Structural Characterization of Gas-Phase Biomolecular Ions with Novel Mass Spectrometry and Computational Methods

Jiahao Wan

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
Matthew F. Bush
Robert E. Synovec

Program Authorized to Offer Degree:
Department of Chemistry

University of Washington

Abstract

Structural Characterization of Gas-Phase Biomolecular Ions with Novel Mass Spectrometry and Computational Methods

Jiahao Wan

Chair of the Supervisory Committee:
František Tureček
Department of Chemistry

This dissertation presents the structural characterization of gas-phase biomolecular ions through novel mass spectrometry-based techniques and computational studies. Ions were generated via electrospray ionization and analyzed using tandem mass spectrometry (MS^n) within an ion trap mass analyzer. Fragmentation methods used were collision-induced dissociation (CID) and ultraviolet photodissociation (UVPD) at various wavelengths. Cyclic ion mobility experiments provided experimental collision cross sections (CCS), revealing three-dimensional shapes of ions. UV-Vis action spectroscopy was used to study the electronic structures of ions. Born-Oppenheimer molecular dynamics (BOMD) and density functional theory (DFT) were used to obtain low-Gibbs-energy structures with theoretical CCS and vibronic absorption spectra. Background for these experimental and theoretical methods is discussed in Chapter 1. Chapter 2 and Chapter 3 focus on the resolution of identity in dissociations in DNA codon cations and cations radicals. Mono- and diprotonated cations were generated by electrospray ionization. Cation radicals were generated by electron transfer (ET). Dissociations in CID-MS included nucleobase loss and backbone cleavage.

Nucleobase loss was found to be highly dependent on both the type and position of the nucleobase within the DNA codon. The $\boldsymbol{w_2}^+$ ion formation was universal in dissociations of cations while the $\boldsymbol{d_2}^+$ ions were dominant in that of cation radicals. In both cases ^{15}N-labeling in nucleobases was used to resolve dissociations in symmetric sequences dXXX and dXYX. Plausible dissociation pathways were proposed based on low-Gibbs-energy structures of isomers and conformers. Chapter 4 investigates the secondary structure transitions of DNA from solution to the gas phase. Cyclic ion mobility was employed to study the dGCGAAAGC heptanucleotide hairpin and its sequence-scrambled isomers. The 2+, 3+, and 4+ ions exhibited mixtures of protomers and conformers, as evidenced by multiple peaks in ion mobility arrival time distributions. Low-energy structures revealed zwitterionic features and disruption of G···C Watson-Crick pairs in the gas-phase ions. The experimental and theoretical CCS for low-energy structures showed reasonable agreement. Chapter 5 describes the development of nitrile imine as a novel photo-cross-linker. Thermal and photodissociation of tetrazole-peptide conjugates generated $-N_2$ ions that crosslinked intrinsic functional groups of peptides, producing unique fragments in CID-MS^3. UV-Vis action spectroscopy and cyclic ion mobility, supported by comprehensive computational studies, was used to characterize the $-N_2$ ions. Additionally, nitrile imines selectively crosslinked with guanine in peptide-dinucleotide complexes, highlighting their potential in detecting non-covalent interactions. Chapter 6 explores the interactions between the tetrazole photo-tag and the N-terminal Asp, Glu, Asn, and Gln residues. Nitrile imines underwent efficient crosslinking to the side-chain amides and carboxyl groups, as indicated by the characteristic $\boldsymbol{v_n}$ type fragment ions and the matching between experimental and theoretical CCS. Other signature fragments were loss of phenylhydrazine and internal alanines. Chapter 7 further explores the effect of basic N-terminal residues Arg, Lys, and His on crosslinking. The proposed mechanism involved proton transfer-assisted nucleophilic attack and oxygen transfer. Chapter 8 summarizes all the work conducted and proposes future directions for exploration.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF SCHEMES

Pages

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Prof. František Tureček for his guidance over the past 4 years. It's an honor to have you as an advisor. You didn't just teach me to be a qualified researcher but also showed me how to be a good person.

I would like to thank all the Tureček group members for your help and support. Dr. Yue Liu and Dr. Shu R. Huang, you served as senior graduate mentors to me and taught me experimental and computational skills that I have been using later all the way through my graduation. Dr. Hongyi Zhu, you have always been a good coworker and a wonderful friend. Haocheng Qian and Mars Wei, you helped me a lot with my projects and I wish you great success in your PhD programs.

I would like to thank the stuff from UW Chemistry Mass Spectrometry Facility, Martin and Brandon, and the stuff from UW Medical Chemistry Mass Spectrometry Center, Dale and James. Thank you for your support and guidance using the instruments.

I would like to thank our collaborators. Dr. Marianna Nytka and Prof. Karel Lemr, it was an honor to work with you. I really enjoyed the trip to Olomouc and look forward visiting it again in the future. Dr. Aleš Marek and Dr. Břetislav Brož, thank you for providing radiolabeled compounds for my projects. Mikuláš and Dr. Václav Zima, it was nice to have you as coworkers and friends. Thank you for your contributions to the group projects!

I would to thank my committee members, Prof. Matthew F. Bush, Prof. Robert E. Synovec and my GSR, Prof. James Bruce. Thank you for serving as my committee members and providing useful suggestions.

I would like to thank my friends. Meng Yu and Qiang Xiao, we didn't get the chance to see each other quite often but always connected online via our small group chat. Guang Yang, it's nice to share the 5-year life experience with you across the Pacific Ocean. My friends in Seattle, Dr. Yusi Chen, Ying, Zhiyao, Ziwen, Zhiyang, I feel so lucky to have you around enjoying the blessed life in PNW. Dr. Yizhou Wang, Dr. Jiayi Li, Dr. Haotian Zhang, Xiangtian and Dr. Yande Peng, I really enjoyed myself every time that I visited bay area to see you guys. They are all unforgettable memories that I will recall again and again in the future. Yushu and Roxanne, thank you for your hosting while I visited NYC. I had so much fun! Haorui and Dr. Yijun Chen, even though you are thousands miles away in Australia, I always look forward seeing you guys again.

I would like to give the special thanks to my partner, Jasmine. I cannot imagine a life without you in this city and this foreign country. Everything happened so fast since we met and sometimes I would be surprised that we have already explored so many places together! I know there are still plenty for us to explore ahead.

At last, I want to thank my parents. Thank you for your love and support. I have always known that, no matter what choice I make, you will always be there for me.

Chapter 1

Introduction

1.1 Structural Analysis of Gas-Phase Biomolecular Ions

In nature, the unique three-dimensional structures of biomolecules are fundamental to their stability and function. For example, the double helix structure of DNA is essential for its replication and transcription processes,[1,2] while the tertiary and quaternary structures of proteins determine their receptor-binding functions.[3,4] Alterations or disruptions in these biomolecular structures within cells can lead to severe diseases, such as cancer[5] and Parkinson's disease.[6] Therefore, accurate structural elucidation of biomolecules provides valuable insights that benefit structural biology, chemistry, and pharmaceutical sciences. Various techniques, such as X-ray crystallography[7] and NMR spectroscopy[8], provide high-resolution, three-dimensional insights of target molecules at the atomic level. However, these methods require meticulous sample purification and preparation, which can be both time-consuming and labor-intensive. Additionally, techniques such as cryogenic electron microscopy, while powerful, are extremely costly and demand expertise in data analysis using advanced software and algorithms, typically accessible only to experienced structural biologists.[9]

On the other hand, mass spectrometry (MS), along with other low-resolution methods, offers an alternative for analyzing the structure of biomolecules. Thanks to soft ionization methods such as electrospray ionization (ESI),[10] many biomolecular ions could be generated and transferred into the gas phase without major dissociation or fragmentation. Mass spectrometry is becoming increasingly popular for characterizing biomolecular structures due to several advantages. Firstly, it does not require highly purified or large quantities of samples, and ions can be easily isolated and stored in the gas phase based on their mass-to-charge ratio (m/z), with well-defined stoichiometry. Secondly, novel gas-phase chemistry techniques such as photodissociation,[11] ion/ion reactions,[12] and covalent crosslinking[13] can be applied to these biomolecular ions to study various aspects of their structure and reactivity. Finally, the highly rarefied gas phase inside mass spectrometers, which is free of solvents, makes it easier to characterize reactive species with short lifetimes in the condensed phase. Computational chemistry methods can be employed to

model the ions of interest, allowing the investigation of properties such as electronic structures and unimolecular reaction rates.

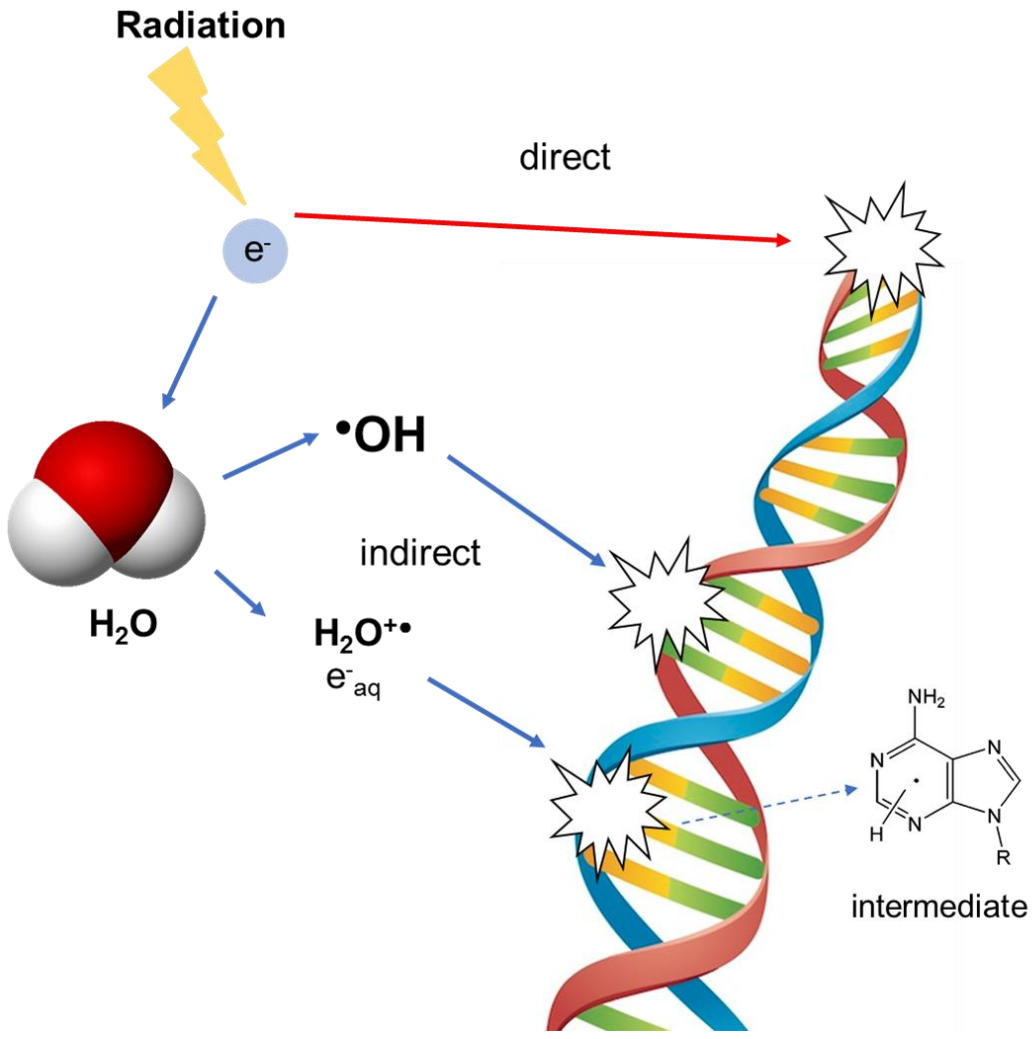
While there is an ongoing debate about whether gas-phase biomolecular ions retain the same structures as they do in the condensed phase, many studies have shown that large molecules, such as protein (complexes), can maintain their key structural features intact for some time in the gas phase.[14] Moreover, studying the gas-phase structure of biomolecules is valuable in its own right. It helps improve our understanding of the electrospray ionization process and links gas-phase conformational information obtained from techniques like collision-induced unfolding (CIU)[15] and hydrogen-deuterium exchange mass spectrometry (HDX-MS)[16] to structures derived from solution or computational modeling. The proposed reaction mechanisms, including dissociations and rearrangements based on elucidated structures, provide guidance for selecting appropriate fragmentation methods to achieve higher sequence coverage in proteomics.

In this dissertation, I will explore the use of novel mass spectrometry techniques, in combination with methods such as UV-Vis action spectroscopy, cyclic ion mobility mass spectrometry (c-IMS), and comprehensive computational modeling, to complete the structural analysis of two biomolecular systems. The first category focuses on DNA cations and cation radicals. Multi-protonated DNA cations exhibit complex intramolecular hydrogen-bonding networks that define their secondary structures, while cation radicals are presumed intermediates in radiative DNA damage processes. The second category examines tetrazole-peptide conjugates, which are employed in inter- and intramolecular photochemical crosslinking reactions. Comprehensive characterization of the reactive intermediates and crosslinked products will provide insights into the fundamental mechanisms of this novel strategy and demonstrate its potential applications in detecting non-covalent interactions. Detailed background information on each system is provided in **Sections 1.1.1** and **1.1.2**.

1.1.1 DNA Cations and Cation Radicals

Deoxyribonucleic acid (DNA) is the molecule that encodes all the essential genetic information in human cells. Any damage to the DNA that disrupts its structure can lead to mutations, malfunctioning, and potentially contribute to carcinogenesis if not properly repaired.[17,18] One of the most common causes of DNA damage is radiative ionization, which can occur through two

primary mechanisms: direct and indirect (**Scheme 1.1**).[19] The direct mechanism involves the interaction of DNA moieties with radiation (e.g., α-, γ-, X-rays and UV light), typically resulting in the ejection of an electron to form a DNA cation radical ($DNA^{+\bullet}$),[19] known as the oxidative pathway. Alternatively, direct DNA ionization can also occur via the capture of a secondary low-energy electron (LEE),[20,21] forming transient negative ions (TNIs) as part of the reductive pathway. These TNIs can decay into electronically excited states, which are often readily dissociative.[22,23]



Scheme 1.1: Direct (red arrow) and indirect (blue arrow) radiation damage to DNA.

Instead of directly ionizing DNA molecules, radiation can also interact with various components in the cellular environment (e.g., water, salts, and proteins), generating reactive

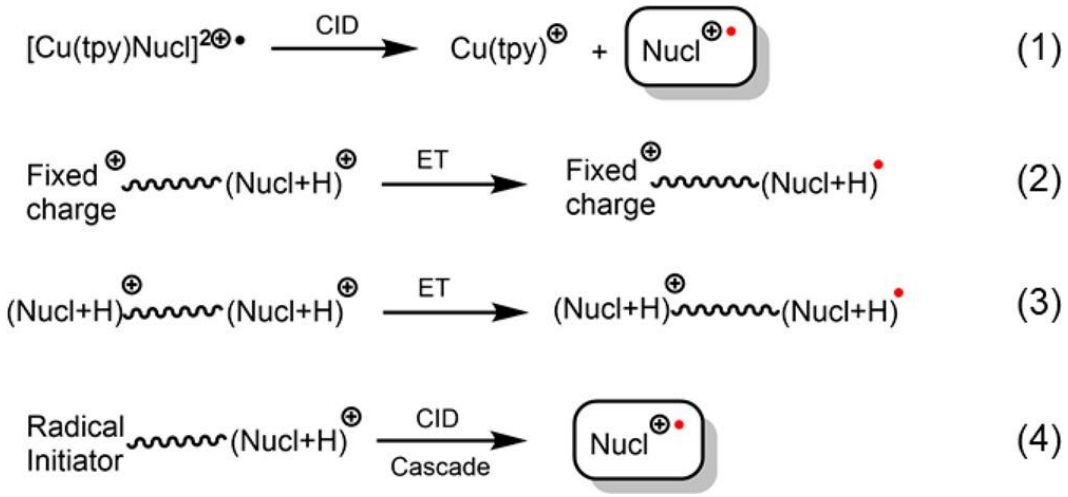
secondary species through what is known as indirect mechanisms.[24] This indirect pathway primarily involves water radiolysis, leading to the formation of highly reactive species like hydroxyl radicals and solvated electrons.[25]

Despite extensive research on the DNA ionization process, the structural characterization of key intermediates, such as DNA cation radicals, remains challenging due to their high reactivity and short lifetimes. Pulse radiolysis studies of DNA in the condensed phase (e.g., in solution or frozen glasses) often yield mixtures of protonated neutral radicals and hydrogen atom adducts.[26-28] The types of products are highly dependent on environmental factors, such as pH and the presence of metal ions, as revealed by techniques like electron paramagnetic resonance (EPR) spectroscopy and fast UV–Vis spectroscopy.[29,30] These factors significantly complicate the characterization of DNA cation radicals in the condensed phase.

Another type of DNA structural damage is depurination under acidic conditions. Although this process is nonenzymatic and less well-known than radiative ionization damage, it is thought to play a role in low-pH environments, such as in lysosomes.[31] Depurination, or the loss of nucleobases, is often followed by glycosidic bond cleavage and single-strand breaks.[32] Therefore, understanding the mechanisms of dissociation based on the structures of protonated DNA under acidic conditions is crucial. However, preparing and characterizing protonated DNA cations, especially those with multiple charge states, is challenging in solution due to the low basicity of pyrimidine nucleobases.

The generation of DNA cations in the gas phase takes advantage of the increasing basicity of nucleobases.[33] The direct electrospray of nucleobases and nucleosides in acidic aqueous solutions (water/methanol/acetic acid = 50: 50 :1) leads to efficient protonation. For nucleotides and nucleosides with additional basic-group modifications, this approach produces multiple-charge cations with a good yield.[34,35] While many studies have focused on negative ions due to the acidic nature of the phosphate diester,[36] investigating positive ions offers two significant advantages. Firstly, multi-protonation of nucleotides, particularly longer sequences, leads to highly internally solvated structures, which can disrupt their native hydrogen-bonding networks, such as Watson-Crick base pairs. I will discuss these details further in the context of the DNA miniloop, d(GCGAAGC), in **Chapter 4**. Secondly, electron transfer (ET) reduction of multi-charged cations provides a feasible approach to generating cation radicals within an ion trap.

This ET reduction, commonly applied to oligonucleotides and fixed-charge-tagged nucleosides, represents a reductive pathway for producing DNA cation radicals (**Scheme 1.2 equation(2), (3)**). Alternatively, the oxidative pathway utilizes intramolecular electron transfer in complexes containing Cu(II) and a co-ligand like terpyridine, where cation radicals are generated via collision-induced dissociation (CID) (**Scheme 1.2 equation(1)**).[37] However, this method has so far been limited to generating radicals in 2′-deoxyguanosine and guanosine.[38,39] Recently, we have successfully created radicals on the ribose moiety using radical-carrying groups in a nucleoside conjugate (**Scheme 1.2 equation(4)**).[40,41] These strategies collectively provide a versatile and robust toolbox for studying the structures of DNA cation radicals in the gas phase.



Scheme 1.2: Formation of DNA cation radicals in the gas phase.

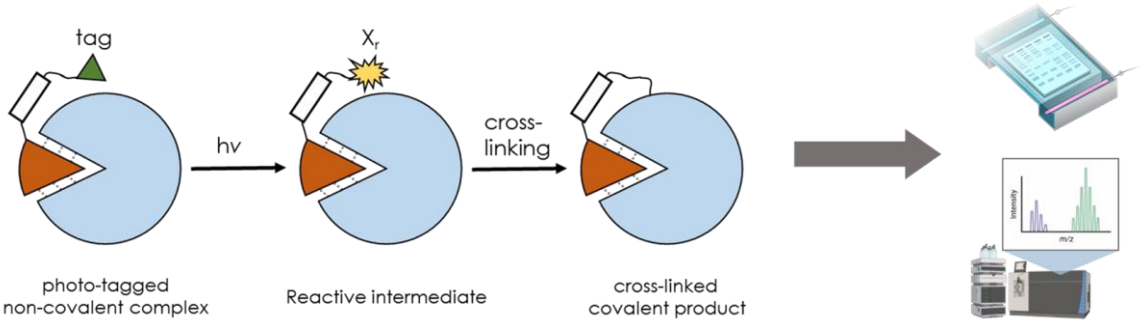
In this dissertation, I will discuss the gas-phase structural characterization of a biologically important system: DNA codons, with a focus on mono- and dications in **Chapter 2** and cation radicals in **Chapter 3**.

1.1.2 Tetrazole-Peptide Conjugates for Photochemical Crosslinking

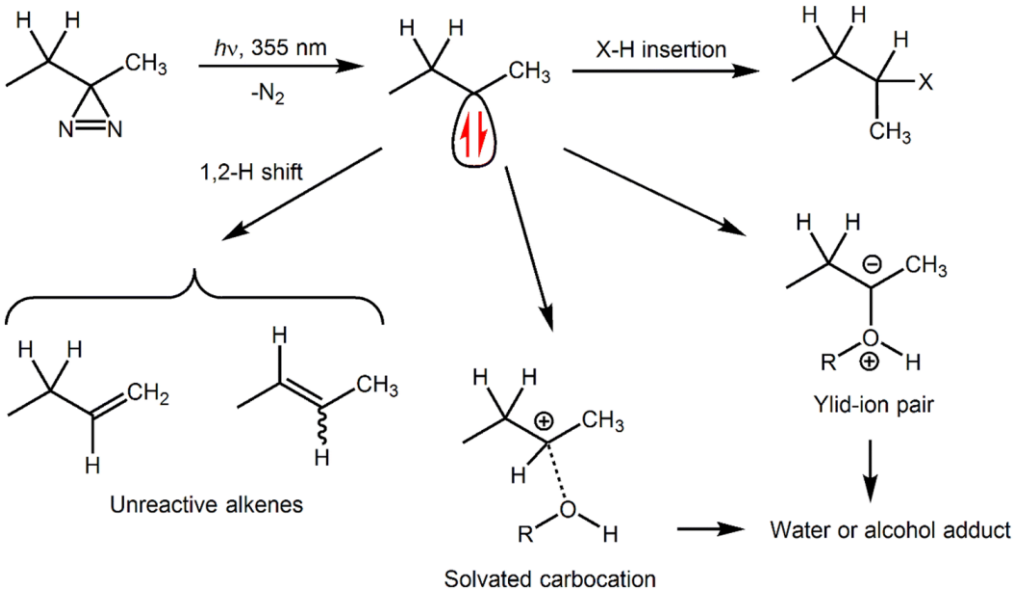
Covalent crosslinking is a well-established technique for probing three-dimensional structures and non-covalent interactions of biomolecular complexes.[13] This method relies on the incorporation of a stable ‘tag’ into the target conjugates or complexes. Under specific conditions, such as UV light irradiation (photochemical crosslinking), a reactive intermediate is generated,

which rapidly crosslinks with nearby functional groups. The resulting products can then be analyzed using mass spectrometry or gel electrophoresis (**Scheme 1.3**). Among the various photo-tags explored, diazirine stands out due to its ability to easily monitor mass changes through N_2 elimination and its versatility in targeting numerous residues via carbene intermediates produced by photodissociation at 355 nm (**Scheme 1.4**).[42,43] It also enjoys the advantage that several amino acids labeled at the side chain are commercially available.[44,45] When used in cells, these labeled amino acids can be recognized by transfer RNAs and incorporated into newly synthesized proteins. In this context, they are referred to as genetically encoded crosslinkers.[46] However, the application of diazirine as photochemical crosslinker is limited by drawbacks such as a low extinction coefficient[47] and the non-specific reactivity of the carbene intermediate[48], which hinder its broader use in photochemical crosslinking studies.

It is noteworthy that most photochemical crosslinking experiments in biomolecules are performed in solution. These experiments require high-concentration samples and often suffer from low yields and non-specific crosslinking, as reactive intermediates like carbenes are easily quenched by the solvent.[50,51] In contrast, gas-phase crosslinking reactions, though less explored, enjoy the feature of higher yields, well-defined stoichiometry and easy monitoring of reaction progress.

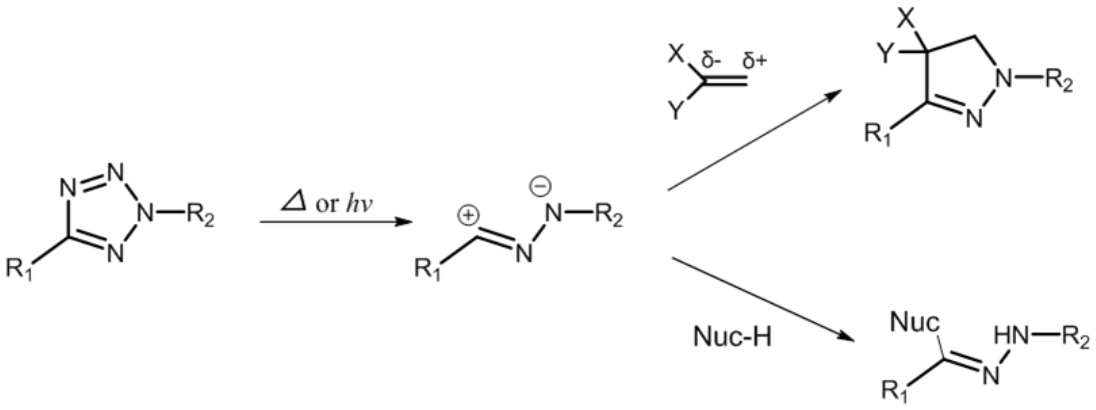


Scheme 1.3: Photochemical crosslinking.



Scheme 1.4: Competitive reactions of carbenes produced by photodissociation of diazirine, reproduced from [49].

In this thesis I will discuss our recently developed tetrazole-tagged peptide conjugates that can be used to study intermolecular or intramolecular photo-crosslinking reactions of various biomolecules in the gas phase. The underlying chemistry involves the thermo- or photodissociation of a 2,5-diaryltetrazole tag, leading to the formation of a nitrile imine intermediate.[52] This dipolar intermediate can selectively react with specific residues or intrinsic groups, such as amide bonds in peptides or nucleobases in DNA, through cycloaddition or nucleophilic attack mechanisms (**Scheme 1.5**). The tetrazole photo-tag can be readily incorporated into the lysine side chain through solid-phase peptide synthesis. The resulting conjugates are then electrosprayed to generate singly charged cations, which can be isolated for detailed analysis using tandem mass spectrometry. The broad absorption band of 2,5-diaryltetrazole derivatives (**Figure 1.1**) allows for easy targeting using our instrument with either fixed- or tunable- wavelength laser sources. In the low-pressure ion trap environment, the photochemical crosslinking yields are significantly improved.



Scheme 1.5: Nitrile imine generation and reactions.

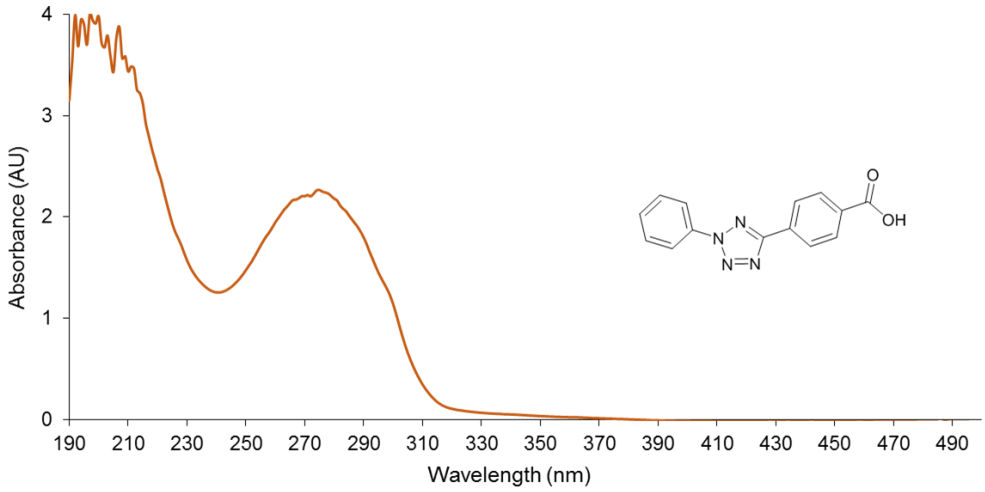


Figure 1.1: UV-Vis absorption of 4-(2-phenyl-2*H*-tetrazol-5-yl)benzoic acid, measured in acetonitrile with 1% acetic acid.

The characterization focused on the structures of nitrile imine intermediates and the various potential crosslinking products. By employing tandem mass spectrometry in combination with UV-Vis action spectroscopy, cyclic ion mobility, and computational analysis, we were able to unambiguously identify the crosslinking sites, providing valuable insights into the mechanisms of this novel reaction. Furthermore, we explored peptide conjugates with different N-terminus and C-terminus residues to investigate their specific non-covalent interactions with the tetrazole photo-tag.

1.2 Novel Mass Spectrometry

Mass Spectrometry (MS) can identify ions based on their mass-to-charge ratio (m/z), thus detecting one or more molecules presented in samples. Although different in shape, size and applications, modern mass spectrometers all consist of three main components: the ionization source, the mass analyzer and the detector. The ionization source converts molecules in the condensed phase to gas-phase ions. The mass analyzer sorts and separates ions based on their m/z. The detector generates mass spectra, plotting the relative ion abundance against m/z. Tandem mass spectrometry (MS^n) provides even more detailed structural information by isolating specific peaks from the MS spectra, activating them, and analyzing the resulting fragments. For example, a peak at -151 Da in a collision-induced dissociation tandem mass spectrum (CID-MS^2) of an oligonucleotide often indicates the elimination of a guanine nucleobase.[53] Additionally, MS^n techniques can offer other information like bond strength and dissociation mechanisms. However, because the mass analyzer differentiates ions only by their m/z values, MS cannot distinguish between molecules with the same mass but different electronic structures or conformations, such as isomers. The electronic structure of isomers relates to their light absorption properties, which correspond to the energy differences between ground states and excited states—information that mass spectrometers alone cannot detect. Molecules with different shapes and sizes can be separated using chromatographic techniques, making instruments like LC-MS invaluable for protein purification and characterization in the biotechnology industry. Unfortunately, many reactive intermediates with short lifetimes in solution are not compatible with chromatography methods. Therefore, to fully characterize the structure and electronic properties of reactive biomolecular ions, it is essential to couple mass spectrometry with complementary techniques.

Mass spectrometers used in the experiments for this thesis include a custom-built instrument comprising a *Bruker* amaZon ion trap with a wavelength-tunable laser and a *Thermo* Orbitrap Ascend equipped with a 213 nm fixed-wavelength laser source. Taking the *Bruker* instrument as an example, I will introduce each component of it in the following sections.

1.2.1 Electrospray Ionization

Electrospray ionization (ESI) is probably one of the most significant advancements in the development and improvement of modern mass spectrometers.[54] It provides gentle, 'soft' ionization for molecules, allowing them to retain much of their structural integrity, which is often not possible with 'hard ionization' techniques like electron ionization[55] or chemical ionization.[56] In addition to preserving the ions' intactness, ESI offers several other advantages. Firstly, it can generate ions with multiple charge states, enabling analytes with extremely high masses to fall within the detection range of mass spectrometers.[57] Secondly, ESI preserves non-covalent interactions, which are crucial for maintaining the higher-order or 'native' structures of biomolecules.[58] This capability has opened new avenues for proteomics and native mass spectrometry studies.

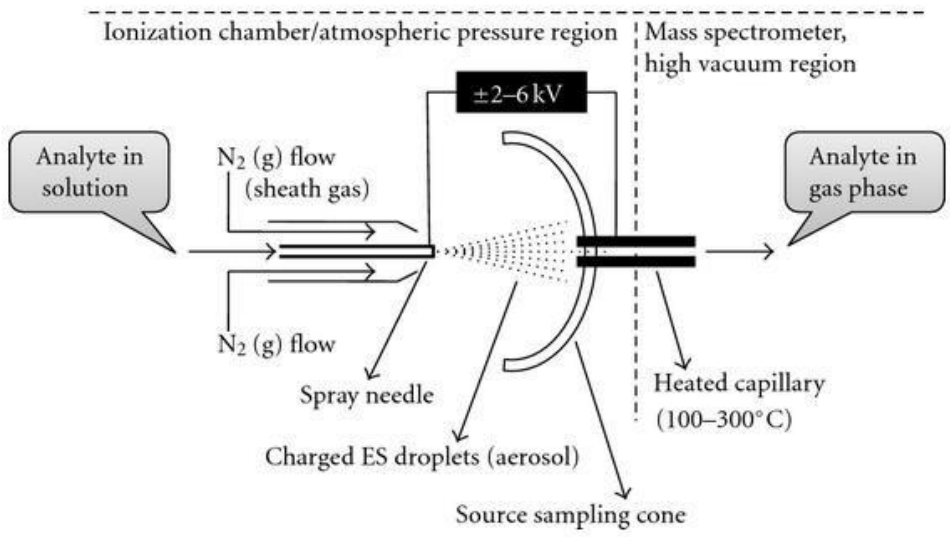


Figure 1.2: A schematic representation of ESI ion source, reproduced from [59].

Despite extensive theoretical modeling and experimental studies, the precise mechanisms of ESI remain ambiguous. Generally, the high voltage applied between the spray needle and the inlet capillary is believed to drive the ion formation, assisted by a stream of heated drying gas. The sample solution, typically at a concentration of 10-100 μM and dissolved in an appropriate solvent, is delivered to the nebulizer via a syringe pump. The nebulizer disperses the solution into small

droplets, and the electric field induces net charges on their surface. As the solvent evaporates, these charges accumulate at the surface of the droplets. When the Coulombic repulsion between the charges exceeds the surface tension, the droplets undergo fission into smaller droplets. This process repeats until the Rayleigh stability limit is reached, at which point bare ions are ejected from the droplets (**Figure 1.2**).[59,60] The ions then pass through a series of ion optics to reach the mass analyzer.

1.2.2 Ion Trap Tandem Mass Spectrometer

Valuable structural information can be extracted from the MS^n analysis of sample molecules. Specifically, various fragmentation methods are applied to the parent ion isolated from the previous stage. The fragmentation techniques used in this thesis include collision-induced dissociation (CID), electron transfer dissociation (ETD), and UV-Vis photodissociation (UVPD). In CID, ions are accelerated and collide with an inert gas, converting kinetic energy into internal energy that excites the ion vibrational states, leading to fragmentation. ETD involves ion-ion reactions between a target multi-cation and a reagent anion, resulting in the transfer of an electron to the target ion. These exothermic reactions release energy that typically causes some degree of fragmentation. In UVPD, a photon with a wavelength corresponding to the electronic excited states is absorbed by the ion, initiating fragmentation.

Tandem mass spectrometers are designed with a series of mass analyzers coupled in space, allowing fragment ions produced at each stage to pass sequentially to the next analyzer.[61] Alternatively, MS^n can be performed using a single mass analyzer by dynamically tuning the electromagnetic field, enabling isolation and fragmentation in a time-sequenced manner.[62] The clear advantage of the latter approach is that it creates a more compact setup while minimizing ion intensity loss during transfer through different ion optics. Among the various mass analyzers, ion traps are particularly well-suited for MS^n due to their versatility in performing multiple functions, such as ion accumulation, selective m/z isolation, excitation, and sequential mass ejection. The basic structure of an ion trap consists of a ring electrode positioned between two end cap electrodes (**Figure 1.3**). These components, with their nearly hyperbolic internal surface shapes, generate a quadrupolar field that traps the ions when a high RF voltage is applied. The trapped ions exhibit periodic motion in both axial and radial directions. By adjusting the RF

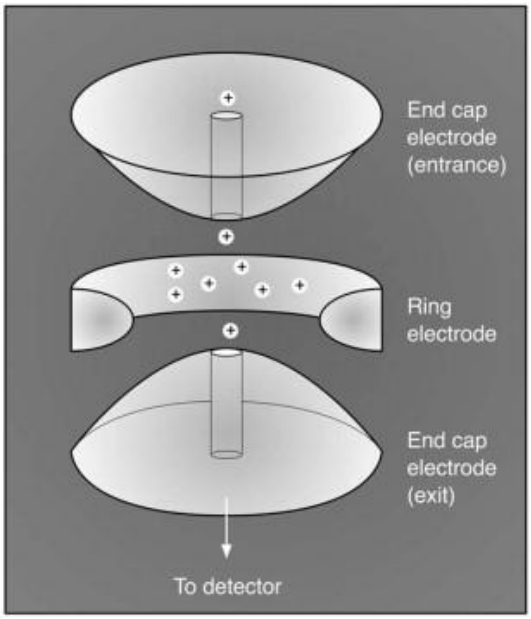


Figure 1.3: The basic structure of an ion trap, reproduced from [64].

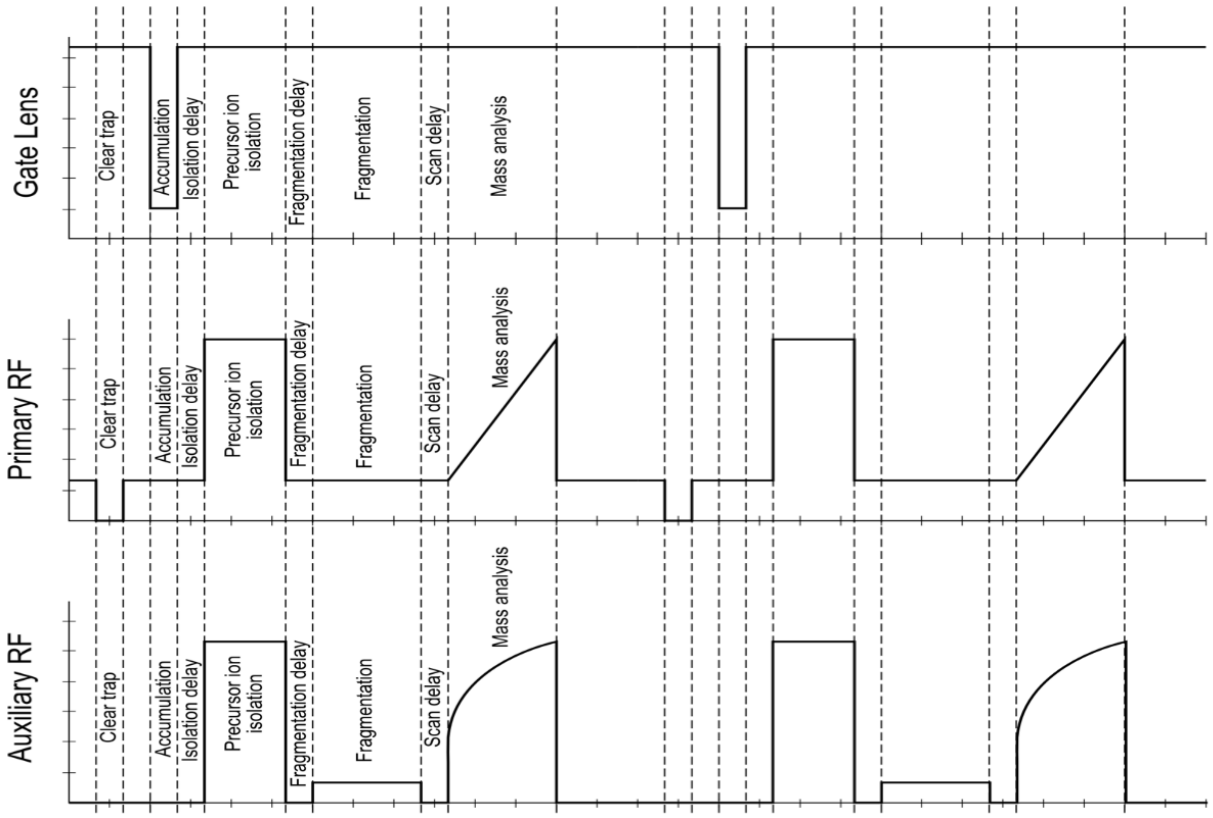


Figure 1.4: The important scan segments for an MS^n scan, reproduced from Figure 3-12 of "amaZon Series User Manual", *Bruker Daltonik GmbH*, 2016.

voltage on the ring electrode and the auxiliary voltage on one or both end cap electrodes, a full MS scan sequence can be executed.[63] For the Bruker amaZon 3D ion trap, an illustration of the complete scan segments is shown in **Figure 1.4**.

At the start of each MS scan, the blocking voltage on the gate lens of the end-cap electrode is lowered so that ions can enter the ion trap where they are cooled down by the damping gas and accumulated for a given time in the low quadrupolar field. To isolate the precursor ion of interest, the electronic system generates a broadband frequency spectrum that contains the resonant frequencies of all ions except the target ion. As a result, all other ions are ejected from the trap, leaving only the precursor ion. To induce fragmentation via CID, the selected ion is accelerated through resonance excitation using a dipole field, causing it to collide with the background gas. During this process, the RF voltage is maintained at an optimal level to ensure that excitation is efficient, while simultaneously trapping all fragment ions within the pseudo-potential well. After fragmentation, the RF field is gradually ramped, enabling all ion products to be ejected sequentially according to their m/z ratios. These ions are detected by a Daly detector, where they are accelerated by an electric field onto a phosphor screen, generating photons that are subsequently captured by a photomultiplier for detection. This process can be repeated to achieve higher stages of MS^n as long as the ion intensity is sufficient and the signal-to-noise ratio remains favorable.

Other types of fragmentation other than CID can be performed. To enable ETD on the Bruker instrument, an auxiliary chemical ionization (CI) source is installed just before the ion trap. During the ETD process, the CI source gate opens, and fluoranthene anion radicals ($C_{16}H_{10}^{-\bullet}$) generated from the CI reagent are introduced into the ion trap to engage in ion-ion reactions with the multiply charged precursor ions. This process facilitates charge reduction and the generation of radical species of interest.

On the modified *Bruker* instrument, UVPD is achieved by directing a wavelength-tunable laser beam into the modified ion trap. Specifically, the laser beam enters and exits the ion trap through two 1-mm holes located on the ring electrodes, positioned at the top and bottom. Two reflection mirrors are equipped with a 45 degree dihedral angle to the optical table surface to guide the laser beams out of the instrument. (**Figure 1.5**). The laser beam is produced by EKSPLA NL301G (Altos Photonics, Bozeman, MT, USA) Nd:YAG laser equipped with a PG142 unit that includes an optical parametric oscillator coupled to a second-harmonic

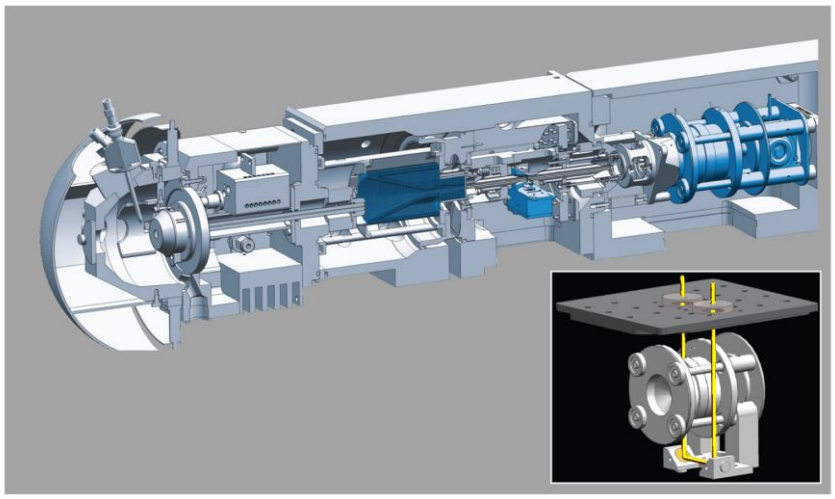


Figure 1.5: The schematic of the *Bruker* ion trap mass spectrometer and electrospray ionization source, with modifications for optical access to the trapped ions highlighted in the inset, reproduced from [66].

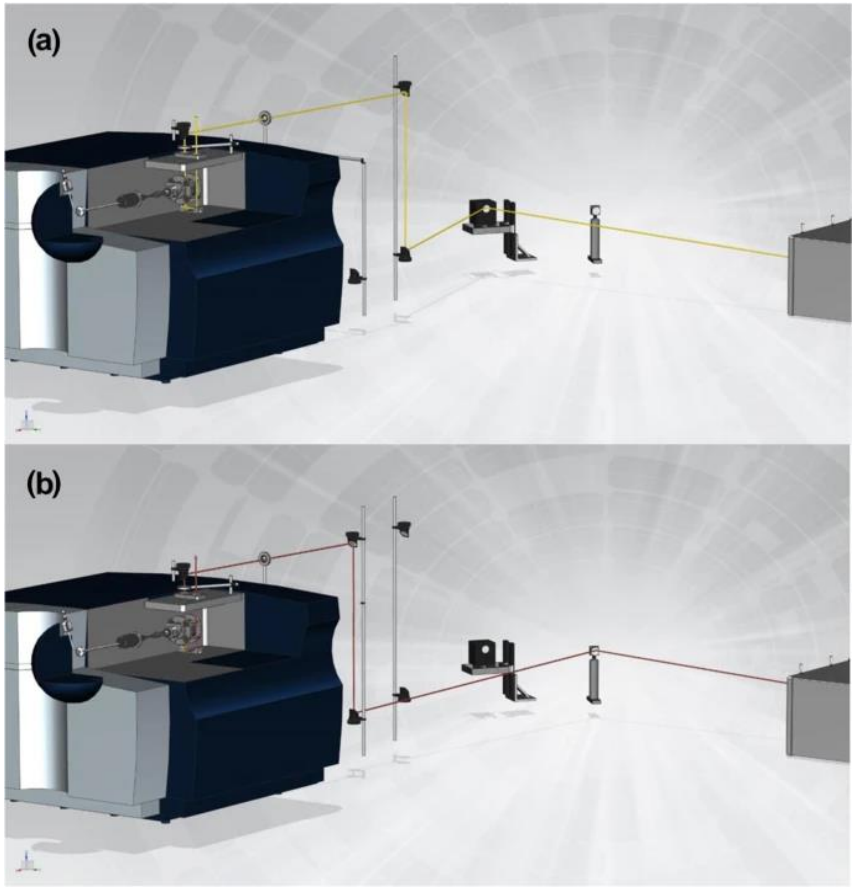


Figure 1.6: Two sets of optics allowing UVPD at (a) 210-700 nm tunable wavelength or (b) 355 nm fixed wavelength at *Bruker* ion trap mass spectrometer.

generator. Two sets of optics are mounted along the optical path from the OPO box to the ion trap, enabling convenient switching between the fixed 355 nm wavelength and the tunable range of 210-700 nm (**Figure 1.6**). During each MS scan, laser pulses are synchronized with the precise time window of the CID stage, with 0 volts of excitation applied, in a process known as 'fake CID'. This setup allows us to control the photodissociation process both spatially and temporally.[65]

Part of the UVPD experiments in this thesis utilized the 213-nm laser on an Orbitrap Ascend Tribrid instrument (ThermoFisher, San Jose, CA). These experiments were conducted either by scanning the ion trap or in the high-resolution mode using the Orbitrap with a resolving power of 100,000. The *Thermo* Ascend instrument differs from the *Bruker* instrument in terms of its mass analyzer types and ion pathway. In the *Thermo* Ascend, ions travel back and forth between the ion optics, moving from the linear ion trap where photodissociation occurs to the Orbitrap for analysis (**Figure 1.7**). Despite these differences in design, the general principles of MS^n on the *Thermo* Ascend are similar to those on the *Bruker* instrument.

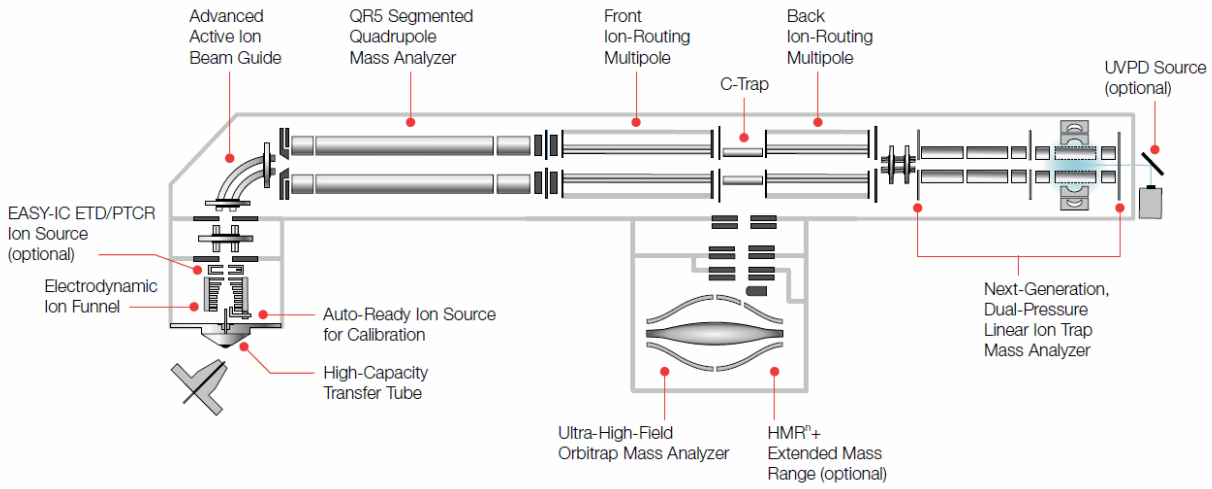


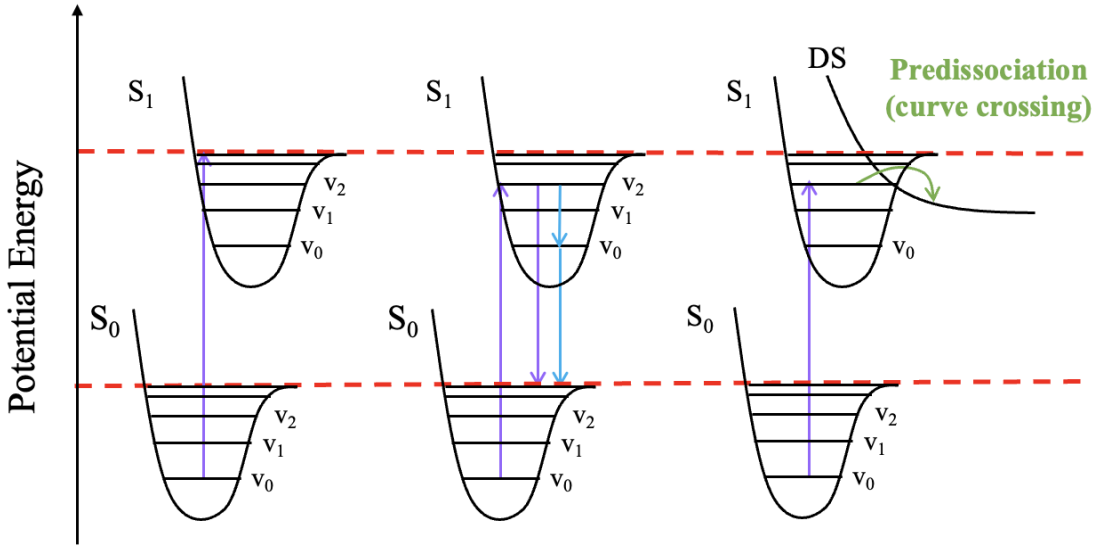
Figure 1.7: The schematic of the *Thermo* Orbitrap Ascend Tribrid instrument components, reproduced from [67].

1.2.3 UV-Vis Photodissociation Action Spectroscopy

Spectroscopic methods provide an orthogonal dimension of structural elucidation for biomolecules by measuring the wavelength-dependent absorption of electromagnetic radiation. The absorption bands can be assigned to certain chromophores within the molecule, such as various functional groups and nuclei. Common spectroscopic techniques, like UV-Vis spectroscopy, are typically performed by measuring the difference between the incident and

transmitted light intensity in appropriately concentrated solutions in the condensed phase. In contrast, ions generated from the ion source and temporarily stored in the high vacuum environment of an ion trap represent optically extremely thin samples. Under these conditions, collecting absorption spectra requires meticulous attention and specialized setups, such as a frozen neon matrix, to ensure accurate measurements.[68]

Instead of directly measuring light absorption, one can measure the ion photodissociation products using a technique known as action spectroscopy.[69] Pioneered by Dunbar and Hettich, this method is now widely employed in gas-phase chemistry to investigate the structures of polyatomic ions.[70,71]



Scheme 1.5: Three potential dissociation pathways following photon absorption by an ion, reproduced from [69].

Considering a single-photon excitation of the ion. **Scheme 1.5** illustrates three different mechanisms of dissociation. In the left panel, the photon energy is high enough to promote a transition from the ground electronic state (S_0) to vibrationally excited electronic excited state (S_1), which lies above the dissociation threshold. In this scenario, absorption of a high-energy photon directly triggers photodissociation. The second mechanism (the middle panel) involves fast internal conversion from an excited electronic state (S_1v_2) to a vibrationally excited ground electronic state. In this case, the conserved energy exceeds the thermodynamic threshold for the lowest-energy dissociation pathway on the ground state potential energy surface. The right panel

depicts the third mechanism where the first excited state (**S_1**) is overlapped with a dissociative higher excited state (**DS**). Nonadiabatic crossing to this dissociative state induces rapid dissociation below the normal energy threshold, practically in a single vibration.

UV-Vis action spectra are generated by monitoring the relative intensity changes of different ion fragments (channels) as a function of wavelength. Ideally, in the absence of multi-photon processes and with a reasonably low dissociation threshold, the action spectra closely resemble the absorption spectra. A significant advantage of UV-Vis action spectroscopy is its multi-channel capability, which provides detailed information about the excited states generated by photon absorption. Since the absorption process is typically a single-photon event, unlike in infrared multiphoton action spectroscopy (IRMPD),[72] it is possible and advantageous to match these action spectra with calculated theoretical spectra at room temperature, aiding in structural assignments.

The UV-Vis action spectroscopy experiments in this thesis were conducted using the modified *Bruker* instrument. Spectra were collected at intervals of 2 nm (from 210 to 409 nm) and 5 nm (from 410 to 700 nm). Key dissociation channels were selected to map the action spectra, and fragment intensities were calibrated based on the laser output energy measured by an EnergyMax-USB J-10 MB energy sensor. Each spectrum was typically collected at least twice to ensure reproducibility.

1.3 Cyclic Ion Mobility–Mass Spectrometry

1.3.1 Mobility and Collision Cross Section of Ions

Ion mobility (IM) emerges as a powerful technique to separate and characterize ions in the gas phase. The core device for ion mobility measurement is a drift cell, which is filled with drift gas ("buffer gas").[73,74] When a uniform electric voltage is applied along the length of the tube, ions are accelerated and move from one side to the other. As the ions travel, they collide with the buffer gas molecules. If the collisions are frequent enough, the damping force from the gas and the electric force balance each other, causing the ions to travel at a constant average speed, known as the drift speed (v_d). The drift speed is proportional to both the electric field strength (E) and the ion mobility

(*K*) (**Equation 1.1**). The ion mobility *K* can be determined from the time it takes for the ion to traverse the drift tube (t_d) and the length of the tube (*L*) (**Equation 1.2**).

$$v_d = K \cdot E \qquad (1.1)$$

$$K = \frac{v_d}{E} = \frac{L}{t_d \cdot E} \qquad (1.2)$$

The mobility describes an ion's ability to travel through the drift tube under specific experimental conditions. The mobility value depends on the ion's mass, charge, and shape, as well as the experimental parameters such as temperature and pressure. Reduced mobility (K_0) refers to the ion mobility normalized to standard conditions (T_0 = 273.15 K, P_0 = 760 Torr) and is widely used for comparison across experiments. It can be calculated using **Equation 1.3**.

$$K_0 = K \frac{P}{P_0} \frac{T_0}{T} \qquad (1.3)$$

By combining **Equation 1.1** to **1.3**, we can derive an expression for the drift time t_d in terms of *P*/*V* (**Equation 1.4**), where *V* is the voltage applied along the drift tube, *P* is the buffer gas pressure, *T* is the temperature, and *L* is the drift tube length.

$$t_d = \frac{L}{v_d} = \frac{L}{K \cdot E} = \frac{L^2 T_0}{K_0 P_0 T} \frac{P}{V} \qquad (1.4)$$

Ion mobility experiments measure the arrival time distribution (ATD) of sample ions. The arrival time (t_a) is the total time an ion takes from entering the drift cell to arriving at the detector, expressed as $t_a = t_d + t_0$, where t_0 accounts for the time spent outside the drift cell. By plotting t_a against *P*/*V*, a linear relationship is obtained, where the intercept represents t_0, and the slope is proportional to $1/K_0$, the inverse of the reduced mobility.

The reduced mobility K_0 of an ion drifting through a buffer gas can be related to its collision cross section (CCS) using kinetic theory (**Equation 1.5**). In this expression, *e* represents the charge on the ion, N_0 is the buffer gas number density at standard pressure P_0 and temperature T_0, μ is the reduced mass of the ion-buffer gas system, and Ω is the CCS. Compact ions with smaller CCS values drift faster and reach the detector sooner, whereas extended ions with larger CCS values

drift more slowly. A useful analogy is a sheet of paper: an unfolded sheet experiences greater air resistance and drifts slowly, while a crumpled piece with the same mass drifts faster due to reduced resistance.

$$K_0 = \frac{3e}{16N_0}\left(\frac{2\pi}{\mu kT}\right)^{1/2}\frac{1}{\Omega} \qquad (1.5)$$

But what does the term Ω truly represent? While 'cross section' suggests that the ion behaves as a rigid sphere, with collision events determined by its surface projection area, the reality is far more complex. Ions consist of nuclei and electrons, which clearly do not resemble hard spheres. Collisions between ions and buffer gas molecules are influenced by both the orientations of the colliding species and the non-covalent interactions between them, such as van der Waals and ion-induced dipole interactions.[75,76] Therefore, Ω represents the momentum transfer collision integral, a more nuanced measure of the interactions.[73] Additionally, the experimental CCS of ions is not an intrinsic property—it depends on the buffer gas used, typically helium or nitrogen. In **Section 1.4**, I will discuss several theoretical methods for calculating the theoretical CCS.

1.3.2 Traveling-Wave c-IMS Instrument and IM-MS Data Analysis

Cyclic ion mobility employs a closed-loop tube instead of a simple linear drift tube, allowing ions to travel multiple cycles within the tube. Since the resolving power (*R*) increases with the path length, or the number of cycles traveled (**Equation 1.6**), cyclic ion mobility instruments can achieve separations that are beyond the capabilities of traditional drift tubes. Additionally, these instruments offer a more compact design. The cyclic ion mobility spectrometer (c-IMS) used in this thesis is the SELECT SERIES cyclic ion mobility spectrometer from Waters Corp. (Wilmslow, UK).[77] The structure and key components are illustrated in **Figure 1.8A**. Typically, ions are first mass-selected using a quadrupole and transferred into the *Trap* region, where they can undergo fragmentation. The fragments are guided through the ion guide (*IG*) and Helium cell before entering the core component: the cyclic ion mobility device (**Figure 1.8B**). The c-IMS device, located orthogonal to the main ion optical path, consists of two main parts: the entry/exit region (**Figure 1.8C**) and the main separation body. The entry/exit region controls the injection and ejection of ions, determining how many cycles they travel. The main separation body contains

608 electrodes supported by printed circuit boards (PCBs), which provide both the structural framework and voltage connections. Each cycle within the c-IMS spans 98 cm, and the device utilizes traveling waves (TW)—a series of periodic voltage pulses—to separate ions.[78] In this process, ions 'surf' on the traveling waves, with higher-mobility ions traveling longer distances before being overtaken by the wave. Ideally, ions can travel through as many cycles as needed for optimal separation. However, in practice, fast-moving ions eventually meet the slower ones again, causing overlap once they exceed a complete cycle. Additionally, ion intensity gradually decreases with multiple cycles. For these reasons, we typically perform 1 to 10 cycles for our biomolecular samples. After exiting the c-IMS, ions pass through two additional PCBs (*IG* and *Transfer*) before reaching the pusher. The pusher injects ion pulses into the time-of-flight (TOF) mass analyzer, where their m/z ratios are detected.

$$R \sim \left(\frac{LEQ}{T}\right)^{1/2} \qquad (1.6)$$

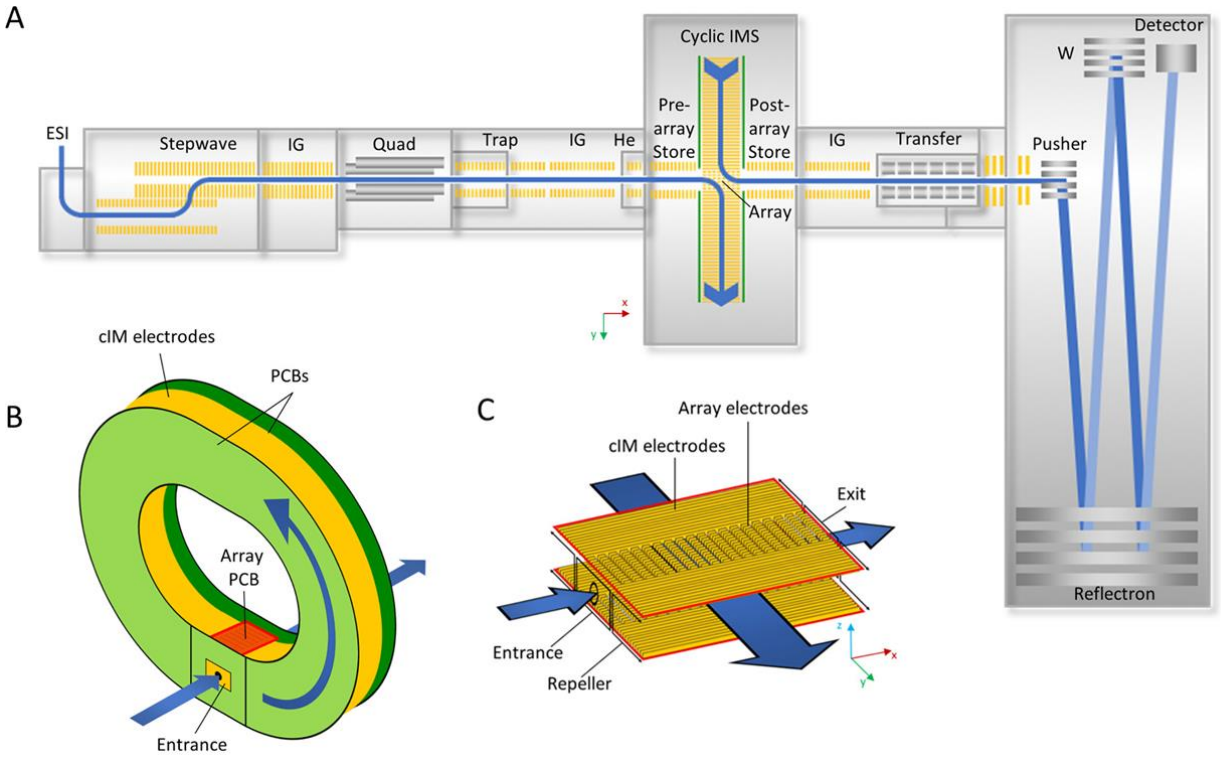


Figure 1.8:. Instrument components of SELECT SERIES c-IMS: (A) instrument overview; (B) c-IMS device; (C) ion entry/exit region, consisting of array electrodes (reproduced from [77]).

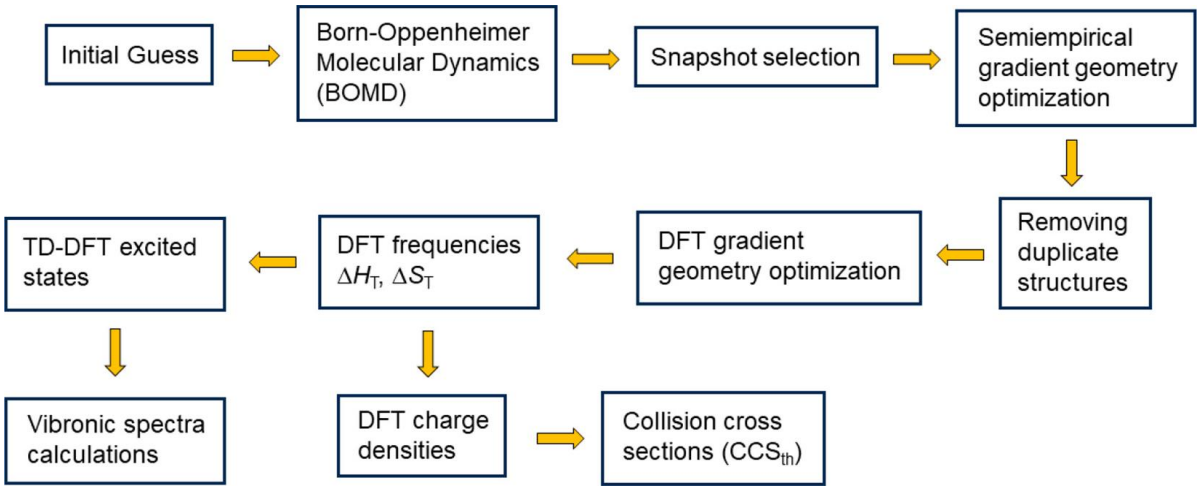
For data measured on the TW-c-IMS instrument, the CCS does not maintain a linear relationship with the drift time (t_d), requiring the use of a calibration curve fitted with a logarithmic function. The calibrants, with known CCS values found in the existing database, should have structural properties similar to those of the analytes to ensure accurate calibration.[79,80] Due to the nature of data collection, arrival time distribution (ATD) diagrams are essentially two-dimensional datasets that combine ion mobility information (t_d) with mass spectrometry data (m/z). Therefore, a careful analysis of isotopic distributions and potential fragmentation events within the c-IMS is essential. For unresolved peaks in the IMS data, Gaussian peak fitting is applied to determine the peak center and width of each component, including any hidden species, ensuring accurate drift time measurement.[81]

1.4 Computational Methods

Experimental results are more convincing when complemented by appropriate theoretical studies. To achieve optimized structures, thermochemistry, and theoretical predictions of spectroscopy and CCS for gas-phase ions, a comprehensive computational workflow was developed (**Scheme 1.6**).

Firstly, several initial structures are constructed using *Gaussview*, with protonation and charge sites assigned based on the chemical properties of functional groups (e.g., basicity) and chemical intuition. These initial structures are then subjected to Born-Oppenheimer Molecular Dynamics (BOMD) simulations using MOPAC within the Cuby4 platform to generate trajectories that produce thousands of structural candidates.[82] The BOMD simulations are typically run at various temperatures (ranging from 310 to 800 K), with total energy stability maintained using the Berendsen thermostat.[83] Notably, since BOMD accounts for both electronic and nuclear motion, the initial protonation sites are not fixed, and proton-transfer isomerization is often observed. Snapshots of protomers and conformers are extracted at regular intervals and fully optimized using semiempirical methods (PM6-D3H4).[84] Distinct conformers within a reasonable energy range are selected for further optimization using density functional theory (DFT). Several rounds of DFT optimizations are typically performed with different levels of theory and progressively larger basis sets. Single-point energies and harmonic frequencies are calculated to determine Gibbs free energy at room temperature. Time-dependent density functional theory (TD-DFT)[85] is used to compute electronic excitation energies and oscillator strengths at 0 K, which are then combined with 300

Boltzmann-weighted vibrational configurations to generate the vibronic absorption spectrum at 310 K using the Newton-X program.[86] Charge densities for each atom are calculated using the Merz–Singh–Kollman (MK) scheme,[87,88] and the theoretical CCS values are then determined with $MobCal_{MPI}$, utilizing the MK charge densities.[89,90]



Scheme 1.6: Workflow of BOMD and density functional theory calculations of structures, excited states, vibronic spectra, and collision cross sections.

There are several methods available for calculating the theoretical CCS (CCS_{theo}) of ions, including the projection approximation,[91,92] elastic hard sphere scattering,[76,93,94] and the trajectory method.[95,96] Among these, the trajectory method, as implemented in $MobCal_{MPI}$, is widely regarded as the most accurate and versatile.[97] To calculate the CCS_{theo} using the trajectory method, one must evaluate the momentum transfer integrals averaged over all possible velocities and geometries of the ion and buffer gas. The scattering angle (χ), used in the calculation of momentum transfer integrals, can only be determined numerically based on the intermolecular potential between the ion and the buffer gas.[98] Therefore, selecting appropriate parameters to represent van der Waals forces, ion-induced dipole interactions, and ion-quadrupole interactions is essential.[99-101] The Merck Molecular Force Field (MMFF94) van der Waals parameters have so far been shown to produce CCS_{theo} values that align most closely with experimental data.[102] Additionally, the Merz–Singh–Kollman (MK) charge densities, based on the electrostatic potential, are preferred for these calculations because they are less sensitive to the choice of basis sets.[103]

1.5 Bibliography

[1] Watson, J. D.; Crick, F .H. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **1953**, *18*, 123–131.

[2] Hsiao, M.-C.; Piotrowski, A.; Alexander, J.; Callens, T.; Fu, C.; Mikhail, F. M.; Claes, K. B. M.; Messiaen, L. Palindrome-Mediated and Replication-Dependent Pathogenic Structural Rearrangements within the *NF*1 Gene. *Human Mutation* **2014**, *35* (7), 891–898.

[3] Aloy, P.; Böttcher, B.; Ceulemans, H.; Leutwein, C.; Mellwig, C.; Fischer, S.; Gavin, A.-C.; Bork, P.; Superti-Furga, G.; Serrano, L.; Russell, R. B. Structure-Based Assembly of Protein Complexes in Yeast. *Science* **2004**, *303* (5666), 2026–2029.

[4] Dey, S.; Pal, A.; Chakrabarti, P.; Janin, J. The Subunit Interfaces of Weakly Associated Homodimeric Proteins. *Journal of Molecular Biology* **2010**, *398* (1), 146–160.

[5] Xu, J.; Reumers, J.; Couceiro, J. R.; De Smet, F.; Gallardo, R.; Rudyak, S.; Cornelis, A.; Rozenski, J.; Zwolinska, A.; Marine, J.-C.; Lambrechts, D.; Suh, Y.-A.; Rousseau, F.; Schymkowitz, J. Gain of Function of Mutant P53 by Coaggregation with Multiple Tumor Suppressors. *Nat. Chem. Biol.* **2011**, *7* (5), 285–295.

[6] Hashimoto, M.; Rockenstein, E.; Crews, L.; Masliah, E. Role of Protein Aggregation in Mitochondrial Dysfunction and Neurodegeneration in Alzheimer's and Parkinson's Diseases. *NeuroMolecular Med.* **2003**, *4* (1), 21–35.

[7] Fraser, J. S.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. Accessing Protein Conformational Ensembles Using Room-Temperature X-Ray Crystallography. *Proceedings of the National Academy of Sciences* **2011**, *108* (39), 16247–16252.

[8] Dyson, H. J.; Wright, P. E. Elucidation of the Protein Folding Landscape by NMR. In *Methods in Enzymology*; Academic Press, 2005; Vol. 394, pp 299–321.

[9] Merk, A.; Bartesaghi, A.; Banerjee, S.; Falconieri, V.; Rao, P.; Davis, M. I.; Pragani, R.; Boxer, M. B.; Earl, L. A.; Milne, J. L. S.; Subramaniam, S. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **2016**, *165* (7), 1698–1707.

[10] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64–71.

[11] Butalewicz, J. P.; Escobar, E. E.; Wootton, C. A.; Theisen, A.; Park, M. A.; Seeley, E. H.; Brodbelt, J. S. Conformational Characterization of Peptides and Proteins by 193 Nm Ultraviolet Photodissociation in the Collision Cell of a Trapped Ion Mobility Spectrometry-Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2024**, *ASAP*.

[12] Stephenson, J. L.; McLuckey, S. A. Ion/Ion Reactions in the Gas Phase: Proton Transfer Reactions Involving Multiply-Charged Proteins. *J. Am. Chem. Soc.* **1996**, *118* (31), 7390–7397.

[13] 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* (8), 7500–7531.

[14] Wyttenbach, T.; Bowers, M. T. Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility–Mass Spectrometry Environment. *J. Phys. Chem. B* **2011**, *115* (42), 12266–12275.

[15] Zhong, Y.; Han, L.; Ruotolo, B. T. Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. *Angewandte Chemie International Edition* **2014**, *53* (35), 9209–9212.

[16] Karch, K. R.; Coradin, M.; Zandarashvili, L.; Kan, Z.-Y.; Gerace, M.; Englander, S. W.; Black, B. E.; Garcia, B. A. Hydrogen-Deuterium Exchange Coupled to Top- and Middle-Down Mass Spectrometry Reveals Histone Tail Dynamics before and after Nucleosome Assembly. *Structure* **2018**, *26* (12), 1651-1663.

[17] JF Ward. Complexity of damage produced by ionizing radiation. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 65, pages 377–382. Cold Spring Harbor Laboratory Press, 2000.

[18] Chen, J.; Potlapalli, R.; Quan, H.; Chen, L.; Xie, Y.; Pouriyeh, S.; Sakib, N.; Liu, L.; Xie, Y. Exploring DNA Damage and Repair Mechanisms: A Review with Computational Insights. *BioTech* **2024**, *13* (1).

[19] Clemens von Sonntag. *Free-radical-induced DNA damage and its repair.* Springer, 2006.

[20] Boudaïffa, B.; Cloutier, P.; Hunting, D.; Huels, M. A.; Sanche, L. Resonant Formation of DNA Strand Breaks by Low-Energy (3 to 20 eV) Electrons. *Science* **2000**, *287* (5458), 1658–1660.

[21] Michael, B. D.; O'Neill, P. A Sting in the Tail of Electron Tracks. *Science* **2000**, *287* (5458), 1603–1604.

[22] Sanche, L. Low Energy Electron-Driven Damage in Biomolecules. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics* **2005**, *35* (2), 367–390.

[23] Alizadeh, E.; Orlando, T. M.; Sanche, L. Biomolecular Damage Induced by Ionizing Radiation: The Direct and Indirect Effects of Low-Energy Electrons on DNA. *Annual Review of Physical Chemistry*, **2015**, *66*, 379–398.

[24] Ward, J. F. DNA Damage Produced by Ionizing Radiation in Mammalian Cells: Identities, Mechanisms of Formation, and Reparability. In *Progress in Nucleic Acid Research and Molecular Biology*; Cohn, W. E., Moldave, K., Eds.; Academic Press, 1988; Vol. 35, pp 95–125.

[25] Alizadeh, E.; Sanz, A. G.; García, G.; Sanche, L. Radiation Damage to DNA: The Indirect Effect of Low-Energy Electrons. *J. Phys. Chem. Lett.* **2013**, *4* (5), 820–825.

[26] Steenken, S. Purine Bases, Nucleosides, and Nucleotides: Aqueous Solution Redox Chemistry and Transformation Reactions of Their Radical Cations and e- and OH Adducts. *Chem. Rev.* **1989**, *89* (3), 503–520.

[27] Candeias, L. P.; Wolf, P.; O'Neill, P.; Steenken, S. Reaction of Hydrated Electrons with Guanine Nucleosides: Fast Protonation on Carbon of the Electron Adduct. *J. Phys. Chem.* **1992**, *96* (25), 10302–10307.

[28] Steenken, S.; Jovanovic, S. V. How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *J. Am. Chem. Soc.* **1997**, *119* (3), 617–618.

[29] Barnes, J.; Bernhard, W. A. The Protonation State of One-Electron Reduced Cytosine and Adenine. 1. Initial Protonation Sites at Low Temperatures in Glassy Solids. *J. Phys. Chem.* **1993**, *97* (13), 3401–3408.

[30] Barnes, J.; Bernhard, W. A. Irreversible Protonation Sites of One-Electron-Reduced Adenine: Comparisons between the C5 and the C2 or C8 Protonation Sites. *J. Phys. Chem.* **1994**, *98* (42), 10969–10977.

[31] An, R.; Jia, Y.; Wan, B.; Zhang, Y.; Dong, P.; Li, J.; Liang, X. Non-Enzymatic Depurination of Nucleic Acids: Factors and Mechanisms. *PLOS ONE* **2015**, *9* (12), e115950.

[32] Lindahl, T.; Andersson, A. Rate of Chain Breakage at Apurinic Sites in Double-Stranded Deoxyribonucleic Acid. *Biochemistry* **1972**, *11* (19), 3618–3623.

[33] Huang, S. R.; Tureček, F. Noncanonical Isomers of Nucleoside Cation Radicals: An Ab Initio Study of the Dark Matter of DNA Ionization. *J. Phys. Chem. A* **2022**, *126* (16), 2480–2497.

[34] Liu, Y.; Huang, S. R.; Tureček, F. Guanine-Adenine Interactions in DNA Tetranucleotide Cation Radicals Revealed by UV/Vis Photodissociation Action Spectroscopy and Theory. *Phys. Chem. Chem. Phys.* **2020**, *22*, 16831– 16842.

[35] Liu, Y.; Dang, A.; Urban, J.; Tureček, F. Charge-Tagged DNA Radicals in the Gas Phase Characterized by UV-Vis Photodissociation Action Spectroscopy. *Angew. Chem., Int. Ed. Engl.* **2020**, *59*, 7772– 7777.

[36] Rodgers, M. T.; Campbell, S.; Marzluff, E. M.; Beauchamp, J. L. Low-Energy Collision-Induced Dissociation of Deprotonated Dinucleotides: Determination of the Energetically Favored Dissociation Pathways and the Relative Acidities of the Nucleic Acid Bases. *International Journal of Mass Spectrometry and Ion Processes* **1994**, *137*, 121–149.

[37] Wee, S.; O'Hair, R. A. J.; McFadyen, W. D. Can Radical Cations of the Constituents of Nucleic Acids Be Formed in the Gas Phase Using Ternary Transition Metal Complexes?. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1797– 1805.

[38] Feketeová, L.; Chan, B.; Khairallah, G. N.; Steinmetz, V.; Maître, P.; Radom, L.; O'Hair, R. A. J. Gas-Phase Structure and Reactivity of the Keto Tautomer of the Deoxyguanosine Radical Cation. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25837– 25844.

[39] Dang, A.; Liu, Y.; Tureček, F. UV-Vis Action Spectroscopy of Guanine, 9-Methylguanine and 2′-Deoxyguanosine Cation Radicals in the Gas Phase. *J. Phys. Chem. A* **2019**, *123*, 3272– 3284.

[40] Zima, V.; Liu, Y.; Tureček, F. Radical Cascade Dissociation Pathways to Unusual Nucleobase Cation Radicals. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (6), 1038–1047.

[41] Zima, V.; Gladwish, O.; Marek, A.; Tureček, F. Nucleoside Cation Radicals: Generation, Radical-Induced Hydrogen Atom Migrations, and Ribose Ring Cleavage in the Gas Phase. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (7), 1594–1608.

[42] Shigdel, U. K.; Zhang, J.; He, C. Diazirine-Based DNA Photo-Cross-Linking Probes for the Study of Protein-DNA Interactions. *Angew. Chem., Int. Ed.* **2008**, *47*, 90– 93.

[43] Dubinsky, L.; Krom, B. P.; Meijler, M. M. Diazirine Based Photoaffinity Labeling. *Bioorg. Med. Chem.* **2012**, *20*, 554– 570.

[44] Suchanek, M.; Radzikowska, A.; Thiele, C. Photo-Leucine and Photo-Methionine Allow Identification of Protein-Protein Interactions in Living Cells. *Nature Methods* **2005**, *2* (4), 261–268.

[45] Yang, T.; Li, X.-M.; Bao, X.; Fung, Y. M. E.; Li, X. D. Photo-Lysine Captures Proteins That Bind Lysine Post-Translational Modifications. *Nature Chemical Biology* **2016**, *12* (2), 70–72.

[46] Aydin, Y.; Coin, I. Genetically Encoded Crosslinkers to Address Protein–Protein Interactions. *Protein Science* **2023**, *32* (5), e4637.

[47] Frey, H. M.; Stevens, I. D. R. 660. The Photolysis of Dimethyldiazirine. *J. Chem. Soc.* **1963**, No. 0, 3514–3519.

[48] Ziemianowicz, D. S.; Bomgarden, R.; Etienne, C.; Schriemer, D. C. Amino Acid Insertion Frequencies Arising from Photoproducts Generated Using Aliphatic Diazirines. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (10), 2011–2021.

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

[50] Sinz, A. Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein–Protein Interactions: Where Are We Now and Where Should We Go from Here? *Angewandte Chemie International Edition* **2018**, *57* (22), 6390–6396.

[51] Yu, C.; Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* **2018**, *90* (1), 144–165.

[52] 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.

[53] McLafferty, F. W.; Turecek, F. *Interpretation Of Mass Spectra*; Organic chemistry series; University Science Books, 1993.

[54] Fenn, J. B. Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angewandte Chemie International Edition* **2003**, *42* (33), 3871–3894.

[55] Maciel, E. V. S.; Pereira dos Santos, N. G.; Vargas Medina, D. A.; Lanças, F. M. Electron Ionization Mass Spectrometry: Quo Vadis? *ELECTROPHORESIS* **2022**, *43* (15), 1587–1600.

[56] Richter, W. J.; Schwarz, H. Chemical Ionization—A Mass-Spectrometric Analytical Procedure of Rapidly Increasing Importance. *Angewandte Chemie International Edition in English* **1978**, *17* (6), 424–439.

[57] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64–71.

[58] Przybylski, M.; Glocker, M. O. Electrospray Mass Spectrometry of Biomacromolecular Complexes with Noncovalent Interactions—New Analytical Perspectives for Supramolecular Chemistry and Molecular Recognition Processes. *Angewandte Chemie International Edition in English* **1996**, *35* (8), 806–826.

[59] Banerjee, S.; Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry* **2012**, *2012* (1), 282574.

[60] Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9.

[61] McLafferty, F. W. Tandem Mass Spectrometry. *Science* **1981**, *214* (4518), 280–287.

[62] March, R. E. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *Journal of Mass Spectrometry* **1997**, *32* (4), 351–369.

[63] March, R. E. Quadrupole Ion Traps. *Mass Spectrometry Reviews* **2009**, *28* (6), 961–989.

[64] Stauffer, E.; Dolan, J. A.; Newman, R. CHAPTER 8 - Gas Chromatography and Gas Chromatography—Mass Spectrometry. In *Fire Debris Analysis*; Stauffer, E., Dolan, J. A., Newman, R., Eds.; Academic Press: Burlington, 2008; pp 235–293.

[65] 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* (9), 1558–1564.

[66] Martens, J.; Berden, G.; Gebhardt, C. R.; Oomens, J. Infrared Ion Spectroscopy in a Modified Quadrupole Ion Trap Mass Spectrometer at the FELIX Free Electron Laser Laboratory. *Review of Scientific Instruments* **2016**, *87* (10), 103108.

[67] He, Y.; Shishkova, E.; Peters-Clarke, T. M.; Brademan, D. R.; Westphall, M. S.; Bergen, D.; Huang, J.; Huguet, R.; Senko, M. W.; Zabrouskov, V.; McAlister, G. C.; Coon, J. J. Evaluation of the Orbitrap Ascend Tribrid Mass Spectrometer for Shotgun Proteomics. *Anal. Chem.* **2023**, *95* (28), 10655–10663.

[68] Leutwyler, S.; Maier, J. P.; Spittel, U. The Electronic Absorption Spectrum of ICCCCI+ in a Ne Matrix. *Chemical Physics Letters* **1983**, *96* (6), 645–648.

[69] Stedwell, C. N.; Polfer, N. C. Spectroscopy and the Electromagnetic Spectrum. In *Laser Photodissociation and Spectroscopy of Mass-separated Biomolecular Ions*; Springer, 2013; pp 1–20.

[70] Dunbar, R. C. Photodissociation of the Methyl Chloride (CH3Cl+) and Nitrous Oxide (N_2O+) Cations. *J. Am. Chem. Soc.* **1971**, *93* (18), 4354–4358.

[71] Hettich, R. L.; Freiser, B. S. Gas-Phase Photodissociation of FeCH2+ and CoCH2+: Determination of the Carbide, Carbyne, and Carbene Bond Energies. *J. Am. Chem. Soc.* **1986**, *108* (10), 2537–2540.

[72] Oomens, J.; Sartakov, B. G.; Meijer, G.; von Helden, G. Gas-Phase Infrared Multiple Photon Dissociation Spectroscopy of Mass-Selected Molecular Ions. *International Journal of Mass Spectrometry* **2006**, *254* (1), 1–19.

[73] Mason, E. A.; McDaniel, E. W. Introduction in *Transport Properties of Ions in Gases*; Wiley Online Library, 1988; Vol. 26.

[74] Kemper, P. R.; Dupuis, N. F.; Bowers, M. T. A New, Higher Resolution, Ion Mobility Mass Spectrometer. *International Journal of Mass Spectrometry* **2009**, *287* (1), 46–57.

[75] Wyttenbach, T.; Helden, G. von; Batka, J. J.; Carlat, D.; Bowers, M. T. Effect of the Long-Range Potential on Ion Mobility Measurements. *Journal of the American Society for Mass Spectrometry* **1997**, *8* (3), 275–282.

[76] Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chemical Physics Letters* **1996**, *261* (1), 86–91.

[77] 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* (13), 8564–8573.

[78] Richardson, K.; Langridge, D.; Giles, K. Fundamentals of Travelling Wave Ion Mobility Revisited: I. Smoothly Moving Waves. *International Journal of Mass Spectrometry* **2018**, *428*, 71–80.

[79] Baker, E. S.; Hoang, C.; Uritboonthai, W.; Heyman, H. M.; Pratt, B.; MacCoss, M.; MacLean, B.; Plumb, R.; Aisporna, A.; Siuzdak, G. METLIN-CCS: An Ion Mobility Spectrometry Collision Cross Section Database. *Nature Methods* **2023**, *20* (12), 1836–1837.

[80] Bush, M. F.; Campuzano, I. D. G.; Robinson, C. V. Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Anal. Chem.* **2012**, *84* (16), 7124–7130.

[81] Kune, C.; Far, J.; De Pauw, E. Accurate Drift Time Determination by Traveling Wave Ion Mobility Spectrometry: The Concept of the Diffusion Calibration. *Anal. Chem.* **2016**, *88* (23), 11639–11646.

[82] Řezáč, J. Cuby: An Integrative Framework for Computational Chemistry. *Journal of Computational Chemistry* **2016**, *37* (13), 1230–1237.

[83] 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. *The Journal of Chemical Physics* **1984**, *81* (8), 3684–3690.

[84] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semiempirical 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* (7), 1749–1760.

[85] Furche, F.; Ahlrichs, R. Adiabatic Time-Dependent Density Functional Methods for Excited State Properties. *The Journal of Chemical Physics* **2002**, *117* (16), 7433–7447.

[86] Barbatti, M.; Ruckenbauer, M.; Plasser, F.; Pittner, J.; Granucci, G.; Persico, M.; Lischka, H. Newton-X: A Surface-Hopping Program for Nonadiabatic Molecular Dynamics. *WIREs Computational Molecular Science* **2014**, *4* (1), 26–33.

[87] Besler, B. H.; Merz Jr., K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *Journal of Computational Chemistry* **1990**, *11* (4), 431–439.

[88] Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *Journal of Computational Chemistry* **1984**, *5* (2), 129–145.

[89] 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* (5), 1660–1670.

[90] Ieritano, C.; Hopkins, W. S. Assessing Collision Cross Section Calculations Using MobCal-MPI with a Variety of Commonly Used Computational Methods. *Materials Today Communications* **2021**, *27*, 102226.

[91] Bleiholder, C.; Wyttenbach, T.; Bowers, M. T. A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (I). Method. *International Journal of Mass Spectrometry* **2011**, *308* (1), 1–10.

[92] Bleiholder, C.; Contreras, S.; Do, T. D.; Bowers, M. T. A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (II). Model Parameterization and Definition of Empirical Shape Factors for Proteins. *International Journal of Mass Spectrometry* **2013**, *345–347*, 89–96.

[93] Shvartsburg, A. A.; Liu, B.; Jarrold, M. F.; Ho, K.-M. Modeling Ionic Mobilities by Scattering on Electronic Density Isosurfaces: Application to Silicon Cluster Anions. *The Journal of Chemical Physics* **2000**, *112* (10), 4517–4526.

[94] Larriba-Andaluz, C.; Fernández-García, J.; Ewing, M. A.; Hogan, C. J.; Clemmer, D. E. Gas Molecule Scattering & Ion Mobility Measurements for Organic Macro-Ions in He versus N2 Environments. *Phys. Chem. Chem. Phys.* **2015**, *17* (22), 15019–15029.

[95] Campuzano, I.; Bush, M. F.; Robinson, C. V.; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I. Structural Characterization of Drug-like Compounds by Ion Mobility Mass Spectrometry: Comparison of Theoretical and Experimentally Derived Nitrogen Collision Cross Sections. *Anal. Chem.* **2012**, *84* (2), 1026–1033.

[96] Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 587–596.

[97] Ross, D. H.; Bhotika, H.; Zheng, X.; Smith, R. D.; Burnum-Johnson, K. E.; Bilbao, A. Computational Tools and Algorithms for Ion Mobility Spectrometry-Mass Spectrometry. *PROTEOMICS* **2024**, *24* (12–13), 2200436.

[98] Li, Z.; Wang, H. Drag Force, Diffusion Coefficient, and Electric Mobility of Small Particles. I. Theory Applicable to the Free-Molecule Regime. *Phys. Rev. E* **2003**, *68* (6), 061206.

[99] Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem. A* **1997**, *101* (5), 968–968.

[100] Kim, H.; Kim, H. I.; Johnson, P. V.; Beegle, L. W.; Beauchamp, J. L.; Goddard, W. A.; Kanik, I. Experimental and Theoretical Investigation into the Correlation between Mass and Ion Mobility for Choline and Other Ammonium Cations in N2. *Anal. Chem.* **2008**, *80* (6), 1928–1936.

[101] Kim, H. I.; Kim, H.; Pang, E. S.; Ryu, E. K.; Beegle, L. W.; Loo, J. A.; Goddard, W. A.; Kanik, I. Structural Characterization of Unsaturated Phosphatidylcholines Using Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **2009**, *81* (20), 8289–8297.

[102] Halgren, T. A. Merck Molecular Force Field. II. MMFF94 van Der Waals and Electrostatic Parameters for Intermolecular Interactions. *Journal of Computational Chemistry* **1996**, *17* (5–6), 520–552.

[103] North, S. C.; Jorgensen, K. R.; Pricetolstoy, J.; Wilson, A. K. Population Analysis and the Effects of Gaussian Basis Set Quality and Quantum Mechanical Approach: Main Group through Heavy Element Species. *Frontiers in Chemistry* **2023**, *11*.

Chapter 2

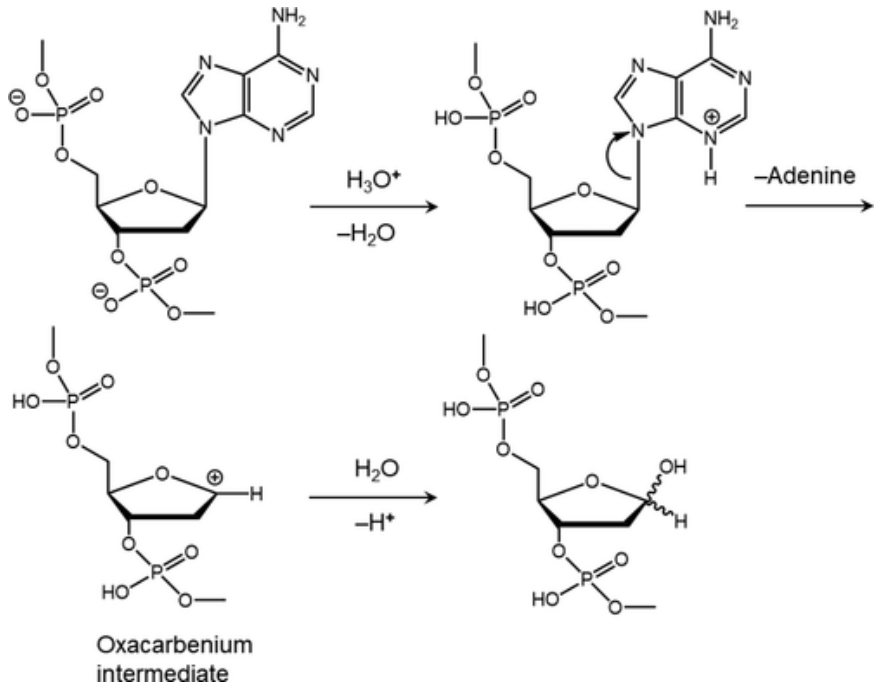
Structures and Dissociations of Mono- and Diprotonated DNA Trinucleotide Codons in the Gas Phase

Reproduced in part with permission from Jiahao Wan, Břetislav Brož, Yue Liu, Shu R. Huang, Aleš Marek and František Tureček. Resolution of Identity in Gas-Phase Dissociations of Mono- and Diprotonated DNA Trinucleotide Codons by ^{15}N-Labeling and Computational Structure Analysis. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 10, 1936–1950.

Abstract *Dissociations of DNA trinucleotide codons as gas-phase singly and doubly protonated ions were studied by tandem mass spectrometry using ^{15}N-labeling to resolve identity in the nucleobase loss and backbone cleavages. The monocations showed different distributions of nucleobase loss from the 5′-, middle, and 3′-positions depending on the nucleobase, favoring cytosine over guanine, adenine, and thymine in an ensemble-averaged 62:27:11:<1 ratio. The distribution for the loss of the 5′-, middle, and 3′-nucleobase was 49:18:33, favoring the 5′-nucleobase, but also depending on its nature. The formation of sequence* $\boldsymbol{w_2^+}$ *ions was unambiguously established for all codon mono- and dications. Structures of low-Gibbs-energy protomers and conformers of* $dAAA^+$, $dGGG^+$, $dCCC^+$, $dTTT^+$, $dACA^+$, *and* $dATC^+$ *were established by Born–Oppenheimer molecular dynamics and density functional theory calculations. Monocations containing guanine favored classical structures protonated at guanine N7. Structures containing adenine and cytosine produced classical nucleobase-protonated isomers as well as zwitterions in which two protonated bases were combined with a phosphate anion. Protonation at thymine was disfavored. Low threshold energies for nucleobase loss allowed extensive proton migration to occur prior to dissociation. Loss of the nucleobase from monocations was assisted by neighboring group participation in nucleophilic addition or proton abstraction, as well as allosteric proton migrations remote from the reaction center. The optimized structures of diprotonated isomers for* $dAAA^{2+}$ *and* $dACA^{2+}$ *revealed combinations of classical and zwitterionic structures. The threshold and transition-state energies for nucleobase-ion loss from dications were low, resulting in facile dissociations involving cytosine, guanine, and adenine.*

2.1 Introduction

Acid hydrolysis of DNA results in elimination of purine nucleobases in a process known as depurination.[1–3] Although nonenzymatic in nature, acid-catalyzed depurination is thought to play a role in nucleic acid metabolism in acidic gastric juice and in other low-pH environments such as in lysosomes.[1] The DNA depurinated sites are vulnerable to strand breaks[4,5] and, if unrepaired, can contribute to carcinogenesis.[6] The stability of DNA is important not only for biological materials but also in connection with the developing data storage technology where loss of a nucleobase would impair information retrieval.[7] Acid degradation of oligonucleotides in aqueous solution proceeds with first-order kinetics[1,8–11] with respect to the nucleotide and depends on pH. This has led to the suggested mechanism of depurination that presumes protonation of the nucleobase in the first step, followed by glycosidic bond cleavage, which forms a deoxyribose C1′ oxocarbenium cation as an intermediate (**Scheme 2.1**).[11]



Scheme 2.1: Previously Suggested Mechanism of Acid-Catalyzed Nucleobase Elimination

However, the presumed oxacarbenium intermediate has not been characterized, as it is rapidly attacked by solvent, introducing a hydroxyl group at sugar C1′ and thereby finishing the hydrolysis. As in other numerous cases of reactive intermediates of chemical reactions,[12] gas-

phase chemistry of isolated ions is ideally suited to tackle this problem by eliminating the solvent and allowing one to study the intrinsic properties of oligonucleotide-related oxocarbenium cations by tandem mass spectrometry. An additional benefit of studying the ion dissociations in the gas phase stems from the possibility of protonating any of the four DNA nucleobases, which is hindered in solution because of the low basicity of pyrimidine nucleobases, especially thymine. Dissociations of simple nucleosides and nucleotides have been addressed by Rodgers and co-workers in their energy-resolved guided ion beam studies of protonated adenine,[13] guanine,[14] and cytosine[15] nucleosides, as well as some fluorine-modified nucleosides.[16,17] However, structures and dissociation energetics of more complex gas-phase oligonucleotide cations have not been addressed. Multiple protonation, in particular, can affect the relative stabilities of nucleobase protomers, as reported for singly protonated 2′-deoxyadenosine[18] and doubly protonated dinucleotide $(dAA+2H)^{2+}$.[19]

We now report a combined experimental and computational study of gas-phase dissociations of trinucleotide mono- and dications covering the DNA genetic code alphabet. A previous thorough study by O'Hair, Reid, and their co-workers has laid out the basic dissociations of DNA codon cations,[20] focusing on the loss of nucleobase and backbone cleavages in dependence on the nature and position of the nucleobases. However, in symmetrical trinucleotides of the dXXX and dXYX type, loss of identical neutral nucleobases X from the different positions was unresolved. The same applies to backbone cleavages occurring between the 5′-, middle, and 3′-nucleotide positions that can produce isobaric fragments of identical elemental composition. We achieve a resolution of this identity by analyzing dissociations of trinucleotides in which the nucleobases are distinguished by site-specific ^{15}N labeling, resulting in mass shifts by 5 Da for [$^{15}N_5$]adenine and [$^{15}N_5$]guanine, 3 Da for [$^{15}N_3$]cytosine, and 2 Da for [$^{15}N_2$]thymine. In addition, for several codons we report a thorough structural analysis of monocation and dication structures and dissociation energetics and kinetics using a combination of Born–Oppenheimer molecular dynamics (BOMD), density functional theory (DFT) calculations, and Rice–Ramsperger–Marcus–Kassel (RRKM) theory. By combining experimental and computational data, we wish to show that intramolecular interactions and neighboring-group participation play an important role in the proton-driven degradation of the codons in the gas phase.

2.2 Experimental Section

2.2.1 Materials

DNA codon trinucleotides (95% pure) were purchased from Integrated DNA Technologies, Inc. (Coralville, IA), and their quality and sequences were checked by tandem mass spectrometry. ^{15}N-labeled trinucleotides of the dXXX type, d[$^{15}N_5$-A]AA, dA[$^{15}N_5$-A]A, dAA[$^{15}N_5$-A], d[$^{15}N_5$-G]GG, dG[$^{15}N_5$-G]G, d[$^{15}N_3$-C]CC, dC[$^{15}N_3$-C]C, d[$^{15}N_2$-T]TT, dT[$^{15}N_2$-T]T, as well as trinucleotides of the dXYX type, d[$^{15}N_5$-A]CA, dAC[$^{15}N_5$-A], d[$^{15}N_5$-A]GA, d[$^{15}N_5$-A]TA, d[$^{15}N_3$-C]AC, d[$^{15}N_3$-C]GC, d[$^{15}N_3$-C]TC, d[$^{15}N_5$-G]AG, d[$^{15}N_5$-G]CG, d[$^{15}N_5$-G]TG, d[$^{15}N_2$-T]AT, d[$^{15}N_2$-T]CT, d[$^{15}N_2$-T]GT, and d[$^{15}N_2$-T]TG, were manually synthesized on a Controlled Pore Glass (CPG) synthesis column (500 Å) loaded with a desired nucleoside (200 nmol) according to standard procedures.[21,22] ^{15}N-labeled phosphoramidites were purchased from Cambridge Isotope Laboratories (Andover, MA); unlabeled phosphoramidites and synthesis columns were obtained from Bioautomation Corporation (Irving, TX); trichloroacetic acid solution (TCA) was purchased from Glen Research Corporation (Sterling, VA). The other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Acetonitrile (HPLC gradient grade, >99.9%, 10 ppm of H_2O) was dried for several days over molecular sieves (4 Å, 500 mL) and filtered (0.22 μm PVDF syringe filter) immediately before use. Detailed synthetic, deprotection, and isolation procedures for the ^{15}N-labeled trinucleotides were described in a separate publication.[23]

2.2.2 Methods

Mass spectra were measured on a Bruker amaZon 3D-ion trap tandem mass spectrometer (Bruker Daltonik, Gmbh, Bremen, Germany). Samples of unlabeled trinucleotides were dissolved in 50:50:1 acetonitrile–water–acetic acid to achieve 10–20 μM concentrations and ionized by electrospray to form singly and doubly protonated ions. Samples of ^{15}N-labeled trinucleotides were dissolved in 1.0 mL of the solvent mixture and electrosprayed. The charge state of interest was selected by mass and subjected to collision induced dissociation (CID) at excitation amplitudes that were varied in 5–6 steps within 0.40–0.60 V and 0.15–0.30 V and for mono and dications,

respectively. Spectra obtained at excitation amplitudes causing the precursor ion relative intensity to decrease to 50 ± 5% were used for comparisons among the different trinucleotides. CID-MS^3 spectra of selected ions were obtained analogously by mass-selecting and dissociating the fragment ions from CID-MS^2.

2.2.3 Calculations

Conformation analysis of monocations and dications was performed using Born–Oppenheimer molecular dynamics (BOMD) calculations as described previously.[24] Briefly, initial structures were generated for multiple protomers of adenine (N1, N3, N7), guanine (N7), cytosine (O2, N3), and thymine (O2 and O4) that were placed in the 5′-, middle, and 3′-nucleobase positions of the trinucleotide ions. BOMD trajectories for each protomer were run for 20 ps in 1 fs steps with Berendsen thermostat[25] set at 510 K. The hybrid semiempirical PM6-D3H4 method[26] was used that includes corrections for hydrogen bonds and dispersion interactions. PM6-D3H4 was run with MOPAC[27] under the high-level Cuby4 platform.[28] Since BOMD includes both nuclear and valence-electron motion, the initial proton positions are not fixed and protons can move among accessible positions in the same or another nucleobase, phosphate oxygen, or deoxyribose hydroxyl groups. Two hundred snapshots were extracted in regular intervals from each 20000-step trajectory and their geometries were gradient optimized with PM6-D3H4. The resulting structures were sorted out to remove duplicates, and 15–20 structures were fully gradient optimized with B3LYP[29] and the 6-31G(d,p) basis set including harmonic frequency analysis. Several low-energy B3LYP structures were further reoptimized with M06-2X[30] and the 6-31+G(d,p) basis set, and the reoptimized geometries were used for single-point energy M06-2X calculations with the 6-311++G(2d,p) basis set. This level of theory has been shown to give very similar relative nucleoside ion energies compared to those obtained by coupled-clusters calculations with single, double, and disconnected triple excitations (CCSD(T)) that were expanded to the complete basis set.[31] In addition, our previous studies indicated that dinucleotide[18,32] and tetranucleotide ion geometries[33,34] that were obtained by M06-2X/6-31+G(d,p) were consistent with ion structures determined by action spectroscopy measurements. Solvation energies were addressed with self-consistent reaction field calculations using the polarizable continuum model (PCM)[35] in water dielectric. Reference nucleoside proton affinities and gas-phase basicities were calculated using

the DFT and CCSD(T) data reported previously[31] and are summarized in **Table 2.1**. Rice–Ramsperger–Kassel–Marcus calculations of unimolecular rate constants were performed with the QCPE 644 program of Zhu and Hase[36] that has been modified by expanding the limit for the number of internal degrees of freedom to 1000 and recompiled to run under Windows 7.[37] Vibrational state densities were obtained by a direct count of quantum states, and the rotational states were treated adiabatically. The microscopic rate constants [k(E,J,K)] were Boltzmann-averaged over the thermal distribution of rotational J and K states at 298 and 473 K.

Table 2.1 DNA Nucleoside Proton Affinities and Gas-Phase Basicities

	Proton Affinity and Gas-Phase Basicity[a,b]									
	M06-2X[c]		M06-2X[d]		CCSD(T)[e]		CCSD(T)[f]		CCSD(T)[g]	
Nucleoside	PA	GB	PA	GB	PA	GB	PA	GB	PA	GB
2'-Deoxyadenosine										
N-1-H	956	927	956	927	964	935	964	935	963	934
N-3-H	986	954	984	953	993	962	993	961	992	961
N-7-H	947	919	947	919	954	925	954	925	953	924
2'-Deoxyguanosine										
N-7-H	1063	1033	1059	1030	1065	1036	1065	1036	1065	1036
2'-Deoxycytidine										
N-3-H	980	952	977	949	983	955	983	955	988	960
O-2-H	961		959							
2'-Deoxythymidine										
O-4-H	916	884	915	883	919	887	918	886	918	886

[a]In kJ mol^{-1}; [b]Referring to 310 K; [c]Calculations with the 6-31+G(d,p) basis set; [d]Calculations with the 6-311++G(2d,p) basis set; [e]From single point-energy calculations extrapolated to the aug-cc-pVTZ basis set; [f]From single point-energy calculations extrapolated to the aug-cc-pVQZ basis set; [g]From single point-energy calculations where the correlation energy was extrapolated to the complete basis set.

2.3 Results and Discussion

2.3.1 Survey of Monocation Spectra

Electrospray ionization produced mono- and diprotonated ions from most trinucleotides. Exceptions were the high-thymine combinations that either did not form dications (dTTT) or whose double protonation was very inefficient, as in dTTA, dTTC, dTAT, dTGT, and dTCT. The general pattern of CID-MS^2 monocation dissociations is illustrated with the spectra of isomeric trinucleotide ions $dACG^+$, $dCAG^+$, and $dAGC^+$ (**Figure 2.1 a–c**), in which the adenine, guanine, and cytosine positions were permutated, and the different nucleobase masses allowed us to unambiguously assign the fragment ions. The main dissociations consisted of losses of the nucleobases that occurred unevenly from the 5′, middle, and 3′ positions, and also depended on the nucleobase.

Loss of the 3′-nucleobase was accompanied by elimination of water which was highly specific for this position for all nucleobases. Backbone dissociations were chiefly represented by cleavages between the 5′ and middle nucleotides, yielding $\boldsymbol{w_2}^+$ ions with a minor formation of $\boldsymbol{w_1}^+$ fragment ions. For oligonucleotide sequence ions nomenclature, see refs[38-40]. The $\boldsymbol{w_2}^+$ ions further lost the 3′-nucleobase, again followed by loss of water. These general features of monocation dissociations were very similar to those reported by O'Hair and co-workers.[20] Evaluation of the fragment ion relative intensities for loss of nucleobase and backbone cleavage from codon monocations is summarized in **Table 2.2**. The relative intensities of nucleobase loss from different positions were normalized to the total ion intensity for nucleobase loss. The percentages for the loss of the 3′ nucleobases were expressed as sum of the $(MH - 3'\text{-base})^+$ and $(MH - 3'\text{-base} - H_2O)^+$ relative intensities. The latter term is given separately in parentheses. The $\boldsymbol{w_2}^+$ ion relative intensities were expressed relative to the sum of the loss-of-base and $\boldsymbol{w_2}^+$ fragment ion intensities. Minor fragment ion intensities were not included in the normalization. To ascertain compatibility of the relative intensities reported in **Table 2.2**, they were measured at excitation amplitudes resulting in 50 ± 5% relative intensity of the precursor ion. The relative intensities for loss of nucleobase for individual bases and their positions in the codons widely varied and depended on the trinucleotide sequence and presence of other bases. On average, loss of 5′-nucleobases was most common, amounting to 49%, followed by loss of 3′-bases at 33%. The middle bases were lost less frequently

at 18%. Cytosine was on average the most vulnerable base to be eliminated (62%), followed by guanine (27%), and adenine (11%). In contrast, loss of thymine was very rare, occurring at <1%. The formation of the w_2^+ ions represented on average 34% of the fragment ions, again showing a 0.4–98% variability among the codon monocations. A plot of the w_2^+ ion intensities versus those for the 5′-base loss was scattered, showing a weak correlation coefficient at $r^2 = 0.38$ (**Figure 2.2**). There were some conspicuous outliers, such as dTTA, dTTT, and dATT, that gave high w_2^+ ion relative intensities in spite of showing a weak loss of the 5′ base. This indicated that the loss of the 5′-nucleobase may not trigger spontaneous elimination of the 5′-deoxyribose moiety to form the w_2^+ ion. This appeared to be in contrast to the mechanism suggested by Rodgers et al. in their study of dinucleotide cation dissociations.[41]

Table 2.2 Fragment Ion Relative Intensities from Monocations

	relative fragment ion intensity[a]					relative fragment ion intensity[a]			
codon[b]	5′	middle	3′ + water[c]	w_2[d]	codon[b]	5′	middle	3′ + water[c]	w_2[d]
AAA	50	13	37 (17)	54	**C**T**C**	48		52 (31)	37
GGG	93	4	3 (0.6)	28	T*CC*			100 (34)	<1
CCC	18	21	61 (23)	11	*CC*G	29	33	38 (21)	24
TTT	27	16	57 (35)	87	**C**G**C**	73	7	20 (5)	37
TTG	4	1	95 (25)	33	G*CC*	68		32 (7)	28
TG**T**	4	94	2	3	*CC*A	44	48	8 (3)	12
G*TT*[e]	97		3 (1.2)	95	**C**A**C**	27	4	69 (30)	29
TTA	12	7	81	75	A*CC*	9		91 (19)	3
TA**T**	4	84	12 (8)	36	AGC	18	8	74 (23)	17
A*TT*	59		41 (20)	98	ACG	20	32	48 (22)	23
*TT*C	1		99 (27)	6	GCA	73	21	6 (2)	41
TC**T**	1	97	2	4	GAC	74		26 (5)	38
C*TT*	97		3 (1)	94	CAG	55	4	41 (13)	44
*AA*T	99		1	73	CGA	92	6	3 (1)	46
AT**A**	61		39 (14)	32	AGT	67	32	1	42
T*AA*			100 (60)	5	ATG	24		76 (38)	52
*AA*G	45		55 (24)	58	GTA	93		7 (4)	58
AG**A**	55	22	23 (8)	46	GAT	100			38
G*AA*	94		6 (2)	59	CGT	96	4		32
*AA*C	14		86 (27)	14	CTG	64		36 (22)	61

relative fragment ion intensity[a]					relative fragment ion intensity[a]				
codon[b]	5′	middle	3′ + water[c]	w_2[d]	codon[b]	5′	middle	3′ + water[c]	w_2[d]
AC**A**	33	30	38 (19)	29	GTC	46		54 (16)	41
C*AA*	67		33 (13)	59	GCT	88		12	56
TGG			100 (9)	0.6	TCG		33	66 (51)	0.3
GTG	88		12 (4)	46	TGC		20	80 (34)	0.4
*GG*T	93	4		25	ACT	26	73	2	42
*GG*C	84	2	14 (2)	32	ATC	9		91 (29)	8
GCG	82	8	10	38	CAT	97		3	44
C*GG*	58		42 (4)	46	CTA	93		7 (5)	56
*GG*A	98		2	39	TCA		65	34 (27)	1
GAG	92	1	7 (1)	37	TAC		4	96 (22)	<0.1
A*GG*	18		82 (12)	26					
*CC*T	99		1	43					

[a]Percent of summed base-loss ion intensities.
[b]^{15}N-labeled nucleosides shown as bold characters;
[c]Combined relative intensities for loss of 3′-base and water. Relative intensities for loss of water in parentheses.
[d]Percent of combined base loss and $\boldsymbol{w_2}^+$ ion intensities.
[e]Nucleobase loss from positions shown in italics was not resolved, and relative intensities were summed for the unresolved positions.

To further address this point, we investigated the dependence of the fragment ion relative intensities on the ion trap excitation amplitude, noting a previous study of oligonucleotide anions that has reported strong effects on product ion spectra.[42] The spectra of trinucleotide monocations obtained within the 0.46–0.58 V range showed only a weak dependence of the relative intensities of the nucleobase loss fragment ions which changed only marginally with the excitation amplitude. Within this interval, the relative intensity of the precursor ions dropped from 95% to <1% for most trinucleotide monocations (**Figure 2.3a-f**). Consecutive dissociations, such as the loss of water following the loss of the 3′-nucleobase and formation of $\boldsymbol{w_2}^+$ ions were enhanced by <20% for most trinucleotide ions over the entire energy interval (**Figure 2.3a-d**). However, dissociations of dGTA$^+$ and dGGG$^+$ showed a larger increase from the very low to highest excitations (**Figure 2.3e, f**). This indicated that the trinucleotide ion composition affected the energetics and kinetics of the dissociations.

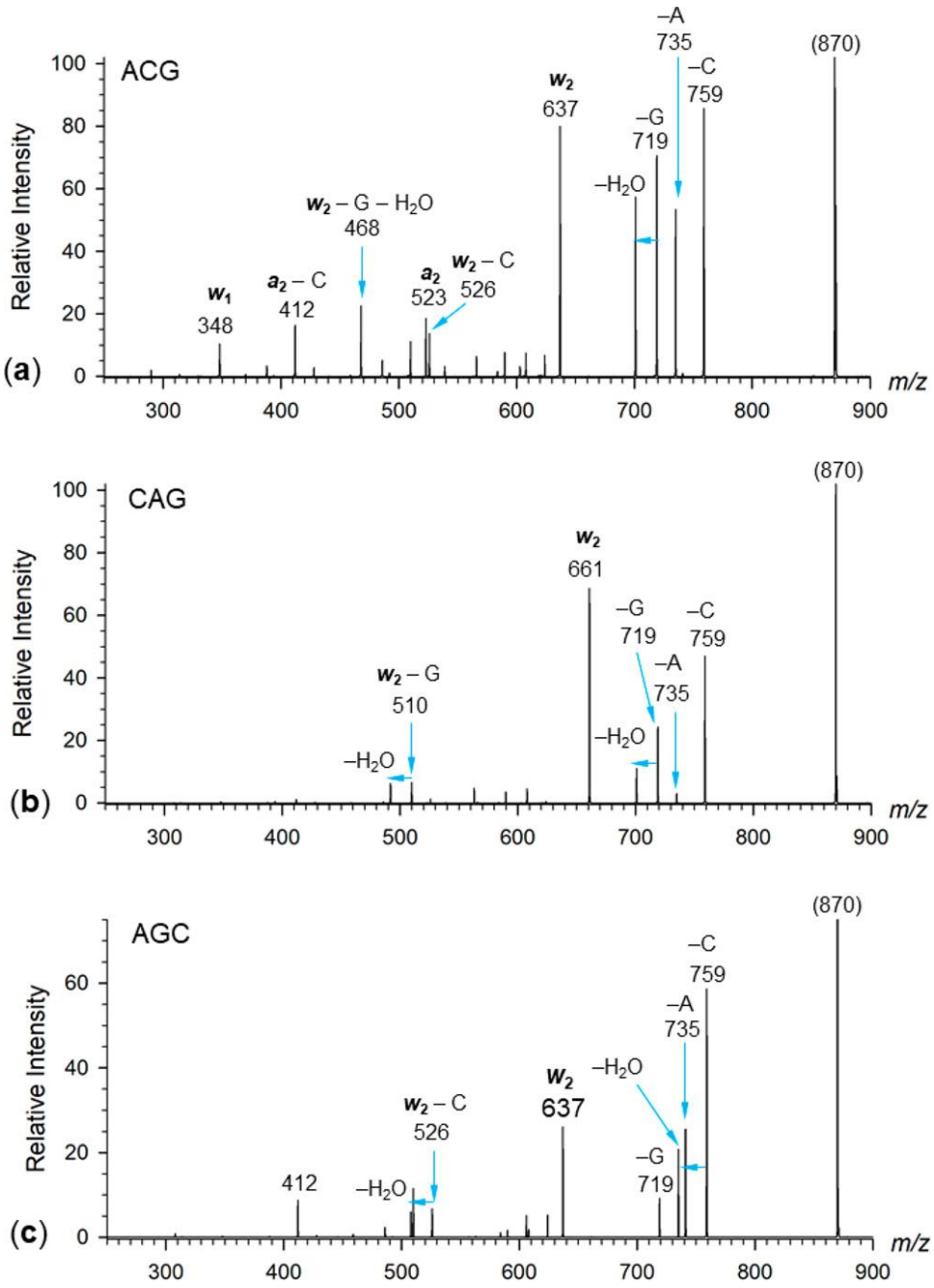


Figure 2.1: CID-MS^2 spectra of singly protonated (a) ACG, (b) CAG, and (c) AGC.

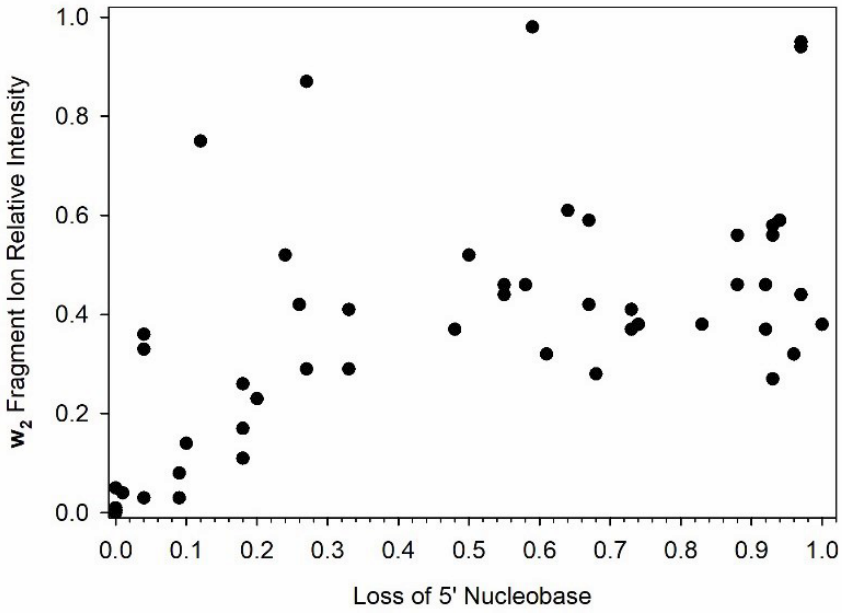


Figure 2.2: Plot of w_2^+ fragment ion relative intensities versus the relative intensity of the (MH - 5'-nucleobase)$^+$ ions from monocations.

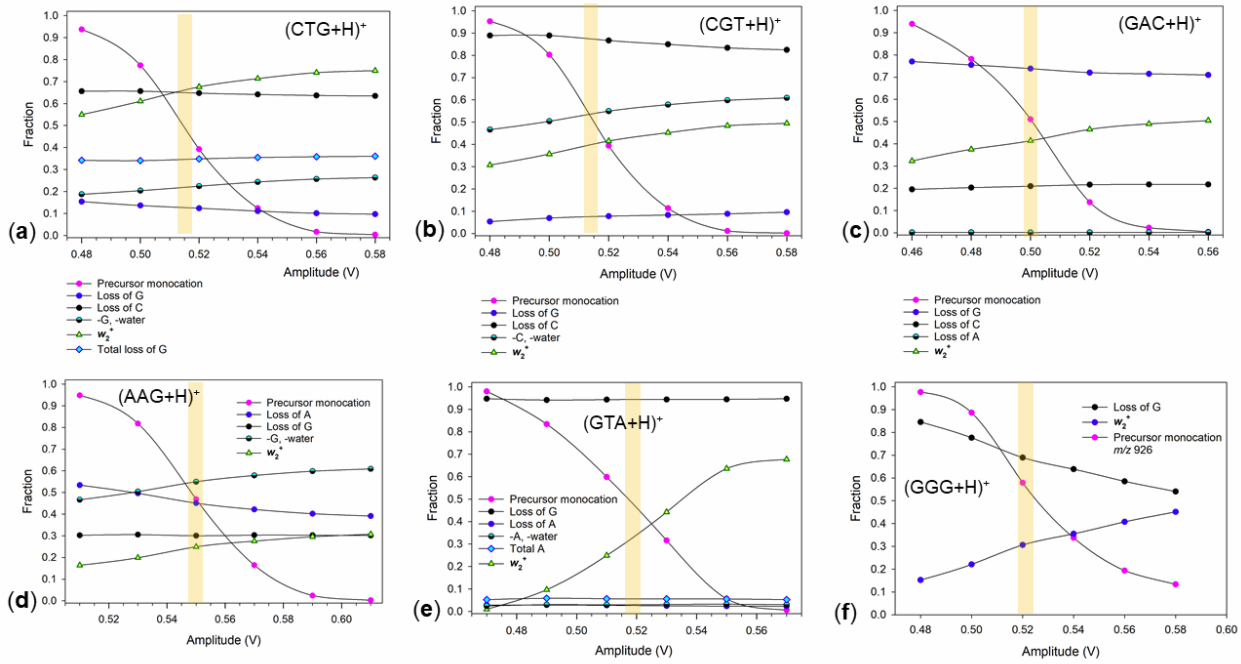


Figure 2.3: Relative intensities in the CID-MS^2 spectra of codon trinucleotide monocations plotted as a function of the ion trap excitation amplitude. (a) (CTG+H)$^+$, (b) (CGT+H)$^+$, (c) (GAC+H)$^+$, (d) (AAG+H)$^+$, (e) (GTA+H)$^+$, (f) (GGG+H)$^+$. The color bars show the intervals of amplitudes used to report the spectra in main text Table 2.2.

Further, we selected $(MH - base)^+$ fragment ions from CID-MS^2 spectra of selected codons for which loss of both the 5′- and 3′-nucleobase was observed, and we obtained their CID-MS^3 spectra. The spectra of the $(MH - 5'\text{-nucleobase})^+$ ions from $dGTC^+$ and $dCAG^+$ and $dCTG^+$ (**Figure 2.4a–c**) showed major $\boldsymbol{w_2}^+$ fragment ions by elimination of 2-hydroxymethylfuran ($C_5H_6O_2$) from the 5′-deoxyribose residue. Hence, when given sufficient internal excitation, the $(MH - 5'\text{-nucleobase})^+$ ions did dissociate to form the sequence fragment ions. The consecutive dissociations of the $\boldsymbol{w_2}^+$ ions depended on the remaining bases. For example, the $\boldsymbol{w_2}(TC)^+$ and $\boldsymbol{w_2}(AG)^+$ ions showed a facile loss of the 3′-nucleobase (C and G, respectively), followed by elimination of water from the 3′-sugar. In contrast, the $\boldsymbol{w_2}(TG)^+$ ion was substantially more resistant to follow-up dissociation (**Figure 2.4c**). These differences indicated that the dissociations may be assisted by interactions between the sugar moiety and the bases remaining in the fragment ions that may exhibit chemical specificity. The specific nature of these interactions is addressed later in **Sections 2.3.3–2.3.8**.

The CID-MS^3 spectra of the $(MH - 3'\text{-nucleobase})^+$ ions from $dGTC^+$ and $dCAG^+$ and $dCTG^+$ (**Figure 2.5a–c**) showed a major loss of water, which was consistent with the analogous fragmentations accompanying the loss of the 3′-nucleobase in the CID-MS^2 spectra of protonated trinucleotides. However, the competitive and consecutive dissociations of these $(MH - 3'\text{-nucleobase})^+$ ions differed depending on the remaining nucleobases. For example, with $dGTC^+$ and $dCTG^+$ these other dissociations resulted in the elimination of the 3′-sugar moiety, forming the corresponding $\boldsymbol{d_2}^+$ fragment ions at m/z 652 and 612, respectively. The $\boldsymbol{d_2}^+$ ions further dissociated by loss of the 5′-nucleobase which was accompanied by a facile loss of water, forming the dominant fragment ions at m/z 483 from both $dGTC^+$ and $dCTG^+$ (**Figure 2.5a, b**). In contrast, the $(MH - G)^+$ ion from $dCAG^+$ underwent competitive loss of water and cytosine (m/z 701 and 608, respectively) while no formation of the $\boldsymbol{d_2}^+$ ion, expected at m/z 621, was observed (**Figure 2.5c**).

The middle-position nucleobase also had an effect on the spontaneous loss of water from the $(MH - 3'\text{-nucleobase})^+$ ions. **Table 2.2** data indicate that loss of water was particularly facile from trinucleotide ions having a pyrimidine base in the middle position. For example, over 50% of $(MH - 3'\text{-nucleobase})^+$ ions from $dTTT^+$, $dCTC^+$, $dCCG^+$, $dGTA^+$, $dTCG^+$, $dCTA^+$, and $dTCA^+$ spontaneously dissociated by loss of water in CID-MS^2 of the monocations.

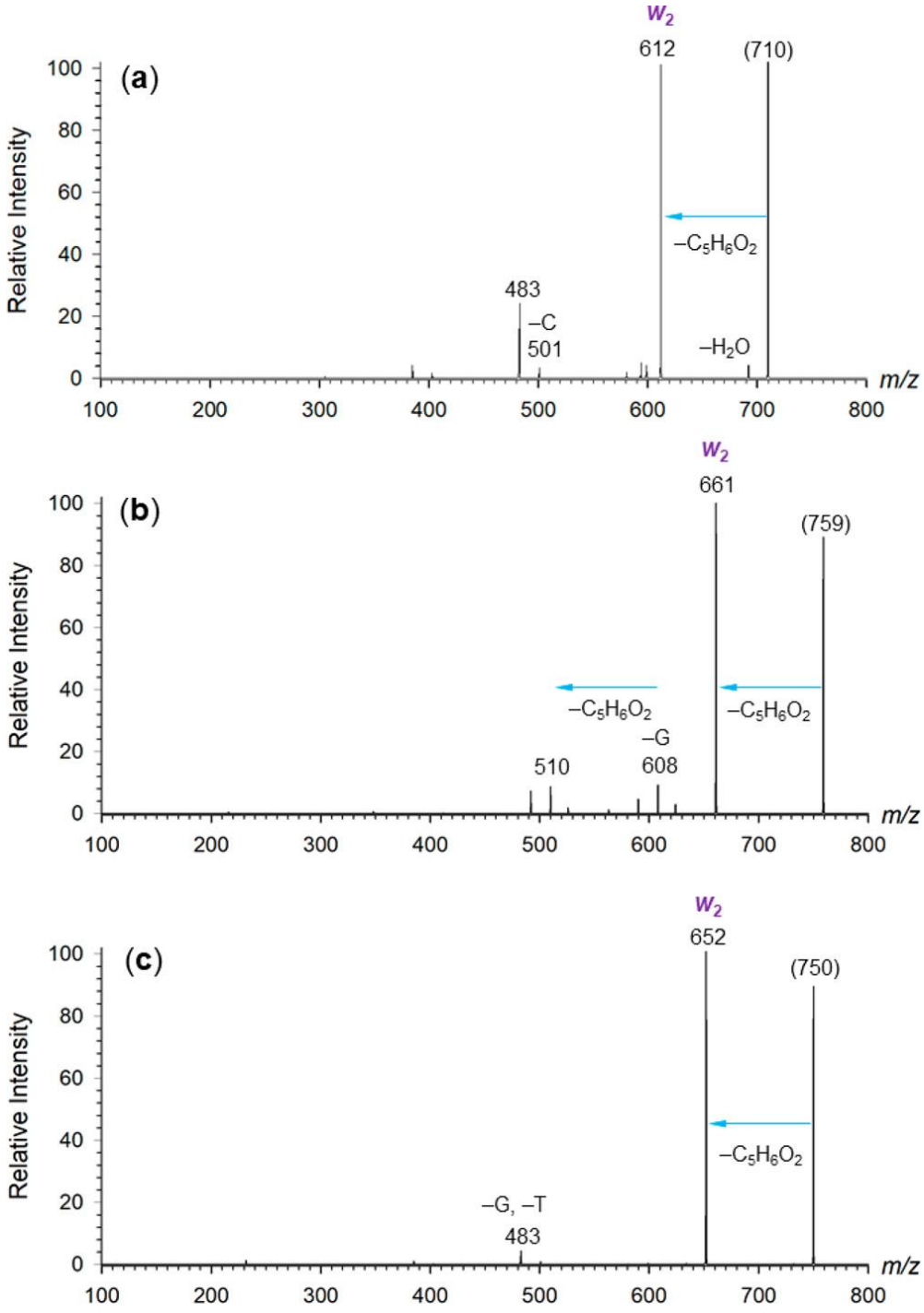


Figure 2.4: CID-MS^3 spectra of (MH – 5′-base)$^+$ ions from (a) dGTC, (b) dCAG, and (c) dCTG.

2.3.2 Resolution of Identity in Dissociations of Codon Monocations

The combined effects of the nucleobase nature and position in codons of the dXYZ type can lead to competitive enhancement or suppression of dissociations originating at any given position.

This makes it difficult to assign specific reactivity to the nucleobase and position. To address this issue, we analyzed the dissociations of trinucleotide ions of the dXXX and dXYX type where the structurally identical nucleobases in different positions were distinguished by ^{15}N labeling. Starting with $dAAA^+$, loss of adenine occurred from the 5′-, middle, and 3′-positions in a 50:13:37 ratio, where the fraction for the loss of 3′-adenine included the ion intensity from subsequent specific loss of water. In contrast, $dGGG^+$ showed a dominant loss of 5′-guanine at 93% whereas the other positions were much less reactive. It is noteworthy that despite the highly abundant loss of 5′-G, the formation of $\boldsymbol{w_2}^+$ fragment ions from $dGGG^+$ was less abundant than from $dAAA^+$. Loss of cytosine from $dCCC^+$ proceeded from the 3′-nucleotide, with the 5′-, middle, and 3′-positions giving an 18:21:61 distribution. Similarly, loss of thymine from $dTTT^+$, albeit overall not very abundant, showed a 27:16:57 distribution for the 5′-, middle, and 3′-positions. The overall low propensity for loss of thymine was underscored by the spectra of $dTGT^+$, $dTAT^+$, and $dTCT^+$, which showed major loss of the middle nucleobases (**Table 2.2**). Insertion of thymine in the middle position, as in $dATA^+$, $dGTG^+$, and $dCTC^+$, did affect the distribution for the loss of the 5′- and 3′-nucleobases. With $dATA^+$, thymine slightly enhanced the loss of 5′-adenine (61%) compared to the same fraction in $dAAA^+$ (50%). In $dGTG^+$, it slightly decreased the loss of 5′-guanine (88%) compared to the same fraction in $dGGG^+$ (93%). In $dCTC^+$, it increased the loss of 5′-cytosine (48%) more visibly when compared to the same fraction in $dCCC^+$ (18%). The effect on the loss of nucleobases of a middle cytosine was chiefly competitive, as shown for $dACA^+$ (33:30:38) and $dGCG^+$ (82:8:10). Combinations of $dCXC^+$ with adenine and guanine in the middle position affected the loss of the 5′- and 3′-cytosine to a different extent. In $dCAC^+$, the nucleobase-loss distribution (27:4:69) resembled that in $dCCC^+$ (16:21:61), showing an enhanced loss of cytosine from both the 5′- and 3′-positions. In contrast, in $dCGC^+$, the presence of the middle guanine resulted in a reversed preference for loss of the 5′-cytosine in a 73:7:20 distribution. The adenine-guanine combination in $dAGA^+$ resulted in an enhanced loss of guanine which was chiefly on account of a diminished loss of 3′-adenine. Finally, $dGAG^+$ showed a nucleobase loss distribution (92:1:7) that was very similar to that in $dGGG^+$ (93:4:3).

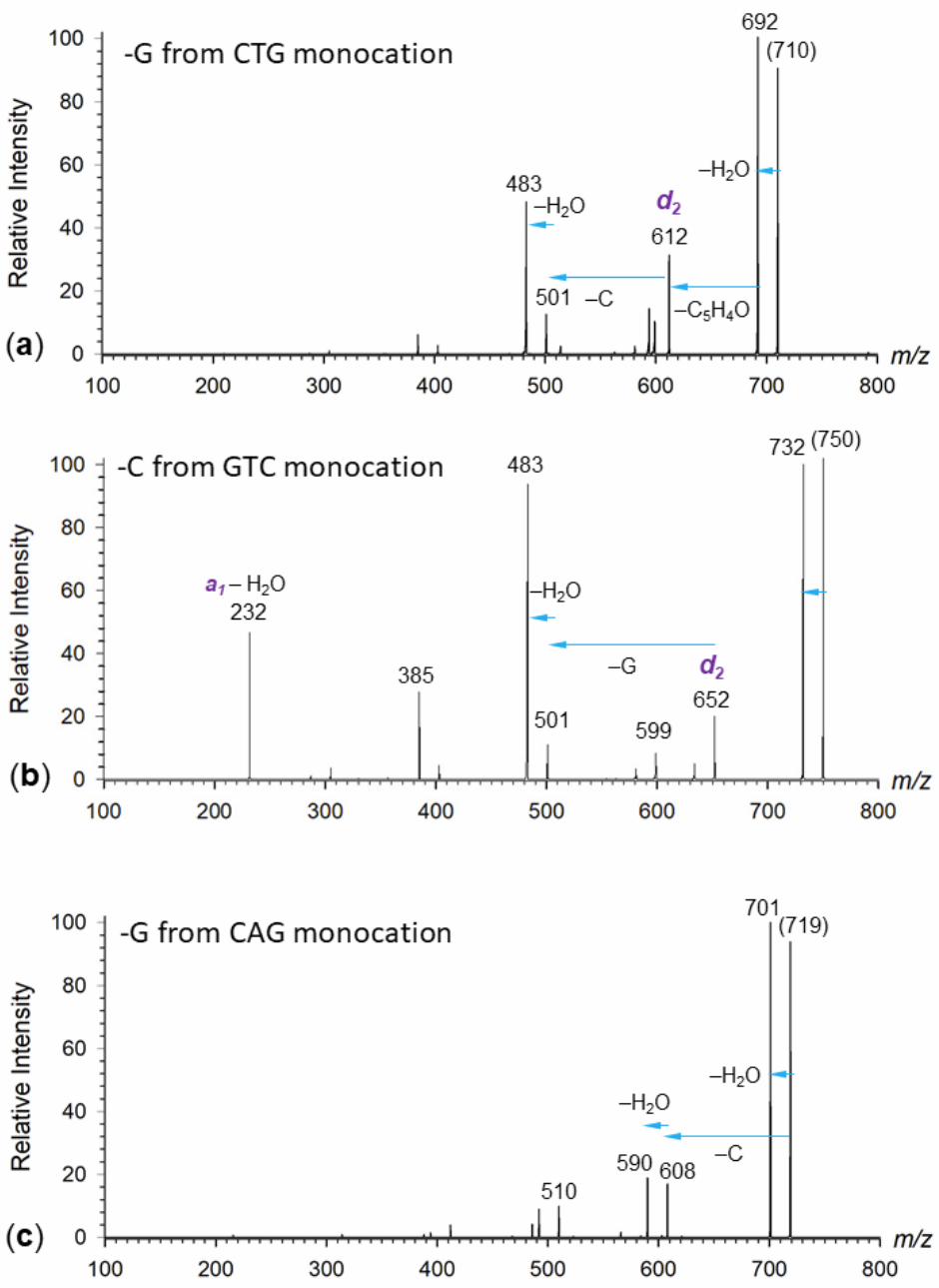


Figure 2.5: CID-MS^3 spectra of (MH – 3'nucleobase)$^+$ ions from (a) dCTG, (b) dGTC, and (c) dCAG.

2.3.3 Codon Monocation Structures and Dissociation Mechanisms

The effects of the nucleobase nature and position that were revealed by resolution of identity in dissociations of trinucleotide ions of the dXXX$^+$ and dXYX$^+$ type pointed to specific interactions

between the nucleobases in the codon monocations which could be of relevance to the reaction mechanisms of acid-induced DNA degradation. To address this point in structural detail at the single-atom resolution level, we obtained fully optimized geometries of multiple protomers and conformers of $dAAA^+$, $dGGG^+$, $dCCC^+$, $dTTT^+$, $dACA^+$, and $dATC^+$. Selected low-Gibbs energy isomers are presented and discussed here.

2.3.4 dAAA⁺ Monocations

BOMD conformational analysis followed by DFT geometry optimizations yielded structures **A1**–**A3** as the lowest Gibbs energy isomers (**Figure 2.6**). The structures differed in both their protonation pattern and conformation but were closely spaced by energy. Structure **A1** was protonated at N1 of 3′-adenine and displayed strong hydrogen bonds involving the phosphoester and sugar hydroxyl groups. Structures **A2** and **A3** were zwitterions in which a proton from the phosphate hydroxyl was transferred to a basic adenine position. This included N1 in both the 5′- and 3′-adenines in **A2**, N3 in the middle adenine, and N7 in the 3′-adenine in **A3**. The phosphate anions in the zwitterions were massively internally solvated by the sugar hydroxyl and nucleobase protons to create elaborate networks of hydrogen bonds. Interestingly, N3-protonated tautomers, which are favored in adenosine ions,[18,43] were not represented among the lowest-energy $dAAA^+$ structures. Nonspecific solvation by water, as estimated by polarizable continuum model calculations including full gradient geometry optimization did not appreciably change the relative Gibbs energies of **A1**–**A3** which remained within 7 kJ mol^{-1} (**Figure 2.6**). The very similar relative energies of **A1**–**A3** are arguably within the accuracy of the DFT calculations used here, indicating no preference for the classical or zwitterionic structures. In addition, the hydrogen-bond networks in both the classical and zwitterionic structures, along with the energy similarity among them, were prone to facilitate rapid proton migration within the ions in the course of dissociation. Accordingly, we found several $dAAA^+$ protomers with different positions of the charging proton that were within 40–80 kJ mol^{-1} above the global energy minimum and may be accessed via proton migration as reactive intermediates upon vibrational excitation.

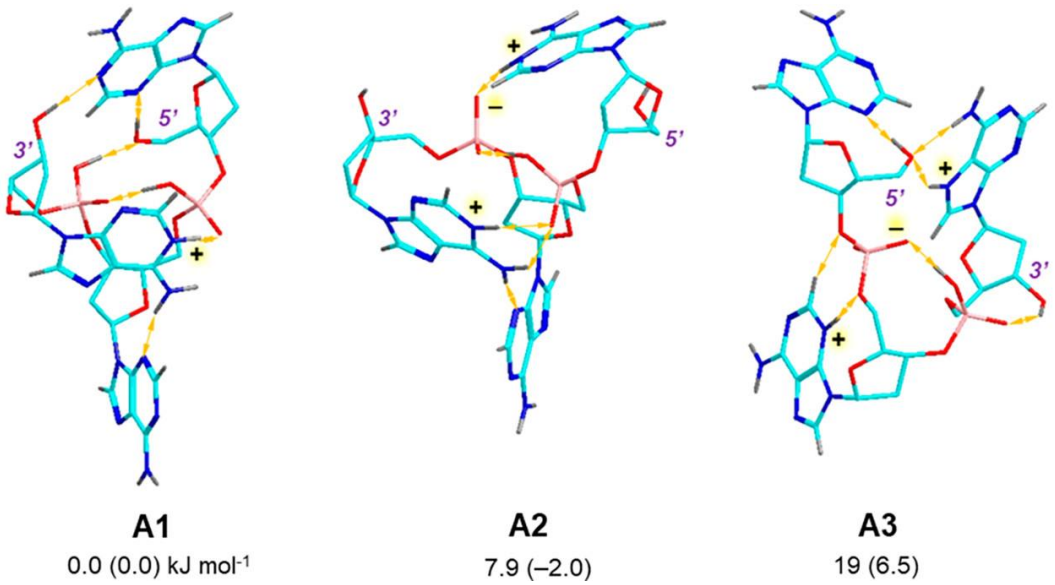
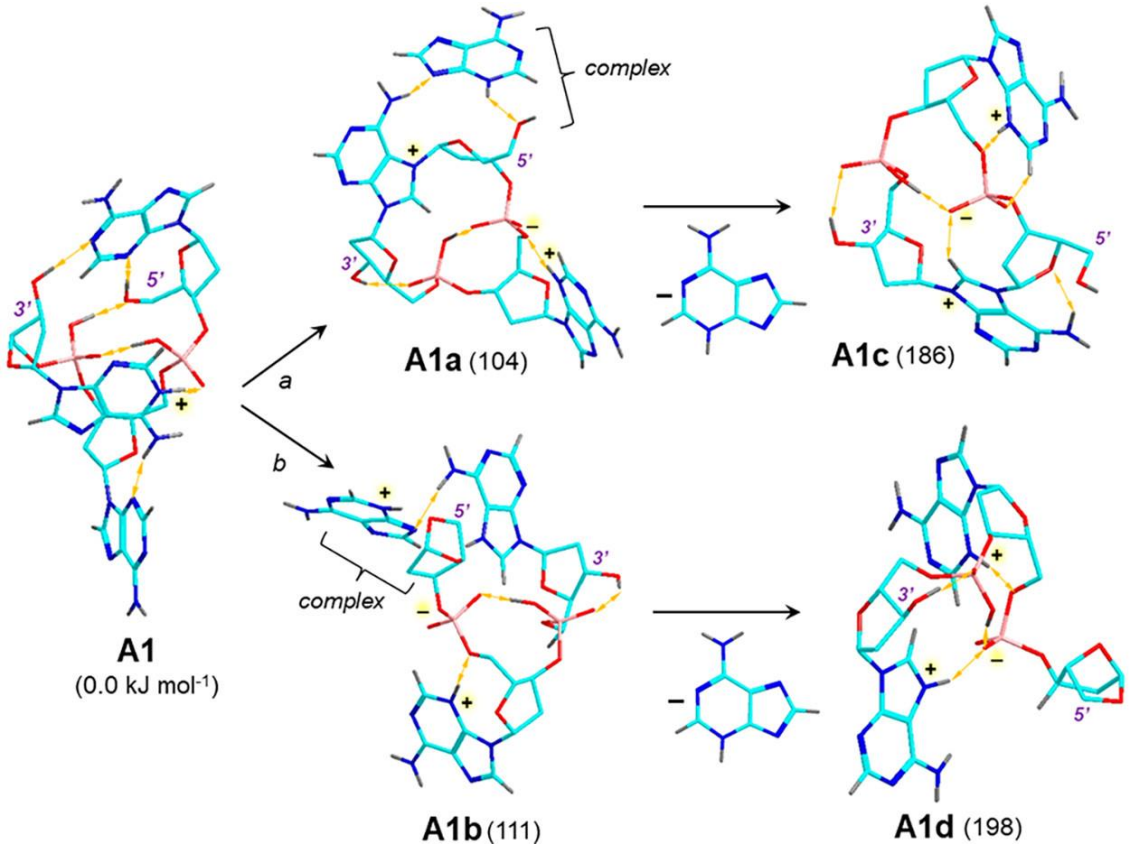


Figure 2.6: M06-2X/6-31+G(d,p)-optimized structures of low Gibbs energy $dAAA^+$ ions. Relative Gibbs energies are from single-point M06-2X/6-311++G(2d,p) calculations including B3LYP/6-31G(d,p) zero-point energies and 310 K enthalpies and entropies. Values in parentheses include PCM solvation energies in water dielectric. Atom color coding: cyan = C, blue = N, red = O, bronze = P, gray = H. Only the nucleobase and exchangeable hydrogen atoms are shown. Protonation and deprotonation sites are labeled with plus and minus signs. Hydrogen bonds are visualized by yellow arrows.

Proton migrations are shown to be involved in the loss of 5′-adenine (**Scheme 2.2**). Dissociation of the N9–C1′ glycosidic bond to 5′-adenine in **A1** was triggered by a relay phosphate → 5′-OH → N3 proton transfer that charged the adenine and increased its leaving group ability. It is worth noting that this dissociation was not predicated on 5′-adenine protonation in the local energy minimum structure of the reactant. The glycosidic bond cleavage was assisted by neighboring group participation by N7 of the 3′-adenine (path a, **Scheme 2.2**), or 5′-OH (path b), forming new covalent bonds at C1′ and leading to complexes **A1a** and **A1b** at 104 and 111 kJ mol^{-1}, respectively, relative to **A1**. Loss of neutral adenine as the less stable N3-H tautomer from these complexes forming isomeric products ions **A1c** and **A1d**, required 186 and 198 kJ mol^{-1}, respectively, although it is possible that the departing adenine molecule can undergo an exergonic isomerization to the canonical N9-H tautomer ($\Delta H_{isom} = -37$ kJ mol^{-1})[44] in the complex and lower the threshold energy. It is interesting to note that the calculated threshold energy for the formation of **A1c** was very similar to that measured for the loss of adenine from protonated 2′-deoxyadenosine (193 kJ mol^{-1}).[13]



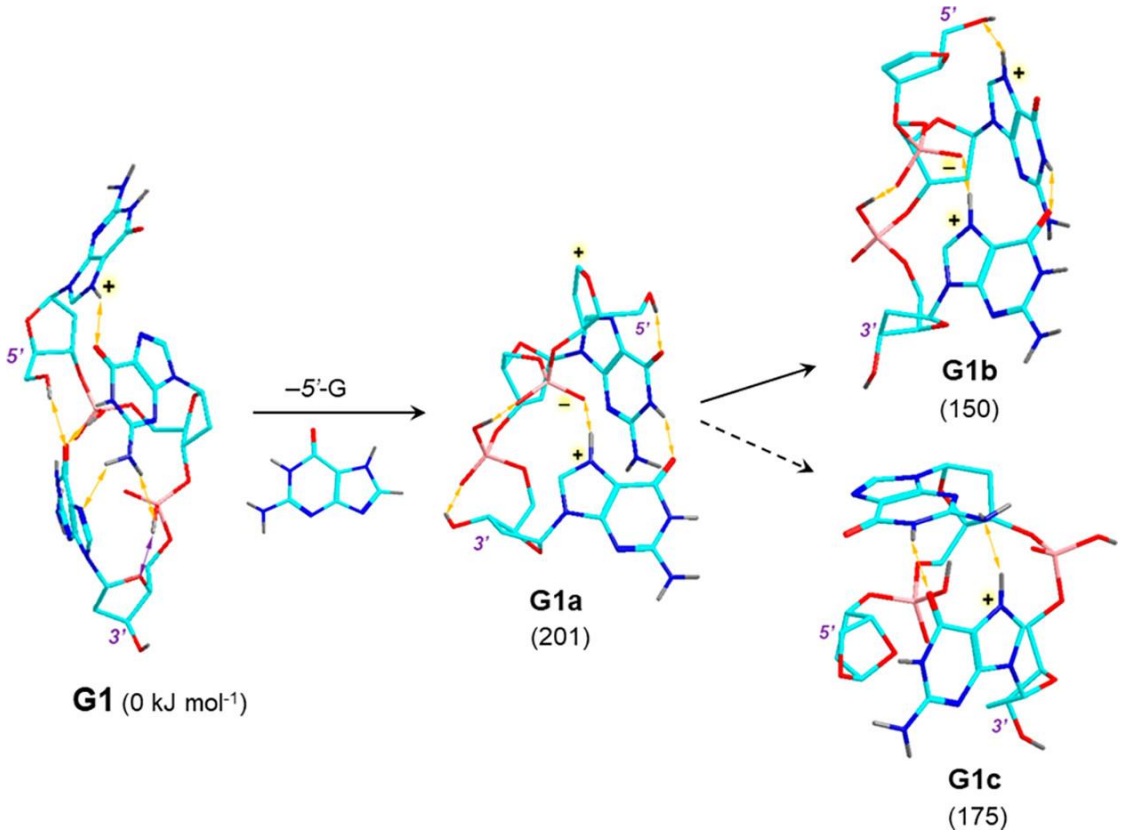
Scheme 2.2: Reaction Pathways for the Loss of 5′-Adenine from $dAAA^{+}$[a]

[a]Relative energies in parentheses in kJ mol^{-1} are from M06-2X/6-311++G(2d,p) single-point energy calculations and include zero-point energies, referring to 0 K.

2.3.5 dGGG⁺ Monocations

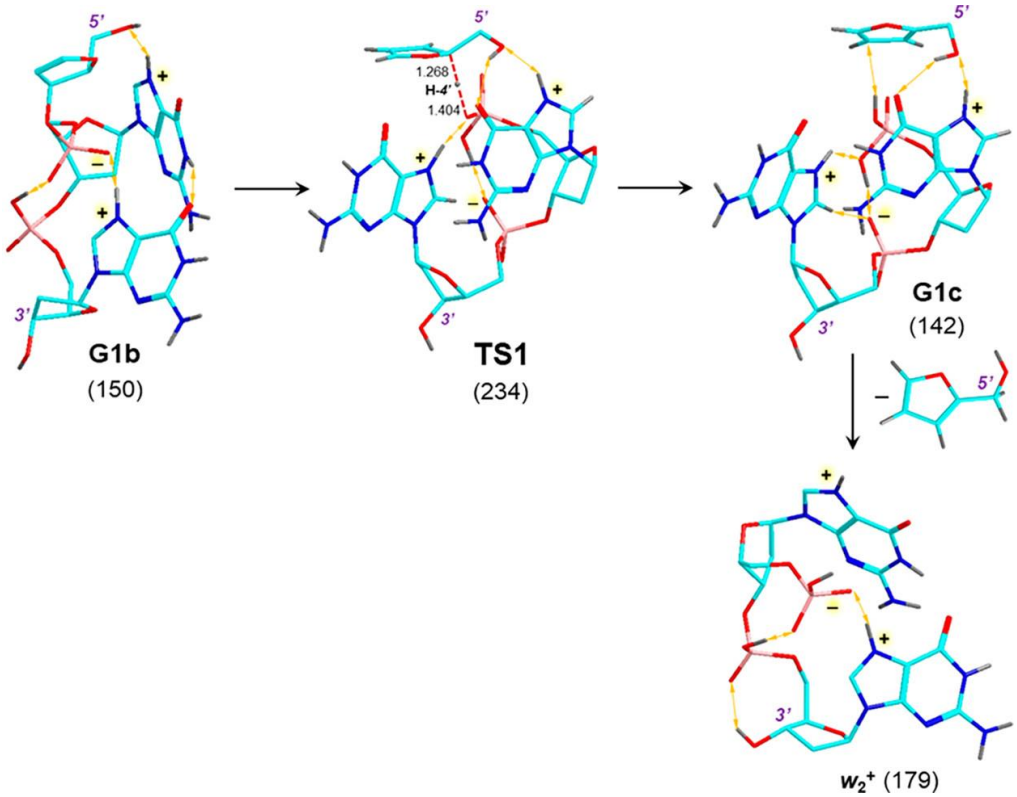
In contrast to $dAAA^{+}$, the $dGGG^{+}$ ions unambiguously favored classical structures protonated at N7 of 5′-guanine, as illustrated by the global energy minimum **G1** (**Scheme 2.3**). Other structures that were protonated at the middle and 3′-guanine were >30 kJ mol^{-1} less stable than **G1**. Structures that were initiated with a zwitterionic proton distribution among the guanine residues and phosphate groups collapsed to N7-protonated protomers upon BOMD and further DFT gradient optimization. This can be related to the exceptional gas-phase basicity of N7 in guanosine (GB = 1036 kJ mol^{-1}, **Table 2.1**) and its orientation with respect to the phosphate groups that disfavored proton transfer and zwitterion formation. Loss of 5′-guanine from **G1** can initially form a C1′ oxocarbenium cation (**G1a**, **Scheme 2.3**) which would be analogous to structures suggested for DNA depurination in water (**Scheme 2.1**). Interestingly, the loss of 5′-guanine was associated with internal proton migration from the proximate phosphate to 3′-guanine, forming the

intermediate ion **G1a** as a zwitterion. **G1a** can get stabilized by deprotonation of the 5′-oxocarbenium ion by the middle guanine residue, forming zwitterionic product **G1b** at 150 kJ mol^{-1} relative to **G1**. Another, albeit somewhat less favorable pathway, can include direct nucleophilic attack at C1′ by the adjacent 5′-OH, protonating the phosphate anion and closing a dioxabicyclo[2.2.1]heptane ring in **G1c** at 175 kJ mol^{-1} relative to **G1** (**Scheme 2.3**). The calculated dissociation energies in **G1** were similar to the threshold energy for loss of guanine from protonated 2′-deoxyguanosine (166 kJ mol^{-1}).[14]



Scheme 2.3: Pathways for Loss of 5′-Guanine from **G1**[a]

[a]Description as in Scheme 2.2.



Scheme 2.4: Formation of the w_2^+ Ion from (GGG-5′-G)$^{+a}$

[a]Description as in Scheme 2.2

The consecutive dissociation of the (MH – 5′-G)$^+$ ion **G1b** forming the w_2^+ fragment ion was investigated by finding the transition state for a proton transfer from C4′ of the 5′-sugar moiety (**TS1**, **Scheme 2.4**). The H4′ proton abstraction was facilitated by a relay proton transfer from the proximate phosphoester group to the other phosphate zwitteranion that increased the proximate phosphate basicity. Deprotonation of the 5′-deoxyribose ring followed the cleavage of the O–C3′ bond to reach **TS1** and proceeded by forming a noncovalent complex of the departing hydroxymethylfuran molecule with the w_2^+ ion (**G1c**, 142 kJ mol^{-1} relative to **G1**). The overall formation of w_2^+ from **G1** was calculated to be 179 kJ mol^{-1} endergonic, not including basis set superposition error. The calculated TS energy was then used for calculations of RRKM rate constants for crossing **TS1**, which, being the highest-energy point at 84 kJ mol^{-1} above **G1b**, was the rate-determining step. The rate constant (**Figure 2.7**, top panel), along with the calculated vibrational energy distribution in **G1b** (**Figure 2.7**, bottom panel) clearly indicated that **G1b** was kinetically stable at the ion trap temperature (310 K). To proceed on the ion trap time scale (50 ms), the dissociation of **G1b** to the w_2^+ ion required internal excitation of 250–300 kJ mol^{-1}, or an

equivalent temperature of 480 K (**Figure 2.7**). This was qualitatively consistent with both the low extent of $dGGG^+$ dissociation to $\boldsymbol{w_2}^+$ following 5′-guanine loss, and facile dissociation of $(MH - 5'\text{-}G)^+$ ions upon further collisional activation.

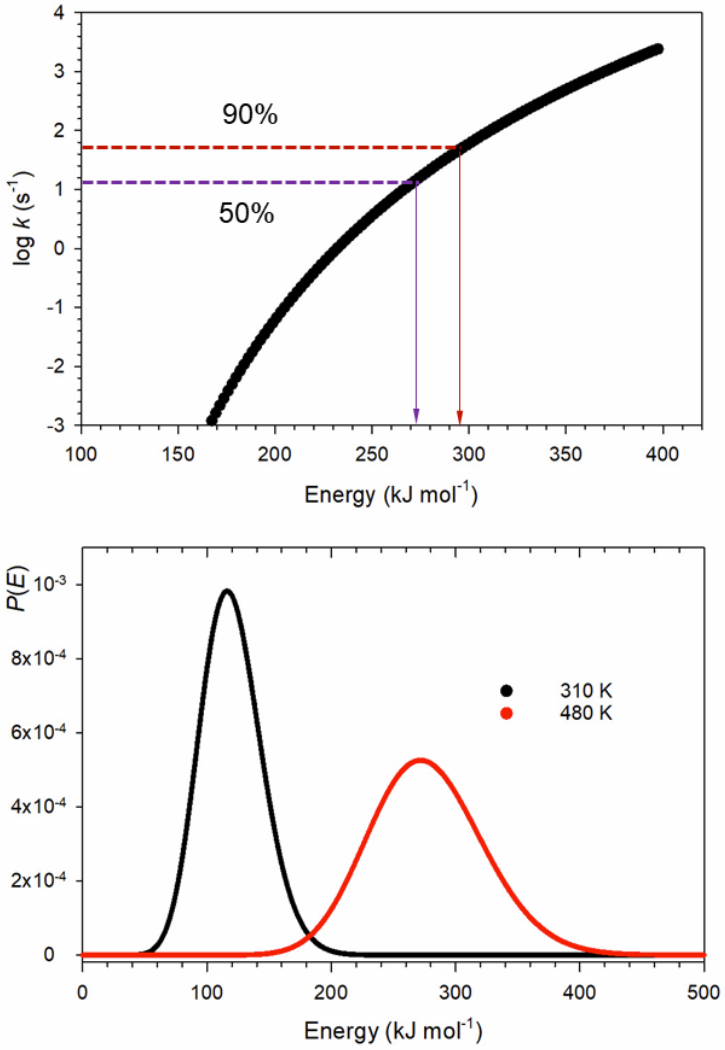


Figure 2.7: Top panel: RRKM rate constants for the dissociation of the **G1b** $(GGG\text{-}5'\text{-}G)^+$ ion via **TS1**. Bottom panel: Vibrational energy distribution in **G1b** at 310 and 480 K.

2.3.6 dCCC⁺ Monocations

Analysis of BOMD trajectories starting with all theoretically possible combinations of cytosine N3- and O2-protonated tautomers and combined with DFT geometry optimizations yielded several low-energy structures for the $dCCC^+$ monocations that are represented by **C1**–**C4** (**Figure 2.8**).

All these low-energy $dCCC^+$ ions favored protonation at cytosine positions N3 over O2. However, in contrast to both $dAAA^+$ and $dGGG^+$, the cytidine trinucleotides did not favor protonation at 5′-cytosine. Two low-energy structures, **C1** and **C2**, were zwitterions showing protonation at the middle and 3′-cytosines, in which the latter protonated base was stabilized by strong hydrogen bonds to the phosphate anion (**Figure 2.8**). The other low-energy $dCCC^+$ ions were classical structures with N3 protonation at the 5′ and middle cytosines in **C3** and **C4**, respectively. Structure **C3** showed a network of hydrogen bonds allowing a facile shuttle of a proton from 5′-cytosine to the 3′-base which may play a role in dissociations resulting in loss of cytosine. Loss of 3′-cytosine from **C1** was assisted by phosphate participation at **C1′**, forming the cyclic phosphate triester **C1a** (**Scheme 2.5**). The calculated threshold energy for the loss of 3′-cytosine depended on the neutral cytosine tautomer. A direct loss of an N3-H tautomer, consistent with the proton distribution in **C1**, led to a dissociation energy threshold at 192 kJ mol^{-1}. A prototropic isomerization of the cytosine N3 to the more stable O2-H tautomer in an ion–molecule complex would lower the threshold energy to 161 kJ mol^{-1}. Overall, the preference for 3′-protonated structures was consistent with the favored loss of the 3′-base upon CID. We note that our calculated threshold energies for the loss of 3′-cytosine from **C1** were higher than the value reported for the loss of cytosine from 2′-deoxycytidine (146 kJ mol^{-1}).[15]

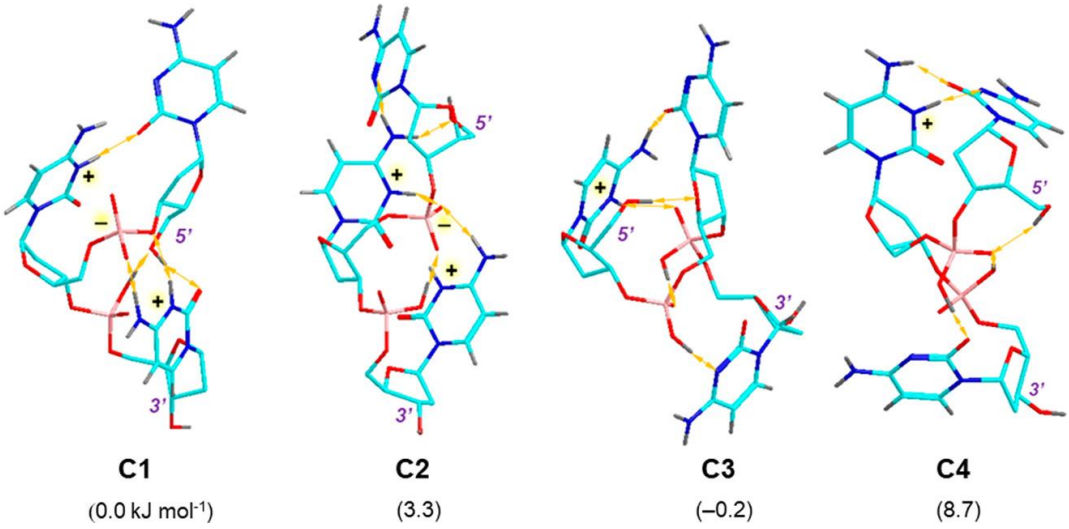
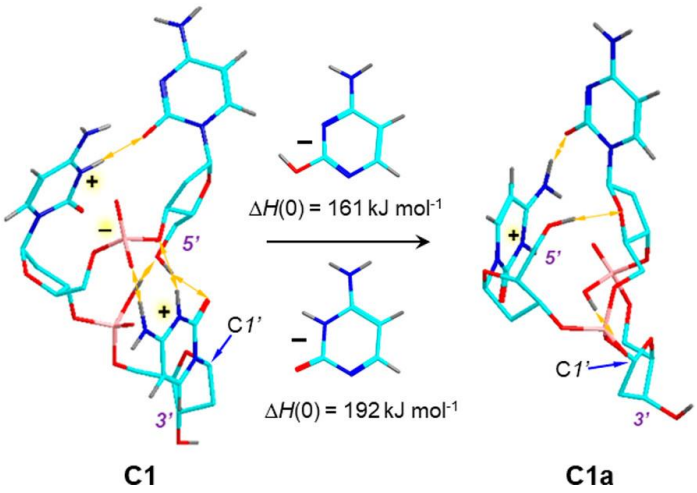


Figure 2.8: Low Gibbs energy $dCCC^+$ ions. Description as in Figure 2.6.

2.3.7 dTTT⁺ Monocations

Initial $dTTT^+$ structures in which combinations of all possible O2 and O4 protonated positions were considered collapsed upon BOMD and geometry optimization to a single lowest energy protomer **T1** (**Figure 2.9**). In stark contrast to $dAAA^+$, $dGGG^+$, and $dCCC^+$, structure **T1** was protonated at the 3′-phosphoester group. Thymine protonated structures, as represented by **T3**, were >40 kJ mol^{-1} less stable than **T1**. Comparing these energy differences with the basicity of the components, one finds the gas-phase basicity of triethyl phosphate (879 kJ mol^{-1})[45] to be slightly lower that the GB of 2′-deoxythymidine (886 kJ mol^{-1}, **Table 2.1**). However, this small difference can be readily overturned by more favorable phosphate hydrogen bonding in $dTTT^+$ ions. Phosphate protonation in $dTTT^+$ was consistent with a previous study of 2′-deoxythymidine that also preferred phosphate protonated structures.[46] The salient feature of **T1** was the elaborate hydrogen-bond network involving the bases, phosphates, 3′-O, and 5′-O that can allow a proton to migrate among the 3′-, middle, and 5′-thymines and induce the base loss. The CID-MS^2 result, which showed a favored loss of the 3′-thymine, may be due to a favorable dissociation thermodynamics. At the same time, loss of thymine from $dTTT^+$ at 13% combined relative intensity was less efficient in competition with the backbone dissociation forming w_2^+ ions (**Table 2.2**).



Scheme 2.5: Loss of 3′-Cytosine from $dCCC^{+a}$

[a]Description as in Scheme 2.2

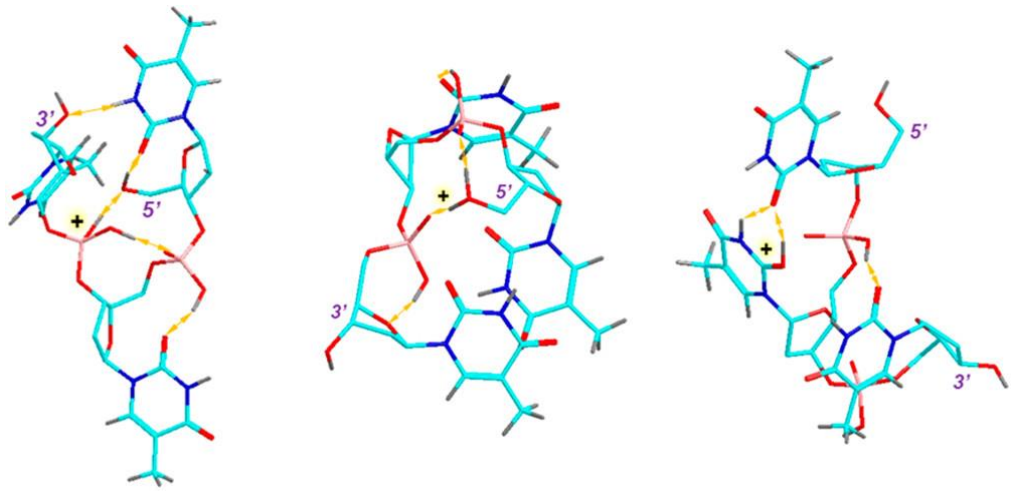


Figure 2.9: Low Gibbs energy dTTT$^+$ ions. Description as in Figure 2.6.

2.3.8 dACA$^+$ and dATC$^+$ Monocations

The prevalence of low energy zwitterionic structures for dAAA$^+$ and dCCC$^+$ made it of interest to examine the proton distribution in dACA$^+$ and dATC$^+$ to establish the most stable protomers for these codon sequences and relate the structures to the dissociations by loss of adenine, cytosine, and backbone cleavage. dACA$^+$ ions favored 5′-N3-protonated adenines **ACA1** and **ACA2** as the lowest energy structures, whereas a doubly adenine- protonated zwitterion **ACA3** was 11 kJ mol^{-1} higher by energy than **ACA1** (**Figure 2.10**). Loss of the 3′- and 5′-adenine was found to occur at similar rates, 33% and 38%, respectively (**Table 2.2**). The calculated threshold energies for the adenine loss from **ACA1** indicated different reaction pathways. Loss of the 3′-base from **ACA1** can be assisted by participation of the proximate phosphate, forming a cyclic phosphoester **ACA1c** at 210 kJ mol^{-1} threshold energy (**Scheme 2.6**). In contrast, loss of 5′-adenine proceeded via a C1′ oxacarbenium ion (**ACA1a**) at 265 kJ mol^{-1}, and the product was stabilized in the final product **ACA1b** (at 210 kJ mol^{-1}) by neighboring group participation by 5′-cytosine O2. Interestingly, this stabilization was in part due to the formation of an internally hydrogen bonded zwitterionic structure in **ACA1b** (**Scheme 2.6**). Thus, multiple reaction pathways for loss of adenine can occur competitively in dACA$^+$.

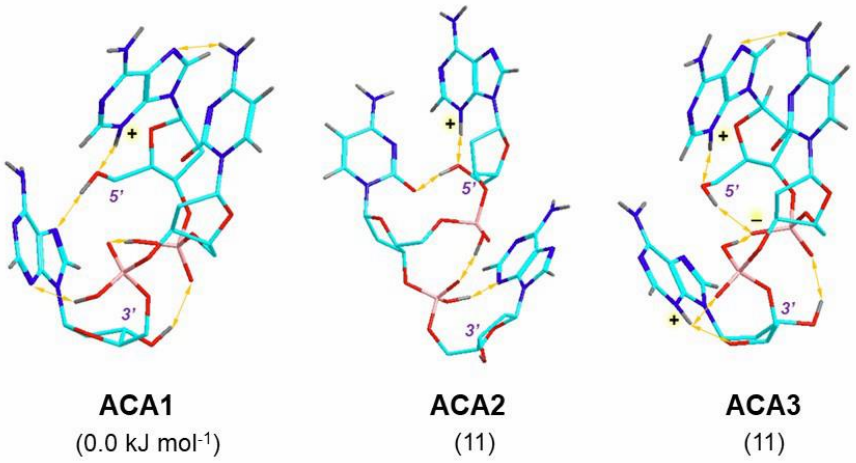
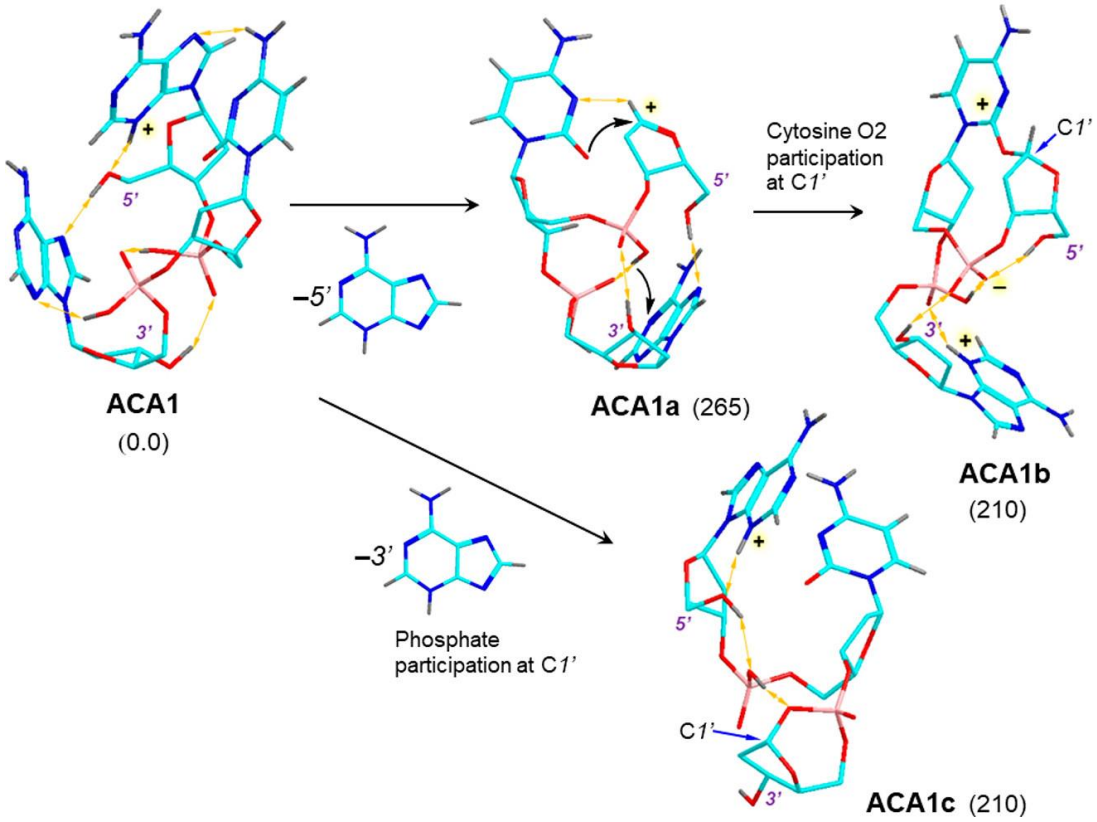


Figure 2.10: M06-2X/6-31+G(d,p) optimized structures of low-Gibbs-energy $dACA^+$ ions. Description as in Figure 2.6.



Scheme 2.6: Loss of 5′- and 3′-Adenine from $dACA^{+a}$

[a]Description as in Scheme 2.2

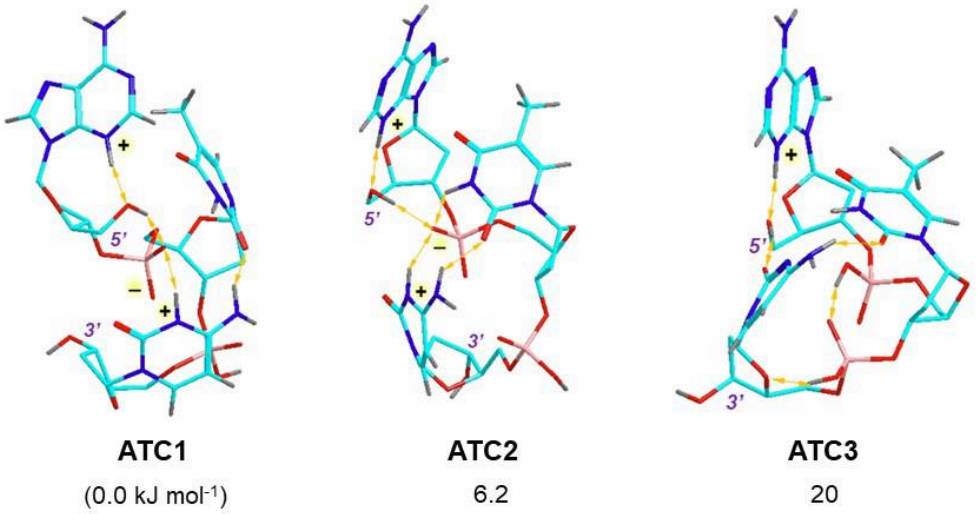
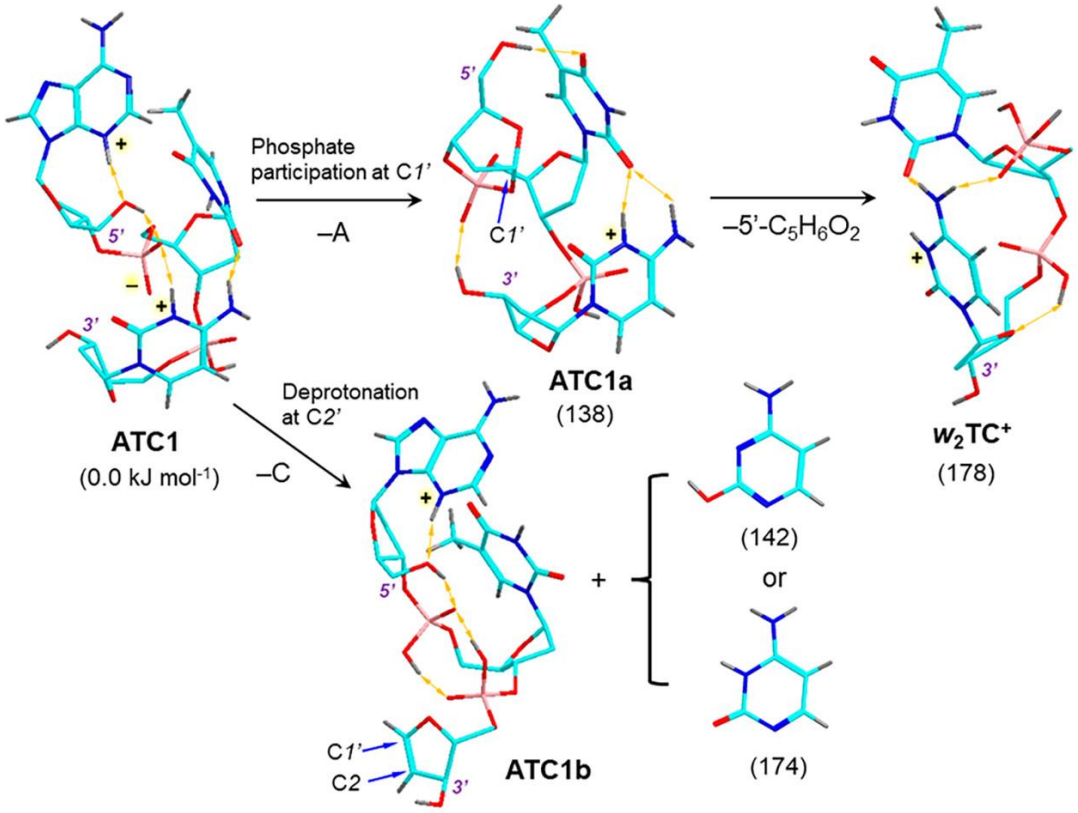


Figure 2.11: M06-2X/6-31+G(d,p) optimized structures of low-Gibbs-energy dATC$^+$ ions. Description as in Figure 2.6.



Scheme 2.7. Dissociations of the dATC$^+$ ions[a]

[a]Description as in Scheme 2.2

Low-energy structures of $dATC^+$ were obtained as zwitterion conformers **ATC1** and **ATC2** and the less stable classical structure **ATC3** (**Figure 2.11**). Both **ATC1** and **ATC2** showed a relay of hydrogen bonds interconnecting adenine and cytosine via the deprotonated phosphate and 5′-hydroxyl groups. Loss of adenine from **ATC1** can proceed with participation by the phosphate anion, forming again a cyclic phosphoester at the 5′-sugar (**ATC1a**, **Scheme 2.7**). This dissociation had a low threshold energy (138 kJ mol^{-1}) but likely involved an activation barrier in the transition state. The subsequent loss of the 5′-sugar residue was only mildly endergonic (ΔH_0 = 40 kJ mol^{-1}) to produce the $\boldsymbol{w_2}(TC)^+$ fragment ion. The competitive loss of cytosine was another low-energy dissociation of **ATC1** that proceeded with proton abstraction by the phosphate anion, forming **ATC1b** with a 1,2-double bond in the 3′-sugar residue at threshold energies ranging from 142 to 174 kJ mol^{-1}, depending on the neutral cytosine tautomer formed (**Scheme 2.7**).

In summary, the monocation structures revealed very different protonation patterns depending on the nucleobase type and position. Guanine-containing trinucleotides favored protonation at the most basic guanine position N7 but without extensive internal solvation by hydrogen bond networks, as illustrated with $dGGG^+$. Adenine- and cytosine-containing ions were more ambiguous, displaying low-energy classical structures with a single protonated nucleobase, as well as zwitterions in which another base was internally protonated by a proximate phosphate and stabilized by an extensive hydrogen bonding network. Thymine residues entirely avoided protonation in their lowest energy protomers which were protonated at the phosphoester groups instead. However, the dissociation energies for the loss of the nucleobase were all substantially above 100 kJ mol^{-1}, allowing access by proton migration to a large number of vibrationally excited monocation protomers and conformers as reactive intermediates within this energy range. Thus, given sufficient internal energy to dissociate, the monocation protomers can isomerize, making the charging proton mobile to access multiple bases in the excited ion. The proton migration can be further facilitated by hydrogen-bond networks in the ions. The dissociation outcome, as to the base loss specificity, appears to be chiefly affected by the stabilization of the intermediate oxacarbenium ions due to neighboring group participation. The stabilizing reactions were identified to include nucleophilic attack and proton abstraction by proximate phosphate groups or nucleobases that appear to be more favorable to occur at the 5′- and 3′-terminal nucleosides.

2.3.9 Resolution of Identity in Dissociations of Codon Dications

The survey of CID-MS^2 spectra of trinucleotide dications is compiled in **Table 2.3**. The survey was limited by the lack of double protonation in dTTT, dTAT, dTGT, and dTCT. Overall, the CID-MS^2 spectra were dominated by competitive losses of protonated nucleobases, AH^+, GH^+ and CH^+, that on average amounted to 84% of the singly charged fragment ion intensities. The formation of singly charged $\boldsymbol{w_2}^+$ and $\boldsymbol{w_1}^+$ ions was observed as the main backbone cleavages, as established by resolution of identity in $dXXX^{2+}$ and $dXYX^{2+}$ dications, as well as read directly from the CID-MS^2 spectra of the $dXYZ^{2+}$ trinucleotides. The average fragment ion relative intensity distribution showed 56%, 8%, and 36% of protonated base loss from the 5′-, middle, and 3′-positions, respectively. Loss of the 3′-protonated base was accompanied by elimination of water, which was less frequent for dications than it was for monocations. The average base-loss frequency favored cytosine (55%) over guanine (33%) and adenine (11%). Loss of protonated thymine was uniformly below 1%. The site specificity for the base loss was further clarified with the help of the ^{15}N-labeled codons. The CID-MS^2 spectrum of $dAAA^{2+}$ showed a predominant loss of 5′-AH^+ at 96%. This was different from the $dAAA^+$ monocation dissociations where loss of neutral 3′-adenine was competitive at 37% (**Table 2.2**). The CID-MS^2 spectrum of $dGGG^{2+}$ showed a more balanced loss of GH^+ from the 5′- and 3′-positions at 59% and 38%, respectively. Again, this contrasted the dissociations of $dGGG^+$ monocations where loss of 5′-guanine prevailed at 93% (**Table 2.2**). The $dAXA^{2+}$ dications (X = C, G, and T) showed loss of AH^+ from the 5′- and 3′-positions that was affected by the middle base, favoring 3′-adenine for X = T and G, while loss of 5′-AH^+ was favored for X = C. The $dGXG^{2+}$ dications showed loss of 5′-GH^+ which paralleled the loss of neutral 5′-guanine from $dGXG^+$ monocations. The loss of protonated cytosine from $dCXC^{2+}$ dications depended on the middle base. For $dCXC^{2+}$ dications with X = A and T, loss of 5′-CH^+ was favored at 63–64%. In contrast, with X = G, the major dissociation was loss of the 3′-CH^+. We note that complementary pairs of $(nucleobase + H)^+$ and loss-of-base ions were observed for AH^+ and GH^+ in the spectra where their m/z was above the ion trap low-mass cutoff.

Table 2.3 Fragment Ion Relative Intensities from Dications

	relative fragment ion intensity[a]					relative fragment ion intensity[a]			
codon[b]	5′	middle	3′ + water[c]	w_2[d]	codon[b]	5′	middle	3′ + water[c]	w_2[d]
AAA	96	1	3	16	A*TT*			100 (7)	1
GGG	59	3	38 (2)	18	G*TT*	10		90 (9)	15
CCC	50	1	48 (1)	5	C*TT*	24		76 (15)	38
AT**A**	47		53 (16)	18	C*AA*	98		2	6
AC**A**	69	7	24 (3)	17	G*AA*	96		4	7
AG**A**	26	13	61 (5)	21	T*AA*	5		95 (59)	13
GA**G**	89	0.4	10	9	ATC	3		97 (3)	0.5
GC**G**	65	1	34 (3)	7	ACT	8	92		3
GT**G**	76		24 (7)	26	CAT	99		1	22
CA**C**	63	0.4	37 (3)	11	CTA	99		1	51
CG**C**	13	0.3	86 (2)	3	TCA	5	50	45 (25)	6
CT**C**	64		36 (8)	55	TAC			100 (2)	
*AA*G[e]	68		32	24	GTC	45		55	3
*AA*C	46		54 (4)	9	GCT	71	28	1	11
*AA*T	100			36	CGT	100			5
*GG*A	88		12	25	CTG	84		16 (3)	46
*GG*C	44		56	5	TCG	2	25	74 (23)	4
*GG*T	100			17	TGC		3	97 (6)	0.2
*CC*A	94		6	14	ACG	50	18	32 (3)	12
*CC*G	78		22 (1)	8	AGC	4		96 (2)	0.7
*CC*T	100			14	GAC	85	1	14 (2)	13
TTG	93	2	5	46	GCA	76	1	22 (2)	8
*TT*A	94		6	18	CGA	71	2	27 (3)	32

*TT*C	94	6	23	CAG	82	5	14 (2)	23
A*CC*	16	84 (3)	3	ATG	13		87 (17)	9
G*CC*	40	60	3	AGT	62	37	1	22
T*CC*	1	99	6	GAT	99	1		18
A*GG*	8	92 (7)	5	GTA	95		5 (1)	12
C*GG*	27	73 (3)	12					
T*GG*		100 (14)	1					

[a]Percent of summed base-loss ion intensities.
[b]^{15}N-labeled nucleosides shown as bold characters;
[c]Combined relative intensities for loss of 3′-base and water. Relative intensities for loss of water in parentheses.
[d]Percent of combined base loss and $\boldsymbol{w_2}^+$ ion intensities.
[e]Nucleobase loss from positions shown in italics was not resolved, and relative intensities were summed for the unresolved positions.

Loss of thymine from the dications showed a dichotomy. In dications having thymine in combination with two basic nucleobases (A, G, or C), loss of thymine was completely suppressed. In contrast, in dications having two thymines, loss of the 5′- or 3′-TH$^+$ was dominant. This may indicate that, in contrast to dTTT$^+$ monocations, thymine was protonated in dications. However, thymine protonation may occur transiently in high-energy structures involved in dissociation rather than in the lowest energy minima.

It is noteworthy that the middle nucleobase also had an effect on the competitive backbone dissociations forming $\boldsymbol{w_2}^+$ ions. Data for dCTC^{2+}, dCTA^{2+}, and dCTG^{2+} showed the relative intensity of $\boldsymbol{w_2}^+$ ions being close to 50% (**Table 2.3**). We attempted to correlate the formation of the $\boldsymbol{w_2}^+$ ions with the loss of protonated 5′-nucleobases as shown in **Figure 2.12**. However, the plot showed a low correlation coefficient, $r^2 = 0.23$, indicating that the formation of $\boldsymbol{w_2}^+$ ions depended on factors other than just the preceding loss of the 5′-nucleobase. The sequential nature of the $\boldsymbol{w_2}^+$ ion formation was indirectly indicated by the absence of the complementary $\boldsymbol{a_1}^+$ ions in the

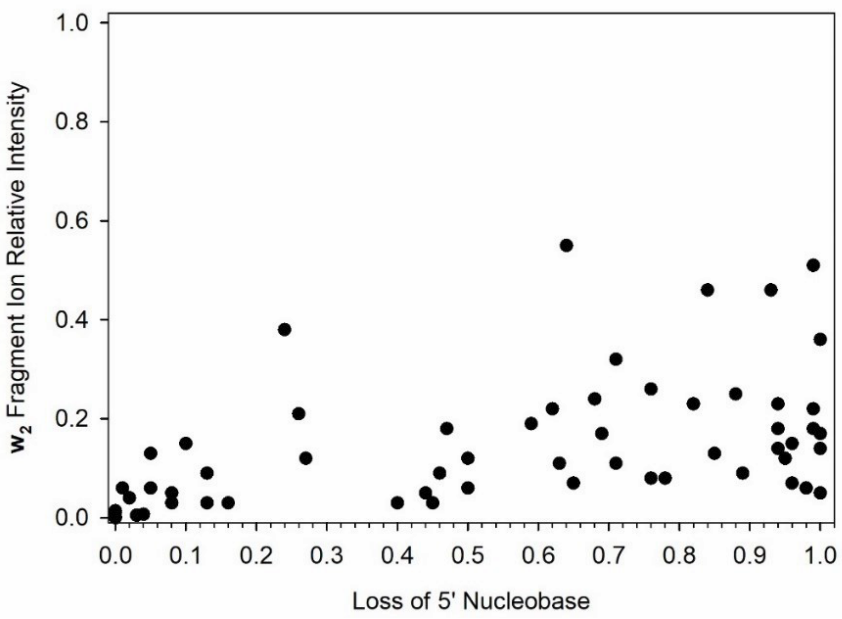


Figure 2.12: Plot of w_2^+ fragment ion relative intensities versus the relative intensity of the $(M^{2+} – 5'\text{-nucleobase}^+)^+$ ions from dications.

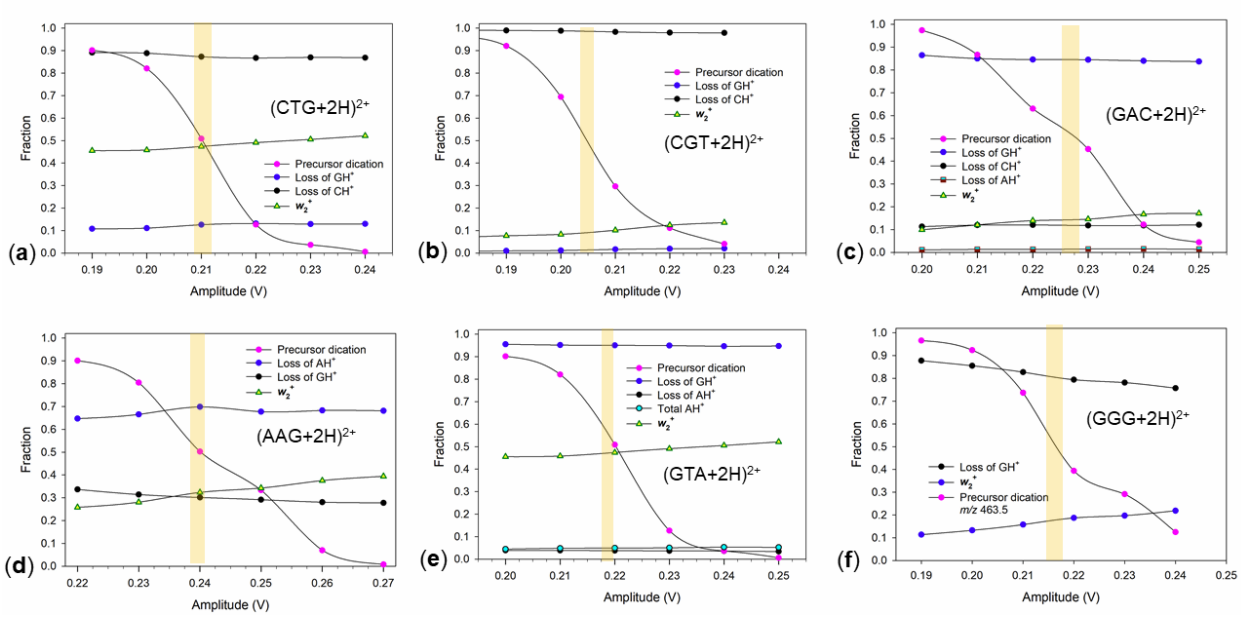


Figure 2.13: Relative intensities in the CID-MS^2 spectra of codon trinucleotide dications plotted as a function of the ion trap excitation amplitude. (a) $(CTG+2H)^{2+}$, (b) $(CGT+2H)^{2+}$, (c) $(GAC+2H)^{2+}$, (d) $(AAG+2H^{)2+}$, (e) $(GTA+2H)^{2+}$, (f) $(GGG+2H)^{2+}$. The color bars show the intervals of amplitudes used to report the spectra in Table 2.3.

CID-MS^2 spectra of the dications. The sequential formation of w_2^+ ions was also consistent with the CID-MS^2 spectra of trinucleotide dications that were measured at several excitation amplitudes and showed a weak increase of the w_2^+ ion relative intensity at the highest excitation (**Figure 2.13**).

In contrast, the relative intensities of fragment ions by competitive loss of the nucleobases were insensitive to the excitation amplitude within the range causing 5 to >95% precursor ion dissociation.

2.3.10 Codon Dication Structures and Dissociation Mechanisms

The effects on trinucleotide dication dissociations of the nucleobase nature and position, as revealed by the CID-MS^2 spectra, were multivariate, showing no simple trends. We attempted to gain insight into the dissociation mechanisms from calculations of structures and relative energies of selected protomers and conformers. For $dAAA^{2+}$, all 18 theoretically possible combinations, resulting from placing two protons at the basic N1, N3, and N7 adenine positions, were considered in the initial structures submitted to BOMD. The calculated lowest Gibbs energy $dAAA^{2+}$ ion structures showed a substantial variability as to the protonation sites and conformations. Among those, the lowest-energy classical isomers **Ad1** and **Ad2** were protonated at N3 and N1 in the 5′- and 3′-adenines. These ions showed extended conformations in which the neutral middle adenine was hydrogen bonded to the phosphate and 3′-adenine ions in **Ad1** and **Ad2**, respectively (**Figure 2.14**). Low energy zwitterionic isomers were obtained that underwent phosphate proton transfer to the third adenine residue. The lowest energy zwitterion (**Ad3**) showed protonation at N3, N1, and N7 in the 5′-, middle, and 3′-adenines, respectively. The phosphate groups in **Ad3** participated in a hydrogen-bonding network, which can facilitate proton transfer that is needed for the elimination of protonated adenine. Similar patterns, albeit with different proton distribution among the adenine basic positions were obtained for low energy zwitterionic protomers **Ad4**-**Ad6** (**Figure 2.14**). Solvation by water, as modeled by PCM calculations, in general favored the zwitterionic structures of which **Ad4** was the global Gibbs energy minimum.

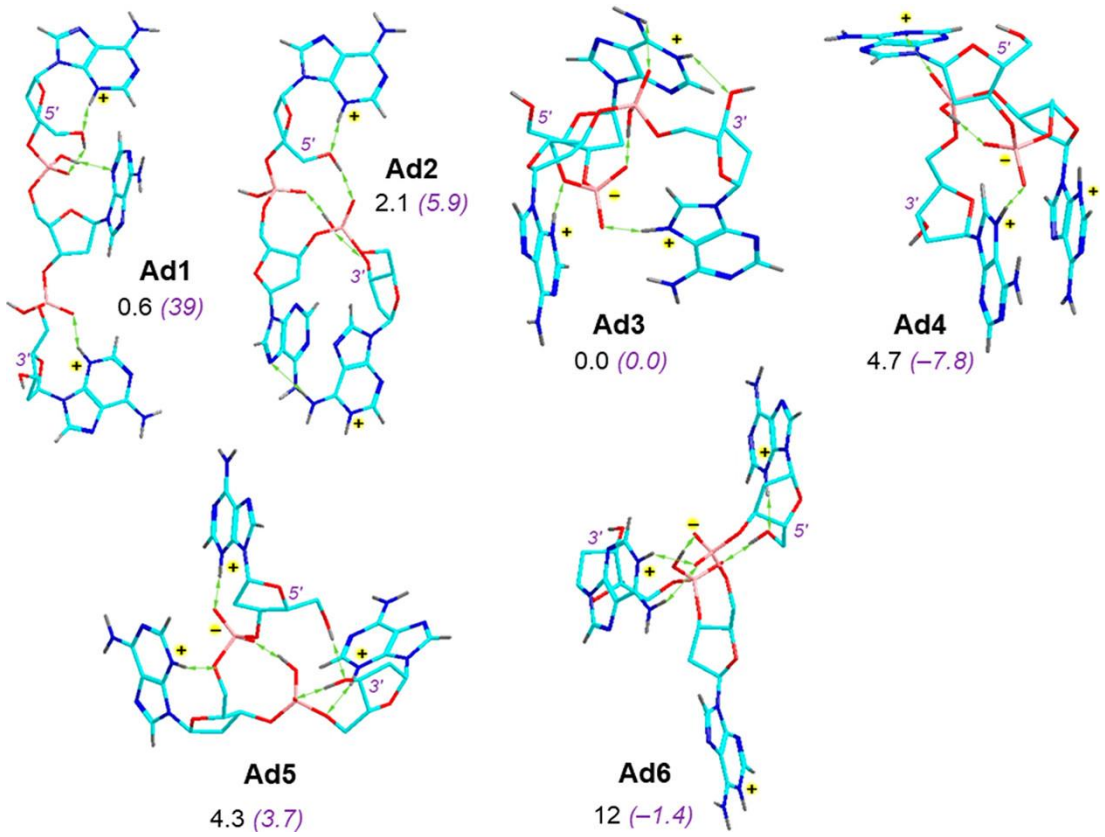


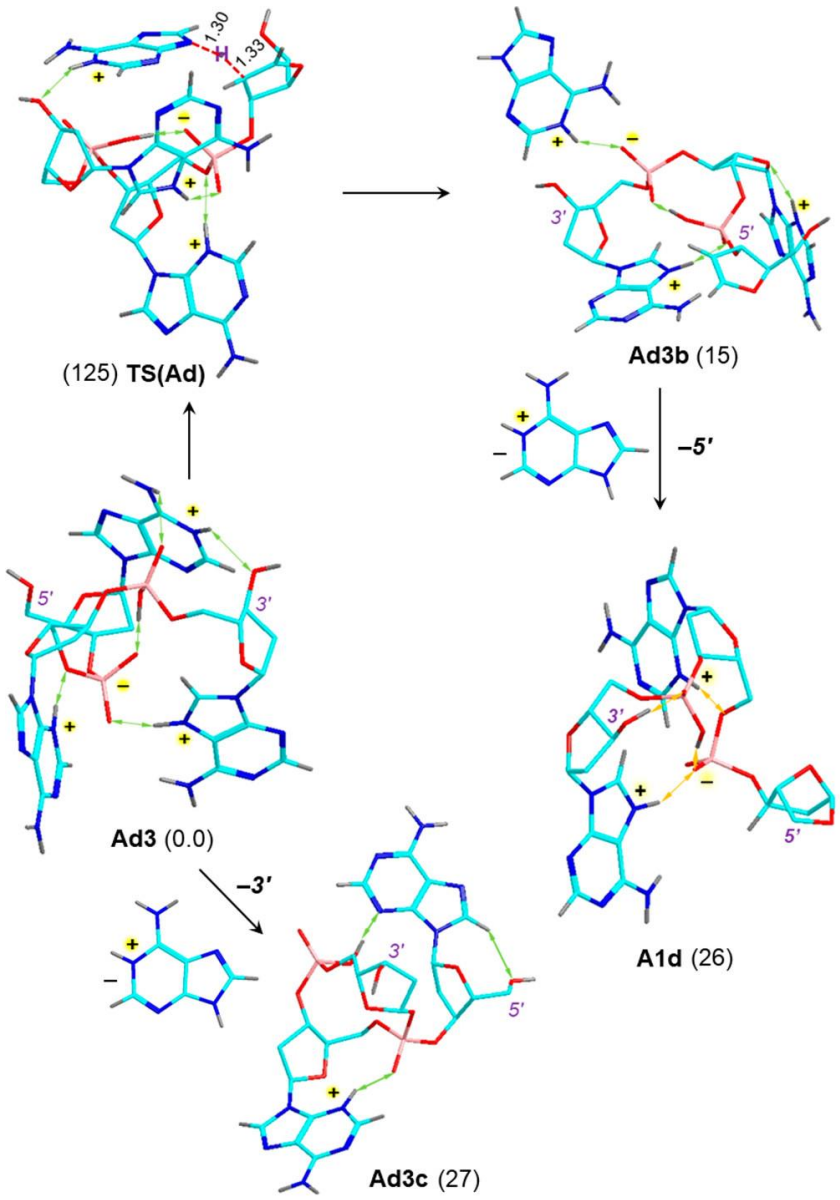
Figure 2.14: Structures of low Gibbs energy $dAAA^{2+}$ ions. Description as in Figure 2.6. Values in parentheses include PCM solvation energies (kJ mol^{-1}) in water dielectric.

The abundant formation of doubly charged $dAAA^{2+}$ and other trinucleotide ions upon electrospray was remarkable when one considers that the trinucleotides were present at equilibrium as phosphate dianions in aqueous solution, and so four protons had to be attached in electrospray microdroplets to form the gas-phase dications. We examined by calculations the gas-phase basicity for proton abstraction from the lowest Gibbs energy dication (**Ad3**) to the lowest energy monocation (**A1**). The GB_{310} value, 781 kJ mol^{-1}, was substantially lower than GB_{310}(2′-deoxyadenosine) (953 kJ mol^{-1} at the same level of theory, **Table 2.1**). This energy lowering may be in part caused by charge repulsion in the multiply charged ion. However, the diminished gas phase basicity of **A1** was sufficient to resist deprotonation of the doubly charged ions by solvent molecules in the electrospray plume.

The facile dissociation of $dAAA^{2+}$ was examined by calculations of threshold energies for the loss of AH^+ from the 3′- and 5′-positions. Protonated adenine has two nearly isoenergetic protomers (N1-H, N9-H, and N3-H, N7-H),[47–49] so considering any of these as fragments gave practically identical threshold energies. The dissociations showed very low threshold energies for

loss of both the 5′- and 3′-adenine ions, forming ions **A1d** and **Ad3c** at ΔH_0 = 26, and 27 kJ mol^{-1}, respectively (**Scheme 2.8**). These dissociation energies were substantially lower than those for the loss of neutral adenine from dAAA$^+$ monocations (186 kJ mol^{-1}, **Scheme 2.2**), indicating that the protonated adenine moiety was a substantially better leaving group than neutral adenine. We note that the classical ion structure **Ad3c** in **Scheme 2.8** was not the global energy minimum among the (AAA^{2+} – AH$^+$)$^+$ ions. Prototropic isomerization to the lower-energy zwitterionic isomer **A1c** (**Scheme 2.2**) would lower the dissociation energy of **Ad3** to 17 kJ mol^{-1}, respectively, further increasing the energy gap between the monocation and dication dissociations. The dissociation thermochemistry was qualitatively consistent with the CID-MS^2 spectra where loss of the protonated nucleobase was the dominating process. However, the slightly lower energy threshold for the loss of the 3′-adenine ion contrasted with the more favorable loss of the 5′-adenine ion in the spectrum. This indicated that the dissociation kinetics was determined by activation energies in the reaction steps preceding fragment ion separation.

To address this issue, we investigated in detail the reaction coordinate for the 5′-adenine loss from **Ad3** as summarized in **Figure 2.15**. One-dimensional elongation of the 5′-adenine N9–C1′ bond was continuously endothermic (**Figure 2.15**, top panel), according to DFT calculations. Adding atomic thermal motion at 510 K, resulted in a reorientation of the departing adenine moiety, starting at an N9–C1′ bond length of 3.1 Å and leading to exothermic transfer of the β-H2′ to adenine N9 (**Figure 2.15**, bottom panel) within a few picoseconds, forming a 2,3-glycal at the 5′-deoxyribose. We were able to find a transition state (**TS$_{Ad}$**) for the proton transfer at 125 kJ mol^{-1} relative to **Ad3** (**Scheme 2.8**). The **TS$_{Ad}$** energy was significantly lower than the TS energies for adenine loss from dAAA$^+$ monocations, attesting to the enhanced leaving group ability of charged adenine, and consistent with the general features of the CID-MS^2 spectra. Interestingly, the newly formed adenine cation did not depart the (MH – 5′-adenine)$^+$ fragment ion (**A1d**) but remained bound in a stable ion–ion complex (**Ad3b**) at 15 kJ mol^{-1} above **Ad3** that was weakly held by ΔH_0 = 22 kJ mol^{-1} (ΔG_{310} = −7.5 kJ mol^{-1}) against dissociation.



Scheme 2.8: Dissociation Pathways for the Loss of 5′- and 3′-Adenine Ions from $dAAA^{2+a}$

[a]Description as in Scheme 2.2

A bound cation–cation complex such as **Ad3b** was unexpected, indeed, in view of the long-range Coulomb repulsion between the ions. Inspection of the **Ad3b** structure revealed that the adenine cation was hydrogen bonded to the negative pole of the phosphate zwitteranion, which contributed to the complex overall stability. Both the product and intermediate complex energies in the dication dissociations were substantially lower than those calculated for the 5′-adenine loss from the singly charged ion (**Scheme 2.2**). This was also consistent with the CID-MS^2 spectra where loss of protonated bases from the dications were the dominant dissociations.

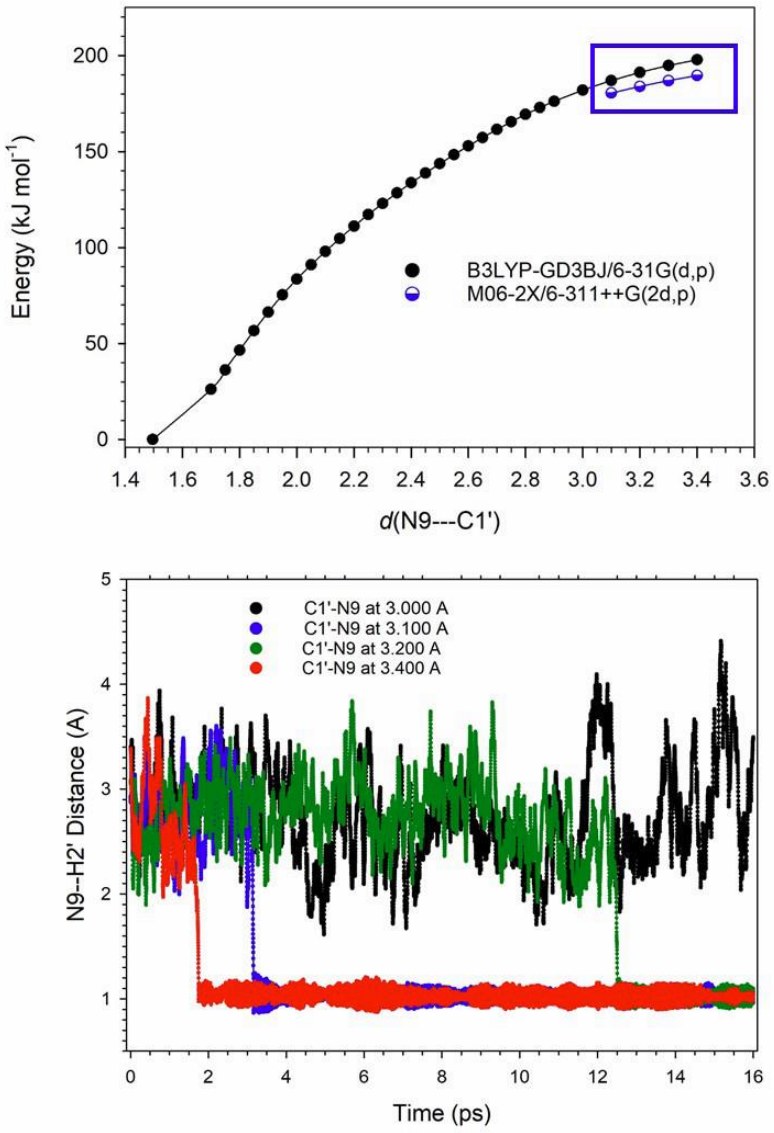
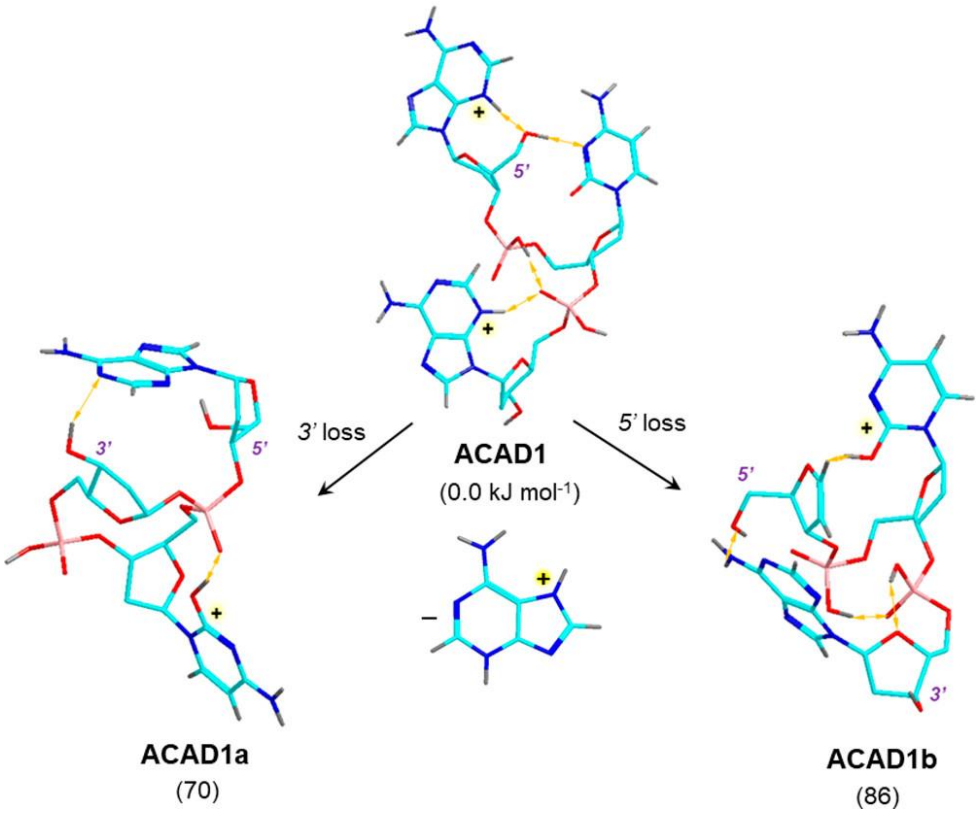


Figure 2.15: Top panel: Calculated relative energies for N9–C1' bond dissociation in dAAA dication **Ad3**. Bottom panel: BOMD trajectories at 510 K starting from the 3.100-3.400 Å points on the potential energy surface in the top panel graph and showing the formation of the adenine N9–H2' bond.

We made another comparison between the structures and dissociation energies for dACA monocations (**Scheme 2.6**) and dications. Among the 12 theoretically possible protomers combining double protonation at adenine N1, N3, and cytosine O2 and N3, the low-energy dication structures obtained by BOMD and DFT geometry optimization all had the 5′-adenine N3-H, 3′-adenine N3-H protonation pattern. The global energy minimum conformer **ACAD1** is shown in **Scheme 2.9**. The other protomeric structures that were protonated at cytosine N3 (**ACAD2**) or 3′-

adenine N1-H were less stable (**Figure 2.16**). No low energy zwitterionic structures were identified for dACA dications.



Scheme 2.9: Dissociations of $dACA^{2+}$ Dications[a]

[a]Description as in Scheme 2.2

Deprotonation of **ACAD1** to the lowest energy monocation **ACA1** was endergonic by ΔG_{310} = 807 kJ mol^{-1}, which represented the gas-phase basicity of the monocation. It is noteworthy to compare the basicity of $dACA^+$ with that of $dAAA^+$ (781 kJ mol^{-1}, vide supra), indicating a lower stabilization of the additional protonated adenine residue in the $dAAA^{2+}$ dications. According to the CID-MS^2 data, protonated adenine accounted for 93% of the base loss from dACA dications, favoring 5′-adenine over 3′-adenine. Therefore, we investigated by calculations the adenine ion loss thermochemistry and structures of the fragment ions. Loss of the 3′-adenine ion was assisted by neighboring group participation by the 5′-phosphate, forming a cyclic phosphotriester **ACAD1a** at 70 kJ mol^{-1} relative to **ACAD1**. Loss of the 5′-adenine ion was accompanied by H2′ proton transfer from the 5′-deoxyribose oxacarbenium ion to the cytosine O2, forming a glycal structure

for the product ion **ACAD1b** at 86 kJ mol^{-1} relative to **ACAD1**. These threshold energies were substantially lower than those for loss of neutral adenine from dACA monocations (210 kJ mol^{-1}, **Scheme 2.4**), indicating again that protonated adenine was a better leaving group than the neutral molecule. The relative dissociation energies for the 5′- and 3′-adenine ion loss from **ACAD1** did not conform to the distribution of the corresponding fragment ions in the identity-resolved CID-MS^2 spectrum, which favored the 5′-base loss over the 3′-base loss by a 69:24 ratio (**Table 2.3**). The structures of the fragment ions indicated complex pathways that included cleavage of the glycosidic C1′-N9 adenine bond, followed by proton transfer to the departing adenine moiety. Stabilization of the putative oxacarbenium ions by nucleophilic attack by the phosphate or proton abstraction by the cytosine O2 could occur in the intermediate complex or after protonated adenine departure.

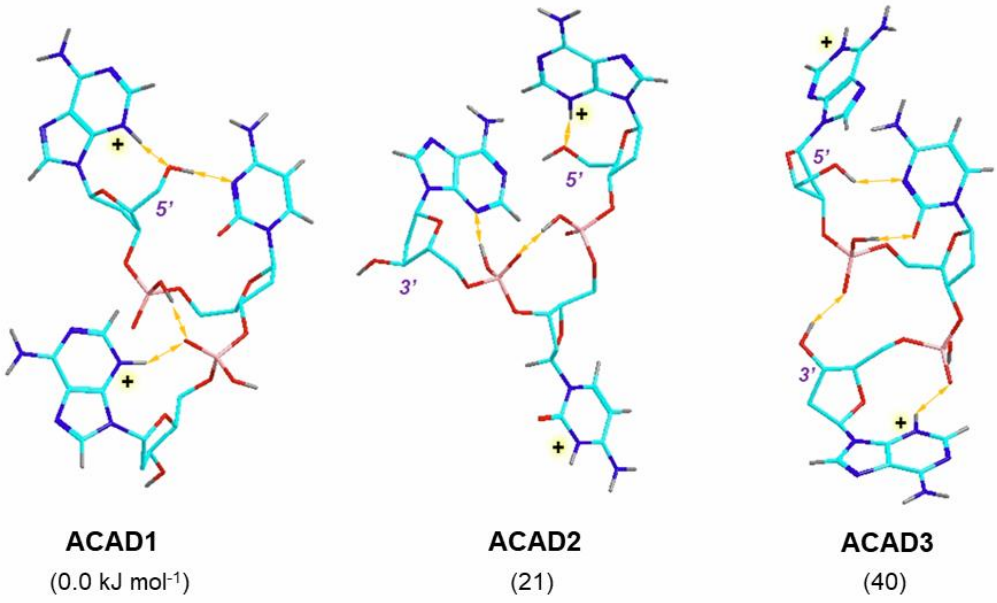


Figure 2.16: M06-2X/6-31+G(d,p) optimized structures of low-Gibbs-energy dACA^{2+} dications. Description as in Figure 2.6.

2.4 Conclusions

The combined results of this experimental and computational study allow us to arrive at the following conclusions. Resolution of identity in dissociations of trinucleotide ions related to DNA codons revealed in detail the positional propensity of adenine, guanine, cytosine, and thymine to dissociate as neutral molecules or protonated nucleobases. Loss of the nucleobase from the

terminal 5′- or 3′-positions was favored. According to combined BOMD and DFT calculations, low-energy protomers of trinucleotide mono- and dications were represented by classical nucleobase-protonated structures as well as zwitterions in which an additional nucleobase was internally protonated by a phosphate group. Dissociations of adenine and cytosine containing ions were accompanied by proton migrations that were induced by vibrational excitation and facilitated by extensive networks of intramolecular hydrogen bonds. These proton migrations occurred not only at the reaction center, but also were associated with the allosteric reorganization of the hydrogen bonding framework. The calculated dissociation and transition-state energies demonstrated that protonated adenine was a substantially better leaving group than neutral adenine, a conclusion that can be generalized for all other nucleobases.

2.5 Bibliography

[1] An, R.; Jia, Y.; Wan, B.; Zhang, Y.; Dong, P.; Li, J.; Liang, X. Non-Enzymatic Depurination of Nucleic Acids: Factors and Mechanisms. *PLoS One* **2014**, *9*, e115950.

[2] Lindahl, T. Instability and Decay of the Primary Structure of DNA. *Nature* **1993**, *362*, 709– 715.

[3] Lindahl, T.; Nyberg, B. Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry* **1972**, *11*, 3610– 3618.

[4] Lindahl, T.; Andersson, A. Rate of Chain Breakage at Apurinic Sites in Double-Stranded Deoxyribonucleic Acid. *Biochemistry* **1972**, *11*, 3618– 3623.

[5] Lindahl, T. Instability and Decay of the Primary Structure of DNA. *Nature* **1993**, *362*, 709– 715.

[6] Schaaper, R. M.; Glickman, B. W.; Loeb, L. A. Role of Depurination in Mutagenesis by Chemical Carcinogens. *Cancer Res.* **1982**, *42*, 3480– 3485.

[7] Matange, K.; Tuck, J. M.; Keung, A. J. DNA Stability: A Central Design Consideration for DNA Data Storage Systems. *Nature Commun.* **2021**, *12*, 1358.

[8] Garrett, E. R.; Mehta, P. Solvolysis of Adenine Nucleosides. I. Effects of Sugars and Adenine Substituents on Acid Solvolysis. *J. Am. Chem. Soc.* **1972**, *94*, 8532– 8541.

[9] Suzuki, T.; Ohsumi, S.; Makino, K. Mechanistic Studies on Depurination and Apurinic Site Chain Breakage in Oligodeoxyribonucleotides. *Nucleic Acids Res.* **1994**, *22*, 4997– 5003.

[10] Garrett, E. R. Kinetics of the Hydrolytic Degradation of a Nucleoside, the Antibiotic Psicofuranine. *J. Am. Chem. Soc.* **1960**, *82*, 827– 832.

[11] Hevesi, L.; Wolfson-Davidson, E.; Nagy, J.; Nagy, O.; Bruylants, A. Contribution to the Mechanism of the Acid-Catalyzed Hydrolysis of Purine Nucleosides. *J. Am. Chem. Soc.* **1972**, *94*, 4715– 4720.

[12] Tureček, F. Transient Intermediates of Chemical Reactions by Neutralization-Reionization Mass Spectrometry. *Top. Curr. Chem.* **2003**, *225*, 77– 129.

[13] Wu, R. R.; Rodgers, M. T. Mechanisms and Energetics for N-Glycosidic Bond Cleavage of Protonated Adenine Nucleosides: N3 Protonation Induces Base Rotation and Enhances N-Glycosidic Bond Stability. *Phys. Chem. Chem. Phys.* **2016**, *18*, 16021– 16032.

[14] Wu, R. R.; Chen, Yu; Rodgers, M. T. Mechanisms and Energetics for N-Glycosidic Bond Cleavage of Protonated 2′-Deoxyguanosine and Guanosine. *Phys. Chem. Chem. Phys.* **2016**, *18*, 2968– 2980.

[15] Wu, R. R.; Rodgers, M. T. O2 Protonation Controls Threshold Behavior for N-Glycosidic Bond Cleavage of Protonated Cytosine Nucleosides. *J. Phys. Chem. B* **2016**, *120*, 4803– 4811.

[16] Devereaux, Z. J.; He, C. C.; Zhu, Y.; Roy, H. A.; Cunningham, N. A.; Hamlow, L. A.; Berden, G.; Oomens, J.; Rodgers, M. T. Structures and Relative Glycosidic Bond Stabilities of Protonated 2′-Fluoro-Substituted Purine Nucleosides. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1521– 1536.

[17] Devereaux, Z. J.; Roy, H.A.; He, C.C.; Zhu, Y.; Cunningham, N.A.; Hamlow, L.A.; Berden, G.; Oomens, J.; Rodgers, M.T. Influence of 2′-Fluoro Modification on Glycosidic Bond Stabilities and Gas-Phase Ion Structures of Protonated Pyrimidine Nucleosides. *J. Fluorine Chem.* **2019**, *219*, 10– 22.

[18] Wu, R. R.; Yang, Bo; Berden, G.; Oomens, J.; Rodgers, M. T. Gas-Phase Conformations and Energetics of Protonated 2′-Deoxyadenosine and Adenosine: IRMPD Action Spectroscopy and Theoretical Studies. *J. Phys. Chem. B* **2015**, *119*, 2795– 2805.

[19] Korn, J. A.; Urban, J.; Dang, A.; Nguyen, H. T. H.; Turecek, F. UV-Vis Action Spectroscopy Reveals a Conformational Collapse in Hydrogen-Rich Dinucleotide Cation Radicals. *J. Phys. Chem. Lett.* **2017**, *8*, 4100– 4107.

[20] Vrkic, A. K.; O'Hair, R. A. J.; Foote, S.; Reid, G. E. Fragmentation Reactions of all 64 protonated Trimer Oligonucleotides Via Tandem Mass spectrometry in an Ion Trap. *Int. J. Mass Spectrom.* **2000**, *194*, 145– 164.

[21] White, H. A.: Manual of Oligonucleotide Synthesis Using the Phosphoramidite Method. In *Methods in Molecular Biology: New Nucleic Acid Techniques*; Walker, J. M., Ed.; Humana Press, 1988; Vol. 4, pp 193– 213.

[22] Pradeepkumar, P. I.; Hobartner, C.; Baum, D. A.; Silverman, S. K. DNA-Catalyzed Formation of Nucleopeptide Linkages. *Angew. Chem., Int. Ed.* **2008**, *47*, 1753– 1757.

[23] North, S. C.; Jorgensen, K. R.; Pricetolstoy, J.; Wilson, A. K. Population Analysis and the Effects of Gaussian Basis Set Quality and Quantum Mechanical Approach: Main Group through Heavy Element Species. *Frontiers in Chemistry* **2023**, *11*.

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

[25] Berendsen, H. J.; Postma, J. V.; van Gunsteren, W. F.; DiNola, A. R. H. J.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684– 3690.

[26] 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.

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

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

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

[30] 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.

[31] Huang, S. R.; Tureček, F. Noncanonical Isomers of Nucleoside Cation Radicals: An Ab Initio Study of the Dark Matter of DNA Ionization. *J. Phys. Chem. A* **2022**, *126*, 2480– 2497.

[32] Liu, Y.; Korn, J. A.; Dang, A.; Turecek, F. Hydrogen-Rich Cation Radicals of DNA Dinucleotides. Generation and Structure Elucidation by UV-Vis Action Spectroscopy. *J. Phys. Chem. B* **2018**, *122*, 9665– 9680.

[33] Huang, S. R.; Liu, Y.; Tureček, F. UV-Vis Photodissociation Action Spectroscopy Reveals Cytosine-Guanine Hydrogen Transfer in DNA Tetranucleotide Cation Radicals upon One-Electron Reduction. *J. Phys. Chem. B* **2020**, *124*, 3505– 3517.

[34] Liu, Y.; Huang, S. R.; Tureček, F. Guanine-Adenine Interactions in DNA Tetranucleotide Cation Radicals Revealed by UV/Vis Photodissociation Action Spectroscopy and Theory. *Phys. Chem. Chem. Phys.* **2020**, *22*, 16831– 16842.

[35] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999– 3093.

[36] Zhu, L., Hase, W. L.: *Quantum Chemistry Program Exchange*. Indiana University, Bloomington, 1994; QCPE 644.

[37] Gregersen, J. A.; Tureček, F. Mass-Spectrometric and Computational Study of Tryptophan Radicals (Trp + H)$^{\bullet}$ Produced by Collisional Electron Transfer to Protonated Tryptophan in the Gas Phase. *Phys. Chem. Chem. Phys.* **2010**, *12*, 13434– 13447.

[38] McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60– 70.

[39] Murray, K. K. DNA Sequencing by Mass Spectrometry. *J. Mass Spectrom.* **1996**, *31*, 1203– 1215.

[40] Schürch, S. Characterization of Nucleic Acids by Tandem Mass Spectrometry-The Second Decade (2004–2013): From DNA to RNA and Modified Sequences. *Mass Spectrom. Rev.* **2016**, *35*, 483– 523.

[41] Rodgers, M. T.; Campbell, S.; Marzluff, E. M.; Beauchamp, J. L. Site-Specific Protonation Directs Low-Energy Dissociation Pathways of Dinucleotides in the Gas Phase. *Int. J. Mass Spectrom. Ion Processes* **1995**, *148*, 1– 23.

[42] Huang, T.; Kharlamova, A.; Liu, J.; McLuckey, S. A. Ion Trap Collision-Induced Dissociation of Multiply Deprotonated RNA: c/y Ions versus (a-B)/w-Ions. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1832– 1840.

[43] Zima, V.; Liu, Y.; Tureček, F. Radical Cascade Dissociation Pathways to Unusual Nucleobase Cation Radicals. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 1038– 1047.

[44] Huang, S. R.; Dang, A.; Tureček, F. Ground and Excited States of Gas-Phase DNA Nucleobase Cation Radicals. A UV-vis Photodissociation Action Spectroscopy and Computational Study of Adenine and 9-Methyladenine. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1271– 1281.

[45] *NIST Chemistry Webbook, Standard Reference Database Number 69*; https://webbook.nist.gov/chemistry (accessed June 2022).

[46] Wu, R. R.; Hamlow, L. A.; He, C. C.; Nei, Y.-w.; Berden, G.; Oomens, J.; Rodgers, M. T. The Intrinsic Basicity of the Phosphate Backbone Exceeds that of Uracil and Thymine Residues: Protonation of the Phosphate Moiety is Preferred over the Nucleobase for pdThd and pUrd. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30351– 30361.

[47] Tureček, F.; Chen, X. Protonated Adenine: Tautomers, Solvated Clusters, and Dissociation Mechanisms. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1713– 1726.

[48] Marian, C.; Nolting, D.; Weinkauf, R. The Electronic Spectrum of Protonated Adenine: Theory and Experiment. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3306– 3316.

[49] Hud, N. V.; Morton, T. H. DFT Energy Surfaces for Aminopurine Homodimers and Their Conjugate Acid Ions. *J. Phys. Chem. A* **2007**, *111*, 3369– 3377.

Chapter 3

Structures and Dissociations of DNA Trinucleotide Codon Cation Radicals in the Gas Phase

Reproduced in part with permission from Jiahao Wan, Břetislav Brož, Yue Liu, Shu R. Huang, Aleš Marek and František Tureček. 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*, 2, 304–319.

Abstract *Sixty DNA trinucleotide cation radicals covering a large part of the genetic code alphabet were generated by electron transfer in the gas phase, and their chemistry was studied by collision-induced dissociation tandem mass spectrometry and theoretical calculations. The major dissociations involved loss of nucleobase molecules and radicals, backbone cleavage, and cross-ring fragmentations that depended on the nature and position of the nucleobases. Mass identity in dissociations of symmetrical trinucleotide cation radicals of the* $(XXX+2H)^{+\bullet}$ *and* $(XYX+2H)^{+\bullet}$ *type was resolved by specific* ^{15}N *labeling. The specific features of trinucleotide cation radical dissociations involved the dominant formation of* $\boldsymbol{d_2}^+$ *ions, hydrogen atom migrations accompanying the formation of* $(\boldsymbol{w_2}+H)^{+\bullet}$, $(\boldsymbol{w_2}+2H)^+$, *and* $(\boldsymbol{d_2}+2H)^+$ *sequence ions, and cross-ring cleavages in the 3′- and 5′-deoxyribose moieties that depended on the nucleobase type and its position in the ion. Born–Oppenheimer molecular dynamics (BOMD) and density functional theory calculations were used to obtain structures and energies of several cation-radical protomers and conformers for* $(AAA+2H)^{+\bullet}$, $(CCC+2H)^{+\bullet}$, $(GGG+2H)^{+\bullet}$, $(ACA+2H)^{+\bullet}$, *and* $(CAA+2H)^{+\bullet}$ *that were representative of the different types of backbone dissociations. The ion electronic structure, protonation and radical sites, and hydrogen bonding were used to propose reaction mechanisms for the dissociations.*

3.1 Introduction

Ionization by high-energy particles or capture of low-energy secondary electrons results in substantial destabilization of the otherwise resistant DNA molecule.[1–4] The major post-ionization processes involve electron[5–12] and proton transfer,[13–16] as well as radical reactions resulting in

nucleobase modification,[17−21] loss, or strand breaks.[22,23] While the overall process resulting in DNA damage has been studied extensively with oligonucleotide models, the nature of the reactive intermediates involved in the early stages of DNA ionization only has been investigated with simple nucleosides and small nucleotides.[13−16,24−29] The principal hurdle to experimental structure elucidation of reactive DNA ion radicals has been their very short lifetime in the condensed phase, solution, or frozen glass, where they undergo fast proton transfer reactions with the environment.[24−29] A more recent development in addressing reactive intermediates related to DNA ionization has consisted of generating them in the rarefied gas phase, where side reactions with solvent, oxygen, and other small molecules are avoided. We have generated several neutral nucleobase radicals by femtosecond electron transfer to gas-phase cations and studied their dissociation kinetics and mechanism.[30−34] Major progress in this area has been due to O'Hair and co-workers, who applied collision-induced electron transfer in ternary complexes consisting of Cu(II), terpyridine, and nucleobases to generate cation radicals in an ion trap,[35,36] allowing their further investigation by tandem mass spectrometry[37] and photodissociation action spectroscopy.[38−40] The Cu(II) complex based collision-induced dissociation (CID) method has an inherent limitation in that it excites the complex thermally and nonspecifically, allowing other competitive dissociations to proceed and dominate. Except for the readily oxidizable guanosine and 2′-deoxyguanosine,[38−40] the side reactions have been shown to thwart the formation of even simple nucleoside cation radicals.

A different redox-based method that has been successfully applied to generate oligonucleotide cation radicals in the gas phase relies on ion–ion reactions of multiply protonated oligonucleotides with an electron donor, such as the fluoranthene anion radical that has been used in electron transfer dissociation (ETD).[41] In the nondissociative version of ETD (called ETnoD),[42] the ion–ion reaction results in one-electron reduction but is not accompanied by dissociation. This has been reported for several hexanucleotide[43] and tetranucleotide dications. [44,45] Having access to oligonucleotide cation radicals then makes it possible for one to apply tandem mass spectrometry with CID or UV–vis photodissociation to investigate the cation radical chemistry and electronic structure, as reviewed recently.[46] We now report the results of a comprehensive study of trinucleotide cation-radical dissociations covering a large part of sequences comprising the genetic code alphabet for amino acids in proteosynthesis. Our interest was in identifying the main dissociation types of these cation radicals and their relationship to the trinucleotide sequence and

electronic structure. Previous comprehensive studies of trinucleotide mono- and dications[47,48] elucidated the tendency for position and base-specific nucleobase loss and phosphate backbone cleavage occurring between the 5′- and middle position, unambiguously forming w_2^+ fragment ions. We now wish to show that dissociations of trinucleotide cation radicals display a much richer chemistry regarding backbone cleavage, proton and hydrogen atom transfer reactions, and 5′- and 3′-deoxyribose cross-ring cleavages. To resolve identity in dissociations of symmetrical trinucleotide cation radicals of the $(XXX + 2H)^{+\bullet}$ and $(XYX + 2H)^{+\bullet}$ type, we apply a set of synthetic trinucleotides specifically labeled with ^{15}N in one of the nucleobases.[49]

3.2 Experimental Section

3.2.1 Materials

Custom-made trinucleotides (95% pure) were purchased from Integrated DNA Technologies, Inc. (Coralville, IA), as reported previously.[48] ^{15}N-labeled trinucleotides were manually synthesized on a 200 nmol scale as reported.[49] Dibenzo-18-crown-6-ether (DBCE), 6,7,9,10,17,18,20,21-octahydrodibenzo[b,k]-[1,4,7,10,13,16]hexaoxacyclooctadecine, was purchased from Sigma-Aldrich (St. Louis, MO).

3.2.2 Methods

All mass spectra were obtained on a Bruker amaZon 3D-ion trap tandem mass spectrometer (Bruker Daltonik, GmbH, Bremen, Germany) equipped with an auxiliary chemical ionization source for proton transfer and electron transfer dissociations.[50] Electrospray ionization was performed in acetonitrile–water–acetic acid solutions at 50:50:1 component ratios and 10–20 μM concentration of the unlabeled trinucleotides and DBCE. The ^{15}N-labeled trinucleotides that were only available in nmol quantities were dissolved in 1 mL of the solvent mixture, and the solutions were electrosprayed as described previously.[48] Mass-selected doubly charged ions were subjected to ion–ion reactions with fluoranthene anion radicals (ETD-MS^2, 150 ms), and the cation radicals formed were selected for collision-induced dissociation (CID-MS^3). The typical excitation amplitude was 0.4–0.5 on the instrument scale to achieve >60% dissociation of the selected

precursor cation radical. A few cation radicals derived from GAA and GCG required a 0.6 amplitude setting to achieve >60% dissociation.

3.2.3 Calculations

Structures of multiple protomers and conformers of trinucleotide cation radicals, (AAA + 2H)$^{+\bullet}$, (GGG + 2H)$^{+\bullet}$, (CCC + 2H)$^{+\bullet}$, (ACA + 2H)$^{+\bullet}$, and (CCA + 2H)$^{+\bullet}$, were obtained by combining Born–Oppenheimer molecular dynamics (BOMD) with density functional theory calculations in the spin-unrestricted format. Initial structures were generated for combinations of low-energy protomers at adenine (protonation at N1, N3, and N7), guanine (protonation at N7), and cytosine (protonation at N3 and O2). Thymine protonation was not considered according to the previous analysis of thymine-rich trinucleotide mono- and dications.[48] BOMD trajectories for each cation-radical protomer were run for 20 ps in 1 fs steps using the Berendsen thermostat[51] with the temperature set at 510 K. BOMD was run with the hybrid semiempirical PM6-D3H4 method[52] including corrections for hydrogen bonds and dispersion interactions. PM6-D3H4 calculations were run with MOPAC[53] under the high-level Cuby4 platform.[54] Since BOMD includes both nuclear and valence-electron motion, the initial proton and hydrogen atom positions were not fixed and could migrate among accessible positions in the same or another nucleobase or phosphate oxygens. Two hundred snapshots were extracted at regular intervals from each 20000-step trajectory, and their geometries were gradient-optimized with PM6-D3H4. The resulting structures were sorted out to remove duplicates, and 15–20 structures were fully gradient optimized with B3LYP[55] and the 6-31G(d,p) basis set, which was complemented by empirical dispersion corrections with Becke–Johnson damping (GD3-BJ).[56] The local energy minima and saddle points were confirmed by harmonic frequency calculations giving the proper number of imaginary frequencies (0 and 1, respectively). Harmonic frequencies calculated by B3LYP-GD3BJ/6-31G(d,p) showed a tight correlation with the more usual B3LYP/6-31+G(d,p) frequencies with a 0.9921 slope, giving a correlation coefficient of r^2 = 0.9999 and root-mean square deviation of *rmsd* = 1.1 cm^{-1}. Several B3LYP-GD3BJ-optimized low-energy structures were further subjected to reoptimization with M06-2X[57] and the 6-31+G(d,p) basis set, and these structures were used for single-point energy calculations with M06-2X/6-311++G(2d,p). All relative energies reported here are based on M06-2X/6-311++G(2d,p) calculations including

scaled B3LYP-GD3BJ/6-31G(d,p) zero-point vibrational energies, enthalpies, and entropies. The reported thermodynamic data refer to 310 K, which is the ion trap ambient temperature. Previous benchmarking of M06-2X/6-311++G(2d,p) relative energies versus those calculated by coupled cluster methods for nucleobase and nucleoside cation radicals showed a close correlation and identical energy ranking for both sets.[58] Atomic charges and spin densities were obtained from natural population analysis[59] (NPA) calculations with M06-2X/6-311++G(2d,p). These standard calculations were performed with the Gaussian 16 suite of programs, Revision B.01, licensed from Gaussian, Inc. (Wallingford, CT).

3.3 Results and Discussion

3.3.1 Generation of Codon Cation Radicals

The cation radicals were prepared from mass-selected dications obtained by electrospray ionization that were subjected to reduction by electron transfer in ion–ion reactions with fluoranthene anions. The electron transfer was calculated to be highly exothermic ($\Delta H_{rxn} < -580$ kJ mol^{-1}), resulting in a substantial dissociation (ETD) of the formed cation radicals. This is illustrated with $(AGC+2H)^{2+}$ (m/z 435.5, **Figure 3.1a**) that upon electron transfer showed loss of cytosine, 3′-cross-ring cleavages, and backbone dissociation forming the $\boldsymbol{d_2}$ ion at m/z 661. More seriously, ETD of codon dications also resulted in deprotonation, forming monocations, such as $(AGC+H)^{+}$ at m/z 870 (**Figure 3.1a**, inset). This side reaction was of concern, because the monocations from the various combinations of nucleosides carried along 33–39% of combined natural ^{13}C and ^{15}N isotopologues that overlapped by mass with the cation radicals. The isotope satellites were difficult to remove by narrowing the mass selection window for the dications because of reduced ion intensity. To improve the formation of codon cation radicals, we used ETD of doubly charged noncovalent complexes of trinucleotides with dibenzo-18-crown-6-ether (DBCE),[60–62] as shown for the $(AGC+DBCE+2H)^{2+}$ ion, m/z 615.5 (**Figure 3.1b**). Complexation with crown ethers such as DBCE has been shown to lower the recombination energy of the dication, thus reducing the electron transfer exergonicity.[63] For example, the recombination energies of dAA dinucleotide dications have been shown to drop from 617 to 485 kJ mol^{-1} in the DBCE complex.[60] DBCE binding to the trinucleotide ion is greatly reduced upon reduction,

chiefly due to weaker hydrogen bonds to the nucleobase radical.[63] In addition, the departing DBCE carries away a fraction of the ET-generated excitation energy, which has been estimated to be roughly proportional to the number of degrees of freedom of the products. [64] For the trinucleotide complexes, DBCE was estimated to retain 32–35% of internal energy after ETD, aiding trinucleotide cation radical stabilization. These combined effects in complex reduction and expulsion of the crown ether resulted in a very clean production of trinucleotide cation radicals, such as $(AGC+2H)^{+\bullet}$ at m/z 871, showing less dissociation and virtually no deprotonation. Interestingly, the formation of $(AGC+H)^{+}$ upon ETD of naked codon dications was entirely due to proton transfer in the ion–ion reaction, and not loss of a hydrogen atom from the cation radicals, as previously discussed for peptide cation radicals.[65] This was corroborated by the CID-MS^3 spectra of the codon cation radicals, such as that of $(AGC+H)^{+\bullet}$ (**Figure 3.1c**), which showed no loss of hydrogen. DBCE complexes were used to generate all cation radicals that showed substantial deprotonation accompanying electron transfer reduction, in particular those from AGC, CAC, CCC, CGC, and CTC. The CID-MS^3 spectra of the cation radicals showed fragment ions arising by loss of a nucleobase and backbone dissociations leading to $\boldsymbol{d_2}^+$ and $\boldsymbol{w_2}^+$ ions. In addition, hydrogen transfer reactions were observed, resulting in the loss of nucleobase radicals, primarily $GH^{\bullet}$, and formation of $(\boldsymbol{w_2} \pm nH)^{+\bullet}$ and $(\boldsymbol{d_2} + 2H)^{+}$ ions. These reactions depended on the nature and position of the nucleobases in the codon cation radical. The dissociations were resolved by ^{15}N labeling, as described systematically below for the different nucleotide sequences.

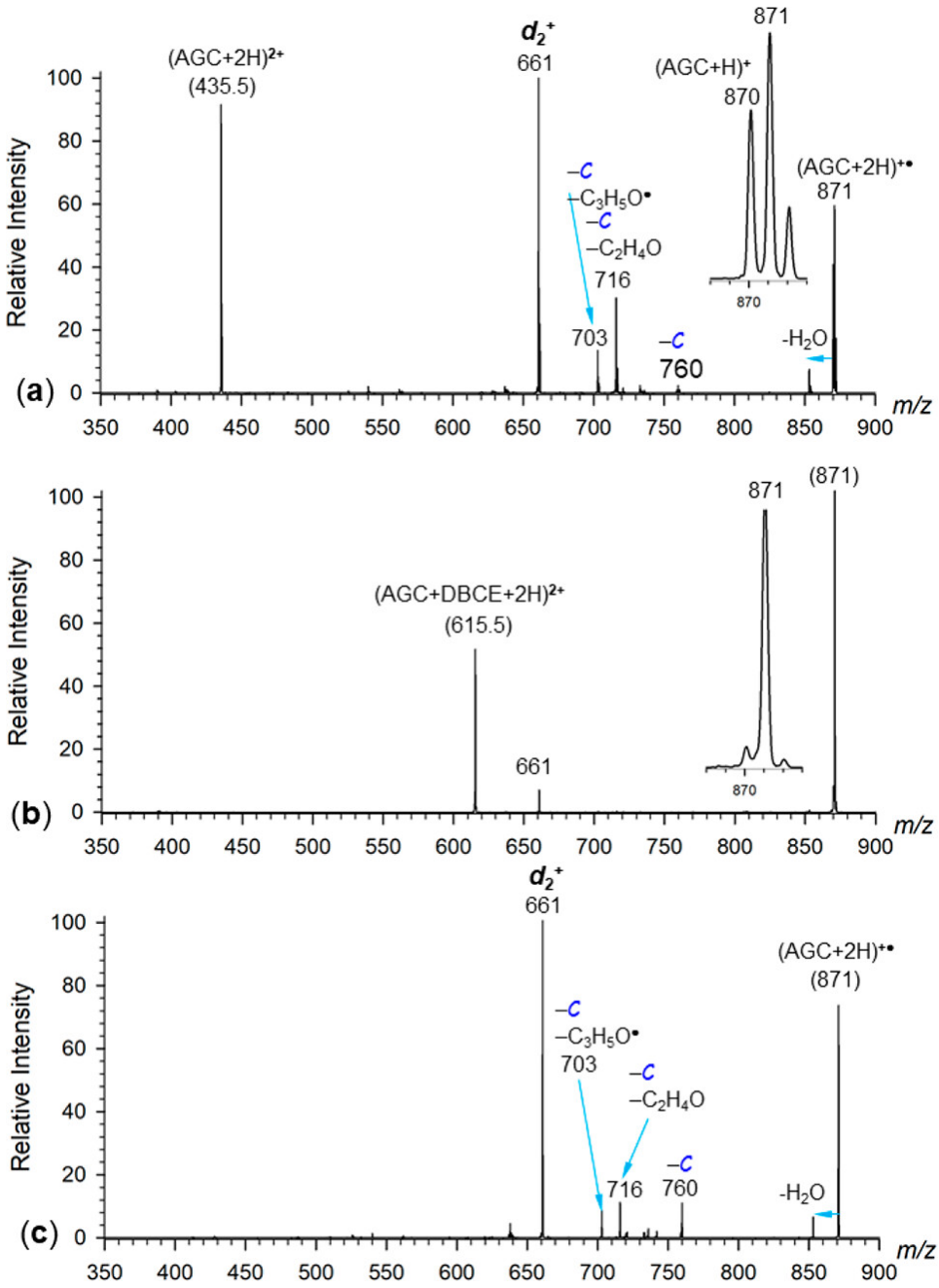


Figure 3.1: ETD-MS^2 spectra of (a) (AGC+2H)$^{2+}$ at m/z 435.5 and (b) noncovalent complex with dibenzo-18-crown-6-ether, (AGC+DBCE+2H)$^{2+}$, at m/z 615.5. (c) CID-MS^3 spectrum of (AGC+2H)$^{+\bullet}$ at m/z 871 that was mass-isolated from the spectrum in (b). Insets in (a) and (b) show the peak profiles of the charge-reduced ions.

3.3.2 Note on Nomenclature

We use the traditional nomenclature of DNA fragment ions that has been established for even-electron ions.[66–68] Thus, $\boldsymbol{a_n}^+$, $\boldsymbol{d_n}^+$, and $\boldsymbol{w_n}^+$ even-electron ions correspond by mass to neutral molecular fragments from backbone phosphate ester cleavage with an attached proton at a nucleobase. Radical cations having the m/z one unit higher or lower than $\boldsymbol{a_n}^+$ and $\boldsymbol{d_n}^+$ are denoted as $(\boldsymbol{a_n}^+ \pm \mathrm{H})^{+\bullet}$ $(\boldsymbol{d_n}^+ \pm \mathrm{H})^{+\bullet}$, respectively. Other differences in the hydrogen atom count are expressed by $\pm n$H added to the fragment ion symbol.

3.3.3 Backbone Dissociation Types in Trinucleotide Cation Radicals

Standard backbone cleavages in trinucleotides by phosphate ester elimination can result in the formation of $\boldsymbol{d_2}^+$ ions retaining the 5′- and middle nucleobases or $\boldsymbol{w_2}^+$ ions retaining the 3′- and middle nucleobases.[47,48,66–68] With hydrogen-rich cation radicals containing an additional hydrogen atom in the radical-carrying nucleobase, one also observes backbone dissociations that are associated with hydrogen atom transfer, forming odd-electron $(\boldsymbol{w_2}+\mathrm{H})^{+\bullet}$ and even-electron $(\boldsymbol{w_2}+2\mathrm{H})^+$ and $(\boldsymbol{d_2}+2\mathrm{H})^+$ ions. The main types of dissociation by nucleobase loss, phosphate ester backbone cleavage, and cross-ring fragmentations for 60 codon trinucleotide cation radicals are summarized in the master table (**Table 3.1**). To illustrate and detail other ion fragmentations, full CID-MS^3 spectra are also shown in **Figures 3.2–3.13**.

Table 3.1 Types of Fragmentations in CID-MS^3 Spectra

sequence[a]	base loss	backbone fragments	cross-ring loss
AAA	5′ > 3′ > middle (52:42:6)	$\boldsymbol{w_2}^+ > \boldsymbol{d_2}^+$ (54:46)	
AAC	*C* > *A*	$\boldsymbol{d_2}^+$	
AAG	*A* > *G*, *G*H$^\bullet$	$\boldsymbol{w_2}^+ \approx (\boldsymbol{w_2}+\mathrm{H})^{+\bullet} > \boldsymbol{d_2}^+$	
AA**T**	*A* > *T*	$\boldsymbol{d_2}^+ \gg \boldsymbol{w_2}^+$, $(\boldsymbol{w_2}+\mathrm{H})^{+\bullet}$, $(\boldsymbol{w_2}+2\mathrm{H})^+$	(*T*+C_2H_4O, C_3H_5O)
ACA	5′-*A* ≈ 3′-*A*, *C*	$(\boldsymbol{d_2}+2\mathrm{H})^+ > (\boldsymbol{w_2}+2\mathrm{H})^+$ (67:33)	
ACC	*C* ≫ *A*	$\boldsymbol{d_2}^+$	
ACG	*C* > *G*, *G*H$^\bullet$, *A*	$(\boldsymbol{d_2}+2\mathrm{H})^+ \approx (\boldsymbol{w_2}+2\mathrm{H})^+ > (\boldsymbol{w_2}+\mathrm{H})^{+\bullet}$, $(\boldsymbol{z_2}+\mathrm{H})^{+\bullet}$	

sequence[a]	base loss	backbone fragments	cross-ring loss
ACT	*C* > *A*	($\boldsymbol{z_2}$+H)$^{+\bullet}$, ($\boldsymbol{w_2}$+2H)$^+$ > ($\boldsymbol{d_2}$+2H)$^+$> $\boldsymbol{d_2}^+$,	
AGA	3′-*A* > 5′-*A*, *C*	$\boldsymbol{d_2}^+$ > $\boldsymbol{w_2}^+$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$	(*A*+C_2H_4O, C_3H_5O)
AGC	*C*	$\boldsymbol{d_2}^+$	(*C*+C_2H_4O, C_3H_5O)
AGG	*G* > *A*	($\boldsymbol{w_2}$+H)$^{+\bullet}$ > $\boldsymbol{d_2}^+$, ($\boldsymbol{a_2}$–H)$^{+\bullet}$	(*G*+C_2H_4O, C_3H_5O)
AGT	*A* ≈ *G*	$\boldsymbol{d_2}^+$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$ > $\boldsymbol{w_2}^+$	(*T*+C_2H_4O, C_3H_5O)
ATA	5′-*A* ≈ 3′-*A*	$\boldsymbol{d_2}^+$ > $\boldsymbol{w_2}^+$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$	
ATC	*C* ≫ *A*	$\boldsymbol{d_2}^+$ ≫ ($\boldsymbol{w_2}$+H)$^{+\bullet}$	(*C*+C_2H_4O, C_3H_5O)
ATG	*A* > *G*	($\boldsymbol{w_2}$+H)$^{+\bullet}$ > $\boldsymbol{w_2}^+$ > $\boldsymbol{d_2}^+$	
ATT	*A* ≈ *T*	$\boldsymbol{d_2}^+$ > $\boldsymbol{w_2}^+$, ($\boldsymbol{a_2}$–H)$^{+\bullet}$	(*T*+C_3H_5O)
CAA	*C*	$\boldsymbol{w_2}^+$	$CH_2OH^\bullet$
CAC	*A*	$\boldsymbol{d_2}^+$, $\boldsymbol{d_2}^+$–*C*	
CAG	*C*	$\boldsymbol{w_2}^+$ > ($\boldsymbol{w_2}$–H)$^{+\bullet}$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$	$CH_2OH^\bullet$
CAT	*C*	$\boldsymbol{w_2}^+$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$	$CH_2OH^\bullet$
CCA	*C*	($\boldsymbol{d_2}$+2H)$^+$ > $\boldsymbol{w_2}^+$	$CH_2OH^\bullet$
CCC	3′ > 5′ ≈ middle > *C*H$^\bullet$ 90:05:05	$\boldsymbol{d_2}^+$ > ($\boldsymbol{d_2}$+2H)$^+$	(*C*+C_2H_4O, C_3H_5O)
CCG	*G*$^\bullet$ > *C*	($\boldsymbol{d_2}$+2H)$^+$ ≈ ($\boldsymbol{w_2}$–H)$^{+\bullet}$ > $\boldsymbol{w_2}^+$> ($\boldsymbol{w_2}$+2H)$^+$	$CH_2OH^\bullet$
CCT	*C*	$\boldsymbol{w_2}^+$ ≫ ($\boldsymbol{w_2}$+H)$^{+\bullet}$	$CH_2OH^\bullet$
CGA	*C* ≫ *A*	$\boldsymbol{w_2}^+$ > ($\boldsymbol{w_2}$–H)$^{+\bullet}$ ≈ ($\boldsymbol{w_2}$+H)$^{+\bullet}$	$CH_2OH^\bullet$
CGC	5′-*C* > 3′-*C*, *G*$^\bullet$	$\boldsymbol{d_2}^+$	(3′-*C*+C_2H_4O)
CGG	*C* ≈ *G*	($\boldsymbol{w_2}$–H)$^{+\bullet}$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$ > $\boldsymbol{w_2}^+$	$CH_2OH^\bullet$
CTA	*C*	$\boldsymbol{w_2}^+$	$CH_2OH^\bullet$
CTC	3′-*C* > *T*$^\bullet$	$\boldsymbol{d_2}^+$ > ($\boldsymbol{d_2}$–H)$^{+\bullet}$	
CTG	*C* > *G*	$\boldsymbol{w_2}^+$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$ > ($\boldsymbol{w_2}$–H)$^{+\bullet}$	$CH_2OH^\bullet$
CTT		$\boldsymbol{d_2}^+$ ≈ $\boldsymbol{w_2}^+$	$CH_2OH^\bullet$
GAA	*A* > *G*, *G*H$^\bullet$, GH_2	$\boldsymbol{w_2}^+$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$ ≈ $\boldsymbol{d_2}^+$	
GAC	*G* > *C*	$\boldsymbol{d_2}^+$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$ > $\boldsymbol{w_2}^+$	(*C*+C_2H_4O, C_3H_5O)
GAG	5′-*G* > 3′-*G*	($\boldsymbol{w_2}$+H)$^{+\bullet}$ > $\boldsymbol{w_2}^+$ ≫ $\boldsymbol{d_2}^+$	
GAT	*G*H$^\bullet$ > *G*	$\boldsymbol{w_2}^+$ > ($\boldsymbol{w_2}$+H)$^{+\bullet}$ > $\boldsymbol{d_2}^+$	

sequence[a]	base loss	backbone fragments	cross-ring loss
GCA	*G* > *C*	$(\boldsymbol{d_2}+2H)^+ > (\boldsymbol{w_2}^+ + H)^{+\bullet} > \boldsymbol{w_2}^+$	
GCC	*G* > *C*	$\boldsymbol{d_2}^+ \gg (\boldsymbol{w_2}+2H)^+$	(*C*+C_2H_4O, C_3H_5O)
GCG	5′-*G* > 3′-*G* > 3′-*G*H$^\bullet$	$(\boldsymbol{w_2}+H)^{+\bullet} > (\boldsymbol{w_2}+2H)^+ > (\boldsymbol{d_2}+2H)^+$	
GCT	*G* > *C*	$(\boldsymbol{w_2}^+ + H)^{+\bullet} > \boldsymbol{w_2}^+$ $(\boldsymbol{w_2}+2H)^+ > (\boldsymbol{d_2}+2H)^+$	
GGA	*G* > *A*, *G*H$^\bullet$, GH_2	$\boldsymbol{w_2}^+ \approx (\boldsymbol{w_2}+H)^{+\bullet} > \boldsymbol{d_2}^+$	
GGC	*G* > *C*, *G*H$^\bullet$	$\boldsymbol{d_2}^+$	(*C*+C_2H_4O, C_3H_5O)
GGG	5′ > 3′ > middle (57:30:13)	$(\boldsymbol{w_2}+H)^{+\bullet} > \boldsymbol{w_2}^+$	
GGT	*G* > *G*H$^\bullet$	$(\boldsymbol{w_2}+H)^{+\bullet} > \boldsymbol{w_2}^+$	(*T*+C_3H_5O)
GTA	*G* > *G*H$^\bullet$ > *G*H_2, *C*	$\boldsymbol{w_2}^+ > (\boldsymbol{w_2}+H)^{+\bullet} > \boldsymbol{d_2}^+$	
GTC	*G* > *C* > *T*$^\bullet$, *G*H$^\bullet$	$\boldsymbol{d_2}^+ > (\boldsymbol{w_2}+H)^{+\bullet} > \boldsymbol{w_2}^+$	(*C*+C_2H_4O, C_3H_5O)
GTG	5′-*G* > 3′-*G*	$(\boldsymbol{w_2}+H)^{+\bullet} > \boldsymbol{w_2}^+$	
GTT	*G* ≈ *T*	$(\boldsymbol{w_2}+H)^{+\bullet} > \boldsymbol{w_2}^+$	98 Da
TAA	*A*	$\boldsymbol{d_2}^+ > > \boldsymbol{w_2}^+ > (\boldsymbol{d_2}+2H)^+ > (\boldsymbol{w_2}+2H)^+$	$CH_2OH^\bullet$
TAC	*C*	$\boldsymbol{d_2}^+ \gg (\boldsymbol{w_2}+2H)^+$	(*C*+C_2H_4O, C_3H_5O)
TAT		$(\boldsymbol{w_2}/\boldsymbol{d_2}+2H)^+ > \boldsymbol{w_2}/\boldsymbol{d_2}$	(41 Da)
TCA	*C* > *A*	$(\boldsymbol{d_2}+2H)^+ \gg (\boldsymbol{w_2}+2H)^+$	
TCC	*C*	$\boldsymbol{d_2}^+ > (\boldsymbol{d_2}-H)^{+\bullet} \approx (\boldsymbol{w_2}+2H)^+, (\boldsymbol{z_2}+H)^{+\bullet}$	
TCG	*C*, *G*	$(\boldsymbol{w_2}+2H)^+ > (\boldsymbol{d_2}+2H)^+, (\boldsymbol{z_2}+H)^{+\bullet}$	
TCT		$(\boldsymbol{w_2}/\boldsymbol{d_2}+2H)^+ > \boldsymbol{w_2}/\boldsymbol{d_2}, (\boldsymbol{a_2}/\boldsymbol{z_2}+H)^{+\bullet}$	
TG**C**	*C*	$\boldsymbol{d_2}^+ > (\boldsymbol{w_2}+2H)^+$	(*C*+C_2H_4O, C_3H_5O)
TGG	*G*	$\boldsymbol{w_2} \approx \boldsymbol{d_2}$	
TGT	*G* > *T*, *T*H$^\bullet$	$\boldsymbol{w_2}/\boldsymbol{d_2} > (\boldsymbol{w_2}/\boldsymbol{d_2}+H)^{+\bullet}$	(*T*+C_2H_4O, C_3H_5O)
TTA		$(\boldsymbol{w_2}+2H)^+ > \boldsymbol{w_2}^+$	
TTC		$(\boldsymbol{w_2}+2H)^+ > (\boldsymbol{w_2}-2H)^+$	
TTG		$\boldsymbol{w_2}^+ > (\boldsymbol{w_2}+H)^{+\bullet}, (\boldsymbol{w_2}-H)^{+\bullet}, \boldsymbol{d_2}^+$	

[a]Sequences for which fragment ion assignments were supported by ^{15}N labeling are shown in boldface.

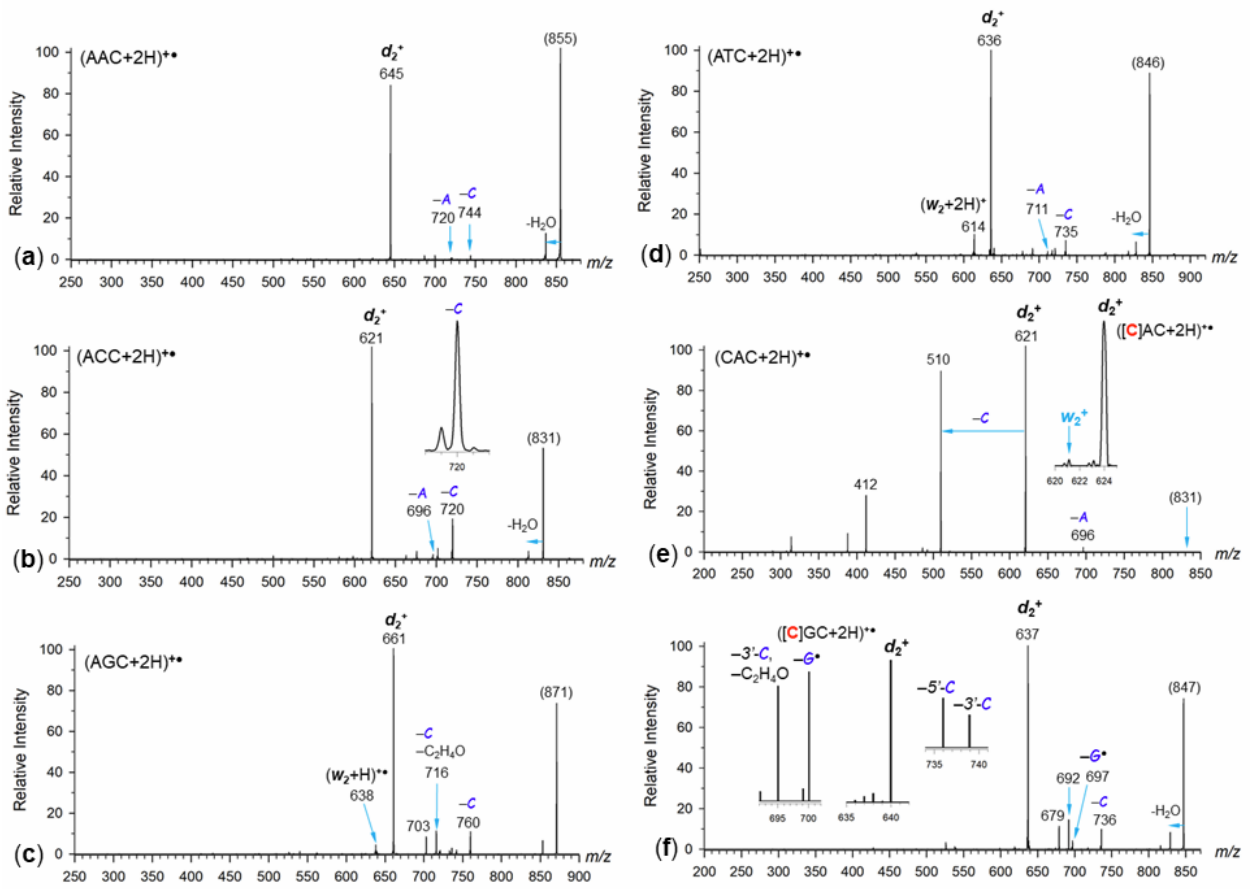


Figure 3.2: CID-MS^3 spectra of the $\boldsymbol{d_2}^+$ group of cation radicals. (a) (AAC+2H)$^{+\bullet}$, (b) (ACC+2H)$^{+\bullet}$, (c) (AGC+2H)$^{+\bullet}$, (d) (ATC+2H)$^{+\bullet}$, (e) (CAC+2H)$^{+\bullet}$, and (f) (CGC+2H)$^{+\bullet}$ with insets showing fragment ion distribution from ([$^{15}N_3$-C]GC+2H)$^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

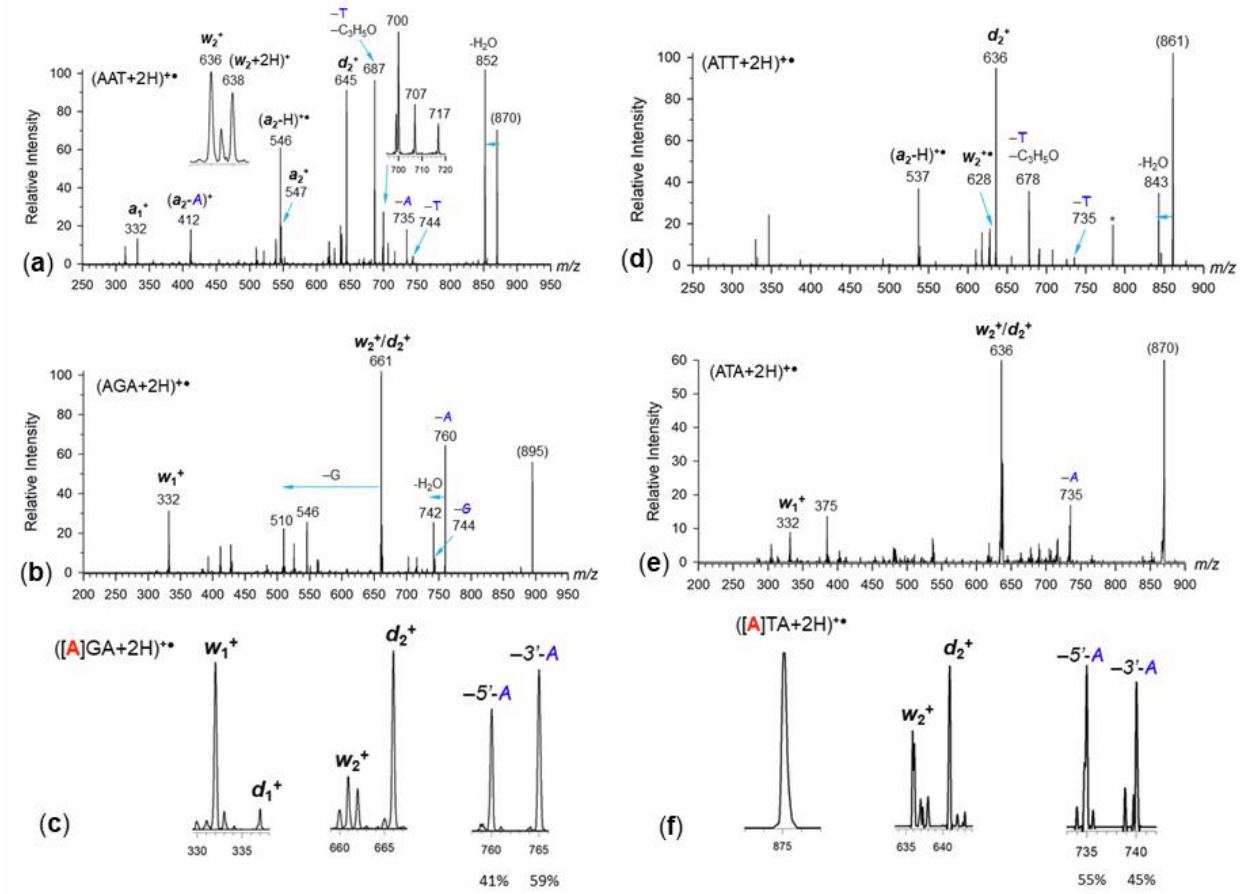


Figure 3.3: CID-MS^3 spectra of the $\boldsymbol{d_2}^+$ group of cation radicals. (a) (AAT+2H)$^{+\bullet}$, (b) (AGA+2H)$^{+\bullet}$, (c) fragment ion distribution from ([$^{15}N_5$-A]GA+2H)$^{+\bullet}$, (d) (ATT+2H)$^{+\bullet}$, (e) (ATA+2H)$^{+\bullet}$, (f) fragment ion distribution from ([$^{15}N_5$-A]TA+2H)$^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

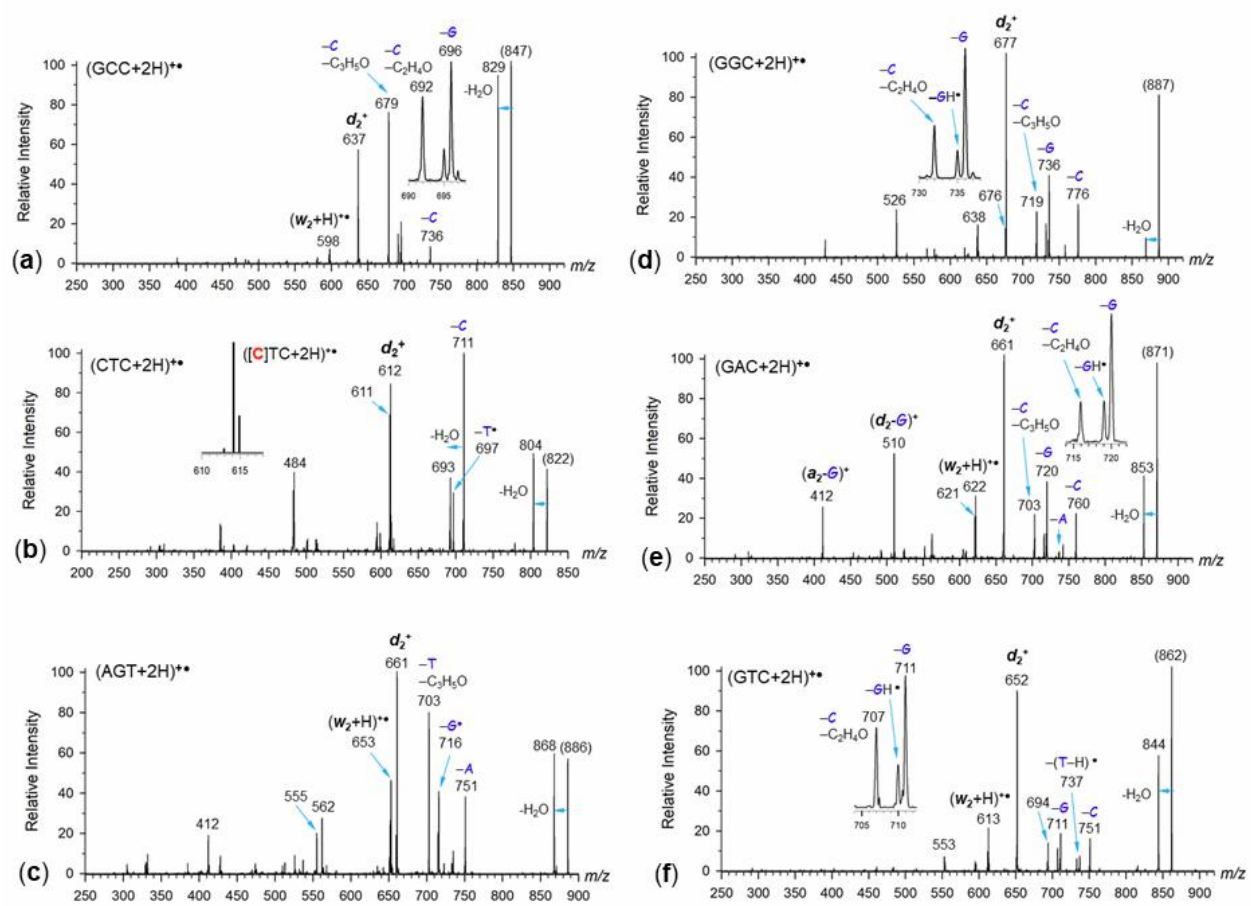


Figure 3.4: CID-MS^3 spectra of the $\boldsymbol{d_2}^+$ group of cation radicals. (a) (GCC+2H)$^{+\bullet}$, (b) (CTC+2H)$^{+\bullet}$ with inset showing the fragment ion distribution from ([$^{15}N_3$-C]TC+2H)$^{+\bullet}$, (c) (AGT+2H)$^{+\bullet}$, (d) (GGC+2H)$^{+\bullet}$, (e) (GAC+2H)$^{+\bullet}$, and (f) (GTC+2H)$^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

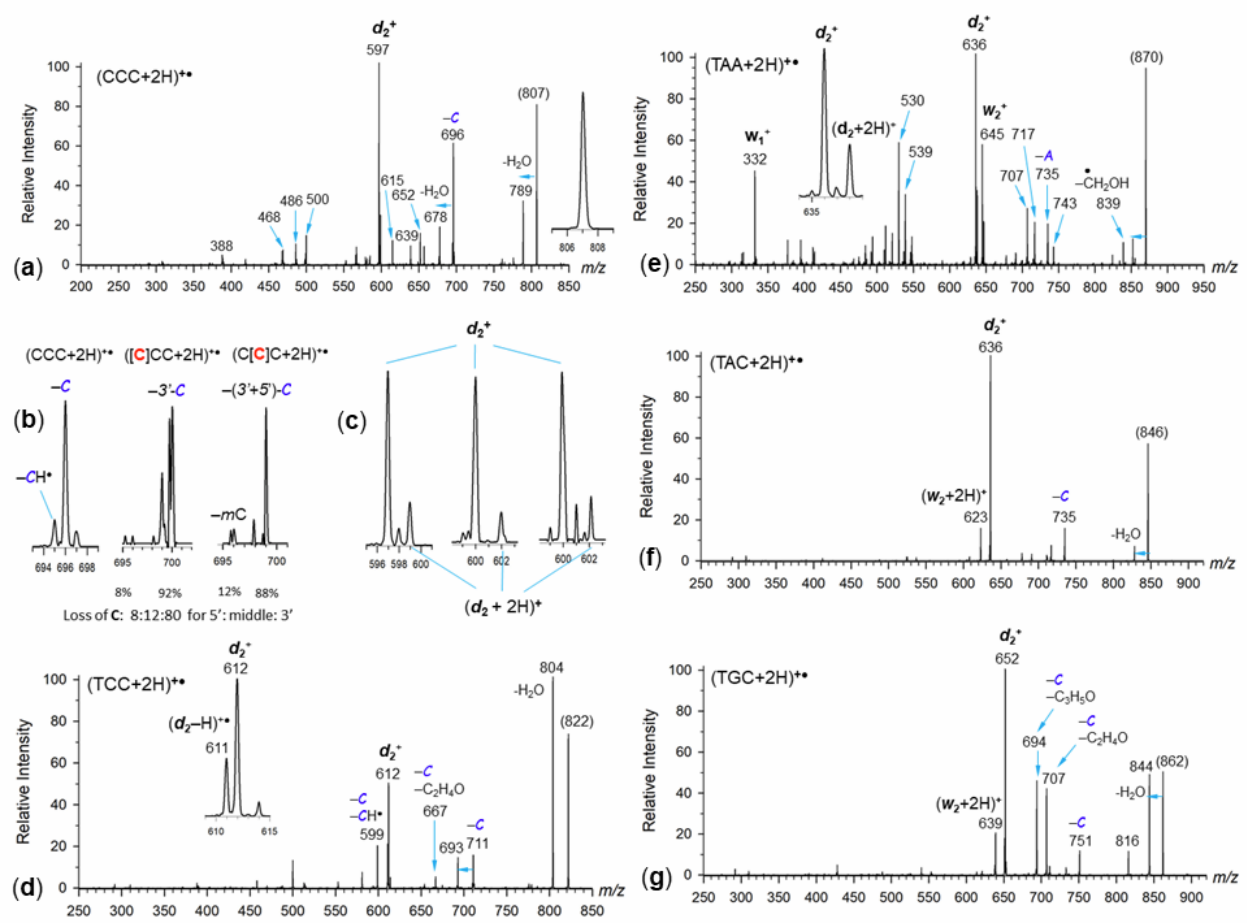


Figure 3.5: CID-MS^3 spectra of the $\boldsymbol{d_2}^+$ group of cation radicals. (a) $(CCC+2H)^{+\bullet}$, (b), (c) fragment ion distribution from $([^{15}N_3\text{-}C]CC+2H)^{+\bullet}$ and $(C[^{15}N_3\text{-}C]C+2H)^{+\bullet}$ (d) $(TCC+2H)^{+\bullet}$, (e) $(TAA+2H)^{+\bullet}$, (f) $(TAC+2H)^{+\bullet}$, and (g) $(TGC+2H)^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

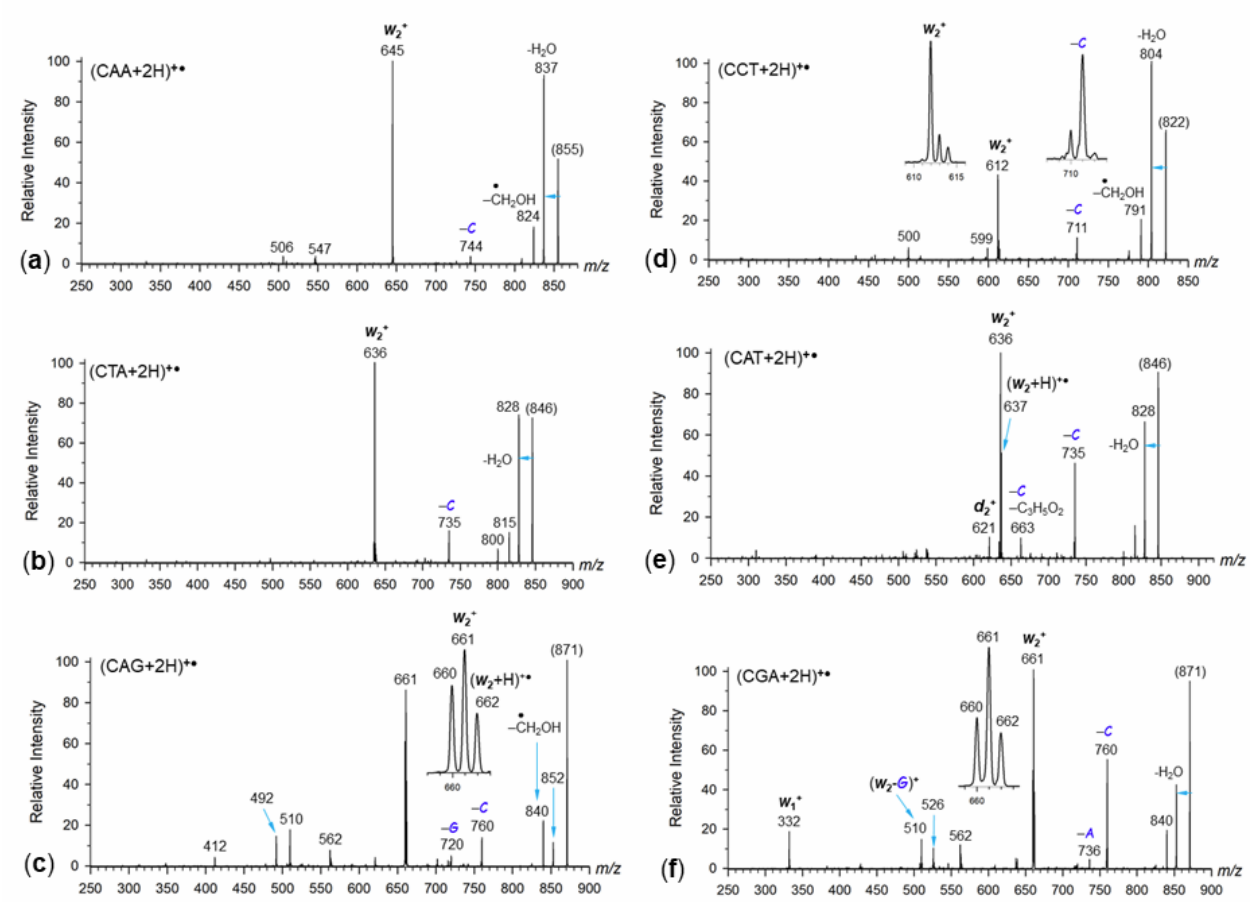


Figure 3.6: CID-MS^3 spectra of the w_2^+ group of cation radicals. (a) (CAA+2H)$^{+\bullet}$, (b) (CTA+2H)$^{+\bullet}$, (c) (CAG+2H)$^{+\bullet}$, (d) (CCT+2H)$^{+\bullet}$, (e) (CAT+2H)$^{+\bullet}$, and (f) (CGA+2H)$^{+\bullet}$.

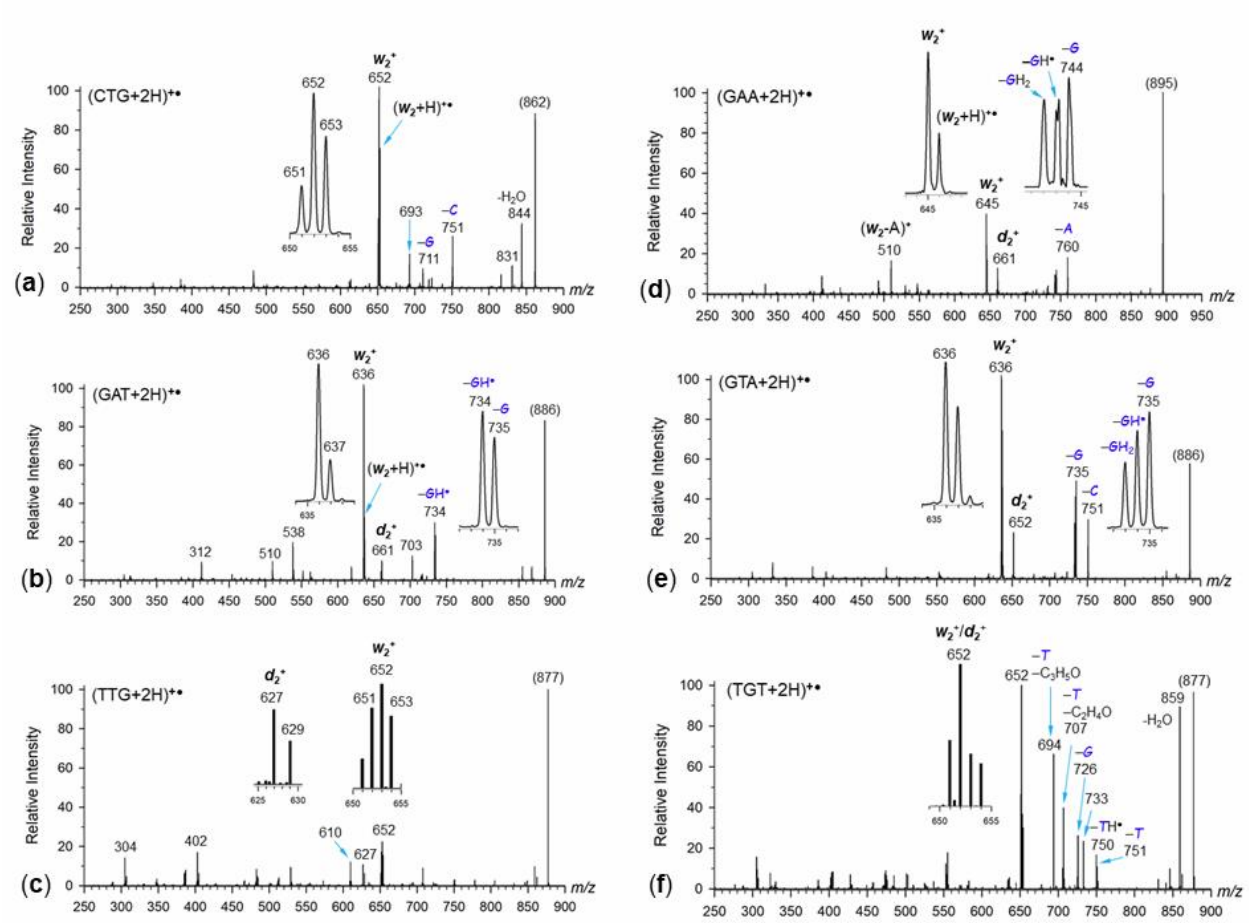


Figure 3.7: CID-MS^3 spectra of the w_2^+ group of cation radicals. (a) $(CTG+2H)^{+\bullet}$, (b) $(GAT+2H)^{+\bullet}$, (c) $(TTG+2H)^{+\bullet}$, (d) $(GAA+2H)^{+\bullet}$, (e) $(GTA+2H)^{+\bullet}$, and (f) $(TGT+2H)^{+\bullet}$.

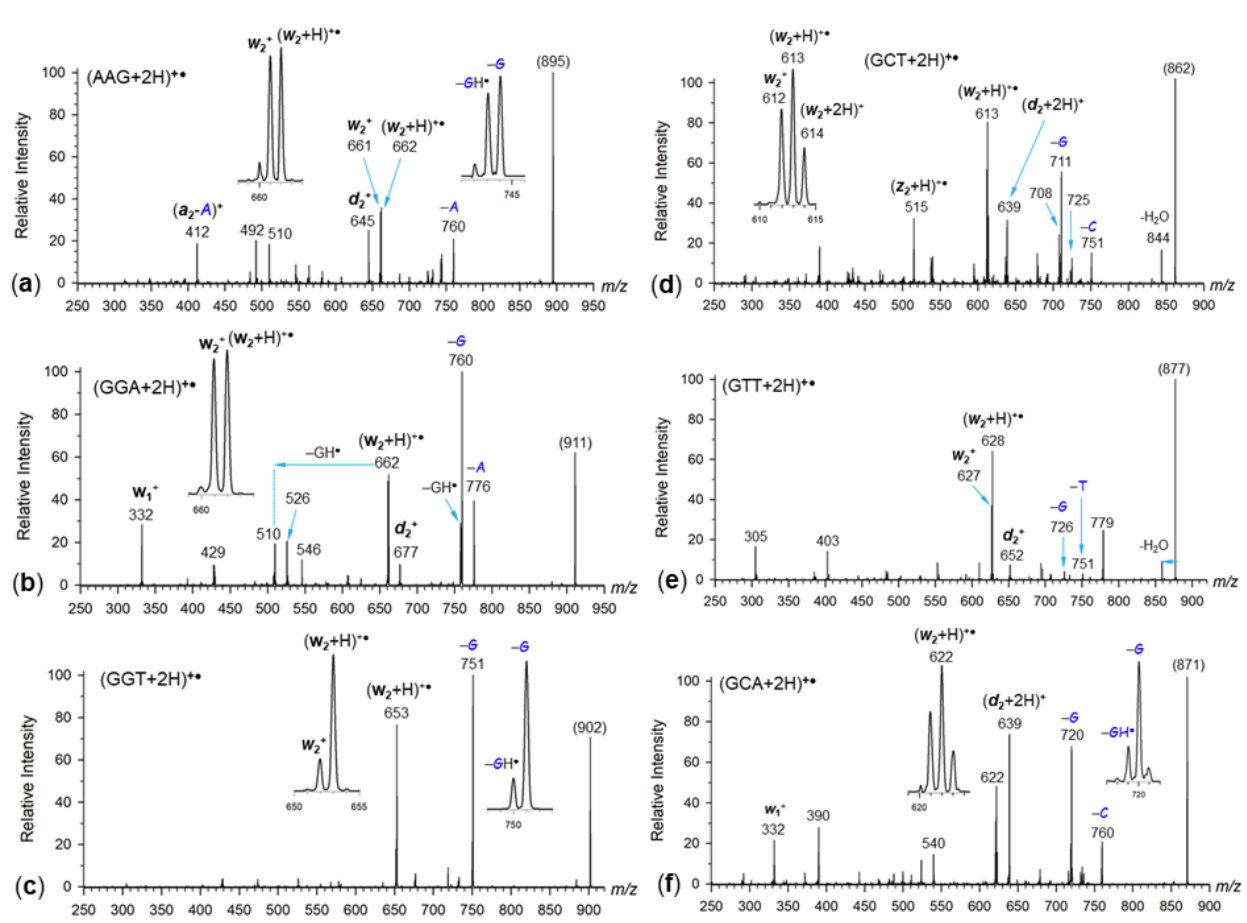


Figure 3.8: CID-MS^3 spectra of the $(w_2+H)^{+\bullet}$ group of cation radicals. (a) $(AAG+2H)^{+\bullet}$, (b) $(GGA+2H)^{+\bullet}$, (c) $(GGT+2H)^{+\bullet}$, (d) $(GCT+2H)^{+\bullet}$, (e) $(GTT+2H)^{+\bullet}$, and (f) $(GCA+2H)^{+\bullet}$.

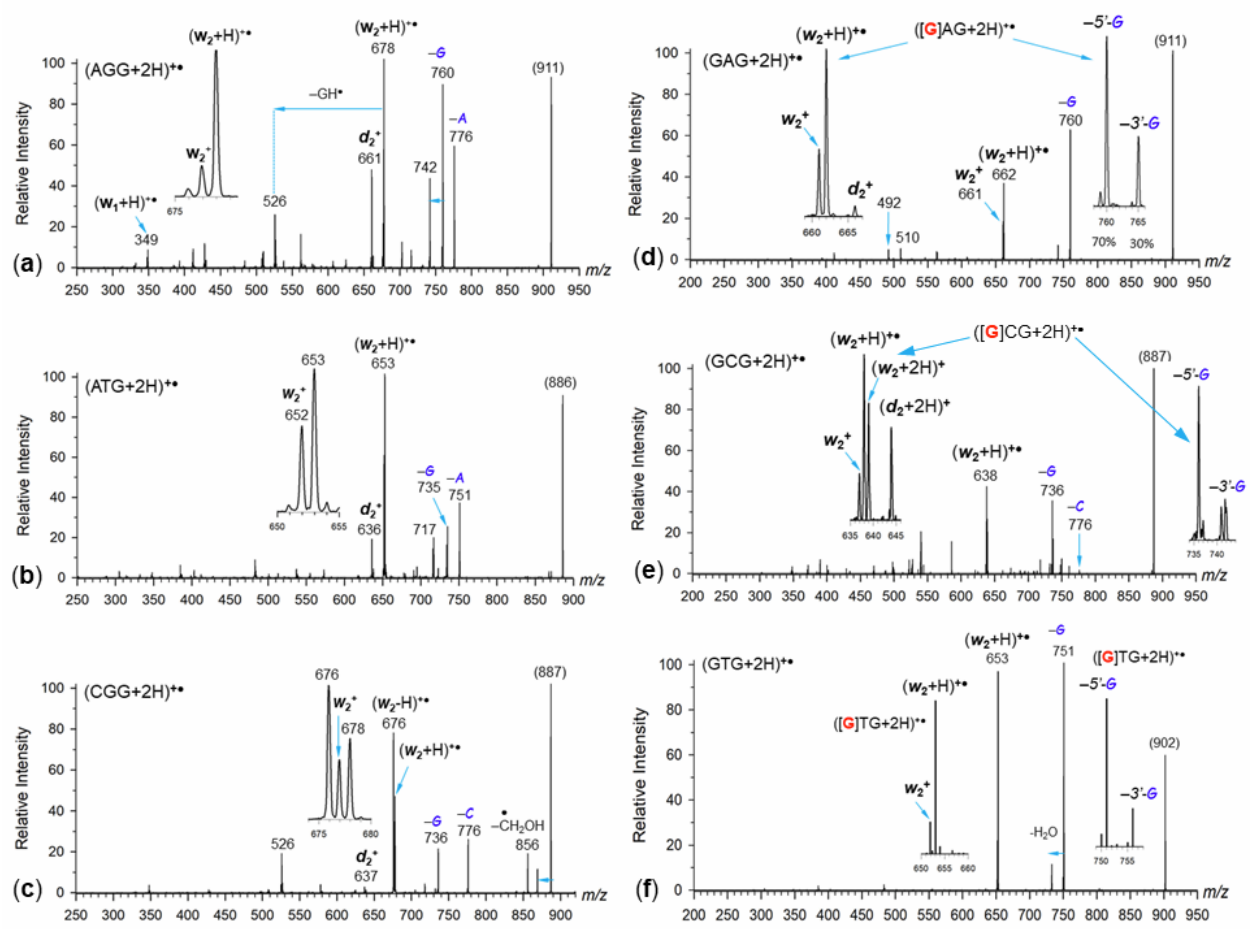


Figure 3.9: CID-MS^3 spectra of the $(\boldsymbol{w}_2+H)^{+\bullet}$ group of cation radicals. (a) $(AGG+2H)^{+\bullet}$, (b) $(ATG+2H)^{+\bullet}$, (c) $(CGG+2H)^{+\bullet}$, (d) $(GAG+2H)^{+\bullet}$ with insets showing the fragment ion distribution from $([^{15}N_5\text{-}G]AG+2H)^{+\bullet}$ (e) $(GCG+2H)^{+\bullet}$ with insets showing the fragment ion distribution from $([^{15}N_5\text{-}G]CG+2H)^{+\bullet}$, (f) $(GTG+2H)^{+\bullet}$ with insets showing the fragment ion distribution from $([^{15}N_5\text{-}G]TG+2H)^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

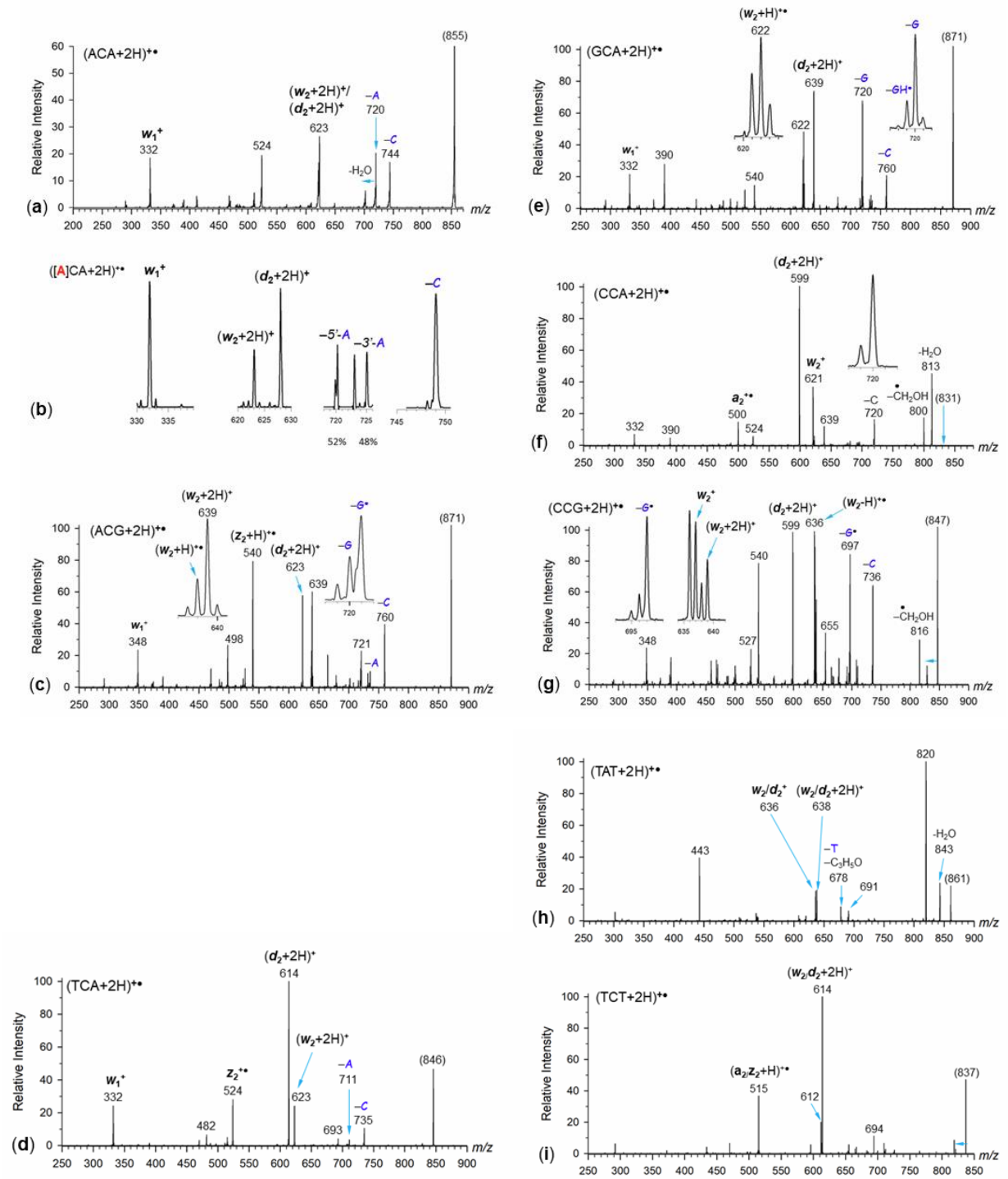


Figure 3.10: CID-MS^3 spectra of the $(\boldsymbol{d}_2+2H)^+$ group of cation radicals. (a) $(ACA+2H)^{+\bullet}$, (b) fragment ion distribution from $([^{15}N_5\text{-}A]CA+2H)^{+\bullet}$, (c) $(ACG+2H)^{+\bullet}$, (d) $(TCA+2H)^{+\bullet}$, (e) $(GCA+2H)^{+\bullet}$, (f) $(CCA+2H)^{+\bullet}$, (g) $(CCG+2H)^{+\bullet}$, (h) $(TAT+2H)^{+\bullet}$, and (i) $(TCT+2H)^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

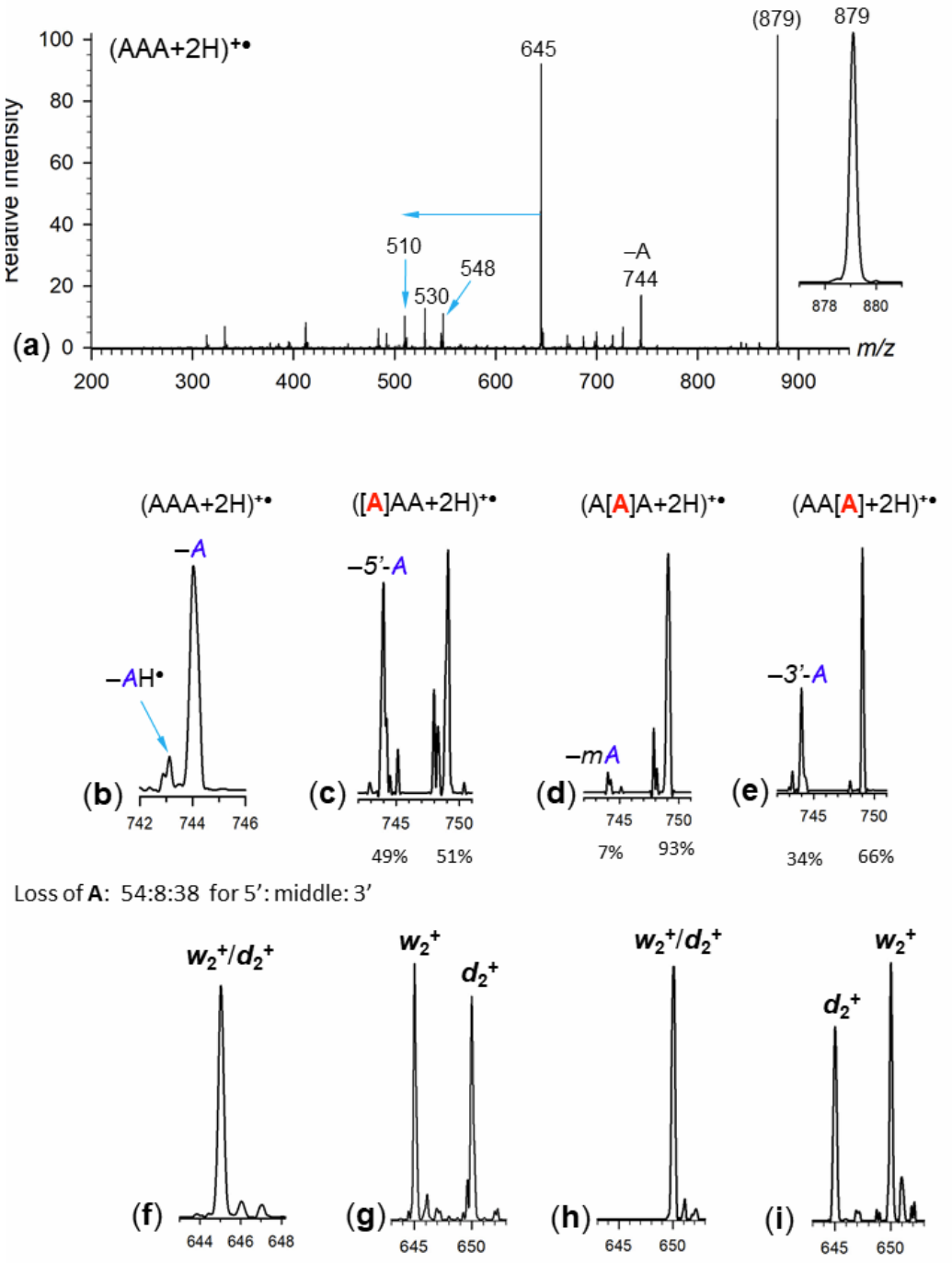


Figure 3.11: Top panel: (a) Reference CID-MS^3 spectrum of $(AAA+2H)^{+\bullet}$. Middle panel: Loss of adenine from (b) $(AAA+2H)^{+\bullet}$, (c) $([^{15}N_5\text{-}A]AA+2H)^{+\bullet}$, (d) $(A[^{15}N_5\text{-}A]A+2H)^{+\bullet}$, and (e) $(AA[^{15}N_5\text{-}A]+2H)^{+\bullet}$. Bottom panel: sequence ions from (f) $(AAA+2H)^{+\bullet}$, (g) $([^{15}N_5\text{-}A]AA+2H)^{+\bullet}$, (h) $(A[^{15}N_5\text{-}A]A+2H)^{+\bullet}$, and (i) $(AA[^{15}N_5\text{-}A]+2H)^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

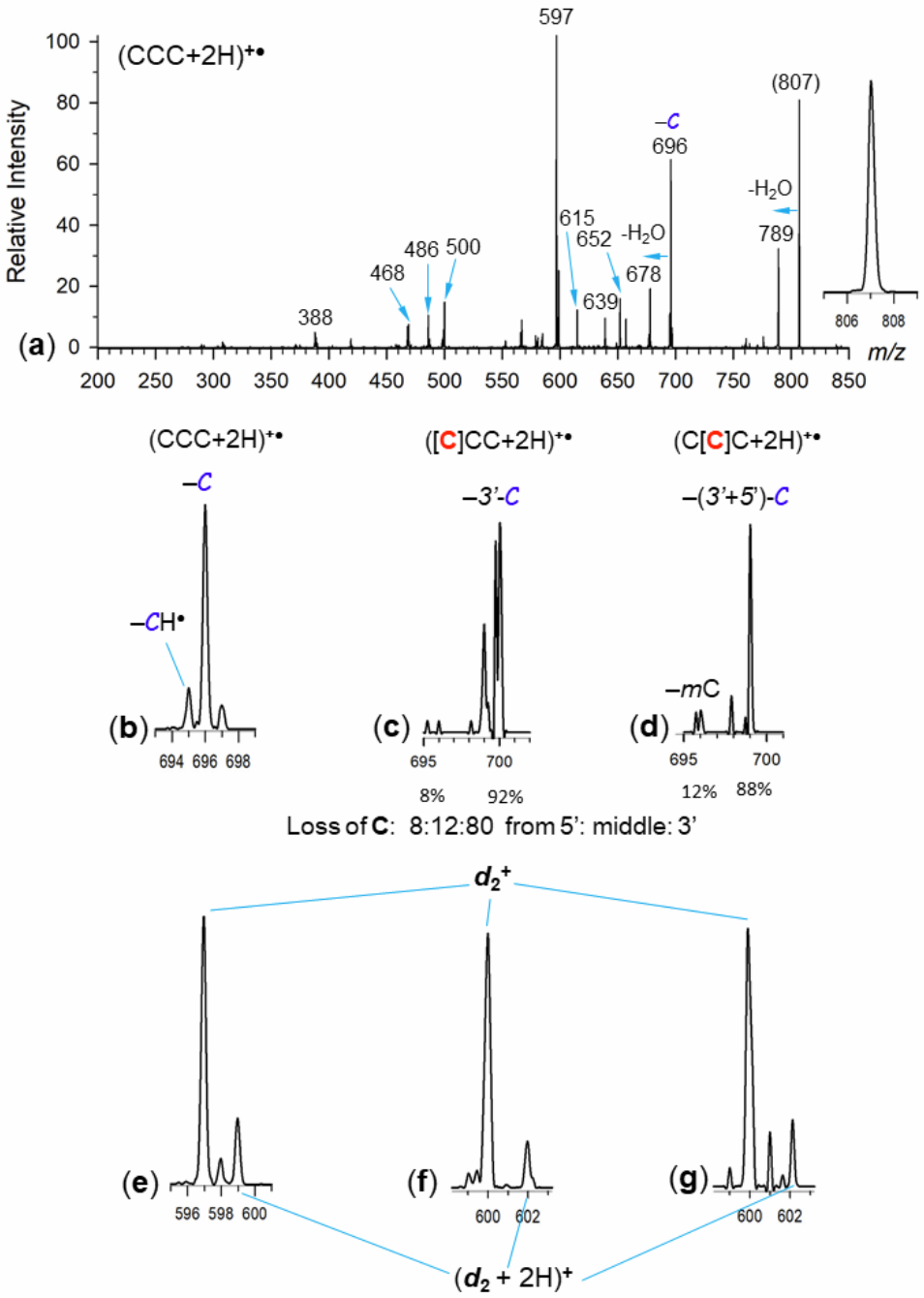


Figure 3.12. Top panel: (a) Reference CID-MS^3 spectrum of $(CCC+2H)^{+\bullet}$. Inset shows the precursor ion peak profile. Middle panel: Loss of cytosine from (b) $(CCC+2H)^{+\bullet}$, (c) $([^{15}N_3\text{-}C]CC+2H)^{+\bullet}$, and (d) $(C[^{15}N_3\text{-}C]C+2H)^{+\bullet}$. Bottom panel: sequence ions from (e) $(GGG+2H)^{+\bullet}$, (f) $([^{15}N_3\text{-}C]CC+2H)^{+\bullet}$, and (g) $(C[^{15}N_3\text{-}C]C+2H)^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

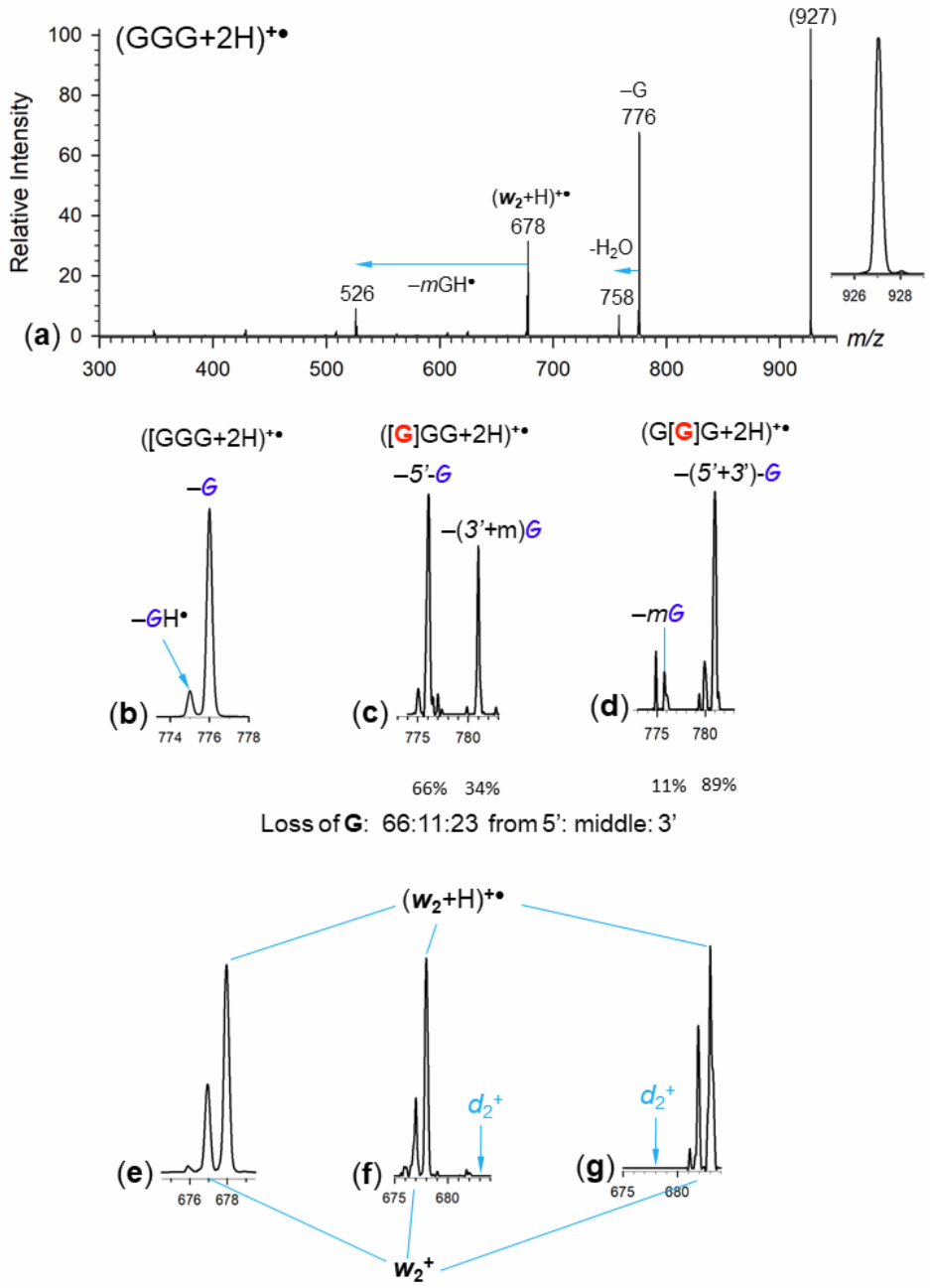
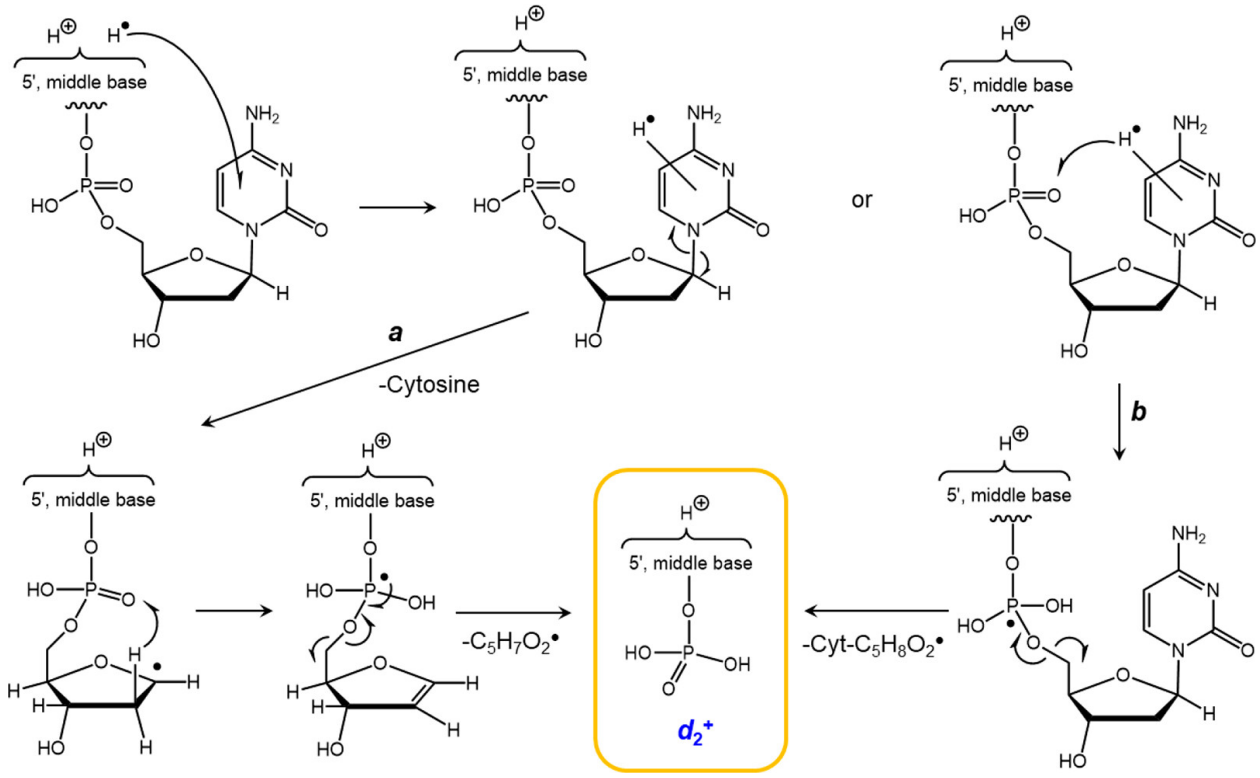


Figure 3.13: Top panel: (a) Reference CID-MS^3 spectrum of $(GGG+2H)^{+\bullet}$. Inset shows the precursor ion peak profile. Middle panel: Loss of guanine from (b) $(GGG+2H)^{+\bullet}$, (c) $([^{15}N_5\text{-}G]GG+2H)^{+\bullet}$, and (d) $(G[^{15}N_5\text{-}G]G+2H)^{+\bullet}$. Bottom panel: sequence ions from (e) $(GGG+2H)^{+\bullet}$, (f) $([^{15}N_5\text{-}G]GG+2H)^{+\bullet}$, and (g) $(G[^{15}N_5\text{-}G]G+2H)^{+\bullet}$. The ^{15}N-labeled nucleobases are shown in red in square brackets.

3.3.4 The d_2^+ Ion Group

The largest group of trinucleotide cation radicals displayed CID-MS^3 spectra with dominant d_2^+ backbone fragment ions, arising by a combined loss of the 3′-nucleobase and a $C_5H_7O_2$ radical

from the 3′-deoxyribose moiety while the charging proton was retained at the 5′- or middle nucleotides. Codon cation radicals undergoing major dissociation to $\boldsymbol{d_2}^+$ ions included the following sequences: AAC, ACC, AGC, ATC, CAC, CGC (**Figure 3.2a–f**), AAT, AGA, [$^{15}N_5$]AGA, ATT, ATA, [$^{15}N_5$]ATA, (**Figure 3.3a–f**), GCC, CTC, AGT, GGC, GAC, GTC (**Figure 3.4a–f**), and CCC, [$^{15}N_3$]CCC, TCC, TAA, TAC, and TGC (**Figure 3.5a–g**). The majority of these sequences shared two structural features. (1) There was a basic (A or G) nucleobase in the 5′-, middle or both positions to sequester the charging proton in the $\boldsymbol{d_2}^+$ ion. (2) The 3′-base was a pyrimidine, C or T, with the exception of AGA, ATA, and TAA. We note that these last three cation radicals also underwent other backbone cleavages; thus, the preference for the $\boldsymbol{d_2}^+$ ion formation was not so prevalent. The formation of $\boldsymbol{d_2}^+$ ions can be explained by the reaction sequence generically shown in **Scheme 3.1**. One-electron transfer to the precursor dication resulted in the formation of a radical at the 3′-nucleobase. This can occur by direct reduction of a protonated 3′-cytosine, or by hydrogen atom transfer from the reduced 5′- or middle nucleobase onto 3′-cytosine or thymine. We note that pyrimidine nucleobases in nucleosides have greater hydrogen atom affinities than purine nucleobases, as studied for nucleobase[69,70] and nucleoside radical conjugates.[61,62] For example, addition of a hydrogen atom at the N3 position in cytosine and 1-methylcytosine has been calculated to be exergonic by 130 kJ mol^{-1} to produce the lowest-energy radical adducts.[69] By comparison, addition of a hydrogen atom to the N1 and C8 positions in adenine was exergonic by 53 and 105 kJ mol^{-1}, respectively.[70] Thus, a hydrogen atom transfer from a purine nucleobase radical to a neutral pyrimidine nucleobase was expected to be exergonic. Hydrogen atom transfer must be the sole mechanism to generate thymine radicals, because the very low thymine basicity in nucleotides prevents it from direct protonation.[48,71] The 3′-nucleobase radical was expected to dissociate by a base loss, forming a C1′-deoxyribose radical (**Scheme 3.1**). Previous studies of nucleoside radicals have indicated that loss of the base was a major dissociation.[61,62] These loss-of-base intermediates were seen in the CID-MS^3 spectra of the cation radicals (**Figures 3.2–3.5**). The further dissociation could be envisioned as a standard phosphate ester elimination that has been considered for dissociations of oligonucleotide cations.[47,48,68,72] However, this interpretation could not account for the facile formation of $\boldsymbol{d_2}^+$ ions from the cation radicals in light of the fact that trinucleotide even-electron mono- and dications do not form $\boldsymbol{d_2}^+$ ions at all.[47,48]

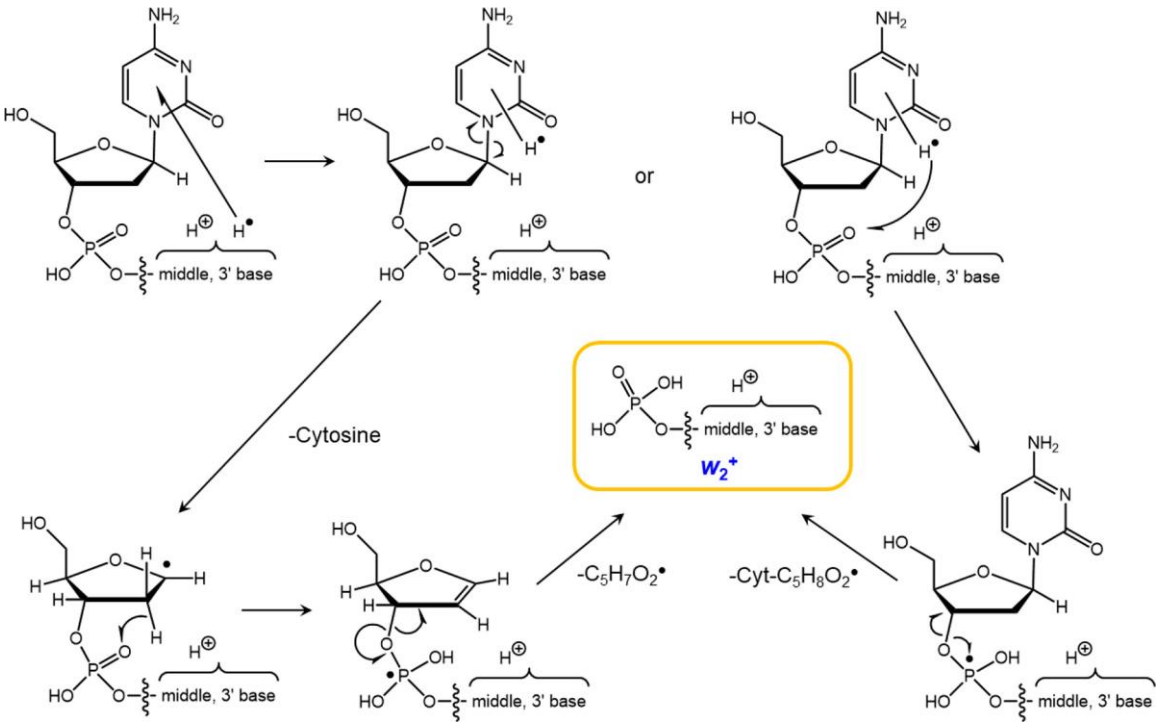


Scheme 3.1: Generic Mechanisms for the Formation of d_2^+ Sequence Fragment Ions

We propose two different mechanisms that take into account the radical nature of the intermediates (**Scheme 3.1**). The loss-of-base pathway (a) presumes hydrogen atom transfer from the deoxyribose radical onto the phosphate group, followed by homolytic cleavage of the C5′–O bond and loss of the $C_5H_7O_2$ radical, forming the d_2^+ ion. The alternative direct pathway (b) includes a hydrogen atom transfer from the nucleobase radical to the phosphate group followed by homolytic cleavage of the C5′–O bond. The phosphate hydrogen atom affinity (37 kJ mol^{-1})[73] indicated that the hydrogen atom migration to the phosphate group should be endergonic and needs to be driven by collisional excitation. These dissociation pathways are more specifically depicted later for $(CCC + 2H)^{+\bullet}$, including fully optimized structures and M06-2X/6-311++G(2d,p) + ZPVE relative and transition-state energies.

3.3.5 The w_2^+ Ion Group

Cation radicals derived from CAA, CTA, CAG, CCT, CAT, CGA (**Figure 3.6a–f**) and CTG, GAT, TTG, GAA, GTA, TGT (**Figure 3.7a–f**) showed major $\boldsymbol{w_2}^+$ fragment ions upon CID-MS^3. These dissociations were to some extent accompanied by other backbone cleavages: namely, formation of minor $\boldsymbol{d_2}^+$ ions (GAA, GAT, GTA) and hydrogen transfers leading to the formation of $(\boldsymbol{w_2}+H)^{+\bullet}$ ions which were observed for CAT, GAA, GTA, and GAT and $(\boldsymbol{w_2}\pm H)^{+\bullet}$ cation radicals seen for CGA and CTG within this group. The formation of $\boldsymbol{w_2}^+$ necessitates radical-induced fragmentation at the 5′-nucleoside involving a combined loss of the 5′-base and sugar $C_5H_7O_2$ radical and retention of the charging proton at the 3′- or middle nucleobase. The trinucleotides having 5′-cytosine satisfied the first criterion whereby formation of a 5′-cytosine radical was expected to trigger the loss of the base. Similarly, the 5′-guanine base in GAA, GAT, and GAT was likely to be protonated in the precursor dication and converted upon electron attachment into a 5′-guanine radical, thus facilitating 5′-nucleobase loss. Alternatively, $\boldsymbol{w_2}^+$ ions could be produced by a direct loss of an $\boldsymbol{a_1}$ neutral radical which was analogous to the formation of $\boldsymbol{d_2}^+$ ions (**Scheme 3.2**).



Scheme 3.2: Generic Mechanisms for the Formation of $\boldsymbol{w_2}^+$ Sequence Fragment Ions

3.3.6 The $(w_2+H)^{+\bullet}$ Ion Group

Another large group of trinucleotide cation radicals showed prominent formation of backbone $(w_2+H)^{+\bullet}$ fragment ions. The sequences undergoing this dissociation involved AAG, GGA, GGT, GCT, GTT, GCA (**Figure 3.8a–f**) and AGG, ATG, CGG, GAG, GCG, GTG (**Figure 3.9a–f**). A common feature of these sequences was that they all contained guanine and most also contained another basic purine nucleobase. GTT was an obvious exception, with the caveat that double protonation of this sequence was very inefficient, resulting in weak doubly charged ions and hence also a low-quality ETD spectrum for CID-MS^3. We note that thymine protonation in trinucleotides is energetically disfavored, so that the second proton in the $(GTT+2H)^{2+}$ precursor dication could be attached to one of the phosphate ester oxygens,[48] thus affecting the structure and chemistry of $(GTT+2H)^{+\bullet}$ following electron transfer. The formation of $(w_2+H)^{+\bullet}$ ions can be viewed as resulting from a combined loss of the 5′-nucleobase and 5′-deoxyribose in the form of a $C_5H_6O_2$ molecule that involved placing a proton and a hydrogen atom at the middle and 3′-bases in the fragment cation radical. This can be achieved by hydrogen atom or proton transfer from the 5′-nucleoside or proton and H atom retention at the middle and 3′-bases in the initial cation radical structure. A possible mechanism for the $(w_2+H)^{+\bullet}$ fragment ion formation will be shown later for $(GGG+2H)^{+\bullet}$ using fully optimized cation-radical structures and M06-2X/6-311++G(2d,p) + ZPVE energies.

3.3.7 The $(w_2+2H)^+$ and $(d_2+2H)^+$ Ion Group

Double hydrogen/proton transfer was observed for dissociations of several trinucleotides. CID-MS^3 of $(ACG+2H)^{+\bullet}$, $(ACT+2H)^{+\bullet}$ (**Figure 3.14a,b**), $(TCA+2H)^{+\bullet}$, and $(TCT+2H)^{+\bullet}$, showed prominent $(w_2+2H)^+$ fragment ions (**Figure 3.10**), which retained the charging proton and gained additional two hydrogen atoms after loss of the 5′-nucleobase and a $C_5H_5O_2$ radical from 5′-deoxyribose. A common structural feature among the trinucleotide cation radicals undergoing this dissociation was the presence of a purine (C or T) nucleobase in the middle position. The $(w_2+2H)^+$ ions were accompanied by other specific backbone cleavages. For example, $(ACG+2H)^{+\bullet}$ and $(ACT+2H)^{+\bullet}$ showed prominent odd-electron $(z_2+H)^{+\bullet}$ fragment ions indicating cleavage of the phosphate O–C5′ bond at the middle nucleotide (**Figure 3.14a,b**). Alternatively, the $(z_2+H)^{+\bullet}$ ions can be formed by phosphoric acid elimination (H_3PO_4, Δm = 98 Da) from the respective $(w_2+H)^{+\bullet}$

ions which also appeared in the CID-MS^3 spectra of these trinucleotide cation radicals (**Figure 3.14**). CID-MS^3 of another group of trinucleotide cation radicals, (ACA+2H)$^{+\bullet}$, (TCA+2H)$^{+\bullet}$, (GCA+2H)$^{+\bullet}$, (CCA+2H)$^{+\bullet}$, (CCG+2H)$^{+\bullet}$, and (TAT+2H)$^{+\bullet}$, showed the formation of $(\boldsymbol{d_2}+2H)^+$ fragment ions by a combined loss of the 3′-base and a $C_5H_5O_2$ radical from 3′-deoxyribose (**Figure 3.10**). The formation of the $(\boldsymbol{d_2}+2H)^+$ ions coincided with that of $(\boldsymbol{w_2}+2H)^+$, as seen for ACA, GCA, ACG, and TCA, but not CCA, which conformed $\boldsymbol{w_2}^+$ ions instead.

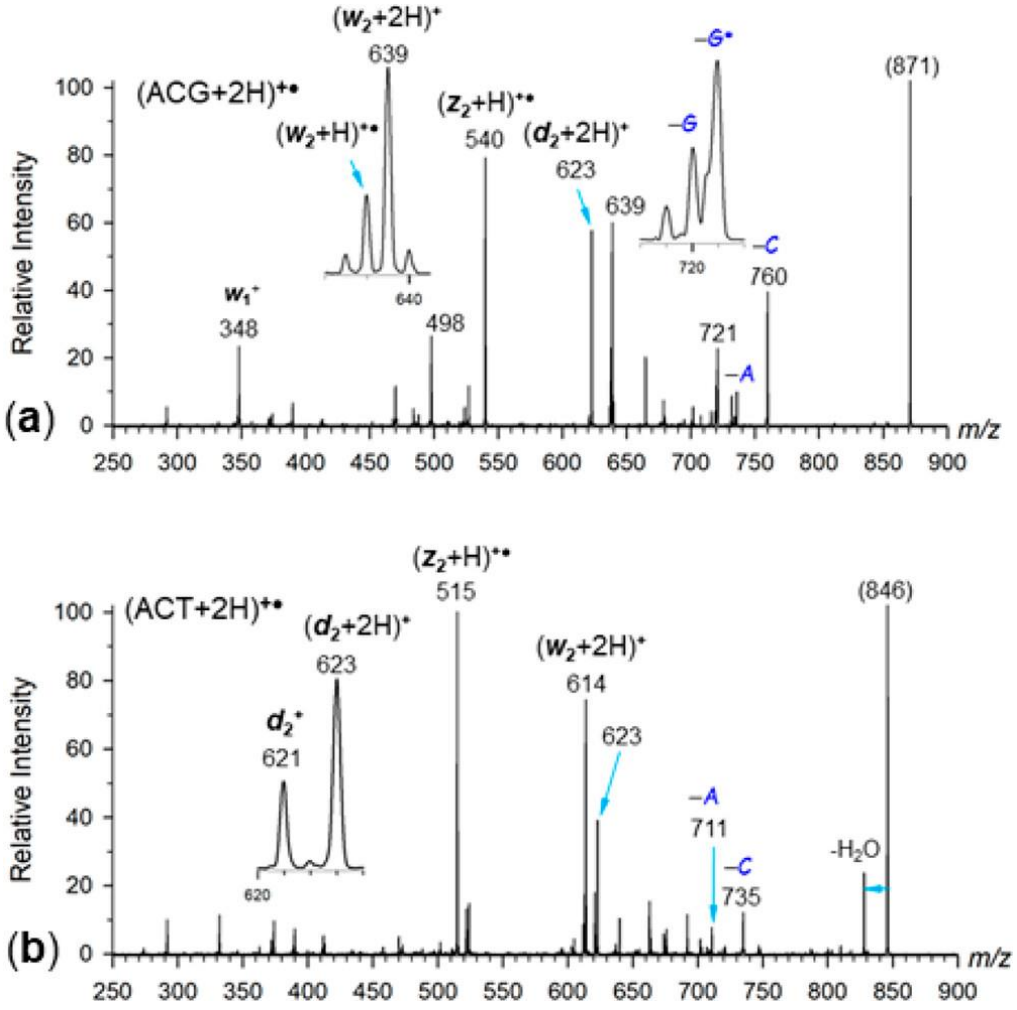
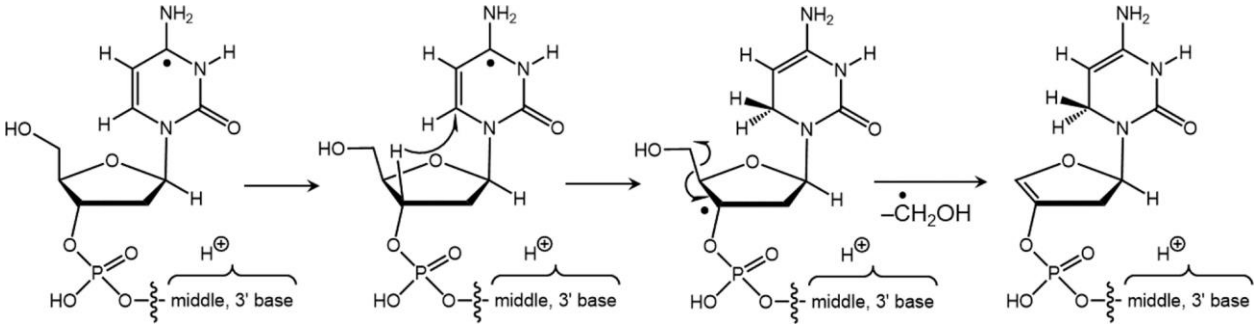


Figure 3.14 CID-MS^3 spectra of (a) (ACG+2H)$^{+\bullet}$ and (b) ACT+2H)$^{+\bullet}$. Insets show the fragment peak profiles.

In addition, the $(\boldsymbol{d_2}+2H)^+$ ions were accompanied by $\boldsymbol{d_2}^+$ but not $(\boldsymbol{d_2}+H)^{+\bullet}$, indicating that the phosphate cleavage between the middle and 3′-nucleoside was always accompanied by loss of a 3′-deoxyribose radical. The consistent feature of the $(\boldsymbol{d_2}+2H)^+$-forming trinucleotides was the presence of a cytosine residue as the middle base and loss of a purine (A or G) 3′-nucleobase.

3.3.8 Cross-Ring Dissociations

In addition to undergoing backbone phosphate cleavages, trinucleotide cation radicals having 3′- and 5′- terminal cytosine or thymine showed radical-induced dissociations in the pertinent terminal deoxyribose moieties (**Figure 3.15a–d**). Loss of the CH_2OH radical was observed for trinucleotides having 5′-cytosine, as in $(CCT+2H)^{+\bullet}$, $(CCA+2H)^{+\bullet}$, $(CAA+2H)^{+\bullet}$, $(CAT+2H)^{+\bullet}$, $(CTA+2H)^{+\bullet}$, $(CTG+2H)^{+\bullet}$, and $(CGA+2H)^{+\bullet}$, but much less so or not at all in $(CCC+2H)^{+\bullet}$, $(CAC+2H)^{+\bullet}$, $(CGC+2H)^{+\bullet}$, and $(CTC+2H)^{+\bullet}$. The latter group of cation radicals was found to produce abundant $\boldsymbol{d_2}^+$ ions which presumably outcompeted dissociations in the 5′-deoxyribose ring. The loss of the CH_2OH radical is likely to be initiated by a hydrogen atom transfer within the 5′-deoxyribose ring to weaken the C4′–C5′ bond, as tentatively shown in **Scheme 3.3**.



Scheme 3.3: Generic Mechanism for the Loss of CH_2OH

The spectra of cation radicals undergoing CH_2OH radical loss also showed prominent fragment ions by the competitive loss of cytosine. This indicated that these dissociations were promoted by the formation of a 5′-cytosine hydrogen atom adduct upon electron transfer. Compared to the above group, trinucleotide cation radicals having 5′-thymine underwent less frequent loss of $^{\bullet}CH_2OH$, as illustrated with $(TAA+2H)^{+\bullet}$ (**Figure 3.5e**).

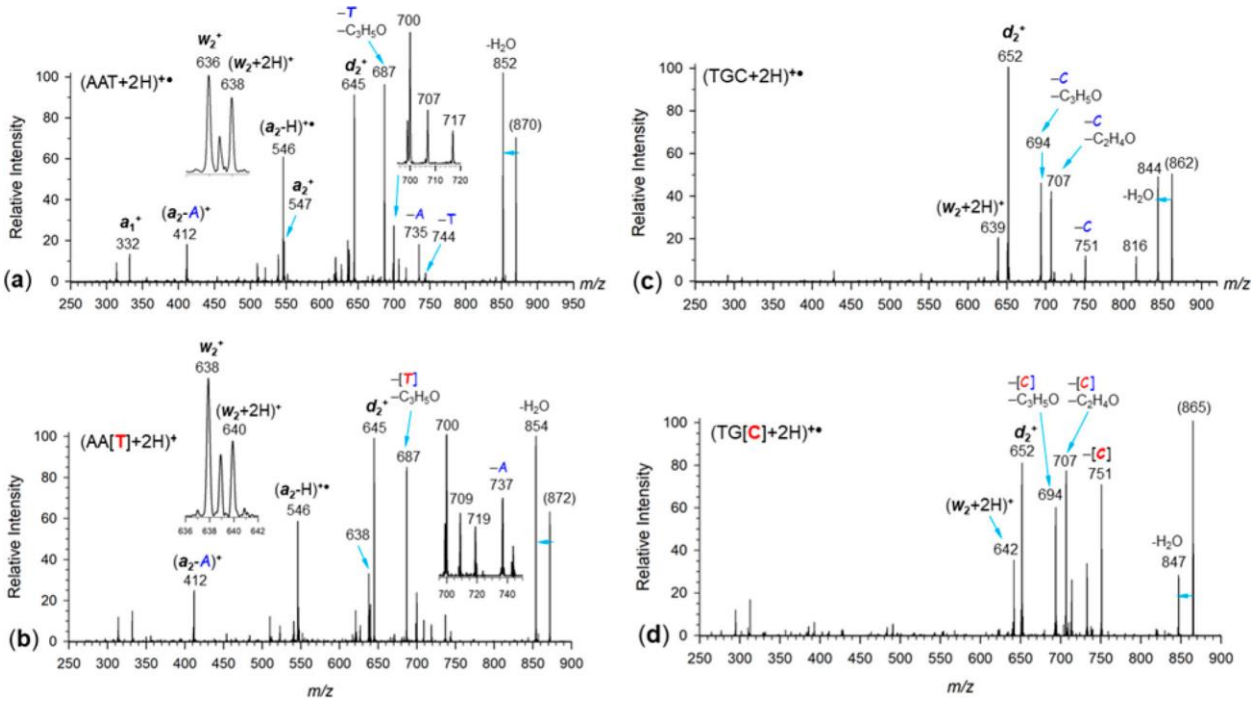
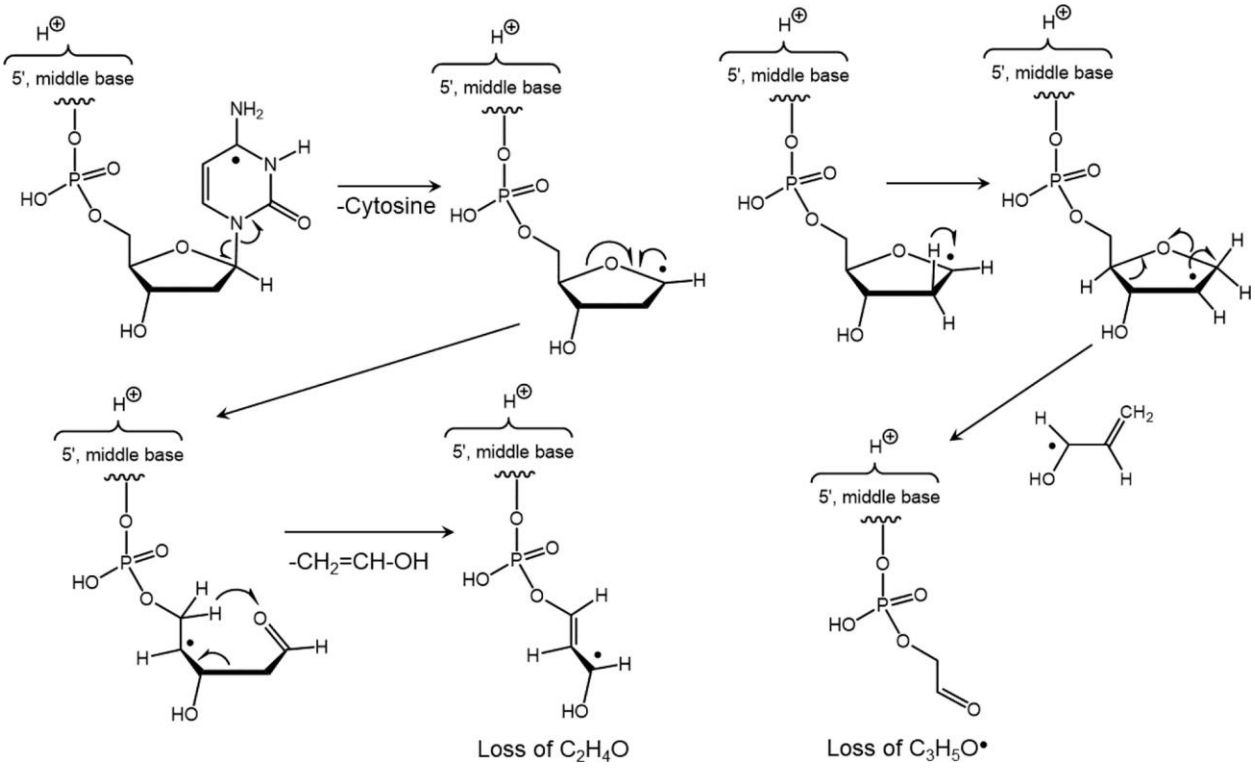


Figure 3.15: CID-MS^3 spectra of (a) $(AAT+2H)^{+\bullet}$, (b) $(AA[^{15}N_2\text{-}T]+2H)^{+\bullet}$, (c) $(TGC+2H)^{+\bullet}$, and (d) $(TG[^{15}N_3\text{–}C]+2H)^{+\bullet}$. Insets show the fragment ion peak profiles. The ^{15}N-labeled nucleobases are shown in red in brackets.

Cytosine and thymine residues at the 3′-terminus showed still different cross-ring dissociations in the 3′-deoxyribose moiety, resulting in the combined loss of the 3′-base and a C_2H_4O molecule or a $C_3H_5O^{\bullet}$ radical. This was observed for AAT, ATT, AGT, GAT, CAT, GGC, GAC, AGC, GTC, GCC, CGC, TGC, TGT, and TAC. The origin of these fragment ions was corroborated by ^{15}N labeling in AA[$^{15}N_2$-T] and TG[$^{15}N_3$–C] (**Figure 3.15b,d**). The spectrum of AA[$^{15}N_2$-T] showed retention of the m/z 687 fragment ion (**Figure 3.15b**) for the loss of ([$^{15}N_2$-T] + C_3H_5O) (**Figure 3.15b**). Similarly, TG[$^{15}N_3$–C] showed retention of m/z 707 and m/z 694 upon combined elimination of 3′-cytosine, C_2H_4O, and $C_3H_5O^{\bullet}$, respectively (**Figure 3.15d**). Because of the nature of these dissociations, it was plausible to propose a mechanism starting from a radical intermediate produced by the loss of the 3′-nucleobase, as shown for cytosine in **Scheme 3.4**. The C1′-radical at the 3′-sugar can undergo ring opening by cleavage of the C4′–O bond forming a C4′ radical. Hydrogen transfer to the newly formed aldehyde carbonyl is followed by cleavage of the C2′–C3′ bond and elimination of C_2H_4O as vinyl alcohol. The competing loss of $C_3H_5O^{\bullet}$ can be presented as starting by hydrogen 1,2-migration bringing the radical at C2′ and facilitating homolytic cleavage of the C1′–O bond and forming an allylic C1′–C2′–C3′ system in the intermediate oxide

radical. This was be expected to undergo a facile dissociation of the C4′–C3′ bond, leading to the loss of CH_2═CH–CHOH radical. It should be noted that the dissociation leading to the loss of ($T+C_3H_5O$) from trinucleotide cation radicals containing 3′-thymine may proceed by a radical-induced deoxyribose ring cleavage directly in the trinucleotide cation radicals. This is indicated by the very low propensity for the loss of thymine hindering the formation of the pertinent deoxyribose radical intermediates.



Scheme 3.4: Generic Mechanisms for Cross-Ring Cleavages

3.3.9 Resolution of Identity in Dissociations of $(AAA+2H)^{+\bullet}$, $(CCC+2H)^{+\bullet}$, and $(GGG+2H)^{+\bullet}$

Loss of the nucleobase was the primary dissociation observed in the CID-ETD-MS^3 spectra of all three symmetrical codon cation radicals (**Figure 3.16a–c**). In addition to the loss of nucleobase molecules, the spectra also showed a minor loss of nucleobase radicals as neutral hydrogen adducts $AH^{\bullet}$, $CH^{\bullet}$, and $GH^{\bullet}$. The formation of these radicals must have involved hydrogen atom migration

to the nucleobase radical produced by electron transfer reduction of the protonated nucleobase. The positional identity for the nucleobase loss was resolved by ^{15}N labeling as shown in **Figures 3.11–3.13**. With $(AAA+2H)^{+\bullet}$, where we had a complete set of 5′-, middle, and 3′-$^{15}N_5$-labeled codons, the distribution was 54:8:38 for the loss of the nucleobase from the corresponding positions. This was similar to the loss of adenine from $(AAA+H)^+$ cations,[48] indicating that the presence of the adenine radical did not have a major effect on this base distribution. Loss of cytosine from $(CCC+2H)^{+\bullet}$ was resolved to show an 8:12:80 ratio for the loss of the 5′-, middle, and 3′-nucleobase. This was slightly more biased toward the loss of the 3′-nucleobase compared to the dissociations of the $(CCC+H)^+$ cations.[48] With $(GGG+2H)^{+\bullet}$ we observed a 66:11:23 ratio for the loss of the 5′-, middle, and 3′-guanine, which represented a broader distribution of positions compared to that for the dissociations of $(GGG+H)^+$ cations.[48] We note that $(TTT+2H)^{+\bullet}$ cation radicals could not be studied because the formation of the $(TTT+2H)^{2+}$ precursor dications was prevented by the low basicity of thymine.[48]

Major differences, depending on the type of the nucleobase, were observed for backbone dissociations where the mass identity of the $\boldsymbol{w_2}^+$ and $\boldsymbol{d_2}^+$ fragment ions was resolved by ^{15}N labeling (**Figures 3.11–3.13**). Backbone cleavage in $(AAA+2H)^{+\bullet}$ was nonspecific, showing a 55:45 ratio for the formation of the $\boldsymbol{w_2}^+$ and $\boldsymbol{d_2}^+$ fragment ions that both appeared at m/z 645 in the **Figure 3.16a** spectrum but were resolved by ^{15}N labeling (**Figure 3.11**). These were both even-electron ions, indicating that they were formed by elimination of neutral radicals. In contrast to $(AAA+2H)^{+\bullet}$, the backbone dissociations of $(CCC+2H)^{+\bullet}$ unequivocally formed even-electron $\boldsymbol{d_2}^+$ ions at m/z 597 (**Figure 3.16b**). This can be realized by one-step elimination of the complementary $\boldsymbol{z_1}^\bullet$ radicals from $(CCC+2H)^{+\bullet}$ or by consecutive loss of 3′-cytosine, followed by elimination of a $C_5H_7O_2$ radical from the 3′-deoxyribose residue. The favored loss of 3′-cytosine (**Figure 3.12**) appeared to be consistent with the latter two-step mechanism. The spectrum also showed hydrogen migrations resulting in the formation of $(\boldsymbol{d_2}+H)^{+\bullet}$ and $(\boldsymbol{d_2}+2H)^+$ ions at m/z 598 and 599, respectively (**Figure 3.16b**). The $\boldsymbol{d_2}^+$ and $(\boldsymbol{d_2}+2H)^+$ ions were identified by ^{15}N-labeling (**Figure 3.12**), which showed that the isomeric $\boldsymbol{w_2}^+$ ions were absent. A specific feature of cation radicals containing 5′-cytosine was a direct elimination of water, as also observed for other codons of the $(CXY+2H)^{+\bullet}$ type. A still different pattern was found for the CID-MS^3 spectra of $(GGG+2H)^{+\bullet}$, which showed odd-electron $(\boldsymbol{w_2}+H)^{+\bullet}$ fragment ions (m/z 678) that were accompanied by their even-electron $\boldsymbol{w_2}^+$ counterparts at m/z 677 (**Figure 3.16c**). The spectra of the ^{15}N-analogues

(**Figure 3.13**) confirmed that the formation of the d_2^+ ions from $(GGG+2H)^{+\bullet}$, to be expected at m/z 683 from the 5′-labeled GGG, was negligibly small. The formation of the $(w_2+H)^{+\bullet}$ ions indicated retention of the added hydrogen atom in one of the remaining middle or 3′-guanine residues, which raised the question of the proton and hydrogen atom sites in the $(GGG+2H)^{+\bullet}$ ions and their possible migrations in the course of dissociation.

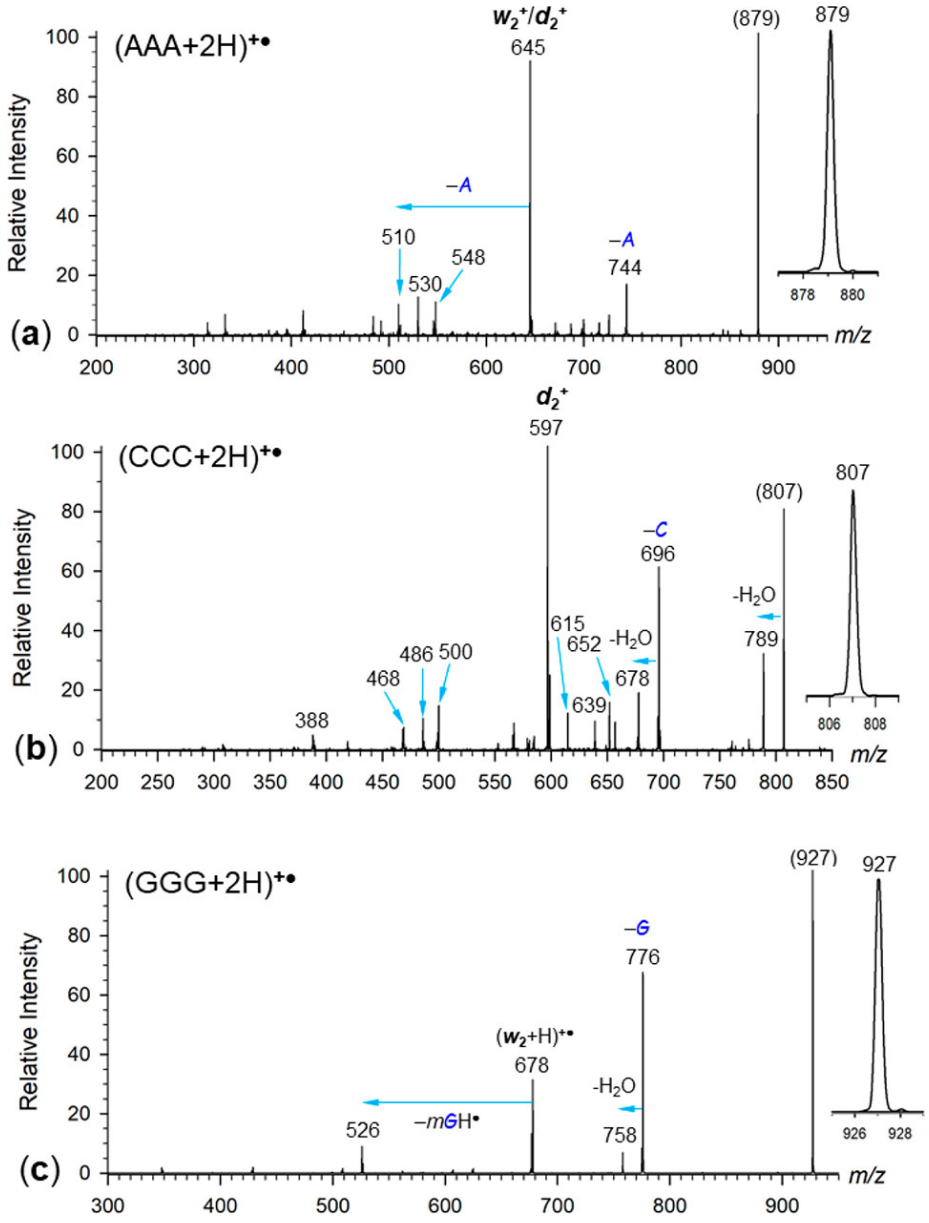


Figure 3.16: CID-MS^3 spectra of (a) $(AAA+2H)^{+\bullet}$ (m/z 879), (b) $(CCC+2H)^{+\bullet}$ (m/z 807), and (c) $(GGG+2H)^{+\bullet}$ (m/z 927) ions. Insets show the precursor ion peak profiles. See text for fragment ion assignment.

Overall, the differences in the dissociations of the $(AAA+2H)^{+\bullet}$, $(CCC+2H)^{+\bullet}$, and $(GGG+2H)^{+\bullet}$ cation radicals were quite remarkable, especially in comparison to backbone dissociations of their closed-shell mono- and dication analogues,[47,48] which unambiguously formed $\boldsymbol{w_2}^+$ backbone ions from most sequences. This pointed out the role of the radical sites in the cation radicals and their effect on steering the dissociations.

3.3.10 Resolution of Identity in Dissociations of Codons of the $(XYX+2H)^{+\bullet}$ Type

Having resolved the dissociations of base-homologous codon cation radicals, we systematically replaced one or more nucleobases to investigate the effect of the distinct base on the dissociations. Starting with the replacement of the middle base, the codon table provides 12 XYX combinations of the centrosymmetric type, for which we were able to generate a complete set of cation radicals and obtain their CID-MS^3 spectra. Mass identity in dissociations due to the 5′- and 3′-nucleobases was resolved by ^{15}N labeling in the 5′-nucleobase. The spectra of $(ACA+2H)^{+\bullet}$ (**Figure 3.10a,b**), $(AGA+2H)^{+\bullet}$ (**Figure 3.3b,c**), and $(ATA+2H)^{+\bullet}$ (**Figure 3.3e,f**) showed a nonspecific loss of 5′- and 3′-adenine that was accompanied by the formation of the $\boldsymbol{w_2}^+$ and $\boldsymbol{d_2}^+$ backbone fragment ions. These were resolved by ^{15}N labeling, showing a slight preference for the $\boldsymbol{d_2}^+$ ion formation (**Figure 3.3**). The spectra also showed prominent $\boldsymbol{w_1}^+$ ions at m/z 332 which, however, were not accompanied by abundant $\boldsymbol{d_1}^+$ counterparts.

The spectra of $(CAC+2H)^{+\bullet}$, $(CGC+2H)^{+\bullet}$ (**Figure 3.2e,f**), and $(CTC+2H)^{+\bullet}$ (**Figure 3.4b**) were qualitatively similar to that of $(CCC+2H)^{+\bullet}$, as evidenced by the facile formation of $\boldsymbol{d_2}^+$ backbone ions. However, the patterns for the loss of nucleobase substantially differed, depending on the middle nucleobase. For $(CAC+2H)^{+\bullet}$, we observed no fragment ions by loss of cytosine and very minor loss of adenine from the precursor cation radical. In contrast, loss of cytosine proceeded abundantly from the $\boldsymbol{d_2}^+$ ion (m/z 621) to produce the m/z 510 secondary ion. The spectrum of the $(CGC+2H)^{+\bullet}$ ion showed an unusual loss of guanine radical ($G^\bullet$, 150 Da, m/z 697) while the loss of cytosine was accompanied by an elimination of 44 and 57 Da neutral fragments, most likely C_2H_4O and $C_3H_5O^\bullet$, respectively, from cross-ring cleavage of the deoxyribose radical. The spectrum of the $(CTC+2H)^{+\bullet}$ ion (**Figure 3.4b**) showed a dominant loss of cytosine and also loss of a thymine radical ($T^\bullet$, 125 Da, m/z 697). These were all radical reactions that had no equivalents in closed-shell ion dissociations. The backbone cleavage yielding the $\boldsymbol{d_2}$ ion at m/z 612 was

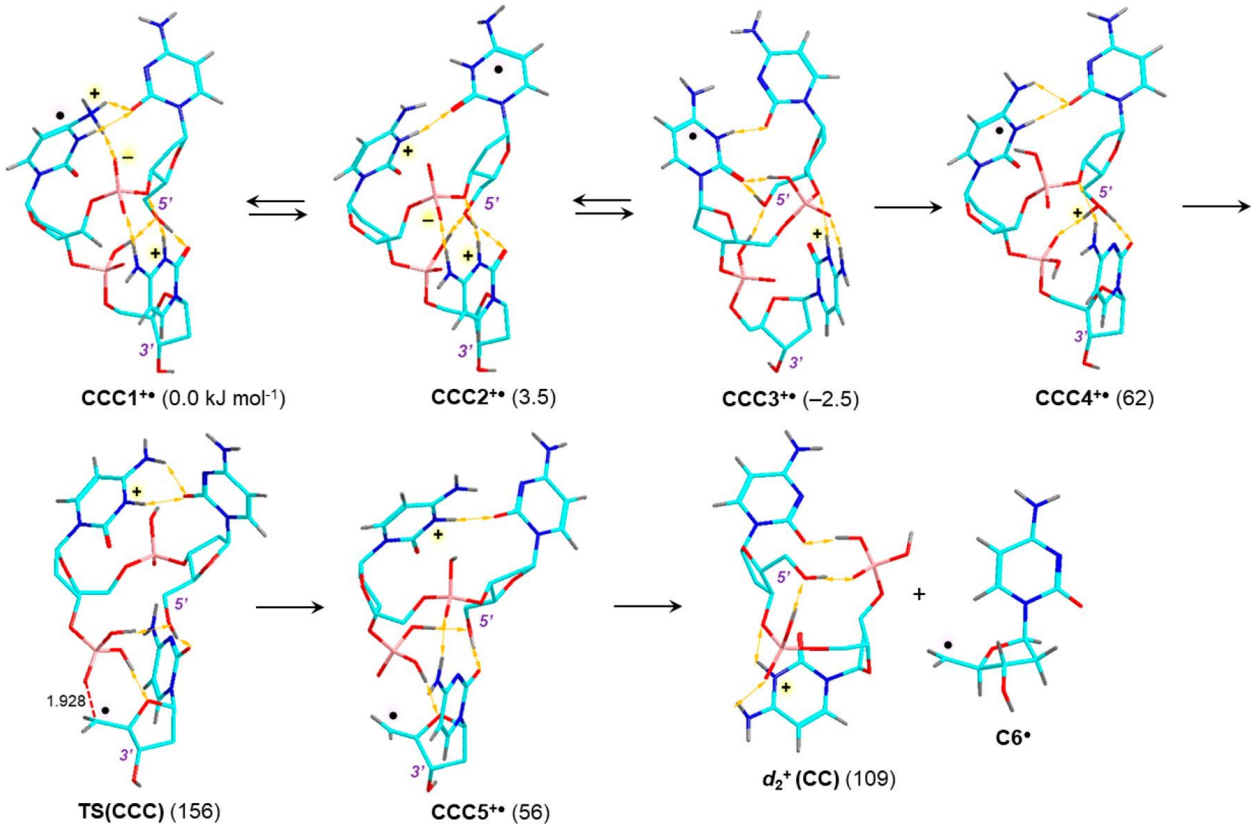
accompanied by hydrogen transfer forming the less usual $(\boldsymbol{d_2}–H)^{+\bullet}$ cation radical at m/z 611. Noteworthy was the direct loss of water from $(CTC+2H)^{+\bullet}$ (**Figure 3.4b**).

The spectra of $(GAG+2H)^{+\bullet}$, $(GCG+2H)^{+\bullet}$, and $(GTG+2H)^{+\bullet}$ (**Figure 3.9d–f**) showed the predominant loss of guanine that occurred chiefly from the 5′-position, as resolved by ^{15}N-labeling. The backbone dissociations involved hydrogen transfers to yield $(\boldsymbol{w_2}+H)^{+\bullet}$ fragment ions that were accompanied by the more common $\boldsymbol{w_2}^+$ ions. This pattern was similar to that observed for $(GGG+2H)^{+\bullet}$ and appeared to be driven by the loss of 5′-guanine. The dissociations of $(GCG+2H)^{+\bullet}$ were outstanding in that they resulted in the formation of even-electron $(\boldsymbol{d_2}+2H)^+$ ions (m/z 644) that must have involved a double hydrogen atom transfer from the complementary $\boldsymbol{z_1}$ fragment.

The spectra of $(TAT+2H)^{+\bullet}$, $(TCT+2H)^{+\bullet}$, and $(TGT+2H)^{+\bullet}$ (**Figures 3.10h,i** and **3.7f**) showed a great diversity in the types of dissociations, strongly depending on the central nucleobase. This effect was understandable, because the central nucleobase was substantially more basic than thymine and was likely to retain the charging proton. Unfortunately, double protonation of these trinucleotides was very inefficient due to the low basicity of thymine and did not allow us to obtain spectra of the ^{15}N-labeled derivatives. For this reason, the fragment ion assignment is only tentative. The spectrum of $(TAT+2H)^{+\bullet}$ (**Figure 3.10h**) showed a major fragment by loss of water (m/z 843) and an unusual ion at m/z 820. This could logically result by elimination of C_2HO by ring cleavage in one of the thymines. Backbone cleavage was observed as $\boldsymbol{w_2}^+/\boldsymbol{d_2}^+$ and $(\boldsymbol{w_2}+2H)^+/(\boldsymbol{d_2}+2H)^+$ even-electron fragment ions. The spectrum of $(TGT+2H)^{+\bullet}$ (**Figure 3.7f**) displayed a major fragment ion by loss of water (m/z 859) and a cluster of $\boldsymbol{w_2}^+/\boldsymbol{d_2}^+$ ions at m/z 652 including their hydrogen-transfer analogues at adjacent m/z values. Deoxyribose cross-ring fragments were present at m/z 707 and 694 that presumably were formed by loss of C_2H_4O and $C_3H_5O^{\bullet}$, respectively, following the loss of thymine (m/z 751). It may be noted that protonation of thymine in trinucleotides has been found to be energetically disfavored, as judged by $(TTT+H)^+$, in which low-energy structures were protonated at the phosphate ester groups instead in thymine.[48] This is expected to have a large effect on the structures of the $(TXT+2H)^{2+}$ precursor dications and $(TXT+2H)^{+\bullet}$ cation radicals.

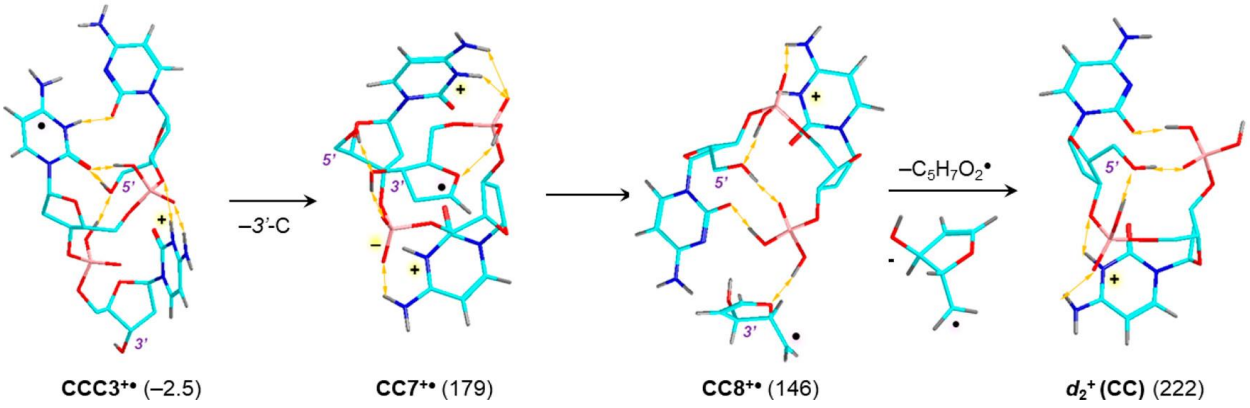
3.3.11 Ion Structures and Energetics

$(CCC+2H)^{+\bullet}$ was the typical representative of the ion group undergoing backbone dissociations to form $\boldsymbol{d_2}^+$ ions. From combined BOMD trajectory and DFT gradient optimization calculations of all theoretically possible protomers carrying protons at cytosine positions N3 and O2, we obtained equilibrium geometries of several isomers; the lowest Gibbs-energy structures **CCC1$^{+\bullet}$–CCC3$^{+\bullet}$** are shown in **Scheme 3.5**. These nearly isoenergetic $(CCC+2H)^{+\bullet}$ ions substantially differed in their electronic structure. **CCC1$^{+\bullet}$** was a zwitterion with a negative phosphate group between the 5′- and middle nucleosides. The 3′-base was protonated at N3, whereas the middle base carried a proton and a hydrogen atom that showed tight hydrogen bonds to the 5′-cytosine O2 and the phosphate anion. **CCC2$^{+\bullet}$** was another zwitterion with the middle and 3′-bases protonated and the 5′-base carrying the hydrogen atom. **CCC3$^{+\bullet}$** was a canonical structure in which the 3′-base was protonated and the middle base was a radical. The nearly isoenergetic nature of these structures, as well as their formation by electron and proton migrations upon BOMD, indicated that these isomerizations were facile, in part due to the interconnecting hydrogen bonds with the phosphate groups that relayed protons within the ions. The lowest energy pathway we found for the dissociation and formation of the $\boldsymbol{d_2}^+$ ion is shown in **Scheme 3.5**. This involved proton migration onto the lower phosphate ester group, where it was shared by a strong hydrogen bond with the 5′-hydroxyl. This intermediate (**CCC4$^{+\bullet}$**) was a local energy minimum at 62 kJ mol^{-1} relative to **CCC1$^{+\bullet}$**. Homolytic cleavage of the O5–C5′ bond was associated with proton and electron migration to reach **TS(CCC)** at 156 kJ mol^{-1} relative to **CCC1$^{+\bullet}$**, representing an unusually low energy barrier. Interestingly, the dissociation was accompanied by electron transfer and spin density migration from the middle cytosine in **CCC4$^{+\bullet}$** to the 3′-deoxyribose methylene group in **TS(CCC)** (**Scheme 3.5**). The dissociation further proceeded to a complex of the incipient $\boldsymbol{z_1}$ radical **C6$^{\bullet}$** and $\boldsymbol{d_2}^+$ (CC) ion that upon separating to the products achieved the 109 kJ mol^{-1} thermochemical threshold. We did not include in the product energy calculations corrections to the basis set superposition error that could further lower the threshold energy by a few kJ mol^{-1}. The low energy threshold and TS energy were consistent with the very facile formation of the $\boldsymbol{d_2}^+$ ions from $(CCC+2H)^{+\bullet}$. We presume that an analogous mechanism can be applied to the formation of $\boldsymbol{d_2}^+$ ions from the other trinucleotide cation radicals having 3′-cytosine.



Scheme 3.5: Calculated Structures and Relative Energies for d_2^+ Ion Formation from $(CCC+2H)^{+\bullet a}$
[a]Atom color coding is as follows: cyan, C; red, O; blue, N; gray, H; bronze, P. Only nucleobase and exchangeable hydrogen atoms are shown to avoid clutter. Relative energies are from M06-2X/6-311++G(2d,p) single-point calculations including zero-point vibrational corrections and referenced to 0 K. Radical locations were assigned from atomic spin densities calculated by a natural population analysis of the M06-2X/6-311++G(2d,p) wave functions.

The alternative pathway a, shown generically in **Scheme 3.1**, was calculated to be substantially less favorable (**Scheme 3.6**). Loss of the 3′-cytosine required 179 kJ mol^{-1} to form the zwitterionic intermediate **CC7**$^{+\bullet}$, which already was above the energy of **TS(CCC)** in pathway b. Homolytic cleavage of the O5–C5′ bond in **CC7**$^{+\bullet}$ led to a complex (**CC8**$^{+\bullet}$) at 146 kJ $mol^{-}1$ relative to **CCC1**$^{+\bullet}$, and further fragment separation required additional 76 kJ mol^{-1}, placing the thermochemical threshold at 222 kJ mol^{-1} relative to **CCC1**$^{+\bullet}$.

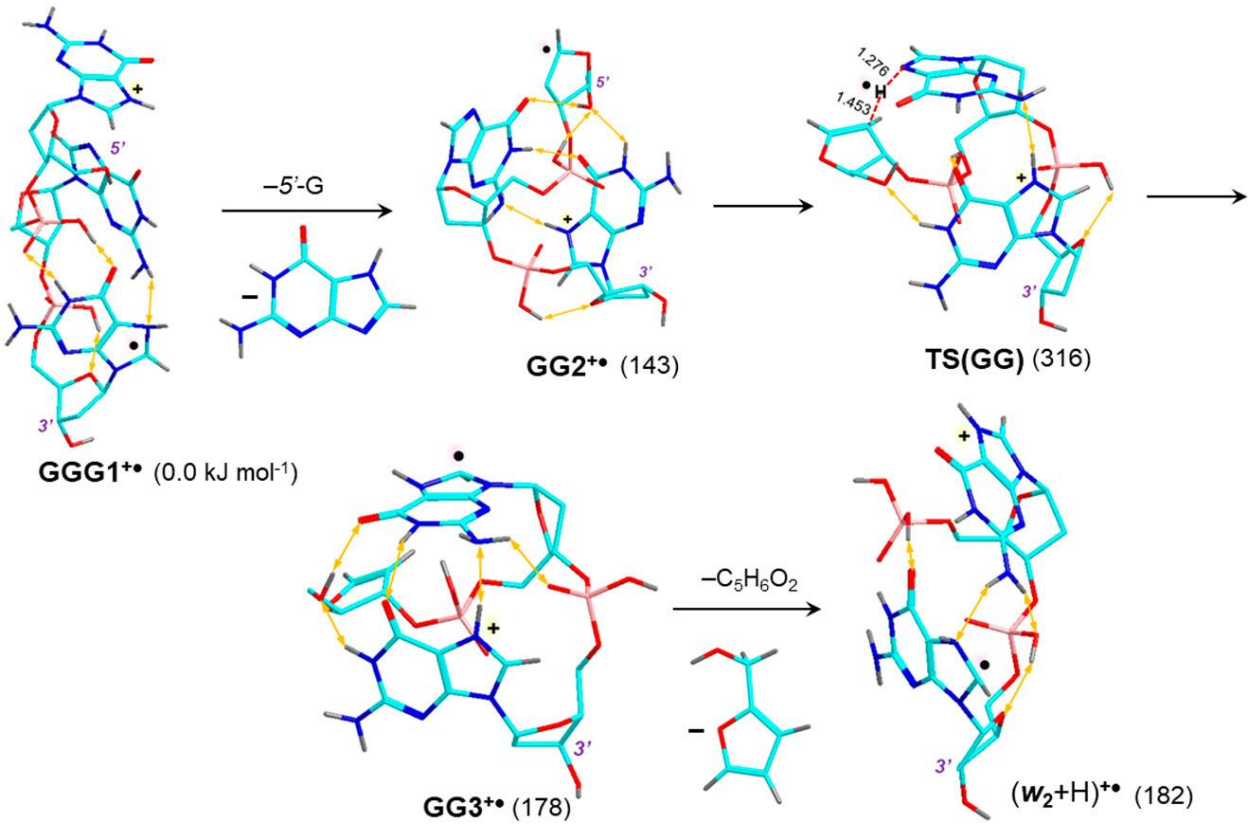


Scheme 3.6: Higher-Energy Pathway to the d_2^+ Ion Formation from $(CCC+2H)^{+\bullet}$ a

[a]Description as in Scheme 3.5.

To elucidate the formation of $(w_2+H)^{+\bullet}$ ions, we obtained optimized structures of low-energy $(GGG+2H)^{+\bullet}$ ions (**Scheme 3.7**). The calculations unequivocally pointed to structure **GGG1**$^{+\bullet}$ as the lowest-energy protomer in which the 5′-guanine was protonated at the most basic N7 position while the 3′-guanine carried the hydrogen radical. There were no zwitterionic isomers among the low-energy $(GGG+2H)^{+\bullet}$ structures. This was consistent with a previous computational study of $(GGG+H)^+$ ions that also found no low-energy zwitterions.[48] Loss of 5′-guanine as a less stable N7–H tautomer was calculated to have a thermochemical threshold at 143 kJ mol^{-1}, forming the ion **GG2**$^{+\bullet}$ as the lowest energy structure, which was a 5′-deoxyribose radical. The correlation of the ion electronic structures in the **GGG1**$^{+\bullet}$ → **GG2**$^{+\bullet}$ reaction step indicated that the reaction must be accompanied by electron transfer from the 3′-guanine radical to 5′-deoxyribose. The timing of this electron transfer was not studied; it may occur in a low-lying excited electronic state of **GGG1**$^{+\bullet}$, which we calculated to be 2.60 eV (251 kJ mol^{-1}) above the ground state and could be accessed by collisional excitation within the internal energy range needed for the dissociation (316 kJ mol^{-1}, **Scheme 3.7**). Alternatively, loss of 5′-guanine may initially proceed via an ionic mechanism by heterolysis of the C1′–N7 bond, and the product or intermediates can undergo electron transfer to reach the lowest energy structure **GG2**$^{+\bullet}$. In the next step, hydrogen transfer to the middle guanine required 316 kJ mol^{-1} in **TS(GG)** to produce an intermediate (**GG3**$^{+\bullet}$) at 178 kJ mol^{-1} relative to **GGG1**$^{+\bullet}$. Loss of furfuryl alcohol by phosphate ester elimination in **GG3**$^{+\bullet}$ was only 4 kJ mol^{-1} endergonic forming the $(w_2+H)^{+\bullet}$ ion, although this dissociation may involve an energy barrier. Overall, the pathway described in **Scheme 3.7** was compatible with the identity-resolved CID-MS^3 spectrum of $(GGG+2H)^{+\bullet}$. The spectrum showed abundant fragment ions by

loss of 5′-guanine, which was conducive to the formation of the $(\boldsymbol{w_2}+H)^{+\bullet}$ ion. Conversely, the $(\boldsymbol{w_2}+H)^{+\bullet}$ relative intensity in the spectrum was relatively low, which was consistent with the calculated high activation energy for **TS(GG)** in the hydrogen migration step, kinetically stabilizing the **GG2**$^{+\bullet}$ intermediate.

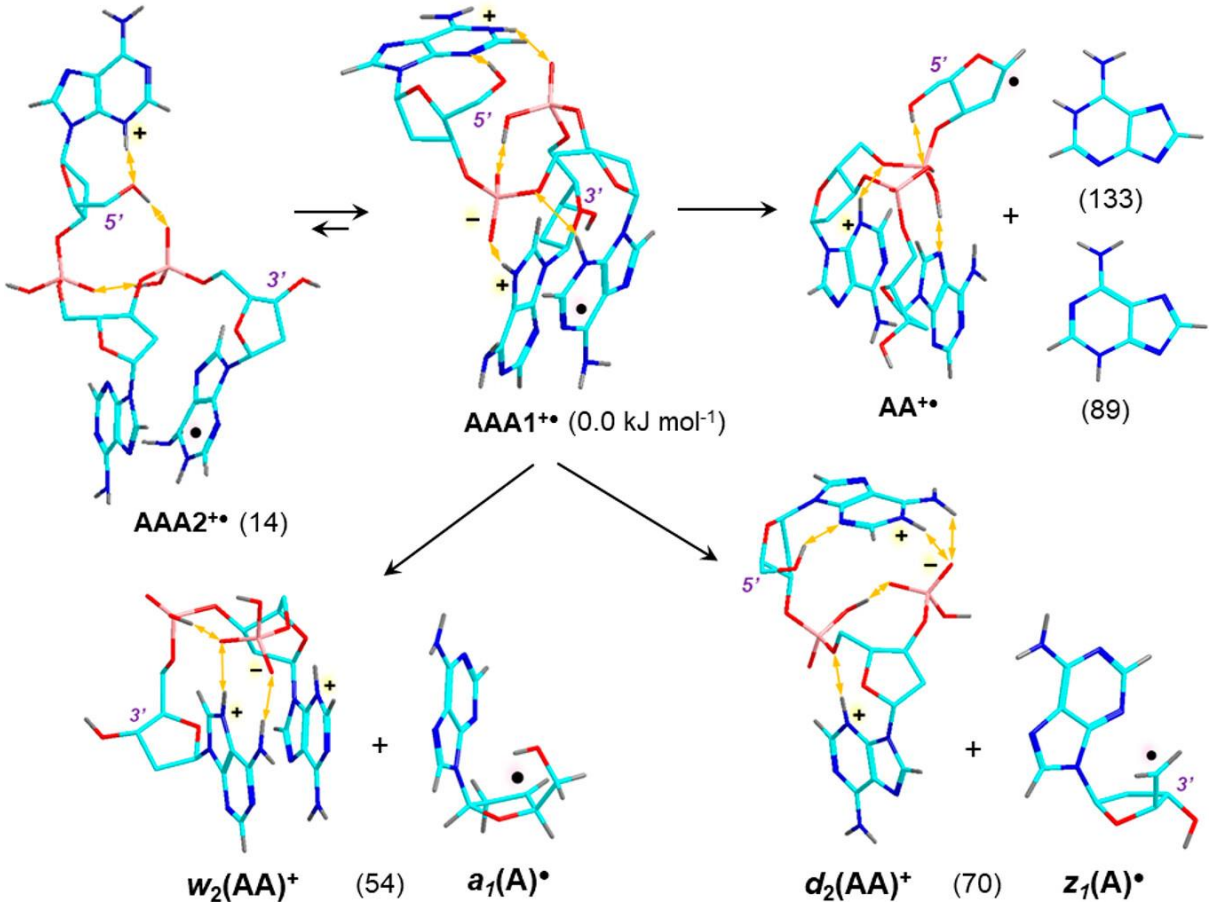


Scheme 3.7: Calculated Structures and Relative Energies of $(GGG+2H)^{+\bullet}$ [a]

[a]Description as in Scheme 3.5.

We suggest that the **Scheme 3.7** mechanism, considering an initial loss of the 5′-base, can be applied to explain dissociations of most trinucleotide cation radicals of the $(\boldsymbol{w_2}+H)^{+\bullet}$ group. **Figures 3.8** and **3.9** document that most of the ions in this group have 5′-guanine, whose loss initiates the subsequent hydrogen transfer and elimination of $C_5H_6O_2$. Formation of $(\boldsymbol{w_2}+H)^{+\bullet}$ ions was also observed for trinucleotide cation radicals having a 5′-adenine, such as $(AAG+2H)^{+\bullet}$ (**Figure 3.8a**), $(AGG+2H)^{+\bullet}$, and $(ATG+2H)^{+\bullet}$ (**Figure 3.9a,b**), although in these cases it competed with other backbone cleavages: namely, those forming $\boldsymbol{w_2}^+$ and $\boldsymbol{d_2}^+$ ions. It may be argued that protonation at 5′-adenine and guanine can occur competitively with these sequences to provide

intermediates for the 5′-adenine loss, which was observed as an abundant dissociation in the spectra. The formation of $(\boldsymbol{w_2}+H)^{+\bullet}$ ions from $(CGG+2H)^{+\bullet}$ (**Figure 3.9c**) was in fact a less abundant process that occurred in competition with dissociations yielding $(\boldsymbol{w_2}–H)^{+\bullet}$ and $\boldsymbol{w_2}^+$ ions. Prominent formation of $(\boldsymbol{w_2}–H)^{+\bullet}$ ions was also observed for $(CCG+2H)^{+\bullet}$ (**Figure 3.10g**), indicating a reverse hydrogen transfer to the departing 5′-cytosine.



Scheme 3.8: Calculated Structures and Relative Energies for the Competitive Formation of $\boldsymbol{w_2}^+$ and $\boldsymbol{d_2}^+$ Ions from $(AAA+2H)^{+\bullet}$ [a]
[a]Description as in Scheme 3.5.

Dissociations of the $(AAA+2H)^{+\bullet}$ ions revealed competitive formation of $\boldsymbol{w_2}^+$ and $\boldsymbol{d_2}^+$ ions, as resolved by ^{15}N-labeling (**Figure 3.11**). Using BOMD and DFT calculations, we obtained optimized structures of the relevant species. $(AAA+2H)^{+\bullet}$ ions were found to prefer zwitterionic structures that were produced by one-electron reduction of low-energy zwitterionic dications.[48] This is illustrated with **AAA1**$^{+\bullet}$ (**Scheme 3.8**) that was a zwitterion protonated at 5′-N1 and 3′-N7

whereas the radical was in the middle adenine base. The nearest-energy canonical structure (**AAA2**$^{+\bullet}$) was protonated at 5′-N3 with the radical being in the 3′-base. Loss of 5′-adenine from **AAA1**$^{+\bullet}$ should lead to ion **AA**$^{+\bullet}$ as the lowest-energy protomer having the radical at 5′-C1′. Although the structure **AAA1**$^{+\bullet}$ correlates with the less stable adenine N1–H tautomer, proton transfer between the 5′-adenine and the proximate phosphate group could deliver the proton to N3, as it is in **AAA2**$^{+\bullet}$, and produce the lower energy N3–H adenine tautomer. The calculated dissociation energy for the latter was rather low at 89 kJ mol^{-1}, which was consistent with the facile adenine loss. Remarkably, the formation of the lowest-energy cation radical **AA**$^{+\bullet}$ necessitated electron transfer from the adenine radical to the C1′ position at 5′-deoxyribose. The competitive formations of the $\boldsymbol{w_2}^+$ and $\boldsymbol{d_2}^+$ ions were calculated to both have low and very similar threshold energies. The branching ratio for these two dissociations obviously depends on the kinetics in the relevant rate-determining steps. Nevertheless, the calculated threshold energies, favoring the formation of $\boldsymbol{w_2}^+$ and $\boldsymbol{a_1}^{\bullet}$, were consistent with the $\boldsymbol{w_2}^+$ and $\boldsymbol{d_2}^+$ fragment ion distribution in the CID-MS^3 spectrum (**Figure 3.11**).

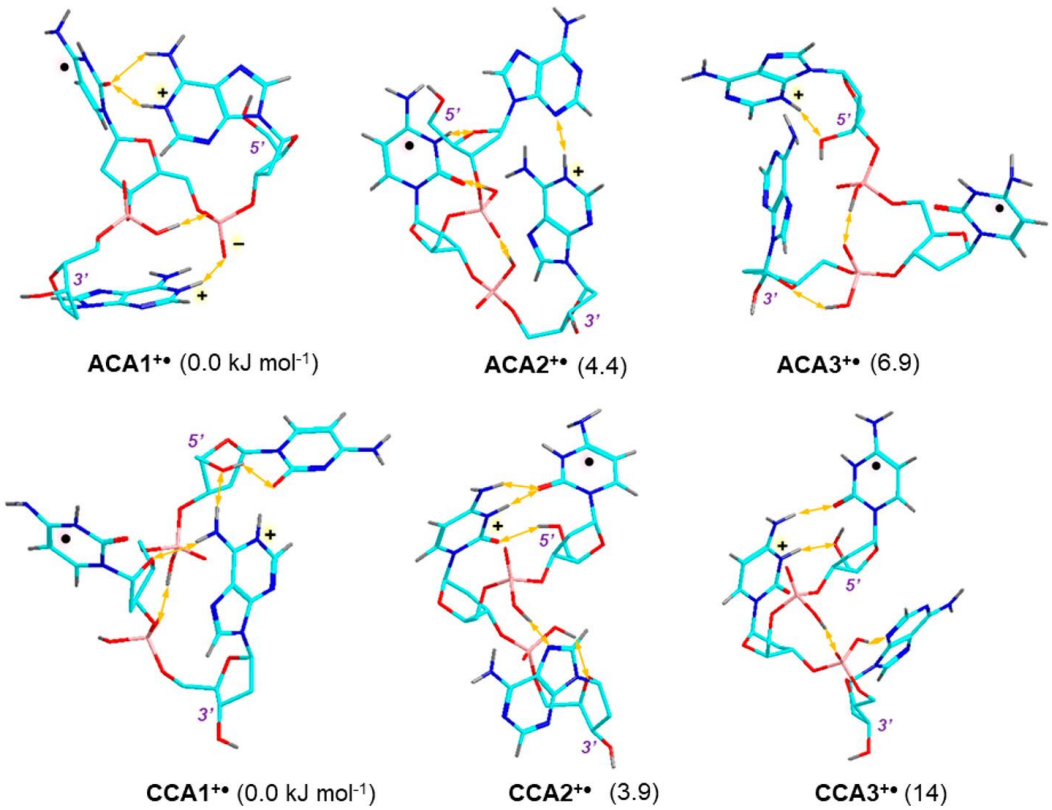
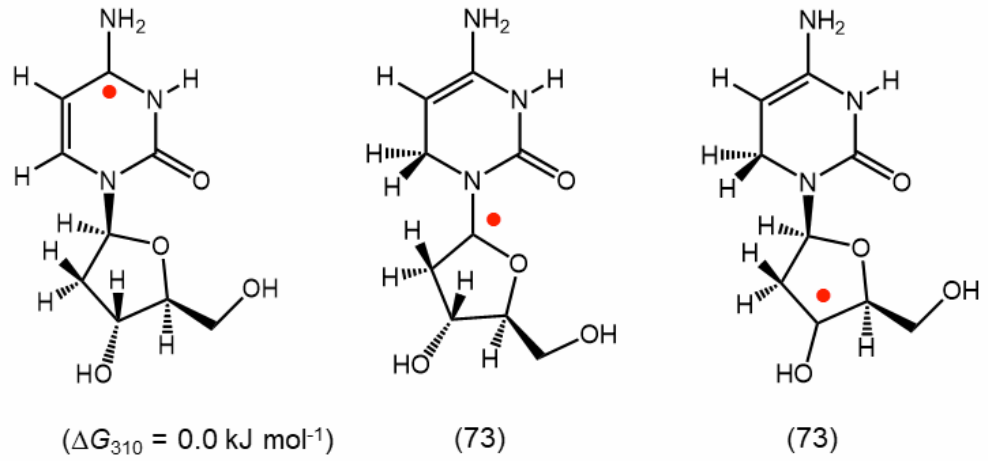


Figure 3.17: Calculated structures of low-energy (ACA+2H)$^{+\bullet}$ and (CCA+2H)$^{+\bullet}$ ions. Structure description as in Scheme 3.5.

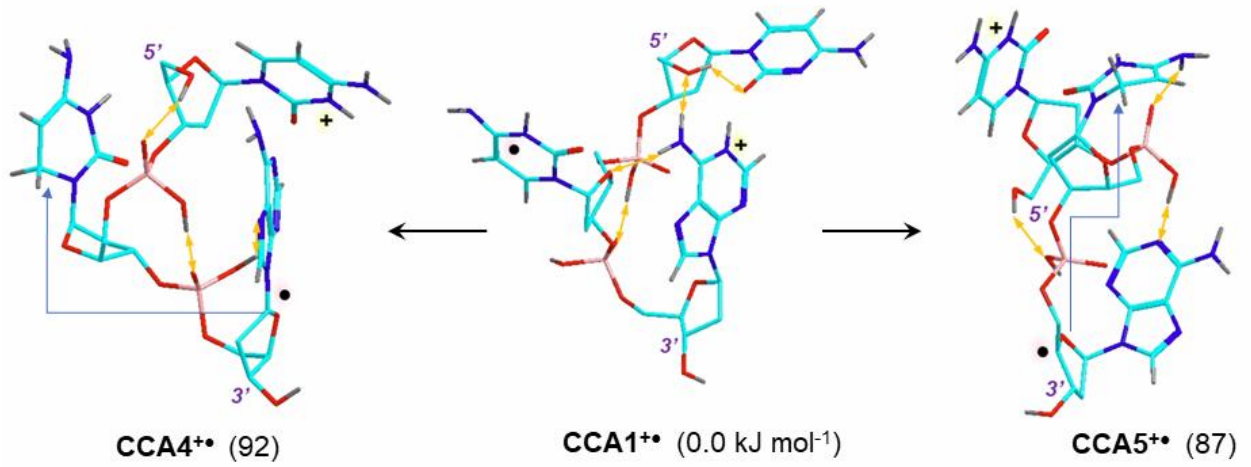
Dissociations forming the $(\boldsymbol{d_2}+2H)^+$ ions were prominently represented with sequences having a middle-cytosine-3′-purine arrangement, such as in $(ACA+2H)^{+\bullet}$, $(ACG+2H)^{+\bullet}$, $(TCA+2H)^{+\bullet}$, $(GCA+2H)^{+\bullet}$, $(CCA+2H)^{+\bullet}$, and $(CCG+2H)^{+\bullet}$ (**Figure 3.10a–g**). To elucidate the electronic effects leading to this unusual dissociation, we addressed by combined BOMD and DFT calculations the structures of $(ACA+2H)^{+\bullet}$ and $(CCA+2H)^{+\bullet}$ (**Figure 3.17**). The three lowest-energy $(ACA+2H)^{+\bullet}$ ions, $\mathbf{ACA1}^{+\bullet}$, $\mathbf{ACA2}^{+\bullet}$, and $\mathbf{ACA3}^{+\bullet}$, were nearly isoenergetic according to M06-2X calculations. All three structures had the radical delocalized between C4 and C6 in the cytosine base, while the proton distribution patterns differed.

$\mathbf{ACA1}^{+\bullet}$ was a zwitterion in which 3′- and 5′-adenine were protonated at N1, and the overall charge balance was maintained by the phosphate anion between the middle and 5′-nucleoside. $\mathbf{ACA2}^{+\bullet}$ and $\mathbf{ACA3}^{+\bullet}$ were canonical cation-radical structures that were protonated at N1-3′- and N3-5′-adenine, respectively. The very similar relative energies of these protomers, as well as their multiple hydrogen bonds connecting the backbone phosphates and nucleobases, indicated that the cation radicals could undergo facile interconversions. The low-energy $(CCA+2H)^{+\bullet}$ ions were all canonical cation radicals. The nearly isoenergetic $\mathbf{CCA1}^{+\bullet}$ and $\mathbf{CCA2}^{+\bullet}$ differed in the position of the radical, which was at the middle cytosine in the former and at 5′-cytosine in the latter. Ion $\mathbf{CCA3}^{+\bullet}$ was a higher-energy conformer of $\mathbf{CCA2}^{+\bullet}$, from which it differed by the orientation of 3′-adenosine. The common feature of the $(ACA+2H)^{+\bullet}$ and $(CCA+2H)^{+\bullet}$ ions was the presence of a cytosine radical in the middle position or its facile formation by hydrogen transfer. This indicated that the cytosine radical played a role in hydrogen transfer from the 3′-nucleoside to trigger dissociation leading to $(\boldsymbol{d_2}+2H)^+$ ions. We examined the hydrogen transfer energetics with model cytidine radicals (**Scheme 3.9**). Transfer to cytosine C6 of deoxyribose H1′ or H3′ was 73 kJ mol^{-1} endergonic, which was well within the range of dissociation energies for trinucleotide cation radicals (vide supra). The energetics for the 3′-deoxyribose hydrogen migrations onto C6 in the middle cytosine radical in **CCA2+•** was similar to that in the model cytidines, 92 and 87 kJ mol^{-1} for H1′ and H4′, in the respective intermediates **CCA4+•** and **CCA5+•** (**Scheme 3.10**). While these energies were within the range of those for other trinucleotide cation radical dissociations, they allowed for competition with the loss of $CH_2OH^{\bullet}$ from the 5′-sugar and backbone dissociation to $\boldsymbol{w_2}^+$ ions from $(CCA+2H)^{+\bullet}$ (**Figure 3.10f**) that are likely to be radical-initiated. The loss of $CH_2OH^{\bullet}$ could originate from the low-energy structure $\mathbf{CCA2}^{+\bullet}$ that had a sterically favorable arrangement of the 5′-cytosine radical and deoxyribose. Similarly, the formation of the $\boldsymbol{w_2}^+$ ion can

be initiated by 5′-cytosine loss in **CCA2**$^{+\bullet}$. Returning back to $(\boldsymbol{d_2}+2H)^+$ ions, the formation of the 3′-sugar radicals in **CCA4**$^{+\bullet}$ and **CCA5**$^{+\bullet}$ was expected to weaken the PO–C5′ bond in 3′-deoxyribose to facilitate elimination of the ($\boldsymbol{z_1}$–H) radical and formation of the $(\boldsymbol{d_2}+2H)^+$ ion.



Scheme 3.9: Relative Gibbs energies of cytidine radicals from M06-2X/6-311++G(2d,p) + ZPVE + ΔH_{310} + ΔS_{310} calculations.



Scheme 3.10: Structures and relative Gibbs energies of $(CCA+2H)^{+\bullet}$ ions after hydrogen atom migration from 3'-deoxyribose to the middle cytosine radical.

3.4 Conclusions

CID-MS^3 spectra of 60 trinucleotide cation radicals, reported here, displayed a variety of dissociations that were triggered by radicals residing in the nucleobases. The radical sites were created by electron attachment to a protonated nucleobase or after hydrogen atom transfer to a neutral nucleobase, which were primarily cytosine or thymine of high hydrogen atom affinities. Dissociations forming standard $\boldsymbol{d_2}^+$ and $\boldsymbol{w_2}^+$ backbone fragments were promoted by hydrogen atom radical migrations that substantially lowered the threshold energies. In addition, multiple hydrogen atom migrations accompanied the backbone dissociations, resulting in the formation of novel $(\boldsymbol{w_2}+H)^{+\bullet}$, $(\boldsymbol{w_2}+2H)^+$, and $(\boldsymbol{d_2}+2H)^+$ fragment ions. These dissociations were linked to the presence and sequence position of particular nucleobases. Guanine residues supported dissociations producing $(\boldsymbol{w_2}+H)^{+\bullet}$ ions. Cytosine and thymine nucleobases in 3′- and 5′-positions triggered cross-ring cleavages and loss of CH_2OH, respectively, in the terminal deoxyriboses. DFT calculations of trinucleotide cation radical structures and energies led us to conclude that adenine-rich trinucleotide cation radicals in the AAA and ACA combinations existed as low-energy zwitterionic structures. In contrast, guanine nucleotides did not favor zwitterionic structures. The calculations also indicated very low threshold energies for the radical dissociations, which was consistent with the rich chemistry of these cation radicals.

3.5 Bibliography

[1] Ward, J. Complexity of Damage Produced by Ionizing Radiation; in *Cold Spring Harbor Symposia on Quantitative Biology*, Cold Spring Harbor Laboratory Press: 2000; pp 377– 382.

[2] von Sonntag, C. *Free-Radical-Induced DNA Damage and Its Repair*; Springer: 2006.

[3] Sanche, L. Low Energy Electron-Driven Damage in Biomolecules. *Eur. Phys. J. D Atomic, Molecular, Optical and Plasma Physics* **2005**, *35*, 367.

[4] Alizadeh, E.; Orlando, T. M.; Sanche, L. Biomolecular Damage Induced by Ionizing Radiation: The Direct and Indirect Effects of Low-Energy Electrons on DNA. *Annu. Rev. Phys. Chem.* **2015**, *66*, 379.

[5] Giese, B. Long-Distance Charge Transport in DNA: The Hopping Mechanism. *Acc. Chem. Res.* **2000**, *33*, 631.

[6] Giese, B. Long-Distance Electron Transfer Through DNA. *Annu. Rev. Biochem.* **2002**, *71*, 51.

[7] Douki, T.; Ravanat, J.-L.; Angelov, D.; Wagner, J. R.; Cadet, J. Effects of Duplex Stability on Charge-Transfer Efficiency within DNA. *Top. Curr. Chem.* **2004**, *236*, 1.

[8] Takada, T.; Kawai, K.; Fujitsuka, M.; Majima, T. Direct Observation of Hole Transfer through Double-Helical DNA over 100 A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14002.

[9] *Charge Transfer in DNA*; Wagenknecht, H.-A., Ed.; Wiley-VCH: 2005.

[10] Gray, H. B. *Charge Transfer in DNA*: from Mechanism to Application; Wiley: 2006.

[11] Genereux, J. C.; Boal, A. K.; Barton, J. K. DNA-Mediated Charge Transport in Redox Sensing and Signaling. *J. Am. Chem. Soc.* **2010**, *132*, 891.

[12] Kawai, K.; Majima, T. Hole Transfer Kinetics of DNA. *Acc. Chem. Res.* **2013**, *46*, 2616.

[13] Adhikary, A.; Collins, S.; Khanduri, D.; Sevilla, M. D. Sugar Radicals Formed by Photoexcitation of Guanine Cation Radical in Oligonucleotides. *J. Phys. Chem. B* **2007**, *111*, 7415.

[14] Chen, H.-Y.; Kao, C.-L.; Hsu, S. C. N. Proton Transfer in Guanine-Cytosine Radical Anion Embedded in B-Form DNA. *J. Am. Chem. Soc.* **2009**, *131*, 15930.

[15] Choi, J.; Yang, C.; Fujitsuka, M.; Tojo, S.; Ihee, H.; Majima, T. Proton Transfer of Guanine Radical Cations Studied by Time- Resolved Resonance Raman Spectroscopy Combined with Pulse Radiolysis. *J. Phys. Chem. Lett.* **2015**, *6*, 5045.

[16] Rodriguez-Santiago, L.; Noguera, M.; Bertan, J.; Sodupe, M. Hydrogen Bonding and Proton Transfer in Ionized DNA Base Pairs, Amino Acids and Peptides. In *Quantum Biochemistry*; Matt, C. F., Ed.; Wiley-VCH: 2010; pp 219– 242.

[17] Joy, A.; Ghosh, A. K.; Schuster, G. B. One-Electron Oxidation of DNA Oligomers That Lack Guanine: Reaction and Strand Cleavage at Remote Thymines by Long-Distance Radical Cation Hopping. *J. Am. Chem. Soc.* **2006**, *128*, 5346.

[18] Ghosh, A.; Joy, A.; Schuster, G. B.; Douki, T.; Cadet, J. Selective One-Electron Oxidation of Duplex DNA Oligomers: Reaction at Thymines. *Org. Biomol. Chem.* **2008**, *6*, 916.

[19] Joseph, J.; Schuster, G. B. Oxidatively Damaged Nucleobases in Duplex DNA Oligomers: Reaction at Thymine-Thymine Mispairs. *J. Am. Chem. Soc.* **2009**, *131*, 13904.

[20] Kanvah, S.; Schuster, G. B. One-Electron Oxidation of DNA: Thymine versus Guanine Reactivity. *Org. Biomol. Chem.* **2010**, *8*, 1340.

[21] Barnett, R. N.; Joseph, J.; Landman, U.; Schuster, G. B. Oxidative Thymine Mutation in DNA: Water-Wire-Mediated Proton-Coupled Electron Transfer. *J. Am. Chem. Soc.* **2013**, *135*, 3904.

[22] Boudaıffa, B.; Cloutier, P.; Hunting, D.; Huels, M. A.; Sanche, L. Resonant Formation of DNA Strand Breaks by Low-Energy (3 to 20 EV) Electrons. *Science* **2000**, *287* (5458), 1658.

[23] Wang, C.-R.; Nguyen, J.; Lu, Q.-B. Bond Breaks of Nucleotides by Dissociative Electron Transfer of Nonequilibrium Prehydrated Electrons: A New Molecular Mechanism for Reductive DNA Damage. *J. Am. Chem. Soc.* **2009**, *131*, 11320.

[24] Steenken, S. Purine Bases, Nucleosides, and Nucleotides: Aqueous Solution Redox Chemistry and Transformation Reactions of Their Radical Cations and e- and OH Adducts. *Chem. Rev.* **1989**, *89*, 503.

[25] Candeias, L. P.; Steenken, S. Electron Adducts of Adenine Nucleosides and Nucleotides in Aqueous Solution: Protonation at Two Carbon Sites (C2 and C8) and Intra- and Intermolecular Catalysis by Phosphate. *J. Phys. Chem. A* **1992**, *96*, 937.

[26] Candeias, L. P.; Wolf, P.; O'Neill, P.; Steenken, S. Reaction of Hydrated Electrons with Guanine Nucleosides: Fast Protonation on Carbon of the Electron Adduct. *J. Phys. Chem. A* **1992**, *96*, 10302.

[27] Barnes, J.; Bernhard, W. A. The Protonation State of One-Electron Reduced Cytosine and Adenine. 1. Initial Protonation Sites at Low Temperatures in Glassy Solids. *J. Phys. Chem. A* **1993**, *97*, 3401.

[28] Barnes, J.; Bernhard, W. A. One-Electron-Reduced Cytosine in Acidic Glasses: Conformational States before and after Proton Transfer. *J. Phys. Chem. A* **1994**, *98*, 10969.

[29] Steenken, S.; Jovanovic, S. V. How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *J. Am. Chem. Soc.* **1997**, *119*, 617.

[30] Wolken, J. K.; Syrstad, E. A.; Vivekananda, S.; Tureček, F. Uracil Radicals in the Gas Phase. Specific Generation and Energetics. *J. Am. Chem. Soc.* **2001**, *123*, 5804.

[31] Syrstad, E. A.; Vivekananda, S.; Tureček, F. Direct Observation of a Hydrogen Atom Adduct to C-5 in Uracil. A Neutralization-Reionization Mass Spectrometric and Ab Initio Study. *J. Phys. Chem. A* **2001**, *105*, 8339.

[32] Wolken, J. K.; Tureček, F. Direct Observation of a Hydrogen Atom Adduct to O-4 in Uracil. A Neutralization-Reionization Mass Spectrometric and Ab Initio Study. *J. Phys. Chem. A* **2001**, *105*, 8352.

[33] Chen, X.; Syrstad, E. A.; Nguyen, M. T.; Gerbaux, P.; Tureček, F. Adenine Radicals in the Gas Phase. An Experimental and Computational Study of Hydrogen Atom Adducts to Adenine. *J. Phys. Chem. A* **2005**, *109*, 8121.

[34] Yao, C.; Cuadrado-Peinado, M.; Polášek, M.; Tureček, F. Specific Generation of 1-Methylcytosine Radicals in the Gas-Phase. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 6708.

[35] Wee, S.; O'Hair, R. A. J.; McFadyen, W. D. Can Radical Cations of the Constituents of Nucleic Acids Be Formed in the Gas Phase Using Ternary Transition Metal Complexes? *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1797.

[36] Lam, A. K.; Abrahams, B. F.; Grannas, M. J.; McFadyen, W. D.; O'Hair, R. A. J. Tuning the Gas Phase Redox Properties of Copper(II) Ternary Complexes of Terpyridines to Control the Formation of Nucleobase Radical Cations. *Dalton Trans.* **2006**, 5051.

[37] Feketeova, L.; Yuriev, E.; Orbell, J. D.; Khairallah, G. N.; O'Hair, R. A. J. Gas-Phase Formation and Reactions of Radical Cations of Guanosine, Deoxyguanosine and Their Homodimers and Heterodimers. *Int. J. Mass Spectrom.* **2011**, *304*, 74.

[38] Feketeova, L.; Chan, B.; Khairallah, G. N.; Steinmetz, V.; Maitre, P.; Radom, L.; O'Hair, R. A. J. Gas-Phase Structure and Reactivity of the Keto Tautomer of the Deoxyguanosine Radical Cation. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25837.

[39] Sun, Y.; Zhou, W.; Moe, M. M.; Liu, J. Reactions of Water with Radical Cations of Guanine, 9-Methylguanine, 2′-Deoxyguanosine and Guanosine: Keto-Enol Isomerization, C8 Hydroxylation, and Effects of N9-Substitution. *Phys. Chem. Chem. Phys.* **2018**, *20*, 27510.

[40] Dang, A.; Liu, Y.; Tureček, F. UV-Vis Action Spectroscopy of Guanine, 9-Methylguanine and 2′-Deoxyguanosine Cation Radicals in the Gas Phase. *J. Phys. Chem. A* **2019**, *123*, 3272.

[41] Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528.

[42] Xia, Y.; Gunawardena, H. P.; Erickson, D. E.; McLuckey, S. A. Effects of Cation Charge-Site Identity and Position on Electron-Transfer Dissociation of Polypeptide Cations. *J. Am. Chem. Soc.* **2007**, *129*, 12232.

[43] Hari, Y. I.; Leumann, C.; Schürch, S. What Hinders Electron Transfer Dissociation (ETD) of DNA Cations? *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2677.

[44] Huang, S. R.; Liu, Y.; Tureček, F. UV-Vis Photodissociation Action Spectroscopy Reveals Cytosine-Guanine Hydrogen Transfer in DNA Tetranucleotide Cation Radicals upon One-Electron Reduction. *J. Phys. Chem. B* **2020**, *124*, 3505.

[45] Liu, Y.; Huang, S. R.; Tureček, F. Guanine-Adenine Interactions in DNA Tetranucleotide Cation Radicals Revealed by UV/Vis Photodissociation Action Spectroscopy and Theory. *Phys. Chem. Chem. Phys.* **2020**, *22*, 16831.

[46] Turecek, F. Flying DNA Cation Radicals in the Gas Phase: Generation and Action Spectroscopy of Canonical and Noncanonical Nucleobase Forms. *J. Phys. Chem. B* **2021**, *125*, 7090.

[47] Vrkic, A. K.; O'Hair, R. A. J.; Foote, S.; Reid, G. E. Fragmentation Reactions of All 64 Protonated Trimer Oligonucleotides Via Tandem Mass Spectrometry in an Ion Trap. *Int. J. Mass Spectrom.* **2000**, *194*, 145.

[48] Wan, J.; Brož, B.; Liu, Y.; Huang, S. R.; Marek, A.; Tureček, F. Resolution of Identity in Gas-Phase Dissociations of Mono and Diprotonated DNA Codons by 15N-Labeling and Computational Structure Analysis. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 1936.

[49] Brož, B.; Tureček, F.; Marek, A. Low Scale Syringe-Made Synthesis of 15N-Labeled Oligonucleotides. *J. Label. Compd. Radiopharm.* **2022**, *65*, 309.

[50] Hartmer, R.; Kaplan, D. A.; Gebhardt, C. R.; Ledertheil, T.; Brekenfeld, A. Multiple Ion/Ion Reactions in the 3D Ion Trap: Selective Reagent Ion Production for ETD and PTR from a Single Compound. *Int. J. Mass Spectrom.* **2008**, *276*, 82.

[51] Berendsen, H. J.; Postma, J. V.; van Gunsteren, W. F.; DiNola, A. R. H. J.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684.

[52] 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.

[53] Stewart, J. J. P. *MOPAC 16*; Stewart Computational Chemistry: 2016.

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

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

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

[57] 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.

[58] Huang, S. R.; Tureček, F. Noncanonical Isomers of Nucleoside Cation Radicals: An Ab Initio Study of the Dark Matter of DNA Ionization. *J. Phys. Chem. A* **2022**, *126*, 2480.

[59] Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735.

[60] Korn, J. A.; Urban, J.; Dang, A.; Nguyen, H. T. H.; Turecek, F. UV-Vis Action Spectroscopy Reveals a Conformational Collapse in Hydrogen-Rich Dinucleotide Cation Radicals. *J. Phys. Chem. Lett.* **2017**, *8*, 4100.

[61] Liu, Y.; Ma, C.; Nováková, G.; Marek, A.; Tureček, F. Charge-Tagged Nucleosides in the Gas Phase: UV-Vis Action Spectroscopy and Structures of Cytidine Cations, Dications, and Cation Radicals. *J. Phys. Chem. A* **2021**, *125*, 6096.

[62] Liu, Y.; Ma, C.; Leonen, C. J. A.; Chatterjee, C.; Nováková, G.; Marek, A.; Tureček, F. Tackling a Curious Case. Generation of Charge-Tagged Guanosine Radicals by Electron Transfer in the Gas Phase and Their Characterization by UV-Vis Photodissociation Action Spectroscopy and Theory. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 772.

[63] Holm, A. I. S.; Larsen, M. K.; Panja, S.; Hvelplund, P.; Nielsen, S. B.; Leib, R. D.; Donald, W. A.; Williams, E. R.; Hao, C.; Tureček, F. Electron Capture, Femtosecond Electron Transfer and Theory: A Study of Noncovalent Crown Ether 1, n-Diammonium Alkane Complexes. *Int. J. Mass Spectrom.* **2008**, *276*, 116.

[64] Viglino, E.; Lai, C. K.; Mu, X.; Chu, I. K.; Tureček, F. Ground and Excited-Electronic State Dissociations of Hydrogen-Rich and Hydrogen-Deficient Tyrosine Peptide Cation Radicals. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1454.

[65] Gunawardena, H. P.; He, M.; Chrisman, P. A.; Pitteri, S. J.; Hogan, J. M.; Hodges, B. D.; McLuckey, S. A. Electron Transfer versus Proton Transfer in Gas-Phase Ion/Ion Reactions of Polyprotonated Peptides. *J. Am. Chem. Soc.* **2005**, *127*, 12627.

[66] McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60.

[67] Murray, K. K. DNA Sequencing by Mass Spectrometry. *J. Mass Spectrom.* **1996**, *31*, 1203.

[68] Schürch, S. Characterization of Nucleic Acids by Tandem Mass Spectrometry-The Second Decade (2004–2013): From DNA to RNA and Modified Sequences. *Mass Spectrom. Rev.* **2016**, *35*, 483.

[69] Tureček, F.; Yao, C. Hydrogen Atom Addition to Cytosine, 1-Methylcytosine, and Cytosine-Water Complexes. A Computational Study of a Mechanistic Dichotomy. *J. Phys. Chem. A* **2003**, *107*, 9221.

[70] Chen, X.; Syrstad, E. A.; Nguyen, M. T.; Gerbaux, P.; Tureček, F. Adenine Radicals in the Gas Phase. An Experimental and Computational Study of Hydrogen Atom Adducts to Adenine. *J. Phys. Chem. A* **2005**, *109*, 8121.

[71] Wu, R. R.; Hamlow, L. A.; He, C. C.; Nei, Y. W.; Berden, G.; Oomens, J.; Rodgers, M. T. The Intrinsic Basicity of the Phosphate Backbone Exceeds That of Uracil and Thymine Residues: Protonation of the Phosphate Moiety Is Preferred over the Nucleobase for PdThd and PUrd. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30351.

[72] Rodgers, M. T.; Campbell, S.; Marzluff, E. M.; Beauchamp, J. L. Site-Specific Protonation Directs Low-Energy Dissociation Pathways of Dinucleotides in the Gas Phase. *Int. J. Mass Spectrom. Ion Processes* **1995**, *148*, 1.

[73] Tureček, F.; Gu, M.; Hop, C. E. C. A. Franck-Condon Dominated Chemistry. Formation and Dissociations of Tetrahydroxyphosphoranyl Radicals Following Femtosecond Reduction of Their Cations in the Gas Phase. *J. Phys. Chem.* **1995**, *99*, 2278.

Chapter 4

Do d(GCGAAGC) Cations Retain the Hairpin Structure in the Gas Phase? A Cyclic Ion Mobility Mass Spectrometry and Density Functional Theory Computational Study

Reproduced in part with permission from Jiahao Wan, Marianna Nytka, Haocheng Qian, Karel Lemr and František Tureček. The DNA Radical Code. Do d(GCGAAGC) Cations Retain the Hairpin Structure in the Gas Phase? A Cyclic Ion Mobility Mass Spectrometry and Density Functional Theory Computational Study. *J. Am. Soc. Mass Spectrom.* **2023**, *34*, 10, 2323–2340.

Abstract *d(GCGAAGC) is the smallest oligonucleotide with a well-defined hairpin structure in solution. We report a study of multiply protonated d(GCGAAGC) and its sequence-scrambled isomers, d(CGAAGCG), d(GCGAACG), and d(CGGAAGC), that were produced by electrospray ionization with the goal of investigating their gas-phase structures and dissociations. Cyclic ion mobility measurements revealed that dications of d(GCGAAGC) as well as the scrambled-sequence ions were mixtures of protomers and/or conformers that had collision cross sections (CCS) within a 439–481 Å^2 range. Multiple ion conformers were obtained by electrospray under native conditions as well as from aqueous methanol. Arrival time distribution profiles were characteristic of individual isomeric heptanucleotides. Extensive Born–Oppenheimer molecular dynamics (BOMD) and density functional theory (DFT) calculations of d(GCGAAGC)$^{2+}$ isomers indicated that hairpin structures were high-energy isomers of more compact distorted conformers. Protonation caused a break up of the C2···G6 pair that was associated with the formation of strong hydrogen bonds in zwitterionic phosphate anion-nucleobase cation motifs that predominated in low energy ions. Multiple components were also obtained for d(GCGAAGC)$^{3+}$ trications under native and denaturing electrospray conditions. The calculated trication structures showed disruption of the G···C pairs in low energy zwitterions. A hairpin trication was calculated to be a high energy isomer. d(GCGAAGC)$^{4+}$ tetracations were produced and separated by c-IMS as two major isomers. All low energy d(GCGAAGC)$^{4+}$ ions obtained by DFT geometry optimizations were zwitterions in which all five purine bases were protonated, and the ion charge was balanced by a phosphate anion. Tetracations of the scrambled sequences were each formed as one dominant isomer. The CCS calculated with the MobCal-MPI method were found to closely match experimental values. Collision-induced dissociation (CID) spectra of multiply charged*

heptanucleotides showed nucleobase loss and backbone cleavages occurring chiefly at the terminal nucleosides. Electron-transfer-CID tandem mass spectra were used to investigate dissociations of different charge and spin states of charge-reduced heptanucleotide cation radicals.

4.1 Introduction

The question of biomolecular ion structure upon transitioning from solution to the gas phase has been of interest to the research community ever since it became possible to obtain large gas-phase ions by electrospray ionization.[1,2] Multiply charged protein cations, in particular, have been a frequent subject of structure-related studies using H/D exchange[2] and ion mobility. [3−8] The major factors affecting protein ion stability in the gas phase have been identified as being due to Coulomb repulsion and the internal solvation of polar and charged groups. It has been recognized that the experimental methods used to tease out gas-phase protein ion structures have limited resolution and are often complemented by molecular dynamics modeling.[9] Compared with peptides and proteins, much less is known about the gas-phase structures of oligonucleotide ions. Due to their acid–base nature, oligonucleotides have been mostly studied as gas-phase polyanions using methods such as temperature induced melting,[10] collisional activation,[11−13] and ion mobility.[14−18] These and related studies have concluded that some features of the oligonucleotide solution structure may be retained in the gas-phase anions. [14,19−21] Even less is known about gas-phase structures of multiply charged oligonucleotide cations, despite that multiple protonation of oligonucleotides by electrospray is remarkably effective, involving both the phosphate groups and nucleobases.[22−27] An ion mobility study of GAA·TCC triplex ions concluded on the basis of force-field molecular dynamics calculations that the oligonucleotide cations retained their solution structure.[28] Remarkably, these authors considered only cytosine-protonated tautomers in their structure assignment.[28] Protonation affects the electrostatic interactions in the gas-phase ions in several ways. Negatively charged phosphates are converted to neutral phosphoester groups that can provide hydroxyl groups to participate in hydrogen bonding networks.[29] Protonation of nucleobases, with the exception of the less basic thymine, affects their proton donor and acceptor sites that can disrupt standard hydrogen bonding patterns, such as the Watson–Crick (WC) pairwise G-C and A-T interactions. In addressing these questions at the fundamental level, it is desirable to investigate oligonucleotide models that have well-established structures in solution

and can be transferred into the gas phase in the form of multiply charged cations. At the same time, a suitable model should have a limited size to be amenable to a rigorous computational investigation of energy and structure that can be utilized to obtain experimentally relevant parameters, such as collision cross sections (CCS). The d(GCGAAGC) miniloop, also called the d(GCGAAGC) hairpin, appeared to be such a suitable model. The miniloop conformation in the condensed phase has been established by X-ray crystallography[30] and NMR spectroscopy,[31] and the structure details have been refined using NMR with residual dipolar couplings.[32] The structure is maintained by WC pairs between terminal G1-C7 and next inner C2-G6, which are supported by π-stacking with G3 and A4 (**Figure 4.1a**).

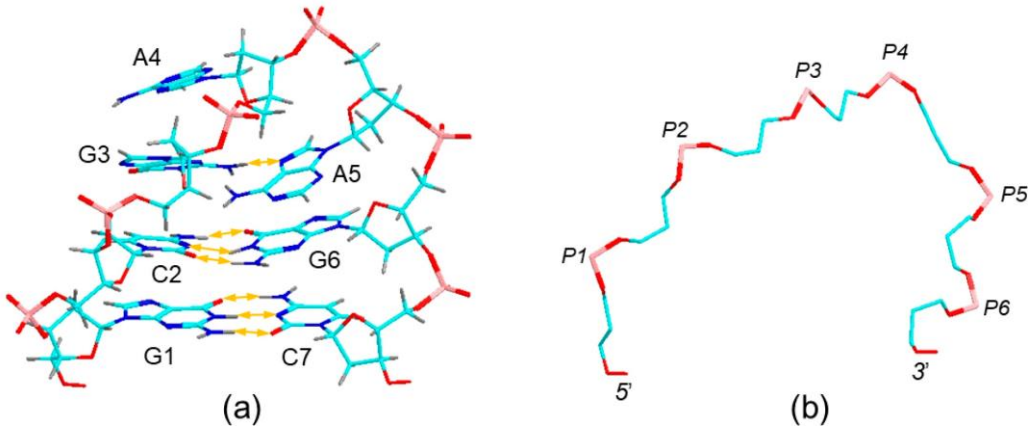


Figure 4.1: (a) d(GCGAAGC) solution structure from refined NMR analysis.[32] Atom color coding is as follows: cyan = C, blue = N, red = O, pink = P, gray = H. Hydrogen bonds between nucleobases are shown as double-headed ochre arrows. (b) Extracted d(GCGAAGC) backbone consisting of O5′-C5′-C4′-C3′-O3′-P segments starting from the 5′ terminus.

The deoxyribose-phosphate backbone makes a smooth loop that allows the WC pairs to assume the energetically most favorable positions (**Figure 4.1b**). Despite its thermodynamic stability in solution, the hairpin can be disrupted when attached to the 3′-terminus of a larger oligonucleotide and hybridized with a complementary strand.[33] Size-wise, the doubly through quadruply protonated forms of d(GCGAAGC) consist of 230–232 atoms, which make them amenable to a computational analysis by Born–Oppenheimer molecular dynamics (BOMD) and gradient geometry optimization using quantum chemical methods. Here, we report a combined experimental and computational study of di-, tri-, and tetracations produced by electrospray protonation of d(GCGAAGC) under native conditions of 5 mM ammonium acetate and from aqueous methanol. As reference models, we include an investigation of polycations produced from

modified heptanucleotide sequences. The reverse sequence, d(CGAAGCG), may theoretically include two WC G-C pairs, although its condensed-phase structure has not been established. Two other sequences that we investigated, d(GCGAACG) and d(CGGAAGC), have the terminal G and C nucleobases in "wrong" mutual positions and thus do not allow WC base pairing in a loop structure. We show that these gas-phase cations can develop a number of canonical and zwitterionic structures within a broad range of relative Gibbs energies that can be characterized by comparing experimental and theoretical collision cross sections. We also report collision-induced dissociation (CID) and tandem electron-transfer-CID (ET-CID-MS^3) spectra of multiply charged heptanucleotide cations to elucidate the effects of sequence and charge on their gas-phase ion chemistry.

4.2 Experimental Section

4.2.1 Materials

Heptanucleotides d(GCGAAGC), d(CGAAGCG), d(GCGAACG), and d(CGGAAGC) were custom synthesized by Integrated DNA Technologies, Inc. (Coralville, IA), and their quality and sequences were checked by tandem mass spectrometry.

4.2.2 Methods

Tandem mass spectra were measured on a Bruker amaZon 3D ion trap mass spectrometer furnished with an auxiliary chemical ionization source for electron and proton transfer studies. [34] Multiply charged ions were generated by electrospray ionization of 10–20 μM solutions in 50:50:1 methanol–water–acetic acid and selected by mass. Collision-induced dissociations were performed with r.f. excitation amplitudes that were adjusted to achieve >50% dissociation of the precursor ions. Electron transfer dissociation was carried out with fluoranthene anion radicals ($C_{16}H_{10}^{-\bullet}$) at 100–150 ms ion–ion reaction times. Proton transfer reactions were performed with fluoranthene anions ($C_{16}H_9^-$) at 100–150 ms reaction times. Ion mobility measurements were made on a Waters Select Series Cyclic IMS Q-TOF instrument equipped with a cyclic traveling wave device (Waters Corp., Wilmslow, U.K.).[35,36] Sample solutions were directly infused into the electrospray ion

source at a rate of 5 µL/min. The instrumental parameters were set as follows: capillary voltage 2 kV, cone voltage 40 V, source offset 10 V, source temperature 100 °C, desolvation temperature 280 °C, cone gas 30 L/h, desolvation gas 600 L/h, nebulizer gas 6.0 bar, step wave body gradient 20 V, head gradient 10 V, trap DC bias 35 V, transfer CE 4 V, helium flow rate 120 mL/min, nitrogen flow rate 40 mL/min, racetrack bias 70 V (for other parameters see **Table 4.4**–**4.11**). The precursor ions were selected by a quadrupole (highmass 15, lowmass 4.9). The IMS experiment sequence consisted of injection, separation, and ejection steps. The separation time was kept at 2 ms. Five TOF pushes per data bin were set for all measurements. Ions for IMS measurements were generated from both 50:50:1 methanol/water/acetic acid and 5 mM ammonium acetate solutions. The latter were used to mimic "native" ionization conditions that are believed to preserve solution conformations of biomolecules.[9] The instrument control was carried out by Masslynx 4.2 with modified DriftScope 2.9 (all Waters Corp., Wilmslow, U.K.) to analyze data. CCS values were determined as averages of six single-pass ion mobility measurements with manual calibration using a standard Waters procedure based on a logarithmic fit[37–39] (**Table 4.5**, **Figure 4.31**). Arrival time profiles obtained by single-pass c-IMS measurements were least-squares fitted with multiple Gaussian peaks using the Excel solver. The fitting parameters were constrained to accommodate the increasing peak broadening due to diffusion as a function of arrival time. The CCS for the maxima of the fitted peaks was calculated by using the experimental calibration curve. The fitted data were compared with arrival time distributions from multipass measurements, increasing the ion path in c-IMS from 98 cm for 1 pass up to 490 cm for 5 passes. The maximum number of passes was chosen to avoid wrap-around overlaps.

4.2.3 Calculations

Conformational analysis was carried out by Born–Oppenheimer molecular dynamics calculations (BOMD) as described previously.[29] Briefly, the semiempirical PM6 method[40] that was augmented by dispersion and hydrogen bonding interactions (D3H4)[41] was used to treat electron energy. The total energy stability was maintained by the Berendsen thermostat.[42] Initial structures of d(GCGAAGC) cations were derived from the reported Cartesian coordinates of the DNA miniloop that were obtained from the NMR spectra.[32] Protons were placed on the phosphate groups and guanine and adenine nucleobases to generate +2 through +4 charge states. On the basis

of previous energy analysis of gas-phase di-,[24] tri-,[29] and tetranucleotide cations, [26,27] we considered protonation at N7 of guanine and N3 of adenine as the most energetically favorable sites. This created 10 initial combinations of doubly charged protomers that are denoted as G1G3, G1G6, G3G6, G1A4, G1A5, G3A4, G3A5, A4G6, A5G6, and A4A5. Trication protomers were considered for the initial G1G3G6, G1G3A4, G3A4G6, and G1A4A5 combinations. Tetracation calculations started from the G1G3A4G6 protomer. BOMD trajectories were run for all of these combinations for 20 ps at 300, 500, 600, and 800 K using the high-end Cuby4 platform.[43] Since BOMD included both electron and nuclear motion, the initial proton positions were not fixed and the ions were allowed to isomerize by proton migration in the course of the BOMD trajectory, as indeed observed and discussed later in the paper. It should be noted, however, that in none of the low-energy structures did we observe proton migration onto a cytosine ring. The 20000 snapshots obtained by BOMD for each initial structure and temperature were used to extract 200 snapshots at regular intervals that were then fully gradient optimized with PM6-D3H4 run under MOPAC.[44] Structures optimized in this cycle were sorted out to remove duplicates and high-energy ions, and the selected candidates were considered for geometry optimization with density functional theory (DFT) calculations. In contrast to our previous experience with DFT calculations of di-, tri-, and tetranucleotide mono and dications [24–27,29] using hybrid density functionals such as B3LYP,[45] M06-2X,[46] and ωB97-XD[47,48] with the 6-31+G(d,p) basis set, these calculations were failing for all charge states of the heptanucleotide ions. The failure typically consisted of a poor wave function and energy gradient guess when starting the next optimization cycle, which resulted in a sudden jump of energy and a distorted structure. Due to the size of these systems (230–232 atoms) and the 6-31+G(d,p) basis set (3176-3187 basis functions), the calculations became intractable. Fortunately, we found out that including in B3LYP calculations empirical corrections of dispersion energy with Becke–Johnson damping (GD3-BJ)[49] resulted in smooth convergence of geometry and energy. Geometry optimizations were therefore run with B3LYP/6-31G(d,p) + GD3-BJ that also provided harmonic frequencies for selected ions. Single-point energies were calculated with B3LYP/6-31++G(d,p) + GD3-BJ that also provided atomic charge densities for collision cross section calculations. The reliability of the B3LYP/6-31G(d,p) + GD3-BJ data was checked by benchmarking against M06-2X/6-31+G(d,p) calculations that were run for a series of trinucleotide cations where relative energies obtained by both methods showed the same trend. [29] The reasons for the failure of standard DFT geometry optimizations of heptanucleotide cations were not clear

but may warrant further analysis. These calculations were carried out with Gaussian 16 (Revision B.01) that was licensed from Gaussian Inc. (Wallingford, CT). The fully optimized geometries were used in calculations of Mulliken and dipole restricted B3LYP-GD3-BJ/6-31++G(d,p) atomic charge densities according to the Mertz–Singh–Kollman (MK) scheme.[50] The ion trajectory method[51] was used to calculate CCS in nitrogen. We used two different sets of parameters; in one we employed the parameters reported by Campuzano et al.[52] and Kim et al.[53] using Mulliken charge densities. The other was based on MK charge densities using the MobCal-MPI package of Ieritano, Hopkins et al.[54,55] In general, we found the CCS calculated by MobCal-MPI to give closer fits than those from standard MobCal.

4.3 Results and Discussion

4.3.1 Ion Formation and Characterization

Electrospray ionization of aqueous methanol solutions of all heptanucleotide sequences under study produced doubly, triply, and quadruply protonated ions at m/z 1065.7, 710.8, and 533.4, respectively. The relative intensities of these multiply charged ions were tuning-dependent and optimized for measurements involving the given charge state. We note that the extent of multiple protonation by electrospray was remarkable, considering that in aqueous solution the heptanucleotides exist as polyanions, so the formation of +4 states in gas-phase ions may require a take up of up to 10 protons. Increased acidity in electrospray microdroplets, as first reported in 1994,[56] may in part account for the facile multiple protonation of the heptanucleotides. Electrospray ionization of d(GCGAAGC) and d(CGAAGCG) solutions in 5 mM ammonium acetate under "native" conditions[7] gave doubly and triply charged ions along with quadruply charged dimers. Quadruply charged monomers were formed at very low intensities under native conditions. The cations were selected by mass and examined by cyclic ion mobility measurements, CID-MS^2 and ET-CID-MS^3 spectra of pertinent charge states.

4.3.2 Ion Mobility Separations and CCS

Heptanucleotide ions in the +2, +3, and +4 charge states were investigated by ion mobility measurements of arrival times using the TWAVE-cyclic IMS. The four sequences showed different arrival time profiles, as discussed for different charge states. The +2 state of the miniloop sequence, d(GCGAAGC)$^{2+}$, showed a composite peak with arrival times between 36 and 42 ms for the single-pass measurement (**Figure 4.2a**). The composite arrival time profile was deconvoluted by fitting with Gaussian peaks having maxima at t_D = 37.2, 38.3, and 40.9 ms and constituting 53, 35, and 7% of the total integrated dication intensity. An improved fit to the experimental profile reaching rmsd = 0.5% was obtained by including another Gaussian at t_D = 39.3 ms (7% intensity), resulting in a near-perfect overlap. The accuracy of the Gaussian fit in **Figure 4.2a** was probed by c-IMS measurements using 2 and 5 passes (**Figure 4.2b** and **4.2c**, respectively). These showed improved separation of the four peaks from the single-pass measurement (**Figure 4.2a**) that appeared with similar relative intensities. A possible fifth component was indicated in **Figure 4.2c** as a shoulder at 233.5 ms. d(GCGAAGC)$^{2+}$ ions produced under "native" conditions by electrospray from 5 mM aqueous ammonium acetate displayed arrival time profiles that were virtually identical with those obtained for ions from aqueous-methanol solutions. The c-IMS data of d(GCGAAGC)$^{2+}$ that were obtained over a period of several months differed in the relative intensity of the band for dimeric tetracations that were readily distinguished by their isotope pattern and did not interfere with the dications (**Figure 4.3a,b**). In contrast, the composite peak of the dication arrival time distribution was unchanged when the ions were produced by electrospray from either aqueous methanol or ammonium acetate solution. This indicated that the electrospray solvent, either "denaturing" or "native", had a negligible effect on the formation of different ion isomers. We further used a composite c-IMS sequence consisting of injecting and separating the ions in one or two cycles, which was followed by storing an ion population in a prearray store (slicing)[35] within a narrow arrival-time range. The slicing (i.e., ion mobility isolation of selected ions) eliminated the interference of quadruply charged dimers in multipass measurements. These ions were reinjected into c-IMS and separated in more cycles. The sequence is visualized in **Figure 4.4**. Comparison of **Figure 4.2b** and **4.2c** arrival time distributions indicated that there were no significant changes in the peak relative positions and intensities upon ion manipulation.

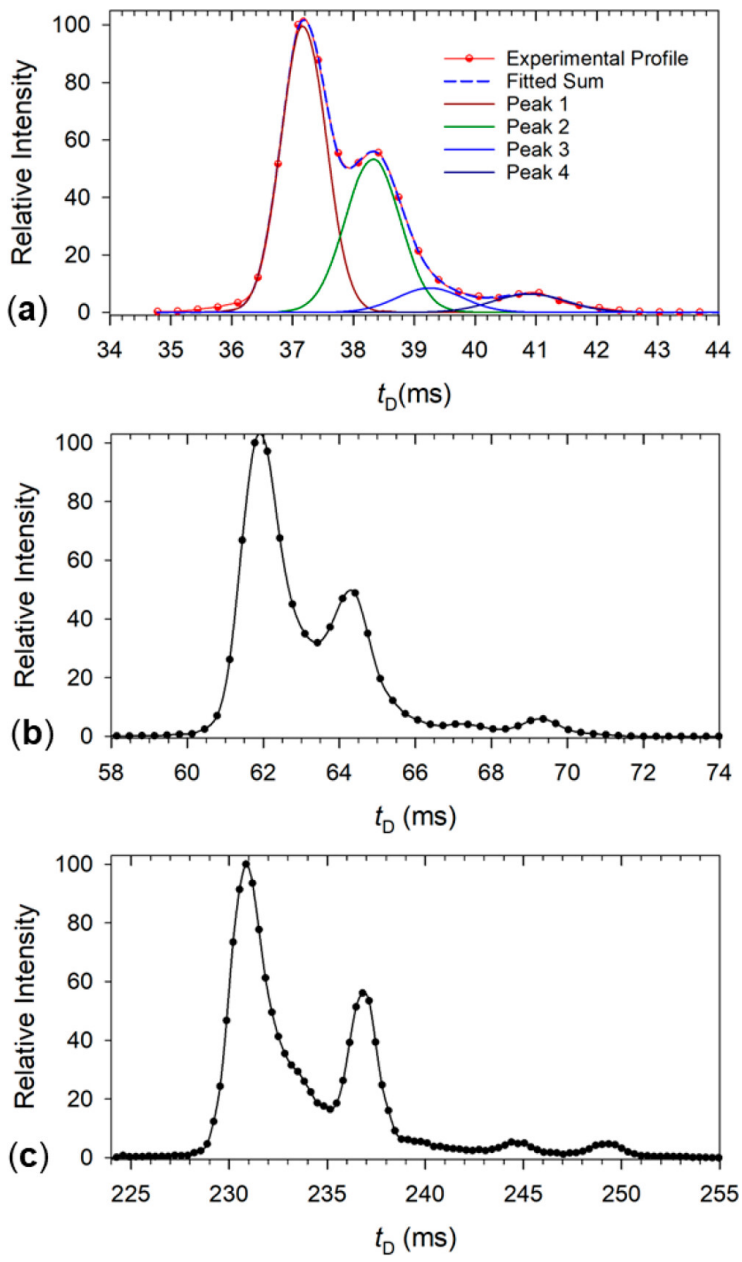


Figure 4.2: Arrival time profiles from c-IMS of d(GCGAAGC)$^{2+}$ dications after (a) one, (b) two, and (c) five passes. The ion population in (c) was stored after two passes and after slicing reinjected for an additional five c-IMS passes. Gaussian fits of partially resolved and unresolved peaks are shown in (a).

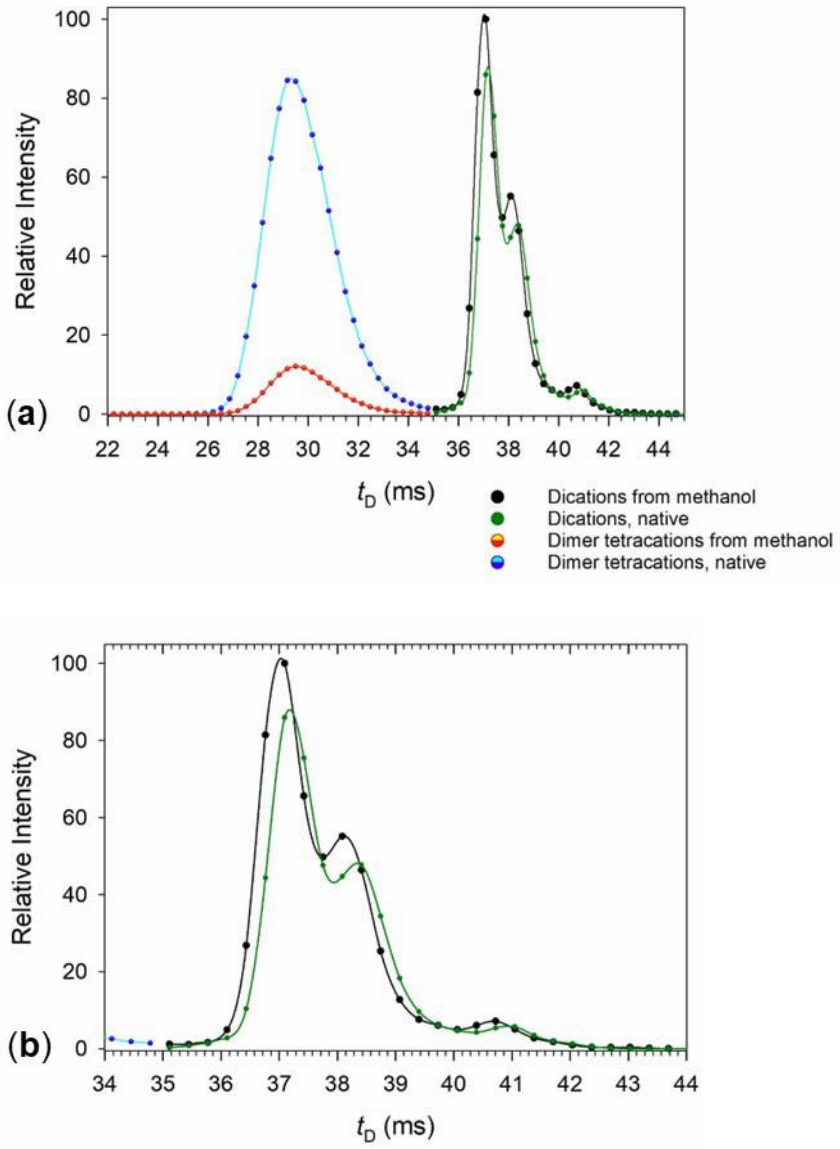


Figure 4.3: d(GCGAAGC)$^{2+}$ arrival time profiles. (a) Dimer tetracations and dications electrosprayed from aqueous methanol and aqueous ammonium acetate. (b) Detailed view of the dication arrival-time profiles.

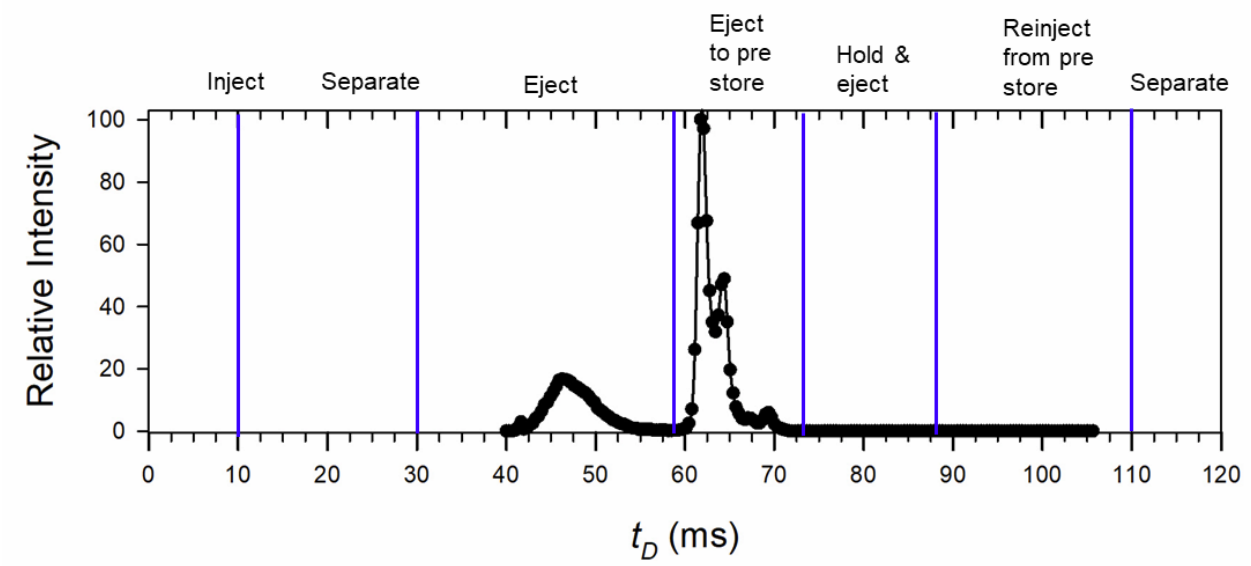


Figure 4.4: Injection and isolation time sequence for slicing and multipass c-IMS measurements of d(GCGAAGC)$^{2+}$.

The inverted dication, d(CGAAGCG)$^{2+}$, also showed a composite peak consisting of several overlapping components (**Figure 4.5a**). In this case, it was more difficult to attain unambiguous deconvolution into individual Gaussian peaks, because of the substantial overlap. Our best fit with *rmsd* = 0.4% showed three peaks with t_D = 35.8, 36.8, 38.0 ms that gave 8, 16, and 50% of total ion intensity. This fit was improved by including another Gaussian at t_D = 38.2 ms, which amounted to 25% of total dication intensity. Analogous results were obtained for c-IMS of d(CGAAGCG)$^{2+}$ that were produced under "native" conditions that also showed composite peak profiles (**Figure 4.6**). Increasing the number of c-IMS passes revealed the presence of two minor components at longer arrival times (**Figure 4.5b**, **c**). For example, after five passes, the data showed at least four components (**Figure 4.5c**). Interestingly, the unresolved shoulder at the lowest t_D in **Figure 4.5a** and **4.5b** appeared to merge into the major peak after the slicing and five-pass separation (**Figure 4.5c**). This could possibly indicate that the d(CGAAGCG)$^{2+}$ isomers interconvert during the passage through the c-IMS. This possibility was investigated by collisionally activating the population of stored (CGAAGCG)$^{2+}$ ions separated by passage through one cycle of IMS. However, subsequent c-IMS by three cycles of separation of the collisionally activated ions revealed no dependence on the excitation voltage in the range 0–70 V (**Figure 4.7**).

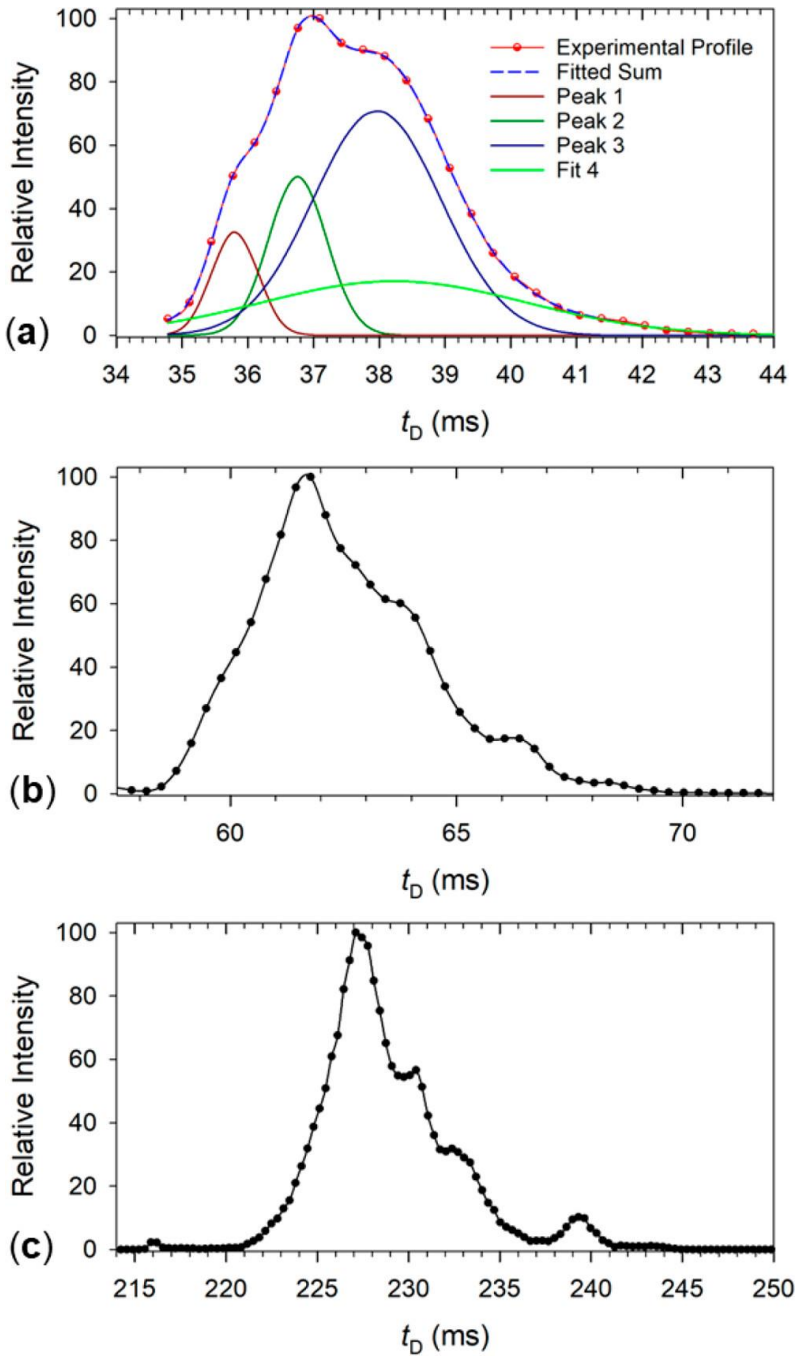


Figure 4.5: Arrival time profiles from c-IMS of d(CGAAGCG)$^{2+}$ dications after (a) one, (b) two, and (c) five passes. The ion population in (c) was stored (slicing) after two passes and reinjected for an additional five c-IMS passes. Gaussian fits of partially resolved and unresolved peaks are shown in (a).

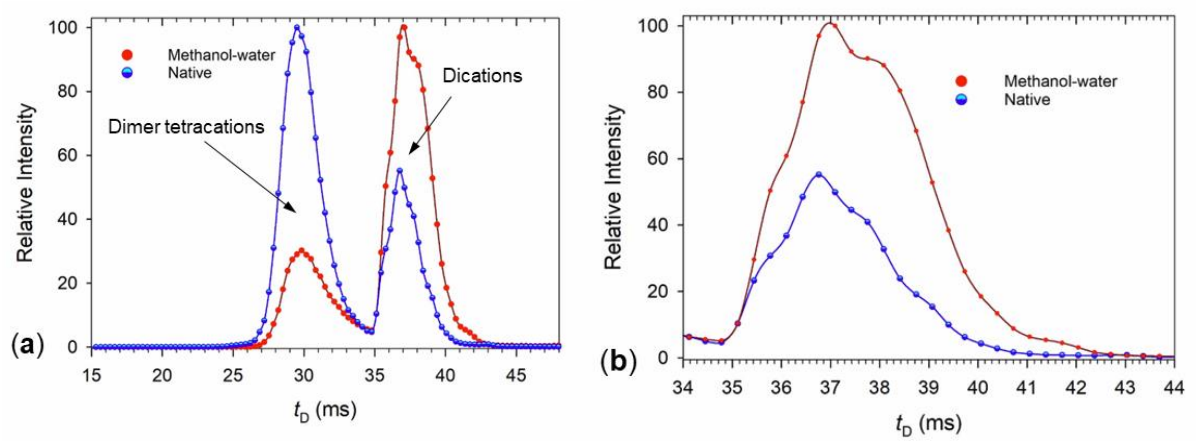


Figure 4.6: d(CGAAGCG)$^{2+}$ arrival time profiles. (a) Dimer tetracations and dications electrosprayed from aqueous methanol and aqueous ammonium acetate. (b) Detailed view of the dication arrival-time profiles.

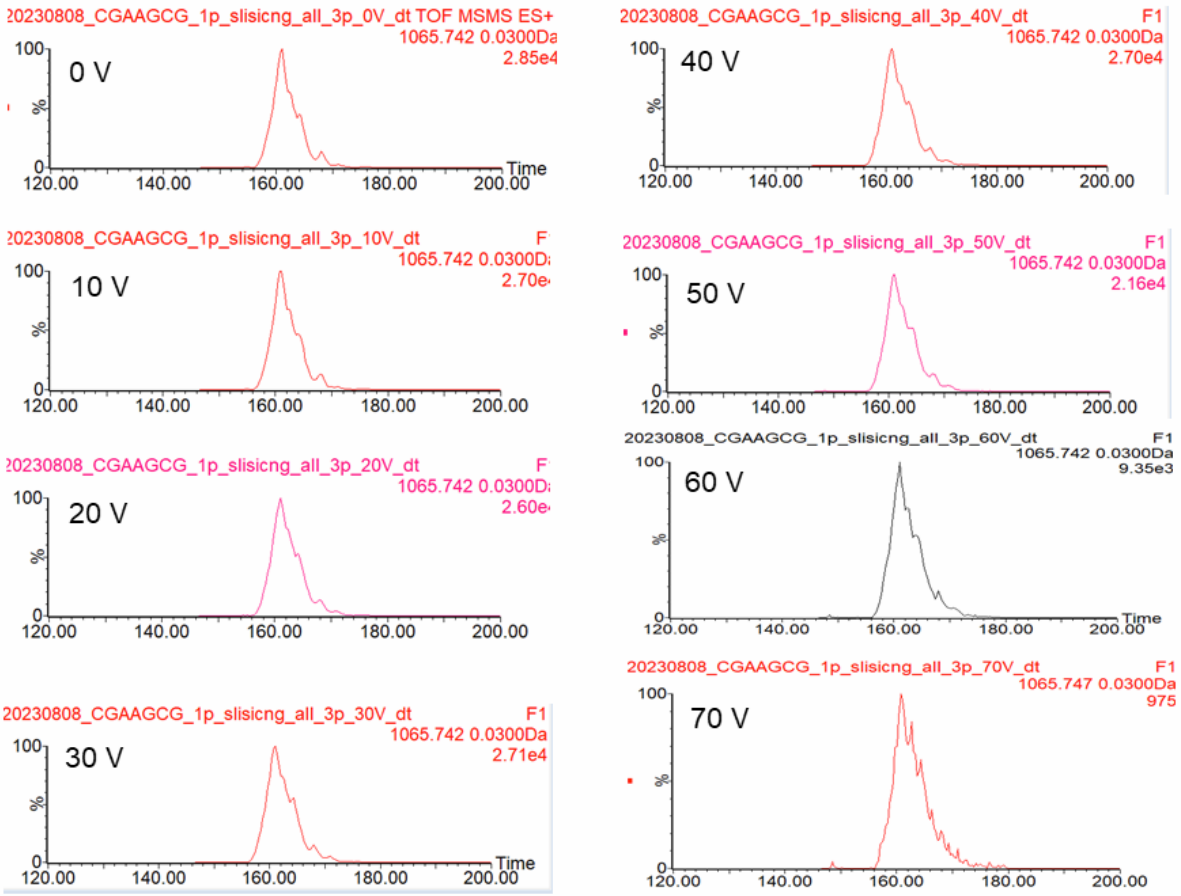


Figure 4.7: Collisional activation of c-IMS separated d(CGAAGCG)$^{2+}$ ions at 0-70 V followed by 3 cycles of c-IMS.

The arrival time profile from c-IMS of d(GCGAACG)$^{2+}$ dications showed a major component that was accompanied by multiple peaks that were partially resolved in the single-pass scan

(**Figure 4.8a**). After deconvolution, the arrival time profile of d(GCGAACG)$^{2+}$ was composed of a major peak with a maximum at t_D = 38.6 ms and side peaks at 36.5, 37.4, 39.8, and 41.3 ms (**Figure 4.8a**), giving a tight fit with rmsd = 0.2%. The Gaussian decomposition of the single-pass profile was corroborated by multiple-pass scans shown in **Figure 4.8b**, **c**. After two passes, the first two components were resolved to show two distinct peak maxima, which was further documented by the five-pass scan, where these peak maxima were further separated by 4 ms **(Figure 4.8c**). In contrast, the two minor peaks appearing at longer arrival times remained incompletely resolved even after five passes despite their stretched time separation window.

The dication with the other scrambled sequence, d(CGGAAGC)$^{2+}$, also showed a composite profile, consisting of a major component with a maximum at t_D = 38.4 ms and minor components at t_D = 36.9 and 41.0 ms (**Figure 4.9a**). Fitting in two more Gaussian peaks with t_D = 38.7 and 41.3 ms allowed us to achieve a rmsd = 0.4%. Arrival time profiles obtained at longer path lengths corroborated the Gaussian fit-assigned peaks for the major components. Extending the ion path length to two and five passes (196 and 490 cm, respectively) allowed us to completely resolve all five components (**Figure 4.9b**, **c**). The only deviation between the Gaussian fit and the resolved scans was in the width of the minor fourth component, which was fitted as a single broad peak. The arrival times of the Gaussian-fitted peaks were converted to CCS that are listed in **Table 4.1**. The major components of d(GCGAAGC)$^{2+}$ and d(CGAAGCG)$^{2+}$ showed very similar CCS values that were within the 448–470 Å^2 range. d(GCGAAGC)$^{2+}$ showed two major components with CCS = 450 and 459 Å^2 that when combined, amounted to 88% of the total ion intensity. These were accompanied by two minor components of CCS = 466 and 478 Å^2, at 7% and 5%, respectively. d(CGAAGCG)$^{2+}$ was composed of three major components with CCS = 448, 459, and 470 Å^2 that when combined, accounted for 91% of ion intensity. A minor peak with CCS = 440 Å^2 was distinguished by arrival-time profile deconvolution. Dications of the scrambled sequences each consisted of one major component that had CCS = 461 and 459 Å^2 at 49% and 54% for d(GCGAACG)$^{2+}$ and d(CGGAAGC)$^{2+}$, respectively (**Table 4.1**). In both cases, the major isomers were accompanied by several minor components within the 445–479 Å^2 range (**Table 4.1**).

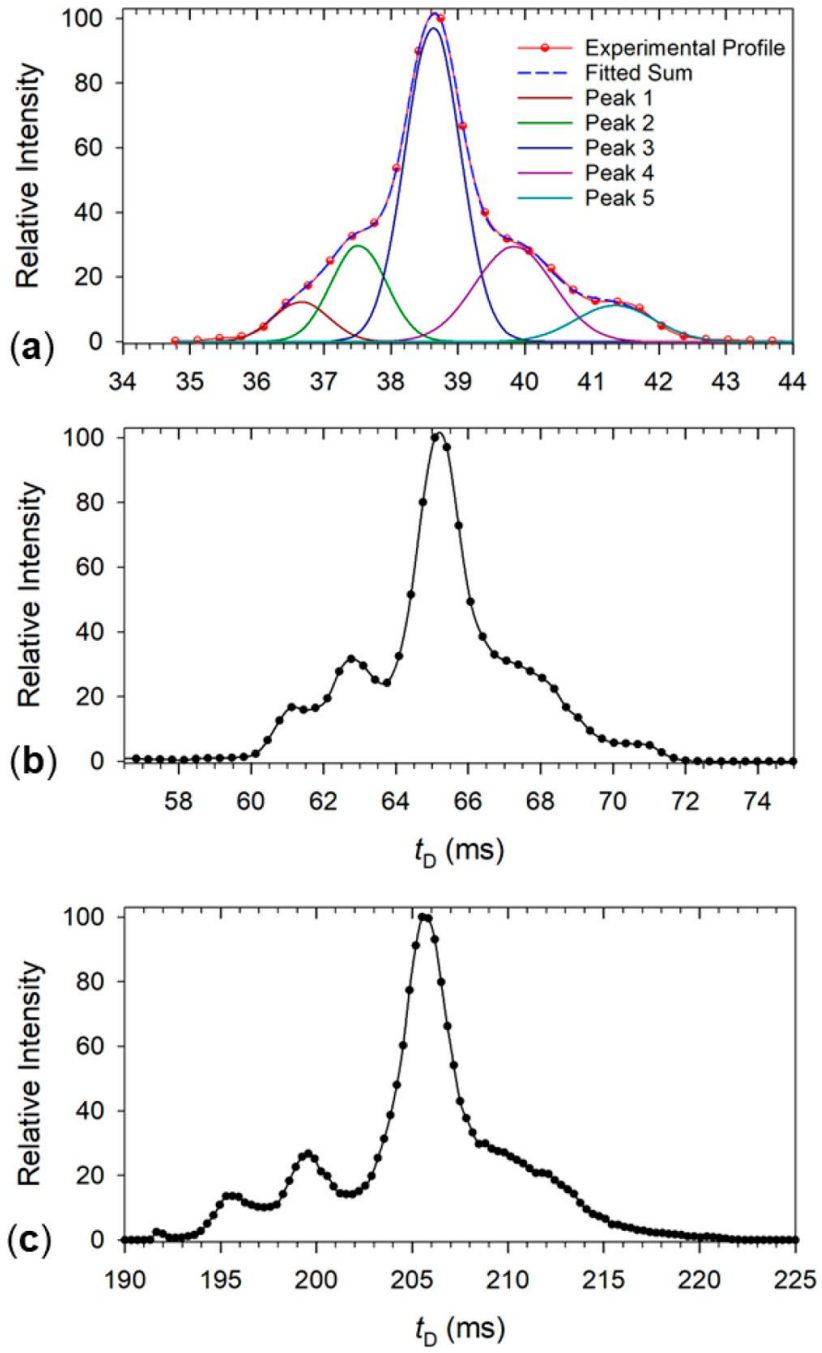


Figure 4.8: Arrival time profiles from c-IMS of d(GCGAACG)$^{2+}$ dications after (a) one, (b) two, and (c) five passes. The ion population in (c) was stored after one pass (slicing) and reinjected for an additional five c-IMS passes. Gaussian fits of partially resolved peaks are shown in (a).

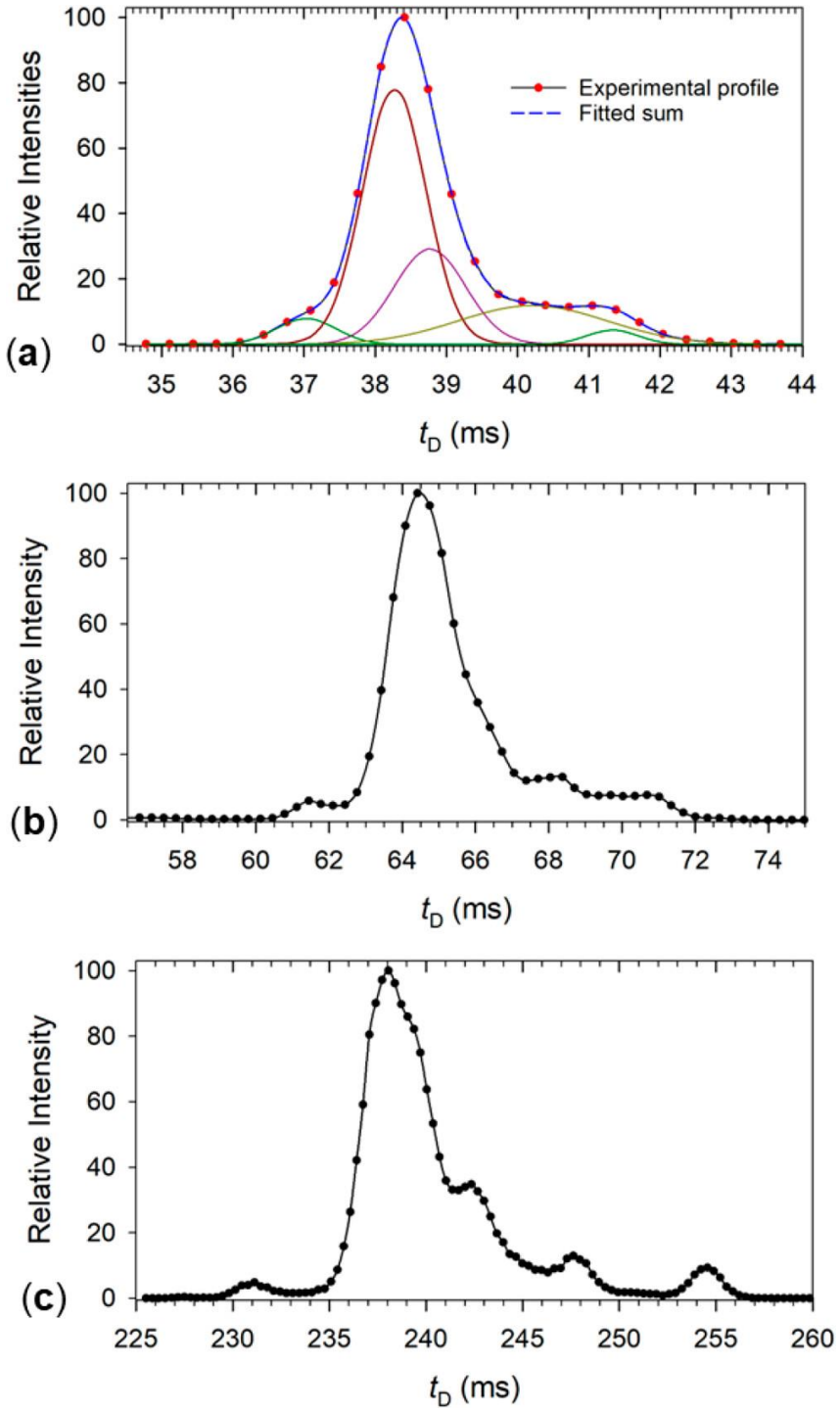


Figure 4.9: Arrival time profiles from c-IMS of d(CGGAAGC)$^{2+}$ dications after (a) one, (b) two, and (c) five passes. The ion population in (c) was stored after two passes (slicing) and reinjected for an additional five c-IMS passes. Gaussian fits of partially resolved peaks are shown in (a).

Table 4.1 Collision Cross Sections of Heptanucleotide Ions

	collision cross section[a]	
ion	**CCSexp**	**CCSfit**
d(GCGAAGC)$^{2+}$	450, 459, 479	450 (53)[b], 459 (35), 466 (7)[c], 478 (5)
d(CGAAGCG)$^{2+}$	448, 467	440 (8), 448 (16), 459 (50), 470 (25)
d(GCGAACG)$^{2+}$	461, 482	445 (6), 452 (15), 461, (49), 470 (21), 479 (8)
d(CGGAAGC)$^{2+}$	459, 480	448 (5), 459 (54), 462 (24)[c], 468 (8)[c], 479 (8)
d(GCGAAGC)$^{3+}$	510	494 (1), 509 (48), 521 (25), 533 (21), 549 (5)
d(CGAAGCG)$^{3+}$	510, 534, 555	510 (5), 522 (9), 531 (15), 535 (59), 550 (12)
d(GCGAACG)$^{3+}$	523, 565	495 (1), 510 (15), 524 (58), 538 (15), 552 (3), 565 (7)
d(CGGAAGC)$^{3+}$	517, 537, 553	489 (1), 513 (16), 523 (13), 532 (19), 542 (23), 554 (28)
d(GCGAAGC)$^{4+}$	617	584 (2)[b], 608 (46), 624 (51)
d(CGAAGCG)$^{4+}$	587, 624	596 (7), 625 (88), 636 (5)[b]
d(GCGAACG)$^{4+}$	624	593 (0.2)[b], 601 (6)[b], 624 (94)
d(CGGAAGC)$^{4+}$	607, 637	608 (83), 627 (17)

[a]In Å^2.
[b]Integrated relative intensities of deconvoluted ion mobility peaks in parentheses.
[c]Broad peaks of unresolved components.

The data indicated that the miniloop and inverted-sequence dications had gas-phase structures of quite similar degrees of compactness that arose from their solution conformations or resulted from conformational changes due to protonation in the gas phase. This was irrespective of the electrospray solvent, suggesting that the ion structures were governed by their gas-phase properties or the nature of the charging events in electrospray rather than their solution properties. The other aspect that could be gleaned from the c-IMS data was the heterogeneity of the ion populations for

all four heptanucleotide sequences, which was particularly salient for d(GCGAAGC)$^{2+}$ that originated from a single stable miniloop solution structure.

Composite peaks were also obtained by c-IMS of the trications (**Figure 4.10a–d**). The miniloop trication gave several components after deconvolution, with a clearly separated major peak at t_D = 25.3 ms and two other peaks at t_D = 26.0 and 26.8 ms. This group was flanked by two minor peaks that were distinguished by Gaussian fitting and had t_D = 24.4 and 27.8 ms. Hence the miniloop trications also showed structure heterogeneity in the gas phase. The inverted trication showed one major peak at 26.9 ms followed by a smaller peak at t_D = 27.8 ms.

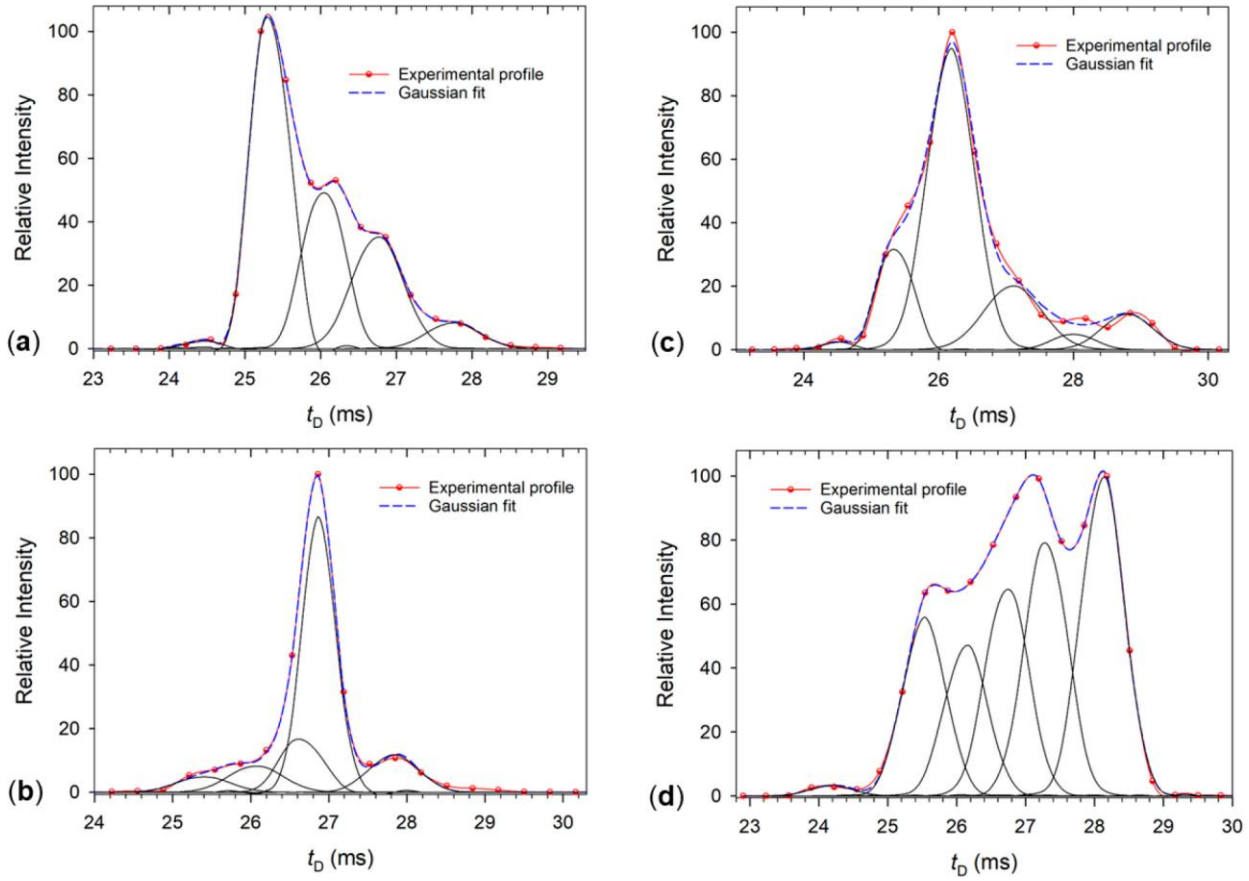


Figure 4.10: Arrival time profiles from single-pass c-IMS of trications. (a) d(GCGAAGC)$^{3+}$, (b) d(CGAAGCG)$^{3+}$, (c) d(GCGAACG)$^{3+}$, and (d) d(CGGAAGC)$^{3+}$ with best Gaussian fits.

The scrambled sequence trication, d(GCGAACG)$^{3+}$, gave a composite peak that was fitted by six Gaussian components at t_D = 24.5, 25.4, and 26.2 (major), 27.1, 28.0, and 28.8 ms. Finally, d(CGGAAGC)$^{3+}$ gave a composite peak consisting of at least three components that were fitted by six Gaussians with t_D = 24.2, 25.5, 26.1, 26.7, 27.3, and 28.1 ms. Multipass measurements gave

arrival time profiles that were consistent with the Gaussian fits in the single-pass spectra. For d(CGAAGCG)$^{3+}$, we obtained three well-separated peaks after five passes (**Figure 4.11a**). Three passes were sufficient to resolve the five peaks of d(GCGAACG)$^{3+}$ (**Figure 4.11b**) as predicted by Gaussian fitting in **Figure 4.10c**. The largely unresolved multiplet of d(CGGAAGC)$^{3+}$ (**Figure 4.10d**) showed partial peak separation, although the previously resolved peaks at t_D = 27.3 and 28.1 ms coalesced after four passes (**Figure 4.11c**).

The trication CCS showed a broader range of values compared to those of the dications. The major components from d(GCGAAGC)$^{3+}$ and d(CGAAGCG)$^{3+}$ were assigned CCS at 509 Å^2 and 535 Å^2, respectively (**Table 4.1**). The scrambled sequences had a larger CCS at 524 Å^2 and 513–554 Å^2 for the major components of d(GCGAACG)$^{3+}$ and d(CGGAAGC)$^{3+}$, respectively.

In contrast to the di- and trications, all four tetracations displayed simpler arrival time profiles (**Figure 4.12a–d**). Thus, d(GCGAAGC)$^{4+}$ gave two nearly resolved peaks of similar relative intensity at t_D = 22.4 and 23.0 ms (**Figure 4.12a**), whereas the inverted tetracation, d(CGAAGCG)$^{4+}$, showed one major peak at t_D = 23.0 ms and only minor peaks at t_D = 21.9 and 23.5 ms (**Figure 4.12b**). The d(GCGAACG)$^{4+}$ ion gave one dominant peak at t_D = 23.0 ms, which, judged by its width, could consist of two unresolved components of very similar arrival times. Finally, d(CGGAAGC)$^{4+}$ gave one major peak at t_D = 22.4 ms, which was accompanied by a minor component at t_D = 23.1 ms. The major components from d(GCGAAGC)$^{4+}$ had CCS = 608 and 624 Å^2 which was similar to the major peak from d(CGAAGCG)$^{4+}$ that had CCS = 625 Å^2. The scrambled sequences showed chiefly one major peak, which was at CCS = 624 and 608 Å^2 for d(GCGAACG)$^{4+}$ and d(CGGAAGC)$^{4+}$, respectively.

The results of these measurements revealed two facts. First, all charge states of all four sequences consisted of mixtures of isomers, which could be represented by different protomers, conformers, or combinations thereof. Second, the CCS of the miniloop ion components was not significantly different from those of the inverted and scrambled sequences. This raised the question of whether or not the gas-phase ions produced by electrospray retained the main features of the miniloop structure as defined by the two WC G-C pairs.[30–32] It is worth noting that arrival time profiles were characteristic of individual heptanucleotides (cf. **Figures 4.2**, **4.5**, **4.8**, **4.9**, **4.10**, **4.12**), as earlier observed for oligosaccharides.[57] The differences in their profiles were evident for all followed charge states and were well stable over time. ATD profiles may be useful in distinguishing isomeric oligonucleotides in general.

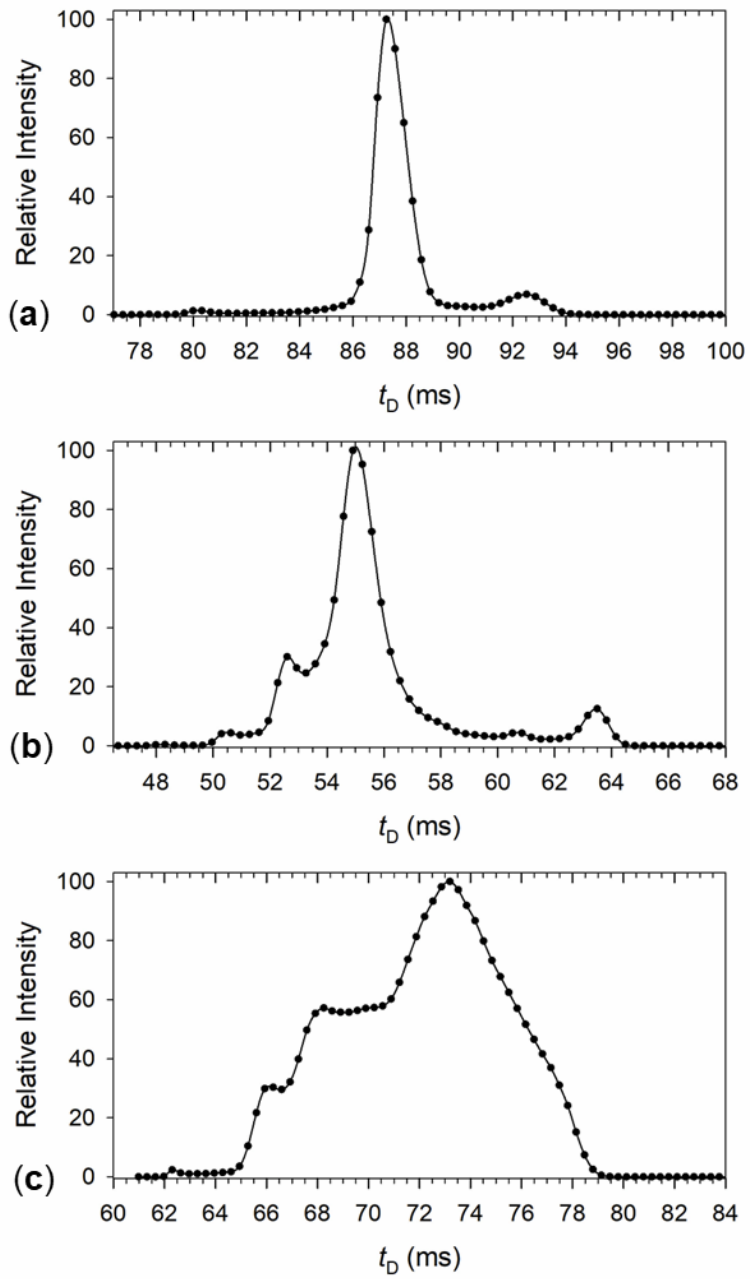


Figure 4.11: Arrival time profiles for multiple passes of trications: (a) d(CGAAGCG)$^{3+}$ (5 passes), (b) d(GCGAACG)$^{3+}$ (3 passes), (c) d(CGGAAGC)$^{3+}$ (4 passes).

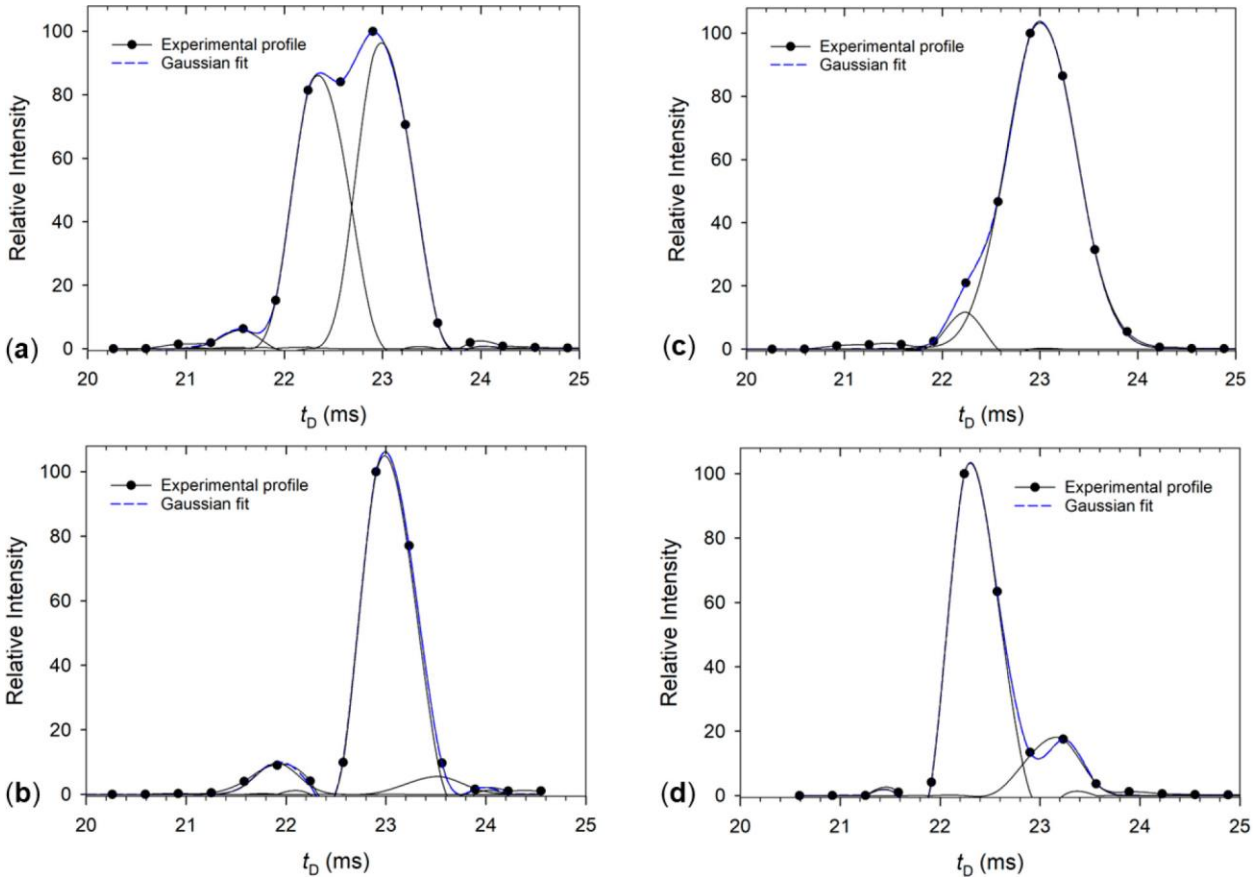


Figure 4.12: Arrival time profiles from single-pass c-IMS of tetracations. (a) d(GCGAAGC)$^{4+}$, (b) d(CGAAGCG)$^{4+}$, (c) d(GCGAACG)$^{4+}$, and (d) d(CGGAAGC)$^{4+}$.

4.3.3 Ion Structures and Theoretical CCS

Because of the aforementioned challenges to obtaining fully optimized structures for the heptanucleotide ions, we focused on calculations of the d(GCGAAGC) miniloop and d(CGAAGCG) reversed miniloop sequence ions. Optimized structures were obtained for dications, trications, and tetracations that along with atomic charge densities were used to calculate theoretical CCS in nitrogen. The relative energies and theoretical CCS are compiled in **Table 4.2**. As mentioned above, the Mobcal-MPI-calculated CCS was smaller than those from the MobCal calculations and, in general, provided better fits with the experimental data. Therefore, we limit the discussion to the MPI theoretical data.

Table 4.2 d(GCGAAGC) Dication Relative Energies and Calculated CCS

dication	$\Delta H0^{a,b}$	$\Delta G_{310}{}^{a,c}$	$CCS_{TM}{}^{d,e}$	$CCS_{MPI}{}^{d,f}$
$G1A4a_{500}$	23	19	444	428
$G1G6_{500}$	166	153	446	431
$G1A4b_{500}$	98	95	449	433
$G1G3b_{500}$	0	0	455	445
$G1G3c_{500}$	57	49	469	458
$G1A5_{300}$	115	101	472	458
$A4G6_{600}$	181	172	471	458
$G1G3a_{300}$	255	243	484	471

[a]In kJ mol^{-1}.
[b]Including zero-point vibrational energies and referring to 0 K.
[c]Including 310 K enthalpies and entropies.
[d]In $Å^2$.
[e]From standard ion trajectory calculations using Mulliken charge densities.
[f]From MobCal-MPI calculations using MK charge densities.

4.3.4 Dications

Out of over two hundred fully optimized $d(GCGAAGC)^{2+}$ dication structures that we obtained, we selected those that had theoretical CCS within 3% of the components that were identified by c-IMS measurements. At the same time, the DFT calculations provided ion relative energies that were used to assess the possibility of protomers and conformers coexisting under equilibrium conditions. The structures are labeled with the initial protonation sites (e.g., G1G3) and temperature that were used in the BOMD trajectory calculations. All the presented optimized d(GCGAAGC) ion structures are oriented with G1 in the lower left corner and the loop going clockwise. Among the several G1G3 dication structures that were optimized from the 300 K BOMD trajectories, the lowest-energy structure (**$G1G3a_{300}$**) preserving the essential features of the miniloop is shown in **Figure 4.13**. **$G1G3a_{300}$** showed WC pairing of G1 with C7 and C2 with G6 which was enforced by multiple hydrogen bonds that are visualized by ochre double headed arrows in **Figure 4.13a**. Hydrogen bonds involving the backbone are shown as light green arrows (**Figure 4.13a**). The deoxyribose-phosphoester backbone formed a flat loop, as shown by the

extracted chain of the O5′-C5′-C4′-C3′-O3′-P segments (**Figure 4.13b,c**). However, the calculated properties of **G1G3a$_{300}$** strongly indicated that it was not the structure belonging to any of the major observed d(GCGAAGC)$^{2+}$ dications. In particular, **G1G3a$_{300}$** was 243 kJ mol^{-1} less stable than the isomer corresponding to the global energy minimum (**G1G3b$_{500}$**), and its CCS$_{MPI}$ (471 Å^2) was larger than the experimental values for the major components (**Table 4.1** and **4.2**). The similarity between the calculated CCS of **G1G3a$_{500}$** and that of the minor components in the c-IMS data (CCS$_{fit}$ = 466 Å^2, Δ_{rel} = 1.1%, and CCS$_{fit}$ = 478 Å^2, Δ_{rel} = −1.5%) may indicate a small fraction of kinetically trapped miniloop ions, although we do not have other supporting evidence for such a conclusion.

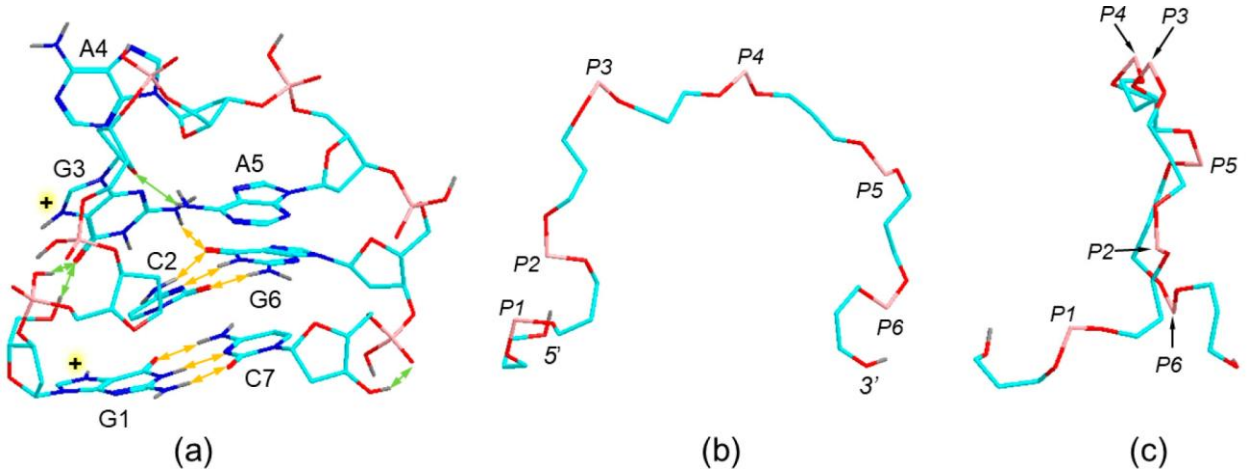


Figure 4.13: (a) B3LYP/6-31G(d,p) + GD3BJ optimized structure of the d(GCGAAGC)$^{2+}$ miniloop dication isomer **G1G3a$_{300}$**. Atom color coding is as follows: cyan = C, blue = N, red = O, pink = P, gray = H. Only exchangeable hydrogen atoms are shown to reduce clutter. Hydrogen bonds between nucleobases are shown as double-headed ochre arrows. Hydrogen bonds involving the backbone are shown as light green arrows. (b) View of the **G1G3a$_{300}$** backbone consisting of the **G1G3a$_{300}$** backbone consisting of the O5–C5–C4–C3–O3–P segments starting from the 5′ terminus. (c) **G1G3a$_{300}$** backbone rotated by 90°.

The lowest-energy isomer (**G1G3b$_{500}$**) had a CCS$_{MPI}$ = 445 Å^2 that closely matched that of the most abundant peak of d(GCGAAGC)$^{2+}$ at a CCS = 450 Å^2 from the ion mobility measurements (**Table 4.1**). Upon BOMD, ion **G1G3b$_{500}$** underwent a proton transfer from P2-OH onto N1 in A5, forming a zwitterion structure (**Figure 4.14**). The P2 anion was stabilized by hydrogen bonds to protonated G3 and A5. A major distortion in **G1G3b$_{500}$** from the miniloop conformation can be seen in (i) the outward rotation of C2 that rearranged its hydrogen bonding to G6, (ii) π-stacking of A4 with charged G3, and (iii) the absence of π-stacking of A5 with G6. Notably, hydrogen

bonds involving the phosphoester and P2-phosphate groups contributed to both a more compact ion structure and very significant energy stabilization compared to the miniloop structure.

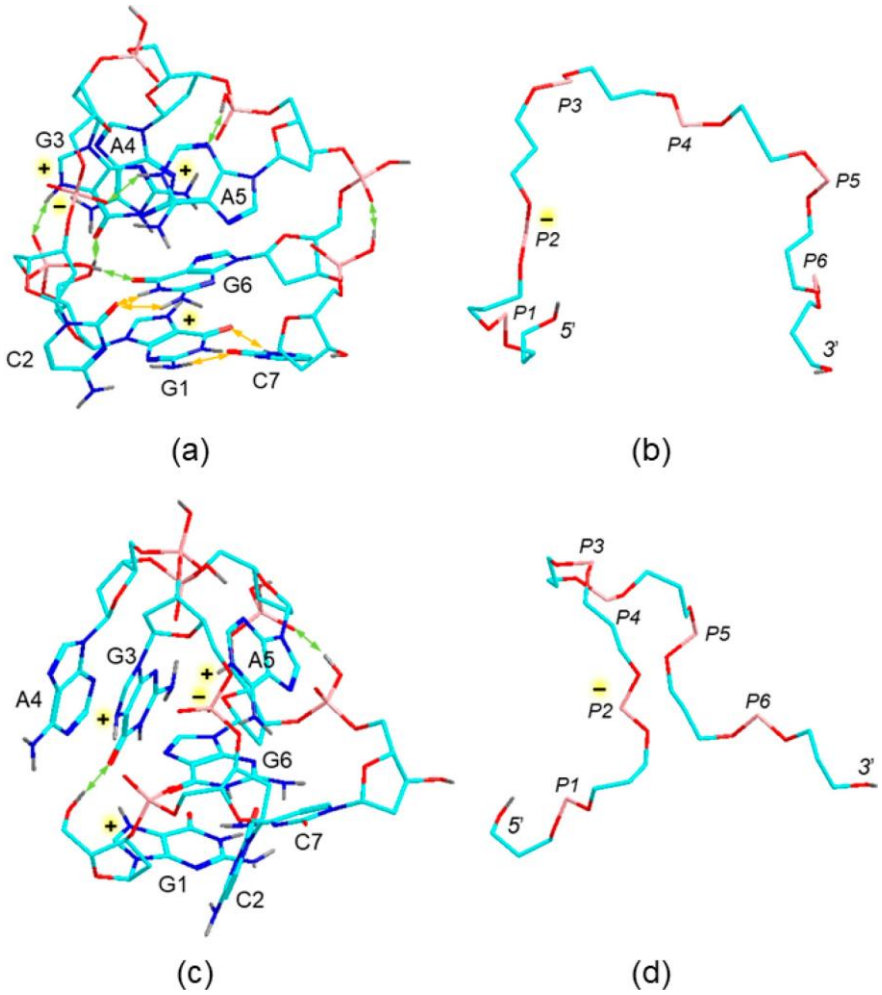


Figure 4.14: (a, c) Different views of the lowest-energy dication **G1G3b$_{500}$**. (b, d) Views of the backbone skeleton. Structure description is as in Figure 4.13. The phosphate anion is marked with a minus sign.

Distortion of hydrogen bonding between C2 and G6 was a major feature of the structure of **G1A4a$_{500}$** which was the second lowest-energy isomer at ΔG_{310} = 19 kJ mol^{-1} relative to **G1G3b$_{500}$** (**Figure 4.15**). Ion **G1A4a$_{500}$** retained the initial protonation sites at N7 of G1 and at N3 of A4. The ion shape can be seen best in a rotated projection (**Figure 4.15b**) that shows C2 that was rotated by almost 90° with respect to G6 and connected to it by a single O2···H–N1 hydrogen bond. Another distinct feature of **G1A4a$_{500}$** was the strong hydrogen bond at 1.354 Å of a Zundel type[58] between the P1-OH and O6 at G6, which was similar to that in **G1G3b$_{500}$**. The CCS$_{MPI}$ of **G1A4a$_{500}$** (428 Å^2) indicated it was the most compact dication structure of those studied; however,

this CCS_{MPI} was outside the experimental CCS range from the ion mobility measurements (**Table 4.1**). Fits to the second IMS peak of CCS = 459 $Å^2$ were realized by structures **G1G3c$_{500}$**, **G1A5$_{300}$**, and **A4G6$_{600}$** that each had CCS_{MPI} = 458 $Å^2$. Ion **G1G3c$_{500}$** which was 49 kJ mol^{-1} above **G1G3b$_{500}$**, was a zwitterion in which the acidic proton from P2-OH migrated to N1 of A5. This isomer showed a distortion of the C2-G6 pair, in which C2 was perpendicular to the G1-C7 pair. In addition, A4, G3, and A5 were arranged in a π-stacking pattern (**Figure 4.16**). Ion **G1A5$_{300}$**, had the best preserved miniloop conformation, which showed WC pairs of G1-C7 and C2-G6. A noteworthy feature of this dication was that its protonation structure isomerized to a zwitterion, in which the P1-OH proton migrated to N7 at G3 (**Figure 4.17**).

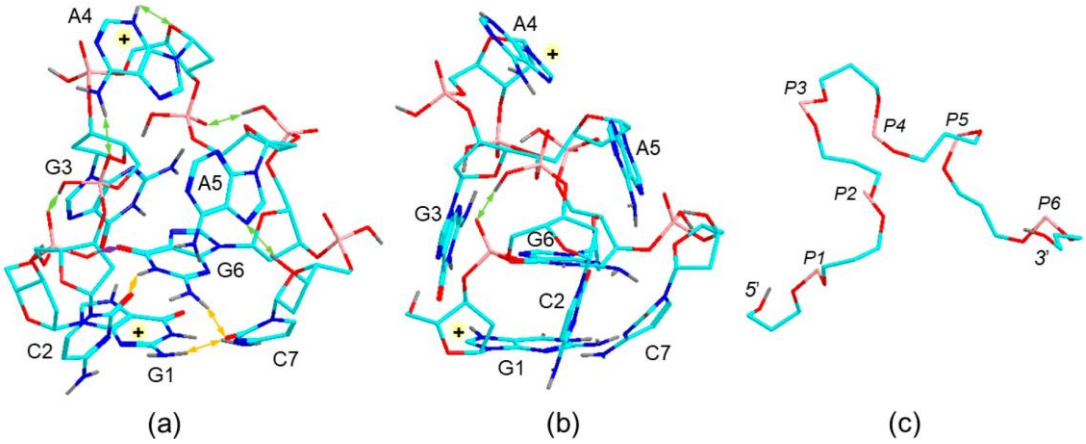


Figure 4.15: (a, b) Different views of d(GCGAAGC)$^{2+}$ dication **G1A4a$_{500}$**. (c) View of the **G1A4a$_{500}$** backbone skeleton in the (b) orientation. Structure description as in Figure 4.13.

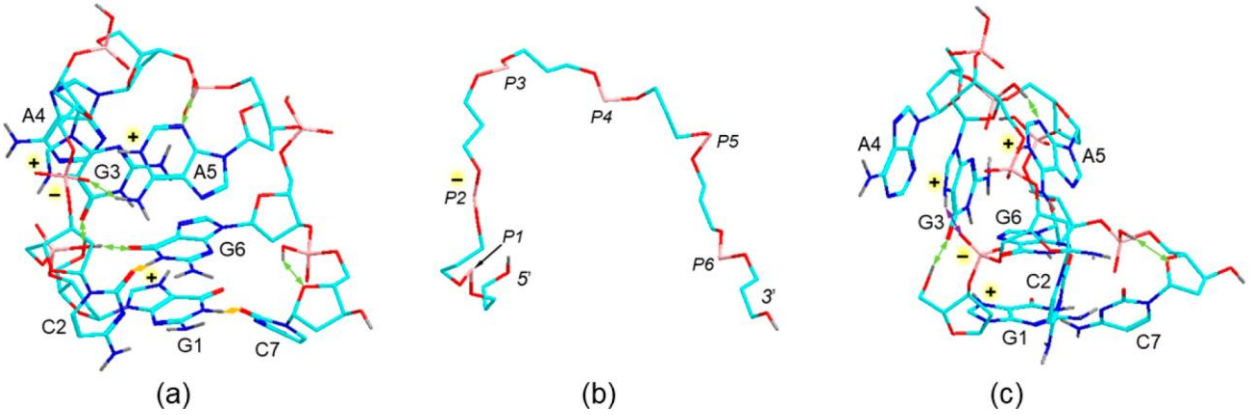


Figure 4.16:. (a, c) Different views of the d(GCGAAGC)$^{2+}$ dication **G1G3c$_{500}$**. (b) View of the **G1G3c$_{500}$** backbone skeleton in the (a) orientation. Structure description as in Figure 4.13.

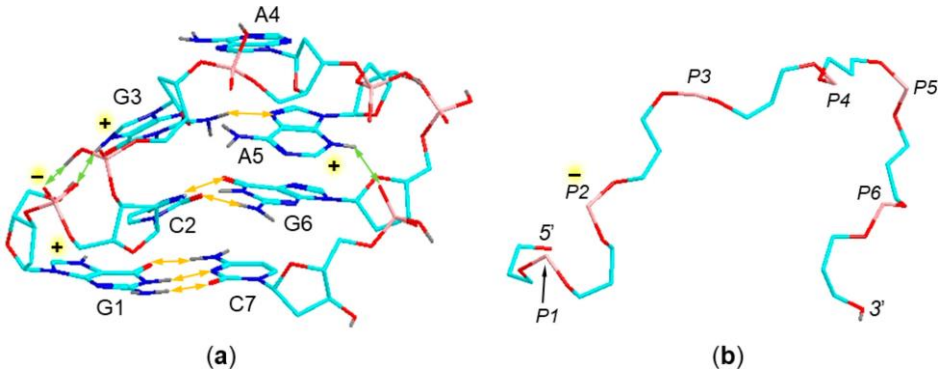


Figure 4.17: (a) Optimized structure of d(GCGAAGC)$^{2+}$ dication **G1A5$_{300}$**. (b) View of the **G1A5$_{300}$** backbone skeleton. Structure description as in Figure 4.13.

However, its high Gibbs energy, ΔG_{310} = 101 kJ mol^{-1} relative to **G1G3b$_{500}$** (**Table 4.2**), clearly indicated that this miniloop dication could not coexist with the low-energy isomers under equilibrium conditions. Its potential presence in the ion mixture could be due to kinetic factors that prevented rearrangement of the miniloop during ion formation and storage. It is noteworthy that in contrast to the other miniloop structure (**G1G3a$_{300}$**), **G1A5$_{300}$** had the P1 anion on the periphery where it was stabilized by hydrogen bonds with P2-OH and N7–H at G3. This prevented protonation at G6 and possibly contributed to the retention of the G1-C7 and C2-G6 WC pairs.

Ion **A4G6$_{600}$** (CCS_{MPI} = 458 Å^2) was another zwitterion that was generated by BOMD at 600 K followed by DFT geometry optimization, which resulted in a substantial structure rearrangement. The optimized structure of **A4G6$_{600}$** indicated that the P3-OH proton migrated to C7, resulting in a severely distorted structure in which C2 was remote from G1 and became hydrogen bonded to A5 (**Figure 4.18**). The structure was dominated by multiple internal solvations of the P3 anion by hydrogen bonds from the protonated nucleobases A4, G6, and C7. The latter was realized as an O2-protonated tautomer in **A4G6$_{600}$**. Overall, ion **A4G6$_{600}$** was 172 kJ mol^{-1} less stable than the global energy minimum and represented another high-energy isomer.

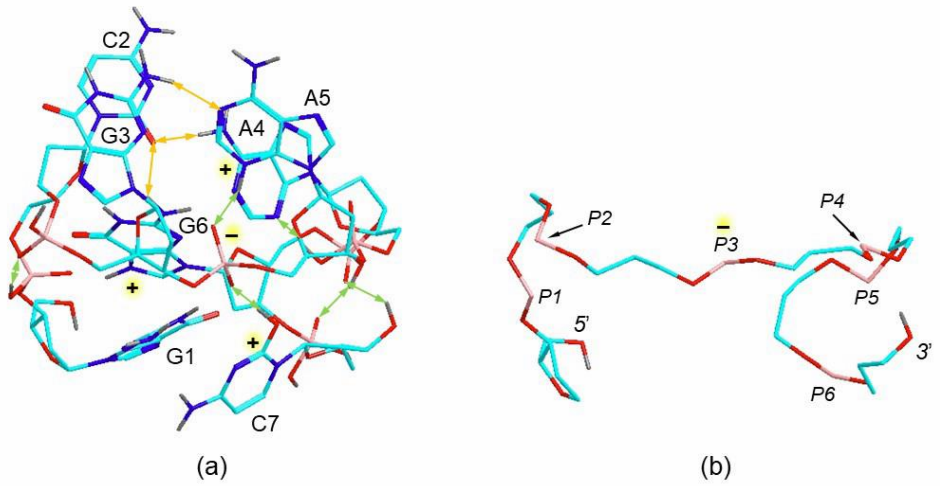


Figure 4.18: (a) B3LYP/6-31G(d,p) + GD3BJ optimized structure of dication **A4G6$_{600}$**. (b) Backbone skeleton of **A4G6$_{600}$**. Structure description as in Figure 4.13.

Two other compact dication structures, **G1A4b$_{500}$** (CCS_{MPI} = 433 Å^2) and **G1G6$_{500}$** (CCS_{MPI} = 431 Å^2) were less stable than **G1G3b$_{500}$** by 95 and 153 kJ mol^{-1}, respectively. Both of these structures showed a complete disruption of the A5-G6 π-stacking (**Figure 4.19**). In addition, partial or complete disruption of G-C hydrogen bonds was apparent for **G1A4b$_{500}$** and **G1G6$_{500}$**, respectively, which probably contributed to their higher Gibbs energy. Note that **G1G6$_{500}$** was a zwitterion formed by the transfer of the acidic proton from P1–H to O6 of G3.

To summarize the aforementioned dication structures, all low-energy d(GCGAAGC)$^{2+}$ isomers showed major deviations from the original miniloop conformation. When preserved, miniloop features with two WC G-C pairs were represented by high-energy isomers in the gas-phase dications. The overall abundance of low-energy zwitterionic structures was remarkable, albeit in line with previous structure studies of tri- and tetranucleotide dications. [24–27,29] This appears to support the general conclusion that the formation of oligonucleotide zwitterions by the transfer of a proton from phosphate ester groups onto nucleobases is energetically favorable in gas-phase ions. Both the zwitterions and canonical dications developed a number of strong hydrogen bonds that involved the deoxyribose and phosphate groups. This contrasted the solution miniloop structure[31,32] in which the phosphates were located on the heptanucleotide perimeter where they were exposed to stabilizing solvation by water. In the absence of solvent, internal solvation in gas-phase dications involving the phosphoester OH groups became the structure determining factor. Furthermore, the calculated CCS indicated that the collapse of the miniloop

structure in gas-phase ions resulted in more compact structures when compared to those in which the miniloop was substantially or partially preserved.

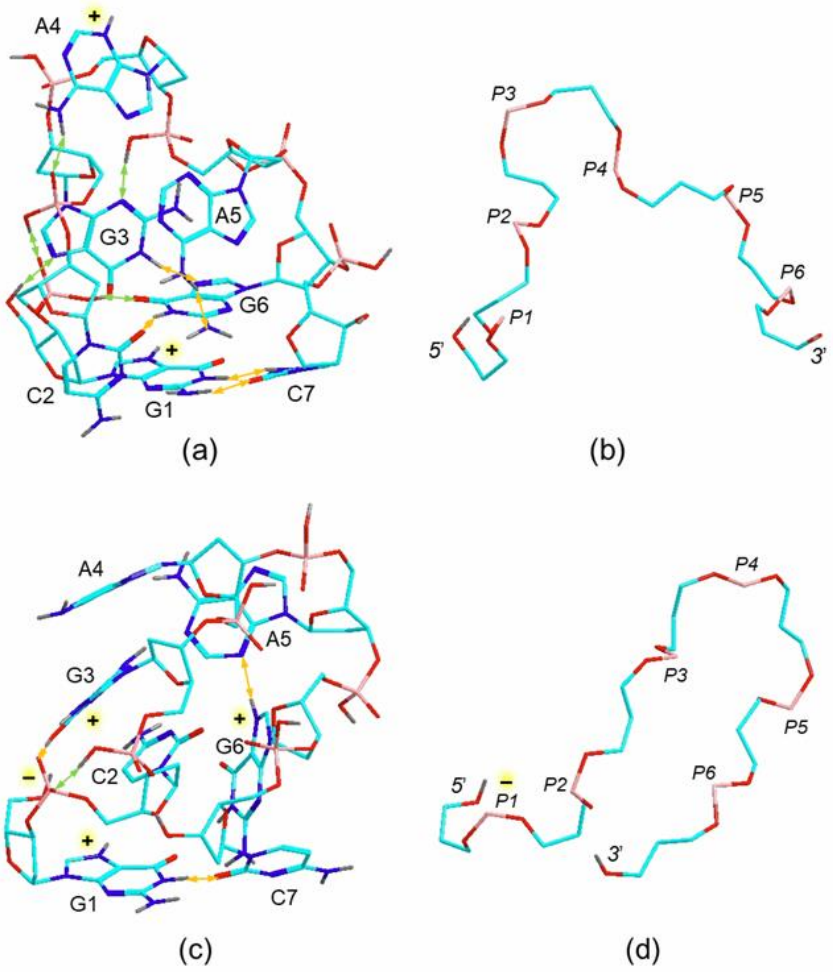


Figure 4.19: (a, c) B3LYP/6-31G(d,p) + GD3BJ optimized structures of d(GCGAAGC) dication isomers **G1A4b$_{500}$** and **G1G6$_{500}$**, respectively. Structure description as in Figure 4.13.

In contrast to d(GCGAAGC), there is no established solution or crystal structure for the reversed d(CGAAGCG) sequence. Our BOMD conformational search of the d(CGAAGCG)$^{2+}$ dications started from random conformations in which the protonation patterns followed the G2G5, G2G7, G5G7, G2A3, G2A4, A3G5, A4G5, A3G7, and A4G7 combinations in which protons were placed at guanine N7 and adenine N3. Several DFT-optimized structures were obtained, which all showed disordered conformations. Selected low-energy structures are displayed in **Figure 4.20**. The MobCal calculated CCS of these low-energy d(CGAAGCG)$^{2+}$ isomers were in the 476–544 Å^2 range, which significantly exceeded the IMS data (440–470 Å^2, **Table 4.1**). A closer agreement

was achieved by implementing the MobCal-MPI scheme that gave CCS_{MPI} in the 460–482 $Å^2$ range (**Figure 4.20**). In particular, the low Gibbs-energy ions G5G7–68 and G2G7–156 had calculated CCS_{MPI} = 460 and 466 $Å^2$, respectively, which were very close to those of the most abundant peak in the c-IMS spectrum at CCS = 459 $Å^2$. With the lowest-energy ion A3G5–185, the MobCal-MPI-calculated CCS_{MPI} = 474 $Å^2$ was close to that of the second most abundant peak from c-IMS with a CCS = 470 $Å^2$. Although the structure assignment for $d(CGAAGCG)^{2+}$ ions may not be definite when based on the CCS alone, the presence of multiple isomers as revealed by c-IMS, as well as the calculated structures, indicated that gas-phase dications preferred disordered geometries with no tendency for multiple WC pairs.

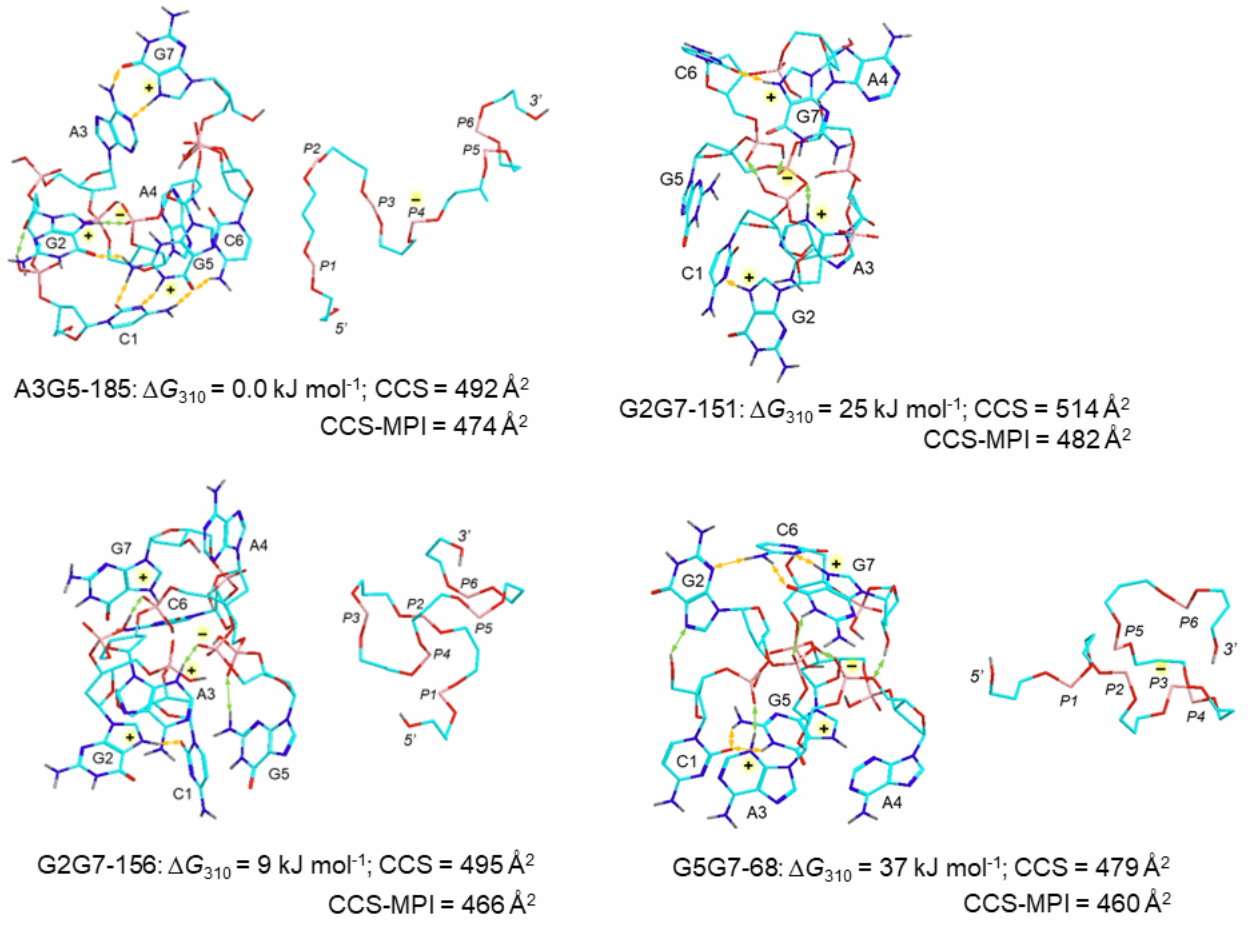


Figure 4.20: (a) B3LYP/6-31G(d,p) + GD3BJ optimized structures of d(CGAAGCG) dications, relative Gibbs energies, and calculated CCS. Structure description as in Figure 4.13.

4.3.5 Trications

d(GCGAAGC)$^{3+}$ initial ion structures were constructed by placing the charging protons on the G and A bases. The optimized structures to be discussed were selected according to the match of their calculated CCS_{MPI} with those from IMS. The lowest-energy structure **GAG$_{500}^{3+}$** had CCS_{MPI} = 508 Å^2 which provided a close match ($\Delta = -0.2\%$) to the most abundant peak in the ion mobility spectrum (509 Å^2). **GAG$_{500}^{3+}$** was a zwitterion in which the P2-OH proton moved to N3 at A5 (**Figure 4.21a**). In addition, the hydrogen-bonding pattern of multiple phosphoester groups favored a proton shift from N7 at G3 to N3 at C7. Multiple hydrogen bonds including the phosphoester OH resulted in a highly coiled conformation, as shown by the backbone projection in **Figure 4.21b**.

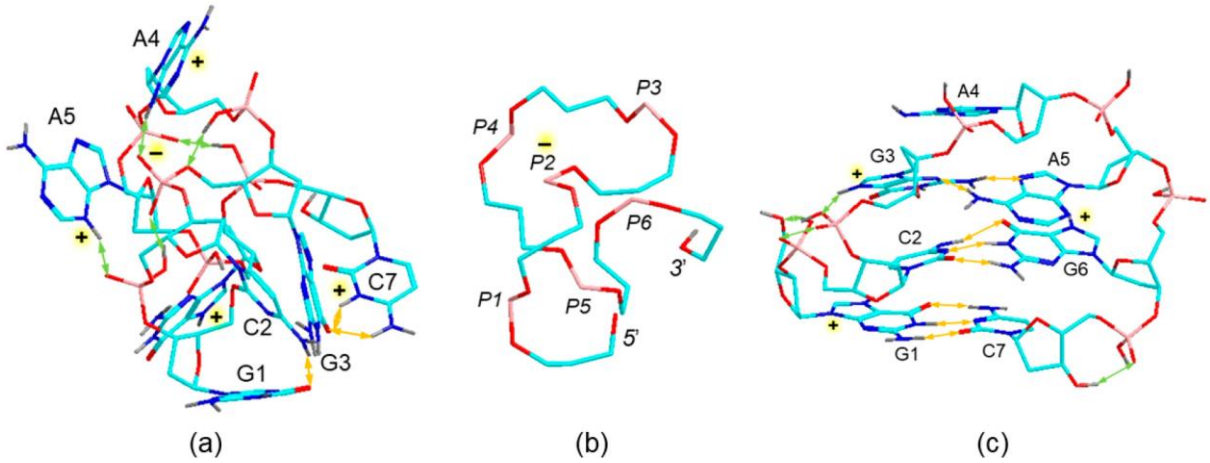


Figure 4.21: Optimized structures of d(GCGAAGC)$^{3+}$ trications (a) **GAG$_{500}^{3+}$** and (c) **GGGa$_{300}^{3+}$**. (b) View of the **GAG$_{500}^{3+}$** backbone skeleton. Structure description as in Figure 4.13.

A stacked trication, **GGGa$_{300}^{3+}$** was obtained whose CCS_{MPI} = 509 Å^2 also was a close match to the major peak in the ion mobility spectrum (CCS = 509 Å^2). Ion **GGGa$_{300}^{3+}$** was a canonical structure with protonation at the N7 positions in G1, G3 and G6 that retained the main features of the miniloop structure, including the G1-C7 and C3-G6 Watson–Crick pairs and A5-G6 ring π-stacking (**Figure 4.21c**). However, **GGGa$_{300}^{3+}$** was a high-energy isomer at ΔG_{310} = 180 kJ mol^{-1} relative to **GAG$_{500}^{3+}$** and, thus, represented a thermodynamically unstable conformation in the gas phase. A related miniloop structure, **GGGb$_{300}^{3+}$** (**Figure 4.22a,b**), which had ΔG_{310} = 126 kJ mol^{-1} relative to **GAG$_{500}^{3+}$**, was thermodynamically more favorable than **GGGa$_{300}^{3+}$** and had a

matching CCS_{MPI} = 511 Å^2 that fitted the second peak in the ion mobility data. Overall, there were several trication structures that gave very close CCS_{MPI} values that would be indistinguishable by c-IMS. Regarding their relative Gibbs energies, the global energy minimum **GAG$_{500}^{3+}$** appeared to be the most likely candidate for the gas-phase ion.

Among the d(CGAAGCG)$^{3+}$ ion structures, we found ion **GA3G$_{500}^{3+}$** that was the global energy minimum at ΔG_{310} = −57 kJ mol^{-1} relative to **GAG$_{500}^{3+}$**. **GA3G$_{500}^{3+}$** was a canonical trication with protonation sites at G2, G5, and G7 (**Figure 4.22c,d**). It showed that A3, A4, and G5 clustered at a backbone loop, whereas C1-G2 and C6-G7 formed local hydrogen-bonded pairs. However, this hydrogen bonding was mediated by the guanine N7 protons rather than the usual WC pattern.

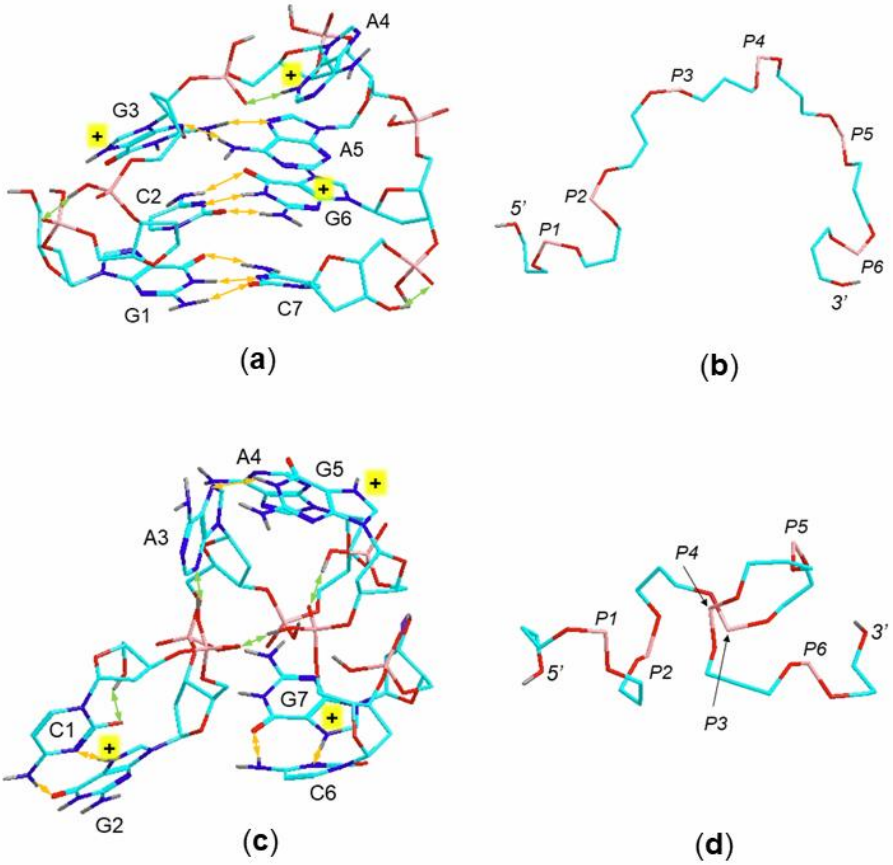


Figure 4.22: B3LYP/6-31G(d,p) + GD3BJ optimized structures of (a) d(GCGAAGC) trication **GGGb$_{300}^{3+}$**, and (c) d(CGAAGCG) trication **GA3G$_{500}^{3+}$** with the respective extracted backbones in (b) and (d). Structure description as in Figure 4.13.

4.3.6 Tetracations

In contrast to dications and trications, the structures of the d(GCGAAGC) and d(CGAAGCG) tetracations did not preserve the miniloop conformations. Two low-energy d(GCGAAGC)$^{4+}$ ions, **GGAG1**$^{4+}$ and **GGAG2**$^{4+}$, both showed a G1-C7 WC pair, while the C2-G6 moiety was disordered (**Figure 4.23**).

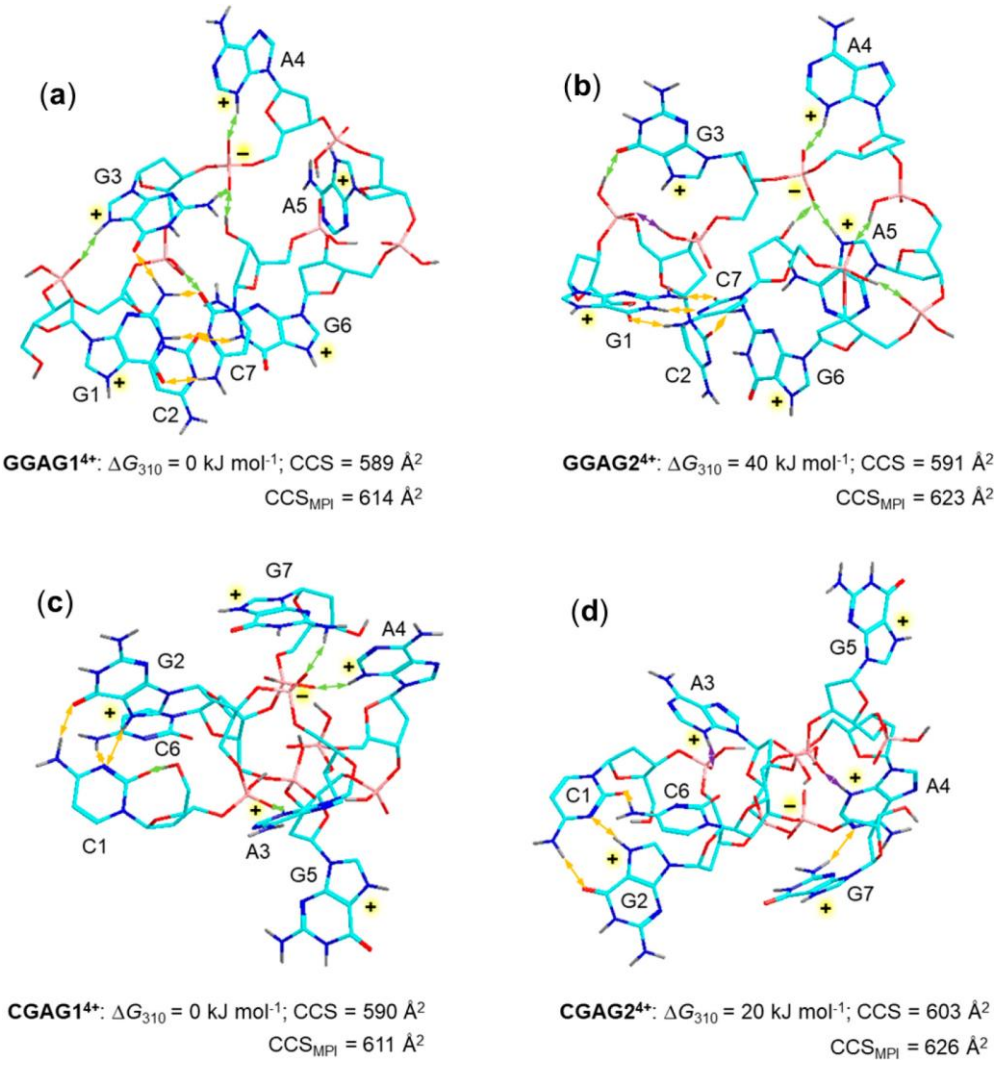


Figure 4.23: Optimized structures of tetracations (a) **GGAG1**$^{4+}$, (b) **GGAG2**$^{4+}$, (c) **CGAG1**$^{4+}$, and (d) **CGAG2**$^{4+}$, with relative Gibbs energies and calculated CCS. Structure description is in Figure 4.13.

Remarkably, and despite the 4-fold protonation, these low energy structures were zwitterions in which all guanines and adenines were protonated while the P3 phosphate was deprotonated. The

calculated CCS_{MPI} of **GGAG1^{4+}** (614 Å^2) was close to that of the first abundant component in c-IMS (608 Å^2; **Figure 4.12a**). The CCS_{MPI} of **GGAG2^{4+}** (623 Å^2) was a close match with the second abundant c-IMS peak at 624 Å^2. These CCS-matching structures indicated a substantial distortion of the ion conformation upon 4-fold protonation. The low-energy reversed-sequence tetracations, **CGAG1^{4+}** and **CGAG2^{4+}**, were also zwitterions (**Figure 4.23**). **CGAG2^{4+}** showed a C1-G2 WC pair, while the other protonated nucleobases were on the ion periphery. The calculated CCS_{MPI} for **CGAG2^{4+}** (626 Å^2) gave a close match with the CCS of the most abundant c-IMS peak at 625 Å^2 (Table 1), while that for the lowest-energy tetracation **CGAG1^{4+}** (CCS_{MPI} = 611 Å^2) was 2.2% underestimated relative to that for the major d(CGAAGCG)$^{4+}$ component.

4.3.7 Ion Dissociations in CID-MS^2 Spectra

Although the ion mobility data indicated isomer mixtures for the heptanucleotide cations, it appeared to be of interest to investigate their dissociations upon collisional activation. The spectra are briefly described here, and the major dissociations are summarized in **Table 4.3**. CID of the cations resulted in the combined loss of nucleobase and backbone cleavages, analogous to those reported earlier for di-, tri-, and tetranucleotide cations.[27,29] The dications showed a loss of neutral nucleobases that preferentially occurred from the terminal positions but also depended on the sequence. Nucleobase loss from the 3′-position can be recognized by the associated loss of water, which is specific for this terminal position.[29] CID-MS^2 of d(GCGAAGC)$^{2+}$ showed a major loss of 5′-guanine whereas losses of cytosine and adenine were minor (**Figure 4.24a**). CID-MS^2 of the reverse-sequence dication d(CGAAGCG)$^{2+}$ showed a major loss of 3′-guanine that was accompanied by loss of cytosine and adenine (**Figure 4.25a**). Similarly, the scrambled-sequence dication d(CGGAAGC)$^{2+}$ showed a loss of cytosine that originated from both the 3′- and 5′-positions (**Figure 4.27a**). In this case the lower relative intensity of the combined loss of cytosine and water indicated that the major fraction of cytosine loss occurred from the 5′-position. In contrast, the other scrambled-sequence dication d(GCGAACG)$^{2+}$ with terminal guanines showed a major loss of adenine from one of the internal positions that was accompanied by the loss of guanine (**Figure 4.26a**). It should be noted that loss of the nucleobase requires that there be an available proton either placed on the base during ionization or transferred from another protonated base upon collisional excitation. These scenarios are not distinguished in the CID spectra, because

proton transfer in oligonucleotide cations is a low energy process that can proceed prior to dissociation.[29] This was further investigated with the current heptanucleotide ion set by generating the lower charge states by proton-transfer reactions [34] of higher-charge states with the fluoranthene anion, $C_{16}H_9^-$. d(GCGAAGC)$^{2+}$ was generated in the ion trap by proton transfer from the mass-selected +3 and +4 states and its CID-MS^3 spectra were obtained. These were found to be nearly identical with the CID-MS^2 spectrum of d(GCGAAGC)$^{2+}$ produced by electrospray, indicating that the proton locations in the dication do not have a major effect on the dissociations after collisional excitation.

Table 4.3 Summary of the Ion Dissociations

Loss of Base

d(GCGAAGC)$^{2+}$	5′-G ≫ 3′-C, A
d(GCGAAGC)$^{3+}$	5′-GH$^+$ > 3′-CH$^+$, very weak loss of AH$^+$
d(GCGAAGC)$^{4+}$	5′-GH$^+$ > 3′-CH$^+$, no loss of AH$^+$
d(CGAAGCG)$^{2+}$	3′-G (-H_2O) > C > A
d(CGAAGCG)$^{3+}$	3′-G (-H_2O) > C > A
d(CGAAGCG)$^{4+}$	5′-C
d(GCGAACG)$^{2+}$	A > 5′-G ≫C
d(GCGAACG)$^{3+}$	GH$^+$ ≫ CH$^+$ > AH$^+$
d(GCGAACG)$^{4+}$	GH$^+$ ≫ CH$^+$, no loss of AH$^+$
d(CGGAAGC)$^{2+}$	C > G > A
d(CGGAAGC)$^{3+}$	3′-CH$^+$ ≫ GH$^+$, no loss of AH$^+$
d(CGGAAGC)$^{4+}$	CH$^+$ dominant

Backbone Cleavages

d(GCGAAGC)$^{2+}$	$\boldsymbol{w_6^{2+}}$; consecutive loss of G, C, and A; $\boldsymbol{w_5/d_5^{2+}}$; $\boldsymbol{w_5/d_5^{+}}$; $\boldsymbol{w_4^{+}}$
d(GCGAAGC)$^{3+}$	$\boldsymbol{w_6^{2+}}$, consecutive loss of C; $\boldsymbol{w_4^{+}}$
d(GCGAAGC)$^{4+}$	$\boldsymbol{d_6^{2+}}$, consecutive loss of G, $\boldsymbol{a_1}$; $\boldsymbol{w_6^{3+}}$

d(CGAAGCG)$^{2+}$	$\boldsymbol{w_6^{2+}}$, consecutive loss of G + H_2O; $\boldsymbol{w_5/d_5^{+}}$; $\boldsymbol{w_4^{+}}$
d(CGAAGCG)$^{3+}$	$\boldsymbol{w_6^{2+}}$, consecutive loss of G + H_2O; $\boldsymbol{w_5/d_5^{2+}}$; $\boldsymbol{w_5/d_5^{+}}$; $\boldsymbol{w_4^{+}}$
d(CGAAGCG)$^{4+}$	$\boldsymbol{w_6^{3+}}$; $\boldsymbol{d_6^{2+}}$, consecutive loss of C + H_2O
d(GCGAACG)$^{2+}$	$\boldsymbol{w_6/d_6^{2+}}$; $\boldsymbol{w_5/d_5^{+}}$
d(GCGAACG)$^{3+}$	$\boldsymbol{w_6/d_6^{2+}}$; $\boldsymbol{w_5/d_5^{2+}}$
d(GCGAACG)$^{4+}$	$\boldsymbol{w_6/d_6^{3+}}$; $\boldsymbol{w_6/d_6^{2+}}$, consecutive loss of C, G; $\boldsymbol{w_5/d_5^{2+}}$
d(CGGAAGC)$^{2+}$	$\boldsymbol{w_6/d_6^{2+}}$, consecutive loss of G; $\boldsymbol{w_5/d_5^{2+}}$
d(CGGAAGC)$^{3+}$	$\boldsymbol{w_6/d_6^{2+}}$, consecutive loss of C; $\boldsymbol{w_5/d_5^{2+}}$
d(CGGAAGC)$^{4+}$	$\boldsymbol{w_6/d_6^{2+}}$; $\boldsymbol{w_6/d_6^{3+}}$

CID of trications showed mainly losses of protonated nucleobases forming doubly charged fragment ions. d(GCGAAGC)$^{3+}$ showed a competitive loss of GH^+ and 3-CH^+, the latter being accompanied by loss of water (**Figure 4.24b**). The inverted trication, d(CGAAGCG)$^{3+}$, showed chiefly loss of GH^+ that was accompanied by consecutive loss of water and backbone cleavage eliminating the a1 neutral fragment and forming the abundant fragment ion at m/z 876.5 (**Figure 4.26b** and **4.27b**). The scrambled trications d(GCGAACG)$^{3+}$ and d(CGGAAGC)$^{3+}$ showed chiefly losses of the terminal nucleobases, GH^+ and CH^+, respectively, that most likely originated from both 3′- and 5′-positions, as judged by the consecutive loss of water. Interestingly, the loss of AH^+ from d(GCGAACG)$^{3+}$ was only minor (**Figure 4.26b**). The loss of a protonated nucleobase requires transfer of two protons. This can be viewed as a two step process proceeding via a complex of the neutral nucleobase with the complementary multiply charged fragment ion, in which the departing nucleobase picks up the charging proton.

CID of the tetracations gave rise to much simpler spectra compared with those of the other charge states. d(GCGAAGC)$^{4+}$ showed loss of GH^+ and 3′-CH^+ that were analogous to the dissociations of the trication (**Figure 4.24c**). The inverted tetracation, d(CGAAGCG)$^{4+}$, showed a very weak nucleobase ion loss, with the main dissociations occurring in the backbone (**Figure 4.25c**). In contrast, d(GCGAACG)$^{4+}$ and d(CGGAAGC)$^{4+}$ showed major losses of the terminal nucleobases, GH^+ and CH^+, respectively (**Figure 4.26c** and **4.27c**), which were analogous to the dissociations of the respective trications. Interestingly, the loss of GH^+ and CH^+ from the

tetracations was accompanied by much less prominent loss of water (**Figure 4.26c**, **4.27c**), suggesting that these dissociations chiefly proceeded from the 5′-positions.

Common features among the dication sequences were also observed for backbone dissociations. Most backbone cleavages occurred at the 5′-terminus, forming $\boldsymbol{w_6^{2+}}$ ions.[29] For dications, this was illustrated by d(GCGAAGC)$^{2+}$ and d(CGAAGCG)$^{2+}$ where the 5′- and 3′-termini were discerned by mass (**Figure 4.24a–c** and **4.25a–c**). With d(GCGAACG)$^{2+}$ and d(CGGAAGC)$^{2+}$ the dominant backbone fragment ions could arise from either terminus and therefore were labeled as $\boldsymbol{w_6/d_6^{2+}}$ (**Figure 4.26a**, **4.27a**). In addition to $\boldsymbol{w_6^{2+}}$, backbone dissociations leading to $\boldsymbol{w_5^{2+}}$ and $\boldsymbol{w_5/d_5^{2+}}$ ions were also observed for all dication sequences. Singly charged fragment ions, e.g., $\boldsymbol{w_5/d_5^{+}}$ and $\boldsymbol{w_4^{+}}$ represented less abundant dissociation products, as shown for d(GCGAAGC)$^{2+}$, d(CGAAGCG)$^{2+}$, and d(GCGAACG)$^{2+}$. The triply charged ions showed a very similar pattern, in which $\boldsymbol{w_6/d_6^{2+}}$ fragment ions dominated the CID-MS^2 spectra. These ions underwent consecutive base loss, e.g., loss of cytosine from $\boldsymbol{w_6}$(CGAAGC)$^{2+}$, and loss of guanine and water from $\boldsymbol{w_6}$(GAAGCG)$^{2+}$.

The quadruply charged ions showed more evident differences depending on the sequence. CID of d(GCGAAGC)$^{4+}$ formed the $\boldsymbol{d_6^{2+}}$ ion as an abundant sequence fragment by cleavage between G6 and C7, in addition to the major $\boldsymbol{w_6^{3+}}$ ion (**Figure 4.24c**). The tetracation of the inverted d(CGAAGCG)$^{4+}$ sequence also produced a $\boldsymbol{d_6^{2+}}$ fragment ion which in this case underwent further consecutive loss of 5′-cytosine and water (m/z 876.5, **Figure 4.25c**). Backbone dissociations of d(GCGAACG)$^{4+}$ and d(CGGAAGC)$^{4+}$ also proceeded near the termini, forming the $\boldsymbol{w_6/d_6^{3+}}$ and $\boldsymbol{w_6/d_6^{2+}}$ ions which were unresolved for these sequences (**Figure 4.26c**, **4.27c**). The formation of $\boldsymbol{w_6/d_6^{2+}}$ ions from the tetracations was interesting because it involved the loss of two charges within the formally complementary mononucleoside fragment. This can be accommodated by consecutive elimination of a protonated nucleobase, forming a glycal cation intermediate that can undergo standard phosphate ester elimination of the deoxyribose $C_5H_7O_2^+$ ion fragment. The loss-of-base and backbone dissociations are summarized in **Table 4.3**.

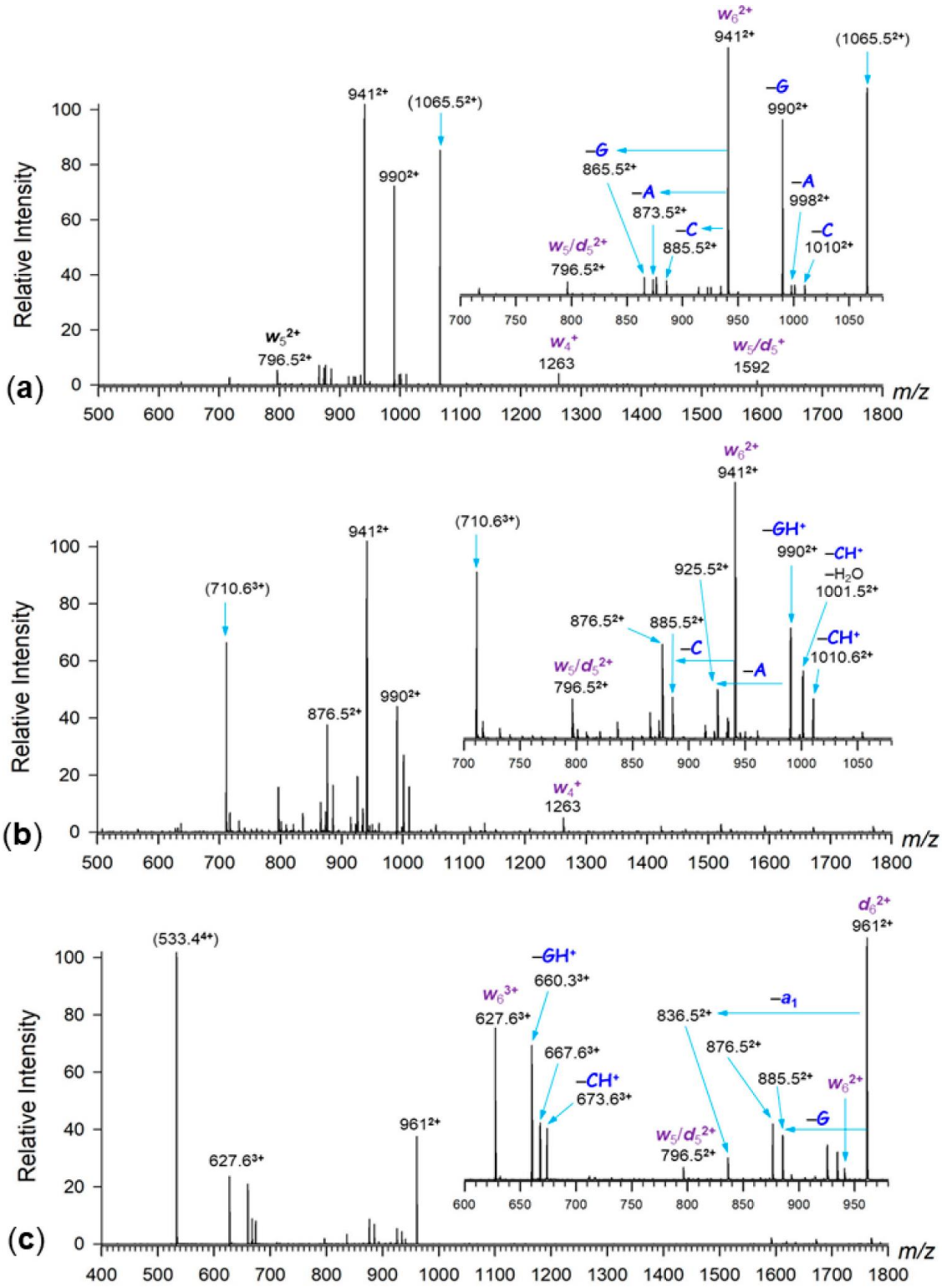


Figure 4.24: CID-MS^2 spectra of (a) GCGAAGC^{2+} at m/z 1065.5, (b) GCGAAGC^{3+} at m/z 710.6, and (c) GCGAAGC^{4+} at m/z 533.4. Insets show the expanded regions of doubly and triply charged fragment ions.

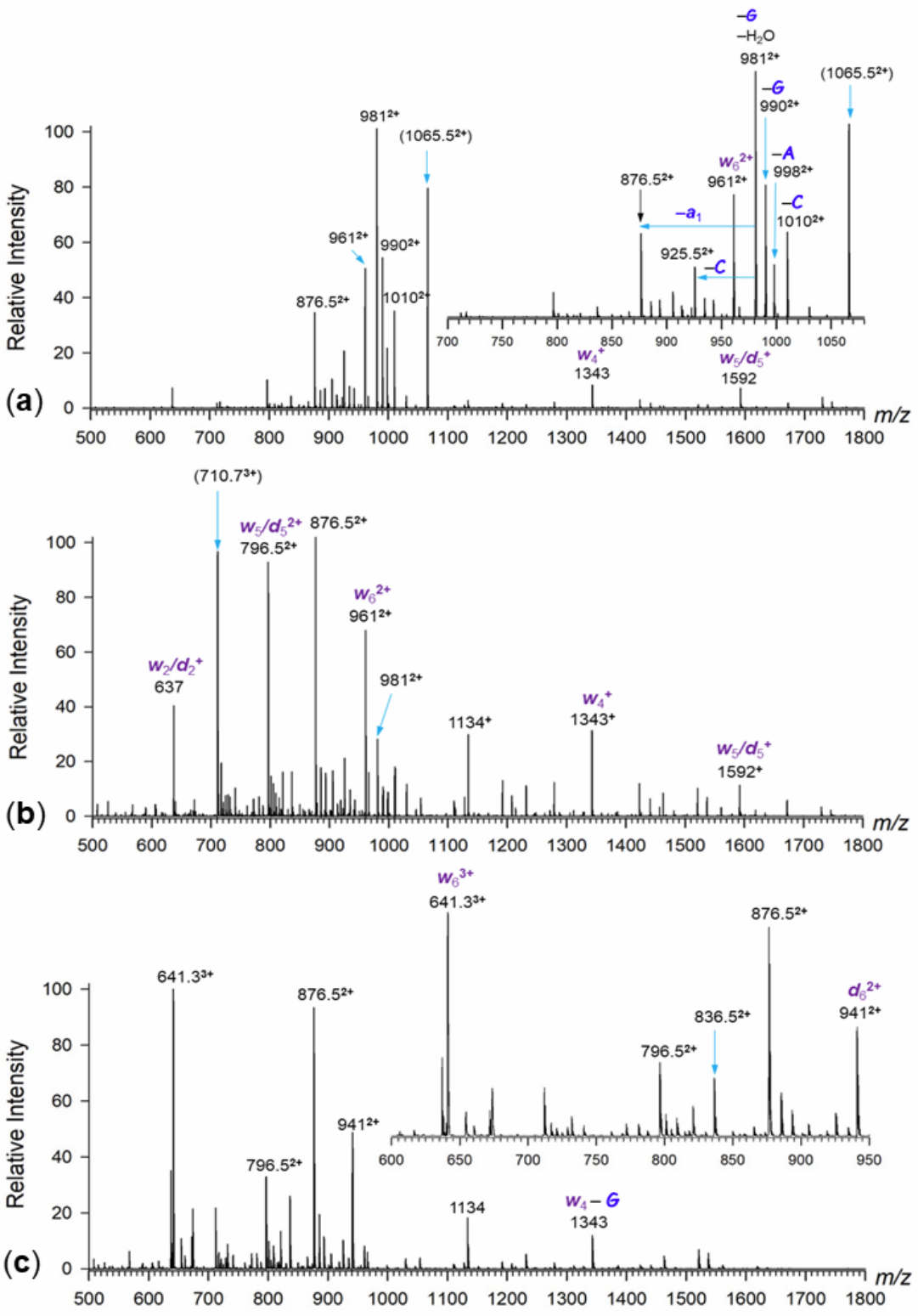


Figure 4.25: CID-MS^2 spectra of (a) d(CGAAGCG)$^{2+}$ at m/z 1065.5; (b) d(CGAAGCG)$^{3+}$ at m/z 710.7; and (c) d(CGAAGCG)$^{4+}$ at m/z 533.4; Insets show the expanded regions of doubly and triply charged fragment ions.

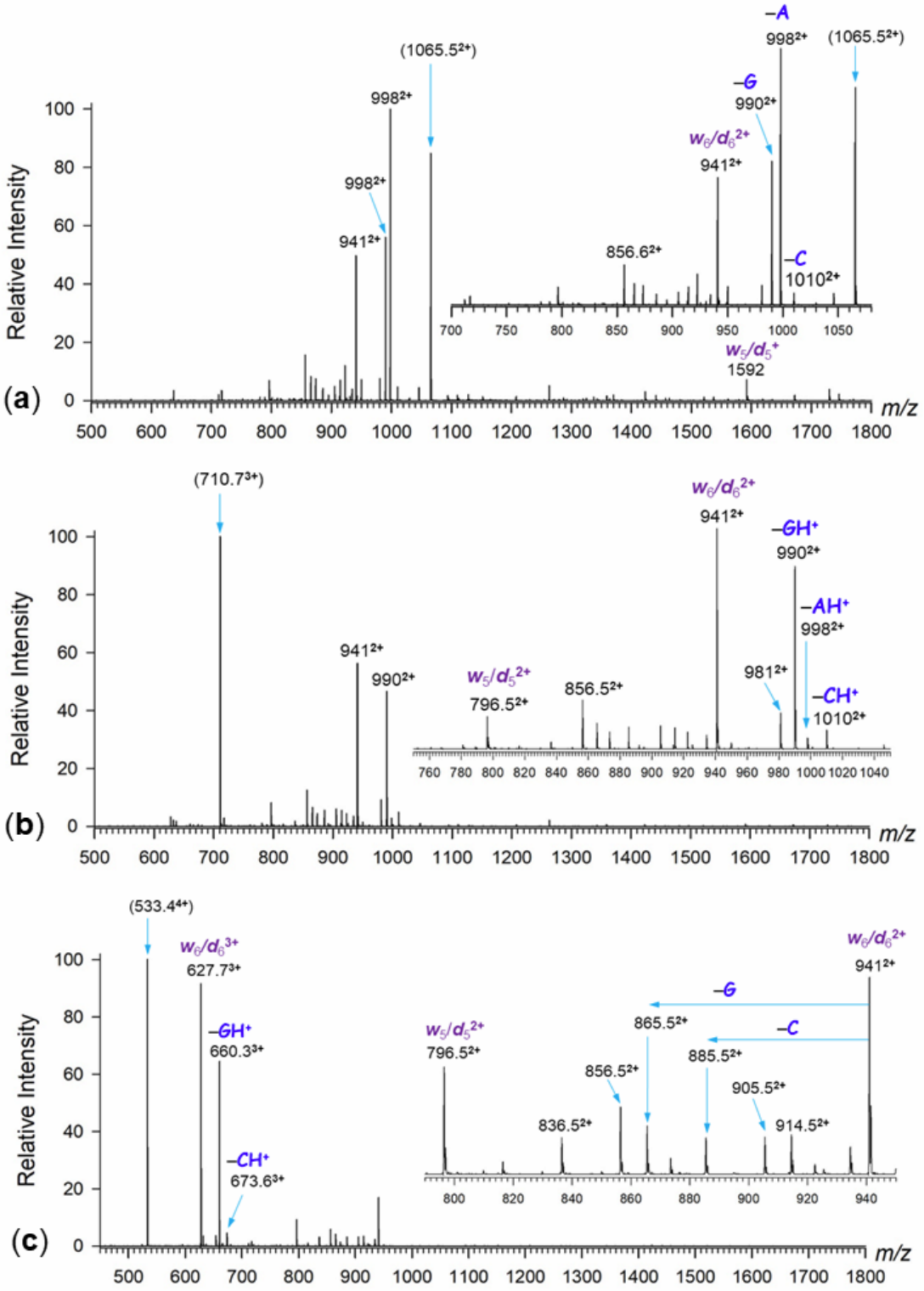


Figure 4.26: CID-MS^2 spectra of (a) d(GCGAACG)$^{2+}$ at m/z 1065.5; (b) d(GCGAACG)$^{3+}$ at m/z 710.7; and (c) d(GCGAACG)$^{4+}$ at m/z 533.4. Insets show the expanded regions of doubly and triply charged fragment ions.

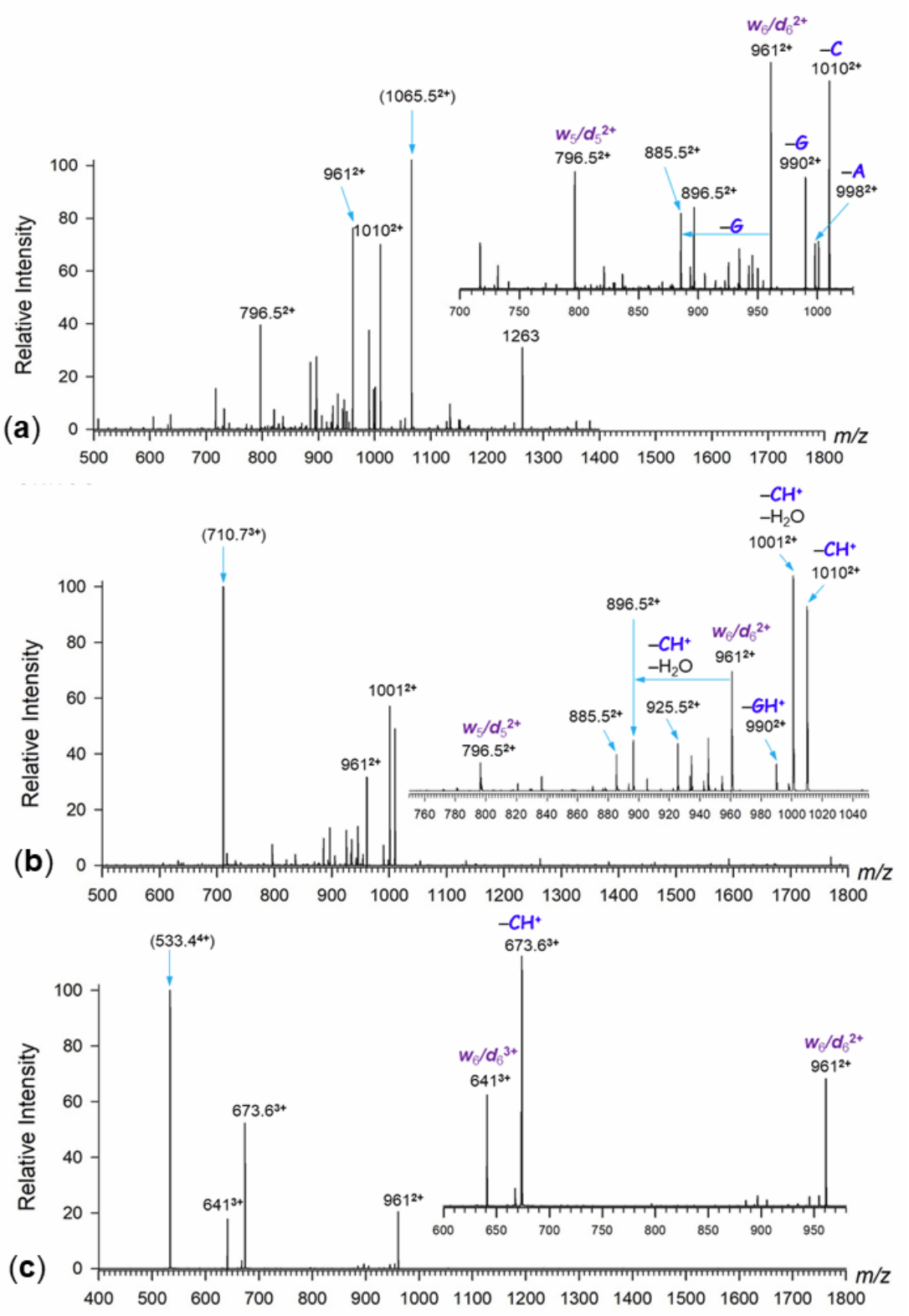


Figure 4.27: CID-MS^2 spectra of (a) d(CGGAAGC)$^{2+}$, (b) d(CGGAAGC)$^{3+}$, and (c) d(CGGAAGC)$^{4+}$. Insets show the expanded regions of doubly and triply charged fragment ions.

4.3.8 Cation-Radical Dissociations in ET-CID-MS³ Spectra

Electron transfer to oligonucleotide cations has been reported to cause dissociation at various levels, which were very low to practically none for hexanucleotides,[23] weak for tetranucleotides,[26,27] and moderate to abundant for di- and trinucleotides.[25,59] In an effort to characterize the present heptanucleotide cations, we obtained ET-CID-MS^3 spectra of various charge states that were generated by nondissociative electron transfer, e.g., one-electron $3^+ \rightarrow 2^{+\bullet}$, $4^+ \rightarrow 3^{+\bullet}$, and two-electron $4^+ \rightarrow 2^+$ charge reducing reactions. The former two one-electron reductions generated open shell systems that were potentially susceptible to radical-induced dissociations, which are briefly described here. CID of d(GCGAAGC+3H)$^{2+\bullet}$ showed two major dissociations which were loss of G and backbone cleavage forming the ($\boldsymbol{w_6}$+H)$^{2+\bullet}$ ions (**Figure 4.28a**). Both dissociations generated product ions that retained the radical sites and were analogous to dissociations observed for even-electron ions d(GCGAAGC)$^{2+}$. ET-CID differed from CID for the formation of $\boldsymbol{a_6}^+$, $\boldsymbol{a_5}/\boldsymbol{z_5}^+$, and $\boldsymbol{d_6}^{2+}$ fragment ions from d(GCGAAGC+3H)$^{2+\bullet}$, which involved loss of neutral radical fragments. CID of d(GCGAAGC+4H)$^{3+\bullet}$ showed losses of 3′-CH$^+$ and GH$^+$ that were also observed for CID of the even-electron trication d(GCGAAGC)$^{3+}$ (**Figure 4.24b**). The major differences pertinent to the radical trication included the formation of $\boldsymbol{d_6}^{2+}$ ions which resulted from the loss of $\boldsymbol{z_1}^{+\bullet}$ radical cations containing (cytosine+2H) as a proton and radical carrier. Hydrogen transfer upon dissociation was also indicated by the loss of (guanine+2H)$^{+\bullet}$ accompanying the loss of GH$^+$, and the formation of $\boldsymbol{w_6}^{2+}$ and $\boldsymbol{w_5}/\boldsymbol{d_5}^{2+}$ ions that were accompanied by ($\boldsymbol{w_6}$+H)$^{2+\bullet}$, and ($\boldsymbol{w_5}$+H)/($\boldsymbol{d_5}$+H)$^{2+\bullet}$ ions by loss of doubly charged fragments. The major distinctive feature of the CID spectrum of d(GCGAAGC+4H)$^{3+\bullet}$ was the formation of $\boldsymbol{a_5}/\boldsymbol{z_5}^{2+}$ ions by backbone cleavage. It is worth noting that the $\boldsymbol{a_5}/\boldsymbol{z_5}^{2+}$ ions were not accompanied by hydrogen atom adduct radicals in the spectrum. The ET-CID-MS^3 spectrum of the inverted sequence dication radical, d(CGAAGCG+3H)$^{2+\bullet}$ showed prominent fragment dication radicals by loss of neutral nucleobases 5′-G > A ≈ C (**Figure 4.29a**). Loss of 5′-G was accompanied by a consecutive elimination of a neutral a1 fragment, forming the prominent m/z 877 dication radical. Backbone cleavages producing the ($\boldsymbol{w_6}$+H)$^{2+\bullet}$ and ($\boldsymbol{w_5}$+H)/($\boldsymbol{d_5}$+H)$^{2+\bullet}$ ions proceeded without the involvement of radical sites that remained in the product ions. The trication radical of the inverted sequence, d(CGAAGCG+4H)$^{3+\bullet}$, showed major $\boldsymbol{w_5}/\boldsymbol{d_5}^{2+}$ ions at m/z 796.5 by loss of ($\boldsymbol{a_2}$+H)/($\boldsymbol{z_2}$+H)$^{+\bullet}$ cation radicals that appeared at m/z 540 (**Figure 4.29b**). The ET-CID-MS^3 spectra of cation radicals with

scrambled sequences, d(GCGAACG)$^{2+\bullet}$, d(GCGAACG)$^{3+\bullet}$, d(CGGAAGC)$^{2+\bullet}$, and d(CGGAAGC)$^{3+\bullet}$ also showed chiefly $\boldsymbol{w_6}$ and $\boldsymbol{w_5}$ type fragment dications and their hydrogen atom adducts, as shown in **Figure 4.30a,b** and **4.31a,b**.

The products of consecutive two-electron reduction of the tetracations, $4^+ \rightarrow 3^{+\bullet} \rightarrow 2^+$, chiefly showed ($\boldsymbol{w_6}$+2H)$^{2+}$ backbone fragment ions (**Figure 4.28c**) that were analogous to those formed from the corresponding dication radicals, allowing for the additional hydrogen atom. The main difference was that the CID-MS^3 dissociations of these +2 ions proceeded with much less hydrogen atom transfer, and all four hydrogen atoms introduced by protonation were retained in the ($\boldsymbol{w_6}$+2H)$^{2+}$ ions. Similarly, the loss of nucleobases from the 2+ ions was devoid of hydrogen atom migrations, and no loss of GH$^\bullet$, CH$^\bullet$, or AH$^\bullet$ was observed (**Figure 4.29c** and **4.30c**). This may indicate that the transfer of the second electron in the $3^{+\bullet} \rightarrow 2^+$ reduction step resulted in the formation of a closed-shell dication and not a dication diradical. Multistep electron transfer to high charge states of proteins has been claimed to produce diradicals on the basis of their reaction with residual oxygen.[60] In contrast, the dissociations of doubly reduced heptanucleotide ions reported here indicated that no diradicals were formed.

In summary, the analysis of the CID-MS^2 and ET-CID-MS^3 spectra revealed the general features of ion backbone dissociations that chiefly occurred at the 5′-terminus regardless of the nature of the nucleobase (G or C) and the heptanucleotide sequence. Similarly, preferential loss of the terminal bases (C and G) generally occurred with the exception of doubly charged d(GCGAACG) that lost one of the internal adenines from d(GCGAACG+2H)$^{2+}$, d(GCGAACG+3H)$^{2+\bullet}$, and d(GCGAACG+4H)$^{2+}$, regardless of the ion formation mode and presence or absence of nucleobase radicals. Loss of the nucleobase is known to be facilitated by base protonation,[29] indicating protonation at adenine in d(GCGAACG) dications. Nevertheless, the proton distribution in these multiply charged ions and the ion conformation were not possible to rationally link to the dissociation patterns that reflected the transition states rather than the initial ion structures.

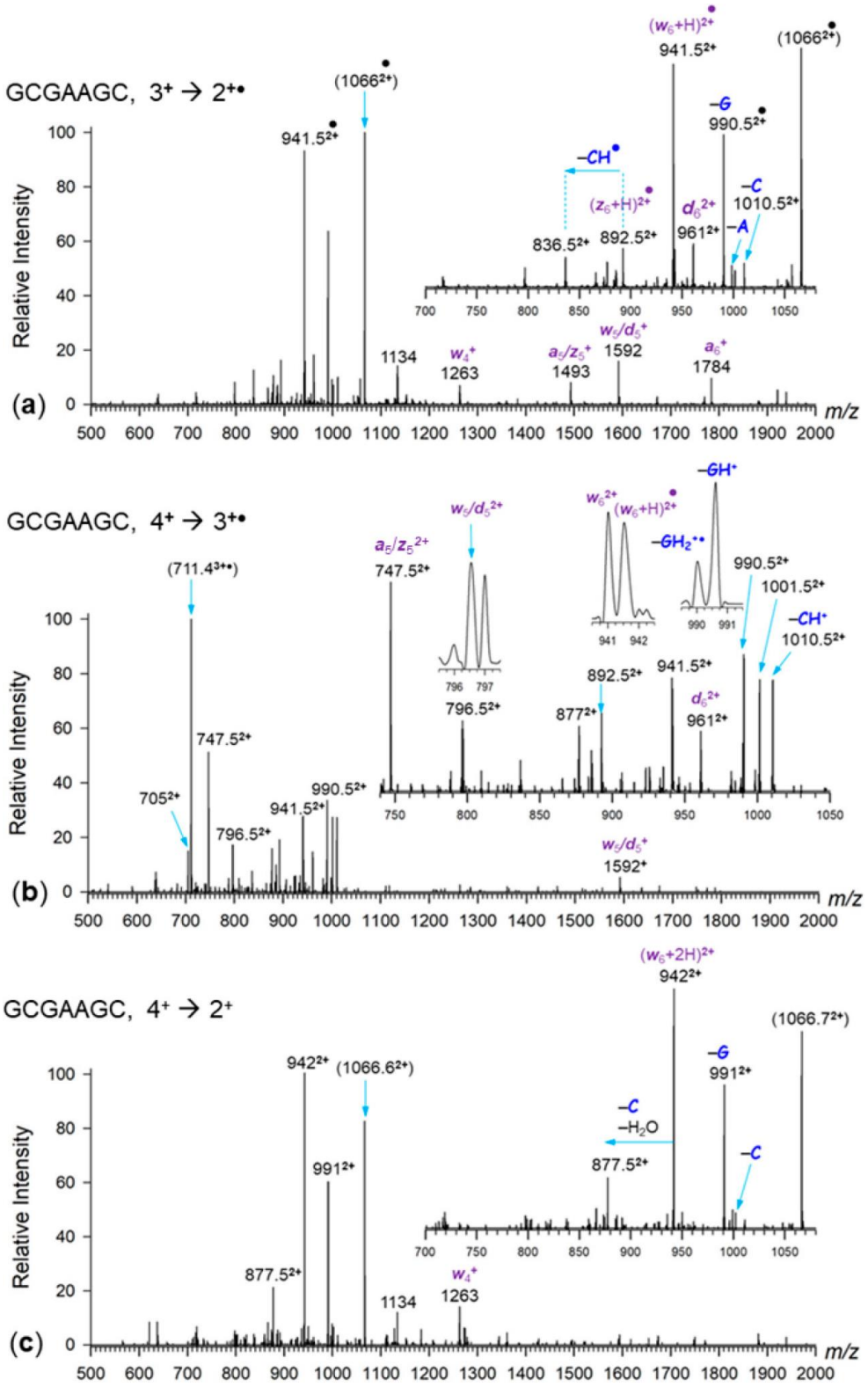


Figure 4.28: ET-CID-MS^3 spectra of d(GCGAAGC) cations: (a) d(GCGAAGC+3H)$^{2+\bullet}$ at m/z 1066; (b) d(GCGAAGC+4H)$^{3+\bullet}$ at m/z 711.4; (c) d(GCGAAGC+4H)$^{2+}$ at m/z 1066.6.

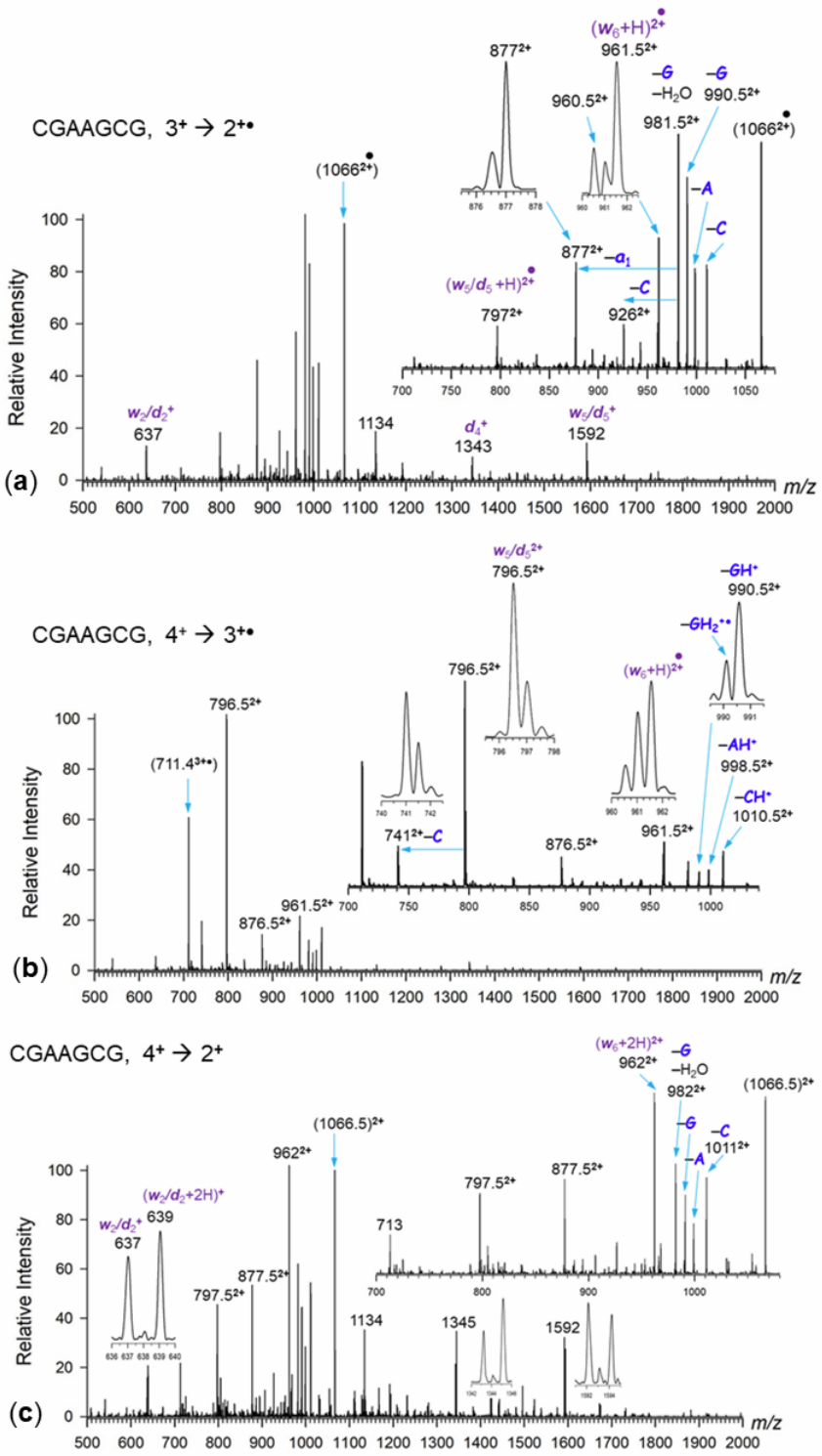


Figure 4.29: ET-CID-MS^3 spectra of (a) d(CGAAGCG+3H)$^{2+\bullet}$ at m/z 1066; (b) d(CGAAGCG+4H)$^{3+\bullet}$ at m/z 711.4; and (c) d(CGAAGCG+4H)$^{2+}$ at m/z 1066.5. Insets show the expanded regions of doubly and triply charged fragment ions.

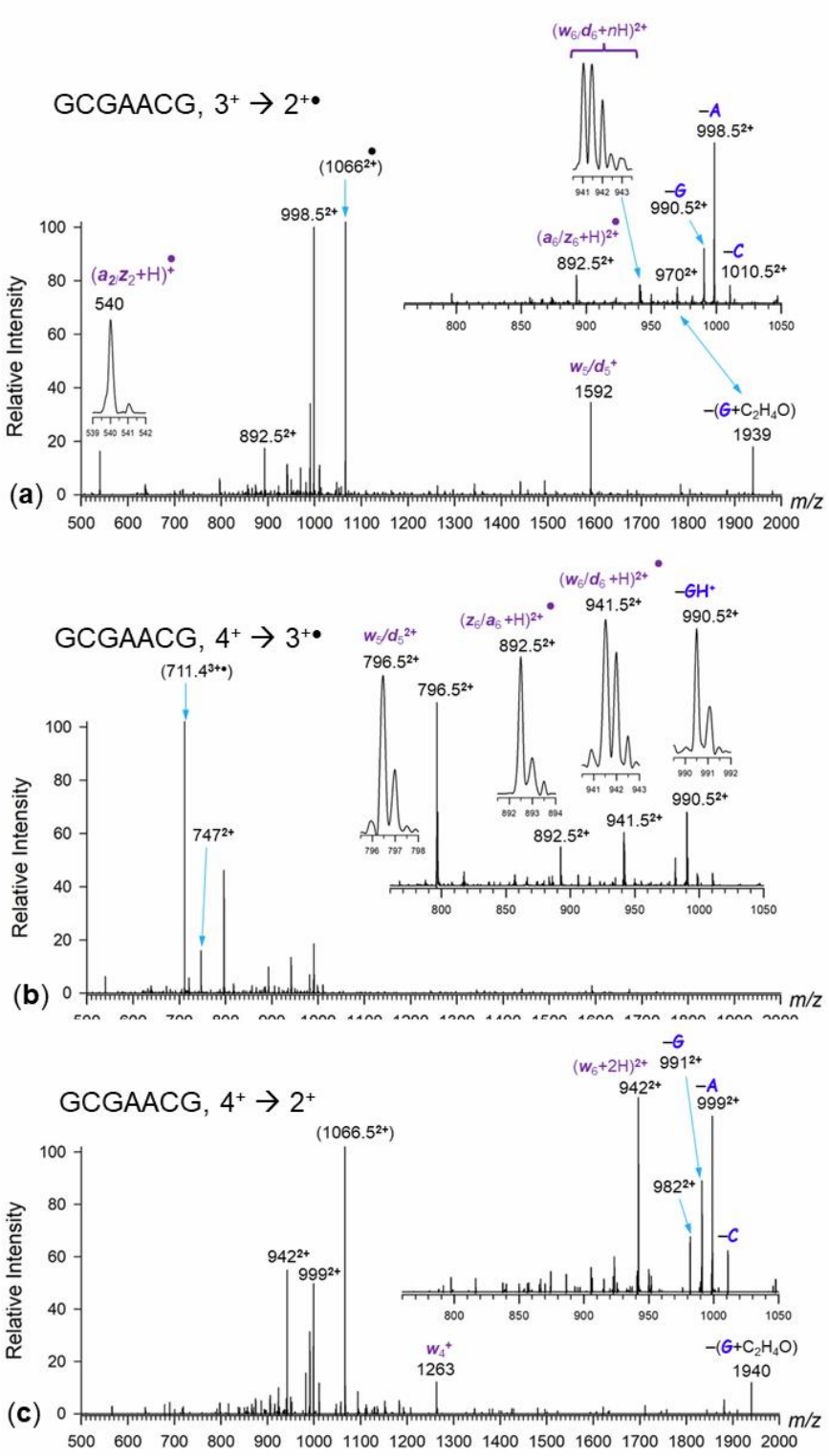


Figure 4.30: ET-CID-MS^3 spectra of (a) d(GCGAACG+3H)$^{2+\bullet}$ at m/z 1066; (b) d(GCGAACG+4H)$^{3+\bullet}$ at m/z 711.4; and (c) d(GCGAACG+4H)$^{2+}$ at m/z 1066.5. Insets show the expanded regions of doubly and triply charged fragment ions.

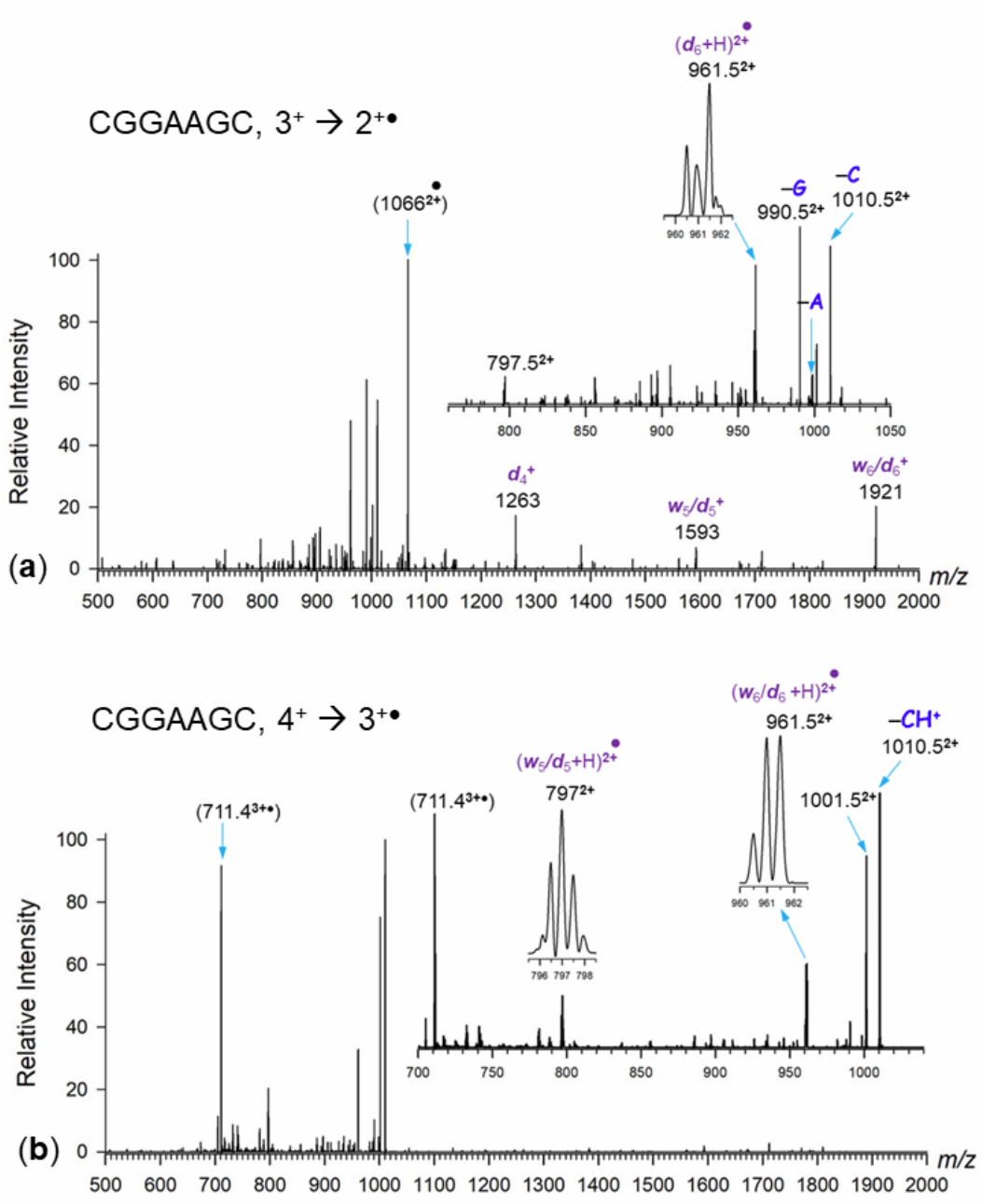


Figure 4.31: ET-CID-MS^3 spectra of (a) d(CGGAAGC+3H)$^{2+\bullet}$ at m/z 1066; (b) d(CGGAAGC+4H)$^{3+\bullet}$ at m/z 711.4. Insets show the expanded regions of doubly and triply charged fragment ions.

4.4 Conclusions

The combined experimental data and calculations allowed us to arrive at the following conclusions. The miniloop structure of GCGAAGC underwent distortions in gas-phase cations generated by electrospray under both native and denaturing conditions. Protonation in di-, tri-, and tetracations resulted in the formation of zwitterionic species in which stabilizing strong hydrogen

bonds in phosphate anion-nucleobase cation motifs disrupted Watson–Crick pairing between C2 and G6. The collapse of the miniloop in $GCGAAGC^{2+}$ manifested itself by compact structures that were indicated by collision cross sections measured by c-IMS. Cations derived from sequence-scrambled heptanucleotides also showed multiple isomers with CCS comparable to those of the GCGAAGC ions.

Table 4.4 c-IMS Parameters for Calibration and Single-Pass Separation

Traveling wave parameters		Sequence			
		Parameter	Inject	Separate	Eject and Acquire
Cyclic TW Velocity (m/s)	375	Time	10	2	automatic
Array TW Velocity (m/s)	375	Time Abs	10	12	automatic
TW static height (V)	15	Pre Array Gradient	85	85	85
TW start height (V)	15	Pre Array Bias	70	70	70
TW limit height (V)	35	Array Entrance	10	30	50
TW ramping rate (V/ms)	2.5	Wave Height	2	0	15
		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

CCS calibration was based on https://doi.org/10.1039/c8sc04396e. Table 4.5 shows the ions that were used for single pass CCS calibration. Theoretical values for the CCS were taken from the Unified CCS Compendium.[61] DriftScope 2.9 software was used for the peak detection. For creating the CCS calibration, Excel file provided by Waters Corp was used. A calibration curve was constructed by plotting ln(CCS' ($Å^2$) vs. ln(d*t*') (**Figure 4.32**).

Table 4.5 Ions Used for CCS Calibration

m/z	z	Mol. mass (Da)	CCS (Å²)
303.17	1	302.20	166.00
374.20	1	373.20	180.30
445.24	1	444.20	194.00
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
800.43	1	799.40	265.70
871.46	1	870.50	278.50
942.50	1	941.50	290.80
1013.54	1	1012.50	302.00
1084.58	1	1083.60	313.70
1155.61	1	1154.60	324.30

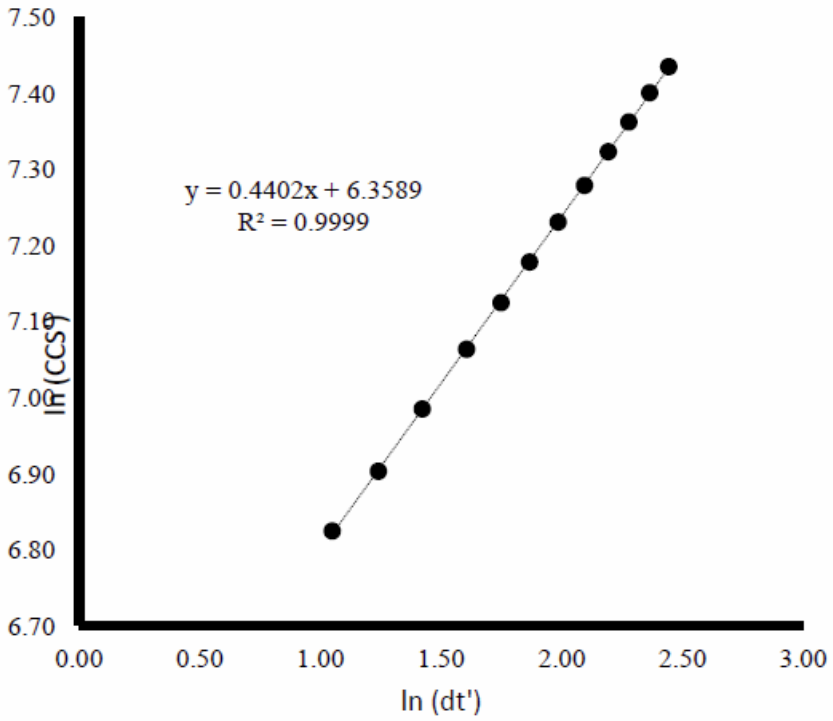


Figure 4.32: Calibration plot of ln(CCS') versus ln(dt'). A logarithmic fit function obtained from the calibration data was used with the corrected drift time to estimate the collision cross section of the studied ions (CCS*=A·dt'').

Table 4.6 c-IMS Parameters for 2 Pass (~2 m) Separation of Dications

Traveling wave parameters		Sequence						
		Parameter	Inject	Separate				Eject and Acquire
Cyclic TW Velocity (m/s)	375	Time	10	GCGAAGC	CGAAGCG	GCGAACG	CGGAAGC	automatic
Array TW Velocity (m/s)	375	Time Abs	10	30	30	30	31	automatic
TW static height (V)	15	Pre Array Gradient	85	85				85
TW start height (V)	15	Pre Array Bias	70	70				70
TW limit height (V)	35	Array Entrance	10	30				50
TW ramping rate (V/ms)	2.5	Wave Height	2	0				15
		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

Table 4.7 c-IMS parameters for $GCGAAGC^{2+}$ after 2 pass separation → slicing → 5 pass separation

Traveling wave parameters		Sequence								
		Parameter	Inject	Separate	Eject	Eject to Pre Store	Hold and Eject	Reinject from Pre Store	Separate	Eject and Acquire
Cyclic TW Velocity (m/s)	375	Time	10	30	17.09	13.09	20.68	13.1	119	automatic
Array TW Velocity (m/s)	375	Pre Array Gradient	85	85	85	45	45	85	85	85
TW static height (V)	15	Pre Array Bias	70	70	70	40	40	70	70	70
TW start height (V)	15	Array Entrance	10	30	50	0	50	10	30	50
TW limit height (V)	35	Wave Height	2	0	25	25	25	2	0	25
TW ramping rate (V/ms)	2.5	Array Offset	45	70	45	55	45	45	70	45
		Array Mode	Forward	Sideways	Forward	Reverse	Forward	Forward	Sideways	Forward Eject
		Array Exit	50	30	2	50	2	50	30	2
		Post Array Gradient	35	35	35	35	35	35	35	35
		Post Array Bias	10	10	10	10	10	10	10	10

Table 4.8 c-IMS parameters for $CGAAGCG^{2+}$ after 2 pass separation → slicing → 5 pass separation

Traveling wave parameters		Sequence								
		Parameter	Inject	Separate	Eject	Eject to Pre Store	Hold and Eject	Reinject from Pre Store	Separate	Enject and Acquire
Cyclic TW Velocity (m/s)	375	Time	10	30	16.60	13.83	17	13.83	113	automatic
Array TW Velocity (m/s)	375	Pre Array Gradient	85	85	85	45	45	85	85	85
TW static height (V)	15	Pre Array Bias	70	70	70	40	40	70	70	70
TW start height (V)	15	Array Entrance	10	30	50	0	50	10	30	50
TW limit height (V)	35	Wave Height	2	0	25	25	25	2	0	15
TW ramping rate (V/ms)	2.5	Array Offset	45	70	45	55	45	45	70	45
		Array Mode	Forward	Sideways	Forward	Reverse	Forward	Forward	Sideways	Forward Eject
		Array Exit	50	30	2	50	2	50	30	2
		Post Array Gradient	35	35	35	35	35	35	35	35
		Post Array Bias	10	10	10	10	10	10	10	10

Table 4.9 c-IMS parameters for $GCGAACG^{2+}$ after 2 pass separation → slicing → 5 pass separation

Traveling wave parameters		Sequence								
		Parameter	Inject	Separate	Eject	Eject to Pre Store	Hold and Eject	Reinject from Pre Store	Separate	Enject and Acquire
Cyclic TW Velocity (m/s)	375	Time	10	2	21.74	8.63	20	8.63	119	automatic
Array TW Velocity (m/s)	375	Pre Array Gradient	85	85	85	45	45	85	85	85
TW static height (V)	15	Pre Array Bias	70	70	70	40	40	70	70	70
TW start height (V)	15	Array Entrance	10	30	50	0	50	10	30	50
TW limit height (V)	35	Wave Height	2	0	25	25	25	2	0	15
TW ramping rate (V/ms)	2.5	Array Offset	45	70	45	55	45	45	70	45
		Array Mode	Forward	Sideways	Forward	Reverse	Forward	Forward	Sideways	Forward Eject
		Array Exit	50	30	2	50	2	50	30	2
		Post Array Gradient	35	35	35	35	35	35	35	35
		Post Array Bias	10	10	10	10	10	10	10	10

Table 4.10 c-IMS parameters for $CGGAAGC^{2+}$ after 2 pass separation → slicing → 5 pass separation

Traveling wave parameters		Sequence								
		Parameter	Inject	Separate	Eject	Eject to Pre Store	Hold and Eject	Reinject from Pre Store	Separate	Enject and Acquire
Cyclic TW Velocity (m/s)	375	Time	10	31	16.85	13.83	20	13.83	120	automatic
Array TW Velocity (m/s)	375	Pre Array Gradient	85	85	85	45	45	85	85	85
TW static height (V)	15	Pre Array Bias	70	70	70	40	40	70	70	70
TW start height (V)	15	Array Entrance	10	30	50	0	50	10	30	50
TW limit height (V)	35	Wave Height	2	0	25	25	25	2	0	15
TW ramping rate (V/ms)	2.5	Array Offset	45	70	45	55	45	45	70	45
		Array Mode	Forward	Sideways	Forward	Reverse	Forward	Forward	Sideways	Forward Eject
		Array Exit	50	30	2	50	2	50	30	2
		Post Array Gradient	35	35	35	35	35	35	35	35
		Post Array Bias	10	10	10	10	10	10	10	10

Table 4.11 c-IMS Parameters for Trications in 5 Pass Separation for $CGAAGCG^{3+}$, 3 Passes for $GCGAACG^{3+}$, and 4 Passes for $CGGAAGC^{3+}$

Traveling wave parameters		Sequence					
		Parameter	Inject	Separate			Eject and Acquire
Cyclic TW Velocity (m/s)	375	Time	10	CGAAGCG	GCGAACG	CGGAAGC	automatic
Array TW Velocity (m/s)	375	Time Abs	10	67	36	51	automatic
TW static height (V)	15	Pre Array Gradient	85	85			85
TW start height (V)	15	Pre Array Bias	70	70			70
TW limit height (V)	35	Array Entrance	10	30			50
TW ramping rate (V/ms)	2.5	Wave Height	2	0			15
		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

4.5 Bibliography

[1] Henry, K. D.; Williams, E. R.; Wang, B. H.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. Fourier-Transform Mass Spectrometry of Large Molecules by Electrospray Ionization. *Proc Natl Acad Sci U. S. A* **1989**, *86*, 9075.

[2] Suckau, D.; Shi, Y.; Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M.; McLafferty, F. W. Coexisting Stable Conformations of Gaseous Protein Ions. *Proc Natl Acad Sci U. S. A.* **1993**, *90*, 790.

[3] Wyttenbach, T.; Bowers, M. T. Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility-Mass Spectrometry Environment. *J Phys Chem B* **2011**, *115*, 12266.

[4] Jurneczko, E.; Barran, P. E. How Useful Is Ion Mobility Mass Spectrometry for Structural Biology? The Relationship between Protein Crystal Structures and Their Collision Cross Sections in the Gas Phase. *Analyst* **2011**, *136*, 20.

[5] Eldrid, C.; Ujma, J.; Kalfas, S.; Tomczyk, N.; Giles, K.; Morris, M.; Thalassinos, K. Gas Phase Stability of Protein Ions in a Cyclic Ion Mobility Spectrometry Traveling Wave Device. *Anal Chem* **2019**, *91*, 7554.

[6] Seo, J.; Hoffmann, W.; Warnke, S.; Bowers, M. T.; Pagel, K.; von Helden, G. Retention of Native Protein Structures in the Absence of Solvent: A Coupled Ion Mobility and Spectroscopic Study. *Angew Chem Int Ed* **2016**, *55*, 14173.

[7] Susa, A. C.; Xia, Z.; Williams, E. R. Native Mass Spectrometry from Common Buffers with Salts That Mimic the Extracellular Environment. *Angew Chem Int Ed* **2017**, *56*, 7912.

[8] Clemmer, D. E.; Russell, D. H.; Williams, E. R. Characterizing the Conformationome: Toward a Structural Understanding of the Proteome. *Acc Chem Res* **2017**, *50*, 556.

[9] Bakhtiari, M.; Konermann, L. Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures. *J Phys Chem B* **2019**, *123*, 1784.

[10] Schnier, P. D.; Klassen, J. S.; Strittmatter, E. F.; Williams, E. R. Activation Energies for Dissociation of Double Strand Oligonucleotide Anions: Evidence for Watson-Crick Base Pairing in Vacuo. *J Am Chem Soc* **1998**, *120*, 9605.

[11] Griffey, R. H.; Greig, M. J.; An, H.; Sasmor, H.; Manalili, S. Targeted Site-Specific Gas-Phase Cleavage of Oligoribonucleotides. Application in Mass Spectrometry-Based Identification of Ligand Binding Sites. *J Am Chem Soc* **1999**, *121*, 474.

[12] Gabelica, V.; Rosu, F.; Houssier, C.; DePauw, E. Gas Phase Thermal Denaturation of an Oligonucleotide Duplex and Its Complexes with Minor Groove Binders. *Rapid Commun Mass Spectrom* **2000**, *14*, 464.

[13] Gabelica, V.; DePauw, E. Collision-Induced Dissociation of 16-Mer DNA Duplexes with Various Sequences: Evidence for Conservation of the Double Helix Conformation in the Gas Phase. *Int J Mass Spectrom* **2002**, *219*, 151.

[14] Gidden, J.; Ferzoco, A.; Baker, E. S.; Bowers, M. T. Duplex Formation and the Onset of Helicity in Poly d(CG)n Oligonucleotides in a Solvent-Free Environment. *J Am Chem Soc* **2004**, *126*, 15132.

[15] Butcher, D.; Chapagain, P.; Leng, F.; Fernandez-Lima, F. Differentiating Parallel and Antiparallel DNA Duplexes in the Gas Phase Using Trapped Ion Mobility Spectrometry. *J Phys Chem B* **2018**, *122*, 6855.

[16] Khristenko, N.; Amato, J.; Livet, S.; Pagano, B.; Randazzo, A.; Gabelica, V. Native Ion Mobility Mass Spectrometry: When Gas-Phase Ion Structures Depend on the Electrospray Charging Process. *J Am Soc Mass Spectrom* **2019**, *30*, 1069.

[17] Saintmont, F.; Hoyas, S.; Rosu, F.; Gabelica, V.; Brocorens, P.; Gerbaux, P. Structural Characterization of Dendriplexes In Vacuo: A Joint Ion Mobility/Molecular Dynamics Investigation. *J Am Soc Mass Spectrom* **2022**, *33*, 1555.

[18] Sipe, S. N.; Fouque, K. J. D.; Garabedian, A.; Leng, F.; Fernandez-Lima, F.; Brodbelt, J. S. Exploring the Conformations and Binding Location of HMGA2·DNA Complexes Using Ion Mobility

Spectrometry and 193 Nm Ultraviolet Photodissociation Mass Spectrometry. *J Am Soc Mass Spectrom* **2022**, *33*, 1092.

[19] Vairamani, M.; Gross, M. L. G-Quadruplex Formation of Thrombin-Binding Aptamer Detected by Electrospray Ionization Mass Spectrometry. *J Am Chem Soc* **2003**, *125*, 42.

[20] Arcella, A.; Portella, G.; Ruiz, M. L.; Eritja, R.; Vilaseca, M.; Gabelica, V.; Orozco, M. Structure of Triplex DNA in the Gas Phase. *J Am Chem Soc* **2012**, *134*, 6596.

[21] Ickert, S.; Hofmann, J.; Riedel, J.; Beck, S.; Pagel, K.; Linscheid, M. W. Charge-Induced Geometrical Reorganization of DNA Oligonucleotides Studied by Tandem Mass Spectrometry and Ion Mobility. *Eur J Mass Spectrom* **2018**, *24*, 225.

[22] Schürch, S. Characterization of Nucleic Acids by Tandem Mass Spectrometry-The Second Decade (2004–2013): From DNA to RNA and Modified Sequences. *Mass Spectrom Rev* **2016**, *35*, 483.

[23] Hari, Y.; Leumann, C. J.; Schürch, S. What Hinders Electron Transfer Dissociation (ETD) of DNA Cations? *J. Am. Soc. Mass Spectrom.* **2017**, *28* (12), 2677–2685. https://doi.org/10.1007/s13361-017-1791-z.

[24] Korn, J. A.; Urban, J.; Dang, A.; Nguyen, H. T. H.; Turecek, F. UV-Vis Action Spectroscopy Reveals a Conformational Collapse in Hydrogen-Rich Dinucleotide Cation Radicals. *J Phys Chem Lett* **2017**, *8*, 4100.

[25] Liu, Y.; Korn, J. A.; Dang, A.; Turecek, F. Hydrogen-Rich Cation Radicals of DNA Dinucleotides. Generation and Structure Elucidation by UV-Vis Action Spectroscopy. *J Phys Chem B* **2018**, *122*, 9665.

[26] Liu, Y.; Huang, S. R.; Tureček, F. Guanine-Adenine Interactions in DNA Tetranucleotide Cation Radicals Revealed by UV/Vis Photodissociation Action Spectroscopy and Theory. *Phys Chem Chem Phys* **2020**, *22*, 16831.

[27] Huang, S. R.; Liu, Y.; Tureček, F. UV-Vis Photodissociation Action Spectroscopy Reveals Cytosine-Guanine Hydrogen Transfer in DNA Tetranucleotide Cation Radicals upon One-Electron Reduction. *J Phys Chem B* **2020**, *124*, 3505.

[28] Li, J.; Begbie, A.; Boehm, B. J.; Button, A.; Whidborne, C.; Pouferis, Y.; Huang, D. M.; Pukala, T. L. Ion Mobility-Mass Spectrometry Reveals Details of Formation and Structure for GAA·TCC DNA and RNA Triplexes. *J Am Soc Mass Spectrom* **2019**, *30*, 103.

[29] Wan, J.; Brož, B.; Liu, Y.; Huang, S. R.; Marek, A.; Tureček, F. Resolution of Identity in Gas-Phase Dissociations of Mono and Diprotonated DNA Codons by 15N-Labeling and Computational Structure Analysis. *J Am Soc Mass Spectrom* **2022**, *33*, 1936.

[30] Sunami, T.; Kondo, J.; Tsunoda, M.; Sekiguchi, T.; Hirao, I.; Watanabe, K.; Miura, K.; Takenaka, A. X-Ray Structure of d(GCGAAGC); Switching of Partner for G:A Pair in Duplex Form. *Nucleic Acids Res* **2002**, *2*, 181.

[31] Hirao, I.; Kawai, G.; Yoshizawa, S.; Nishimura, Y.; Ishido, Y.; Watanabe, K.; Miura, K. Most Compact Hairpin-Turn Structure Exerted by a Short DNA Fragment, d(GCGAAGC) in Solution: An Extraordinarily Stable Structure Resistant to Nucleases and Heat. *Nucleic Acids Res* **1994**, *22*, 576.

[32] Padrta, P.; Štefl, R.; Králík, L.; Žídek, L.; Sklenář, V. Refinement of d(GCGAAGC) Hairpin Structure Using One- and Two-Bond Residual Dipolar Couplings. *J Biomol NMR* **2002**, *24*, 1.

[33] Jolles, B.; Refregiers, M.; Laigle, A. Opening of the Extraordinarily Stable Mini-Hairpin d(GCGAAGC). *Nucleic Acids Res* **1997**, *25*, 4608.

[34] Hartmer, R.; Kaplan, D. A.; Gebhardt, C. R.; Ledertheil, T.; Brekenfeld, A. Multiple Ion/Ion Reactions in the 3D Ion Trap: Selective Reagent Ion Production for ETD and PTR from a Single Compound. *Int J Mass Spectrom* **2008**, *276*, 82.

[35] 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.

[36] Ujma, J.; Ropartz, D.; Giles, K.; Richardson, K.; Langridge, D.; Wildgoose, J.; Green, M.; Pringle, S. Cyclic Ion Mobility Mass Spectrometry Distinguishes Anomers and Open-Ring Forms of Pentasaccharides. *J Am Soc Mass Spectrom* **2019**, *30*, 1028.

[37] Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.; Robinson, C. V. Ion Mobility Mass Spectrometry Analysis of Large Protein Complexes. *Nat Protoc* **2008**, *3*, 1139.

[38] Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal Chem* **2010**, *82*, 9557.

[39] Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A. Collision Cross Section Compendium to Annotate and Predict Multi-Omic Compound Identities. *Chem Sci* **2019**, *10*, 983.

[40] Stewart, J. J. P. Optimization of Parameters for Semi-Empirical Methods. V. Modification of NDDO Approximations and Application to 70 Elements. *J Mol Model* **2007**, *13*, 1173.

[41] 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.

[42] Berendsen, H. J.; Postma, J. V.; van Gunsteren, W. F.; DiNola, A. R. H. J.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, *81*, 3684.

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

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

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

[46] 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.

[47] Chai, J. D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J Chem Phys* **2008**, *128*, 084106.

[48] Chai, J. D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion. *Phys Chem Chem Phys* **2008**, *10*, 6615.

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

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

[51] Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long Range Potential. *J Phys Chem* **1996**, *100*, 16082.

[52] Campuzano, I. D. G.; Bush, M. F.; Robinson, C. V.; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I. Structural Characterization of Drug-like Compounds by Ion Mobility Mass Spectrometry: Comparison of Theoretical and Experimentally Derived Nitrogen Collision Cross-Sections. *Anal Chem* **2012**, *84*, 1026.

[53] Kim, H.; Kim, H. I.; Johnson, P. V.; Beegle, L. W.; Beauchamp, J. L.; Goddard, W. A.; Kanik, I. Experimental and Theoretical Investigation into the Correlation between Mass and Ion Mobility for Choline and Other Ammonium Cations in N2. *Anal Chem* **2008**, *80*, 1928.

[54] 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.

[55] Ieritano, C.; Hopkins, W. S. Assessing Collision Cross Section Calculations Using MobCal-MPI with a Variety of Commonly Used Computational Methods. *Mater Today Commun* **2021**, *27*, 102226.

[56] Gatlin, C. L.; Tureček, F. Acidity Determination in Droplets Formed by Electrospraying Methanol-Water Solutions. *Anal Chem* **1994**, *66*, 712.

[57] Borovcová, L.; Hermannová, M.; Pauk, V.; Šimek, M.; Havlíček, V.; Lemr, K. Simple Area Determination of Strongly Overlapping Ion Mobility Peaks. *Anal Chim Acta* **2017**, *981*, 71.

[58] Zündel, G.; Metzger, H. Energiebänder Der Tunnelnden Uberschuß-Protonen in Flüssigen Säuren. Eine IR-Spektroskopische Untersuchung Der Natur Der Gruppierungen H5O2+. *Z Phys Chem* **1968**, *58*, 225.

[59] 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* (2), 304–319. https://doi.org/10.1021/jasms.2c00322.

[60] Baba, T.; Campbell, J. L. Capturing Polyradical Protein Cations after an Electron Capture Event: Evidence for Their Stable Distonic Structures in the Gas Phase. *J Am Soc Mass Spectrom* **2015**, *26*, 1695.

[61] Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A. Collision Cross Section Compendium to Annotate and Predict Multi-Omic Compound Identities. *Chem. Sci.* **2019**, *10* (4), 983–993.

Chapter 5

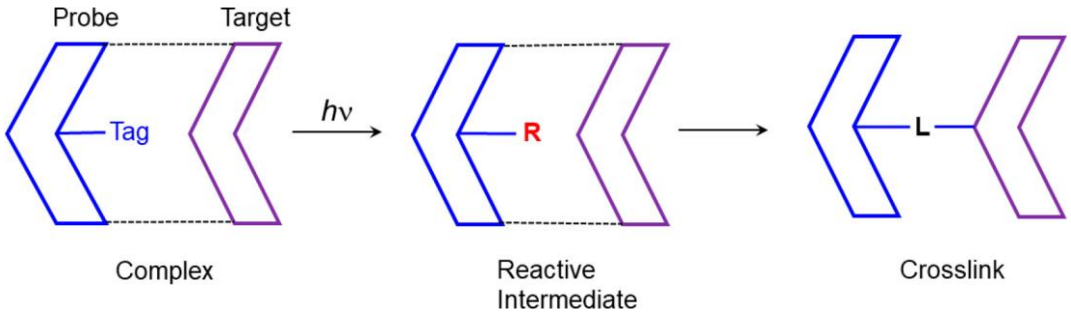
Nitrile Imines as Peptide and Oligonucleotide Photo-Cross-Linkers in Gas-Phase Ions

Reproduced in part with permission from Jiahao Wan, Marianna Nytka, Haocheng Qian, Kim Vu, Karel Lemr and František Tureček. Nitrile Imines as Peptide and Oligonucleotide Photo-Cross-Linkers in Gas-Phase Ions. *J. Am. Soc. Mass Spectrom.* **2024**, *35*, 2, 344–356.

Abstract *Nitrile imines produced by photodissociation of 2,5-diaryltetrazoles undergo cross-linking reactions with amide groups in peptide-tetrazole (tet-peptide) conjugates and a tet-peptide-dinucleotide complex. Tetrazole photodissociation in gas-phase ions is efficient, achieving ca. 50% conversion with 2 laser pulses at 250 nm. The formation of cross-links was detected by CID-MS^3 that showed structure-significant dissociations by loss of side-chain groups and internal peptide segments. The structure and composition of cross-linking products were established by a combination of UV–vis action spectroscopy and cyclic ion mobility mass spectrometry (c-IMS). The experimental absorption bands were found to match the bands calculated for vibronic absorption spectra of nitrile imines and cross-linked hydrazone isomers. The calculated collision cross sections (CCS_{th}) for these ions were related to the matching experimental CCS_{exp} from multipass c-IMS measurements. Loss of N_2 from tet-peptide conjugates was calculated to be a mildly endothermic reaction with ΔH_0 = 80 kJ mol^{-1} in the gas phase. The excess energy in the photolytically formed nitrile imine is thought to drive endothermic proton transfer, followed by exothermic cyclization to a sterically accessible peptide amide group. The exothermic nitrile imine reaction with peptide amides is promoted by proton transfer and may involve an initial [3 + 2] cycloaddition followed by cleavage of the oxadiazole intermediate. Nucleophilic groups, such as cysteine thiol, did not compete with the amide cyclization. Nitrile imine cross-linking to 2′-deoxycytidylguanosine was found to be >80% efficient and highly specific in targeting guanine. The further potential for exploring nitrile-imine cross-linking for biomolecular structure analysis is discussed.*

5.1 *Introduction*

Photochemical cross-linking is an established method for structure elucidation of protein–protein[1] and protein–nucleic acid complexes.[2,3] Cross-linking relies on highly reactive intermediates that are transiently produced by photolysis of a stable precursor group (Tag), which is introduced into one component (Probe) of the complex (**Scheme 5.1**). Upon formation, the reactive group (R) undergoes rapid insertion into the sterically accessible X–H bonds in the target component (X = C, N, O), thus forming a covalent bond via a bridging group (L) and converting the complex into a molecule.



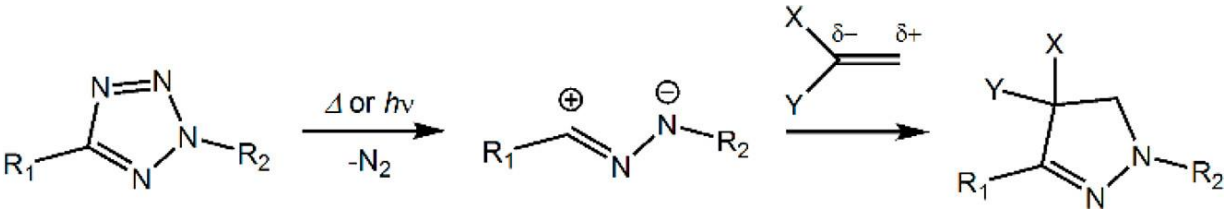
Scheme 5.1: Photodissociative Cross-Linking

The choice of reactive groups has been limited by the requirement that the precursor is stable and can be incorporated into the complex in a site-specific manner. Carbenes[4,5] and benzophenone excited states[6] have been the most frequently used cross-linkers in condensed-phase applications. Cross-links can be detected by mass analysis, such as gel electrophoresis or mass spectrometry. For protein cross-linking, enzymatic digestion followed by peptide analysis by tandem mass spectrometry can be used to identify the cross-links and thus assign the contact sites in the complex.[1]

We have introduced photochemical cross-linking in gas-phase peptide–peptide,[7–12] peptide–nucleotide,[13] and peptide–ligand complexes,[14] as recently reviewed.[15] This has relied on carbene intermediates that were generated by UV–vis photodissociation (UVPD) of diazirine-tagged peptides or ligands.[9] Diazirine photodissociation is readily monitored by mass change in the gas-phase complex ion by N_2 elimination. Cross-links are analyzed by CID-MS^3 sequencing that allows one to identify the amino acid residues that were targeted by the carbene. [7] This scheme is fundamentally different from other methods of gas-phase covalent bond formation, such

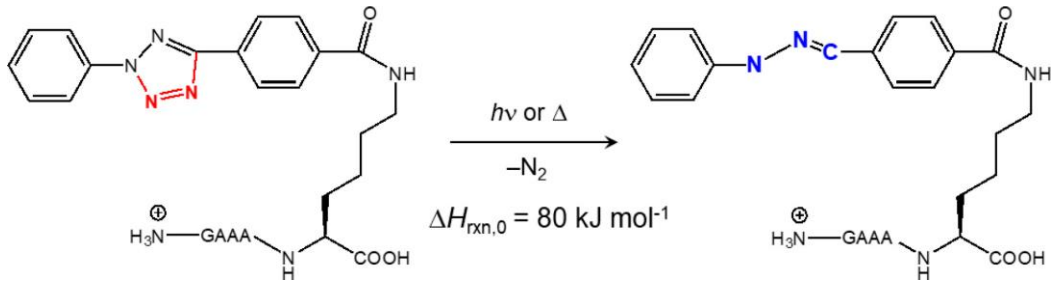
as nonselective photodissociation of small peptide complexes at 157 nm[16] or gas-phase anion–cation reactions of active esters with peptide N-terminus.[17,18] Photoexcitation at 157 nm is inherently nonselective because it does not target a particular chromophore, and thus cannot be used for structure analysis of the complex. Ion–ion reactions do not address noncovalent complexes, whereas the interaction between the gas-phase reagents is dominated by strong ion–ion attractive interactions. While carbene-based cross-linking can be considered a general method of wide applicability, it also has inherent limitations.[15] One of those is the very low extinction coefficient for the electronic excitation in the diazirine ring. This involves the dipole forbidden $\pi_{xy} \rightarrow \pi_z$ transition[19,20] that has a very low oscillator strength, which adversely affects the photodissociation conversion. Another drawback is the overall indiscriminate carbene reactivity that can attack any proximate X–H bond in the target moiety, which complicates product analysis. An alternative method of producing carbenes by CID of diazomalonate complexes has also been reported.[21]

We now report another, rather unexpected, photochemical reaction that we apply to achieve efficient intramolecular cross-linking in peptides and intermolecular cross-linking in peptide–nucleotide complexes. The new reaction is based on nitrile imines. Nitrile imines are reactive intermediates that are known to undergo [3 + 2] cycloadditions[22] to enones, alkylidene malonates, and a broad range of other dipolarophiles, that have been used in the synthesis of various five-membered heterocyclic systems (**Scheme 5.2**).[23–28] As reactive intermediates, nitrile imines are readily accessible by photolysis or thermolysis of tetrazoles, by base-induced elimination from hydrazonyl halides, and by other methods.[29] 2,5-Diaryltetrazole carboxylates have been used by Song et al. to conjugate the tetrazole group to the lysine ε-amino group in peptides for further modifications by [3 + 2] cycloaddition with dipolarophiles.[30] Despite the broad range of dipolarophiles used for reactions with nitrile imines, there has been a paucity of reports of reactions with functional groups pertinent to biomolecules, such as amide groups in proteins and purine and pyrimidine heterocycles in nucleic acids. Chlorinated nitrile imines have been reported to undergo [3 + 2] cycloaddition to enamides, but addition to an inactivated amide group has not been reported. The electronic properties of nitrile imines have been studied by Wentrup and co-workers.[31,32]



Scheme 5.2: Formation of Nitrile Imines from Tetrazoles and [3 + 2] Addition to Dipolarophiles

The basic thermochemistry of nitrile imine generation is shown in **Scheme 5.3** for a 2,5-diaryltetrazolecarboxamide (tet) conjugate with the pentapeptide Gly-Ala-Ala-Ala-Lys (GAAA(tet)K + H)$^+$, based on density functional theory (DFT) calculations. The calculated reaction enthalpy indicates that photodissociation at 250 nm, providing photons with 4.96 eV (479 kJ mol^{-1}) energy, is more than sufficient to break the tetrazole ring. As opposed to reactions in the condensed phase, the energy excess in the nitrile imine formation, $\Delta H_{exc} \leq 479–80 = 399$ kJ mol^{-1}, is temporarily stored in the gas-phase ion because of slow collisional cooling[33,34] and can be used to drive further reactions with proximate functional groups. Here, we report nitrile-imine cross-linking reactions with several peptides tagged with 2,5-diaryltetrazole. We wish to show that nitrile-imine-based reactions have the potential to be developed into a universal method for cross-linking in biomolecular ions.



Scheme 5.3: Nitrile Imine Thermochemistry

5.2 Experimental Section

5.2.1 Materials and Methods

The tet-peptide conjugates, in which the 2,5-diaryltetrazole group was attached as a carboxamide to the lysine ε-amine, were prepared by standard methods of peptide synthesis, and

their sequences were confirmed by tandem mass spectrometry with accurate mass measurements (**Tables 5.1–5.4**).

Typically, synthesis of the photo-tagged peptides included the following steps: (1) deprotection of the lysine side chain. (2) amide coupling of the tetrazole photo-tag with the lysine side chain. (3) standard solid phase synthesis of the rest of the peptide sequence. (4) cleavage of the synthesized peptide from the resin.

Fmoc-lysine(Mtt)-Wang resin (80 mg) was weighed into a 12 mL syringe cylinder. For deprotection of the lysine methyltrityl (Mtt) group, inject 6 mL of a mixture of methylene chloride (DCM)/methanol (MeOH)/trifluoroacetic acid (TFA) (98:1:1) and immerse the beads for 5 minutes. Remove the reaction solution and inject another 6 mL of the deprotection solution to immerse the beads overnight to fully remove the protecting group. After deprotection, wash the beads in turn with the following solvents: 5 mL DCM twice; 5 mL MeOH twice; 5 mL DCM twice again; 5 mL 1% N,N-diisopropylethylamine (DIEA) in DMF twice; then 5 mL DMF 4 times to fully remove any traces of acid. Prepare the coupling cocktail solution by dissolving 4-(2-phenyl-2H-tetrazol-5-yl)benzoic acid (80 mg, 6 equiv.), benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate (PyBOP)(149 mg, 6 equiv.), hydroxybenzotriazole (HOBt) (55 mg, 8.4 equiv.), N-methylmorpholine (NMM)(80 μL) in 4 mL DMF. Inject the cocktail into the syringe and allow it to react overnight at room temperature. Then wash the beads with 5 mL DMF 6 times. For Fmoc deprotection, inject 5 mL 20% piperidine in DMF and shake for 10 minutes. Wash the beads with DMF 6 times. Prepare the cocktail solution for peptide chain elongation by dissolving Fmoc protected amino acid (3 equiv.), PyBOP (225 mg, 3 equiv.), HOBt (82 mg, 4.2 equiv.) and 120 μL NMM in 5 mL DMF. Inject the solution and shake for 20 minutes for the reaction to complete, then eject the liquid phase, and wash the beads with 5 mL DMF 6 times. Repeat the Fmoc deprotection, washing, coupling, and washing again until the desired peptide sequence is obtained. After the last Fmoc deprotection and washing step, wash the beads with DCM 5 times to remove DMF. Strip the peptide from resin by injecting a solution mixture of TFA/water/triethylsilane (95:2.5:2.5) twice. Collect the TFA solution containing the peptide product in a small glass vial and evaporate the solvent with air flow to yield 12 – 20 mg peptide product. For protected lysine, cysteine, asparagine and glutamine introduced into peptide sequences during the synthesis, the methyltrityl and trityl protecting groups on the side chains are

orthogonal with Fmoc protecting group and are stable under basic conditions. They are easily removed during the peptide-resin cleavage process.

UV–vis photodissociation tandem mass spectrometry (UVPD-MS^2) was performed on a Bruker amaZon ion trap mass spectrometer that was furnished with windows and focusing optics allowing the beam from an NL301G (Altos Photonics, Bozeman, MT) Nd:YAG laser to be focused into the ion trap.[35] The laser-ion trap mass spectrometer system was described in detail previously.[36] Following the previous report by Blanksby, Trevitt, and coauthors on the photodissociation of 2,5-diaryltetrazoles,[37] the excitation wavelength was tuned to 250 nm. Accurate mass measurements were performed on an Orbitrap Ascend Tribrid instrument (ThermoFisher, San Jose, CA) at a 100 000 resolving power. Ion mobility measurements were carried out on a SELECT SERIES cyclic ion mobility spectrometer (c-IMS) (Waters Corp., Wilmslow, UK)[38] with direct infusion into a normal flow electrospray ion source at a flow rate of 5 μL/min. Each sample and calibrant were measured six times in a positive mode. Typically, ion mobility separation over five cycles was used at the total ion path of 490 cm. Details of all measurement parameters and calibration for collision cross section (CCS) calculations are given in **Table 5.5** and **5.6** and in **Figure 5.1**.

Table 5.1 Accurate Mass Measurements of (AAFA(tet)K + H)$^+$ ions.

m/z	Ion formula	error (mu)	neutral loss	ion assignment
755.3639	$C_{38}H_{47}N_{10}O_7$	1.5		
727.3570	$C_{38}H_{47}N_8O_7$	0.8	N_2	
709.3463	$C_{38}H_{45}N_8O_6$	-0.07	H_2O	
699.3628	$C_{37}H_{47}N_8O_6$	-0.7	CO	
684.3149	$C_{36}H_{42}N_7O_7$	-0.1	C_2H_5N	
656.3197	$C_{35}H_{42}N_7O_6$	0.2	C_3H_5NO	-Ala
636.3154	$C_{32}H_{42}N_7O_7$	-0.6	C_6H_5N	
635.3068	$C_{32}H_{41}N_7O_7$	0.2	C_6H_6N	
619.2879	$C_{32}H_{39}N_6O_7$	0.4	$C_6H_8N_2$	-phenylhydrazine
591.2935	$C_{31}H_{39}N_6O_6$	-0.2	$C_7H_8N_2O$	

585.2837	$C_{32}H_{37}N_6O_5$	1.2	$C_6H_{10}N_2O_2$	656 - Ala
548.2522	$C_{29}H_{34}N_5O_6$	-1.1	$C_9H_{13}N_3O$	
509.2512	$C_{26}H_{33}N_6O_5$	0.3	$C_{12}H_{14}N_2O_2$	-Ala-Phe
438.2144	$C_{23}H_{28}N_5O_4$	-0.04	$C_{15}H_{19}N_3O_3$	***y₂***
416.1937	$C_{20}H_{26}N_5O_5$	0.3	$C_{18}H_{21}N_3O_2$	509 – C_6H_7N
375.1836	$C_{24}H_{25}NO_3$	0.2	$C_{14}H_{22}N_5O_4$	
367.1773	$C_{20}H_{23}N_4O_3$	0.3	$C_{18}H_{24}N_4O_4$	***y₁***
345.1564	$C_{17}H_{21}N_4O_4$	0.1	$C_{21}H_{26}N_4O_3$	416 - Ala
292.1098	$C_{17}H_{14}N_3O_2$	-0.9	$C_{21}H_{33}N_5O_5$	438 – $C_6H_{14}N_2O_2$
282.1243	$C_{16}H_{16}N_3O_2$	0	$C_{22}H_{31}N_5O_5$	416 – $C_4H_{10}N_2O_3$

Table 5.2 Accurate Mass Measurements of (GAAA(tet)K + H)$^+$ ions.

m/z	Ion formula	error (mu)	neutral loss	ion assignment
665.3155	$C_{31}H_{41}N_{10}O_7$	–0.5		
637.3096	$C_{31}H_{41}N_8O_7$	–0.2	N_2	
619.3001	$C_{31}H_{39}N_8O_6$	0.8	H_2O	
608.2833	$C_{30}H_{38}N_7O_7$	0	CH_3N	
602.2736	$C_{31}H_{36}N_7O_6$	0.9	NH_5O	619 – NH_3
566.2736	$C_{28}H_{36}N_7O_6$	0.9	C_3H_5NO	-Ala
529.2413	$C_{25}H_{33}N_6O_7$	0.2	$C_6H_8N_2$	-phenylhydrazine
495.2354	$C_{25}H_{31}N_6O_5$	–0.2	$C_6H_{10}N_2O_2$	566-Ala
438.2136	$C_{23}H_{28}N_5O_4$	–0.5	$C_8H_{13}N_3O_3$	***y₂***
424.1982	$C_{22}H_{26}N_5O_4$	–0.3	$C_9H_{15}N_3O_3$	495-Ala
367.1777	$C_{20}H_{23}N_4O_3$	0.7	$C_{11}H_{18}N_4O_4$	***y₁***

Table 5.3 Accurate Mass Measurements of (AAFA(tet)K-COONa+Na)$^+$ Ions.

m/z	Ion formula	error (mu)	neutral loss	ion assignment
799. 3265	$C_{38}H_{45}N^{10}O^7Na^2$	0.2		
771.3175	$C_{38}H_{45}N^8O^7Na^2$	–2.6	N_2	
727.3284	$C_{37}H_{45}N^8O^5Na^2$	–1.8	CO_2	
679.2686	$C_{32}H_{39}N_7O_7Na_2$	–1.5	C_6H_6N	
629.2444	$C_{32}H_{35}N_6O_5Na_2$	–1.5	$C_6H_{10}N_2O_2$	y_3
482.1763	$C_{23}H_{25}N_5O_4Na_2$	–1.2	$C_{15}H_{19}N_3O_3$	y_2
421.1239	$C_{21}H_{19}N_4O_3Na_2$	–0.8	$C_{17}H_{26}N_4O_4$	
411.1394	$C_{20}H_{21}N_4O_3Na_2$	–1.0	$C_{18}H_{24}N_4O_4$	y_1
377.1343	$C_{20}H_{19}N_4ONa_2$	–0.6	$C_{18}H_{26}N_4O_6$	421 – CO_2
351.1397	$C_{15}H_{21}N_4O_3Na_2$	–0.7	$C_{23}H_{24}N_4O_4$	

Table 5.4 Accurate Mass Measurements (GAAA(tet)K-COONa+Na)$^+$ Ions.

m/z	Ion formula	error (mu)	neutral loss	ion assignment
709.2793	$C_{31}H_{39}N_{10}O_7Na_2$	–0.6		
681.2716	$C_{31}H_{39}N_8O_7Na_2$	–2.1	N_2	
663.2626	$C_{31}H_{37}N_8O_6Na_2$	–1.5	H_2O	
637.2824	$C_{30}H_{39}N_8O_5Na_2$	–1.5	CO_2	
553.2137	$C_{26}H_{31}N_6O_5Na_2$	–1.4	$C_5H_8N_2O_2$	y_3
535.2031	$C_{26}H_{29}N_6O_4Na_2$	–1.5	$C_5H_{10}N_2O_3$	$y_3 – H_2O$
509.2240	$C_{25}H_{31}N_6O_3Na_2$	–1.3	$C_6H_8N_2O_4$	$y_3 – CO_2$
482.1768	$C_{23}H_{26}N_5O_4Na_2$	–1.2	$C_8H_{13}N_3O_3$	y_2
464.1664	$C_{23}H_{24}N_5O_3Na_2$	–1.2	$C_8H_{15}N_3O_4$	$y_2 – H_2O$
421.1243	$C_{21}H_{19}N_4O_3Na_2$	–1.0	$C_{10}H_{20}N_4O_4$	
411.1399	$C_{20}H_{21}N_4O_3Na_2$	–1.0	$C_{11}H_{18}N_4O_4$	y_1

377.1346	$C_{20}H_{19}N_4ONa_2$	–0.8	$C_9H_{20}N_4O_6$	421 – CO_2
261.0931	$C_8H_{15}N_4O_3Na_2$	–0.9	$C_{23}H_{24}N_4O_4$	

Table 5.5 c-IMS Parameters for Calibration in one (two) pass separation.

Traveling wave parameters		Sequence			
		Parameter	Inject	Separate	Eject and Acquire
Cyclic TW Velocity (m/s)	375	Time	10	2	automatic
Array TW Velocity (m/s)	375	Time Abs	10	12	automatic
TW static height (V)	15	Pre Array Gradient	85	85	85
TW start height (V)	15	Pre Array Bias	70	70	70
TW limit height (V)	35	Array Entrance	10	30	50
TW ramping rate (V/ms)	2.5	Wave Height	2	0	15
		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

Calibrant arrival times (t_a) corresponding to peak maxima were obtained from one and two pass experiments (98 and 196 cm path length). The calibration drift times were calculated as t_d = t_a(two passes) – t_a(one pass). The dead time was t_0 = t_a(one pass) – t_d. For determining CCS values, the drift times of analytes (t_d(analyte) = t_a(analyte) – t_0 (an average of 6 measurements)) were divided by the number of passes. DriftScope 2.9 software detected peak maxima. An Excel file provided by Waters Corp. was employed evaluating CCS values. A calibration curve was constructed by plotting ln(CCS', Å^2) vs. ln(t_d') (Figure 5.1) and used along with the corrected drift times to estimate the collision cross sections of the ions under study (CCS*=A × t_d").

Table 5.6 Ions Used for CCS Calibration

(*from the Unified CCS Compendium; https://doi.org/10.1039/c8sc04396e)

m/z	z	Mol. mass (Da)	CCS (Å²)*
587.315	1	586.31	226.20
658.352	1	657.35	239.90
729.390	1	728.38	252.50
800.427	1	799.42	265.79
871.464	1	870.46	278.50

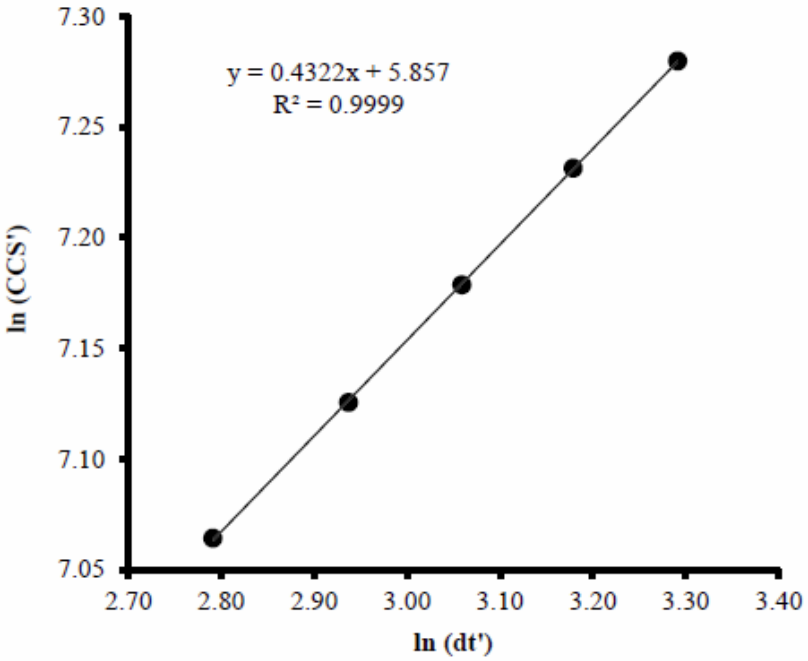
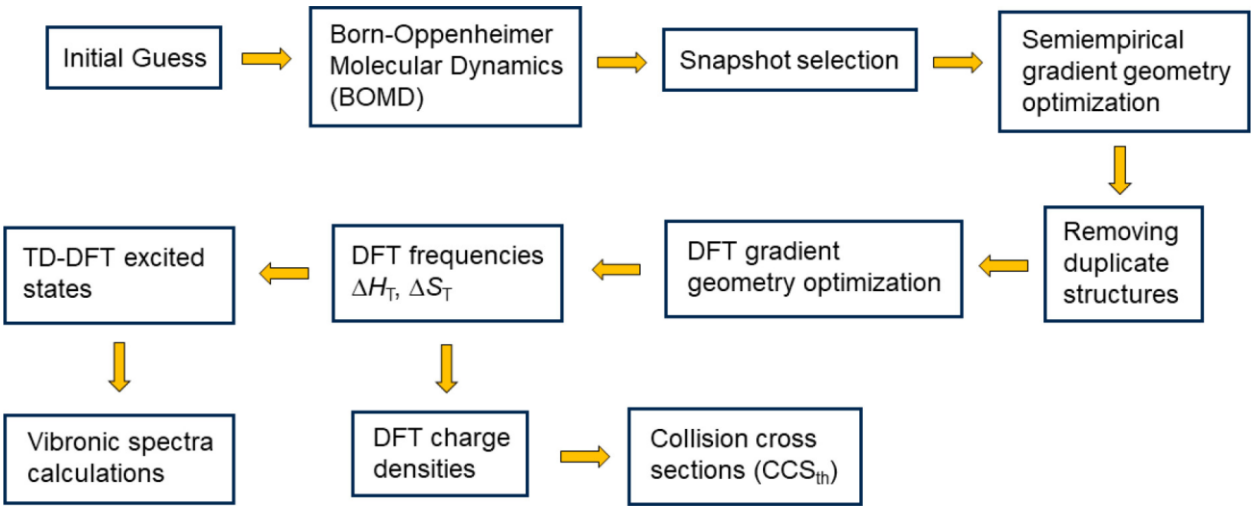


Figure 5.1: Calibration plot of ln(CCS') versus ln(d*t*').

5.2.2 Calculations

Ion structures were obtained by a multistep procedure shown in **Scheme 5.4**. Briefly, conformers were generated by Born–Oppenheimer molecular dynamics (BOMD) calculations of 20 ps trajectories with 1 fs steps that were performed by PM6-D3H4.[39] These calculations were run under the high-level Cuby4 platform[40] with the Berendsen bath at 510 K.[41] PM6 calculations were run under MOPAC.[42] It should be noted that because both electron and nuclear motions are treated explicitly by BOMD, the proton positions are not fixed and can move in the course of the

trajectory calculations. Two hundred snapshots of each trajectory were taken at regular intervals and fully gradient-optimized with PM6-D3H4. The resulting structures were compacted by removing duplicates and submitted to gradient geometry optimization with B3LYP [43] that was augmented by empirical dispersion terms[44] with Becke–Johnson damping (GD3-BJ).[45] B3LYP-GD3-BJ was also used to calculate harmonic frequencies. Additional sets of optimized geometries were obtained by M06-2X calculations[46] with the 6-31+G(d,p) basis set that were also used for single-point energy calculations with M06-2X/6-311++G(2d,p). Charge distributions were calculated according to Merz, Singh, and Kollman (MK).[47,48] Collision cross sections were calculated by the modified ion trajectory method ($MobCal_{MPI}$) [49,50] using the MK charge densities. Excited states were calculated by time-dependent density functional theory[51] using M06-2X/6-31+G(d,p). Typically, 15–20 excited states were considered. Vibronic absorption spectra were calculated for electronic excitations from 300 Wigner configurations[52,53] that were generated from B3LYP vibrational frequencies and sorted out by their Boltzmann factors. These calculations were run using the NewtonX software.[54]



Scheme 5.4: Workflow of BOMD and Density Functional Theory Calculations of Structures, Excited States, Vibronic Spectra, and Collision Cross Sections

5.3 Results and Discussion

5.3.1 Photodissociation of Tetrazole-Peptide Conjugates

The tet-peptide conjugates were protonated by electrospray ionization, selected by mass, and subjected to laser photodissociation. In addition, sodiated ions, (tet-peptide-COONa + Na)$^+$, were generated from solutions containing sodium acetate and used for UVPD. Photodissociation of all tet-peptide ions at 250 nm was efficient, achieving up to 50% conversions after just two 2 mJ laser pulses. This is illustrated with the (AAFA(tet)K + H)$^+$ ions (m/z 755) that were generated by electrospray and submitted to photodissociation (**Figure 5.2a**). Similar to previous studies in solution,[22] UVPD-MS^2 primarily resulted in N_2 loss, forming the transient nitrile imine intermediate at m/z 727. Loss of N_2 upon photodissociation of gas-phase 2,5-diaryltetrazole ions has been reported by Blanksby, Trevitt, and coauthors who measured the extinction coefficients as $\varepsilon \cong 2 \times 10^4$ L mol^{-1} cm^{-1} at λ_{max} = 255 nm.[37] Our UVPD efficiencies at 250 nm for tet-tagged peptides were consistent with the previous measurements. In addition to loss of N_2, the UVPD-MS^2 spectrum showed consecutive loss of C_2H_5N (m/z 684), which is an atypical peptide ion dissociation. Loss of N_2 was also the main dissociation of (AAFA(tet)K + H)$^+$ that was triggered by collision-induced dissociation (CID) that allowed us to measure accurate m/z and assign elemental composition to the ions (**Table 5.1**).

5.3.2 CID-MS^3 of Nitrile Imine Intermediates

Further analysis of the (AAFA(tet-N_2)K + H)$^+$ ions was carried out by UVPD-CID-MS^3 (**Figure 5.2b,c**). The spectrum showed internal sequence fragment ions by loss of Ala-Phe and Ala-Phe-Ala neutral internal residues, which are typically not produced from linear peptide chains,[55] while being indicative of cyclic peptide structures.[56] Accordingly, the reference CID spectrum of the (AAFA(tet)K+H)$^+$ ion showed a series of standard C-terminal $\boldsymbol{y_1}$–$\boldsymbol{y_3}$ fragment ions but no internal losses (**Figure 5.3**). Fragment ions of the $\boldsymbol{y_n}$ type were also present in the UVPD-CID-MS^3 (**Figure 5.2b**) and CID-CID-MS^3 spectra of (AAFA(tet-N_2)K + H)$^+$ (**Figure 5.4**). However, the pattern of relative intensities of these sequence ions from (AAFA(tet-N_2)K + H)$^+$ markedly differed from those in the **Figure 5.3** reference spectrum, indicating that they may be

formed by loss of internal neutral peptide residues instead. Another indication of cross-linking was by loss of neutral lysine ($C_6H_{14}N_2O_2$) from the m/z 438 ion (m/z 292, **Figures 5.2b** and **5.4**), which was incompatible with a $\boldsymbol{y_2}$ structure, but could be accommodated if the N-terminal Ala was covalently linked to the nitrile imine moiety. In addition to internal fragments, the **Figure 5.2b** spectrum also showed losses of C_2H_5N from the Ala N-terminus and $C_6H_8N_2$ from the nitrile imine group. The CID-MS^3 spectrum of (AAFA(tet-N_2)K + H)$^+$ ions that were generated by collision-induced loss of N_2 from the tetrazole ring was entirely similar to that of ions produced by photodissociation (**Figure 5.4**). The assignment of the neutral losses as being internal peptide fragments was corroborated by the UVPD-CID-MS^3 spectrum of (AGFA(tet-N_2)K + H)$^+$ that showed a loss of Gly-Phe and not Phe-Ala (**Figure 5.5**). All of these dissociations indicated that the nitrile imine group produced by photodissociation or CID was engaged in a ring-forming reaction with the peptide chain. The neutral fragments were identified by deuterium labeling. The ([D_9]-AAFA(tet-N_2)K + H)$^+$ ion (m/z 736, **Figure 5.2c**), generated from ([D_9]-AAFA(tet)K + H)$^+$ after exhaustive H/D exchange, showed loss of C_2H_4DN and $C_6H_5D_3N_2$ that pointed to acetaldimine and phenyl hydrazine for the respective neutral molecules.

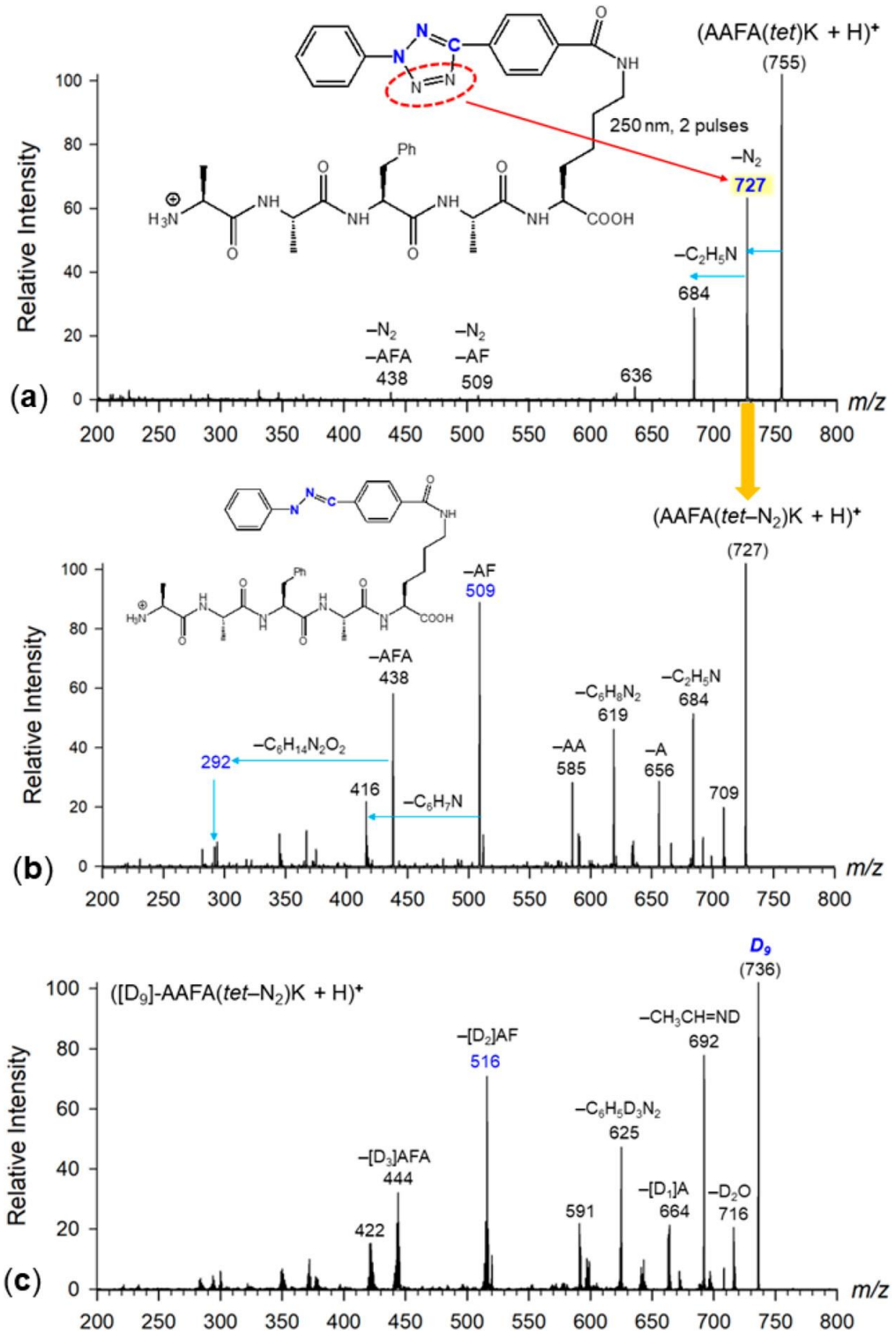


Figure 5.2: (a) UVPD-MS^2 spectrum of (AAFA(tet)K + $H)^+$ (m/z 755). (b) CID-MS^3 spectrum of (AAFA(tet-N_2)K + $H)^+$ (m/z 727) from (a). (c) CID-MS^3 spectrum of ([D_9]-AAFA(tet-N_2)K + $H)^+$ (m/z 736). For ion assignment by accurate mass measurements, see Table 5.1

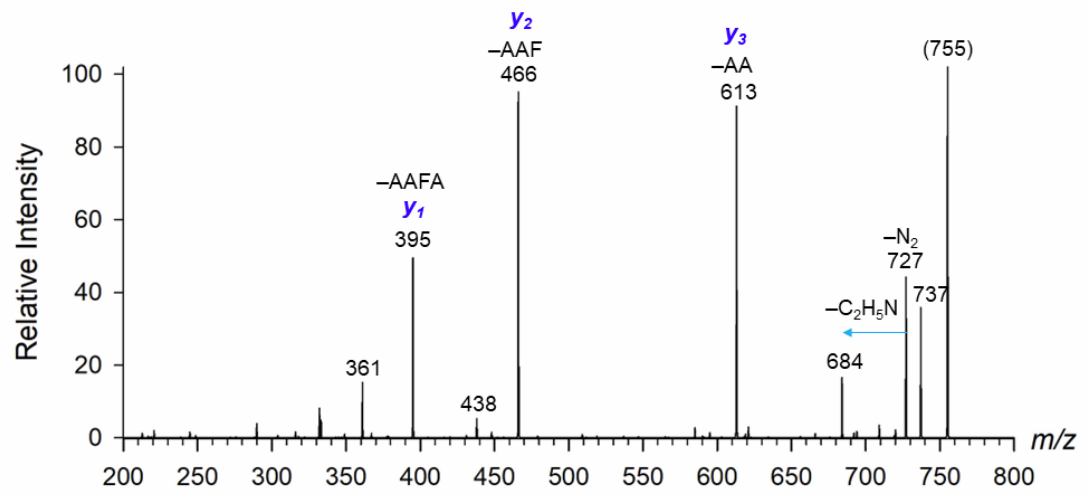


Figure 5.3: Reference CID-MS^2 spectrum of (AAFA(tet)K + H)$^+$, m/z 755.

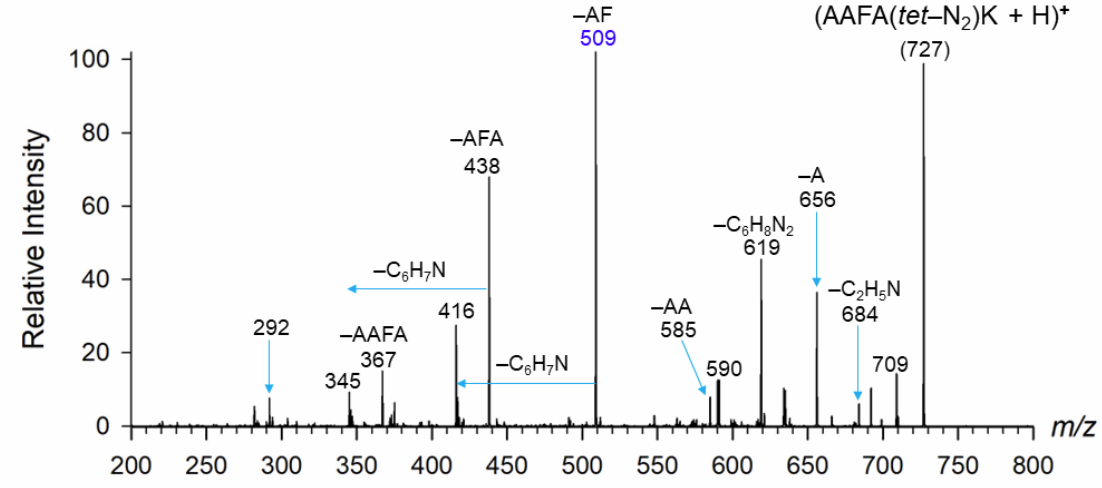


Figure 5.4: CID-CID-MS^3 spectrum of (AAFA(tet–N$_2$)K + H)$^+$, m/z 727.

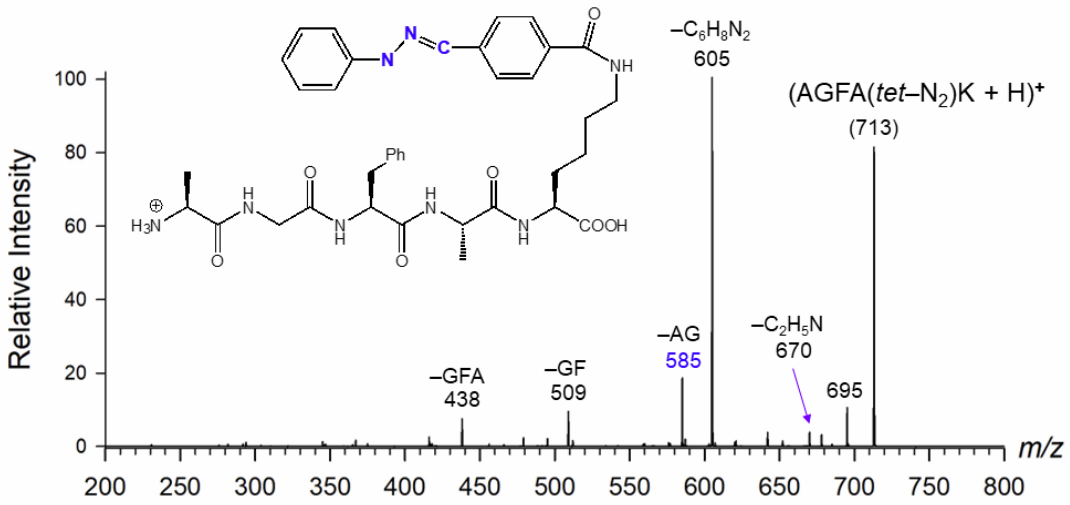


Figure 5.5: UVPD-CID-MS^3 spectrum of (AGFA(tet–N_2)K + H)$^+$, m/z 713.

We have further investigated the effect on nitrile imine formation and dissociations of modifying the C-terminus and N-terminal amino acid residues in the peptide-tet conjugates. Blocking the carboxyl terminus as a methyl ester in (AAFA(tet-N_2)K-OCH_3 + H)$^+$ gave CID-MS^3 that was analogous to that of (AAFA(tet-N_2)K + H)$^+$, showing loss of acetaldimine, phenylhydrazine, and internal Ala-Phe and Ala-Phe-Ala fragments (**Figure 5.6**). Modifications of the N-terminal residues, such as in (GAAA(tet)K + H)$^+$, (CAAA(tet)K + H)$^+$, (NAAA(tet)K + H)$^+$, and (QAAA(tet)K + H)$^+$, did not affect UVPD, resulting in an efficient loss of N_2 and nitrile imine formation in each case (**Figure 5.7**). Similarly, placing the tet-tagged lysine at the N-terminus as in ((tet)KAAAG + H)$^+$ resulted in an efficient N_2 loss upon UVPD (**Figure 5.8a**). The CID-MS^3 dissociations showed variations in the relative intensities of fragment ions, which, however, followed the general features displayed by (AAFA(tet-N_2)K + H)$^+$ (**Figure 5.2b**). For example, CID-MS^3 of the m/z 637 ion from (GAAA(tet-N_2)K + H)$^+$ showed prominent loss of CH_2═NH from the Gly N-terminus in addition to elimination of phenylhydrazine (**Figure 5.9a**). Internal fragments by loss of Ala-Ala and Ala-Ala-Ala at m/z 495 and 424, respectively, attested to a ring structure formed by nitrile imine cross-linking. The ion assignments were corroborated by accurate mass measurements (**Table 5.2**). CID-MS^3 of (CAAA(tet-N_2)K + H)$^+$ (**Figure 5.9b**) showed a very prominent loss of phenyl hydrazine in addition to the loss of C_2H_5NS from the N-terminal cysteine, represented as a $HSCH_2CH$═NH imine fragment, and internal fragments by loss of Ala-Ala and Ala-Ala-Ala that were accompanied by loss of H_2S. These dissociations indicated that the cysteine thiol group was not engaged in the ring-forming cross-linking. Note that nucleophilic

additions to nitrile imines have been observed in solution,[22,23] and hence the lack of reactivity for the cysteine thiol was remarkable. CID-MS^3 in the presence of side chain amide groups, such as in $(NAAA(tet\text{-}N_2)K + H)^+$ and $(QAAA(tet\text{-}N_2)K + H)^+$, resulted in an enhanced loss of ammonia. However, dissociations leading to the loss of the N-terminal group, $HN{=}CH{-}CH_2CONH_2$ and $HN{=}CH{-}CH_2CH_2CONH_2$ from asparagine and glutamine, respectively, were observed, as were losses of internal Ala-Ala and Ala-Ala-Ala fragments (**Figure 5.9c,d**). These common dissociations of the N-terminal aldimine fragments, HN=CH(R), pointed to cross-linking to the nitrile imine by an amide group within the peptide sequence. Some level of participation by the side-chain amide groups in Asn and Gln could not be completely excluded and may be the reason for the enhanced loss of ammonia from these peptide conjugates.

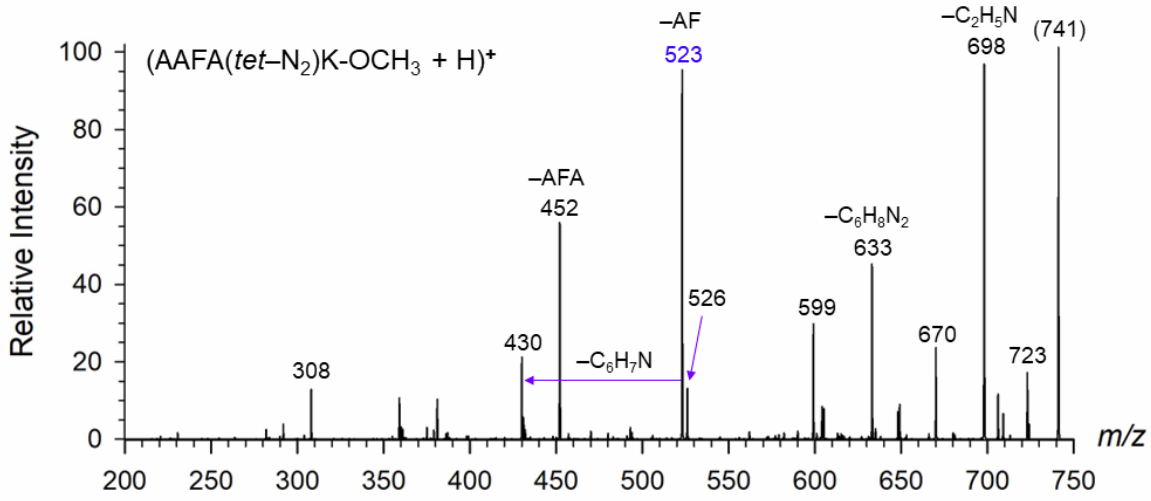


Figure 5.6: UVPD-CID-MS^3 spectrum of $(AAFA(tet{-}N_2)K\text{-}OCH_3 + H)^+$, m/z 741.

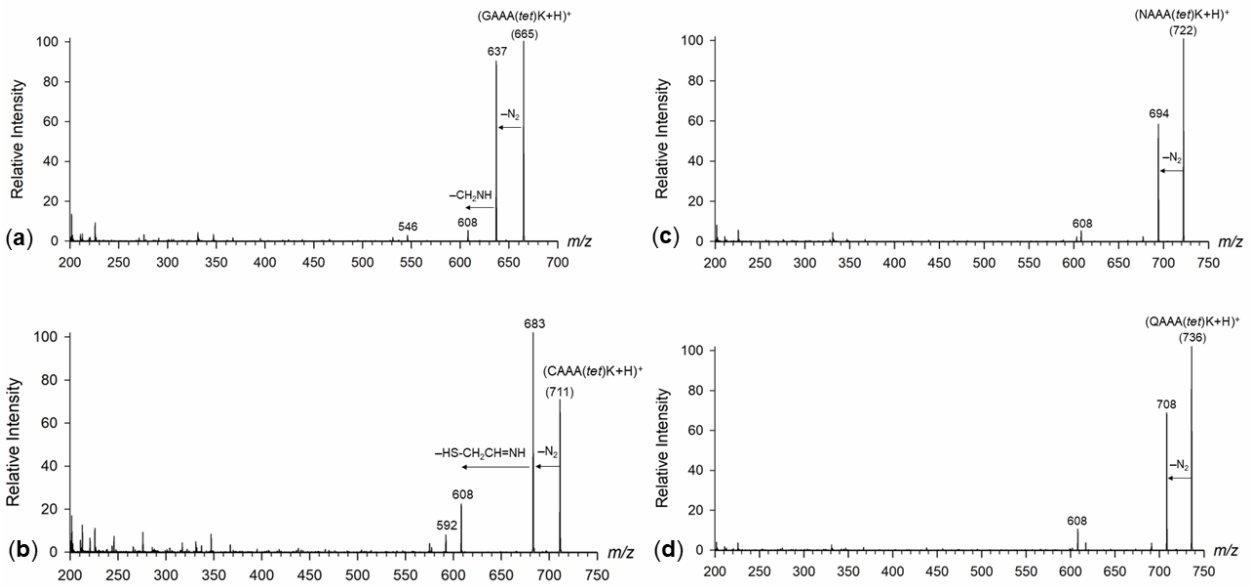


Figure 5.7: UVPD-MS^2 spectra, 2 laser pulses at 250 nm, of (a) (GAAA(tet)K + H)$^+$, m/z 665; (b) (CAAA(tet)K + H)$^+$, m/z 711; (c) (NAAA(tet)K + H)$^+$, m/z 722; (d) (QAAA(tet)K + H)$^+$, m/z 736.

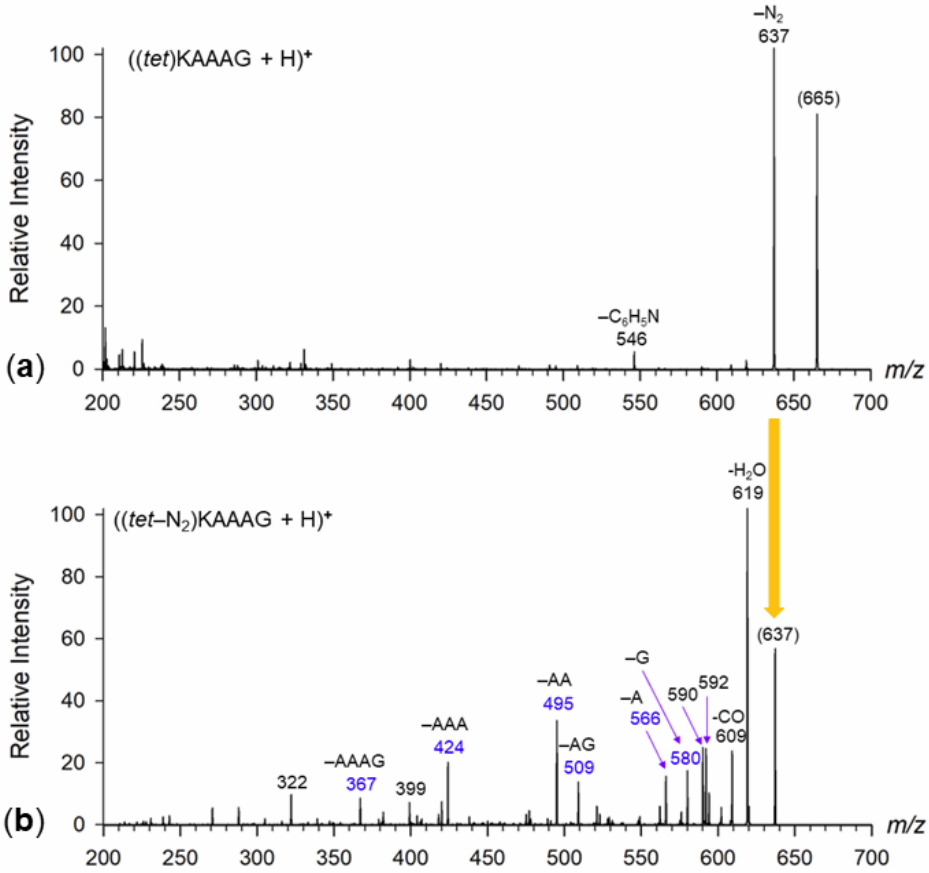


Figure 5.8: (a) UVPD-MS^2 spectrum, 2 laser pulses at 250 nm, of ((tet)KAAAG + H)$^+$, m/z 665; (b) UVPD-CID-MS^3 spectrum of ((tet–N$_2$)KAAAG + H)$^+$, m/z 637.

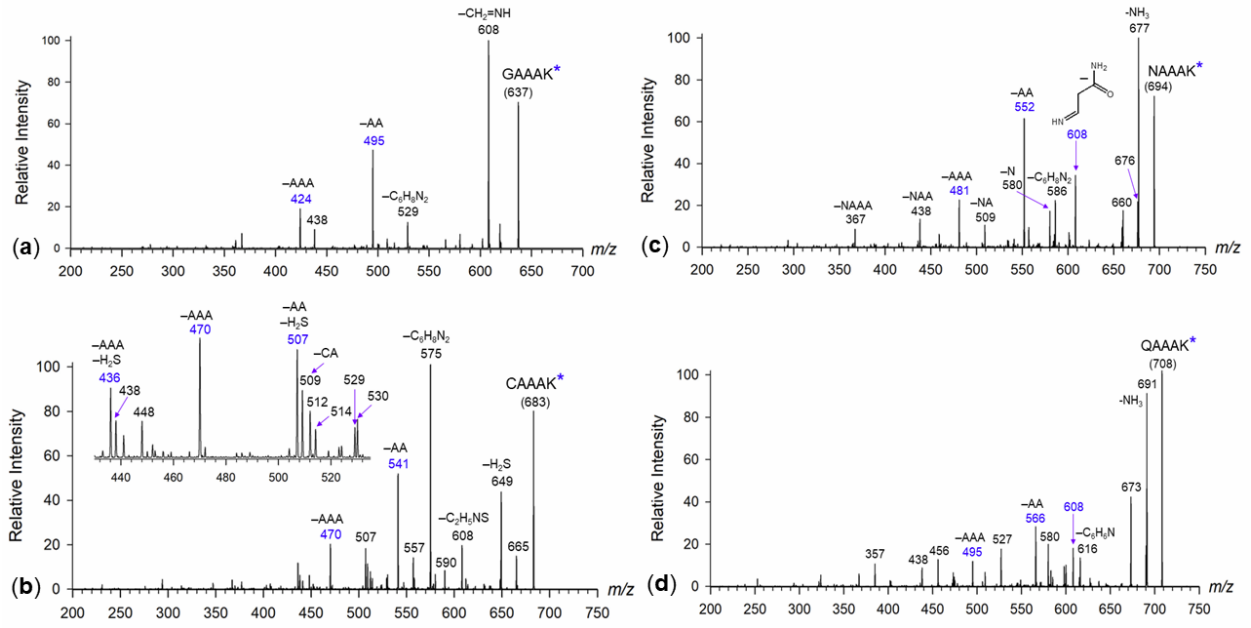


Figure 5.9: UVPD-CID-MS^3 spectra of (a) (GAAA(tet–N_2)K +H)$^+$, m/z 637; (b) (CAAA(tet–N_2)K + H)$^+$, m/z 683; (c) (NAAA((tet–N_2))K + H)$^+$, m/z 694; (d) (QAAA((tet–N_2))K + H)$^+$, m/z 708.

CID-MS^3 of the inverted sequence ion, ((tet-N_2)KAAAG + H)$^+$, showed a major loss of water and internal fragments by loss of Gly, Ala, Ala-Gly, Ala-Ala, and Ala-Ala-Ala (**Figure 5.8b**). Fragment ions of the bn type, which are characteristic of glycine C-terminated N-linked conjugates,[57] were absent in the CID-MS^3 spectrum of ((tet-N_2)KAAAG + H)$^+$, indicating nearly complete cyclization involving the glycine residue. In particular, the loss of Gly as $HNCH_2CO$ (57 Da) and Ala-Gly (128 Da) was indicative of a carboxyl linkage to the nitrile imine.

We further investigated the effect of the charging ion using sodiated conjugates, (AAFA(tet-N_2)K-COONa + Na)$^+$, m/z 771 (**Figure 5.10a**), and (GAAA(tet-N_2)K-COONa + Na)$^+$, m/z 681 (**Figure 5.10b**), that were produced by UVPD at 250 nm of the sodiated precursor ions. In both cases, the CID-MS^3 spectra showed a series of $\boldsymbol{y_1}$–$\boldsymbol{y_3}$ sequence ions that were identified by accurate mass measurements (**Tables 5.3** and **5.4**). In contrast to the above-described protonated ions, no internal fragments were found for the sodiated ions that would indicate cross-linking. This pointed to the possible role of the charging proton in promoting the peptide reactions with the nitrile imine group.

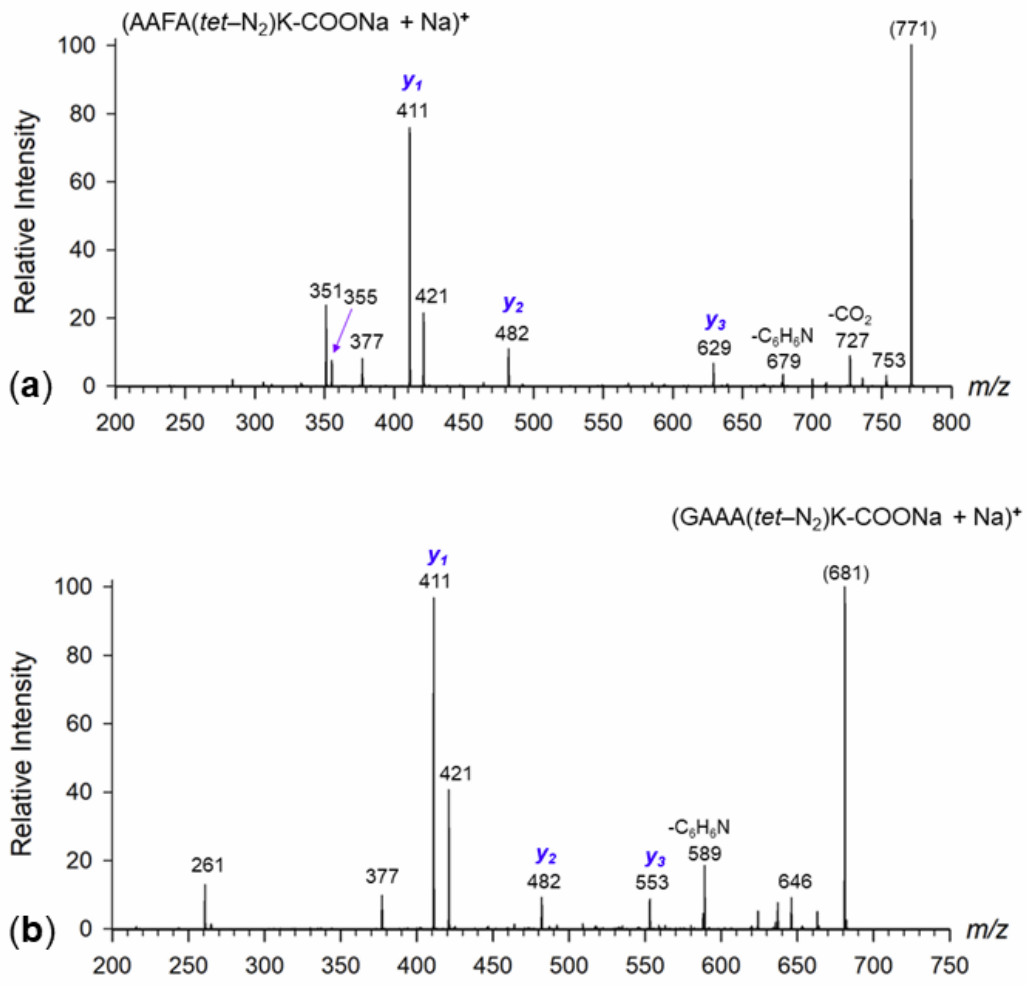


Figure 5.10: UVPD-CID-MS^3 spectra of (a) (AAFA(tet–N_2)K-COONa + Na)$^+$, m/z 771; (b) (GAAA(tet–N_2)K-COONa + Na)$^+$, m/z 681.

5.3.3 Structures and UV–Vis Action Spectra of Nitrile-Imine Cross-Links

The fundamental question regarding the observed cyclizations in protonated ions was whether they occurred upon the photochemical nitrile imine formation or were promoted by subsequent collisional activation of isolated stable nitrile imine ions. To answer this question, we used action spectroscopy to obtain the UV–vis spectra of the (AAFA(tet-N_2)K + H)$^+$ and (GAAA(tet-N_2)K + H)$^+$ ions. The spectra were obtained in the CID-UVPD mode where the nitrile imine ions used for the action spectra measurements were generated by CID-MS^2 from the tetrazole conjugates. Because CID in the ion trap is performed in a mass-resonant mode, the resulting (peptide(tet-N_2) + H)$^+$ ions are not further activated. The action spectrum of (GAAA(tet-N_2)K + H)$^+$, which has

only a nitrile-imine related chromophore absorbing in the 210–700 nm region, is representative (**Figure 5.11a**).

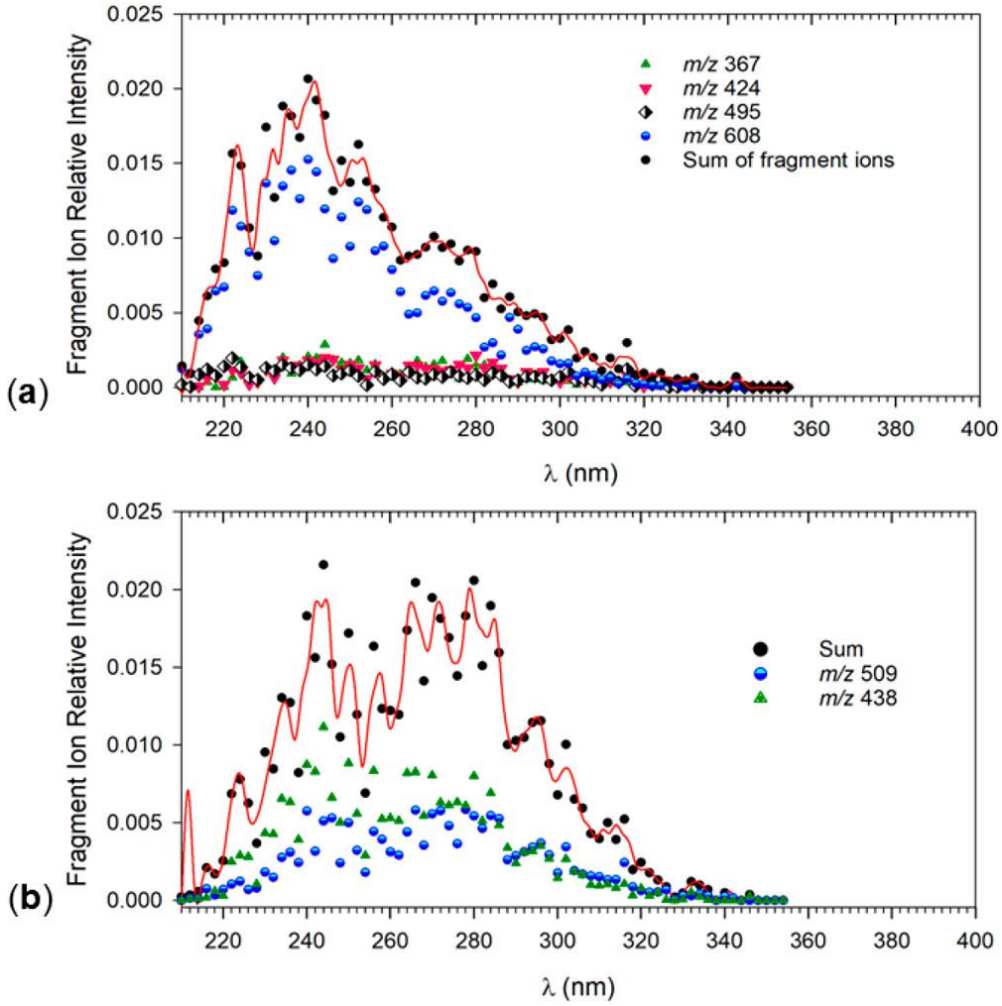


Figure 5.11: UVPD action spectra of (a) $(GAAA(tet\text{-}N_2)K + H)^+$ and (b) $(AAFA(tet\text{-}N_2)K + H)^+$. The red outline is a fit to the sum of photofragment ion intensities.

The action spectrum was monitored for the m/z 608 (loss of $CH_2{=}NH$, major fragment ion), m/z 495 (loss of Ala-Ala), m/z 424 (loss of Ala-Ala-Ala), and m/z 367 (loss of Gly-Ala-Ala-Ala) photofragment ion channels. The spectrum showed a composite band with maxima at 220, 240, and 250 nm, and another weaker broad band at 270 nm. These maxima appeared in at least two photofragment ion channels, namely, m/z 608 and m/z 495, which were indicative of cross-linked structures.

No absorption was detected above 350 nm. The action spectrum of (AAFA(tet-N_2)K + H)+ (**Figure 5.11b**) was monitored for the m/z 509 (loss of Ala-Phe) and m/z 439 (loss of Ala-Phe-Ala) channels. The spectrum showed a broad band consisting of three maxima at 240, 265–280, and 290–320 nm. The third band was visible in the m/z 509 ion channel. No absorption was observed above 350 nm.

To interpret the spectra and assign the absorption bands, we carried out extensive BOMD and density functional theory (DFT) calculations of ion structures and excited electronic states for (GAAA(tet-N_2)K + H)$^+$ and (AAFA(tet-N_2)K + H)$^+$. The initial guesses in these calculations were nitrile imine protomers and conformers, as well as cyclized structures involving functional groups that can potentially give rise to the major fragment ions in the CID-MS^3 spectra. These features involved a free neutral or protonated Gly CH_2NH_2 group to account for the loss of CH_2═NH, and a retained phenyl–N–N segment for the loss of phenylhydrazine. The lowest-energy structures of each type are shown in **Figure 5.12**. For several (GAAA(tet-N_2)K + H)$^+$ isomers, we also calculated vibronic absorption spectra at 350 K that included excitations from 300 configurations representing the vibrational states of the ions. The low-energy nitrile imines (**G1** and **G2)** showed vibronic spectra that strongly depended on the ion conformation. The lowest-energy conformer **G1** (**Figure 5.12**) showed a very weak transition to the first excited state at 438 nm (**Figure 5.13a**). The main transitions at 272, 269, and 246 nm gave rise to the compact band in the vibronic spectrum with a maximum at 270 nm. Vibronic broadening of the 269 and 272 nm bands then resulted in a band tail extending to 360 nm (**Figure 5.13a**). The other low-energy conformer (**G2**, 13 kJ mol^{-1} relative to **G1**, **Table 5.7**) showed a strong transition to the second excited state at 290 nm that was spread over a broad range of transitions in the thermal ion reaching beyond 400 nm (**Figure 5.13b**). Structures produced by simple proton transfer to the nitrile imine group (**G3**, 28 kJ mol^{-1} relative to **G1**) showed a strong transition to the first excited state at 330 nm (**Figure 5.13c**) that was incompatible with the action spectrum. The lowest energy structures belonged to hydrazidines (e.g., **G4**, −84 kJ mol^{-1} relative to **G1**) arising by nucleophilic attack at the nitrile imine by the Gly amine group. However, the vibronic spectrum of **G4** showed major bands at 220 nm and an absorption profile that was incompatible with the action spectrum (**Figure 5.13d**). Oxadiazoles were another type of product that could have arisen by the formal [3 + 2] addition of the Gly amide to the nitrile imine. The lowest energy oxadiazole isomer (**G5**, −68 kJ mol^{-1} relative

to **G1**) had a vibronic spectrum that showed a strong transition at 310 nm that resulted in bands at 315 and 345 nm that, however, were absent in the action spectrum (**Figure 5.13e**).

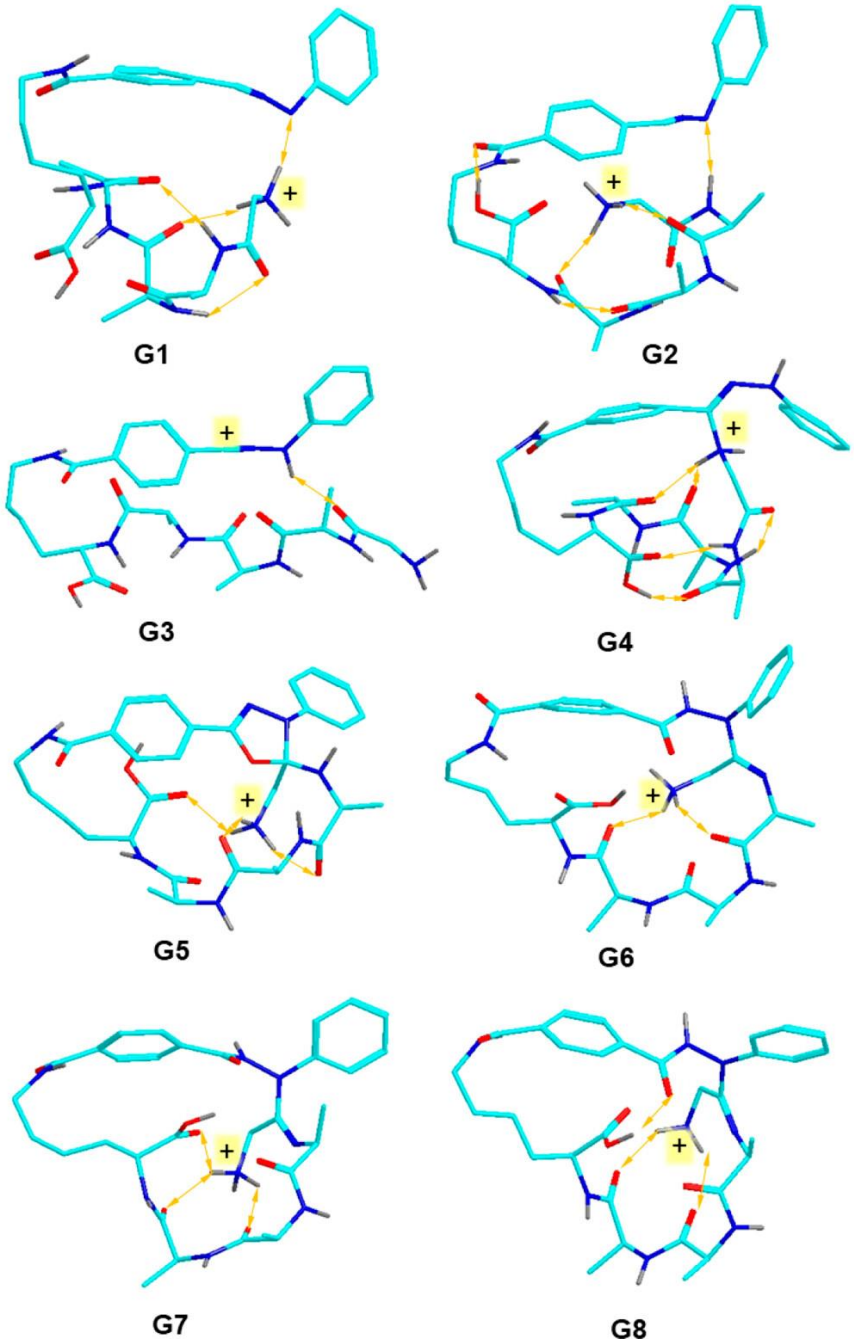


Figure 5.12: M06-2X/6-31+G(d,p) optimized structures of representative low-energy (GAAA(tet-N_2)K + H)$^+$ ions **G1**–**G8**. Atom color coding is as follows: cyan = C, blue = N, red = O, gray = H. Only exchangeable hydrogens are shown to avoid clutter. Hydrogen bonds are visualized by ochre double-headed arrows.

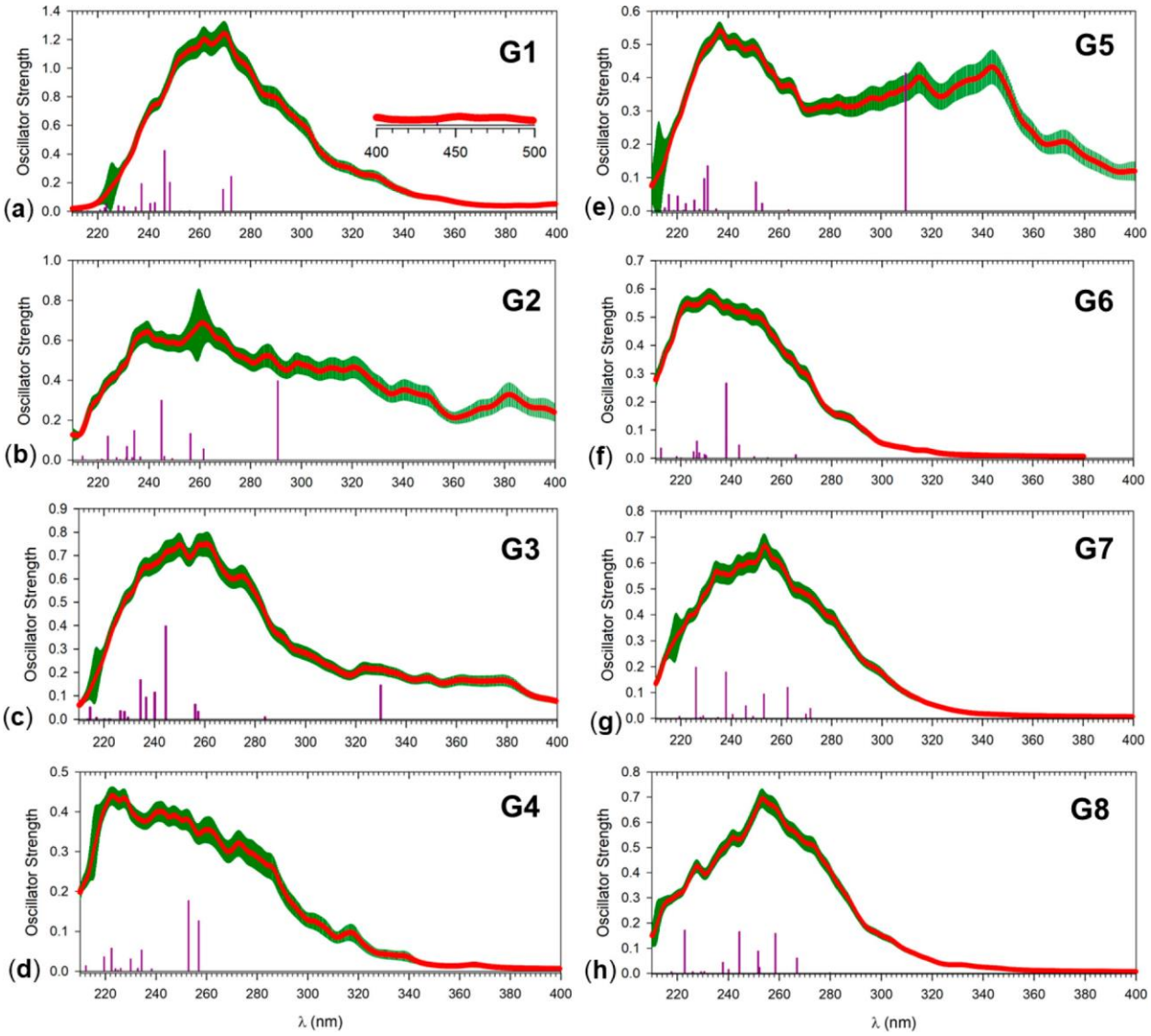


Figure 5.13: (a–h) M06-2X/6-31+G(d,p) vibronic spectra of (GAAA(tet-N_2)K + H)$^+$ isomers **G1**–**G8**. Vertical lines represent TD-DFT electron transitions at 0 K. For ion structures, see Figure 5.12.

Table 5.7 Calculated Relative Energies and Collision Cross Sections of (GAAA(tet-N_2)K + H)$^+$ Ions

ion	relative energy[a,b,c]	CCS[d]
G1	0 (0)	249.6 ± 1.0[e]
G2	13 (19)	245.3 ± 1.2
G3	28 (17)	257.5 ± 1.4
G4	–84 (−65)	231.5 ± 1.0

G5	–67 (−54)	241.3 ± 0.6
G6	–43 (−33)	251.3 ± 1.4
G7	–44 (−30)	241.5 ± 1.0
G8	–32 (−18)	242.9 ± 1.0

[a]In kJ mol^{-1}.
[b]From M06-2X/6-311++G(2d,p) single-point energy calculations including zero-point energies and referring to 0 K.
[c]Relative Gibbs energies at 310 K in parentheses.
[d]In Å^2.
[e]Standard deviations from multiple runs (n = 10).

Since, according to the calculations, the oxadiazole C–O bond in **G5** was weak (ΔH_{rxn} = 24 kJ mol^{-1}), breaking it was considered to be facile. Opening of the oxadiazole ring, followed by proton transfer, formed N–C linked hydrazones that are represented as conformers **G6** and **G7** (−44 and −43 kJ mol^{-1}, respectively, relative to **G1**). The respective vibronic spectra showed bands in the 240–260 nm region (**Figure 5.13f,g**), and the position of the maxima and the shape of these bands were in good agreement with the bands in the action spectrum. Another low-energy conformer of this type (**G8**, **Figure 5.12**) had a vibronic spectrum (**Figure 5.13h**) that was very similar to those of **G6** and **G7**, illustrating the general features of the diarylhydrazone chromophore. In summary, structures **G1** and **G6**–**G8** provided the best match with the action spectrum of the (GAAA(tet-N_2)K + H)$^+$ ion.

Electronic excited states were also analyzed using TD-DFT calculations of several potential isomers of the (AAFA(tet-N_2)K + H)$^+$ ions. Because of the larger size of these ions, we were unable to calculate vibronic spectra, and so the assignment relied on the 0 K vertical excitations only, whereas absorption profiles due to thermal band shifts and broadening were not included. The optimized structures and calculated TD-DFT transitions are shown in **Figure 5.14**. Overall, the structure assignment for (AAFA(tet-N_2)K+ H)$^+$ was consistent with that discussed above for (GAAA(tet-N_2)K + H)$^+$. In particular, the lowest-energy nitrile imine conformer (**F1**) had a strong transition at 339 nm that was absent in the action spectrum. Protonated diarylhydrazine (**F2**) formed by N-terminal amine addition to the nitrile imine had strong transitions in the 210 nm region that were not present in the action spectrum. Oxadiazole adducts, as represented by structures **F3** and **F4**, showed strong transitions at 285 nm that were expected to be red-shifted at

310 K and would not match the action spectrum of $(AAFA(tet\text{-}N_2)K + H)^+$. A tentative match can be expected for hydrazidine **F5** where the electron transitions at 230–260 nm (**Figure 5.14**) are likely to be red-shifted in a vibronic spectrum to form bands at 240–280 that were observed in the action spectrum of $(AAFA(tet\text{-}N_2)K + H)^+$. Note that structure **F5** was protonated in the hydrazidine group, which differed from **G6**–**G8** that were protonated in the N-terminal amine (**Figure 5.12**).

5.3.4 Cyclic Ion Mobility Measurements

In an effort to further characterize the ions of interest, we investigated the composition and structures of the peptide tetrazole conjugates and their cross-linked products using cyclic ion mobility spectrometry (c-IMS).[38] This allowed us to separate isomers by their arrival times and obtain collision cross sections (CCS_{exp}) in N_2. The $(AAFA(tet)K + H)^+$ and $(GAAA(tet)K + H)^+$ ions showed single peaks after five passes (490 cm path length) of c-IMS, indicating structurally homogeneous ions (**Figure 5.15**). The measured CCS_{exp} = 273 and 256 $Å^2$ for $(AAFA(tet)K + H)^+$ and $(GAAA(tet)K + H)^+$, respectively, were matched by the calculated CCS_{th} for the lowest Gibbs-energy conformers of these ions (**Figure 5.15**). The ions of both sequences were protonated at the N-terminus regardless of the initial proton position in structures that were submitted to BOMD. The optimized structures showed globular peptide moieties that were maintained by multiple intramolecular hydrogen bonds. The 2,5-diaryl tetrazole moiety was positioned at the ion periphery and, with the exception of its carbonyl, was not linked by hydrogen bonds to the peptide polar groups. This indicated that the tetrazole group may be viewed as a spectator without significantly affecting the peptide ion conformation. However, the rigid structure of the 2,5-diaryl tetrazole moiety limits access of the incipient nitrile imine to amide bonds and other functional groups to be located close to the pentapeptide N-terminus. From the optimized structures, one can assess the minimum number of four residues that separate the tetrazole ring and the incipient nitrile imine from the nearest sterically possible cross-linking site.

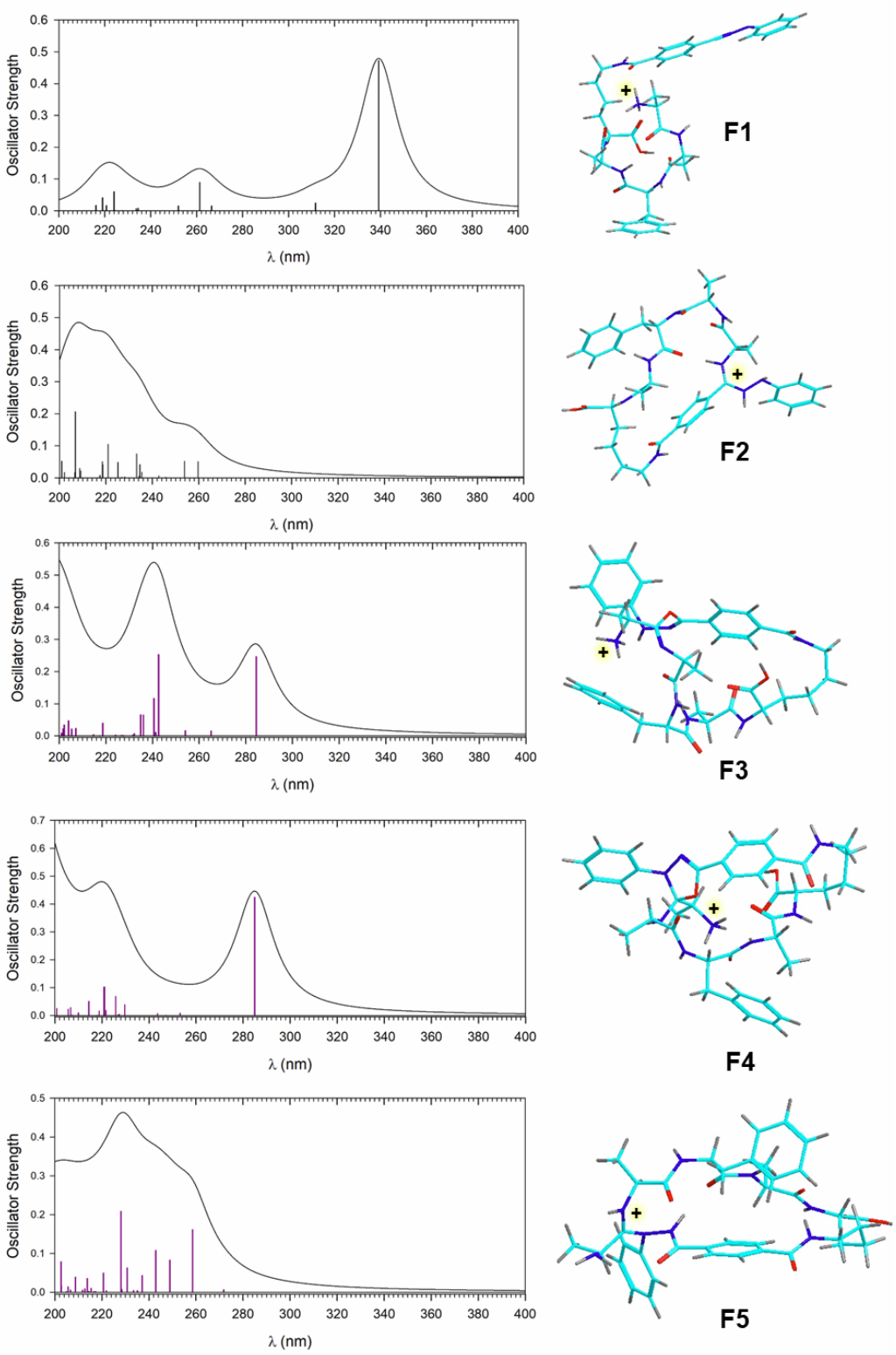


Figure 5.14: M06-2X/6-31+G(d,p) optimized structures and TD-DFT electron transitions of (AAFA(tet–N_2)K + H)$^+$ ions. The band shapes are from artificial convolution with Lorenztian functions with a full width at half maximum of 12 nm.

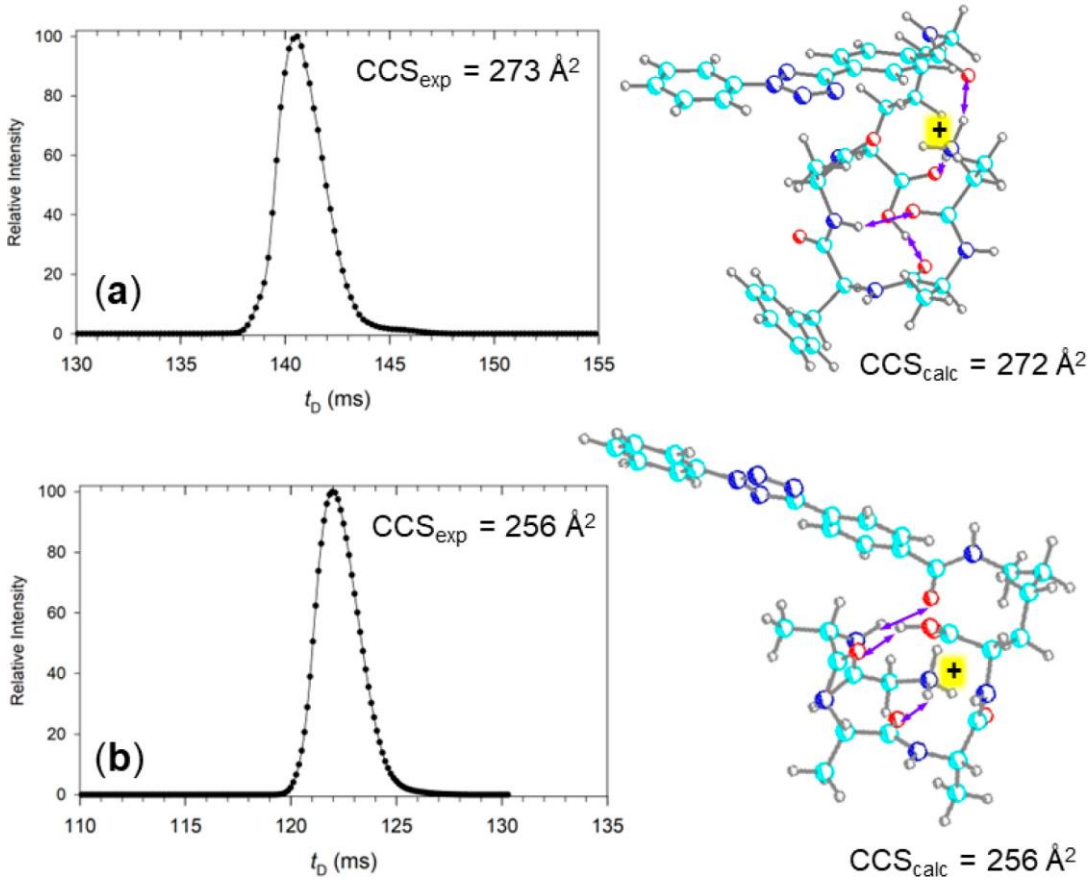


Figure 5.15: Arrival time profiles (5 cycles), optimized structures, and CCS of (a) (AAFA(tet)K + H)$^+$ and (b) (GAAA(tet)K + H)$^+$. Atom color coding is as in Figure 5.12. Hydrogen bonds are visualized by double-headed purple arrows.

In contrast to the precursor ions, the (AAFA(tet-N_2)K + H)$^+$ and (GAAA(tet-N_2)K + H)$^+$ ions generated by CID were mixtures of isomers (**Figure 5.16**). According to c-IMS, (AAFA(tet-N_2)K + H)$^+$ produced one major peak of CCS$_{exp}$ = 268 Å^2, amounting to ca. 80% of the total ion intensity.

This major peak was flanked by two minor peaks (**Figure 5.16a**). (GAAA(tet-N_2)K + H)$^+$ produced two major peaks of CCS$_{exp}$ = 246 and 254 Å^2 that were accompanied by two additional minor peaks (**Figure 5.16b**). The CCS$_{exp}$ values of (GAAA(tet-N_2)K + H)$^+$ were compared to those of CCS$_{th}$ that were calculated for several structure types (**G1**–**G8**, **Table 5.7**). Nitrile imines **G1** and **G2** had CCS$_{th}$ in the 246–250 Å^2 range that overlapped with the 246 Å^2 peak in **Figure 5.16b**. Considering the minor band at 270 nm in the **Figure 5.11a** action spectrum that can be attributed to a nitrile imine, the combined data indicated the presence of an unreacted nitrile imine in the mixture. Consistent with this conclusion, the CID-MS^3 spectrum of (GAAA(tet-N_2)K + H)$^+$ (**Figure 5.9a**) showed a minor $\boldsymbol{y_2}$ ion at m/z 438, which indicated loss of a Gly-Ala-Ala $\boldsymbol{b_3}$ neutral

fragment from the straight peptide chain. The CCS_{th} of the protonated nitrile imine isomer **G3** (257.5 $Å^2$) was close to that of the second major peak in the c-IMS arrival time profile. However, isomer **G3** was excluded on the basis of its incompatible absorption spectrum (**Figure 5.13c**). The low-energy hydrazidine **G4** had a CCS_{th} that did not match any of the experimental values and, in accordance with its nonmatching vibronic spectrum (**Figure 5.13d**), was excluded. The CCS_{th} of oxadiazole **G5** was at the low end of the CCS_{exp} values; however, its vibronic spectrum (**Figure 5.13e**) did not match the action spectrum. The CCS_{th} of hydrazones **G6** (251.3 $Å^2$) and **G7** (241.5 $Å^2$) matched within −1.1% and −1.8%, respectively, the CCS_{exp} of the major peaks from c-IMS. In addition, another conformer of the same type (**G8**) had CCS_{th} = 243 $Å^2$, which was within −1.1% of the CCS_{exp} of the major c-IMS peak. In addition to having closely related CCS, structures **G6**–**G8** also had vibronic spectra that match the major bands in the action spectrum of $(GAAA(tet\text{-}N_2)K + H)^+$. Hence, the combination of c-IMS and UV–vis action spectroscopy provided data that upon analysis allowed us to assign probable structures to the reaction products while excluding those that showed close fits in only one type of experimental data.

The formation of cross-linked structures **G6**–**G8** and **F5** raised the question of the pertinent reaction mechanism(s). These structures showed the formation of a C═O bond by amide oxygen transfer to the nitrile imine carbon, and a new C–N bond to the imine nitrogen. We proposed a tentative mechanism that accounts for these changes in $(GAAA(tet\text{-}N_2)K + H)^+$, which is shown in **Scheme 5.5**. A formal [3 + 2] addition of the Gly amide to the nitrile imine was presumed to initially form intermediate **3**. This step may be facilitated by proton transfer to the nitrile imine; note that the role of catalytical initial proton transfer was corroborated by the absence of cross-linking in sodium ion adducts (**Figure 5.10**) that lacked active protons. The protonated oxapyrazoline ring in **3** presumably opened by breaking the quaternary C–O bond, forming a hydrazide (**4**), which in $(GAAA(tet\text{-}N_2)K + H)^+$ was finally stabilized by proton migration back to the N-terminal amine group, producing conformers **2a**,**2b**. Structures **2a**,**2b** were compatible with the major dissociations of $(GAAA(tet\text{-}N_2)K + H)^+$ ions upon CID that showed loss of CH_2═NH and $PhNHNH_2$ (**Figure 5.9a**) and analogously for the other XAAAK-tet sequences (**Figure 5.9b–d**). The final step of proton migration (**4** → **2a,b**) may be reversible, depending on the basicity of the pertinent nitrogen atoms and proton stabilization by hydrogen bonding. For example, the $(AAFA(tet\text{-}N_2)K + H)^+$ structure **F5** was protonated in the hydrazidine group (**Figure 5.14**), analogous to intermediate **4** in **Scheme 5.5**.

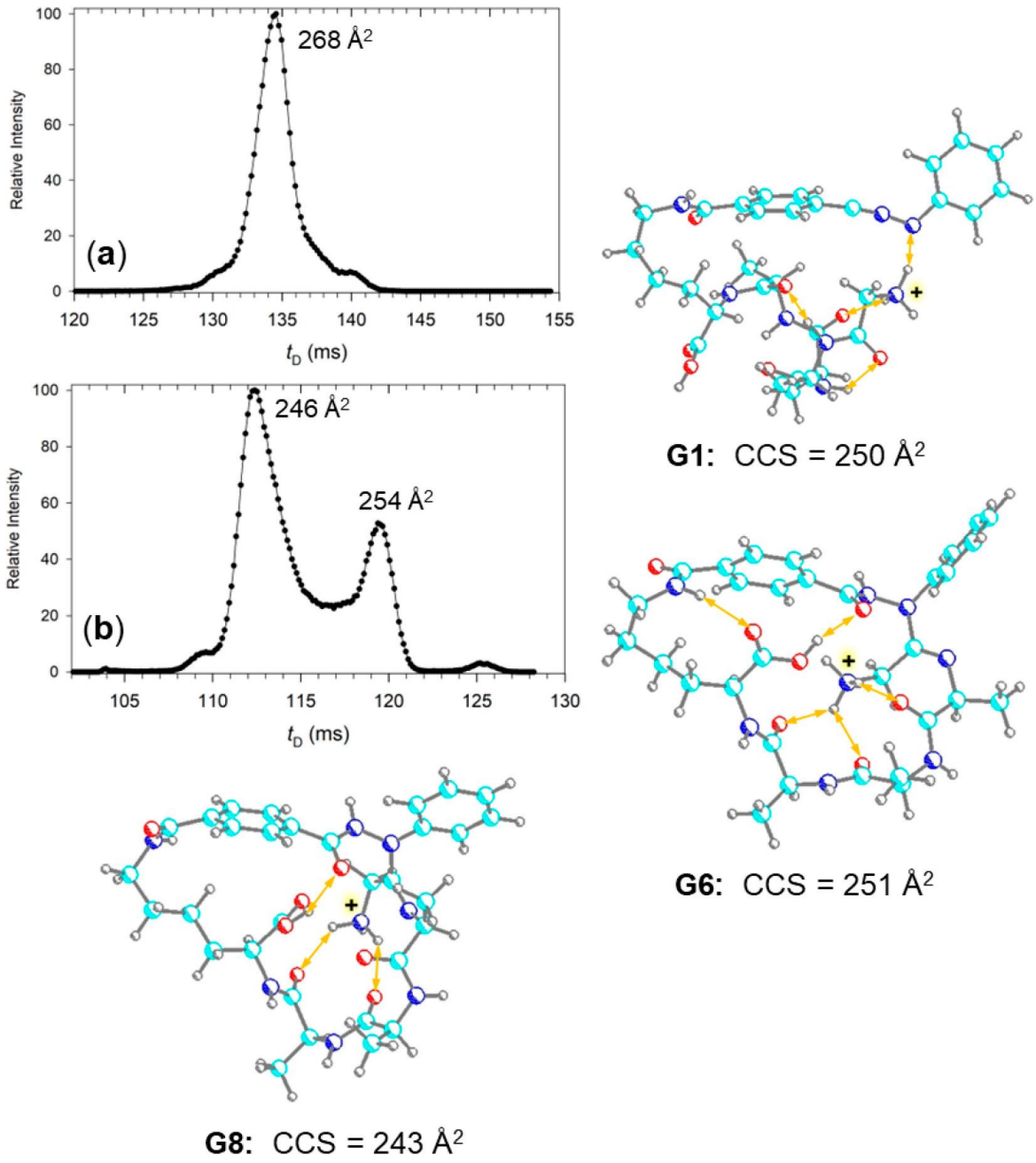
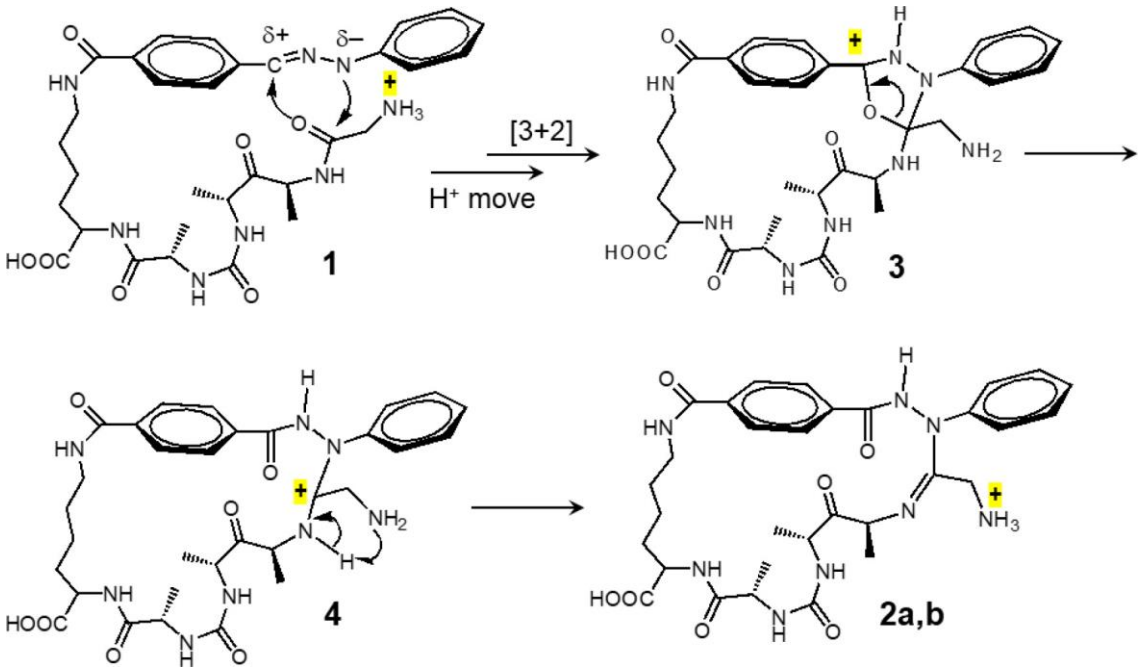
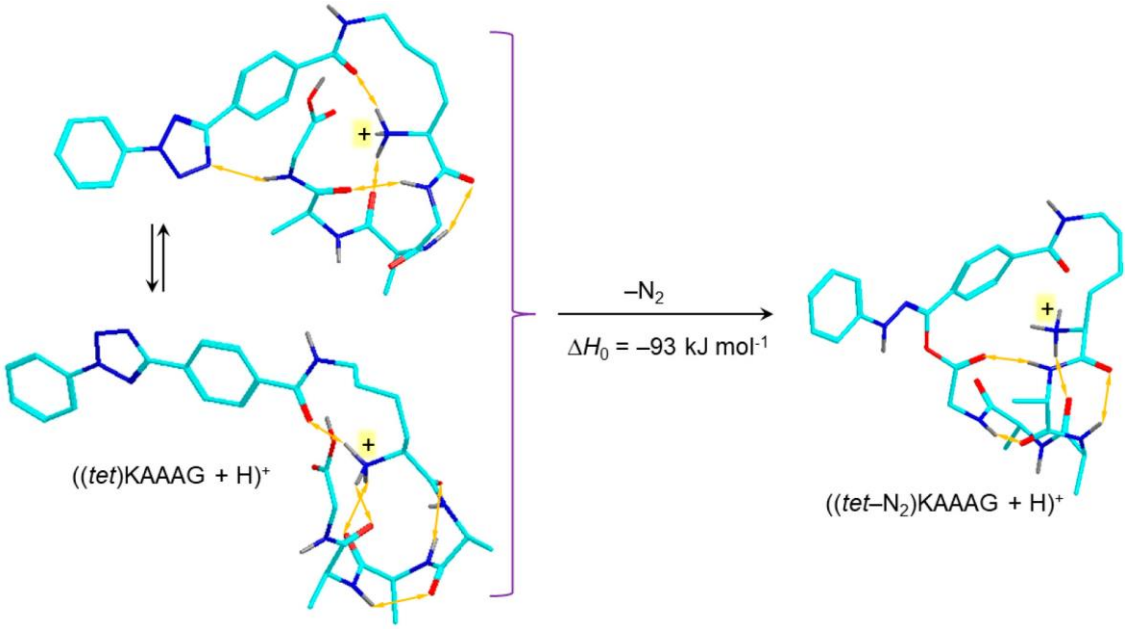


Figure 5.16: Arrival time distributions after five passes of (a) $(AAFA(tet\text{-}N_2)K + H)^+$ and (b) $(GAAA(tet\text{-}N_2)K + H)^+$ with experimental CCS_{exp} for the major peaks and matching calculated ion structures.



Scheme 5.5: Proposed Mechanism for Nitrile Imine-Amide Cross-Linking

Cross-linking to the C-terminal residue, as in $((tet\text{-}N_2)KAAAG + H)^+$, may involve simple nucleophilic attack by the carboxyl at the nitrile imine that is promoted by proton transfer. **Scheme 5.6** indicates that the overall reaction from $((tet)KAAAG + H)^+$ to the cyclized $((tet\text{-}N_2)KAAAG + H)^+$ product is substantially exothermic and thus possible following photodissociation. The conformations of the lowest-energy $((tet)KAAAG + H)^+$ precursor ions in **Scheme 5.6** allow a facile access of the carboxyl proton and oxygen to the nitrile imine for proton transfer and ring closure. The role of the charged N-terminal NH_3 group in reversibly providing the proton has not been established and would require an extensive transition-state search to be elucidated.



Scheme 5.6. Loss of N_2 and Cyclization in ((tet)KAAAG + H)$^{+a}$

[a]The reaction enthalpy is from M06-2X/6-311++G(2d,p) + ZPVE calculations and refers to 0 K. Atom color coding is as in Figure 5.12.

5.3.5 Nitrile Imine Cross-Linking in a Peptide-Dinucleotide Complex

To further explore the nitrile-imine reactivity toward other functional groups, we investigated the noncovalent ion complex of dinucleotide 2′-deoxycytidylguanosine (***dCG***) with the GAAA(tet)K peptide. The singly protonated complex, (***dCG***-GAAA(tet)K + H)$^+$, was produced by electrospray ionization, selected by mass (m/z 1221), and submitted to UVPD. **Figure 5.17a** shows that efficient (>50%) photodissociation was achieved in 2 laser pulses at 250 nm that triggered loss of N_2 from the tetrazole ring. CID-MS^3 of the mass-selected photoproduct (m/z 1193, **Figure 5.17b**) resulted in a highly selective loss of cytosine (***C***, m/z 1082) and deoxycytidine (***dC***, m/z 984) by standard phosphoric ester elimination.[58,59] In contrast, no loss of guanosine was observed, as evidenced by the lack of a m/z 944 fragment ion. Only a very minor fraction of (***dCG***-GAAA(tet-N_2)K + H)$^+$ underwent dissociation to the components, as documented by the very weak peak of (GAAA(tet-N_2)K + H)$^+$ at m/z 637 (**Figure 5.17b**). The spectra can be unequivocally interpreted as evidence for highly efficient and selective nitrile imine cross-linking to the guanine ring. BOMD + DFT calculations of (***dCG***-GAAA(tet)K+ H)$^+$ precursor complexes indicated low-energy structures in which cytosine was protonated at N-3, and the guanine ring was wedged in a

loop formed by GAAA(tet)K (**Figure 5.17a**, inset). We note that protonation of free ***dCG*** favors the guanine N-7 position,[60] so the low energy of cytosine-protonated (***dCG***-GAAA(tet)K + H)$^+$ must be due to the nucleobase coordination to the tetrazole system. This complex geometry may favor proton transfer to the nitrile imine, followed by addition to the guanine ring. The guanine bonds engaged in the nitrile-imine reaction (C═O, N7–C8, etc.), structures of the cross-link(s), and the mechanism of this new reaction are currently being investigated.

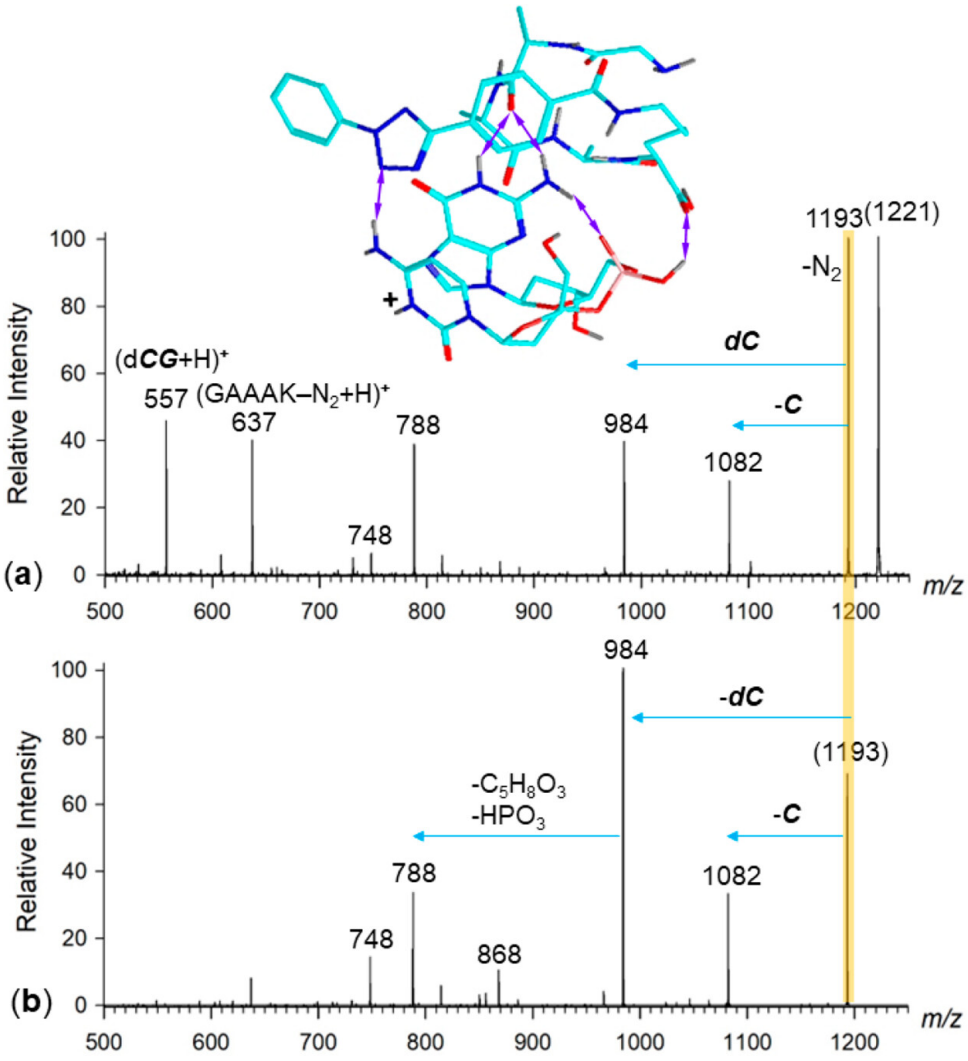


Figure 5.17: (a) UVPD-MS^2 spectrum of the (***dCG***-GAAA(tet)K + H)$^+$ complex. Inset shows the M06-2X/6-31+G(d,p)-optimized structure of a low-energy complex ion. Only exchangeable hydrogens are shown to avoid clutter. (b) CID-MS^3 spectrum of (***dCG***-GAAA(tet-N_2)K + H)$^+$. Atom color coding is as in Figure 5.12. Hydrogen bonds are visualized by double-headed purple arrows.

5.4 Conclusions

Nitrile imine intermediates that were generated from peptide-2,5-diaryltetrazole conjugates were found to undergo intramolecular cyclizations with peptide amide and carboxyl groups. The generation of the nitrile imine intermediates in conjugates can be done efficiently by CID or UVPD at 250 nm. UVPD is the preferred method for nitrile imine generation in noncovalent ion complexes, which is likely to be the major future application of these novel reactions in gas-phase ions. In particular, the discovered reactivity of peptide amide groups toward nitrile imines appears promising as it provides potential clues to cross-link structure elucidation and conformational analysis. The role of multiple charging, amino acid side-chain functional groups, post-translational modifications, and nucleobases other than guanine in the complexes and their effect on nitrile-imine based cross-linking are the subjects of current research directions in order to explore the applicability of this new method.

5.5 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.

[2] Shigdel, U. K.; Zhang, J.; He, C. Diazirine-Based DNA Photo-Cross-Linking Probes for the Study of Protein-DNA Interactions. *Angew Chem Int Ed* **2008**, *47*, 90.

[3] 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.

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

[5] Dubinsky, L.; Krom, B. P.; Meijler, M. M. Diazirine Based Photoaffinity Labeling. *Bioorg Med Chem* **2012**, *20*, 554.

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

[7] 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.

[8] 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.

[9] Pepin, R.; Shaffer, C. J.; Tureček, F. Position-Tunable Diazirine Tags for Peptide-Peptide Ion Crosslinking in the Gas-Phase. *J Mass Spectrom* **2017**, *52*, 557.

[10] 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.

[11] 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.

[12] 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.

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

[14] Liu, Y.; Liu, Y.; Nytka, M.; Huang, S. R.; Lemr, K.; Tureček, F. Probing D- and L-Adrenaline Binding to B2-Adrenoreceptor Peptide Motifs by Photodissociation Crosslinking and Ion Mobility Mass Spectrometry. *J Am Soc Mass Spectrom* **2021**, *32*, 1041.

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

[16] Lee, S.; Valentine, S. J.; Reilly, J. P.; Clemmer, D. E. Controlled Formation of Peptide Bonds in the Gas Phase. *J Am Chem Soc* **2011**, *133*, 15834.

[17] McGee, W. M.; McLuckey, S. A. Efficient and Directed Peptide Bond Formation in the Gas Phase via Ion/Ion Reactions. *Proc Natl Acad Sci U A* **2014**, *111*, 1288.

[18] Cheung See Kit, M.; Webb, I. K. Application of Multiple Length Cross-Linkers to the Characterization of Gaseous Protein Structure. *Anal Chem* **2022**, *94*, 13301.

[19] Liu, M. T. H. *Chemistry of Diazirines*; 1987; Vol. I.

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

[21] Julian, R. R.; May, J. A.; Stoltz, B. M.; Beauchamp, J. E. Biomimetic Approaches to Gas Phase Peptide Chemistry: Combining Selective Binding Motifs with Reactive Carbene Precursors to Form Molecular Mousetraps. *Int J Mass Spectrom* **2003**, *228*, 851.

[22] Sharp, J. T.; Padwa, A.; Pearson, W. H. *Chemistry of Heterocyclic Compounds 59: Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products*; 2002.

[23] Shawali, A. S. Reactions of Heterocyclic Compounds with Nitrilimines and Their Precursors. *Chem Rev* **1993**, *93*, 2731.

[24] Su, Y.; Zhao, Y.; Chang, B.; Zhao, X.; Zhang, R.; Liu, X.; Huang, D.; Wang, K.-H.; Huo, C.; Hu, Y. [3 + 2] Cycloaddition of Para-Quinone Methides with Nitrile Imines: Approach to Spiro-Pyrazoline-Cyclohexadienones. *J Org Chem* **2019**, *84*, 6719.

[25] Voronin, V. V.; Ledovskaya, M. S.; Gordeev, E. G.; Rodygin, K. S.; Ananikov, V. P. [3 + 2]-Cycloaddition of in Situ Generated Nitrile Imines and Acetylene for Assembling of 1,3-Disubstituted Pyrazoles with Quantitative Deuterium Labeling. *J Org Chem* **2018**, *83*, 3819.

[26] Guo, C.-X.; Zhang, W.-Z.; Zhang, N.; Lu, X.-B. 1,3-Dipolar Cycloaddition of Nitrile Imine with Carbon Dioxide: Access to 1,3,4-Oxadiazole-2(3H)-Ones. *J Org Chem* **2017**, *82*, 7637.

[27] Spiteri, C.; Keeling, S.; Moses, J. E. New Synthesis of 1-Substituted-1H-Indazoles via 1,3-Dipolar Cycloaddition of in Situ Generated Nitrile Imines and Benzyne. *Org Lett* **2010**, *12*, 3368.

[28] Tu, L.; Gao, L.; Wang, X.; Shi, R.; Ma, R.; Li, J.; Lan, X.; Zheng, Y.; Liu, J. [3 + 2] Cycloaddition of Nitrile Imines with Enamides: An Approach to Functionalized Pyrazolines and Pyrazoles. *J Org Chem* **2021**, *86*, 559.

[29] Fournier-Le Ray, N.; Joly, N.; Fillaut, J. L. Optimising the Synthesis of 2,5-Diaryltetrazoles: The Decisive Choice of the Reaction Solvent. *Tetrahedron* **2023**, *143*, 133560.

[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* (15), 2832–2835. https://doi.org/10.1002/anie.200705805.

[31] Bégué, D.; Qiao, G. G.; Wentrup, C. Nitrile Imines: Matrix Isolation, IR Spectra, Structures, and Rearrangement to Carbodiimides. *J Am Chem Soc* **2012**, *134*, 5339.

[32] Bégué, D.; Wentrup, C. Carbenic Nitrile Imines: Properties and Reactivity. *J Org Chem* **2014**, *79*, 1418.

[33] Moriwaki, Y.; Tachikawa, M.; Maeno, Y.; Shimizu, T. Collision Cooling of Ions Stored in Quadrupole Radio-Frequency Trap. *Jpn J Appl Phys* **1992**, *31*, 1640.

[34) Pepin, R.; Tureček, F. Kinetic Ion Thermometers for Electron Transfer Dissociation. *J Phys Chem B* **2015**, *119*, 2818.

[35] Martens, J.; Berden, G.; Gebhardt, C. R.; Oomens, J. Infrared Ion Spectroscopy in a Modified Quadrupole Ion Trap Mass Spectrometer at the FELIX Free Electron Laser Laboratory. *Rev Sci Instrum* **2016**, *87*, 103108/1.

[36] 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.

[37] 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.

[38] 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.

[39] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semi-Empirical 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.

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

[41] Berendsen, H. J.; Postma, J. V.; van Gunsteren, W. F.; DiNola, A. R. H. J.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, *81*, 3684.

[42] Stewart, J. J. P. *MOPAC 16*; 2016.

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

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

[45] 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**, *126*, 8712.

[46] 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.

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

[48] Besler, B. H.; Merz, K. M.; Kollman, P. Atomic Charges Derived from Semiempirical Methods. *J Comput Chem* **1990**, *11*, 431.

[49] 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.

[50] Ieritano, C.; Hopkins, W. S. Assessing Collision Cross Section Calculations Using MobCal-MPI with a Variety of Commonly Used Computational Methods. *Mater Today Commun* **2021**, *27*, 102226.

[51] Furche, F.; Ahlrichs, R. Adiabatic Time-Dependent Density Functional Methods for Excited State Properties. *J Chem Phys* **2002**, *117*, 7433.

[52] Wigner, E. On the Quantum Correction for Thermodynamic Equilibrium. *Phys Rev* **1932**, *40*, 749.

[53] Bonacic-Koutecky, V.; Mitric, R. Theoretical Exploration of Ultrafast Dynamics in Atomic Clusters: Analysis and Control. *Chem Rev* **2005**, *105*, 11.

[54) Barbatti, M.; Ruckenbauer, M.; Plasser, F.; Pittner, J.; Granucci, G.; Persico, M.; Lischka, H. Newton-X: A Surface-Hopping Program for Nonadiabatic Molecular Dynamics. Wiley Interdiscip. *Rev Comput Mol Sci* **2014**, *4*, 26.

[55 Bleiholder, C.; Osburn, 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.

[56] Novak, J.; Lemr, K.; Schug, K. A.; Havliček, V. CycloBranch: De Novo Sequencing of Nonribosomal Peptides from Accurate Product Ion Mass Spectra. *J Am Soc Mass Spectrom* **2015**, *26*, 1780.

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

[58] Vrkic, A. K.; O'Hair, R. A. J.; Foote, S.; Reid, G. E. Fragmentation Reactions of All 64 Protonated Trimer Oligodeoxynucleotides and 16 Mixed Base Tetramer Oligodeoxynucleotides via Tandem Mass

Spectrometry in an Ion Trap. *Int. J. Mass Spectrom.* **2000**, *194* (2–3), 145–164. https://doi.org/10.1016/S1387-3806(99)00150-5.

[59] Wan, J.; Brož, B.; Liu, Y.; Huang, S. R.; Marek, A.; Tureček, F. Resolution of Identity in Gas-Phase Dissociations of Mono- and Diprotonated DNA Trinucleotide Codons by 15N-Labeling and Computational Structure Analysis. *J. Am. Soc. Mass Spectrom.* **2022**. https://doi.org/10.1021/jasms.2c00194.

[60] Liu, Y.; Korn, J. A.; Dang, A.; Turecek, F. Hydrogen-Rich Cation Radicals of DNA Dinucleotides. Generation and Structure Elucidation by UV-Vis Action Spectroscopy. *J Phys Chem B* **2018**, *122*, 9665.

Chapter 6

Photochemical and Collision-Induced Crosslinking of Asp, Glu, Asn, and Gln Residues in Peptide-Nitrile Imine Conjugate Ions in the Gas Phase

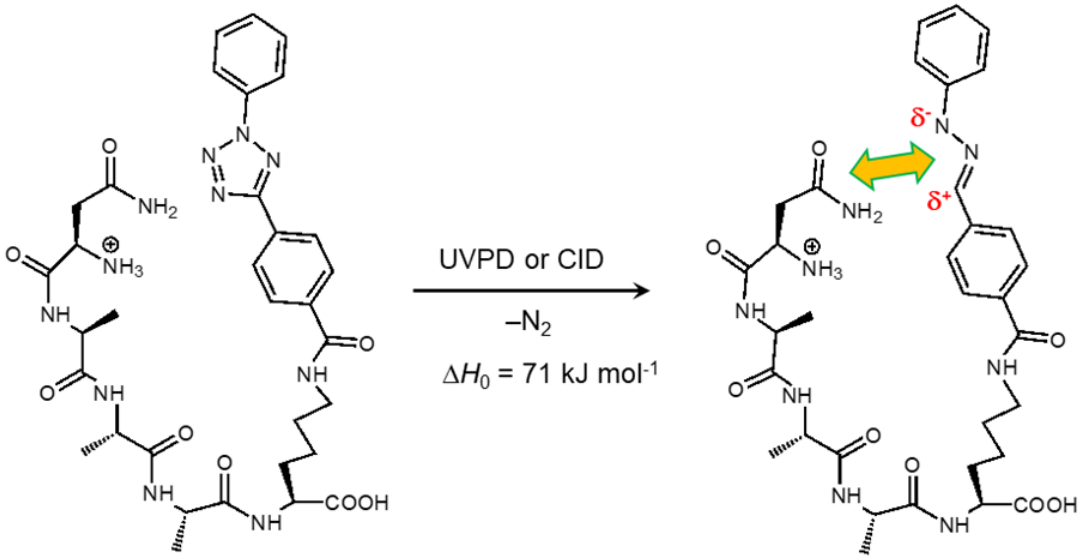
Reproduced in part with permission from Mikuláš Vlk, Jiahao Wan, Marianna Nytka, Tuan Ngoc Kim Vu, Karel Lemr and František Tureček. Photochemical and Collision-Induced Crosslinking of Asp, Glu, Asn, and Gln Residues in Peptide-Nitrile Imine Conjugate Ions in the Gas Phase. *Accepted by J. Am. Soc. Mass Spectrom.*

Abstract *Peptide conjugates furnished with a 2,5-diaryltetrazolecarbonyl tag at the C-terminal lysine, that we call peptide-tet-K, were found to undergo efficient crosslinking of Asp, Glu, Asn, and Gln residues to transient nitrile-imine intermediates produced by photodissociation and collision-induced dissociation (CID) of the tetrazole ring in gas-phase ions. UV photodissociation (UVPD) at 213 nm achieved crosslinking conversion yields of 37 and 61% for DAAAK-tet-K and EAAAK-tet-K, respectively. The yields for NAAAK-tet-K and QAAAK-tet-K were 29 and 57%, respectively. Even higher crosslink yields were found for CID-MS^3 of stable denitrogenated ions, (peptide-tet-K–N_2 + H)$^+$, that were in the 69-83% range. Different types of crosslinks were distinguished by CID-MS^n that showed a distinct series of backbone fragment ions, loss of N-terminal groups, and loss of phenylhydrazine from the modified nitrile imines. The Asp and Glu side-chain carboxyl groups were major participants in crosslinking that resulted in proton and oxygen transfer to the nitrile imine group. Other types of crosslinking involved Asn and Gln $CONH_2$ groups and backbone amides. Cyclic ion mobility mass spectrometry was used to separate NAAAK-tet-K and QAAAK-tet-K ion conformers and products of their collision-induced denitrogenation. Linear nitrile-imine and crosslinked ion structures were identified by comparing the experimental collision cross sections (CCS_{exp}) with those for structures obtained by combined Born-Oppenheimer molecular dynamics and density functional theory (DFT) calculations. The formation of crosslinks was found to be energetically favorable and involved proton-facilitated nucleophilic attack at the nitrile imine carbon atom.*

6.1 Introduction

Whereas nitrile imines represent well-known and used reactive intermediates in organic synthesis of pyrazolines and other heterocyclic compounds,[1,2] their use for covalent crosslinking and footprinting of biomolecules has been much less explored. Lin and coworkers have reported 2-aryl-5-carboxytetrazole-lysine analogues that have been found to crosslink to proximate nucleophile groups such as the Glu carboxyl in tagged proteins.[3,4] Crosslinking was initiated by the formation of nitrile-imine intermediates by photodissociative cleavage of the tetrazole ring, as shown for peptide-tetrazole conjugates (**Scheme 6.1**). More recently, Zhang, Li, and their coworkers have explored the reaction of tetrazoles with primary amines that, when accompanied by spurious oxidation in solution, led to cyclized triazole products. These authors have suggested that the new reaction involved nitrile-imine intermediates.[5] We have used photochemically and thermally generated nitrile-imine conjugates to achieve efficient crosslinking to peptides and dinucleotides in gas-phase ions.[6-8] The gas-phase reactions were driven by the substantial internal excitation in nitrile imines that was acquired by photon absorption. For example, loss of N_2 from tetrazole conjugates with peptides was calculated to have a low threshold energy, as shown in **Scheme 6.1** for a NAAA-tet-K ion, leaving the nitrile-imine intermediate internally excited and promoting its intramolecular reactions resulting in covalent bond formation and crosslinking. An important factor favoring crosslinking was the slow dissipation of internal energy under the low-pressure regime,[9,10] as applied in the ion trap.[11] The novel reactions occurred efficiently even in peptides that did not contain standard nucleophilic groups.[6,7] For example, the protonated conjugate of peptide GAAAK with the 2,5-diaryltetrazolecarbonyl substituent at the lysine side-chain amine, $(GAAA\text{-}tet\text{-}K + H)^+$, was converted to a transient nitrile imine in gas-phase ions that underwent crosslinking to the G-A amide group.[6] The reaction was found to be catalyzed by intramolecular proton transfer, based on experiments with gas-phase ions charged by sodium ion adduction that showed no crosslinking.[6] In an effort to further explore and potentially exploit these novel ion crosslinking reactions, we now investigated UV photodissociation at 213 nm (UVPD) and CID of conjugate ions containing N-terminal Asp, Glu, Asn, and Gln residues in $(DAAA\text{-}tet\text{-}K + H)^+$, $(EAAA\text{-}tet\text{-}K + H)^+$, their dimethyl esters $(D(OCH_3)AAA\text{-}tet\text{-}K(OCH_3) + H)^+$, $(E(OCH_3)AAAtet\text{-}K(OCH_3) + H)^+$, $(NAAA\text{-}tet\text{-}K + H)^+$, and $(QAAA\text{-}tet\text{-}K + H)^+$. Our goal was to generate nitrile imines by UVPD and CID and use CID-MS^n to distinguish linear and

cyclized reaction products. To characterize selected structures we employed high-resolution ion mobility mass spectrometry and theoretical calculations, based on Born-Oppenheimer molecular dynamics and density functional theory, to obtain collision cross sections (CCS). We wish to show that the carboxyl groups in Asp and Glu participate in efficient crosslinking to nitrile imines. Unexpectedly, crosslinking was also observed for the carboxamide groups in Asn and Gln.



Scheme 6.1: Formation of a nitrile imine intermediate as exemplified by the (NAAA-tet-K + H)$^+$ ion. The reaction enthalpy is from M06-2X/def2qzvpp calculations.

6.2 Experimental Section

6.2.1 Materials and Methods

Peptide conjugates were synthesized on solid-support (Wang resin) by standard coupling procedures, starting with lysine side-chain derivatization with 4-(2-phenyl-2H-tetrazol-5-yl)benzoic acid, as described previously in Chapter 5.[6] The peptide conjugates were characterized by high resolution mass spectrometry as ions obtained by protonation in electrospray. Accurate mass measurements are summarized in **Tables 6.1-6.4**. Photodissociation at 250 nm (UVPD) and collision-induced dissociation of selected photoproducts from (DAAA-tet-K) was performed on a Bruker amaZon 3D ion trap mass spectrometer as described previously.[12] The typical laser pulse energies at 250 nm were 1.8-2.0 mJ per pulse. UVPD spectra of all ions were obtained at 213 nm

on a Orbitrap Ascend Tribrid instrument (ThermoFisher, San Jose, CA) with scanning the ion trap, or in the high-resolution mode in the Orbitrap at 100 000 resolving power. Ion mobility measurements were carried out on a SELECT SERIES Cyclic Ion Mobility Spectrometer (c-IMS) (Waters Corp., Wilmslow, U.K.)[13] with direct infusion into a normal flow electrospray ion source at a flow rate of 5μL/min. Each sample and calibrant were measured six times in a positive mode. Typically, ion mobility separation over several cycles (*n*) was used at the total ion path of *n* ✖ 98 cm. Details of all measurement parameters and calibration for collision cross section (CCS) determination have been reported previously.[6]

Table 6.1 Accurate Mass Measurements for (DAAA-*tet*-K+H)$^+$ Ions in UVPD-CID-MS^3.

m/z	ion formula	neutral loss	error (mmu)	ion assignment
695.3122	$C_{33}H_{43}N_8O_9$	N_2	-2.6	M+H-N_2
678.2858	$C_{33}H_{40}N_7O_9$	NH_3	-2.4	M+H-N_2-NH_3
677.3018	$C_{33}H_{41}N_8O_8$	H_2O	-2.4	M+H-N_2-H_2O
660.2755	$C_{33}H_{38}N_7O_8$	H_5NO	-2.2	M+H-N_2-H_2O-NH_3
624.2755	$C_{30}H_{38}N_7O_8$	C_3H_5NO	-2.2	M+H-N_2-Ala
608.2807	$C_{30}H_{38}N_7O_7$	$C_3H_5NO_2$	-2.1	M+H-N_2-Asp(87)
587.2440	$C_{27}H_{35}N_6O_9$	$C_6H_8N_2O_2$	-2.0	M+H-N_2-phenylhydrazine
580.2859	$C_{29}H_{38}N_7O_6$	$C_4H_5N_3O_3$	-1.9	M+H-N_2-Asp
553.2386	$C_{27}H_{33}N_6O_7$	$C_6H_{10}N_2O_2$	-1.9	M+H-N_2-Ala-Ala
527.2595	$C_{26}H_{35}N_6O_6$	$C_7H_8N_2O_3$	-1.7	M+H-N_2-Asp(87)-C-Ala
509.2491	$C_{26}H_{33}N_6O_5$	$C_7H_{10}N_2O_4$	-1.6	M+H-N_2-Asp-Ala
482.2019	$C_{24}H_{28}N_5O_6$	$C_9H_{15}N_3O_3$	-1.6	M+H-N_2-Ala-Ala-Ala
456.2226	$C_{23}H_{30}N_5O_5$	$C_{10}H_{13}N_3O_4$	-1.6	M+H-N_2-Asp(87)-C-Ala-Ala

m/z	ion formula	neutral loss	error (mmu)	ion assignment
438.2121	$C_{23}H_{28}N_5O_4$	$C_{10}H_{15}N_3O_5$	-1.4	M+H-N_2-Asp-Ala-Ala
416.1915	$C_{20}H_{26}N_5O_5$	$C_{13}H_{17}N_3O_4$	-1.4	M+H-N_2-Asp-Ala-93
385.1858	$C_{20}H_{25}N_4O_4$	$C_{13}H_{18}N_4O_5$	-1.3	M+H-N_2-Asp(87)-C-Ala-Ala-Ala
367.1753	$C_{20}H_{23}N_4O_3$	$C_{13}H_{20}N_4O_6$	-1.2	M+H-N_2-Asp-Ala-Ala-Ala

Table 6.2 Accurate Mass Measurements for (EAAA-*tet*-K+H)$^+$ Ions in UVPD-CID-MS^3.

m/z	ion formula	error (mmu)	neutral loss	ion assignment
709.3276	$C_{34}H_{45}N_8O_9$	-2.8		
691.3169	$C_{34}H_{43}N_8O_8$	-2.9	H_2O	M+H-N_2-H_2O
638.2908	$C_{31}H_{40}N_7O_8$	-2.5	$C_3H_5N_3O$	M+H-N_2-Ala
618.2859	$C_{28}H_{40}N_7O_9$	-2.3	$C_6H_5N_3$	M+H-N_2-91
617.2780	$C_{28}H_{39}N_7O_9$	-2.4	$C_6H_6N_3$	M+H-N_2-92
608.2804	$C_{30}H_{38}N_7O_7$	-2.3	$C_4H_7N_3O_2$	M+H-N_2-Glu(101)
601.2593	$C_{28}H_{37}N_6O_9$	-2.3	$C_6H_8N_4$	M+H-N_2-phenylhydrazine
580.2857	$C_{29}H_{38}N_3O_3$	-2.2	$C_5H_7N_5O_6$	M+H-N_2-Glu
567.2540	$C_{28}H_{35}N_6O_7$	-2.2	$C_6H_{10}N_4O_2$	M+H-N_2-Ala-Ala
537.2563	$C_{26}H_{33}N_8O_5$	-0.5	$C_8H_{12}N_2O_4$	M+H-Glu-Ala
527.2593	$C_{26}H_{35}N_6O_6$	-2.0	$C_8H_{10}N_4O_3$	M+H-N_2-Glu(101)-C-Ala
509.2489	$C_{26}H_{33}N_6O_5$	-1.8	$C_8H_{12}N_4O_4$	M+H-N_2-Glu-Ala
496.2172	$C_{25}H_{30}N_5O_6$	-1.9	$C_9H_{15}N_5O_3$	M+H-N_2-Ala-Ala-Ala

466.2192	$C_{23}H_{28}N_7O_4$	-0.5	$C_{11}H_{17}N_3O_5$	M+H-Glu-Ala-Ala
456.2223	$C_{23}H_{30}N_5O_5$	-1.8	$C_{11}H_{15}N_5O_4$	M+H-N_2-Glu(101)-C-Ala-Ala
438.2119	$C_{23}H_{28}N_5O_4$	-1.7	$C_{11}H_{17}N_5O_5$	M+H-N_2-Glu-Ala-Ala
395.1822	$C_{20}H_{23}N_6O_3$	-0.4	$C_{14}H_{22}N_4O_6$	M+H-Glu-Ala-Ala-Ala
385.1856	$C_{20}H_{25}N_4O_4$	-1.5	$C_{14}H_{20}N_6O_5$	M+H-N_2-Glu(101)-C-Ala-Ala-Ala
367.1750	$C_{20}H_{23}N_4O_3$	-1.4	$C_{14}H_{22}N_6O_6$	M+H-N_2-Glu-Ala-Ala-Ala

Table 6.3 Accurate Mass Measurements for (NAAA-*tet*-K+H)$^+$ Ions in UVPD-CID-MS^3.

m/z	ion formula	error (mmu)	neutral loss	ion assignment
694.3292	$C_{33}H_{44}N_9O_8$	-1.5		M+H–N_2
677.3027	$C_{33}H_{41}N_8O_8$	-1.5	NH_3	
659.2922	$C_{33}H_{39}N_8O_7$	-1.4	NH_3 + H_2O	
623.2922	$C_{30}H_{39}N_8O_7$	-1.4	C_3H_5NO	M+H–N_2–[Ala]
608.2817	$C_{30}H_{38}N_7O_7$	-1.0	$C_3H_6N_2O$	M+H–N_2–Asn (86)
601.2719	$C_{27}H_{37}N_8O_8$	-1.0	$C_6H_5NH_2$	
586.2609	$C_{27}H_{36}N_7O_8$	-1.1	$C_6H_8N_2$	M+H–N_2–phenylhydrazine
580.2869	$C_{29}H_{38}N_7O_6$	-1.1	$C_4H_6N_2O_2$	M+H–N_2–Asn, [$\boldsymbol{y_4}$ + 2H]$^+$
557.2821	$C_{26}H_{37}N_8O_6$	-1.1	$C_7H_7NO_2$	
552.2559	$C_{27}H_{34}N_7O_6$	-0.6	$C_6H_{10}N_2O_2$	M+H–N_2–[AlaAla]
509.2500	$C_{23}H_{33}N_6O_5$	-0.7	$C_{10}H_{11}N_3O_3$	M+H–N_2–AsnAla, [$\boldsymbol{y_3}$ + 2H]$^+$
481.2186	$C_{24}H_{29}N_6O_5$	-0.8	$C_9H_{15}N_3O_3$	M+H–N_2–[AlaAlaAla]

459.1979	$C_{21}H_{27}N_6O_6$	-0.8	$C_{12}H_{17}N_3O_2$	
438.2129	$C_{23}H_{28}N_5O_4$	-0.7	$C_{10}H_{16}N_4O_4$	M+H–N_2–AsnAlaAla, $[\boldsymbol{y_2} + 2H]^+$
367.1757	$C_{20}H_{23}N_4O_3$	-0.7	$C_{13}H_{21}N_5O_5$	$[\boldsymbol{y_1} + 2H]^+$

Table 6.4 Accurate Mass Measurements for (QAAA-*tet*-K+H)$^+$ Ions in UVPD-CID-MS^3.

m/z	ion formula	error (mmu)	neutral loss	ion assignment
708.3439	$C_{34}H_{46}N_9O_8$	-2.5		M+H–N_2
691.3177	$C_{34}H_{43}N_8O_8$	-2.1	NH_3	
673.3073	$C_{34}H_{41}N_8O_7$	-2.0	NH_3+ H_2O	
616.2946	$C_{28}H_{40}N_8O_8$	-1.8	C_6H_6N	
608.2812	$C_{30}H_{38}N_7O_7$	-1.5	$C_4H_8N_2O$	M+H–N_2–Asn (100)
600.2762	$C_{28}H_{38}N_7O_8$	-1.4	$C_6H_8N_2$	M+H–N_2–phenylhydrazine
598.2606	$C_{28}H_{36}N_7O_8$	-1.4	$C_6H_{10}N_2$	M+H–N_2–NH_3–C_6H_7N
580.2864	$C_{29}H_{38}N_7O_6$	-1.4	$C_5H_8N_2O_2$	M+H–N_2–Gln, $[\boldsymbol{y_4} + 2H]^+$
566.2708	$C_{28}H_{36}N_7O_6$	-1.4	$C_6H_{10}N_2O_2$	M+H–N_2–[AlaAla]
527.2602	$C_{26}H_{35}N_6O_6$	-1.1	$C_8H_{11}N_3O_2$	$\boldsymbol{v_3}$
509.2498	$C_{26}H_{33}N_6O_5$	-0.9	$C_8H_{13}N_3O_3$	M+H–N_2–GlnAla, $[\boldsymbol{y_3} + 2H]^+$
495.2341	$C_{25}H_{31}N_6O_5$	-0.9	$C_9H_{15}N_3O_3$	M+H–N_2–[AlaAlaAla]
473.2133	$C_{22}H_{29}N_6O_6$	-1.0	$C_{12}H_{17}N_3O_2$	566–C_6H_7N
456.1868	$C_{22}H_{26}N_5O_6$	-1.0	$C_{12}H_{20}N_3O_2$	$\boldsymbol{v_2}$
438.2126	$C_{23}H_{28}N_5O_4$	-1.0	$C_{11}H_{18}N_4O_4$	M+H–N_2–GlnAlaAla, $[\boldsymbol{y_n} + 2H]^+$

385.1497	$C_{19}H_{21}N_4O_5$	-0.9	$C_{15}H_{25}N_5O_3$	$\boldsymbol{v_1}$
367.1755	$C_{20}H_{23}N_4O_3$	-1.0	$C_{14}H_{23}N_5O_5$	$[\boldsymbol{y_1} + 2H]^+$

6.2.2 Calculations

We used the previously developed protocol to obtain ion structures.[8] Briefly, this consisted of Born-Oppenheimer molecular dynamics (BOMD) calculations of 20 ps trajectories with 1 fs steps, using the Berendsen thermostat[14] at 510, 610, and 810 K, that were performed with PM6-D3H4[15] which complements the semiempirical Hamiltonian with dispersion and hydrogen-bonding interactions. These calculations were carried out using the high-end Cuby4 platform[16] and MOPAC.[17] Two hundred snapshots were selected at regular 100 fs intervals from each BOMD trajectory, and the structures were fully gradient-optimized with PM6-D3H4. Several low energy structures were selected and further optimized with B3LYP[18] with the 6-31+G(d,p) basis set, including empirical dispersion corrections (GD3-BJ).[19,20] All structures were established as local energy minima by harmonic frequency analysis. Several low-energy structures were selected and reoptimized with M06-2X[21] 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[22] basis set (4200-4400 basis functions). 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).[23,24] 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 (MobCalMPI)[25,26] using the MK charge densities.

6.3 Results and Discussion

6.3.1 DAAA-tet-K Spectra and Crosslink Identification

CID and UVPD-MS^2 at 213 nm of the DAAA-tet-K ion ((DAAAK-tet-K + H)$^+$, m/z 723) resulted in loss of N_2 from the tetrazole (m/z 695) that was accompanied by standard backbone

cleavages forming a series of $[\boldsymbol{y_1}+2H]^+$-$[\boldsymbol{y_4}+2H]^+$ ions at m/z 395, 466, 537, and 608 (**Figure 6.1a,b**). For the standard fragment ions, we use the modified peptide ion nomenclature with explicit numbering of hydrogen atoms and protons.[27] The loss of N_2 was more efficient on UVPD (**Figure 6.1b**), indicating the prevalently photochemical nature of this dissociation. In addition, even at the UVPD-MS^2 level the loss of N_2 was followed by the elimination of a HN=CH-CH_2COOH neutral fragment from the Asp residue. The resulting m/z 608 ion showed a doublet consisting of the $[\boldsymbol{y_4}+2H]^+$ ion as a minor component, and the major fragment ion due to the HN=CH-CH_2COOH loss, as established from the high-resolution CID-MS^2 spectrum (**Figure 6.1a**, inset). All fragment ion assignments were corroborated by accurate mass measurements and H/D exchange that provided the expected retention of deuterium in the fragment ions originating from the fully exchanged D_{10} precursor ion (**Figure 6.2a-c**). The loss HN=CH-CH_2COOH, which is uncommon in dissociations of straight-chain peptide ions, was analogous to aldimine losses from peptide-tetrazole conjugates that were N-terminated with other amino acid residues (Gly, Ala), as reported earlier.[6] This dissociation indicated crosslinking to the nitrile imine at a position other than the Asp side-chain and amine group. Other uncommon fragment ions were found at m/z 527, 456, 385, 329, and 322 that were more abundant in the CID-MS^3 spectrum of the denitrogenated ion (m/z 695) that was generated by UVPD (**Figure 6.1c**). Very similar dissociations were found for CID-MS^3 of the $(DAAAK\text{-}tet\text{-}K - N_2 + H)^+$ ion that was generated by CID (**Figure 6.3**). The **Figure 6.1c** and **Figure 6.3** spectra showed two prominent series of fragment ions that were representative of crosslinked ion structures. Loss of internal Ala residues gave rise to ions at m/z 624, 553, and 482 that indicated crosslinking by the N-terminal Asp residue to leave the Ala residues free for elimination. We denote these internal neutral fragments with square brackets, [Ala], to distinguish them from regular backbone fragments. The loss of internal residues is a hallmark of dissociations of cyclic peptide ions,[28,29] which in this case pointed to structures formed by crosslinking to the nitrile imine group. The other fragment ion series indicating crosslinking consisted of C-terminal ions carrying the Lys-attached $C_{14}H_{11}N_2O_2$ group that we denoted as $\boldsymbol{\upsilon_1}$-$\boldsymbol{\upsilon_3}$ (m/z 385, 456, 527) to distinguish them from standard $[\boldsymbol{y_n} + 2H]^+$ ions. Although, according to their formulas, the $\boldsymbol{\upsilon_n}$ ions formally corresponded to water adducts to $[\boldsymbol{y_n} + 2H]^+$ ions, we found them to be covalently bound. This was corroborated by the CID-MS^4 spectrum of the $\boldsymbol{\upsilon_3}$ ion (m/z 527) that showed backbone fragment ions by loss of [Ala] and [AlaAla], but only a very weak peak for loss of water that would have been expected as a major

product from a noncovalent water adduct (**Figure 6.4**). Furthermore, the CID-MS^2 spectrum of $(DAAAK\text{-}tet\text{-}K + H)^+$ showed no $[\boldsymbol{y_n} + 2H + H_2O]^+$ fragment ions (**Figure 6.1a**), confirming that water addition to $[\boldsymbol{y_n} + 2H]^+$ ions did not occur as an ion-molecule reaction in the ion trap. Hence we conclude that the $\boldsymbol{\upsilon_n}$ ions were covalently bound. The formation and nature of the $\boldsymbol{\upsilon_n}$ ions was further probed by obtaining a UVPD-CID-MS^3 spectrum of the DAAAK-tet-K dimethyl ester ($(D(OCH_3)AAA\text{-}tet\text{-}K(OCH_3)–N_2 + H)^+$, m/z 723) in which the Asp carboxyl was blocked (**Figure 6.5a**). The spectrum showed no n ions, indicating that a free Asp carboxyl was required for this type of crosslinking. The formation of $\boldsymbol{\upsilon_n}$ ions can be viewed as proceeding from crosslinks in which the Asp carboxyl proton was transferred to the negatively charged nitrile imine nitrogen, and the electron-deficient carbon was attacked by the carboxyl oxygen forming a covalent C–O bond (**Scheme 6.2**). The formation of the $\boldsymbol{\upsilon_n}$ fragment ions upon CID can be initiated by proton-promoted cleavage of the Asp CO–O bond which can be assisted by participation of the amide oxygen forming a lactone intermediate. Further amide bond dissociations then led to the C-terminal $\boldsymbol{\upsilon_n}$ ions in which the former Asp carboxyl oxygen and two hydrogen atoms were embedded in the nitrile imine moiety as a hydrazide (**Scheme 6.2**).

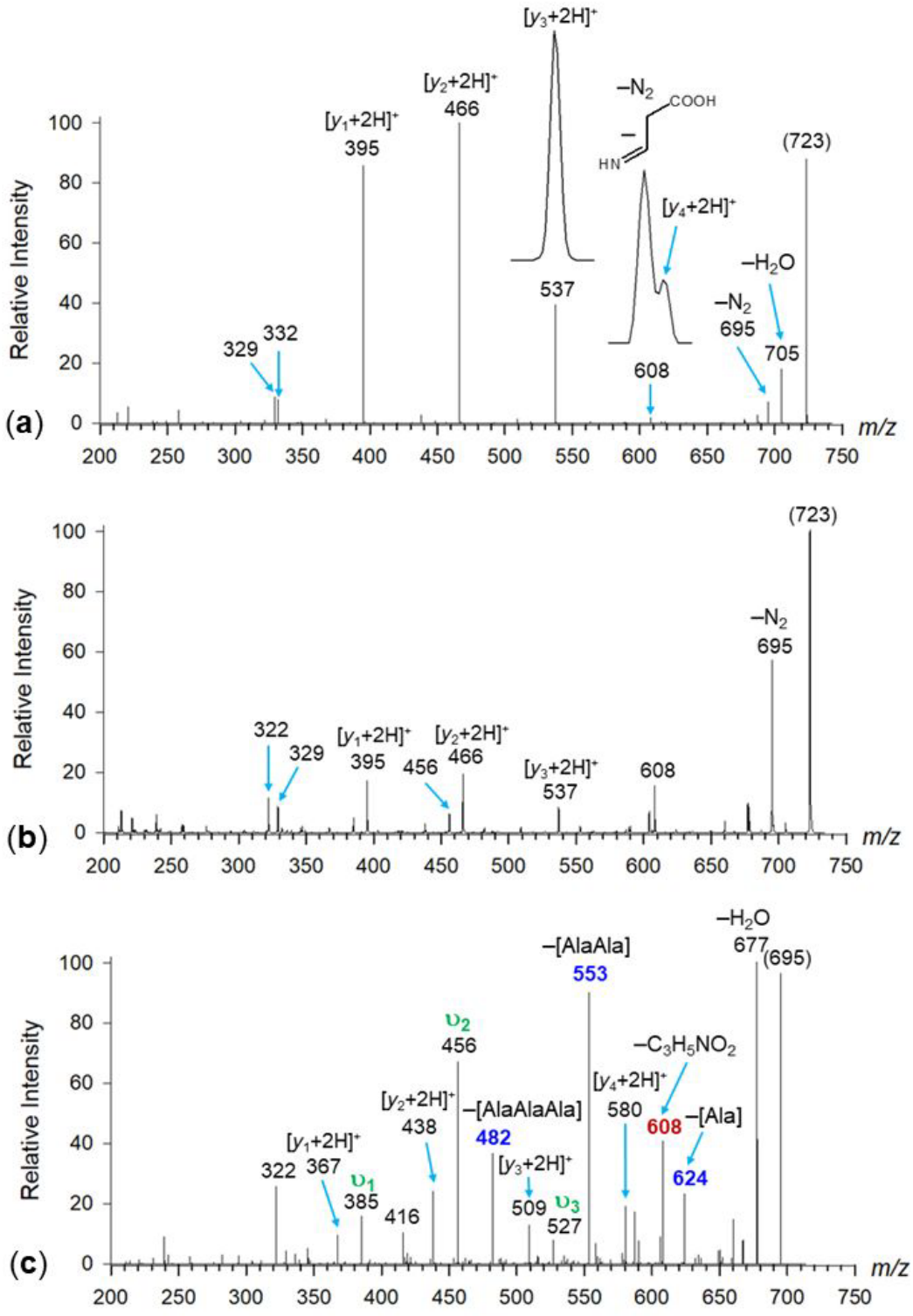


Figure 6.1: (a) CID-MS^2 spectrum of $(DAAA\text{-}tet\text{-}K + H)^+$ (m/z 723); (b) UVPD-MS^2 spectrum of $(DAAA\text{-}tet\text{-}K + H)^+$ (m/z 723) at 213 nm; (c) CID-MS^3 spectrum of $(DAAA\text{-}tet\text{-}K - N_2 + H)^+$ (m/z 695) generated by 213 nm UVPD of $(DAAA\text{-}tet\text{-}K + H)^+$.

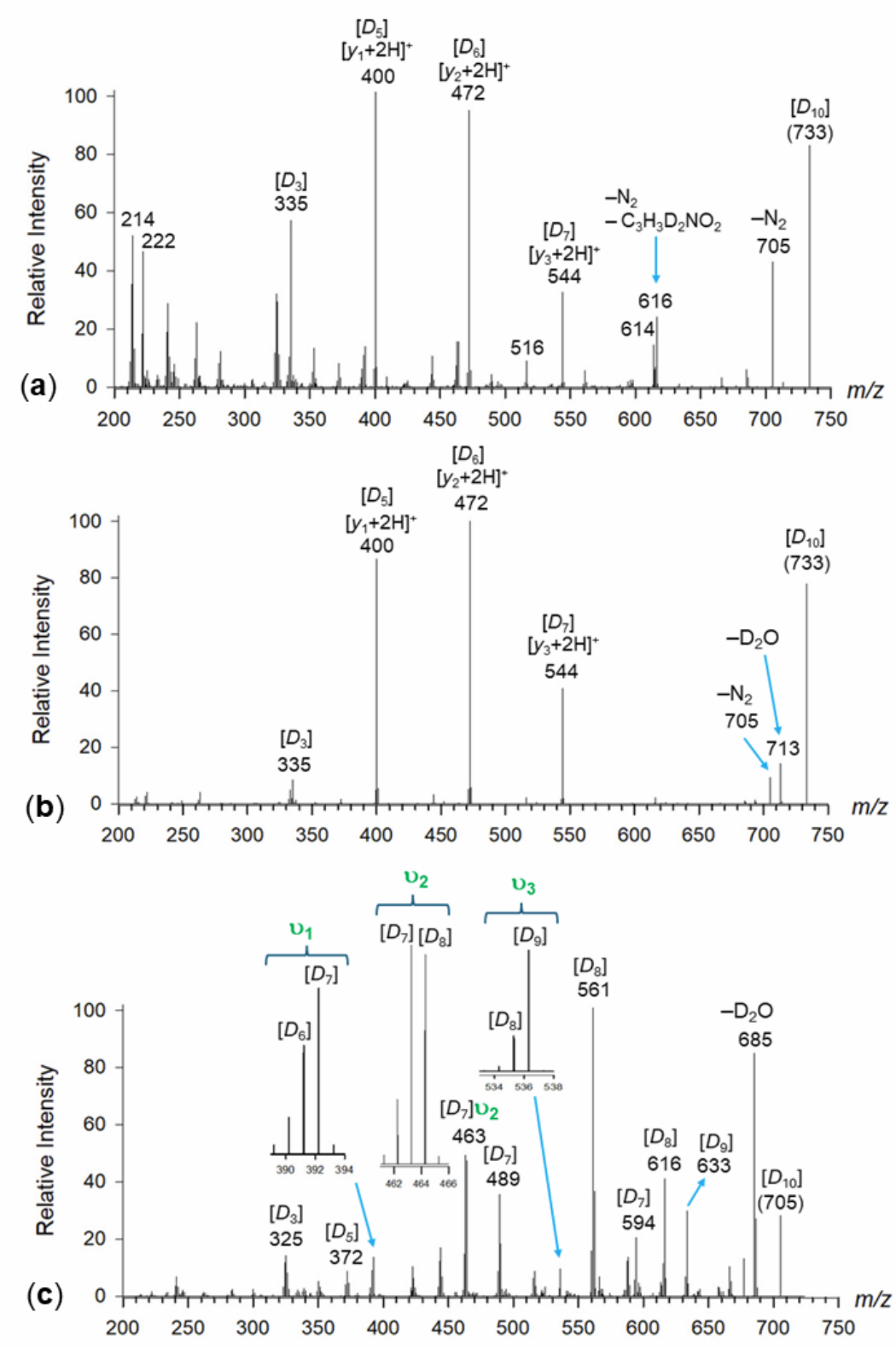


Figure 6.2: (a) UVPD-MS^2 and (b) CID-MS^2 of H/D exchanged ([D$_9$]DAAA-tet-K + D)$^+$ at m/z 733. (c) CID-MS^3 of H/D exchanged ([D$_9$]DAAA-tet-K–N$_2$ + D)$^+$ at m/z 705.

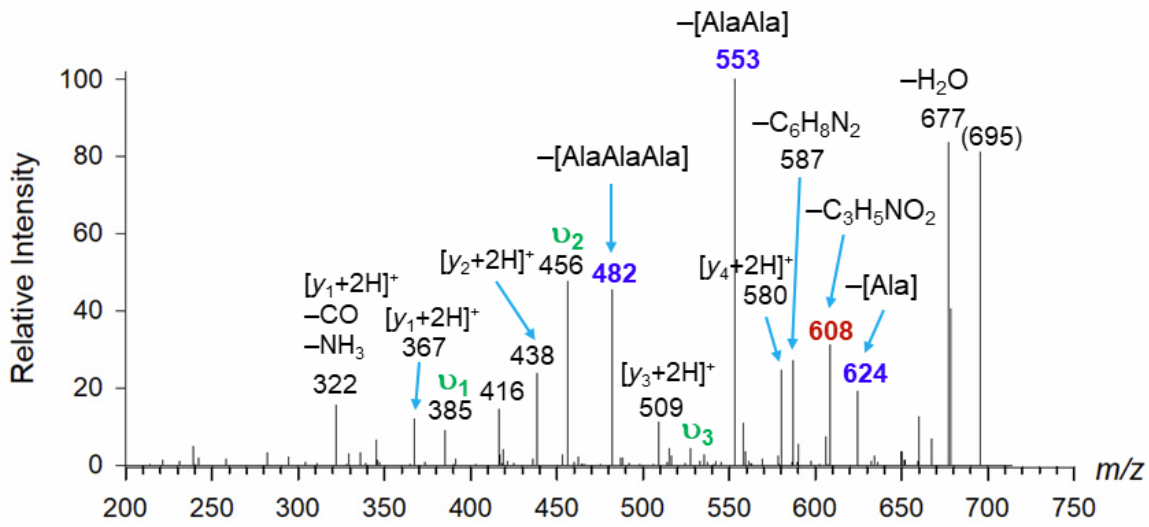


Figure 6.3: CID-MS^3 of (DAAAK-tet-K – N_2 + H)$^+$ (m/z 695) generated by CID of (DAAAK-tet-K + H)$^+$.

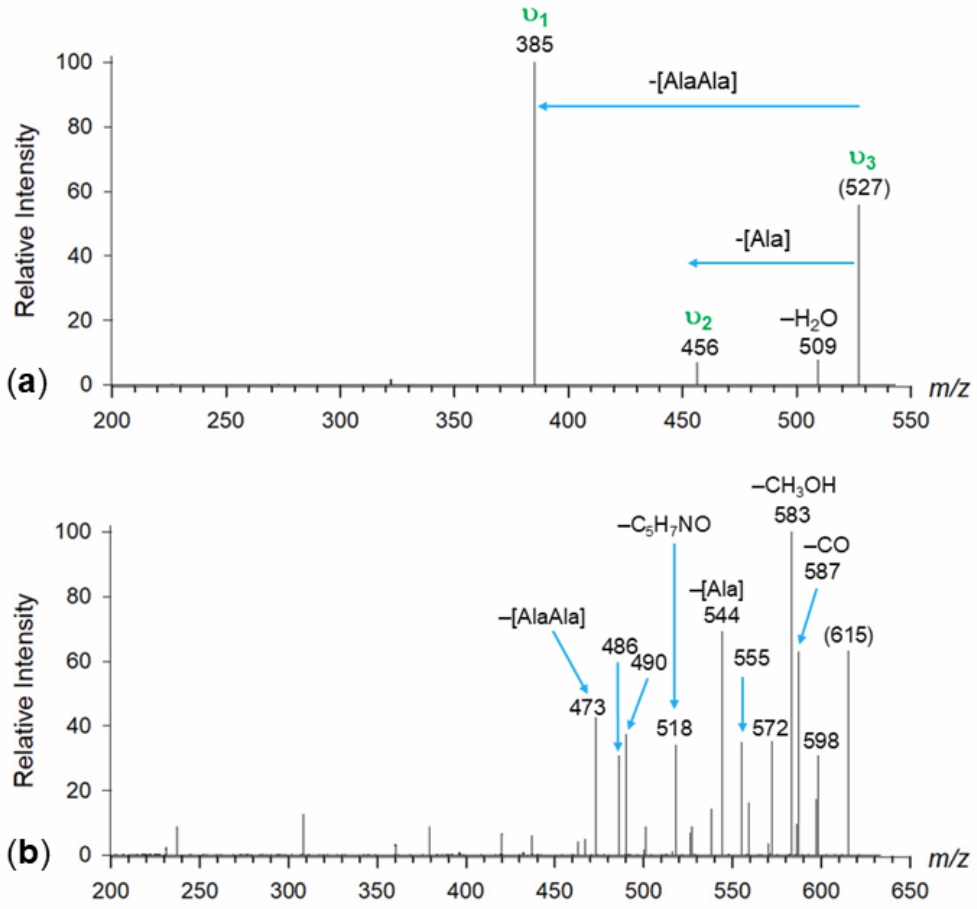


Figure 6.4: CID-MS^4 of (a) the m/z 527 (υ_3 ion) from (DAAAK-tet-K – N_2 + H)$^+$ (b) m/z 615 (loss of phenylhydrazine) from (D(OCH_3)AAA-tet-K(OCH_3) – N_2 + H)$^+$.

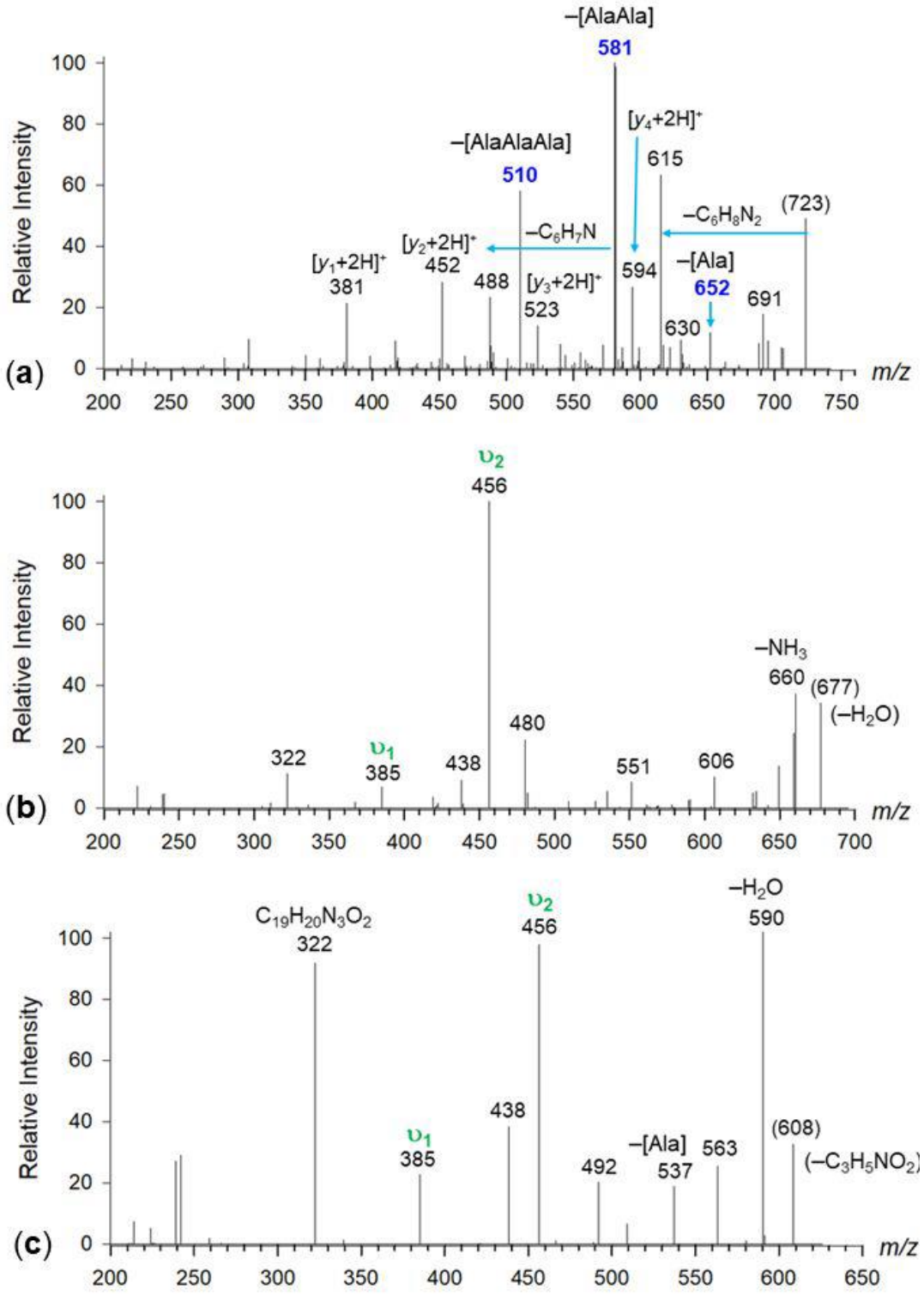
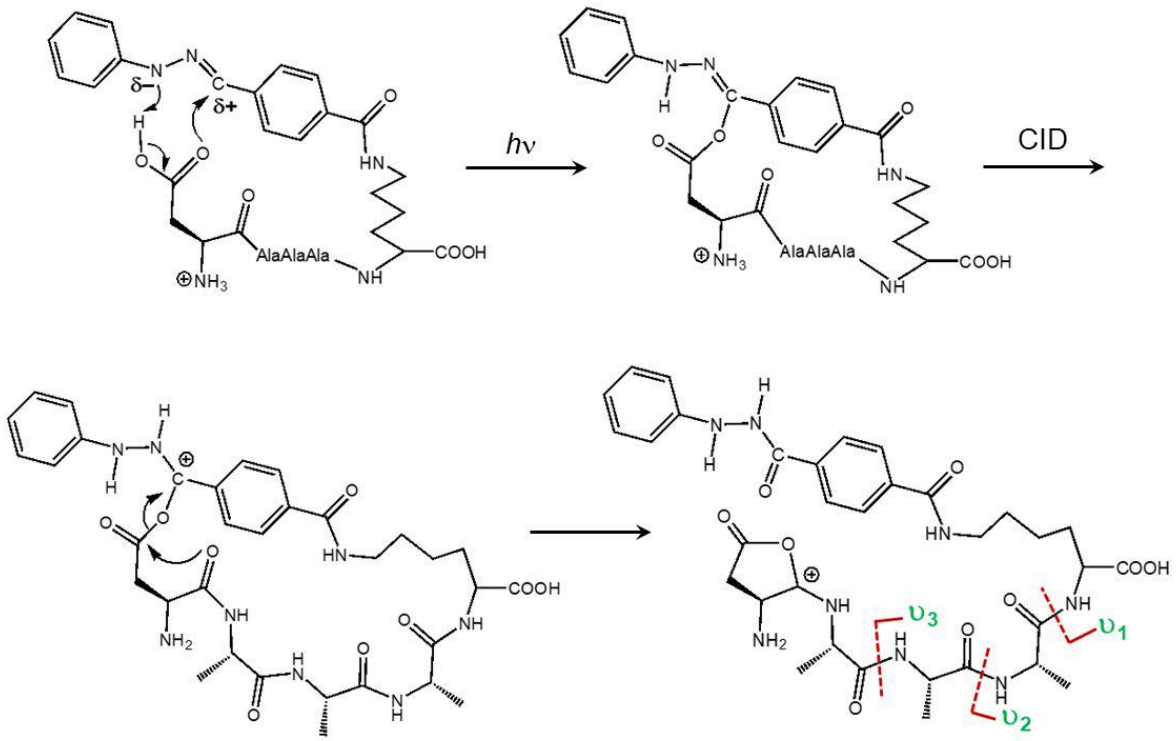


Figure 6.5: (a) UVPD-CID-MS^3 spectrum of $(D(OCH_3)AAA\text{-}tet\text{-}K(OCH_3) - N_2 + H)^+$ (m/z 723); (b) CID-MS^4 spectra (b) m/z 677 and (c) m/z 608 ions from Figurer 6.1c spectrum of $(DAAA\text{-}tet\text{-}K - N_2 + H)^+$.



Scheme 6.2: Proposed Asp Carboxyl Crosslinking Mechanism and Backbone Dissociations in the Crosslinks.

The formation of υ_n ions was also probed by CID-MS^4 of the m/z 677 (loss of water), and m/z 608 (loss of $C_3H_5NO_2$) primary fragment ions from the UVPD-CID-MS^3 spectrum of $(DAAAK\text{-}tet\text{-}K + H)^+$. Both these CID-$MS^4$ spectra (**Figure 6.5b,c**) showed prominent υ_2 ions (m/z 456). Since the m/z 608 ion no longer had the Asp residue, the formation from it of the υ_2 ion indicated that the oxygen atom transferred to the nitrile imine can also originate from other accessible positions, e.g., the amide groups. Consistent with this, while carboxyl esterification stopped the formation of υ_n ions, it did not prevent internal backbone fragmentation, resulting in the loss of $[Ala_n]$ that indicated crosslinked structures. The spectrum of the methyl ester further showed enhanced loss of $C_6H_8N_2$ (most likely phenyl hydrazine), giving the ion at m/z 615 that had only a minor homologue at m/z 587 in the CID-MS^3 spectrum of $(DAAAK\text{-}tet\text{-}K + H)^+$. Further CID-$MS^4$ of the m/z 615 ion (**Figure 6.4b**) showed losses of internal [Ala] (m/z 544) and [AlaAla] (m/z 473), indicating that the ion was crosslinked when formed by loss of phenylhydrazine. Overall, the DAAA-tet-K data showed that UVPD resulted in crosslinking to the nitrile imine, favoring $(DAAAK\text{-}tet\text{-}K\text{–}N_2 + H)^+$ structures dissociating to n ions and methyl ester structures dissociating by loss of phenylhydrazine. The overall crosslink yields in the CID-MS^3 spectra were similar at

83% and 77% for (DAAAK-tet-K–N_2 + H)$^+$ and its methyl diester analogue, respectively (**Table 6.5**).

Table 6.5 Crosslinking Yields

	crosslinking yield (%)[a]				
		CID-MS^3			
sequence	UVPD-MS^2	υ_n	Ala	$C_6H_8N_2$	total
DAAA-*tet*-K	37	35	44	5	83
D(OCH_3)AAA-*tet*-K(OCH_3)	50	2	44	31	77
EAAA-*tet*-K	61	40	34	6	80
E(OCH_3)AAA-*tet*-K(OCH_3)	62	2	40	25	67
NAAA-*tet*-K	29	4	52	16	72
QAAA-*tet*-K	57	23	37	9	69

[a]Sum of identified crosslink fragment ion intensities relative to the sum of all backbone fragment ion intensities. Ions by loss of water, ammonia, and combinations thereof were not included.

6.3.2 EAAA-tet-K Spectra and Crosslink Identification

CID and UVPD of the homologous EAAA-tet-K ion ((EAAAK-tet-K + H)$^+$, m/z 737) gave different results (**Figure 6.6a,b**). CID resulted in regular backbone cleavages giving rise to standard [y_n +2H]$^+$ ions whereas loss of N_2 was relatively less abundant, as were backbone n fragment ions resulting from crosslinked nitrile-imine intermediates (**Figure 6.6a**). In contrast, UVPD gave a prominent series of υ_n ions (m/z 385, 456, 527) in addition to the [y_n +2H]$^+$ ions (**Figure 6.6b**). Products of deep photodissociation containing an oxygenated nitrile imine moiety, such as $C_{19}H_{20}N_3O_2$ (m/z 322) and $C_{14}H_{11}N_2O_2$ (m/z 239) were also formed upon UVPD (**Figure 6.6b**). The differences can be ascribed to different modes of ion excitation. Slow heating on CID, resulting in ion vibrational excitation, chiefly promoted low-energy dissociations, such as peptide backbone cleavage forming [y_n +2H]$^+$ ions without N_2 loss from the tetrazole group. In contrast, absorption of the 213-nm photon (561 kJ mol^{-1}) by the diaryltetrazole chromophore initiated tetrazole ring cleavage from an excited electronic state. Comparing the typical reaction enthalpy

for the tetrazole N_2 loss (50-80 kJ mol^{-1})[6,7] with the 213-nm photon energy, one can expect the nitrile-imine intermediates formed by UVPD to be substantially excited. Further exothermic crosslinking with oxygen atom and proton transfer to the nitrile-imine intermediate can further energize the crosslinks to trigger backbone cleavage. The dissociations originating from crosslinks were further emphasized in the CID-MS^3 spectrum of the $(EAAAK\text{-}tet\text{-}K–N_2 + H)^+$ ion (m/z 709) , that was generated by UVPD (**Figure 6.6c**). The **Figure 6.6c** spectrum showed $[\boldsymbol{y_n} +2H]^+$ ions at m/z 367, 438, 509, and 580 that were presumed to represent structures with linear peptide chains. In addition, ions formed by internal losses of $[Ala_n]$, indicating crosslinked structures, were prominent at m/z 638, 567, and 496. Another ion series pointing to crosslinks by the Glu carboxyl was represented by n ions at m/z 527, 456, and 385. We note that the CID-MS^3 spectrum of the $(EAAAK\text{-}tet\text{-}K– N_2 + H)^+$ ion (m/z 709) that was generated by CID was similar to the Figure 3c spectrum, showing identical fragment ions (**Figure 6.7**).

We also investigated the effect on UVPD and CID of carboxyl methylation in EAAAK-tet-K dimethyl ester, as illustrated by the UVPD-CID-MS^3 spectrum of the denitrogenated ion at m/z 737 (**Figure 6.8**). The spectrum showed very minor formation of $\boldsymbol{\upsilon_n}$ ions that were represented only by a weak $\boldsymbol{\upsilon_2}$ ion at m/z 470. Nevertheless, loss of phenylhydrazine (m/z 629) and internal $[Ala_n]$ residues (m/z 666, 595, and 524) attested to a major fraction of crosslinked structures. Linear peptide structures were indicated by the standard $[\boldsymbol{y_n} +2H]^+$ ions at m/z 381, 452, 523, and 594. These figures were reflected by the crosslinking yields for EAAAK-tet-K and its dimethyl ester (**Table 6.5**). The crosslinking yields upon UVPD were very similar for EAAAK-tet-K and its dimethyl ester at 61% and 62%, respectively (**Table 6.5**). The crosslinking yields upon CID of denitrogenated ions were slightly higher for $(EAAAK\text{-}tet\text{-}K –N_2 + H)^+$ than its methyl ester at 80% and 67%, respectively. The lower yield for the methyl ester was chiefly due to the near absence of the $\boldsymbol{\upsilon_n}$ ions. We presumed that the mechanism of the $\boldsymbol{\upsilon_n}$ ion formation from EAAAK-tet-K was analogous to that shown in **Scheme 6.2** and hinged on the availability of a free Glu carboxyl group.

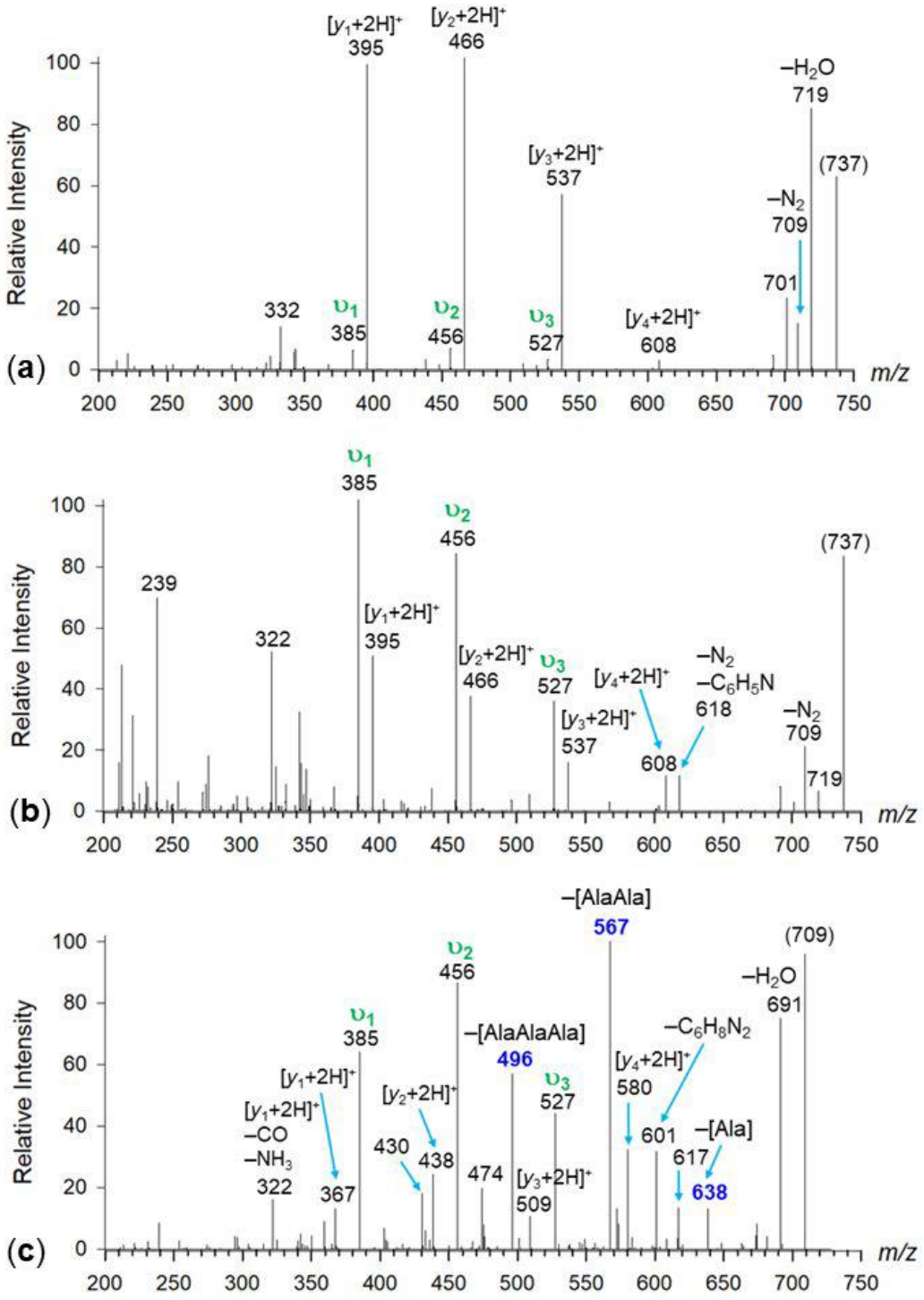


Figure 6.6: (a) CID-MS^2 spectrum of $(EAAA\text{-}tet\text{-}K + H)^+$ (m/z 737); (b) UVPD-MS^2 spectrum of $(EAAA\text{-}tet\text{-}K + H)^+$ (m/z 737) at 213 nm; (c) CID-MS^3 spectrum of $(EAAA\text{-}tet\text{-}K - N_2 + H)^+$ (m/z 709) generated by 213 nm UVPD of $(EAAA\text{-}tet\text{-}K + H)^+$.

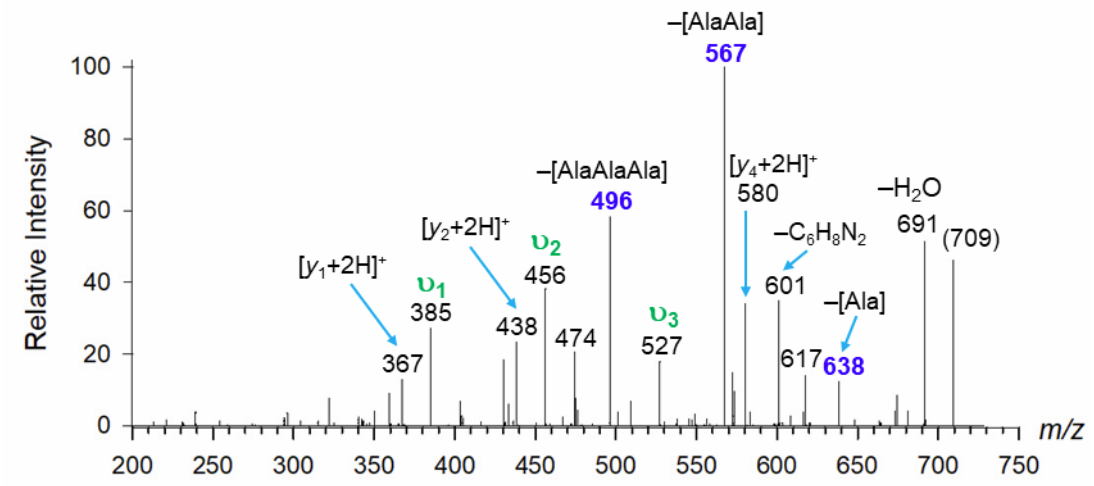


Figure 6.7: CID-MS^3 of (EAAAK-tet-K – N_2 + H)$^+$ (m/z 709) generated by CID of (EAAAK-tet-K + H)$^+$.

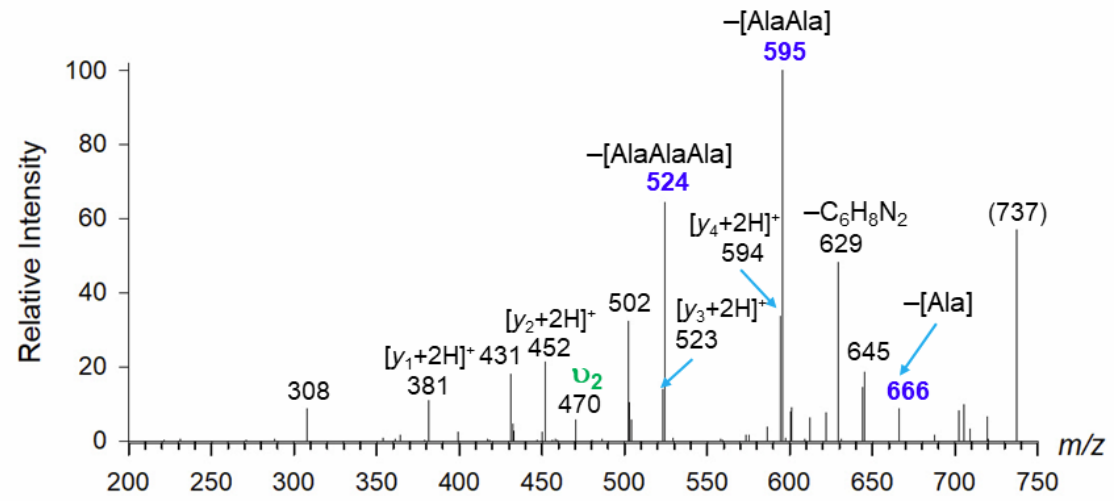


Figure 6.8: UVPD-CID-MS^3 of the m/z 737 ion generated by N_2 loss from dimethyl ester (E(OCH_3)AAAK-tet-K(OCH_3) + H)$^+$.

6.3.3 NAAA-tet-K and QAAA-tet-K Spectra and Crosslink Identification

The effect on crosslinking of the side-chain functional groups was further investigated with NAAA-tet-K and QAAAK-tet-K to provide comparison of the side-chain amide groups with the corresponding carboxyl groups in DAAA-tet-K and EAAA-tet-K and their methyl esters. The UVPD-MS^2 spectra of (NAAA-tet-K + H)$^+$ (m/z 722) and (QAAA-tet-K + H)$^+$ (m/z 736) showed a series of [$\boldsymbol{y_n}$ +2H]$^+$ ions in addition to the N_2 loss peaks at m/z 694 and 708, respectively **(Figure 6.9a,b)**. Crosslinking upon UVPD was indicated by the common m/z 608 ion resulting from the

loss of the respective N-terminal $HN{=}CHCH_2CONH_2$ and $HN{=}CHCH_2CH_2CONH_2$ neutral fragments from the Asn and Gln residues. These dissociations have also been observed for UVPD of the Asn and Gln conjugates at 250 nm which, however, resulted in less backbone dissociation.[6] The unusual dissociation of the C–CO bond in the Asn and Asp residues did not involve radical intermediates, as opposed to analogous backbone cleavages leading to x-type fragment ions in peptide cation radicals,[30,31] and upon peptide ion photofragmentation at 157 nm[32] and 192 nm.[33] Although the mechanism of this dissociation, involving transfer of two hydrogens, was unknown, its occurrence pointed to crosslinking off the N-terminal residue.

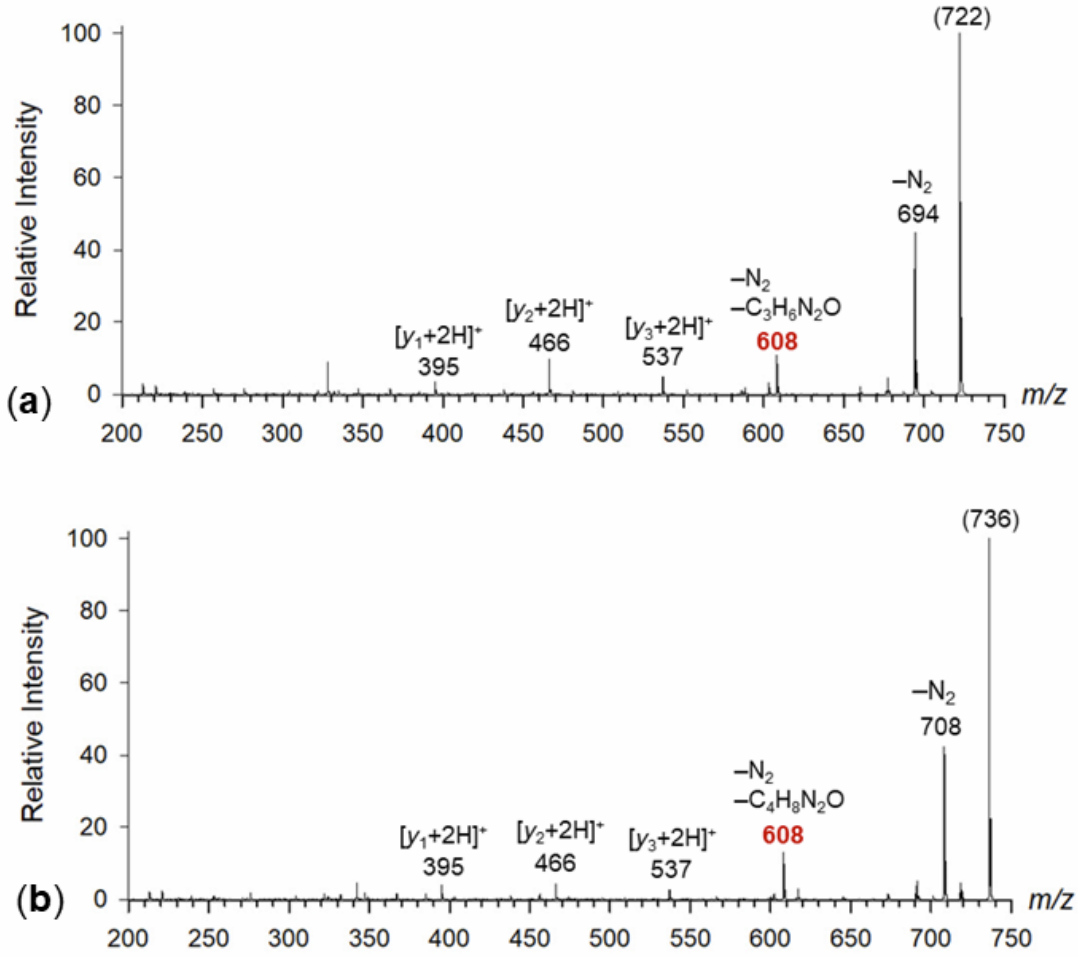


Figure 6.9: UVPD-MS^2 spectra at 213 nm of (a) $(NAAAK\text{-}tet\text{-}K + H)^+$ (m/z 722) and (b) $(QAAAK\text{-}tet\text{-}K + H)^+$ (m/z 736).

Further information was obtained from CID-MS^3 of the denitrogenated ions (**Figure 6.10a,b**). The spectrum of $(NAAA\text{-}tet\text{-}K{-}N_2 + H)^+$ (m/z 694, **Figure 6.10a**) showed a major loss of ammonia

(m/z 677) that was analogous to the loss of water from (DAAA-tet-K–N_2 + H)$^+$ (**Figure 6.1c**). Although ammonia and water could originate from various positions in the ions, the coincidence of their elimination from Asn and Asp, respectively, seemed to indicate they came from the carboxamide and carboxyl group. Note that loss of ammonia was not observed on UVPD-MS^2 of (NAAA-tet-K + H)$^+$ (**Figure 6.9a**), and its prominent presence in the CID-MS^3 spectrum of the denitrogenated ion suggested that it originated from crosslinked structures. The **Figure 6.10a** spectrum showed two series of backbone fragment ions. One was the [$\boldsymbol{y_n}$ +2H]$^+$ ions at m/z 367, 438, 509, and 580 that we interpreted as markers of linear nitrile-imine structures. The other series involved losses of internal [Ala] residues producing ions at m/z 481, 552, and 623. These unequivocally represented cyclic crosslinked ion structures. In addition, the spectrum displayed an m/z 586 ion by loss of $C_6H_8N_2$ (phenylhydrazine) and the m/z 608 ion due to loss of $HN{=}CHCH_2CONH_2$. CID-MS^4 of the latter ion (**Figure 6.10c**) showed dominant $\boldsymbol{v_2}$ and $\boldsymbol{v_1}$ ions (m/z 456 and 385, respectively) indicating structures that were crosslinked off the Asn residue and involved oxygen atom transfer onto the nitrile imine group. Crosslinking in the m/z 608 ion was corroborated by the presence of the m/z 537 ion due to loss of [Ala]. It is noteworthy that although the m/z 608 ions from (NAAA-tet-K–N_2 + H)$^+$ (**Figure 6.10c**) and (DAAAtet-K–N_2 + H)$^+$ (**Figure 6.5c**) had the same elemental composition, their spectra differed, indicating different isomers or mixtures thereof.

CID-MS^3 of the Gln ion, (QAAA-tet-K–N_2 + H)$^+$ (m/z 708, **Figure 6.10b**) also displayed a peak by loss of ammonia (m/z 691), as well as the [$\boldsymbol{y_n}$ +2H]$^+$ and loss of [Ala] ion series. In addition, the spectrum showed $\boldsymbol{v_1}$-$\boldsymbol{v_3}$ ions at m/z 385, 456, and 527, that were analogous to those from (EAAA-tet-K–N_2 + H)$^+$ (**Figure 6.6c**). As discussed above (**Scheme 6.2**), the formation of n ions was associated with crosslinking reactions resulting in oxygen transfer to the nitrile imine group. With (QAAA-tet-K–N_2 + H)$^+$, the oxygen originated from the carboxamide and peptide amide groups. The role in oxygen transfer of the Asp, Glu, and Gln side chain groups was consistent with previous data on crosslinking in peptide conjugates N-terminated with Gly and Ala that lacked reactive sidechain groups and showed no oxygen transfer reactions.

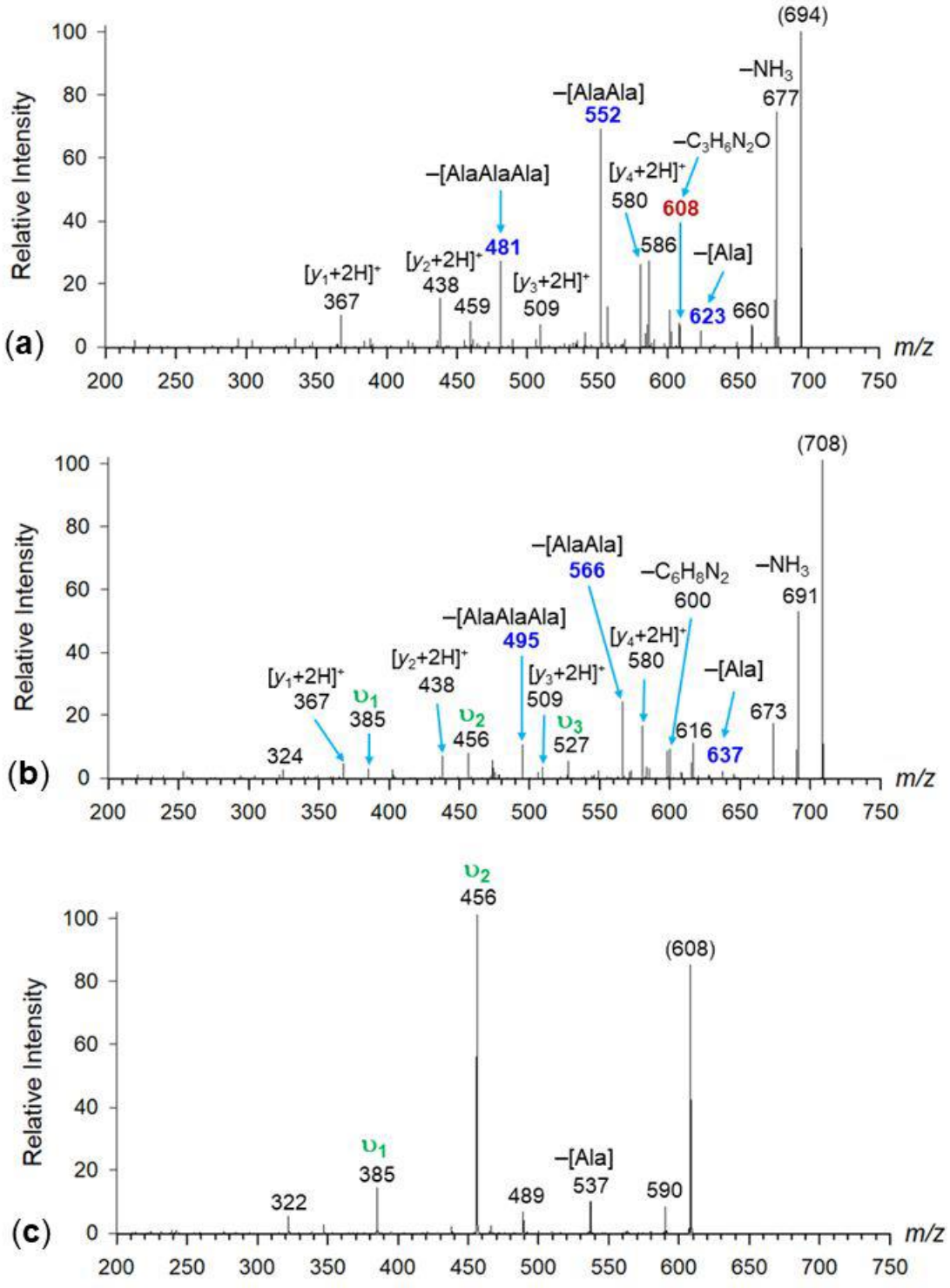


Figure 6.10: CID-MS^3 spectra of (a) (NAAA-tet-K – N_2 + H)$^+$ (m/z 694); (b) (QAAA-tet-K – N_2 +H)$^+$ (m/z 708). (c) CID-MS^4 spectrum of the m/z 608 ion from Figure 6.10a.

The crosslinking yields in UVPD-MS^2 for NAAA-tet-K and QAAA-tet-K, 29 and 57%, respectively, closely followed those for DAAA-tet-K and EAAA-tet-K (**Table 6.5**). The CID-MS^3 data for $(NAAA\text{-}tet\text{-}K–N_2 + H)^+$ and $(QAAA\text{-}tet\text{-}K–N_2 + H)^+$ were similar at 72 and 69% respectively. The larger contribution of υ_n ions from $(QAAA\text{-}tet\text{-}K–N_2 + H)^+$ was balanced by the larger fractions of ions due to loss of $[Ala_n]$ and phenylhydrazine from $(NAAA\text{-}tet\text{-}K–N_2 + H)^+$ (**Table 6.5**).

6.3.4 NAAA-tet-K and QAAA-tet-K Ion Structures and Ion Mobility

The spectra indicated that the polar Asp, Glu, Asn, and Gln side-chain groups participated in crosslinking to the nitrile-imine intermediates in gas-phase ions. It was of interest to examine the structures of the precursor ions and denitrogenated products to further identify the reactive groups and explore potential relationships between the precursor and product ion conformations. To that end we used BOMD and DFT calculations to thoroughly map the conformational space of the precursor ions to establish low-Gibbs energy structures that can be relevant for gas-phase ion populations. For selected $(NAAA\text{-}tet\text{-}K + H)^+$, $(QAAA\text{-}tet\text{-}K + H)^+$, $(NAAA\text{-}tet\text{-}K–N_2 + H)^+$, and $(QAAA\text{-}tet\text{-}K–N_2 + H)^+$ ions, we also used ion mobility measurements to obtain collision cross sections (CCS_{exp}) to be compared with the calculated data (CCS_{calc}) to aid structure assignment. Crosslinking in these ion mobility experiments was triggered by CID. The CID-MS^3 data of ions generated by UVPD and CID revealed no major differences in the spectra, indicating that these two fragment ion sets were comparable.

The arrival time distribution profile of $(NAAA\text{-}tet\text{-}K + H)^+$ showed a single peak of 2.2 ms full width at half maximum (fwhm) after 5 cycles (490 cm path length). The measured CCS_{exp} = 269.5 $Å^2$ was matched by the lowest-Gibbs energy ions **N1** and **N2** that had CCS_{calc} = 274 and 271 $Å^2$, respectively (**Figure 6.11**). The calculated values differed from the experimental one by only 1.7% and 0.6%, which did not allow us to prioritize **N2** over **N1**. Structures **N1** and **N2** had a very similar folding pattern of the peptide chain with an internally solvated N-terminal ammonium group, and developed major hydrogen bonds involving the Lys carboxyl. The main difference distinguishing **N1** and **N2** was the rotational orientation of the tetrazole ring which was N_2-syn in **N1** and N_2-anti in **N2** (**Figure 6.11**). According to the analysis of the BOMD trajectories, the phenyltetrazole moiety was undergoing a facile rotation about the connecting C–C and N–C bonds, exchanging

the N_2-syn and N_2-anti orientation multiple times in the course of 20 ps. It can be presumed that structures **N1** and **N2** rapidly interconverted during the ion mobility travel, so that the measured CCS_{exp} was a population-averaged value.

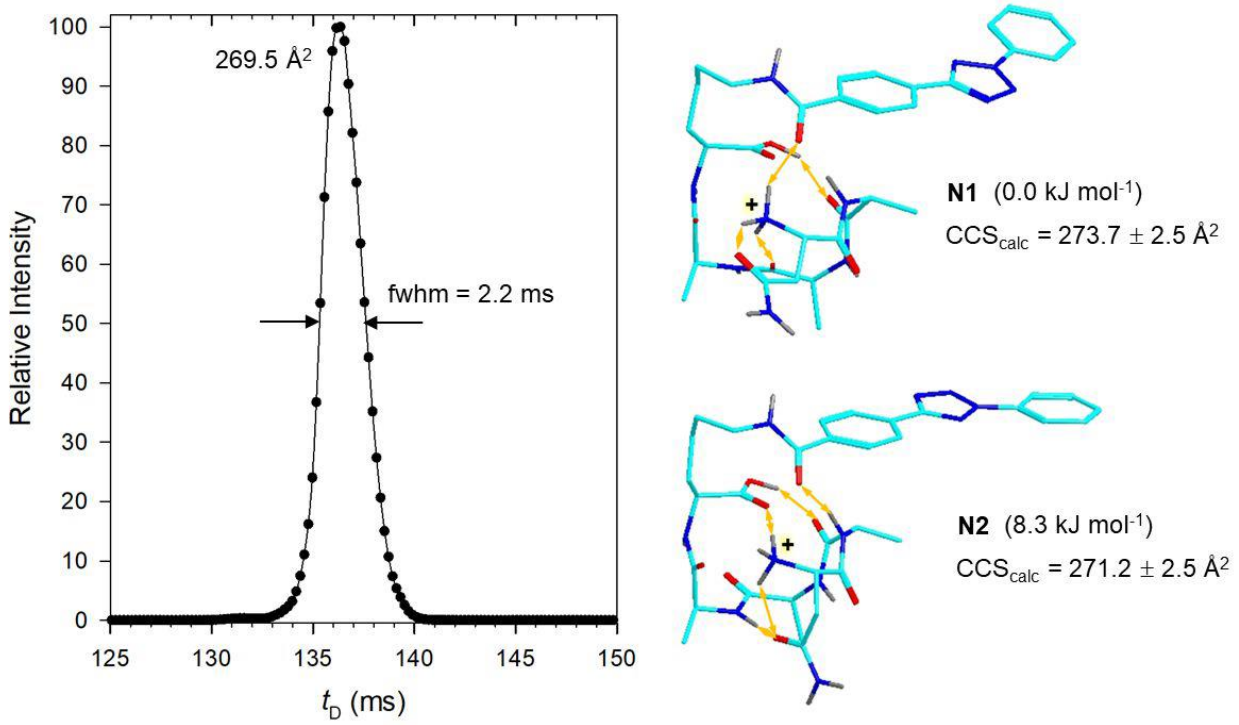


Figure 6.11: Left panel: Arrival time profile of $(NAAA\text{-}tet\text{-}K + H)^+$ after 5 cycles. Right panel: M06-2X/6-31+G(d,p) optimized geometries of ions **N1** and **N2**. Atom color coding is as follows: cyan = C, blue = N, red = O, gray = H. Only exchangeable O-H and N-H hydrogens are displayed to avoid clutter. Relative Gibbs energies are from M062X/def2qzvpp single-point energy calculations including B3LYP/6-31+G(d,p) zero-point corrections, and 310 K enthalpies and entropies.

The arrival time distribution profile of $(QAAA\text{-}tet\text{-}K + H)^+$ showed after 5 cycles a broader, unsymmetrical peak that indicated multiple components. The peak maximum was assigned CCS_{exp} = 270.7 Å^2. Fitting with Gaussian functions gave two major components of CCS_{fit} = 270.2 and 273.0 Å^2 and a more compact minor component with CCS_{fit} = 266.5 Å2 (**Figure 6.12**). The two lowest- Gibbs energy ion structures (**Q1** and **Q2**) had the calculated cross sections at CCS_{calc} = 278.9 and 277.6 Å^2 which were ca. 2.7 % larger than the fitted experimental values. A third, more compact, component of a higher ΔG_{310} (**Q3**) had CCS_{calc} = 266.1 Å^2, matching the CCS_{exp} of the minor component in the ion mixture. Structures **Q1** and **Q2** differed in their folding patterns that were determined by different hydrogen bonds. In particular, the H-bonding of the Gln amide to

the ammonium group in **Q2** resulted in a more compact structure than that of **Q1** that was reflected by a relatively smaller CCS_{calc} for the former isomer (**Figure 6.12**). Structure **Q3** had the Gln residue folded towards the diaryltetrazole moiety, resulting in a more compressed structure of a higher energy and smaller CCS_{calc}.

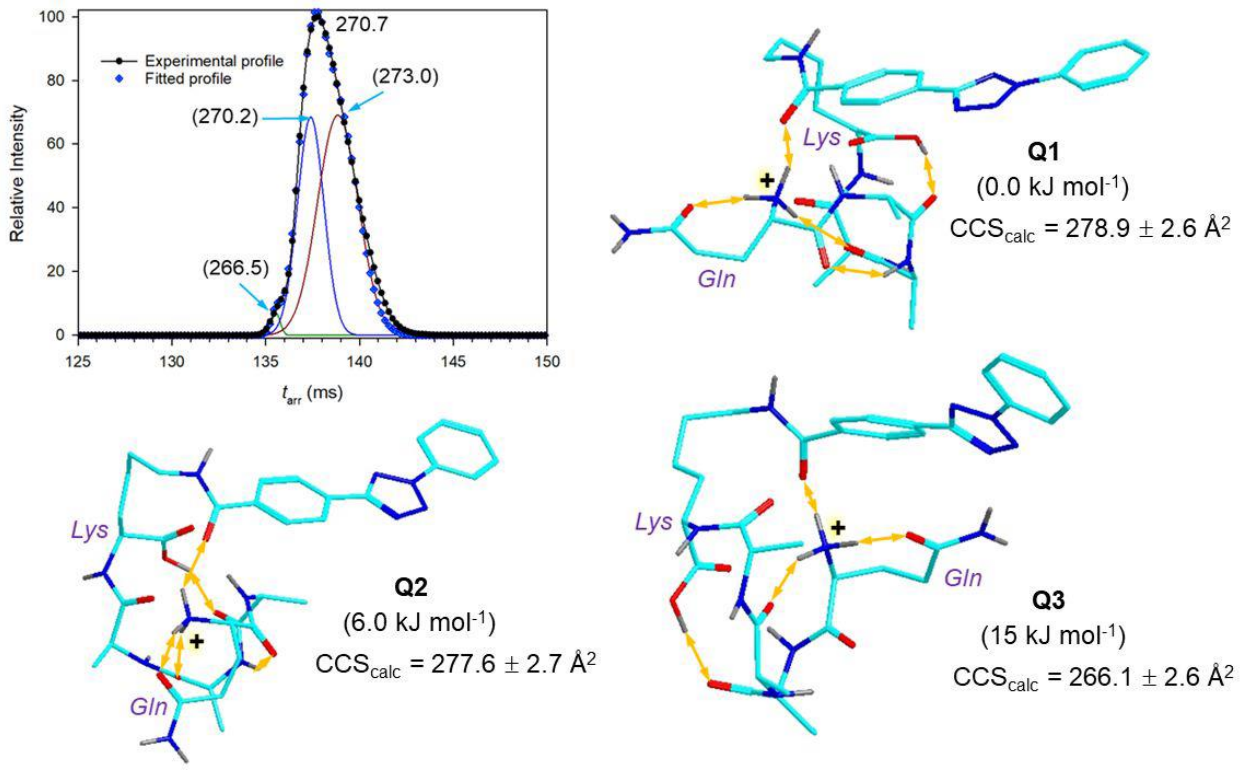


Figure 6.12: Top left panel: Arrival time profile of $(QAAA\text{-}tet\text{-}K + H)^+$ after 5 cycles with fitted Gaussian peaks and CCS_{exp} in Å^2. M06-2X/6-31+G(d,p) optimized geometries of ions **Q1**-**Q3**. Atom color coding and relative Gibbs energies are as in Figure 6.11.

Ion mobility measurements of the CID-generated $(NAAA\text{-}tet\text{-}K\text{–}N_2 + H)^+$ showed multiple components in the arrival time profile that had CCS_{exp} = 249.9, 255.9, 260.0, 264.6, and 267.2 Å^2, with more partially resolved components, as shown in **Figure 6.13**. To assign the peaks, we considered several types of structures in our BOMD and DFT computational analysis. Among the nitrile imines, the lowest-energy conformer **N3** had CCS_{calc} = 255.2 Å^2 that coincided with the second peak in the arrival time profile. It should be mentioned that nitrile imines were calculated as high-energy intermediates whose presence in the ion mixture was given by kinetic constraints to cyclization and crosslinking. Asn amide O-crosslinked isomers **N4a** and **N4b** had CCS_{calc} = 249.4 and 260.6 Å^2, respectively, that coincided with the first and third peak in the arrival time

profile. We note that structures of the **N4** type were formed by oxygen transfer to the nitrile imine group and conducive to the formation of υ_n fragment ions upon activation. Another type of Asn amide crosslinked isomers were represented by structures **N5**. These were the lowest-energy crosslinks, e.g., cyclization of **N3** to **N5c** was exergonic by ΔG_{310} = –113 kJ mol^{-1} (**Scheme 6.3**). Several **N5** structures were considered, such as **N5a**-**N5c** that had CCS_{calc} close to the CCS_{exp} of the peaks in the arrival time profile (**Figure 6.13**). For additional conformers of the **N5** type, see **N5d**-**N5f** (**Figure 6.14**). Structures **N5** had strong N–C bonds forming the crosslinks, and upon excitation were expected to undergo peptide backbone cleavages resulting in [Ala] neutral losses. Another structure type was needed for ions undergoing loss of $HN{=}CHCH_2CONH_2$ that indicated crosslinking off the Asn residue. These can be represented by structures **N6a**-**N6c** in which crosslinking was realized by the Asp-Ala chain amide and was followed by further rearrangement (**Figure 6.14**). Similar structures were identified in crosslinks of $(GAAA\text{-}tet\text{-}K - N_2 + H)^+$ where their presence was supported by ion mobility and UV-action spectroscopy.[6] According to their CCS_{calc} = 261.1 and 261.9 Å^2 for **N6a** and **N6b**, respectively (**Figure 6.14**), structures of the **N6** type were likely to belong to the minor components of the crosslinked products, as shown in the arrival time profile (**Figure 6.13**).

Having obtained the energies of $(NAAA\text{-}tet\text{-}K{-}N_2 + H)^+$ ions, we were able to assess the energetics of some of the less usual dissociations of the crosslinked ions. **Scheme 6.3** shows the M06-2X/def2qzvpp energy, corrected for zero-point vibrational terms, for the ring closure in nitrile imine **N3** forming the lowest-energy crosslink **N5c**. The ring closure was 113 kJ mol^{-1} exothermic, virtually preventing reverse ring opening to the nitrile imine. Ion **N5c**, which retained the N-N-phenyl linkage, was the presumed precursor for the loss of phenylhydrazine that required a dissociation of the C–N bond and transfer of two hydrogen atoms onto the leaving group. Several product peptide ions were considered having cyclized structures. In **N7a**, cyclization involved the N-terminal amine group, forming the protonated amidine ring (**Scheme 6.3**). In **N7b**, the crosslinking N–C bond was disrupted as a result of the phenyl hydrazine loss, forming a nitrile that contained the original Asp side-chain amide nitrogen in the C-N group. The exposed Asn sidechain carbonyl then cyclized with the N-terminal amine, forming a four-membered lactam ring. The salient feature of these cyclization-involving dissociations was that the calculated threshold energies (230-235 kJ mol^{-1}, **Scheme 6.3**) were comparable to those for typical peptide backbone cleavages that have been reported ranging between 162-263 kJ mol^{-1}.[34-37] This was

consistent with the observed competition between the two types of ion dissociations (**Figure 6.9a, Figure 6.10**).

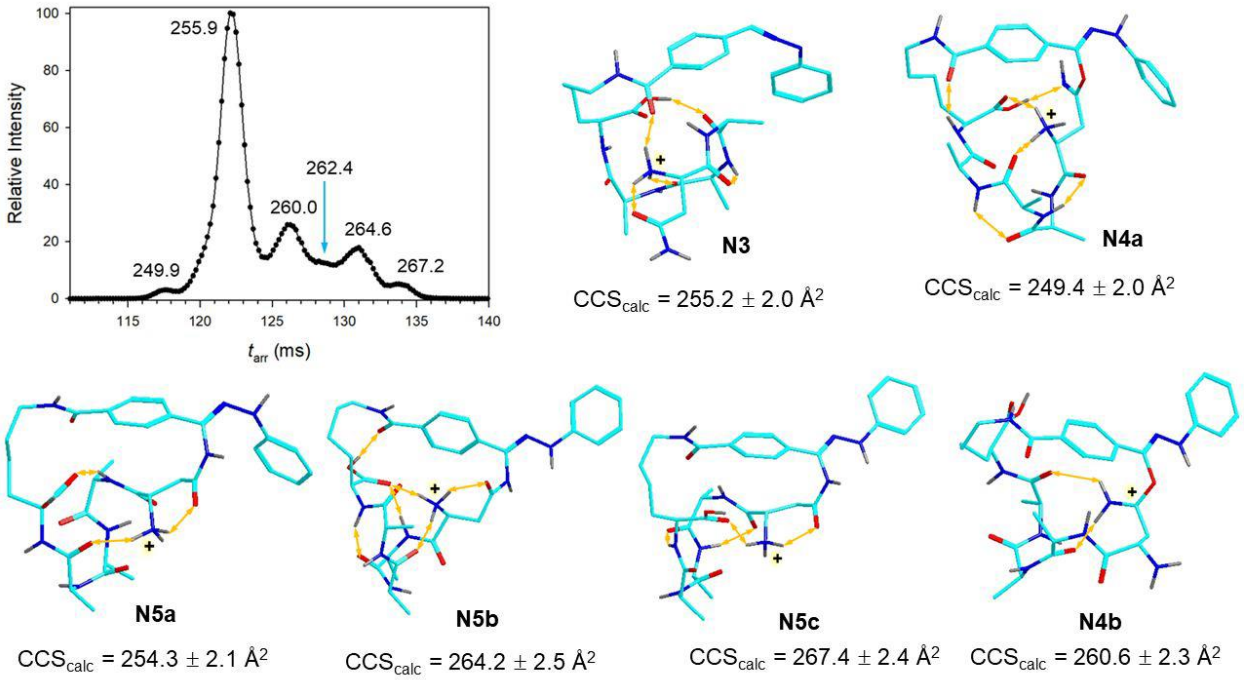


Figure 6.13: Top left panel: Arrival time profile of (NAAA-tet-K-N_2 + H)$^+$ after 5 cycles with CCS_{exp} in Å^2. M06-2X/6-31+G(d,p) optimized geometries of ions **N3**-**N5b**. Atom color coding and relative Gibbs energies are as in Figure 6.11.

Multiple isomers were also observed for the arrival time distribution of (QAAA-tet-K–N_2 + H)$^+$ ions produced by CID of (QAAA-tet-K + H)$^+$, giving peaks at CCS_{exp} = 254.4, 259.6, 262.6, 265.9, and 268.6 Å^2 (**Figure 6.15**). Our survey of optimized structures for the (QAAA-tet-K–N_2 + H)$^+$ ions revealed nitrile imines **Q4a** and **Q4b** of the respective CCS_{calc} = 256.5 and 269.1 Å^2, matching the first and fifth peak, respectively, in the arrival time distribution. Regarding cyclic structures, ions **Q5a**-**Q5c** were isomers crosslinked by the Gln amide that were 34-65 kJ mol^{-1} more stable than the lowest-energy nitrile imine **Q4b**. That the less stable nitrile imines were detected in the ion mixture pointed to kinetic constraints to crosslinking cyclization. The amide-N-linked structures showed CCS_{calc} = 256.5, 259.7, and 264.0 Å^2 which were close to the CCS_{exp} of the first through fourth peaks in the arrival time distribution (**Figure 6.15**). The amide-N-linked structures were considered to be representative of the fraction of (QAAA-tet-K–N_2 + H)$^+$ ions that

underwent loss of internal [Ala] which was the dominant dissociation on collisional activation (**Figure 6.10b**).

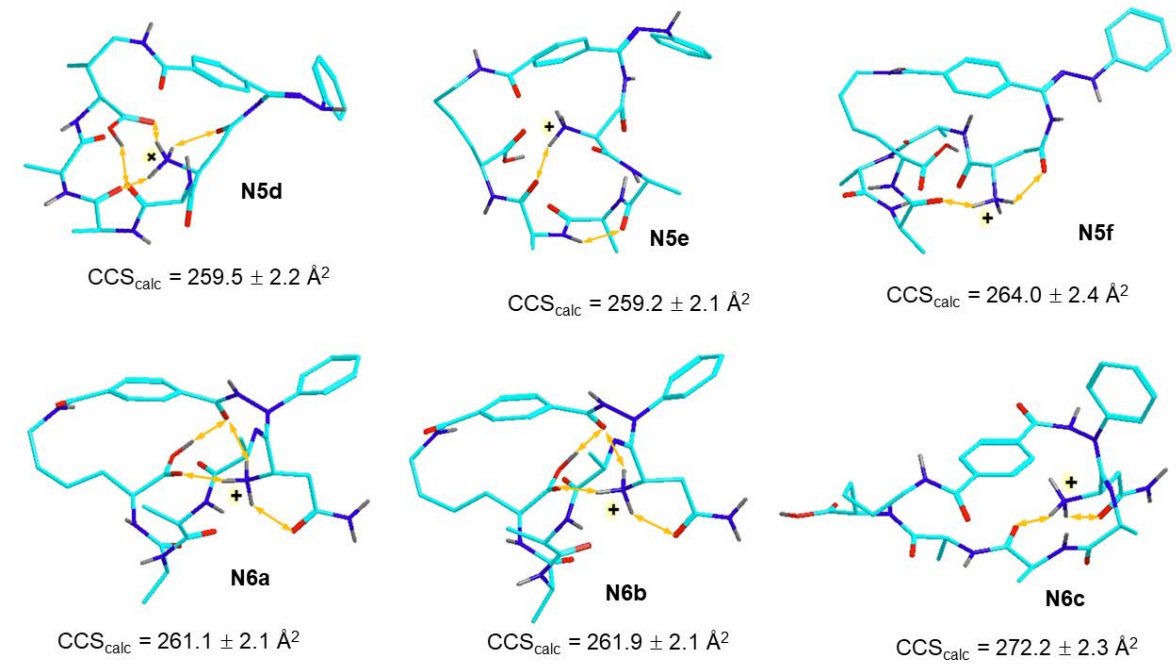


Figure 6.14: M06-2X/6-31+G(d,p) optimized structures of (NAAAK-tet-K – N_2 + H)$^+$ ions. Atom color coding is as in Figure 6.11.

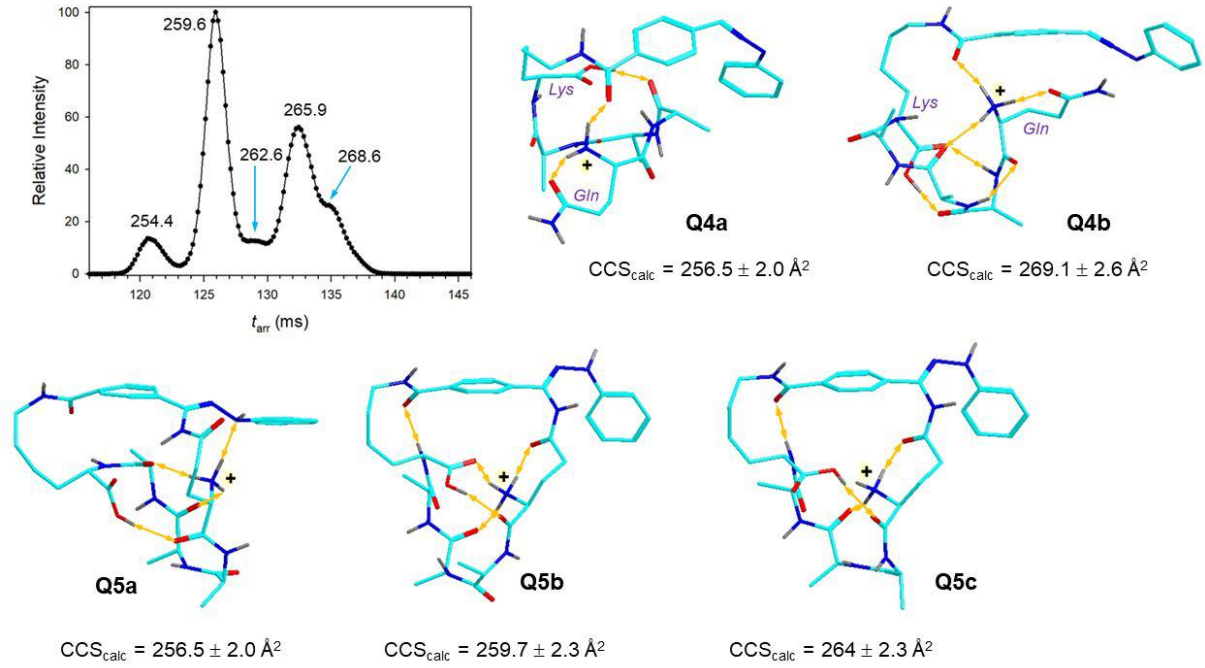
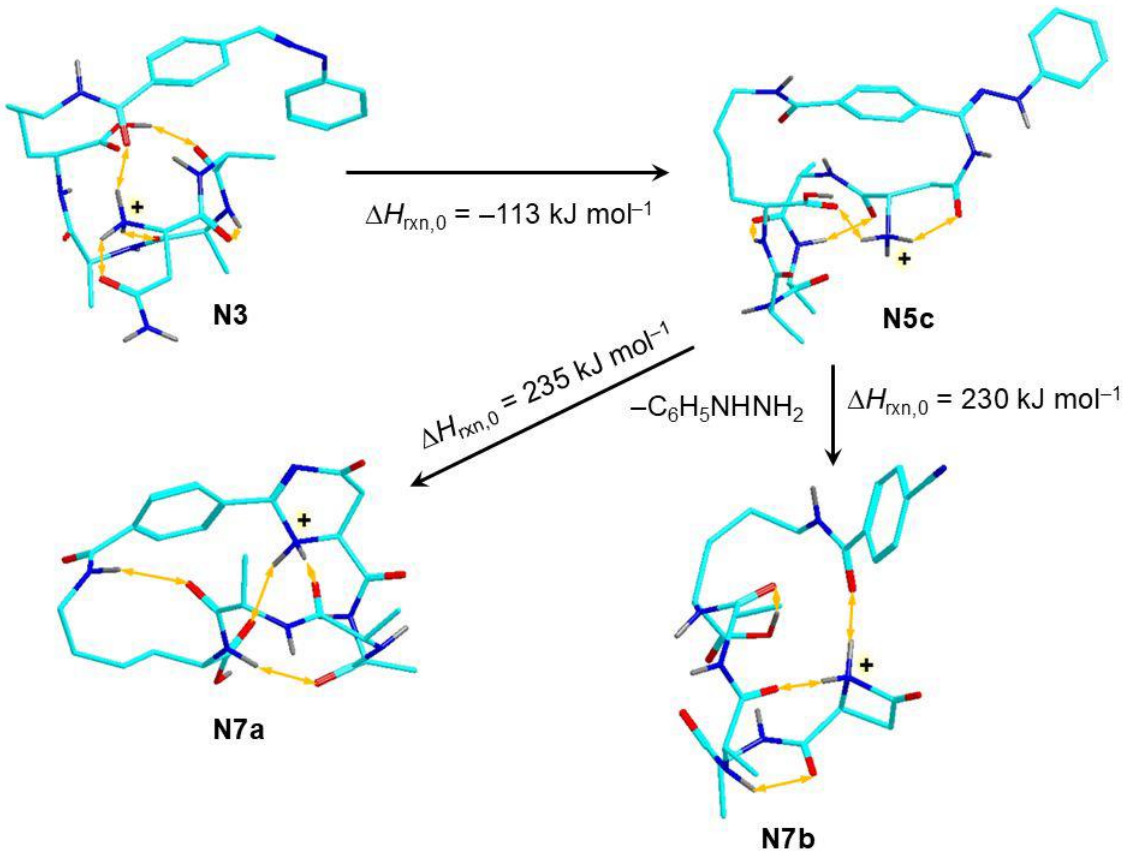


Figure 6.15: Top left panel: Arrival time profile of (QAAA-tet-K–N_2 + H)$^+$ after 5 cycles with CCS$_{exp}$ in Å^2. M06-2X/6-31+G(d,p) optimized geometries of ions **Q4a,b** and **Q5a-c**. Atom color coding is as in Figure 6.11.



Scheme 6.3: Intermediates and Reaction Enthalpies at 0 K ($\Delta H_{rxn,0}$) for Loss of Phenylhydrazine from $(NAAA\text{-}tet\text{-}K\text{–}N_2 + H)^+$.

6.3.5 DAAA-tet-K and EAAA-tet-K Ion Structures

The CID-MS^3 spectra of $(DAAA\text{-}tet\text{-}K - N_2 + H)^+$ and $(EAAA\text{-}tet\text{-}K - N_2 + H)^+$ showed series of fragment ions indicating mixtures of nitrile-imine and crosslinked structures of different types (**Figure 6.1c** and **6.6c**, respectively). An important general question regarding nitrile-imine crosslinking was whether the conjugate precursor ion conformation can be related to the formation and composition of the crosslinks. Our previous analysis of peptide-nitrile imine scaffolds indicated that crosslinking was governed by the ion dynamics rather than initial static ion structures.[7] We now tested this hypothesis by obtaining and comparing low-energy structures for $(DAAA\text{-}tet\text{-}K + H)^+$ and $(EAAA\text{-}tet\text{-}K + H)^+$ that according to the spectra gave similar composition of crosslinks on UVPD-MS^2 and CID-MS^3. The fully optimized structures of the lowest energy conformers are shown in **Figure 6.16**. The lowest-energy $(DAAA\text{-}tet\text{-}K + H)^+$ ion **D1** showed the peptide moiety folded on itself and its conformation was held by multiple hydrogen

bonds. The Asp carboxyl oxygens were moderately remote, d(C–O) = 5.5 Å at closest approach, from the incipient nitrile imine carbon that was the presumed center of nucleophilic attack in the nitrile imine. The position of the diaryltetrazole with respect to the peptide was controlled by two hydrogen bonds, one between the N-terminal NH_3 charged group and the aryl amide C=O, and the other between a tetrazole nitrogen and the Asp amide NH. The other low energy structures, **D2** and **D3**, had even longer distances for the Asp carboxyl oxygens at d(C–O) = 10.4 and 8.8 Å, respectively (**Figure 6.16**). In contrast, all the lowest-energy $(EAAA\text{-}tet\text{-}K + H)^+$ ions, **E1-E3**, showed close proximity of the Glu carboxyl oxygens to the future nitrile-imine carbon at d(C–O) = 3.7, 3.7, and 3.3 Å, respectively. These features were due to a favorable hydrogen bond of the Glu COOH to the tetrazole ring nitrogen atoms (**Figure 6.16**) that was enabled by the longer Glu side chain. Hence, it would seem that the $(EAAA\text{-}tet\text{-}K + H)^+$ ions had a higher propensity for crosslinking by the Glu carboxyl. The **Table 6.5** data for UVPD-MS^2 showed that $(EAAA\text{-}tet\text{-}K + H)^+$ produced a significantly higher yield of crosslinks, as judged by the prominent n fragment ions (**Figure 6.6b**) that involved O–C bond formation to the nitrile imine.

6.4 Conclusions

The results of this combined experimental and computational study allowed us to arrive at the following conclusions. Nitrile-imine intermediates were efficiently produced by UVPD or CID of peptide-tetrazole conjugates containing N-terminal Asp, Glu, Asn, and Gln residues. The nitrile imine formation by loss of N_2 from the tetrazole ring was found to be a facile dissociation of a low threshold energy that competed with peptide backbone cleavages and was particularly promoted by photoexcitation. Asp, Glu, Asn, and Gln side chains participated in covalent bonding to the nitrile imine group, forming cyclized crosslinks with yields reaching 83%. The carboxyl groups in Asp and Glu reacted with the nitrile imines as nucleophiles following proton transfer and forming O=C-O–C bonds. The side-chain amide groups in Asn and Gln reacted to form new CONH–C bonds with the nitrile-imine carbon atom. Upon collisional activation, the cyclized products dissociated by two major pathways. One was cleavage of the peptide amide bonds resulting in the ejection of neutral internal amino acid residues [Ala], [AlaAla], and [AlaAlaAla]. The other pathway commenced by proton-promoted dissociation of the crosslinking O=C–O-C bond, retaining the former carboxyl oxygen and two hydrogen atoms in the nitrile imine, that was followed by standard peptide amide bond dissociations leading to modified C-terminal sequence

fragment ions. Another specific dissociation promoted by crosslinking to nitrile imines was the loss of phenylhydrazine. These specific reactions allowed us to distinguish the linear and cyclized peptide products and assess the crosslinking yields. Nitrile-imine crosslinking has emerged as a general method for the investigation of peptide structure and reactivity in gas-phase ions.

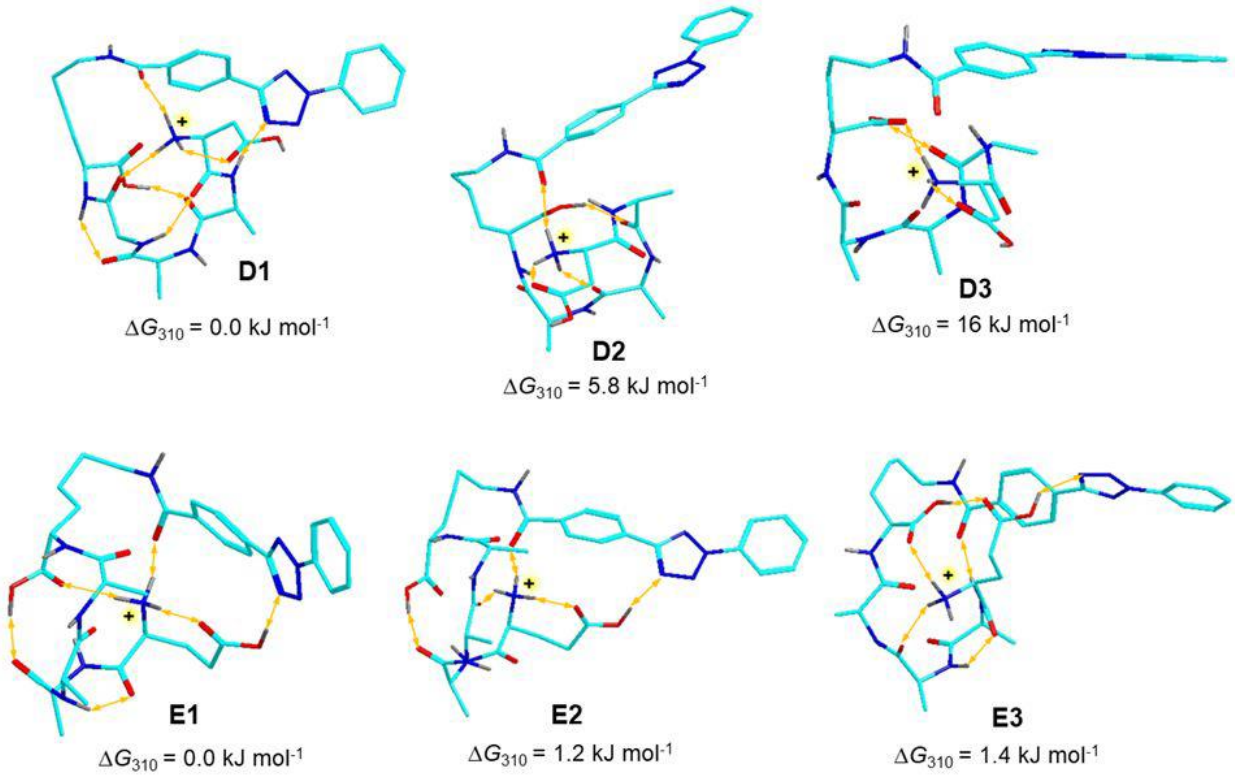


Figure 6.16: M06-2X/6-31+G(d,p) optimized structures of low-energy $(DAAA\text{-}tet\text{-}K + H)^+$ and $(EAAA\text{-}tet\text{-}K + H)^+$ ions. Relative Gibbs energies in each set are from M076-2X/def2qzxvpp single point energy calculations and include zero-point energy corrections, enthalpies and entropies. Atom color coding is as in Figure 6.11.

6.5 Bibliography

[1] 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.

[2] Shawali, A. S. Reactions of Heterocyclic Compounds with Nitrilimines and Their Precursors. *Chem. Rev.* **1993**, *93*, 2731–2777.

[3] 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.

[4] 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.

[5] Zhang, J.; Liu, J.; Li, X.; Ju, Y.; Li, Y.; Zhang, G.; Li, Y. Unexpected Cyclization Product Discovery from the Photoinduced Bioconjugation Chemistry between Tetrazole and Amine. *J. Am. Chem. Soc.* **2024**, *146*, 2122-2131.

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

[7] Zhu, H.; Nytka, M.; Vu, T. N. K.; Lemr, K.; Tureček, F. Photochemical and Collision- Induced Crosslinking in Stereochemically Distinct Scaffolds of Peptides and NitrileImines in Gas-Phase Ions. *J. Am. Soc. Mass Spectrom.* submitted July 2024.

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

[9] Dove, J. E.; Hippler, H.; Troe, J. Direct Study of Energy Transfer of Vibrationally Highly Excited Carbon Disulfide Molecules. *J. Chem. Phys.* **1985**, *82*, 1907−1919.

[10] Heymann, M.; Hippler, H.; Troe, J. Collisional Deactivation of Vibrationally Highly Excited Polyatomic Molecules. IV. Temperature Dependence of ⟨ΔE⟩. *J. Chem. Phys.* **1984**, *80*, 1853−1860.

[11] Pepin, R.; Tureček, F. Kinetic Ion Thermometers for Electron Transfer Dissociation. *J. Phys. Chem. B*, **2015**, *119*, 2818-2826.

[12] 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.

[13] 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.

[14] 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.

[15] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semiempirical 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.

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

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

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

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

[20] 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.

[21] 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.

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

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

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

[25] 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.

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

[27] 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.

[28] Bleiholder, C.; Osburn, 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.

[29] 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.

[30] Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130*, 351–358.

[31] Tureček, F.; Julian, R. R. Peptide Radicals and Cation-Radicals in the Gas Phase. *Chem. Rev.* **2013**, *113*, 6691-6733.

[32] Zhang, L.; Reilly, J. P.. Peptide de Novo Sequencing Using 157 nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82*, 898–908.

[33] Brodbelt. J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43*, 2757–2783.

[34] Klassen, J. S.; Kebarle, P. Collision-Induced Dissociation Threshold Energies of Protonated Glycine, Glycinamide, and Some Related Small Peptides and Peptide Amino Amides. *J. Am.Chem. Soc.* **1997**, *119*, 6552-6563.

[35] Paizs, B.; Suhai, S. Fragmentation Pathways of Protonated Peptides. *Mass Spectrom. Rev.* **2005**, *24*, 508–548.

[36] Laskin, J.; Yang, Z.-B.; Song, T.; Lam, C.; Chu, I. K. Effect of the Basic Residue on the Energetics, Dynamics, and Mechanisms of Gas-Phase Fragmentation of Protonated Peptides. *J. Am. Chem. Soc.* **2010**, *132*, 16006–16016.

[37] Mookherjee, A.; Armentrout, P. B. Thermodynamics and Reaction Mechanisms for Decomposition of a Simple Protonated Tripeptide, H^+GGA: From H^+GGG to H^+GAG to H^+GGA. *J. Am. Soc.Mass Spectrom*, **2022**, *33*, 355-368.

Chapter 7

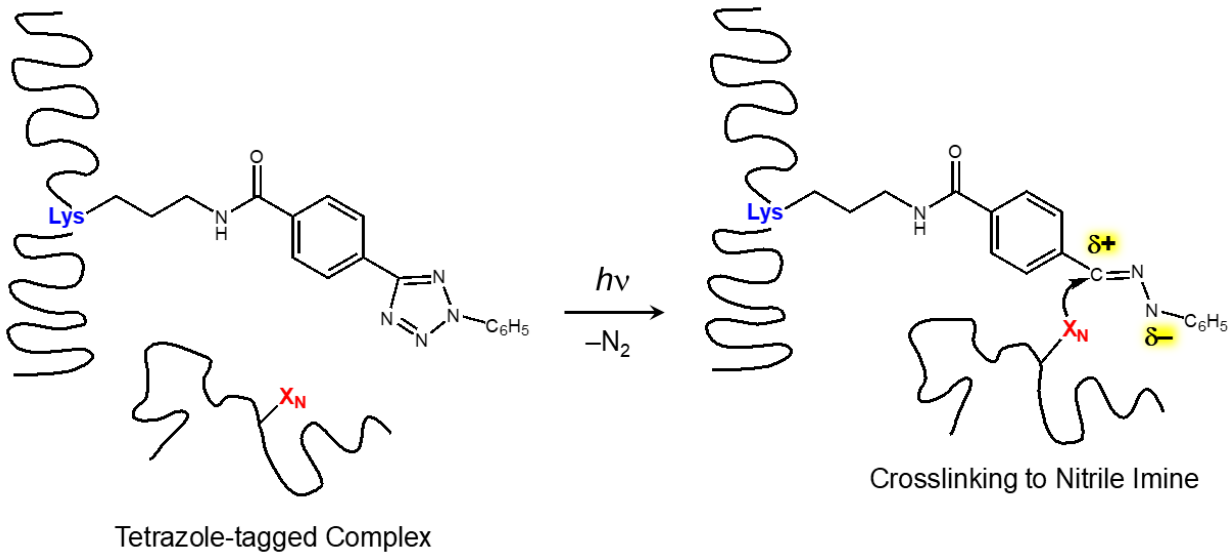
Photochemical and Collision-Induced Crosslinking of Lys, Arg, and His to Nitrile Imines in Peptide Conjugate Ions in the Gas Phase

Reproduced in part with permission from Jiahao Wan, Mikuláš Vlk, Marianna Nytka, Tuan Ngoc Kim Vu, Karel Lemr and František Tureček. Photochemical and Collision-Induced Crosslinking of Lys, Arg, and His to Nitrile Imines in Peptide Conjugate Ions in the Gas Phase. *Submitted to J. Am. Soc. Mass Spectrom.*

Abstract *We report a study of internal covalent crosslinking with photolytically generated diarylnitrile imines of N-terminal arginine, lysine and histidine residues in peptide conjugates. Conjugates in which a 4-(2-phenyltetrazol-5-yl)benzoyl group was attached to C-terminal lysine, that we call RAAA-tet-K, KAAA-tet-K, and HAAA-tet-K, were ionized by electrospray and subjected to UV photodissociation (UVPD) at 213 nm. UVPD triggered loss of N_2 and proceeded by covalent crosslinking to nitrile imine intermediates that involved the side chains of N-terminal arginine, lysine and histidine, as well as the peptide amides. Crosslinking yields were determined from UVPD-MS^2 measurements as 67%, 66%, and 84% for RAAA-tet-K, KAAA-tet-K, and HAAA-tet-K, respectively. CID-MS^3 of the denitrogenated intermediates from RAAA-tet-K, KAAA-tet-K, and HAAA-tet-K indicated overall crosslinking yields of 80%, 89% and 80%, respectively. The nature of the crosslinking reactions and crosslink structures were investigated for RAAA-tet-K by high-resolution cyclic ion mobility mass spectrometry that identified precursor ion conformers and multiple dissociation products. All sequences were subjected to conformational analysis by Born-Oppenheimer molecular dynamics and energy analysis by density functional theory calculations with M06-2X/def2qzvpp that provided relative and dissociation energies for several crosslink structural types. The crosslinking reactions were substantially exothermic, driving the efficient conversion of nitrile imine intermediates to cyclic products. The principal steps in covalent crosslinking involved proton transfer onto the nitrile imine group accompanied by nucleophilic attack by the peptide side-chain and amide groups. Blocking the proton transfer and nucleophile resulted in a loss of crosslinking abilities.*

7.1 Introduction

Nitrile imines are reactive intermediates that have been used in organic chemistry for the synthesis of various nitrogen heterocycles.[1] Nitrile imines show particular reactivity to polar double bonds, such as those in methylene malonates and nitriles, but also have been found to react with nucleophilic groups.[2] In addition to their broad utility for organic synthesis, nitrile imines have been proposed as reagents in biomolecular crosslinking.[3] The nitrile imine intermediate is generated by photodissociation of a suitable precursor, such as a diaryltetrazole group attached to the biomolecule, and allowed to react with a nucleophilic group (X_N) in the target compound to form a covalent bond (**Scheme 7.1**). The Scheme 1 reactions have been employed in a few studies of protein crosslinking where the diaryltetrazole was attached as an amide to the lysine ε-NH_2 group, and the reactive nucleophile was an aspartic acid side-chain carboxyl.[4,5]



Scheme 7.1: Crosslinking of Nitrile-Imine Intermediates with Target Nucleophiles.

Nitrile imines have been reported to be efficient crosslinkers in gas-phase peptide conjugates and non-covalent complexes.[6] In contrast to their condensed phase reactions, the nitrile imine reactivity in the gas phase cations is thought to be enhanced by intramolecular proton transfer forming transient nitrile-iminium ions of increased electrophilicity. This has been shown to activate nitrile imines to react with weak nucleophiles, such as peptide amide groups and nucleobases in complexes with oligonucleotides.[6] We have previously investigated nitrile-imine

crosslinking to N-terminal amino acid residues in singly charged peptide ions containing Asp and Glu carboxyl and Asn and Gln carboxamide groups.[7] These have shown dual reactivity, whereby crosslinking was mediated to different degrees by the carboxyl or carboxamide, as well as by the peptide chain amide groups. Nitrile-imine crosslinking to C-terminal carboxyl groups has been reported for stereochemically distinct peptide scaffolds.[8]

Here, we explore the reactivity of N-terminal residues possessing basic and nucleophilic functionalities, such as a primary amine group in lysine, a guanidine group in arginine, and an imidazole ring in histidine. The peptide conjugates we investigated carried the 2,5-diaryltetrazole group (tet) attached as an amide at the C-terminal lysine residue. These are denoted as KAAA-tet-K, RAAA-tet-K, and HAAA-tet-K. The basic amino acid residues in peptides are known to lower the peptide ion reactivity toward backbone cleavage by sequestering the charging proton and making it energetically less favorable for amide protonation.[9-11] From the crosslinking point of view, this could be a favorable feature because it may competitively enhances tetrazole ring cleavage upon non-specific, collision-induced excitation. In this way, nitrile imines can be generated by both UV photodissociation (UVPD) and collision-induced dissociation (CID), which expands the options for further ion structure analysis by UVPD and CID tandem mass spectrometry, as well as by high-resolution ion mobility measurements. Conversely, the high basicity of Lys, Arg, and His functional groups may interfere with proton transfer to the nitrile imine to hamper crosslinking reactivity. We wish to show by experiment and theory that the arginine, lysine, and histidine basic/nucleophilic groups are reactive towards nitrile imines. To that end, we use a combination of tandem and cyclic ion mobility mass spectrometry and theoretical conformational and energy analysis to detect and identify the likely crosslink structures.

7.2 Experimental Section

7.2.1 Materials and Methods

Peptide conjugates with 4-(2-phenyl-2H-tetrazol-5-yl)benzoic acid (Sigma-Aldrich, St. Louis, MO), were synthesized by standard Fmoc coupling methods as reported previously.[6] The peptides were characterized by accurate m/z measurements and sequencing of electrosprayed ions in high-resolution mass spectrometry. The ε-N,N,N-trimethyl lysine derivative was prepared by

exhaustive methylation of N-t-Boc-protected lysine that was built into the peptide sequence. To prepare the coupling solution, dissolve Boc-Lys(Fmoc)-OH (3 equiv.), benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) (3 equiv.), and hydroxybenzotriazole (HOBt) (4.2 equiv.) in dry *N*,*N*-dimethylformamide (DMF). Inject the solution into the reaction syringe containing on-resin AAA-*tet*-K and shake for 20 minute. After the coupling step, remove the liquid phase and wash the beads with DMF six times. Deprotection of the Fmoc group on the lysine side chain was carried out by injecting 20% piperidine in DMF and shaking for 10 minutes. Then wash the beads again with DMF six times to ensure complete removal of piperidine residues. The trimethylation cocktail was prepared by mixing methyl iodide (30 equiv.) and *N*,*N*-diisopropylethylamine (10 equiv.) in DMF. Inject this cocktail into the reaction syringe and shake for 2 hours to allow the reaction to reach completion. Next, eject the liquid phase and wash the beads with DMF six times, followed by six washes with dichloromethane. Strip the peptide from the resin by injecting a mixture of TFA/water/triethylsilane (95:2.5:2.5) twice. The Boc protecting group was effectively removed during the peptide-resin cleavage process. Trimethylation was confirmed by HRMS and further validated by the detection of trimethylamine elimination in CID-MS^2. Multistage UV photodissociation (UVPD) and CID-MS^n spectra were measured on an Orbitrap Ascend Tribrid instrument (ThermoFisher, San Jose, CA). For UVPD at 213 nm, we used multiple scans with a 20 ms stage time (33 laser pulses per scan) at 1.5 ± 0.2 μJ/pulse. Spectra were acquired by scanning the ion trap, or in the high-resolution mode in the Orbitrap at 125 000 resolving power. Accurate m/z and ion assignments are shown in **Tables 7.1-7.4**. Ion mobility measurements were carried out on a SELECT SERIES Cyclic Ion Mobility Spectrometer (c-IMS) (Waters Corp., Wilmslow, U.K.)[12] with direct infusion into a normal flow electrospray ion source at a flow rate of 5μL/min. Each sample and calibrant were measured six times in a positive mode. Typically, ion mobility separation over several cycles (n) was used at the total ion path of n × 98 cm. Details of all measurement parameters and calibration for collision cross section (CCS) determination have been reported previously.[6]

7.2.2 Calculations

Born-Oppenheimer molecular dynamics (BOMD) calculations were carried out as described previously.[13] Briefly, multiple initial ion conformers were built in which the charging proton was placed on the basic residue (R, K, or H). BOMD trajectories were run for 20 ps at 510, 610, and 810 K with PM6-D3H4 that included empirical dispersion and hydrogen bonding interactions.[14] These calculations were run on the high-level Cuby4 platform[15] using MOPAC.[16] Inspection of the trajectories revealed 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, and so no longer trajectory runs were necessary because they generated only duplicate structures. The Berendsen thermostat[17] was used for energy stabilization. Two hundred snapshots were extracted from each trajectory at regular 100 fs intervals, and the geometries were gradient-optimized with PM6-D3H4. These were sorted out to weed out duplicates and 20-30 lowest-energy structures were optimized with B3LYP[18] and M062X[19] with the 6-31+G(d,p) basis set including B3LYP/6-31+G(d,p) harmonic frequency analysis. The B3LYP calculations were augmented by including empirical dispersion corrections (GD3-BJ).[20,21] The M06-2X/6-31+G(d,p) optimized geometries were used to calculate single-point energies; these calculations used M06-2X and the def2qzvpp[22] basis set (up to 4635 basis functions). Basis set superposition error (BSSE) corrections were not applied when calculating ion dissociation energies because the def2qzvpp basis set nearly reaches the complete basis set limit where BSSE becomes negligible.[23] 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).[24,25] 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 (CCS_{calc}) were calculated by the modified ion trajectory method ($MobCal_{MPI}$)[26,27] using the MK charge densities. Ten trajectories were run amounting to 595840 points. Average CCS_{calc} values are reported with standard deviations. Standard van der Waals parameters for different atom types were obtained from the MMFF94 data set.[28]

Table 7.1 Accurate Mass Measurements for (RAAA-*tet*-K + H)$^+$ Ions.

m/z	ion formula	error (mmu)	neutral loss	ion assignment
764.3940	$C_{35}H_{50}N_{13}O_7$	-1.1		
736.3880	$C_{35}H_{50}N_{11}O_7$	-0.9	N_2	
719.3621	$C_{35}H_{47}N_{10}O_7$	-0.2	NH_3	
719.3621	$C_{35}H_{48}N_{11}O_6$	-0.5	H_2O	
645.3465	$C_{29}H_{45}N_{10}O_7$	-0.3	C_6H_8N	
622.3094	$C_{30}H_{40}N_9O_6$	-0.2	$C_5H_{10}N_2O$	$\boldsymbol{v}_4$
551.2724	$C_{27}H_{35}N_8O_5$	-0.1	$C_8H_{15}N_3O_2$	$\boldsymbol{v}_3$
480.2352	$C_{24}H_{30}N_7O_4$	-0.2	$C_{11}H_{20}N_4O_3$	$\boldsymbol{v}_2$
409.1982	$C_{21}H_{25}N_6O_3$	-0.1	$C_{14}H_{15}N_5O_2$	$\boldsymbol{v}_1$
388.2301	$C_{15}H_{30}N_7O_5$	-0.2	$C_{20}H_{20}N_4O_2$	$\boldsymbol{b}_4 + H_2O$
371.2035	$C_{15}H_{27}N_6O_5$	-0.2	$C_{20}H_{23}N_5O_2$	
346.1660	$C_{20}H_{20}N_5O$	-0.2		
263.0927	$C_{15}H_{11}N_4O$	0.05	$C_{20}H_{39}N_7O_6$	
221.0709	$C_{14}H_9N_2O$	0.04	$C_{21}H_{41}N_9O_6$	

Table 7.2 Accurate Mass Measurements for (KAAA-*tet*-K + H)$^+$ Ions.

m/z	ion formula	error (mmu)	Neutral loss	ion assignment
736.3881	$C_{35}H_{50}N_{11}O_7$	-0.9		

708.3821	$C_{35}H_{50}N_9O_7$	-0.7	N_2	
608.2823	$C_{30}H_{38}N_7O_7$	-0.4	$N_2 + C_5H_{12}N_2$	
602.3294	$C_{29}H_{44}N_7O_7$	-0.3	$C_6H_6N_4$	
600.3137	$C_{29}H_{42}N_7O_7$	-0.5	$C_6H_8N_4$	$-N_2$ -phenylhydrazine
537.2567	$C_{26}H_{33}N_8O_5$	-0.1	$C_9H_{17}N_3O_2$	$[\boldsymbol{y_3} + 2H]^+$
529.2765	$C_{26}H_{37}N_6O_6$	-0.4	$C_9H_{13}N_3O$	600-[Ala]
466.2195	$C_{23}H_{27}N_7O_4$	-0.2	$C_{12}H_{22}N_4O_3$	$[\boldsymbol{y_2} + 2H]^+$
458.2395	$C_{23}H_{32}N_5O_5$	-0.3	$C_{12}H_{18}N_4O_2$	600-[AlaAla]
436.2187	$C_{20}H_{30}N_5O_6$	-0.3		
395.2024	$C_{20}H_{27}N_4O_4$	-0.3	$C_{15}H_{27}N_5O_4$	$[\boldsymbol{y_1} + 2H]^+$
387.2022	$C_{20}H_{27}N_4O_4$	-0.5	$C_{15}H_{23}N_5O_3$	600-[AlaAlaAla]
367.1762	$C_{20}H_{23}N_4O_3$	-0.2		
365.1816	$C_{17}H_{25}N_4O_5$	-0.3		
347.1712	$C_{17}H_{23}N_4O_4$	-0.2		
342.2135	$C_{15}H_{28}N_5O_4$	-0.3		
294.1447	$C_{14}H_{20}N_3O_4$	-0.1		
276.1341	$C_{14}H_{18}N_3O_3$	-0.1		
271.1765	$C_{12}H_{23}N_4O_3$	0.0		
254.1983	$C_{14}H_{26}N_2O_2$	-0.6		
241.0970	$C_{14}H_{13}N_2O_2$	-0.1		
231.1127	$C_{13}H_{15}N_2O_2$	-0.1		
221.0708	$C_{14}H_9N_2O$	-0.1		

213.1021	$C_{13}H_{13}N_2O$	-0.1		

Table 7.3 Accurate Mass Measurements for (ε-$(CH_3)_3$KAAA-*tet*-K + H)$^+$ Ions.

m/z	ion formula	error (mmu)	neutral loss	ion assignment
CID-MS^2				
778.43378	$C_{38}H_{56}N_{11}O_7$	-2.1		
750.42780	$C_{38}H_{56}N_9O_7$	-1.9	N_2	
719.36096	$C_{35}H_{47}N_{10}O_7$	-1.4	C_3H_9N	$–(CH_3)_3N$
691.35456	$C_{35}H_{47}N_{10}O_7$	-1.7	$C_3H_9N_3$	$–(N_2 + (CH_3)_3N)$
UVPD-CID-MS^3				
750.42780	$C_{38}H_{56}N_9O_7$	-1.9		
580.28698	$C_{29}H_{38}N_7O_6$	-0.8	$C_9H_{18}N_2O$	$[\boldsymbol{y_4} + 2H]^+$
509.24983	$C_{26}H_{33}N_6O_5$	-0.9	$C_{12}H_{21}N_3O_2$	$[\boldsymbol{y_3} + 2H]^+$
438.21259	$C_{23}H_{28}N_5O_4$	-1.0	$C_{15}H_{26}N_4O_3$	$[\boldsymbol{y_2} + 2H]^+$
367.17574	$C_{20}H_{23}N_4O_3$	-0.7	$C_{18}H_{33}N_5O_4$	$[\boldsymbol{y_1} + 2H]^+$
325.18656	$C_{15}H_{25}N_4O_4$	-0.5	$C_{23}H_{31}N_7O_3$	$\boldsymbol{b_4} – (CH_3)_3N$
254.14960	$C_{12}H_{20}N_3O_3$	-0.3	$C_{26}H_{36}N_8O_4$	$\boldsymbol{b_3} – (CH_3)_3N$

Table 7.4 Accurate Mass Measurements (HAAA-*tet*-K + H)$^+$ Ions.

m/z	ion formula	error (mu)	neutral loss	ion assignment
745.35184	$C_{35}H_{45}N_{12}O_7$	-1.0		
717.34580	$C_{35}H_{45}N_{10}O_7$	-0.9	N_2	
609.27710	$C_{29}H_{37}N_8O_7$	-0.9	$C_6H_8N_4$	-phenylhydrazine
608.28201	$C_{30}H_{38}N_7O_7$	-0.7	$C_5H_7N_5$	-His
591.26676	$C_{29}H_{35}N_8O_6$	-0.6	$C_6H_{10}N_4O$	609-H_2O
538.24040	$C_{26}H_{32}N_7O_6$	-0.5	$C_9H_{13}N_5O$	609-Ala
537.25644	$C_{26}H_{33}N_8O_5$	-0.4	$C_9H_{12}N_4O_2$	608-Ala
472.21849	$C_{23}H_{30}N_5O_6$	-0.6	$C_{12}H_{15}N_7O$	609-His
467.20328	$C_{23}H_{27}N_6O_5$	-0.5	$C_{12}H_{18}N_6O_2$	538-Ala
466.21935	$C_{23}H_{28}N_7O_4$	-0.4	$C_{12}H_{17}N_5O_3$	537-Ala
395.18239	$C_{20}H_{23}N_6O_3$	-0.2	$C_{15}H_{22}N_6O_3$	466-Ala
360.22382	$C_{15}H_{30}N_5O_5$	-0.3	$C_{20}H_{15}N_7O_2$	
351.17722	$C_{15}H_{23}N_6O_4$	-0.3	$C_{20}H_{22}N_6O_3$	
347.17107	$C_{17}H_{23}N_4O_4$	-0.3	$C_{18}H_{22}N_8O_3$	
262.12976	$C_{12}H_{16}N_5O_2$	-0.1	$C_{23}H_{29}N_7O_5$	
238.08473	$C_{13}H_{10}N_4O$	-0.2	$C_{22}H_{35}N_8O_6$	
221.07077	$C_{14}H_9N_2O$	-0.1	$C_{21}H_{36}N_{10}O_6$	

213.10209	$C_{13}H_{13}N_2O$	-0.1	$C_{22}H_{32}N_{10}O_6$
209.10314	$C_9H_{13}N_4O_2$	-0.2	$C_{26}H_{32}N_8O_5$

7.3 Results and Discussion

Electrospray ionization of KAAA-*tet*-K, RAAA-*tet*-K, and HAAA-*tet*-K chiefly produced singly charged ions (**Figure 7.1**) that were selected by mass and investigated by tandem mass spectrometry. These sequences will be discussed separately, starting with RAAA-*tet*-K.

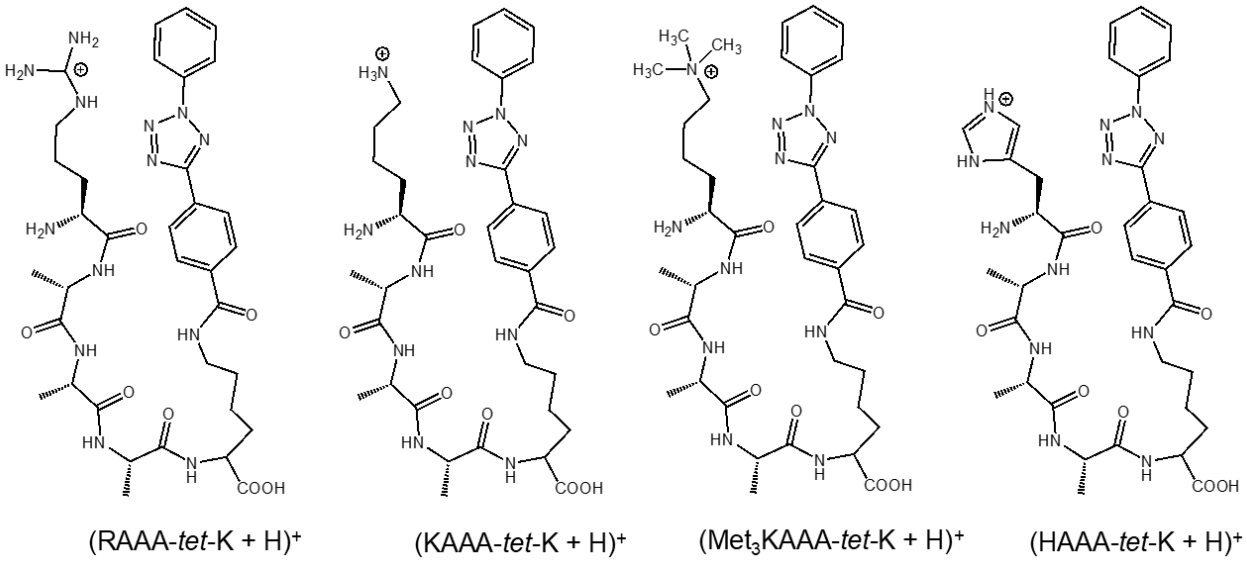


Figure 7.1: Peptide-tetrazole conjugate ions.

7.3.1 RAAA-tet-K

UVPD and CID-MS^2 of the $(RAAA\text{-}tet\text{-}K + H)^+$ ions (m/z 764) resulted in efficient loss of N_2 initially forming the m/z 736 nitrile-imine ion (**Figure 7.2a**, **7.3a**). Both UVPD and CID triggered further dissociations of the $(RAAA\text{-}tet\text{-}K –N_2 + H)^+$ ions that produced two types of fragment ions. One type, which is present in particular upon UVPD, was by loss of C_6H_5N from the nitrile imine group (m/z 645). According to our previous data and interpretation,[6] this fraction was assigned to linear nitrile-imine intermediates. In addition, there was a $C_{14}H_9N_2O$ fragment ion that was

characteristic of the free nitrile imine moiety. The main ion series commenced with the m/z 622 ion that was formed by loss of $C_5H_{10}N_2O$ from the denitrogenated ion. The other ions in this series were at m/z 551, 480, and 409 that sequentially differed by an [Ala] residue. UVPD-MS^2 further displayed a major $C_{15}H_{11}N_4O$ fragment ion in which the nitrile imine contained a covalently linked CH_2N_2 group originating from the Arg side chain. The nature of the $(RAAA\text{-}tet\text{-}K - N_2 + H)^+$ ions was further probed by UVPD-CID-MS^3 (**Figure 7.2b**) and CID-CID-MS^3 (**Figure 7.3b**) that gave the same types of fragment ions. The chief fragment ions, besides the m/z 719 by loss of NH_3, were the m/z 622, 551, 480, and 409 C-terminal sequence ions in which a CH_2N_2 group from the Arg side chain was embedded in the nitrile imine group. We considered ions in this series, which are annotated as $\boldsymbol{v_1}$-$\boldsymbol{v_4}$ to distinguish them from standard $[\boldsymbol{y_i} + 2H]^+$ ions,[29] as evidence for crosslinking by the Arg guanidine group. Notably, the CID-MS^3 spectra displayed only minor ions due to linear, non-crosslinked structures, such as the $[\boldsymbol{y_2} + 2H]^+$ (m/z 438) and $[\boldsymbol{b_4} + H_2O]^+$ (m/z 388) ions. All assignments were corroborated by accurate mass measurements (**Table 7.1**) and CID-MS^n of H/D exchanged $([D_{12}]\text{-}RAAA\text{-}tet\text{-}K + D)^+$ and $([D_{12}]\text{-}RAAA\text{-}tet\text{-}K - N_2 + D)^+$ ions (**Figure 7.4a,b**). Dissociations of $(RAAA\text{-}tet\text{-}K - N_2 + H)^+$ ions were also studied under much harsher energy conditions of 213 nm UVPD-MS^3 (**Figure 7.3c**). In addition to the formation of the $\boldsymbol{v_1}$-$\boldsymbol{v_4}$ ions, UVPD resulted in deep dissociation forming the $C_{15}H_{11}N_4O$ ion (m/z 263) and a number of low-mass fragment ions. Overall, the information obtained from CID-UVPD was equivalent from that provided by UVPD-CID and CID-CID. The $\boldsymbol{v_4}$ sequence ion (m/z 622) that was indicative of arginine crosslinking was further studied by CID-MS^4. The spectrum (**Figure 7.2c**) showed dominant $\boldsymbol{v_2}$ ions by loss of [AlaAla]. We presumed that the facile loss of [AlaAla] was enabled by the formation of a cyclic 3,6-dimethyl-2,5-diketopiperazine neutral fragment that can arise from the combination of Ala2-Ala3 or Ala3-Ala4 residues.

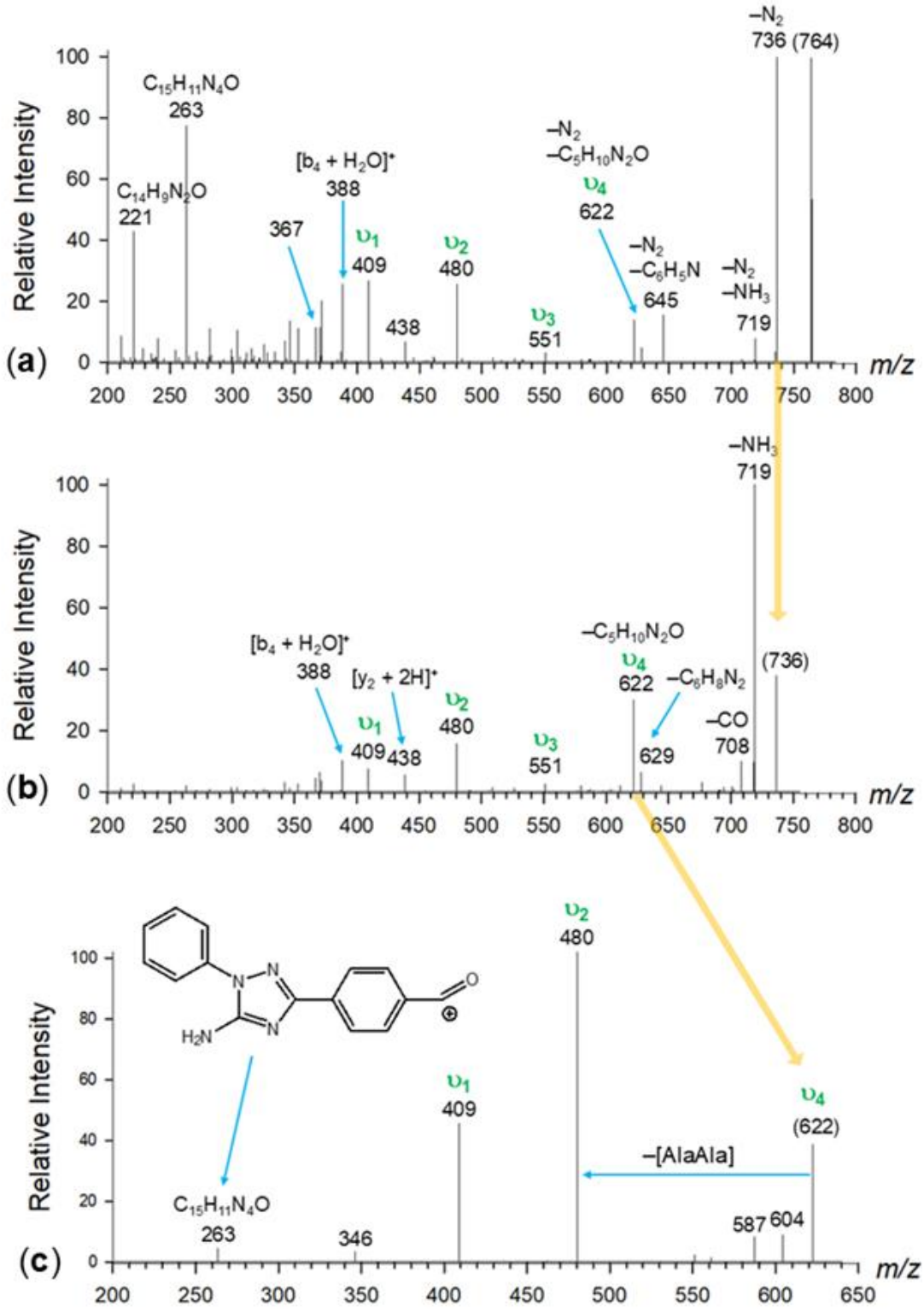


Figure 7.2: (a) UVPD-MS^2 of (RAAA-*tet*-K + H)$^+$ (*m/z* 764); (b) UVPD-CID-MS^3 of (RAAA-*tet*-K $-N_2$ + H)$^+$ (*m/z* 736); (c) CID-MS^4 of the υ_4 ions (*m/z* 622) from the Figure 2b spectrum.

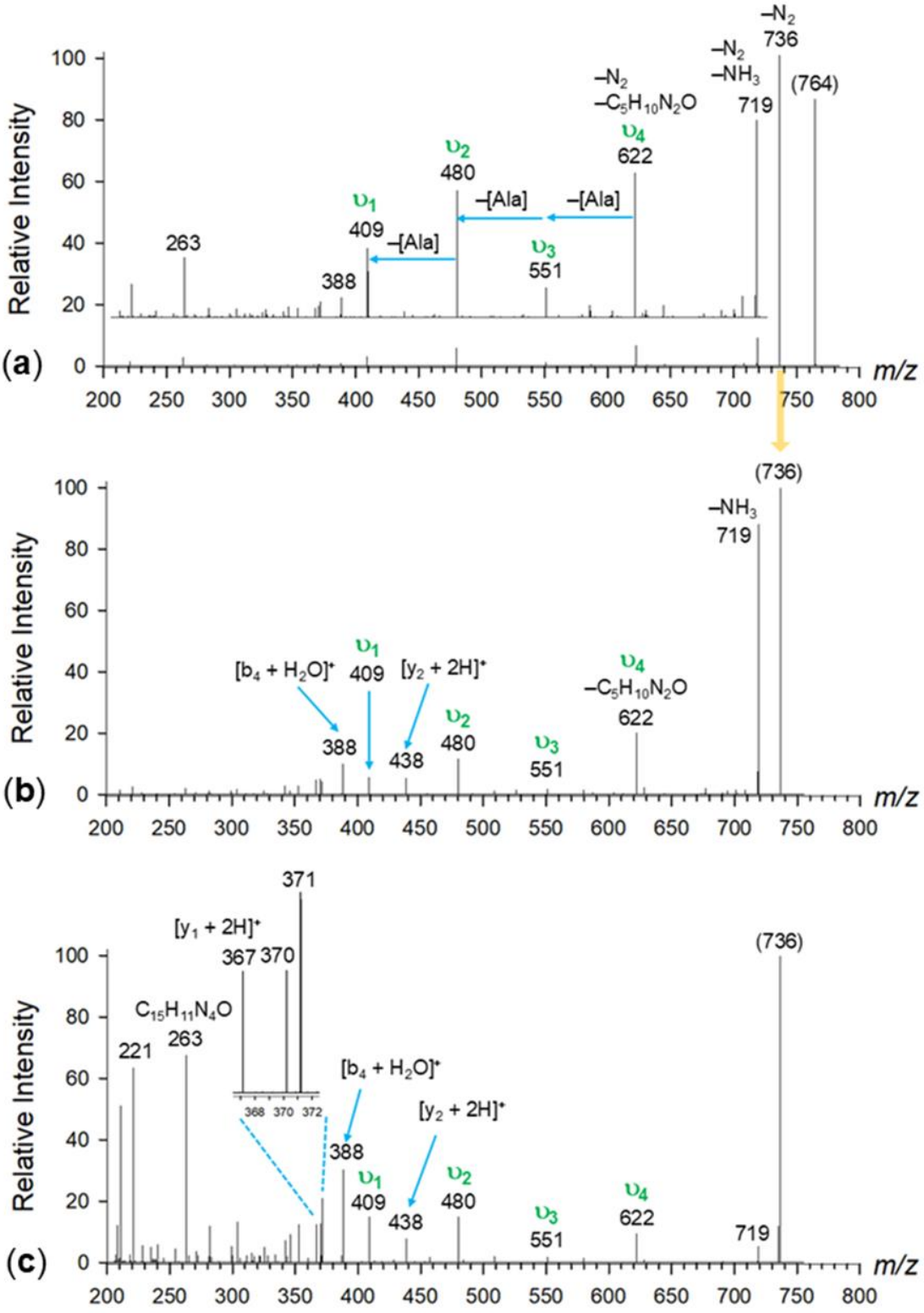


Figure 7.3: (a) CID-MS^2 of (RAAA-*tet*-K + H)$^+$ (*m/z* 764); (b) CID-CID-MS^3 of (RAAA-*tet*-K –N_2 + H)$^+$ (*m/z* 736); (c) CID-UVPD-MS^3 of (RAAA-*tet*-K –N_2 + H)$^+$ (*m/z* 736).

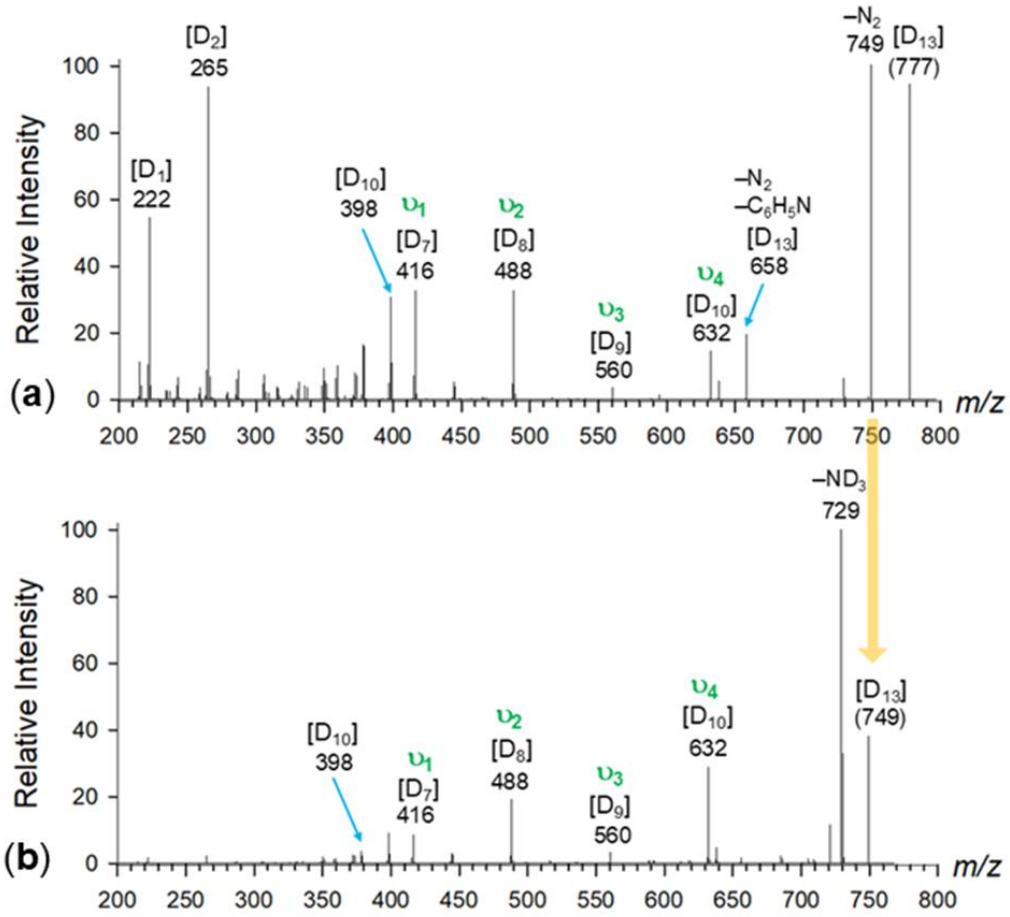
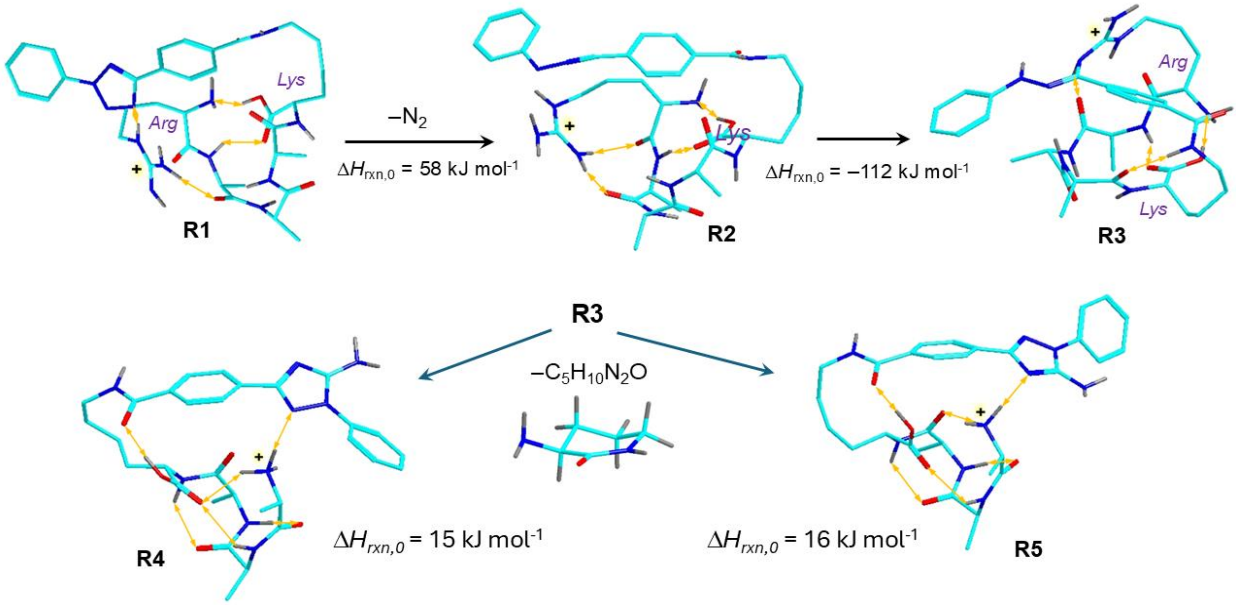


Figure 7.4: (a) CID-MS^2 of ([D_{12}]-RAAA-*tet*-K + D)$^+$ (*m/z* 777); (b) CID-CID-MS^3 of ([D_{12}]-RAAA-*tet*-K –N_2 + D)$^+$ (*m/z* 749).

To elucidate the structure and formation of the **υ4** sequence ions, we studied the dissociation energetics using DFT calculations following BOMD conformational analysis of the reactants and products. Loss of N_2 from the lowest-energy (RAAA-tet-K + H)$^+$ conformer (**R1**) was 58 kJ mol^{-1} endothermic, forming nitrile-imine intermediate **R2** which was the lowest-energy conformer among these ions. Crosslinking by the Arg guanidine group in R2 was substantially exothermic, $\Delta H_{rxn,0}$ = –112 kJ mol^{-1}, forming ion **R3** (**Scheme 7.2**). Further loss of the $C_5H_{10}N_2O$ molecule from **R3**, as indicated by CID-MS^3, had very low threshold energies of 15-16 kJ mol^{-1}, producing conformers **R4** and **R5**, respectively. This was established for a cyclic lactam structure for the Arg $C_5H_{10}N_2O$ fragment. In addition, the H_2N-C=N group that was transferred to the nitrile imine in the course of crosslinking was found to spontaneously cyclize, forming a 1,2,4-triazole ring in the virtually isoenergetic conformers of the **υ4** ion **(Scheme 7.2**). The intermediate reaction steps

leading to the υ_4 ion and $C_5H_{10}N_2O$ fragment formation, such as the Arg-Ala peptide bond and guanidine N–C bond cleavages, and accompanying proton migrations, likely had energy barriers exceeding the threshold energies that we did not address by calculations. Nevertheless, the very low threshold energies were consistent with the prominent $\boldsymbol{\upsilon_4}$ ion formation.



Scheme 7.2: Structures and Energies for the Formation of Nitrile Imines and $C_5H_{10}N_2O$ Loss from Arg-Crosslinked Ions.[a]

[a]Atom color coding is as follows: cyan = C, blue = N, red = O, gray = H. Only exchangeable N-H and O-H hydrogens are displayed to avoid clutter. Hydrogen bonds are annotated by ochre double-headed arrows. The relative Gibbs free energies are from M06-2X/def2qvzpp calculations and include B3LYP/6-31+G(d,p) zero-point energies.

To further analyze the $(RAAA\text{-}tet\text{-}K + H)^+$ and $(RAAA\text{-}tet\text{-}K –N_2 + H)^+$ ions we used cyclic ion mobility to separate isomers and obtain collision cross sections (CCS_{exp}).

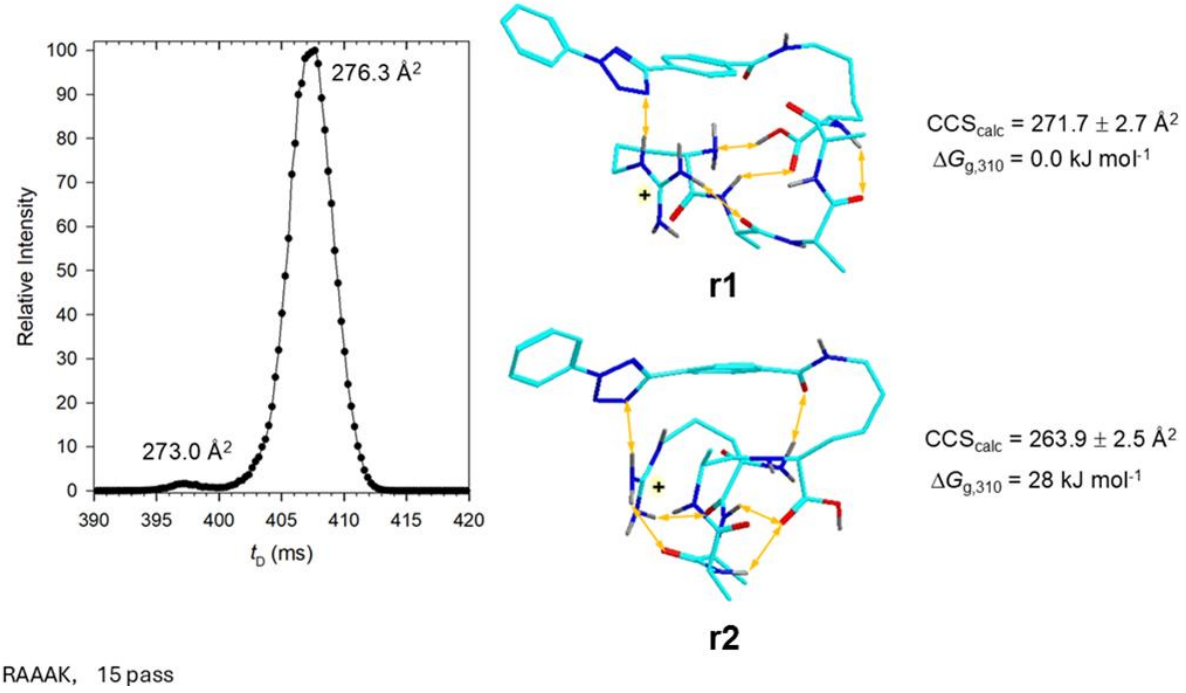


Figure 7.5: Left panel: Arrival time distribution of (RAAA-*tet*-K + H)$^+$ ions after 15 cycles (1.47 m pathlength). Right panel: M06-2X/6-31+G(d,p) optimized structures of lowest energy (RAAA-*tet*-K + H)$^+$ ions. Atom color coding is as follows: cyan = C, blue = N, red = O, gray = H. Only exchangeable N-H and O-H hydrogens are displayed to avoid clutter. Hydrogen bonds are annotated by ochre double-headed arrows. The relative Gibbs free energies are from M06-2X/def2qvzpp calculations and include B3LYP/6-31+G(d,p) zero-point energies, enthalpies and entropies.

Ion mobility showed a major isomer (94%) with CCS$_{exp}$ = 276.3 Å^2 that was accompanied by a more compact minor isomer with CCS$_{exp}$ = 273.0 Å^2 (**Figure 7.5**, left panel). According to the calculations, the lowest energy structure **r1** had a CCS$_{calc}$ = 271.7 ± 2.7 Å^2. A more compact but substantially less stable isomer (**r2**) had CCS$_{calc}$ = 263.9 ± 2.5 Å^2. The agreement between the experimental and calculated CCS for **r1** (Δ_{rel} = 1.7%), albeit still remarkably good, showed a slightly larger difference than those reported for other peptide ions.[6-8] The minor component remained unidentified and may correspond to a higher-energy conformer. The lowest-energy structure **r1** showed a tightly folded peptide chain that was supported by multiple hydrogen bonds. Those included guanidinium group H-bonding to the Ala2 amide and the tetrazole ring (**Figure 7.5**). The conformer near-homogeneity of the (RAAA-tet-K + H)$^+$ ions, as displayed by the arrival time distribution, was consistent with the relatively large gap in Gibbs free energies (28 kJ mol^{-1}) separating **r1** from **r2** and other higher energy isomers.

The $(RAAA\text{-}tet\text{-}K - N_2 + H)^+$ ions used for ion mobility measurements were generated by CID-MS^2. Owing to the similarity of the CID-MS^3 spectra of $(RAAA\text{-}tet\text{-}K - N_2 + H)^+$ made by CID and UVPD (**Figure 7.2b** and **7.3b**, respectively), we presume that the formation and composition of isomers were similar in both activation modes. The arrival time distribution of $(RAAA\text{-}tet\text{-}K - N_2 + H)^+$ ions, obtained after 5 cycles (490 cm pathlength), showed multiple components (**Figure 7.6**, left panel).

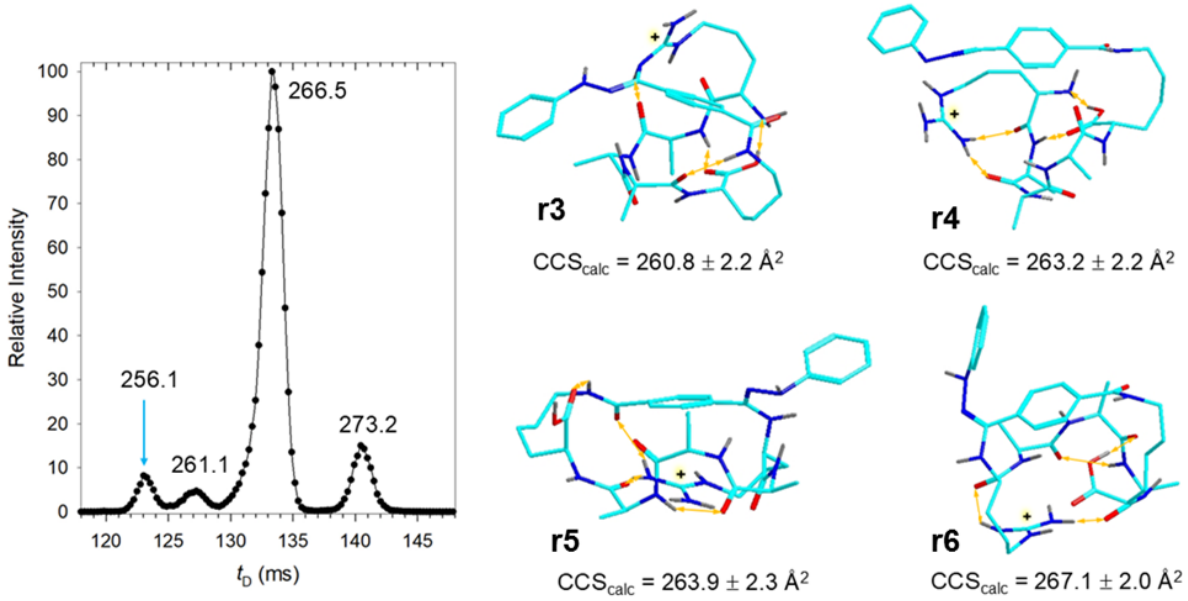


Figure 7.6: Left panel: Arrival time distribution of $(RAAA\text{-}\textit{tet}\text{-}K - N_2 + H)^+$ ions after 5 cycles (490 cm pathlength). Right panel: M06-2X/6-31+G(d,p) optimized structures of representative lowest energy $(RAAA\text{-}\textit{tet}\text{-}K - N_2 + H)^+$ ions. Atom color coding is as in Figure 7.5.

The dominant component had CCS_{exp} = 266.5 Å^2 that was flanked by more compact minor components of CCS_{exp} = 256.1 Å^2 and 261.1 Å^2 and a more expanded component of CCS_{exp} = 273.2 Å^2. To assign plausible structures for the $(RAAA\text{-}tet\text{-}K - N_2 + H)^+$ ions, we considered linear nitrile-imine conformers, as well as crosslinks by the Arg side-chain guanidine and N-terminal amine groups. This selection was guided by the dissociations observed in the UVPD-MS^2 and CID-MS^3 spectra, as discussed above. When considering low-energy conformers in each group, the CCS_{calc} of 17 isomers ranged from 255.9 Å^2 through 267.1 Å^2 which covered the range of CCS_{exp}. Individual peak assignment was less specific because of the multitude of structures and

the uncertainty in matching CCS_{calc} with CCS_{exp} that can be estimated as within ca. 2% or 5 Å^2 (see above). The lowest energy structures in each group are depicted in **Figure 7.6**. The Arg-crosslinked ion **r3** had CCS_{calc} = 260.8 Å^2 which was on the low side of the major peak in the mobilogram. Ion **r3** had a compact structure that was secured by the Arg crosslink and included proton transfer onto the nitrile imine in addition to cyclization. The nitrile-imine isomer **r4** of CCS_{calc} = 263.2 Å^2 made it closer to the major peak in the mobilogram. Two other isomers, **r5** and **r6**, were crosslinked by the N-terminal amine group. Their CCS_{calc} = 263.9 and 267.1 Å^2 for **r5** and **r6**, respectively, also fell within the experimental range and were close to the CCS_{exp} of the main peak. We note that ion **r5** was the global energy minimum of the entire set at ΔG_{310} = –30 kJ mol^{-1} relative to **r3**. Nitrile imine **r4** was at ΔG_{310} = 106 kJ mol^{-1} relative to **r3**, indicating the high exothermicity of crosslink formation. Two items are noteworthy in this respect. First, the relative energies of crosslinks of a different type, such as **r3** and **r5**, are not indicative of their competitive formation from nitrile imines. This is because the crosslinking reactions were determined by the pertinent activation energies and, once formed, the isomers were unlikely to interconvert. Second, the match between the hypothetical CCS_{calc} and measured CCS_{exp} does not guarantee structure assignment. This was particularly evident from the CID-MS^3 spectra that did not indicate the presence of N-terminally crosslinked structures, for which a loss of [Ala] would be expected upon CID.

7.3.2 KAAA-tet-K

Both UVPD and CID of (KAAA-tet-K + H)$^+$ (m/z 736) resulted in an efficient loss of N_2 (**Figure 7.7a**, **Figure 7.8**, respectively). Higher excitation upon UVPD then led to consecutive dissociations of the denitrogenated m/z 708 ions; only a few of these dissociations were indicative of linear nitrile-imine structures, e.g. the m/z 617 by loss of C_6H_5N, and m/z 438 that according to its elemental composition was a denitrogenated [$\boldsymbol{y_2}$ + 2H]$^+$ fragment ion. The chief dissociation channel was loss of $C_6H_8N_2$ that, according to deuterium labeling in ([D_{10}]-KAAA-tet-K + D)$^+$ (m/z 747, **Figure 7.9**), contained three exchangeable protons and was identified as phenylhydrazine. This primary dissociation was followed by internal backbone cleavages expelling neutral [Ala] fragments and forming the m/z 529, 458, and 367 ions (**Figure 7.7a**). Loss of phenyl hydrazine was also the predominant channel in CID-MS^3 of the m/z 708 ions forming

the m/z 600 ion (**Figure 7.7b**). This dissociation required transfer of three hydrogen atoms onto the nitrile imine group and must have been associated with crosslinking to the remaining nitrile carbon atom to avoid a high energy methine structure in the fragment ion. This was consistent with the internal backbone cleavages and loss of [Ala]. These were particularly evident in the UVPD-MS^3 spectrum of the m/z 708 ion (**Figure 7.10**) that also displayed fragment ions resulting from deep secondary dissociations that were identified by their elemental composition (**Table 7.2**). The nature of the crosslinks was further elucidated by CID-MS^4 of the m/z 600 ion **(Figure 7.7c**). This showed losses of [Ala] and [AlaAla], m/z 529 and 458, respectively, indicating crosslinked structures. In addition, we identified a series of fragment ions separated by [Ala] at m/z 294, 365, and 436 that provided a clue to the crosslink bond connectivity. The m/z 294 ion, $C_{14}H_{20}N_3O_4$, was composed of the lysine residue that was attached to the benzoyl moiety containing a $CONH_2$ group that was not originally present in the nitrile imine. This group resulted from a combined or consecutive transfer of O and NH_2 unto the nitrile-imine carbon atom in the course of crosslink formation.

A possible reaction sequence for this transformation is sketched in **Scheme 7.3**, showing proton transfer to the nitrile-imine nitrogens accompanied by the formation of the ε-N–C and O–C bonds to the nitrile imine carbon atom. This produces an intermediate with a weak N–C bond that is prone for phenyl hydrazine elimination which is accompanied by proton transfer. The m/z 600 ion, bottom structure in **Scheme 7.3**, can break any of the backbone amide bonds and eliminate CO, [Ala], or [AlaAla], which are the main dissociations. In addition, cleavage of the lysine C–ε-N bond, possibly assisted by N-terminal NH_2 participation, can open the macrocyclic ring and enable the formation of modified y-type ions that we denote as $\boldsymbol{v_1}$-$\boldsymbol{v_3}$ at m/z 294, 365, and 436, respectively, to distinguish them from standard $[\boldsymbol{y_i} + 2H]^+$ ions. Notably, the absence of the lysine carbonyl prevents the formation of the $\boldsymbol{v_4}$ ion that was absent in the **Figure 7.7c** spectrum. It should be noted that the spectra alone did not allow us to distinguish between structures crosslinked by the side chain and N-terminal amine groups, and additional calculations and experiments were needed to resolve this issue.

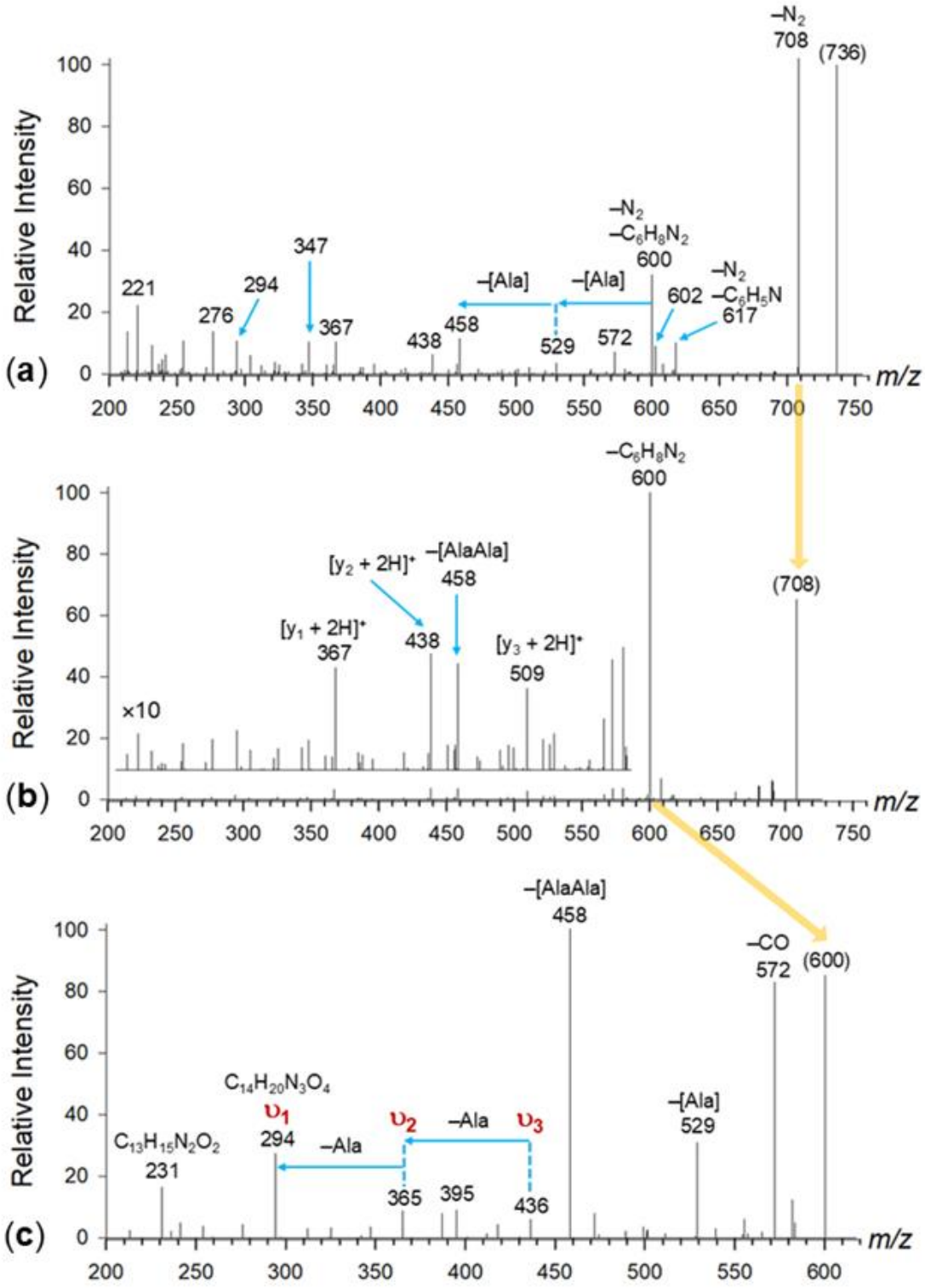


Figure 7.7: (a) UVPD-MS^2 of (KAAA-*tet*-K + H)$^+$ (*m/z* 736); (b) UVPD-CID-MS^3 of (KAAA-*tet*-K –N_2 + H)$^+$ (*m/z* 708); (c) CID-MS^4 of the *m/z* 600 ion from the Figure 7.7b spectrum.

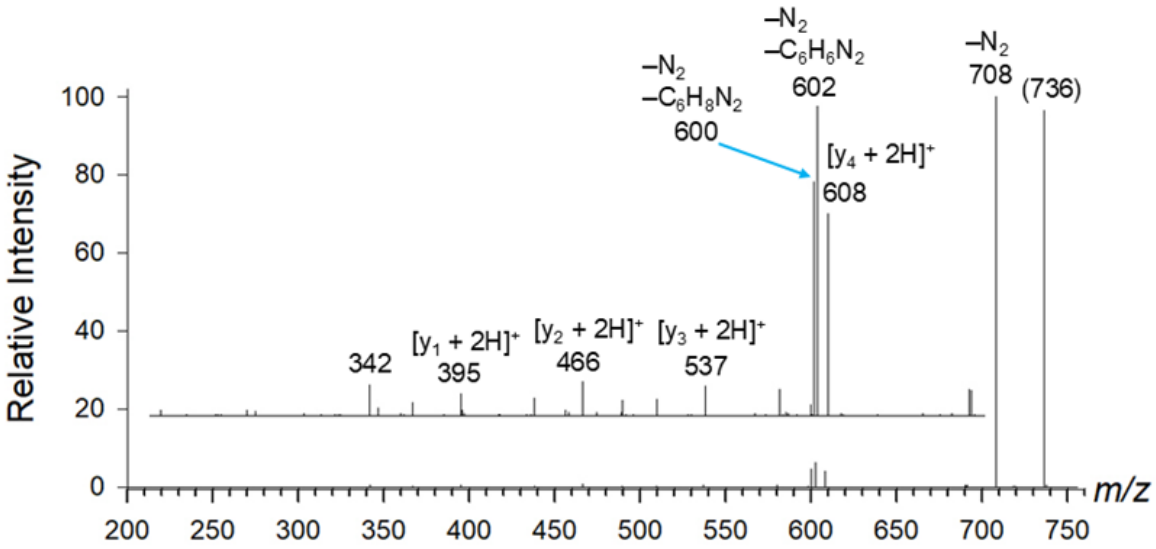


Figure 7.8: CID-MS^2 of (KAAA-*tet*-K + H)$^+$, *m/z* 736.

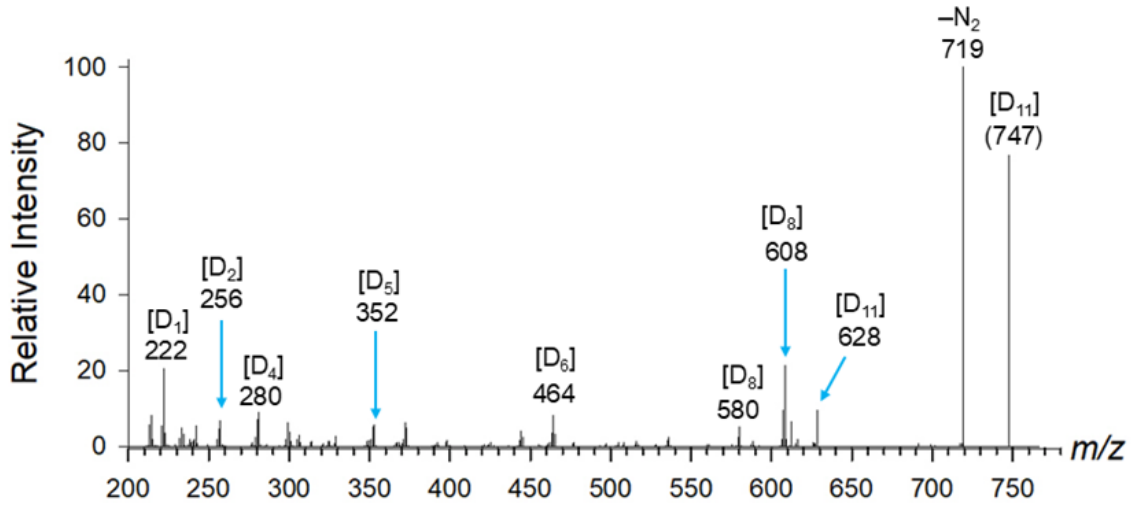


Figure 7.9: UVPD-MS^2 of fully H/D exchanged ([D$_{10}$]-KAAA-*tet*-K + D)$^+$, *m/z* 747.

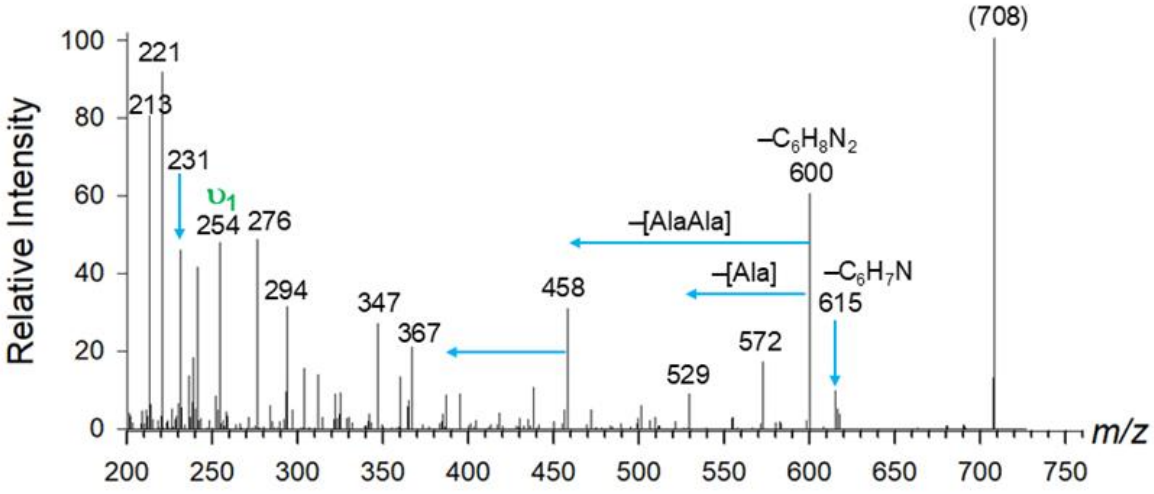
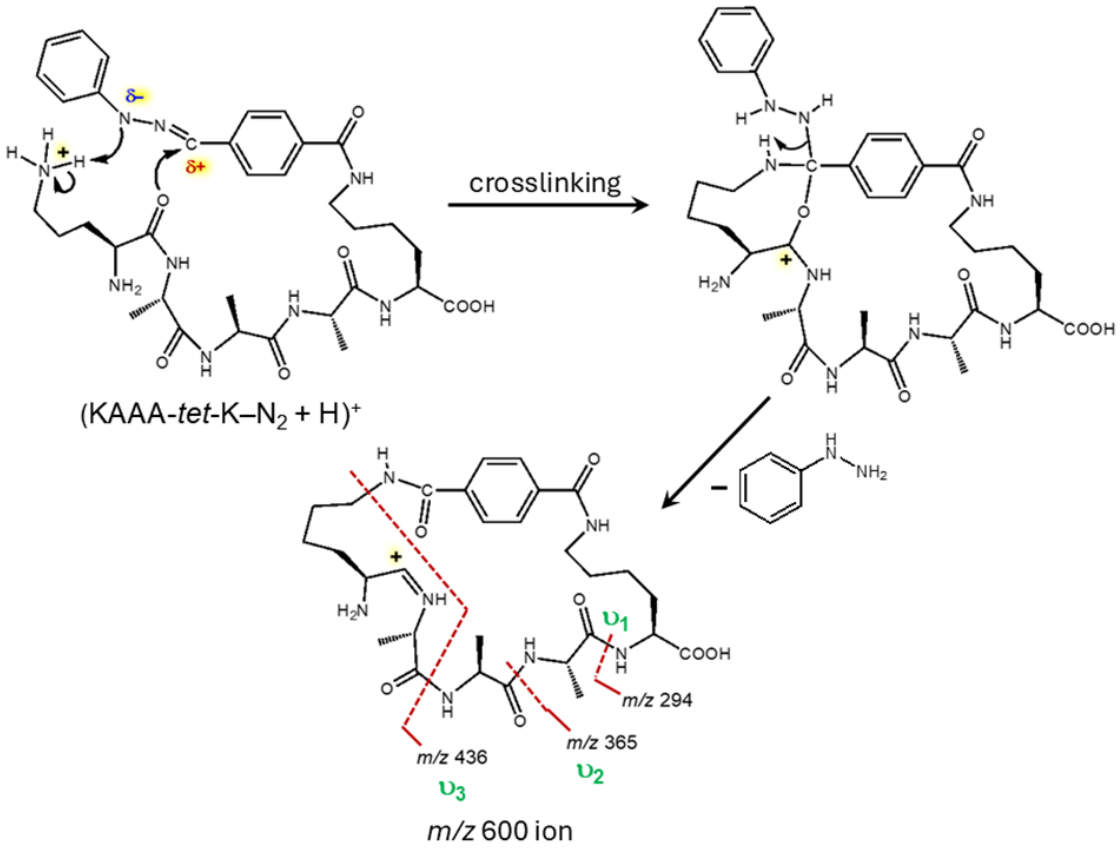
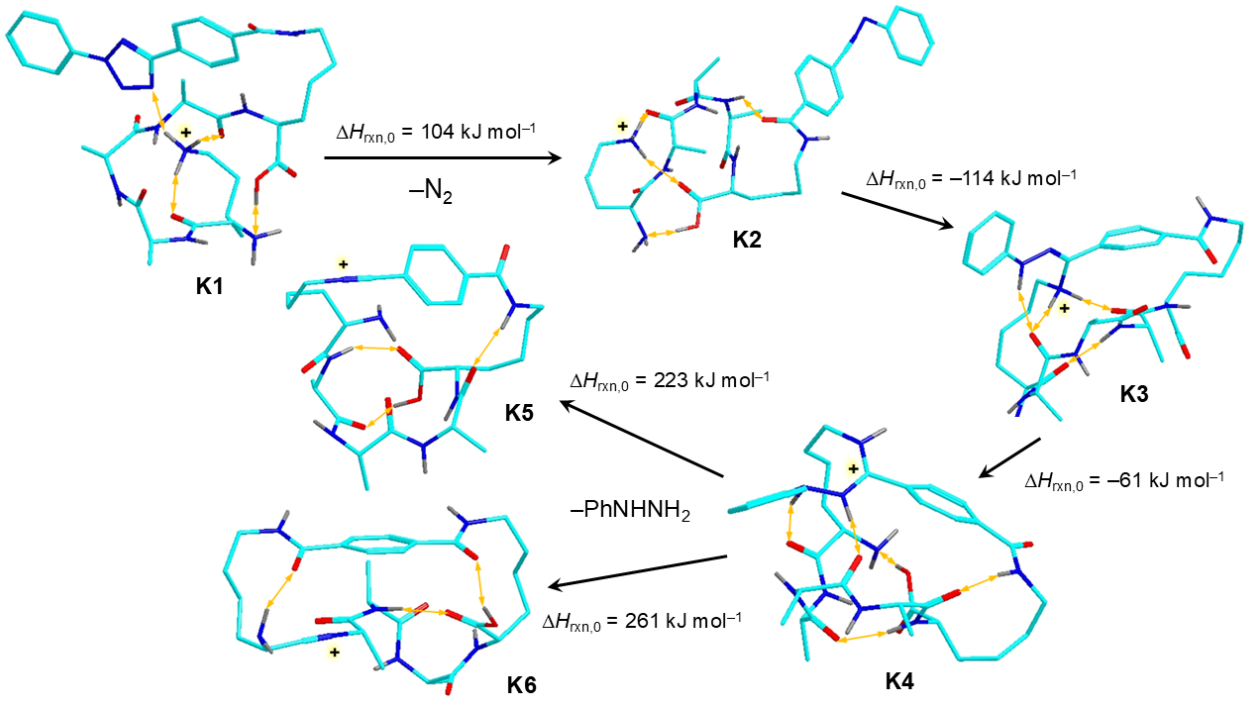


Figure 7.10: CID-UVPD-MS3 of (KAAA-tet-K –N2 + H)+, m/z 708.



Scheme 7.3: Crosslinking and Backbone Dissociations in (KAAA-*tet*-K $-N_2$ + H)$^+$.



Scheme 7.4: Structures and M06-2X/def2qzvpp Relative Energies of KAAAK-*tet*-K Ions.

We mapped the relevant parts of the KAAAK-tet-K ion potential energy surface to determine the dissociation and isomerization energies **(Scheme 7.4**). Starting from the lowest-energy $(KAAA\text{-}tet\text{-}K + H)^+$ conformer (**K1**), loss of N_2 was calculated to require 104 kJ mol^{-1} to form the lowest-energy nitrile imine **K2**. Ring closure in **K2** involving the Lys-ε-amine group was 114 kJ mol^{-1} exothermic, forming crosslink **K3** that could further isomerize by exothermic proton migration to the most stable isomer **K4**. We found two pathways for further dissociation of **K4** by loss of phenylhydrazine. In the first one, proton migration would weaken the $PhNHNH_2$–C bond, resulting in its dissociation forming ion **K5** that was 223 kJ mol^{-1} endothermic relative to **K4**. Ion **K5** was presumed to dissociate by [Ala] and [AlaAla] internal cleavages upon CID (**Figure 7.7c**). The other pathway included migration of an amide proton to trigger $PhNHNH_2$ loss with participation by the Lys amide, forming isomer **K6**. This reaction was 261 kJ mol^{-1} endothermic. Ion **K6** was a natural precursor for the formation of the $\boldsymbol{v_1}$-$\boldsymbol{v_3}$ fragment ions upon further CID (**Figure 7.7c**).

The role of the lysine ε-NH_2 group and proton transfer was investigated by using a derivative that was trimethylated at the side chain nitrogen, $(\varepsilon\text{-}(CH_3)_3KAAA\text{-}tet\text{-}K)^+$ which was a quaternary ammonium ion, m/z 778. The UVPD and CID-MS^2 spectra showed facile loss of N_2 (m/z 750,

Figure 7.11a,b). However all backbone dissociation upon MS^2 as well as CID-MS^3 gave standard $[\boldsymbol{y_n} + 2H]^+$ fragment ions, m/z 367, 438, 509, and 580 (**Figure 7.11c**). No fragment ions indicating crosslinks were observed. A side-reaction observed for $(\varepsilon\text{-}(CH_3)_3KAAA\text{-tet-K} - N_2 + H)^+$ was the loss of $(CH_3)_3N$, giving rise to the m/z 691 fragment ion. CID-MS^4 of this ion showed a complete series of $[\boldsymbol{y_n} + 2H]^+$ fragment ions in addition to complementary $\boldsymbol{b_n}$ $-(CH_3)_3N$ ions, whereas no crosslinked ions were identified (**Figure 7.11d**). Since lysine trimethylation accomplished the functions of blocking both the nucleophilic attack and proton transfer to the nitrile imine, these effects were not clearly distinguished. However, as reported previously, N-terminal amino acid residues lacking nucleophilic groups did undergo crosslinking to nitrile imines by using the peptide amide group. Since this reaction was absent with $(\varepsilon\text{-}(CH_3)_3KAAA\text{-tet-K})^+$, we tend to conclude that it was the lack of a transferable proton that blocked crosslinking in these gas-phase ions. This conclusion was consistent with previous results with sodiated ions that also did not undergo crosslinking to nitrile imines.[6]

7.3.3 HAAA-tet-K

UVPD and CID-MS^2 of $(HAAA\text{-tet-K} + H)^+$ (m/z 745) revealed a prominent loss of phenylhydrazine (m/z 609) after loss of N_2 (m/z 717) (**Figure 7.12a**, **Figure 7.13**). In addition, UVPD resulted in deep dissociations within both the peptide chain and nitrile imine moiety, producing among others the $C_{14}H_9N_2O$ (m/z 221) and $C_{12}H_{16}N_6O_2$ (m/z 262) fragment ions (**Table 7.4**). Both UVPD and CID displayed consecutive dissociations of the m/z 609 ions by expulsion of internal [Ala] residues (m/z 537, 472) that suggested crosslinked structures. In addition to the loss of $C_6H_8N_2$, the spectra also showed loss of $C_5H_7N_3$ from the His residue giving rise to the m/z 608 ion. This was a characteristic dissociation of N-terminal amino acid residues in ions that were crosslinked to the nitrile imine by the peptide amide bond and rearranged to hydrazone structures.[6] The hydrazone crosslink can account for loss of both

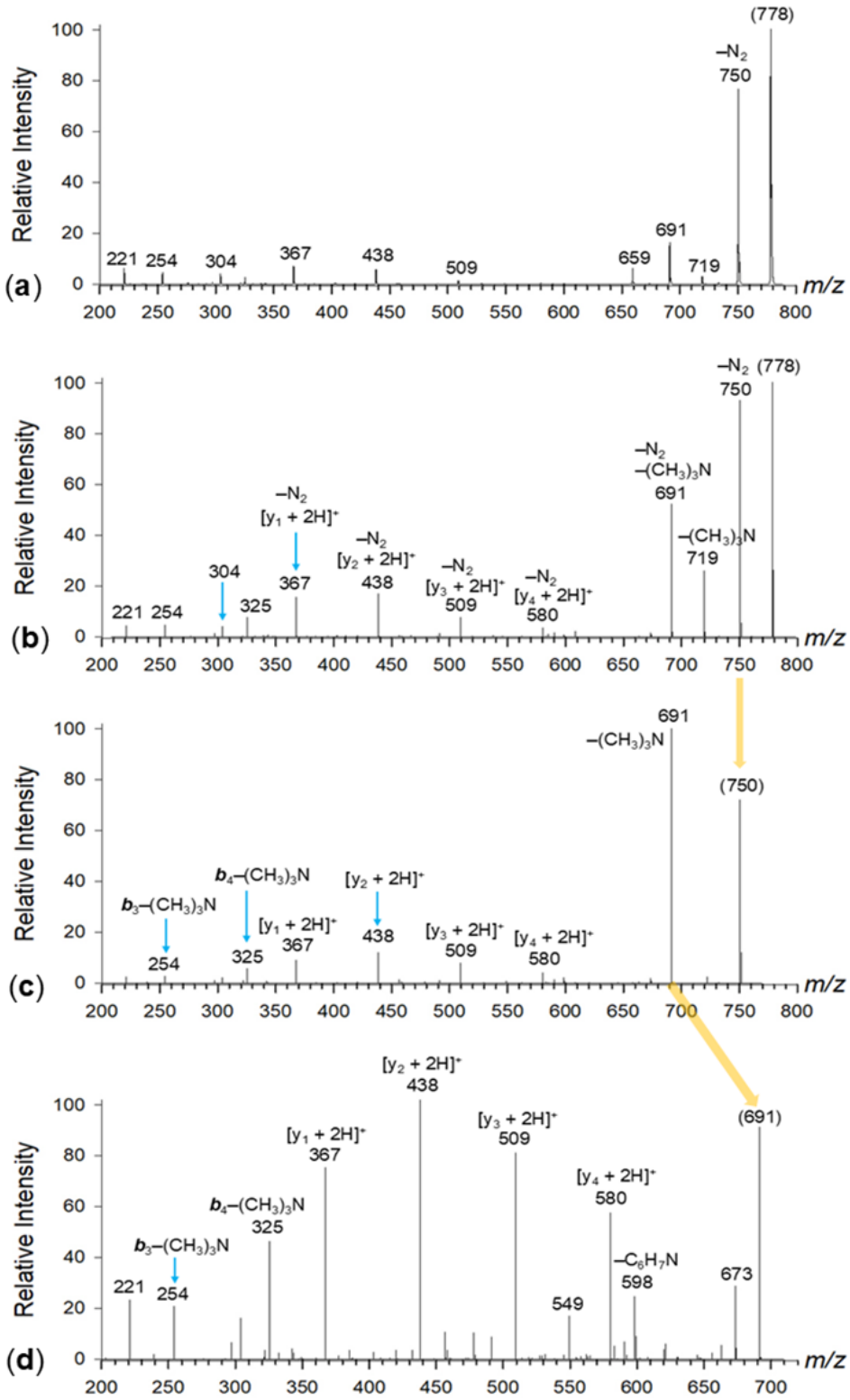
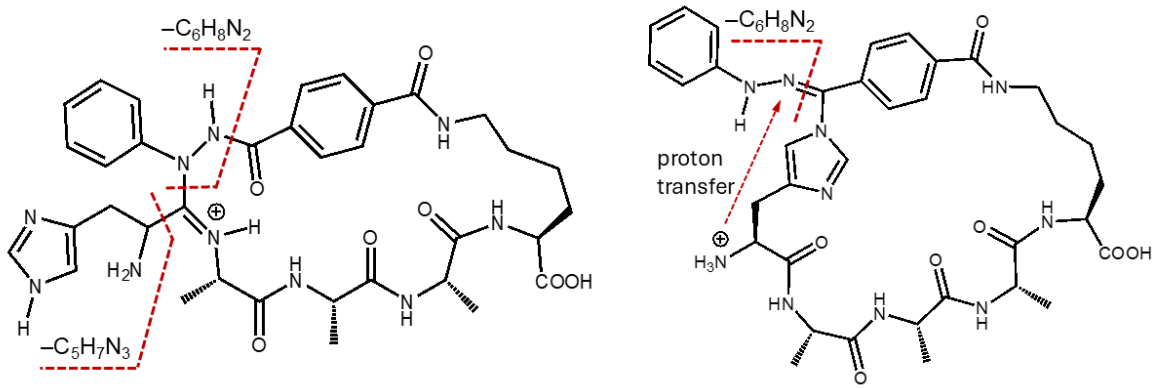


Figure 7.11: (a) CID-MS^2 of (e-$(CH_3)_3$KAAA-*tet*-K + H)$^+$ (*m/z* 778); (b) UVPD-MS^2 of (e-$(CH_3)_3$KAAA-*tet*-K + H)$^+$ (*m/z* 778); (c) UVPD-CID-MS^3 of (e-$(CH_3)_3$KAAA-*tet*-K –N_2 +H)$^+$ (*m/z* 750); (d) CID-MS^4 of (e-$(CH_3)_3$KAAA-*tet*-K –N_2 – $(CH_3)_3$N +H)$^+$ (*m/z* 769).

$C_5H_7N_3$ containing the His side chain, and phenylhydrazine as $C_6H_8N_2$ (**Scheme 7.5**). Both dissociations must be accompanied by proton migrations either from or to the neutral fragment. The ion assignment was corroborated by CID-MS^3 of the $([D_9]HAAAK - N_2 + D)^+$ ion (m/z 727) that showed the fragment ions by loss of $C_6H_5D_3N_2$ and $C_5H_5D_2N_3$ as a partially resolved doublet at m/z 616.3177 and m/z 616.3290, respectively, in the high-resolution mass spectrum (**Figure 7.14a**). This represented transfer of three exchangeable deuterium atoms onto the phenylhydrazine molecule in $C_6H_5D_3N_2$ and transfer of one exchangeable deuterium atom from the His amine group in $C_5H_5D_2N_3$ to the fragment ion. Loss of phenylhydrazine was also prominent in CID-MS^3 of the $(HAAAK - N_2 + H)^+$ ion (m/z 717) producing the m/z 609 fragment ion (**Figure 7.12b**). There was only weak further fragmentation that revealed the presence of crosslinked and linear nitrile-imine structures. Crosslinks were represented by loss of [AlaAla] and [AlaAlaAla] internal residues (m/z 575 and 504, respectively). Linear structures were indicated by the $[\boldsymbol{y_i} + 2H]^+$ ions at m/z 580, 509, 438, and 367 (**Figure 7.12b**). Further analysis of the m/z 609 ion by CID-MS^4 revealed loss of internal [Ala] and [AlaAla] fragments, but not [AlaAlaAla]. This was consistent with the suggested structure of the crosslinked ion where the Ala2 residue was bonded to the ring bridgehead. The side-chain dissociations starting from amide and His crosslinked structures are sketched in **Scheme 7.5** and further investigated computationally.



Scheme 7.5: Diagnostic Side-Chain Dissociations in Crosslinked $(HAAA\text{-}\textit{tet}\text{-}K - N_2 + H)^+$ Ions.

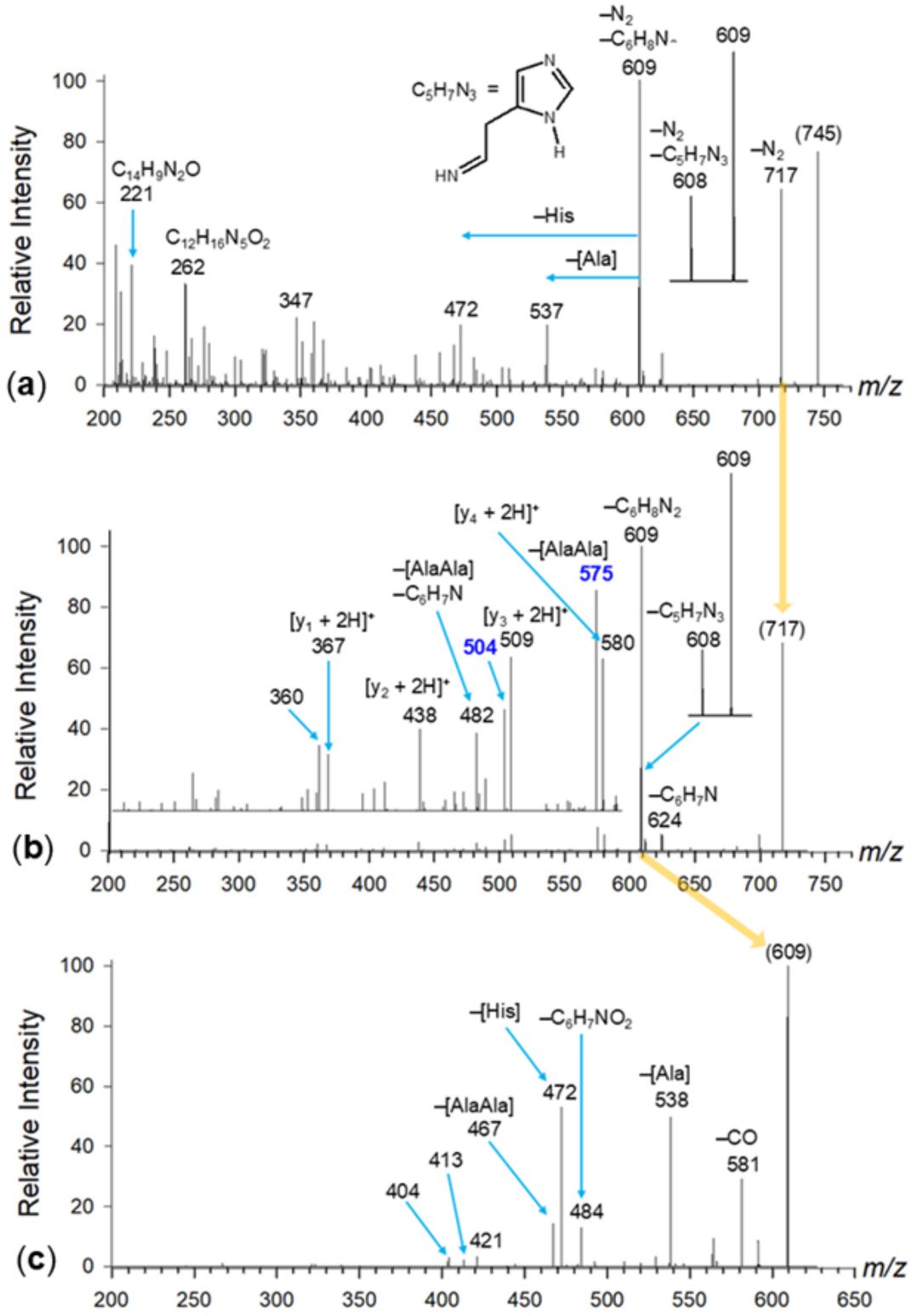


Figure 7.12: (a) UVPD-MS^2 of (HAAA-*tet*-K + H)$^+$ (*m/z* 745); (b) CID-MS^3 of (HAAA-*tet*-K –N_2 + H)$^+$ (*m/z* 717); (c) CID-MS^4 of (HAAA-*tet*-K –N_2 –$C_6H_8N_2$ + H)$^+$ (*m/z* 609).

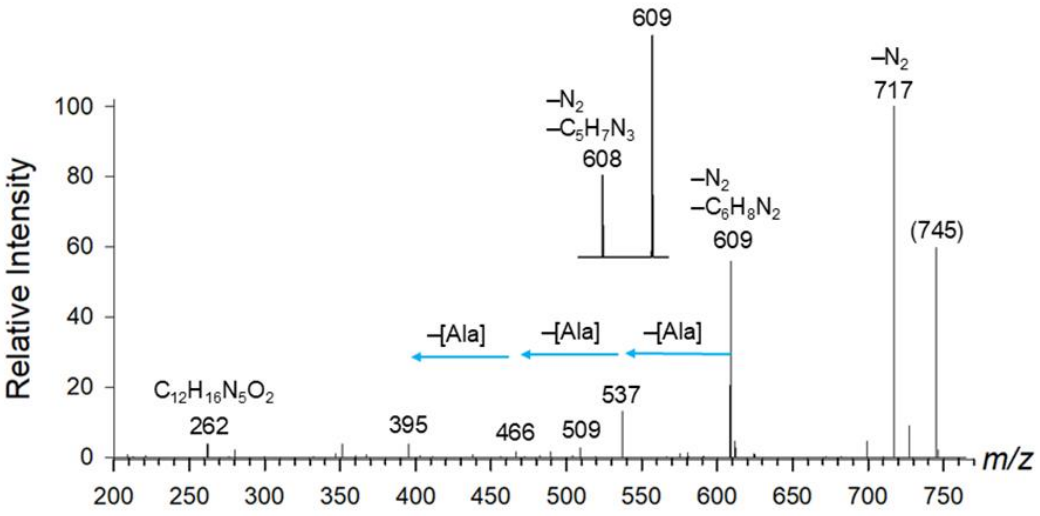


Figure 7.13: CID-MS^2 of (HAAA-*tet*-K + H)$^+$ (*m/z* 745).

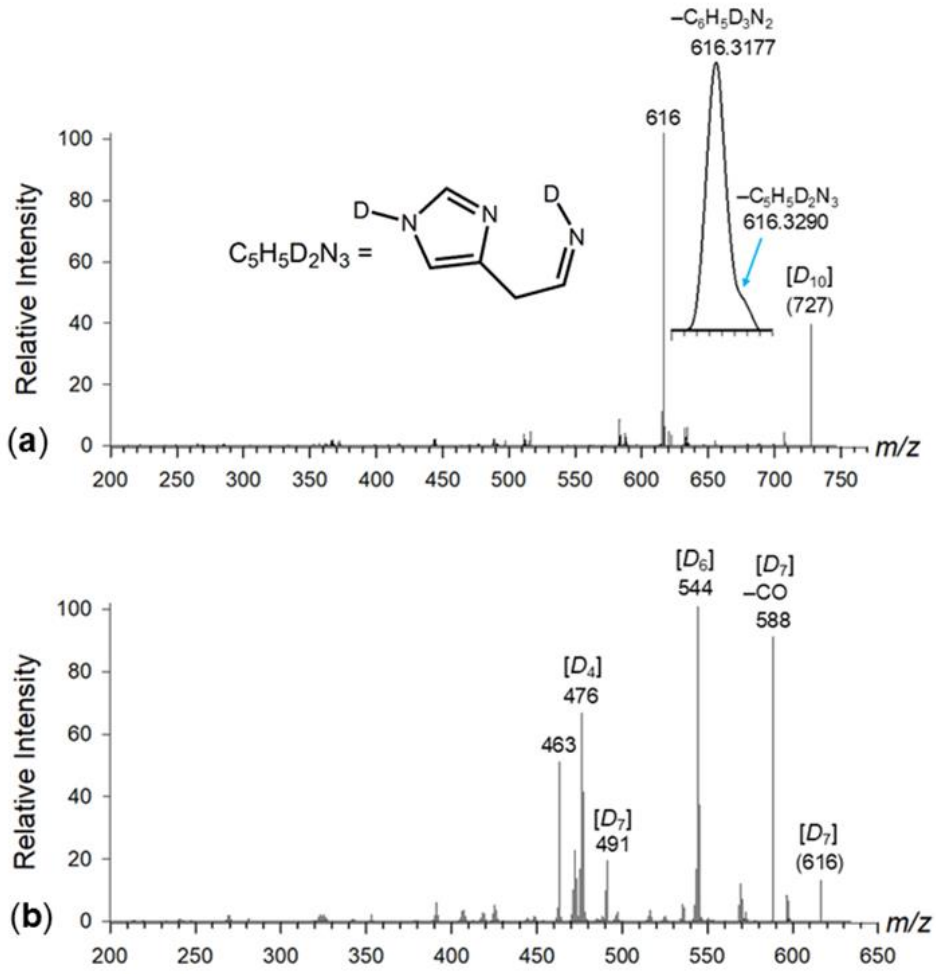
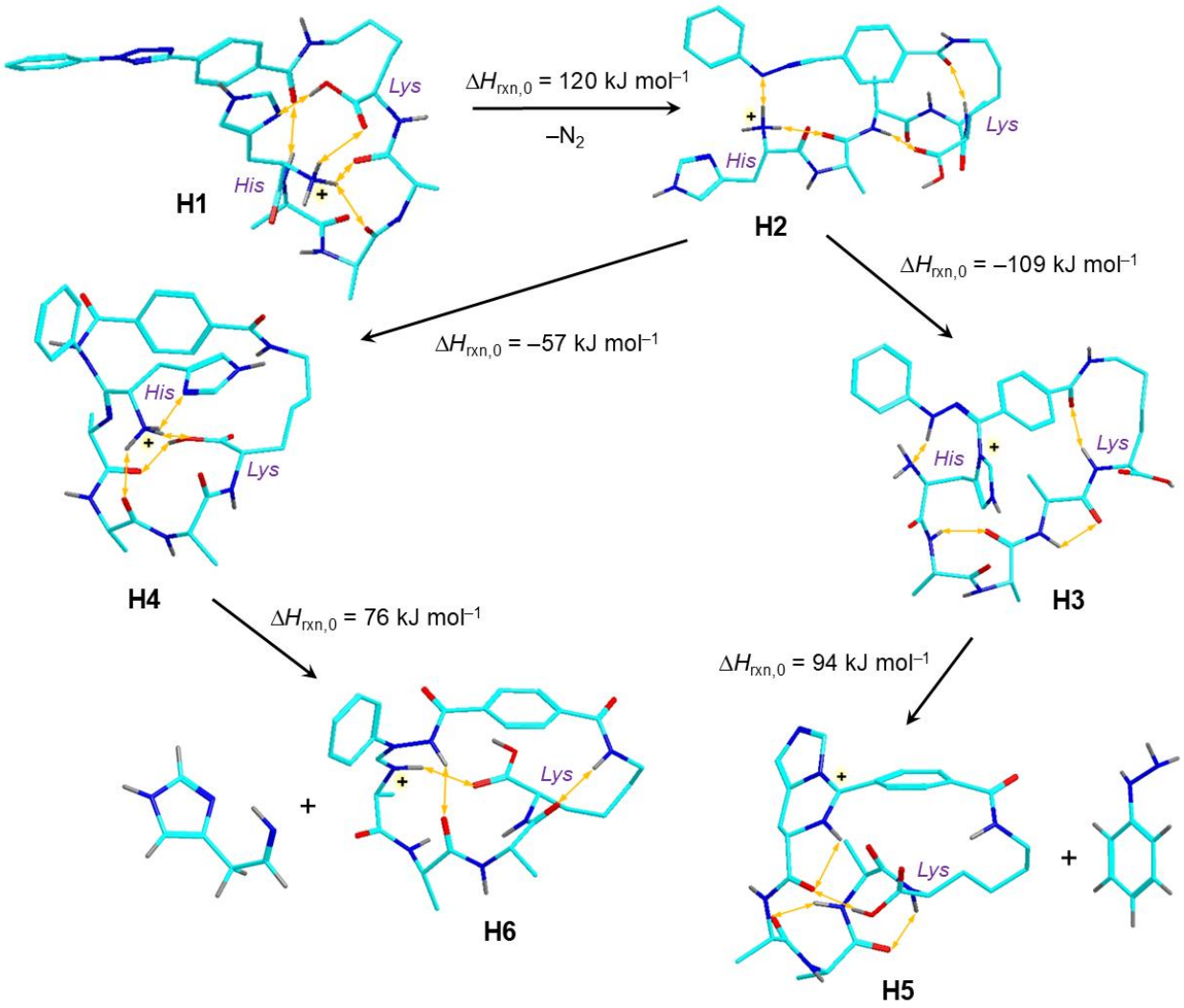


Figure 7.14: (a) CID-MS^3 of ([D$_9$]HAAA-*tet*-K –N$_2$ + D)$^+$ (*m/z* 727). (b) CID-MS^4 of ([D$_6$]HAAA-*tet*-K –N$_2$ –C$_6$H$_8$N$_2$ + D)$^+$ (*m/z* 616).



Scheme 7.6: Structures and Dissociation Energies of HAAA-*tet*-K Ions.

We used BOMD and DFT calculations to assess the thermochemistry of crosslinking in $(HAAA\text{-}tet\text{-}K - N_2 + H)^+$ ions. The lowest energy $(HAAA\text{-}tet\text{-}K + H)^+$ precursor ion (**H1**) was protonated at the N-terminus. Structures that were started with a protonated His imidazole collapsed by proton transfer to N-terminally protonated ones upon BOMD. Loss of N_2 ($\Delta H_{rxn,0}$ = 120 kJ mol^{-1}) proceeded via nitrile imine **H2**, which was the lowest-energy conformer, showing a strong hydrogen bond between the charged N-terminal NH_3 group and one of the nitrile imine nitrogen atoms (**Scheme 7.6**). Crosslinking in **H2** by the imidazole ring could produce ion **H3**, which was the lowest-energy conformer of this type. As opposed to **H1**, protonation in **H3** was favored at the crosslinked His imidazole ring whereas the ammonium proton was transferred to

the hydrazidine moiety where it formed a strong hydrogen bond to the N-terminal amine (**Scheme 7.6**). The formation of **H3** from **H2** was highly exothermic, $\Delta H_{rxn,0}$ = –109 kJ mol^{-1}, making the overall N_2 loss and crosslinking only 11 kJ mol^{-1} endothermic, and leaving this dissociation well within the range of excitation energies in **H1** after photon absorption or collisional activation. The alternative crosslinking that was associated with amide oxygen transfer onto the nitrile imine could produce ion **H4**. Although of somewhat higher energy than **H3**, the formation of **H4** from **H1** required 63 kJ mol^{-1} that was thermochemically facile considering the **H1** excitation energies.

Structure **H3** was a natural intermediate for the elimination of phenylhydrazine forming the m/z 609 ion **H5** (**Scheme 7.6**), which was the dominant dissociation of both (HAAA-tet-K + H)$^+$ and (HAAA-tet-K –N_2 + H)$^+$ (**Figure 7.12a,b**). The calculated enthalpy for the **H3** reaction dissociating to **H5** and phenylhydrazine, $\Delta H_{rxn,0}$ = 94 kJ mol^{-1}, was significantly lower than typical energies for peptide ion backbone dissociations, as was the overall enthalpy needed for the **H1** → **H5** conversion with the consecutive loss of N_2 and phenylhydrazine, $\Delta H_{rxn,0}$ = 105 kJ mol^{-1}. This was consistent with the facile loss of phenylhydrazine that outcompeted all other dissociations, such as peptide backbone cleavages that typically require threshold energies in the 162-263 kJ mol^{-1} range.[30-33] Structure **H4** was a plausible precursor for the side-chain loss of imidazolylacetaldimine ($C_5H_6N_2$), giving rise to the m/z 608 fragment ion (**Figure 7.12a,b**). This dissociation, **H4** → **H6** + $C_5H_6N_2$, required a threshold energy of $\Delta H_{rxn,0}$ = 76 kJ mol^{-1} (**Scheme 7.6**) which was well within the range of the precursor ion excitation energies. The combined threshold energies for the formation from **H1** of **H5** and **H6**, $\Delta H_{rxn,0}$ = 105 and 139 kJ mol^{-1}, respectively, were in a qualitative agreement with the formation of the corresponding m/z 609 and m/z 608 fragment ions where the lower-energy former dissociation prevailed.

The data showed that all three basic residues engaged in crosslinking interactions with the nitrile-imine group. The types of chemical reactions were different for the arginine, lysine, and histidine residues, leading to products that showed different reactivity upon further collisional excitation, as discussed above. The most salient difference was found for the dissociation type that was elimination of phenylhydrazine for the lysine and histidine crosslinks whereas the arginine crosslinks broke up the guanidine group. Since the phenylhydrazine loss involved multiple proton transfers, its lower representation with arginine crosslinks may be due to the substantially higher guanidine group basicity compared to the lysine NH_2 and histidine imidazole ring.

To assess the proclivity of the basic residues for the crosslinking reactions, we calculated the relative intensities of ions identified as originating from crosslinks and reported them as crosslink yields (**Table 7.5**). The data showed that the arginine, lysine, and histidine residues did not substantially differ in their crosslinking reactivity. A large difference was observed for ε-$(CH_3)_3$KAAAK where crosslinking was severely hampered by the lack of a mobile proton. The similar overall crosslinking yields pointed to the high reactivity of the nitrile imine group. Judged by the optimized ion structures, the interacting groups can be remote in the lowest-energy nitrile imine conformers. However, owing to the high internal energy acquired by photodissociation or CID, the nitrile imine intermediates can undergo fast conformational changes, and those that bring the proton-donating and nucleophilic groups close to the nitrile imine are bound to result in exothermic crosslinking. This was evident from the analysis of BOMD trajectories for the nitrile imines that showed progressions of multiple conformations within 20 ps.

Table 7.5 Yields of Crosslinked Products.

		crosslink yield[a]	
Sequence	UVPD conversion[b]	UVPD-MS^2	UVPD-CID-MS^3
RAAA-*tet*-K	36	67	80
KAAA-*tet*-K	40	66	89
ε-$(CH_3)_3$KAAA-*tet*-K	31	4.8	5.2
HAAA-*tet*-K	25	84	80

[a]Calculated as the percentage of identified crosslinked fragment ions relative to all sequence fragment ions.
[b]Sum of identified UVPD product ion intensities relative to the precursor ion.

7.4 Conclusions

Basic amino acid residues with nucleophilic side-chain groups, as in Arg, Lys, and His, were found to react with nitrile imines in peptide conjugates, forming internally crosslinked products. This finding expands the scope of nitrile-mine based crosslinking to a variety of peptides ranging from those with non-basic (Gly, Ala, Phe) through acidic (Asp, Glu) to weakly nucleophilic (Asn,

Gln) amino acid residues. Experiments in which the proton donor and/or nucleophiles were blocked showed substantially diminished reactivity towards nitrile imines. Crosslinking in peptide conjugates is readily recognized by the formation of internal fragments upon ion activation.

7.5 Bibliography

[1] 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.

[2] Shawali, A. S. Reactions of Heterocyclic Compounds with Nitrilimines and Their Precursors. *Chem. Rev.* **1993**, *93*, 2731–2777.

[3] 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.

[4) 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.

[5] Zhang, J.; Liu, J.; Li, X.; Ju, Y.; Li, Y.; Zhang, G.; Li, Y. Unexpected Cyclization Product Discovery from the Photoinduced Bioconjugation Chemistry between Tetrazole and Amine. *J. Am. Chem. Soc.* **2024**, *146*, 2122-2131.

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

[7] Vlk, M.; Wan, J.; Nytka, M.; Vu, T. N, K,.; Lemr, K.; Turecek, F. Photochemical and Collision-Induced Crosslinking of Asp, Glu, Asn, and Gln Residues in Peptide-Nitrile Imine Conjugate Ions in the Gas Phase. *J. Am. Soc. Mass Spectrom.* submitted for publication, September 2024.

[8] Zhu, H.; Nytka, M.; Vu, T. N. K.; Lemr, K.; Tureček, F.: "Photochemical and Collision-Induced Crosslinking in Stereochemically Distinct Scaffolds of Peptides and Nitrile-Imine Conjugate Ions in the Gas-Phase. *J. Am. Soc. Mass Spectrom.* **2024**, in press.

[9] Alexander, A. J.; Thibault, P.; Boyd, R. K. Collision-Induced Dissociations of Peptide Ions 2. Remote Charge-Site Fragmentations in a Tandem, Hybrid Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **1989**, *3*, 30–34.

[10] 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 Processes* **1991**, *111*, 173–189.

[11] Dongre, A. R.; Jones, J. L.; Somogyi,A.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* **1996**, *118*, 8365–8374.

[12] 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.

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

[14] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semiempirical 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.

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

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

[17] 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.

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

[19] 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.

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

[21] 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.

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

[23] Fishman, V.; Semidalas, E.; Martin, J. M. L. Basis Set Extrapolation from the Vanishing Counterpoise Correction Condition. *J. Phys. Chem. A* **2024**, *128*, 7462-7470.

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

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

[26] 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.

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

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

[29] 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.

[30] Klassen, J. S.; Kebarle, P. Collision-Induced Dissociation Threshold Energies of Protonated Glycine, Glycinamide, and Some Related Small Peptides and Peptide Amino Amides. *J. Am.Chem. Soc.* **1997**, *119*, 6552-6563.

[31] Paizs, B.; Suhai, S. Fragmentation Pathways of Protonated Peptides. *Mass Spectrom. Rev*. **2005**, *24*, 508–548.

[32] Laskin, J.; Yang, Z.-B.; Song, T.; Lam, C.; Chu, I. K. Effect of the Basic Residue on the Energetics, Dynamics, and Mechanisms of Gas-Phase Fragmentation of Protonated Peptides. *J. Am. Chem. Soc*. **2010**, *132*, 16006–16016.

[33] Mookherjee, A.; Armentrout, P. B. Thermodynamics and Reaction Mechanisms for Decomposition of a Simple Protonated Tripeptide, H^+GGA: From H^+GGG to H^+GAG to H^+GGA. *J. Am. Soc.Mass Spectrom*, **2022**, *33*, 355-368.

Chapter 8

Conclusions and Future Work

In this thesis, I have demonstrated the application of advanced gas-phase analytical techniques for the structural characterization of biomolecular ions. Tandem mass spectrometry serves as the core methodology, enabling structural insights by fragmenting selected precursor ions. Techniques such as electron transfer and UVPD were employed to generate reactive intermediates, including oligonucleotide cation radicals and nitrile imines. Action spectroscopy provided valuable insights into electronic structures, while cyclic ion mobility distinguished ions based on their CCS, aiding in the identification of isomers and conformers. Often, these techniques were combined to achieve a more comprehensive characterization. Computational methods, including BOMD and DFT, contributed by yielding optimized structures, theoretical spectra, and CCS_{theo}, enhancing the interpretation of experimental data. For gas-phase reactions, such as dissociations and cross-linking events, we proposed plausible mechanisms informed by experimental observations, thermochemistry, and kinetic modeling.

Despite the relatively modest size of the biomolecular systems studied—ranging from a few hundred to 3000 Da for oligonucleotides and peptide conjugates—our work emphasized the importance of secondary structure and non-covalent interactions beyond basic fragmentation and sequence analysis (Chapter 4). We highlighted processes such as proton (Chapters 5, 6, and 7) and hydrogen transfers (Chapter 3), given their significance in real biological systems. Through the use of powerful experimental methods, such as isotope labeling, combined with meticulous handling of protomer and conformer populations in computational analysis, we could achieve atomic-level resolution in investigating these transfer processes.

For each system we have studied, certain challenges remain unresolved. In the following sections, I will briefly discuss future directions and potential strategies to address these challenges for each system.

8.1 DNA Cation Radicals

A key challenge for advancing the characterization of DNA cation radicals is developing new strategies for radical generation at various sites within the ion. Although reductive pathways have

offered a versatile approach to create hydrogen-rich cation radicals and diradicals in oligonucleotides, oxidative pathways remain largely restricted to producing cation radicals from guanine and guanosine. Recently, our group have successfully generated nucleoside ions with a 2′ or 5′-O-acetyl radical on the ribose, facilitated by a distant radical initiator and triggered by CID.[1,2] These cation radicals exhibited site-selective hydrogen transfer from various positions on the ribose ring, as revealed by CID-MS^n of selectively deuterium-labeled samples. Additionally, with 2′,3′-OH-protected nucleosides containing 5′-O-acetyl radicals, we achieved the generation of non-canonical adenine and guanine cation radicals, which were otherwise inaccessible via oxidative pathways.[3]

Another important area for improvement is the resolution of characterization techniques. While Chapter 3 explored dissociation pathways for cation radicals in oligonucleotides, the specific electronic structures of each radical intermediate within these mechanisms remain uncharacterized. For smaller molecules, such as nucleobases and nucleosides, UV-Vis action spectroscopy is an excellent tool for distinguishing isomers with distinct electronic structures. However, peak broadening resulting from thermal ions diminishes the technique's resolution, as observed in larger systems. Therefore, conducting IR and UV-Vis action spectroscopy in cold ion traps would enhance the precision of isomer and conformer identification by yielding better-resolved spectral peaks. Alongside experimental spectra, advancements in computational methods are crucial for producing reliable theoretical absorption spectra to match. The structural optimization of DNA heptanucleotides discussed in Chapter 4 nearly approached the upper size limit of systems that can be feasibly calculated using our current BOMD and DFT methodologies. Additionally, the intricate hydrogen-bonding networks in oligonucleotides present challenges in calculating the harmonic frequencies required for generating accurate theoretical spectra. Thus, new methods and algorithms in computational chemistry must be developed to overcome these limitations and enable more accurate spectral predictions for complex biomolecular systems.

8.2 Nitrile Imines as Photo-Cross-Linkers

Future work on developing nitrile imines as photo-cross-linkers involves several aspects. The first is to further explore the reaction of nitrile imines with a broader range of functional groups in biomolecular ions. This thesis has examined tetrazole-peptide conjugates with a variety of N-terminal amino acid residues, including non-basic (Gly, Ala, Phe), basic and nucleophilic (Arg,

Lys, His), acidic (Asp, Glu), and weakly nucleophilic (Asn, Gln) types. These residues demonstrated different reactivity in crosslinking reactions, with proton-transfer-assisted nucleophilic attack playing a central role across cases, as analyzed through tandem mass spectrometry and ion mobility studies. Next steps include synthesizing new tetrazole-peptide conjugates, specifically XAAA-tet-K with X as Phe, Tyr, or Trp. These aromatic residues may engage in π-π interactions with the tetrazole tag in gas-phase ions, making them promising probes for studying non-covalent interactions in biomolecules. Additionally, incorporating the tetrazole tag within the middle of a linear peptide chain, flanked by different N-terminal and C-terminal amino acids, could reveal competitive crosslinking behaviors influenced by nucleophilicity, proton transfer, and steric factors. Since intrinsic functional groups in peptides also participate in crosslinking—such as internal amides highlighted in Chapters 5, 6, and 7 or C-terminal carboxyl groups in peptide scaffolds[4]—their competitive behaviors will also be investigated. Further interest lies in nitrile imine's potential as a cross-linker for DNA, given its observed reactivity toward guanine in [*dCG*-GAAA-tet-K+H]$^+$ complexes. A synthetic strategy to attach the photo tag onto a hydroxyl group within the ribose ring via ester coupling will allow investigation of crosslinking reactions in nucleotide conjugates, which feature multiple mobile protons and zwitterionic structures, as discussed in Chapter 2. This approach could broaden our understanding of nitrile imine's applications in biomolecular crosslinking.

The second focus area is on fine-tuning the reactivity of photo-cross-linkers. A straightforward approach is to block potential crosslinking sites; we have employed trimethylation for amines and esterification for carboxyl groups within peptides to reduce their reactivity with nitrile imines. Additionally, sodiation was used to eliminate available protons, providing a general method to inhibit crosslinking. Another adjustable parameter is the absorption wavelength, which can be modulated by modifying the phenyl ring with electron-donating or -withdrawing groups to shift its absorption band. This strategy is particularly advantageous when multiple photo tags are used within the same system, allowing selective crosslinking by simply adjusting the wavelength. In larger systems, such as protein complexes, precise control over crosslinking efficiency and site specificity would be highly beneficial. Achieving this control requires continuous exploration of the crosslinker in various systems and a deep understanding of its reaction mechanisms.

A long-term goal is to evaluate the performance of nitrile imines in crosslinking mass spectrometry (XL-MS) for proteins, as they offer milder and more tunable reactivity compared to

carbene cross-linkers. Key challenges include the synthetic incorporation of tetrazole tags into proteins, addressing the low crosslinking yields often resulting from side reactions with solvents, and developing methods to quantitatively translate mass spectral data into structural information.

8.3 Bibliography

[1] Zima, V.; Gladwish, O.; Marek, A.; Tureček, F. Nucleoside Cation Radicals: Generation, Radical-Induced Hydrogen Atom Migrations, and Ribose Ring Cleavage in the Gas Phase. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (7), 1594–1608.

[2] Zima, V.; Marek, A.; Tureček, F. Competitive Radical Migrations and Ribose Ring Cleavage in Adenosine and 2′-Deoxyadenosine Cation Radicals. *J. Phys. Chem. A* **2024**, *128* (6), 1109–1123.

[3] Zima, V.; Liu, Y.; Tureček, F. Radical Cascade Dissociation Pathways to Unusual Nucleobase Cation Radicals. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (6), 1038–1047.

[4] Zhu, H.; Nytka, M.; Vu, T. N. K.; Lemr, K.; Tureček, F. Photochemical and Collision-Induced Cross-Linking in Stereochemically Distinct Scaffolds of Peptides and Nitrile Imines in Gas-Phase Ions. *J. Am. Soc. Mass Spectrom.* **2024**. *In Press*.