Structural Elucidation of Biomolecular Ions in the Gas Phase Using Novel Mass Spectrometric and Computational Methods

Yang Liu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

František Tureček, Chair

Matthew F. Bush

Bo Zhang

Program Authorized to Offer Degree:

Department of Chemistry

University of Washington

Abstract

Structural Elucidation of Biomolecular Ions in the Gas Phase Using Novel Mass Spectrometric and Computational Methods

Yang Liu

Chair of the Supervisory Committee:
Professor František Tureček
Department of Chemistry

Biological molecules such as protein and DNA play critical roles in various cellular reactions and carry out essential biological functions. But the mechanisms of these reactions, and the essential molecular structures to facilitate them are often poorly understood. Many efforts have been applied to these problems, but due to their complexity and various limitations of existing methods, the elucidation of biological reaction mechanisms and biological molecule or complex structures is still considered a challenge today. Therefore, the scientific community has been highly motivated to invent new methods to tackle these problems devise new angles of approach to grain insights into biomolecular structure. This dissertation summarizes some recent work in

characterizing cation-radical reactions, and structures of gas-phase molecular complexes using novel mass spectrometry methods in combination with theoretical computational modeling.

The novel mass spectrometric methods presented in this work are gas-phase photo-dissociative crosslinking techniques and UV-visible photodissociation action spectroscopy. They have several unique advantages in tackling structure elucidation of weakly bound complexes and transient radical species. 1) The mass spectrometer (MS) is a universal detector which is widely used to characterize various kinds of biological molecules. 2) MS works with ions of interest that are generated and stored in the gas phase at low pressure of an inert gas (He). It is the perfect system to study reaction mechanisms because of low interference from the ambient environment, such as solvent, counterions, surfaces, etc. 3) Methods exist to generate cation radicals in the gas phase, and thanks to the low pressure inside a mass spectrometer, the produced species are kinetically stable on the experimental time scale of several milliseconds. 4) MS is a perfect tool for conducting gas-phase reactions because it allows one to manipulate the ion population and select and focus the ions. It is advantageous for crosslinking reactions that often suffer from low concentrations of reactants when the reaction is attempted in condensed phase. 5) The experimental action spectra can be interpreted using sophisticated computational techniques providing vibronic transitions of multiple isomeric candidates. The new data generated with these new mass spectrometry methods offer unique insights, but also pose challenges to experimental data interpretation. Various computational approaches are used to supplement the experimental data, and the results from the computations are used to guide the interpretation. This dissertation also includes several novel approaches on the computational front, introducing the customized modeling pipeline with a combination of Born-Oppenheimer dynamics, density functional theory calculations, and machine learning techniques.

The first chapter introduces some basic terms and outlines the background for the topic of study, including a quick review of the current challenges and techniques in the field of structure elucidation of biomolecules. Also introduced are the fundamentals of mass spectrometer, commonly used and newly developed ion activation methods, and computational modeling. This chapter lays the foundation of the work described in later chapters.

A small neuroprotective peptide Cys-Ala-Gln-Lys (CAQK) has recently been discovered to mitigate adverse effects of brain trauma in mice, possibly because of specific interactions with yet-to-be identified proteins. In Chapter 2, an application of photo-crosslinking is described to study the noncovalent interactions of CAQK with several model target peptide motifs. The experimental results in combination with Born–Oppenheimer molecular dynamics revealed the structural preferences for binding to the amino acid residues of potential target peptides.

In Chapter 3, inspired by the biological significance of CAQK, we employed the photo-dissociative crosslinking techniques to probe its interactions with several dinucleotides as structure motifs of nuclear DNA. This lysine-containing peptide can be viewed as a simplified surrogate of a histone interacting with DNA. We were able to show that in interactions with CAQK ions, even simple dinucleotides differ in their binding efficiency and stereochemistry. We provided structures of selected complexes obtained by electronic structure theory calculations using density functional theory (DFT).

UV and energetic particle radiation can ionize DNA creating electron deficiency (hole) at nucleobases. These holes can migrate along the strand, leading to lesions and DNA damage by follow up radical reactions. Chapter 4 and chapter 5 are focused on the characterization of DNA cation radicals in the guanine and cytosine containing dinucleotide as model systems in probing the electron transfer mechanisms in DNA radiation damage. Experimental action spectra were

obtained for these small nucleotides and the absorption bands were interpreted by finding the closest match from the calculated spectra.

The last chapter features an ongoing project in which we have made an attempt, using photo-crosslinking techniques, to probe noncovalent interactions within a complex consisting of a chiral agonist and its binding motifs. The experimental results revealed high binding affinity of the native agonist regardless of chirality, however the crosslinking fragment was not observed. Several DFT optimized low energy structure had shown a consistently low “contact” rate of the phototag with the target peptide. A further characterization of ion structures will be complemented from the collisional cross section analysis by ion mobility (IM) measurements.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank my research advisor Dr. František Tureček for your strong support. I truly appreciate the guidance and insight you provided over these years. During this long journey, for all those moments that I doubted myself, you've always had faith in me. Thank you for creating such an encouraging workspace to promote self-paced learning and problem-solving. Inspired by your enthusiasm towards research, and motivated by your curiosity in various technology frontiers, I have found the learning and exploring process quite enjoyable!

Secondly, I would like to thank my undergraduate research advisor Dr. Yuzhong Wang and Dr. Sichong Chen who introduced me into the world of research. Thank you for encouraging me to study abroad and pursue different research topics.

I would like to thank Dr. Michael Gelb for the fantastic opportunity that converts analytical chemistry techniques into real-world clinical applications. Taking such high impact project (newborn screening) as a starting project in graduate school has been a very rewarding experience and have helped me build confidence in research.

There are many people at UW who had helped me along the way. Thanks to Martin for all the trainings on the instruments and organizing all these regional meetings to help us make connections. I would like to thank my coworkers Xinying, Fan, Arun and Sophia in the Gelb lab for all your generous help in assay development and synthesis. I would also want to thank my amazing colleagues in the Turecek group: Chris and Rob, thanks for all the time and patience you spent teaching me the fundamentals. Angelo, thank you for encouraging me to take "hands-on"

actions and conquer my fears to make mistakes. Ivy, thank you for all the kick-start trainings on computation modeling! Emilie and Andy, I feel so lucky to be friends with such wonderful people like you! Thank you for always being there for me! I treasure all the moment we spent together in high or low spirits. Wing, I can’t thank you enough for having my back in those last-minute emergencies and for helping out with my projects when nobody else could! And Yue, I really appreciate your can-do spirit and your positive attitude towards the unknown future. All these people I met at UW have influenced my life and helped me become a better person.

Lastly, I would like to thank UW DIRECT for broadening my horizons, allowing me to explore the diversity of research ideas and providing me with various career options.

DEDICATION

This dissertation is dedicated to my amazing husband, Chunxiang for all your love, understanding and encouragement during this long journey.

I wish also to dedicate this to my wonderful parents for pushing me to do the best and trusting me to make my own life decisions.

Chapter 1. INTRODUCTION

1.1 A Brief Overview of Structure Elucidation of Biomolecules

The three-dimensional structures of biomolecules, such as proteins and nucleic acids, are essential to perform their biological function. Disruption of these biomolecular structures could result in cellular and/or bodily malfunction and diseases such as Alzheimer's[1], Parkinson's[1] and cancer[2]. To address these biological and clinical challenges, one needs to study the structure of pertinent biomolecules. Structure elucidation of biomolecules, therefore has drawn a lot of attention in the fields of medicine[3], biotechnology[4] and fundamental research[5]. A variety of different research techniques have been applied to study biomolecule structures, including high-resolution technologies such as X-ray crystallography[6, 7] and NMR[8], as well as relatively low-resolution methods such as spectroscopy[9], radical probes[10] and chemical crosslinking coupled with mass spectrometric analysis[11]. Mass spectrometry, in particular, provides fast experiment turnaround times to achieve *in vivo* structure elcucidation[11]. In addition to experimental methods, recent developments in computer-assisted methods such as theoretical calculations[12,13] and bioinformatics[14] can also provide structural information from a different prospective. The combination of these different approaches has shown great potential in structural elucidation, zooming from low-resolution information obtained with bigger systems to much higher-resolution structures achieved with modeling.

In this dissertation, I will discuss two novel techniques for structure elucidation as well as the application of these techniques to several biomolecular systems of biological significance. These techniques leverage mass spectrometry for its versatile capability of studying gas phase ions, and combine various computational methods for additional insight and interpretation of the

experimental data. Before we deep dive into the new technology, let's first review some key terms and background of this topic.

1.1.1 *Non-covalent Interactions*

Non-covalent interactions are critical for maintaining the sophisticated 3D structure of biomolecule complexes, in particular, for heterogeneous complexes. These interactions also provide flexibility and specificity in most important biological processes, such as DNA replication[15], enzyme inhibition[16], and neuron transmission pathway[17]. A minor change in the structure could result in changes of noncovalent interactions which then in turn could cause dramatic changes to biological functions[1]. The conventional approach here includes high-resolution approaches such as X-ray, NMR or cryo-electron microscopy[18], have been established to study the binding sites of these non-covalent complexes. However, newly discovered biological effects typically wait years before being studied by direct high-resolution "imaging approach". These approaches usually study one biological molecule or complex in each experiment, which is considered low-throughput. Another limitation of these methods is the lack of characterization of dynamic and transient[19] complexes as X-ray requires crystalline samples. Large-scale and high-throughput methods such as yeast-2 hybrid[20], and affinity purification[21] can quickly identify the units that are involved in forming complexes but little structure information are provided with these approaches.

A notable breakthrough of measuring dynamic complex structures in biological systems owes to the development mass spectrometry based characterizations: Stoichiometry, a critical piece of information that determines the mass of biomolecules associated noncovalently could be obtained by "native"[22] mass spectrometry techniques; Collisional cross-section measured by ion mobility mass spectrometry (IM-MS) are used to separate ions by their shapes to provide insights

into protein folding[5]; H/D exchange[10] and hydroxyl radical labeling are used to probe the conformational dynamics, folding and binding by "exchanging" and "marking" the surface of complexes. Crosslinking-mass spectrometry specializes at probing site-specific interactions at the interfaces, extending into *in vivo* analysis[11]. Although these mass-spectrometry based methods alone cannot offer structure information at the level of resolution provided by NMR and X-ray, they have a higher throughput and the capability of dealing with weak complexes, potentially in *vivo*. Therefore, these mass spectrometry-based methods can offer information that other conventional technologies cannot provide. With the development of computational modeling techniques, one can employ theory to aid the interpretation of experimental findings by the mass spectrometry-based technologies and therefore reconstruct the detailed 3-D structure.

This work describes photochemical crosslinking-mass spectrometry as the main methodology of mapping non-covalent interactions; a detailed description of the crosslinking experimental design and computational methods will be provided in later sections. Projects related to this subject will be described in Chapters 2, 3 and 6.

1.1.2 *Unusual Radical Species*

Radicals are highly reactive chemical species bearing an unpaired electron. Unlike the traditional image of self-regulated equilibria of cyclic biological processes known by biochemists for at least half a century, radical reactions seem "un-biological" as they are usually irreversible and explosive[23]. Although being unstable and energetically unfavored, radical reactions are involved in various essential biological processes. Several amino-acid radicals, such as those derived from tyrosine, tryptophan, glycine and cysteine, play an essential role in the function of redox enzymes that maintain biological metabolism, DNA synthesis, and immune regulation.[24]

Unwanted radicals, however, are destructive and can cause damage to cellular components, such as proteins, lipids and DNA, leading to degenerative diseases such as diabetes and cancer.[25]

Oxidative stress and ultraviolet light radiation can cause radical damage to DNA. The formation of a highly reactive cation radical (or a "hole") is primarily initiated at a nucleobase. The hole can then propagate along the chain to cause chemical damage, resulting in DNA strand breakage and nucleobase release[26], as shown on Figure 1.1 .

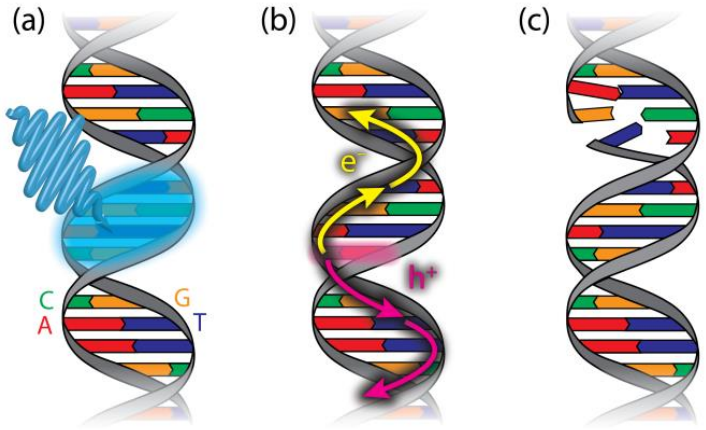


Figure 1.1. Schematics of DNA ionization damage and hole transport.
Image source: https://taaoliver.com/research/

Studying the mechanism of electron transport in DNA cation radicals contributes to the understanding of DNA damage. The electron transfer kinetics have been studied by indirect approaches such as reaction product identification [27]. However, the short transient lifetime of ion radicals in solution has been the main challenge of achieving spectroscopic characterization. Experimental spectroscopic characterization of these transient intermediates has been performed with nanosecond UV flash photolysis, electron paramagnetic resonance (EPR) and picosecond time-resolved IR spectroscopy.[28] Fast UV-vis spectroscopy photolysis revealed rapid deprotonation of nucleosides and resulted in the formation of neutral radicals. EPR has been used to study the radical mechanisms but was hampered by the similarity of the spectra of charged

radicals and neutral radicals.[29] Theoretical modeling with density functional theory (DFT) calculations has been useful in identifying these transient intermediates but the theoretical description alone was often overly simplified.[30]

The characterization and investigation of cation-radical species in the gas-phase can be advantageous, when compared to the characterization of cation-radical species in condensed phase. For example, radical side reactions, such as interactions with solvent, are avoided in the gas-phase and the experimental results are readily combined with computational modeling for interpretation. One method to generate these DNA cation radicals is by electron transfer to a multiply charged precursor ion and further characterized by mass spectrometry[31]. Infrared multiphoton dissociation (IRMPD) and UV-vis action spectroscopy are modern mass spectrometric techniques used to characterize the structures and absorption bands of transient radical intermediates. This work will focus on combining UV-vis action spectroscopy and theory to study the structures and electron transfer mechanisms of dinucleotide model systems. The projects related to the characterization of these transient cation radical species are described in Chapter 4 and 5.

1.2 Brief Introduction of Mass Spectrometry

Mass spectrometry has become a powerful tool in analytical chemistry with applications in environmental[32], forensic[33] and clinical[34] analysis. The robust and universal detection capabilities of mass spectrometry, aided by significant technological advances, have also enhanced its ability in structural characterization. The history of mass spectrometry dates back to the 1890s. J.J. Thomson and E. Everett developed an apparatus that measured the mass of the electron, which led to Thomson's Nobel Prize in Physics in early 20th century. The first mass spectrometer capable of measuring atomic weight, was constructed by Thomson and his disciple, and was named the

"parabola spectrograph". The parabola spectrograph laid the foundation for future designs of mass spectrometers[35].

Before diving deeper into the technological development of mass spectrometers, some fundamental concepts about mass spectrometry need to be reviewed as they serve the main function of mass "recognition".

1.2.1 *Ionization Source*

Mass spectrometers measure the mass-to-charge ratios of charged species generated from an ionization source. Many types of ionization sources have been developed and applied to different fields of research, depending on the purpose of research and analyte types.

Electron impact (EI)[35] ionization has been the most common and widely-used ionization methods for gaseous samples. EI is implemented by collisions of accelerated electrons with analyte molecules in a chamber, resulting in ejected electrons and cations. These ions can be then separated by their mass-to-charge ratio in a mass analyzer. Though robust and popular in the past decades, the drawbacks of EI are not negligible. EI requires the analyte to be volatile as the ionization takes place in gas phase on the vapor form of the analyte. Also, EI is a "hard" ionization method[36] due to high energy transferred from ionizing electron to the analyte molecule, breaking the molecule ion into small fragments prior to detection. This can make the identification of the analytes very difficult for the molecular ion cannot be detected in many cases.

Chemical Ionization (CI)[37, 38] is another commonly used ionization method that is similar to EI as proceeding in the gas phase, but different in that CI requires a reagent gas. At ~1 Torr, the reagent gas, e.g. CH_4, can be ionized by high-energy electrons to form highly reactive cation radical $CH_4^{+\bullet}$. These cation radicals can further react with neutral CH_4 to generate $CH_3^{\bullet}$ radicals and CH_5^+ ions. The CH_5^+ cation is a strong acid that can transfer a proton to an analyte molecule

to generate the molecular ion; hence, a "soft" ionization method. In the case of negative ion mode, e.g. with an electron transfer dissociation (ETD) reagent source, the ionization relies on electron affinity. Take fluoranthene as an example—when neutral methane is hit by a high energy electron, lower energy (thermal) electrons are ejected, and can be absorbed by analyte molecules that have a finite (positive) electron affinity, such as fluoranthene which is commonly used as an ETD reagent[39].

Mass spectrometry analysis had been limited to low mass, gaseous volatile samples until the invention of new soft ionization methods. Electrospray Ionization (ESI), developed by John Fenn in 1984, opened a new era of mass spectrometry by turning what was then "impossible" into possible: that is the ionization of high molecular weight, low volatility samples.[40] ESI, a "soft" ionization method, is beneficial to biological and material science research due to its capability of transferring analytes of high masses and low volatilities into gas phase, with no or only little bond breakage. In addition to preserving the molecular ion (or pseudo-molecular ion) of analytes, the ESI source can generate ions in multiple charge states which is desirable for the study of biomolecules (e.g. proteins, nucleic acids and carbohydrates) and synthetic polymers. By generating ions of multiple charge states, ESI is capable of bringing analytes of high molecular mass (kDa - MDa)[41] into the mass spectrometer. These highly charged ions can retain some noncovalent interactions that contribute to forming higher order structures of proteins and enzyme complexes[42]. ESI can also be conveniently coupled with separation techniques such as liquid chromatography (LC) to increase its robustness for quantitative analysis in applications such as proteomics[43] and metabolomics[44].

ESI facilitates the formation of analyte ions from liquid-phase to the gas-phase. This ionization is driven by a high electrical potential (2~3 kV) difference between the mass

spectrometer inlet and the capillary tip that contains analyte sample, as shown in Figure 1.2. The sample flow is dispersed by the electric field into an aerosol, and the de-solvated charged analyte ions are released with the aid of electrostatic repulsion.[45] A typical ion generated by ESI is charged by the adduct of one or more protons, however, other common adducts, such as sodium, potassium or ammonium, are quite possible[46].

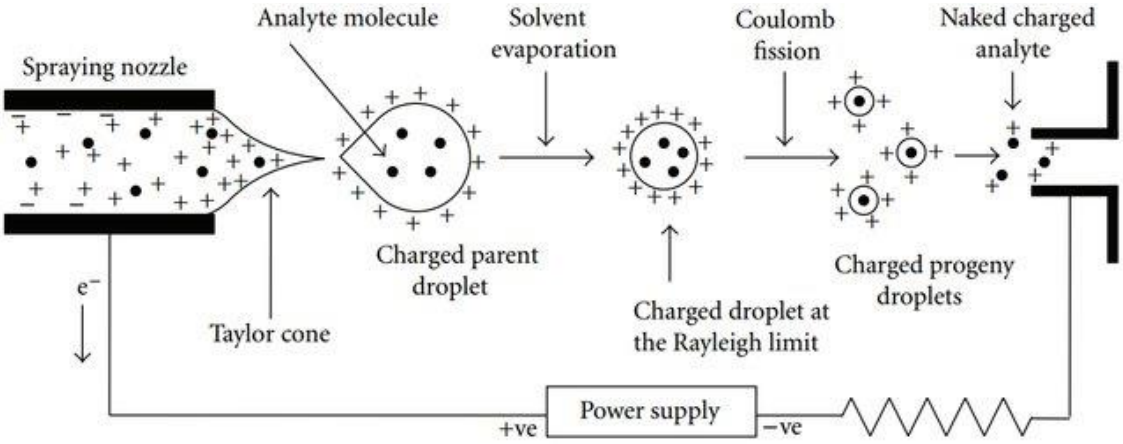


Figure 1.2. Schematics of gas-phase ion generation via Electrospray Ionization (ESI). Image source: https://www.researchgate.net/figure/Schematic-representation-of-the-electrospray-ionization-process-Adopted-from-33_fig5_307606559

Other modern ionization methods such as matrix assisted laser desorption (MALDI), fast atom bombardment (FAB), desorption electrospray (DESI), are used in mass spectrometric analyses[47] of different types of samples but will not be discussed here in this work.

1.2.2 *Mass Analyzers and Ion Traps*

There are many types of mass analyzers that can accomplish the task of distinguishing and separating ions of different mass-to-charge ratios. The original mass spectrometer created by Thomson used a combination of electrostatic and magnetic fields to direct the ions of different mass-to-charge ratios to move along different trajectories. A similar approach is still used in

modern mass spectrometers such as magnetic sectors[48]. The ions are separated according to their momentum to follow different trajectories. The ion flight times are usually in microseconds[49].

The quadrupole mass filter is one of the most commonly used mass analyzers. The quadrupole is composed of two pairs of metal rods, in which alternating AC and DC potentials are applied. The two pairs are placed opposite to each other and charged with a potential of opposite polarity. The region confined within the four rods serves as a mass filter as only ions of a certain m/z can pass through the quadrupole while others will have unstable trajectories and ejected (Figure 1.3). By scanning a range of AC and DC potentials, one can achieve the separation and detection of a mixture of ions within a selected mass-to-charge range to obtain a mass spectrum[47]. Quadrupole devices can also serve as reaction cells in which ions are accelerated to collide with a reaction gas to induce dissociation, by tweaking the potential on the rods. The ion activation will be discussed in the next section.

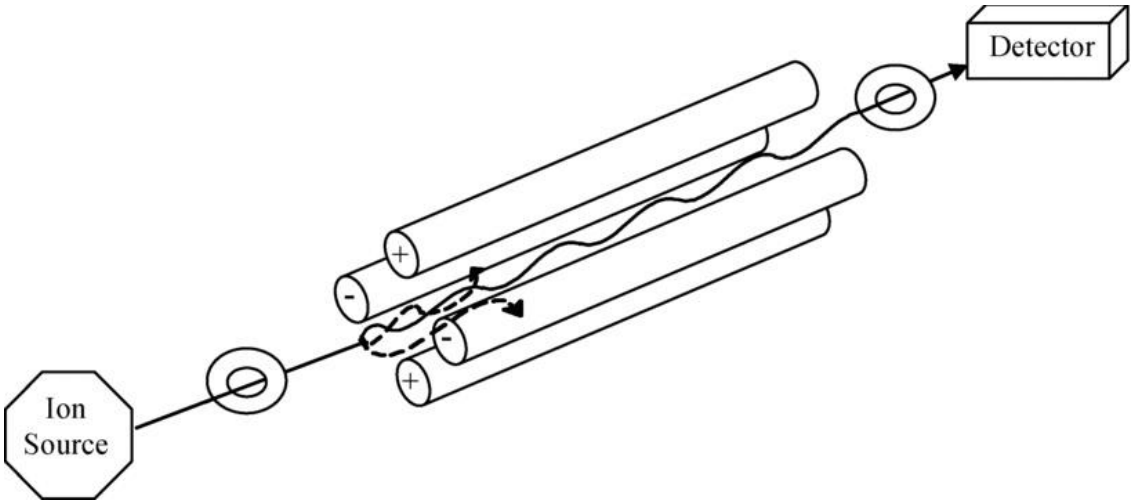


Figure 1.3. Schematics diagram of a quadrupole mass analyzer[47].

Source: El-Aneed, A.; Cohen, A.; Banoub, J., *Applied Spectroscopy Reviews* **2009,** *44* (3), 210-230

Ion traps are modified quadrupole devices that allow multiple functionality such as ion filtering, ion storage and activation[50]. Based on the design, ion traps can be categorized into 3-D traps and 2-D linear traps. The 3-D trap, also called the Paul trap, after the inventor, 1989 Nobel Prize winner Wolfgang Paul, has become a standard ion trapping device for many years. The 3-D trap consists of two metal electrodes (i.e. end electrodes) facing each other and a ring electrode placed in between. Ions are trapped in the region confined by these electrodes in an RF electric field (Figure 1.4), and move along complex periodic trajectories[51].

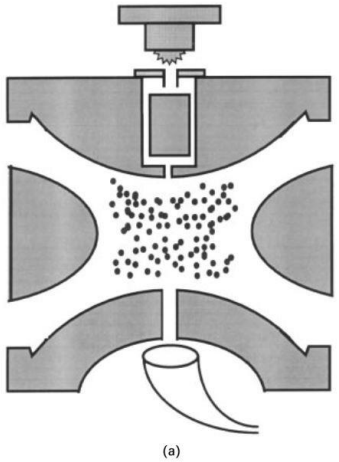


Figure 1.4. Schematics diagram of ion trapping and storing in a 3D ion trap[51].
Source: March, R. E., *Journal of Mass Spectrometry* **1997,** *32* (4), 351-369

The 2-D linear ion trap works similarly to the 3-D trap with the ability of resonantly exciting and ejecting ions from the trap. Instead of trapping the ions in the center of the 3-D trap, a linear ion trap adapted from the quadrupole in which ions are trapped along the z axis. Such a structure enables the linear ion trap to contain a larger number of ions because it is less affected by space charge repulsion of the trapped ions. The ion motion in the ion trap is driven by the quadrupole field. In a typical linear ion trap (Thermo LTQ (Figure 1.5)), the rods are segmented into three parts: the front section, center and back section[52]. The ions are confined radially in the

xy plane by a radio frequency applied at the center section and along the z axis by DC potentials applied on the front and back sections. The ions are selectively ejected by an auxiliary dipolar frequency applied on the X rods that have a thin slit cut in them to allow ion passage. When the applied dipolar frequency equals the frequency of motion of trapped ion, it gains resonance and is moved farther away from the center, and is subsequently ejected through the slits to the detectors.

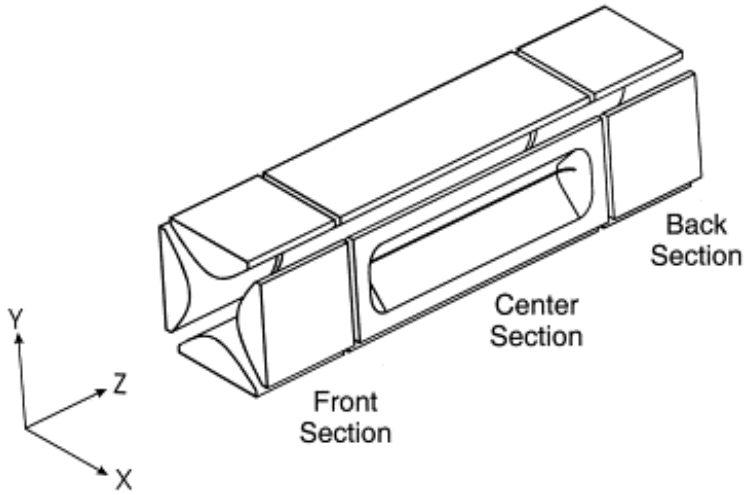


Figure 1.5. Schematics diagram of a 2-D trap[52].

Source: Schwartz, J. C.; Senko, M. W.; Syka, J. E. P., **2002,** *13* (6), 659-669

1.2.3 *Ion Activation*

With soft ionization methods such as ESI that produce molecular (or pseudo-molecular) ions, little structural information can be obtained just by ionization. However, with subsequent ion activation with tandem mass spectrometry (MS/MS), this disadvantage is easily resolved due to the capability of dissecting ion structures. Three major ion activation methods that will be discussed here are Collision Induced Dissociation (CID), Electron Transfer Dissociation (ETD) and UV-Vis photodissociation (UVPD).

Collision Induced Dissociation (CID) is the most frequently used method for ion structure elucidation. In an ion trap, AC potential is applied in addition to trapping potentials, to resonantly

excite the trapped ions. This increases the ion kinetic energy from inelastic collisions with buffer gas in the trap. The center-of-mass collision energy is translated into internal energy, which is then distributed throughout the ion's vibrational modes. When the ion internal energy acquired by cumulative collisions exceeds the dissociation threshold, fragmentation occurs. The extent of fragmentation can be tuned by adjusting the AC amplitude and the activation time[53]. In this work, most CID measurements were done with the activation of peptides and oligonucleotides ions. The standard fragmentation scheme of peptide (Figure 1.6) and oligonucleotide (Figure 1.7) is shown below. The most common fragments from peptide ions result from amide bond cleavage to yield *b/y* type product ions[54]. Sequencing of nucleotide ions results in all sorts of bond cleavages[55] with fragment ions of the *a/w* and *d/z* series being the most prominent products.

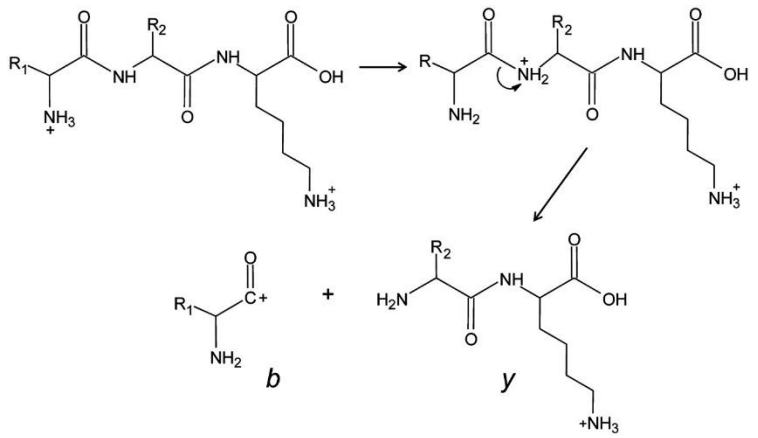


Figure 1.6. Schematics diagram of peptide fragmentation by CID[56].

Figure is adapted from: Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., *Proceedings of the National Academy of Sciences* **2004,** *101* (26), 9528-9533

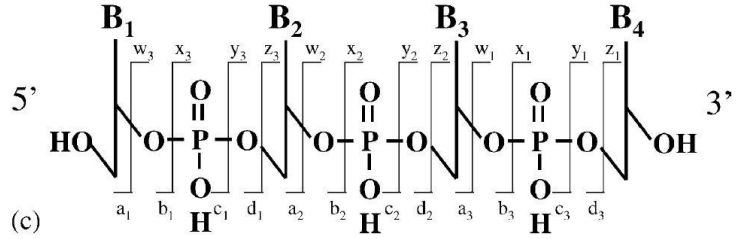


Figure 1.7. Oligonucleotide fragment ion nomenclature by McLuckey et al[55].
Figure is adapted from: Wu, J.; McLuckey, S. A., *International Journal of Mass Spectrometry* **2004,** *237* (2-3), 197-241.

Electron transfer dissociation (ETD) is another ion activation method employed for MS^n structure elucidation. It is implemented by the transfer of an electron from an electron donor to multiply-charged cations of interest. The electron donors are usually anions radicals, such as fluoranthene, which are generated in a CI source. ETD produces analyte cation radicals that typically cannot be made by CID. This technique is particularly useful in peptide sequencing[56] and in locating post-translational modifications (PTMs)[57], as it fragments the N-Cα bond along the backbone (*c/z* ions) while mostly preserving the PTMs. In contrast, traditional CID mainly produces *b/y* series while also causing losses of PTM groups (phosphate, carbohydrate) that are difficult to locate in the peptide. ETD is also useful for identifying some amino acid side-chain characteristics that cannot be identified by CID.[58] In some cases the electron transfer does not result in dissociation (i.e. ET-no-D). We utilized such reactions for the generation of transient intermediate such as DNA cation radicals[59] and peptides cation radicals[60] to mimic the short-lived radical species in biological systems. These unusual species are difficult to obtain via a standard ESI ionization source or CID, but ET-no-D techniques have given us access to the world of transient species.

Ultraviolet-Visible Photo-Dissociation (UVPD) is a relatively new activation method that is utilized in this work. High energy photons are used to trigger bond activation and applied for gas-phase reactions or structure elucidation. UV-vis photons are introduced into ion trap to activate the trapped ions. Rapid electronic excitations occur when the ion absorbs a photon and can result in bond dissociation based on the following three mechanisms (shown in Figure 1.8): the first mechanism involves a transition from the ground electronic state to an unbound vibrational level in excited state that is above the dissociation threshold, resulting in fragmentation from the excited state; In the second case, the electronic energy from photon absorption is converted to the vibrational energy via internal conversion, therefore promoting dissociation from the ground state; In the third scenario, the fragmentation is caused by excitation followed by curve crossing to a dissociative state.[61] In order to qualify for UVPD activation, the ions must have chromophores that absorb light within the UV-vis region. UVPD often requires instrumentation modifications to allow a laser beam to enter the mass spectrometer and interact with ions. In addition, a software communication between the laser control system and the ion trap must be synchronized for the correct trapping time window. The instrumentation that has been built and used in the Turecek group is demonstrated below[62] (Figure 1.9).

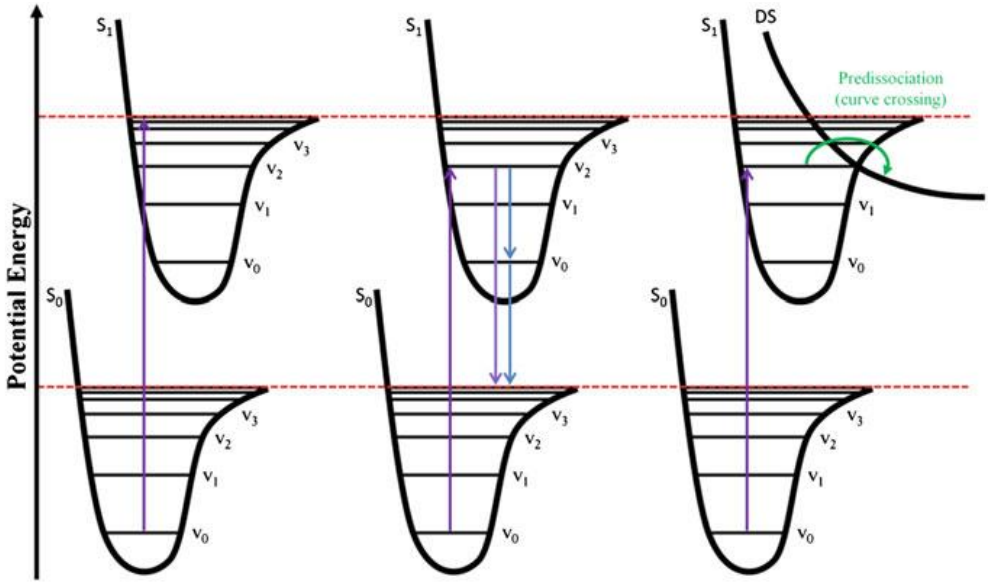


Figure 1.8. Three potential dissociation mechanisms of UV Photodissociation [61].

Figure is adapted from: Polfer, N. C.; Dugourd, P., *Laser photodissociation and spectroscopy of mass-separated biomolecular ions*. Springer: 2013; Vol. 83.

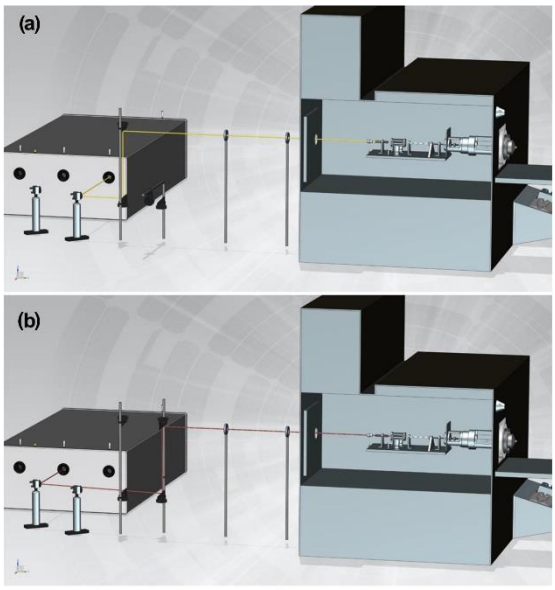


Figure 1.9. Instrumentation of Thermo LTQ mass spectrometer for the application of MS^n-UVPD [62].

(a) tunable wavelength 210 nm- 700 nm (b) single wavelength 355 nm

Figure is adapted from: Dang, A.; Korn, J. A.; Gladden, J.; Mozzone, B.; Tureček, F., *Journal of The American Society for Mass Spectrometry* **2019,** *30* (9), 1558-1564.

Most applications of UVPD-MS/MS can be classified into two categories: (1) detection of photo-reaction products based on fixed single-wavelength activation; (2) measurement of wavelength dependent absorption spectra with a tunable photon source.

Single-wavelength activation using 157, 193, 213 and 266 nm photons[63] has been used as a structural elucidation tools for peptide ions as these wavelengths targets the bond breakage of amides (193 and 213 nm), tyrosine/tryptophan side chains (266nm) and arginine (157nm)[64]. Photons of single wavelength 355 nm are outside the absorption region of most naturally-occurring peptides or DNA and are used for characterization of synthetic chromophore-modified peptides[65]. Photons of 355 nm can also be utilized as photon-switches for photo-crosslinkers such as diazirines. This thesis will summarize some novel applications of probing non-covalent interactions in biomolecular complexes in the gas-phase by incorporating the synthetic diazirine photo-crosslinkers. Photochemical crosslinking will be discussed in later sections and the relevant projects will be further discussed in Chapters 2, 3, 6.

The second UVPD-MS/MS scenario, often referred to as action spectroscopy, can be pictured as conducting gas-phase spectroscopy within a mass spectrometer, and will be discussed below in section 1.2.4.

1.2.4 *Action Spectroscopy*

Action spectroscopy, also known as consequence spectroscopy, is a powerful characterization technique[66]. This measurement is based on the “consequence” of light activation such that the dissociation of the targeted ion is an indication of analogous absorption in the condensed phase. However, in the gas phase, the ion population is too low for absorption measurements via the decrease of transmitted light intensity alone. Instead of measuring the change in light intensity, the method focuses on the “action” of intense light activation. When an

ion of interest is irradiated with light, the absorbed photon can cause fragmentation if the absorbed energy is sufficient for bond cleavage. The dissociation yield is wavelength-dependent and reveals some of the electronic or vibrational absorption properties of the parent ion. Plotting the relative abundances of photo-fragmented ions as mass-resolved channels as a function of radiation wavelength results in action spectra (demonstrated in Figure 1.10). The absorption bands are representative of different electronic or vibrational transitions in the ion population. These combined transitions are interpreted with computer-aided theoretical calculations. Some examples of action spectroscopy are infrared multiphoton dissociation (IRMPD) action spectroscopy[67] for vibrational transition characterization, ultraviolet-visible photodissociation (UVPD) action spectroscopy[68] for vibronic transitions, and infrared ultraviolet ion dip spectroscopy for the vibrational transition measurement of UV-selected conformers[69] .

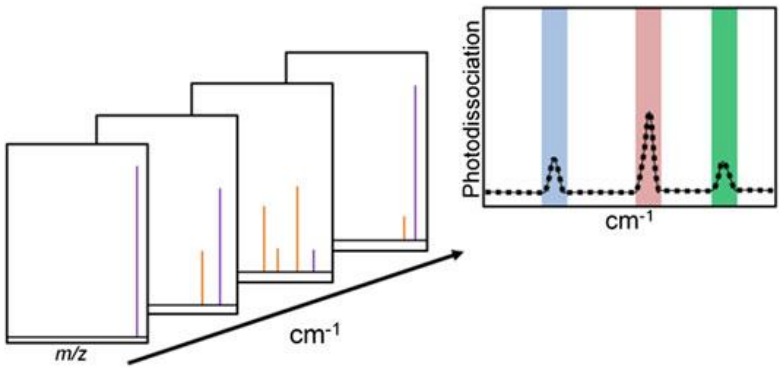


Figure 1.10. Schematics of photodissociation action spectroscopy.
The precursor ion is dissociated into multiple product ions via resonant photon activation[61].
Figure is adapted from: Figure is adapted from: Polfer, N. C.; Dugourd, P., *Laser photodissociation and spectroscopy of mass-separated biomolecular ions*. Springer: 2013; Vol. 83.

While infrared multiphoton dissociation (IRMPD) action spectroscopy involves multiple low-energy photons for “slow heating” of the target ion into dissociation, UV-vis photo-dissociation (UVPD) action spectroscopy is often achieved with a single photon that carries

sufficient energy to cross the dissociation threshold, following the three main mechanisms mentioned in the "Ion Activation Section", Figure 1.8. Due to the multitude of vibronic transitions available at room temperature, the UVPD action spectra usually display broad bands[70], and therefore this method alone is better suited for electronic structure characterization. A thorough understanding of the conformation, electronic structure and thermodynamics of these transient species can be achieved by the combination of UV action spectroscopy and high level quantum theory calculations.[68, 71, 72] With dinucleotides as model systems, conformational structure elucidation and charge transfer mechanism studies will be described in detail in Chapter 4 and 5.

1.2.5 *Ion Detection*

The ions that are separated in the mass analyzer are detected to generate a mass spectrum. The ion detector records the charge or current when an ion hits it. Faraday cups, electron multipliers, and electro-optical ion detectors are the most popular modern detectors. Faraday cups are metal cups linked with a current amplifier and voltage convertor. The current brought in by the impinging ions charges a resistor circuit, and the resulting voltage is amplified by an operational amplifier and measured by the metering circuit. Electron multipliers (EM) are the most common ion detectors because of their high sensitivity and fast response times which makes them compatible with the fast scanning speed of modern mass spectrometers. EMs work on the principle of secondary electron emission. When an ion strikes the surface, electrons are ejected and repeatedly accelerated by the potential between the first surface and next surface, releasing more secondary electrons until they reach the collector electrode. EMs achieve an amplification gain of up to ~10^6 of the original ion current, and the final signal is sent to a computer for signal processing.[73]

1.3 PHOTOCHEMICAL CROSSLINKING AND GAS-PHASE EXPERIMENTS

Crosslinking was introduced to map the non-covalent interactions in protein complexes via an indirect approach.[11, 74] It utilizes a crosslinker which is a reactive group used to convert noncovalent interactions to covalent bonds. A traditional crosslinker has two reactive functionalities and a spacer arm that can be adjusted according to the target. Once two sites within a non-covalent complex are crosslinked, the distance between them are known to be within a limit defined by the crosslinker. Therefore, distance constrains in three dimensions are measured and can be used to reconstruct the 3-D structure of the complex. Chemical reaction (i.e. acid/base) based crosslinkers have been widely used but suffer from limitations such as finding reactive residues.[75] Photon-activation based crosslinkers can increase the coverage of crosslink mapping such as membrane-spanning regions, as these crosslinkers react with targets without the limitation of polarity and hydrophilicity.

The chromophores that are frequently used in photo-crosslinking are diazirines, aryl azides and benzophenones. Upon photon activation, each of these moieties generate highly reactive intermediates that quickly react with the target molecule, forming covalent linkages. Among these chromophores, aryl azides are easy to synthesize but require photon activation at wavelength of <300nm which can denature proteins[76]; benzophenones are easy to handle and are resistant to water deactivation but the crosslinker size can sterically alter the non-covalent interactions within the complexes.[77] Diazirine is better suited for the purpose of our study mainly because of its activation wavelength (330-370nm) that is outside the absorption by naturally occurring biomolecules, chemical stability and its small steric hindrance.[78] Diazirines were first introduced as photolabeling reagent by Knowles in the 1970s.[79] When activated with 330-370 nm photons, diazirine eliminates N_2 to generate a highly reactive carbene intermediate. The singlet carbene acts

as the active reagent for crosslinking as it inserts rapidly with any proximate X-H (X= C, N, O) bonds. For the carbenes with adjacent CH groups, the insertion competes with the exothermic carbene isomerization forming unreactive olefins, (Figure 1.11). The sub-microsecond half-lives[80] of carbenes provide a kinetic constraint for the cross-linking reaction that captures the transient conformational change in dynamic systems.

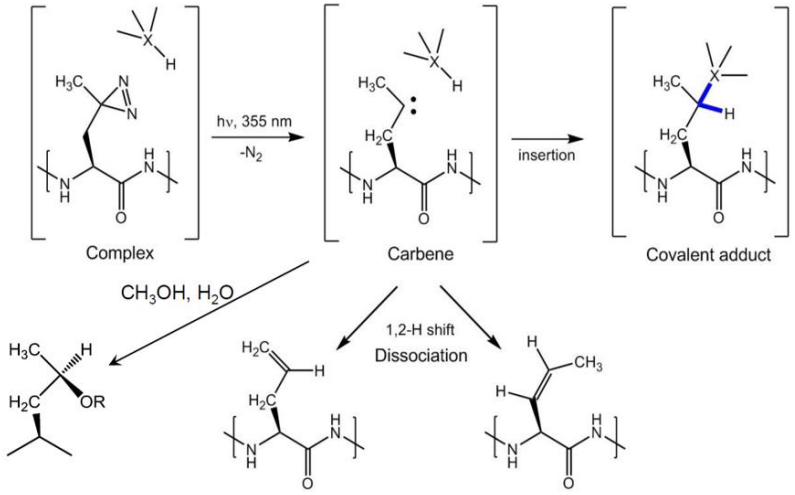


Figure 1.11. Schematics of photoactivated diazirine crosslinking and side reactions.

Applications of diazirine photo-crosslinking have been reported with photo-reactive amino acid analogues that mimic the natural amino acids. In these studies, photoleucine (Leu*) and photomethionine (Met*) have been incorporated into the primary sequence of a protein during a peptide synthesis or protein expression, which enabled interaction of the labeled protein and targeted counterpart without adversely affecting their functionality in the cell. [81] [82] Recently, Pepin *et al.* reported a tunable synthetic diazirine tag that can be selectively incorporated into different peptide residues, as demonstrated below in Figure 1.12 .[83] [84] [85] This tag not only can be used to elucidate protein-protein interactions, but also in probing interactions in DNA-protein and protein-agonist complexes.

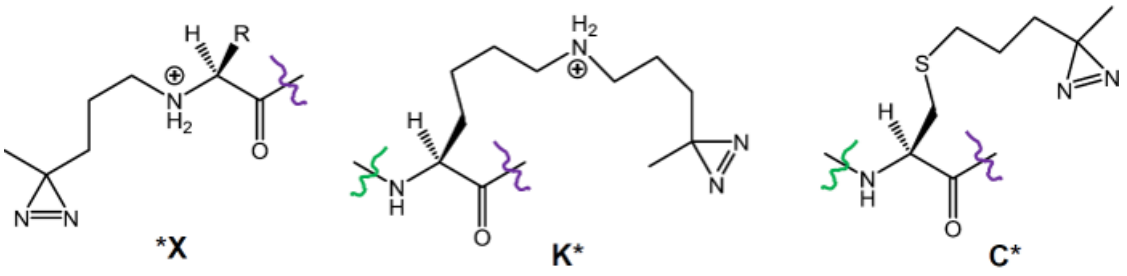


Figure 1.12. The tagging of positional tunable diazirine label
to N-terminus (*X), Lysine side chain (K*) and Cysteine sidechain (C*).

The traditional crosslinking mass spectrometry workflow includes a solution-phase reaction, followed by digestion and enrichment steps to amplify the crosslinked interface.[86 87] This laborious procedure usually requires high concentration (mM to M) of reactants which sometimes leads to non-specific binding. The very low efficiency of these reactions are also described as "finding a needle in the haystack" and requires tedious separations, enrichment, and detection of high sensitivity.

An alternative approach is gas-phase crosslinking. Gas-phase biomolecular ions are analogous to ions existing in solution, as most biomolecules are charged and held closely by intermolecular/intramolecular forces. Gas-phase ions also represent the structure driven by their own conformational preference without interference of solvents and salts. A main subject of this work is based on the techniques of photo-crosslinking in the gas phase, which requires the generation of gaseous biomolecule noncovalent complexes and photo-activation of the embedded photo-crosslinkers in a mass spectrometer. The gas-phase ions are also conveniently modeled by computers, providing a more accurate comparison across theory and experiment. This work is going to introduce the application of these techniques into a larger scope of study such as probing the interaction of physiologically active peptides (Chapter 2), peptide - DNA (Chapter 3) and peptide - agonist (Chapter 6) complexes.

1.4 COMPUTER MODELING AND CALCULATIONS

Gas-phase crosslinking and action spectroscopy experiments provide experimental information that can be complemented by computational tools. Molecular dynamics is a useful tool to track conformation trajectories of non-covalent binding. The calculated absorption spectra are useful in the interpretation of the ion structures and isomer populations that contribute to the experimental action spectra. However, these results need to be carefully evaluated with a full conformation search and high-level thermodynamics calculations in order to consider all possible low-energy isomers at a high-confidence level of theory[88]. In this work, a customized modeling pipeline is described in order to fulfill this task. The pipeline includes the following processes: (1) initial structures guess, (2) extensive search in conformational space, (3) geometry optimization with high-level theory gradient method, (4) thermodynamics calculations and (5a) theoretical absorption spectra calculations or (5b) contact analysis modeling.

1.4.1 *Customized Conformation Search Pipeline*

The conformational space searching process described here starts with building distinct initial isomers with different 3-D structures to provide diverse initial starting points on the potential energy surface (PES). Molecular dynamics (MD) is a computer simulation method that studies nuclear motion in molecules. The trajectories are determined by solving the Newton equation of motion. There are two main families of MD methods: The "classic" mechanics assumes the "ball-stick" model, known as the force field or molecular mechanics (MM) that applies empirical potentials to simulates the nuclear movement at a nanosecond scale. The quantum mechanics (QM) MD simulations take into account the quantum nature of the chemical bond by using quantum

equations to calculate the electron density of valence electrons that determine the bonding, in addition to treating the nuclei with their inner electrons classically. QM simulations are usually run on picosecond time scales, but they are much more computationally expensive than force-field MM. [89] The MD in this work uses Born-Oppenheimer dynamics at an augmented semi-empirical level (PM6-D3H4) of quantum theory including hydrogen bonding and dispersion interactions.[90] This approximation approach is fast and efficient, providing quick sampling of the total energy surface and generates high-quality geometries. Once a short sampling trajectory (e.g. 20 ps, 20000 1-fs steps) is generated, we sample the trajectory in 100-step intervals to generate a subset of 200 snapshots of the 3-D structures for each initial isomer. In this work, each PES sampling requires at least 8 initial isomers, which gives thousands of snapshots. These snapshots are optimized at the same semi-empirical level to the corresponding local minima.

This process generates a large number of 3-D coordinates, and the majority of which are redundant and highly similar in terms of conformation. Due to the limitations of PM6 theory, the overall electronic energy ranking at this level is not reliable, and so a higher-level method (i.e. DFT) with a larger basis-set is required to benchmark the energy levels and select the lowest-energy hypothetical structures. However, it is not practical to perform high-level geometry optimizations for all these structures since QM calculations are expensive and time-consuming. A customized conformation clustering method is therefore implemented in order to extract “useful information”, while reducing the conformation pool from thousands of candidates into a smaller set containing a minimum amount (60-80) of distinct and representative conformational clusters. This reduced set improves the efficiency of further investigations of downstream geometry optimizations and thermodynamic calculations with high level functionals.

The method is based on a machine learning technique called K-Means clustering, a center-based partitioning clustering method. In this algorithm, K conformations as cluster centroids are initialized randomly and the remaining conformations are assigned to their closest cluster centroid. The algorithm aims to partition all structures into K non-overlapping clusters (C_1, C_2, …, C_k) such that the sum of squares of distances of each structure x to its cluster centroid (geometric means, μ_i) is minimized. The objective function is written as:

$$\min \sum_{i=1}^{K} \sum_{x \in C_i} ||x - \mu_i||^2 \quad (1.1)$$

Where x denotes the representation vector of a structure within the *i*th cluster C*i,* and μ_i is the corresponding cluster centroid. The final state (equilibrium state) is achieved through several iterations of dynamic partitioning and updates, until the above function reaches convergence. [91]

Since the subject of this work is biomolecules such as peptides and DNA, all isomers are grouped into clusters based on their secondary and super secondary structures. For each structure, the atoms that define the skeleton are selected as "key atoms". Take a peptide as an example: C_α, C_β and all heavy atoms of the side chain will be selected; if a rigid functional group is present, only three atoms that define the plane are necessary. Combining the selections, each structure is then represented as a vector of intramolecular distances of "key atoms" to counter the effect of rotation and translation. Without making any pre-assumption of the number of clusters in the pool, the best number of clusters is selected based on the "elbow method" approach [92]. A self-defined score referred to as "root mean inertia (sum of squared distances of samples to their closest cluster center) score" is used to evaluate the performance of different Ks of choice. The smaller the value, the closer each structure is to its cluster centroid. A wide range of Ks (1, 2 , 3, …m; m being the maximum number of cluster possible in the pool) are explored and the corresponding "root mean inertia score" is plotted as a function of K, in order to observe the downward elbow-trend of the

score. The K at the elbow is the point where adding an extra cluster will not give a better "grouping" of the molecular structures and is considered the optimal K. A detailed implementation code can be found in the Appendix.

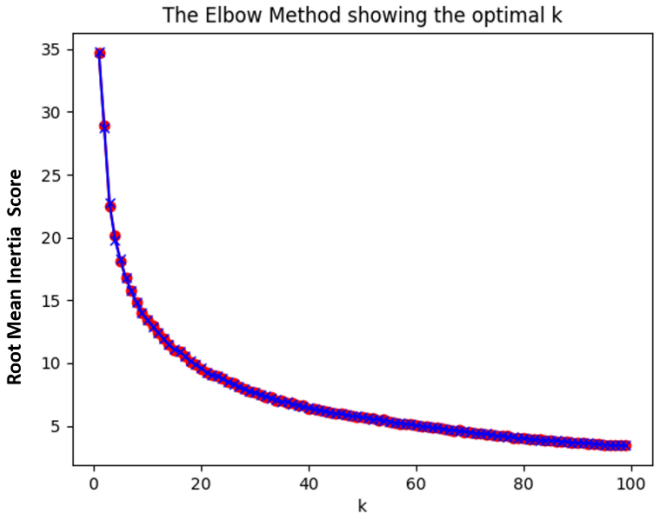


Figure 1.13. The elbow method showing the decreasing trend of score vs. number of clusters from 400 structures as an example, the optimal K here is 14.

Since hydrogen bonds are critical in the stabilization of secondary structures, the hydrogen bonding pattern (HB pattern) is used as an extra feature to represent each structure within its cluster. Therefore, structures within each cluster can then be further classified or filtered by the similarity of HB pattern. The HB pattern can be auto-generated with the same script found in Appendix.

1.4.2 *Geometry Optimization and Thermodynamics Calculations*

The next step is to perform geometry optimizations with density functional theory (DFT) calculations for selected structures, with the goal of finding the equilibrium geometry (local energy minimum) of each pre-selected structure as a new initial guess. These points are determined by the

condition that the first derivatives of potential energy with respect to the nuclei coordinates vanish (i.e. the total force acting on each nucleus vanishes) and the second derivatives are all positive at local minima. The DFT hybrid functionals (i.e. ωB97X-D [93], M06-2X [94]) used in this work accounts for dispersion interactions. The electronic, vibrational, and rotational energies and entropies are derived from the harmonic frequency calculations of the optimized structures. The combined electronic, enthalpy, and entropy terms were used to produce free energies at the experimental temperature which are ranked and filtered.

1.4.3 *Contact Analysis*

For the work described in Chapter 2, 3 and 6 where we investigate the noncovalent interactions in biomolecular complexes, the selected low energy isomers are subjected to contact analysis. The modeled "contact" is defined as a distance of two atoms in the optimized structure, or as a result of thermal motion that is less than or equal to to the sum of their van der Waals radii. The goal is to generate a 100 ps MD trajectory to mimic the conformational motion of the photo-sensitive tag on a time scale which is similar to the half-life of the transient carbene in the gas-phase. The total contact ratio is defined as the number of contacts between the diazirine carbon to any proximal atoms that are available for carbene insertion divided by the total number of steps in the MD trajectory. The calculated contacts are then compared with the experimentally determined cross-link sites. [83 84 95 96]

1.4.4 *Calculated Absorption Spectra*

For the work described in Chapters 4 and 5, where we acquire the experimental action spectra on DNA dinucleotide radicals and calculate thermodynamically-favored geometries in the gas-phase, we employ Time-dependent density functional theory (TD-DFT) calculations to

generate calculated absorption spectra to provide deeper understanding of the action spectra. The calculated excitation energies and oscillator strength generate the resonant lines, which are converted to the theoretical absorption spectra by convoluting with an artificial pre-fixed Lorentzian broadening function. The resulting spectra are compared the experimental action spectra. These calculated absorption bands correspond to vertical excitations at 0 K whereas the experiments are conducted at 300 K. Therefore, vibronic broadening due to electronic excitations between vibrationally states that are populated at the experimental temperature is expected to broaden and red-shift the resonant lines by 10-30 nm. [97] [68] [98]

1.5 Enzymatic Assay for Inborn Error Screening

In addition to my work on the structural elucidation of biomolecules outlined in this dissertation, my interest in other applications of mass spectrometry also motivated me to participate in projects involving the development of enzymatic assays for the early screening and diagnostics of inborn errors of metabolism using tandem mass spectrometry. This work has been published,[99, 100] but because it does not fit the main topic of this dissertation, further discussion is excluded from the dissertation.

1.6 References:

1. Hashimoto, M.; Rockenstein, E.; Crews, L.; Masliah, E., *Neuromolecular Med* **2003,** *4* (1-2), 21-36.
2. 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., *Nature Chemical Biology* **2011,** *7*, 285.
3. Mauser, H.; Guba, W., (1367-6733 (Print)).
4. Klebe, G., *Journal of Molecular Medicine* **2000,** *78* (5), 269-281.
5. Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T., *Nature chemistry* **2011,** *3* (2), 172-177.
6. McPherson Jr, A., *Methods of Biochemical Analysis* **1976**, 249-345.
7. Fraser, J. S.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T., *Proceedings of the National Academy of Sciences* **2011,** *108* (39), 16247.
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. Taraska, J. W., *Current opinion in structural biology* **2012,** *22* (4), 507-513.
10. Konermann, L.; Tong, X.; Pan, Y., *Journal of Mass Spectrometry* **2008,** *43* (8), 1021-1036.

11. Singh, P.; Panchaud, A.; Goodlett, D. R., *Analytical Chemistry* **2010,** *82* (7), 2636-2642.
12. Mulder, F. A. A.; Filatov, M., *Chem. Soc. Rev.* **2010,** *39* (2), 578-590.
13. Munoz, V.; Eaton, W. A., *Proceedings of the National Academy of Sciences* **1999,** *96* (20), 11311-11316.
14. Holm, L.; Kaariainen, S.; Rosenstrom, P.; Schenkel, A., *Bioinformatics* **2008,** *24* (23), 2780-2781.
15. Waga, S.; Hannon, G. J.; Beach, D.; Stillman, B., *Nature* **1994,** *369* (6481), 574-578.
16. Milani, M.; Mastrangelo, E.; Bollati, M.; Selisko, B.; Decroly, E.; Bouvet, M.; Canard, B.; Bolognesi, M., *Antiviral Research* **2009,** *83* (1), 28-34.
17. El Far, O.; Betz, H., *Biochemical Journal* **2002,** *365* (2), 329-336.
18. 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., *Cell* **2016,** *165* (7), 1698-1707.
19. Sharon, M.; Robinson, C. V., *Annual Review of Biochemistry* **2007,** *76* (1), 167-193.
20. Fields, S.; Song, O.-k., *Nature* **1989,** *340* (6230), 245-246.
21. Gingras, A.-C.; Gstaiger, M.; Raught, B.; Aebersold, R., *Nature Reviews Molecular Cell Biology* **2007,** *8* (8), 645-654.
22. Boeri Erba, E.; Petosa, C., *Protein Science* **2015,** *24* (8), 1176-1192.
23. Dormandy, T. L., *Annals of the Royal College of Surgeons of England* **1980,** *62* (3), 188-194.
24. Stubbe, J.; van der Donk, W. A., *Chemical Reviews* **1998,** *98* (2), 705-762.
25. Florence, T. M., *Australian and New Zealand Journal of Ophthalmology* **1995,** *23* (1), 3-7.
26. Balasubramanian, B.; Pogozelski, W. K.; Tullius, T. D., *Proceedings of the National Academy of Sciences* **1998,** *95* (17), 9738.
27. O'Neill, M. A. B., J. K., Sequence-Dependent DNA Dynamics: The Regulator of DNA-Mediated Charge Transport. In *Charge Transfer in DNA*, Wagenknecht, H. A., Ed. Wiley-VCH: Weinheim, Germany, 2005; pp 27-31.
28. Parker, A. W.; Lin, C. Y.; George, M. W.; Towrie, M.; Kuimova, M. K., *The Journal of Physical Chemistry B* **2010,** *114* (10), 3660-3667.
29. Bachler, V.; Hildenbrand, K., *International Journal of Radiation Applications and Instrumentation. Part C. Radiation Physics and Chemistry* **1992,** *40* (1), 59-68.
30. Close, D. M.; Crespo-Hernández, C. E.; Gorb, L.; Leszczynski, J., *The Journal of Physical Chemistry A* **2008,** *112* (49), 12702-12706.
31. Yao, C. X.; Cuadrado-Peinado, M. L.; Polasek, M.; Turecek, F., *Angewandte Chemie-International Edition* **2005,** *44* (41), 6708-6711.
32. Medved, M., *Rapid Communications in Mass Spectrometry* **1991,** *5* (1), 11-14.
33. Jackson, G. P.; Barkett, M. A., A History of the Forensic Applications of Mass Spectrometry. In *The Encyclopedia of Mass Spectrometry*, Gross, M. L.; Caprioli, R. M., Eds. Elsevier: Boston, 2016; pp 271-284.
34. Gelb, M. H.; Turecek, F.; Scott, C. R.; Chamoles, N. A., *Journal of Inherited Metabolic Disease* **2006,** *29* (2-3), 397-404.
35. Thomson, J. J., *Rays of positive electricity and their application to chemical analyses*. Longmans, Green and Company: 1921; Vol. 1.
36. McLafferty, F. W.; Turecek, F., *Interpretation of Mass Spectra Forth Edition.* 1993; p 1-17.
37. Munson, M. S. B.; Field, F. H., *Journal of the American Chemical Society* **1966,** *88* (12), 2621-2630.
38. Munson, B., *Analytical Chemistry* **1971,** *43* (13), 28A-43a.
39. Martens, J.; Berden, G.; Oomens, J., *Analytical Chemistry* **2016,** *88* (12), 6126-6129.
40. Yamashita, M.; Fenn, J. B., *The Journal of Physical Chemistry* **1984,** *88* (20), 4451-4459.
41. Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L., *The Clinical biochemist. Reviews* **2003,** *24* (1), 3-12.
42. Smith, R. D.; Light-Wahl, K. J.; Winger, B. E.; Loo, J. A., *Organic Mass Spectrometry* **1992,** *27* (7), 811-821.
43. Karpievitch, Y. V.; Polpitiya, A. D.; Anderson, G. A.; Smith, R. D.; Dabney, A. R., *The annals of applied statistics* **2010,** *4* (4), 1797-1823.
44. Nordström, A.; O'Maille, G.; Qin, C.; Siuzdak, G., *Analytical Chemistry* **2006,** *78* (10), 3289-3295.
45. Wilm, M., *Molecular & Cellular Proteomics* **2011,** *10* (7), M111.009407.
46. Kruve, A.; Kaupmees, K.; Liigand, J.; Oss, M.; Leito, I., *Journal of Mass Spectrometry* **2013,** *48* (6), 695-702.
47. El-Aneed, A.; Cohen, A.; Banoub, J., *Applied Spectroscopy Reviews* **2009,** *44* (3), 210-230.
48. Glish, G. L.; Burinsky, D. J., *Journal of the American Society for Mass Spectrometry* **2008,** *19* (2), 161-172.
49. Støchkel, K.; Milne, B. F.; Nielsen, S. B., *The Journal of Physical Chemistry A* **2011,** *115* (11), 2155-2159.
50. March, R. E.; Todd, J. F., *Quadrupole ion trap mass spectrometry*. John Wiley & Sons: 2005; Vol. 165.
51. March, R. E., *Journal of Mass Spectrometry* **1997,** *32* (4), 351-369.
52. Schwartz, J. C.; Senko, M. W.; Syka, J. E. P., *Journal of the American Society for Mass Spectrometry* **2002,** *13* (6), 659-669.
53. Sleno, L.; Volmer, D. A., *Journal of Mass Spectrometry* **2004,** *39* (10), 1091-1112.
54. Biemann, K., *Biomedical & environmental mass spectrometry* **1988,** *16* (1-12), 99-111.
55. Wu, J.; McLuckey, S. A., *International Journal of Mass Spectrometry* **2004,** *237* (2-3), 197-241.
56. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., *Proceedings of the National Academy of Sciences* **2004,** *101* (26), 9528-9533.
57. Wiesner, J.; Premsler, T.; Sickmann, A., *PROTEOMICS* **2008,** *8* (21), 4466-4483.
58. Turecek, F.; Julian, R. R., *Chemical Reviews* **2013,** *113* (8), 6691-6733.
59. Hari, Y.; Leumann, C. J.; Schürch, S., *Journal of The American Society for Mass Spectrometry* **2017,** *28* (12), 2677-2685.
60. Lyon, Y. A.; Beran, G.; Julian, R. R., *Journal of The American Society for Mass Spectrometry* **2017,** *28* (7), 1365-1373.
61. Polfer, N. C.; Dugourd, P., *Laser photodissociation and spectroscopy of mass-separated biomolecular ions*. Springer: 2013; Vol. 83.
62. Dang, A.; Korn, J. A.; Gladden, J.; Mozzone, B.; Tureček, F., *Journal of The American Society for Mass Spectrometry* **2019,** *30* (9), 1558-1564.
63. R. Julian, R., *Journal of The American Society for Mass Spectrometry* **2017,** *28* (9), 1823-1826.
64. Kim, T.-Y.; Thompson, M. S.; Reilly, J. P., *Rapid Communications in Mass Spectrometry* **2005,** *19* (12), 1657-1665.
65. Wilson, J. J.; Brodbelt, J. S., *Analytical Chemistry* **2007,** *79* (20), 7883-7892.
66. Cismesia, A. P.; Bailey, L. S.; Bell, M. R.; Tesler, L. F.; Polfer, N. C., *Journal of The American Society for Mass Spectrometry* **2016,** *27* (5), 757-766.

67. Filippi, A.; Fraschetti, C.; Rondino, F.; Piccirillo, S.; Steinmetz, V.; Guidoni, L.; Speranza, M., *International Journal of Mass Spectrometry* **2013,** *354*, 54-61.
68. Korn, J. A.; Urban, J.; Dang, A.; Nguyen, H. T. H.; Turecek, F., *Journal of Physical Chemistry Letters* **2017,** *8* (17), 4100-4107.
69. Polfer, N. C. D., P., Laser Photodissociation and Spectroscopy of Mass Separated Biomolecular Ions. In *Lecture Notes in Chemistry*, Springer: Cham, 2013; Vol. 83, pp 13-20.
70. Marcum, J. C.; Halevi, A.; Weber, J. M., *Phys Chem Chem Phys* **2009,** *11* (11), 1740-51.
71. Liu, Y.; Korn, J. A.; Dang, A.; Turecek, F., *Journal of Physical Chemistry B* **2018,** *122* (42), 9665-9680.
72. Dang, A.; Liu, Y.; Turecek, F., *Journal of Physical Chemistry A* **2019,** *123* (15), 3272-3284.
73. Dietz, L. A., *Review of Scientific Instruments* **1965,** *36* (12), 1763-1770.
74. Sinz, A., *Analytical and Bioanalytical Chemistry* **2005,** *381* (1), 44-47.
75. Belsom, A.; Mudd, G.; Giese, S.; Auer, M.; Rappsilber, J., *Analytical Chemistry* **2017,** *89* (10), 5319-5324.
76. A. Fleming, S., *Tetrahedron* **1995,** *51* (46), 12479-12520.
77. Weber, P. J. A.; Beck-Sickinger, A. G., *The Journal of Peptide Research* **1997,** *49* (5), 375-383.
78. Das, J., *Chemical Reviews* **2011,** *111* (8), 4405-4417.
79. Bayley, H.; Knowles, J. R., *Biochemistry* **1978,** *17* (12), 2420-2423.
80. Pezacki, J. P.; Couture, P.; Dunn, J. A.; Warkentin, J.; Wood, P. D.; Lusztyk, J.; Ford, F.; Platz, M. S., *Journal of Organic Chemistry* **1999,** *64* (12), 4456--4464.
81. Suchanek, M.; Radzikowska, A.; Thiele, C., *Nature Methods* **2005,** *2* (4), 261-268.
82. Vila-Perelló, M.; Pratt, M. R.; Tulin, F.; Muir, T. W., *Journal of the American Chemical Society* **2007,** *129* (26), 8068-8069.
83. Pepin, R.; Shaffer, C. J.; Turecek, F., *Journal of Mass Spectrometry* **2017,** *52* (8), 557-560.
84. Nguyen, H. T. H.; Andrikopoulos, P. C.; Rulek, L.; Shaffer, C. J.; Tureek, F., *Journal of the American Society for Mass Spectrometry* **2018,** *29* (8), 1706--1720.
85. Liu, Y.; Ramey, Z.; Turecek, F., *Chemistry - A European Journal* **2018,** *24* (37), 9259--9263.
86. Yu, C.; Huang, L., *Analytical Chemistry* **2018,** *90* (1), 144-165.
87. Sinz, A., *Angewandte Chemie International Edition* **2018,** *57* (22), 6390-6396.
88. Wodrich, M. D.; Corminboeuf, C.; Schreiner, P. R.; Fokin, A. A.; Schleyer, P. v. R., *Organic Letters* **2007,** *9* (10), 1851-1854.
89. Alder, B. J.; Wainwright, T. E., *The Journal of Chemical Physics* **1959,** *31* (2), 459-466.
90. Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P., *Journal of Chemical Theory and Computation* **2009,** *5* (7), 1749-1760.
91. Forgy, E., *Biometrics* **1965,** *21*, 768-780.
92. Thorndike, R. L., *Psychometrika* **1953,** *18* (4), 267-276.
93. Chai, J. D.; Head-Gordon, M., *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615--6620.
94. Zhao, Y.; Truhlar, D. G., *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.
95. Shaffer, C. J.; Andrikopoulos, P. C.; Rezac, J.; Rulisek, L.; Turecek, F., *J Am Soc Mass Spectrom* **2016,** *27* (4), 633-45.
96. Huang, S. R.; Liu, Y.; Turecek, F., *Phys Chem Chem Phys* **2019,** *21* (4), 2046-2056.
97. Lesslie, M.; Lawler, J. T.; Dang, A.; Korn, J. A.; Bím, D.; Steinmetz, V.; Maître, P.; Tureček, F.; Ryzhov, V., *ChemPhysChem* **2017,** *18* (10), 1293-1301.
98. Barbatti, M.; Ruckenbauer, M.; Plasser, F.; Pittner, J.; Granucci, G.; Persico, M.; Lischka, H., *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014,** *4* (1), 26-33.
99. Liu, Y.; Yi, F.; Kumar, A. B.; Chennamaneni, N. K.; Hong, X.; Scott, C. R.; Gelb, M. H.; Turecek, F., *Clinical chemistry* **2017,** *63* (6), 1118-1126.
100. Khaledi, H.; Liu, Y.; Masi, S.; Gelb, M. H., *Analytical chemistry* **2018,** *90* (20), 12168-12171.

Chapter 2. NON-COVALENT INTERACTIONS OF A NEUTROPROTECTIVE PEPTIDE REVEALED BY PHOTODISSOCIATIVE CROSS-LINKING IN THE GAS PHASE

Reproduced in part with permission from Liu, Y.; Ramey, Z.; Turecek, F., *Chemistry - A European Journal* **2018,** *24* (37), 9259--9263.

Abstract

The small neuroprotective peptide Cys-Ala-Gln-Lys (CAQK) was selectively tagged with a diazirine group and its photochemical cross-linking was studied with model target peptides. Experimental results in combination with Born–Oppenheimer molecular dynamics revealed the structural preferences for binding to the amino acid residues in the target peptides. Effects of the target peptide sequence and N- and C-terminal modifications are discussed.

2.1 INTRODUCTION

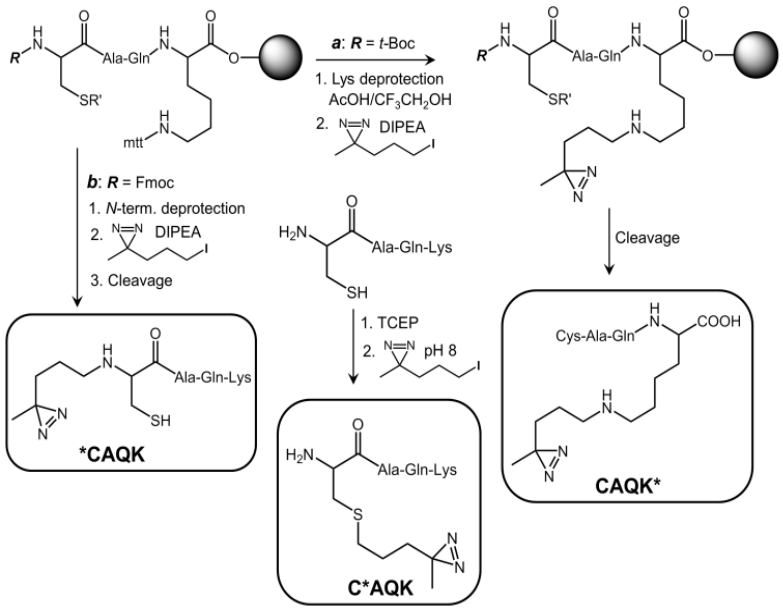
Non-covalent interactions of small molecules with proteins underlie many biological processes such as enzyme inhibition, immune recognition, and drug binding, among others. High resolution methods such as X-ray diffraction crystallography, NMR spectroscopy or, lately, cryo-electron microscopy [1, 2] have been used to establish the macromolecular structure and binding sites in non-covalent complexes. However, newly discovered biological effects must typically wait for several years before being studied using high-resolution structure methods. A recent example is the small peptide Cys-Ala-Gln-Lys (CAQK) that has been reported to exhibit unexpected neuroprotective effects in alleviating brain trauma in test animals[3]. Neither the binding site of CAQK nor the involved proteins are known accurately, with chondroitin sulfate proteo-glycan 4

precursor (CSPG4 RAT)[4] having been suggested as the possible binding target. Whereas binding to the protein target is the ultimate goal in structure biology, model studies of selected peptide binding motifs are valuable because they simplify the problem and deal with well-defined components. [5, 6] Non-covalent binding of small peptides in charged complexes can be enhanced in droplets produced by electrospray ionization at the level where bulk solution complexes may be too weak to be analyzed. The non-covalent complexes that are transferred to the gas phase can be studied by photo-dissociative cross-linking in combination with tandem mass spectrometric sequence analysis of covalent cross-links that provides the cross-link positions.[7] This approach allows one to study non-covalent interactions with small peptides of selected binding motifs in proteins We now apply this methodology to study binding in complexes of CAQK with peptide motifs LLSPGH, ALLVRST, and FGENHL from the respective (2–7), (20–26), and (33–38) positions on the N-terminal chain of CSPG4 RAT.[3]

2.2 METHODS

Our approach relies on the specific position labeling of CAQK with the photoactive diazirine group[8] that, upon photodissociation at 355 nm, eliminates N_2, forming a reactive carbene. The carbene can undergo insertion into proximate X-H bonds (X=C, N, O, S) or self-destruct by 1,2-H migration, forming unreactive olefins.[9] The nanosecond kinetics of the 1,2-H migration is commensurable with the time scale for the conformational motion in thermal gas-phase peptide–peptide complexes. This allows one to survey the contact points between the incipient carbene and the target peptide X-H bonds using all-valence-electron molecular dynamics and to employ theory to aid interpretation of experimental cross-linking data. CAQK peptides position-tagged at the N-terminus, Cys thiol, and the Lys ε-NH2 group (denoted as *CAQK, C*AQK, and CAQK*, respectively), were synthesized by a combination of solid-phase and solution methods[10] (Scheme

2.1; for synthetic chemistry details, see the Supporting Information) and characterized by high-resolution tandem mass spectrometry. Selective S-alkylation in C*AQK was further confirmed by the absence of Cys thiol oxidation with air in the alkylated peptide.



Scheme 2.1. Synthetic steps for peptide diazirine tagging.

2.3 Results and Discussions

Non-covalent complexes were generated as singly charged ions by electrospray ionization of a mixture of the photopeptide and the target peptide, as shown for $(C^*AQK+LLSPGH+ H)^+$ (*m/z* 1167, Figure 2.1 a). The dimer ions are denoted by $(M^*+m)^+$ where M* and m stand for the photo-peptide and target peptide, respectively. The mass-selected complex (Figure 2.1 b) was photodissociated at 355 nm (Figure 2.1 c) to form the $(C^*AQK\text{-}N_2+LLSPGH+H)^+$ ion at m/z 1139, which is denoted as $(M+m)^+$. In addition to the loss of N_2, photodissociation also triggered complex dissociation to LLSPGH and $(M^*\text{-}N_2)$ (denoted as M) that appeared as singly charged ions at m/z 623 and 517, respectively, competing for the available proton. The peptide monomers represented

the fraction of $(M+m)^+$ that did not undergo covalent cross-linking upon photodissociation. This is further illustrated by the collision-induced dissociation tandem mass spectrum (CID-MS^2) of $(M^*+m)^+$ (Figure 2.1 d) that showed a predominant dissociation to the LLSPGH and C*AQK monomers, but negligible loss of N2 from the complex. The minor fragment ion at m/z 517 (Figure 2.1 d) was formed by non-photochemical loss of N_2 from vibrationally excited $(M^*+ H)^+$ ions. The relative intensities in the CID-MS^2 spectra of $(M^*+H)^+$ and $(m+H)^+$ ions were used to estimate the relative basicities of M* and m and protonation sites in the complexes, as described in detail in the Supporting Information (Table S1). The data on proton distribution in the complexes were then used for initial guesses of ion structures, as described below.

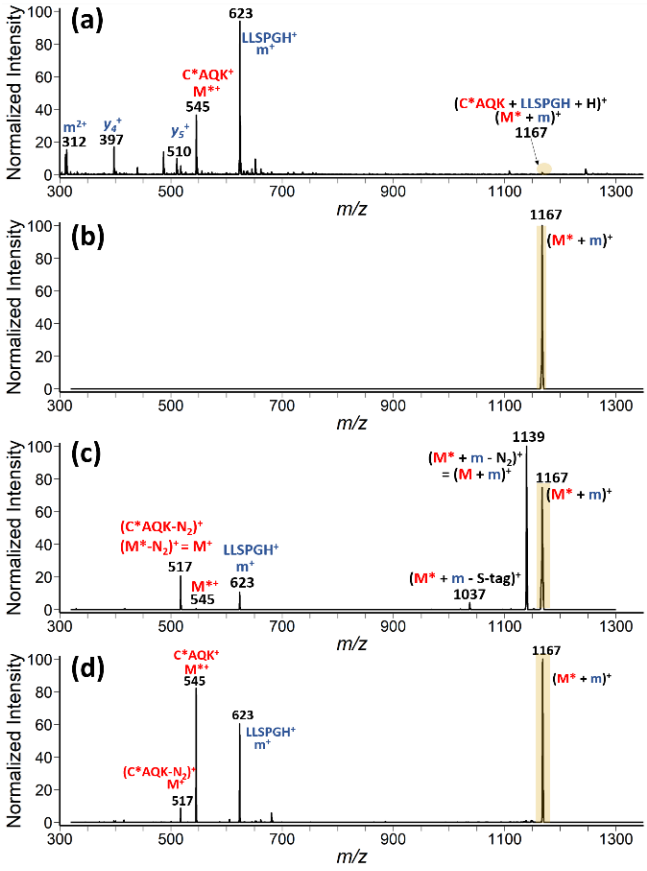


Figure 2.1. (a) Mass spectrum of the $(C^*AQK + LLSPGH + H)^+$ complex ion. (b) MS^2 isolation of the complex. (c) MS^2 photodissociation spectrum of the complex. (d) MS^2 collision-induced dissociation of the complex.

The $(M+m)^+$ ions from photodissociation of $(C^*AQK+ LLSPGH+H)^+$ were further analyzed by tandem CID-MS^3 giving the spectrum shown in Figure 2.2. The UVPD-CID-MS^3 spectra of the other combinations of M* and m peptides are depicted in Figures S1–S8 (Supporting Information).

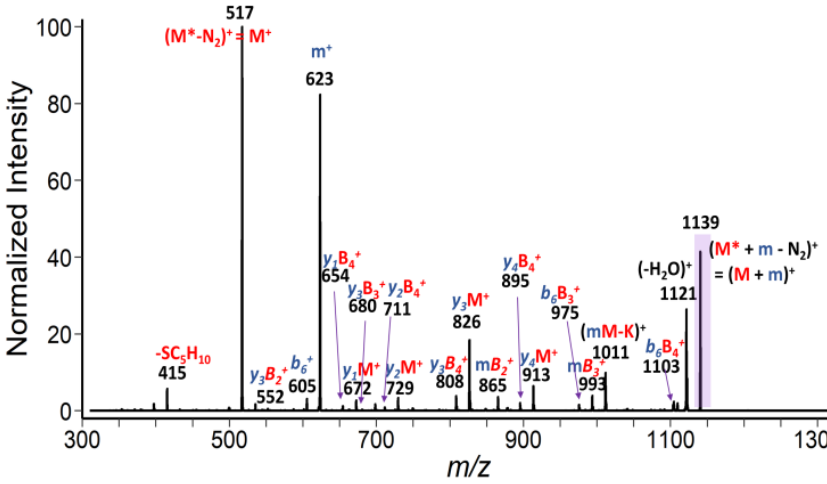


Figure 2.2. CID-MS^3 spectrum of $(C^*AQK - N_2 +LLSPGH + H)^+$.

In addition to dissociation to monomeric peptides, the CID-MS^3 displayed fragment ions originating from covalent bond cleavages leading to loss of water and backbone fragment ions. All these dissociations indicated covalently cross-linked complexes. The facile distinction of non-covalent and crosslinked complexes by UVPD-MS^2 and CID-MS^3 allowed us to quantitate the mole fractions of both types. Table 2.1 summarizes the mole fractions (R_{MS2}) of stable $(M+m)^+$ ions produced by UVPD-MS^2. The covalent fractions are expressed by R_{MS3}, and the total efficiency of covalent cross-linking is given as $R_{tot}=R_{MS2} \times R_{MS3}$. The data indicate effects on R_{MS2}, R_{MS3}, and R_{tot} of both the position of the photo-active tag and the structure modification in the target peptide by C-terminal amidation and N-terminal acetylation.

Table 2.1. Crosslinking efficiencies for LLSPGH

Table 1. Cross-linking efficiencies for LLSPGH.

Target peptide	C*AQK			*CAQK			CAQK*		
	R_{MS2}	R_{MS3}	R_{tot}	R_{MS2}	R_{MS3}	R_{tot}	R_{MS2}	R_{MS3}	R_{tot}
LLSPGH	0.78	0.28	0.21	0.70	0.22	0.15	0.26	0.10	0.026
LLSPGH-NH_2	0.62	0.04	0.02	0.46	0.02	0.01	0.18	0.08	0.015
N-Ac-LLSPGH	0.66	0.22	0.15	0.68	0.14	0.10	0.25	0.12	0.030

These modifications were primarily used to exclude zwitterionic tautomers of the target peptide. In particular, C-terminal amidation in LLSPGH resulted in a substantial decrease of R_{MS2} in complexes with C*AQK, *CAQK, and CAQK*, whereas N-terminal acetylation had a notably weaker effect (Table 2.1). Similar effects were observed for R_{tot} quantitating the covalent crosslink fractions. The effect of C-terminal amidation was most pronounced for *CAQK (15-fold decrease of R_{tot}) and less so for CAQK*. The tag position also affected the R_{tot} where the highest overall cross-linking efficiency (21%) was found for Cys-S-tagged peptide C*AQK. It should be noted that the quantitation with R_{MS3} may underestimate those covalent cross links that are vulnerable to CID, such as ester groups formed by carbene insertion into the carboxyl O-H.[7] Thus, the R_{MS3} and R_{tot} efficiencies must be viewed as lower bounds of the actual covalent cross-links in complexes from these peptide combinations.

The CID-MS^3 spectra provided information on the residue specific cross-linking through sequence analysis of backbone dissociations of the cross-linked complexes. Before discussing the dissociations, we briefly describe the cross-linked fragment ion notation,[7] which is based on the standard Biemann–Roepstorff–Fohlman nomenclature of peptide fragment ions.[11, 12] The target peptide and associated backbone fragments are labeled by a lowercase letter (m, b_n, and y_n), whereas the photo-peptide and its backbone fragments are labeled by capital letters (M, B_n, and

Y_n). The combination of fragments from both the target and photo-peptides is labeled according to the observed *m/z*. A simple representation of this naming system is illustrated in Figure 2.2 for the dimer complex $(C^*AQK\text{-}N_2+ LLSPGH+H)^+$. For example, a photo-cross-linked fragment ion containing the C-terminal y_3 segment (PGH) of LLSPGH and the N-terminal B_2 segment from $C^*\text{-}N_2AQK$ ($C^*\text{-}N_2A$) is denoted as y_3B_2, and likewise for the other residue combinations.

Analysis of the backbone fragment ions of $(LLSPGH+ C^*AQK\text{-}N_2+H)^+$ produced by UVPD-CID-MS^3 provided information regarding the possible sites of contact between the transient carbene and X-H bonds in the target peptide amino acid residues that led to cross-linking. For example, the presence of the y_1M fragment ion (m/z 672, Figure 2.2) indicated a covalent bond formed by cross-linking the photopeptide at the His-6 residue. The y_2M (m/z 729) and y_3M (m/z 826) fragment ions indicated covalent bond formation at the GH and PGH residues without further distinction. The data indicated predominant interactions with the C-terminal residues of the target peptides. Topical analysis of cross-linking was possible for the LLS and PGH segments because of the proline effect that competitively promotes backbone dissociations at the N-terminal side of the Pro residue.[13, 14] In contrast, residue specific distinction within the LLS and PGH segments was less reliable for the same reason.

The results of sequence analysis are summarized in Figure 2.3 a-c. Cross-linking in LLSPGH occurred predominantly (>78%) near the C-terminus for all three photo-peptides (Figure 2.3 a). With C*AQK specifically, the distribution was 14% and 86% in LLS and PGH, respectively. Photo-peptides with diazirine tags at the N-terminus and Cys side chain gave nearly identical results, indicating very similar contacts with the LLSPGH residues. The Lys-tagged photo-peptide showed a higher fraction of cross-links with the N-terminal residues of LLSPGH. Crosslinking of $LLSPGH\text{-}NH_2$ resulted in an increased participation of the N-terminal residues

compared to those in the carboxyl- terminated target peptide (Figure 2.3 b). For example, the distribution of cross-links with C*AQK in the LLS segment (42%) was close to that for PGH-NH_2 (58 %), and a similar breakdown, 34:66, was obtained for cross-links with *CAQK. Cross-linking with CAQK* displayed a more pronounced bias toward the C-terminal residues (PGH-NH_2), which, when combined, amounted to 74% of cross-links. Interactions of N-acetyl- LLSPGH with all three photo-peptides resulted in prevalent (>75%) cross linking in the C-terminal PGH residues (Figure 2.3 c). Interestingly, no substantial enhancement of cross-linking at the N-terminus was observed for CAQK* in this case (Figure 2.3 c). Cross-linking efficiencies and cross-link distributions were also determined for the other two peptide sequence analogs, ALLVRSTP and SSFGENHL, and the data are presented and discussed in the Supporting Information (Figures S9 and S10).

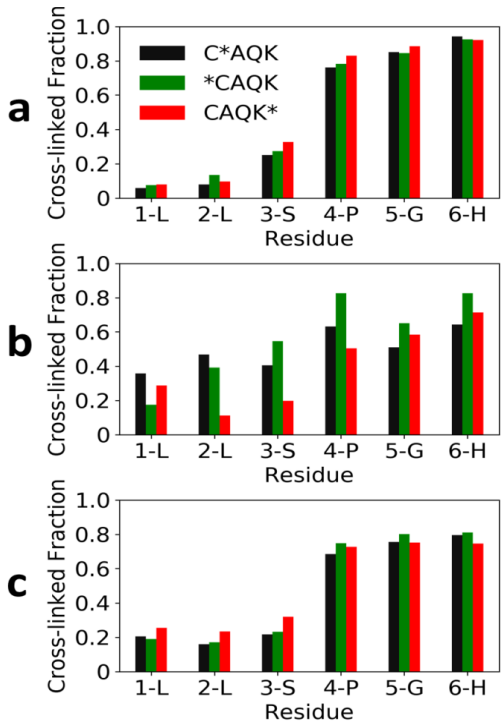


Figure 2.3. Cross-link distributions in (a) LLSPGH, (b) LLSPGH-NH_2, and (c) N-Ac-LLSPGH.

The above-described results illustrated the range of non-covalent binding and cross-linking in complexes composed of different photo- and target peptides. To gain insight into the structural basis of non-covalent interactions in the above-discussed complexes, we carried out a detailed computational study of the $(C^*AQK+LLSPGH+H)^+$ system. Initial ion structures were selected with different distributions of charge sites. Zwitterionic initial structures included two protonation sites at the photo-peptide Lys, target peptide His or either peptide N-terminus in combination with a carboxylate anion to yield the overall charge state +1. Canonical initial ion structures were protonated at the photo-peptide Lys or target peptide His residues. Born–Oppenheimer molecular dynamics (BOMD) with PM6-D3H4[15] at 410 K for 20 ps resulted in a substantial consolidation of the initial protomers to a few structure types, and the fifteen lowest-energy structures were further refined by full DFT optimization with ωB97X-D/6–31+G(d,p)[16] (Table S3). The C*AQK conformations in the complexes were compared to those of free $(C^*AQK+H)^+$ and $(CAQK+H)^+$ ions to assess the effect of the diazirine tag and binding on the CAQK ion secondary structure (Figure 2.4). The optimized structures of low-energy conformers indicated that the S-azipentyl substituent had only a small effect on the ion conformation, as both $(CAQK+H)^+$ and $(C^*AQK+H)^+$ displayed dominant internal solvation of the lysine NH_3 groups. This finding was important to show that tagging the CAQK peptide did not substantially affect its conformational preferences. In contrast, C*AQK noncovalent binding in the complexes was realized in a number of conformations, as illustrated by the lowest free-energy conformers **1–3** (Figure 2.4; for other, higher free-energy conformer structures, see Table S3 and Figure S11). It is of note that the free-energies of complexes **1** and **2** with the canonical form of LLSPGH and that of **3** with a zwitterionic form of the target peptide were within 7 $kJmol^{-1}$, which can be considered equivalent at the DFT level of theory, and there were five more complexes within 20 $kJmol^{-1}$ (Table S3). This indicated

a considerable potential structure diversity of the non-covalent complexes in the gas phase even for local energy minima representing 0 K structures. To investigate the conformations in thermal gas-phase complexes, we ran long (100 ps) BOMD trajectory calculations at 310 K for eight lowest free-energy complex conformers to determine close contacts between the incipient carbene atom and X-H bonds in the target peptide resulting from thermal motion.

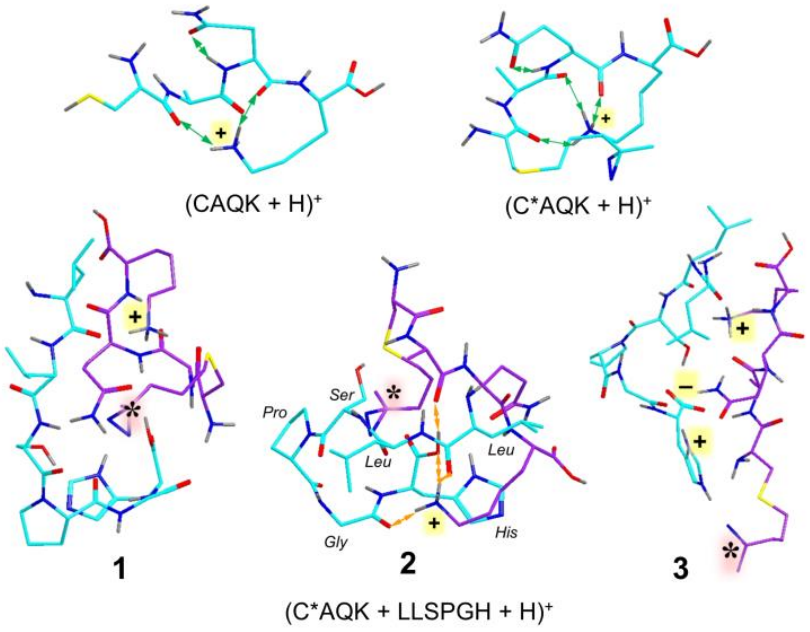


Figure 2.4. ωB97X-D/6-31+G(d,p) optimized structures of $(CAQK + H)^+$, $(C^*AQK + H)^+$, and $(C^*AQK + LLSPGH + H)^+$ complexes.

Atom color coding is as follows: Cyan or magenta = C, blue = N, red = O, yellow = S, gray = H. Only exchangeable (N-H, O-H) hydrogens are displayed. The diazirine rings are annotated with asterisks.

Multiple trajectories for **1–8** were analyzed for close contacts of the C*AQK diazirine carbon (denoted by an asterisk in Figure 2.5 and Figure S11) with X-H bonds in LLSPGH, and the data were consolidated for the individual residues (Figure 2.5). The contact analysis identified two groups of complexes. In **2, 3, 5**, and **7**, the majority of thermal contacts occurred at the His-6 residue. The other group showed substantial fractions of contacts at Leu-1 (**4, 6, 8**) and Leu-2 (**1**).

Distinction of these theoretical possibilities was made by comparing the contacts with cross-linking data for $(C^*AQK+LLSPGH+H)^+$, focusing on the distribution between the LLS and PGH segments (14:86 from experiment, Figure 2.3 a). These indicated the best match for structure **2** (14:86) followed by **3** (9:91), whereas 1 did not match and showed an opposite trend (73:27). The best-matching complex **2** showed a yin-yang-like structure of intertwined $(C^*AQK+H)^+$ and LLSPGH peptide units (Figure 2.4). The strongest non-covalent interactions in **2** involved the Lys $NH3^+$ hydrogen bonding to the His-6-COOH, Gly-5, and Leu-1 carbonyls of LLSPGH, and the His-6-COOH hydrogen bonding to the Ala carbonyl of C*AQK. The LLSPGH moiety in **2** showed a tightly folded proline b-turn, which was enforced by the above-listed intermolecular hydrogen bonds. We note that, at solution pH 7, the Lys side-chain group of pKa=10.53 is protonated to form an ion analogous to that produced in the gas-phase. Hence, the binding characteristics that we determined for the gas-phase ion may indicate that structure motifs analogous to the b-turn cleft of LLSPGH may play a role in the CAQK peptide binding to proteins.

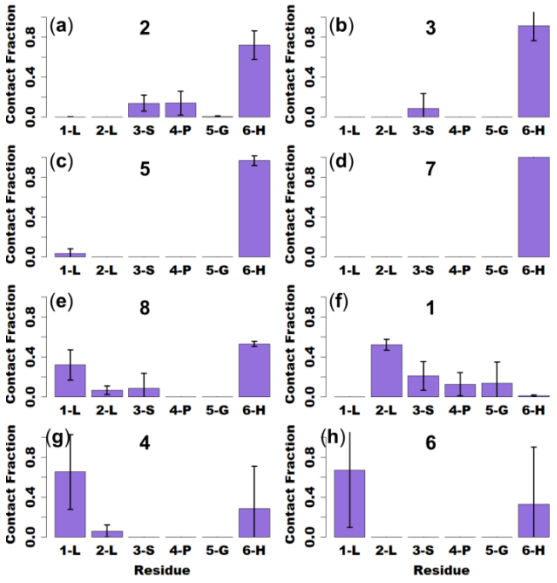


Figure 2.5. Analysis of close contacts within 4.5 Å between the C* diazirine carbon atom and X-H bonds of the amino acid residues of the LLSPGH target peptide. For structure numbering see Figure 2.4 and Figure S11.

2.4 CONCLUSION

In conclusion, carbene cross-linking to several target peptides of selectively diazirine-tagged CAQK has revealed different structures of non-covalent gas-phase complexes. The bonding in the complexes is dominated by inter-fragment hydrogen bonds that cause major changes in the secondary structure of the components. BOMD conformational analysis is presented as a useful tool for structural assignment, providing detailed information on intermolecular atom–atom contacts in thermal gas-phase complexes and facilitating the interpretation of experimental data.

2.5 EXPERIMENTAL SECTION

Synthetic procedures for diazirine tag introduction and product characterization are described in the Supporting Information section. Photodissociative cross-linking was performed on a modified LTQ-XL-ETD (ThermoElectron Fisher, San Jose, SA, USA) linear ion trap mass spectrometer equipped with a Nd-YAG laser/OPO laser system.[17] Further experimental and computational details are provided in the Supporting Information.

2.6 REFERENCES

1. Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H., *Journal of Molecular Biology* **1990,** *213* (4), 899-929.
2. 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., *Cell* **2016,** *165* (7), 1698-1707.
3. Mann, A. P.; Scodeller, P.; Hussain, S.; Joo, J.; Kwon, E.; Braun, G. B.; Molder, T.; She, Z.-G.; Kotamraju, V. R.; Ranscht, B.; Krajewski, S.; Teesalu, T.; Bhatia, S.; Sailor, M. J.; Ruoslahti, E., *Nature Communications* **2016,** *7*.
4. https://www.uniprot.org/blast (accessed April 23 2018).
5. Arlt, C.; Flegler, V.; Ihling, C. H.; Schaefer, M.; Thondorf, I.; Sinz, A., *Angewandte Chemie-International Edition* **2017,** *56* (1), 275-279.
6. Koelbel, K.; Warnke, S.; Seo, J.; von Helden, G.; Moretti, R.; Meiler, J.; Pagel, K.; Sinz, A., *Chemistryselect* **2016,** *1* (13), 3651-3656.
7. Shaffer, C. J.; Andrikopoulos, P. C.; Rezac, J.; Rulisek, L.; Turecek, F., *J Am Soc Mass Spectrom* **2016,** *27* (4), 633-45.
8. Das, J., *Chemical Reviews* **2011,** *111* (8), 4405-4417.
9. Pezacki, J. P.; Couture, P.; Dunn, J. A.; Warkentin, J.; Wood, P. D.; Lusztyk, J.; Ford, F.; Platz, M. S., *Journal of Organic Chemistry* **1999,** *64* (12), 4456--4464.
10. Pepin, R.; Shaffer, C. J.; Turecek, F., *Journal of Mass Spectrometry* **2017,** *52* (8), 557-560.
11. Roepstorff, P.; Fohlman, J., *Biomedical Mass Spectrometry* **1984,** *11* (11), 601-601.

12. Biemann, K., *Methods in Enzymology* **1990,** *193*, 886-887.
13. Vaisar, T.; Urban, J., *Journal of Mass Spectrometry* **1996,** *31* (10), 1185-1187.
14. Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A., *Angewandte Chemie-International Edition* **2006,** *45* (32), 5301-5303.
15. Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P., *Journal of Chemical Theory and Computation* **2009,** *5* (7), 1749-1760.
16. Chai, J.-D.; Head-Gordon, M., *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615-6620.
17. Shaffer, C. J.; Marek, A.; Nguyen, H. T. H.; Turecek, F., *Journal of the American Society for Mass Spectrometry* **2015,** *26* (8), 1367-1381.

Chapter 3. PHOTODISSOCIATIVE CROSS-LINKING OF DIAZIRINE-TAGGED PEPTIDES WITH DNA DINUCLEOTIDES IN THE GAS PHASE

Reproduced in part with permission from Liu, Y.; Turecek, F., *Journal of The American Society for Mass Spectrometry* **2019,** *30* (10), 1992-2006.

Abstract:

Non-covalent complexes of DNA dinucleotides dAA, dAT, dGG, dGC, and dCG with diazirine-tagged Cys-Ala-Gln-Lys peptides were generated as singly charged ions in the gas phase. Laser photodissociation at 355 nm of the diazirine ring in the gas-phase complexes created carbene intermediates that underwent covalent cross-linking to the dinucleotides. The dinucleotides differed in the cross-linking yields, ranging from 27 to 36% for dAA and dAT up to 90–98%for dGG, dGC, and dCG. Collision-induced dissociation tandem mass spectrometry (CID-MS^3) of the crosslinked conjugates revealed that fragmentation occurred chiefly in the dinucleotide moieties, resulting in a loss of a nucleobase and backbone cleavages. The CID-MS^3 spectra further revealed that cross-links were primarily formed in the 3′-nucleotides for the dAT, dGC, and dCG combinations. Gas-phase and solution structures of dGG complexes with S-tagged CAQK were investigated by Born-Oppenheimer molecular dynamics (BOMD) and density functional theory calculations. The low free-energy complexes had zwitterionic structures in which the peptide was protonated at the N-terminus and in the Lys residue whereas the carboxyl or dGG phosphate were deprotonated, corresponding to the respective $(Cys^+, Lys^+, COO^-)^+$ and $(Cys^+, Lys^+, phosphate^-)^+$ protomeric types. Both types preferred structures in which the peptide N-terminal cysteine carrying the S-photo-tag was aligned with the 3′-guanine moiety. BOMD trajectories at 310 K were analyzed for close contacts of the incipient peptide carbene with the positions in dGG that pointed

to frequent contacts with the N-1, NH_2, and N-7 atoms of 3′-guanine, in agreement with the cross-linking results. Carbene insertion to the guanine N-1-H and NH_2 bonds was calculated by density functional and Møller-Plesset perturbational theory to be 350–380 kJ mol^{-1} exothermic. Based on calculations, we proposed a mechanism for the carbene reaction with guanine starting with an exothermic attack at N-7 to form a dipolar intermediate that can close an aziridine ring in another exothermic reaction, forming a stable covalent cross link.

3.1 INTRODUCTION

Interactions of DNA with proteins are essential for life processes such as DNA replication and repair. The complexity of DNA-protein supermolecular assemblies, as revealed by X-ray crystallography[1], makes it challenging to study the components of DNA-protein interactions at high resolution and to investigate the interaction dynamics on a time scale relevant for biochemical processes. Mass spectrometry has been used to aid identification of DNA-protein cross-links formed in solution by specific chemical reactions[2-7], as well as by ionizing radiation[8], and UV light [9-12, 13-16], and the various approaches to deal with the complexity of the resulting mixtures have been reviewed[17-20]. The complexity problem can be simplified by bringing the DNA and protein components into the gas phase as non-covalent ion complexes and study their structure and dynamics in the absence of solvent, additives, reagents, and other intervening effects. Previous studies demonstrated the usefulness of such an approach, showing that oligonucleotide-peptide ion complexes can be generated in the gas phase [21-23], isolated by mass-to-charge ratios, and studied by tandem mass spectrometry [24-26]. However, the previous gas-phase studies were limited to using collision-induced dissociation (CID) to probe the interactions in the complexes, whereas structural details of nucleotide-peptide binding were not obtained.

We have shown recently that related peptide-peptide interactions in non-covalent ion complexes can be probed by combining site-specific photodissociative cross-linking in the gas phase with Born-Oppenheimer computational analysis of the molecular dynamics, revealing the interactions in the complexes and indicating the formation of covalent cross-links. By combining experiment and theory, we have been able to assign interactions in gas-phase complexes of the small peptide Cys-Ala-Gln-Lys (CAQK) with selected peptide motifs[27].

CAQK has recently been discovered to mitigate adverse effects of brain trauma in mice[28], possibly because of specific interactions with yet-to-be identified proteins. Because of the biological importance of CAQK, we now set out to investigate its interactions with several dinucleotides as surrogates of nuclear DNA. This small lysine-containing peptide can be viewed as a simplified surrogate of a histone interacting with DNA. Interaction with DNA and RNA of histone lysine side chains is the foundation of nucleic acids coiling in chromatin[29]. Modifications of histone lysines by alkylation and acylation are important tools for chromatin regulation [30]. However, the pattern of these modifications, the so-called histone code[31], has not been fully elucidated in spite of numerous studies[32]. In our approach, we use modified CAQK peptides which are specifically tagged with a 4,4-azipentyl group at the N-terminus (*CAQK)[33], cysteine thiol group (C*AQK)[27], and lysine side chain (CAQK*)[33], introducing a diazirine ring that can be selectively photodissociated at 355 nm (Figure 3.1). Diazirine photodissociation generates a transient carbene[34] that can undergo non-selective insertion into the proximate C–H, N–H, or O–H bonds of the partner molecule, forming a covalent bond and converting the complex into a molecular ion. The photodissociation products, both covalent and noncovalent, are further analyzed by CID tandem mass spectrometry (CID-MS^3) to probe the formation and location of the covalent bonds connecting the peptide and the DNA. Since CID-MS^3 often results in limited

sequence coverage[35], the experimental data do not allow us to achieve atomic-level resolution in determining the cross-links and identifying all interactions in the complexes. To improve resolution, we complement experimental cross-linking data with Born-Oppenheimer molecular dynamics (BOMD) calculations that involve all valence-electron interactions along with nuclear motion at the experimental temperature. BOMD provides time-resolved trajectories that reveal contacts between the peptide and DNA atoms that may result in cross-linking in photochemically produced carbene intermediates. We now apply this approach to the study of CAQK complexes with DNA dinucleotides dAA, dAT, dGG, dGC, and dCG. We wish to show that in interactions with CAQK ions, even simple dinucleotides differ in their binding efficiency and stereochemistry, and we provide structures of selected complexes obtained by electronic structure theory calculations using density functional theory (DFT).

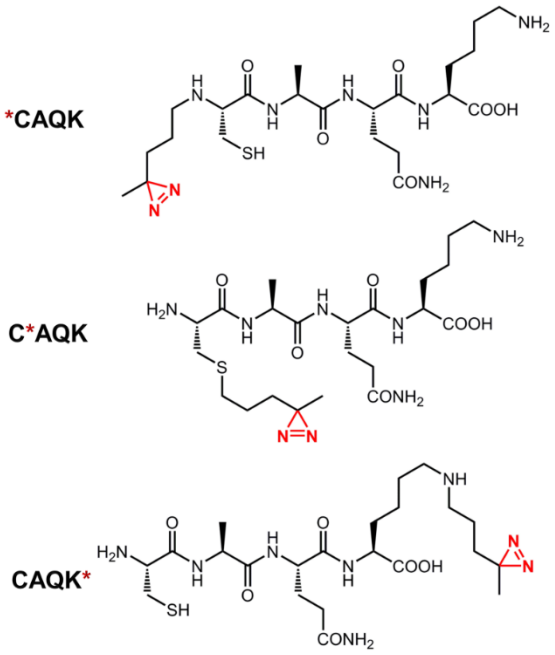


Figure 3.1. Diazirine labeled Cys-Ala-Gln-Lys peptides.

3.2 EXPERIMENTAL SECTION

3.2.1 *Materials and Methods*

Peptides C*AQK, *CAQK, and CAQK* were synthesized according to previously reported procedures[27]. Mass spectra were measured on an LTQ-XL-ETD (Thermo, San Jose, USA), and Bruker amaZon Speed (Bruker, Bremen, Germany) ion trap mass spectrometers equipped with an Nd-YAG laser EKSPLA laser[36]. Ions were produced by electrospray ionization of ca. 50 μM solutions in methanol-water-acetic acid, selected by *m/z* and stored in the ion trap.

3.2.2 *Calculations*

Born-Oppenheimer molecular dynamics (BOMD) trajectories were run with semiempirical all-valence-electron quantum chemistry calculations using the Berendsen thermostat algorithm[37]. Temperature was set at 310 K to represent the experimental conditions in the ion trap. For each complex type, several (> 5) initial structures were constructed from PM6-optimized peptide subunits in which the diazirine tag and dGG were in different spatial orientations. These complexes were subjected to a preliminary BOMD using PM6[38] that was augmented by including dispersion interactions (D3H4[39]). These calculations were run by MOPAC[40] that was coupled to the Cuby4 framework[41,42]. Running trajectories with 1 fs steps for 20 ps furnished 20,000 snapshots per each initial structure from which 200 snapshots were extracted at 100 fs intervals. The extracted snapshot structures were fully gradient-optimized with PM6-D3H4 and sorted out by their secondary and super-secondary structural similarities to compact duplicates and reduce the size of the selection. This yielded 60–70 distinct structures whose geometries were fully optimized by DFT gradient calculations. B3LYP/6-31G(d,p)[43] optimized structures were used for frequency calculations to provide vibrational enthalpies and entropies whereas ωB97X-D/6–31 + G(d,p)[44]

optimized structures were used to evaluate electronic energies including dispersion interactions. The electronic, vibrational, and rotational terms were combined to yield relative free energies for the gas-phase complexes. In addition, solvent effects were included by single-point energy self-consistent reaction field calculations using the polarizable continuum model[45] with water as the dielectric and the gas-phase optimized structures. All the thermochemical calculations were performed using the Gaussian 16 (Revision A.03)[46] suite of programs. The final optimized structures with relative free energies within 40 kJ mol^{-1} of the global energy minimum were then subjected to a full BOMD with PM6-D3H4 for 100 ps at 310 K. Cuby4 was also used to extract close contacts between the carbon atom of the diazirine ring and the dinucleotide atoms. Close contacts were limited to X–H…C distance of less than 4.5 Å on the basis of the van der Waals radii of the diazirine and X–H atoms[36].

3.3 Results and Discussion

3.3.1 *Complex Formation and Photodissociation*

Electrospray ionization of equimolar mixtures of the labeled CAQK peptides (abbreviated as M*) and dinucleotides (dXX, abbreviated as m) produced the pertinent singly charged $(M^*+dXX + H)^+$ complexes in ca. 1% yield relative to the dominant singly charged $(M^*+ H)^+$ and $(m+H)^+$ ions. This is illustrated by the spectra of (C*AQK + dXX) complexes in Fig. S1a-e (Supporting Information). The complexes were isolated by mass, *m/z* 1109, 1100, 1141, and 1101 for dAA, dAT, dGG, and dGC, dCG, respectively, stored in the ion trap and subjected to several laser pulses at 355 nm. This resulted in loss of N_2 and partial dissociation as shown by the UVPD-MS^2 mass spectra (Figure 3.2 a–e). With dAA and dAT complexes, UVPD resulted in substantial dissociation following the loss of N_2, producing $(M^*-N_2+H)^+$ peptide ions at *m/z* 517 whereas the complementary dinucleotide $(m + H)^+$ ions gave only minor peaks (*m/z* 565 and 556, Figure 3.2 a,

b). In contrast, UVPD of the $(M^* + m + H)^+$ ions from dGG, dGC, and dCG, following the loss of N_2, produced $(M^* - N_2 + m + H)^+$ complexes with only minor dissociation to peptide and dXX monomers (Figure 3.2 c–e). The UVPD-MS^2 results for C*AQK, *CAQK, and CAQK* complexes are summarized in Table 3.1. Comparing the UVPD-MS^2 spectra across the three photopeptides, the photolyzed C*AQK and *CAQK complexes of dGG, dGC, and dCG showed comparably high resistance to dissociation. This was expressed as a percent fraction R(MS^2) of surviving $(M^* - N_2 + m + H)^+$ ions (Table 1)[35 36]. The complexes of CAQK* were somewhat less stable, giving 82% survivor ions for dGG. The complexes of dAA displayed more variable stabilities, giving 36%, 27%, and 19% survivor complexes for C*AQK, *CAQK, and CAQK*, respectively.

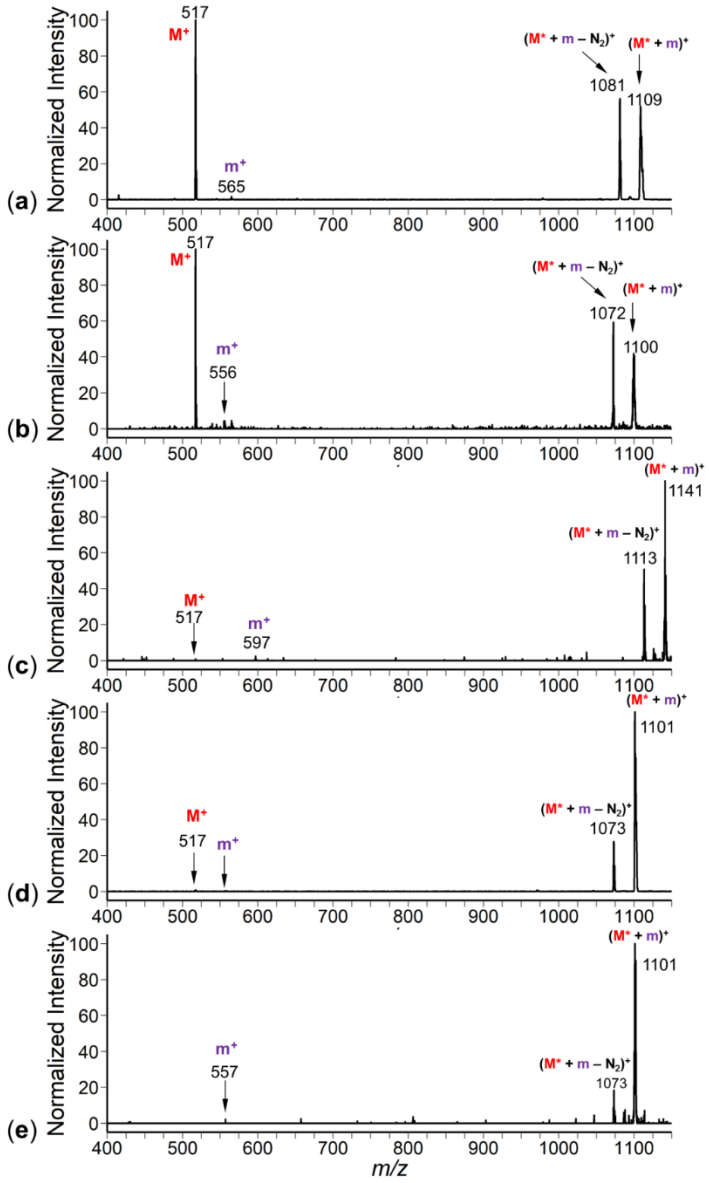


Figure 3.2. UVPD-MS^2 spectra of C*AQK complexes with (a) dAA, (b) dAT, (c) dGG), (d) dGC, and (e) dCG.

The dissociations of the photolyzed complexes are in part driven by the highly exothermic isomerization of the carbene intermediate to an olefin ($\Delta H_{isom} \approx 200$ kJ mol^{-1}) [47] that occurs in the absence of fast X–H insertion[48-51]. The UVPD-MS^2 dissociations were compared with CID-MS^2 of (M* +m+ H)$^+$ ions that were not exposed to diazirine photolysis. The spectra in Fig. S2a-e (Supporting Information) indicated that among the C*AQK complexes, that with dAA was most

sensitive to CID, primarily forming the $(M^* + H)^+$ ion at *m/z* 545. CID of the dAT, dGG, and dGC complexes produced protonated dinucleotide and peptide at comparable relative intensities. The complex of dCG was most resistant to CID, forming small but nearly equal fractions of protonated dCG and C*AQK. We note that CID did not result in elimination of N_2, the only significant dissociation besides cleavage to the monomers being a minor loss of water (Fig. S2a-e).

3.3.2 *CID-MS^3 of Photolyzed Conjugates of C*AQK*

The $(M^*\text{-}N_2+m+H)^+$ ions formed by photodissociation were selected by mass and subjected to CID. In the absence of covalent cross-linking, the MS^3 fragmentation was expected to result in the formation of monomeric units, competing for the single available proton according to their relative basicities. In contrast, the covalently bound ions, which we call conjugates, can presumably undergo a combination of peptide and nucleotide backbone fragmentations, as well as water, ammonia, side chain, and nucleobase losses. To describe the fragment ions, we used the standard DNA ion nomenclature [52 53], which, reading from the 5′ to the 3′ terminus, assigns the fragments as a, b, c, d, w, x, y, z, and those by loss of the neutral 5′and 3′ base as $(-B_{1H}(X_H))$ and $(-B_{2H}(X_H))$, respectively. The protonated 5′ and 3′ bases are denoted as $(B_1H_2^+)$ and $(B_2H_2^+)$, respectively (Scheme 3.1). Backbone cleavage in the peptide was described by the standard Biemann-Roepstorff-Pohlman nomenclature for the C-terminal (x_n, y_n) and N-terminal (a_m, b_m) fragments [54 55]. For combined dissociations in both the dXX and peptide units, we used a combined notation, e.g., $w_1\_B_1$ refers to an ion containing the w_1 nucleobase fragment linked to the B_1 peptide fragment. A table of all theoretical fragment ion combinations was assembled and compared against the fragment ions observed in the CID-MS^3 spectra. The identified fragment ions from all CID-MS^3 spectra of photo-cross-linked complexes are shown in Table S1.

CID of the (dAA + C*AQK − N_2+H)$^+$ ion (*m/z* 1081) resulted in the formation of the peptide monomer (M* − N_2+H)$^+$ (*m/z* 517) and its secondary fragment (M* − N_2 - S-tag + H)$^+$ (*m/z* 415, Figure 3.3a). In addition, the CID-MS^3 spectrum showed numerous fragment ions indicating covalently linked species. We succeeded in assigning most CID-MS^3 fragment ions from (dAA + C*AQK - N_2+H)$^+$ to backbone cleavages in dAA combined with loss of a nucleobase and water. Consecutive loss of water is indicated by green arrows in Figure 3.3 a–e. A few fragment ions, such as *m/z* 659, did not correspond to logical backbone cleavages and were unassigned. The ion series starting at *m/z* 946 was formed by loss of adenine (B_{1H} or B_{2H}) and followed by successive eliminations of water (*m/z* 928 and *m/z* 910). Another prominent fragment ion series started at *m/z* 848 (w_1/d_1_M), indicating backbone cleavage in dAA and loss of a mononucleoside unit. In this case, as with dGG, the w_1 and d_1 fragments were isomers and thus indistinguishable by mass. Two low-intensity fragment ion series were particularly diagnostic for locating covalent cross-links. The *m/z* 811 ion was formed by losses of both adenine nucleobases, indicating peptide cross-linking to the deoxyribose-phosphate framework. Consecutive eliminations of water from the *m/z* 811 ion produced secondary fragment ions at *m/z* 793, 775, and 757. Likewise, the *m/z* 713 fragment ion and its secondary fragments by loss of water (*m/z* 695, 677, and 659) indicated loss of both nucleobases in precursors cross-linked to the deoxyribose-phosphate framework. The formation of the *m/z* 713 ion can be viewed as converging from multiple pathways, for example, loss of H_3PO_4 from *m/z* 811, loss of adenine from *m/z* 848, or loss of adenine-dideoxyribose (233 Da) from *m/z* 946. These are standard dissociations of nucleotide ions [52 53 56 57]. In contrast to the above fragment ions, the *m/z* 652 and its loss-of-water secondary fragments were assigned to adenine-peptide cross-links. This indicated that the population of (dAA + C*AQK - N_2+H)$^+$ ions was heterogeneous, containing isomers cross-linked to the nucleobases

as well as to the deoxyribose-phosphate framework. The cross-linking efficiency for (dAA + C*AQK + H)$^+$ was R(MS^3) = 58% (Table 3.1), as expressed by a ratio of the intensities of the identified crosslinked fragment ions to the sum of all fragment ion intensities in the CID-MS^3 spectrum[35, 36]. Combined with the 36% fraction of (dAA + C*AQK - N^2+H)$^+$ ions surviving UVPD-MS^2, the overall efficiency R(total) was 21%.

Table 3.1. Crosslinking percent efficiencies

Photopeptide	Dinucleotide	R(MS^2)[a]	R(MS^3)[b]	R(total)[c]
C*AQK	dAA	36	58	21
	dAT	30	67	20
	dGG	98	47	46
	dGC	96	39	37
	dCG	90	38	34
*CAQK	dAA	27	58	9
	dAT	26	67	13
	dTA	36	40	14
	dGG	98	37	36
	dGC	93	27	25
	dCG	97	18	17
CAQK*	dAA	19	60	11
	dGG	82	40	33

[a]R(MS^2) = 100 × [M*+m-N$_2$]/[[M$^+$]+[m$^+$]+[M*+m-N$_2$]]

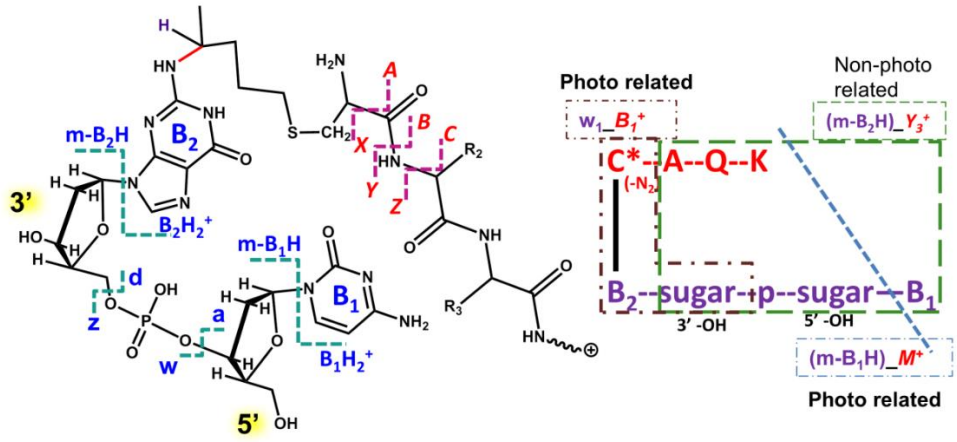
[b]R(MS^3) = 100 × Σ[Cross-linked fragments]/[Σ [Cross-linked fragments][M$^+$]+[m$^+$]]

[c]R(total) = [a]R(MS^2) × [b]R(MS^3)

The CID-MS^3 spectrum of the (dAT + C*AQK − N$_2$ + H)$^+$ ion (Figure 3.3b, *m/z* 1072) was analyzed in a similar fashion, as briefly discussed below. A comparison of the monomer (M* − N$_2$+H)$^+$ ion relative intensity with those of the cross-linked fragment ions in the *m/z* 530–1000 region indicated a significant fraction, R(MS^3) = 67%, of covalently cross-linked conjugates of dAT. The structurally most significant fragment ions were at *m/z* 937 (loss of adenine) and *m/z* 839 (w$_1$_M) that both indicated peptide cross-linking to the 3′-dT moiety.

A significant fraction of cross-linked products, R(MS^3) = 47% (Table 3.1) was also indicated by the CID-MS^3 spectrum of the (dGG + C*AQK - N$_2$+H)$^+$ ion (Figure 3.3c, *m/z* 1113). Major dissociations were a loss of guanine (*m/z* 962) and a backbone cleavage giving the w$_1$/d$_1$_M

ion at *m/z* 864. The less intense ions at *m/z* 811 and *m/z* 713 and their secondary fragments by loss of water pointed to cross-links within the deoxyribose-phosphate framework, as discussed above for dAA. Specific cross-linking to the nucleobases was indicated by minor fragment ions at *m/z* 668, pertinent to G_M.



Scheme 3.1. Nomenclature of fragment ions from dinucleotide-peptide cross-links.

The CID-MS^3 spectra of the isomeric (dGC +C*AQK - N_2 + H)$^+$ and (dCG + C*AQK − N_2 + H)$^+$ ions showed striking differences (Figure 3.3d, e) despite similar cross-linking efficiencies, R(MS^3) = 39% and 38%, respectively (Table 3.1). Starting with dGC (*m/z* 1073, Figure 3.3d), the CID-MS^3 spectrum displayed fragment ions by loss of both cytosine (*m/z* 962) and guanine (*m/z* 922) and their loss-of-water satellites. The loss-of-G peak was substantially more intense than that for the loss of C. Another significant fragment ion series (*m/z* 824, w_1_M, and its loss-of-water satellites) indicated peptide cross-linking within the 3′-dC moiety. In addition, the *m/z* 811 and *m/z* 713 fragment ions revealed the presence of isomers in which the peptide was cross-linked to the deoxyribose-phosphate framework. Peptide cross-linking to the nucleobases was indicated by minor fragment ions at *m/z* 668 and *m/z* 628 that were assigned to (G_M) and (C_M), respectively.

Contrasting the spectrum of the GC conjugate, the CID-MS^3 spectrum of (dCG + C*AQK $- N_2+H)^+$ showed fragment ions by loss of cytosine (*m/z* 944) whereas loss of guanine (*m/z* 962) was undetectable (Figure 3.3e). The other important fragment ions were represented by *m/z* 864 that was assigned as w_1_M and indicated peptide cross-linking to the 3′-dG moiety. The *m/z* 713 and 677 ions pointed to loss of guanine and water from the *m/z* 864 ion, indicating that the peptide was cross-linked to the 3′-phosphodeoxyribose moiety. However, the weaker fragment ion at *m/z* 668 also indicated a minor fraction of isomers where the peptide was cross-linked to the guanine residue.

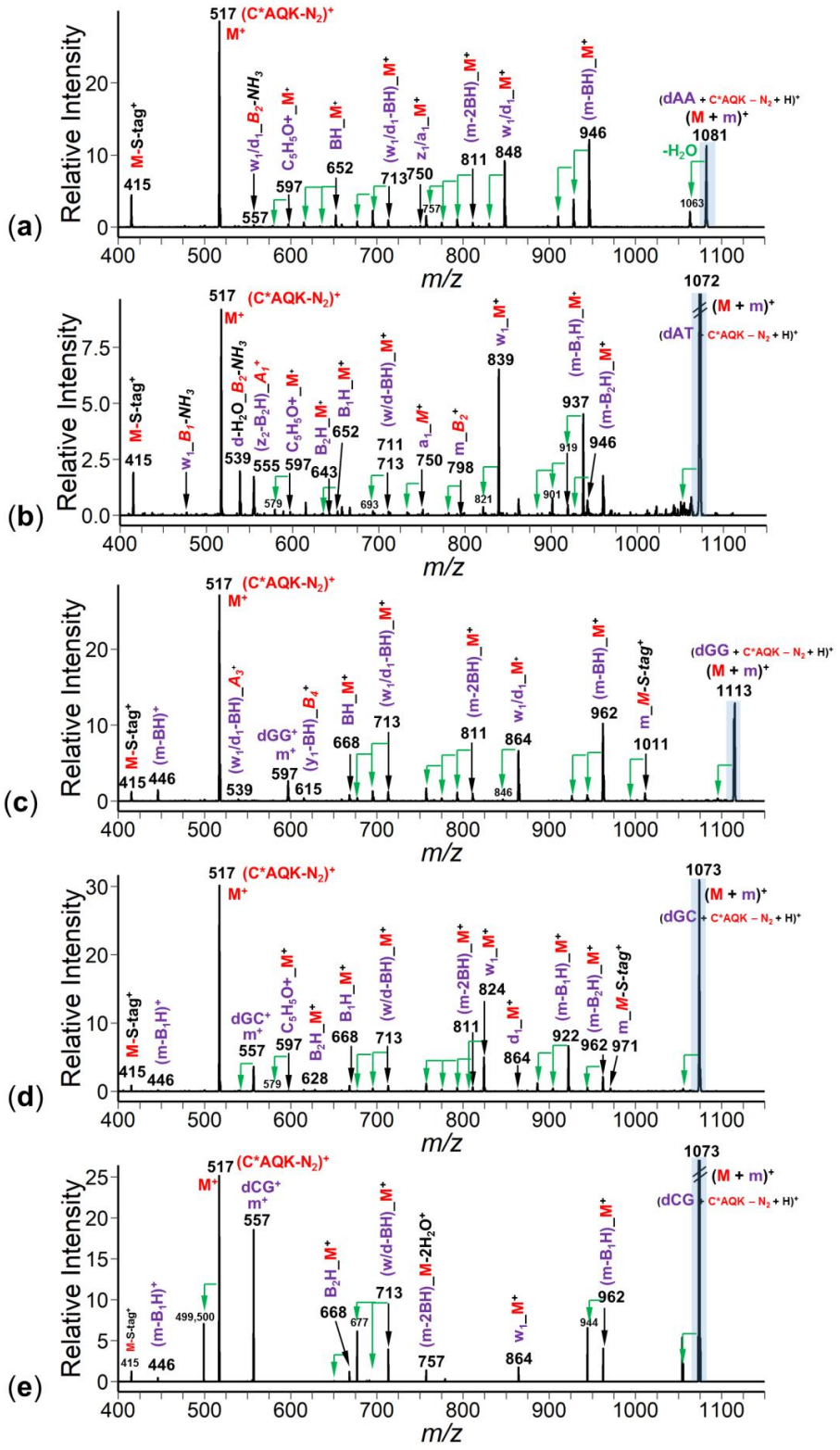


Figure 3.3. CID-MS^3 spectra of (C*AQK – N_2) complexes with (a) dAA, (b) dAT, (c) dGG, (d) dGC, and (e) dCG. Green arrows indicate consecutive dissociations by loss of water.

3.3.3 *Conjugates of C*AQK and CAQK**

Photodissociation of dinucleotide complexes with *CAQK gave results that were analogous to those for C*AQK. Upon photodissociative loss of N_2, the complexes of AA, AT, and TA underwent substantial dissociation to monomers, chiefly represented by the peptide ion M^+ (*m/z* 517, Fig. S3a-c). The survivor $(dXX + C^*AQK - N_2+H)^+$ conjugates within this group amounted to $R(MS^2)$ = 26–36% of the ion intensities (Table 3.1). In contrast, conjugates of GG, GC, and CG dissociated much less (Fig. S3d-f), generating 93–98% of survivor ions (Table 3.1). Photodissociation of complexes with CAQK* was studied for dAA and dGG only and produced results that were analogous to those for the other tagged peptides. Table 3.1 shows 19% and 82% survivor ion yields in the CID-MS^3 spectra of conjugates with dAA and dGG, respectively. The UVPD spectra of the *CAQK complexes were compared with CID spectra that resulted in dissociation to monomers (Fig.S4a-f).

The CID-MS^3 spectrum of $(dAA + ^*CAQK - N_2+H)^+$ (*m/z* 1081, Fig. S5a) showed major cross-linked fragment ions by loss of adenine (*m/z* 946), backbone cleavage (w_1/d_1_M, *m/z* 848), and B_H_M (*m/z* 652). The quality of this and Fig. S5b spectrum was affected by the presence of near-isobaric contaminants of ions from *CAQK disulfide that we found forming readily in solutions of peptides with free cysteine. These ions contained two diazirine groups and underwent loss of the second N_2 molecule upon CID-MS^3, giving rise to fragment ions at *m/z* 1053 and 1044 in the respective Figs. S5a and S5b. The CID-MS^3 spectrum of $(dTA + ^*CAQK - N_2+H)^+$ (*m/z* 1072, Fig. S5c) showed fragment ions by loss of adenine (*m/z* 937), but also an adenine-peptide adduct (*m/z* 652), indicating non-specific cross-linking in the 5′ and 3′ parts of the dinucleotide. The CID-MS^3 spectrum of $(dGG + ^*CAQK - N_2+H)^+$ (*m/z* 1113, Fig. S5d) was dominated by fragment ions formed by loss of guanine (*m/z* 962), the w_1/d_1_M backbone ion (*m/z* 864), and their

dehydrated satellites. Also in this series, the CID-MS^3 spectra of the dGC and dCG complexes displayed marked differences in the fragment ion intensities (Fig. S5e, f). The dinucleotides had different propensities for competing for the proton against CAQK*–N_2, with $(dCG + H)^+$ showing higher ion relative intensity than $(dGC + H)^+$, implying a higher gas-phase basicity for dCG. Regarding the cross-linked fragments, loss of the 5′ nucleobase was prevalent, as evidenced by the relative intensities of the *m/z* 922 and *m/z* 962 ions from the dGC and dCG conjugates, respectively. Dinucleotide backbone cleavage formed the cross-linked w_1_M ions, retaining the 3′-portion of the dinucleotide for both dGC and dCG. These results indicated that the peptide N-terminal tag in the complexes was in the vicinity of the 3′-residue regardless of whether it was guanine or cytosine.

The CID-MS^3 spectrum of the $(dAA + CAQK^* - N_2+H)^+$ complex was similar to that of the C*AQK complex, showing again fragment ions by loss of adenine (*m/z* 946), formation of w_1/d_1_M (*m/z* 848), *m/z* 811, and *m/z* 713 ions along with their dehydrated satellites (Fig. S6a). These results indicated that in the dAA complexes, the position of the peptide diazirine tag did not play a critical role to affect cross-linking.

3.3.4 *Reference CID of Dinucleotide Ions*

The CID-MS^3 of peptide-dinucleotide conjugates revealed that most dissociations occurred in the dinucleotide moieties whereas peptide backbone dissociations were scarce or absent. To assess the effect of peptide covalent binding on the CID spectra of the conjugates, we obtained reference CID-MS^2 of protonated dinucleotides under the same conditions of slow heating in the ion trap, typically achieving > 80% precursor ion depletion following excitation. The reference spectra (Supplementary Figs. S7a-f) are now briefly described, and the observed fragment ions are compared with their analogues from the peptide conjugates.

CID-MS^2 of $(dAA + H)^+$ (Fig. S7a) showed a $(m\text{-}BH)^+ > w_1/d_1{}^+$ pattern of fragment ion intensities that was analogous to that for CID-MS^3 of the C*AQK, *CAQK, and CAQK* conjugates. The $[(m\text{-}BH)^+]/[w_1/d_1{}^+]$ intensity ratio for $(dAA + H)^+$(1.4) was slightly higher than that shown for $(dAA + C^*AQK - N_2 + H)^+$ (1.2) in Figure 3.3a. CID-MS^2 of $(dGG + H)^+$ (Fig. S7b) showed a dominant loss of guanine but relatively much less backbone cleavage to the $w_1/d_1{}^+$ ions than was observed in CID-MS^3 of the C*AQK and *CAQK conjugates. The differences for dAA and dGG may be caused by peptide crosslinking to the nucleobases, as indicated by the CID-MS^3 spectra, which can be expected to hamper the loss of a free nucleobase from the conjugate. CID-MS^2 of $(dAT + H)^+$ (Fig.S7c) chiefly produced the $w_1{}^+$ fragment ion, which was also prominent in the CID-MS^3 of the dAT conjugate (Figure 3.3b). The CID-MS^2 spectrum of $(dGC + H)^+$ (Fig. S7d) displayed fragment ions by loss of G and C in a 1.1:1 ratio. This was different from the CID-MS^3 spectrum of the dGC conjugate (Figure 3.3d) where the loss of G was fourfold more prominent. In addition, the conjugate showed a higher propensity for backbone cleavage forming the $w_1\_M^+$ fragment ion compared to the formation of the analogous $w_1{}^+$ ion from $(dGC + H)^+$. Regarding dCG, both $(CG + H)^+$ (Fig. S7e) and the conjugate ion with C*AQK (Figure 3.3e) preferred loss of cytosine. The main difference was in the formation of the $w_1{}^+$ fragment ion which was more prominent in the CID-MS^3 spectrum of the conjugate (*m/z* 713, Figure 3.3e). Summarizing these comparisons, covalent attachment of CAQK peptides did not change the nature of CID of the dinucleotide conjugates and indicated the prevalence of 5′-base losses and formation of w ions. The differences in the relative intensities of analogous nucleotide and conjugate fragment ions can be related to the presence of the peptide moiety, in particular in blocking the loss of free nucleobases.

3.3.5 *Complex Ion Structures*

The experimental data indicated a high overall efficiency of gas-phase cross-linking in CAQK-dXX complexes that reached 46% for $(dGG + C^*AQK + H)^+$ (Table 3.1). At the same time, the CID-MS^3 data provided only limited information on the location of covalent cross-links, which, for dAA and dGG, was hampered by their near-symmetrical nature, leading to indistinguishable isomeric fragments. To improve resolution and relate the observed cross-links to the gas-phase ion structures, we carried out extensive computational analysis of the $(dGG + C^*AQK + H)^+$ system using Born-Oppenheimer molecular dynamics (BOMD) at the semi-empirical level of quantum theory that included all valence-electron interactions. Thermodynamic calculations were also carried out that relied on density functional theory including dispersion interactions and solvation effects in water.

Initial complex ion structures were built for two different $(dGG + C^*AQK + H)^+$ protomers that were protonated at the Cys N-terminus and Lys side chain, in combination with peptide carboxylate or dGG phosphate anions to produce the overall charge of + 1. Several initial structures of each protomer type were built at different relative orientations of the peptide and dGG counterparts. In total, 11 initial structures were considered as starting points for the BOMD trajectory studies (Fig.S8). BOMD was run at 410 K for 20 ps to produce a total of 11 × 20,000 = 220,000 snapshots out of which 2200 structures were extracted at 100 fs intervals, optimized by PM6-D3H4, and sorted by their secondary and supersecondary structural similarities to compact duplicates and reduce the size of the selection. In the course of BOMD runs, the conformation sampling developed the initial complexes into five groups of canonical and zwitterionic structures out of which 60–70 distinct structures were selected for full geometry optimization by DFT gradient calculations. B3LYP/6-31G(d,p) optimized structures were used for frequency

calculations to provide vibrational enthalpies and entropies whereas ωB97X-D/6–31 + G(d,p) optimized structures provided electronic energies. The electronic, vibrational, and rotational terms were then combined to yield relative free energies for the complexes. In addition, solvent effects were included in single-point energy calculations via the polarizable continuum model[45] using water as the dielectric. The calculated relative free energies are compiled in Table 3.2, the optimized geometries of complexes **a-d, m**, and **j** are given in supplementary Tables S3-S8.

Table 3.2. Relative free energies of $(dGG+C^*AQK+H)^+$ complexes

		Relative free energy[a]
Complex	$\Delta G^\circ_{310,\ g}$	$\Delta G^\circ_{310,\ aq}$
a	7.7[b]	1.5[c]
b	6.3	0.6
c	0.0	0.0
d	14	−4.8
e	11	26
f	5.3	13
g	4.1	19
h	16	17
i	14	32
j	−5.8	5.0
k	1.1	36
l	−2.2	33
m	21	3.5
n	21	31
o	18	30
p	−3.8	36

[a]In kJ mol^{-1}
[b]From ωB97X-D/6–31 + G(d,p) calculations including B3LYP/6-31G(d,p) zero-point vibrational energies, 310 K enthalpies and entropies
[c]From ωB97X-D/6–31 + G(d,p) calculations including solvation energies in the water dielectric

The distribution of free energies for the gas-phase complexes $\Delta G_{g,310K}$) is shown in a box-plot format (Fig. S9) for the five protomer types. The Cys-protonated canonical structures encompassed 10 conformers within a 22–237 kJ mol^{-1} energy range and a median of 106 kJ mol^{-1}. The Lys protonated canonical structures had 8 conformers ranging from 21 to 150 kJ mol^{-1}, with a 63 kJ mol^{-1} median. The three zwitterionic protomer types showed distinct free-energy distributions. The $(Cys^+, Lys^+, COO^-)^+$ combination had 20 structures within a 5.8–170 kJ mol^{-1} range and a 63 kJ mol^{-1} median. The $(Cys^+, Lys^+, phosphate^-)^+$ protomers included ~ 16 structures within a narrow range of 0–69 kJ mol^{-1} and a median of ~ 19 kJ mol^{-1}. The $(3'\text{-}G^+, Cys^+,$

phosphate^{-})$^{+}$ combination had 7 structures within a 35–146 kJ mol^{-1} range with a 107 kJ mol^{-1} median. The energies of the gas-phase complexes indicated overall preference for zwitterionic protomers of the (Cys^{+}, Lys^{+}, COO^{-})$^{+}$ and (Cys^{+}, Lys^{+}, phosphate^{-})$^{+}$ types.

Since the complexes were formed in solution before or during electrospray, we also strived to assess their free energies in aqueous solution. To this end, we selected structures within 66 kJ mol^{-1} of the lowest-energy gas-phase structure for single-point energy calculations that included solvent effects. These showed that most low-energy solvated structures belonged to the zwitterionic protomer types. Solvation changed the ranking of some complexes as illustrated by the free-energy data for 16 low-energy structures **a-p** in Table 3.2. In particular, complex **d**, which is a (Cys^{+}, Lys^{+}, phosphate^{-})$^{+}$ zwitterion, was stabilized by 19 kJ mol^{-1} in water relative to the other complexes, becoming the global free-energy minimum. On the contrary, complexes **l** and **p** which were low free-energy zwitterions of the (Cys^{+}, Lys^{+}, COO^{-})$^{+}$ type in the gas phase were relatively destabilized by 35–40 kJ mol^{-1} in water (Table 3.2).

The binding between the peptide and dGG in complexes **a-p** was realized in a few general types (Figure 3.4). Structures **b, c, d**, and **m** had strong hydrogen bonds between the charged phosphate and both peptide ammonium groups. In addition, the Lys NH_3 $^{+}$ developed a stabilizing hydrogen bond to the 5′-hydroxyl in these low-energy complexes (**b-d, m**, Figure 3.4) that were particularly favored by solvation. The hydrogen bonding to 5′-OH enforced the orientation of the peptide with respect to dGG in which the N-terminal tag was in the vicinity of the 3′-guanine. Another group of (Cys^{+}, Lys^{+} , phosphate^{-})$^{+}$ zwitterionic complexes, **e, f, g, h**, and **n**, were characterized by hydrogen bonds between the N-terminal NH_3^{+} and the 3′-guanine whereas the Lys NH_3 $^{+}$ developed hydrogen bonds to the 5′-guanine and OH group (Fig. S10). The Table 3.2 data indicate that this H-bonding arrangement was less energetically favorable, leading to higher

free energies in the gas phase and upon solvation. The low-energy structure a (Figure 3.4) represented a hybrid in which the N-terminal NH_3^+ forms a hydrogen bond to the phosphate whereas the Lys NH_3^+ bound to the 3′-guanine.

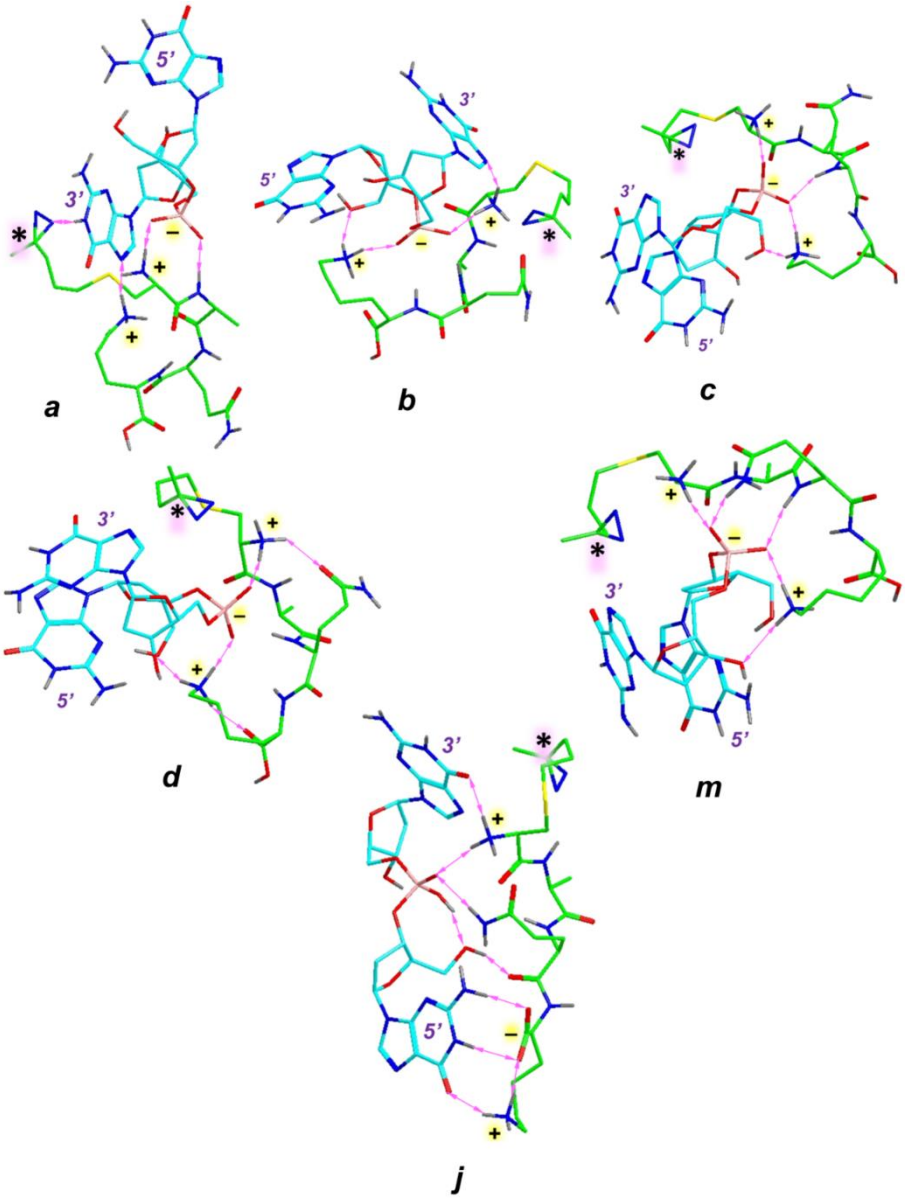


Figure 3.4. ωB97X-D/6-31+G(d,p) optimized structures of zwitterionic (C*AQK + dGG + H)$^+$ complexes **a**, **b**, **c**, **d**, and **m** and **j**.

Atom color coding is as follows: C in peptide = light green, C in dGG = cyan, O = red, N = blue, P = pink, S = yellow, gray = H. Only exchangeable hydrogens are shown. Double-headed

arrows indicate hydrogen bonds between the peptide and dGG. The diazirine ring is labeled with an asterisk.

The $(Cys^+, Lys^+, COO^-)^+$ zwitterionic complexes, **i, j, k, l, o,** and **p** offered a variety of hydrogen-bonding patterns (Figure 3.5). In some (**i, j, k**, and **o**), the COO^- anion participated in H-bonding to dGG, and in the others, it did not (Figure 3.5). There was no obvious effect of COO^- hydrogen bonding on the complex free energy, as structures of both types belonged to low-energy gas-phase complexes (**j, k, l**, and **p**, Table 3.2). However, only complex **j** was favored by solvation to compete with **a-d** and **m**. It appears from the energy data that complexes **a-d, j**, and **m** were thermodynamically most favored to be formed in aqueous solution and transferred to the gas phase. Conformational changes caused by thermal motion in gas-phase ions can favor low-energy gas-phase structures (**p**) and disfavor others (**m**). These effects are discussed later in the paper in connection with BOMD trajectory analysis.

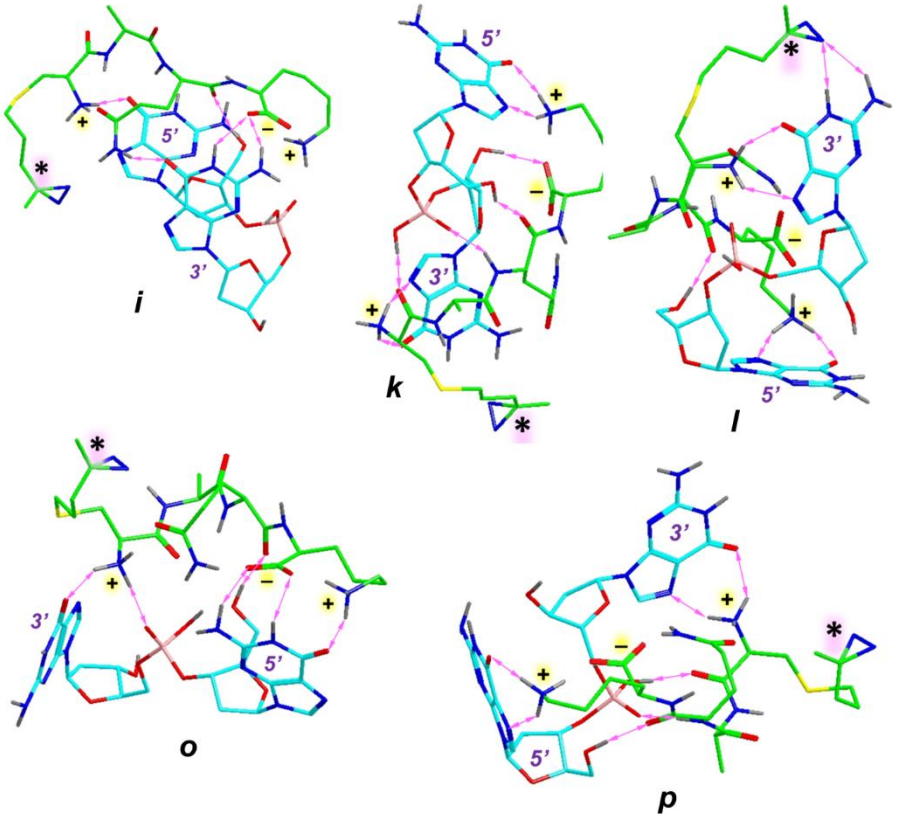


Figure 3.5. ωB97X-D/6-31+G(d,p) optimized structures of zwitterionic (C*AQK + dGG + H)$^+$ complexes **i**, **k**, **l**, **o**, and **p** of the (Cys^+, Lys^+, COO^-)$^+$ type.

To summarize the structure analysis, the predominant interactions determining most of the low-energy dGG complex structures were due to ion-ion hydrogen bonding between the charged groups. Another important feature was hydrogen bonding of the Lys NH_3^+ to the dGG 5′-hydroxyl that assisted the alignment of the dGG and peptide ion moieties where the N-terminus carrying the diazirine tag was close to the 3′-nucleobase. Peptide hydrogen bonding to the nucleobases, which could be considered important for DNA sequence recognition, was important in several structures **e, f, g, h, i, j, k, l, n, o**, and **p**. However, according to the calculated free energies, these complexes, except **j**, were disfavored by solvation with water. We presumed that these results can be generalized to the other dinucleotides and used to interpret their crosslinking.

3.3.6 *Born-Oppenheimer Molecular Dynamics and Contact Analysis*

Complexes **a-p** were analyzed for close contacts between the diazirine carbon atom (C-136) and the guanine, 2'-deoxyribose, and phosphate oxygen atoms (Figure 3.6). dGG atoms carrying hydrogen (X – H) were considered as potential targets for carbene insertion, while the basic N-7 and N-3 positions in the guanines could potentially react with the transient carbene. First, we inspected close contacts within 4.5 A in optimized structures **a-p** representing 0 K geometries. Structures **b, f, n, o**, and **p** displayed no 0 K contacts.

The majority of close contacts in **a, c, d, e, h, j, k, l**, and **m** were realized in the 3′-guanine positions N-1 (N61) and NH_2 (N62) for potential insertion to the N – H bonds, and at C-5 (C56), N-7 (N55), and C-8 (C52) for carbene addition. The 5′-nucleoside showed only two 0 K contacts, one at the NH_2 (N21 in g) and the other with the deoxyribose C-2 (C28 in **c**).

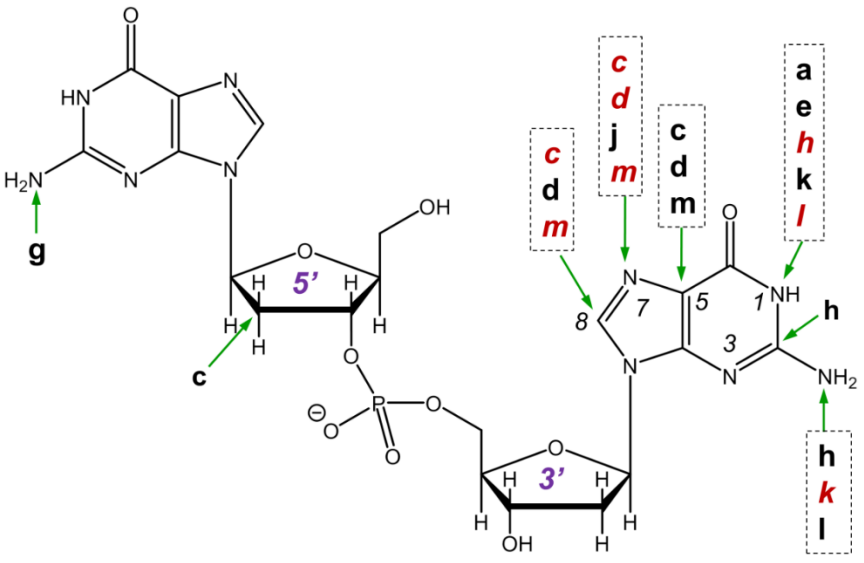


Figure 3.6. Close contacts of the incipient carbene atom in optimized (0 K) structures of $(dGG + C^*AQK + H)^+$. Bold: contacts within 4.5 Å; red italics: contacts within 4.0 Å. There were no 0 K contacts in complexes **b, i, n, o**, and **p**.

BOMD trajectories were run for 100 ps at 310 K for complexes **a-p**. The distribution of contacts between the incipient carbene (C136) and the guanine and deoxyribose units is shown in Figure 3.7 and further specified for the positions in the frequently visited 3′-guanine base, such as C-8 (C52), C-4 (C53), C-5 (C56), C-2 (C59), N-7 (N55), N-3 (N-57), N-1 (N61), and NH_2 (N62) that are ordered left to right in the distribution bar graphs in Figure 3.7. The overall contact frequency varied from 0.2% in complex **f** through 150% in complex **a**. The high figure exceeding 100% was due to multiple simultaneous contacts in the course of the 100,000-step trajectory. We focus the discussion of the contact data on the thermodynamically most stable complexes **a-d, j, k, l, m**, and **p**. Complexes **b, c, d**, and **m** showed a high proportion of contacts at the 3′-guanine N-7 position that correlated with the proximity of N-7 and the diazirine C136 in the 0 K structures (Figure 3.6).

The other positions developing contacts in **b, c, d,** and **m** were C-8 and C-5 (Figure 3.7). The total contact counts in **b, c, d**, and **m** were similar, ranging between 25 and 37% (Figure 3.7). The contacts in complexes c and m showed similar distributions indicating similar ranges of thermal motion of the 4,4-azipentyl substituent in these complexes. Thermal motion in complexes **a, j, k, l**, and **p** created contacts with the N-1 and NH_2 positions in 3′-guanine that were inaccessible in **b, c,** d, and **m**. In addition, **a** was the only low-energy complex that developed contacts in the 5′-nucleoside (Figure 3.7).

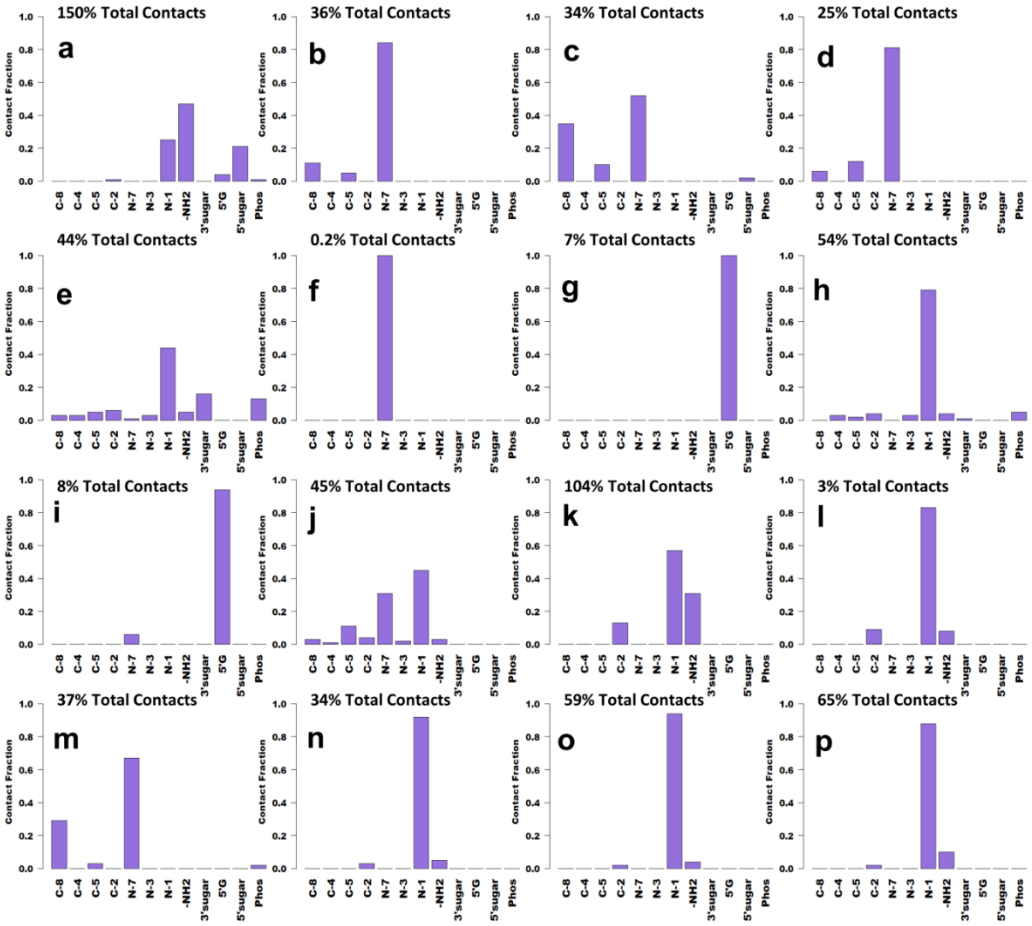


Figure 3.7. Distribution of close contacts in complexes **a**-**p** from BOMD trajectories of (dGG + C*AQK + H)$^+$ complexes. Positions C-8, C-4, C-5, C-2, N-7, N-1, and NH_2 refer to the 3'-guanine base.

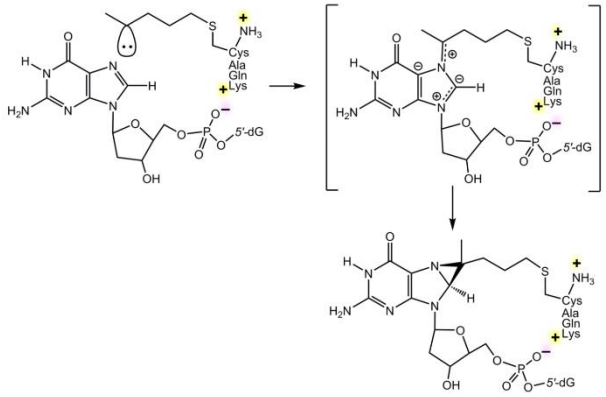
3.3.7 *Reaction Mechanisms for Covalent Bond Formation*

The high overall cross-linking efficiency in (dGG + C*AQK + H)$^+$ (46%, Table 3.1) indicated that a substantial fraction of the peptide carbenes that were formed at close contact with dGG reacted by forming covalent bonds. The trajectory data indicated that the majority of contacts occurred at 3′-guanine. This result was also consistent with the cross-linking data in the GC and CG complexes that showed preferential covalent bond formation in the 3′-nucleoside. The high incidence of close contacts at the 3′-nucleobase raised the question of the reaction mechanism for

the covalent bond formation. Contacts with the 3′-guanine N-1-H and NH_2 groups in complexes **j, k, l**, and **p** can proceed by the standard insertion mechanism[58] involving the N – H bonds. These insertions were expected to be 350–380 kJ mol^{-1} exothermic and thus thermodynamically driven. The exothermicity for the carbene insertion was estimated from the calculated reaction enthalpies for a model system, which was the insertion of propane-2-carbene (2) into the N-1-H and NH_2 bonds in 9-methyl guanine (**1**) forming 2-propylguanines **3** and **4** (Scheme S1, Supporting Information). In contrast, carbene reactions with heterocycles in general[59], and nucleosides in particular[60 61 62], have been studied only sparsely. Our calculations of a model propane-2-carbene addition to N-7 in 9-methylguanine revealed that the reaction was substantially exothermic, $\Delta H_{0,rxn}$ = − 152 to − 184 kJ mol^{-1} at different levels of theory (Scheme S1, Table S2), and therefore thermodynamically favorable. The addition initially formed a dipolar intermediate in the imidazole part of the nucleobase that can be characterized as a N-7-C-8-C-9 ylide (**5**). A similar intermediate has been observed in the dichlorocarbene reaction with pyridine[59]. Intermediate 5 can further react by closing a three-membered ring. The ring closure forming a N-7-C-8 fused azirine was calculated to be 66–81 kJ mol^{-1} exothermic, leading to a stable cross-linked product **6**. The activation energy for the ring closure, 112–119 kJ mol^{-1} for TS1 relative to **5** (Scheme S1) can be readily provided by the exothermic addition.

We envisaged an analogous mechanism for the carbene attack on guanine in the $(dGG + C^*AQK - N_2+H)^+$ complex as sketched in Scheme 3.2. Because of the substantial exothermicity of these reactions, we did not expect the 3′- and 5′-guanine rings to display distinct reactivity toward the carbene intermediate. Rather, we argue that the preferential attack at the 3′-nucleobase was guided by the conformations of the complexes that stemmed from the hydrogen bonding interactions between the polar phosphate, ammonium, and 5′-OH groups. At the same time, the

tight hydrogen bonding in the complexes allowed for only a limited conformational motion of the diazirine-carrying side chain whereas the core of the complex was not affected. This interpretation was consistent with the contact analysis of the 0 K structures and 310 K trajectories. For example, the low-energy complex **d** showed 0 K contacts with C-5, N-7, and C-8 in 3′-guanine (Figure 3.6), which developed into the most frequent contacts with 3′-N-7 at 310 K whereas the other positions were visited much less (Figure 3.7).



Scheme 3.2. Reaction of carbene with 3'-guanine in the $(dGG + C^*AQK\text{-}N_2 + H)^+$ complex.

3.3.8 *Effects of Peptide Complexation on the Dinucleotide Structure*

The optimized structures of low-energy complexes indicated strong dipolar interactions between dGG and C*AQK that were likely to also exist in complexes involving the other combinations of peptides and dinucleotides. To specify the effect of the peptide ion binding, we undertook a conformational search of dGG anions and compared their low-energy structures with those in the complexes. Three different initial monomer conformations were selected as starting points, which were of stacked, extended, and S-shape types. After conformation sorting, 13 non-degenerate structures were submitted to DFT geometry optimizations and frequency calculations to obtain the relative free energies. Based on this analysis, the lowest free-energy structures were of the stacked type (**dGG1** and **dGG2**, Fig. S11), whereas the semi-folded S-shaped **dGG3** and

extended **dGG4** were 28 and 31 kJ mol^{-1} higher in energy. The low-energy dGG structures displayed a stacked motif that was similar to that found in complexes **c, d**, and **m** (Figure 3.4) where the exposed phosphate anion was solvated by the peptide dication. The hydrogen bond between the phosphate and 5-OH in dGG1 and dGG2 was replaced by an H-bond to the lysine NH_3^+ group in complexes **c, d**, and **m**, which endowed them with additional stability and directed the lysine C-terminus toward the 5′-nucleobase. However, the energy data suggested that coordination with the peptide ion can also be realized with other low-energy complex structures (**a, b, j**) in which guanine stacking was absent. This led us to the conclusion that non-bonding interactions in these peptide-dinucleotide complexes could produce low free-energy protomers and conformers displaying substantial structural variability while conserving or disrupting nucleobase stacking interactions in the dinucleotides.

3.4 CONCLUSION

Singly charged complexes of diazirine-tagged peptides CAQK with several DNA dinucleotides were generated in the gas phase. Peptide carbene intermediates formed by photodissociation were found to undergo efficient covalent cross-linking that largely occurred in the 3′-nucleoside. Electronic structure and Born-Oppenheimer molecular dynamics calculations of the $(dGG + C^*AQK + H)^+$ complex revealed a variety of low-energy conformers of zwitterionic types having doubly protonated peptides along with deprotonated dinucleotide phosphate or peptide carboxyl groups. A common feature revealed by the cross-linking data in accord with the structure analysis was that the peptide and dinucleotide preferred orientations allowing close approach of the diazirine tag at the N-terminal cysteine to the 3′-nucleoside. Thermal motion at 310 K in the complexes resulted in an extended range of close contacts between the tag and the

nucleotide, but the core structure of the complex, as determined by hydrogen bonding, was not disrupted.

3.5 ACKNOWLEDGEMENTS

This research has received funding from the National Science Foundation Division of Chemistry (Grants CHE-1661815 and CHE-1624430). F.T. thanks the Klaus and Mary Ann Saegebarth Endowment for support. Y. L. thanks Dr. Emilie Viglino for generous help with peptide synthesis.

3.6 REFERENCES

1. von Hippel, P. H., *Annu. Rev. Biophys. Biomol. Struct.* **2007,** *36*, 79 - 105.
2. Raindlova, V.; Pohl, R.; Hocek, M., *Chemistry* **2012,** *18* (13), 4080-4087.
3. Carrette, L. L. G.; Morii, T.; Madder, A., *Bioconjugate Chemistry* **2013,** *24* (12), 2008-2014.
4. Flett, F. J.; Walton, J. G. A.; Mackay, C. L.; Interthal, H., *Analytical Chemistry* **2015,** *87* (19), 9595--9599.
5. Wickramaratne, S.; Boldry, E. J.; Buehler, C.; Wang, Y. C.; Distefano, M. D.; Tretyakova, N. Y., *Journal of Biological Chemistry* **2015,** *290* (2), 775--787.
6. Ming, X.; Groehler, A.; Michaelson-Richie, E. D.; Villalta, P. W.; Campbell, C.; Tretyakova, N. Y., *Chemical Research in Toxicology* **2017,** *30* (4), 980--995.
7. Buxton, K. E.; Kennedy-Darling, J.; Shortreed, M. R.; Zaidan, N. Z.; Olivier, M.; Scalf, M.; Sridharan, R.; Smith, L. M., *J Proteome Res* **2017,** *16* (9), 3433-3442.
8. Weir Lipton, M. S.; Fuciarelli, A. L.; Springer, D. L.; Hofstadler, S. A.; Edmonds, C. G., *Rapid Communications in Mass Spectrometry* **1997,** *11* (15), 1673-1676.
9. Golden, M. C.; Resing, K. A.; Koch, T. H.; Collins, B. D.; Willis, M. C., *Protein Science* **2008,** *8* (12), 2806--2812.
10. Wong, D. L.; Reich, N. O., *Biochemistry* **2000,** *39* (50), 15410--15417.
11. Rieger, R. A.; McTigue, M. M.; Kycia, J. H.; Gerchman, S. E.; Grollman, A. P.; Iden, C. R., *Journal of the American Society for Mass Spectrometry* **2000,** *11* (6), 505--515.
12. Steen, H.; Petersen, J.; Mann, M.; Jensen, O. N., *Protein Science* **2001,** *10* (10), 1989--2001.
13. Wagenknecht, H. A.; Rajski, S. R.; Pascaly, M.; Stemp, E. D. A.; Barton, J. K., *Journal of the American Chemical Society* **2001,** *123* (19), 4400--4407.
14. Doneanu, C. E.; Gafken, P. R.; Bennett, S. E.; Barofsky, D. F., *Analytical Chemistry* **2004,** *76* (19), 5667--5676.
15. Pourshahian, S.; Limbach, P. A., *J Mass Spectrom* **2008,** *43* (8), 1081-8.
16. Rosenfeld, K. K.; Ziv, T.; Goldin, S.; Glaser, F.; Manor, H., *J Mol Biol* **2011,** *410* (1), 77-92.
17. Steen, H.; Jensen, N., *Mass Spectrometry Reviews* **2002,** *21* (3), 163-82.
18. Tacheny, A.; Dieu, M.; Arnould, T.; Renard, P., Mass spectrometry-based identification of proteins interacting with nucleic acids. 2013; Vol. 94, pp 89--109.
19. Laughlin, S. a., May the best molecule win: Competition ESI mass spectrometry. 2015; Vol. 16, pp 24506--24531.
20. Groehler, A.; Degner, A.; Tretyakova, N. Y., Mass Spectrometry-Based Tools to Characterize DNA–Protein Cross-Linking by Bis-Electrophiles. 2017; Vol. 121, pp 63--77.
21. Lin, S.; Cotter, R. J.; Woods, A. S., *Proteins* **1998,** *Suppl 2*, 12-21.
22. Veenstra, T. D., Electrospray ionization mass spectrometry: A promising new technique in the study of protein/DNA noncovalent complexes. 1999; Vol. 257, pp 1--5.
23. Beck, J. L.; Colgrave, M. L.; Ralph, S. F.; Shei, M. M., *Mass Spectrometry Reviews* **2001,** *20* (2), 61--87.
24. Alves, S.; Woods, A.; Tabet, J. C., *J Mass Spectrom* **2007,** *42* (12), 1613-22.
25. Alves, S.; Woods, A.; Delvolvé, A.; Tabet, J. C., *International Journal of Mass Spectrometry* **2008,** *278* (2-3), 122-128.
26. Brahim, B.; Tabet, J. C.; Alves, S., *Eur J Mass Spectrom (Chichester)* **2018,** *24* (1), 168-177.

27. Liu, Y.; Ramey, Z.; Turecek, F., *Chemistry - A European Journal* **2018,** *24* (37), 9259--9263.
28. Mann, A. P.; Scodeller, P.; Hussain, S.; Joo, J.; Kwon, E.; Braun, G. B.; Mlder, T.; She, Z. G.; Kotamraju, V. R.; Ranscht, B.; Krajewski, S.; Teesalu, T.; Bhatia, S.; Sailor, M. J.; Ruoslahti, E., *Nature Communications* **2016,** *7*.
29. Huang, R. C.; Bonner, J., *Proceedings of the National Academy of Sciences of the United States of America* **1962,** *48*, 1216--1222.
30. Bannister, A. J.; Kouzarides, T., Regulation of chromatin by histone modifications. 2011; Vol. 21, pp 381--395.
31. Strahl, B. D.; Allis, C. D., The language of covalent histone modifications. 2000; Vol. 403, pp 41--45.
32. Jenuwein, T.; Allis, C. D., Translating the histone code. 2001; Vol. 293, pp 1074--1080.
33. Pepin, R.; Shaffer, C. J.; Turecek, F., *Journal of Mass Spectrometry* **2017,** *52* (8), 557-560.
34. Das, J., *Chemical Reviews* **2011,** *111* (8), 4405-4417.
35. Nguyen, H. T. H.; Andrikopoulos, P. C.; Rulek, L.; Shaffer, C. J.; Tureek, F., *Journal of the American Society for Mass Spectrometry* **2018,** *29* (8), 1706--1720.
36. Shaffer, C. J.; Andrikopoulos, P. C.; Řez, *Journal of the American Society for Mass Spectrometry* **2016,** *27* (4), 633--645.
37. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., *Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.
38. Stewart, J. J. P., *Journal of Molecular Modeling* **2007,** *13* (12), 1173-1213.
39. Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P., *Journal of Chemical Theory and Computation* **2009,** *5* (7), 1749-1760.
40. Stewart, J. J. P. *MOPAC 16.*, Stewart Computational Chemistry, Colorado Springs, CO, 2016.
41. Rezac, J., *Journal of Computational Chemistry* **2016,** *37* (13), 1230-1237.
42. Řezáč, J. *Cuby—ruby framework for computational chemistry*, 4.
43. Becke, A. D., *Physical Review A* **1988,** *38* (6), 3098-3100.
44. Chai, J.-D.; Head-Gordon, M., *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615-6620.
45. Tomasi, J.; Mennucci, B.; Cammi, R., *Chemical Reviews* **2005,** *105* (8), 2999-3093.
46. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. A.01*, Wallingford, CT, 2016.
47. Marek, A.; Turecek, F., *Journal of the American Society for Mass Spectrometry* **2014,** *25* (5), 778--789.
48. Seburg, R. A.; McMahon, R. J., *Journal of the American Chemical Society* **1992,** *114* (18), 7183--7189.
49. Jackson, J. E.; Soundararajan, N.; White, W.; Liu, M. T. H.; Bonneau, R.; Platz, M. S., *Journal of the American Chemical Society* **1989,** *111* (17), 6874--6875.
50. Pezacki, J. P.; Couture, P.; Dunn, J. A.; Warkentin, J.; Wood, P. D.; Lusztyk, J.; Ford, F.; Platz, M. S., *Journal of Organic Chemistry* **1999,** *64* (12), 4456--4464.
51. Stevens, I. D. R.; Liu, M. T. H.; Soundararajan, N.; Paike, N., *Tetrahedron Letters* **1989,** *30* (4), 481--484.
52. Murray, K. K., *Journal of Mass Spectrometry* **1996,** *31* (11), 1203-1215.
53. McLuckey, S. A.; Berkel, G. J.; Glish, G. L., *Journal of the American Society for Mass Spectrometry* **1992,** *3* (1), 60-70.
54. Roepstorff, P.; Fohlman, J., *Biological Mass Spectrometry* **1984,** *11* (11), 601--601.
55. Biemann, K., *Methods in Enzymology* **1990,** *193*, 886-887.
56. Wu, J.; McLuckey, S. A., Gas-phase fragmentation of oligonucleotide ions. 2004; Vol. 237, pp 197--241.
57. Schurch, S., *Mass Spectrom Rev* **2016,** *35* (4), 483-523.
58. Liu, M. T., *Chemistry of diazirines*. CRC: 1987; Vol. 1.
59. Tippmann, E. M.; Platz, M. S.; Svir, I. B.; Klymenko, O. V., *Journal of the American Chemical Society* **2004,** *126* (18), 5750--5762.
60. Mohan, U.; Burai, R.; McNaughton, B. R., *Tetrahedron Letters* **2014,** *55* (22), 3358--3360.
61. Nowak, I.; Robins, M. J., *Journal of Organic Chemistry* **2007,** *72* (9), 3319--3325.
62. Jerbi, J.; Springborg, M., *International Journal of Quantum Chemistry* **2018,** *118* (11).

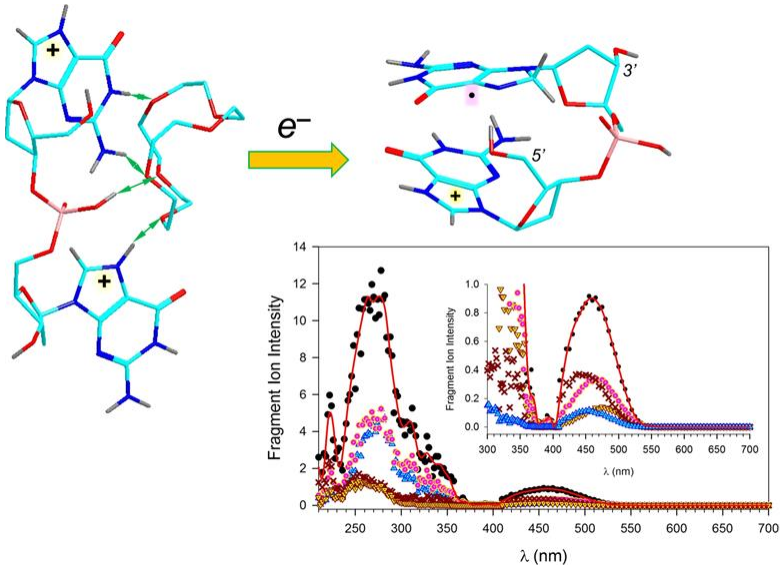
Chapter 4. HYDROGEN-RICH CATION RADICALS OF DNA DINUCLEOTIDES: GENERATION AND STRUCTURE ELUCIDATION BY UV-VIS ACTION SPECTROSCOPY

Reproduced in part with permission from Liu, Y.; Korn, J. A.; Dang, A.; Turecek, F., *Journal of Physical Chemistry B* **2018,** *122* (42), 9665-8680.

Abstract

Hydrogen-rich DNA dinucleotide cation radicals $(dGG + 2H)^{+\bullet}$, $(dCG + 2H)^{+\bullet}$, and $(dGC + 2H)^{+\bullet}$ represent transient species comprising protonated and hydrogen atom adducted nucleobase rings that serve as models for proton and radical migrations in ionized DNA. These DNA cation radicals were generated in the gas phase by electron-transfer dissociation of dinucleotide dication−crown−ether complexes and characterized by UV−vis photodissociation action spectra, ab initio calculations of structures and relative energies, and time-dependent density functional theory calculations of UV−vis absorption spectra. Theoretical calculations indicate that $(dGG + 2H)^{+\bullet}$ cation radicals formed by electron transfer underwent an exothermic conformational collapse that was accompanied by guanine ring stacking and facile inter-nucleobase hydrogen atom transfer, forming 3′-guanine C-8−H radicals. In contrast, exothermic hydrogen transfer from the 5′-cytosine radical onto the guanine ring in $(dCG + 2H)^{+\bullet}$ was kinetically hampered, resulting in the formation of a mixture of 5′-cytosine and 3′-guanine radicals. Conformational folding and nucleobase stacking were energetically unfavorable in $(dGC + 2H)^{+\bullet}$ that retained its structure of a 3′-cytosine radical, as formed by one-electron reduction of the dication. Hydrogen-rich guanine $(G + H)^{\bullet}$ and cytosine $(C + H)^{\bullet}$ radicals were calculated to have vastly different basicities in water,

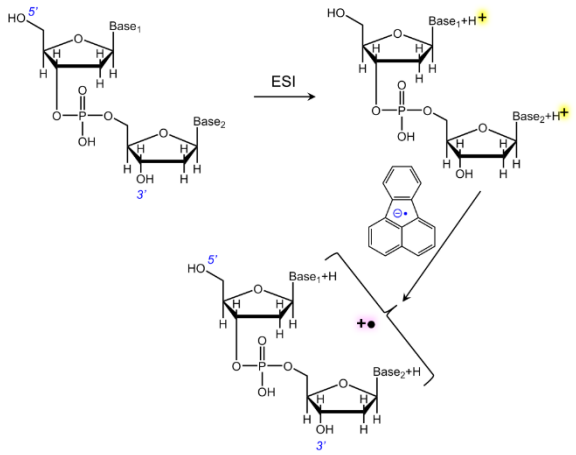
as illustrated by the respective pK_a values of 20.0 and 4.6, which is pertinent to their different abilities to undergo proton-transfer reactions in solution.



4.1 INTRODUCTION

DNA ionization by high-energy particles is a major cause of chemical changes leading to the complex process of DNA damage[1].1 Ionization randomly creates an electron defect (a "hole") in a nucleobase that can propagate by charge transfer along the double strand to generate reactive intermediates undergoing further radical reactions. [2-7] Although the chemistry of hole generation and propagation dynamics has been studied for over two decades with ingenious model systems, the intrinsic electronic properties of the reactive cation radical intermediates have been elucidated only recently. [8-14] For example, Schuster and co-workers examined adenine−thymine-rich nucleic acid models that upon one-electron oxidation underwent reactions, indicating proton-coupled electron transfer between the ionized adenine and thymine nucleobases. [8] Proton transfer between ionized cytosine and guanine nucleobases in Watson−Crick pairs has been studied computationally[15-23] and by gas-phase experiments. [24]

In addition to standard studies of DNA ionization in solution, there have been several studies aimed at generating and characterizing nucleobase, [25-32] and nucleoside[24] radicals and cation radicals in the gas phase. By excluding interactions with solvent and counterions, gas-phase studies allow one to address the intrinsic electronic and chemical properties of the radical species of interest. [33] In addition, by using tandem mass spectrometry, one can achieve specific generation and isolation of cation radicals that can be further probed by UV−vis photodissociation action spectroscopy (UVPD) in the infrared and UV−visible regions of the spectrum. Recently, we have reported a novel type of DNA and chimeric RNA dinucleotide cation radicals that were studied in the gas phase. [34] These novel radical species were generated by ion−ion reactions of doubly protonated dinucleotides with molecular anions,[35] specifically fluoranthene anion radicals[36] (Scheme 4.1), whereby one-electron reduction in the gas phase was utilized as a versatile synthetic tool. We call these species hydrogen-rich DNA cation radicals because they contain two additional hydrogen atoms, one attached as the charging proton and the other introducing the unpaired electron into the DNA molecule. The term “hydrogen-rich” was coined previously for peptide cation radicals analogously generated by electron transfer to multiply protonated peptides and proteins. [37 38]



Scheme 4.1. Formation of Hydrogen-Rich Dinucleotide Cation Radicals by One-Electron Reduction of Dications.

Hydrogen-rich DNA cation radicals having adenine nucleobases, such as $(dAA + 2H)^{+\bullet}$, have been found to exhibit interesting electronic properties and reactivity.[34] In particular, electron transfer to the dications caused major conformational changes associated with the stacking of the adenine cation and radical moieties that was followed by spontaneous inter-base proton transfer.

Hydrogen-rich DNA dinucleotide radicals are related to reactive intermediates produced by the attachment of low-energy electrons to DNA nucleobases[39]. With purine nucleobases, fast electron attachment forms a transient anion radical, which is protonated by solvent on a microsecond time scale[40]. The resulting radical is a hydrogen-atom adduct to the nucleobase, which is analogous to the gas-phase species generated by electron transfer to the protonated nucleobase. Hence, by studying hydrogen-rich DNA dinucleotide cation radicals in the gas phase, one can shed light on the intrinsic electronic properties of transient radicals formed in solution.

We now extended the study of DNA dinucleotide cation radicals to involve guanine and cytosine nucleobases, as in $(dGG + 2H)^{+\bullet}$, $(dCG + 2H)^{+\bullet}$, and $(dGC + 2H)^{+\bullet}$. Our goal was to determine the structures of the dication precursors and characterize the hydrogen-rich cation radicals formed by electron transfer using tandem mass spectrometry and UV−vis photodissociation action spectroscopy (UVPD). UVPD is based on resonant absorption of a photon over a scanned range of wavelengths, leading to dissociation to fragment ions. The wavelength-dependent profile of the fragment ion intensities is used to track absorbance of the cation radical, producing an action spectrum, which is analogous to an absorption spectrum. [41,42] The mass-resolved channels of the action spectrum can further provide information on dissociations from particular electronic excited states of the cation radical. Action spectra are assigned to cation

radical structures by comparison with absorption spectra obtained by quantum chemistry methods, such as time-dependent density functional theory (TD-DFT) [43] or equation-of-motion coupled clusters calculations.[44] Previous UVPD studies have found a close agreement between the experimental UV−vis action spectra and the calculated absorption spectra for several peptide and nucleobase cation radicals, [31, 32, 34 45-47] suggesting UVPD action spectroscopy as a powerful tool for structure elucidation of transient radicals. Here, we wish to show by experiment and theory that purine and pyrimidine nucleobases in dinucleotide cation radicals exhibit different properties regarding nucleobase stacking, hydrogen bonds, and propensity for undergoing intramolecular proton transfer.

4.2 Experimental Section

4.2.1 *Materials and Methods*

Dinucleotides dGG, dGC, and dCG were custom-synthesized by Integrated DNA Technologies (Coralville, IA) and used as received. Dibenzo-18-crown-6-ether (DBCE, CAS Number 14187-32-7) was purchased from Sigma-Aldrich (Milwaukee, WI). All solvents were of high-performance liquid chromatography analytical grade. Mass spectra were measured on an LTQ-XL-ETD linear ion trap mass spectrometer (Thermo Fisher, San Jose, CA) equipped with an auxiliary chemical ionization source for the production of fluoranthene anions. Doubly charged precursor ions were generated in a home-built micro-electrospray ion source consisting of a fused silica capillary pulled to a 10 μm diameter tip and maintained at 2.2 kV. Methanol/water/acetic acid (50/50/1) solutions containing the dinucleotide and DBCE in a 1:1.2 molar ratio at 10−20 μM concentrations were sprayed at 1.5−2 μL min^{-1} flow rates. $(dGC + DBCE + 2H)^{2+}$ precursor ions were generated from acetonitrile/water/acetic acid (50/50/1) solutions at 10−20 μM

concentrations. Electron-transfer dissociation (ETD) mass spectra were obtained by ion−ion charge-transfer reactions with fluoranthene anion radicals at 150−200 ms reaction times. The charge-reduced dinucleotide cation radicals were selected by mass with 2−3 mass unit isolation windows. UVPD was performed as described previously. [45] Briefly, the mass-isolated and trapped ions were irradiated by light pulses (3−6 ns pulse width, 20 Hz repetition rate) produced by an EKSPLA NL301G (Altos Photonics, Bozeman, MT) Nd:YAG laser source equipped with a PG142C optical parametric oscillator. This provided wavelength tuning in the three ranges, 210−354, 355−400, and 400−700 nm, with a laser power of 0.52−12.69 mJ per pulse. The laser power was measured at each wavelength, typically in 2 nm steps. The intensities of the resulting UVPD MS^3 fragment ions were monitored as a function of wavelength, normalized to the laser output power at each wavelength, and plotted to provide action spectra.

4.2.2 *Calculations*

Precursor dication structures were obtained for noncovalent complexes with 18-crown-6-ether (CE) as a surrogate for DBCE. Previous structure analysis has indicated very similar structures and relative energies for dication complexes of dAA with CE and DBCE, [34] justifying the use of the smaller CE in the computational analysis of precursor ion structures. Conformation analysis of the dication CE complexes was performed by Born−Oppenheimer molecular dynamics (BOMD) calculations using the semiempirical all-valence-electron PM6 method[48] supplemented with dispersion corrections, [49] PM6-D3H4, and run by MOPAC [50] under the Cuby4 platform,[51] as described previously.[52] Born−Oppenheimer molecular dynamics (BOMD) trajectories were run at 310−410 K using the Berendsen thermostat algorithm.[53] The conformational search was carried out for several protomers of the dinucleotides, as described later in the paper. Selected low-energy conformers from the BOMD runs of each complex protomer were reoptimized by density

functional theory (DFT) calculations using the B3LYP [54] and ωB97X-D[55] hybrid functionals. The B3LYP calculations were performed with the 6-31G(d,p) basis set to provide harmonic frequencies that were used to calculate ion enthalpies and entropies at the ion trap temperature (310 K). The ωB97X-D calculations were carried out with the 6-31+G(d,p) basis set to provide the electronic terms. The combined electronic, enthalpy, and entropy terms were combined to produce free energies that were used to rank conformers of each protomeric complex. In addition, solvation energies in water, methanol, and acetonitrile were calculated with ωB97X-D/6-31+G(d,p) using the polarizable continuum model. [56] The lowest-free-energy ions were used for the selection of protomers in the charge-reduced dinucleotide cation radicals. Cation radical structures were optimized by B3LYP, ωB97X-D, and M06-2X [57] calculations with the 6-31+G(d,p) basis set that were run within the spin-unrestricted formalism. Single-point energies were obtained by B3LYP, ωB97X-D, M06-2X, and Møller−Plesset [58] (MP2-(frozen core)) calculations with the larger 6-311++G(2d,p) basis set. High-spin states in spin-unrestricted UMP2 energies were annihilated using the standard procedures[59] [60] that reduced the spin expectation values to $\langle S^2 \rangle < 0.76$. Additionally, restricted open-shell (ROMP2) single-point energy calculations were performed to assess the effect of spin contamination on the calculated energies[61]. Atomic spin densities were calculated using the natural population analysis [62] of the M06-2X/6-311++G(2d,p) wave functions. Excitation energies and oscillator strengths were obtained for 45−55 excited states by time-dependent DFT calculations[43] with ωB97X-D and M06-2X/6-31+G(d,p), according to previous benchmarking studies.[31,32] Excited states having spin expectation values <S2> > 1.5 were discarded. All of the electronic structure calculations were performed with the Gaussian 16 (revision A.03) suite of programs.[63] Rice−Ramsperger−Kassel−Marcus (RRKM) calculations of unimolecular rate constants [64] were performed using the QCPE program[65] that was recompiled for

Windows 7. [66] Rotations were treated adiabatically, and the rate constants obtained by direct count of quantum states were Boltzmann-averaged over the distribution of ion rotational states at 310 K.

4.3 Results and Discussion

4.3.1 *Dinucleotide Cation Radical Generation*

Electrospray ionization of dinucleotides dGG, dCG, and dGC produced doubly protonated cations $(dGG + 2H)^{2+}$, $(dCG + 2H)^{2+}$, and $(dGC + 2H)^{2+}$, respectively, which were selected by mass and subjected to ion−ion reactions with fluoranthene anion radicals. Electron transfer to dinucleotide dications resulted in predominant dissociation, yielding only minor fractions of cation radicals. This is illustrated by the electron-transfer dissociation (ETD) mass spectrum of $(dGG + 2H)^{2+}$ (Figure S1) that displayed fragment ions formed by loss of H, nucleobase, and the formation of protonated nucleobases. Hence, ETD of the dinucleotide dications was unsuitable for the generation of stable hydrogen-rich cation radicals. We note that larger oligonucleotides have been shown to form very stable cation radicals upon electron transfer to multiply charged ions, which has been deemed detrimental to using ETD for DNA sequence analysis. [67] To efficiently generate the pertinent cation radicals from dinucleotides, we resorted to a previously employed method, which is based on noncovalent ion−molecule complexes of the dication with crown ethers. [34 68] These are formed as doubly charged ions by electrospray and subjected to ETD. Electron transfer results in charge reduction and elimination of a neutral crown ether molecule, forming intact hydrogen-rich dinucleotide cation radicals in excellent yield. Figure 4.1a−c illustrates the formation by ETD of $(dGG + 2H)^{+\bullet}$, $(dCG + 2H)^{+\bullet}$, and $(dGC + 2H)^{+\bullet}$ from complexes with dibenzo-18-crown-6-ether (DBCE). With dGG, the doubly charged complex $(dGG + DBCE + 2H)^{2+}$ (*m/z* 479, Figure 4.1a) underwent efficient ETD to form the $(dGG + 2H)^{+\bullet}$ cation radical at

m/z 598, which was accompanied by minor $(dGG + H)^+$ (*m/z* 597), $(dGG + DBCE + H)^+$ (*m/z* 957), and $(DBCE + NH_4)^+$ (*m/z* 378) fragment ions. Similar results were found for the other dinucleotide complexes.

Cation radicals with cytosine at the 5′-terminus, $(dCG + 2H)^{+\bullet}$, showed a slightly higher propensity for side reactions, such as elimination of H and water, as indicated by the corresponding *m/z* 557, 517, 540, and 500 fragment ions in the ETD spectra of the complexes (Figure 4.1b).

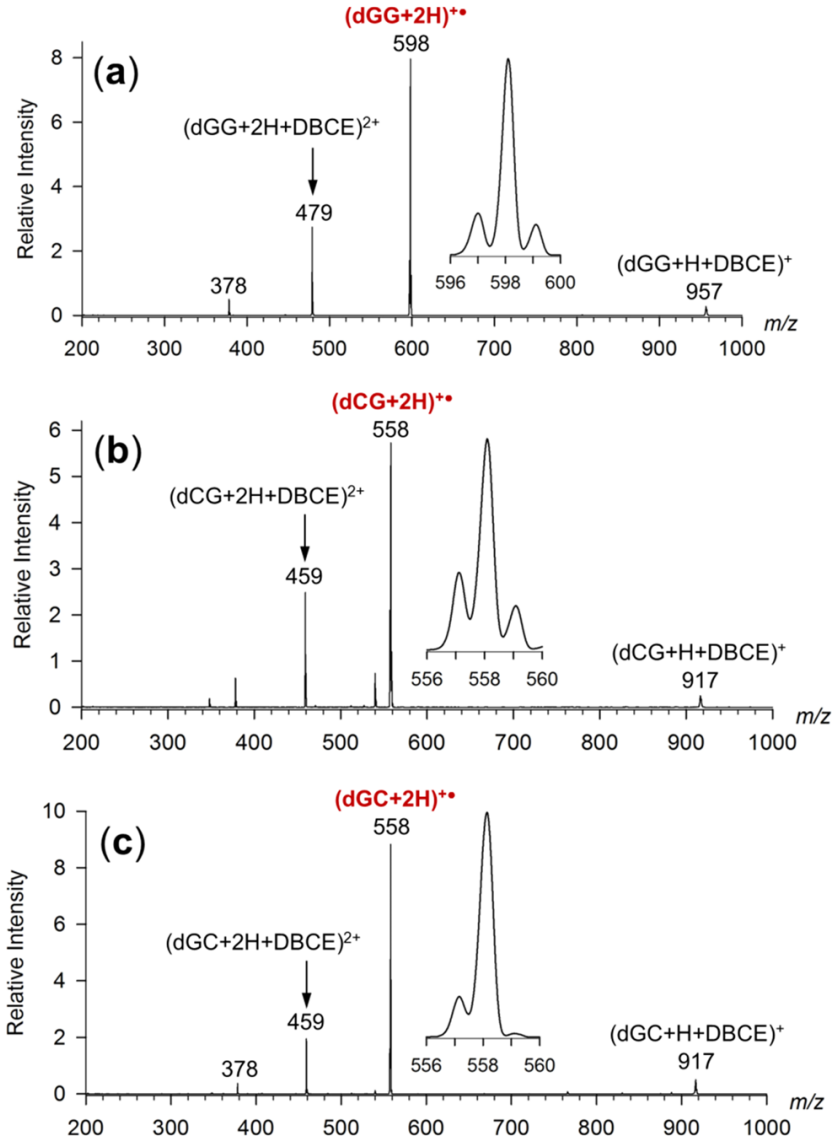


Figure 4.1. ETD mass spectra (fluoranthene anion, reaction time 150 ms) of dinucleotide dication complexes with dibenzocrown ether(DBCE) (a) $(dGG + DBCE + 2H)^{2+}$, (b) $(dCG + DBCE + 2H)^{2+}$,and (c) $(dGC + DBCE + 2H)^{2+}$.

4.3.2 *Cation Radical Collision-Induced Dissociation and Photodissociation Spectra*

The dinucleotide cation radicals produced by ETD of the crown−ether complexes were selected and isolated by their mass-to-charge ratios and characterized by collision-induced dissociation ($CID\text{-}MS^3$) and photodissociation mass spectra to identify and compare the major fragmentation pathways. For the DNA backbone fragment ion nomenclature, see ref [69]. $CID\text{-}MS^3$ of $(dGG + 2H)^{+\bullet}$ resulted in a dominant competitive loss of guanine and $(guanine + H)^{\bullet}$ radical to give the respective *m/z* 447 and 446 fragment ions (Figure 4.2a). Backbone cleavage forming the isobaric d_1 and w_1 ions (*m/z* 348) and their H-atom adducts (*m/z* 349) was also observed. $CID\text{-}MS^3$ of $(dGC + 2H)^{+\bullet}$ and $(dCG + 2H)^{+\bullet}$ yielded dramatically different spectra (Figure 4.2b,c). The cation radical with 3′-cytosine underwent a number of dissociations, including a dominant loss of water (*m/z* 540), cytosine (*m/z* 447), and guanine (*m/z* 406), 3′-deoxyribose cross-ring cleavages (labeled xx, *m/z* 512, 484, 403, 390), and backbone cleavages with charge retention on the 5′-guanosine fragment (d_1, *m/z* 348) and 3′-cytidine fragments, $(w_1 + 2H)^+$ (*m/z* 310), and $(z_1 + H)^{+\bullet}$ (*m/z* 211). Contrasting this, $CID\text{-}MS^3$ of $(dCG + 2H)^{+\bullet}$ showed major loss of water (*m/z* 540), $(cytosine + H)^{\bullet}$ radical (*m/z* 446), and cytosine (*m/z* 447), whereas loss of guanine was very minor (*m/z* 406). Backbone cleavage resulted in the retention of charge on the 3′-guanine fragment (w_1, *m/z* 348) and its H-atom adduct, $(w_1 + H)^{+\bullet}$, at *m/z* 349. The cationradical dissociations allowed us to deduce some common features of the nucleobase nature and position affecting the spectra. First, a cytosine at the 3′-terminus was highly conducive to facile elimination of water (Figure 4.2b). Second, a cytosine at the 5′-terminus was chiefly lost as a hydrogen adduct radical (Figure 4.2c) rather than a molecule. Third, the loss of cytosine outcompeted loss of guanine. Finally, the

cytosine fragments effectively competed for the charging proton when the nucleobase was at the 3′-terminus (Figure 4.2c).

Photodissociation of the dinucleotide cation radicals produced fragment ions that were similar to those observed by CID. This is illustrated by UVPD mass spectra obtained at 280 nm, where $(dGG + 2H)^{+\bullet}$, $(dCG + 2H)^{+\bullet}$, and $(dGC + 2H)^{+\bullet}$ exhibited strong absorption (Figure S2a−c, Supporting Information). UVPD of $(dGG + 2H)^{+\bullet}$ resulted in loss of guanine and $(guanine + H)^{\bullet}$ radicals (*m/z* 447 and 446, respectively) and formation of backbone d_1/w_1 fragment ions and their hydrogen adducts (*m/z* 348, 349, Figure S2a). UVPD of $(dCG + 2H)^{+\bullet}$ resulted in loss of cytosine and $(cytosine + H)^{\bullet}$ radicals, and the formation of backbone 5′-guanine containing w_1 ions (*m/z* 348), indicating prevalent dissociation in the cytidine moiety (Figure S2b). UVPD of $(dGC + 2H)^{+\bullet}$ gave rise to several fragment ions, namely, *m/z* 447 (loss of C), *m/z* 403 (cross-ring fragment), *m/z* 310 ($w_1 + 2H$), and *m/z* 308 ($w_1 - 2H$) (Figure S2c). It should be noted that strong absorption at 280 nm resulting in photodissociation was also observed for closed-shell cations $(dGG + H)^{+}$, $(dCG + H)^{+}$, dications $(dGG + 2H)^{2+}$, and $(dCG + 2H)^{2+}$, as evidenced by the respective ion action spectra (Figures S3 and S4, Supporting Information). Similar absorption has been reported for aqueous dG-5′-phosphate.[70] [71]Thus, electronic excitation at 280 nm in the gas-phase cation radicals cannot be unambiguously assigned to chromophores associated with the nucleobase radical moieties.

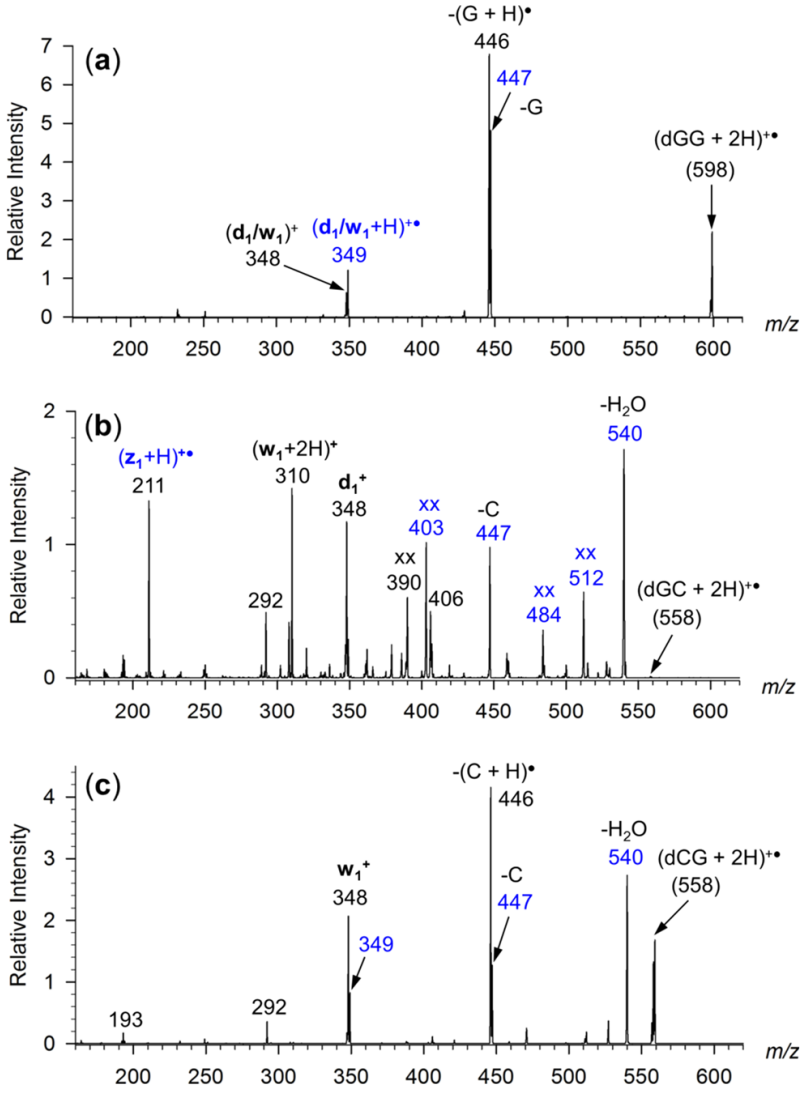


Figure 4.2. CID-MS^3 spectra of dinucleotide cation radicals: (a) $(dGG + 2H)^{+\bullet}$, (b) $(dGC + 2H)^{+\bullet}$, and (c) $(dCG + 2H)^{+\bullet}$. Cation-radical fragments are labeled blue.

4.3.3 *$(dGG + 2H)^{+\bullet}$ Action Spectroscopy and Ion Structures*

The main photodissociation fragment ions were used to monitor absorption by the gas-phase cation radicals in the 210−700 nm region of the spectrum. Starting with $(dGG + 2H)^{+\bullet}$, the cation radical displayed a long-wavelength band at 450−470 nm that was composed of contributions from all major dissociation channels (Figure 4.3a). The mass-resolved channels showed a slightly shifted absorption maxima that appeared at 450 nm for *m/z* 446 and 349, whereas

at 470 nm for *m/z* 447 and 348. The action spectrum further showed two bands at 330 and 310 nm and a major band centered at 270 nm. The latter band was chiefly composed of contributions from the dissociation channels by loss of G and $(G + H)^{\bullet}$ radical. The backbone dissociations forming the $(d_1/w_1)^+$ and $(d_1/w_1 + H)^{+\bullet}$ fragment ions gave a band that was slightly blue-shifted, peaking at 260 nm. Another band appeared at 210 nm, which received contributions from all major dissociation channels.

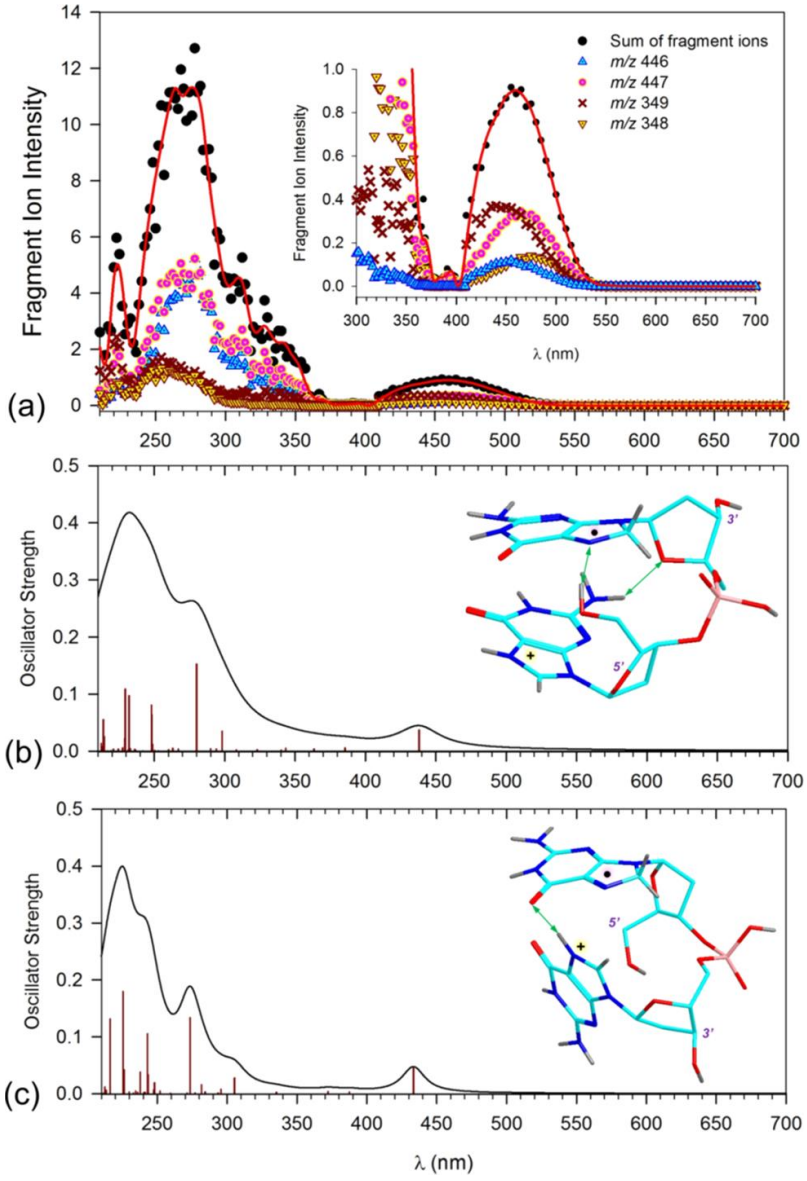


Figure 4.3. (a) Photodissociation action spectrum of $(dGG+ 2H)^{+\bullet}$.

The inset shows an expanded 300−700 nm region. M06-2X/6-31+G(d,p) TD-DFT calculated absorption spectra of (b) GG1e$^{+\bullet}$ and(c) GG1f$^{+\bullet}$. The bars represent the calculated absorption wavelengths and oscillator strength. The band shapes are from artificial line broadening with Lorenztian functions at 12 nm full width at half-maximum.

To aid interpretation of the action spectra, we carried out extensive calculations of structures and energies of dication precursors and cation radicals. A detailed study of the dication−crown complexes was necessary to identify low-energy dinucleotide protomers that could be expected to be formed by electrospray. The (dGG + 2H)$^{2+}$ cations are discussed first. With dGG, we considered structures with nucleobases protonated at N-7 of the canonical tautomer, which is the most basic position in guanine. [72 73 74] Another set had the 6-OH tautomer of guanine protonated at N-3, which is the second most basic position.[72] The crown ether ligand was coordinated by the 3′- or 5′-nucleobase cations of either protomeric type. As established previously for dAA complexes, [34] coordination of dibenzo-18-crown-6-ether is very similar to that of 18-crown-6-ether and can be established by computations for complexes of the latter as surrogates. BOMD of these protomers and crown−ether positional isomer combinations, followed by DFT optimization yielded structures that were ranked by free energy. Several low-free-energy structures were reoptimized with ωB97X-D, including polarizable continuum model to estimate solvation energies of the complexes in water and acetonitrile (Table S1, Supporting Information). The lowest-free-energy structures of each type are depicted in Figures 4 and S5 (Supporting Information). Among the positional isomers, those having the crown ether ligand at the 5′-terminus were consistently more stable than the 3′-terminus isomers. According to DFT calculations, the canonical N-7 protonated ion **dGG1a^{2+}** and its conformer **dGG1b^{2+}** were the most stable structures in both the gas phase and solution (Table S1).

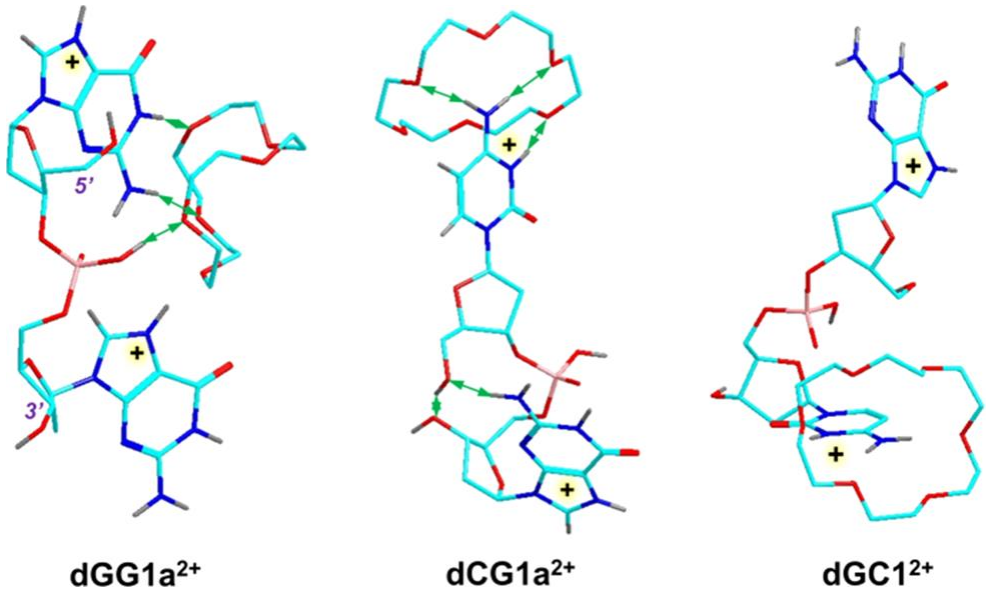


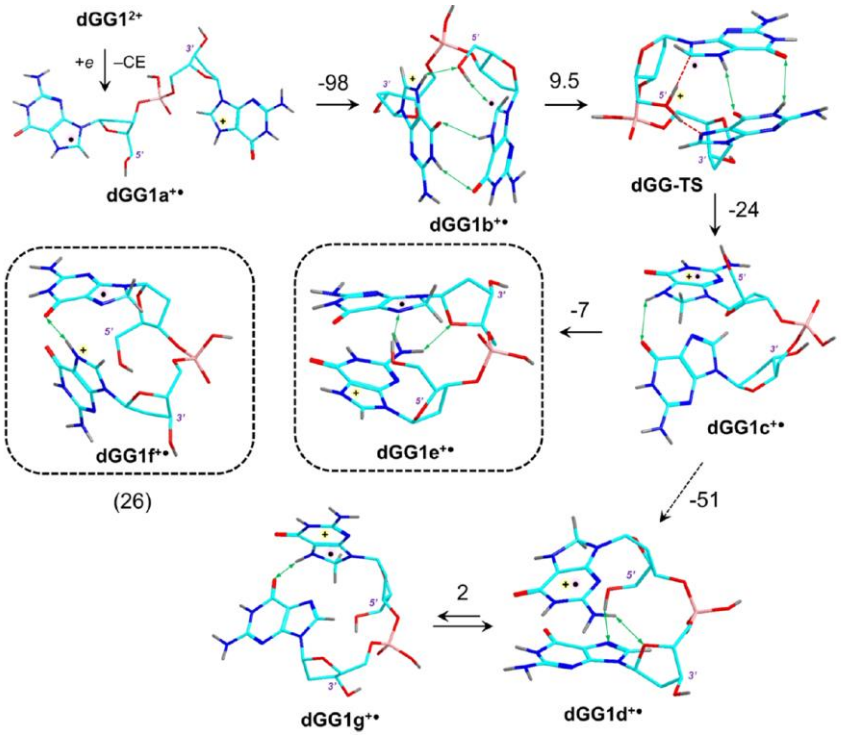
Figure 4.4. Lowest-free-energy dication dinucleotide−crown−ether complexes from ωB97X-D/6-31+G(d,p) geometry optimizations.

Atom color coding is as follows: cyan = C, blue = N, red = O, orange = P, gray = H. Only exchangeable and nucleobase hydrogens are shown..

The structures combining N-7 protonation on one guanine with N-3 protonation of the 6-OH enol tautomer of the other guanine (**dGG2**$^{2+}$ and **dGG3**$^{2+}$, respectively) were close enough in energy to be considered as potential candidates for $(dGG + 2H)^{+\bullet}$ cation radicals produced by ETD.

Electron transfer to **dGG1**$^{2+}$ followed by crown−ether loss can be viewed as originally producing cation radical **dGG1a**$^{+\bullet}$ in which the nucleobases were remote and the incoming electron was attached to the 5′-nucleobase (Scheme 4.2). Interestingly, attempts to fix the radical within the 3′-guanine nucleobase by puckering the ring in the initial guess geometry resulted in a relaxation upon gradient geometry optimization, yielding local energy minima that were 5′-guanine radicals. The calculated relative energies of $(GG + 2H)^{+\bullet}$ cation radicals are displayed in Table 4.1. In a further conformational search, ion **dGG1a**$^{+\bullet}$ underwent a highly exothermic folding to form conformer **dGG1b**$^{+\bullet}$, in which the radical and cation guanine nucleobases were stacked.

The further development of this structure likely involved a proton transfer between the N-7 and C-8 guanine positions in **dGG1b**$^{+\bullet}$ that was relayed by the 5′-OH group and proceeded through a low-energy transition state (**dGG-TS1**), exothermically yielding an isomeric cation radical **dGG1c**$^{+\bullet}$. The latter isomer can further refold to other low-energy structures, such as **dGG1d**$^{+\bullet}$ or its rotamer **dGG1g**$^{+\bullet}$. As a result of the proton transfer, ions **dGG1c**$^{+\bullet}$, **dGG1d**$^{+\bullet}$, and **dGG1g**$^{+\bullet}$ had 3′-guanine as a neutral moiety, whereas 5′-nucleobase was turned into a 7,8-dihydroguanine cation radical (Scheme 4.2). An inverse proton transfer from 5′-7,8-dihydroguanine to N-7 of the 3′-nucleobase can form isomer **dGG1e**$^{+\bullet}$ with a reversed position of the neutral radical and cation nucleobase moieties. We also considered an alternative isomerization process resulting in the formation of a 3′-guanine radical and 5′-cation (**dGG1f**$^{+\bullet}$), representing another low-energy isomer.

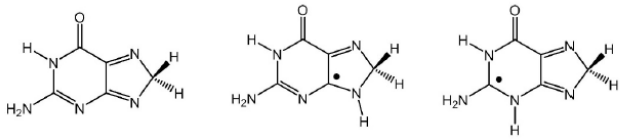


Scheme 4.2. Formation of (GG + 2H)$^{+\bullet}$ Cation Radicals[a].

[a]Reaction enthalpies (kJ mol^{-1}) are from M06-2X/6-311++G(2d,p) calculations including zero point vibrational energies and refer to 0 K. For energies from other calculations, see Table 4.1.

Table 4.1. Relative energies of $(dGG + 2H)^{+\bullet}$ ions

Ion	Relative Energy[a,b] B3LYP	B97X-D[c]	M06-2X[c]	PMP2[c]	ROMP2[c]
dGG1a$^{+\bullet}$	56	-	98	129	-
dGG1b$^{+\bullet}$	0	0	0	0	0
dGG1c$^{+\bullet}$	–39	–19	–24	–16	–25
dGG1d$^{+\bullet}$	–46	–72	–75	–80	–88
dGG1e$^{+\bullet}$	–9	–32	–30	–31	–31
dGG1f$^{+\bullet}$	–15	3	2	24	22
dGG2a$^{+\bullet}$	114	68	107	-	-
dGG2b$^{+\bullet}$	68	55	74	-	-
dGG3a$^{+\bullet}$	61	75	66	-	-
dGG3b$^{+\bullet}$	35	57	55	-	-
dGG4a$^{+\bullet}$	–32	–19	–20	2	–8
dGG4b$^{+\bullet}$	–29	–49	–41	–23	–32
dGG-TS	6	12	9	–9	–19
dGG1f$^{+\bullet}$ → loss of G	147[d]	205[d]	201[d]	189[d]	187[d]
dGG1f$^{+\bullet}$ → loss of (9H-G)$^{\bullet}$	77[e]	146[e]	140[e]	137[e]	155[e]
dGG1f$^{+\bullet}$ → loss of (3H-G)$^{\bullet}$	176[f]	251[f]	243[f]	233[f]	252[f]



G 9H-G$^{\bullet}$ 3H-G$^{\bullet}$

[a]In kJ mol^{-1}. [b]Including B3LYP/6-31+G(d,p) zero-point vibrational corrections. [c]Single-point calculations with the 6-311++G(2d,p) basis set. [d]Relative to **dGG1f**$^{+\bullet}$. [e]Formation of the 8,9-dihydroguanine radical (9H-G)$^{\bullet}$. [f]Formation of the 3,8-dihydroguanine radical (3H-G)$^{\bullet}$.

TD-DFT calculations of excitation energies and oscillator strength in **dGG1a**$^{+\bullet}$**–dGG1f**$^{+\bullet}$ provided the absorption spectra (Figure S6) to be compared to the experimental action spectrum (Figure 4.3a). The calculated spectra of the low-energy isomers **dGG1c**$^{+\bullet}$**, dGG1d**$^{+\bullet}$, and **dGG1g**$^{+\bullet}$ having the 7,8-dihydroguanine cation radical chromophore (Figure S6) did not display the prominent 450−470 nm band shown in the action spectrum while having a band at 380−390 nm that was absent in the action spectrum. In contrast, the calculated spectra of isomers **dGG1e**$^{+\bullet}$ and **dGG1f**$^{+\bullet}$ were compatible with the action spectrum, showing bands at 240, 280−300, and 433−438 nm (Figure 4.3b,c). It should be noted that the action spectrum was obtained at the ion trap temperature (310 K), whereas the TD-DFT calculated bands correspond to vertical excitation in 0 K structures. Vibronic broadening due to electronic excitations from vibrationally excited states

present at 310 K is expected to both broaden the bands and red-shift the absorption maxima by 10−30 nm, [75] as observed previously for nucleobase31,32 and dinucleotide cation radicals.34

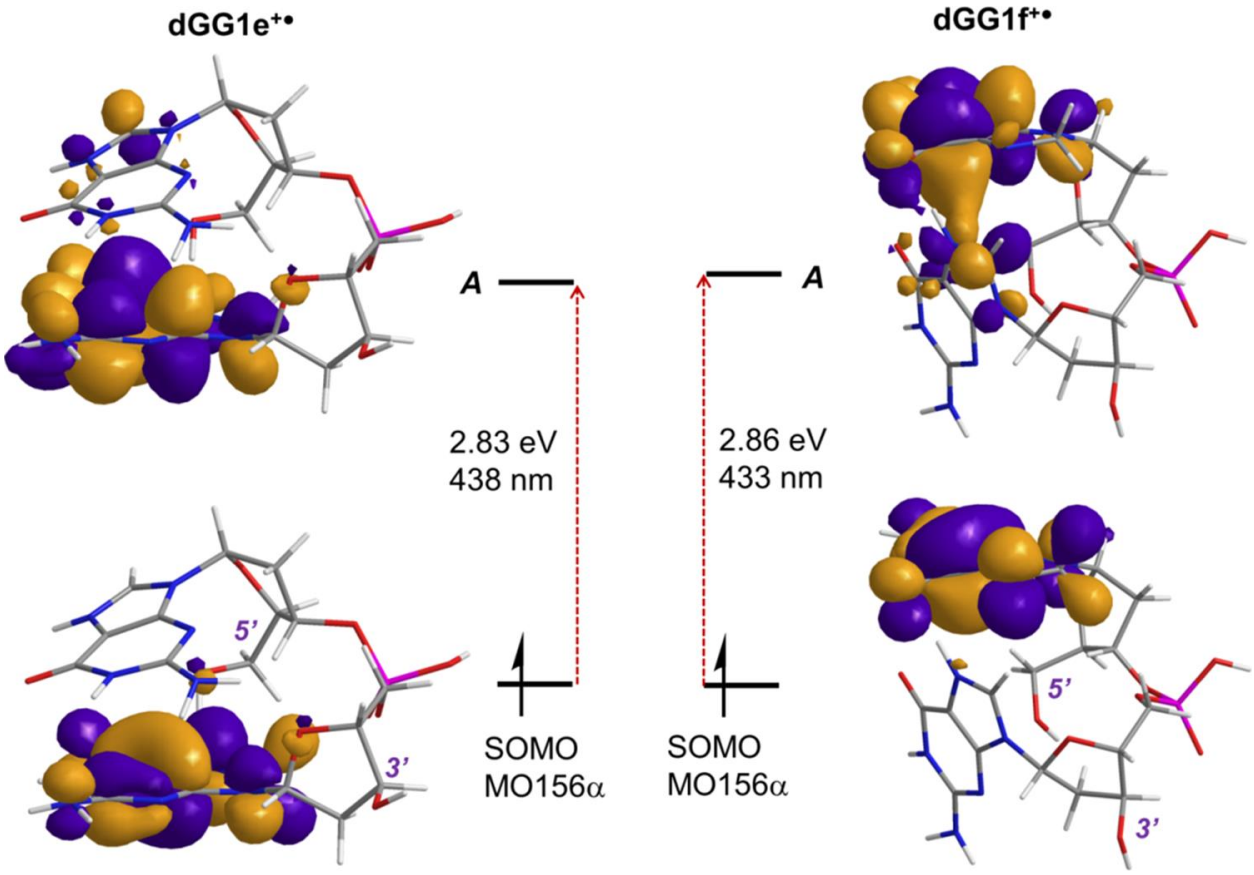
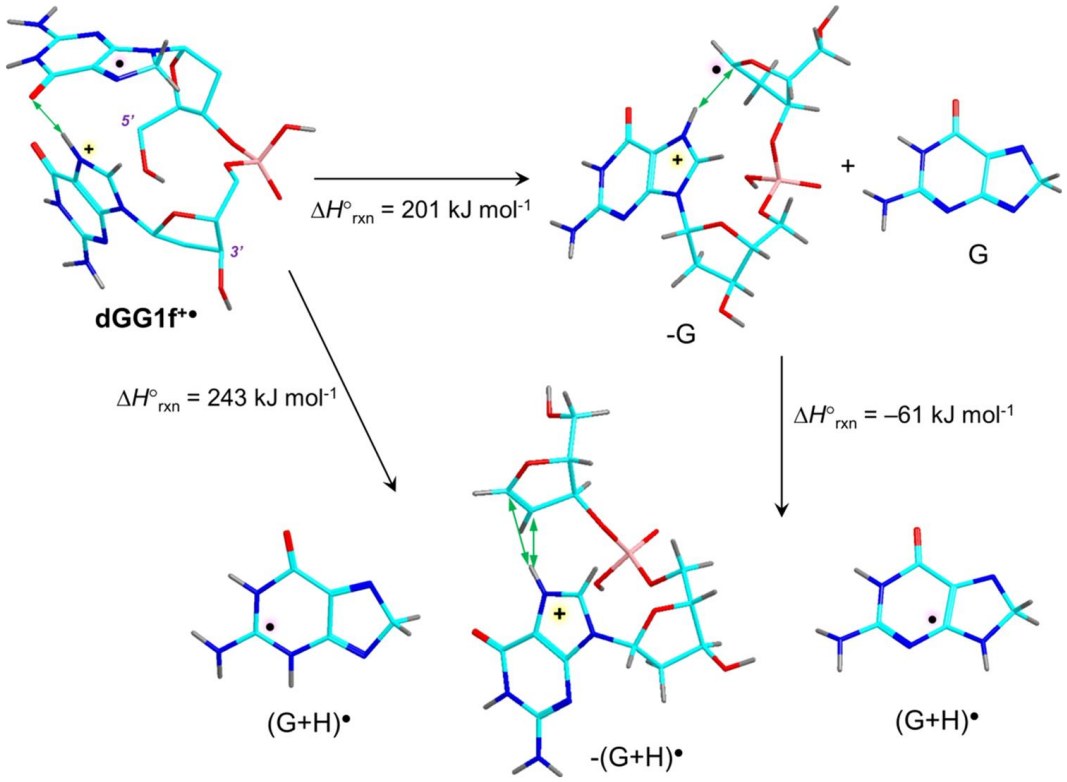


Figure 4.5. Molecular orbitals for the first excited state of (dGG + 2H)$^{+\bullet}$.

The nature of the diagnostic transition giving rise to the 450−460 nm band was elucidated by analyzing the molecular orbitals involved in the electron excitations in **dGG1e**$^{+\bullet}$ and **dGG1f**$^{+\bullet}$ (Figure 4.5). In both these radicals, the pertinent transitions corresponded to the first (A) excited state. The transitions in **dGG1e**$^{+\bullet}$ occurred from the singly occupied π-molecular orbital (SOMO, 156α) to a combination of π^* orbitals, all located at the 3′-guanine radical. Charge transfer to the π-electron system of the 5′-guanine cation was very minor. The transitions in **dGG1f**$^{+\bullet}$ occurred from the π-SOMO located at the 5′-radical to a combination of π^* orbitals chiefly

located at the same ring. In addition, the A excited state in **dGG1f**$^{+\bullet}$ involved a small portion of electron transfer to the protonated imidazole ring of the 3′-guanine cation (Figure 4.5). The assignment of the excited states in $(GG + 2H)^{+\bullet}$ was further corroborated by TD-DFT calculations of a model 7,8-dihydroguanosine radical, $(G + H)^{\bullet}$, having an analogous π-electron system (Figure S7, Supporting Information). In particular, the TD-DFT absorption spectrum of $(G + H)^{\bullet}$ displayed a prominent band at 439 nm, which was analogous to the 438 and 433 nm bands of **dGG1e**$^{+\bullet}$ and **dGG1f**$^{+\bullet}$, respectively. With $(G + H)^{\bullet}$, we were able to obtain a TD-DFT spectrum that was broadened by including 600 vibronic transitions at 310 K to estimate the thermal shift and broadening of the nucleobase bands. In particular, the 439 nm band corresponding to excitation to the first excited state exhibited a substantial red shift and broadening (Figure S7).

In addition to N-7-diprotonated guanine tautomers, we also considered cation radicals originating from the N-7, N-3-protonated O-6 enol isomers **dGG2**$^{2+}$ and **dGG3**$^{2+}$. However, none of these cation radicals (**dGG1b**$^{+\bullet}$**–dGG1f**$^{+\bullet}$, **dGG2a**$^{+\bullet}$, **dGG2b**$^{+\bullet}$, **dGG3a**$^{+\bullet}$, **dGG3b**$^{+\bullet}$ **dGG4a**$^{+\bullet}$, and **dGG4b**$^{+}$) provided a match with the experimental action spectrum (for spectra and discussion, see Figures S6, S8, and S9 in the Supporting Information). Hence, we concluded that these low-energy cation radicals were not produced by electron transfer in the gas phase.



Scheme 4.3. Dissociation of dGG1f$^{+\bullet}$ by loss of G and (G+H)$^{\bullet}$ [a].

[a]Reaction step enthalpies, $\Delta H°_{rxn}$, kJ mol^{-1}, are from M06-2X/6-311++G(2d,p) calculations and refer to 0 K.

To summarize the structure assignment of (GG + 2H)$^{+\bullet}$ ions, the 7-H-guanine protomers in the cation radicals initially produced by ETD were determined by the protonation sites in the lowest-energy crown−ether complex **dGG1a**$^{2+}$ that was expected to dominate in both solution and gas phase. Favored protonation at both N-7 positions in **dGG1a**$^{2+}$ was consistent, with N-7 being the most basic site in guanine,72 and gas-phase 2′-deoxyguanosine.74 Charge reduction resulted in a major change of structure. Loss of Coulomb repulsion triggered exothermic nucleobase stacking that was followed by combined N-7-proton transfers yielding 8,8-H guanine radicals that showed a signature absorption band in the action spectrum. The position of the guanine radical at

the 3′- or 5′-terminus could not be unequivocally determined from the action spectrum alone. However, energy considerations favored isomer **dGG1e**$^{+\bullet}$ with a 3′-guanine radical.

To gauge the range of photon energies capable of causing single-photon photodissociation of $(dGG + 2H)^{+\bullet}$ cation radicals, we calculated the dissociation energies for loss of guanine and $(G + H)^{\bullet}$ radical from **dGG1f**$^{+\bullet}$. The departing guanine molecule was considered as a less stable C-8−H2 tautomer, which originated by direct cleavage of the 5′-guanine radical moiety in **dGG1f**$^{+\bullet}$. Likewise, the $(G + H)^{\bullet}$ radical was taken as a C-8−H2 tautomer (Scheme 4.3). The loss of guanine had a threshold energy of 187−205 kJ mol^{-1} at different levels of theory (Table 4.1). The loss of $(G + H)^{\bullet}$ was found to be kinetically competitive and can be presumed to have a similar threshold energy. In our energy analysis, we considered the molar enthalpy of **dGG1f**$^{+\bullet}$ at 310 K (106 kJ mol^{-1}) and an estimated kinetic shift of 120 kJ mol^{-1} for the dissociation to occur on the 50 ms ion trap time scale. Hence, excitation of approximately 200 + 120 − 105 = 215 kJ mol^{-1} should be sufficient to drive dissociation. Such energy can be supplied by a single photon of $\lambda < 550$ nm. This indicated that the bands observed in the action spectrum of $(dGG + 2H)^{+\bullet}$ (Figure 4.3a) could be produced by single photon absorption.

4.3.4 *$(dCG + 2H)^{+\bullet}$ Action Spectroscopy and Ion Structures*

The action spectrum of $(dCG + 2H)^{+\bullet}$ showed a weak band at 450−480 nm and a stronger one at 350 nm that were chiefly composed of contributions from the sequence d1 fragment ions (*m/z* 348 and 349, Figure 4.6a). Another strong band at 280 nm had a major contribution from the *m/z* 446 fragment ions formed by loss of the $(C + H)^{\bullet}$ radical. To identify the $(dCG + 2H)^{+\bullet}$ isomers with matching absorption spectra, we first carried out extensive analysis of the $(dCG + crown + 2H)^{2+}$ precursor ions with the goal of sorting out by energy the guanine and cytosine ion tautomers as well as isomers differing in the position of the crown ligand. In addition to the above-mentioned

guanine protonation tautomers, cytosine protonation was considered at the O-2 and N-3 positions which are the most and nearly equally basic sites in cytosine. [72, 76, 77] BOMD trajectories followed by full optimization with ωB97X-D/6-31+G(d,p) yielded low-energy gas-phase and solvated structures for several combinations of protomers and crown−ether complexes (Figure S10) (Table S2, Supporting Information). The energy data indicated a favorable formation of N-3-protonated cytosine at the 5′-terminus, which coordinated the crown ether, while the 3′-guanine was protonated at N-7, as in **dCG1a**$^{2+}$ (Figure 4.4) and its conformers **dCG1b**$^{2+}$ and **dCG1c**$^{2+}$. Of the other protomeric and positional combinations (**dCG2**$^{2+}$**–dCG5**$^{2+}$, Figure S10), the only one that was energetically competitive was the positional isomer **dCG2**$^{2+}$ having the same nucleobase protonation pattern as did **dCG1a**$^{2+}$ but with the crown−ether coordination switched to the 3′-guanine ion.

Electron attachment to **dCG1a**$^{2+}$ followed by crown−ether loss produced a cation radical (**dCG1**$^{+\bullet}$) with the unpaired electron occupying the cytosine base (Figure 4.6b). An alternative structure having the unpaired electron in the guanine ring (**dCG2**$^{+\bullet}$, Figure S11a) was a high-energy species corresponding to an excited electronic state of **dCG1**$^{+\bullet}$. For the calculated relative energies of $(CG + 2H)^{+\bullet}$ ions, see Table 4.2. Other alternative structures were derived from **dCG1**$^{+\bullet}$ by proton or hydrogen atom migrations (Scheme 4.4).

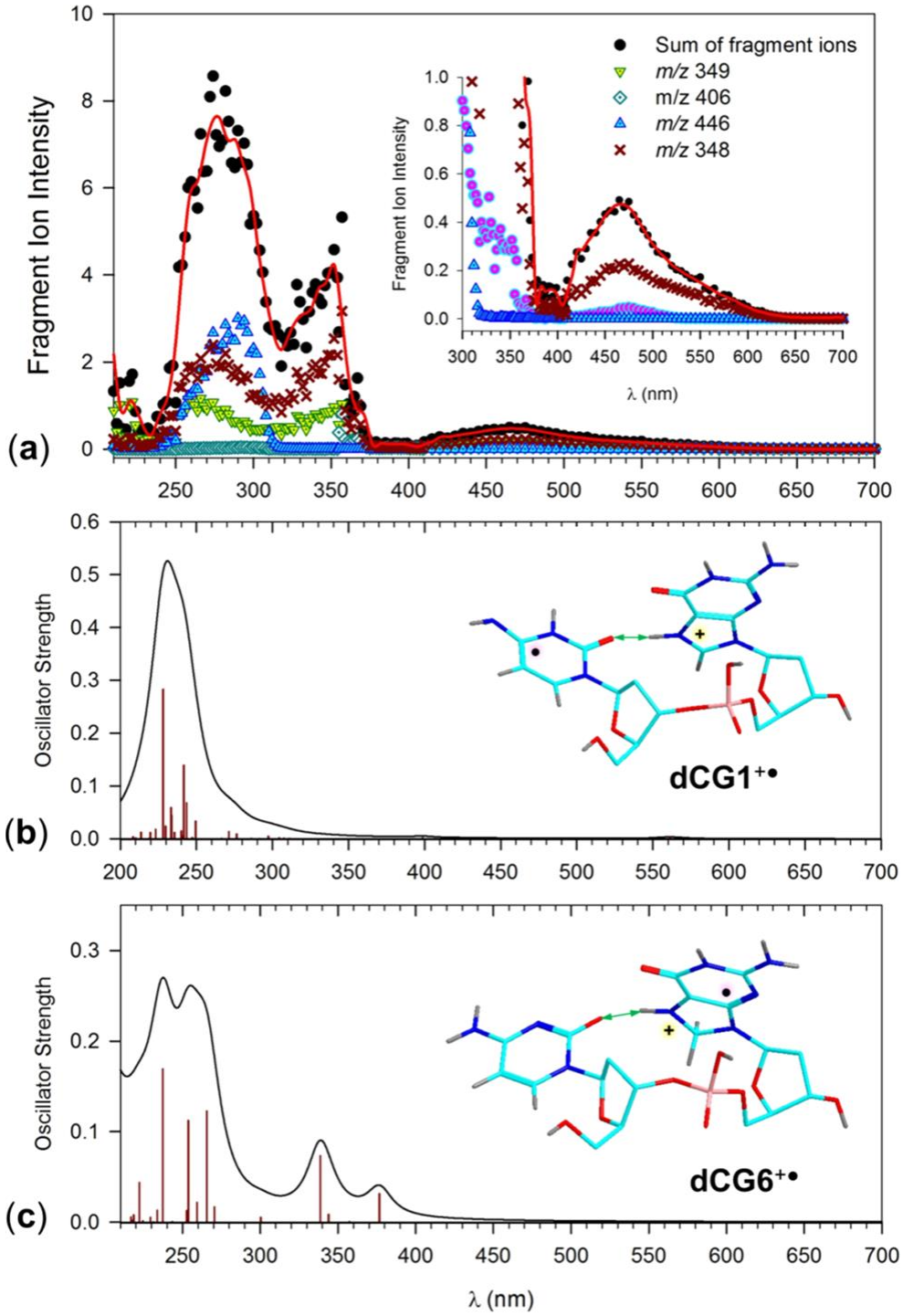
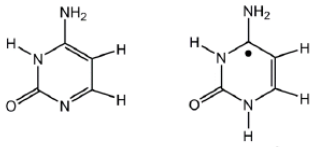


Figure 4.6. (a) Action spectrum of $(dGG+2H)^{+\bullet}$. The inset shows an expanded 300−700 nm region. M06-2X/6-31+G(d,p) TD-DFT absorption spectra of (b) $dCG1^{+\bullet}$ and(c) $dCG6^{+\bullet}$. The bars represent the calculated absorption wavelengths and oscillator strength. The band shapes are from artificial line broadening with Lorenztian functions at 12 nm full width at half-maximum.

This produced an O, N-7-dihydroguanine cation radical (**dCG3**$^{+\bullet}$, Figure S11b), O, N-3-dihydrocytosine cation radical (**dCG4**$^{+\bullet}$, Figure S11c), a guanine protonation isomer **dCG5**$^{+\bullet}$ (Figure S11d), and a 7,8-dihydroguanine cation radical (**dCG6**$^{+\bullet}$, Figure 4.6c). With the exception of the last structure, all of these isomers were less stable than **dCG1**$^{+\bullet}$, so there was no thermodynamic drive for isomerizations forming them. Isomerization of **dCG1**$^{+\bullet}$ to **dCG6**$^{+\bullet}$ was exothermic and involved a transition state (**dCG-TS1**) that was at 98−106 kJ mol^{-1} relative to **dCG1**$^{+\bullet}$ according to different levels of theory (Table 4.2).

Table 4.2. Relative energies of (dCG + 2H)$^{+\bullet}$ ions

	Relative Energy[a,b]				
Ion	B3LYP[c]	B97X-D[c]	M06-2X[c]	PMP2[c]	ROMP2[c]
dCG1$^{+\bullet}$	0	0	0	0	0
dCG2$^{+\bullet}$	-[d]	127	123	-	-
dCG3$^{+\bullet}$	26	30	18	31	17
dCG4$^{+\bullet}$	35	26	11	28	27
dCG5$^{+\bullet}$	66	65	55	-	-
dCG6$^{+\bullet}$	–31	–23	–23	–9	–9
dCG7$^{+\bullet}$	19	23	11	34	42
dCG-TS1	131	103	106	98	101
dCG-TS2	92	54	58	70	76
dCG-TS3	186	194	189	180	171
loss of isoC	109	148	145	165	171
loss of (C+H)$^{\bullet}$	97	141	137	159	161



isoC (C+H)$^{\bullet}$

[a] In kJ mol^{-1}. [b] Including B3LYP/6-31+G(d,p) zero-point vibrational corrections. [c] Single-point calculations with the 6-311++G(2d,p) basis set. [d] Collapsed to dCG1 upon B3LYP/6-31+G(d,p) gradient geometry optimization.

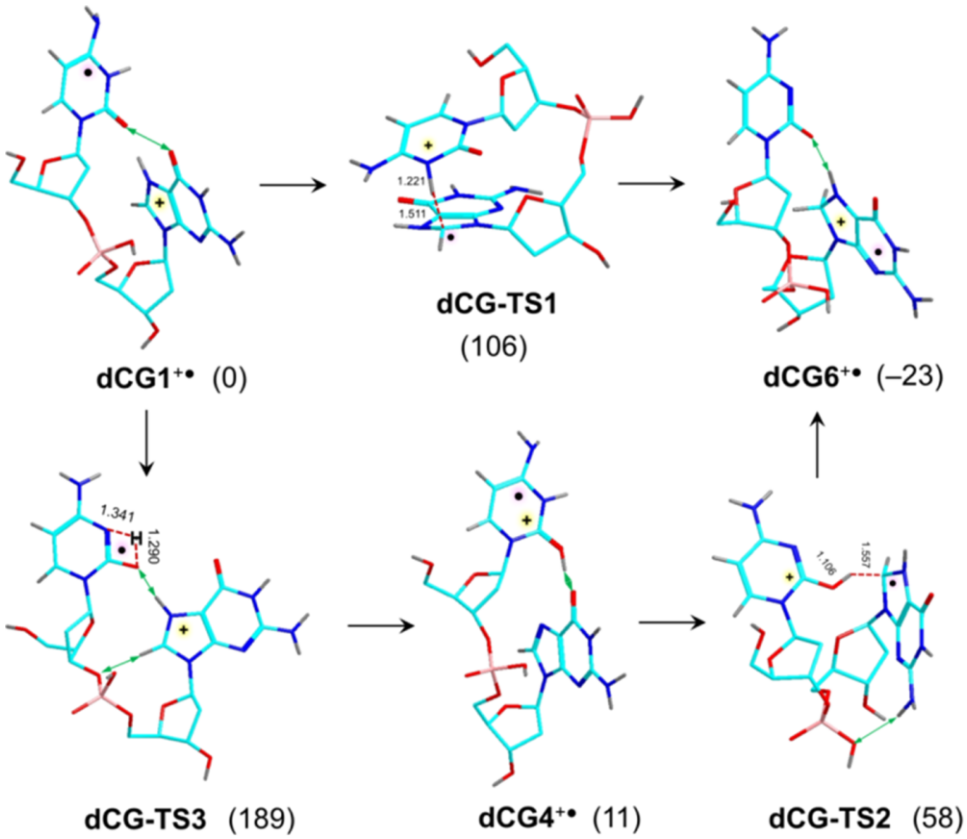
Comparing the action spectrum of (CG + 2H)$^{+\bullet}$ with the calculated absorption spectra of dCG1$^{+\bullet}$–dCG6$^{+\bullet}$, we found no complete agreement for any single isomer (Figures 6b,c and S11a–d). However, the weak band at 450 nm in the spectrum of **dCG1**$^{+\bullet}$ (Figure 4.6b) and the

bands at 339 and 377 nm in the spectrum of the lowest-energy isomer **dCG6**$^{+\bullet}$ (Figure 4.6c) could account for the 470 and 350 nm diagnostic features in the action spectrum. Hence, we assigned the action spectrum to a mixture of these two isomers as the closest match. We note that electron transfer to the precursor dication exothermically formed **dCG1**$^{+\bullet}$ as a primary product, which was connected to **dCG6**$^{+\bullet}$ via **dCG-TS1**, thus allowing isomerization if kinetically feasible. To assess the isomerization kinetics, we carried out RRKM calculations that provided unimolecular rate constants (k, s^{-1}) as a function of internal energy (Figure S12a). The calculated rate constants for the rearrangement in **dCG1**$^{+\bullet}$ displayed a slow rise with the ion internal energy. According to the ωB97X-D-based calculations, 10% conversion on the experimental 50 ms time scale was achieved at 310 kJ mol^{-1} internal energy, whereas 50% conversion required 360 kJ mol^{-1} (Figure S12b). Considering the ion thermal energy (E_{therm} = 102 kJ mol^{-1} at 310 K), isomerization via dCG-TS1 would require >200 kJ mol^{-1} additional excitation in **dCG1**$^{+\bullet}$ acquired by ETD. The excitation in **dCG1**$^{+\bullet}$ (E_{exc}) was estimated as a fraction of the ETD exothermicity that was partitioned in a 0.57:0.43 ratio between **dCG1**$^{+\bullet}$ and the departing DBCE according to their heat capacities. ETD exothermicity (ΔE_{ETD}) was estimated on the basis of the calculated recombination energy of the precursor complex dication (RE = 520 kJ mol^{-1}), fluoranthene electron affinity (EA = 61 kJ mol^{-1}), [78] and the binding energy of DBCE in the reduced complex (E_{dis} = 130 kJ mol^{-1}) as ΔEETD = (RE − EA − E_{dis}) ≈ 330 kJ mol^{-1}, from which we obtained $E_{exc} = 0.57 \times \Delta E_{ETD}$ = 188 kJ mol^{-1}. Hence, the estimated internal energy of **dCG1**$^{+\bullet}$, $E_{int} = E_{exc} + E_{therm}$ = 188 + 102 = 290 kJ mol^{-1}, fell close to the kinetically relevant energy region (Figure S12b), allowing partial isomerization to proceed on the ion trap time scale. This analysis was consistent with the action spectrum of $(CG + 2H)^{+\bullet}$ that indicated the formation of a mixture of **dCG1**$^{+\bullet}$ and **dCG6**$^{+\bullet}$.

CID and UVPD of $(dCG + 2H)^{+\bullet}$ showed an abundant loss of cytosine and $(C + H)^{\bullet}$ radicals (Figure 4.2c). We estimated the pertinent dissociation threshold energies by calculations carried out for isomer **dCG1**$^{+\bullet}$ having a radical cytosine ring. The neutral cytosine molecule was taken as the less stable N-3−H tautomer (iso-C), whose structure correlates with that of the cytosine nucleobase in **dCG**$^{+\bullet}$. The presumed $(C + H)^{\bullet}$ radical product was the most stable N-1−H, N-3−H tautomer (Scheme S1 Supporting Information). The calculated dissociation energies indicated very similar thresholds for loss of iso-C and $(C + H)^{\bullet}$, E_{diss} = 145 and 137 kJ mol^{-1}, respectively (Table 4.2). Dissociations by loss of C and $(C + H)^{\bullet}$ were notably less endothermic than those by loss of G and $(G + H)^{\bullet}$ (vide supra). This was consistent with the CID and UVPD-MS^3 spectra of $(dCG + 2H)^{+\bullet}$, which showed a nearly exclusive loss of $(C + H)^{\bullet}$ and C (Figures 2 and S2, Supporting Information).

4.3.5 *$(dGC + 2H)^{+\bullet}$ Action Spectroscopy and Ion Structures*

The action spectrum of $(dGC + 2H)^{+\bullet}$ (Figure 4.7a) was different from the spectra of both $(dGG + 2H)^{+\bullet}$ and $(dCG + 2H)^{+\bullet}$. Extremely weak broad bands at 620 and 420−500, and 360 nm resulted from contributions of all major dissociation channels. In the short-wavelength region of the spectrum, there was a single strong band carried by all major channels with a maximum at 285 nm. This band had an asymmetric profile, possibly indicating an overlapping minor band with an absorption maximum at 250−260 nm. The 620 and 420−500 nm bands were likely associated with nucleobase radical chromophores that showed similar maxima albeit different intensities for $(dGG + 2H)^{+\bullet}$ and $(dCG + 2H)^{+\bullet}$. The 285 nm band was likely associated with excitations in the radical nucleobase rings, as well as those not affected by radical addition. To provide a more specific interpretation of the action spectra, we carried out calculations of ion structures and excitation energies.



Scheme 4.4. Isomerization of $(dCG + 2H)^{+\bullet}$ cation radicals[a]

[a]Relative energies (kJ mol^{-1}) are from M06-2X/6-311++G(2d,p) calculations including zero-point vibrational energies and and refer to 0 K.

Several $(dGC+CE + 2H)^{2+}$ structures were considered as dication precursors, which were protonated at the basic positions of guanine (N-7 and O-6, N-3) and cytosine (N-3 and O-2) and differed in the location of the crown−ether ligand. All of the dGC protomer types ($GC12^{+}$–$GC82^{+}$, Figure S13) favored crown−ether coordination on the protonated 3′-cytosine ring. BOMD conformational analysis followed by ωB97X-D geometry optimization yielded the lowest-energy gas-phase structures ($GC12^{+}$–$GC52^{+}$). These ions were cytosine N-3-protonated, whereas the guanine was either 6-oxo-N-7−H or 6-OH−N-3−H tautomer. Structure $GC12^{+}$ was the global energy minimum (Figure 4.4). The free-energy differences notably increased upon including solvation energy in the calculations, making $GC12^{+}$ the by far lowest-energy structure (Table S3).

Hence, according to the gas-phase and solvation energy data, GC12+ should be formed preferentially by electrospray.

Table 4.3. Relative energies of $(dGC + 2H)^{+\bullet}$ ions

	relative energy[a,b]				
ion	B3LYP[c]	ωB97X-D[c]	M06-2X[c]	PMP2[c]	ROMP2[c]
$dGC1^{+\bullet}$	0 (0)[d]	0 (0)	0 (0)	0 (0)	0 (0)
$dGC2^{+\bullet}$	22 (21)	6 (5)	7 (6)	16 (15)	17 (16)
$dGC3^{+\bullet}$	8 (6)	9 (7)	17 (15)	16 (14)	17 (15)
$dGC4^{+\bullet}$	13 (17)	12 (15)	13 (17)	11 (14)	12 (15)
$dGC5^{+\bullet}$	−5 (10)	2 (17)	14 (29)	−30 (−15)	−37 (−22)
$dGC6^{+\bullet}$	26 (39)	−1 (12)	11 (24)	−7 (6)	−14 (−1)
$dGC7^{+\bullet}$	−15 (−4)	−34 (−23)	−60 (−50)	−29 (−18)	−27 (−16)
dGC-TS	94	82	95	69	71

[a] In kJ mol^{-1}. [b] Including B3LYP/6-31+G(d,p) zero-point vibrational energy corrections. [c] Single-point calculations with the 6-311++G(2d,p) basis set. [d] Values in parentheses are relative free energies at 310 K.

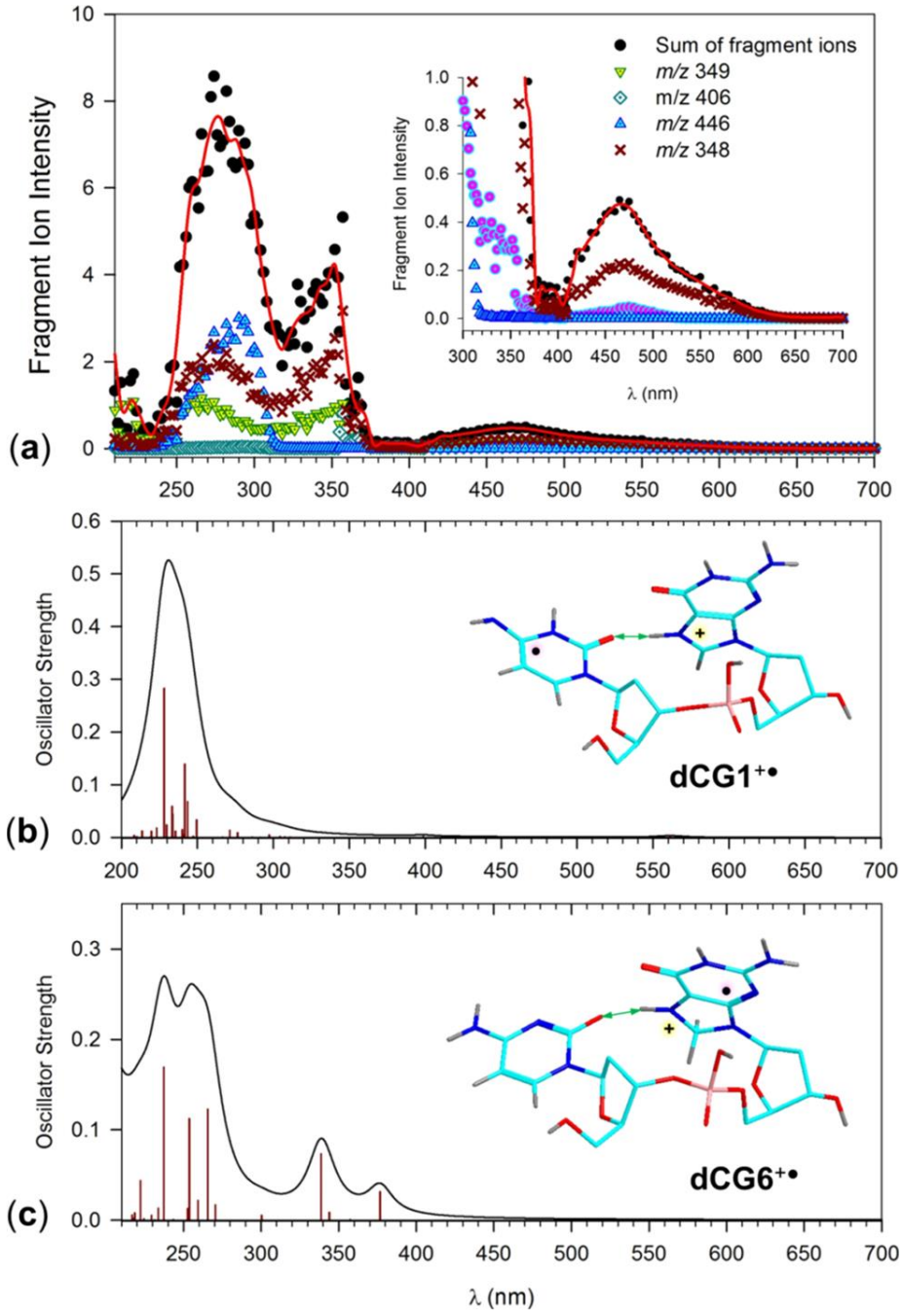
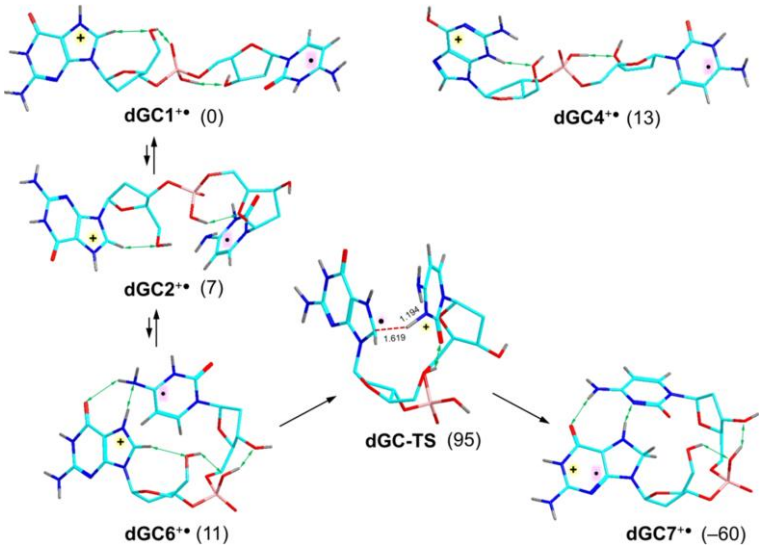


Figure 4.7. (a) Action spectrum of $(dGC+2H)^{+\bullet}$. The inset shows an expanded 300−700 nm region. M06-2X/6-31+G(d,p) TD-DFT absorption spectra of (b) $dGC1^{+\bullet}$ and(c) $dGC4^{+\bullet}$. The bars represent the calculated absorption wavelengths and oscillator strength. The band shapes are from artificial line broadening with Lorenztian functions at 12 nm full width at half-maximum.

Electron transfer to **dGC1**$^{2+}$ followed by crown−ether ligand loss produced a cytosine radical (**dGC1**$^{+\bullet}$) as a lowest-energy conformer of this structural type (Figure 4.7b). Ion **dGC1**$^{+\bullet}$ had an extended conformation that did not display nucleobase stacking. The TD-DFT absorption spectrum of **dGC1**$^{+\bullet}$ displayed strong bands in the 235−260 nm region, a weak band at 300 nm and very weak bands at 420 and 515 nm. The overall absorption pattern of **dGC1**$^{+\bullet}$ agreed with that observed in the action spectrum, when allowing for wavelength shifts and band broadening because of vibronic effects. A similar absorption spectrum (Figure 4.7c) was calculated for the guanine O-6, N-3−H tautomer **dGC4**$^{+\bullet}$ which was another cytosine radical with an extended conformation. Isomer **dGC4**$^{+\bullet}$ was marginally (14−17 kJ mol^{-1}) less stable than **dGC1**$^{+\bullet}$ according to all our calculations (Table 4.3). For the absorption spectra of the other low-energy isomers **dGC2**$^{+\bullet}$**, dGC3**$^{+\bullet}$**,** and **dGC5**$^{+\bullet}$, see the respective Figure S14a−c in the Supporting Information. Conformational collapse in **dGC1**$^{+\bullet}$ resulting in nucleobase stacking formed isomer **dGC6**$^{+\bullet}$ that was a cytosine radical comparably stable as **dGC1**$^{+\bullet}$. The absorption spectrum of **dGC6**$^{+\bullet}$ (Figure S14d, Supporting Information) was similar to the spectra of **dGC1**$^{+\bullet}$ and **dGC4**$^{+\bullet}$ with the exception of weaker bands in the 230−250 nm region. The lowest-energy (dGC + 2H)$^{+\bullet}$ isomer was obtained by a hydrogen atom transfer from the cytosine N-3 position to C-8 at guanine, yielding the 7,8-dihydroguanine cation radical **dGC7**$^{+\bullet}$ (Scheme 4.4). However, the calculated absorption spectrum of **dGC7**$^{+\bullet}$ (Figure S14e) showed bands at 336 and 376 nm that did not have equivalents in the action spectrum of (dGC+ 2H)$^{+\bullet}$. The spectroscopic data led us to conclude that the (dGC + 2H)$^{+\bullet}$ formed by ETD were cytosine radicals. A more specific assignment of the cation radical conformation and the protomeric form of the charged guanine ring relied on the energy data alone. The precursor dications preferred guanine protonation at N-7, disfavoring the formation of the protomeric isomers **GC4**$^{2+}$**−GC6**$^{2+}$ that would have formed cation radical **dGC4**$^{+\bullet}$ upon

electron transfer. Hence, based on both relative energies and action spectra, ion **dGC1**$^{+\bullet}$ and its conformers appear to be the best match for the $(GC + 2H)^{+\bullet}$ cation radicals formed by electron transfer.



Scheme 4.5. Isomerization of $(dGC + 2H)^{+\bullet}$ cation radicals[a]

[a]Relative energies (kJ mol^{-1}) are from M06-2X/6-311++G(2d,p) calculations including zero-point vibrational energies and and refer to 0 K.

4.3.6 *Discussion*

The above-described structural analysis raised several points regarding the formation and properties of the hydrogen-rich dinucleotide cation radicals. The site of electron attachment to the dications was found to depend on both the nucleobase nature and position. Electron attachment to the cytosine−guanine combinations $(CG + 2H)^{2+}$ and $(GC + 2H)^{2+}$ favored the formation of cytosine radicals, e.g., **dCG1**$^{+\bullet}$ and **dGC1**$^{+\bullet}$. Low-energy guanine-centered radicals such as **dCG6**$^{+\bullet}$ and **dGC7**$^{+\bullet}$ were formed as secondary products of hydrogen atom or proton migration targeting the guanine C-8 position. The preferential electron attachment to protonated cytosine can be related to its higher adiabatic recombination energy, RE_{adiab} = 5.17 eV for the N-3−H

tautomer,76 compared to that of the N-7, N-9−H tautomer of protonated guanine, RE_{adiab} = 4.46 eV. Both these RE_{adiab} values were calculated on the same level of *ab initio* theory (CCSD(T)/6-311++G(3df,2p); Table S4, Supporting Information). In the case of nearly identical nucleobases such as the N-7-protonated guanines in $(GG + 2H)^{2+}$, electron attachment favored the 5′-ring (**dGG1a**$^{+\bullet}$). This outcome was probably caused by the stabilization of the 3′-cation by a hydrogen bond between the guanine H-8 and phosphodiester O=P oxygen, which was facilitated by their syn arrangement on the 3′-deoxyribose ring. We note that a similar dipolar interaction has been reported by Rodgers et al. for singly protonated guanosine and 2′-deoxyguanosine cations.74 In contrast, the anti-arrangement of the 5′-guanine cation and the P=O bond prevents the formation of a C-8−H···O=P hydrogen bond.

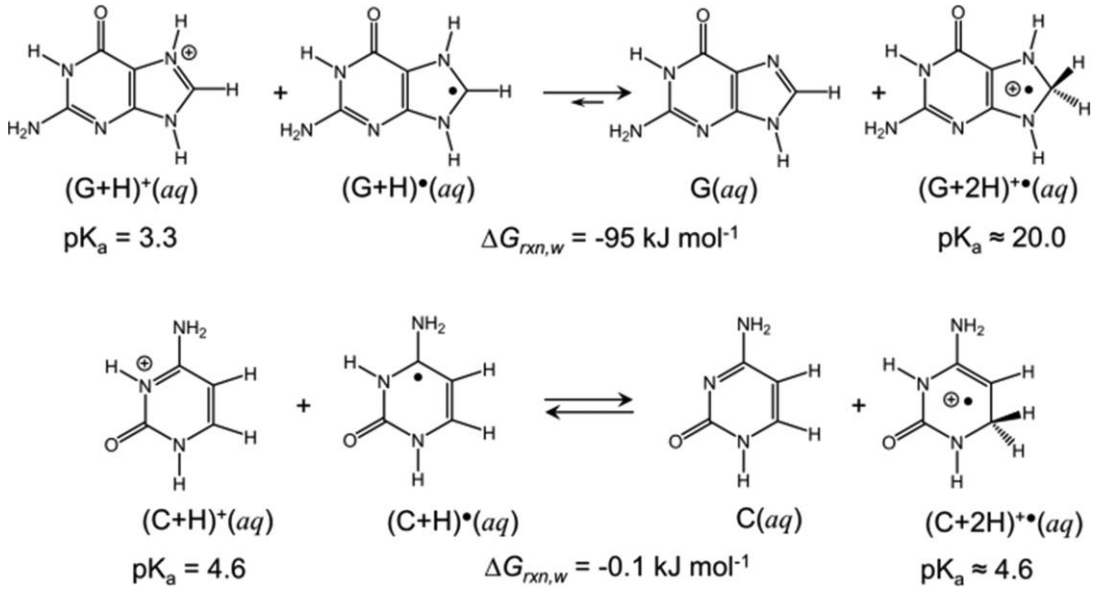
Another feature distinguishing the guanine and cytosine cation radicals was their different tendencies to undergo ring stacking and isomerize by proton transfer. With **dGG1**$^{+\bullet}$ that was initially formed by one-electron reduction of the dication, stacking to **dGG1b**$^{+\bullet}$ was 98 kJ mol^{-1} exothermic, which was greater than the stacking energy calculated by Hobza and Šponer for a neutral guanine−guanine pair (54 kJ mol^{-1}). [79] The nucleobase stacking in **dGG1b**$^{+\bullet}$ was constrained by the deoxyribose and phosphate backbone to reach a 60° angle between the rings. The large stabilization of **dGG1b**$^{+\bullet}$ can be ascribed to the formation of multiple hydrogen bonds involving N-7−H and O-6 of the 5′-guanine and N-1−H, NH_2, and O-6 of the 3′-guanine (Scheme 4.2). Interestingly, the 5′-hydroxyl developed a H-bond to the C-8 radical, which was analogous to H-bonds in alkyl radicals. [80] The angle between the rings and guanine hydrogen bonding was largely preserved in **dGG-TS**, involving N-7 proton transfer from the 3′-guanine ion to the C-8 position on the 5′-guanine radical, which was catalytically relayed by the 5′-hydroxyl, resulting in an extremely low energy transition state (Scheme 4.2). Ring stacking was further enhanced in 3′-

radicals **dGG1d**$^{+\bullet}$ and **dGG1e**$^{+\bullet}$, in which the guanine rings were positioned at a 30° angle. However, another low-energy 5′-radical**, dGG1f**$^{+\bullet}$, had rings close to 90° and the stabilizing inter-ring interaction was chiefly due to a strong hydrogen bond between the 5′-guanine O-6 and 3′-guanine N-7−H at 1.66 Å (Scheme 4.2).

Stacking interactions between the nucleobases were absent in **dCG**$^{+\bullet}$ and **dGC**$^{+\bullet}$. Both the 5′-cytosine radical **dCG1**$^{+\bullet}$ and its isomerized 3′-guanine radical product **dCG6**$^{+\bullet}$ showed a nearly perpendicular arrangement of the nucleobase rings (Figure 4.6b,c). The attractive inter-ring interaction in **dCG1**$^{+\bullet}$ was chiefly due to a single strong H-bond (at d(O···H) = 1.50 Å) between O-2 of the 5′-cytosine radical and N-7−H of the guanine ion. This H-bonding interaction (at d(O···H) = 1.61 Å) was preserved in **dCG6**$^{+\bullet}$. Regarding the isomerization of **dCG1**$^{+\bullet}$, spin density analysis of the reactant (**dCG1**$^{+\bullet}$) and the transition state (**dCG-TS1**) indicated a proton-coupled electron-transfer process (Scheme 4.4). According to the natural population analysis of **dCG1**$^{+\bullet}$ and **dCG-TS1**, the unpaired electron density that was initially localized in the cytosine ring in **dCG1**$^{+\bullet}$ moved to guanine in **dCG-TS1**, which was followed by a transfer of the cytosine N-3 proton. The transition-state energy for **dCG-TS1** (106 kJ mol^{-1}) was notably higher than that for **dGG-TS** (9 kJ mol^{-1}, Scheme 4.2). We identified another isomerization pathway consisting of an endothermic migration of the guanine N-7−H to the cytosine 2-O (**dCG4**$^{+\bullet}$), followed by reverse proton transfer to guanine C-8 (**dCG-TS2**, Scheme 4.4). This can be viewed as a cytosineradical-catalyzed migration of guanine N-7−H to the stabilized C-8 position. This pathway may avoid a direct 1,2-H migration of guanine N-7−H, which was presumed to require a much higher energy in the pertinent transition state. However, the formation of **dCG4**$^{+\bullet}$ was unlikely to proceed without internal catalysis because the transition state for a direct **dCG1**$^{+\bullet}$ **→dCG4**$^{+\bullet}$ isomerization via a

1,3-hydrogen shift was calculated to require a high-energy transition state (**dCG-TS3**, Scheme 4.4).

Hydrogen bonding between the nucleobases was not realized in $(dGC + 2H)^{+\bullet}$ isomers **dGC1**$^{+\bullet}$ and **dGC4**$^{+\bullet}$, which were identified as the major products of one-electron reduction. Conformational folding in **dGC1**$^{+\bullet}$, resulting in nucleobase cation-radical interactions, did not produce a lower-energy conformer. The folded structures **dGC5**$^{+\bullet}$ and **dGC6**$^{+\bullet}$ were disfavored against **dGC1**$^{+\bullet}$ by entropy and were not likely to dominate under thermal equilibrium conditions. A large negative change of entropy (-34 J mol^{-1} K^{-1}) was also associated with attaining the transition state for proton migration, **dGC1**$^{+\bullet}$ → **dGC-TS**, which may explain slow isomerization to **dGC7**$^{+\bullet}$ and retention of **dGC1**$^{+\bullet}$ conformers after ETD.



Scheme 4.6. Photon Transfer in $(G+H)^{\bullet}$ and$(G+H)^{\bullet}$

The facile hydrogen transfer between the guanine nucleobases in gas-phase $(dGG+ 2H)^{+\bullet}$ raised the question of proton or hydrogen atom migration in DNA cation radicals in solution, which was relevant to chemical changes in DNA following ionization. Hydrogen-rich DNA radicals are

formed by protonation of anion radicals following fast capture of a low-energy electron.40 We addressed this topic with model lowest-energy guanine and cytosine radicals (Scheme 4.5, Scheme 4.6and Table S4, Supporting Information) and their respective 9-methylguanine and 1-methylcytosine homologues (Table S5) to account for the effect of the pertinent deoxyribose substituents. According to our DFT calculations, proton transfer from $(G + H)^+(aq)$ to $(G + H)^{\bullet}(aq)$ was highly exothermic, $\Delta G_{rxn,w} = -95$ kJ mol^{-1} (average value from Table S4, Supporting Information), producing G(aq) and the hydrogen-rich cation radical $(G + 2H)^{+\bullet}(aq)$. Combining ΔGrxn,w with the pK_a of guanine ($pK_a = 3.3$), [81] [82] one obtains pK_a for $(G + 2H)^{+\bullet}$ ranging from 19.8 to 20.2. This indicates that $(G + H)^{\bullet}$ should be a superbase in water and react rapidly with solvent, forming $(G + 2H)^{+\bullet}$ over a range of pH values. In contrast, DFT calculations of the hydrogen-rich cytosine cation radical, $(C + 2H)^{+\bullet}(aq)$, indicated facile deprotonation, yielding an average $pK_a = 4.6$ when adjusted to the pK_a of guanine (Table S4, Supporting Information). Thus, protonation of $(C + H)^{\bullet}$ by water can be expected to be sensitive to the solution pH and cytosine environment in DNA (Scheme 4.6).[83] The effect of the deoxyribose substituent in 2′-deoxyguanosine and 2′-deoxycytidine on pK_a was modeled by the pertinent methyl-substituted nucleobases. Averaged computational data from Table S5 (Supporting Information) indicated a slight increase of basicity, as expressed by the pK_a's of 20.8 and 5.4 for 9-methylguanosine and 1-methylcytosine, respectively. We note that various aspects of proton transfer involving guanine−cytosine base pairs in the cation-radical[15,16,19] or anion-radical forms[22,23] have been studied and discussed.[20,21] However, to the best of our knowledge, pK_a values of neutral nucleobase radicals have not been determined, and thus our reported values provide insight into the protonation energetics of these transient species.

4.4 CONCLUSIONS

The experimental and computational data reported here allowed us to arrive at the following conclusions. DNA dinucleotide cation radicals of a new hydrogen-rich type were prepared in high yield and purity by electron-transfer dissociation of dication−crown−ether complexes. Crown−ether complexation in the dications depended on the nucleobase and its 3′or 5′ position and stabilized dinucleotide protomers with N-7-protonated guanine and N-3-protonated cytosine.

Photodissociation action spectroscopy provided spectra for $(dGG + 2H)^{+\bullet}$, $(dCG + 2H)^{+\bullet}$, and $(dGC + 2H)^{+\bullet}$ that were interpreted by TD-DFT calculations and assigned to nucleobase radical structural types formed directly by electron attachment or after rearrangement. Radical−cation nucleobase interactions in $(dGG + 2H)^{+\bullet}$ resulted in exothermic conformational collapse that commenced by nucleobase stacking and was followed by hydrogen migration between the nucleobases, forming stable guanine-8,8-H_2 radicals. Contrasting this, conformational folding in $(dCG + 2H)^{+\bullet}$ and $(dGC + 2H)^{+\bullet}$ was slightly endoergic and did not involve nucleobase stacking. Hydrogen transfer from cytosine to guanine, albeit exothermic, was kinetically hampered.

4.5 ACKNOWLEDGEMENTS

Financial support from the Chemistry Division of the National Science Foundation (Grants CHE-1661815 for experiments and CHE-1624430 for calculations) is gratefully acknowledged. F.T. acknowledges the Klaus and Mary Ann Saegebarth Endowment for general support.

4.6 REFERENCES

1. Khanna, K. K.; Jackson, S. P., *Nature Genetics* **2001,** *27* (3), 247-254.
2. Wagenknecht, H. A., *Charge Transfer in DNA*. Wiley-VCH: Weinheim, Germany, 2005; p 1-23.
3. Schuster, G. B., *Accounts of Chemical Research* **2000,** *33* (4), 253-260.
4. O'Neill, M. A. B., J. K., Sequence-Dependent DNA Dynamics: The Regulator of DNA-Mediated Charge Transport. In *Charge Transfer in DNA*, Wagenknecht, H. A., Ed. Wiley-VCH: Weinheim, Germany, 2005; pp 27-31.
5. Giese, B., *Accounts of Chemical Research* **2000,** *33* (9), 631-636.
6. Giese, B.; Amaudrut, J.; Kohler, A. K.; Spormann, M.; Wessely, S., *Nature* **2001,** *412* (6844), 318-320.
7. Douki, D.; Ravanat, J. L.; Angelov, D.; Wagner, J. R.; Cadet, J., Effects of duplex stability on charge-transfer efficiency within DNA. In *Long-Range Charge Transfer in DNA I*, 2004; Vol. 236, pp 1-25.
8. Joy, A.; Ghosh, A. K.; Schuster, G. B., *Journal of the American Chemical Society* **2006,** *128* (16), 5346-5347.
9. Ghosh, A.; Joy, A.; Schuster, G. B.; Douki, T.; Cadet, J., *Organic & Biomolecular Chemistry* **2008,** *6* (5), 916-928.
10. Joseph, J.; Schuster, G. B., *Journal of the American Chemical Society* **2009,** *131* (39), 13904-+.
11. Kanvah, S.; Schuster, G. B., *Organic & Biomolecular Chemistry* **2010,** *8* (6), 1340-1343.
12. Kanvah, S.; Joseph, J.; Schuster, G. B.; Barnett, R. N.; Cleveland, C. L.; Landman, U., *Accounts of Chemical Research* **2010,** *43* (2), 280-287.
13. Kravec, S. M.; Kinz-Thompson, C. D.; Conwell, E. M., *Journal of Physical Chemistry B* **2011,** *115* (19), 6166-6171.
14. Barnett, R. N.; Joseph, J.; Landman, U.; Schuster, G. B., *Journal of the American Chemical Society* **2013,** *135* (10), 3904-3914.
15. Cauet, E.; Lievin, J., Radical cations of the nucleic bases and radiation damage to DNA: Ab initio study. In *Advances in Quantum Chemistry, Vol 52*, 2007; Vol. 52, pp 121-147.
16. Rodriguez-Santiago, L. N., 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: Weinheim, Germany, 2010; pp 219-242.
17. Li, X. F.; Cai, Z. L.; Sevilla, M. D., *Journal of Physical Chemistry B* **2001,** *105* (41), 10115-10123.
18. Wang, M.; Zhao, J.; Bu, Y., *Physical Chemistry Chemical Physics* **2013,** *15* (42), 18453-18463.
19. Choi, J.; Yang, C.; Fujitsuka, M.; Tojo, S.; Ihee, H.; Majima, T., *Journal of Physical Chemistry Letters* **2015,** *6* (24), 5045-5050.
20. Sevilla, M. D.; Kumar, A.; Adhikary, A., *Journal of Physical Chemistry B* **2016,** *120* (11), 2984-2986.
21. Choi, J.; Yang, C.; Fujitsuka, M.; Tojo, S.; Ihee, H.; Majima, T., *Journal of Physical Chemistry B* **2016,** *120* (11), 2987-2989.
22. Lin, Y.; Wang, H.; Gao, S.; Li, R.; Schaefer, H. F., III, *Journal of Physical Chemistry B* **2012,** *116* (30), 8908-8915.
23. Hsu, S. C. N.; Wang, T.-P.; Kao, C.-L.; Chen, H.-F.; Yang, P.-Y.; Chen, H.-Y., *Journal of Physical Chemistry B* **2013,** *117* (7), 2096-2105.
24. Feketeova, L.; Chan, B.; Khairaah, G. N.; Steinmetz, V.; Maitre, P.; Radom, L.; O'Hair, R. A. J., *Journal of Physical Chemistry Letters* **2017,** *8* (13), 3159-3165.
25. Wolken, J. K.; Syrstad, E. A.; Vivekananda, S.; Turecek, F., *Journal of the American Chemical Society* **2001,** *123* (24), 5804-5805.
26. Chen, X. H.; Syrstad, E. A.; Nguyen, M. T.; Gerbaux, P.; Turek, F., *Journal of Physical Chemistry A* **2004,** *108* (42), 9283-9293.
27. Yao, C. X.; Cuadrado-Peinado, M. L.; Polasek, M.; Turecek, F., *Angewandte Chemie-International Edition* **2005,** *44* (41), 6708-6711.
28. Lam, A. K. Y.; Abrahams, B. F.; Grannas, M. J.; McFadyen, W. D.; O'Hair, R. A. J., *Dalton Transactions* **2006,** (42), 5051-5061.
29. Wee, S.; O'Hair, R. A. J.; McFadyen, W. D., *Rapid Communications in Mass Spectrometry* **2005,** *19* (13), 1797-1805.
30. Feketeova, L.; Khairallah, G. N.; Chan, B.; Steinmetz, V.; Maitre, P.; Radom, L.; O'Hair, R. A. J., *Chemical Communications* **2013,** *49* (66), 7343-7345.
31. Lesslie, M.; Lawler, J. T.; Dang, A.; Korn, J. A.; Bím, D.; Steinmetz, V.; Maître, P.; Tureček, F.; Ryzhov, V., *ChemPhysChem* **2017,** *18* (10), 1293-1301.
32. Dang, A.; Nguyen, H. T. H.; Ruiz, H.; Piacentino, E.; Ryzhoy, V.; Turecek, F., *Journal of Physical Chemistry B* **2018,** *122* (1), 86-97.
33. Turecek, F., Transient intermediates of chemical reactions by neutralization-reionization mass spectrometry. In *Modern Mass Spectrometry*, 2003; Vol. 225, pp 77-129.
34. Korn, J. A.; Urban, J.; Dang, A.; Nguyen, H. T. H.; Turecek, F., *Journal of Physical Chemistry Letters* **2017,** *8* (17), 4100-4107.
35. McLuckey, S. A.; Huang, T.-Y., *Analytical Chemistry* **2009,** *81* (21), 8669-8676.
36. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., *Proceedings of the National Academy of Sciences* **2004,** *101* (26), 9528-9533.
37. Zubarev, R. A., *Mass Spectrometry Reviews* **2003,** *22* (1), 57-77.
38. Turecek, F.; Julian, R. R., *Chemical Reviews* **2013,** *113* (8), 6691-6733.
39. Westphal, K.; Wiczk, J.; Miloch, J.; Kciuk, G.; Bobrowski, K.; Rak, J., *Organic & Biomolecular Chemistry* **2015,** *13* (41), 10362-10369.
40. Steenken, S., *Chemical Reviews* **1989,** *89* (3), 503-520.
41. Antoine, R. D., P., UV−Visible Activation of Biomolecular Ions. In *Laser Photodissociation and Spectroscopy of Mass-separated Biomolecular Ions*, Polfer, N., Dugourd, P., Ed. Springer: Cham, 2013; Vol. 83.
42. Polfer, N. C. D., P., Laser Photodissociation and Spectroscopy of Mass Separated Biomolecular Ions. In *Lecture Notes in Chemistry*, Springer: Cham, 2013; Vol. 83, pp 13-20.
43. Furche, F.; Ahlrichs, R., *Journal of Chemical Physics* **2002,** *117* (16), 7433-7447.
44. Comeau, D. C.; Bartlett, R. J., *Chemical Physics Letters* **1993,** *207* (4-6), 414-423.
45. Shaffer, C. J.; Pepin, R.; Turecek, F., *J Mass Spectrom* **2015,** *50* (12), 1438-42.
46. Nguyen, H. T. H.; Shaffer, C. J.; Pepin, R.; Turecek, F., *Journal of Physical Chemistry Letters* **2015,** *6* (23), 4722-4727.
47. Viglino, E.; Shaffer, C. J.; Turecek, F., *Angewandte Chemie-International Edition* **2016,** *55* (26), 7469-7473.
48. Stewart, J. J. P., *Journal of Molecular Modeling* **2007,** *13* (12), 1173-1213.
49. Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P., *Journal of Chemical Theory and Computation* **2009,** *5* (7), 1749-1760.
50. Stewart, J. J. P. *MOPAC 16.*, Stewart Computational Chemistry, Colorado Springs, CO, 2016.
51. Rezac, J., *Journal of Computational Chemistry* **2016,** *37* (13), 1230-1237.
52. Nguyen, H. T. H.; Andrikopoulos, P. C.; Bim, D.; Rulisek, L.; Dang, A.; Turecek, F., *Journal of Physical Chemistry B* **2017,** *121* (27), 6557-6569.

53. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., *Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.
54. Becke, A. D., *Physical Review A* **1988,** *38* (6), 3098-3100.
55. Chai, J.-D.; Head-Gordon, M., *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615-6620.
56. Tomasi, J.; Mennucci, B.; Cammi, R., *Chemical Reviews* **2005,** *105* (8), 2999-3093.
57. Zhao, Y.; Truhlar, D. G., *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.
58. Moller, C.; Plesset, M. S., *Physical Review* **1934,** *46* (7), 0618-0622.
59. Schlegel, H. B., *Journal of Chemical Physics* **1986,** *84* (8), 4530-4534.
60. Mayer, I., *International Journal of Quantum Chemistry* **1978,** *14* (1), 29-38.
61. Chan, B.; Radom, L., *Theoretical Chemistry Accounts* **2011,** *130* (2-3), 251-260.
62. Reed, A. E.; Weinstock, R. B.; Weinhold, F., *Journal of Chemical Physics* **1985,** *83* (2), 735-746.
63. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. A.01*, Wallingford, CT, 2016.
64. Gilbert, R. G.; Smith, S. C.
65. Zhu, L.; Hase, W. L., *Indiana University, Bloomington. Indiana, USA* **1994**.
66. Frank, A. J.; Sadilek, M.; Ferrier, J. G.; Turecek, F., *Journal of the American Chemical Society* **1997,** *119* (50), 12343-12353.
67. Hari, Y.; Leumann, C. J.; Schürch, S., *Journal of The American Society for Mass Spectrometry* **2017,** *28* (12), 2677-2685.
68. Viglino, E.; Lai, C. K.; Mu, X.; Chu, I. K.; Turecek, F., *Journal of the American Society for Mass Spectrometry* **2016,** *27* (9), 1454-1467.
69. Murray, K. K., *Journal of Mass Spectrometry* **1996,** *31* (11), 1203-1215.
70. Petroselli, G.; Dantola, M. L.; Cabrerizo, F. M.; Capparelli, A. L.; Lorente, C.; Oliveros, E.; Thomas, A. H., *Journal of the American Chemical Society* **2008,** *130* (10), 3001-3011.
71. Zhang, Y.; Improta, R.; Kohler, B., *Physical Chemistry Chemical Physics* **2014,** *16* (4), 1487-1499.
72. Colominas, C.; Luque, F. J.; Orozco, M., *Journal of the American Chemical Society* **1996,** *118* (29), 6811-6821.
73. Halder, A.; Bhattacharya, S.; Datta, A.; Bhattacharyya, D.; Mitra, A., *Physical Chemistry Chemical Physics* **2015,** *17* (39), 26249-26263.
74. Wu, R. R.; Yang, B.; Berden, G.; Oomens, J.; Rodgers, M. T., *Journal of Physical Chemistry B* **2014,** *118* (51), 14774-14784.
75. Barbatti, M.; Ruckenbauer, M.; Plasser, F.; Pittner, J.; Granucci, G.; Persico, M.; Lischka, H., *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014,** *4* (1), 26-33.
76. Yao, C. X.; Cuadrado-Peinado, M. L.; Polasek, M.; Turecek, F., *Journal of Mass Spectrometry* **2005,** *40* (11), 1417-1428.
77. Yao, C.; Turecek, F.; Polce, M. J.; Wesdemiotis, C., *International Journal of Mass Spectrometry* **2007,** *265* (2-3), 106-123.
78. Michl, J., *Journal of Molecular Spectroscopy* **1969,** *30* (1), 66-&.
79. Hobza, P.; Sponer, J., *Journal of the American Chemical Society* **2002,** *124* (39), 11802-11808.
80. Hammerum, S., *Journal of the American Chemical Society* **2009,** *131* (24), 8627-8635.
81. Fasman, G. D., *Handbook of biochemistry and molecular biology. Nucleic acids-v. 1-3*. 1975.
82. Dawson, R. M. C.; Elliott, D. C.; Elliott, W. H.; Jones, K. M., *Data for Biochemical Research*. 3rd ed.; Oxford University Press: Oxford, 1986.
83. Verdolino, V.; Cammi, R.; Munk, B. H.; Schlegel, H. B., *Journal of Physical Chemistry B* **2008,** *112* (51), 16860-16873.

Chapter 5. UV-VIS ACTION SPECTROSCOPY AND STRUCTURES OF HYDROGEN-RICH 2′-DEOXYCYTIDINE DINUCLEOTIDE CATION RADICALS. A DIFFICULT CASE

Reproduced in part with permission from Liu, Y.; Korn, J. A.; Tureček, F., *International Journal of Mass Spectrometry* **2019,** *443*, 22-31.

Abstract

Hydrogen-rich cation radicals of 20-deoxycytidine dinucleotide, $(dCC + 2H)^{+\bullet}$, were generated by electron transfer dissociation of their doubly charged complexes with crown ethers and found to be stable in the gas phase. The $(dCC + 2H)^{+\bullet}$ ions were calculated to be formed with ca. 280 kJ mol^{-1} internal energy that was rapidly dissipated below a 130 kJ mol^{-1} energy threshold for the lowest-energy dissociation by loss of water. UV-vis action spectroscopy was applied to monitor several photodissociation channels of $(dCC + 2H)^{+\bullet}$ and to compose spectra that showed distinct absorption bands at 220, 270, 290, and 350 nm, and a broad featureless band at 500-700 nm. UV-vis action spectra were also obtained for the $(dCC – H2O + 2H)^{+\bullet}$ cation radical that showed bands at 225, 270, 330, 430, and 530 nm. Born-Oppenheimer molecular dynamics and density functional theory (DFT) calculations were employed to generate a complete set of tautomeric (dCC + crown + 2H)$^{2+}$ precursor dications and identify the lowest-energy structures in the gas phase and aqueous solution that were the cytosine N-3-H tautomers. Calculations also indicated that cytosine cations and radicals in the lowest-energy $(dCC + 2H)^{+\bullet}$ ions were N-3-H tautomers. Absorption spectra of several $(dCC + 2H)^{+\bullet}$ tautomers were obtained by time-dependent DFT calculations that provided a limited match with the action spectrum. In contrast, a satisfactory match was obtained for the

action and absorption spectra of $(dCC - H2O + 2H)^{+\bullet}$ fragment ions. $(dCC + 2H)^{+\bullet}$ represents a difficult case because of a multitude of cation and radical tautomers to be analyzed and the propensity for consecutive photofragment dissociation upon photon absorption.

5.1 INTRODUCTION

Capture of low-energy secondary electrons is one of the principal mechanisms of DNA damage [1]. When studied in frozen glasses, electron capture by nucleotides was found to result in nucleobase loss from the transient anion radicals [2] , forming 10-deoxyribose radicals that could undergo further reactions. When formed in a protic medium, the basic nucleobase anion radicals can undergo rapid protonation, forming radical hydrogen atom adducts[3]. DNA-electron interactions have been extensively studied by pulse radiolysis that produced complex mixtures and most studies relied on electron paramagnetic resonance spectroscopy to characterize the radical species [4-6]. Hydrogen-rich nucleobase radicals have also been generated by femtosecond electron transfer to mass-selected nucleobase cations and characterized by neutralization-reionization mass spectrometry[7] [8] [9] [10] and ab initio theoretical calculations[11]. Hydrogen-rich DNA cation radicals represent a novel type of transient intermediates in which one nucleotide carries a charge in the form of a protonated nucleobase while another nucleotide has been converted to a hydrogen atom adduct. By having both the charged and radical moieties, hydrogen-rich DNA cation radicals, denoted as $(dXX + 2H)^{+\bullet}$, X = nucleobase, are particularly suited for mass spectrometric studies. Recently, we have reported that hydrogen-rich cation radicals can be efficiently generated by electron transfer dissociation of doubly charged DNA dinucleotides with crown ethers. Cation radicals of $(dAA + 2H)^{+\bullet}$, chimeric RNA dinucleotides[12], $(dGG + 2H)^{+\bullet}$, $(dGC + 2H)^{+\bullet}$, and $(dGC + 2H)^{+\bullet}$ [13] have been generated and characterized by UV-vis photodissociation action spectroscopy [14] [15] [16].

The cytosine rings in dideoxycytidine ions and radicals present a particular challenge for the generation of dications and their reduction by electron transfer. Neutral cytosine is known to exist as N-1-H, O-2-H, N-3-H, and N-4-imine tautomers of very similar relative energies in the gas phase, as elucidated by VUV synchrotron ionization[17 18] , microwave spectroscopy [19] , and *ab initio* calculations [20-24]. Cytosine cation radicals that were generated by intramolecular electron transfer in gas-phase Cu complexes have been shown by infrared multiphoton dissociation (IRMPD) and single-photon UV-Vis action spectroscopies to consist of a mixture of tautomers[25]. *Ab initio* calculations of cytosine cation radicals pointed to N-3-H and O-2-H tautomers as the lowest-energy structures[26]. Cytosine protonation by electrospray ionization has been shown by IRMPD spectroscopy to produce gas phase ions as a mixture of O-2-H and N-3-H tautomers[27 28]. Hydrated cytosine ions[29 30], cytidine nucleosides[31] and nucleotides[32] have all been shown by IRMPD spectroscopy to consists of N-3-H and O-2-H protonated forms in the gas phase. Hydrogen-rich cytosine and 1-methylcytosine radicals have been generated in the gas phase by femtosecond electron transfer and characterized by tandem mass spectra after reionization[33 34]. *Ab initio*[35] and density functional theory[36] calculations of cytosine radicals have identified the N-1-H, N-3-H tautomer as the lowest-energy structure whereas hydrogen atom adducts to C-5, C-6, and several other tautomeric combinations were >30 kJ mol^{-1} higher in energy. Photoexcitation of cytosine has attracted substantial interest, and the studies have revealed that the excited electronic states underwent fast decay with rates and branching ratios that depended on the tautomer structure[37-44]. The previous data indicated that the formation of cytosine ions, cation radicals, and their electronic structure and behavior may be complicated, in particular if the cytosine rings are incorporated in di- or larger oligonucleotides. Here we report an experimental and computational

study addressing the complexities of the formation, action spectra, and comprehensive structure and energy analysis of hydrogen-rich 20-deoxycytidine dinucleotide ions and radicals.

5.2 Experimental Section

5.2.1 *Materials and Methods*

20-deoxycytidine dinucleotide (dCC) was custom made by Integrated DNA Technologies (Coralville, IA) and used as received. 18-Crown-6-ether (CE) and 2,3:11,12-dibenzo-18-crown-6-ether (DBCE) were purchased from Sigma-Aldrich (Milwaukee, WI) and used as received. The dinucleotide and crown ethers in an approximately 1:1.2 molar ratio were dissolved in 80:20:1 mixture of acetonitrile:water:acetic acid to achieve concentrations in the 20-50 mM range. The solutions were electrosprayed into the ion trap mass spectrometer. Electron transfer dissociation (ETD) and photodissociation mass spectra were measured on two different instruments. One was a modified ThermoElectron Fisher (San Jose, CA, USA) LTQ-XL ETD linear ion trap mass spectrometer that was furnished with an external EKSPLA NL301G (Altos Photonics, Bozeman, MT) Nd-YAG laser source operating at a frequency of 20 Hz with a 3-6 ns pulse width, and an optical parametric oscillator system, as described previously[14]. The other instrument was a Bruker amaZon Speed 3D ion trap mass spectrometer that was equipped with windows allowing laser beam entrance to the ion trap and furnished with the same EKSPLA NL301G laser. The operating procedures for obtaining UV-vis action spectra in the 210-700 nm region on both instruments have been described[45 46]. The electrospray ion source, and ion transfer ion optical elements were tuned to optimize the formation of doubly charged complexes. Doubly charged ions were isolated by mass in the ion trap and allowed to react with fluoranthene anions for 100-200 ms. The charge-reduced cation radicals formed were isolated by mass and exposed to laser pulses for

photodissociation. The number of laser pulses used for photodissociation varied from 1 pulse in the 400-700 nm region, 1-5 pulses in the 210-350nm region, to 5 pulses in the 353-400 region. Each wavelength point in the UVPD spectrum consisted of >100 accumulated and averaged scans. The reported action spectra are averages of two reproduced measurements obtained on different days.

5.2.2 *Calculations*

Born-Oppenheimer molecular dynamics of doubly protonated (dCC + crown ether) complexes was run with semiempirical quantum PM6 calculations[47] that were augmented by including dispersion and hydrogen bonding interactions, D3H4[48]. BOMD trajectories were run at 1 fs steps for 20 ps with energy conservation by weak coupling to a Berendsen thermostat[49], generating 20000 conformer snapshots for each tautomeric ion. These calculations were run with MOPAC[50] that was coupled to the Cuby4 platform[51]. Two hundred snapshots were selected from each trajectory, the structures were fully gradient-optimized with PM6-D3H4 and sorted out by their secondary and super-secondary structural similarities. Several low-energy complex structures were selected for each tautomer and their geometries were optimized with B3LYP[52] and the 6-31G (d,p) basis set to yield harmonic frequencies. The complex structures were further reoptimized with ωB97X-D[53] and M06-2X[54], both with the 6-31+G (d,p) basis set, to evaluate electronic energies including dispersion interactions. Further sets of energies were obtained by single-point M06-2X/6-311++G(2d,p) and MP2(frozencore)/6-311++G(2d,p) calculations[55]. Contamination by higher spin states in the UMP2 calculations was treated by standard spin annihilation methods (PMP2[56] [57]) and further checked by restricted open-shell MP2 (ROMP2) single-point calculations[58]. The electronic, vibrational, and rotational energies and entropies were combined to yield relative free energies at the experimental temperature of 310 K. To assess solvent effects, we

performed single-point energy self-consistent reaction field calculations using the polarizable continuum model[59] with van der Waals surfaces, water as the dielectric, and structures optimized in the gas-phase and dielectric. Vertical excitation energies and oscillator strength were calculated by time-dependent density functional theory (TD-DFT[60]) using M06-2X/6-31+G (d,p). Our previous investigations of electronic excitations in nucleobase, nucleoside, and dinucleotide radicals have indicated that TD-DFT M06-2X/6-31+G (d,p) calculations provide reliable excitation parameters when benchmarked against higher-level, equation-of-motion coupled-cluster (CCSD), calculations[61]. All the thermochemical calculations were performed using the Gaussian 16 (Revision A.03) suite of programs [62].

5.3 Results and Discussion

5.3.1 *Ion formation*

Electrospray ionization of dCC-crown ether mixture solutions yielded doubly charged complexes that were isolated by mass and subjected to ion-ion reactions with fluoranthene anions. ETD[63] resulted in charge reduction and elimination of the crown ligand, as shown for the complex with dibenzocrown ether (DBCE), $(dCC + 2H + DBCE)^{2+}$ at *m/z* 439 (Figure 5.1a). The complex with 18-crown-6-ether (CE) gave similar results. In both cases ETD generated the $(dCC + 2H)^{+\bullet}$ cation radical at *m/z* 518 as a major product in an excellent yield. In contrast, ETD of $(dCC + 2H)^{2+}$ resulted in complete dissociation (Figure S1, Supplementary Information). The other products from ETD of the crown ether complex were $(dCC + H)^{+}$ (*m/z* 517) and $(dCC + H + crown)^{+}$ at m/z 877 and 781 for DBCE and CE, respectively. The $(dCC + 2H)^{+\bullet}$ cation radical was isolated by mass and probed by collision-induced dissociation, CID-MS^3 (Figure 5.1b). This produced a

dominant fragment ion by loss of water (*m/z* 500), and further fragment ions by loss of $(C + H)^{\bullet}$ radicals (*m/z* 406), and their secondary products by loss of water (*m/z* 388). Backbone dissociations formed the $(d_1/w_1 + 2H)^+$ (*m/z* 310), $(d_1/w_1)^+$ (*m/z* 308), and a_1/z_1 (*m/z* 210) ions. The backbone ions assignment followed the standard nomenclature[64 65]. Interestingly, most of the backbone fragment ions were even-electron species, indicating loss of neutral radicals from $(dCC + 2H)^{+\bullet}$.

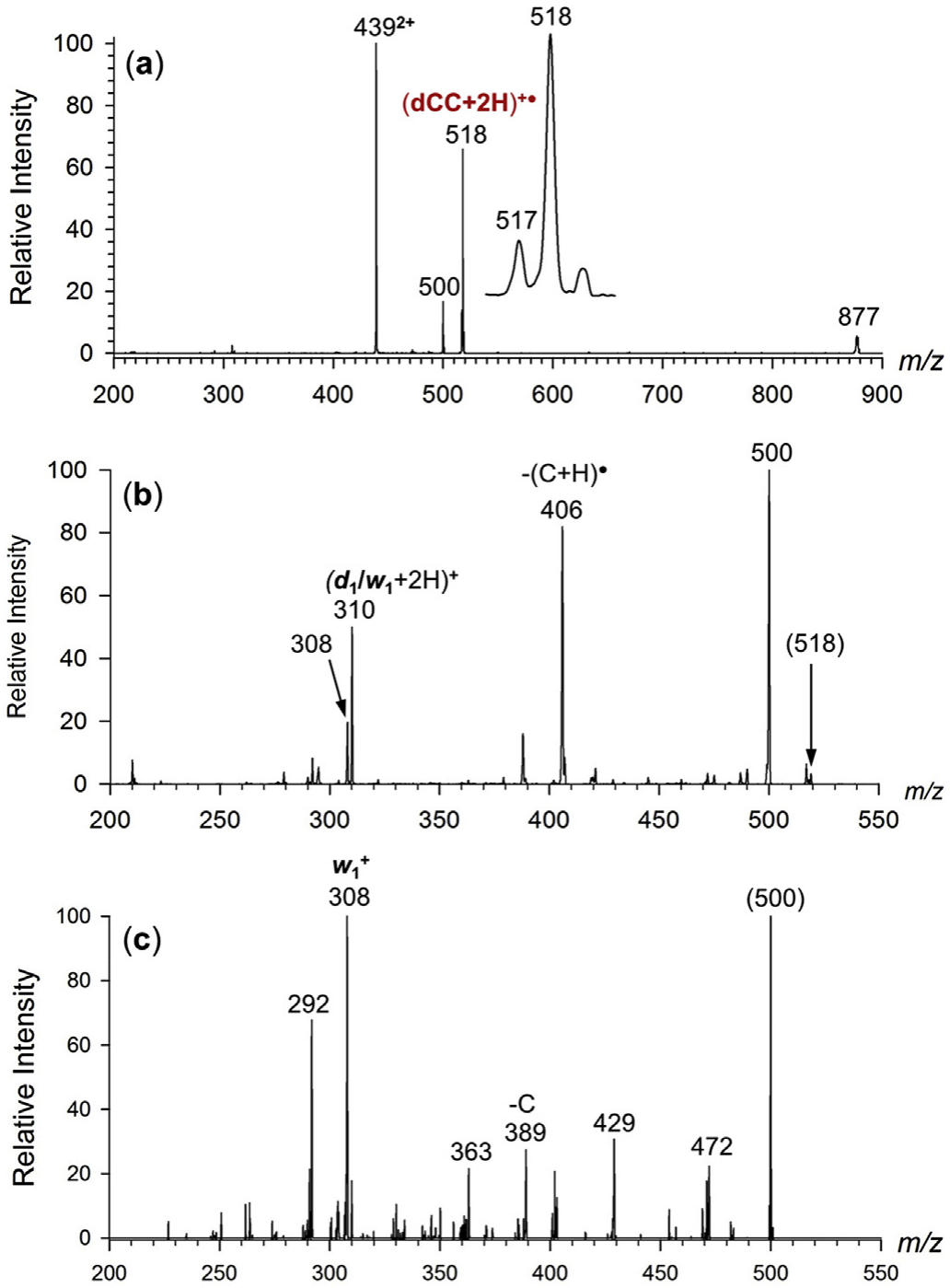


Figure 5.1. ETD spectrum (fluoranthene, 150 ms) of (dCC + DBCE + 2H)$^{2+}$ complex at m/z 439. (b) CID-MS3 spectrum of (dCC + 2H)$^{+\bullet}$ ion at m/z 518. (c) CID-MS4 spectrum of the (dCC – H2O + 2H)$^{+\bullet}$ ion at m/z 500.

The prominent $(dCC + 2H - H_2O)^{+\bullet}$ cation-radicals were further investigated by CID-MS^4 that yielded the mass spectrum shown in Figure 5.1c. This displayed a dominant ion at m/z 308 that we assigned as d_1/w_1, and several other fragment ions formed by combined losses of CO, C_2H_3O, and cytosine. Interestingly, the $(d_1/w_1 + 2H)^+$ ion from $(dCC + 2H)^{+\bullet}$ appeared as $(d_1/w_1 - H_2O + 2H)^+$ at *m/z* 292 in the CID spectrum of $(dCC + 2H - H_2O)^{+\bullet}$ (Figure 5.1c). These fragmentations involving double hydrogen transfer have not been observed in the CID spectra of closed-shell DNA cations[65] and appear to be associated with the radical cytosine rings in $(dCC + 2H)^{+\bullet}$ and $(dCC+2H-H_2O)^{+\bullet}$. We will discuss the possible reaction sequence leading to these fragmentations later in the paper.

5.3.2 *Photodissociation action spectra*

Photodissociation of mass-selected $(dCC + 2H)^{+\bullet}$ ions was investigated in the wavelength range of 210-700 nm, covering the valence-electron excitations in the cation-radicals. The major photodissociation products were observed at *m/z* 500, 406, 310, 308, and 292 that were analogous to fragment ions produced by CID. The *m/z* 406 and 500 channels turned out to be problematic in following the absorption by $(dCC + 2H)^{+\bullet}$, as they showed unusual characteristics over most of the covered wavelength region. The dominant *m/z* 500 channel further displayed gradually increasing dissociation at $l > 500$ nm which was unprecedented in the action spectra of other dinucleotide cation radicals[12 13]. The action spectrum that was combined from the *m/z* 292, 308, 310, 406 and 500 channels showed distinct bands at 350, 290, 270, and 220 nm (Figure 5.2a), in addition to a broad featureless band at 500-700 nm. Light absorption at 350 and 270 nm was pronounced in the *m/z* 308 and 310 channels (Figure 5.2b), whereas the 220 nm band was in part contributed by the *m/z* 292 channel (Figure 5.2c).

The photodissociation characteristics of the fragment ion by loss of water (*m/z* 500) was further investigated by UVPD-MS^4. The action spectrum showed bands at 330, 270 and 225 nm that were analogous to those found for $(dCC + 2H)^{+\bullet}$. In addition, weak bands were observed at 530 and 430 nm (Figure 5.3a). The main bands in the action spectra of $(dCC + 2H)^{+\bullet}$ and the *m/z* 500 fragment ion indicated similar chromophores that were associated with the pertinent radical moieties.

5.3.3 *Dication structures*

To characterize the absorption properties of $(dCC + 2H)^{+\bullet}$ we undertook an extensive computational analysis of the structures involved in the ion formation, photoexcitation, and dissociation. Our goal was to first establish the thermodynamically most stable isomers of the $(dCC + crown + 2H)^{2+}$ precursor ions to determine the protonation sites, cytosine tautomers and crown-ether attachment sites. This was followed by conformational analysis of energy-selected $(dCC + 2H)^{+\bullet}$ cation radicals and several products of their isomerization by proton and hydrogen atoms migration. In the last step, we performed TD-DFT calculations of 45 excited states of several representative $(dCC + 2H)^{+\bullet}$ isomers.

Table 5.1. Relative energies of $(dCC + CE + 2H)^{2+}$ ions[a]

Type	Conformer	Relative Energy[b,c]			
		Gas phase		Water	
		ΔH (0)[d]	ΔG (310)[e]	ΔG_{aq} (310)[f]	
N-3-H, N-3-H, 3′-crown	**1a**	0.3	–11	7.1[g]	8.5[h]
	1b	0	0	0	0
N-3-H, N-3-H, 5′-crown	**2a**	17	7.5	15	19
	2b	26	15	19	13
O-2-H, N-3-H, 3′-crown	**3a**	14	–3.2	42	43
	3b	10	–0.5	50	55
N-3-H, O-2-H, 5′-crown	**4a**	14	7.1	40	46
N-3-H, O-2-H, 3′-crown	**5**	67	58		
O-2-H, N-3-H, 5′-crown	**6**	72	60		
O-2-H, O-2-H, 3′-crown	**7**	99	89		
O-2-H, O-2-H, 5′-crown	**8**	74	73		

[a] CE stands for 18-crown-6-ether.
[b] Energies in kJ mol^{-1} for ωB97X-D/6-31 + G (d,p) optimized structures.
[c] Including B3LYP zero-point vibrational energies.
[d] Relative enthalpies in the gas phase at 0 K.
[e] Relative free energies in the gas phase at 310 K.
[f] Including solvation energies in the water dielectric.
[g] Optimized with inclusion of solvation by water via the polarizable continuum model.
[h] PCM single-point energies on gas-phase optimized structures.

To begin with the (dCC + 2H + crown)$^{2+}$ precursor ions, we used BOMD to generate conformers of all 8 combinations of complexes consisting of 4 cytosine O-2 and N-3-protonated tautomers, each with a 30- or 50-crown ether ligand. As reported previously[13 14], dinucleotide dication complexes with 18-crown-6-ether and 2,3:11,12-dibenzo-18-crown-6-ether showed very similar energy rankings of tautomer and coordination isomers, and so we used the smaller 18-crown-6-ether as an economical surrogate in these extensive calculations. The DFT-calculated free energies of fully optimized gas-phase complexes revealed a distinct preference for structures in which the crown ether was coordinated to the N-3-H protonated cytosine in the 3′ (**1a,b**) or 5′ positions (**2a,b**) (Figure 5.4, Table 5.1). This distinction was further amplified by solvent effects

in water where structures **1a, 1b** represented the lowest free-energy isomers (Table 5.1). Complexation to the N-3-H cytosine cation at the 30 and 50 termini in tautomers having the other cytosine protonated at O-2-H also gave rise to low-energy gas-phase structures (**3a, 3b** and 4a, Figure 5.4). The reason for the stability of the N-3-H coordinated complexes in the gas phase was evident from the hydrogen bonding to the crown ether, as the N-3-H protonated cytosine ring can develop three strong hydrogen bonds to the crown-ether oxygens, one by N-3-H and two by protons of the NH_2 group (Figure 5.4). In contrast, coordination to the O-2 protonated cytosine ring in **5-8** was less favorable, as the cytosine ring developed only two hydrogen bonds to the crown ligand, one by the O-2-H and the other by one of the NH_2 protons (Figure 5.4). The reason for the preferential coordination to the 3′-N-3-H protonated cytosine, e.g. **1a, 1b** versus **2a**, was less obvious and was likely due to a combination of several subtle effects. In summary, coordination by the crown ether in combination with solvent effects steered the cytosine protonation in dCC to the N-3 positions favoring a single type of the dinucleotide ion tautomer.

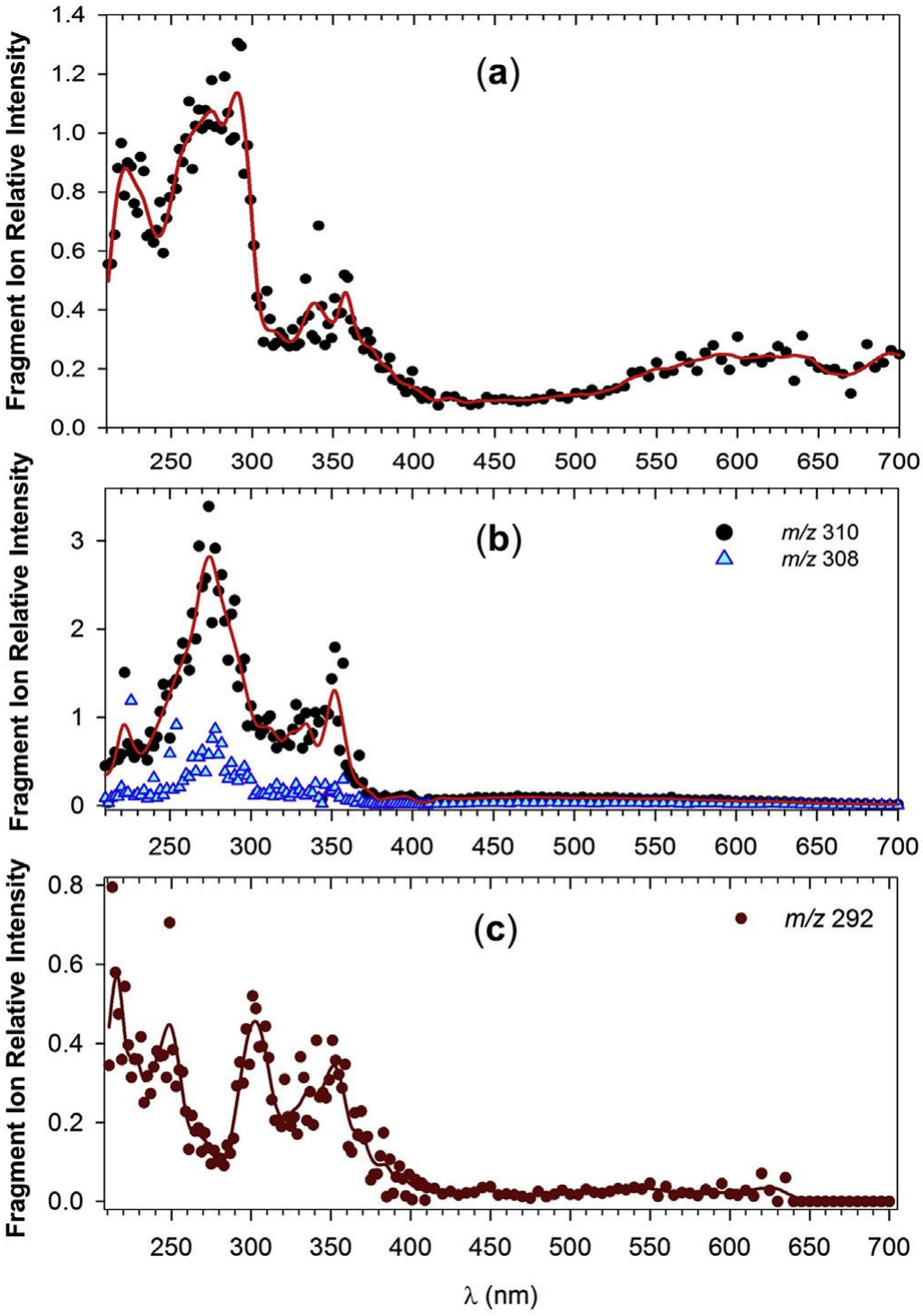


Figure 5.2. UV-Vis action spectra of $(dCC + 2H)^{+\bullet}$. (a) Sum of photofragment ion intensities. (b) Relative intensities of m/z 308 and 310 photofragment ions. (c) Relative intensities of m/z 292 photofragment ions.

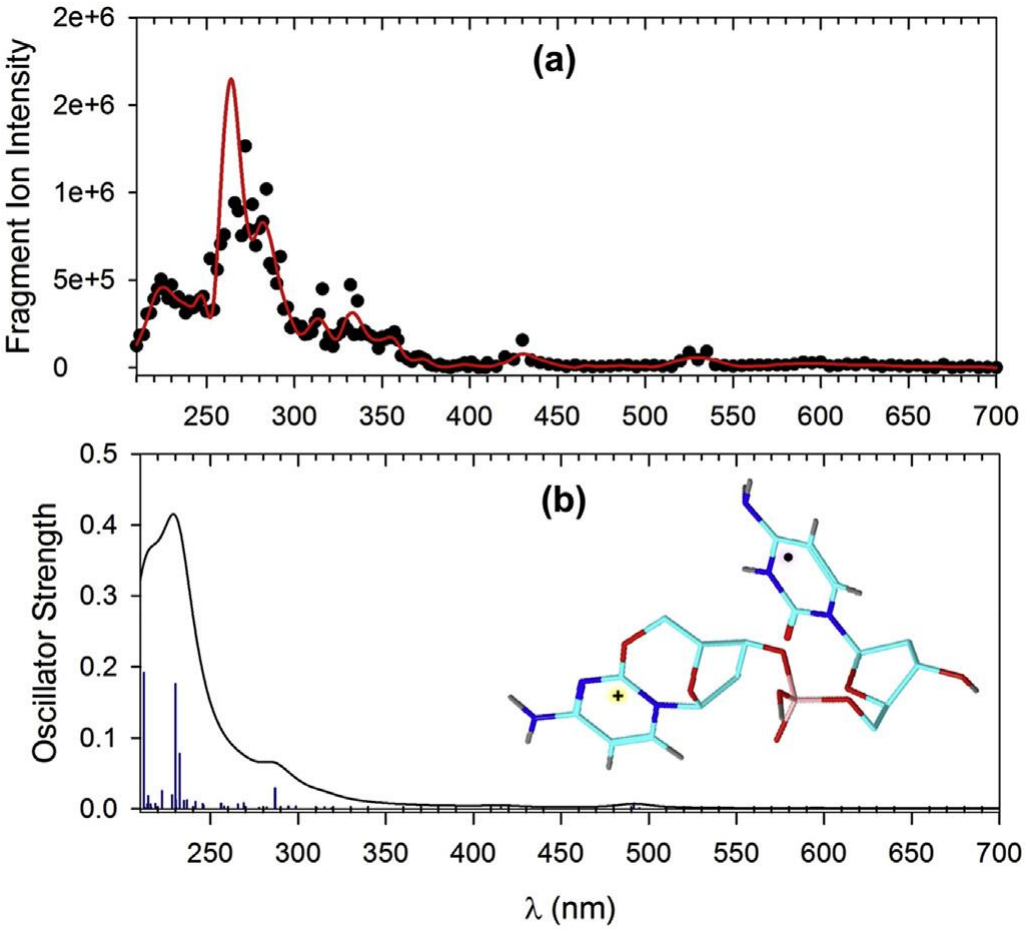


Figure 5.3. (a) Action spectrum of the $(dCC - H2O + 2H)^{+\bullet}$ from ETD-CID-MS4. (b) TD-DFT M06-2X/6-31+G (d,p) absorption spectrum of 17. The lines were artificially broadened by convolution with a Lorentzian function at 12 nm full-width at half maximum.

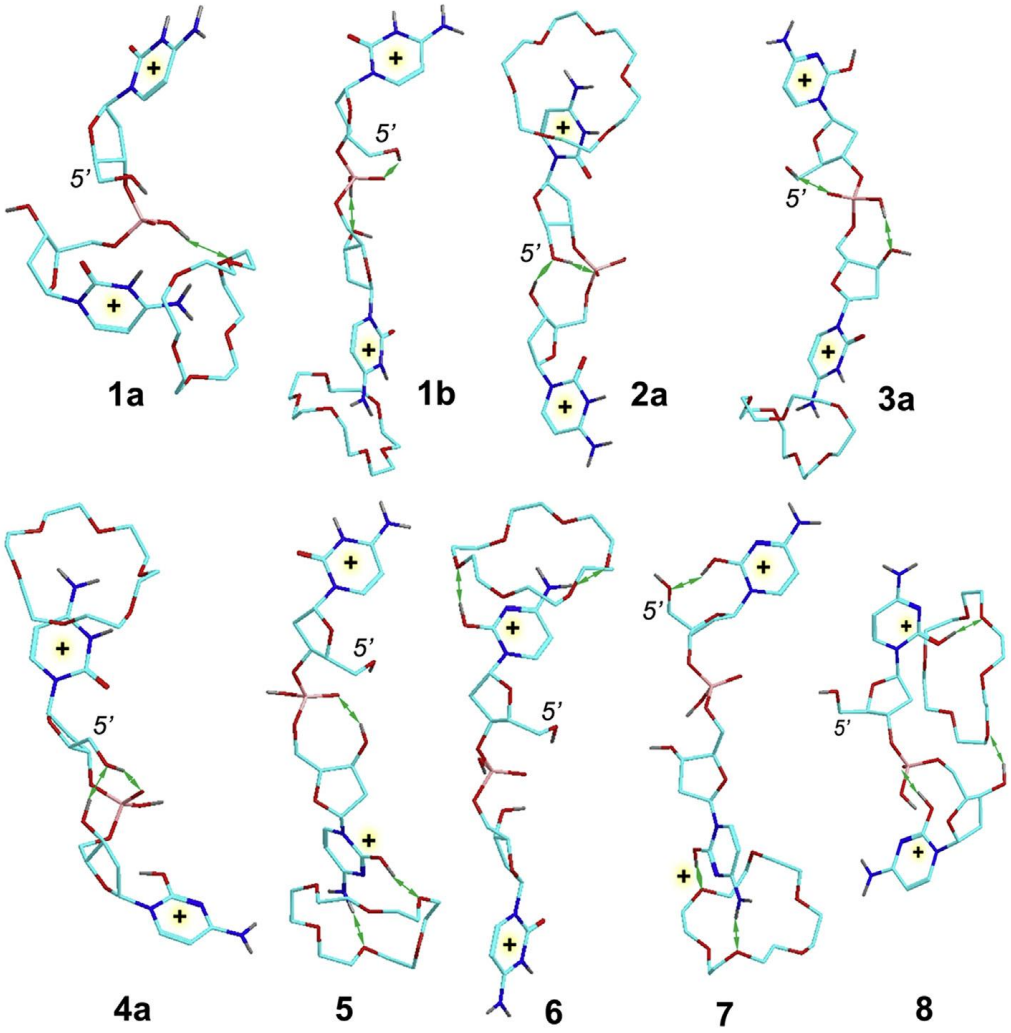


Figure 5.4. ωB97X-D/6-31+ G (d,p) optimized structures of selected low-energy (dCC + crown +2H)$^{2+}$ ions. Atom color coding is as follows: cyan = C, gray = H, blue = N, red = O, Pink = P. Only exchangeable (N-H, O-H) hydrogens are shown. Major hydrogen bonds are indicated by green arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

5.3.4 *Cation radical formation and structures*

Electron attachment to the (dCC + 2H + crown)$^{2+}$ ions was investigated for the lowest energy structure **1a**. According to ωB97X-D calculations, the reducing electron entered the non-coordinated cytosine ring (1r, Scheme 5.1). This was analogous to electron capture by other

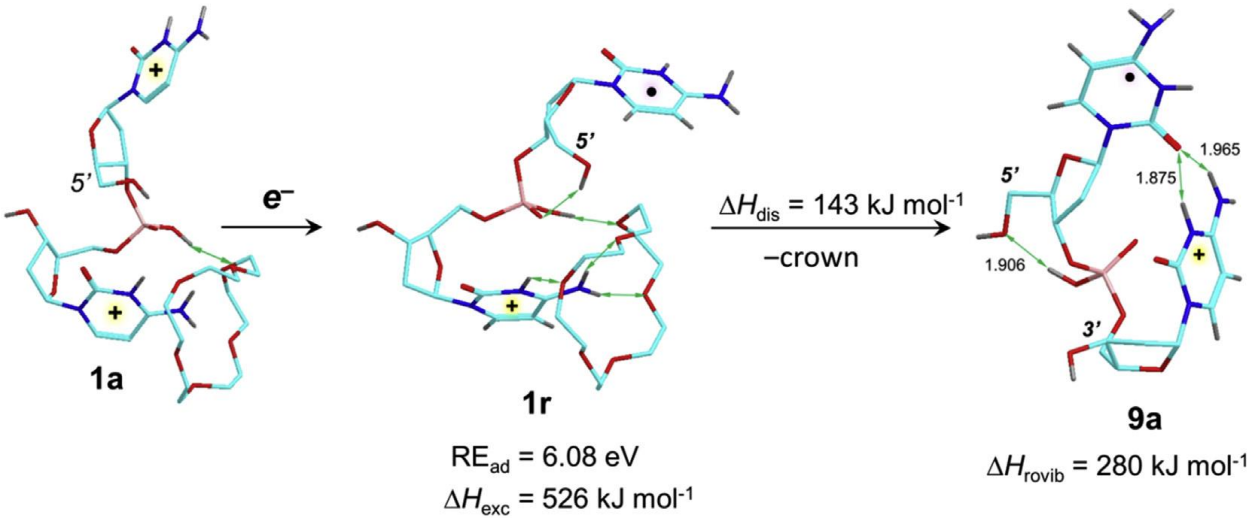
dicationic crown-ether complexes where solvation by the electron-donating crown ether was shown to stabilize the ion, decreasing its recombination energy, and steering the incoming electron into the non-solvated charged group[66]. Electron attachment to $(dCC + crown + 2H)^{2+}$ was associated with an adiabatic recombination energy that was calculated as REadiab =6.08 eV. Considering the electron affinity of the fluoranthene electron donor, EA = 0.63 eV [67], the vibrational excitation energy in the $(dCC + crown + 2H)^{+\bullet}$ intermediate complex **1r** can be estimated[68] as $E_{excz}(6.08 - 0.63)$ = 5.45 eV (526 kJ mol^{-1}) which is added to the 310 K ro-vibrational enthalpy in **1a** (120 kJ mol^{-1}). The high vibrational energy of charge-reduced **1r** drove the dissociation of **1r**, overcoming the binding energy of the crown ether ($\Delta H_{dis,310}$ =143 kJ mol^{-1}, $\Delta G_{dis,310}$=70 kJ mol^{-1}, by ωB97X-D/6-31+G(d,p) including counterpoise corrections), and resulting in ca. 503 kJ mol^{-1} non-fixed energy which was partitioned between $(dCC + 2H)^{+\bullet}$ and the crown ether. From the energy partition terms based on the number of vibrational degrees of freedom in $(dCC + 2H)^{+\bullet}$ and the crown ether, 180 and 144, respectively[69], we estimated the initial excitation energy in $(dCC + 2H)^{+\bullet}$ formed by ETD to be ca. 503 ×180/(180+144) ≋ 280 kJ mol^{-1}.

Table 5.2. Relative energies of $(dCC + 2H)^{+\bullet}$ ions[a]

Ion/Reaction	Radical Site	Relative Energy[a,b]			
		ωB97X-D	M06-2X	PMP2	ROMP2
9a	5′	0ω	0	0	0
9b	5′	19	16	17	18
9c	5′	59	25	21	22
9d	5′	36	39		
10	3′	31	29	28	27
11a	3′	54	35	34	33
11b	3′	53	48		
12	5′	63	49		
13	5,5-H_2	37	34	31	38
14	6,6-H_2	56	58		
15	5′	73	61		
16	3′	97	87		
TS (9a → 13)		97	103	92	96
9a → 17 + H_2O		149	130	127	126
9a → 18 + H_2O		199	188	189	196
17 → *m/z* 389 + cytosine		146	130	154	161
17 → 19 + 20		170	154	167	156
9a → 19 + 20		318	284	294	282

[a] Energies in kJ mol^{-1}.
[b] Including B3LYP zero-point vibrational energies and referring to 0 K in the gas phase.

Under the conditions in the ion trap, the excitation can drive dissociations and isomerizations of the $(dCC + 2H)^{+\bullet}$ cation radical, or be dissipated by collisions with the bath gas[69] [70].



Scheme 5.1. Electron transfer and dissociation of the $(dCC + crown + 2H)^{2+}$ complex **1a.**

Geometry optimization of $(dCC + 2H)^{+\bullet}$ cation radicals yielded several structures that differed in their conformation and position of the charging proton and odd electron. In general, one-electron reduction was expected to occur in one of the O-2 or N-3-protonated cytosine rings which was consistent with the calculated cation-radical structures. The relative energies quoted in the text refer to M06-2X/6-311++G (2d,p) calculations, energies from the other calculations are listed in Table 5.2. The calculations pointed to the 5′-N-3-H-cytosine radicals **9a** (Figure 5.5) and **9b** (Figure S2) as the lowest-energy isomers (Table 5.2). Note that the unpaired electron distribution in **9a** and **9b** directly correlated with that in the charge-reduced crown complex **1r** (Scheme 5.1). The position of the radical was evident from the puckered conformation of the charge reduced ring and confirmed by the calculated atomic spin densities that showed >98% of

the unpaired spin to be located in the 50-cytosine ring. The 30-N-3-H-cytosine radicals were ca. 30 kJ mol^{-1} higher in energy for the most stable conformer (**10**, Figure 5.5). Isomeric cation-radicals with an O-2-protonated ring in the 3′- (**12**) or 5′-position (**11a**) had the radical in the complementary N-3-H ring, but were overall >35 kJ mol^{-1} less stable than **9a**. The ordering of relative energies for the cytosine radical tautomers in $(dCC + 2H)^{+\bullet}$ followed that of cytosine radicals where the N-1-H, N-3-H tautomer is the most stable structure[33 36]. It is noteworthy that all the low-energy $(dCC + 2H)^{+\bullet}$ structures comprised hydrogen bonds between the charged ring as a proton donor and the reduced ring as a proton acceptor. The dramatic structure change from the mostly extended dications to folded cation radicals can be attributed to the combined effects of canceling the Coulomb repulsion between the charged cytosine rings upon reduction, and allowing hydrogen bonding between the charged and radical moieties, as in structures **9-11, 15**, and **16** (Figure 5.5). This was consistent with the previously reported conformational collapse in other dinucleotide cation radicals[12 13].

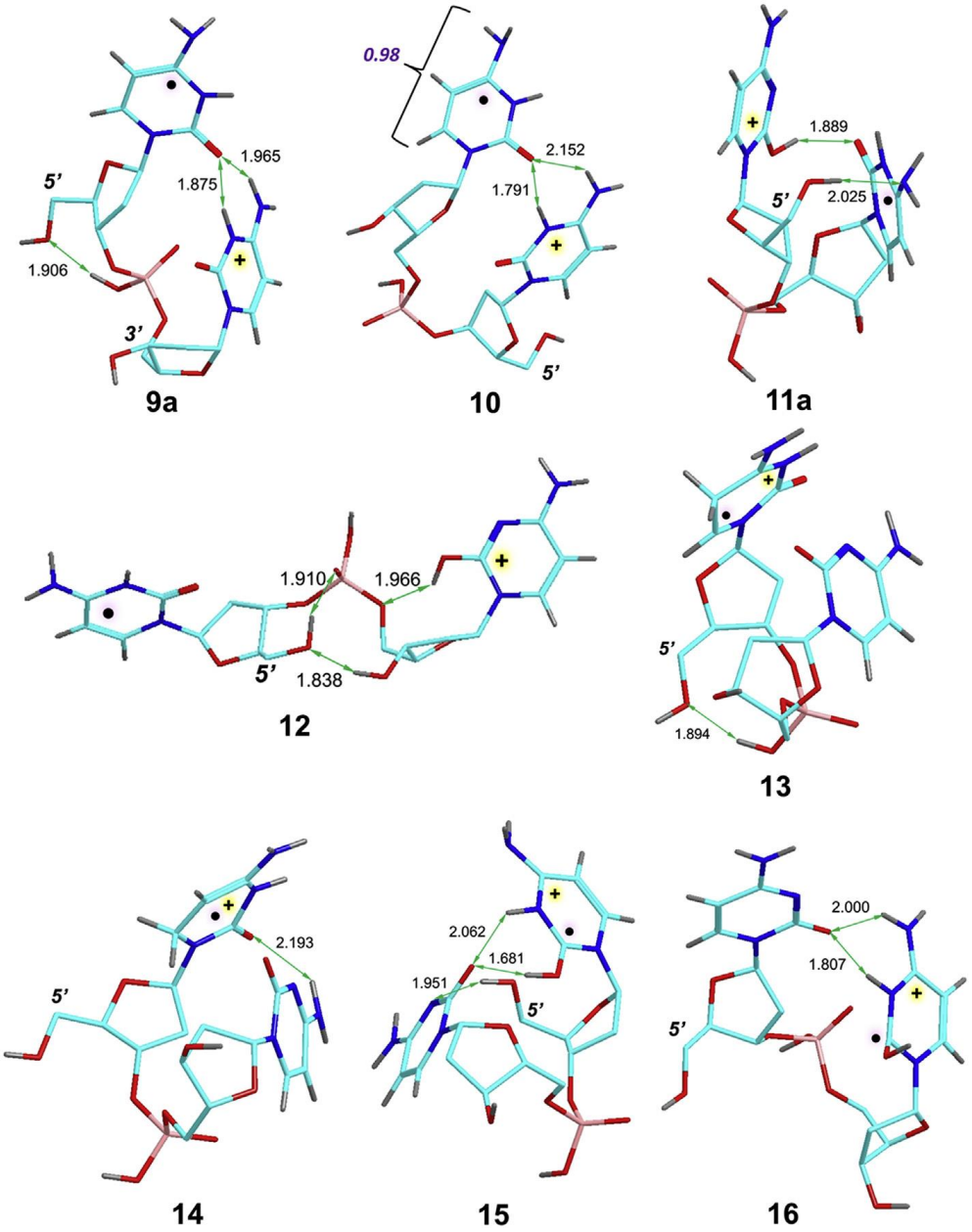
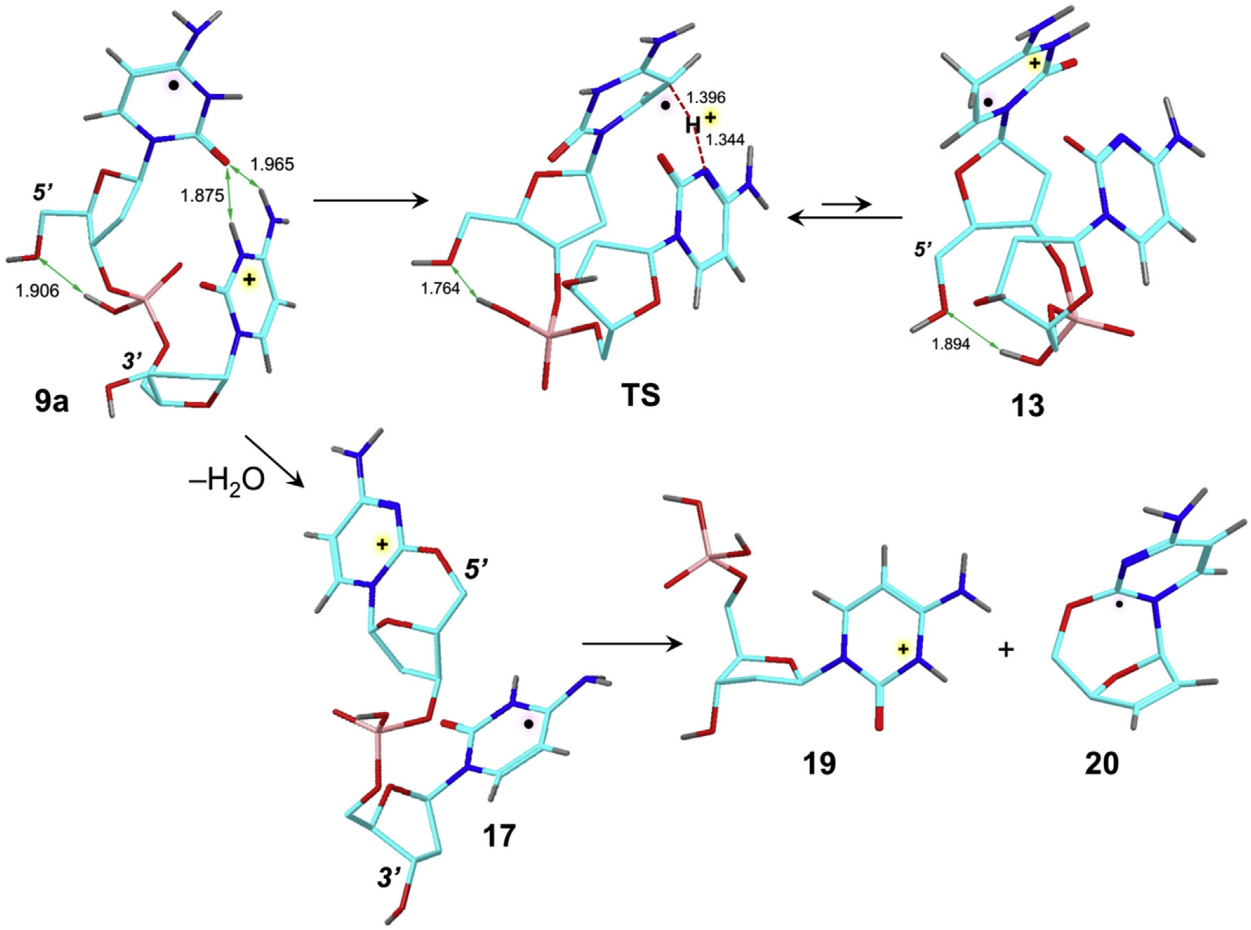


Figure 5.5. M06-2X/6-31+G (d,p) optimized structures of low-energy (dCC + 2H)$^{+\bullet}$. Color coding is as in Figure 5.4. Major hydrogen bonds are indicated by green arrows with distances in Ångstrøms. The sum of atomic spin densities is shown for structure 10. (For interpretation of the references to color in this figure legend, the reader is referred to theWeb version of this article.)

In addition to the canonical cation radicals that can be directly produced by electron attachment to the dications, we also investigated structures resulting from proton or hydrogen atom transfer between the cytosine units. Hydrogen atom migration to C-5 of the 5′-cytosine formed cation radical **13** (Figure 5.5) that was 34 kJ mol^{-1} less stable than **9a**. An analogous H-atom migration to the C-6 position formed an even less stable radical **14** at 58 kJ mol^{-1} relative to **9a**. Proton migration to the 5′- and 3′-radical cytosine ring produced dihydrocytosine cation radicals **15** and **16**, respectively (Figure 5.5). However, these were substantially less stable than **9a** (Table 5.2).

The reaction energies of $(dCC + 2H)^{+\bullet}$ were calculated for the main channels and referenced to the lowest energy structure **9a**. Isomerization by a migration of the 3′-N-3-H cytosine proton to C-5 in the 5′-cytosine (**13**, Scheme 5.2) was calculated to have a low TS energy (103 kJ mol^{-1}) and can proceed in **9a** at the initial mean vibrational energy of 280 k mol^{-1} provided by ETD (vide supra). RRKM calculations of the rate constant for the isomerization in **9a** (Figure S3a) indicated a relatively slow reaction that would achieve ca. 20% conversion at 280 kJ mol^{-1}. However, the reverse isomerization of **13** to **9a** was 34 kJ mol^{-1} exothermic and fast, favoring **9a** by a factor of >100 over a broad range of internal energies (Figure S3b). Loss of water was the lowest-energy dissociation of $(dCC + 2H)^{+\bullet}$, requiring a threshold energy of 130 kJ mol^{-1} from **9a** to form ion **17** (Scheme 5.2). An isomer of $(dCC - H_2O + 2H)^{+\bullet}$ in which the 30-cytosine was cyclized to the 50-position (18) was also considered, but it was calculated to be 58 kJ mol^{-1} less stable than ion **17**. The low threshold energy for the formation of **17** was consistent with the dominant loss of water from $(dCC + 2H)^{+\bullet}$ upon CID (Figure 5.1b). Considering the estimated initial excitation of the ETD-formed $(dCC + 2H)^{+\bullet}$ (280 kJ mol^{-1}, vide supra), it was not surprising that upon ETD of the complex, a fraction of the cation radicals underwent water elimination (cf.

Figure 5.1a) in competition with collisional cooling. A further dissociation of ion **17** forming the w_1 ion (**19**, *m/z* 308) and a_1 radical (**20**) was a low-energy process for which we calculated a threshold energy of 154 kJ mol^{-1}. This backbone cleavage competed with the loss of the 3′-guanine base (*m/z* 389, Figure 5.1c) with a threshold energy of 130 kJ mol^{-1} (Table 5.2). We note that the energetics of these competing dissociations was more realistically captured by MP2 calculations (Table 5.2).



Scheme 5.2. Isomerization and dissociation of $(dCC+2H)^{+\bullet}$ cation radical **9a.**

5.3.5 *Calculated absorption spectra and action spectra assignment*

We used the optimized structures of $(dCC + 2H)^{+\bullet}$ ions for TDDFT calculations of excitation energies and oscillator strength. M06-2X and ωB97X-D TD-DFT calculations gave very

similar results for the excited states, and so only the former are presented here. Likewise, applying the larger 6-311++G (2d,p) basis set in the TDDFT M06-2X calculations gave data that were very similar to those from calculations with the 6-31 + G (d,p) basis set, as illustrated with **15** (Figure S4). The absorption spectra of the lowest-energy 5′-N-3-H-cytosine radicals (**9a** and **9b**) showed weak bands in the 550-600 and 390-410 nm regions. Several weak bands were also found in 270-330nm region (Figure 5.6). The main absorption lines appeared in the 230-260 nm region of the spectrum. The line wavelengths and oscillator strength differed for **9a** and **9b**, indicating that the spectra were sensitive to the cation radical conformation. However, band broadening by vibronic transitions was likely to obscure the differences, as indicated by the simulated absorption peaks that were artificially broadened by convolution of Lorentzian functions at 12 nm full width at half maximum (Figure 5.6). The spectrum of the isomeric 3′-N-3-H-cytosine radical (**10**) displayed the main line at 260 nm which resulted in a different band pattern even after broadening. The absorption spectra of the 3′-N-3-H$^{\bullet}$ radicals having the 5′-ring protonated at O-2 (**11a, 11b**) showed distinct features in the 210-300 nm region. In particular, the spectra displayed a blue shift of the main lines and different pattern of the weak lines at 250-330 nm. A blue shift in the 210-250 nm region was also observed for the 5′-N-3-H$^{\bullet}$ radical having the 3′-ring protonated at O-2 (**12**). The spectrum of the 5,5-H_2 isomer **13** had an absorption band at 450 nm and stronger bands at l < 250 nm (Figure 5.6). The 6,6-H_2 isomer (**14**) gave a still different spectrum, showing prominent bands at 476 and 350 nm and other bands in the 300-400 nm region. Finally, the 5′-N-3, O-2-H cation radical (**15**) showed no bands above 360 nm, with main absorption appearing in the 230-250 nm region and at wavelengths < 200 nm (Figure 5.6).

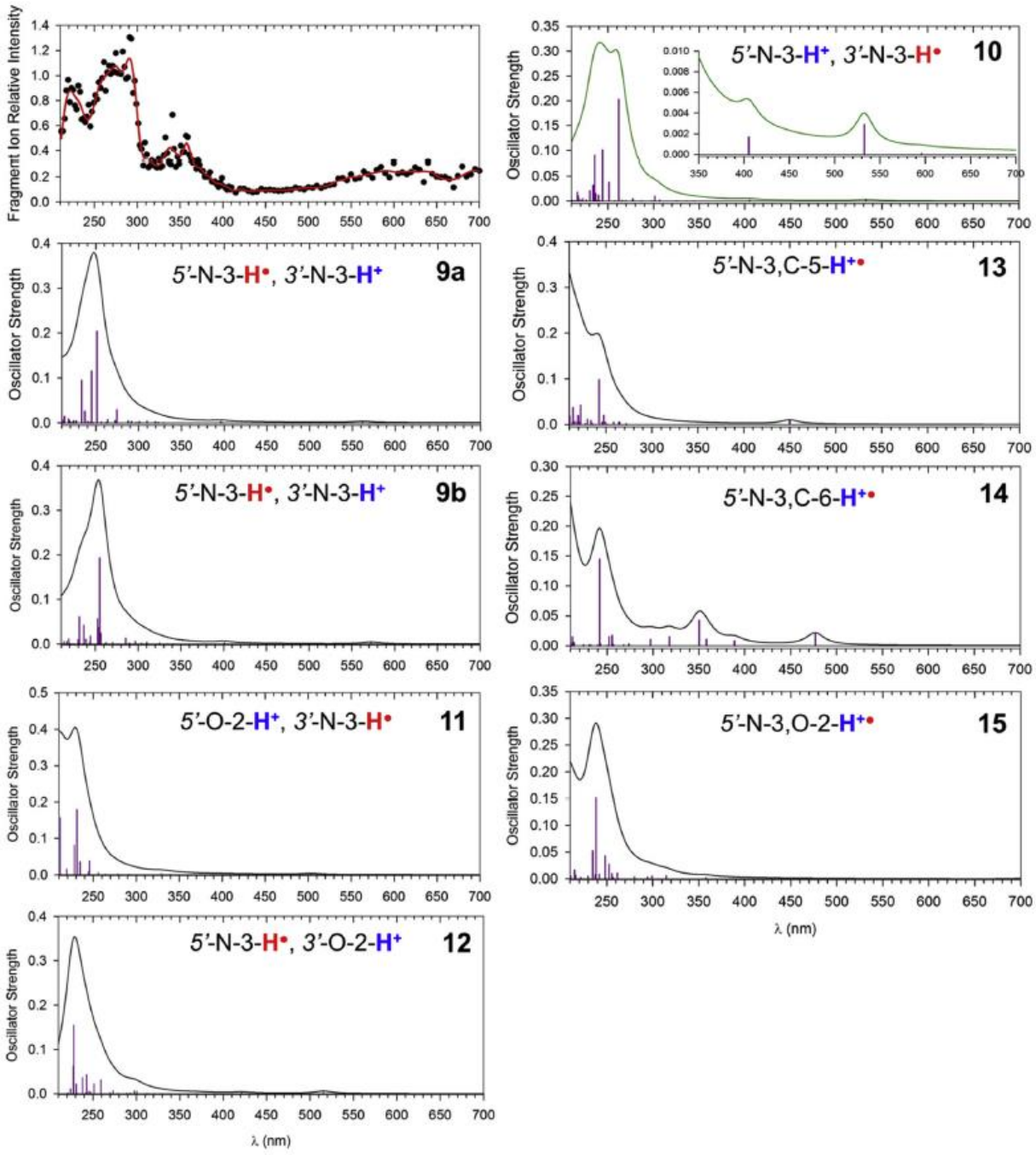


Figure 5.6. Action and TD-DFT absorption spectra of $(dCC + 2H)^{+\bullet}$ ions. The nucleobase 5' or 3' positions are denoted by italics.

A comparison of the action spectrum with the calculated absorption spectra revealed both matches and discrepancies. The 220-nm band had counterparts in the absorption spectra of **11** and **13**. Although we did not calculate multiple vibronic spectra for the isomers, we note that vibronic

transitions typically result in a red shift and band broadening. Thus electron transitions in several isomers with l < 200 nm may appear at 220 nm in the action spectrum. The 270 and 290 nm bands in the action spectrum are usually not much diagnostic for DNA oligonucleotide cation radicals l > 500 nm which was unprecedented in the action spectra of other dinucleotide cation radicals[12 13]. Considering the vibronic red shift, the absorption spectrum of **10** showed the closest match of the dual-peak band at 240 and 260 nm with the corresponding bands in the action spectrum. The 350-nm band in the action spectrum was the most difficult one to match. Although most of the low-energy $(dCC + 2H)^{+\bullet}$ isomers showed multiple transitions in this region of the spectrum, the bands had low oscillation strength to match the experimental band. The C-6-H radical **14** was the only isomer displaying a strong band at 350nm in the absorption spectrum; however, **14** was a high-energy isomer (Table 5.2) that would be difficult to form by a rearrangement of an ETD-produced cation-radical. The broad bands at l > 550 nm in the action spectrum had analogical absorption bands in the spectra of **9a, 9b, 10, 11**, and **12**, although those were less intense. We observed previously that the calculated intensities of long-wavelength bands of DNA radicals were very sensitive to the vibrational states and often displayed a substantial increase due to vibronic transitions[61]. In summary, none of the calculated TDDFT absorption spectra showed a match with the action spectrum that would be comparable to results obtained for other DNA radicals[13 14].

The radical energetics and correlation with precursors and intermediates suggested ions **9a, 9b** as the most likely species produced by ETD. A notable feature of their chemistry was the facile loss of water forming the *m/z* 500 ion (Figure 5.1a and b). We used this fragment ion and its possible structure relationship with **9a, 9b** to characterize the $(dCC + 2H)^{+\bullet}$ ions. The UV-Vis action spectrum of the $(dCC - H_2O + 2H)^{+\bullet}$ ion displayed bands at 220, 270, 330, 430, and 530 nm (Figure 5.3a) that were composed of the photofragment ion intensities at *m/z* 308 and 292 (Figure

5.1c). The action spectrum showed a close match with the TD-DFT absorption spectrum of a cation radical produced by loss of the 5′-OH group with cyclization onto the 5′-cytosine ring, forming a cyclocytidine structure (**17**, Scheme 5.2, Figure 5.3b). In contrast, the absorption spectrum of the 3′/5′ cyclized isomer (**18**) was distinctly different and did not exhibit matching features with the action spectrum of the $(dCC - H_2O + 2H)^{+\bullet}$ ion. The reactant (**9a**) and product (**17**) displayed different electron distributions in the 3′- and 5′-cytosine rings. Although we did not study the details of the ring formation and the pertinent transition state, the development of the electron density in **17** reflected the ground electronic state that favored the charged 5′-cyclocytosine and a radical 3′-cytosine ring. The close match of the absorption spectrum of **17** and its reaction correspondence with structure **9a** and its conformers provided corroborative evidence for the $(dCC + 2H)^{+\bullet}$ ions having a 5′-N-3-H$^{\bullet}$, 3′-N-3-H^{+} arrangement of the charged and radical cytosine rings.

5.4 CONCLUSIONS

The experimental and computational data reported here allowed us to arrive at the following conclusions. Stable hydrogen-rich deoxycytidine dinucleotide cation radicals can be generated in the gas phase by electron transfer dissociation of doubly charged dCC-crown-ether complexes. Upon activation, the cation radicals undergo dissociations involving radical losses. Photodissociation in the 210-700 nm wavelength range yielded action spectra that were interpreted after extensive modeling of precursor dication and cation-radical structures. Although the experimental and calculated spectra did not achieve a flawless match, corroborating evidence from fragment ion action spectra and energy analysis led us to conclude that the $(dCC + 2H)^{+\bullet}$ ions had the N-3-H tautomeric pattern in both the protonated and H-atom adducted cytosine rings. The 5′-N-3-H$^{\bullet}$, 3′-N-3-H^{+} distribution of the cytosine radical and charge sites was preferred energetically,

but could not be resolved by UV-vis action spectroscopy alone because of the similarity of the electronic transitions in isomers differing in the radical and charge position. [71]

5.5 ACKNOWLEDGEMENTS

Financial support was provided by the Chemistry Division of the National Science Foundation, Grants CHE-1661815 for experiments and CHE-1624430 for computations. F.T. acknowledges the Klaus and Mary Ann Saegebarth Endowment for general support.

5.6 REFERENCES

1. Boudaiffa, B.; Cloutier, P.; Hunting, D.; Huels, M. A.; Sanche, L., *Science* **2000,** *287* (5458), 1658-1660.
2. Aflatooni, K.; Gallup, G. A.; Burrow, P. D., *Journal of Physical Chemistry A* **1998,** *102* (31), 6205-6207.
3. Jalbout, A. F.; Adamowicz, L., Electron attachment to DNA base complexes. In *Advances in Quantum Chemistry, Vol 52*, 2007; Vol. 52, pp 231-+.
4. Myers, L. S.; Hollis, M. L.; Theard, L. M., *Advances in Chemistry Series* **1968,** (81), 345-&.
5. Nucifora, G.; Smaller, B.; Avery, E. C.; Remko, R., *Radiation Research* **1972,** *49* (1), 96-&.
6. Hildenbrand, K.; Schultefrohlinde, D., *Free Radical Research Communications* **1990,** *11* (4-5), 195-206.
7. Wolken, J. K.; Syrstad, E. A.; Vivekananda, S.; Turecek, F., *Journal of the American Chemical Society* **2001,** *123* (24), 5804-5805.
8. Syrstad, E. A.; Vivekananda, S.; Turecek, F., *Journal of Physical Chemistry A* **2001,** *105* (36), 8339-8351.
9. Wolken, J. K.; Turecek, F., *Journal of Physical Chemistry A* **2001,** *105* (36), 8352-8360.
10. Chen, X.; Syrstad, E. A.; Nguyen, M. T.; Gerbaux, P.; Tureček, F., *The Journal of Physical Chemistry A* **2005,** *109* (36), 8121-8132.
11. Turecek, F., Computational studies of radicals relevant to nucleic acid damage. In *Advances in Quantum Chemistry, Vol 52*, 2007; Vol. 52, pp 89-120.
12. Korn, J. A.; Urban, J.; Dang, A.; Nguyen, H. T. H.; Turecek, F., *Journal of Physical Chemistry Letters* **2017,** *8* (17), 4100-4107.
13. Liu, Y.; Korn, J. A.; Dang, A.; Turecek, F., *Journal of Physical Chemistry B* **2018,** *122* (42), 9665-9680.
14. Shaffer, C. J.; Pepin, R.; Turecek, F., *J Mass Spectrom* **2015,** *50* (12), 1438-42.
15. Polfer, N. C. D., P., Laser Photodissociation and Spectroscopy of Mass Separated Biomolecular Ions. In *Lecture Notes in Chemistry*, Springer: Cham, 2013; Vol. 83, pp 13-20.
16. Antoine, R. D., P., UV−Visible Activation of Biomolecular Ions. In *Laser Photodissociation and Spectroscopy of Mass-separated Biomolecular Ions*, Polfer, N., Dugourd, P., Ed. Springer: Cham, 2013; Vol. 83.
17. Kostko, O.; Bravaya, K.; Krylov, A.; Ahmed, M., *Physical Chemistry Chemical Physics* **2010,** *12* (12), 2860-2872.
18. Chen, Z.; Lau, K.-C.; Garcia, G. A.; Nahon, L.; Bozanic, D. K.; Poisson, L.; Al-Mogren, M. M.; Schwell, M.; Francisco, J. S.; Bellili, A.; Hochlaf, M., *Journal of the American Chemical Society* **2016,** *138* (51), 16596-16599.
19. Alonso, J. L.; Vaquero, V.; Pena, I.; Lopez, J. C.; Mata, S.; Caminati, W., *Angewandte Chemie-International Edition* **2013,** *52* (8), 2331-2334.
20. Scanlan, M. J.; Hillier, I. H., *Journal of the Chemical Society-Chemical Communications* **1984,** (2), 102-103.
21. Kwiatkowski, J. S.; Bartlett, R. J.; Person, W. B., *Journal of the American Chemical Society* **1988,** *110* (8), 2353-2358.
22. Gould, I. R.; Green, D. V. S.; Young, P.; Hillier, I. H., *Journal of Organic Chemistry* **1992,** *57* (16), 4434-4437.
23. Kobayashi, R., *Journal of Physical Chemistry A* **1998,** *102* (52), 10813-10817.
24. Trygubenko, S. A.; Bogdan, T. V.; Rueda, M.; Orozco, M.; Luque, F. J.; Sponer, J.; Slavicek, P.; Hobza, P., *Physical Chemistry Chemical Physics* **2002,** *4* (17), 4192-4203.
25. Lesslie, M.; Lawler, J. T.; Dang, A.; Korn, J. A.; Bím, D.; Steinmetz, V.; Maître, P.; Tureček, F.; Ryzhov, V., *ChemPhysChem* **2017,** *18* (10), 1293-1301.
26. Wolken, J. K.; Yao, C.; Turecek, F.; Polce, M. J.; Wesdemiotis, C., *International Journal of Mass Spectrometry* **2007,** *267* (1-3), 30-42.
27. Filippi, A.; Fraschetti, C.; Rondino, F.; Piccirillo, S.; Steinmetz, V.; Guidoni, L.; Speranza, M., *International Journal of Mass Spectrometry* **2013,** *354*, 54-61.
28. Salpin, J.-Y.; Guillaumont, S.; Tortajada, J.; MacAleese, L.; Lemaire, J.; Maitre, P., *Chemphyschem* **2007,** *8* (15), 2235-2244.
29. Bakker, J. M.; Salpin, J.-Y.; Maitre, P., *International Journal of Mass Spectrometry* **2009,** *283* (1-3), 214-221.
30. Broquier, M.; Soorkia, S.; Pino, G.; Dedonder-Lardeux, C.; Jouvet, C.; Gregoire, G., *Journal of Physical Chemistry A* **2017,** *121* (34), 6429-6439.
31. Wu, R. R.; Yang, B.; Frieler, C. E.; Berden, G.; Oomens, J.; Rodgers, M. T., *Journal of Physical Chemistry B* **2015,** *119* (18), 5773-5784.

32. Wu, R. R.; Hamlow, L. A.; He, C. C.; Nei, Y. W.; Berden, G.; Oomens, J.; Rodgers, M. T., *Journal of the American Society for Mass Spectrometry* **2017,** *28* (8), 1638-1646.
33. Yao, C.; Turecek, F.; Polce, M. J.; Wesdemiotis, C., *International Journal of Mass Spectrometry* **2007,** *265* (2-3), 106-123.
34. Yao, C. X.; Cuadrado-Peinado, M. L.; Polasek, M.; Turecek, F., *Angewandte Chemie-International Edition* **2005,** *44* (41), 6708-6711.
35. Turecek, F.; Yao, C. X., *Journal of Physical Chemistry A* **2003,** *107* (43), 9221-9231.
36. Zhang, J. D.; Xie, Y.; Schaefer, H. F., III; Luo, Q.; Li, Q. S., *Molecular Physics* **2006,** *104* (13-14), 2347-2366.
37. Tomic, K.; Tatchen, J.; Marian, C. M., *Journal of Physical Chemistry A* **2005,** *109* (37), 8410-8418.
38. Bazso, G.; Tarczay, G.; Fogarasi, G.; Szalay, P. G., *Physical Chemistry Chemical Physics* **2011,** *13* (15), 6799-6807.
39. Kosma, K.; Schroeter, C.; Samoylova, E.; Hertel, I. V.; Schultz, T., *Journal of the American Chemical Society* **2009,** *131* (46), 16939-16943.
40. Ho, J.-W.; Yen, H.-C.; Chou, W.-K.; Weng, C.-N.; Cheng, L.-H.; Shi, H.-Q.; Lai, S.-H.; Cheng, P.-Y., *Journal of Physical Chemistry A* **2011,** *115* (30), 8406-8418.
41. Domingo, A.; Rodriguez-Fortea, A.; de Graaf, C., *Journal of Chemical Theory and Computation* **2012,** *8* (1), 235-244.
42. Lobsiger, S.; T'Rachsel, M. A.; Frey, H.-M.; Leutwyler, S., *Journal of Physical Chemistry B* **2013,** *117* (20), 6106-6115.
43. Nakayama, A.; Harabuchi, Y.; Yamazaki, S.; Taketsugu, T., *Physical Chemistry Chemical Physics* **2013,** *15* (29), 12322-12339.
44. Triandafillou, C. G.; Matsika, S., *Journal of Physical Chemistry A* **2013,** *117* (46), 12165-12174.
45. Martens, J.; Berden, G.; Gebhardt, C. R.; Oomens, J., *Review of Scientific Instruments* **2016,** *87* (10).
46. Dang, A.; Korn, J. A.; Gladden, J.; Mozzone, B.; Tureček, F., *Journal of The American Society for Mass Spectrometry* **2019,** *30* (9), 1558-1564.
47. Stewart, J. J. P., *Journal of Molecular Modeling* **2007,** *13* (12), 1173-1213.
48. Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P., *Journal of Chemical Theory and Computation* **2009,** *5* (7), 1749-1760.
49. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., *Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.
50. Stewart, J. J. P. *MOPAC 16.*, Stewart Computational Chemistry, Colorado Springs, CO, 2016.
51. Řezáč, J. *Cuby—ruby framework for computational chemistry*, 4.
52. Becke, A. D., *Physical Review A* **1988,** *38* (6), 3098-3100.
53. Chai, J.-D.; Head-Gordon, M., *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615-6620.
54. Zhao, Y.; Truhlar, D. G., *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.
55. Moller, C.; Plesset, M. S., *Physical Review* **1934,** *46* (7), 0618-0622.
56. Schlegel, H. B., *Journal of Chemical Physics* **1986,** *84* (8), 4530-4534.
57. Mayer, I., *International Journal of Quantum Chemistry* **1978,** *14* (1), 29-38.
58. Knowles, P. J.; Andrews, J. S.; Amos, R. D.; Handy, N. C.; Pople, J. A., *Chemical Physics Letters* **1991,** *186* (2-3), 130-136.
59. Tomasi, J.; Mennucci, B.; Cammi, R., *Chemical Reviews* **2005,** *105* (8), 2999-3093.
60. Furche, F.; Ahlrichs, R., *Journal of Chemical Physics* **2002,** *117* (16), 7433-7447.
61. Dang, A.; Liu, Y.; Turecek, F., *Journal of Physical Chemistry A* **2019,** *123* (15), 3272-3284.
62. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. A.01*, Wallingford, CT, 2016.
63. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., *Proceedings of the National Academy of Sciences* **2004,** *101* (26), 9528-9533.
64. Murray, K. K., *Journal of Mass Spectrometry* **1996,** *31* (11), 1203-1215.
65. McLuckey, S. A.; Berkel, G. J.; Glish, G. L., *Journal of the American Society for Mass Spectrometry* **1992,** *3* (1), 60-70.
66. 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.; Turecek, F., *International Journal of Mass Spectrometry* **2008,** *276* (2-3), 116-126.
67. Michl, J., *Journal of Molecular Spectroscopy* **1969,** *30* (1), 66-&.
68. Turecek, F.; Julian, R. R., *Chemical Reviews* **2013,** *113* (8), 6691-6733.
69. Pepin, R.; Turecek, F., *Journal of Physical Chemistry B* **2015,** *119* (7), 2818-2826.
70. Goeringer, D. E.; McLuckey, S. A., *International Journal of Mass Spectrometry* **1998,** *177* (2-3), 163-174.
71. Liu, Y.; Korn, J. A.; Tureček, F., *International Journal of Mass Spectrometry* **2019,** *443*, 22-31.

Chapter 6. GAS-PHASE PHOTODISSOCIATIVE CROSSLINKING OF DIAZIRINE-TAGGED ADRENALINE WITH BINDING MOTIFS FROM BETA-2 ADRENERGIC RECEPTOR

Abstract

Gas-phase studies of noncovalent complexes in combination with Born–Oppenheimer molecular dynamics provide useful data for evaluating and elucidating intermolecular binding in biological systems. We have been developing gas-phase footprinting methods that utilize reactive carbenes, generated from diazirine-tagged molecules via photodissociation at 355 nm, to rapidly cross-link to proximate bonds, enabling to capture conformational changes in dynamic systems. Membrane proteins, such as β-2-adrenoceptor (β_2AR), bind native endogenous agonists in a dynamic fashion. Here, we report a photo-crosslinking study using diazirine-labeled chiral D- and L-adrenalines as ligands to elucidate noncovalent bonding of agonists to β_2AR binding motifs. We chose two peptide sequences SSIVSFY and VYILLNWIGY, as model peptides representing the active binding site of β_2AR to study the different interactions with proper and non-proper ligands. The experimental results were further interpreted by Born-Oppenheimer molecular dynamics (BOMD) computer modeling to elucidate the binding dynamics.

6.1 INTRODUCTION

Molecular recognition, referring to the specific interactions within biomolecular complexes such as the ligand-receptor, antigen-antibodies, enzymes-substrates etc., plays an important role in biological systems. Among these essential biological processes, neural transmission, mediated by neurotransmitters, carries out the communication throughout the body. During the communication

processes, neurotransmitters are trafficked between neurons while being recognized by selective receptors to allow passing along specific signals. The binding interactions of neurotransmitters and their receptor are sometimes dynamic, resulting in different conformational states. [1] The G-protein-coupled receptors (GPCRs), being one example, are integral membrane proteins that can bind different ligands and have shown dynamic behavior. Thanks to the recent progress in structural biology, a few conformational states have been recognized[2] [3] [4] but little was known about the dynamics and the stability of the complexes. The β_2-adrenergic receptor (β_2AR) is one of the most exhaustively studied model systems[1] [5] [6] for the investigation of the GPCR protein dynamics in signaling. Nevertheless, mechanisms of selective recognition among different ligands remain poorly understood, owing to the difficulty of characterizing weakly bound complexes with conventional "imaging" methods such as X-ray diffraction[7] and NMR[8]. These studies have indicated a partial peptide sequence that is directly involved with agonist recognition, providing a simplified access point of a dynamic binding and modeling study. The peptide sequence is expected to be located in the low dielectric environment (membrane), making it suitable for a binding study in the gas phase. Recent advances on this topic have been reported by Ishiuchi *et al.*, and Sekiguchi *et al.*, who utilized mass spectrometry and laser spectroscopy to characterize multiple complexes of a peptide and various ligands. [9] [10] These studies, that utilized vibronic bands assigned to hypothetical transitions, have indicated that binding motif recognizes proper ligands due to the differences in the catechol side chain.

Here we report a different approach using photodissociative crosslinking in the gas phase to map the noncovalent interactions of the partial binding motif with different. We have successfully implemented similar methodologies to study site-specific peptide-peptide[11] [12] [13], and peptide-nucleotide[14] binding, as model systems representing protein-protein and protein-DNA

interactions. The agonists studied in this work are (L)-adrenaline, (D)-adrenaline and dopamine, on the hypothesis that they might have different binding affinity to β_2AR and result in different activity. [15, 7] By introducing a position tunable diazirine crosslinker to the agonist (i.e. adrenaline), a highly reactive carbene intermediate is generated by photon activation that can rapidly insert to any proximate X-H (X = C, O, N) bond on the target peptide. This competes with exothermic isomerization of the carbene to an unreactive olefin, providing a kinetic constraint for the crosslinking to capture dynamic conformations. We hope to further analyze the crosslinked product by CID tandem mass spectrometry (CID-MS^3) to probe the location of crosslinks. We conduct Born-Oppenheimer molecular dynamics (BOMD) calculations at experimental temperature to investigate the mechanism of transient binding, under the distance constraints in the 3D space provided by the crosslinking experiments.

6.2 Experiments and Calculations

6.2.1 *Materials and Methods*

Singly tagged photo-agonists (L)*-adrenaline, (D)*-adrenaline, Dopamine* and doubly tagged Dopamine** (Figure 6.1) were synthesized in lab, using a position-tunable diazirine tag [16].

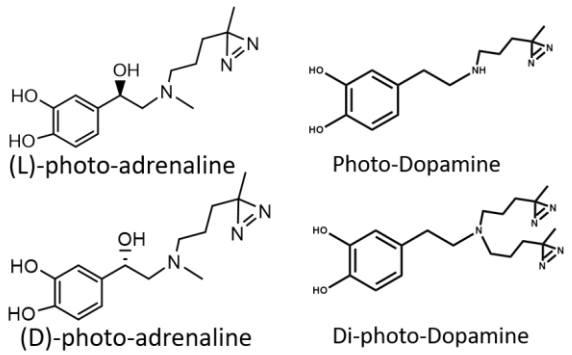


Figure 6.1. The chemical sturcture of photo-agonists: (L)*-adrenaline, (D)*-adrenaline, Dopamine* and Dopamine**.

The synthesis route of photo-adrenaline is summarized in (Figure 6.2): In a tinted glass vial, 3mL mix solvent system (3% DMSO in DMF), 4.7 µmol (1 eq.) of (L)-adrenaline (from Sigma-Aldrich) or (D)-adrenaline (from Toronto Research Chemicals) was dissolved completely; 5.6 µmol (1.2 eq.) 4,4-diazirine-1-iodopentane was added; the reaction mixture was purged with argon as protecting gas and the container was capped and wrapped with aluminum foil to avoid exposure to light. The reaction vessel was incubated in a shaker at 37C for 48h, forming an iodide salt. The photo-dopamine synthesis utilized a dopamine hydrochloride (from Sigma-Aldrich, 1 eq.) salt as reactant, dissolved in DMF. K_2CO_3 base was introduced to strip the HCl, and 4,4-diazirine-1-iodopentane (1.2 eq.) was added to the solution and was left to react for 48 h.

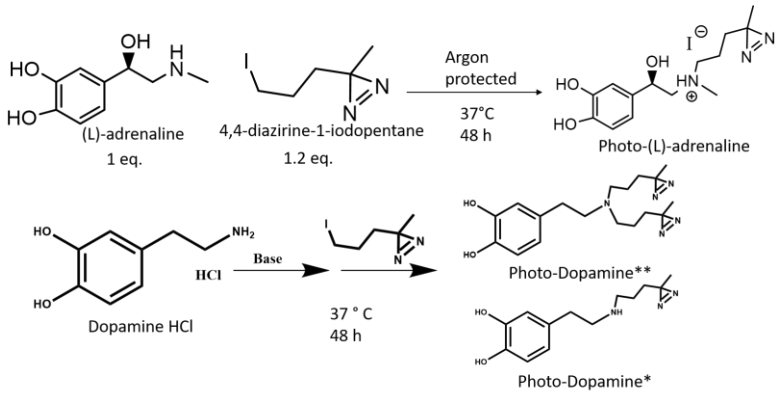


Figure 6.2. The schematics of photo-agonists synthesis.

Two N-acetylated, C-amidated peptides with sequences $S_{(203)}$SIVSFY$_{(209)}$ and $V_{(307)}$YILLNWIGY$_{(317)}$ were custom-synthesized by Genscript to represent the two sequences that were suggested to be related to ligand binding in previous study[7]. The peptide-ligand complex mixture was made by diluting the photo-agonist reaction mixture (~1mM) and the peptide stock solution (mM concentration) by a spray solvent (acetonitrile: water = 1:1 with 1% acetic acid) into final molar concentration ca. 50 µM: 50 µM. Mass spectra were measured on a Bruker amaZon Speed (Bruker, Bremen, Germany) ion trap mass spectrometer (MS) equipped with an Nd-YAG

EKSPLA laser. [17] [18] Complex ions produced by electrospray ionization were mass-selected and stored in the ion trap, followed by UV photo-dissociative crosslinking and analyzed by MS^3-CID.

Condensed-phase photolysis was carried out as control experiment. Samples were prepared as described above and then placed under a UV-lamp (365nm) for 4~6 hours. The product mixture was electrosprayed into MS and analyzed by MS^2-CID.

6.2.2 *Calculations*

Born-Oppenheimer molecular dynamics (BOMD) trajectories were run with semiempirical all-valence-electron quantum chemistry calculations using the Berendsen thermostat algorithm[19]. For the protonated (L)*-adrenaline and N-acetyl-SSIVSFY-amidated complexes, 12 initial structures were constructed in which the ligand and the peptide were in different spatial orientations. These complexes were subjected to a preliminary BOMD using PM6[20] that was augmented by including dispersion interaction and hydrogen bonding interactions (D3H4 [21]). These calculations were run by MOPAC[22] that was coupled to the Cuby4 framework.[23,24] Temperature was elevated to 410 K to enable a broader sampling of the total conformation space and generating more conformations. Running trajectories with 1 fs steps for 20 ps furnished 20,000 snapshots per each initial structure from which 200 snapshots were extracted at 100 fs intervals. The extracted snapshot structures were fully gradient-optimized with PM6-D3H4 and sorted out by their secondary and supersecondary structural similarities to compact duplicates and reduce the size of the selection. This yielded 60–70 distinct structures whose geometries were fully optimized by DFT gradient calculations. B3LYP/6-31G(d,p)[25] optimized structures were used for frequency calculations to provide vibrational enthalpies and entropies whereas ωB97X-D/6–31 + G(d,p)[26] optimized structures were used to evaluate electronic energies including dispersion interactions. The electronic, vibrational, and rotational terms were combined to yield relative free energies for

the gas-phase complexes. In addition, solvent effects were included by single-point energy self-consistent reaction field calculations using the polarizable continuum model[27] with water as the dielectric and the gas-phase optimized structures. All the thermochemical calculations were performed using the Gaussian 16 (Revision A.03) [28] suite of programs. The final optimized structures with relative free energies within 40 kJ mol^{-1} of the global energy minimum were then subjected to a full BOMD with PM6-D3H4 for 100 ps at 310 K to represent the experimental temperature. Cuby4 was also used to extract close contacts between the carbon atom of the diazirine ring and the dinucleotide atoms. Close contacts were limited to X–H…C distance of less than 4 Å on the basis of the van der Waals radii of the diazirine and X–H atoms [29]. Full conformation search and energy minimization at same level of theory was conducted for monomer photo-ligand and peptide to calculate binding energy (with counterpoise corrections). Theoretical collision cross section (Ω_{th}) analysis were calculated by MOBCAL [30] program.

6.3 Results and Discussion

6.3.1 *Synthetic Photo-agonist Characterization*

The (L)*-adrenaline was successfully synthesized and characterized by ESI-MS/MS. The ESI mass spectrum showed the existence of unreacted protonated (L)-adrenaline (*m/z* = 184) that is denoted as $(M+H)^+$ and its dehydrated ion $(M+H-H_2O)^+$ in light orange, the protonated (L)*-adrenaline $(M^*+H)^+$ (*m/z* = 280) and some other related fragments in red. (Figure 6.3 left)

The ESI spectrum of synthetic (D)*-adrenaline showed a similar distribution of reactants and product with slightly lower yield of the product (Figure 6.3 right).

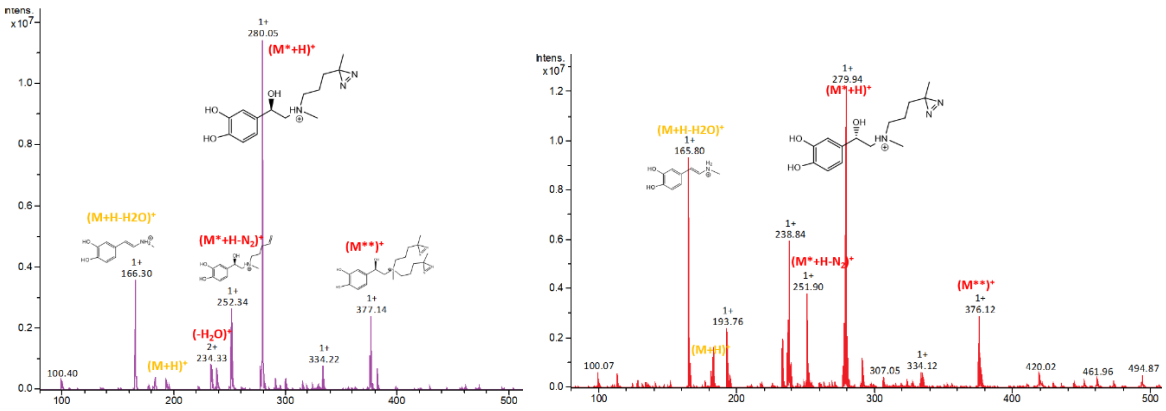


Figure 6.3. ESI spectra of the synthetic (L)*-adrenaline (left) and (D)*-adrenaline (right).

The chemical structure of the photo-agonist was verified by MS^2-CID that showed fragment ions indicating that the derivatization occurred on secondary amine. The (D)* (Figure 6.4a) and (L)* (Figure 6.4b) showed the same fragment series and relative abundancies upon collision induced dissociation at same activation energy and activation time.

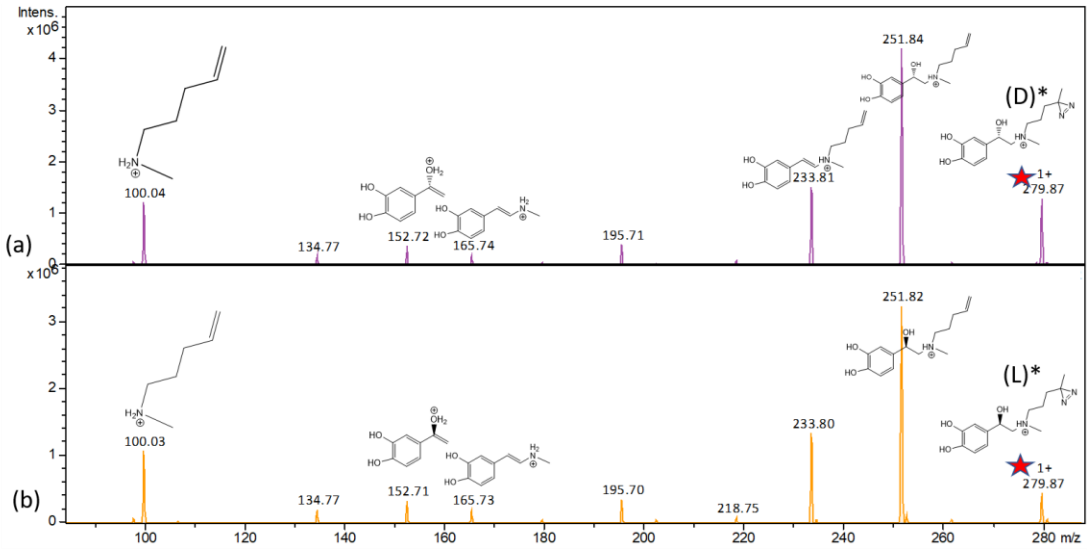


Figure 6.4. The MS^2-CID spectra of isolated (a) (D)*-adrenaline and (b) (L)*-adrenaline

The UV Photodissociation were also bench marked by radiating the isolated photo-adrenaline ions $(M^*+H)^+$ at 355 nm in the ion trap. The (D)* (Figure 6.5a) and (L)* (Figure 6.5b) showed the same fragment series. Interestingly, the product due to loss of N_2 (*m/z* 252) was not

observed, possibly because of the secondary dissociations. Therefore, only the most stable ion *m/z* = 100 was observed. A wavelength dependent dissociation experiment was also carried out to track the production of the *m/z* 100 photofragment ion, and the highest yield (after wavelength photon energy correction) was observed at in the 340~355 nm region, which was consistent with the diazirine absorption. To check for the formation of a diazo isomer, wavelength-dependent experiment was also conducted at 500 nm -600 nm region[31] but no precursor ion depletion was observed at those wavelengths.

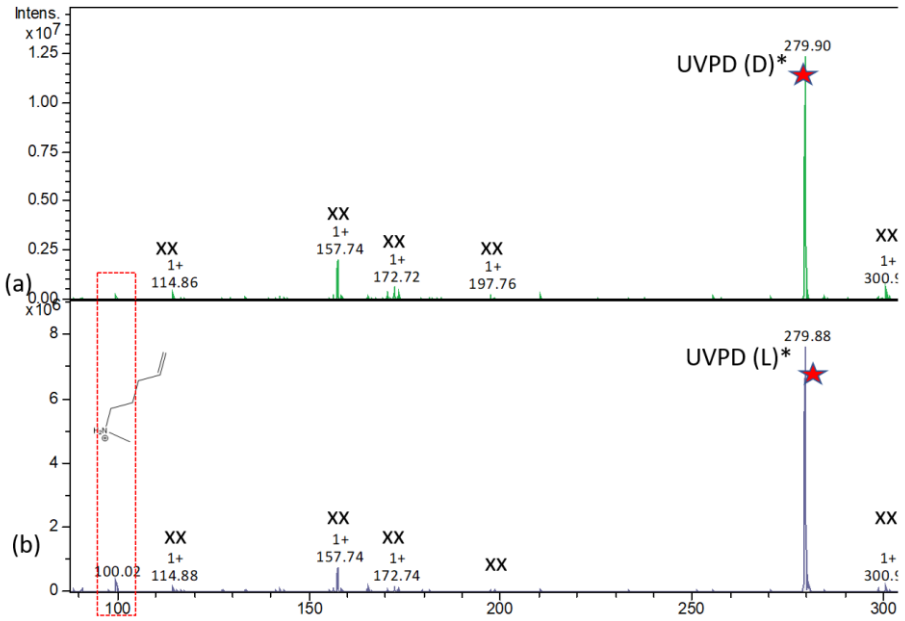


Figure 6.5. The MS^2-UVPD spectra of isolated (a) (D)*-adrenaline and (b) (L)*-adrenaline. The ions denoted as xx are background peaks from UVPD that are confirmed from blank control.

The photo-dopamine synthesis resulted in doubly alkylated product Dopamine** in a high yield (Figure 6.6a). The MS^2-CID spectrum showed that the alkylation occurred at the amino chain (Figure 6.6b). The doubly alkylated dopamine is denoted as M**.

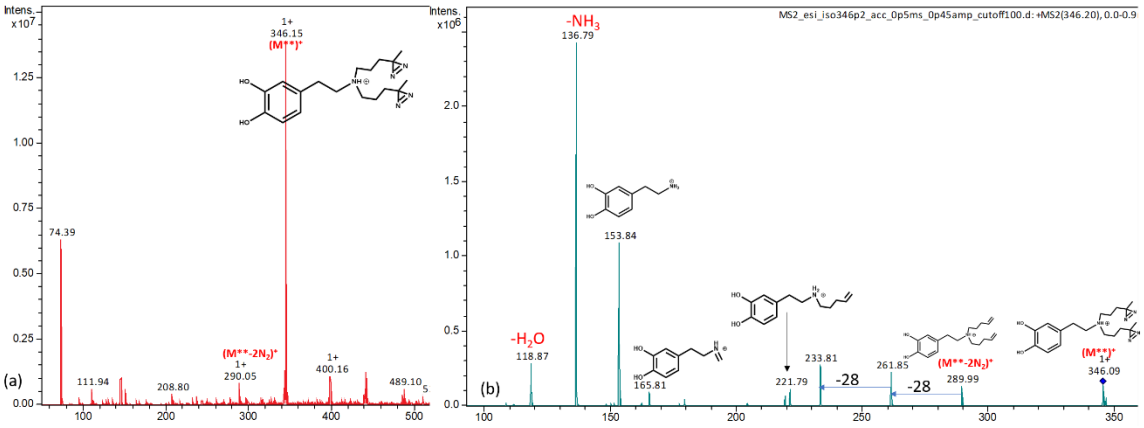


Figure 6.6. (a)The ESI spectrum of photo-dopamine synthesis mixture and (b) MS^2-CID of isolated *m/z* 346.

6.3.2 *Complex Formation and Photodissociation*

(A) Complex with SSIVSFY sequence

Electrospray ionization of equimolar mixtures of labeled photo-agonist (abbreviated as M*, labeled in red in Figure 6.7) and the target peptide (m) produced the singly charged $(M^* + m + H)^+$ dimer complexes (*m/z* 1122) in ca. 60% relative abundance for the photo-adrenaline (Figure 6.7). Note that smaller amount of unreacted adrenaline (denoted as M, in light orange) was present in the mixture and have shown to form complexes $(M + m + H)^+$ with peptide (*m/z* 1026) at a comparable yield, resulting in slightly higher abundance compared to the photo-adrenaline peptide complexes. The ESI of the (L)*-isomer $(M^* + m + H)^+$ is shown in (Figure 6.7). A very similar ion distribution was observed for the (D)*- isomer complexes where both the $(M^* + m + H)^+$ and $(M + m + H)^+$ ion complexes were present at similar ion intensities, indicating strong binding of the peptide motif with both the (D)- and (L)- adrenaline in its natural form. The phototag derivatization did not result in altering the binding affinity.

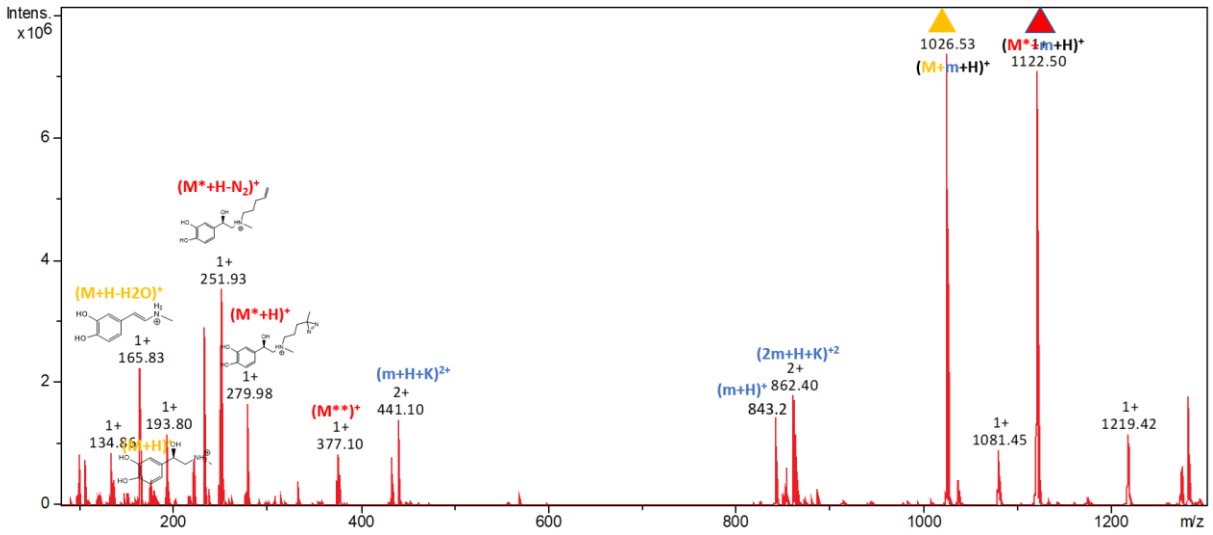


Figure 6.7. The ESI spectrum of equimolar mixture of (L)*-adrenaline and N-acetylated $SSIVSFY_{(NH2)}$ complexes.

The complex formation was investigated by MS^2-CID and MS^2-UVPD. The complexes were isolated by mass, *m/z* 1122, and stored in the ion trap and subjected to several laser pulses at 355 nm. This resulted in loss of N_2, $(M^*\text{-}N_2\text{+m+H})^+$ at *m/z* 1094 and a partial dissociation of the photo-dissociative product into monomer $(M^*\text{-}N_2\text{+H})^+$ ion from photo-adrenaline at *m/z* 252. Interestingly, the complementary peptide $(m + H)^+$ ion at *m/z* 843 was not observed in the spectrum. Both the (D)- and (L)- complexes showed similar UVPD spectra (Figure 6.8) where the (D)* and (L)* label denotes the corresponding chirality of the photo-labeled compound. The peaks denoted with "xx" are background ions from UVPD. The fraction $R(MS^2)$ of surviving $(M^*\text{-}N_2\text{+m+H})^+$ ions was 64% for (D)* complex and 62% for the (L)* complex

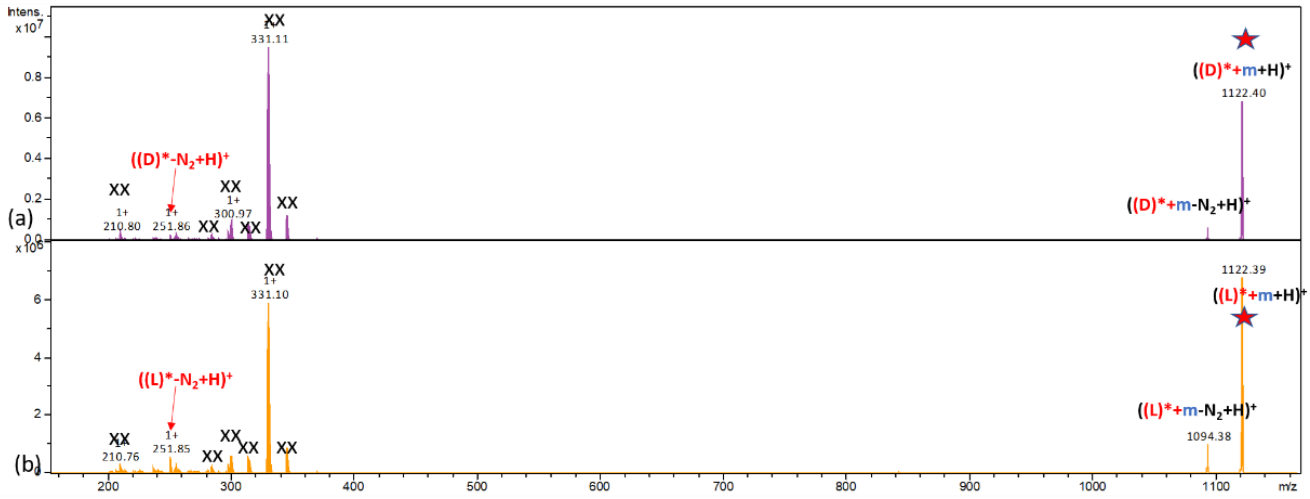


Figure 6.8. MS^2-UVPD spectra of isolated (a) (D)*-adrenaline and (b) (L)*-adrenaline. of (L)*-adrenaline and *Acetyl*-$SSIVSFY_{(NH2)}$ complexes.

CID of *m/z* 1122 (both (D)* and (L)*), however, showed different results: the peptide monomer ions $(m + H)^+$ *m/z* 843, its fragments (b_6, *m/z* 663) and their subsequent neutral loss ions ($(m + H\text{-}NH_3)^+$, *m/z* 826; $(b_6 - H_2O)^+$, *m/z* 645) are the most abundant product ions while the protonated photo-adrenaline *m/z* 280 abundance was only observed at < 5%. (Figure 6.9)

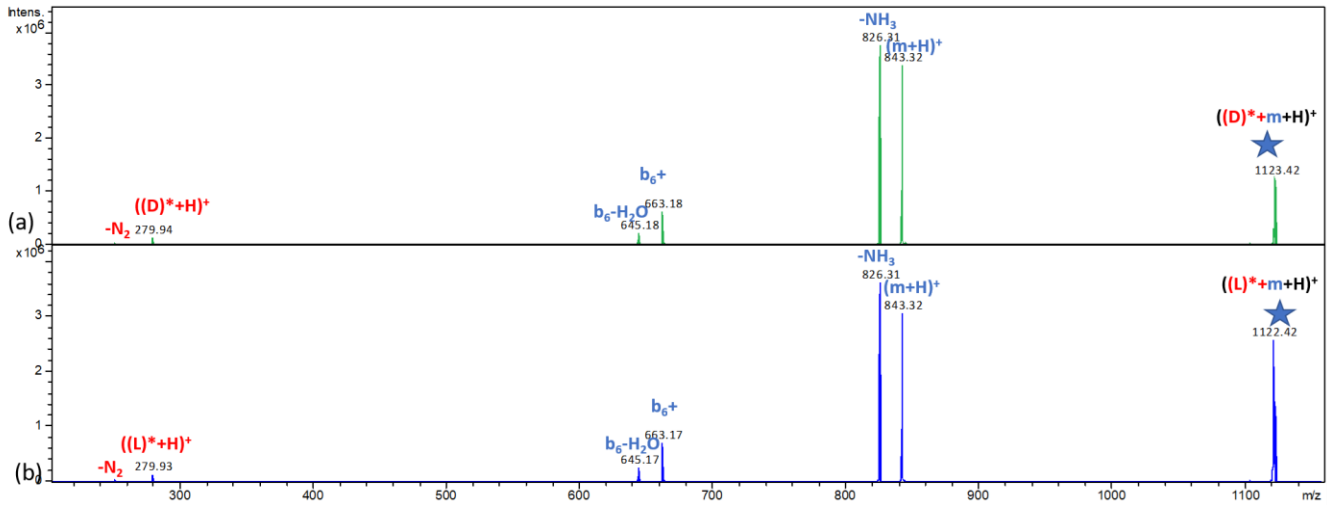


Figure 6.9. MS^2-CID spectra of isolated (a) (D)*-adrenaline and (b) (L)*-adrenaline. of (L)*-adrenaline and *Acetyl*-$SSIVSFY_{(NH2)}$ complexes.

For photo-dopamine complexes, the ESI spectra indicated a weaker binding interaction of the doubly alkylated dopamine complexes $(M^{**}+m+H)^+$ compared to that of the adrenaline complexes, regardless of the tuning of the optimized mass region. (Figure 6.10 a1 and Figure 6.10 a2). MS^2-UVPD spectra (Figure 6.10 c) showed a subsequent loss of N_2. Due to the prominent UVPD background ion contamination in the mass region 200-350 *m/z*, the formation of monomer photo-dopamine $(M^{**}+H- N_2)$ observed at *m/z* 318 could not be identified with confidence. MS^2-CID spectra (Figure 6.10 b) indicated the formation of protonated peptide ions $(m + H)^+$ and photo-dopamine ion $(M^{**}+H)^+$.

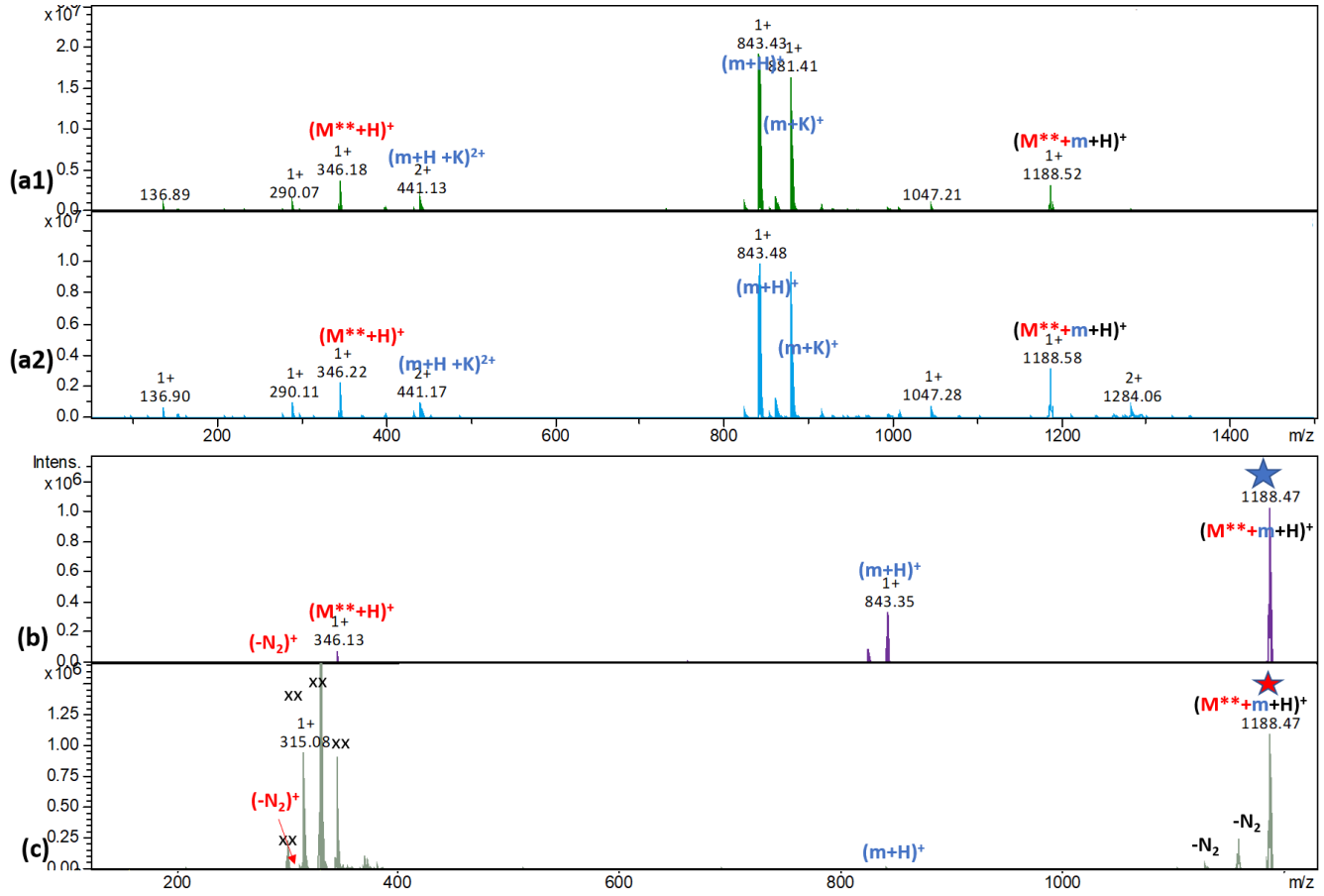


Figure 6.10. ESI mass spectra of ion complexes of photo-dopamine and *Acetyl*-SSIVSFY$_{(NH2)}$ (a1) tuned to optimize *m/z* 843, (a2) optimize to *m/z* 1189; (b) MS^2-CID spectrum of isolated photo-dopamine and peptide complex $(M^{**}+m+H)^+$; (c) MS^2-UVPD spectrum of $(M^{**}+m+H)^+$

(B) Complex with VYILLNWIGY sequence

The ESI of the (L)*-isomer $(M^* + m + H)^+$ is shown in (Figure 6.11). A very similar ion abundance distribution was observed for the (D)*- isomer complexes where both the $(M^* + m + H)^+$ and $(M + m + H)^+$ ion complexes were present, indicating noncovalent binding of the *Acetyl*-VYILLNWIGY$_{(NH2)}$ motif were stronger with the natural form of both (D)- and (L)- adrenaline when compared to their photo-derivatized products.

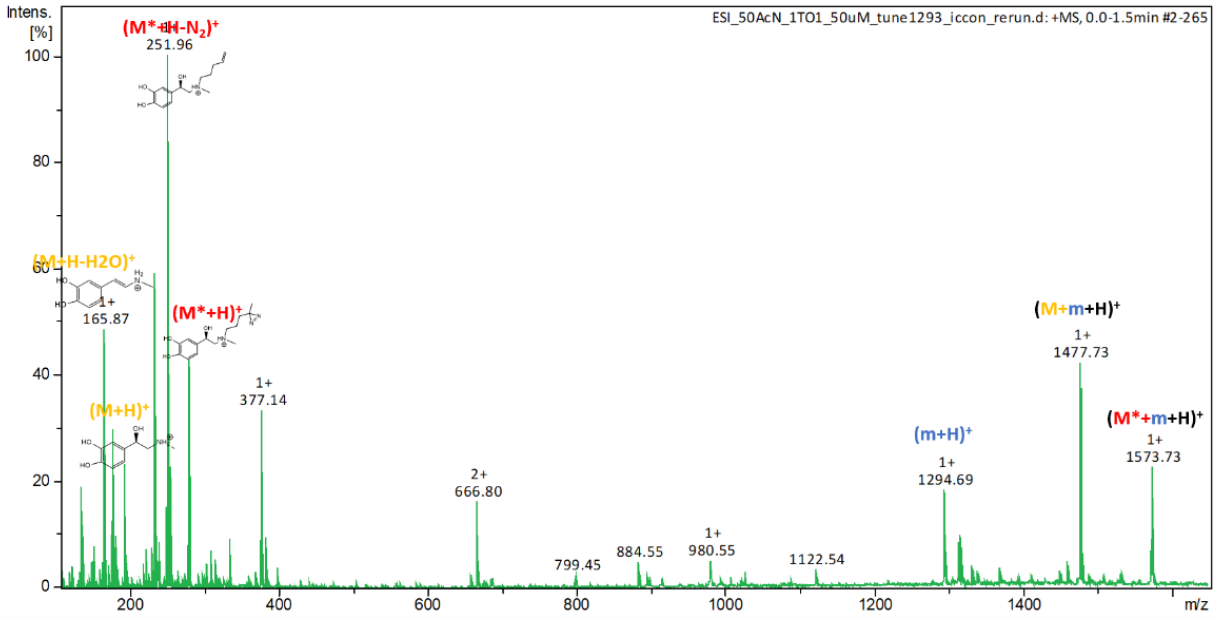


Figure 6.11. ESI spectrum of equimolar mixture of (L)*-adrenaline and *N-acetylated* VYILLNWIGY$_{(NH2)}$ complexes.

The complex formation was investigated by MS^2-CID and MS^2-UVPD. The complexes were isolated by mass, *m/z* 1573, stored in the ion trap, and subjected to several laser pulses as 355 nm. This resulted in loss of N_2 $(M^*\text{-}N_2\text{+}m\text{+}H)^+$ at *m/z* 1545 and the partial dissociation of the photo-dissociative product into the monomer peptide ions $(m + H)^+$ at *m/z* 1294 and the formation of $(M^*\text{-}N_2\text{+}H)^+$ at *m/z* 252. The low abundance of both monomer peaks indicated a high survival rate of $(M^*\text{-}N_2\text{+}m\text{+}H)^+$ upon UVPD. The R(MS^2) for (D)- and (L)- isomer complexes were similar, 73% and 78%, repectively. Similar UVPD spectra as shown in (Figure 6.12 a1, a2). CID

of *m/z* 1573 (both (D)* and (L)*), showed prominent peptide ion formation at *m/z* but the counterpart ion (protonated photo-adrenaline) was not observed (Figure 6.12 b1, b2)

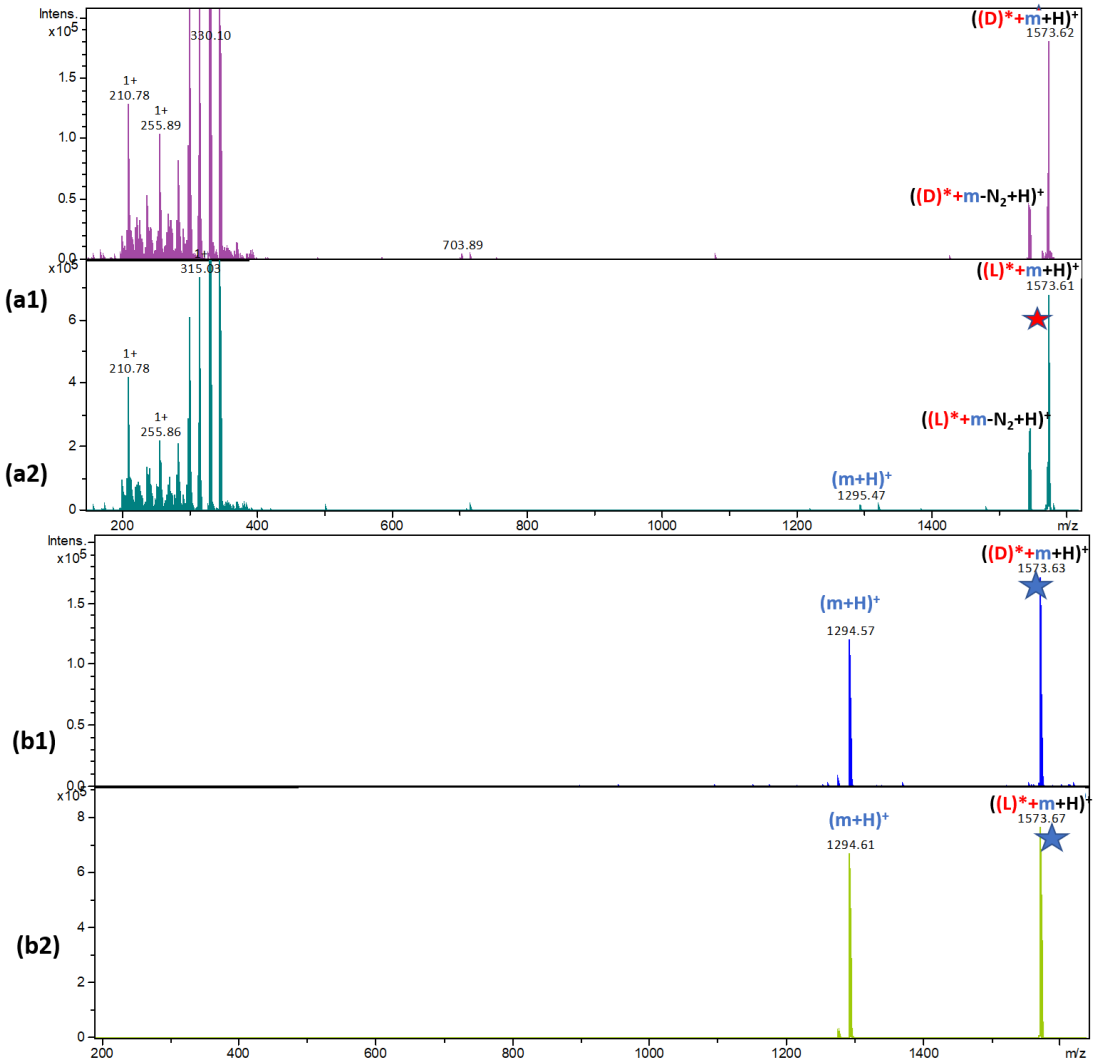


Figure 6.12. MS^2-CID spectra of isolated (a) (D)*-adrenaline and (b) (L)*-adrenaline. of (L)*-adrenaline and (L)*-adrenaline and *N-acetylated*-VYILLNWIGY$_{(NH2)}$ complexes.

For photo-dopamine complexes, the ESI spectra indicated weak binding interactions giving relative ion abundance of the doubly alkylated dopamine complexes $(M^{**}+m+H)^+$ *m/z* 1639 of ca. 5%. Due to the low abundance of dimer complex, MS^2- UVPD activation produced noisy mass spectrum with the -N_2 complex ion at *m/z* 1611, peptide ion $(m+H)^+$ at 1294, along with other un-

identified peaks. The possible product formation from M** could not be observed due to the ion trap low-mass cutoff. CID spectrum showed stronger signal of *m/z* 1294 and its neutral loss ion (*m/z* 1277).

Compared to the serine-rich sequence discussed in section (A), the ESI spectra showed that relative abundances for the singly charged ligand-peptide complex ions was much lower for the leucine-rich sequence studied in this section, among all photo-agonist studied. The result indicated weaker interaction between the ligand and the VYILLNWIGY peptide.

6.3.3 *CID-MS^3 of Photolyzed Conjugates of (L)*-adrenaline and (D)*-adrenaline*

The $(M^*\text{-}N_2+m+H)^+$ ions formed by photodissociation were selected by mass and subjected to CID. If crosslinked product was present, fragment ions containing both components were expected to be observed, with some exceptions such that the newly formed covalent bond would undergo collisional dissociation. For the complexes of *N-Acetyl*-SSIVSFY$_{(NH2)}$ with (D)*-adrenaline (Figure 6.13(a)) and *N-Acetyl*-SSIVSFY$_{(NH2)}$ with (L)*-adrenaline (Figure 6.13(b)), the MS^3-UVPD-CID spectra had shown the formation of monomeric units only, at near-identical activation energy and activation time. The spectra were quite similar in the ion abundance distribution, with only slight difference at *m/z* 1076 (loss of water from parent ion). The water loss could be accredited to different solvation effect of the benzylic alcohol of (D)- and (L)- that the hydroxyl group in the (D) isomer might not be solvated within the complex and was easily fragmented, without interrupting the binding of the dimer complex. The spectra were zoomed in 5 to 6 times (Figure 6.14 a and b) and compared to the MS^2-CID spectrum of the protonated monomeric peptide ((Figure 6.14 c). The fragment series of peptide were almost identical in the three spectra. Only *m/z* 429 (1% intensity) could be mass overlapped with potential conjugate fragment (protonated 2'-amine- fragment *m/z* 100 from ligand with peptide b_3 ion) (Figure 6.14 a

and b). Some other possible conjugated fragments are *m/z* 600 (protonated 1'-amine *m/z* 85 from ligand and b_5 ion), *m/z* 666 as $((L)^*\text{-}N_2\_y_3)^+$ ion and *m/z* 767 as $((L)^*\text{-}N_2\_b_5)^+$ ion, which are at very low relative intensities (< 1%) in Figure 6.14.

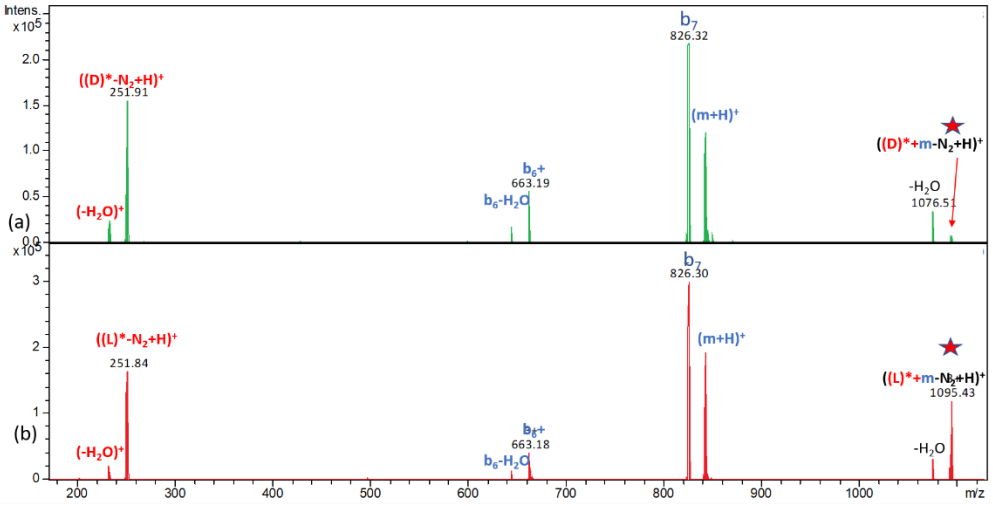


Figure 6.13. MS^3-UVPD-CID spectra of isolated photodissociative product of dimer complexes of *N-Acetyl*-$SSIVSFY_{(NH2)}$ and (a) (D)*-adrenaline and (b) (L)*-adrenaline.

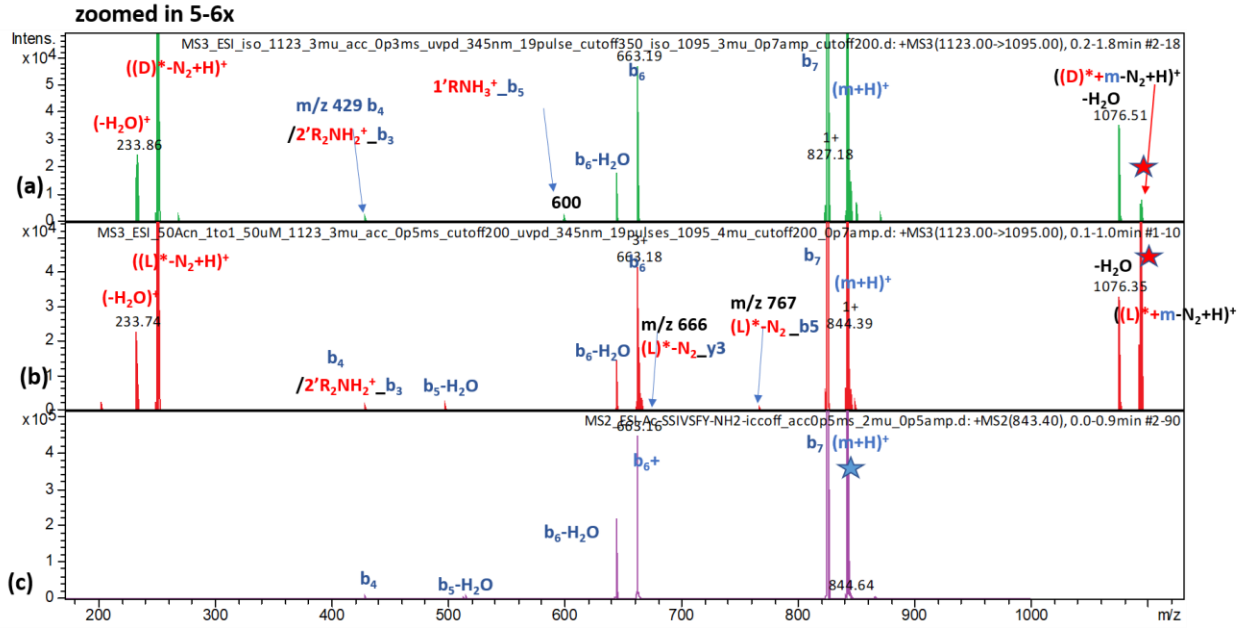


Figure 6.14. Zoomed MS^3-UVPD-CID spectra peak assignment of isolated photodissociative product of dimer complexes of (a) *N-Acetyl*-$SSIVSFY_{(NH2)}$ and (D)*-adrenaline; (b) *N- Acetyl*-$SSIVSFY_{(NH2)}$ and (L)*-adrenaline; (c) MS^2-CID of monomer peptide *N- Acetyl*-$SSIVSFY_{(NH2)}$

In the complexes of *N-Acetyl*-VYILLNWIGY$_{(NH2)}$ with (D)*-adrenaline (Figure 6.15 (a)) and *N-Acetyl*-VYILLNWIGY$_{(NH2)}$ with (L)*-adrenaline (Figure 6.15 (b)), the MS^3-UVPD-CID spectra had shown the formation of neutral loss (-NH_3) from the precursor ion, as well as peptide monomeric units, followed by the peptide fragment series. The photo-adrenaline monomer ion was below the low-mass cutoff so was not observed. The relative ion intensity of the (D)* and (L)* containing complexes were differed that the 1000 – 1115 mass region where the -NH_3 product of b_8 ion was observed for (D)*- in Figure 6.15 (a) but nonexistent in Figure 6.15 (b). Instead, b_9-NH_3 ion was observed in Figure 6.15 (b) but not in (a). *m/z* 944 could be assigned as b_7 ion or a conjugate ion consisted of 1'-amine (*m/z* 85) and y_7-NH_3. Compared to the monomer peptide fragmentation (Figure 6.15 (c)) where only consecutive ammonia loss from precursor ion, and b7, b8, b9 were observed, CID of the complex produced more fragments.

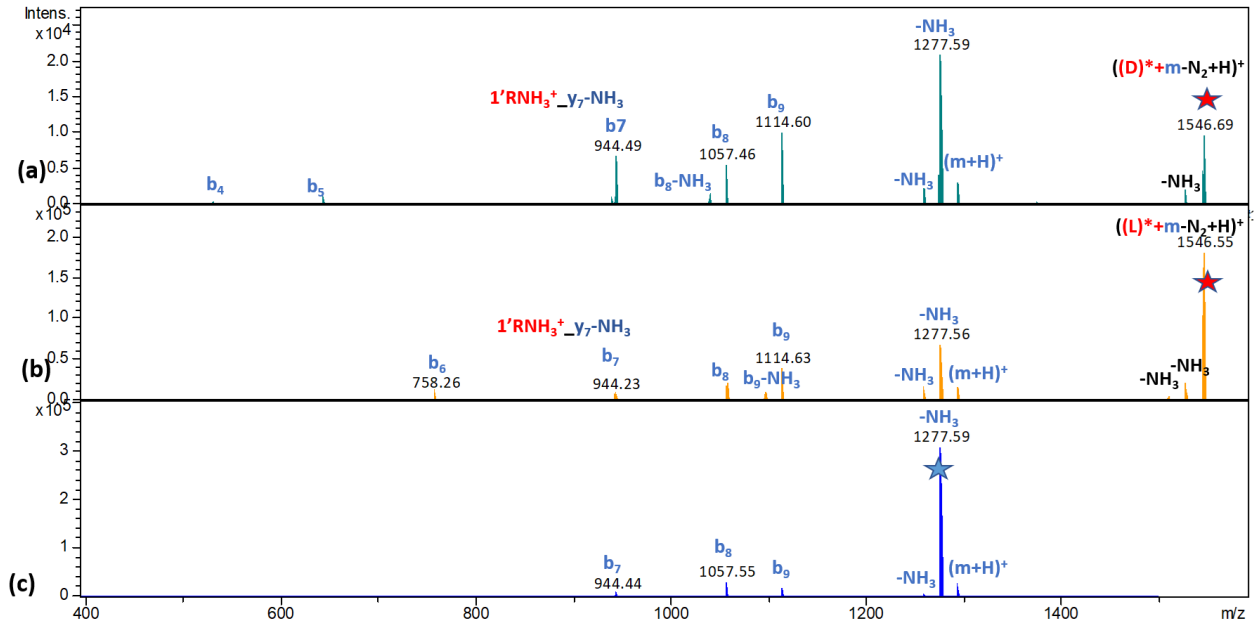


Figure 6.15. MS^3-UVPD-CID spectra peak assignment of isolated photodissociative product of dimer complexes (a) *N-Acetyl*-VYILLNWIGY$_{(NH2)}$ and (D)*-adrenaline; (b) *N-Acetyl*-VYILLNWIGY$_{(NH2)}$ and (L)*-adrenaline; (c) MS^2-CID of monomer peptide *N-Acetyl*-VYILLNWIGY$_{(NH2)}$

6.3.4 *CID-MS^3 of Photolyzed Conjugates of Dopamine***

Same experiments were carried out for photo-dopamine ligand and the two target peptides. In the complex of *N-Acetyl*-SSIVSFY$_{(NH2)}$ and doubly alkylated dopamine** (M**-N_2+m+H)$^+$, no obvious loss of water was observed close to the precursor mass, which could indicate that in the adrenaline complexes that the loss of water at *m/z* 1076 in Figure 6.14 (a) and (b) were accredited to the benzylic alcohol. The fragment series of the serine-rich peptide (Figure 6.16 a) complex was similar to that of the monomer peptide, but the relative abundance of b_7 was much lower in the dopamine complex, which could indicate the C-terminal amidation of the peptide would be coordinated by some other functional groups in the complex. Also, fragments with *m/z* below b_6 series were observed in the complex fragmentation at much higher relative intensities to the b_6 ion than the ones observed in the monomer (Figure 6.14 (c)). This could be attributed by the coordination of the ligand that promoted the formation of smaller fragments.

In the complex of *N-Acetyl*-VYILLNWIGY$_{(NH2)}$ and doubly alkylated dopamine** (M**-N_2+m+H)$^+$, CID spectrum showed loss of N_2 from the precursor. The fragment series from the peptide were similar to that from the monomer (Figure 6.15 (c)) but the relative intensity of b_9 (*m/z* 1114) ion was much higher in the complex (Figure 6.16 b). Unlike the adrenaline complexes that promoted the ammonia loss from b_8 or b_9 ions, the dopamine complexes did not show these series, indicating different binding interactions.

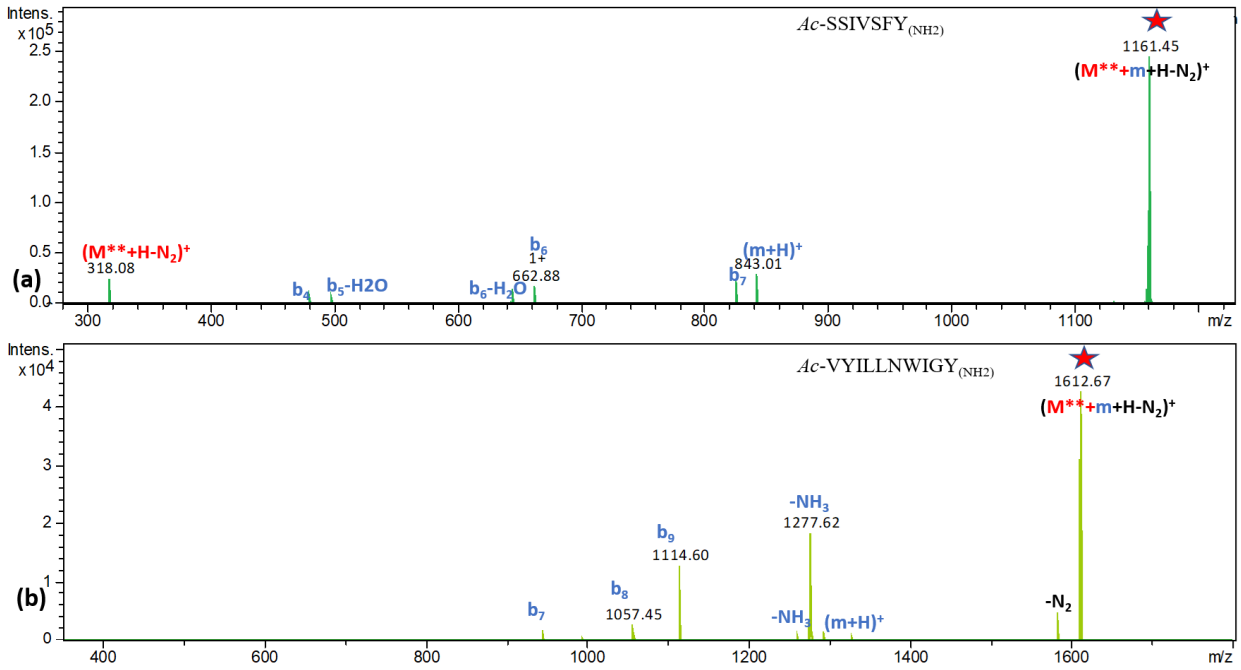


Figure 6.16. MS^3-UVPD-CID spectra peak assignment of isolated photodissociative product of dimer complexes (a) *N-Acetyl*-SSIVSFY$_{(NH2)}$ and dopamine**; (b) *N-Acetyl*-VYILLNWIGY$_{(NH2)}$ and dopamine**

6.3.5 *Condensed-phase photolysis analysis of (L)*-adrenaline Complex*

To further analyze the photodissociative cross-linking product, condensed phase photolysis experiment was conducted for (L)*-adrenaline and *N-Acetyl*-SSIVSFY$_{(NH2)}$ complex at low concentration (50 µM) as control experiment, in hope of depleting the photo-adrenaline into the photolysis product with loss of N_2. At such low concentration, the analyte molecules are surrounding by solvent molecules and are not expected to have close interactions, therefore little or no crosslinking would be produced by the UV photolysis. Once the photolysis was done, the mixture was analyzed by electrospray ionization (ESI). As observed in the ESI, the *m/z* 280 was depleted into forming the $(M^*\text{-}N_2\text{+}H)^+$ product (Figure 6.17 b). The corresponding non-covalent complexes $(M^*\text{-}N_2\text{+}m\text{+}H)^+$ was formed at *m/z* 1094 in Figure 6.17b but was not observed in Figure 6.17 a. The $(M^*\text{-}N_2\text{+}m\text{+}H)^+$ ion was isolated and subjected to CID analysis.

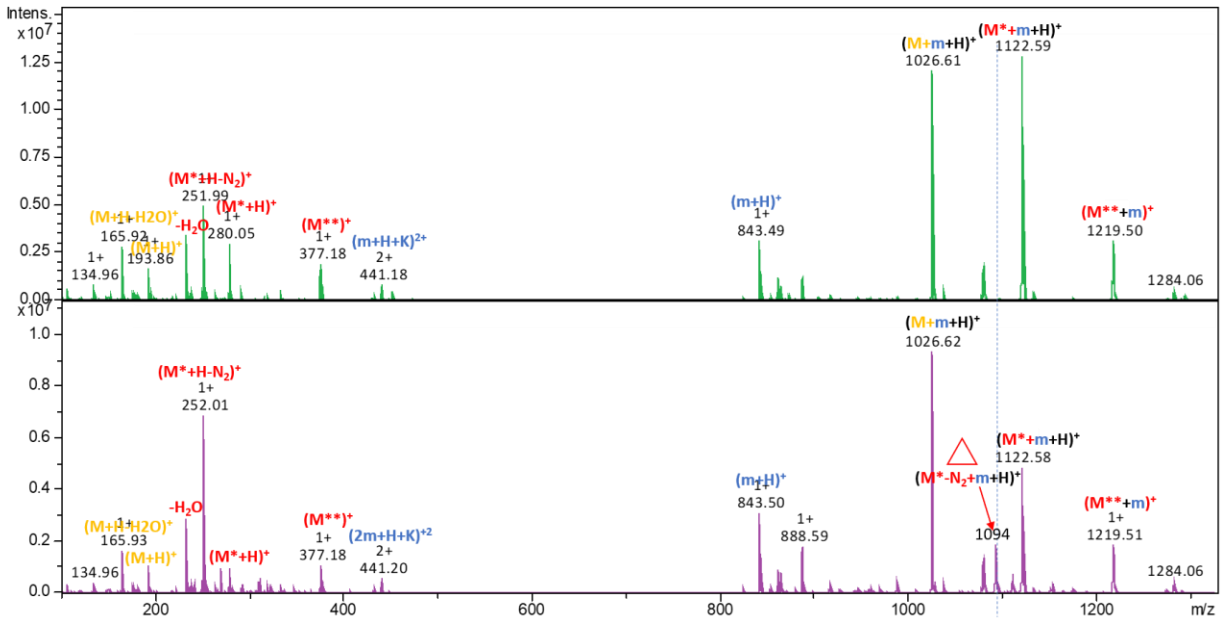


Figure 6.17. The ESI spectrum of equimolar mixture of (L)*-adrenaline and N-acetylated $SSIVSFY_{(NH2)}$ complexes (a) before condensed-phase photolysis (b) after condensed-phase photolysis

The Figure 6.18 (a) showed the MS^2-CID spectrum of $(M^*\text{-}N_2\text{+m+H})^+$ from condensed-phase UV photolysis sample, compared against the CID spectrum of gas-phase UV photodissociative product from MS^2-UVPD Figure 6.18 (b). Except for the *m/z* at 1077 that was not existed in (a), the ion series were the same across the comparison. Another difference was that the relative intensities of the monomeric ions were less abundant to the precursor ion in (a), under the same activation energy. This could indicate that the two complexes were bonded slightly different since (a) was formed directly from a olefin alkylated ligand ($M^*\text{-}N_2$) in solution whereas (b) was a photo-dissociative product from $(M^*\text{+m+H})^+$ complex.

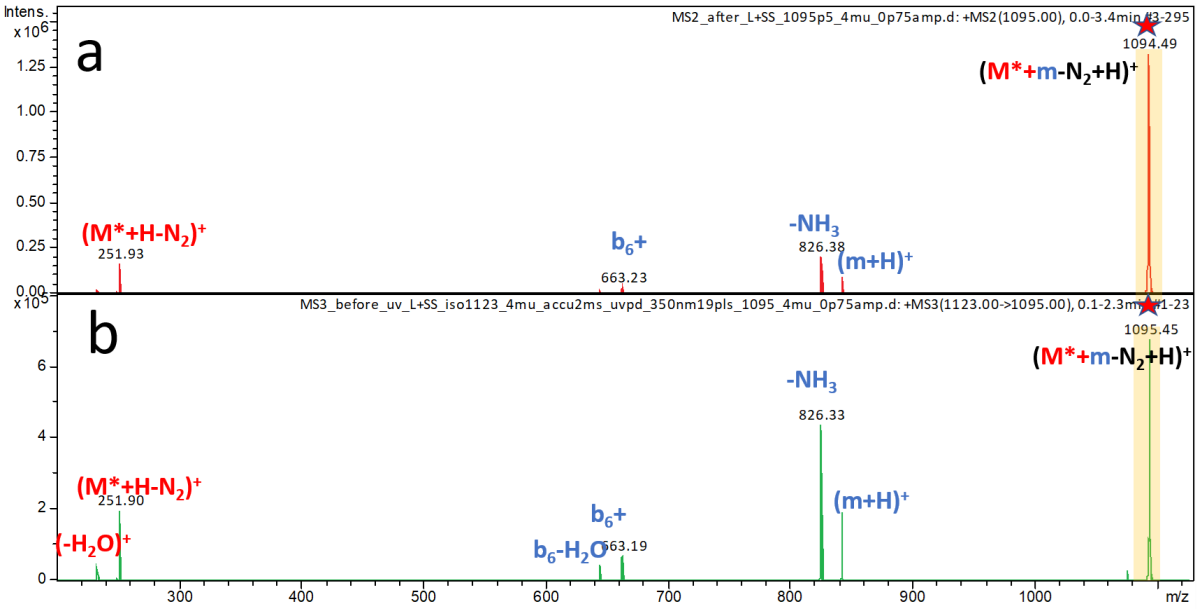


Figure 6.18. CID spectra of $(M^*\text{-}N_2\text{+}m\text{+}H)^+$ *m/z* 1094 (a) isolated from ESI of condensed phase photolysis experiment; (b) isolated from ESI-UVPD gas-phase crosslinking experiment of *N-Acetyl*-SSIVSFY$_{(NH2)}$ and (L)*-adrenaline

6.3.6 *Complex Structures*

The experimental data did not provide much constraint of crosslinking site, even though a high survival yield was observed for the MS^2-UVPD. To improve the resolution and relate the result to the gas-phase ion structures, we carried out extensive computational analysis for the ((L)* + Acetyl-SSIVSFY$_{(NH2)}$ + H)$^+$ system using Born-Oppenheimer molecular dynamics (BOMD) at the semi-empirical level of quantum theory that included all valence electron interactions. Thermodynamic calculations were also carried out that relied on DFT including dispersion interactions.

12 initial complex ion structures were built by arranging the catechol and the amino group towards different polar residue of the peptide counterpart. The BOMD was run at 410K for 20 ps to produce 200,000 snapshots out of which 2400 structures were extracted at 100 fs intervals,

optimized by PM6-D3H4, are sorted by their secondary and super secondary similarities to compact duplicates and reduce the size of the selection. All structures had shown consistent protonation on the amino group. These 60-70 distinct structures were selected for full geometry optimization by DFT gradient calculations with B3LYP/6-31G(d,p). The optimized structures were sorted again by conformation similarities to reduce duplicates and reduce the size to 48 isomers. The selective optimized structures were used for frequency calculations to provide vibrational enthalpies and entropies, while the electronic energies were provided by ωB97X-D/6–31+G(d,p) re-optimization. The electronic, vibrational and rotational terms were then combined to yield relative free energies for the complexes. Since these complexes are formed in solution, solvent effects were included in single-point energy calculations via the polarizable continuum model[27] using water as the dielectric. The calculated relative free energies of complex 0-47 are plotted in (Figure 6.19). The solvent effect had altered the conformational preference in the gas-phase. The optimized geometries of complexes that have relative gas-phase free energies under 25 kJ/mol **(#0 ~ #6, #8, #9, #14**) are given further discussion. These structures, although being energetically favorable in the gas-phase, of which many of them are destabilized by up to 20 kJ/mol in water.

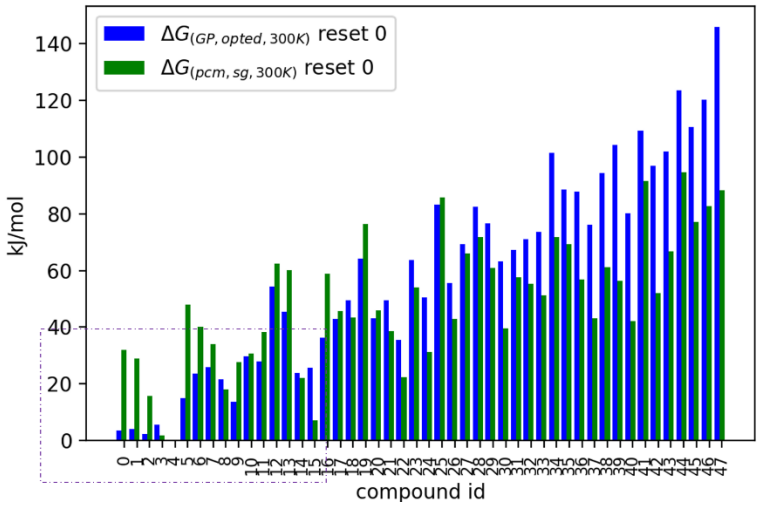


Figure 6.19. Relative free energies of hypothetical structure candidates. The blue bars and the green bars correspond to gas-phase relative energies, and solvated free energy, respectively.

The binding between the peptide and the (L)*-adrenaline complexes was realized in a few types summarized in (Table 6.1). The structures are illustrated in Figure 6.20, revealing the possible binding site of the ligand are the catechol and the amino side chain. These hydrogen bonds between the ligand and the peptide could be important for the agonist recognition. And the different binding patterns indicated different conformation states in the ligand bonded β_2AR. Among these optimized complexes displayed here in the work, **complex # 0, 3, 5, 6, 9, 14** showed no close contact (within 4.0 Å) at 0 K between the diazirine carbon (C151) to any atoms carrying hydrogen (X-H) in the peptide counterpart; **complex #1** and **#2** showed close contact to 3 atoms at the Phenylalanine side chain; **complex #4** showed a conformation with diazirine carbon located closely to 2 atoms at the Tyrosine aromatic ring; **complex #8** had shown 0 K close location of diazirine carbon to the O-H in Serine-5 and O-H and C-H on Tyrosine sidechain at 0 K.

Table 6.1. Relative free energies and binding types of low energy candidates

COMPLEX #	$\Delta G_{g,\ 300K}$ (kJ/mol)	Binding groups
0	3.5	2 cat-, amino-
1	4.1	2 cat-, amino-
2	2.3	1 cat-, amino-
3	5.6	amino-
4	0.0	amino-
5	14.8	2 cat-, amino-
6	23.7	1 cat-, amino-
8	21.6	2 cat-, amino-
9	13.7	1 cat-, amino-
14	23.8	amino-

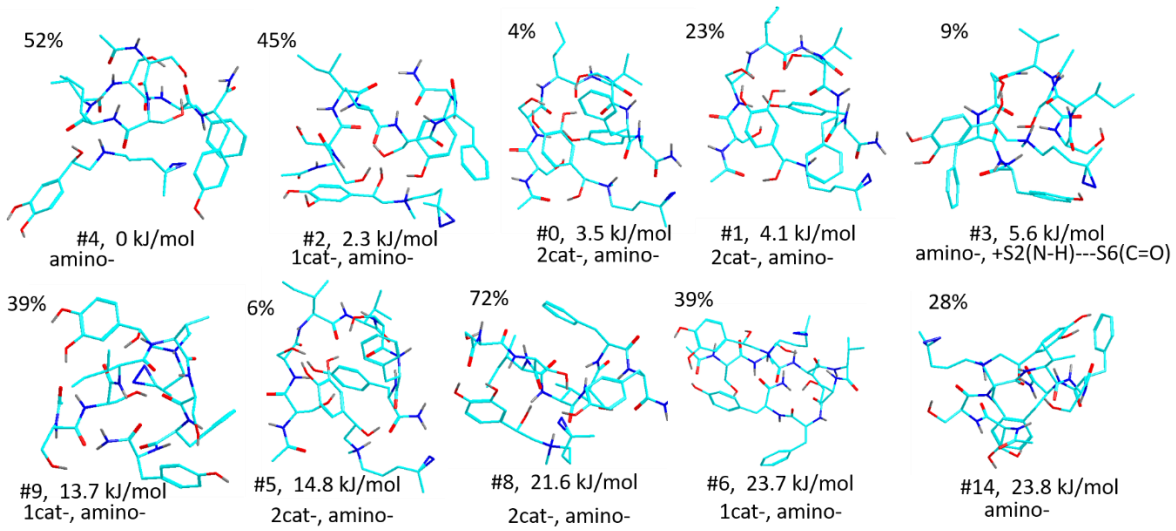


Figure 6.20. ωB97X-D/6–31+G(d,p) optimized structures of ((L)* + Acetyl-SSIVSFY$_{(NH2)}$ + H)$^{+}$ complexes.

Only low energy structures were displayed. Non-exchangeable hydrogens are hidden.

6.3.7 *Dynamic Binding Analysis and Binding Energy Analysis*

The close contact analysis under thermal motion were also evaluated by BOMD trajectory analysis, run for 100 ps at 310 K for the complexes. The overall contact frequency varied from 4% to 173%. The frequency exceeding 100% was due to multiple simultaneous contacts in the course of 100,000-step trajectory. The total contact frequency was plotted against the relative free energy in graph (Figure 6.21), and the plots were color coded by number of close contact (target X-H atoms and the carbene incipient carbon) at 0 K.

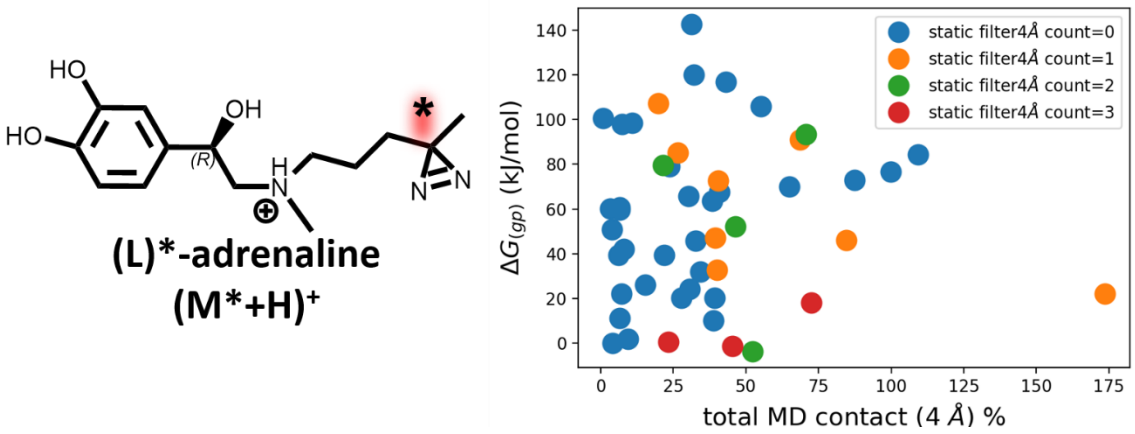


Figure 6.21. The total close contact rates of highlighted diazirine carbon C151 to the target peptide in 47 complexes studied by BOMD trajectories. The corresponding 0 K contact atom count was color-coded by blue (0), orange (1), green (2) and red (3).

The complex listed in Table 6.1 have showed a smaller range of dynamic contacts (4% for **complex #0** to 72% for **complex #8**), labeled at each conformation (Figure 6.20). It has revealed a trend that more contacts were developed near the 0K close contact site: F-6 and Y-7 were developed for **complex #1** and **#2**; for **complex #4**, 75% of contacts occurred at Y-7, and the secondary close contact was developed by thermal motion at the I-3 position, which was considered far (> 7 Å) in the 0 K conformation; **complex #8** showed developed close contact to S-2, in addition to S-5 and S-7 that was observed at 0 K. The contact at S-2 was not due to a complete conformational rearrangement since side chain of O-H group in S-2 was accessible (within 5 Å) by C151 under small bond rotation.

For the rest of the complexes that had no contact at 0 K, **#0** and **#5** only developed 4% and 6% contacts with emphasis on Y-7. This could be explained by the 0 K conformation that the two -OH on catechol and the amino group was tightly bonded to the peptide, pushing the diazirine chain to pointing at one direction and limiting the space it could approach. **Complex #3** also revealed low contact frequency (9%), possibly due to the strong interaction between residue S-2

to S-6 that had "wrapped" the ligand in a fixed position, preventing large conformation change and alkyl bond movement to develop contacts with the peptide. **Complex #9** and **#14** had different scenarios than the complexes **#0, 1, 3, and 5**. Both of them were loosely bonded to the peptide by the protonated amino group via the hydrogen bonding. The orientation of the ligand in both cases was like "floating" on the surface of the peptide, therefore allowing high flexibility of thermal motion. All the low energy structures discussed here had hydrogen bonding involving the benzylic alcohol (O137-H138…O) to oxygens in the peptide, which could be considered contributing in the chiral agonist recognition.

The recombination energy was calculated on fully optimized **dimer complex #0 #2, #4 and #5** with counterpoise correction. The counterpoise correction provided adjustment ranging from 15 kJ/mol to 29 kJ/mol. In MS^2-CID spectrum (Figure 6.22), when the dimer was activated by collision, it produced the protonated peptide ion which was more abundant than the protonated ligand. This indicates that the peptide stabilized the available proton when competing with the ligand. Therefore, extensive conformation search was applied to the protonated peptide and neutral ligand in order to find global minima of both products.

For the peptide monomer search, 7 protomer was constructed with proton placed at oxygen of different carbonyl groups as initial structures and BOMD at 410K was used to generate large amount of structures. The 1400 monomer structures extracted from sampling trajectories were de-duplicated and optimized by DFT gradient calculations and frequency calculations to provide vibrational enthalpies and entropies, performed in the same way as in the dimer conformer search. The optimized peptide structures had wide range of relative free energy distributions (0 ~ 151 kJ/mol) and only the ones within 30 kJ/mol were considered. In all low energy structures, the proton was coordinated by two carbonyl groups on the peptic backbone and had displayed a wide

variety of protomers that is denoted as the bonded residues (i.e. F-6—H…I-3, S-5—H…V4, N-acetyl—H…S-1 etc.) The neutral photo-ligand was also searched to find its global minimum. The low energy structure was coiled to a "C-shaped" conformation with the diazirine closed pointing to the catechol. This conformation was not observed in the ligand bonded in the dimer, possibly due to that the noncovalent interactions with the peptide had forced the ligand to stay extended to make the complex stable. The global minima of all optimized structures were selected to calculate recombination energy.(Figure 6.22) The recombination energy ΔH(0 K) of **complex #0** was calculated to be -150 kJ/mol, comparable to that of **complex #4** at -152 kJ/mol. Small absolute value indicated weaker binding, which were the cases for **complex #2** (-144 kJ/mol) and **complex #5** (-133 kJ/mol).

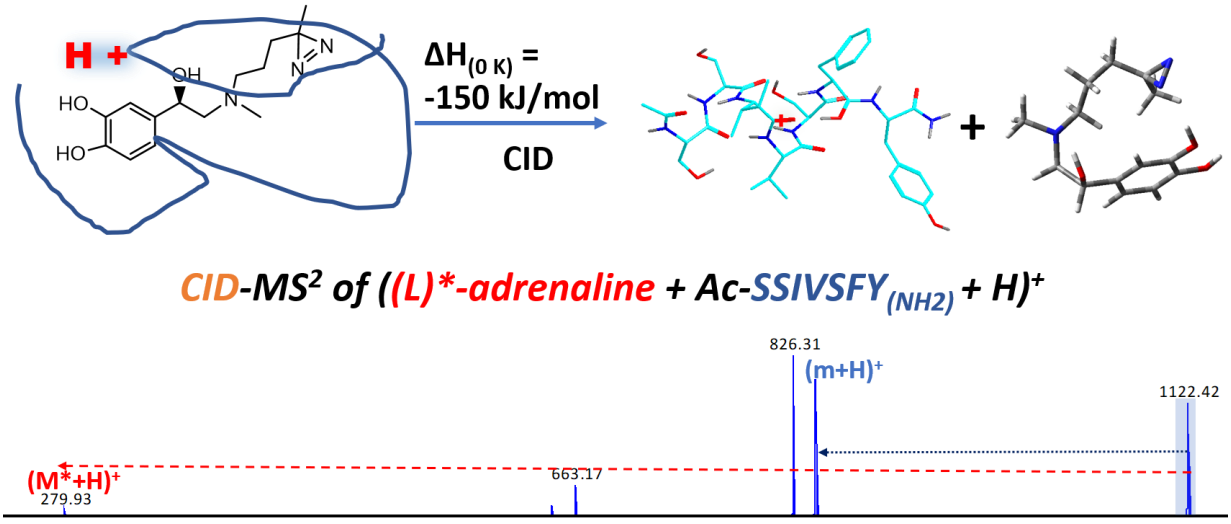


Figure 6.22.The recombination energy calculation for optimized dimer complexes.

6.4 FUTURE WORK

Since the photo-dissociative experimental results did not provide direct constraints of crosslink sites, additional characterization of the complex such as collision cross section will be

useful in distinguishing complexes that are bonded differently. The collision cross section of the ligand-peptide complexes will be measured by ion-mobility mass spectrometry, acting as an extra dimension of experimental constraint. The results are to be compared to theoretical collision cross section (Ω_{th}) analysis, and to select the best hypothetical structures. The best structures will be used in explaining the mechanisms of dynamic binding.

6.5 REFERENCES

1. Nygaard, R.; Zou, Y.; Ron; Thomas; Daniel; Manglik, A.; Albert; Corey; Juan; Michael; Foon; Tong; David; Mueller, L.; R; Brian, *Cell* **2013,** *152* (3), 532-542.
2. Deupi, X.; Kobilka, B. K., *Physiology* **2010,** *25* (5), 293-303.
3. Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. Ø.; Piana, S.; Shaw, D. E., *Proceedings of the National Academy of Sciences* **2009,** *106* (12), 4689-4694.
4. Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E., *Proceedings of the National Academy of Sciences* **2011,** *108* (46), 18684-18689.
5. Strosberg, A. D., *Protein Science* **1993,** *2* (8), 1198-1209.
6. Goral, V.; Jin, Y.; Sun, H.; Ferrie, A. M.; Wu, Q.; Fang, Y., *PloS one* **2011,** *6* (4), e19282.
7. Ring, A. M.; Manglik, A.; Kruse, A. C.; Enos, M. D.; Weis, W. I.; Garcia, K. C.; Kobilka, B. K., *Nature* **2013,** *502* (7472), 575-579.
8. Bokoch, M. P.; Zou, Y.; Rasmussen, S. G. F.; Liu, C. W.; Nygaard, R.; Rosenbaum, D. M.; Fung, J. J.; Choi, H.-J.; Thian, F. S.; Kobilka, T. S.; Puglisi, J. D.; Weis, W. I.; Pardo, L.; Prosser, R. S.; Mueller, L.; Kobilka, B. K., *Nature* **2010,** *463* (7277), 108-112.
9. Ishiuchi, S.; Yamada, K.; Oba, H.; Wako, H.; Fujii, M., *Phys Chem Chem Phys* **2016,** *18* (33), 23277-84.
10. Sekiguchi, T.; Tamura, M.; Oba, H.; Carcarbal, P.; Lozada-Garcia, R. R.; Zehnacker-Rentien, A.; Gregoire, G.; Ishiuchi, S. I.; Fujii, M., *Angew Chem Int Ed Engl* **2018,** *57* (20), 5626-5629.
11. Nguyen, H. T. H.; Andrikopoulos, P. C.; Rulek, L.; Shaffer, C. J.; Tureek, F., *Journal of the American Society for Mass Spectrometry* **2018,** *29* (8), 1706--1720.
12. Liu, Y.; Ramey, Z.; Turecek, F., *Chemistry - A European Journal* **2018,** *24* (37), 9259--9263.
13. Huang, S. R.; Liu, Y.; Turecek, F., *Phys Chem Chem Phys* **2019,** *21* (4), 2046-2056.
14. Liu, Y.; Turecek, F., *Journal of The American Society for Mass Spectrometry* **2019,** *30* (10), 1992-2006.
15. Stepensky, D.; Chorny, M.; Dabour, Z.; Schumacher, I., *Journal of Pharmaceutical Sciences* **2004,** *93* (4), 969-980.
16. Pepin, R.; Shaffer, C. J.; Turecek, F., *Journal of Mass Spectrometry* **2017,** *52* (8), 557-560.
17. Dang, A.; Korn, J. A.; Gladden, J.; Mozzone, B.; Tureček, F., *Journal of The American Society for Mass Spectrometry* **2019,** *30* (9), 1558-1564.
18. Shaffer, C. J.; Andrikopoulos, P. C.; Rezac, J.; Rulisek, L.; Turecek, F., *J Am Soc Mass Spectrom* **2016,** *27* (4), 633-45.
19. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., *Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.
20. Stewart, J. J. P., *Journal of Molecular Modeling* **2007,** *13* (12), 1173-1213.
21. Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P., *Journal of Chemical Theory and Computation* **2009,** *5* (7), 1749-1760.
22. Stewart, J. J. P. *MOPAC 16.*, Stewart Computational Chemistry, Colorado Springs, CO, 2016.
23. Rezac, J., *Journal of Computational Chemistry* **2016,** *37* (13), 1230-1237.
24. Řezáč, J. *Cuby—ruby framework for computational chemistry*, 4.
25. Becke, A. D., *Physical Review A* **1988,** *38* (6), 3098-3100.
26. Chai, J.-D.; Head-Gordon, M., *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615-6620.
27. Tomasi, J.; Mennucci, B.; Cammi, R., *Chemical Reviews* **2005,** *105* (8), 2999-3093.
28. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. A.01*, Wallingford, CT, 2016.
29. Shaffer, C. J.; Andrikopoulos, P. C.; Řez, *Journal of the American Society for Mass Spectrometry* **2016,** *27* (4), 633--645.
30. Shvartsburg, A. A.; Jarrold, M. F. *MobCal-A Program to Calculate Mobilities*.
31. Moritani, I.; Murahashi, S.-I.; Nishino, M.; Kimura, K.; Tsubomura, H., *Tetrahedron Letters* **1966,** *7* (4), 373-378.

BIBLIOGRAPHY

1. Hashimoto, M.; Rockenstein, E.; Crews, L.; Masliah, E., *Neuromolecular Med* **2003,** *4* (1-2), 21-36.
2. 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., *Nature Chemical Biology* **2011,** *7*, 285.
3. Mauser, H.; Guba, W., (1367-6733 (Print)).
4. Klebe, G., *Journal of Molecular Medicine* **2000,** *78* (5), 269-281.
5. Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T., *Nature chemistry* **2011,** *3* (2), 172-177.
6. McPherson Jr, A., *Methods of Biochemical Analysis* **1976**, 249-345.
7. Fraser, J. S.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T., *Proceedings of the National Academy of Sciences* **2011,** *108* (39), 16247.
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. Taraska, J. W., *Current opinion in structural biology* **2012,** *22* (4), 507-513.
10. Konermann, L.; Tong, X.; Pan, Y., *Journal of Mass Spectrometry* **2008,** *43* (8), 1021-1036.
11. Singh, P.; Panchaud, A.; Goodlett, D. R., *Analytical Chemistry* **2010,** *82* (7), 2636-2642.
12. Mulder, F. A. A.; Filatov, M., *Chem. Soc. Rev.* **2010,** *39* (2), 578-590.
13. Munoz, V.; Eaton, W. A., *Proceedings of the National Academy of Sciences* **1999,** *96* (20), 11311-11316.
14. Holm, L.; Kaariainen, S.; Rosenstrom, P.; Schenkel, A., *Bioinformatics* **2008,** *24* (23), 2780-2781.
15. Waga, S.; Hannon, G. J.; Beach, D.; Stillman, B., *Nature* **1994,** *369* (6481), 574-578.
16. Milani, M.; Mastrangelo, E.; Bollati, M.; Selisko, B.; Decroly, E.; Bouvet, M.; Canard, B.; Bolognesi, M., *Antiviral Research* **2009,** *83* (1), 28-34.
17. El Far, O.; Betz, H., *Biochemical Journal* **2002,** *365* (2), 329-336.
18. 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., *Cell* **2016,** *165* (7), 1698-1707.
19. Sharon, M.; Robinson, C. V., *Annual Review of Biochemistry* **2007,** *76* (1), 167-193.
20. Fields, S.; Song, O.-k., *Nature* **1989,** *340* (6230), 245-246.
21. Gingras, A.-C.; Gstaiger, M.; Raught, B.; Aebersold, R., *Nature Reviews Molecular Cell Biology* **2007,** *8* (8), 645-654.
22. Boeri Erba, E.; Petosa, C., *Protein Science* **2015,** *24* (8), 1176-1192.
23. Dormandy, T. L., *Annals of the Royal College of Surgeons of England* **1980,** *62* (3), 188-194.
24. Stubbe, J.; van der Donk, W. A., *Chemical Reviews* **1998,** *98* (2), 705-762.
25. Florence, T. M., *Australian and New Zealand Journal of Ophthalmology* **1995,** *23* (1), 3-7.
26. Balasubramanian, B.; Pogozelski, W. K.; Tullius, T. D., *Proceedings of the National Academy of Sciences* **1998,** *95* (17), 9738.
27. O'Neill, M. A. B., J. K., Sequence-Dependent DNA Dynamics: The Regulator of DNA-Mediated Charge Transport. In *Charge Transfer in DNA*, Wagenknecht, H. A., Ed. Wiley-VCH: Weinheim, Germany, 2005; pp 27-31.
28. Parker, A. W.; Lin, C. Y.; George, M. W.; Towrie, M.; Kuimova, M. K., *The Journal of Physical Chemistry B* **2010,** *114* (10), 3660-3667.
29. Bachler, V.; Hildenbrand, K., *International Journal of Radiation Applications and Instrumentation. Part C. Radiation Physics and Chemistry* **1992,** *40* (1), 59-68.
30. Close, D. M.; Crespo-Hernández, C. E.; Gorb, L.; Leszczynski, J., *The Journal of Physical Chemistry A* **2008,** *112* (49), 12702-12706.
31. Yao, C. X.; Cuadrado-Peinado, M. L.; Polasek, M.; Turecek, F., *Angewandte Chemie-International Edition* **2005,** *44* (41), 6708-6711.
32. Medved, M., *Rapid Communications in Mass Spectrometry* **1991,** *5* (1), 11-14.
33. Jackson, G. P.; Barkett, M. A., A History of the Forensic Applications of Mass Spectrometry. In *The Encyclopedia of Mass Spectrometry*, Gross, M. L.; Caprioli, R. M., Eds. Elsevier: Boston, 2016; pp 271-284.
34. Gelb, M. H.; Turecek, F.; Scott, C. R.; Chamoles, N. A., *Journal of Inherited Metabolic Disease* **2006,** *29* (2-3), 397-404.
35. Thomson, J. J., *Rays of positive electricity and their application to chemical analyses*. Longmans, Green and Company: 1921; Vol. 1.
36. McLafferty, F. W.; Turecek, F., *Interpretation of Mass Spectra Forth Edition*. 1993; p 1-17.
37. Munson, M. S. B.; Field, F. H., *Journal of the American Chemical Society* **1966,** *88* (12), 2621-2630.
38. Munson, B., *Analytical Chemistry* **1971,** *43* (13), 28A-43a.
39. Martens, J.; Berden, G.; Oomens, J., *Analytical Chemistry* **2016,** *88* (12), 6126-6129.
40. Yamashita, M.; Fenn, J. B., *The Journal of Physical Chemistry* **1984,** *88* (20), 4451-4459.
41. Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L., *The Clinical biochemist. Reviews* **2003,** *24* (1), 3-12.
42. Smith, R. D.; Light-Wahl, K. J.; Winger, B. E.; Loo, J. A., *Organic Mass Spectrometry* **1992,** *27* (7), 811-821.
43. Karpievitch, Y. V.; Polpitiya, A. D.; Anderson, G. A.; Smith, R. D.; Dabney, A. R., *The annals of applied statistics* **2010,** *4* (4), 1797-1823.
44. Nordström, A.; O'Maille, G.; Qin, C.; Siuzdak, G., *Analytical Chemistry* **2006,** *78* (10), 3289-3295.
45. Wilm, M., *Molecular & Cellular Proteomics* **2011,** *10* (7), M111.009407.
46. Kruve, A.; Kaupmees, K.; Liigand, J.; Oss, M.; Leito, I., *Journal of Mass Spectrometry* **2013,** *48* (6), 695-702.

47. El-Aneed, A.; Cohen, A.; Banoub, J., *Applied Spectroscopy Reviews* **2009,** *44* (3), 210-230.
48. Glish, G. L.; Burinsky, D. J., *Journal of the American Society for Mass Spectrometry* **2008,** *19* (2), 161-172.
49. Støchkel, K.; Milne, B. F.; Nielsen, S. B., *The Journal of Physical Chemistry A* **2011,** *115* (11), 2155-2159.
50. March, R. E.; Todd, J. F., *Quadrupole ion trap mass spectrometry*. John Wiley & Sons: 2005; Vol. 165.
51. March, R. E., *Journal of Mass Spectrometry* **1997,** *32* (4), 351-369.
52. Schwartz, J. C.; Senko, M. W.; Syka, J. E. P., *Journal of the American Society for Mass Spectrometry* **2002,** *13* (6), 659-669.
53. Sleno, L.; Volmer, D. A., *Journal of Mass Spectrometry* **2004,** *39* (10), 1091-1112.
54. Biemann, K., *Biomedical & environmental mass spectrometry* **1988,** *16* (1-12), 99-111.
55. Wu, J.; McLuckey, S. A., *International Journal of Mass Spectrometry* **2004,** *237* (2-3), 197-241.
56. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., *Proceedings of the National Academy of Sciences* **2004,** *101* (26), 9528-9533.
57. Wiesner, J.; Premsler, T.; Sickmann, A., *PROTEOMICS* **2008,** *8* (21), 4466-4483.
58. Turecek, F.; Julian, R. R., *Chemical Reviews* **2013,** *113* (8), 6691-6733.
59. Hari, Y.; Leumann, C. J.; Schürch, S., *Journal of The American Society for Mass Spectrometry* **2017,** *28* (12), 2677-2685.
60. Lyon, Y. A.; Beran, G.; Julian, R. R., *Journal of The American Society for Mass Spectrometry* **2017,** *28* (7), 1365-1373.
61. Polfer, N. C.; Dugourd, P., *Laser photodissociation and spectroscopy of mass-separated biomolecular ions*. Springer: 2013; Vol. 83.
62. Dang, A.; Korn, J. A.; Gladden, J.; Mozzone, B.; Tureček, F., *Journal of The American Society for Mass Spectrometry* **2019,** *30* (9), 1558-1564.
63. R. Julian, R., *Journal of The American Society for Mass Spectrometry* **2017,** *28* (9), 1823-1826.
64. Kim, T.-Y.; Thompson, M. S.; Reilly, J. P., *Rapid Communications in Mass Spectrometry* **2005,** *19* (12), 1657-1665.
65. Wilson, J. J.; Brodbelt, J. S., *Analytical Chemistry* **2007,** *79* (20), 7883-7892.
66. Cismesia, A. P.; Bailey, L. S.; Bell, M. R.; Tesler, L. F.; Polfer, N. C., *Journal of The American Society for Mass Spectrometry* **2016,** *27* (5), 757-766.
67. Filippi, A.; Fraschetti, C.; Rondino, F.; Piccirillo, S.; Steinmetz, V.; Guidoni, L.; Speranza, M., *International Journal of Mass Spectrometry* **2013,** *354*, 54-61.
68. Korn, J. A.; Urban, J.; Dang, A.; Nguyen, H. T. H.; Turecek, F., *Journal of Physical Chemistry Letters* **2017,** *8* (17), 4100-4107.
69. Polfer, N. C. D., P., Laser Photodissociation and Spectroscopy of Mass Separated Biomolecular Ions. In *Lecture Notes in Chemistry*, Springer: Cham, 2013; Vol. 83, pp 13-20.
70. Marcum, J. C.; Halevi, A.; Weber, J. M., *Phys Chem Chem Phys* **2009,** *11* (11), 1740-51.
71. Liu, Y.; Korn, J. A.; Dang, A.; Turecek, F., *Journal of Physical Chemistry B* **2018,** *122* (42), 9665-9680.
72. Dang, A.; Liu, Y.; Turecek, F., *Journal of Physical Chemistry A* **2019,** *123* (15), 3272-3284.
73. Dietz, L. A., *Review of Scientific Instruments* **1965,** *36* (12), 1763-1770.
74. Sinz, A., *Analytical and Bioanalytical Chemistry* **2005,** *381* (1), 44-47.
75. Belsom, A.; Mudd, G.; Giese, S.; Auer, M.; Rappsilber, J., *Analytical Chemistry* **2017,** *89* (10), 5319-5324.
76. A. Fleming, S., *Tetrahedron* **1995,** *51* (46), 12479-12520.
77. Weber, P. J. A.; Beck-Sickinger, A. G., *The Journal of Peptide Research* **1997,** *49* (5), 375-383.
78. Das, J., *Chemical Reviews* **2011,** *111* (8), 4405-4417.
79. Bayley, H.; Knowles, J. R., *Biochemistry* **1978,** *17* (12), 2420-2423.
80. Pezacki, J. P.; Couture, P.; Dunn, J. A.; Warkentin, J.; Wood, P. D.; Lusztyk, J.; Ford, F.; Platz, M. S., *Journal of Organic Chemistry* **1999,** *64* (12), 4456--4464.
81. Suchanek, M.; Radzikowska, A.; Thiele, C., *Nature Methods* **2005,** *2* (4), 261-268.
82. Vila-Perelló, M.; Pratt, M. R.; Tulin, F.; Muir, T. W., *Journal of the American Chemical Society* **2007,** *129* (26), 8068-8069.
83. Pepin, R.; Shaffer, C. J.; Turecek, F., *Journal of Mass Spectrometry* **2017,** *52* (8), 557-560.
84. Nguyen, H. T. H.; Andrikopoulos, P. C.; Rulek, L.; Shaffer, C. J.; Tureek, F., *Journal of the American Society for Mass Spectrometry* **2018,** *29* (8), 1706--1720.
85. Liu, Y.; Ramey, Z.; Turecek, F., *Chemistry - A European Journal* **2018,** *24* (37), 9259--9263.
86. Yu, C.; Huang, L., *Analytical Chemistry* **2018,** *90* (1), 144-165.
87. Sinz, A., *Angewandte Chemie International Edition* **2018,** *57* (22), 6390-6396.
88. Wodrich, M. D.; Corminboeuf, C.; Schreiner, P. R.; Fokin, A. A.; Schleyer, P. v. R., *Organic Letters* **2007,** *9* (10), 1851-1854.
89. Alder, B. J.; Wainwright, T. E., *The Journal of Chemical Physics* **1959,** *31* (2), 459-466.
90. Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P., *Journal of Chemical Theory and Computation* **2009,** *5* (7), 1749-1760.
91. Forgy, E., *Biometrics* **1965,** *21*, 768-780.
92. Thorndike, R. L., *Psychometrika* **1953,** *18* (4), 267-276.
93. Chai, J. D.; Head-Gordon, M., *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615--6620.
94. Zhao, Y.; Truhlar, D. G., *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.
95. Shaffer, C. J.; Andrikopoulos, P. C.; Rezac, J.; Rulisek, L.; Turecek, F., *J Am Soc Mass Spectrom* **2016,** *27* (4), 633-45.
96. Huang, S. R.; Liu, Y.; Turecek, F., *Phys Chem Chem Phys* **2019,** *21* (4), 2046-2056.
97. Lesslie, M.; Lawler, J. T.; Dang, A.; Korn, J. A.; Bím, D.; Steinmetz, V.; Maître, P.; Tureček, F.; Ryzhov, V., *ChemPhysChem* **2017,** *18* (10), 1293-1301.
98. Barbatti, M.; Ruckenbauer, M.; Plasser, F.; Pittner, J.; Granucci, G.; Persico, M.; Lischka, H., *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014,** *4* (1), 26-33.
99. Liu, Y.; Yi, F.; Kumar, A. B.; Chennamaneni, N. K.; Hong, X.; Scott, C. R.; Gelb, M. H.; Turecek, F., *Clinical chemistry* **2017,** *63* (6), 1118-1126.
100. Khaledi, H.; Liu, Y.; Masi, S.; Gelb, M. H., *Analytical chemistry* **2018,** *90* (20), 12168-12171.
101. Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H., *Journal of Molecular Biology* **1990,** *213* (4), 899-929.
102. Mann, A. P.; Scodeller, P.; Hussain, S.; Joo, J.; Kwon, E.; Braun, G. B.; Molder, T.; She, Z.-G.; Kotamraju, V. R.; Ranscht, B.; Krajewski, S.; Teesalu, T.; Bhatia, S.; Sailor, M. J.; Ruoslahti, E., *Nature Communications* **2016,** *7*.
103. https://www.uniprot.org/blast (accessed April 23 2018).

104. Arlt, C.; Flegler, V.; Ihling, C. H.; Schaefer, M.; Thondorf, I.; Sinz, A., *Angewandte Chemie-International Edition* **2017,** *56* (1), 275-279.
105. Koelbel, K.; Warnke, S.; Seo, J.; von Helden, G.; Moretti, R.; Meiler, J.; Pagel, K.; Sinz, A., *Chemistryselect* **2016,** *1* (13), 3651-3656.
106. Roepstorff, P.; Fohlman, J., *Biomedical Mass Spectrometry* **1984,** *11* (11), 601-601.
107. Biemann, K., *Methods in Enzymology* **1990,** *193*, 886-887.
108. Vaisar, T.; Urban, J., *Journal of Mass Spectrometry* **1996,** *31* (10), 1185-1187.
109. Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A., *Angewandte Chemie-International Edition* **2006,** *45* (32), 5301-5303.
110. Chai, J.-D.; Head-Gordon, M., *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615-6620.
111. Shaffer, C. J.; Marek, A.; Nguyen, H. T. H.; Turecek, F., *Journal of the American Society for Mass Spectrometry* **2015,** *26* (8), 1367-1381.
112. von Hippel, P. H., *Annu. Rev. Biophys. Biomol. Struct.* **2007,** *36*, 79 - 105.
113. Raindlova, V.; Pohl, R.; Hocek, M., *Chemistry* **2012,** *18* (13), 4080-4087.
114. Carrette, L. L. G.; Morii, T.; Madder, A., *Bioconjugate Chemistry* **2013,** *24* (12), 2008-2014.
115. Flett, F. J.; Walton, J. G. A.; Mackay, C. L.; Interthal, H., *Analytical Chemistry* **2015,** *87* (19), 9595--9599.
116. Wickramaratne, S.; Boldry, E. J.; Buehler, C.; Wang, Y. C.; Distefano, M. D.; Tretyakova, N. Y., *Journal of Biological Chemistry* **2015,** *290* (2), 775--787.
117. Ming, X.; Groehler, A.; Michaelson-Richie, E. D.; Villalta, P. W.; Campbell, C.; Tretyakova, N. Y., *Chemical Research in Toxicology* **2017,** *30* (4), 980--995.
118. Buxton, K. E.; Kennedy-Darling, J.; Shortreed, M. R.; Zaidan, N. Z.; Olivier, M.; Scalf, M.; Sridharan, R.; Smith, L. M., *J Proteome Res* **2017,** *16* (9), 3433-3442.
119. Weir Lipton, M. S.; Fuciarelli, A. L.; Springer, D. L.; Hofstadler, S. A.; Edmonds, C. G., *Rapid Communications in Mass Spectrometry* **1997,** *11* (15), 1673-1676.
120. Golden, M. C.; Resing, K. A.; Koch, T. H.; Collins, B. D.; Willis, M. C., *Protein Science* **2008,** *8* (12), 2806--2812.
121. Wong, D. L.; Reich, N. O., *Biochemistry* **2000,** *39* (50), 15410--15417.
122. Rieger, R. A.; McTigue, M. M.; Kycia, J. H.; Gerchman, S. E.; Grollman, A. P.; Iden, C. R., *Journal of the American Society for Mass Spectrometry* **2000,** *11* (6), 505--515.
123. Steen, H.; Petersen, J.; Mann, M.; Jensen, O. N., *Protein Science* **2001,** *10* (10), 1989--2001.
124. Wagenknecht, H. A.; Rajski, S. R.; Pascaly, M.; Stemp, E. D. A.; Barton, J. K., *Journal of the American Chemical Society* **2001,** *123* (19), 4400--4407.
125. Doneanu, C. E.; Gafken, P. R.; Bennett, S. E.; Barofsky, D. F., *Analytical Chemistry* **2004,** *76* (19), 5667--5676.
126. Pourshahian, S.; Limbach, P. A., *J Mass Spectrom* **2008,** *43* (8), 1081-8.
127. Rosenfeld, K. K.; Ziv, T.; Goldin, S.; Glaser, F.; Manor, H., *J Mol Biol* **2011,** *410* (1), 77-92.
128. Steen, H.; Jensen, N., *Mass Spectrometry Reviews* **2002,** *21* (3), 163-82.
129. Tacheny, A.; Dieu, M.; Arnould, T.; Renard, P., Mass spectrometry-based identification of proteins interacting with nucleic acids. 2013; Vol. 94, pp 89--109.
130. Laughlin, S. a., May the best molecule win: Competition ESI mass spectrometry. 2015; Vol. 16, pp 24506--24531.
131. Groehler, A.; Degner, A.; Tretyakova, N. Y., Mass Spectrometry-Based Tools to Characterize DNA–Protein Cross-Linking by Bis-Electrophiles. 2017; Vol. 121, pp 63--77.
132. Lin, S.; Cotter, R. J.; Woods, A. S., *Proteins* **1998,** *Suppl 2*, 12-21.
133. Veenstra, T. D., Electrospray ionization mass spectrometry: A promising new technique in the study of protein/DNA noncovalent complexes. 1999; Vol. 257, pp 1--5.
134. Beck, J. L.; Colgrave, M. L.; Ralph, S. F.; Shei, M. M., *Mass Spectrometry Reviews* **2001,** *20* (2), 61--87.
135. Alves, S.; Woods, A.; Tabet, J. C., *J Mass Spectrom* **2007,** *42* (12), 1613-22.
136. Alves, S.; Woods, A.; Delvolvé, A.; Tabet, J. C., *International Journal of Mass Spectrometry* **2008,** *278* (2-3), 122-128.
137. Brahim, B.; Tabet, J. C.; Alves, S., *Eur J Mass Spectrom (Chichester)* **2018,** *24* (1), 168-177.
138. Mann, A. P.; Scodeller, P.; Hussain, S.; Joo, J.; Kwon, E.; Braun, G. B.; Mlder, T.; She, Z. G.; Kotamraju, V. R.; Ranscht, B.; Krajewski, S.; Teesalu, T.; Bhatia, S.; Sailor, M. J.; Ruoslahti, E., *Nature Communications* **2016,** *7*.
139. Huang, R. C.; Bonner, J., *Proceedings of the National Academy of Sciences of the United States of America* **1962,** *48*, 1216--1222.
140. Bannister, A. J.; Kouzarides, T., Regulation of chromatin by histone modifications. 2011; Vol. 21, pp 381--395.
141. Strahl, B. D.; Allis, C. D., The language of covalent histone modifications. 2000; Vol. 403, pp 41--45.
142. Jenuwein, T.; Allis, C. D., Translating the histone code. 2001; Vol. 293, pp 1074--1080.
143. Shaffer, C. J.; Andrikopoulos, P. C.; Řez, *Journal of the American Society for Mass Spectrometry* **2016,** *27* (4), 633--645.
144. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., *Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.
145. Stewart, J. J. P., *Journal of Molecular Modeling* **2007,** *13* (12), 1173-1213.
146. Stewart, J. J. P. *MOPAC 16.*, Stewart Computational Chemistry, Colorado Springs, CO, 2016.
147. Rezac, J., *Journal of Computational Chemistry* **2016,** *37* (13), 1230-1237.
148. Řezáč, J. *Cuby—ruby framework for computational chemistry*, 4.
149. Becke, A. D., *Physical Review A* **1988,** *38* (6), 3098-3100.
150. Tomasi, J.; Mennucci, B.; Cammi, R., *Chemical Reviews* **2005,** *105* (8), 2999-3093.
151. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. A.01*, Wallingford, CT, 2016.
152. Marek, A.; Turecek, F., *Journal of the American Society for Mass Spectrometry* **2014,** *25* (5), 778--789.

153. Seburg, R. A.; McMahon, R. J., *Journal of the American Chemical Society* **1992,** *114* (18), 7183--7189.
154. Jackson, J. E.; Soundararajan, N.; White, W.; Liu, M. T. H.; Bonneau, R.; Platz, M. S., *Journal of the American Chemical Society* **1989,** *111* (17), 6874--6875.
155. Stevens, I. D. R.; Liu, M. T. H.; Soundararajan, N.; Paike, N., *Tetrahedron Letters* **1989,** *30* (4), 481--484.
156. Murray, K. K., *Journal of Mass Spectrometry* **1996,** *31* (11), 1203-1215.
157. McLuckey, S. A.; Berkel, G. J.; Glish, G. L., *Journal of the American Society for Mass Spectrometry* **1992,** *3* (1), 60-70.
158. Roepstorff, P.; Fohlman, J., *Biological Mass Spectrometry* **1984,** *11* (11), 601--601.
159. Wu, J.; McLuckey, S. A., Gas-phase fragmentation of oligonucleotide ions. 2004; Vol. 237, pp 197--241.
160. Schurch, S., *Mass Spectrom Rev* **2016,** *35* (4), 483-523.
161. Liu, M. T., *Chemistry of diazirines*. CRC: 1987; Vol. 1.
162. Tippmann, E. M.; Platz, M. S.; Svir, I. B.; Klymenko, O. V., *Journal of the American Chemical Society* **2004,** *126* (18), 5750--5762.
163. Mohan, U.; Burai, R.; McNaughton, B. R., *Tetrahedron Letters* **2014,** *55* (22), 3358--3360.
164. Nowak, I.; Robins, M. J., *Journal of Organic Chemistry* **2007,** *72* (9), 3319--3325.
165. Jerbi, J.; Springborg, M., *International Journal of Quantum Chemistry* **2018,** *118* (11).
166. Khanna, K. K.; Jackson, S. P., *Nature Genetics* **2001,** *27* (3), 247-254.
167. Wagenknecht, H. A., *Charge Transfer in DNA*. Wiley-VCH: Weinheim, Germany, 2005; p 1-23.
168. Schuster, G. B., *Accounts of Chemical Research* **2000,** *33* (4), 253-260.
169. Giese, B., *Accounts of Chemical Research* **2000,** *33* (9), 631-636.
170. Giese, B.; Amaudrut, J.; Kohler, A. K.; Spormann, M.; Wessely, S., *Nature* **2001,** *412* (6844), 318-320.
171. Douki, D.; Ravanat, J. L.; Angelov, D.; Wagner, J. R.; Cadet, J., Effects of duplex stability on charge-transfer efficiency within DNA. In *Long-Range Charge Transfer in DNA I*, 2004; Vol. 236, pp 1-25.
172. Joy, A.; Ghosh, A. K.; Schuster, G. B., *Journal of the American Chemical Society* **2006,** *128* (16), 5346-5347.
173. Ghosh, A.; Joy, A.; Schuster, G. B.; Douki, T.; Cadet, J., *Organic & Biomolecular Chemistry* **2008,** *6* (5), 916-928.
174. Joseph, J.; Schuster, G. B., *Journal of the American Chemical Society* **2009,** *131* (39), 13904-+.
175. Kanvah, S.; Schuster, G. B., *Organic & Biomolecular Chemistry* **2010,** *8* (6), 1340-1343.
176. Kanvah, S.; Joseph, J.; Schuster, G. B.; Barnett, R. N.; Cleveland, C. L.; Landman, U., *Accounts of Chemical Research* **2010,** *43* (2), 280-287.
177. Kravec, S. M.; Kinz-Thompson, C. D.; Conwell, E. M., *Journal of Physical Chemistry B* **2011,** *115* (19), 6166-6171.
178. Barnett, R. N.; Joseph, J.; Landman, U.; Schuster, G. B., *Journal of the American Chemical Society* **2013,** *135* (10), 3904-3914.
179. Cauet, E.; Lievin, J., Radical cations of the nucleic bases and radiation damage to DNA: Ab initio study. In *Advances in Quantum Chemistry, Vol 52*, 2007; Vol. 52, pp 121-147.
180. Rodriguez-Santiago, L. N., 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: Weinheim, Germany, 2010; pp 219-242.
181. Li, X. F.; Cai, Z. L.; Sevilla, M. D., *Journal of Physical Chemistry B* **2001,** *105* (41), 10115-10123.
182. Wang, M.; Zhao, J.; Bu, Y., *Physical Chemistry Chemical Physics* **2013,** *15* (42), 18453-18463.
183. Choi, J.; Yang, C.; Fujitsuka, M.; Tojo, S.; Ihee, H.; Majima, T., *Journal of Physical Chemistry Letters* **2015,** *6* (24), 5045-5050.
184. Sevilla, M. D.; Kumar, A.; Adhikary, A., *Journal of Physical Chemistry B* **2016,** *120* (11), 2984-2986.
185. Choi, J.; Yang, C.; Fujitsuka, M.; Tojo, S.; Ihee, H.; Majima, T., *Journal of Physical Chemistry B* **2016,** *120* (11), 2987-2989.
186. Lin, Y.; Wang, H.; Gao, S.; Li, R.; Schaefer, H. F., III, *Journal of Physical Chemistry B* **2012,** *116* (30), 8908-8915.
187. Hsu, S. C. N.; Wang, T.-P.; Kao, C.-L.; Chen, H.-F.; Yang, P.-Y.; Chen, H.-Y., *Journal of Physical Chemistry B* **2013,** *117* (7), 2096-2105.
188. Feketeova, L.; Chan, B.; Khairaah, G. N.; Steinmetz, V.; Maitre, P.; Radom, L.; O'Hair, R. A. J., *Journal of Physical Chemistry Letters* **2017,** *8* (13), 3159-3165.
189. Wolken, J. K.; Syrstad, E. A.; Vivekananda, S.; Turecek, F., *Journal of the American Chemical Society* **2001,** *123* (24), 5804-5805.
190. Chen, X. H.; Syrstad, E. A.; Nguyen, M. T.; Gerbaux, P.; Turek, F., *Journal of Physical Chemistry A* **2004,** *108* (42), 9283-9293.
191. Lam, A. K. Y.; Abrahams, B. F.; Grannas, M. J.; McFadyen, W. D.; O'Hair, R. A. J., *Dalton Transactions* **2006,** (42), 5051-5061.
192. Wee, S.; O'Hair, R. A. J.; McFadyen, W. D., *Rapid Communications in Mass Spectrometry* **2005,** *19* (13), 1797-1805.
193. Feketeova, L.; Khairallah, G. N.; Chan, B.; Steinmetz, V.; Maitre, P.; Radom, L.; O'Hair, R. A. J., *Chemical Communications* **2013,** *49* (66), 7343-7345.
194. Dang, A.; Nguyen, H. T. H.; Ruiz, H.; Piacentino, E.; Ryzhoy, V.; Turecek, F., *Journal of Physical Chemistry B* **2018,** *122* (1), 86-97.
195. Turecek, F., Transient intermediates of chemical reactions by neutralization-reionization mass spectrometry. In *Modern Mass Spectrometry*, 2003; Vol. 225, pp 77-129.
196. McLuckey, S. A.; Huang, T.-Y., *Analytical Chemistry* **2009,** *81* (21), 8669-8676.
197. Zubarev, R. A., *Mass Spectrometry Reviews* **2003,** *22* (1), 57-77.
198. Westphal, K.; Wiczk, J.; Miloch, J.; Kciuk, G.; Bobrowski, K.; Rak, J., *Organic & Biomolecular Chemistry* **2015,** *13* (41), 10362-10369.
199. Steenken, S., *Chemical Reviews* **1989,** *89* (3), 503-520.
200. Antoine, R. D., P., UV−Visible Activation of Biomolecular Ions. In *Laser Photodissociation and Spectroscopy of Mass-separated Biomolecular Ions*, Polfer, N., Dugourd, P., Ed. Springer: Cham, 2013; Vol. 83.
201. Furche, F.; Ahlrichs, R., *Journal of Chemical Physics* **2002,** *117* (16), 7433-7447.
202. Comeau, D. C.; Bartlett, R. J., *Chemical Physics Letters* **1993,** *207* (4-6), 414-423.
203. Shaffer, C. J.; Pepin, R.; Turecek, F., *J Mass Spectrom* **2015,** *50* (12), 1438-42.
204. Nguyen, H. T. H.; Shaffer, C. J.; Pepin, R.; Turecek, F., *Journal of Physical Chemistry Letters* **2015,** *6* (23), 4722-4727.
205. Viglino, E.; Shaffer, C. J.; Turecek, F., *Angewandte Chemie-International Edition* **2016,** *55* (26), 7469-7473.
206. Nguyen, H. T. H.; Andrikopoulos, P. C.; Bim, D.; Rulisek, L.; Dang, A.; Turecek, F., *Journal of Physical Chemistry B* **2017,** *121* (27), 6557-6569.
207. Moller, C.; Plesset, M. S., *Physical Review* **1934,** *46* (7), 0618-0622.
208. Schlegel, H. B., *Journal of Chemical Physics* **1986,** *84* (8), 4530-4534.
209. Mayer, I., *International Journal of Quantum Chemistry* **1978,** *14* (1), 29-38.

210. Chan, B.; Radom, L., *Theoretical Chemistry Accounts* **2011,** *130* (2-3), 251-260.
211. Reed, A. E.; Weinstock, R. B.; Weinhold, F., *Journal of Chemical Physics* **1985,** *83* (2), 735-746.
212. Gilbert, R. G.; Smith, S. C.
213. Zhu, L.; Hase, W. L., *Indiana University, Bloomington. Indiana, USA* **1994**.
214. Frank, A. J.; Sadilek, M.; Ferrier, J. G.; Turecek, F., *Journal of the American Chemical Society* **1997,** *119* (50), 12343-12353.
215. Viglino, E.; Lai, C. K.; Mu, X.; Chu, I. K.; Turecek, F., *Journal of the American Society for Mass Spectrometry* **2016,** *27* (9), 1454-1467.
216. Petroselli, G.; Dantola, M. L.; Cabrerizo, F. M.; Capparelli, A. L.; Lorente, C.; Oliveros, E.; Thomas, A. H., *Journal of the American Chemical Society* **2008,** *130* (10), 3001-3011.
217. Zhang, Y.; Improta, R.; Kohler, B., *Physical Chemistry Chemical Physics* **2014,** *16* (4), 1487-1499.
218. Colominas, C.; Luque, F. J.; Orozco, M., *Journal of the American Chemical Society* **1996,** *118* (29), 6811-6821.
219. Halder, A.; Bhattacharya, S.; Datta, A.; Bhattacharyya, D.; Mitra, A., *Physical Chemistry Chemical Physics* **2015,** *17* (39), 26249-26263.
220. Wu, R. R.; Yang, B.; Berden, G.; Oomens, J.; Rodgers, M. T., *Journal of Physical Chemistry B* **2014,** *118* (51), 14774-14784.
221. Yao, C. X.; Cuadrado-Peinado, M. L.; Polasek, M.; Turecek, F., *Journal of Mass Spectrometry* **2005,** *40* (11), 1417-1428.
222. Yao, C.; Turecek, F.; Polce, M. J.; Wesdemiotis, C., *International Journal of Mass Spectrometry* **2007,** *265* (2-3), 106-123.
223. Michl, J., *Journal of Molecular Spectroscopy* **1969,** *30* (1), 66-&.
224. Hobza, P.; Sponer, J., *Journal of the American Chemical Society* **2002,** *124* (39), 11802-11808.
225. Hammerum, S., *Journal of the American Chemical Society* **2009,** *131* (24), 8627-8635.
226. Fasman, G. D., *Handbook of biochemistry and molecular biology. Nucleic acids-v. 1-3*. 1975.
227. Dawson, R. M. C.; Elliott, D. C.; Elliott, W. H.; Jones, K. M., *Data for Biochemical Research*. 3rd ed.; Oxford University Press: Oxford, 1986.
228. Verdolino, V.; Cammi, R.; Munk, B. H.; Schlegel, H. B., *Journal of Physical Chemistry B* **2008,** *112* (51), 16860-16873.
229. Boudaiffa, B.; Cloutier, P.; Hunting, D.; Huels, M. A.; Sanche, L., *Science* **2000,** *287* (5458), 1658-1660.
230. Aflatooni, K.; Gallup, G. A.; Burrow, P. D., *Journal of Physical Chemistry A* **1998,** *102* (31), 6205-6207.
231. Jalbout, A. F.; Adamowicz, L., Electron attachment to DNA base complexes. In *Advances in Quantum Chemistry, Vol 52*, 2007; Vol. 52, pp 231-+.
232. Myers, L. S.; Hollis, M. L.; Theard, L. M., *Advances in Chemistry Series* **1968,** (81), 345-&.
233. Nucifora, G.; Smaller, B.; Avery, E. C.; Remko, R., *Radiation Research* **1972,** *49* (1), 96-&.
234. Hildenbrand, K.; Schultefrohlinde, D., *Free Radical Research Communications* **1990,** *11* (4-5), 195-206.
235. Syrstad, E. A.; Vivekananda, S.; Turecek, F., *Journal of Physical Chemistry A* **2001,** *105* (36), 8339-8351.
236. Wolken, J. K.; Turecek, F., *Journal of Physical Chemistry A* **2001,** *105* (36), 8352-8360.
237. Chen, X.; Syrstad, E. A.; Nguyen, M. T.; Gerbaux, P.; Tureček, F., *The Journal of Physical Chemistry A* **2005,** *109* (36), 8121-8132.
238. Turecek, F., Computational studies of radicals relevant to nucleic acid damage. In *Advances in Quantum Chemistry, Vol 52*, 2007; Vol. 52, pp 89-120.
239. Kostko, O.; Bravaya, K.; Krylov, A.; Ahmed, M., *Physical Chemistry Chemical Physics* **2010,** *12* (12), 2860-2872.
240. Chen, Z.; Lau, K.-C.; Garcia, G. A.; Nahon, L.; Bozanic, D. K.; Poisson, L.; Al-Mogren, M. M.; Schwell, M.; Francisco, J. S.; Bellili, A.; Hochlaf, M., *Journal of the American Chemical Society* **2016,** *138* (51), 16596-16599.
241. Alonso, J. L.; Vaquero, V.; Pena, I.; Lopez, J. C.; Mata, S.; Caminati, W., *Angewandte Chemie-International Edition* **2013,** *52* (8), 2331-2334.
242. Scanlan, M. J.; Hillier, I. H., *Journal of the Chemical Society-Chemical Communications* **1984,** (2), 102-103.
243. Kwiatkowski, J. S.; Bartlett, R. J.; Person, W. B., *Journal of the American Chemical Society* **1988,** *110* (8), 2353-2358.
244. Gould, I. R.; Green, D. V. S.; Young, P.; Hillier, I. H., *Journal of Organic Chemistry* **1992,** *57* (16), 4434-4437.
245. Kobayashi, R., *Journal of Physical Chemistry A* **1998,** *102* (52), 10813-10817.
246. Trygubenko, S. A.; Bogdan, T. V.; Rueda, M.; Orozco, M.; Luque, F. J.; Sponer, J.; Slavicek, P.; Hobza, P., *Physical Chemistry Chemical Physics* **2002,** *4* (17), 4192-4203.
247. Wolken, J. K.; Yao, C.; Turecek, F.; Polce, M. J.; Wesdemiotis, C., *International Journal of Mass Spectrometry* **2007,** *267* (1-3), 30-42.
248. Salpin, J.-Y.; Guillaumont, S.; Tortajada, J.; MacAleese, L.; Lemaire, J.; Maitre, P., *Chemphyschem* **2007,** *8* (15), 2235-2244.
249. Bakker, J. M.; Salpin, J.-Y.; Maitre, P., *International Journal of Mass Spectrometry* **2009,** *283* (1-3), 214-221.
250. Broquier, M.; Soorkia, S.; Pino, G.; Dedonder-Lardeux, C.; Jouvet, C.; Gregoire, G., *Journal of Physical Chemistry A* **2017,** *121* (34), 6429-6439.
251. Wu, R. R.; Yang, B.; Frieler, C. E.; Berden, G.; Oomens, J.; Rodgers, M. T., *Journal of Physical Chemistry B* **2015,** *119* (18), 5773-5784.
252. Wu, R. R.; Hamlow, L. A.; He, C. C.; Nei, Y. W.; Berden, G.; Oomens, J.; Rodgers, M. T., *Journal of the American Society for Mass Spectrometry* **2017,** *28* (8), 1638-1646.
253. Turecek, F.; Yao, C. X., *Journal of Physical Chemistry A* **2003,** *107* (43), 9221-9231.
254. Zhang, J. D.; Xie, Y.; Schaefer, H. F., III; Luo, Q.; Li, Q. S., *Molecular Physics* **2006,** *104* (13-14), 2347-2366.
255. Tomic, K.; Tatchen, J.; Marian, C. M., *Journal of Physical Chemistry A* **2005,** *109* (37), 8410-8418.
256. Bazso, G.; Tarczay, G.; Fogarasi, G.; Szalay, P. G., *Physical Chemistry Chemical Physics* **2011,** *13* (15), 6799-6807.
257. Kosma, K.; Schroeter, C.; Samoylova, E.; Hertel, I. V.; Schultz, T., *Journal of the American Chemical Society* **2009,** *131* (46), 16939-16943.
258. Ho, J.-W.; Yen, H.-C.; Chou, W.-K.; Weng, C.-N.; Cheng, L.-H.; Shi, H.-Q.; Lai, S.-H.; Cheng, P.-Y., *Journal of Physical Chemistry A* **2011,** *115* (30), 8406-8418.
259. Domingo, A.; Rodriguez-Fortea, A.; de Graaf, C., *Journal of Chemical Theory and Computation* **2012,** *8* (1), 235-244.
260. Lobsiger, S.; T'Rachsel, M. A.; Frey, H.-M.; Leutwyler, S., *Journal of Physical Chemistry B* **2013,** *117* (20), 6106-6115.
261. Nakayama, A.; Harabuchi, Y.; Yamazaki, S.; Taketsugu, T., *Physical Chemistry Chemical Physics* **2013,** *15* (29), 12322-12339.
262. Triandafillou, C. G.; Matsika, S., *Journal of Physical Chemistry A* **2013,** *117* (46), 12165-12174.
263. Martens, J.; Berden, G.; Gebhardt, C. R.; Oomens, J., *Review of Scientific Instruments* **2016,** *87* (10).
264. Knowles, P. J.; Andrews, J. S.; Amos, R. D.; Handy, N. C.; Pople, J. A., *Chemical Physics Letters* **1991,** *186* (2-3), 130-136.

265. 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.; Turecek, F., *International Journal of Mass Spectrometry* **2008,** *276* (2-3), 116-126.
266. Pepin, R.; Turecek, F., *Journal of Physical Chemistry B* **2015,** *119* (7), 2818-2826.
267. Goeringer, D. E.; McLuckey, S. A., *International Journal of Mass Spectrometry* **1998,** *177* (2-3), 163-174.
268. Liu, Y.; Korn, J. A.; Tureček, F., *International Journal of Mass Spectrometry* **2019,** *443*, 22-31.
269. Nygaard, R.; Zou, Y.; Ron; Thomas; Daniel; Manglik, A.; Albert; Corey; Juan; Michael; Foon; Tong; David; Mueller, L.; R; Brian, *Cell* **2013,** *152* (3), 532-542.
270. Deupi, X.; Kobilka, B. K., *Physiology* **2010,** *25* (5), 293-303.
271. Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. Ø.; Piana, S.; Shaw, D. E., *Proceedings of the National Academy of Sciences* **2009,** *106* (12), 4689-4694.
272. Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E., *Proceedings of the National Academy of Sciences* **2011,** *108* (46), 18684-18689.
273. Strosberg, A. D., *Protein Science* **1993,** *2* (8), 1198-1209.
274. Goral, V.; Jin, Y.; Sun, H.; Ferrie, A. M.; Wu, Q.; Fang, Y., *PloS one* **2011,** *6* (4), e19282.
275. Ring, A. M.; Manglik, A.; Kruse, A. C.; Enos, M. D.; Weis, W. I.; Garcia, K. C.; Kobilka, B. K., *Nature* **2013,** *502* (7472), 575-579.
276. Bokoch, M. P.; Zou, Y.; Rasmussen, S. G. F.; Liu, C. W.; Nygaard, R.; Rosenbaum, D. M.; Fung, J. J.; Choi, H.-J.; Thian, F. S.; Kobilka, T. S.; Puglisi, J. D.; Weis, W. I.; Pardo, L.; Prosser, R. S.; Mueller, L.; Kobilka, B. K., *Nature* **2010,** *463* (7277), 108-112.
277. Ishiuchi, S.; Yamada, K.; Oba, H.; Wako, H.; Fujii, M., *Phys Chem Chem Phys* **2016,** *18* (33), 23277-84.
278. Sekiguchi, T.; Tamura, M.; Oba, H.; Carcarbal, P.; Lozada-Garcia, R. R.; Zehnacker-Rentien, A.; Gregoire, G.; Ishiuchi, S. I.; Fujii, M., *Angew Chem Int Ed Engl* **2018,** *57* (20), 5626-5629.
279. Liu, Y.; Turecek, F., *Journal of The American Society for Mass Spectrometry* **2019,** *30* (10), 1992-2006.
280. Stepensky, D.; Chorny, M.; Dabour, Z.; Schumacher, I., *Journal of Pharmaceutical Sciences* **2004,** *93* (4), 969-980.
281. Shvartsburg, A. A.; Jarrold, M. F. *MobCal-A Program to Calculate Mobilities*.
282. Moritani, I.; Murahashi, S.-I.; Nishino, M.; Kimura, K.; Tsubomura, H., *Tetrahedron Letters* **1966,** *7* (4), 373-378.

APPENDIX A

Data processing scripts and repositories related to this work can be found on https://github.com/liu1992yang/ , owned by Yang Liu. All computer modeling jobs were completed on Hyak, a shared scalable super-computing infrastructure operated by UW-IT (https://wiki.cac.washington.edu/display/hyakusers/WIKI+for+Hyak+users).

CONFORMATION SEARCH PIPELINE SCRIPT

1. Geometry Clustering Scripts

This script is used for conformation clustering of molecules that have identical atom's numeric labeling.
Required input: Z-matrix coordinates; "key atom" list; geometry file name list
(input can be generated using script from 2.(C) below)
Output: Best number of clusters; clustering result; hydrogen-bonding patterns

```
import numpy as np
import pandas as pd
import matplotlib as mpl
mpl.use('Agg')
import matplotlib.pyplot as plt
import functools
import math
import os,re,sys

from itertools import combinations, compress
from scipy.spatial import distance
from scipy import signal
from sklearn.cluster import KMeans
from sklearn import metrics

X_H_COVALENT = 1.1 #distance threshold for X-H covalent bonds, X = C, O, N
X_H_HB = 2.5 #distance threshold for X-H noncovalent bonds, X = O,N
wdir = os.getcwd()


def read_reference(key_atom_fname):
  """
  Read from a ',' split key-atom file and return a list of int
  remove duplicates
  """
  reference_atom_num = set()
  try:
```

```
    with open(key_atom_fname, "r") as fin :
      for line in fin:
        content = line.strip()
        if content != '':
          reference_atom_num.update([int(i) for i in content.split(',')])
      return list(reference_atom_num)
  except OSError:
    print('OS error')
    sys.exit()
  except ValueError:
    print('Could not convert data to integer')
    sys.exit()


def get_indexes(reference_atom_num):
  reference_atom_num.sort()
  return [x-1 for x in reference_atom_num]

def get_index_pair(indexes_of_interest):
  #make pairs of all index combinations
  return list(combinations(indexes_of_interest,2))


def gen_distance_colnames(pairs):
  #generate column names of pairs
  headers = ['file']
  for pair in pairs:
    headers.append(str(pair[0]+1) + "_" + str(pair[1]+1))
  return headers


def read_energy_file(energy_fname):
  """
  read energy file and generate list of pairs[(comfile, energy),(),...]
  """
  file_energy = []
  try:
    with open(os.path.join(wdir,energy_fname),'r') as fin_energy:
      for line in fin_energy:
        arr = line.strip().split()
        if len(arr) < 2 or '' in arr:
          continue
        comfile, curr_energy = get_comfile_name(arr[0]), float(arr[1])
        file_energy.append((comfile, curr_energy))
      return file_energy
  except IOError:
    print("Could not read " + energy_fname)
    sys.exit()


def get_comfile_name(energy_name):
  #keep extracted file format
  #also allow compatibility of the old version that file ended in *.com
  if '.' in energy_name:
    return energy_name.strip()
  else:
    return energy_name.strip() + ".com"
```

```
def read_geomfile(filename):
  """
  read one geometry file and return nested list
  """
  data=[]
  with open(os.path.join(wdir, filename),'r') as fin:
    for line in fin:
      if line.strip() == '':
        continue
      if line.startswith("  "):
        arr = re.split(" +", line.strip())
        if len(arr) != 4:
            continue
        data.append((arr[0], float(arr[1]), float(arr[2]),float(arr[3])))
    return {idx: atom for idx, atom in enumerate(data)}
  raise ValueError("Error parsing geom file")



def get_coords(coords_dict,index):
  """
  type coords_dict: List[str, float, float, float]
  rtype: tup(float, float, float)
  """
  return coords_dict[index][1:]


def get_type(coords_dict,index):
  """
  Need to get atom type to be considered for hydrogen bonding
  rtype: str
  """
  return coords_dict[index][0]


def subset_by_type(coords_dict):
  """
  subset the coords_dict indexes by atom type
  only need H, N, O, F
  """
  indexes_h = []
  indexes_nof = []
  for idx in coords_dict.keys():
    type_idx = get_type(coords_dict, idx)
    if type_idx == 'H':
      indexes_h.append(idx)
    if type_idx in ('N', 'O', 'F'):
      indexes_nof.append(idx)
  return indexes_h, indexes_nof


def dist(coords_dict, index1, index2):
  """
  Given index1, index2, calculate distance: float
  """
```

```
  return distance.euclidean(get_coords(coords_dict,index1),
get_coords(coords_dict,index2))


def x_h_link(distance):
  """
  Determine whether X(C,O,N) and H atom is covalently linked
  type distance: float
  rtype: bool
  """
  return distance <= X_H_COVALENT

def cos_ahb_dists(d_ah, d_hb, d_ab):
  """
  given three distances, calculate cosine angle a-h-b
  d_ah: distance of atom a to hydrogen
  """
  return (d_ah**2 + d_hb**2 - d_ab**2)/(2*d_ah*d_hb)


def cos_index(coords_dict, index_a, index_b, d_ah, d_hb):
  """
  calculates the cosine of bond angle a-h-b based on
  bond lengh a-h and h-b
  """
  d_ab = dist(coords_dict,index_a, index_b)
  return cos_ahb_dists(d_ah, d_hb, d_ab)


def filter_hx_dists(h_index, x_index_list, coords_dict):
  """
  given ONE hydrogen index, loop through the x_indexes (N,O,F)
  calculate distance and
  (1)filter by X_H_HB to rule out impossible H-bond
  (2)filter by X_H_COVALENT to rule out impossible Hydrogen case
  rtype: [(x1, dis_x1-H),(x2, dist_x2-H)...] filtered
  """
  hx_dists = []
  for x_idx in x_index_list:
    dist_xh = dist(coords_dict, h_index, x_idx)
    if dist_xh >= X_H_HB:
      continue
    hx_dists.append((x_idx, dist_xh))
  #if qualified number of distance < 2
  # return empty list
  if len(hx_dists) < 2:
    return []
  #put the lowest distance first, which is covalent case
  hx_dists.sort(key = lambda elem: elem[1])
  if x_h_link(hx_dists[0][1]):
    return hx_dists
  return []

def gen_legit_list(h_index_list, x_index_list, coords_dict):
  """
  based on whole h_list, get legit combinations
```

```
  {(h1,coval_x, coval_dist):[(x2, dist2), ..], (h4 ,coval_x1,
coval_dist1):[(x3, dist3),..]..}
  """
  hx_legit = {}
  for h_idx in h_index_list:
    hx_dists = filter_hx_dists(h_idx, x_index_list, coords_dict)
    if hx_dists == []:
      continue
    coval_idx, coval_dist = hx_dists[0]
    hx_legit[(h_idx, coval_idx, coval_dist)] = hx_dists[1:]
  return hx_legit


def get_hbond(hx_legit, coords_dict):
  """
  {(h1,coval_x1): [y1, y2], (h2, coval_x2): [y4]...}
  """
  hbonds = {}
  for h_a_pair, h_b_dists in hx_legit.items():
    idx_h, idx_a, d_ah = h_a_pair
    #creat filter that cosine a-h-b <= -0.5 (>=120degree) that qualifies h-
bond angle
    filt = list(map(lambda x: cos_index(coords_dict, idx_a, x[0], d_ah,
x[1])<= -0.5, h_b_dists))
    if sum(filt) > 0: #not empty, at least one hbond
      hbonds[(idx_h, idx_a)] = list(map(lambda x: x[0], compress(h_b_dists,
filt)))
  return hbonds



def hb_pattern(coords_dict):
  """
  get total h_bonding_expression for one geometry file
  """
  h_index_list, x_index_list = subset_by_type(coords_dict)
  hx_legit = gen_legit_list(h_index_list, x_index_list, coords_dict)
  hbonds = get_hbond(hx_legit, coords_dict)
  group_pattern = []
  for key, value in hbonds.items():
    group_pattern.extend((key[0],key[1], x_idx) for x_idx in value)
  return ','.join(sorted(list(map(lambda x: format_hb(x[0],x[1],x[2],
coords_dict), group_pattern))))



def format_hb(h_idx, index_a, index_b, coords_dict):
  a_type = get_type(coords_dict, index_a)
  b_type = get_type(coords_dict, index_b)
  return '{0}{1}-H{2}_{3}{4}'.format(a_type, str(index_a+1), str(h_idx+1),
b_type, str(index_b+1))


def get_pair_distances(coords_dict, idx_pairs):
  """
  get index pair distances
  return List[float]
```

```
    """
    pair_dist = []
    for pair in idx_pairs:
      pair_dist.append(dist(coords_dict,pair[0],pair[1]))
    return pair_dist


### MACHINE LEARNING FUNCTION

def get_rmsd_score(dists_data,list_k):
    """
    calculate root mean inertia (sum of squared distances of samples to their
    closest cluster center) score
    type dists_data: DataFrame
    type K: List[int]
    """
    rmsd = []
    score =[]
    assert 0 <len(list_k) <= dists_data.shape[0]
    for k in list_k:
      kmeanModel = KMeans(n_clusters=k,max_iter=500).fit(dists_data)
      kmeanModel.fit(dists_data)
      rmsd.append(math.sqrt(kmeanModel.inertia_/len(dists_data)))
      score.append(kmeanModel.score(dists_data))
    return rmsd, score

def max_dist_to_line_k(score, list_k):#list
    """
    draw a line between the first and last point on the curve
    and measure the distance of each point to the line
    return the k corresponding to the max distance to the line
    """
    n_points = len(list_k)
    all_coords = np.vstack((range(n_points),score)).T
    np.array([range(n_points), score])
    first_point = all_coords[0]
    line_vec = all_coords[-1] - first_point
    lineVecNorm = line_vec / np.sqrt(np.sum(line_vec**2))
    vecFromFirst = all_coords - first_point
    scalarProduct = np.sum(vecFromFirst * np.matlib.repmat(lineVecNorm,
n_points, 1), axis=1)
    vecFromFirstParallel = np.outer(scalarProduct, lineVecNorm)
    vecToLine = vecFromFirst - vecFromFirstParallel
    distToLine = np.sqrt(np.sum(vecToLine ** 2, axis=1))
    return list_k[np.argmax(distToLine)]


def get_best_k(rmsd, score, list_k, max_k):
    """
    based on polynomial fitted curve, find the best k correspond to elbow point
    """
    assert len(rmsd) == len(list_k)
    rmsd_smoothed = signal.savgol_filter(rmsd,window_length=7,polyorder=5)
    score_smoothed = signal.savgol_filter(score,window_length=7,polyorder=5)
    k_names = ['rmsd','score', 'rmsd_smoothed','score_smoothed']
```

```
  k_values = list(map(lambda x: max_dist_to_line_k(x,
list_k),[rmsd,score,rmsd_smoothed,score_smoothed]))
  best_k = math.ceil(np.array(k_values).mean())
  #PLOT ELBOW
  plt.figure()
  plt.plot(list_k,rmsd,"ro-")
  plt.plot(list_k,rmsd_smoothed, "bx-")
  plt.xlabel('k')
  plt.ylabel('RMSD to centroid')
  plt.title('The Elbow Method showing the optimal k')

  plt.savefig(os.path.join(wdir,"elbow_maxk_"+str(max_k)+".png"))
  plt.close()

  for name,value in zip(k_names, k_values):
     print('best k from {0}:{1};'.format(name, str(value)))
  print('Use {} for k mean clustering'.format(str(best_k)))
  return best_k


def best_k_cluster(best_k,max_iteration, data):
  """
  Based on pre-determined K, perform kmeans clustering
  return clustering group assignment and
  H-bond pattern
  """
  km = KMeans(n_clusters = best_k, max_iter = max_iteration)
  result = km.fit_predict(data)
  result_df = pd.DataFrame({"cluster":result})
  hbond_kmean = pd.concat([hb_df.iloc[:,[0,1]], result_df,hb_df.iloc[:,2]],
axis=1)
  hbond_kmean['hb_id'] = pd.Categorical(hbond_kmean['H-
bond'].astype(str)).codes
  hbond_kmean = hbond_kmean.iloc[:,[0,1,2,4,3]]
  hbond_kmean.sort_values(by = ['cluster','hb_id','energy'], inplace = True)
#sort by cluster, then hb_id, then energy
  return hbond_kmean



if __name__ == '__main__':
  if len(sys.argv) < 4 :
    print("Usage: python hb_cluster_sys_demo.py energyFile keyAtomFile maxK")
    sys.exit()

  energy_file = sys.argv[1]
  key_atom_fname = sys.argv[2]
  MAX_K = int(sys.argv[3])

  hb_energy_file = "hb_energy_" + energy_file
  distance_file_out = "distance_read_from_energy"
  ##get reference atoms
  reference_atom_num = read_reference(key_atom_fname)
  index_pairs = get_index_pair(get_indexes(reference_atom_num))
  ##Read the energy file and store (filename, energy) pairs
  #process filename_energy and generate list of coordinates
  energy_table = read_energy_file(energy_file)
```

```
  fns, energies = zip(*energy_table) #unzip to get two tuples
  file_coords = list(map(read_geomfile, fns))

  ##apply h-bond function on each filename (map), convert to df1, then write
out[filename, energy, h-bond]
  #combine fname, energy, hbond pattern
  hb_df = pd.DataFrame(list(zip(fns, energies,
list(map(hb_pattern,file_coords)))))
  hb_df.columns = ['file','energy','H-bond']
  hb_df.to_csv(os.path.join(wdir,hb_energy_file), sep="\t",index = False)

  ##apply get-distances pairs on each filename(map), convert to df2, then
write[filename, pair1, pair2,..]
  #filename from fns, list[dists] from file_coords
  pair_dists = list(map(functools.partial(get_pair_distances,idx_pairs =
index_pairs), file_coords))
  fn_dists = [[fns[i]] + pair_dists[i] for i in range(len(fns))] #extend to
flattened rows
  dists_df = pd.DataFrame(fn_dists)
  dists_df.columns = gen_distance_colnames(index_pairs)
  dists_df.to_csv(os.path.join(wdir,distance_file_out), sep = '\t', index =
False)

  ##Machine learning based on existing df
  print("Now starting clustering")
  list_k = list(range(1,MAX_K))
  dists_data = dists_df.iloc[:,1:]
  #print(dists_data.shape)
  labels = dists_df.iloc[:,0]
  rmsd, score = get_rmsd_score(dists_data, list_k)
  best_k = get_best_k(rmsd, score, list_k, MAX_K)
  hbond_kmean = best_k_cluster(best_k , 500, dists_data)

hbond_kmean.to_csv(os.path.join(wdir,"kmean"+str(best_k)+"_hb_energy_sorted_w
_hbid.tsv"),sep="\t",index=False )
```

2. *Molecular Dynamics Preparation and Processing Scripts*

Three processing scripts to (A) prepare geometries (built from GaussView) to run molecular dynamics trajectories by MOPAC program interfaced by cuby4 framework; (B) extract desired snapshots by interval steps and format for geometry optimization jobs by MOPAC program interfaced by cuby4 framework; (C) extract atomic coordinates from optimized geometry output files, and format to Gaussian Z-matrix input files that are compatible with geometry clustering script and subsequent Gaussian calculations.

(A) Prepare input files to run sampling MD trajectories in parallel

Required input: *.gjf or *.com (Gaussian) input files
Parameters: original filetype, max cycles and temperature
Output: individual folders that contain corresponding *.xyz and *.yaml files; a parallel task list; a submission script to Hyak's scheduler to allow parallel computing.

```
#!/bin/bash

FILE_TYPE=$1
MAX_CYCLE=$2
TEMPERATURE=$3
module load anaconda3_4.3.1

PYTHON_FILE="$FILE_TYPE" PYTHON_CYCLE="$MAX_CYCLE" PYTHON_TEMP="$TEMPERATURE"
/gscratch/sw/anaconda-4.3.1/python3/bin/python3.6  - << END

import sys
import os
import subprocess

FILETYPE = os.environ['PYTHON_FILE']
MAXCYCLE = os.environ['PYTHON_CYCLE']
TEMP = os.environ['PYTHON_TEMP']
assert FILETYPE and MAXCYCLE and TEMP, "python required environment variable
empty\nUsage ./gen_task_md.sh filetype max_md_cycle temperature(K)"


def get_list():
  listofFiles = []
  try:
    for entry in os.listdir(os.getcwd()):
      #print(entry)
      if entry.endswith(FILETYPE):
        listofFiles.append(entry)
    return listofFiles

  except FileNotFoundError:
    print(FILETYPE+"NOT FOUND IN "+os.getcwd())
    sys.exit(1)


# read charge multiplicity
def convert_file(fname):
  '''
  input type str
  rtype tuple
  '''
  fn_out = fname.split('.')[0]+'.xyz'
  content=[]
  with open(fname) as fin:
    for line in fin:
      stripped_line = line.strip()
      if stripped_line.startswith('%') or stripped_line.startswith('#'):
        continue
      if stripped_line == '':
        continue
      arr = stripped_line.split()
      if len(arr) == 4:
        content.append(arr)
      if len(arr) == 2 and len(arr[1]) == 1:
        try:
```

```
          multiplicity = int(arr[1])
          charge = int(arr[0])
        except ValueError:
          print("Can not find charge/multiplicity")
          sys.exit()
  atom_num = len(content)
  with open(fn_out, 'w') as fout:
    fout.write(str(atom_num)+'\n')
    fout.write('\n')
    fout.write('\n'.join('\t'.join(item for item in line) for line in
content))
  return charge, multiplicity

def write_yaml(folder, charge, multiplicity):
  with open(os.path.join(folder,'anneal.yaml'), 'w') as fout:
    fout.write('''\
job: dynamics
geometry: {0}.xyz
maxcycles: {1}
charge: {2}
multiplicity: {3}
timestep: 0.001
interface: mopac
mopac_precise: yes
mopac_peptide_bond_fix: yes
method: pm6
modifiers: dispersion3, h_bonds4
modifier_h_bonds4:
  h_bonds4_scale_charged: no
  h_bonds4_extra_scaling: {{}}
init_temp: {4}
thermostat: berendsen
thermostat_tc: 0.05
temperature: {4}
'''.format(folder,str(MAXCYCLE),str(charge),str(multiplicity), str(TEMP)))
    if multiplicity != 1:
      fout.write("spin_restricted: uhf")
      fout.write("scf_cycles: 1000")

def prep_file_md(fname,charge, multiplicity):
  folder = fname.split('.')[0]
  fxyz = folder + '.xyz'
  subprocess.call(['mkdir',folder])
  subprocess.call(['mv',fxyz, folder])
  write_yaml(folder, charge, multiplicity)

def write_sbatch(files):
  with open('multiple_md.sh','w') as batch:
    batch.write('''\
#!/bin/bash
#SBATCH --job-name=mds
#SBATCH --nodes=1
#SBATCH --time=72:00:00
#SBATCH --mem=100Gb
#SBATCH --workdir={0}
#SBATCH --partition=chem
#SBATCH --account=chem
```

```
module load parallel_sql
module load contrib/mopac16
source {1}/.rvm/scripts/rvm

cat md_tasklist|psu --load

ldd /sw/contrib/cuby4/cuby4/classes/algebra/algebra_c.so > ldd.log

parallel_sql --sql -a parallel --exit-on-term -j {2}
'''.format(os.getcwd(),os.environ['HOME'],str(len(files))))

def write_tasklist(files):
   with open('md_tasklist','w') as tasklist:
     tasklist.write('\n'.join('cd {}; cuby4 anneal.yaml
&>LOG'.format(os.getcwd()+'/'+fname.split('.')[0]) for fname in files))


if __name__ == '__main__':

  #print(os.environ)
  print(FILETYPE)
  files = get_list()

  for fname in files:
    charge, multiplicity = convert_file(fname)
    prep_file_md(fname, charge, multiplicity)

  write_tasklist(files)
  write_sbatch(files)


  #also writes a filelist
  print("write out a 'filelist'")
  with open('filelist', 'w') as flist:
    flist.write('\n'.join(list(map(lambda x: x.split('.')[0],files))))



END
if [[ $? = 0 ]]; then
  echo "Please keep 'filelist' for future use"
  echo "Please run sbatch multiple_md.sh to submit for parallel md"
  echo "change partition name and runtime if needed"
else
  echo "failure:$?"
fi
```

(B) Extract snapshots at desired interval steps from sampling MD trajectories and prepare files for parallel batch optimization

Required input: trajectory anneal files from cuby4; file name list
Parameters: document name that contains list of filenames; interval steps

Output: individual folders that contain corresponding snapshot geometry *.xyz and *.yaml files; a file name list; a submission script to Hyak's scheduler to allow parallel computing.

```
#!/bin/bash
module load anaconda3_4.3.1

PYTHON_FILELIST="$1" PYTHON_INTERVAL="$2" /gscratch/sw/anaconda-
4.3.1/python3/bin/python3.6  - << END

import os, subprocess, sys
import functools, itertools

FILELIST= os.environ['PYTHON_FILELIST']
INTERVAL = os.environ['PYTHON_INTERVAL']
assert FILELIST and INTERVAL,'usage: traj_process.sh filelist interval_steps'


def get_prefix(filelist):
  flist = []
  try:
    with open(filelist) as fin:
      for line in fin:
        if not line.strip():
          continue
        flist.append(line.strip())
    return flist
  except OSError:
    print(filelist+ 'not found!')
    sys.exit(1)

def file_exist(fname):
  '''
  check if a path exists
  '''
  if not os.path.exists(fname):
    print(fname + ' does not exist')
    return False
  return True

def format_tasklist(tasklist):
  path = os.getcwd()
  if not tasklist:
    return #return None type if tasklist is empty
  return map(lambda x : 'cd {}; cuby4
inp.yaml&>LOG'.format(os.path.join(path,x)),tasklist)

def charge_multp_temp(prefix,anneal):
  #by default, if not found specified in the anneal
  '''
  rtype: charge: str(int)
  rtype: multipl: str(int)
  rtype: temp: str(int)
  '''
  multipl = '1'
  with open(os.path.join(prefix,anneal)) as fin:
    for line in fin:
```

```
      curr_line = line.strip()
      if line.startswith('charge'):
        charge = line.split(':')[1].strip()
        continue
      if line.startswith('temperature'):
        temp = line.split(':')[1].strip()
        continue
      if line.startswith('multiplicity'):
        multipl = line.split(':')[1].strip()
        continue
    return charge, multipl, temp


def split_traj(prefix,traj,temp,interval_steps, charge, multp):
  """
  interval_steps: int, others: str
  """
  with open(os.path.join(prefix,traj), 'r') as fin:
    counter = 1
    l1 = fin.readline().strip()
    assert l1, "empty 1st line"
    atom_number = int(l1)
    print(atom_number)
    current_snap = list(itertools.islice(fin,1 , atom_number+1))
    #the original 1st line has been read, so skip current 1st line
    skip = (atom_number+2)*(interval_steps-1)+2
    tasklist = []
    while fin:
      if not current_snap:
        break
      sub_folder = '{}_{}K_snap_{}'.format(prefix, temp, str(counter))
      #(1)mkdir of new folder
      subprocess.run(['mkdir',sub_folder])
      #(2)save geom to prefix_temp_snap_n.xyz
      write_sub(current_snap, atom_number, sub_folder)
      #(3)write inp.yaml
      write_yaml(sub_folder, charge, multp)
      #(4)add one task to tasklist
      tasklist.append(sub_folder)
      counter +=1
      current_snap = list(itertools.islice(fin,skip,skip+atom_number))
  return tasklist

def write_sub(snap, atom_num, sub_folder):
  content = map(lambda x: x.strip().split()[:4], snap)
  if not file_exist(sub_folder):
    print(sub_folder + 'has not been made yet')
    return #return nonetype
  with open(os.path.join(sub_folder, sub_folder +'.xyz'),'w') as fout:
    fout.write(str(atom_num)+'\n'+'\n')
    fout.write('\n'.join('\t'.join(elem for elem in line) for line in
content))

def write_yaml(sub_folder, charge, multipl):
  if not file_exist(sub_folder):
    return
  with open(os.path.join(sub_folder,'inp.yaml'),'w') as fout:
```

```
    fout.write('''job: optimize
geometry: {0}
charge: {1}
multiplicity: {2}
maxcycles: 2000
print: timing
interface: mopac
mopac_precise: yes
mopac_peptide_bond_fix: yes
method: pm6
modifiers: dispersion3, h_bonds4
modifier_h_bonds4:
  h_bonds4_scale_charged: no
  h_bonds4_extra_scaling: {{}}
'''.format(sub_folder + '.xyz',charge, multipl))


def process_one_traj(prefix, interval_steps):
  """
  interval_steps: int
  """
  #check if prefix directory exists
  if not file_exist(prefix):
    return #nonetype
  origin_flist = os.listdir(prefix)
  anneal, traj = None, None #initial nonetype
  for fn in origin_flist:
    if fn.startswith('anneal'):
      anneal = fn
    if fn.startswith('trajectory'):
      traj = fn
  if not anneal or not traj:
    print('no anneal or traj file found for {}'.format(prefix))
    return
  #get charge mulplicity and  temperature
  charge, multp, temp = charge_multp_temp(prefix, anneal)
  #read traj, split up and write subfiles, return tasks for final tasks write
out
  return split_traj(prefix, traj, temp, interval_steps, charge, multp)


def write_sbatch(task_number):
  with open('pm6_parallel.sh', 'w') as fout:
    fout.write('''#!/bin/bash
#SBATCH --job-name=pm6s
#SBATCH --nodes=1
#SBATCH --time=72:00:00
#SBATCH --mem=100Gb
#SBATCH --workdir={}
#SBATCH --partition=ilahie
#SBATCH --account=ilahie

#module load parallel_sql
module load parallel-20170722
module load contrib/mopac16
source {}/.rvm/scripts/rvm
```

```
ldd /sw/contrib/cuby4/cuby4/classes/algebra/algebra_c.so > ldd.log
cat pm6_tasks | parallel -j 28

'''.format(os.getcwd(),os.environ['HOME']))

if __name__ == '__main__':
  file_list = FILELIST
  interval = int(INTERVAL)
  prefix_list = get_prefix(file_list)
  print(prefix_list)
  all_tasks = list(map(functools.partial(process_one_traj, interval_steps =
interval),prefix_list))
  task_number = sum(len(i) for i in all_tasks if i is not None)
  with open('pm6_tasks','w') as fout:
    fout.write('\n'.join('\n'.join(format_tasklist(i)) for i in all_tasks if
i is not None))
  write_sbatch(task_number)
  sys.exit()

END
if [[ $? = 0 ]]; then
  echo "Please run sbatch pm6_parallel.sh for parallel pm6 optimization"
  echo "change partition name and run time if needed"
  echo "tasklist is stored in 'pm6_tasks'"
else
  echo "failure:$?"
fi
```

(C) Batch extract coordinates from optimized snapshots and format as Gaussian input files

Required input: individual folder that contains optimized *.xyz file, LOG file
Parameters: output file extension, desired functional and basis set
Output: formatted optimized geometries compatible with Gaussian calculations and conformation clustering

```
#!/bin/bash

module load anaconda3_4.3.1

PYTHON_FTYPE="$1" PYTHON_FUNCTIONAL="$2" PYTHON_BASISSET="$3"
/gscratch/sw/anaconda-4.3.1/python3/bin/python3.6  - << END

import sys
import os
import subprocess

FTYPE = os.environ['PYTHON_FTYPE']
FUNCTIONAL = os.environ['PYTHON_FUNCTIONAL']
BASISSET = os.environ['PYTHON_BASISSET']
assert FTYPE and FUNCTIONAL and BASISSET, "python required variable
empty\nUsage ./extract_opted_pm6.sh filetype functional basis-set"
```

```
def file_exist(fname):
  '''
  check if a path exists
  '''
  if not os.path.exists(fname):
    print(fname + ' does not exist')
    return False
  return True

def read_charge_multp(folder):
  if not file_exist(folder):
    return
  flist = os.listdir(folder)
  inp = None
  for fn in flist:
    if fn.startswith('inp'):
      inp = fn
  if not inp:
    print('no original inp file found')
    return
  multipl = '1'
  with open(os.path.join(folder,inp)) as fin:
    for line in fin:
      curr_line = line.strip()
      if curr_line.startswith('charge'):
        charge = curr_line.split(':')[1].strip()
      if curr_line.startswith('multiplicity'):
        multipl = curr_line.split(':')[1].strip()
    return charge, multipl

def get_snap_folders():
  return sorted([d_name for d_name in os.listdir(os.getcwd()) if
os.path.isdir(d_name) and 'snap' in d_name])


def get_opt_fname(folder):
  if not file_exist(folder):
    return
  flist = os.listdir(folder)
  opt_fn = [i for i in flist if i.startswith('optimized')]
  if not opt_fn:
    print("{} not optimized".format(folder))
    return
  opt_fn = ''.join(opt_fn)
  return opt_fn

def get_xyz_energy(folder, charge, multipl):
  opt_fn = get_opt_fname(folder)
  if not opt_fn:
    return
  energy = None
  opt_xyz = []
  with open(os.path.join(folder,opt_fn)) as fin:
    for line in fin:
      if line.startswith('Energy'):
        energy = line.strip().split()[1]
        continue
```

```
      if not line.startswith('  '):
        continue
      if line.strip() == '':
        continue
      opt_xyz.append(line)
  with open(os.path.join('dftsubs',folder + '.' + FTYPE), 'w') as fout:
    fout.write('''%mem=128gb
%nproc=28
%Chk={0}.chk
#p opt {1}/{2}  scf=(xqc, tight)  pop=min

Complex {3}

{4} {5}
'''.format(folder,FUNCTIONAL,BASISSET,folder,charge, multipl))
    fout.write(''.join(opt_xyz))
    fout.write('\n')
  return energy

def format_energy(folder,energy):
  if not energy:
    return
  return '\t'.join([folder + '.' + FTYPE, energy])

def extract_all(folder_list):
  energy_content = []
  if not file_exist('dftsubs'):
    print('dir dftsubs not found!')
    sys.exit(1)
  for folder in folder_list:
    energy = get_xyz_energy(folder, charge, multipl)
    energy_content.append(format_energy(folder,energy))
  with open(os.path.join('dftsubs','pm6_energy_combine.txt'),'w') as fout:
    fout.write('\n'.join(line for line in energy_content if line))


if __name__ == '__main__':
  folder_list = get_snap_folders()
  if not folder_list:
    print('no snap folder found!')
    sys.exit(1)
  #only need to read charge, multiplicity once for all files
  charge, multipl = read_charge_multp(folder_list[0])
  print(charge, multipl)
  subprocess.call(['mkdir','dftsubs'])
  extract_all(folder_list)
  sys.exit()
END

if [[ $? = 0 ]]; then
  echo "extracted optimized geometries and energy info are in dftsubs folder"
  echo "please see 'pm6_energy_combine.txt' for energy profile"
else
  echo "failure:$?"
fi
```

CONTACT ANALYSIS PROCESSING SCRIPT

1. *Prepare input files to run measure jobs by cuby4 framework*

```
import sys
import numpy as np
import math
import re


def read_c_others(atom_file):
  try:
    with open(atom_file, 'r') as fin:
      raw_input = fin.read().splitlines()
  except FileNotFoundError:
    print('file not found')
  center_number = int(raw_input[0].split(':')[1])
  others = []
  block_pair = {}
  for block in raw_input[1:]:
    block_name = block.split(':')[0]
    block_atoms = list(map(int,block.split(':')[1].split(',')))
    block_pair.setdefault(block_name,
list(zip([center_number]*len(block_atoms), block_atoms)))
  return block_pair


def read_atoms(filename):
  '''
  rtype = list[str]
  '''
  with open(filename, 'r') as fin:
    atom_list = []
    atom_num = 1
    for line in fin:
      if line.strip() == '':
        continue
      if line.startswith("  "):
        arr = re.split(" +", line.strip())
        if len(arr) == 4:
          atom = arr[0]
          atom_list.append(atom + str(atom_num))
          atom_num += 1
    return atom_list


def get_name_number(num, atom_list):
  return atom_list[num-1]


def get_pair_names(block_pair,atom_list):
  '''
```

```
  rtype = dict
  '''
  pair_dict = {}
  for key, pairs in block_pair.items():
    name_corr_pairs ={tuple(map(lambda x : get_name_number(x, atom_list),
pair)): pair for pair in pairs}
    pair_dict.setdefault(key, name_corr_pairs)
  return pair_dict

def write_yaml_label(pair_dict):
  with open('measure.yaml','w') as yaml:
    yaml.write("job: measure\n")
    yaml.write("geometry: trajectory_anneal.xyz\n\n")
    yaml.write("measurements:\n")
    with open('labels_by_block.txt', 'w') as label:
      for block, vs in pair_dict.items():
        label.write(block + ':')
        label_list = []
        for pair,num in pair_dict[block].items():
          name = '-'.join(pair)
          num_to_write = '; '.join(tuple(map(str, num)))
          yaml_line = '    {}: distance({})\n'.format(name,num_to_write)
          yaml.write(yaml_line)
          label_list.append(name)
        label.write(','.join(label_list)+ '\n')


if __name__ == '__main__' :
  if len(sys.argv) < 3:
    print('usage: python write_measure.py atom_file example_input(gjf style
file)')
    sys.exit()
  atom_file = sys.argv[1]
  standard_input = sys.argv[2]
  block_pair= read_c_others(atom_file)
  atom_tp_num = read_atoms(standard_input)
  pair_dict=get_pair_names(block_pair, atom_tp_num)
  write_yaml_label(pair_dict)
```

2. *Process contact data by atom blocks*

```
import sys

FILTER = 4 #distance threshold for close contact

def read_label(labelFile):
  print('reading label file')
  with open(labelFile, "r") as fin :
    blocks = fin.read().splitlines()
    block_label = {}
    for line in blocks:
      block = line.split(':')[0]
      labels = line.split(':')[1].split(',')
```

```
      block_label.setdefault(block,labels)
    return block_label

def get_center(block_label):
  for v in block_label.values():
    pair=list(v)[0]
    common=pair.split('-')[0]
    return common


def read_log_filter(datalog, common):
  print("Reading data file")
  col_sum_dict = {}
  listCounts = []
  with open(datalog, "r") as fin:
    for line in fin:
      if line.startswith(" "):
        continue
      if line.startswith(common):
        colNames = line.strip().split()

        print("Got " + str(len(colNames)) + " columns!")
        continue
      strip_line = line.strip()
      if strip_line == '':
        break
      numbers = strip_line.split()
      counts = list(map(lambda x: float(x) <= FILTER, numbers))
      listCounts.append(counts)
  step_contact_bool = list(map(max, listCounts))
  count_col = list(zip(*listCounts)) #*sign to unpack nested row list/tuple
into columns
  sum_col = list(map(sum, count_col))
  #print(step_contact_bool)
  for i in range(len(colNames)):
    col_sum_dict[colNames[i]] = sum_col[i]
  return col_sum_dict, step_contact_bool


def total_contact_steps(step_contact_bool):
  return sum(step_contact_bool)


def sum_by_residue(block_label, col_sum_dict):
  block_sum={}
  label_sum = 0
  for block, labels in block_label.items():
    label_sum = sum([col_sum_dict[label] for label in labels])
    block_sum.setdefault(block, label_sum)
  return block_sum


def by_ratio(block_sum):
  block_fraction = {}
  print(list(block_sum.values()))
  sum_all_block = sum(list(block_sum.values()))
  print('sum = {}'.format(str(sum_all_block)))
```

```
  if sum_all_block == 0:
    return block_sum, sum_all_block
  for block, count in block_sum.items():
    block_fraction[block]=count/sum_all_block
  return block_fraction, sum_all_block

###TEMP WRITE OUT STYLE
def write_out_flat(block_fraction,sum_all_block,contact_steps, output):
  with open(output, 'w') as fout:
    fout.write('\t'.join(block_fraction.keys())+'\tTOTAL\tCONTACT_STEPS\n')
    fout.write('\t'.join(list(map(lambda x: "{0:.3f}".format(x),
block_fraction.values()))+[str(sum_all_block),str(contact_steps)]) +'\n')

if __name__ == '__main__':
  if len(sys.argv) < 4:
    print("python countResidue.py label_by_block.txt datalog output.txt")
    sys.exit()
  labelFile = sys.argv[1]
  datalog = sys.argv[2]
  output = sys.argv[3]
  filter_str = str(FILTER).replace('.','p')
  fout_name = output.split('.')[0]+ '_' + filter_str + '.txt'
  block_label = read_label(labelFile)
  center = get_center(block_label)
  col_sum_dict, step_contact_bool = read_log_filter(datalog, center)
  block_sum=sum_by_residue(block_label, col_sum_dict) #total count
  block_frac, sum_all_block = by_ratio(block_sum)
  contact_steps = total_contact_steps(step_contact_bool)
  print('contact_steps = {}'.format(str(contact_steps)))
  write_out_flat(block_frac, sum_all_block, contact_steps, fout_name)
```