemistry." *J. Comput. Chem.* **2016**, *37*, 1230-1237.

[8] Stewart, J. J. P. *MOPAC 16*; Stewart Computational Chemistry: Colorado Springs, CO, **2016**.

[9] Becke, A. D. "Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior." *Phys. Rev. A* **1988**, *38*, 3098-3100.

[10] Grimme, S.; Ehrlich, S.; Goerigk, L. "Effect of the Damping Function in Dispersion Corrected Density Functional Theory." *J. Comput. Chem.* **2011**, *32*, 1456-1465.

[11] Nickerson, C. J.; Bryenton, K. R.; Price, A. J. A.; Johnson, E. R. "Comparison of Density-Functional Theory Dispersion Corrections for the DES15K Database." *J. Phys. Chem. A* **2023**, *127*, 8712-8722.

[12] Zhao, Y.; Truhlar, D. G. "The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals." *Theor. Chem. Acc.* **2008**, *120*, 215-241.

[13] Weigend, F. "Accurate Coulomb-Fitting Basis Sets for H to Rn." *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057-1065.

[14] Gray, M.; Bowling, P. E.; Herbert, J. M. "Comment on "Benchmarking Basis Sets for Density Functional Theory Thermochemistry Calculations: Why Unpolarized Basis Sets and the Polarized 6-311G Family Should Be Avoided"." *J. Phys. Chem. A* **2024**, *128*, 7739-7745.

[15] Singh, U. C.; Kollman, P. A. "An Approach to Computing Electrostatic Charges for Molecules." *J. Comput. Chem.* **1984**, *5*, 129-145.

[16] Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. "Atomic Charges Derived from Semiempirical Methods." *J. Comput. Chem.* **1990**, *11*, 431-439.

[17] Ieritano, C.; Crouse, J.; Campbell, J. L.; Hopkins, W. S. "A Parallelized Molecular Collision Cross Section Package with Optimized Accuracy and Efficiency." *Analyst* **2019**, *144*, 1660-1670.

[18] Ieritano, C.; Hopkins, W. S. "Assessing Collision Cross Section Calculations Usi ng MobCal-MPI with a Variety of Commonly Used Computational Methods." *Mater. Today Commun.* **2021**, *27*, No. 102226.

[19] Halgren, T. A. “Merck Olecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMFF94.” *J. Comput. Chem.* **1996**, *17*, 490-519.

[20] Wan, J; Nytka, M.; Vu, K.; Qian, H.; Lemr, K.; Tureček, F. “Nitrile Imines as Peptide and Oligonucleotide Photocrosslinkers in Gas-Phase Ions.” *J. Am. Soc. Mass Spectrom.* **2024**, *35*, 344-356.

[21] Goeringer, D. E.; McLuckey, S. A. “Evolution of Ion Internal Energy during Collisional Excitation in the Paul Ion Trap: A Stochastic Approach. ” *J. Chem. Phys.* **1996**, *104*, 2214-2221.

[22] Green, M. M. “Mass Spectrometrry – a Sensitive Probe of Molecular Structure.” *Pure Appl. Chem.* **1978**, *50*, 185-196.

[23] Mandelbaum, A. “Stereochemical Effects in Mass Spectrometry.” *Mass Spectrom. Rev.* **1983**, *2*, 223-284.

[24] Tureček, F. “Stereochemistry of Organic Ions in the Gas Phase: A Review.” *Coll. Czech. Chem. Commun.* **1987**, *52*, 1928-1984.

[25] Splitter, J. S.; Tureček, F. *Applications of Mass Spectrometry to Organic Stereochemstry*; VCH Publishers, New York: **1994**, pp. 123-128.

[26] Marek, A.; Nguyen, H. T. H.; Brož, B.; Tureček, F. “Stereospecific Control of Peptide Gas-Phase Ion Chemistry with cis and trans Cyclo Ornithine Residues.” *J. Mass Spectrom.* **2018**, *53*, 124-137.

[27] Squillacote, M.; Sheridan, R. S.; Chapman, O. L.; Anet, F. A. L. “Spectroscopic Detection of the Twist-Boat Conformation of Cyclohexane. Direct Measurement of the Free Energy Difference between the Chair and the Twist-Boat.” *J. Am. Chem. Soc.* **1975**, *97*, 3244-3246.

[28] Johnson, W. S.; Bauer, V. J.; Margrave, J. L.; Frisch, M. A.; Dreger, L. H.; Hubbard, W. N. “The Energy Difference between the Chair and Boat Forms of Cyclohexane. The Twist Conformation of Cyclohexane.” *J. Am. Chem. Soc.* **1961**, *83*, 606-614.

Table 5.27: CID-MS^4 Fragments of (1,4-*s*-GAAAK + H – N_2)$^+$ *m/z* 762 from 1,4-*s*-GAAAK Following UVPD at 213 nm

744 $-H_2O$	634 β_5	563 β_4	492 β_3
726	606	535	464
661	563	518	447
616	492	492	421
563			
545			
525			
492			
454			
421			
383			

Table 5.29: CID-MS^4 Fragments of $(MH - N_2)^+$ *m/z* 790 from 1,2-*s*-GAAAR Following UVPD

762 –CO	653 $-C_7H_7NO_2$	563 β_4	492 β_3	421 β_2	364 β_1
745	636	492	421	364	256
634	525	421	364	256	239
563	454	364	135	166	135
492	383			135	120
421	312			120	109
364					

Table 5.30: CID-MS^4 Fragments of $(MH - N_2)^+$ *m/z* 790 from 1,4-*s*-GAAAR Following UVPD

772 $-H_2O$	762 –CO	730 $-NH_3$–CHNO	653 $-C_7H_7NO_2$	563 β_4	545 $\mathbf{b}_4$	492 β_3	474 $\mathbf{b}_3$	447 β_4-NH_3-CO	421 β_2	403 $\mathbf{b}_2$
755	745	712	636	535	517	464	446	419	403	385
745	719	702	542	518	474	447	403	390	393	375
730	634	616	525	492		421	135	362	376	357
701	563	545	454					313	364	346
684	517	474	383					265	336	318
630	492	454						245	313	221
615	446	383						135	166	166
544	375							120	135	135
477									120	120