The Structure and Properties of Weakly Bound Clusters

by

Ce Zhou

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemistry

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner   Dr. Derek Wilson

                             Associate Professor

                             York University, Toronto, ON, Canada

Supervisor   Dr. W. Scott Hopkins

                             Associate Professor

                             University of Waterloo, Waterloo, ON, Canada

Internal-external Member   Dr. Anh Pham

                             Assistant Professor

                             University of Waterloo, Waterloo, ON, Canada

Advisory Committee Member(s) Dr. J. Larry Campbell

                             Adjunct Associate Professor of Chemistry

                             University of Waterloo, Waterloo, ON, Canada

                             Dr. Marcel Nooijen

                             Professor

                             University of Waterloo, Waterloo, ON, Canada

AUTHOR'S DECLARATION

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis contains works which have been published in peer reviewed publications. These works have been acknowledged in the references section of this thesis. My contributions, relating to this thesis are summarized below:

The structure similarity function related content in Chapter 2 and the Chapter 3. The Protonated Serine Dimer in Gas Phase is reproduced from the following publication:

Zhou, C.; Ieritano, C.; Hopkins, W. S. Augmenting Basin-Hopping With Techniques From Unsupervised Machine Learning: Applications in Spectroscopy and Ion Mobility. *Front. Chem.* **2019**, *7* (August). https://doi.org/10.3389/fchem.2019.00519.

My contribution to this work includes conducting the first case study: The IR Spectrum of the Protonated Serine Dimer, and authoring the first draft of the article except for the part of case study 2.

Abstract

In this thesis, two novel methods are introduced to advance the study of gas phase clusters. The structure similarity method is a computational technique that is able to quantify the structure difference for a pair of isomers, with a structure interpolation technique capable of finding intermediates in-between the isomer pair. A new experimental method, which couples differential mobility spectrometry with ultraviolet photodissociation spectroscopy (DMS-UVPD), is also developed and tested. Three test cases are discussed herein. These test cases showcase new theoretical techniques for mapping and visualizing potential energy surface (PES) and finding transition state (TS) structures, as well as experimental techniques of measuring UVPD spectra of DMS-MS isolated ion populations. Introduce of structure similarity, a technique developed for unsupervised machine learning (ML), enables effective domain of mapping PESs, which may subsequently be used to interpret experimental observations for systems of high geometric complexity. The experimental DMS-UVPD technique is shown capable of isolating ion species such that UVPD spectra may be recorded for characterization of analytes of interest. For the test cases described herein, these new methods provide meaningful (sometimes anti-intuitive) directions for future work.

For the structure similarity method, its PES mapping capability is tested in Chapter 3 with a collection of protonated serine dimer cations, $[Ser_2 + H]^+$ to rationalize its infrared multiphoton dissociation (IRMPD) spectrum. Eventually, the spectral carrier is assigned to a non-global minimum (GM) isomer based on the partitioning information of the PES and spectral similarity. In Chapter 4, the accompanying structural interpolation method is employed to find TSs that can rationalize a regioselective alkylation reaction between a barbituric acid derivative and an alkyl-tricarbastannatrane complex. By combining the interpolation method together with chemical intuition, a total of 3 reaction channels are found, and the regioselectivity of the alkylation is identified as a kinetic effect. In Chapter 5, an acylhydrazone (AY) derivative, a photoswitch candidate, is examined using the DMS-UVPD technique. Experimentally, the protonated $[AY + H]^+$ cation is injected into the instrument for DMS separation and laser interrogation, while theoretically, a number of neutral and protonated isomers are sampled. Eventually, separation of the ion population is observed and attributed to some ion-solvent cluster. Four isomers are found from theoretical calculation that may account for the UVPD spectra.

Acknowledgements

First, I would like to thank Dr. W. Scott Hopkins for offering the chance to work and study in his group as a graduate student. His one-liner word of courage grants enormous confidence in me when I tried to build up that seemingly crazy structural interpolation method, which I used later in two of my works presented in this thesis. He is also a very patient person, not only to the mistakes I made, but also to the rest of the research group. Besides, I would also like to thank Dr. Terry McMahon, who interviewed me when I was back in the Shandong Normal University in China as a third-year undergraduate and gave me a second chance to touch Chemistry. Otherwise I may become a programmer somewhere without touching any of the chemistry related work I have done here. I would also thank Dr. Larry Campbell for his help in the DMS work and Dr. Marcel Nooijen for his excellent courses in modern theoretical chemistry. In addition, I would like to thank Dr. Eric Fillion and Dr. Mónica Barra for providing excellent research topics for my thesis.

Besides, I would also thank the past and present members of the Hopkins group. Dr. Steve Walker was one source of my motivation to write the DPE program to extract DMS data, Dr. Neville Coughan taught me a lot on the principles and the practices of the OPO laser, which helped me eventually to write up an automation script for laser experiment data extraction. Dr. Jeff Crouse helped me quite a lot in understanding the theoretical concepts of quantum mechanics via constructive conversations in his office. Weiqiang (future Dr. Fu) is a reliable person when I need help in my personal social affairs and is the first user of my machine learn algorithm in his research. I would also thank Dr. Xinda Lu, Songhao Bao, and Feng Li for being my close friends during my course of graduation.

Last but not least, I sincerely appreciate the concerns from my families and friends during the hard time when this thesis is completed. I would also like to thank my parent who gave the initial push on me to studying abroad and funded me through my undergraduate study.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIM	Atom In Molecule
AH	Adiabatic Hessian
AHAS	Adiabatic Hessian After Step
APCI	Atmospheric Pressure Chemical Ionization
BCP	Bond Critical Point
BH	Basin Hopping
BIRD	Blackbody Infrared Radiative Dissociation
CBS	Complete Basis Set
CC	Coupled Cluster
CCP	Cage Critical Point
CCSD	Coupled Cluster Single and Double
CCSD(T)	Coupled Cluster Single-Double and Perturbative Triple
CID	Collision Induced Dissociation
CISD	Configuration Interaction Singles and Doubles
CP	Critical Point
CV	Compensation Voltage
DCE	1,2-dichloroethane
DFT	Density Functional Theory
DG	Disconnectivity Graph
DIANA	Divisive Analysis
DIIS	Direct Inversion of the Iterative Subspace
DM	Distance Matrix
DMS	Differential Mobility Spectrometry

DLPNO	Domain-based Local Pair Natural Orbital
ECD	Electron Capture Dissociation
EI	Electron Ionization
EOM-CC	Equation Of Motion Coupled Cluster
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
FCI	Full Configuration Interaction
FTICR	Fourier Transform Ion Cyclotron Resonane
FWHM	Full Width at Half Maximum
GC	Gas Chromatography
GEDIIS	Generalized Energy-represented Direct Inversion of the Iterative Subspace
GGA	General Gradient Approximation
GM	Global Minimum
GTO	Gaussian Type Orbital
HC	Hierarchical Clustering
HF	Hartree-Fock
HWHM	Half Width at Half Maximum
ICP	Inductively Coupled Plasma
IMHB	Intramolecular Hydrogen Bond
IMS	Ion Mobility Spectrometry
IR	Infrared
IRC	Intrinsic Reaction Coordinate
IRMPD	Infrared Multiple Photon Dissociation
IVR	Intramolecular Vibrational (Energy) Redistribution
LC	Liquid Chromatography

LDA	Local Density Approximation
MALDI	Matrix-Assisted Laser Desorption/Ionization
MDS	Multi-Dimensional Scaling
ML	Machine Learning
MM	Molecular Mechanics
MO	Molecular Orbitals
MS	Mass Spectrometry
NCI	Non-Covalent Interaction
NCP	Nuclear Critical Points
NDO	Natural Difference Orbital
NMR	Nuclear Magnetic Resonance
OPO	Optical Parametric Oscillator
PAO	Projected Atomic Orbital
PCM	Polarizable Continuum Model
PELE	Python Energy Landscape Explorer
PES	Potential Energy Surface
PNO	Pair Natural Orbital
QST	Quadratic Synchronous Transit
QTAIM	Quantum Theory of Atom In Molecules
RCP	Ring Critical Point
RDG	Reduced Density Gradient
SCF	Self-Consistent Field
SID	Surface-Induced Dissociation
STEOM	Similarity Transformed Electron Of Motion
SV	Separation Voltage

TDDFT	Time-Dependent Density Functional Theory
TOF	Time-of-Flight
Triptych	Tricarbastannatrane
TS	Transition State
UVPD	Ultraviolet and Visible Photodissociation
UV-Vis	Ultraviolet and Visible
VG	Vertical Gradient
VH	Vertical Hessian
ZPE	Zero-Point Energy

Chapter 1
Introduction

Clusters are aggregates of atoms, ions, or molecules, with a size ranging between a diatomic system and bulk matter. Formation of weakly bound clusters is driven mainly by intermolecular electrostatic interactions such as charge-dipole, hydrogen bonding, and cation-π interactions,[1–3] in contrast to the relatively strong intramolecular interactions present within a molecule. Understanding the impact of these interactions is crucial in determining the structural evolution and properties of cluster series, and eventually in developing new materials.

Research on weakly-bound clusters usually involves both experimental and theoretical approaches in tandem. Experimentally, mass spectrometry (MS) is the central method to the detection of analyte species owning to its high resolution and low detection limit.[4] A variety of research scheme can be generated by coupling MS with other components such as ion source,[5] ion mobility,[6] and action spectroscopy.[7] The results of these experimental investigations are usually interpreted and understood by comparing to quantum chemical calculations, where information on the thermodynamics, electronic structure, and energetics of the cluster systems can be obtained. In this thesis, a new methodology for studying cluster systems in the gas phase is presented. First, an overview of the experimental techniques used throughout the thesis and the instrumentation setup required for these experiments to be performed is provided. Afterwards, the theoretical models employed within this thesis is outline, to gain further understanding of the experimental results.

The focus of this research centres around ions in the gas phase as most compounds can be easily ionized and gas phase ions have negligible matrix effects. Whereas ions can be quantified with great sensitivity and selectivity using MS techniques, coupling MS with other means for orthogonal characterization enables the investigation of isomeric species, which are otherwise indistinguishable by MS alone. Differential mobility spectrometry (DMS), a variant of ion mobility spectrometry, allows for the study of ion behaviour under alternating electric field strength. More importantly, the study of ions under atmospheric pressure in the DMS and the incorporation of solvent molecules in the carrier gas provide opportunities to investigate the gas-phase behavior and properties of the bare ion and its corresponding ion-solvent cluster species. The inclusion of action spectroscopy techniques to the DMS-MS setup further expands the ability to characterize isomeric species and clusters by probing their electronic structures. In this work, a state-of-the-art instrumentation is demonstrated, in which a

commercial mass spectrometer that supports DMS is coupled with a bench-top optical parametric oscillator (OPO) laser system. This instrumentation setup enables ultraviolet-visible photodissociation (UVPD) experiments over DMS-selected ions. In addition, an accompanying software package is also developed to enable high-performance data extraction and processing in different projects.

Extensive use of computational chemistry calculations throughout the thesis allows for rationalization of experimental observations. In most instances, structural sampling is performed to identify energetically low-lying conformers of a given species. The structures of the low-lying isomers are then obtained and used in further ab initio calculations to determine valuable structural information (*e.g.*, binding motifs) for the characterization of cluster species. Extraction of structural information from experimental studies to complement computational findings requires geometric analysis, *i.e.*, finding structural similarities. Herein, to enable rapid structural comparison, a pairwise dissimilarity function between molecular structures is proposed as an alternative means to accomplish quantitative structure comparison in a consistent and automatic manner. In addition, unsupervised machine learning (ML) methods, namely hierarchical clustering (HC) and multidimensional scaling (MDS) is introduced to visualize the relative distribution of the isomers within the conformation space. MDS provides a simplified method for the classification of isomers and their relative distribution, providing information that can be used as an orthogonal argument to the relative energies of the isomers. In addition, a structure interpolation mechanism is also introduced, with its capability in finding transition states (TSs) demonstrated.

The rest of this thesis is structured as follows: Chapter 2 introduces the theoretical and experimental approaches that are involved. In Chapters 3-5, three different molecular systems are analysed with the newly implemented theoretical and experimental methods applied. Details of each work are discussed in the corresponding chapters.

In Chapter 3, the dissimilarity functions are tested with protonated serine dimer ($[Ser_2+H]^+$) clusters, a system with well documented IRMPD spectra.[8,9] However, its global minimum (GM) structure has not been firmly determined as calculations under different level of theory suggests a different GM structure. In this work, the dimer structures in literature are collected and augmented with additional BH simulations. The isomers are further optimized with high level DFT calculations and corrected the electronic energy with that from single point energy corrected at DLPNO-CCSD/def2-TZVPP level. The electronic calculation results in a total of 40 low-lying isomers and the resulting GM matches that

form Seo, *et al.*[8] However, it is also found that the second lowest-energy isomer (Isomer 2) matches the GM proposed by Kong, *et al.*[10] Its relative Gibbs energy is only +5.6 kJ/mol above the calculated global minimum at the DFT level, and +1.3 kJ/mol with CCSD corrected electronic energy. This partially explains the discrepancy in the identity of the GM from different literature sources, as the energy between the first and the second lowest isomers might be quite close. Considering the possible error in the relative energy, however, there is no deterministic assignment for the GM from a pure relative energy argument. Additionally, the structural dissimilarities of the 40 isomers are calculated using the distance matrix (DM) distance defined earlier and projected the resulting DM using hierarchical clustering (HC) and multidimensional scaling (MDS). The HC gave four principle groups, from which four binding motifs are extracted from the representative structures. Both the HC and MDS suggests that the GM and Isomer 2 belong to basins that are distant from one another on the potential energy landscape. By calculating the cosine distance between the scaled harmonic IR spectrum of the 40 isomers with the IRMPD spectra, Isomer 2 is found to give the best match among all isomers and is thus assigned as the spectral carrier.

In Chapter 4, a theoretical investigation is carried out in explaining the regioselectivity of a recently reported alkylation reaction between barbituric acid with 3-methylbut-2-en-1-yl tricarbastannatrane.[11] The purpose of choosing this project is to test the structural interpolation method as well as to demonstration its capability in dealing complicated TS search problem. The reaction shown here tests the method in two ways: (1) the system is geometrically large for quantum calculation but is electronically simple, and (2) its reactive center is proposed as a 6- or 8-membered ring, and the initial guess is not easy to construct from intuition.

Experimentally, the formation of the thermodynamically favored product is not observed.[11,12] Theoretically, the initial guesses of the TSs are constructed with the aid of a dissimilarity function-based structural interpolation approach, and three types of TSs are found after geometric optimizations using a high-level density functional theory (DFT) method. The result indicates that the absence of the thermodynamically favored product is a kinetic effect.

In Chapter 5, a joint theoretical and DMS-UVPD experiment is conducted for (E)-N'-benzylidenebenzohydrazide, an acylhydrazone photoswitch (abbreviated AY) that has been studied previously in the condensed-phase.[13] The phothswich is a type of molecule that can undergo rapid photoisomerization upon exposure to light of certain wavelength.[13] They have wide application in

scenarios in which the state/property of a system is controlled with light irradiation.[14–16] Meanwhile, development of new photoswitches suffers from the absent of effective theoretical model, which involves understanding of the PES.

Experimentally, separation of the gaseous $[AY + H]^+$ ions are achieved using differential mobility spectrometry (DMS), under both a pure nitrogen and a methanol (1.5%) modified environment. Further experimental investigations indicate that the resolved ion populations show almost identical response to the experiments conducted and are thus attributed to some ion solvent clusters of $[AY + H]^+$. Theoretically, a number of $[AY + H]^+$ low energy conformers under both gas phase and methanol solvated condition are examined using a combination of DFT calculations with coupled cluster singles and doubles (CCSD) single point energy corrections. A total of nine relevant $[AY + H]^+$ isomers and eight TSs are found, and the resulting PES model suggests four possible species that are subject to kinetic trapping. Further comparison of their UV-Vis absorption spectra with the experimental UVPD spectra show that the UVPD can be effectively covered by the theoretical spectra.

To summarize, this thesis presents the efforts to develop novel experimental and theoretical methodologies to study weakly bound clusters, including coupling DMS-MS with UVPD, and introducing unsupervised machine learning into the routine of mapping PESs. This work demonstrates the efficacy of these approaches and provides directions for future efforts in this regard.

Chapter 2
Experimental and Theoretical Methods

2.1 Mass Spectrometry and Associated Experimental Techniques

There are two essential tasks in the paradigm of gas phase species analysis: one is to bring the analyte of interest into gas phase, and the other is its characterization. In principle, the analyte can be brought to the gas phase in its neutral form and characterized spectroscopically, as is in the analysis of iodine vapor.[17] However, this scheme cannot be generalized to arbitrary analytes as most compounds are not volatile and chemically stable enough. In contrast, gas phase ions can be easily generated from solution for most substances and guided with electro-magnetic force through vacuum. This grants mass spectrometry (MS) and its associated techniques tremendous flexibility in analyzing gaseous ions.

2.1.1 A General Introduction to Mass Spectrometry.

A diagram of modern MS analysis scheme is shown in Figure 2-1. When fitted to the two-component paradigm discussed earlier, the two slot-fitting tasks of modern MS turn out to be the generation of gaseous analyte ions and their separation based on the mass to charge ratio (m/z). The ion characterization capability of MS can be further enhanced by the tandem mass spectrometer (MS/MS) setup. MS can also be coupled with other sample preparation or separation techniques, such as time resolved mixing,[18] gas/liquid chromatography (GC/LC),[4] and ion mobility spectrometry (IMS)[6], to gain additional dimensions in the analysis scheme.
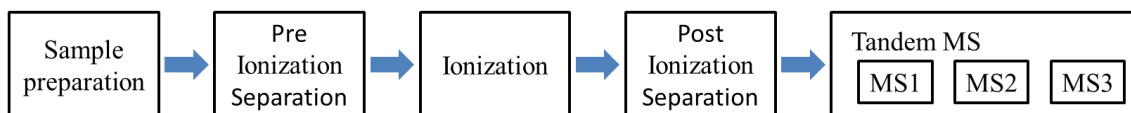
Figure 2-1 General analysis scheme of modern MS system

To generate gas phase ions, the electrospray ionization (ESI) technique[5] is a popular choice. As is shown in Figure 2-2, large charged analyte droplets are formed when applying high voltage between the spray needle and the counter-electrode. These droplets undergo evaporation and explosion when hit their Rayleigh limit[19] until eventually become desired analyte ions. There are two ends for the formation mechanism of gaseous ions from charged droplets: one is the ion evaporation mechanism,[20] during which light analyte ions are ejected from the charged droplet as it evaporates, and the other is the charge residue mechanism,[21] stating that heavy gaseous ions are the remaining cores of the evaporated droplet.

5

Most of the time, the analyte is sprayed as protonated cations or deprotonated anions from its acidic or basic solution. When spraying analyte solution contains some inorganic salt, such as sodium chloride, the ion-analyte complex can also be formed in the gas phase.[22,23]
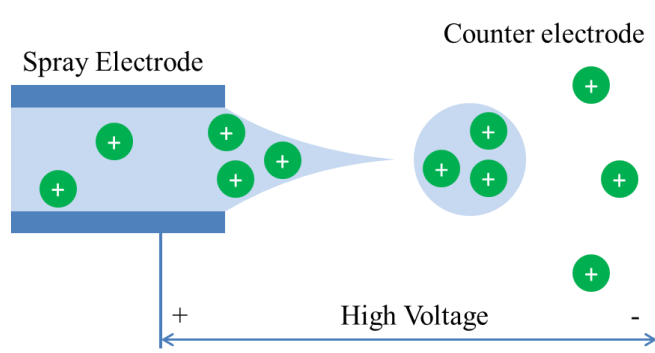
Figure 2-2 Schematic diagram of ESI

In addition to ESI, there are other means to generate gas phase ions. For instance, chemical ionization techniques, like atmospheric pressure chemical ionization (APCI)[24] utilizes gas phase acid-base reactions to generate protonated or deprotonated analyte ions. Another famous technique to ionize large biomolecules is matrix-assisted laser desorption/ionization (MALDI),[25] in which the analyte is crystallized on an inert surface and is then ionized via laser irradiation. Moreover, hard ionization techniques like electron ionization (EI)[26] or inductively coupled plasma (ICP)[27] could also generate gas phase ions, but with the high likeliness of significant analyte decomposition.

Mass analyzers can resolve gas phase ions based on their m/z utilizing rules of electromagnetic interactions. Considering that most small (<1000 m/z) analyte ions in gas phase are singly or doubly charged cations or anions, peak resolution (Δm) of 0.5 m/z unit is the minimum requirement to resolve peaks of the doubly charged isotopologues. This in turn gives a lower boundary of 2000 at m=1000 m/z (calculated using $m/\Delta m$) for the resolving power requirement of routine MS analysis. This requirement can be met by the transmission quadrupole mass filter,[4] which can cover up to 4000 m/z with a constant resolution of 0.3 m/z unit. The orbitrap ion analyser can cover up to 6000 m/z with a resolving power of up to 150,000[4] and is powerful in processing highly charged molecules, such as small peptides and protein, within its covered m/z range. Molecules with large m/z (10^5-10^6) can be handled by either a time-of-flight (TOF) mass analyser ($\Delta m \approx 0.001$, mass range up to 10^6)[4] or by Fourier transform ion cyclotron resonance (FTICR).[28]

In the routines of tandem mass spectrometry (MS/MS) analysis, two or more mass analysis steps are sequentially linked by ion fragmentation steps. Effective execution of such routines is usually ensured in modern MS instruments by integrating several ion analysers along the ion beam pathway within the instrument. These routines give the fragmentation pattern of an analyte ion, which is useful in identifying and characterizing the analyte. There are several types of fragmentation techniques. The most popular one is collision induced dissociation (CID), in which the precursor ions collides with a neutral buffer gas and decomposes under the influence of external electric field.[29] Another collision-based fragmentation method is surface-induced dissociation (SID), in which the analyte ions collide with a solid surface.[30] Alternatively, the analyte ions can also be dissociated via charge transfer. Analyte ions can be effectively fragmented by both the Coulomb energy released when they collide with the oppositely charged species and the potential destabilization upon being oxidized or reduced. Concrete examples of such method include the electron capture dissociation (ECD) and the electron transfer dissociation (ETD).[31,32] Lastly, photo excitations can also be utilized to dissociate the analyte ions. Examples of this category includes the infrared multiple photon dissociation (IRMPD)[33], ultraviolet photodissociation (UVPD),[34,35] and the blackbody infrared radiative dissociation (BIRD)[36] method.

2.1.2 Differential Mobility Spectrometry

In the analysis scheme shown in Figure 2-1, differential mobility spectrometry (DMS) is a post ionization separation technique. DMS can resolve gaseous ion species under atmospheric pressure based on the non-linear dependence of their ion mobility on the external electric field.[37] Ion mobility, K(E), is defined as the ratio between the drift velocity, v_d, of the ion and the field strength when ions drift through a pressurized region under external electric field, E:[38]

$$K(E) = \frac{v_d}{E} \tag{2.1}$$

Under low field strength, K is usually (essentially) a constant, and the drift velocity depends linearly on the field strength.[6] However, K is field dependent, and this can be observed as differences in ion mobility between high and low field conditions. Typically, K is expressed as:[37]

$$K(E) = K(0)(1 + \alpha(E)) \tag{2.2}$$

The $\alpha(E)$ term describes the non-linear change of ion mobility as a function of electric field and it is usually expressed as an empirical polynomial that is fitted to experimental measurements.[6] As is

shown in Figure 2-3, the on-board DMS cell of the Qtrap 5500 system consists of two parallel planar electrodes of dimensions 3 cm × 1 cm with a gap separation of 0.1cm.[39,40] The cell is usually filled with nitrogen carrier gas at atmospheric pressure, but solvent vapors may also be doped in the N_2 carrier to "chemically modify" the environment (*vide infra*). Analyte ions generated from the ESI source are introduced and travel axially through the cell along with the carrier gas towards the detector. Perpendicular to the carrier gas flow, the separation voltage (SV), an AC voltage combined from two sinusoidal waves, is applied between the two electrodes.[37] The SV parameter is commonly described by the peak-to-peak voltage of the waveform. Each period of the waveform is divided into two portions: high field and low field. The high field portion of the waveform is of greater magnitude and acts for shorter duration than the low field portion. The two conditions are of opposite polarity and timed such that integration of the electric field over a complete SV period is zero.[37]
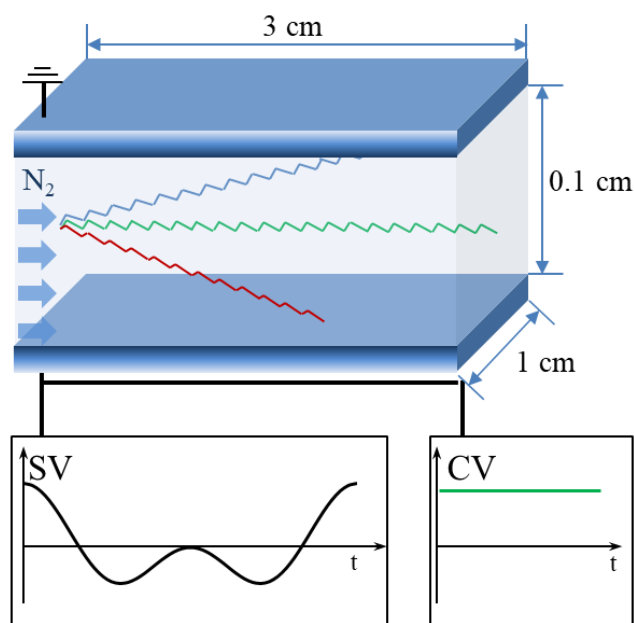
Figure 2-3 Schematic diagram of a DMS cell.

During the SV cycle, analyte ions migrate towards one of the electrodes, due to the variation of the ion mobility as a function of field strength. Accumulation of the displacements for each duty cycle leads to the separation of the analyte ions different in differential ion mobility behaviour. Analyte ions that collide with an electrode while traversing the cell are neutralize and lost to detection. To selectively allow an analyte ion which adopts a unique trajectory to pass through the DMS cell, an additional DC voltage, referred to as the compensation voltage (CV), can be applied.[37]

There are three typical working modes for DMS. The first is the transmittance mode, in which both SV and CV are fixed at certain specific values to selectively transmit a target ion through the DMS. This mode is meant for sample filtration / purification before the MS analysis. The second mode of operation is single scan mode, in which the CV is ramped at a fixed value of SV. The output of this mode is an ionogram (see Figure 2-4a), a record showing the abundance of the eluted ions at different values of CV for the selected value of SV. The third mode is full scan mode wherein ionograms are collected for a series of SV values. For the eluted peaks in each ionogram, their optimal CV values and full width at half maximum (FWHM) are determined and then plotted as a function of SV. Full scan mode yields a dispersion plot (see Figure 2-4b), which provides information that can be interpreted in terms of the type of clustering behavior experienced by each of the observed ion species.
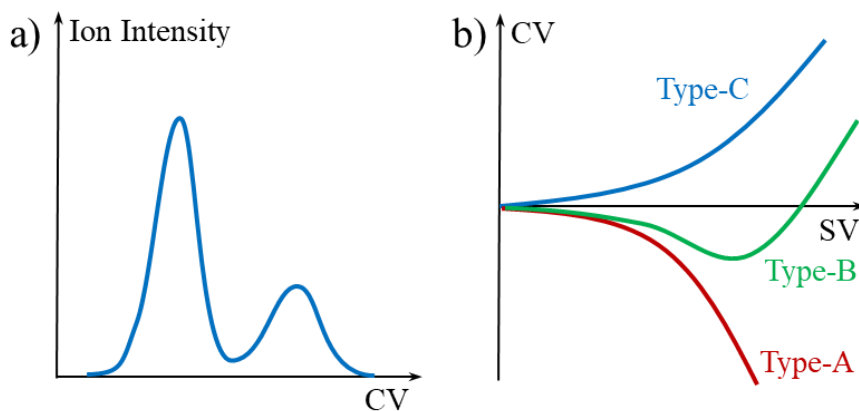
Figure 2-4. a) An example ionogram and b) Three types of DMS behavior

Three ion clustering conditions / behaviours are typically observed in DMS, as is shown in Figure 2-4b.[40] Under pure nitrogen conditions, field dependent ion differential mobility arises mostly from the non-linear relation between the ion's speed and the collision with carrier gas molecules. Ions in a pure N_2 environment usually exhibit hard-sphere collisions, referred to here as (Type C) behaviour. In such a collision environment, the mass of the ion can be treated as constant since the charge-induced interaction between the ion and nitrogen is negligible. The drift velocity of the ion, and thus the ion-nitrogen collision frequency, is determined by and proportional to the electric field strength. At the high field portion of the wave form, the increased collision frequency due to the higher field strength would eventually decrease the ion mobility given negligible ion deformation. Example ions of this category can be found in[41]. In contrast, when an appropriate chemical modifier is introduced to the carrier gas, the analyte ions form ion-solvent clusters under the low field condition and their low field mobility is

thus reduced compared to the high field condition (where solvent evaporates) due to the increased cluster size.[41] If this behaviour persists across the entire range of SV values, it is termed Type A (strong clustering) behaviour. In some cases, however, the high field condition imparts enough energy such that ion-solvent clusters cannot form even under the low field component of the SV waveform. The resulting dispersion curve adopts Type A behaviour at low SV values, reaches a minimum value of CV at some relatively high value of SV, then adopts Type C behaviour at the upper end of the SV range. This intermediate behaviour is known as Type B (weak clustering) behaviour. It should be noted that ion differential mobility behavior is not limited to Types A-C. Certain species, like the tricarbastannatrane cation, exhibit very strong clustering with modifier molecules and produce ion-solvent clusters that can survive the high field portion of the wave form at low SV, giving rise to the so called Type D behavior.[42]

As is discussed above, an ion's DMS behavior is correlated with its clustering ability with modifier. To evaluate this ability at molecular level, one has to consider the population and properties of the ion solvent clusters, including the composition of possible clusters, the possible conformation and the relative energy of each cluster, and the collision cross section of each relevant species. In addition, the temperature dependence of the above properties is also required (*e.g.* [43]). The two-temperature theory is a convenient model to account for the field induced heating in the high field and low field portion of a working cycle.[41] An example first principle model with the above considerations is forwarded by Haack et al., in which the DMS behavior of tetramethylammonium cation is predicted.[44] The theoretical dispersion plot of tetramethylammonium cation agree qualitatively with experimental measurements, but still deviates in the CV value required for ion transmission.

In principle, the DMS resolved ion populations are kinetically trapped species under the experimental separation condition, otherwise, under ergodic assumption, all the ion species should convert to the lowest energy structure, and no separation will occur. Several concrete examples demonstrate the resolution, characterization, and re-equilibration of such kinetically trapped ions.[45–47] This implies that exploring theoretically those kinetically trapped species may be useful in refining the theoretical model.

A further application of the DMS behavior is that the turning point (or SV@CV$_{min}$) of an ion's dispersion plot is correlated with the ion-solvent binding energy.[48] This binding interaction is related not only the electronic and the steric effect of the analyte ion,[48,49] but also the binding affinity of the

10

modifier.[50] This binding affinity information, together with the collision cross section of each ion species, can be used to predict other molecular properties of pharmaceutical relevance.[49,51]

2.1.3 Action Spectroscopy

Action spectroscopy is a technique that probes the spectroscopic features of gas phase ions. In conventional infrared (IR) or ultraviolet-visible (UV-Vis) absorption spectroscopy, samples are usually prepared in solid or liquid phase, and the spectrum is generated by monitoring the depletion of incident IR or UV-Vis light.[4,52] However, this scheme is only effective for a limited number of neutral gas-phase species since most analytes do not yield vapor concentrations high enough to induce significant levels of light absorption. The response of action spectroscopy originates from the photo-induced fragmentation of the analyte ions, which can be captured with high sensitivity using the MS technique. As is shown in Figure 2-5, the procedure for a laser-induced photodissociation analysis usually involves three steps: (1) analyte ion isolation within an ion trap, (2) irradiation with a laser of frequency ω and, (3) photon absorption and subsequent dissociation of the parent ions, followed by detection of the product ions and remaining parent species. Based on the extent of dissociation (*i.e.*, the ratio of remaining parent ions and fragments), the spectroscopic response at frequency ω can be determined. Repeating the above three steps for a range of laser frequencies yields an action spectrum for the ion. In this work, two types of action spectroscopic techniques, IRMPD and UVPD (see Figure 2-6), are utilized.
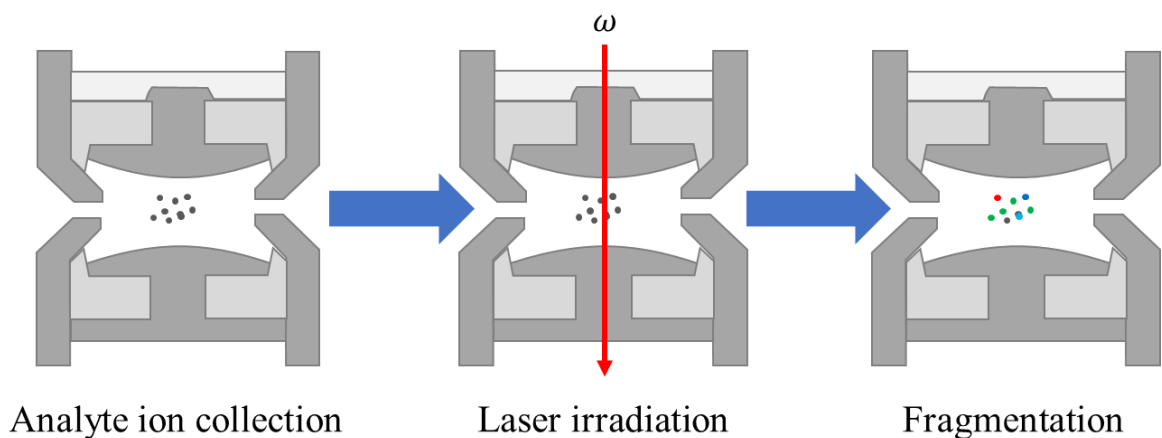
Figure 2-5. The process of photo-induced dissociation
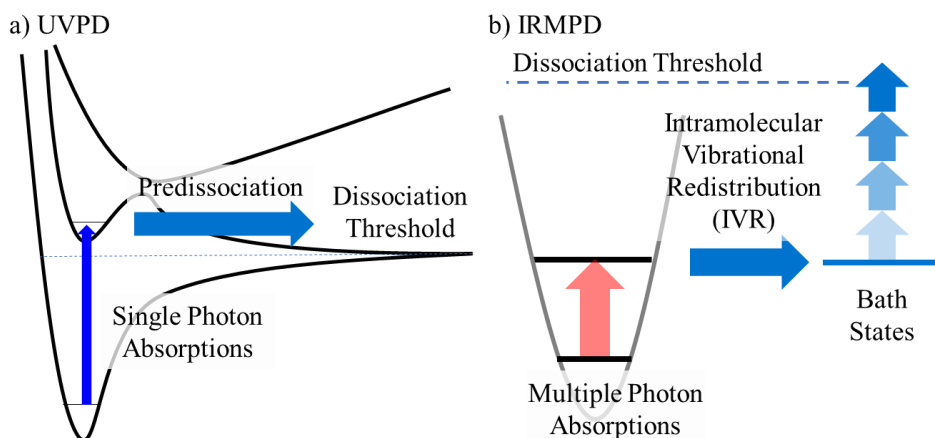
Figure 2-6. The a) single photon absorption and the b) multiple-photon absorption process

The fundamental difference between IRMPD and UVPD is the photon energy of the probe laser. IRMPD employs an IR laser and probes the ions' vibrational spectrum, while UVPD uses a UV-Vis laser and probes the ions' vibronic spectrum. This difference directly influences the efficiency of ion fragmentation. For UVPD, photofragmentation of an analyte can be initiated by absorbing a single photon[34], while for IRMPD fragmentation involves both absorption of multiple photons and the intramolecular vibrational energy redistribution (IVR) process,[53] as is shown in Figure 2-6. Consequently, the power of a bench-top laser is usually adequate to handle UVPD experiments,[54] while for some species IRMPD experiments requires the use of powerful free electron lasers[55] which are only available in a limited number of facilities.

For the work described in this thesis, a modified commercial Qtrap 5500 mass spectrometer (SCIEX) has been coupled with a bench-top Nd:YAG pumped Horizon optical parametric oscillator (OPO) laser system (Continuum Horizon II, 10 Hz). As is shown in Figure 2-7, the Qtrap 5500 system features a triple quadrupole mass analyzer that has been fitted with a Selexion DMS cell. An optical viewport, P, is opened at the rear of the vacuum chamber, and two sets of reflective mirrors, M_1 and M_2, are fitted on opposite sides of Q_3 to create an optical path that intersects perpendicularly with ions trapped within the linear ion trap. This enables UVPD spectroscopy of DMS-separated mass-selected species. For routine UVPD experiments, ions are collected in Q_3 for $1 - 3$ ms depending on their abundancy. Then the trapped ions are exposed to ten retro-reflected laser pulses per second for 500 ms. The resulting fragments are ejected to the detector using a mass selective axial ion ejection protocol[56]. Further details of the instrumentation can be found elsewhere.[57]
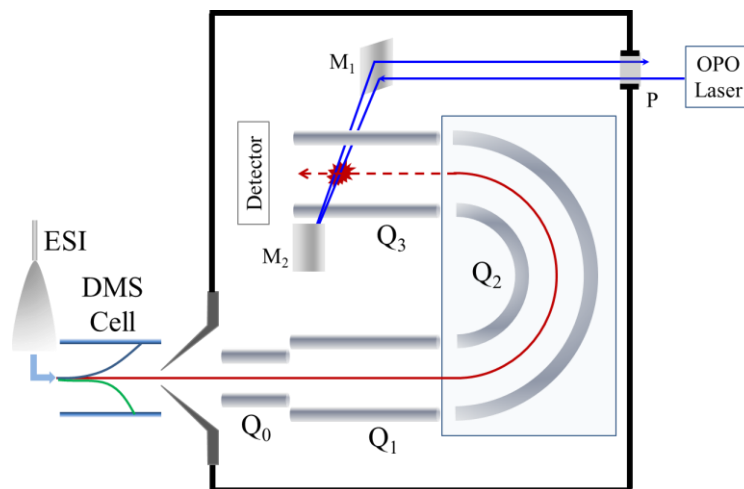
Figure 2-7. Schematic Diagram of Qtrap 5500 Coupled with Horizon OPO Laser

After each round of laser interrogation, ion intensities for the parent ($I_{par,\lambda}$) and all fragment ($I_{frag,\lambda}$) ions are determined via MS. The optical response, $I(\lambda)$, for this measurement is calculated as a fragmentation efficiency:

$$I(\lambda) = -\log\left(\frac{I_{par,\lambda}}{I_{par,\lambda} + \Sigma I_{frag,\lambda}}\right) \tag{2.3}$$

The signal-to-noise ratio of the spectrum can be improved by averaging repeated measurements at the same wavelength. The power of the laser pulses is measured at each wavelength by inserting a power meter (Gentec QE25LP/ Maestro) in the optical pathway between P and the OPO laser. The averaged laser power and its standard deviation are determined and the response (calculated from eq. 2.3) is scaled using the laser power, $P(\lambda)$ in mJ/pulse. Often, the spectrum is then normalized to a range of 0 to 1. Further details regarding data acquisition and processing can be found in Section 2.1.4

$$I_{scaled}(\lambda) = \frac{I(\lambda)\lambda}{P(\lambda)} \tag{2.4}$$

$$I_{norm}(\lambda) = \frac{I_{scaled}(\lambda)}{max\,(I_{scaled})} \tag{2.5}$$

2.1.4 Automated DMS-MS Data Extraction and Processing

A single dispersion plot incorporates data from several (about a dozen) ionograms, and can be extracted, transferred, and processed manually in about four hours. A more efficient means of data extraction and processing is desirable for several reasons. Firstly, the DMS behavior of an ion is usually

13

studied against several factors such as sample preparation conditions (e.g. [58]), DMS conditions (e.g. modifiers), and MS conditions (e.g. DMS-UVPD). Secondly, it is demonstrated that DMS data can be incorporated in databases for supervised machine learning models to predict pharmaceutically relevant quantities such as pKa and cell permeability.[49] Improving the accuracy of such model usually requires massive amount of data. Lastly, one may need to integrate over multiple ranges along certain axes (*e.g.* improve the quality of an ionogram by excluding signal from irrelevant mass channels). To improve the efficiency of data extraction and processing, a customized multidimensional data array (named RawData) was designed. With a supporting gadget from SCIEX, RawData was implemented in a python program named dispersion plot extractor, or DPE, that can directly process the .wiff files produced by the DMS-MS controlling software, Analyst.

The RawData array has three components: (1) an n-dimensional data array, (2) a list of 1D array storing axis values of the corresponding dimension of the data array, and (3) a list of axis names. For storage efficiency and data accuracy, the axis values may not be equally spaced. Based on the data structure, the multidimensional array have two featured methods; one is the "truncation" method that gives another RawData object whose data range along each axis is truncated based on the input parameter; The other is the "integration" method which integrates the ion signal, y, over all dimensions except one that is given, and converts the integrated data into x-y arrays. In the following two examples, the effectiveness of this RawData in the DMS related tasks is demonstrated.

Example 1, which stems from the original purpose of the DPE program, is automation of the data extraction process. This was instigated by the desire to create a database of DMS data for a variety of biologically active small molecules for the purpose of machine learning.[59] This process involves DMS data extraction, ionogram peak fitting, and dispersion plot fitting. As is discussed earlier, in the first step, the RawData is constructed as a 3d Raw data array from the .wiff files containing three axes: CV, MS, and ion intensity. One can then select the mass channel of interest with the 'truncation" method and "integrate" to generate the ionogram data array. With the ionogram data collection at hand, the data smoothing and peaks fitting tasks are mostly delegated to the scipy package.[60] For ionogram fitting, Gaussian smoothing is applied to the raw data points from last step and the peaks are fitted to Gaussian peaks of the form:

$$G(x; A, \mu, w) = \frac{A}{w} \sqrt{\frac{2}{\pi}} e^{-2\left(\frac{(x-\mu)}{w}\right)^2} \tag{2.6}$$

By adapting this form, the peak area is A, the peak center is μ, and the full width at half maximum (FWHM) is $\sqrt{2ln2}w$. The fitting is assessed in reduced χ^2:

$$\chi^2 = \frac{1}{d} \sum_i \left(x_{i,data} - x_{i,fit}\right)^2 \tag{2.7}$$

In eq. 2.7, d is the degree of freedom of the fit, which is the difference between the number of data points and the number of parameters in the Gaussian peak functions (3 per function; i.e., A, μ, w). The initial guesses of the peaks are determined utilizing the first order derivative information of the smoothed data, and the optimization is delegated to the scipy module.[60] When a desired fit is found with satisfactory χ^2, the center and the FWHM of the peak are extracted for the next step to generate dispersion plot.

With the peak data from ionograms extracted, the next step is to construct dispersion plots. Prior to commencing any data fitting, the first task is to determine the number of data series and the peaks included. The number of series can be determined from the maximum number of peaks within one ionogram but categorizing the peaks in each ionogram can be tricky if there are more than three data series present. Currently, a scoring system has been implemented that considers the consistency in the position, the area, and the FWHM of the peaks. However, there are still circumstances in which the scoring system fails; in these cases, the fit must be done manually.

Based on the discussion in section 2.1.2, two empirical functions are employed to fit the dispersion data. One is shared by Type-A and Type-B ions, having the following form:

$$f(x; a, b, c, d) = \frac{1}{d}\left(1 - (a - x)e^{-\frac{a-x}{b+cx}}\right) \tag{2.8}$$

The other is for Type-C:

$$f(x; A, R) = A\left(e^{Rx^2} - 1\right) \tag{2.9}$$

The initial guess for the two functions is hard-coded, and the fitting is again delegated to scipy.[60] Note that for Type-A ions, the inflection point from the fitted function may not be accurate. With all the three steps coded, the time required to process one set of monster mix data (including ionograms,

dispersion plots, and peak information summary of 205 compounds) is approximately 10 min, in contrast with the half-day-time scale needed per compound (*i.e., ca.* 103 days total per experimental run).

Example 2 is the automatic generation of UVPD spectra from experimental data. The structure of the experimental data has three dimensions: the UV laser wavelength, the MS data for the parent ion and its dissociated product under laser irradiation, and the multiple measurements at each wavelength (to improve S/N). Note that a set of power measurement data as is described in section 2.1.3 are also recorded.

To begin generation of a UVPD spectrum, the MS data (*ca.* 10-12 repeated measurements) at each UV wavelength is archived into a single RawData array containing three axes: m/z, index of measurement, and wavelength. With the fragment mass of the dissociated product determined, a truncation is performed using mass ranges of length 1 m/z centered at the fragment masses. Then, for each MS measurement, integration over each desired mass range is conducted to find the total ion count for the species of interest. Following this, the optical response is calculated as described in eq. 2.3. With the N responses calculated from different MS spectrum, the average, $\bar{r}$, and the standard deviation,[4] s_r, of the response is determined using the following equation:

$$s_r = \sqrt{\frac{\sum_{i=1}^{N} |r_i - \bar{r}|^2}{N-1}} \tag{2.10}$$

The above procedure is repeated for data at each wavelength. Eventually, a crude UVPD spectrum is obtained along with standard deviation in fragmentation efficiency. Similar calculations can be done for laser power. Using the error propagation rule for division[4] (eq. 2.11b, re means relative error), one can eventually generate a power normalized UVPD spectrum. For better visualization, UVPD spectra are smoothed with a 7-point Savitzky-Golay filter.[61]

$$e(a \pm b) = \sqrt{e(a)^2 + e(b)^2} \tag{2.11a}$$

$$re(a \times b) = re(a/b) = \sqrt{re(a)^2 + re(b)^2} \tag{2.11b}$$

2.2 Theoretical Calculation Methods

Quantum calculations are conducted to both rationalize at microscopic level the experimental observations and reveal the nature of the problem under investigation. The most useful information drawn from quantum calculations is the geometry and the relative energies of the low energy species, especially of the GM whose energy is the lowest among all isomers. This information can then be used to predict the equilibrium ion population under the experimental conditions. If kinetic effects are to be investigated, the intermediate isomers and the TSs interconnecting the low energy isomers are to be found as well. In this chapter, the theoretical methods used to search and determine these structures are introduced.

2.2.1 The Basin Hopping Simulation

The first step of a theoretical investigation is to sample the molecular structures of interest. For simple systems, this can be done with chemical intuition. For complicated molecular systems, whose sampling is beyond human intuition, the basin-hopping (BH)[62] simulation can be employed. The pre-screened structures are further optimized under higher level of theory. The BH algorithm is a biased Monte-Carlo sampling method toward low energy structures. As is shown in Figure 2-8:
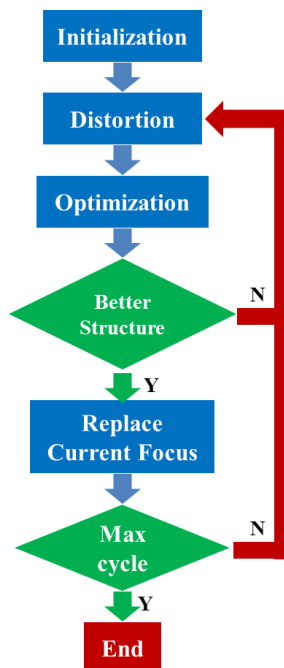
Figure 2-8. General procedure of a BH simulation.

A BH cycle has two major steps: one is finding new structures from the currently focused structure from random distortion and optimization, and the other is evaluating if the new structure is valuable enough to be a new focus. A focus replacement is definite if the new structure has lower energy than the current GM or conditional if its energy lies within a predefined energy threshold. In practice, the energy threshold is defined as a Boltzmann window (in K), and a focus replacement is commenced if the following equation is True:

$$random(0,1) < e^{-\frac{E_{new}-E_{GM}}{k_B T}} \tag{2.12}$$

where the random function generates a number between 0 and 1, E_{new} and E_{GM} is the energy for the newly obtained structure and the GM, and the $k_B T$ is the Boltzmann window. Additionally, three constrains are imposed to the new isomers: 1) newly obtained isomers should be unique to the existing isomers, 2) the interatomic distance should be no less than 0.7 Å, and 3) the overall cluster should be enclosed within a sphere of a given box size. For each BH simulation, around 10^5 to 10^6 BH cycles are conducted depending on the acceptance ratio of individual simulation. Considering the tremendous number of optimizations, the molecular mechanics (MM) force field is employed most of the time. In a MM force field, molecules are usually described in a classical ball-and-spring picture, in which the nuclei are treated as charged point mass and the chemical bonds are described in springs. The total energy of the molecule is the sum of the potential energy in each spring and other interactions such as van der Waals interactions and electrostatic attraction and repulsion. For instance, the AMBER force field has the following form:[63]

$$E_{tot} = \sum_{bond} K_r(r - r_{eq})^2 + \sum_{angle} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedral} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \tag{2.13}$$

Despite an efficient algorithm sampling molecular structures, obtaining the true GM structure is not deterministic. This is due to both the nature of the Monte-Carlo simulation and the systematic deficiencies. For instance, because the MM force field does not allow bond breaking, sampling different protonic isomers is impossible within a single BH simulation. Another example is that if the specified Boltzmann window is too low, the search may be kinetically trapped within a local region of the PES and is thus unable to sample structures belonging to other PES regions. To partially overcome the systematic deficiencies, unsupervised machine learning (ML) methods are inserted into the original BH algorithm.

2.2.2 Structure Similarity and Unsupervised Machine Learning

If an analogy is drawn between the isomer search problem and an adventurer searching for the lowest basin of a region, the basin hopping algorithm does not seem an intelligent strategy: the adventurer simply keeps walking around, record the lowest basin visited so far, and occasionally climb up a hill. However, one obvious necessity for the adventurers, the map, is completely missing. In the context of isomer search, the adventurer's map should give the "location" of an isomer with respect to the previously located isomers. It should also be able to show a reasonable path if the adventurer wishes to travel from one isomer to another. Lastly, it should keep a record of the visited PES regions and direct the adventurer to those that have not. The idea of building this adventurer's map based on the pairwise distances between isomers coincide with the essence of unsupervised ML methods, the dissimilarity function.

Unsupervised machine learning methods rely on the intrinsic dissimilarities between data entities to achieve classification, and an interpolation mechanism between entities to enable model evolution. In a number of global optimization and unsupervised ML algorithms, such as particle swarm optimization, differential evolution, self-organizing map, and growing neural gas,[64-67] a commonly shared feature is the distance, or the dissimilarity between entities. To enable treating molecular structures as entities of unsupervised ML data, a dissimilarity function and an interpolation mechanism is essential. Thus, the focus of this work is quantifying molecular structure differences and extrapolation methods to find intermediate structures. Here two types of dissimilarity functions are proposed: one is the distance matrix-based dissimilarity (or DM distance), and the other is the so-called cosine distance.[68]

To calculate the DM distance, the pair of molecules are converted to their distance matrix form:

$$\boldsymbol{D}_{ij} = |\vec{r}_i - \vec{r}_j| \tag{2.14}$$

With the two distance matrices, the DM distance can be calculated in the following:

$$d(\boldsymbol{D}_1, \boldsymbol{D}_2) = \sum_{i,j>i} |\boldsymbol{D}_{1,ij} - \boldsymbol{D}_{2,ij}| \tag{2.15}$$

Note that the permutation for D_1 and D_2 in eq. 2.15 is aligned such that the minimum d is generated. Similarly, in calculating the CoM distance, the molecule is converted to a mass-weighted vector, $\vec{R}$.

$$\vec{R}_{COM} = \frac{\sum_i m_i \vec{r}_i}{\sum_i m_i} \tag{2.16}$$

$$\vec{R}_i = m_i |\vec{r}_i - \vec{R}_{COM}| \tag{2.17}$$

The cosine distance is then given by:

$$d(\vec{R}_1, \vec{R}_2) = \frac{\cos^{-1}\left(s(\vec{R}_1, \vec{R}_2)\right)}{\pi} \tag{2.18}$$

$$s(\vec{R}_1, \vec{R}_2) = \frac{\vec{R}_1 \cdot \vec{R}_2}{|\vec{R}_1||\vec{R}_2|} \tag{2.19}$$

Both distances can quantify the structural difference between two molecules. In addition, if the response of two spectra within the same frequency region are vectorized, the cosine distance can also give a quantitative difference between two normalized spectra. This quantitative difference grants certain level of automation in spectra comparison, which is particularly helpful when the number of isomers grows large.[68] Given the distance matrix representation of two molecules, G_1 and G_2, an interpolation mechanism via linear addition of G_1 and G_2 is proposed:

$$G_{inter} = \lambda G_1 + (1 - \lambda)G_2, (0 < \lambda < 1) \tag{2.20}$$

Finding the corresponding real space structure of G_{inter} is difficult or even impossible for most large molecular systems. Instead, one can optimize over a real space structure, C, to best reproduce G_{inter}. This is equivalent to minimizing a residual force-field of the form:

$$V(G_C; \chi) = \frac{\chi}{r_{ij}}\left(G_{C,ij} - G_{inter,ij}\right)^2 \tag{2.21}$$

The χ parameter can be an arbitrary positive number to facilitate calculation, and G_C is the DM representation of C. This technique is particularly useful when an exact interpolated structure is not required, such as generating initial guess for transition state optimization.

With the pairwise distance between structures determined, one can eventually construct the multidimensional map for the adventurers. To better visualize this map, two common data processing techniques, namely hierarchical clustering (HC)[69] and multidimensional scaling (MDS),[70] are employed. HC is a type of data analysis technique that groups data entity based on their common feature defined by a similarity function. The HC process can be performed in two strategies: agglomerative clustering, in which singletons of data entity merge to ultimately form a single collection, and divisive clustering,

in which a set containing all data entities is formed initially and is then split into desired number of subsets.

For agglomerative approaches, the decision of linking a data set is controlled by the linkage method, which gives the distance between merged data sets based on the elements they contain. there are simple linkage methods like single and complete, which uses the minimum and the maximum pairwise element distance of two data sets, respectively. There are also several complicated algorithms such as the UPGMA/WPGMA[71] and Ward's method.[69] On the contrary, the divisive approach is implemented in the DIANA algorithm.[72] The result of HC is usually represented in a dendrogram, as is shown in Figure 2-9. Note that in this thesis, the following equation is used to update the Ward's distance (v, s, t is the data sets before grouping, $u = s \cup t$, $|v|$ is the number of elements in v, T is the total number of elements in v, s, t):

$$d(u,v) = \sqrt{\frac{|v| + |s|}{T}d(v,s)^2 \frac{|v| + |t|}{T}d(v,t)^2 - \frac{|v|}{T}d(s,t)^2} \qquad (2.22)$$

In addition to visualization, one could also find the common features among the clustered structures in a much more efficient and systematic manner than manually iterating though each structure. The common structure features from small molecules and clusters can potentially help predict structures of large ones if they have similar components that could constitute the features.
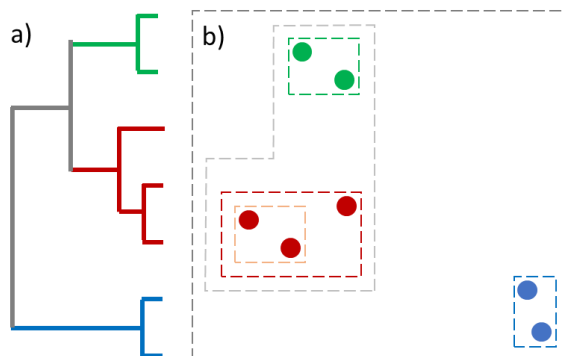
Figure 2-9. a) An example dengrogram and b) the corresponding data points on a 2D plane.

In parallel with HC, MDS is an alternative technique to visualize the multidimensional map on a 2D plane. The essence of MDS is to find a mapping between a set of points $x_i \in \mathbb{R}^2$ and the structures in the multidimensional map such that the distance between x_i and x_j best reproduces D_{ij}, the

dissimilarity between structure i and j. Mathematically, this criteria is met by minimizing the stress function, S:[73]

$$S(x, D) = \sum_{i<j}(|x_i - x_j| - D_{ij})^2 \qquad (2.23)$$

Optimization of this stress function is achieved by iterative majorization method.[74,75] Since the result of the optimization is sensitive to the initial position of x_i, the initialization uses principle coordination analysis (also known as Torgerson-Gower scaling).[76] It should be noted that the axis of the resulting 2D plot may not have implication, as the distances between x_i and x_j are invariant to arbitrary translation and rotation. However, the distribution of the projected points provides visualization of the clustering of the structures.

2.2.3 The Paradigm of Quantum Calculation

The traditional picture of molecules comes from the Born-Oppenheimer approximation,[77] in which motions of nuclei and electrons can be treated separately. The nuclei within molecules are simply represented by positional vectors, while the electrons are represented by (complex) wave functions, $\Psi(r_1, r_2, ..., r_n)$. Since electrons are fermions, the wave function must be anti-symmetric with respect to the exchange of any two variables:

$$\Psi(r_1, r_2, ..., r_a, ..., r_b, ..., r_n) = -\Psi(r_1, r_2, ..., r_b, ..., r_a, ..., r_n,) \qquad (2.24)$$

To evaluate the total electronic energy of a molecule at their equilibrium, the time independent Schrödinger equation is used:

$$\hat{H}\Psi = E\Psi \qquad (2.25)$$

In eq. 2,25, $\hat{H}$ is the Hamiltonian operator for total energy. A typical electronic Hamiltonian shown below includes both the kinetic energy of the electrons and the potential energy arises from the electrostatic interactions.

$$\hat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \qquad (2.26)$$

Currently, there are two branches of ab initio calculation methods: one is the density functional theory (DFT), and the other is Hartree-Fock (HF) and post HF methods. Each branch of method has different

22

mathematical representations in both the wave function and the Hamiltonian and is briefly introduced in the subsequent sections.

2.2.4 Density Functional Theory and Time-dependent DFT

In DFT, the many-body total wave function is projected as one-body electron density, $\rho(r)$:

$$\rho(r) = N \int d^3r_2 \ldots \int d^3r_N \Psi^*(r, r_2, \ldots, r_N)\Psi(r, r_2, \ldots, r_N) \tag{2.27}$$

where N is the total number of electrons. The two Hohenberg-Kohn theorems[78] state that the correspondence between the electron density of a non-degenerate ground state and the external potential is also unique, and that the electron density can be optimized variationally. In the Kohn-Sham formalism, the total density, $\rho(r)$, is reproduced by a collection of non-interacting particles, represented in single particle orbitals, $\varphi_j(r)$:

$$\rho(r) = \sum_{j=1}^{N} |\varphi_j(r)^2| \tag{2.28}$$

The orbitals satisfy the time independent Schrödinger equation:

$$\left(-\frac{\nabla^2}{2} + v_s[\rho](r) \right) \varphi_j(r) = \varepsilon_j \varphi_j(r) \tag{2.29}$$

The potential energy, v_s, consists the external electrostatic potential (usually the electron-nuclei attraction), the Hartree potential, and the exchange-correlation potential, all of which are a functional of the total electron density, $\rho(r)$:

$$v_s[\rho] = v_{ne}[\rho] + v_H[\rho] + v_{xc}[\rho] \tag{2.30}$$

$$v_{ne}[\rho] = \sum_A \frac{Z_A}{|r - r_A|} \tag{2.31}$$

$$v_H[\rho] = \int dr' \frac{\rho}{|r - r'|} \tag{2.32}$$

23

The solution to the above equation is the one that minimizes the total energy of the system, which can then be written as:

$$E[n(r)] = -\frac{1}{2}\sum_i \varphi_j^*(r)\nabla^2\varphi_j(r) + \int v_{ne}(r)\,n(r)dr + \frac{1}{2}\iint \frac{n(r)n(r')}{|r-r'|}drdr' + E_{xc}[n(r)] \quad (2.33)$$

The only unknown term in the total energy is the exchange-correlation energy, E_{xc}, and approximating E_{xc} is central to the development of a DFT method. According to John Perdew's 'Jacob's Ladder',[79] the density functionals can be divided into five hierarchical levels. The first level is the local density approximation (LDA), which treats the electron as uniform electron gas:

$$E_x^{LDA} = -\frac{3}{2}\left(\frac{3}{8\pi}\right)^{\frac{1}{3}}\int n(r)^{\frac{3}{4}}dr \quad (2.34)$$

The next level of improvement over the LDA approximation includes the gradient of the electron density, $\nabla n(r)$, and is referred to as the general gradient approximation (GGA). Furthermore, inclusion of electron density Laplacian ($\nabla^2 n$) and electron kinetic energy density in the functional brings the third level known as the meta-GGA. Functionals in the next hyper (or hybrid) – GGA level features the partial inclusion of the exact Hartree-Fock exchange functional. In the last double-hybrid functionals, wavefunction based expressions are used both in the electron correlation and wave function exchange.

In this work, two density functionals are employed: one is the B3LYP functional,[80,81] and the other is the ωB97XD/ωB97XD3 functionals.[82,83] Both functionals belong to the hybrid-GGA level. Specially, the ωB97XD and the ωB97XD3 are range separated functionals, having a HF exact exchange of 22.2% at short range, and up to 100% at long range. In comparison, the B3LYP functional has a constant 20% HF exact exchange at all ranges.[84]

To find the excitation energy of a molecule, the time dependent DFT (TDDFT) method is also employed. Similar to the Hohenberg-Kohn theorem, the Ronge-Gross theorem proves the unique correspondence between a time dependent electron density, $n(r,t)$, and a time dependent potential, $v(r,t)$.[85–87] Like the ground state Kohn-Sham equation, its time-dependent version is written as[86,87]

$$i\frac{\partial}{\partial t}\varphi_j(r,t) = \left(-\frac{\nabla^2}{2} + v_s[n(r,t),\Psi_0,\Phi_0](r,t)\right)\varphi_j(r,t) \tag{2.35}$$

$$n(r,t) = \sum_{j=1}^{N}|\varphi_j(r,t)^2| \tag{2.36}$$

Note that the v_s potential is a functional of all the historical electron density (from 0 to t), n, the initial interacting wave function, Ψ_0 (in wave function of N electrons), and the initial Kohn-Sham wave function, Φ_0 (constructed from φ_j).

$$v_s(n(r,t),\Psi_0,\Phi_0) = v_{ext}(n(r,t)) + v_H(n(r,t)) + v_{xc}(n(r,0\sim t),\Psi_0,\Phi_0) \tag{2.37}$$

If the initial state is the ground state, the dependence on the initial states are removed, as they themselves are a functional of $n(r,t)$. In addition, adiabatic approximation can remove the dependence of v_{xc} on the past electron density, which means at arbitrary time t, the system is always in its ground state.

$$v_{xc}(n(r,t),\Psi_0,\Phi_0) \rightarrow v_{xc}^{GS}(n(r,0\sim t)) \rightarrow v_{xc}^{GS,ADIA}(n(r,t)) \tag{2.38}$$

Fundamentally, the electronic spectrum of a molecule is a measurement of how strong the resonance between the electron density change and the frequency of the external field. In the context of TDDFT, this is achieved by finding the poles of the density-density response function:

$$\chi(r,r',\omega) = \chi_{KS}(r,r',\omega) + \int d^3r_1 \int d^3r_2\, \chi_{KS}(r,r_1,\omega)\left\{\frac{1}{|r_1-r_2|} + f_{xc}(r_1,r_2,\omega)\right\}\chi(r_2,r',\omega) \tag{2.39}$$

$$\chi_{KS}(r,r',\omega) = 2\lim_{\eta\to0^+}\sum_{ia}\left\{\frac{\varphi_i^*(r)\varphi_a(r)\varphi_a^*(r')\varphi_i(r')}{\omega - \varepsilon_a + \varepsilon_i + i\eta} - \frac{\varphi_a^*(r)\varphi_i(r)\varphi_i^*(r')\varphi_a(r')}{\omega + \varepsilon_a - \varepsilon_i - i\eta}\right\} \tag{2.40}$$

$$f_{xc}(n_{GS}(r))(r,r',t-t') = \left.\frac{\delta v_{xc}(r,t)}{\delta n(r',t')}\right|_{n_{GS}} \tag{2.41}$$

In quantum calculation programs, this process is converted to an eigenvalue problem using the method suggested by Casida and coworkers.[88]

2.2.5 Hartree-Fock and Post Hartree-Fock Methods

Alongside the electron density approaches discussed above, the many-body problem of the electrons can also be expressed in an explicit wavefunction form. In modern quantum calculations, the total molecular wavefunction is usually expressed as a linear combination of primitive Gaussian type orbitals (GTOs) centered at the nuclei.[89] Adopted from atomic orbitals, these functions usually have the following form:[89]

$$\phi(\boldsymbol{r}) = R(r)Y(\theta, \phi) \tag{2.42}$$

In eq. 2.42, $R(r)$ and $Y(\theta, \phi)$ are the radial and the angular portion of the function in spherical coordinate space. To perfectly simulate the true wavefunction, one in principle needs an infinite number of GTO. However, only a limited number of GTOs are employed in concrete calculations due to computation resource limitations. The collection of employed GTOs is termed the basis set. Due to the antisymmetric nature of electrons, the total wavefunction is usually written as Slater determinants:

$$\Psi(r_1, r_2, \dots, r_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(r_1) & \chi_2(r_1) & \cdots & \chi_n(r_1) \\ \chi_1(r_2) & \chi_2(r_2) & \cdots & \chi_n(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(r_n) & \chi_2(r_n) & \cdots & \chi_n(r_n) \end{vmatrix} \equiv |\chi_1 \chi_2 \cdots \chi_n\rangle \tag{2.43}$$

In eq. 2.43, the $\chi_i(r)$ are one-electron orthonormal wavefunctions, or in the word of a chemist, molecular orbitals (MOs) constructed from the GTOs. Note that if one swaps the position of two arbitrary electrons, the sign of the wavefunction flips.

Among the wavefunction based methods, the Hartree-Fock (HF)[89] method lies is the most fundamental. In the HF method, the Hamiltonian is written as the sum of single electron contributions:

$$\hat{H} = \sum_i f_i = \sum_i \left[-\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + v^{HF}(i) \right] \tag{2.44}$$

The $v^{HF}(i)$ is the averaged electron-electron repulsion felt by the ith orbital from the other orbitals and is inaccurate due to the ignorance of dynamic electron correlations. With both the Hamiltonian setup and the wavefunction initialized, the Schrödinger equation is then solved in an iterative manner, known as the self-consistent field (SCF) approach.[89] Despite the resulting total energy being inaccurate,

the HF calculation gives a series of orthonormal MOs and associated energies that is reusable in post HF methods.

As is discussed earlier, the source of inaccuracy of the HF method is due to the ignorance of electron correlation in the $v^{HF}(i)$ term. In a classic system containing multiple charged particles, the term correlation means the force exerted on one particle depends not only on its own location, but also on the position of other particles in the system. In the context of post HF calculations, the quantum analogy to this classic interpretation suggests that the total wavefunction should not contain only one Slater determinant, but a linear combination of them with different electron configurations. As is shown in the rest of this section, the scheme of linear combination is central and unique to each post HF calculation method.

To effectively iterate through the possible determinants, the second quantization[90] approach is usually employed. The idea of second quantization is to construct and interconvert the determinants using creation and annihilation operators:

$$\hat{i}^\dagger |VAC\rangle = |\chi_i\rangle$$
$$\hat{i}|\chi_i\rangle = |VAC\rangle$$
$$\hat{a}^\dagger \hat{i}|\chi_i\rangle = |\chi_a\rangle$$

(2.45)

As is shown in eq. 2.45, the effect of these creation and annihilation operator is to create or eliminate an election from a specified orbital. Conventionally, index label i, j, k, l… are for occupied orbitals and a, b, c, d… are for virtual orbitals. The $|VAC\rangle$ is the vacuum state which satisfies $\langle VAC|VAC\rangle = 1$. Following the Pauli's exclusion principle,[89] there are several cases in which the determinant will be zeroized:

$$\hat{i}|VAC\rangle = \hat{a}|VAC\rangle = \hat{i}^\dagger |\chi_i\rangle = \hat{a}^\dagger |\chi_a\rangle = 0$$

(2.46)

Additionally, these creation and annihilation operators follow anticommutation relations:

$$\hat{i}\hat{j} + \hat{j}\hat{i} = 0$$
$$\hat{i}^\dagger \hat{j}^\dagger + \hat{j}^\dagger \hat{i}^\dagger = 0$$
$$\hat{i}\hat{j}^\dagger + \hat{j}^\dagger \hat{i} = \delta_{ij}$$

(2.47)

27

In most calculations, the HF ground state, $|HF\rangle$ is usually used as the reference state while the remaining determinants are expressed in normal ordered[90] product (in curved brackets) of creation and annihilation operators. For simplicity, the hat over the creation and annihilation operator is omitted in the following discussion.

Of the various post HF methods, one straightforward improvement is configuration interaction (CI)[89]:

$$|\Psi_{CI}\rangle = \hat{C}|\Psi_{HF}\rangle = (1 + \hat{C}_1 + \hat{C}_2 + \cdots + \hat{C}_N)|\Psi_{HF}\rangle$$

$$C_1 = \left(\frac{1}{1!}\right)\sum_{ia} c_i^a\{a^\dagger i\}, C_2 = \left(\frac{1}{2!}\right)\sum_{ijab} c_{ij}^{ab}\{a^\dagger i b^\dagger j\}, \ldots \qquad (2.48)$$

The coefficients, c, is determined by solving a series of equations constructed in the following:

$$\langle\Psi_{HF}|\hat{H}\hat{C} - E_{HF}|\Psi_{HF}\rangle = E_{corr}$$
$$\langle\Psi_i^a|\hat{H}\hat{C} - E_{HF}|\Psi_{HF}\rangle = c_i^a E_{corr}$$
$$\langle\Psi_{ij}^{ab}|\hat{H}\hat{C} - E_{HF}|\Psi_{HF}\rangle = c_{ij}^{ab} E_{corr} \qquad (2.49)$$
$$\ldots$$

Once the coefficients are determined, one can find the correlation energy using the first equation in eq 2.49. The upper limit of the correlation energy can be found if the projection outlined in eq. 2.49 includes all the possible excited state determines, and the resulting method is termed full CI (FCI).[89] However, the FCI approach quickly becomes computationally prohibitive as the number of electrons and the size of the basis set increases. To simplify the calculation, the CI singles and doubles (CISD) approach includes only the singly and doubly excited determinants:

$$|\Psi_{CISD}\rangle = (1 + \hat{C}_1 + \hat{C}_2)|\Psi_{HF}\rangle \qquad (2.50)$$

The truncated CISD method suffers from size extensivity issue,[89] *i.e.*, the total energy of a system containing two infinitely-far-away fragments A and B does not equal to the sum of their individual total energy:

$$E_{AB} \neq E_A + E_B \qquad (2.51)$$

To resolve the size extensivity issue, the exponential ansatz, which is the core to the coupled cluster (CC) method,[90] is introduced:

$$|\Psi_{CC}\rangle = e^{\hat{T}}|\Psi_{HF}\rangle \tag{2.52}$$

The $e^{\hat{T}}$ ansatz can be expanded in Maclaurin series (Taylor series at x=0):

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \cdots \tag{2.53}$$

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots$$

Like CISD, the series is usually truncated at the second order for computation efficiency. The resulting method is the coupled cluster singles and doubles (CCSD):[90]

$$\hat{T} = \hat{T}_1 + \hat{T}_2 = \sum_{i,a} t_i^a \{a^{\dagger}i\} + \frac{1}{4}\sum_{ij,ab} t_{ij}^{ab}\{a^{\dagger}ib^{\dagger}j\} \tag{2.54}$$

The exponential ansatz can also be viewed as a transformation of the HF Hamiltonian. By applying the exponential ansatz to the Schrödinger equation involving the Hartree-Fock Hamiltonian, one gets eq. 2.55a:

$$\hat{H}_{HF}e^{\hat{T}}|\Psi_{HF}\rangle = E_{CCSD}e^{\hat{T}}|\Psi_{HF}\rangle \tag{2.55a}$$

Multiplying $e^{-\hat{T}}$ to both sides of eq. 2.55a:

$$e^{-\hat{T}}\hat{H}_{HF}e^{\hat{T}}|\Psi_{HF}\rangle = E_{CCSD}e^{-\hat{T}}e^{\hat{T}}|\Psi_{HF}\rangle = E_{CCSD}|\Psi_{HF}\rangle \tag{2.55b}$$

Defining $\bar{H} = e^{-\hat{T}}\hat{H}_{HF}e^{\hat{T}}$, one gets:

$$\bar{H}|\Psi_{HF}\rangle = E_{CC}|\Psi_{HF}\rangle \tag{2.55c}$$

The t amplitudes can be solved by projection methods, or more precisely, by solving the following series of equations:

$$\langle \Psi_i^a|\bar{H}|\Psi_{HF}\rangle = 0$$
$$\langle \Psi_{ij}^{ab}|\bar{H}|\Psi_{HF}\rangle = 0 \tag{2.56}$$

The canonical CCSD method outlined above has a time complexity of $O(N^6)$,[90,91] which means this method is not very affordable for medium sized molecular systems (composed of 1-2 dozen of first and second row atoms). As is expected from eq. 2.56, the major contributors to the overall $O(N^6)$ time

complexity comes from the number doubly excited determinants. Recently, however, the importation of pair natural orbitals could significantly reduce the time complexity of the coupled cluster methods. Introduction of this approach will come in the later sections.

To improve the CCSD correlation energy, an obvious move is to include contribution from the triply excited determinants (set $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$ and add $\langle \Psi_{ijk}^{abc} | \bar{H} | \Psi_{HF} \rangle = 0$ to eq. 2.56). However, the expected time complexity for the resulting CCSDT method would be $O(N^8)$[91,92] as a consequence of iterating through the triply excited determinants. To compensate between the accuracy of the correlation energy and the time complexity, Raghavachari et al. proposes an augmentation to account for the missing $E_{ST}^{(5)}$ and $E_{DT}^{(5)}$ component in the correlation energy.[91,92] With the t amplitudes from CCSD calculation, the energy correction, ΔE_T, can be calculated in a non-iterative way:[91]

$$\Delta E_T = \frac{1}{36} \sum_{ijk} \sum_{abc} \left(\Delta_{ijk}^{abc} \right)^{-1} \left[\bar{u}_{ijk}^{abc} + \bar{\bar{u}}_{ijk}^{abc} \right] \bar{\bar{u}}_{ijk}^{abc} \tag{2.57}$$

$$\bar{u}_{ijk}^{abc} = \frac{1}{4} \mathcal{A}_{ijk}^{abc} t_i^a \langle jk || bc \rangle \tag{2.57a}$$

$$\bar{\bar{u}}_{ijk}^{abc} = \frac{1}{4} \mathcal{A}_{ijk}^{abc} \left(\sum_e t_{ij}^{ae} \langle bc || ek \rangle - \sum_m t_{im}^{ab} \langle mc || ji \rangle \right) \tag{2.57b}$$

Note that $\mathcal{A}_{ijk}^{abc}$ is the permutation operator that can iterate through all possible triple excitations with proper signs. The leading term in the time complexity is the evaluation of $\bar{\bar{u}}_{ijk}^{abc}$, which is $O(N^7)$. The resulting CCSD(T) method is well balanced between time complexity and accuracy and is thus considered the "Gold Standard" in quantum chemistry calculations.[93]

Recently, pair natural orbital (PNO) based method, particularly the domain-based local pair natural orbital (DLPNO)[94–98] approach has been integrated with the canonical CC method. The resulting DLPNO-CC method has a significantly reduced time complexity[95,96,99] that scales linearly with molecular size. As a demonstration, the DLPNO-CCSD(T) method is applied to Crambin, a small protein having more than 600 atoms.[96] The idea of PNO is first forwarded by Meyer *et al.* to speed up CI.[94,100] For a linear combination of slater determinates, they can be written in a single compact determinant:

$$\sum_a C^a|\psi_1 \dots \psi_a \dots \psi_N\rangle = |\psi_1 \dots \left(\sum_a C^a \psi_a\right) \dots \psi_N\rangle \tag{2.58}$$

The summation shown in eq. 2.58 is equivalent to defining a new pseudo natural orbital. Obviously, the new orbitalis no longer orthogonal to the original basis, ψ. The sum of doubles in the CI wavefunction then becomes:

$$\sum_{ij}\sum_{ab} C_{ij}^{ab}|\Psi_{ij}^{ab}\rangle = \sum_{ij}\sum_a C_{ij}^a|\Psi_{ij}^{aa_{ij}}\rangle \tag{2.59}$$

An obvious improvement by introducing PNO applied in eq. 2.59 is that the time complexity to iterate over the doubles reduces from $O(N^4)$ to $O(N^3)$. However, one compensation of the PNO approach is that one need additional storage space for the extra PNOs. Another drawback of PNO is the additional recanonicalization procedure needed if one wishes to obtain the canonical coefficients. In the DLPNO scheme, the PNOs are constructed over projected atomic orbitals (PAO).[96] These PNO orbitals are highly localized and thus most of the correlation energy comes from electrons occupying neighbouring PNOs. By employing this domain localization, the DLPNO-CCSD method could achieve linear scaling with respect to the size of the molecule.[96]

The excited states can be treated with the equation of motion coupled cluster (EOM-CC)[101] method. The excited state wave functions can be expressed in an excitation operator, $\hat{R}$, act on the ground state:

$$|\Psi_{excited}\rangle = \hat{R}|\Psi_0\rangle = \hat{R}e^{\hat{T}}|\Psi_{HF}\rangle, \qquad \hat{R} = \hat{R}_0 + \hat{R}_1 + \hat{R}_2 + \cdots \tag{2.60}$$

Since $\hat{R}$ and $\hat{T}$ commutes with each other, the Schrödinger equation then becomes:

$$\hat{H}\hat{R}e^{\hat{T}}|\Psi_{HF}\rangle = E_{excited}\hat{R}e^{\hat{T}}|\Psi_{HF}\rangle \tag{2.61}$$

Left multiplying $e^{-\hat{T}}$ on both sides,

$$Left = e^{-\hat{T}}\hat{H}\hat{R}e^{\hat{T}}|\Psi_{HF}\rangle = e^{-\hat{T}}\hat{H}e^{\hat{T}}\hat{R}|\Psi_{HF}\rangle = \bar{H}\hat{R}|\Psi_{HF}\rangle \tag{2.62a}$$

$$Right = e^{-\hat{T}}E_{excited}\hat{R}e^{\hat{T}}|\Psi_{HF}\rangle = E_{excited}e^{-\hat{T}}e^{\hat{T}}\hat{R}|\Psi_{HF}\rangle = E_{excited}\hat{R}|\Psi_{HF}\rangle \tag{2.62b}$$

$$Left = Right \rightarrow \bar{H}\hat{R}|\Psi_{HF}\rangle = E_{excited}\hat{R}|\Psi_{HF}\rangle \tag{2.62c}$$

At this step, one could find that the excitation energy can be found by solving for the eigenvalues of $\bar{H}$. In practice, only a few excitation energies are found using the Davidson method.[102] Since the ground state CC and the EOM-CC method are both finding (different) eigenvalues and eigenvectors of the same Hamiltonian, one should expect the time complexity for both methods to be $O(N^6)$. However, implementing DLPNO for excited state calculation is more complicated than to the ground state.[101] Given the canonical solution to the ground state, the excited state calculation can be facilitated by the similarity transformed EOM (STEOM) procedure.[101,103,104] The heart of STEOM is a second similarity transformation over $\bar{H}$:

$$\hat{G} = \{e^S\}^{-1} \bar{H} \{e^S\} \tag{2.63}$$

The $\hat{S}$ operator has the following components:

$$S = S^- + S^+ \tag{2.64}$$

$$S^- = \sum_{i',m} S_{i'}^m \{m^\dagger i'\} + \frac{1}{2} \sum_{i,m,j,b} S_{ij}^{mb} \{m^\dagger i b^\dagger j\} \tag{2.64a}$$

$$S^+ = \sum_{e,a'} S_e^{a'} \{a'^\dagger e\} + \frac{1}{2} \sum_{e,a,j,b} S_{ej}^{ab} \{a^\dagger e b^\dagger j\} \tag{2.64b}$$

Note that the $\hat{S}$ operator is defined over an active space whose active occupied and virtual orbitals are labeled m, n, ... and e, f, , respectively The inactive orbitals, on the other hand, have a prime (') at the top-right corner of their labels. This second similarity transformation can be viewed mainly as a pre-diagonalization procedure of the $\langle D|\bar{H}|S \rangle$ block of the transformed Hamiltonian, $\bar{H}$. However, the $\{e^S\}$ transformation could also account for the contribution of triples (via the $1/2S^2$ term) and is thus more accurate than the canonical EOM-CCSD method.[105] The determination of the S coefficients is converted to solving IP-EOM-CC and EA-EOM-CC problems defined over Hilbert space of rank N-1 and N+1 (N is the number of electrons of the reference state), respectively. By selecting a relatively small active space, the time complexity of the excited state calculation (and thus the overall calculation) can be reduced to $O(N^5)$. After the second transformation, the excited eigenstates can be found by diagonalizing a portion of the $\langle S|\hat{G}|S \rangle$ block. The IP-EOM-CC and EA-EOM-CC can be also accelerated with the DLPNO approach,[101] and the resulting STEOM-DLPNO-CCSD is quite affordable for the molecular systems of interest in this work.

2.2.6 Modelling the Potential Energy Surface

With the Born-Oppenheimer approximation, the energy profile of a molecular system can be summarized as a multi-dimensional function of the nuclei conformation. If one plots the energy function with respect to independent geometric coordinates, one could eventually get a potential energy surface (PES), as is illustrated in Figure 2-10.
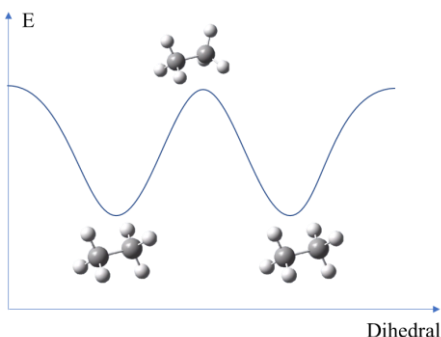
Figure 2-10. The PES of ethane with respect to the H-C-C-H dihedral angle.

To model a PES, one usually needs two components: the minimum structure of the surface, and the transition states (TS) that interconnects the minima. To obtain the conformation of minima and TS, one need to perform geometric optimizations, which utilizes the energy gradient, $\boldsymbol{g}$, and the energy Hessian, $\boldsymbol{H}$, of the molecular conformation, $\boldsymbol{R}$:

$$g_i = \frac{\partial E(\boldsymbol{R})}{\partial R_i} \tag{2.65}$$

$$H_{ij} = \frac{\partial^2 E(\boldsymbol{R})}{\partial R_i \partial R_j} \tag{2.66}$$

To perform a geometric optimization, one usually starts with a reasonable initial guess structure generated from chemical intuition. Then in each subsequent step, one calculates (or approximates) the energy gradient and energy Hessian of the current structure, and updates the structure based on the gradient and Hessian. The loop is terminated if the norm of the gradient falls below the predefined threshold, or the maximum number of iterations is hit. Mathematically, both the minimum and the TS have an energy gradient of zero on the PES. For a minimum structure, all the eigenvalues of $\boldsymbol{H}$ must be positive, while for TS structures, there is one and only one negative eigenvalue. Additionally, for a TS optimization, an intrinsic reaction coordinate (IRC)[106] calculation is conducted to determine if the TS

33

connects the desired minima on the PES. The geometric optimization algorithms generally fall into three categories: the Newtown and quasi-Newtown method (*e.g.* the BFGS update method [107] and the quadratic synchronous transit (QST) method[108]), the conjugate gradient method (*e.g.* the BERNY algorithm[109]) and the direct inversion of the iterative subspace (DIIS) (*e.g.* generalized energy-represented DIIS (GEDIIS)[110]).

With the optimized minima and TSs, one can either construct an energy diagram or a disconnectivity graph.[111] The energy diagram constructed by simply connecting the TSs with the corresponding minima pair that they connect. It can visualize the highest energy barrier along a reaction pathway. The disconnectivity graph is generated by ramping up the energy from the global minimum by a constant interval and group the isomers that can interconvert at each energy threshold. The disconnectivity graph can reveal the partition of the PES by the transition energy barrier and isomeric species that are subject to local kinetic trapping.[112] An illustrative energy diagram and disconnectivity graph is shown in Figure 2-11.

a) Energy Diagram   b) PES   c) Disconnectivity Graph
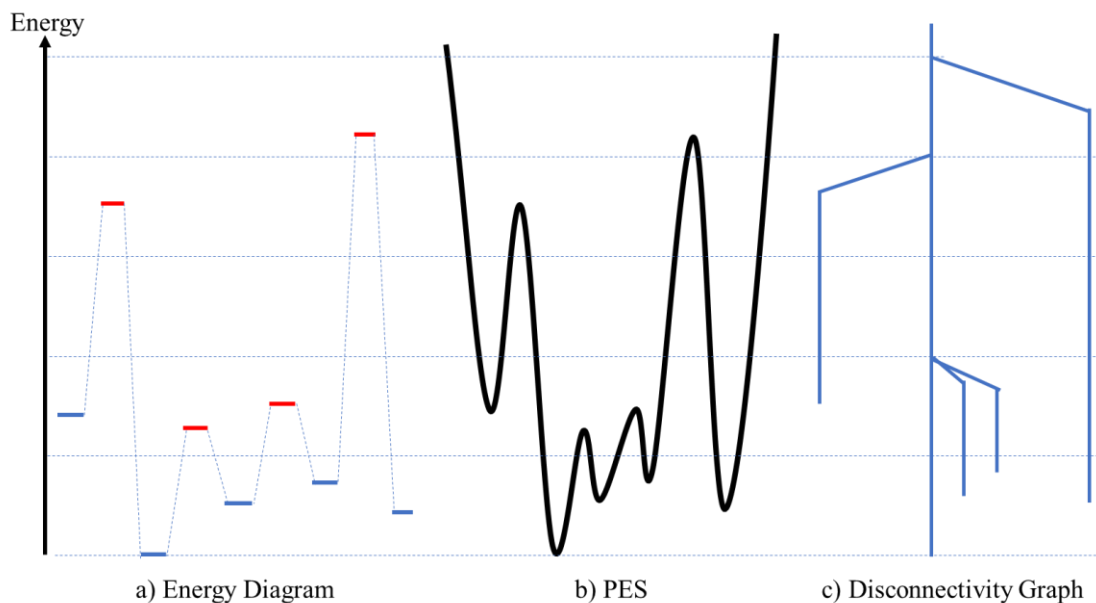
Figure 2-11. The correspondence between a) energy diagram, b) PES, and c) disconnectivity graph. The energy diagram (a) and the disconnectivity graph (c) is drawn from the PES in b.

2.2.7 Simulating vibrational and vibronic spectra

In addition to the ground state properties, the spectroscopic behavior of a molecule is also of great interest. In conjugation with the gas phase photodissociation spectroscopy, the vibrational and vibronic absorption spectra are usually simulated theoretically. However, the photodissociation process is a convolution of both photon absorption and molecular dissociation, and thus absorption spectrum alone may not be adequate to fully model the experimental spectrum. Consequently, a difference in the relative signal intensity between experimental and theoretical spectra is inevitable. Depending on the individual project, the vibrational and the electronic spectra are simulated to compare against either the IRMPD or the UVPD spectra. Thus, in this section, theory of vibrational and vibronic spectra simulation is covered.

The IR absorption spectrum originates from the resonance between the incident IR light and the vibration of the nuclei. To find the vibrational frequencies, one need to construct and diagonalize the second order Hessian of the PES at the minimum structure. To model the PES at the vicinity of the minimum structure, R_e, the harmonic oscillator approximation is usually employed, in which the PES is expressed in terms of a multi-dimensional quadradic potential:

$$V(\boldsymbol{R}) = \frac{1}{2}(\boldsymbol{R} - \boldsymbol{R_e})^T \boldsymbol{K}(\boldsymbol{R} - \boldsymbol{R_e}) \tag{2.67}$$

$$K_{ij} = \frac{\partial^2 E}{\partial R_i \partial R_j}\bigg|_{R=R_e} \tag{2.67a}$$

In the eq. 2.67a, $\boldsymbol{K}$ is the Hessian matrix at molecular conformation $\boldsymbol{R_0}$. This matrix is symmetric and can be calculated either analytically if the calculation method permits or numerically by performing single point energy calculations over the 6N perturbed structures (N is the number of nuclei in the molecule). Substituting the quadratic potential into the time-independent Schrödinger equation, the resulting Hamiltonian for nuclear motion is:

$$\hat{H} = E(\boldsymbol{R_e}) - \frac{\hbar^2}{2}\sum_{i=1}^{3N}\frac{\partial^2}{M_i \partial R_i^2} + \frac{1}{2}\sum_{i,j=1}^{3N}(R_i - R_{e,i})K_{ij}(R_j - R_{e,j}) \tag{2.68}$$

First by introducing $Q_i = \sqrt{M_i}(R_i - R_{e,i})$, eq. 2.68 can be simplified into

$$\hat{H} = E(\boldsymbol{R_e}) - \frac{\hbar^2}{2} \sum_{i=1}^{3N} \frac{\partial^2}{M_i \times \frac{1}{M_i} \partial Q_i^2} + \frac{1}{2} \sum_{i,j=1}^{3N} Q_i \frac{1}{\sqrt{M_i}} K_{ij} \frac{1}{\sqrt{M_j}} Q_j$$

$$= E(\boldsymbol{R_e}) - \frac{\hbar^2}{2} \sum_{i=1}^{3N} \frac{\partial^2}{\partial Q_i^2} + \frac{1}{2} \sum_{i,j=1}^{3N} Q_i G_{ij} Q_j \tag{2.69}$$

$$= E(\boldsymbol{R_e}) - \frac{\hbar^2}{2} \nabla_Q^2 + \frac{1}{2} \boldsymbol{Q}^T \boldsymbol{G} \boldsymbol{Q}$$

$\boldsymbol{\nabla_Q}$ is the gradient operator under Q coordinate, whose result is a column vector. The matrix $\boldsymbol{G}$ is simply the transformed force constant under Q coordinate. In the second step, $\boldsymbol{G}$ is decomposed into its diagonal form:

$$\boldsymbol{G} = \boldsymbol{U\Lambda U}^T \tag{2.70}$$

For a symmetric matrix $\boldsymbol{G}$, the matrix $\boldsymbol{U}$ is composed of orthonormal basis, and has the following property:

$$\boldsymbol{UU}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I} \tag{2.71}$$

Now introducing the $\boldsymbol{q} = \boldsymbol{U}^T\boldsymbol{Q}$, since the $\boldsymbol{U}$ matrix is constant,

$$\boldsymbol{\nabla_q} = \boldsymbol{U}^T \boldsymbol{\nabla_Q}, \qquad \left(\boldsymbol{\nabla_q}\right)^T = \left(\boldsymbol{U}^T\boldsymbol{\nabla_Q}\right)^T = \left(\boldsymbol{\nabla_Q}\right)^T \boldsymbol{U} \tag{2.72}$$

Given that q is column vector, one gets:

$$\nabla_q^2 = \left(\dot{\boldsymbol{\nabla_q}}\right)^T \boldsymbol{\nabla_q} = \left(\boldsymbol{\nabla_Q}\right)^T \boldsymbol{UU}^T \boldsymbol{\nabla_Q} = \left(\boldsymbol{\nabla_Q}\right)^T \boldsymbol{I}\boldsymbol{\nabla_Q} = \left(\dot{\boldsymbol{\nabla_q}}\right)^T \boldsymbol{\nabla_q} = \nabla_Q^2 \tag{2.73}$$

Substituting eq. 2.70, 2.72 and 2.73 to eq. 2.69, the Hamiltonian is written in a normal mode coordinate:

$$\tag{2.74}$$

$$\hbar 2 \qquad 1$$

36

$$= E(\mathbf{R}_e) + \sum_k \hbar\omega_k \left(-\frac{1}{2}\frac{\partial^2}{\partial\tilde{q}_k^2} + \frac{1}{2}\tilde{q}_k^2 \right) \quad \left(\omega_k = \sqrt{\lambda_k}, \tilde{q}_k = \sqrt{\frac{\omega_k}{\hbar}}q_k \right)$$

Using the harmonic approximation, the many-body vibrational problem can be decomposed into multiple one-dimensional harmonic oscillator problems. Using the solution for a 1D-harmonic oscillator,[38] the vibrational energy for the multidimensional harmonic potential is:

$$E_{vib} = \sum_k \hbar\omega_k \left(v_k + \frac{1}{2} \right) \tag{2.75}$$

For non-linear molecules, the number of vibrational degrees of freedom (DoF) is 3N-6. The remaining 6 modes correspond to translational and rotational motions and should be treated separately. For linear molecule, the vibrational DoF is 3N-5. To go beyond the harmonic oscillator approximation, one could introduce anharmonicity to the harmonic potential by adding higher order terms to eq. 2.67.[113]

The UV-Vis absorption spectrum of a molecule is associated with possible electronic excitations, which can be qualitatively pictured as the resonance between the incident light and the change in electron density distribution within a molecule. The transition rate, k, from an initial state, Ψ_i to a final state, Ψ_f under incident light of wavelength ω is given by the Fermi's golden rule:[114,115]

$$k(\omega)_{i\to f} = \frac{4\omega^3 n^2}{3\hbar c^3}\left|\langle\Psi_i|\hat{\mu}|\Psi_f\rangle\right|^2 \delta(E_i - E_f \pm \hbar\omega) \tag{2.76}$$

In eq. 2.76, n is the refractive index of the medium, which is 1 in vacuum, and $\hat{\mu}$ is the electric dipole operator. For a molecular system, both the electrons and the nuclei contribute to the overall dipole moment:

$$\hat{\mu} = \hat{\mu}_e(\mathbf{r}) + \hat{\mu}_N(\mathbf{R}) = -e\left(\sum_i \mathbf{r}_i + \sum_\alpha Z_\alpha \mathbf{R}_\alpha \right) \tag{2.77}$$

Assuming the Born-Oppenheimer approximation holds, the initial and the final state wavefunction can be decoupled into the following form:

$$|\Psi_i\rangle = |\chi_i(\mathbf{R})\psi_i(\mathbf{r};\mathbf{R})\rangle, |\Psi_f\rangle = |\chi_f(\mathbf{R})\psi_f(\mathbf{r};\mathbf{R})\rangle \tag{2.78}$$

Substituting eq. 2.77 and 2.78 into $\langle\Psi_i|\hat{\mu}|\Psi_f\rangle$, one gets:

$$\begin{aligned}
\langle \Psi_i | \hat{\mu} | \Psi_f \rangle &= \langle \chi_i \psi_i | \hat{\mu}_e + \hat{\mu}_N | \chi_f \psi_f \rangle \\
&= \langle \chi_i \psi_i | \hat{\mu}_e | \chi_f \psi_f \rangle + \langle \chi_i \psi_i | \hat{\mu}_N | \chi_f \psi_f \rangle \\
&= \langle \chi_i | \chi_f \rangle \langle \psi_i | \hat{\mu}_e | \psi_f \rangle + \langle \psi_i | \psi_f \rangle \langle \chi_i | \hat{\mu}_N | \chi_f \rangle
\end{aligned} \tag{2.79}$$

Obviously, $\langle \psi_i | \psi_f \rangle = 0$ since they are different electronic states at the same nuclear conformation. Thus, the term $\langle \Psi_i | \hat{\mu} | \Psi_f \rangle$ eventually becomes:

$$\langle \Psi_i | \hat{\mu} | \Psi_f \rangle = \langle \chi_i | \chi_f \rangle \langle \psi_i | \hat{\mu}_e | \psi_f \rangle \tag{2.80}$$

Most of the time, the equilibrium conformation of the initial and the final state is defined over a different molecular frame, while photon absorption occurs in lab frame. In the context of this thesis, the molecular frame refers to a singular coordinate system with its origin at the centre of mass of the molecule in either its ground or excited state, whereas the lab frame refers to a universal coordinate system that encompasses both the ground state and the excited states. Thus, the first step before calculating the inter-state overlap integrals is to properly align the initial and the final state conformation in a common lab frame. Herein, the two conformations are superimposed at their center of mass. Their relative orientation satisfies the Eckart condition:[116,117]

$$\sum_\alpha m_\alpha \boldsymbol{T} \boldsymbol{R_\alpha} \times \overline{\boldsymbol{R}}_\alpha = 0 \tag{2.81}$$

Once the rotation matrix, $\boldsymbol{T}$, is determined, the atomic coordinates and the total wavefunction of the two states is also transformed accordingly. After the transformation, the electronic and the nuclear portion of the integral in eq. 2.80 are treated separately. To facilitate discussion, $\boldsymbol{Q}$ is used to represent molecular conformation in mass weighted coordinates ($\boldsymbol{Q_0}$ is the equilibrium conformation), $\boldsymbol{q}$ is used to represented normal mode coordinates, and $\boldsymbol{U}$ is used for the vibrational normal mode eigenvectors. $\boldsymbol{q}, \boldsymbol{Q}$, and $\boldsymbol{U}$ satisfies $\boldsymbol{q} = \boldsymbol{U}^T \boldsymbol{Q}$. Variables for the final state is represented with a bar over the variables.

The electronic portion of the integral, $\langle \psi_i | \hat{\mu}_e | \psi_f \rangle$, is denoted the transition dipole moment (μ_e), whose magnitude influences the overall transition rate and thus the intensity of an absorption band on a UV-Vis spectrum. Since the dipole operator is related to the position operator ($\hat{\mu}_e = -e\hat{r}$), the parity of the electronic states plays a crucial role in the magnitude of the integral. The dipole operator, $\hat{\mu}_e$, can be written in a Taylor expansion centered at the equilibrium conformation of the initial state, $\boldsymbol{q_0}$:

$$\hat{\mu}_e(\boldsymbol{q}) = \hat{\mu}_e(\boldsymbol{q_0}) + \sum_i \left.\frac{\partial\hat{\mu}_e(\boldsymbol{q})}{\partial q_i}\right|_{\boldsymbol{q_0}} q_i + \cdots \tag{2.82}$$

If only the first term is included, one approximates that the transition dipole is constant with respect to nuclear motions, obtaining the Franck-Condon approximation.[38,118] If the initial and the final electronic state has the same parity, or the transition is symmetry forbidden, inclusion of the linear (second) term is crucial in accounting for the Hertzberg-Teller effect.[119]

In principle, the conformational wavefunction, χ, consists of translational, rotational, and vibrational motions. Since the rotational and translational energy level spacing is quite small in the energy scale of electronic excitation, only the vibrational part is considered when calculating the UV-Vis absorption spectrum. However, it is worth noting that the change in rotational state is one possible placeholder for the angular momentum of an absorbed photon. Using the harmonic oscillator approximation, χ can be expanded into the product of individual 1D harmonic oscillator wavefunctions:

$$|\chi_i(\boldsymbol{q})\rangle = \prod_k |\psi_k(q_k)\rangle, |\chi_f(\overline{\boldsymbol{q}})\rangle = \prod_k |\bar{\psi}_k(\bar{q}_k)\rangle \tag{2.83}$$

In eq. 2.83, the index k iterates through the normal modes, q_k is the vibrational displacement in the kth mode. Most of the time, the q coordinates in the initial and the final state are not aligned in laboratory frame. To calculate overlap integrals, one need to determine the offset between the two q coordinates. This can be achieved by a linear transformation:

$$\boldsymbol{q} = \boldsymbol{J}\overline{\boldsymbol{q}} + \boldsymbol{K} \tag{2.84}$$

The J matrix is the Duschinsky rotation matrix[113] and K is the displacement between the initial and the final state equilibrium. With the Eckart condition satisfied, the matrix $\boldsymbol{J}$ and $\boldsymbol{K}$ can be determined:[116,117]

$$\boldsymbol{J} = \boldsymbol{U}^T\overline{\boldsymbol{U}}, \boldsymbol{K} = \boldsymbol{U'}^T(\overline{\boldsymbol{Q}}_0 - \boldsymbol{Q}_0) \tag{2.85}$$

$\langle\chi_i|\chi_f\rangle$ then becomes:

$$\langle\chi_i(\boldsymbol{q})|\chi_f(\overline{\boldsymbol{q}})\rangle = \langle\chi_i(\boldsymbol{J}\overline{\boldsymbol{q}} + \boldsymbol{K})|\chi_f(\overline{\boldsymbol{q}})\rangle = \prod_k \int \psi_k^*((\boldsymbol{J}\overline{\boldsymbol{q}} + \boldsymbol{K})_k)\bar{\psi}_k(\bar{q}_k)d\bar{q}_k \tag{2.86}$$

In eq. 2.86, the shape of each ψ_k is determined by its own vibrational quantum number, ν_k. Since most spectroscopic experiments are conducted under room temperature, ψ is usually in its ground state ($\nu_k = 0$) while $\bar{\psi}$ may be in some vibrationally excited state ($\bar{\nu}_k \geq 0$). By iterating through all meaningful vibrational levels of the excited state, one can generate a list of integrals following eq. 2.86, whose square are termed Franck-Condon factors. This factor gives the relative intensity of individual vibronic peak within one electronic excitation band series and is thus related to the band shape.

To successfully apply Fermi's golden rule in eq. 2.76 following the Adiabatic Hessian (AH) approach listed above, structure and Hessian matrix for both the initial and the final state minimum is required. However, due to the complexity in both the geometric and the electronic structure, proper minimum on the excited state PES may not be available. Alternatively, since the Franck-Condon window is localized in real space, one can simulate the FC spectrum using an artificial harmonic PES for the excited state. The orientation of the surface in lab frame can be determined using vertical approaches with energy and the energy gradient of the vertical excitation, while the excited state Hessian can also be approximated in a variety of ways.[120] For instance, the simplest vertical approach is the Vertical Gradient (VG), which assumes the initial and the final state Hessian are of the same. More complicated methods include using final state Hessian at the initial state equilibrium geometry, or geometry optimized from the initial state by one step. These two strategies are employed in the Vertical Hessian (VH) and the Adiabatic Hessian After Step (AHAS) approach, respectively. However, since the final state Hessian is not from a minimum structure, it is prone to imaginary frequencies. Finally, it should be noted that methods for simulating vibronic spectrum discussed so far are still within the Born-Oppenheimer approximation. For systems with intense surface crossing (e.g. in [121]), the BO approximation breaks down and off-diagonal non-adiabatic coupling should be considered.

2.2.8 Errors in the *Ab Initio* Methods

As is discussed in the previous sections, two molecular properties are of particular interest in this work: ground state thermochemistry and vibrational/electronic absorption spectroscopy. From thermodynamic study, one eventually obtains the relative population, P_i, of each contributing isomer, which is determined using Boltzmann distribution using the relative free energy of the isomer to the global minimum:

$$P_i = e^{-\frac{\Delta G}{k_B T}} \tag{2.87}$$

The accuracy of the overall free energy depends on that of its individual components, namely translational, rotational, vibrational, and electronic energy. Since the rotational and translational energy contribution depends mostly on the temperature and is thus likely the same among different isomers, the major source of error in the total energy difference is from the electronic and the vibrational (zero-point) energy. However, before any discussion on the potential errors in each component, it is important to note that the error in the relative population is extremely sensitive to the error in the free energy difference. Taking the logarithm of both sides of eq. 2.87, one gets:

$$lnP_i = -\frac{\Delta G}{k_B T} \tag{2.88}$$

Taking the derivative of both sides with respect to ΔG, and rearrange, one gets:

$$\frac{\partial P_i}{P_i} = -\frac{\partial \Delta G}{k_B T}, \qquad re(P_i) = \frac{e(\Delta G)}{k_B T} \tag{2.89}$$

In eq. 2.89, e(X) means the absolute error of X, while re(X) means relative error. Eq. 2.89 shows that a seemingly small absolute error in the relative energy can cause a huge deviation in the resulting isomer population. Quantitatively, at 298.15K, one kJ/mol deviation from the true free energy difference will result in a 40% error in the population, P_i, while for one kcal/mol deviation, the uncertainty in P_i grows to 170%.

To benchmark some theoretical method, the ideal data source is some experimental measurements. However, for ground state ab initio calculations, a direct measurement of the corresponding quantity, i.e. the total energy of a molecular system, is not possible. Thus instead, results from high-level theoretical calculations are usually employed. In addition, the error in the calculation method are systematic error due to ignoring some trivial interactions. Although one can argue that the single point energy should deviate in a similar way among isomeric species, error cancellation in calculating relative energies can hardly be assessed without knowing the exact value. In the following discussion, the error in the total energy are treated as random error across different molecular species, and their accumulation are assumed additive (following eq. 2.11a). For DFT methods, a popular choice is to benchmark against results from CCSD(T) calculations at MP2 minima (*e.g.* in [122]). For post HF methods, the reference data usually come from FCI.[90,123] For excited state calculations, spectroscopic measurement can be used

41

as a good experimental data source[124]. However, depending on the spectroscopic region, the quality of the experimental spectra may vary, and theoretical data may also be used instead.[93,124]

In a recent publication, Mardirossian et al. evaluated the accuracy over 200 density functionals in four scenarios, including non-covalent interaction, isomerization, thermochemistry, and barrier heights.[84] Their reference values are mainly from CCSD(T)/CBS calculations, which they claim to have much better accuracy than the DFT methods. Of the two functionals employed in this work, namely B3LYP-D3 and ωB97X-D/ωB97X-D3, the energy obtained using def2-QZVPPD basis set deviation from reference level is generally smaller than 3 kcal/mol for all the isomerization cases except for one particular case involving C_{20} and C_{24} clusters.[125] The authors also show that the deviation using def2-TZVPPD level is of similar magnitude. For barrier heights, ωB97X-D/ωB97X-D3 shows an energy deviation of 2.3 kcal/mol, which outperforms that of B3LYP-D3 functional, 8.7 kcal/mol. Thus, for the cases in this project, a good estimation of the expected energy deviation from DFT calculations is 3 kcal/mol (12.6 kJ/mol).

As the "gold standard" method that provides benchmark data, CCSD(T) is usually assessed in two perspectives: one is the comparison with higher level calculation , such as CCSDT and full CI, and the other is the comparison with the energy at complete basis set (CBS) limit. In a benchmark study by Eriksen et al., it is found that the CCSD(T) method could in average reproduce 100.2% of the triples contribution of CCSDT with 0.9% standard deviation.[126] Benchmarks against FCI, however, are only performed on very few diatomic or triatomic molecules due its high computational cost,[124,127] and thus a general conclusion regarding the deviation of CCSD(T) energy from the exact value is not viable. Meanwhile, several basis set extrapolation methods are proposed to determine CCSD(T) energy at the CBS limit within an error of 1 kJ/mol.[128–130] If the CCSD(T) electronic energies are trusted as the true energy in an ab initio calculation, the above discussion indicates a maximum error of 1.4 kJ/mol is expected in the electronic energy difference of isomers.

Due to the $O(N^7)$ time complexity, canonical CCSD(T) is only practical for molecules having up to about a dozen of first and second row atoms. Recently, with the introduction of the DLPNO approach, CCSD(T) becomes affordable for larger molecular systems. Thus, the associated error brough by DLPNO needs to be considered. The expected deviation brought by DLPNO is reported in literature[95] as 0.4 - 0.5 kcal/mol (1.7 - 2.1 kJ/mol) from canonical result using the aug-cc-pVTZ basis set and the

TightPNO threshold. In another publication,[131] the authors found that the DLPNO-CCSD(T) energy reproduces that from canonical CCSD(T) within a deviation of 0.65 kcal/mol (2.7 kJ/mol) using cc-pVTZ basis. Since the reference of these deviations is the canonical energy, error from DLPNO should be added to that from canonical CCSD(T). The error in the DLPNO-CCSD(T) energy can then be estimated as 2.2 kJ/mol ($\sqrt{1^2 + 2^2}$ kJ/mol) if the calculations are conducted using a TightPNO setup and CCSD(T) energies are extrapolated to the CBS limit. For relative isomer energy, the error grows to 3.1 kJ/mol ($\sqrt{2} \times 2.2$ kJ/mol)

Besides the electronic energy difference, another significant source of error of the isomer energy comes from vibrational contribution, i.e. zero-point energy (ZPE). Frequency calculation in this thesis is conducted with DFT as the size of the molecule considered in this project is not applicable for coupled cluster theory. With the harmonic approximation, one can expect that the error in the ZPE is less than 0.4 kJ/mol[132,133] for molecules without hydrogen. For molecules containing hydrogen bonds, however, inclusion of anharmonicities is important.[133] To sum up, when error in ZPE is included, the deviation in relative Gibbs energy from DFT calculations is approximately 13 kJ/mol, while that from the DLPNO-CCSD(T) calculations are reduced to 3.1 kJ/mol. The much smaller error from CCSD(T) energies demonstrates its superiority over the DFT methods. However, even the CCSD(T) corrected relative Gibbs energy is far from accurate in determining the relative isomer population (124% error). Thus, the relative energy of isomers serves more as a qualitative tool to hint the existence of certain isomer in the ion population rather than as a quantitative predictor of the relative isomer weight that one must comply when comparing theoretical-experimental results.

Besides relative isomer energy, spectroscopic behavior of the ions is another powerful tool to elucidate the structure information of the isomers. The infrared spectra are predicted by vibrational calculations, while the vibronic spectra are a convolution of electronic excitations and vibrations. Electronic excitations predicted by TDDFT generally have a *ca.* 0.3 eV deviation,[134] while that from EOM-CCSD method is 0.14 eV.[103] Depending on the benchmark dataset, the STEOM-CCSD has comparable[93,103] or less[105] error than conventional EOM-CCSD. Particularly, due to the implicit inclusion of triples, STEOM-CCSD is charge transfer separable and outperforms EOM-CCSD for charge transfer excitations.[93] With inclusion of DLPNO technique, another ca. 0.007 eV error is introduced using the TghtPNO threshold (0.03 eV for NormalPNO).[101] Each excitation can be

simulated either as pure FC band or as FCHT band. However, an accurate reproduction of the experimental band shape is difficult.[135]

2.2.9 Electron Density Based Analysis Methods

In addition to the total energy, ab initio calculations also give the real space electron density distribution of within a molecular system, which can be utilized to quantify the chemistry concepts such as bonding and weak interactions. In this section, the atom-in-molecules (AIM) analysis, the non-covalent interaction (NCI) analysis (also known as the reduced density gradient (RDG) analysis), and the natural difference orbitals (NDO) analysis are introduced.

The foundation of the AIM analysis is Bader's quantum theory of atom in molecules (QTAIM).[136,137] Briefly, this method aims to build a topological network over the real space electron density, $\rho(r;R)$, of a molecule obtained from either a DFT or a post HF calculation. The vertices of the network are chosen those whose electron density gradient, $\nabla\rho$, is zero, and is termed the critical points (CPs). These CPs are classified into four categories based on the number of negative eigenvalues in their electron density hessian, namely the nuclear CP (NCP), bond CP (BCP), ring CP (RCP), and cage CP (CCP). The name of the category is based on the typical location of their appearance. With the CPs found, the edges of the topological map are the viral paths, or the maximal density gradient paths connecting neighbouring NCPs and BCPs. These bonding paths usually suggest some bonding interactions that can range from the strong covalent bonds to the weaker interactions like hydrogen bonding. Conveniently, interactions between different moieties of a molecule, like hydrogen bonding, can be read from the topological analysis. In addition, by collecting different observables, such as the electron density, ρ, and the density Laplacian, $\nabla^2\rho$, at the CPs and compare them against known literature values (e.g.[137–141]) from other well-known system, the nature of the interaction at the CP can be assigned.

Despite a powerful method identifying interactions, AIM analysis may fail to identify potential weak interactions as CPs in some cases.[142] Thus, the NCI method[143] is included as a parallel method to visualize weak interactions. The NCI method utilizes the magnitude of the reduced density gradient (RDG), s, to determine the strength of the weak interaction and the sign of the second largest eigenvalue of the electron density hessian, $sign(\lambda_2)$, to determine whether the interaction is attractive or repulsive. A positive λ_2 implies a repulsive, non-bonding interaction, and vice versa.

44

$$s = sign(\lambda_2) \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}} \tag{2.90}$$

As eq. 2.48 implies, in regions associated with intramolecular interactions, like bonding or van der Waals interactions, small RDG value could be obtained from near-zero electron density gradient and electron density of significant magnitude. When outside the boundary of a system, however, the RDG value goes to infinity due to the near-zero electron density. The weak interaction region of interest can be separated from the strong ones by applying an upper limit to the electron density. The resulting reduced density can be visualized in a 3D isosurfaces alongside with the AIM topological network.

For most molecular systems, an electronic transition may receive comparable contributions from multiple electronic configurations (Slater determinants), and characterization of such excitation with a list of molecular orbital pair is tedious. To efficiently demonstrate the overall electron density change for a particular transition, the natural difference orbital (NDO) analysis is introduced.[144,145] To characterize ground state electronic excitations, the electron density difference, ρ, between the ground and the excited states are of central interest:

$$\rho_\Delta^{0I}(r) = \rho_I(r) - \rho_0(r) \tag{2.91}$$

$$\rho_I(r) = n \int \Psi_I^*(r, r_2, \dots, r_n)\Psi_I(r, r_2, \dots, r_n)dr_2 \dots dr_n \tag{2.91a}$$

$$\rho_0(r) = n \int \Psi_0^*(r, r_2, \dots, r_n)\Psi_0(r, r_2, \dots, r_n)dr_2 \dots dr_n \tag{2.91b}$$

in the above equation, n refers to the total number of electrons, the index '0' refers to the ground state, and the index 'I' refers to the Ith excited state. The resulting density is plotted as isosurfaces and gives information on the spatial distribution of electron detachment and attachment after the excitation. Note that since ionization or electron attachment is not likely during the electronic transitions of interest, the overall density change should equal to 0:

$$\int \rho_\Delta^{0I}(r)\, dr = 0 \tag{2.92}$$

Chapter 3

The Protonated Serine Dimer in the Gas Phase

3.1 Introduction

Among the gaseous amino acid clusters, the serine clusters, especially the serine octamer, have drawn considerable research attention.[2,146] The preference for homo-chiral octamer is believed a crucial property that accounts for the chirality of biomolecules.[147,148] In recent study, Scutelnic et al.[149] found that the global minimum (GM) of the homochiral serine octamer is an elliptical cluster of purely zwitterionic serine units interconnected by a dense hydrogen bonding network. However, information from GM alone may not be adequate to fully understand the role of serine octamer in prebiotic chiral separation. Consequently, a complete topological mapping of the octamer PES will eventually be required. To better model PES of such large cluster, knowledge of the hydrogen bonding schemes from other serine containing species can be potentially useful.

The protonated serine dimer, $[Ser_2 + H]^+$, is a good candidate for studying the role of H-bonding in amino acid structure owing to its potential complexity in hydrogen bonding network (each serine unit has four H-bond donor and four H bond acceptor site). Moreover, $[Ser_2 + H]^+$, is one of the dominant fragments produced during dissociation of protonated serine octamers, suggesting that it may be an important structural unit of $[Ser_8 + H]^{+2}$. IRMPD spectra of the serine dimer in the mid IR ($1000 - 1900 \, cm^{-1}$ and $3200 - 3800 \, cm^{-1}$) have been reported by Seo et al.[8] and Sunahori et al.[9], respectively. These studies proposed isomer structural assignments based on comparison with theoretical calculations. However, the proposed spectral carriers and computed low-energy structures vary between studies owing to variable results arising from different levels of theory.[8–10] Two of the proposed structures are shown in Figure 3-1.
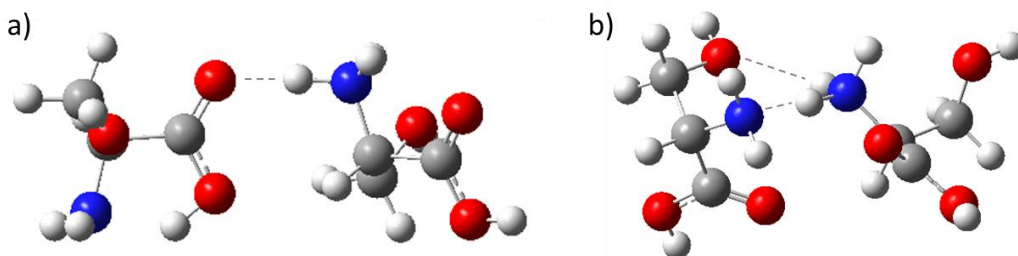

Figure 3-1. GM proposed a) by Kong *et al.*[150] and b) Sunahori *et al.*[9] and Seo *et al.*[8]

To address the discrepancy between the global minima proposed in the various studies, the current investigation surpasses previous work in two aspects: one is the higher-level *ab initio* DFT calculation with CCSD corrected single point energy. The other is the utilization of the PES topology information generated using the literature GMs and their neighbouring dimer isomers. Specifically, many isomers are sampled by BH and optimized to the same level of theory as is used in the literature. The PES topology is built with the dissimilarity function of isomers introduced in section 2.2.2 and is analysed using HC and MDS. By considering the similarities between computed and experimental IR spectra and the localized distribution of the computed isomers, additional insights are garnered for more confident spectral assignment.

3.2 Method

In this study, conventional BH simulations are used to sample $[Ser_2 + H]^+$ isomers, as is described in section 2.2.1. In particular, the AMBER force-field[63] is used as the model potential for rapid energy evaluation. To prepare reasonable building blocks for the protonated dimer, a number of neutral and protonated serine monomers are optimized under B3LYP/6-31I++G(d,p) level of theory. Their atomic partial charges are fitted using the CHelpG partition scheme[151] and are then used in the electrostatic interaction portion of the molecular mechanics force field. To search the potential energy landscape, dihedral angles in both moieties are given random rotations of $-5° \leq \phi \leq +5°$ in each geometric distortion. The neutral moiety is also given random rotations of $-5° \leq \theta \leq +5°$ around its body-fixed x–, y–, and z –axes, and random translations of -0.5 Å $\leq \eta \leq +0.5$ Å in each of the x–, y–, and z –directions. This ensures that the relative orientations of the two moieties are sampled. Following an initial run of 1,000 steps at a thermal energy threshold of E $\approx$ 0.43 eV (T = 5,000 K) to generate candidate structures, several parallel BH runs of 10,000 steps are run at a thermal energy threshold of E $\approx$ 0.09 eV (T = 1,000 K) to search the PES. In total, more than 50,000 cluster geometries were sampled. Following the BH simulation, the first 200 unique lowest energy structures are carried forward to re-optimization at the B3LYP/6-31I++G(d,p) + GD3 level of theory. Note that the GD3 term refers to the Grimme's D3 empirical correction for dispersion interaction.[152] Harmonic frequency calculations are conducted for the low energy structures obtained to ensure that these structures are local minima on the PES (rather than TSs). These calculations also give the vibrational (viz. IR) spectra as well as the thermal corrections to the isomer's energy. To get more accurate relative energy, the electronic energy of the 40 isomers is further corrected using those determined at DLPNO-CCSD/def2-TZVPP level. The

molecular mechanics and the DFT calculations, including the CHelpG charge fitting, are conducted with the Gaussian 09 Rev.D01[153] software package, while the DLPNO-CCSD calculation is conducted with the ORCA 4.2.0 [154–156] software package. The corrected Gibbs energy is calculated in the following equation:

$$G_{CCSD} = G_{DFT} - E_{DFT} + E_{CCSD} \qquad (3.1)$$

Using the DFT optimized geometries, the pairwise DM distance of the 40 dimer isomers are calculated. Linkages for hierarchical clustering were then determined using Ward's minimum variance method, which at each step finds the pair of clusters that leads to the minimum increase in total within-cluster variance after merging.[157] To better visualize the data, the multi-dimensional scaling (MDS) is employed to create a 2D plot of the clustered data.[158] To determine which (if any) of the computed [Ser$_2$ + H]$^+$ isomers are observed experimentally, calculated harmonic vibrational spectra were compared against the experimental IRMPD spectrum using the methodology outlined by Fu *et al.*[68] The experimental IRMPD spectrum are digitized from figures in their respective publications, interpolated in 2 cm^{-1} interval, and then normalized such that the maximum intensity in each region was set to 1. Calculated IR spectra are first scaled using appropriate frequency scaling factors and broadened with a Lorentzian line shape of 15 cm^{-1} FWHM,[68,132] and then are similarly interpolated and normalized. The intensity vectors are constructed by appending the normalized spectrum of the high IR region after that of the mid IR region. The similarity between the experimental and theoretical spectra is evaluated by taking the Euclidian distance (d_{Euc}) between the intensity vectors and assigning a scaled similarity index as per:

$$Scaled\ Similarity = 1 - \frac{d_{Euc} - d_{Euc}^{Min}}{d_{Euc}^{Max} - d_{Euc}^{Min}} \qquad (3.2)$$

where d_{Euc}^{Min} and d_{Euc}^{Max} are the minimum and maximum Euclidean distance amongst the set of vectors respectively. This treatment generates a scaled similarity index that ranges between 0 (worst match) and 1 (best match).

3.3 Results

3.3.1 Binding Motifs of the [Ser$_2$ + H]$^+$ Isomers

Following the procedure described in the method section, a total number of 40 unique isomers are obtained within an energy window of 55 kJ/mol. A detailed list of their relative energies is available in

Appendix A. The resulting dendrogram from the hierarchical clustering, which is plotted in Figure 3-2, clearly shows four distinct groups of geometric structures; these groups are highlighted in blue, red, green, and orange. Based on this hierarchical clustering analysis, it is clearly seen that the BH algorithm identified several local minima associated with four distinct regions of the $[Ser_2 + H]^+$ PES. The lowest energy isomer in each of these four regions (*viz.* isomers 1, 2, 13, and 21) are highlighted and labeled on the MDS plot. This type of analysis provides insight with respect to how thoroughly a region of the PES has been searched. For example, if only one or two data points were identified in the blue region of the MDS plot, one might decide to initialize an additional BH run starting from one of the previously identified geometries. Moreover, this analysis can help guide interpolation efforts to identify TSs or geometries associated with stable intermediates between two previously identified minima. For example, upon inspection of the MDS plot shown in Figure 3-2, one can identify two outliers associated with the red group (in the top left of the red section) and one outlier associated with the green group (bottom left of the green section).

Having identified four low energy geometric groupings associated with the $[Ser_2 + H]^+$ PES, one can then visually inspect the structures to rationalize their association via hierarchical clustering. In doing so, one finds that the clustered species are associated with four distinct binding motifs, which is labeled motifs 1 (orange), 2 (blue), 3 (green), and 4 (red). The 3D structures and 2D chemical structures for the lowest energy isomer in each group is provided in Figure 3-3. Motifs 1 and 3 are associated with bidentate complexation between the ammonium group of the protonated moiety and the neutral moiety. In the case of motif 1, the ammonium group forms intermolecular hydrogen bonds with the amino group and the hydroxyl group of the neutral moiety. In contrast, motif 3 forms intermolecular hydrogen bonds with the hydroxyl group and the carboxylic acid group of the neutral moiety. Motifs 2 and 4 are associated with monodentate complexation between the ammonium group of the protonated moiety and the neutral moiety. These two binding motifs differ in terms of the relative orientations of the two serine moieties and with respect to the presence of a O–H•••N intramolecular hydrogen bond (IMHB) in the neutral moiety (motif 2) versus a O–H•••O IMHB in the neutral moiety (motif 4) as seen on Figure 3-3 (Right).
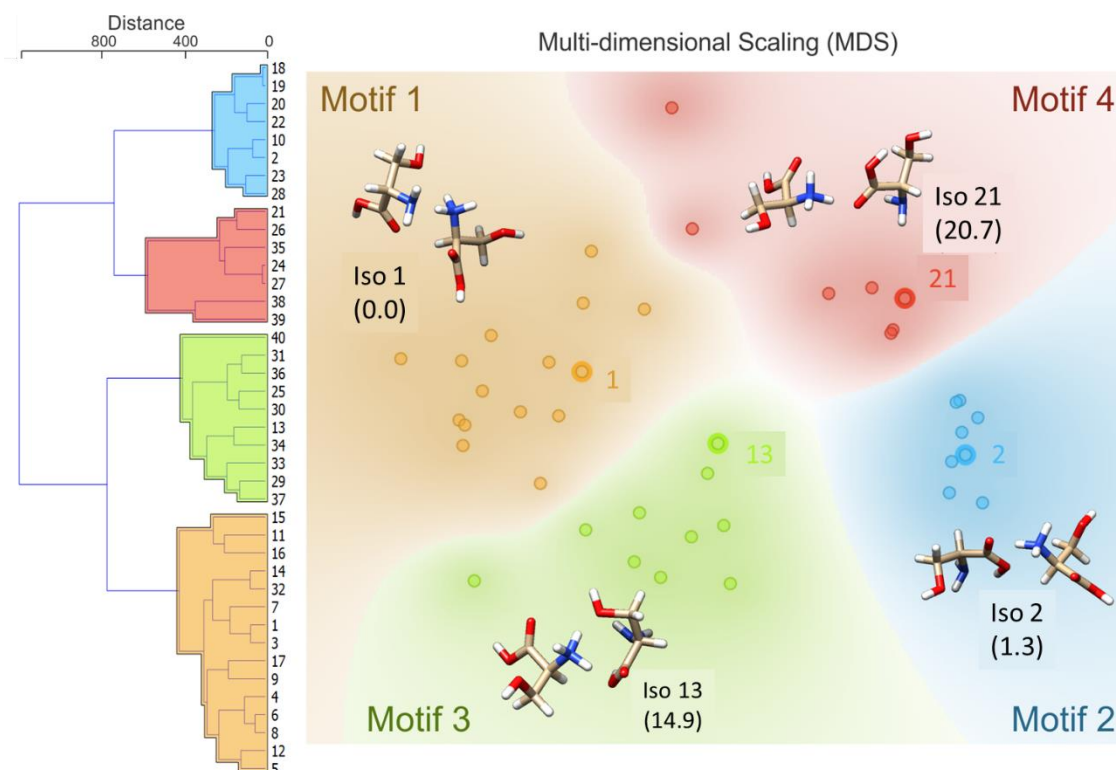
Figure 3-2. (Left) The distance dendrogram for the protonated serine dimer. Isomer numbers are indicated for each branch of the dendrogram. (Right) A MDS 2D projection of the hierarchical clustered data. Isomers are numbered in order of increasing relative energy. Standard Gibbs' energies from eq. 3.1 (in parentheses) are reported in kJ mol^{-1}.
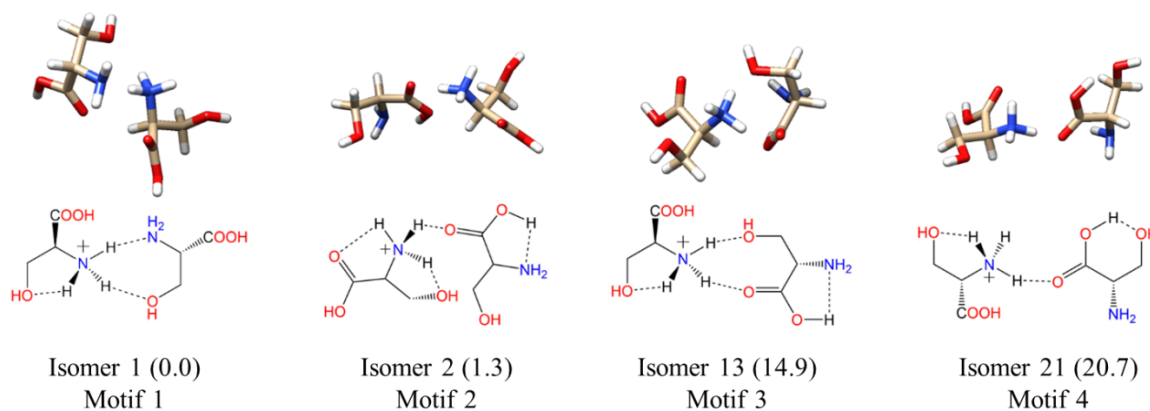
Figure 3-3. The lowest energy isomers for each low energy binding motif of the protonated serine dimer. Motifs 1 and 3 show bidentate coordination between the two moieties, whereas motifs 2 and 4 exhibit monodentate coordination between the two moieties. Standard Gibbs' energies from eq. 3.1 (in parentheses) are reported in kJ mol^{-1}.

3.3.2 Spectral Assignment of the Experimental IRMPD

The scaled similarities for the computed $[Ser_2 + H]^+$ isomer spectra are plotted in Figure 3-4. Inspection of Figure 3-4 indicates that Isomer 2 yields a significantly better match to the experimental spectrum than do other isomers. Moreover, it is found that four of the five best matches are provided by isomers associated with binding motif 2. This suggests that, despite the fact that motif 1 is associated with the lowest energy region of the $[Ser_2 + H]^+$ PES at T = 298 K and P = 1 atm according to free energy calculated from pure DFT and from CCSD correction, the region of the PES associated with motif 2 is predominantly populated in ion trap experiments.
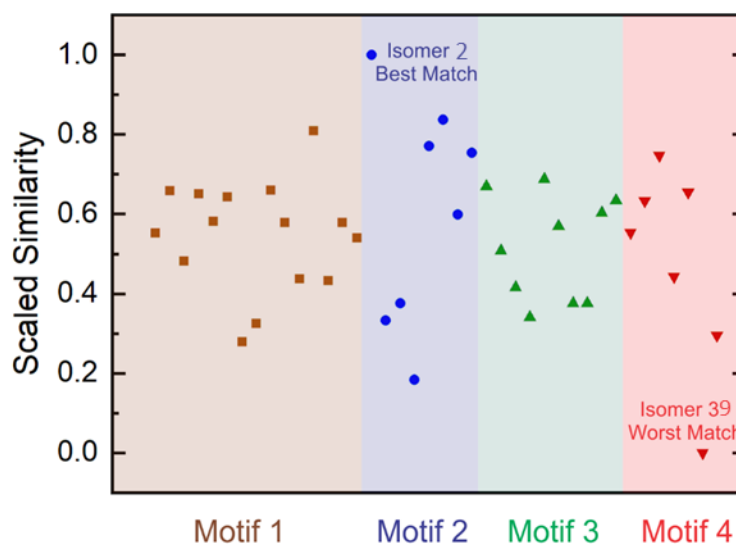
Figure 3-4. Scaled Euclidean similarities of computed harmonic vibrational spectra to experimental IRMPD spectra for the protonated serine dimer. Isomer 2 gives the best match and Isomer 39 gives the worst match amongst the 40-isomer set. Isomers are ordered in increasing energy from left to right in each motif.

Figure 3-5 plots the experimental IRMPD spectrum for $[Ser_2 + H]^+$ and the computed spectra for isomers 1, 2 (best match), 13, and 21 – the lowest energy isomers associated with each of the four binding motifs. The diagnostic peaks, which are highlighted in blue in Figure 3-5, are associated with the HNH angle bending motions (ca. 1450 cm^{-1}) and N–H bond stretching motions (ca. 3250 cm^{-1}) of the ammonium and amino groups. Although isomer 1 is the global minimum structure based on standard Gibbs' energies, the spectrum of isomer 2 (+1.3 kJ mol^{-1}) is more consistent with the experimental spectrum in both regions. This was also noted by Sunahori *et al.*, who identified isomer

2 in their study.[9] Kong *et al.* also identified isomer 6 in their work,[10] but apparently did not consider it in their spectral assignment.
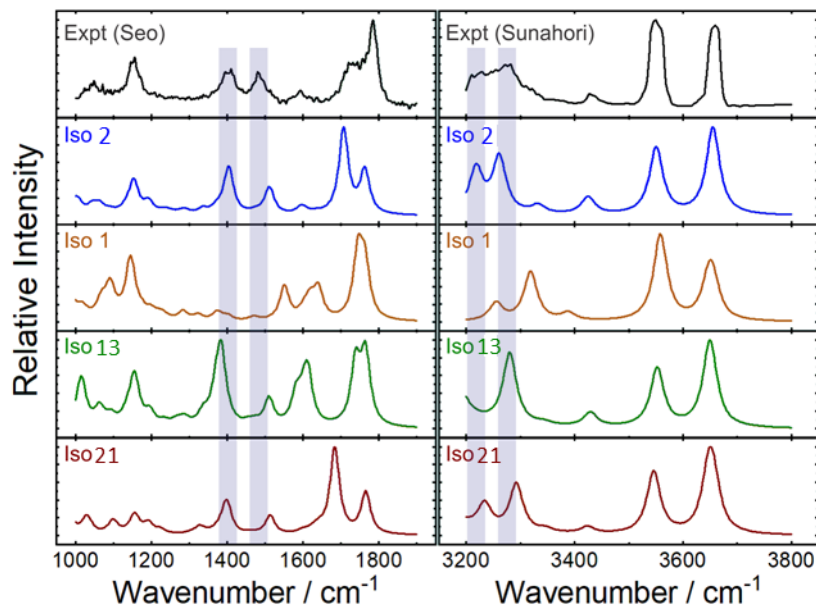
Figure 3-5. Experimental IRMPD spectra and computed harmonic vibrational spectra for the protonated serine dimer. The experimental spectra were adapted from Seo *et al.*[8] and Sunahori *et al.*[9] The computed IR spectra are associated with the lowest energy isomer for each of the four binding motifs. Scaling factors of 0.9679 and 0.95 were employed for the 1000 – 1900 cm^{-1} and 3200 - 3800 cm^{-1} region, respectively.[68,132]

To best determine the relative energy of isomer 1 and isomer 2, single point energy calculations at DLPNO-CCSD(T)/def2-QZVPP level of theory is performed for each of the isomers. Replacing the CCSD energy correction in eq. 3.1 with that from CCSD(T), it is eventually found that Isomer 2 becomes the global minimum, while Isomer 1 is only 0.1 kJ/mol above Isomer 2. Since the energy difference in Isomer 1 and 2 is 1/10 of the so called sub-chemical accuracy, Boltzmann weighted spectra may not be reliable to judge whether Isomer 2 or Isomer 1 is subject to kinetic trapping. To formally determine the presence of kinetic trapping, one need to fully map the PES by finding isomerization pathways that interconnects the 40 minima. If the CCSD(T) energy is trusted, however, the structure similarity argument is proved useful at least in this particular case and can be potentially useful in other similar scenarios. Additionally, the protonated serine dimer can be a curious test case for upper-hierarchy methods like CCSDT.

It is necessary to highlight three caveats for the above example of identifying the spectral carrier of [Ser$_2$ + H]$^+$. First, to assemble the experimental spectrum used in the assignment, the results of two

separate studies are collated. It is not necessarily true that the same ensemble populations were produced under the experimental conditions employed in both studies. However, given that isomer 2 provides the best match to both regions of the experimental spectrum, it seems to be that instrument conditions were similar in these two cases. A second consideration is the fact that peak intensities in IRMPD spectra are not necessarily well-modelled by computed absorption spectra owing to the fact that IRMPD intensities are dependent on absorption cross sections and the coupling efficiency for accessing dissociative channels.[159] The methodology outline above assumes that the computed linear absorption intensities are representative of IRMPD intensities or, barring that, that the IRMPD intensities for a given band vary similarly from the computed intensity for all isomeric species. Finally, the above treatment also assumes that the computed harmonic frequencies suitably model the experimental spectrum. The validity of this assumption depends on the accuracy of the model chemistry and on the anharmonicity of the system being studied. While the $[Ser_2 + H]^+$ is apparently well-modelled by the joint DFT and CCSD approach employed here, one should in general be aware of the anharmonic nature of hydrogen bonds and shared protons.[160–163]

3.4 Chapter 3 Summary

In this chapter, a total of 40 isomers are located in the theoretical investigation of the $[Ser_2 + H]^+$ cluster within an energy threshold of 55 kJ/mol. Hierarchical clustering analysis over the geometric similarity of the 40 isomers reveals four regions on the PES, whose representative bonding motif comes from the lowest energy isomer of each cluster. In the subsequent comparison of calculated spectra against the experimental IRMPD spectra, it is found that isomer 2 (+1.3 kJ/mol) most likely the spectral carrier, regardless of being a higher energy isomer than the global minimum. Single point energy calculation at CCSD(T) level reveals that Isomer 2 is the actual GM, with Isomer 1 is only +0.1 kJ/mol above the GM. However, this energy difference and thus the resulting Boltzmann population may not be reliable considering the potential error in the CCSD(T) energy being 1 kJ/mol even at the complete basis set limit. Considering the best relative energy obtained so far and the spectrum resemblance, Isomer 2 is assigned the primary carrier of the experimental IRMPD spectrum. To further investigate this effect, a thorough investigation of the reaction pathways interconnecting the four regions of the PES is required.

Chapter 4

Regioselective Alkylation with Alkyl-tricarbastannatrane Complex

4.1 Introduction

Tricarbastannatrane, $[N(CH_2CH_2CH_2)_3Sn]^+$ (referred to as triptych herein), is a cationic organotin complex that exhibits a highly constrained cage structure, as is shown in Figure 4-1. The first synthesis of this cage structure is achieved via the reaction between a Grignard reagent with SnCl₄,[164] while more efficient synthesis can be achieved via thermal redistribution and Schwartz's reagent.[12] The product of the synthesis is chloro-triptych, which can then be used to prepare other triptych salts[165] and alkyl-triptych complexes.[165,166]
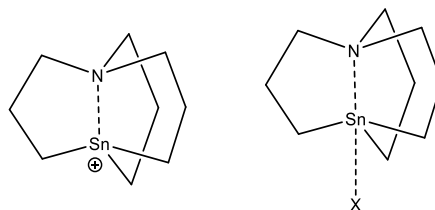
Figure 4-1. Tricarbastannatrane cation (left) and its complex (right)

The sterically hindered cage structure of triptych gives a particularly short N-Sn distance, and thus facilitates intramolecular electron pair donation from N to Sn to form a transannular N-Sn dative bond. As a result of this N-Sn bonding and the cage structure, the triptych cation has a strong tendency to form pentavalent organotin complexes with Lewis bases; the resulting complexes usually adopt triangular bipyramidal structures.[167,168]

The transannular bonding interaction has been investigated in several condensed phase studies, with characterization methods such as X-ray crystallography,[167,168] nuclear magnetic resonance (NMR),[165,169] vibrational spectroscopy (IR, Raman).[170–172] Crystallographic data reveals longer Sn-X and shorter N-Sn bonds than those in linear organotin complexes, while solid phase IR and Raman spectroscopy estimates that the Sn-X bond energy is in the order of 10^2 kJ/mol.[172] The ^{119}Sn NMR study of a series of triptych salts[165] gives a semi-quantitative standard to estimate whether the Sn center is in a free or bound chemical environment.

The bonding interactions have also been characterized in gas phase in studies utilizing IRMPD[173] and DMS.[42] In the former study, experimental IRMPD spectra and ^{119}Sn NMR shifts of a few triptych-

solvent complexes are reproduced from the corresponding theoretical calculation using DFT. The theoretical interaction energy (including both bond dissociation and accompanied geometry distortion) for the N-Sn and Sn-X bonding has a range of 250 - 420 kJ/mol and 335 - 550 kJ/mol, respectively, while the overall cluster dissociation energy is estimated 63 - 126 kJ/mol, which is in good agreement with literature.[172] In the latter study, the ion-solvent clustering behaviour of the bare triptych cation with acetone and acetonitrile was studied. Interestingly, the molecular cation was found to exhibit an atypical Type D dispersion behavior. Cluster dissociation free energies of ~50 kJ/mol were found at gas temperatures of $150°C$.

Current research interest in triptych complexes is associated with two main directions. First is the potential of triptych for forming hypervalent Sn complexes having coordination numbers greater than five at the Sn center. The second direction of interest is in the application of triptych in organic synthesis. Being a group IV element, the Sn center in linear tin and organotin compounds usually adopt quadrivalent coordination with a tetrahedral conformation. However, depending on the ligand of the complex, organotin compounds adopting penta-,[174,175] hexa-,[175,176] and even heptavalent[176] coordination are found. Particularly for the triptych complexes, hexavalent coordination is found in a fluoro-triptych water complex.[169] Of the compounds studied, however, the determination of coordination number relies heavily on the interatomic distances from crystallographic data, while little is done from the theoretical side to characterize the bonding interactions. To fill this knowledge gap, this work employs atom-in-molecule (AIM) and the non-bonding interaction (NCI) analyses are conducted to characterize the bonding interactions in a different point of view.

In organic synthesis, Alkylation with metal-alkyl complexes (*e.g.*, Grignard reagent,[11] alkyl copper and lithium[177]) is a foundational approach to carbon-carbon bond formation in organic synthesis. In recent studies, the alkyl-triptych complexes are found superior in several cases to the metal-alkyl complexes as an alkylation reagent in terms of both stability and reactivity.[178–181] Concrete examples include the palladium catalysed Stille cross coupling reaction,[166,182–184] conjugative addition with Meldrum acid derivatives,[165] and 1,4-hydrostannylation of α,β-unsaturated carbonyls.[11] Particularly, regioselective alkylation (shown in Scheme 4.1) carried out in 1,2-dichloroethane (DCE) has been demonstrated[185] between a barbituric acid derivative, (5-benzylidene-1,3-dimethylpyrimidine-2,4,6(1H,3H,5H)-trione), **1**, and the 3-methylbut-2-en-1-yl tricarbastannatrane complex, **2**:
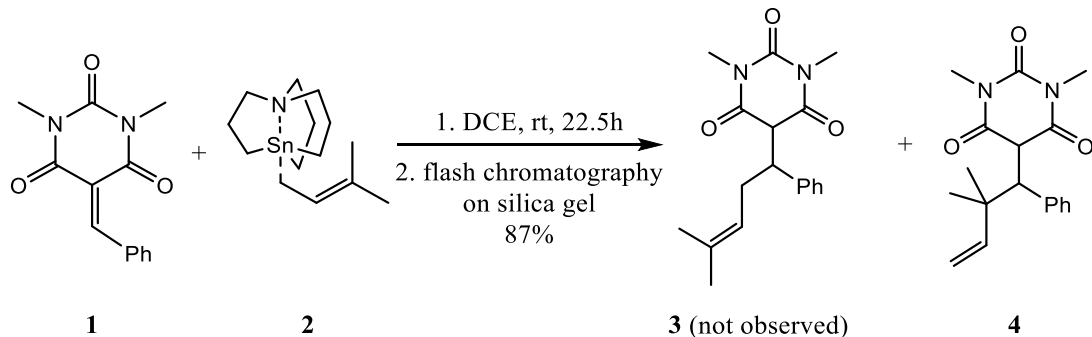
Scheme 4-1. The regioselective alkylation reaction.

While products **3** and **4** are both theoretically possible via eight-membered and six-membered cyclic transition states (as is shown in Figure 4-2), respectively, only **4** is observed experimentally. Putting aside the main goal of rationalizing this observation, there are two reasons for this particular reaction being selected to test the structure interpolation method: one is that the species involved in the reaction are of significant geometric complexity and intuitively guessing the TS structure is not feasible (even for an experienced personnel); The other is that the ligand transfer process is a Lewis acid-base reaction, and thus closed-shell *ab initio* calculations are adequate. Following the experimental work of Baierl et al.,[11] a theoretical investigation is conducted to determine the reaction mechanism for the alkylation utilizing the structural interpolation method.
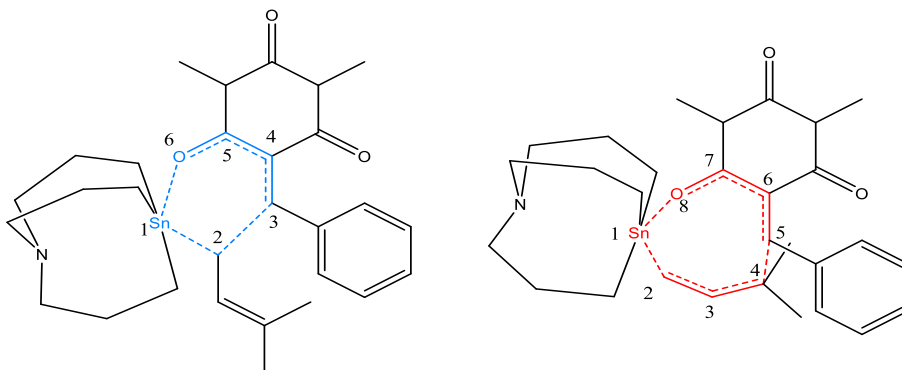
Figure 4-2. The proposed TS to form product 3 (left) and 4 (right)

4.2 Method

The experimental work of Baierl *et al.* is described elsewhere[11] and summarized in scheme 4-1. The focus for this work is the first reaction step in scheme 4-1. To simplify the discussion for the rest of this chapter, compound 1 of Scheme 4.1 is labelled moiety B while the alkyl group and the triptych moiety in compound **2** is labelled moiety A and moiety T, respectively. The complex of triptych with product **3** and **4** are labelled as T3 and T4, as is used in the scheme.

To perform a structural interpolation described in section 2.2.2, a pair of molecular structures for the reagent and the product are required. As is shown in Figure 4-3, to determine these two structures, sampling the reagent and product clusters is required. To do this, 10,000 step BH simulation with molecular mechanics is conducted each over a B+AT cluster, T3, and T4. For the BH of the B+AT cluster, the relative orientation of the B and the AT moiety and the rotation of free dihedrals are sampled. Note that free dihedrals are those whose rotation axes are along the bonds that connects to at least one sp^3 carbon and are not constituting edge of a ring. For the T3 and the T4 complex, clusters in which the Sn atom is attached to the O^6 and O^8 atom are both included. All the free dihedrals as well as the dihedral along Sn-O bond is considered during conformation sampling.
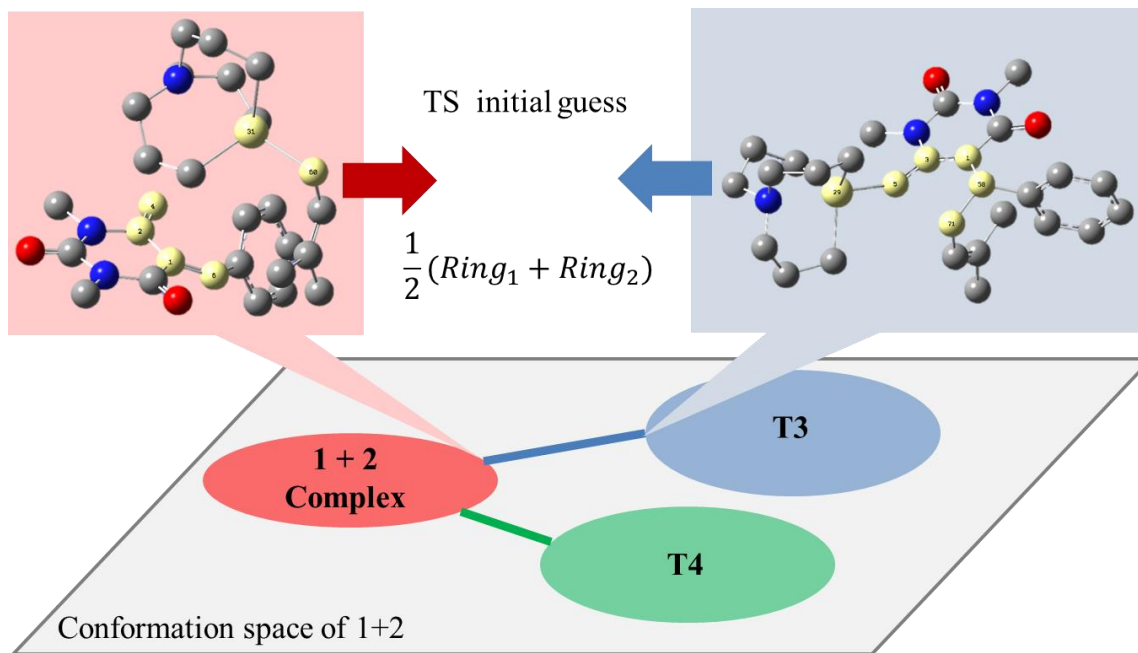
Figure 4-3. Schematic diagram of locating candidate structure pairs for extrapolation.

The resulting isomers of the above three BH simulations are further optimized at PM6 level, and the resulting unique structures spawns the red, blue, and green regions in Figure 4-3, respectively. Unlike conventional BH simulations that seeks the lowest energy structure, reagent-product pairs having highest geometric similarity in their reactive ring conformation are selected. This is done by constructing for each reagent and product conformer a DM representation over their reactive ring, as is highlighted in Figure 4-3, and then evaluate the pairwise geometric similarity. After finding the candidate pairs for Complex-T3 (the blue line) and Complex-T4 (the green line), the structure interpolation is then performed to find an intermediate structure as initial guess. Due to the size of the system, the structure interpolation is not performed directly over the whole structure, but only between the rings. The orientation of the rest moiety to the center of the ring is reconstructed in a sensible manner. The interpolated structure is set as the initial guess of the TS between the corresponding pair of species and is subject to further optimizations.

With the initial guess obtained, unimolecular TS optimizations and subsequent frequency calculations are performed under the UB3LYP/6-311G(d,p)+GD3 level of theory with the application of the polarizable continuum model (PCM)[186] for DCE and the def2-TZVPPD/ECP-28 basis set[187,188] for the Sn atoms. The optimized TSs are verified by the unique imaginary frequency and the IRC calculations. With the first two TSs obtained through a complicated process, other possible TSs can be iteratively generated following chemical intuition suggested by the two initial TSs and then optimized under the same level of theory. The lowest energy BH unique for the two product clusters are also processed under the same level of theory. For a more accurate reaction energy profile, the electronic energy for the product clusters and the TSs are corrected utilizing single point energy calculations at UB3LYP/6-311++G(d,p)+GD3 level, again with the same PCM model and basis set for Sn applied. Meanwhile, full geometric optimizations and frequency calculations are conducted at this level for the two reagents (B and AT). With all the optimizations done, the reaction energy profile for all the TSs and product clusters is recorded in the relative Gibbs energy at 298K from the sum of the two reagents.

Using the structures and the electron density from the high-level calculations, a geometric analysis and a joint AIM+NCI analysis (described in section 2.2.6) are conducted to reveal the nature of the bonding interactions within the clusters. There are two properties of interested in this analysis: one is the dissociation and formation of the bonds involved in the transfer of the A moiety as well as the behavior of the N-Sn transannular bond, and the other is the weak bonding interactions in between the B, A, and T moieties. For these purposes, several associated geometric parameters are collected from

the optimized TS structures. The electron density and the density Laplacian at the corresponding BCPs are calculated for each frame along the intrinsic reaction coordinate (IRC) pathway, forming a so-called IRC profile.

In this study, the Gaussian 09 Rev.E01 software package[141] is employed for all molecular mechanics and *ab initio* quantum calculations while the BH simulation and the TS search routine is implemented with Python 3.6. The wavefunction analysis (AIM and NCI) are conducted with the Mulitwfn software[189] and visualized with the VMD software.[190] For the AIM analysis, the CPs initial guesses are generated from a reasonable subset of nuclei positions, diatomic midpoints, triatomic centers, and tetratomic centers, as is implemented in Multiwfn. Post real space function values are also calculated utilizing the default settings. For NCI analysis, the default upper limit of 0.05 a.u. for electron density in Multiwfn is maintained. To ensure the quality of the isosurfaces, cuboid grids of 0.04 a.u. grid spacing are utilized.

4.3 Results

4.3.1 Structures and Relative Energies of the Three TS Series

With the search technique described earlier, a total of 3 TS series, namely 3-linear (TS3L), 4-linear (TS4L), and 4-cyclic (TS4C), are found. Thermodynamic parameters of the TS channels are summarized in Table 4-1. The label "3" and "4" in the series name is inherited from the resulting product while the linear (L) and cyclic (C) is defined by whether moiety T and moiety B stays on the different or the same side of the plane defined by moiety A. The representative lowest energy TS of each series and the resulting reaction energy diagram are shown in Figure 4-4 and Figure 4-5 respectively. Additional details of the transition state can be found in Appendix B.

As is shown in Table 4-1, TS4L has the lowest energy barrier of all the reaction channels, followed by the TS3L channel whose energy is 7.3 kJ/ mol higher. This result is in qualitative agreement with the experimentally determined regioselectivity of the substitution reaction. Meanwhile, optimizations on the T3 and T4 show that the former is the more thermodynamically stable product, which agrees with the chemical intuition that C_β is less substituted thus has less steric hindrance than C_δ. Another trend particularly worth noting is that the relative energy ranking of the TSs within one TS series are mostly dominated by their electronic energy difference, even with the entropic partition working against

it. This implies tracking the bond breaking and formation at the reactive sites is of crucial for a fundamental understanding of the regioselective alkylation.

TS3L   TS4L   TS4C
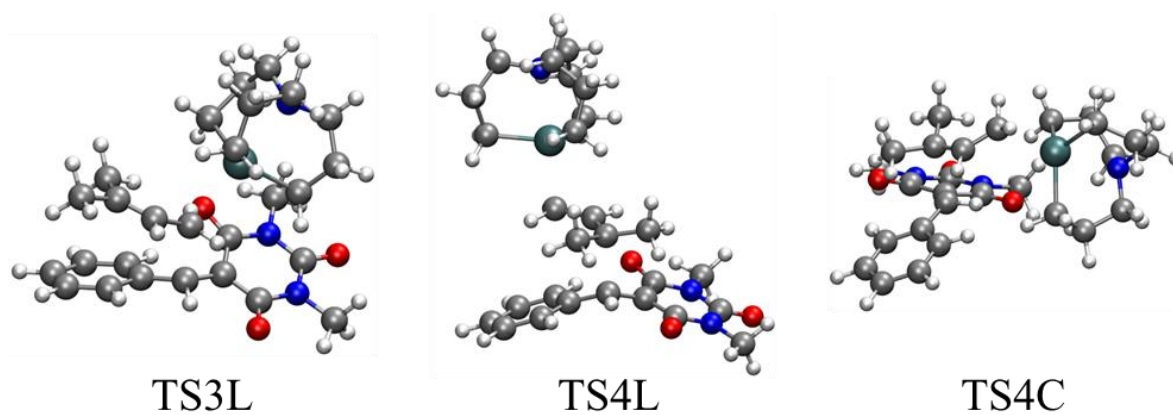
Figure 4-4. Representative TSs of the three TS series.
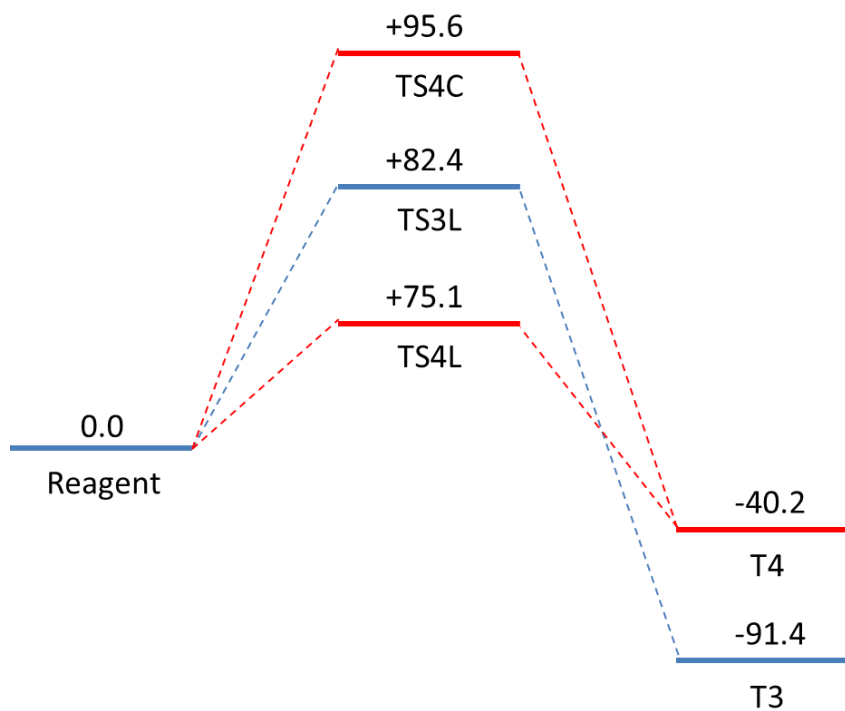
Figure 4-5. The relative Gibbs energy profile of the alkylation reaction. Energies are in kJ/mol

Table 4-1 Thermodynamic properties of the transition states and the product clusters. All properties are relative values to the sum of two reagent molecules.

		ΔE_{el} (mHartree)	ΔE_{ZPE} (mHartree)	ΔE_{el+ZPE} (kJ mol^{-1})	ΔH (kJ mol^{-1})	ΔS (J mol^{-1} K^{-1})	$\Delta G_{298.15K}$ (kJ mol^{-1})
TS3L							
	1	3.9	3.1	18.2	7.3	-217.8	82.4
	2	7.7	2.5	26.8	6.7	-196.7	85.5
	3	9.3	3.2	32.8	7.8	-195.4	90.6
TS4L							
	1	0.7	3.5	11.0	7.3	-221.2	75.1
	2	2.5	3.3	15.1	7.2	-212.6	77.1
	3	2.8	3.7	16.9	8.0	-216.5	79.8
	4	3.0	3.4	16.7	7.0	-219.0	80.1
	5	5.1	3.5	22.5	7.6	-207.1	82.7
TS4C							
	1	5.9	3.9	25.7	7.8	-242.6	95.6
	2	9.1	3.9	34.2	8.2	-235.5	102.3
T3		-62.5	5.8	-148.8	-149.9	-225.1	-82.8
T4		-51.7	6.5	-118.5	-121.4	-259.2	-44.1

4.3.2 Geometric Analysis of the Three TS Series

As a further step of the investigation, selected geometric parameters are collected and summarized in Table 4-2. In addition to the typical bond lengths, angles and dihedrals, a new pucker parameter, which is defined by the reactive center and 3 atoms that bond to it, is also introduced. As its name implies, this parameter measures the extent of puckering at a reactive site. For instance, the pucker parameter at C_δ is calculated in the following way if taken the two methyl groups connected to C_δ as M_1 and M_2 and the angle between them as $A(M_1,C_\delta,M_2)$:

$$Pucker(C_\delta) = abs(360 - A(M_1,C_\delta,M_2) - A(C_\gamma,C_\delta,M_2) - A(M_1,C_\delta, C_\gamma))$$

Similarly, $Pucker(C_\beta)$ is calculated utilizing the two hydrogen atoms and C_γ. With this parameter, one obvious difference between the TS3 and TS4 channels is that both the reactive carbons (C_β and C_δ) in TS4 channel have a puckering of about $10°$, while in TS3, puckering at the reactive site (C_β) is impeded. In addition, the NSn and Snβ bond length in TS4 are on average 0.08 Å longer and 0.1 Å shorter than those of TS3, which is evidence that the dissociation of T and A is more reluctant in TS4

61

compared to TS3 due to the puckering flexibility granted by its βδ bonding scheme. Looking at TS4 channels alone, the higher energy of the TS4C channel is mainly due to the steric repulsion imposed by the relative position of the B and T moiety in the "C" scheme. This argument is supported by the Snβγ angle being almost 20° larger in TS4C than in TS4L.
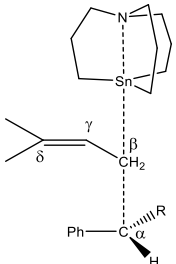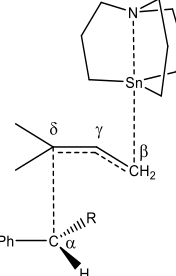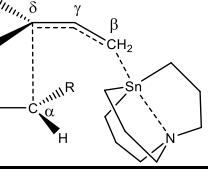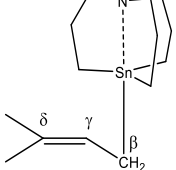
Within one channel, the TSs are differentiated by the relative orientation of the A and B moiety (in terms of Hαβγ and αβγδ). Generally, the Gibbs free energy variance for the TSs within each channel is about 10kJ/mol, whose largest contribution is the electronic energy difference. For each channel, the relative energy ranking of the TSs can be rationalized from different aspects. Note that in the following discussion, the TSs within one channel is named based on their relative energy ranking, as is labelled in Table 4-2.

For the TS4C channel, only two reasonable structures are obtained due to the steric hindrance between the T and B moiety. With intuition from general organic chemistry, one could notice that the Sn atom in T and the ortho-oxygens of C_α in B will bear a normal positive and negative charge as the reaction process. Thus, the relative energy order can be rationalized with the Sn-O distance (3.46 Å in 1 vs. 4.29 Å in 2).

For the TS4C channel, only two reasonable structures are obtained due to the steric hindrance between the T and B moiety. With intuition from general organic chemistry, one could notice that the Sn atom in T and the ortho-oxygens of C_α in B will bear a normal positive and negative charge as the reaction process. Thus, the relative energy order can be rationalized with the Sn-O distance (3.46 Å in 1 vs. 4.29 Å in 2).

Like the case of TS4L, the 3 TSs of the TS3L channel also adopts a sandwiched structure. However, for these TSs, the conformation at C_β deviates strongly from the ideal triangular bipyramid shape. This distortion is maximized in the lowest energy TS, resulting an Snβα angle of 135°. The source of the distortion could be either a steric repulsion between moiety A and B or an electrostatic attraction between moiety T and B, the same type between Sn and O in the case of TS4C analyzed above. If one further examine the Snβα, Snβγ, αβγ angles of the three TSs, however, one could notice that compared to TS 2 and 3, the 25° Snβα decrease in 1 is redistributed to the Snβγ and αβγ angles in an almost equal manner. This observation supports the latter argument that the attractive Sn-O interaction between motif T and B is the cause of the distortion.

Table 4-2 Selected geometric parameters of the transition state

TS Series		Bond(Å)			Angle (°)				Dihedral (°)		Pucker (°)	
TS3L		NSn	Snβ	βα	NSnβ	Snβα	Snβγ	αβγ	Hαβγ	αβγδ	β	δ
	1	2.42	2.53	2.22	178	135	108	116	−73	−82	0.1	0.0
	2	2.41	2.54	2.18	178	157	97	106	95	−79	0.7	0.1
	3	2.41	2.59	2.15	177	161	96	103	158	−88	1.0	0.0
TS4L		NSn	Snβ	δα	NSnβ	Snβδα	Snβγ	αδγ	Hαδγ	αδγβ	β	δ
	1	2.51	2.37	2.15	176	172	104	95	−168	96	12.2	8.0
	2	2.47	2.43	2.13	175	175	102	94	64	89	8.9	11.0
	3	2.48	2.42	2.11	175	180	105	95	−36	−95	10.3	10.9
	4	2.49	2.39	2.10	175	176	105	96	160	−90	11.5	9.9
	5	2.49	2.41	2.12	176	177	104	94	−35	93	11.0	10.3
TS4C		NSn	Snβ	δα	NSnβ	Snβδα	Snβγ	αδγ	Hαδγ	αδγβ	β	δ
	1	2.48	2.40	2.24	174	29	121	97	−25	109	16.8	9.9
	2	2.50	2.42	2.14	171	27	121	96	22	−107	15.0	12.7
A-T		NSn	Snβ		NSnβ		Snβγ				β	δ
		2.68	2.23		179		111				27.3	0.0

4.3.3 AIM-NCI Analysis

The AIM-NCI analysis results on the reagents and the representative TSs of all channels are shown in Figure 4-6 and Figure 4-7, while the electron density, ρ and the density Laplacian, $\nabla^2\rho$, at selected BCPs are tabulated in Table 4-3. Generally, the AIM analysis reproduces the covalent bond network and additionally signals some unintuitive weak bonding interactions. The NCI analysis, on one hand, gives consistent results with AIM regarding bond breaking and formation, but on the other hand, signals additional weak interactions missing from the bonding topological network. As a proof of completeness,

for each structure analyzed, the number of different CPs satisfies the Poincaré-Hopf relation.[191] Searching within the signified regions by the NCI analysis gives no additional CPs to the best effort.

As is shown in Figure 4-6, the central Sn atom adopts a pentavalent bonding scheme. All five bounds are classified dative bonds after comparing their ρ and $\nabla^2\rho$ values at BCP with literature.[190] However, the transannular SnN bond is significantly weakened due to the steric repulsion imposed by the triptych cage, resulting a smaller ρ at its BCP compared to that of the Sn-C bonds. With the help of NCI analysis, one can visualize this steric repulsion as red isosurfaces inside the cage. For the barbituric acid, regions of weak interactions are found between the carbonyls and their neighbouring groups while regions for repulsive interaction centered around the RCPs. To simplify the TS results in the coming discussion, trivial intra-fragment CPs, viral paths, and interaction surfaces, as is originated from the two reagent molecules, are omitted.

To ease visualization, the AIM-NCI results for the TSs in Figure 4-7 are separated into the strong bonding interactions (above) along the N-Sn, Sn-β, and $\beta(\delta)$-α bonds, and the weak ones (below) whose isosurfaces spreads in between the fragments within the system, intercepting with several inter-fragment bonding paths at their CPs. Comparing the ρ and $\nabla^2\rho$ at BCPs in the TSs with those within the bare A-T complex in Table 4-3, an increase in these values at BCP_{NSn} and a decrease at $BCP_{Sn\beta}$ is observed, as is expected for the A moiety to transfer. In addition, the ρ at BCP_{NSn} and $BCP_{Sn\beta}$ in the TS4 channel is in average 0.8 a.u. lower and 1.7 a.u. higher than those in the TS3 channel, which justifies the order of reluctance in the Sn-β dissociation. Between the TS4L and the TS4C channel, however, these values do not have a significant difference, which means it is the steric repulsion in between the B, A, and T moieties that account for the energy differences of the two channels.

All the representative TSs share common types of weak interaction between moiety B and A, and moiety A and T, namely AB and AT interactions. For TS4L alone, the BT interaction is missing due to the separation of moiety A, and the AT and AB interaction retains approximately shape across all the TSs. This explains the small energy difference between the TSs in TS4L series. Another interesting discovery is that each of the five TSs in this channel has five bright green attractive regions on their AB interaction surface, which is self-consistent with the fact that only 5 TSs have been found for this channel regardless the geometric parameters predicting 6.
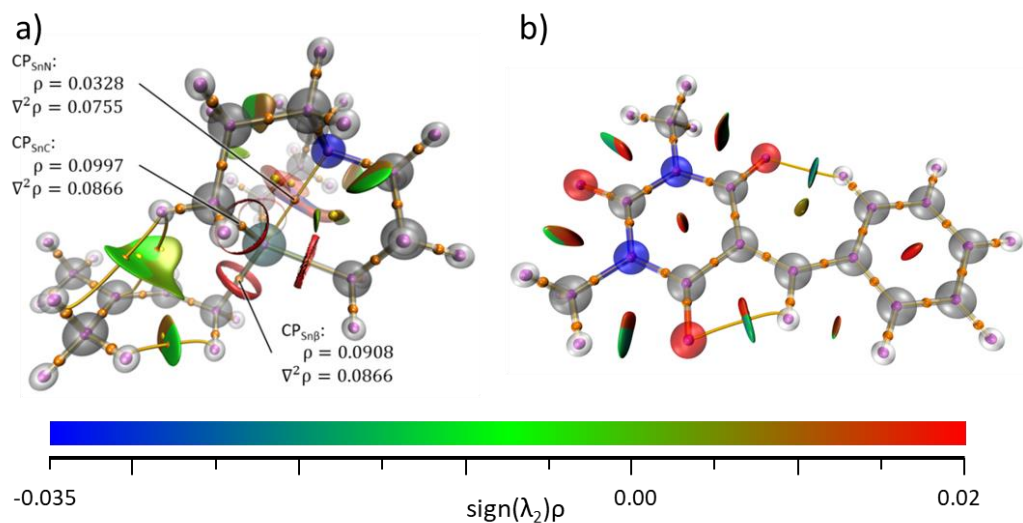
Figure 4-6. AIM and NCI analysis of two reagents: a) A-T complex and b) B, values are in au.

Table 4-3 Electron density (in a.u.) and density Laplacian (in 10^{-2} a.u.) at selected critical points. Notation follows those used in the geometric analysis.

		CP_{NSn}		$CP_{Sn\beta}$		$CP_{\beta\alpha}/CP_{\delta\alpha}$	
		ρ	$\nabla^2\rho$	ρ	$\nabla^2\rho$	ρ	$\nabla^2\rho$
TS3L							
	1	5.4	13.5	4.5	6.2	5.5	4.9
	2	5.5	13.6	4.3	5.8	6.0	3.8
	3	5.5	13.8	4.0	5.4	6.4	3.5
TS4L							
	1	4.6	11.0	6.5	7.1	6.5	2.5
	2	4.9	11.9	5.7	6.4	6.9	1.8
	3	4.8	11.7	5.8	6.4	7.1	1.4
	4	4.7	11.3	6.1	6.8	7.2	1.5
	5	4.7	11.4	6.0	6.6	7.0	1.6
TS4C							
	1	4.8	11.6	6.1	6.8	5.5	2.9
	2	4.6	11.1	5.8	6.5	6.8	1.6
	A-T	3.3	7.6	9.1	8.7	-	-
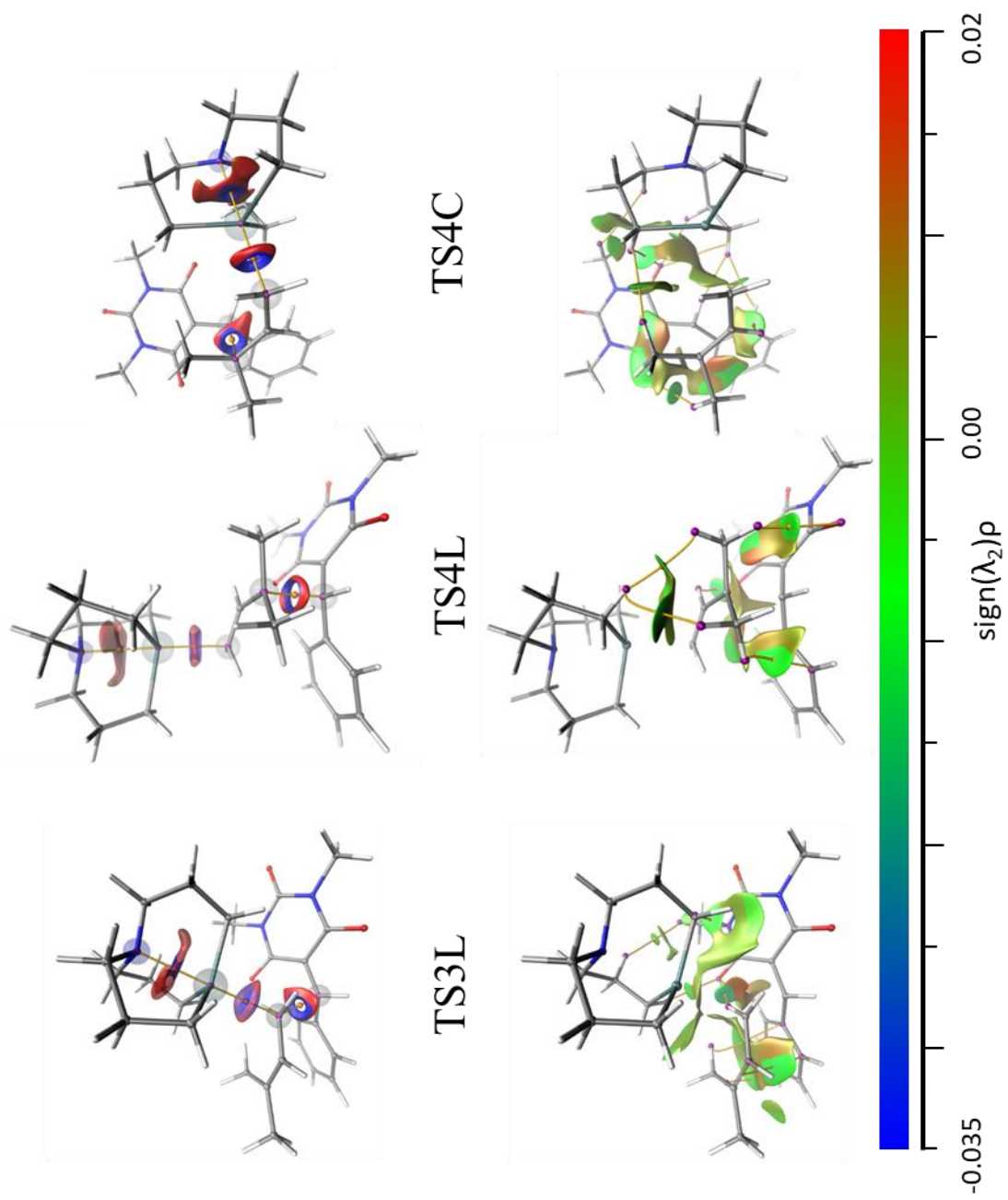
Figure 4-7. AIM and NCI analysis of the representative TSs. Top is the primary bonding and bottom is secondary inter-fragment interactions.

sign(λ_2)ρ

TS3L

TS4L

TS4C

-0.035

0.00

0.02

For the TS4C channel, the BT interaction can be observed because of the 'C' binding scheme. In agreement with the finding previously, this interaction is a crucial part of the energy difference between TS4L and TS4C. For the TSs in this channel, 1 has 3 bonding paths between fragment T and the oxygen site on fragment B, while for 2, there is only 1 (see Figure 4-7). This observation partially verifies the conclusion drawn earlier that the BT interaction dominates the TSs' energy difference. In addition, the bonding paths shows no evidence for the existence of additional Sn-O bonding interactions, thus the triptych is still pentavalent coordinated in all TSs.

For TS in the TS3L channel, their relative energy ranking is inversely related to the volume of the BT interaction surface and the number of associated bonding paths, which is a good indication that the BT interaction is a good descriptor of the relative energy, and the source of distortion to the $Sn\beta\alpha$ angle.

4.3.4 AIM-NCI Analysis Along the IRC Pathway

To further examine the reluctance in the ligand dissociation in the TS4L channel, the ρ_{BCP} of NSn, $Sn\beta$, and $\beta(\delta)\alpha$ as a function of IRC coordinate are collected for the three representative structures and plotted in Figure 4-8. The bond dissociation limit is chosen when the electron density has committed 50% of its total variance over the IRC steps of consideration, and label them by the dashed droplines. However, this definition may not be a suitable choice for the $\beta(\delta)\alpha$ bond as is shown in the study of Popelier et al.[192] showing this bonding threshold occurs at about 40% of the overall variance for a symmetric X-C-X S_N2 transition states. Despite the systematic deficiencies, two observations are still worth noting. One is the synchrony between the NSn bond formation and $Sn\beta$ bond dissociation, as is demonstrated in Figure 4-8 a) and b) that approximately both the NSn and the $Sn\beta$ threshold lies at IRC index of 0 for the two TS4 channels and IRC index of -0.5 for the TS3 channel. The other is the overall reluctance of $Sn\beta$ dissociation in the TS4 channels compared to that in the TS3 channel, given that the $\beta(\delta)\alpha$ bond formation is of comparable progression. An indication from this observation is that the charge separation associated with the $Sn\beta$ bond breaking is key to the activation energy difference between the TS3L and TS4L channel, which implies there is a trade-off between regioselectivity and the rate of the alkylation.
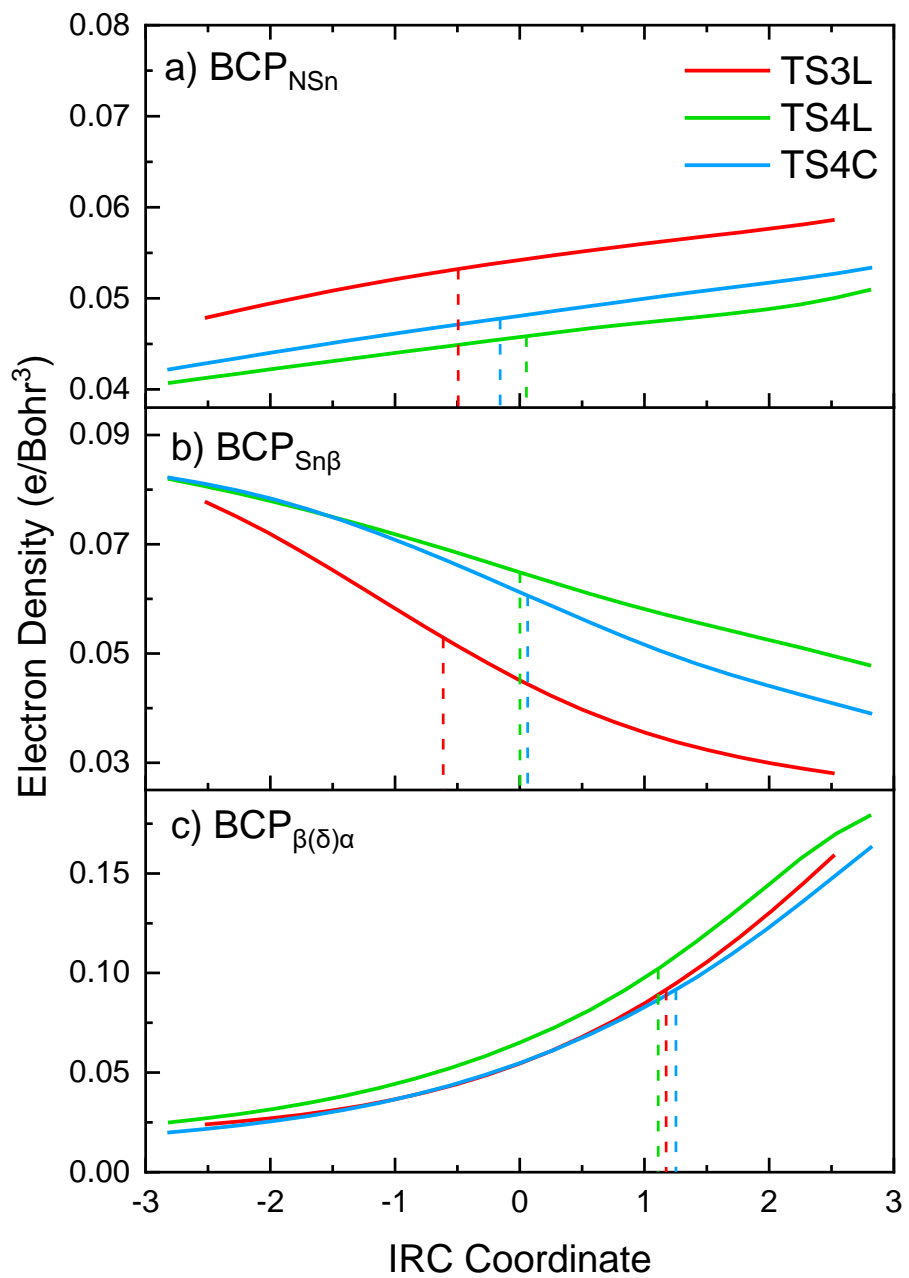
Figure 4-8. IRC profile of the electron density at BCP for the three representative TSs.

4.4 Chapter 4 Summary

In this study, three reaction channels have successfully been identified for the alkylation of a barbituric acid derivative with the 3-methylbut-2-en-1-yl triptych complex utilizing the structure interpolation technique. Despite an initially presumed cyclic structure, one eventually found that in all TSs, the $N - Sn - C_\beta$ atoms tend to remain linear rather than adopting hexavalent coordination as the substitution proceeds. The absence of the more thermodynamically stable product **3** is rationalized by the identification of a higher activation energy of the TS3L reaction channel than in the TS4L channel. Based on comparison with literature results, five dative bonds are found in the complex, with CPs and bonding paths originating from the Sn atom, which is consistent with its pentavalent coordination nature in the complex. Further examination of the electron density values at the CP of the $N - Sn$, $Sn - C_\beta$, and $C_\beta/C_\delta - C_\alpha$ bonds reveal that the bonding strength of the $Sn - C_\beta$ and $C_\beta/C_\delta - C_\alpha$ bonds are stronger in the TS4L channel, while the $N - Sn$ bond is stronger in the TS3L channel. This suggests that in the TS4L channel, the ligand transfer has less impact to the dissociation of $Sn - C_\beta$ bond than in the TS3L channel, which can be rationalized by the larger extent of puckering at C_β and C_δ in TS4. While the energy difference between TS4L and TS3L is attributed to an electronic effect, energetic differences between TS4L and TS4C predominantly arise from steric hindrance.

Furthermore, the ligand transfer process of the alkyl-triptych complex is monitored by electron density at the BCP of the $N - Sn$, $Sn - C_\beta$, and $C_\beta/C_\delta - C_\alpha$ bonds as the transfer proceeds along the IRC coordinate. From the electron density profile, the $N - Sn$ transannular donation, the $Sn - C_\beta$ bond dissociation, and the $C_\beta/C_\delta - C_\alpha$ bond formation is found. With a similar extent of $C_\beta/C_\delta - C_\alpha$ formation, further progression in the $N - Sn$ donation and $Sn - C_\beta$ dissociation is observed in the TS3L channel than in the TS4L/TS4C channel. To sum up, this work suggests that the preference of β substitution may be enhanced by increasing steric hindrance at C_δ and introducing electron withdrawing groups at C_β, whereas enhancing C_δ substitution may require the opposite.

Chapter 5

Spectroscopic Characterization of an Acylhydrazone Photoswitch

5.1 Introduction

Photochromic compounds are those which can reversibly transform between a thermally stable ground state and some metastable excited state(s) upon light irradiation.[13,193,194] Depending on the mechanism of the back transform to the ground state, these compounds can be classified as T-type (thermal-induced) or P-type (photo-induced).[13] Examples of such compounds include natural proteins like the visual pigment Rhodopsin[195] and synthetic photoswitches like diarylethene, azobenzene, and acylhydrazone derivatives.[196] By attaching photoswitch moieties to other systems such as protein molecules[15,197,198] and supramolecular systems,[199] or embed them into condensed phase material,[14,200] one can modify certain properties of the system, *e.g.* the conformation of a protein[198] or mechanic properties of a material,[14] via photo-isomerization of these switches with controlled light irradiation. The performance of artificial photoswitches are assessed in terms of addressability, thermal stability, efficiency, and reliability.[13] In terms of molecular properties, the performance of a photoswitch molecule is closely related to: the position of the isomer UV-Vis absorption bands, their respective intensities (or molar extinction coefficients), and the kinetics of photo- and thermal-isomerization.

Despite the usefulness of photoswitch molecules, tunning a photoswitch molecule for desired properties is generally difficult.[13] Some of the difficulty in this regard is due to the lack of guidance from theoretical models. To establish such models for the photoswitches, a good starting point is to explore their gas phase behavior since the gas phase ions are not subject to matrix effects and thus may be well-modeled with the ab initio calculations. In this work, DMS-UVPD-MS studies are conducted, and experimental observations are supported with high level ab initio calculations. The DMS technique has been shown capable of separating prototropic isomers in several cases[46,47,201] and may thus be employed to select different conformers in a given ion population. Subsequent laser interrogation of the DMS- and MS-selected species then yields their distinctive UVPD spectra.

In this work, (E)-N'-benzylidenebenzohydrazide ($t_{1/2} = 145 min$, $\lambda_{max,E} = 297 nm$, $\lambda_{max,Z} = 388 nm$)[202–204] is chosen as the subject compound. Its neutral and protonated forms are denoted AY and [AY + H]$^+$ henceforth. There are several considerations in choosing this compound. Firstly, preliminary tests indicate the compound has short term (*ca.* 1 min) stability in an acidic environment, and can be

ionized by ESI and detected by MS. Secondly, the half-life for thermal back-isomerization, $t_{1/2}$ of this compound reported in literature is comparable to the experimental timescale (at least 1-2 h for each UVPD run). It is also reported that the cis and trans isomeric forms have distinctive λ_{max} in the UV region, which can facilitate unambiguous spectroscopic characterization. Thirdly, the compound isomer can adopt several skeletal and prototropic forms that can be potentially separated by DMS. Finally, the chosen compound is of moderate molecular size and geometric complexity, and theoretical calculations are feasible.

5.2 Method

5.2.1 Experimental Method

The AY ((E)-N'-benzylidenebenzohydrazide) sample in this study is generously provided by Dr. Mónica Barra and Ho Yin Chan, Department of Chemistry, University of Waterloo. In this experiment, the on-board ESI source of the Qtrap 5500 system is used to generate gas phase ions. The spraying voltage is tuned between 4500 V and 5500 V for optimal ion intensity. The AY stock solution is prepared by dissolving 0.1-1.0 mg of AY crystals in pure acetonitrile. To form an optimal $[AY + H]^+$ ion intensity of $10^6 \sim 10^7$ via ESI, the AY solution is usually diluted in methanol (with 1% acetic acid) to 1 - 5 μg/mL. This sample preparation procedure results from a balance between the requirement of ESI for an acidic and protic solution to stably spray cations and the chemical instability of acylhydrazone derivatives under such environment.

As a precaution for potential sample degradation (See Appendix C), a late mixing setup is employed in the experiments to ensure the ion composition from source is constant over the laser scan time frame. As is shown in Figure 5-1, two 1mL gas tight syringes are utilized, one containing the diluted methanol solution of AY, and the other containing 2% acetic acid solution in methanol. Using two segments of PEEK tube of similar length (*ca.* 1.2m long), the two syringes are connected to the two equivalent inlets of the Tee joint. The outlet of the Tee joint is connected to the ESI source with another 62 cm long tube. The two syringes are pumped simultaneously by one syringe pump, and the analyte and the acid solution is mixed at the Tee joint. There are two associated consequences with this setup: one is that the concentrations of both the analyte and the acid are halved after mixing, and the other is that the flow rate at the ESI source is twice the apparent pumping rate. To account for the dilution during the mixing, the concentration of the analyte and the acid solution are doubled according. Usually, the flow rate of

71

the pump is set to $4 - 6 \, \mu L/min$ for optimal ion signal quality, giving a pumping rate of 8-12 $\mu L/min$ at the inlet of the ESI source. Given the 62 cm tube length and the 0.127 mm inner diameter (with a 10% error),[205] the inner volume of the tube is approximately 7.8 μL. Using the pump rate stated earlier, it can be estimated that prior to ESI, the acidified sample solution is in a constant state of "1-min-after-mixing". Utilizing the setup in Figure 5-1, The $[AY + H]^+$ ion is observed under positive mode ESI when spraying from a methanol solution of 1-2 µg/mL AY with 1% acetic acid. Additionally, sodiated and chlorinated complexes ($[AY + Na]^+$ and $[AY + Cl]^-$) are also found when spraying solution of AY in acetonitrile saturated with NaCl. However, no deprotonated $[AY - H]^-$ ions are found when spraying its acetonitrile solution mixed with up to 0.5% ammonia.
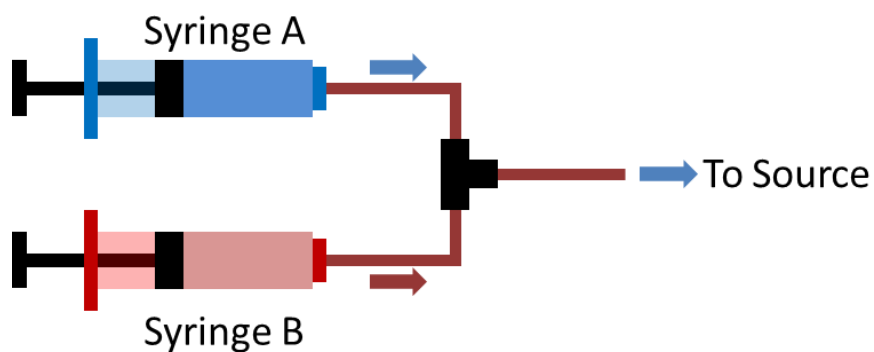
Figure 5-1 The late mixing setup

Upon ESI, the $[AY + H]^+$ ions are guided through the on-board DMS cell to the triple quadrupole mass spectrometer for further analysis. Unless otherwise specified, the default parameter setup (as is defined in Appendix D) is employed for each experiment. Prior to any DMS related experiments, the existence of the $[AY + H]^+$ cations is determined by preliminary Q1 scans within the mass range from 220.0 to 230.0, which reveals a nominal mass of around 225.1 for the parent $[AY + H]^+$ ion. Then the ion at m/z 225.1 is further verified by CID experiments under the EPI scan mode, which examines the fragment between 70.0 – 230.0 nominal mass from the 225.1 parent ion. The CE of this experiment ramps from 5 eV to 100 eV at a step of 1 eV. The result of the CID experiments shows fragments of m/z 77.0, 104.0, 105.0, 122.0, 147.0, and 207.1. It is also determined that the CE for optimal parent ion transmission is 10 eV, which is the default setup for the DMS scans. Meanwhile, CE = 20 eV gives a satisfactory fragmentation pattern, which is used to confirm the identity of the parent ion at the beginning of each experiment run.

Utilizing the optimal parameters for the EPI mode determined in the preliminary experiments, conventional DMS analysis is conducted to search the viable separation conditions for the latter spectroscopic experiments. For all DMS experiments, the default SV parameter ranges from 0 V to 4000 V unless the instrument stalls or the signal to noise ratio is unacceptable. To better track the behavior of the separated ion population at high SV, the SV ramping step is 500V when below SV = 2000V halved to 250 V when above. The ramping range of CV is adjusted individually under each SV to include all the elucidated peaks. To achieve a better resolving power, 3 modifiers, namely methanol, water, and acetonitrile are chosen to enhance the resolving power of DMS. These modifiers are injected to the instrument via the outlet of the modifier pipeline using an Agilent G1312A pump such that the modifier concentration in the DMS cell retains 1.5 % (See Appendix E entry (a) for modifier flow rate). For separations with no modifier (*i.e.*, the N_2 environment), the DR parameter is set medium to enhance peak resolution.

For the DMS resolved species, both the CID and the UV-Vis photodissociation experiments are conducted to characterize their thermo- and photo-dissociation properties. The choice of DMS separation condition depends both on the peak resolution of the separated species and on the intensity of each elucidated peak. To get a high-resolution CID profile for each of the resolved species, the CE is ramped from 5 eV to 30 eV at a step of 0.1 eV. The CE upper bound of 30 eV is chosen since the intermediate fragments are most abundant within this range according to the preliminary scans. To account for species of low ion signal intensity, the LIT time is adjusted accordingly to a maximum of 5 ms, and a sum of up to 10 scans is adopted.

For the UV-Vis photodissociation experiments, Q_1 gates on the parent ion of m/z 225.1 while the Q_3 trap collects fragment signal whose nominal mass lies between m/z 70.0 and 230.0. The LIT is set from 1 to 10 ms depending on the ion signal abundancy such that the parent ion intensity lies between $10^6 \sim 10^7$ without laser irradiation. To minimize possible collision induced heating prior to the laser irradiation, the DP, EP, and CE adopts their minimum setup, which is 0 V, 2 V, and 5 eV, respectively. The oscillator delay of the pump laser generally ranges from 180 to 200 ms, and the optical scan range is set from 208 nm to 380 nm (3.26 – 5.96 eV) at a step of 1 nm. To power-normalize the spectra, the power of the laser pulse at each scanned wavelength is found by an average of 100-time measurement during the same experiment period.

Data acquisition for the above experiments is accomplished with the Analyst software that accompanies the Qtrap 5500 system. The OPO laser is remotely controlled with the LABVIEW software. Synchronization between the Qtrap 5500 system and the OPO laser is achieved with a custom python script using the pyautogui package. Post data extraction and processing for both the CID and the UV-Vis photodissociation experiments are achieved using a home-built python program interfaced with the ProteoWizard software[154–156]. Details of data processing can be found in Section 2.1.4.

For both the CID and the UV-Vis photodissociation experiments, the candidate fragment mass channels being monitored are chosen from those determined from the preliminary CID experiment and are then adjusted within 1 m/z range for individual experiments based on their raw data. In principle, the adjustment should account for the inevitable peak shift in the MS instrument, the ion signal continuity, and the interference from contaminants. To compare the spectra from different experiments, they are normalized against the maximum signal intensity to a range of 0 to 1.

5.2.2 Theoretical Calculations

As is shown in the skeletal form of AY shown in Figure 5-2, most of the skeletal carbon and nitrogen atoms are sp^2 hybridized according to the classical VSEPR theory. This means the number of possible conformations of the $[AY+H]^+$ isomers is quite limited due to the possible $\pi - \pi$ conjugations within the molecules. By an exhaustive combination of all the possible protonation sites (on the O^1, N^3, and N^4 sites) with all the Z/E conformations along C^2-N^3, N^3-N^4, and N^4-C^5, one could potentially find 24 (3×2^3) AY and $[AY+H]^+$ isomers. With the energetically disfavored isomers omitted (*e.g.* isomers exhibiting Z conformation at both N^3-N^4, and N^4-C^5), nine $[AY+H]^+$ isomers are left for further investigation in this research. To facilitate in-context structural recognition, the skeletal structure of the $[AY+H]^+$ isomers are encoded in the Z/E conformation along the C^2-N^3, N^3-N^4, and N^4-C^5 bond. Meanwhile the protonation scheme is defined as follows: 'O' means protonating on site O^1 and N^3, 'N_A' means protonating on site N^3 and N^4, and 'N_B' means protonating on site O^1 and N^4. A shortened name for each isomer is also assigned, as is defined in Table 5-1. Based on the nomenclature established above, the conformation shown in Figure 5.2 is encoded AY-ZEE-N_A, and it has a shortened name of 1a.

For the $[AY+H]^+$ isomers selected above, they are first optimized under ωB97XD/def2-TZVPP level of theory, as have been done in similar studies involving UV spectrum prediction.[111] Following the geometry optimization, a frequency calculation is performed to both verify the minimum and to enable

thermodynamic calculations. The [AY+H]⁺ isomers are optimized both with and without methanol PCM to account for the corresponding environment when these isomers are in the mass spectrometer for laser irradiation and in the sample solution prior to electro-spraying. The geometry optimization and frequency calculation in this step is achieved using the Gaussian 16 Rev. A03 software package.[206] The solvent model at this step is chosen the IEFPCM model implemented in the Gaussian 16.
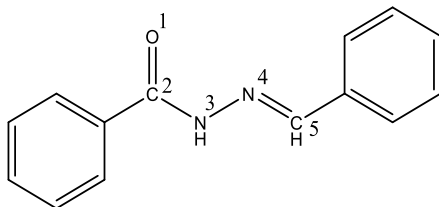
Figure 5-2 The skeletal formula of AY

After the first step, the relative Gibbs energy of the AY and the [AY + H]⁺ isomers are calculated at T = 298.15 K. Then isomers lying within a Gibbs energy threshold of 40 kJ mol⁻¹ of the global minimum are carried forward for DLPNO-CCSD/def2-TZVPP single point energy calculation, which is performed using ORCA 4.2.0[144,145] with a def2-TZVPP/C auxiliary basis and the default NormalPNO setup. To facilitate UV spectra simulation in latter steps, the chosen isomers are also optimized under ωB97XD3/def2-TZVPP level of theory as is implemented in ORCA, following a frequency calculation for verification and thermal correction. To account for the solvent environment, the CPCM model built in ORCA is employed. DLPNO-CCSD single point energy calculations are performed on both the ωB97XD and the ωB97XD3 minima to correct their electronic energy. The corrected Gibbs energy is calculated in the following equation:

$$G_{CCSD} = G_{DFT} - E_{DFT} + E_{\text{CCSD}} \tag{5.1}$$

In eq. 5.1, G_{CCSD} is the corrected Gibbs energy that combines the electronic energy, E_{CCSD}, from the DLPNO-CCSD calculation with the thermal energy ($G_{DFT} - E_{DFT}$) from the DFT frequency calculation.

As is discussed in section 2.1.2, DMS resolved species are kinetically trapped and cannot freely interconvert due to large conversion energy barrier. To account for the possible kinetic trapping effect, a transition state (TS) search is performed for the [AY + H]⁺ isomers in gas phase. The level of theory chosen is ωB97XD3/def2-TZVPP, the same to that for the minima optimization. Given the simplicity of the [AY + H]⁺ isomer structures, the initial guess of the transition states can be found intuitively or

by constrained optimizations. In addition to the one-imaginary-frequency criteria introduced by the frequency calculation, all the TSs are also verified for connecting the desired pair of minima. This is achieved by an intrinsic reaction coordinate (IRC) calculation or optimizations of slightly distorted TS structures along the imaginary vibration mode. Once desired TSs are located, the relative Gibbs energy of those transition states are also corrected with the DLPNO-CCSD calculation as is done to the minima structures. With both the energies of the minima and the transition states, the potential energy landscape is represented in a disconnectivity graph[187] (see section 2.2.6 for more details) plotted using the pele package[13]

To rationalize the results of the gas phase laser experiments, theoretical UV-Vis spectra of the $[AY + H]^+$ isomers are simulated via both TD-DFT and EOM-CCSD methods. For both the TD-DFT and the EOM-CCSD calculations, the basis set and the solvent model is inherited from those used in the optimization procedure. For the EOM-CCSD calculations, the STEOM-DLPNO-CCSD is employed for its low computation cost. The solution phase TDDFT spectra is represented in the form of stick spectrum while the STEOM-CCSD ones are simulated with a Gaussian peak of 58.8 m^{-1} half width at half maximum (corresponds to an INLINEW parameter of 50 in the ORCA program). For gas phase isomers, the vibronic Franck-Condon spectra are simulated in the vertical gradient (VG) mode using a INLINEW setup of 50. In the VG mode, it is assumed that the hessian matrix for the ground state is identical to that of the excited state, which is useful for this case since attempts to consistently locate the minima on the first, second, and third excited state potential energy surface fail in the preliminary trial calculations. Specially in the VG simulation, a combination of the ground state hessians obtained under ωB97XD3/def2-TZVPP level of theory with the STEOM-CCSD vertical transition energies and dipole moments is achieved using the DELE and the TDIP keyword of the ESD module in ORCA.

To further characterize and assign the electronic transitions being explored, natural difference orbital (NDO) analysis is performed on the major STEOM-CCSD vertical excitations of the low energy isomers. In summary, the NDO analysis calculates the spatial electron density difference between the ground state and the excited state of interest.[13] In this study, the density difference is calculated against a $300 \times 200 \times 200$ data grid that encapsulates the molecule. The step size of the grid fluctuates around 0.1 Bohr due to the varying volume of the individual isomers. Using the VMD software package,[13] the iso-surface of both electron detachment and attachment density are simulated using a threshold of 0.0004 a.u.

76

5.3 Results and Discussion

5.3.1 Generation and Verification of the [AY + H]$^+$ Ions in the Gas Phase

Prior to the DMS and UV-Vis experiments, the key fragments of [AY + H]$^+$ are determined via CID using a CE of 20 eV. As is shown in Figure 5-3a, the parent peak is observed at m/z 225.1. The most abundant fragment peak occurs at around m/z 104.0, which can be assigned to the [PhCHN]$^+$ cation. This fragment can be easily formed by breaking the N^3-N^4 bond (see Figure 5-2). The peak at m/z 105.0 can be assigned to [PhCO]$^+$, [PhCHNH]$^+$, and/or isotopologues of [PhCHN]$^+$, while the peak at m/z 77.0 is attributed to [C$_6$H$_5$]$^+$. The three mass peaks at m/z 122.0, 147.0, and 207.0 likely arise from the loss of [PhCN], [C$_6$H$_6$], and [H$_2$O] from the parent ion. Major fragments are labelled on Figure 5-3a. The breakdown curve for [AY + H]$^+$, which is shown in Figure 5-3b, reveals that the m/z 77, 104, and 105 fragments are the dominant product ions up to collision energies of ca. 80 V.

5.3.2 DMS Investigations

Dispersion plots of [AY + H]$^+$ under different modifier conditions are shown in Figure 5-4. In a pure nitrogen environment (Figure 5-4a), two Type-C ion populations are observed. Baseline resolution for these populations is achieved at SV = 2750 V by setting the resolving gas (DR) flow to medium. Figures 5.4b-d show the dispersion behaviour of [AY + H]$^+$ in N$_2$ environments that have been modified with 1.5% water, methanol, and acetonitrile, respectively. A progressively stronger clustering behaviour is observed for the subpopulation associated with the black trendline. The subpopulation associated with the red trendline remain Type-C for water and methanol modifiers and is not observed for an ACN-modified environment. Of the four modifier condition tested, only the pure N$_2$ (SV = 2750 V, DR = Medium, CV = +1.9/+3.8 V) and the methanol modified (SV = 2000 V, DR = Off, CV = −6.3/+0.9 V) environments yield baseline ion separations that is suitable for UVPD experiments,. These separation conditions are highlighted in Figure 5-4 and the corresponding ionograms are shown in Figure 5-5. Note that under pure N$_2$ environment (Figure 5-5a), the optimal CV is slightly shifted to +3.8 V from the peak center (+3.5 V) to avoid potential contamination from the other separated ion population centered at +1.9 V. Separation obtained under water modifier is not reproducible, as is documented in Appendix C.
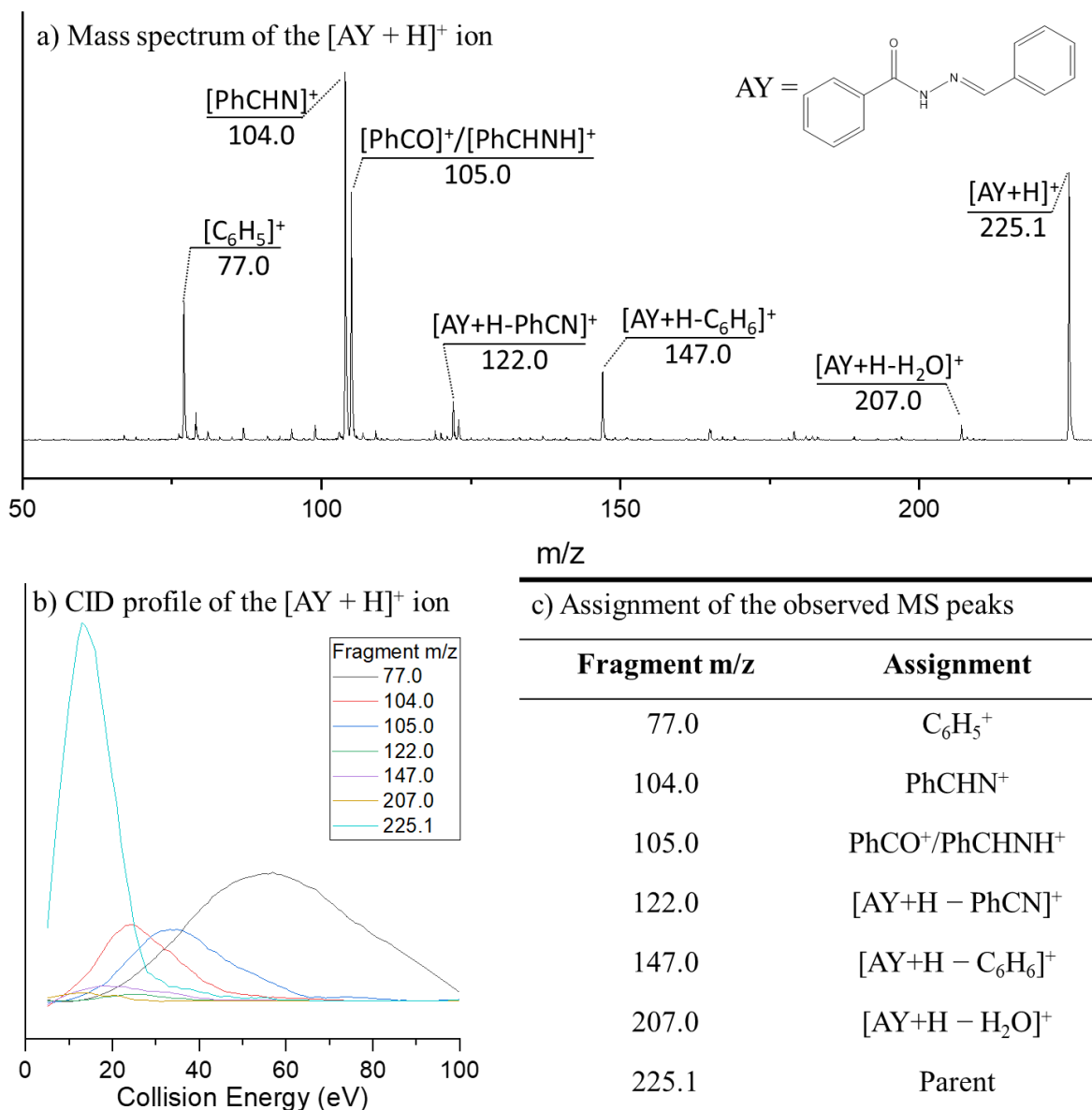
Figure 5-3 a) The mass spectrum, b) the breakdown curve of the [AY + H]$^+$ ion and c) the assignment for observed mass peaks

78
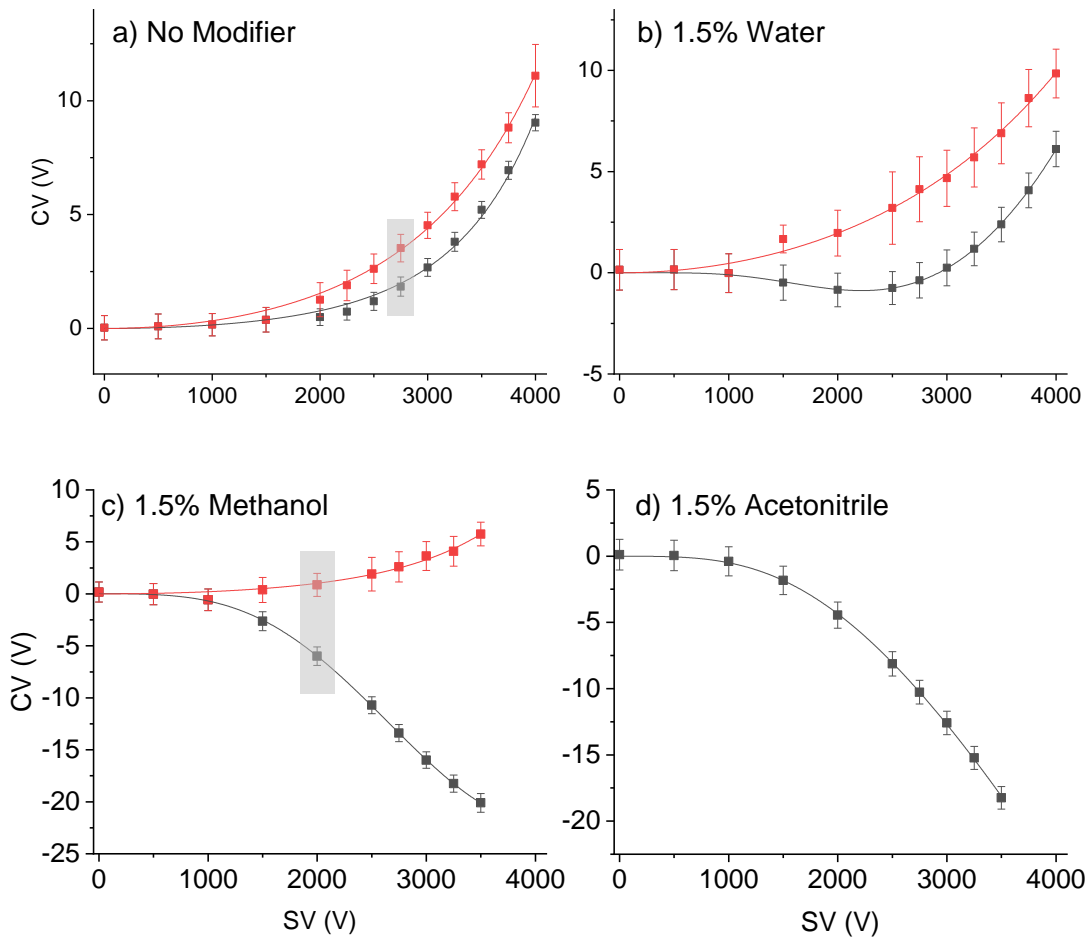
Figure 5-4. The dispersion plot of the [AY + H]+ ion with different modifiers applied.
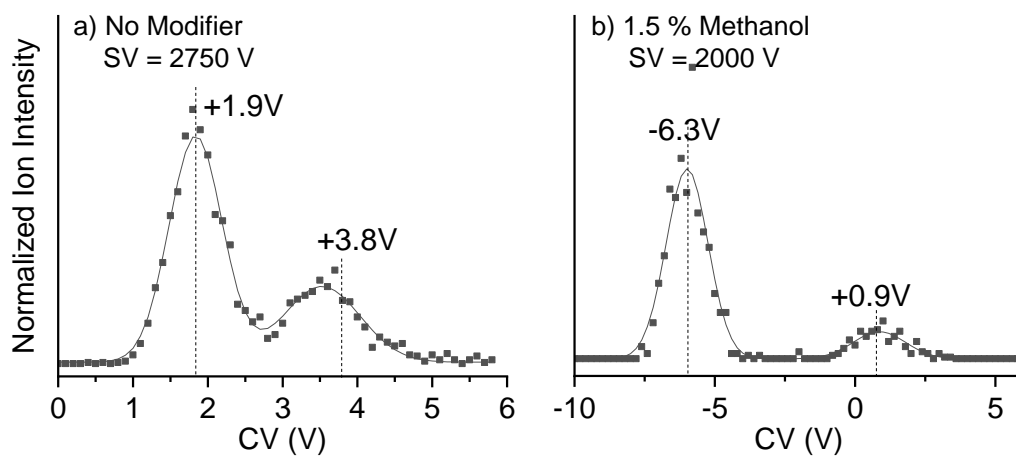
Figure 5-5. The ionograms at optimal separation conditions

79

To characterize the resolved ion populations, the CID breakdown curves for each separated ion populations are shown in Figure 5-6. Note that the m/z 122.0 fragment is present in the mass spectra but is excluded from the breakdown curve data series because of its relatively low intensity and high signal to noise ratio. For both ion populations separated either under pure N_2 or under 1.5% methanol modifier condition, the formation of the m/z 104.0 fragment, which correspond to breaking the C^2-N^3 bond, is dominant compared to the formation of other fragments.

Despite being resolved under different DMS conditions, the two observed subpopulations exhibit essentially identical fragmentation patterns. The minor difference in the breakdown curve shapes overlap within the statistical confidence interval of the data. This is further evidence that the red subpopulation is associated with a dissociating cluster which yields $[AY + H]^+$ in the same geometric arrangement as the bare ion feature (i.e., the black subpopulation).

5.3.3 The Experimental UVPD Spectra of the DMS Resolved Ions

To characterize the resolved populations under two modifiers, UVPD experiments are conducted over each resolved species. In addition, an overall UVPD with no SV and CV applied is also generated for each modifier. Figure 5-7 shows the UVPD spectra of $[AY + H]^+$ obtained from the two separate subpopulations. A summary of the experimental parameters employed to record these spectra can be found in Appendix E, entry (c), and details of generating the spectra are outlined in example 2, section 2.1.4. Note that the area of each mass peak is found by integrating over a ± 0.5 m/z interval centered on the nominal mass, and the solid traces are obtained via 5-point Savitzky–Golay smooth of the experimental data. Similar to in the CID experiments, the m/z 104 fragment is the most abundant across the optical range, followed by the m/z 105 fragment. The m/z 77, 122, 147, and 207 products are also generated by photo-fragmentation, but their abundance is much lower compared to that observed in the CID experiments. In addition to these product ions, a new fragment (m/z 79), which is assigned to the protonated benzene ($[C_6H_7]^+$), is observed via photodissociation. The resulting UVPD spectrum from this product channel is similar to that from other channels and its signal intensity is of the same magnitude with that of the m/z 77 channel.

Within the spectral range of 3.2 - 6.0 eV (208 – 380 nm), a satisfying signal-to-noise (S/N) ratio is achieved for all the separated ion populations (two ion populations at each modifier). A broad band is observed within the 3.4 - 4.5 eV spectral region for the spectra recorded via all product channels, and the signal maximum occurs around 3.6 – 3.8 eV. Fine spectral features associated with vibronic

80

structure are not resolved owing to the apparent density of states. It is also noteworthy that, the difference in the resulting spectra for the two DMS-resolved ion subpopulations are indistinguishable within experimental error. This is consistent with the CID results and again is indicative that one of the resolved species may be attributed to some ion-solvent cluster that gives analyte ions upon dissociation.
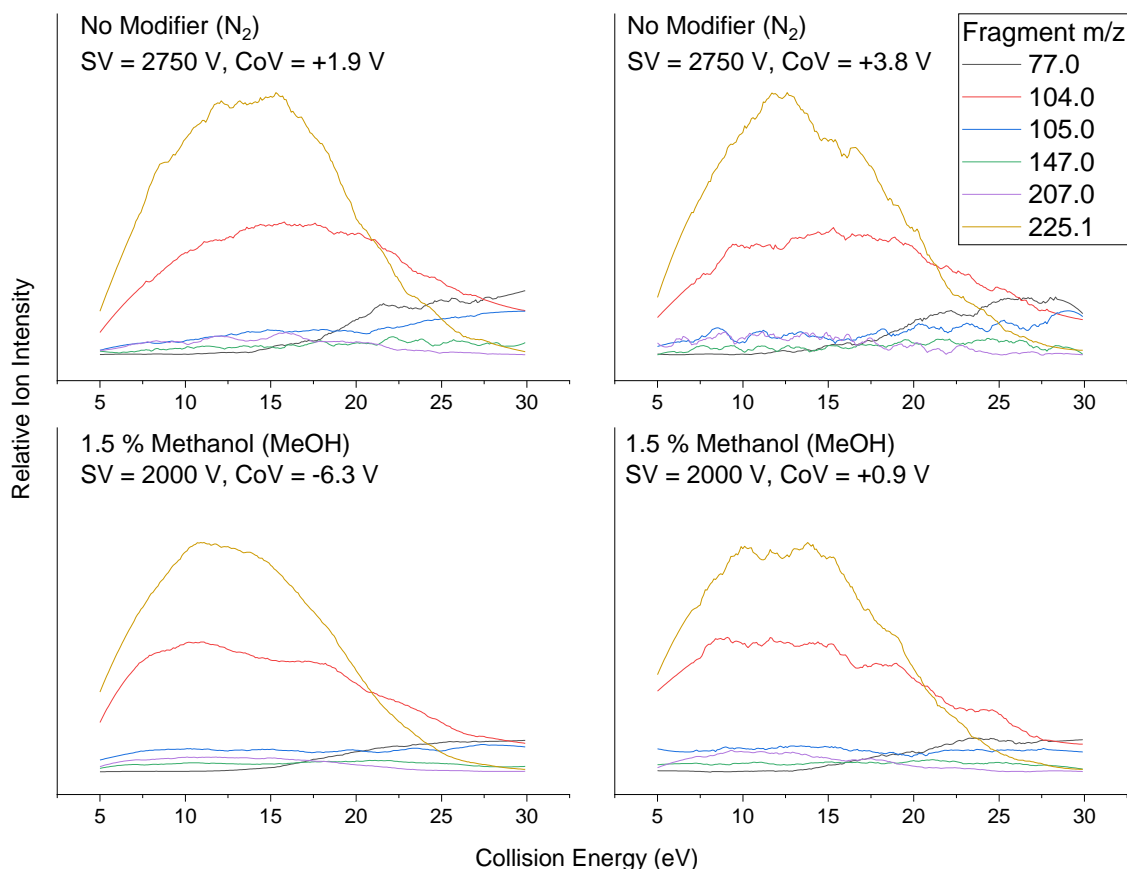
Figure 5-6. The breakdown curve of the DMS resolved ion populations

5.3.4 The [AY + H]+ Isomers and Transition States

Following the procedures outlined in section 5.2.2, the energy profile for the $[AY + H]^+$ isomers is summarized in Table 5-1 and Figure 5-8. Several low-lying isomers occur within 10 kJ/mol of the computed global minimum. Generally, the structures of most $[AY + H]^+$ isomers deviate from planarity, both in solution and in the gas phase, owing mostly to the hybridization change at the protonation site (O^1, N^3, and N^4). In methanol solution, the N_A protonation scheme is preferred for all three skeletal forms (*i.e.*, 1a, 2a, and 3a are the respective lowest energy isomer of the ZEE, ZEZ, and ZZE series), as was calculated for the acetonitrile environment. The ZEE-N_A (1a) conformation is the global

81

minimum structure, however, the ZEE-O (1c) and ZZE-N$_A$ (3a) structures are expected to have significant contribution to the overall ensemble population at room temperature (298 K) owing to their relatively low energies. In contrast, the ZEE-O (1c) conformer is the global minimum in the gas phase. However, at T = 423 K, which is the lowest temperature setting of the DMS cell, both the ZEE and ZEZ conformations are expected to contribute significantly to the overall ion population (assuming thermal equilibrium).

To model the thermal interconversion between the $[AY + H]^+$ isomers, the geometries and relative energies of the interconnecting TSs were calculated. These are summarized in Figure 5-9. In general, the computed transition states can be classified into three categories: proton hopping, dihedral scissoring, and nitrogen inversion. The proton hopping TSs are the simplest to visualize; imaginary normal modes feature a proton oscillating between two spatially adjacent protonation sites. Members in this category include TS1, TS3, and TS6, whose relative Gibbs energy ranges from around 40 kJ/mol (TS1; $O^1 - N^4$ hopping) to 270 kJ/mol (TS6; N^3-N^4 hopping). The dihedral scissoring TSs feature a direct skeletal isomerization by a large dihedral rotation along the N^3-N^4 bond. TS4 and TS7, which exhibit Gibbs energies of 20–30 kJ/mol, belong to this category. TS5 (76 kJ/mol) exhibits a scissoring motion that is coupled with a hydroxyl group rotation along O^1-C^2. The imaginary mode of TS2, TS8, and TS9 all involves an inversion on the N^4 atom, and thus the three TSs belongs to the nitrogen inversion category. The relative Gibbs energies of these TSs falls in the range of 180 - 200 kJ/mol. It is interesting to note that the dihedral rotation along the N^3-N^4 bond has a comparable energy barrier to that of O^1-N^4 proton hopping, while similar dihedral rotation along the adjacent N^4-C^5 bond has much higher energy barrier. This implies that the π conjugation along N^3-N^4 is much weaker than along N^4-C^5, an apparent double bond in the skeletal formula (see Figure 5-2).
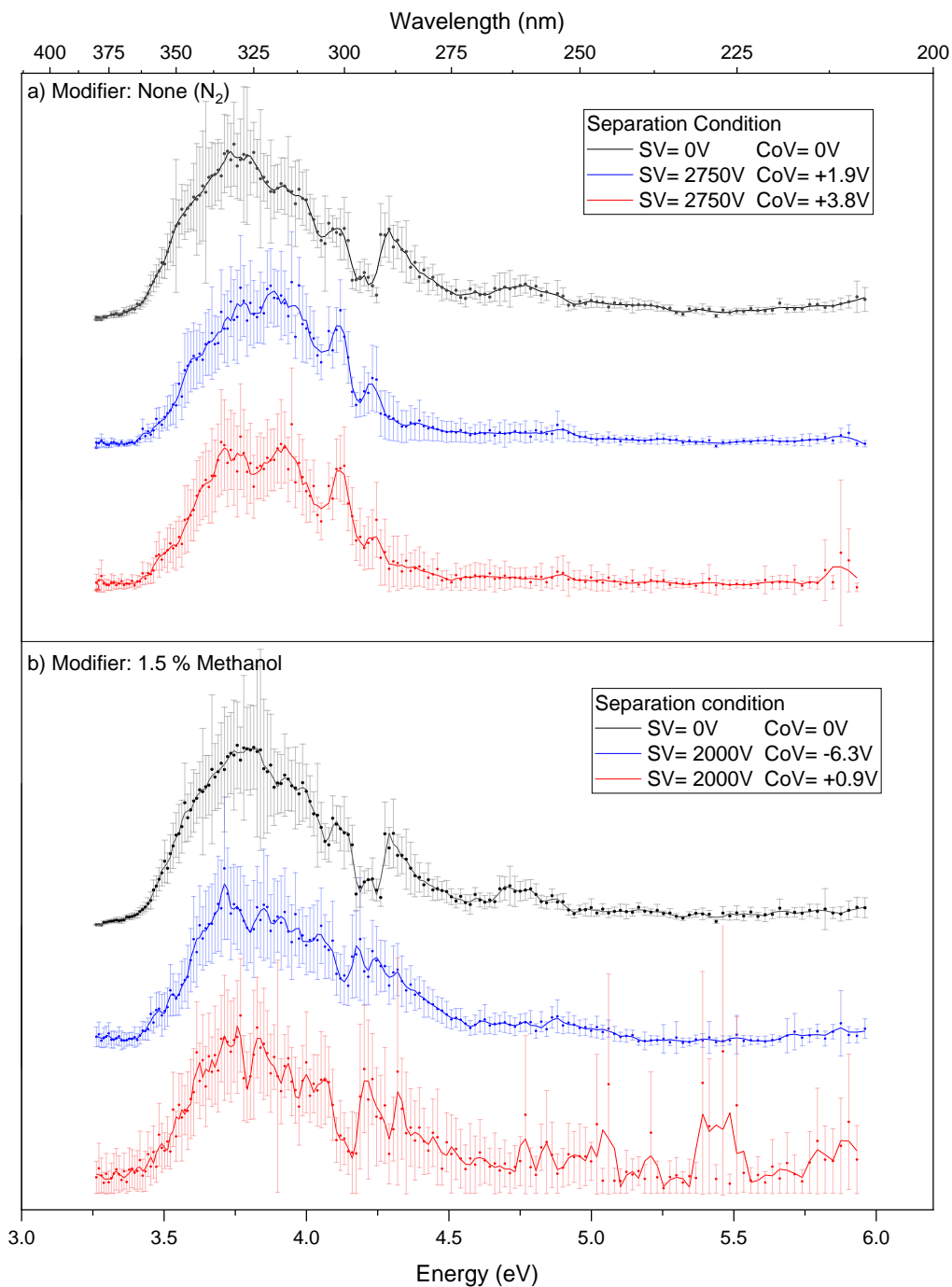
Figure 5-7. UVPD spectra of the DMS resolved [AY + H]$^+$ species under a) pure nitrogen and b) 1.5 % methanol. DMS parameter condition is outlined in the legend. Error bar is generated out of 12 replicate MS measurement and 100 replicate power measurement at each wavelength.

Table 5-1. Summary of Relative Gibbs Energies (in kJ/mol) for [AY + H]⁺ Isomers

Isomer Label	Bonding Scheme		Gas Phase				MeOH	
			wB97XD	CCSD	wB97XD3	CCSD	wB97XD	CCSD
1a	ZEE	N_A	7.75	13.94	8.98	14.33	**0.00**	**0.00**
1b		N_B	10.25	8.17	12.87	7.50	21.90	14.34
1c		O	**0.00**	**0.00**	**0.00**	**0.00**	7.78	2.18
2a	ZEZ	N_A	13.40	18.71	12.88	17.55	7.64	7.14
2b		N_B	13.41	13.24	14.21	13.71	22.08	15.92
2c		O	8.79	6.71	8.43	6.77	26.30	24.78
3a	ZZE	N_A	16.65	23.49	16.61	23.52	7.43	3.61
3b		N_B	33.95	31.45	34.89	32.21	43.01	32.73
3c		O	37.86	35.41	37.58	35.14	39.25	26.69

1a (14.33)   1b (7.50)   1c (0.00)

2a (17.55)   2b (13.71)   2c (6.77)

3a (23.52)   3b (32.21)   3c (35.14)
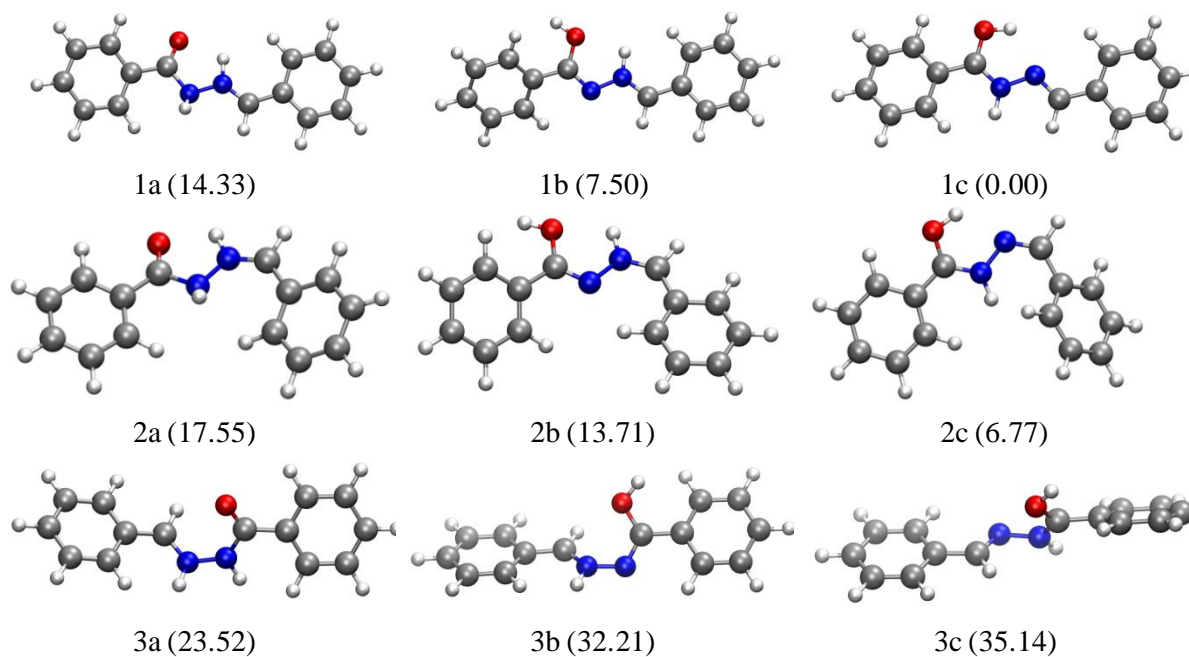
Figure 5-8. The structures and the relative Gibbs energies (in kJ/mol) of gas phase [AY + H]⁺

TS1 (2a – 2c, 43.71) TS2 (1a – 2a, 173.02) TS3 (1a – 1c, 38.40)

TS4 (1a – 3a, 24.72) TS5 (1c – 3c, 76.00) TS6 (3b – 3c, 273.32)

TS7 (1b – 3b, 33.94) TS8 (1b – 2b, 203.74) TS9 (2a – 3a, 185.67)
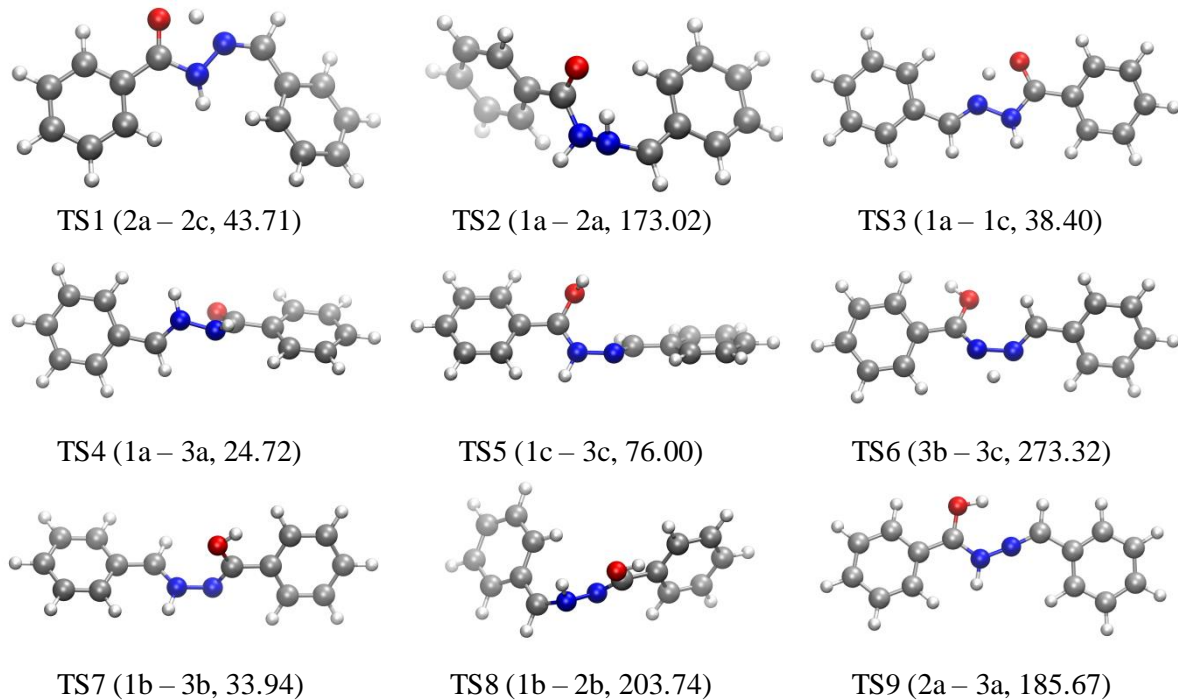
Figure 5-9. The structures and the relative Gibbs energies (in kJ/mol) of [AY + H]$^+$ TSs

Having computed the isomer and the TSs energies, a disconnectivity graph (DG) is constructed to aid in visualization of the [AY + H]$^+$ PES. The DG and the associated isomerization pathways are shown in Figures 5-10. and 5-11, respectively. As shown in Figure 5-10, the lowest energy funnel on the PES includes the 1a, 1c, 3a, and 3c isomers., with the highest energy barrier to isomerization within the funnel (to produce 3c) being 76.0 kJ/mol. Two secondary funnels are also identified. The funnel that captures the 2a and 2c isomers is isolated from the global minimum funnel by TS2 = 173 kJ/mol. The funnel that captures all three of the N_B isomers, on the other hand, is isolated by a barrier of more than 250 kJ/mol from the lowest energy basin. This situation, wherein multiple deep funnels are found on a PES, can result in kinetic trapping of high energy structures for electro-sprayed ensembles.[112]

85
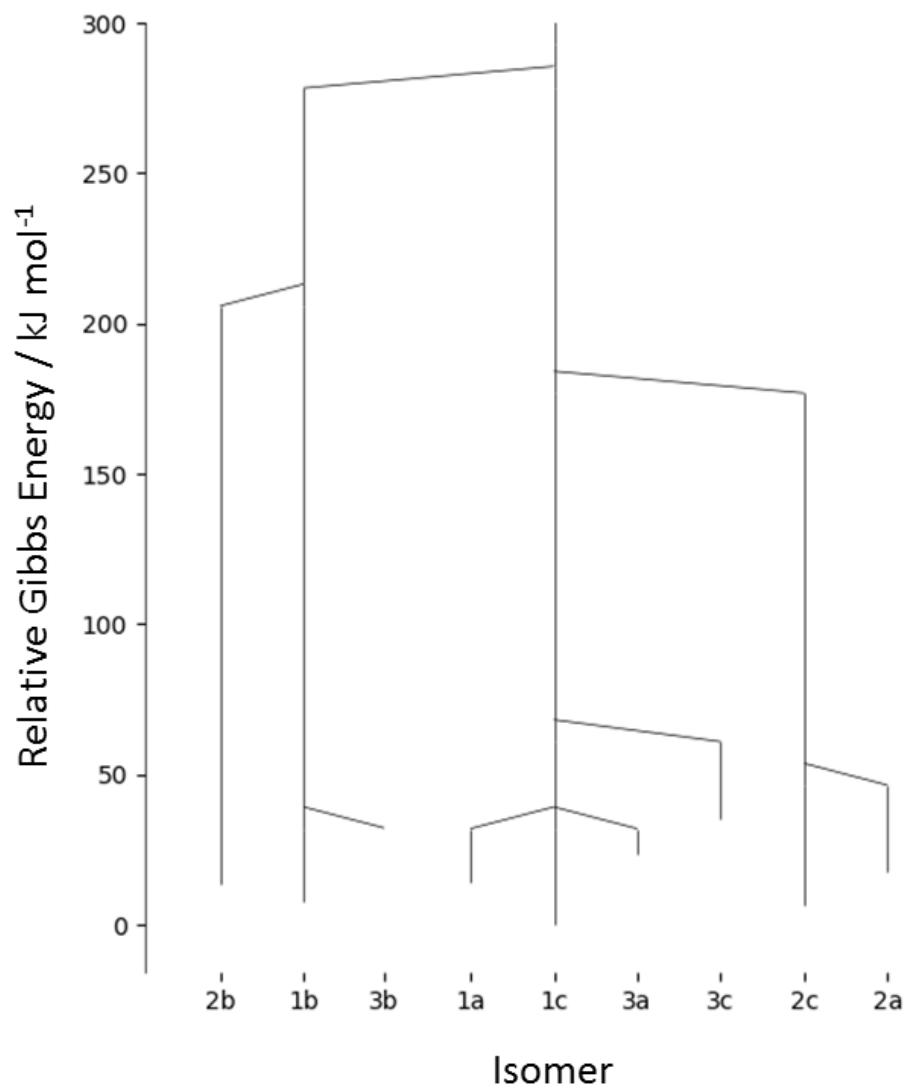
Figure 5-10. The disconnectivity graph of the [AY + H]$^+$ ions
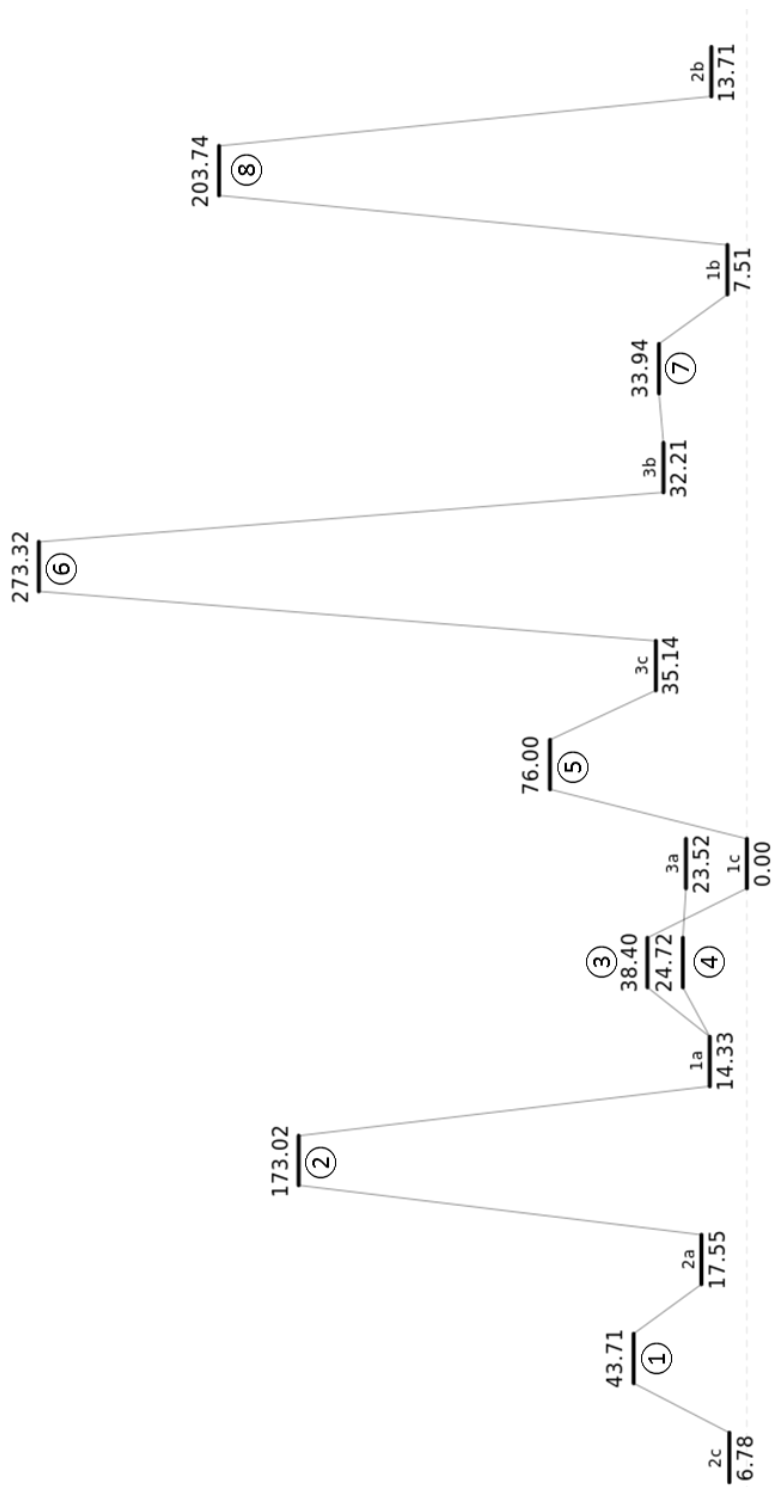
Figure 5-11. The energy profile of the reaction pathways involved in the DG. Energy is in kJ/mol. Isomers are labeled using the short notation while the TSs are labeled in circled numbers.

5.3.5 Theoretical Spectra of the [AY + H]⁺ Ions

The calculated UV-Vis absorption spectra of the low-lying $[AY + H]^+$ isomers are shown in Figure 5-12. These spectra show vertical excitation energies calculated at the (ωB97XD/STEOM-CCSD levels of theory and the spectral prediction from vertical gradient Frank-Condon simulations. For all the low-lying isomers, the primary band is associated with the $S_0 \rightarrow S_1$ transition. To reveal the type of the $S_0 \rightarrow S_1$ transition, natural difference orbital (NDO) analysis, which finds the electron density difference between the excited and the ground state, is conducted. As is shown in Figure 5-13, the shape of the NDO for the $S_0 \rightarrow S_1$ transitions of isomer 1a, 1b, 1c indicates that these transitions are $\pi \rightarrow \pi^*$ transitions as both electron enriched (blue) and the electron deficient (red) region resembles a molecular π orbital. The position of the primary band for the $[AY + H]^+$ ions is better classified by the protonation scheme than by the skeletal conformation. For the N_A and O protomers, the $S_0 \rightarrow S_1$ transition usually lies within the 300-320 nm (3.8-4.1 eV) range, while for the N_B protomers, the transition is red-shifted to around 350 nm (3.5 eV).

Of the protonated and neutral isomers being investigated, the DFT (ωB97XD/ωB97XD3) and the CCSD (DLPNO-STEOM-CCSD) methods predict the same global minimum. The relative electronic energies of other higher energy isomers calculated by the two methods usually differ by less than 7 kJ/mol. Particularly for the protonated species, the ωB97XD and ωB97XD3 functionals generally give negligible differences in the resulting equilibrium geometry and relative energy (< 1 kJ/mol). In preliminary calculations, similar geometries and relative energies are also obtained for the four test isomers (1a, 1c, 2a, and 2c) using the B3LYP/6-311++G(d,p) level of theory with GD3 empirical dispersion. For gas phase UV-Vis spectra, the DFT method tends overestimate the band position by around 0.25 eV when compared to results from the STEOM-DLPNO-CCSD method. For the spectra of the neutral species, the prediction from two methods generally differs by less than 0.1 eV except in the case of 1b-3b.
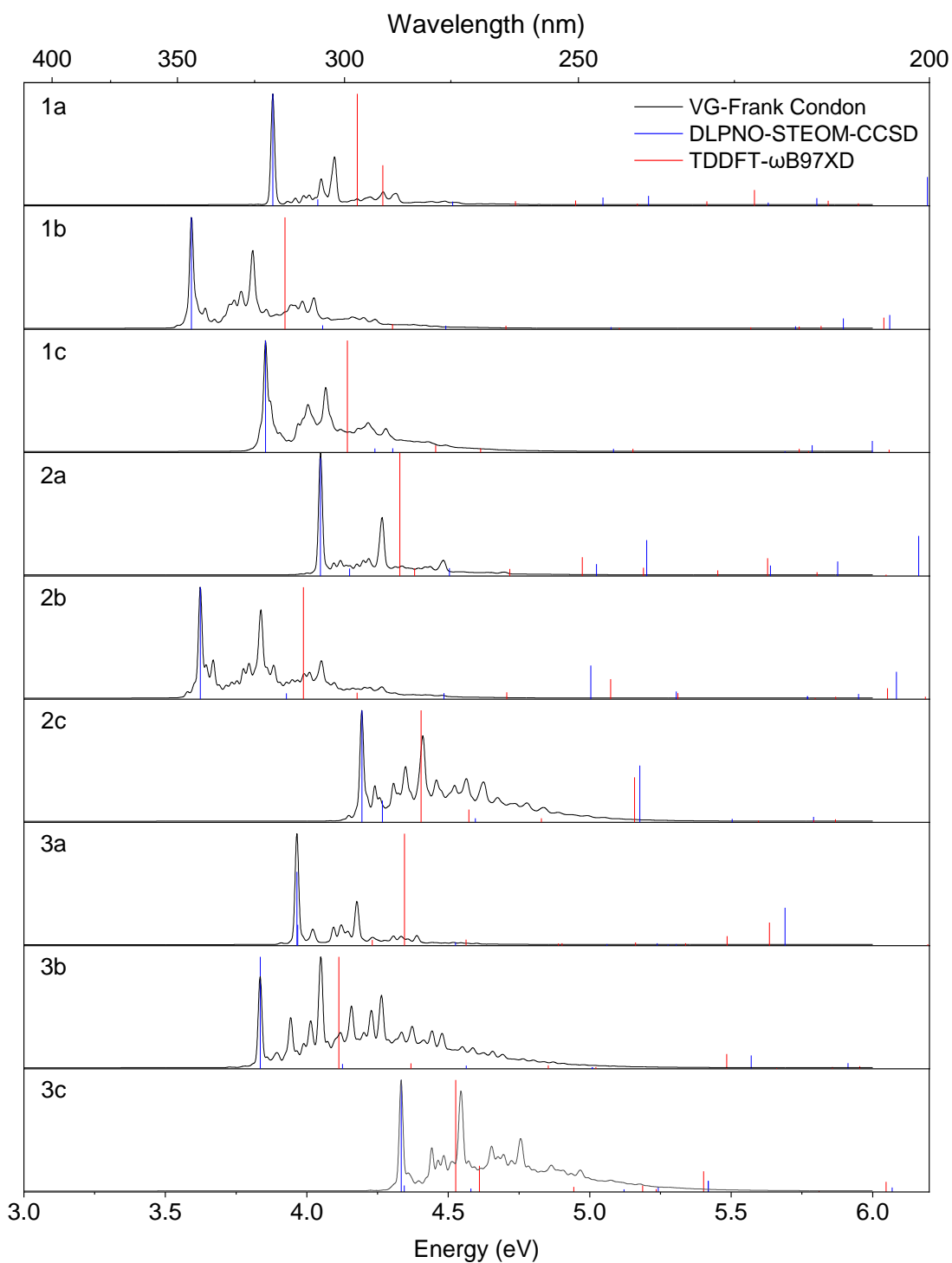
Figure 5-12. The vertical and Frank-Condon excitation spectra of the [AY + H]⁺ isomers
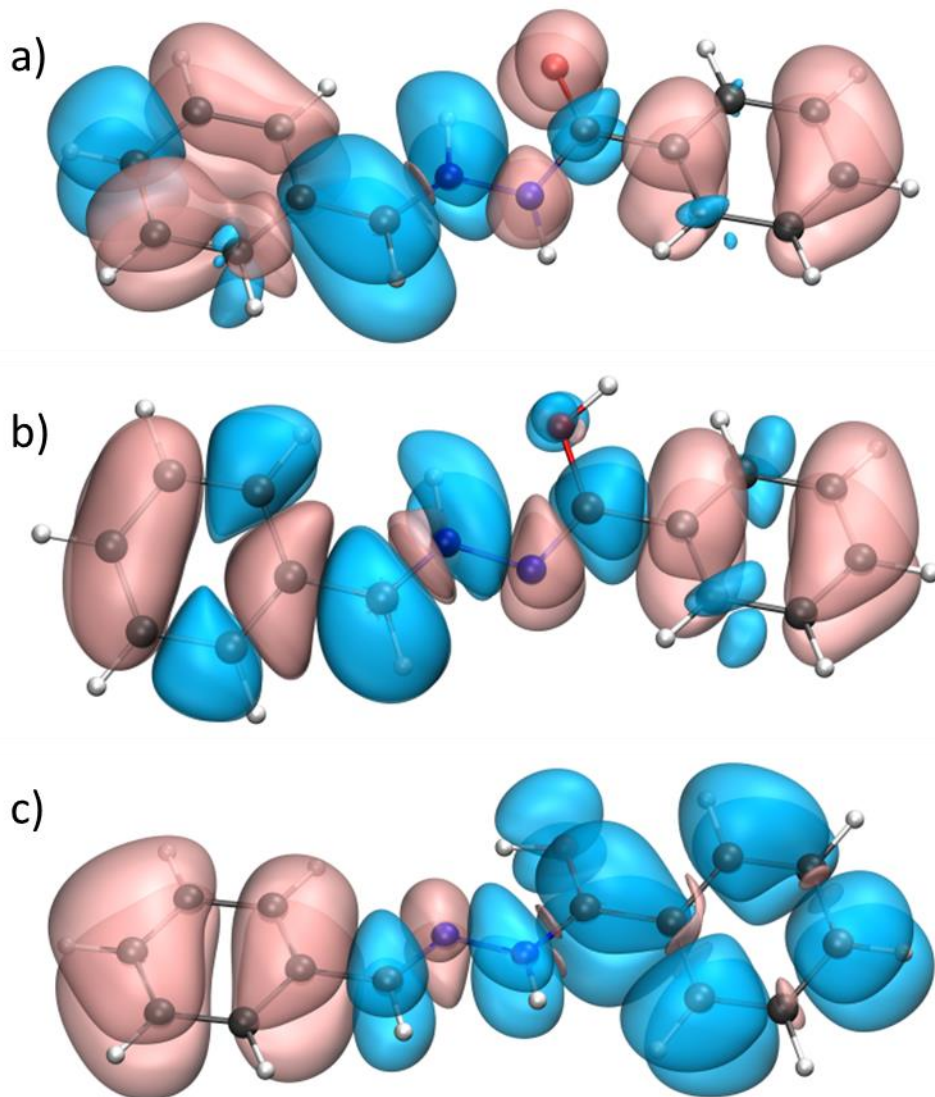
Figure 5-13. The NDO analysis of the $S_0 \rightarrow S_1$ transition of the [AY + H]$^+$ isomer a)1a, b)1b, and c)1c. Blue and red corresponds to charge enriched and depleted regions, respectively, after the excitation.

5.3.6 Rationalization of the Experimental Observations of [AY + H]$^+$.

Based on the similarities in both the CID profile and the UVPD spectra, the DMS resolved population with pure nitrogen and methanol modifier (as is shown in Figure 5-4a and 5-4c) are attributed to the [AY + H]$^+$ bare ions and its ion-solvent clusters. The bare ion is assigned to the most abundant ion population (black) while the ion-solvent cluster is assigned to the other (red). One important assumption

for the assignment is that the composition of the bare ion population is near identical to that from decomposed ion solvent clusters. The disconnectivity graph in Figure 5-11 clearly shows that four kinetically trapped species may coexist in the ion population, separated by interconversion energy barrier of at least 150 kJ/mol. These four species are not being resolved by DMS, which may be due to the similarity in their collision cross section and in their binding interaction with modifiers.

With the above rationalization of the ion population, the experimental UVPD spectra is a convolution of the individual UV spectra of each $[AY + H]^+$ isomers. As the UVPD spectra obtained from all separation conditions are almost identical, the representative experimental spectrum is chosen the one measured with pure nitrogen carrier gas and 0 SV (Figure 5-7a, black). Figure 5-14 shows a comparison between the chosen experimental spectrum and the theoretical electronic absorption spectra of the four local minima (1b, 2b, 1c, and 2c). The experimental band maximum at 3.7 eV can be attributed to the contribution from isomer 1c with a 0.1 eV deviation. The response at around 3.5 eV can be accounted by 1b and 2b, while the minor features at 4.3-4.4 eV can be attributed to the contribution of 2c, whose primary band locates at 4.2 eV. However, the overall band structure of the experimental spectrum cannot be reproduced from the Boltzmann weighted electronic absorption spectrum. There are two possible explanation: one is that the Boltzmann population from theoretical calculation can be potentially have significant error, as is discussed in section 2.2.8; the other is that the ion population in the solution phase or in the charged droplet during ESI is retained in the gas phase due to the kinetic trapping effect.

To further verify the solvent cluster assignment, one eventually need a theoretical investigation of the ion-solvent clusters of $[AY + H]^+$ isomers and the binding energies. The discrepancy in the peak position may be fixed by employing larger basis set for the excited state calculation. To better simulate the band shape of individual excitations, and also to eventually develop a general model to predict the efficiency of a photoswitch, one may also need to go beyond the Born-Oppenheimer approximation and take account the potential energy surface crossing. On the experimental side, a brute-force way to improve the quality of the experimental spectra is to increase the number of repetitions when measuring the response at each wavelength. One could further design optimized data acquisition procedure that dynamically adjust the balance between signal quality and data acquisition time. Long term effort should focus on the separation of the means to separate $[AY + H]^+$ isomers so that the UVPD spectrum is deconvoluted. Moreover, other metallic-AY complexes are also good subjects for further studies.
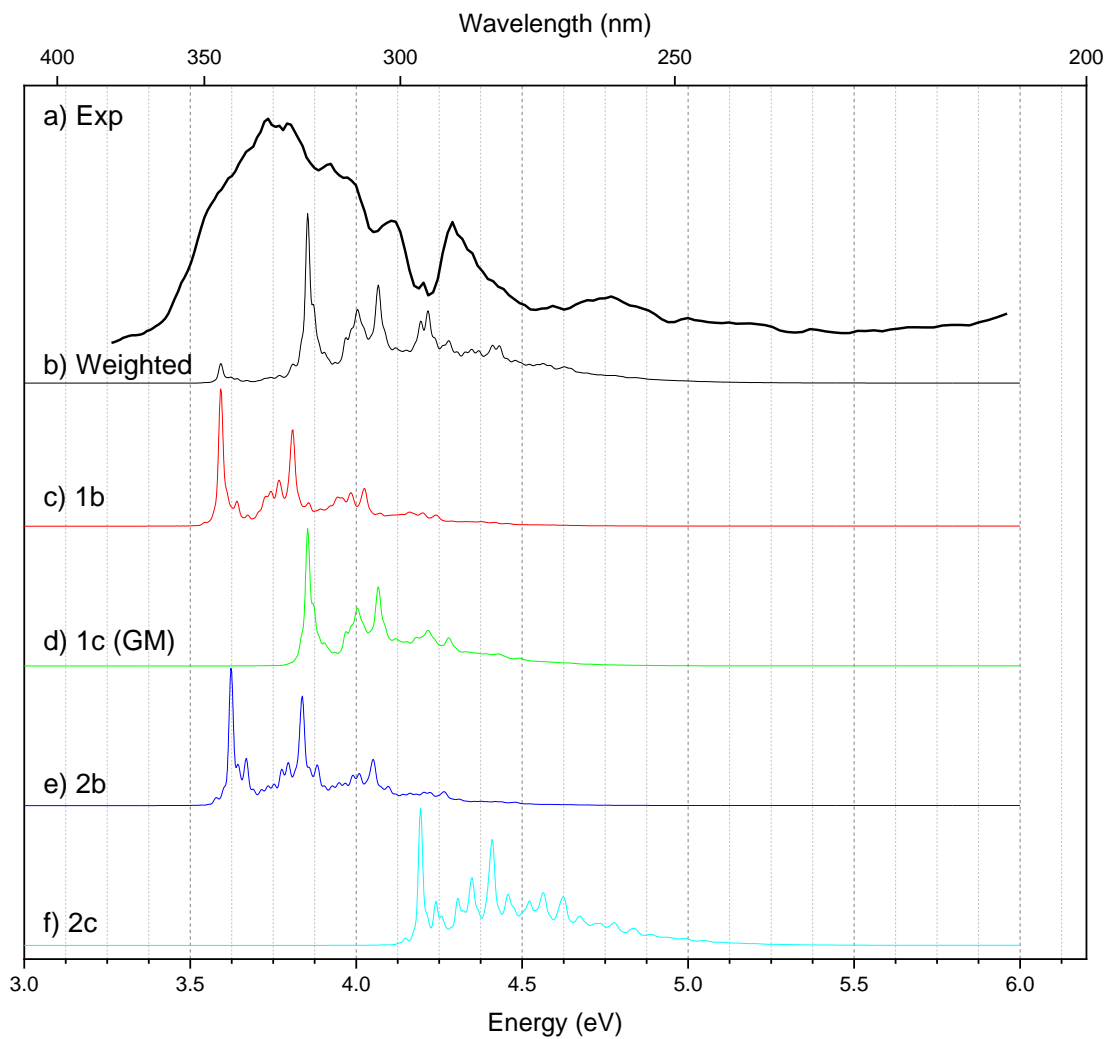
Figure 5-14. Comparison of Experimental (a) and Theoretical (b-f) UV-Vis Spectra of the [AY + H]+ ions. Experimental spectrum is taken from Figure 5-7a (No modifier, SV = 0 V). Spectrum b is the Boltzmann weighted spectrum of all 9 [AY + H]$^+$ isomers at 298 K. Spectra c-f are the normalized $S_0 \rightarrow S_1$ electronic absorption spectrum of isomer 1b, 1c, 2b, and 2c.

5.4 Chapter 5 Summary

In this work, the $[AY + H]^+$ gaseous ions are produced from electro-spraying of acidified analyte solution in methanol. Using the DMS technique, resolution of the gas phase $[AY + H]^+$ ions has been observed with no modifier, methanol modifier, and water modifier. For the ion populations resolved under no modifier and methanol modifier, their CID profile are recorded in the CE range of $5 - 30$ eV, while their UVPD spectra are recorded within the spectral range of 3.5-6.0 eV (208-380nm). Both the CID profile and the UVPD spectra shows indistinguishable difference, which suggest the resolved species are bare analyte ions and its ion-solvent cluster of some kind. Theoretically, isomers of $[AY + H]^+$ in gas phase and methanol are investigated using DFT calculations with CCSD electronic energy corrections. The transition states connecting the $[AY + H]^+$ isomers in the gas phase are also determined and the resulting disconnectivity graph of the $[AY + H]^+$ ions suggest the presence of four kinetically trapped species. The vertical excitation spectrum of the $[AY + H]^+$ isomers are calculated using both the DFT and the STEOM- DLPNO-CCSD method. The vibronic spectra of the isomers are simulated using a combination of DFT Hessian with STEOM- DLPNO-CCSD electronic excitation energy. For the global minimum, the position of its primary electronic excitation from STEOM-DLPNO-CCSD calculation is only 0.1 eV off from the experimental peak maximum, while the deviation from DFT calculation is 0.3 eV. The band shape of the experimental UVPD spectra does not resemble that of the UV-Vis spectrum for individual isomers but can be potentially simulated with a convolution of 4 UV-Vis spectra from the kinetically trapped species. In addition, details of the neutral AY species is available in Appendix D.

Chapter 6
Conclusion

This thesis advances the experimental and theoretical tools that are available for studying gas phase ions. The effectiveness of utilizing structure similarity functions is demonstrated by mapping the PES in the case of protonated serine dimer clusters. The accompanying means of structure interpolation is demonstrated capable of determining TS structures in the investigation of an alkylation reaction between an alkyl-tricarbastannatrane complex and a barbituric acid derivative. Moreover, the DMS-UVPD technique is introduced and is demonstrated to be an effective technique for characterizing gas phase ions. For the three studies described herein, the application of complementary experimental and computational tools is shown to be critical. As these tools continue to improve, more precise and detailed studies of larger molecular systems will become possible.

In chapter 3, the capability of the structure similarity function to map a PES is examined using the model system of $[\text{Ser}_2 + \text{H}]^+$ gas phase clusters. A total of 40 low energy conformers within 55 kJ/mol of the computed global minimum (at T = 298 K) are collected from literature and from BH simulation. Further geometric classification with the aid of DM distance reveals four primary bonding motifs between the two serine units in $[\text{Ser}_2 + \text{H}]^+$, each having a representative isomer. A comparison between the scaled harmonic spectra of the four representative isomers and the experimental IRMPD spectrum shows that isomer 2, rather than the computed GM, yields the best match. Further high-level calculation reveals that GM and Isomer 2 are close in Gibbs energy (*ca.* 0.1 kJ/mol difference). Considering that the GM and Isomer 2 belong to different regions of the associated PES, it is apparent that $[\text{Ser}_2 + \text{H}]^+$ may be subject to local kinetic trapping during formation. To further study this kinetic trapping effect, a thorough search of TSs interconnecting the isomers is necessary.

In chapter 4, a regioselective alkylation reaction between the barbituric acid and the 3-methylbut-2-en-1-yl triptych complex is investigated. To generate the TS structure, a structure interpolation technique based on structure similarity, is employed. A total of three reaction pathways are located. The activation energy barriers of these three reaction pathways rationalizes the experimental fact that the T3 product is not formed despite it being thermodynamically favored. The geometric parameters of the 3-methylbut-2-en-1-yl moiety for the TSs reveals that the lower energy of the 8-member-ring reaction channel is related to the steric freedom of puckering at the substitution sites (C_β and C_δ). AIM

94

and NCI analyses enable visualization the bonding path network and the weak interactions within the TSs. Based on the results of the AIM analysis, there is no evidence of a hyper-valent tin center within any TS found in this work. The electron density profile of the N-Sn, Sn-C_β, and $C_{\beta/\delta}$-C_α BCPs along the IRC coordinate indicates that dissociation from the triptych moiety along the 8-member-ring reaction channel is disfavoured compared to the 6-member-ring reaction channel, given a similar $C_{\beta/\delta}$-C_α bond formation progress. To further investigate the regioselectivity of this alkylation reaction, a good starting point would be explorations of the substitution effect on C_β and C_δ. In chapter 5, a joint study combining the DMS-UVPD experiments with theoretical calculations is conducted on the (E)-N'-benzylidenebenzohydrazide, an acylhydrazone photoswitch abbreviated as AY. Experimentally, the gaseous $[AY + H]^+$ ions are resolved with DMS under both pure nitrogen and methanol (1.5%) modified carrier gas environment. CID and UVPD measurements are conducted on each of the separated ion species, but neither the break down curve nor the UV spectra show distinctive difference. Thus, the experimental data suggest the resolved ion populations are attributed to the bare $[AY + H]^+$ ions and its ion-solvent cluster. In the theoretical counterpart, 9 $[AY + H]^+$ isomers are examined under ωB97XD/ωB97XD3 method with electronic energy corrections from DLPNO-CCSD calculations. Further more, the disconnectivity graph of the $[AY + H]^+$ PES suggests four candidate isomers that may experience local kinetic trapping. The vibronic spectra for the $[AY + H]^+$ isomers are simulated using the DFT Hessian and the vertical energies from STEOM-DLPNO-CCSD calculation. A comparison between the experimental UVPD and the vibronic spectra of the four isomers suggest that the UVPD may be a convolution of the four spectra.

In summary, with the work presented in this thesis, the potential of the novel structure similarity methods and the DMS-UVPD method is demonstrated. These original techniques developed in this work could potentially enhance the arsenal of gas phase cluster research and enable dealing with more geometrically and electronically convoluted cases.

Bibliography

(1)   Iskra, A.; Gentleman, A. S.; Kartouzian, A.; Kent, M. J.; Sharp, A. P.; Mackenzie, S. R. Infrared Spectroscopy of Gas-Phase $M^+(CO_2)_n$ (M = Co, Rh, Ir) Ion–Molecule Complexes. *The Journal of Physical Chemistry A* **2017**, *121* (1), 133–140. https://doi.org/10.1021/acs.jpca.6b10902.

(2)   Cooks, R. G.; Zhang, D.; Koch, K. J.; Gozzo, F. C.; Eberlin, M. N. Chiroselective Self-Directed Octamerization of Serine: Implications for Homochirogenesis. *Analytical Chemistry* **2001**, *73* (15), 3646–3655. https://doi.org/10.1021/ac010284l.

(3)   Singh, P.; Wangoo, N.; Sharma, R. K. Phenylalanine Dimer Assembly Structure as the Basic Building Block of an Amyloid like Photoluminescent Nanofibril Network. *Soft Matter* **2020**, *16* (17), 4105–4109. https://doi.org/10.1039/d0sm00387e.

(4)   Harris, D. C. *Quantitative Chemical Analysis*; W. H. Freeman, 2010.

(5)   *Applied Electrospray Mass Spectrometry*, 1st ed.; Pramanik, B. N., Ganguly, A. K., Gross, M. L., Eds.; CRC Press: 270 Madison Avenue, New York, 2002. https://doi.org/10.1201/9780203909270.

(6)   Cumeras, R.; Figueras, E.; Davis, C. E.; Baumbach, J. I.; Gràcia, I. Review on Ion Mobility Spectrometry. Part 1: Current Instrumentation. *Analyst* **2015**, *140* (5), 1376–1390. https://doi.org/10.1039/c4an01100g.

(7)   Oh, H. bin; Lin, C.; Hwang, H. Y.; Zhai, H.; Breuker, K.; Zabrouskov, V.; Carpenter, B. K.; McLafferty, F. W. Infrared Photodissociation Spectroscopy of Electrosprayed Ions in a Fourier Transform Mass Spectrometer. *Journal of the American Chemical Society* **2005**, *127* (11), 4076–4083. https://doi.org/10.1021/ja040136n.

(8)   Seo, J.; Hoffmann, W.; Malerz, S.; Warnke, S.; Bowers, M. T.; Pagel, K.; von Helden, G. Side-Chain Effects on the Structures of Protonated Amino Acid Dimers: A Gas-Phase Infrared Spectroscopy Study. *International Journal of Mass Spectrometry* **2018**, *429*, 115–120. https://doi.org/10.1016/j.ijms.2017.06.011.

(9)   Sunahori, F. X.; Yang, G.; Kitova, E. N.; Klassen, J. S.; Xu, Y. Chirality Recognition of the Protonated Serine Dimer and Octamer by Infrared Multiphoton Dissociation Spectroscopy. *Phys. Chem. Chem. Phys.* **2013**, *15* (6), 1873–1886. https://doi.org/10.1039/C2CP43296J.

(10)   Kong, X.; Tsai, I. A.; Sabu, S.; Han, C. C.; Lee, Y. T.; Chang, H. C.; Tu, S. Y.; Kung, A. H.; Wu, C. C. Progressive Stabilization of Zwitterionic Structures in [H(Ser) 2-8]+ Studied by Infrared Photodissociation Spectroscopy. *Angewandte Chemie - International Edition* **2006**, *45* (25), 4130–4134. https://doi.org/10.1002/anie.200600597.

(11)   Baierl, R.; Fillion, E. 1,4-Allylstannylation of Electron-Deficient Olefins – Mechanistic Studies. University of Waterloo 2016, p 26.

(12)   Kavoosi, A. Reactivity of Tricarbastannatranes, 2016.

(13)   van Dijken, D. J.; Kovaříček, P.; Ihrig, S. P.; Hecht, S. Acylhydrazones as Widely Tunable Photoswitches. *Journal of the American Chemical Society* **2015**, *137* (47), 14982–14991. https://doi.org/10.1021/jacs.5b09519.

(14)   Gupta, P.; Panda, T.; Allu, S.; Borah, S.; Baishya, A.; Gunnam, A.; Nangia, A.; Naumov, P.; Nath, N. K. Crystalline Acylhydrazone Photoswitches with Multiple Mechanical Responses. *Crystal Growth and Design* **2019**, *19* (5), 3039–3044. https://doi.org/10.1021/acs.cgd.8b01860.

(15)   Broichhagen, J.; Trauner, D. The in Vivo Chemistry of Photoswitched Tethered Ligands. *Current Opinion in Chemical Biology* **2014**, *21*, 121–127. https://doi.org/10.1016/j.cbpa.2014.07.008.

(16)   Khan, A.; Kaiser, C.; Hecht, S. Prototype of a Photoswitchable Foldamer. *Angewandte Chemie - International Edition* **2006**, *45* (12), 1878–1881. https://doi.org/10.1002/anie.200503849.

(17)   Kiefer, W.; Bernstein, H. J. Vibrational-Rotational Structure in the Resonance Raman Effect of Iodine Vapor. *Journal of Molecular Spectroscopy* **1972**, *43* (3), 366–381. https://doi.org/10.1016/0022-2852(72)90048-3.

(18)   Chen, Y. C.; Urban, P. L. Time-Resolved Mass Spectrometry. *TrAC - Trends in Analytical Chemistry* **2013**, *44*, 106–120. https://doi.org/10.1016/j.trac.2012.11.010.

(19)   Rayleigh, Lord. XX. On the Equilibrium of Liquid Conducting Masses Charged with Electricity. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1882**, *14* (87), 184–186. https://doi.org/10.1080/14786448208628425.

(20)   Iribarne, J. v.; Thomson, B. A. On the Evaporation of Small Ions from Charged Droplets. *The Journal of Chemical Physics* **1976**, *64* (6), 2287–2294. https://doi.org/10.1063/1.432536.

(21)   Konermann, L. A Simple Model for the Disintegration of Highly Charged Solvent Droplets during Electrospray Ionization. *Journal of the American Society for Mass Spectrometry* **2009**, *20* (3), 496–506. https://doi.org/10.1016/j.jasms.2008.11.007.

(22)   Seo, J.; Warnke, S.; Pagel, K.; Bowers, M. T.; von Helden, G. Infrared Spectrum and Structure of the Homochiral Serine Octamer-Dichloride Complex. *Nature Chemistry* **2017**, *9* (12), 1263–1268. https://doi.org/10.1038/nchem.2821.

(23)   Domingues, P.; Domingues, M. R. M.; Amado, F. M. L.; Ferrer-Correia, A. J. Characterization of Sodiated Glycerol Phosphatidylcholine Phospholipids by Mass Spectrometry. *Rapid Communications in Mass Spectrometry* **2001**, *15* (10), 799–804. https://doi.org/10.1002/rcm.300.

(24)   Carroll, D. I.; Dzidic, I.; Stillwell, R. N.; Horning, M. G.; Horning, E. C. Subpicogram Detection System for Gas Phase Analysis Based upon Atmospheric Pressure Ionization (API) Mass Spectrometry. *Analytical Chemistry* **1974**, *46* (6), 706–710. https://doi.org/10.1021/ac60342a009.

(25)   Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Analytical Chemistry* **1991**, *63* (24), 1193A-1203A. https://doi.org/10.1021/ac00024a716.

(26)   Barker, J.; Ando, D. J. *Mass Spectrometry: Analytical Chemistry by Open Learning*; Analytical Chemistry by Open Learning (Cloth); Wiley, 1999.

(27)   Montaser, A.; Golightly, D. W. *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*; VCH Publishers, 1992.

(28)   Marshall, A. G.; Hendrickson, C. L. Fourier Transform Ion Cyclotron Resonance Detection: Principles and Experimental Configurations. *International Journal of Mass Spectrometry* **2002**, *215* (1–3), 59–75. https://doi.org/10.1016/S1387-3806(01)00588-7.

(29)   Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods in Enzymology* **2005**, *402* (1993), 148–185. https://doi.org/10.1016/S0076-6879(05)02005-7.

(30)   Mabud, M. A.; Dekrey, M. J.; Graham Cooks, R. Surface-Induced Dissociation of Molecular Ions. *International Journal of Mass Spectrometry and Ion Processes* **1985**, *67* (3), 285–294. https://doi.org/10.1016/0168-1176(85)83024-X.

(31)   Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society* **1998**, *120* (13), 3265–3266. https://doi.org/10.1021/ja973478k.

(32)   Hart-Smith, G. A Review of Electron-Capture and Electron-Transfer Dissociation Tandem Mass Spectrometry in Polymer Chemistry. *Analytica Chimica Acta* **2014**, *808*, 44–55. https://doi.org/10.1016/j.aca.2013.09.033.

(33)   Polfer, N. C. Infrared Multiple Photon Dissociation Spectroscopy of Trapped Ions. *Chemical Society Reviews* **2011**, *40* (5), 2211–2221. https://doi.org/10.1039/c0cs00171f.

(34)   Wilkinson, I.; Garcia, I. A.; Whitaker, B. J.; Hamard, J. B.; Blanchet, V. The Photodissociation of NO2 by Visible and Ultraviolet Light. *Physical Chemistry Chemical Physics* **2010**, *12* (48), 15766–15779. https://doi.org/10.1039/c0cp01551b.

(35)   Karas, M.; Bachmann, D.; Hillenkamp, F. Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. *Analytical Chemistry* **1985**, *57* (14), 2935–2939. https://doi.org/10.1021/ac00291a042.

(36)   Dunbar, R. C. Bird (Blackbody Infrared Radiative Dissociation): Evolution, Principles, and Applications. *Mass Spectrometry Reviews* **2004**, *23* (2), 127–158. https://doi.org/10.1002/mas.10074.

(37)   Krylov, E. v.; Nazarov, E. G.; Miller, R. A. Differential Mobility Spectrometer: Model of Operation. *International Journal of Mass Spectrometry* **2007**, *266* (1–3), 76–85. https://doi.org/10.1016/j.ijms.2007.07.003.

(38)   Atkins, P. W.; de Paula, J. *Atkins' Physical Chemistry*; Macmillan Higher Education, 2006.

(39)   Graydon, I.; Beatty, E.; Paul, S.; Us, M. N.; Hauck, J. A. Atmospheric Pressure Charged Particle Discriminator for Mass Spectrometry. US7098452, 2006.

(40)   Schneider, B. B.; Nazarov, E. G.; Londry, F.; Vouros, P.; Covey, T. R. Differential Mobility Spectrometry/Mass Spectrometry History, Theory, Design Optimization, Simulations, and Applications. *Mass Spectrometry Reviews* **2016**, *35* (6), 687–737. https://doi.org/10.1002/mas.21453.

(41)   Krylov, E. V.; Nazarov, E. G. Electric Field Dependence of the Ion Mobility. *International Journal of Mass Spectrometry* **2009**, *285* (3), 149–156. https://doi.org/10.1016/j.ijms.2009.05.009.

(42)   Crouse, J.; Haack, A.; Benter, T.; Hopkins, W. S. Understanding Nontraditional Differential Mobility Behavior: A Case Study of the Tricarbastannatrane Cation,

N(CH2CH2CH2)3Sn. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (4), 796–802. https://doi.org/10.1021/jasms.9b00042.

(43)   Zhou, C.; Ieritano, C.; Hopkins, W. S. Augmenting Basin-Hopping With Techniques From Unsupervised Machine Learning: Applications in Spectroscopy and Ion Mobility. *Frontiers in Chemistry* **2019**, *7* (August). https://doi.org/10.3389/fchem.2019.00519.

(44)   Haack, A.; Crouse, J.; Schlüter, F. J.; Benter, T.; Hopkins, W. S. A First Principle Model of Differential Ion Mobility: The Effect of Ion-Solvent Clustering. *Journal of the American Society for Mass Spectrometry* **2019**, *30* (12), 2711–2725. https://doi.org/10.1007/s13361-019-02340-1.

(45)   Campbell, J. L.; le Blanc, J. C. Y.; Schneider, B. B. Probing Electrospray Ionization Dynamics Using Differential Mobility Spectrometry: The Curious Case of 4-Aminobenzoic Acid. *Analytical Chemistry* **2012**, *84* (18), 7857–7864. https://doi.org/10.1021/ac301529w.

(46)   Xia, H.; Attygalle, A. B. Untrapping Kinetically Trapped Ions: The Role of Water Vapor and Ion-Source Activation Conditions on the Gas-Phase Protomer Ratio of Benzocaine Revealed by Ion-Mobility Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2017**, *28* (12), 2580–2587. https://doi.org/10.1007/s13361-017-1806-9.

(47)   Walker, S. W. C.; Mark, A.; Verbuyst, B.; Bogdanov, B.; Campbell, J. L.; Hopkins, W. S. Characterizing the Tautomers of Protonated Aniline Using Differential Mobility Spectrometry and Mass Spectrometry. *Journal of Physical Chemistry A* **2018**, *122* (15), 3858–3865. https://doi.org/10.1021/acs.jpca.7b10872.

(48)   Liu, C.; le Blanc, J. C. Y.; Shields, J.; Janiszewski, J. S.; Ieritano, C.; Ye, G. F.; Hawes, G. F.; Hopkins, W. S.; Campbell, J. L. Using Differential Mobility Spectrometry to Measure Ion Solvation: An Examination of the Roles of Solvents and Ionic Structures in Separating Quinoline-Based Drugs. *Analyst* **2015**, *140* (20), 6897–6903. https://doi.org/10.1039/c5an00842e.

(49)   Liu, C.; Yves Le Blanc, J. C.; Schneider, B. B.; Shields, J.; Federico, J. J.; Zhang, H.; Stroh, J. G.; Kauffman, G. W.; Kung, D. W.; Shapiro, M.; Ieritano, C.; Shepherdson, E.; Verbuyst, M.; Melo, L.; Hasan, M.; Naser, D.; Janiszewski, J. S.; Hopkins, W. S.; Campbell, J. L. Assessing Physicochemical Properties of Drug Molecules via Microsolvation Measurements with Differential Mobility Spectrometry. *ACS Central Science* **2017**, *3* (2), 101–109. https://doi.org/10.1021/acscentsci.6b00297.

(50)   Coughlan, N. J. A.; Liu, C.; Lecours, M. J.; Campbell, J. L.; Hopkins, W. S. Preferential Ion Microsolvation in Mixed-Modifier Environments Observed Using Differential Mobility Spectrometry. *Journal of the American Society for Mass Spectrometry* **2019**, *30* (11), 2222–2227. https://doi.org/10.1007/s13361-019-02332-1.

(51)   Walker, S. W. C.; Anwar, A.; Psutka, J. M.; Crouse, J.; Liu, C.; le Blanc, J. C. Y.; Montgomery, J.; Goetz, G. H.; Janiszewski, J. S.; Campbell, J. L.; Hopkins, W. S. Determining Molecular Properties with Differential Mobility Spectrometry and Machine Learning. *Nature Communications* **2018**, *9* (1), 1–7. https://doi.org/10.1038/s41467-018-07616-w.

(52)   Petrucci, R. H.; Herring, F. G.; Madura, J. D.; Bissonnette, C. *General Chemistry Principles and Modern Applications*, 10th editi.; Pearson Canada Inc: Toronto, Ontario, 2011.

(53)   Uzer, T.; Miller, W. H. Theories of Intramolecular Vibrational Energy Transfer. *Physics Reports* **1991**, *199* (2), 73–146. https://doi.org/10.1016/0370-1573(91)90140-H.

(54)   Dang, A.; Korn, J. A.; Gladden, J.; Mozzone, B.; Tureček, F. UV–Vis Photodissociation Action Spectroscopy on Thermo LTQ-XL ETD and Bruker AmaZon Ion Trap Mass Spectrometers: A Practical Guide. *Journal of the American Society for Mass Spectrometry* **2019**, *30* (9), 1558–1564. https://doi.org/10.1007/s13361-019-02229-z.

(55)   Lemaire, J.; Boissel, P.; Heninger, M.; Mauclaire, G.; Bellec, G.; Mestdagh, H.; Simon, A.; Caer, S. le; Ortega, J. M.; Glotin, F.; Maitre, P. Gas Phase Infrared Spectroscopy of Selectively Prepared Ions. *Physical Review Letters* **2002**, *89* (27), 273002. https://doi.org/10.1103/PhysRevLett.89.273002.

(56)   Londry, F. A.; Hager, J. W. Mass Selective Axial Ion Ejection from a Linear Quadrupole Ion Trap. *Journal of the American Society for Mass Spectrometry* **2003**, *14* (10), 1130–1147. https://doi.org/10.1016/S1044-0305(03)00446-X.

(57)   Coughlan, N. J. A.; Carr, P. J. J.; Walker, S. C.; Zhou, C.; Guna, M.; Campbell, J. L.; Hopkins, W. S. Measuring Electronic Spectra of Differential Mobility-Selected Ions in the Gas Phase. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (2), 405–410. https://doi.org/10.1021/jasms.9b00039.

(58)   Zhang, Q. Characterizing Humulone Content in Beer Using Differential Mobility Spectrometry, 2018.

(59)   Bowman, Z. A. Determining Molecular Physicochemical Properties Using Differential Mobility Spectrometry By, 2019.

(60)   Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. pietro; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. v.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G. L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J.

L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17* (3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.

(61)   Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **1964**, *36* (8), 1627–1639. https://doi.org/10.1021/ac60214a047.

(62)   Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *The Journal of Physical Chemistry A* **1997**, *101* (28), 5111–5116. https://doi.org/10.1021/jp970984n.

(63)   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995**, *117* (19), 5179–5197. https://doi.org/10.1021/ja00124a002.

(64)   Eberhart; Yuhui Shi. Particle Swarm Optimization: Developments, Applications and Resources. In *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*; IEEE, 2001; Vol. 1, pp 81–86. https://doi.org/10.1109/CEC.2001.934374.

(65)   Storn, R.; Price, K. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* **1997**, *11* (4), 341–359. https://doi.org/10.1023/A:1008202821328.

(66)   Kohonen, T. The Self-Organizing Map. *Proceedings of the IEEE* **1990**, *78* (9), 1464–1480. https://doi.org/10.1109/5.58325.

(67)   Fritzke, B. A Growing Neural Gas Network Learns Topologies. In *Proceedings of the 7th International Conference on Neural Information Processing Systems*; NIPS'94; MIT Press: Cambridge, MA, USA, 1994; pp 625–632.

(68)   Fu, W.; Hopkins, W. S. Applying Machine Learning to Vibrational Spectroscopy. *Journal of Physical Chemistry A* **2018**, *122* (1), 167–171. https://doi.org/10.1021/acs.jpca.7b10303.

(69)   Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **1963**, *58* (301), 236–244. https://doi.org/10.1080/01621459.1963.10500845.

(70)   Tagilayev, A. R. *Modern Multidimensional Scaling*; Springer Series in Statistics; Springer New York: New York, NY, 2005; Vol. 36. https://doi.org/10.1007/0-387-28981-X.

(71)   Sokal, R. R. *A Statistical Method for Evaluating Systematic Relationships*; 1958; Vol. 38.

(72)   Patnaik, A. K.; Bhuyan, P. K.; Krishna Rao, K. v. Divisive Analysis (DIANA) of Hierarchical Clustering and GPS Data for Level of Service Criteria of Urban Streets. *Alexandria Engineering Journal* **2016**, *55* (1), 407–418. https://doi.org/10.1016/j.aej.2015.11.003.

(73)   Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; Štajdohar, M.; Umek, L.; Žagar, L.; Žbontar, J.; Žitnik, M.; Zupan, B. Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research* **2013**, *14*, 2349–2353.

(74)   Shah, S. A.; Koltun, V. Robust Continuous Clustering. *Proceedings of the National Academy of Sciences* **2017**, *114* (37), 9814–9819. https://doi.org/10.1073/pnas.1700770114.

(75)   Borg, I.; Groenen, P. J. F. *Modern Multidimensional Scaling*; Springer Series in Statistics; Springer New York: New York, NY, 2005. https://doi.org/10.1007/0-387-28981-X.

(76)   Torgerson, W. S. *Theory and Methods of Scaling*; Wiley, 1958.

(77)   Born, M.; Oppenheimer, R. Zur Quantentheorie Der Molekeln. *Annalen der Physik* **1927**, *389* (20), 457–484. https://doi.org/10.1002/andp.19273892002.

(78)   Sahni, V. *Quantal Density Functional Theory*; 2004. https://doi.org/10.1007/978-3-662-09624-6.

(79)   van Doren, V. E.; van Alsenoy, C.; Geerlings, P.; of Physics, A. I. *Density Functional Theory and Its Application to Materials: Antwerp, Belgium, 8-10 June 2000*; AIP conference proceedings; American Institute of Physics, 2001.

(80)   Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *The Journal of Chemical Physics* **1993**, *98* (7), 5648–5652. https://doi.org/10.1063/1.464913.

(81)   Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *Journal of Physical Chemistry®* **1994**, *98* (45), 11623–11627. https://doi.org/10.1021/j100096a001.

(82)   Chai, J. da; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Physical Chemistry Chemical Physics* **2008**, *10* (44), 6615–6620. https://doi.org/10.1039/b810189b.

(83)   Lin, Y. S.; Li, G. de; Mao, S. P.; Chai, J. da. Long-Range Corrected Hybrid Density Functionals with Improved Dispersion Corrections. *Journal of Chemical Theory and Computation* **2013**, *9* (1), 263–272. https://doi.org/10.1021/ct300715s.

(84)   Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Molecular Physics* **2017**, *115* (19), 2315–2372. https://doi.org/10.1080/00268976.2017.1333644.

(85)   Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters* **1984**, *52* (12), 997–1000. https://doi.org/10.1103/PhysRevLett.52.997.

(86)   Marques, M. A. L.; Ullrich, C. A.; Nogueira, F.; Rubio, A.; Burke, K.; Gross, E. K. U. *Time-Dependent Density Functional Theory*; Marques, M. A. L., Ullrich, C. A., Nogueira, F., Rubio, A., Burke, K., Gross, E. K. U., Eds.; Lecture Notes in Physics; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; Vol. 706. https://doi.org/10.1007/b11767107.

(87)   Ullrich, C. A.; Yang, Z. A Brief Compendium of Time-Dependent Density Functional Theory. *Brazilian Journal of Physics* **2014**, *44* (1), 154–188. https://doi.org/10.1007/s13538-013-0141-2.

(88)   Casida, M. E. Time-Dependent Density Functional Response Theory for Molecules; 1995; pp 155–192. https://doi.org/10.1142/9789812830586_0005.

(89)   Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Books on Chemistry; Dover Publications, 1996.

(90)   Bartlett, R. J.; Musiał, M. Coupled-Cluster Theory in Quantum Chemistry. *Reviews of Modern Physics* **2007**, *79* (1), 291–352. https://doi.org/10.1103/RevModPhys.79.291.

(91)   Raghavachari, K.; Pople, J. A.; Replogle, E. S.; Head-Gordon, M. Fifth-Order Møller-Plesset Perturbation Theory: Comparison of Existing Correlation Methods and

Implementation of New Methods Correct to Fifth Order. *Journal of Physical Chemistry* **1990**, *94* (14), 5579–5586. https://doi.org/10.1021/j100377a033.

(92)   Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chemical Physics Letters* **1989**, *157* (6), 479–483. https://doi.org/10.1016/S0009-2614(89)87395-6.

(93)   Izsák, R. Single-Reference Coupled Cluster Methods for Computing Excitation Energies in Large Molecules: The Efficiency and Accuracy of Approximations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, *10* (3), 1–29. https://doi.org/10.1002/wcms.1445.

(94)   *Methods of Electronic Structure Theory*; Schaefer, H. F., Ed.; Springer US: Boston, MA, 1977. https://doi.org/10.1007/978-1-4757-0887-5.

(95)   Liakos, D. G.; Neese, F. Is It Possible To Obtain Coupled Cluster Quality Energies at near Density Functional Theory Cost? Domain-Based Local Pair Natural Orbital Coupled Cluster vs Modern Density Functional Theory. *Journal of Chemical Theory and Computation* **2015**, *11* (9), 4054–4063. https://doi.org/10.1021/acs.jctc.5b00359.

(96)   Riplinger, C.; Neese, F. An Efficient and near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *Journal of Chemical Physics* **2013**, *138* (3). https://doi.org/10.1063/1.4773581.

(97)   Neese, F.; Wennmohs, F.; Hansen, A. Efficient and Accurate Local Approximations to Coupled-Electron Pair Approaches: An Attempt to Revive the Pair Natural Orbital Method. *Journal of Chemical Physics* **2009**, *130* (11). https://doi.org/10.1063/1.3086717.

(98)   Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *Journal of Chemical Physics* **2013**, *139* (13). https://doi.org/10.1063/1.4821834.

(99)   Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse Maps - A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory. *Journal of Chemical Physics* **2016**, *144* (2). https://doi.org/10.1063/1.4939030.

(100)   Meyer, W. PNO–CI Studies of Electron Correlation Effects. I. Configuration Expansion by Means of Nonorthogonal Orbitals, and Application to the Ground State and Ionized States of Methane. *The Journal of Chemical Physics* **1973**, *58* (3), 1017–1035. https://doi.org/10.1063/1.1679283.

(101)   Dutta, A. K.; Neese, F.; Izsák, R. Towards a Pair Natural Orbital Coupled Cluster Method for Excited States. *Journal of Chemical Physics* **2016**, *145* (3). https://doi.org/10.1063/1.4958734.

(102)   Davidson, E. R. The Iterative Calculation of a Few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices. *Journal of Computational Physics* **1975**, *17* (1), 87–94. https://doi.org/10.1016/0021-9991(75)90065-0.

(103)   Nooijen, M.; Bartlett, R. J. Similarity Transformed Equation-of-Motion Coupled-Cluster Theory: Details, Examples, and Comparisons. *Journal of Chemical Physics* **1997**, *107* (17), 6812–6830. https://doi.org/10.1063/1.474922.

(104)   Nooijen, M. Many-Body Similarity Transformations Generated by Normal Ordered Exponential Excitation Operators. *Journal of Chemical Physics* **1996**, *104* (7), 2638–2651. https://doi.org/10.1063/1.470988.

(105)   Sous, J.; Goel, P.; Nooijen, M. Similarity Transformed Equation of Motion Coupled Cluster Theory Revisited: A Benchmark Study of Valence Excited States. *Molecular Physics* **2014**, *112* (5–6), 616–638. https://doi.org/10.1080/00268976.2013.847216.

(106)   Fukui, K. The Path of Chemical Reactions - The IRC Approach. *Accounts of Chemical Research* **1981**, *14* (12), 363–368. https://doi.org/10.1021/ar00072a001.

(107)   Eckert, F.; Pulay, P.; Werner, H. J. Ab Initio Geometry Optimization for Large Molecules. *Journal of Computational Chemistry* **1997**, *18* (12), 1473–1483. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1473::AID-JCC5>3.0.CO;2-G.

(108)   Peng, C.; Bernhard Schlegel, H. Combining Synchronous Transit and Quasi-Newton Methods to Find Transition States. *Israel Journal of Chemistry* **1993**, *33* (4), 449–454. https://doi.org/10.1002/ijch.199300051.

(109)   Schlegel, H. B. Optimization of Equilibrium Geometries and Transition Structures_WhatGaussianIsBasedOn_.Pdf. *Journal of computational chemistry* **1982**, *3* (2), 214–218.

(110)   Li, X.; Frisch, M. J. Energy-Represented Direct Inversion in the Iterative Subspace within a Hybrid Geometry Optimization Method. *Journal of Chemical Theory and Computation* **2006**, *2* (3), 835–839. https://doi.org/10.1021/ct050275a.

(111)   Smeeton, L. C.; Oakley, M. T.; Johnston, R. L. Visualizing Energy Landscapes with Metric Disconnectivity Graphs. *Journal of Computational Chemistry* **2014**, *35* (20), 1481–1490. https://doi.org/10.1002/jcc.23643.

(112)   Scott Hopkins, W.; Marta, R. A.; Steinmetz, V.; McMahon, T. B. Mode-Specific Fragmentation of Amino Acid-Containing Clusters. *Physical Chemistry Chemical Physics* **2015**, *17* (43), 28548–28555. https://doi.org/10.1039/c5cp03517a.

(113)   Baiardi, A.; Bloino, J.; Barone, V. General Formulation of Vibronic Spectroscopy in Internal Coordinates. *Journal of Chemical Physics* **2016**, *144* (8). https://doi.org/10.1063/1.4942165.

(114)   de Souza, B.; Neese, F.; Izsák, R. On the Theoretical Prediction of Fluorescence Rates from First Principles Using the Path Integral Approach. *Journal of Chemical Physics* **2018**, *148* (3). https://doi.org/10.1063/1.5010895.

(115)   Griffiths, D. J. *Introduction to Quantum Mechanics*; Cambridge University Press, 2017.

(116)   Eckart, C. Some Studies Concerning Rotating Axes and Polyatomic Molecules. *Physical Review* **1935**, *47* (7), 552–558. https://doi.org/10.1103/PhysRev.47.552.

(117)   Dymarsky, A. Y.; Kudin, K. N. Computation of the Pseudorotation Matrix to Satisfy the Eckart Axis Conditions. *Journal of Chemical Physics* **2005**, *122* (12), 5–7. https://doi.org/10.1063/1.1864872.

(118)   Condon, E. A Theory of Intensity Distribution in Band Systems. *Physical Review* **1926**, *28* (6), 1182–1201. https://doi.org/10.1103/PhysRev.28.1182.

(119)   Avila Ferrer, F. J.; Barone, V.; Cappelli, C.; Santoro, F. Duschinsky, Herzberg-Teller, and Multiple Electronic Resonance Interferential Effects in Resonance Raman Spectra and Excitation Profiles. The Case of Pyrene. *Journal of Chemical Theory and Computation* **2013**, *9* (8), 3597–3611. https://doi.org/10.1021/ct400197y.

(120)   Avila Ferrer, F. J.; Santoro, F. Comparison of Vertical and Adiabatic Harmonic Approaches for the Calculation of the Vibrational Structure of Electronic Spectra. *Physical Chemistry Chemical Physics* **2012**, *14* (39), 13549–13563. https://doi.org/10.1039/c2cp41169e.

(121)   Köuppel, H.; Domcke, W.; Cederbaum, L. S. Multimode Molecular Dynamics Beyond the Born-Oppenheimer Approximation; 2007; pp 59–246. https://doi.org/10.1002/9780470142813.ch2.

(122)   Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *Journal of Chemical Theory and Computation* **2011**, *7* (8), 2427–2438. https://doi.org/10.1021/ct2002946.

(123)   Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga, J. Non-Iterative Fifth-Order Triple and Quadruple Excitation Energy Corrections in Correlated Methods. *Chemical*

Physics Letters **1990**, *165* (6), 513–522. https://doi.org/10.1016/0009-2614(90)87031-L.

(124)   Varandas, A. J. C. CBS Extrapolation in Electronic Structure Pushed to the End: A Revival of Minimal and Sub-Minimal Basis Sets. *Physical Chemistry Chemical Physics* **2018**, *20* (34), 22084–22098. https://doi.org/10.1039/c8cp02932f.

(125)   Manna, D.; Martin, J. M. L. What Are the Ground State Structures of C20 and C24? An Explicitly Correlated Ab Initio Approach. *Journal of Physical Chemistry A* **2016**, *120* (1), 153–160. https://doi.org/10.1021/acs.jpca.5b10266.

(126)   Eriksen, J. J.; Jørgensen, P.; Gauss, J. On the Convergence of Perturbative Coupled Cluster Triples Expansions: Error Cancellations in the CCSD(T) Model and the Importance of Amplitude Relaxation. *Journal of Chemical Physics* **2015**, *142* (1). https://doi.org/10.1063/1.4904754.

(127)   Watts, J. D.; Gauss, J.; Bartlett, R. J. Coupled-Cluster Methods with Noniterative Triple Excitations for Restricted Open-Shell Hartree-Fock and Other General Single Determinant Reference Functions. Energies and Analytical Gradients. *The Journal of Chemical Physics* **1993**, *98* (11), 8718–8733. https://doi.org/10.1063/1.464480.

(128)   Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-Accuracy Extrapolated Ab Initio Thermochemistry. III. Additional Improvements and Overview. *Journal of Chemical Physics* **2008**, *128* (11). https://doi.org/10.1063/1.2835612.

(129)   Varandas, A. J. C. Straightening the Hierarchical Staircase for Basis Set Extrapolations: A Low-Cost Approach to High-Accuracy Computational Chemistry. *Annual Review of Physical Chemistry* **2018**, *69*, 177–203. https://doi.org/10.1146/annurev-physchem-050317-021148.

(130)   Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *Journal of Chemical Physics* **1997**, *106* (23), 9639–9646. https://doi.org/10.1063/1.473863.

(131)   Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *Journal of Chemical Theory and Computation* **2015**, *11* (4), 1525–1539. https://doi.org/10.1021/ct501129s.

(132)   Andersson, M. P.; Uvdal, P. New Scale Factors for Harmonic Vibrational Frequencies Using the B3LYP Density Functional Method with the Triple-ζ Basis Set 6-311+G(d,p). *Journal of Physical Chemistry A* **2005**, *109* (12), 2937–2941. https://doi.org/10.1021/jp045733a.

(133)   Pfeiffer, F.; Rauhut, G.; Feller, D.; Peterson, K. A. Anharmonic Zero Point Vibrational Energies: Tipping the Scales in Accurate Thermochemistry Calculations? *Journal of Chemical Physics* **2013**, *138* (4). https://doi.org/10.1063/1.4777568.

(134)   Adamo, C.; Jacquemin, D. The Calculations of Excited-State Properties with Time-Dependent Density Functional Theory. *Chemical Society Reviews* **2013**, *42* (3), 845–856. https://doi.org/10.1039/c2cs35394f.

(135)   Muniz-Miranda, F.; Pedone, A.; Battistelli, G.; Montalti, M.; Bloino, J.; Barone, V. Benchmarking TD-DFT against Vibrationally Resolved Absorption Spectra at Room Temperature: 7-Aminocoumarins as Test Cases. *Journal of Chemical Theory and Computation* **2015**, *11* (11), 5371–5384. https://doi.org/10.1021/acs.jctc.5b00750.

(136)   Bader, R. F. W. Atoms in Molecules. *Accounts of Chemical Research* **1985**, *18* (1), 9–15. https://doi.org/10.1021/ar00109a003.

(137)   Kumar, P. S. V.; Raghavendra, V.; Subramanian, V. Bader's Theory of Atoms in Molecules (AIM) and Its Applications to Chemical Bonding. *Journal of Chemical Sciences* **2016**, *128* (10), 1527–1536. https://doi.org/10.1007/s12039-016-1172-3.

(138)   Monteiro, N. K. V.; Firme, C. L. Hydrogen-Hydrogen Bonds in Highly Branched Alkanes and in Alkane Complexes: A DFT, Ab Initio, QTAIM, and ELF Study. *Journal of Physical Chemistry A* **2014**, *118* (9), 1730–1740. https://doi.org/10.1021/jp500131z.

(139)   Matta, C. F.; Hernández-Trujillo, J.; Tang, T. H.; Bader, R. F. W. Hydrogen - Hydrogen Bonding: A Stabilizing Interaction in Molecules and Crystals. *Chemistry - A European Journal* **2003**, *9* (9), 1940–1951. https://doi.org/10.1002/chem.200204626.

(140)   Lepetit, C.; Fau, P.; Fajerwerg, K.; Kahn, M. L.; Silvi, B. Topological Analysis of the Metal-Metal Bond: A Tutorial Review. *Coordination Chemistry Reviews* **2017**, *345*, 150–181. https://doi.org/10.1016/j.ccr.2017.04.009.

(141)   Bianchi, R.; Gervasio, G.; Marabello, D. Experimental Electron Density Analysis of Mn2(CO)10: Metal-Metal and Metal-Ligand Bond Characterization. *Inorganic Chemistry* **2000**, *39* (11), 2360–2366. https://doi.org/10.1021/ic991316e.

(142)   Rosenberg, R. E. Can 2-X-Ethanols Form Intramolecular Hydrogen Bonds? *The Journal of Physical Chemistry A* **2019**, *123* (35), 7651–7660. https://doi.org/10.1021/acs.jpca.9b06801.

(143)   Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *Journal of the American Chemical Society* **2010**, *132* (18), 6498–6506. https://doi.org/10.1021/ja100936w.

(144)   Plasser, F.; Wormit, M.; Dreuw, A. New Tools for the Systematic Analysis and Visualization of Electronic Excitations. I. Formalism. *Journal of Chemical Physics* **2014**, *141* (2). https://doi.org/10.1063/1.4885819.

(145)   Plasser, F.; Bäppler, S. A.; Wormit, M.; Dreuw, A. New Tools for the Systematic Analysis and Visualization of Electronic Excitations. II. Applications. *The Journal of Chemical Physics* **2014**, *141* (2), 024107. https://doi.org/10.1063/1.4885820.

(146)   Nanita, S. C.; Cooks, R. G. Serine Octamers: Cluster Formation, Reactions, and Implications for Biomolecule Homochirality. *Angewandte Chemie - International Edition* **2006**, *45* (4), 554–569. https://doi.org/10.1002/anie.200501328.

(147)   Yang, P.; Xu, R.; Nanita, S. C.; Cooks, R. G. Thermal Formation of Homochiral Serine Clusters and Implications for the Origin of Homochirality. *Journal of the*

American Chemical Society **2006**, *128* (51), 17074–17086. https://doi.org/10.1021/ja064617d.

(148)   Vandenbussche, S.; Vandenbussche, G.; Reisse, J.; Bartik, K. Do Serine Octamers Exist in Solution? Relevance of This Question in the Context of the Origin of Homochirality on Earth. *European Journal of Organic Chemistry* **2006**, No. 14, 3069–3073. https://doi.org/10.1002/ejoc.200600370.

(149)   Scutelnic, V.; Perez, M. A. S.; Marianski, M.; Warnke, S.; Gregor, A.; Rothlisberger, U.; Bowers, M. T.; Baldauf, C.; von Helden, G.; Rizzo, T. R.; Seo, J. The Structure of the Protonated Serine Octamer. *Journal of the American Chemical Society* **2018**, *140* (24), 7554–7560. https://doi.org/10.1021/jacs.8b02118.

(150)   Ma, L.; Ren, J.; Feng, R.; Zhang, K.; Kong, X. Structural Characterizations of Protonated Homodimers of Amino Acids: Revealed by Infrared Multiple Photon Dissociation (IRMPD) Spectroscopy and Theoretical Calculations. *Chinese Chemical Letters* **2018**, *29* (9), 1333–1339. https://doi.org/10.1016/j.cclet.2018.02.008.

(151)   Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *Journal of Computational Chemistry* **1990**, *11* (3), 361–373. https://doi.org/10.1002/jcc.540110311.

(152)   Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *The Journal of Chemical Physics* **2010**, *132* (15), 154104. https://doi.org/10.1063/1.3382344.

(153)   M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V., W. C. Gaussian 09 Rev. D.01. 2016.

(154)   Neese, F. The ORCA Program System. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2* (1), 73–78. https://doi.org/10.1002/wcms.81.

(155)   Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Physical Chemistry Chemical Physics* **2005**, *7* (18), 3297. https://doi.org/10.1039/b508541a.

(156)   Hellweg, A.; Hättig, C.; Höfener, S.; Klopper, W. Optimized Accurate Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations for the Atoms Rb to Rn. *Theoretical Chemistry Accounts* **2007**, *117* (4), 587–597. https://doi.org/10.1007/s00214-007-0250-5.

(157)   Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **1963**, *58* (301), 236–244. https://doi.org/10.1080/01621459.1963.10500845.

(158)   Wickelmaier, F. *An Introduction to MDS*; Aalborg Universitetsforlag: Aalborg, 2003.

(159)   Parneix, P.; Basire, M.; Calvo, F. Accurate Modeling of Infrared Multiple Photon Dissociation Spectra: The Dynamical Role of Anharmonicities. *Journal of Physical Chemistry A* **2013**, *117* (19), 3954–3959. https://doi.org/10.1021/jp402459f.

(160)   Schofield, D. P.; Kjaergaard, H. G.; Matthews, J.; Sinha, A. The OH-Stretching and OOH-Bending Overtone Spectrum of HOONO. *Journal of Chemical Physics* **2005**, *123* (13). https://doi.org/10.1063/1.2047574.

(161)   Oomens, J.; Steill, J. D.; Redlich, B. Gas-Phase IR Spectroscopy of Deprotonated Amino Acids. *Journal of the American Chemical Society* **2009**, *131* (12), 4310–4319. https://doi.org/10.1021/ja807615v.

(162)   Steill, J. D.; Szczepanski, J.; Oomens, J.; Eyler, J. R.; Brajter-Toth, A. Structural Characterization by Infrared Multiple Photon Dissociation Spectroscopy of Protonated Gas-Phase Ions Obtained by Electrospray Ionization of Cysteine and Dopamine.

Analytical and Bioanalytical Chemistry **2011**, *399* (7), 2463–2473. https://doi.org/10.1007/s00216-010-4582-y.

(163)   Ieritano, C.; Carr, P. J. J.; Hasan, M.; Burt, M.; Marta, R. A.; Steinmetz, V.; Fillion, E.; McMahon, T. B.; Scott Hopkins, W. The Structures and Properties of Proton-and Alkali-Bound Cysteine Dimers. *Physical Chemistry Chemical Physics* **2016**, *18* (6), 4704–4710. https://doi.org/10.1039/c5cp07414b.

(164)   Jurkschat, K.; Tzschach, A. 1-Aza-5-Stanna-5,5-Dimethylbicyclo[3.3.01,5] Octan Und 1-Aza-5-Stanna-5-Methyltricyclo[3.3.3.01,5] Undecan, Pentakoordinierte Tetraorganozinnverbindungen. *Journal of Organometallic Chemistry* **1984**, *272* (1), 160–162. https://doi.org/10.1016/0022-328X(84)80450-7.

(165)   Kavoosi, A.; Fillion, E. Synthesis and Characterization of Tricarbastannatranes and Their Reactivity in B(C6F5)3-Promoted Conjugate Additions. *Angewandte Chemie - International Edition* **2015**, *54* (18), 5488–5492. https://doi.org/10.1002/anie.201500983.

(166)   Vedejs, E.; Haight, A. R.; Moss, W. O. Internal Coordination at Tin Promotes Selective Alkyl Transfer in the Stille Coupling Reaction. *Journal of the American Chemical Society* **1992**, *114* (16), 6556–6558. https://doi.org/10.1021/ja00042a044.

(167)   Neelakantan, L.; Molin-Case, J. A. Crystal and Molecular Structure of 2-p-Bromophenyl-3,4-Dimethyl-5-Phenyloxazolidine. *The Journal of Organic Chemistry* **1984**, *36* (16), 2261–2262. https://doi.org/10.1021/jo00815a013.

(168)   Jurkschat, K.; Tszchach, A.; Meunier-Piret, J. Crystal and Molecular Structure of 1-AZA-5-STANNA-5-Methyltricyclo[3.3.3.01,5]Undecane.Evidence for a Transannular Donor-Acceptor Interaction in a Tetraorganotin Compound. *Journal of Organometallic Chemistry* **1986**, *315* (1), 45–49. https://doi.org/10.1016/0022-328X(86)80409-0.

(169)   Jurkschat, K.; Kolb, U.; Dräger, M.; Dargatz, M. Unusual Hexacoordination in a Triorganotin Fluoride Supported by Intermolecular Hydrogen Bonds. Crystal and Molecular Structures of 1-Aza-5-Stanna-5-Halogenotricyclo[3.3.3.01.5]Undecanes

N(CH2CH2CH2)3SnF·H2O and N(CH2CH2CH2)3SnX (X = Cl, Br, I). *Organometallics* **1995**, *14* (6), 2827–2834. https://doi.org/10.1021/om00006a031.

(170)   Schenzel, K.; Kolbe, A.; Reich, P. Zum Schwingungsverhalten Pentakoordinierter Zinn(IV)-Organischer Verbindungen. IR- Und Raman-Untersuchungen an 2,8,9-Tricarbastannatranen. *Monatshefte für Chemie Chemical Monthly* **1990**, *121* (8–9), 615–623. https://doi.org/10.1007/BF00809764.

(171)   Ignatyev, I. S.; Schenzel, K.; Sundius, T. Normal Coordinate Analysis of the Vibrational Spectra of Stannatranes. *Spectrochimica Acta Part A: Molecular Spectroscopy* **1992**, *48* (4), 489–493. https://doi.org/10.1016/0584-8539(92)80039-Y.

(172)   Kolb, U.; Dräger, M. Hypervalent Tin-Organic Compounds: Vibrational Spectroscopy in the Solid as a Tool for Structure Determination. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy* **1997**, *53* (4 PART A), 517–529. https://doi.org/10.1016/s1386-1425(96)01822-7.

(173)   Simidzija, P.; Lecours, M. J.; Marta, R. A.; Steinmetz, V.; McMahon, T. B.; Fillion, E.; Hopkins, W. S. Changes in Tricarbastannatrane Transannular N-Sn Bonding upon Complexation Reveal Lewis Base Donicities. *Inorganic Chemistry* **2016**, *55* (19), 9579–9585. https://doi.org/10.1021/acs.inorgchem.6b01185.

(174)   Shujah, S.; Zia-Ur-Rehman; Muhammad, N.; Shah, A.; Ali, S.; Meetsma, A.; Hussain, Z. Homobimetallic Organotin(IV) Complexes with Hexadentate Schiff Base: Synthesis, Crystal Structure and Antimicrobial Studies. *Journal of Organometallic Chemistry* **2014**, *759*, 19–26. https://doi.org/10.1016/j.jorganchem.2014.02.010.

(175)   Gericke, R.; Wagler, J. Molecular Structures of Sn(II) and Sn(IV) Compounds with Di-, Tri- and Tetramethylene Bridged Salen* Type Ligands. *Main Group Metal Chemistry* **2014**, *37* (1–2), 1–9. https://doi.org/10.1515/mgmc-2014-0004.

(176)   Sharma, R.; Ravikanth, M. Synthesis, Structure and Properties of the First Examples of Hexacoordinate SnIVComplexes of Pyrrolyldipyrrins. *European Journal of Inorganic Chemistry* **2017**, *2017* (4), 829–834. https://doi.org/10.1002/ejic.201601179.

(177)   Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilizaion of AB Initio Molecular Potentials for the Prevision of Solvent Effects. *Chemical Physics* **1981**, *55* (1), 117–129. https://doi.org/10.1016/0301-0104(81)85090-2.

(178)   Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *Journal   of   Chemical   Information   and   Modeling*   **2019**. https://doi.org/10.1021/acs.jcim.9b00725.

(179)   Metz, B.; Stoll, H.; Dolg, M. Small-Core Multiconfiguration-Dirac–Hartree–Fock-Adjusted Pseudopotentials for Post- d Main Group Elements: Application to PbH and PbO. *The Journal of Chemical Physics* **2000**, *113* (7), 2563–2569. https://doi.org/10.1063/1.1305880.

(180)   Rappoport, D.; Furche, F. Property-Optimized Gaussian Basis Sets for Molecular Response Calculations. *The Journal of Chemical Physics* **2010**, *133* (13), 134105. https://doi.org/10.1063/1.3484283.

(181)   Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Physical   Chemistry   Chemical   Physics*   **2005**,   *7*   (18),   3297. https://doi.org/10.1039/b508541a.

(182)   Fillion, E.; Taylor, N. J. Cine-Substitution in the Stille Coupling: Evidence for the Carbenoid   Reactivity   of   Sp3-Gem-Organodimetallic   Iodopalladio-Trialkylstannylalkane Intermediates. *Journal of the American Chemical Society* **2003**, *125* (42), 12700–12701. https://doi.org/10.1021/ja037409j.

(183)   Sebahar, H. L.; Yoshida, K.; Hegedus, L. S. Effect of Adjacent Chiral Tertiary and Quaternary Centers on the Metal-Catalyzed Allylic Substitution Reaction. *Journal of Organic Chemistry* **2002**, *67* (11), 3788–3795. https://doi.org/10.1021/jo0200318.

(184)   Li, L.; Wang, C. Y.; Huang, R.; Biscoe, M. R. Stereoretentive Pd-Catalysed Stille Cross-Coupling Reactions of Secondary Alkyl Azastannatranes and Aryl Halides. *Nature Chemistry* **2013**, *5* (7), 607–612. https://doi.org/10.1038/nchem.1652.

(185)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. v; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision E.01. Gaussian, Inc.: Wallingford CT 2009.

(186)   Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *Journal of Computational Chemistry* **2012**, *33* (5), 580–592. https://doi.org/10.1002/jcc.22885.

(187)   Humphrey, W.; Dalke, A.; Schulten, K. {VMD} -- {V}isual {M}olecular {D}ynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(188)   Stone, J. An Efficient Library for Parallel Ray Tracing and Animation, 1998.

(189)   Alkorta, I.; Thacker, J. C. R.; Popelier, P. L. A. An Interacting Quantum Atom Study of Model SN2 Reactions (X−···CH3X, X = F, Cl, Br, and I). *Journal of Computational Chemistry* **2018**, *39* (10), 546–556. https://doi.org/10.1002/jcc.25098.

(190)   Bouas-Laurent, H.; Dürr, H. Organic Photochromism (IUPAC Technical Report). *Pure and Applied Chemistry* **2001**, *73* (4), 639–665. https://doi.org/10.1351/pac200173040639.

(191)   Roy, D.; Balanarayan, P.; Gadre, S. R. An Appraisal of Poincaŕ-Hopf Relation and Application to Topography of Molecular Electrostatic Potentials. *Journal of Chemical Physics* **2008**, *129* (17). https://doi.org/10.1063/1.2999558.

(192)   Yeagle, P. L.; Albert, A. D. G-Protein Coupled Receptor Structure. *Biochimica et Biophysica Acta - Biomembranes* **2007**, *1768* (4), 808–824. https://doi.org/10.1016/j.bbamem.2006.10.002.

(193)   Irie, M.; Fukaminato, T.; Matsuda, K.; Kobatake, S. Photochromism of Diarylethene Molecules and Crystals: Memories, Switches, and Actuators. *Chemical Reviews* **2014**, *114* (24), 12174–12277. https://doi.org/10.1021/cr500249p.

(194)   Merino, E.; Ribagorda, M. Control over Molecular Motion Using the Cis-Trans Photoisomerization of the Azo Group. *Beilstein Journal of Organic Chemistry* **2012**, *8*, 1071–1090. https://doi.org/10.3762/bjoc.8.119.

(195)   Solid Color-Coded PEEK Tubing https://www.vici.com/tube/solid-peek.php (accessed May 14, 2020).

(196)   Holman, J. D.; Tabb, D. L.; Mallick, P. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; Vol. 135, pp 13.24.1-13.24.9. https://doi.org/10.1002/0471250953.bi1324s46.

(197)   Velema, W. A.; Szymanski, W.; Feringa, B. L. Photopharmacology: Beyond Proof of Principle. *Journal of the American Chemical Society* **2014**, *136* (6), 2178–2191. https://doi.org/10.1021/ja413063e.

(198)   Sinicropi, A. Biomimetic Photoswitches. *Chim. Ind.* **2010**, 102–109.

(199)   Klajn, R. Immobilized Azobenzenes for the Construction of Photoresponsive Materials. *Pure and Applied Chemistry* **2010**, *82* (12), 2247–2276. https://doi.org/10.1351/PAC-CON-10-09-04.

(200)   Patel, S. K.; Cao, J.; Lippert, A. R. A Volumetric Three-Dimensional Digital Light Photoactivatable Dye Display. *Nature Communications* **2017**, *8*, 1–8. https://doi.org/10.1038/ncomms15239.

(201)   Patrick, A. L.; Cismesia, A. P.; Tesler, L. F.; Polfer, N. C. Effects of ESI Conditions on Kinetic Trapping of the Solution-Phase Protonation Isomer of p-Aminobenzoic Acid in the Gas Phase. *International Journal of Mass Spectrometry* **2017**, *418*, 148–155. https://doi.org/10.1016/j.ijms.2016.09.022.

(202)   Galangau, O.; Nakashima, T.; Maurel, F.; Kawai, T. Substituent Effects on the Photochromic Properties of Benzothiophene-Based Derivatives. *Chemistry - A European Journal* **2015**, *21* (23), 8471–8482. https://doi.org/10.1002/chem.201500647.

(203)   Ejuh, G. W.; Tchangnwa Nya, F.; Ottou Abe, M. T.; Jean-Baptiste, F. F.; Ndjaka, J. M. B. Electronic Structure, Physico-Chemical, Linear and Non Linear Optical Properties Analysis of Coronene, 6B-, 6N-, 3B3N- Substituted C24H12 Using RHF, B3LYP and WB97XD Methods. *Optical and Quantum Electronics* **2017**, *49* (11), 1–14. https://doi.org/10.1007/s11082-017-1221-2.

(204)   Scholz, M. S.; Bull, J. N.; Coughlan, N. J. A.; Carrascosa, E.; Adamson, B. D.; Bieske, E. J. Photoisomerization of Protonated Azobenzenes in the Gas Phase. *Journal of Physical Chemistry A* **2017**, *121* (34), 6413–6419. https://doi.org/10.1021/acs.jpca.7b05902.

(205)   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.; Hratch, D. J. Gaussian 16 Revision A.03. Gaussian, Inc.: Wallingford CT 2016.

(206)   Stevenson, J.; Ruehle, V. Python Energy Landscape Explorer (PELE). 2012.

Appendix A

The Energy Profile of the [Ser$_2$ + H]$^+$ Isomers

| Isomer	E(B3LYP) (Hartree)	G(B3LYP) (Hartree)	Relative G (B3LYP, kJ/mol)	E(CCSD) (Hartree)	Relative G* (CCSD, kJ/mol)
1	-798.630777	-798.433571	0.00	-796.895968	0.00
2	-798.626007	-798.431451	5.57	-796.895470	1.31
3	-798.629181	-798.433136	1.14	-796.895343	1.64
4	-798.629491	-798.432144	3.75	-796.894215	4.60
5	-798.628154	-798.431592	5.20	-796.893967	5.25
6	-798.628322	-798.431752	4.78	-796.893628	6.14
7	-798.628894	-798.430528	7.99	-796.892852	8.18
8	-798.627909	-798.430057	9.23	-796.891454	11.85
9	-798.626617	-798.428397	13.58	-796.890973	13.12
10	-798.625465	-798.430240	8.75	-796.890654	13.95
11	-798.621675	-798.426875	17.58	-796.890578	14.15
12	-798.623024	-798.427932	14.81	-796.890411	14.59
13	-798.624573	-798.427493	15.96	-796.890297	14.89
14	-798.624011	-798.426738	17.94	-796.889976	15.73
15	-798.623929	-798.427893	14.91	-796.889751	16.32
16	-798.623595	-798.427484	15.98	-796.889690	16.48
17	-798.623620	-798.426479	18.62	-796.889380	17.30
18	-798.618667	-798.425654	20.79	-796.889153	17.89
19	-798.617912	-798.425460	21.30	-796.888937	18.46
20	-798.618065	-798.423523	26.38	-796.888464	19.70
21	-798.616086	-798.424478	23.87	-796.888082	20.70
22	-798.618066	-798.424217	24.56	-796.888021	20.86
23	-798.624037	-798.427849	15.02	-796.887779	21.50
24	-798.618788	-798.424285	24.38	-796.887509	22.21
25	-798.617500	-798.423944	25.28	-796.887144	23.17
26	-798.616358	-798.423624	26.12	-796.887132	23.20
27	-798.617830	-798.423837	25.56	-796.887095	23.30
28	-798.622660	-798.426836	17.68	-796.886790	24.10
29	-798.620575	-798.424226	24.54	-796.886775	24.14
30	-798.616415	-798.422553	28.93	-796.886176	25.71
31	-798.616198	-798.422751	28.41	-796.885890	26.46
32	-798.620515	-798.422833	28.19	-796.885715	26.92
33	-798.619504	-798.422846	28.16	-796.885098	28.54
34	-798.620249	-798.421713	31.13	-796.884440	30.27
35	-798.616597	-798.421613	31.40	-796.884207	30.88
36	-798.616269	-798.420925	33.20	-796.883743	32.10
37	-798.617017	-798.421145	32.62	-796.883656	32.32
38	-798.615622	-798.418609	39.28	-796.880891	39.58
39	-798.613853	-798.418668	39.13	-796.877728	47.89
40	-798.608799	-798.413712	52.14	-796.875754	53.07

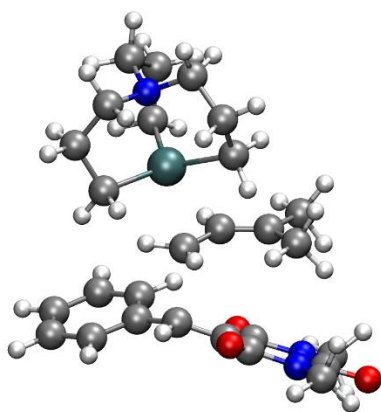*The relative CCSD energy is calculated using equation 3.1
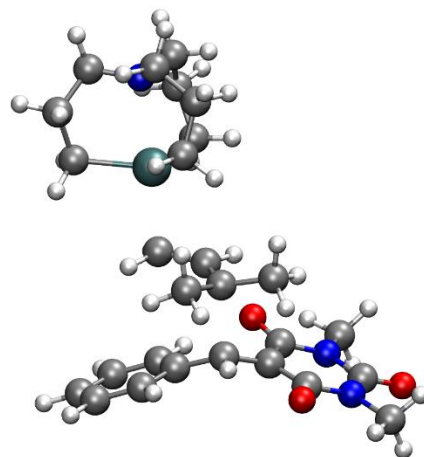
Appendix B
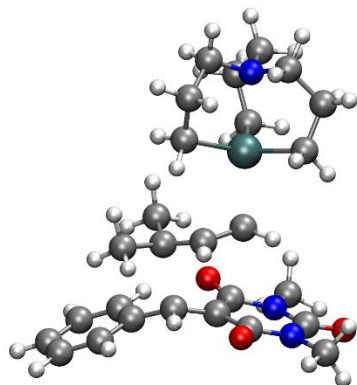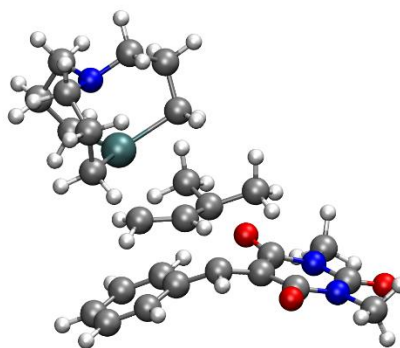
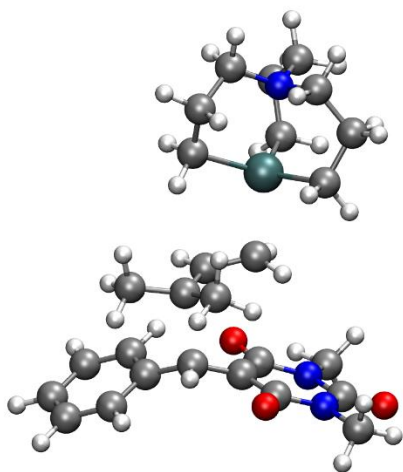## The TS Geometries of the Alkyl-triptych Alkylation Reaction.
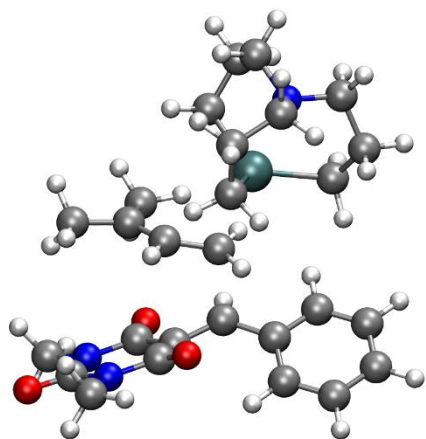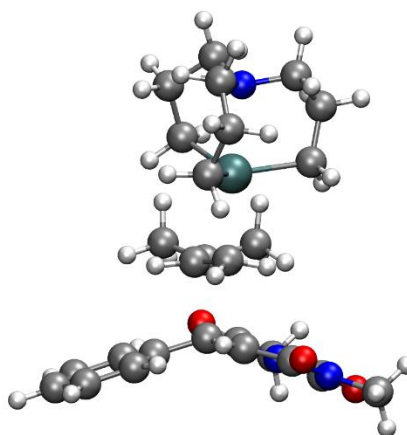

TS3L-1
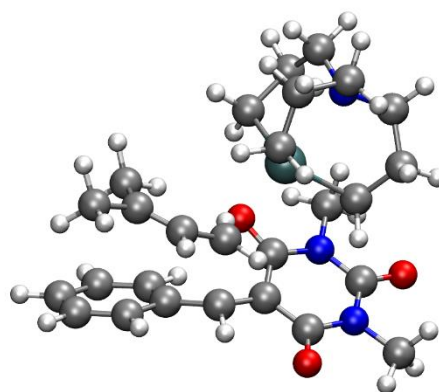
TS3L-2
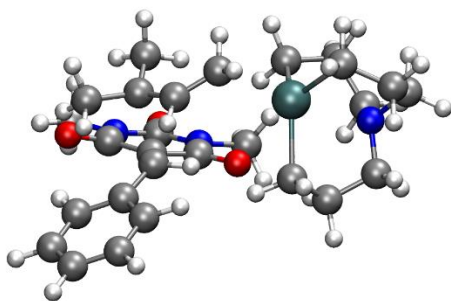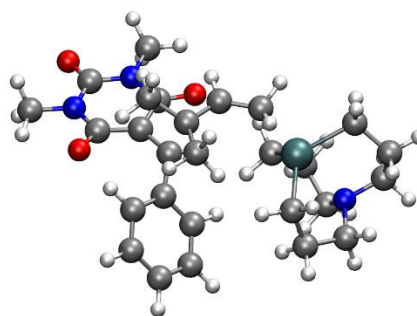

TS3L-3


TS4L-1

TS4L-2

TS4L-3

TS4L-4

TS4L-5

TS4C-1

TS4C-2

126

Appendix C

## AY Sample Degradation Observed in Preliminary Experiments

In one of the preliminary scans, two DMS scans have been conducted at SV =2000 V with 1.5% methanol modifier. One is done when the AY sample (~2 μg/mL) is freshly prepared for the day's work (Figure C-1a, blue), the other is conducted after conducting one round of laser scans that lasts 20 min (Figure C-1a, green). Both a significant ion intensity loss and a change in the relative ratio of the two ion populations (from ~1:6 to ~1:10) are observed after the AY sample is mixed with acid. Using another sample to same experiment (Figure C-1, red), a decrease in ion signal is found as well. This observation promotes the use of a late mixing setup described in section 5.2.1.
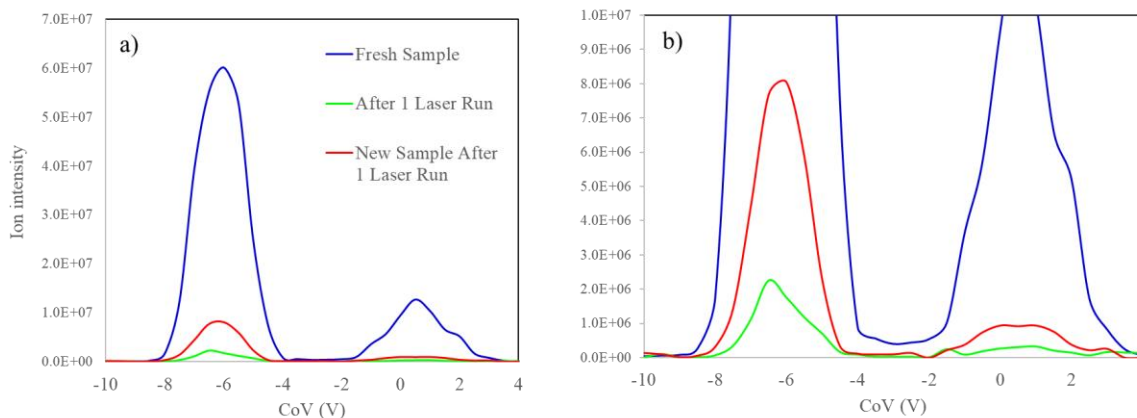
Figure C-1. AY dissociation observed in preliminary experiments. a) Ion intensity ranges from 0 to 7×10^7, and b) Ion intensity ranges from 0 to 10^7.

In later attempts to reproduce the AY dissociation, another AY sample solution of *ca.* 2 μg in Methanol (1% AcOH v/v) is used and a series of DMS measurements are conducted with the same experimental setup. However, the AY dissociation observed earlier is not reproduced, as is shown in Figure C-2. Note that the 'Fresh' dataset is generated from the start of the day's experiment, which is 5 hours ahead of the dissociation measurements, at the beginning of which a new sample solution is prepared out of the stock AY solution in acetonitrile. As a

precaution that the ion composition may change during one laser scan, we decided to employ the late mixing setup.
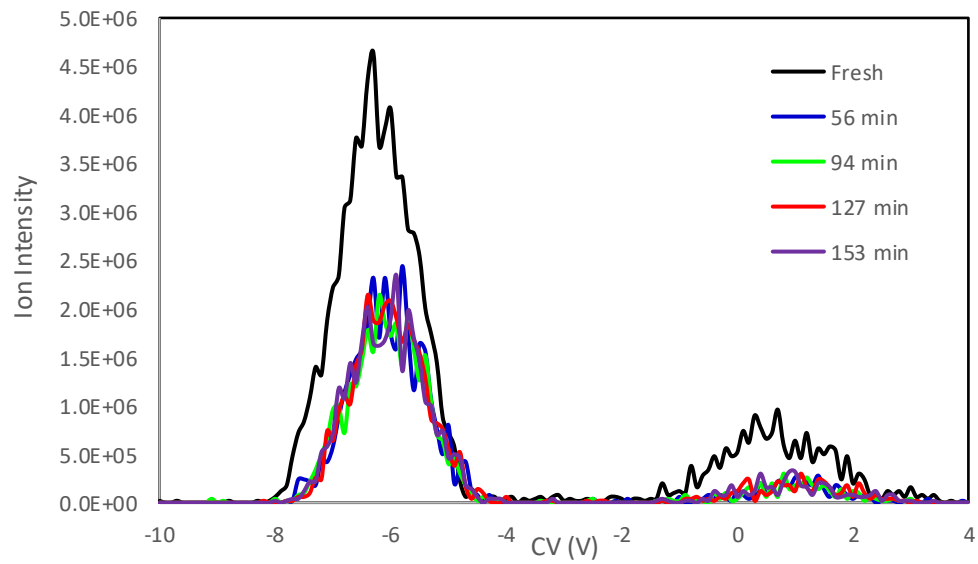
Figure C-2. The second attempt to reproduce the AY degradation.

Appendix D
## The Neutral Acylhydrazone (AY) Species

In parallel with the protonated $[AY + H]^+$ ions, the ACN solvated neutral AY species are also investigated to rationalize the experimental work of Van Dijken *et al.*[13] Results of the investigation as well as a brief discussion is documented here for further reference.

The nomenclature for the neutral AY species inherits that of protonated species (as is defined in section 5.2.2), with differences in the definition of protonation scheme. For the AY isomers, 'O' protonation scheme is defined as protonation on O^1, while 'N_A' and 'N_B' refers to protonation on N^3 and N^4, respectively.

Following the procedures outlined in section 5.2.2, the energy profile and isomer conformations for the AY isomers is summarized in Table D-1 and Figure D-1, respectively. Note that the PCM model for ACN is used for geometric optimizations, frequency calculations, and electronic spectrum simulations. For neutral AY isomers solvated by acetonitrile, the global minimum structure is 1a. A second most abundant isomer, 3a, has a relative Gibbs energy of 8.59 kJ/mol, while all other isomers being investigated are more than 10 kJ/mol higher in energy. As is shown in Figure D-1, the N_A protonated isomers (1a-3a) deviate from planarity at the phenyl group bound to the carbonyl, due to the steric hindrance of the proton on N^3. However, the remaining molecular isomers are planar, as is expected based on the hybridization of the C and N atoms. It should be noted that when searching for a 3c isomer candidate, geometry optimizations were conducted using initial guesses of a ZZE-O skeletal input structure. However, the resulting optimized geometries adopted a ZEE-O form rather than the desired 3c isomer.

**Table D-1. Relative Gibbs Energies (in kJ/mol) of AY Isomers**

Isomer Label	Bonding Scheme		Acetonitrile	
			wB97XD	CCSD
1a	ZEE	N_A	**0.00**	**0.00**
1b		N_B	41.73	36.19
1c		O	23.71	28.80
2a	ZEZ	N_A	18.62	15.56
2b		N_B	42.03	34.96
2c		O	40.75	43.40
3a	ZZE	N_A	11.77	8.59
3b		N_B	67.97	62.54

1a (0.00)   1b (36.19)   1c (28.80)

2a (15.56)   2b (34.96)   2c (43.40)

3a (8.59)   3b (62.54)
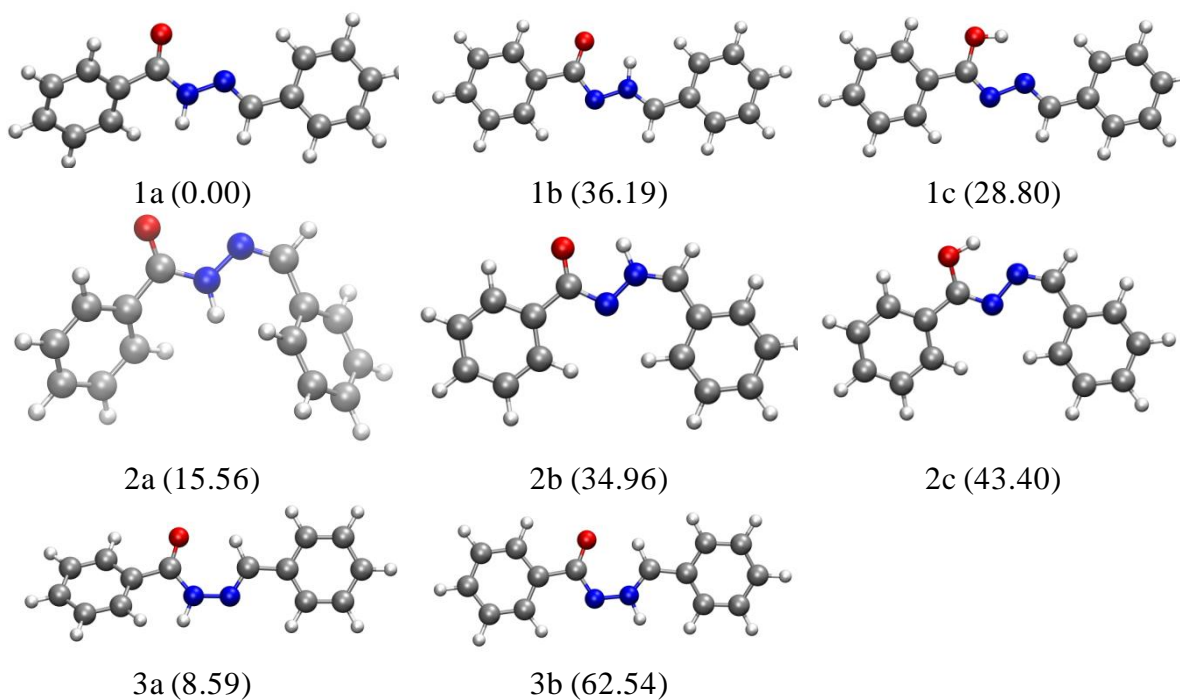
Figure D-1 The structures and the relative Gibbs energies (in kJ/mol) of AY in ACN

The vertical-excitation UV-Vis absorption spectra of the AY isomers as calculated at the ωB97XD and DLPNO-STEOM-CCSD levels of theory are summarized in Figure D-2. The STEOM-CCSD spectra are simulated with Gaussian peaks having line width of 500 cm^{-1}. The most intense band of most AY isomers originates from the $S_0 \rightarrow S_1$ or $S_0 \rightarrow S_2$ transitions, except for the 2a isomer, whose primary band system is associated with the $S_0 \rightarrow S_3$ electronic transition. The secondary band of these isomers is generally less than one third the intensity of the primary bands and thus has little contribution to the spectra within the region of interest. One notable observation from the calculated spectra is that the band positions are more strongly depend on the protonation scheme than on the skeletal conformation. The excitations for the N_A-protonated isomers usually fall in the region of 240 - 280 nm (4.4 - 5.1 eV), whereas those of the O-protonated isomers within 280-300 nm (4.1 - 4.4 eV) region. The N_B-protonated isomers yield bands at relatively long wavelength (*ca.* 350 nm, 3.5 eV).

So far, the calculated vertical excitations disagree with the work of Van Dijken *et al.*[13] in the peak assignment. As is shown in Figure D-3, the experimental λ_{max} for 1a and 2a are found at 297 nm and 388/365 nm, respectively. Considering that the UV measurements are carried out with AY acetonitrile solution at room temperature, only the 1a and the 3a isomers are of importance according to the isomer energy profile. The primary band of 1a and 3a are located at 276 nm and 279 nm, respectively, which agrees well with the experimental λ_{max} for the E form at 297 nm. Note that although 1c and 2c give better match in the primary band position, they are not chosen the assignment structure due to their high relative energy. More interestingly, the λ_{max} at 388 nm, which is assigned to 2a in literature,[13] best matches the spectra of the N_B-protonated isomers, including 1b, 2b, and 3b, while the theoretical λ_{max} for 2a is at *ca.* 250 nm, a shorter wavelength than the λ_{max} of the 1a. This assignment explains the disproportional absorbance change during the photo-isomerization experiments considering the similarity of the Z/E isomers in their structure and thus in their molar absorptivity at λ_{max}. In conclusion, the literature λ_{max} for the E isomers may be a misassignment.
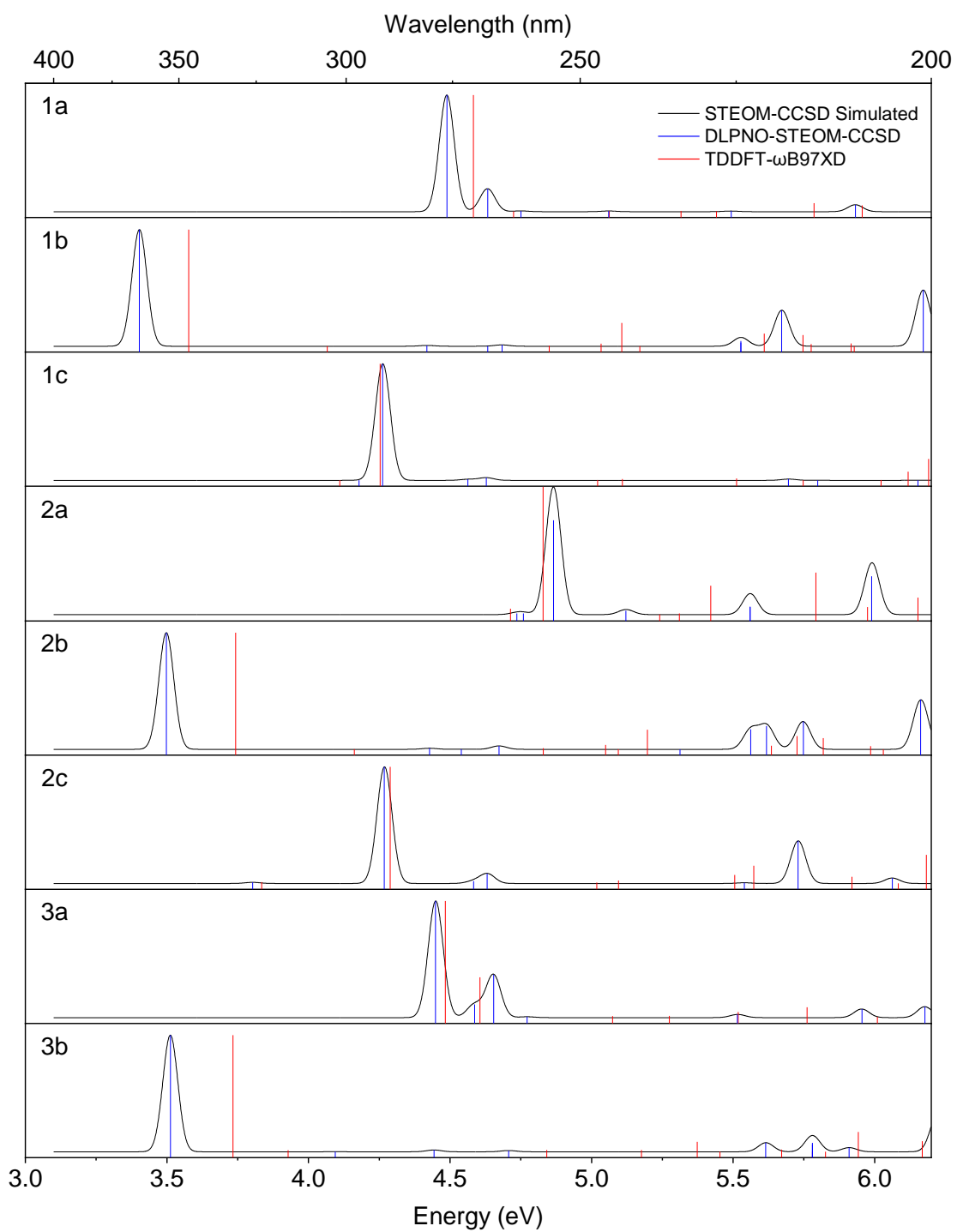
Figure D-2 The vertical excitation spectra of the AY isomers
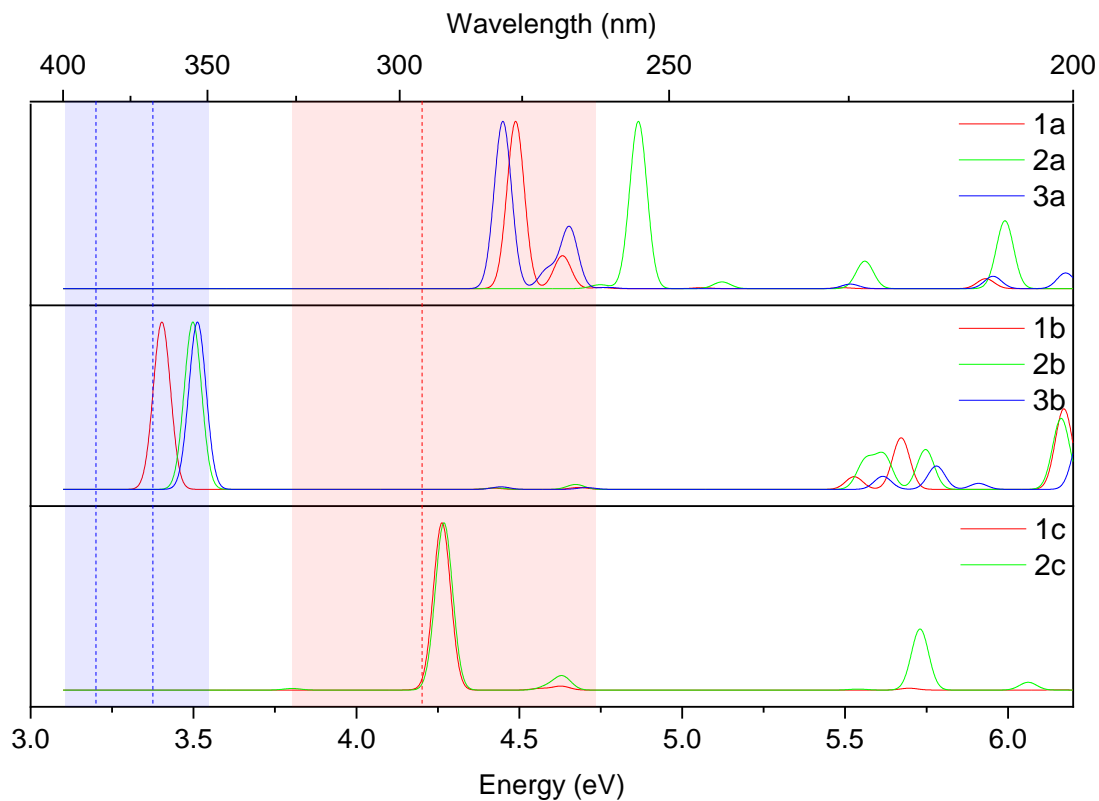
Figure D-3 Vertical excitation UV-Vis spectra of the neutral AY isomers, as is ported from Figure 15. Experimental UV-Vis spectra of AY in acetonitrile by Van Dijken et al.[13] is labelled in colored squares (for the estimated range of the peak) and dashed lines (for peak maximum). Red elements correspond to the primary band feature before UV induced isomerization while the blue ones correspond to the new feature that arise after isomerization.
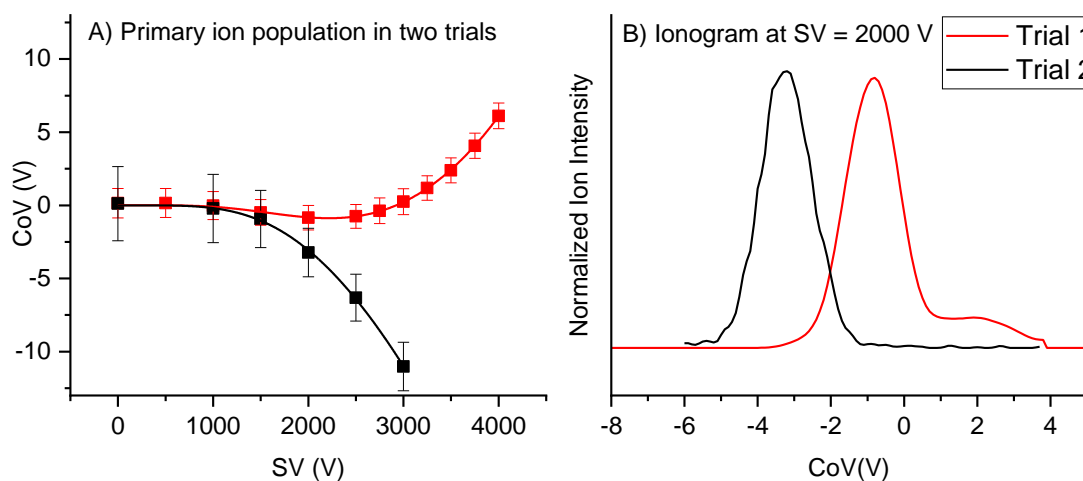
Appendix E

Miscellaneous Items for Chapter 5

a) The flow rates of the modifiers employed in the DMS separation

Modifier	MeOH	H2O	ACN
Flow Rate (µL/min)	137	61	177

b) The discrepancy observed in the water modifier resolved species in two different DMS runs.

Figure E-1. Discrepancy in the ion population under 1.5% water modifier

c) Details of Laser data extraction.

Table E-6-1 Summary of Laser and Data Extraction Setup

Modifier	DMS setup SV(CV) (V)	Max. Laser Pulse Power (mJ)	Fragments taken account (m/z)
None (N$_2$)	0 (0)	2.2	77.5, 79.5, 104.4, 105.5,122.5,147.7, 206.9
	2750 (+1.9)	0.9	77.5, 79.5, 104.4, 105.5,122.5,147.7, 206.9
	2750 (+3.8)	0.9	77.5, 79.5, 104.4, 105.5,122.5,147.7, 206.9
MeOH	0 (0)	2.2	77.5, 79.5, 104.4, 105.5,122.5,147.7, 206.9
	2000 (-6.3)	3.4	104.5,105.5
	2000 (+0.9)	2.2	104.5, 105.5, 147.6

Appendix F

Default Qtrap 5500 Instrument Parameters in the Acylhydrazone Photoswitch Experiments.

Source/Gas	
Curtain Gas (CUR)	20.0
Collision Gas	Medium
IonSpray Voltage (IS)	5000
Ion Source Gas 1 (GS1)	20
Ion Source Gas 2 (GS2)	0
Temperature (TEM)	0 (Room temperature)
Compound	
Declustering Potential (DP)	100
Entrance Potential (EP)	10
Collision Energy (CE)	10
DMS	
DMS Temperature (TEM)	Low (150°C)
Modifier	None
Separation Voltage (SV)	0
Compensation Voltage (CV)	0
DMS offset (DMO)	-3
DMS Resolution Enhancement (DR)	off
Resolution	
Ion Energy 1 (IE1)	1.0
Detector	
CEM	1900
Mode specific parameters	
Enhanced Product Ion (EPI) for DMS	
Sum of Scan	1
Parent mass	225.2
Scan mass range	70.0 – 230.0
Enhanced Multi Charge (EMC) for photodissociation	
Sum of Scan	1-3 depends on ion intensity
LIT Time	1-5ms depends on ion intensity
Q3 Entry Barrier	3V
Q3 Empty Time	500 ms