

# Physicochemical properties of isolated biomolecule clusters

by

Ahdia Anwar

A thesis  
presented to the University of Waterloo  
in fulfillment of the  
thesis requirements for the degree of  
Master of Science  
in  
Chemistry

Waterloo, Ontario, Canada, 2017

© Ahdia Anwar 2017

## **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

I would like to acknowledge the contributions of my co-authors to the research described in this thesis.

Chapter 4 is comprised of a manuscript that is ready to be submitted. It is titled: *Differential Mobility Spectrometry and Machine Learning Predict Chemical Reaction Rates*. The authors, in order as they appear on the manuscript, are as follows:

Ahdia Anwar  
Jarrod Psutka  
Stephen W.C. Walker  
Gilles Goetz  
John S. Janiszewski  
J. Larry Campbell  
W. Scott Hopkins

Chapter 5 is comprised of a recently accepted manuscript in the *International Journal of Mass Spectrometry*. It is titled: *Separating and Probing Tautomers of Protonated Nucleobases using Differential Mobility Spectrometry*. The authors, in order as they appear on the manuscript, are as follows:

Ahdia Anwar  
Jarrod Psutka  
Stephen W.C. Walker  
Thorsten Dieckmann  
John S. Janiszewski  
J. Larry Campbell  
W. Scott Hopkins

## **Abstract**

An investigation of the physicochemical properties of isolated biomolecule clusters is herein described. Three distinct subprojects examine (a) the role several complex organic ligands may play in the promotion or disruption of guanine quadruplex structure, (b) the solvent clustering behavior of biologically and pharmaceutically relevant ions, and (c) the differential mobility spectrometry behavior of protonated nucleobase tautomers. A joint computational and experimental approach has been taken. Experimental results have been obtained through techniques such as infrared multiple photon dissociation (IRMPD) spectroscopy, differential mobility spectrometry (DMS), and circular dichroism (CD) spectroscopy. Support for these experimental outcomes are provided by computations involving molecular dynamics simulations and high-level quantum mechanical calculations. Ultimately, findings of the research conducted will not only impact drug discovery, but will also provide invaluable information that will aid in the development of a fundamental description of DMS.

## Acknowledgments

First and foremost, I would like to thank my supervisor, Dr. Scott Hopkins. I am extremely grateful for the support, guidance, and opportunities he has provided, as well as the immense amount of patience he has shown over the last several years. It truly has been a wonderful experience being a part of the Hopkins group.

I would also like to thank the members of my committee, Dr. German Sciaini and Dr. Thorsten Dieckmann, for taking the time to critique my work. All comments and suggestions are a useful part of my learning process. I am also very appreciative of Dr. Terry McMahon for being a part of my proposal committee.

I am very appreciative of Dr. Valerie Gabelica of the Université de Bordeaux for allowing me the opportunity to conduct research in her laboratory. I'd also like to thank her graduate student Adrien Marchand for teaching me the research methods used at the Institut Européen de Chimie et Biologie facility.

Thank you to the current and former members of the Hopkins Lab for making the lab such an enjoyable place to work. Special thanks to Dr. Steve Walker, Jarrod Psutka, and Mike Lecours for their efforts towards our research collaborations.

Lastly, I would like to thank my family and friends for the encouragement they've given me. Without them, I would not be where I am today.

## Table of Contents

Author's Declaration .....	ii
Statement of Contributions .....	iii
Abstract .....	iv
Acknowledgments.....	v
List of Figures.....	viii
List of Tables .....	xi
List of Abbreviations.....	xii
1 Introduction.....	1
2 Methodology.....	4
2.1 Introduction to Computational Methods .....	4
2.1.1 Molecular Mechanics .....	4
2.1.2 Density Functional Theory .....	5
2.1.3 Basin Hopping .....	6
2.2 Introduction to Experimental Methods.....	8
2.2.1 Infrared Multiple Photon Dissociation (IRMPD).....	8
2.2.2 Circular Dichroism.....	11
2.2.3 Differential Mobility Spectrometry .....	14
3 G-Quadruplexes.....	16
3.1 Introduction.....	16
3.2 Methodology.....	17
3.2.1 Computational .....	18
3.2.2 Experimental.....	21
3.3 Results and Discussion.....	24
3.3.1 Ligand structures .....	24
3.3.2 G-ligand interactions.....	27
3.3.3 Antiparallel Tetramolecular G-quadruplexes .....	31
3.4 Conclusions.....	37
4 Pharmaceutically Relevant Ions .....	39
4.1 Introduction.....	39
4.2 Methodology.....	41
4.3 Results and Discussion.....	42
4.4 Conclusions.....	47

5 Nucleobase Tautomers.....	48
5.1 Introduction.....	48
5.2 Methods.....	50
5.2.1 Experimental Details.....	50
5.2.2 Computational Details.....	52
5.3 Results.....	53
5.3.1 Protonated Cytosine (C + H) <sup>+</sup> .....	53
5.3.2 (A + H) <sup>+</sup> , (T + H) <sup>+</sup> , (U + H) <sup>+</sup> , and (G + H) <sup>+</sup> .....	59
5.4 Conclusions.....	64
6 Concluding Remarks.....	66
References.....	69
Appendix I: G-Quadruplexes – Supporting Information.....	77
Structural Formulae of HK21 Ligand.....	77
XYZ Coordinates of Optimized Geometries.....	77
NMR Spectra.....	83
Appendix II: Pharmaceutically Relevant Ions – Supporting Information.....	84
Structural Formulae.....	84
Calculation Results.....	85
XYZ Coordinates of Optimized Geometries.....	86
Dispersion Plots.....	95
Appendix III: Nucleobase Tautomers – Supporting Information.....	104

## List of Figures

Figure 2.1 Flow chart of basin hopping algorithm.....	7
Figure 2.2 Vibrational energy levels of a fictional species void of anharmonic effects (left) and a species displaying anharmonicity (right). .....	9
Figure 2.3 Model of IR multiphoton absorption. ....	10
Figure 2.4 Schematic diagram of free electron laser. <sup>38</sup> .....	11
Figure 2.5 Linear (left) and circular (right) polarized light.....	12
Figure 2.6 Schematic of CD spectrophotometer. ....	13
Figure 2.7 Schematic diagram of differential mobility spectrometer. Adapted from reference <sup>47</sup> . .....	14
Figure 2.8 Possible exhibited DMS behaviors. Adapted from reference <sup>47</sup> . ....	15
Figure 3.1 (Left) Structure of cation-stabilized G-tetrad, and (right) structure of cation-stabilized G-quadruplex.....	16
Figure 3.2 Structural formulae of ligands examined in the computational portion of this investigation (a) 360A (2-N,6-N-bis(1-methylquinolin-1-ium-3-yl)pyridine-2,6-dicarboxamide), (b) PhenDC3 (3,3'-[1,10-phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methyl-Quinolinium]), (c) PDS (4-(2-aminoethoxy)-N2,N6-bis[4-(2-aminoethoxy)-2-quinolinyl]-2,6-pyridinedicarboxamide), (d) TMPyP4 (5,10,15,20-tetra-(N-methyl-4-pyridyl)porphyrin), (e) TrisQ (triazoniatrinanaphthylene). ....	19
Figure 3.3 Basin hopping study of PhenDC3-G monomer clusters. Here, 10,000 BH steps identified 14 unique candidate geometries. ....	20
Figure 3.4 The three lowest-energy isomers of the PhenDC3 ligand. Relative Gibbs' energies as calculated using B3LYP/6-311++G(d,p) level of theory at 298K are provided in kJ/mol.....	25
Figure 3.5 Comparison of the experimental IRMPD spectrum of PhenDC3 to the calculated IR spectra of (A) the global minimum and the next two lowest-energy isomers with relative Gibbs' energies of (B) 15.8 kJ/mol, and (C) 20.9 kJ/mol, respectively. Computational spectra have been calculated at the B3LYP/6-311+G(d,p) level of theory, and a DFT scaling factor of 0.9679 has been applied. ....	26
Figure 3.6 The lowest-energy conformation for each ligand clustered with guanine, as calculated at the B3LYP/6-311++G(d,p) level of theory. ....	28
Figure 3.7 Comparison of the experimental IRMPD and calculated IR spectra of (A) protonated guanosine (B) PhenDC3 ligand, and (C) protonated guanosine-PhenDC3 complex, respectively. Experimental spectra is shown in black, calculated spectra for the lowest-energy isomer of each is shown in red. Computational spectra have been calculated at the B3LYP/6-311+G(d,p) level of theory, and a DFT scaling factor of 0.9679 has been applied. ....	30
Figure 3.8 Mass spectra of solutions containing 40 $\mu\text{mol/L}$ DNA oligonucleotide (sequence denoted), 100 mmol/L TMAA, 1 mmol/L KCl, 7.5 $\mu\text{mol/L}$ PhenDC3, and 2.5 $\mu\text{mol/L}$ T6.....	32
Figure 3.9 CD spectra of solutions containing 40 $\mu\text{mol/L}$ DNA oligonucleotide, 100 mmol/L TMAA, 1 mmol/L KCl, and 7.5 $\mu\text{mol/L}$ PhenDC3.....	33



Figure 3.10 CD spectra of solutions containing 40 $\mu\text{mol/L}$ GGGTTA DNA oligonucleotide, 100 $\text{mmol/L}$ TMAA, 1 $\text{mmol/L}$ KCl, and 7.5 $\mu\text{mol/L}$ ligand. ....	35
Figure 3.11 Mass spectra of solutions containing 40 $\mu\text{mol/L}$ GGGTTA DNA oligonucleotide, 100 $\text{mmol/L}$ TMAA, 1 $\text{mmol/L}$ $\text{Sr}(\text{OAc})_2$ , 7.5 $\mu\text{mol/L}$ PhenDC3, and 2.5 $\mu\text{mol/L}$ T6. ....	36
Figure 4.1 The structural motifs of the acrylamide CRGs. ....	42
Figure 4.2 (Top panel) The correlation between calculated $D_0$ for ion-solvent binding and the predictions of the Random Forest ML model when using only DMS data as the input. (Bottom panel) The correlation between calculated $D_0$ for ion-solvent binding and the predictions of the Random Forest ML model when using DMS data and calculated collision cross section as the inputs. ....	44
Figure 4.3 (Top panel) The correlation between measured $\text{pK}_a$ and the predictions of the Random Forest ML model. (Bottom panel) The correlation between measured $\text{pK}_b$ and the predictions of the Random Forest ML model. Both fits used DMS data and calculated collision cross section as the inputs. ....	45
Figure 4.4 (Top panel) The correlation between measured $\text{LogD}$ and the predictions of the Random Forest ML model. (Bottom panel) The correlation between measured $t_{1/2}$ for reaction with GSH and the predictions of the Random Forest ML model. Both fits used DMS data and calculated collision cross section as the inputs. ....	46
Figure 5.1 A schematic diagram of the DMS cell and the QJet region of the mass spectrometer. Adapted from reference <sup>44</sup> . ....	52
Figure 5.2 The dispersion plot obtained for $(\text{C} + \text{H})^+$ ( $m/z$ 112) with a DMS cell containing (A) a pure $\text{N}_2$ environment, and a $\text{N}_2$ environment seeded with 1.5% (mole ratio) (B) methanol vapor, and (C) isopropyl alcohol vapor. Error bars are $2\sigma$ obtained from Gaussian fits to the ionogram peaks. (D) The ionogram recorded for the $m/z$ 112 peak in a pure $\text{N}_2$ environment with $\text{SV} = 3500$ V (highlighted green in A). (Inset) The three lowest energy tautomers of $(\text{C} + \text{H})^+$ as calculated at the $\text{CCSD}(\text{T})/6\text{-}311++\text{G}(\text{d},\text{p})//\text{B3LYP}/6\text{-}311++\text{G}(\text{d},\text{p})$ level of theory. Energies are reported as standard Gibbs' energies in $\text{kJ mol}^{-1}$ . ....	55
Figure 5.3 The breakdown curves obtained for $(\text{C} + \text{H})^+$ ( $m/z$ 112) when isolating on the ionogram peaks plotted in Figure 5.2D. (A) peak I ( $\text{CV} = -6$ V), (B) peak II ( $\text{CV} = 1$ V), and (C) peak III ( $\text{CV} = 4$ V). Collision energy was increased in 2.5 V increments from 0 to 50 V. Ion kinetic energy is calculated with respect to the center of mass frame for the collision partners. ....	55
Figure 5.4 (A) The ionogram recorded when gating on $(\text{C} + \text{H})^+$ ( $m/z$ 112). The results of HDX experiments when isolating on (B) peak I, (C) peak II, and (D) peak III. The black traces show the observed mass distributions in the absence of HDX reagent. The blue traces show the effect of introducing a low vapor pressure of $\text{D}_2\text{O}$ , and the red traces are observed following HDX in $\text{N}_2$ at 18 $^\circ\text{C}$ seeded with a saturated partial pressure of $\text{D}_2\text{O}$ . ....	56
Figure 5.5 The ionogram recorded for $\text{C}\cdot\text{H}^+$ ( $m/z$ 112) in a pure $\text{N}_2$ environment at $\text{SV} = 3000$ V as the declustering potential, DP, is stepped from 0–300 V in 50 V increments. ....	58
Figure 5.6 The lowest energy proton-bound cytosine-methanol clusters, $(\text{C} + \text{H})^+\cdot\text{MeOH}$ and $(\text{C} + \text{H})^+\cdot(\text{MeOH})_2$ . Proton transfer along the intermolecular hydrogen-bond network could	

facilitate interconversion of the two lowest energy isomers of  $(C + H)^+$ . Standard Gibbs' energies were calculated at the B3LYP/6-311++G(d,p) level of theory..... 59

Figure 5.7 The dispersion plots obtained for (A)  $(T + H)^+$  (m/z 117), (B)  $(U + H)^+$  (m/z 113), (C)  $(A + H)^+$  (m/z 136), and (D)  $(G + H)^+$  (m/z 152) for a pure  $N_2$  environment with DP set to 150 V. Curves that are greyed-out are associated with larger clusters which fragment to produce the ion of interest post-DMS. The numeric labels indicate the tautomer associated with a particular dispersion plot (see Figure 5.2)..... 61

Figure 5.8 The four lowest energy tautomeric forms of  $(C + H)^+$ ,  $(T + H)^+$ ,  $(U + H)^+$ ,  $(A + H)^+$ , and  $(G + H)^+$ . Electronic energies were calculated at the CCSD(T)/6-311++G(d,p) level of theory and thermochemical corrections were calculated at the B3LYP/6-311++G(d,p) level of theory. Standard Gibbs' energies are reported in  $\text{kJ mol}^{-1}$ ..... 63

## List of Tables

Table 3.1 Oligonucleotide-ligand solution compositions.....	23
Table 3.2. G-quadruplex-ligand binding energies of the lowest-energy cluster conformations, as calculated at the B3LYP/6-311++G(d,p) level of theory.....	29
Table 5.1 The percentage of the ensemble population of the three lowest energy tautomeric forms of (C + H) <sup>+</sup> , (T + H) <sup>+</sup> , (U + H) <sup>+</sup> , (A + H) <sup>+</sup> , and (G + H) <sup>+</sup> . Experimental populations were determined by Gaussian fits of the ionogram distributions recorded at DP = 0 V and at DP = 150 V. Errors (1σ) are reported in parentheses. The calculated population percentages are based on relative standard Gibbs' energies which were calculated at the CCSD(T)//B3LYP level of theory employing a 6-311++G(d,p) basis set. ....	64

## List of Abbreviations

B3LYP	Becke, three parameter, Lee-Yang-Parr
BE	Binding energy
BH	Basin hopping
CCS	Collision cross section
CD	Circular dichroism
CHELPG	Charges from electrostatic potentials using a grid-based method
CID	Collision-induced dissociation
CLIO	Centre Laser Infrarouge D'Orsay
CPL	Circularly polarized light
CRG	Covalent reactive group
CV	Compensation voltage
DFT	Density functional theory
DMS	Differential mobility spectrometry
DNA	Deoxyribonucleic acid
ESI	Electrospray ionization
FEL	Free electron laser
G	Guanine
GGA	Generalized gradient approximation
GM	Global minimum
GSH	Glutathione
HDX	Hydrogen-deuterium exchange
HF	Hartree-Fock
IECB	Institut Européen de Chimie et Biologie
IMS	Ion mobility spectrometry
IR	Infrared
IRMPD	Infrared multiphoton dissociation
IVR	Intramolecular vibrational energy redistribution
K <sup>+</sup>	Potassium cation
L-CPL	Left-handed circularly polarized light
LDA	Local density approximation
ML	Machine learning
MM	Molecular mechanics
MQOH	2-methylquinonlin-8-ol
MS	Mass spectrometry
m/z	Mass-to-charge ratio
Na <sup>+</sup>	Sodium cation
PES	Potential energy surface
QIT	Quadrupole ion trap
R-CPL	Right-handed circularly polarized light

SV	Separation voltage
TMAA	Trimethylammonium acetate
UFF	Universal force field
UV	Ultraviolet

# 1 Introduction

Atomic and molecular clusters provide a bridge between molecules, which have clearly defined quantum states, and the condensed phase, where bulk properties emerge.<sup>1</sup> Interest in the field of cluster chemistry expands beyond this function however, as unique physicochemical properties not displayed in either molecules or the bulk phase, are exhibited by clusters.<sup>2</sup> Although a number of cluster types exist, the focus on biomolecules is of particular interest, owing to their ability to model and interpret biological processes. This sub-field encompasses the study of amino acids, proteins, nucleic acids, and drug candidates. Directions of study were evaluated based on their relevance and potential impact in the field of drug discovery and development. This thesis outlines the three distinct subprojects that were chosen in pursuit of this ambition. Despite the varied nature of the topics explored, together these studies demonstrate the role cluster chemistry plays in bridging the study of small molecules and complex biological processes.

The first subproject, discussed in Chapter 3, examines the effects of a number of complex organic ligands on the promotion or disruption of guanine (G) quadruplex structure. G-quadruplexes are structures found in guanine-rich sequences of eukaryotic DNA.<sup>3</sup> Owing to their characteristic locations on the promoter region of the oncogenes, and telomeric region of eukaryotic chromosomes, their study has implications in both the fields of anti-aging, as well as cancer research.<sup>4</sup> An increasing large interest is currently being shown in the study of quadruplex-binding ligands.<sup>5-7</sup> These large organic molecules, characterized by the presence of an aromatic core and basic side chains, can be either naturally-occurring or synthetically produced.<sup>8</sup> A better understanding of the interactions of these compounds with G-quadruplexes would allow for the facilitation of rational drug design for G-quadruplex-selective binding. In this investigation, we determine the structures and properties of a number of known G-quadruplex

binding ligands, examine their interactions with individual guanine nucleosides, as well as explore their resulting effect on G-quadruplex structures. A bottom-up approach is taken here, in that data obtained regarding individual ligand molecules guides the interpretation of ligand-nucleoside interactions, and ultimately the rationalization of ligand-oligonucleotide three-dimensional topologies.

The second subproject, outlined in Chapter 4, explores the potential application of an emerging analytical technique known as differential mobility spectrometry (DMS), with mass spectrometry-based studies of pharmaceutically relevant drug-like molecules. Drugs that possess the ability to covalently modify their biological target have approval in the treatment of a number of physiological conditions.<sup>9-11</sup> In fact, nearly 30% of drugs available in the market today act via a covalent mechanism.<sup>12</sup> Advantages of these compounds over traditional, noncovalent binding drugs include greater biological efficiency, lower required dose, longer duration of action, as well as the potential to avoid the development of drug resistance.<sup>13</sup> Therefore, interest in covalent inhibitors, as well as their characterization, continues to grow. In the pursuit of improving covalent inhibitor selectivity, characteristics such as electrophilic reactivity, bioavailability, and  $pK_a$  must be evaluated.<sup>14,15</sup> In this investigation, the cluster-binding behaviors of a number of closely-related acrylamide molecules were determined by DMS, and supported by quantum mechanical calculations. Here we demonstrate that gas phase DMS clustering behaviour, when treated with a Random Forest supervised machine learning algorithm, can be used to quantitatively predict a number of properties associated with the characterization and evaluation of drug candidates. Application of this technique in the drug discovery process will aim to increase both the accuracy and efficiency of current methodologies.

Lastly, Chapter 5 discusses the use of protonated nucleobase molecules as a vehicle to further explore both the abilities and limitations of DMS as an analytical method. In this work,

DMS is used to select for individual tautomeric forms of protonated nucleobases. Following isolation, these molecules were further characterized through both hydrogen-deuterium exchange (HDX) and collision-induced dissociation (CID) experiments. Upon electrospray ionization (ESI) of the protonated nucleobase solutions, DMS ionograms showed the presence of a number of tautomeric species, albeit preference for the creation of a single form of each nucleobase was evident. These experiments demonstrated the necessity of additional characterization following DMS isolation; as a number of minor ion signals were found to be attributed to larger nucleobase-containing clusters which fragment post-DMS, rather than tautomeric forms of the bare protonated nucleobases. The computational portion of this study involved electronic structure calculations, which allowed for the assignment of ion signal peaks. Findings of this study will serve in the development of a fundamental description of differential mobility spectrometry.

An exhaustive literature review of these systems has identified key areas of study where current research methods remain insufficient. The application of a joint computational and experimental approach to the study of these clusters serves to overcome such deficiencies. Experimental techniques used in these investigations include infrared multiple photon dissociation spectroscopy, differential mobility spectrometry, and circular dichroism spectroscopy. Parallel computational work involving both high-level quantum mechanical calculations, as well as relatively low level molecular dynamics simulations, provide support and rationalization of experimental outcomes.



## 2 Methodology

### 2.1 Introduction to Computational Methods

The computational methods employed in this investigation are particularly useful in the guidance and interpretation of experiment. These molecular dynamics and quantum mechanical calculations yield an assortment of information including molecular geometries, charge distributions, harmonic vibrational frequencies, thermodynamic corrections, and cluster binding energetics. For complex cluster systems, it is necessary to conduct an exhaustive search of the potential energy surface to identify stable, chemically-relevant structures (i.e., isomers, conformers, tautomers). To do this, we incorporate the use of molecular mechanics (MM) as an optimization method in a custom Monte Carlo-type algorithm which guides the search. To refine the predictions of our MM-based search of the PES, more accurate electronic structure calculations are conducted. In general, these are undertaken at the density functional level of theory (DFT) due to the fact that this method strikes a desirable balance between accuracy and computational efficiency.

#### 2.1.1 Molecular Mechanics

Molecular mechanics (MM) is a modelling method which makes use of classical mechanics to predict the potential energies of molecular systems.<sup>16,17</sup> This technique involves treating each atom within the molecular system as a single particle, assigning each particle a particular Van der Waals radius, and treating each bond interaction as a harmonic spring.<sup>18</sup> MM can be described with the following simplified equation.

$$E^{\text{Total}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{i < j} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right] \quad (1)$$

In the above equation,  $K$  defines the force constants for bond stretching and angle bending and torsional motions.  $r$  is the bond lengths,  $\theta$  defines bond angles,  $A$  and  $B$  are both Van der Waals parameters,  $q$  defines the atomic partial charges, and  $\epsilon$  is equal to the vacuum permittivity. The first term in the equation describes the harmonic stretching or compression of each bond, while the second term describes the potential energy contribution as a result of harmonic bond angle deformations. Both of these terms are truncated from their corresponding Taylor Series expansions. The last term in the above equation amalgamates the contributions from both non-bonded terms with Lennard-Jones and Coulomb potentials; electrostatic and Van der Waals contributions are summed here.<sup>19</sup> Molecular mechanics is a coarse model for molecular properties, but it does provide a very fast means of assessing many (in our case thousands) of structures for stability and relative energy. This allows for very large molecular systems to be studied, or even thousands of iterations to be performed within several hours, thus making MM ideal for use in the scanning of a potential energy surface using the basin hopping method described below.

### 2.1.2 Density Functional Theory

Density functional theory (DFT) is an *ab initio* computational method used in the modelling of electronic structure of atoms and molecules. This quantum mechanical method provides a desirable balance between accuracy and computational cost; DFT addresses the inaccuracy associated with the Hartree-Fock (HF) method, as well as the prohibitive computational demands of post-HF methods.<sup>20</sup> By accounting for electron-correlation effects, this method improves upon HF which simply approximates these forces as an average Coulomb repulsion.<sup>21</sup> DFT bypasses the problem of solving the many-body Schrödinger equation by replacing the electronic wavefunction with a function describing the electronic density as the fundamental quantity. The objective of this method becomes the minimization of the spatially

dependent density functional.<sup>22,23</sup> Thus, the density functional approach can be summarized by the following sequence:

$$n(\mathbf{r}) \Rightarrow \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \Rightarrow v(\mathbf{r}) \quad (2)$$

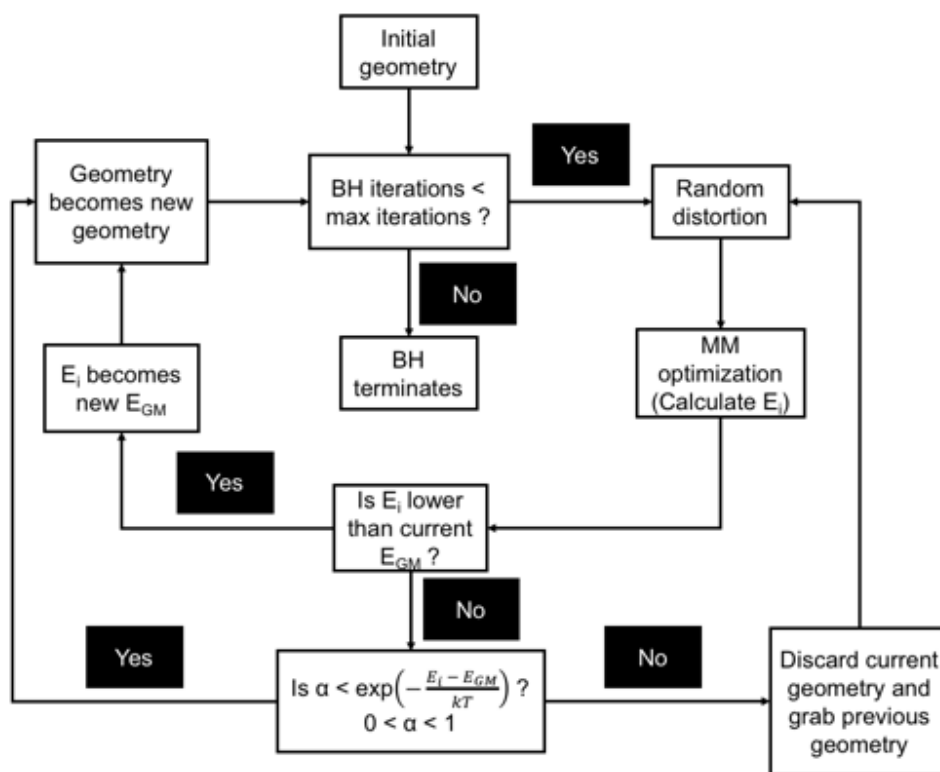
To elaborate, knowledge of the particle density,  $n(\mathbf{r})$ , implies knowledge of the wavefunction,  $\psi(\mathbf{r})$ , and the potential,  $V(\mathbf{r})$ , and hence all other observables. This is based on the fundamental concept within quantum mechanics which states that the system's wavefunction,  $\psi$ , contains all possible information about the system.

There exist three broad categories of density functional methods. Local density approximation (LDA) methods depend only on the value of the electron density at the point where the functional is evaluated. LDA operates on the assumption that the molecule is uniform, and thus extends this density value to the entire molecule.<sup>24</sup> Generalized gradient approximation (GGA) methods improve upon the description of molecules with nonhomogeneous electron densities by incorporating the gradient of the electron density as well.<sup>25</sup> Hybrid methods improve over these standard density functionals by introducing individual electronic wavefunctions to compute the total energy of a system; the accuracy of these energy calculations improves upon inclusion of the HF exact exchange component. The Becke, three parameter, Lee-Yang-Parr (B3LYP) functional, which combines GGA protocols and 20% HF exact exchange,<sup>26</sup> currently remains the most popular of the hybrid methods, and is the functional utilized within this thesis.

### 2.1.3 Basin Hopping

Basin hopping (BH) is a global optimization technique relying on the use of a stochastic-type algorithm.<sup>27</sup> BH samples the potential energy surface (PES) of a molecular system in search of its lowest energy conformation. This process involves the random displacement of

coordinates, local minimization, and subsequent acceptance or rejection of the new coordinates based on comparison to the current global minimum (GM) structure.<sup>28</sup> It is not necessary for the new optimized structure to be lower in energy than the current GM; rather, the new coordinates must simply fall within a specified thermal distribution of the GM structure. A flow chart outlining the basin hopping algorithm is provided below (Figure 2.1).



**Figure 2.1** Flow chart of basin hopping algorithm.

In Figure 2.1,  $E_i$  denotes the system's potential energy for that particular iteration,  $E_{GM}$  is the system's lowest potential energy of all iterations performed to that point in the search, and  $\alpha$  is a random number assigned a value between 0 and 1. The 'Random distortion' shown is not entirely random; a number of geometric distortion parameters (i.e. maximum translation and rotation steps) are defined as part of the BH input.

In order to perform an exhaustive sampling of a system's potential energy surface, a large number of BH iterations (typically 10,000 for the cluster sizes studied herein) must be completed. To allow for this to be done using available resources, and within a reasonable timeframe, BH exploits the low computational cost associated with the MM modelling method), BH is able to identify a comprehensive ensemble of unique structures for each system of interest. These candidate geometries can then be further optimized at higher levels of theory to yield predictions comparable to experimental data.

## **2.2 Introduction to Experimental Methods**

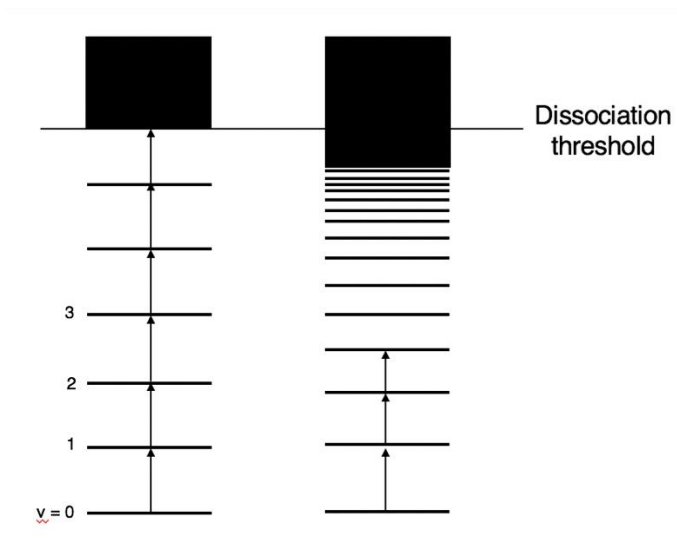
A fundamental description of the experimental techniques utilized in this investigation is provided. Each method is unique in both its underlying principles and data offered, such that a benefit is conferred when a combination of experimental approaches is taken.

### **2.2.1 Infrared Multiple Photon Dissociation (IRMPD)**

Infrared multiple photon dissociation (IRMPD) is a technique that combines mass spectrometry and laser spectroscopy whereby molecules are photo-fragmented in the gas phase as a means of structural analysis.<sup>29</sup> Although this unimolecular fragmentation can be achieved either through an increase in temperature, or through the absorption of photons in the ultraviolet (UV), visible, or infrared (IR) regions of the electromagnetic spectra, IRMPD strictly relies on the use of IR photoexcitation and dissociation.<sup>30</sup>

In order for absorption to take place, a vibrational mode must exist with the same frequency as the laser source.<sup>31</sup> Theoretically, a fictional species may exhibit successive vibrational transitions of the same frequency, and thus an intense source of photons at this resonant frequency will lead to dissociation. However, this is not the case in practice due to the fact that all molecules display anharmonicity.<sup>32</sup> In reality, a lower frequency will be observed for

each subsequent transition between vibrational energy levels, such that the energy requirement of the  $v = 0$  to  $v = 2$  transition is less than twice of that of the  $v = 0$  to  $v = 1$  transition.<sup>33</sup>



**Figure 2.2** Vibrational energy levels of a fictional species void of anharmonic effects (left) and a species displaying anharmonicity (right).

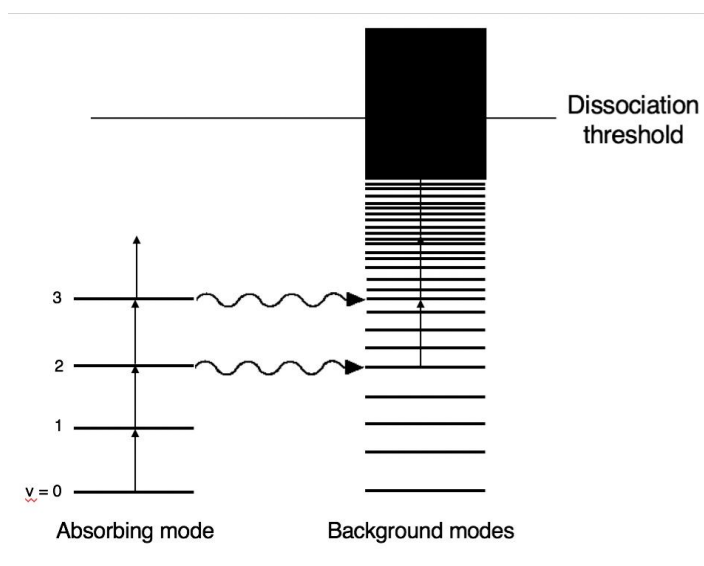
This anharmonic behaviour can be expressed by the Morse potential. This model describes the potential energy of diatomic molecules as a function of interatomic distance<sup>34</sup>:

$$V(r) = D_e(1 - e^{-a(r-r_e)})^2 \quad (3)$$

Here  $r$  is the distance between the atoms,  $r_e$  is the equilibrium bond distance,  $D_e$  is the depth of the well (sum of the zero-point and dissociation energies), and  $a$  is a parameter that controls the width of the well. The vibrational structure of these molecules can be approximated by the following expression, where  $v$  is the vibrational quantum number, and  $\omega_e$  and  $\omega_e\chi_e$  are constants directly related to the parameters of the Morse potential.<sup>35</sup>

$$E(v) = \omega_e \left( v + \frac{1}{2} \right) - \omega_e\chi_e \left( v + \frac{1}{2} \right)^2 \quad (4)$$

Absorption of the first photon is purely a vibrational transition, however all subsequent photon absorptions must rely on a combination of vibrational excitation, ro-vibrational mode excitation, anharmonic compensation, power broadening, and/ or direct multiple photon absorption.<sup>36,37</sup> Following absorption of the first few IR photons, the absorption modes of the molecule are able to mix with the high state density background modes. Favorable mixing of vibrational states at this stage allows for the depopulation of the absorbing modes and further absorption of photons in a process called intramolecular vibrational energy redistribution (IVR). Efficient IVR allows for the system to absorb photons until the dissociation threshold is reached; the molecule has sufficient energy to fragment.<sup>36,37</sup>



**Figure 2.3 Model of IR multiphoton absorption.**

For the work described here, IRMPD experiments are performed through the use of the tunable free electron laser (FEL) available at the Centre Laser Infrarouge D'Orsay (CLIO). Unlike conventional lasers which utilize excited atoms or molecules, the FEL makes use of a high energy electron beam passing through a magnetic undulator to produce light.<sup>38</sup> Retro-reflecting the axial component of the emission serves to stimulate emission and produce coherent (laser) light.

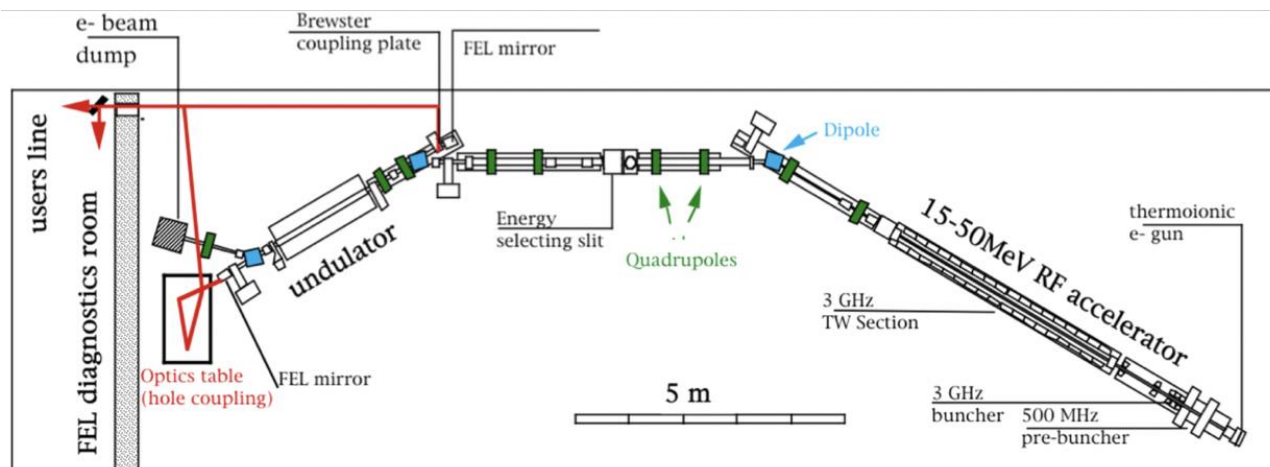


Figure 2.4 Schematic diagram of free electron laser.<sup>38</sup>

Initially, the electron beam, sourced from the thermoionic gun, is accelerated to near the speed of light through the linear accelerator. After being steered through the quadrupoles, the beam makes its way to the undulator where it oscillates through a series of alternating magnets.<sup>38</sup> The centripetal acceleration that occurs within the undulator results in the production of the necessary IR photons; these can now interact with the molecule/cluster of interest, previously ionized by electrospray ionization (ESI), and stored within the quadrupole ion trap (QIT). It is within the QIT where the IRMPD process takes place, and resulting mass spectra are obtained.<sup>38</sup>

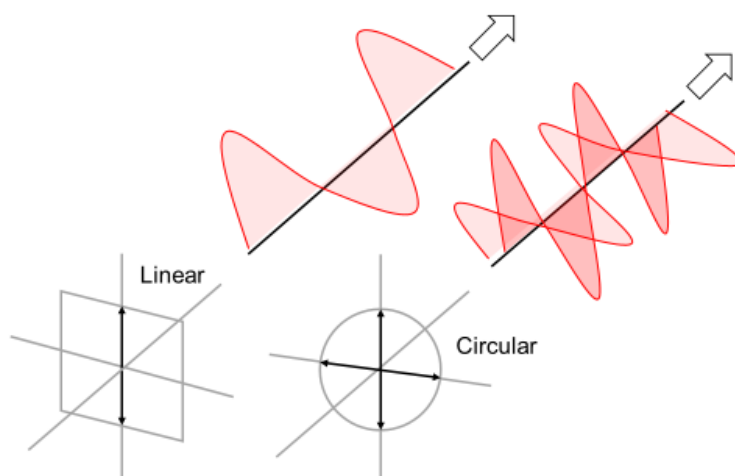
## 2.2.2 Circular Dichroism

Circular dichroism (CD) spectroscopy is a method often employed in the structural determination of large biological molecules.<sup>39</sup> CD is defined as the difference in the absorption of left-handed circularly polarised light (L-CPL) and right-handed circularly polarised light (R-CPL). It occurs only when a molecule contains one or more chiral chromophores.<sup>40</sup> CD spectroscopy involves the measurement of the CD of the molecule of interest over a range of wavelengths.

$$CD = A(\lambda)_{L-CPL} - A(\lambda)_{R-CPL} \quad (5)$$

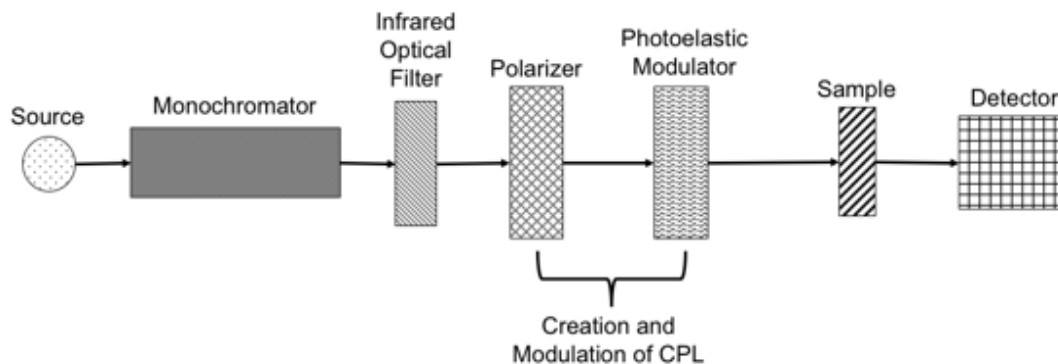


All polarized light can be described as the sum of two polarized states at a right angle to each other. When these polarized light waves of equal amplitude are in-phase, the result is linearly polarized light. If the two waves are out of phase to any degree, the resultant light is no longer linearly polarized.<sup>41</sup> Circularly polarized light (CPL) is a consequence of a helix formed when the two polarized states are a quarter wave out-of-phase.<sup>41</sup>



**Figure 2.5 Linear (left) and circular (right) polarized light.**

Biological compounds lend themselves well to study by CD spectroscopy due to their predominantly chiral nature. The main attraction of this spectroscopic method is its ability to discriminate between the various topologies of large macromolecules such as proteins, DNA, and G-quadruplexes.<sup>39</sup> This is due to the fact that the CD spectra of these compounds is largely dependent on their 3-dimensional structures rather than composed of a sum of their individual chiral components. Applications of this method include the study of structural changes as a result of various factors including temperature, pH, and introduction of ligands of interest.<sup>42</sup>



**Figure 2.6 Schematic of CD spectrophotometer.**

CD experiments are performed through the use of a spectrometer, a schematic of which is provided in Figure 2.6. Initially, light from a broad emission source is passed through a monochromator which allows for the selection of a desired wavelength of light. The selected linearly polarized monochromatic light then passes through a filter, followed by passage through a polarizer and photo elastic modulator. It is within these last two components where the creation and modulation of circularly polarized light takes place. Finally, the CPL is allowed to pass through the sample, and a signal is recorded on the detector. In the absence of an optically-active molecule, neither L-CPL nor R-CPL will be absorbed preferentially, thus only a steady output will be measured by the detector. A chiral molecule is necessary in order to yield any interpretable results.<sup>39</sup>

Although the use of CD spectroscopy confers many benefits including speed of acquisition and the ability to differentiate various topologies, a drawback of this technique lies in the fact that the results are largely qualitative. Typically, the detected signal from a CD spectrometer must be compared to a previously-constructed library of reference topologies in order to perform any data interpretation.<sup>39</sup> An example of this includes the reference library used in the interpretation of results of the G-quadruplex subproject.<sup>43</sup>

### 2.2.3 Differential Mobility Spectrometry

Ion mobility spectrometry (IMS) widely refers to the identification or characterization of ions by some property of their transport through a medium under the influence of electromagnetic fields.<sup>44-46</sup> Unlike traditional IMS methods, differential mobility spectrometry (DMS) is dependent on the change in ion transport as a function of electric field at atmospheric pressure.<sup>47</sup>

DMS involves the transport of sample ions by flow of a carrier gas between two parallel plate electrodes. These electrodes apply an asymmetric electric field perpendicular to the flow of the transport gas whose amplitude in one polarity is referred to as the separation voltage (SV).<sup>48</sup> This modulated asymmetric waveform leads to the “zigzagging” motion of the sample ions as they move toward the channel’s exit. In order to correct ion trajectories, a counterbalancing compensation voltage (CV) is applied within the DMS cell.<sup>49</sup> Ions that experience a net zero voltage will pass through the channel and be detected, whereas all others will migrate towards an electrode and subsequently be neutralized. Therefore, differences in ion mobility in high and low electric fields can be described by the CV required for trajectory correction at a particular SV.<sup>50</sup>

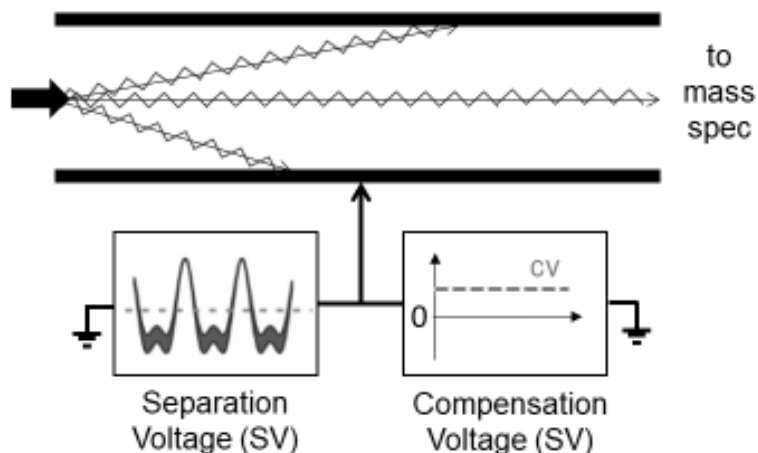
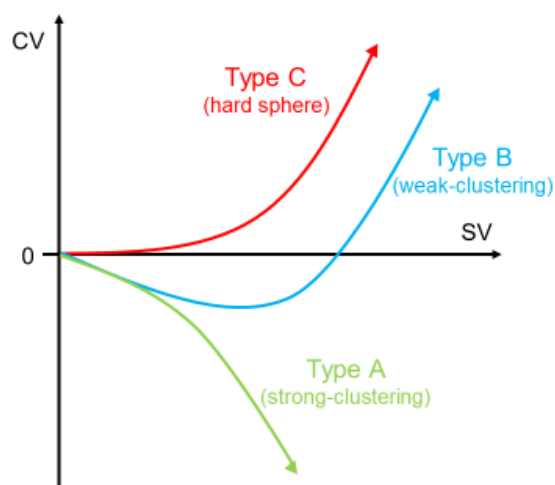


Figure 2.7 Schematic diagram of differential mobility spectrometer. Adapted from reference 47.

The addition of a volatile solvent to the DMS carrier gas has demonstrated the ability to alter ion mobility further.<sup>44,51,52</sup> This effect arises as a result of dynamic clustering and de-clustering within the DMS environment. Ion solvation takes place under the low-field portion of the SV waveform, while the high-field portion lends itself to spontaneous de-clustering.<sup>52</sup> It is this dynamic process that allows for the separation of isomers with even minor variations in solvent binding energies.



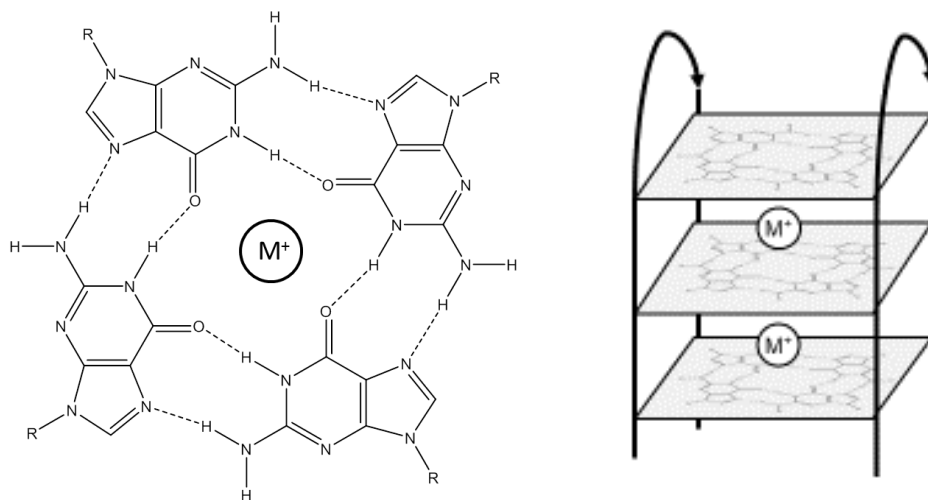
**Figure 2.8 Possible exhibited DMS behaviors. Adapted from reference 47.**

In order to visualize and interpret DMS behavior, the CV at which the maximum ion transmission takes place, is graphed as a function of SV, in what is referred to as a dispersion plot.<sup>47</sup> As illustrated in Figure 2.8, a number of prototypical DMS behaviors can be observed. In Type A, CV is seen to decrease with an increase in SV; this comes as a result of strong clustering between cations and the solvent modifier.<sup>44,47</sup> Weak-clustering ions exhibit Type B behaviour, whereby an initial decrease in CV is observed as SV increases, until an extremum is reached, followed by an increase in CV as SV is increased. Lastly, Type C describes DMS behaviour whereby no clustering takes place, and subsequently only a positive relationship between CV and SV is observed.<sup>44,47</sup>

## 3 G-Quadruplexes

### 3.1 Introduction

Guanine (G) quadruplexes are comprised of two or more G-tetrads, and are often found in guanine-rich sequences of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).<sup>53</sup> These planar, stacked structures are formed by Hoogsteen hydrogen bonding of guanine bases, and are further stabilized by the presence of a monovalent metal ion.<sup>54</sup> By occupying the central cavity between stacked tetrads, the metal cation is able to neutralize the electrostatic repulsion of the inward-facing O6 oxygens, ultimately stabilizing the G-quadruplex. Owing to their respective ionic radii, the potassium ion ( $K^+$ ) promotes the formation of the most stable G-quadruplexes,<sup>55</sup> whereas the sodium ion ( $Na^+$ ) has been shown to result in the most stable G-tetrad.<sup>56</sup>



**Figure 3.1 (Left) Structure of cation-stabilized G-tetrad, and (right) structure of cation-stabilized G-quadruplex.**

G-quadruplex conformation is influenced by both the DNA sequence as well as the folding reaction conditions, thus giving rise to extensive structural polymorphism.<sup>53</sup> These structures may be formed from one (intramolecular) or two or more (intermolecular) DNA strands, and can be further described in terms of the strand orientations (parallel, antiparallel, or hybrid).<sup>4,53</sup> It is difficult to predict which combinations of DNA and folding conditions will result in a particular structure, therefore each combination must be characterized empirically. The

formation of G-quadruplexes has been shown to be thermodynamically favorable under physiological conditions.<sup>4</sup> Potential sites of quadruplex formation have been identified in G-rich sequences of eukaryotic telomeres, as well as recently in promoter regions of non-telomeric genomic DNA.<sup>57</sup>

These structures have been implicated in a number of biological processes, including the inhibition of telomerase activity and the ability to control gene expression.<sup>6</sup> The telomerase inhibition function is of particular interest, as this ribonucleoprotein is active in 80 to 85% of cancer cells. Activation of this enzyme within certain cells essentially confers immortality, a largely undesirable trait when associated with cancer. Limiting the proliferative capacity of malignant cells through telomerase repression is a prominent focus of current research in oncology.<sup>5,6</sup>

A number of studies have established the potential of G-quadruplex structures as anticancer drug targets.<sup>5,7,8</sup> For example, several cationic porphyrins have demonstrated the ability to inhibit telomerase activity in human cancer cells by binding to, and consequently stabilizing, G-quadruplexes.<sup>8,58</sup> Although a great deal of effort has been directed toward the research and identification of potent G-quadruplex binders, the exact binding mode of these ligands remains largely unknown.<sup>59</sup> Potential applications in the field of anticancer drug development makes the study of G-quadruplex-ligand interactions promising.

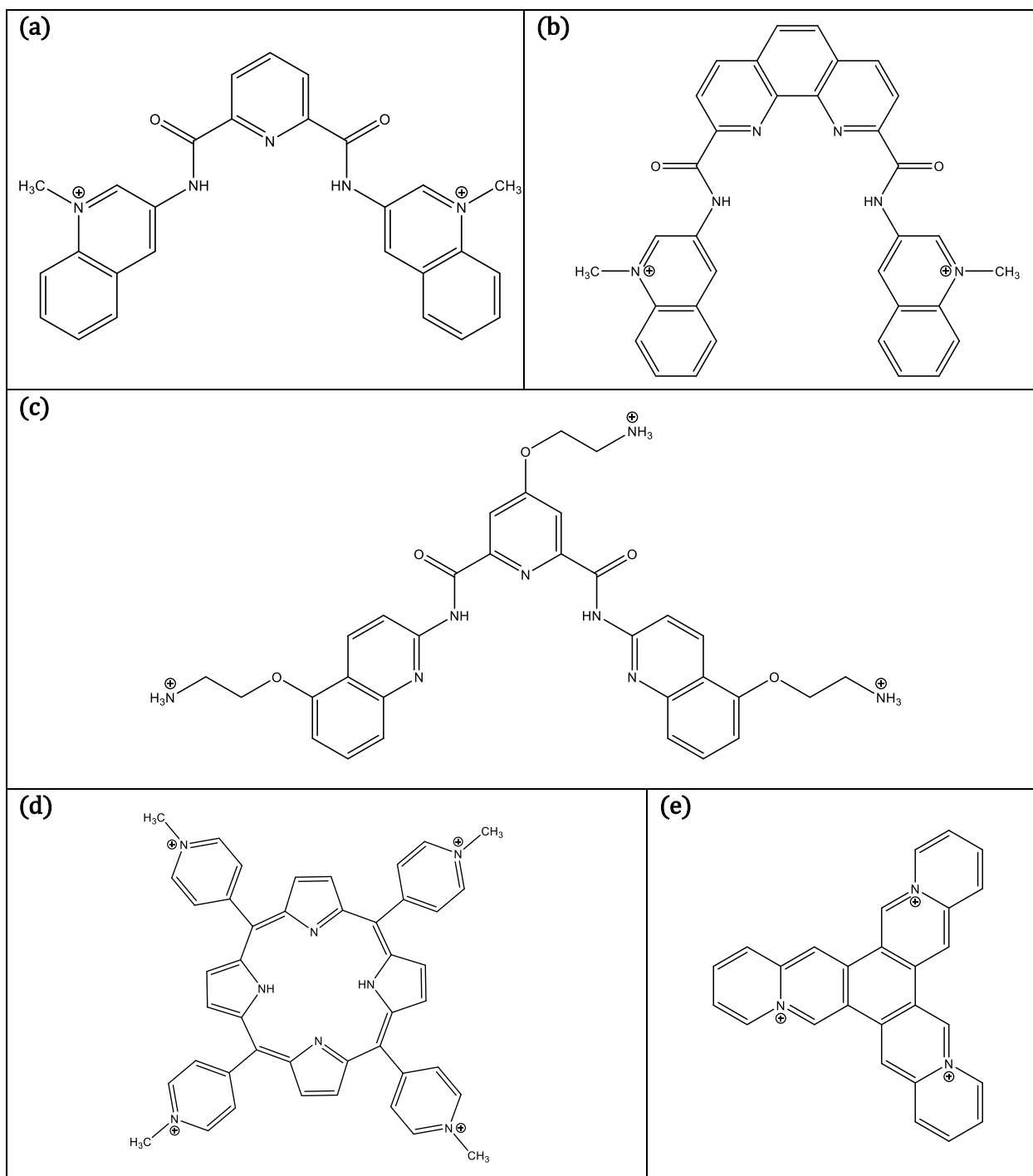
### **3.2 Methodology**

A joint computational and experimental approach was taken to study the role several organic ligands may play in the promotion or disruption of guanine quadruplex structures. The computational portion of this investigation focused on examining the specific interactions between a number of known G-quadruplex binders and guanine, at the molecular scale. This was done using a series of alternating high- and low-level calculations, which allowed for the

determination of molecular geometries, clustering interactions, and binding energies. The experimental portion relied on the use of IRMPD spectroscopy, electrospray ionization mass spectrometry (ESI-MS), and CD spectroscopy. The IRMPD experiments yielded frequency spectra of deoxy- and guanosine-ligand clusters which, when combined with computational frequency predictions, identify the nature and locations of the interactions taking place. In contrast, the MS and CD experiments were performed on solutions of ligands with oligonucleotides, to yield information regarding quadruplex formation at the macromolecular scale. Details of both computational and experimental methodologies are outlined below.

### **3.2.1 Computational**

The computational portion of this investigation involved a series of molecular dynamics simulations and high-level quantum mechanical calculations, all of which were performed with the Gaussian09 program.<sup>60</sup> The process of alternating between high- and low-level calculations allows for both the sampling of the potential energy surface (PES), as well as acquiring accurate electronic energies. Although the ultimate goal of this study focuses on the interaction of these complex ligands with G-quadruplexes, a bottom-up approach was taken in order to better predict the resulting physicochemical properties. This approach involves calculations examining the ligands of interest<sup>61-64</sup> (Figure 3.2) clustered with a neutral guanine molecule.

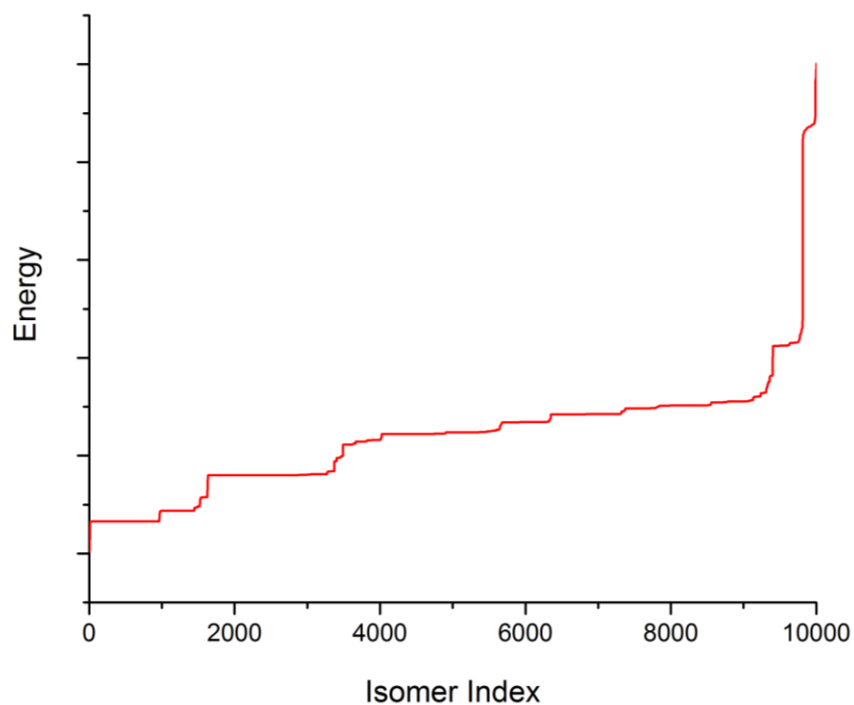


**Figure 3.2** Structural formulae of ligands examined in the computational portion of this investigation (a) 360A (2-N,6-N-bis(1-methylquinolin-1-ium-3-yl)pyridine-2,6-dicarboxamide), (b) PhenDC3 (3,3'-[1,10-phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methyl-Quinolinium]), (c) PDS (4-(2-aminoethoxy)-N2,N6-bis[4-(2-aminoethoxy)-2-quinolinyl]-2,6-pyridinedicarboxamide), (d) TMPyP4 (5,10,15,20-tetra-(N-methyl-4-pyridyl)porphyrin), (e) TrisQ (triazoniatrianaphthylene).

Initially, all five ligands and a neutral G molecule were optimized using DFT with the Becke, three parameter, Lee-Yang-Parr (B3LYP) hybrid functional, and 6-311++G(d,p) basis set.



Atomic partial charges were also calculated at this stage using the charges from electrostatic potentials using a grid-based method (CHELPG). BH with the universal force field (UFF) prediction then samples the PES of all molecules whose structures are not rigidly defined; dihedral angle manipulations are explored for all ligands excluding TrisQ, which displays zero degrees of freedom with respect to its conformation. Each BH routine involves 10,000 steps and a relatively high thermal energy of 0.5 eV. Dihedral angle rotations are set at a maximum of 3° per BH step. Subsequently, analyses are performed in order to identify a much smaller number of candidate structures from the 10,000 output files yielded by each BH routine. This analysis process involves the rank-ordering of all output files by their electronic energies, and the specification of a minimum derivative value, whereby all structures that vary by an electronic energy less than this value are considered identical. In Figure 3.3, plateau regions depict a large number of files with essentially identical structures and electronic energies.



**Figure 3.3 Basin hopping study of PhenDC3-G monomer clusters. Here, 10,000 BH steps identified 14 unique candidate geometries.**

Next, DFT optimization of the candidate structures takes place using the B3LYP functional and 6-311++G(d,p) basis set, and the lowest energy structure is identified. The next stage of BH involves the sampling of each G-ligand cluster PES in five individual routines. BH analyses are once again performed in the determination of unique structures. Finally, the optimization of all candidate geometries at the DFT/B3LYP/6-311++G(d,p) level of theory completes the computational process.

### **3.2.2 Experimental**

The experimental portion of this investigation involved the use of IRMPD spectroscopy at the Centre Laser Infrarouge D'Orsay (CLIO) in Orsay, France, as well as ESI-MS and CD spectroscopy experiments performed at the Institut Européen de Chimie et Biologie (IECB) in Bordeaux, France. IRMPD yielded frequency spectra of the PhenDC3 and 360A ligands individually, as well as those of their respective clusters with guanosine and deoxyguanosine. In contrast, the MS and CD experiments explored the resulting three-dimensional G-quadruplex topologies as a number of parameters were manipulated in oligonucleotide-ligand solutions.

#### **CLIO Experiments**

IRMPD spectra were recorded using the free electron laser (FEL) at the University of Paris' CLIO facility.<sup>38</sup> Aqueous solutions were prepared in (50/50 vol%) methanol/water with 0.1% formic acid, at concentrations of approximately 100  $\mu\text{mol/L}$ . Solid guanosine and deoxyguanosine (Sigma Aldrich), mixed with stoichiometric quantities of 360A (iodide salt) and PhenDC3 (methylsulfonate salt) ligands (provided by Marie-Paule Tellaude-Fichou at the University of Paris), were used without further purification. Solutions were continuously injected at 100  $\mu\text{L/hour}$  into a Bruker 3000+ quadrupole ion trap mass spectrometer using an electrospray ionization source (Agilent) in positive ion mode. The gas-phase ions were then mass-selected and subsequently irradiated by the FEL over a range of 800-1800  $\text{cm}^{-1}$ . Vibrational

spectra were generated by monitoring the depletion of the parent ion signal, and corresponding daughter ion signal enhancement. The reported spectra illustrate IRMPD efficiencies ( $-\log(I_{parent})/(I_{parent} + I_{fragments})$ ) as a function of photon wavenumber.

## **IECB Experiments**

Oligonucleotides were purchased from Eurogentec in reverse-phase purified lyophilized form, and subsequently dissolved in nuclease-free water (Ambion, Life technologies SAS). Seven DNA sequences were selected for study based on their presence in, or relevance to, the human telomeric sequence: GGGTA, TTAGGGT, AGGGTT, TGGGT, TAGGGT, TTAGGG, and GGGTTA (where T is thymine, and A is adenine). Ligands PhenDC3 and HK21 (structure provided in supplementary information), supplied by Marie-Paule TELAUDE-FICHOU at the University of Paris, were also dissolved in nuclease-free water.

UV absorption of the oligonucleotide stock solutions were measured at 260nm (Uvikon XS) and concentrations were subsequently determined after applying molar absorption coefficients and Cavaluzzi-Borer corrections<sup>65</sup> obtained from the Integrated DNA Technologies website. Oligonucleotides were then combined with trimethylammonium acetate (TMAA, Fluka Analytical), a salt solution (either potassium chloride or strontium acetate), ligand solution (PhenDC3 or HK21), and nuclease-free water. These solutions contained higher concentrations of the DNA sequence, salt solution, and ligand, than were ultimately measured by MS or CD. These 'folding' solutions were left undisturbed for various lengths of time, in order to allow for quadruplex folding. Herein, all results describe solutions which were left to fold for one day. Immediately preceding MS or CD experiments, 'analyzed' solutions were diluted with TMAA, TTTTTT (T6) oligonucleotide solution, and nuclease-free water, to prepare the 'analyzed' solutions. Salt solutions were included for the purpose of introducing a quadruplex-stabilizing

cation, TMAA was added to maintain physiological ionic strength, and T6 is included as an internal standard. The composition of the ‘folding’ and ‘analyzed’ solutions are as follows.

**Table 3.1 Oligonucleotide-ligand solution compositions.**

	<b>‘Folding’ solution</b>	<b>‘Analyzed’ solution</b>
Single-stranded DNA	1.00 mmol/L	40.0 $\mu\text{mol/L}$
Quadruple-stranded DNA	250 $\mu\text{mol/L}$	10.0 $\mu\text{mol/L}$
TMAA	100 mmol/L	100 mmol/L
Salt solution (KCl or Sr(OAc) <sub>2</sub> )	25 mmol/L	1 mmol/L
Ligand (PhenDC3 or HK21)	750 $\mu\text{mol/L}$	30.0 $\mu\text{mol/L}$
T6	-	2.50 $\mu\text{mol/L}$

CD experiments were performed with a JASCO J-815 spectropolarimeter using a quartz cell of 2 mm path length. All reported spectra are the sum of three scans, acquired at 20°C with a scan speed of 50 nm/min and integration time of 0.5 s in the range of 240 nm to 340 nm. The CD were normalized to molar circular-dichroic absorption ( $\Delta\epsilon$ ) based on DNA concentration using the following equation:

$$\Delta\epsilon = \frac{\theta}{32980 \times c \times l} \quad (6)$$

Here  $\theta$  is the CD ellipticity in millidegrees,  $c$  is the DNA concentration in mol/L, and  $l$  is the path length in cm. Baseline spectra were recorded for solutions containing 100 mmol/L TMAA and 1 mmol/L KCl, and subsequently subtracted from oligonucleotide-ligand spectra.

All mass spectra were obtained in negative ion mode using an LCT Premier mass spectrometer (Waters). The electrospray ionization source had a set voltage of 2200 V, a desolvation temperature of 60°C, and a source pressure of 35 mBar. ‘Analyzed’ solutions shown in Table 1, were injected at a rate of 200  $\mu\text{L/h}$ , and the resulting ion signal intensities with a mass-to-charge ( $m/z$ ) ratio between 800-2400 were recorded. Interpretation of the spectra requires determining the charge states of the ion signals, according to the following equation:

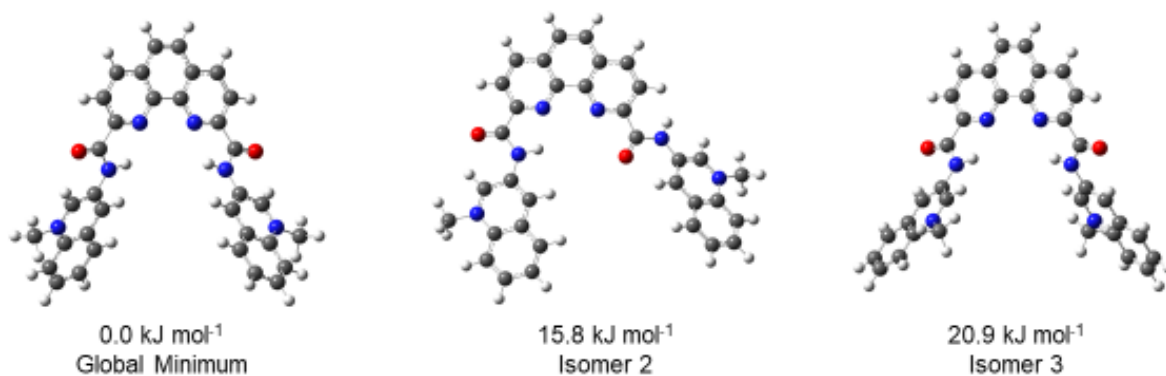
$$|z| = \frac{1}{|\Delta(\frac{m}{z})|} \quad (7)$$

Here the charge state of the ion signal ( $z$ ) is calculated at the inverse of the apparent separation between isotopologue peaks ( $\Delta(\frac{m}{z})$ ) for the peak of interest. By multiplying the  $m/z$  ratio of a given peak by its corresponding charge state, the mass of the ionic complex is obtained. Stoichiometries of DNA: cation: ligand can now be assigned to ion signals. The cation stoichiometry provides insight into the number of stable guanine quartets present; given that cations intercalate between adjacent quartets, the number of stable quartets is equal to one more than the number of cations.<sup>42</sup>

### 3.3 Results and Discussion

#### 3.3.1 Ligand structures

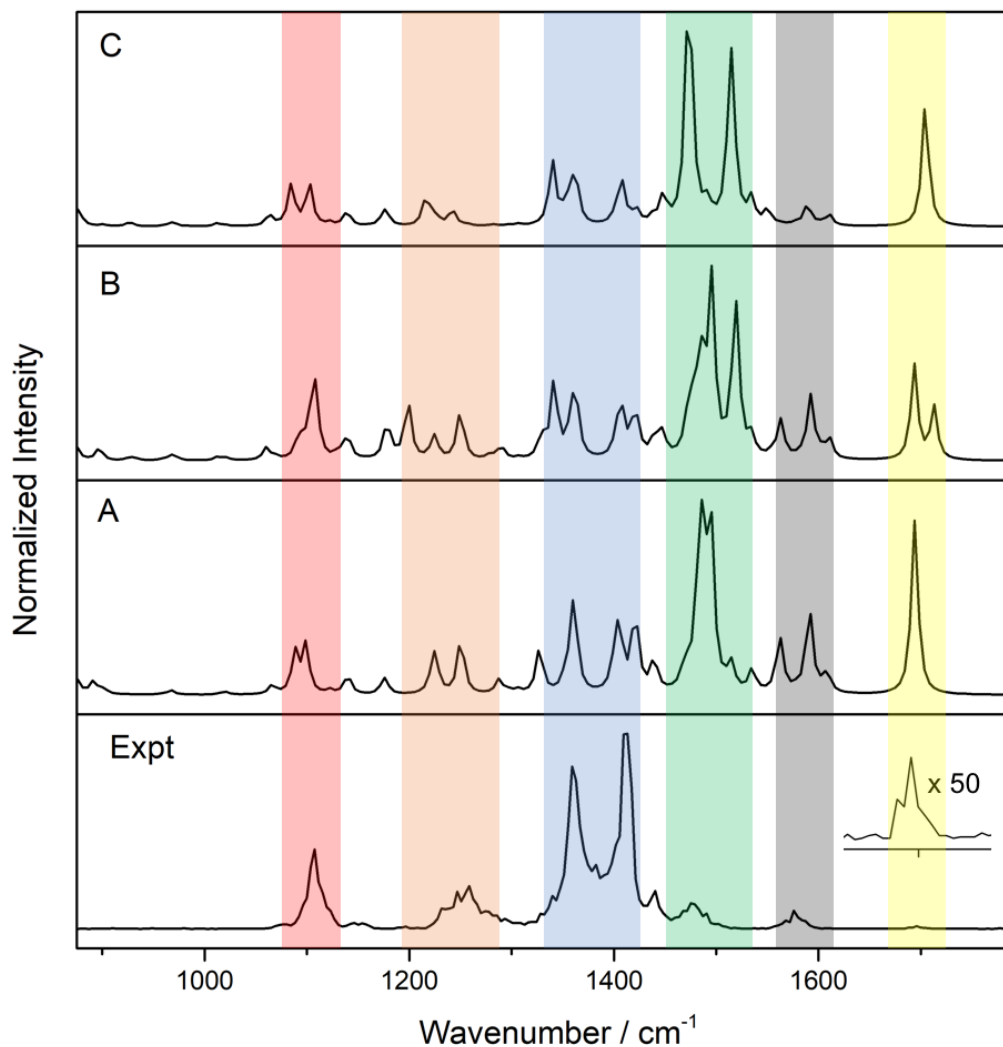
The calculated global minimum (GM) of the PhenDC3 ligand, as well as the next two lowest-energy isomers, are shown in Figure 3.4. Dihedral angles within this molecule were distorted in order to find the most stable conformation. The GM shown here was determined to be a roughly symmetrical molecule, with respect to the orientation of the quinolinium moieties. The dihedral angles between these moieties and the phenanthroline group allow for hydrogen bonding to form a six-membered ring, which appears to confer stability within this molecule. Unlike the optimized higher-energy isomers of this compound, the next two most stable forms also show the quinolinium moieties orientated away from the phenanthroline subgroup. However, neither isomer 2 or 3 exhibit two hydrogen-bond-stabilized six-membered rings. Given that all three of isomers fall within a range of  $\sim 20$  kJ/mol, a range which is considered thermodynamically stable, it is possible that all are present in the gas phase. This assumption is validated by the IRMPD spectra shown in Figure 3.5.



**Figure 3.4** The three lowest-energy isomers of the PhenDC3 ligand. Relative Gibbs' energies as calculated using B3LYP/6-311++G(d,p) level of theory at 298K are provided in kJ/mol.

The XYZ coordinates of the GM structure for each ligand in Figure 3.2 are provided in Appendix I. The lowest energy conformation for all ligands displayed a high degree of symmetry and planarity, as well as a characteristic side-chain orientation resulting in the forming of hydrogen-bond-stabilized six membered rings.

In order to evaluate the accuracy of our computational predictions, calculated frequency spectra of the PhenDC3 and 360A ligands were compared to IRMPD spectra obtained using a FEL. Figure 3.5 plots the experimental IRMPD spectra against the computed frequencies of the three lowest-energy isomers of PhenDC3, as calculated at the B3LYP/6-311+G(d,p) level of theory. Due to the fact that these harmonic frequency calculations do not account for the effects of anharmonicity, the calculated spectra predict vibrational peaks at wavenumbers higher than those displayed in the experimental spectra. In order to correct for this, a commonly used DFT scaling factor of 0.9679 has been applied to the computed spectra displayed.<sup>66</sup> The calculated spectra of the three lowest-isomers of PhenDC3 display good agreement with the experimental spectra shown. This suggests that all three are likely present, and thermodynamically stable, in the gas phase.



**Figure 3.5** Comparison of the experimental IRMPD spectrum of PhenDC3 to the calculated IR spectra of (A) the global minimum and the next two lowest-energy isomers with relative Gibbs' energies of (B) 15.8 kJ/mol, and (C) 20.9 kJ/mol, respectively. Computational spectra have been calculated at the B3LYP/6-311+G(d,p) level of theory, and a DFT scaling factor of 0.9679 has been applied.

Based on the comparison between the calculated and experimental spectra shown in Figure 3.5, we are able to assign the observed vibrational bands to normal modes associated with the lowest-energy isomers of PhendDC3. The peak at  $\sim 1100\text{ cm}^{-1}$ , shown highlighted in red, can be assigned to H-scissoring and ring N-stretching motions. The vibrational peaks highlighted in blue can be confidently associated with the stretching of bonds within the ring systems. The vibrational peak observed at  $\sim 1475\text{ cm}^{-1}$ , shown highlighted in green, is attributed to a combination of H wagging and ring stretching motions. Lastly, C=O stretching results in the

vibrational band observed at approximately  $\sim 1700\text{ cm}^{-1}$  as highlighted in yellow. Both the green and yellow highlighted bands exhibit low intensities relative to the predicted spectra. Inefficient coupling between these normal modes and the dissociative threshold leading to fragmentation at the sites of these vibrations within the molecule, are likely leading to the evident difference in intensities.

### 3.3.2 G-ligand interactions

The high level of agreement between our predicted IR spectra, and that obtained experimentally, instills confidence in the computational methods used to study the structures of these ligands with guanine. We can now turn our attention to studying G-ligand interactions. Figure 3.6 illustrates the lowest-energy conformations of the clusters containing the quadruplex-binding ligands with guanine. Here, the sites of interaction between the ligands and G are highlighted, and the resulting H-bonds are shown as dotted lines. We observe that the carbonyl oxygen of the guanine is always oriented to interact with the positively-charged site on the ligand molecule. All ligands, except for PhenDC3, show the G molecule oriented roughly in the same plane as the largely planar ligands. In contrast, the lowest energy-conformation of the G-PhenDC3 cluster exhibits guanine positioned at a perpendicular angle, and inserted between the quinolinium moieties of the ligand. The manner in which the ligands interact with guanine shown here, can be used to predict the binding sites of these ligands on G-quadruplexes. The nature of the interaction between ligands 360A, PDS, TMPyP4, and TrisQ, with guanine, predicts non-quartet G-quadruplex binding; these ligands will likely cluster with the DNA backbone of the quadruplex, rather than stacking with the G-quartet itself. In contrast, PhenDC3 should display intercalation between quartets of the G-quadruplex.



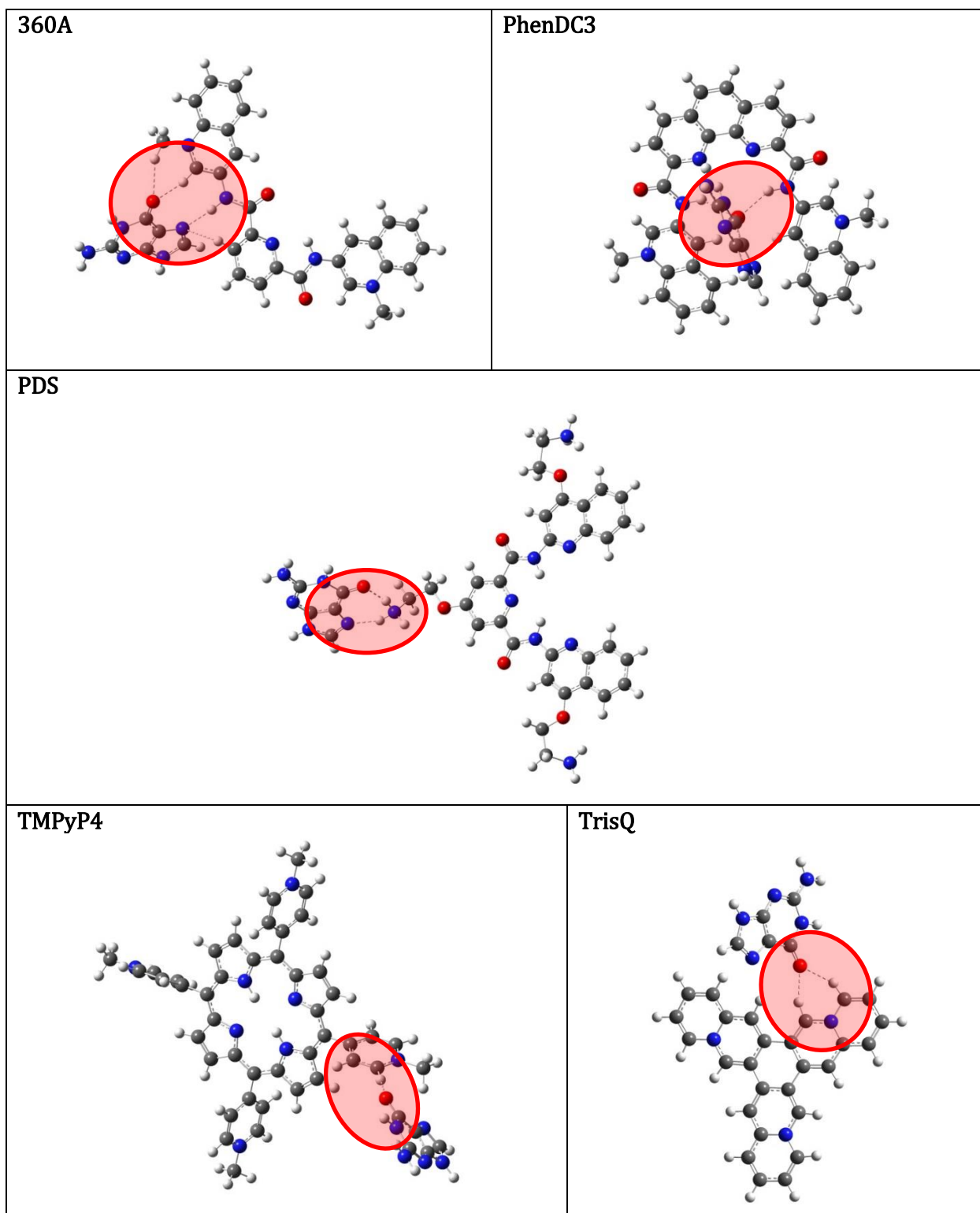


Figure 3.6 The lowest-energy conformation for each ligand clustered with guanine, as calculated at the B3LYP/6-311++G(d,p) level of theory.

Binding energies (BE) of the ligands with guanine were calculated according to equation 8. Zero-point energies were included for each species. Thus, the reported BEs are equivalent to the dissociation energies of the clusters, which represent the maximum BEs.

$$BE = (E_{ligand} + ZPE_{ligand} + E_{solvent} + ZPE_{solvent}) - (E_{cluster} + ZPE_{cluster}) \quad (8)$$

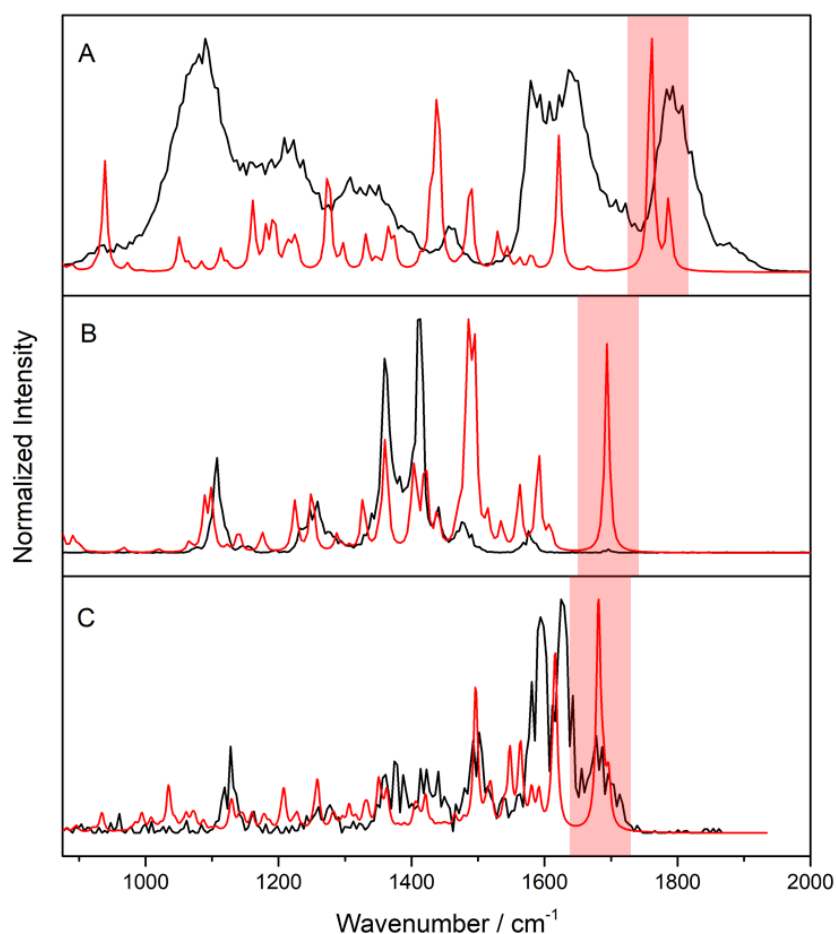
Table 3.2 summarizes the zero-point corrected binding energies for the lowest-energy conformation of each G-ligand cluster. Here, we see that all ligands exhibit strong binding to guanine; each cluster has a dissociation energy of greater than 1 eV. The strongest binding is observed in the G-PDS cluster; a BE of 2.2 eV is relatively unsurprising, as the interaction between guanine and PDS involves hydrogen bonding of a positively-charged site on the ligand, with two electronegative atoms on the nucleobase, to result in a six-membered ring. In contrast, the lowest binding energy is observed in the G-PhenDC3 cluster, which may be due to changes in the ligand structure caused by nucleobase insertion.

**Table 3.2. G-quadruplex-ligand binding energies of the lowest-energy cluster conformations, as calculated at the B3LYP/6-311++G(d,p) level of theory.**

Ligand	Binding Energy (eV)
360A	1.24
PDS	2.22
PhenDC3	1.13
TMPyP4	1.51
TrisQ	1.88

The determination of binding energies plays an important role in the study of quadruplex-binding ligands. By binding to G-quadruplexes, ligands are able to stabilize these structures,<sup>7,8</sup> and ultimately prevent the expression of the largely undesirable enzyme telomerase. Thus, ligands that are predicted to bind more strongly to quadruplexes will likely be more difficult to dislodge or dissociate, and thus will confer greater stability to the G-quadruplex structure.

A comparison of the IR frequency spectra of protonated guanosine, PhenDC3 ligand, and their resulting complex, is shown in Figure 3.7. Here the red-highlighted region of each spectra encompasses the vibrational peaks associated with the carbonyl stretching motion. As expected by the G-ligand calculations, we see a shift to a lower frequency for this vibrational peak upon complexation. This C=O stretching band in the protonated guanosine spectrum is observed at a frequency of  $\sim 1770\text{ cm}^{-1}$ , however upon clustering with the PhenDC3 ligand, we see that the complex results in a shift of  $\sim 100\text{ cm}^{-1}$ . This indicates that it is the carbonyl site of the guanine, as well as the protonated guanosine, that is the location of interaction for the cluster.



**Figure 3.7** Comparison of the experimental IRMPD and calculated IR spectra of (A) protonated guanosine (B) PhenDC3 ligand, and (C) protonated guanosine-PhenDC3 complex, respectively. Experimental spectra is shown in black, calculated spectra for the lowest-energy isomer of each is shown in red. Computational spectra have been calculated at the B3LYP/6-311+G(d,p) level of theory, and a DFT scaling factor of 0.9679 has been applied.

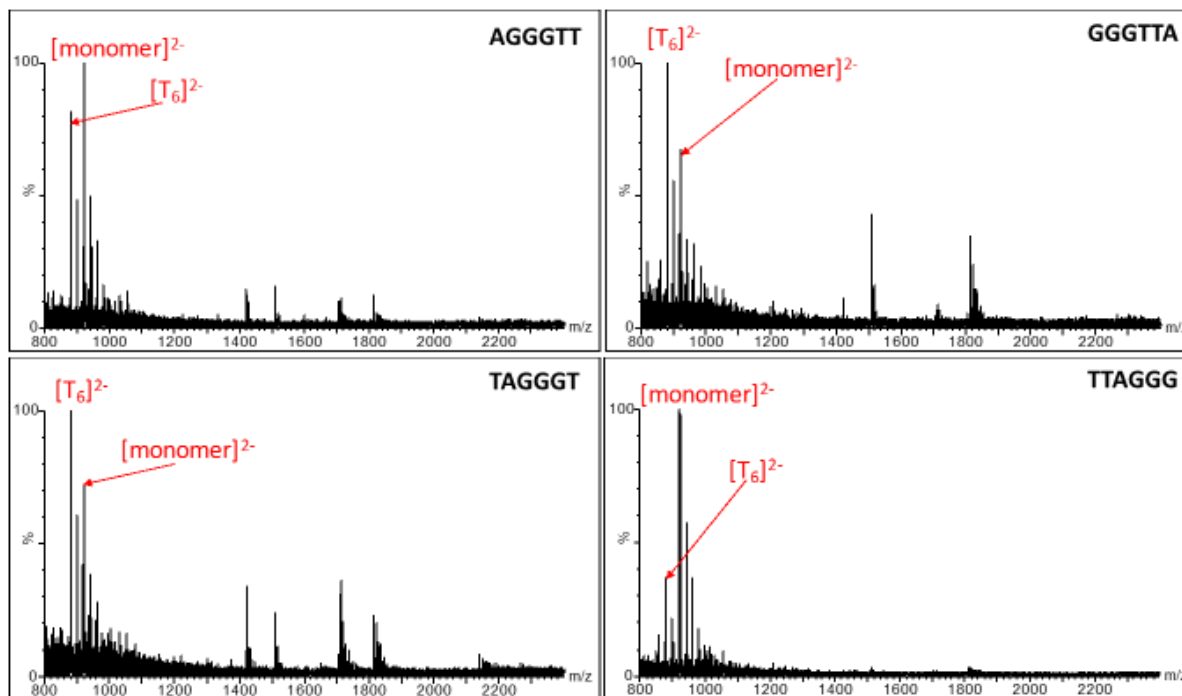
### 3.3.3 Antiparallel Tetramolecular G-quadruplexes

Despite the extensive structural polymorphism exhibited by intramolecular and bimolecular G-quadruplexes, those composed of four individual nucleotide strands (tetramolecular) have only ever been observed in a single conformation.<sup>53,57</sup> Specifically, these tetrameric structures are said to only exist with all strands parallel to one another, such that all strands have the same 5' to 3' orientation, and the guanine glycosidic torsional angles are all in the *anti* conformation.<sup>53</sup> Interestingly, as part of a collaboration with the Gabelica group at the University of Bordeaux, a member of the Hopkins Lab identified a possible example of an antiparallel tetramolecular G-quadruplex. This classification has not been observed previously, thus significant efforts were put forth to verify the observed structure. Here, the resulting G-quadruplex topology is explored as a number of parameters (i.e. nucleotide sequence, cation, quadruplex-binding ligand, stoichiometries) are manipulated. The parameters observed to result in this antiparallel tetramolecular topology include the GGGTTA DNA sequence, K<sup>+</sup> cation, with 3 equivalents of PhenDC3 per quadruplex.

#### Nucleotide Sequence

Seven DNA oligonucleotide sequences which either contain, or are closely related to, the human telomeric sequence repeat (GGGTTA) were initially evaluated. ESI-MS spectra of these sequences with TMAA, KCl, and PhenDC3 (prepared according to specifications in Table 1), were evaluated to determine the degree of quadruplex formation. With the use of an oligonucleotide internal standard (2.5 μmol/L T6) that does not form quadruplex structures,<sup>67</sup> the extent of quadruplex formation for the oligonucleotide sequence of interest can be elucidated. Comparison of the peak height of the internal standard, to those corresponding to the monomer as well as the ligand-bound and unbound quadruplexes of the sequence of interest, yields the

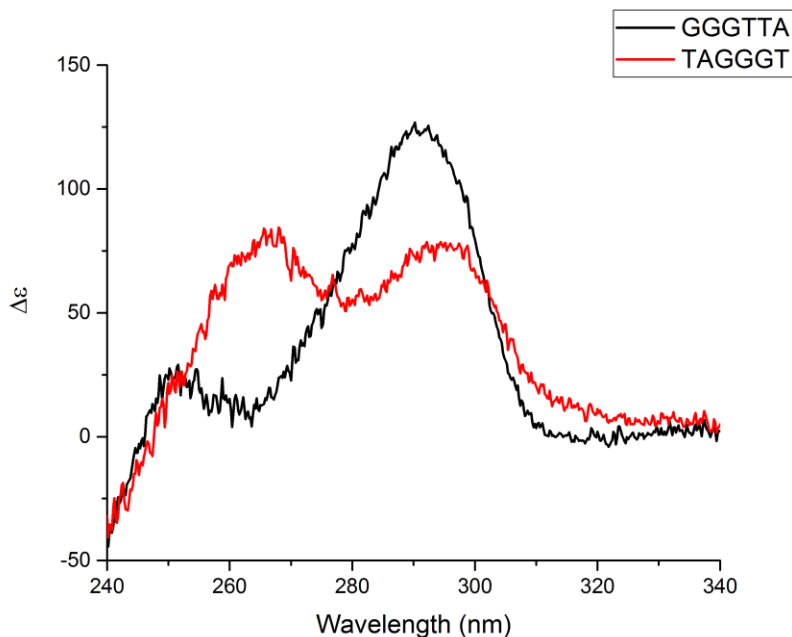
degree of quadruplex formation. Figure 3.8 displays the resulting ESI-MS results of the four DNA oligonucleotides containing the human telomeric sequence.



**Figure 3.8** Mass spectra of solutions containing 40  $\mu\text{mol/L}$  DNA oligonucleotide (sequence denoted), 100  $\text{mmol/L}$  TMAA, 1  $\text{mmol/L}$  KCl, 7.5  $\mu\text{mol/L}$  PhenDC3, and 2.5  $\mu\text{mol/L}$  T6.

Each DNA sequence shown in Figure 3.8 has a molecular weight of 1847.3 g/mol, and carries a  $[-2]$  negative charge, to give an observed  $m/z$  ratio of 923.7. Peaks corresponding to this monomer are labelled. Similarly, the T6 sequence has a molecular weight of 1763.2 g/mol, and carries as  $[-2]$  negative charge, for an  $m/z$  ratio of 881.6. Peaks corresponding to quadruplex structures vary depending on both their charge and number of bound ligands, however the most dominant quadruplex signals are attributed to those carrying a combination of  $[-5]$  or  $[-6]$  charge with either two or three bound PhenDC3 ligands. Zooming in on these signals elucidates the number of cations ( $\text{K}^+$  here) present. Based on the intensities of the quadruplex ion signals relative to the internal standard, two ligand sequences were selected for further evaluation; GGGTTA and TAGGGT both displayed a high degree of quadruplex formation under the stated

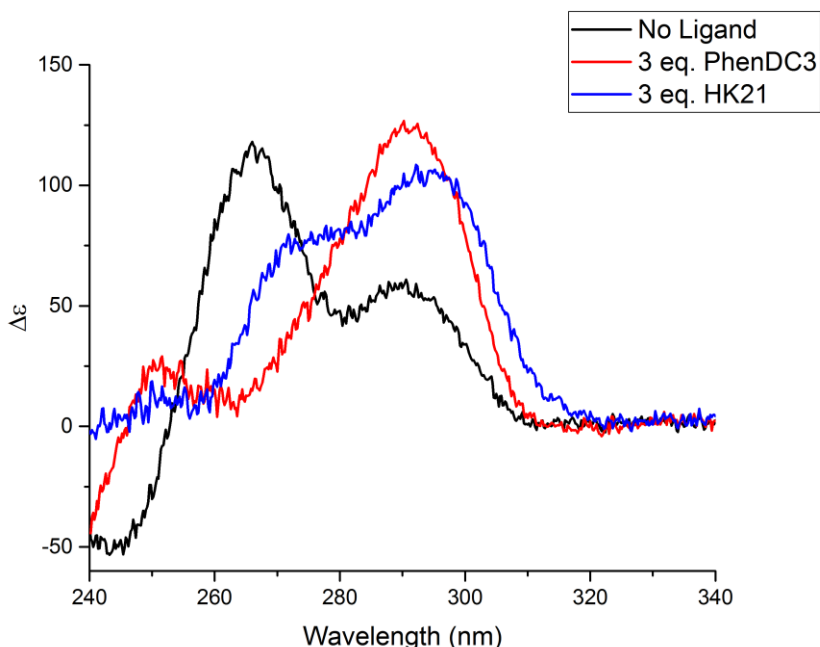
conditions. In order to evaluate the three-dimensional conformation of these structures, CD spectra were obtained.



**Figure 3.9** CD spectra of solutions containing 40  $\mu\text{mol/L}$  DNA oligonucleotide, 100 mmol/L TMAA, 1 mmol/L KCl, and 7.5  $\mu\text{mol/L}$  PhenDC3.

The CD spectra shown in Figure 3.9 illustrate the difference in quadruplex topology as a result of the DNA sequence used. The spectra of parallel quadruplexes are characterized by the presence of a negative peak at  $\sim 240$  nm and a positive peak at  $\sim 260$  nm.<sup>43</sup> Alternatively, the spectra of antiparallel G-quadruplexes display a local minimum at  $\sim 260$  nm and a global maximum at  $\sim 290$  nm.<sup>43</sup> Here, we see that under the described conditions, the solution containing the TAGGGT sequence displays bands corresponding to a parallel quadruplex structure. In contrast, the GGGTTA sequence results in the formation of a previously unobserved antiparallel tetramolecular quadruplex, as evident by the characteristic antiparallel bands previously described. Thus, we turn our focus to experiments involving the GGGTTA sequence exclusively.

Next, the effect of the ligand on the quadruplex topology was explored. Here, mass spectra were obtained for solutions containing either no ligand, or PhenDC3, TrisQ, or HK21, in a 3:1 stoichiometric ratio of ligand to quadruplex concentration. The resulting spectra indicated that, regardless of the ligand conditions used, quadruplex formation took place. However, upon further inspection of the ion signals, it was determined that the TrisQ ligand was not bound to the quadruplex, and thus evaluation of solutions containing this ligand were unwarranted. CD experiments were performed on the solutions containing either no ligand, or PhenDC3, or HK21. Figure 3.10 captures the resulting three-dimensional quadruplex topology as a result of the ligand used. Here, we see that the solution containing no ligand results in a parallel G-quadruplex, as evident by the characteristic negative peak at ~240 nm and the global maximum positive peak at ~260 nm. In contrast to this, and the antiparallel structure exhibited with PhenDC3, the solution containing HK21 ligand resulted in a hybrid topology. That is, the resulting HK21-bound quadruplex exhibits G-quartets where three guanines are in an *anti* conformation and one guanine is in a *syn* orientation, or vice versa. Thus, we proceed with experiments involving only the PhenDC3 ligand hereafter.



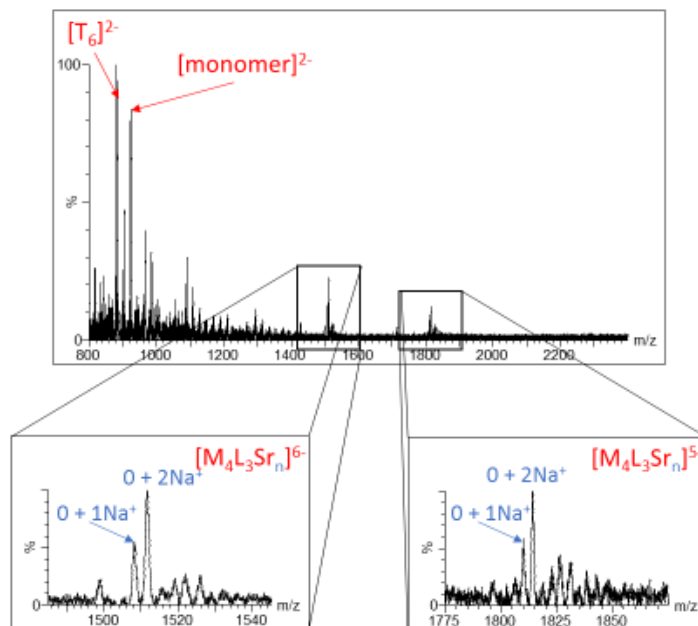
**Figure 3.10 CD spectra of solutions containing 40  $\mu\text{mol/L}$  GGGTTA DNA oligonucleotide, 100 mmol/L TMAA, 1 mmol/L KCl, and 7.5  $\mu\text{mol/L}$  ligand.**

At this stage, the ratio of ligand to DNA oligonucleotide concentration within the solutions was assessed. Solutions containing either 0, 1, 3, or 5 equivalents of PhenDC3 per quadruplex, were evaluated. Both spectra containing either no ligand or 1 equivalent of ligand to quadruplex, resulted in a parallel topology. In contrast, solutions containing either 3 or 5 equivalents produced a CD spectra characteristic of antiparallel topology. Thus, to further evaluate the existence of antiparallel tetramolecular quadruplexes, a minimum of 3 ligand equivalents per quadruplex was used.

The last parameter to be evaluated was the quadruplex-stabilizing cation used. Here, we included either potassium chloride or strontium acetate to provide  $\text{K}^+$  or  $\text{Sr}^{2+}$ , respectively. Mass spectra obtained of solutions containing these salts both displayed a high degree of quadruplex formation. However, upon further examination by zooming in on the mass spectra, we observed that despite the inclusion of strontium acetate, no strontium ions had complexed with the quadruplexes. Rather, sodium ions, which were not intentionally introduced to the solution,



were performing as the required stabilizing cation. Experiments with strontium acetate as the salt solution were repeated several times, however sodium contamination could not be avoided, and all resulting mass spectra illustrated the presence of these cations.



**Figure 3.11** Mass spectra of solutions containing 40  $\mu\text{mol/L}$  GGGTTA DNA oligonucleotide, 100 mmol/L TMAA, 1 mmol/L  $\text{Sr}(\text{OAc})_2$ , 7.5  $\mu\text{mol/L}$  PhenDC3, and 2.5  $\mu\text{mol/L}$  T6.

In the above figure, M denotes the number of oligonucleotide monomers ( $M_4 =$  quadruplex), L is the ligand, and Sr depicts the number of strontium ions present.

These experiments have established the existence of the antiparallel tetramolecular quadruplex, as well as the DNA sequence, ligand, and cation, required to form this topology. Further studies using nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography to determine the exact structure did not prove to be fruitful. The resulting NMR data, as provided in Appendix I, displayed a very convoluted spectrum, pointing to the presence of multiple structures. The X-ray crystallography experiments also did not yield any useful information, as no crystals suitable for analysis had formed within six months (at the time of writing this thesis) following their preparation. Collaboration with another group well-versed in

X-ray crystallography is currently being pursued, in order to establish structure of these quadruplexes.

### 3.4 Conclusions

The potential energy surface of a number of known G-quadruplex-binding ligands were explored using the basin hopping search algorithm. The unique isomers identified by this search were then optimized at the density functional level of theory, and their respective IR frequency spectra were obtained. Optimized geometries of these ligands exhibited roughly planarly-oriented sidechains. Experimental IRMPD spectra of two of these ligands were obtained using a free electron laser, and subsequently compared to computational predictions. A good degree of agreement was found between the experimental and calculated spectra, thus allowing for the assignment of normal modes to observed vibrational peaks. We determined that the computational methods employed here are suitable for the study of G-quadruplex binders. Basin hopping was then used to explore the potential energy surface of guanine-ligand clusters. Lowest-energy conformations of these clusters were then evaluated to determine the nature of the guanine-ligand interactions, as well as the resulting cluster dissociation energies. Ligands 360A, PDS, TMPyP4, and TrisQ all displayed guanine molecules in a roughly planar orientation to the ligand, whereas PhenDC3 showed a perpendicular orientation resulting in insertion of the guanine in its quinolinium sidechains. Data obtained regarding these interactions predicts PhenDC3 as the only ligand to intercalate between G-quartets within a quadruplex.

As part of a collaboration with the Gabelica group at the University of Bordeaux, the existence of the previously unobserved antiparallel tetramolecular quadruplex was evaluated. Manipulation of a number of parameters including the DNA oligonucleotide sequence, stabilizing cation, ligand, and ligand stoichiometries, were explored. Studies on several DNA oligonucleotides related to the human telomeric sequence established the extent to which they

fold into G-quadruplexes. Of the sequences displaying the greatest quantity of folded quadruplex, only the GGGTTA sequence produced the desired antiparallel structure. It was determined that potassium should be used as the quadruplex-stabilizing cation, as studies with strontium displayed only sodium cation insertion, even when no sodium was intentionally added to the solution. We also found that the PhenDC3 ligand, when used in combination with the other optimized parameters resulted in an antiparallel topology. Although these experiments have established the existence of the antiparallel tetramolecular quadruplex, further studies must be conducted to unequivocally determine the exact structure.

## 4 Pharmaceutically Relevant Ions

This chapter is comprised of a manuscript which ready for submission. My role involved the completion of all differential mobility spectrometry experiments, as well as all calculations prior to treatment with the Random Forest supervised machine learning algorithm.

### 4.1 Introduction

There is a great deal of interest in drugs which covalently modify their biological target.<sup>9-11,68,69</sup> This interest stems from the potential for increased selectivity by targeting specific nucleophilic residues in proteins, and the possibility that covalent drugs might exhibit enhanced efficacy due to the silencing of enzymatic activity until protein re-synthesis can occur.<sup>68</sup> To date, dozens of covalent drugs have been approved for treatment of, e.g., hyperlipidemia, infectious diseases, and cancer.<sup>9</sup> These drugs contain electrophilic moieties such as carbamates, acetates,  $\beta$ -lactones,  $\beta$ -lactams, and acrylamides, and several reviews have described the efforts to develop covalent inhibitor therapeutics.<sup>9-11,69</sup> Covalent inhibitors offer several advantages over non-covalent analogues, including longer duration of action, increased biochemical efficiency, and the potential to avoid some mechanisms of drug resistance.<sup>13</sup> However, employing a covalent approach for drug interactions does present additional risk for hypersensitivity and toxicity arising from covalent modification of unintended targets and/or high reactivity.<sup>70-72</sup>

A 2013 study by Dahal et al. showed that high rates of reaction with glutathione (GSH) correlated with high covalent binding burdens in hepatocytes.<sup>73</sup> It has also been shown that the degree of covalent binding burden observed in hepatocytes (when combined with the daily dose of drug) is predictive of human hepatotoxicity.<sup>74,75</sup> Consequently, the reactivity of covalent inhibitors with GSH has been utilized in the assessment of biological activity and selectivity.<sup>70</sup> In 2014, Flanagan et al. outlined chemical and computational methods to characterize covalent

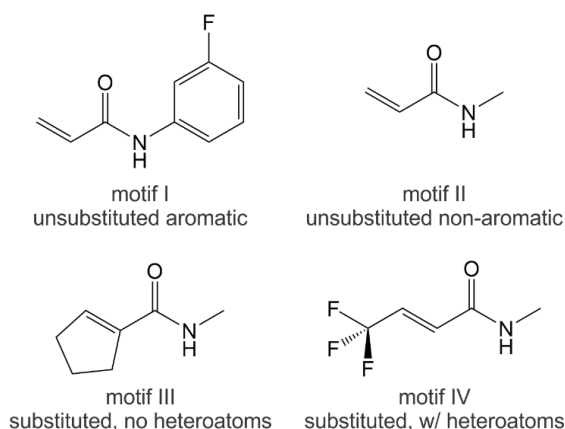
reactive groups (CRGs) for the prospective design of irreversible covalent inhibitors.<sup>68</sup> Their experimental methods employed mass spectrometry [method A; reference 30] and NMR spectroscopy [method B; reference 30] to determine pseudo-first-order kinetics for a series of GSH reactions with various CRGs. Their computational method employed a Monte Carlo conformational search followed by DFT calculations at the B3LYP/6-311+G(d,p) level of theory to estimate the kinetic reaction barrier ( $\Delta G^\ddagger$ ) for acrylamide CRGs interacting with methanethiolate ( $\text{MeS}^-$ ), which was used as a computational surrogate for the thiolate ion  $\text{GS}^-$ . A global analysis of the experimental and computational data showed that the calculated values of  $\Delta G^\ddagger$  correlated strongly with the measured reaction half-life ( $R^2 = 0.915$ ).<sup>68</sup> Besides introducing an accurate computational method for in silico screening, the work of Flanagan et al. also provided insight into the rate limiting step for the reaction of acrylamide CRGs with GSH. Specifically, calculations showed that the activation energy of the conjugate reactions were well approximated by calculating the Gibbs' energy difference between the precursor molecules and the transition state associated with complexation, and that the transition state energy could be tuned by manipulating the stereo-electronic properties of the acryloyl group.<sup>68</sup> Stated in another way, the chemical reactivities of the acrylamide CRGs are strongly influenced by their interaction potential energy surface.

Recently, we have demonstrated the use of differential mobility spectrometry (DMS) for assessing molecular properties in the gas phase (e.g., ion-solvent binding energy),<sup>45,47</sup> which we have shown also correlate with condensed phase physicochemical properties (e.g., solubility, cell permeability).<sup>76</sup> The outcomes of these recent studies suggest that the strong correlations observed between DMS behavior and molecular properties occur because DMS provides an indirect measurement of the interaction potential between an analyte and its environment.<sup>76</sup> A detailed description of DMS is available in references <sup>44,47,48,50,51,77,78</sup>. In brief, the critical output

of the DMS instrument is the relationship between the separation voltage (SV) and the compensation voltage (CV) that is required for optimal transmission of an analyte through the DMS cell. A global view of this behavior across the full SV and CV range is known as a dispersion plot.<sup>44,47,79</sup> Analytes which interact strongly with the collision gas in the DMS cell require increasingly negative CV with increasing SV for optimal transmission (termed Type-A behavior), whereas those which exhibit hard sphere collisions require increasingly positive CV with increasing SV (termed Type-C behavior). Weakly interacting species initially exhibit increasingly negative CV with increasing SV, but reach a minimum value of CV before adopting a positive trend (Type-B behavior).<sup>44,47,79</sup> We have demonstrated previously that DMS is able to distinguish subtle stereo-electronic effects based on differences in an analyte's trajectory as it transits the DMS cell.<sup>76</sup> We have also recently demonstrated that DMS data can be treated with supervised machine learning (ML) to assess quantitatively a variety of molecular properties which are related to an analyte's interaction potential. Here we show that DMS dispersion data can be treated with ML to generate a predictive model of the physicochemical properties acrylamide CRGs, including their reactivities with GSH.

## 4.2 Methodology

In total, twenty-four acrylamide CRGs were studied. These species were selected from the set of those characterized in reference 68, and included molecules which can be classified into four different structural motifs: (I) unsubstituted aromatic derivatives, (II) unsubstituted non-aromatic derivatives, (III) substituted derivatives which do not contain heteroatoms, and (IV) substituted derivatives which contain heteroatoms. Examples of these four structural motifs are shown in Figure 4.1, and the remaining molecules in our test set are provided in the supplementary information.



**Figure 4.1** The structural motifs of the acrylamide CRGs.

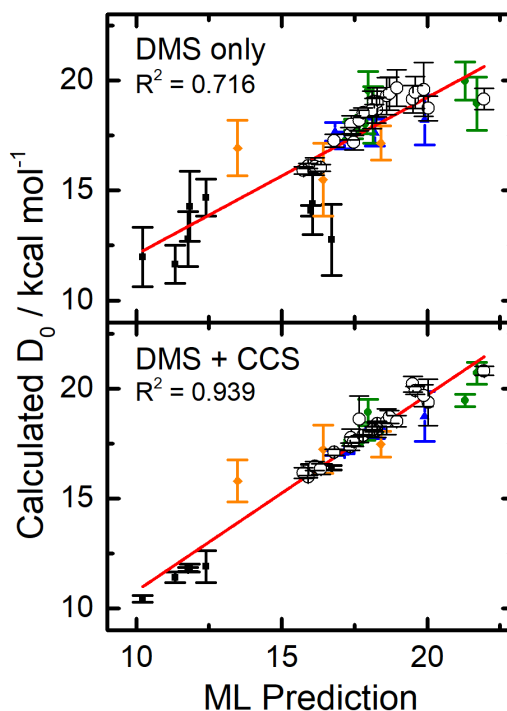
The DMS behaviors of the twenty-four CRGs were recorded in an  $N_2$  environment (1 atm, 150 °C) that was seeded with 1.5 % isopropyl alcohol vapor. The dispersion plot data were then combined with calculated ion-solvent binding energies ( $D_0$ ) and collision cross sections (CCSs),<sup>80</sup> and measured condensed phase physicochemical properties (pKa, pKb, LogD,  $t_{1/2}$ ). This combined data set was subsequently analyzed with a Random Forest supervised machine learning (ML) algorithm, for which we provide details in the supplementary material. In previous work, we conducted a survey of ML models and identified the Random Forest algorithm as implemented in the Python Orange data mining package as the best model for treating DMS data based on fit  $R^2$  values.<sup>81</sup>

### 4.3 Results and Discussion

Figure 4.2 shows the results of a Random Forest model which predicts ion-solvent binding energy ( $D_0$ ) using only DMS data (Figure 4.2; top panel) and the combination of DMS data and CCS (Figure 4.2; bottom panel). Note that the fits shown in Figure 4.2 also included data for twenty 2-methylquinonlin-8-ol (MQOH) derivatives, which we published previously.<sup>76</sup> The MQOH data set was included to ensure that the DMS methodology was valid across a variety of molecule types/chemistries. We have shown previously that DMS behavior is a strong predictor

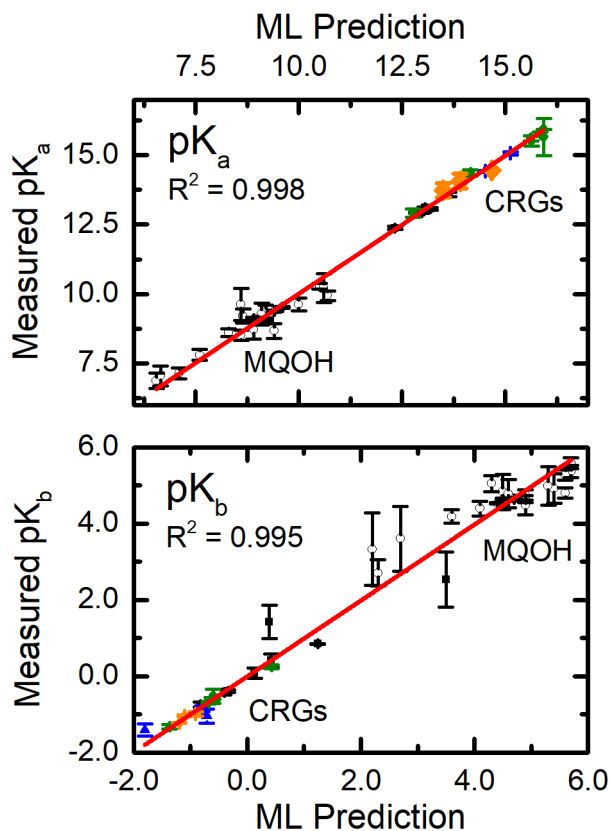
of ion-solvent binding energy,<sup>44,45,47,76</sup> and it is clear from Figure 4.2 that this is also the case for the acrylamide CRGs. However, the correlation between  $D_0$  and the observed DMS behavior is weaker for the CRGs than was observed in studies of other molecular systems. We hypothesized that poorer correlation observed for the CRGs is due to the significant variation in molecular size across the data set; in principle, DMS behavior also depends on the CCS of the analytes, but this was neglected in previous studies due to negligible size differences across the studied test sets.<sup>45,76</sup> Here we are studying a collection of molecules which vary significantly in molecular size from the smallest to largest species (ca. 15 % difference). The bottom panel of Figure 4.2 demonstrates that accounting for molecular size in the ML fit of the DMS data yields a marked improvement of the correlation between calculated ion-solvent binding energy and ML predictions. Note that if a vector of random numbers is included in place of the CCS data in the ML fit, the  $R^2$  value increases marginally to 0.757 (viz. CCS has a significant impact on the quality of the fit).





**Figure 4.2 (Top panel)** The correlation between calculated  $D_0$  for ion-solvent binding and the predictions of the Random Forest ML model when using only DMS data as the input. **(Bottom panel)** The correlation between calculated  $D_0$  for ion-solvent binding and the predictions of the Random Forest ML model when using DMS data and calculated collision cross section as the inputs.

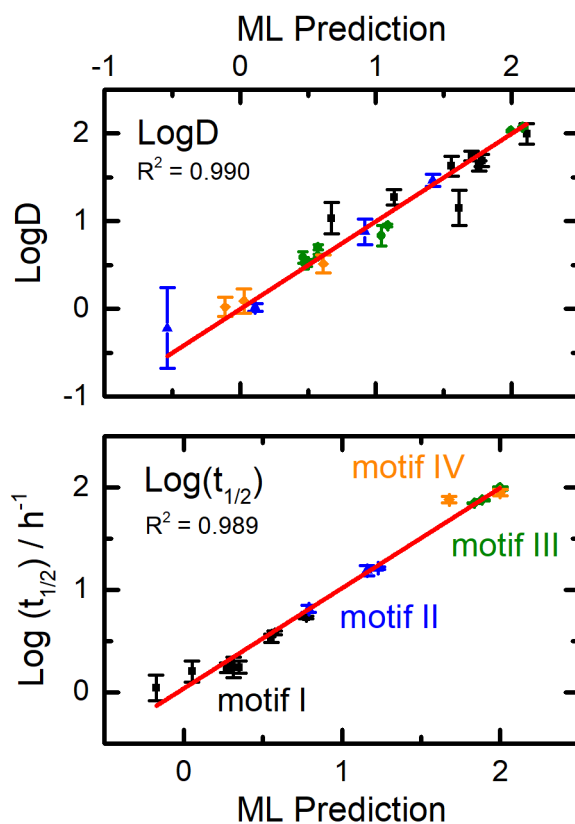
To further explore the correlations between DMS behavior and molecular physicochemical properties, we treated the combined CRG and MQOH data set with the Random Forest algorithm to predict solution phase  $pK_a$  and  $pK_b$ . The results of these fits, which again included DMS data and calculated CCSs, are shown in Figure 4.3. Clearly, there are exceptionally strong correlations between the gas phase DMS measurements and the experimental  $pK_a$  /  $pK_b$  values. It is notable that the ML model easily distinguishes between the acrylamide and MQOH sub-sets, and predicts equally accurate  $pK_a$  /  $pK_b$  values for both groups of molecules. It is also worth noting that, in general, species which exhibit stronger ion-solvent binding energies tend to exhibit higher values of  $pK_a$  (compare CRGs in the top pane of Figure 4.3 to Figure 4.2). We made this observation previously in our original study of the MQOH data set.<sup>76</sup>



**Figure 4.3 (Top panel)** The correlation between measured pK<sub>a</sub> and the predictions of the Random Forest ML model. **(Bottom panel)** The correlation between measured pK<sub>b</sub> and the predictions of the Random Forest ML model. Both fits used DMS data and calculated collision cross section as the inputs.

Figure 4.4 shows the correlations between ML predictions and experimentally determined values of LogD (water/octanol distribution coefficient) and Log ( $t_{1/2}$ ), the half-life for reactions between GSH and the acrylamide CRGs. Note that MQOH data was not included in these fits because these properties have not been measured for the 2-methylquinolin-8-ol derivatives. We find that ML treatment of the CCS and DMS data also yields a very accurate predictive model for LogD coefficients (Figure 4.4; top panel) and reaction kinetics for GSH with the acrylamide CRGs (Figure 4.4; bottom panel). Further consideration of the correlation plot for Log ( $t_{1/2}$ ) seems to reveal additional insight concerning acrylamide CRG reactivity. The substituted species (motifs III and IV) exhibit reaction rates which are approximately an order of magnitude less than those of the unsubstituted non-aromatic species (motif II) and two orders

of magnitude less than those of the unsubstituted aromatic CRGs (motif I). Flanagan et al. showed that this trend roughly correlated with increasing values of  $\Delta G^\ddagger$ .<sup>68</sup> Here we see that the trend in CRG reactivity also roughly correlates with ion-solvent binding energy ( $D_0$ ) which, like  $\Delta G^\ddagger$ , is clearly influenced by the stereo-electronic properties, and thus interaction potential, of the acryloyl group.



**Figure 4.4** (Top panel) The correlation between measured LogD and the predictions of the Random Forest ML model. (Bottom panel) The correlation between measured  $t_{1/2}$  for reaction with GSH and the predictions of the Random Forest ML model. Both fits used DMS data and calculated collision cross section as the inputs.

It is important to highlight the fact that DMS measurements probe the charged molecules (either protonated or de-protonated), whereas the condensed phase measurements probe the neutral species. In all cases, protonation occurred at the carbonyl group and deprotonation at the amine. Since the DMS ion-solvent interactions, which occur at the site of charging, directly sample the interaction potential of the acryloyl moiety, the charged species are useful representations

for the hydrogen bonding experienced by the neutral CRGs in bulk protic solution; intermolecular H-bonding is the initial step in protonation or deprotonation of a molecule, and it has been shown that these structures resemble the early stages of the proton transfer process.<sup>82-84</sup>

#### 4.4 Conclusions

The observed correlations between the DMS behavior of the acrylamide CRGs and their solution phase physicochemical properties is very appealing because the DMS behavior of an analyte can be recorded in minutes using only picograms of sample. This suggests that DMS might find use as a new technique for fast and accurate physicochemical properties measurements. However, there are still open questions and caveats to be addressed. Despite the fact that we have here conducted an ML-based analysis on a combined CRG and MQOH data set (where possible), there is still much work to be done to unambiguously demonstrate the efficacy of using DMS to assess the properties of a wide variety of molecular types / chemistries. Furthermore, although DMS data were shown to accurately correlate with reaction kinetics for CRGs with GSH, it is important to note that GSH is not necessarily an accurate simulacrum for the active sites of enzymatic systems. Nevertheless, the work reported here does demonstrate the *potential* for DMS measurements to be incorporated into the suite of measurement techniques which are currently utilized to assay molecular properties for drug discovery and development.

## 5 Nucleobase Tautomers

This chapter is comprised of a manuscript which has recently been accepted for publication. It will appear in the *International Journal of Mass Spectrometry*, as part of a special edition in honor of Prof. Terry McMahon.

The outlined study involves the use of differential mobility spectrometry (DMS) to separate tautomers of protonated nucleobases, prior to probing with HDX and CID. We demonstrate that population distributions of these tautomers can be selectively tuned using a declustering voltage ramp post-DMS. Here, we find that some ion signals that would typically be assigned to a higher-energy species, arise from fragmentation of larger nucleobase-containing clusters. Findings of this study illustrate the need for characterization prior to the assignment of ion peaks, and ultimately furthers the development of a fundamental description of DMS.

### 5.1 Introduction

Owing to the importance of DNA and RNA in encoding and expressing genetic information, and the central role that nucleobases play in establishing the structure and functionality of nucleic acid sequences, a great deal of experimental<sup>85-88</sup> and theoretical<sup>89-92</sup> effort has gone into determining the structures and properties of cytosine (C), guanine (G), adenine (A), thymine (T), and uracil (U). Of importance are the sites of protonation and the tautomeric forms that the nucleobases exhibit, since these variations are thought to impact mutagenic processes (*e.g.*, point mutation during nucleic acid replication)<sup>93</sup> and the stabilization of triplex structures.<sup>94</sup> Mass spectrometry and quantum chemical calculations have been employed to great success in determining nucleobase properties such as gas-phase acidity and basicity.<sup>95-101</sup> However, it has been shown previously that several different tautomers are likely to exist simultaneously in a given nucleobase ensemble.<sup>102,103</sup> For example, Salpin *et al.* used infrared multiple photon

dissociation (IRMPD) spectroscopy to demonstrate the presence of at least two tautomeric forms of  $(C + H)^+$ ,  $(T + H)^+$ , and  $(U + H)^+$  in ion populations generated by ESI.<sup>104</sup> Comparison of the experimental IRMPD spectra with IR spectra that were calculated at the B3LYP/6-31++G(d,p) level of theory indicated that the three protonated nucleobases existed predominantly as enolic tautomers, with a small sub-population of oxo tautomers. Subsequent work by Bakker *et al.* showed that the vibrational spectra of monohydrated protonated uracil,  $(U + H)^+ \bullet OH_2$ , and cytosine,  $(C + H)^+ \bullet OH_2$ , were also consistent with the presence of two tautomeric species arising from the production of two protonated forms of the associated nucleobases via ESI.<sup>105,106</sup> This suggests that the molecular properties of protonated nucleobases as determined by mass spectrometry are likely to correspond to an ensemble average for the various tautomeric structures that are present under the experimental conditions employed during measurement. It is therefore desirable to separate the tautomeric species prior to mass spectrometric or spectroscopic interrogation.

Various forms of ion mobility spectrometry have been employed to separate tautomers prior to MS analysis.<sup>107-109</sup> For example, the Attygalle laboratory recently reported on the characterization of tautomer populations of deprotonated hydroxybenzoic acid with travelling wave ion mobility spectrometry.<sup>110</sup> This work challenges the notion that ESI-MS results reflect solution phase population distributions, and demonstrates that tautomer populations can be tuned by varying ESI source conditions. We have also recently reported on the use of ion mobility to characterize tautomer populations generated via ESI by using differential mobility spectrometry (DMS)<sup>51,111</sup> to separate and probe the nitrogen- and oxygen-protonated tautomers of *para*-aminobenzoic acid.<sup>44,79</sup> By taking advantage of the different DMS behaviors of the two tautomers, we could examine the MS/MS fragmentation patterns and HDX behaviors of each species individually and demonstrate that each structure did, indeed, exhibit its own

characteristic physicochemical properties. We also demonstrated that a great deal of care had to be taken in HDX experiments since high vapor pressures of HDX reagent can drive *in situ* tautomerization via proton-transfer relay networks established upon ion-solvent clustering.<sup>79</sup> Studies like these show that ESI solvent effects are not necessarily the primary criteria that determine relative tautomer (or, by extension, isomer or conformer) population distributions. Instead, a variety of post-ESI instrument conditions could potentially contribute to the observed sub-populations within a gas phase ensemble.

Here, we utilize the DMS technique to separate and study the individual tautomeric forms of protonated adenine, (A + H)<sup>+</sup>, guanine, (G + H)<sup>+</sup>, cytosine, (C + H)<sup>+</sup>, thymine, (T + H)<sup>+</sup>, and uracil, (U + H)<sup>+</sup> that are generated via ESI. The various tautomers of these molecules are studied individually by HDX and CID, and we show that the relative tautomer populations can be manipulated post-ESI and post-DMS by using the instrument declustering potential to selectively fragment high-energy, kinetically trapped tautomers prior to MS characterization.

## 5.2 Methods

### 5.2.1 Experimental Details

A SelexION differential mobility spectrometer was used in conjunction with a QTRAP 5500 system (SCIEX; Concord, ON).<sup>44,46,48,77</sup> Instrument parameters included a ESI probe voltage of 5500 V, a source temperature of 32 °C, nebulizing gas pressure of 20 psi, and auxiliary gas pressure of 0 psi. The DMS was set to a temperature of 150 °C, and nitrogen was used as both the curtain gas (20 psi) and collisionally activated dissociation gas (~9 mTorr) for all experiments. Nucleobase solids were purchased from Sigma-Aldrich and subsequently dissolved in a 50:50 mixture of ultrapure water and methanol with 0.1% formic acid to yield solutions of 10 ng/mL. Analyte solutions were pumped into the ESI source at 7  $\mu$ L/min. HPLC-grade methanol,

isopropanol, and deuterium oxide were also purchased from Sigma-Aldrich and used without further purification or dilution.

DMS experiments involved the stepping of the separation voltage (SV) from 0 to 4000 V in 500 V increments. At each SV, the compensation voltage (CV) was scanned from  $-80$  V to 15 V in increments of 0.1 V to produce an ionogram. A dispersion plot,<sup>47,50</sup> which plots optimal conditions for ion transmission as a function of SV and CV, was then generated. Dispersion plots enable the identification of the DMS behavior of particular ions according to known patterns.<sup>47,50,77</sup> These data were acquired for each nucleobase in a pure N<sub>2</sub> DMS environment, as well as with DMS environments that had been seeded with 1.5% (mole ratio) methanol (MeOH) and isopropanol (IPA) chemical modifiers. Hydrogen-deuterium exchange (HDX) experiments were conducted through the infusion of deuterium oxide into the throttle gas. These experiments were conducted under two different HDX conditions. In the first implementation, the throttle gas was bubbled through D<sub>2</sub>O to saturate the N<sub>2</sub> with HDX reagent. This yields maximum rates of HDX in the junction chamber between the DMS cell and the orifice of the mass spectrometer, as described in reference 79. In the second implementation, the throttle gas sampled only the headspace above the D<sub>2</sub>O HDX reagent vessel, resulting in a lower D<sub>2</sub>O partial pressure and slower rates of HDX.<sup>79</sup> In this way, the DMS cell was used to select a specific tautomer prior to HDX, which was monitored by recording a full scan mass spectrum (Q1).

Enhanced product ion (EPI) scans were also conducted for each of the separated nucleobase tautomers. Following DMS isolation of a given tautomer, the collision energy (CE) of the Q2 ion trap was ramped from 0 V to 60 V in 0.25 V increments, while recording the complete mass spectrum at each interval. By plotting the fraction of the parent and each fragment ion present as a function of collision energy, breakdown curves were produced.<sup>112</sup> A schematic diagram of the DMS region is shown in Figure 5.1.



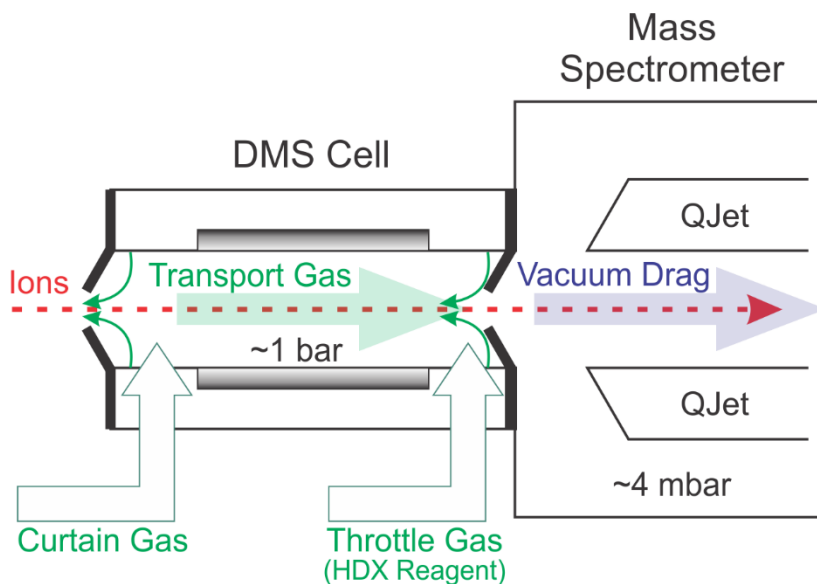


Figure 5.1 A schematic diagram of the DMS cell and the QJet region of the mass spectrometer. Adapted from reference 44.

## 5.2.2 Computational Details

All possible tautomeric forms of the protonated nucleobases (C, G, A, T, and U) were considered. Optimization and frequency calculations ( $T = 298.15 \text{ K}$ ,  $P = 1 \text{ atm}$ ) were performed at the B3LYP level of theory using a 6-311++G(d,p) basis set as implemented in Gaussian 09.<sup>60</sup> Harmonic frequency calculations were conducted for all tautomers to estimate thermochemical corrections to the DFT electronic energies. These calculations also generated harmonic vibrational spectra for the tautomers of  $(C + H)^+$ ,  $(T + H)^+$ , and  $(U + H)^+$  for comparison with the experimental IRMPD spectra reported in reference 104 as a means of validating our computational methodology. Using the calculated standard Gibbs' energies, the various tautomers were sorted energetically to determine the species most likely to be present in the probed ensembles. The four lowest energies tautomers of each protonated nucleobase were then carried forward for treatment with the coupled cluster single, double, and perturbative triple excitations method (*i.e.*, CCSD(T)/6-311++G(d,p) level of theory). These improved electronic energies were combined with the DFT thermochemical corrections to produce the standard

Gibbs' energies that we report in this manuscript. Calculated structures and thermodynamic data are provided in the supporting information.

## 5.3 Results

### 5.3.1 Protonated Cytosine (C + H)<sup>+</sup>

The dispersion plots recorded for C•H<sup>+</sup> (m/z 112) in a pure N<sub>2</sub> environment, and in environments seeded with 1.5 % (mole ratio) MeOH or IPA vapor are shown in Figures 5.2A–C, respectively. Three major peaks are observed for (C + H)<sup>+</sup>, thus suggesting the presence of at least three different tautomers in the sample. Upon seeding the DMS cell with MeOH vapor, the observed Type B ion trajectories are deflected to lower values of CV compared to those observed in the pure N<sub>2</sub> environment.<sup>51,52</sup> This is expected due to the ion-solvent clustering interactions experienced by the (C + H)<sup>+</sup> tautomers in the modified environment.<sup>113</sup> When the stronger clustering IPA modifier is introduced to the DMS cell, the tautomers all adopt Type A behavior. Interestingly, the two weakest features diminish in intensity and disappear by SV ≈ 2500 V in the IPA-modified environment. This is likely an indication of *in situ* tautomerization induced by the IPA clustering,<sup>52</sup> or of proton scavenging by the IPA at higher values of SV (*i.e.*, IPA has a higher gas phase basicity than those two tautomers at high field). The errors (2σ) given on the dispersion plots are determined from Gaussian fits of the peaks observed in the associated ionograms. An example is provided in Figure 5.2D, which plots the ionogram recorded when monitoring the m/z 112 peak with the separation voltage set to SV = 3500 V.

The three lowest energy tautomers of (C + H)<sup>+</sup> as identified by CCSD(T)//B3LYP calculations are shown in the inset of Figure 5.2. These are the same lowest energy species reported by Salpin *et al.* in reference 104. It is worth highlighting the fact that Salpin *et al.* observed only two tautomers via IRMPD, whereas our DMS results show three peaks in the ionogram for m/z 112. There are several possible explanations for this discord, which include the fact that different ion

sources might produce different population distributions and that similar structures (*e.g.*, (C + H)<sup>+</sup> isomers 1 and 3) might exhibit similar spectra in the region studied by Salpin *et al.* Indeed, the calculated vibrational spectra for isomers 1 and 3 exhibit very similar spectra in the 1000 – 2000 cm<sup>-1</sup> region. However, to explore further the properties of the separated species, we examined the CID behavior of the (C + H)<sup>+</sup> species associated with each peak in the ionogram. Figure 5.3 plots the breakdown curves obtained for each of the three peaks shown in the ionogram in Figure 5.2D. As expected, (C + H)<sup>+</sup> exhibits fragmentation channels associated with loss of NH<sub>3</sub>, H<sub>2</sub>O, and HNCO.<sup>114–116</sup> While all three DMS-separated species displayed the same fragmentation channels in roughly the same distributions, peak I showed an onset of fragmentation 4–5 V lower in energy than peaks II and III. Although we expect that these measurements are somewhat coarse, they are reproducible and provide some evidence that the structure associated with peak I in the ionogram is, indeed, a unique tautomeric species.

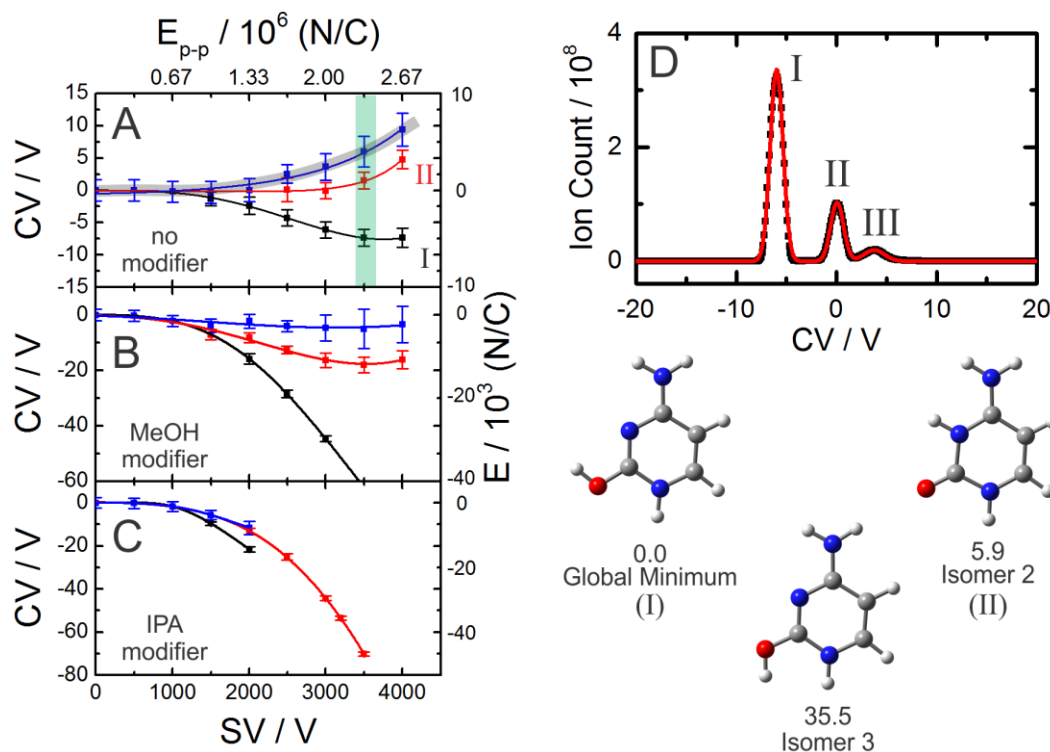


Figure 5.2 The dispersion plot obtained for  $(C + H)^+$  ( $m/z$  112) with a DMS cell containing (A) a pure  $N_2$  environment, and a  $N_2$  environment seeded with 1.5% (mole ratio) (B) methanol vapor, and (C) isopropyl alcohol vapor. Error bars are  $2\sigma$  obtained from Gaussian fits to the ionogram peaks. (D) The ionogram recorded for the  $m/z$  112 peak in a pure  $N_2$  environment with  $SV = 3500$  V (highlighted green in A). (Inset) The three lowest energy tautomers of  $(C + H)^+$  as calculated at the CCSD(T)/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory. Energies are reported as standard Gibbs' energies in  $\text{kJ mol}^{-1}$ .

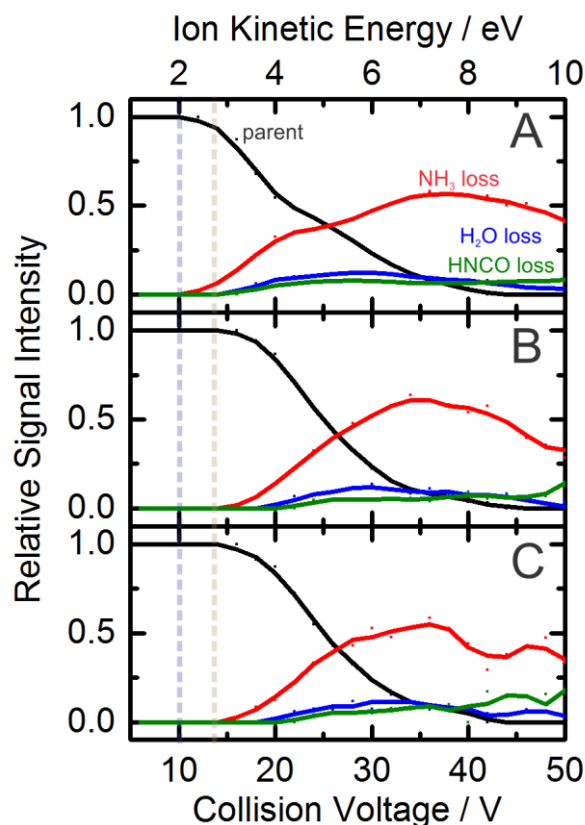
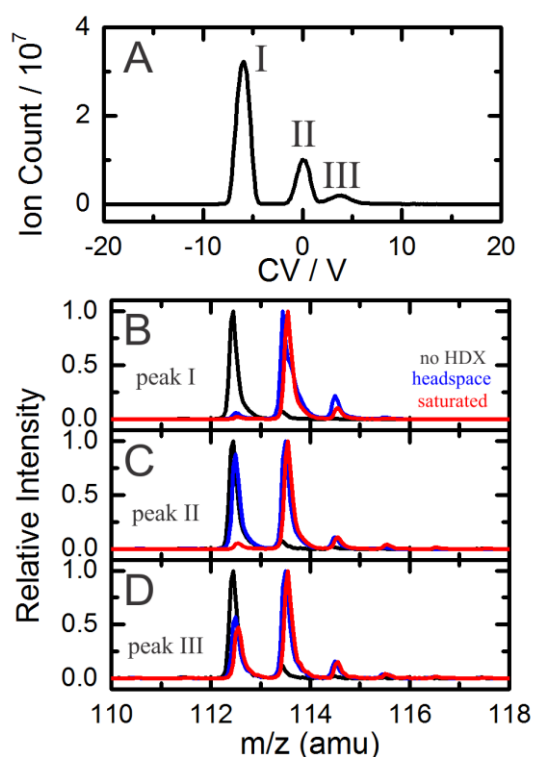


Figure 5.3 The breakdown curves obtained for  $(C + H)^+$  ( $m/z$  112) when isolating on the ionogram peaks plotted in Figure 5.2D. (A) peak I ( $CV = -6$  V), (B) peak II ( $CV = 1$  V), and (C) peak III ( $CV = 4$  V). Collision energy was increased in 2.5 V increments from 0 to 50 V. Ion kinetic energy is calculated with respect to the center of mass frame for the collision partners.

To further probe the separated species, the relative rates of HDX for each  $(C + H)^+$  structure were assessed.<sup>79,117</sup> Figure 5.4 plots the results of the HDX experiments for  $(C + H)^+$ . It is clear from Figures 5.4B-D that the  $(C + H)^+$  species associated with the three peaks in the ionogram all exhibit different HDX behavior. Although  $(C + H)^+$  has four exchangeable protons, in all three cases we see little evidence of HDX of the third and fourth nuclei. When gating the DMS on peak I, rapid exchange of one H atom is observed under both the high and low  $D_2O$  partial pressure

conditions. This accords with the expected acidity of the protonated carbonyl of the global minimum structure. The structure associated with peak II shows a relatively low rate of exchange for the first H atom, going to completion under high  $D_2O$  partial pressure, but remaining relatively unaffected in the low pressure  $D_2O$  environment. This might be evidence for saturation of the HDX environment under high partial pressure, or evidence of HDX-induced tautomerization.<sup>79</sup> The lower rate of HDX for peak II accords with the lower relative acidity expected for the protonated nitrogen atoms of isomer 2.

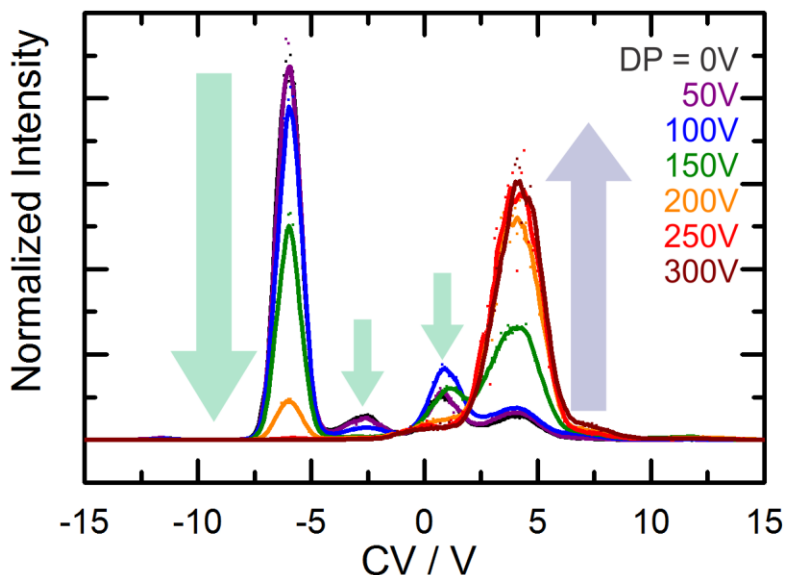


**Figure 5.4** (A) The ionogram recorded when gating on  $(C + H)^+$  ( $m/z$  112). The results of HDX experiments when isolating on (B) peak I, (C) peak II, and (D) peak III. The black traces show the observed mass distributions in the absence of HDX reagent. The blue traces show the effect of introducing a low vapor pressure of  $D_2O$ , and the red traces are observed following HDX in  $N_2$  at 18 °C seeded with a saturated partial pressure of  $D_2O$ .

The HDX profile of peak III is somewhat difficult to interpret in the context of peaks I and II. Unlike the species associated with peaks I and II, the structure responsible for peak III does not undergo complete exchange of the first proton in the saturated  $D_2O$  environment. This suggests that the rate of HDX for peak III is lower than that of peaks I and II. However, in the low pressure

D<sub>2</sub>O environment, the rate of HDX for peak III is higher than that of peak II. Moreover, the HDX profile for peak III is very nearly identical in both the high and low D<sub>2</sub>O partial pressure environments. This intriguing behavior led us to rethink the DMS sampling procedure, and specifically, the identity of the molecules giving rise to the (C + H)<sup>+</sup> (m/z 112) peak in the mass spectrum. We began by investigating ESI source conditions, but ultimately found that there was little variation in the observed ionogram across the parameter range that we explored (*e.g.*, solution flow rate, ESI voltage, etc.). However, significant variation of the ionogram was observed upon variation of the instrument declustering potential (DP). Figure 5.5 plots the ionogram recorded for C•H<sup>+</sup> (m/z 112) as DP is stepped from 0–300 V in 50 V increments. At DP = 0 V, four peaks are observed in the ionogram – a new, weak feature is observed at CV = –3 V. As DP is increased, peaks I and II, and the new feature at CV = –3 V deplete, indicating that the (C + H)<sup>+</sup> parent ion is fragmenting at the higher DP voltages. This behavior is expected since this voltage-ramp experiment can be viewed as a coarse version of CID; following selection by the DMS, the ions are accelerated/heated by the DP en route to the mass analyzer. The use of high DP voltages has previously been employed to activate ions prior to MS analysis in a form of ersatz “in-source MS/MS” fragmentation.<sup>118</sup> Interestingly, peak III, which showed the anomalous HDX behavior, *grows* in intensity with increasing DP voltage. We attribute this behavior to the fragmentation of larger clusters to produce (C + H)<sup>+</sup>. This is supported by the mass spectra at low DP voltages where signals are observed at masses corresponding to (C + H)<sup>+</sup> clustered with, *e.g.*, water and formic acid. If a larger cluster fragments post-DMS to yield (C + H)<sup>+</sup>, the trajectory of that larger cluster will appear in the (C + H)<sup>+</sup> ionogram and dispersion plot. Consequently, care must be taken to ensure that species separated by the DMS cell are attributed to the correct parent ion. This is an important consideration that has gone unaddressed in DMS-based studies to date. Ion behavior as a function of DP seems to be a satisfactory means of distinguishing target

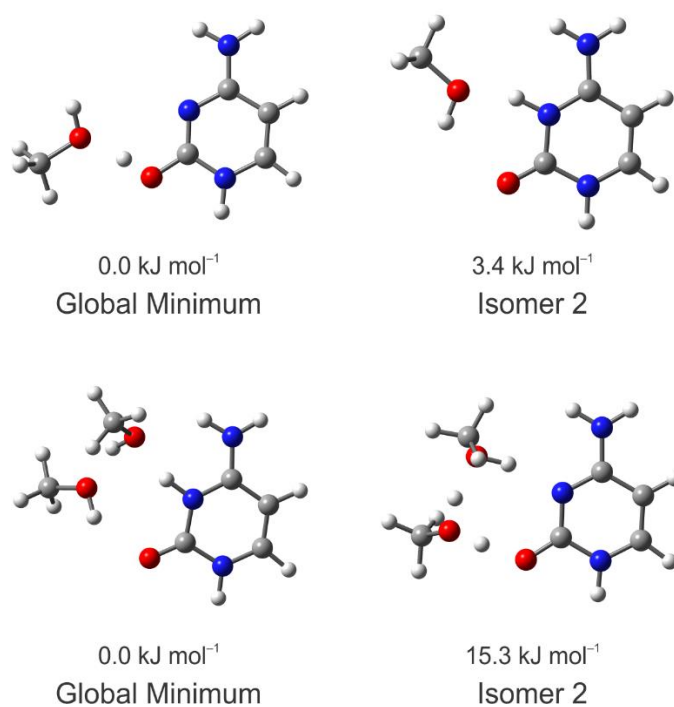
ions from larger clusters which contain the ions of interest. Note that there is a slight initial increase in the signals attributed to the bare  $(C + H)^+$  ions at low DP voltages; we interpret this depletion at low DP voltages to clustering with trace amounts solvent vapor in the region between the DMS and the mass spectrometer (see Figure 5.1).



**Figure 5.5** The ionogram recorded for  $C\bullet H^+$  ( $m/z$  112) in a pure  $N_2$  environment at  $SV = 3000$  V as the declustering potential, DP, is stepped from 0–300 V in 50 V increments.

Based on the analysis described above, we assign peaks I and II, and the weak feature at  $CV = -3$  V (observed at low DP) to the three lowest energy isomers of  $(C + H)^+$  (shown in Figure 5.2; highlighted with green downward arrows in Figure 5.5). These species are completely resolved by DMS, and they exhibit different breakdown curves and HDX profiles. Our calculations also suggest that the two lowest energy tautomers observed here are the same structures that were identified by Salpin *et al.* with IRMPD spectroscopy.<sup>104</sup> Note that these two species differ structurally by a simple proton transfer between adjacent proton acceptor sites. This proton transfer process is easily facilitated by a relay mechanism via an intermolecular hydrogen-bonding network with a protic solvent molecule.<sup>79,119–121</sup> For example, the two lowest energy

proton-bound cytosine-methanol complexes,  $(C + H)^+ \cdot (MeOH)_n$  ( $n = 1, 2$ ), are shown in Figure 5.6. Given that these structures are likely to be present in a methanol-modified DMS environment, and that similar structures are likely to be present in the high partial pressure HDX environment, interconversion of isomers 1 and 2 is likely to occur prior to MS characterization under these conditions. This interpretation is supported by the loss in signal intensity for the weaker features in the alcohol-modified environments. Similar observations were previously reported in our study of protonated 4-aminobenzoic acid.<sup>79</sup>



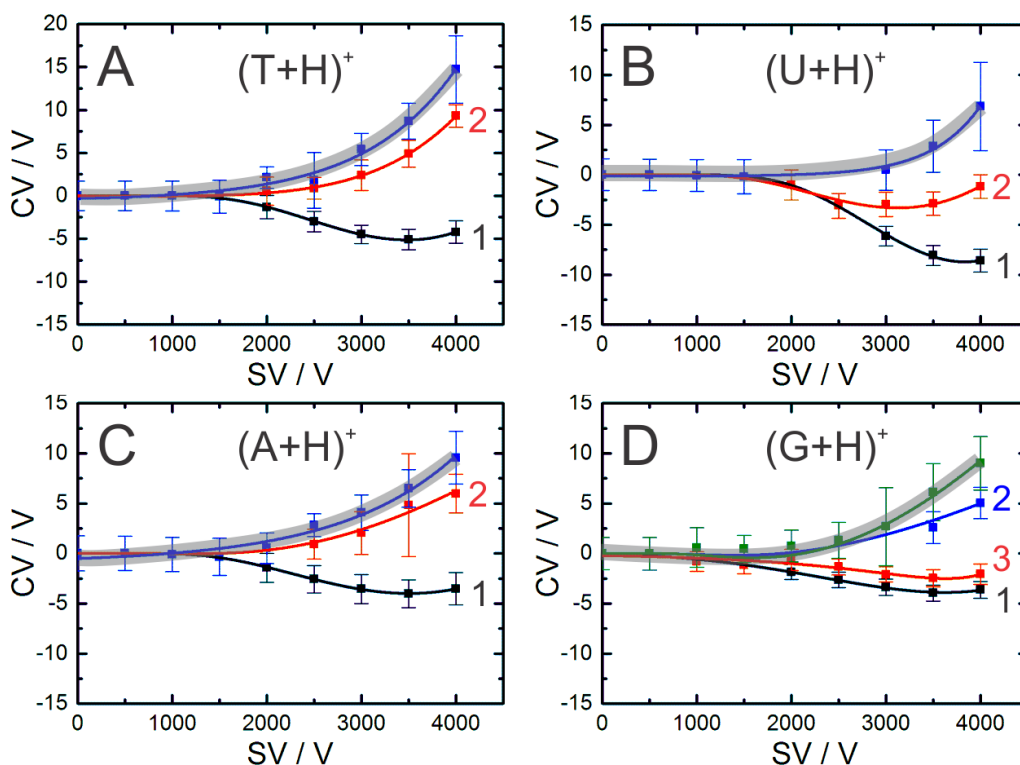
**Figure 5.6** The lowest energy proton-bound cytosine-methanol clusters,  $(C + H)^+ \cdot MeOH$  and  $(C + H)^+ \cdot (MeOH)_2$ . Proton transfer along the intermolecular hydrogen-bond network could facilitate interconversion of the two lowest energy isomers of  $(C + H)^+$ . Standard Gibbs' energies were calculated at the B3LYP/6-311++G(d,p) level of theory.

### 5.3.2 $(A + H)^+$ , $(T + H)^+$ , $(U + H)^+$ , and $(G + H)^+$

Having established a method to clearly identify bare- and clustered-ion signal in the DMS data, we proceeded to conduct analogous studies for  $(A + H)^+$ ,  $(T + H)^+$ ,  $(U + H)^+$ , and  $(G + H)^+$ . The dispersion plots that were recorded for these species in an unmodified  $N_2$  environment are



plotted in Figure 5.7. The dispersion plots for the MeOH- and IPA-modified environments, the DP-scans, the HDX data, and the breakdown curves for these species are all available in the supporting information that accompanies this manuscript. The greyed-out traces in Figure 5.7 are associated with larger clusters, which produce the ion of interest upon fragmentation post-DMS cell. When running with a declustering potential of  $DP = 150 \text{ V}$  (the standard instrument setting), we can clearly resolve two tautomers for  $(A + H)^+$ ,  $(T + H)^+$ , and  $(U + H)^+$ , and three tautomers for  $(G + H)^+$  (see supporting information for additional details). This accords with the work of Salpin *et al.*, who identified contributions from two tautomers in the IRMPD spectra of  $(T + H)^+$ , and  $(U + H)^+$ . Moreover, as was the case with  $(C + H)^+$ , trace amounts of a third tautomeric species are observed for  $(A + H)^+$ ,  $(T + H)^+$ , and  $(U + H)^+$  under low declustering potential conditions (see DP scans in supporting information). This suggests that  $(C + H)^+$ ,  $(A + H)^+$ ,  $(T + H)^+$ , and  $(U + H)^+$  all have two tautomeric forms at relatively low energy, and a third at higher energy which fragments under low-to-moderate declustering potential voltages. In the case of  $(G + H)^+$ , a single tautomer accounts for most (*ca.* 80%) of the total ion signal. Two other weaker features persist to higher DP voltages (*ca.* 150-200 V), indicating the presence of two additional higher-energy tautomers in the ensemble (see Figure S26). At low DP, these two features exhibit intensities that are approximately equal, but the structure associated with the  $CV = -1 \text{ V}$  ionogram peak (at  $SV = 3500 \text{ V}$ ) depletes at a substantially higher rate as DP is increased.

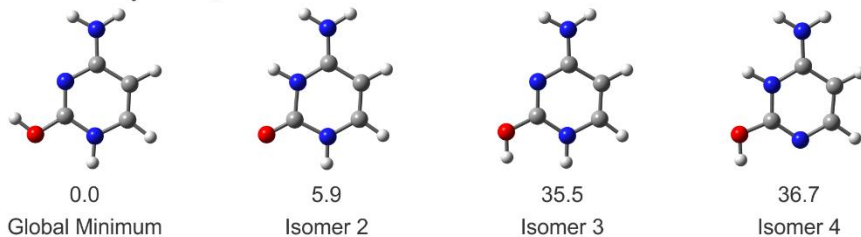


**Figure 5.7** The dispersion plots obtained for (A)  $(T + H)^+$  ( $m/z$  117), (B)  $(U + H)^+$  ( $m/z$  113), (C)  $(A + H)^+$  ( $m/z$  136), and (D)  $(G + H)^+$  ( $m/z$  152) for a pure  $N_2$  environment with DP set to 150 V. Curves that are greyed-out are associated with larger clusters which fragment to produce the ion of interest post-DMS. The numeric labels indicate the tautomer associated with a particular dispersion plot (see Figure 5.2).

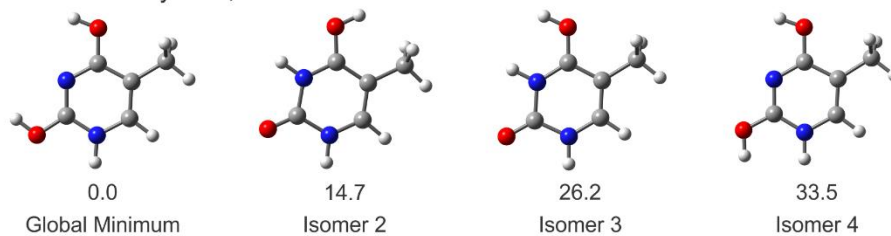
To estimate the relative standard Gibbs' energies of the various protonated nucleobase tautomers, electronic structure calculations were undertaken at the CCSD(T)/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory. The four lowest energy tautomers of the protonated nucleobases are shown in Figure 5.8. As expected based on the experimental results,  $(C + H)^+$ ,  $(T + H)^+$ , and  $(U + H)^+$  exhibit two tautomers at relatively low energy, which can interconvert via solvent-mediated proton-transfer between adjacent basic sites on the nucleobase. The third lowest energy tautomer for these species lies at least 20  $\text{kJ mol}^{-1}$  above the global minimum structure. This suggests that the higher energy tautomers are kinetically trapped during the ESI process,<sup>110,120</sup> since they are expected to have a negligible contribution to the ensemble population in a stochastic Boltzmann distribution at the experimental temperature and pressure (see Table 1). In the case of  $(A + H)^+$ , calculations suggest that three low-energy

tautomers are likely to be observed in the probed ensemble. Our DMS experiments clearly separate two structures. At low declustering potential we observe an asymmetry/skewing to the low-CV side of the dominant ionogram peak (see Figure S25), which suggests the presence of a third, higher-energy tautomer, but attempts to fully resolve this signal have been unsuccessful. For  $(G + H)^+$ , our calculations indicate that the population distribution should be dominated by a single tautomer (as was observed experimentally). The second and third lowest energy tautomers of  $(G + H)^+$  are calculated to lie  $16.9 \text{ kJ mol}^{-1}$  and  $19.1 \text{ kJ mol}^{-1}$  above the global minimum, respectively. This suggests that the two weak features observed in the  $(G + H)^+$  ionogram are metastable species which were kinetically trapped during production. To compare the calculated fractional populations to those observed experimentally, the ionogram peaks were fit to Gaussian distributions and peak areas were extracted. The relative population percentages of the lowest energies tautomers for each protonated nucleobase are reported in Table 1. These values are provided for experiments where DP was set to 0 V and 150 V to illustrate how sensitive nucleobase tautomer populations are to the declustering potential. Although experiment and theory are in relatively good agreement when it comes to  $(U + H)^+$  tautomer populations, for the most part there are significant differences between the observed and calculated tautomer populations of the protonated nucleobases. This, taken together with the tautomer population variability as a function of DP, indicates that the gas phase ensembles of the protonated nucleobases are generated and trapped in non-equilibrium conditions, and that these populations can be manipulated post-production via instrument conditions. It is also worth noting that, in the case of  $(T + H)^+$ , there is a significant difference between the relative tautomer populations at DP = 0 V and DP = 150 V (see Table 1). This is likely due to loss of the global minimum tautomer signal due to ion-solvent clustering at low DP since the expected relative populations are re-established at DP = 150 V.

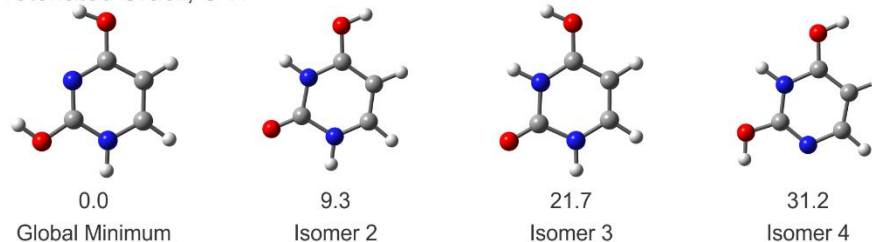
Protonated Cytosine, C•H<sup>+</sup>



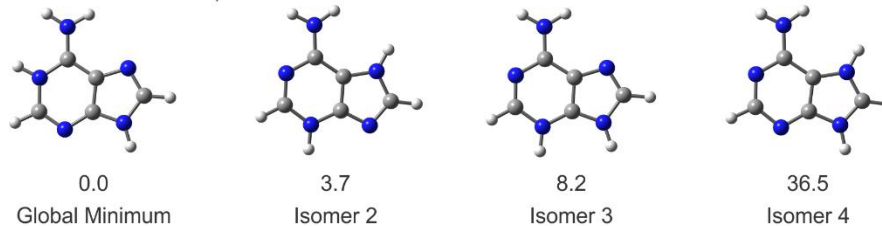
Protonated Thymine, T•H<sup>+</sup>



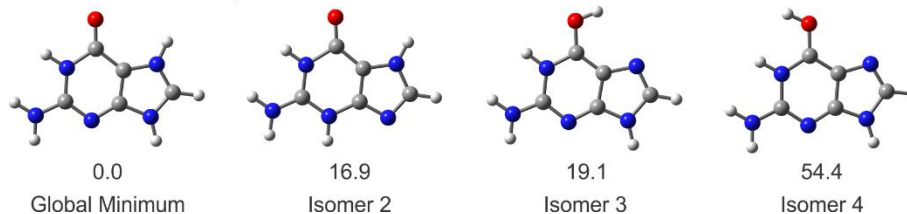
Protonated Uracil, U•H<sup>+</sup>



Protonated Adenine, A•H<sup>+</sup>



Protonated Guanine, G•H<sup>+</sup>



**Figure 5.8** The four lowest energy tautomeric forms of (C + H)<sup>+</sup>, (T + H)<sup>+</sup>, (U + H)<sup>+</sup>, (A + H)<sup>+</sup>, and (G + H)<sup>+</sup>. Electronic energies were calculated at the CCSD(T)/6-311++G(d,p) level of theory, and thermochemical corrections were calculated at the B3LYP/6-311++G(d,p) level of theory. Standard Gibbs' energies are reported in kJ mol<sup>-1</sup>.

**Table 5.1** The percentage of the ensemble population of the three lowest energy tautomeric forms of (C + H)<sup>+</sup>, (T + H)<sup>+</sup>, (U + H)<sup>+</sup>, (A + H)<sup>+</sup>, and (G + H)<sup>+</sup>. Experimental populations were determined by Gaussian fits of the ionogram distributions recorded at DP = 0 V and at DP = 150 V. Errors (1  $\sigma$ ) are reported in parentheses. The calculated population percentages are based on relative standard Gibbs' energies which were calculated at the CCSD(T)//B3LYP level of theory employing a 6-311++G(d,p) basis set.

Species	Percentage of Population		
	DP = 0 V	DP = 150 V	Calculated
<u>(C + H)<sup>+</sup></u>			
Global Minimum	79.5 (0.4)	73.5 (0.5)	91.5
Tautomer 2	8.1 (0.5)	26.5 (0.9)	8.5
Tautomer 3	12.4 (0.5)	0.0	0.0
<u>(A + H)<sup>+</sup></u>			
Global Minimum	73.6 (4.2)	100.0	79.3
Tautomer 2	17.7 (1.2)	0.0	17.8
Tautomer 3	8.7 (4.1)	0.0	2.9
<u>(G + H)<sup>+</sup></u>			
Global Minimum	74.7 (0.4)	79.5 (0.4)	99.8
Tautomer 2	10.8 (4.5)	20.5 (1.3)	0.1
Tautomer 3	14.6 (1.8)	trace	0.0
<u>(T + H)<sup>+</sup></u>			
Global Minimum	30.9 (0.5)	86.6 (0.9)	99.7
Tautomer 2	39.4 (1.4)	13.4 (1.0)	0.3
Tautomer 3	29.7 (1.5)	0.0	0.0
<u>(U + H)<sup>+</sup></u>			
Global Minimum	91.1 (0.3)	94.6 (0.3)	97.7
Tautomer 2	6.2 (0.3)	5.4 (0.4)	2.3
Tautomer 3	2.7 (0.5)	0.0	0.0

## 5.4 Conclusions

Differential mobility spectrometry has been used to separate the various tautomers present in gas phase ensembles of (C + H)<sup>+</sup>, (T + H)<sup>+</sup>, (U + H)<sup>+</sup>, (A + H)<sup>+</sup>, and (G + H)<sup>+</sup>. We find that these populations are dominated by contributions from a single tautomeric species for each protonated nucleobase, but the observation of weaker features in the DMS data indicates that higher-energy, metastable species are present in the sample. The unique identity of these tautomers has been confirmed via differences in the observed HDX profiles and breakdown curves, and structures are assigned based on CCSD(T)//B3LYP calculations. The observation of multiple protonated nucleobase tautomers is in accordance with an IRMPD spectroscopy study by Salpin *et al.*, who identified contributions from at least two tautomeric forms of (C + H)<sup>+</sup>, (T

+ H)<sup>+</sup>, and (U + H)<sup>+</sup>.<sup>104</sup> Supporting electronic structure calculations yield results for relative tautomer energies which are in general agreement with the observed populations, and we find the same low energy structures for (C + H)<sup>+</sup>, (T + H)<sup>+</sup>, and (U + H)<sup>+</sup> as were determined spectroscopically.<sup>104</sup> This suggests that DMS could find use as a pre-filter to separate tautomeric species prior to laser interrogation, thereby facilitating deconvolution of spectra. Indeed, similar methodology has already been employed in studies of isomeric lipid and saccharide species.<sup>122,123</sup>

It is, however, important to note that the DMS data for all five protonated nucleobases exhibited spurious ion signals that arose from fragmentation of larger nucleobase-containing clusters following DMS-separation and prior to MS detection. These species could not be distinguished from the bare protonated nucleobases simply by monitoring the ion signals corresponding to the bare protonated nucleobase or its product ions. Instead, we could identify contributions from larger clusters by monitoring ion signal as post-DMS DP voltages were ramped. The ion signals of the bare protonated nucleobases increased slightly at low DP voltages (DP < 50 V), and then decreased dramatically in intensity as the DP voltage was ramped up to 300 V. It should be noted that depletion of the various tautomer signals did not occur at the same rate, presumably due to relative differences in fragmentation energies. We are currently investigating this more detail. In contrast, the ion signals arising from fragmentation of larger clusters continued to increase in intensity well into the DP ramp, depleting only at very high voltages (DP > 200 V). This is an important consideration for future experiments wherein DMS is used to separate isomers, conformers, or tautomers prior to spectroscopic or mass spectrometric characterization, since ion populations that are produced from declustering processes are likely to yield convoluted results owing to contributions from multiple structures.

## 6 Concluding Remarks

Cluster chemistry provides a bridge between molecules and the condensed phase. The study of biomolecule clusters allows for the modelling and interpretation of complex biological processes, and thus is an example of the practicality of cluster chemistry. Three distinct subprojects evaluating the physicochemical properties of isolated biomolecule clusters were herein described. Topics of study were selected based on their relevance and potential impact in the drug discovery and development process. The three subprojects examine (a) the role several complex organic ligands may play in the promotion or disruption of guanine quadruplex structure, (b) the solvent clustering behavior of biologically and pharmaceutically relevant ions, and (c) the differential mobility spectrometry behavior of protonated nucleobase tautomers. A joint computational and experimental approach was taken in each study, and a brief description of these methodologies is provided in Chapter 2.

Chapter 3 discusses the effect several complex organic ligands may have on the promotion or disruption of G-quadruplex structure. G-quadruplexes are structures commonly found in guanine-rich sequences of eukaryotic DNA, namely the telomeric region of chromosomes. Due to the ability of G-quadruplexes to inhibit the activity of telomerase, an enzyme largely associated with conferring immortality to cancer cells, a great deal of effort is being shown toward their study. Specifically, the identification and characterization of ligands which selectively bind and stabilize these structures, is currently of interest. This study involved both computational predictions and experimental data to explore the interactions of a number of ligands with G-quadruplexes. Following a search of the potential energy surface of each quadruplex-binding ligand, and subsequent optimization at the density functional level of theory, the global minimum structure of each compound was found. Good agreement between the calculated and experimental IRMPD spectra of these ligands allowed for the assignment of

the observed IR peaks to vibrational normal modes. Guanine-ligand interactions were then evaluated computationally, and the resulting data yielded predictions for the location and strength of binding of these ligands to G-quadruplexes. A number of mass spectrometry and circular dichroism spectroscopy experiments were also performed in an effort to establish the existence of the antiparallel tetramolecular G-quadruplex. Although the specific conditions that yield this previously unobserved topology were determined, further studies must be performed to determine the exact structure.

Chapter 4 discusses the potential application of differential mobility spectrometry in the drug discovery process. Here, the clustering behavior of pharmaceutically relevant drug-like molecules was examined both computationally and experimentally. A series of low-level (molecular mechanics) and high-level (density functional theory) calculations were performed in order to determine the lowest-energy structure of each acrylamide molecule. Sites of protonation or deprotonation were also determined, as was the binding strength of each compound when clustered with a solvent molecule. Experimentally, the DMS clustering behavior was determined in a variety of DMS cell environments. When these data were treated with a Random Forest supervised learning algorithm, a number of properties evaluated in the drug development process were quantitatively predicted. Application of this technique in the drug discovery process will aim to increase both the accuracy and efficiency of current methodologies.

The final project is discussed in Chapter 5. Here, we use protonated nucleobase molecules as a model to further the development of a fundamental description of DMS theory and application. In this study, individual tautomeric forms of protonated nucleobases are isolated with DMS, and characterized using hydrogen-deuterium exchange and collision-induced dissociation experiments. Upon electrospray ionization of the protonated nucleobase solutions, a number of tautomeric forms of each molecule are produced. Density functional theory



calculations determine the lowest-energy isomers of each species, and allow for the assignment of peaks within the observed ionograms. We demonstrate the necessity of additional characterization following DMS isolation; minor ion signals typically associated with higher-level tautomeric species, were found to be attributed to larger nucleobase-containing clusters which fragment post-DMS. Findings of this study elicit important considerations that should be taken in future experiments using DMS as an isolation method for closely-related species.

The diversity of the projects explored in this thesis illustrate the role cluster chemistry plays in bridging the study of small molecules and complex biological processes. The joint computational and experimental approach to each study allowed for a comprehensive analysis and interpretation of data. Results of these subprojects each have unique implications for the future of drug discovery and development.

## References

1. Luo, Z. & Castleman, A. W. Special and general superatoms. *Acc. Chem. Res.* **47**, 2931–2940 (2014).
2. Cox, A. J., Louderback, J. G., Apsel, S. E. & Bloomfield, L. A. 4d-transition metal clusters. *Phys. Rev. B* **49**, (1994).
3. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366 (1988).
4. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* **19**, 414–422 (2009).
5. Neidle, S. Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **277**, 1118–1125 (2010).
6. Rodriguez, R. *et al.* A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **130**, 15758–15759 (2008).
7. Gomez, D. *et al.* Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. *J. Biol. Chem.* **281**, 38721–38729 (2006).
8. Pagano, B. *et al.* Looking for efficient G-quadruplex ligands: Evidence for selective stabilizing properties and telomere damage by drug-like molecules. *ChemMedChem* **10**, 640–649 (2015).
9. Singh, J., Petter, R. C., Baillie, T. a & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
10. Noe, M. C. & Gilbert, A. M. *Targeted Covalent Enzyme Inhibitors. Annual Reports in Medicinal Chemistry* **47**, (Elsevier Inc., 2012).
11. Kalgutkar, A. S. & Dalvie, D. K. Drug discovery for a new generation of covalent drugs. *Expert Opin. Drug Discov.* **7**, 561–581 (2012).
12. Mah, R., Thomas, J. R. & Shafer, C. M. Drug discovery considerations in the development of covalent inhibitors. *Bioorganic Med. Chem. Lett.* **24**, 33–39 (2014).
13. Johnson, D. S., Weerapana, E. & Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med. Chem.* **2**, 949–964 (2010).
14. Krishnan, S. *et al.* Design of reversible, cysteine-targeted michael acceptors guided by kinetic and computational analysis. *J. Am. Chem. Soc.* **136**, 12624–12630 (2014).
15. Jöst, C., Nitsche, C., Scholz, T., Roux, L. & Klein, C. D. Promiscuity and selectivity in covalent enzyme inhibition: A systematic study of electrophilic fragments. *J. Med. Chem.* **57**, 7590–7599 (2014).
16. Poltev, V. Molecular Mechanics: Method and Applications. *Handb. Comput. Chem.* 259–291 (2012). doi:10.1007/978-94-007-0711-5\_9
17. Allinger, N. L., Li, F. & Yan, L. Molecular mechanics. The MM3 force field for alkenes. *J.*

- Comput. Chem.* **11**, 848–867 (1990).
18. Box, V. G. S. The Molecular Mechanics of Quantized Valence Bonds. *J. Mol. Model.* **3**, 124–141 (1997).
  19. Vanommeslaeghe, K., Guvench, O., MacKerell, A. D. & Jr. Molecular mechanics. *Curr. Pharm. Des.* **20**, 3281–92 (2014).
  20. Car, R. & Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
  21. Orio, M., Pantazis, D. A. & Neese, F. Density functional theory. *Photosynth. Res.* **102**, 443–453 (2009).
  22. Kohn, W. & Sham, L. J. Density Functional Theory. *Phys. Rev.* **140**, A1133 (1965).
  23. Engel, E. & Dreizler, R. M. *Density Functional Theory. Perspectives in Electronic Structure Theory* (2011). doi:10.1007/978-3-642-14090-7
  24. Kristyán, S. & Pulay, P. Can (semi) local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.* **229**, 175–180 (1994).
  25. Zhao, Y. & Truhlar, D. G. Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: The MPW1B95 and MPWB1K models and comparative assessments for hydrogen bonding and van der Waals interactions. *J. Phys. Chem. A* **108**, 6908–6918 (2004).
  26. Tirado-Rives, J. & Jorgensen, W. L. Performance of B3LYP density functional methods for a large set of organic molecules. *J. Chem. Theory Comput.* **4**, 297–306 (2008).
  27. 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. *J. Phys. Chem. A* **101**, 5111–5116 (1997).
  28. Wales, D. J. & Scheraga, H. a. Global Optimization of Clusters, Crystals, and Biomolecules. *Science (80-. )*. **285**, 1368–1372 (1999).
  29. Håkansson, K. *et al.* Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptide to yield complementary sequence information. *Anal. Chem.* **73**, 4530–4536 (2001).
  30. Wu, R. & McMahon, T. B. Infrared multiple photon dissociation spectroscopy as structural confirmation for GlyGlyGlyH<sup>+</sup> and AlaAlaAlaH<sup>+</sup> in the gas phase. Evidence for amide oxygen as the protonation site. *J. Am. Chem. Soc.* **129**, 11312–11313 (2007).
  31. Fridgen, T. D., Macaleese, L., McMahon, T. B., Lemaire, J. & Maitre, P. Gas phase infrared multiple-photon dissociation spectra of methanol, ethanol and propanol proton-bound dimers, protonated propanol and the propanol/water proton-bound dimer. *Phys. Chem. Chem. Phys.* **8**, 955–66 (2006).
  32. Anderson, O. L., Isaak, D. G. & Yamamoto, S. Anharmonicity and the equation of state for gold. *J. Appl. Phys.* **65**, 1534–1543 (1989).

33. Kong, C. L. Combining rules for intermolecular potential parameters. II. Rules for the Lennard-Jones (12-6) potential and the Morse potential. *J. Chem. Phys.* **59**, 2464-2467 (1973).
34. Girifalco, L. A. & Weizer, V. G. Application of the morse potential function to cubic metals. *Phys. Rev.* **114**, 687-690 (1959).
35. Zhou, Y., Karplus, M., Ball, K. D. & Berry, R. S. The distance fluctuation criterion for melting: Comparison of square-well and Morse potential models for clusters and homopolymers. *J. Chem. Phys.* **116**, 2323-2329 (2002).
36. Eyler, J. R. Infrared multiple photon dissociation spectroscopy of ions in penning traps. *Mass Spectrom. Rev.* **28**, 448-467 (2009).
37. Oomens, J., Sartakov, B. G., Meijer, G. & von Helden, G. Gas-phase infrared multiple photon dissociation spectroscopy of mass-selected molecular ions. *Int. J. Mass Spectrom.* **254**, 1-19 (2006).
38. Ghodke, A. D., Sasaki, S., Miyata, K. & Takada, T. Operation and extension to far-infrared of the CLIO FEL facility.
39. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751**, 119-139 (2005).
40. Petersen, C. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **98**, 11624-11627 (1964).
41. Martin, S. R. & Bayley, P. M. Absorption and circular dichroism spectroscopy. *Methods Mol Biol* **173**, 43-55 (2002).
42. Marchand, A. *et al.* Ligand-induced conformational changes with cation ejection upon binding to human telomeric DNA G-quadruplexes. *J. Am. Chem. Soc.* **137**, 750-756 (2015).
43. Karsisiotis, A. I. *et al.* Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chemie - Int. Ed.* **50**, 10645-10648 (2011).
44. Campbell, J. L., Zhu, M. & Hopkins, W. S. Ion-molecule clustering in differential mobility spectrometry: Lessons learned from tetraalkylammonium cations and their isomers. *J. Am. Soc. Mass Spectrom.* **25**, 1583-1591 (2014).
45. Liu, C. *et al.* Using differential mobility spectrometry to measure ion solvation: An examination of the roles of solvents and ionic structures in separating quinoline-based drugs. *Analyst* **140**, 6897-6903 (2015).
46. 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. *Anal. Chem.* **84**, 7857-7864 (2012).
47. Hopkins, W. S. Determining the properties of gas-phase clusters. *Mol. Phys.* **113**, 3151-3158 (2015).
48. Krylov, E. V., Nazarov, E. G. & Miller, R. A. Differential mobility spectrometer: Model of operation. *Int. J. Mass Spectrom.* **266**, 76-85 (2007).

49. Shvartsburg, A. A. *Differential Ion Mobility: Non-linear Ion Transport and Fundamentals of FAIMS*. (CRC Press, 2009).
50. Levin, D. S., Vouros, P., Miller, R. A., Nazarov, E. G. & Morris, J. C. Characterization of gas-phase molecular interactions on differential mobility ion behavior utilizing an electrospray ionization-differential mobility-mass spectrometer system. *Anal. Chem.* **78**, 96–106 (2006).
51. Schneider, B. B., Covey, T. R., Coy, S. L., Krylov, E. V. & Nazarov, E. G. Chemical effects in the separation process of a differential mobility/Mass spectrometer system. *Anal. Chem.* **82**, 1867–1880 (2010).
52. Schneider, B. B., Covey, T. R. & Nazarov, E. G. DMS-MS separations with different transport gas modifiers. *Int. J. Ion Mobil. Spectrom.* **16**, 207–216 (2013).
53. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415 (2006).
54. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* **59**, 871–880 (1989).
55. Bhattacharyya, D., Mirihana Arachchilage, G. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **4**, 38 (2016).
56. Azargun, M. *et al.* Guanine tetrads: an IRMPD spectroscopy, energy resolved SORI-CID, and computational study of M(9-ethylguanine)<sub>4</sub><sup>+</sup> (M = Li, Na, K, Rb, Cs) in the gas phase. *Phys. Chem. Chem. Phys.* **17**, 25778–25785 (2015).
57. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36**, 5482–5515 (2008).
58. Han, H. & Hurley, L. H. G-quadruplex DNA: A potential target for anti-cancer drug design. *Trends Pharmacol. Sci.* **21**, 136–142 (2000).
59. Parkinson, G. N., Lee, M. P. H. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–80 (2002).
60. Frisch, M. J. *et al.* Gaussian 09. *Gaussian Inc.* (2009). doi:10.1159/000348293
61. Koirala, D., Dhakal, S., Ashbridge, B., Sannohe, Y. & Rodriguez, R. Europe PMC Funders Group Europe PMC Funders Author Manuscripts A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. **3**, 782–787 (2012).
62. Bertrand, H. *et al.* Recognition of G-quadruplex DNA by triangular star-shaped compounds: With or without side chains? *Chem. - A Eur. J.* **17**, 4529–4539 (2011).
63. De Cian, A., DeLemos, E., Mergny, J. L., Teulade-Fichou, M. P. & Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.* **129**, 1856–1857 (2007).
64. Granotier, C. *et al.* Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.* **33**, 4182–4190 (2005).
65. Cavaluzzi, M. J. & Borer, P. N. Revised UV extinction coefficients for nucleoside-5'-

- monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.* **32**, e13 (2004).
66. Andersson, M. P. & Uvdal, P. New scale factors for harmonic vibrational frequencies using the B3LYP density functional method with the triple- $\zeta$  basis Set 6-311+G(d,p). *J. Phys. Chem. A* **109**, 2937–2941 (2005).
  67. Valerie Gabelica and Edwin De Pauw, F. R. A Simple Method to Determine Electrospray Response Factors of Noncovalent Complexes. *Anal. Chem.* **81**, 6708–6715 (2009).
  68. Flanagan, M. E. *et al.* Chemical and computational methods for the characterization of covalent reactive groups for the prospective design of irreversible inhibitors. *J. Med. Chem.* **57**, 10072–10079 (2014).
  69. Potashman, M. H. & Duggan, M. E. Covalent Modifiers : An Orthogonal Approach to Drug Design Perspective Covalent Modifiers : An Orthogonal Approach to Drug Design. **52**, (2009).
  70. Lanning, B. R. *et al.* HHS Public Access. **10**, 760–767 (2015).
  71. Uetrecht, J. Immune-mediated adverse drug reactions. *Chem. Res. Toxicol.* **22**, 24–34 (2009).
  72. Zhang, X., Liu, F., Chen, X., Zhu, X. & Uetrecht, J. Involvement of the immune system in idiosyncratic drug reactions. *Drug Metab. Pharmacokinet.* **26**, 47–59 (2011).
  73. Dahal, U. P., Obach, R. S. & Gilbert, A. M. Benchmarking in vitro covalent binding burden as a tool to assess potential toxicity caused by nonspecific covalent binding of covalent drugs. *Chem. Res. Toxicol.* **26**, 1739–1745 (2013).
  74. Bauman, J. N. *et al.* Can in vitro metabolism-dependent covalent binding data distinguish hepatotoxic from nonhepatotoxic drugs? An analysis using human hepatocytes and liver S-9 fraction. *Chem. Res. Toxicol.* **22**, 332–340 (2009).
  75. Shintaro Nakayama, Ryo Atsumi, Hideo Takakusa, Yoshimasa Kobayashi, A. K. & Yoko Nagai, Daisuke Nakai, and O. O. A Zone Classification System for Risk Assessment of Idiosyncratic Drug Toxicity Using Daily Dose and Covalent Binding. *Drug Metab. Dispos.* **37**, 1970–1977 (2009).
  76. Liu, C. *et al.* Assessing Physicochemical Properties of Drug Molecules via Microsolvation Measurements with Differential Mobility Spectrometry. *ACS Cent. Sci.* **3**, 101–109 (2017).
  77. Purves, R. W. & Guevremont, R. Electrospray ionization high-field asymmetric waveform ion mobility spectrometry-mass spectrometry. *Anal. Chem.* **71**, 2346–2357 (1999).
  78. Schneider, B. B., Covey, T. R., Coy, S. L., Krylov, E. V & Nazarov, E. G. NIH Public Access. **298**, 45–54 (2010).
  79. Campbell, J. L., Yang, A. M. C., Melo, L. R. & Hopkins, W. S. Studying Gas-Phase Interconversion of Tautomers Using Differential Mobility Spectrometry. *J. Am. Soc. Mass Spectrom.* **27**, 1277–1284 (2016).
  80. Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**, 86–91 (1996).

81. Demšar, J. *et al.* Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **14**, 23492353 (2013).
82. Buerger, H. B. & Dunitz, J. D. From crystal statics to chemical dynamics. *Acc. Chem. Res.* **16**, 153–161 (1983).
83. Perrin, C. L. & Nielson, J. B. 'STRONG' HYDROGEN BONDS. (1997).
84. Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed.* **41**, 49–76 (2002).
85. Kubota, M. & Kobayashi, T. Electronic structure of uracil and uridine derivatives studied by photoelectron spectroscopy. *J. Electron Spectros. Relat. Phenomena* **82**, 61–70 (1996).
86. Kwiatkowski, J. S. & Leszczynski, J. Molecular structure and vibrational IR spectra of the 2-hydroxypyridine 2(1H)-pyridinone system and its thio and seleno analogs: Density functional theory versus conventional ab initio calculations. *J. Mol. Struct.* **376**, 325–342 (1996).
87. Smets, J. & Adamowicz, L. Matrix-Isolation FT-IR Studies and. 6387–6400 (1995).
88. Sowers, L. C., Fazakerley, G. V., Eritja, R., Kaplan, B. E. & Goodman, M. F. Base pairing and mutagenesis: observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 5434–5438 (1986).
89. Baei, M. T., Taghartapeh, M. R., Lemeski, E. T. & Soltani, A. A computational study of adenine, uracil, and cytosine adsorption upon AlN and BN nano-cages. *Phys. B Condens. Matter* **444**, 6–13 (2014).
90. Dong, C., Yang, J., Ning, H. & Li, C. Studies on structures, energetics, and electron affinities of As-nucleobases and their anions with density functional theory. *J. Mol. Struct. THEOCHEM* **950**, 64–71 (2010).
91. Fornaro, T., Biczysko, M., Monti, S. & Barone, V. Dispersion corrected DFT approaches for anharmonic vibrational frequency calculations: nucleobases and their dimers. *Phys. Chem. Chem. Phys.* **16**, 10112–28 (2014).
92. Panigrahi, S., Bhattacharya, A., Banerjee, S. & Bhattacharyya, D. Interaction of nucleobases with wrinkled graphene surface: Dispersion corrected DFT and AFM studies. *J. Phys. Chem. C* **116**, 4374–4379 (2012).
93. Sinden, R. R. *DNA Structure and Function*. (Academic Press, Inc., 1994).
94. Frank-Kamenetskii, M. D. & Mirkin, S. M. Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95 (1995).
95. Chandra, A. K., Michalska, D., Wysokisky, R. & Zeegers-Huyskens, T. Theoretical study of the acidity and basicity of the cytosine tautomers and their 1:1 complexes with water. *J. Phys. Chem. A* **108**, 9593–9600 (2004).
96. Di Donna, L., Napoli, A., Sindona, G. & Athanassopoulos, C. A comprehensive evaluation of the kinetic method applied in the determination of the proton affinity of the nucleic acid molecules. *J. Am. Soc. Mass Spectrom.* **15**, 1080–1086 (2004).

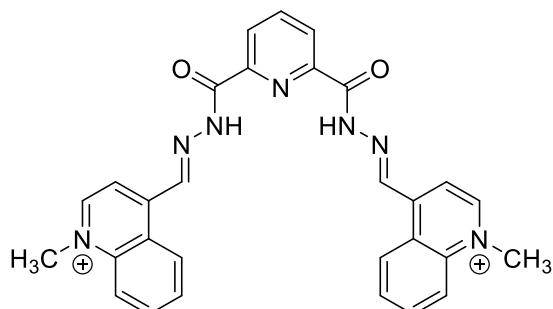
97. Kryachko, E. S., Nguyen, M. T. & Zeegers-Huyskens, T. Theoretical Study of Tautomeric Forms of Uracil. 1. Relative Order of Stabilities and Their Relation to Proton Affinities and Deprotonation Enthalpies. *J. Phys. Chem. A* **105**, 1288–1295 (2001).
98. Kurinovich, M. A., Phillips, L. M., Sharma, S. & Lee, J. K. The gas phase proton affinity of uracil: measuring multiple basic sites and implications for the enzyme mechanism of orotidine 5'-monophosphate decarboxylase. *Chem. Commun. (Camb)*. 2354–5 (2002).
99. Moustafa, H., El-Taher, S., Shibl, M. F. & Hilal, R. Equilibrium geometry and gas-phase proton affinity of 2-thiouracil derivatives. *Int. J. Quantum Chem.* **87**, 378–388 (2002).
100. Tho Nguyen, M., Chandra, A. K. & Zeegers-Huyskens, T. Protonation and deprotonation energies of uracil Implications for the uracil–water complex. *J. Chem. Soc. Faraday Trans.* **94**, 1277–1280 (1998).
101. Russo, N., Toscano, M., Grand, A. & Jolibois, F. Protonation of thymine, cytosine, adenine, and guanine DNA nucleic acid bases: Theoretical investigation into the framework of density functional theory. *J. Comput. Chem.* **19**, 989–1000 (1998).
102. Colominas, C., Luque, F. J. & Orozco, M. Tautomerism and protonation of guanine and cytosine. Implications in the formation of hydrogen-bonded complexes. *J. Am. Chem. Soc.* **118**, 6811–6821 (1996).
103. Podolyan, Y., Gorb, L. & Leszczynski, J. Energetics and Proton Affinities. 7346–7352 (2000).
104. Salpin, J. Y. *et al.* Infrared spectra of protonated uracil, thymine and cytosine. *ChemPhysChem* **8**, 2235–2244 (2007).
105. Bakker, J. M., Salpin, J. Y. & Maître, P. Tautomerism of cytosine probed by gas phase IR spectroscopy. *Int. J. Mass Spectrom.* **283**, 214–221 (2009).
106. Bakker, J. M. *et al.* Tautomerism of uracil probed via infrared spectroscopy of singly hydrated protonated uracil. *J. Phys. Chem. A* **112**, 12393–12400 (2008).
107. Karpas, Z., Berant, Z. & Stimac, R. M. An ion mobility spectrometry/mass spectrometry (IMS/MS) study of the site of protonation in anilines. *Struct. Chem.* **1**, 201–204 (1990).
108. Lalli, P. M. *et al.* Protomers: Formation, separation and characterization via travelling wave ion mobility mass spectrometry. *J. Mass Spectrom.* **47**, 712–719 (2012).
109. Schröder, D., Buděšínský, M. & Roithová, J. Deprotonation of p-hydroxybenzoic acid: Does electrospray ionization sample solution or gas-phase structures? *J. Am. Chem. Soc.* **134**, 15897–15905 (2012).
110. Xia, H. & Attygalle, A. B. Effect of Electrospray Ionization Source Conditions on the Tautomer Distribution of Deprotonated p-Hydroxybenzoic Acid in the Gas Phase. *Anal. Chem.* **88**, 6035–6043 (2016).
111. Eiceman, G. A. & Karpas, Z. *Ion Mobility Spectrometry*. (CRC Press, 2014).
112. Vekey, K. Internal Energy Effects in Mass Spectrometry.pdf. *J. Mass Spectrom.* **31**, 445–463 (1996).



113. Schneider, B. B., Nazarov, E. G., Londry, F. & Covey, T. R. Comparison of the peak capacity for DMS filters with various gap height: experimental and simulations results. *Int. J. Ion Mobil. Spectrom.* **18**, 159–170 (2015).
114. Cheng, P., Li, Y., Li, S., Zhang, M. & Zhou, Z. Collision-induced dissociation (CID) of guanine radical cation in the gas phase: an experimental and computational study. *Phys. Chem. Chem. Phys.* **12**, 4667–4677 (2010).
115. Nelson, C. C. & McCloskey, J. A. Collision-Induced Dissociation of Adenine. *J. Am. Chem. Soc.* **114**, 3661–3668 (1992).
116. Nelson, C. C. & McCloskey, J. A. Collision-Induced Dissociation of Uracil and Its Derivatives. *J. Am. Soc. Mass Spectrom.* **5**, 339–349 (1994).
117. Blagojevic, V. & Bohme, D. K. Differential mobility spectrometer as an ion/molecule reactor: Peptide H-D exchange in mobility separation. *Int. J. Mass Spectrom.* **378**, 180–185 (2015).
118. Buckley, J. A., French, J. B. & Reid, N. M. A multi-purpose trace atmospheric gas analyser. *Can. Aeronaut. Sp. J.* **20**, 231–233 (1974).
119. de Grotthuss, C. J. T. Sur la decomposition de l'eau et des corps qu'elle tient en dissolution a l'aide de l'electrcite galvanique. *Ann. Chim.* **58**, 54–73 (1806).
120. Scott Hopkins, W., Marta, R. a., Steinmetz, V. & McMahon, T. B. Mode-specific fragmentation of amino acid-containing clusters. *Phys. Chem. Chem. Phys.* **17**, 28548–28555 (2015).
121. Kaila, V. R. I. & Hummer, G. Energetics and dynamics of proton transfer reactions along short water wires. *Phys. Chem. Chem. Phys.* **13**, 13207 (2011).
122. Levin, D. S., Vouros, P., Miller, R. A. & Nazarov, E. G. Using a Nanoelectrospray-Differential Mobility Spectrometer-Mass Spectrometer System for the Analysis of Oligosaccharides with Solvent Selected Control Over ESI Aggregate Ion Formation. *J. Am. Soc. Mass Spectrom.* **18**, 502–511 (2007).
123. Lintonen, T. P. I. *et al.* Differential Mobility Spectrometry-Driven Shotgun Lipidomics. *J. Am. Chem. Soc.* **86**, 9662–9669 (2014).

## Appendix I: G-Quadruplexes – Supporting Information

### Structural Formulae of HK21 Ligand



### XYZ Coordinates of Optimized Geometries

Coordinate data of lowest energy isomer of each quadruplex-binding ligand, as calculated at the B3LYP/6-311++G(d,p) level of theory.

#### 360A

C	-5.480821	-4.081971	1.328271
C	-4.393121	-3.247432	1.294881
C	-4.472886	-1.985461	0.647507
C	-5.707235	-1.607314	0.036356
C	-6.815630	-2.477384	0.079963
C	-6.694687	-3.692615	0.716990
H	-5.417121	-5.042268	1.824808
H	-3.459970	-3.535284	1.764532
H	-7.756622	-2.206272	-0.376223
H	-7.547123	-4.360196	0.751052
C	-4.717832	0.458827	-0.625671
C	-3.484909	0.131445	-0.026632
C	-3.383648	-1.098129	0.604985
N	-5.764238	-0.371883	-0.584694
H	-2.456549	-1.388214	1.088746
H	-4.852542	1.406214	-1.122519
N	-2.395310	1.006366	-0.059931
H	-1.510695	0.676376	0.305407
C	-7.025579	0.068864	-1.230701
H	-6.870878	1.055879	-1.656740
H	-7.294374	-0.630814	-2.021412
H	-7.818536	0.116704	-0.484992
C	-2.422434	2.343110	-0.411407
O	-3.418196	2.905756	-0.825638
C	-1.125313	3.080852	-0.210115
C	-1.171650	4.476583	-0.236549
C	-0.000023	5.185274	-0.000033
C	1.171633	4.476636	0.236505
C	1.125348	3.080903	0.210116
N	0.000031	2.383745	0.000012
H	-0.000044	6.268497	-0.000051
H	-2.114053	4.970906	-0.431048
H	2.114017	4.971003	0.430984
C	5.480654	-4.082006	-1.328303
C	4.392999	-3.247409	-1.294886
C	4.472842	-1.985449	-0.647498

C	5.707219	-1.607377	-0.036361
C	6.815567	-2.477506	-0.079995
C	6.694550	-3.692722	-0.717037
H	5.416894	-5.042293	-1.824851
H	3.459825	-3.535204	-1.764526
H	7.756580	-2.206450	0.376180
H	7.546949	-4.360348	-0.751119
C	4.717933	0.458809	0.625707
C	3.484981	0.131506	0.026675
C	3.383652	-1.098059	-0.604949
N	5.764293	-0.371958	0.584706
H	2.456532	-1.388092	-1.088699
H	4.852704	1.406177	1.122571
N	2.395400	1.006458	0.059981
H	1.510782	0.676463	-0.305346
C	7.025665	0.068711	1.230704
H	6.871021	1.055730	1.656757
H	7.294430	-0.630991	2.021405
H	7.818617	0.116517	0.484988
C	2.422498	2.343217	0.411415
O	3.418232	2.905925	0.825632

### PhenDC3

C	0.646451	3.824971	-0.337117
C	1.200826	5.073905	-0.743083
C	2.402832	5.056639	-1.478929
C	3.034595	3.859316	-1.723884
C	2.460565	2.695615	-1.183677
N	1.305183	2.665274	-0.529937
H	2.829802	5.990215	-1.827673
H	3.963999	3.794492	-2.273274
C	-0.647338	3.824827	0.337201
C	-1.202127	5.073663	0.742906
C	-2.404151	5.056154	1.478717
C	-3.035544	3.858676	1.723873
C	-2.461119	2.695055	1.183916
N	-1.305706	2.664957	0.530220
H	-2.831436	5.989662	1.827259
H	-3.964952	3.793662	2.273236
C	0.565006	6.303954	-0.374758
C	-0.566701	6.303844	0.374346
C	3.243928	1.417732	-1.276540
C	-3.244082	1.416942	1.276972
N	2.792856	0.426150	-0.418609
N	-2.792799	0.425454	0.419043
O	4.215901	1.290821	-1.998583
O	-4.216273	1.289987	1.998714
H	1.977650	0.697036	0.125051
H	-1.977569	0.696478	-0.124513
C	4.539044	-1.226241	-0.865175
C	3.467981	-0.743613	-0.084613
C	3.118967	-1.470756	1.043824
C	3.815575	-2.638596	1.399931
C	4.905411	-3.079556	0.589486
N	5.215562	-2.330106	-0.531735
H	2.308788	-1.129368	1.679865
H	4.853981	-0.693817	-1.748788
C	6.341532	-2.742958	-1.404301
H	6.423193	-2.034505	-2.223511
H	7.267627	-2.739531	-0.830218
H	6.150125	-3.738997	-1.802144
C	5.628854	-4.237585	0.942279

C	3.486585	-3.388566	2.561149
C	5.275663	-4.935436	2.075764
C	4.201764	-4.511108	2.891344
C	-4.538826	-1.227232	0.865256
C	-3.467726	-0.744379	0.084878
C	-3.118452	-1.471383	-1.043567
C	-3.814836	-2.639298	-1.399865
C	-4.904721	-3.080482	-0.589610
N	-5.215137	-2.331169	0.531628
H	-2.308235	-1.129825	-1.679469
H	-4.853964	-0.694930	1.748869
C	-3.485570	-3.389125	-2.561097
C	-5.627945	-4.238588	-0.942601
C	-4.200537	-4.511745	-2.891486
C	-5.274491	-4.936295	-2.076093
C	-6.341170	-2.744256	1.404001
H	-6.423050	-2.035883	2.223258
H	-7.267184	-2.740916	0.829788
H	-6.149674	-3.740300	1.801788
H	1.022649	7.236572	-0.684200
H	-1.024652	7.236376	0.683594
H	6.457006	-4.584153	0.341060
H	5.834351	-5.822751	2.347729
H	3.951042	-5.075973	3.780688
H	2.666563	-3.051310	3.184160
H	-2.665509	-3.051697	-3.183964
H	-3.949606	-5.076500	-3.780840
H	-5.833010	-5.823669	-2.348210
H	-6.456132	-4.585327	-0.341529

## PDS

C	-3.810372	-5.396452	-0.345148
C	-3.004042	-4.286448	-0.263289
C	-3.569958	-2.990725	-0.176503
C	-4.992334	-2.860523	-0.166743
C	-5.798121	-4.020257	-0.262322
C	-5.217272	-5.264225	-0.349222
H	-3.368588	-6.383346	-0.414683
H	-1.924322	-4.366380	-0.268113
H	-6.876263	-3.925888	-0.289511
H	-5.837621	-6.148749	-0.428823
C	-5.496966	-1.534830	-0.065850
C	-4.655963	-0.451525	-0.007098
C	-3.259483	-0.720018	-0.042153
N	-2.734921	-1.918278	-0.113564
H	-4.990981	0.571384	0.060648
O	-6.873386	-1.425167	-0.027960
C	-7.477398	-0.139630	0.035290
C	-8.973392	-0.401915	-0.042292
N	-9.341496	-1.451358	0.994427
H	-9.291589	-1.084490	1.949824
H	-8.661474	-2.220487	0.919190
H	-10.284838	-1.823089	0.850820
H	-9.563331	0.493007	0.152046
H	-9.247871	-0.821364	-1.009804
H	-7.186134	0.489611	-0.811125
H	-7.201092	0.375419	0.963540
N	-2.298921	0.307976	-0.000284
H	-1.343920	-0.031932	-0.026159
C	-2.460458	1.656288	0.037533
O	-3.529862	2.250921	0.067413
C	-1.152214	2.430928	0.036130

C	-1.235770	3.826388	0.041004
C	-0.045180	4.542713	0.034058
C	1.169404	3.861512	0.023411
C	1.139012	2.471436	0.020829
N	0.001410	1.763300	0.026217
H	-2.220365	4.272784	0.048533
H	2.120802	4.376282	0.024561
O	0.025916	5.919418	0.037620
C	-1.174750	6.683801	0.126800
C	-0.728310	8.125000	0.308236
N	0.263622	8.488135	-0.784915
H	-1.563996	8.822024	0.253410
H	-0.201607	8.256696	1.253077
H	-1.783566	6.555799	-0.776229
H	-1.774510	6.397103	0.995658
H	-0.180758	8.541251	-1.707158
H	0.985158	7.756678	-0.826156
H	0.717019	9.389365	-0.606156
C	4.050025	-5.279605	-0.002740
C	3.207502	-4.194418	0.036978
C	3.728703	-2.877319	0.017268
C	5.145037	-2.699114	-0.035764
C	5.988246	-3.834692	-0.087713
C	5.450109	-5.100588	-0.070629
H	3.642121	-6.283305	0.010709
H	2.131978	-4.310976	0.080481
H	7.060527	-3.705615	-0.162171
H	6.098812	-5.967013	-0.116112
C	5.604827	-1.353532	-0.037803
C	4.728091	-0.297978	-0.013179
C	3.341592	-0.614321	0.021075
N	2.858061	-1.831813	0.043402
H	5.028475	0.737773	-0.018873
O	6.976899	-1.194162	-0.061084
C	7.532751	0.113535	-0.113998
C	9.033185	-0.099091	-0.238617
N	9.485059	-1.054626	0.854183
H	9.459669	-0.619824	1.781634
H	10.435858	-1.400904	0.696852
H	8.834107	-1.852050	0.865298
H	9.284165	-0.578388	-1.184471
H	9.594440	0.828825	-0.135377
H	7.275488	0.683317	0.787360
H	7.182077	0.669913	-0.988383
N	2.348650	0.382719	0.035157
H	1.405209	0.012109	0.053299
C	2.469523	1.736944	0.013560
O	3.520501	2.361996	-0.009321

### TMPP4

C	-0.686584	-4.255411	0.046611
C	0.679448	-4.256598	0.044382
C	1.127535	-2.895391	-0.015411
C	-1.132493	-2.893437	-0.012072
N	-0.001834	-2.112555	-0.043144
C	2.455834	-2.438157	-0.012784
C	3.521389	-3.493975	-0.006896
C	3.746142	-4.314113	-1.122149
C	4.738194	-5.276620	-1.092539
C	4.334747	-3.702190	1.116767
C	5.307698	-4.684022	1.101831
N	5.502121	-5.455571	0.008521

C	6.588427	-6.475937	0.002370
C	2.871971	-1.089338	-0.012777
C	4.272580	-0.679170	-0.081552
C	4.273915	0.670815	-0.079456
C	2.874057	1.083506	-0.009666
N	2.048909	-0.002145	0.017296
C	2.460494	2.433175	-0.007172
C	3.528100	3.486811	0.003482
C	3.758340	4.308208	-1.109687
C	4.342075	3.687599	1.128125
C	5.322600	4.661748	1.115131
C	4.757836	5.263045	-1.078110
N	5.524370	5.432570	0.022545
C	6.566603	6.497646	0.047547
C	1.133053	2.892864	-0.012456
C	0.687147	4.254823	0.045986
C	-0.678895	4.256010	0.044156
C	-1.126994	2.894787	-0.015116
N	0.002393	2.111963	-0.043047
C	-2.455293	2.437532	-0.011637
C	-3.520823	3.493375	-0.005280
C	-3.746389	4.313123	-1.120586
C	-4.738395	5.275747	-1.090539
C	-4.333328	3.702011	1.118993
C	-5.306170	4.683874	1.104478
N	-5.501403	5.455115	0.011015
C	-6.587753	6.475441	0.006212
C	-2.459921	-2.433752	-0.006513
C	-2.873515	-1.084119	-0.008319
C	-2.871472	1.088720	-0.011137
N	-2.048361	0.001543	0.017996
C	-4.273430	-0.671484	-0.076380
C	-4.272152	0.678516	-0.078379
C	-3.527480	-3.487475	0.004409
C	-3.756932	-4.309704	-1.108375
C	-4.338141	-3.691536	1.130756
C	-5.312981	-4.671571	1.120795
C	-4.750688	-5.270262	-1.073780
C	-6.600052	-6.463656	0.028845
N	-5.511881	-5.445271	0.029831
H	-6.743158	-6.830311	1.042910
H	-6.313321	-7.288985	-0.619022
H	-7.518931	-6.003372	-0.336070
H	-6.718766	6.858289	1.015983
H	-7.511434	6.011325	-0.341300
H	-6.306916	7.289860	-0.657748
H	6.931539	6.660210	-0.964139
H	7.388364	6.174888	0.682955
H	6.128626	7.416263	0.439544
H	7.508256	-6.015387	-0.359804
H	6.730515	-6.848353	1.014492
H	6.300372	-7.297181	-0.650055
H	4.204849	-3.114226	2.016129
H	5.133905	-1.324206	-0.150283
H	5.136537	1.314326	-0.146378
H	3.163561	-4.198397	-2.027009
H	4.950244	-5.918428	-1.937751
H	4.210654	3.095696	2.024681
H	5.962665	4.853237	1.966583
H	4.975117	5.904686	-1.922057
H	3.178261	4.195453	-2.016519
H	1.324190	5.121861	0.105101
H	-1.314602	5.124154	0.101698

H	-3.164528	4.197077	-2.025868
H	-4.951006	5.917224	-1.935841
H	-5.943923	4.882595	1.955962
H	-4.202744	3.114327	2.018439
H	-5.133575	1.323548	-0.145910
H	-5.136041	-1.315127	-0.142032
H	-3.176690	-4.197119	-2.015131
H	-4.966335	-5.913600	-1.916921
H	-4.204640	-3.101759	2.028402
H	-5.949338	-4.866696	1.974125
H	1.315151	-5.124736	0.102066
H	-1.323558	-5.122476	0.106334
H	5.946083	-4.882334	1.952915
H	0.001639	1.098439	-0.048947
H	-0.001050	-1.099034	-0.049257

### TrisQ

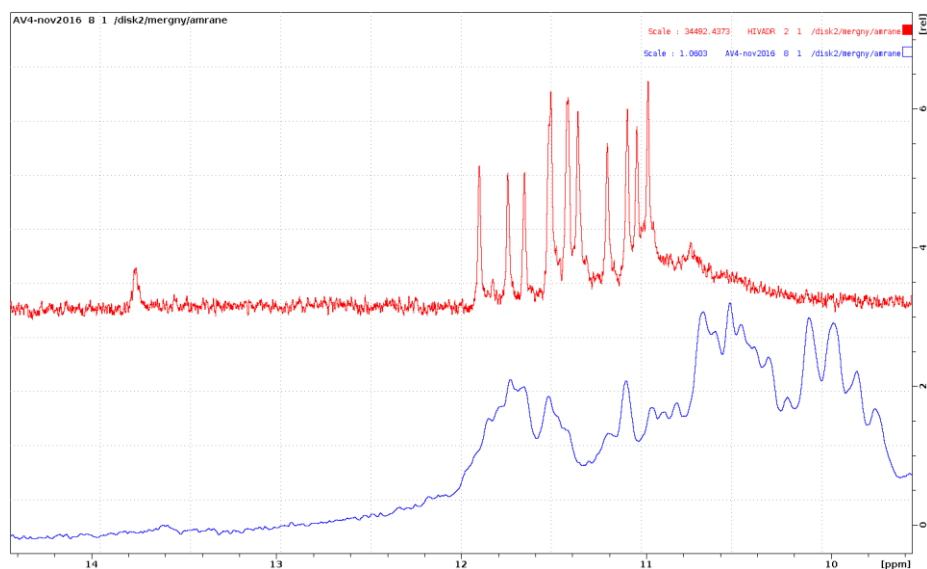
C	1.040126	1.027789	0.000076
C	1.397038	-0.405915	0.000022
C	0.369975	-1.414752	-0.000074
C	-1.050083	-1.006956	-0.000086
C	-1.410338	0.386903	-0.000006
C	-0.347067	1.412756	0.000061
C	1.648781	3.402734	0.000183
C	2.606701	4.437076	0.000247
C	2.217404	5.755096	0.000293
C	0.840048	6.069964	0.000274
C	-0.088101	5.071268	0.000211
N	0.299138	3.741320	0.000166
H	2.955925	6.547887	0.000344
H	3.654529	4.164402	0.000260
H	0.502345	7.098972	0.000309
H	-1.154099	5.252214	0.000194
C	1.983289	2.032878	0.000133
C	-0.655668	2.749704	0.000102
C	2.709157	-0.807054	0.000059
C	0.769033	-2.734061	-0.000143
C	-2.752338	0.701138	0.000005
C	-2.053649	-1.942646	-0.000159
C	2.122620	-3.129214	-0.000113
C	2.539627	-4.475936	-0.000182
C	3.875710	-4.797716	-0.000140
C	4.837026	-3.762302	-0.000026
C	4.436044	-2.459170	0.000040
N	3.090652	-2.129672	-0.000004
H	4.193054	-5.833672	-0.000194
H	1.779699	-5.247215	-0.000269
H	5.897041	-3.984264	0.000010
H	5.125674	-1.626386	0.000126
C	-3.771462	-0.273415	-0.000074
C	-5.146275	0.038762	-0.000070
C	-6.093015	-0.957420	-0.000153
C	-5.676849	-2.307662	-0.000241
C	-4.347832	-2.611924	-0.000244
N	-3.389782	-1.611528	-0.000160
H	-7.148859	-0.714312	-0.000150
H	-5.434264	1.082487	-0.000001
H	-6.399040	-3.114726	-0.000308
H	-3.971345	-3.625504	-0.000309
H	-1.869823	-3.005771	-0.000220
H	-3.100029	1.724356	0.000077
H	3.043255	1.822446	0.000139
H	-1.668326	3.121845	0.000083

H	0.057026	-3.546979	-0.000225
H	3.537774	-0.116182	0.000142

## NMR Spectra

The resulting NMR spectra (shown in blue) obtained for a solution containing 87.5  $\mu\text{mol/L}$  GGGTTA DNA, 35  $\text{mmol/L}$  TMAA, 8.75  $\text{mmol/L}$  KCl, and 262.5  $\mu\text{mol/L}$  PhenDC3 ligand.

Sample NMR spectra of a well-resolved quadruplex is shown for comparison (in red).

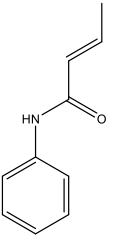
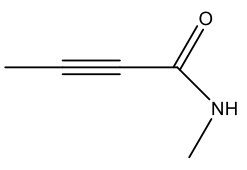
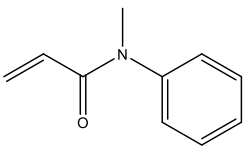
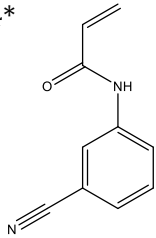
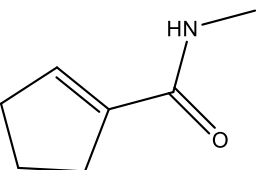
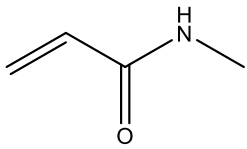
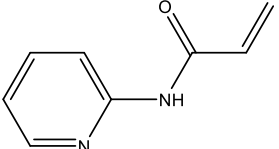
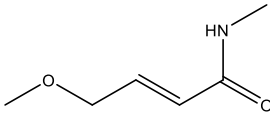
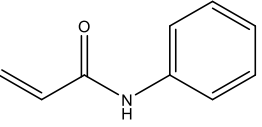
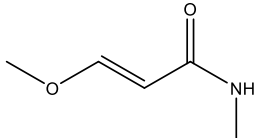
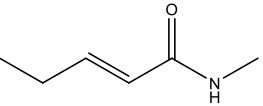
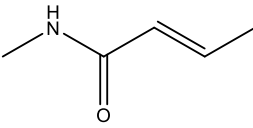
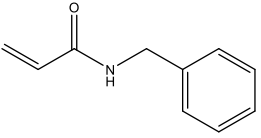
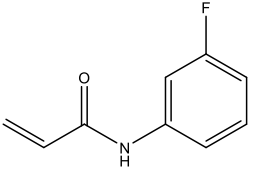
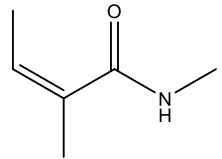
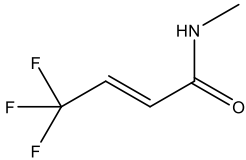
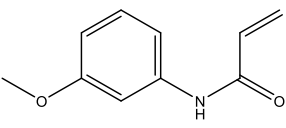
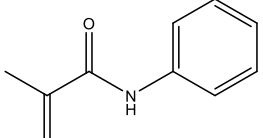
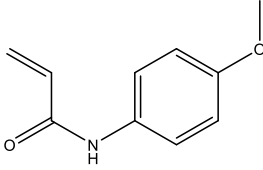
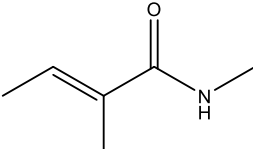
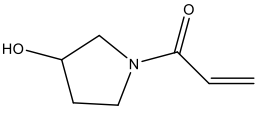
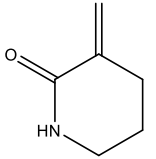
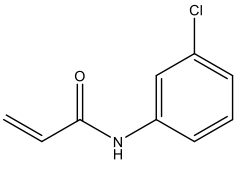
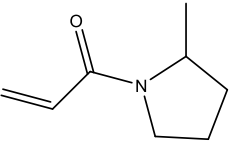




## Appendix II: Pharmaceutically Relevant Ions – Supporting Information

### Structural Formulae

Asterisks denote compounds that will be deprotonated; all others will be protonated.

A01 	A02 	A03 	A04* 
A05 	A06 	A07 	A08 
A09* 	A10 	A11 	A12 
B01 	B02* 	B03 	B04* 
B05 	B06 	B07 	B08 
B12 	C01 	C02* 	C03 

## Calculation Results

Thermochemical data for the protonated or deprotonated acrylamide CRGs, as calculated at the B3LYP/6-311++G(d,p) level of theory at a temperature of  $T = 298$  K and a pressure of  $P = 1$  atm.

Acrylamide CRG	Electronic Energy (hartrees)	Zero Point Corrected Energy (hartrees)	Gibbs Energy (hartrees)	Enthalpy (hartrees)
A01	-518.1733318	-517.9727518	-518.0115608	-517.9603158
A02	-325.1173493	-324.9936263	-325.0252413	-324.9854793
A03	-518.1493725	-517.9492685	-517.9885725	-517.9366465
A04*	-570.2032574	-570.0589554	-570.0972184	-570.0468064
A05	-403.8239084	-403.6392174	-403.6748384	-403.6286434
A06	-287.0451822	-286.9248782	-286.9566462	-286.9166682
A07	-494.9106571	-494.7488171	-494.7839031	-494.7384321
A08	-440.9350759	-440.7538199	-440.7915029	-440.7417739
A09*	-477.9198091	-477.7743961	-477.8098871	-477.7640511
A10	-401.6210212	-401.4684792	-401.5031262	-401.4581812
A11	-365.7096243	-365.5329683	-365.5696593	-365.5218493
A12	-326.3825199	-326.2347229	-326.2680199	-326.2249689
B01	-518.1573530	-517.9559270	-517.9950660	-517.9436290
B02*	-577.1958128	-577.0586478	-577.0954448	-577.0474518
B03	-365.7072028	-365.5307368	-365.5661678	-365.5197138
B04*	-623.2829060	-623.1854050	-623.2222640	-623.1744660
B05	-593.3952402	-593.1899512	-593.2292782	-593.1766012
B06	-518.1610511	-517.9600061	-517.9985031	-517.9477711
B07	-593.3960541	-593.1909801	-593.2311131	-593.1775191
B08	-365.7114059	-365.5351879	-365.5706359	-365.5240439
B12	-479.0538810	-478.8643780	-478.8998910	-478.8532940
C01	-364.4933826	-364.3357376	-364.3672466	-364.3273836
C02*	-937.5515230	-937.4155610	-937.4532990	-937.4039530
C03	-443.1411319	-442.9279879	-442.9634989	-442.9166409

*Note: Asterisks (\*) denote deprotonated compounds carrying a -1 charge; all other compounds were protonated and carried a +1 charge.*

Zero point corrected binding energies (equivalent to  $D_0$  dissociation energies), as calculated at the B3LYP/6-311++G(d,p) level of theory, according to equation (8) in the maintext.

Acrylamide CRG	Isopropanol BE (kcal/mol)	Methanol BE (kcal/mol)	Water BE (kcal/mol)
A01	6.0530822	5.6693014	5.5303722
A02	23.9519778	21.7047640	19.0634784
A03	18.1236044	16.7178094	14.6615448
A04*	11.5071500	11.3331434	10.6714415
A05	19.5797879	17.8694029	15.6863200
A06	21.6868802	19.9062777	17.3978972
A07	10.9477311	10.2197649	9.5681030
A08	19.9387823	18.4036590	16.1759606
A09*	12.6909344	12.4023458	11.6421887
A10	17.7293443	16.4150392	14.4011939
A11	19.7328358	18.1436845	15.8594481
A12	19.4211551	17.9657874	15.6815510
B01	19.7302003	18.1933200	15.8612051
B02*	12.0063287	11.8419856	11.1746362
B03	19.2761392	17.4848063	15.0542362
B04*	13.9911206	13.4815255	12.5326154
B05	17.4529919	16.0577390	13.4941383
B06	18.8773610	17.4448972	15.3427621
B07	17.3145020	15.9972477	13.5463465
B08	19.2793394	17.5898503	15.2888596
B12	18.3506977	16.8325170	5.4135939
C01	23.3350168	21.2928083	18.7415066
C02*	10.9818045	11.7798628	11.0823305
C03	18.5398271	17.1500335	15.0271281

*Note: Asterisks (\*) denote deprotonated compounds carrying a -1 charge; all other compounds were protonated and carried a +1 charge.*

## XYZ Coordinates of Optimized Geometries

Coordinate data of lowest energy isomer of each acrylamide CRG, as calculated at the B3LYP/6-311++G(d,p) level of theory.

### A01

C	3.129251	-1.227296	-0.240932
C	3.874478	-0.161631	0.261339
C	3.286141	1.091283	0.435026
C	1.945742	1.286091	0.112106
C	1.209808	0.211018	-0.388547
C	1.788783	-1.046174	-0.577078
H	3.590649	-2.196487	-0.385318
H	4.917884	-0.306256	0.514320

H	3.868948	1.918261	0.821681
H	1.478256	2.254926	0.244389
H	1.208930	-1.860055	-0.997752
N	-0.180539	0.412567	-0.742672
H	-0.394534	0.916493	-1.596163
C	-1.201908	0.001701	-0.008460
C	-2.561458	0.238382	-0.416805
H	-2.700635	0.759547	-1.357922
C	-3.615647	-0.157896	0.323460
H	-3.422992	-0.677169	1.259072
C	-5.036554	0.060332	-0.039444
H	-5.158442	0.583989	-0.987547
H	-5.557718	-0.902644	-0.089466
H	-5.537559	0.628171	0.753116
O	-0.988050	-0.631335	1.118949
H	-0.036431	-0.730920	1.304691

### A02

C	1.477765	1.798457	-0.000233
H	0.930602	2.080281	-0.899665
H	2.439918	2.305241	0.011964
H	0.909953	2.076879	0.887264
N	1.735942	0.345175	0.000366
H	2.706476	0.051845	-0.000533
O	1.081023	-1.866261	-0.000209
C	0.799755	-0.580075	0.000097
C	-0.560977	-0.251477	0.000341
C	-1.747912	-0.019257	0.000140
C	-3.166714	0.246027	-0.000145
H	-3.368788	1.319549	-0.001226
H	-3.628843	-0.203963	-0.884437
H	-3.628873	-0.202179	0.885038
H	2.028274	-2.075842	-0.000490

### A03

C	-0.665350	0.329556	0.003354
C	-1.327230	0.171927	-1.211722
C	-2.645670	-0.279407	-1.203751
C	-3.282874	-0.560136	0.004397
C	-2.606682	-0.391779	1.211970
C	-1.287781	0.058701	1.219233
H	-0.824066	0.393354	-2.145971
H	-3.171682	-0.410878	-2.141443
H	-4.308343	-0.909336	0.004769
H	-3.102598	-0.609562	2.149996
H	-0.755074	0.194412	2.153467
N	0.699845	0.846953	0.004964
C	1.744626	0.043200	-0.005859
O	2.970688	0.526982	0.021289
C	1.632362	-1.405425	-0.052477
H	0.632334	-1.811907	-0.096255
C	2.712360	-2.196528	-0.043693
H	3.719988	-1.801547	0.001708
H	2.600417	-3.273190	-0.082272
C	0.868712	2.318052	0.037737
H	1.332586	2.633890	0.976782
H	1.457195	2.660020	-0.818606
H	-0.117249	2.768879	-0.025675
H	3.017247	1.492380	0.063312

### A04\*

C	-2.615838	1.224110	-0.000015
---	-----------	----------	-----------

C	-2.463993	-0.176171	-0.000008
C	-1.195624	-0.758309	0.000027
C	-0.009298	0.021495	0.000072
C	-0.189354	1.435984	0.000031
C	-1.462223	2.002214	0.000015
H	-3.603676	1.668275	-0.000048
H	-1.087601	-1.836242	0.000022
H	0.695548	2.055048	0.000015
H	-1.554898	3.084938	-0.000004
C	-3.628597	-1.008904	-0.000037
N	-4.579570	-1.670522	-0.000049
N	1.169186	-0.679313	0.000069
C	2.367409	-0.062768	0.000009
C	3.494921	-1.063053	0.000034
H	3.194635	-2.106888	0.000169
C	4.782212	-0.711248	-0.000083
H	5.579219	-1.448749	-0.000043
H	5.052530	0.340111	-0.000206
O	2.639906	1.159031	-0.000039

### A05

C	-2.591856	-0.907593	0.229745
C	-1.096104	-1.255541	0.007748
C	-0.431177	0.112982	-0.001107
C	-1.338262	1.103700	-0.120409
C	-2.736819	0.577981	-0.184281
H	-2.838474	-1.017779	1.287243
H	-0.942162	-1.761228	-0.953696
H	-0.710015	-1.912056	0.792539
H	-1.122558	2.164433	-0.216389
H	-3.112616	0.695333	-1.210060
H	-3.420885	1.146565	0.451776
H	-3.254532	-1.564819	-0.331473
C	1.000787	0.313366	0.030503
N	1.837541	-0.689761	-0.089084
H	1.432140	-1.602614	-0.248725
C	3.304630	-0.577792	-0.062856
H	3.713790	-1.564192	0.143690
H	3.673648	-0.221791	-1.026314
H	3.605528	0.116757	0.719238
O	1.565947	1.494814	0.181589
H	0.918565	2.188587	0.366985

### A06

C	-2.519458	-0.067968	0.084366
H	-2.633751	0.997272	0.264703
H	-3.441519	-0.635409	0.052274
C	-1.341486	-0.678035	-0.072456
H	-1.294507	-1.751223	-0.221506
C	-0.057206	0.012559	-0.034909
O	0.066229	1.318514	-0.039392
N	1.062704	-0.661601	0.008037
H	0.987167	-1.670246	0.058751
C	2.411975	-0.071962	0.044080
H	3.115772	-0.821621	-0.310655
H	2.445066	0.802888	-0.602027
H	2.665594	0.217565	1.065485
H	-0.775534	1.776308	-0.174632

### A07

C	-0.669624	0.443476	-0.000031
C	-1.715973	1.374704	0.000139

C	-3.021417	0.919619	0.000103
C	-3.301784	-0.456710	-0.000104
C	-2.249292	-1.337044	-0.000176
N	-0.975213	-0.874910	-0.000169
H	-3.834398	1.635882	0.000233
H	-1.490884	2.433538	0.000309
H	-4.317609	-0.826489	-0.000182
H	-2.360060	-2.412903	-0.000303
N	0.653512	0.816116	-0.000197
H	0.836128	1.810519	0.000265
C	1.761856	-0.054852	0.000153
C	3.068549	0.619231	-0.000075
H	3.089165	1.704251	-0.000088
C	4.197276	-0.094239	-0.000265
H	4.169935	-1.178035	0.000004
H	5.166376	0.389124	-0.000306
O	1.591346	-1.265540	0.000504
H	-0.155066	-1.505114	0.000124

### A08

C	-2.260620	-0.791823	-0.101752
H	-2.694044	-1.407493	0.707168
H	-2.662252	-1.205726	-1.044034
O	-2.567661	0.561840	0.049359
C	-3.972030	0.837172	0.052662
H	-4.076226	1.913169	0.174959
H	-4.468070	0.326398	0.885082
H	-4.431909	0.531378	-0.893306
C	-0.785616	-0.992515	-0.096051
H	-0.471001	-2.027280	-0.223918
C	1.528427	-0.153176	0.034197
O	2.135784	-1.319331	0.105744
C	0.092387	0.014823	0.035169
H	-0.284821	1.027096	0.125925
N	2.325710	0.885174	-0.042367
H	1.882965	1.788778	-0.151110
C	3.795850	0.832953	-0.039882
H	4.134437	0.079386	0.668416
H	4.166834	1.811173	0.258043
H	4.166630	0.588105	-1.036912
H	1.522117	-2.045866	0.279375

### A09\*

C	-0.614869	-0.326877	-0.000095
C	-1.613449	-1.338566	0.000055
C	-2.969398	-1.043526	0.000128
C	-3.409490	0.285478	0.000057
C	-2.449159	1.299374	-0.000062
C	-1.084447	1.017295	-0.000127
H	-1.269929	-2.368096	0.000097
H	-3.693751	-1.854632	0.000229
H	-4.469419	0.520905	0.000108
H	-2.769005	2.339198	-0.000101
H	-0.350893	1.810041	-0.000185
N	0.692142	-0.763459	-0.000147
C	1.734056	0.085138	-0.000032
O	1.758062	1.339411	0.000140
C	3.043732	-0.663964	-0.000092
H	2.961893	-1.747443	-0.000295
C	4.234758	-0.060025	0.000112
H	4.286497	1.024464	0.000317
H	5.164712	-0.621485	0.000083

**A10**

C	-3.760171	0.242162	0.000002
H	-4.575396	-0.475094	0.000252
H	-3.796829	0.856299	-0.900491
H	-3.796647	0.856720	0.900216
O	-2.544206	-0.555949	0.000054
C	-1.395555	0.063013	0.000015
H	-1.403719	1.152013	-0.000006
C	-0.226716	-0.640412	0.000037
H	-0.252374	-1.722454	0.000065
C	1.030236	0.010973	0.000012
N	2.170483	-0.667100	-0.000062
H	2.097151	-1.675309	-0.000169
C	3.514212	-0.079042	-0.000091
H	3.681361	0.520684	0.898698
H	3.681457	0.520417	-0.899041
H	4.238205	-0.890630	0.000078
O	1.047343	1.334839	0.000079
H	1.936279	1.715774	-0.000095

**A11**

C	3.707854	0.224374	-0.552784
H	3.391302	0.129520	-1.593465
H	3.798143	1.286874	-0.313456
H	4.697879	-0.223656	-0.457356
C	2.725035	-0.488847	0.398878
H	2.675706	-1.557362	0.176798
H	3.096529	-0.387757	1.427159
C	1.362888	0.110946	0.353172
H	1.328991	1.177550	0.584553
C	0.230675	-0.554205	0.060543
H	0.264319	-1.618727	-0.146824
C	-1.082882	0.052299	0.022872
N	-2.163643	-0.680589	-0.103281
H	-2.035178	-1.683850	-0.125139
C	-3.538791	-0.159159	-0.161515
H	-3.576466	0.705731	-0.821573
H	-4.179848	-0.946587	-0.551081
H	-3.875926	0.129088	0.835579
O	-1.298670	1.347475	0.109185
H	-0.479277	1.861050	0.107288

**A12**

C	2.976646	-0.248028	0.000017
H	3.625291	-1.120819	-0.000186
H	3.192069	0.334910	0.899227
H	3.192045	0.335270	-0.898969
N	1.583442	-0.713338	-0.000056
H	1.420326	-1.712001	-0.000138
C	0.512552	0.057570	0.000011
C	-0.811992	-0.503365	-0.000007
H	-0.878846	-1.586000	-0.000033
O	0.626429	1.367380	0.000120
C	-1.922040	0.262232	0.000024
H	-1.804639	1.343091	0.000073
C	-3.309657	-0.258004	-0.000059
H	-3.848627	0.127920	0.873079
H	-3.848162	0.127219	-0.873813
H	-3.357637	-1.346810	0.000322
H	1.539594	1.689119	-0.000040

**B01**

C	-1.949027	-0.023707	-1.222624
C	-3.226768	-0.572430	-1.146136
C	-3.888247	-0.631399	0.079804
C	-3.274905	-0.138011	1.230395
C	-1.997104	0.411679	1.157557
C	-1.327317	0.466954	-0.069339
H	-1.444727	0.033168	-2.182159
H	-3.708156	-0.943767	-2.042802
H	-4.884935	-1.052545	0.136567
H	-3.793451	-0.171647	2.180973
H	-1.530433	0.808390	2.053712
C	0.067554	1.033415	-0.144904
H	0.245327	1.486479	-1.125476
H	0.221359	1.787561	0.633302
N	1.074443	-0.050542	0.053408
H	0.697869	-0.972551	0.247486
C	2.378402	0.072029	-0.002658
C	3.251541	-1.069988	0.203812
H	2.762170	-2.016417	0.403937
C	4.584219	-0.969970	0.151238
H	5.082587	-0.028858	-0.048359
H	5.207417	-1.841577	0.309270
O	2.948476	1.229262	-0.248279
H	2.325976	1.960025	-0.376943

**B02\***

C	0.270095	0.616087	0.000570
C	0.892302	-0.665756	0.000248
C	2.270589	-0.746912	-0.000098
C	3.124756	0.345104	-0.000204
C	2.519447	1.609746	-0.000350
C	1.140701	1.742951	-0.000081
H	0.278269	-1.553979	0.000060
H	4.199147	0.209500	-0.000689
H	3.145236	2.497880	-0.000631
H	0.676812	2.722954	-0.000248
N	-1.076319	0.887046	0.000675
C	-2.004026	-0.088401	0.000401
O	-1.863466	-1.332990	0.000350
C	-3.397958	0.486986	-0.000081
H	-3.455829	1.571840	0.000003
C	-4.499793	-0.266650	-0.000768
H	-4.409724	-1.348533	-0.000964
H	-5.494836	0.168682	-0.001254
F	2.845134	-2.001409	-0.000181

**B03**

C	2.897615	-0.430535	0.192061
H	2.951724	-0.989762	-0.740361
H	2.960542	-1.121900	1.033913
H	3.717294	0.282551	0.242351
N	1.634721	0.324301	0.244588
H	1.660938	1.268746	0.605187
C	0.460409	-0.146058	-0.097398
C	-0.742794	0.689151	-0.069061
O	0.467533	-1.400664	-0.497785
C	-1.980818	0.164299	0.072658
H	-2.789020	0.891050	0.048438
C	-2.444683	-1.239977	0.304669
H	-3.240136	-1.229373	1.054383
H	-2.907944	-1.646071	-0.604062



H	-1.678687	-1.931159	0.659249
C	-0.526964	2.185640	-0.168314
H	-1.484451	2.694144	-0.272904
H	-0.055681	2.593282	0.733805
H	0.091036	2.453705	-1.029191
H	-0.385516	-1.665131	-0.868333

#### B04\*

N	-2.705775	0.805180	-0.000163
C	-1.815799	-0.175654	-0.000480
C	-0.409697	0.357740	-0.000176
H	-0.304391	1.437306	-0.000125
C	0.657865	-0.446353	-0.000090
H	0.519747	-1.522994	-0.000166
O	-1.997584	-1.426123	0.000026
C	-4.079300	0.341925	0.000297
H	-4.314699	-0.283847	-0.877253
H	-4.314207	-0.283610	0.878150
H	-4.756151	1.204758	0.000385
C	2.055415	0.023386	0.000043
F	2.201730	1.372508	0.000374
F	2.767987	-0.430145	1.084766
F	2.768048	-0.429603	-1.084876

#### B05

C	0.106077	0.124677	-0.484838
C	-0.106476	1.507410	-0.454613
C	-1.397829	1.959810	-0.189376
C	-2.445996	1.069995	0.026271
C	-2.214449	-0.313609	-0.022714
H	-1.594265	3.024902	-0.169817
H	0.688732	2.204468	-0.690979
H	-3.436745	1.456023	0.221720
N	1.433934	-0.384079	-0.746706
H	1.559729	-1.001546	-1.541630
C	2.491300	-0.150760	0.007016
C	3.783987	-0.718854	-0.325460
H	3.825171	-1.333166	-1.217885
C	4.870863	-0.502919	0.424160
H	4.839134	0.111598	1.315980
H	5.822512	-0.943791	0.153900
O	2.402947	0.591321	1.078555
C	-0.916905	-0.787751	-0.274775
H	-0.742560	-1.856622	-0.291940
O	-3.146909	-1.260582	0.163700
C	-4.504467	-0.876618	0.428639
H	-5.051199	-1.809572	0.538218
H	-4.572184	-0.300394	1.355031
H	-4.917023	-0.305119	-0.406692
H	1.496234	0.927575	1.217131

#### B06

C	3.170489	-1.212992	-0.029935
C	3.764524	0.048404	0.009448
C	2.968661	1.192358	0.014735
C	1.580296	1.099651	-0.015531
C	0.995017	-0.170520	-0.047024
C	1.788007	-1.325693	-0.060736
H	3.781230	-2.107147	-0.037474
H	4.843638	0.138587	0.032023
H	3.428710	2.172572	0.039078
H	0.982484	1.996304	-0.019052

H	1.326808	-2.308184	-0.090619
N	-0.415296	-0.405921	-0.083740
H	-0.652415	-1.381226	-0.226794
C	-1.458998	0.382660	0.035627
C	-2.822053	-0.172650	-0.012715
C	-3.851627	0.648998	-0.266365
H	-3.745291	1.707160	-0.483836
O	-1.228156	1.668308	0.194493
H	-4.865454	0.270195	-0.301452
C	-2.995586	-1.657463	0.198605
H	-4.054849	-1.908955	0.222301
H	-2.554154	-2.240489	-0.618187
H	-2.551937	-1.988710	1.141723
H	-2.038843	2.148358	0.415881

### B07

C	0.229787	0.209839	-0.498307
C	-0.292863	-1.079313	-0.679524
C	-1.637032	-1.307904	-0.449918
C	-2.481525	-0.253296	-0.055264
C	-1.951207	1.037141	0.108085
H	-2.066642	-2.292198	-0.586326
H	0.344286	-1.888795	-1.018627
N	1.631241	0.457602	-0.755456
H	1.891979	0.975667	-1.588326
C	2.608864	0.054432	0.032365
C	3.996999	0.318465	-0.295699
H	4.178395	0.853987	-1.220870
C	5.003395	-0.072045	0.494295
H	4.830910	-0.610264	1.418777
H	6.030187	0.143042	0.224950
O	2.347178	-0.593185	1.136530
C	-0.597395	1.263671	-0.113814
H	-0.190055	2.260081	0.014660
H	-2.580841	1.863644	0.405970
O	-3.766497	-0.580402	0.134589
C	-4.711818	0.428502	0.518116
H	-5.665401	-0.086507	0.600107
H	-4.445902	0.863984	1.484951
H	-4.779919	1.207710	-0.245497
H	1.385647	-0.701827	1.267468

### B08

C	3.094152	0.090594	-0.156969
H	3.269819	-0.593201	0.671469
H	3.719122	0.973681	-0.044896
H	3.331371	-0.410218	-1.097003
N	1.687573	0.525135	-0.151063
H	1.490352	1.493470	-0.364859
C	0.657282	-0.258585	0.054415
C	-0.714306	0.256527	0.067800
O	0.964122	-1.525007	0.248441
C	-1.725812	-0.621330	-0.116195
H	-1.483611	-1.669756	-0.289592
C	-3.185850	-0.342874	-0.167060
H	-3.577689	-0.623381	-1.151563
H	-3.706884	-0.980112	0.555912
H	-3.447694	0.694622	0.027349
C	-0.880375	1.750039	0.222711
H	-1.921640	2.012863	0.390771
H	-0.308658	2.132455	1.072496
H	-0.564702	2.288288	-0.678641

H	0.203692	-2.040831	0.550267
---	----------	-----------	----------

### B12

C	-2.021792	-1.122346	-0.593118
C	-0.578866	-1.357014	-0.136226
C	-1.002660	1.043423	-0.475058
C	-2.312110	0.320449	-0.153678
H	-2.104365	-1.208932	-1.679944
H	-2.710182	-1.838248	-0.143917
H	-0.038229	-2.058279	-0.769838
H	-0.532024	-1.682519	0.903817
H	-0.870287	1.893533	0.197904
H	-0.967376	1.367741	-1.519102
H	-3.148960	0.769084	-0.695266
N	0.036249	0.005641	-0.235881
O	-2.476569	0.430028	1.252385
H	-3.357197	0.134842	1.509761
C	1.310313	0.267318	-0.084304
C	2.297065	-0.758734	0.214128
H	1.933359	-1.767996	0.345638
C	3.599739	-0.476624	0.336451
H	3.986879	0.527421	0.213330
H	4.310323	-1.260861	0.567089
O	1.776005	1.498031	-0.195033
H	1.098687	2.161421	-0.386294

### C01

C	-0.698065	1.441995	0.292558
C	-1.809439	0.640512	-0.393883
C	-1.810337	-0.797434	0.107224
H	-0.672733	2.466493	-0.079909
H	-1.671198	0.654876	-1.479128
H	-2.439547	-1.442149	-0.507416
H	-0.896077	1.500739	1.369447
H	-2.786591	1.080073	-0.186779
H	-2.157571	-0.864990	1.141976
C	0.652077	0.794599	0.068913
C	1.804112	1.464542	-0.088560
H	1.814597	2.548205	-0.072027
H	2.751305	0.961709	-0.234743
C	0.673122	-0.665013	0.025118
O	1.849442	-1.251618	-0.045647
N	-0.436476	-1.367283	0.057789
H	-0.371768	-2.380072	0.071054
H	1.820560	-2.216168	-0.130034

### C02\*

C	-0.122258	0.058767	0.000360
C	0.048556	1.473207	-0.000032
C	1.319552	2.037828	-0.000150
C	2.482325	1.261893	0.000065
C	2.310045	-0.120376	0.000031
H	1.415113	3.120674	-0.000256
H	-0.837460	2.089996	-0.000249
H	3.469866	1.704637	-0.000103
N	-1.297472	-0.650556	-0.000002
C	-2.500557	-0.045266	-0.000090
C	-3.617148	-1.058951	-0.000335
H	-3.304141	-2.099173	-0.001010
C	-4.908997	-0.723072	0.000221
H	-5.191381	0.325204	0.000785
H	-5.697509	-1.469897	0.000004

O	-2.789476	1.173122	-0.000018
C	1.068931	-0.722710	0.000166
H	0.968087	-1.800313	0.000237
Cl	3.770167	-1.157066	-0.000039

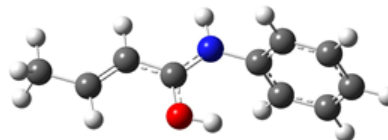
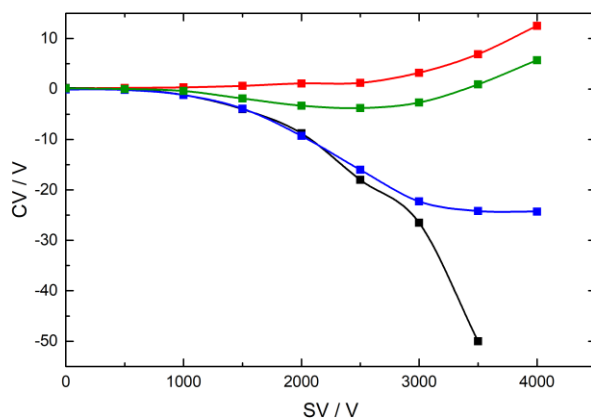
### C03

C	-0.591979	-1.536016	0.186758
C	-2.115414	-1.379584	0.262939
C	-2.406070	-0.167094	-0.634635
C	-1.249778	0.802519	-0.347670
H	-0.142906	-1.900664	1.110286
H	-0.298196	-2.189681	-0.638484
H	-2.426655	-1.188248	1.291513
H	-2.621351	-2.284758	-0.071493
H	-2.398532	-0.460253	-1.687749
H	-3.371468	0.292572	-0.423729
H	-0.987382	1.406674	-1.216832
C	-1.477578	1.709576	0.864936
H	-0.599481	2.318394	1.082249
H	-1.737443	1.135792	1.757167
H	-2.305989	2.385495	0.643357
N	-0.106663	-0.153029	-0.109984
C	1.146702	0.186356	-0.195443
C	2.239906	-0.772420	-0.031216
H	1.991910	-1.816636	-0.171056
C	3.484814	-0.413712	0.296623
H	3.781678	0.611261	0.499024
H	4.262754	-1.159460	0.404180
O	1.387193	1.465344	-0.445886
H	2.318528	1.620209	-0.655216

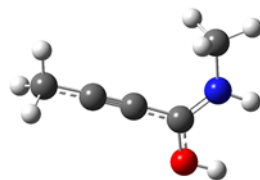
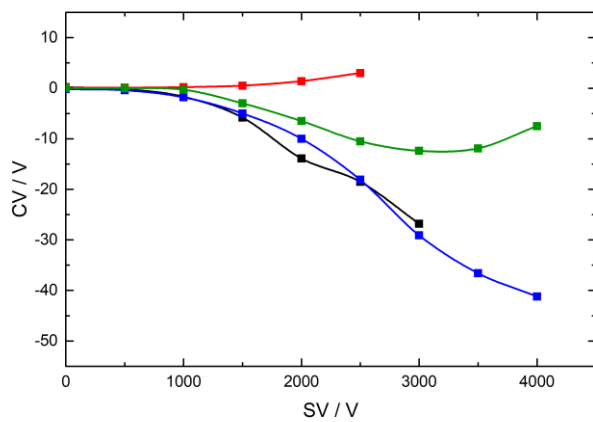
## Dispersion Plots

The dispersion plots obtained for protonated or deprotonated CRGs with a DMS cell (300 °C) containing **(red)** a pure N<sub>2</sub> environment, and a N<sub>2</sub> environment seeded with 1.5% (mole percent) **(green)** water vapor, **(blue)** methanol vapor, and **(black)** isopropyl alcohol vapor. The lowest energy isomer of each charged CRG is shown, as calculated at the B3LYP/6-311++G(d,p) level of theory.

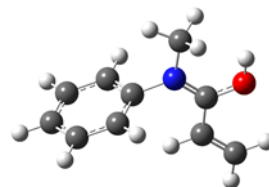
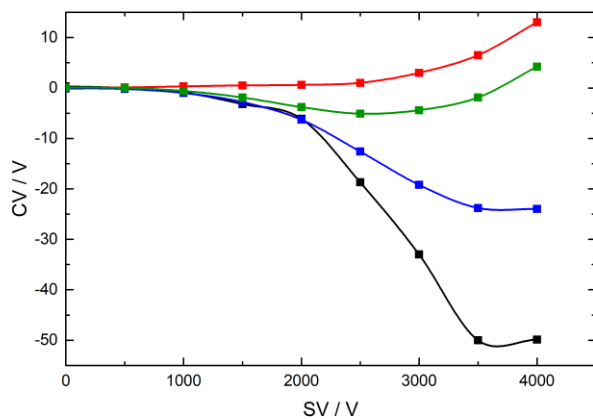
### (A01 + H)<sup>+</sup> (m/z = 162)



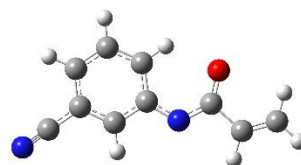
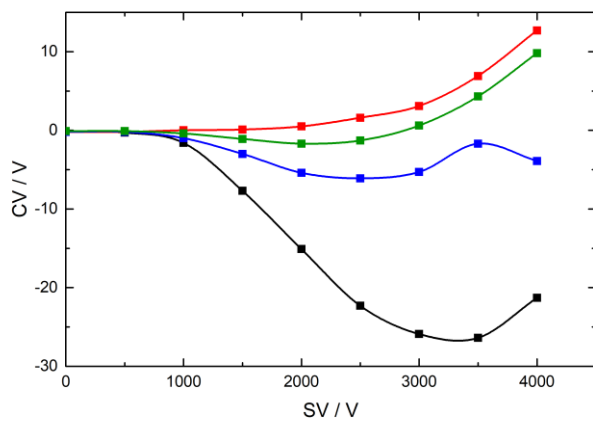
**(A02 + H)<sup>+</sup> (m/z = 98)**



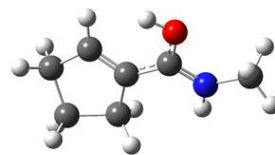
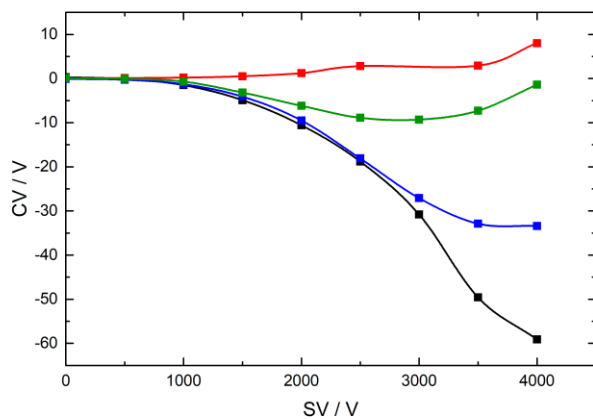
**(A03 + H)<sup>+</sup> (m/z = 162)**



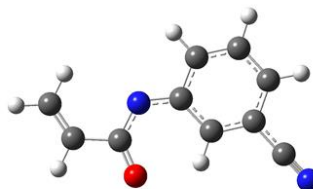
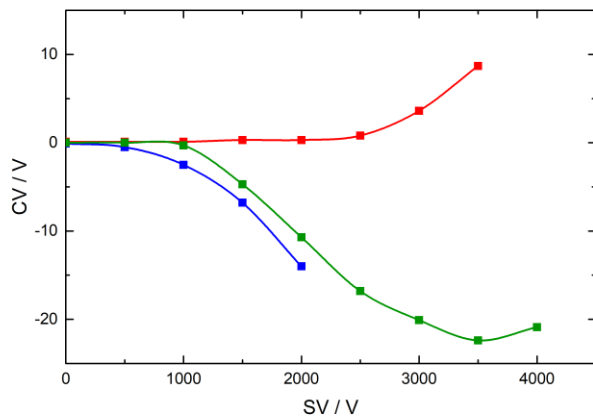
**(A04 - H)<sup>-</sup> (m/z = 171)**



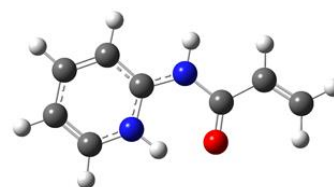
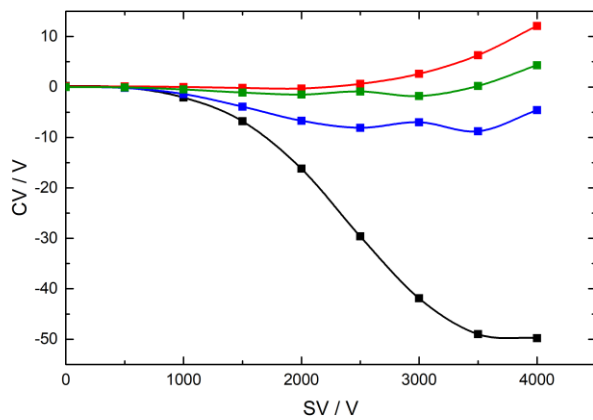
**(A05 + H)<sup>+</sup> (m/z = 126)**



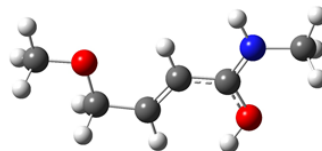
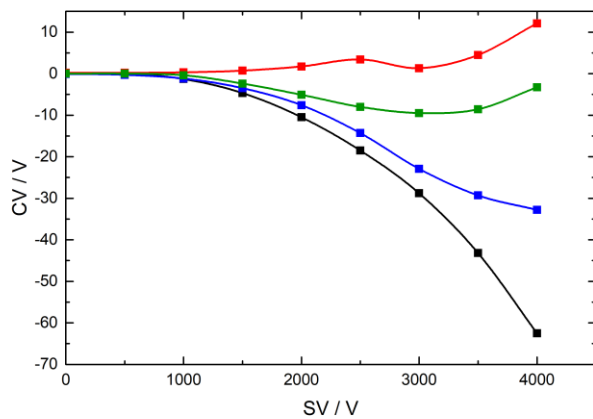
**(A06 + H)<sup>+</sup> (m/z = 86)**



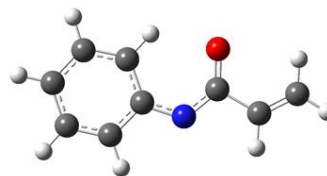
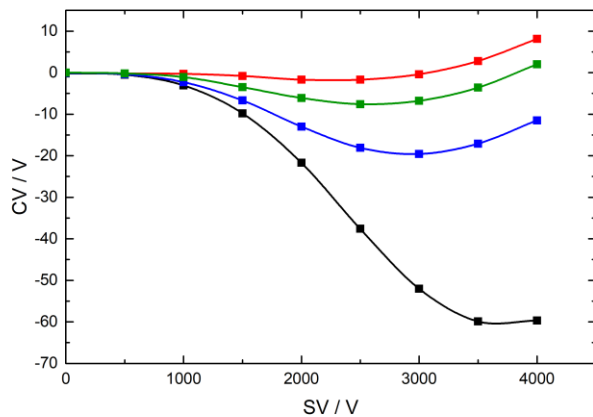
**(A07 + H)<sup>+</sup> (m/z = 149)**



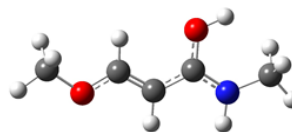
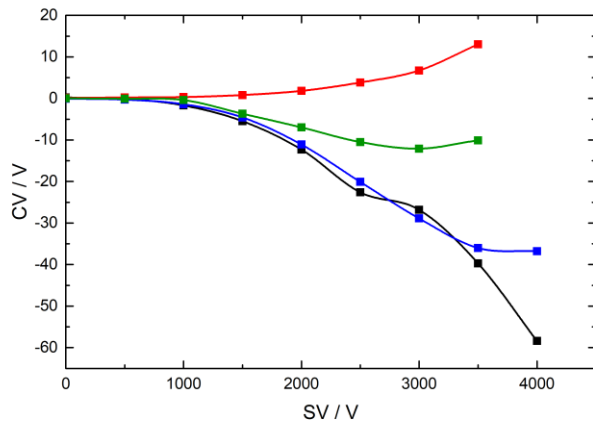
**(A08 + H)<sup>+</sup> (m/z = 130)**



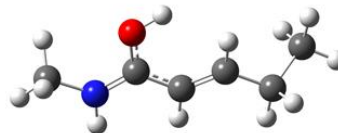
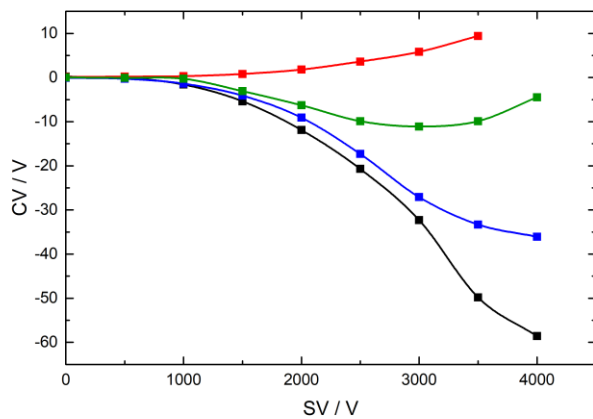
**(A09 - H)<sup>-</sup> (m/z = 146)**



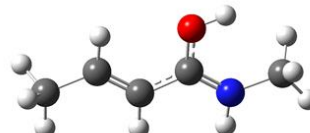
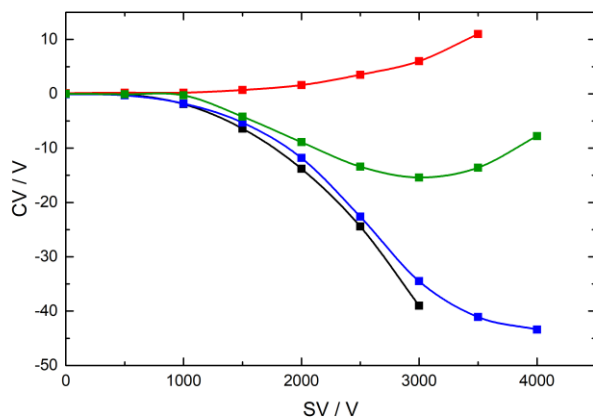
**(A10 + H)<sup>+</sup> (m/z = 116)**



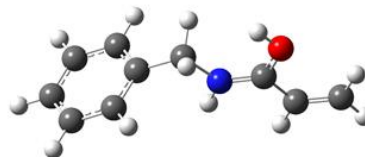
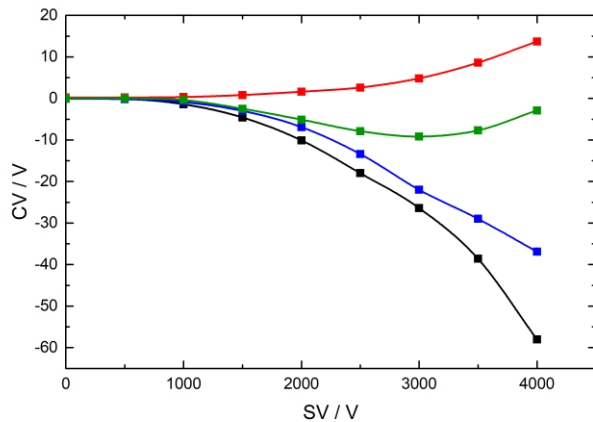
**(A11 + H)<sup>+</sup> (m/z = 114)**



**(A12 + H)<sup>+</sup> (m/z = 100)**

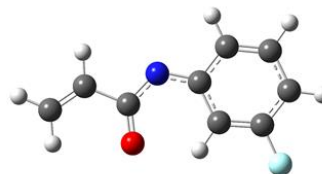
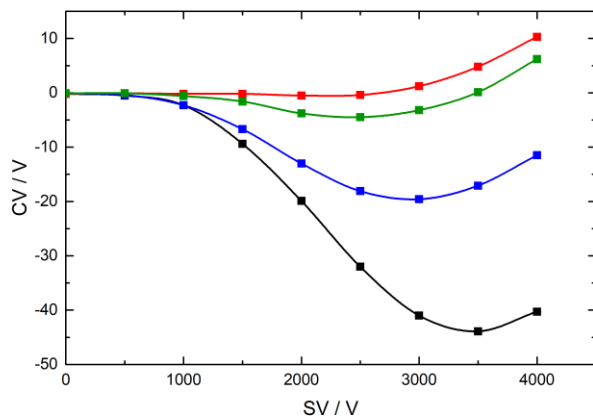


**(B01 + H)<sup>+</sup> (m/z = 162)**

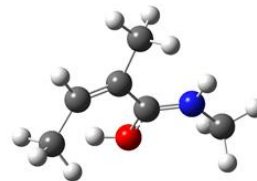
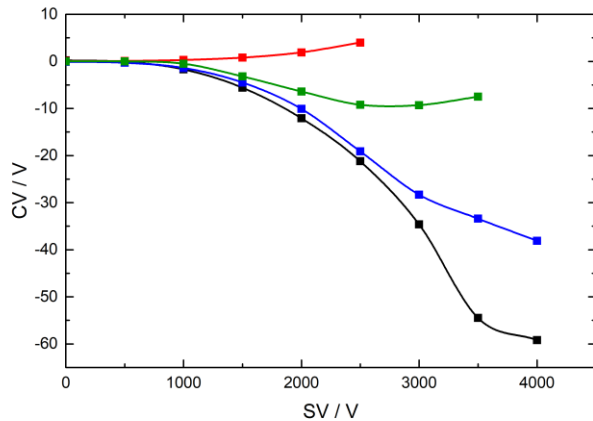




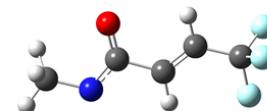
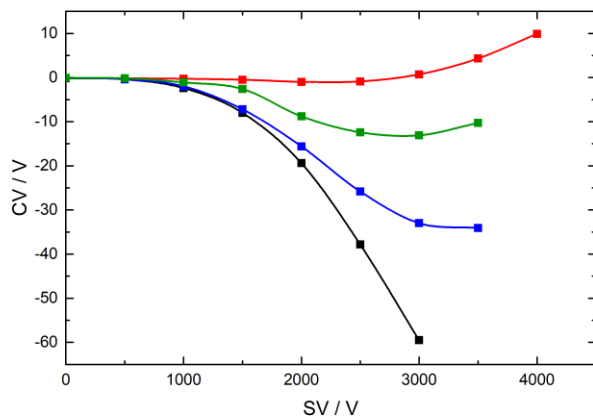
**(B02 - H)<sup>-</sup> (m/z = 164)**



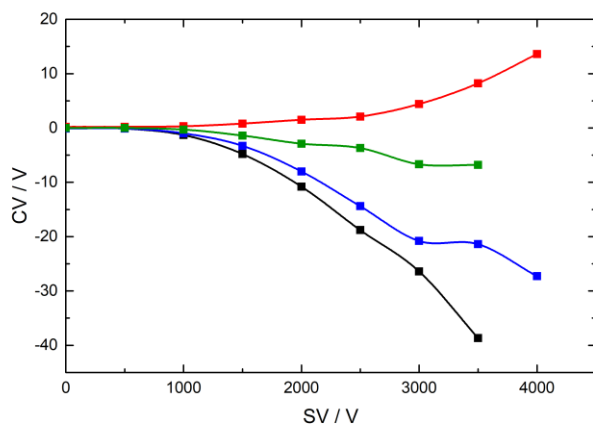
**(B03 + H)<sup>+</sup> (m/z = 114)**



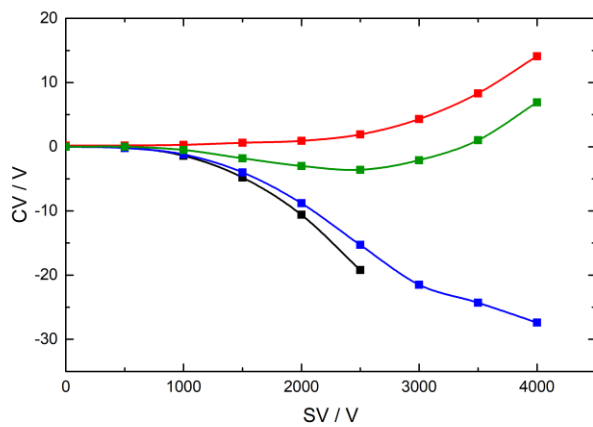
**(B04 - H)<sup>-</sup> (m/z = 152)**



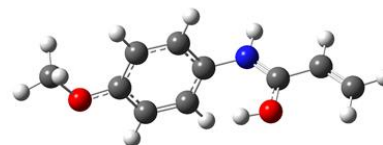
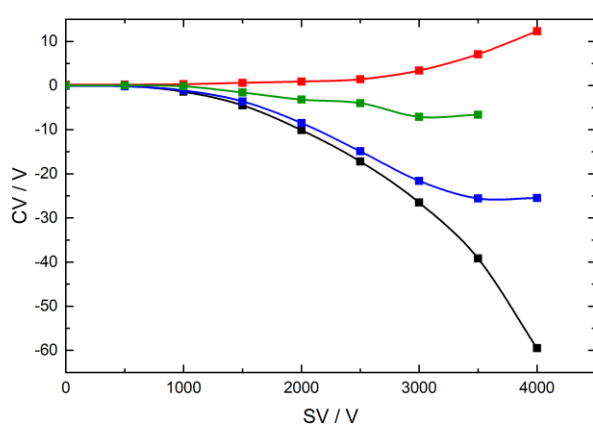
**(B05 + H)<sup>+</sup> (m/z = 178)**



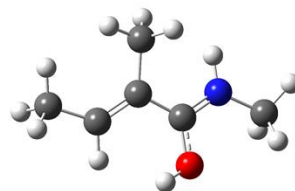
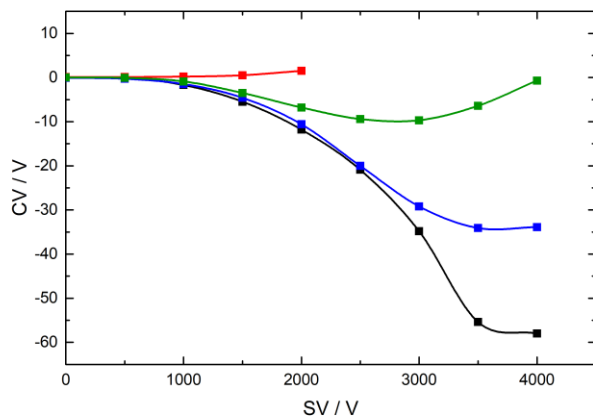
**(B06 + H)<sup>+</sup> (m/z = 162)**



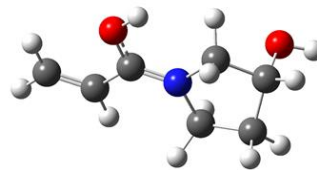
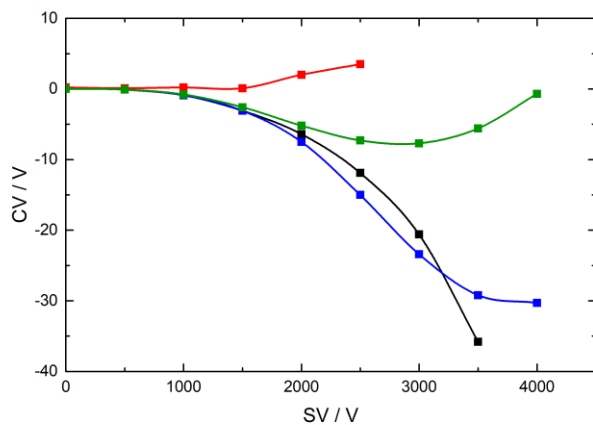
**(B07 + H)<sup>+</sup> (m/z = 178)**



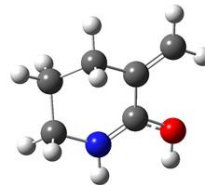
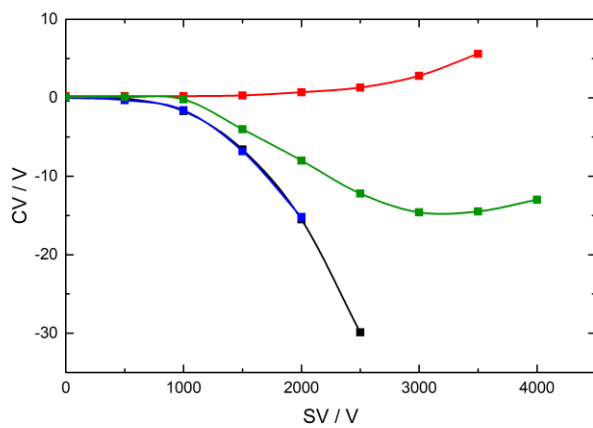
**(B08 + H)<sup>+</sup> (m/z = 114)**



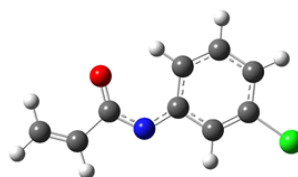
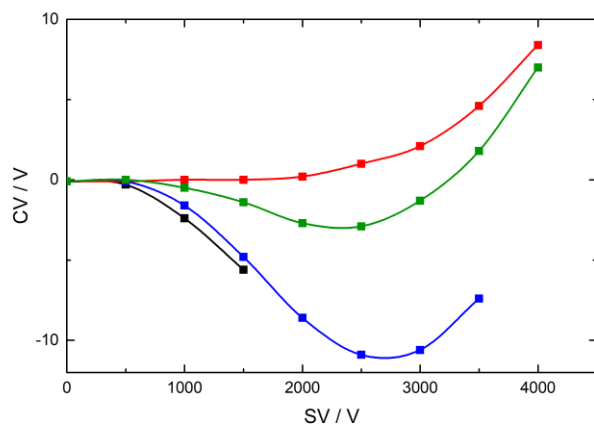
**(B12 + H)<sup>+</sup> (m/z = 142)**



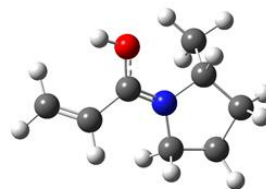
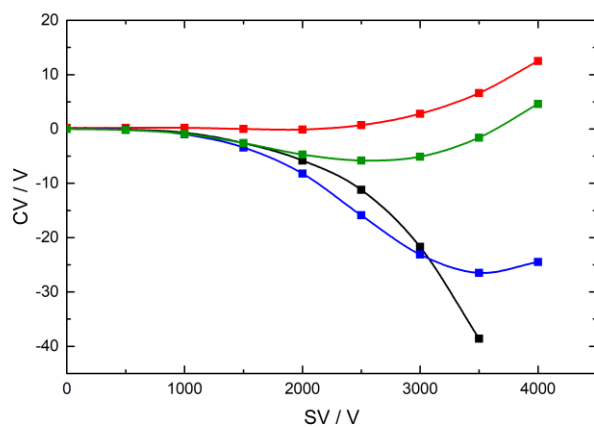
**(C01 + H)<sup>+</sup> (m/z = 112)**



**(C02 - H)<sup>-</sup> (m/z = 180)**



**(C03 + H)<sup>+</sup> (m/z = 140)**



## **Appendix III: Nucleobase Tautomers – Supporting Information**

The supporting information associated with this article can be found, in the online version, at [<https://doi.org/10.1016/j.ijms.2017.08.008>].