

**THE STRUCTURES OF PROTONATED AND ALKALI  
METAL CATIONIZED NUCLEOBASE SELF-ASSEMBLIES  
AND BASE PAIRS BY INFRARED MULTIPHOTON  
DISSOCIATION SPECTROSCOPY AND  
COMPUTATIONAL METHODS**

by © Ruodi Cheng

A Dissertation Submitted to the School of Graduate Studies in partial fulfillment of the  
requirements for the degree of

**Doctor of Philosophy Department of Chemistry**

Memorial University of Newfoundland

December 2019

## ABSTRACT

Nucleobases are the bricks of nucleic acids such as deoxyribonucleic acid (DNA)/ribonucleic acid (RNA) molecules. In the current research, ionic nucleobase self-assemblies and base pairs were studied using gas-phase ion techniques in either a Fourier transform ion cyclotron resonance (FTICR) or ‘paul-type’ quadrupole ion-trap (QIT) mass spectrometers. Sustained off-resonance irradiation collision-induced dissociation (SORI-CID) and infrared multiphoton dissociation (IRMPD) were used to fragment the target ions. IRMPD spectroscopy was employed to collect spectra by using tunable IR lasers, either an optical parametric oscillator or amplifier (OPO/A) or a free-electron laser (FEL). Density functional theory (DFT) was mainly used to study the structural information, calculate thermodynamic results, and perform IR frequency calculations for isomers. Besides, computed IR intensities were compared to experimental IRMPD spectra to explore their consistency.

The study of uracil with  $\text{Ca}^{2+}$  clusters was presented in Chapter 3. There were agreements between the global minima IR spectra and the experimental results. Uracil tetramer, pentamer and hexamer with  $\text{Ca}^{2+}$  are composed of both tautomerized and canonical uracils, which were not proposed by previous work. Further research on discovering the structures of 1-methylcytosine dimers with alkali metal cations has revealed two possible structures; a new one which is in planar geometry containing the interbase hydrogen bonding as well as being bound by the metal cation (Chapter 4). There was a conclusion that, as alkali metal cations’ radii increased, the ion-dipole interaction weakened. In Chapter 5 the research dug into the effects of alkali metal cations and proton on

guanine:cytosine (G:C) base pair. The heavier metal cations were found less likely to interrupt the hydrogen bonds between Watson-Crick G:C base pairs. Moreover, an unexpected great abundance of protonated cytosine for the dissociation of protonated G:C molecule, termed an *anomaly* by previous works, was explained because of the high proton transfer barrier but not the proposed thermochemistries. Guanine-involved base pair mismatches with protons were discussed in Chapters 6. For (9eG:1mT)H<sup>+</sup> and (9eG:9eG)H<sup>+</sup>, the lowest energy structures were sufficient to explain their IRMPD spectra while the global minimum of (9eG:9eG)H<sup>+</sup> presented only one *classical* hydrogen bond. For (9eG:9mA)H<sup>+</sup> the lowest energy structures' weighted-average spectrum was substantially consistent with its IRMPD spectrum.

## ACKNOWLEDGEMENTS

There are so many people supporting and guiding me throughout this fantastic and unreplicable journey. First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Travis D. Fridgen. Without his contributions of time, guidance, support, inspiration, criticism, and funding, I will not be able to finish all my research and accomplish this dissertation, especially during hard times. He creates the environment fulfilled with liberty and equality in our research group, which allowed me to demonstrate ideas freely without worrying about making mistakes. Also, Travis has provided me with plenty of precious opportunities to travel, for performing experiments at other facilities or presenting our work at various international conferences. During these trips, I have learnt from other scientists in the mass spectrometry field and interacted with lots of well-known chemists (and made friends) in diversified fields worldwide. Apart from the scientific knowledge he has imparted, Dr. Fridgen also influenced me with his excellent human qualities as being a patient, passionate, responsible and kind person. To my supervisor, Prof. Travis D. Fridgen: many thanks to your help and encouragement from the very beginning to the end.

I am grateful to Fridgen group members, previous and present, who always share academic ideas, help to polish my presentation slides, and correct my speaking and writing English mistakes (especially to Kelsey). I would also like to thank my committee member, Prof. Peter G. Pickup for all his time, advice and help to this dissertation. I am also grateful to CLIO and FELIX staff for their excellent support for several publications. I would like to extend my sincere thanks to all my examiners, Dr. Yuming Zhao, Dr. Karl Jobst and the

external examiner Dr. Sang Yun Han, for their critical and valuable feedback on my thesis. I would also like to acknowledge my co-authors, Dr. Estelle Loire, Dr. Jonathan Martens, Victoria E. Rose, and Dr. Barry Power, for their committed efforts on collecting fingerprint region spectra and manuscripts editing.

I am lucky to have such caring people in the general office of Chemistry Department, Mary, Debbie, Rosalind, Ebony, and Melissa for making this a great workplace.

I owe my special thanks to the closest friends, Yuchao, Shoupeng, Fuchen, and Tianyu, for their unconditional help, valuable contribution and encouragement when I was in depression, especially the first year I moved to Canada.

Finally, I would like to thank the people who mean a lot to me, my parents, for instilling me with a strong passion for learning and giving me the freedom to choose all the things I desired. I would never be able to pay back for their endless love, care, support and sacrifices, which shaped my life.

## TABLE OF CONTENTS

ABSTRACT.....	II
ACKNOWLEDGEMENTS .....	IV
TABLE OF CONTENTS.....	VI
LIST OF FIGURES .....	XI
LIST OF SCHEMES.....	XVI
LIST OF TABLES .....	XVII
LIST OF ABBREVIATIONS AND SYMBOLS .....	XVIII
LIST OF APPENDICES.....	XXI
CO-AUTHORSHIP STATEMENT.....	XXII
<b>Chapter 1 Introduction.....</b>	<b>1</b>
1.1 Overview.....	1
1.2 The Application of Cationic Complexes of Nucleobases.....	2
1.2.1 The Biological Role of Nucleobases and the Impacts of Metal Cations. ....	2
1.2.2 Materials Applications of Nucleobase/Metal Ion Complexes. ....	8
1.3 Gas-Phase Studies of Cationized Nucleobases.....	11
1.3.1 Protonated Nucleobase Complexes.....	12
1.3.2 Metal Cationized Nucleobase Complexes. ....	23
1.4 The Content of This Thesis.....	32
References.....	34
<b>Chapter 2 Experimental and Computational Methods.....</b>	<b>43</b>

2.1 General Introduction to Chapter 2 .....	43
2.2 Experimental Methods.....	43
2.2.1 Electrospray Ionization. ....	44
2.2.2 Overview of Mass Analyzers and Fragmentation Methods.....	46
2.2.3 Fourier-Transform Ion Cyclotron Resonance Mass Spectrometer (FT-ICR-MS) .....	46
2.2.3.1 Trapping of FT-ICR-MS.....	47
2.2.3.2 Excitation and Detection of FT-ICR-MS.....	50
2.2.4 Quadrupole Ion-Trap (QIT) Mass Spectrometer .....	56
2.2.4.1 Trapping of Ions in QIT.....	57
2.2.4.2 Excitation and Isolation of QIT. ....	61
2.2.4.3 Detection of Ions in QIT. ....	62
2.2.5 IRMPD Fragmentation.....	64
2.2.5.1 Description of IRMPD Process.....	64
2.3 Computational Methods .....	67
2.3.1 DFT .....	68
2.3.1.1 B3LYP. ....	70
2.3.1.2 Dispersion Correction. ....	71
2.3.1.3 M06-2X.....	72
2.3.2 Basis Sets .....	73
2.3.2.1 Minimal Basis Sets and Split-Valence Basis Sets. ....	74
2.3.2.2 Polarization and Diffuse Functions.....	75
2.3.3 Basis Set Superposition Error (BSSE).....	76
2.3.4 Single Point Energy Calculation .....	77

References.....	79
<b>Chapter 3 Self-Assembled Uracil Complexes Containing Tautomeric Uracil: An IRMPD Spectroscopic and Computation Study of the Structures of Gaseous Uracil<sub>n</sub>Ca<sup>2+</sup> (n=4, 5, or 6) Complexes .....</b>	<b>86</b>
3.1 Introduction.....	88
3.2 Methods. ....	91
3.2.1 Computational Methods.....	91
3.2.2 Experimental Methods. ....	92
3.3 Results and Discussion. ....	92
3.3.1 Infrared Multiple Photon Dissociation of U <sub>n</sub> Ca <sup>2+</sup> .....	92
3.3.2 IRMPD Spectra for U <sub>n</sub> Ca <sup>2+</sup> .....	93
3.3.3 U <sub>4</sub> Ca <sup>2+</sup> .....	96
3.3.4 U <sub>6</sub> Ca <sup>2+</sup> .....	99
3.3.5 U <sub>5</sub> Ca <sup>2+</sup> .....	102
3.3.6 Other Observed U <sub>5,6</sub> Ca <sup>2+</sup> Structures.....	103
3.3.7 Computed Energies .....	105
3.4 Conclusions .....	105
Conflicts of interest.....	108
References.....	109
<b>Chapter 4 Hydrogen Bonding in Alkali Metal Cation-Bound Dimers of 1-Methyl Cytosine: An IRMPD Spectroscopic and Computational Study .....</b>	<b>112</b>
4.1 Introduction.....	114
4.2 Methods .....	117
4.2.1 Computational Methods .....	117
4.2.2 Experimental Methods .....	118

4.3 Results and Discussion .....	119
4.3.1 Computed Structures for $(1\text{-mCyt})_2\text{M}^+$ .....	119
4.3.2 Vibrational (IRMPD) Spectra of $(1\text{-mCyt})_2\text{M}^+$ .....	122
4.3.3 Comparison of $(1\text{-mCyt})_2\text{M}^+$ and $(\text{Cyt})_2\text{M}^+$ .....	127
4.4 Conclusions .....	128
Acknowledgements.....	129
Conflicts of interest.....	130
References.....	131
<b>Chapter 5 A vibrational spectroscopic and computational study of gaseous protonated and alkali metal cationized G-C base pairs .....</b>	<b>136</b>
5.1 Introduction.....	138
5.2 Methods.....	143
5.2.1 Experimental Methods .....	143
5.3 Results and Discussion .....	145
5.3.1 $(9\text{eG:1mC})\text{H}^+$ .....	145
5.3.1.1 Structure of $(9\text{eG:1mC})\text{H}^+$ by IRMPD Spectroscopy.....	145
5.3.1.2. Dissociation of $(9\text{eG:1mC})\text{H}^+$ .....	150
5.3.2 $(9\text{eG:1mC})\text{M}^+$ ( $\text{M}=\text{Li}^+, \text{Na}^+, \text{K}^+, \text{Rb}^+, \text{Cs}^+$ ) .....	153
5.3.2.1 $(9\text{eG:1mC})\text{Li}^+$ .....	154
5.3.2.2 $(9\text{eG:1mC})\text{Na}^+$ .....	157
5.3.2.3 $(9\text{eG:1mC})\text{K}^+$ and $(9\text{eG:1mC})\text{Rb}^+$ .....	159
5.3.2.4 $(9\text{eG:1mC})\text{Cs}^+$ .....	163
5.3.3 The effects of the alkali metal cation on the computed structure of the G:C base pair. ....	164
5.4. Conclusion .....	164

Acknowledgments .....	166
References.....	167
<b>Chapter 6 An IRMPD Spectroscopic and Computational Study of Protonated Guanine-Containing Mismatched Base Pairs in the Gas Phase .....</b>	<b>173</b>
6.1 Introduction.....	175
6.2 Methods. ....	178
6.2.1 Experimental Methods. ....	178
6.2.2 Computational Methods.....	179
6.3 Results and Discussion .....	180
6.3.1 (9eG:1mT)H <sup>+</sup> . ....	180
6.3.2 (9eG:9mA)H <sup>+</sup> . ....	183
6.3.3 (9eG:9eG)H <sup>+</sup> . ....	187
6.3.4 Computed Energies Comparison. ....	189
6.4 Conclusions .....	190
Acknowledgements.....	192
References.....	193
<b>Chapter 7 Summary and Outlook .....</b>	<b>199</b>
References.....	206
<b>Appendices A .....</b>	<b>208</b>
<b>Appendices B .....</b>	<b>222</b>
<b>Appendices C .....</b>	<b>257</b>

## LIST OF FIGURES

<b>Figure 1.1</b> Schematics of five natural nucleobases and canonical Watson-Crick (WC) pairing, G:C and A:T. ....	3
<b>Figure 1.2</b> Side view of A-form (a), B-form (b), Z-form (c) DNA and the duplex structure (d) held by the Hoogsteen base pairs. From reference 26 with permission. ....	4
<b>Figure 1.3</b> The main biochemical activity of cisplatin. (a) cisplatin first entered the cell; one water molecule is incorporated into the complex to displace one chlorine and form the aqua complex(b). (b) attacks the N7 positions of the adjacent G:C base pairs (c) to form the Pt–DNA intra-strand cross-link adduct (d). Reproduced from reference 43 with permission. ....	8
<b>Figure 1.4</b> (A)The self-assembled supramolecular polymer via hydrogen bonds between nucleobases. (B) The chains linked by a hydrogen bond between nucleobases to form the supramolecular polymer. (C) The supramolecular polymers cross-linked by the side-chain nucleobases' hydrogen bonds. Reprinted from reference 12 with permission .....	10
<b>Figure 1.5</b> The structure of the borate ester stabilized guanosine quartets monomer with $K^+$ located at the center. Reprinted from reference 69 with permission. ....	11
<b>Figure 1.6</b> The spectra of $UH^+$ on the left and the spectra of $TH^+$ on the right. In both cases, isomer a) is the second lowest energy isomer and isomer b) is the predominant structure. Reprinted from reference 85 with permission .....	13
<b>Figure 1.7</b> The spectra comparison of $CH^+$ . a) The second lowest energy isomer, C1_hb. b) The concluded predominate and most stable isomer, C1_hc. c) The IRMPD spectrum of the $CH^+$ molecule. The figure is reprinted from reference 85 with permission. ....	15
<b>Figure 1.8</b> A schematic of the protonated dimeric cytosine (On the top) (a) The optimized global minima of the cytosine dimer with a proton bound to N3 (b) The transition-state of the dimeric cytosine structures. Reprinted from reference 86 with permission. ....	17
<b>Figure 1.9</b> Experimental IRMPD spectrum of the protonated monomer of 1-methylcytosine, $m/z$ 126, at the top panel: black trace from the dissociation of the proton-bound dimer by $CO_2$ laser photolysis; red trace sourced from electrospray ionization. The spectra comparison with normal modes calculated at B3LYP/6-31G** (for which	

frequencies $>850\text{cm}^{-1}$ have been scaled by 0.97, with $20\text{ cm}^{-1}$ Gaussian broadenings) for four possible tautomers. Reprinted from reference 87 with permission. ....	18
<b>Figure 1.10</b> The predicted structure of the (a) $(1\text{mC}:1\text{mG})\text{H}^+$ and b) $(1\text{mC}:9\text{mG})\text{H}^+$ at the B3LYP/6-311+G(2d,2p) level with the Gaussian 09 suite of programs. Reprinted from reference 97 with permission. ....	19
<b>Figure 1.11</b> Computed dissociation diagrams of (a) $(1\text{mC}:9\text{eG})\text{H}^+$ and (b) $(1\text{mC}:9\text{eG})\text{H}^+$ from their lowest energy structures. The predicted energies, $D_0/\Delta H/\Delta G$ (dissociation energies, relative enthalpies and relative Gibbs energies), were obtained by using the CBS-QB3 theory in kJ/mol at 298.15 K. $\Delta PA = \Delta PA(1\text{mC}) - \Delta PA(x\text{mG})$ . Reprinted from reference 97 with permission. ....	21
<b>Figure 1.12.</b> The Hoo N(3)CH <sup>+</sup> is the global minima, followed by Hoo N(7)GH <sup>+</sup> . The WC N(7)GH <sup>+</sup> is the third lowest 298K Gibbs energy structure concluded for GCH <sup>+</sup> produced from the pH=5.8 solution. The computed structures were optimized at the M06-2X/6-311G++(d,p) level. Reprinted from reference 89 with permission. ....	22
<b>Figure 1.13.</b> Experimental IRMPD spectra of the protonated dimeric molecule, GCH <sup>+</sup> , in top panels: black trace and pink trace from the dissociated proton-bound G:C dimer by FEL laser for solution A, pH=3.2, and solution B, pH=5.8. Bottom panels: the red, green and blue curves stand for the computed IR intensities for Hoo N(3)CH <sup>+</sup> , Hoo N(7)CH <sup>+</sup> and WC N(7)GH <sup>+</sup> , respectively, at the M06-2X/6-311G++(d,p) level with the Gaussian 09 suite of programs. Reprinted from reference 89 with permission. ....	22
<b>Figure 1.14</b> The computed structure of alkali metal cationized G-tetrad, top view (left) and side view (right). $>\text{OOM}$ is the angle of two opposite Os and alkali metal cation. Reprinted from reference 108 with permission. ....	24
<b>Figure 1.15</b> BIRD breakdown plots for $(9\text{eG})_8\text{Na}^+$ (top) and $(9\text{eG})_8\text{K}^+$ (bottom) at 341 K and 399 K, respectively. Reprinted from reference 109 with permission. ....	26
<b>Figure 1.16.</b> Three lowest energy conformers of divalent complex, $(\text{H}_2 - \text{AGAG})^{2+}$ . The stability of three isomers is in the sequence of $(\text{H} \cdot \text{I})_2^{2+} \approx (\text{H} \cdot \text{II})_2^{2+} > (\text{III}')_2^{2+}$ . Reprinted from reference 93 with permission. ....	26
<b>Figure 1.17.</b> (a) Two preferred structures, $(\text{M} \cdot \text{I})_2^{2+}$ and $(\text{M} \cdot \text{II})_2^{2+}$ , where $\text{M} = \text{Na}, \text{K}$ , in the gas phase and water phase. (b) Two preferred structures, $(\text{Li} \cdot \text{I})_2^{2+}$ and $(\text{Li} \cdot \text{II})_2^{2+}$ , of	

(Li<sub>2</sub> – AGAG)<sup>2+</sup> in the gas phase and water phase. Both the top and side views of the structures are shown. Gas-phase and COSMO stand for the structures optimized in the gas phase and water phase, respectively. Reprinted from reference 93 with permission. ....31

**Figure 1.18.** Schematic of the UracilCa<sup>2+</sup> (left) and the (Uracil-H)(Uracil)Ca<sup>2+</sup> (right)...31

**Figure 3.1** Comparison of the experimental IRMPD spectra for U<sub>n</sub>Ca<sup>2+</sup> (n=4,5,6) in the fingerprint and the CH/NH/OH stretching regions. ....94

**Figure 3.2** Comparison of the experimental IRMPD spectrum (grey trace) for U<sub>4</sub>Ca<sup>2+</sup> and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for four different isomers. The 298 K enthalpies and Gibbs energies relative to structure C<sub>2</sub>T<sub>2</sub>a are also shown (also computed using B3LYP/6-31+G(d,p)), kJ mol<sup>-1</sup>. ....98

**Figure 3.3** Comparison of the experimental IRMPD spectrum (grey trace) for U<sub>6</sub>Ca<sup>2+</sup> and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure C<sub>4</sub>T<sub>2</sub>a are also shown (also computed using B3LYP/6-31+G(d,p)), kJ mol<sup>-1</sup>. ....100

**Figure 3.4** Comparison of the experimental IRMPD spectrum (grey trace) for U<sub>5</sub>Ca<sup>2+</sup> and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure C<sub>3</sub>T<sub>2</sub>a are also shown (also computed using B3LYP/6-31+G(d,p)), kJ mol<sup>-1</sup>. ....104

**Figure 4.1** B3LYP/6-31+G(d,p) computed structures for the SAP and AAP structures for (1-mCyt)<sub>2</sub>M<sup>+</sup> complexes. The 298 K enthalpies and Gibbs energies (parentheses) of the AAP structures relative to the SAP structures are also provided at the B3LYPD3/6-31+G(d,p) level (top), B3LYPD3/6-311+G(3df, 3pd)//B3LYPD3/6-31+G(d,p) (middle), and M06-2X/6-31+G(d,p) level (bottom), kJ mol<sup>-1</sup>. ....121

.....123

**Figure 4.2** Experimental IRMPD spectra of (1-mCyt)<sub>2</sub>M<sup>+</sup> complexes, where M= Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Rb<sup>+</sup>, Cs<sup>+</sup> (grey traces) and computed IR spectra for the SAP complexes (black traces). ....123

**Figure 4.3** Experimental IRMPD spectra of 1-mCyt)<sub>2</sub>M<sup>+</sup> complexes, where M= Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Rb<sup>+</sup>, Cs<sup>+</sup> (grey traces) and computed IR spectra for the AAP complexes (black traces). ....125

.....125

<b>Figure 5.1</b> Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint and the CH/NH/OH stretching regions for (9eG:1mC)H <sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers. The 298 K enthalpies and Gibbs energies relative to structure Hoo N7(9eG)H <sup>+</sup> are also shown (also computed using B3LYP/6-31+G(d,p)). All energies are shown in kJ mol <sup>-1</sup> .....	147
<b>Figure 5.2</b> Infrared multiple photon dissociation mass spectra of all six complexes at the specified wavenumber. The <i>m/z</i> and name of the precursor ions and fragments are also shown. The vertical axis is % relative abundance of the largest feature. ....	153
<b>Figure 5.3</b> A contour potential energy surface for (9eG:1mC)H <sup>+</sup> . The vertical axis is the N3—N7 distance, and the horizontal is N-H <sup>+</sup> distance (N3 of 1mC). The contours are the electronic energies relative to the lowest energy structure in kJ mol <sup>-1</sup> . ....	154
<b>Figure 5.4</b> Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)Li <sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for six isomers. The weighted average spectra of all six structures are shown in both regions. ....	156
<b>Figure 5.5</b> Structures of (9eG:1mC)Li <sup>+</sup> and (9eG:1mC)Na <sup>+</sup> isomers whose IR spectra are compared to the experimental spectra in <b>Figures 5.4</b> and <b>5.6</b> . The structures and relative 298 K enthalpies and Gibbs energies were computed using B3LYPD3/6-31+G(d,p) and are in kJ mol <sup>-1</sup> . For the Li <sup>+</sup> structures, some side views are shown for non-planar structures. ....	157
<b>Figure 5.6</b> Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)Na <sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers. ....	159
<b>Figure 5.7</b> Structures of (9eG:1mC)K <sup>+</sup> , (9eG:1mC)Rb <sup>+</sup> , and (9eG:1mC)Cs <sup>+</sup> isomers whose IR spectra are compared to the experimental spectra in <b>Figures 5.8</b> , <b>5.9</b> and <b>5.10</b> . The structures and relative 298 K enthalpies and Gibbs energies were computed using B3LYPD3/6-31+G(d,p) and are in kJ mol <sup>-1</sup> . For the Cs <sup>+</sup> structures, some side views are shown for non-planar structures.....	160
.....	161

<b>Figure 5.8</b> Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)K <sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers. ....	161
<b>Figure 5.9</b> Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)Rb <sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers. ....	162
.....	163
<b>Figure 5.10</b> Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)Cs <sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers. ....	163
<b>Figure 6.1</b> Comparison of the experimental IRMPD spectrum (grey trace) for (G:T)H <sup>+</sup> and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure GN7_GN7-TO4_TN3-GO6 are also shown (also computed using B3LYP/6-31+G(d,p))......	182
<b>Figure 6.2</b> Comparison of the experimental IRMPD spectrum (grey trace) for (G:A)H <sup>+</sup> and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for four different isomers. The 298 K enthalpies and Gibbs energies relative to structure AN1_AN1-GO6_AN6-GN7 are also shown (also computed using B3LYP/6-31+G(d,p))......	184
<b>Figure 6.3</b> Infrared multiple photon dissociation mass spectra of three complexes at their maxima in dissociation in the fingerprint region. The <i>m/z</i> and name of the precursor ions and fragments are also shown. ....	187
<b>Figure 6.4</b> Comparison of the experimental IRMPD spectrum (grey trace) for (G:G)H <sup>+</sup> and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure GN7_GN7-GN7_GC8-GO6 are also shown (also computed using B3LYP/6-31+G(d,p))......	189

## LIST OF SCHEMES

<b>Scheme 3.1.</b> (A) The lowest energy $U_4Ca^{2+}$ composed of all diketo uracils determined in previous works. <sup>17,19,20</sup> (B) Dimeric $Ca^{2+}$ /uracil structure composed of one diketo and one N3O4 tautomer comprising another $U_4Ca^{2+}$ isomer. ....	95
<b>Scheme 3.2.</b> Four tautomers of uracil including the numbering convention for pyrimidine bases such as uracil. ....	96
<b>Scheme 3.3</b> Lowest energy trimeric $Ca^{2+}$ /uracil structure composed of two canonical uracils and one N3O4 tautomer comprising the lowest energy $U_6Ca^{2+}$ isomer.....	101
<b>Scheme 3.4</b> Possible structure of one half of $U_{14}Ca^{2+}$ with the 6 mer core composed of two trimeric structures. ....	108
<b>Scheme 4.1</b> Hydrogen bonding in a) the G-quadruplex and b) the i-motif. c) The reported symmetric anti-parallel (SAP) structure of the alkali metal cation bound dimer of cytosine, $(Cyt)_2M^+$ , $M=Li, Na,$ and $K$ . The conventional ring numbering for cytosine is shown in b) .....	115
<b>Scheme 5.1</b> Watson-Crick G:C and A:T base pairs .....	139
<b>Scheme 5.2</b> Watson-Crick and Hoogsteen structures for protonated G:C base pairs. ....	142
<b>Scheme 5.3</b> The structures of protonated dimers of methanol and acetonitrile, and of 9-ethylguanine and 1-methylcytosine. In both cases the lower proton affinity base is the site of protonation.....	149
<b>Scheme 6.1</b> Watson-Crick G:C and A:U(T) base pairs.....	176

## LIST OF TABLES

<b>Table 3.1</b> Comparison of basis set on the 298 K energetics (top relative enthalpies, bottom relative Gibbs energies) in $\text{kJ mol}^{-1}$ of some $\text{U}_4\text{Ca}^{2+}$ structures. ....	106
<b>Table.4.1</b> Computed 298 K energetics of dissociation (B3LYP/6-311+G(3df,3pd)//B3LYP/6-31+G(d,p)) for the $(1\text{-mCyt})_2\text{M}^+$ complexes. Counterpoise corrections for BSSE are included. ....	126
<b>Table 4.2.</b> The 298 K relative enthalpies and Gibbs energies, in $\text{kJ mol}^{-1}$ , of the cytosine dimeric complexes by three different computational methods. ....	128
<b>Table 6.1.</b> 298 K proton affinities (PA) and gas basicities (GB) for 9-ethylguanine, 9-methyladenine, and 1-methylthymine. Protonation site is indicated in parentheses. a: CBS-QB3 b: B3LYP/6-31+G(d,p) .....	186

## LIST OF ABBREVIATIONS AND SYMBOLS

$\Delta\nu$	Excitation Frequency
(9eG) <sub>4</sub>	9eG Tetrads
(1mC)H <sup>+</sup>	Protonated 1-Methylcytosine
(1mC:1mG)H <sup>+</sup>	Protonated 1-Methylcytosine:1-Methylguanine
(1mC:9eG)H <sup>+</sup>	Protonated 1-Methylcytosine:9-Ethylguanine
(9eG)H <sup>+</sup>	Protonated 9-Ethylguanine
1mC	1-Methylcytosine
5'-AMP	5'-Adenosine Monophosphate
9eG	9-Ethylguanine
A	Adenine
ATP	Adenosine Triphosphate
B3LYP	Becke, 3-Parameter, Lee-Yang-Parr Hybrid Functional
BIRD	Blackbody Infrared Radiative Dissociation
BO	Born-Oppenheimer
BSSE	Basis Set Superposition Errors
C	Cytosine
C <sub>2</sub> H <sup>+</sup>	I-motif
CBS	Complete Basis Set
CH <sup>+</sup>	Protonated Cytosine
CID	Collision-Induced Dissociation

<b>CLIO</b>	Centre Laser Infrarouge D'orsay
<b>DFT</b>	Density Functional Theory
<b>DNA</b>	Deoxyribonucleic Acid
<b>ECP</b>	Effective Core Potentials
<b>EM</b>	Electron Multiplier
<b>ESI</b>	Electrospray Ionization
$f_b$	Beat Frequency
$f_c$	Cyclotron Frequency
<b>FC</b>	Faraday Cup
<b>FEL</b>	Free-Electron Laser
<b>FELIX</b>	Free-Electron Lasers at Infrared Experiments
$f_{exc}$	Excitation Frequency
<b>FID</b>	Free Induction Decay
<b>FTICR</b>	Fourier Transform Ion Cyclotron Resonance
<b>G</b>	Guanine
<b>GCH<sup>+</sup></b>	Protonated G:C Base Pair
<b>GMP</b>	Guanosine Monophosphate
<b>GTOs</b>	Gaussian-Type Orbitals
<b>HF</b>	Hartree-Fork
<b>H-K</b>	Hohenberg-Kohn Theorem
<b>IRMPD</b>	Infrared Multiphoton Dissociation

<b>IVR</b>	Vibrational Energy Redistribution
<b>K-S</b>	Kohn-Sham
<b>LCAO</b>	Linear Combination of Atomic Orbitals
<i>m/z</i>	Mass to Charge Ratio
<b>MO</b>	Molecular Orbital
<b>MP</b>	Møller–Plesset
<b>PA</b>	Proton Affinities
<b>QIT</b>	Quadrupole Ion Trap
<b>RF</b>	Radio Frequency
<b>RNA</b>	Ribonucleic Acid
<b>SCF</b>	Self-Consistent-Field
<b>SORI-CID</b>	Sustained Off-Resonance Irradiation Collision-Induced Dissociation
<b>STOs</b>	Slater-Type Orbitals
<b>T</b>	Thymine
<b>TEBPs</b>	Telomere End-Binding Proteins
<b>TH<sup>+</sup></b>	Protonated Thymine
<b>TOF-MS</b>	Time of Flight Mass Spectrometer
<b>U</b>	Uracil
<b>UH<sup>+</sup></b>	Protonated Uracil

**LIST OF APPENDICES**

Appendices A .....208  
Appendices B .....222  
Appendices C .....257

## CO-AUTHORSHIP STATEMENT

### **Chapter 3 Self-Assembled Uracil Complexes Containing Tautomeric Uracil: An IRMPD Spectroscopic and Computation Study of the Structures of Gaseous Uracil<sub>n</sub>Ca<sup>2+</sup> (n=4, 5, or 6) Complexes**

**Article:** R. Cheng, V. E. Rose, B. Power and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2018, **20**, 572–580.

The *principal author* (*Ruodi Cheng*) contributed to all aspects of the project as the primary researcher, including literature review, data collection and analysis, presenting and discussing data with corresponding authors, writing the manuscript, and preparing answers to the questions and comments of peer reviewers.

The *corresponding author* (*Victoria E. Rose*) contributed to collect the experimental data and facilitated the final version of the manuscript.

The *corresponding author* (*Barry Power*) contributed to some computational aspects and facilitated the final version of the manuscript.

The *corresponding author* (*Travis D. Fridgen*) proposed the initial research and contributed to several aspects of this chapter, including some computational aspects, the revision of several manuscript drafts, final manuscript submission to the journal, and supervision of the principal author.

### **Chapter 4 Hydrogen Bonding in Alkali Metal Cation-Bound Dimers of 1-Methyl Cytosine: An IRMPD Spectroscopic and Computational Study**

**Article:** R. Cheng, E. Loire and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2019, **21**, 11103–11110.

The *principal author* (*Ruodi Cheng*) contributed to all aspects of the project as the primary researcher, including literature review, data collection and analysis, presenting and discussing data with corresponding authors, writing the manuscript, and preparing answers to the questions and comments of peer reviewers.

The *corresponding author* (*Estelle Loire*) collaborated in the collection of the experimental data at CLIO and facilitated the final version of the manuscript.

The *corresponding author* (*Travis D. Fridgen*) proposed the initial research and contributed to several aspects of this chapter, including the revision of several manuscript drafts, final manuscript submission to the journal, and supervision of the principal author.

## **Chapter 5 A Vibrational Spectroscopic and Computational Study of Gaseous Protonated and Alkali Metal Cationized G-C Base Pairs**

**Submitted Article:** R. Cheng, J. Martens, and T. D. Fridgen, *Phys. Chem. Chem. Phys.*  
(Manuscript ID: CP-ART-11-2019-005962)

The *principal author* (*Ruodi Cheng*) contributed to all aspects of the project as the primary researcher, including proposing the initial research, literature review, data collection and analysis, presenting and discussing data with corresponding authors, writing the manuscript, and preparing answers to the questions and comments of peer reviewers.

The *corresponding author* (*Jonathan Martens*) collaborated in the collection of the experimental data at FELIX and facilitated the final version of the manuscript.

The *corresponding author* (*Travis D. Fridgen*) contributed to several aspects of this chapter, including some computational aspects, the revision of several manuscript drafts, final manuscript submission to the journal, and supervision of the principal author.

## Chapter 6 An IRMPD Spectroscopic and Computational Study of Protonated Guanine-Containing Mismatched Base Pairs in the Gas Phase

**Submitted Article:** R. Cheng, E. Loire, J. Martens and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2020, **22**, 2999–3007.

The *principal author* (*Ruodi Cheng*) contributed to all aspects of the project as the primary researcher, including proposing the initial research, literature review, data collection and analysis, presenting and discussing data with corresponding authors, writing the manuscript, and preparing answers to the questions and comments of peer reviewers.

The *corresponding author* (*Estelle Loire*) collaborated in the collection of the experimental data at CLIO and facilitated the final version of the manuscript.

The *corresponding author* (*Jonathan Martens*) collaborated in the collection of the experimental data at FELIX and facilitated the final version of the manuscript.

The *corresponding author* (*Travis D. Fridgen*) contributed to several aspects of this chapter, including some computational aspects, the revision of several manuscript drafts, final manuscript submission to the journal, and supervision of the principal author.

# Chapter 1 Introduction

## 1.1 Overview

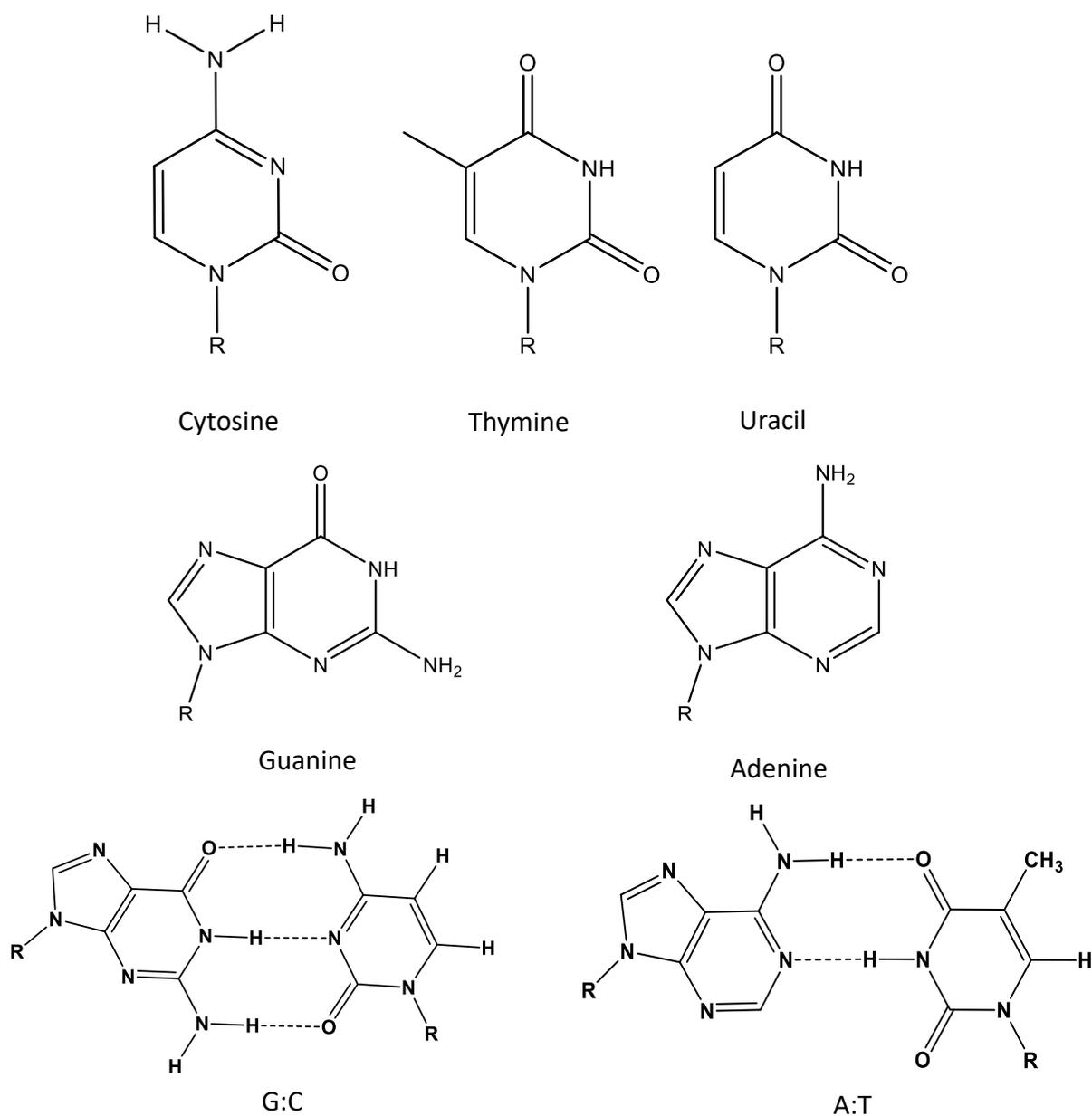
*In vivo*, the complementary pairing interactions between nucleobases provide the possibility to form DNA/RNA duplexes, triplexes,<sup>1</sup> quadruplexes,<sup>2</sup> and hairpins.<sup>3</sup> Moreover, base pairs can self-assemble into homo-adducts and different types of hetero-adducts.<sup>4-6</sup> An example of one of these self-assembled structures within nucleic acids is called G-quadruplex. The G-quadruplex is a three-layered sandwiched structure of guanine tetrads which are better stabilized with the presence of alkali metal cations, in one, two or four strands of nucleic acids containing guanine-rich regions.<sup>7-11</sup> In addition to the well-known roles of nucleobases in biology, they are also widely applied in synthesizing biomaterials,<sup>12-15</sup> which stems from their ability to form large clusters by characteristic non-covalent interactions. To reveal the intrinsic binding properties of biomolecules interacting with metal cations and protons, gas-phase studies employing mass spectrometric methods have become essential. The rapid development of computational methods provides further insights into the structural information and thermodynamic properties of various isomers by comparing with gaseous experimental results.

This chapter is a brief introduction to previous studies of nucleobase self-assemblies with either metal cations or protons. These studies have significant impacts on understanding the biological roles of nucleobase adducts and designing bio-polymers utilizing the clustering abilities of nucleobases.

## 1.2 The Application of Cationic Complexes of Nucleobases

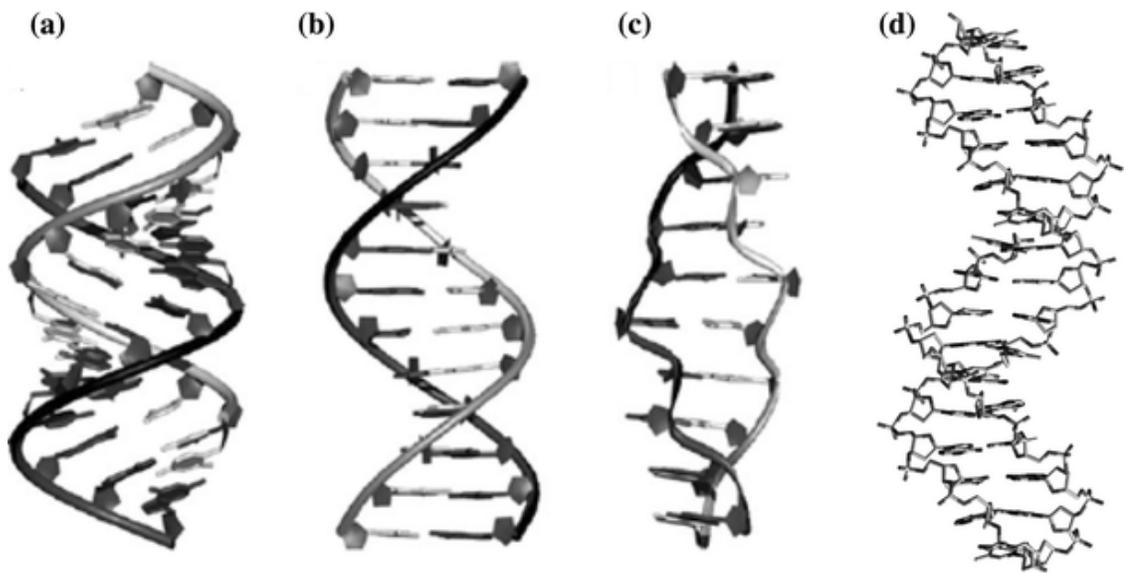
**1.2.1 The Biological Role of Nucleobases and the Impacts of Metal Cations.** Unveiling the mysteries of deoxyribonucleic acid (DNA) double-helix structure, also called the B-form helix, is widely believed to be one of the most important scientific discoveries in the 20<sup>th</sup> century. The structure of DNA was first published in 1953 by James Watson and Francis Crick in *Nature*.<sup>16,17</sup> The complementary base pairs, coined Watson-Crick base pairs, are guanine (G) with cytosine (C) and adenine (A) with thymine (T). Nucleobases bound to deoxyribose are called nucleosides, and the addition of phosphate group to a nucleoside turns it to a nucleotide which is the monomeric component of nucleic acids.

Another polymeric molecule that is almost identical to DNA is ribonucleic acid (RNA). However, there are several seemingly small, but significant differences between DNA and RNA. First, instead of pairing with T, A pairs with a similar base, uracil (U). Secondly, the sugar of RNA nucleotide is ribose instead of deoxyribose. Thus, the presence of the 2' hydroxy group in RNA sugar increases the stability of the A-form helix (**Figure 1.2**) and does not allow it to form the B-form helix that dominates the structure of DNA. Another difference is that RNA is single-stranded, which can form complicated secondary or tertiary structures by self-assembly. The functions of RNA are coding,<sup>18</sup> decoding,<sup>19</sup> regulation,<sup>20</sup> and expression<sup>21</sup> of genes, while DNA plays the central role as a storage of the genetic information.<sup>22</sup> The structures of guanine(G), adenine (A), cytosine (C), thymine (T), uracil (U), and the Watson-Crick base pairs in DNA are shown in **Figure 1.1**.



**Figure 1.1** Schematics of five natural nucleobases and canonical Watson-Crick (WC) pairing, G:C and A:T.

Various conformations of DNA have different biological functions. B-form DNA is the most common structure existing in cells,<sup>23</sup> which is right-handed and having the base pair perpendicular to the double-helix axis. A-form DNA is a shorter chain with a smaller twist angle compared to the B-form; this is considered as a self-protection transition form of DNA under dehydrating conditions.<sup>24</sup> Z-form DNA is a left-handed form in a zigzag pattern. These three forms of DNA contain Watson-Crick base pairs, and the structure of duplex held by Hoogsteen base pairs is shown in **Figure 1.2**. Hoogsteen base pairing was reported after Watson-Crick base pairing and provided another base-pairing method for A:T and G:C base pairs.<sup>25</sup> The formation of G-quadruplex secondary structures, one of the self-assembled nucleobases adducts, relies on the Hoogsteen base pairing between guanines.



**Figure 1.2** Side view of A-form (a), B-form (b), Z-form (c) DNA and the duplex structure (d) held by the Hoogsteen base pairs. From reference 26 with permission.

Metal cations can affect the interactions between base pairs and the functions of nucleic acids.<sup>27–29</sup> Previous works have illustrated that the presence of metal cations such as  $\text{Cu}^{2+}$  and  $\text{Zn}^{2+}$  can induce changes in the DNA double-helical structures between A-form, B-form and Z-form DNA.<sup>26,30</sup> The impacts of metal cations to nucleic acids are affected by the properties of the metal itself,<sup>27,31</sup> the concentration of metal ions,<sup>31</sup> the size and the form of nucleic acids,<sup>32</sup> the polarity of solvents,<sup>33</sup> the pH value of solutions, and some other physical environmental parameters like temperature.<sup>34</sup>

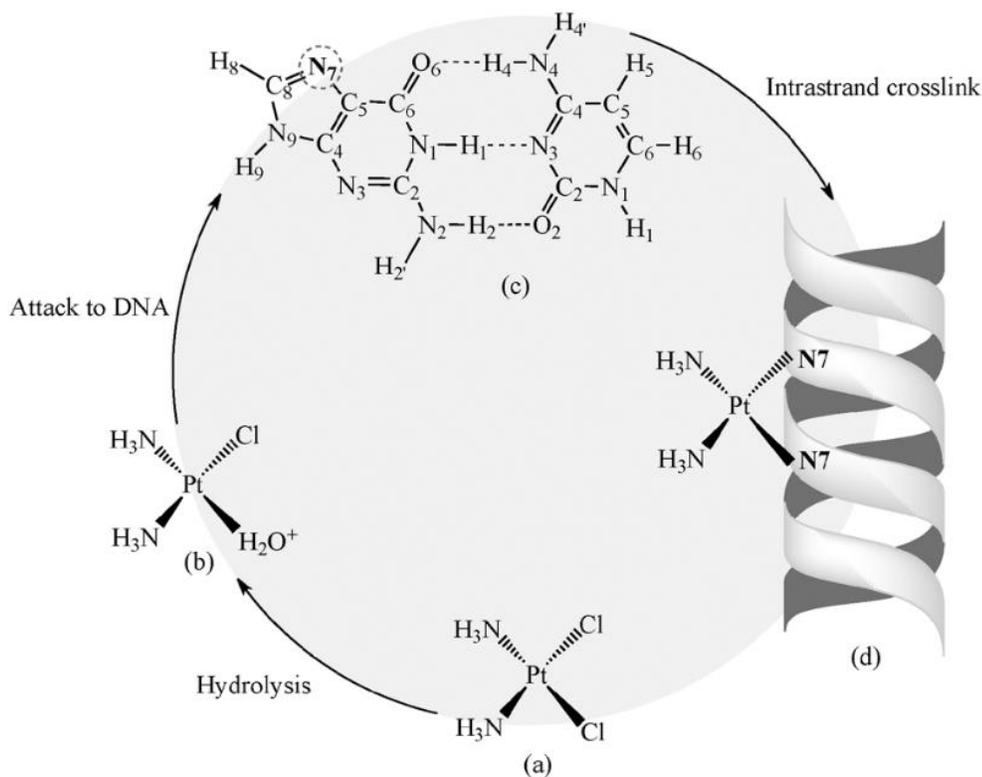
The first discovered metal ion-nucleobase interaction was published in 1956,<sup>35</sup> which was concerned with the pH-dependent interactions between  $\text{Mg}^{2+}$  and adenosine triphosphate (ATP). Since then,  $\text{Mg}^{2+}$  has been considered as one of the most important divalent metal cations *in vivo*. It also contributes to the interaction with oxyanions of RNA – non-bridging oxygens of phosphate of the nucleic acid backbone.<sup>32,36,37</sup>  $\text{Mg}^{2+}$  binds bidentate to 3' and 5' phosphate of RNA affected by strong electrostatic interactions as well as complicated non-electrostatic components, including the charge transfer, the polarization of  $\text{PO}^-$  and the weak exchange interaction. The O's coordinate with  $\text{Mg}^{2+}$  in its first coordination sphere with the strong binding energy.  $\text{Mg}^{2+}$  is selected by nature not only because the radius of  $\text{Mg}^{2+}$  (0.65 Å) is suitable for the vacancies between oxyanions on RNA backbones, but also its higher charge density compared to  $\text{Na}^+$  and  $\text{Ca}^{2+}$ .

Nucleobases are also ligands for metal-containing complexes, such as platinum (Pt).<sup>38</sup> The electron pairs on bare oxygen and nitrogen can form coordination bonds to transition metals. Cisplatin ( $\text{Pt}(\text{NH}_3)_2\text{Cl}_2$ ), an anti-cancer chemotherapy drug containing Pt, has captured scientists' attention for several decades. The structure of cisplatin is depicted in **Figure 1.3(a)**. Cisplatin can kill cancer cells by damaging their DNA and inhibiting

DNA synthesis.<sup>13,39–42</sup> There are many components in the cell that can react with Pt, including RNA and proteins. However, the interaction between DNA and cisplatin is the most crucial step to realize the tumour cure. Once the cisplatin enters the cell, hydrolysis of cisplatin will substitute a water molecule for  $\text{Cl}^-$  (**Figure 1.3(b)**), which allows the hydrolyzed cisplatin to interact with G:C base pair (N7 of G) in the DNA helix (**Figure 1.3(c)**). The formation of cisplatin-DNA adducts impacts the hydrogen bonds between base pairs in the DNA double helix and the  $\pi$ - $\pi$  stacking interaction between nucleotides of the DNA strand, which is shown in **Figure 1.3(d)**. The interaction between cisplatin and the DNA double helix can cause damage and even cancer cell death.<sup>43</sup> However, there are many side effects of cisplatin, such as neurotoxicity, ototoxicity, and tumour resistance.<sup>44</sup> Many other complexes with transition metals,  $\text{Ti}^{4+}$ ,  $\text{Ru}^{3+}$  *etc.*, have also been studied to reveal their anti-cancer capabilities.<sup>45</sup>

Alkali metals are reactive and regularly exist as monocations, which are in great abundance both in nature and organisms, and bound to nucleic acids by ion-dipole interactions.<sup>46</sup> At low concentrations, they prefer to bind to the negatively charged phosphate backbones of nucleic acids by ionic interactions. However, at high concentrations, alkali metal cations interact with the nucleobases and can cause the structural distortion of base pairs by interrupting the hydrogen bonds holding base pairs together.<sup>5,47,48</sup> Additionally, alkali metal cations can modulate the structures of the nucleic acids. For example, RNA synthesis is potentially affected by the presence of alkali metal cations; metal cations can inhibit the chain initiation process of RNA by interacting with the transcribing enzyme.<sup>49,50</sup> Lots of research on alkali metal cationized DNA/RNA structures has been dedicated to understanding the alkali metal-nucleobase interactions,

which may induce permanent mutations. For example, base pair mismatches are regarded as the highest potential reason to cause the destabilization of DNA double helix. The presence of a mismatch can destabilize a duplex by  $7.1 - 41.8 \text{ kJ mol}^{-1}$ , which has been investigated in particular dependence on the concentration of  $\text{Na}^+$ .<sup>26,51,52</sup> It is also manifested by a lowering of the melting temperature of DNA double helix. Alkali metal cations are functional for the formation of quadruplexes in DNA. There was the discovery of a stabilized G-quartet by forming the strong interactions between  $\text{Na}^+$  and  $\text{K}^+$  with the carbonyl groups of four guanines.<sup>53</sup> G-quadruplexes were also found to be most stable in the presence of  $\text{K}^+$ ,<sup>7,10,54-56</sup> which exists at the end region of the chromosomes called telomeres. Telomeres preserve the integrity of genetic information during the replication process and obstruct the fusion of individual chromatids.<sup>57</sup> Also, many kinds of cancer cells can preserve telomere length during replication, which can prevent the death of cancer cells. Thus, telomeres are also considered as a potential valid drug target by cancer therapists.<sup>58</sup>



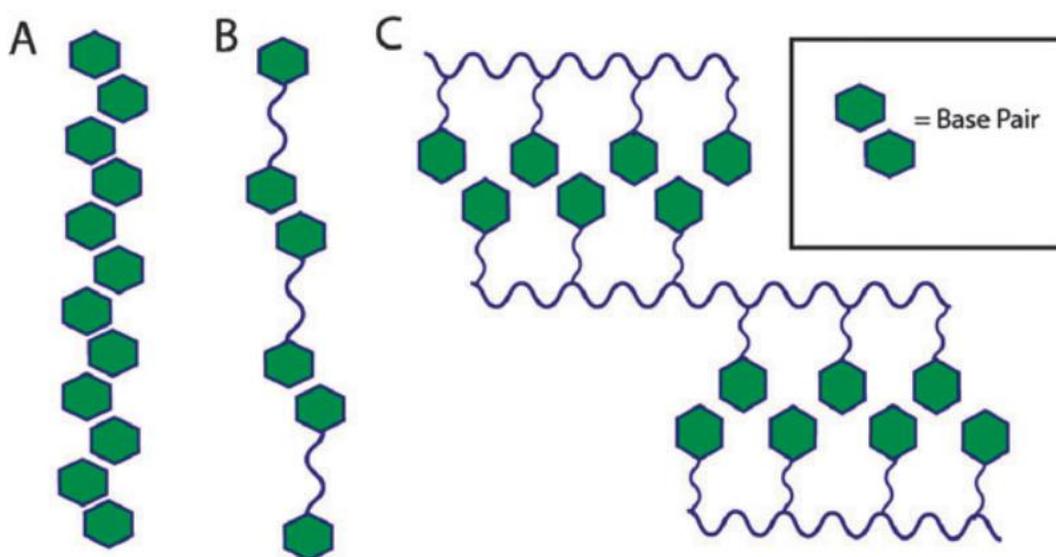
**Figure 1.3** The main biochemical activity of cisplatin. (a) cisplatin first entered the cell; one water molecule is incorporated into the complex to displace one chlorine and form the aqua complex (b). (b) attacks the N7 positions of the adjacent G:C base pairs (c) to form the Pt–DNA intra-strand cross-link adduct (d). Reproduced from reference 43 with permission.

**1.2.2 Materials Applications of Nucleobase/Metal Ion Complexes.** The research of base pairing not only revolutionized the subjects of biology and medicine, but it also inspired the development of supramolecular chemistry. It is based on the unique intrinsic properties of nucleobases, such as the high fidelity and their versatility to form hydrogen bonds in a number of ways (Watson-Crick, Hoogsteen and Wobble).<sup>12</sup> Researchers focus on the use of non-covalent interactions between nucleobases to create supramolecular polymers.<sup>12,13,59–61</sup> Research on employing non-covalent interactions between nucleobases

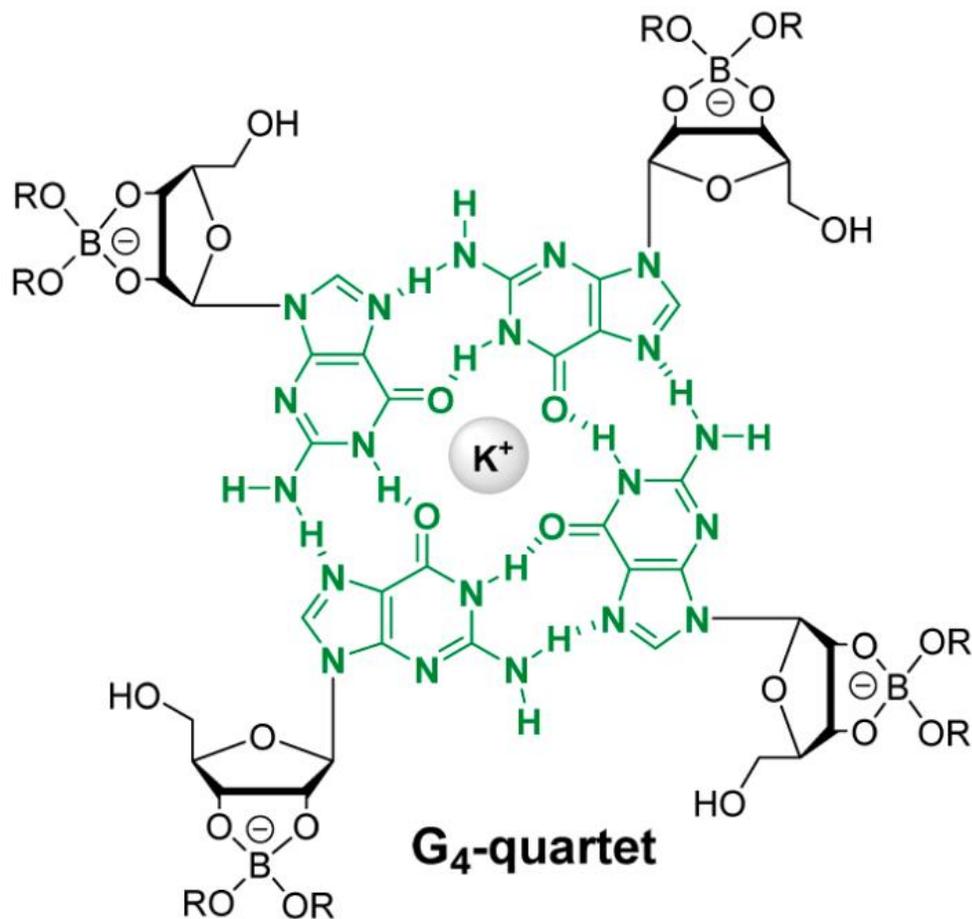
to design biomaterials is on account of the discovered characteristics of nucleobases; they can self-assemble by forming non-covalent interactions. These materials are ‘smarter’ than the covalently bound structures; they can rapidly respond to the change of extrinsic stimuli such as temperature<sup>12</sup> and pH value.<sup>59</sup> Non-covalently bonded structures can still preserve some physical properties of covalent bond concatenated materials. Examples of the chains formed by nucleobase monomers and the created polymers cross-linked by the hydrogen bonds between nucleobases<sup>12</sup> (side-chain) are shown in **Figure 1.4**. Supramolecular materials containing nucleobases were discovered to have other advantages. For example, crosslinked supramolecular hydrogels designed by utilizing the hydrogen bonding between G:C base pairs have demonstrated faster and better drug release performance *in vitro*.<sup>62</sup>

Besides the materials designed by utilizing non-covalent interactions of nucleobases, metal ions and metal-containing complexes can also play a role in the material construction by forming coordination bonds and/or ion-nucleobases interactions. The combination of the coordination bond formed between metal cations and nucleobases and the hydrogen bond existing between nucleobases provides the possibility of creating metal-involved supramolecular structures, if there are suitable spaces and appropriate configurations for metals and nucleobases.<sup>15,63</sup>  $Zn^{2+}$  is proposed to interact with both the N7 of adenine and an oxide of the 5' phosphate group of 5'-adenosine monophosphate (5'-AMP) to form the  $Zn^{2+}$ -bound 5'-AMP dimer. 5'-AMP can interact with  $Zn^{2+}$  to form a hydrogel, which is a fibrous structure and has been imaged by transmission electron microscopy.<sup>59,64,65</sup> In another study, the significance of alkali metal cations on guanosine monophosphate (GMP) molecules were examined and its potential to stabilize the supramolecular structures.<sup>66</sup> The presence of the  $K^+$  located at the center of GMP-quartets

was first attempted to create a transparent gel, even though it is not able to be crystallized.<sup>67</sup> The utilization of borate esters helps to stabilize the guanosine quartets, and it forms the stable nanofibers in the hydrogel with the  $K^+$  located at the center of  $G_4$ . The structure of the stabilized guanosine quartets monomer is shown in **Figure 1.5**.<sup>68,69</sup>



**Figure 1.4** (A) The self-assembled supramolecular polymer via hydrogen bonds between nucleobases. (B) The chains linked by a hydrogen bond between nucleobases to form the supramolecular polymer. (C) The supramolecular polymers cross-linked by the side-chain nucleobases' hydrogen bonds. Reprinted from reference 12 with permission



**Figure 1.5** The structure of the borate ester stabilized guanosine quartets monomer with  $K^+$  located at the center. Reprinted from reference 69 with permission.

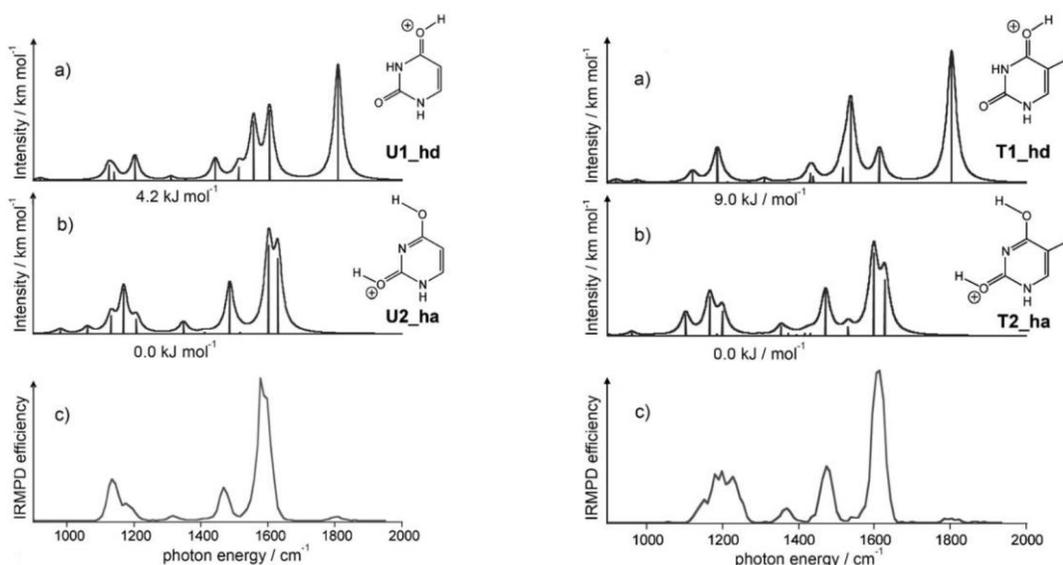
### 1.3 Gas-Phase Studies of Cationized Nucleobases.

There have been increasing numbers of works published exploring gas-phase properties of nucleobase adducts with different metal cations. Since the late 1980s,<sup>70</sup> electrospray ionization (ESI) has allowed researchers to study gas-phase biomolecules without destroying their weak non-covalent interactions. Furthermore, computational

methods can be used to help provide a deeper understanding of the experimental results. The main point of this section is to review the studies of nucleobase complexes with protons, alkali metal cations and calcium cation ( $\text{Ca}^{2+}$ ). The reason to study these cations with nucleobases is due to their high abundance in nature, especially for the alkali metal cations and protons. There are numerous exciting works accomplished concerning nucleobase complexes with other metal cations which are not included in this chapter.<sup>71-78</sup>

**1.3.1 Protonated Nucleobase Complexes.** The first accurate calculation of proton affinities (PA), as well as deprotonation enthalpies of the N-H bond of all five nucleobases using density functional theory (DFT) were conducted and compared by *Chandra et al.* in 1999.<sup>79,80</sup> They employed the B3LYP exchange correction functional<sup>81-83</sup> coupled with 6-31+G(d,p) basis set, and corrected the basis set superposition errors (BSSE)<sup>84</sup> using counterpoise method. The results illustrated that the highest PA sites for uracil, thymine, guanine and adenine are O4(C5 side), O4(C5 side), N7, and N1, respectively. For cytosine, there is only a slight difference between the PA's at O2(N3 side) and N3. Both the intrinsic basicities and acidities of nucleobases are crucial for further understanding of the properties of hydrogen bonds between nucleobases, and the potential solvent effects on the base pair interactions. Several years later, vibrational spectra in the fingerprint region of protonated uracil ( $\text{UH}^+$ ), thymine ( $\text{TH}^+$ ) and cytosine ( $\text{CH}^+$ ) were collected using infrared multiple photon dissociation (IRMPD) spectroscopy at the Centre Laser Infrarouge d'Orsay (CLIO) free-electron laser facility (FEL) in France.<sup>85</sup> The B3LYP/6-31++G(d,p) computed spectra of the lowest energy isomers of all three protonated molecules were in good agreement with the experimental IRMPD spectra. The comparison of the experimental and computed

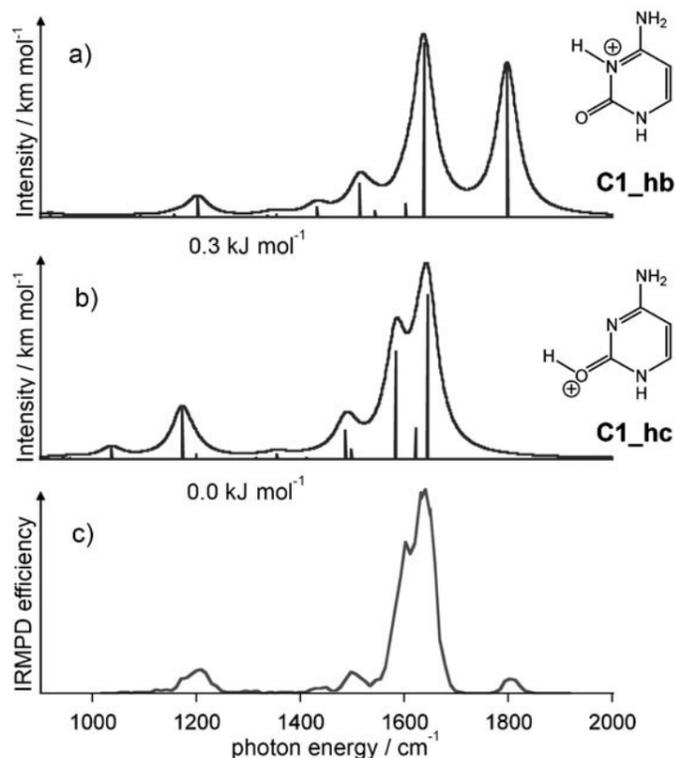
spectra of  $\text{UH}^+$  and  $\text{TH}^+$  are in **Figure 1.6**.  $\text{U2\_ha}$  and  $\text{T2\_ha}$  are conformers that illustrated the proton bound to O2 (N3 side) accompanying the tautomerization of H3 from N3 to O4(N3 side). The computed IR spectra of these two lowest energy isomers reproduce their experimental IRMPD spectra, respectively, except a weak absorption at  $1800\text{ cm}^{-1}$ . This weak band at  $1800\text{ cm}^{-1}$  could be some of the a) isomers (first panel),  $\text{U1\_ha}$  and  $\text{T1\_hd}$ , in both cases.  $\text{U1\_hd}$  and  $\text{T1\_hd}$  are the isomers consistent with the conclusions by *Chandra et al.*, being protonated at O4(C5 side), which has a weak  $1800\text{ cm}^{-1}$  absorption indicating a free C=O stretching vibration. It is noteworthy that the lowest energy structure observed in each case is an enol tautomer by N3O4 tautomerism.



**Figure 1.6** The spectra of  $\text{UH}^+$  on the left and the spectra of  $\text{TH}^+$  on the right. In both cases, isomer a) is the second lowest energy isomer and isomer b) is the predominant structure. Reprinted from reference 85 with permission

The proton affinities of cytosine were reported as  $955.5 \text{ kJ mol}^{-1}$  at N3 and  $956.8 \text{ kJ mol}^{-1}$  at O2(N3 side), which are slightly lower than that of N7 of guanine ( $960.1 \text{ kJ mol}^{-1}$ ) computationally.<sup>79</sup> The N3 and O2(N3 side) of cytosine show almost identical basicities, which were further explored using IRMPD spectroscopy. Computed IR spectra of two lowest-energy species of  $\text{CH}^+$  and their comparison with IRMPD spectra are shown in **Figure 1.7**. The isomer C1\_hc is protonated at O2(N3 side) and is almost isoenergetic with C1\_hb showing protonation at N3; C1\_hc shows the slightly higher stability than C1\_hb,  $0.3 \text{ kJ mol}^{-1}$ , which is consistent with the previously computed PA results. However, C1\_hc better reproduces the spectra of  $\text{CH}^+$  except for a small absorption at  $1800 \text{ cm}^{-1}$ , similar to  $\text{UH}^+$  and  $\text{TH}^+$ . It was concluded that C1\_hc is the predominant isomer of  $\text{CH}^+$  mainly because of its consistency with the IRMPD spectrum, and there is a small amount of contribution by C1\_hb.

The slightly higher energy isomers for  $\text{UH}^+$ ,  $\text{TH}^+$  and  $\text{CH}^+$  were not able to be ruled out as existing either on a thermodynamic nor an experimental basis. The enol isomers of all three cases have been shown experimentally and computationally to be the lowest energy structures and the most abundant species, whereas the isomers consistent with the PA calculations originally done by *Chandra* also seem to exist to a small extent for protonated bases in the gas-phase.



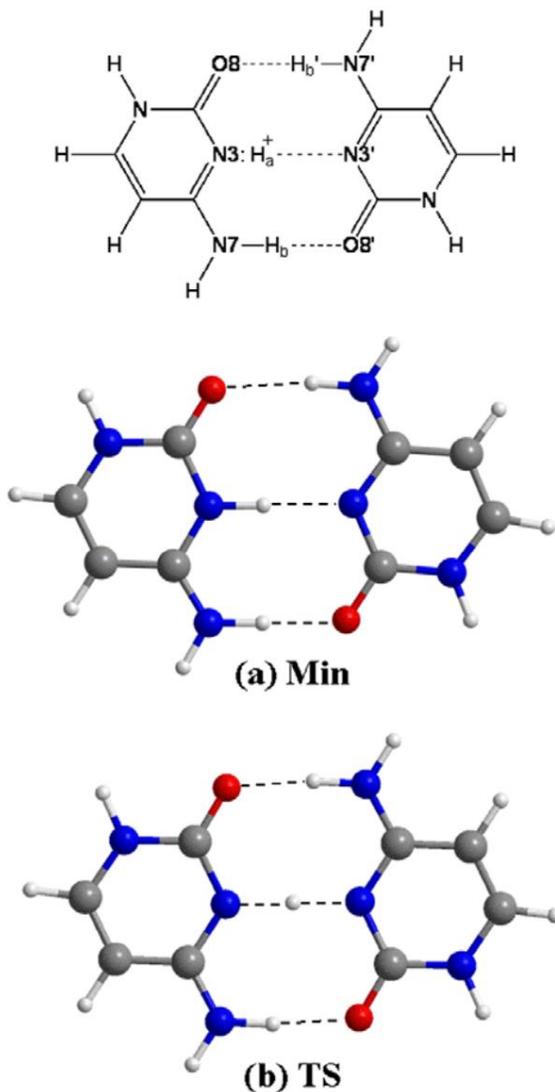
**Figure 1.7** The spectra comparison of  $\text{CH}^+$ . a) The second lowest energy isomer, C1\_hb. b) The concluded predominant and most stable isomer, C1\_hc. c) The IRMPD spectrum of the  $\text{CH}^+$  molecule. The figure is reprinted from reference 85 with permission.

Some research has also been conducted to explore the protonated homo- and hetero-dimeric nucleobase complexes.<sup>6,86-93</sup> The stability of cytosine dimer, the basis of the i-motif ( $\text{C}_2\text{H}^+$ ) structure in nucleic acids, relies on the presence of protons. The i-motif structure is considered to be unstable and virtually non-existent at basic physiological pH.<sup>86</sup> However, *in vivo* i-motif structures may be in charge of gene expression in the regulation and telomeric regions in human DNA.<sup>94</sup> A theoretical study of the protonated cytosine dimer was first published in 2006,<sup>86</sup> long before the discovery of i-motif *in vivo* in 2018.

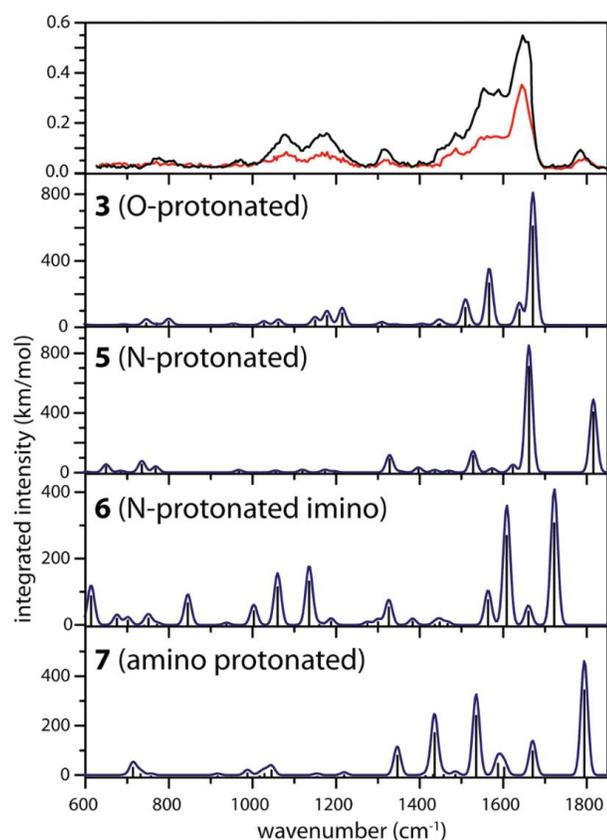
B3LYP/6-311+G(2d,p) and MP2 computational methods were used to calculate the dissociation energy and the energy barrier for proton transfer between two cytosines in i-motif, which were found to be  $172.4 \text{ kJ mol}^{-1}$  and  $6.7 \text{ kJ mol}^{-1}$ , respectively. The lowest energy structure of protonated cytosine dimer was determined to be in the symmetric planar geometry with two cytosines anti-parallel to one another. Further, the proton is bound to the N3 of one cytosine with a bridging hydrogen bond to N3 of the other monomer. The optimized protonated cytosine dimer and its computed transition state are shown in **Figure 1.8**.

IRMPD spectrum of the protonated dimer of 1-methylcytosine (1mC) was collected in 2010, and the N1 position is blocked because it is the binding site to sugar in DNA.<sup>87</sup> The results revealed that the structure of the protonated dimer is the one where proton is located between two N3s. As well, a CO<sub>2</sub> laser was used to irradiate and dissociate the protonated dimer complex to the protonated 1mC monomer, (1mC)H<sup>+</sup> (and neutral 1mC), and collect the IRMPD spectrum of (1mC)H<sup>+</sup>. This spectrum was compared to another spectrum of (1mC)H<sup>+</sup> electrosprayed directly from the solution. These two IRMPD spectra are compared with computed IR spectra of (1mC)H<sup>+</sup> isomers in **Figure 1.9**. The black trace in the first panel is the (1mC)H<sup>+</sup> IR spectrum born by dissociation of the dimer, and the red trace is the spectrum of the electrosprayed protonated monomer. The two spectra are consistent. By comparison with the IRMPD spectra, the O4 protonated isomer is predominant (second panel in **Figure 1.9**), and the N3 protonated isomer also seems to present in small abundance (third panel in **Figure 1.9**) as the result of the presence of an absorption at  $1800 \text{ cm}^{-1}$ . The N-protonated imino and amino isomers are unlikely to exist based on comparisons of their computed IR spectra with the experimental IRMPD spectrum.

This study reveals that there is a proton transfer from N3 to O2 that occurs as the protonated dimer dissociates to  $\text{CH}^+$  and neutral cytosine.



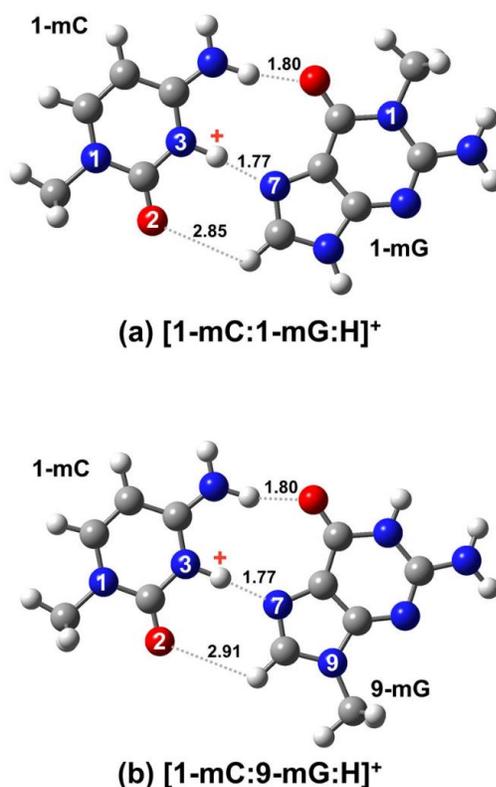
**Figure 1.8** A schematic of the protonated dimeric cytosine (On the top) (a) The optimized global minima of the cytosine dimer with a proton bound to N3 (b) The transition-state of the dimeric cytosine structures. Reprinted from reference 86 with permission.



**Figure 1.9** Experimental IRMPD spectrum of the protonated monomer of 1-methylcytosine,  $m/z$  126, at the top panel: black trace from the dissociation of the proton-bound dimer by  $\text{CO}_2$  laser photolysis; red trace sourced from electrospray ionization. The spectra comparison with normal modes calculated at B3LYP/6-31G\*\* (for which frequencies  $>850\text{cm}^{-1}$  have been scaled by 0.97, with  $20\text{ cm}^{-1}$  Gaussian broadenings) for four possible tautomers. Reprinted from reference 87 with permission.

Protonated base pairs have also attracted significant attentions.<sup>88-90,93,95,96</sup> A theoretical investigation of the protonated GC base pair ( $\text{GCH}^+$ ) showed that the  $\text{C}^{\text{---}}\text{H}^+\text{G}$  isomer is less stable than  $\text{CH}^+\text{---G}$  isomer, despite guanine having a slightly higher PA than cytosine.<sup>88</sup> Collision induced dissociation (CID) methods<sup>90,97</sup> with different collision energies and IRMPD spectroscopy experiments<sup>89</sup> in the fingerprint region were performed

on the protonated GC base pair. The most recent study of  $GCH^+$  system by Han's group<sup>97</sup> concluded that there are different dissociation fragmentation ratios for protonated 1-methylcytosine:9-ethylguanine ( $1mC:9eG$ ) $H^+$  and protonated 1-methylcytosine:1-methylguanine ( $1mC:1mG$ ) $H^+$ . The predicted structures of ( $1mC:9mG$ ) $H^+$  and ( $1mC:1mG$ ) $H^+$  were found to have geometries analogous to the Hoogsteen G:C base pair with the proton bound to cytosine in both cases shown in **Figure 1.10**, despite the higher proton affinity for N7 of guanine.

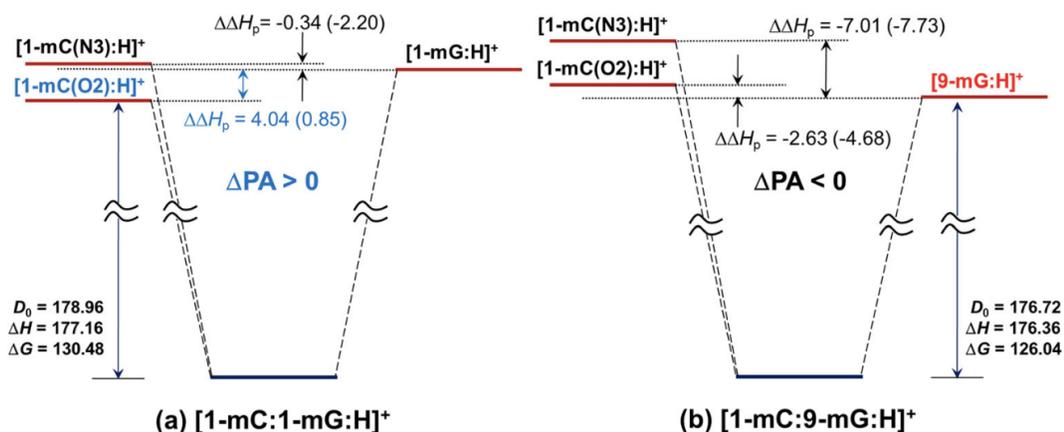


**Figure 1.10** The predicted structure of the (a) ( $1mC:1mG$ ) $H^+$  and b) ( $1mC:9mG$ ) $H^+$  at the B3LYP/6-311+G(2d,2p) level with the Gaussian 09 suite of programs. Reprinted from reference 97 with permission.

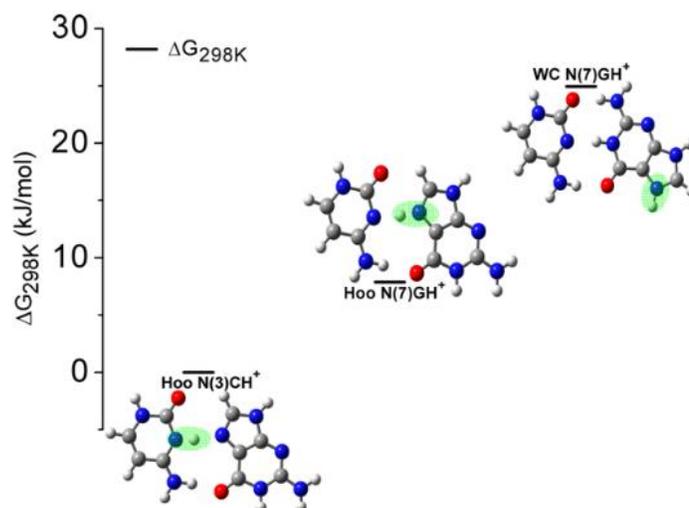
Furthermore, based on the computed gas-phase basicities of 1mC, 1mG and 9mG, the expected abundance of (1mC)H<sup>+</sup> should only be about 55% and 10% for CID of (1mG:1mC)H<sup>+</sup> and (9mG:1mC)H<sup>+</sup>. However, in these systems the abundance of (1mC)H<sup>+</sup> was shown to be 80% for (1mC:1mG)H<sup>+</sup> and 40% for (1mC:9mG)H<sup>+</sup> dissociation. As the computed stability of fragments is in the order of (9mG)H<sup>+</sup> > (1mC(O2))H<sup>+</sup> > (1mG)H<sup>+</sup> > (1mC(N3))H<sup>+</sup> on the potential energy diagram, shown in **Figure 1.11**, the difference between the experimental fragmentation ratios in two cases was explained by the proton transfer process of (1mC)H<sup>+</sup> from N3 to O2, which is observed by previous protonated cytosine research.<sup>87</sup> (9mG)H<sup>+</sup>, which is 60% of the fragment products of (1mC:9mG)H<sup>+</sup> dissociation, is more stable relative to (1mC)H<sup>+</sup>. In the (1mC:1mG)H<sup>+</sup> case, the (1mC)H<sup>+</sup> with proton bound to O2(N3 side) is preferred than (1mG)H<sup>+</sup>. However, the *anomalous* CID abundance is not able to be predicted based on their computed gas-phase basicities.

In another study, the IRMPD spectra of GCH<sup>+</sup> were electrosprayed from solutions at two pH values, 5.8 and 3.2, in the 900-1900 cm<sup>-1</sup> region.<sup>89</sup> At pH=5.8, the precursor ion was fragmented by the FEL laser to produce GH<sup>+</sup> and CH<sup>+</sup> products in an 85:15% ratio. Based on the IRMPD spectrum, this ion was the protonated Watson Crick G:C base pair with proton at N7 of guanine, (WC N(7)GH<sup>+</sup>), despite it being ~20 kJ mol<sup>-1</sup> higher in 298K Gibbs energy compared to the lowest energy Hoogsteen G:C base pair with the proton bound to N3 of cytosine (Hoo N(3)CH<sup>+</sup>). The relative Gibbs energy diagram of the three lowest energy isomers is shown in **Figure 1.12**. On the other hand, at pH=3.2, the dissociation of the electrosprayed dimer presented a GH<sup>+</sup>: CH<sup>+</sup> ratio near 50:50. Moreover, the IRMPD spectrum of GCH<sup>+</sup> in this solution is reproduced by the lowest energy

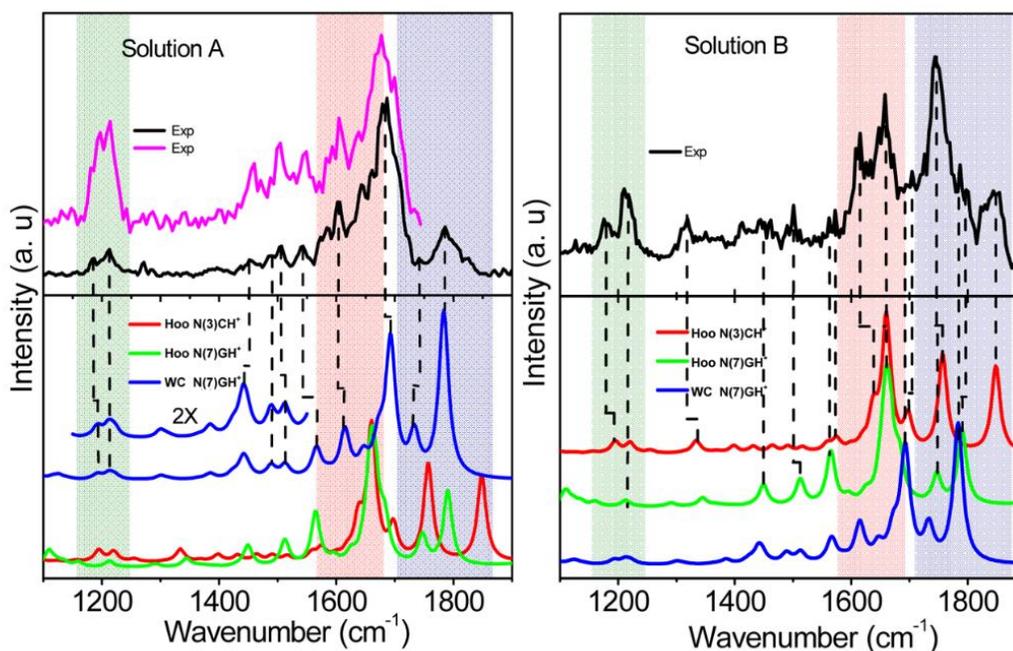
Hoogsteen isomer, Hoo N(3)CH<sup>+</sup>. The spectra of GCH<sup>+</sup> from the two different pH solutions are shown in **Figure 1.13**.



**Figure 1.11** Computed dissociation diagrams of (a) (1mC:9eG)H<sup>+</sup> and (b) (1mC:9eG)H<sup>+</sup> from their lowest energy structures. The predicted energies,  $D_0/\Delta H/\Delta G$  (dissociation energies, relative enthalpies and relative Gibbs energies), were obtained by using the CBS-QB3 theory in kJ mol<sup>-1</sup> at 298.15 K.  $\Delta PA = \Delta PA(1mC) - \Delta PA(xmG)$ . Reprinted from reference 97 with permission.



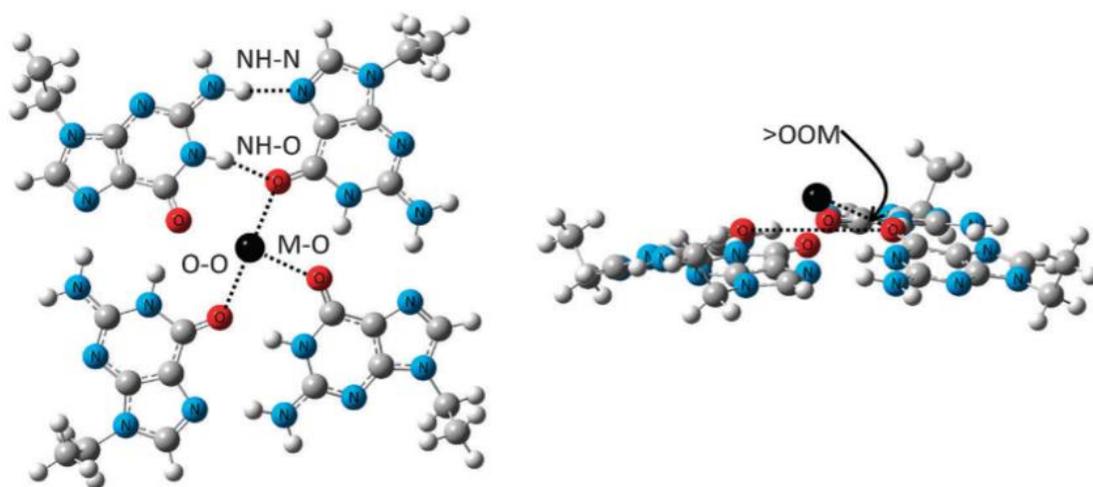
**Figure 1.12.** The Hoo N(3)CH<sup>+</sup> is the global minima, followed by Hoo N(7)GH<sup>+</sup>. The WC N(7)GH<sup>+</sup> is the third lowest 298K Gibbs energy structure concluded for GCH<sup>+</sup> produced from the pH=5.8 solution. The computed structures were optimized at the M06-2X/6-311G++(d,p) level. Reprinted from reference 89 with permission.



**Figure 1.13.** Experimental IRMPD spectra of the protonated dimeric molecule, GCH<sup>+</sup>, in top panels: black trace and pink trace from the dissociated proton-bound G:C dimer by FEL laser for solution A, pH=3.2, and solution B, pH=5.8. Bottom panels: the red, green and blue curves stand for the computed IR intensities for Hoo N(3)CH<sup>+</sup>, Hoo N(7)GH<sup>+</sup> and WC N(7)GH<sup>+</sup>, respectively, at the M06-2X/6-311G++(d,p) level with the Gaussian 09 suite of programs. Reprinted from reference 89 with permission.

**1.3.2 Metal Cationized Nucleobase Complexes.** Alkali metal cations are known to be in great abundance in nature and organisms. As we previously mentioned, alkali metal cations have been discovered to be capable of stabilizing the G-quadruplex in DNA *in vivo*. Many studies<sup>7,9,53,98–101</sup> on G-quartets and G-quadruplexes in the presence of different alkali metal cations have been published since its first discovery in 1962.<sup>102</sup> To understand the intrinsic intermolecular bonding properties and thermostabilities, the first spectroscopic study of the gas-phase sodium cationized G-tetrads were published in 2014 by *Fraschetti* and co-workers<sup>103</sup> following several previous gas-phase computational works on G-tetrads and G-quadruplexes with alkali metal cations.<sup>7,101,104–107</sup> The IRMPD spectra of gaseous 9-ethylguanine tetrads bound with alkali metal cations ((9eG)<sub>4</sub>M<sup>+</sup>, M=Li, Na, K, Rb, Cs) were collected in both the fingerprint and C-H/N-H/O-H regions and were used to determine the structures of alkali metal cationized G-tetrads. It was suggested that alkali metal cations were centrally bound to the carbonyl oxygen of guanines that were in turn bound to two other neighbouring guanines by two hydrogen bonds with each.<sup>108</sup> The computed structure is illustrated in **Figure 1.14**. Energy-resolved CID was also employed to determine the relative gas-phase stabilities of the tetrads. The (9eG)<sub>4</sub>Na<sup>+</sup> was found to be the most stable to CID, following by (9eG)<sub>4</sub>Li<sup>+</sup> > (9eG)<sub>4</sub>K<sup>+</sup> > (9eG)<sub>4</sub>Rb<sup>+</sup> > (9eG)<sub>4</sub>Cs<sup>+</sup>, while there is only slight difference of the collision energies needed among K<sup>+</sup>, Rb<sup>+</sup> and Cs<sup>+</sup> complexes. However, the expected ion-dipole interactions between metal cations and nucleobases follow a sequence of Li<sup>+</sup>>Na<sup>+</sup>>K<sup>+</sup>>Rb<sup>+</sup>>Cs<sup>+</sup> as a result of the decreasing electron density of the metal cation; Li<sup>+</sup> would be expected to bind most strongly to 9eG in tetrad among all five metals. The *anomaly* was probed further by computing the distortion or destabilization energies of (9eG)<sub>4</sub>M<sup>+</sup> due to the metal cation. This destabilization energy was defined as

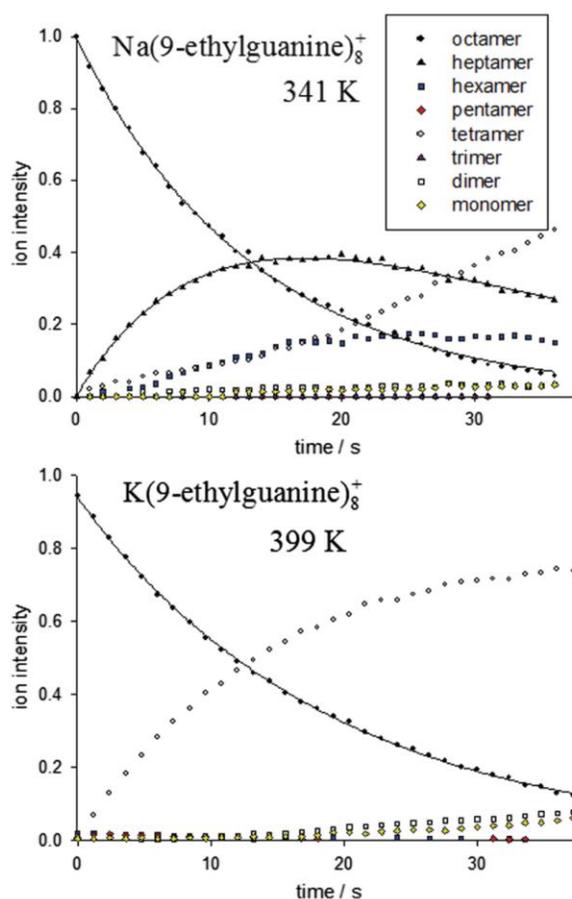
the difference between the energy of the neutral G-tetrad and the single point energy of each  $(9eG)_4M^+$  with the metal cation removed. As  $Li^+$  has the largest impact on G-tetrad by twisting the geometry of the hydrogen bound guanines, it also destabilizes the hydrogen bonding in the tetrad, which offsets the strong ion-dipole interactions.  $Na^+$  occupies the centre of a planar tetrad and destabilizes the hydrogen bonding interactions much less than  $Li^+$  resulting in the higher overall stability of  $(9eG)_4Na^+$  to dissociation.



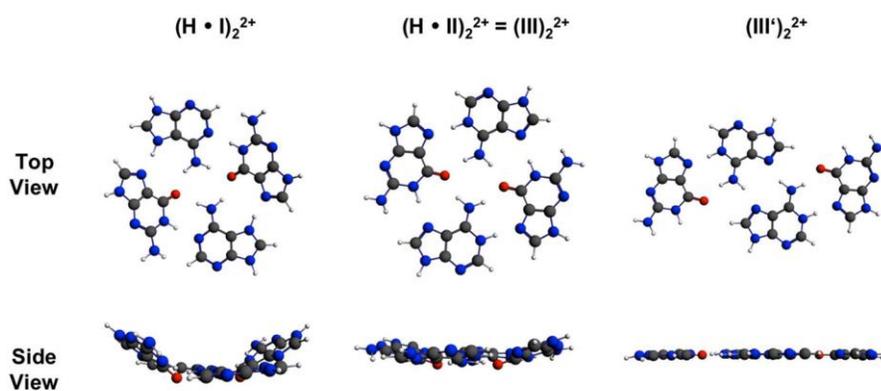
**Figure 1.14** The computed structure of alkali metal cationized G-tetrad, top view (left) and side view (right).  $>OOM$  is the angle of two opposite Os and an alkali metal cation. Reprinted from reference 108 with permission.

Further studies on the gas-phase G-quadruplex,  $(9eG)_8M^+$ , revealed that the  $K^+$  complex was by far the most stable structure to thermal blackbody infrared radiative dissociation (BIRD) followed by the  $Rb^+$  complex and the  $Cs^+$  complex, and finally, the least stable complex was the  $(9eG)_8Na^+$  (the  $Li^+$  quadruplex was not observed).<sup>109</sup> The experimental BIRD breakdown plots for  $(9eG)_8K^+$  and  $(9eG)_8Na^+$  are shown in **Figure 1.15**.

Comparing to  $(9eG)_8Na^+$ ,  $(9eG)_8K^+$  has a higher decomposition temperature (over 50 K higher) over the same time frame and similar dissociation. Another difference is that while  $(9eG)_4Na^+$  dissociates via loss of a single  $9eG$  monomer,  $(9eG)_8K^+$  dissociates by losing a neutral tetrad. The first-order BIRD kinetics plots indicate only one conformer exists for each  $(9eG)_8M^+$  complex. The lowest energy conformer for both  $M=Na$  and  $K$  is a sandwiched structure with metal cation between parallel tetrads. The Arrhenius activation energy for  $(9eG)_8K^+$  was found to be much higher than that of  $(9eG)_8Na^+$  (176 vs 74 kJ mol<sup>-1</sup>), which was verified computationally. In conclusion, the  $(9eG)_8K^+$  has higher stability than  $Na^+$ -mediated octamer, which is different from the results of G-tetrad with alkali metal cation. It is explained: the cavity of G-tetrad is in the best fit for  $Na^+$  leading to the presence of  $Na^+$  at the center of the planar G-tetrad. For  $Na^+$  to form quadruplex, it must be pulled from the centre of the tetrad to interact with the other tetrad. However, the heavier metal,  $K^+$ , sits above the tetrad is capable of binding another G-tetrad strongly.  $K^+$  has been found to be associated with the G-quadruplex in nature despite lower concentration in the cell compared to  $Na^+$ . As such the reason for the presence of  $K^+$  has garnered much attention. Previous computational work<sup>7</sup> in the solution phase determined the reason that  $Na^+$  is not associated with G-quadruplexes is that it has larger hydration energy than  $K^+$  such that  $Na^+$  would rather be solvated than bound in the quadruplex. However, the gas-phase work done by Azargun *et al.*<sup>109</sup> where the  $K^+$  tetrad is far more stable to dissociation in the gas phase, in the absence of solvent, contradicts to some extent that the presence of  $K^+$  in quadruplexes is due solely to the higher hydration energy of  $Na^+$ .



**Figure 1.15** BIRD breakdown plots for  $(9eG)_8\text{Na}^+$  (top) and  $(9eG)_8\text{K}^+$  (bottom) at 341 K and 399 K, respectively. Reprinted from reference 109 with permission.



**Figure 1.16.** Three lowest energy conformers of divalent complex,  $(\text{H}_2 - \text{AGAG})_2^{2+}$ . The stability of three isomers is in the sequence of  $(\text{H} \cdot \text{I})_2^{2+} \approx (\text{H} \cdot \text{II})_2^{2+} > (\text{III}')_2^{2+}$ . Reprinted from reference 93 with permission.

Gaseous nucleobase monomers, dimers, tetramers and larger clusters with alkali metal cations have been investigated by different groups.<sup>5,93,101,107,110–116</sup> The guanine-adenine mixed tetramer (AGAG), which is analogous to G-tetrads, constituted by two guanines and two adenines and bound by either protons or alkali metal cations ( $H^+$ ,  $Li^+$ ,  $Na^+$  and  $K^+$ ) has been investigated computationally.<sup>93</sup> The structures of the three lowest energy structures of the  $(H_2-AGAG)^{2+}$  are shown in **Figure 1.16**.  $(H \cdot II)_2^{2+}$  is in a similar geometry as the most stable conformer for neutral AGAG. It is protonated at adenine's most basic site, N1, which is hydrogen-bonded to N7 of guanine, the most basic site of guanine. Conformer  $(H \cdot I)_2^{2+}$  is protonated at N7 of each adenine and is hydrogen bonded to N7 of guanine. Protonating at N7 of adenine reduces the repulsive forces between lone pairs of nitrogens (N7 of adenine and N7 of guanine), leading to  $(H \cdot I)_2^{2+}$  being a competitive conformer compared to  $(H \cdot II)_2^{2+}$  with the existence of protons. The planar  $(III')_2^{2+}$  structure is less stable than  $(H \cdot II)_2^{2+}$  by  $28.0 \text{ kJ mol}^{-1}$  and  $24.7 \text{ kJ mol}^{-1}$  298K Gibbs energy in the aqueous and gas phase, respectively, and it is also computed to be less stable in the solid state according to the previous calculation.<sup>117</sup>

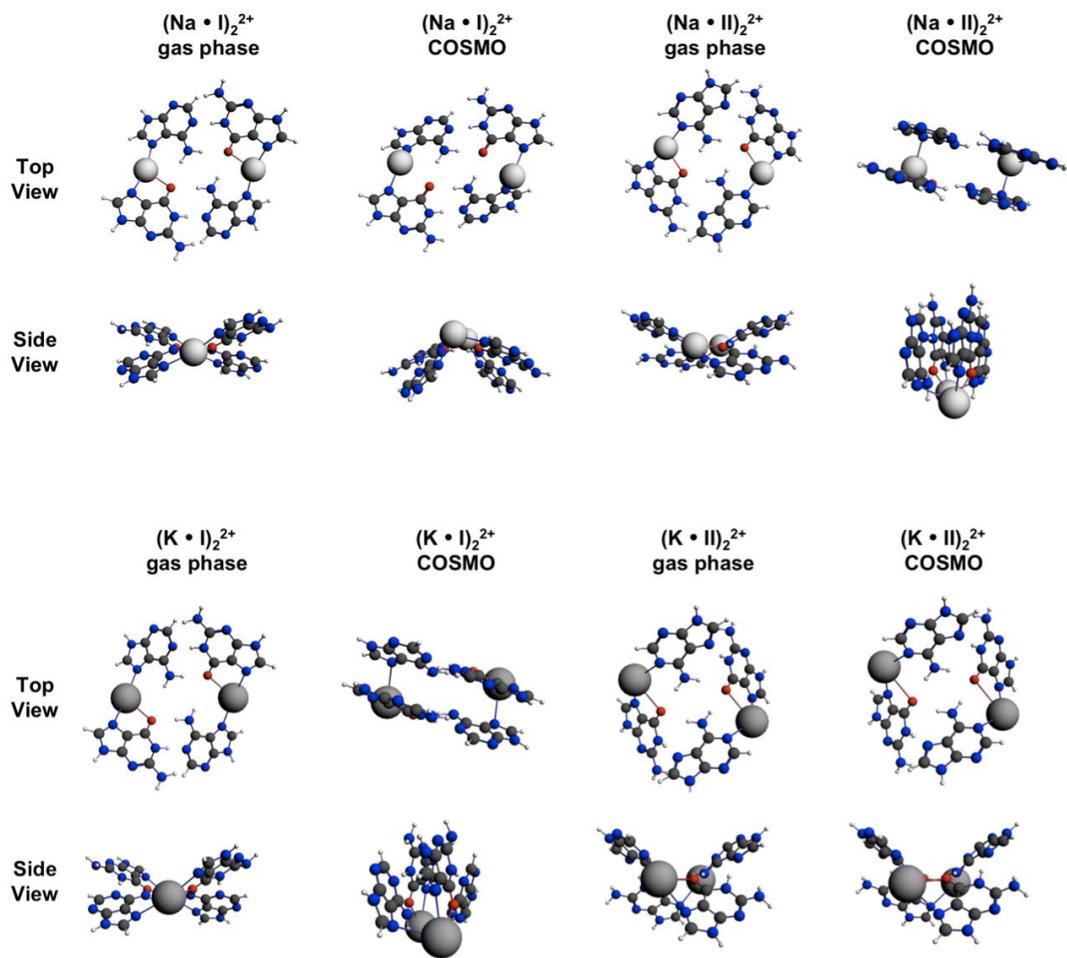
Other attempts to avoid the repulsion forces between the nitrogens of adenine and guanine used alkali metal cations in the AGAG quartets instead of protons.<sup>93</sup>  $Na^+$  and  $K^+$  both induce distortions, leading to non-planar structures shown in **Figure 1.17(a)**. These structures have significantly longer and weaker intramolecular hydrogen bonds compared to the protonated AGAG quartets in the gas phase. In the aqueous phase, the structures  $(K \cdot I)_2^{2+}$  and  $(Na \cdot II)_2^{2+}$  are twisted into two paralleled mismatches interacted by  $\pi$ - $\pi$  interactions, but not hydrogen bonds.  $Li^+$  has fewer distortion effects on the AGAG quartet structure compared to  $Na^+$  and  $K^+$ . The near-planar structures are stabilized by two  $Li^+$  in

both the gas and solution phase, analogous to the protonated AGAG tetramer. The inter-base hydrogen bonds are still observed, and  $\text{Li}^+$  is tridentate to N7 and O6 of guanine and N1 or N7 of adenine. The two optimized semi iso-energetic conformers of  $(\text{Li}_2 - \text{AGAG})^{2+}$  in the gaseous and aqueous phases are shown in **Figure 1.17(b)**. Based on the computed thermodynamic results and structural parameters such as bond distance and bond angle,  $\text{Li}^+$  is predicted to be the most preferred alkali metal cation to stabilize the AGAG-quartet, which can stabilize the structure when there are two alkali metal cations introduced into the tetrads.

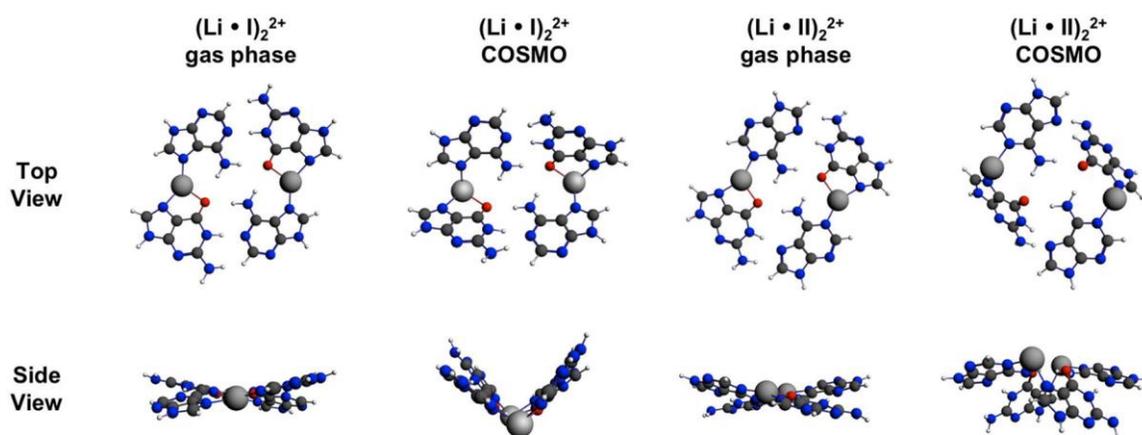
The interactions between metal cations and nucleobases, especially the effects of metal cations on nucleobase tautomerization, have been studied in the gas-phase by both experimental and computational strategies.<sup>118</sup> The adenine- $\text{M}^+$  was determined to exist in the tautomeric form which has the hydrogen bound to N7 (N7 tautomer) instead of N9. The N7 tautomer was found to exist when adenine complexed with  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Cs}^+$ ; while the metal cations were found to bidentate to N3 and N9 of adenine. This was explored by both the IRMPD spectroscopy and theoretical methods (B3LYP/6-31+G(d,p) and MP2).<sup>118,119</sup> So far, there has been no evidence provided for tautomerization of guanine induced by complexing with alkali metal cations. For purines, cytosine shows the highest preference to keep the canonical form with  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Rb}^+$  and  $\text{Cs}^+$  bound to O2 and N3, which was verified by IRMPD spectroscopy and computational methods.<sup>118,120</sup> Both the experimental and computational methods were also employed to explore thymine and uracil monomer with  $\text{Li}^+$  and  $\text{Na}^+$ , which were both determined to be mono-coordinated by  $\text{Li}^+$  or  $\text{Na}^+$  at the same binding position, O4, of the canonical isomer.<sup>50,118,121,122</sup> However, the divalent cations, such as  $\text{Mg}^{2+}$  and  $\text{Zn}^{2+}$ , have the capability to induce tautomerization in thymine

theoretically; the canonical thymine-cation isomer is  $\sim 100 \text{ kJ mol}^{-1}$  higher in energy than the amide tautomer with the cation, while the neutral keto-enol tautomer is  $\sim 42.7 - 114.6 \text{ kJ mol}^{-1}$  higher than canonical thymine before metal complexation.<sup>118</sup> Thus, the higher energy uracil tautomer needs to be considered when it interacts with divalent metal cations. Even though the uracil monomer with alkali metals has not shown the preference of tautomerization, complexation with divalent cations such as  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Sr}^{2+}$ ,  $\text{Ba}^{2+}$ ,  $\text{Pb}^{2+}$  can tautomerize uracil.<sup>72,123–125</sup> A theoretical study for the complex between uracil and  $\text{Ca}^{2+}$  illustrates that the proton on N3 prefers to tautomerize to O4, which allows  $\text{Ca}^{2+}$  to interact with O2 and N3 bidentate.<sup>125</sup> The IRMPD spectrum in the mid-infrared region of deprotonated dimeric uracil with  $\text{Ca}^{2+}$ ,  $(\text{Uracil})(\text{Uracil-H})\text{Ca}^{2+}$  was collected.<sup>123</sup> Schematics of preferred thermodynamic structures of  $\text{UracilCa}^{2+}$  and  $(\text{Uracil-H})(\text{Uracil})\text{Ca}^{2+}$  are depicted in **Figure 1.18**. The spectra comparison is also consistent with the computational results. The lowest energy structure of  $\text{UracilCa}^{2+}$  and  $(\text{Uracil})(\text{Uracil-H})\text{Ca}^{2+}$  contains the N3O4 iminol tautomer, which allows  $\text{Ca}^{2+}$  bidentate to this tautomerized uracil by N3O2 (It is also referred to keto-enol tautomerization in Chapter 3). In  $(\text{Uracil})(\text{Uracil-H})\text{Ca}^{2+}$ , the deprotonated uracil is formed by losing a proton at N3. Thus, the metal cation is tetradentate to N3O4 (deprotonated uracil)/ N3O2 (tautomerized uracil), and there is a hydrogen bond between monomers, O4-H4 (tautomer)---O2 (deprotonated uracil). Furthermore, larger clusters of uracil self-assemblies,  $\text{U}_4$  to  $\text{U}_{14}$ , with  $\text{Ca}^{2+}$  were examined by both theoretical and experimental methods.<sup>126</sup> Master equation modelling of blackbody infrared radiative dissociation (BIRD) kinetics data were performed and compared to experimental BIRD values to determine Arrhenius parameters and threshold dissociation energies. The obtained threshold dissociation energies and theoretically calculated binding

energies of  $U_6Ca^{2+}$  are higher than other more massive clusters, which indicates that  $U_6Ca^{2+}$  can be the inner core for  $U_nCa^{2+}$  ( $n=7-14$ ).

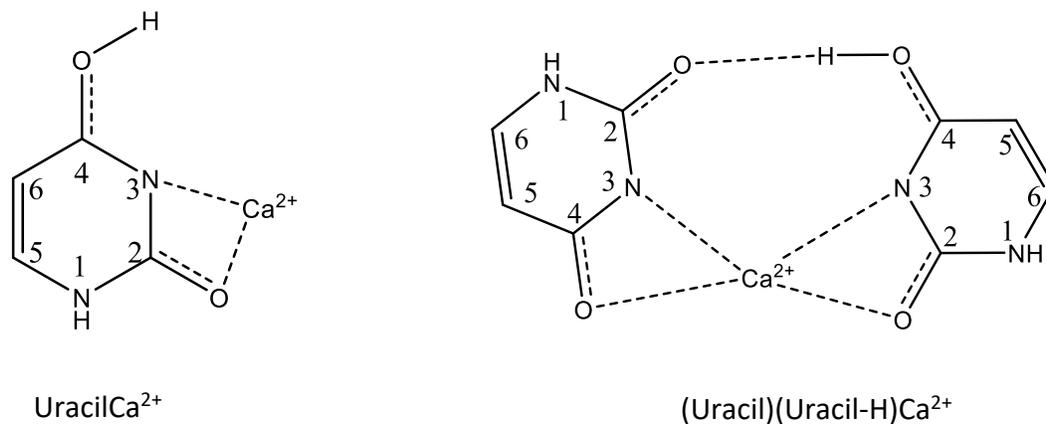


(a)



(b)

**Figure 1.17.** (a) Two preferred structures,  $(M \cdot I)_2^{2+}$  and  $(M \cdot II)_2^{2+}$ , where  $M = Na, K$ , in the gas phase and water phase. (b) Two preferred structures,  $(Li \cdot I)_2^{2+}$  and  $(Li \cdot II)_2^{2+}$ , of  $(Li_2 - AGAG)^{2+}$  in the gas phase and water phase. Both the top and side views of the structures are shown. Gas-phase and COSMO stand for the structures optimized in the gas phase and water phase, respectively. Reprinted from reference 93 with permission.



**Figure 1.18.** Schematic of the UracilCa<sup>2+</sup> (left) and the (Uracil-H)(Uracil)Ca<sup>2+</sup> (right).

## 1.4 The Content of This Thesis.

In this thesis, mass spectrometric and computational methods are applied to study protonated and metal cationized nucleobase complexes. In Chapter 2, there are detailed descriptions of the methods.

In chapter 3 and chapter 4 self-assemblies of nucleobases with cations are discussed. Previous studies on the 4- to 14-mer uracil complexes with  $\text{Ca}^{2+}$  pointed out that 6-mer uracil cluster played as the center of larger molecules, and proposed structures of the tetramer, pentamer and hexamer were constituted only by the canonical uracils. However, uracil has shown the preference to form N3O4 keto-enol tautomer when interacting with  $\text{Ca}^{2+}$  (monomer and deprotonated dimer). In Chapter 3, the structures of uracil tetramer, pentamer and hexamer bound by calcium dication were investigated by IRMPD spectroscopy in both fingerprint and C-H/O-H/N-H regions and computational methods.

The symmetric anti-parallel planar structure was concluded as the only contributor to the cytosine dimer with  $\text{Li}^+$ ,  $\text{Na}^+$  and  $\text{K}^+$  in the gas phase experimentally and thermodynamically.<sup>110</sup> However, in the potassium cationized cytosine dimer the red-shifted shoulder absorption in the high energy region (experimentally) attracted our attention. Research on the self-assembled 1mC dimers with alkali metal cations are discussed in Chapter 4. The red-shifted  $\text{NH}_2$  absorption in CH/OH/NH region spectra can be interpreted by intrinsic binding properties, thermodynamic results and spectra comparison.

The contents of chapter 5 and chapter 6 focus on mixed base pairs interacting with cations. The *anomalous* abundance of fragments for protonated G:C molecules (including the alkyl G:C) indicates the direction to discover the reason for this *anomaly* further.

Furthermore, the effect of alkali metal cations to G:C base pair has not been discussed. In chapter 5 guanine and cytosine base pairs, employing 9-ethylguanine and 1-methylcytosine to eliminate the hydrogen bonding possibilities from the sites bound to sugar in nucleic acids, with  $H^+$ ,  $Li^+$ ,  $Na^+$ ,  $K^+$ ,  $Rb^+$  and  $Cs^+$  complexes are described in Chapter 5, studied by both computational and experimental methods. While the protonated guanine-containing mismatches are discussed in Chapter 6 because these molecules have been investigated as the most stable mismatches structures, and the proton is in great abundance in nature.

## References

- 1 L. Betts, J. A. Josey, J. M. Veal and S. R. Jordan, *Science.*, 1995, **270**, 1838–1841.
- 2 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402–5415.
- 3 K. Sakamoto, H. Gouzu, K. Komiya, D. Kiga, S. Yokoyama, T. Yokomori and M. Hagiya, *Science*, 2000, **288**, 1223–1226.
- 4 T. Shimizu, R. Iwaura, M. Masuda, T. Hanada and K. Yase, *J. Am. Chem. Soc.*, 2001, **123**, 5947–5955.
- 5 B. A. Cerda and C. Wesdemiotis, *J. Am. Chem. Soc.*, 1996, **118**, 11884–11892.
- 6 B. Yang, A. R. Moehlig, C. E. Frieler and M. T. Rodgers, *J. Phys. Chem. A*, 2015, **119**, 1857–1868.
- 7 F. Zaccaria, G. Paragi and C. Fonseca Guerra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20895–20904.
- 8 M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770–780.
- 9 J. T. Davis, *Angew. Chemie - Int. Ed.*, 2004, **43**, 668–698.
- 10 A. Wong, R. Ida and G. Wu, *Biochem. Biophys. Res. Commun.*, 2005, **337**, 363–366.
- 11 S. Seifert, *Annu. Rev. Microbiol.*, 2018, **72**, 49–69.
- 12 M. Fathalla, C. M. Lawrence, N. Zhang, J. L. Sessler and J. Jayawickramarajah, *Chem. Soc. Rev.*, 2009, **38**, 1608–1620.
- 13 S. Sivakova and S. J. Rowan, *Chem. Soc. Rev.*, 2004, **34**, 9–21.

- 14 H. Yang and W. Xi, *Polymers (Basel)*., 2017, **9**, 666.
- 15 P. Amo-Ochoa and F. Zamora, *Coord. Chem. Rev.*, 2014, **276**, 34–58.
- 16 J. D. Watson and F. H. C. Crick, in *Cold Spring Harb Symp Quant Biol*, Cold Spring Harbor Laboratory Press, 1953, vol. 18, pp. 123–131.
- 17 J. D. WATSON and F. H. C. CRICK, *Nature*, 1953, **171**, 964–967.
- 18 J. H. Matthaei, O. W. Jones, R. G. Martin and M. W. Nirenberg, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 666–677.
- 19 J. S. Mattick and I. V Makunin, *Hum. Mol. Genet.*, 2006, **15**, R17–R29.
- 20 J. S. Mattick, *Nat. Rev. Genet.*, 2004, **5**, 316.
- 21 A. E. Pasquinelli, B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kuroda, B. Maller, D. C. Hayward, E. E. Ball, B. Degnan and P. Müller, *Nature*, 2000, **408**, 86.
- 22 C. Bancroft, T. Bowler, B. Bloom and C. T. Clelland, *Science*, 2001, **293**, 1763–1765.
- 23 T. J. Richmond and C. A. Davey, *Nature*, 2003, **423**, 145.
- 24 D. R. Whelan, T. J. Hiscox, J. I. Rood, K. R. Bambery, D. McNaughton and B. R. Wood, *J. R. Soc. Interface*, 2014, **11**, 20140454.
- 25 K. Hoogsteen, *Acta Crystallogr.*, 1963, **16**, 907–916.
- 26 M. H. Shamsi and H. B. Kraatz, *J. Inorg. Organomet. Polym. Mater.*, 2013, **23**, 4–23.
- 27 D. R. Williams, *Chem. Rev.*, 1972, **72**, 203–213.
- 28 F. R. Keene, J. A. Smith and J. G. Collins, *Coord. Chem. Rev.*, 2009, **253**, 2021–2035.

- 29 P. Belmont, J. F. Constant and M. Demeunynck, *Chem. Soc. Rev.*, 2001, **30**, 70–81.
- 30 B. Spingler and C. Da Pieve, *Dalt. Trans.*, 2005, 1637–1643.
- 31 B. Lippert, *J. Chem. Soc., Dalt. Trans.*, 1997, **21**, 3971–3976.
- 32 M. Pechlaner and R. K. O. Sigel, in *Characterization of Metal Ion-nucleic Acid Interactions in Solution*, 2012.
- 33 B. Lippert and D. Gupta, *Dalt. Trans.*, 2009, 4619–4634.
- 34 A. Ono, H. Torigoe, Y. Tanaka and I. Okamoto, *Chem. Soc. Rev.*, 2011, **40**, 5855–5866.
- 35 O. H. Gaebler, in *Enzymes: Units of Biological Structure and Function*, Academic Press New York, 1956.
- 36 A. S. Petrov, J. C. Bowman, S. C. Harvey and L. D. Williams, *RNA*, 2011, **17** 291–297.
- 37 R. K. O. Sigel and H. Sigel, *Acc. Chem. Res.*, 2009, **43**, 974–984.
- 38 B. Rosenberg, L. Van Camp and T. Krigas, *Nature*, 1965, **205**, 698–699.
- 39 A. H.-J. Wang, J. Reedijk, J. H. van Boom, S. S. G. E. van Boom and D. Yang, *Biochemistry*, 1995, **34**, 12912–12920.
- 40 S. Cai, X. Tian, L. Sun, H. Hu, S. Zheng, H. Jiang, L. Yu and S. Zeng, *Anal. Chem.*, 2015, **87**, 10542–10546.
- 41 P. Das, N. Das, S. Bhowmick, S. Kumari, A. Jana and K. Singh, *Polyhedron*, 2018, **157**, 267–275.
- 42 Y. Jung and S. J. Lippard, *Chem. Rev.*, 2007, **107**, 1387–1407.
- 43 J. P. Cerón-Carrasco, D. Jacquemin and E. Cauët, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12457–12464.

- 44 E. Wong and C. M. Giandomenico, *Chem. Rev.*, 1999, **99**, 2451–2466.
- 45 A. M. Pizarro and P. J. Sadler, in *Nucleic Acid-Metal Ion Interactions*, The Royal Society of Chemistry, 2009, pp. 350–416.
- 46 J. Anastassopoulou, *J. Mol. Struct.*, 2003, **651–653**, 19–26.
- 47 S. Lippard, in *Principles of Bioinorganic Chemistry*, University Science Books, 1994.
- 48 W. Kaim, B. Schwederski and A. Klein, John Wiley & Sons, 2013.
- 49 L. A. Loeb and A. R. Zakour, in *Nucleic Acid-Metal Ion Interactions*, Spiro TG, Ed, John Wiley & Sons, New York, 1980, pp. 115–144.
- 50 N. Russo, M. Toscano and A. Grand, *J. Phys. Chem. B*, 2001, **105**, 4735–4741.
- 51 E. O. Otokiti and R. D. Sheardy, *Biochemistry*, 1997, **36**, 11419–11427.
- 52 H. Urakawa, P. A. Noble, S. El Fantroussi, J. J. Kelly and D. A. Stahl, *Appl. Environ. Microbiol.*, 2002, **68**, 235 LP – 244.
- 53 S. N. Georgiades, N. H. A. Karim, K. Suntharalingam and Ramon Vilar, *Angew. Chemie. Int. Ed.*, 2010, **49**, 4020–4034.
- 54 D. Bhattacharyya, G. Mirihana Arachchilage and S. Basu, *Front. Chem.*, 2016, **4**, 38.
- 55 K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2006, **128**, 9963–9970.
- 56 S. Neidle and M. A. Read, *Biopolym. Orig. Res. Biomol.*, 2000, **56**, 195–208.
- 57 M. A. Jafri, S. A. Ansari, M. H. Alqahtani and J. W. Shay, *Genome Med.*, 2016, **8**, 69.
- 58 H. Han and L. H. Hurley, *Trends Pharmacol. Sci.*, 2000, **21**, 136–142.

- 59 G. M. Peters and J. T. Davis, *Chem. Soc. Rev.*, 2016, **45**, 3188–3206.
- 60 A. J. Wilson, *Soft Matter*, 2007, **3**, 409–425.
- 61 L. Brunsveld, B. J. B. Folmer, E. W. Meijer and R. P. Sijbesma, *Chem. Rev.*, 2001, **101**, 4071–4098.
- 62 L. Liu, Y. Pei, Y. Zhang, J. Wang and L. Chen, *Eur. Polym. J.*, 2018, **102**, 75–81.
- 63 J. A. R. Navarro and B. Lippert, *Coord. Chem. Rev.*, 1999, **185–186**, 653–667.
- 64 H. Liang, Z. Zhang, Q. Yuan and J. Liu, *Chem. Commun.*, 2015, **51**, 15196–15199.
- 65 A. Kumar and S. K. Gupta, *Green Chem.*, 2015, **17**, 2524–2537.
- 66 T. J. Pinnavaia, C. L. Marshall, C. M. Mettler, C. L. Fisk, H. T. Miles and E. D. Becker, *J. Am. Chem. Soc.*, 1978, **100**, 3625–3627.
- 67 D. J. Adams, K. Morris, L. Chen, L. C. Serpell, J. Bacsá and G. M. Day, *Soft Matter*, 2010, **6**, 4144–4156.
- 68 G. M. Peters, L. P. Skala, T. N. Plank, H. Oh, G. N. Manjunatha Reddy, A. Marsh, S. P. Brown, S. R. Raghavan and J. T. Davis, *J. Am. Chem. Soc.*, 2015, **137**, 5819–5827.
- 69 G. M. Peters, L. P. Skala, T. N. Plank, B. J. Hyman, G. N. Manjunatha Reddy, A. Marsh, S. P. Brown and J. T. Davis, *J. Am. Chem. Soc.*, 2014, **136**, 12596–12599.
- 70 J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Science*, 1989, **246**, 64–71.
- 71 S. Guillaumont, J. Tortajada, J. Y. Salpin and A. M. Lamsabhi, *Int. J. Mass Spectrom.*, 2005, **243**, 279–293.
- 72 A. A. Power, O. Y. Ali, M. B. Burt and T. D. Fridgen, *Int. J. Mass Spectrom.*, 2012, **332–332**, 233–240.

- 73 N. Russo, M. Toscano and A. Grand, *J. Mass Spectrom.*, 2003, **38**, 265–270.
- 74 J.-Y. Salpin, S. Guillaumont, J. Tortajada and M. Al Lamsabhi, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 359–369.
- 75 J.-L. Sun, H. Liu, H.-M. Wang, K.-L. Han and S. Yang, *Chem. Phys. Lett.*, 2004, **392**, 285–290.
- 76 M. Berdakin, V. Steinmetz, P. Maitre and G. A. Pino, *J. Phys. Chem. A*, 2014, **118**, 3804–3809.
- 77 A. Sen and C. E. H. Dessent, *J. Phys. Chem. Lett.*, 2014, **5**, 3281–3285.
- 78 C. A. Schalley, *Mass Spectrom. Rev.*, 2001, **20**, 253–309.
- 79 A. K. Chandra, M. T. Nguyen, T. Uchimaru and T. Zeegers-Huyskens, *J. Phys. Chem. A*, 1999, **103**, 8853–8860.
- 80 N. Russo, M. Toscano, A. Grand and F. Jolibois, *J. Comput. Chem.*, 1998, **19**, 989–1000.
- 81 M. Elstner, T. Frauenheim, E. Kaxiras, G. Seifert and S. Suhai, *Phys. Status Solidi*, 2000, **217**, 357–376.
- 82 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 83 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.
- 84 S. F. Boys and F. de Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.
- 85 J. Y. Salpin, S. Guillaumont, J. Tortajada, L. MacAleese, J. Lemaire and P. Maitre, *ChemPhysChem*, 2007, **8**, 2235–2244.
- 86 S. Y. Han and H. Bin Oh, *Chem. Phys. Lett.*, 2006, **432**, 269–274.
- 87 J. Oomens, A. R. Moehlig and T. H. Morton, *J. Phys. Chem. Lett.*, 2010, **1**, 2891–2897.

- 88 J. Jun and S. Y. Han, *Theor. Chem. Acc.*, 2017, **136**, 1–10.
- 89 A. F. Cruz-Ortiz, M. Rossa, F. Berthias, M. Berdakin, P. Maitre and G. A. Pino, *J. Phys. Chem. Lett.*, 2017, **8**, 5501–5506.
- 90 Y. Seong, S. Y. Han, S. C. Jo and H. Bin Oh, *Mass Spectrom. Lett.*, 2011, **2**, 73–75.
- 91 K. Rajabi, K. Theel, E. A. L. Gillis, G. Beran and T. D. Fridgen, *J. Phys. Chem. A*, 2009, **113**, 8099–8107.
- 92 B. Yang, R. R. Wu, G. Berden, J. Oomens and M. T. Rodgers, *J. Phys. Chem. B*, 2013, **117**, 14191–14201.
- 93 D. A. Megger, P. M. Lax, J. Paauwe, C. Fonseca Guerra and B. Lippert, *J. Biol. Inorg. Chem.*, 2018, **23**, 41–49.
- 94 M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger and D. Christ, *Nat. Chem.*, 2018, **10**, 631–637.
- 95 K. N. Jarzemska, A. M. Goral, R. Gajda and P. M. Dominiak, *Cryst. Growth Des.*, 2013, **13**, 239–254.
- 96 S. Y. Han, S. H. Lee, J. Chung and H. Bin Oh, *J. Chem. Phys.*, 2007, **127**, 245102.
- 97 J. J. Park and S. Y. Han, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 846–854.
- 98 W. Guschlbauer, J.-F. Chantot and D. Thiele, *J. Biomol. Struct. Dyn.*, 1990, **8**, 491–511.
- 99 J. R. Williamson, M. K. Raghuraman and T. R. Cech, *Cell*, 1989, **59**, 871–880.
- 100 E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco Jr and E. H. Blackburn, *Cell*, 1987, **51**, 899–908.
- 101 T. Aggerholm, S. C. Nanita, K. J. Koch and R. G. Cooks, *J. Mass Spectrom.*, 2003,

- 38**, 87–97.
- 102 M. Gellert, M. N. Lipsett and D. R. Davies, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 2013–2018.
- 103 C. Frascchetti, M. Montagna, L. Guarcini, L. Guidoni and A. Filippi, *Chem. Commun.*, 2014, **50**, 14767–14770.
- 104 J. Gu and J. Leszczynski, *J. Phys. Chem. A*, 2002, **106**, 529–532.
- 105 M. Meyer, M. Brandl and J. Sühnel, *J. Phys. Chem. A*, 2001, **105**, 8223–8225.
- 106 M. Meyer, T. Steinke, M. Brandl, J. Sühnel and M. E. T. Al, *J. Comput. Chem.*, 2000, **22**, 109–124.
- 107 M. Meyer, A. Hocquet and J. Sühnel, *J. Comput. Chem.*, 2005, **26**, 352–364.
- 108 M. Azargun and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25778–25785.
- 109 M. Azargun, Y. Jami-Alahmadi and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2017, **19**, 1281–1287.
- 110 J. Gao, G. Berden, M. T. Rodgers and J. Oomens, *Phys. Chem. Chem. Phys.*, 2016, **18**, 7269–7277.
- 111 K. J. Koch, T. Aggerholm, S. C. Nanita and R. G. Cooks, *J. Mass Spectrom.*, 2002, **37**, 676–686.
- 112 M. T. Rodgers and P. B. Armentrout, *J. Am. Chem. Soc.*, 2000, **122**, 8548–8558.
- 113 T. Fujii, *Mass Spectrom. Rev.*, 2000, **19**, 111–138.
- 114 N. Russo, M. Toscano and A. Grand, *J. Am. Chem. Soc.*, 2001, **123**, 10272–10279.
- 115 B. Qiu, J. Liu, Z. Qin, G. Wang and H. Luo, *Chem. Commun.*, 2009, 2863–2865.
- 116 E. A. L. Gillis and T. D. Fridgen, *Int. J. Mass Spectrom.*, 2010, **297**, 2–8.
- 117 P. Amo-Ochoa, P. J. Sanz Miguel, P. Lax, I. Alonso, M. Roitzsch, F. Zamora and

- B. Lippert, *Angew. Chemie - Int. Ed.*, 2005, **44**, 5670–5674.
- 118 M. Kabeláč and P. Hobza, *J. Phys. Chem. B*, 2006, **110**, 14515–14523.
- 119 K. Rajabi, E. A. L. Gillis and T. D. Fridgen, *J. Phys. Chem. A*, 2010, **114**, 3449–3456.
- 120 B. Yang, R. R. Wu, N. C. Polfer, G. Berden, J. Oomens and M. T. Rodgers, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 1523–1533.
- 121 E. A. L. Gillis, K. Rajabi and T. D. Fridgen, *J. Phys. Chem. A*, 2009, **113**, 824–832.
- 122 Y. W. Nei, T. E. Akinyemi, C. M. Kaczan, J. D. Steill, G. Berden, J. Oomens and M. T. Rodgers, *Int. J. Mass Spectrom.*, 2011, **308**, 191–202.
- 123 B. Power, V. Haldys, J. Y. Salpin and T. D. Fridgen, *Int. J. Mass Spectrom.*, 2015, **378**, 328–335.
- 124 C. Trujillo, A. M. Lamsabhi, O. Mó, M. Yáñez and J. Y. Salpin, *Int. J. Mass Spectrom.*, 2011, **306**, 27–36.
- 125 C. Trujillo, A. M. Lamsabhi, O. Mó, M. Yáñez and J. Y. Salpin, *Org. Biomol. Chem.*, 2008, **6**, 3695–3702.
- 126 E. a. L. Gillis, M. Demireva, K. Nanda, G. Beran, E. R. Williams and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 3304.

## Chapter 2 Experimental and Computational Methods

### 2.1 General Introduction to Chapter 2

Fourier transform (FT) ion cyclotron resonance (ICR) and quadrupole ion trap (QIT) mass spectrometers are methods developed to trap the ions in the gas phase for a relatively long period of time, and analyze them based on their mass to charge ratios ( $m/z$ ). Prior to mass detection, trapped ions can be internally excited and dissociated. In this chapter excitation methods such as collision-induced dissociation (CID) and infrared multiple photon dissociation (IRMPD) are described in detail. IRMPD conducted with a tunable infrared laser can be used to collect vibrational spectra of complexes. IRMPD spectroscopy will also be described in this chapter.

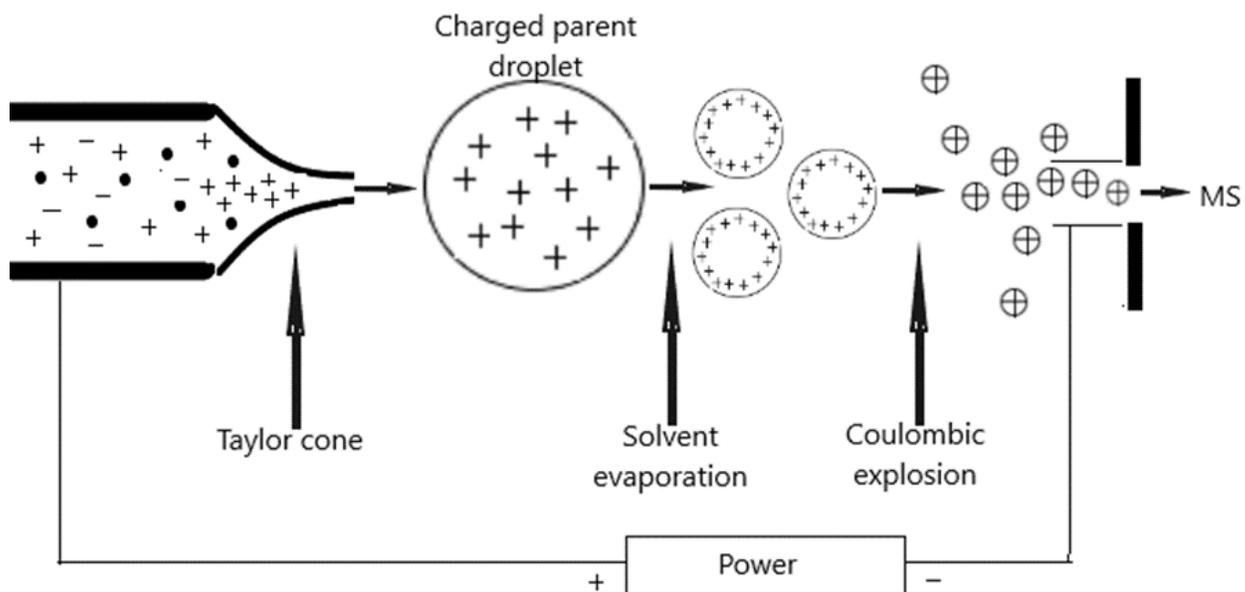
Besides experimental techniques, density functional theory (DFT) is the primary theoretical method used for performing structural optimizations and thermodynamic calculations of the gaseous complexes in this thesis, including various isomers of complexes to determine the most preferred isomers as well as transition states to gain insight on preferred reaction pathways. The computed IR intensities of each structure can be used to compare with vibrational IRMPD spectra collected. Various approximation functionals and basis sets have been used for thermostability comparisons.

### 2.2 Experimental Methods

The mass spectrometric methods used in this thesis are described below, beginning with the electrospray ionization used for volatilizing the ionic complexes of interest. This

is followed by the descriptions of instruments used for trapping and determining the  $m/z$  of gaseous ions. Then, general descriptions of various structure determination techniques are also provided.

**2.2.1 Electrospray Ionization.** Electrospray ionization (ESI) is a type of ion source<sup>1</sup> used to generate gaseous ions from the solution phase for their introduction into mass spectrometers. The ions in solution assemble at the tip of capillary to which is applied a high electric potential at the counter electrode; a potential difference of about 3-5 kV is typically applied between the capillary tip and the entrance to the mass spectrometer. The assembled charge forms a Taylor cone,<sup>2</sup> from which charged droplets containing analyte ions are formed. In the spray chamber, a heated drying gas (*ie.* N<sub>2</sub>) flowing counter to the ion's motion is applied to help remove and evaporate neutral molecules and solvent from droplets. As the droplet undergoes the process of evaporation, it has an ever-decreasing droplet size, which increases the charge density on droplet's surface. When the repulsive force of charges on the surface of the droplet is sufficiently high (surpass the Rayleigh limit), droplet fission, also called Coulombic explosion, generates smaller droplets. This process continues until individual singly or multiply charged ions are present in the gas phase. These single gas-phase ions then enter into the mass analyzer. The ESI process is depicted in **Figure 2.1**. Both positive and negative ions can be produced by the ESI method depending on the electronic fields supplied between the sprayer and counter electrode.



**Figure 2.1** Schematic of ESI method for the case of producing positive ions.

ESI is considered a very soft ionization technique because it can introduce biomolecules and complexes containing non-covalent interactions into the gas phase without breaking it apart. ESI is widely used as an ion source coupled with FTICR or ion-trap mass spectrometers to produce ionic complexes for the bio-molecules that are non-volatile and thermally labile. Another advantage of ESI method is its ability to produce multiply charged ions. The molecular weight of analyzed precursor ions extends from kDa to MDa, which can be applied to study the accurate structural information of large molecules. The structural completeness of large biomolecules, such as proteins, peptides and nucleotides, can be kept during the ionization process of ESI, although the present work does not deal with such large species.

## 2.2.2 Overview of Mass Analyzers and Fragmentation Methods

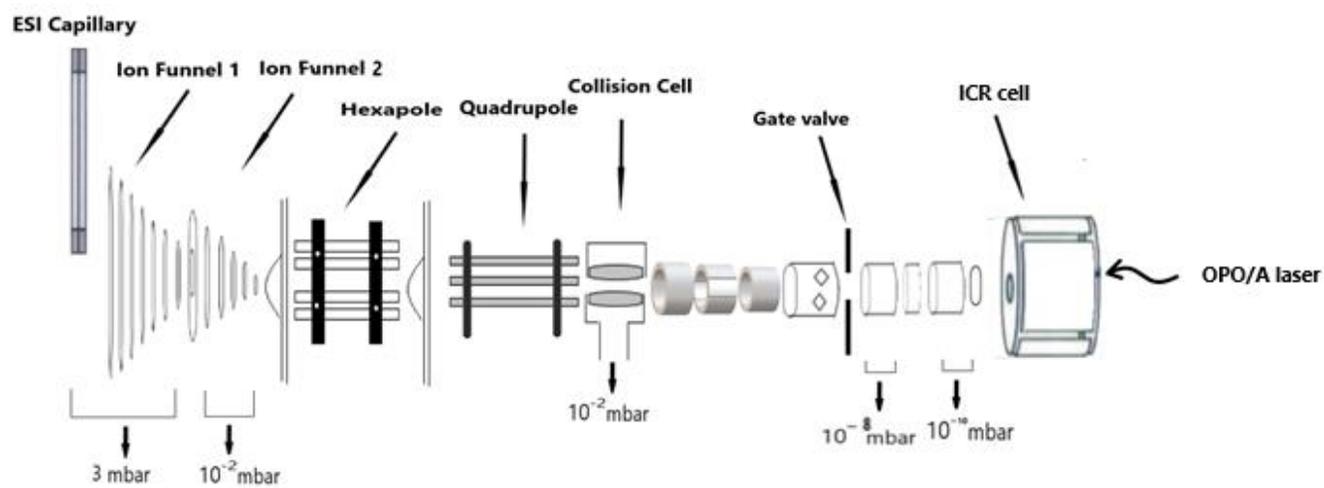
Two main fragmentation methods with two types of ion trapping mass analyzers were used for performing experiments in this thesis. In both FT-ICR and QIT mass spectrometers, ions can be stored for a long period of time. A combination of static magnetic and electric fields is used to trap ions in the FT-ICR, whereas the QIT employs oscillating electric fields to trap ions. These analyzers, as well as fragmentation techniques, are described next.

### 2.2.3 Fourier-Transform Ion Cyclotron Resonance Mass Spectrometer (FT-ICR-MS)

Ion cyclotron resonance was first introduced by Sommer *et al.*<sup>3</sup> to be applied in a mass spectrometer in 1951 to determine the charge-to-mass ratio of the proton. In the early 1970s, *McIver* and other groups introduced the trapped ion cell, also called ICR cell, to trap ions.<sup>4,5</sup> It allows the ions to be trapped, manipulated and detected within the same space, differentiated from the conventional drift cell. Since the mid-1970s, Fourier Transform was applied to ICR by *MB Comisarow* and *AG Marshall* at the University of British Columbia.<sup>6</sup> The advantage of using the Fourier transform is to speed up the data analysis process. The application of Fourier transform to raw data, collected as a time-domain signal, can analyze the cyclotron frequencies of all ions in the ICR cell at one time, which dramatically increases the speed of data analysis by a factor of  $\sim 10^5$  faster than scanning the spectrum to obtain a mass spectrum as a function of frequency. By adapting the Fourier transform to ion cyclotron resonance, FT-ICR became a matured, widely used, high-resolution analytical instrument applied to study gaseous molecules.<sup>7-12</sup> A schematic of the ESI-FTICR-MS used for present studies presents in **Figure 2.2 (a)**, which contains ion source

(ESI), ion funnels (to focus and transmit a beam of ions), hexapole and quadrupole ion trap (for ion accumulation), collision cell (for ion fragmentation), and ICR cell. A photo of optical parametric oscillator/amplifier (OPO/A) laser coupled to Bruker Apex-Qe 7T FT-ICR-MS in the Laboratory for the Study of Energetics, Structures, and Reactions of Gaseous Ions at Memorial University is presented in **Figure 2.2 (b)**.

**2.2.3.1 Trapping of FT-ICR-MS.** Firstly, positive (or negative) ions are volatilized by ESI and accelerated by an electric field and then focused into ICR cell. The presence of a 7 T magnetic field traps the ions in two dimensions (x-y plane), perpendicular to the magnetic field (z-axis), by causing ions to travel in a circular motion around the z-axis. Ergo, the ions are not able to escape the magnetic field in two dimensions. In order to achieve trapping in the third dimension, a small electric potential, ~1-2 V, is applied to two trapping plates, perpendicular to the magnetic field.<sup>13</sup> The mass dependence of trapped-ion motion in the ICR cell is described below.



(a)



(b)

**Figure 2.2** (a) Schematic description of the structure of FT-ICR-MS with ESI ion source and (b) Photo of the ESI-FT-ICR with OPO/A laser at Memorial University

When ions of mass,  $m$ , moving with a constant velocity,  $\vec{v}$ , and charge,  $q$ , enter into a uniform magnetic field,  $\vec{B}$ , they are influenced by the Lorentz force  $\vec{F}$ ,

$$\vec{F} = q\vec{v} \times \vec{B} \quad (2.1)$$

As ions enter the ICR cell along the magnetic field, ions are "kicked" by an electric field, thereby given a velocity perpendicular to the magnetic field. The path of ion is a circle of radius  $r$  in the ICR cell around the z-axis. The Lorentz force is balanced by the centrifugal force of ion's circular motion,

$$qvB = \frac{mv^2}{r} \quad (2.2)$$

In the static magnetic field, the relation between cyclotron frequency  $\omega_c$ , linear velocity  $v$  and the radius of the charged moving particle is defined by the equation

$$v = \omega_c r \quad (2.3)$$

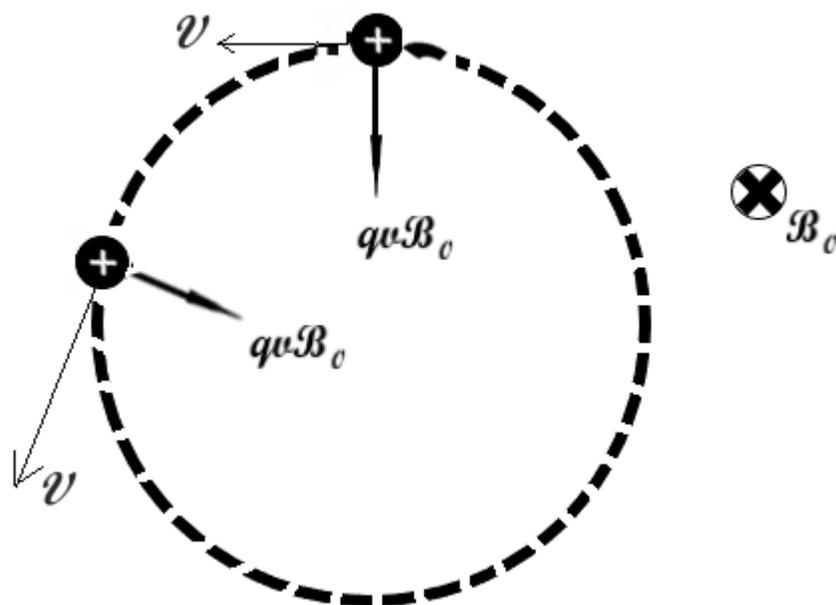
The relation between  $m/q$  and the magnetic field strength and cyclotron frequency are acquired by rearranging equation 2.2 and combining with equation 2.3,

$$\frac{m}{q} = \frac{B}{\omega_c} \quad (2.4)$$

Or, since  $q = ze$ , where  $z$  is the number of charge and  $e$  is the charge

$$\frac{m}{z} = \frac{Be}{\omega_c} \quad (2.5)$$

According to equation 2.5, for a fixed magnetic field, the  $m/z$  ratio of the ion can be determined by measuring the cyclotron frequencies of ions orbiting the magnetic field. The electric potential along the z-axis applied to trapped ions causes a harmonic oscillation at a very low frequency and is not considered here. The cyclotron motion of a positive ion is shown in **Figure 2.3**.

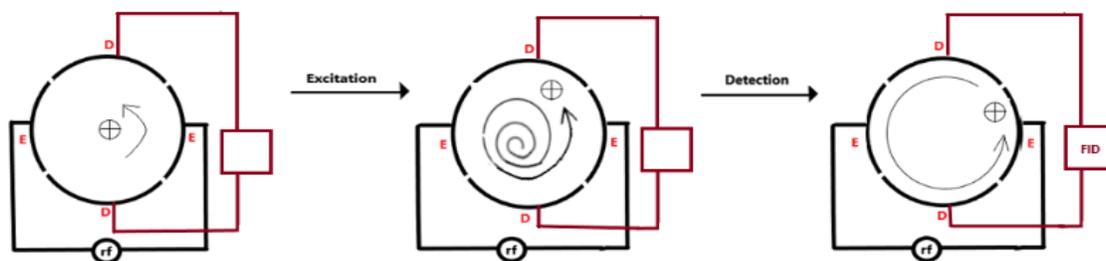


**Figure 2.3** Schematic of the circular motion trajectory of a positive ion in the constant magnetic field,  $B_0$ .

After trapping in the ICR cell, the ions can be excited by applying resonant radio frequency voltage to excitation plates and then detected by a pair of detection plates. Before, the ions can be trapped in the ICR cell for several hours; while after excitation, trapping can still last for up to several minutes.

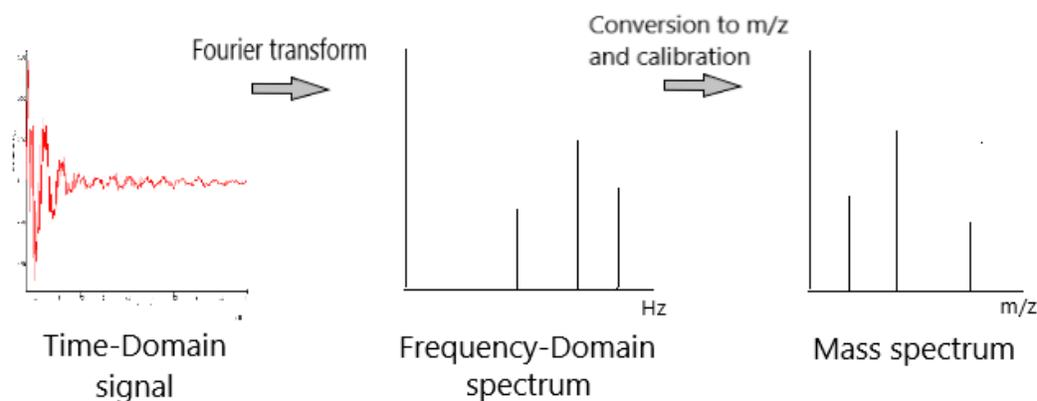
**2.2.3.2 Excitation and Detection of FT-ICR-MS.** There are three reasons to excite the ions in the ICR cell.<sup>14</sup> The first, discussed here, is to increase the cyclotron radius making ions detectable in the ICR cell. A pair of excitation (E) plates and a pair of detection (D) plates are orthogonal to each other and composed of the ICR cell, as seen in **Figure 2.4**. First, a resonant radio frequency (RF) excitation pulse is alternately applied to the excitation plates to stimulate ions to have higher kinetic energy; since the frequency

remains the same, this higher kinetic energy means ions have larger orbits after excitation. The frequencies of pulses are swept over the full range of different ions' frequencies (typically corresponding to  $m/z$  ~20 to ~1000) in order to excite all ions trapped in the ICR cell. Ions having the same  $m/z$  with increased radii are called an ion packet, and each ion packet has its characteristic cyclotron frequency according to equation 2.4 and 2.5. The increased radii of ion packets' orbits allow them to be in proximity to a detection plate and induce an image charge of electrons to one detection plate. As an ion packet leaves the vicinity of the first plate and moves to that of the second plate, the electrons composing the image charge travels through a detector to the second electrode. The signal collected by the detector is called an image current, which is a sinusoidal oscillation and measured as a function of time as the ions continue to orbit the magnetic field in the ICR cell. The frequency of the induced image current is the same as that of an ion packet's cyclotron motion. Motions of an ion packet during the excitation and detection processes are shown in **Figure 2.4**.



**Figure 2.4** Schematic of the cross-section in the ICR cell while the ion packet is experiencing the excitation to increase the radius and detection. The Ds are the detection plates, and the Es are the excitation plates.

All ion packets, each composed of different mass ions and thus different frequencies, are detected simultaneously. The magnitude of image current is proportional to the number of ions in each ion packet, and the signal of oscillated image current is called free induction decay (FID), a time-domain signal. The FID signal is amplified and then translated by Fourier transform into the frequency-domain spectrum of different ion packets. The useful mass spectrum of various ion packets is then obtained by the conversion and calibration of the frequency-domain spectrum. **Figure 2.5** depicts the process of data processing.



**Figure 2.5** Data analysis process to obtain the mass spectrum from the original FID signal of the detector.

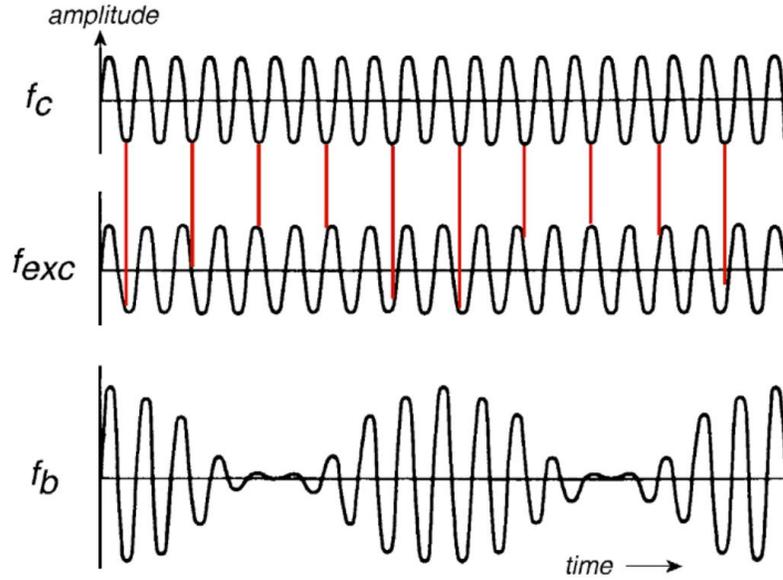
The second use of the excitation process is to remove some (or all) trapped ions out of the ICR cell. The ions will be accelerated to a cyclotron radius larger than the radius for detection and then be discharged on the ICR cell walls. This process is used for isolating ions. By exciting and removing all ion packets but one, the desired ions are preserved and isolated in the ICR cell. Isolation is a crucial process prior to all fragmentation techniques for the characterization of ions as it guarantees the origins of fragment ions.

The third reason to excite an ion is to increase the kinetic energy of ions in the ICR cell to collide them into a neutral "target" atom or molecule to convert their translational energy into internal energy. The collision-induced dissociation that happened in the ICR cell is called sustained off-resonance irradiation collision-induced dissociation (SORI-CID) and is described below. It is a process of gradually depositing a large amount of energy to ions in a long activation time, which enables ions to go through its dissociation pathway.

In SORI-CID, ions are irradiated slightly off-resonance; and the difference between ion's natural cyclotron frequency and RF excitation frequency ( $\Delta\nu$ ) can be around 500-1500 Hz. The RF pulse is applied to the trapped ions for a period of time (0.1-1s)<sup>15</sup> (250 ms for work in this thesis), hence it is the term 'sustained', and increases the kinetic energy of the trapped ions. Thus, by applying the off-resonance potential to the excitation plates, the orbital radius of ion cyclotron motion expands and contracts rather than simply expanding during resonant irradiation pulses described previously. The interference between cyclotron frequency ( $f_c$ ) and excitation frequency ( $f_{exc}$ ) is termed as a beat frequency ( $f_b$ ),

$$f_b = f_{exc} - f_c \quad (2.6)$$

and the amplitude of  $f_b$  is proportional to the kinetic energy obtained by ions. The relation between  $f_c$ ,  $f_{exc}$  and  $f_b$  are shown in **Figure 2.6**.



**Figure 2.6** The frequency of ion cyclotron motion  $f_c$ , and that of excitation pulse  $f_{exc}$ . The interference of two frequencies is  $f_b$ . All three are the time-domain signals. The higher amplitude of  $f_b$  indicates larger kinetic energy of the ion. Reprinted from reference 15 with permission.

The kinetic energy is steadily changed, and the average kinetic energy of ions is expected to be about 2/3 the maximum lab frame kinetic energy which can be calculated by the equation,<sup>16</sup>

$$E_{lab} = \left( \frac{\beta^2 q^2 V_{p-p}}{32\pi^2 m d^2 \Delta v^2} \right) \quad (2.7)$$

where  $m$  is the mass of the ion,  $q$  is the charge of the ion,  $V_{p-p}$  is the peak to peak excitation voltage,  $\beta$  is a geometrical factor of the ICR cell (0.92 in the present case),  $d$  is the diameter of the ICR cell (6 cm for the Bruker Apex-Qe 7T), and  $\Delta v$  is the difference between the

ion's natural cyclotron frequency and RF excitation frequency (~500 Hz). To convert the lab frame energy to the center of mass frame energy (kinetic energy) transferred by each collision, there is a need to multiply  $E_{\text{lab}}$  by  $m_{\text{Ar}} / (m_{\text{Ar}} + m_{\text{complex}})$  where Ar is the collision gas.

During excitation, accelerated ions (<10eV) undergo multiple (tens to hundreds) collisions with the target gas (typically Ar) that is pulsed into the ICR cell at a pressure of about  $10^{-5}$  to  $10^{-6}$  mbar.<sup>16</sup> It is a relatively high pressure compared to the regular  $10^{-8}$  to  $10^{-10}$  Torr in the ICR cell. The maximum energy transferred to the ion in a particular collision ranges from a few tens to a few hundreds  $\text{kJ mol}^{-1}$ . With each collision, the ion's internal energy is increased, followed by the rapid process of intramolecular vibrational energy redistribution (IVR), which randomizes the energy among other internal coordinates. Once the total internal energy accumulated by multiple collisions surpasses the lowest energy dissociation threshold, the dissociation of ion occurs, and its observation depends on molecular kinetics.<sup>12</sup> Often, products of different competing dissociation reactions or consecutive fragmentations can be observed, which can sometimes be distinguished by multiple stages of isolation or tandem mass spectrometry.

SORI-CID a soft technique and can dissociate the complex typically through its lowest dissociation pathways, which promotes the widespread use of this method to couple with the ESI-FTICR-MS to study the weak intermolecular non-covalent interactions in large biomolecules.<sup>17-24</sup>

There is a wide range of applications of FTICR-MS for the study of biomolecules, like nucleobases,<sup>8,25-29</sup> nucleotides<sup>26,30-33</sup> and DNA/RNA-related molecules<sup>28,31-36</sup> due to some advantages of this technique compared with classical mass spectrometers. First, the

superconducting electromagnet of ICR cell, made from coils of superconducting wire, presents better stability than the created for other mass spectrometers, such as quadrupole ion trap and time of flight mass spectrometer (TOF-MS), which contributes to the higher resolving power (~10-100X higher). The features of high accuracy of  $m/z$  detection, high resolution and high sensitivity of FTICR-MS also benefit from an extended trapping time (seconds) without colliding with unrelated background molecules, thanks to the very low pressure in ICR cell.<sup>9</sup> Thus, FT-ICR-MS can deal with the complicated mixtures of ions because of its high resolution expressed by the sharp peaks and its potential to perform the tandem MS experiments. Secondly, besides the advantage of resolution, the FTICR-MS technique is different from some other mass analyzers, *ie.* TOF-MS, which typically eject ions to hit a detector. In this case, ions are detected by the detection plates and can be trapped in the ICR cell for a much longer time.

There are some other fragmentation techniques also executed in the ICR cell besides SORI-CID, such as infrared multiphoton dissociation (IRMPD) spectroscopy and blackbody infrared radiative dissociation (BIRD). The IRMPD fragmentation method is introduced in 2.2.5, while the BIRD is not described in this chapter as it was not used in this work.

#### **2.2.4 Quadrupole Ion-Trap (QIT) Mass Spectrometer**

Besides the FTICR-MS, the other mass spectrometer used in this research is the QIT, also called the 3D quadrupole or Paul-type ion trap. Both free-electron lasers at the FELIX (free electron laser for infrared experiments) facility in Amsterdam and CLIO (Centre Laser Infrarouge d'Orsay) facility in Paris are equipped with 3D QIT instruments.

Bruker amazon speed ETD Paul-type ion trap is coupled to FELIX, while the QIT mass spectrometer at CLIO is a Bruker Esquire 3000+ 3D ion trap.

**2.2.4.1 Trapping of Ions in QIT.** Earnshaw's theorem states that charged particles cannot be trapped in the static electric fields in three dimensions (3D).<sup>37</sup> However, it is possible to limit the motions of ions to a specific range by trapping them in oscillating electric fields. This is the fundamental theory of the 3D quadrupole ion trap. The use of quadrupole electric fields to manipulate ions was initially investigated by *Wolfgang Paul* et al. in the early 1950s;<sup>38</sup> the research resulted in the joint Nobel Prize in Physics in 1989 with *Hans Georg Dehmelt* who won it for his work on the Penning trap which is closely related to the FT-ICR mass spectrometer. The QIT was commercialized in the mid-1980s.<sup>39</sup>

The QIT is composed of two hyperbolic metal end-cap electrodes with their foci facing each other, and a ring electrode (r) placed symmetrically between them, as shown in **Figure 2.7**. This geometry establishes a potential field when an oscillating AC potential is applied between the hyperbolic electrodes. Thus, ions are stabilized in the axial direction but unstable in the radial direction (x-y plane). Then, the applied AC potential of constant frequency (RF) but variable amplitude to the ring electrode can create a 3D potential well to confine ions in all three dimensions.

However, in this quadrupole field, the overall force that ion felt is not averaged to zero but towards the center of the oscillating field. Thus, ions will experience a focusing power back to the center of the device as the ion's oscillating trajectory deviates far from the quadrupole field. The cell is filled with inert gas (e.g. He) at ~1 mTorr to help dampen the kinetic energies of ions and shrink their trajectories to the center of the cell. A sample trajectory of a single ion in the QIT is in **Figure 2.8**.

This quadrupole field can be described by solutions to the second-order linear differential equation expressed initially by *Mathieu* in interpreting the mathematics of vibrating skins (*ie.* drums).<sup>40</sup> The Mathieu equation is:

$$\frac{d^2u}{d\xi^2} + (a_u - 2q_u \cos 2\xi)u = 0 \quad (2.8)$$

where  $u$  is a function of three-dimensional axes  $x, y$  and  $z$ ,  $\xi$  is a dimensionless parameter which is equal to  $\Omega t/2$ , and  $\Omega$  is the radial frequency of RF potential applied to the ring electrode while  $t$  is time,  $a_u$  and  $q_u$  are two trapping parameters without considering dimensional information.

Solutions to the Mathieu equation provide stability and instability regions. Without getting into the mathematics, the solutions to the Mathieu equation provide trapping parameters,  $a_z$  and  $q_z$ , as follows,

$$a_z = -16 \frac{eU_0}{m(r_0^2 + 2z_0^2)\omega_0^2} \quad (2.9)$$

$$q_z = -8 \frac{eV_0}{m(r_0^2 + 2z_0^2)\omega_0^2} \quad (2.10)$$

as well as  $a_r$  and  $q_r$ ,

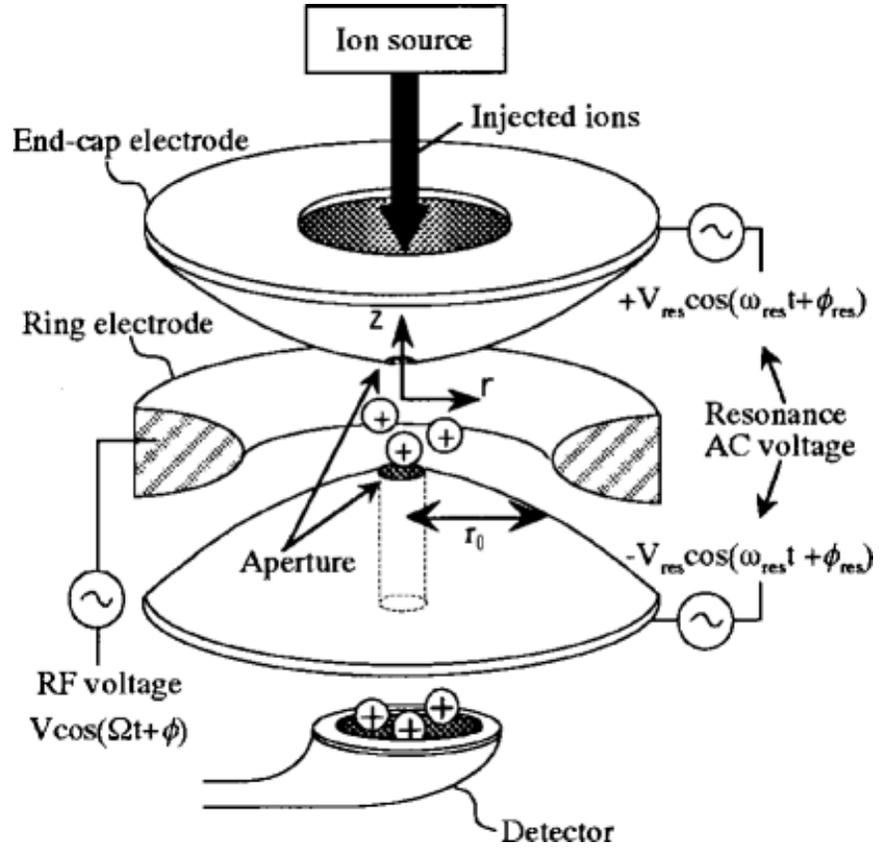
$$a_r = 8 \frac{eU_0}{m(r_0^2 + 2z_0^2)\Omega^2} \quad (2.11)$$

$$q_r = -4 \frac{eV_0}{m(r_0^2 + 2z_0^2)\Omega^2} \quad (2.12)$$

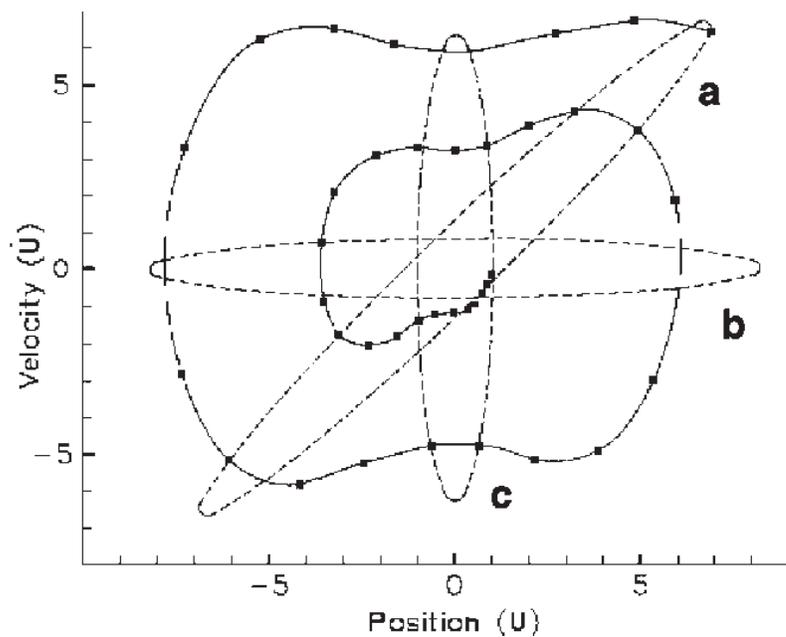
$$\Phi_0 = U_0 + V_0 \cos \Omega t \quad (2.13)$$

where  $\Phi_0$  is the applied AC potential to the ring electrode,  $V_0 \cos \Omega t$  is the RF potential, and  $U_0$  is the static DC potential,  $r_0$  and  $z_0$  represents the distance between the central axis of cylindrical symmetry to the ring electrode and either end-cap electrode, respectively,  $\Omega$  is

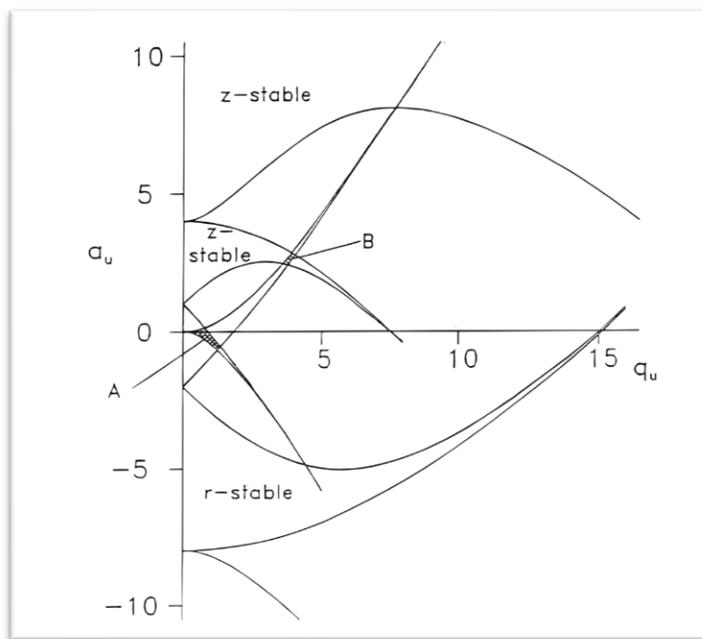
the radial frequency of the RF potential applied to the ring electrode,  $a_z$ ,  $q_z$  and  $a_r$ ,  $q_r$  are different dimensional trapping parameters.



**Figure 2.7** Depicts of the three-dimensional quadrupole field of QIT. Figure reproduced from reference 41 with permission.



**Figure 2.8** The sample trajectory of an ion in the QIT is reprinted from reference 42 with permission.



**Figure 2.9** The overlapped regions of r-stable and z-stable, A and B, are the regions of 3D ion stability. Reproduced from reference 43 with permission.

The regions where the stored ions are stable within the QIT are those simultaneous overlapped regions (A and B) in  $a_r/q_r$  and  $a_z/q_z$  spaces shown in **Figure 2.9**. The labelled areas, 'r-stable', provide the regions where the stored ions are stable in r-direction, and the other parts labelled 'z-stable' means ions are stabilized in z-direction in these fields. Commercial instruments use the stability region close to the origin of this  $a_u/q_u$  plot.

At FELIX, after 0.05 - 50 ms of ion accumulation, ions enter the trap field through the first hyperbolic electrode. The RF field is of 400-800 V<sub>pp</sub> at 781 kHz. In the trapping field, the cell is filled with inert background gas, He, tuned from at about 10<sup>-3</sup> mbar pressure to cool ions to the center of the trap.<sup>44</sup> To enable optical access to ions in the cell, the ring electrode is modified to have two 3 mm holes centred in its top and bottom to give access to IRMPD (described in 2.2.5).

**2.2.4.2 Excitation and Isolation of QIT.** As ions are confined in both axial and radial axis, resonance excitation can affect ions by changing their secular frequencies at either or both axes. Secular frequency is the oscillating frequency of ions, which is lower than the fundamental frequency applied to the trap. Resonance excitation with the axial secular frequencies of ions is a powerful technique used to executive the excitation and isolation in QIT at FELIX.

Before performing resonance excitation to ions, the 'ion cooling' process is used to dampen the kinetic energies of ions to ~ 0.1 eV by colliding with the background buffer gas molecules (He), which makes the ions' trajectories skewing from the center of the cell less than 1 mm. Then, oscillating potentials same to ion's secular frequencies is applied to the end-cap electrodes with an amplitude of a few hundreds mVs, exciting the translational motion of ions along the z-axis. The applied oscillating potentials allow ions

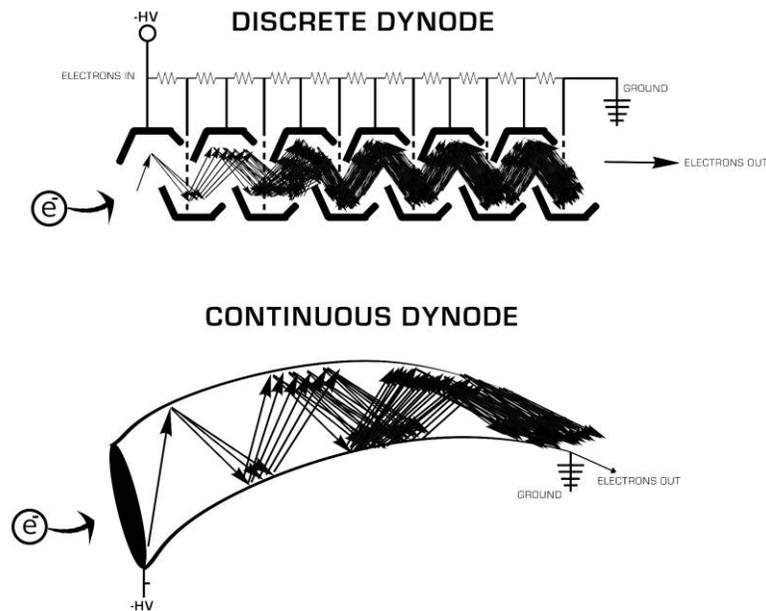
to move further from the center of the trapping field and, thus, to experience more acceleration from the trapping field until the kinetic energies of ions reach tens of eVs.

Ion excitation can be used for multiple processes. First, to isolate a particular ion species, different resonance frequencies are applied to hyperbolic electrodes to excite the unwanted ions in a wide range of  $m/z$ , which can amplify the magnitude of motion and remove these ions out of the trap field. There is a single or a small range of  $m/z$  of ions isolated in the cell. Also, by ions excitation, the kinetic energy of ions can be increased. The collisions between the excited ions and neutral gas, typically He, can be used to affect collision-induced dissociation (CID). Details about the CID in QIT are not described as it is not reported in this thesis.

**2.2.4.3 Detection of Ions in QIT.** Different to the detection plates existed in the ICR cell, ions are detected by an external detector outside of the QIT. There are two ways for ion detection, either a Faraday Cup (FC) or an Electron Multiplier (EM). As EM is a type of detector used at FELIX, it is introduced below.

Ions first hit the conversion dynode to which is applied a high voltage (5-20 kV) opposite to the charge of ions after sequentially ejecting them out of ion trap from the second end-cap electrode. After hitting the conversion dynode, the signal intensity of ions is increased because the electrons are released from the conversion dynode and join the ions' stream. The EM is a vacuum-tube structure with high negative voltage composed of either several discrete dynodes or a cone-shaped dynode, which is shown in **Figure 2.10**. As electrons or positive ions are accelerated to strike the surface of dynode with high velocity, there are more electrons ejected from the surface of the dynode. This process is called secondary electrons emission. A cascade of electrons is accelerated in the direction

of the end of EM and hits the surface of the dynode repeatedly to cause more electrons release. Thus, electrons are multiplied by several strikes and finally emitted from the end of EM. For each entering ion, there is a minimum of  $10^5$  and a maximum of  $10^8$  electrons emitted. The ion signal is produced by detecting the electron current which is differentiated by  $m/z$  of ions.



**Figure 2.10** The figure on the top illustrates the process of electron accumulation in the discrete EM and the bottom one is that in the continuous EM. Reprinted from reference 45 with permission.

The QIT is an extraordinary mass spectrometer with high efficiency and high sensitivity. Even though the resolution of QIT is only about 1 Da at regular (high) scan speed, which is relatively poor, it is not as expensive as the FT-ICR-MS. Even though the

principle of the QIT is different from FTICR, it also can trap ions for a long time, which allows various ion-molecule reactions to be studied for the isolated ions, such as IRMPD and CID.

### **2.2.5 IRMPD Fragmentation**

Fragmentation techniques are used to perform gas-phase ion dissociation to determine the intrinsic molecular structures and thermochemical characteristics of ionic complexes. Once the internal energy of the precursor ion surpasses the dissociation energy threshold, it can fragment. Fragments of IRMPD are typically similar to that of SORI-CID, but the infrared photons are the source of energy instead of energetic collisions.<sup>12,44,46</sup>

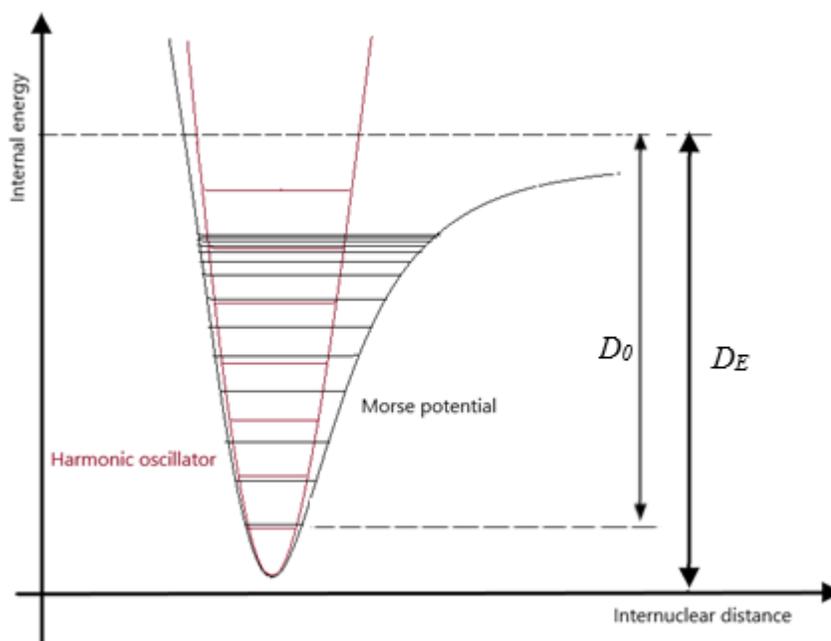
IRMPD is a technique to fragment ions in the gas phase and is coupled to different mass spectrometers, such as the FTICR and QIT mentioned above. Infrared multiphoton dissociation spectroscopy is a type of consequence spectroscopy employing tunable IR lasers. It provides more direct evidence of gaseous molecular structures by collecting their vibrational spectra (infrared spectra). Tunable laser systems are used to execute the dissociation process for spectroscopic purposes.

**2.2.5.1 Description of IRMPD Process.** The essential characteristic of ions that can undergo the IRMPD fragmentation is that they are IR-active, which means there is a change in dipole moment upon the absorption of IR radiation. The first step of IRMPD process is to have a photon at certain wavenumber produced by laser absorbed by precursor ion. It means the photon must be resonant with an ion's ground-state vibrational mode. However, as the molecular vibrations are not harmonic, energy differences between different

vibrational states decrease as the internal energy increase. The anharmonic vibrational ladder is compared to that of an imaginary harmonic ladder for diatomic molecule is presented in **Figure 2.11**, showing that the vibrational energy levels are not evenly spaced as they are for the anharmonic oscillator. Thus, for the anharmonicity of vibrations, the process of multiple photons absorption could not be based on the absorption process as below

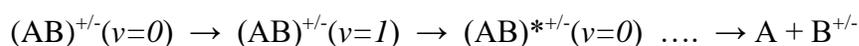


which is called ‘ladder-climbing’ style process.

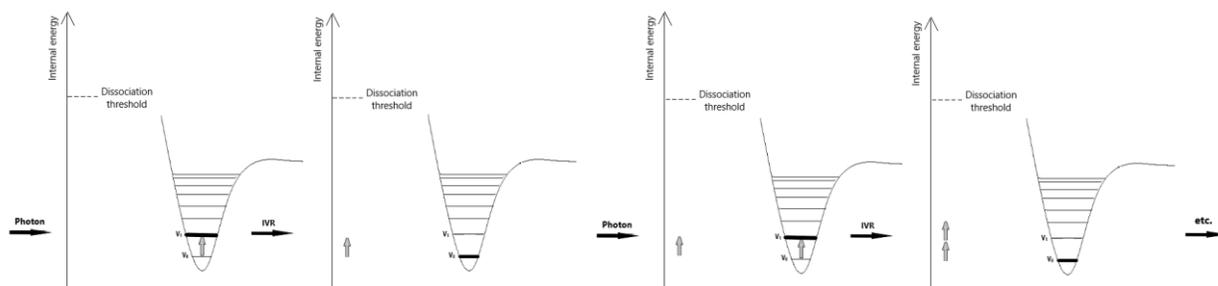


**Figure 2.11** Schematic of the vibrational energy levels. The red one is the harmonic oscillator, and the black one is the Mores Potential, which can undergo dissociation once the internal energy over the dissociation threshold.  $D_E$  is the well depth,  $D_0$  is dissociation energy.

As the numerous energy absorption is required to increase the internal energy of the ion to cause dissociation, a key step in IRMPD is intramolecular vibrational energy redistribution (IVR) that happens immediately after a photon is absorbed, which means the energy of the photon is redistributed to and stored within other vibrational degrees of freedom of the molecule. Since the absorbing vibrational mode relaxes back to ground-state,  $\nu=1$  to  $\nu=0$ , the original vibrational mode is able to absorb another resonant photon carrying the same energy. The process of IRMPD at certain wavenumber can be described as,



where \* signifies that the energies of the photons are redistributed in other internal modes of the ion through IVR. As more and more photons are absorbed, the molecule will go through the dissociation pathway once the accumulated internal energy is over the dissociation threshold of the molecule. The energy accumulation process as a molecule irradiated by IRMPD is also shown in **Figure 2.12**.



**Figure 2.12** Schematic of the IRMPD process. After absorbing the energy of a photon, the IVR process redistributes the energies to increase the internal energy of the molecule and excited-state vibration relaxes back to the ground-state to irradiate with the next photon.

A tunable IR laser is used to perform IRMPD spectroscopy by scanning the IR regions. An IRMPD spectrum is collected by tuning a laser at different frequencies and produce the photons having different energies. IRMPD spectra in this thesis were collected by scanning the fingerprint (1800-600  $\text{cm}^{-1}$ ) and C-H/O-H/N-H (3800-2800  $\text{cm}^{-1}$ ) regions of the infrared. Then the wavelength-dependent IRMPD efficiencies were calculated as the negative of the logarithm of fragments' intensities divided by the sum of the total ion intensities shown as Equation 2.15.

$$\text{IRMPD efficiencies} = -\log \left( \frac{\sum_i \text{Fragment } i}{\sum_i \text{Fragement } i + \text{Parent ion}} \right) \quad (2.14)$$

Several different types of tunable infrared laser systems have been employed to perform IRMPD spectroscopy, such as free-electron laser (FEL)<sup>12,47</sup> and optical parametric oscillator/amplifier (OPO/A).<sup>12</sup> Laser systems are not introduced in detail in this thesis, and the application of IRMPD to determine structures of nucleobase adducts with metals and protons will be discussed in chapters 3, 4, 5 and 6.

### 2.3 Computational Methods

Computational chemistry is a branch of methods of solving chemical problems by theoretical simulations. It can be used for, but is not limited to, studying the structural properties (such as bond length, bond angles, and non-covalent interactions), absolute and relative energies, dipole moments, HOMO-LUMO gap energies and vibrational frequencies. It is considered a powerful approach that can also be used to investigate the

kinetics and dynamics of chemical reactions. The purpose of this part of Chapter 2 is to introduce the different functionals under density functional theory (DFT), which is an important computational modelling method mainly used to deeply probe the physical chemistry information of systems researched in later chapters.

### 2.3.1 DFT

The wavefunctions for many-body systems contain all the quantum information of the molecule. However, it is fairly costly to solve the accurate Schrödinger equation of the complicated N-body system to obtain its accurate wavefunction. The Born-Oppenheimer (BO) is an approximation assumed that the nuclei and electrons could be treated separately.<sup>48</sup> Thus, as the nuclei are assumed to be static relative to electrons motion, the terms of the nuclear-nuclear Coulombic energy and the kinetic energy of the nuclei is approximated to be constant and zero, respectively. However, even though the electrons are not affected by the motions of nuclei in their lattice sites, they still can feel the potential force from other electrons. There are several terms of energy in the Hamiltonian, including the nuclei-electron potential energy, electron-electron potential energy, and the kinetic energy of electrons. The molecular energies can be broken down into components:

$$E_{\text{total}} = E_{\text{electronic}} + E_{\text{vibrational}} + E_{\text{rotational}} + E_{\text{translational}}. \quad (2.15)$$

Along with the success of BO assumption, molecular orbital (MO) theory uses the states of molecular orbitals, also named as the states of electrons, to calculate the electronic structures of the molecule. It stems from the linear combination of atomic orbitals (LCAO)

approximation. LCAO approximation considers the wavefunction of the molecule as a simple linear sum of weighted atomic orbitals of every atom.

Several theories have different approximations, including the Hartree-Fock method and density functional theory (DFT). DFT has been used to perform nearly all the calculations in this thesis. Other methods, such as the complete basis set (CBS), will not be introduced in detail in this chapter.

In the Hartree-Fock approximation, the electrons are assumed to move independently of each other, and every single electron feels the average static field from all the other  $N-1$  particles in the  $N$ -body system. The spin orbital is defined as two electrons containing two types of spin function, which can fill one orbital satisfying the Pauli exclusion principle. It ensures that only two anti-spin electrons fill in one orbital. The wavefunction of the molecule is written in the form of a single determinant called Slater Determinant by the spin orbitals of  $N$  atoms, which can guarantee the wavefunction is kept anti-symmetric when the interexchange of two electrons happens. The Hartree-Fock method, also termed as a self-consistent-field (SCF) method, is a solution to find the lowest energy of the ground-state atom from the Slater determinant, and that of system, with the consideration of electron correlation.

DFT is a method moving beyond the Hartree-Fock method. Hohenberg and Kohn<sup>49</sup> first determined the usage of DFT in theoretical chemistry by a theorem called Hohenberg-Kohn theorem (H-K). The H-K theorem states that the electron density can be used to determine the properties of the ground-state  $N$ -body system by using three spatial coordinates for every single electron. Thus, the electronic wavefunction in DFT method is represented by a function of  $3N$  variables,  $x, y, z$  coordinates of the  $N$ -body system, which

is greatly simplified than the traditional wavefunction of the system. It is further developed as a statement of the Kohn-Sham (K-S) theory. K-S theory provides the solution for all the complicated interacting electrons in the system by creating an imaginary system for every single particle, which considers the static effective potential which can have the same impacts as the original (real) interacting electrons. This means every electron is resolved to be affected by an assumed effective potential that includes the external potential and interelectronic interaction. The interelectronic interaction, also called Coulomb interactions, incorporates the exchange and correlation interactions. The most challenging assumptions of DFT is the exchange and correlation interactions, which derives from the solution of an idealized problem (a free electron gas). Overall, DFT is considered as a semi-empirical technique as there is the inclusion of the empirical parameters. This semi-empirical technique significantly lowers the computational cost compared to the H-F method, thus it is more efficient, which makes it attractive in the practical use of computational chemistry research. DFT is capable of calculating the properties of large systems containing hundreds of atoms.

In practice, the DFT computational method is widely applied to investigate the structural, magnetic, electronic, and thermodynamic properties of systems. Many examples have shown that DFT functionals, such as “Becke, 3-parameter, Lee-Yang-Parr” hybrid functional (B3LYP), can provide robust results and generally performs with better computational expediency than most of the post Hartree-Fock methods.

**2.3.1.1 B3LYP.** B3LYP is a type of hybrid functionals. The term hybrid functional stands for the hybridization of implicit exchange energy derived from the H-F theory, and

correlation-exchange energy obtained from the empirical data or fit from *ab initio* methods.

The exchange-correlation functional of B3LYP is :

$$E_{xc}^{B3LYP} = E_x^{LDA} + a_0 (E_x^{HF} - E_x^{LDA}) + a_x (E_x^{GGA} - E_x^{LDA}) + E_c^{LDA} + a_c (E_c^{GGA} - E_c^{LDA})$$

(2.16)

where  $a_0 = 0.20$ ,  $a_x = 0.72$  and  $a_c = 0.81$ . The subscripts 'x' and 'c' of all terms refer to the 'exchange' and 'correlation', respectively. The superscript LDA stands for the local-density approximation, which approximates the exchange and/or correlation energy functional and only depends upon the electron density at coordinates. The superscript GGA, generalized gradient approximation, stands for a better approximation compared to LDA, which expands the LDA based on the gradient change of electron density in the coordination. The  $E_x^{GGA}$  and  $E_c^{GGA}$  are the exchange functional of Becke (B)<sup>50</sup> and the correlation functional of Lee, Yang and Parr (LYP),<sup>51</sup> respectively. The first three correction terms in the equation above stand for the number "3". Thus, this method is termed "B3LYP". B3LYP is considered as the most popular hybrid functional in the DFT family.<sup>52</sup>

**2.3.1.2 Dispersion Correction.** London dispersion force is a weak non-covalent interaction between two particles. It is not included in the standard DFT methods because the fluctuation of the electron distribution is not described. Thus, the dispersion correction, which is not included in the B3LYP method, is needed to provide accurate computational results. D3 is the third version of the developed dispersion correction for B3LYP by Grimme in 2010.<sup>53</sup> It has been shown to provide a higher accuracy for intramolecular

interactions and thermodynamic results compared to the uncorrected B3LYP and D2 version correction.

Other than B3LYP, the Minnesota 06 series method, M06-2X,<sup>54</sup> was also employed for structural optimization and confirming the relations of thermodynamic results of isomers in this thesis. It was described as one of the best methods for studying the main group molecules' thermodynamics, kinetics, and non-covalent interactions such as the hydrogen bond in biomolecules without additional dispersion correction needed.<sup>55</sup>

**2.3.1.3 M06-2X.** As it was mentioned, generalized gradient approximation (GGA) is a better exchange-correlation approximation comparing to LDA. Whereas the more accurate approximation is called meta-GGA which contains the second derivative of electron density and all other terms of GGA.<sup>56</sup> Thus, this extension approximation of GGA includes electron density, electron density gradient and the non-interacting kinetic energy density as inputs to the functionals.

M06-2X<sup>54</sup> belongs to the family of Minnesota functionals developed by the group of Prof. Donald Truhlar at the University of Minnesota in 2007. It is a global hybrid functional using 29 fitted parameters<sup>57</sup> for optimization on a training set. M06-2X is a functional based on the meta-GGA with 54% non-local (or called exact) exchange treatment. This method can provide great performance for main-group thermochemistry, kinetics and noncovalent interactions, especially the reasonable accuracy in standard noncovalent interactions' tests. However, it is not recommended for transition-metal chemistry because of its high percentage of non-localized exchange hybridization. The other weakness for all Minnesota functionals, including M06-2X, is the strong

underestimation of binding energy.<sup>58</sup> Some work presented that M06-2X functionals have a better performance than B3LYP on the systems containing dispersion and ionic hydrogen bond.<sup>59</sup>

### 2.3.2 Basis Sets

As mentioned above, the LCAO approximation considers the wavefunction of the molecule to be a linear accumulation of the weighted atomic orbitals. The basis set is made up of a finite number of basis functions by using LCAO approximation to represent the electronic wavefunction, which is centred at every nucleus of the molecule. The earliest numerical basis functions are the nodeless Slater-type orbitals (STOs), which are the solutions to the Schrödinger equation for hydrogen-like atoms. However, STOs lead to the problematic integral calculation. The Gaussian-type orbitals (GTOs), a type of approximation to STOs by the linear combinations method, were used to expand the atomic orbitals and to save the calculation cost.<sup>60-63</sup> Gaussian functions are in the form

$$g_{ijk}(r) = N x^i y^j z^k e^{-\alpha r^2} \quad (2.17)$$

where  $x$ ,  $y$ , and  $z$  are the position coordinates measured from the nucleus;  $i$ ,  $j$ , and  $k$  are nonnegative integers, and the  $\alpha$  is an orbital exponent.

Different subshells are calculated with different  $i$ ,  $j$ , and  $k$  numbers. For example, the sum of  $i+j+k$  is equal to 0, 1 and 2, meaning the  $s$ ,  $p$  and  $d$  orbitals. In the practical computational chemistry process, the basis set is not constituted by several independent Gaussian functions but constructed by the normalized linear combinations of GTOs with the specific coefficients, which can provide more precise descriptions of orbitals. There are

several different types of basis sets, including minimal basis sets, split-valence basis sets, polarization basis sets, and the diffuse functions incorporated basis sets introduced below.

**2.3.2.1 Minimal Basis Sets and Split-Valence Basis Sets.** The smallest basis set is minimal basis sets, which is corresponding to use the minimum number of functions to describe the atoms – a single basis function for one orbital. However, the shortcoming of the minimal basis set arises from the molecular bonding, which makes the three Cartesian coordination need to be treated individually as inner shell and outer shell have the different coefficients when the valence electrons form a bond.<sup>64</sup>

Split-valence basis sets address the deficiency of minimal basis sets. The split-valence basis set is commonly used to provide the solution by combining different numbers of sets of functions and different numbers of Gaussian functions on the core atomic orbitals and valence atomic orbitals, separately. For example, in 6-31G split-valence basis sets, the six before the hyphen stands for six Gaussian functions used for the core orbitals, and the three and one after the hyphen means the valence electrons are described by two functions which contain three and one Gaussian components each. The expansion coefficients and Gaussian exponents used in 6-31G basis sets have been determined by the H-F method on ground-state atoms. For all the split-valence basis sets, the bigger basis sets can be distinguished by the more numbers after the hyphen; how many numbers after hyphen represents how many sets of Gaussian functions corresponding to valence electron orbitals used. Two, three and four sets of Gaussian functions described the valence electron orbitals are also named as double, triple, and quadruple-zeta( $\zeta$ ) basis sets.<sup>65</sup>

**2.3.2.2 Polarization and Diffuse Functions.** Furthermore, polarization functions are introduced into the basis sets to move the center of the electron distribution to the bonding positions, the *d*-orbital on the main group elements and the *p*-orbital on hydrogen, which has expanded valence electrons to *p* and *s* orbital, respectively. The addition of the polarization functions gives more flexibility to the wavefunction to change the shape which can be more asymmetric. The addition of (d,p), which is synonymous to \*\*, stands for a set of five *d*-type Gaussian polarization functions adds to the Li to Ca atoms (valence *p* orbitals) and a set of three *p*-type Gaussian polarization functions to H (valence *s* orbitals). The *p*-type polarization is essential for the accurate description of some H-involved bonds. By including \* in the basis set, *d*- and *f*- type polarizations are added to main group atoms and some transition metals.

The diffuse functions are crucial for the electrons loosely associated with the nuclei, which are generally related to the hydrogen bonds and lone pairs. The accurate computed intramolecular hydrogen bonds, dipole moment and anions molecule highly rely on including diffuse functions. Notation + uses for indicating the inclusive diffuse functions, which integrates a set of diffused *s*-type and *p*-type functions on non-hydrogen atoms. Furthermore, notation ++ means the hydrogens are also described by adding on the *s*-type diffuse functions.

Both the polarization and diffuse functions can lead to more reliable geometries, energies, and vibrational frequency calculations. For biomolecules forming the intramolecular hydrogen bonds in this research, the 6-31+G(d,p) were employed as the primary basis sets to optimize the structures and calculate the vibrational frequencies of

molecules accompanying with the B3LYP-D3 or M06-2X. It has been determined to be efficient and accurate in many previous works.<sup>66-70</sup>

As 6-31+G(d,p) is only defined for the atoms of H through Kr, the other split-valence basis sets with polarization functions, Def2-SVP, was introduced to couple with 6-31+G(d,p) for heavy atoms. The inner shell orbitals are replaced by the effective core potentials (ECP) by this double-zeta functional for Rb<sup>+</sup> and Cs<sup>+</sup> in Chapters 4 and 5, which is a method of using pseudopotential instead of regular Gaussian functions for inner core orbitals.

### 2.3.3 Basis Set Superposition Error (BSSE).

Interaction energies, or binding energies, between monomer A and B can be calculated by

$$E_{\text{int}} = E_{AB}^{AB}(\text{AB}) - E_A^A(\text{A}) - E_B^B(\text{B}) \quad (2.18)$$

where the subscript indicates the geometry, the superscript indicates the basis set, and in the parenthesis indicates the component. However, in computational chemistry  $E_{\text{int}}$  predicted by equation 2.18 is always bigger than the experimental/theoretical estimation on interaction energy, indicating the weaker interaction, and the distance of monomers A and B is larger than expected. This is as the result of the increasing basis set size between A and B in complex AB; the overlap between monomers A and B drives more basis sets used for describing it in complex AB comparing to separate monomer. The more accessible basis sets are used for describing the wavefunctions, the lower energy is evaluated. Such problem is referred to basis set superposition error, BSSE.<sup>71</sup>

Counterpoise Method (CP) is an approximation solution to BSSE. The basis sets applied for wavefunction description of each product is the same as the complex, which can be presented as equation 2.19

$$E_{\text{int}} = E_{AB}^{AB} (AB) - E_A^{AB} (A) - E_B^{AB} (B) \quad (2.19)$$

where the subscript indicates the geometry, the superscript indicates the basis set, and in the parenthesis indicates the component. While for each monomer, the BSSE can be described as

$$E_{\text{BSSE}} (A) = E_A^{AB} (A) - E_A^A (A) \quad (2.20)$$

$$E_{\text{BSSE}} (B) = E_B^{AB} (B) - E_B^B (B) \quad (2.21)$$

It is essential to consider BSSE in the dimeric system containing hydrogen bonds.<sup>72</sup> In this thesis, it was decided not to use BSSE for proton affinity calculations as they are generally very small due to the small basis sets used for H.

### 2.3.4 Single Point Energy Calculation

Single point energy calculations were carried out starting from the structures optimized by B3LYPD3/6-31+G(d,p) or B3LYPD3/6-31+G(d,p)/Def2-SVP. 6-311+G(3df,3pd) is a larger basis set with more polarization function, providing more accurate computed energetic results comparing to 6-31+G(d,p), or at least a point of comparison. In this research, the triple-zeta Karlsruhe, Def2-TZVP, is used for the larger atoms when single point energy calculation is conducted with 6-311+G(3df,3pd) for C, N,

O and H. The following thermochemical relations determine the absolute energy of isomers obtained by larger basis sets:

$$H = E_{\text{electronic}} + E_{\text{enthalpy correction}} \quad (2.22)$$

$$G = E_{\text{electronic}} + E_{\text{Gibbs energy correction}} \quad (2.23)$$

where the enthalpy ( $E_{\text{enthalpy correction}}$ ) and Gibbs energy ( $E_{\text{Gibbs energy correction}}$ ) corrections are computed by 6-31+G(d,p) and the electronic energy ( $E_{\text{electronic}}$ ) is calculated by 6-311+G(3df,3pd). This is because the optimization and frequency calculations are not sensitive to the size of basis sets.<sup>73</sup>

There are numerous theories and basis sets used to conduct calculations on different systems, such as Møller–Plesset (MP) perturbation theory and quantum chemistry composite methods, which are beyond the scope of this thesis and not described in this chapter.

## References

- 1 M. Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson and M. B. Alice, *J. Chem. Phys.*, 1968, **49**, 2240–2249.
- 2 G. I. Taylor, *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.*, 1964, **280**, 383–397.
- 3 H. Sommer, H. A. Thomas and J. A. Hipple, *Phys. Rev.*, 1951, **82**, 697.
- 4 T. B. McMahon and J. L. Beauchamp, *Rev. Sci. Instrum.*, 1972, **43**, 509–512.
- 5 R. T. McIver Jr, *Rev. Sci. Instrum.*, 1970, **41**, 555–558.
- 6 M. B. Comisarow and A. G. Marshall, *J. Chem. Phys.*, 1975, **62**, 293–295.
- 7 R. M. A. Heeren, A. J. Kleinnijenhuis, L. A. McDonnell and T. H. Mize, *Anal. Bioanal. Chem.*, 2004, **378**, 1048–1058.
- 8 R. Wu and T. B. McMahon, *Mass Spectrom. Rev.*, 2009, **28**, 546–585.
- 9 A. G. Marshall, *Int. J. Mass Spectrom.*, 2000, **200**, 331–356.
- 10 J. Lemaire, S. Thomas, A. Lopes, E. Louarn, H. Mestdagh, H. Latappy, J. Leprovost and M. Heninger, *Sensors (Switzerland)*, 2018, **18**, 1–15.
- 11 A. Marshall, C. Hendrickson and G. Jackson, *Mass Spectrom. Rev.*, 1998, **17**, 1–35.
- 12 N. C. Polfer, *Chem. Soc. Rev.*, 2011, **40**, 2211–2221.
- 13 Alan G. Marshall, FAQ about Fourier Transform Ion Cyclotron Resonance (FTICR)

Mass Spectrometry,  
[https://medicine.yale.edu/keck/proteomics/technologies/mass\\_spectrometry/fticrfaq/](https://medicine.yale.edu/keck/proteomics/technologies/mass_spectrometry/fticrfaq/).

- 14 P. J. Amorim Madeira, P. A. and C. M., in *Fourier Transform - Materials Analysis*, 2012.
- 15 J. H. Gross, in *Mass Spectrometry (2 edition)*, Springer, 2011, p. 753.
- 16 C. Peltz, L. Drahos and K. Vékey, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 2119–2126.
- 17 Y. Jami-Alahmadi and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2016, **18**, 2023–2033.
- 18 X. Guo, M. C. Duursma, A. Al-Khalili and R. M. A. Heeren, *Int. J. Mass Spectrom.*, 2003, **225**, 71–82.
- 19 M. Azargun and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25778–25785.
- 20 R. Mihalca, Y. E. M. van der Burgt, A. J. R. Heck and R. M. A. Heeren, *J. Mass Spectrom.*, 2007, **42**, 450–458.
- 21 H. J. Cooper, K. Håkansson and A. G. Marshall, *Mass Spectrom. Rev.*, 2005, **24**, 201–222.
- 22 A. J. Kleinnijenhuis, A. J. R. Heck, M. C. Duursma and R. M. A. Heeren, *J. Am. Soc. Mass Spectrom.*, 2005, **16**, 1595–1601.

- 23 J. M. Wells and S. A. McLuckey, *Methods Enzymol.*, 2005, **402**, 148–185.
- 24 H. Zhang, E. S. Paulsen, K. A. Walker, K. E. Krakowiak and D. V. Dearden, *J. Am. Chem. Soc.*, 2003, **125**, 9284–9285.
- 25 B. Power, V. Haldys, J. Y. Salpin and T. D. Fridgen, *Int. J. Mass Spectrom.*, 2015, **378**, 328–335.
- 26 A. K. Vrkic, R. A. J. O’Hair and C. B. Lebrilla, *Eur. J. Mass Spectrom.*, 2003, **9**, 563–577.
- 27 M. S. Weir Lipton, A. L. Fuciarelli, D. L. Springer, S. A. Hofstadler and C. G. Edmonds, *Rapid Commun. Mass Spectrom.*, 1997, **11**, 1673–1676.
- 28 J. Mo, G. C. Todd and K. Håkansson, *Biopolym. Orig. Res. Biomol.*, 2009, **91**, 256–264.
- 29 O. Y. Ali and T. D. Fridgen, *ChemPhysChem*, 2012, **13**, 588–596.
- 30 P. Lorkiewicz, R. M. Higashi, A. N. Lane and T. W.-M. Fan, *Metabolomics*, 2012, **8**, 930–939.
- 31 D. Bullinger, R. Fux, G. Nicholson, S. Plontke, C. Belka, S. Laufer, C. H. Gleiter and B. Kammerer, *J. Am. Soc. Mass Spectrom.*, 2008, **19**, 1500–1513.
- 32 C. S. Giam, T. L. Holliday, J. L. Williams, R. R. Weller, J. A. Mayernik and D. E. Hinton, *Mar. Environ. Res.*, 1989, **28**, 309–312.

- 33 J. H. Banoub, R. P. Newton, E. Esmans, D. F. Ewing and G. Mackenzie, *Chem. Rev.*, 2005, **105**, 1869–1916.
- 34 Y. Xu, C. Afonso, R. Wen and J. Tabet, *J. mass Spectrom.*, 2008, **43**, 1531–1544.
- 35 J. W. Flora, J. C. Hannis and D. C. Muddiman, *Anal. Chem.*, 2001, **73**, 1247–1251.
- 36 S. A. Hofstadler, K. A. Sannes-Lowery and J. C. Hannis, *Mass Spectrom. Rev.*, 2005, **24**, 265–285.
- 37 S. Earnshaw, *Trans. Camb. Phil. Soc.*, 1842, **7**, 97–112.
- 38 W. Paul, *Rev. Mod. Phys.*, 1990, **62**, 531–540.
- 39 G. C. Stafford, P. E. Kelley, J. E. P. Syka, W. E. Reynolds and J. F. J. Todd, *Int. J. Mass Spectrom. Ion Process.*, 1984, **60**, 85–98.
- 40 É. Mathieu, *J. Math. Pures Appl.*, 1868, **13**, 137–203.
- 41 K. Yoshinari, *Rapid Commun. Mass Spectrom.*, 2000, **14**, 215–223.
- 42 P. H. Dawson, in *Quadrupole Mass Spectrometry and its Applications*, Elsevier, 1976.
- 43 R. E. March, *J. Mass Spectrom.*, 1997, **32**, 351–369.
- 44 J. Martens, G. Berden, C. R. Gebhardt and J. Oomens, *Rev. Sci. Instrum.*, 2016, **87**, 103108.

- 45 Wikipedia contributors, Electron multiplier,  
[https://en.wikipedia.org/w/index.php?title=Electron\\_multiplier&oldid=907734097](https://en.wikipedia.org/w/index.php?title=Electron_multiplier&oldid=907734097).
- 46 T. D. Fridgen, *Mass Spectrom. Rev.*, 2009, **28**, 586–607.
- 47 D. Oepts, A. F. G. van der Meer and P. W. van Amersfoort, *Infrared Phys. Technol.*, 1995, **36**, 297–308.
- 48 M. Born and R. Oppenheimer, *Ann. Phys.*, 1927, **389**, 457–484.
- 49 P. Hohenberg and W. Kohn, *Phys. Rev. B*, 1964, **136**, 864–871.
- 50 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.
- 51 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.
- 52 K. K. Irikura and D. J. Frurip, in *ACS Symposium Series*, ACS Publications, 1998, pp. 2–18.
- 53 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 54 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
- 55 E. G. Hohenstein, S. T. Chill and C. D. Sherrill, *Methods*, 2008, **4**, 1996–2000.
- 56 Minnesota functionals,  
[https://en.wikipedia.org/wiki/Minnesota\\_functionals#Minnesota\\_06](https://en.wikipedia.org/wiki/Minnesota_functionals#Minnesota_06).
- 57 N. Mardirossian and M. Head-Gordon, *J. Chem. Theory Comput.*, 2016, **12**, 4303–

- 4325.
- 58 G. A. Dilabio and A. Otero-de-la-Roza, *Rev. Comput. Chem.*, 2016, **29**, 1–97.
- 59 M. Walker, A. J. A. Harvey, A. Sen and C. E. H. Dessent, *J. Phys. Chem. A*, 2013, **117**, 12590–12600.
- 60 R. K. Nesbet, *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.*, 1955, **230**, 312–321.
- 61 K. Oohata, H. Taketa and S. Huzinaga, *J. Phys. Soc. Japan*, 1966, **21**, 2306–2313.
- 62 S. Huzinaga, *J. Chem. Phys.*, 1965, **42**, 1293–1302.
- 63 F. Jensen, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2013, **3**, 273–295.
- 64 T. Engel, in *Physical Chemistry*, Pearson Education India, 2006.
- 65 E. R. Davidson and D. Feller, *Chem. Rev.*, 1986, **86**, 681–696.
- 66 A. A. Power, O. Y. Ali, M. B. Burt and T. D. Fridgen, *J. Phys. Chem. B*, 2017, **121**, 58–65.
- 67 C. Trujillo, A. M. Lamsabhi, O. M6, M. Y6ñez and J. Y. Salpin, *Org. Biomol. Chem.*, 2008, **6**, 3695–3702.
- 68 M. B. Burt and T. D. Fridgen, *J. Phys. Chem. A*, 2013, **117**, 1283–1290.
- 69 C. Trujillo, A. M. Lamsabhi, O. M6, M. Y6ñez and J. Y. Salpin, *Int. J. Mass Spectrom.*, 2011, **306**, 27–36.

- 70 K. Rajabi, E. A. L. Gillis and T. D. Fridgen, *J. Phys. Chem. A*, 2010, **114**, 3449–3456.
- 71 F. B. Van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt and J. H. van Lenthe, *Chem. Rev.*, 1994, **94**, 1873–1885.
- 72 S. Simon, M. Duran and J. J. Dannenberg, *J. Chem. Phys.*, 1996, **105**, 11024–11031.
- 73 J. Leszczynski, in *Handbook of Computational Chemistry*, Springer, 2012.

# Chapter 3 Self-Assembled Uracil Complexes Containing Tautomeric Uracil: An IRMPD Spectroscopic and Computation Study of the Structures of Gaseous $\text{Uracil}_n\text{Ca}^{2+}$ ( $n=4, 5, \text{ or } 6$ ) Complexes

**This chapter has been published:**

R. Cheng, V. E. Rose, B. Power and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2018, **20**, 572–580.

**Authors:**

Ruodi Cheng, Victoria E. Rose, Barry Power and Travis D. Fridgen\*

**See Appendix A for additional results.**

**Abstract:**

The structures of doubly-charged uracil (U) complexes with  $\text{Ca}^{2+}$ ,  $\text{U}_n\text{Ca}^{2+}$  ( $n=4,5,6$ ), were studied by infrared multiphoton dissociation (IRMPD) spectroscopy and theoretical methods. Ions were produced by electrospray ionization (ESI) and were isolated in the gas phase in a Fourier-transform ion cyclotron resonance mass spectrometer (FT-ICR-MS). The recorded IRMPD spectra in both the fingerprint and the C-H/N-H/O-H stretching regions and computed vibrational spectra reveal that the structures present in the greatest abundance consist of both canonical uracil as well as the lactim (or colloquially “enol”) tautomer of uracil.  $\text{U}_4\text{Ca}^{2+}$  consists of two hydrogen-bonded dimers of uracil, one canonical

and one tautomer, with each uracil interacting with  $\text{Ca}^{2+}$  through a carbonyl oxygen. The structures most consistent with the vibrational spectra of  $\text{U}_6\text{Ca}^{2+}$  consist of two hydrogen-bonded uracil trimers, in turn composed of two canonical and one tautomeric uracil, with each uracil also interacting with  $\text{Ca}^{2+}$  through carbonyl oxygen. The computed structures whose vibrational spectra best agree with the experimental vibrational spectra are also the lowest-energy structures for all three complexes.

### 3.1 Introduction.

Adenine, guanine, thymine, cytosine, and uracil are nucleobases comprising deoxyribonucleic acid (DNA) and/or ribonucleic acid (RNA). During the process of DNA and RNA replication, metal cations exhibit different effects depending on their concentrations. At high concentrations, they can interact with the nucleobases and disrupt normal hydrogen bonding patterns between base pairs. Their presence at high concentrations, therefore, can cause the mismatching of nucleobases which is considered to be the main reason for mutations, as they interfere with the native conformations of the nucleic acid polymers.<sup>1,2</sup> However, at lower concentrations, metal cations interact with the phosphate groups of the nucleic acids, stabilizing the double helix of DNA.<sup>3</sup> In the 1960s, guanosine 5'-monophosphate was found to form planar G-tetrad complexes that exhibit Hoogsteen paired motifs.<sup>4</sup> Further research in the 1980s revealed that these structures could form in DNA, stabilized by monovalent cations.<sup>5,6</sup> Studies of G-tetrads were of interest due to their biological relevance as they are found within the guanine-rich telomeric regions near the ends of chromosomes<sup>1,5,7</sup> and within the promoter regions of oncogenes.<sup>8</sup> Therefore, the G-tetrad has been investigated as a potential target for cancer therapies.<sup>9</sup> Whether these structures exist *in vivo* was controversial; there is evidence that they are formed in living cells, as proteins that recognize the G-tetrad have been found.<sup>10,11</sup> More recently, there has been evidence found that telomere end-binding proteins (TEBPs) regulate G-quadruplex DNA formation *in vivo*.<sup>12</sup>

Cation-bound self-assembled nucleobase complexes have attracted the attention of scientists not only in biological fields but also materials chemistry. Specific knowledge of

self-assembled nucleobase structures bound to metal cations can be applied in the fields of supramolecular chemistry and nanotechnologies where the ability to selectively produce molecular-scale structures is pivotal. By better understanding non-covalent interactions such as hydrogen bonds, ion-dipole interactions and  $\pi$ - $\pi$  interactions in the complexes of nucleobases and metal cations, this knowledge can be implemented for analogous self-assembly and molecular recognition processes to form supramolecular polymers,<sup>13</sup> nanofibres,<sup>14</sup> and nano-scale electronics.<sup>15</sup> Stable uracil tetramers containing metal cations at their centers, with Hoogsteen-bonding patterns analogous to the guanine quartets, have been found to exist in four-stranded RNA molecules.<sup>16</sup>

Uracil complexes in the gas phase have been studied by density functional theory (DFT) and electrospray ionization-mass spectrometry (ESI-MS) by Zins *et al.*<sup>17</sup> Lithium, sodium, and calcium cations ( $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ) were found to bond to uracil producing a series of different adducts; notable among these was a decameric complex that appeared to have unusual stability for such a complex structure. The decamer was identified as a magic number cluster, which has a high relative intensity compared to the other ions present. Potassium was not capable of forming the uracil decamer, but the calcium cation led to the formation of particularly large clusters such as 18-mers ( $\text{U}_{18}\text{Ca}^{2+}$ ) and even 24-mers ( $\text{U}_{24}\text{Ca}^{2+}$ ). These results indicate that both the size and the charge of the metal cation affect uracil aggregation. Theoretical calculations showed that a low energy decamer structure of  $\text{Na}^+$  and  $\text{Ca}^{2+}$  were the same, with two quasiplanar uracil pentamers on two sides with cations in the center. The uracil pentamer is stabilized by ten NH-O hydrogen bonds between uracil molecules. The metal cation interacts with two uracil pentamers by cation dipole interactions. These calculations also showed that the lithium cation may be too small

to form sufficient ion-dipole interactions to occupy the space between two uracil pentamers so the  $\text{Li}^+$  ion forms octameric complexes with the cation existing between two uracil tetramers. Of course, these mass spectral observations did not provide evidence to support the theoretically determined structures.

Further research into the structure of uracil pentamer complex  $[\text{U}_5\text{M}]^+$  ( $\text{M}=\text{K}^+, \text{Na}^+, \text{Cs}^+$ ) was conducted using electrospray ionization mass spectrometry (ESI-MS).<sup>18</sup> The proposed structure for these pentamers contains two hydrogen bonds between two neighbouring uracils forming a cyclic structure with the metal cation in the middle. The uracil pentamer system has a non-planar geometry with  $\text{K}^+$  in the center while  $\text{Na}^+$  bonded complex is not exactly a metal-centered structure. The  $\text{U}_5\text{Na}^+$  is not as thermodynamically stable as  $\text{U}_5\text{K}^+$  because there are fewer interactions with  $\text{Na}^+$  as a result of its off-center structure. The results also indicated that the suitable size of the metal cation is vital for stabilizing such a complex. Work by Gillis *et al.*<sup>19</sup> investigated the energetics of dissociation of trapped gas-phase uracil cluster ions consisting of 4-14 uracil units and a divalent calcium ion center ( $\text{U}_n\text{Ca}^{2+}$  where  $n=4-14$ ) by blackbody infrared radiative dissociation (BIRD) kinetics and theoretical methods. It was determined that the energy of dissociation for  $\text{U}_6\text{Ca}^{2+}$  is significantly greater than that for the larger clusters ( $n=7-14$ ) consistent with collision induced dissociation (CID) studies by Zins *et al.*<sup>20</sup> and indicated that the clusters are composed of an inner solvent shell of six uracil molecules bound directly to the  $\text{Ca}^{2+}$ , with additional uracils bound in a second solvation shell through hydrogen bonding to the inner shell uracil molecules. Theoretical and experimental results also revealed an odd-even alternation of dissociation threshold energies for the larger clusters, where complexes with an even number of uracil molecules were found to be more

stable to dissociation, indicating that the large clusters of uracil, beyond  $n=6$ , may be composed of hydrogen-bonded dimeric units of uracil bound to the inner  $U_6Ca^{2+}$  core. In this work, plausible structures were suggested, but only those containing canonical uracil were contemplated, and no spectroscopic evidence for the structures was available.

The present work aims to provide more direct experimental insight into the previously proposed structures of the  $U_nCa^{2+}$  ( $n=4-6$ ), using infrared multiple photon dissociation (IRMPD) spectroscopy in the  $1000-1900\text{ cm}^{-1}$  and  $2800-3800\text{ cm}^{-1}$  regions. The experimental IRMPD spectra are compared to theoretical infrared (IR) spectra from electronic structure calculations. Different computation methods and basis sets are used to compare the energies of isomeric structures. By comparing theoretical and experimental results, the precise structures of uracil complexes are determined.

## **3.2 Methods.**

**3.2.1 Computational Methods.** Geometry optimizations and frequency calculations were performed using the Gaussian 09 suite of programs using B3LYP density functional theory and the 6-31+G(d,p) split-valance basis set for all atoms from which the relative 298 K Gibbs energies and enthalpies of various isomers were compared. The computed IR spectra were scaled by 0.97 and 0.95 in the lower ( $900-2000\text{ cm}^{-1}$ ) and higher ( $2700-4000\text{ cm}^{-1}$ ) energy regions, respectively, to compare with the experimental IRMPD spectra. The single point energies for all lowest energy structures are performed by B3LYP/6-311+G(3df,3pd) for comparison as well. For all calculations, an empirical dispersion correction was done using Grimme's D3 version with the original D3 damping function, B3LYPD3.<sup>21</sup>

**3.2.2 Experimental Methods.** For the IRMPD spectroscopy, two different instruments were used, one for each of the fingerprint and C-H/N-H/O-H stretching region. For both, a Bruker Apex-Qe 7 Fourier-transform ion cyclotron resonance mass spectrometer (FT-ICR-MS) with an Apollo II electrospray ion source was used. Uracil and calcium chloride were purchased from Sigma-Aldrich and used without further purification. The solution producing the uracil tetramer ( $U_4Ca^{2+}$ ) was 1.7 mmol L<sup>-1</sup> uracil in 50/50 water and methanol and 5-10 drops of 10 mmol L<sup>-1</sup> in CaCl<sub>2</sub>.  $U_5Ca^{2+}$  and  $U_6Ca^{2+}$  were prepared to concentrations of 2.5 mmol L<sup>-1</sup> uracil in a 50/50 mixture of water and acetonitrile and contained 130 μmol L<sup>-1</sup> CaCl<sub>2</sub>. Solutions were injected using a 1 mL syringe at a flow rate of 112 μL h<sup>-1</sup>. The temperature of the dry gas was set at 90 °C. Prior to IRMPD, ions were isolated by standard FT-ICR isolation techniques.

For IRMPD experiments in the fingerprint region, the FT-ICR-MS was coupled to a free electron laser (FEL) at the Centre Laser Infrarouge d'Orsay.<sup>22,23</sup> Irradiation times were between 1 and 2 s and the FEL was scanned at 5 cm<sup>-1</sup> intervals. In the C-H/N-H/O-H stretch region, isolated ions were irradiated with an optical parametric oscillator/amplifier (OPO/A) in the Laboratory for the Study of Energetics, Structures, and Reactions of Gaseous Ions at Memorial University.<sup>24</sup> Ions were irradiated by the laser for 1 s, and the scan step was set as 2 cm<sup>-1</sup>. The IRMPD efficiencies (intensities) are the negative of the logarithm of the product ion intensities divided by the sum of the total ion intensities.

### **3.3 Results and Discussion.**

**3.3.1 Infrared Multiple Photon Dissociation of  $U_nCa^{2+}$ .** Mass spectra following 3420 cm<sup>-1</sup> irradiation of the isolated  $U_4Ca^{2+}$ ,  $U_5Ca^{2+}$ , and  $U_6Ca^{2+}$  complexes are displayed in Figure

S10 A, B, and C, respectively.  $U_4Ca^{2+}$  ( $m/z$  244) undergoes charge separation to  $UH^+$  ( $m/z$  113) and  $(U_3Ca-H)^+$  ( $m/z$  375), the latter of which loses uracil to form  $(U_2Ca-H)^+$  ( $m/z$  263).  $U_5Ca^{2+}$  ( $m/z$  300) undergoes loss of uracil ( $m/z$  244) which undergoes the same fragmentation as just described for  $U_4Ca^{2+}$ .

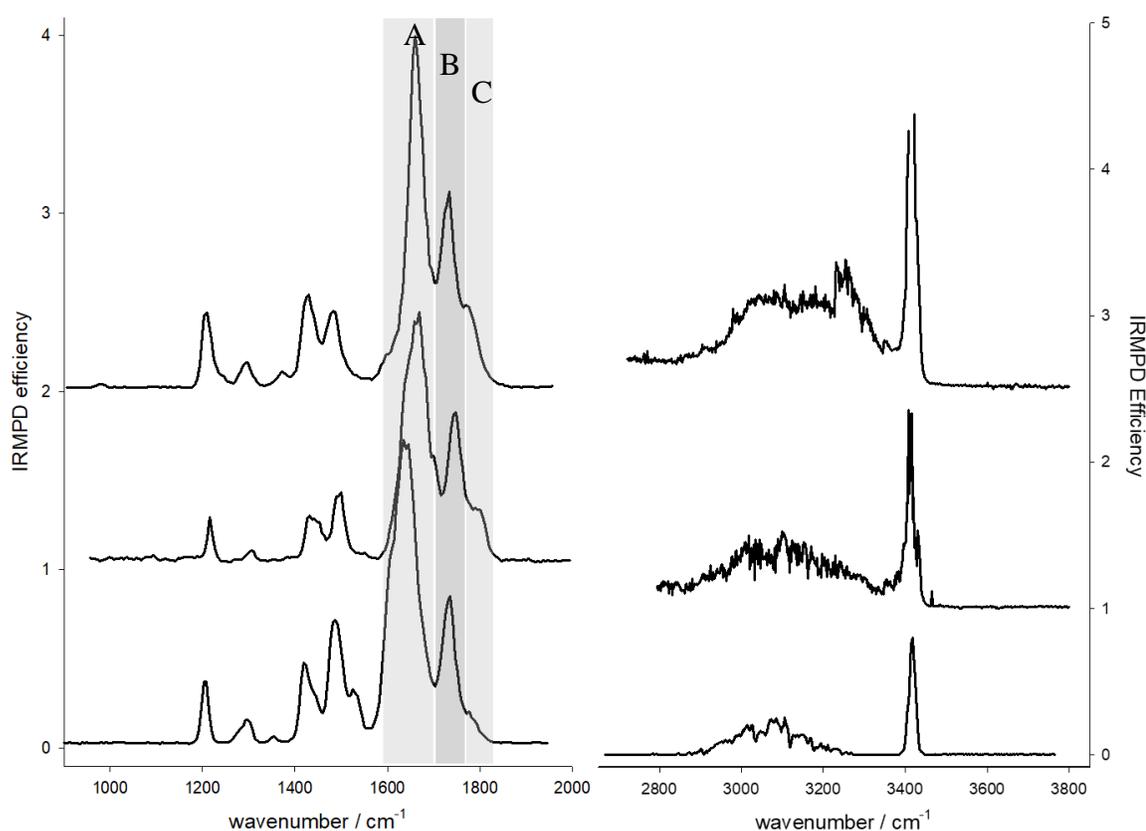
$U_5Ca^{2+}$  also undergoes charge separation forming  $UH^+$  and  $(U_4Ca-H)^+$  ( $m/z$  487) the latter of which undergoes successive losses of uracil forming  $m/z$  375, 263 and 151.

$U_6Ca^{2+}$  ( $m/z$  356) presumably undergoes loss of uracil to form  $m/z$  300 exclusively as no  $U_5Ca-H)^+$  ( $m/z$  599) is observed. The  $U_5Ca^{2+}$  then undergoes charge separation and loss of uracil as just described.

All parent and fragment ions are accounted for when determining the IRMPD efficiency to construct the IRMPD spectra which will be discussed below.

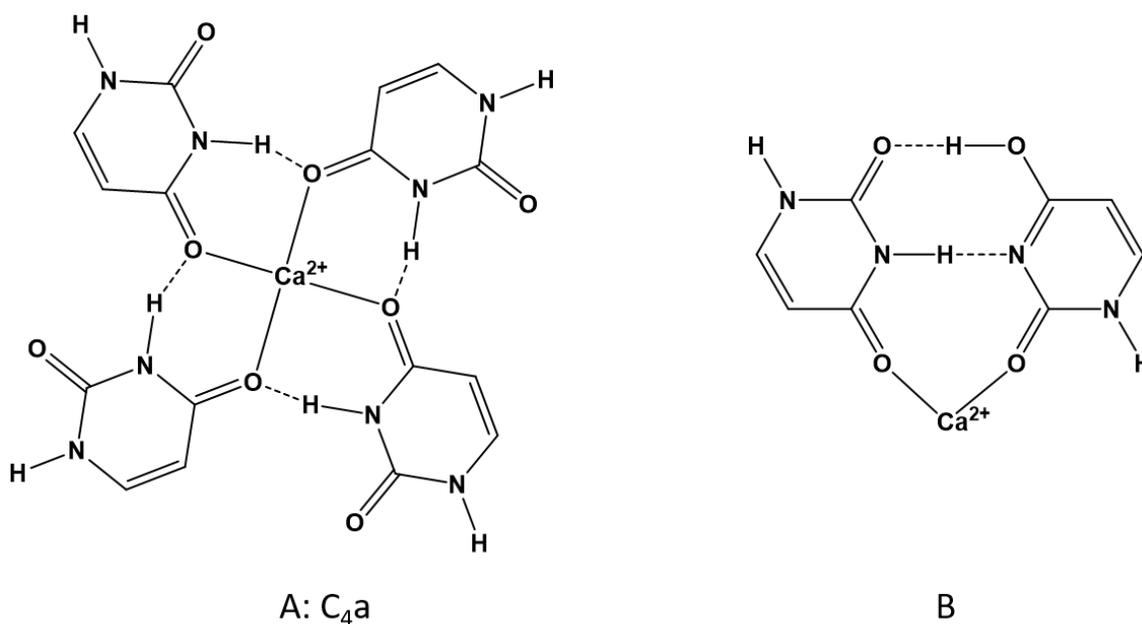
**3.3.2 IRMPD Spectra for  $U_nCa^{2+}$ .** In **Figure 3.1**, the IRMPD spectra in the fingerprint and C-H/N-H/O-H stretching region for  $U_4Ca^{2+}$ ,  $U_5Ca^{2+}$ , and  $U_6Ca^{2+}$  are presented. Clearly, in both regions, the spectra of each complex are quite similar to one another, so similar structures might be expected. In the high energy region, sharp N-H stretching bands are observed at about  $3420\text{ cm}^{-1}$  and broad hydrogen bonded N-H (or O-H—*vide infra*) are observed between about  $3400$  and  $2800\text{ cm}^{-1}$ . In the fingerprint region, the most intense bands are observed between about  $1575\text{ cm}^{-1}$  and  $1800\text{ cm}^{-1}$  correspond to strong C=O stretching for uracil. These bands are quite indicative of the structure of the complex and can be broken down into three regions. The first, indicated by ‘A’ in **Figure 3.1**, is the most red-shifted of the three regions and the band centred at about  $1635\text{ cm}^{-1}$  is the stretching of the C=O bound to the metal cation. The second region, B, has absorptions centred around

1730  $\text{cm}^{-1}$  and the third, C, weak shoulders are observed at about 1775  $\text{cm}^{-1}$ . Region C is where strong free C=O stretching vibrations would be expected and just to the red, region B, C=O stretching that has been slightly redshifted by an interaction such as a hydrogen bond is expected.



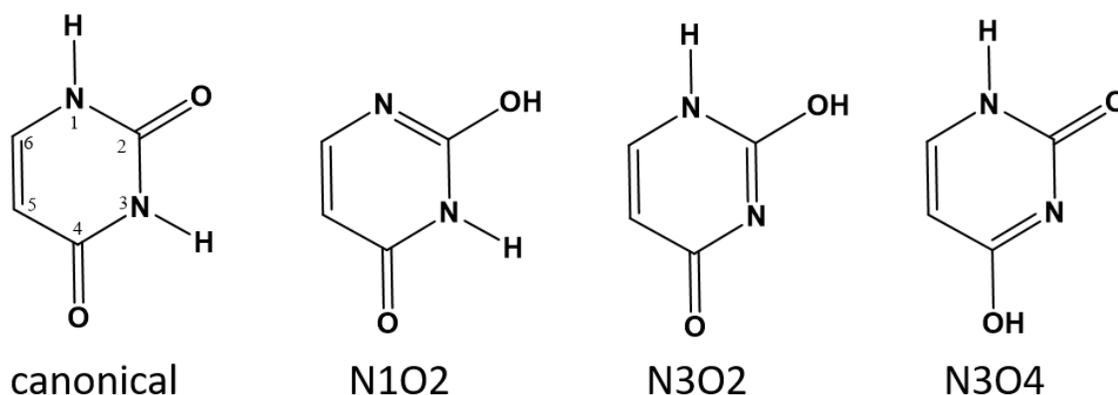
**Figure 3.1** Comparison of the experimental IRMPD spectra for  $\text{U}_n\text{Ca}^{2+}$  ( $n=4,5,6$ ) in the fingerprint and the CH/NH/OH stretching regions.

In previous work,<sup>17,19,20</sup> only canonical (or diketo) uracil was considered when trying to obtain the lowest energy structures of  $\text{Ca}^{2+}$ /uracil complexes. The lowest energy structure found for  $\text{U}_4\text{Ca}^{2+}$  (**Scheme 3.1 A**) had each of the uracils bound to  $\text{Ca}^{2+}$  through O4 and a hydrogen bond between N3H and O4 of the neighbouring uracil. This structure is called  $\text{C}_4\text{a}$ , where  $\text{C}_4$  indicates four canonical uracils and ‘a’ represents the lowest energy structure. While this structure is consistent with the spectra in the higher energy region—a free N-H group and hydrogen bonding N-H stretching—the four free C=O groups would be expected to produce intense features in region C at about  $1800\text{ cm}^{-1}$  in the fingerprint region of the infrared spectrum. The structures of the  $\text{U}_4\text{Ca}^{2+}$ ,  $\text{U}_5\text{Ca}^{2+}$ , and  $\text{U}_6\text{Ca}^{2+}$  complexes will be discussed in more detail below.



**Scheme 3.1.** (A) The lowest energy  $\text{U}_4\text{Ca}^{2+}$  composed of all diketo uracils determined in previous works.<sup>17,19,20</sup> (B) Dimeric  $\text{Ca}^{2+}$ /uracil structure composed of one diketo and one N3O4 tautomer comprising another  $\text{U}_4\text{Ca}^{2+}$  isomer.

**3.3.3 U<sub>4</sub>Ca<sup>2+</sup>.** In some recent studies, it was shown that in the presence of ions uracil<sup>25–30</sup> and thymine<sup>31</sup> tautomerize preferentially where the N3 hydrogen is instead on O4 (N3O4 tautomer) and the ion interacts with O2 and N3. While the canonical structure for uracil and thymine is significantly lower in energy than any of the enol tautomers, the N3O4 tautomer of thymine bound to a dication can be lower in energy by some 100 kJ mol<sup>-1</sup> than the canonical tautomer bound to a dication.<sup>31</sup> Therefore, the normally low energy tautomers of uracil, seen in **Scheme 3.2** need to be considered.



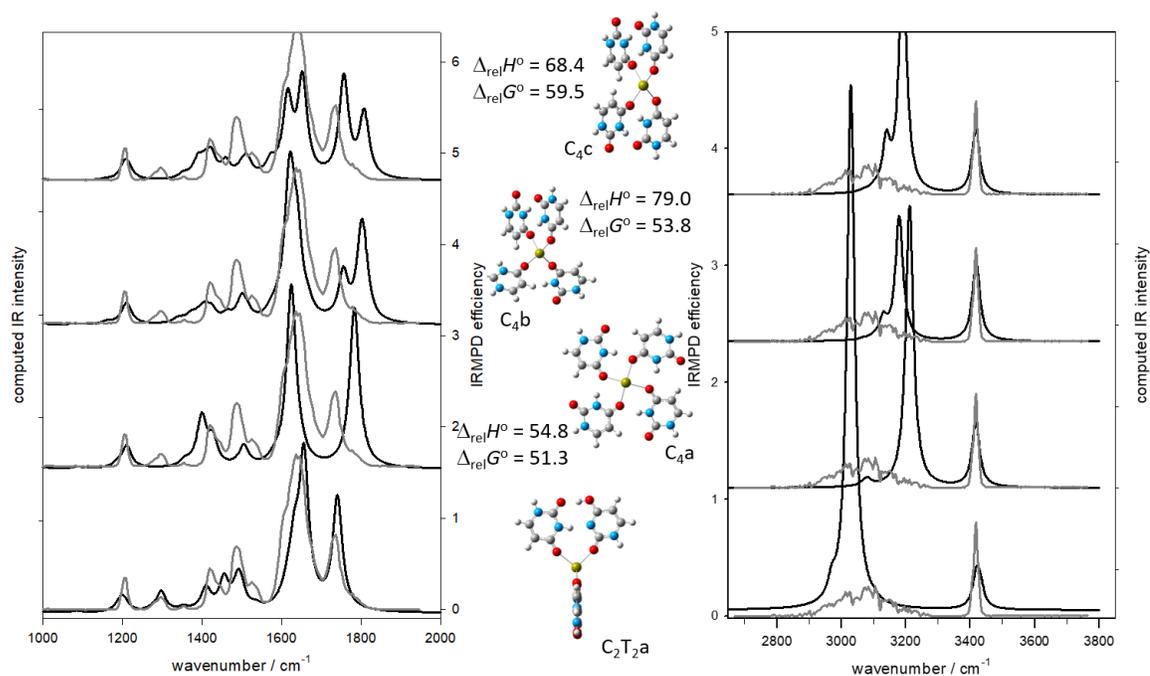
**Scheme 3.2.** Four tautomers of uracil including the numbering convention for pyrimidine bases such as uracil.

Based on this previous work and armed with the spectroscopy in **Figure 3.1**, a structure for U<sub>4</sub>Ca<sup>2+</sup> composed of two dimers of uracil, each with one canonical and one N3O4 tautomer, both bound to Ca<sup>2+</sup> (see **Scheme 3.1 B**) was constructed. In this structure, canonical uracil is bound to Ca<sup>2+</sup> through O4 and the N3O4 tautomer is bound to Ca<sup>2+</sup> through O2 providing the strong band in region A. There is a hydrogen bonded N-H group and a hydrogen bonded O-H group, as well as two free N-H groups, consistent with the

spectra in the high energy region. There is a hydrogen bonded carbonyl providing the strong band in region B and no free carbonyl group, again consistent with the infrared spectrum. The  $U_4Ca^{2+}$  complex based on these dimers is named  $C_2T_2a$  (C for canonical uracil and T for tautomeric uracil) is calculated to be lower in 298 K Gibbs energy by  $51.3 \text{ kJ mol}^{-1}$  (see **Figure 3.2**) than the lowest energy structure determined with all canonical uracils.

In **Figure 3.2**, the experimental IRMPD spectra (grey traces) are compared with the computed IR spectra (black traces) for the structures depicted in the centre of the figure. It is clear that the computed spectrum for the lowest energy structure,  $C_2T_2a$ , compares very well with the entirety of the recorded IRMPD spectrum. It is equally apparent that the computed infrared spectrum for the  $C_4a$  structure, discussed above, is not consistent with the C=O stretching region of the experimental IRMPD spectrum. The obvious disagreement is the absence of a strong free C=O stretch in the experimental IRMPD spectrum, which is predicted by the calculations, and the presence of the hydrogen bonded C=O stretch in the IRMPD spectrum, which is not predicted by the calculations. **Figure 3.2** also shows the second and third lowest energy structures from Gillis *et al.*,<sup>19</sup> which are also high in energy compared to  $C_2T_2a$ , and have similar inconsistencies with the IRMPD spectrum mainly in the  $1700\text{-}1800 \text{ cm}^{-1}$  region.

There are other isomers of  $U_4Ca^{2+}$  consisting of two dimers composed of one canonical and one tautomeric uracil.  $C_2T_2b$  is similar to  $C_2T_2a$  except the tautomeric uracil has the N3 hydrogen shifted to O2 and this uracil is bound to  $Ca^{2+}$  through O4, see Figure S1 (supplementary material). This results in a structure that is only  $8.0 \text{ kJ mol}^{-1}$  higher in 298 K Gibbs energy than  $C_2T_2a$ . The IRMPD spectrum in the fingerprint region is also



**Figure 3.2** Comparison of the experimental IRMPD spectrum (grey trace) for  $\text{U}_4\text{Ca}^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for four different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $\text{C}_2\text{T}_2\text{a}$  are also shown (also computed using B3LYP/6-31+G(d,p)),  $\text{kJ mol}^{-1}$ .

consistent with the computed IR spectrum for  $\text{C}_2\text{T}_2\text{b}$ . In the higher energy region, there is a pronounced red-shift computed for the hydrogen bonded O-H stretch; however, this is not enough to rule out a contribution to the experimental spectrum from  $\text{C}_2\text{T}_2\text{b}$ .  $\text{C}_2\text{T}_2\text{a/b}$  contains one dimer each as the ones in  $\text{C}_2\text{T}_2\text{a}$  and  $\text{C}_2\text{T}_2\text{b}$  and results in a structure that is computed to be only  $3.9 \text{ kJ mol}^{-1}$  higher in 298 K Gibbs energy than the lowest energy  $\text{C}_2\text{T}_2\text{a}$  isomer. Assuming a thermal distribution, it would be expected that  $\text{C}_2\text{T}_2\text{a/b}$  contributes about 20% of the mixture of ions. It is clear that  $\text{C}_2\text{T}_2\text{a/b}$  also reproduces the

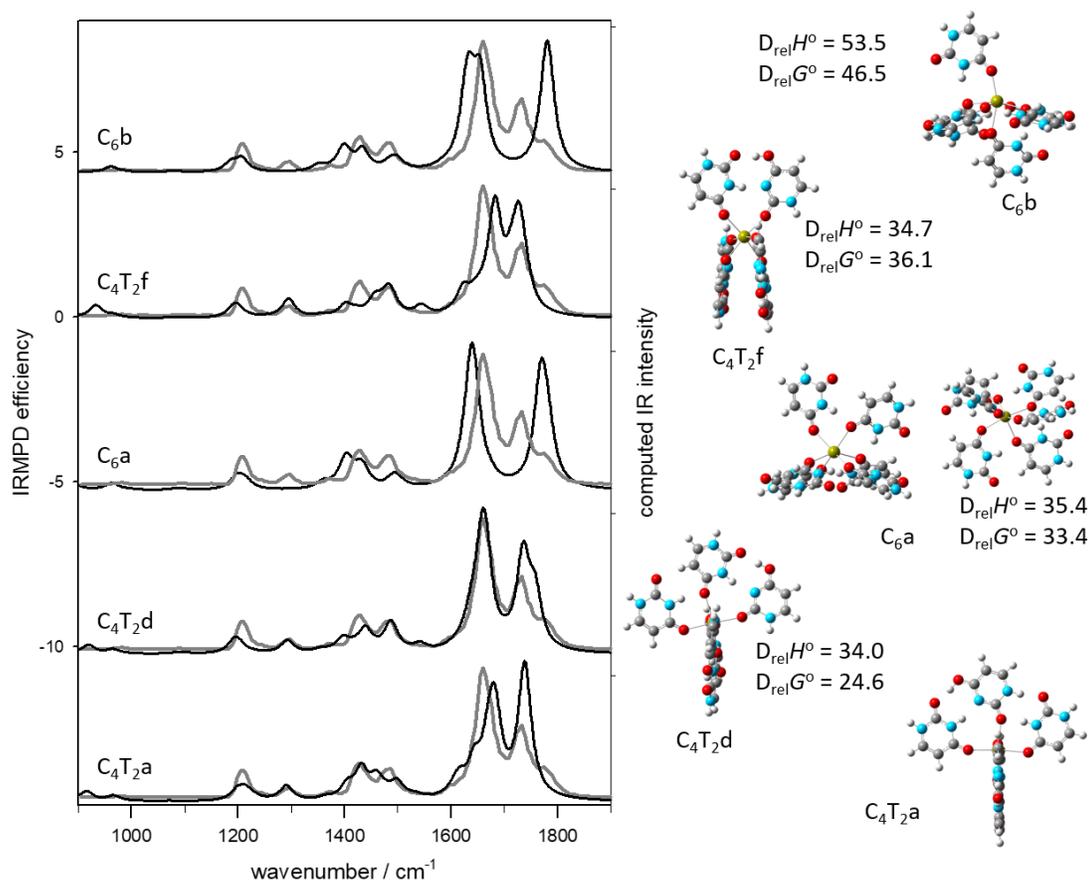
experimental spectrum quite well and unlike C<sub>2</sub>T<sub>2a</sub> predicts the weak band observed at about 1530 cm<sup>-1</sup>.

C<sub>2</sub>T<sub>2c</sub> and C<sub>2</sub>T<sub>2d</sub> both have the canonical uracil bound to Ca<sup>2+</sup> through O2 and differ from each in the same way that C<sub>2</sub>T<sub>2a</sub> and C<sub>2</sub>T<sub>2b</sub> differ from one another. C<sub>2</sub>T<sub>2c</sub> and C<sub>2</sub>T<sub>2d</sub> are more than 20 kJ mol<sup>-1</sup> higher in energy than C<sub>2</sub>T<sub>2a</sub> and their computed spectra do not compare as well with the experimental IRMPD spectrum. It is concluded from these experiments and calculations that the main contributor to the experimental IRMPD spectrum is the C<sub>2</sub>T<sub>2a</sub> structure with a tetrahedral arrangement of two uracil hydrogen bonded dimers with one member of the dimer canonical and one which is a keto-enol tautomer.

Some other higher energy structures for U<sub>4</sub>Ca<sup>2+</sup> are shown in **Figure S7** along with their computed energies and a comparison of the computed and experimental vibrational spectra. These include structures based on four N3O4 or N1O4 tautomeric uracils whose computed vibrational spectra do not reproduce the experimental spectra.

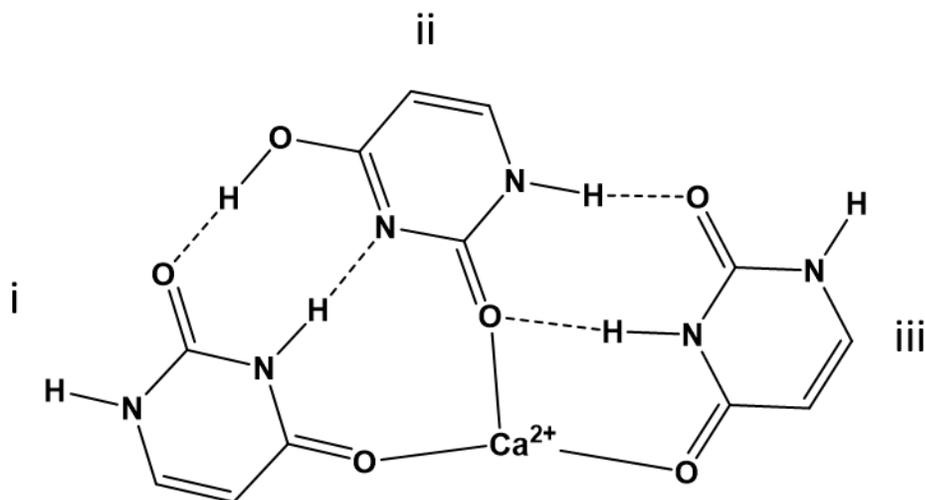
**3.3.4 U<sub>6</sub>Ca<sup>2+</sup>.** To construct a structure for U<sub>6</sub>Ca<sup>2+</sup>, like U<sub>4</sub>Ca<sup>2+</sup>, that is consistent with its experimental vibrational spectrum, a third dimer as in **Scheme 3.1 B** is added to U<sub>4</sub>Ca<sup>2+</sup> to form an octahedral structure. The complex optimizes to the octahedral structure labelled C<sub>3</sub>T<sub>3a</sub> in **Figure 3.3**. This isomer, based on the three dimers, “collapses” such that two of the dimers have a  stacking interaction. The agreement between the computed IR spectrum for this structure and the IRMPD spectrum is certainly not poor enough to rule the structure out.

Another way to form the  $U_6Ca^{2+}$  is to add canonical uracil to each of the dimers, constructing two trimers with two canonical uracils and one of the tautomers in each, as in **Scheme 3.3**, where both canonical uracils are hydrogen bonded to the tautomeric uracil. The added uracil in this trimer (iii in **Scheme 3.3**) is also bound to  $Ca^{2+}$  through O4. The



**Figure 3.3** Comparison of the experimental IRMPD spectrum (grey trace) for  $U_6Ca^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $C_4T_2a$  are also shown (also computed using B3LYP/6-31+G(d,p)), kJ mol<sup>-1</sup>.

addition of this third uracil to each of the dimers in the lowest energy tetramer structure, C<sub>2</sub>T<sub>2</sub>a, results in a structure where the two trimers are perpendicular to each other, forming a roughly octahedral complex, and is lower in 298 K Gibbs energy than C<sub>4</sub>T<sub>2</sub>f by 36.1 kJ mol<sup>-1</sup>. Distortion of the complexes from an octahedral structure is due to the inter-trimer uracil hydrogen bonding. The computed infrared spectrum for this lowest-energy structure, C<sub>4</sub>T<sub>2</sub>a, can be seen in **Figure 3.3** and compares very well—better than C<sub>4</sub>T<sub>2</sub>f—with the experimental spectrum in the fingerprint region and the higher energy region (see **Figure S2**). This structure, along with C<sub>4</sub>T<sub>2</sub>b, C<sub>4</sub>T<sub>2</sub>c, and C<sub>4</sub>T<sub>2</sub>e, compared in **Figure S3**, maximize the number of hydrogen bonds with two N-H to O=C, one O-H to O=C, and one N-H to N. However, C<sub>4</sub>T<sub>2</sub>a matches the IRMPD spectrum in the fingerprint region much better. C<sub>4</sub>T<sub>2</sub>b differs in that one canonical uracil (iii in **Scheme 3.3**) is bound through O2 rather than O4 and is 13 kJ mol<sup>-1</sup> higher in energy. C<sub>4</sub>T<sub>2</sub>c and C<sub>4</sub>T<sub>2</sub>e have the two canonical uracils hydrogen bonded together and are higher in energy by 16 and 30 kJ mol<sup>-1</sup>, respectively.



**Scheme 3.3** Lowest energy trimeric Ca<sup>2+</sup>/uracil structure composed of two canonical uracils and one N3O4 tautomer comprising the lowest energy U<sub>6</sub>Ca<sup>2+</sup> isomer.

Another structure that has a computed IR spectrum that compares favourably with the IRMPD spectrum is C<sub>4</sub>T<sub>2</sub>d (**Figure 3.3** and **Figure S2**). This structure, like C<sub>4</sub>T<sub>2</sub>c and C<sub>4</sub>T<sub>2</sub>e, differs from C<sub>4</sub>T<sub>2</sub>a in that the two canonical uracils in each of the trimers are hydrogen bonded together, in the case of C<sub>4</sub>T<sub>2</sub>d via an N3-H to O4 hydrogen bond and a C5-H to O2 interaction. While this structure is consistent with the IRMPD spectrum, it is 24.6 kJ mol<sup>-1</sup> higher in 298 K Gibbs energy which makes it unlikely to be observed.

As was the case for the tetramers, the previously determined hexameric structures based solely upon canonical uracils,<sup>19</sup> *ie.* C<sub>6</sub>a and C<sub>6</sub>b in **Figures 3.3** and **S2**, are higher in energy than C<sub>4</sub>T<sub>2</sub>a and their computed IR spectra do not compare well with the IRMPD spectra. The main problem with the computed spectra is that they contain a free C=O stretch (region C) and do not predict the hydrogen bonded C=O stretch (region B) which is observed experimentally. Some other higher energy structures and their computed spectra are shown in **Figure S9**, including those where the two trimers are mixed in that they contain a trimer of C<sub>4</sub>T<sub>2</sub>a and one of either C<sub>4</sub>T<sub>2</sub>b, c, or e. These ‘mixed’ complexes are intermediate in energy; for example, C<sub>4</sub>T<sub>2</sub>a/b is higher in energy than C<sub>4</sub>T<sub>2</sub>a, but lower in energy than C<sub>4</sub>T<sub>2</sub>b. They are not nearly as good a fit to the IRMPD spectra in comparison to C<sub>4</sub>T<sub>2</sub>a (**Figure S9**).

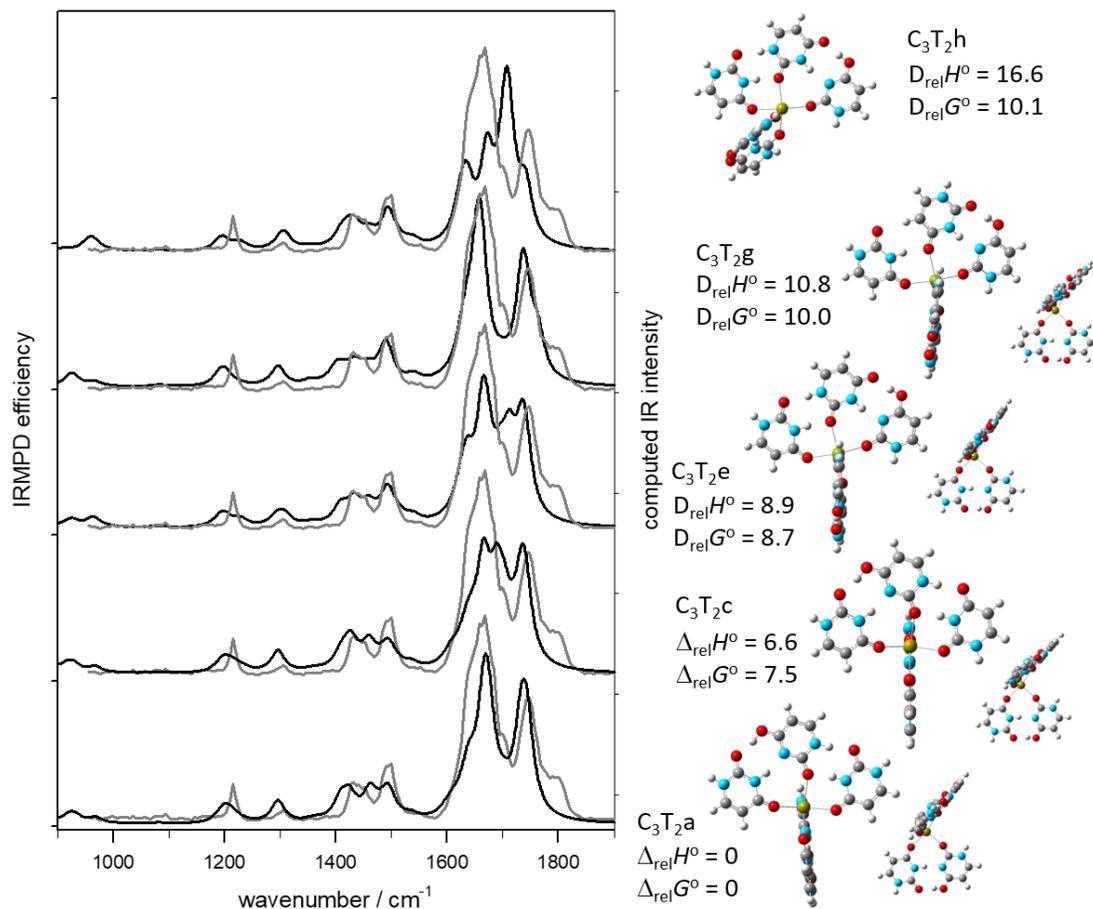
**3.3.5 U<sub>5</sub>Ca<sup>2+</sup>.** The IRMPD spectrum of U<sub>5</sub>Ca<sup>2+</sup> is compared to the computed IR spectra of several isomers in **Figures 3.4, S4, S5, S6, and S8**. Removal of one of the canonical uracils (iii in **Scheme 3.3**) from the lowest energy U<sub>6</sub>Ca<sup>2+</sup> structure results in two very similar structures, C<sub>3</sub>T<sub>2</sub>a and C<sub>3</sub>T<sub>2</sub>b (in **Figures 3.4** and **S4**, respectively) which only differ in 298

K Gibbs energy by  $0.6 \text{ kJ mol}^{-1}$ . They do not differ significantly in structure, simply whether tautomeric or canonical uracil in the remaining dimer is closest to the trimer. The IR spectrum of  $\text{C}_3\text{T}_2\text{a}$  is consistent with the experimental IRMPD spectrum, but certainly,  $\text{C}_3\text{T}_2\text{b}$  cannot be ruled out spectroscopically or energetically.  $\text{C}_3\text{T}_2\text{c}$  and  $\text{C}_3\text{T}_2\text{d}$  differ from  $\text{C}_3\text{T}_2\text{a}$  and  $\text{C}_3\text{T}_2\text{b}$  in that uracil iii in the trimer is bound through O2 rather than O4 and results in structures that are a little less than  $8 \text{ kJ mol}^{-1}$ , higher in 298 K Gibbs energy. While neither of these structures can be ruled out spectroscopically, they do not agree as well as the two lower energy structures in the C=O stretching region.  $\text{C}_3\text{T}_2\text{e}$  and  $\text{C}_3\text{T}_2\text{f}$  differ in that the trimeric moiety is composed of tautomeric uracil bound to a canonical uracil bound to  $\text{Ca}^{2+}$  via O2 and the second canonical uracil is bound through O4, but hydrogen bonded to canonical uracil. This results in structures just slightly higher than the previously discussed structure and their computed IR spectra are not as good a match to the experimental vibrational spectrum in the fingerprint region. Other computed IR spectra for higher energy structures are compared in **Figure 3.4** and the supplementary material.

In **Figure S6**, the computed IR spectra for the three lowest energy C5-based structures that were determined previously<sup>19</sup> are compared to the experimental IRMPD spectrum. Besides being between  $23$  and  $33 \text{ kJ mol}^{-1}$  higher in energy than  $\text{C}_2\text{T}_3\text{a}$ , spectroscopically these isomers can be ruled out based on the poor agreement in the C=O stretching region as discussed previously for the  $\text{U}_4\text{Ca}^{2+}$  and  $\text{U}_6\text{Ca}^{2+}$  complexes.

**3.3.6 Other Observed  $\text{U}_{5,6}\text{Ca}^{2+}$  Structures.** In the IRMPD spectra for  $\text{U}_5\text{Ca}^{2+}$  and  $\text{U}_6\text{Ca}^{2+}$ , there is a weak but pronounced shoulder observed in region C of **Figure 3.1**, just to the red

of  $1800\text{ cm}^{-1}$ . The band can only be attributed to a free C=O stretch which is not present in any of the lowest energy structures. The only structures found that contain a free C=O



**Figure 3.4** Comparison of the experimental IRMPD spectrum (grey trace) for  $U_5Ca^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $C_3T_2a$  are also shown (also computed using B3LYP/6-31+G(d,p)),  $\text{kJ mol}^{-1}$ .

stretch are those not based on the dimer or trimer containing an N3O4 tautomer. Because the lowest energy structures of pentamer and hexamer clusters lack the free C=O stretch, the existence of higher energy clusters is likely due to the presence of this shoulder peak at

1800  $\text{cm}^{-1}$  observed in the experimental spectrum. For example, the lowest energy isomer containing a free C=O stretch for  $\text{U}_6\text{Ca}^{2+}$ , is  $\text{C}_{6a}$ . This isomer is also some 33  $\text{kJ mol}^{-1}$  higher in 298 K Gibbs energy than  $\text{C}_{4\text{T}2a}$ , and if present, it must be kinetically trapped, probably during the electrospray process. Tautomerization of the all keto tautomer of uracil to the enol-keto tautomer proceeds with a very high activation energy, some 200  $\text{kJ mol}^{-1}$  for neutral uracil.<sup>32</sup> The tautomerization barrier is computed to be lower, but still considerable (70  $\text{kJ mol}^{-1}$ , 298 K Gibbs energy barrier) for both the protonated and water-assisted tautomerization<sup>33</sup> and an intramolecularly assisted uracil tautomerization in  $(\text{U}_2\text{-H})\text{Cu}^+$ .<sup>28</sup>

**3.3.7 Computed Energies.** The B3LYPD3/6-31+G(d,p) lowest energy structures for  $\text{U}_4\text{Ca}^{2+}$  as well as for the lowest energy  $\text{C}_3\text{T}_1$  and  $\text{C}_4$  structures, were submitted to a single point calculation at the same level of theory but with the larger 6-311+G(3df,3pd) basis. The calculations are compared in **Table 3.1**, using the thermal corrections to the 298 K enthalpy and Gibbs energy from the vibrational calculations with a smaller basis. Clearly, there is very little difference between the computed energies using the smaller or larger basis sets. A similar set of calculations were performed on various  $\text{U}_5\text{Ca}^{2+}$  and  $\text{U}_6\text{Ca}^{2+}$  isomers with the same conclusion (see supplementary **Tables S2** and **S3**).

### 3.4 Conclusions

A combination of IRMPD spectroscopy and computational chemistry was used to show that the structures of three doubly charged uracil/calcium clusters:  $\text{U}_4\text{Ca}^{2+}$ ,  $\text{U}_5\text{Ca}^{2+}$ ,

and  $U_6Ca^{2+}$ , all contain two N3O4 enol-tautomerized uracils. The lowest energy structure for the  $U_4Ca^{2+}$  complex consists of two uracil dimers, each consisting of one canonical and one tautomeric uracil, with two hydrogen bonds between them. The lowest energy  $U_6Ca^{2+}$

**Table 3.1** Comparison of basis set on the 298 K energetics (top relative enthalpies, bottom relative Gibbs energies) in  $\text{kJ mol}^{-1}$  of some  $U_4Ca^{2+}$  structures.

Structure	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-311+G(3df,3pd)
	0.0	0.0
$C_2T_2a$	0.0	0.0
	7.3	6.8
$C_2T_2b$	8.0	7.5
	21.7	21.7
$C_2T_2c$	22.4	22.3
	24.6	24.3
$C_2T_2d$	25.6	25.4
	35.9	35.2
$C_3T_1a$	31.8	31.2
	54.8	58.0
$C_4a$	51.3	54.5

complex consists of two uracil trimers, the same dimeric structure as in  $U_4Ca^{2+}$  but a second canonical uracil doubly hydrogen bonded to the other side of the tautomeric uracil. The

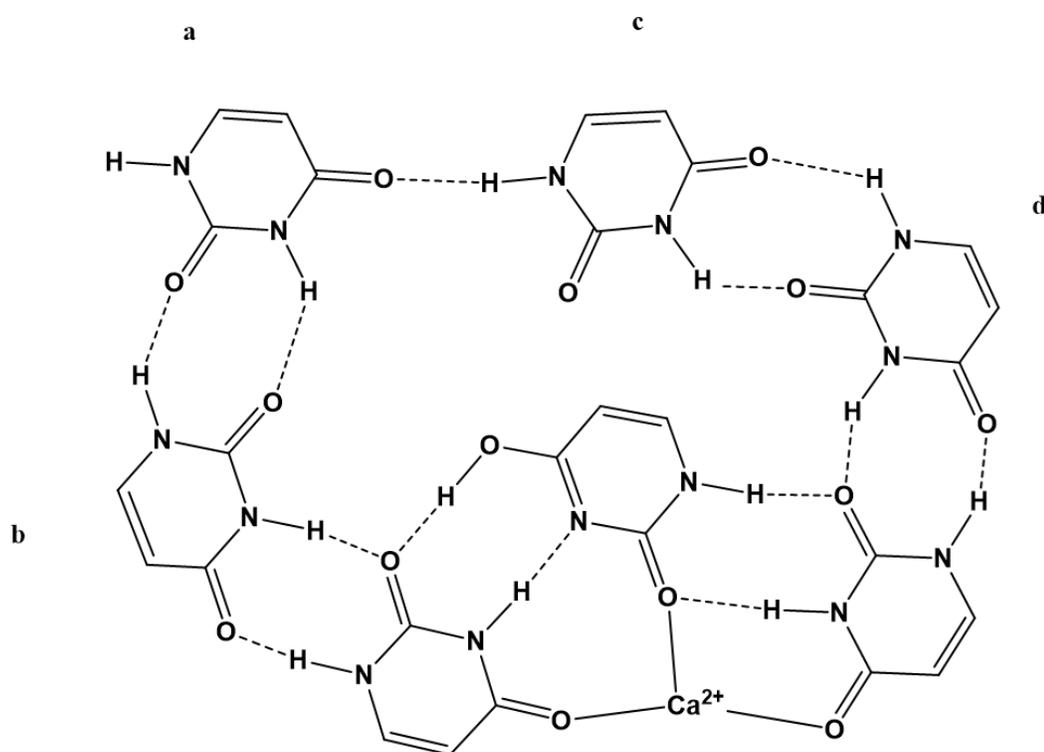
lowest energy  $U_5Ca^{2+}$  consists of one dimer and one trimer as in the  $U_4Ca^{2+}$  and  $U_6Ca^{2+}$ , respectively. The computed IR spectra for the lowest energy structures for each of the complexes were consistent with the experimental IRMPD spectra in all cases.

For  $U_4Ca^{2+}$  a second structure, only a few  $\text{kJ mol}^{-1}$  higher than the lowest energy structure, with an IR spectrum also consistent with the IRMPD spectrum also accounts for a small absorption at about  $1530\text{ cm}^{-1}$ . A weak shoulder observed in the free C=O stretch region for both  $U_4Ca^{2+}$  and  $U_4Ca^{2+}$  are thought to be due to a small amount of higher energy, kinetically trapped complexes with only canonical uracils, as these are the only ones with predicted absorptions in these regions.

The structures determined for  $U_4Ca^{2+}$ ,  $U_5Ca^{2+}$ , and  $U_6Ca^{2+}$  are vastly different than those previously determined and only composed of canonical uracils.<sup>17,19,20</sup> This is easily seen in the C=O stretch region where there is the absence of a strong absorption in the IRMPD spectra due to a free C=O group which is predicted for canonical uracil based structures and the presence of a hydrogen bonded C=O stretch in the IRMPD spectrum which is not predicted for the canonical uracil based structures.

In previous experiments<sup>19</sup> the BIRD dissociation kinetics of  $U_nCa^{2+}$  were observed where  $n=5-14$ . The thermochemistries of dissociation led to the conclusion that  $U_{14}Ca^{2+}$  was composed of a  $U_6Ca^{2+}$  core surrounded by four hydrogen bonded dimers of uracil in the second solvation shell, each hydrogen bonded to the core. This conclusion was based on the even uracil ( $n=14, 12, 10, 8$ ) containing complexes being more strongly bound than the odd uracil containing complexes ( $n=13, 11, 9, 7$ ), the difference between the successive dissociations, *ie.*  $n=14$  and  $13$  being about the strength of a hydrogen bond,  $10\text{ kJ mol}^{-1}$ . While a detailed computational study of these larger complexes is beyond the scope of this

study, it is interesting to contemplate a structure for the  $U_{14}Ca^{2+}$  complex that also explains this even-odd dissociation energy observation. One such structure is proposed in **Scheme 3.4** (only showing half the uracils in the complex) based on a core containing the  $C_4T_2a$  isomer of  $U_6Ca^{2+}$ . In this structure, each uracil dimer is bound to the complex via three interactions. The uracil labelled 'a' in **Scheme 3.4** is bound by three hydrogen bonds, one to the other dimer, and might be the first to be lost, leaving the uracil labelled 'b' bound by only two hydrogen bonds and the second to be lost. Uracil 'c' is bound by three interactions and would be the third uracil lost followed by 'd' which is only bound to the core by two hydrogen bonds.



**Scheme 3.4** Possible structure of one half of  $U_{14}Ca^{2+}$  with the 6 mer core composed of two trimeric structures.

**Conflicts of interest.** There are no conflicts to declare.

## References

- (1) Lippert, B. *J. Chem. Soc. Dalt. Trans.* **1997**, *21*, 3971–3976.
- (2) Reed, J. E.; Arnal, A. A.; Neidle, S. *J. Am. Chem. Soc.* **2006**, *128* (18), 5992–5993.
- (3) Eichhorn, G. L.; Berger, N. A.; Butzow, J. J.; Clark, P.; Heim, J.; Pitha, J.; Richardson, C.; Rifkind, J. M.; Shin, Y.; Tarien, E. In *Metal Ions in Biological Systems: Studies of Some Biochemical and Environmental Problems*; Dhar, S. K., Ed.; Springer US: Boston, MA, 1973; pp 43–66.
- (4) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48* (12), 2013–2018.
- (5) Sundquist, W. I.; Klug, A. *Lett. to Nat.* **1989**, *342*, 825–829.
- (6) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871–880.
- (7) Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703–730.
- (8) Yuan, L.; Tian, T.; Chen, Y.; Yan, S.; Xing, X.; Zhang, Z.; Zhai, Q.; Xu, L.; Wang, S.; Weng, X.; Yuan, B.; Feng, Y.; Zhou, X. *Sci. Rep.* **2013**.
- (9) Kelland, L. R. *Eur. J. Cancer* **2005**, *41*, 971–979.
- (10) Erlitzki, R.; Fry, M. *J. Biol. Chem.* **1997**, *272* (25), 15881–15890.
- (11) Horvath, M. P.; Schultz, S. C. *J. Mol. Biol.* **2001**, *310*, 367–377.
- (12) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. *J. Nat. Struct. Mol. Biol.* **2005**, *12* (10), 847–854.
- (13) Rowan, S. J.; Suwanmala, P.; Sivakova, S. *J. Polym. Sci. Part A Polym. Chem.* **2003**, *41*, 3589–3596.
- (14) Shimizu, T.; Iwaura, R.; Masuda, M.; Hanada, T. *J. Am. Chem. Soc.* **2001**, *123* (8),

5947–5955.

- (15) Eichen, Y.; Braun, E.; Sivan, U. *Acta Polym.* **1998**, *49*, 663–670.
- (16) Cheong, C.; Moore, P. B. *Biochemistry* **1992**, *31*, 8406–8414.
- (17) Zins, E.; Rochut, S.; Pepe, C. *J. Mass Spectrom.* **2009**, *2008* (August 2008), 40–49.
- (18) Qiu, B.; Liu, J.; Qin, Z.; Wang, G.; Luo, H. *Chem. Commun.* **2009**, No. 20, 2863.
- (19) Gillis, E. A. L.; Demireva, M.; Nanda, K.; Beran, G.; Williams, E. R.; Fridgen, T. *D. Phys. Chem. Chem. Phys.* **2012**, *14* (10), 3304–3315.
- (20) Zins, E.; Rochut, S.; Pepe, C. *J. Mass Spectrom.* **2009**, *2009* (September 2008), 813–820.
- (21) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132* (15).
- (22) Bakker, J. M.; Besson, T.; Scuderi, D.; Mai, P. *J. Phys. Chem. A* **2007**, *111*, 13415–13424.
- (23) Prazeres, R.; Glotin, F.; Insa, C.; Jaroszynski, D. A.; Ortega, J. M. *Eur. Phys. J. D* **1998**, *3*, 87–93.
- (24) Azargun, M.; Jami-Alahmadi, Y.; Fridgen, T. D. *Phys. Chem. Chem. Phys.* **2017**, *19*, 1281–1287.
- (25) Power, B.; Haldys, V.; Salpin, J. Y.; Fridgen, T. D. *Int. J. Mass Spectrom.* **2015**, *378*, 328–335.
- (26) Power, A. A.; Ali, O. Y.; Burt, M.; Fridgen, T. D. *Interna* **2012**, *330*, 233–240.
- (27) Ali, O. Y.; Randell, N. M.; Fridgen, T. D. *ChemPhysChem* **2012**, *13*, 1507–1513.
- (28) Ali, O. Y.; Fridgen, T. D. *ChemPhysChem* **2012**, *13*, 588–596.
- (29) Power, B.; Rowe, S.; Fridgen, T. D. *J. Phys. Chem. B* **2017**, *121*, 58–65.
- (30) Akinyemi, T. E.; Wu, R. R.; Nei, Y.-W.; Cunningham, N. A.; Roy, H. A.; Steill, J.

- D.; Berden, G.; Oomens, J.; Rodgers, M. T. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (11), 2438–2453.
- (31) Kabeláč, M.; Hobza, P. *J. Phys. Chem. B* **2006**, *110* (29), 14515–14523.
- (32) Hu, X.; Li, H.; Liang, W.; Han, S. *J. Phys. Chem. B* **2005**, *109* (12), 5935–5944.
- (33) Bakker, J. M.; Sinha, R. K.; Besson, T.; Brugnara, M.; Tosi, P.; Salpin, J.; Mai, P. *J. Phys. Chem. A* **2008**, 12393–12400.

# **Chapter 4 Hydrogen Bonding in Alkali Metal Cation-Bound Dimers of 1-Methyl Cytosine: An IRMPD Spectroscopic and Computational Study**

**This chapter has been published:**

R. Cheng, E. Loire and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2019, **21**, 11103–11110.

**Authors:**

Ruodi Cheng, Estelle Loire and Travis D. Fridgen\*

**Abstract:**

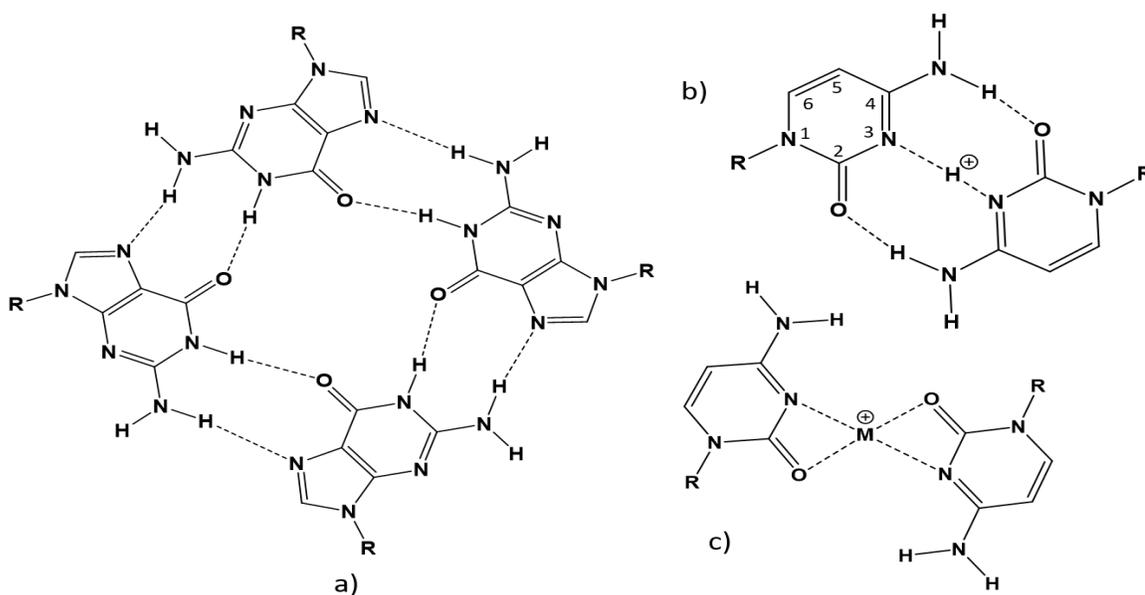
The structures of alkali metal cation bound 1-methylcytosine (1-mCyt) dimers were explored using vibrational spectroscopy in the form of infrared multiple photon dissociation (IRMPD) spectroscopy and by computational methods. For the smaller alkali metal cations,  $\text{Li}^+$  and  $\text{Na}^+$ , only non-hydrogen bonded symmetric anti-parallel structures were observed in agreement with the lowest energy computed structures. For  $\text{K}^+$ ,  $\text{Rb}^+$ , and  $\text{Cs}^+$  the vibrational spectra in the N-H stretch region showed strong evidence for hydrogen bonding in agreement with the lowest energy structures which contained hydrogen bonding interactions between the amine group of one cytosine and the carbonyl oxygen of the other cytosine. The lowest energy structures for these complexes were compared to previously studied cytosine complexes  $[(\text{Cyt})_2\text{M}]^+$  where  $\text{M} = \text{Li}, \text{Na}, \text{and K}$ . The calculations are in

agreement that only the non-hydrogen bonded structures would be observed for these cytosine complexes.

## 4.1 Introduction

Large biomolecules, such as DNA can twist themselves into rather interesting self-assembled structures that are largely governed by intramolecular interactions within the molecule itself. Their structures are also affected by solvent molecules, other biomolecules, as well as the presence of metal cations. As the crucial part of G-quadruplex formation, the interactions and stabilization of G-quadruplexes by monovalent metal cations have been studied<sup>1-4</sup> since the discovery of G-tetrads in the sequence of human telomere *in vitro* in the 1980s.<sup>2,5-8</sup> The structures of, and intramolecular interactions within, G-quadruplexes are formed by folding in single strands or multi-strands of guanine-rich regions of DNA/RNA.<sup>9-14</sup> Alkali metal cations were found to stabilize the structure of the 9-ethylguanine (9eG) tetrad where  $\text{Na}^+$  is greatly preferred over the other alkali metal cations in the solution phase.<sup>15</sup> A subsequent study conducted on gas-phase metal cationized tetrads<sup>16</sup> agreed with the stability ordering of the tetrads by alkali metal cations,  $\text{Na}^+ > \text{Li}^+ > \text{K}^+, \text{Rb}^+, \text{Cs}^+$  determined in the solution phase. Studies on G-quadruplexes have concluded that stronger hydration of  $\text{Na}^+$  compared to  $\text{K}^+$  was responsible for the fact that it is  $\text{K}^+$  that is associated with G-quadruplexes in biology.<sup>17</sup> However a study of gas-phase quadruplexes,  $(9\text{eG})_8\text{M}^+$ , composed of two tetrads bound by  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Rb}^+$ , or  $\text{Cs}^+$  revealed that even in the absence of solvent the quadruplex stabilized by  $\text{K}^+$  is significantly more strongly bound than  $(9\text{eG})_8\text{Na}^+$ , meaning that there is an intrinsic stability of the  $\text{K}^+$ -G-quadruplex<sup>18</sup> and shows the importance that gas-phase studies can have on the fundamental understanding of biomolecule interactions.

Similar to guanine, cytosine is a nucleobase in DNA and RNA which can form highly ordered structures. One such structure is called the i-motif which can be formed from four nucleic acid strands or by a single strand that can fold into four-stranded structures in cytosine rich regions.<sup>19–22</sup> While guanine quadruplexes are formed from tetrads, the i-motif is composed of pairs of cytosine protonated dimers; the guanine tetrad and protonated dimer of cytosine are shown in **Scheme 4.1 (a)** and **(b)**, respectively. The i-motif is not particularly stable under basic physiological pH due to the requirement of protonation.<sup>23</sup> At lower pH though, the i-motif is quite stable and is predicted to have dissociation energy of  $172 \text{ kJ mol}^{-1}$  and a  $6.7 \text{ kJ mol}^{-1}$  proton transfer barrier between the two cytosines.<sup>24</sup> The i-motif has been found in the regulatory regions and telomeres of human cells<sup>25</sup>, and it plays a key role in regulating gene expression *in vivo*.



**Scheme 4.1** Hydrogen bonding in a) the G-quadruplex and b) the i-motif. c) The reported symmetric anti-parallel (SAP) structure of the alkali metal cation bound dimer of cytosine,  $(\text{Cyt})_2\text{M}^+$ ,  $\text{M}=\text{Li}, \text{Na}, \text{and K}$ . The conventional ring numbering for cytosine is shown in b)

It has been shown<sup>26–28</sup> that methyl group substitution donates more electrons to certain sites of nucleobases which can affect the proton or electron affinity. For example, the proton affinity of the N3 and O2 sites of cytosine are predicted to be practically the same whereas methyl substitution at N1 enhances the proton affinity of the N3 site.<sup>29</sup> Infrared multiple photon dissociation (IRMPD) spectroscopy has been used to confirm the structure of the protonated dimers of both cytosine and 1-methylcytosine shown in **Scheme 4.1 (b)**.<sup>26,30</sup>

The effect of metal cations on i-motif DNA has been studied for several years and depends on metal cation concentration and solution temperature.<sup>31</sup> In their interactions with nucleobases, metal cations can destroy the normal hydrogen bond,<sup>32</sup> can induce stronger interactions,<sup>33</sup> and as well as tautomerization.<sup>34,35</sup> In fact, metal cations are considered as a main cause of mutations in DNA.<sup>36,37</sup> Uracil, for example, has been shown, both theoretically and experimentally, to be most stable as an imide tautomer when interacting with some metal cations.<sup>35,38–41</sup>

Besides the biological interest in i-motifs, it has been studied for its potential to develop materials, such as nanowires, DNA supermolecular hydrogels and DNA motors due to its sensitivity to pH.<sup>42–44</sup> In the gas phase, the alkali metal cations lithium, sodium, and potassium-bound dimer of cytosine have been studied and compared to the protonated, Cu<sup>+</sup>, and Ag<sup>+</sup> bound dimers.<sup>45</sup> The results show that the lowest energy structures for the three alkali metal cation-coordinated dimers of cytosine are tetra-dentate, with the metal cations bound to N3 and O2 of both cytosines, and are of planar symmetric anti-parallel (SAP) geometry as shown in **Scheme 4.1 (c)**. The Cu<sup>+</sup> bound dimer, on the other hand, is bi-coordinate—bound

to N3 of each cytosine and is similar in geometry to the protonated dimer of cytosine **Scheme 4.1 (b)**.

In this work, the alkali metal cationized 1-methylcytosine (1-mCyt) dimers,  $(1\text{-mCyt})_2\text{M}^+$  ( $\text{M}=\text{Li}, \text{Na}, \text{K}, \text{Rb}, \text{and Cs}$ ), were examined in the gas phase by infrared multiple photon dissociation (IRMPD) spectroscopy in both the  $1000 - 1800 \text{ cm}^{-1}$  and  $2800 - 3800 \text{ cm}^{-1}$  regions. The methylation of cytosine in the 1 position blocks the site attached to ribose/deoxyribose in DNA/RNA. Electronic structure calculations are compared to the gas-phase experimental spectra. A comparison of thermodynamic stability between  $(1\text{-mCyt})_2\text{M}^+$  structures is provided by computational methods.

## 4.2 Methods

**4.2.1 Computational Methods.** B3LYP density functional theory has been used to reliably model bioorganic and bioinorganic systems with hydrogen bonds successfully.<sup>46</sup> Geometry optimizations and frequency calculations were performed using the Gaussian 16 (G16) suite of programs<sup>47</sup> using B3LYP density functional theory and the 6-31+G(d,p) split-valence basis set for all atoms except Rb and Cs for which the Def2-SVP basis set and relativistic core potential was employed. Single point energy calculations for all optimized structures were performed using B3LYP/6-311+G(3df,3pd) on all atoms except for Rb and Cs for which Def2-TZVP was used. For all B3LYP calculations, an empirical dispersion correction was done using Grimme's D3 version with the original D3 damping function, B3LYPD3.<sup>48</sup>

M06-2X<sup>49,50</sup> is also considered to be highly reliable for the gas-phase ionic biomolecules with intramolecular interactions. Thus, calculations using the M06-2X method with the 6-31+G(d,p) (Def2-SVP on Rb and Cs) were also performed to compare with the B3LYP thermochemical results. The computed IR spectra were scaled by 0.97 and 0.945 in the lower (900-2000 cm<sup>-1</sup>) and higher (2800-3800 cm<sup>-1</sup>) energy regions, respectively<sup>51-53</sup>, and convoluted with a Lorentzian profile with a width (FWHM) of 15 cm<sup>-1</sup> to compare with the experimental IRMPD spectra. The relative 298 K Gibbs energies and enthalpies for structures were compared. Reported dissociation energies were calculated by the equations below and corrected for BSSE using the counterpoise correction in G16.

$$\Delta_{\text{diss}}H^{\circ} = \Delta H^{\circ} (\text{dimer}) - \Delta H^{\circ} (\text{monomer}) - \Delta H^{\circ}((\text{monomer})M^{+}) + \text{BSSE}$$

$$\Delta_{\text{diss}}G^{\circ} = \Delta G^{\circ} (\text{dimer}) - \Delta G^{\circ} (\text{monomer}) - \Delta G^{\circ}((\text{monomer})M^{+}) + \text{BSSE}$$

**4.2.2 Experimental Methods.** A Bruker Apex-Qe 7 Fourier-transform ion cyclotron resonance mass spectrometer (FT-ICR-MS) with an Apollo II electrospray ion source was used for all experimental work. To conduct IRMPD spectroscopy in the fingerprint region, the FT-ICR-MS was coupled to a free electron laser (FEL) at the Centre Laser Infrarouge d'Orsay (CLIO).<sup>54</sup> The ions of interest were first mass-selected and then irradiated for between 1 and 2 s. The FEL was scanned at 5 cm<sup>-1</sup> intervals from 900 cm<sup>-1</sup> to 2000 cm<sup>-1</sup>. In the 2800 – 3800 cm<sup>-1</sup> region, IRMPD experiments were conducted in the Laboratory for the Study of Energetics, Structures, and Reactions of Gaseous Ions at Memorial University,<sup>18</sup> where the isolated ions were irradiated with an optical parametric oscillator/amplifier (OPO/A,

LaserSpec, 1 W max power) for 2 s. The OPO/A was scanned from 3800 to 2800  $\text{cm}^{-1}$  with a 2  $\text{cm}^{-1}$  stepsize. The IRMPD efficiencies (*ie.* infrared intensities) are the negative of the logarithm of the precursor ion intensities divided by the sum of the precursor and product ion intensity.

1-mCyt and alkali metal chlorides were purchased from Sigma-Aldrich and used without further purification. The solution producing the 1-methylcytosine dimers was 1.3  $\text{mmol L}^{-1}$  in 1-methylcytosine in 50/50 water/methanol. In 1 mL of this solution 1-2 drops of 10  $\text{mmol L}^{-1}$  metal chloride solution was added. Solutions were injected using a 1 mL syringe at a flow rate of 0.15  $\text{mL h}^{-1}$ . The temperature of the dry gas was set at 200  $^{\circ}\text{C}$ .

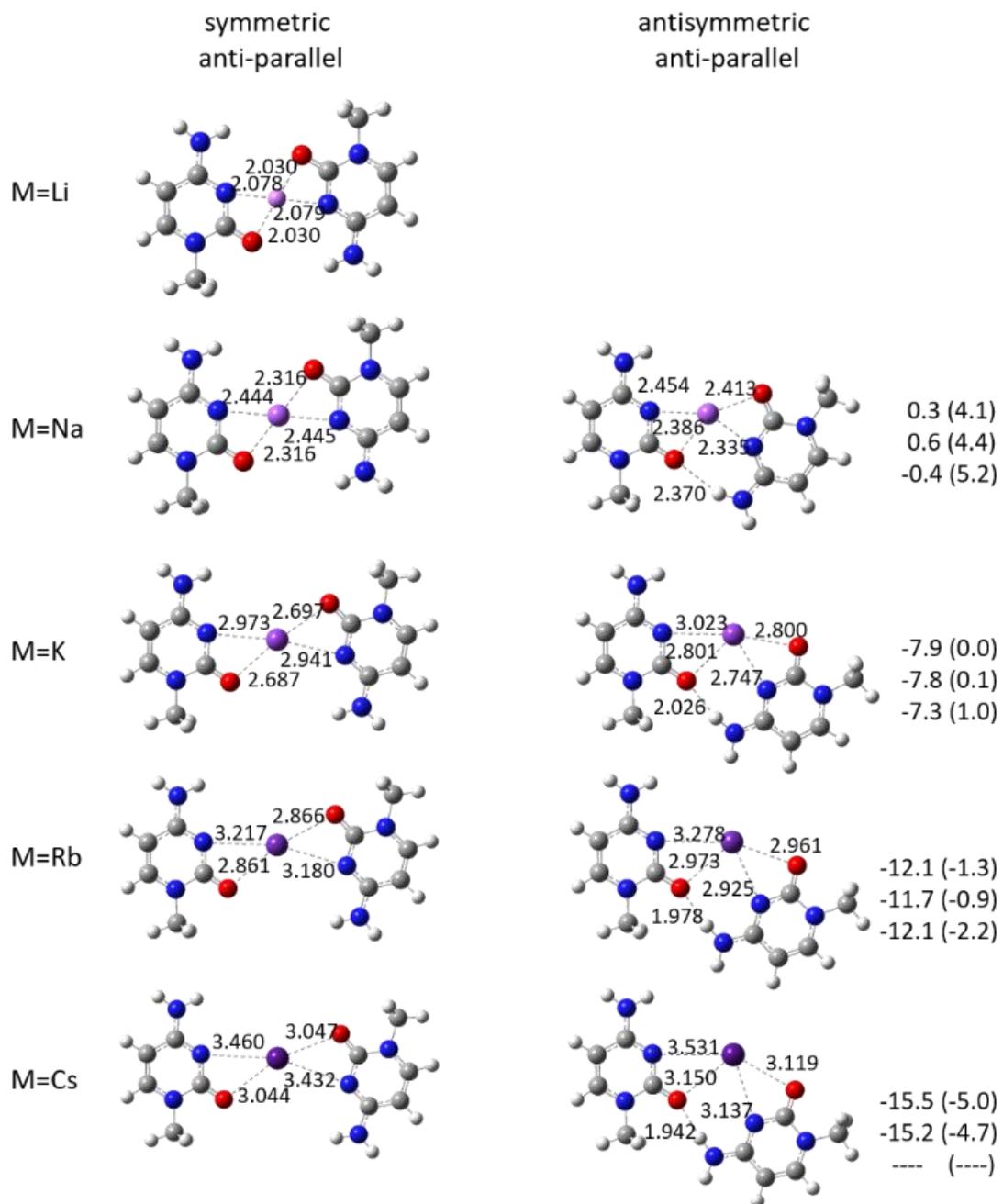
## 4.3 Results and Discussion

**4.3.1 Computed Structures for  $(1\text{-mCyt})_2\text{M}^+$ .** Previous studies have shown that the 1:1 complexes between alkali metal cations and cytosine exist as bidentate structures with the metal cation bound to the carbonyl oxygen and N3 of cytosine in its canonical (keto/amine) form.<sup>34,55</sup> The B3LYP/6-31+G(d,p) calculations on  $(1\text{-mCyt})_2\text{M}^+$  resulted in only one structure for  $\text{M}=\text{Li}$ , and two each for  $\text{M}=\text{Na}$ ,  $\text{K}$ ,  $\text{Rb}$ , and  $\text{Cs}$ . These structures are shown in **Figure 4.1** with the bond lengths of all electrostatic interactions. The first type of structure, and the only one for  $\text{M}=\text{Li}$ , is the symmetric anti-parallel (SAP) structure in the terminology used previously<sup>45</sup> and can be seen in **Figure 4.1**. The SAP structures for  $(1\text{-mCyt})_2\text{Li}^+$  and  $(1\text{-mCyt})_2\text{Na}^+$  are fairly symmetric (virtually  $\text{C}_{2h}$  symmetry), but those for  $(1\text{-mCyt})_2\text{K}^+$ ,  $(1\text{-$

mCyt)<sub>2</sub>Rb<sup>+</sup>, and (1-mCyt)<sub>2</sub>Cs<sup>+</sup> are distorted by a slight rotation about the metal cation; the term SAP will still be used for these complexes. In all cases, the SAP structures are tetradentate and, like the 1:1 alkali metal:cytosine complexes mentioned above, are bound to N3 and the carbonyl oxygen of both 1-mCyt. The second structure found for all complexes except (1-mCyt)<sub>2</sub>Li<sup>+</sup> is similar to the SAP structure except that they are significantly distorted, rotated around the metal cation, so as to allow for a hydrogen bond between the amine group of one 1-mCyt and the carbonyl of the other. This asymmetric anti-parallel (AAP) structure is also tetradentate. As might be expected, as the central ion increases in size and the ion-dipole interaction with cytosine weakens, the structures can distort more easily to form shorter (and stronger) ionic hydrogen bonds to help stabilize the structure.

Note that no other structures involving the keto/amine tautomer were found. Attempts to optimize a non-planar, or planar parallel structure resulted in optimization to the AAP structures.

The relative 298 K enthalpies and Gibbs energies of the complexes are also compared in **Figure 4.1** using three different computational methods: B3LYPD3/6-31+G(d,p) (top), B3LYPD3/ 6-311+G(3df,3pd)//B3LYPD3/6-31+G(d,p) (middle), and M06-2X/ 6-31+G(d,p) (bottom). For (1-mCyt)<sub>2</sub>Li<sup>+</sup> and (1-mCyt)<sub>2</sub>Na<sup>+</sup>, the SAP structures are the lowest in Gibbs energy and would be expected to be observed, in fact the AAP conformer is not expected to exist, based on this theory, for the Li<sup>+</sup> complex. For (1-mCyt)<sub>2</sub>K<sup>+</sup>, the SAP and AAP structures are basically isoenergetic in terms of Gibbs energy. For (1-mCyt)<sub>2</sub>Rb<sup>+</sup> and (1-mCyt)<sub>2</sub>Cs<sup>+</sup>, the AAP structures are favoured. As the metal cation increases in size, the ion-cytosine interaction—



**Figure 4.1** B3LYP/6-31+G(d,p) computed structures for the SAP and AAP structures for  $(1\text{-mCyt})_2\text{M}^+$  complexes. The 298 K enthalpies and Gibbs energies (parentheses) of the AAP structures relative to the SAP structures are also provided at the B3LYPD3/6-31+G(d,p) level (top), B3LYPD3/6-311+G(3df, 3pd)//B3LYPD3/6-31+G(d,p) (middle), and M06-2X/6-31+G(d,p) level (bottom),  $\text{kJ mol}^{-1}$ .

which is predominantly an ion-dipole interaction—decreases in strength and the structure- stabilizing hydrogen bonding interaction becomes favoured. Note that for the M06-2X/6-31+G(d,p) calculations, the SAP structure of  $(1\text{-mCyt})_2\text{Cs}^+$  minimized to the transition state with a small imaginary frequency. All attempts to get rid of the imaginary frequency were unsuccessful.

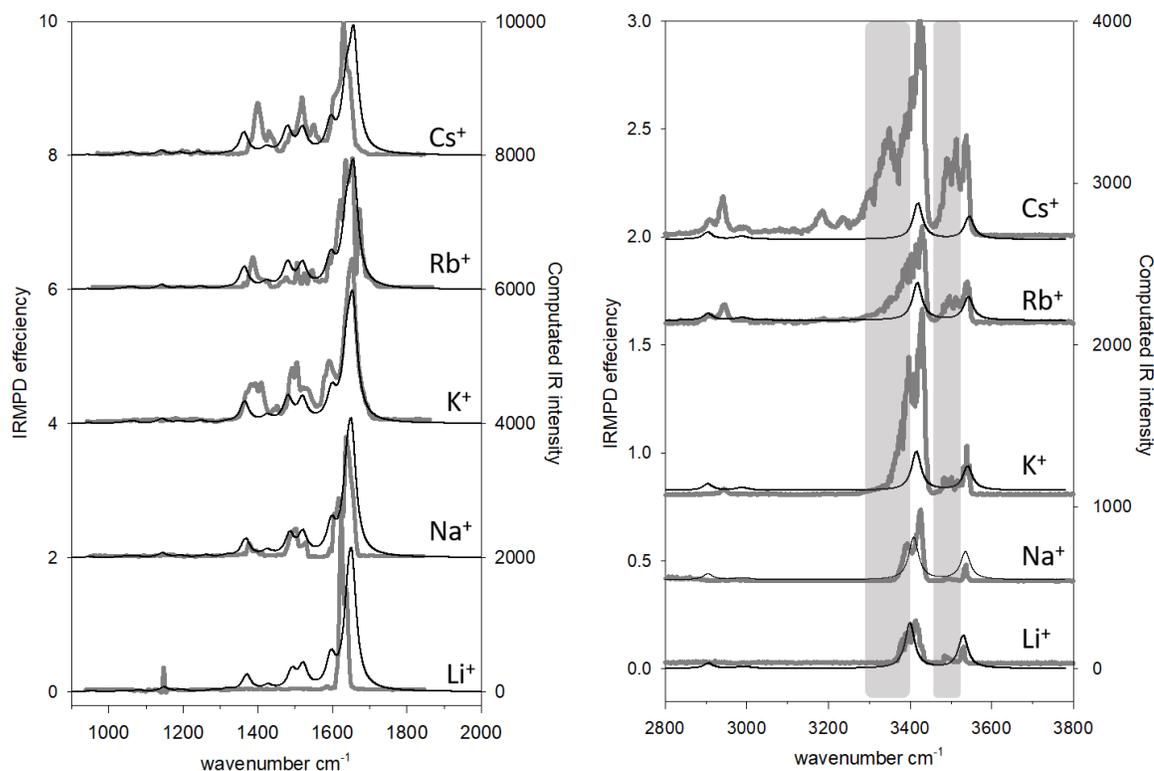
#### 4.3.2 Vibrational (IRMPD) Spectra of $(1\text{-mCyt})_2\text{M}^+$

Upon absorption of multiple infrared photons resonant with a vibrational mode, all  $(1\text{-mCyt})_2\text{M}^+$  complexes fragmented by losing a neutral 1-mCyt.

The experimental spectra are compared with the computed IR spectra for the SAP structures in **Figure 4.2**. The experimental spectra for all  $(1\text{-mCyt})_2\text{M}^+$  in the fingerprint region are all quite similar exhibiting metal-coordinated C=O stretching and NH<sub>2</sub> bending between 1600 and 1700 cm<sup>-1</sup> as well as complex C-C/C-N stretching of the ring and the appendages centred at about 1400 and 1500 cm<sup>-1</sup>. The computed IR spectra for the SAP structures agree very well with the experimental spectra in the fingerprint region. It is worth mentioning that due to the stronger bonding for  $(1\text{-mCyt})_2\text{Li}^+$ , the photofragment yield is significantly reduced that for the other complexes resulting only one observed band just above 1600 cm<sup>-1</sup>.

In the higher energy region, bands at 3540 and 3430 cm<sup>-1</sup> can be ascribed to the NH<sub>2</sub> asymmetric and symmetric stretching modes, respectively, and the computed spectra for the SAP structures reproduce these observed features. On the low energy side of these bands, there are also very weak absorptions or shoulders observed for the  $(1\text{-mCyt})_2\text{Li}^+$  and  $(1\text{-mCyt})_2\text{Na}^+$  that broaden and grow to

prominence for the  $(1\text{-mCyt})_2\text{K}^+$ ,  $(1\text{-mCyt})_2\text{Rb}^+$ , and  $(1\text{-mCyt})_2\text{Cs}^+$  complexes. It should be noted that a splitting of the asymmetric absorption was also observed for the  $(\text{Cyt})_2\text{M}^+$  ( $\text{M}=\text{Li}, \text{Na}, \text{K}$ ) complexes.<sup>45</sup>



**Figure 4.2** Experimental IRMPD spectra of  $(1\text{-mCyt})_2\text{M}^+$  complexes, where  $\text{M}=\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Rb}^+$ ,  $\text{Cs}^+$  (grey traces) and computed IR spectra for the SAP complexes (black traces).

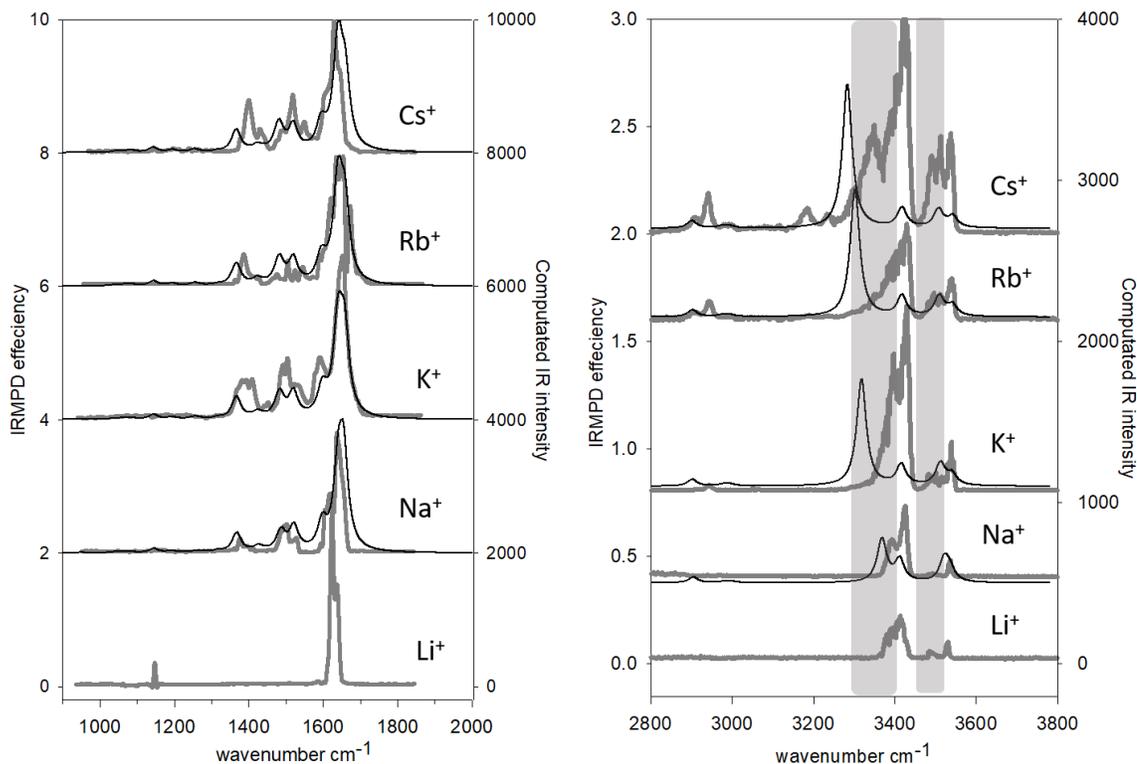
Since there are two  $\text{NH}_2$  groups, one for each 1-mCyt, there are in phase and out of phase modes for each of the symmetric and asymmetric stretches. These might be considered the carriers for the shoulders observed—at least the weak ones for  $(1\text{-mCyt})_2\text{Li}^+$  and  $(1\text{-mCyt})_2\text{Na}^+$ —since none of the complexes are completely

symmetric. However, these bands are computed to be nearly degenerate, ranging from as little as  $0.2 \text{ cm}^{-1}$  to as much as  $1.2 \text{ cm}^{-1}$  difference for the more asymmetric SAP complexes of  $(1\text{-mCyt})_2\text{Cs}^+$ .

In **Figure 4.3**, the experimental spectra are compared to the computed IR spectra for the AAP structures of the  $(1\text{-mCyt})_2\text{M}^+$  complexes. The predicted spectra for the fingerprint region are virtually identical for the SAP (**Figure 4.2**) and AAP (**Figure 4.3**) isomers which are both in very good agreement with the experimental spectra. The similarity in the position of the C=O stretching band for the SAP and AAP structures is surprising since there is a significant hydrogen bond in the AAP structures which would lead one to expect a red shift. However, concurrent with the addition of the hydrogen bonding interaction is a lengthening, and presumably weakening, of the C=O--M<sup>+</sup> bond by approximately 0.1 angstrom. This weakening of the C=O--M<sup>+</sup> bond would counteract any expected redshift due to the additional hydrogen bond.

In the higher energy region, due to the hydrogen bonding interactions, the computed in phase and out of phase NH<sub>2</sub> stretching modes are no longer degenerate. The hydrogen bonded symmetric and asymmetric NH<sub>2</sub> stretches are shifted significantly to the red. These hydrogen bonded NH<sub>2</sub> stretching vibrations of the AAP complexes explain the observed low energy shoulders that broaden and redshift as the metal gets larger, and the hydrogen bonding interaction gets stronger. That the computed hydrogen bonded NH<sub>2</sub> stretching vibrations do not accurately reproduce the experimental spectra is due to the anharmonicity of the hydrogen bonded N-H

stretching and the harmonic nature of the calculations; this has been well documented.<sup>56–61</sup>



**Figure 4.3** Experimental IRMPD spectra of  $(1\text{-mCyt})_2\text{M}^+$  complexes, where  $\text{M} = \text{Li}^+, \text{Na}^+, \text{K}^+, \text{Rb}^+, \text{Cs}^+$  (grey traces) and computed IR spectra for the AAP complexes (black traces).

The computed relative energetics provided in **Figure 4.1** (computed dissociation energies are provided in **Table 4.1**) are also in agreement with the experimentally observed spectra. Based on the computed energies, we would expect to predominantly observe the SAP complexes for  $(1\text{-mCyt})_2\text{Li}^+$  and  $(1\text{-mCyt})_2\text{Na}^+$ , whereas we would expect to see both SAP and AAP structures for  $(1\text{-mCyt})_2\text{K}^+$ . For

(1-mCyt)<sub>2</sub>Rb<sup>+</sup> and (1-mCyt)<sub>2</sub>Cs<sup>+</sup> we would expect to see predominantly the hydrogen bonded AAP structures, although we cannot rule out the existence of the SAP structures neither spectroscopically nor based on the computed energies. However, it is clear that the hydrogen bonded structures are a significant presence for M=K, Rb, and Cs, but not for M=Li and Na.

**Table.4.1** Computed 298 K energetics of dissociation (B3LYP/6-311+G(3df,3pd)//B3LYP/6-31+G(d,p)) for the (1-mCyt)<sub>2</sub>M<sup>+</sup> complexes. Counterpoise corrections for BSSE are included.

Complex	$\Delta_{\text{diss}}H^{\circ} / \text{kJ mol}^{-1}$	$\Delta_{\text{diss}}G^{\circ} / \text{kJ mol}^{-1}$
SAP_Li	184.2	144.0
SAP_Na	159.3	122.2
AAP_Na	158.5	117.6
SAP_K	123.3	86.6
AAP_K	130.7	86.2
SAP_Rb	112.4	80.9
AAP_Rb	123.4	81.1
SAP_Cs	99.7	69.3
AAP_Cs	114.6	73.7

### 4.3.3 Comparison of (1-mCyt)<sub>2</sub>M<sup>+</sup> and (Cyt)<sub>2</sub>M<sup>+</sup>

In the study by *Gao et al.*<sup>45</sup> on (Cyt)<sub>2</sub>M<sup>+</sup> where M=Li, Na and K, it was concluded that only the SAP structures exist as there was no evidence for hydrogen bonded structures. This contrasts the results here that show the (1-mCyt)<sub>2</sub>M<sup>+</sup> complexes where M=K, Rb, and Cs that clearly show hydrogen bonding AAP structures. Computations on the SAP and AAP structures of (Cyt)<sub>2</sub>M<sup>+</sup> (M= Na, K, Rb and Cs) were performed and the results are in **Table 4.2**.

Like for the (1-mCyt)<sub>2</sub>Li<sup>+</sup> complex, only the SAP structure was optimized and attempts to find an AAP structure for (Cyt)<sub>2</sub>Li<sup>+</sup> were unsuccessful. For the rest of the complexes, the B3LYP calculations predict that only for (Cyt)<sub>2</sub>Cs<sup>+</sup> is the AAP structure slightly predominant in terms of Gibbs energy, while for (Cyt)<sub>2</sub>K<sup>+</sup> and (Cyt)<sub>2</sub>Rb<sup>+</sup>, the SAP structure is predicted to be lower in energy, but only slightly. For the M06-2X calculations, the AAP structures are lower in energy for both (Cyt)<sub>2</sub>Rb<sup>+</sup> and (Cyt)<sub>2</sub>Cs<sup>+</sup>. These calculations are consistent with the observations that no hydrogen bonded structures exist for (Cyt)<sub>2</sub>Li<sup>+</sup>, (Cyt)<sub>2</sub>Na<sup>+</sup>, and (Cyt)<sub>2</sub>K<sup>+</sup>. It should be noted, however, that for the (Cyt)<sub>2</sub>K<sup>+</sup> IRMPD spectra,<sup>45</sup> there are diffuse shoulders observed that extend to lower energy from both the symmetric and asymmetric stretching bands compared to the (Cyt)<sub>2</sub>Li<sup>+</sup> and (Cyt)<sub>2</sub>Na<sup>+</sup> IRMPD spectra.

**Table 4.2.** The 298 K relative enthalpies and Gibbs energies, in kJ mol<sup>-1</sup>, of the cytosine dimeric complexes by three different computational methods.

	B3LYPD3/ 6-31+G(d,p)		B3LYPD3/ 6-311+G(3df,3pd)		M06-2X/ 6-31+g(d,p)	
	$\Delta_{rel}H^{\circ}$	$\Delta_{rel}G^{\circ}$	$\Delta_{rel}H^{\circ}$	$\Delta_{rel}G^{\circ}$	$\Delta_{rel}H^{\circ}$	$\Delta_{rel}G^{\circ}$
SAP_Na	0.0	0.0	0.0	0.0	0.0	0.0
AAP_Na	0.1	5.6	0.8	6.4	-0.9	5.0
SAP_K	0.0	0.0	0.0	0.0	0.0	0.0
AAP_K	-7.7	1.6	-7.4	2.2	-8.0	2.4
SAP_Rb	0.0	0.0	0.0	0.0	0.0	0.0
AAP_Rb	-11.9	0.3	-11.0	1.2	-12.6	-7.3
SAP_Cs	0.0	0.0	0.0	0.0	0.0	0.0
AAP_Cs	-14.9	-2.8	-14.0	-1.9	-15.9	-8.6

#### 4.4 Conclusions

IRMPD spectroscopy in the fingerprint and the higher energy N-H stretching region was used to probe the structures of alkali metal cation bound 1-methylcytosine (1-mCyt) dimers. A comparison of these spectra with computed IR spectra showed that for (1-mCyt)<sub>2</sub>K<sup>+</sup>, (1-mCyt)<sub>2</sub>Rb<sup>+</sup>, and (1-mCyt)<sub>2</sub>Cs<sup>+</sup>, broad

shoulders extend from the low energy side of the symmetric and asymmetric  $\text{NH}_2$  stretching bands. This broadening is explained as being attributed to hydrogen bonding between the amine group of one cytosine and the carbonyl oxygen of the other cytosine. The IRMPD spectra compare very well with the computed IR spectra and the observed results are consistent with the computed energetics between the hydrogen bonded and non-hydrogen bonded complexes. The lowest energy structures for the 1-mCyt complexes were compared to previously studied cytosine complexes  $(\text{Cyt})_2\text{M}^+$  where  $\text{M} = \text{Li}, \text{Na}, \text{and K}$ . The calculated energetics are in agreement that only the non-hydrogen bonded structures would be observed for these cytosine complexes. However, the IRMPD spectrum for  $(\text{Cyt})_2\text{K}^+$ , does display broadened N-H bands extending to lower energy from both the symmetric and asymmetric stretching bands compared to the  $(\text{Cyt})_2\text{Li}^+$  and  $(\text{Cyt})_2\text{Na}^+$  IRMPD spectra;<sup>45</sup> this could indicate that the hydrogen bonding complexes (AAP structures) do exist for  $(\text{Cyt})_2\text{K}^+$  as they do here for  $(1\text{-mCyt})_2\text{K}^+$ .

### **Acknowledgements**

The CLIO staff are gratefully acknowledged for their help in obtaining IRMPD spectra in the fingerprint region. The authors acknowledge the computational resources provided by Compute Canada for these studies. The authors also wish to thank the financial contributions from NSERC, CFI, and Memorial University.

### **Conflicts of interest**

There are no conflicts to declare.

## References

- 1 J. R. Williamson, M. K. Raghuraman and T. R. Cech, *Cell*, 1989, **59**, 871–880.
- 2 D. Bhattacharyya, G. Mirihana Arachchilage and S. Basu, *Front. Chem.*, 2016, **4**, 38.
- 3 J. T. Davis, *Angew. Chemie - Int. Ed.*, 2004, **43**, 668–698.
- 4 S. N. Georgiades, N. H. A. Karim, K. Suntharalingam and Ramon Vilar, *Angew. Chemie. Int. Ed.*, 2010, **49**, 4020–4034.
- 5 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263–282.
- 6 C. W. Greider and E. H. Blackburn, *Cell*, 1985, **43**, 405–413.
- 7 M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770–780.
- 8 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402–5415.
- 9 Y. Wang and D. J. Patel, *Biochemistry*, 1992, **31**, 8112–8119.
- 10 B. Lippert, *J. Chem. Soc., Dalt. Trans.*, 1997, 3971–3976.
- 11 W. I. Sundquist and A. Klug, *Lett. to Nat.*, 1989, **342**, 825–829.
- 12 J. R. Williamson, *Annu. Rev. Biophys. Biomol. Struct.*, 1994, **23**, 703–730.
- 13 F. W. Smith and J. Feigon, *Nature*, 1992, **356**, 164–168.
- 14 R. Hänsel-Hertsch, M. Di Antonio and S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 279–284.
- 15 K. Fukushima and H. Iwahashi, *Chem. Commun.*, 2000, 895–896.
- 16 M. Azargun and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25778–25785.

- 17 F. Zaccaria, G. Paragi and C. Fonseca Guerra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20895–20904.
- 18 M. Azargun, Y. Jami-Alahmadi and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2017, **19**, 1281–1287.
- 19 J. Leroy, M. Gueron and K. Gehring, *Letters to Nature*, 1993, **363**.
- 20 C. Kang, I. Berger, C. Lockshin, R. Ratliff, R. Moyzis, C. Kang, I. Berger, C. Lockshin, R. Ratliff, R. Moyzist and A. Rich, 1994, **91**, 11636–11640.
- 21 M. Guéron and J. L. Leroy, *Curr. Opin. Struct. Biol.*, 2000, **10**, 326–331.
- 22 A. T. Phan and J.-L. Mergny, *Nucleic Acids Res.*, 2002, **30**, 4618–4625.
- 23 S. Benabou, A. Aviñó, R. Eritja, C. González and R. Gargallo, *RSC Adv.*, 2014, **4**, 26956–26980.
- 24 S. Y. Han and H. Bin Oh, *Chem. Phys. Lett.*, 2006, **432**, 269–274.
- 25 M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger and D. Christ, *Nat. Chem.*, 2018, **10**, 631–637.
- 26 B. Yang, A. R. Moehlig, C. E. Frieler and M. T. Rodgers, *J. Phys. Chem. A*, 2015, **119**, 1857–1868.
- 27 C. Desfrancois, H. Abdoul-Carime, V. P. S. Carles, J. P. Schermann, D. M. A. Smith and L. Adamowicz, *J. Chem. Phys.*, 1999, **110**, 11876.
- 28 M. Etinski and C. M. Marian, *Phys. Chem. Chem. Phys.*, 2009, **12**, 4915–4923.
- 29 M. Liu, T. Li, F. S. Amegayibor, D. S. Cardoso, Y. Fu and J. K. Lee, *J. Org. Chem.*, 2008, **73**, 9283–9291.
- 30 J. Oomens, A. R. Moehlig and T. H. Morton, *J. Phys. Chem. Lett.*, 2010, **1**, 2891–2897.
- 31 A. Ono, H. Torigoe, Y. Tanaka and I. Okamoto, *Chem. Soc. Rev.*, 2011, **40**, 5855–

5866.

- 32 S. Menzer, M. Sabat and B. Lippert, *J. Am. Chem. Soc.*, 1992, **114**, 4644–4649.
- 33 M. H. Shamsi and H. B. Kraatz, *J. Inorg. Organomet. Polym. Mater.*, 2013, **23**, 4–23.
- 34 M. Kabeláč and P. Hobza, *J. Phys. Chem. B*, 2006, **110**, 14515–14523.
- 35 R. Cheng, V. E. Rose, B. Power and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2017, **20**, 572–580.
- 36 G. L. Eichhorn, N. A. Berger, J. J. Butzow, P. Clark, J. Heim, J. Pitha, C. Richardson, J. M. Rifkind, Y. Shin and E. Tarien, in *Metal Ions in Biological Systems*, ed. S. K. Dhar, Springer US, Boston, MA, 1973, pp. 43–66.
- 37 B. Lippert, *J. Chem. Soc. Dalt. Trans.*, 1997, **21**, 3971–3976.
- 38 E. a. L. Gillis, M. Demireva, K. Nanda, G. Beran, E. R. Williams and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 3304.
- 39 K. J. Koch, T. Aggerholm, S. C. Nanita and R. G. Cooks, *J. Mass Spectrom.*, 2002, **37**, 676–686.
- 40 J. M. Bakker, R. K. Sinha, T. Besson, M. Brugnara, P. Tosi, J. Y. Salpin and P. Maître, *J. Phys. Chem. A*, 2008, **112**, 12393–12400.
- 41 E. L. Zins, S. Rochut and C. Pepe, *J. Mass Spectrom.*, 2009, **44**, 40–49.
- 42 P. Alberti, A. Bourdoncle, B. Saccà, L. Lacroix and J. L. Mergny, *Org. Biomol. Chem.*, 2006, **4**, 3383–3391.
- 43 D. Liu and S. Balasubramanian, *Angew. Chemie - Int. Ed.*, 2003, **42**, 5734–5736.
- 44 Y. Dong, Z. Yang and D. Liu, *Acc. Chem. Res.*, 2014, **47**, 1853–1860.
- 45 J. Gao, G. Berden, M. T. Rodgers and J. Oomens, *Phys. Chem. Chem. Phys.*, 2016,

- 18**, 7269–7277.
- 46 M. Meyer, T. Steinke, M. Brandl, J. Sühnel and M. E. T. Al, 2000, **22**, 109–124.
- 47 M. H. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, *Gaussian, Inc. Wallingford, CT*.
- 48 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 49 M. Walker, A. J. A. Harvey, A. Sen and C. E. H. Dessent, *J. Phys. Chem. A*, 2013, **117**, 12590–12600.
- 50 E. G. Hohenstein, S. T. Chill and C. D. Sherrill, *Methods*, 2008, **4**, 1996–2000.
- 51 Y. Jami-Alahmadi and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2016, **18**, 2023–2033.
- 52 B. Power, V. Haldys, J. Y. Salpin and T. D. Fridgen, *Int. J. Mass Spectrom.*, 2015, **378**, 328–335.
- 53 R. Cheng, V. E. Rose, B. Power and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2018, **20**, 572–580.
- 54 R. Prazeres, F. Glotin, C. Insa, D. A. Jaroszynski and J. M. Ortega, *Eur. Phys. J. D*, 1998, **3**, 87–93.
- 55 B. Yang, R. R. Wu, N. C. Polfer, G. Berden, J. Oomens and M. T. Rodgers, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 1523–1533.
- 56 E. A. L. Gillis and T. D. Fridgen, *Int. J. Mass Spectrom.*, 2010, **297**, 2–8.
- 57 A. Cimas, T. D. Vaden, T. S. J. A. De Boer, L. C. Snoek and M. Gaigeot, *J. Chem. Theory Comput.*, 2009, **5**, 1068–1078.
- 58 C. Fraschetti, L. Guarcini, M. Speranza and A. Filippi, *Int. J. Mass Spectrom.*, 2019,

**438**, 148–156.

- 59 B. D. Linford, A. Le Donne, D. Scuderi, E. Bodo and T. D. Fridgen, *Eur. J. Mass Spectrom.*, 2018, **0**, 1–9.
- 60 M. B. Moghaddam and T. D. Fridgen, *J. Phys. Chem. B*, 2013, **117**, 6157–6164.
- 61 M. B. Burt and T. D. Fridgen, *J. Phys. Chem. A*, 2013, **117**, 1283–1290.

# **Chapter 5 A vibrational spectroscopic and computational study of gaseous protonated and alkali metal cationized G-C base pairs**

**This chapter has been submitted for publication as:**

R. Cheng, J. Martens, and T. D. Fridgen, *Phys. Chem. Chem. Phys.* (Manuscript ID: CP-ART-11-2019-005962)

**Authors:**

Ruodi Cheng, Jonathan Martens and Travis D. Fridgen\*

**See Appendix B for additional results.**

**Abstract:**

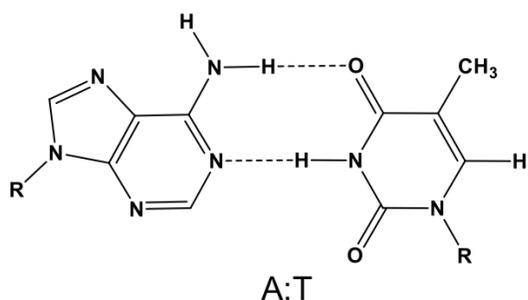
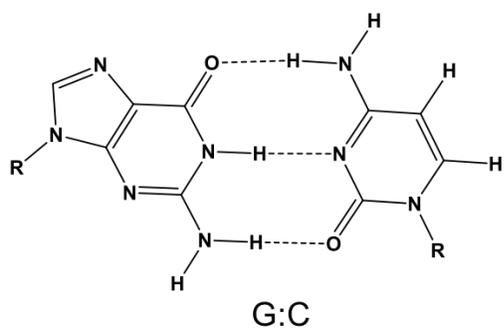
The structures and properties of metal cationized complexes of 9-ethylguanine (9eG) and 1-methylcytosine (1mC), (9eG:1mC) $M^+$ , where  $M^+ = Li^+, Na^+, K^+, Rb^+, Cs^+$  as well as the protonated complex, (9eG:1mC) $H^+$ , have been studied using a combination of IRMPD spectroscopy and computational methods. For (9eG:1mC) $H^+$ , the dominant structure is a Hoogsteen type complex with the proton covalently bound to N3 of 1mC despite this being the third best protonation site of the two bases; based on proton affinities N7 of 9eG should be protonated. However, this structural oddity can be explained considering both the number of hydrogen bonds that can be formed when N3 of 1mC is protonated as well as

the strong ion-induced dipole interaction that exists between an N3 protonated 1mC and 9eG due to the higher polarizability of 9eG. The anomalous dissociation of  $(9eG:1mC)H^+$ , forming much more  $(1mC)H^+$  than would be predicted based on the computed thermochemistry, can be explained as being due to the structural oddity of the protonation site and that the barrier to proton transfer from N3 of 1mC to N7 of 9eG grows dramatically as the base pair begins to dissociate. For the  $(9eG:1mC)M^+$ ;  $M= Li^+, Na^+, K^+, Rb^+, Cs^+$  complexes, single unique structures could not be assigned. However, the experimental spectra were consistent with the computed spectra. For  $(9eG:1mC)Li^+$ , the lowest energy structure is one in which  $Li^+$  is bound to O6 of 9eG and both O2 and N3 of 1mC; there is also an interbase hydrogen bond from the amine of 1mC to N7 of 9eG. For  $Na^+, K^+$ , and  $Rb^+$ , similar binding of the metal cation to 1mC is calculated but, unlike  $Li^+$ , the lowest energy structure is one in which the metal cation is bound to N7 of 9eG; there is also an interbase hydrogen bond between the amine of 1mC and the carbonyl of 9eG. The lowest energy structure for the Cs complex is the Watson-Crick type base pairing with  $Cs^+$  binding only to 9eG through O6 and N7 and with three hydrogen bonds between 9eG and 1mC. It is also interesting to note that the Watson-Crick base pairing structure gets lower in Gibbs energy relative to the lowest energy complexes as the metal gets larger. This indicates that the smaller, more densely charged cations have a greater propensity to interfere with Watson-Crick base pairing than do the larger, less densely charged metal cations.

## 5.1 Introduction.

Thousands of research papers on the topic of non-covalent interactions between base pairs have been published following the discovery of Watson-Crick base pairing patterns,<sup>1</sup> A:T, G:C in DNA and A:U, G:C in RNA. Arguably, this was one of the most influential biochemical discoveries of the twentieth century. The sequence of base pairs is nature's method of storing genetic information<sup>2</sup> and guiding protein synthesis.<sup>3</sup> The stability of base pair interactions can be affected by the solvent as well as the presence of metal cations. When solvated, both the common conformations of DNA, B-DNA as well as zigzag or Z-DNA, show disordering at A:T base pairs, but not G:C pairs<sup>4</sup> due to the extra stability of the latter which contains an extra hydrogen bond, three vs two for the A:T base pair as can be seen in **Scheme 5.1**.

Interest in the study of metal ion/nucleobase interactions stems from the fact that metal cations strongly affect the structure of base-pairs and, therefore, the DNA duplex which can result in the mistranslation of genetic information.<sup>5,6</sup> However, the effect of metal cations on these biological molecules is not completely understood. Metal ions in low concentration are thought to stabilize the DNA duplex by binding to the phosphate backbone which decreases the repulsive forces between the negative charges on the nucleotides.<sup>7</sup> However, in higher concentrations, the metal cations can destroy the hydrogen bonds between Watson-Crick base pairs, and even cause the tautomerization of nucleobases.<sup>8</sup> There is an ever-increasing amount of research attempting to uncover the intrinsic physical and chemical properties of metal-bonded nucleobase complexes by both experimental and/or computational methods in both the solution and gas phase.<sup>5-7,9-25</sup>



**Scheme 5.1** Watson-Crick G:C and A:T base pairs

The biological role of G:C base pairs with metal cations has captured the attention of scientists for numerous reasons. For example, there have been many studies concerning the G:C base pair coordinated with Pt-containing complexes following the 1969 discovery of cisplatin.<sup>26–29</sup> Cisplatin is an anti-cancer chemotherapy drug; it kills cancer by damaging DNA molecules and inhibiting DNA synthesis. Interestingly, cisplatin shows a kinetic preference to interact with G:C rich regions in transfer RNA (tRNA<sup>Ala</sup>), and one of its acceptors, Microhelix<sup>Ala</sup> (Mh<sup>Ala</sup>).<sup>30</sup> It also potentially induces a permanent mutation by stabilizing rare tautomers of guanine and/or cytosine in DNA.<sup>31</sup> Besides the studies of G:C base pairs in biological systems, the potential applications for DNA-based materials have also been investigated. A recent report was published on the study of crosslinked supramolecular hydrogels designed by utilizing the hydrogen bonds between the G:C base

pair as a junction in the polymer structures, and its ability to self-assemble to host a drug molecule.<sup>32</sup> The G:C base pair was shown to have improved rheological activity *in vitro*, making it possible to establish faster drug release when it arrives at the cancerous tissue in organisms.

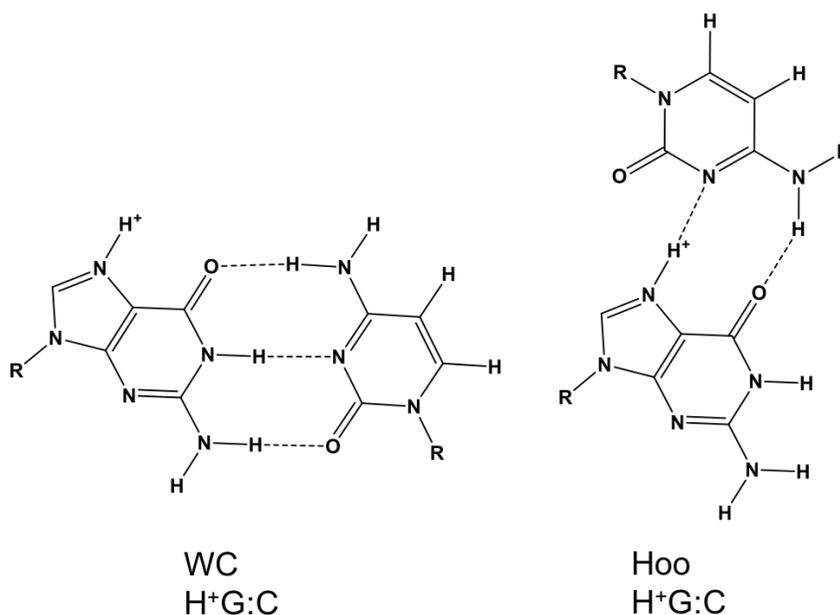
Research on alkali metal cationized and proton bound nucleobase complexes is considerable, not only due to these cations' abundances in nature, but also because they are essential elements for living organisms.<sup>13,22,25,33-37</sup> Gas-phase studies of alkali metal cations with nucleobases proved that the self-aggregation of nucleobases were promoted by the existence of alkali metal cations and can form magic number clusters.<sup>11,22</sup> For uracil and thymine,  $\text{Li}^+$  can stabilize trimers and  $\text{Na}^+$  can stabilize tetramers, whereas other larger alkali metal cations prefer to stabilize pentamers. Guanine quartets/quadruplexes bound to alkali metal cations have received much attention as alkali metal cations locate at the center of the quartets or hold the quadruplex, greatly stabilizing the structures,<sup>6,9,22,35,38-41</sup> similar to quadruplexes formed in oligonucleotides.<sup>42-44</sup> Similar to studies conducted in solution,<sup>45</sup> gas-phase studies showed that  $\text{Na}^+$ -bound G-tetrads are more stable than  $\text{Li}^+$ -bound tetrads because of weaker hydrogen bonds caused by an almost tetrahedral distortion of the tetrad, offsetting its ion-dipole interaction advantage.<sup>25</sup> However,  $\text{K}^+$  is the alkali metal cation most often observed to stabilize the G-quadruplexes in telomeres or other guanine-rich regions of DNA. Recently, this was revealed to be due to the greater intrinsic thermochemical stability of  $\text{K}^+$ -associated quadruplexes over any other alkali metal cation, including  $\text{Na}^+$ , due to the  $\text{K}^+$ -tetrad complex being non-planar and its ability to accept another G-tetrad to form the sandwiched structure of the G-quadruplex.<sup>9</sup>

Cytosine self-assemblies prefer to form tetra-dentate dimeric structures with the alkali metal cations bound to O2 and N3 of both cytosines. This is in contrast to the bi-coordinated i-motif structure where the proton binds to N3 of both cytosines.<sup>13,17,46</sup>

Gas-phase protonated G-C base pairs were studied previously by infrared multiphoton dissociation (IRMPD) spectroscopy in the fingerprint region and by computational methods.<sup>15,18,47,48</sup> It was concluded that the protonated G:C base pair existed as one of two different isomers depending on the pH value of the solution from which it was electrosprayed. At pH 5.8, Watson-Crick base pairing dominated whereas at pH 3.2, the Hoogsteen base pair was observed as the major structure.<sup>15</sup> The structures of protonated Watson-Crick and Hoogsteen G:C base pairs are depicted in **Scheme 5.2**. In an interesting set of studies using collision induced dissociation (CID) and computational methods, the dissociation of protonated Hoogsteen guanine:cytosine (G:C) base pairs was found to yield predominantly protonated cytosine and neutral guanine, despite guanine having a reportedly higher proton affinity.<sup>18,49</sup> This *anomalous* dissociation pattern was initially explained to be due to two routes to protonated cytosine, one leading to N3-protonated cytosine, the other leading to O2-protonated cytosine effectively doubling the rate constant for protonated cytosine formation.<sup>49</sup> The branching ratios for CID of protonated 1-methylguanine:1-methylcytosine (1mG:1mC) and protonated 9-methylguanine:1-methylcytosine (9mG:1mC) Hoogsteen base pairs were compared, and while they both show this anomalous behaviour, the latter shows significantly more protonated 9-methylguanine product than 1-methylcytosine. High level calculations showed that O2 protonated 1-methylcytosine actually has a slightly higher proton affinity than 1-methylguanine, providing a more reasonable explanation for the anomalous CID behaviour;

O2 protonated 1-methylcytosine is the lowest energy product. 9-methylguanine, however, has a higher proton affinity and is the lowest energy product for dissociation of the protonated 1-methylcytosine:9-methylguanine base pair, yet not in a ratio that would be predicted based on the differences in proton affinities.<sup>18</sup>

The goal of the present work is to discover the intrinsic structures and unimolecular dissociation properties of the protonated or alkali metal cationized 9-ethylguanine:1-methylcytosine (9eG:1mC) base pairs by both experimental and computational methods. Here 9eG and 1mC were used to eliminate the possibility of metal cation interactions at, or binding between nucleobases at N1 of cytosine and N9 of guanine which is bound to ribose or deoxyribose in nucleic acids.



**Scheme 5.2** Watson-Crick and Hoogsteen structures for protonated G:C base pairs.

## 5.2 Methods

**5.2.1 Experimental Methods.** To obtain IRMPD spectra in the CH/NH/OH stretching regions, ions were first trapped and isolated by standard isolation techniques in a Bruker Apex Qe7 Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (MS). They were then exposed to the radiation from an optical parametric oscillator/amplifier (OPO/A) in the Laboratory for the Study of Energetics, Structures and Reactions of Gaseous Ions at Memorial University which has been described elsewhere.<sup>50</sup> Ions were electrosprayed from an Apollo II source from 1 mL of 1 mmol L<sup>-1</sup> solutions of 9eG in 50:50 H<sub>2</sub>O: methanol to which 2-3 drops of 100 mmol L<sup>-1</sup> 1mC in 50:50 H<sub>2</sub>O: methanol and 2-3 drops of aqueous 10 mmol L<sup>-1</sup> alkali metal chloride or 1% formic acid were added. The pH of the solutions from which the protonated complexes were electrosprayed was about 3.3. The laser was scanned at 2 cm<sup>-1</sup> steps and irradiation was for 1s at each wavelength. The laser power was between 600 and 800 mJ over the 2700 – 3800 cm<sup>-1</sup> region.

To obtain infrared spectra in the fingerprint region, the free electron laser (FEL) for infrared experiments (FELIX) was used in the Netherlands. IRMPD spectra were collected of ionic complexes trapped (stored waveform inverse Fourier transform, SWIFT) in a modified 3D quadrupole ion trap mass spectrometer (Bruker, AmaZon Speed ETD).<sup>51</sup> Ions were electrosprayed from solutions described above, but which were diluted 100-fold with acetonitrile. The FEL was scanned at 3 cm<sup>-1</sup> intervals with 2 pulses of tunable infrared radiation (between 45 and 120 mJ per pulse) from FELIX in the 650 – 1850 cm<sup>-1</sup> region. The wavelength was calibrated using an online grating spectrometer.

The reported IRMPD efficiencies (intensities) are the negative of the natural logarithm of the product ion intensities divided by the sum of the total ion intensities. Spectra presented are raw in that no corrections for the laser power was done.

All chemicals were purchased from Sigma-Aldrich and used without further purification.

**5.2.2 Computational Methods.** Geometry optimizations, dipole moments, polarizabilities, and IR frequency calculations were performed using Gaussian 09<sup>52</sup> suite of programs using B3LYP density functional theory which has been proved to be used successfully for nucleobase clusters.<sup>23,53,54</sup> The 6-31+G(d,p) split-valance basis set<sup>55</sup> was used for all atoms except Cs and Rb for which the Def2-SVP basis and effective core potential<sup>56,57</sup> was used. The relative 298 K Gibbs energies and enthalpies of various isomers with respect to the lowest energy structure were reported. The computed IR spectra were scaled by 0.975 and 0.945 in the lower (900-2000  $\text{cm}^{-1}$ ) and higher (2800-3800  $\text{cm}^{-1}$ ) energy regions, respectively, to compare with the experimental IRMPD spectra. The computed IR spectra were convoluted with a full width at the half max of 20  $\text{cm}^{-1}$  using Gaussview, and the units of the convoluted computed IR spectra are  $\text{L mol}^{-1} \text{cm}^{-1}$  in all cases. For all calculations, an empirical dispersion correction was done using Grimme's D3 version with the original D3 damping function, B3LYPD3.<sup>58,59</sup>

Proton affinities and gas basicities for 9-ethylguanine were computed at the CBS-QB3 level of theory in order to compare these quantities with those computed for cytosine, guanine, and the other alkylated bases which were computed previously.<sup>25</sup>

Calculations of proton-transfer transition states were using the QST3 keyword and verified by the animation of the imaginary frequency, which, in all cases, corresponding to the motion of the proton between the two bases.

## 5.3 Results and Discussion

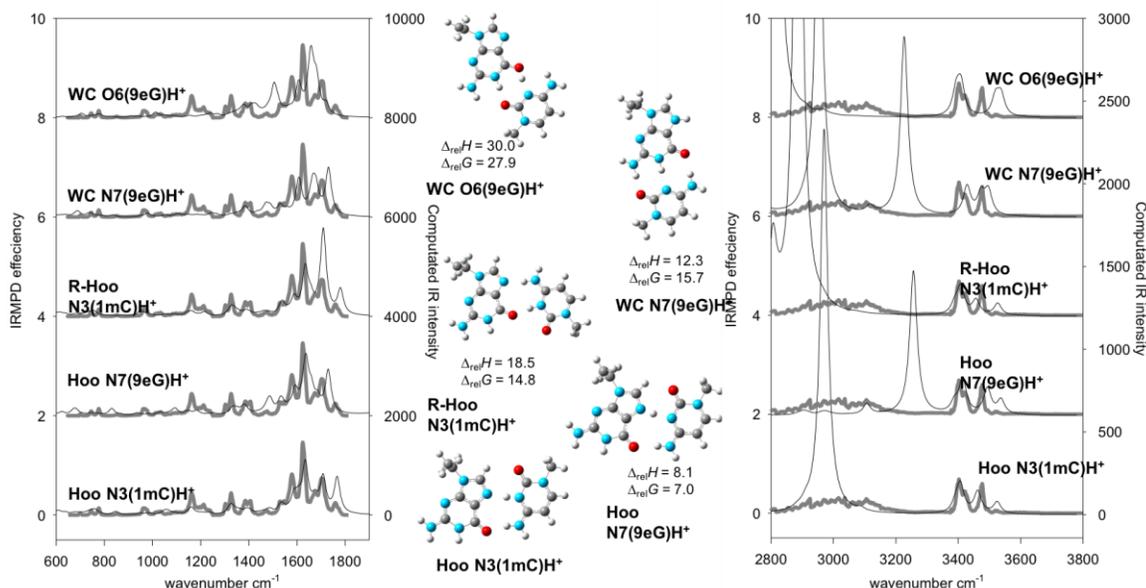
### 5.3.1 (9eG:1mC)H<sup>+</sup>.

**5.3.1.1 Structure of (9eG:1mC)H<sup>+</sup> by IRMPD Spectroscopy.** Upon absorption of tunable infrared radiation, the (9eG:1mC)H<sup>+</sup>, dissociated into both protonated 9eG and 1mC. The ratio of these products will be discussed in more detail below. The IRMPD spectrum of (9eG:1mC)H<sup>+</sup>, in both the fingerprint and hydrogen stretching region, is compared to the computed IR spectra of the five lowest energy structures in **Figure 5.1**. Using the same nomenclature as previously used for the (G:C)H<sup>+</sup> structures,<sup>15</sup> the lowest energy structure (Hoo N3(1mC)H<sup>+</sup>) is the Hoogsteen (Hoo) structure where the proton is covalently bonded to N3 of 1mC and is hydrogen bonded to N7 of 9eG. A second hydrogen bond occurs between the amino group of 1mC and O6 of 9eG. The IRMPD spectrum for this structure is entirely consistent with the experimental IRMPD spectrum in both regions. In the higher energy region, the predicted band just below 3000 cm<sup>-1</sup> is the N-H stretch of the amino group of 1mC that is hydrogen bonded to O6 of guanine. Experimentally, these strongly hydrogen-bonded and anharmonic absorptions typically show up as broad absorptions as they do in the present spectrum and are not well reproduced by harmonic calculations. The other hydrogen bonded N-H stretch in Hoo N3(1mC)H<sup>+</sup> is even more strongly red-shifted to about 2460 cm<sup>-1</sup>, out of the observable window of both lasers used.

A second Hoo structure, Hoo N7(9eG)H<sup>+</sup>, differs from the first mainly in that the proton is covalently bound to N7 of 9eG and hydrogen bonded to N3 of 1mC. This structure is 7.0 kJ mol<sup>-1</sup> higher in Gibbs energy than the lowest energy structure. The barrier to proton transfer between these complexes is computed to be low, 12.6 kJ mol<sup>-1</sup> using B3LYP/6-31+G(d,p). This proton transfer barrier is lower than the zero-point energy of the proton transfer mode which is about 15 kJ mol<sup>-1</sup> based on the computed position of the band, 2460 cm<sup>-1</sup>. However, B3LYP is known to substantially underestimate proton transfer barriers<sup>60,61</sup> by as much as 10 kJ mol<sup>-1</sup>, meaning that the actual proton transfer barrier is probably still above the zero-point energy resulting in two wells. Spectroscopically, one of the main features discriminating against these two structures is the free C=O (O2 of 1mC) stretch which is predicted in the Hoo N3(1mC)H<sup>+</sup> structure and observed at 1760 cm<sup>-1</sup>. However, the same C=O stretch of Hoo N7(9eG)H<sup>+</sup> is predicted significantly to the red at 1685 cm<sup>-1</sup>. Based on the computed energetics and the comparison of computed and experimental vibrational spectra, Hoo N7(9eG)H<sup>+</sup> is not likely a major contributor to the experimental spectrum.

The next higher energy structure is a reverse Hoogsteen structure (R-Hoo N3(mC)H<sup>+</sup>) and also has a general agreement with the observed IRMPD spectrum but is almost 15 kJ mol<sup>-1</sup> higher in Gibbs energy. The next lowest energy structures are a Watson-Crick (WC) base pair (WC N7(9eG)H<sup>+</sup>) and a similar WC base pairing structure but instead of being protonated at N7 of 9eG, it is protonated at O6 (WC O6(9eG)H<sup>+</sup>); there are only two hydrogen bonds in the latter structure, the bond donors being the O6 and N1 of guanine and acceptors are the N3 and O2 1mC positions, respectively. From the comparisons of the experimental and the computed spectra for WC N7(9eG)H<sup>+</sup> and WC O6(9eG)H<sup>+</sup> (**Figure**

5.1) and that they are almost 16 and 28 kJ mol<sup>-1</sup> higher, respectively, in Gibbs energy than the lowest energy Hoo structure, it is clear that they are not major contributors to the (9eG:1mC)H<sup>+</sup> complex. Structures and spectra for ten higher energy isomers presented in Figures S1-S2.



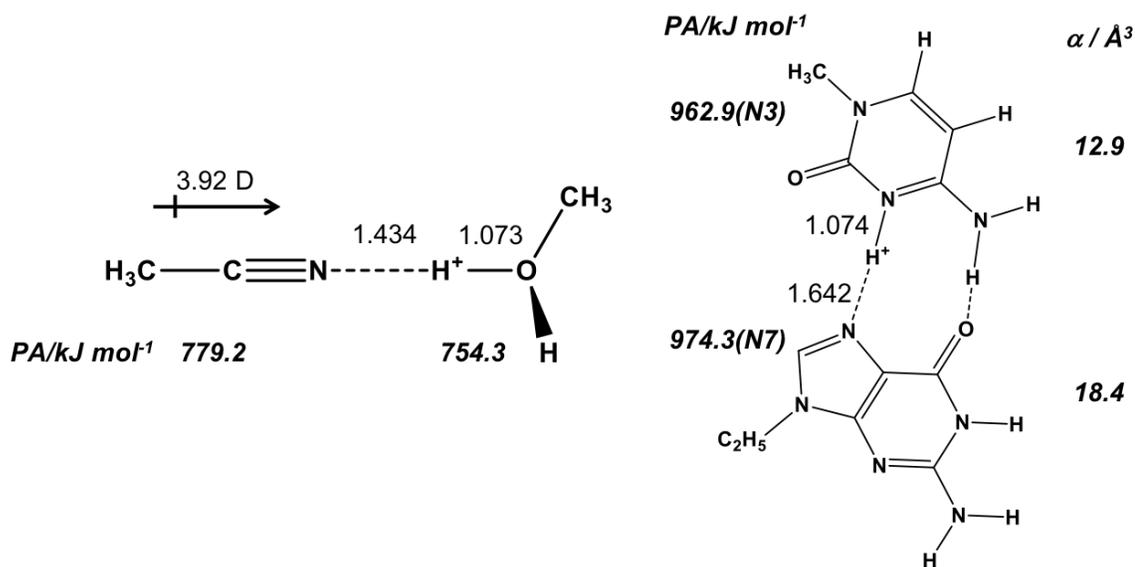
**Figure 5.1** Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint and the CH/NH/OH stretching regions for (9eG:1mC)H<sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers. The 298 K enthalpies and Gibbs energies relative to structure Hoo N7(9eG)H<sup>+</sup> are also shown (also computed using B3LYP/6-31+G(d,p)). All energies are shown in kJ mol<sup>-1</sup>.

At this point, it is interesting to further discuss the structure of the protonated dimers of cytosine and guanine and their alkylated analogues in more detail. The proton in the lowest energy (9eG:1mC)H<sup>+</sup> structure, Hoo N3(1mC)H<sup>+</sup>, is covalently bound to N3 of 1mC with a bond distance of 1.074 Å and hydrogen bonded to N7 of 9eG 1.642 Å, despite the proton affinity of 9eG being computed to be 11.4 kJ mol<sup>-1</sup> higher (CBS-QB3) than N3 in

1mC (**Scheme 5.3**). The structures of mixed proton bound dimers containing one high dipole moment monomer have been studied previously and may shed light on the structures of the guanine/cytosine (and alkylated analogues) proton bound dimers.<sup>62,63</sup> For example, the proton in the computed minimum energy structure for the protonated methanol/acetonitrile dimer was covalently bound to methanol and protonated methanol was hydrogen bonded to acetonitrile despite acetonitrile having a proton affinity that is almost 25 kJ mol<sup>-1</sup> higher than methanol (see **Scheme 5.3**). Since acetonitrile has such a high dipole moment, 3.92 vs 1.70 D for methanol, the ion-dipole interaction between protonated methanol and acetonitrile more than makes up for the 25 kJ mol<sup>-1</sup> energy deficit incurred by having the proton bound to methanol rather than acetonitrile. It was also shown that the structural anomaly, the degree of sharing of the proton or the difference in base-H<sup>+</sup> distance, depends on both the difference in proton affinity and the difference in dipole moment.<sup>58</sup>

The B3LYP/6-31+G(d,p) dipole moments of G and C are similar, 6.82 and 6.86 D, respectively (Table S1). These compare well with the previously computed 7.26 and 6.93 D values<sup>64</sup> for G and C, respectively, and the experimental value for C of 7.0 D.<sup>65</sup> For 9eG and 1mC, the dipole moments are also predicted to be not too dissimilar, 7.36 vs 6.41 D. In both cases, (G:C)H<sup>+</sup> and (9eG:1mC)H<sup>+</sup>, to observe this structural anomaly based on the difference in proton affinities between N7 of G or 9eG and N3 of C or 1mC, respectively, differences in dipole moments would have to be more than about 3 D.<sup>58</sup> Instead, we offer an alternative, yet similar explanation for the position of proton in (9eG:1mC)H<sup>+</sup>. In both cases, (G:C)H<sup>+</sup> and (9eG:1mC)H<sup>+</sup>, the guanine is computed to be significantly more polarizable (Table S1).<sup>65,66</sup> Using B3LYP/6-31+G(d,p) 1mC has computed polarizability

( $\alpha$ ) of  $12.9 \text{ \AA}^3$  while that of 9eG is  $18.4 \text{ \AA}^3$  (Table S1). Ion-induced dipole interactions can be quite strong, for example, the  $\text{Na}^+ \cdots \text{CH}_4$  complex has a binding energy<sup>67</sup> of some  $30 \text{ kJ mol}^{-1}$  and the polarizability of  $\text{CH}_4$  is only  $3.0 \text{ \AA}^3$ , far less than the difference between the computed polarizabilities of 1mC and 9eG. The position of the proton in  $(9\text{eG}:1\text{mC})\text{H}^+$  (and  $(\text{G}:\text{C})\text{H}^+$  etc.), significantly closer to the lower proton affinity N3 of 1mC than N7 of 9eG is because the difference in proton affinity,  $11.4 \text{ kJ mol}^{-1}$  using CBS-QB3, is made up for by stronger ion-induced dipole interactions between  $1\text{mCH}^+$  and neutral 9eG. That N3 of 1mC is protonated rather than O2 is because protonating at N3 allows two hydrogen bonds to be formed in  $(9\text{eG}:1\text{mC})\text{H}^+$ .



**Scheme 5.3** The structures of protonated dimers of methanol and acetonitrile, and of 9-ethylguanine and 1-methylcytosine. In both cases the lower proton affinity base is the site of protonation.

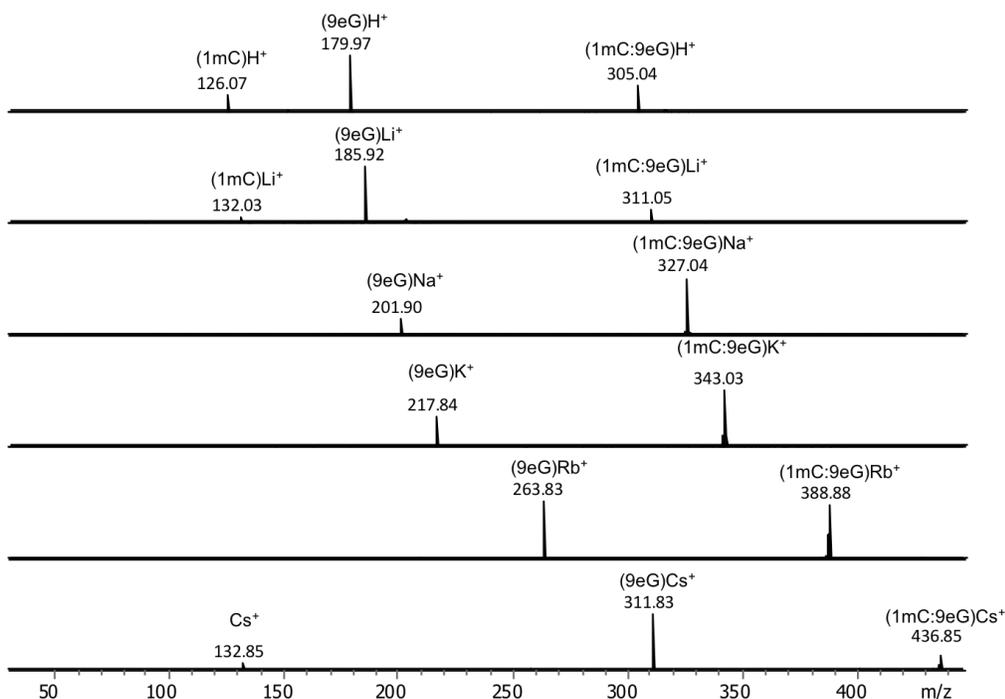
**5.3.1.2. Dissociation of (9eG:1mC)H<sup>+</sup>.** Interestingly, under IRMPD conditions (9eG:1mC)H<sup>+</sup> undergoes two competing dissociations losing either 9eG or 1mC to leave (1mC)H<sup>+</sup> and (9eG)H<sup>+</sup> at *m/z* 126 and 180, respectively. The ratio of the dissociation products is about 30% (1mC)H<sup>+</sup>. This has been observed in the past for (G:C)H<sup>+</sup> under both CID and IRMPD conditions where the ratio of products is close to 50:50.<sup>15,49,68</sup> Based on the experimental gas basicities (927.6 and 918 kJ mol<sup>-1</sup> for G and C, respectively)<sup>69</sup> and assuming a negligible barrier for proton transfer between the two monomers, it is expected that the CH<sup>+</sup> should not be present in a greater abundance than 2%. This is the reason that having 50% CH<sup>+</sup> upon CID and IRMPD has been termed an anomaly. However, based on the CBS-QB3 computed gas basicities<sup>25</sup> between most basic sites, O2 of C and N7 of G (921.2 and 921.1 kJ mol<sup>-1</sup>, respectively, Table S1), the expected ratio is, in fact, closer to 50% CH<sup>+</sup>, but lower if the proton remains on N3 of cytosine. For (1mG:1mC)H<sup>+</sup> about 80% of the product of CID is (1mC)H<sup>+</sup>, but for (9mG:1mC)H<sup>+</sup> only 40% of the product is (1mC)H<sup>+</sup>.<sup>18</sup> Based on the CBS-QB3 gas basicities (Table S1), the expected abundance of (1mC)H<sup>+</sup> should only be about 55% and 10% for CID of (1mG:1mC)H<sup>+</sup> and (9mG:1mC)H<sup>+</sup>, respectively; significantly lower than the observed abundances. Even more surprising is the observed 30% (1mC)H<sup>+</sup> (**Figure 5.2**) in the present work for IRMPD of (9eG:1mC)H<sup>+</sup> given that the amount expected based on the CBS-QB3 gas basicities (943.3 kJ mol<sup>-1</sup> for 9eG and 933.2 kJ mol<sup>-1</sup> for 1mC, Table S1) is only 2% (or 0.6% based on the proton affinity at N3 of 1mC, 930.5 kJ mol<sup>-1</sup>). The difference in Gibbs energy between two Hoogsteen structures Hoo N3(1mC)H<sup>+</sup> and Hoo N7(9eG)H<sup>+</sup> is 7.0 kJ mol<sup>-1</sup> meaning that the latter is expected to be present in a mixture at about 6%, which still does not account for the 30% (1mC)H<sup>+</sup> that is observed.

In the lowest energy structure, Hoo N3(1mC)H<sup>+</sup>, the proton is covalently bound to N3 of 1mC and there are two hydrogen bonds between (1mC)H<sup>+</sup> and neutral 9eG along with a degree of ion-induced dipole interaction. To affect IRMPD or CID, energy in the form of multiple infrared photons or multiple collisions, respectively, is deposited into (9mG:1mC)H<sup>+</sup>. The electronic energy barrier for the transfer of proton from 1mC to 9eG is computed to be 12.6 kJ mol<sup>-1</sup>; the N7(9eG) – N3(1mC), (abbreviated N--N) distance was 2.569 Å in the transition state compared to 2.713 Å at the global minimum; in order to traverse the minimum proton-transfer energy barrier, the N--N distance must shrink from its optimal value.

The energy barrier to proton transfer was computed at a few different N--N distances, where N--N was held fixed and all other internuclear degrees of freedom were optimized. At an N--N distance of 2.713 Å, the global minimum, the electronic energy barrier has already increased to 23 kJ mol<sup>-1</sup>. Above the dissociation thresholds, computed to be 195 kJ mol<sup>-1</sup> to produce (1mC)H<sup>+</sup> and 185 kJ mol<sup>-1</sup> to produce (9eG)H<sup>+</sup>, there is more than enough energy to traverse these proton transfer barriers, so if thermodynamics was guiding this reaction, there should be virtually no (1mC)H<sup>+</sup>. At N--N distances of 2.863, 3.163, 3.263, and 3.463 Å, however, the barrier rises to 48, 122, 150, and 208 kJ mol<sup>-1</sup>, respectively; this latter point already in excess of the dissociation threshold.

A potential energy surface was scanned along the N--N distance, representing the dissociation of the complex, and the N-H<sup>+</sup> distance (N3-H<sup>+</sup> of 1mC) and is presented as a contour plot in **Figure 5.3**. It can be seen that there is a relatively narrow channel from the global minimum at N--N and N-H<sup>+</sup> of 2.713 and 1.074 Å, respectively, to the proton transfer product where the Hoo N7(9eG)H<sup>+</sup> has an N--N and N-H<sup>+</sup> of 2.651 and 1.551 Å,

respectively. It is quite a shallow well near the global minimum energy complex. Proton transfer to form Hoo N7(9eG)H<sup>+</sup> requires very little energy and stretching the N-N bond by almost an angstrom (i.e. along the N3-N7) axis is merely 30 kJ mol<sup>-1</sup> higher in energy. If by increasing the internal energy, the complex begins to increase the N--N distance (i.e. dissociate) the barrier to proton transfer rises quite dramatically. This can be seen if one moves from the global minimum on the surface up to an N--N distance of, say 3.4 Å which is only about 27 kJ mol<sup>-1</sup> above the minimum, then move out along the N-H<sup>+</sup> axis; the energy of transferring a proton when the N--N distance has increased to 3.4 is close to 190 kJ mol<sup>-1</sup>. The explanation for such a large amount of (1mC)H<sup>+</sup> being observed, far beyond what is expected based on thermodynamics, is due to a large kinetic barrier to proton transfer *once the dynamics of dissociation has already begun*. In light of this argument, and that in the lowest energy structure of (9eG:1mC)H<sup>+</sup> where the proton is covalently bound to 1mC, it might be more astonishing that any (9eG)H<sup>+</sup> be formed. However, that (9eG)H<sup>+</sup> observed means that the lifetime of the excited complex, above the dissociation threshold, is long enough that multiple instances of inter-base bond lengthening and contracting as well as proton transfer back and forth between the bases, is possible. Molecular dynamics simulations should be able to shed more light on the dynamics of (9eG:1mC)H<sup>+</sup> dissociation.



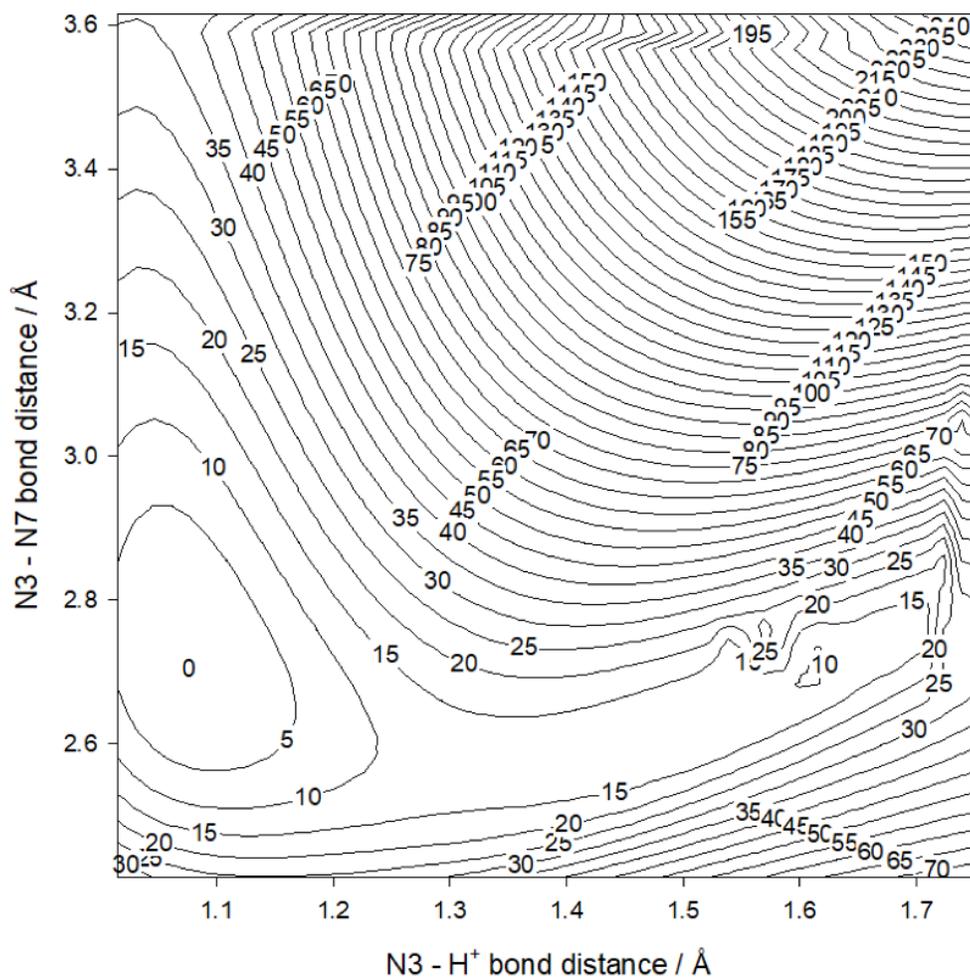
**Figure 5.2** Infrared multiple photon dissociation mass spectra of all six complexes at the specified wavenumber. The  $m/z$  and name of the precursor ions and fragments are also shown. The vertical axis is the % relative abundance of the largest feature.

### 5.3.2 (9eG:1mC)M<sup>+</sup> (M=Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Rb<sup>+</sup>, Cs<sup>+</sup>)

Different isomers for each (9eG:1mC)M<sup>+</sup> complex are named according to the position that the metal cation is bound to on 9eG and 1mC, separated by a hyphen. Multiple structures with the same binding positions are distinguished with numbers; 1, 2, 3...etc; at the end which follows the sequence of increasing 298 K Gibbs energy.

Upon absorption of infrared radiation from the OPO or FEL, all of the alkali metal cationized dimers dissociate to lose 1mC, except (9eG:1mC)Li<sup>+</sup> where about 1-2% of the fragmentation products were (1mC)Li<sup>+</sup> (**Figure 5.2**). A small amount of Cs<sup>+</sup> was also

observed during IRMPD of (9eG:1mC)Cs<sup>+</sup> which is presumably secondary dissociation of (9eG)Cs<sup>+</sup>.



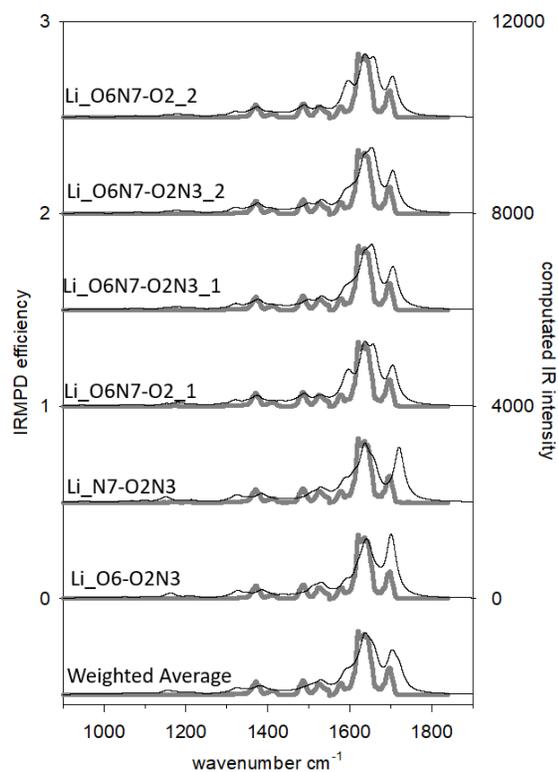
**Figure 5.3** A contour potential energy surface for (9eG:1mC)H<sup>+</sup>. The vertical axis is the N3—N7 distance and the horizontal is N—H<sup>+</sup> distance (N3 of 1mC). The contours are the electronic energies relative to the lowest energy structure in kJ mol<sup>-1</sup>.

**5.3.2.1 (9eG:1mC)Li<sup>+</sup>.** The IRMPD spectra for (9eG:1mC)Li<sup>+</sup> in the fingerprint region is presented in **Figure 5.4** along with the computed IR spectra and thermochemistry for the

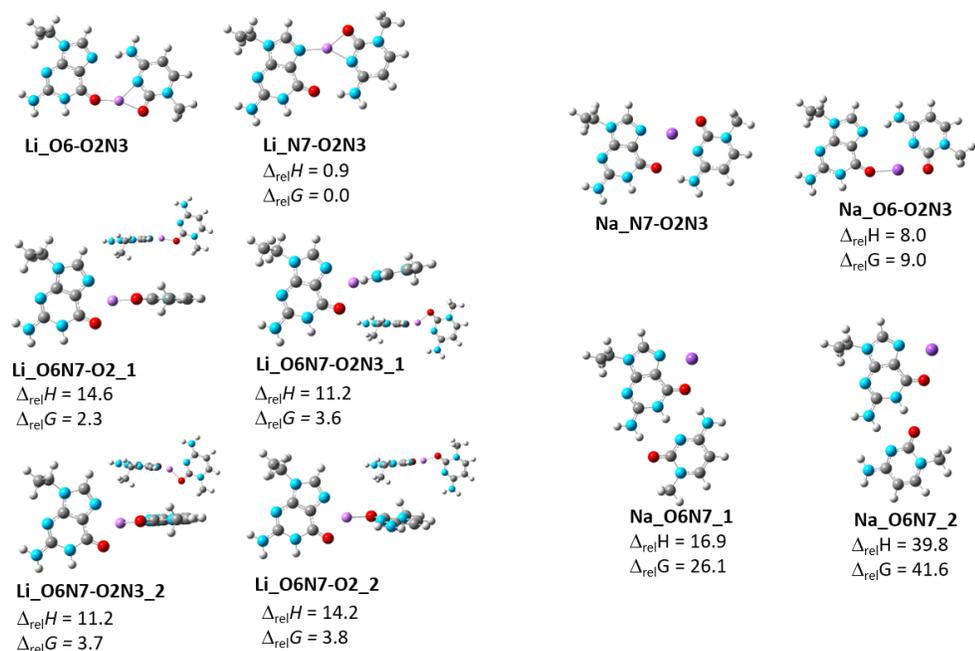
six lowest energy structures in **Figure 5.5**, all of which are within 4 kJ mol<sup>-1</sup> in Gibbs energy. The lowest energy structure, Li\_O6-O2N3, is a reverse Hoogsteen type structure containing a 1mC amine group to N7 (9eG) hydrogen bond. An isoenergetic isomer, N7-O2N3, is similar to a Hoogsteen structure where the hydrogen bond is now between the amine group of 1mC and O6 of 9eG. In both structures, the metal is bound to both N3 and O2 of 1mC. There is a general agreement between the features of both predicted IR spectra and the experimental IRMPD spectrum. In Li\_O6-O2N3, the 9eG C=O stretch is red shifted to 1701 cm<sup>-1</sup> due to binding to Li<sup>+</sup> and is in better agreement than the hydrogen bonded 9eG C=O stretch predicted at 1720 cm<sup>-1</sup> for Li\_N7-O2N3. The N-H stretching region is presented in Figure S28 – S32 for all the metal coordinated complexes. In the N-H stretching region for (9eG:1mC)Li<sup>+</sup> (Figure S28), the free N-H stretching is observed at 3400 cm<sup>-1</sup>, but the hydrogen bonded N-H stretching absorptions predicted below 3400 cm<sup>-1</sup> are not observed. The absence of these absorptions or the presence of very broad absorptions is sometimes the case for these very anharmonic hydrogen bonded modes when they are present in strongly bound complexes.<sup>70-73</sup>

Li\_O6N7-O2\_1 and Li\_O6N7-O2\_2 are like one another with Li bound to O6 and N7 of 9eG as well as O2 of 1mC, and the two bases are almost perpendicular to one another. They only differ by a flip of the bases with respect to one another. Another set, Li\_O6N7-O2N3\_1 and Li\_O6N7-O2N3\_2 are also similar to one another and the previously mentioned two, except that the Li<sup>+</sup> is also bound to N3 of cytosine. Like Li\_O6-O2N3, the carbonyl groups are all bound to Li<sup>+</sup> and their predicted spectra are all in general agreement with IRMPD spectra. It is not possible based on the spectroscopy to rule out any of these lowest energy structures.

Based on the computed 298 K Gibbs energies, these six (9eG:1mC)Li<sup>+</sup> structures in **Figure 5.5** would be present in a ratio of 1 : 1 : 0.40 : 0.23 : 0.22 : 0.22 assuming an equilibrium distribution at 298 K. A weighted average of the six spectra using these ratios as weighting factors is provided at the bottom of **Figure 5.4** and is consistent with the experimental IRMPD spectrum.



**Figure 5.4** Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)Li<sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for six isomers. The weighted average spectra of all six structures are shown in both regions.



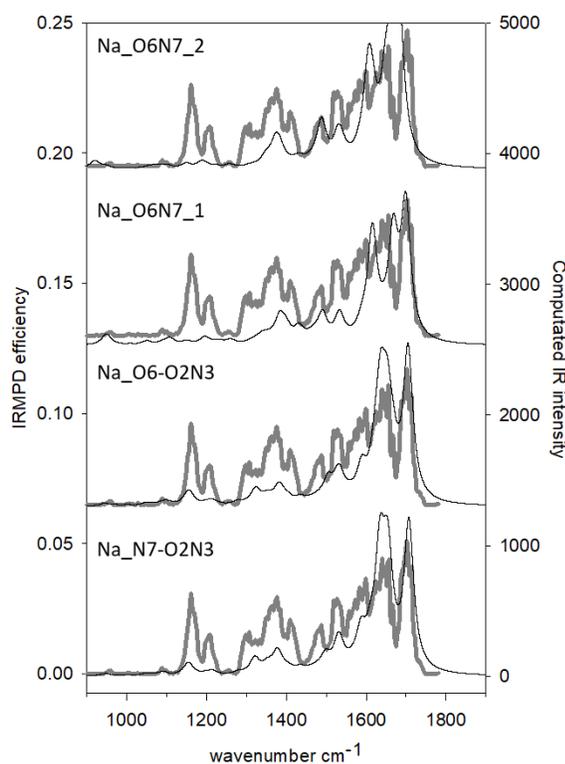
**Figure 5.5** Structures of (9eG:1mC)Li<sup>+</sup> and (9eG:1mC)Na<sup>+</sup> isomers whose IR spectra are compared to the experimental spectra in **Figures 5.4** and **5.6**. The structures and relative 298 K enthalpies and Gibbs energies were computed using B3LYPD3/6-31+G(d,p) and are in kJ mol<sup>-1</sup>. For the Li<sup>+</sup> structures, some side views are shown for non-planar structures.

**5.3.2.2 (9eG:1mC)Na<sup>+</sup>.** The IRMPD spectrum in the fingerprint region for (9eG:1mC)Na<sup>+</sup> is shown in **Figure 5.6** along with the computed IR spectra for the four lowest energy structures. All four structures are displayed in **Figure 5.5**. The lowest energy structure is the Hoogsteen type structure, Na\_N7-O2N3, similar to Li\_N7-O2N3 structure for (9eG:1mC)Li<sup>+</sup> except in the Na<sup>+</sup> structure, the metal is shared more between both N7 and O6 of 9eG. For example in Li\_N7-O2N3, O6 of 9eG is hydrogen bonded to the amine group of 1mC with an O6-Li<sup>+</sup> distance of 3.227 Å whereas in Na\_N7-O2N3 the O6\_Na<sup>+</sup> distance is 2.845 Å. The reverse Hoogsteen type structure, Na\_O6-O2N3, is 9.0 kJ mol<sup>-1</sup> higher in Gibbs energy and would only be expected to be present in an abundance of 2.6 % of that of the lowest energy structure based on an equilibrium distribution at 298 K. No

non-planar structures could be found as in (9eG:1mC)Li<sup>+</sup>, instead, they optimized to the structures where the bases were co-planar. Due to the greater interaction between the metal cation and O6 of 9eG in Na\_N7-O2N3 the red shift of the 9eG C=O stretch in Na\_N7-O2N3 is greater than that in Li\_N7-O2N3 and is predicted to occur at 1706 cm<sup>-1</sup>, in good agreement with the experimental spectrum. In fact, the combined hydrogen bond and binding of Na<sup>+</sup> to the carbonyl of 9eG, the red shift is similar to that in Na\_O6-O2N3 which is predicted to occur at 1704 cm<sup>-1</sup> and renders the two C=O stretches indistinguishable.

The next two lowest energy structures found have Gibbs energy differences of 26.1 and 41.6 kJ mol<sup>-1</sup> relative to Na\_N7-O2N3. Na\_O6N7\_1 is a Watson-Crick base pair with Na<sup>+</sup> bonded to O6 and N7 of 9eG. Na\_O6N7\_2 is a reverse Watson-Crick base pair with the same metal binding. Neither of these two structures, both very high in energy, are as consistent with the IRMPD spectrum as the Hoogsteen structure.

Spectroscopically, we cannot distinguish between the two lowest energy structures for (9eG:1mC)Na<sup>+</sup> as both their predicted spectra are consistent with the experimental vibrational spectrum, but we can rule out the higher energy structures.

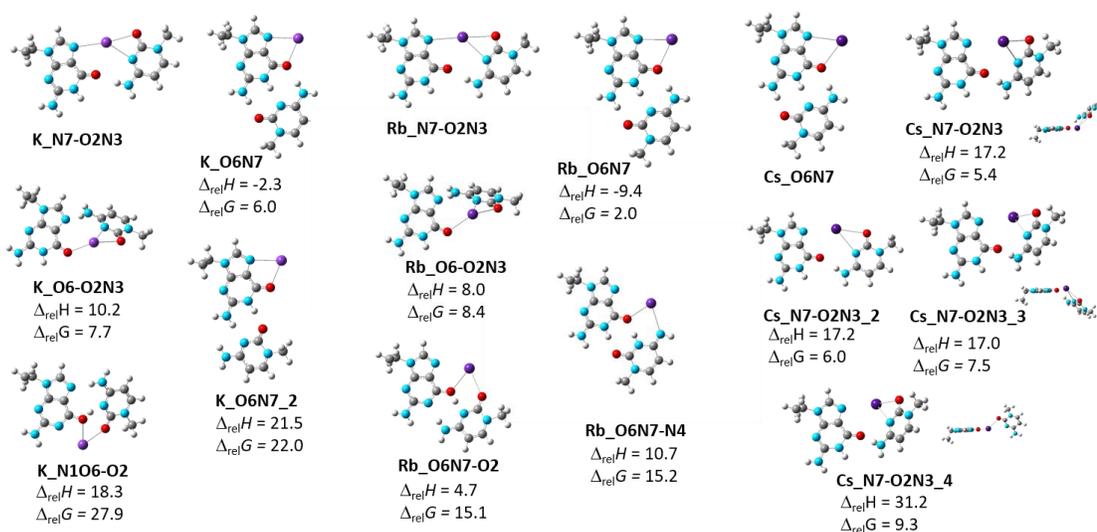


**Figure 5.6** Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)Na<sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers.

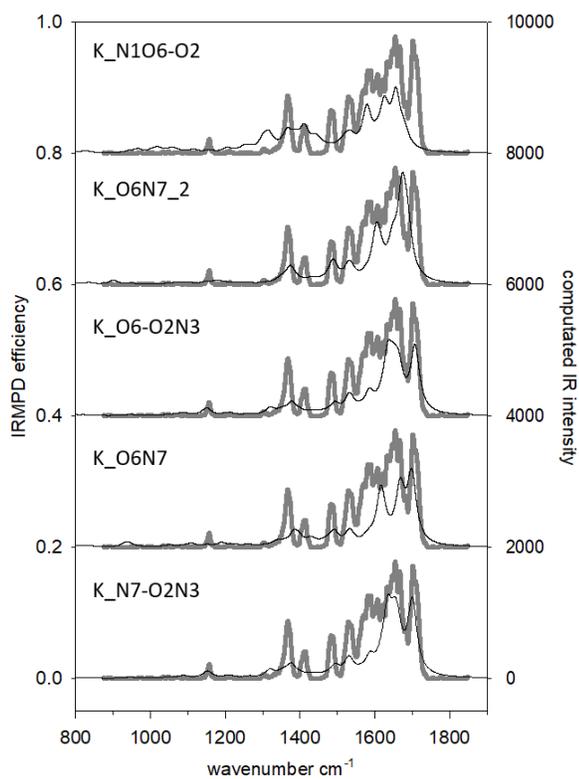
**5.3.2.3 (9eG:1mC)K<sup>+</sup> and (9eG:1mC)Rb<sup>+</sup>.** All the computed lowest energy structures for K<sup>+</sup>, Rb<sup>+</sup> and Cs<sup>+</sup> are shown in **Figure 5.7**. The experimental IRMPD spectra and the computed IR spectra of the five lowest energy structures for (9eG:1mC)K<sup>+</sup> and (9eG:1mC)Rb<sup>+</sup> are presented in **Figures 5.8** and **5.9**, respectively. Both complexes share the same three lowest energy structures. The lowest in energy are the Hoogsteen-type structures, M\_N7-O2N3, with the metal cation bound to N7 and O6 of 9eG as well as N3 and O2 of 1mC. The next lowest energy structure, M\_O6N7, has Watson-Crick base pairing with the metal bound to N7 and O6 of 9eG and not interacting with a cytosine at

all. The computed IR spectra of both of these structures are consistent with the experimental spectrum in the fingerprint region, however, in the higher-energy region (Figures S30 and S31), the weakly red-shifted hydrogen bonded N-H stretching region is in better agreement with the rather broad band observed in experimental vibrational spectra. It is not possible to rule these two lowest energy structures out conclusively.

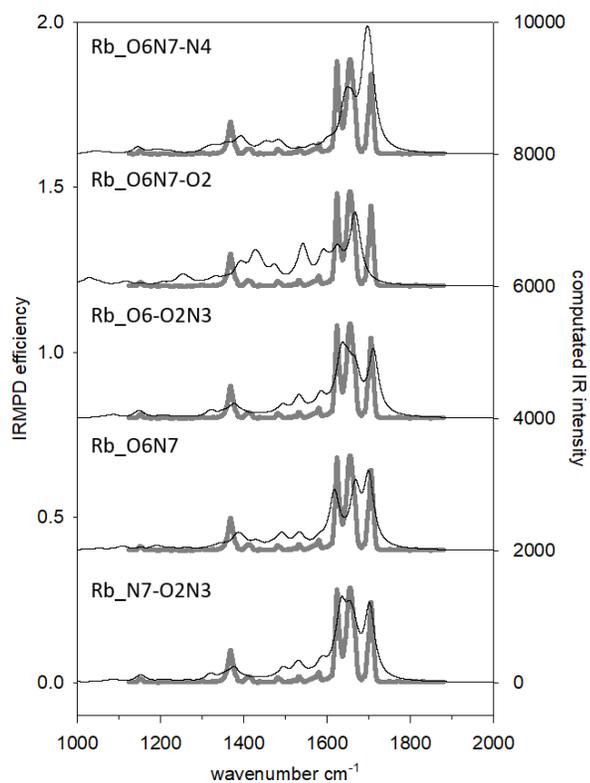
The third lowest-energy structure for both metals are non-planar complexes with the metal bound to O6 of 9eG and O2 and N3 of 1mC and contains a hydrogen bond between the amine group of 1mC and N7 of 9eG; for both metals this structure about 8 kJ mol<sup>-1</sup> higher in Gibbs energy than the lowest energy structure. Spectroscopically, the M\_N7-O2N3 and M\_O6-O2N3 cannot be ruled out. The next two highest energy structures for both metals are significantly higher in relative Gibbs energy, and their computed IR spectra are not particularly consistent with the experimental spectrum.



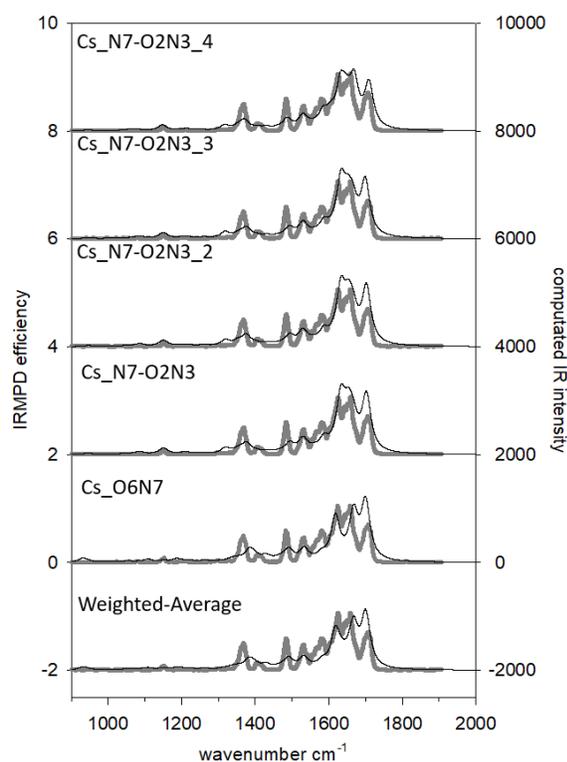
**Figure 5.7** Structures of (9eG:1mC)K<sup>+</sup>, (9eG:1mC)Rb<sup>+</sup>, and (9eG:1mC)Cs<sup>+</sup> isomers whose IR spectra are compared to the experimental spectra in **Figures 5.8, 5.9** and **5.10**. The structures and relative 298 K enthalpies and Gibbs energies were computed using B3LYPD3/6-31+G(d,p) and are in kJ mol<sup>-1</sup>. For the Cs<sup>+</sup> structures, some side views are shown for non-planar structures.



**Figure 5.8** Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for  $(9eG:1mC)K^+$  with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers.



**Figure 5.9** Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)Rb<sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers.



**Figure 5.10** Comparison of the experimental IRMPD spectra (grey traces) in the fingerprint region for (9eG:1mC)Cs<sup>+</sup> with the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five isomers.

**5.3.2.4 (9eG:1mC)Cs<sup>+</sup>.** All five lowest energy structures for (9eG:1mC)Cs<sup>+</sup> are presented in **Figure 5.7**, while their computed IR spectra are compared to the experimental IRMPD spectrum in **Figure 5.10**. The lowest energy structure, Cs\_O6N7, is computed to be the Watson-Crick structure with the metal cation bound solely to 9eG, and cytosine bound to 9eG by three hydrogen bonds. There are also four N7-O2N3 isomers higher in Gibbs energies by 5.4 kJ mol<sup>-1</sup>, 6.0kJ mol<sup>-1</sup>, 7.5kJ mol<sup>-1</sup> and 9.3kJ mol<sup>-1</sup> relative to O6N7. In **Figure 5.7**, the difference between the isomers is visualized by their side view, where the four structures are different due to the orientations of 1mC relative to 9eG. The weighted-average computed IR spectrum is not inconsistent with the experimental spectrum in the

fingerprint region. In CH/OH/NH region (Figure S32), Cs\_O6N7 is probably responsible for the shoulders on the low energy side of the free N-H stretching bands whose positions are well represented by the N7-O2N3 structures. Spectroscopically, none of the structures shown in **Figure 5.7** can be ruled out.

**5.3.3 The effects of the alkali metal cation on the computed structure of the G:C base pair.** In Figure S27, we compare some computed results on the structures of the alkali metal cations 9eG:1mC base pairs. It is interesting to note that the lighter metal cations have an affinity for binding the two bases together in Hoogsteen type structures. Cs<sup>+</sup> does not exist at the B3LYP/6-31+G(d,p) level of theory in the M\_06-O2N3 which is the lowest energy structure for Li<sup>+</sup>. In fact, the lowest energy structure for the cesiated complex is the WC structure. Of the three structures shown in Figure S27, the WC structure decreases in relative energy as the metal cation increases in size.

## 5.4. Conclusion

A study of the protonated and metal cationized complexes between 9eG and 1mC, (9eG:1mC)H<sup>+</sup> and (9eG:1mC)M<sup>+</sup>, M= Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Rb<sup>+</sup>, Cs<sup>+</sup> was undertaken using IRMPD spectroscopy and computational methods.

For (9eG:1mC)H<sup>+</sup>, both experiment and theory show that the predominant structure was a Hoogsteen type complex with the proton covalently bound to N3 of 1mC despite this being the third best protonation site of the two bases. This structural oddity is explained as being due to a combination of being able to form more hydrogen bonds when the proton is bound to N3 rather than O2 of 1mC and being able to form a strong ion-induced dipole

interaction between N3 protonated 1mC and 9eG, the latter of which is expected to have a much higher polarizability than 1mC. Furthermore, the anomalous dissociation of (9eG:1mC)H<sup>+</sup>, forming much more (1mC)H<sup>+</sup> than would be predicted based on computed thermochemistry, was explained as being due to a larger barrier to proton transfer from N3 of 1mC to N7 of 9eG as the base pair begins to dissociate.

The (9eG:1mC)H<sup>+</sup> Hoogsteen structure determined in this study, is identical to a (G:C)H<sup>+</sup> motif found to exist in RNA by comparing base pair structures from X-ray crystallography measurements in the protein database (PDB) with computed structures from electronic structure calculations.<sup>74</sup> The present work provides infrared spectroscopic evidence for and characterization of a gas phase self-assembled protonated base pair which is also observed in real cellular nucleic acids.

While single unique structures for the (9eG:1mC)M<sup>+</sup>; M= Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Rb<sup>+</sup>, Cs<sup>+</sup> complexes could not be assigned, the experimental spectra were consistent with the computed spectra. In all cases the experimental spectra are consistent with the metal cation being bound to canonical 1mC through both O2 and N3 which is consistent with the conclusions of previous IRMPD spectroscopic<sup>10</sup> and bond dissociation<sup>75</sup> measurements by Rodger's group concluding that the alkali metal cation is bound to O2 and N3 of cytosine. For (9eG:1mC)Li<sup>+</sup>, the lowest energy structure is one in which the O2 and N3 bound 1mC/Li<sup>+</sup> is bound to O6 of 9eG. For Na<sup>+</sup>, K<sup>+</sup>, and Rb<sup>+</sup>, similar binding of the metal cation to 1mC is calculated but they are also bound to N7 of 9eG. For Cs, the lowest energy structure is the Watson-Crick type structure with Cs<sup>+</sup> only binding to 9eG through O6 and N7 but the structure similar to those observed for the lighter alkali metals cannot be ruled out. It is also interesting to note that the Watson-Crick base pairing structure gets lower in

Gibbs energy relative to the lowest energy complexes as the metal gets larger. This indicates that the smaller, more densely charged cations have a greater propensity to interfere with Watson-Crick base pairing than do the larger, less densely charged metal cations.

### **Acknowledgments**

We are grateful for the excellent support from the FELIX technical staff. This research was enabled in part by support provided by ACENET, Compute Ontario, and Westgrid as well as Compute Canada. Financial support from NSERC, Bruker, CFI, and Memorial University.

## References

- 1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
- 2 C. Y. Switzer, S. E. Moroney and S. A. Benner, *Biochemistry*, 1993, **32**, 10489–10496.
- 3 J. L. Sessler, C. M. Lawrence and J. Jayawickramarajah, *Chem. Soc. Rev.*, 2007, **36**, 314–325.
- 4 A. H. J. Wang, T. Hakoshima, G. van der Marel, J. H. van Boom and A. Rich, *Cell*, 1984, **37**, 321–331.
- 5 Y. Takezawa and M. Shionoya, *Acc. Chem. Res.*, 2012, **45**, 2066–2076.
- 6 M. H. Shamsi and H. B. Kraatz, *J. Inorg. Organomet. Polym. Mater.*, 2013, **23**, 4–23.
- 7 B. Lippert, *J. Chem. Soc., Dalt. Trans.*, 1997, **21**, 3971–3976.
- 8 J. Anastassopoulou, *J. Mol. Struct.*, 2003, **651–653**, 19–26.
- 9 M. Azargun, Y. Jami-Alahmadi and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2017, **19**, 1281–1287.
- 10 B. Yang, R. R. Wu, N. C. Polfer, G. Berden, J. Oomens and M. T. Rodgers, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 1523–1533.
- 11 K. J. Koch, T. Aggerholm, S. C. Nanita and R. G. Cooks, *J. Mass Spectrom.*, 2002, **37**, 676–686.
- 12 G. M. Peters and J. T. Davis, *Chem. Soc. Rev.*, 2016, **45**, 3188–3206.
- 13 J. Gao, G. Berden, M. T. Rodgers and J. Oomens, *Phys. Chem. Chem. Phys.*, 2016, **18**, 7269–7277.

- 14 J. Oomens, A. R. Moehlig and T. H. Morton, *J. Phys. Chem. Lett.*, 2010, **1**, 2891–2897.
- 15 A. F. Cruz-Ortiz, M. Rossa, F. Berthias, M. Berdakin, P. Maitre and G. A. Pino, *J. Phys. Chem. Lett.*, 2017, **8**, 5501–5506.
- 16 J. Y. Salpin, S. Guillaumont, J. Tortajada, L. MacAleese, J. Lemaire and P. Maitre, *ChemPhysChem*, 2007, **8**, 2235–2244.
- 17 R. Cheng, E. Loire and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2019, **1**, 11103–11110.
- 18 J. J. Park and S. Y. Han, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 846–854.
- 19 N. Russo, M. Toscano and A. Grand, *J. Phys. Chem. B*, 2001, **105**, 4735–4741.
- 20 D. A. Megger, P. M. Lax, J. Paauwe, C. Fonseca Guerra and B. Lippert, *J. Biol. Inorg. Chem.*, 2018, **23**, 41–49.
- 21 P. Amo-Ochoa and F. Zamora, *Coord. Chem. Rev.*, 2014, **276**, 34–58.
- 22 T. Aggerholm, S. C. Nanita, K. J. Koch and R. G. Cooks, *J. Mass Spectrom.*, 2003, **38**, 87–97.
- 23 E. A. L. Gillis, M. Demireva, K. Nanda, G. Beran, E. R. Williams and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 3304–3315.
- 24 M. Pechlaner and R. K. O. Sigel, University of Zurich, 2012.
- 25 M. Azargun and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25778–25785.
- 26 A. H.-J. Wang, J. Reedijk, J. H. van Boom, S. S. G. E. van Boom and D. Yang, *Biochemistry*, 1995, **34**, 12912–12920.
- 27 S. Cai, X. Tian, L. Sun, H. Hu, S. Zheng, H. Jiang, L. Yu and S. Zeng, *Anal. Chem.*, 2015, **87**, 10542–10546.

- 28 P. Das, N. Das, S. Bhowmick, S. Kumari, A. Jana and K. Singh, *Polyhedron*, 2018, **157**, 267–275.
- 29 S. Sivakova and S. J. Rowan, *Chem. Soc. Rev.*, 2004, **34**, 9–21.
- 30 P. Papsai, J. Aldag, T. Persson and S. K. C. Elmroth, *Dalt. Trans.*, 2006, 3515–3517.
- 31 J. P. Cerón-Carrasco, D. Jacquemin and E. Cauët, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12457–12464.
- 32 L. Liu, Y. Pei, Y. Zhang, J. Wang and L. Chen, *Eur. Polym. J.*, 2018, **102**, 75–81.
- 33 K. Rajabi, E. A. L. Gillis and T. D. Fridgen, *J. Phys. Chem. A*, 2010, **114**, 3449–3456.
- 34 N. Russo, M. Toscano and A. Grand, *J. Am. Chem. Soc.*, 2001, **123**, 10272–10279.
- 35 M. Meyer, A. Hocquet and J. Sühnel, *J. Comput. Chem.*, 2005, **26**, 352–364.
- 36 N. Russo, M. Toscano, A. Grand and F. Jolibois, *J. Comput. Chem.*, 1998, **19**, 989–1000.
- 37 R. Hänsel-Hertsch, M. Di Antonio and S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 279–284.
- 38 F. Zaccaria, G. Paragi and C. Fonseca Guerra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20895–20904.
- 39 M. Meyer, T. Steinke, M. Brandl, J. Sühnel and M. E. T. Al, *J. Comput. Chem.*, 2000, **22**, 109–124.
- 40 G. F. Salgado, C. Cazenave, A. Kerkour and J. L. Mergny, *Chem. Sci.*, 2015, **6**, 3314–3320.
- 41 S. Seifert, *Annu. Rev. Microbiol.*, 2018, **72**, 49–69.

- 42 C. Frascchetti, M. Montagna, L. Guarcini, L. Guidoni and A. Filippi, *Chem. Commun.*, 2014, **50**, 14767–14770.
- 43 D. Paul, A. Marchand, D. Verga, M.-P. Teulade-Fichou, S. Bombard, F. Rosu and V. Gabelica, *Analyst*, 2019, **144**, 3518–3524.
- 44 M. J. Lecours, A. Marchand, A. Anwar, C. Guetta, W. S. Hopkins and V. Gabelica, *Biochim. Biophys. Acta (BBA)-General Subj.*, 2017, **1861**, 1353–1361.
- 45 K. Fukushima and H. Iwahashi, *Chem. Commun.*, 2000, 895–896.
- 46 B. Yang, A. R. Moehlig, C. E. Frieler and M. T. Rodgers, *J. Phys. Chem. A*, 2015, **119**, 1857–1868.
- 47 S. Y. Han, S. H. Lee, J. Chung and H. Bin Oh, *J. Chem. Phys.*, 2007, **127**, 245102.
- 48 J. Jun and S. Y. Han, *Theor. Chem. Acc.*, 2017, **136**, 1–10.
- 49 J. J. Park, C. S. Lee and S. Y. Han, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 2368–2379.
- 50 P. Modrich, *Annu. Rev. Genet.*, 1991, **25**, 229–253.
- 51 J. Martens, G. Berden, C. R. Gebhardt and J. Oomens, *Rev. Sci. Instrum.*, 2016, **87**, 103108.
- 52 M. H. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, *Electron. Suppl. Mater. Chem. Sci.*, 2015, S1–S20.
- 53 A. A. Power, O. Y. Ali, M. B. Burt and T. D. Fridgen, *J. Phys. Chem. B*, 2017, **121**, 58–65.
- 54 Y. W. Nei, T. E. Akinyemi, C. M. Kaczan, J. D. Steill, G. Berden, J. Oomens and

- M. T. Rodgers, *Int. J. Mass Spectrom.*, 2011, **308**, 191–202.
- 55 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
- 56 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 57 T. Leininger, A. Nicklass, W. Küchle, H. Stoll, M. Dolg and A. Bergner, *Chem. Phys. Lett.*, 1996, **255**, 274–280.
- 58 L. Goerigk and S. Grimme, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6670–6688.
- 59 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 60 S. Sadhukhan, D. Muñoz, C. Adamo and G. E. Scuseria, *Chem. Phys. Lett.*, 1999, **306**, 83–87.
- 61 S. Andersson and M. Grüning, *J. Phys. Chem. A*, 2004, **108**, 7621–7636.
- 62 M. B. Burt and T. D. Fridgen, *J. Phys. Chem. A*, 2007, **111**, 10738–10744.
- 63 T. D. Fridgen, *J. Phys. Chem. A*, 2006, **110**, 6122–6128.
- 64 P. G. Jasien and G. Fitzgerald, *J. Chem. Phys.*, 1990, **93**, 2554–2560.
- 65 C. F. J. Böttcher, in *Theory of Electric Polarization*, Elsevier Pub. Co., Amsterdam, 1973.
- 66 R. C. Johnson, T. D. Power, J. S. Holt, B. Immaraporn, J. E. Monat, A. A. Sissoko, M. M. Yanik, A. V. Zagorodny and S. M. Cybulski, *J. Phys. Chem.*, 1996, **100**, 18875–18881.
- 67 A. W. Castleman, K. I. Peterson, B. L. Upschulte and F. J. Schelling, *Int. J. Mass Spectrom. Ion Physics.*, 1983, **47**, 203–206.
- 68 Y. Seong, S. Y. Han, S. C. Jo and H. Bin Oh, *Mass Spectrom. Lett.*, 2011, **2**, 73–75.
- 69 E. P. L. Hunter and S. G. Lias, *J. Phys. Chem. Ref. Data*, 1998, **27**, 413–656.

- 70 F. Lanucara, B. Chiavarino, M. E. Crestoni, D. Scuderi, R. K. Sinha, P. Maître and S. Fornarini, *Int. J. Mass Spectrom.*, 2012, **330**, 160–167.
- 71 M. B. Burt and T. D. Fridgen, *J. Phys. Chem. A*, 2013, **117**, 1283–1290.
- 72 C. G. Atkins, K. Rajabi, E. A. L. Gillis and T. D. Fridgen, *J. Phys. Chem. A*, 2008, **112**, 10220–10225.
- 73 R. R. Wu, B. Yang, G. Berden, J. Oomens and M. T. Rodgers, *J. Phys. Chem. B*, 2015, **119**, 2795–2805.
- 74 M. Chawla, P. Sharma, S. Halder, D. Bhattacharyya and A. Mitra, *J. Phys. Chem. B*, 2011, **115**, 1469–1484.
- 75 B. Yang and M. T. Rodgers, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16110–16120.

# **Chapter 6 An IRMPD Spectroscopic and Computational Study of Protonated Guanine-Containing Mismatched Base Pairs in the Gas Phase**

**This chapter has been published:**

R. Cheng, E. Loire, J. Martens and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2020, **22**, 2999–3007.

**Authors:**

Ruodi Cheng, Estelle Loire, Jonathan Martens and Travis D. Fridgen\*

**See Appendix C for additional results.**

**Abstract:**

Infrared multiple photon dissociation (IRMPD) spectroscopy has been used to probe the structures of the three protonated base-pair mismatches containing 9-ethylguanine (9eG) in the gas phase. Computational chemistry has been used to determine the relative energies and compute the infrared spectra of these complexes. By comparing the IRMPD spectra with the computed spectra, in all cases, it was possible to deduce that what was observed experimentally was the lowest energy computed structures. The protonated complex between 9eG and 1-methylthymine (1mT) is protonated at N7 of 9eG—the most basic site of all three bases in this study—and bound in a Hoogsteen type structure with two hydrogen bonds. The experimental IRMPD spectrum for the protonated complex between 9eG and

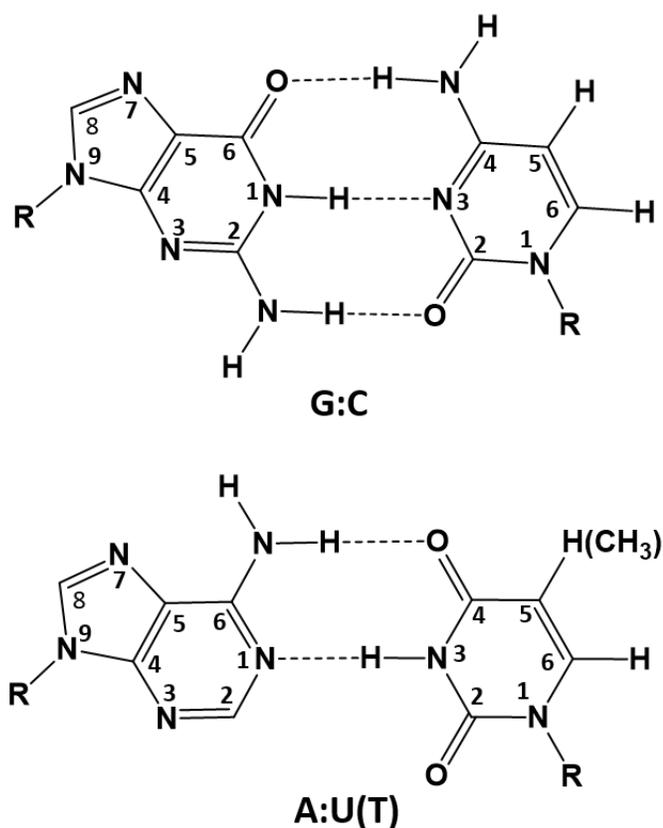
9-methyladenine (9mA) is described as arising from a combination of the two lowest energy structures, both of which are protonated at N1 of adenine and each containing two hydrogen bonds with 9eG being the acceptor of both. The protonated dimer of 9eG is protonated at N7 with an N7-H<sup>+</sup>--N7 ionic hydrogen bond. It also contains interaction between a C-H of protonated guanine and the O6 carbonyl of neutral guanine which is manifested in a slight red shift of that carbonyl stretch. The protonated 9eG/9mA structures have been previously identified by X-Ray crystallography in RNA and are contained within the protein database.

## 6.1 Introduction.

There are approximately 6 billion nucleobases composing the human genome, existing as base pairs and constituting the genes that replicate during the process of cell division. The nucleobases in DNA double-stranded helices and the folded single-stranded RNA typically form canonical base pairs by non-covalent interactions; guanine(G) matches with cytosine(C) and adenine(A) matches with thymine (T) in DNA and with uracil (U) in RNA (**Scheme 6.1**). Nearly 1 out of  $10^7$  nucleobases is estimated to occur as natural mismatches<sup>1,2</sup> which can cause mutagenesis, carcinogenesis or cell death.<sup>2-4</sup> While checking mechanisms exist for incorrect base insertion, spontaneous mutagenesis has been shown to be influenced by the mismatched base pairs (ie. GT or CA) adopting geometries other than Watson-Crick (WC) geometries,<sup>5</sup> but which mimic the shape of the WC base-pair, and go unnoticed.<sup>6</sup> Sometimes tautomerization of nucleobases can be the cause for these mismatches.<sup>6</sup> Because the frequency of mutations in DNA and RNA molecules is susceptible to the stabilities and shapes of these nucleobases' mismatches,<sup>2,7-9</sup> there is extensive scientific literature on mismatched base pairs which show that they don't follow the complementary principles discovered by Watson and Crick.<sup>10</sup>

Among all five nucleobases, guanine has been shown to have a tendency to self-assemble and form G-quartets,<sup>11-13</sup> G-quadruplexes,<sup>14-19</sup> guanine-guanine(GG) mismatches,<sup>20-22</sup> and some other guanine-containing adducts.<sup>23,24</sup> This is possibly due to guanine involved base pair mismatches and self-assemblies being more thermodynamically stable than others which are supported by the higher melting temperature of high guanine containing double-stranded helices.<sup>8</sup> Also, DNA duplexes are reported to be efficient

charge carriers due to  $\pi$ - $\pi$  interactions in their close stacked base pairs.<sup>25</sup> The efficiency of long-distance electron transfer in DNA was considered to be sensitive to guanine and G-containing mismatches because the electrons were found to be generated along with the guanine radical cation,<sup>26-28</sup> and mismatches might disrupt the integrity of  $\pi$ - $\pi$  stacking of regular base pairs leading to lower electron transfer rates.<sup>29-31</sup> However, recently there have been claims that neither guanine nor G-containing mismatches affect the rate of electron migration over long distances in DNA.<sup>31</sup>



**Scheme 6.1** Watson-Crick G:C and A:U(T) base pairs

The importance of complexes containing guanine has raised our interests in

revealing their intrinsic molecular structures and has also attracted a plethora of theoreticians to study the structures of metal cation-mediated complexes and the electrostatic interactions governing their structures. Research on Ruthenium(III) ( $\text{Ru}^{3+}$ ) containing complexes with regular base pairs and transversion mismatches, GG, AA, CC, and TT, was explored using density functional theory (DFT).<sup>22</sup> It was illustrated that the GG mismatches with two different types of  $\text{Ru}^{3+}$  complexes has the highest interaction energies compared to any other base pairs. It was also revealed that G-containing mismatches and the regular GC pair are relatively more stable than all the other base pairs both in the absence of and when complexed with  $\text{Ru}^{3+}$ . G-quartets and mixed AGAG-quartets have also been studied in the presence of proton and metal cations.<sup>13,32–35</sup> The interbase hydrogen bonds were found to be stabilized by two protons, and the  $[\text{AGAG}+2\text{H}]^{2+}$  dication is considered as two protonated GA mismatches. The two protons in  $[\text{AGAG}+2\text{H}]^{2+}$  are bound to N1 or N7 of adenine and overcome the repulsive interactions between the lone pairs on N7 of guanine and N1 or N7 of adenine.<sup>36</sup> In addition, mismatches were found to be stabilized in acidic environment.<sup>37</sup> It is crucial, therefore, to understand the influence of protonation on the structure of G-containing mismatches based on experimental evidence.

Vibrational spectroscopy, specifically infrared multiple photon dissociation (IRMPD),<sup>38–41</sup> has been used to measure the vibrational spectra of various nucleobase adducts with cations,<sup>42–46</sup> including the G-tetrads,<sup>13,47</sup> uracil complexes and self-assemblies,<sup>46,48–50</sup> i-motif structures,<sup>51–53</sup> GC base pairs<sup>54,55</sup> and many other biomolecules in the gas-phase. A recent study on alkali metal cationized 1-methylcytosine (1-mCyt) dimers revealed an asymmetric structure due to an interbase hydrogen bonding interaction

to be a major contributor to the gas phase vibrational spectra of  $K^+$ ,  $Rb^+$  and  $Cs^+$  complexes and likely even contributing to the  $Li^+$  and  $Na^+$  complexes.<sup>52</sup> Also, the gas phase IRMPD spectra of the protonated GC base pairs were found to have either the Hoogsteen structures or the less thermodynamically stable Watson-Crick structure when electrosprayed from solutions of pH 3.2 or 5.8, respectively.<sup>54</sup>

In this paper, vibrational spectroscopy was employed to uncover the structures of protonated G-containing mismatches in gas-phase. Specifically,  $(9eG:9mA)H^+$ ,  $(9eG:1mT)H^+$ , and  $(9eG:9eG)H^+$  (where 9eG = 9-ethylguanine, 9mA = 9-methyladenine, 1mT = 1-methylthymine) were studied by IRMPD spectroscopy in the fingerprint region and computational methods. Methyl- and ethyl- groups on the nucleobases block the site that would be attached to the ribose/deoxyribose in nucleic acids.

## **6.2 Methods.**

**6.2.1 Experimental Methods.** The IRMPD spectroscopy experiments in the fingerprint region were performed at two different free electron laser (FEL) facilities coupled to ion trapping mass spectrometers. The  $(9eG:9mA)H^+$  and  $(9eG:9eG)H^+$  spectra were collected at the free electron laser for infrared experiments (FELIX) in the Netherlands which is coupled to a modified 3D quadrupole ion trap mass spectrometer (Bruker AmaZon Speed ETD).<sup>56</sup> These ions were electrosprayed from the solutions described below, which were first diluted 100-fold with acetonitrile. The trapped and isolated ions were irradiated with FEL radiation scanned at  $3\text{ cm}^{-1}$  intervals with 2 pulses of tunable infrared radiation at FELIX in the  $900 - 1900\text{ cm}^{-1}$  region. At the centre laser infrarouge d'Orsay (CLIO) a

Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker Apex-Qe 7T) coupled to the FEL was used to collect the spectrum for (9eG:1mT)H<sup>+</sup>. These ions were isolated in the FTICR and irradiated with FEL radiation by for 250 ms at 5 cm<sup>-1</sup> intervals in the 900 – 1900 cm<sup>-1</sup> region.

All chemicals were purchased from Sigma-Aldrich and used without further purification. All solvents described below were a 50/50 mix of 18 MΩ•cm (Millipore) water and methanol. Solutions for the (9eG)<sub>2</sub>H<sup>+</sup> were prepared by adding two drops of 1% formic acid into 0.75 mmol L<sup>-1</sup> 9-ethylguanine solution. The (9eG:1mT)H<sup>+</sup> complex was prepared by adding 3-4 drops 1% fresh formic acid to a solution containing 0.25 mmol L<sup>-1</sup> 9eG and 1 mmol L<sup>-1</sup> 1mT. The (9eG:1mA)H<sup>+</sup> complex was prepared by adding 3-4 drops of 1% fresh formic acid into a solution containing 0.25 mmol L<sup>-1</sup> 9eG and 1 mmol L<sup>-1</sup> 1mA. Several attempts to prepare (9eG:1mU)H<sup>+</sup> with different concentrations of 9eG, 1mU, and 1% formic but were unsuccessful.

The IRMPD efficiencies (intensities) are calculated as the negative of the logarithm of the product ion intensities divided by the sum of the total ion intensities.

**6.2.2 Computational Methods.** B3LYP density functional theory and 6-31+G(d,p) split-valence basis set has been used to reliably model and compare the thermochemistries of isomeric bioorganic and bioinorganic systems with hydrogen bonds successfully.<sup>42,46,57–60</sup> Geometry optimizations and frequency calculations were performed using the Gaussian 09/16 suite of programs.<sup>61,62</sup> All B3LYP calculations an empirical dispersion correction was included using Grimme's D3 version with the original D3 damping function, B3LYPD3.<sup>63</sup> The computed vibrational frequencies were corrected by a factor of 0.975 in

the fingerprint region ( $900\text{-}1900\text{ cm}^{-1}$ ),<sup>46,48,64</sup> and convoluted with a Lorentzian profile with a full width at half max of  $15\text{ cm}^{-1}$  to compare with the experimental IRMPD spectra. For comparison, single point energy calculations were performed using 6-311+G(3df, 3pd) on all the optimized structures of B3LYP/6-31+G(d,p) and are reported as B3LYPD3/6-311+G(3df,3pd)//6-31+G(d,p). M06-2X<sup>65,66</sup> is also considered to perform well for the thermodynamic calculations of complexes containing non-covalent interactions. For comparison, M06-2X/6-31+G(d,p) was used to optimize the lowest energy structures and calculate the thermochemical parameters. The thermochemistries reported are all 298 K values and in  $\text{kJ mol}^{-1}$ . Finally, CBS-QB3 was used to compute proton affinities. No attempt was made to correct for basis set superposition error in these calculations.

### 6.3 Results and Discussion

All the complexes are named by giving the site of protonation first then the hydrogen bond interactions with the donor first and the acceptor second. For example, GN7\_GN7-TO4\_TN3-GO6 shows that the 9eG is protonated at N7 and there is the hydrogen bond from N7 of 9eG to O4 of 1mT and from N3 of 1mT to O6 of 9eG.

In all figures showing computed and experimental vibrational spectra, the bolded grey traces are experimental IRMPD spectra while the black traces are computed IR spectra for the indicated structure.

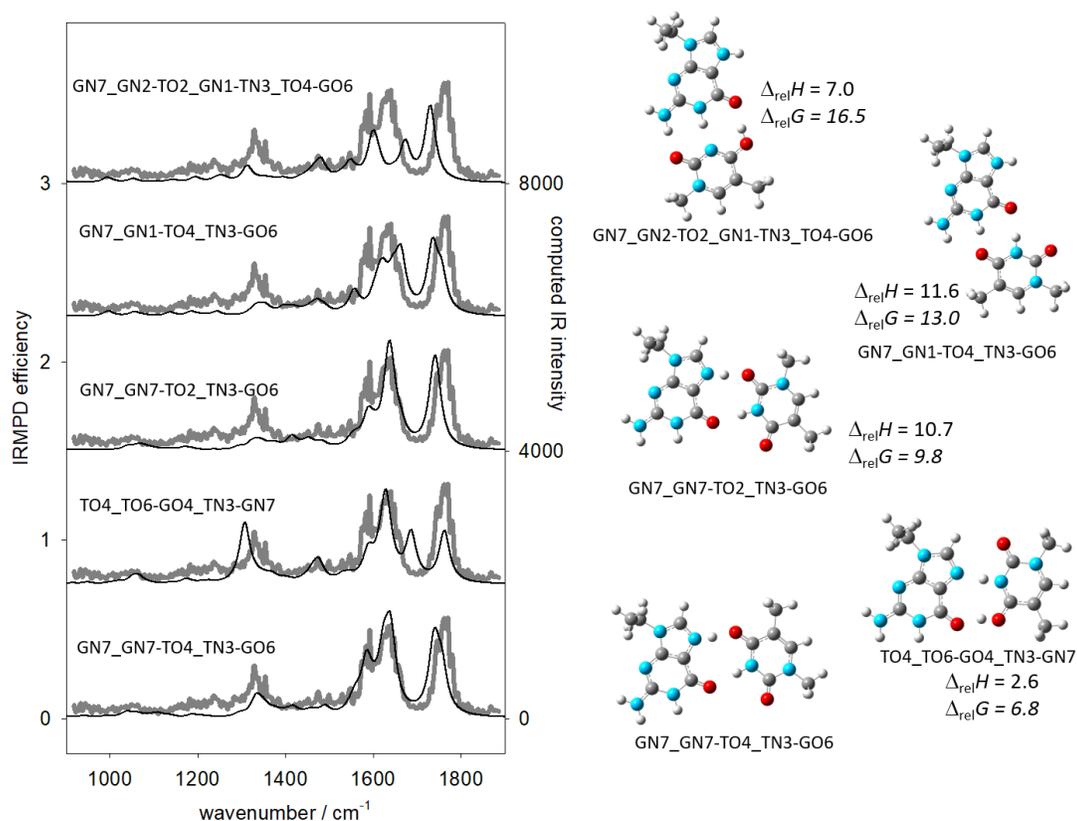
**6.3.1 (9eG:1mT)H<sup>+</sup>.** The experimental IRMPD spectrum of (9eG:1mT)H<sup>+</sup> in the fingerprint region collected at CLIO is illustrated in **Figure 6.1** and is compared with computed IR spectra of the five lowest energy isomers. The computed IR spectrum of the

lowest energy structure, GN7\_GN7-TO4\_TN3-GO6, a Hoogsteen type structure, is clearly consistent with the experimental spectrum. The observed absorption band at 1772  $\text{cm}^{-1}$  is reproduced by the predicted free C=O2 stretch of 1mT at 1757  $\text{cm}^{-1}$  and the weakly hydrogen bonded C=O6 stretch of 9eG at 1740  $\text{cm}^{-1}$ . The experimental absorption centred at 1641  $\text{cm}^{-1}$  can be attributed to three predicted vibrations at 1644, 1639, and 1666  $\text{cm}^{-1}$  which are dominated by  $\text{NH}_2$  bending of 9eG, C5=C6 stretching of 1mT, as well as hydrogen bonded C=O4 stretching, and HNC bending of 1mT. The computed TN3-GO6 and GN7-TO4 hydrogen bonds are predicted to be 1.90 and 1.48 Å, respectively. The 1593  $\text{cm}^{-1}$  shoulder is well reproduced by the computed 1586  $\text{cm}^{-1}$  band which is composed of predominantly  $\text{NH}_2$ , HN1C2, and HN7C8 bending in 9eG. The weaker absorption at 1328  $\text{cm}^{-1}$  is also very nicely reproduced by the calculations and consists of numerous complicated stretching and bending motions involving ring atoms of both bases.

The second-lowest energy isomer, TO4\_TO4-GO6\_TN3-GN7, is protonated at O4 of 1mT despite the significantly higher proton affinity of 9eG than that of O4 of 1mT, by some 103.9  $\text{kJ mol}^{-1}$ . This isomer is only 6.8  $\text{kJ mol}^{-1}$  higher in Gibbs energy than the lowest energy isomer. The lower than expected relative energy of this complex, due to being 1mT protonated, is most likely due to the strong ion-dipole and ion-induced dipole interactions formed between  $1\text{mTH}^+$  and 9eG.<sup>55,67,68</sup> The main disagreement between the experimental IRMPD spectrum and the computed spectrum for this isomer is the hydrogen bonded C=O6 stretch of 9eG predicted at 1687  $\text{cm}^{-1}$ , but which is not observed experimentally. The hydrogen bond to this carbonyl oxygen is computed to be 1.49 Å.

The GN7\_GN7-TO2\_TN3-GO6 isomer differs from the lowest energy isomer by a flip of 1mT and is computed to be 9.8  $\text{kJ mol}^{-1}$  higher in Gibbs energy. Spectroscopically,

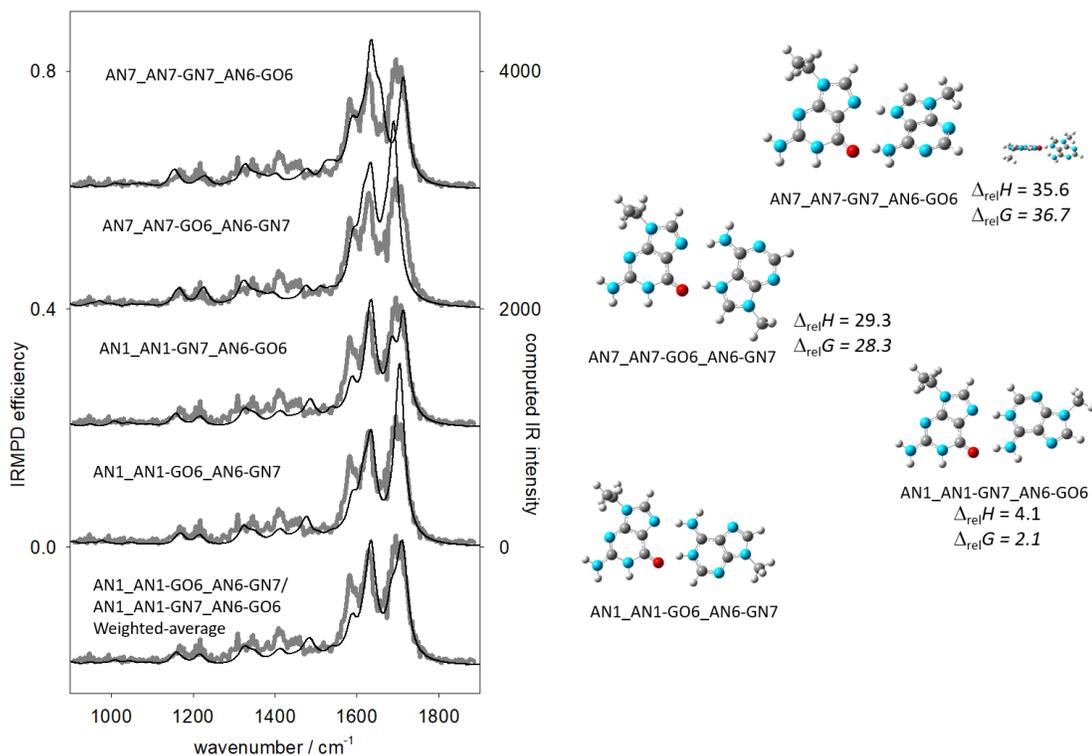
it cannot be ruled out due to a very similar predicted IRMPD spectrum to the lowest energy isomer. However, that it is higher in energy than the second lowest energy isomer which is not contributing to the experimental spectrum, it is unlikely that this one is contributing either. Two other higher energy structures are also shown in **Figure 6.1** are Watson-Crick type structures and are not consistent with the experimental spectrum. Other higher energy structures with their computed spectra are presented in **Figure S1** and compared with the (9eG:1mT)H<sup>+</sup> experimental IRMPD spectrum.



**Figure 6.1** Comparison of the experimental IRMPD spectrum (grey trace) for (G:T)H<sup>+</sup> and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure GN7\_GN7-TO4\_TN3-GO6 are also shown (also computed using B3LYP/6-31+G(d,p)), kJ mol<sup>-1</sup>.

**6.3.2 (9eG:9mA)H<sup>+</sup>.** In **Figure 6.2**, the IRMPD spectrum of (9eG:9mA) is compared to the four lowest energy structures. Interestingly, the four lowest energy structures are all 9mA protonated, despite the significantly larger proton affinity of 9eG (Table 1). The two lowest energy structures, AN1\_AN1-GO6\_AN6-GN7 and AN1\_AN1-GN7\_AN6-GO6 are complexes where the Hoogsteen face of 9eG is interacting with the WC face of 9mA; both are protonated at N1 of 9mA which is predicted to have a proton affinity almost 25 kJ mol<sup>-1</sup> lower than N7 of guanine. This type of structural oddity has been discussed before with respect to protonated complexes containing one high dipole moment monomer<sup>67,68</sup> and the protonated 1-methylcytosine/9-ethylguanine complex, (9eG:1mC)H<sup>+</sup>.<sup>55</sup> In these species the proton is covalently bound to the lower proton affinity monomer and then energy deficit is made up by forming a strong ion-dipole and/or ion-induced dipole interaction. In the present case, 9eG has a dipole moment of 7.36 D, almost three times that of 9mA (Table 1). The 25 kJ mol<sup>-1</sup> deficit by protonating 9mA over 9eG is more than made up by the very strong ion-dipole interaction between 9mAH<sup>+</sup> and 9eG. The larger polarizability of 9eG than 9mA (18.4 Å<sup>3</sup> vs 15.7 Å<sup>3</sup>) means that a stronger ion-induced dipole interaction would favour the 9mA protonated structure over the 9eG protonated one. Unlike (9eG:1mC)H<sup>+</sup> where the 9eG protonated complex was an optimized local minimum, in the present example with (9eG:9mA)H<sup>+</sup> it was not possible to locate a local minimum similar to AN1\_AN1-GN7\_AN6-GO6 where 9eG was protonated at N7 instead of 9mA being protonated at N1. Optimization calculations on the GN7 protonated complex were repeated

using B3LYPD3/6-311+G(3df,3pd) and M06-2X/6-31+G(d,p), but were unsuccessful, the proton shifted resulting in the AN1\_AN1-GN7\_AN6-GO6 structure.



**Figure 6.2** Comparison of the experimental IRMPD spectrum (grey trace) for (G:A)H<sup>+</sup> and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for four different isomers. The 298 K enthalpies and Gibbs energies relative to structure AN1\_AN1-GO6\_AN6-GN7 are also shown (also computed using B3LYP/6-31+G(d,p)), kJ mol<sup>-1</sup>.

The AN1\_AN1-GO6\_AN6-GN7 and AN1\_AN1-GN7\_AN6-GO6 structures differ by only 2.1 kJ mol<sup>-1</sup> in Gibbs energy and neither can be ruled out spectroscopically, both are consistent with the experimental spectrum. At the bottom of **Figure 6.2**, the weighted average based on the difference in Gibbs energy of these two lowest energy structures is compared with the experimental IRMPD spectrum for (9eG:9mA)H<sup>+</sup> and better reproduces

the spectrum. The most pronounced experimental features in the IRMPD spectrum are in the 1500-1800  $\text{cm}^{-1}$  region. The band at 1705  $\text{cm}^{-1}$  is well reproduced by 9eG C=O stretch combined with 9mA  $\text{NH}_2$  bending. The band at 1630  $\text{cm}^{-1}$  is the  $\text{NH}_2$  bend combined with C2-N2 stretching, both of 9eG. The 1580  $\text{cm}^{-1}$  feature belongs predominantly to 9eG C2-N3/N3-C4 stretching. This latter band is better resolved experimentally than is predicted by the calculations.

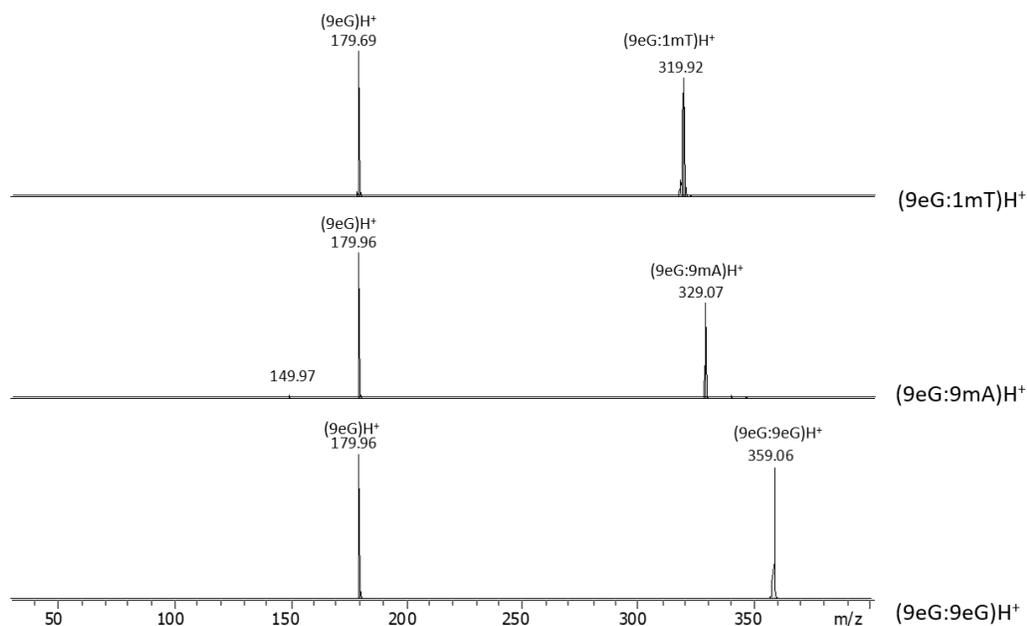
The next two structures shown in **Figure 6.2** are 28.3 and 36.7  $\text{kJ mol}^{-1}$  higher in energy, and therefore are unlikely to be present in great abundance. Certainly, they cannot be ruled out as contributing to the experimental spectrum based on a comparison of their computed spectra. The lowest energy structure composed of canonical bases (ie. not tautomeric) with an N7 protonated 9eG, GN7\_GN1-AN1\_GO6-AN6, is 54  $\text{kJ mol}^{-1}$  higher in Gibbs energy than the lowest energy structure (Figure S2) and its computed IR spectrum is not consistent with the experimental spectrum. Other higher energy structures are compared to the experimental IRMPD spectrum in Figures S2 and S3; none of their computed IR spectra reproduce the observed spectrum.

The IRMPD dissociation mass spectra for (9eG:1mT) $\text{H}^+$ , (9eG:9mA) $\text{H}^+$ , and (9eG:9eG) $\text{H}^+$  are presented in **Figure 6.3**. In all cases, the main dissociation product is protonated 9eG at  $m/z$  180 which is expected based on the proton affinities and gas-phase basicities of the bases (see computed values in Table 1), with N7 of 9eG having the highest proton affinity. Interestingly, however, for (9eG:9mA) $\text{H}^+$  there is a small, but not insignificant amount of protonated 9mA. Based on the difference in gas basicities, for the two bases, the intensity of protonated 9mA should be no more than  $2 \times 10^{-5}$  compared to that of protonated 9eG. Experimentally, the ratio is 0.02:1. This is similar to the anomaly

reported in previous work for the dissociation of protonated guanine-cytosine base pairs where protonated cytosine was observed in a significantly higher abundance than expected.<sup>54,55,69–71</sup> For (9eG:1mC)H<sup>+</sup>, the observed 1mCH<sup>+</sup>:9eGH<sup>+</sup> ratio was 0.3:1 despite an expected ratio of only 0.006:1 depending on the computed gas basicities. This was explained by the dynamics of dissociation of the energized (9eG:1mC)H<sup>+</sup> system. (9eG:1mC)H<sup>+</sup>, is protonated at N3 of 1mC with a low-lying energy barrier for proton transfer to 9eG. The surface is very shallow around the minimum energy structure in both the dissociation and proton transfer degrees of freedom. The proton transfer energy barrier, however, grows significantly as distance between the two bases increases during dissociation. In order to transfer the proton from 1mC to 9eG, the energized complex must adopt a configuration where the distance between the two bases is similar to that of the minimum energy structure to make proton transfer energetically feasible. It is expected that a similar dynamics problem is at play in the present system such that much more 9mAH<sup>+</sup> is observed in the dissociation of (9eG:1mA)H<sup>+</sup> than expected based on gas basicities.

**Table 6.1.** 298 K proton affinities (PA) and gas basicities (GB) for 9-ethylguanine, 9-methyladenine, and 1-methylthymine. Protonation site is indicated in parentheses. a: CBS-QB3 b: B3LYP/6-31+G(d,p)

	PA <sup>a</sup> / kJ mol <sup>-1</sup>	GB <sup>a</sup> / kJ mol <sup>-1</sup>	dipole moment <sup>b</sup> / D of neutral base	Polarizability <sup>b</sup> / Å <sup>3</sup> of neutral base
9-ethylguanine (N7)	974.3	943.3	7.36	18.4
9-methyladenine (N1)	949.4	916.1	2.67	15.7
9-methyladenine (N7)	924.1	893.1		
1-methylthymine (O4)	870.4	847.3	4.97	13.7
1-methylthymine (O2)	852.2	832.1		

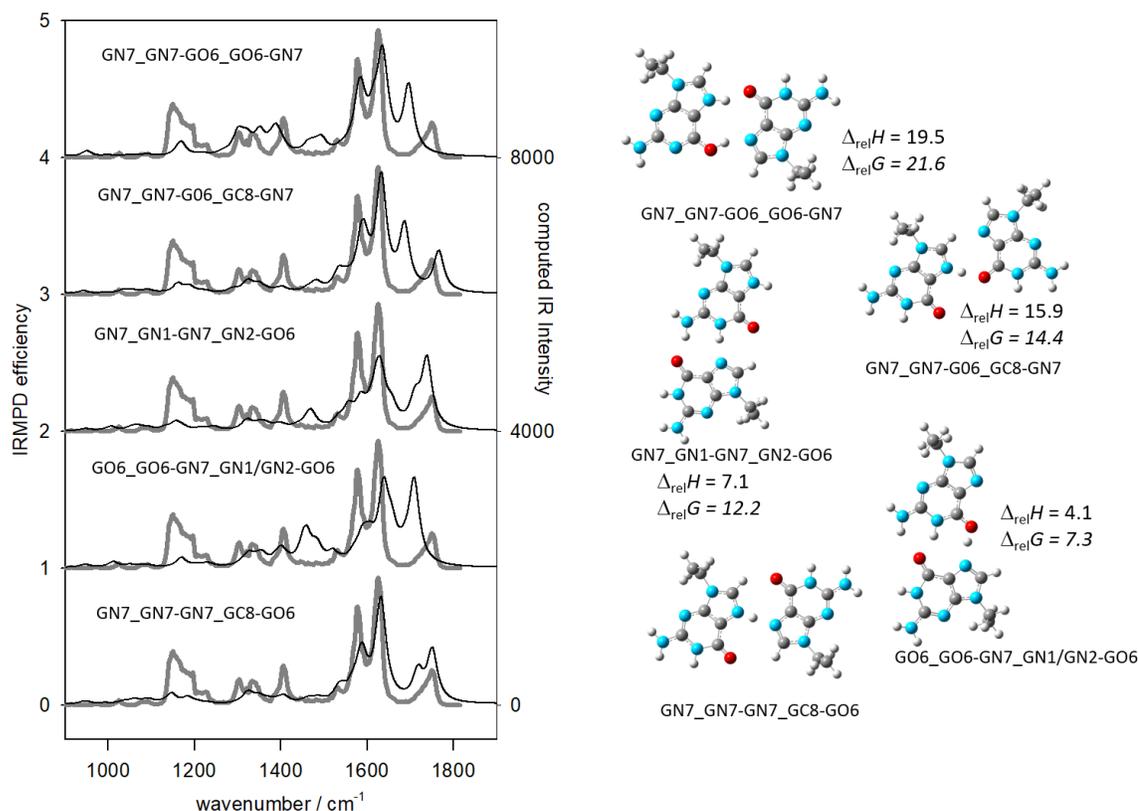


**Figure 6.3** Infrared multiple photon dissociation mass spectra of three complexes at their maxima in dissociation in the fingerprint region. The  $m/z$  and name of the precursor ions and fragments are also shown.

**6.3.3 (9eG:9eG)H<sup>+</sup>.** The IRMPD spectra for the (9eG:9eG)H<sup>+</sup> complex and the computed IR spectra for the 5 lowest energy structures are compared in **Figure 6.4**. The lowest energy structure for (9eG:9eG)H<sup>+</sup> (GN7\_GN7-GN7\_GC8-GO6) is one where the Hoogsteen faces of both 9eG are interacting but with only one *classical* hydrogen bond. It is protonated at N7 of one 9eG with a hydrogen bond to the N7 of the other 9eG and a hydrogen bonding-type interaction between C8 of the protonated 9eG and the O6 of the other. The computed spectrum for this structure is in excellent agreement with the experimental IRMPD spectrum. The predicted free C=O6 stretch of the protonated 9eG at 1751 cm<sup>-1</sup> and the red-shifted shoulder predicted at 1718 cm<sup>-1</sup> due to the C=O6 stretch of the neutral 9eG

interacting with the C8-H is in excellent agreement with the experimental band at 1753 and the unresolved shoulder to the red of that band. The weak interaction, with a C-H—O=C bond distance of 2.25 is enough to slightly shift the C=O stretch to slightly lower energy. The strong absorptions at 1628 and 1579  $\text{cm}^{-1}$  agree well with the  $\text{NH}_2$  bending absorptions predicted to occur at 1635, 1590, and 1584  $\text{cm}^{-1}$ . The observed bands at lower energy also agree well with the positions of predicted bands for modes involving ring stretches. This is the first reported dimeric base pair containing only a single *classical* hydrogen bond but not follow the regular Watson-Crick, Hoogsteen, or Wobble base pair principle, with the confident conclusion supported by both thermodynamic and spectroscopic results.

The second lowest energy structure, GO6\_GO6-GN7\_GN1/GN2-GO6, has O6 protonated 9eG interacting via its WC face with the Hoogsteen face of neutral 9eG and is computed to be higher in Gibbs energy by 7.3  $\text{kJ mol}^{-1}$ , and clearly its calculated IR spectrum is not in good agreement with the experimental spectrum. Foremost, there is the absence of a free C=O in the computed structure, instead having C=O stretches computed to be at 1709 and 1687  $\text{cm}^{-1}$  due to hydrogen bonding and protonation. The third lowest energy structure, GN7\_GN1-GN7\_GN2-GO6, is also one where N7 protonated 9eG interacts via its WC face to the Hoogsteen face of neutral 9eG and cannot be completely ruled out by spectroscopic means, but is 12.2  $\text{kJ mol}^{-1}$  higher in Gibbs energy than the lowest energy structure. Its computed infrared spectrum is nearly as good a match to the experimental spectrum. The next two higher energy structures, the highest containing an N7-protonated N1 to O6 tautomer, are clearly not a good match to the experimental spectrum. Other structures are shown in Figure S4-S5. None of these higher energy conformers can reproduce the experimental IRMPD spectrum.



**Figure 6.4** Comparison of the experimental IRMPD spectrum (grey trace) for  $(G:G)H^+$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure GN7\_GN7-GN7\_GC8-GO6 are also shown (also computed using B3LYP/6-31+G(d,p)), kJ mol<sup>-1</sup>.

**6.3.4 Computed Energies Comparison.** Relative thermochemistries computed using B3LYPD3/6-31+G(d,p), M06-2X/6-31+G(d,p), and B3LYPD3/6-311+G(3df,3pd)//6-31+G(d,p) are compared in Tables S1, S2 and S3 for the lowest energy structures of  $(9eG:1mT)H^+$ ,  $(9eG:9mA)H^+$ , and  $(9eG:9eG)H^+$ , respectively. All computed energies provide a similar picture as those using B3LYP/6-31+G(d,p). One notable exception is in the TO4\_TO6-GO4\_TN3-GN7 structure for the  $(9eG:1mT)H^+$  complex which is nearly

isoenergetic with the lowest energy GN7\_GN7-TO4\_TN3-GO6 structure using M06-2X/6-31+G(d,p). However, TO4\_TO6-GO4\_TN3-GN7 is ruled out spectroscopically as seen in **Figure 6.1**.

## 6.4 Conclusions

An IRMPD spectroscopic and computational study was undertaken to probe the structures of protonated guanine-containing mismatch complexes, specifically of (9eG:1mT)H<sup>+</sup>, (9eG:9mA)H<sup>+</sup> and (9eG:9eG)H<sup>+</sup>. For (9eG:1mT)H<sup>+</sup> and (9eG:9eG)H<sup>+</sup>, the lowest energy structures were sufficient to explain the IRMPD spectra, while for (9eG:9mA)H<sup>+</sup> the two lowest energy structures were only 2.1 kJ mol<sup>-1</sup> apart in Gibbs energy and their weighted average is consistent with the IRMPD spectrum.

While the DNA bases have pK<sub>a</sub> values which indicate that they are unlikely to be protonated, at physiological pH, it has also been determined that their chemical environments in nucleic acids—and potentially through interactions with metal cations—shift their pK<sub>a</sub>'s to the physiological regime,<sup>72,73</sup> raising the interest toward understanding protonation of the nucleobases in nucleic acids.<sup>74-79</sup> It would be interesting to compare our model protonated mismatch complexes of guanine to those known to occur in nucleic acids. Atomic coordinates for proteins and RNA contained within the protein data base, PDB, are from X-Ray crystallography measurements which do not all contain positions of hydrogens or protons.<sup>75</sup> However, computational chemistry has been used to compare with the potential structures from PDB for protonated base pairs which have been identified in RNA.<sup>75</sup> In a set of 19 base pairs where protonation is thought to occur,<sup>75</sup> two of them are N1 protonated adenine bound to guanine, identical to the two structures found to be the

lowest-energy structures in this study and to be consistent with the experimental vibrational spectrum. This is the first infrared spectroscopic evidence for the occurrence of a self-assembled protonated base pair, observed in the gas phase which has also been proposed to exist in cells.

While two protonated guanine dimers were also identified in the PDB, they were bound via the Hoogsteen face of neutral guanine to the sugar face of N3 protonated guanine, unlike that found in the present study. However, it is worth mentioning that the crystal structure of 9-ethylguanine hemihydrochloric acid<sup>80,81</sup> was found to exist as protonated dimers such as that found here in the gas phase, albeit without the C-H---O=C interaction that the present spectroscopic evidence shows exists at least in the gas phase.

Only one protonated guanine (N7) complex with uracil has been identified<sup>74</sup> but the hydrogen bond is between the N2 of guanine and O4 of uracil, very different than the structure concluded to exist for (9eG:1mT)H<sup>+</sup>. Given that the (9eG:9mA)H<sup>+</sup> identified in this study and the (9eG:1mC)H<sup>+</sup> identified in a recent publication from the protein data bank, it would be interesting to see if (9eG:1mT)H<sup>+</sup> and (9eG:9eG)H<sup>+</sup> complexes like the ones found here will turn up in RNA.

In a previous study spectroscopic study from this group which included the gaseous protonated complex of 9-ethylguanine and 1-methylcytosine, (9eG:1mC)H<sup>+</sup>, the lowest energy structure which was also consistent with the experimental vibrational spectrum was the Hoogsteen complex with the cytosine protonated at N3.<sup>55</sup> A protonated complex with the same structure, between guanine and cytosine, has also been proposed to occur in RNA in the protein data base.<sup>75</sup>

## **Acknowledgements**

We are grateful for the excellent support from the FELIX and CLIO technical staff. This research was enabled in part by support provided by ACENET, Compute Ontario, Westgrid, and Compute Canada. Financial support from NSERC, Bruker, CFI, and Memorial University is greatly appreciated.

## References

- 1 J. Westbrook, A. Smithers and T. Herrick, *J. Educ. Teach.*, 2010, **36**, 251–256.
- 2 A. F. El-Yazbi, A. Wong and G. R. Loppnow, *Anal. Chim. Acta*, 2017, **994**, 92–99.
- 3 T. Lindahl, *Nature*, 1993, **362**, 709–715.
- 4 B. N. Ames, L. S. Gold and W. C. Willett, *Proc. Natl. Acad. Sci.*, 1995, **92**, 5258 LP – 5265.
- 5 K. Bebenek, L. C. Pedersen and T. A. Kunkel, *Proc. Natl. Acad. Sci.*, 2011, **108**, 1862–1867.
- 6 W. Wang, H. W. Hellinga and L. S. Beese, *Proc. Natl. Acad. Sci.*, 2011, **108**, 17644–17648.
- 7 M. T. Russo, G. De Luca, P. Degan and M. Bignami, *Mutat. Res. Mol. Mech. Mutagen.*, 2007, **614**, 69–76.
- 8 F. H. Martin and J. Farecd Aboul-ela, David Koh and Ignacio Tinoco, *Nucleic Acids Res.*, 1985, **13**, 4811–4824.
- 9 T. A. Kunkel and D. A. Erie, *DNA Repair Cancer*, 2005, **74**, 681–710.
- 10 J. D. Watson and F. H. Crick, *Nature*, 1953, **171**, 3–12.
- 11 M. Meyer, M. Brandl and J. Sühnel, *J. Phys. Chem. A*, 2001, **105**, 8223–8225.
- 12 M. Meyer, A. Hocquet and J. Sühnel, *J. Comput. Chem.*, 2005, **26**, 352–364.
- 13 M. Azargun and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25778–25785.
- 14 G. F. Salgado, C. Cazenave, A. Kerkour and J. L. Mergny, *Chem. Sci.*, 2015, **6**, 3314–3320.
- 15 M. Azargun, Y. Jami-Alahmadi and T. D. Fridgen, *Phys. Chem. Chem. Phys.*,

- 2017, **19**, 1281–1287.
- 16 J. Gu and J. Leszczynski, *J. Phys. Chem. A*, 2002, **106**, 529–532.
- 17 S. Seifert, *Annu. Rev. Microbiol.*, 2018, **72**, 49–69.
- 18 F. Zaccaria, G. Paragi and C. Fonseca Guerra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20895–20904.
- 19 M. Azargun, P. J. Meister, J. W. Gauld and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2019, **21**, 15319–15326.
- 20 J. Li, H. He, X. Peng, M. Huang, X. Zhang and S. Wang, *Anal. Sci.*, 2015, **31**, 663–667.
- 21 N. Peyret, P. A. Seneviratne, H. T. Allawi and J. SantaLucia, *Biochemistry*, 1999, **38**, 3468–3477.
- 22 D. Das, A. Dutta and P. Mondal, *Comput. Theor. Chem.*, 2015, **1072**, 28–36.
- 23 A. Calzolari, R. Di Felice, E. Molinari and A. Garbesi, *Phys. E Low-dimensional Syst. Nanostructures*, 2002, **13**, 1236–1239.
- 24 J. T. Davis and G. P. Spada, *Chem. Soc. Rev.*, 2007, **36**, 296–313.
- 25 A. R. Arnold, M. A. Grodick and J. K. Barton, *Cell Chem. Biol.*, 2016, **23**, 183–197.
- 26 B. Giese, *Annu. Rev. Biochem.*, 2002, **71**, 51–70.
- 27 B. Giese, B. Giese, S. Wessely and S. Wessely, *Angew. Chem. Int. Ed. Engl.*, 2000, **39**, 3490–3491.
- 28 B. Giese and M. Spichthy, *ChemPhysChem*, 2002, **1**, 195–198.
- 29 J. D. Slinker, N. B. Muren, S. E. Renfrew and J. K. Barton, *Nat. Chem.*, 2011, **3**, 228.

- 30 S. O. Kelley, N. M. Jackson, M. G. Hill and J. K. Barton, *Angew. Chemie Int. Ed.*, 1999, **38**, 941–945.
- 31 P. Dauphin-Ducharme, N. Arroyo-Currás and K. W. Plaxco, *J. Am. Chem. Soc.*, 2019, **141**, 1304–1311.
- 32 J. R. Williamson, M. K. Raghuraman and T. R. Cech, *Cell*, 1989, **59**, 871–880.
- 33 D. Bhattacharyya, G. Mirihana Arachchilage and S. Basu, *Front. Chem.*, 2016, **4**, 38.
- 34 J. T. Davis, *Angew. Chemie - Int. Ed.*, 2004, **43**, 668–698.
- 35 S. N. Georgiades, N. H. A. Karim, K. Suntharalingam and Ramon Vilar, *Angew. Chemie. Int. Ed.*, 2010, **49**, 4020–4034.
- 36 D. A. Megger, P. M. Lax, J. Paauwe, C. Fonseca Guerra and B. Lippert, *J. Biol. Inorg. Chem.*, 2018, **23**, 41–49.
- 37 T. Brown, G. A. Leonard, E. D. Booth and G. Kneale, *J. Mol. Biol.*, 1990, **212**, 437–440.
- 38 N. C. Polfer, *Chem. Soc. Rev.*, 2011, **40**, 2211–2221.
- 39 L. Jašíková and J. Roithová, *Chem. – A Eur. J.*, 2018, **24**, 3374–3390.
- 40 T. D. Fridgen, *Mass Spectrom. Rev.*, 2009, **28**, 586–607.
- 41 M. B. Burt and T. D. Fridgen, *Eur. J. Mass Spectrom.*, 2012, **18**, 235–250.
- 42 J. Gao, G. Berden, M. T. Rodgers and J. Oomens, *Phys. Chem. Chem. Phys.*, 2016, **18**, 7269–7277.
- 43 H. U. Ung, A. R. Moehlig, R. A. Kudla, L. J. Mueller, J. Oomens, G. Berden and T. H. Morton, *Phys. Chem. Chem. Phys.*, 2013, **15**, 19001–19012.
- 44 J. Y. Salpin, S. Guillaumont, J. Tortajada, L. MacAleese, J. Lemaire and P. Maitre,

- ChemPhysChem*, 2007, **8**, 2235–2244.
- 45 Y. W. Nei, T. E. Akinyemi, C. M. Kaczan, J. D. Steill, G. Berden, J. Oomens and M. T. Rodgers, *Int. J. Mass Spectrom.*, 2011, **308**, 191–202.
- 46 B. Power, V. Haldys, J. Y. Salpin and T. D. Fridgen, *Int. J. Mass Spectrom.*, 2015, **378**, 328–335.
- 47 C. Fraschetti, M. Montagna, L. Guarcini, L. Guidoni and A. Filippi, *Chem. Commun.*, 2014, **50**, 14767–14770.
- 48 R. Cheng, V. E. Rose, B. Power and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2017, **20**, 572–580.
- 49 A. A. Power, O. Y. Ali, M. B. Burt and T. D. Fridgen, *J. Phys. Chem. B*, 2017, **121**, 58–65.
- 50 E. A. L. Gillis, K. Rajabi and T. D. Fridgen, *J. Phys. Chem. A*, 2009, **113**, 824–832.
- 51 F. Rosu, V. Gabelica, L. Joly, G. Grégoire and E. De Pauw, *Phys. Chem. Chem. Phys.*, 2010, **12**, 13448–13454.
- 52 R. Cheng, E. Loire and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2019, **1**, 11103–11110.
- 53 B. Yang, R. R. Wu, G. Berden, J. Oomens and M. T. Rodgers, *J. Phys. Chem. B*, 2013, **117**, 14191–14201.
- 54 A. F. Cruz-Ortiz, M. Rossa, F. Berthias, M. Berdakin, P. Maitre and G. A. Pino, *J. Phys. Chem. Lett.*, 2017, **8**, 5501–5506.
- 55 R. Cheng, J. Martens and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, submitted.
- 56 J. Martens, G. Berden, C. R. Gebhardt and J. Oomens, *Rev. Sci. Instrum.*, 2016, **87**,

- 103108.
- 57 M. Meyer, T. Steinke, M. Brandl, J. Sühnel and M. E. T. Al, *J. Comput. Chem.*, 2000, **22**, 109–124.
- 58 E. L. Zins, S. Rochut and C. Pepe, *J. Mass Spectrom.*, 2009, **44**, 40–49.
- 59 B. D. Linfood, A. Le Donne, D. Scuderi, E. Bodo and T. D. Fridgen, *Eur. J. Mass Spectrom. (Chichester, Eng.)*, 2019, **25**, 133–141.
- 60 J. Oomens, A. R. Moehlig and T. H. Morton, *J. Phys. Chem. Lett.*, 2010, **1**, 2891–2897.
- 61 M. H. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, *Electron. Suppl. Mater. Chem. Sci.*, 2015, S1–S20.
- 62 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson and H. Nakatsuji, *Revis. A*.
- 63 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 64 Y. Jami-Alahmadi and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2016, **18**, 2023–2033.
- 65 M. Walker, A. J. A. Harvey, A. Sen and C. E. H. Dessent, *J. Phys. Chem. A*, 2013, **117**, 12590–12600.
- 66 E. G. Hohenstein, S. T. Chill and C. D. Sherrill, *Methods*, 2008, **4**, 1996–2000.
- 67 T. D. Fridgen, *J. Phys. Chem. A*, 2006, **110**, 6122–6128.
- 68 M. B. Burt and T. D. Fridgen, *J. Phys. Chem. A*, 2007, **111**, 10738–10744.
- 69 J. J. Park and S. Y. Han, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 846–854.

- 70 Y. Seong, S. Y. Han, S. C. Jo and H. Bin Oh, *Mass Spectrom. Lett.*, 2011, **2**, 73–75.
- 71 J. J. Park, C. S. Lee and S. Y. Han, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 2368–2379.
- 72 B. Lippert, *Chem. Biodivers.*, 2008, **5**, 1455–1474.
- 73 C. L. Tang, E. Alexov, A. M. Pyle and B. Honig, *J. Mol. Biol.*, 2007, **366**, 1475–1496.
- 74 A. Halder, S. Bhattacharya, A. Datta, D. Bhattacharyya and A. Mitra, *Phys. Chem. Chem. Phys.*, 2015, **17**, 26249–26263.
- 75 M. Chawla, P. Sharma, S. Halder, D. Bhattacharyya and A. Mitra, *J. Phys. Chem. B*, 2011, **115**, 1469–1484.
- 76 A. Halder, S. Vemuri, R. Roy, J. Katuri, D. Bhattacharyya and A. Mitra, *ACS Omega*, 2019, **4**, 699–709.
- 77 A. Halder, S. Halder, D. Bhattacharyya and A. Mitra, *Phys. Chem. Chem. Phys.*, 2014, **16**, 18383–18396.
- 78 J. L. Wilcox, A. K. Ahluwalia and P. C. Bevilacqua, *Acc. Chem. Res.*, 2011, **44**, 1270–1279.
- 79 G. B. Goh, J. L. Knight and C. L. Brooks, *J. Chem. Theory Comput.*, 2013, **9**, 935–943.
- 80 G. S. Mandel and R. E. Marsh, *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.*, 1975, **B31**, 2862–2867.
- 81 R. Taylor and O. Kennard, *J. Mol. Struct.*, 1982, **78**, 1–28.

## Chapter 7 Summary and Outlook

In this work, biomolecules containing non-covalent interactions were studied using mass spectrometry, infrared multiphoton dissociation (IRMPD) spectroscopy, sustained off-resonance irradiation collision-induced dissociation (SORI-CID) in conjunction with DFT computations. Theoretical methods were used to calculate the electronic structures, enthalpies, Gibbs energies, polarizabilities, gas basicities, proton affinities and binding energies of many different isomers of the various complexes studied in this thesis. The calculations provide the structural information and relative thermochemical properties of these complexes, and the insightful analysis of the experimental results. B3LYP/6-31+G(d,p) was employed as the primary computational method, because of its cost-efficiency, to perform structural optimizations as well as IR frequency and thermodynamic calculations. Even though the thermostability trends were found to be slightly different from the results computed by other methods in few cases—especially for higher energy isomers—the global minima determined by B3LYP/6-31+G(d,p) were identical to those computed by other higher-level computational methods in all cases. SORI-CID and IRMPD provided the fragmentation pathway and product ratios of ionic complexes experimentally. The vibrational bands in IRMPD spectra, such as C=O, O-H, and N-H stretches, were observed to be characteristic of their binding properties and bonding environments, showing that IRMPD spectroscopy is an experimental technique suitable for determining the structures of ionic complexes.

Self-assembled uracil (U) complexes with  $\text{Ca}^{2+}$  were studied in Chapter 3 following previous experimental work on dissociation and thermal stability of  $\text{U}_n\text{Ca}^{2+}$  ( $n=4-$

14) clusters.<sup>1</sup>  $U_4Ca^{2+}$ ,  $U_5Ca^{2+}$ , and  $U_6Ca^{2+}$  were all revealed to be composed of both canonical and tautomerized nucleobases, which were not proposed in any previous works on these large complexes. In the lowest energy isomers of three molecules, tautomerized uracil shows the hydrogen on N3 tautomerized to O4, called a keto-enol tautomer. This discovery is consistent with the previous predictions on the lowest energy uracil monomer interacting with divalent metal cations.<sup>2</sup> There are other isomers with similar energetic properties (several  $\text{kJ mol}^{-1}$  higher Gibbs energy) that can not be ruled out spectroscopically or thermodynamically in all three ionic complexes. Also, higher Gibbs energy isomers ( $>30 \text{ kJ mol}^{-1}$ ) containing free carbonyl stretch may contribute to the observed weak absorptions at  $1800 \text{ cm}^{-1}$  in the fingerprint region of IRMPD spectra. These higher energy isomers are constituted by canonical uracil only, and thus the activation energies needed to form these structures are presumably much lower than the respective lowest energy structures. As  $U_6Ca^{2+}$  is considered as the inner core of more massive clusters, larger uracil self-assemblies can be designed by adding uracils to this center.

The research in Chapter 3 disclosed the presence of both tautomerized and canonical uracil in the  $Ca^{2+}$ -uracil complexes. Further investigation on nucleobase self-assemblies with metal cations can elucidate more about the effects of significant abundant cations on the structures of these bio-complexes. In Chapters 4, 5 and 6, nucleobase self-assemblies were investigated by having either an ethyl- or methyl- group blocking the site that would be attached to the ribose/deoxyribose in nucleic acids.

The research in Chapter 4 focuses on exploring the structures of 1-methylcytosine dimer affected by alkali metal cations. Unlike uracil, cytosine tautomers are accessible at room temperature. Previously, cytosine has been concluded to be stabilized as a canonical

monomer when it interacts with alkali metal cation.<sup>3,4</sup> In Chapter 4, 1-methylcytosine (1mC) dimer with  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Rb}^+$  and  $\text{Cs}^+$  has been discovered to exist as two conformers, symmetric anti-parallel (SAP) and asymmetric anti-parallel (AAP) geometries, the latter was not indicated by previous work on cytosine dimer with  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ .<sup>3</sup> Spectroscopically, there are the increasingly diffuse and intense shoulder absorptions of symmetric and asymmetric  $\text{NH}_2$  stretches in the  $3800\text{-}2800\text{ cm}^{-1}$  region from  $\text{Li}^+$  to  $\text{Cs}^+$  complexes. It indicates the presence of red-shifted N-H stretches. However, the red-shifted shoulders do not belong to the previously reported SAP structure. As AAP isomers contain the interbase hydrogen bonds, hydrogen-bonded amino groups can be the reason for the shoulder absorptions. Thermochemical results show that AAP structures in  $\text{K}^+$ ,  $\text{Rb}^+$  and  $\text{Cs}^+$  complexes are isoenergetic or lower in energy than the SAP structures. In conclusion, SAP structure is the predominant isomer for  $\text{Li}^+$  and  $\text{Na}^+$  complexes; however, the AAP conformer is the major contributor to  $\text{K}^+$ ,  $\text{Rb}^+$  and  $\text{Cs}^+$  cases.

I-motif structures (protonated cytosine dimer) have been observed *in vivo*.<sup>5</sup> As the great abundance of alkali metal cations in the organism, the i-motif-like molecules (cytosine dimers stabilized by alkali metal cations) might be detected. Following the research in Chapter 4, the structures of protonated and alkali metal cationized dimeric nucleobases were discussed in Chapters 5 and 6; the contents of these two chapters focused on the structures of guanine-containing nucleobase dimers.

Guanine:cytosine (G:C) base pairs exist in both DNA and RNA polymers. The first part of Chapter 5 is the research on protonated G:C base pair. The lowest energy structure of protonated 9-ethylguanine/1-methylcytosine dimer,  $(9\text{eG}:1\text{mC})\text{H}^+$ , has been probed to be a Hoogsteen-type G:C base pair, showing the proton is covalently bound to N3 of 1mC

and hydrogen bonded to N7 of 9eG. This structure is consistent with the previous discovery by biologists that the presence of proton is most likely to induce and form the protonated Hoogsteen G:C base pair.<sup>6-8</sup> However, the proton affinity (PA) of N3 of 1mC was calculated to be much lower than that of N7 of 9-ethylguanine (9eG) by computational methods; it was not expected to have the proton closer to 1mC because of its lower PA. In this work, the PA deficit between monomers was presumed to be offset by a stronger ion-dipole interaction and ion-induced dipole interaction in the global minimum, due to 9eG having higher polarizability and dipole moment than 1mC.

Furthermore, the experimental ratio of fragments,  $(1mC)H^+:(9eG)H^+$ , in the IRMPD dissociation mass spectrum was observed to be much higher than that expected by the computed gas basicities of 1mC and 9eG, which was also observed in analogous works previously and termed an *anomaly*.<sup>6,9-11</sup> In this chapter, the *anomaly* has been explained as the result of the high kinetic barrier for proton transfer when  $(9eG:1mC)H^+$  started to dissociate, which was studied by performing relaxed scanning on the potential energy surface of this system. In order to traverse through the lowest energy barrier, 9eG and protonated 1mC need to shrink their N7-N3 distance to assist proton transfer through the barrier ( $\sim 12 \text{ kJ mol}^{-1}$ ) from the ground state of the global minimum. Otherwise, it is less likely to observe protonated 9eG once the dissociation starts, as the damped thermal energy is not enough to surmount the reaction barrier.

Further work on alkali metal cationized 9eG:1mC shows that the impacts of alkali metal cations are various. As  $Li^+$  has the highest charge density and smallest radius among all five metal cations, it is most likely to disrupt the hydrogen bond between G:C base pair and induce the Hoogsteen structure. As the size of metal increases, the relative Gibbs

energy of alkali metal cationized Watson-Crick (WC) pairing structure is lowered. And WC structure has been found to be the major contributor to  $\text{Cs}^+$  - 9eG:1mC complex, showing  $\text{Cs}^+$  bound to O6N7 of 9eG. Alkali metal cations and proton do not induce any tautomerization of 9eG or 1mC but can disrupt the hydrogen bonds between Watson-Crick G:C base pair and form the Hoogsteen-style structure, which relies on different alkali metal cations' radii and their charge densities.

Moreover, research in Chapters 4 and 5 both generate the conclusion that, as alkali metal cations increase in size, ion-dipole interaction between metal cation and nucleobases becomes weaker supported by the observations of bond distance lengthening.

Lately, because guanine-involved mismatches are the most stable mismatched base pairs *in vivo*,<sup>12</sup> protonated 9eG:9mA, 9eG:9eG, and 9eG:1mT, were the focus of Chapter 6, where 9mA is 9-methyladenine, and 1mT is 1-methylthymine. It has been concluded that two existing isomers of  $(9\text{eG}:1\text{mA})\text{H}^+$  both present the structures as the proton close to 9mA rather than 9eG, which can be contributed by the strong ion-dipole and ion-induced dipole interaction, analogous to  $(9\text{eG}:1\text{mC})\text{H}^+$ . The protonated 9eG:1mT shows the proton is covalently bound to 9eG and hydrogen bound to 1mT, which is consistent with their thermodynamic properties. Besides, there is a first discovered nucleobase pair stabilized by a single *classic* hydrogen bond,  $(9\text{eG}:9\text{eG})\text{H}^+$ , which is neither in Watson-Crick or Hoogsteen style.

Previously investigations on 18+1 protonated base pairs, eighteen distinct protonated base pairs from RNA crystal structure data set and one presumed protonated base pair, has been performed by computational methods, including structural optimization, thermostability calculations and binding energy decomposition.<sup>13</sup> It has been concluded

that there is the preferred interaction between protonated A/G/C with neutral C/G. In these cases, some structures show the proton covalently bound to lower PA monomer, which was presumed as a result of strong ion-dipole interaction. In Chapters 5 and 6, the combination of IRMPD spectroscopic method and computed quantum theory validated the presence of these structures. Also, computed polarizabilities provided a hint for considering the induction effects between neutral and protonated monomers. NMR and X-ray crystallography methods were mentioned to be not suitable to study the structures of protonated base pairs due to the exchangeable proton in solution and proton's low scattering cross-section, respectively.<sup>13</sup> In order to determine the existence and position of an extra proton, IRMPD spectroscopy can be used as an experimental method of providing real evidence for the structures of protonated base pairs. Moreover, two protonated 9eG:9mA isomers are consistent with the previous identification of protonated guanine:adenine (G:A) structures in RNA crystal structures; we would expect to see more structures as we discussed in these two chapters to be disclosed in the protein data bank.

Based on the insights provided in this dissertation, there are multiple directions to take for future work. Firstly, following the work in Chapter 5, it is expecting to dig into the molecular dynamics of protonated G:C base pair, which will assist a more comprehensive understanding of the *anomaly*—the unexpected higher intensity of (1mC)H<sup>+</sup> in (9eG:1mC)H<sup>+</sup> dissociation mass spectrum. Furthermore, this study opens up an area in determining the structures of protonated and metal cationized base pairs, homo and hetero, by both spectroscopic identification and thermodynamic comparison, which may shed light on the mysterious structures of base pairs in DNA and RNA strands *in vivo*. In terms of the insights offered by this work and previous disclosures, the structures of protonated

cytosine-containing mismatches are attractive to explore further. It will be exciting to compare the structural information of these biomolecules to that discovered in DNA and RNA, and even boost the findings of protonated and metal cationized base pairs and mismatches. As FTICR-MS is capable of trapping and detecting the high molar mass biomolecules, further work will be interesting to use FTICR-MS to provide more information on biomolecules on a large scale, including nucleobases, nucleotides, and even nucleic acids. It will help with further discovery on the structures of DNA molecules, determine the binding positions of ions to nucleotides and even nucleic acids, understand the effects of ions on geometries and thermostabilities of these biomolecules, and may boost the finding of ionic drugs for targeted genes of cancer therapy.

## References

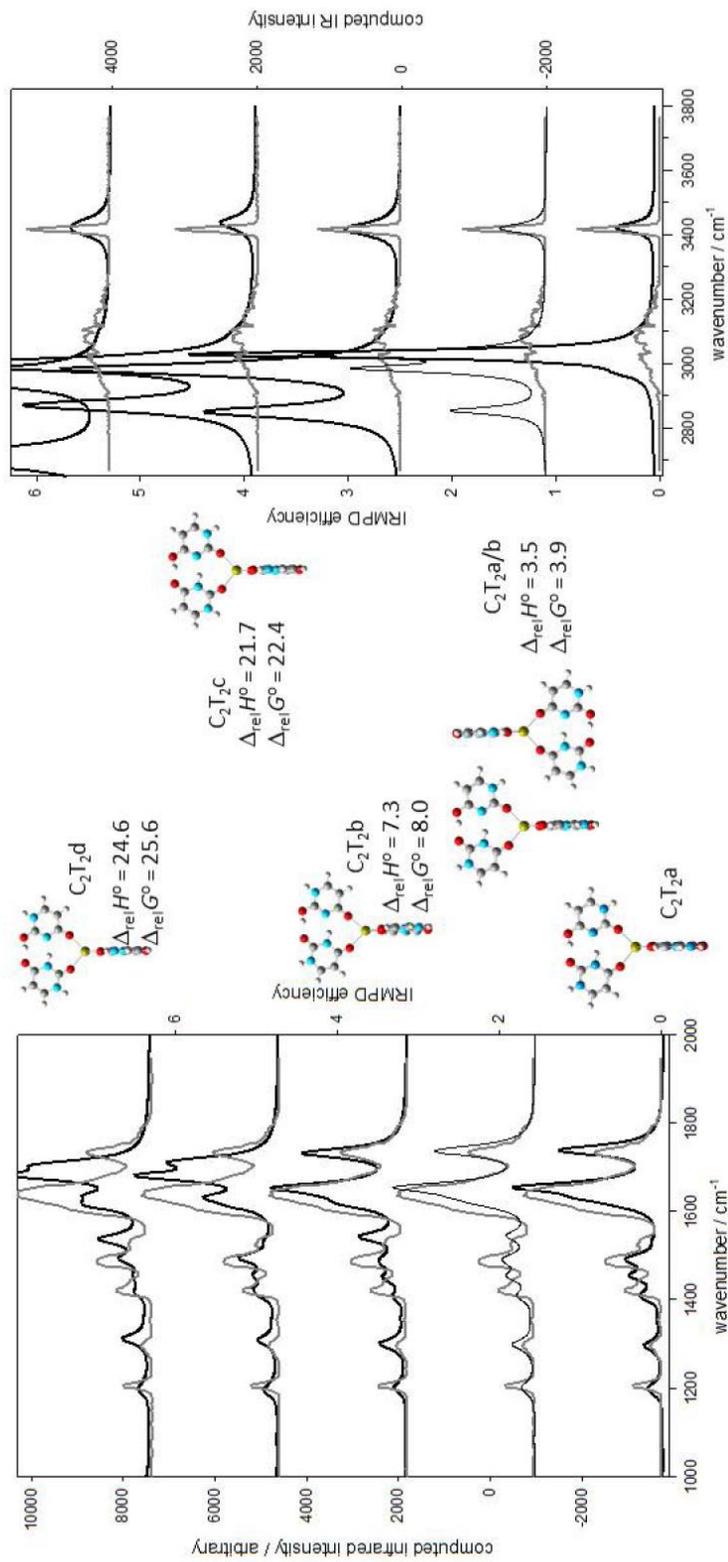
- 1 E. a. L. Gillis, M. Demireva, K. Nanda, G. Beran, E. R. Williams and T. D. Fridgen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 3304.
- 2 C. Trujillo, A. M. Lamsabhi, O. Mó, M. Yáñez and J. Y. Salpin, *Int. J. Mass Spectrom.*, 2011, **306**, 27–36.
- 3 B. Yang, R. R. Wu, N. C. Polfer, G. Berden, J. Oomens and M. T. Rodgers, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 1523–1533.
- 4 B. Yang and M. T. Rodgers, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16110–16120.
- 5 M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger and D. Christ, *Nat. Chem.*, 2018, **10**, 631–637.
- 6 A. F. Cruz-Ortiz, M. Rossa, F. Berthias, M. Berdakin, P. Maitre and G. A. Pino, *J. Phys. Chem. Lett.*, 2017, **8**, 5501–5506.
- 7 M. D. Frank-Kamenetskii and S. M. Mirkin, *Annu. Rev. Biochem.*, 1995, **64**, 65–95.
- 8 S. Y. Han, S. H. Lee, J. Chung and H. Bin Oh, *J. Chem. Phys.*, 2007, **127**, 245102.
- 9 J. J. Park and S. Y. Han, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 846–854.
- 10 Y. Seong, S. Y. Han, S. C. Jo and H. Bin Oh, *Mass Spectrom. Lett.*, 2011, **2**, 73–75.
- 11 J. J. Park, C. S. Lee and S. Y. Han, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 2368–2379.
- 12 F. H. Martin and J. Farecd Aboul-ela, David Koh and Ignacio Tinoco, *Nucleic Acids Res.*, 1985, **13**, 4811–4824.

- 13 M. Chawla, P. Sharma, S. Halder, D. Bhattacharyya and A. Mitra, *J. Phys. Chem. B*, 2011, **115**, 1469–1484.

# **Appendices A**

## **(Chapter 3)**

Figure S1



Comparison of the experimental IRMPD spectrum (grey trace) for  $U_4Ca^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $C_2T_2a$  are also shown (also computed using B3LYP/6-31+G(d,p)).

Figure S2

Comparison of the experimental IRMPD spectrum in the C-H/N-H/O-H stretch region (grey trace) for  $U_6Ca^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $C_4T_2a$  are also shown (also computed using B3LYP/6-31+G(d,p)). The experimental spectrum is complementary to that in the fingerprint region seen in Figure 3.

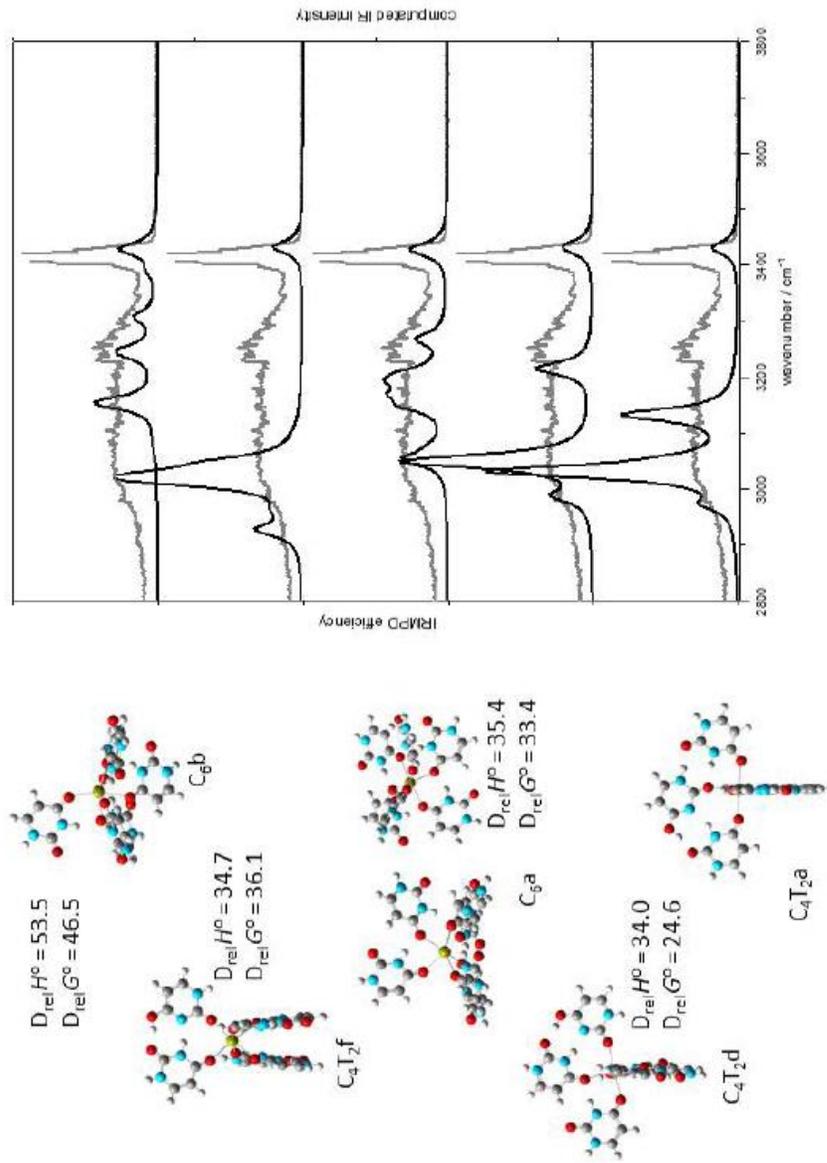
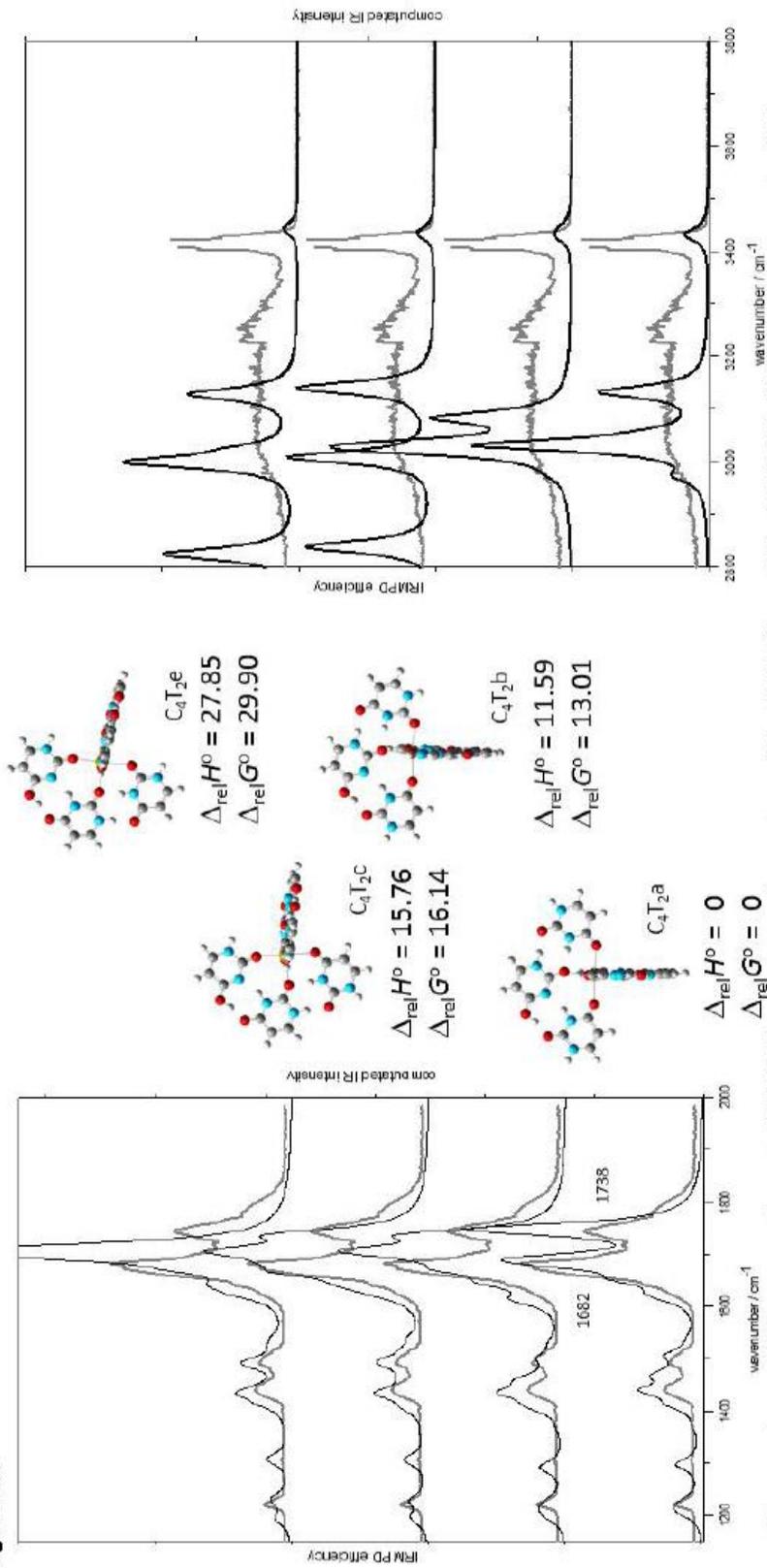
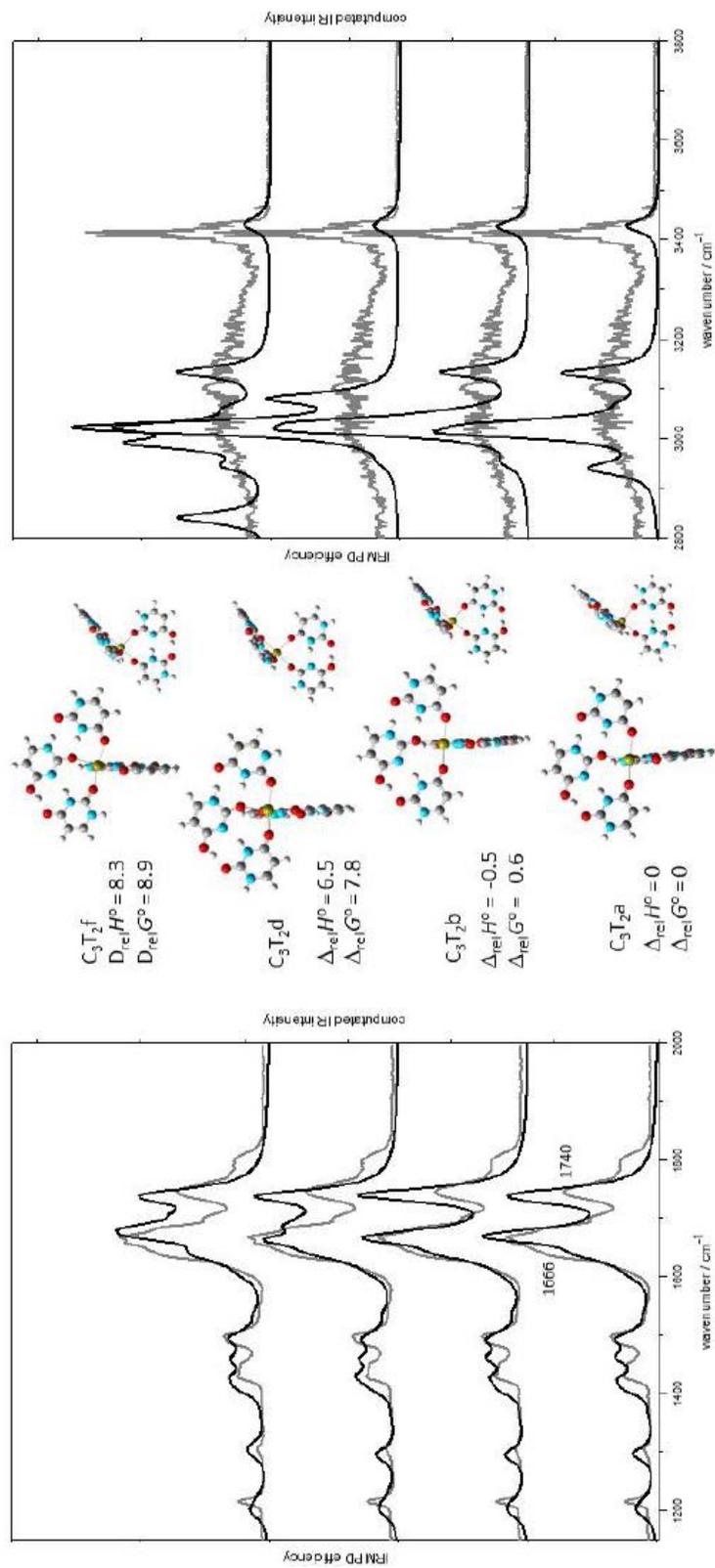


Figure S3



Comparison of the experimental IRMPD spectrum (grey trace) for  $U_6Ca^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for four different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $C_4I_2a$  are also shown (also computed using B3LYP/6-31+G(d,p)).

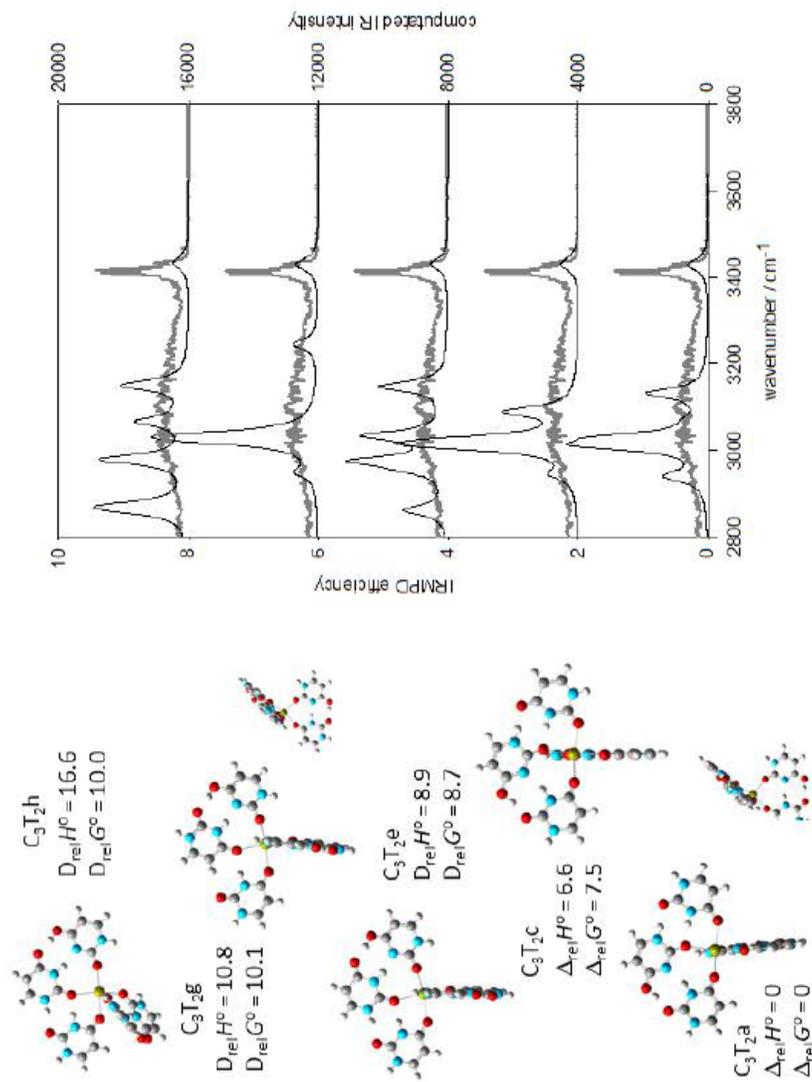
Figure S4

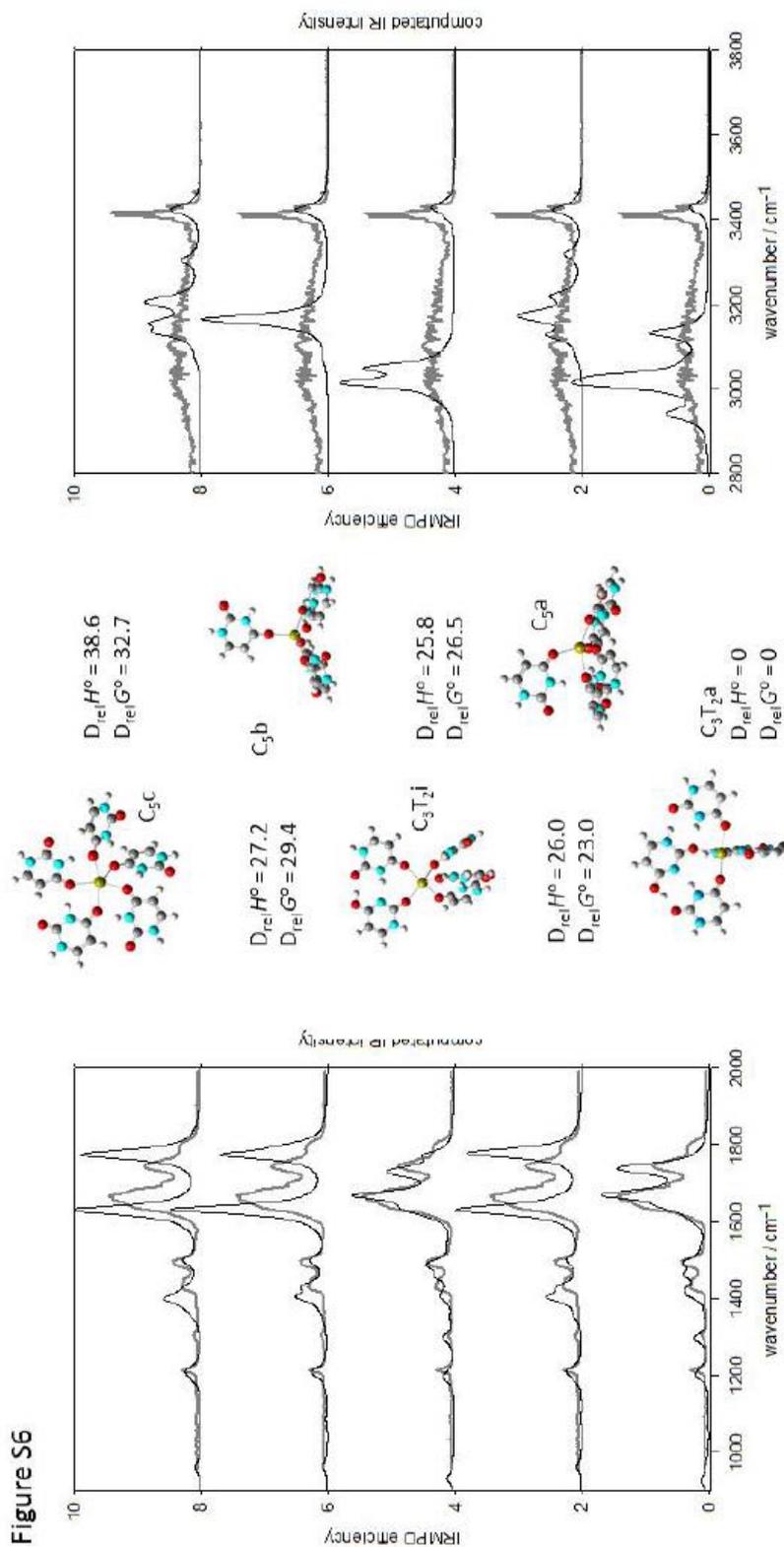


Comparison of the experimental IRMPD spectrum (grey trace) for  $U_5Ca^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for four different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $C_3T_2a$  are also shown (also computed using B3LYP/6-31+G(d,p)).

Figure S5

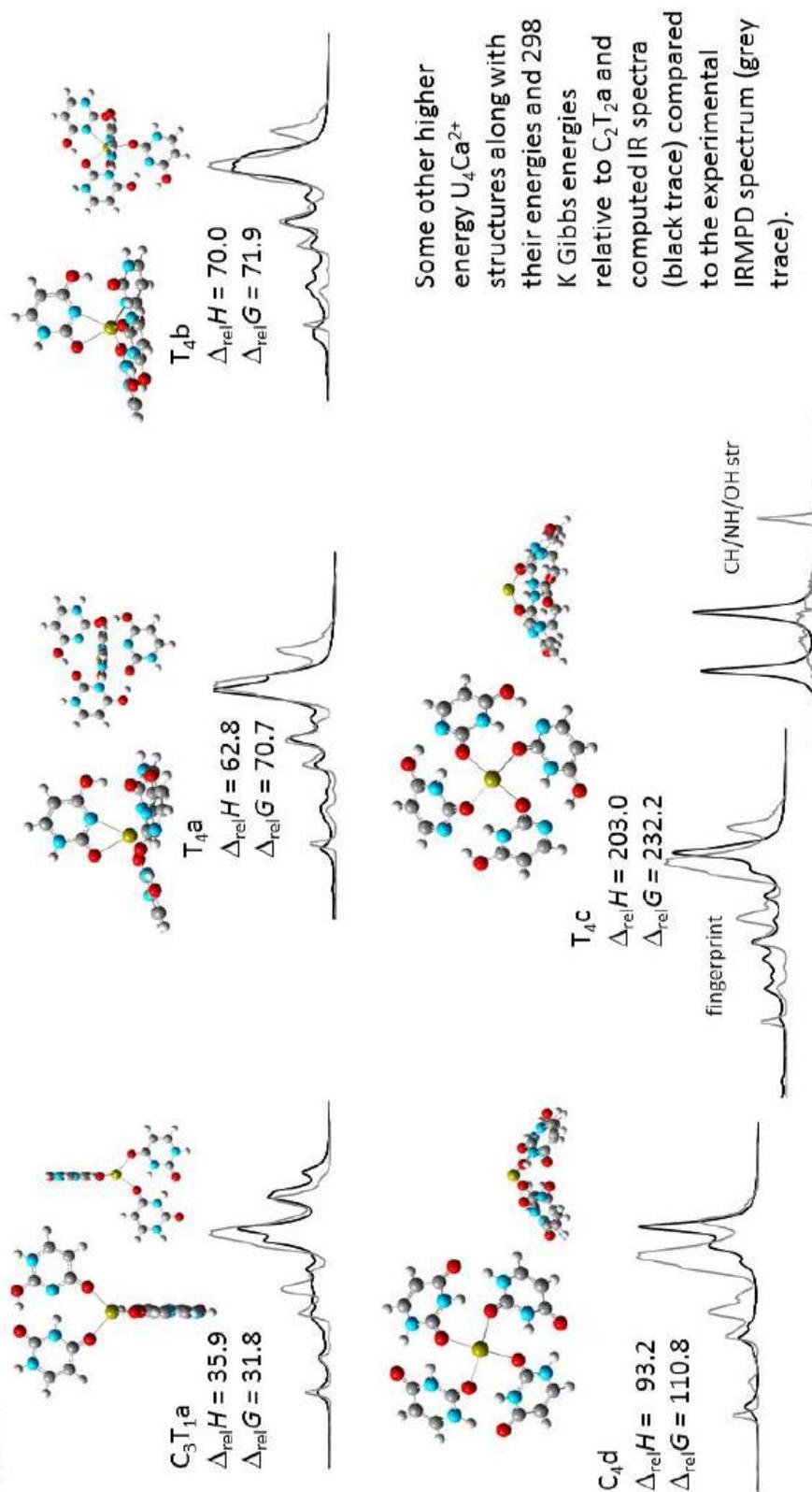
Comparison of the experimental IRMPD spectrum in the C-H/N-H/O-H stretch region (grey trace) for  $U_5Ca^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $C_3T_2a$  are also shown (also computed using B3LYP/6-31+G(d,p)). The experimental spectrum is complementary to that in the fingerprint region seen in Figure 4.

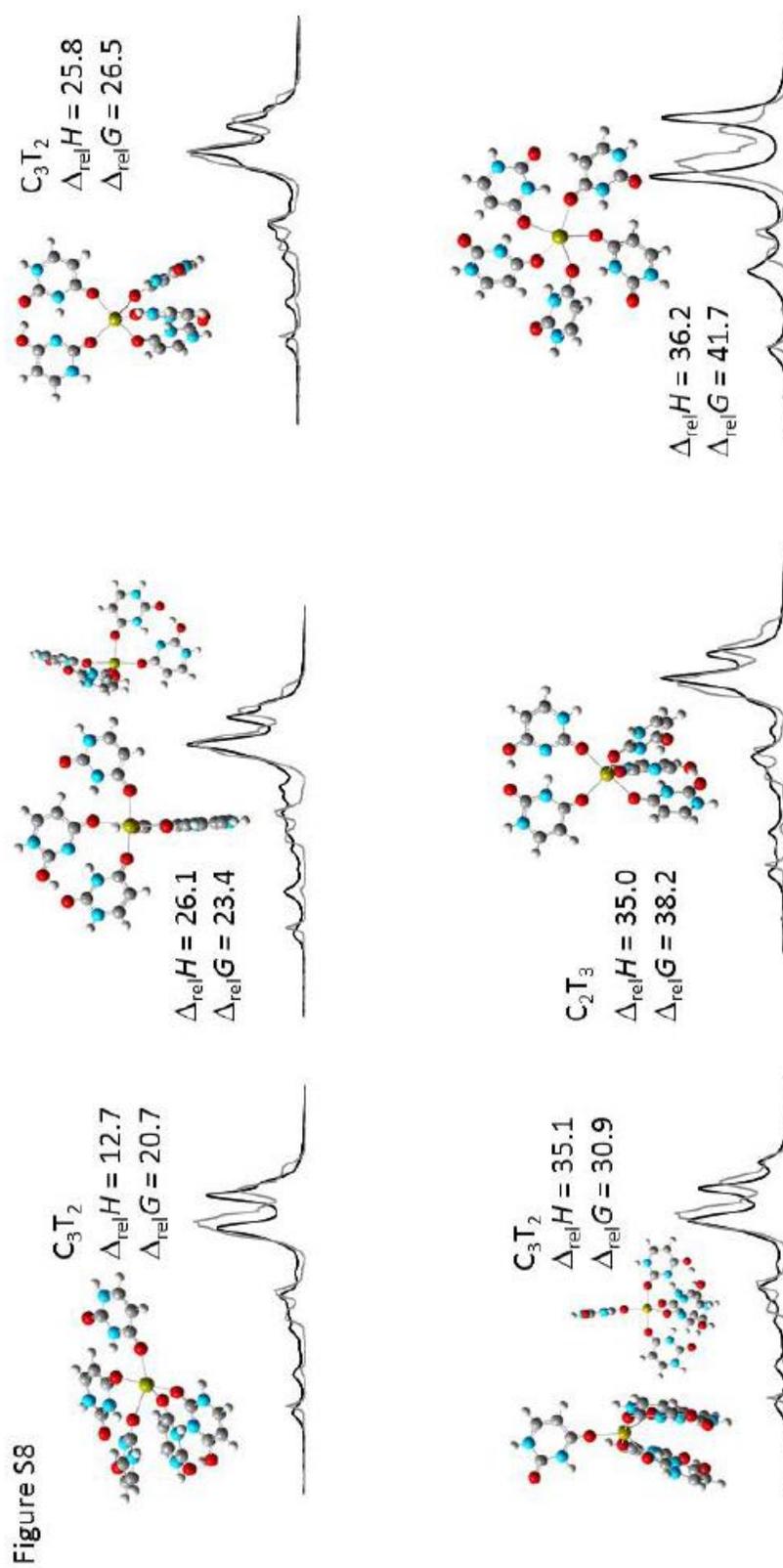




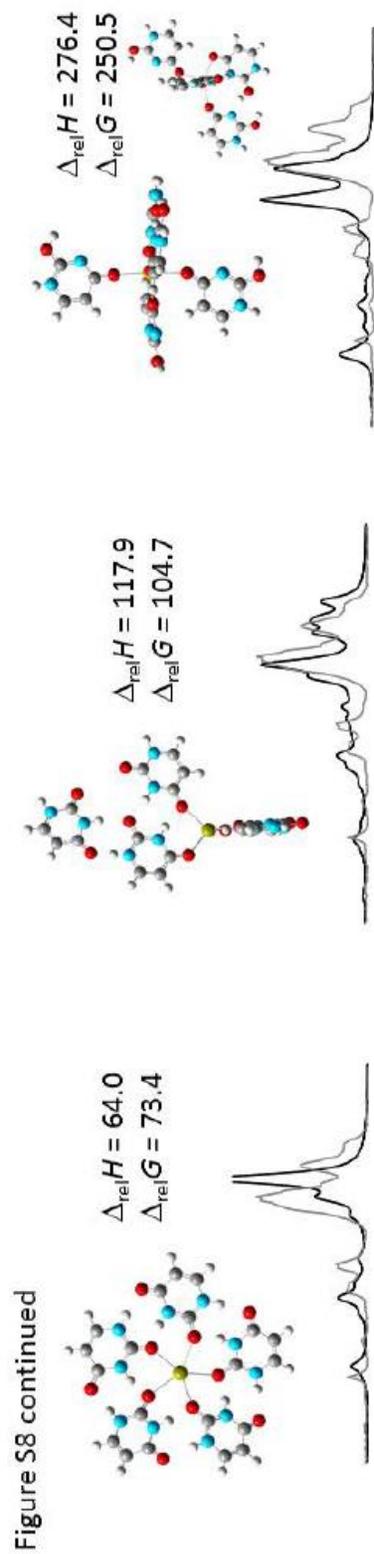
Comparison of the experimental IRMPD spectrum (grey trace) for  $U_5Ca^{2+}$  and the B3LYP/6-31+G(d,p) computed IR spectra (black traces) for five different isomers. The 298 K enthalpies and Gibbs energies relative to structure  $C_3T_2a$  are also shown (also computed using B3LYP/6-31+G(d,p)).

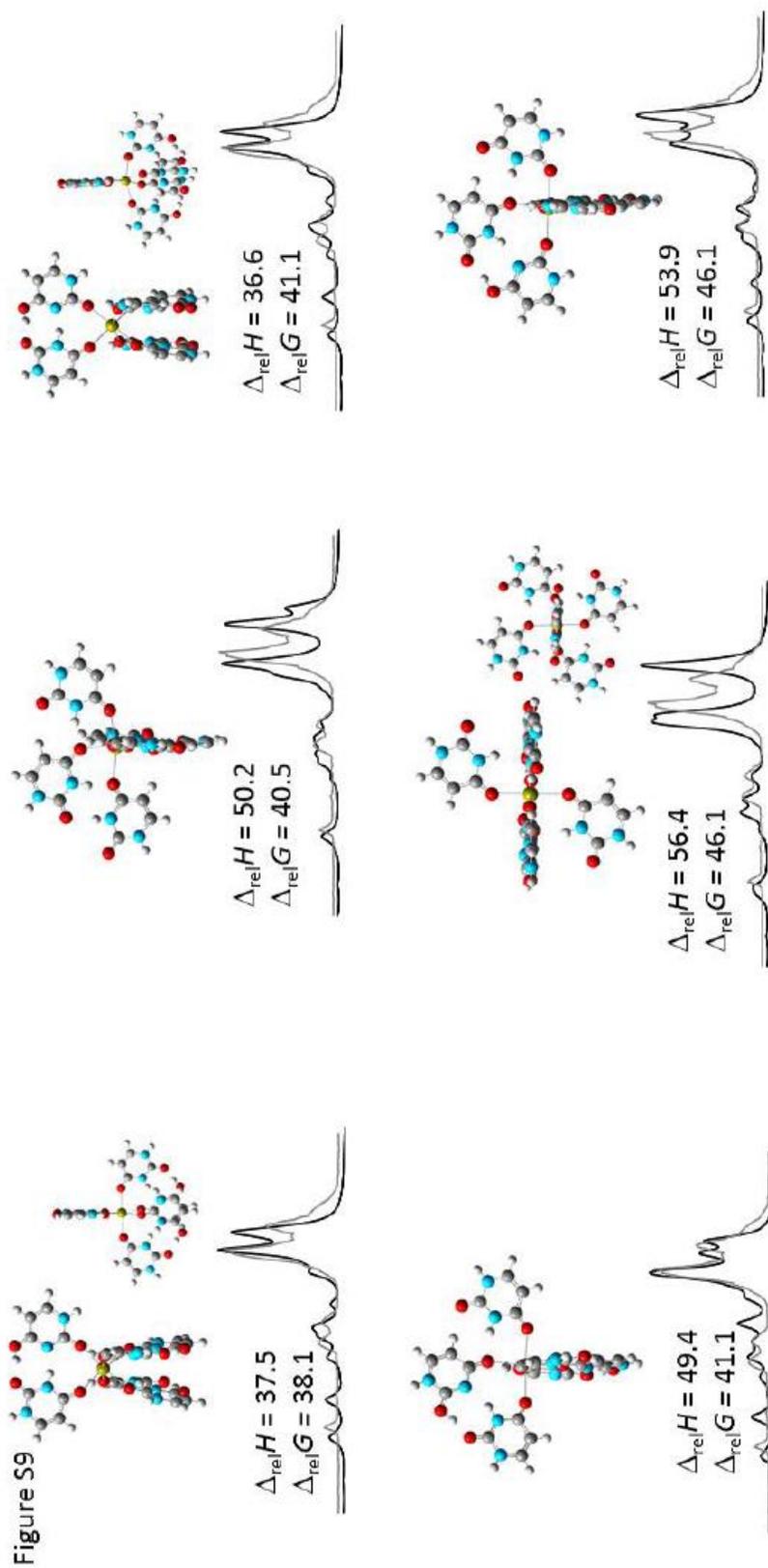
Figure S7





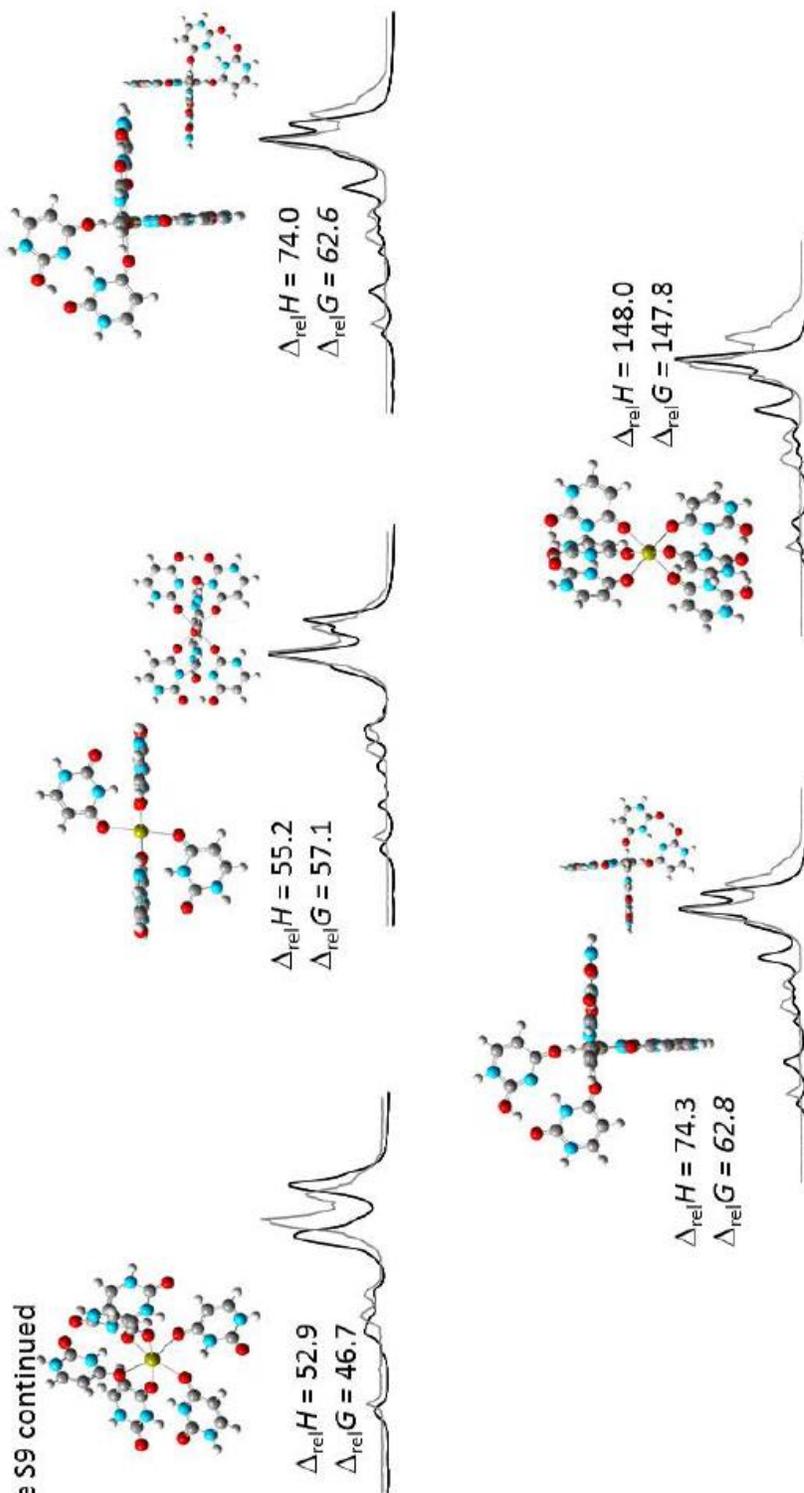
Some other higher energy  $U_5Ca^{2+}$  structures along with their energies and 298 K Gibbs energies relative to  $C_3T_2a$  and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace).





Some other higher energy  $U_6Ca^{2+}$  structures along with their energies and 298 K Gibbs energies relative to  $C_4T_2a$  and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace).

Figure S9 continued



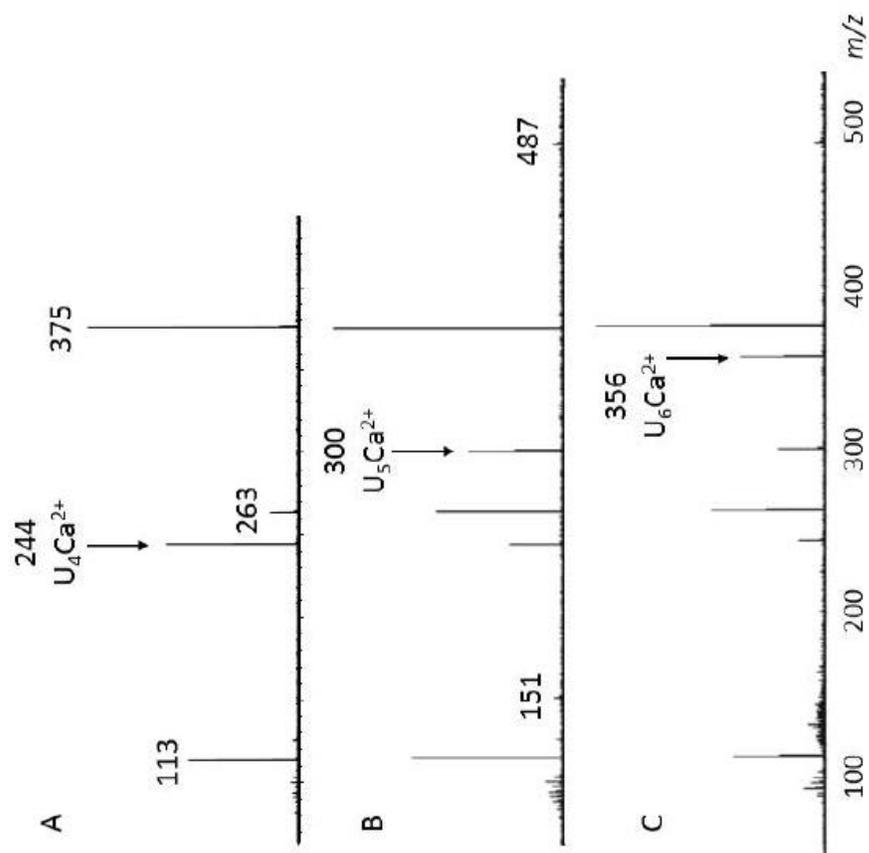


Figure S10. Mass spectra following  $3420\text{ cm}^{-1}$  irradiation of isolated A)  $U_4Ca^{2+}$ , B)  $U_5Ca^{2+}$ , and C)  $U_6Ca^{2+}$ .

Table S1: Comparison of basis set on the 298 K energetics (top relative enthalpies, bottom relative Gibbs energies) in  $\text{kJ mol}^{-1}$  of some  $\text{U}_5\text{Ca}^{2+}$  structures.

Structure	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-311+G(3df,3pd)
C <sub>3</sub> T <sub>2</sub> a	0.0	0.0
	0.0	0.0
C <sub>3</sub> T <sub>2</sub> b	-0.5	-0.4
	+0.6	+0.6
C <sub>3</sub> T <sub>2</sub> d	6.5	7.0
	7.8	8.3
C <sub>3</sub> T <sub>2</sub> c	6.6	7.1
	7.5	8.0
C <sub>3</sub> T <sub>2</sub> f	8.3	8.6
	8.9	9.3
C <sub>3</sub> T <sub>2</sub> e	8.9	8.8
	8.7	8.6

Table S2: Comparison of basis set on the 298 K energetics (top relative enthalpies, bottom relative Gibbs energies) in  $\text{kJ mol}^{-1}$  of some  $\text{U}_6\text{Ca}^{2+}$  structures.

Structure	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-311+G(3df,3pd)
T <sub>4</sub> C <sub>2</sub> a	0.0	0.0
	0.0	0.0
T <sub>4</sub> C <sub>2</sub> b	11.6	12.2
	13.1	13.7
T <sub>4</sub> C <sub>2</sub> c	15.8	15.7
	16.1	16.1
T <sub>4</sub> C <sub>2</sub> d	34.0	32.7
	24.6	23.3
T <sub>4</sub> C <sub>2</sub> e	27.9	28.7
	29.9	30.8
T <sub>4</sub> C <sub>2</sub> f	34.7	31.9
	36.1	33.2

## **Appendices B**

**(Chapter 5)**

Table S1. Table of computed proton affinities and gas basicities (kJ mol<sup>-1</sup>) as well as computed (6-31+G(d,p) and literature values of dipole moments and polarizabilities of the neutral base.

Neutral monomer (protonation site)	$\Delta_p H^0$ ( $\Delta_p G^0$ ), (298 K) <sup>a</sup>	dipole moment / D of neutral base	Polarizability / Å <sup>3</sup> of neutral base
C(N3)	949.3 (915.9)	6.86, 6.93, <sup>b</sup> (7.0)	11.0, 11.8, <sup>b</sup> 11.1, <sup>c</sup> (10.3) <sup>d</sup>
C(O2)	955.5 (921.2)		
1mC(N3)	962.9 (930.5)	6.41	12.9
1mC(O2)	967.0 (933.2)		
G(N7)	952.1 (921.1)	6.82, 7.26 <sup>b</sup>	14.7, 15.8, <sup>b</sup> 15.7, <sup>c</sup> (13.6) <sup>d</sup>
1mG (N7)	963.5 (932.7)		
9mG(N7)	970.2 (938.7)		
9eG(N7)	974.3 (943.3)	7.36	18.4

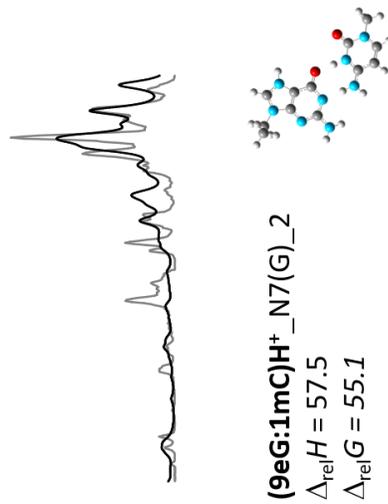
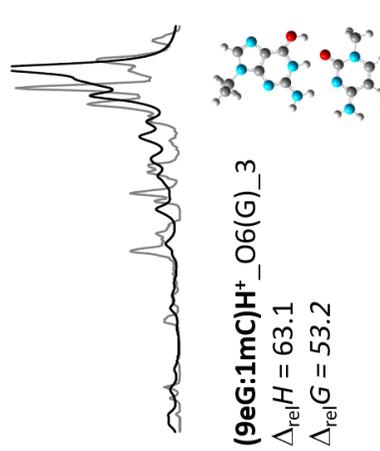
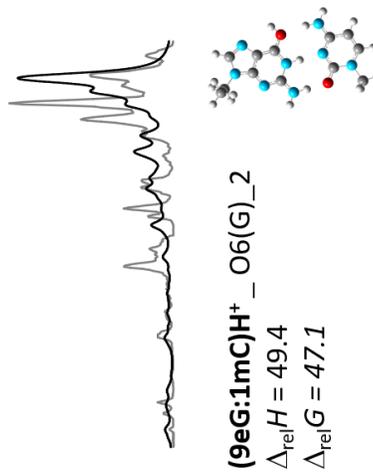
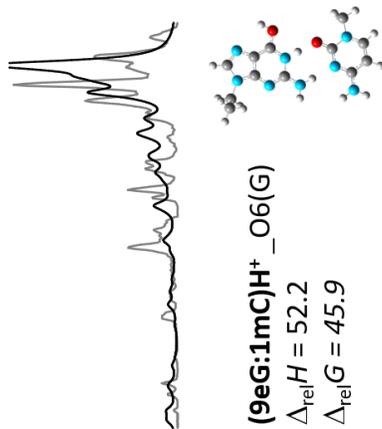
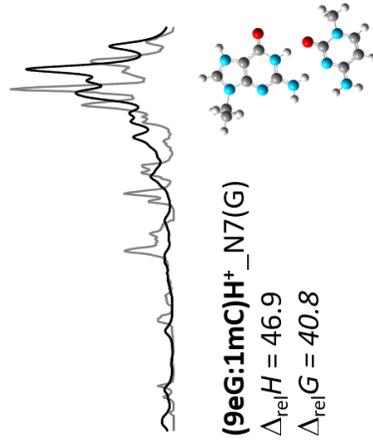
a: CBS-QB3.

b: "Molecular dipole moments and polarizabilities from local density functional calculations: Application to DNA base pairs" P.G. Jasien, G. Fitzgerald, J. Chem. Phys. 93, 1990, 2554-2560.

c: "Calculation of the Molecular Polarizability Tensor" K. J. Miller, J. Am. Chem. Soc. 112, 1990, 8543- 8551.

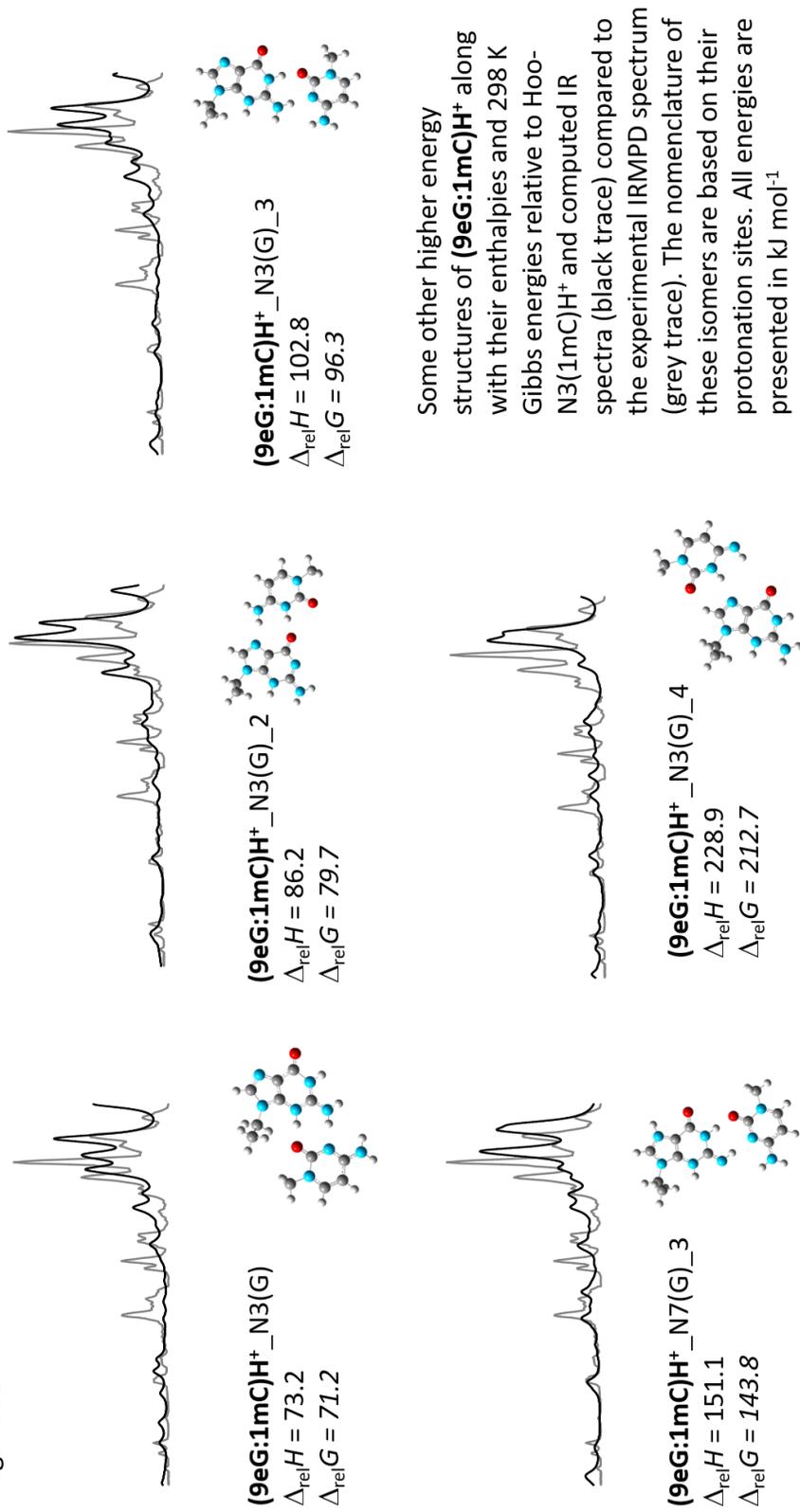
d: Experimental values from C.F.J. Bottcher "Theory of Electric Polarization" Elsevier, Amsterdam (1952), p. 289.

Figure S1



Some other higher energy structures of **(9eG:1mC)H<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to Hoo-N3(1mC)H<sup>+</sup> and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are based on their protonation sites. All energies are presented in kJ mol<sup>-1</sup>

Figure S2



Some other higher energy structures of **(9eG:1mC)H<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to Hoo-N3(1mC)H<sup>+</sup> and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are based on their protonation sites. All energies are presented in kJ mol<sup>-1</sup>

Figure S3

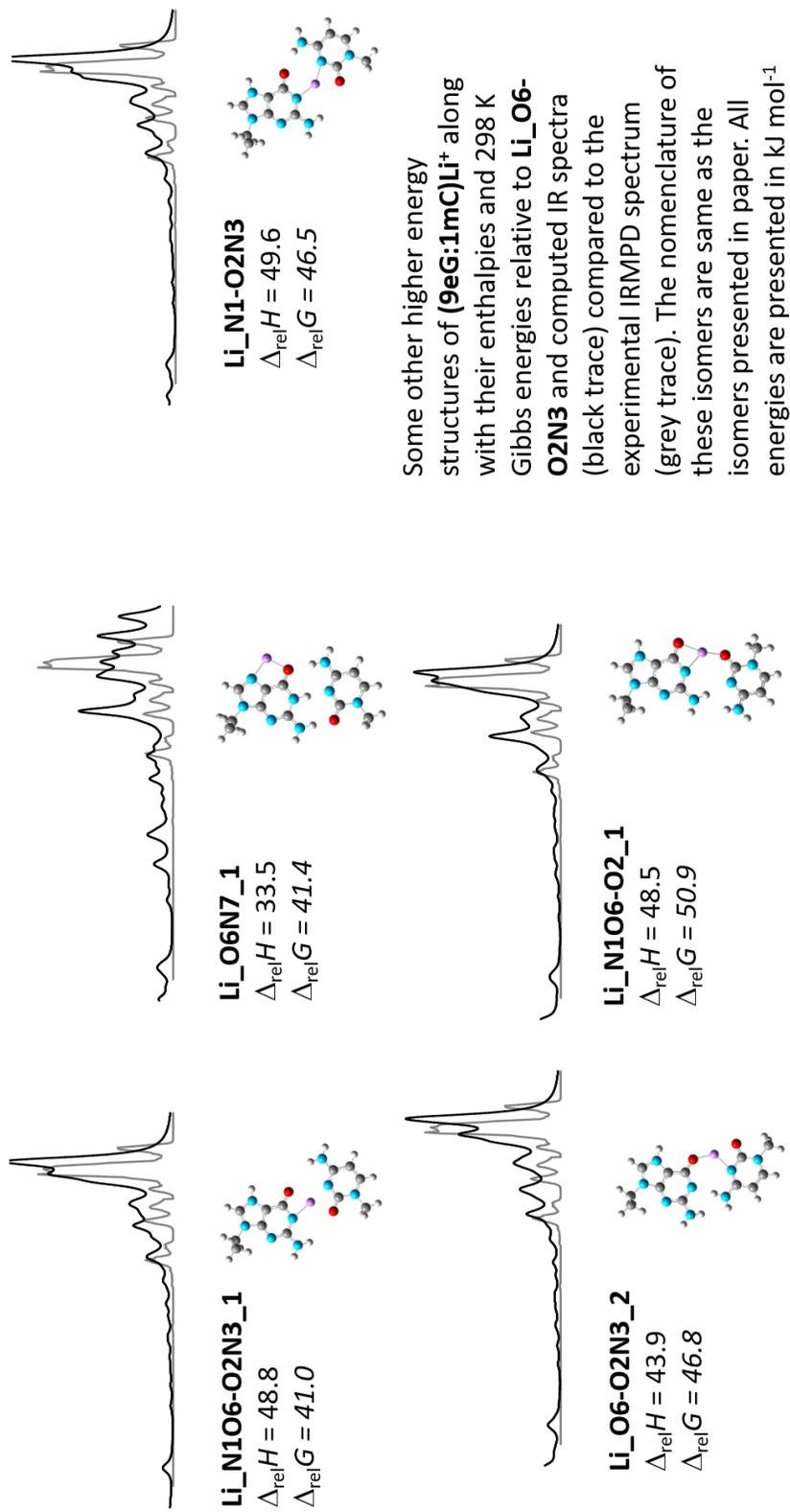
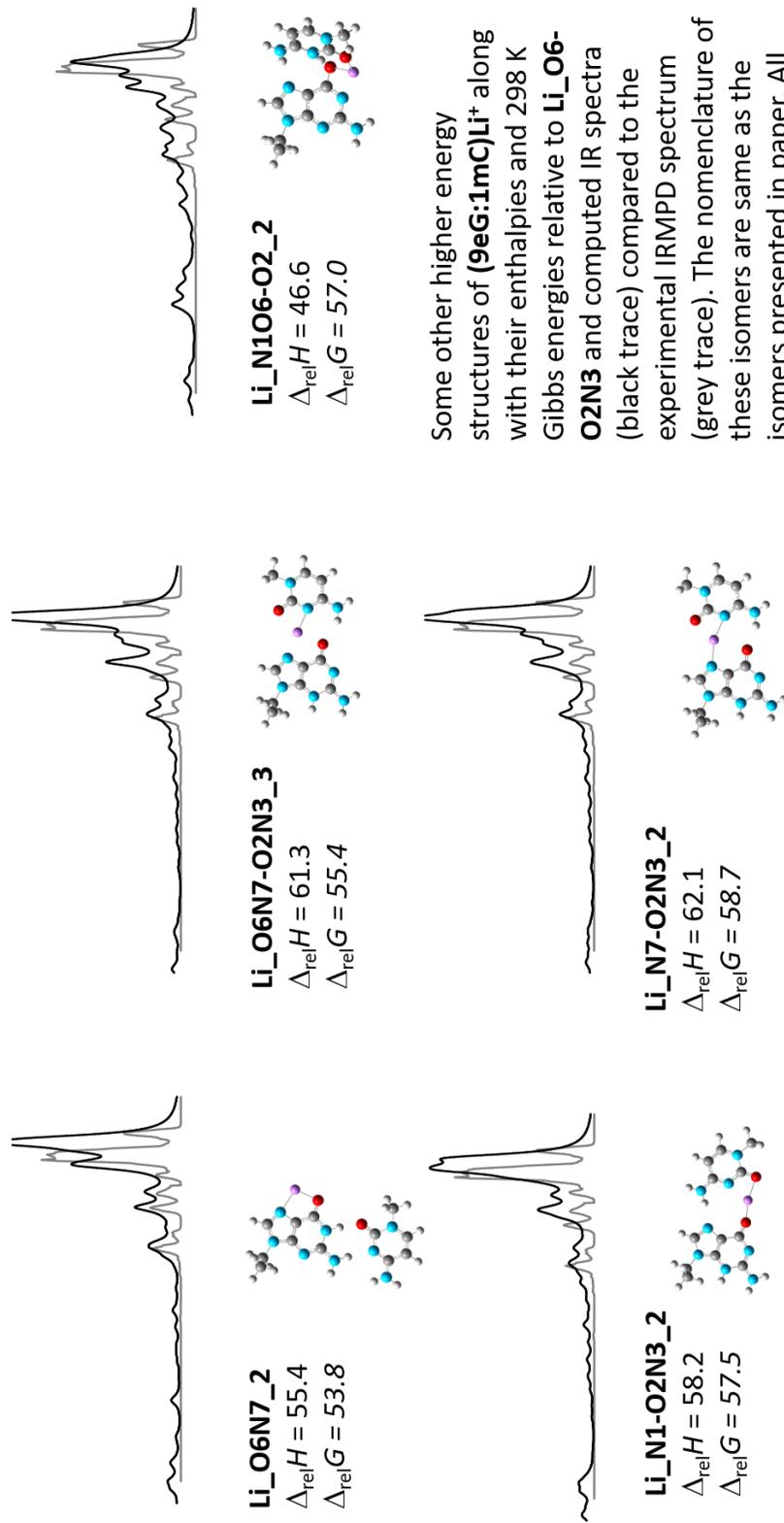


Figure S4



Some other higher energy structures of **(9eG:1mC)Li<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Li\_O6-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in kJ mol<sup>-1</sup>

Figure S5

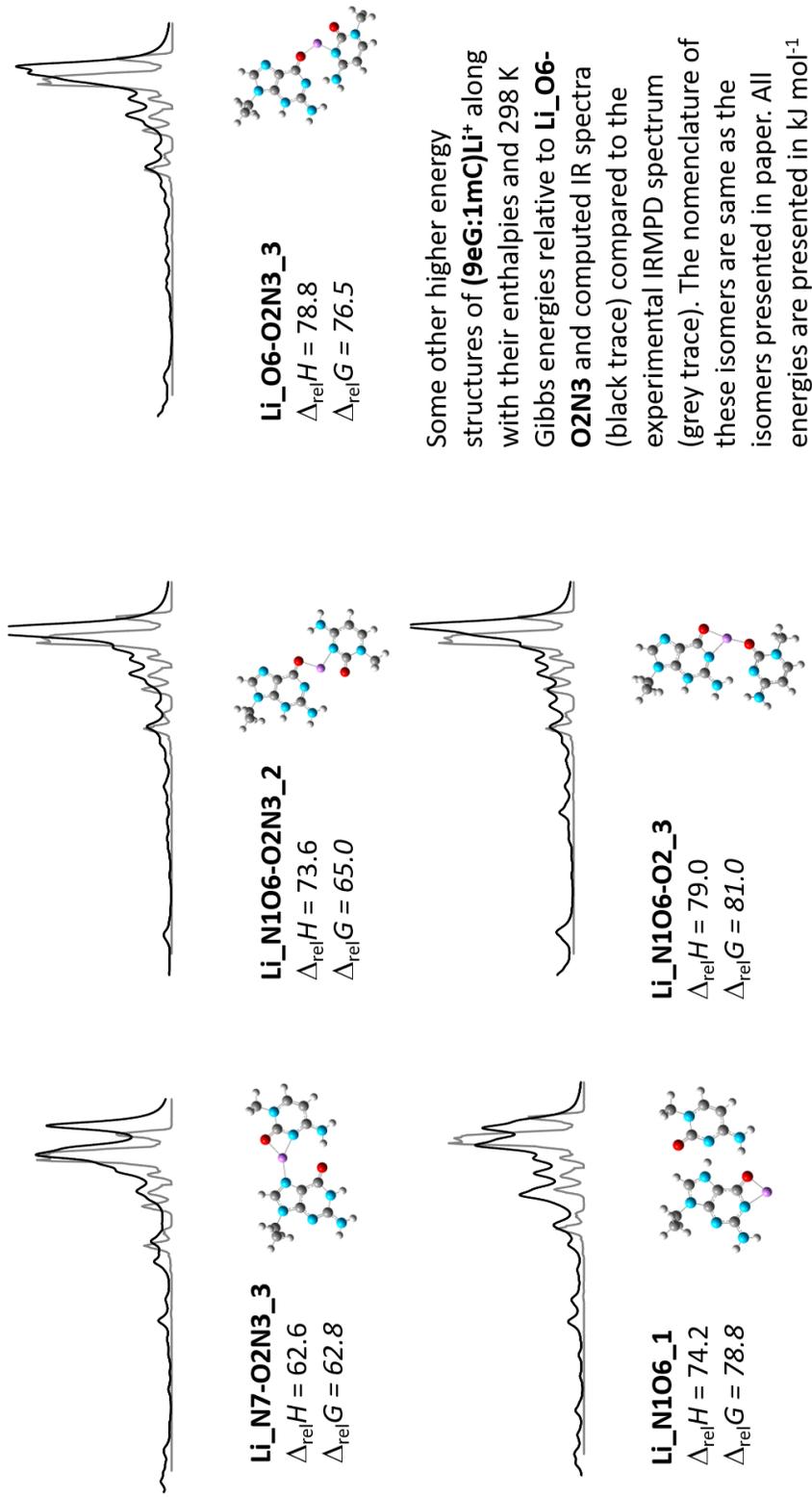


Figure S6

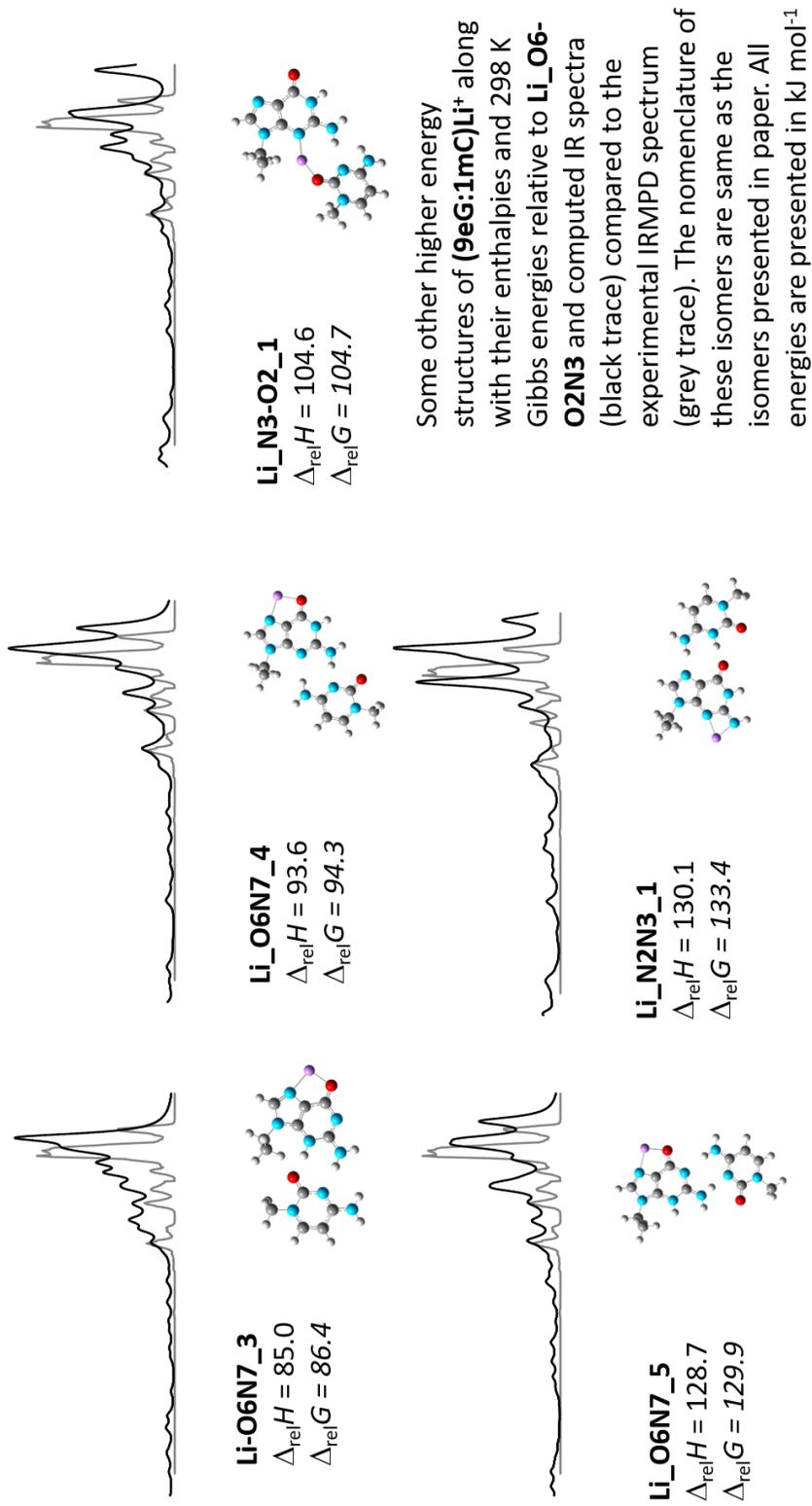
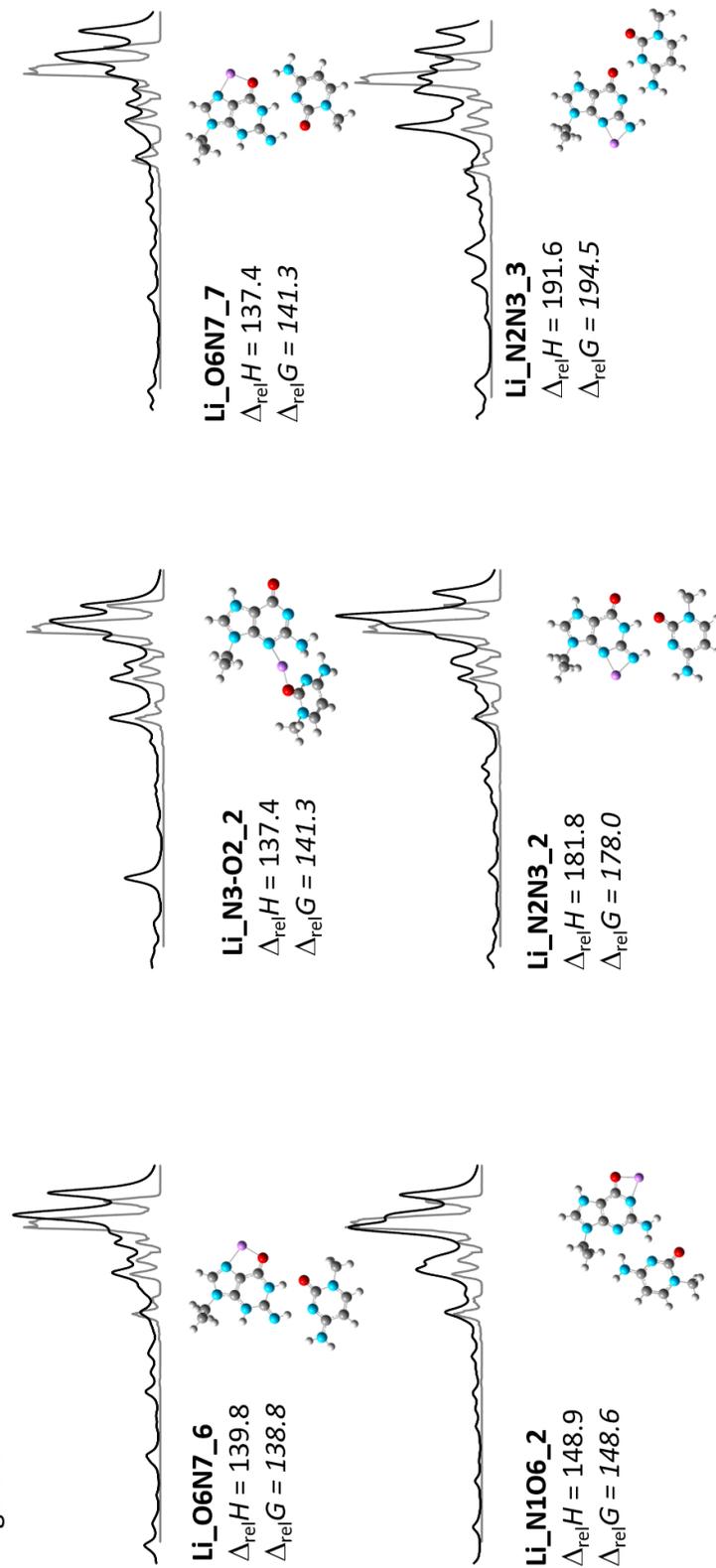
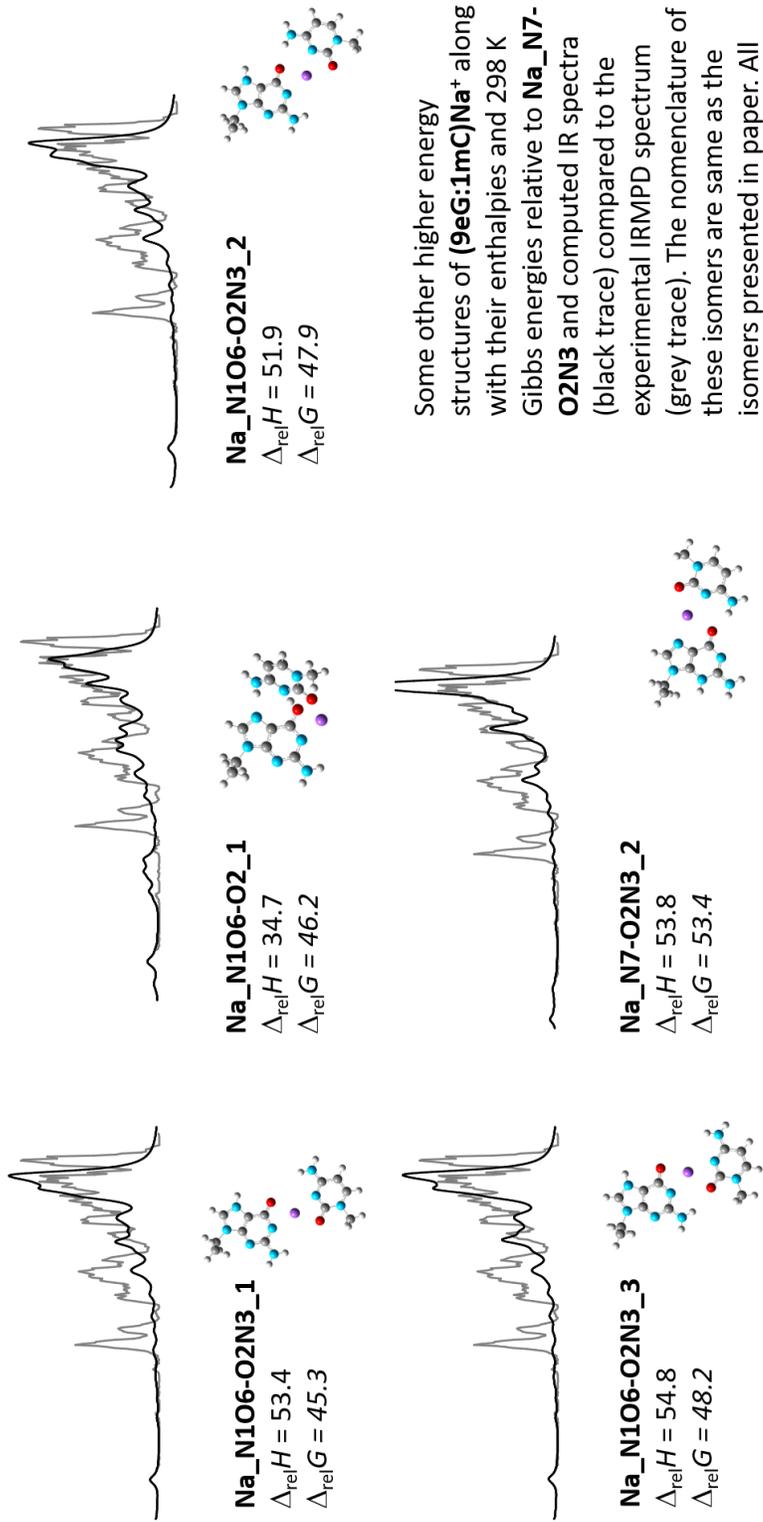


Figure S7



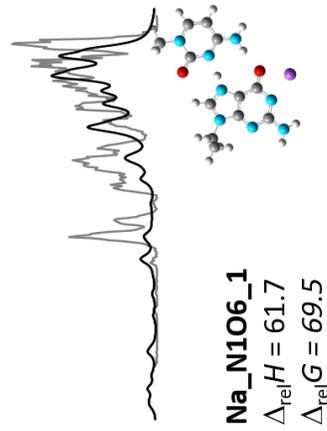
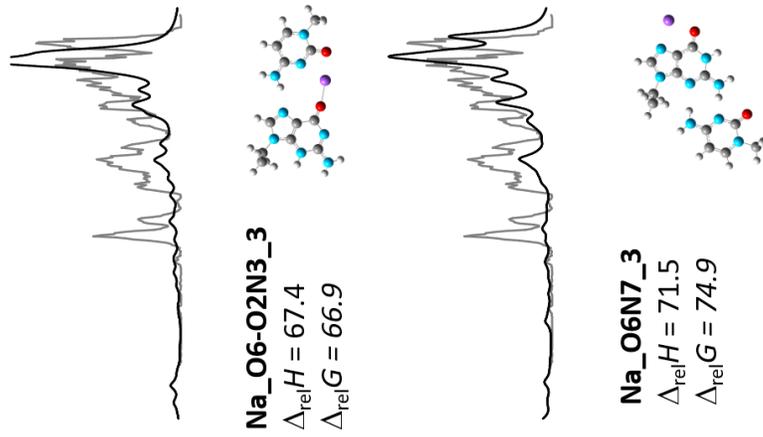
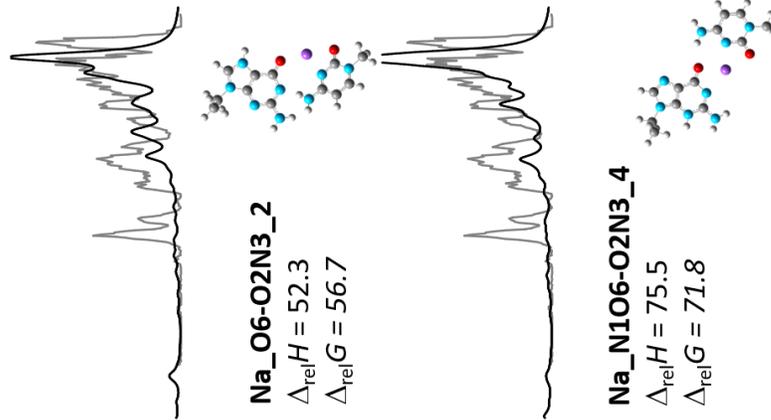
Some other higher energy structures of **(9eG:1mC)Li<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Li\_O6-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S8



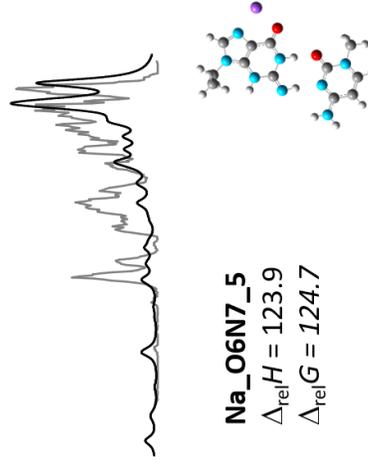
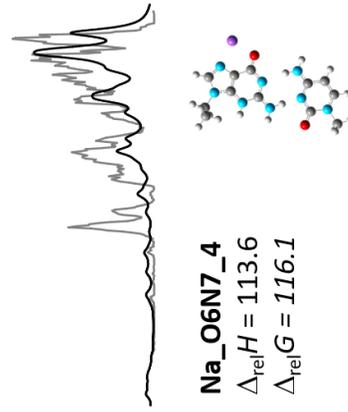
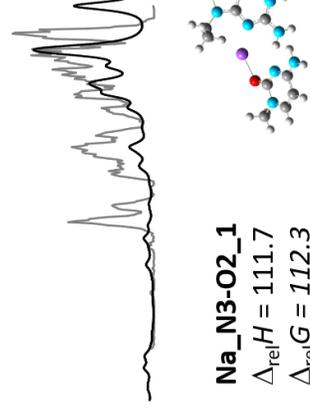
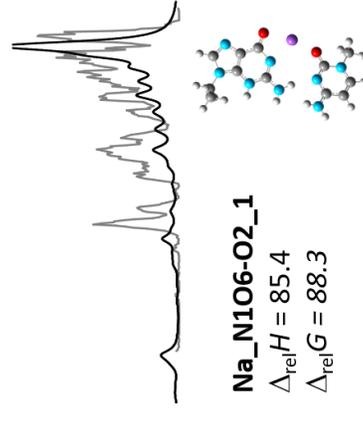
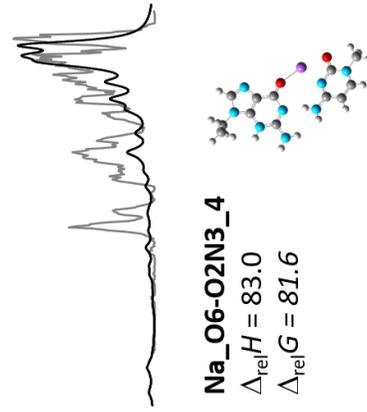
Some other higher energy structures of **(9eG:1mC)Na<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Na\_N7-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in kJ mol<sup>-1</sup>

Figure S9



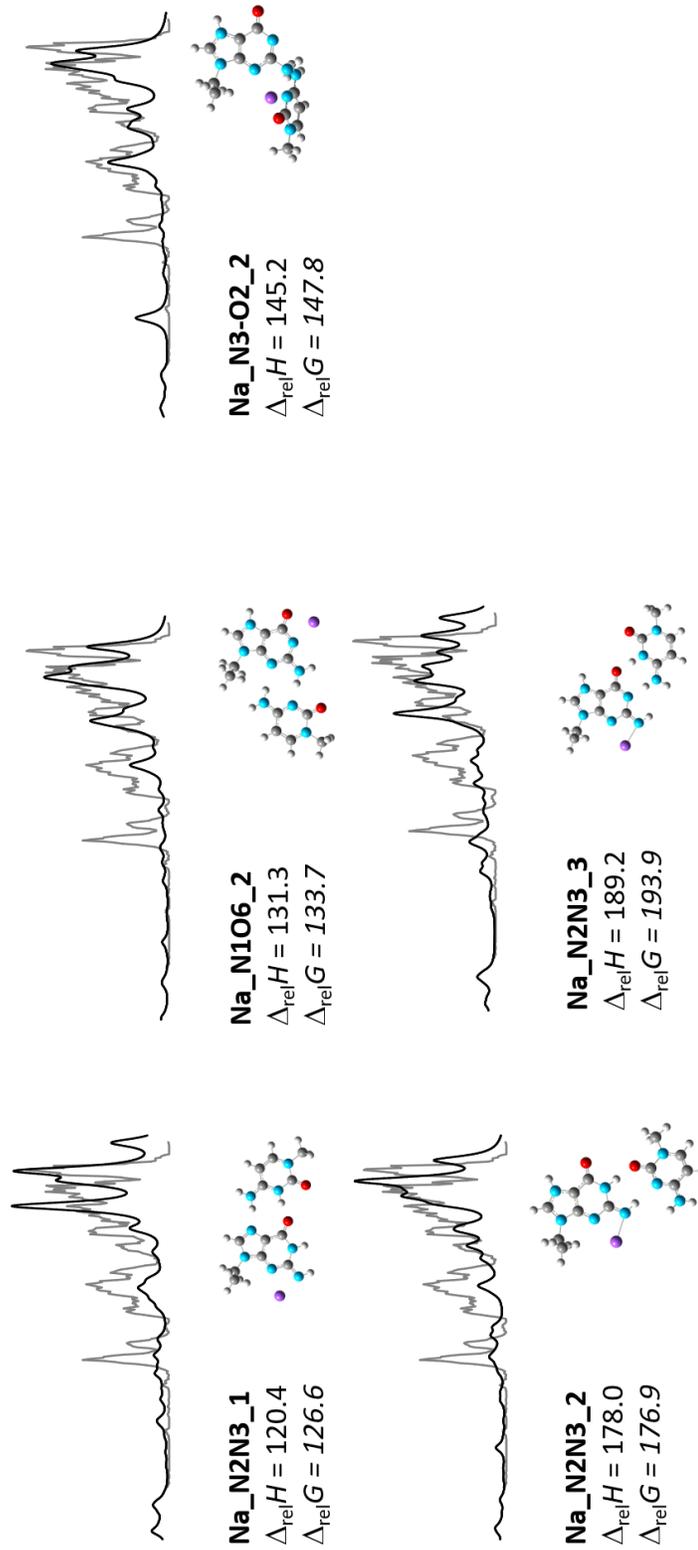
Some other higher energy structures of **(9eG:1mC)Na<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Na\_N7-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S10



Some other higher energy structures of **(9eG:1mC)Na<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Na\_N7-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S11



Some other higher energy structures of **(9eG:1mC)Na<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Na\_N7-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S12

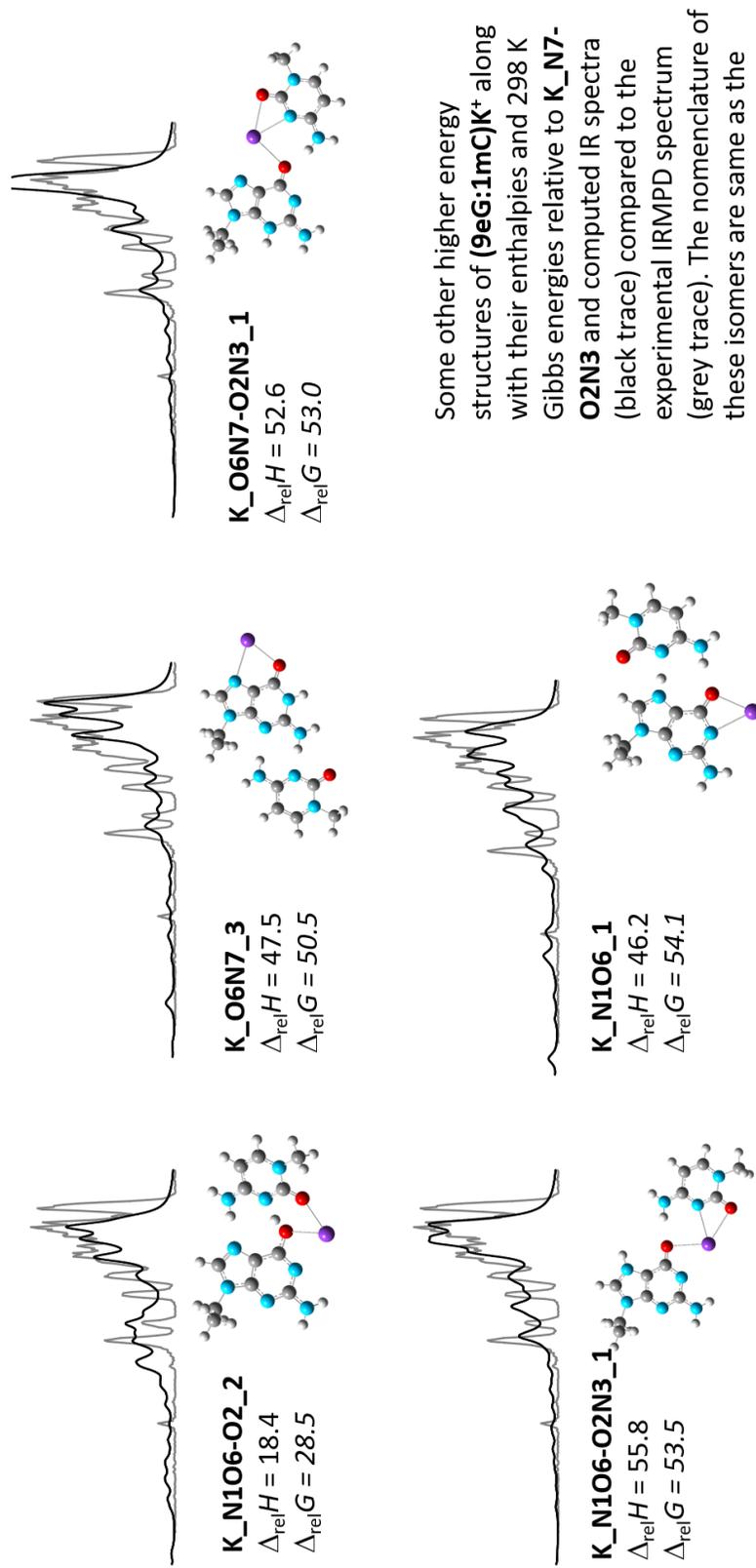


Figure S13

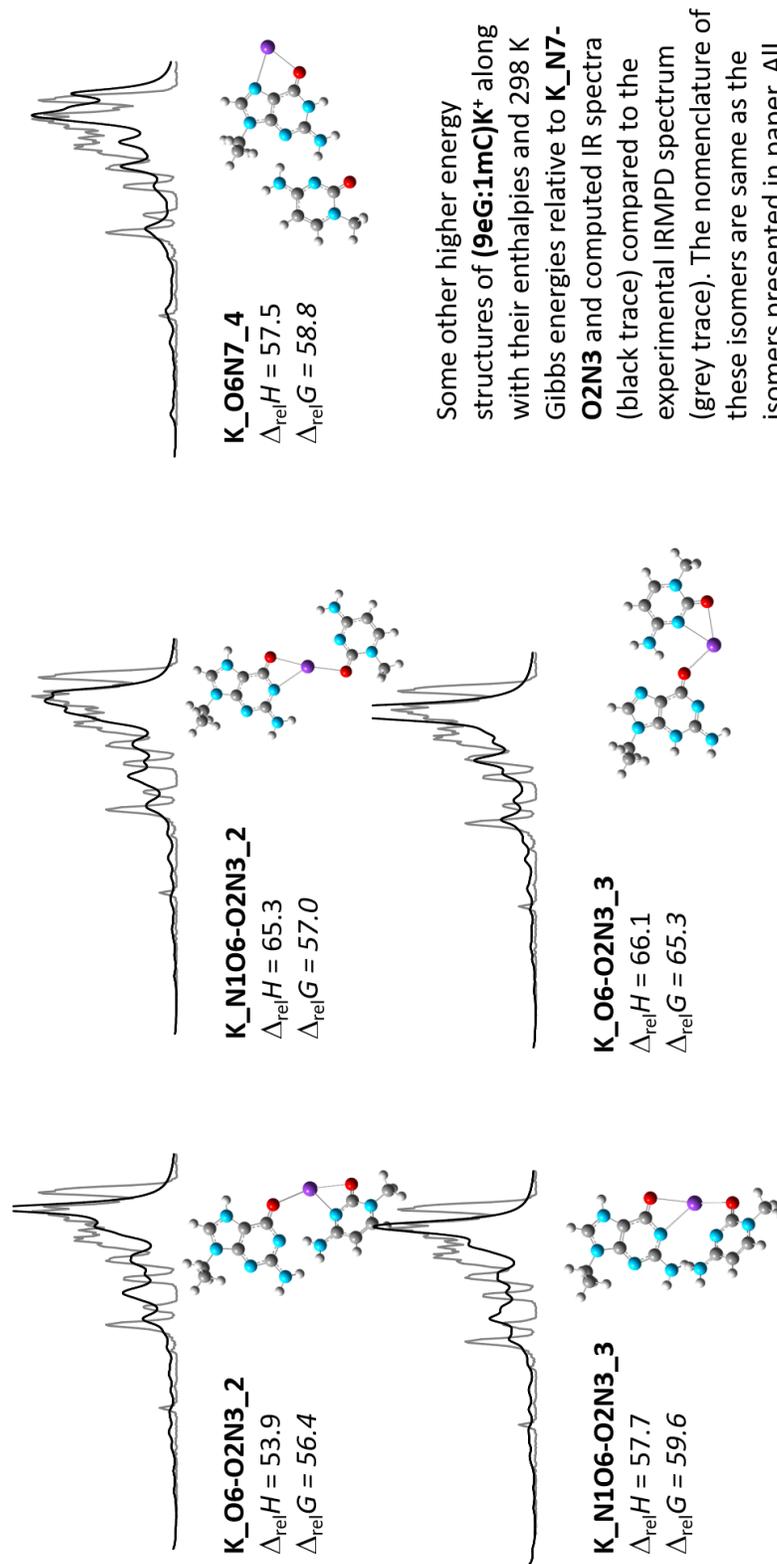


Figure S14

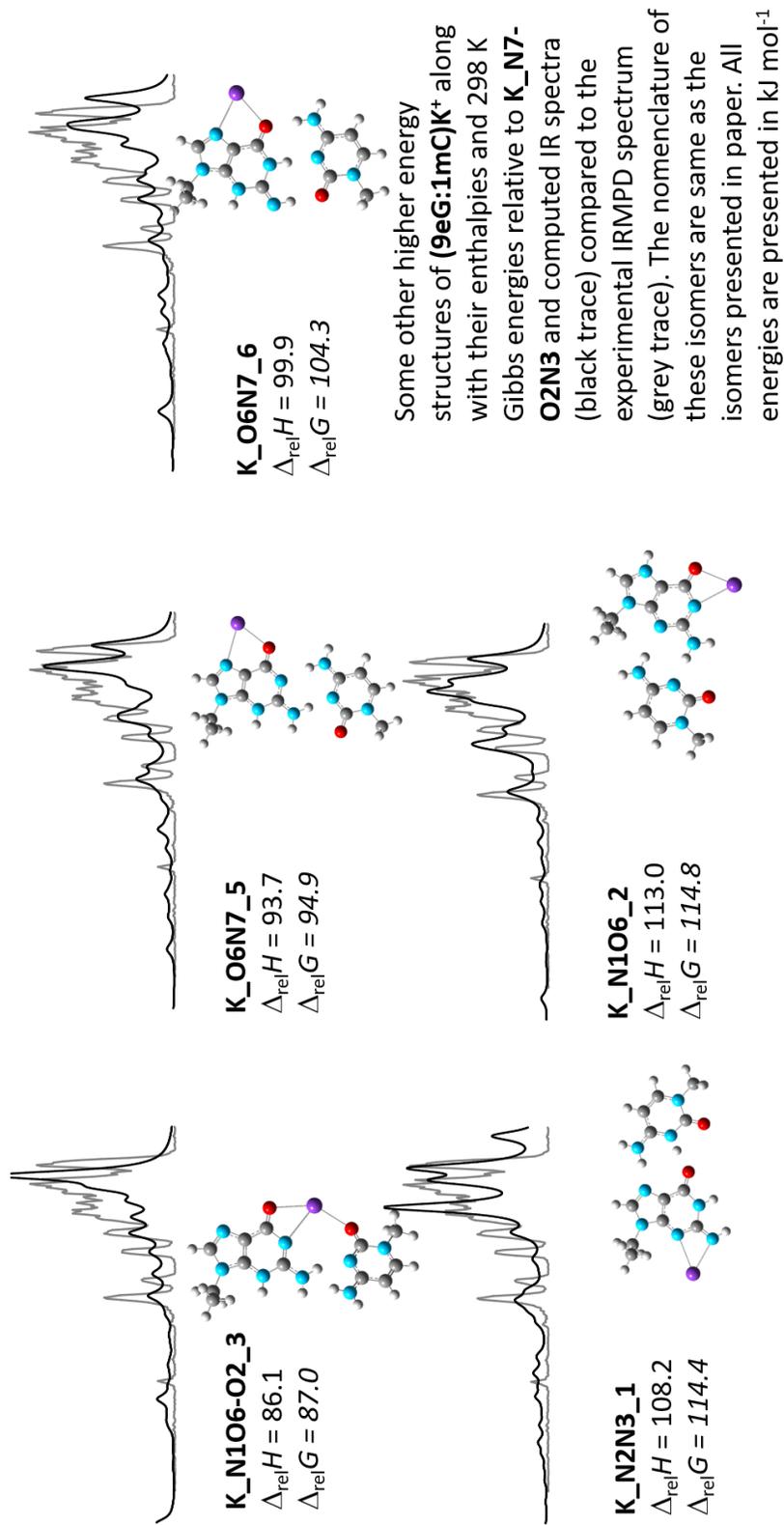
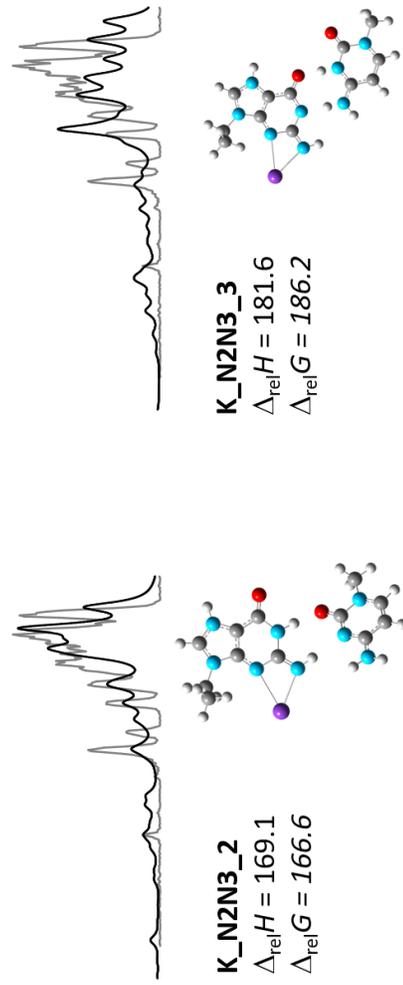


Figure S15



Some other higher energy structures of **(9eG:1mC)K<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **K\_N7-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S16

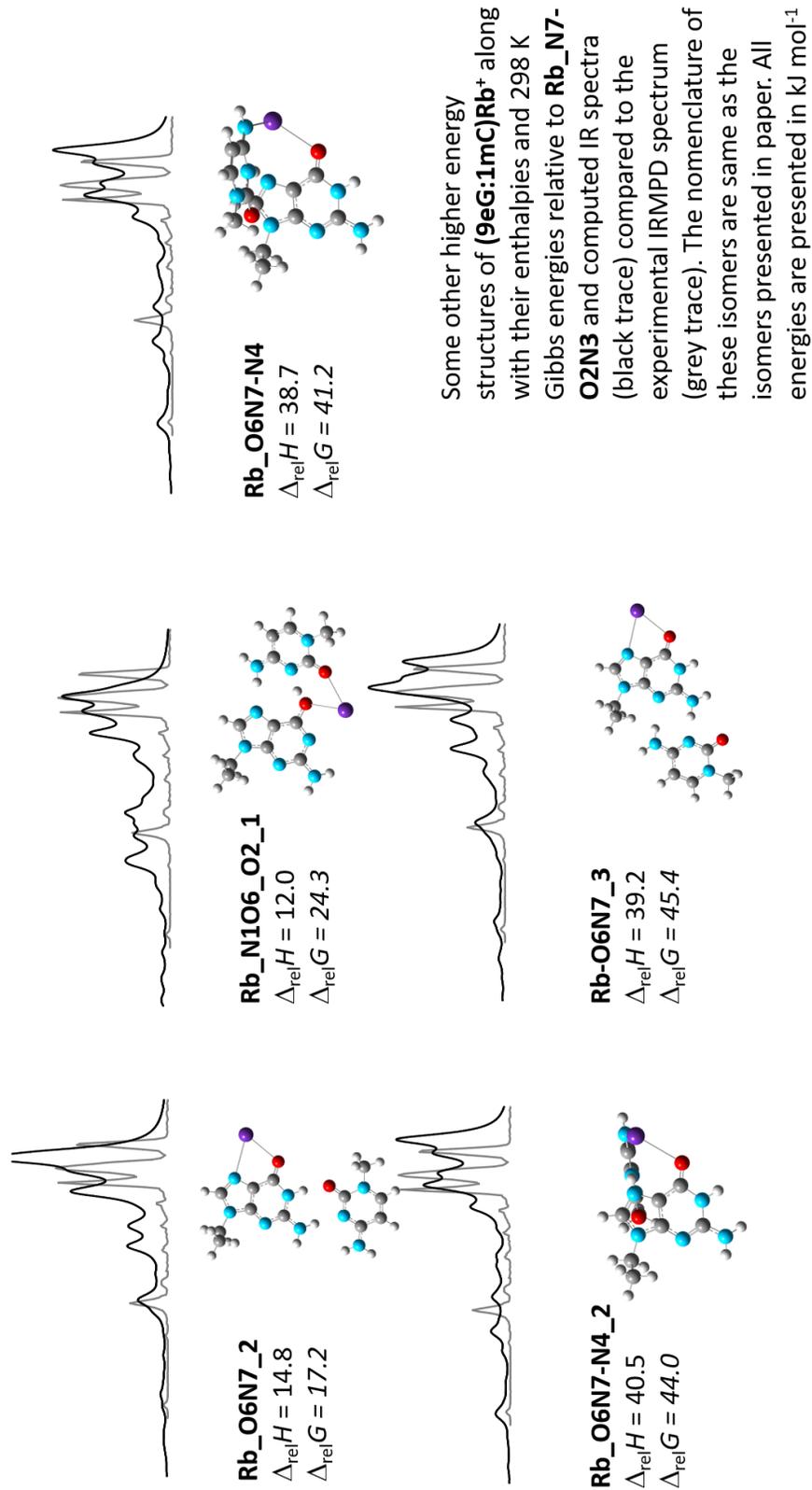


Figure S17

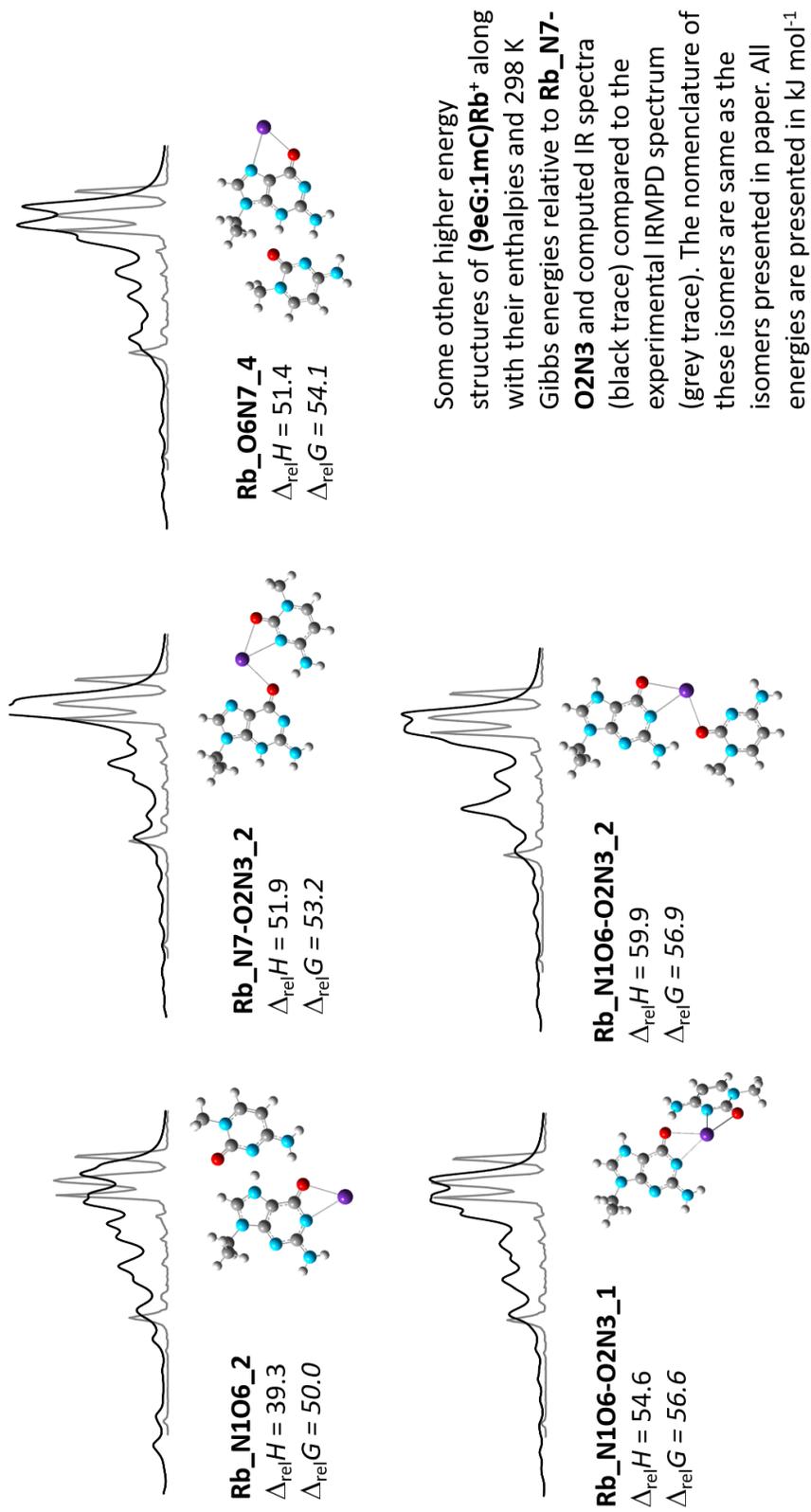
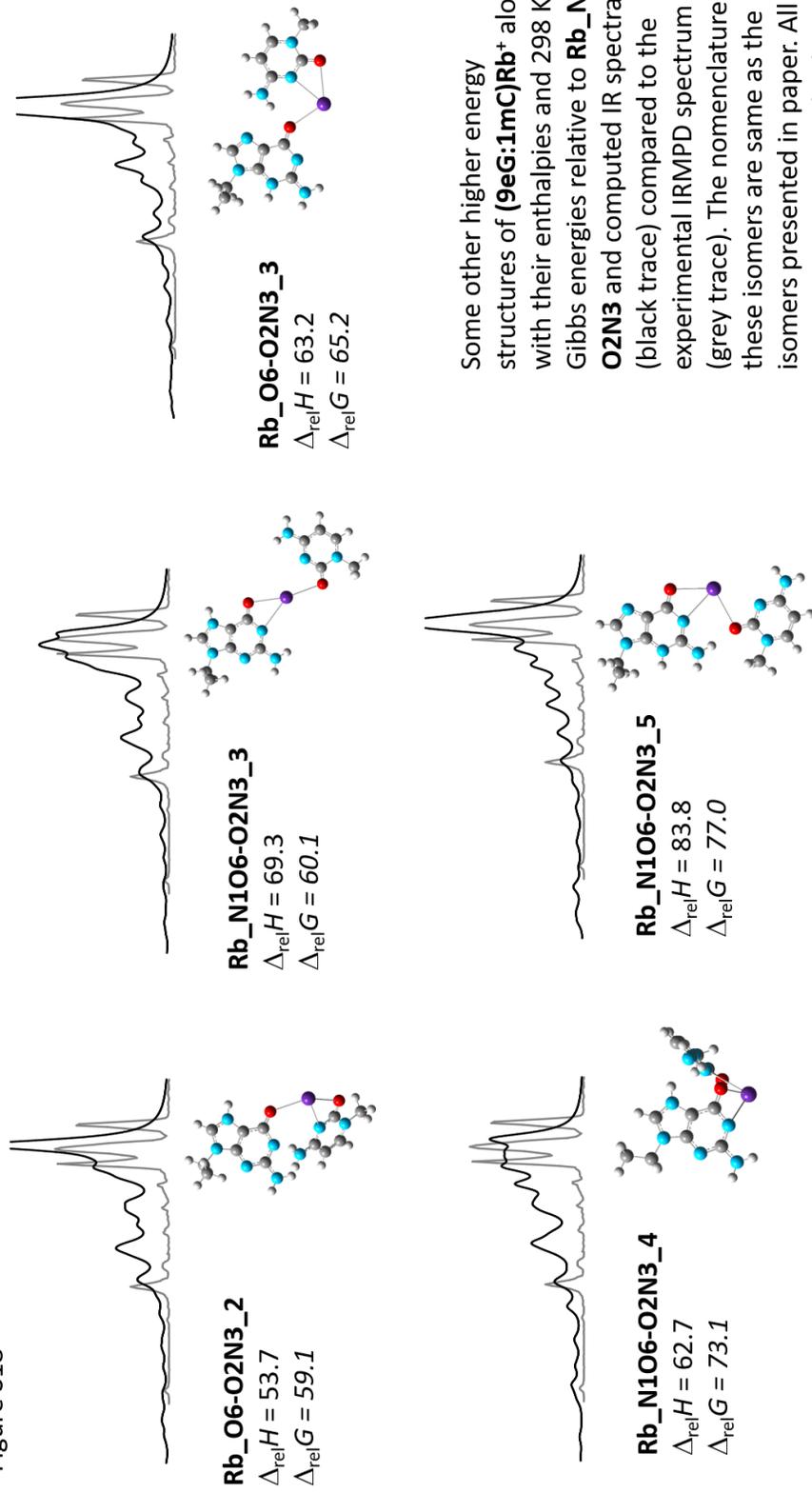


Figure S18



Some other higher energy structures of **(9eG:1mC)Rb<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Rb\_N7-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S19

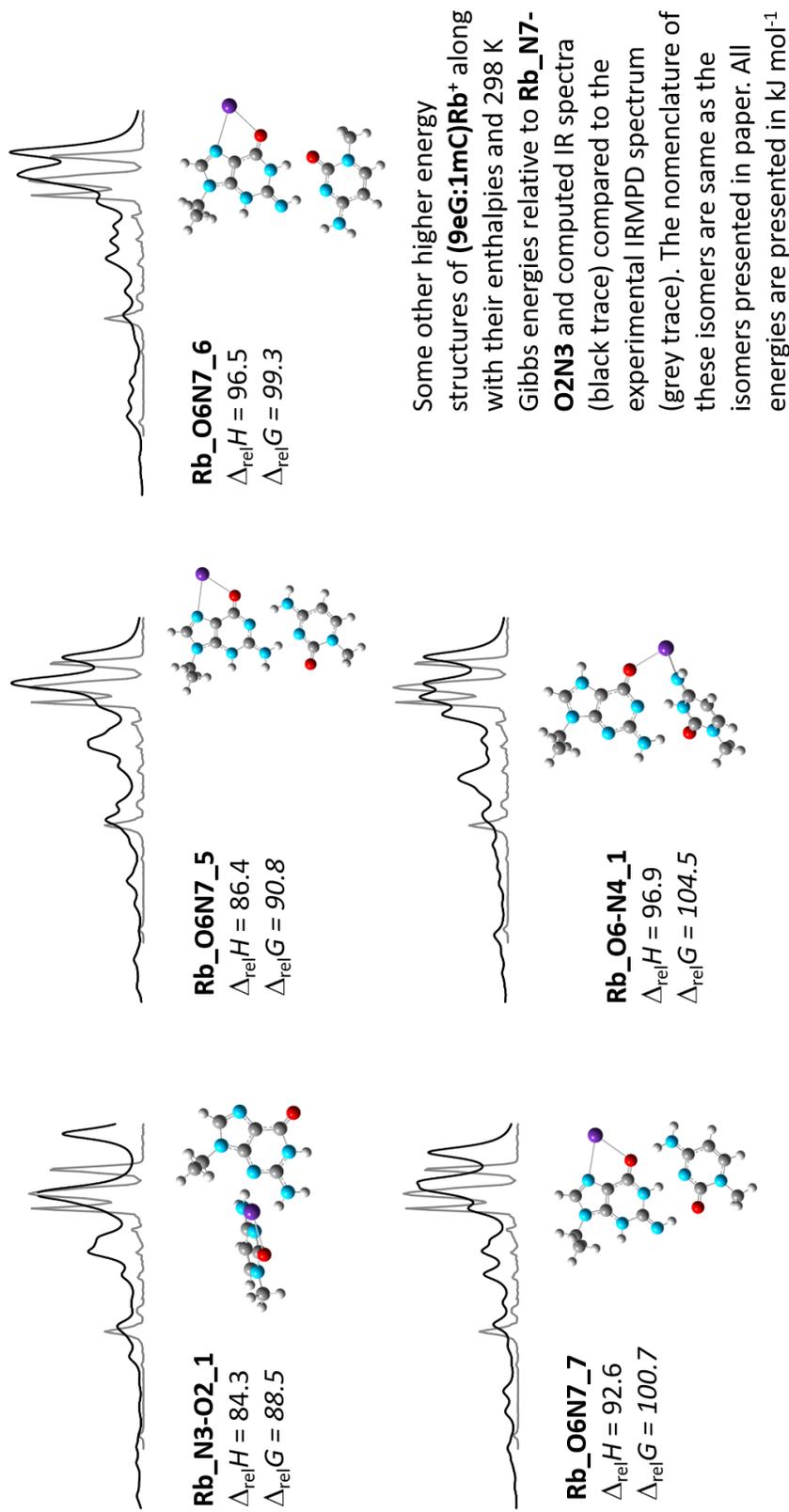
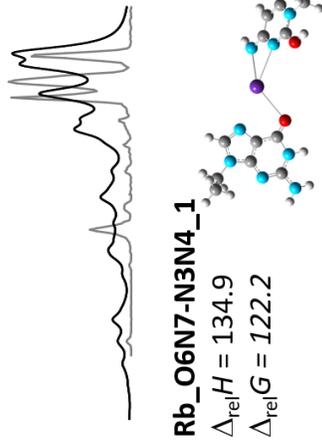
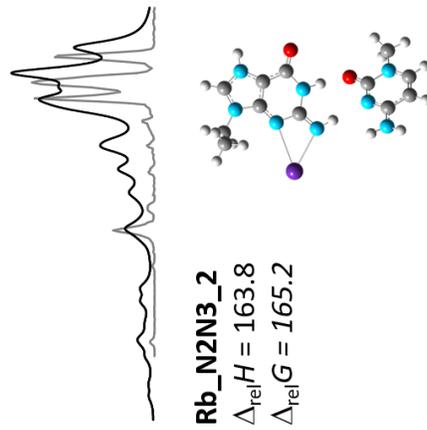
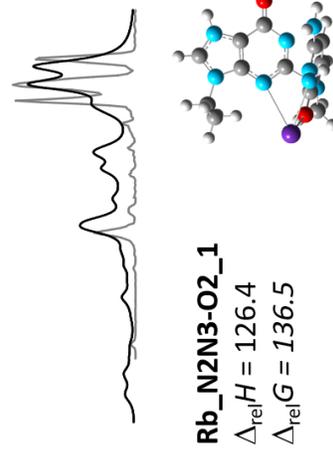
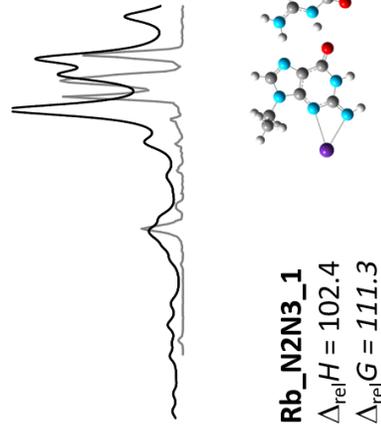
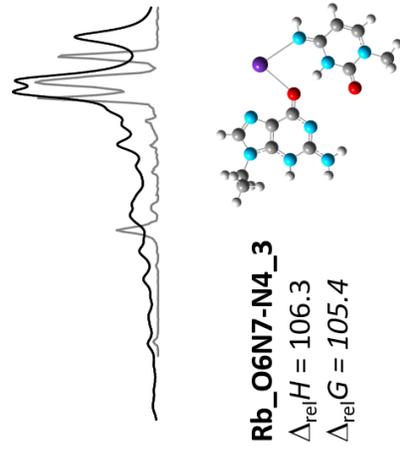
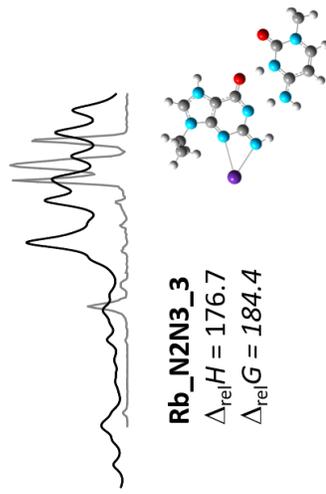


Figure S20



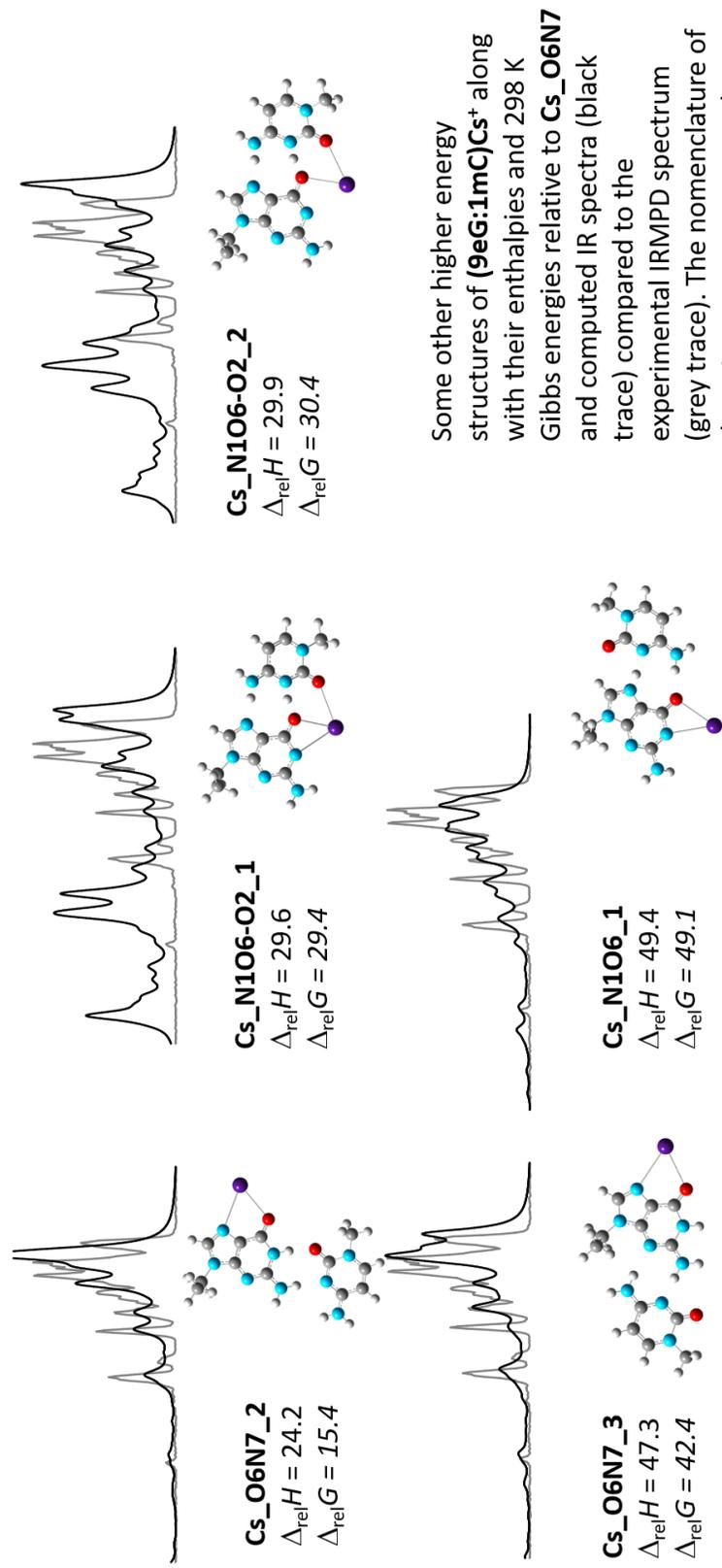
Some other higher energy structures of **(9eG:1mC)Rb<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Rb\_N7-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S21



Some other higher energy structures of **(9eG:1mC)Rb<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Rb\_N7-O2N3** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S22



Some other higher energy structures of **(9eG:1mC)Cs<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Cs\_06N7** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S23

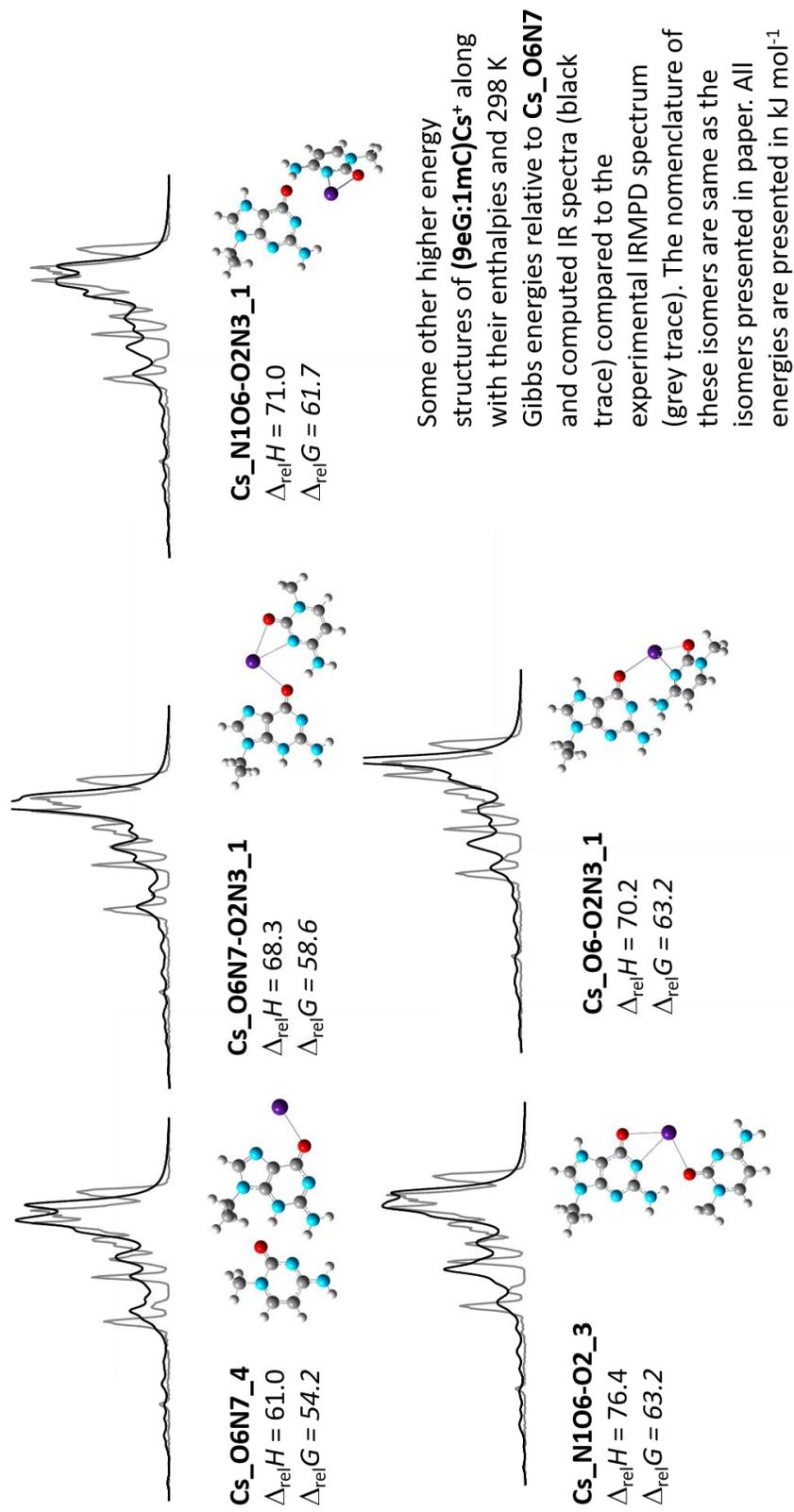


Figure S24

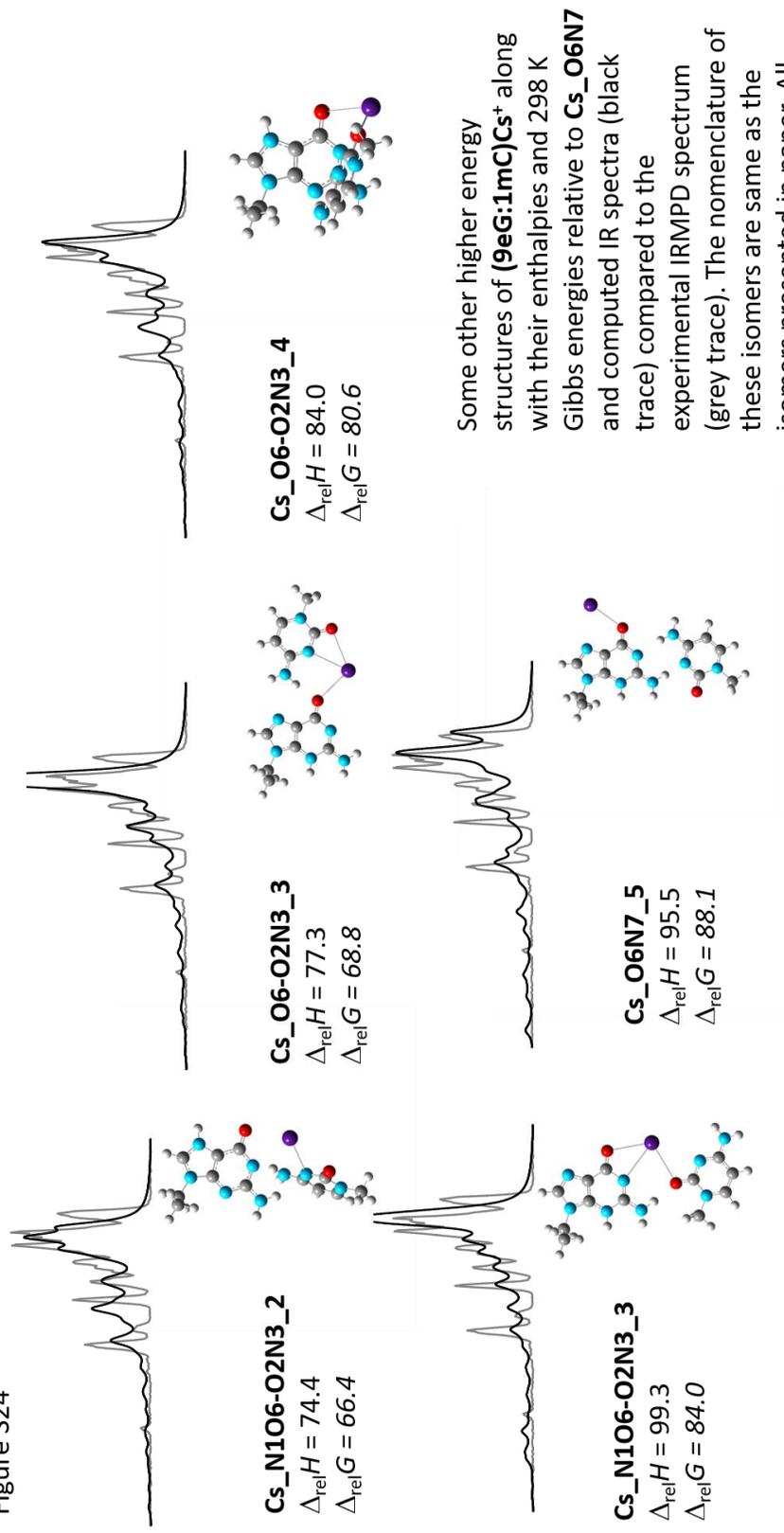


Figure S25

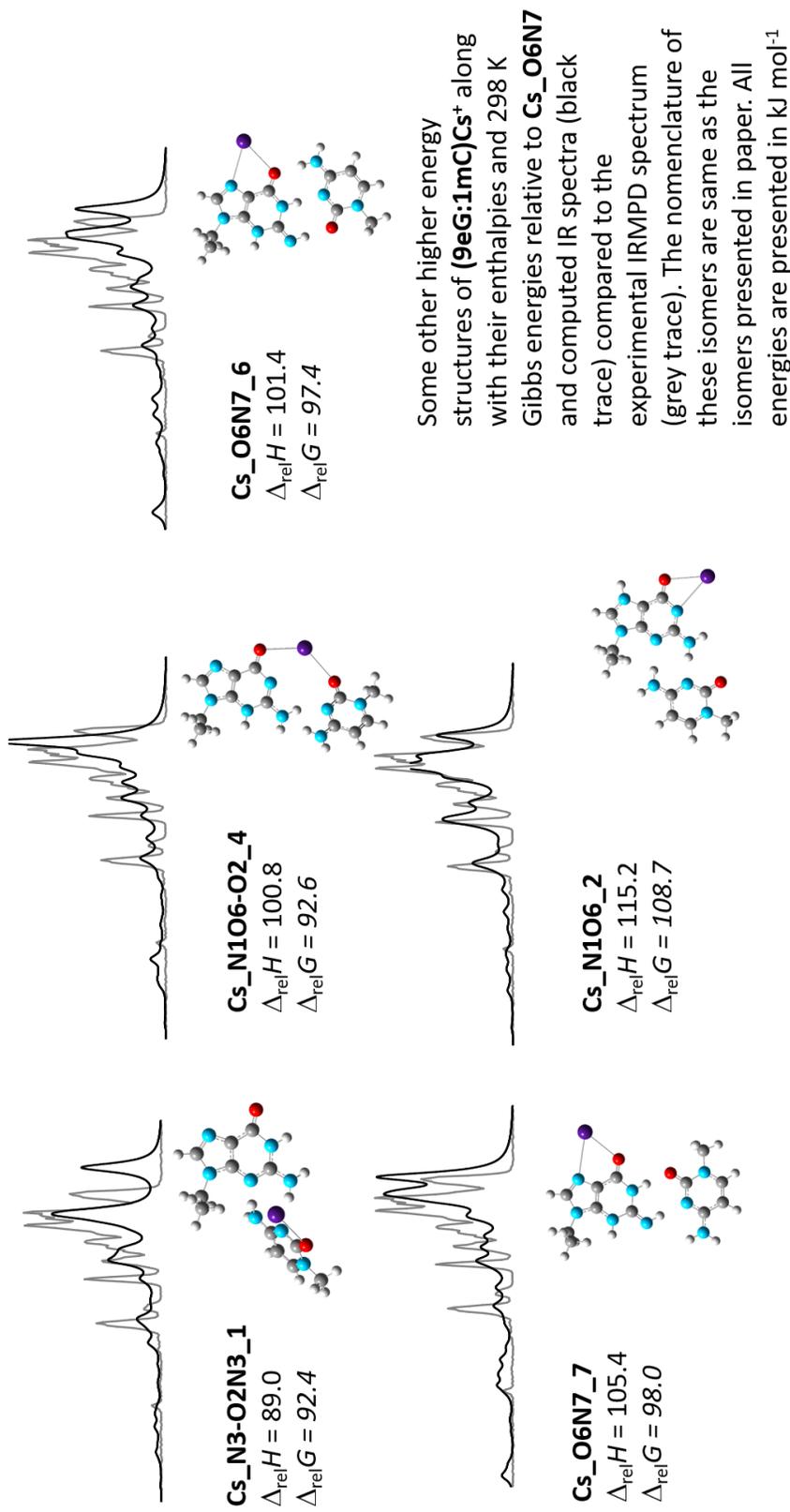
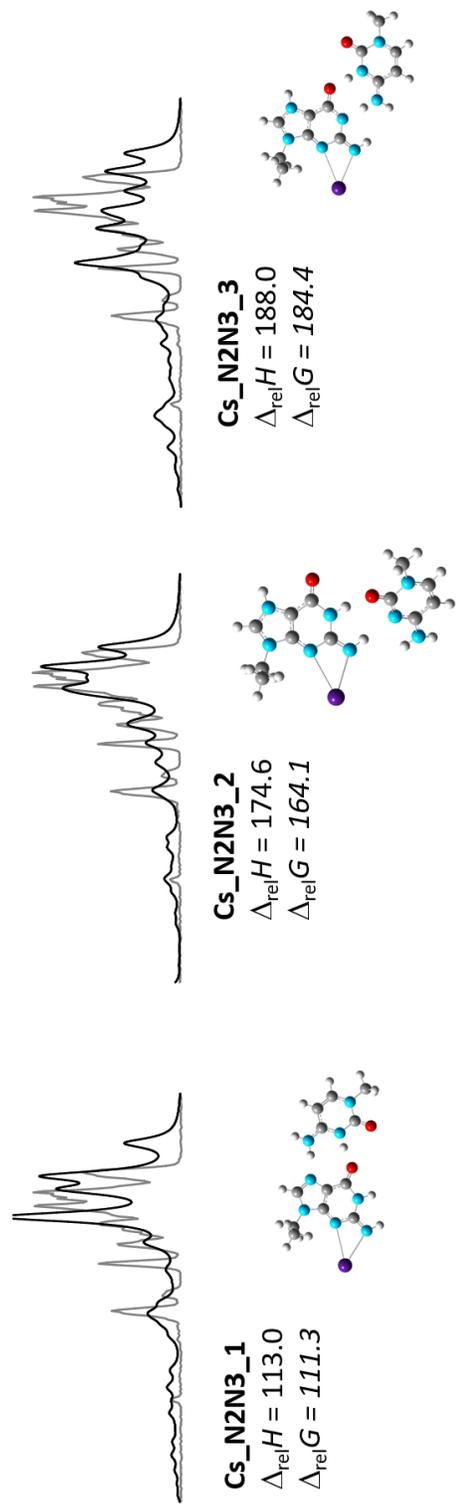
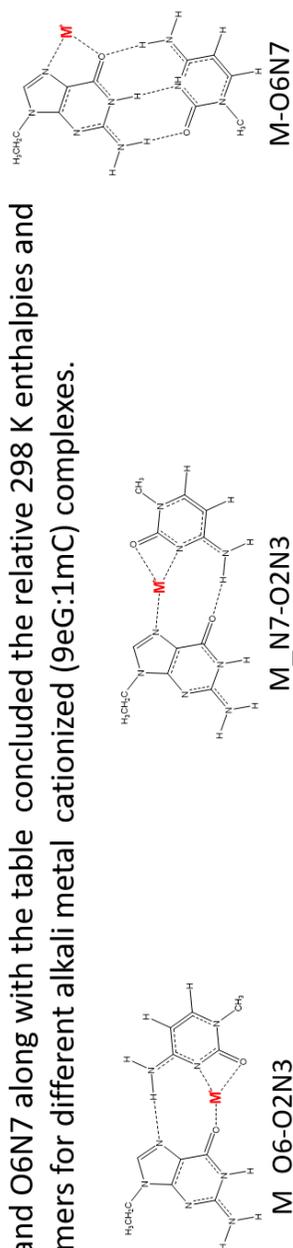


Figure S26



Some other higher energy structures of **(9eG:1mC)Cs<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **Cs\_O6N7** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). The nomenclature of these isomers are same as the isomers presented in paper. All energies are presented in  $\text{kJ mol}^{-1}$

Figure S27. Schematic of metal cationized O6-O2N3, N7-O2N3 and O6N7 along with the table concluded the relative 298 K enthalpies and Gibbs energies of isomers for different alkali metal cationized (9eG:1mC) complexes.



<b>M</b>	<b>M_O6-O2N3</b>		<b>M_O6_O2N3 Derivatives</b>		<b>M_N7-O2N3</b>		<b>M_O6N7 Derivatives</b>	
	$\Delta_{ref}H$	$\Delta_{ref}G$	Total Number of similar Isomers	$\Delta_{ref}H$	$\Delta_{ref}G$	Total Number of similar Isomers	$\Delta_{ref}H$	$\Delta_{ref}G$
Li	<b>0.0</b>	<b>0.0</b>	6	0.0	0.9	0	16.9	26.1
Na	8.0	9.0	0	<b>0.0</b>	<b>0.0</b>	0	26.1	16.9
K	10.2	7.7	0	<b>0.0</b>	<b>0.0</b>	0	-2.3	6.0
Rb	8.0	8.4	0	<b>0.0</b>	<b>0.0</b>	0	-9.4	2.0
Cs	N/A	N/A	0	17.2	5.4	5	<b>0.0</b>	<b>0.0</b>

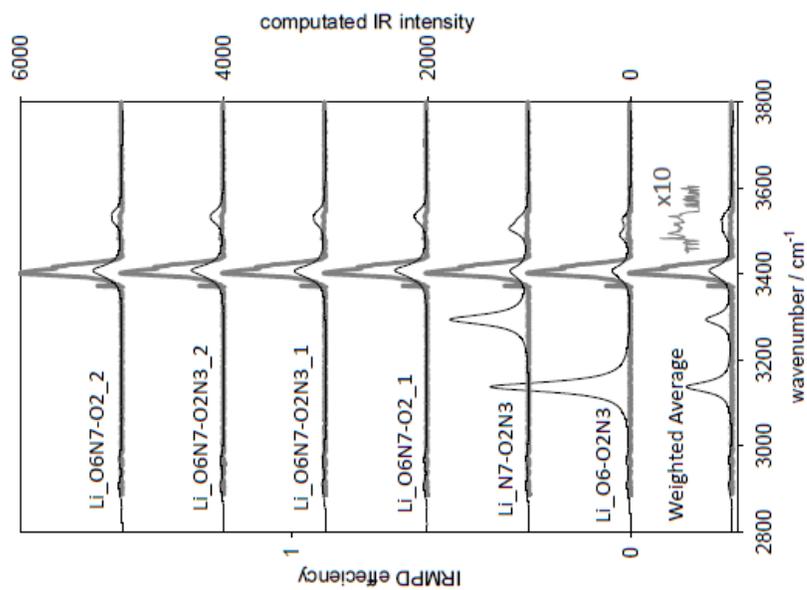


Figure S28. OPO IRMPD spectra for (9eG:1mC)Li<sup>+</sup>

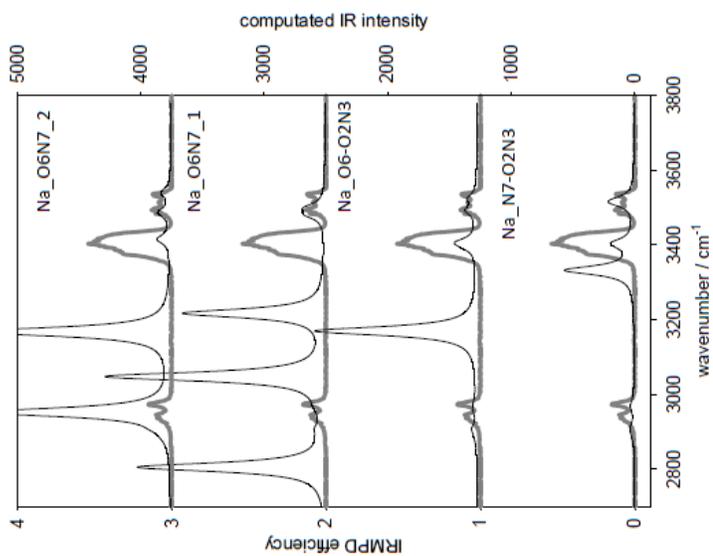


Figure S29. OPO IRMPD spectra for (9eG:1mC)Na<sup>+</sup>

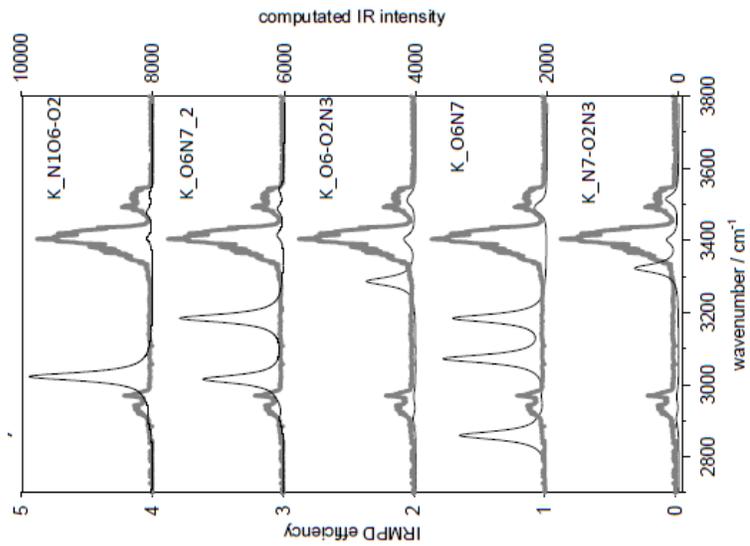


Figure S30. OPO IRMPD spectra for (9eG:1mC)K<sup>+</sup>

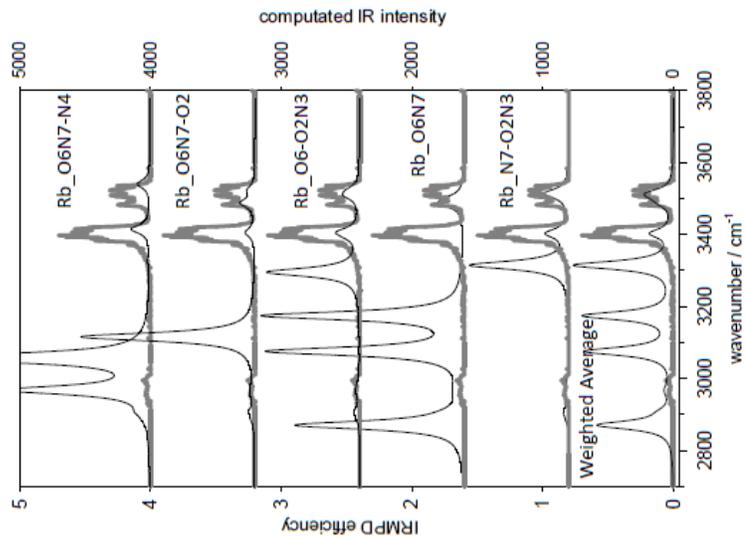


Figure S31. OPO IRMPD spectra for (9eG:1mC)Rb<sup>+</sup>

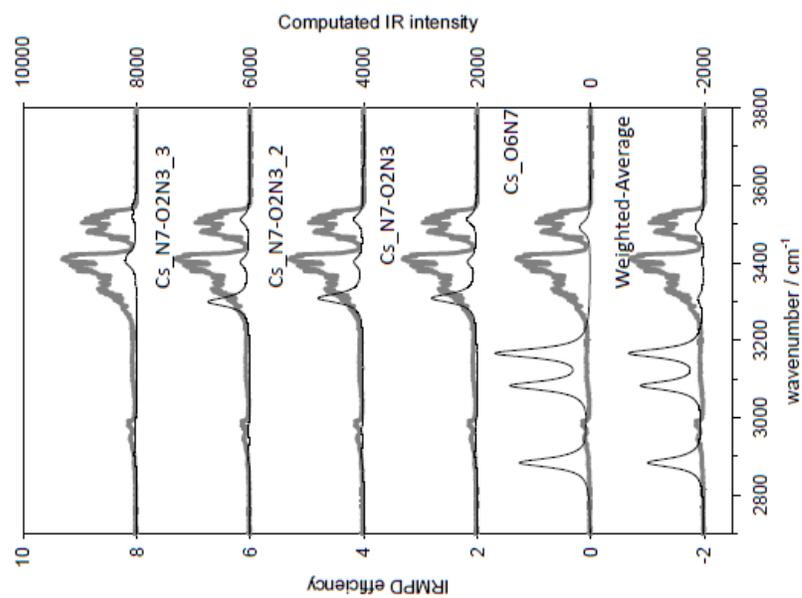


Figure S32. OPO IRMPD spectra for (9eG:1mC)Cs<sup>+</sup>

Structures	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-31+G(3df,3pd)	Structures	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-31+G(3df,3pd)
Hoo N3(1mC)H+	0.0 0.0	0.0 0.0	Li_O6-O2N3	0.0 0.0	0.0 0.0
Hoo N7(9eG)H+	8.1 7.0	8.9 7.8	Li_N7-O2N3	0.9 0.0	1.7 0.7
R-Hoo N3(1mC)H+	18.5 14.8	16.7 13.0	Li_O6N7-O2_1	14.6 2.3	14.0 1.8
WC N7(9eG)H+	12.3 15.7	11.8 15.3	Li_O6N7-O2N3_1	11.2 3.6	10.0 2.4
WC O6(9eG)H+	30.0 27.9	29.8 27.8	Li_O6N7-O2N3_2	11.2 3.7	10.0 2.5
			Li_O6N7-O2_2	14.7 3.8	14.2 3.3

Table S1 Comparison of basis set on the 298K energetics (top relative enthalpies, bottom relative Gibbs energies ) in kJ mol<sup>-1</sup> of five lowest energy (9eG:1mC)H<sup>+</sup>.

Table S2 Comparison of basis set on the 298K energetics (top relative enthalpies, bottom relative Gibbs energies ) in kJ mol<sup>-1</sup> of six lowest energy (9eG:1mC)Li<sup>+</sup>.

Structures	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-31+G(3df,3pd)	Structures	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-31+G(3df,3pd)
Na_N7-O2N3	0.0 0.0	0.0 0.0	K_N7-O2N3	0.0 0.0	0.0 0.0
Na_O6-O2N3	8.0 9.0	7.7 8.6	K_O6N7	-2.3 6.0	-1.3 7.1
Na_O6N7_1	16.9 26.1	18.3 27.6	K_O6-O2N3	10.2 7.7	8.7 6.2
Na_O6N7_2	39.8 41.6	39.4 41.2	K_O6N7_2	21.5 22.0	20.2 20.8
			K_N1O6-O2	18.3 27.9	20.9 30.5

Table S3 Comparison of basis set on the 298K energetics (top relative enthalpies, bottom relative Gibbs energies ) in kJ mol<sup>-1</sup> of four lowest energy (9eG:1mC)Na<sup>+</sup>.

Table S4 Comparison of basis set on the 298K energetics (top relative enthalpies, bottom relative Gibbs energies ) in kJ mol<sup>-1</sup> of five lowest energy (9eG:1mC)K<sup>+</sup>.

Structures	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-31+G(3df,3pd)	Structures	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-31+G(3df,3pd)
Rb_N7-O2N3	0.0 0.0	0.0 0.0	Cs_O6N7	0.0 0.0	0.0 0.0
Rb_O6N7	-9.4 2.0	-7.1 4.3	Cs_N7-O2N3	17.2 5.4	14.9 3.1
Rb_O6-O2N3	8.0 8.4	7.7 8.1	Cs_N7-O2N3_2	17.2 6.0	15.5 4.3
Rb_O6N7-O2	4.7 15.1	9.4 19.8	Cs_N7-O2N3_3	17.0 7.5	14.8 5.3
Rb_O6N7-N4	10.7 15.2	11.9 16.4	Cs_N7-O2N3_4	31.2 9.3	29.0 7.1

Table S5 Comparison of basis set on the 298K energetics (top relative enthalpies, bottom relative Gibbs energies ) in  $\text{kJ mol}^{-1}$  of five lowest energy (9eG:1mC)Rb<sup>+</sup>.

Table S6 Comparison of basis set on the 298K energetics (top relative enthalpies, bottom relative Gibbs energies ) in  $\text{kJ mol}^{-1}$  of five lowest energy (9eG:1mC)Cs<sup>+</sup>.

## **Appendices C**

**(Chapter 6)**

Structures of (9eG:1mT)H <sup>+</sup>	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-311+G(3df,3pd) //B3LYPD3/6-31+G(d,p)	M06-2X/6-31+G(d,p)
GN7_GN7-TO4_TN3-GO6	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
TO4_TO4-GO6_TN3-GN7	2.6 (6.8)	3.4 (7.6)	-3.7 (0.2)
GN7_GN7-TO2_TN3-GO6	10.7 (9.8)	10.2 (9.3)	9.9 (9.2)
GN7_GN1-TO4_TN3-GO6	11.6 (13.0)	11.5 (12.9)	13.6 (14.7)
GN7_GN2-TO2_GN1-TN3_TO4-GO6	7.0 (16.5)	7.9 (17.4)	4.8 (13.2)

Table S1 298K relative enthalpies and Gibbs energies, in kJ mol<sup>-1</sup>, of five lowest energy isomers of (9eG:1mT)H<sup>+</sup> by three different computational methods. Relative Gibbs energies are indicated in parentheses.

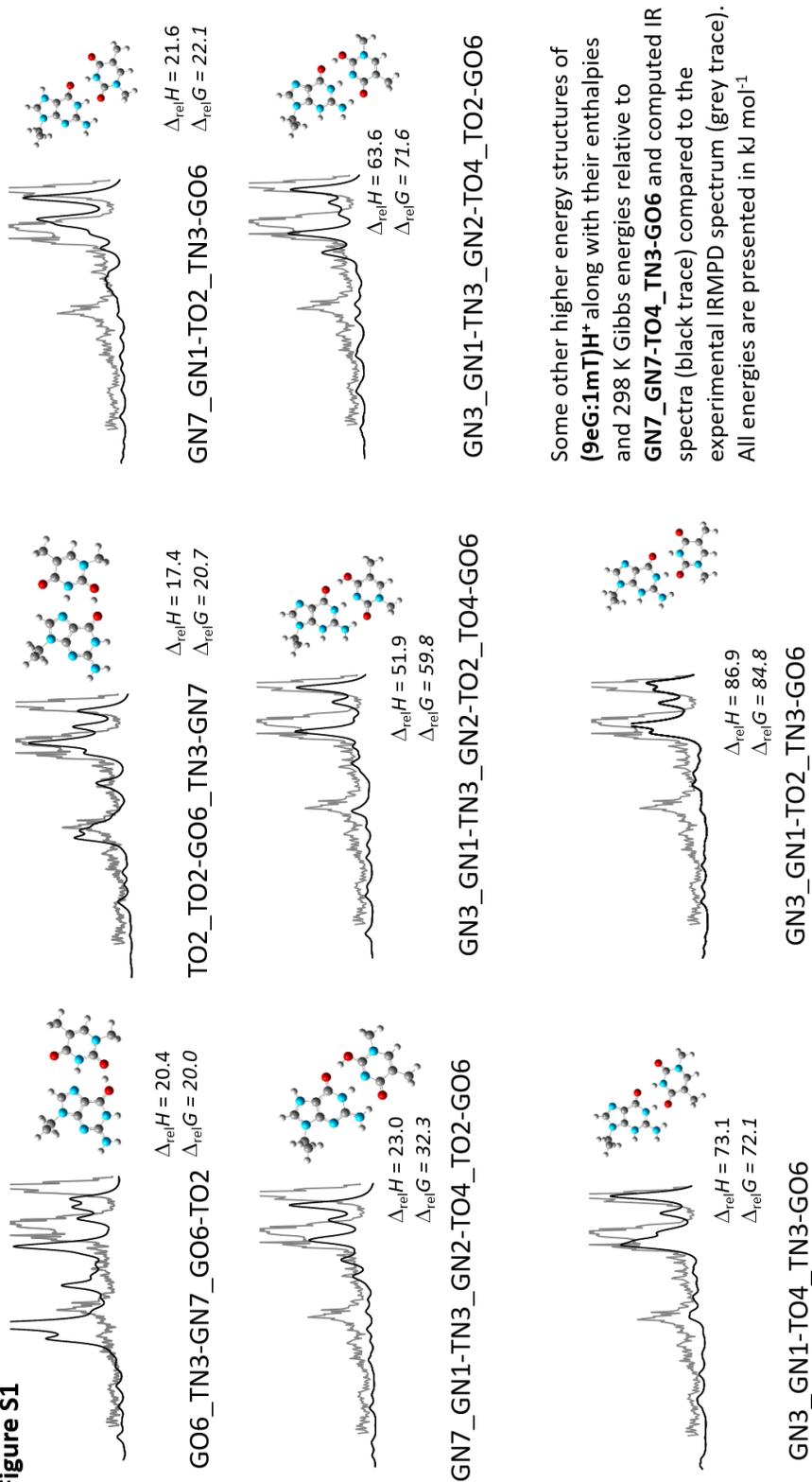
Structures of (9eG:1mA)H <sup>+</sup>	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-311+G(3df,3pd) //B3LYPD3/6-31+G(d,p)	M06-2X/6-31+G(d,p)
AN1_AN1-GO6_AN6-GN7	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
AN1_AN1-GN7_AN6-GO6	2.1 (4.1)	3.8 (1.8)	4.5 (2.6)
AN7_AN7-GO6_AN6-GN7	28.3 (29.3)	28.8 (27.8)	30.7 (31.7)
AN7_AN7-GN7_AN6-GO6	35.6 (36.7)	34.9 (36.0)	33.2 (35.5)

Table S2 298K relative enthalpies and Gibbs energies, in kJ mol<sup>-1</sup>, of four lowest energy isomers of (9eG:1mA)H<sup>+</sup> by three different computational methods. Relative Gibbs energies are indicated in parentheses.

Structures of (9eG:9eG)H <sup>+</sup>	B3LYPD3/6-31+G(d,p)	B3LYPD3/6-311+G(3df,3pd) //B3LYPD3/6-31+G(d,p)	M06-2X/6-31+G(d,p)
GN7_GN7-GN7_GC8-GO6	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
GO6_GO6-GN7_GN1/GN2-GO6	4.1 (7.3)	3.2 (6.4)	-0.2 (3.9)
GN7_GN1-GN7_GN2-GO6	7.1 (12.2)	6.3 (11.5)	10.7 (15.7)
GN7_GN7-GO6_GC8-GN7	15.9 (14.4)	14.4 (12.8)	15.2 (9.8)
GN7_GN7-GO6_GO6-GN7	19.5 (21.6)	17.0 (19.0)	16.7 (15.0)

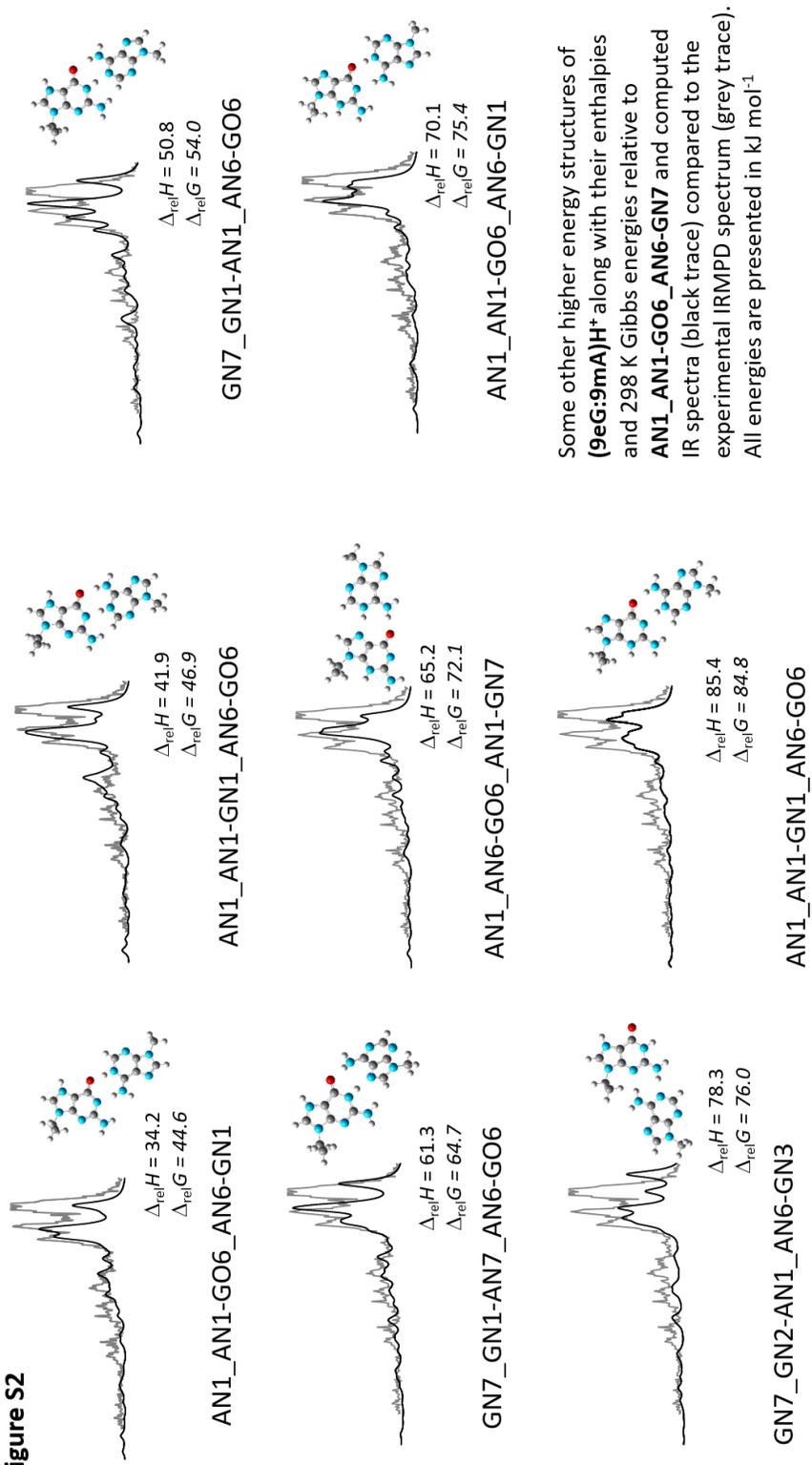
Table S3 298K relative enthalpies and Gibbs energies, in kJ mol<sup>-1</sup>, of five lowest energy isomers of (9eG:9eG)H<sup>+</sup> by three different computational methods. Relative Gibbs energies are indicated in parentheses.

Figure S1



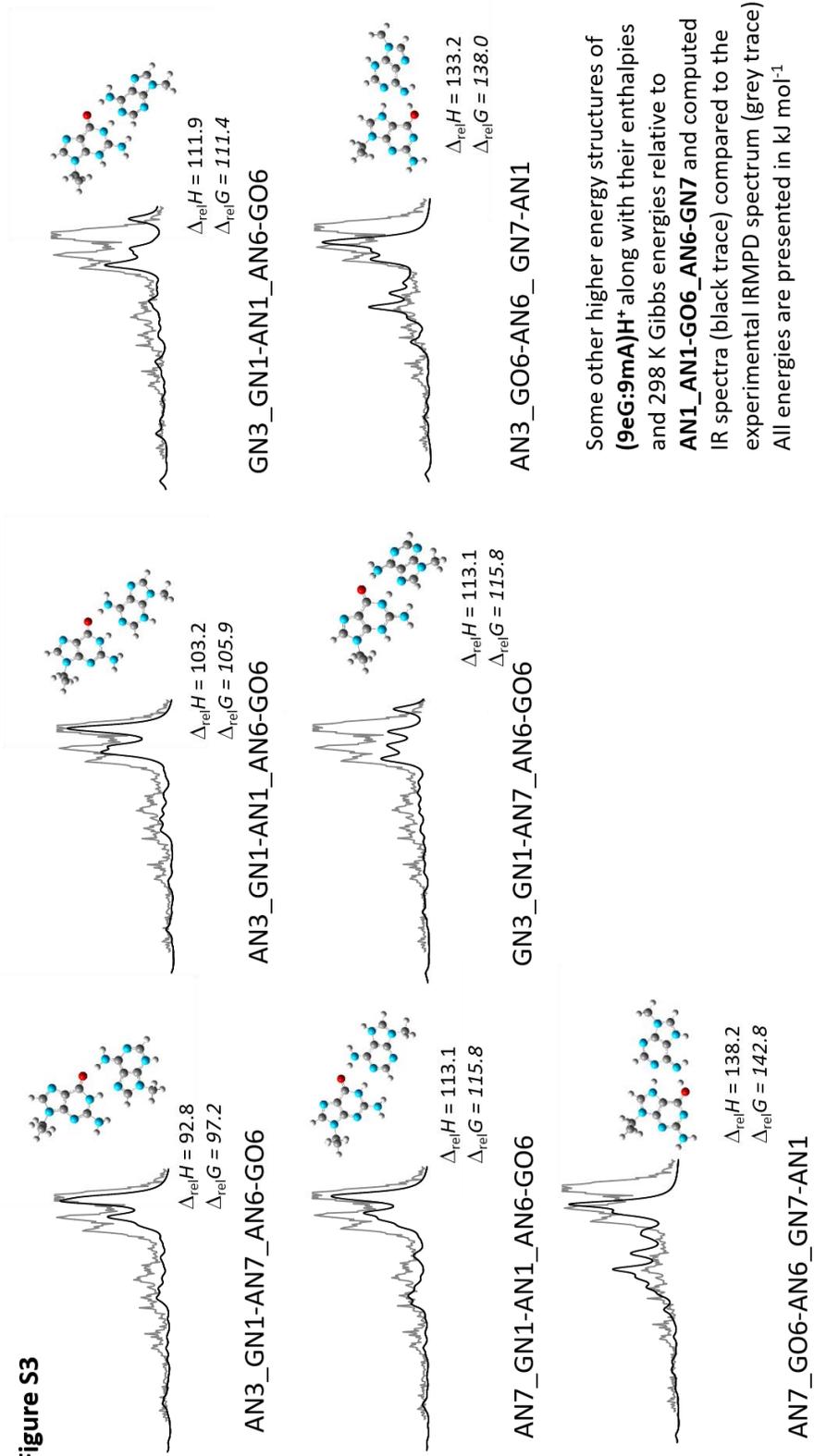
Some other higher energy structures of **(9eG:1mT)H\*** along with their enthalpies and 298 K Gibbs energies relative to **GN7\_GN7-TO4\_TN3-GO6** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in  $\text{kJ mol}^{-1}$ .

Figure S2



Some other higher energy structures of **(9eG:9mA)H<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **AN1\_AN1-GO6\_AN6-GN7** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in  $\text{kJ mol}^{-1}$ .

Figure S3



Some other higher energy structures of **(9eG:9mA)H<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **AN1\_AN1-GO6\_AN6-GN7** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in  $\text{kJ mol}^{-1}$

Figure S4

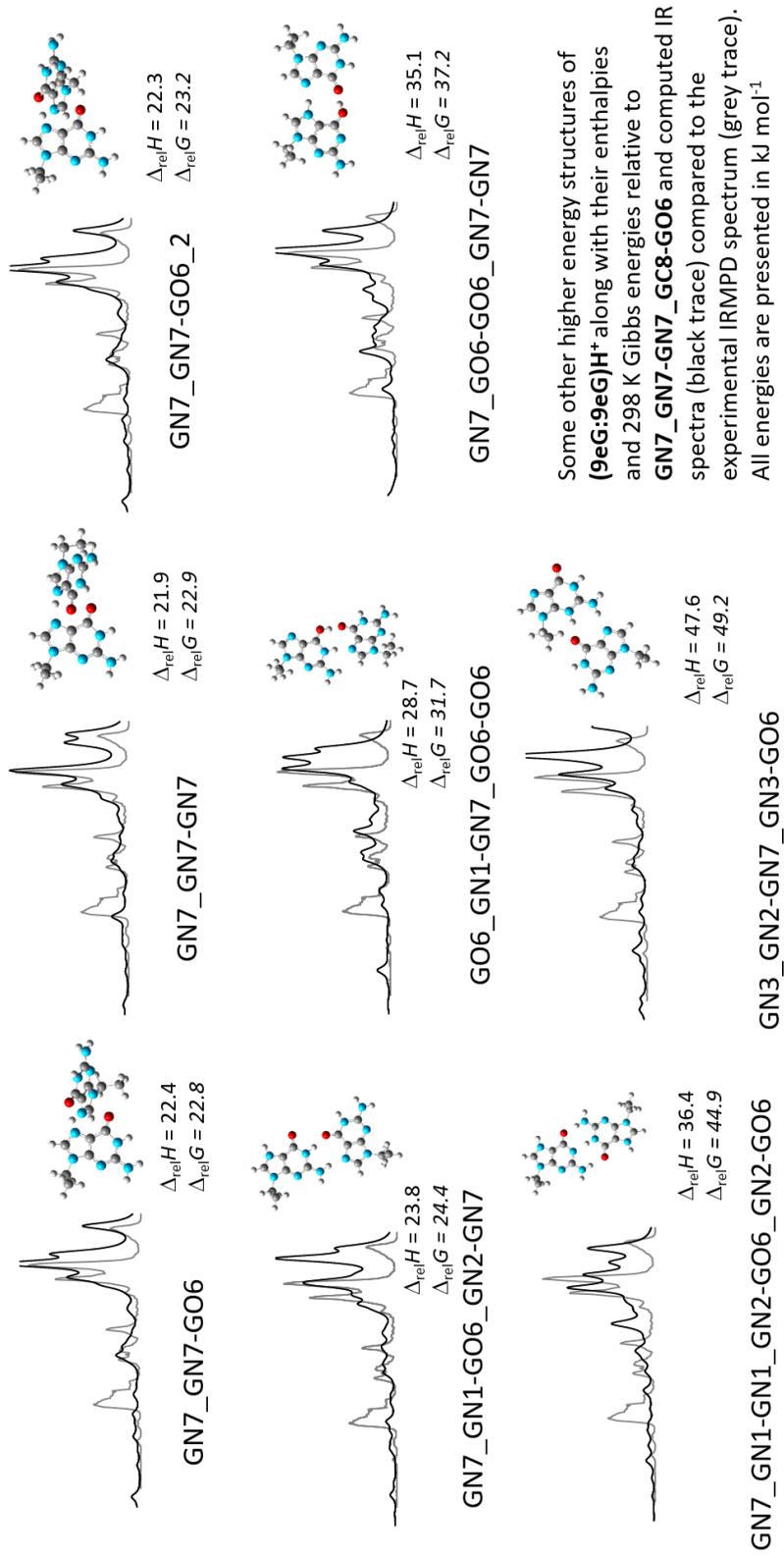
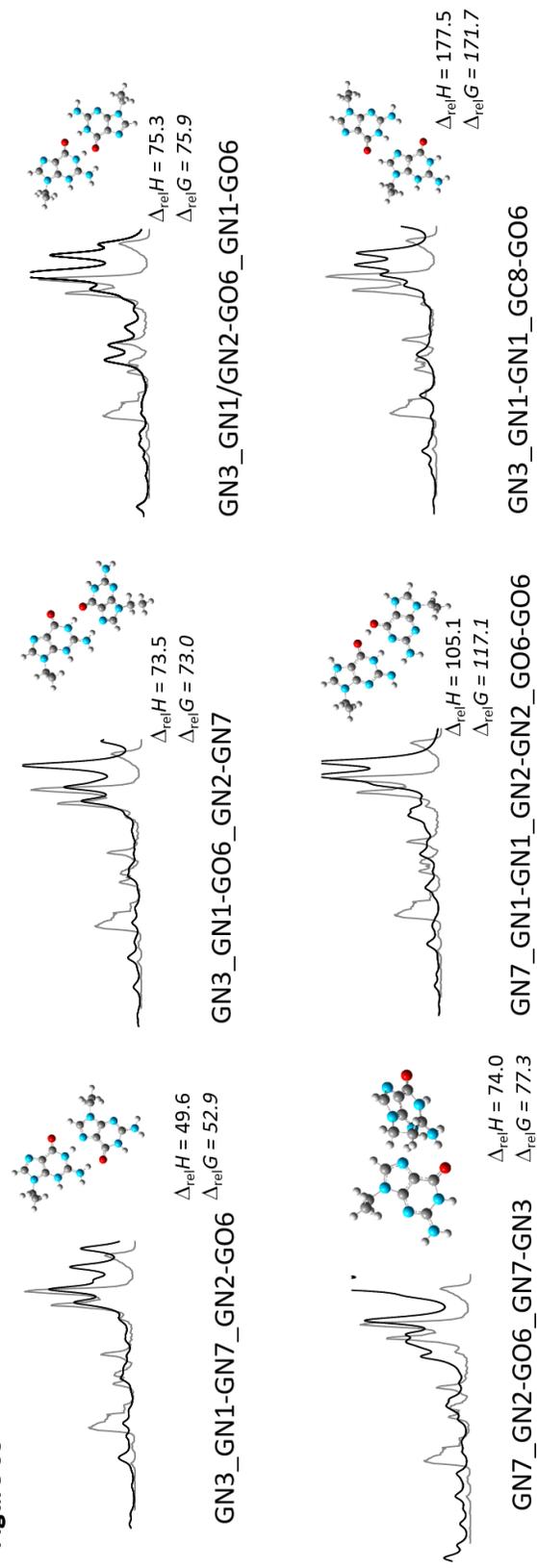


Figure S5



Some other higher energy structures of **(9eG:9eG)H<sup>+</sup>** along with their enthalpies and 298 K Gibbs energies relative to **GN7\_GN7-GN7\_GC8-GO6** and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in kJ mol<sup>-1</sup>