January 2019

Modification Of Bruker Amazon Etd And Solarix Mass Spectrometers Towards Infrared Multiple Photon Dissociation: Structural Characterization Of Modified Nucleosides

Lucas Ash Hamlow
Wayne State University

MODIFICATION OF BRUKER AMAZON ETD AND SOLARIX MASS SPECTROMETERS TOWARDS INFRARED MULTIPLE PHOTON DISSOCIATION: STRUCTURAL CHARACTERIZATION OF MODIFIED NUCLEOSIDES

by

LUCAS A. HAMLOW

DISSERTATION

Submitted to the Graduate School

Wayne State University,

Detroit, Michigan

In partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2019

MAJOR: CHEMISTRY (Analytical)

Approved By:

_____

Advisor Date

_____

_____

_____

ACKNOWLEDGEMENTS

I would like to first thank my advisor Professor Mary T. Rodgers. Not only has she introduced me to the field of mass spectrometry, but also further ignited my passion for building tools to facilitate better examination and understanding of the world around me. Her consistent drive to improve not only her research but also how that research is communicated has been a constant source of inspiration. The breadth of her knowledge and experience has also become a significant aspirational goal, driving me to continue learn new things and expanding my skill set.

I also need to thank Dr. Cliff E. Frieler, Dr. Peter B. Armentrout, and Dr. Chris McNary for many interesting and inspiring conversations pertaining not only to repairing instrumentation and maintaining the laboratory environment, but also kick starting me in my journey into programming and the phenomenal opportunities it can provide. I also need to thank the staff members of the C&IT department at Wayne State University for facilitating the computational work critical to parts of this work.

I would like to thank my committee members, Dr. Colin F. Poole, Dr. Aaron S. Rury, and Dr. Andre R. Venter for their time and assistance. I greatly appreciate the assistance of and discussions with our collaborators and the supporting staff at the FELIX facility, particularly Dr. Giel Berden, Dr. Jonathan Martens, and Dr. Jos Oomens. Their assistance and example has been immensely valuable in several facets of the work presented here.

I thank all the previous and present members of the Rodgers group, but must especially acknowledge Dr. Chenchen He, Dr. Yuan-Wei Nei, Harrison Roy, and Zachary Deveraux for their significant contributions to my knowledge, understanding, and direction. I also must give special thanks to the close friends I have made here at Wayne

State, particularly Dr. Marissa M. Kerrigan, Dr. Habib Baydoun, for their constant encouragement.

Finally, I would like to thank my parents Philip and Stacy Hamlow, and my sister Joan for their continued love and support, and the constant encouragement. The love and support of the rest of my family is also deeply appreciated for helping me keep pushing forward.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

adenine, Ade

adenine arabinoside, araAdo

adenosine, Ado

collision-induced dissociation, CID

cytosine, Cyt

cytosine arabinoside, araCyd

cytidine, Cyd

2′-deoxyadenosine, dAdo

2′-deoxycytidine, dCyd

2′-deoxyguanosine, dGuo

2′-deoxythymidine, dThd

2′-deoxyuridine, dUrd

2′,3′-dideoxyadenosine, ddAdo

2′,3′-dideoxycytidine, ddCyd

2′,3′-dideoxyguanosine, ddGuo

2′,3′-dideoxythymidine, ddThd

2′,3′-dideoxyuridine, ddUrd

deoxyribonucleic acid, DNA

electrospray ionization, ESI

electron transfer dissociation, ETD

Fourier transform ion cyclotron resonance mass spectrometer, FT-ICR MS

free electron laser, FEL

free electron laser for infrared experiments, FELIX

guanine, Gua

guanosine, Guo

guided ion beam mass spectrometer, GIBMS

infrared multiple photon dissociation, IRMPD

intramolecular vibrational redistribution, IVR

5-methyluridine, Thd

optical parametric oscillator/amplifier, OPO

ribonucleic acid, RNA

threshold collision-induced dissociation, TCID

thymine, Thy

thymidine, dThd

transfer RNA, tRNA

uracil, Ura

uridine, Urd

CHAPTER 1 INTRODUCTION

1.1 Nucleic Acid Constituents

Nucleic acid polymers play a critical role in cell function, as they are responsible for both data storage and handling.[1,2] Two different types of nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), are responsible for these functions. Nucleotides are the largest component of all nucleic acids, comprised of a phosphate group, nucleobase residue, and sugar ring as shown in **Figure 1.1**. The nucleoside is a component of the nucleotide consisting of only the nucleobase residue and sugar moiety. The primary difference between DNA and RNA is the composition of their sugar moieties. The monomers of DNA contain a 2′-deoxyribose sugar moiety, whereas the RNA monomers possess a ribose moiety. DNA and RNA share a largely parallel set of standard nucleobase residues. The purine nucleobases adenine (Ade) and guanine (Gua) as well as the pyrimidine nucleobase cytosine (Cyt) are consistent between DNA and RNA. However, uracil (Ura) present in RNA is replaced by thymine (5-methyluracil, Thy) in DNA. The structures of the common DNA and RNA nucleosides described here, adenosine, 2′-deoxyadenosine, guanosine, 2′-deoxyguanosine, cytidine, 2′-deoxycytidine, thymidine and uridine, are shown in **Figure 1.2** as dNuo and Nuo respectively, along with atom numbering for the nucleobase and sugar moieties.

Although the primary structures of DNA and RNA are highly similar, differing in sugar moiety and nucleobase composition as described above, substantial differences are observed in the secondary, tertiary, and quaternary structures adopted by DNA and RNA.[3] DNA generally adopts a double-stranded helix conformation, with several common, and more than a dozen helix geometries available.[4,5] The two primary helix

geometries of double-stranded DNA, A-DNA and B-DNA, differ by the puckering of their sugar rings, resulting in notably different secondary structures.[3] B-DNA displays C2′-*endo* sugar puckering resulting in a longer 3′-phosphate to 5′-phosphate distance and a longer, narrower helix.[6] Whereas A-DNA, with C3′-*endo* sugar puckering, displays a shorter 3′-phosphate to 5′-phosphate distance, which results in a shorter, wider helix structure.[3] The change in helix geometry also results in a change in accessibility to the base-pairs through the major and minor grooves of the helix, and consequently a change in reactivity.[7-11] In contrast to DNA, RNA is primarily single stranded, and able to form double helical regions by folding back on itself, although double-stranded helical structures do exist in RNA.[12,13] The common secondary structural elements of RNA include hairpin loops and stems, bulges, internal loops, and junctions, all of which contribute to the broad variety of functions performed by RNA, and are commonly facilitated by a single strand.[14,15] An excellent example of the complex tertiary structures adopted by single stranded RNA is tRNA, an example of which is shown in **Figure 1.3**.[16] tRNA exhibits four distinct helical stem regions, three of these regions also contain loops as well as a variable loop region, each of which serves a distinct purpose.[16] The diversity of structure present in tRNA facilitates interesting and unique function, that is often facilitated by nucleotide modification.[17,18]

Naturally-occurring nucleotide modifications are crucial to a variety of processes involving DNA and RNA, from repair to structure and function. In particular, RNA is highly modified, which is crucial to its tertiary structure(s) and the myriad of functions it performs. Methylation is one of the most common modifications, at either the nucleobase or 2′-position of the sugar moiety and within tRNA contributes to stability. At certain positions

on the nucleobase however, much more chemically diverse modifications are also observed that facilitate unique chemical reactivity or help stabilize alternative structures.[19] Leveraging and expanding upon the chemical diversity of nucleic acids, synthetically modified nucleosides have been pursued as pharmaceutical targets for decades.[20-24] Idoxuridine (iododeoxyuridine, 5-iodo-2′-deoxyuridine) becoming the first pharmaceutical antiviral agent in 1962[25] for the treatment of herpes simplex keratitis. Since then, nucleoside analogues have become an important group of pharmaceutical compounds.[26-29]

This thesis work expands the characterization of the gas-phase structures of protonated nucleosides by IRMPD action spectroscopy and complimentary theoretical calculations towards the ever-expanding set of modified nucleosides that play crucial roles biologically. The protonated gas-phase structures of three families of modified nucleosides are examined. These families include the naturally-occurring thiated uridines, 2-thiouridine and 4-thiouridine; the pharmaceutically relevant arabinoside analogues of adenosine, guanosine, cytidine, and uridine; as well as the analytically and pharmaceutically relevant synthetic 2′,3′-dideoxyribose analogues of adenosine, guanosine, cytidine, thymidine, and uridine.

1.2 Nucleoside Structural Parameters

Extensive work towards thorough characterization of the structures of the canonical DNA and RNA nucleosides and a wide range of synthetic and modified nucleosides was performed using crystallography and NMR in the 1970s through the 1990s.[30-42] An important outcome of these extensive studies was the development of a set of straightforward parameters to efficiently describe specific nucleoside

conformations.[42] Several of those parameters are used in this research to characterize the gas-phase conformations adopted by protonated nucleosides and are detailed in the following sections.

1.2.1 Sugar Puckering

The furanose sugar ring is especially flexible and its conformation is described by 40 distinct possible conformations. These conformations are effectively described by a pseudorotation angle (P) calculated from the five dihedral angles of the ring: ν_0, ν_1, ν_2, ν_3, and ν_4 (∠C4′O4′C1′C2′, ∠O4′C1′C2′C3′, ∠C1′C2′C3′C4′, ∠C2′C3′C4′O4′, and ∠C3′C4′O4′C1′), respectively. P is then calculated using **Equation 1.1**.

$$\tan P = \frac{(\nu_4+\nu_1) - (\nu_3+\nu_0)}{2\nu_2(\sin 36° + \sin 72°)} \qquad (1.1)$$

Together, P and ν_2 are used with the pseudorotation wheel shown in **Figure 1.4** to determine the specific sugar puckering mode.[42] Each mode is designated as either envelope (E) or twisted (T) where either one or two atoms, respectively, are puckered out of the plane defined by the other three atoms of the sugar ring. Atoms puckered above the ring (on the same side as the nucleobase) are indicated using superscripts to the left (major pucker) or the right (minor pucker) of the E/T designation. Likewise, atoms puckered below the ring (on the opposite side from the nucleobase) are indicated using subscripts in the same fashion. Therefore, sugar pucker labels such as $^2T_3$ indicate that the C2′ atom is the major pucker above the plane of the ring and the C3′ atom is the minor pucker and below the plane of the furanose ring. These 40 specific sugar puckering modes are also inclusive of the common sugar puckering labels such as C2′-*endo*, which encompasses the range of ^2E, $^2T_1$, and $^2T_3$ conformations. These labels are also included

in **Figure 1.4** and can be generally applied to pseudorotation angles within ±18° of the labeled sugar pucker.

1.2.2 Nucleobase Orientation

The orientation of the nucleobase is described by the glycosidic bond angle, ∠O4′C1′N1C2 for the pyrimidine nucleosides and ∠O4′C1′N7C6 for the purine nucleosides. The glycosidic bond angle is used to identify either an *anti* nucleobase orientation facilitating Watson-Crick base pairing, or a *syn* orientation facilitating Hoogsteen base-pairing. The nucleobase orientation and the hydrogen-bonding edges involved in base pairing are depicted in **Figure 1.5**. The *anti* orientation is defined for conformations having a glycosidic bond angle of 90°±90°, whereas the *syn* orientation is defined as 270°±90°.The *anti* and *syn* designations each span 180° of the range of glycosidic bond angles, however, typical glycosidic bond angles generally exhibit much narrower ranges.[43] Intermediate glycosidic bond angles lying close to the border between *anti* and *syn*, such as *intermediate-syn* from 90° to 130°, or *intermediate-anti* from 270° to 310°, may also be described as *high-anti* and *high-syn*, respectively. These additional descriptions indicate that although the nucleobase orientation may fall within the strict *anti* or *syn* designation, the specific conformer may facilitate different inter- or intramolecular interactions than those observed for the more common *anti* or *syn* orientations.

1.2.3 5′-Hydroxy Orientation

Another structural parameter commonly examined is the orientation of the 5′-hydroxy moiety.[43] This group is particularly important for gas-phase protonated nucleosides as it is readily available for intramolecular hydrogen bonding, often to stabilize the nucleobase orientation. There are three distinct designations for the position

of the 5′-hydroxy substituent, measured as the ∠O4′C4′C5′C6′ dihedral angle and depicted in **Figure 1.6**. From 0 to 120° the 5′-hydroxy substituent lies above the plane of the ring and is described as *gauche*$^+$, from 120 to 240° the 5′-hydroxy group is pointed away from O4′ and near C3′ and is designated the *trans* orientation. From 240° to 360° the 5′-hydroxy points away from C3′ and near O4′ and is designated as the *gauche*$^-$ orientation. Prior studies of the gas-phase conformations of the protonated canonical DNA and RNA nucleosides found that the 5′-hydroxy substituent is the most favorable intramolecular hydrogen-bond acceptor, preferentially in the *gauche*$^+$ orientation.[44-50] As such, changes in the 5′-hydroxy orientation might be expected to reflect heavily in the intramolecular binding modes accessible to a particular conformer.

1.2.4 Hydrogen-Bonding Interactions

In the absence of intermolecular interactions with solvent or nearby nucleosides, as found in the gas-phase, intramolecular interactions are crucial to the stabilization of the nucleobase orientation, sugar puckering, and 5′-hydroxy orientation. The most common intramolecular interactions in isolated gas-phase protonated nucleosides are hydrogen-bonding interactions between the nucleobase and sugar, specifically the 5′-hydroxy moiety of the sugar.[44-55] Alternative nucleobase-sugar hydrogen-bonding interactions specifically with the 2′-hydroxy moiety have also been studied for the RNA nucleosides,[44-55] but typically lie higher in relative Gibbs energy. Secondary hydrogen-bonding interactions may occur between the sugar hydroxy moieties, such as the 2′- and 3′-hydroxy, or the 3′- and 5′-hydroxy substituents. As the intramolecular hydrogen-bonding interactions are often better descriptions of protonated gas-phase nucleoside conformers, the descriptors of nucleobase orientation and 5′-hydroxy

orientation are incorporated into the hydrogen bonding designations. For example, an intramolecular noncanonical hydrogen-bonding interaction between the 5′-hydroxy and C6 of a pyrimidine nucleobase might be designated $C6H_a\cdots O5'_{g+}$, where $C6H_a$ indicates the nucleobase orientation is *anti*, C6H is the hydrogen-bond donor, whereas the hydrogen-bond acceptor, the 5′-hydroxy oxygen atom, is in the *gauche*$^+$ orientation.

1.3 Gas-Phase Protonated DNA and RNA Nucleosides

Increased accessibility in recent years to effective tools for the detailed study of gas-phase ion structure has led to efforts towards extending the characterization of nucleoside structure in the absence of intermolecular interactions. A comprehensive characterization of nucleic acid monomers from nucleobase to mononucleotide and under several states of ionization ranging from deprotonation and protonation, to metal cationization, has followed.[51,52,56-68] Protonation plays a role in DNA repair[69] as well as altering base-pairing[70,71] in unique secondary structures such as the i-motif.[72] Protonation provides a readily accessible method for providing access to these biomolecules in the mass spectrometer, and provides insight into protonation sites. Thus far, the work of the Rodgers group and others has primarily focused on characterization of the DNA and RNA analogues. Brief summaries of the known gas-phase intrinsic structures of the protonated canonical DNA and RNA nucleoside analogues of adenine, guanine, cytosine, thymine, and uracil are given in the following sections.

1.3.1 Protonated Adenine Nucleosides

Gas-phase experiments and theory have determined the preferred protonation site of adenine (Ade) to be the N1 position.[73,74] However, subsequent experiments by Wu et al.[52] and Touboul et al.[75] by IRMPD, and the kinetic method, each supported by theoretical

calculations, determined N3 to be the preferred site of protonation for 2′-deoxyadenosine (dAdo) and adenosine (Ado). Protonation at the N3 position is found to be stabilized by a strong $N3H^{+}_{s}\cdots O5'_{g+}$ hydrogen-bonding interaction and results in a *syn* nucleobase orientation. Protonation at the N1 or N7 positions is calculated to be much less favorable in the gas-phase for dAdo and Ado. However, the IRMPD experiments of Wu et al.[52] found that the predicted IR spectra of low-energy N1-protonated conformers to be both highly parallel and complementary to those of the lowest-energy N3 protonated conformers. This indicates that protonation at both N3 and N1 are relevant experimentally. In contrast to protonation at the N3 position, protonation at the N1 position is stabilized by a noncanonical $C8H_{a}\cdots O5'_{g+}$ hydrogen-bonding interaction with an *anti* nucleobase orientation. Additional theoretical calculations in a polarizable continuum indicated that although N1 protonation is relatively high in Gibbs energy compared to N3 protonation in the gas-phase, the relative stability order in solution may be reversed. Overall, the low-energy conformers observed in IRMPD experiments and found via theoretical calculations encompass both N3 and N1 protonation, exhibiting C2′-*endo* and C3′-*endo* sugar puckering, with C2′-*endo* preferred energetically, but with both sugar puckers displaying highly parallel spectral signatures. These low-energy conformers also exhibit *gauche*$^{+}$ 5′-hydroxy substituents, which facilitate stronger intramolecular nucleobase-sugar hydrogen-bonding interactions.

1.3.2 Protonated Guanine Nucleosides

The IRMPD action spectra of protonated 2′-deoxyguanosine (dGuo) and Guanosine (Guo), along with complementary theoretical calculations were previously reported by Wu et al..[51] Protonation at the N7 position was observed to be highly preferred

over protonation at the N3 or O6 positions. Low-energy conformers protonated at the N7 position stabilize an *anti* nucleobase via noncanonical C8H$_a$···O5′$_{g+}$ hydrogen-bonding interactions and dominate the experimentally observed population. C3′-*endo* sugar puckering was calculated to be slightly preferred energetically over C2′-*endo* with the spectroscopic analysis in agreement. The absence of the 2′-hydroxy moiety in dGuo results in far fewer low-energy conformers calculated than for Guo, with a larger gap in Gibbs energy between the calculated ground conformer (C3′-*endo*) and next most stable conformer (C2′-*endo*). The presence of the 2′-hydroxy substituent facilitates additional and/or alternative intramolecular hydrogen-bonding interactions between the 2′- and 3′-hydroxy substituents, and leads to a greater number of low-energy conformers calculated for Guo and nucleobase-sugar hydrogen-bonding interactions between the 2′-hydroxy and nucleobase, particularly the N3 position. *Gauche*$^+$ 5′-hydroxy substituents are preferred by both Guo and dGuo, to facilitate the noncanonical C8H$_a$···O5′$_{g+}$ hydrogen-bonding interaction, but conformers displaying a *gauche*$^-$ or *trans* 5′-hydroxy substituent are found within 20 kJ/mol of the calculated ground conformer.

1.3.3 Protonated Cytosine Nucleosides

The gas-phase conformations of protonated 2′-deoxycytidine (dCyd) and cytidine (Cyd) determined by IRMPD spectroscopy and complimentary theoretical calculations were reported by Wu et al.[54] and Filippi et al.[49] Those experiments find that both N3 and O2 protonation are important within the conformers populated. The most stable conformers, protonated at either the N3 or O2 positions, prefer *anti* nucleobase orientations stabilized by noncanonical C6H$_a$···O5′$_{g+}$ hydrogen-bonding interactions with *gauche*$^+$ 5′-hydroxy substituents. Conformers, stabilized by the C6Ha···O5′$_{g+}$

hydrogen-bonding interaction, exhibiting C2′-*endo* sugar puckering and protonated at either the N3 or O2 position were the only conformers populated for dCyd. Although structurally parallel conformers are also primarily responsible for the measured IRMPD spectrum of Cyd, analogous conformers displaying C3′-*endo* sugar puckering may also be accessed by Cyd. These low-energy N3-protonated and O2-protonated conformers all prefer a *gauche*$^+$ 5′-hydroxy substituent, which provides the greatest stability to the noncanonical nucleobase-sugar hydrogen-bonding interaction that stabilizes these conformers. Conformers displaying *syn* nucleobase orientations were found higher in Gibbs energy, with O2 protonation preferentially stabilized by the ionic $O2H^+_s\cdots O5'_{g+}$ hydrogen-bonding interaction vs N3 protonation stabilized by a $O5'H_{g+}\cdots O2_s$ hydrogen-bonding interaction. No conformers displaying *syn* nucleobase orientations were populated by ESI as determined by synergistic IRMPD spectroscopy and theoretical calculations.

1.3.4 Protonated Thymine Nucleosides

The intrinsic gas-phase conformations of protonated 2′-deoxythymidine (dThd) were studied by IRMPD and theoretical calculations by Salpin et al.,[76] and by Wu et al.[55] in conjunction with protonated 5-methyluridine (Thd). Protonation of these thymine nucleosides preferentially occurs at the O2 or O4 position, or induces tautomerization to the 2,4-dihydroxy tautomer. Both prior IRMPD experiments on protonated dThd indicate a strong preference for formation of the 2,4-dihydroxy tautomer, with a minor presence of conformers protonated at the O2 position. The experimental IRMPD spectrum of protonated Thd also indicates the dominant presence of 2,4-dihydroxy tautomers, with at least two distinct conformers protonated at the O2 position also observed. Although

conformers protonated at the O4 position were calculated to be reasonably low in Gibbs energy, their predicted IR spectra preclude them from being populated in the experiments. Several conformers, for both dThd and Thd, exhibiting *anti*-oriented nucleobases and either C2′-*endo* or C3′-*endo* sugar puckering were found within 3 kJ/mol in relative Gibbs energy of their respective ground conformers. Conformers of dThd and Thd possessing C2′-*endo* sugar puckering were calculated to be the ground conformers. These low-energy conformers are stabilized by noncanonical $C6H_a\cdots O5'_{g+}$ intramolecular hydrogen-bonding interactions, very similar to those observed for dCyd and Cyd. Several relatively low-energy conformers with *syn* nucleobases were also calculated that are stabilized by $O2H_s\cdots O5'_{g+}$ hydrogen-bonding interactions. The work of Wu et al. suggests that these *syn* conformers may also be present experimentally, providing reasonable spectroscopic agreement in both the IR fingerprint region and hydrogen-stretching region. *Gauche*$^+$ 5′-hydroxy orientations are preferred energetically by both the experimentally populated conformers displaying *anti* nucleobase orientations, as well as the low-energy *syn* oriented conformers.

1.3.5 Protonated Uracil Nucleosides

Wu et al. also characterized the gas-phase conformations of protonated 2′-deoxyuridine (dUrd) and uridine (Urd) by IRMPD and theoretical calculations.[53] Like Thd and dThd as described in **Section 1.3.4**, protonation of dUrd and Urd is favorable at the O2 or O4 positions, and can induce tautomerization to the 2,4-dihydroxy tautomer. Tautomerization to the 2,4-dihydroxy tautomer is calculated to be energetically preferred, and the measured IRMPD spectra of dUrd and Urd support the presence of several tautomeric conformers in the experiments. However, much like dThd and Thd, several

conformers protonated at the O2 or O4 position are calculated to have reasonably low relative Gibbs energies. In contrast to dThd and Thd, the predicted IR spectra of the O4-protonated conformers provide a better spectral match to the experimental spectrum than those of the O2-protonated conformers and are concluded as being present in the experiments. However, several 2,4-dihydroxy tautomers provide excellent representation of most of the measured spectral features and small features unique to the O4-protonated conformers observed in both the IR fingerprint and hydrogen-stretching regions indicates that these tautomers are likely the primary conformers populated in the experiments. These low-energy tautomers and O4-protonated conformers display an energetic preference for the *anti* nucleobase orientation with C2′-*endo* sugar puckering, nearly all of which are stabilized by a noncanonical C6H$_a$···O5′$_{g+}$ intramolecular hydrogen-bonding interaction. The presence of the 2′-hydroxy moiety allows for the calculated ground conformer to display an alternative O2H$_a$···O2′H···O3′ dual hydrogen-bonding interaction that is only ~0.3 kJ/mol Gibbs energy less stable than the ground conformers. Tautomeric conformers with *syn* nucleobase orientations and C2′-*endo* sugar puckering may also be observed in the experiments for both dUrd and Urd. These tautomeric conformers are stabilized by an O2H$_s$···O5′$_{g+}$ or O5′$_{g+}$···O2$_s$ intramolecular hydrogen-bonding interaction and display highly parallel spectra to the lowest-energy *anti*, C2′-*endo* puckered conformers, but lie higher in Gibbs energy such that if present are present in low abundance. As with the other nucleosides, *gauche*$^+$ 5′-hydroxy orientations are preferred by all of the conformers experimentally populated.

1.4 Nucleoside Modification

A wide variety of nucleoside modifications are observed naturally in RNA, and even more have been pursued synthetically. As mentioned in **Section 1.1**, naturally occurring modification plays a particularly important role in several processes[77-80], and synthetically modified nucleosides have been demonstrated to be powerful pharmaceutical tools[20,21,26,28]. Three specific modifications, including naturally occurring and synthetic, were chosen for study in this work, introduced in **Sections 1.4.1–1.4.3**.

1.4.1 Thiated Uridine Nucleosides

As discussed in **Section 1.1**, tRNA is a highly modified, structurally complex single stranded RNA. Modified nucleosides in tRNA play roles, ranging from facilitation of the appropriate secondary structure for interaction with other macromolecules, to altering the binding specificity of those interaction sites, and protecting the tRNA from damage.[77,80,81] Thiation of uridine, specifically at the 2- or 4-position (**Figure 1.2**), along with several other modifications, results in nucleosides that serve important, but very different roles in tRNA. 2-Thiouridine and its many derivatives, further modified at the 5-position of the nucleobase, play important roles in base pairing at the wobble position (position 34) of the codon region, critical to codon recognition and ultimately protein synthesis.[79,82-84] Occupation of the wobble position by 2-thiouridine or its derivatives allows for greater flexibility in base pairing, such that base pairing with both Ado or Guo is observed, instead of the strong preference that canonical uridine exhibits for Ado.

4-Thiouridine plays a very different role in tRNA than 2-thiouridine, and can be found at position 8, the beginning of the acceptor stem.[85,86] 4-Thiouridine is less commonly found in naturally-occurring tRNA, but is thought to afford some resistance to

UV-induced damage by crosslinking with a nearby Cyd at position 13.[87] This photoactivity is also the basis for broader synthetic use of 4-thiouridine to introduce crosslinking in nucleic acids, such as in the PAR-CLIP procedure.[88,89]

The gas-phase structures of protonated 2- and 4-thiouracil were studied by Nei et al.[59] by IRMPD action spectroscopy and complementary theoretical calculations to determine the favorable sites of protonation. The spectroscopy and predicted IR spectra of 2-thiouracil indicate that formation of the 2-sulfhydryl-4-hydroxy tautomer is favored, and conformers protonated at the O4-position may also be accessed, but their presence was not conclusively shown. In contrast, the spectroscopy and predicted IR spectra of 4-thiouracil indicate that protonation at the S4-position is favored over formation of the 2-hydroxy-4-sulfhydryl tautomer, which is less populated in the experiments. This work extends that study to the analogous nucleosides 2- and 4-thiouridine. In particular, the impact of thiation on nucleoside structure, any impact the sugar moiety may have on protonation preference, and the influence of thiation on the gas-phase conformation of protonated uridine is examined.

1.4.2 Arabinose-Based Nucleosides

Several nucleoside analogs were isolated from sea sponges in the mid-1950s. The first of these was designated spongothymidine, as well as spongouridine.[90,91] These nucleoside analogues were later identified as arabinose-based nucleosides, with the stereochemistry at the 2-position of the classical ribose sugar ring inverted, resulting in the 2′-hydroxy moiety being on the same side of the sugar ring as the nucleobase, shown in **Figure 1.2** as araNuo. Two arabinose nucleosides have found pharmaceutical application, cytosine arabinoside and adenine arabinoside, known as cytarabine and

vidarabine, respectively.[23,92] The pharmaceutical applications of cytarabine and vidarabine are quite different. Cytarabine is a crucial component in the treatment of several cancers such as acute myeloid leukemia (AML), acute lymphocytic leukemia (ALL), chronic myelogenous leukemia (CML) and non-Hodgkin's lymphoma.[93,94] In contrast, vidarabine, though it is no longer marketed in the US, was used to treat the herpes simplex and varicella zoster viruses.[92] Given the relevance of and interest in nucleoside analogues with the arabinose moiety towards pharmaceutical applications, extensive theoretical,[95] crystallographic,[96,97] and NMR studies[30,98] have sought to understand the impact of the stereochemical inversion on nucleoside structure and function. In solution, NMR experiments observed repulsive interactions between the 2′- and 5′-hydroxy moieties. In contrast, crystallography and early gas-phase theoretical studies found evidence for a strong intramolecular hydrogen-bonding interaction between the 2′- and 5′-hydroxy moieties.[95,97] In the previous gas-phase studies of the canonical protonated DNA and RNA nucleosides, summarized in **Section 1.3**, the most common intramolecular hydrogen-bonding interaction observed involved the nucleobase and 5′-hydroxy substituent. Understanding the impact of stereochemical inversion at the 2′-position on intrinsic gas-phase structure and the intramolecular hydrogen-bonding networks of the protonated nucleosides may provide additional insight into intrinsic nucleoside stabilization relevant to environments providing alternative modes of stabilization.

1.4.3 Dideoxyribose-Based Nucleosides

A family of synthetically modified nucleosides that has a significant presence pharmaceutically is the 2′,3′-dideoxyribose family. Reverse transcriptase inhibitor

nucleoside analogue drugs are important contributors to the antiretroviral drugs used to treat HIV and AIDS.[21] Of these nucleoside reverse transcriptase inhibitors, many do not possess hydroxy substituents at the 2′- and 3′-positions of the ribose. In fact, the most well known nucleoside reverse transcriptase inhibitor, zidovudine (ZDV, or azidothymidine, AZT) is 3′-azido-2′,3′-dideoxythymidine. However, 2′,3′-dideoxycytidine (Zalcitabine, ddC) and 2′,3′-dideoxyinosine (didanosine, ddI) are more conventional examples of the 2′,3′-dideoxyribose family of nucleoside reverse transcriptase inhibitors.[99,100] Beyond their pharmaceutical use, 2′,3′-dideoxyribose nucleoside analogues also serve an important role in the study of DNA. The Sanger sequencing method, or chain termination method, of sequencing DNA relies upon incorporation of labeled 2′,3′-dideoxynucleotides by DNA polymerase into DNA chains.[101-103] Upon incorporation of the 2′,3′-dideoxynucleotide, the absence of the 3′-hydroxy moiety prevents further chain elongation. Both applications rely upon the conformations of the 2′,3′-dideoxynucleosides being similar enough to the 2′-deoxynucleosides so as to allow their recognition by several enzymes.

Due to their pharmaceutical and research relevance, a substantial body of research into the solution phase structure of the 2′,3′-dideoxyribose nucleosides has been published.[35,36,39,40,104] Several 2′,3′-dideoxyribose nucleoside analogues have been observed by NMR to adopt N-type (C2′-*exo*/C3′-*endo*) sugar puckering [35,40,104]. This contrasts with the S-type (C3′*exo*/C2′-*endo*) sugar puckering observed for many nucleoside analogues by crystallography.[40] Prior study by gas-phase calculations of several 2′,3′-dideoxynucleoside analogues indicated that the 5′-hydroxy orientation may be more flexible than observed for those nucleosides that did not display reverse

transcriptase inhibition. These 2′,3′-dideoxynucleoside analogues indicated a higher propensity for the *gauche*$^-$ 5′-hydroxy orientation.[39] Study of the 2′,3′-dideoxynucleosides by gas-phase ion spectroscopy, and comparison to the canonical DNA and RNA nucleosides, allows for detailed examination of the impact of the 2′- and 3′-hydroxy substituents on the orientation of the 5′-hydroxy moiety and sugar puckering on protonated nucleoside structure.

1.5 Tandem Mass Spectrometry

Mass spectrometry is an extraordinarily powerful analytical tool allowing for the accurate determination of the mass-to-charge (m/z) ratio of an ion. At its most basic level, this data can be used to identify the elemental composition of an ion, and with careful experimental control its concentration in a sample. Careful experimental control and clever experimental designs also allow for detailed study of the intrinsic properties of an ion in the high vacuum, controlled environment of a mass spectrometer.[105-108]

Tandem mass spectrometry (MS^n), involving multiple stages of mass selection and mass analysis, can allow for further identifying information to be extracted about an ion. A general scheme of a tandem mass spectrometry experiment is shown in **Figure 1.7**. Activation of a precursor ion via ion-molecule reactions, collision-induced dissociation, photodissociation, or some other process is commonly employed between stages of mass analysis. Determination of the elemental composition of resulting product ions, in conjunction with data regarding the precursor ion, provides more detailed information regarding the precursor ion, generally related to its chemical structure, or stability. In a typical MS^n experiment, precise control over the activation method is not necessary, as the primary goal is to induce fragmentation. However, carefully controlling the energy

available to collision-induced dissociation, or the frequency of light used for photodissociation, can allow for more detailed information to be extracted regarding ion stability or structure.

This thesis work contributes to the development of instrumentation for IRMPD action spectroscopy and methods for the analysis of IRMPD spectroscopic data. The development of an experimental setup for infrared multiple photon dissociation (IRMPD) action spectroscopy, a frequency-dependent photodissociation tandem mass spectrometry technique, is detailed in **Chapter 2**. Software tools and scripts were also developed to facilitate several aspects of the complimentary theoretical calculations and data analysis that greatly assist in the analysis of these tandem mass spectrometry experiments and are detailed in **Chapter 3**. The use of IRMPD action spectroscopy to characterize the gas-phase structures of the protonated modified nucleosides described in **Section 1.4** is also detailed in **Chapter 4** and in **Chapter 5**.

Figure 1.1

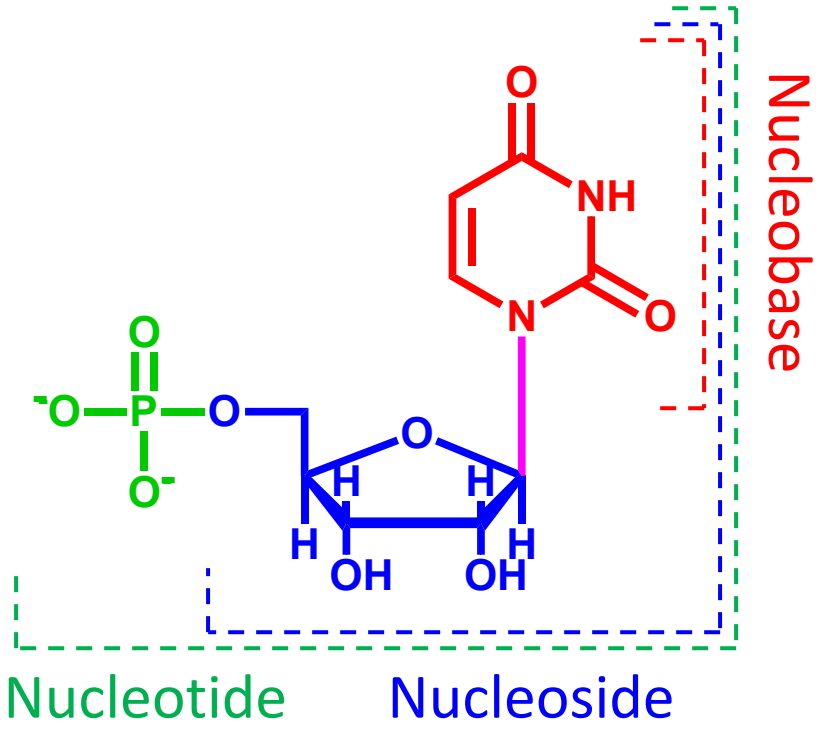


Figure 1.1 Structure of uridine-5′-monophosphate, a mononucleotide of RNA, delineated to also display the structure of uridine, the corresponding nucleoside (omitting the phosphate moiety shown in green), and the uracil nucleobase (further omitting the sugar ring shown in blue).

Figure 1.2

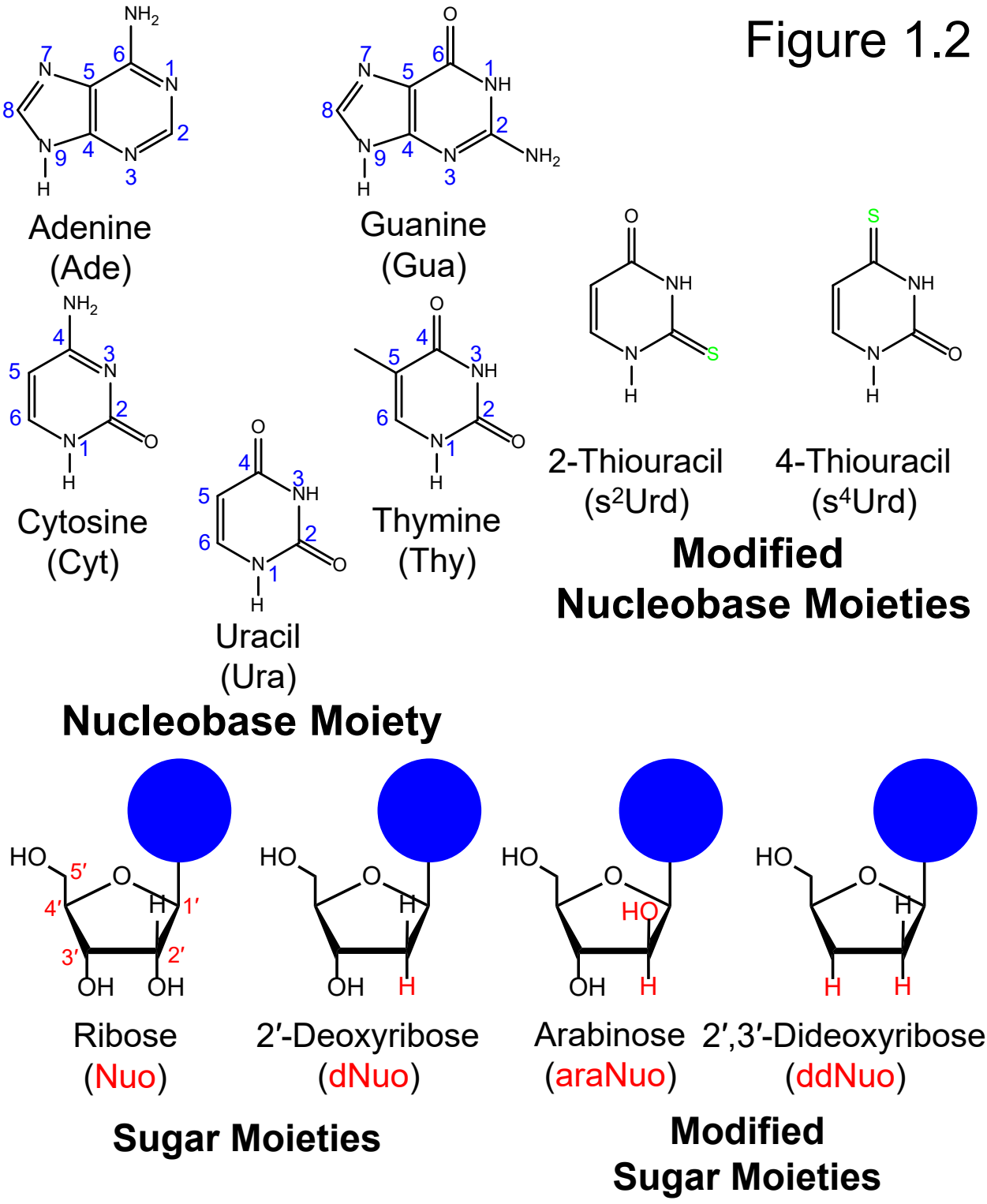


Figure 1.2 Canonical and modified nucleoside components. The nucleobase moieties examined are shown with their nucleoside three letter designations. The canonical and modified sugar moieties studied are also shown, with their designation modifiers shown in parentheses. The sugar modification is listed before the nucleoside designation (xNuo).

Figure 1.3

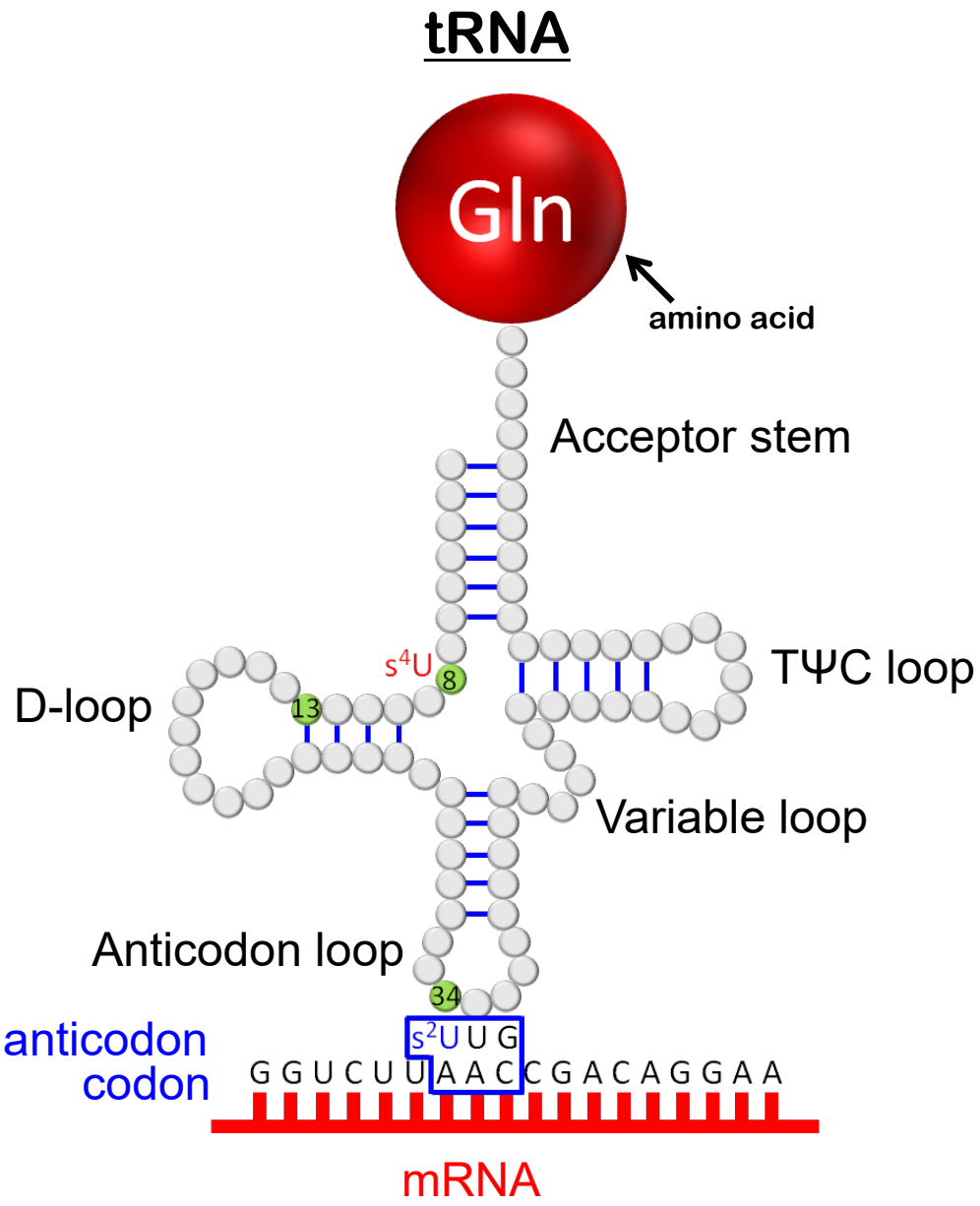


Figure 1.3 Example of a tRNA used by the ribosome for translation of an mRNA to a protein with the locations of 4-thiouridine and 2-thiouridine shown. The location(s) of the cytidine residue 4-thiouridine may crosslink with upon UV irradiation is shown. Also shown is an example of the codon:anticodon interaction region, which is important to understand the role of 2-thiouridine in the wobble position.[109]

Figure 1.4

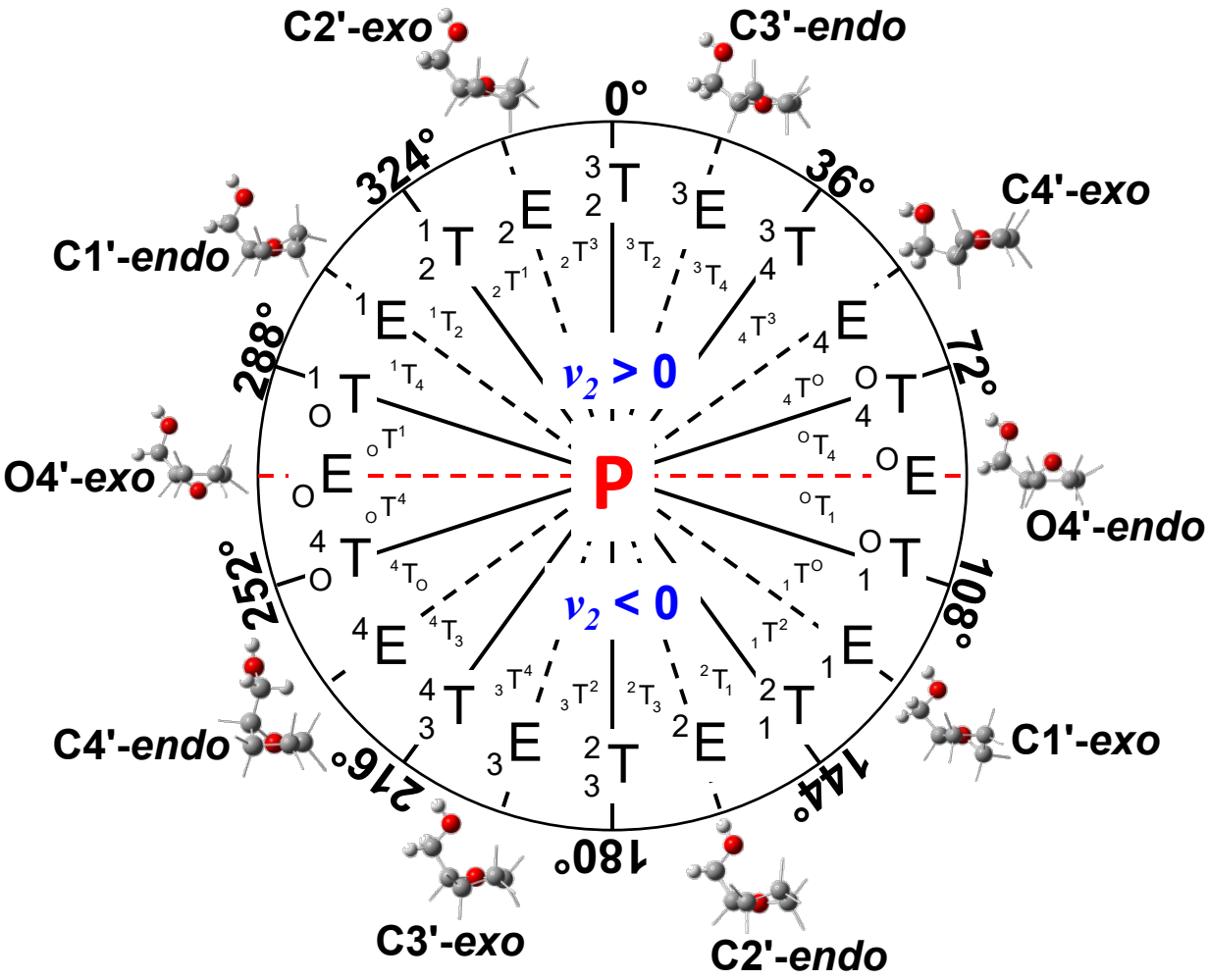
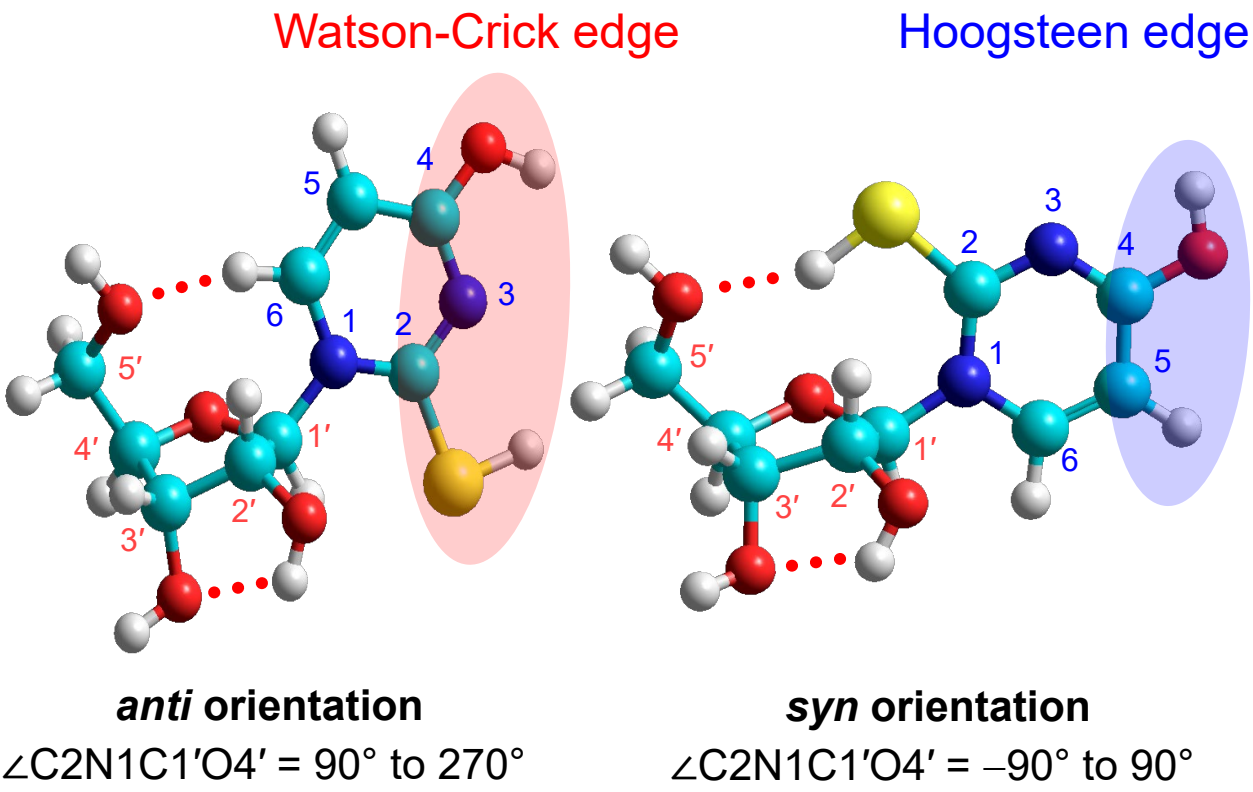


Figure 1.4 The pseudorotation wheel used to designate the specific nucleoside sugar puckering modes. Examples of the major envelope sugar-puckering modes are shown pictorially with their common *endo*/*exo* designations.

Figure 1.5



***anti* orientation**
∠C2N1C1′O4′ = 90° to 270°

***syn* orientation**
∠C2N1C1′O4′ = –90° to 90°

Figure 1.5 The two general orientations of the nucleobase about the glycosidic bond, *anti* and *syn* are shown. The Watson-Crick and Hoogsteen edges are also labeled. The glycosidic bond angles used to identify these nucleobase orientations are also detailed.

Figure 1.6

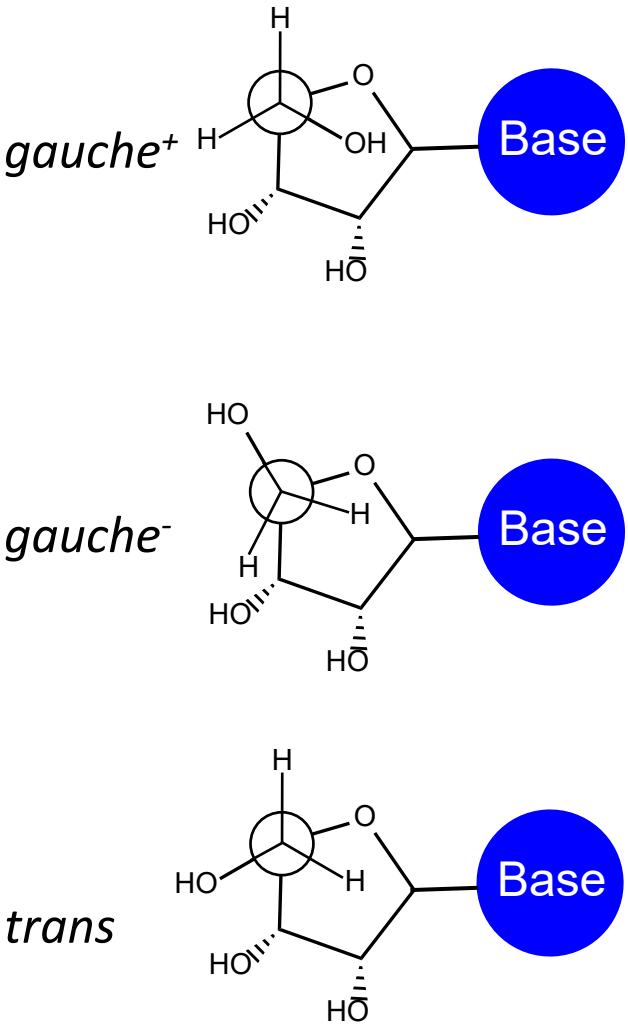


Figure 1.6 Depictions of the three general orientations of the 5′-hydroxy moiety of a nucleoside, *gauche*$^{+}$, *gauche*$^{-}$, and *trans*.

Figure 1.7

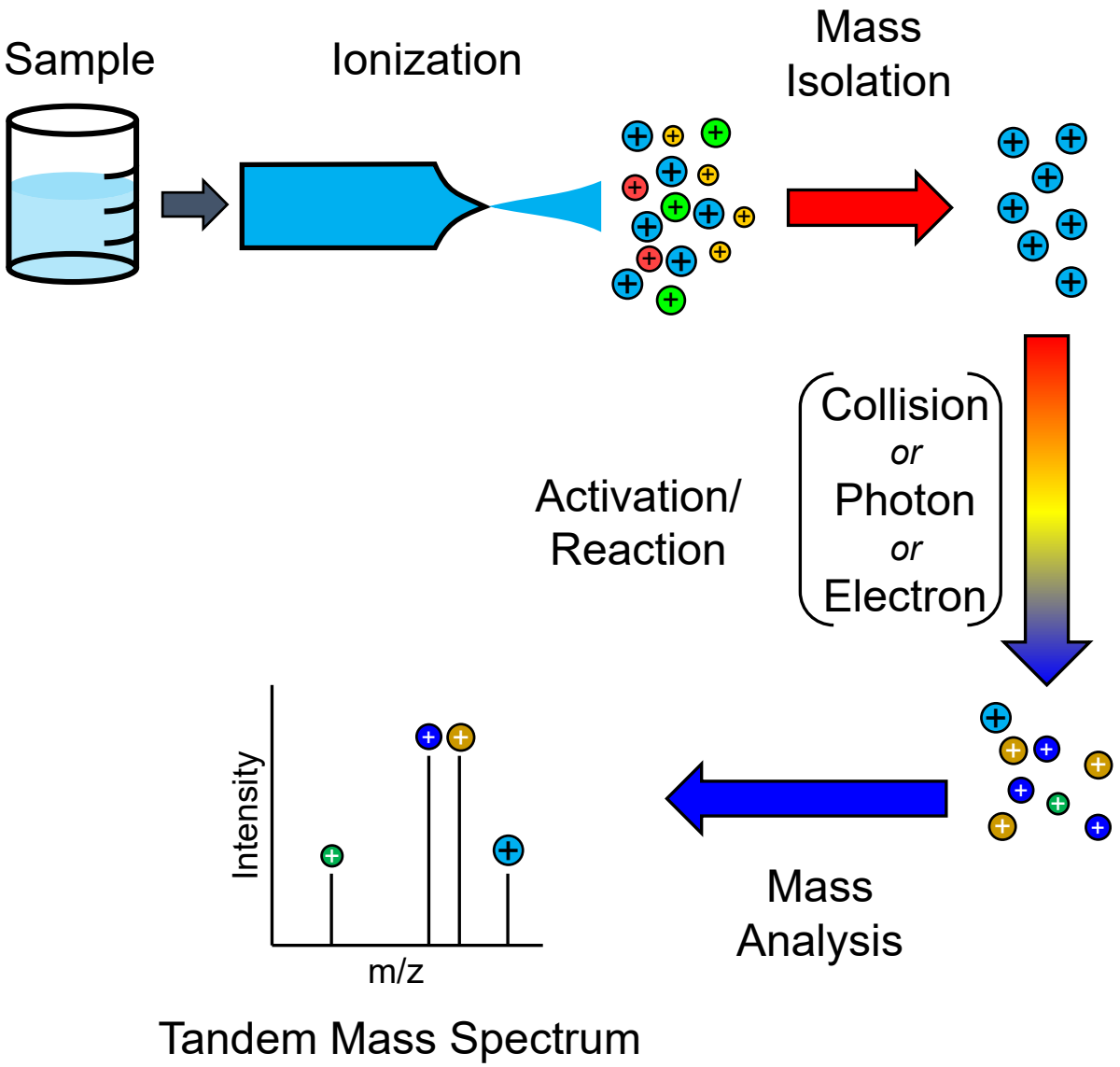


Figure 1.7 General scheme for a tandem mass spectrometry experiment. A sample is converted to gas-phase ions, an ion of interest is isolated, activated by one of several techniques, then the resulting product ions are mass analyzed.

CHAPTER 2 TOOLS FOR IRMPD ACTION SPECTROSCOPY

Portions of this chapter were reprinted with permission from reference 151: Hamlow, L. A.; Zhu, Y.; Devereaux, Zachary J.; Cunningham, N.A.; Berden, G.; Oomens, J.; Rodgers, M. T. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 2125.

2.1 Infrared Multiple Photon Dissociation (IRMPD) Action Spectroscopy

In a tandem mass spectrometry experiment, photodissociation can offer several advantages over other activation methods. Photodissociation is also a relatively unique activation method in that it requires the ion of interest to contain a suitable chromophore that is sensitive to the frequency of the exciting photons. The identity of the chromophore(s) can allow for ultraviolet (UV) photodissociation,[110] which excites ions electronically, or infrared (IR) photodissociation,[111] which excites ions vibrationally. The frequency dependence of the chromophores within an ion also allows for spectroscopic measurements to be made. The low number density of ions within a mass spectrometer prevents detection of absorption by transmission. However, detection of the products of photodissociation, fragment ions, allows for action spectroscopy experiments to be performed. The energy available for activation is also readily tunable, not only by the frequency of light used, but also by the irradiation time or laser power. The use of gas-phase ions offers compelling advantages for spectroscopy by removing matrix effects. The ability to trap ions in the mass spectrometer allows for acquisition of spectra for species that would be difficult to acquire in more traditional gas-phase spectroscopic experiments.

Infrared ion spectroscopy has become an important tool in the study of intrinsic gas-phase structures over the last thirty years.[112-115] Specifically, infrared multiple photon dissociation (IRMPD) action spectroscopy has found substantial use thanks to the efforts of several free electron laser facilities.[116-120] An experimental workflow for IRMPD action spectroscopy experiments is illustrated in **Figure 2.1**. The ions of interest are isolated and irradiated by an IR laser that vibrationally excites the ion. If the ion contains an appropriate chromophore for the frequency of the irradiating IR laser the ion may photodissociate by infrared multiple photon dissociation (IRMPD) as described below, after which a tandem mass spectrus is collected. Tandem mass spectrum of the ion of interest is collected across the IR frequency range examined. Intensities of the precursor (I_p) and fragment ions (I_f) are converted to an IRMPD yield via **Equation 2.1**, corrected for frequency-dependent variations in laser power, and plotted against IR frequency to generate an IRMPD action spectrum.

$$\text{IRMPD yield} = \sum_i I_{f_i} \Big/ \left(I_p + \sum_i I_{f_i} \right) \tag{2.1}$$

2.1.1 IRMPD Mechanism

The first requirement for IRMPD is a resonant vibrational mode (a chromophore), which allows the ion of interest to absorb energy from the incident IR photon into the resonant vibrational mode. Ideally this now energized vibrational mode will lead to dissociation of the relevant bond. However, a single IR photon typically does not deposit enough energy into the resonant vibrational mode to exceed the dissociation threshold of a room temperature ion. Unfortunately, due to the anharmonicity present in real vibrational modes, there is a limit to the number of photons that can be absorbed by the resonant vibrational mode before it is no longer resonant. Luckily, intramolecular

vibrational redistribution (IVR)[121] occurs more rapidly than the typical IR laser systems used for IRMPD can pump energy into the ion. As the name suggests, IVR redistributes the energy absorbed by the resonant vibrational mode into the bath of background vibrational modes available in the ion,[122] returning that vibrational mode to a state that is again resonant and capable of absorbing another photon. This process is typically repeated tens to thousands of times for each ion, slowly increasing the internal energy of the ion until the dissociation threshold is reached or exceeded.[123] This slow heating typically results in dissociation along the lowest energy pathway. Different levels of anharmonicity present in the resonant vibrational modes, compounded with the multiple photon nature of the technique, can result in changes in the measured IRMPD yield and even shifts the resonant frequencies when compared with the linear IR spectra predicted by theoretical calculations based on the harmonic approximation. Despite this, comparison of an experimental IRMPD action spectrum and predicted IR spectra has proven to be a robust method for characterization of gas-phase ionic species.[112,113,115,118,124-131]

2.1.2 IR Fingerprint Region Experimental Method

The modified nucleosides 2-thiouridine, 4-thiouridine, and 2′,3′-dideoxythymidine were purchased from Carbosynth (San Diego, CA, USA). 2′,3′-Dideoxyadenosine, 2′,3′-dideoxyguanosine, 2′,3′-dideoxycytidine, and 2′,3′-dideoxyuridine were purchased from Chem Impex (Wood Dale, IL, USA). Adenine arabinoside and guanine arabinoside were purchased from Metkinen (Kuopio, Finland). Cytosine arabinoside and uracil arabinoside were purchased from Sigma Aldrich (Zwijndrecht, The Netherlands and St. Louis, MO, USA).

IRMPD action spectra in the IR fingerprint region were measured using a custom-built 4.7 T Fourier-transform ion cyclotron resonance mass spectrometer (FT-ICR MS) at the FELIX facility in Nijmegen, The Netherlands, coupled to the FELIX free electron laser (FEL, 10 Hz repetition rate, bandwidth 0.3% of the central frequency, energy up to 70 mJ/pulse).[105] A schematic of this experimental setup is shown in **Figure 2.2**. Solutions of the ions of interest were diluted to ~1 mM in 50%:50% methanol:water and acidified by ~1% acetic acid. However, due to the low basicity of uridine and its analogues, between 1 and 10% HCl was used instead of acetic acid to generate adequate ion intensities for the uridine analogues. The solutions were sprayed by a "Z-spray" electrospray ionization (ESI) source at a flow rate of ~10 μL/min, with assistance from N_2 nebulizing and heating gas flows to generate the protonated ions of interest. Ions enter the first vacuum stage orthogonally through a skimmer cone and are then orthogonally sampled into a hexapole ion guide. Ions are accumulated by the hexapole ion guide before injection into the ICR cell through a quadrupole-bender and octopole ion guide. Stored waveform inverse Fourier transform (SWIFT) techniques are used to isolate the ions of interest in the FT-ICR cell where they are subsequently irradiated by the FEL for several seconds before mass analysis and detection in the FT-ICR cell. Tandem mass spectra are acquired at 4-5 cm^{-1} steps over the frequency range of ~600–1900 cm^{-1}. Intensities of the resultant fragment ions and remaining precursor ions are extracted from these tandem mass spectra and shown in **Figure 2.1** as a precursor depletion spectrum in blue and fragment appearance spectrum in red. The IRMPD yield is calculated by **Equation 2.1** and corrected for the observed frequencies and frequency-dependent power across the range measured and is plotted in grey in **Figure 2.1**. The IRMPD action spectrum in the

hydrogen-stretching region of protonated adenine arabinoside was also measured using this 4.7 T FT-ICR MS, coupled to a table-top OPO laser system (LaserVision, Bellevue, WA, 10 Hz repetition rate, bandwidth 3 cm^{-1}, energy up to 15 mJ/pulse). The output power of this OPO system was sufficiently constant across the 3300–3800 cm^{-1} frequency range studied as to not require correction for frequency dependent variation in the laser power.

2.2 Optical Parametric Oscillator/Amplifier Laser System

The LaserVision (Bellevue, WA, USA) OPO laser system is commonly used in infrared ion spectroscopy instrumental set ups, such as for the IRMPD spectroscopy experiments in the hydrogen-stretching region performed at the FELIX facility and the Rodgers laboratory and described in **Section 2.1.2**. The LaserVision OPO offers enough power for IRMPD with the appropriate pump laser, a broadly tunable frequency range, and a reasonably narrow bandwidth. An extension into the IR fingerprint region is also available for the system, although the power drop in the IR fingerprint region is large enough to prevent IRMPD from being as routinely effective. External control of the output frequency is straightforward through a series of LabVIEW™ subVIs provided with the system. These subVI communicate over TCP/IP with a dedicated control unit provided with the LaserVision OPO system.

The OPO system provided by LaserVision consists of an optical parametric oscillator and optical parametric amplifier (OPA), shown schematically in **Figure 2.3**, and is designed to be pumped by an external Nd:YAG laser at the fundamental frequency of 1064 nm. A Continuum Surelite III EX Nd:YAG laser (San Jose, CA, USA) is installed in the system described here and can deliver up to ~7 W of power at 1064 nm at a 10 Hz repetition rate of 7-9 ns pulses. The beam is steered into the entrance of the OPO system

by two 1” Nd:YAG laser line mirrors (Thorlabs, Newton, NJ, USA). As shown in **Figure 2.3** the beam first passes through an iris diaphragm and telescope to resize it, before a 2:1 beam splitter directs 33% of the power towards the optical parametric oscillator and 66% towards the optical parametric amplifier. Rotating half-wave plates intersect both paths allowing for independent control over laser power in each pathway. On the oscillator beam path, the Nd:YAG fundamental is frequency doubled to 532 nm and any remaining 1064 nm light is filtered out. The oscillator stage itself is comprised of two counter-rotating potassium titanyl phosphate (KTP) crystals and results in three outputs: the 532 nm pump, 710–880 nm (near IR) and 2.1–1.35 μm (intermediate IR). The 532 nm and near IR components are filtered out and the intermediate IR is sent to the OPA stage. In the amplifier stage the intermediate IR interacts with the remaining 66% of the original 1064 nm pump beam, and is amplified, with a mid IR component (2.2–5.0 μm) generated by difference frequency generation in two stages by two pairs of counter-rotating potassium titanyl arsenate (KTA) crystals. The residual 1064 nm pump is removed to a beam dump and a stack-of-plates Si polarizer is used to select either the intermediate IR or mid IR components at the exit of the enclosure. The mid IR can be scanned from 1970–4690 cm^{-1} with a bandwidth of ~3–4 cm^{-1} and 50–170 mW of power. The power of the IR beam can be further tuned by adjusting the power of the Nd:YAG pump or adjusting the rotating half-wave plates at any given output frequency.

2.2.1 Beam Paths and Enclosures

The LaserVision OPO system relies upon a Class IV Nd:YAG laser, the Continuum Surelite III EX, of which the fundamental 1064 nm output is invisible and extremely dangerous without the appropriate safety gear. The IR output of the OPO is also fully

capable of burning items in the beam path under sustained irradiation. To maintain a practical level of safety in a laboratory designed for mass spectrometers, it was important that the laser system be housed in an appropriate enclosure. In addition to the safety concerns, the mid IR beam used for IRMPD experiments lies within the hydrogen-stretching region, which is particularly prone to absorption by water vapor in the air, and as such would ideally be transmitted in a relatively inert atmosphere to preserve laser power. To meet reasonable safety requirements for the laboratory environment, enclose the beam path for atmospheric purging, and maintain access to the OPO system a series of enclosures shown in **Figure 2.4** were constructed.

The enclosure for the Nd:YAG pump laser and OPO system is based upon a 4'x6' 2.5" thick optical breadboard from Newport with ¼"-20 tapped holes on a 1" grid across the top surface for mounting of optical components. The enclosure was constructed from T-slot extruded aluminum (Extruded Aluminum Fabricators, Ferndale, MI, USA) to fit snugly to the outside of the breadboard and result in the output of the OPO at roughly the height of the beam inlet for the FT-ICR MS, the height of which is fixed. Inexpensive Sintra® panels are mounted within the T-slots of the extruded aluminum frame, with rubber gaskets to provide sealing for region B, shown in **Figure 2.4**. A laser curtain and removable black aluminum sheet provide access to the OPO and Nd:YAG head, respectively, in region A of **Figure 2.4**. In areas with higher risk for accidental ablation, such as beyond the Nd:YAG output, a small panel of black Trespa® was added to protect the relatively fragile Sintra® panels. The internal wall separating regions A and B in this enclosure allows for immediate introduction of the mid IR output to a relatively inert environment of dry N_2 gas and drastically reduces the size of the enclosure that needs to

be sealed. The optics in region B are accessible via double doors at the end of the enclosure, constructed from T-slot aluminum extrusions, Sintra® panels, and rubber seals.

An optical shutter (SH05, Thorlabs) is placed at the entrance of region B and is commanded by the currently active mass spectrometer to only open during a defined activation window. Initial alignment of the OPO with the two mass spectrometers is done with the 532 nm pump of the OPO stage after removing the 532 nm filter and turning off the 1064 pump power to the OPA. This beam is an intense green and reasonably well aligned with mid IR output, providing good visibility and straightforward alignment. More precise alignment is done by blocking this intense 532 nm light and using the small amount of 532 nm light produced by interaction of the 1064 nm pump in the OPA stage, which is much better aligned with the mid IR output. This beam is much weaker and typically requires a relatively dark room for visibility. Unfortunately, neither of these visible 532 nm beams share the same beam divergence as the mid IR output, and are therefore not useful measures of focus, which is instead estimated by a mid IR liquid crystal card or sheet.

As the two mass spectrometers connected to this laser system are located on either side of the large enclosure, two smaller enclosures (C and D in **Figure 2.4**), based upon 30”x30” aluminum optical breadboards (MB30, Thorlabs) and constructed from T-slot aluminum extrusions, were designed. These enclosures contain the focusing optics appropriate for the corresponding mass spectrometer and are just large enough to also house Synrad Firestar v40 CO_2 lasers. These CO_2 lasers produce a fixed 10.6 μm wavelength infrared beam with up to 40 W of power allowing for access to most organic

molecules for general purpose IRMPD, but not action spectroscopy. As the CO_2 lasers pose a greater hazard to the enclosure panels, with much greater power, focus, and a more appropriate wavelength, each of these smaller enclosures use Trespa® side and top panels. Both mid-IR beam paths use 2” protected gold optical mirrors due to the relatively large beam size. Flat mirrors from Thorlabs are used to steer the beam path, with an appropriate focusing mirror towards the end of each beam path, circled in green on **Figure 2.4**, to produce a suitable beam width at the respective ion clouds.

2.3 Bruker amaZon ETD Quadrupole Ion Trap Mass Spectrometer (QIT-MS)

The 3D quadrupole ion trap offers several advantages as a mass spectrometry platform for IRMPD. The primary advantage of the QIT-MS for IRMPD is the trapping of the ion cloud for extended periods of time and fragmentation events occurring directly within this trapping volume. A trapped ion cloud allows for reasonably long irradiation times of the ion cloud, such that photodissociation can be observed even for relatively weak chromophores. The QIT-MS also benefits from relatively high pressures (~1 mTorr) of an inert gas, typically He, in the trapping volume. The presence of the inert gas improves trapping efficiency by collisional cooling of the ions thereby reducing the kinetic energy they acquire from the RF field responsible for trapping them, but also helps focus the ion cloud within the trapping volume. The smaller ion cloud then allows for excellent overlap with an incident laser to be achieved, which can result in shorter irradiation times, and therefore shorter experiments or the opportunity for more signal averaging in a given period of time. Unfortunately, interactions with the buffer gas in the ion trap cool not only the kinetic energy, but also the internal energy, which works against IRMPD activation.

The Bruker amaZon ETD QIT-MS used in this work generates ions with an Apollo II ESI source, assisted by nebulizing N_2 gas, and a countercurrent of heated N_2 as a drying gas with flow rates of ~3.0 L/min. Ions are transferred into the mass spectrometer by a metallized glass capillary, the entrance of which is biased to ~−4000 V, with the ESI emitter held at ground. The exit of the metallized glass capillary is biased to ~100−150 V for injection into the first of two RF ion funnels. Ions leaving the second ion funnel enter a differential pumping stage and a hexapole ion guide that transmits them into the ion trap where they are accumulated. The ion trap consists of two end-cap electrodes and a ring electrode. The primary RF amplitude is applied to the ring electrode to trap the ions, and an auxiliary RF amplitude can be applied to the second end-cap electrode to excite trapped ions. The scan sequence of the QIT-MS is shown in **Figure 2.5**. To begin a scan the primary RF voltage is turned off to clear the ion trap. The primary voltage is then turned back on and the gate lens is opened to allow ions to accumulate in the trap. An isolation delay can be configured with only the primary RF active to allow for ions to cool through interactions with the buffer gas. The primary RF and auxiliary RF are then used to isolate the ion of interest, followed by a delay with only the primary RF active to allow the ions to cool again before activation. Shown in **Figure 2.5** is the use of the auxiliary RF during the activation window for CID, however, this segment can also be used for other activation methods including electron transfer dissociation (ETD) or IRMPD in the present set up. A further delay can be added after activation, before the ions are finally scanned out of the ion trap by the primary and auxiliary RF for mass analysis.

2.3.1 Instrument Modifications

Physical modification of the Bruker amaZon ETD platform for acceptance of an incident laser was quite straightforward. A vacuum flange with two through holes for mounting of two optical windows, and M6 blind tapped holes on a 1” grid was purchased from Bruker. Either ZnSe or BaF_2 windows can be used to provide access for the mid IR output of the OPO, but 1” ZnSe windows from Thorlabs are used here to also allow for transmission of the 10.6 μm output from a CO_2 laser. A second ion trap assembly was also purchased from Bruker with a modified ring electrode with two holes machined through the center of the electrode and aligned vertically within the ion trap assembly. The optical windows on the vacuum flange are aligned with the ion trap assembly such that one of the windows is aligned with the additional holes through the ion trap. A custom mirror mount was also purchased from Bruker. This mirror mount locates two ½” protected gold mirrors (Thorlabs) beneath the ion trap assembly in the vacuum chamber to redirect the laser path out of the second window to a beam dump. This assembly is shown overlaid on a cutaway view of the QIT-MS in **Figure 2.6** with the laser path.

To mount the optical windows to the modified vacuum flange a custom holder was machined from Delrin™ to mate with the vacuum flange. This holder maintains alignment of the windows and applies gentle pressure to the two Viton® CF gaskets (for 1.33” CF flanges) that sandwich the optical window. The vacuum in the instrument creates an adequate seal with the bottom gasket and window to allow the instrument to operate at ~3.75×10^{-6} Torr, nearly the same pressures as the trap assembly without windows or through holes.

2.3.2 Hardware Synchronization

Synchronization of timing between firing of the Nd:YAG pump laser, scanning of the mass spectrometer, and the optical shutter is important to ensure that only the ions of interest are irradiated. Synchronization of the optical shutter with the QIT-MS activation segment, shown in **Figure 2.5**, is accomplished via TTL signal from the auxiliary connector of the QIT-MS. This TTL signal is set to be active-high during a chosen activation segment and can be configured via the QIT-MS web interface or via the XML interface discussed in greater detail in the following section, **Section 2.3.3**. The TTL signal is optically isolated via the circuit shown in **Figure 2.7** to ensure it has enough power to trigger the optical shutter and allow the mid IR beam into the ion trap.

Synchronization between the scan sequence of the QIT-MS and firing of the Nd:YAG is also accomplished via TTL signal. An active-low TTL signal associated with the flashlamp discharge of the Nd:YAG laser, offset by approximately 100 ns, is converted to active-high via the circuit depicted in **Figure 2.8**. This signal is then passed into the auxiliary interface of the QIT-MS. The QIT-MS can be configured to wait between individual scans such that receiving the active-high TTL signal from the Nd:YAG starts the next scan sequence from segment 1 in **Figure 2.5**. This timing synchronization ensures that the same number of laser pulses are present within the activation window of each tandem mass scan acquired. Both TTL synchronization signals are monitored on a digital oscilloscope to verify the length of the activation segment and the number of laser pulses per segment.

2.3.3 Data Acquisition Software

During an IRMPD experiment, control of the QIT-MS is handled by the Bruker Trap Control software in a nearly identically manner to a typical CID experiment. The only change in the configuration of Trap Control during an IRMPD experiment versus a standard CID experiment is that the auxiliary RF activation energy is kept at 0.00 V for the activation segment used for IRMPD. Without RF activation the trapped ions remain in the center of the trap, thereby enhancing overlap with the laser. A data point at a given IR irradiation frequency is typically comprised of several averaged tandem mass scans. During a typical IRMPD experiment 10-15 scans are averaged. The number of scans averaged can be readily adjusted via the Trap Control interface. During a single tandem mass scan several activation segments, up to 10, can be utilized with mass isolation steps between each step and IRMPD can be activated for any of the steps via the web interface. Activating IRMPD for more than one segment is generally not useful as the OPO cannot be scanned meaningfully between activation segments but could be useful for two laser experiments. The use of multiple segments allows multiple activation methods to be used during a single experiment such that the IRMPD action spectrum of product ions of CID or electron transfer dissociation (ETD) could be measured.

Custom LabVIEW™ software was developed to manage synchronization between the OPO and QIT-MS and control scanning of the OPO output frequency. This software communicates with Trap Control via the XML interface accessible with the Trap Control service mode, and with the LaserVision OPO via LabVIEW™ drivers supplied with the system. Custom LabVIEW™ subVIs were acquired from Bruker (courtesy of Christoph Gebhardt) and are used to create and save custom XML files to a local directory

monitored by Trap Control, containing commands for the configuration of the QIT-MS. These XML commands are not limited to those options available on the Trap Control interface. Control over the activation window length and He pressure, for instance, are more customizable via the XML interface than by the Trap Control interface. This XML interface is also used to configure the QIT-MS to both generate the TTL output during a selected activation window, as well as accept the TTL input from the Nd:YAG flashlamp discharge, allowing for hardware synchronization between the QIT-MS and OPO to be established. The XML interface also configures Trap Control to pause at the end of each averaged scan and wait for a specific XML command before beginning acquisition of the next averaged scan. This software synchronization mode allows the OPO to scan to the next output frequency before triggering acquisition of the next averaged scan. However, this mode is somewhat problematic as it can make Trap Control appear unresponsive if it does not receive the correct XML trigger to acquire the next set of scans. Therefore, a timeout is typically set for this mode. In a normal IRMPD experiment this timeout is set to ~3 min, which is sufficiently long that there is little risk of the timeout removing the software synchronization in the middle of acquiring a spectrum. This timeout is also long enough that the OPO has enough time to scan to the starting frequency after software synchronization and the OPO scan are started.

Hardware synchronization with the Nd:YAG and optical shutter is typically established well before an experiment begins as it has little impact on the operation of the QIT-MS. Triggering of the optical shutter by the QIT-MS is also required for photodissociation when tuning irradiation time and/or laser power to prevent irradiation of the ion cloud before the ion of interest is isolated. In a future version of Trap Control, a

separate TTL signal might be sent from the QIT-MS upon collection of a data point and passed into the software to trigger the OPO to move to the next frequency. However, with this feature unavailable in the current version of Trap Control a timed loop is instead used to trigger scanning of the OPO. Therefore, before software synchronization is established the acquisition of a full averaged data point under the full experimental conditions is timed for input to the OPO scan. After acquisition of the averaged tandem mass scans at each IR frequency across the region of interest, the resulting data file is processed in the Bruker Data Analysis software by a custom Visual Basic script modified slightly from that of Martens et. al.[128] This script extracts the precursor and fragment ion intensities and the OPO frequency at each data point, and sends the OPO frequency to Trap Control with the XML trigger. This data is then written to a comma-delimited file for any further analysis or processing in separate graphing software.

2.3.4 QIT-MS IRMPD Performance

Several operational parameters for IRMPD in the QIT-MS are the neutral gas pressure, laser power, laser irradiation time, and laser overlap with the ion cloud. The He gas pressure within the trap is regulated through the primary gas controller (GC_{He}), which can be used to provide 0 to 100% gas flow through a proportional valve. Typically, GC_{He} values of ~20% correspond to an increase in background pressure of 0.75×10^{-6} Torr vs. the background pressure of ~3.00×10^{-6} Torr with no He introduced to the trap. To examine the effects of this pressure on IRMPD yield, spectra of protonated tryptophan (a well characterized system)[128,132] were collected while varying GC_{He} from 5 to 80% (**Figure 2.9**). Reduced collisional cooling at low pressures resulted in higher dissociation yields, but reduced trapping efficiency, and led to greater noise. As pressure in the trap

increased, the IRMPD yield decreased in concert with the noise, though not linearly across the pressure range. A pressure-dependent balance of IRMPD yield and noise was found for GC_{He} values of ~25% for this QIT MS instrument. Ion intensity also plays an important role in determining the signal-to-noise ratio of the IRMPD yield and can be improved by increasing GC_{He} up to a limit after which ion intensities may decrease. It is important to note that the very small feature observed in **Figure 2.9** around 3300 cm^{-1} was still observable until around ~60% GC_{He} and decreases very little from ~20% GC_{He}. Below ~20% GC_{He} the size of this small feature is comparable to the noise making it difficult to distinguish. Likewise, at GC_{He} > 60%, the feature observed at ~3300 cm^{-1} is small enough to be indistinguishable from noise. Conveniently, the calibrated GC_{He} value under the 'maximum resolution' scanning mode built-into Trap Control is typically ~25%. Therefore, manual changes to GC_{He} via the XML or service interface are generally limited to systems that are either difficult to fragment or are found to undergo fragmentation without irradiation. For systems such as protonated tryptophan that undergo very facile dissociation, where there may be notable fragmentation of the reactant ion prior to irradiation, increasing the He pressure beyond ~20% was helpful to increase collisional cooling. For systems where IRMPD is less efficient, slightly decreasing the He pressure in the trap was found to enhance the IRMPD yield, but at a cost of increased noise. While He pressure in the trap certainly affects the resultant IRMPD action spectrum, there appears to be a reasonably wide range of pressures over which its effects are subtle. Small features remained distinguishable as pressure was increased until approximately 60% GC_{He}, whereas noise is reduced to a greater degree until about 50% GC_{He}.

Figure 2.10 displays a series of IRMPD spectra of protonated 2-thiouridine, $[s^2Urd+H]^+$, in the hydrogen-stretching region collected at different laser powers and irradiation times that demonstrate the importance of tuning the energy available for dissociation. Operating at full laser power resulted in saturation of the two most intense features and significant broadening of all features. Decreasing the irradiation time and laser power until the most intense peak has ~50% IRMPD yield results in even more substantial peak narrowing, increases the signal-to-noise ratio, and reduces the chance of fragmenting the precursor ion to its limit resulting in more reliable relative intensities. In the current work, reduction of the OPO power from 127 mW to 60 mW at 3500 cm^{-1} enabled acquisition of spectra exhibiting much narrower features, while also reducing peak saturation and retaining the less intense features observed at higher power.

Laser power and irradiation time were found to have a greater effect on the IRMPD spectrum acquired than He pressure, while also being more straightforward to tune. The available power from the OPO used here was more than sufficient for the collection of spectra over the hydrogen-stretching region of the protonated nucleosides examined in this work. Use of the maximum OPO power available can result in poorer representation of relative intensities for the major features and subsequent peak broadening. Irradiation time is more limited in adjustment and does not appear to have as substantial an impact as laser power. As observed for He pressure, both major and minor spectral features are often distinguishable across a wide range of power and irradiation times. This flexibility allows an IRMPD spectrum to be acquired that displays features that are close to the dissociation threshold alongside those that are well above it. Collection of a spectrum under multiple conditions affords the opportunity to either splice spectra together after

correcting for the relative energy available for dissociation or create overlays to highlight a feature that may otherwise be difficult to distinguish. Choosing appropriate conditions is also critical to feature width, which can have a large impact on the ease of spectral interpretation.

2.3.5 IR Hydrogen-Stretching Region Experimental Method

The modified Bruker amaZon ETD QIT-MS described here was used to measure the IRMPD action spectra in the hydrogen-stretching region for all of the protonated nucleoside analogues discussed in **Section 1.4** except for protonated adenosine arabinoside, the acquisition of which was discussed in **Section 2.1.2**. The gas-phase protonated nucleoside analogues examined here were generated by ESI from ~20 μM solutions of the nucleoside analogue of interest in 50%:50% methanol:water that was acidified by ~1% acetic acid, and introduced to the ESI source at a flow rate of 3.0 μL/min. The ESI emitter is held at ground potential versus a counter electrode held at −4.5 kV and assisted by a nebulizing N_2 gas flow at 3.0 L/min. Large droplets and neutrals are rejected by a counter-current of heated N_2 gas flowing around the inlet capillary, which also assists in desolvation. Ions are transferred into the mass spectrometer by this metallized-glass inlet capillary and injected into the first of two RF ion funnels. Following the second ion funnel ions are transferred into and accumulated in the 3D quadrupole ion trap. Within the ion trap the ions of interest are mass isolated before irradiation by the LaserVision OPO. The isolated protonated nucleoside analogues were irradiated for 0.1 to 2 seconds to produce ~50% dissociation of the precursor ion at the most intense spectral feature. The IRMPD action spectra were collected from 3300−3800 cm^{-1} at increments of 4 cm^{-1}. The frequency dependent power of this OPO system was sufficiently linear across the

frequency range measured such that application of a frequency-dependent power correction was unnecessary.

2.4 Bruker SolariX Fourier-Transform Ion Cyclotron Resonance Mass Spectrometer (FT-ICR MS)

The Fourier-transform ion cyclotron resonance mass spectrometer (FT-ICR MS) is also a particularly capable platform for IRMPD action spectroscopy. The greatest advantage of the FT-ICR MS for IRMPD action spectroscopy is the ultra-high vacuum present in the trapping region, which for the Bruker solariX FT-ICR MS discussed here is ~10^{-10} Torr. The intense magnetic field within the ICR cell also allows for extremely long trapping times. The ultra-high vacuum in the ICR cell renders counter-productive collisional cooling during an IRMPD experiment largely irrelevant, unlike that observed in the QIT-MS. Longer trapping times are also important as the overlap between the incident laser and the ion cloud is more difficult to optimize given the much larger ion cloud in comparison to the QIT-MS. A potential drawback of the FT-ICR MS platform, particularly with respect to the QIT-MS, is the inherently greater complexity in both instrumentation and software control that can make IRMPD experiments more challenging to technically implement and perform.

Ions are generated in the solariX FT-ICR MS by an Apollo II ESI source assisted by nebulizing and counter-current N2 gas with flow rates of ~3.0 L/min each. Ions are transferred into the mass spectrometer by a metallized glass capillary the entrance of which is biased at ~−4000 V vs the ESI emitter. Ions are injected orthogonally by the glass capillary, the exit of which is biased to ~−150 V, into the first of two ions funnels that collect and transmit the ions to an RF octopole ion guide leading to a quadrupole mass

filter. An initial stage of mass analysis can be performed in the quadrupole mass filter, before the ions are injected into the collision cell where CID or ETD can occur. Ions are then transferred into the ICR cell by a hexapole ion guide. Within the ICR cell several stages of mass isolation and fragmentation can occur before the ions are mass analyzed and detected within the ICR cell.

2.4.1 Instrument Modifications

The geometry of the FT-ICR MS allows for a straightforward incident laser path within the ICR cell. As seen in **Figure 2.11**, the ICR cell resides at the end of the ion path, serving as both mass analyzer and detector, which leaves the entrance to the ICR cell readily accessible. Typically, this entrance to the ICR cell, opposite the ion entrance, is used for an electron capture dissociation (ECD) cathode, and on the Bruker solariX 7T discussed here this ECD cathode assembly facilitates laser access via a through hole. This is likely a left over from the design for general purpose IRMPD using a CO_2 laser that was available with the previous version of the Bruker FT-ICR MS platform. This ECD assembly allows for introduction of the OPO laser to the instrument through a 1.33” CF BaF_2 viewport mounted at the end of the ICR cell vacuum chamber, providing better transmission of the OPO than the ZnSe windows used by the QIT-MS, but poorer transmission of the CO_2 laser. A 40” focal length, 2” protected gold mirror (Edmund Optics, Barrington, NJ, USA) focuses the OPO such that the focal point is approximately 8-10” in front of the viewport, resulting in a beam narrow enough to pass through the ECD cathode without risk of damage to the viewport due to high power density. Alignment of the OPO with the center of the ICR cell is accomplished using two transparent targets attached at either end of the magnet bore with the ICR cell vacuum chamber removed

from the magnet, allowing for alignment of the laser precisely through the center of the magnet.

To isolate the beam path between the small enclosure described in **Section 2.2.1** and the viewport on the FT-ICR MS a telescoping tube was fabricated from 3D printed seals, bearings, and schedule 40 black polypropylene pipe. The telescoping motion allows the tube to be installed between the enclosure and bore of the magnet easily while also providing a reasonably tight fit to the viewport, isolating the beam from atmospheric water vapor. The end of the tube fits around the CF viewport, which provides some alignment of the tube to the FT-ICR MS, but a flange was also machined from a polypropylene sheet to fit snugly in the end of the magnet bore to secure the tube in the center of the magnet and helping to protect the magnet bore from debris.

2.4.2 Hardware Synchronization

During an IRMPD action spectroscopy experiment the Bruker solariX FT-ICR MS is operated in a nearly identical manner as typical for mass analysis acquisition. Behind the Bruker solariXcontrol software, the event sequence of the FT-ICR MS is controlled via pulse programming through BASIC (Bruker Acquisition Sequence for Instrument Control). Text files, typically labeled ‘BASIC’, describe the full sequence of events during a mass analysis via event keywords. Each of these keywords is associated with an event definition, many of which are accessible through an associated ‘IPSO.properties’ file.

The pulse programming tutorial for the solariX FT-ICR MS provides an example for the configuration of a custom user event for triggering an external laser via TTL signal from a specific port on the auxiliary interface on the FT-ICR MS. A copy of the default ‘IPSO.properties’ file was created and named ‘IPSO_IRMPD_ON.properties’. Following

the pulse programming tutorial, the empty ‘USER_EVENT_1’ was configured as shown in **Figure 2.12**, leverages an unused delay variable ‘d25’ to determine the length of this trigger, and is conveniently available via solariXcontrol, allowing for the irradiation time to be readily adjusted during the setup phase of an experiment. The commands ‘setnmr4|9’ and ‘setnmr4^9’ turn on and off the TTL signal at the pin corresponding to ‘setnmr4(9)’, with the delay specified by ‘d25’ inserted between them. This signal is optically isolated by the circuit shown in **Figure 2.13** Passing a TTL signal into the FT-ICR MS is less clear and not well addressed in the pulse programming tutorial. Fortunately, assistance from Steve VanOrden of Bruker facilitated the pulse programming of ‘USER_EVENT_2’, also shown in **Figure 2.12**. The ‘trigpe1’ command will wait for the positive edge of a signal on the corresponding pin, with no timeout, effectively pausing the scan indefinitely until the Nd:YAG signal is received.

A copy of the standard BASIC file (‘BASIC_IRMPD_ON’) was made for injection of these custom events to prevent tampering with the standard operation of the FT-ICR MS. ‘USER_EVENT_2’, is inserted before the ICR cell is quenched following the previous scan, a snippet of which is shown in **Figure 2.14**. ‘USER_EVENT_1’, is inserted into the BASIC event sequence immediately after the first automatically generated mass isolation/fragmentation event, shown in **Figure 2.15**. If the IRMPD action spectra of a CID or ECD fragment ion is desired, the custom BASIC file must be modified and selected via solariXcontrol to change this sequence. The selection of this modified BASIC file and modification of the ‘master.properties’ file to point at the modified ‘IPSO_IRMPD_ON.properties’ event definition file is required to configure the FT-ICR MS for acquisition of an IRMPD action spectrum.

2.4.3 Data Acquisition Software

Unfortunately, solariXcontrol does not have a readily accessible software interface for external commands. Therefore, a custom LabVIEW™ program running on the FT-ICR MS workstation configures a simple timed scan of the OPO across the selected frequency range at a speed calculated to step approximately one bandwidth of the laser (~4 cm^{-1}) in the time it takes the FT-ICR MS to acquire a given number of scans. This scan is then started at approximately the same time as an acquisition is started on the FT-ICR MS, providing a rough synchronization between their scanning. While the OPO is scanning, a file is created with the elapsed time and the current frequency for use in creating the IRMPD action spectrum after the acquisition. The precursor and fragment ion intensities are extracted from the data file via Visual Basic script in Bruker Data Analysis, parallel to that used in **Section 2.3.3**.

2.4.4 FT-ICR MS IRMPD Performance

As discussed previously, the ICR cell provides an excellent environment for IRMPD action spectroscopy. The high-vacuum in the ICR cell prevents the collisional cooling found in the QIT-MS. However, the different shape and focusing of the ion cloud in the FT-ICR MS is expected to result in longer irradiation times. The IRMPD action spectrum of protonated tryptophan, $[Trp+H]^+$, was acquired on this instrument to verify both alignment and functionality of the FT-ICR MS as a tool for IRMPD. This spectrum is compared against a parallel spectrum acquired on the QIT-MS in **Figure 2.16**. The ions were irradiated for 2 seconds to acquire the spectrum on the FT-ICR MS, whereas similar IRMPD yield was achieved with irradiation for 0.5 seconds in the QIT-MS at similar laser powers. The FT-ICR MS has also been used to acquire a set of spectra for cisplatin

analogues that could not be acquired on the QIT-MS due to their extremely weak chromophores.[133] The acquisition of these spectra required irradiation times of ~20 seconds, but the same ions were not measurable on the QIT-MS as collisional cooling during the long irradiation times required prevented photodissociation. Although the acquisition of an IRMPD action spectrum on this FT-ICR MS platform is somewhat more challenging, and can take substantially longer, the ability to irradiate for extremely long times can allow for acquisition of spectra that simply are not possible on the QIT-MS platform described in **Section 2.3**.

Figure 2.1

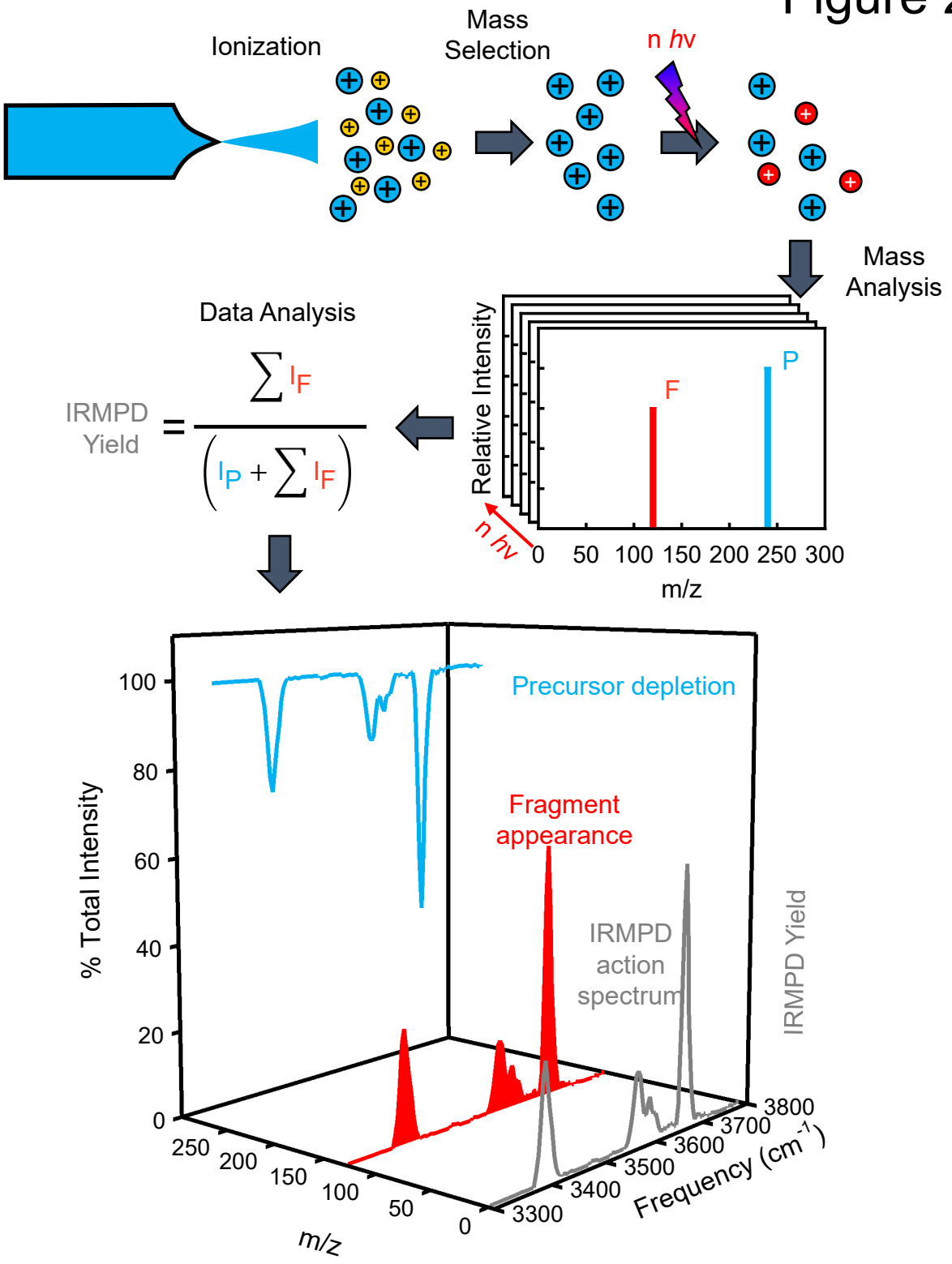


Figure 2.1 General workflow for an IRMPD action spectroscopy experiment. The precursor depletion spectrum and a fragment appearance spectrum are shown alongside the resultant IRMPD action spectrum.

Figure 2.2

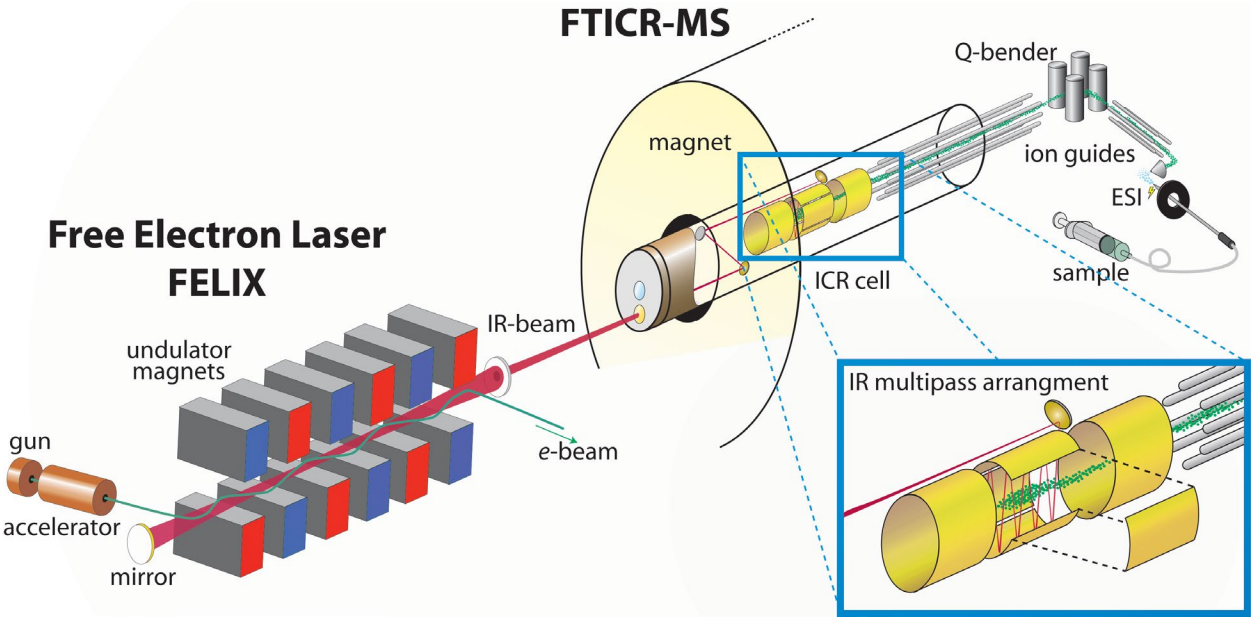


Figure 2.2 Schematic of the custom-built 4.7 T Fourier-transform ion cyclotron resonance mass spectrometer coupled to the FELIX free electron laser for IRMPD action spectroscopy. Alternatively the FT-ICR MS can be coupled to an OPO laser system (not shown).

Figure 2.3

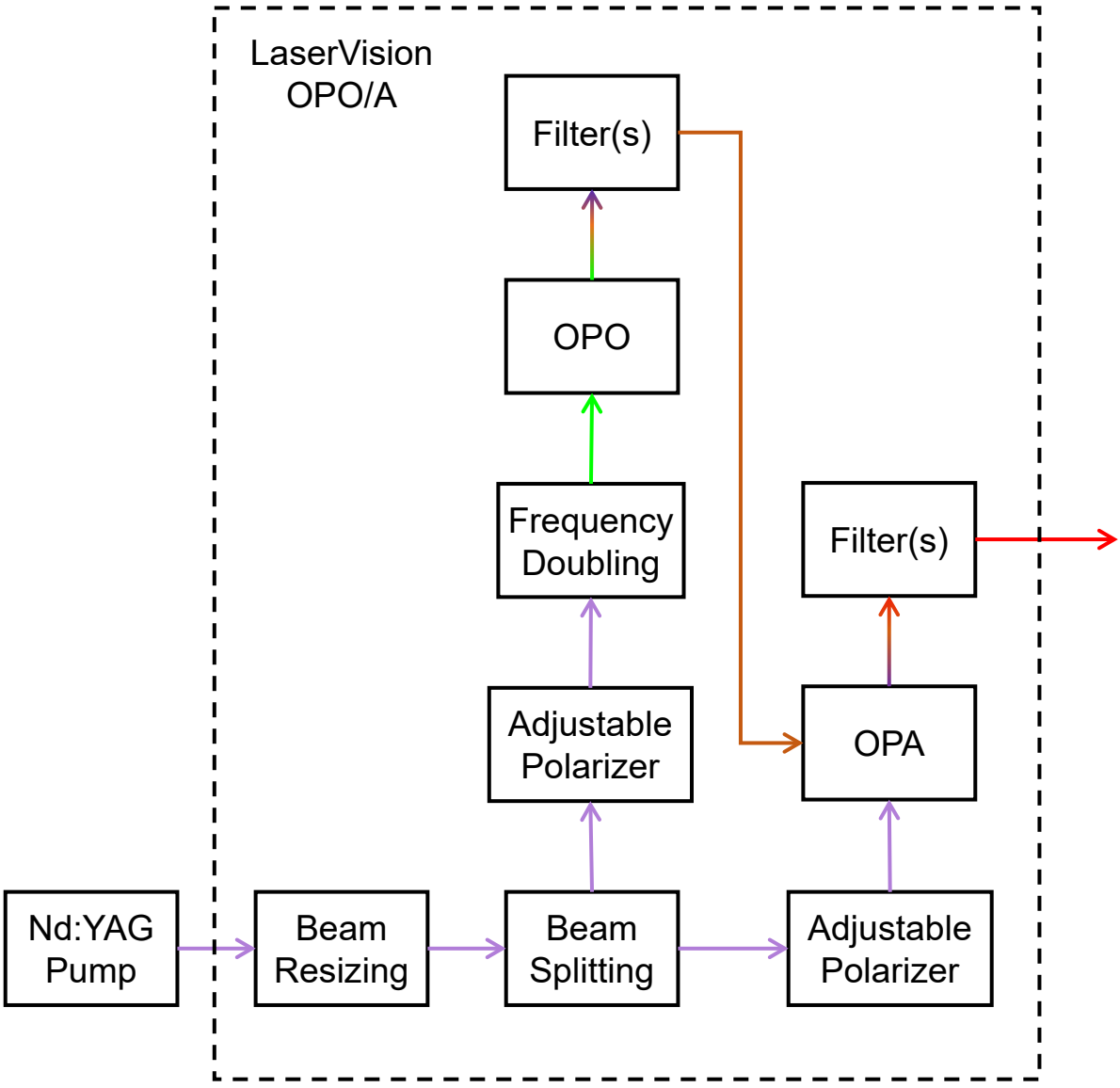


Figure 2.3 Schematic of the LaserVision optical parametric oscillator and optical parametric amplifier system (OPO). This system is pumped by the 1064 nm output of a Nd:YAG laser with ~7 W of power. For IRMPD action spectroscopy in the IR hydrogen-stretching region, the OPO is configured to output light in the range of 1970–4690 cm^{-1}.

Figure 2.4

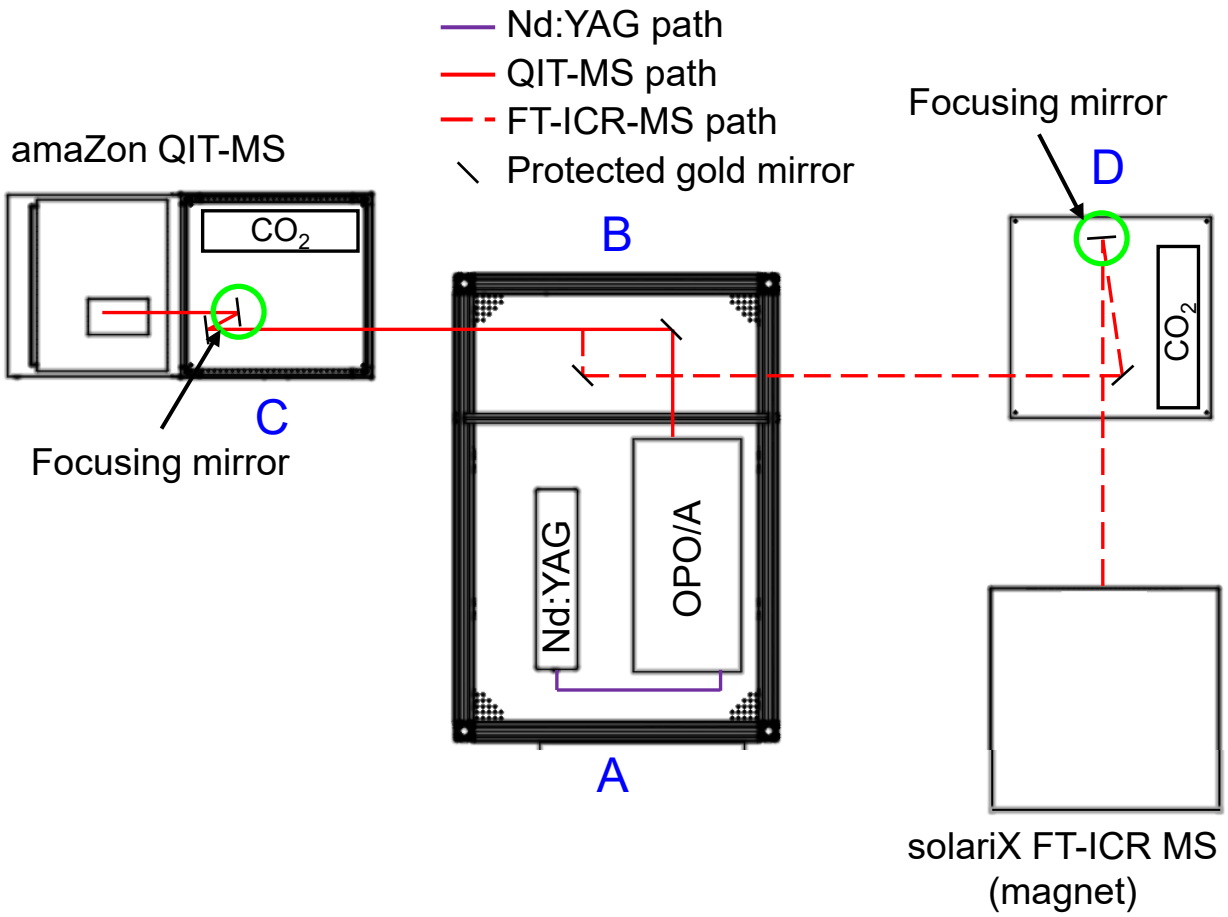
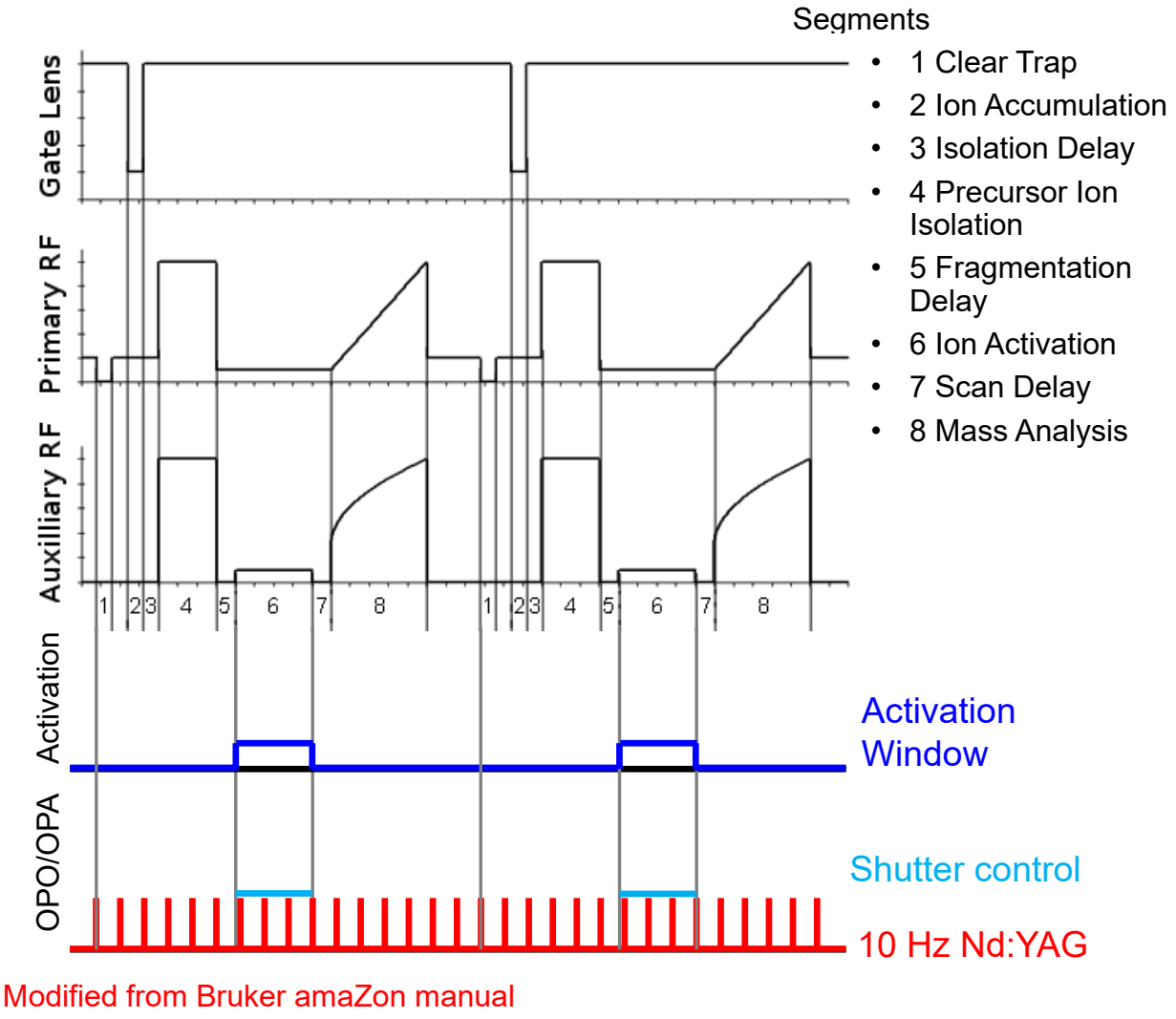


Figure 2.4 Schematic of the custom enclosures and beam paths that service the two modified commercial mass spectrometers coupled to the LaserVision OPO. Beam alignment is facilitated by 2" protected gold mirrors, with a single 2" convex focusing mirror at the end of each path to focus the beam onto the ion cloud.

Figure 2.5



Modified from Bruker amaZon manual

Figure 2.5 The scan sequence of the QIT-MS with important instrument signals shown with the segments of an individual scan. The activation window used to trigger an optical shutter is shown in blue and the Nd:YAG flashlamp signal is shown in red.

Figure 2.6

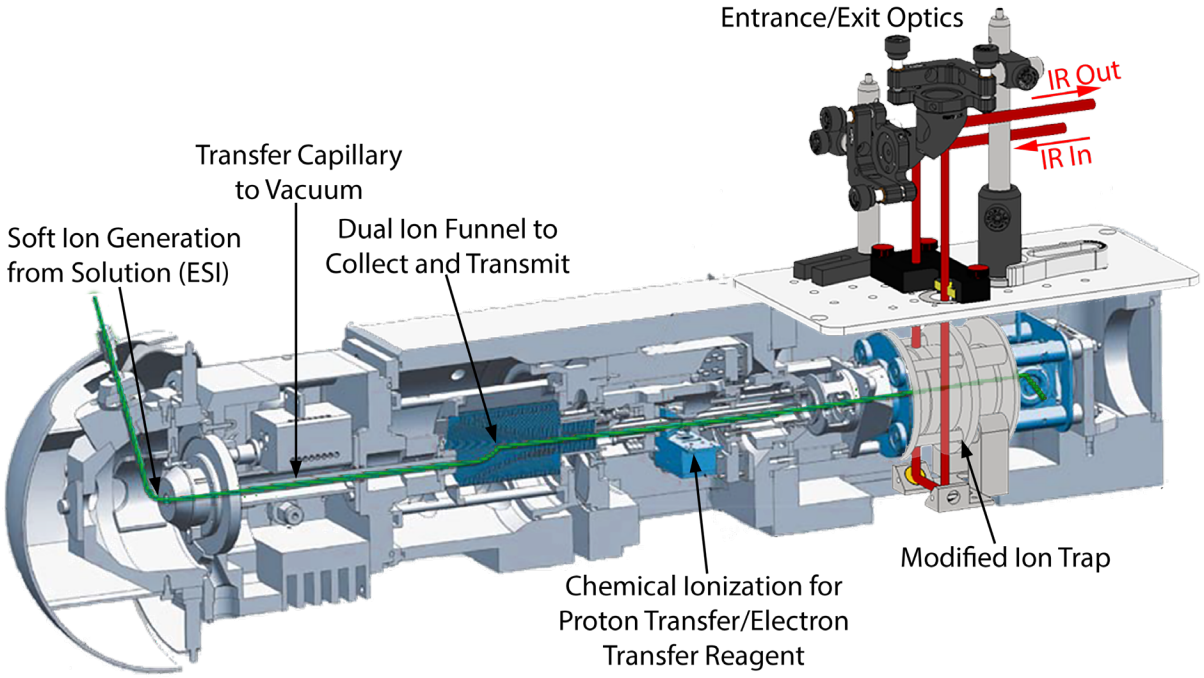


Figure 2.6 Schematic overview of the modified amaZon ETD quadrupole ion trap mass spectrometer and electrospray ionization source. The flange to the high vacuum region has been replaced with a small optical breadboard having two ZnSe windows for passage of the laser beam into and out of the vacuum region. The ion trap has been modified with a 2 mm hole machined through the ring electrode, oriented vertically in the instrument. Mirror mounts have been installed to guide the beam into and out of the vacuum region.

Figure 2.7

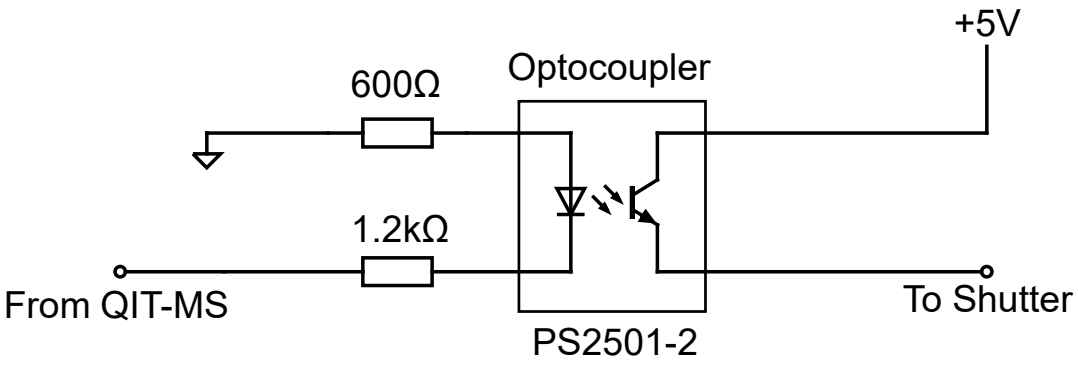


Figure 2.7 Simple optical isolation circuit to isolate the TTL signal from the QIT-MS during an activation window, and ensure it has sufficient power to trigger an optical shutter.

Figure 2.8

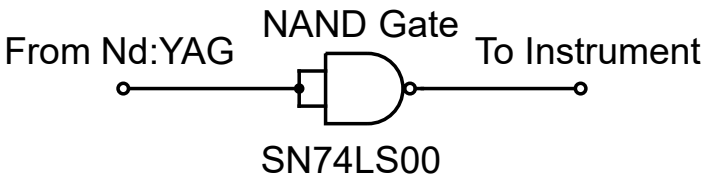


Figure 2.8 Simple circuit to convert the active-low TTL signal corresponding to the Nd:YAG flashlamp discharge, to an active-high TTL signal suitable for signaling the start of an acquisition sequence in either the QIT-MS or FT-ICR MS.

Figure 2.9

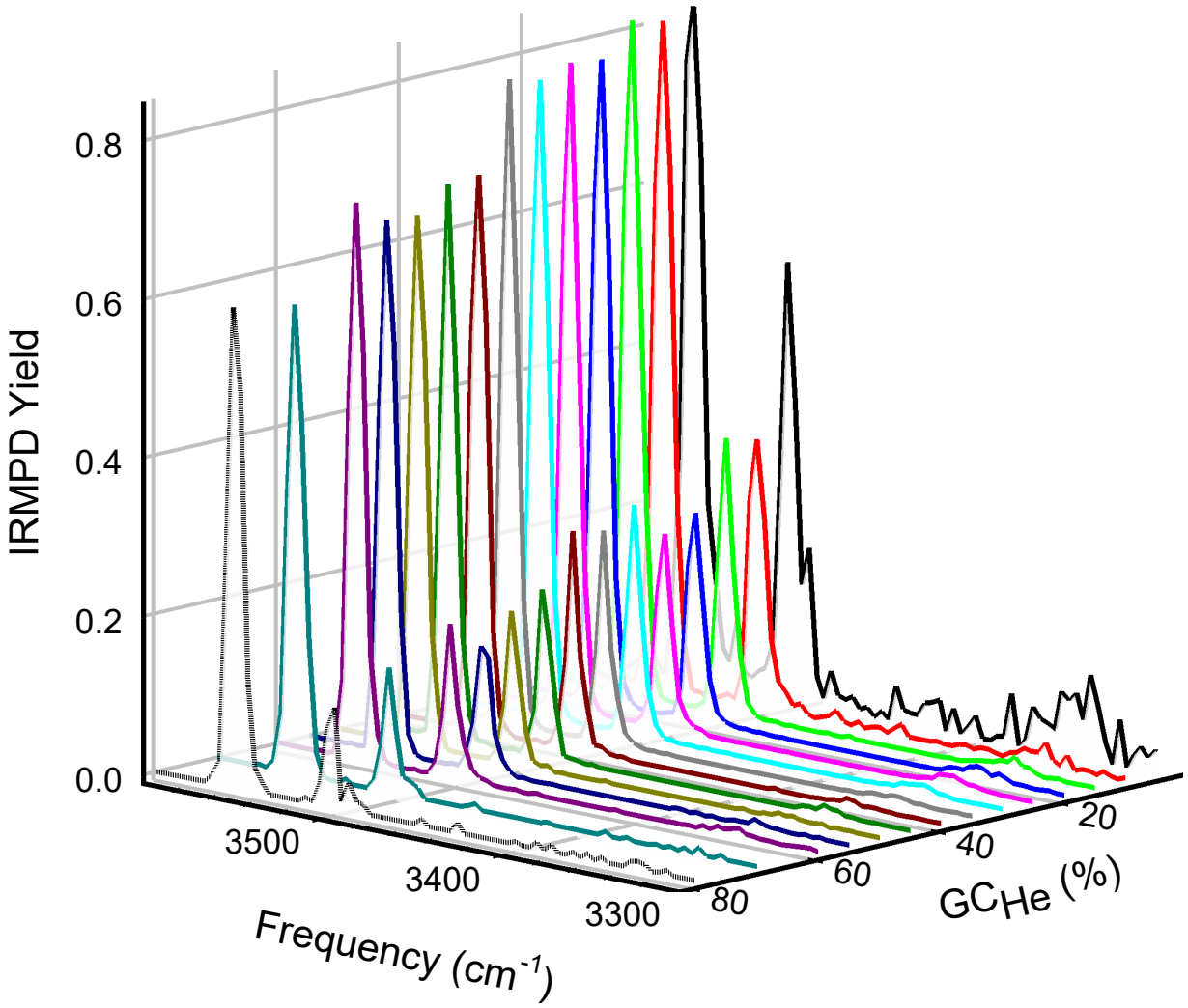


Figure 2.9 IRMPD action spectra of $[Trp+H]^+$ in the hydrogen-stretching region at a series of He pressures within the ion trap, controlled by the built-in He valve (GC_{He}) set from 5% to 80%. Laser power was de-tuned to roughly half of the maximum power to prevent saturation of the spectral features above 3500 cm^{-1}. Ions were irradiated for 0.5 seconds. The data represent averages of 10 MS/MS scan sequences.

Figure 2.10

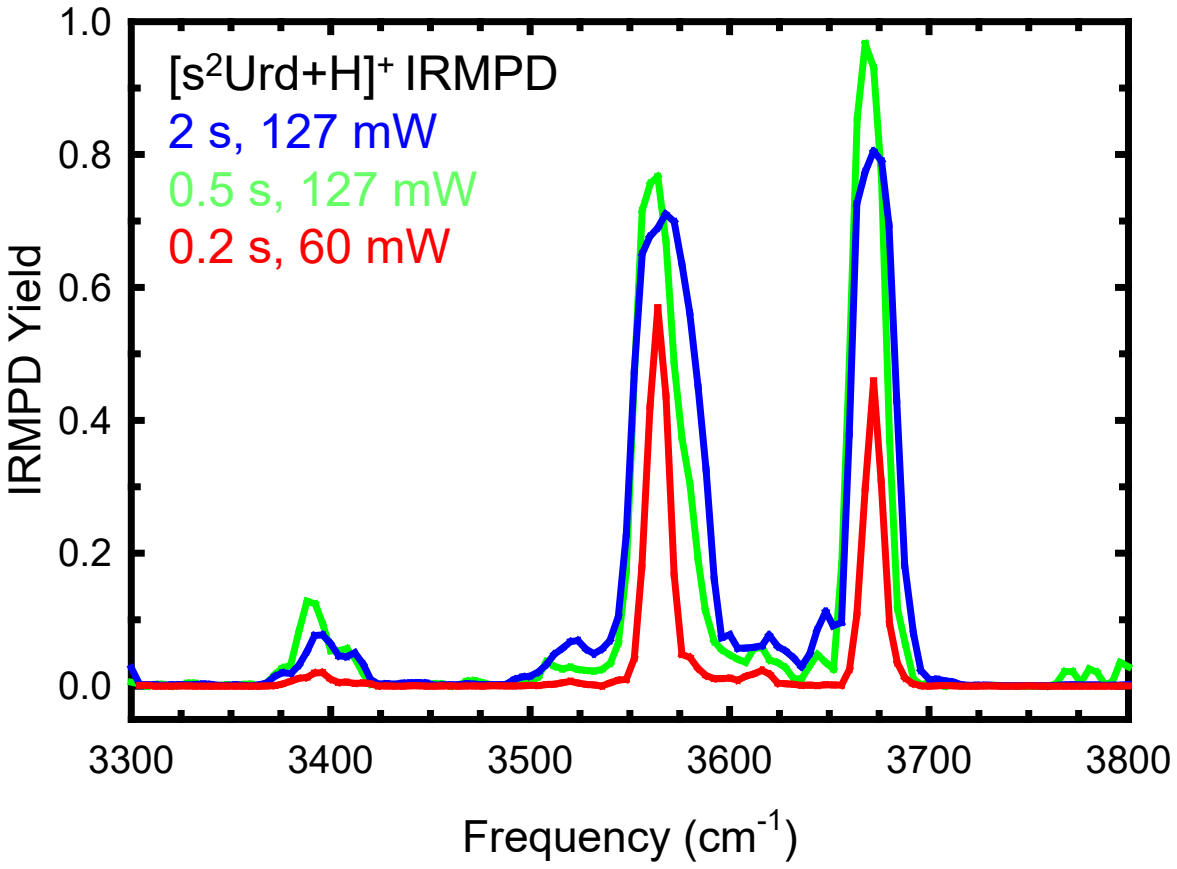


Figure 2.10 IRMPD action spectra in the hydrogen-stretching region of $[s^2Urd+H]^+$ acquired at different irradiation times and laser powers. The measured feature widths, relative intensities, and noise demonstrate the importance of choosing appropriate conditions in obtaining high quality, reproducible spectra that minimize analysis uncertainty.

Figure 2.11

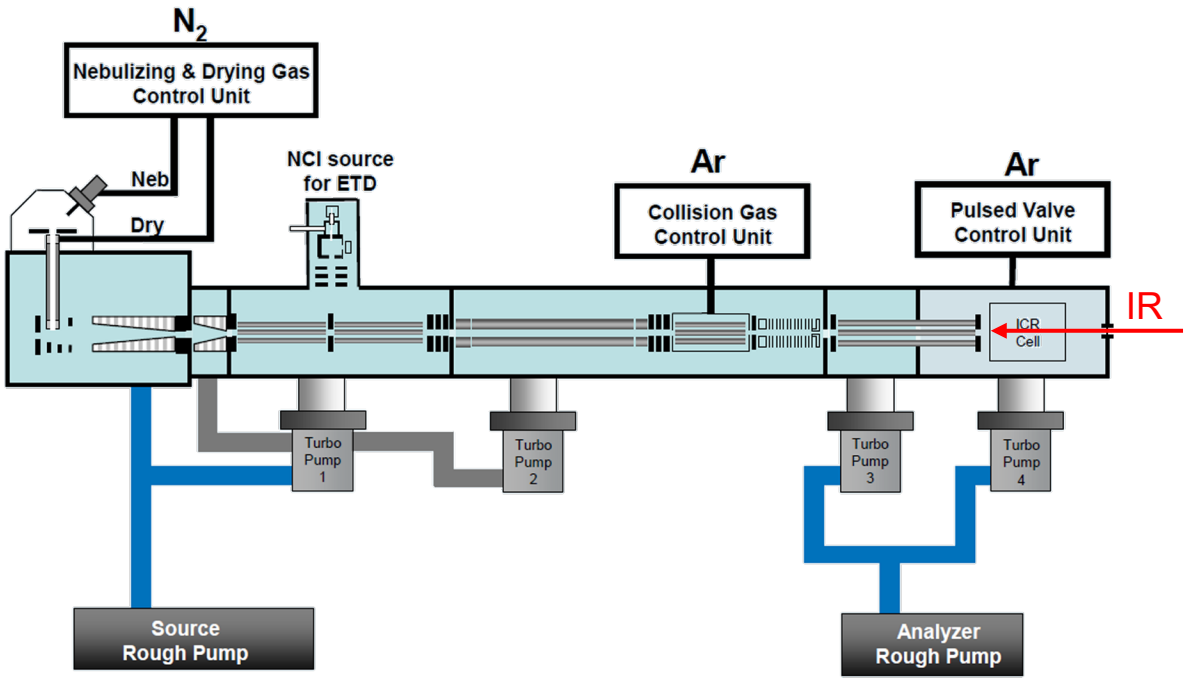


Figure 2.11 Schematic overview of the modified solariX Fourier transform ion cyclotron resonance mass spectrometer. A BaF_2 viewport is mounted to the end of the ICR cell to allow the OPO laser access to the ion cloud.

Figure 2.12

Synchronize Optical Shutter

```
# USER_EVENT_1_KEY:
USER_EVENT_1.lines = 1
USER_EVENT_1.1 = " (d25 setnmr4|9 setnmr4^9) ;initiate IR laser pulse,
d25 = duration "
```

Synchronize Nd:YAG and MS Scan

```
# USER_EVENT_2_KEY:
USER_EVENT_2.lines = 1
USER_EVENT_2.1 = " 10 u trigpe1 ;wait indefinitely for trigger pulse
from Nd:YAG  "
```

Figure 2.12 Custom user event definitions in BASIC to pass out a signal for triggering the optical shutter during the activation window, and accepting the trigger signal from the Nd:YAG.

Figure 2.13

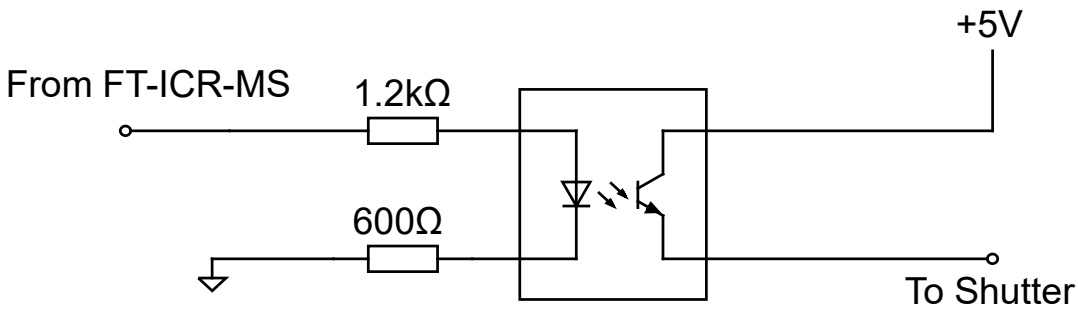


Figure 2.13 Optical isolation circuit to isolate the activation window TTL signal from the FT-ICR and ensure it can trigger the optical shutter at the output of the OPO.

Figure 2.14

```
;-----------------------------------------------
;Initialization Block
;-----------------------------------------------
Standard commands

;-----------------------------------------------
;Hystar autosuspend Block
;-----------------------------------------------
Standard commands

;-----------------------------------------------
;Dataset Acquisition Block ("2D" loop
;-----------------------------------------------
Standard commands

;-----------------------------------------------
;Scan Accumulation Block (NS loop)
;-----------------------------------------------
Standard commands

;-----------------------------------------------
;Source & Infinity Cell Quench Block
;-----------------------------------------------
USER_EVENT_2
ION_QUENCH
```

Figure 2.14 An example of the placement of 'USER_EVENT_2' within the BASIC_IRMPD_ON file. This configures the scan to wait for the trigger from the Nd:YAG before quenching the ICR cell following the previous scan.

Figure 2.15

```
;-------------------------------------------------
;Ion Accumulation Block (end)
;(repeat Accumulation blocks for Source, Collision Cell, Infinity Cell)
;-------------------------------------------------
Standard commands

;============================================================
; INTERNAL MS/MS
; --> Possible Events automatically generated
;  1) INFINITY_CELL_ISOLATION
;  2) PULSED_VALVE_1, PULSED_VALVE_2 (for post-isolation/pre-SORI)
;  3) SORI
;  4) ECD
;============================================================
INTERNAL_MSMS
USER_EVENT_1
```

Figure 2.15 An example of the placement of 'USER_EVENT_1' within the BASIC_IRMPD_ON file. This places the IRMPD activation window immediately after the first MS/MS event configurable from the solariXcontrol user interface.

Figure 2.16

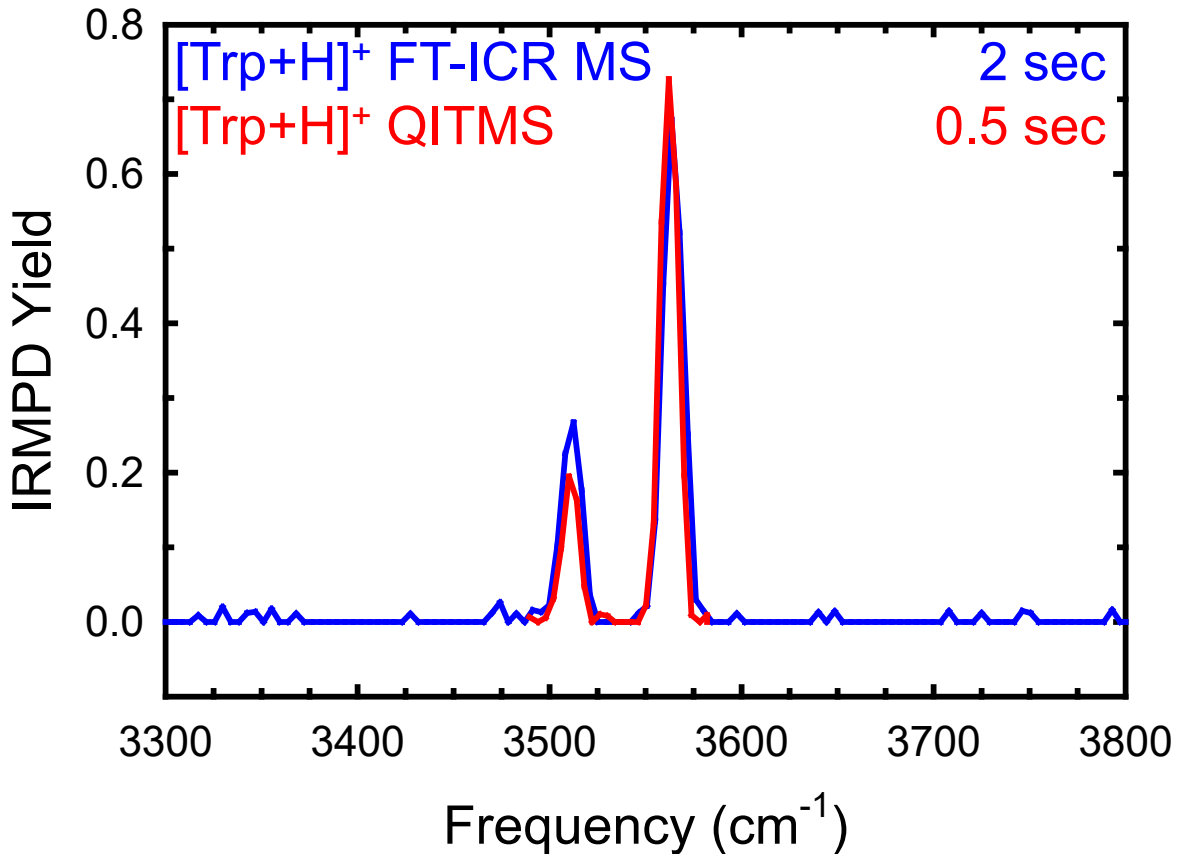


Figure 2.16 IRMPD action spectra of protonated tryptophan, $[Trp+H]^+$ collected on the Bruker amaZon ETD QIT-MS and Bruker solariX FT-ICR MS.

CHAPTER 3 DEVELOPMENT OF SOFTWARE AND SCRIPTS FOR ANALYSIS AND AUTOMATION OF COMPUTATIONAL CHEMISTRY CALCULATIONS

3.1 Computational Chemistry Calculations

Computational chemistry is an extremely valuable tool in the analysis of experiments probing the intrinsic properties of molecules such as the spectroscopic experiments described in **Chapter 2**. The analysis of an IRMPD action spectrum typically relies upon many geometry optimization and frequency analysis calculations followed by comparisons with those calculated vibrational frequencies. To provide useful spectroscopic comparisons for the analysis of an experimental IRMPD action spectrum a conformational search process is used to broadly explore the potential energy landscape and find stable, low-energy conformers that can be further optimized, and their vibrational frequencies calculated. An overview of the conformational search and optimization process used in this work is shown in **Figure 3.1**.

3.1.1 Conformational Search

To inexpensively explore the conformational space available to an ion of interest, a molecular dynamics simulated annealing based conformational search is performed using molecular mechanics. Molecular mechanics uses a parameterized force field to simulate Newtonian molecular motion. The simulated annealing procedure simulates heating of the ion, allowing it to equilibrate at an elevated temperature and explore conformational space accessible at the internal energy available, before slowly cooling it to encourage relaxation to a low-energy conformation. As molecular dynamics does not readily allow for bond breaking or bond formation, individual conformers representing each of the favorable protonation sites of a given protonated nucleoside analogue are

examined individually. The protonated 2- and 4-thiouridine nucleoside analogues discussed in **Chapter 4** and the arabinose nucleosides examined in **Chapter 5** were parameterized for the molecular dynamics simulations with parameters from the Amber 3 force field in the Hyperchem 8.0 suite of programs.[134,135] The 2′,3′-dideoxyribose nucleosides examined in **Chapter 5** were instead parameterized in the Antechamber program[136] using parameters from the *ff14SB*[137] and *GAFF*[138] force fields to examine any changes in the resulting conformations as a result of parameterization. Each simulated annealing procedure begins with the simulation temperature at 0 K which is then increased to 1000 K over 0.3 ps, held at 1000 K for 0.2 ps, and then reduced back to 0 K over 0.3 ps. At 1000 K there is enough thermal energy in the rovibrational modes of the ion to facilitate relatively comprehensive exploration of the conformational space available to the ion.[55,57,139-144] This temperature cycling is repeated for at least 300 cycles, and up to 3000 cycles generating as many candidate conformers as temperature cycles simulated. The computed energies of each of the resulting conformers are extracted and can be used alongside automated or manual evaluation of the candidate conformers to select structurally-unique, low-energy conformers to carry forward for further electronic structure calculations. Automated methods for selecting unique, low-energy conformers from the conformational search are detailed in **Section 3.5**, with manual selection methods relying heavily upon relative energetics and manual examination and structure manipulation.

3.1.2 Electronic Structure Optimization and Frequency Analysis

Conformers from the molecular dynamics simulations frequently display structural features that are obviously not ideal, such as slightly bent aromatic rings. Further, the

molecular dynamics simulations do not provide vibrational frequencies or accurate energetics. Although expensive compared to molecular dynamics calculations, electronic structure calculations can provide a wealth of useful information for detailed study of the intrinsic properties of a molecule or ion. Amongst the myriad of available electronic structure methods, density functional theory (DFT) methods tend to be relatively inexpensive yet still provide reasonably accurate relative energetics and vibrational frequencies.[145] The geometries of the conformers selected from the conformational search are optimized with the B3LYP DFT functional and a 6-311+G(d,p) basis set; the vibrational frequencies and Raman intensities are also calculated at this level of theory. Previous work on protonated nucleosides has demonstrated that this theoretical approach provides vibrational frequencies in good agreement with the measured IRMPD spectrum, and represents a good compromise of quality and cost for geometry optimization.[51-55] Single point energies are calculated at the B3LYP/6-311+G(2d,2p) level of theory, which provides better relative energetics than the geometry optimization/frequency analysis calculations, without significantly increasing the overall cost of the electronic structure calculations. Relative Gibbs energies at 298 K are calculated by including the zero-point energy (ZPE) and thermal corrections calculated at the B3LYP/6-311+G(d,p) level of theory.

3.1.3 Predicted IR Spectra

Inherent to the frequency analyses performed using electronic structure methods is a harmonic approximation for all of the vibrational modes of the ion. Treatment of these vibrational modes as harmonic oscillators substantially decreases the cost of the electronic structure calculations. However, all of the normal vibrational modes of a system

are affected by differing levels of anharmonicity and if the predicted vibrational frequencies assume fully harmonic vibrational modes, differences between the predicted and measured spectral features are expected.[105,123,146,147] Fortunately, scaling of the predicted harmonic vibrational frequencies is a reasonably robust and well accepted practice for treating predicted vibrational frequencies for comparison to experimental vibrational spectra.[131,148,149] Due to the varying anharmonicity of different vibrational modes, separate scale factors are selected to treat similar vibrational modes. For example, in IRMPD action spectroscopy the vibrational frequencies predicted in the IR fingerprint region (600–1900 cm^{-1}) typically involve heavier atoms than those in the hydrogen-stretching region (3300–3800 cm^{-1}), and are affected by anharmonicity differently, such that the two regions are best treated with distinct scale factors to produce the best agreement between predicted IR spectra and the experimental IRMPD action spectrum.

Each protonated nucleoside is also affected by anharmonicity uniquely, and as such is treated individually with an individual set of scale factors to produce the best agreement with the measured IRMPD action spectrum. The choice of scale factors is an important component in the analysis of the measured IRMPD spectrum by the predicted IR spectra, as it can easily introduce non-spectroscopic biases. Therefore, all of the predicted IR spectra calculated are considered when choosing a scale factor, not just the most stable among them. However, assignment of a scale factor is further refined following this initial assessment by considering only those conformers that present the best agreement with the measured IRMPD spectrum and are low in Gibbs energy. To facilitate visual comparison with the measured IRMPD spectrum for selection of

vibrational scale factors, and for analysis of the conformers present in the experiment, the predicted vibrational frequencies and intensities are convoluted with Gaussian peak shapes. A full width at half maximum (fwhm) of 20 cm^{-1} is used in the IR fingerprint region to best represent the experimental peak widths in this region and a fwhm of 15 cm^{-1} is used in the hydrogen-stretching region.

The process of analyzing the experimental IRMPD spectrum by comparison with the predicted IR spectra to determine those conformers present in the experiments is shown diagrammatically in **Figure 3.2**. The major spectral features in each region of a predicted IR spectrum, typically the most intense and frequently relatively simple features, are compared with the experimental IRMPD spectrum first. If the major features provide poor representation of the central frequency and peak shape of the experimental spectrum, the respective conformer is not relevant to the experiment. Important minor features to consider are typically less intense and sometimes more complex than major features. However, some of the most important minor conformers to consider are those that are unique whether that be a specific protonation site or some other unique structural feature. As expected, conformers that display good agreement with the major features as well as good agreement with the position, shape, and intensity of important minor features are likely the most important conformers in the experiments. Conformers that agree well with the major spectroscopic features and have good agreement with the position and shape of minor features, but disagree with the measured intensities of these features, are generally only expected to be minor contributors to the experiments depending on their relative stabilities. The nature of the vibrational mode associated with the minor features that agree well is also important in determining if a conformer contributes to the

experimental spectrum. Conformers that agree well with the major features, but not the important minor features, are less likely to be important contributors to the experiment, especially if they are high in Gibbs energy.

3.2 LabView Graphical Programming Environment

National Instruments LabVIEW™ (laboratory virtual instrument engineering workbench), is a general-purpose programming environment. LabVIEW™ particularly excels in data acquisition, analysis, and presentation along with instrument control, with a great number of diverse built-in tools facilitating these applications. National Instruments also produces a variety of hardware with built-in driver support in LabVIEW™ that enables rapid and robust development of computer-instrument interfaces. Some other compelling advantages of the LabVIEW™ professional development system are the ease of designing and implementing interactive, intuitive graphical user interfaces (GUI), and the straightforward creation and distribution of standalone executables and installers.

A typical LabVIEW program, or Virtual Instrument (VI), is comprised of two parts, a front panel, containing the GUI, and a block diagram, which contains the functions it performs, represented graphically similarly to a data flow diagram. The front panel relies upon a system of controls (input), constants (display), and indicators (output), with the majority of common application controls such as sliders, buttons, text and numeric inputs, and graphs readily available. Custom controls (XControls) can also be created, allowing for new GUI functionality beyond the standard controls. The controls, constants, and indicators on the front panel are represented on the block diagram for use, and properties such as value, appearance, and behavior can be customized by editing the front panel in

the development environment or programmatically via property nodes on the block diagram.

A significant advantage of the LabVIEW™ platform is the ease with which executables and installers can be generated. A LabVIEW™ executable will run on the operating system in which it was built, and only requires the appropriate version of the LabVIEW™ runtime software, and any other necessary components to be installed. LabVIEW™ installers allow for distribution of the executable, runtime, and any other required components, greatly improving the user experience, and particularly useful for instrument control applications where the necessary device drivers can be managed within in the development environment and distributed with the control software.

3.3 Automation of Management and Analysis of Theoretical Calculations

The preparation of input files for computational chemistry and the extraction of information from output files largely consists of simple text manipulation and string parsing. This type of task becomes especially arduous when performed for hundreds to thousands of files at a time. Further, human errors can be difficult to identify after the calculation or subsequent analysis. A preliminary attempt to automate parsing of the output files from electronic structure calculations was developed as a simple bash script, compatible with the Unix environment of the Wayne State University Grid where the electronic structure calculations are performed.

In order to expand the functionality of the automation tools available, a series of subVI were developed in LabVIEW™ to address the specific set of tasks presented by the most common conformational search workflow at the time. Those tasks were parsing the output file from a simulated annealing procedure in Hyperchem 8.0, conversion of the

output structure file type into a more useful format, and incorporation of that structure into an input file for electronic structure calculations in Gaussian 09. This strict sequence of tasks was built into a rigid graphical interface that lacked any meaningful opportunity for further customization. Although it was used across several projects to analyze and prepare several hundred calculations, this lack of flexibility greatly limited use of this early tool outside of that specific set of tasks.

In order to serve more diverse needs these subVIs were revisited to expand their flexibility and functionality. Instead of a static interface with limited availability for expansion, a new framework for access to these subVI and more was also developed. Grid Helper, as this new development was labeled, is ultimately just a framework for access to these subVI, which allows for new subVI and functionality to be added more readily. The Grid Helper interface and the subVI used to create input files for Gaussian 09, are shown in **Figure 3.3** along with the general workflow facilitated by Grid Helper. Grid Helper relies upon the start asynchronous call node,[150] a subpanel on the main front panel, and queued message handlers to dynamically call individual subVI that perform specific tasks with little to no dependence on Grid Helper itself. The listbox on the left of the front panel is populated by a set of XML files that list the primary subVI available in the Grid Helper installation directory. Upon a change in listbox selection a message queue instructs the currently active subVI to stop, then finds the subVI corresponding to this new selection and asynchronously initializes the new subVI within the subpanel on the front panel. Each primary subVI is independent of each other, even if their functionalities are tied together in the workflow. Further, each primary subVI is only dependent upon Grid Helper as a launcher. Any transfer of data between the subVI must occur through a file

that is saved in a human-readable format and visible to the user. The human-readable format allows for visibility into the process and manual editing of the files when necessary. The addition of new subVI, and therefore new functionality was designed to be straightforward. A template used to develop the existing and new subVI is readily available within the LabVIEW™ project. When a new subVI is developed it can simply be added to the Grid Helper installation directory. The appropriate XML file can be edited to include the new subVI such that re-compiling or re-installation of Grid Helper is not required. At the time of writing, six subVI are available within Grid Helper, and handle nearly all the batched tasks required for the most common workflows used in the computational approach to IRMPD data analysis as described in **Section 3.1**. These subVI are described in **Sections 3.3.1-3.3.3**. Although Grid Helper is available as an installer with the LabVIEW™ runtime included, a dependency on Open Babel[151,152] for file conversion requires a separate installation of the Open Babel package for that functionality, although the other subVI will work without it.

3.3.1 Input Generation Tools

The conformational search process performed in HyperChem 8.0, described in **Section 3.1.1**, produces a set of structure files in the HyperChem HIN format, and a log file of the recorded simulation temperatures and energetics. The *Extract Energies* subVI parses that log file of energies and extracts the energy and file name from the last step of each simulated annealing cycle. This information is written to a separate text file for later use and manual inspection, if required. The *Convert *.hin to *.com* subVI utilizes the Open Babel package[151,152] to convert the HyperChem structure files to Gaussian cartesian input files that have the correct formatting for copying directly into a Gaussian

09 *.com input file. *Make OptFreq/SP Inputs*, which is shown in **Figure 3.3**, is then used to create these input files. Either specific structure files or a directory of structure files can be used to populate the listbox control, which allows for manual selection amongst these files or automated selection based upon the relative energies written by *Extract Energies*. The *Extract Energies* subVI can be used to automatically select the N most stable structures generated by a simulated annealing run, for example. A variety of string inputs are used to configure parameters for the electronic structure calculations and submission to the Wayne State University Grid. These parameters and the Gaussian cartesian input file(s) are then converted into Gaussian 09 input file(s) using a template file. This template file uses a set of regular expressions of the form *%filename%* associated with the string inputs. This template is a simple text file, allowing for substantial flexibility beyond the default geometry optimization, frequency analysis, and single point energy calculations it is configured for by default.

3.3.2 Log File Parsing

Following electronic structure calculations by Gaussian 09, the resultant output files are parsed to extract their relevant energetics as a component of further analysis. The *Parse Log Files* subVI, shown in **Figure 3.4**, extracts the energetics from large sets of Gaussian 09 output files. Atomic positions can also be extracted and used to calculate dihedral angles, bond angles, and bond distances. This data is passed into a Microsoft Excel workbook for further use, and creates a comma-delimited file as a backup. Several recent publications have leveraged this tool to gather energetic and structural information for large numbers of conformers, with substantially decreased effort relative to the manual extraction procedure previous employed.[53,63-65,153-156]

3.3.3 IRMPD Data Processing and Preliminary Analysis

The *Spectra* subVI, shown in **Figure 3.5**, is the first step in processing electronic structure calculations for comparison with experimental IRMPD action spectra. The vibrational frequencies and intensities are extracted from the frequency analysis calculation and used to generate the convoluted spectrum described in **Section 3.1.3**. An experimental IRMPD action spectrum can also be loaded as a comma-delimited XY file. Comparison of the convoluted spectra with the experimental IRMPD spectrum in this interface facilitates not only selection of the appropriate scale factor, as described in **Section 3.1.3**, but also an initial analysis to identify the predicted IR spectra for the structures most likely present in the experimental measurements. The extracted vibrational frequencies and intensities, as well as the convoluted spectra can then be saved as a comma-delimited file for a more careful analysis in other graphing software.

3.4 Python

Python[157] is a widely used, high-level, interpreted programming language. In recent years it has gained substantial attention both as a web-development tool, and as a powerful platform for data analysis. The latter of which is enabled by packages that add substantial mathematical and scientific functions and modify the existing high-level data structures of Python into nomenclature and functionality more familiar to the scientific community. Popular Python packages adding substantial high-level mathematical functionality such as NumPy[158] or those that facilitate high-level scientific analyses such as SciPy[159] allow for substantial code reuse and subsequently significantly faster prototyping of data processing and analysis scripting in Python. Python heavily emphasizes readability, making it relatively straightforward to learn and understand.

Python being an interpreted language allows for rapid development and prototyping of small code segments and for interactive data analysis using Python notebooks. The Python interpreter operates on most major operating systems, allowing for code reuse across platforms, from Windows personal computers to Unix-based grids.

3.5 Conformational Search Tools

Relying upon conformer energies calculated at the relatively simple molecular mechanics level employed in this and similar work as indicators of structural uniqueness or the quality of a low-energy conformer can be very unreliable. If important conformers are not sampled by the conformational search then they must be discovered and generated manually, which can be particularly challenging. Historically, a more energy biased approach has worked reasonably well, but this is likely a result of the number of repeated conformational searches performed and a reliance upon previous studies of related systems to guide the search process. A more robust method for sampling unique conformers from the conformational search without manual intervention would greatly benefit systems where previous work is not available or incomplete.

3.5.1 Cartesian Coordinate Similarity

The root mean square deviation (RMSD) between two structures is a common measure of structural similarity. Several computational chemistry packages use the Kabsch algorithm[160] to calculate the minimum RMSD between two structures.[161-163] Fortunately, portable implementations of the Kabsch algorithm to calculate the optimal rotation matrix to minimize RMSD between two sets of points are available in several programming languages. A relatively straightforward Python package of the Kabsch

algorithm[160,164] was used to begin development of a structural comparison tool specifically for sampling unique conformers from the population found from a conformational search.

To identify structurally unique conformers an RMSD cutoff is used. To reduce the time required to calculate RMSD values for large sets of conformers, conformers are only compared with those that have already been identified as unique. This process begins by calculating the minimum RMSD between the first two conformers in the set. If this minimum RMSD exceeds the set cut-off value, the conformers are considered *unique* and are placed on a list of *unique* structures with which to compare additional conformers. The minimum RMSD values are then calculated for the next conformer against those that have been identified as *unique*. If all of the RMSD values exceed the cut-off then this conformer is considered *unique* and added to the list. This process is repeated through the entire set of conformers. If a conformer is found to be similar to an existing *unique* conformer, their calculated energies can be compared and the most stable conformer retained as representative of that structure. Through this process two output files are created: one containing the *unique* conformers and their calculated energies, and the other containing the minimum RMSD calculated for each conformer against the *unique* conformers. These files are used alongside manual visualization of the structures as necessary to choose an appropriate RMSD cutoff for the system of interest, making this an iterative process.

3.5.2 Structural Parameter Similarity

Another method for identifying conformers with unique structures, specifically for the nucleosides examined in this work, is to focus on the structural parameters that characterize the structure of the nucleoside. The most prominent of these parameters are

the glycosidic bond angle, pseudorotation angle of the sugar moiety, and O5′ hydroxy orientation, which are all described in detail in **Section 1.2**. Each of these parameters is either a dihedral angle or is calculated from several dihedral angles. Fortunately, dihedral angles are readily calculated from atomic coordinates, which have already been extracted during the Cartesian coordinate RMSD method described in **Section 3.5.1**. An RMSD can then be calculated from the respective dihedral angles of the two structures being compared using **Equation 3.1**.

$$\text{RMSD}=\sqrt{\sum_{i=0}^{N}|X_i - Y_i|^2/N} \tag{3.1}$$

Where X_i and Y_i are the same parameter measured for the two structures being compared and N is the number of parameters examined. Calculating an RMSD value between two structures via structural parameters is a drop-in replacement for the Kabsch RMSD calculation in the script detailed in **Section 3.5.1** and can be used in the same manner to determine unique conformers.

3.5.3 Structural Filtering Performance

Ideally, calculating the minimum RMSD value for each conformer examined would reveal a step-function clearly indicating an RMSD value above which structures are unique. In order to assess the performance of each of the two RMSD calculation methods described in **Sections 3.5.1** and **3.5.2,** both methods were used to calculate the minimum RMSD values for all of the conformers within three separate sets of calculations. These sets of calculations are taken from different stages within the conformational search process described in **Section 3.1**. Examining the performance of the automated structural

filtering process at multiple stages during the conformational search process not only helps reveal when it performs best, but may also help guide further development.

The minimum RMSD values of conformers directly from the molecular dynamics conformational search of protonated cytosine arabinoside, $[araCyd+H]^+$, are shown in **Figure 3.6**. This conformational search was performed in Hyperchem 8.0 with the Amber 3 force field and no explicit molecular dynamics optimization of the conformers after their temperature cycles. The minimum RMSD values of all 300 conformers examined calculated by the Cartesian coordinate method are shown in red, and those calculated using the structural parameter method are shown in blue. Along with histograms of the RMSD values calculated, the % of conformers with RMSD values below a given value are also shown for both methods. The minimum RMSD values calculated based on Cartesian coordinates display a nearly normal distribution around 1 and do not clearly indicate a good cut-off value that would sample a reasonably small number of conformers. The RMSD values calculated by the structural parameter method are much larger, a result of measuring dihedral angles instead of coordinates, which have a much larger range. This distribution of RMSD values is far less symmetric and tails noticeably towards larger RMSD values. This distribution is better for arbitrarily selecting a RMSD cut-off while still maintaining a reasonable number of sampled conformers for the more expensive DFT calculations, but does not display as clear an RMSD cut-off as desired. For both distributions in **Figure 3.6**, a dominant concern when selecting an RMSD cut-off is the cost of the following DFT calculations, and as such values of 1.2–1.3 and ~20 would likely be suitable, respectively.

The second set of conformers examined are of protonated 2′,3′-dideoxycytidine, $[ddCyd+H]^+$. These conformers are also taken directly from molecular dynamics simulations, but the dynamics simulations were performed using the AMBER program with the *ss14SB* force field. These conformers were then subjected to an explicit molecular dynamics optimization step following the temperature cycle before analysis. The Cartesian coordinate RMSD values are shown in red in **Figure 3.7**, and those calculated based on structural parameters are shown in blue. A very different distribution of minimum RMSD values is now observed for the Cartesian coordinate method, with a somewhat more even distribution between 0 and 2.2. Several potential RMSD cut-off values can be chosen around 1.0, 1.5, and 2.0. However, as there are 3000 conformers present in this analysis, typically 10 to 15% of these conformers would be sampled for further electronic structure calculations, to minimize the cost. The efficiency of this analysis might be further enhanced by also considering the calculated energetics, but this might also result in ignoring interesting or important conformers that could be kinetically trapped in the experiments. The RMSD values calculated using the structural parameter method present a very different picture. Not only are the RMSD values dramatically smaller than those calculated by the same method in **Figure 3.6**, but nearly all of the conformers examined fall within 0.4 RMSD units of one another. This is a clear indication that the addition of a structural optimization step following the molecular dynamics simulation substantially condenses the number of structures observed. This also allows for selection of an RMSD cut-off that provides a reasonable limit to the number of conformers sampled during the following electronic structure calculations.

The final set of conformers examined follow directly from the $[ddCyd+H]^+$ conformers sampled from **Figure 3.7**. An RMSD cut-off of 0.35 was used with the structural parameter RMSD values to sample 76 unique conformers, which were subjected to a geometry optimization at the B3LYP/6-31+G(d) level of theory. These further optimized conformers were then analyzed by the two RMSD methods with the results of these analyses shown in **Figure 3.8**. The Cartesian coordinate RMSD values clearly indicate two separate distributions, with ~40% of the calculated conformers displaying minimum RMSD values ≤1.2, indicating that several similar conformers were still sampled following the earlier filtering step. This is even more clearly seen in the structural parameter RMSD values, with ~50% of the conformers sampled displaying minimum RMSD values ≤0.1. Ideally, the first structural filtering step would sample no similar conformers, maximizing the effectiveness of the more expensive electronic structure calculations. However, a preliminary DFT optimization, as displayed here, is relatively inexpensive, and can then guide the more expensive DFT optimization, frequency analysis, and single point energy calculations to the most interesting and *unique* conformers.

For the conformational search of $[ddCyd+H]^+$, a two-step approach with the automated structural parameter filtering method captured nearly all of the important low-energy conformers. However, a single absent conformer was identified in the subsequent data analysis. Fortunately, this was found by intuition and an unexpected gap in the glycosidic bond dihedral angle distribution of the calculated conformers. This is not ideal, as a less familiar system would be more likely to result in missing this manual check such

that further improvements to the selection of RMSD cut-offs should continue to be pursued.

Figure 3.1

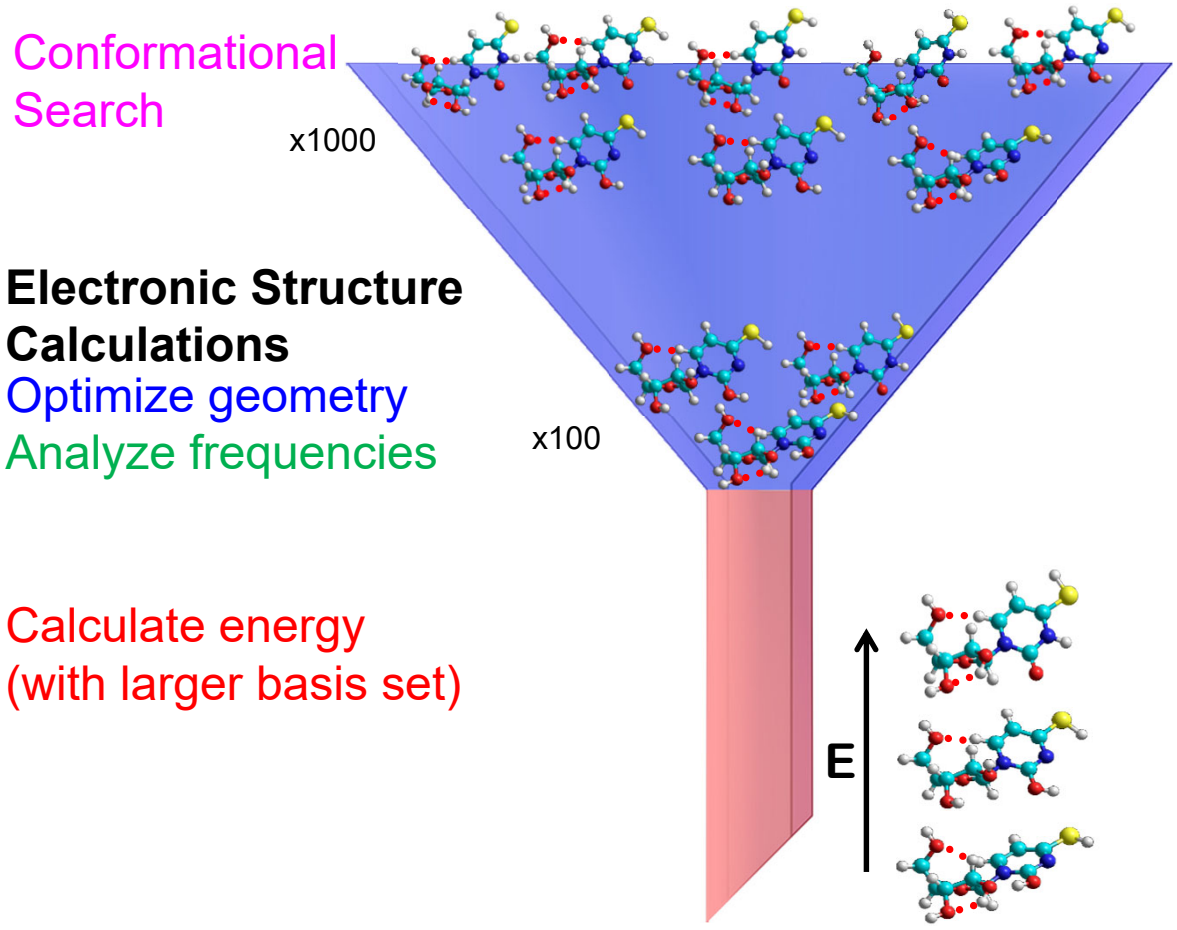


Figure 3.1 Overview of the computational chemistry calculations utilized for generating high-quality low-energy stable conformers of small molecules. A conformational search generates a wide-range of candidate conformers that are filtered to select the most representative conformers for further optimization and frequency analysis by electronic structure calculations. A final energy calculation is performed with a larger basis set to improve energetic predictions.

Figure 3.2

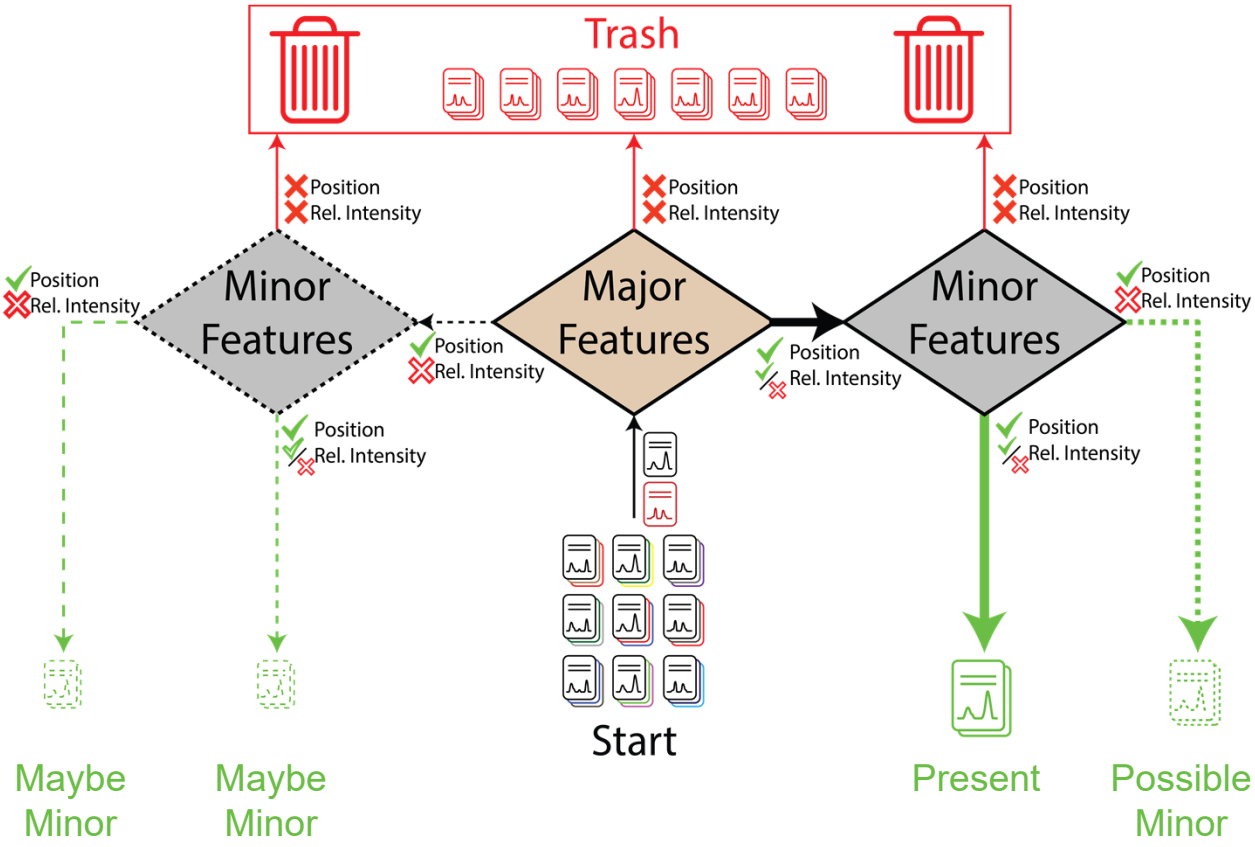


Figure 3.2 Decision tree for visual comparison and analysis of predicted IR spectra versus the experimental IRMPD spectrum to elucidate conformers present in the experiments. Conformers that agree well with the major and minor measured features are expected to be present and important to the experiments. Conformers that present disagreement with the measured IRMPD yield may be present in the experiment, pending unique features indicating their presence in the experiments. Conformers that do not agree with the major features are likely not present in the experiments.

Figure 3.3

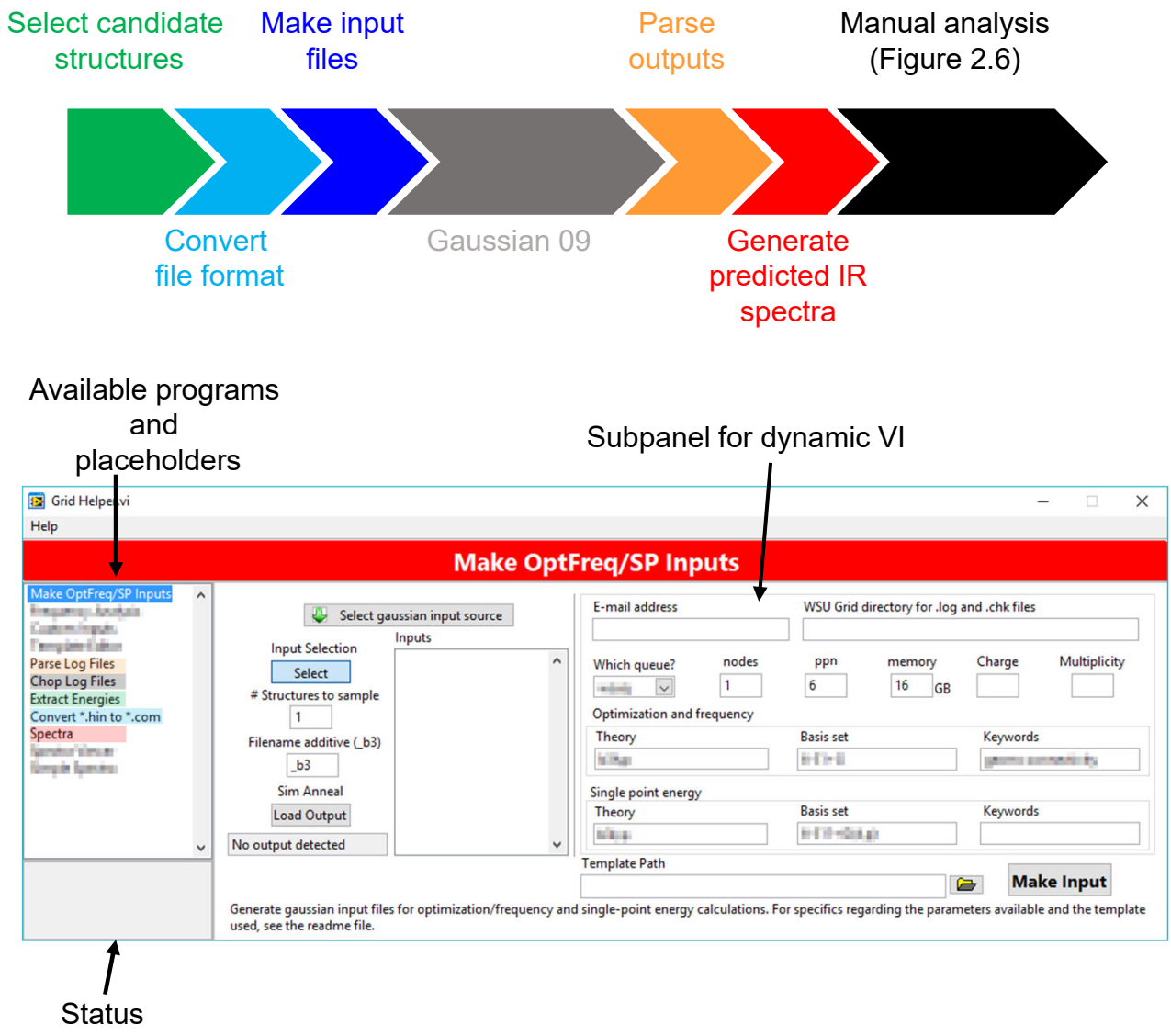


Figure 3.3 The workflow and user interface of GridHelper.vi, with Make OptFreq/SP Inputs shown as an example of the individual subVI that assist in the computational chemistry workflow. Steps in the workflow and the corresponding subVI are highlighted by the same colors.

Figure 3.4

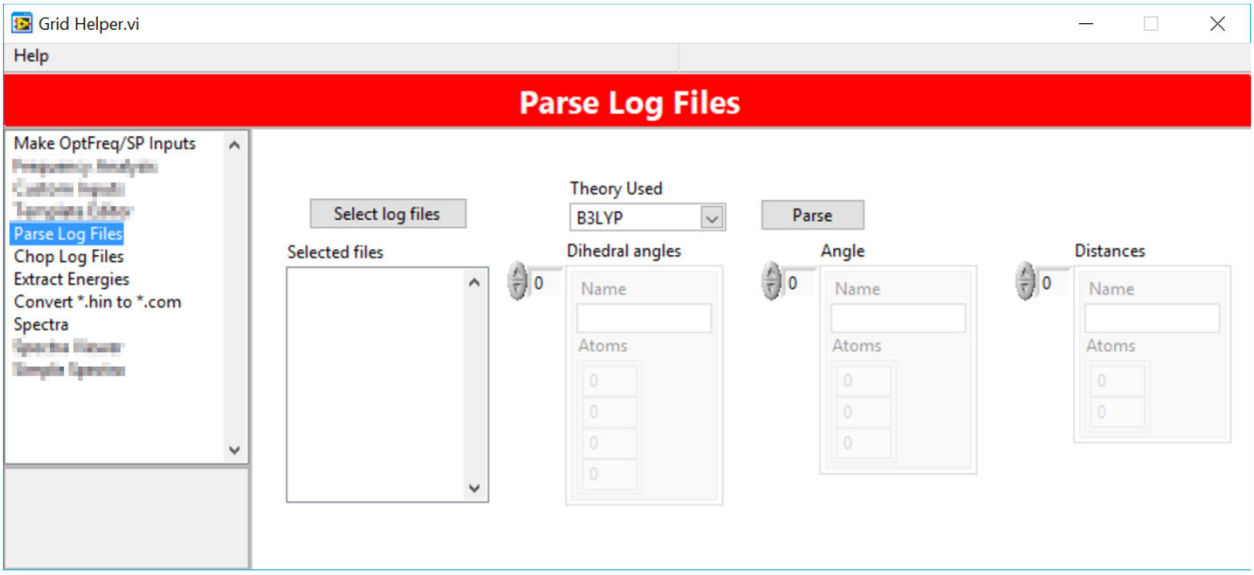


Figure 3.4 The user interface for parsing Gaussian 09 output files. Multiple files can be parsed in one operation, with desired dihedral angles, bond angles, and bond distances also extracted. The level of theory is currently limited to B3LYP and MP2 as their energetics are printed in the output file differently, but can be expanded by adding definitions to the corresponding configuration file.

Figure 3.5

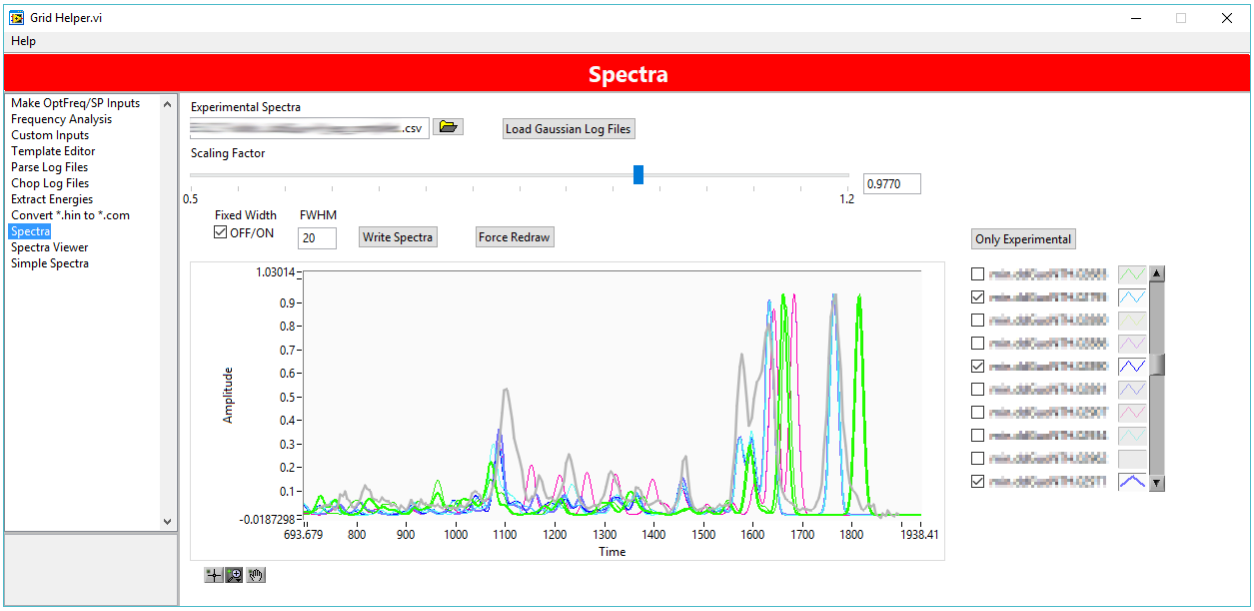


Figure 3.5 The user interface of Spectra.vi, used to generate convoluted IR spectra and begin analysis of the predicted IR spectra and experimental IRMPD spectrum. Several predicted IR spectra can be displayed simultaneously, allowing for this interface to also serve as an initial stage of data analysis.

Figure 3.6

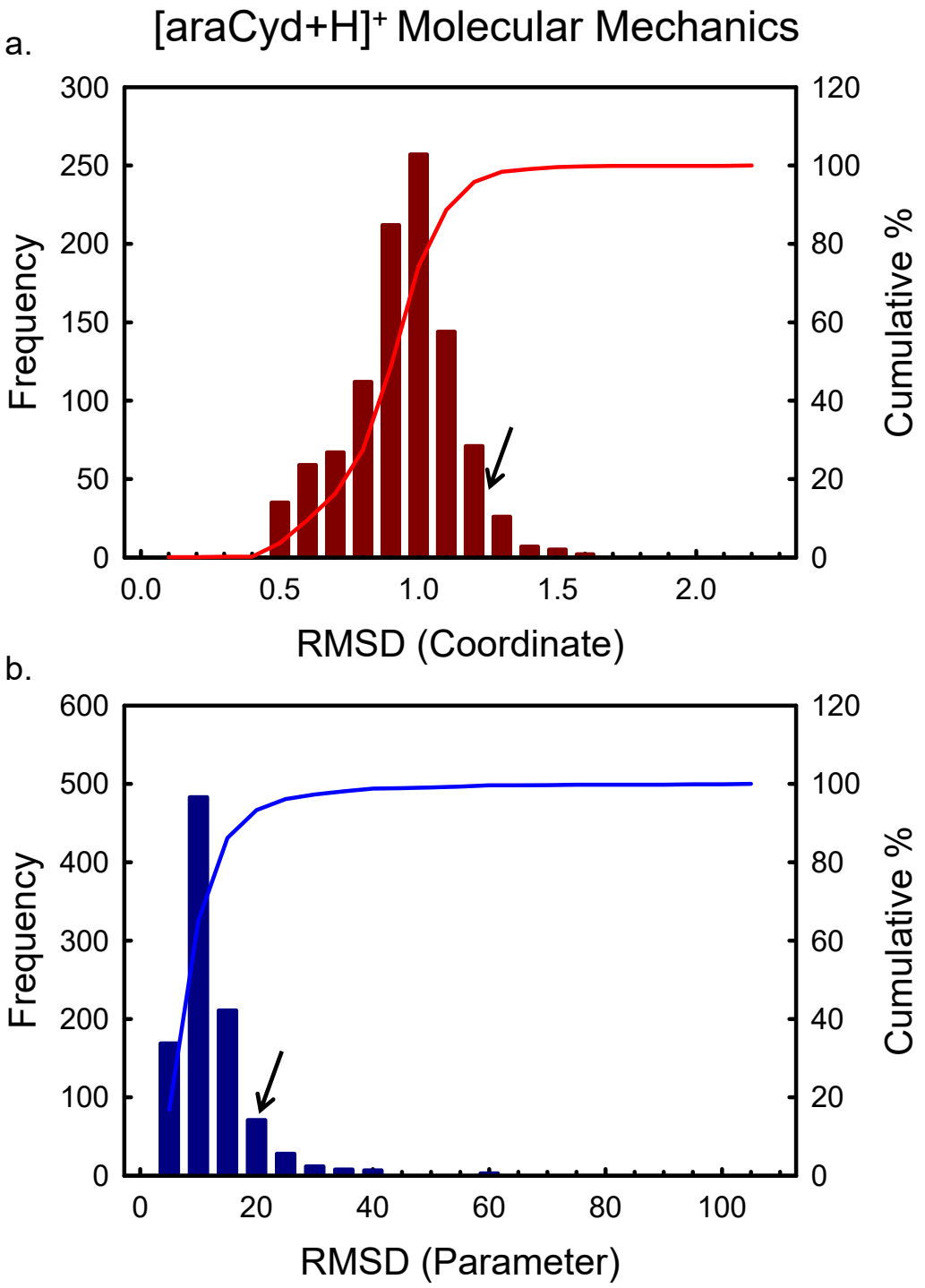


Figure 3.6 Histograms of the minimum RMSD values calculated using the Cartesian coordinate (part a) and structure parameter (part b) methods for structures of protonated cytosine arabinoside directly from the molecular mechanics simulated annealing procedure, with no explicit optimization step. The cumulative % of the histogram is also plotted. Reasonable RMSD cut-offs indicated for each with arrows.

Figure 3.7

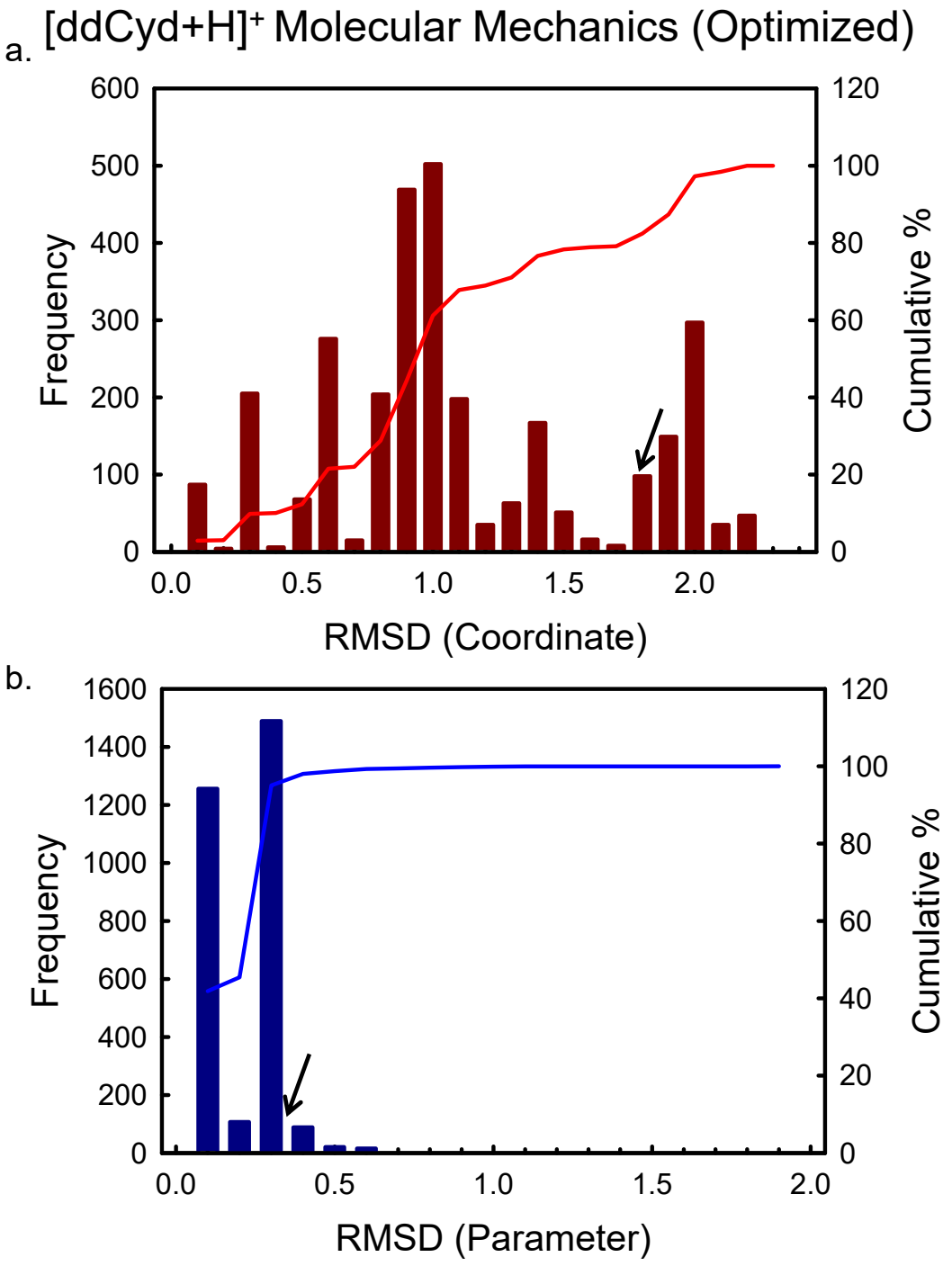


Figure 3.7 Histograms of the minimum RMSD values calculated using the Cartesian coordinate (part a) and structure parameter (part b) methods for structures of protonated 2′,3′-dideoxycytidine that were explicitly optimized via molecular mechanics after each simulated annealing cycle. The cumulative % of the histogram is also plotted. Reasonable RMSD cut-offs indicated for each with arrows.

Figure 3.8

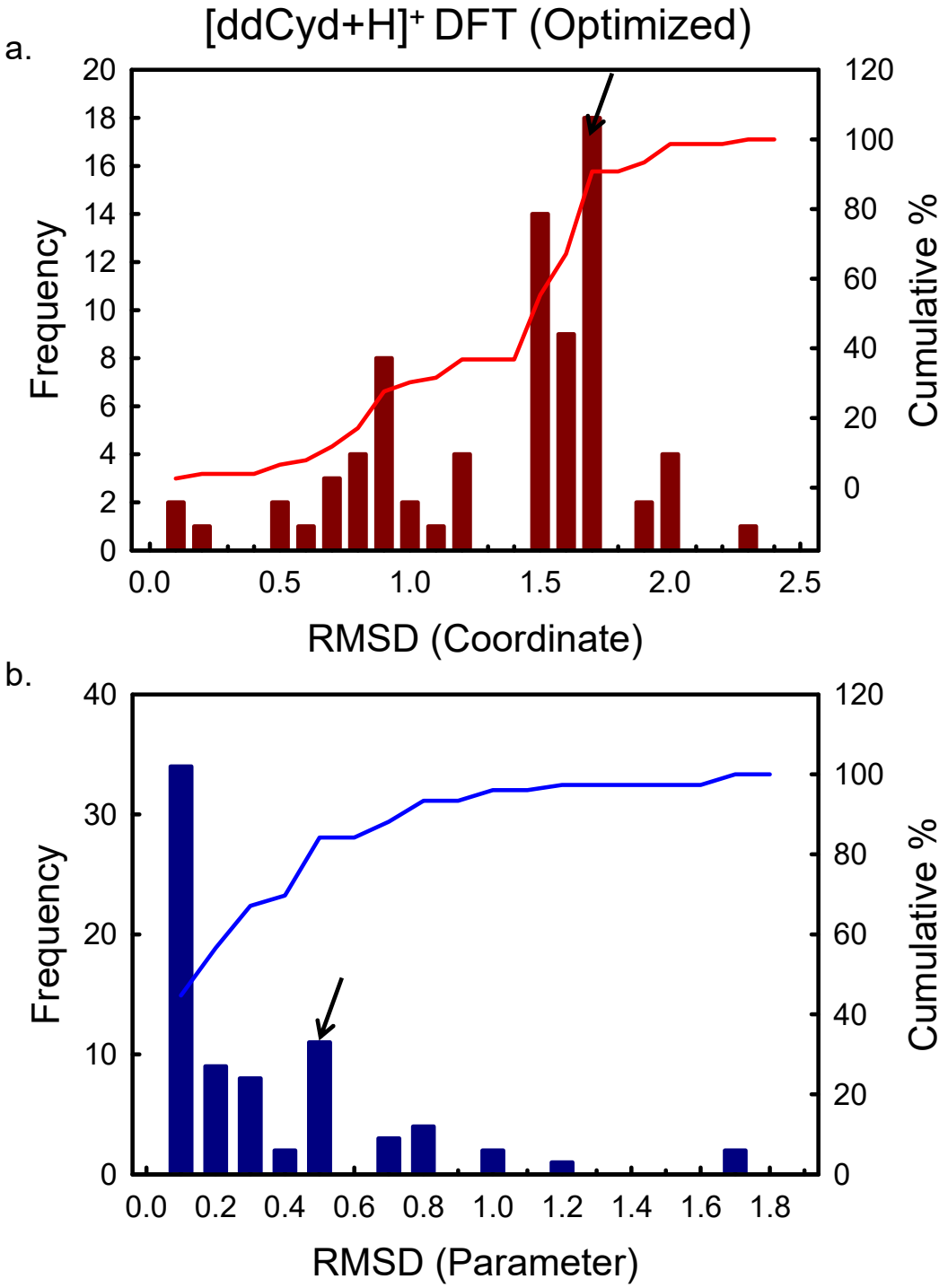


Figure 3.8 Histograms of the minimum RMSD values calculated using the Cartesian coordinate (part a) and structure parameter (part b) methods for structures of protonated 2′,3′-dideoxycytidine optimized by DFT at the B3LYP/6-31+G(d) level of theory. The cumulative % of the histogram is also plotted. Reasonable RMSD cut-offs indicated for each with arrows.

CHAPTER 4 CHARACTERIZATION OF GAS-PHASE STRUCTURES OF PROTONATED THIATED URIDINES

Portions of this chapter were reprinted with permission from reference 156: Hamlow, L. A.; Zhu, Y.; Devereaux, Zachary J.; Cunningham, N.A.; Berden, G.; Oomens, J.; Rodgers, M. T. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 2125. Copyright 2018 American Society for Mass Spectrometry.

4.1 Introduction and Methods

The motivations for studying the gas-phase conformations of protonated 2-thiouridine, $[s^2Urd+H]^+$, and 4-thiouridine, $[s^4Urd+H]^+$, are discussed in detail in **Chapter 1**. The IRMPD action spectroscopy and complimentary theoretical approaches used to examine the gas-phase conformations of $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ are discussed in detail in this chapter. Experimental IRMPD action spectra are compared to the predicted B3LYP/6-311+G(d,p) linear IR spectra for the low-energy conformers calculated at this level of theory. These comparisons between experimental IRMPD spectra and predicted IR spectra allow for determination of the conformations populated by ESI. The IRMPD experimental methods are detailed in **Chapter 2** and the theoretical methods used are detailed in **Chapter 3**. The conformational searches for $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ were performed in the Hyperchem 8.0 suite with the Amber 3 force field and no explicit optimization before low-energy stable conformers were sampled manually. The 40 most stable conformers were selected from each conformational search and these structures were supplemented by manually generated analogues of the low-energy conformers of $[Urd+H]^+$.[53]

4.2 Results

4.2.1 IRMPD Action Spectroscopy

The measured IRMPD action spectra of $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ are compared in **Figure 4.1**. While these ions are isomeric and impossible to differentiate by mass spectrometry alone, IRMPD action spectroscopy readily distinguishes these species. Perhaps the most obvious deviation between the spectra of these isomeric nucleosides is the intense peak at 1785 cm^{-1} in the spectrum of $[s^4Urd+H]^+$, a feature not observed for $[s^2Urd+H]^+$. The region between 1000 and 1500 cm^{-1} also displays a number of notably obvious differences in the measured IRMPD spectra. In the hydrogen-stretching region, many of the frequencies observed for these isomeric nucleosides are similar, but significant disparity is seen in the relative intensities. Similar disparities in the relative intensities can also be seen in the features between 1550 and 1650 cm^{-1}.

4.2.2 Theoretical Results

Relative B3LYP and MP2(full) Gibbs free energies (ΔG_{298}) and important structural parameters for all low-energy structures within ~25 kJ/mol of the ground conformers representing unique structures and each protonation site determined for $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ are listed in **Tables 4.1** and **4.2**. The nomenclature used to label these structures is based on the protonation site/tautomeric conformation, T for the 2-hydroxy-4-sulfhydryl or 2-sulfhydryl-4-hydroxy tautomers, or O2, S2, O4, or S4 for the protonated canonical species. This designation is followed by a capital letter indicating relative stability alphabetically within the family of conformers for that protonation site based upon the calculated B3LYP 298 K Gibbs free energies. Finally, a subscript S2 or S4 is used to distinguish conformers of $[s^2Urd+H]^+$ or $[s^4Urd+H]^+$, respectively.

Many of the lowest energy optimized structures exhibit C2′-*endo* or C3′-*endo* sugar puckering, although several alternate sugar puckering conformations are found among the stable structures computed, such as C1′-*exo*, C4′-*exo,* O4′-*exo,* and C3′-*exo*. However, conformers exhibiting these less common sugar puckering modes typically lie higher in free energy. Images of these structures can be seen in **Figures 4.2** and **4.3** accompanied by conformer names, relative energetics calculated at the B3LYP and MP2 levels of theory, sugar puckering mode, and nucleobase orientation.

2-Sulfhydryl-4-hydroxy tautomers are the only structures found within 20 kJ/mol of the most stable conformer found by both B3LYP and MP2 for $[s^2Urd+H]^+$. The two most stable calculated 2-sulfhydryl-4-hydroxy tautomers of $[s^2Urd+H]^+$, $\mathbf{TA_{S2}}$ and $\mathbf{TB_{S2}}$, display C2′-*endo* and C3′-*endo* sugar puckering, respectively, and are stabilized by O2′H···O3′ intramolecular hydrogen-bonding interactions. Both also display *anti* nucleobase orientations stabilized by the noncanonical C6H···O5′ hydrogen-bonding interaction between the nucleobase and sugar. Conformers of $[s^2Urd+H]^+$ protonated at the 4-position such as $\mathbf{O4A_{S2}}$ lie relatively high in energy but display a thioamide group not present in the 2-sulfhydryl-4-hydroxy tautomers. However, these conformers lie ≥23.9 kJ/mol (B3LYP) and ≥24.9 kJ/mol (MP2) higher in Gibbs energy than the 2-sulfhydryl-4-hydroxy ground conformer. The most stable O4 protonated conformers of $[s^2Urd+H]^+$ display a mixture of C3′-*endo* and C2′-*endo* sugar puckering with *anti* nucleobase orientations. The most stable conformer protonated at the 2-position, $\mathbf{S2A_{S2}}$, displays C2′-*endo* sugar puckering with an *anti* nucleobase orientation and a dual S2H···O2′H···O3′ hydrogen-bonding interaction. This provides stability through an ion-dipole interaction, but

S2A$_{S2}$ is predicted to be 21.4 kJ/mol (B3LYP) and 20.6 kJ/mol (MP2) less stable than the calculated **TA$_{S2}$** ground conformer.

The calculated conformers of $[s^4Urd+H]^+$ display a more even distribution of 2-hydroxy-4-sulfhydryl tautomers and S4 protonated conformers. MP2(full) predicts a 2-hydroxy-4-sulfhydryl ground conformer, **TA$_{S4}$**, whereas B3LYP predicts an S4 protonated ground conformer, **S4A$_{S4}$**. **TA$_{S4}$** is highly parallel to **TA$_{S2}$** of $[s^2Urd+H]^+$, displaying C2′-*endo* sugar puckering, an *anti* nucleobase orientation, and the same hydrogen-bonding interactions. Although B3LYP and MP2(full) disagree on the identity of the ground conformer, both levels of theory predict several S4 protonated conformers with 4.2 kJ/mol of the calculated ground conformer. All of these low-energy S4 protonated conformers display *anti* nucleobase orientations and C2′-*endo* or C3′-*endo* sugar puckering and the same intramolecular hydrogen-bonding interactions as the most stable 2-hydroxy-4-sulfhydryl tautomers found for $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$. The most stable O2 protonated conformer of $[s^4Urd+H]^+$, **O2A$_{S4}$**, displays a dual O2H···O2′H···O3′ hydrogen-bonding interaction parallel to that found for the analogous **S2A$_{S2}$** conformer of $[s^2Urd+H]^+$. A more detailed discussion of the theoretical results and additional low-energy conformers of $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ can be found in reference 156.

4.3 Discussion

4.3.1 $[s^2Urd+H]^+$

A mixture of structures of $[s^2Urd+H]^+$ were accessed in the experiments, and the predicted IR spectra of these are compared with the measured IRMPD action spectrum in **Figure 4.4**. The two dominant conformers, **TA$_{S2}$** and **TB$_{S2}$**, both minor tautomers, display C2′-*endo* and C3′-*endo* puckering, respectively. Both conformers have *anti*

nucleobase orientations that favor O2′H···O3′ hydrogen-bonding interactions in the sugar, and noncanonical C6H···O5′ hydrogen-bonding between the sugar and nucleobase. These two conformers display excellent agreement with the measured spectrum in both the IR fingerprint and hydrogen-stretching regions, but do not represent the small features at ~3380 or 3608 cm^{-1}. The small feature at ~3380 cm^{-1} is only represented in conformers protonated at the O4 or S2 positions. The predicted IR spectra of **O4B$_{S2}$** compared in **Figure 4.4** displays reasonable agreement with the measured spectrum, but disagreement with a few of the major features suggests that it is only a minor contributor to the measured IRMPD spectrum. The S2 protonated conformers compared in reference 151 display much greater disagreement in the IR fingerprint region, eliminating them from contribution to the measured IRMPD spectrum. Two slightly higher energy minor tautomers, **TC$_{S2}$** and **TD$_{S2}$**, are likely also present in the experiments in small quantities due to good agreement in the IR fingerprint region and their representation of the small features at ~3608 and ~3513 cm^{-1} in the hydrogen-stretching region, respectively. The structures accessed all demonstrate C2′-*endo* or C3′-*endo* sugar puckering with *anti* nucleobase orientations. More detailed spectroscopic analyses of the computed conformers and their representation of the measured IRMPD action spectrum are found in reference 151.

4.3.2 $[s^4Urd+H]^+$

Several structures of $[s^4Urd+H]^+$ were accessed in the experiments, and the predicted IR spectra for representatives of these structures are compared with the measured IRMPD action spectrum in **Figure 4.5**. A mixture of four low-energy S4 protonated conformers, **S4A$_{S4}$**, **S4B$_{S4}$**, **S4C$_{S4}$**, and **S4E$_{S4}$** are present in the experiments.

The predicted IR spectra of **S4C$_{S4}$** and **S4E$_{S4}$** are highly parallel to those of **S4A$_{S4}$** and **S4B$_{S4}$**, respectively, and are therefore not shown in **Figure 4.5**. However, their predicted IR spectra are compared in Reference 149. The most stable 2-hydroxy-4-sulfhydryl tautomer conformer, **TA$_{S4}$**, is also present in the experiments. The four S4 protonated conformers exhibit good agreement with the measured spectrum across the IR fingerprint region and hydrogen-stretching region. The two pairs of parallel S4 protonated conformers, **S4A$_{S4}$** and **S4C$_{S4}$**, and **S4B$_{S4}$** and **S4E$_{S4}$**, display two distinct spectra with the feature at ~1785 cm^{-1} shifting between them. The width of the feature at ~1785 cm^{-1} indicates that both pairs of S4 protonated conformers are present in the experiment. The small shoulder at ~1640 cm^{-1}, however, is not represented by any of the S4 protonated conformers but is a major feature of **TA$_{S4}$**. Several major features are absent from the predicted IR spectra of **TA$_{S4}$**, but it otherwise presents good agreement. The absence of these major spectral features and the relative intensity of the unique feature at ~1640 cm^{-1} indicate that **TA$_{S4}$** is present, but a minor contributor to the experiments. Both C2′-*endo* and C3′-*endo* sugar puckering are present among the five low-energy conformers present in the experiments. No structures displaying a *syn* nucleobase orientation contribute to the experiments. More detailed spectroscopic analysis and vibrational assignments can be found in reference 151.

4.3.3 Comparison to IRMPD and Theoretical Studies of Protonated Thiouracils

The measured IRMPD spectra of the protonated forms of the 2-thiouracil and 4-thiouracil nucleobases, [s^{2}Ura+H]$^+$ and [s^{4}Ura+H]$^+$,[59] in the IR fingerprint region are compared with their nucleoside counterparts in **Figure 4.6**. There are notable parallels between these spectra, most evident in the range of 1400–1900 cm^{-1}, a region largely

attributed to the vibrational modes of the nucleobase. Peak positions are reasonably well reproduced between the analogous nucleoside and nucleobase with some minor shifting. The measured features are more intense and broader in the nucleobase spectrum and have been scaled down to facilitate comparison. A previous study[59] identified the preferred protonation state of the 2-thiouracil nucleobase, $[s^2Ura+H]^+$, as the minor 2-sulfhydryl-4-hydroxy tautomer, with the most stable O4 protonated conformer lying 20.5 kJ/mol higher in free energy. Although the IR spectrum in the IR fingerprint region predicted for the O4 protonated conformer matched much of the measured $[s^2Ura+H]^+$ spectrum reasonably well, one small feature was not well represented. This disagreement and the relatively high Gibbs energy led to the conclusion that the O4 protonated conformer was not populated in the IR fingerprint experiment. Analysis of the IR fingerprint spectrum of $[s^2Urd+H]^+$ would suggest that this conclusion is also valid for the nucleoside, however the addition of spectral information in the hydrogen-stretching region provides clear evidence for a small population of an O4 protonated species.

The measured IR fingerprint region spectrum of protonated 4-thiouracil, $[s^4Ura+H]^+$, is best represented by the S4 protonated conformers calculated in previous work.[59] However, the most stable 2-hydroxy-4-sulfhydryl minor tautomers calculated display most of the major features of the measured spectra without substantial disagreement and are calculated to be only 3.0 kJ/mol higher in Gibbs energy. The absence of any distinctive spectral features attributed to the minor tautomer and the absence of one of the major features in the predicted spectrum suggests that the minor tautomer conformers might be present in the experiments. However, the lack of any distinctive features unique to the minor tautomers results in that assessment relying upon

the calculated energetics instead of the spectroscopic results. In contrast, the predicted IR spectra of the 2-hydroxy-4-sulfhydryl tautomers and O4 protonated conformers of the nucleoside $[s^4Urd+H]^+$ display more distinctive differences. This is readily observed in the IR fingerprint region where the shoulder attributed to the tautomer at ~1640 cm^{-1} is more distinct and provides clear evidence of the presence of the **TA_{S4}** conformer.

4.3.4 Comparison to IRMPD and Theoretical Studies of Protonated Canonical Uridine

The spectrum of $[Urd+H]^+$ [53] is compared with those measured for $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ in **Figure 4.6**. Many of the major features are highly parallel, with the notable exception being the red shifting of several nucleobase ring stretching modes of the thiouridines. The structures of these features are largely preserved however and major differences beyond this are attributed to changes in the protonation preferences as compared to the more competitive mixture of O4 protonated and T conformers found for $[Urd+H]^+$. Differences in the relative intensities of the features between ~1400 and 1700 cm^{-1} and that measured at ~3380 cm^{-1} are observed between the measured spectra of $[Urd+H]^+$ and $[s^2Urd+H]^+$. These differences can be attributed to a decrease in the importance of the O4 conformers in $[s^2Urd+H]^+$ versus $[Urd+H]^+$. In contrast, the much larger relative intensity of the carbonyl stretch at ~1785 cm^{-1} in $[s^4Urd+H]^+$ and decrease in the relative intensity of the features at ~1466 cm^{-1} versus $[Urd+H]^+$ reflects the change in protonation preference of $[s^4Urd+H]^+$ shifting towards S4 protonation. The sugars in all three systems exhibit a mixture of C2′-*endo* and C3′-*endo* puckering with difficulty in differentiating these features spectroscopically. A *syn* tautomer structure involving rotation of the substituent at the 2-position to hydrogen bond with O5′ was found relatively

low in free energy for $[Urd+H]^+$, but is not energetically favorable or observed spectroscopically for either protonated thiouridine.

4.3.5 Comparison of $[s^2Urd+H]^+$ to Studies of Neutral 2-Thiouridine

Previous studies of neutral 2-thiouridine by NMR,[109,165,166] crystallography,[167] and electronic structure calculations[78,168] all indicate a preference for C3′-*endo* sugar puckering. Structure characterization by NMR has shown that uridine has a 0.4 kJ/mol preference for C3′-*endo* over C2′-*endo*, while 2-thiouridine enhances that stabilization to 4.2 kJ/mol. This preference was also found in a crystallography study in which a C3′-*endo*, *anti* structure was found. The consensus of these studies was that thiation has little effect on the conformational parameters, with only minor stabilization of C3′-*endo* sugar puckering. Consistent with this conclusion, the largest differences in conformational parameters between $[Urd+H]^+$ and $[s^2Urd+H]^+$ in the gas phase are a result of protonation; the sugar puckering and nucleobase orientation are largely conserved. However, the lack of spectroscopic distinction between C2′-*endo* and C3′-*endo* puckering combined with calculated energetics that place the two sugar puckering modes very close in energy, cannot definitively confirm a preference for C3′-*endo* puckering in the protonated gas-phase species.

4.4 Conclusions

IRMPD action spectra were measured for $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ in the IR fingerprint and hydrogen-stretching regions. Synergistic calculations were used to generate candidate structures and predict their linear IR spectra. Comparisons between predicted IR spectra and measured IRMPD spectra were used to determine the conformers populated by ESI. Mixtures of O4 and S4-protonated conformers and the

2-sulfhydryl-4-hydroxy or 2-hydroxy-4-sulfhydryl minor tautomers of $[s^2Urd+H]^+$ or $[s^4Urd+H]^+$, respectively, are found experimentally. $[s^2Urd+H]^+$ displays a strong preference for 2-hydroxy-4-sulfhydryl conformers, whereas $[s^4Urd+H]^+$ demonstrates a preference for protonation at the S4-position. Structures exhibiting a mixture of C2′-*endo* and C3′-*endo* sugar puckering all with *anti* nucleobase orientations were populated in the experiments for both species.

Table 4.1 Relative free energies of stable low-energy conformers of $[s^2Urd+H]^+$ at 298 K in kJ/mol. [a]

Conformer	ΔG_{298} B3LYP	ΔG_{298} MP2	P (°)	Nucleobase Orientation	Sugar puckering
TA_{S2}	0.0	0.0	164.9	*anti*	C2′-*endo* ($^2T_3$)
TB_{S2}	2.9	2.4	19.4	*anti*	C3′-*endo* ($^3T_4$)
TC_{S2}	4.1	2.4	167.5	*anti*	C2′-*endo* ($^2T_3$)
TD_{S2}	6.0	5.7	169.2	*anti*	C2′-*endo* ($^2T_3$)
TE_{S2}	9.6	8.7	15.5	*anti*	C3′-*endo* ($^3T_2$)
TF_{S2}	14.3	19.4	136.2	*anti*	C1′-*exo* ($_1T^2$)
TG_{S2}	14.5	11.9	154.8	*syn*	C2′-*endo* ($^2T_1$)
TH_{S2}	16.6	16.7	40.1	*anti*	C4′-exo ($_4T^3$)
TI_{S2}	17.1	17.3	133.7	*syn*	C1′-*exo* ($_1T^2$)
TJ_{S2}	17.3	17.9	127.6	*syn*	C1′-*exo* ($_1T^2$)
TK_{S2}	20.1	21.2	135.9	*syn*	C1′-*exo* ($_1T^2$)
TL_{S2}	21.4	22.5	164.1	*anti*	C2′-*endo* ($^2T_3$)
$S2A_{S2}$	21.4	20.6	169.2	*anti*	C2′-*endo* ($^2T_3$)
TM_{S2}	21.6	21.9	163.7	*anti*	C2′-*endo* ($^2T_3$)
TN_{S2}	22.7	28.2	52.1	*anti*	C4′-*exo* ($_4T^3$)
TO_{S2}	23.1	30.2	132.6	*anti*	C1′-*exo* ($_1T^2$)
$O4A_{S2}$	23.9	24.9	15.5	*anti*	C3′-*endo* ($^3T_2$)
TP_{S2}	24.1	29.9	47.2	*anti*	C4′-*exo* ($_4T^3$)
$O4B_{S2}$	25.0	25.3	16.6	*anti*	C3′-*endo* ($^3T_2$)
$O4C_{S2}$	25.6	28.0	163.8	*anti*	C2′-*endo* ($^2T_3$)
$S2B_{S2}$	25.7	25.9	164.6	*anti*	C2′-*endo* ($^2T_3$)

[a] Structures and frequencies were calculated at the B3LYP/6-311+G(d,p) level of theory. Gibbs free energies were calculated at the B3LYP/6-311+G(2d,2p) and MP2(full)/6-311+G(2d,2p) levels of theory. The pseudorotation angle, nucleobase orientation, and sugar puckering for each structure are also listed.

Table 4.2 Relative free energies of stable low-energy conformers of $[s^4Urd+H]^+$ at 298 K in kJ/mol. [a]

Conformer	ΔG_{298} B3LYP	ΔG_{298} MP2	P (°)	Nucleobase Orientation	Sugar puckering
S4A$_{S4}$	0.0	1.0	165.6	*anti*	C2′-*endo* ($^2T_3$)
S4B$_{S4}$	0.5	1.1	12.5	*anti*	C3′-*endo* ($^3T_2$)
S4C$_{S4}$	1.2	2.6	144.5	*anti*	C2′-*endo* (^2_1T)
S4D$_{S4}$	1.9	2.2	14.2	*anti*	C3′-*endo* ($^3T_2$)
S4E$_{S4}$	2.1	3.1	12.4	*anti*	C3′-*endo* ($^3T_2$)
S4F$_{S4}$	2.9	0.8	196.6	*anti*	C3′-*exo* ($_3T^2$)
S4G$_{S4}$	3.2	4.1	14.3	*anti*	C3′-*endo* ($^3T_2$)
S4H$_{S4}$	4.2	2.7	196.3	*anti*	C3′-*exo* ($_3T^2$)
TA$_{S4}$	5.9	0.0	165.2	*anti*	C2′-*endo* ($^2T_3$)
S4I$_{S4}$	6.6	6.2	168.4	*anti*	C2′-*endo* ($^2T_3$)
S4J$_{S4}$	7.8	7.8	168.5	*anti*	C2′-*endo* ($^2T_3$)
TB$_{S4}$	8.1	2.2	171.6	*anti*	C2′-*endo* ($^2T_3$)
TC$_{S4}$	8.4	2.5	16.9	*anti*	C3′-*endo* ($^3T_2$)
TD$_{S4}$	11.2	3.8	167.8	*anti*	C2′-*endo* ($^2T_3$)
TE$_{S4}$	11.3	5.0	150.7	*syn*	C2′-*endo* ($^2T_1$)
TF$_{S4}$	13.2	6.6	13.2	*anti*	C3′-*endo* ($^3T_2$)
S4K$_{S4}$	15.0	21.4	137.4	*anti*	C1′-*exo* ($_1T^2$)
S4L$_{S4}$	17.0	18.4	33.6	*anti*	C3′-*endo* (^3_4T)
S4M$_{S4}$	18.2	19.9	154.2	*syn*	C2′-*endo* ($^2T_1$)
S4N$_{S4}$	22.1	20.0	51.5	*anti*	C4′-*exo* ($_4T^3$)
O2A$_{S4}$	26.2	17.2	169.5	*anti*	C2′-*endo* ($^2T_3$)
O2B$_{S4}$	26.6	17.1	153.3	*syn*	C2′-*endo* ($^2T_1$)

[a] Structures and frequencies were calculated at the B3LYP/6-311+G(d,p) level of theory. Gibbs free energies were calculated at the B3LYP/6-311+G(2d,2p) and MP2(full)/6-311+G(2d,2p) levels of theory. The pseudorotation angle, nucleobase orientation, and sugar puckering for each structure are also listed.

Figure 4.1

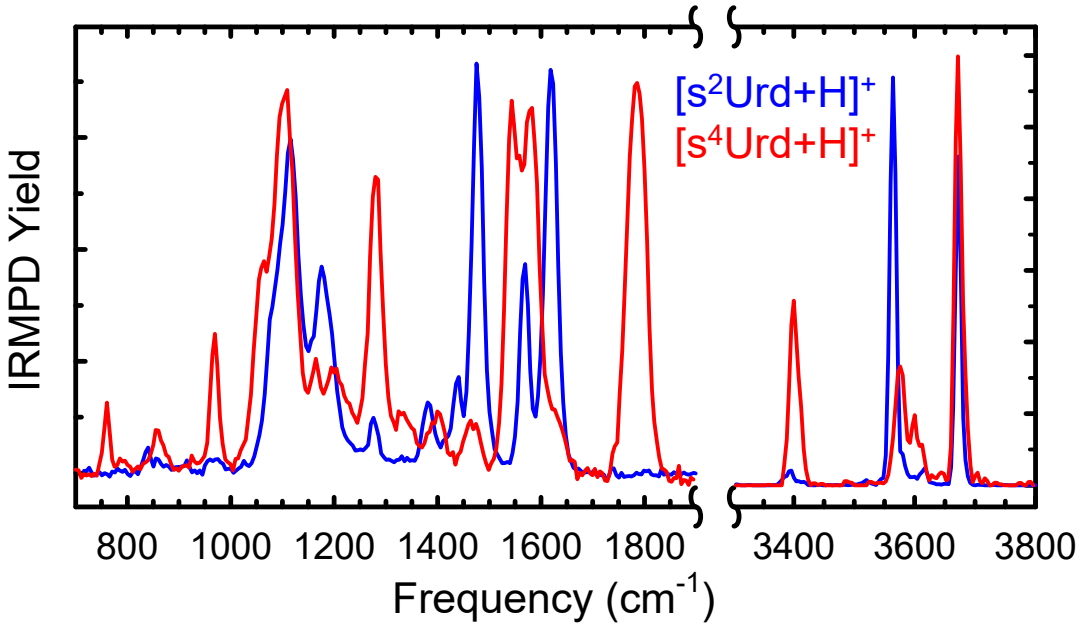


Figure 4.1 Experimental IRMPD action spectra of $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ over the IR fingerprint and hydrogen-stretching regions.

Figure 4.2

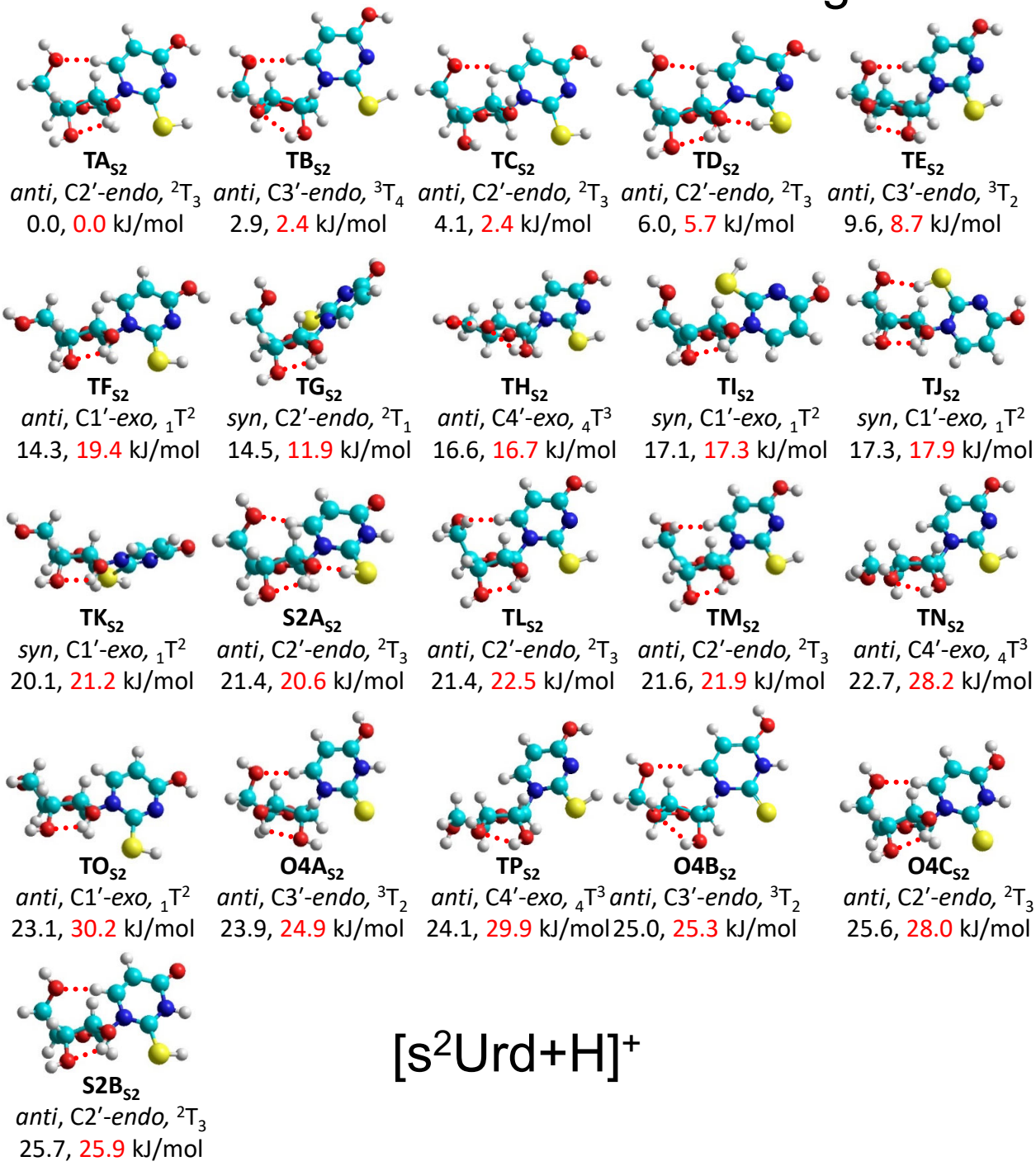


Figure 4.2 B3LYP/6-311+G(d,p) low-energy conformers of [s^{2}Urd+H]$^{+}$ with relative B3LYP/6-311+G(2d,2p) (black) and MP2/6-311+G(2d,2p) (red) Gibbs free energies at 298 K. The site of protonation or tautomeric form, orientation of the nucleobase, and specific puckering of the sugar are noted for each conformer.

Figure 4.3

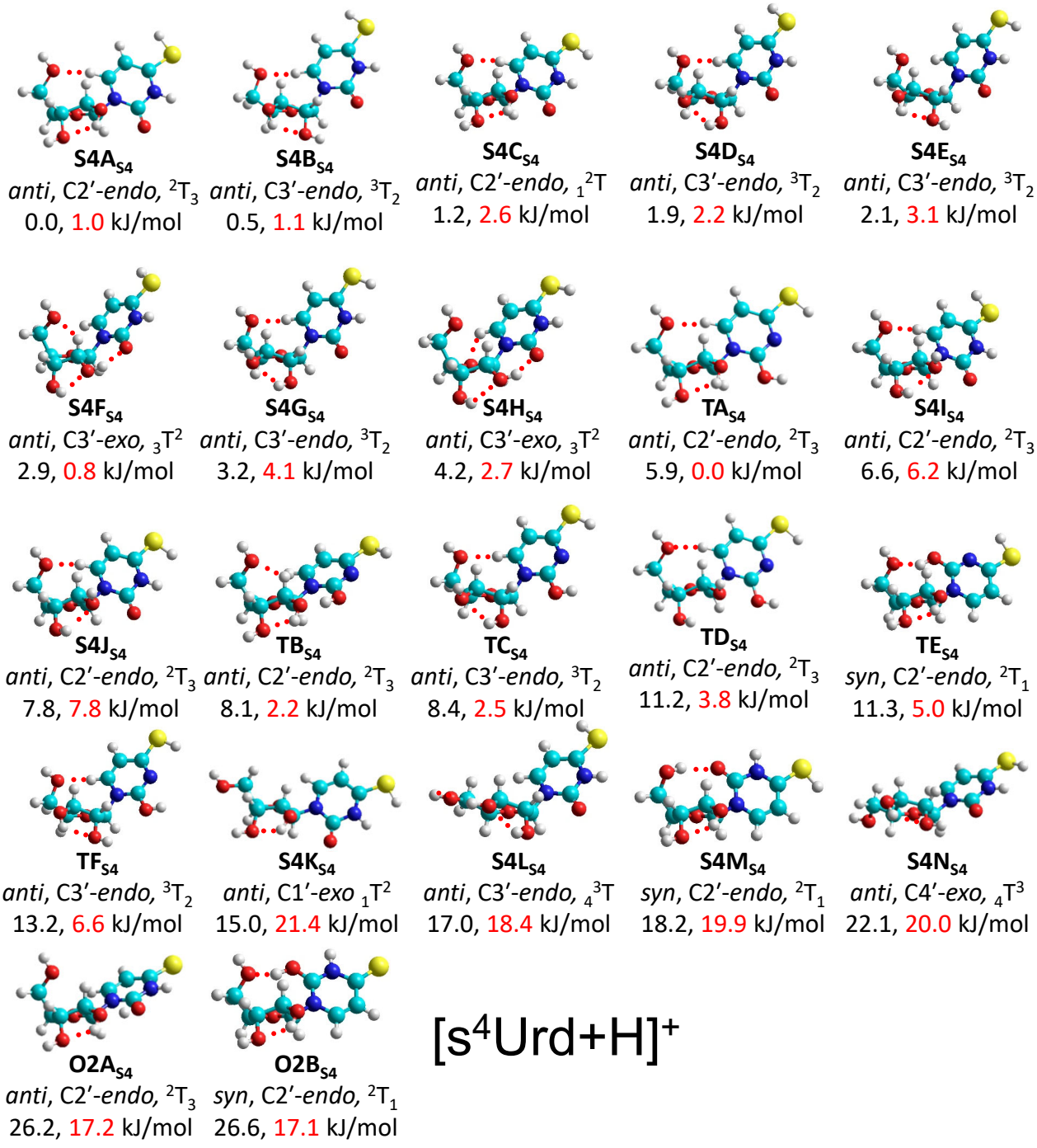


Figure 4.3 B3LYP/6-311+G(d,p) low-energy conformers of [s^{4}Urd+H]$^{+}$ with relative B3LYP/6-311+G(2d,2p) (black) and MP2/6-311+G(2d,2p) (red) Gibbs free energies at 298 K. The site of protonation or tautomeric form, orientation of the nucleobase, and specific puckering of the sugar are noted for each conformer.

Figure 4.4

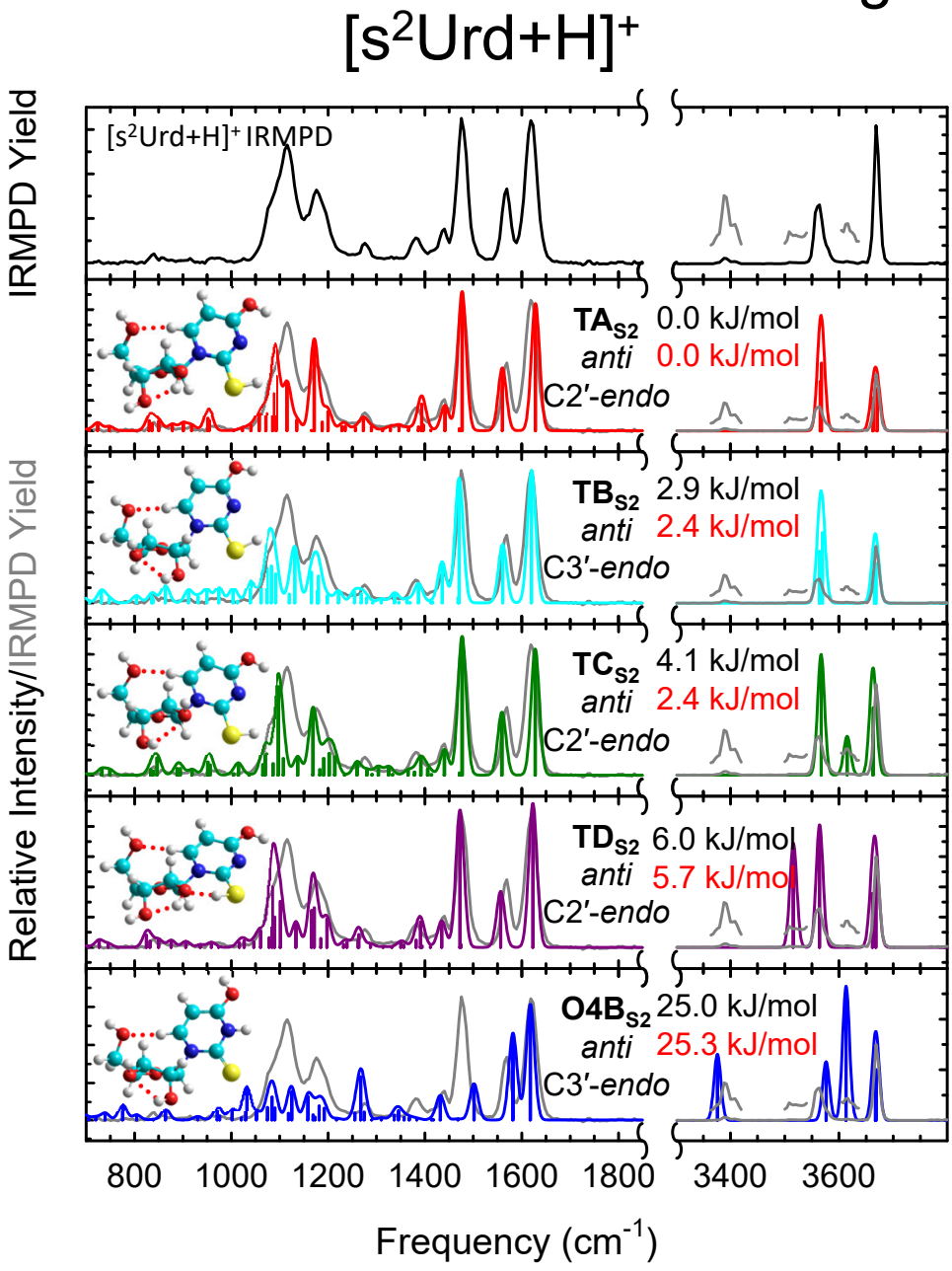


Figure 4.4 Comparison of the measured IRMPD action spectrum of $[s^2Urd+H]^+$ with the B3LYP/6-311+G(d,p) predicted linear IR spectra for low-energy conformers of $[s^2Urd+H]^+$ that are populated in the experiments. The protonation site, nucleobase orientation, sugar puckering, and relative B3LYP/6-311+G(2d,2p) (black) and MP2/6-311+G(2d,2p) (red) Gibbs free energies at 298 K are indicated. The measured IRMPD action spectrum is superimposed with the calculated spectra and the intensities scaled to facilitate comparisons. Features recollected with greater laser power are overlaid and offset from the baseline to better illustrate these small features.

Figure 4.5

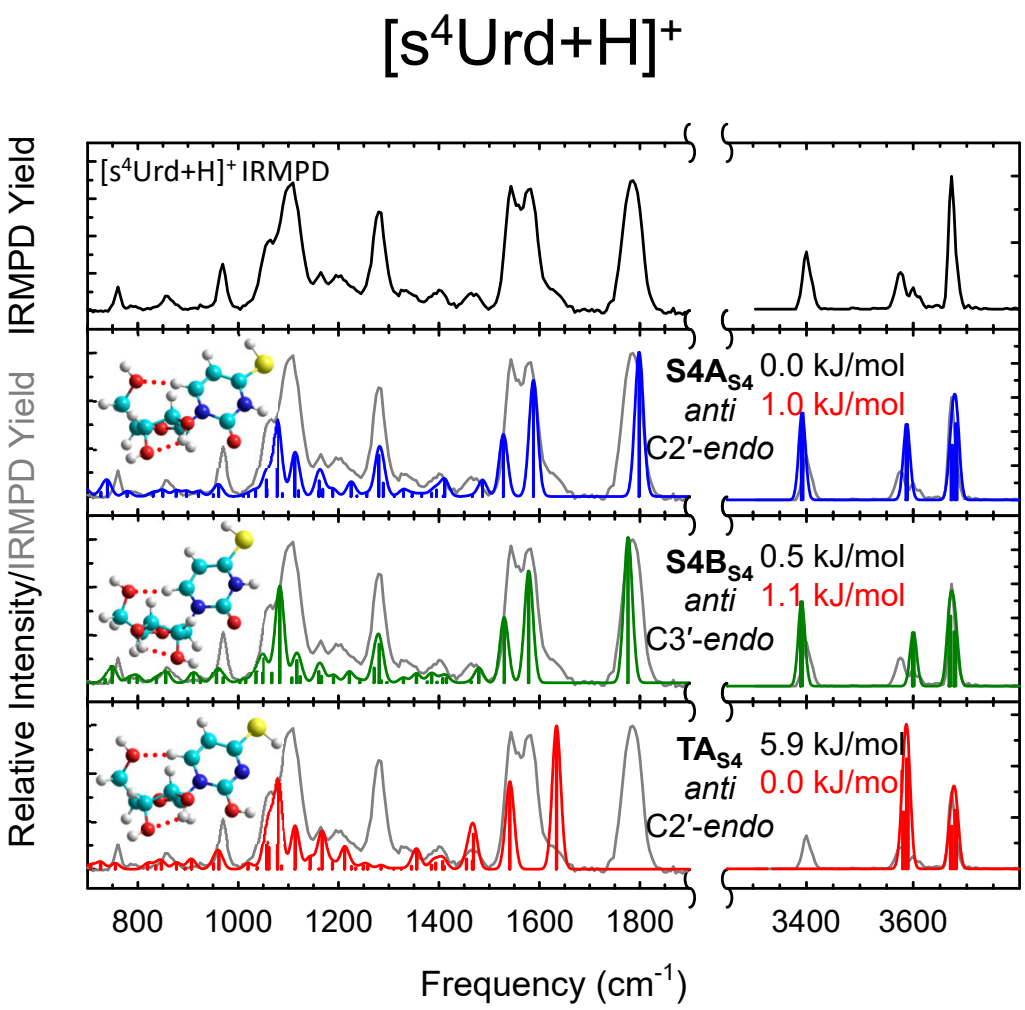


Figure 4.5 Comparison of the measured IRMPD action spectrum of $[s^4Urd+H]^+$ with the B3LYP/6-311+G(d,p) predicted linear IR spectra for low-energy conformers of $[s^4Urd+H]^+$ that are populated in the experiments. The protonation site, nucleobase orientation, sugar puckering, and relative B3LYP/6-311+G(2d,2p) (black) and MP2/6-311+G(2d,2p) (red) Gibbs free energies at 298 K are indicated. The measured IRMPD action spectrum is superimposed on the calculated spectra and the intensity scaled to facilitate comparisons.

Figure 4.6

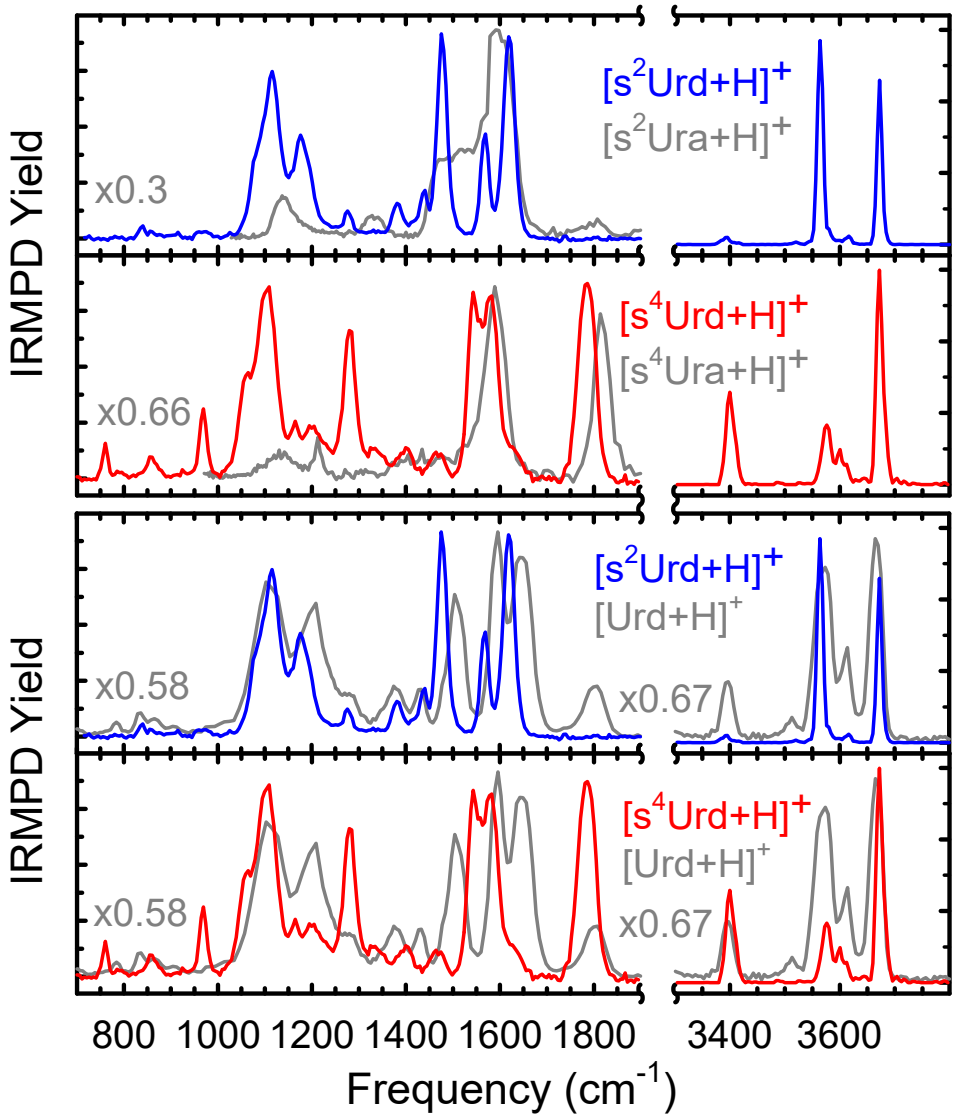


Figure 4.6 Comparison of experimental IRMPD spectra of $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ in both the IR fingerprint and hydrogen-stretching regions with spectra in the IR fingerprint region for the corresponding protonated nucleobases, $[s^2Ura+H]^+$ and $[s^4Ura+H]^+$, taken from reference 59, and over both regions for the protonated canonical nucleoside, $[Urd+H]^+$, taken from reference 53.

CHAPTER 5 IMPACT OF THE 2′- AND 3′- SUGAR HYDROXY MOIETIES ON GAS-PHASE PROTONATED NUCLEOSIDE STRUCTURE

Portions of this chapter were reprinted with permission from reference 169: Hamlow, L. A.; He, C. C.; Devereaux, Zachary J.; Roy, H. A.; Cunningham, N. A.; Soley, Erik O.; Berden, G.; Oomens, J.; Rodgers, M. T. *Int. J. Mass Spectrom.* **2019**, *438*, 124. Copyright 2019 Elsevier B.V. and reference 170: Hamlow, L. A.; Devereaux, Zachary J.; Roy, H. A.; Cunningham, N. A.; Berden, G.; Oomens, J.; Rodgers, M. T. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 832. Copyright 2019 American Society for Mass Spectrometry.

5.1 Introduction and Methods

Motivations for the extension of studies of the understanding of the intrinsic structures of gas-phase nucleosides towards nucleosides modified at the sugar moiety are discussed in detail in **Chapter 1**. Infrared multiple photon dissociation (IRMPD) action spectroscopy and complimentary theoretical calculations are used to examine the gas-phase structures of two families of protonated modified nucleoside analogues. The arabinose analogues of adenosine, $[araAdo+H]^+$, guanosine, $[araGuo+H]^+$, cytidine, $[araCyd+H]^+$, and uridine, $[araUrd+H]^+$, as well as the 2′,3′-dideoxyribose analogues of adenosine, $[ddAdo+H]^+$, guanosine, $[ddGuo+H]^+$, cytidine, $[ddCyd+H]^+$, thymidine, $[ddThd+H]^+$, and uridine, $[ddUrd+H]^+$ are examined in this work.[169,170] Comparisons between the measured IRMPD action spectra and predicted IR spectra of candidate conformers allow for elucidation of the conformations populated by electrospray ionization (ESI) in the experiments. Comparison between the arabinose and 2′,3′-dideoxyribose nucleoside analogues, as well as to the canonical ribose and 2′-deoxyribose nucleosides also provide some insight into the impact of the 2′- and 3′- sugar hydroxy moieties on

intrinsic structure. Further details on the IRMPD action spectroscopy method are discussed in **Chapter 2**, and the theoretical methods utilized are discussed in detail in **Chapter 3**.

5.2 Results

5.2.1 Theoretical Results

The calculated relative energetics and structural parameters for representative low-energy, stable conformers for all of the $[araNuo+H]^+$ and $[ddNuo+H]^+$ analogues examined in this work are listed in **Tables 5.1–5.9**. Conformers are labeled by the site of protonation or tautomeric state, followed by an alphabetical designation based upon their B3LYP/6-311+G(2d,2p) relative energetics within the set of conformers displaying that protonation site/tautomeric state.

5.2.1.1 Adenosine Analogues

Images of representative low-energy conformers of $[araAdo+H]^+$ and $[ddAdo+H]^+$ are shown in **Figures 5.1** and **5.2**. The relative energetics and structural parameters of these conformers are listed in **Tables 5.1** and **5.2**. Both $[araAdo+H]^+$ and $[ddAdo+H]^+$ display a preference for protonation at the N3-position in the gas-phase, in agreement with previous studies of the DNA and RNA analogues. Protonation at the N1-position is the next most stable, with $\mathbf{N1A_{araAdo}}$ and $\mathbf{N1A_{ddAdo}}$ lying 18.3, and 25.3 kJ/mol (B3LYP), respectively, above the ground **N3A** conformers. Protonation at the N7-position is even less favored, with $\mathbf{N7A_{araAdo}}$ and $\mathbf{N7A_{ddAdo}}$ lying 29.6 and 27.2 kJ/mol above **N3A**, respectively. However, calculations in a polarizable continuum indicate that N1 and N3 protonated conformers are much closer in relative stability in aqueous solution. Low-energy conformers protonated at the N3-position prefer to adopt conformations stabilized

by a N3H$^+$···O5′ ionic hydrogen-bonding interaction. These low-energy N3 protonated conformers prefer to adopt *syn* nucleobase orientations. In contrast, the most stable conformers protonated at the N1- or N7-positions prefer *anti* nucleobase orientations stabilized by C8H···O5′ noncanonical hydrogen-bonding interactions. The most stable conformer of [araAdo+H]$^+$, **N3A$_{araAdo}$**, prefers C4′-*exo* sugar puckering, whereas the most stable conformer of [ddAdo+H]$^+$, **N3A$_{ddAdo}$**, prefers a structurally similar C3′-*endo* sugar puckering. More detailed discussions of the calculated conformers of [araAdo+H]$^+$ and [ddAdo+H]$^+$ can be found in references 169 and 170, respectively.

5.2.1.2 Guanosine Analogues

Images of representative low-energy conformers of [araGuo+H]$^+$ and [ddGuo+H]$^+$ are shown in **Figures 5.3** and **5.4**. Structural parameters and relative energetics are listed for these conformers in **Tables 5.3** and **5.4**. Protonation is strongly preferred by both guanosine analogues at the N7-position. The **N7A$_{araGuo}$** ground conformer of [araGuo+H]$^+$ exhibits C2′-*endo* sugar puckering and is stabilized by an O2′H···O5′ hydrogen-bonding interaction. In contrast, the **N7A$_{ddGuo}$** ground conformer of [ddGuo+H]$^+$ has C3′-*endo* sugar puckering and is stabilized by a noncanonical C8H···O5′ hydrogen-bonding interaction. Both of these ground conformers have *anti* oriented nucleobases and adopt relatively similar conformations despite differences in their hydrogen-bonding stabilization. Conformers protonated at the O6- or N3-positions lie much higher in Gibbs energy. The conformers protonated at the O6-position generally display similar conformations to the N7 protonated ground conformers, whereas those protonated at N3 display *syn* nucleobase orientations more similar to the low-energy N3 protonated conformers of [araAdo+H]$^+$ and [ddAdo+H]$^+$. More detailed discussions of the conformers

calculated for $[araGuo+H]^+$ and $[ddGuo+H]^+$ can be found in references 169 and 170, respectively.

5.2.1.3 Cytidine Analogues

Images of representative low-energy conformers of $[araCyd+H]^+$ and $[ddCyd+H]^+$ are shown in **Figures 5.5** and **5.6**. Relative energetics and structural parameters for these conformers are listed in **Tables 5.5** and **5.6**. Protonation at the N3-position is energetically preferred over protonation at the O2-position at the B3LYP level of theory. However, the most stable conformers of each lie relatively close in Gibbs energy for both $[araCyd+H]^+$ and $[ddCyd+H]^+$. Similar to $[araAdo+H]^+$ and $[araGuo+H]^+$, the lowest energy N3 and O2 protonated conformers, $\mathbf{N3A}_{\mathbf{araCyd}}$ and $\mathbf{O2A}_{\mathbf{araCyd}}$, of $[araCyd+H]^+$ prefer the O2′H···O5′ hydrogen-bonding interaction, which stabilizes C2′-*endo* sugar puckering. However, conformers of $[araCyd+H]^+$ displaying noncanonical C6H···O5′ hydrogen-bonding interactions, such as $\mathbf{N3B}_{\mathbf{araCyd}}$, lie only slightly higher in Gibbs energy. The most stable N3 and O2 protonated conformers of $[ddCyd+H]^+$, $\mathbf{N3A}_{\mathbf{ddCyd}}$ and $\mathbf{O2A}_{\mathbf{ddCyd}}$, both display C3′-*endo* sugar puckering and C6H···O5′ noncanonical hydrogen-bonding interactions. Both protonated cytidine analogues prefer *anti* nucleobase orientations. The most stable conformers displaying *syn* nucleobase orientations, $\mathbf{O2J}_{\mathbf{araCyd}}$, and $\mathbf{O2C}_{\mathbf{ddCyd}}$, lie 19.9 and 17.2 kJ/mol above their respective ground conformers. More detailed discussions of the computed conformers of $[araCyd+H]^+$ and $[ddCyd+H]^+$ can be found in references 169 and 170, respectively.

5.2.1.4 Uridine Analogues

Representative low-energy conformers of $[araUrd+H]^+$ and $[ddUrd+H]^+$ are shown in **Figures 5.7** and **5.8**. The relative energetics and structural parameters for these

conformers are listed in **Tables 5.7** and **5.8**. Protonation at the O4-position is preferred energetically by both $[araUrd+H]^+$ and $[ddUrd+H]^+$, however tautomerization to the minor 2,4-dihydroxy **TA** tautomer lies only slightly higher in Gibbs energy than the **O4A** ground conformers for both systems. Protonation at the O2-position is also relatively stable, with $\mathbf{O2A_{araUrd}}$ and $\mathbf{O2A_{ddUrd}}$ lying only 9.0 and 14.8 kJ/mol higher in Gibbs energy than the calculated ground conformers of $[araUrd+H]^+$ and $[ddUrd+H]^+$, respectively. As with the other arabinose analogues, $[araUrd+H]^+$ displays a slight preference for C2′-*endo* sugar puckering over C3′-*endo* puckering. Similarly, $[ddUrd+H]^+$ displays a preference for C3′-*endo* puckering over C2′-*endo* puckering, as observed for the other 2′,3′-dideoxyribose analogues. The lowest energy O4-protonated and minor tautomer conformers prefer *anti* nucleobase orientations, but the most stable O2-protonated conformers display *syn* nucleobase orientations. Detailed discussions of the low-energy conformers of $[araUrd+H]^+$ and $[ddUrd+H]^+$ are found in references 169 and 170, respectively.

5.2.1.5 2′,3′-Dideoxythymidine

Representative low-energy conformers of $[ddThd+H]^+$ are shown in **Figure 5.9**. Relative energetics and structural parameters are listed for these conformers in **Table 5.9**. Upon protonation, tautomerization to the minor 2,4-dihydroxy tautomer $\mathbf{TA_{ddThd}}$ is preferred over protonation at the O4-position in $\mathbf{O4A_{ddThd}}$ by 2.5 kJ/mol. The most stable O2 protonated conformer, $\mathbf{O2A_{ddThd}}$, lies only 9.7 kJ/mol above $\mathbf{TA_{ddThd}}$. Both $\mathbf{TA_{ddThd}}$ and $\mathbf{O4A_{ddThd}}$ adopt *anti* nucleobase orientations stabilized by C6H···O5′ hydrogen-bonding interactions and C3′-*endo* sugar puckering. The most stable O2-protonated conformer instead adopts a *syn* nucleobase orientation stabilized by an $O2H^+\cdots O5'$ ionic hydrogen-

bonding interaction and C2′-*endo* sugar puckering. A more detailed discussion of the calculated low-energy conformers of $[ddThd+H]^+$ is found in reference 170.

5.2.2 IRMPD Action Spectroscopy

Photodissociation of each protonated modified nucleoside analogue led to loss of the protonated nucleobase via cleavage of the glycosidic bond as follows,

$$[xNuo+H]^+ + n\ hv \rightarrow [Nua+H]^+ + (xNuo\text{-}Nua) \quad (2)$$

where xNuo is the relevant modified nucleoside analogue and Nua is the corresponding nucleobase. The measured IRMPD action spectra of the protonated arabinose and 2′,3′-dideoxyribose nucleoside analogues are compared with those of the canonical DNA and RNA nucleosides in **Figures 5.10–5.14**. The predicted linear IR spectra of similar low-energy conformers are often indistinct with respect to major features, and thus resulting in several conformers identified as contributing to the measured spectrum.

5.2.2.1 Adenosine Analogues

The measured IRMPD spectrum of $[ddAdo+H]^+$ and $[araAdo+H]^+$ are displayed with those measured in previous work of $[dAdo+H]^+$ and $[Ado+H]^+$ in **Figure 5.10**.[52] The predicted linear IR spectra of the conformers primarily responsible for the experimental populations of $[Ado+H]^+$, $[araAdo+H]^+$, $[dAdo+H]^+$, and $[ddAdo+H]^+$ are also shown in **Figure 5.10**.[52] Comparison of the measured IRMPD action spectra for the adenosine analogues reveals significant parallels, with the major features above 1400 cm^{-1} in the IR fingerprint region well reproduced across all four IRMPD spectra displayed. The hydrogen-stretching region is also extremely parallel between the four adenine nucleosides, although $[araAdo+H]^+$ does display a small shoulder at ~3450 cm^{-1} that is not present for $[ddAdo+H]^+$, $[dAdo+H]^+$, and $[Ado+H]^+$. The calculated gas-phase ground

conformers, **N3A**, all provide good agreement with the measured IMRPD spectrum across both regions measured. However, the width of the major features at ~1675 cm^{-1} all indicate that some N1-protonated conformers, such as **N1A**, are likely present in the experiments. Detailed discussions of all of the conformers populated in the experiments, and the corresponding vibrational assignments for the experimental spectra of the modified adenosine analogues are provided in references 169 and 170, respectively.

5.2.2.2 Guanosine Analogues

The measured IRMPD spectrum of $[ddGuo+H]^+$ and $[araGuo+H]^+$ are compared with those measured in prior work of $[dGuo+H]^+$ and $[Guo+H]^+$ in **Figure 5.11**.[51] Predicted linear IR spectra of conformers found to be important contributors to these measured spectra are also shown. The measured IRMPD spectra of the guanosine nucleosides are extremely similar, suggesting that the conformers accessed in the experiments are also similar. This is in agreement with the theoretical results discussed previously as well the prior work examining $[dGuo+H]^+$ and $[Guo+H]^+$. The more thorough comparisons of $[araGuo+H]^+$ and $[ddGuo+H]^+$ found in References 164 and 163 indicate that N7-protonated conformers dominate the experimental population, with no evidence that higher-energy N3- or O6-protonated conformers have a measurable presence in the experiments. The **N7A** conformer of each system generally provides the best agreement with the corresponding measured IRMPD spectrum, but several low-energy N7-protonated conformers provide reasonable agreement with the measured spectrum. Vibrational assignments for $[araGuo+H]^+$ and $[ddGuo+H]^+$ are listed in references 169 and 170, respectively.

5.2.2.3 Cytidine Analogues

The measured IRMPD action spectra of $[ddCyd+H]^+$ and $[araCyd+H]^+$ are compared with those measured for $[dCyd+H]^+$ and $[Cyd+H]^+$ in previous work in **Figure 5.12**.[54] The predicted IR spectra for representative conformers of those populated in the experiments are also shown for comparison. Comparison of the measured IRMPD action spectra reveals significant parallels between the systems in both peak position and relative intensity. The thorough comparisons with the predicted IR spectra presented in References 164 and 163 reveal that both N3- and O2-protonated conformers must be present in the experiments, in agreement with the conclusions for the DNA and RNA analogues. Low-energy N3 protonated conformers, such as **N3A**, must be present in the experiment to contribute the feature measured at ~1800 cm^{-1}, attributed to the C=O stretch of the nucleobase, whereas O2 protonated conformers such as **O2A** generally predict the complex set of features between 1500 and 1700 cm^{-1} more accurately. A similar situation is observed in the hydrogen-stretching region, with neither N3- nor O2-protonated conformers provide excellent representation of the measured spectrum individually. Therefore, a combination of the low-energy N3- and O2-protonated conformers must be present in the experiments, with the relative intensities of their unique features indicating an approximately equal population of each for each system. The most stable N3- and O2-protonated conformers, **N3A** and **O2A**, respectively, are shown in **Figure 5.12** for comparison, as they provide the best overall representation of the measured IRMPD spectrum. However, several additional conformers may be present in the experiments. Detailed discussions of these conformers and those that are not

populated in the experiments for $[araCyd+H]^+$ and $[ddCyd+H]^+$ are located in references 169 and 170, respectively.

5.2.2.4 Uridine Analogues

The measured IRMPD action spectra of $[ddUrd+H]^+$ and $[araUrd+H]^+$ are compared with those measured for $[dUrd+H]^+$ and $[Urd+H]^+$ in previous work in **Figure 5.13**.[53] Conformers with predicted IR spectra that present good agreement with the measure IRMPD spectrum are also compared in **Figure 5.13**. Parallel to that observed for the cytidine analogues previously, the measured feature at ~1800 cm^{-1} is attributed to a C=O stretch, present in the O4- or O2-protonated conformers. Similar to the DNA and RNA analogues, the O2-protonated conformers of $[araUrd+H]^+$ and $[ddUrd+H]^+$ provide relatively poor representation of the measured spectrum, indicating that O4 protonated conformers must be present in the experiment to contribute this measured feature. The **O4A** conformers of both $[ddUrd+H]^+$ and $[araUrd+H]^+$ provide reasonable agreement with the measured spectrum in both the IR fingerprint and hydrogen-stretching regions, but cannot contribute every measured feature. The predicted spectra of low-energy minor tautomer conformers, such as **TA**, are highly complementary to these low-energy O4-protonated conformers, neatly contributing those features that were missing, and indicating a significant contribution to the experiment. The relative intensities of the unique measured features indicate that for both $[araUrd+H]^+$ and $[ddUrd+H]^+$ a mixture of low-energy O4-protonated and minor tautomer conformers are present in the experiments. For a more detailed discussion of the specific conformers present in the experiments for $[araUrd+H]^+$ and $[ddUrd+H]^+$ see references 169 and 170, respectively.

5.2.2.5 2′,3′-Dideoxythymidine

The measured IRMPD spectrum of $[ddThd+H]^+$ is compared with those measured in previous work for the thymidine DNA and RNA analogues and the predicted linear IR spectra of conformers important to the experiments in **Figure 5.14**.[55] Like the cytidine and uridine analogues discussed previously, the measured IRMPD spectrum of $[ddThd+H]^+$ agrees well with those measured for the DNA and RNA analogues, and has unique features that indicate the presence of at least two distinct conformers in the experiments. The calculated ground conformer, $\mathbf{TA_{ddThd}}$, provides excellent agreement with the measured IRMPD spectrum in the IR fingerprint region, and also agrees well with the hydrogen-stretching region. However, the measured feature at ~1800 cm^{-1} indicates that either an O4- or O2-protonated conformer must be present in the experiments. The low-energy O4-protonated conformers of $[ddThd+H]^+$ don't present good agreement with the measured IRMPD spectrum, whereas the low-energy O2 protonated conformers present reasonable and complimentary agreement to the minor tautomers, indicating that at least $\mathbf{O2A_{ddThd}}$ is present in the experiments. Detailed discussion of all of the conformers present in the experiments and corresponding vibrational assignments are given in reference 170.

5.3 Discussion

5.3.1 Impact of 2′ and 2′,3′-Hydroxy Substituents on Structure of Purine Nucleosides

The parameters most indicative of structure (glycosidic bond angle, 5′-hydroxy orientation, and pseudorotation angle) for all conformers calculated within 25 kJ/mol Gibbs energy of the respective ground conformer for the 2′,3′-dideoxyribose,

2′-deoxyribose, ribose, and arabinose nucleosides of the purines, adenine and guanine, are shown on polar plots in **Figure 5.15**.[170] As discussed previously but not displayed in **Figure 5.15**, the preferred protonation sites for the adenine and guanine nucleoside analogues are not altered by the presence of the 2′- or 2′,3′-hydroxys or inversion of the stereochemistry at the 2′-position. Altogether, the purine nucleosides are quite resistant to notable structural changes induced by the 2′- or 2′,3′-hydroxy substituents, as observed in the quite similar distributions of the structural parameters of low-energy conformers shown in **Figure 5.15**. The absence of the 2′-hydroxy results in the largest shift in accessible conformers by eliminating a sugar–sugar hydrogen-bonding interaction, either between the 2′- and 3′-hydroxy substituents (RNA) or the 2′- and 5′-hydroxy moieties (arabinosides). The absence of both the 2′- and 3′-hydroxy groups removes the next most accessible sugar–sugar hydrogen-bonding interaction between the 3′- and 5′-hydroxy substituents, which competes directly with the most common nucleobase–5′-hydroxy interaction. However, these changes in the intramolecular hydrogen-bonding stabilization have relatively minimal impact on the structures of the low-energy conformers calculated. This is a result of the primary intramolecular hydrogen-bonding interactions remaining consistent between the 2′,3′-dideoxy and the DNA, and RNA nucleosides, $N3H^{+}\cdots O5'$ for adenine analogues and $N7H\cdots O5'$ for guanine analogues, resulting in largely conserved low-energy conformations. In the absence of both the 2′- and 3′-hydroxys a shift in the calculated sugar puckering preference, supported by some spectroscopic evidence, is observed for $[ddAdo+H]^{+}$, from C2′-*endo* in the DNA and RNA analogues, to C3′-*endo*. Although $[ddGuo+H]^{+}$ also prefers C3′-*endo* sugar puckering, the DNA and RNA analogues also demonstrated this preference, so this preference is likely driven by the

nucleobase identity rather than the 2′- or 3′- sugar hydroxy moieties. However, $[araGuo+H]^+$ displays a preference for C2′-*endo* sugar puckering due to the unique O2′H···O5′ sugar–sugar hydrogen-bonding interaction. Further discussion of **Figure 5.15** and the impact of the 2′- and 2′,3′-hydroxys on the structures of purine nucleosides can be found in reference 170. The impact of inverting the 2′-stereochemistry of the purine nucleosides is discussed in more detail in reference 169.

5.3.2 Impact of 2′ and 2′,3′-Hydroxys on Structure of Pyrimidine Nucleosides

Structural parameters for the calculated conformers of the protonated 2′,3′-dideoxyribose, 2′-deoxyribose, ribose, and arabinose nucleosides of the pyrimidine nucleobases cytosine, thymine, and uracil with relative stability within 25 kJ/mol Gibbs energy of their respective ground conformers are shown in **Figure 5.16**.[170] The most notable change as a function of 2′- or 2′,3′-hydroxy moiety is the shift in protonation preference of the uracil nucleosides from the minor tautomer for $[Urd+H]^+$ to O4-protonation in $[araUrd+H]^+$, $[dUrd+H]^+$, and $[ddUrd+H]^+$. However, this change has relatively little impact on the structural parameters discussed here as the resulting rearrangement of relative stabilities is well within the stability range displayed in **Figure 5.16**. Unlike the protonated purine nucleosides, the protonated pyrimidine nucleosides generally adopt highly parallel conformations due to the prevalence of the C6H···O5′ and $O2H^{(+)}$···O5′ nucleobase–sugar hydrogen-bonding interactions and protonation modes that facilitate the latter. The presence of the 2′-hydroxy generally results in a greater number of similar conformers, due to the additional sugar-sugar hydrogen-bonding interactions, O2′H···O3′ and O3′H···O2′ in $[Nuo+H]^+$ and O2′H···O5′ and O5′H···O2′ in $[araNuo+H]^+$. Both $[Nuo+H]^+$ and $[araNuo+H]^+$ also offer unique sugar-nucleobase

hydrogen-bonding interactions through O2′. Absence of the 2′-hydroxy greatly reduces the number of low-energy conformers available by removing all of these potential hydrogen-bonding interactions, but sugar-sugar hydrogen-bonding is still available through O3′ and O5′. The absence of both 2′- and 3′-hydroxys shifts sugar puckering preference from C2′-*endo* to C3′-*endo*, and although there is spectroscopic evidence to support this shift experimentally, low-energy conformers with each sugar pucker generally display only small differences in minor features. More detailed discussions regarding the impact of the 2′- and 2′,3′-hydroxy moieties on nucleoside structure are provided in reference 170. The impact of 2′-stereochemistry inversion is discussed in greater detail in reference 169.

5.4 Conclusions

Complimentary theoretical calculations and IRMPD action spectroscopy of the protonated arabinosides and 2′,3′-dideoxynucleosides indicate that protonation of each nucleoside analogue generally corresponds to that found for the DNA and RNA forms. The notable exception to this is $[Urd+H]^+$, which displays an energetic preference for the minor tautomer, instead of O4-protonation as observed for $[araUrd+H]^+$, $[dUrd+H]^+$, and $[ddUrd+H]^+$. The presence of the 2′-hydroxy moiety above the plane of the sugar ring alters the preferred sugar puckering mode for $[araAdo+H]^+$ and $[araGuo+H]^+$ to C4′-*exo* and C2′-*endo* respectively, versus the C2′-*endo* and C3′-*endo* sugar puckerings preferred by their DNA and RNA counterparts. The absence of the 2′,3′-hydroxy moieties shifts the energetic preference for sugar puckering to C3′-*endo* for all of the nucleoside analogues examined. The theoretical calculations and IRMPD action spectroscopy studies of the protonated arabinosides indicate the presence of an alternative mode of hydrogen-

bonding not found in their canonical analogues via an O2′H···O5′ interaction in at least one of the conformers populated in the experiments for each of the protonated arabinosides. These results indicate that sugar–sugar stabilizing interactions are in direct competition with nucleobase–sugar interactions, revealing that stabilizing the flexible sugar puckering is competitive, and in all systems studied except $[araAdo+H]^+$, provides greater overall stability than stabilizing the nucleobase orientation about the glycosidic bond. The absence of the 2′,3′-hydroxy moieties in $[ddNuo+H]^+$ doesn't significantly change the structures populated experimentally vs the DNA and RNA analogues. However, conformers of the pyrimidine $[ddNuo+H]^+$ analogues were calculated relatively low in Gibbs energy with *trans* 5′-hydroxy moieties, which was not observed within the calculated conformers of the DNA and RNA pyrimidine analogues. These conformers are not thought to be populated in these experiments, but their presence among the low-energy conformers computed indicates a greater relative stability for the *trans* 5′-hydroxy moiety in the absence of other sugar–sugar hydrogen-bonding interactions. Inversion of the stereochemistry at the 2′-position provides a greater impact, introducing a unique sugar–sugar hydrogen-bonding interaction that directly competes with the nucleobase–sugar interactions present in the other analogues, and displays a generally stronger preference for C2′-*endo* sugar puckering. Stereochemical inversion at the 2′-position produces a more notably different set of structural parameters than the absence of the 2′- or 2′,3′-hydroxy moieties, but also helps demonstrate the largely conserved preferences in intrinsic structure, primarily the site of protonation, nucleobase orientation, and intramolecular hydrogen-bonding interactions between nucleosides that differ by the 2′- and 3′- sugar hydroxy moieties.

Table 5.1 Relative Gibbs energies and structural parameters of representative low-energy conformers of $[araAdo+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
N3A	0	33.4	36.5	298.9	*syn*	C4'-*exo* ($_4T^3$)
N3B	1.2	43.4	138.7	300.5	*syn*	C1'-*exo* ($_1T^2$)
N3C	2.7	39.9	162.1	306.8	*syn*	C2'-*endo* ($^2T_3$)
N3D	3.8	90.7	144.4	289.1	*anti*	C2'-*endo* ($^2T_1$)
N3F	7.8	41.4	151.2	301.5	*syn*	C2'-*endo* ($^2T_1$)
N1A	18.3	216.8	152.6	294.1	*anti*	C2'-*endo* ($^2T_1$)
N3I	18.8	54.1	135.2	178.9	*syn*	C1'-*exo* ($_1T^2$)
N3J	19.4	340.9	346	39.5	*syn*	C2'-*exo* ($_2T^3$)
N1C	21.4	208.6	7.5	293.3	*anti*	C3'-*endo* ($^3T_2$)
N3K	23.1	179	0.2	188.1	*anti*	C3'-*endo* ($^3T_2$)
N3M	24.9	184.5	352.6	296.5	*anti*	C2'-*exo* ($_2T^3$)
N3N	29.2	52.5	134.6	63.3	*syn*	C1'-*exo* ($_1T^2$)
N3O	29.4	330	169.3	295.3	*syn*	C2'-*endo* ($^2T_3$)
N3P	29.5	325.1	3.3	187.5	*syn*	C3'-*endo* ($^3T_2$)
N7A	29.6	200.6	5.9	297.4	*anti*	C3'-*endo* ($^3T_2$)
N1E	29.9	225.3	39.6	189.4	*anti*	C4'-*exo* ($_4T^3$)
N3Q	30.5	12.2	163.6	40	*syn*	C2'-*endo* ($^2T_3$)
N7B	30.8	213.8	151.6	294.9	*anti*	C2'-*endo* ($^2T_1$)

[a] Structures and frequencies were calculated with the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C4), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering (P) are also listed.

Table 5.2 Relative Gibbs energies and structural parameters of low-energy conformers of $[ddAdo+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
N3A	0.0	28.9	35.4	298.3	*syn*	C3'-*endo* ($^3T_4$)
N3B	0.4	47.0	150.2	298.1	*syn*	C2'-*endo* ($^2T_1$)
N3C	10.0	18.0	175.1	41.3	*syn*	C2'-*endo* ($^2T_3$)
N3D	11.6	11.7	174.0	43.6	*syn*	C2'-*endo* ($^2T_3$)
N3E	14.5	342.7	348.3	39.3	*syn*	C2'-*exo* ($_2T^3$)
N3F	19.1	333.1	195.5	296.6	*syn*	C3'-*exo* ($_3T^2$)
N1A	24.0	211.7	7.8	292.2	*anti*	C3'-*endo* ($^3T_2$)
N1B	25.8	239.7	179.8	292.7	*anti*	C2'-*endo* ($^2T_3$)
N7A	27.2	199.7	5.3	296.1	*anti*	C3'-*endo* ($^3T_2$)
N3G	28.4	329.6	11.7	178.9	*syn*	C3'-*endo* ($^3T_2$)
N3H	29.0	330.4	171.3	175.5	*syn*	C2'-*endo* ($^2T_3$)
N3I	31.0	212.2	8.2	292.1	*anti*	C3'-*endo* ($^3T_2$)
N3J	31.6	248.9	183.0	291.4	*anti*	C3'-*exo* ($_3T^2$)
N7B	35.3	223.7	179.8	298.2	*anti*	C2'-*endo* ($^2T_3$)
N3K	35.8	174.5	202.6	293.9	*anti*	C3'-*exo* ($_3T^4$)
N1C	36.7	189.7	16.4	178.9	*anti*	C3'-*endo* ($^3T_2$)

[a] Structures and frequencies were calculated at the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C4), pseudorotation angle (P), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering are also listed.

Table 5.3 Relative Gibbs energies and structural parameters of low-energy conformers of $[araGuo+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
N7A	0	213.1	151	295.4	*anti*	C2'-*endo* ($^2T_1$)
N7B	1.7	213.2	154.2	295.8	*anti*	C2'-*endo* ($^2T_1$)
N7C	3.1	211.7	9.2	293.4	*anti*	C3'-*endo* ($^3T_2$)
N7D	4.2	204.6	6.7	296.3	*anti*	C3'-*endo* ($^3T_2$)
N7E	9.7	211.5	7.8	295.6	*anti*	C3'-*endo* ($^3T_2$)
N7F	10.7	222.9	159.4	297.7	*anti*	C2'-*endo* ($^2T_1$)
N7G	13.8	227.1	142.6	177.4	*anti*	C1'-*exo* ($_1T^2$)
N7H	16.2	234.3	79	188.5	*anti*	O4'-*endo* ($^OT_4$)
N7I	19.8	197.5	203.8	40.6	*anti*	C3'-*exo* ($_3T^4$)
N7K	23.6	232	135.2	63.6	*anti*	C1'-*exo* ($_1T^2$)
N7L	24.3	230.4	146.3	64	*anti*	C2'-*endo* ($^2T_1$)
N7M	29	45.1	159.5	301	*syn*	C2'-*endo* ($^2T_1$)
N7N	30	45.2	161.6	301	*syn*	C2'-*endo* ($^2T_1$)
O6A	38.2	219.2	153.3	294.2	*anti*	C2'-*endo* ($^2T_1$)
O6C	43.3	213.9	10.2	290.7	*anti*	C3'-*endo* ($^3T_2$)
N3A	47	38.8	156.5	307.4	*syn*	C2'-*endo* ($^2T_1$)

[a] Structures and frequencies were calculated with the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C4), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering (P) are also listed.

Table 5.4 Relative Gibbs energies and structural parameters for low-energy conformers of $[ddGuo+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
N7A	0.0	202.1	6.0	295.3	*anti*	C3'-*endo* ($^3T_2$)
N7B	7.6	228.1	179.5	296.9	*anti*	C2'-*endo* ($^2T_3$)
N7C	16.8	194.8	211.0	42.8	*anti*	C3'-*exo* ($_3T^4$)
N7D	17.4	192.9	209.5	44.6	*anti*	C3'-*exo* ($_3T^4$)
N7E	20.9	187.0	19.9	179.6	*anti*	C3'-*endo* ($^3T_4$)
N7F	21.1	184.1	16.4	61.7	*anti*	C3'-*endo* ($^3T_2$)
N7G	23.5	187.0	162.2	175.7	*anti*	C2'-*endo* ($^2T_3$)
N7H	25.0	189.0	20.2	67.1	*anti*	C3'-*endo* ($^3T_4$)
N7I	31.3	57.1	157.0	290.9	*syn*	C2'-*endo* ($^2T_1$)
N7J	33.6	76.1	34.3	180.9	*syn*	C3'-*endo* ($^3T_4$)
N7K	35.5	76.8	39.0	69.5	*syn*	C4'-*exo* ($_4T^3$)
N7L	36.3	74.2	38.0	66.6	*syn*	C4'-*exo* ($_4T^3$)
N7M	38.6	71.7	169.8	179.0	*syn*	C2'-*endo* ($^2T_3$)
O6A	44.2	213.6	8.1	291.9	*anti*	C3'-*endo* ($^3T_2$)
O6B	44.3	213.6	8.0	291.9	*anti*	C3'-*endo* ($^3T_2$)
N3A	54.1	44.1	145.3	299.4	*syn*	C2'-*endo* ($^2T_3$)

[a] Structures and frequencies were calculated at the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C4), pseudorotation angle (P), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering are also listed.

Table 5.5 Relative Gibbs energies and structural parameters of low-energy conformers of $[araCyd+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
N3A	0.0	207.2	150.6	295.6	*anti*	C2'-*endo* ($^2T_1$)
N3B	2.8	200.3	12.4	297.2	*anti*	C3'-*endo* ($^3T_2$)
N3C	3.3	209.6	153.1	295.9	*anti*	C2'-*endo* ($^2T_1$)
O2A	5.4	207.4	152.3	295.8	*anti*	C2'-*endo* ($^2T_1$)
O2B	7.7	207.6	155.3	296.2	*anti*	C2'-*endo* ($^2T_1$)
O2C	8.5	204.7	14.5	295.4	*anti*	C3'-*endo* ($^3T_2$)
N3G	13.3	210.7	153.6	295.6	*anti*	C2'-*endo* ($^2T_1$)
N3H	13.4	201.6	147.3	176.8	*anti*	C2'-*endo* ($^2T_1$)
N3I	15.0	204.1	61.7	190.0	*anti*	C4'-*exo* ($_4T^O$)
N3J	15.9	198.5	153.4	181.1	*anti*	C2'-*endo* ($^2T_1$)
N3K	18.2	191.7	200.8	43.1	*anti*	C3'-*exo* ($_3T^4$)
O2F	18.4	204.1	148.9	177.2	*anti*	C2'-*endo* ($^2T_1$)
O2G	18.9	218.4	151.1	297.1	*anti*	C2'-*endo* ($^2T_1$)
N3L	19.0	191.1	199.3	46.1	*anti*	C3'-*exo* ($_3T^4$)
O2H	19.1	202.0	47.9	189.9	*anti*	C4'-*exo* ($_4T^3$)
O2J	19.9	77.0	143.6	287.3	*syn*	C1'-*exo* ($_1T^2$)
N3P	22.1	190.6	143.2	282.7	*anti*	C1'-*exo* ($_1T^2$)
N3Q	22.2	195.5	72.3	62.6	*anti*	O4'-*endo* ($^OT_4$)

[a] Structures and frequencies were calculated with the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C2), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering (P) are also listed.

Table 5.5.6 Relative Gibbs energies and structural parameters of low-energy conformers of $[ddCyd+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
N3A	0.0	199.3	10.4	296.1	*anti*	C3'-*endo* ($^3T_2$)
O2A	4.8	199.4	11.5	295.3	*anti*	C3'-*endo* ($^3T_2$)
N3B	6.8	224.9	169.9	296.8	*anti*	C2'-*endo* ($^2T_3$)
O2B	10.9	225.1	169.6	295.5	*anti*	C2'-*endo* ($^2T_3$)
N3C	14.2	192.4	206.5	43.9	*anti*	C3'-*exo* ($_3T^4$)
O2C	17.2	56.9	151.4	299.7	*syn*	C2'-*endo* ($^2T_1$)
N3D	17.6	189.1	18.9	61.8	*anti*	C3'-*endo* ($^3T_4$)
N3E	18.2	192.3	21.2	179.7	*anti*	C3'-*endo* ($^3T_4$)
O2D	20.0	193.4	204.5	43.9	*anti*	C3'-*exo* ($_3T^4$)
N3F	21.2	190.4	166.9	176.2	*anti*	C2'-*endo* ($^2T_3$)
O2E	21.8	189.4	18.6	62.6	*anti*	C3'-*endo* ($^3T_4$)
O2F	22.4	193.2	20.5	179.5	*anti*	C3'-*endo* ($^3T_4$)
N3G	28.4	71.7	45.3	293.8	*syn*	C4'-*exo* ($_4T^3$)
O2G	28.9	339.0	205.3	297.0	*syn*	C3'-*exo* ($_3T^4$)
O2H	28.9	320.0	335.7	297.2	*syn*	C2'-*exo* ($_2T^1$)
N3H	30.6	69.6	41.3	68.5	*syn*	C4'-*exo* ($_4T^3$)

[a] Structures and frequencies were calculated at the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C4), pseudorotation angle (P), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering are also listed.

Table 5.5.7 Relative Gibbs energies and structural parameters of low-energy conformers of $[araUrd+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
O4A	0.0	206.5	150.4	295.8	*anti*	C2'-*endo* ($^2T_1$)
O4B	2.2	200.3	12.3	297.6	*anti*	C3'-*endo* ($^3T_2$)
O4C	3.2	208.9	152.9	296.0	*anti*	C2'-*endo* ($^2T_1$)
TA	3.9	207.6	157.1	296.7	*anti*	C2'-*endo* ($^2T_1$)
TB	4.6	209.1	16.7	293.6	*anti*	C3'-*endo* ($^3T_2$)
O2A	9.0	84.1	143.7	288.7	*syn*	C1'-*exo* ($_1T^2$)
O4D	10.9	208.4	17.6	297.1	*anti*	C3'-*endo* ($^3T_2$)
O4E	11.3	206.3	151.7	296.0	*anti*	C2'-*endo* ($^2T_1$)
TC	11.4	206.2	15.9	296.2	*anti*	C3'-*endo* ($^3T_2$)
TD	12.1	82.4	143.8	287.9	*syn*	C1'-*exo* ($_1T^2$)
O4F	14.2	200.7	12.1	297.3	*anti*	C3'-*endo* ($^3T_2$)
O4G	14.6	208.5	154.4	296.3	*anti*	C2'-*endo* ($^2T_1$)
O4H	16.0	208.6	17.6	294.3	*anti*	C3'-*endo* ($^3T_2$)
O4I	16.1	200.0	64.0	190.0	*anti*	C4'-*exo* ($_4T^O$)
TE	16.3	203.7	44.9	190.0	*anti*	C4'-*exo* ($_4T^3$)
TF	16.3	201.7	151.5	177.6	*anti*	C2'-*endo* ($^2T_1$)
TG	18.2	34.8	147.1	299.2	*syn*	C2'-*endo* ($^2T_1$)

[a] Structures and frequencies were calculated with the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C2), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering (P) are also listed.

Table 5.5.8 Relative Gibbs energies and structural parameters of low-energy conformers of $[ddUrd+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
O4A	0.0	200.6	9.8	296.7	*anti*	C3'-*endo* ($^3T_2$)
TA	1.4	200.7	11.0	295.6	*anti*	C3'-*endo* ($^3T_2$)
O4B	7.1	223.7	170.3	298.1	*anti*	C2'-*endo* ($^2T_3$)
TB	8.2	225.8	170.2	295.9	*anti*	C2'-*endo* ($^2T_3$)
TC	11.3	55.6	151.8	299.9	*syn*	C2'-*endo* ($^2T_1$)
O2A	14.8	53.4	153.6	300.7	*syn*	C2'-*endo* ($^2T_1$)
O4C	14.8	193.3	207.5	43.4	*anti*	C3'-*exo* ($_3T^4$)
O2B	17.1	43.9	41.4	301.9	*syn*	C4'-*exo* ($_4T^3$)
TD	17.6	194.3	205.1	43.3	*anti*	C3'-*exo* ($_3T^4$)
O4D	19.1	190.3	18.7	60.6	*anti*	C3'-*endo* ($^3T_4$)
O4E	20.2	192.6	21.5	180.1	*anti*	C3'-*endo* ($^3T_4$)
TE	20.6	193.6	19.8	179.7	*anti*	C3'-*endo* ($^3T_4$)
TF	23.3	339.5	206.8	298.5	*syn*	C3'-*exo* ($_3T^4$)
TG	23.4	318.3	335.3	297.8	*syn*	C2'-*exo* ($_2T^1$)
O4F	23.9	189.7	167.6	176.5	*anti*	C2'-*endo* ($^2T_3$)
TH	24.3	192.7	170.5	177.1	*anti*	C2'-*endo* ($^2T_3$)

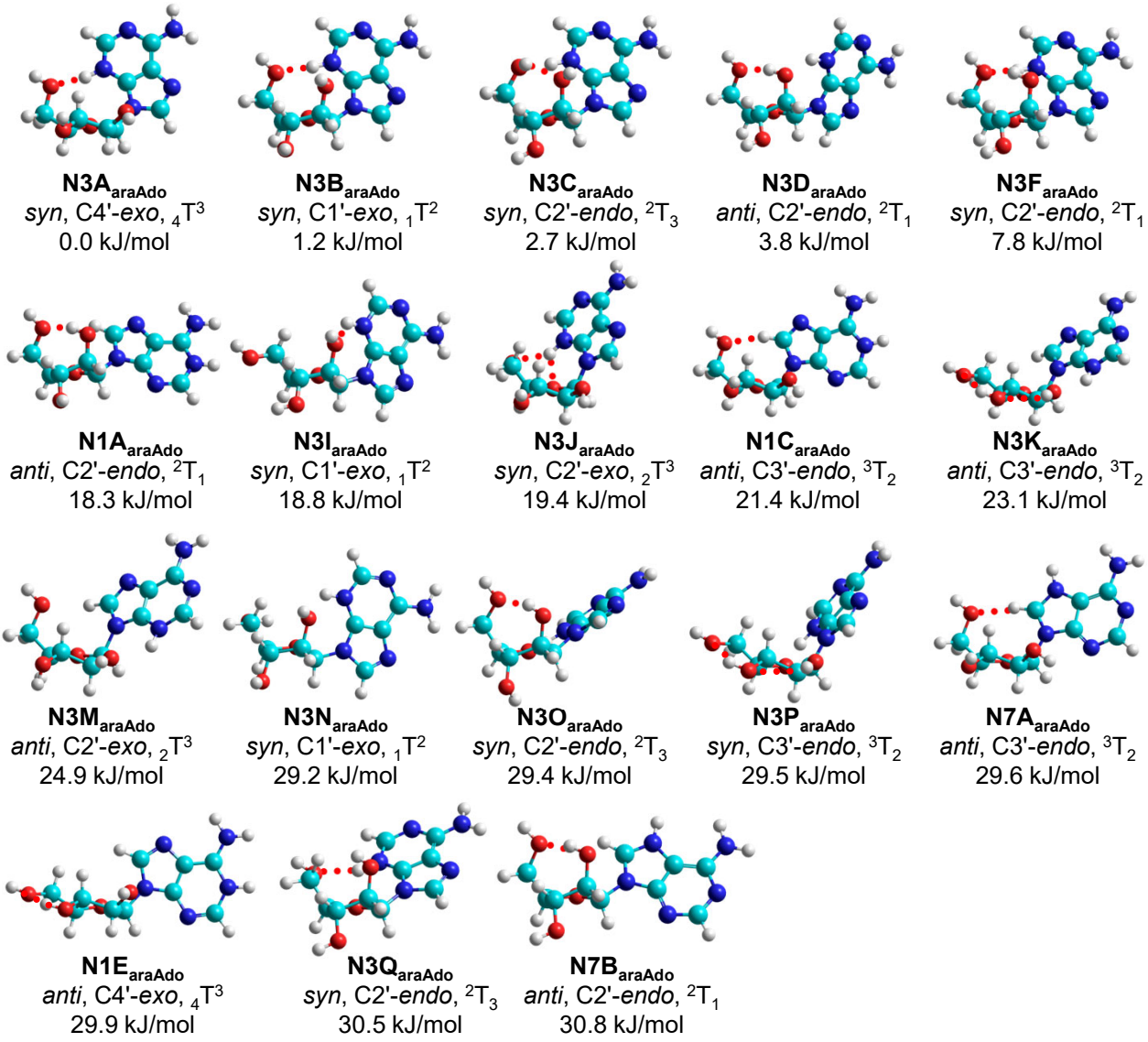
[a] Structures and frequencies were calculated at the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C4), pseudorotation angle (P), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering are also listed.

Table 5.5.9 Relative Gibbs energies and structural parameters for low-energy conformers of $[ddThd+H]^+$.[a]

Conformer	ΔG_{298} (kJ/mol)	χ (degrees)	P (degrees)	τ (degrees)	Nucleobase Orientation	Sugar Puckering
TA	0.0	200.5	12.0	295.2	*anti*	C3'-*endo* ($^3T_2$)
O4A	2.5	200.3	10.7	296.4	*anti*	C3'-*endo* ($^3T_2$)
TB	6.0	225.4	168.6	295.8	*anti*	C2'-*endo* ($^2T_3$)
O4B	9.5	222.8	168.8	298.1	*anti*	C2'-*endo* ($^2T_3$)
TC	9.5	55.8	151.8	299.8	*syn*	C2'-*endo* ($^2T_1$)
O2A	9.7	53.5	153.5	300.8	*syn*	C2'-*endo* ($^2T_1$)
O2B	12.1	40.0	39.3	302.3	*syn*	C4'-*exo* ($_4T^3$)
TD	16.1	199.7	11.8	295.4	*anti*	C3'-*endo* ($^3T_2$)
TE	17.0	193.0	203.6	46.5	*anti*	C3'-*exo* ($_3T^4$)
O4C	17.7	194.1	206.8	44.0	*anti*	C3'-*exo* ($^3T^4$)
TF	18.2	190.6	18.6	62.1	*anti*	C3'-*endo* ($^3T_4$)
TG	18.5	193.3	20.6	179.8	*anti*	C3'-*endo* ($^3T_4$)
O4D	20.9	189.9	20.0	60.7	*anti*	C3'-*endo* ($^3T_4$)
TH	21.8	340.1	205.7	298.1	*syn*	C3'-*exo* ($_3T^4$)
TI	22.2	186.2	330.6	38.7	*anti*	C2'-*exo* ($_2T^1$)
TJ	22.3	318.3	335.9	297.2	*syn*	C2'-*exo* ($_2T^1$)
TK	22.9	193.6	167.6	176.8	*anti*	C2'-*endo* ($^2T_3$)
O4E	22.9	192.5	22.0	180.4	*anti*	C3'-*endo* ($^3T_4$)
TL	23.0	194.2	20.8	67.1	*anti*	C3'-*endo* ($^3T_4$)
O2C	24.1	343.8	201.6	299.6	*syn*	C3'-*exo* ($_3T^4$)

[a] Structures and frequencies were calculated at the B3LYP/6-311+G(d,p) level of theory, and Gibbs energies at the B3LYP/6-311+G(2d,2p) level of theory. The nucleobase orientation (χ, ∠O4′C1′N1C4), pseudorotation angle (P), 5′-hydroxyl orientation (τ, ∠O4′C4′C5′O5′) and sugar puckering are also listed.

Figure 5.1



[araAdo+H]$^+$

Figure 5.1 Representative B3LYP/6-311+G(d,p) low-energy conformers of [araAdo+H]$^+$ with relative B3LYP/6-311+G(2d,2p) Gibbs energies at 298 K. The protonation site, orientation of the nucleobase, and specific puckering of the sugar are noted for each conformer.

Figure 5.2

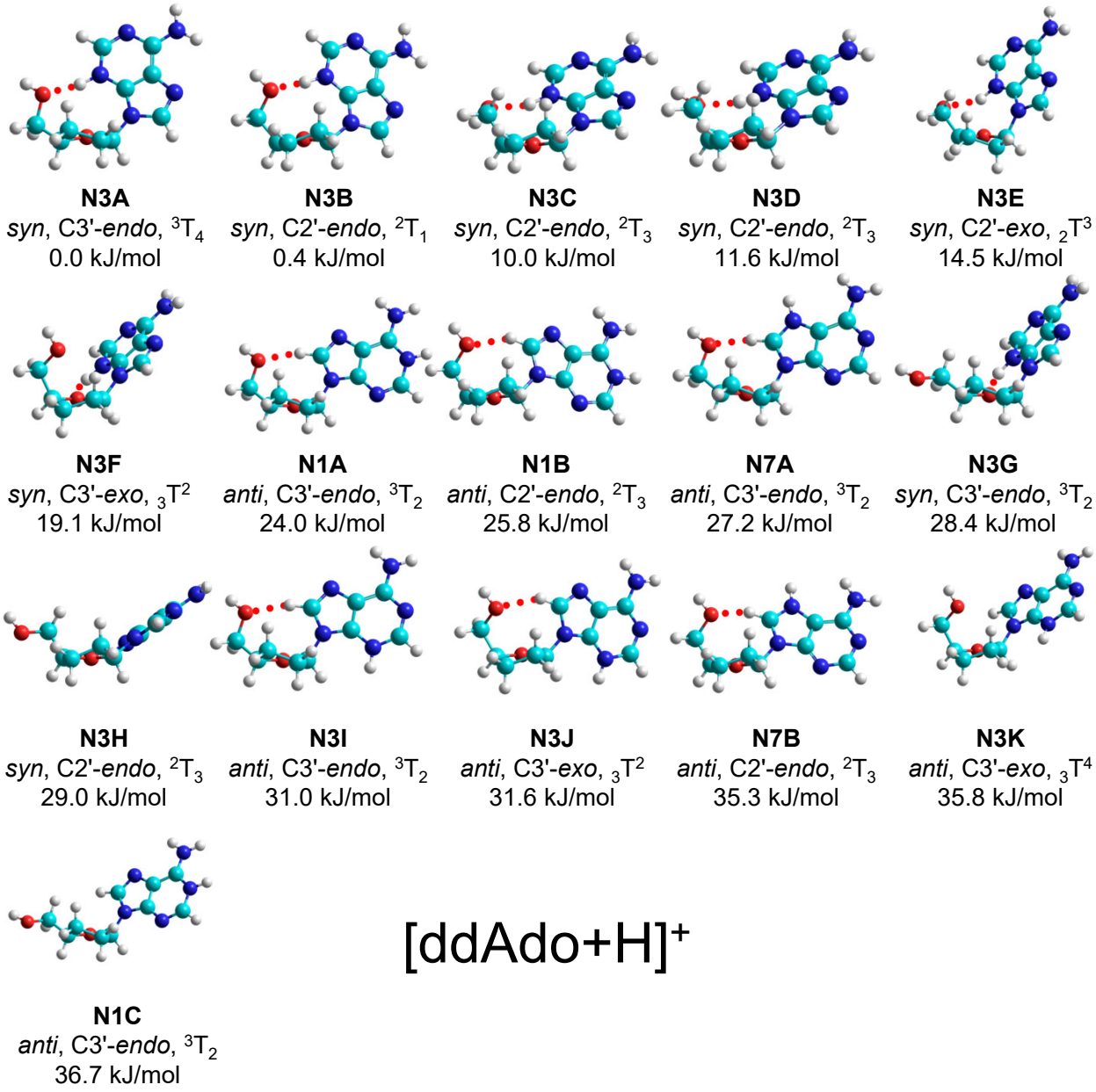


Figure 5.2 Representative B3LYP/6-311+G(d,p) low-energy conformers of $[ddAdo+H]^+$ with relative Gibbs energetics calculated at the B3LYP/6-311+G(2d,2p) level of theory at 298 K. The conformer designation, nucleobase orientation, and sugar puckering are noted for each conformer.

Figure 5.3

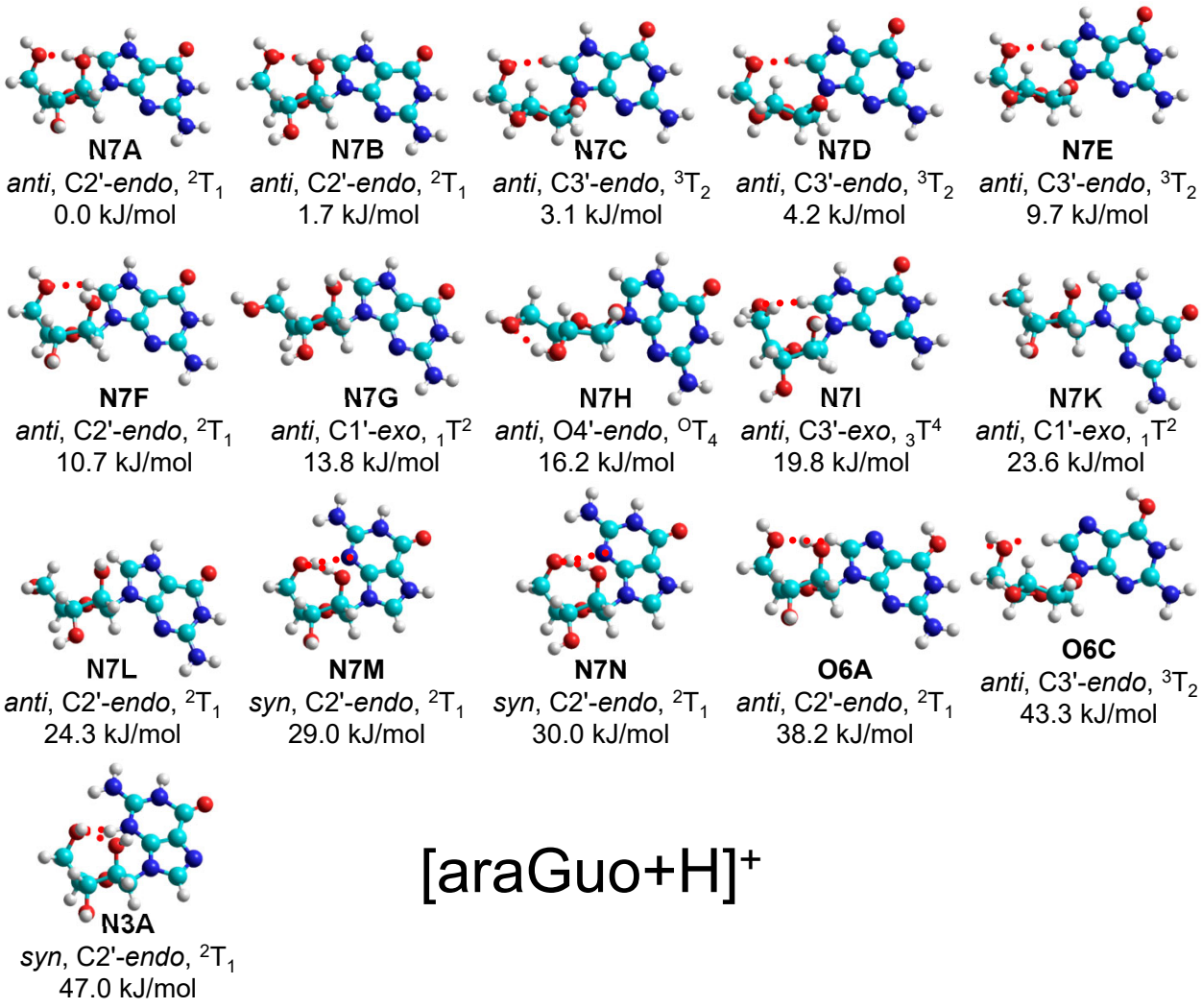


Figure 5.3 Representative B3LYP/6-311+G(d,p) low-energy conformers of $[araGuo+H]^+$ with relative B3LYP/6-311+G(2d,2p) Gibbs energies at 298 K. The protonation site, orientation of the nucleobase, and specific puckering of the sugar are noted for each conformer.

Figure 5.4

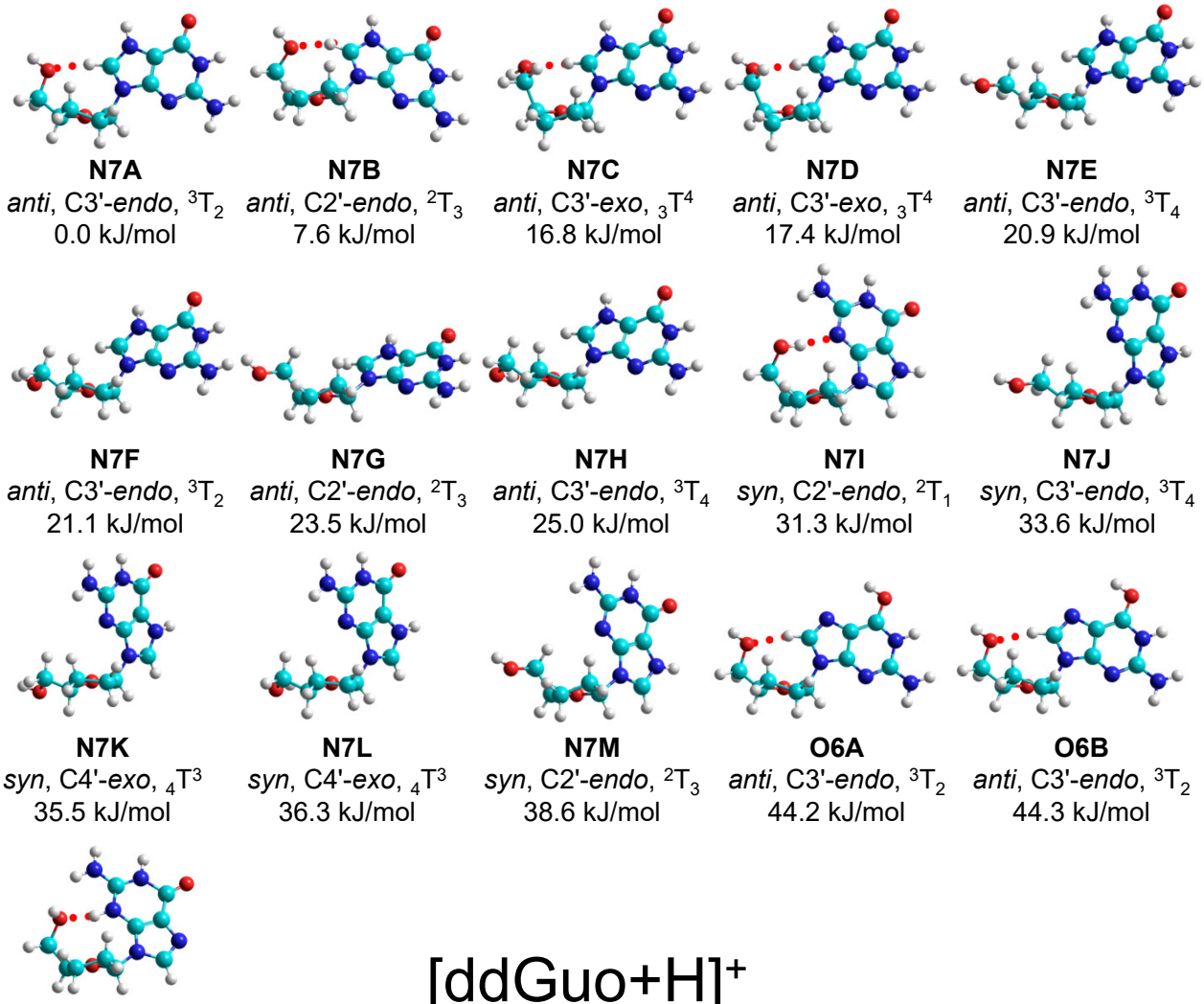
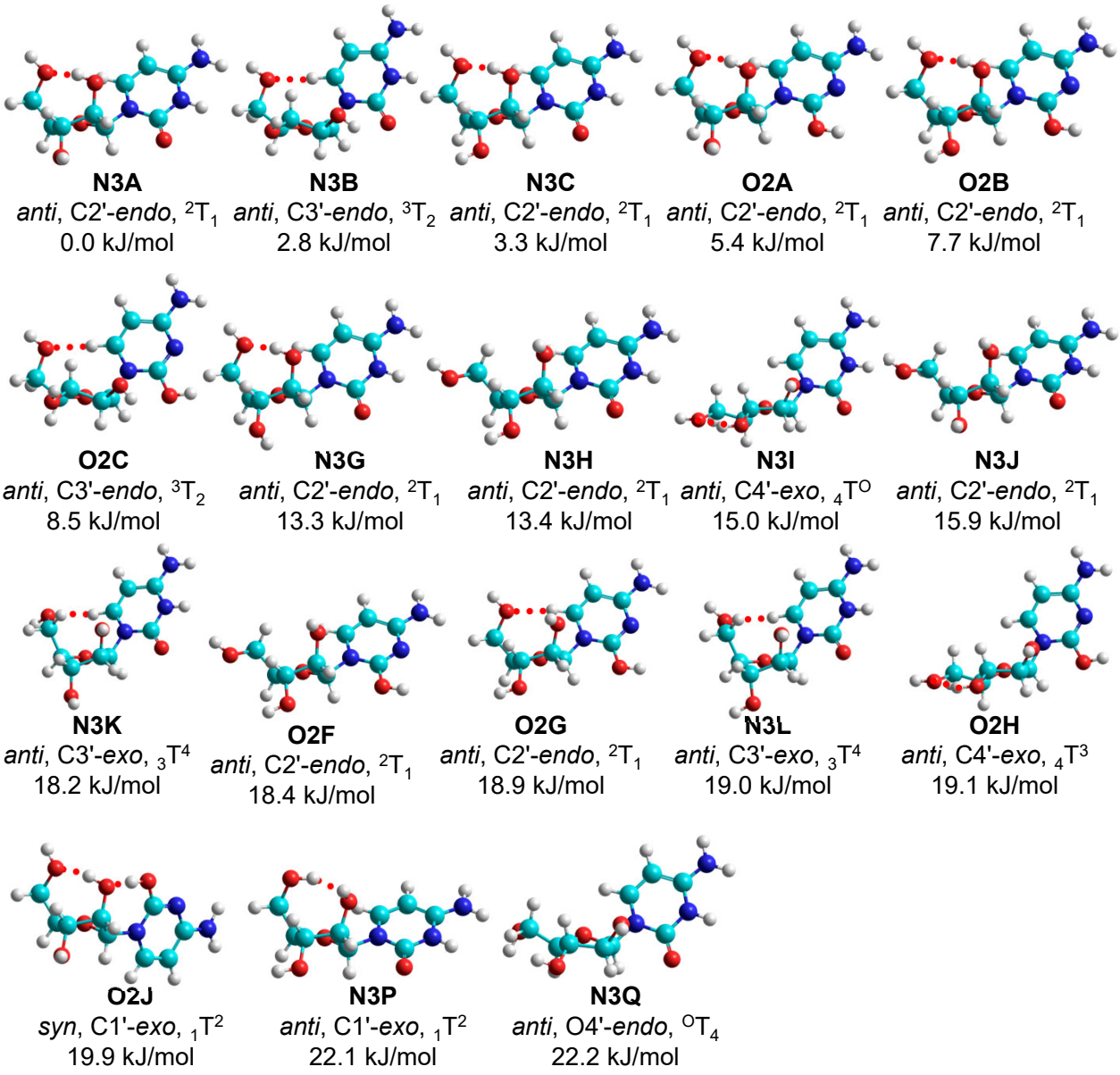


Figure 5.4 Representative B3LYP/6-311+G(d,p) low-energy conformers of $[ddGuo+H]^{+}$ with relative Gibbs energetics calculated at the B3LYP/6-311+G(2d,2p) level of theory at 298 K. The conformer designation, nucleobase orientation, and sugar puckering are noted for each conformer.

Figure 5.5



[araCyd+H]+

Figure 5.5 Representative B3LYP/6-311+G(d,p) low-energy conformers of [araCyd+H]+ with relative B3LYP/6-311+G(2d,2p) Gibbs energies at 298 K. The protonation site, orientation of the nucleobase, and specific puckering of the sugar are noted for each conformer.

Figure 5.6

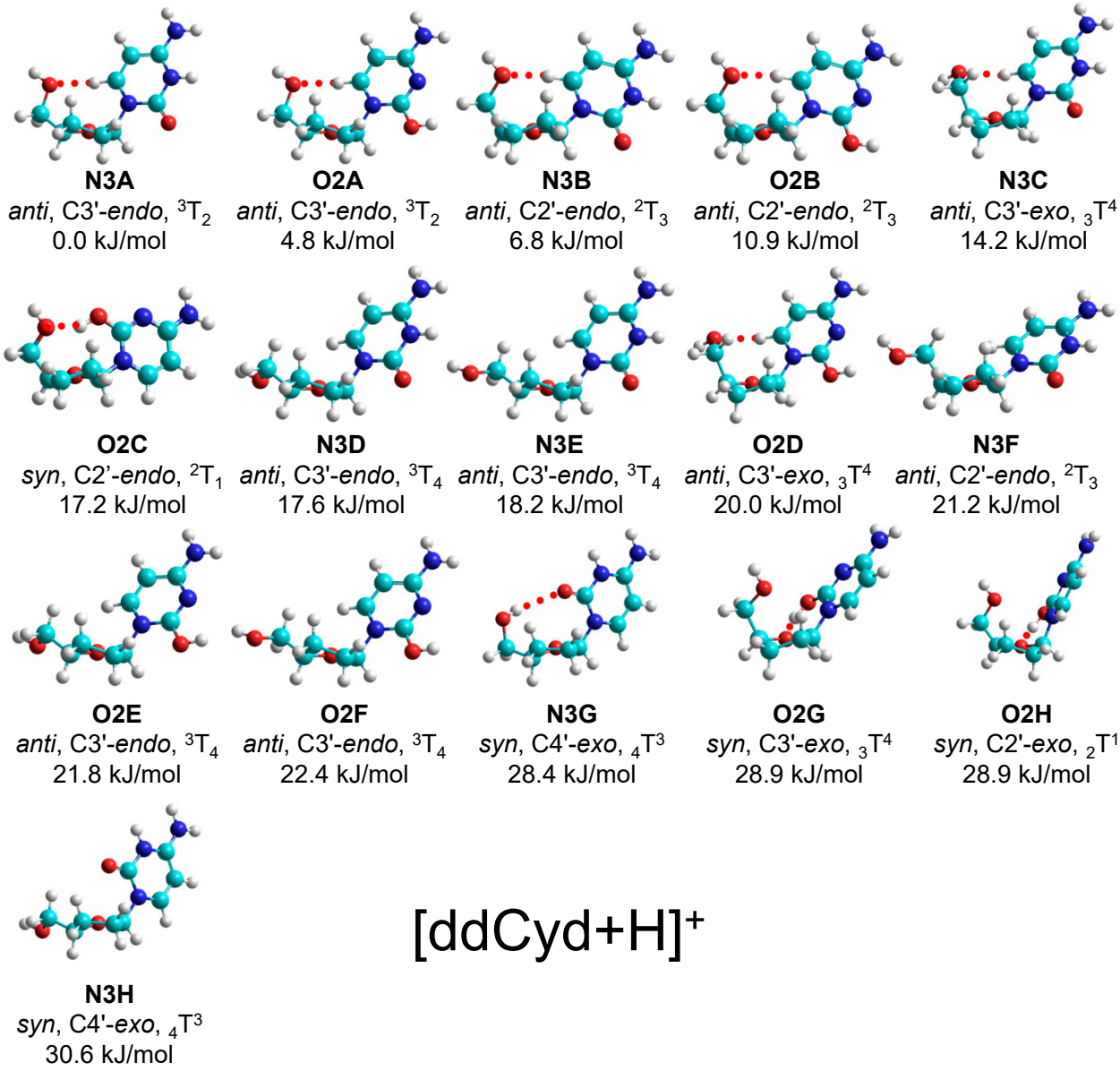


Figure 5.6 Representative B3LYP/6-311+G(d,p) low-energy conformers of $[ddCyd+H]^+$ with relative Gibbs energetics calculated at the B3LYP/6-311+G(2d,2p) level of theory at 298 K. The conformer designation, nucleobase orientation, and sugar puckering are noted for each conformer.

Figure 5.7

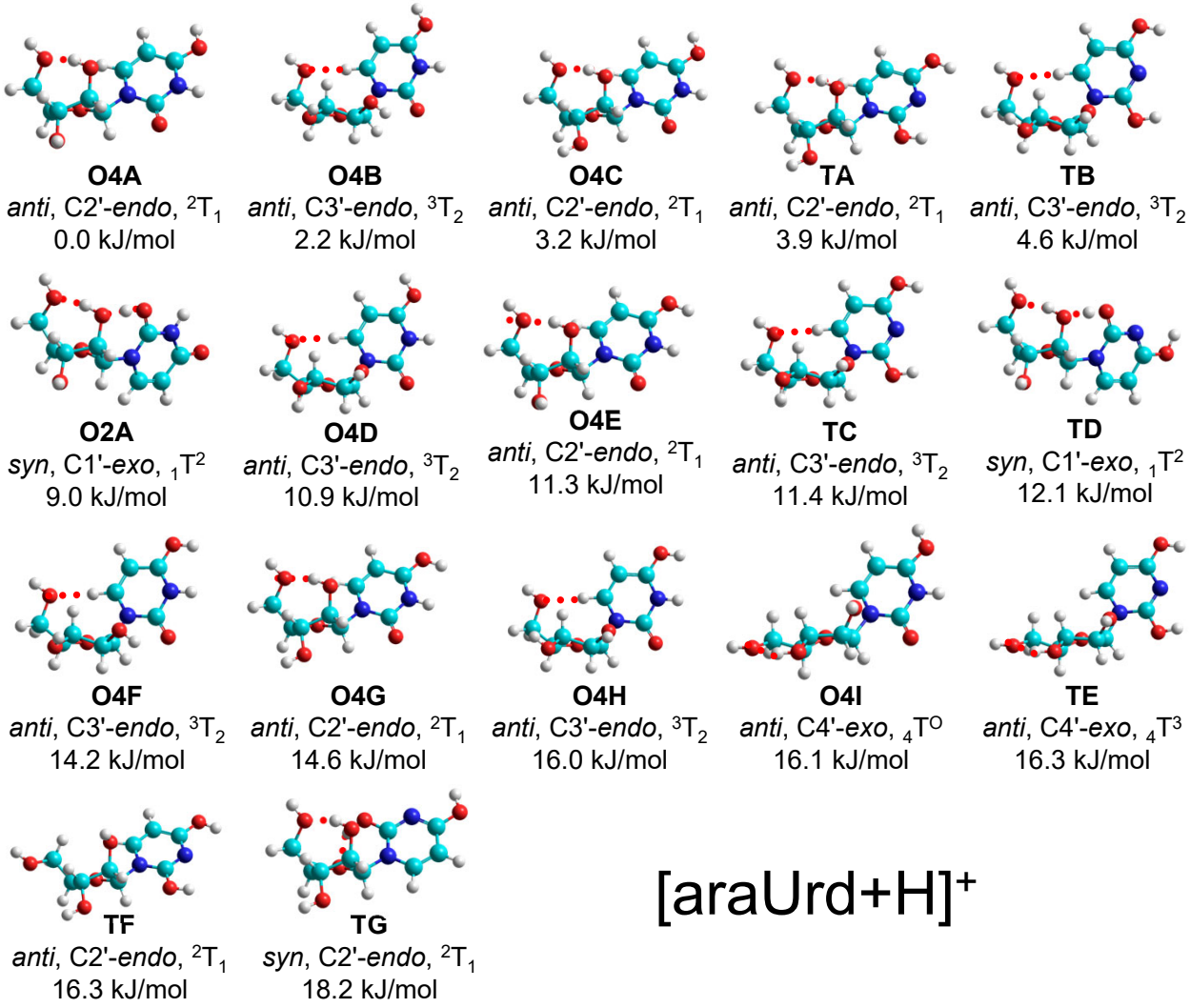


Figure 5.7 Representative B3LYP/6-311+G(d,p) low-energy conformers of $[araUrd+H]^+$ with relative B3LYP/6-311+G(2d,2p) Gibbs energies at 298 K. The protonation site, orientation of the nucleobase, and specific puckering of the sugar are noted for each conformer.

Figure 5.8

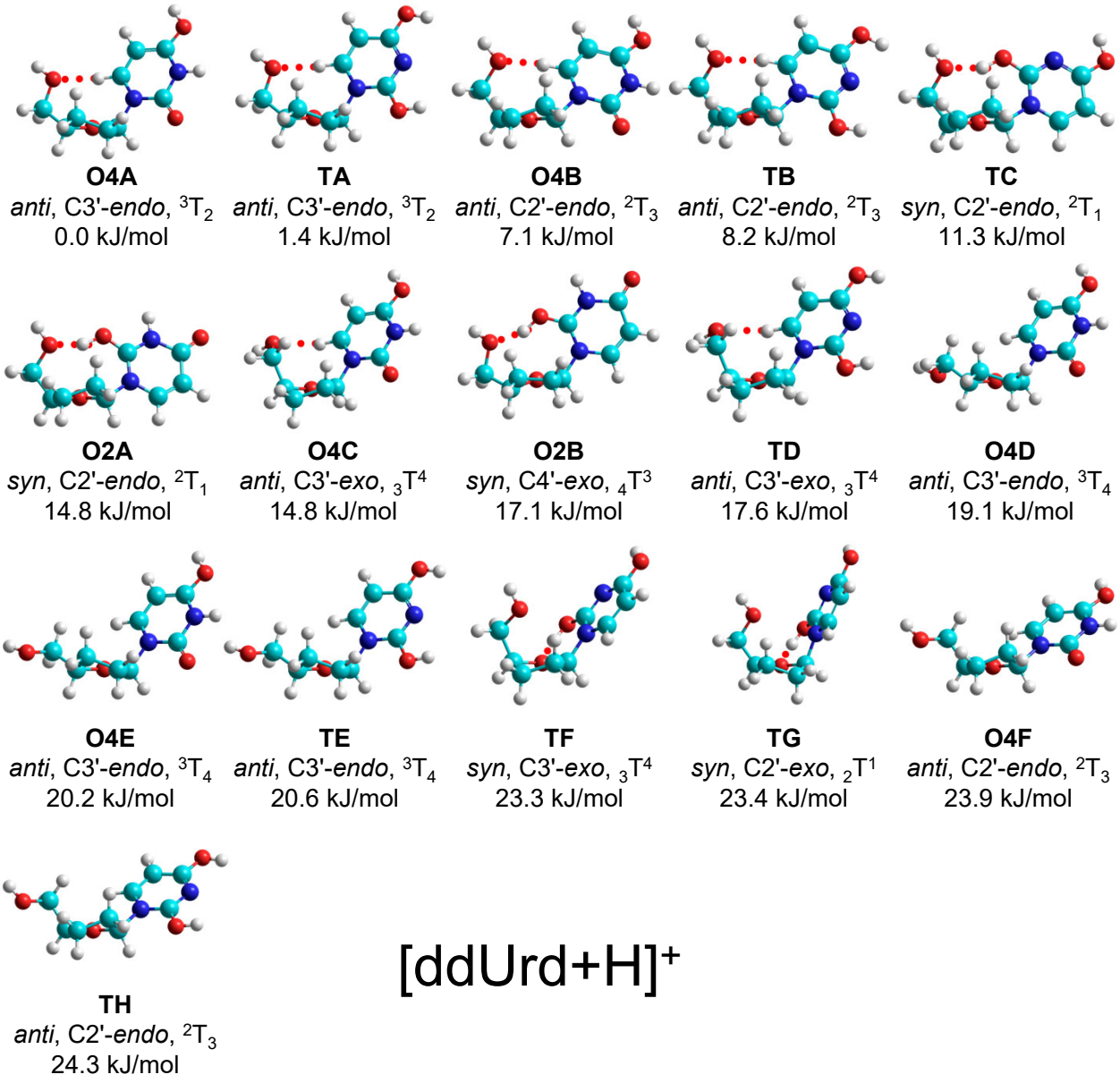
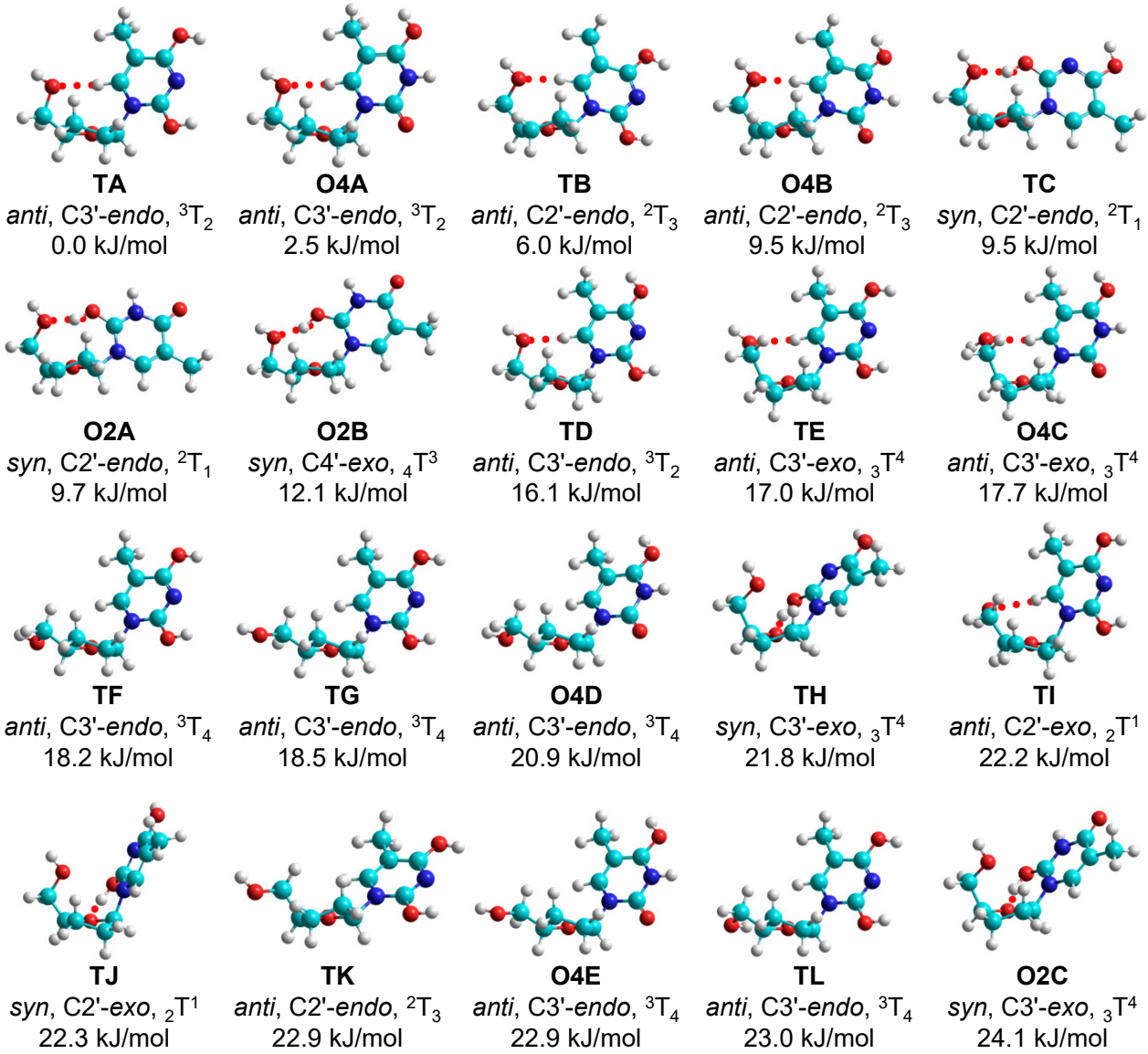


Figure 5.8 Representative B3LYP/6-311+G(d,p) low-energy conformers of $[ddUrd+H]^+$ with relative Gibbs energetics calculated at the B3LYP/6-311+G(2d,2p) level of theory at 298 K. The conformer designation, nucleobase orientation, and sugar puckering are noted for each conformer.

Figure 5.9



[ddThd+H]$^+$

Figure 5.9 Representative B3LYP/6-311+G(d,p) low-energy conformers of [ddThd+H]$^+$ with relative Gibbs energetics calculated at the B3LYP/6-311+G(2d,2p) level of theory at 298 K. The conformer designation, nucleobase orientation, and sugar puckering are noted for each conformer.

Figure 5.10

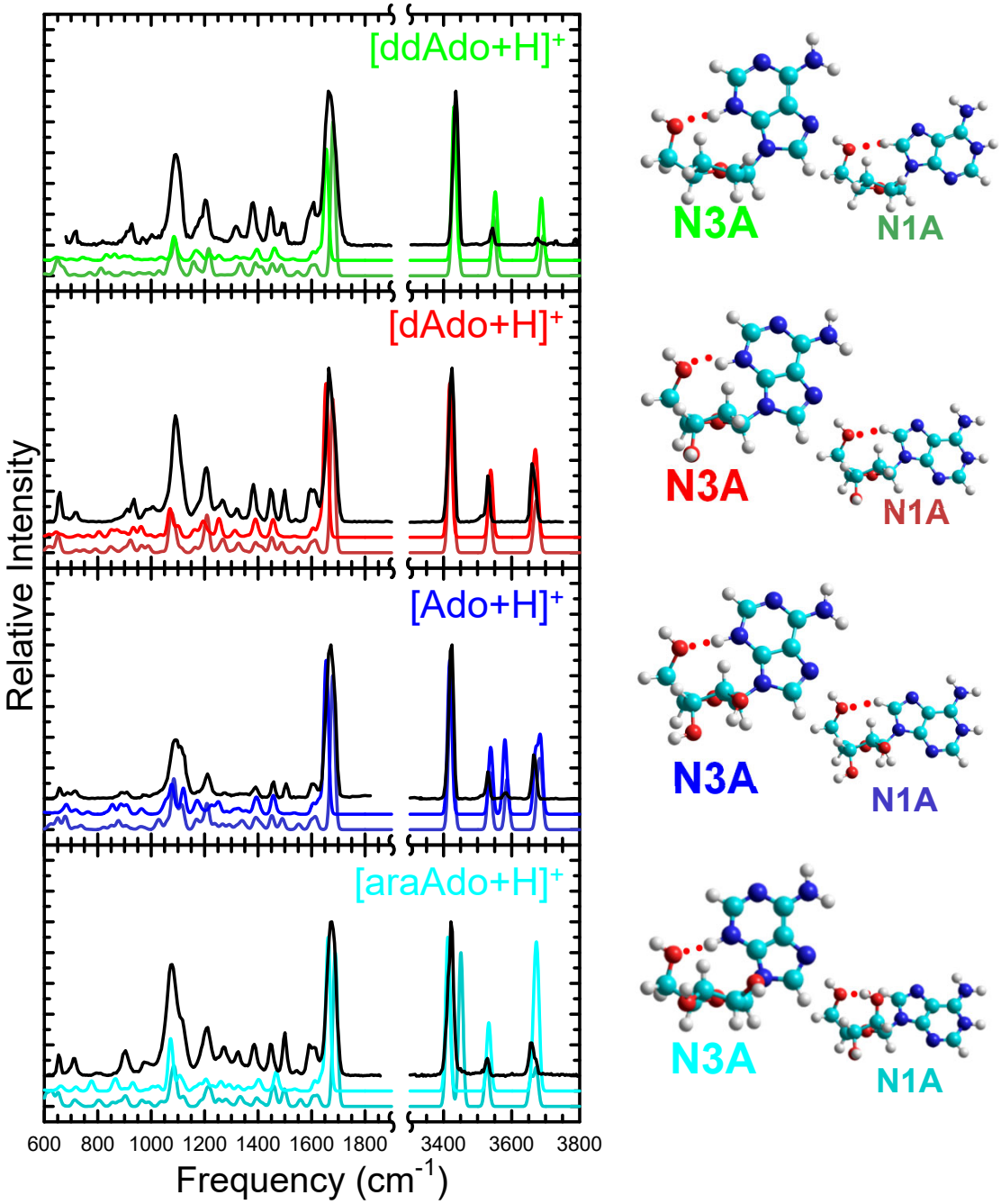


Figure 5.10 Comparison of the experimental IRMPD spectra of 2′,3′-dideoxyadenosine, 2′-deoxyadenosine, adenosine, and adenosine arabinoside. Also compared are the predicted linear IR spectra and images of the conformers identified as the primary contributors to the experimental spectrum as calculated at the B3LYP/6-311+G(d,p) level of theory. The intensities of the experimental and predicted spectra are normalized to facilitate comparison. The larger structures are likely the more important contributors to the experimental spectrum.

Figure 5.11

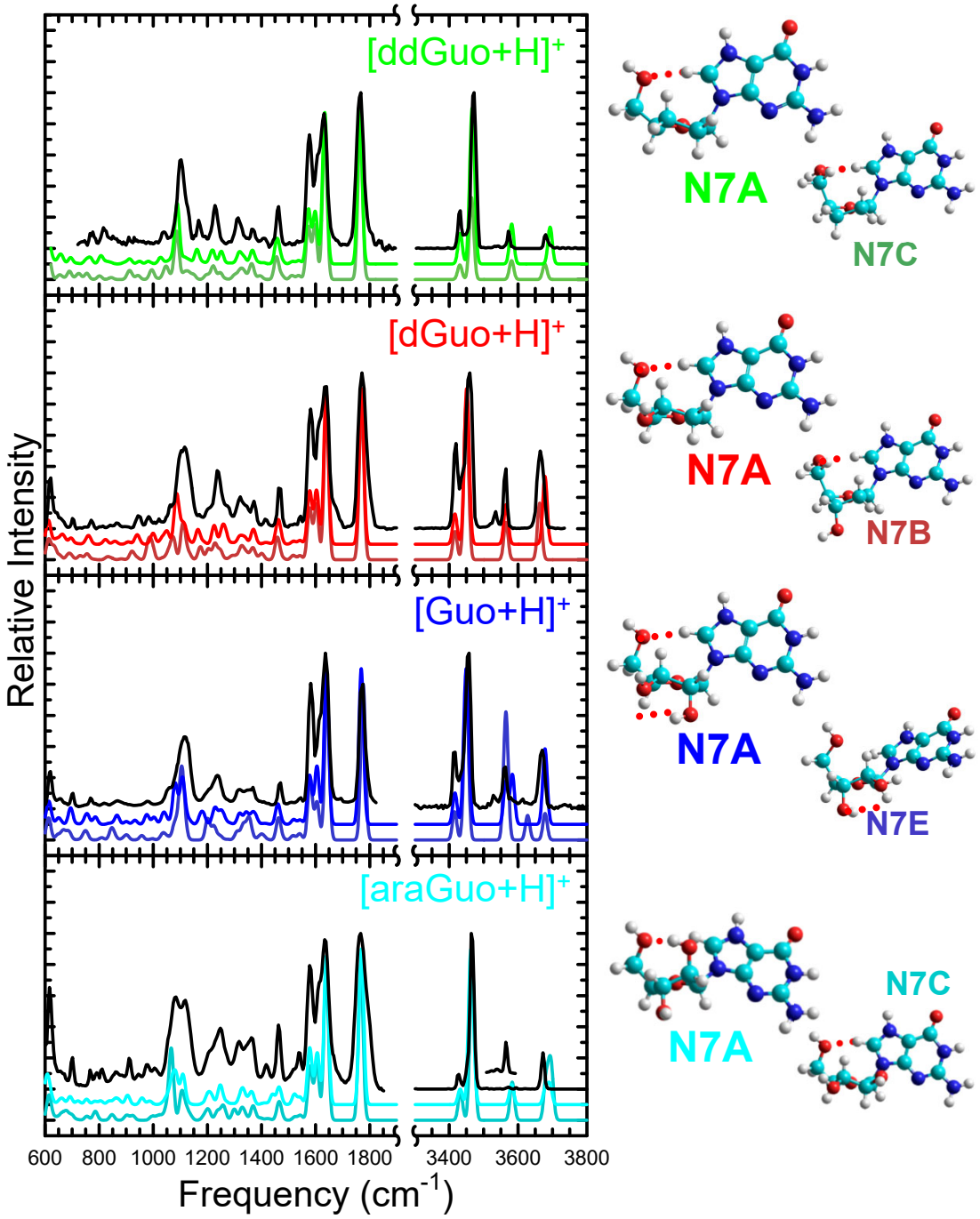


Figure 5.11 Comparison of the experimental IRMPD spectra of 2′,3′-dideoxyguanosine, 2′-deoxyguanosine, guanosine, and guanosine arabinoside. Also compared are the predicted linear IR spectra and images of the conformers identified as the primary contributors to the experimental spectrum as calculated at the B3LYP/6-311+G(d,p) level of theory. The intensities of the experimental and predicted spectra are normalized to facilitate comparison. The larger structures are likely the more important contributors to the experimental spectrum.

Figure 5.12

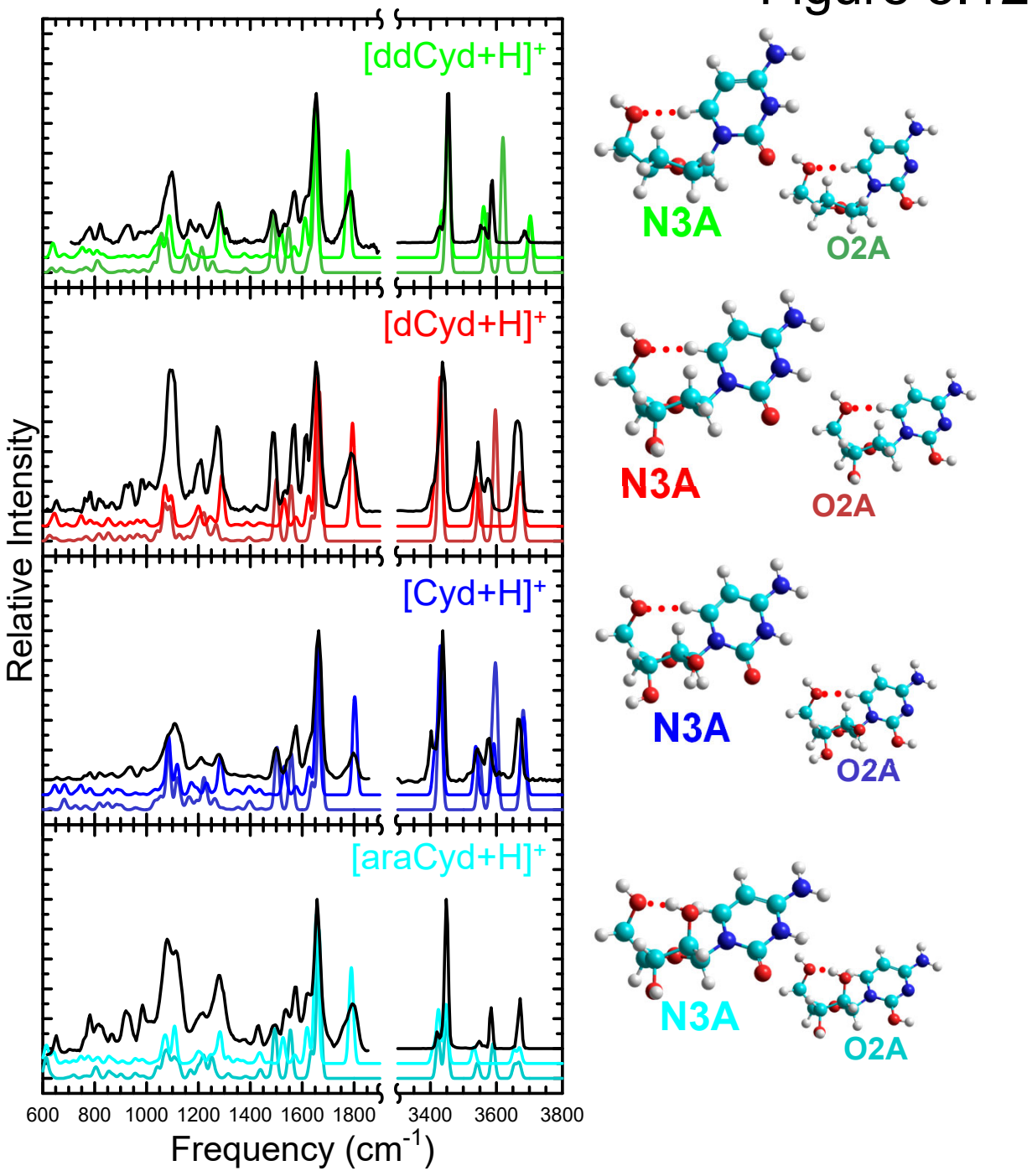


Figure 5.12 Comparison of the experimental IRMPD spectra of 2′,3′-dideoxycytidine, 2′-deoxycytidine, cytidine and cytosine arabinoside. Also compared are the predicted linear IR spectra and images of the conformers identified as the primary contributors to the experimental spectrum as calculated at the B3LYP/6-311+G(d,p) level of theory. The intensities of the experimental and predicted spectra are normalized to facilitate comparison. The larger structures are likely the more important contributors to the experimental spectrum.

Figure 5.13

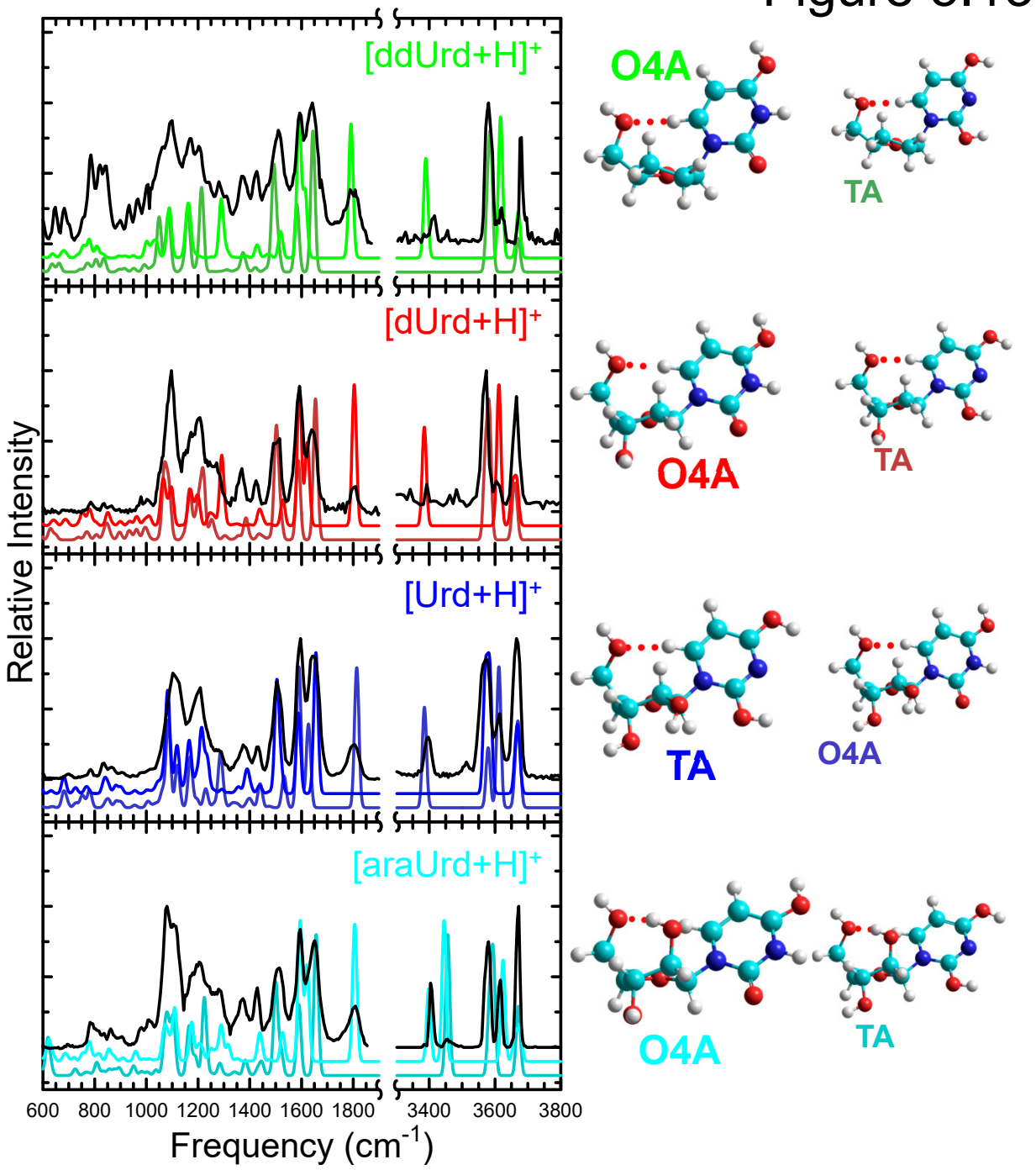


Figure 5.13 Comparison of the experimental IRMPD spectra of 2′,3′-dideoxyuridine, 2′-deoxyuridine, uridine, and uracil arabinoside. Also compared are the predicted linear IR spectra and images of the conformers identified as the primary contributors to the experimental spectrum as calculated at the B3LYP/6-311+G(d,p) level of theory. The intensities of the experimental and predicted spectra are normalized to facilitate comparison. The larger structures are like the more important contributors to the experimental spectrum

Figure 5.14

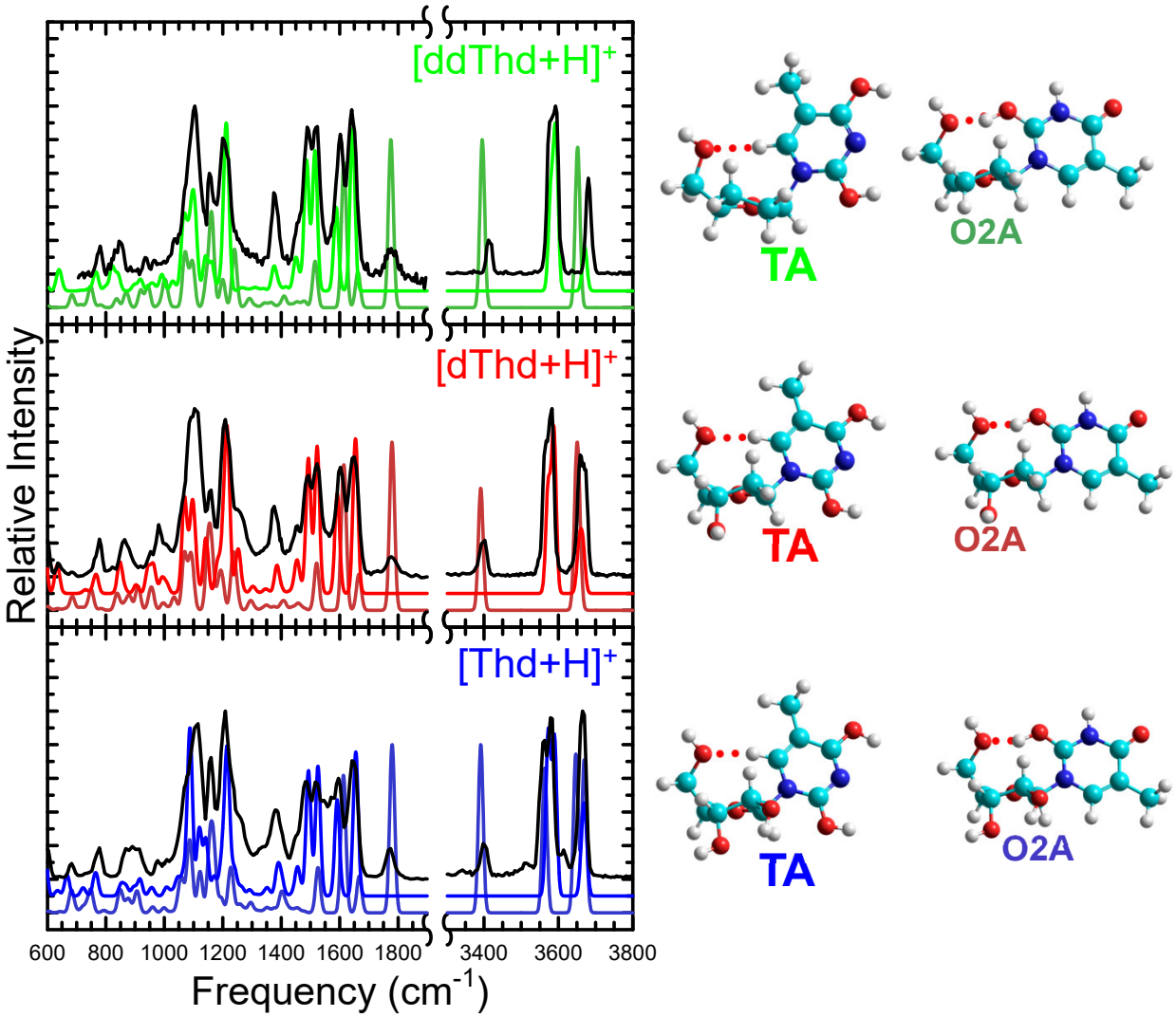


Figure 5.14 Comparison of the experimental IRMPD spectra of 2′,3′-dideoxythymidine, 2′-deoxythymidine and thymidine. Also compared are the predicted linear IR spectra and images of the conformers identified as the primary contributors to the experimental spectrum as calculated at the B3LYP/6-311+G(d,p) level of theory. The intensities of the experimental and predicted spectra are normalized to facilitate comparison. The larger structures are likely the more important contributors to the experimental spectrum.

Figure 5.15

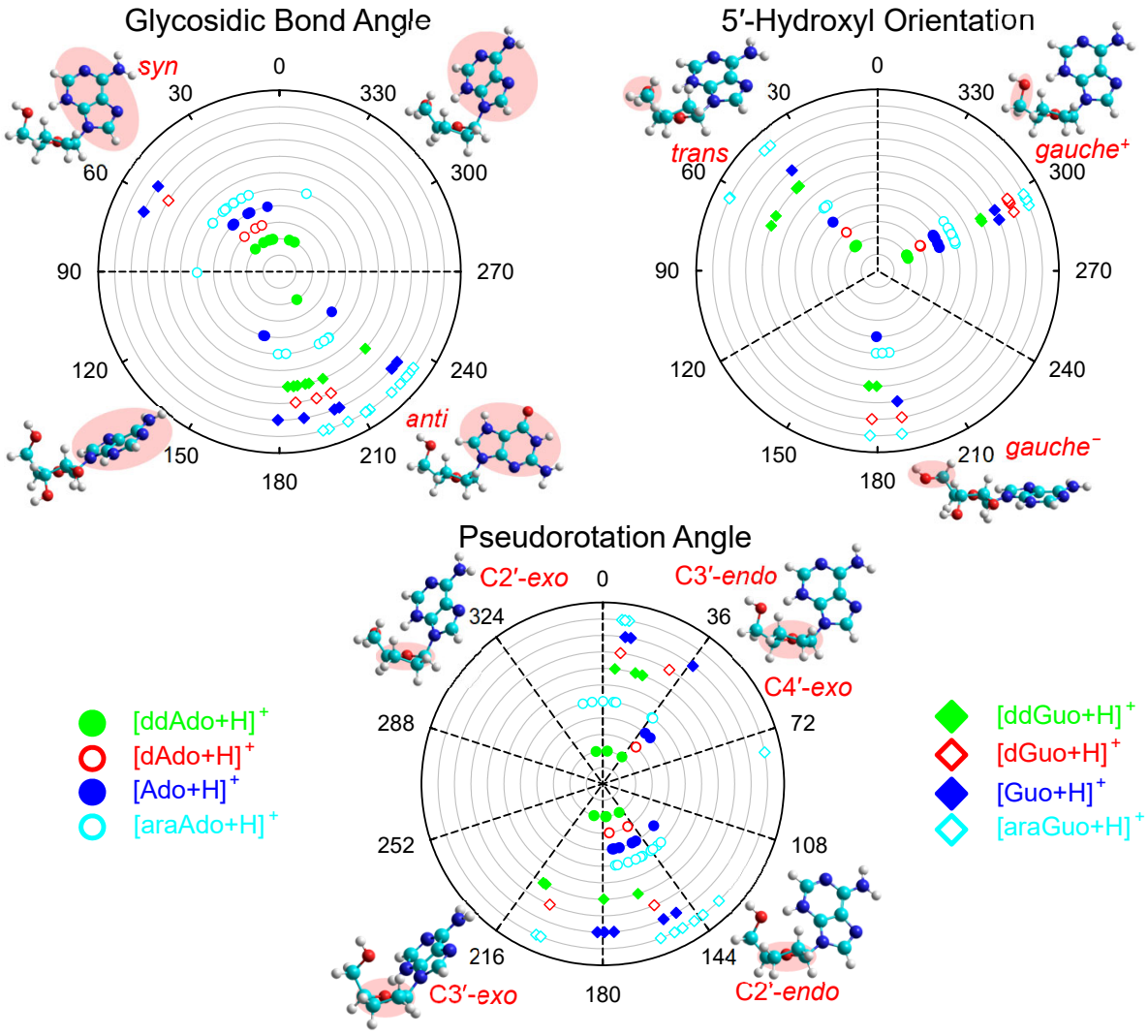


Figure 5.15 The three primary parameters that govern the structure of the conformers of the protonated purine nucleoside analogues calculated, glycosidic bond angle, 5′-hydroxyl orientation and pseudorotation angle (sugar puckering), are displayed for all conformers with relative stabilities predicted within 25 kJ/mol Gibbs energy of the corresponding ground conformers.

Figure 5.16

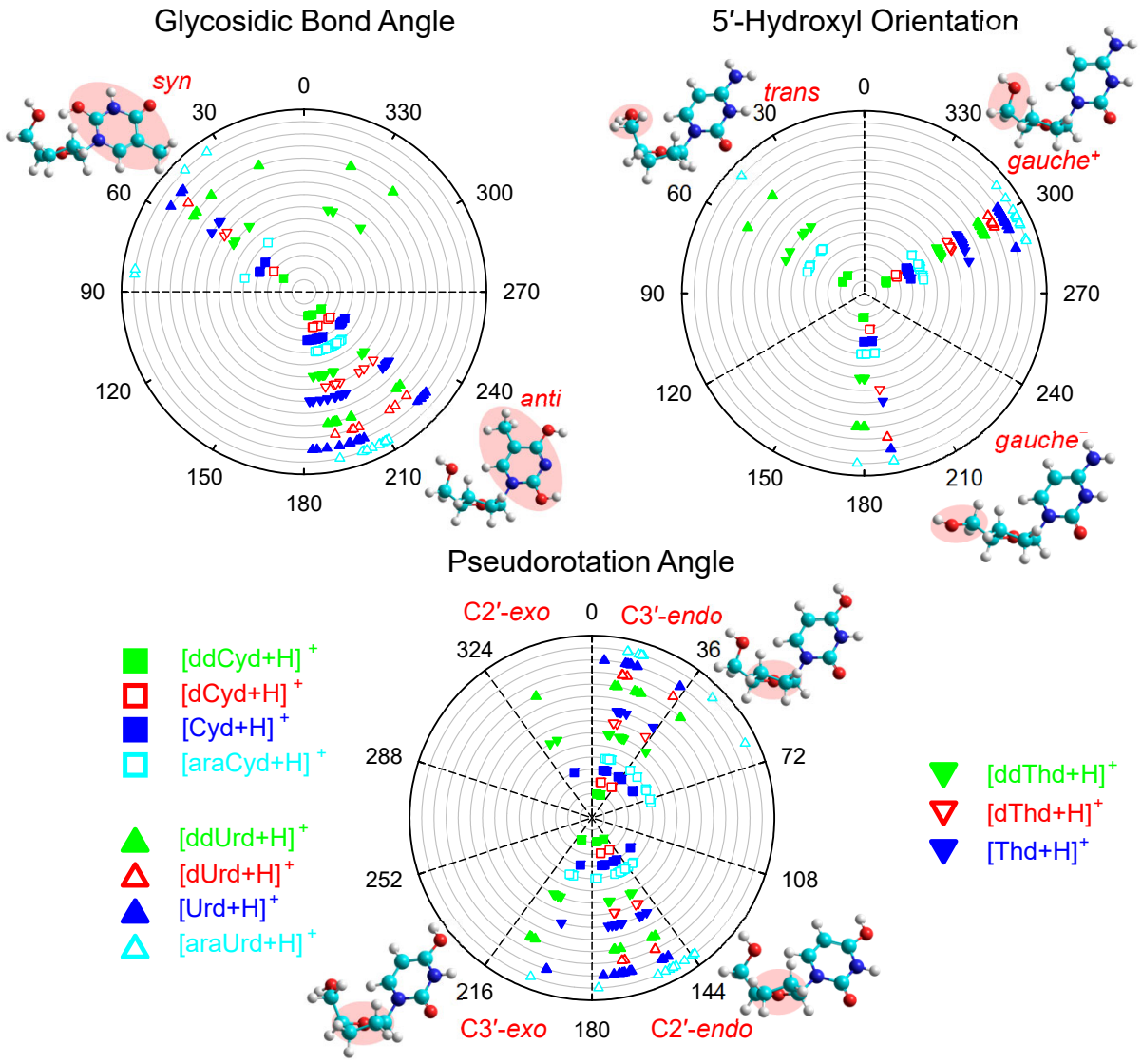


Figure 5.15 The three primary parameters that govern the structure of the conformers of the protonated pyrimidine nucleoside analogues calculated, glycosidic bond angle, 5′-hydroxyl orientation and pseudorotation angle (sugar puckering), are displayed for all conformers with relative stabilities predicted within 25 kJ/mol Gibbs energy of the corresponding ground conformers.

CHAPTER 6 CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

In this thesis work I have developed hardware and software tools, for use in IRMPD tandem mass spectrometry experiments. I have also developed a suite of software tools to assist with simple but extremely repetitive file management and parsing tasks involved in computational chemistry methods common within the Rodgers research group. I have used these software tools and the IRMPD action spectroscopy experimental setup I developed to examine the gas-phase intrinsic structures of several modified protonated nucleosides. In particular, I examined the gas-phase structures of 2-thiouridine and 4-thiouridine, as well as the arabinose and 2′,3′-dideoxyribose analogues of adenosine, guanosine, cytidine, and uridine, as well as 2′,3′-dideoxythymidine. These results were carefully compared with the previously published results pertaining to their DNA and RNA analogues to examine the impact of these modifications on nucleoside structures in the experiments, and any impact they may have on the potential energy landscape of the modified nucleosides.

In **Chapter 2** the development of an IRMPD action spectroscopy system coupled to two commercial mass spectrometers is detailed. Access to IRMPD action spectroscopy in the hydrogen-stretching region on more readily available instrumentation is important to increasing access to this powerful structural analysis tool. Historically, custom-built mass spectrometers were the preferred platform for IRMPD action spectroscopy, in order to provide the necessary hardware and software control to acquire high-quality spectra. However, commercially available mass spectrometers often offer better overall performance at reasonable cost and can be readily modified to allow access to the ion

cloud for an incident laser, an important feature not only for IRMPD action spectroscopy but also UV photodissociation. I coupled the LaserVision OPO system and the dedicated controller for the system to a Bruker amaZon ETD QIT-MS and Bruker solariX FT-ICR MS. The scanning of the mass spectrometers, opening of the optical shutter, and firing of the pump laser are all synchronized to ensure that ions are only irradiated during a set activation window, and each activation window contains the same number of laser pulses. On the QIT-MS, further software synchronization is established via a custom LabVIEW™ program to control acquisition on the QIT-MS and scanning of the OPO. Critical performance parameters for the QIT-MS are characterized to ensure that it is operated in such a way as to acquire the highest-quality spectra possible. The FT-ICR MS offers no easily accessible software synchronization options, which complicates data acquisition, but the only compromise for IRMPD action spectroscopy is the shape of the ion cloud which has reduced overlap with the incident laser compared to the QIT-MS, resulting in much longer irradiation times and therefore experiments.

In **Chapter 3** the computational chemistry method used to generate candidate structure and linear IR spectra for analysis of IRMPD action spectra is described. Also detailed are the software tools and scripts I developed to facilitate not only more rapid creation of the hundreds of individual calculations, but also the analysis of the results. Two Python scripts for the automated analysis of the initial conformational search are described. These scripts seek to greatly accelerate and improve the sampling of low-energy candidate structures from the molecular mechanics conformational search in order to improve exploration of conformational space and improve the variety of structures optimized, and linear IR spectra predicted by the more expensive DFT calculations that

follow. The performance of these two scripts, which address the issue of identifying structural uniqueness in small molecules by two different methods are characterized and compared.

In **Chapter 4** the IRMPD action spectroscopy setup detailed in **Chapter 2** is utilized in conjunction with experiments at the FELIX facility and the computational methods described in **Chapter 3** to examine protonated 2-thiouridine and 4-thiouridine. Comparisons of these results to previous work characterizing the intrinsic gas-phase conformations of 2′-deoxyuridine and uridine, as well as 2-thiouracil and 4-thiouracil enable examination of the impact of thiation on the intrinsic gas-phase structure, and the impact of the sugar moiety on protonation of the nucleobase. It was observed that 2-thiouridine and 4-thiouridine display unique protonation preferences with protonated 2-thiouridine strongly preferring formation of the minor 2-sulfhydryl-4-hydroxy tautomer with some O4-protonated conformer present, whereas 4-thiouridine strongly prefers O4-protonation and some minor tautomer may be present in the experiments. This is generally in agreement with the available experimental data for 2-thiouracil and 4-thiouracil.[59] Both differ from the DNA and RNA uridine analogues, which generally display a more even distribution of minor tautomer and O4-protonated conformers.

The IRMPD action spectra of the arabinose and 2′,3′-dideoxyribose analogues of adenosine, guanosine, cytidine, and uridine, as well as 2′,3′-dideoxythymidine are compared with predicted IR spectra of candidate low-energy conformers in **Chapter 5**. These experiments allow for examination of the impact of the 2′- and 2′,3′- sugar hydroxy moieties on intrinsic nucleoside structure and protonation preference. No significant changes in the experimentally relevant structures were obvious, but changes in the

intramolecular hydrogen-bonding interactions and sugar puckering modes accessed and preferred are observed. The arabinonucleosides all exhibit some evidence, either spectroscopic or energetic, for the presence of an O2′H···O5′ hydrogen-bonding interaction. This interaction preferentially stabilizes C2′-*endo* sugar puckering, which is therefore generally preferred, although [araAdo+H]$^+$, like the adenosine DNA and RNA analogues, is energetically dominated by N3-protonation allowing for the N3H$^+$···O5′ ionic hydrogen-bonding interaction with which C3′-*endo* puckering is preferred. However, like the adenosine DNA and RNA analogues, N1-protonated conformers are also present for [araAdo+H]$^+$ where an O2′H···O5′ hydrogen-bonding interaction and C2′-*endo* sugar puckering are observed. The absence of the 2′- and 3′-hydroxy substituents does not alter the preferred protonation preferences of the DNA nucleoside analogues. However, spectroscopic evidence suggests that the absence of both the 2′- and 3′-hydroxy substituents shifts the sugar puckering preference from generally C2′-*endo* to C3′-*endo*. The lack of additional opportunities for sugar-sugar hydrogen-bonding stabilization greatly reduces the number of similar conformers found in comparison to the arabinoside, DNA, and RNA nucleoside analogues. The computational methods used in **Chapter 5** leverage the conformational search analysis scripts described in **Chapter 3** to largely automate the candidate structure selection process. In particular, the structural parameter RMSD method described in **Section 3.5.2**, used for the calculations of the 2′,3′-dideoxyribose analogues performed very well, with only a single important low-energy conformer identified as absent from the automated selection.

6.2 Future Work

This thesis work sought to develop tools for tandem mass spectrometry experiments, analysis, and complimentary theoretical calculations. This work also uses some of these tools to examine the protonated gas-phase structures of several modified nucleic acid building blocks. This work can be expanded upon in several ways, discussed below:

In **Chapter 2** a tunable IR laser system is coupled to two commercial mass spectrometers, a QIT MS and a FT-ICR MS. However, a component of the original scope of this project involved the additional setup of two fixed frequency CO_2 lasers for general purpose IRMPD photodissociation. During the design and alignment of the tunable IR laser to the two mass spectrometers the additional beam paths for the CO_2 lasers and the compromises that needed to be made for them became infeasible. Currently, the tunable IR OPO laser is readily coupled to both mass spectrometers and implementation of a CO_2 laser on either instrument would require dismantling part or all of the current OPO beam path. It is possible that with further consideration a suitable beam path for the CO_2 lasers that does not compromise that of the OPO laser may be found that will allow for either simultaneous operation, or just more readily switching between beam paths, without disassembly or realignment. If the appropriate contacts can be made at Bruker, it is also possible that custom software might be written to better facilitate software synchronization between the OPO laser controller and the FT-ICR MS.

In **Chapter 3** two methods for calculating a structural RMSD between two conformers are described. The goal of these methods is to largely enable automation of the candidate structure selection process so as to greatly accelerate and improve the

computational method used in conjunction with IRMPD action spectroscopy measurements. Although the structural parameter RMSD is generally more effective than the more common Cartesian coordinate RMSD for relatively small and rigid molecules such as those examined in this work, there remains a great deal of room for improvement. Currently, the structural parameters are limited to dihedral angles, a result of the scripts development specifically for the protonated nucleosides, where the proton is relatively fixed at the temperatures examined. However, in the sodiated nucleosides for instance, the sodium ion can chelate readily, and structural parameters that deal with proximity to a given atom might be extremely useful in identifying similar structures that originate from notably different starting points, which is a drawback to the structural parameter RMSD versus the Cartesian coordinate RMSD. The interface for these scripts is quite rudimentary, and along with expanding upon the acceptable structural parameters, a new interface for defining these parameters would likely be extremely helpful.

In **Chapter 4** the gas-phase structures of two protonated modified uridines are examined by IRMPD action spectroscopy and theoretical calculations. Although 2-thiouridine is found in naturally occurring tRNA, it is more commonly found with additional modifications at the 5-position of the nucleobase. Examining the gas-phase structures of these further modified 2-thiouridines would be of interest to understand if they further change protonation preference or preferred structure, which might help explain their importance at the wobble position.

In **Chapter 5** the impact of the 2′- and 2′,3′- sugar hydroxy moieties on protonated gas-phase nucleoside structure is examined through the arabinonucleosides and 2′,3′-dideoxynucleosides. This could be further extended to include a set of

3′-deoxynucleoside analogues, which would then provide a comprehensive understanding of the impact of the individual sugar hydroxy moieties. The 2′,3′-dideoxynucleosides are also a foundation for further synthetic modifications, many of which are used pharmaceutically as nucleoside reverse transcriptase inhibitors. Understanding how further modification at the 2′- and 3′- positions on the sugar moiety might impact gas-phase structure may help further elucidate the impact of sugar modification on nucleoside structure.

APPENDIX A SOFTWARE AND HARDWARE FOR THE GIBMS

A.1 Guided Ion Beam Tandem Mass Spectrometer (GIBMS)

The threshold collision-induced dissociation (TCID) analysis of energy-resolved collision-induced dissociation (ER-CID) tandem mass spectrometry data relies upon careful control over the kinetic and internal energy available to ions before and during a single collision with a gaseous atom or molecule. Careful control of these energies and modeling of the collision cross sections allows for determination of the threshold energy at which dissociation occurs along a specific pathway for an ion. The level of control over the kinetic and internal energy of an ion necessary to acquire particularly high-quality energy-resolved CID data over experimental parameters is not available with commercial instrumentation. To facilitate the high-quality ER-CID experiments necessary for TCID analysis a custom-built guided ion beam tandem mass spectrometer (GIBMS) is used, which is described in detail elsewhere, with a brief summary given below.[171-173]

Managing the internal energy of an ion is particularly challenging in the mass spectrometer. Therefore, controlling the influence of the ion optics on internal energy is key to the construction and operation of the GIBMS. Schematics of the GIBMS are shown in **Figures A.1** and **A.2** as a top-down drawing and a 3D rendered half-section view, respectively. Ions are generated using a custom-built electrospray ionization (ESI) source, which is described in greater detail in **Section A.2**. ESI is well known to be able to produce ions without depositing significant internal energy into their internal modes.[174,175] Ions are transferred from the ESI source into the first vacuum stage by a 5.5” long stainless steel capillary (0.020” inner diameter, ID) wrapped in a heating filament allowing for operation between ~100–125 °C, and biased at 20–50 V. Ions exiting this

capillary are captured by a radio frequency (RF) ion funnel.[176,177] DC potentials are applied to the first and last plates of the ion funnel with a resistor chain connecting the intervening plates to create a linear DC gradient along the funnel that draws ions towards the exit of the funnel. RF voltage is applied to the plates of the ion funnel with the voltage applied to adjacent plates being 180° out of phase. The frequency of the RF voltage applied is in the range of 0.6–1.2 kHz with an amplitude 10–30 V_{pp} to trap ions radially within the funnel. Ions exit the ion funnel into a RF hexapole ion guide that is surrounded by a collision cell, where they are thermalized by collisions with background gasses. Ions leaving the hexapole ion guide are collected and focused by a set of DC differential lenses (DL) that capture and focus the ion beam into the first stage of focusing and shaping optics (FS1). FS1 prepares the ions for the following magnetic sector momentum analyzer by converting the cylindrical symmetry of the ion beam to a ribbon shape and accelerating the ions into the magnetic sector. The use of DC lenses and the magnetic sector momentum analyzer are important for preserving the internal energy distribution of the ion population established in the hexapole ion guide, as RF optics or analyzers would introduce RF heating. Following mass selection of the ion of interest in the magnetic sector, the ion beam is reshaped to cylindrical symmetry and focused by another set of DC lenses (FS2) into an exponential retarder. The exponential retarder is comprised of a series of thin lenses with DC voltages applied that carefully slows the ions following the magnetic sector. Ions leave the exponential retarder into a set of DC focusing lenses (FS3) that inject the ions into an RF octopole ion guide. The DC-bias on the RF octopole ion guide is controlled to set the desired ion kinetic energy. The octopole ion guide passes through a collision cell where a low pressure of Xe gas (~0.05–0.20 mTorr) ensures that

multiple ion neutral collisions are improbable.[178-180] Xenon is chosen as the collision gas because it is monoatomic, heavy, and polarizable, leading to efficient kinetic to internal energy transfer during CID.[178,179] The octopole ion guide serves as an effective trap for ions in the radial direction, largely eliminating the loss of scattered reactant or product ions.[181] The product and remaining reactant ions exit the octopole ion guide into another set of DC focusing lenses (FS4) that focus and prepare the ions for injection into a quadrupole mass analyzer for mass analysis. Ions leaving the quadrupole enter a final stage of ion optics that focus the ions onto a Daly detector where they are detected by a secondary electron scintillation detector and standard pulse counting techniques.[182]

A.2 Electrospray Ionization

Electrospray ionization (ESI), as developed and popularized by John Fenn,[183,184] has contributed to a significant expansion in the types of samples readily available for analysis by mass spectrometry. Large biomolecules have particularly benefitted from the introduction of ESI to the toolbox of mass spectrometry ionization techniques.[185-188] The ability of ESI to produce ions with relatively low internal energies is especially important when producing relatively fragile ions. The soft nature of ESI is also particularly important to studying intact small molecules, as it greatly reduces the propensity for metastable fragmentation following ionization, and gives more control over the ion under investigation. With appropriate care it has been demonstrated that a well-tuned ESI source and interface can produce thermalized ion populations, an important consideration when making careful measurements of the intrinsic properties of an ion.[174,175]

A.2.1 Original GIBMS ESI Source

A previous student, Yu Chen, developed an ESI source for our custom-built GIBMS, similar in design to that developed by Moison et al., the details of both are found elsewhere.[174,189] This ESI source, shown in **Figure A.3**, greatly expanded the systems of study available to the GIBMS, from the predominantly metal-ligand complexes studied with the DC discharge flow tube source or protonated and radical cations of moderately volatile organic molecules with the microwave discharge flow tube source originally associated with the instrument, to small biomolecules in varied states of cationization. With this ESI source, thermalization previously achieved using the flow tube, is now accomplished with an RF ion funnel and hexapole ion guide/collision cell interface that is described in **Section A.1**.

The ESI emitter assembly was constructed on a triangular optical rail and platform system mounted to the source flange. An XYZ translation stage (Line Tool Co, model A RH-1/2” travel) was mounted to a custom platform, allowing for easy removal and coarse control of the position of the emitter assembly along the Z-axis. A custom shaft collar assembly was machined for the translation stage to hold a MicroTight® adapter (IDEX P-770). This union coupled the 0.063” plastic tubing from a syringe pump, typically fluorinated ethylene polypropylene (FEP) with an ID of 0.030”, to a 35-gauge stainless steel hypodermic tube and polyether ether ketone (PEEK) sleeve serving as the ESI emitter. An ESI voltage (1.7–3.0 kV) was applied to a custom copper collar with a tight slip fit in the PEEK MicroTight® fitting, with an 0-80 hole tapped in the radial direction to allow a screw to gently “pinch” the hypodermic tubing without deforming or piercing it, providing electrical connection directly to the ESI needle.

This ESI source greatly expanded the systems that could be readily examined on the GIBMS, with a shift from metal-ligand complexes to crown ether-amino acid complexes, nucleic acid monomers, base pairs, and more recently ionic liquid clusters.[190-195] However, after nearly constant use for several years, some major and minor issues became apparent. Key issues where quality of life could be readily improved included safety, signal stability, ease-of-use, and the management of environmental factors. Improvements to the full ESI source assembly, including the in-vacuo components, to target performance gains, are generally more expensive to address as they require more substantial machining and the purchase or modification of further ion optics. The subsequent **Sections A.2.2** and **A.2.3** describe two alternative ESI emitter assemblies that have been created for the GIBMS to try and address the less expensive quality of life improvements and provide opportunities for further expansion.

A.2.2 Prototype ESI Source

The first new emitter platform created for the GIBMS ESI source sought to begin addressing some of the safety and ease-of-use issues that became apparent after extended use, and offers the opportunity to more easily adapt the emitter to new ESI needles and sprayer assemblies. Several factors are likely responsible for the occasionally unstable spray observed with the original ESI source for the GIBMS and include: the relatively high analyte concentration (0.5–1.5 mM) required to generate usable signal, the sensitivity of the spray to the ESI voltage, and the highly dynamic air flow around the ESI needle. A new emitter platform was designed to address the safety concerns associated with the ESI voltage, and to facilitate greater exploration of spray generation and orientation.

A platform based upon the existing triangular optical rail from the original GIBMS ESI source was designed as both a functional replacement and test bed for exploring more emitter positions than the original emitter assembly gave access to. This platform was based upon the plate to which the XYZ translation stage mounts in the original GIBMS ESI source. The two arc slots machined into this new plate accommodate a 0.1875” pin and 8-32 binding head screw, which guide and secure a 0.5” square aluminum bar that serves as a rail for the ESI emitter, and the full assembly is shown in **Figure A.4**. This plate mounts to a custom carriage machined for the original GIBMS ESI source that was modified slightly to remove interferences with the protruding pin and screw that secure the aluminum rail to the plate. A custom H-channel holder was machined from a 0.625” thick Delrin® bar to slide and clamp onto the aluminum tail, into which a stainless steel zero dead volume union from IDEX (part #: U-402) is set. Tapped holes were added to allow for 6-32 screws to secure the position of this holder on the aluminum rail and to secure and electrically connect the ESI voltage to the stainless steel union. The ESI voltage is connected to the analyte solution through the stainless steel union, allowing for removal of the custom copper collar and 0-80 hardware. This change in application of the ESI voltage not only removes the chance of pinching and ruining the ESI emitter by the 0-80 screw, but also greatly improves the robustness of the electrical connection, as the coaxial cable now has little to no risk of disconnecting accidentally. This also greatly improves the ease with which the ESI emitter is assembled. Machine shop drawings of the new or modified components of this source are given in **Figure A.5–A.7**

To allow for increased decoupling of the electrospray from its proximity to the stainless steel capillary interface, which acts as the counter electrode in the original ESI

source, a secondary counter electrode was fabricated. A stainless steel tube and custom Delrin® holder were designed and fabricated to mount to the aluminum rail that also holds the ESI emitter assembly. The counter electrode voltage can then be applied to this electrode to decouple the electrospray from the capillary interface, allowing the ESI emitter to be located much farther from the capillary interface without greatly increasing the ESI voltage.

A simple pneumatic nebulization assembly was also designed, depicted in **Figure A.8**. A longer (≥ 3”) ESI needle was fabricated from 35 gauge hypodermic tubing and connected to the syringe pump via stainless steel union as described previously, also serving as the electrical connection to the ESI voltage. The ESI needle passes through a 0.063” OD X 0.007” ID PEEK sleeve connecting this union to a Swagelok stainless steel tee (part #: SS-100-3), which the needle also passes through. A stainless steel tube, 0.063” OD X 0.030” ID”, is attached to the tee, coaxial to the needle, cut to allow ≤ 0.1” of the ESI needle to extend beyond it. A nitrogen gas flow is connected to the orthogonal port of the tee and flows coaxially to the needle within the stainless steel tubing. The gas flow can be controlled manually by a needle-leak valve, or by a spare MKS mass flow controller (Model 1159B).

Testing of this pneumatic nebulization prototype revealed that a spray could be formed without voltage, although creation and maintenance of a spray was difficult. The spray produced droplets that appeared larger than those generated by electrostatic nebulization alone, as observed on a camera focused on the capillary inlet. Application of a voltage to the stainless steel union or the secondary counter electrode with a grounded union also produced a visible spray, with greater ease and consistency. However, the

applied voltage had little to no impact on the ion signal observed at the detector, which was about three orders of magnitude lower than the ion signal expected from the electrostatic ESI source. Another significant problem, for which a solution was not found, arises in the form of rapid vibrations of the ESI needle inside the coaxial tube, the likely culprit of the visual fluctuations observed in the spray. Use of 32-gauge hypodermic tubing as the ESI needle, to stiffen the emitter, was pursued as the material was readily available, however it was not observed to dramatically reduce this vibration. This issue is likely compounded by difficulties in adapting the needle fabrication method to longer needles, and challenges in the assembly of the pneumatic emitter assembly.

Although the prototype pneumatic nebulizer was fraught with challenges and didn't appear to offer any performance improvements without further development, the increased ease-of-use and more robust electrical connection led to continued use of this ESI platform as a purely electrostatic ESI source. Although this design lacks the precise X,Y,Z positional control of the original ESI source, the distance and angle control provided by the aluminum rail and arc slots are sufficient to achieve comparable results. A key point of failure was discovered only after extended use; the walls of the Delrin® union holder are too thin to effectively hold the 6-32 threads used to secure the union in the channel. An extra hole was drilled and tapped to secure the union. The electrical connection was changed to a spring clip from an IDEX insulating mounting bracket (part # M-447) to decouple securing the union from the electrical connection. The holder itself is relatively inexpensive and as such replacement is not prohibitive. The simplicity and relatively low cost of this design also make it relatively inexpensive to maintain, despite its shortcomings. Drawings of each component are available in the Appendix.

A.1.3 Enclosed ESI Source

As mentioned in **Section A.2.1**, one of the issues that began to arise with the original GIBMS ESI source involved environmental factors. These arose in several forms, such as inconsistent airflow from a nearby air conditioning vent, to people walking near the exposed spray, each having visible impact on spray stability. Although these issues are important to consistent ion signal, isolating the spray environment such that it can be flushed with an inert gas, such as N_2, also provides a more favorable environment for easily oxidizable samples. To address this issue and continue to improve the safety and ease-of-use of the ESI emitter, an enclosed ESI source, loosely based upon the design of McNary et. al,[196] was developed. To speed up development and substantially reduce the cost of development and manufacturing, 3D printing was chosen to both prototype and fabricate as much of this source as was amenable to the method.

The 3D printed components of this source were designed within Fusion 360 (Autodesk, San Rafael, CA, USA), and then 3D printed via fused deposition modeling (FDM) on a Prusa i3 mk2 3D printer (Prusa Research, Prague, CZ). Polyethylene terephthalate glycol (PETG) was chosen as the 3D printing material due to its availability, chemical resistance, and strength. Printed components were exported from Fusion 360 to STEP files via the built-in 3D printing utility, then imported into Simplify3D where they were sliced with 0.2 mm layer heights and between 30% and 60% infill, and sent to the Prusa i3 mk2 printer using the default preset temperatures for PETG of 240 °C for the hotend and 90 °C for the bed. Typically, three outside layers were used so that material could be removed if necessary for assembly with the non-printed components. Support structures were generated automatically by Simplify3D where necessary, although in

some cases custom support structures were designed with the components to produce more even surfaces where the automatically generated support material allowed excessive drooping on overhanging surfaces that contacted non-printed parts in the assembly.

A section view of this source is shown in **Figure A.9**. Aluminum standoffs were machined from 0.5” aluminum round bar to offset the 3D printed assembly from the existing source flange, replacing the 3/8”-16 bolts that had mounted the capillary interface to this flange, and providing a fixed distance from the emitter assembly to the capillary inlet. A cast acrylic tube (4.0” OD, 3.5” ID) was cut to length to span the gap between the source flange and ESI emitter providing an adequate seal to the external environment while maintaining the ability to visually monitor the spray. Machine shop drawings of every component of this source are shown in **Figures A.10–A.15**

A primary goal of this ESI source was to allow nearly the same amount of movement in the X,Y, and Z directions as the original GIBMS ESI source allowed, and maintaining the parallel orientation of capillary inlet and ESI needle in order to retain the original functionality but improve isolation of the spray from the environment and the user from the ESI voltage. The main body is attached to the ESI source flange by four aluminum stand offs and the intervening space is surrounded by a cast acrylic tube. The main body provides a platform for a carriage to travel in the XY plane up to 0.1875” in any direction. Motion along the Z axis is precisely controlled by an optical lens tube from Thorlabs (part #: CML25 and CMV10) with the fine thread (1.00”-32). A separate carrier for the stainless steel union was designed to fit within this lens tube and allow for the ESI voltage to connect via spade terminal, 4-40 nut and set screw. A spring is placed within

the lens tube, maintaining contact between the union holder and lens tube to ensure that motion along the Z axis is smooth in both directions. Motion of the carriage in the XY plane is controlled by nylon thumbscrews and 10-32 nuts that are set into the 3D printed housing during the printing process. Although threads can be modeled during the printing process, or cut into the plastic after printing, the use of steel nuts is an inexpensive way to ensure smooth and reproducible motion, with less risk of damage due to misuse. Flat planes are designed in the XY carriage, inset slightly into the body to effectively trap the nylon thumbscrews in the Z-axis, holding the XY carriage flush to the body to reduce airflow into and out of the enclosed volume. A guide for the needle is also 3D printed, which aligns with the union holder via a 3D printed hexagonal keyway, to ensure alignment of the needle with the Z axis and provide a degree of protection against bending, which can easily destroy the 35-gauge needles, when inserting the union and needle into the assembly.

Assembly of this ESI source onto the source flange is slightly more challenging than either of the ESI sources described in **Section A.2.1** or **A.2.2** with more pieces requiring a relatively tight fit and assembly in a single step. Cleaning the source is also more challenging as removing the sprayer from proximity to the capillary inlet requires removal of either the sprayer assembly or the carriage it rides in, and access to the interior of the acrylic tube is also somewhat restricted. However, adjustment of spray position is more precise and straightforward than the prototype ESI source described in **Section A.2.2**, and the sprayer assembly is safer and more robust. Despite the greater challenge in servicing compared to the previous ESI sources, maintenance is still quite rapid. The performance of this ESI source is not markedly different from either previous iteration,

with no notable changes in sprayer geometry, inlet capillary, or the in-vacuum assembly. This source does however allow for more reproducible and robust positioning of the sprayer assembly, removes concerns regarding an unstable spray environment, and is safer to use than the original GIBMS ESI source with regards to the ESI voltage.

A.3 Data Acquisition Software

Acquisition of TCID data on the GIBMS is facilitated by computer control of the Xe gas valves, reaction cell pressure gauge, quadrupole mass analyzer, octopole DC-bias power supply, and a digital counter/timer. The original data acquisition software for the GIBMS was developed by Dr. Cliff Frieler in Fortran 90, based upon the Fortran data acquisition software originally developed by Armentrout et al.[106,173] However, an issue arose during the summer of 2014 involving the RF generator for the octopole ion guide in the reaction region of the GIBMS. This issue was traced to rapid changes in voltage coming from the octopole DC-bias power supply, generally when sent a command by the data acquisition software. This rapid change in octopole bias voltage resulted in instability in the octopole rf voltage, which severely impacted ion transmission, to the extent that data collection became nearly impossible. This led to replacement of the computer-controlled octopole bias power supply.

Unfortunately, updates to the GPIB specification used to communicate with the original power supply (Kepco BOP 100-1M bipolar linear power supply and BIT 488B interface card) meant that a drop-in replacement was not commercially available. A newer version of this power supply/interface was acquired (Kepco BOP 100-1M, with a BIT 4886 interface card), with the corresponding device drivers. To incorporate this new power

supply, and substantially update the computer hardware and software interface, new data acquisition software was developed.

A.3.1 Original Data Acquisition Software Overview

The original Fortran 90 data acquisition suite provided two separate programs to perform different scanning modes. MSCAN, or *mass scan*, performed a straightforward tandem mass spectrometry experiment to determine the specific masses for precursor and fragment ions to be used during later data acquisition. EMP, or *energy, mass, and pressure*, served as the primary data acquisition tool for ER-CID experiments on the GIBMS. EMP scanned the collision energy, quadrupole mass analyzer, and Xe gas pressure in the foreground (collision cell) and background (chamber) to acquire precursor and fragment ion intensities as a function of collision energy, respectively.

Nearly every function/feature of MSCAN and EMP is required for rapid and robust acquisition of high-quality TCID data. Therefore, the function of each of the three distinct scanning modes present between MSCAN and EMP needed to be replicated exactly. During the mass scan performed by MSCAN the collision energy is set to a desired value and Xe gas is introduced to the reaction cell at around ~0.20 mTorr to facilitate fragmentation. The quadrupole mass filter is then scanned to provide mass analysis of the precursor and any fragment ions produced. This tandem mass spectrum provides important m/z information for identifying the fragment ions produced. This m/z data is also important for selection of the m/z vlues that will be acquired in the later data scans as representative of the precursor and fragment ions.

EMP is performed in two distinct scanning modes. The first scan mode, a so-called *E_0 scan*, is responsible for measuring the kinetic energy distribution of the precursor ion

beam prior to data acquisition. The quadrupole mass filter is configured such that the m/z value corresponding to the precursor ion, as determined by MSCAN, is transmitted to the detector. Xe gas is then introduced to the background (the reaction chamber), or not at all, to prevent precursor ion dissociation or scattering. Finally, the collision energy is scanned from a retarding potential, where no precursor ions pass through the octopole ion guide to the detector, until all precursor ions have sufficient kinetic energy to reach the detector. Acquisition of this energy-dependent mass spectrum allows for determination of the kinetic energy distribution of the precursor ions as they enter the reaction region. This step is important for inclusion of this kinetic energy distribution into modeling of the final TCID data. Knowledge of the kinetic energy distribution is critically important to the data analysis, but is also useful for evaluating tuning conditions so that the most appropriate choices can be made with regard to data acquisition.

The final scanning mode, also performed by EMP, is known as a *data scan*. This scan mode is responsible for acquisition of the primary TCID data. In a *data scan*, the collision energy is scanned over a broader range than in an *E0 scan*, typically from a value that results in no fragmentation until the observed reactivity levels off. Xe gas is introduced to the collision cell (foreground) and the reaction chamber (background) in subsequent scans, at a low pressure typically in the range of 0.03–0.20 mTorr to minimize the likelihood of any ion undergoing more than a single collision. A complete *scan* comprises two distinct scans, a foreground scan in which Xe gas is present in the collision cell, and a background scan with Xe gas directed to the background (reaction chamber). Although not critical, we have chosen to minimize time delays and hysteresis in the collision energy set by scanning unidirectionally during each scan, from low (or retarding)

potentials to high collision energies in the foreground scan and from high to low collision energies in the background scans. The pressure in the collision cell is measured before each foreground and background scan. The pressure is repeatedly measured until it stabilizes (typically ~15 seconds to stabilize within 0.002 mTorr) which initiates the next scan. At each collision energy examined, a mass spectrum is acquired at the m/z points selected based on the MSCAN, such that no time is wasted collecting data where no signal exists or where the signal for a given ion is less intense. The data are averaged and saved after each scan to allow evaluation of the data during acquisition and to prevent data loss should a catastrophic event terminate acquisition prematurely.

A.3.2 Development of MS-EMP

To simplify and unify the user interface, a single program combining the functionality and controls of MSCAN and EMP was developed. This choice also provides the opportunity to build a framework in which the addition of instrument hardware-capabilities, such as ion mobility, can be more readily accommodated while maintaining a consistent interface. Representing the combination of MSCAN and EMP, new data acquisition software was designated as MS-EMP. MS-EMP is built on a state machine-like architecture in LabVIEW™. The state machine architecture[197] was chosen as it is relatively straightforward to understand and follow for those with little to no programming experience. The state machine architecture is combined with a tab control for easy compartmentalization of the controls for different scanning modes. In hindsight, an alternate method of compartmentalization should have been pursued to limit the rate at which block diagram complexity is increased as more scanning modes and controls are

added. Unfortunately, more elegant solutions, such as subpanels appeared too daunting to implement at the time.

An overview of the MS-EMP front panel is shown in **Figure A.16**. Data is plotted on a standard XY graph control, the formatting of which is largely programmatically controlled. However, the standard zoom, pan, and selection controls of a LabVIEW graph control are retained. The legend to the right of the graph automatically populates with labeled graph channels depending on the scanning mode that is active, and the way that the scan is configured. Universal scan controls such as ‘start’ and ‘stop’ are located in the bottom left with file handling variables that are used for all scanning modes. The scan controls ‘halt’ and ‘abort’ both cause the currently active scan to stop. However, ‘halt’ waits until the end of the current scan and retains the data collected during that scan, whereas ‘abort’ stops the scan as soon as possible and discards the data collected during the aborted scan. The ‘operator’ variable is selected via pop-up dialog box upon program start. The ‘operator’ variable selects the researcher appropriate directory in the data structure, with the ‘file path’, ‘file name’, ‘file suffix’, and ‘scan suffix’ variables used by the different scanning modes to generate unique and informative file names. Both ‘file suffix’ and ‘scan suffix’ increment automatically based upon scan behavior and scan mode. Controls and indicators in the bottom right status region are also largely shared between the scanning modes. Linear or logarithmic scaling for the graph axes is controlled via buttons for ease of use. Scan progress, monitored via a queued message handler,[198] is displayed with the relevant scan variables such as the current values of the collision energy, m/z, intensity, and reaction cell pressure. Also included in this section are controls for selecting the precursor and product ions via the graph cursor, only active when *mass*

scan is the active scan mode, which allow for more rapid population of the configuration information for a *data scan*.

The loop structures of the three scanning modes, *mass scan*, E_0 *scan*, and *data scan*, are readily compartmentalized into their own subVI, with further subVI handling tasks such as file handling, data management, and communication with the instrument. Each scan mode also has one or two tabs on the GUI populated by their relevant controls and indicators. These controls are part of the main front panel, and their values are passed into the corresponding scan subVI when a scan is started. To decouple GUI performance from scan performance, the scanning subVI are called asynchronously on demand.[150] However, communication between the scanning subVI and the GUI for the purposes of scan controls such as halting and aborting a scan, as well as displaying scan status on the GUI, is necessary. Communication between the front panel and asynchronously run scanning subVI is therefore handled by queued message handlers[198] which are initiated by the main subVI upon initialization and passed into the scan subVI with the rest of the scan configuration.

A.3.3 Communication Interfaces

An overview of the communication interfaces used by MS-EMP to communicate with the GIBMS is shown in **Figure A.17**. A pair of PCI cards are installed in a custom-built Windows 7 workstation to handle communication with several components of the GIBMS. A National Instruments PCI-GPIB card handles communication over GPIB with a Kepco BOP 100-1M/BIT 4886 DC power supply that provides voltage for the collision energy and a Canberra 2071A dual counter timer for acquisition of ion intensity from the Daly detector. A National Instruments PCI-6221 data acquisition (DAQ) PCI card with a

BNC-2110 breakout box controls the gas-control valves that direct the Xe gas into the collision cell or reaction chamber, reads the pressure from a MKS Type 270 signal conditioner, and sends serial commands to a custom digital-to-analog converter (DAC) that commands the quadrupole mass controller.

The input and output pin assignments of the PCI-6221 DAQ are listed in **Table A.1**. A custom driver communicates with the custom DAC via serial commands which then commands the Extrel C60 quadrupole controller by a 0–10V signal. Two digital pins, port 0/line 0 and port 1/line 1 are connected to the User 1 and User 2 connectors on the BNC-2110 breakout box. This provides backwards compatibility with the BNC cables connected to the relays that control the Xe gas control valves for the foreground (reaction cell) and background (reaction chamber). An analog input, AI 0, from the DAQ is connected to a MKS Type 270 signal conditioner (MKS Instruments, Andover, MA, USA), reading a 0-10 V signal corresponding to the pressure in the reaction cell in mm Hg, measured by a MKS Baratron 690A capacitance manometer. The BIT-4886 control card of the Kepco BOP 100-1M power supply has readily accessible GPIB drivers available through the Kepco website.[199] However, the Canberra 2071A does not have readily accessible drivers and as such the specific functions required were developed from scratch using the native GPIB communication subVI of LabVIEW™.

A.3.4 Mass Scan Interface

The *mass scan* interface, shown within **Figure A.16**, handles the configuration of several mass analyses simultaneously. The ‘channel’ control selects the active data plot on the graph, can create new tandem mass spectrum scan configurations and manages the display configurations of previous tandem mass spectra. The ‘voltage’ control

determines the octopole DC-bias during the tandem mass spectrum acquisition. The m/z controls and '# of masses' determine the m/z array that is scanned by the quadrupole mass filter during the tandem mass scan. Within the acquisition of a tandem mass spectrum there are two distinct ways to configure the acquisition of multiple averaged scans. When the 'scan mode' button is set to 'accumulate', the 'scans' control determines the number of individual tandem mass cans accumulated before automatically stopping. If the scan mode is set to 'continuous', scans will continue accumulating and averaging until the scan is either 'halted' or 'aborted' manually via the scan controls. In either of these modes, more scans can be accumulated and averaged into the tandem mass spectra via the 'resume' scan control. Saving data to a file with *mass scan* as the active scanning mode saves all of the acquired tandem mass analyses to a single data file. The 'save' button for each channel is used to determine if the selected channel is saved to this comma delimited data file. A Microsoft Word template is available for automated report generation using the built-in LabVIEW™ subVI. This report contains the averaged tandem mass spectra, their scan configurations for record keeping, and any further comments incorporated by the user. As mentioned in **Section A.3.3** the 'select precursor' and 'select product' buttons are only active while *mass scan* is the active scan mode. In conjunction with the built-in cursor functionality of the XY graph, these buttons enable the selection of m/z values in the active tandem mass spectrum as representative m/z ratios for the corresponding precursor or product ion intensities and automatically configures these points in the 'data configuration' tab for use during a *data scan*.

A.3.5 E0 Scan Interface

The *E_0 scan* interface is split across two tabs shown in **Figure A.18** to differentiate the controls that are relevant to scan configuration and those relevant to scan analysis. On the ‘E0 Scan’ tab, the collision energy controls, parallel to the the tandem mass scan controls in the *mass scan* interface, are used to establish the array of collision energies scanned during an *E_0 scan*. Only a single m/z is collected by the quadrupole mass filter during an *E_0 scan*. Therefore the ‘observed m/z’, determined manually or from a prior *mass scan,* and the entered ‘charge’ value are used to convert the collision energy array to a voltage array for commanding the octopole DC-bias power supply. The ‘observed m/z’ is also used to set the quadrupole mass filter. The ‘exact mass’ and ‘neutral’ inputs are used to store the exact masses of the reactants for inclusion in the saved data file. The ‘neutral’ input is typically the exact mass of the collision gas, and the ‘exact mass’ is the exact mass of the precursor ion, which is not necessarily the m/z at which data is acquired as determined from via a mass scan. The same ‘continuous’ or ‘accumulation’ scan modes found in *mass scan* are also available in *E_0 scan*, although the default for *E_0 scan* is ‘continuous’ to encourage collection of better-quality data as the data acquired in an *E_0 scan* is directly relevant in the subsequent modeling of the *data scans*.

The ‘Fit E0’ tab is only active after a completed E0 scan has been collected and contains the array of energies scanned (X-axis). Upon switching to the ‘Fit E0’ tab a subVI attempts to fit a Gaussian peak shape to the first derivative of the acquired E_0 data. However, noise either before or after the linear part of the data can make this automated fit challenging. Therefore, the array of collision energies collected during the *E_0 scan* are populated in the ‘energies’ array control. The ‘start fit’ and ‘stop fit’ controls allow for

selection of a specific range of collision energies across which to fit the Gaussian peak shape, allowing for a better fit to the first derivative. The resulting center of the Gaussian peak, E0, and full-width at half-maximum (FWHM) are displayed.

The data of an *E_0 scan* is saved in a custom file-format for loading into the data acquisition software. This also includes the determined E_0 and FWHM from the Gaussian fitting procedure. A Microsoft Word template is also available for *E_0 scan* to record the scan configuration, E_0, FWHM and comments to a report for easy record keeping. An *E_0 scan* is an important preparation step for a *data scan* from the perspective of data analysis. Therefore, a feature added to MS-EMP that was missing from the original EMP data acquisition software is the requirement for an *E_0 scan* with the current 'file suffix' to be present before a *data scan* can be configured. This ensures that each *data scan* has associated with it a recent assessment of the kinetic energy characteristics of the precursor ion beam.

A.3.6 Data Scan Interface

As described for an *E_0 scan* in **Section A.3.6**, the controls for the *data scan* mode are separated into two tabs, 'data scan' and 'data configuration', both shown in **Figure A.19**. This is due to the more involved configuration of the two distinct variables scanned during a data scan, and issues with the interface scaling to different screen sizes. The 'data scan' tab configures scanning of the collision energy and the number of individual scans to average, while also collecting important information for the resulting data file, such as the 'neutral' mass and 'exact mass' of the precursor ion, similar to an *E_0 scan*. Configuration of the collision energy scan for a *data scan* is more complex than the energy or mass scanning in the other two scan modes. A conventional linear scan is readily

configured using the ‘data initial’, ‘data increment’, ‘data final’, and ‘# of points’ controls, typically with ~200 points per scan. However, certain experiments might benefit from greater control over spacing in certain ranges, generally tighter energy point spacing near the threshold for dissociation. The ‘dual’ button allows for configuration of a two-part energy scan, with finely spaced energy points from ‘data initial’ to ‘dual initial’, corresponding to the ‘data increment’ control, and coarsely spaced energy points from ‘dual initial’ to ‘data final’ corresponding to the ‘dual increment’ control. The ‘log e’ button configures a logarithmic energy scale instead of a linear scale, achieving a similar result to the ‘dual’ configuration. Both alternate configurations for the collision energy scan generally abide by the preference for the simple linear scan for ~200 points per scan. The ‘continuous’ and ‘accumulate’ modes are both available for configuration prior to a scan, the same as for *mass scan* and *E_0 scan*, but like an *E_0 scan*, the default is ‘continuous’ to encourage more averaging and the acquisition of higher-quality data.

The ‘data configuration’ tab configures scanning of the quadrupole mass filter at each collision energy. There are several ways to populate this configuration, from full manual entry, to selection of precursor and product ions from prior mass scans. The precursor ion intensity is always collected and is always the first item in the list on the left. Below this is an array of product ions to be collected during the data scan. The array to the right stores all of the ions that have been configured for the current system, whether they are collected during that specific scan or not. The observed m/z, color and symbol, of these fragment ions are stored such that the corresponding button can be selected, and that configuration will automatically be added or removed from the array of fragment ions to be collected during the data scan.

The file system variables ‘file suffix’ and ‘scan suffix’ are particularly important for *data scan*. As mentioned in **Section A.3.3**, these variables are used with ‘file name’ to generate a unique file name that indicates the number of data acquisitions performed previously that day, but also the individual scan within that data acquisition. For instance, during a TCID experiment on the GIBMS, *data scan* is used to acquire data under several Xe pressures in the collision cell to facilitate extrapolation to zero pressure and single collision conditions. Acquisition at each pressure is a single data acquisition, with its own E_0 *scan* and *data scan* and associated ‘file suffix’, which increments from ‘aa’ to ‘ab’, on to ‘zz’, which allows for easy association between a data file and a Xe pressure and specific data acquisition. Each data scan within a data acquisition is saved, incrementing the ‘scan suffix’ for the next scan, allowing for the correction of scan issues during data analysis.

Table A.1 List of ports and lines of the PCI-6221 DAQ and BNC-2110 breakout box and their associated functions.

BNC-2110			
Port Name	**Signal Type**	**Signal Name**	**Purpose**
port0/line0	Digital (5V)	Chamber Out	Signals relay, which controls gas valve
port1/line1	Digital (5V)	Cell Out	
ai0	Analog (0–10V)	Baratron In	Read baratron pressure
port0/line5	Digital (5V)	Sin	Serial communication with custom DAC for quadrupole control
port0/line6	Digital (5V)	CS	
port0/line7	Digital (5V)	LDAC	

Figure A.1

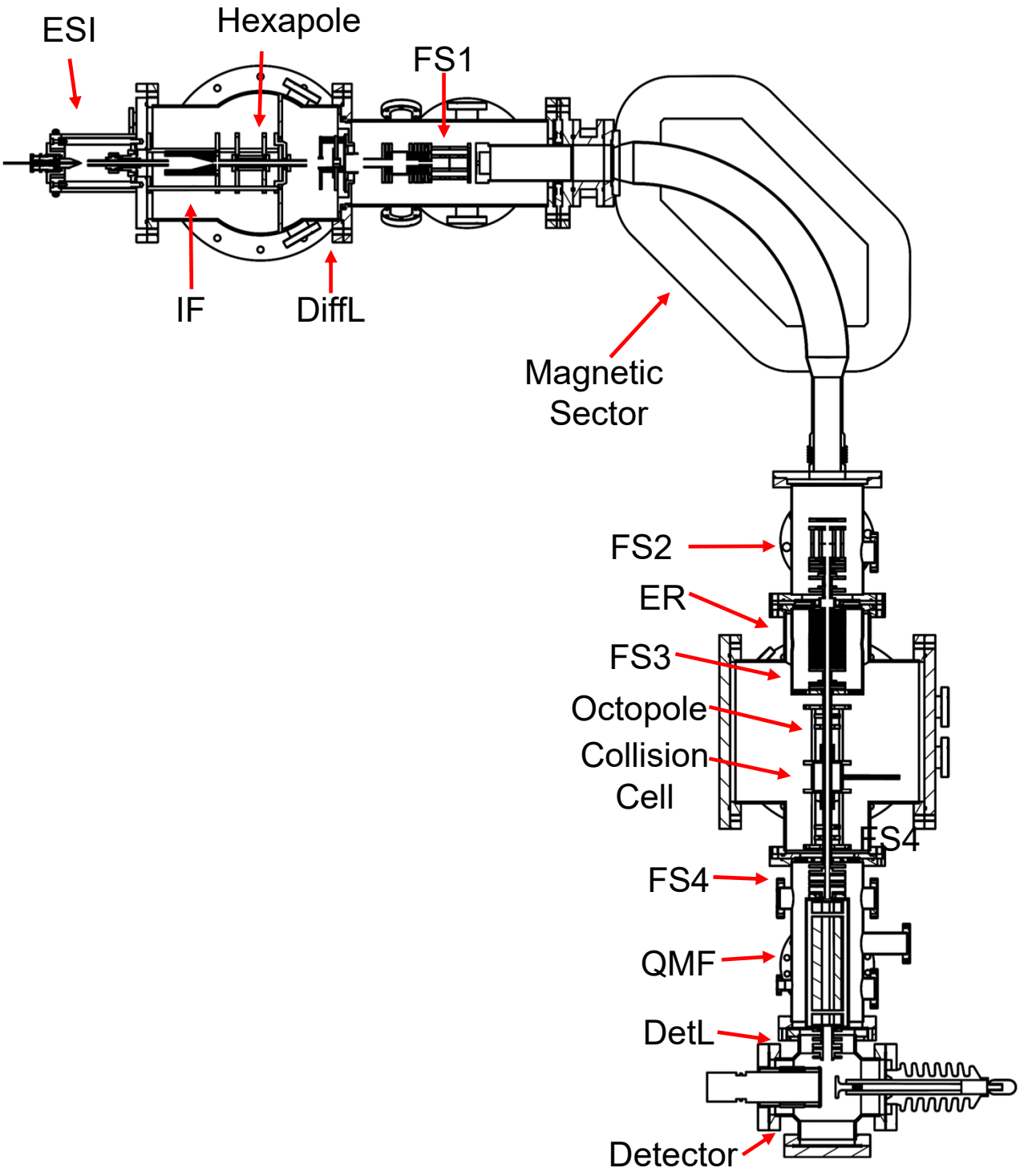


Figure A.1 Schematic of the custom-built guided ion beam tandem mass spectrometer (GIBMS) located in the Rodgers laboratory. Shown are the ESI source, ion funnel (IF), hexapole ion guide, differential lenses (DiffL), focusing stages 1, 2, 3, and 4 (FS1-4), an exponential retarder (ER), octopole ion guide, collision cell, quadrupole mass filter (QMF), detector lenses (DetL), and a Daly detector.

Figure A.2

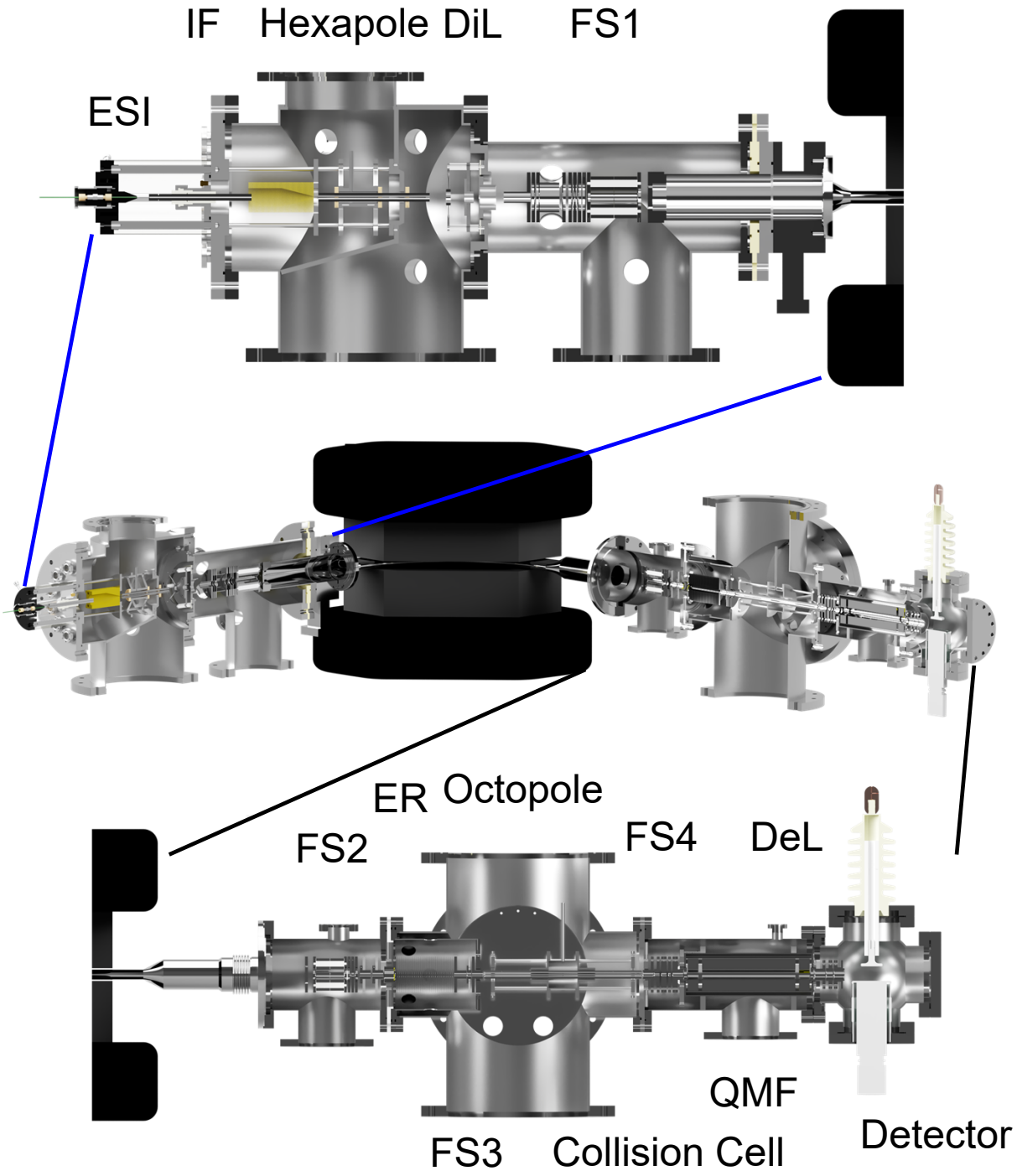


Figure A.2 3D renderings of a section view of the GIBMS. Shown are the ESI source, ion funnel (IF), hexapole ion guide, differential lenses (DfL), focusing stages 1, 2, 3, and 4 (FS1-4), an exponential retarder (ER), octopole ion guide, collision cell, quadrupole mass filter (QMF), detector lenses (DeL), and a Daly detector.

Figure A.3

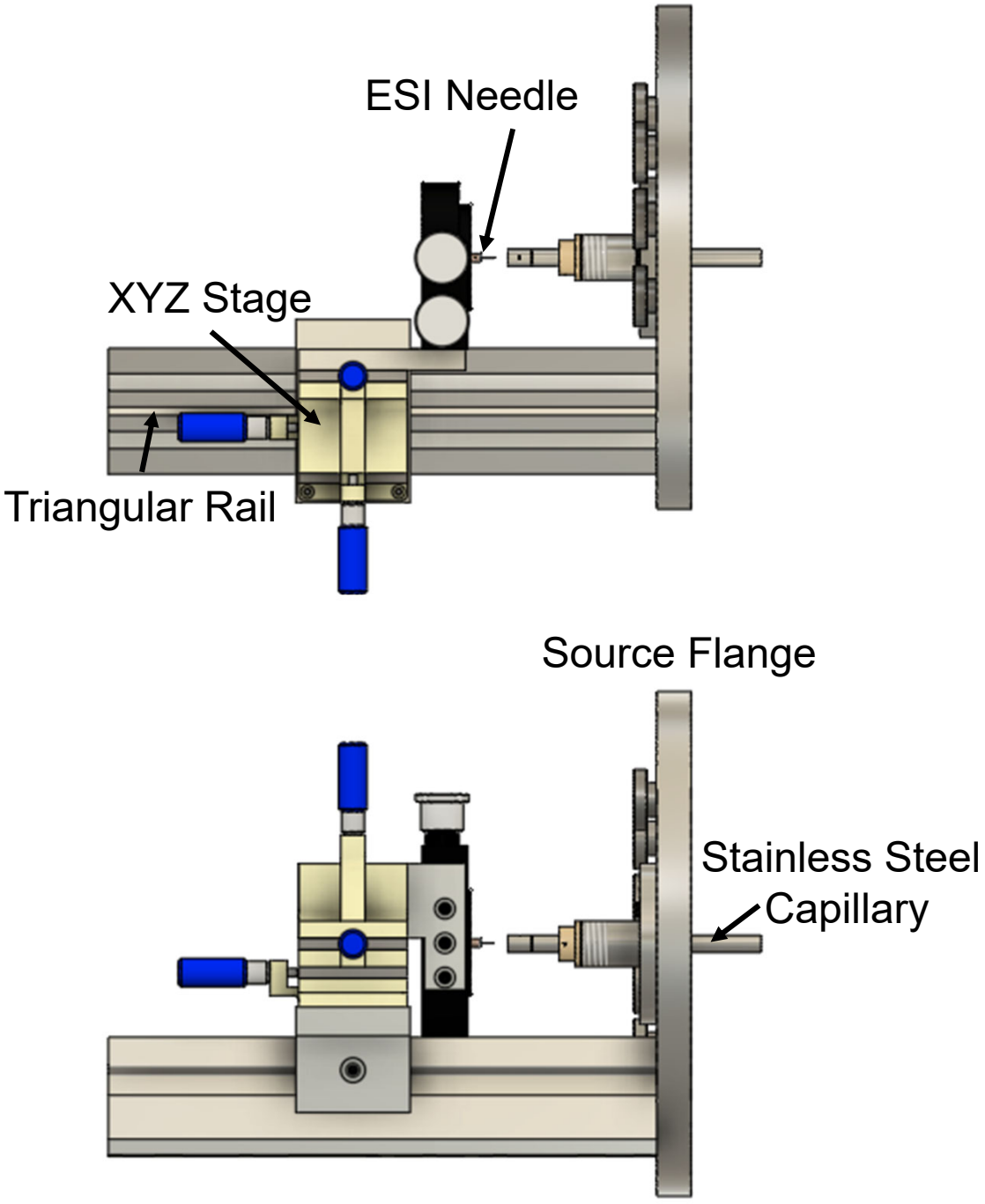


Figure A.3 Schematic of the original GIBMS ESI source designed and constructed by Chen, Y. et. al. in reference 184, and based upon the designs of Moison, B. et. al. in reference 169. A triangular optical rail and custom mount support an XYZ translation stage and a 2” optics holder. A MicroTight union is mounted in the optics holder to connect the ESI needle to the solution flow from a syringe pump.

Figure A.4

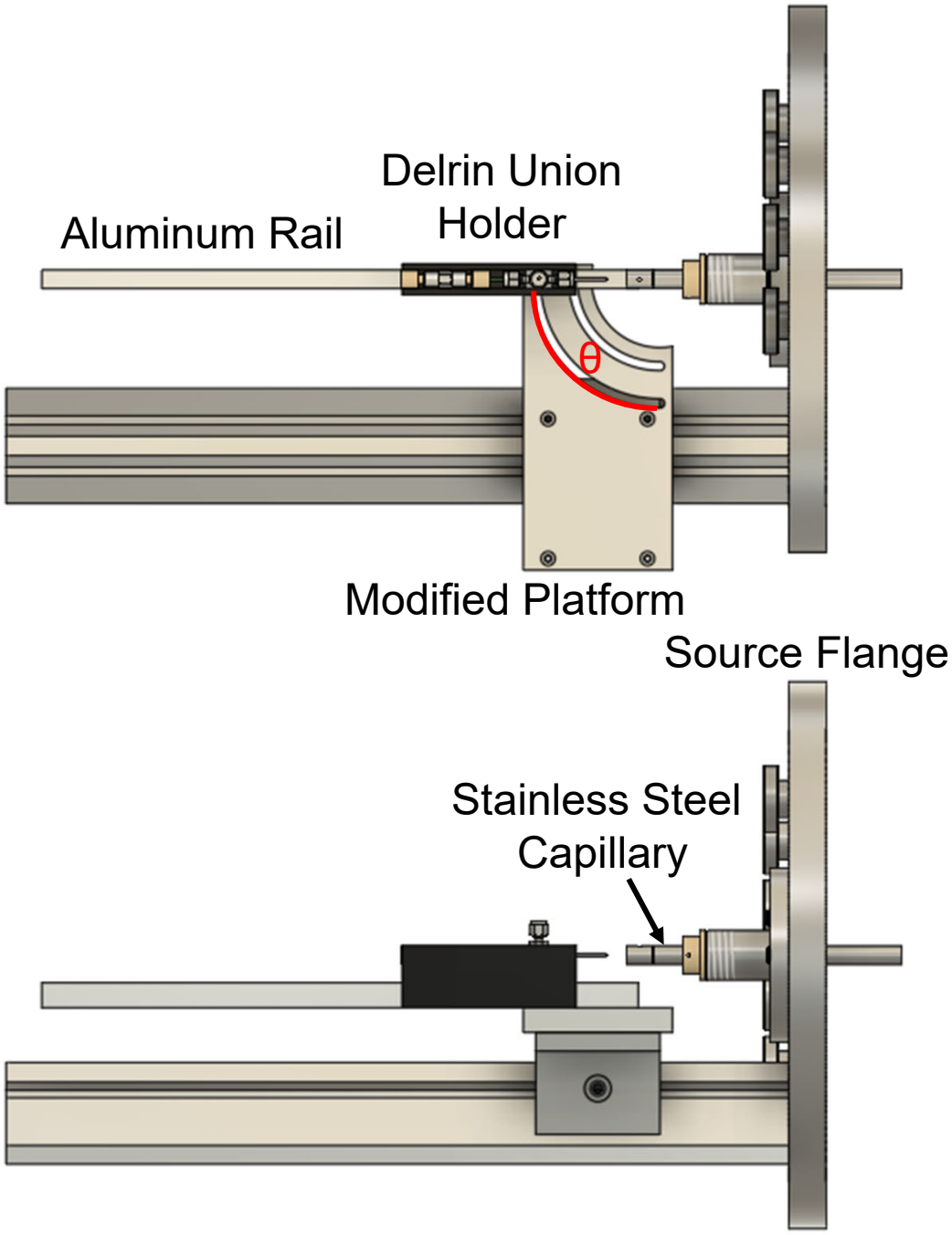


Figure A.4 Schematic of the first prototype ESI source to further explore angle and position of the ESI emitter and improve safety when in use. The same source flange and triangular rail as the original GIBMS ESI source are used. The triangular rail mount is modified with a new platform that supports an aluminum rail and a Delrin H-channel holder. A stainless steel union connects an ESI needle to the solution from the syringe pump. A proposed pneumatic nebulizer is also shown.

Figure A.5

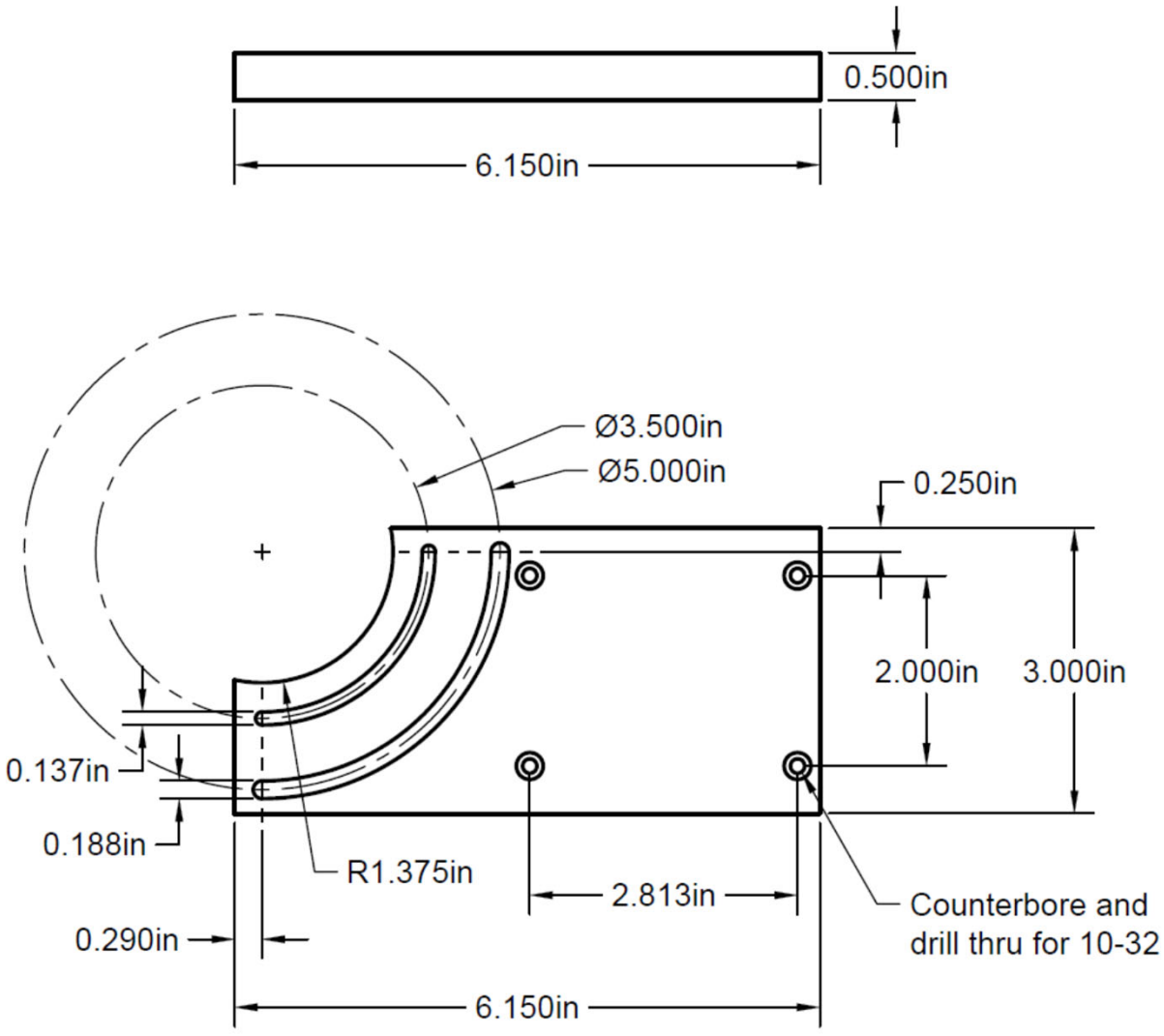
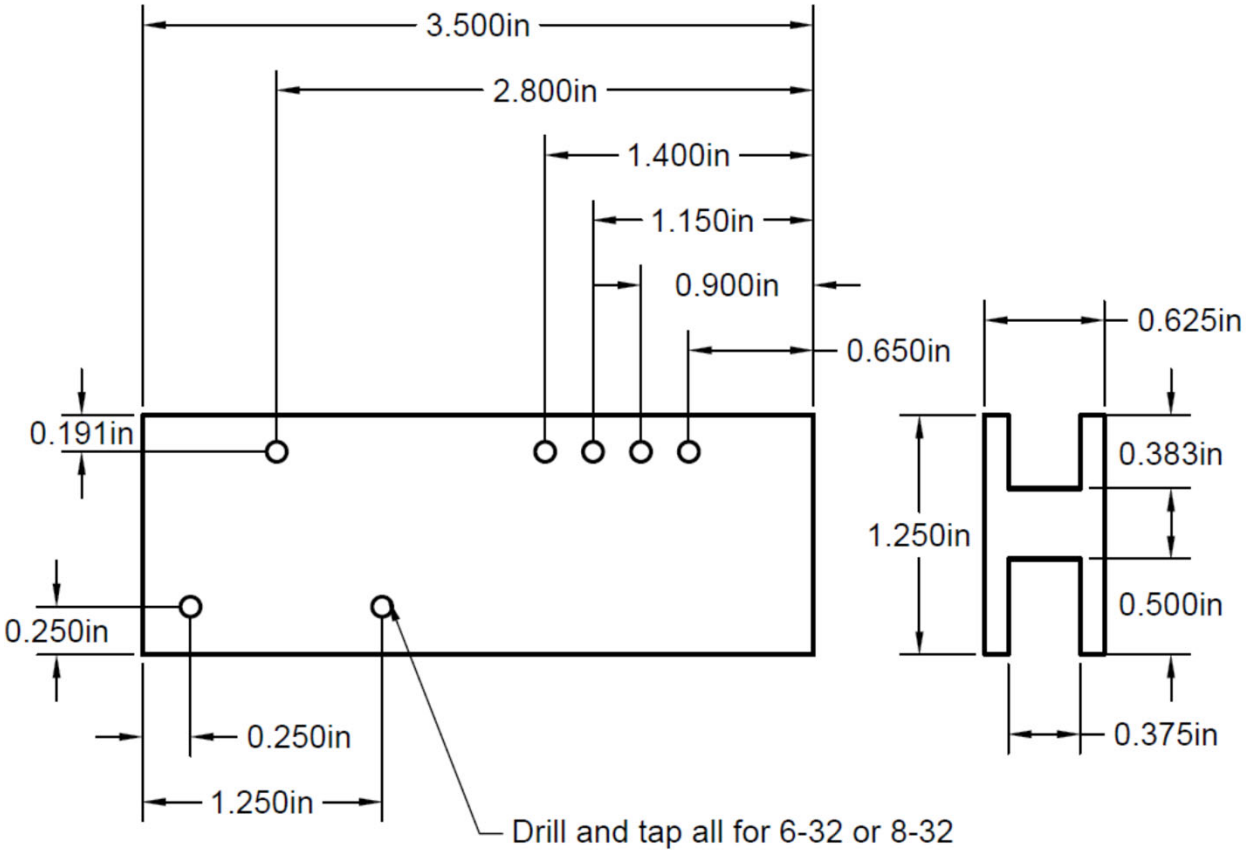


Figure A.5 Machine shop drawing of the prototype ESI source modified baseplate.

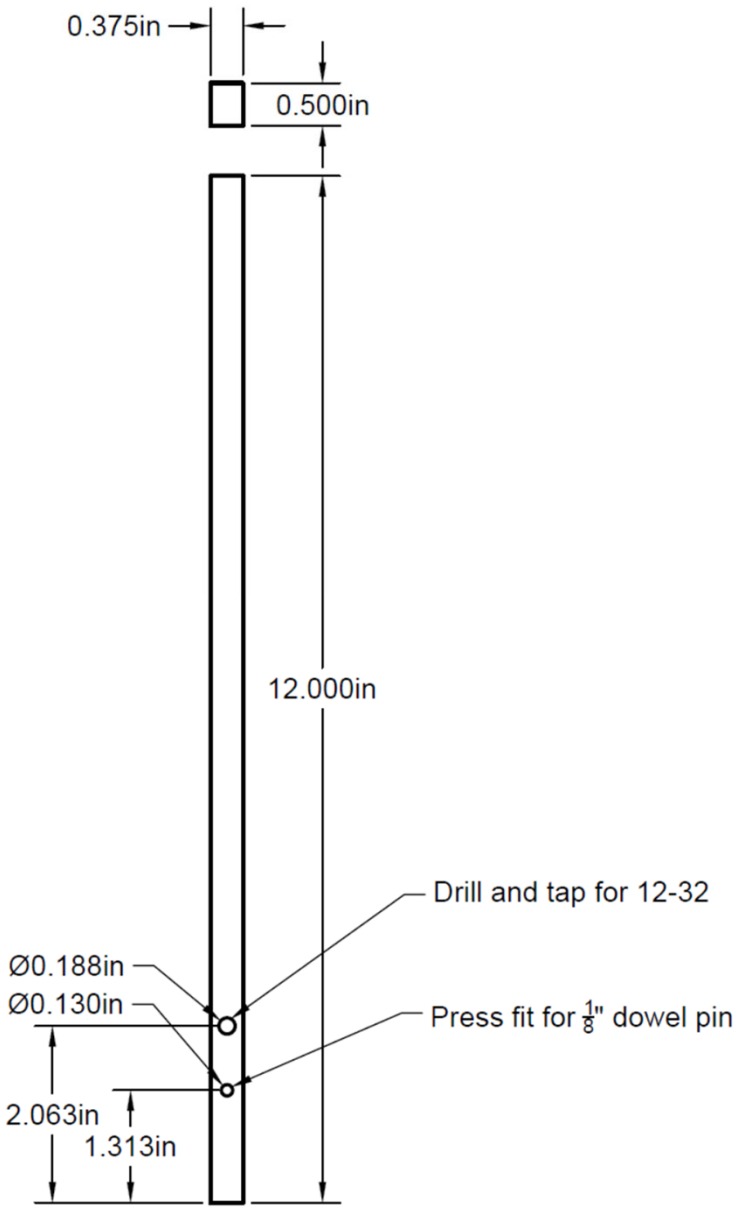
Figure A.6



Prototype ESI Source Union Mount
Material: Delrin
Scale 1:1

Figure A.6 Machine shop drawing of the prototype ESI source union mount.

Figure A.7



Prototype ESI Source Rail
Material: Aluminum
Scale 1:2

Figure A.7 Machine shop drawing of the prototype ESI source rail.

Figure A.8

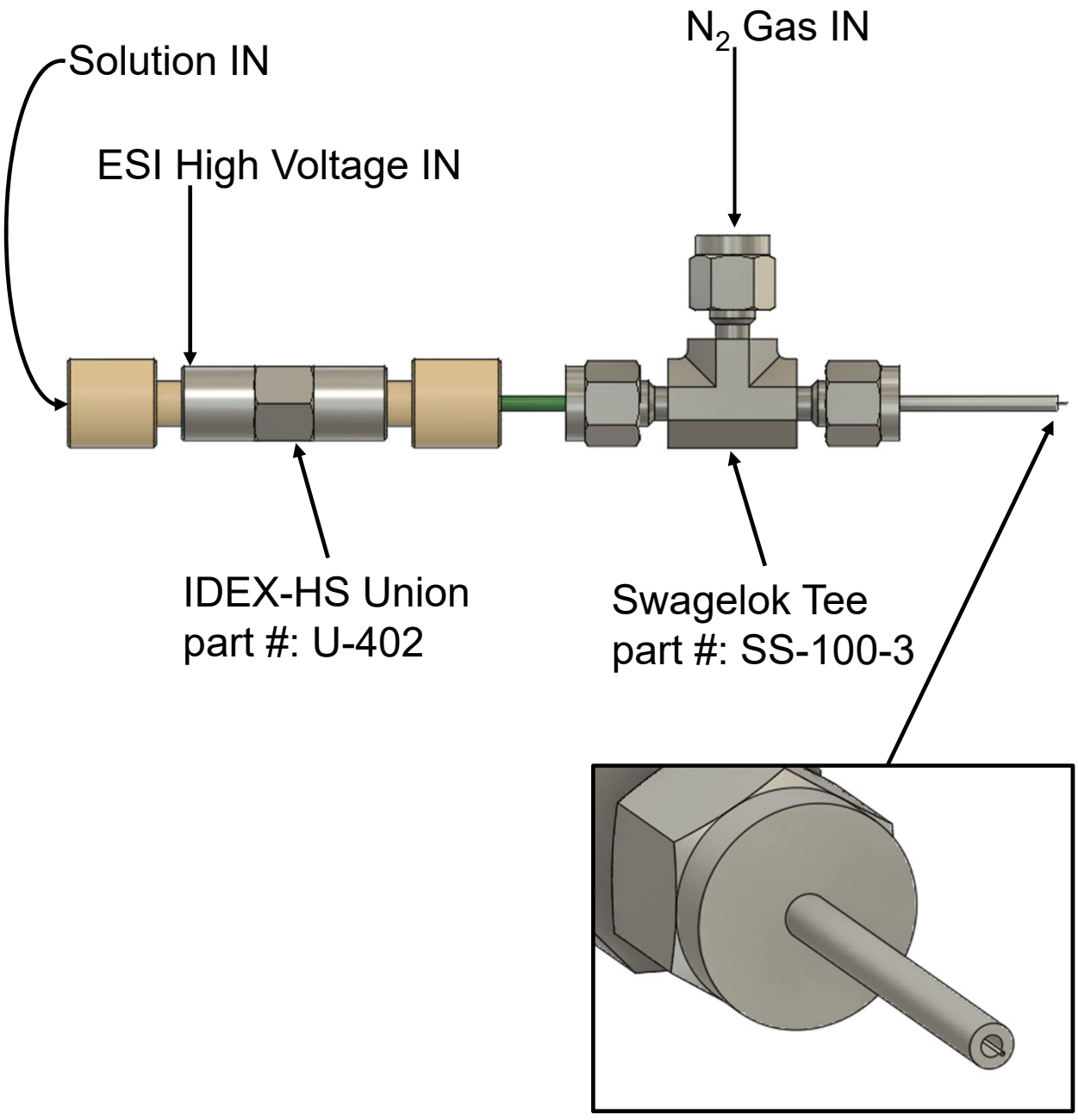


Figure A.8 Prototype pneumatic nebulization assembly. A standard HPLC stainless steel union is used to make an electrical connection between the ESI voltage and solution. A T-junction allows the ESI emitter to pass straight through and a nebulizing gas to be introduced coaxially to the solution.

Figure A.9

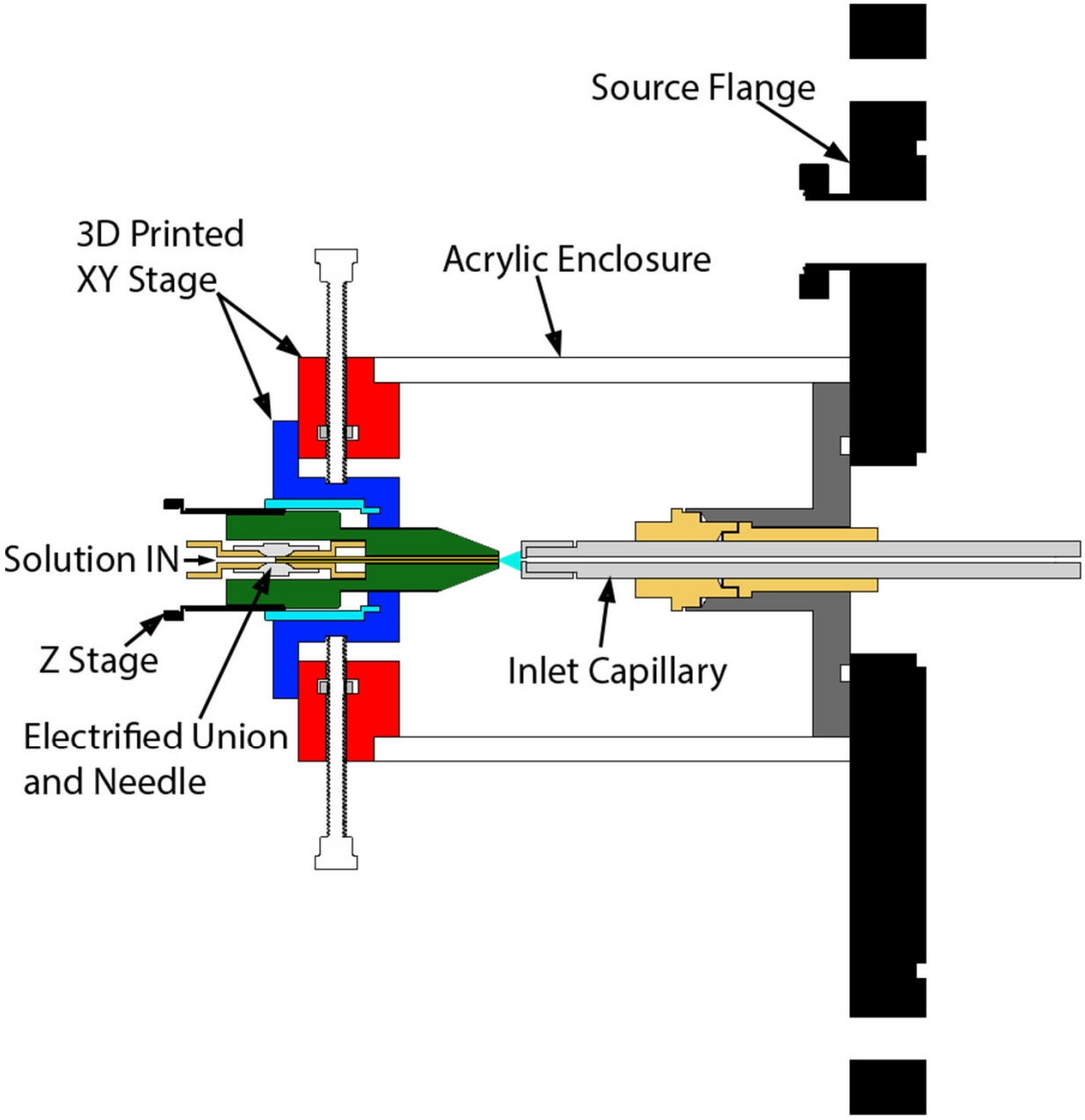


Figure A.9 Cross sectioned view of the enclosed ESI source. The components in red, dark blue, and green are 3D printed in PETG. Custom aluminum standoffs supporting the assembly from the flange are not shown.

Figure A.10

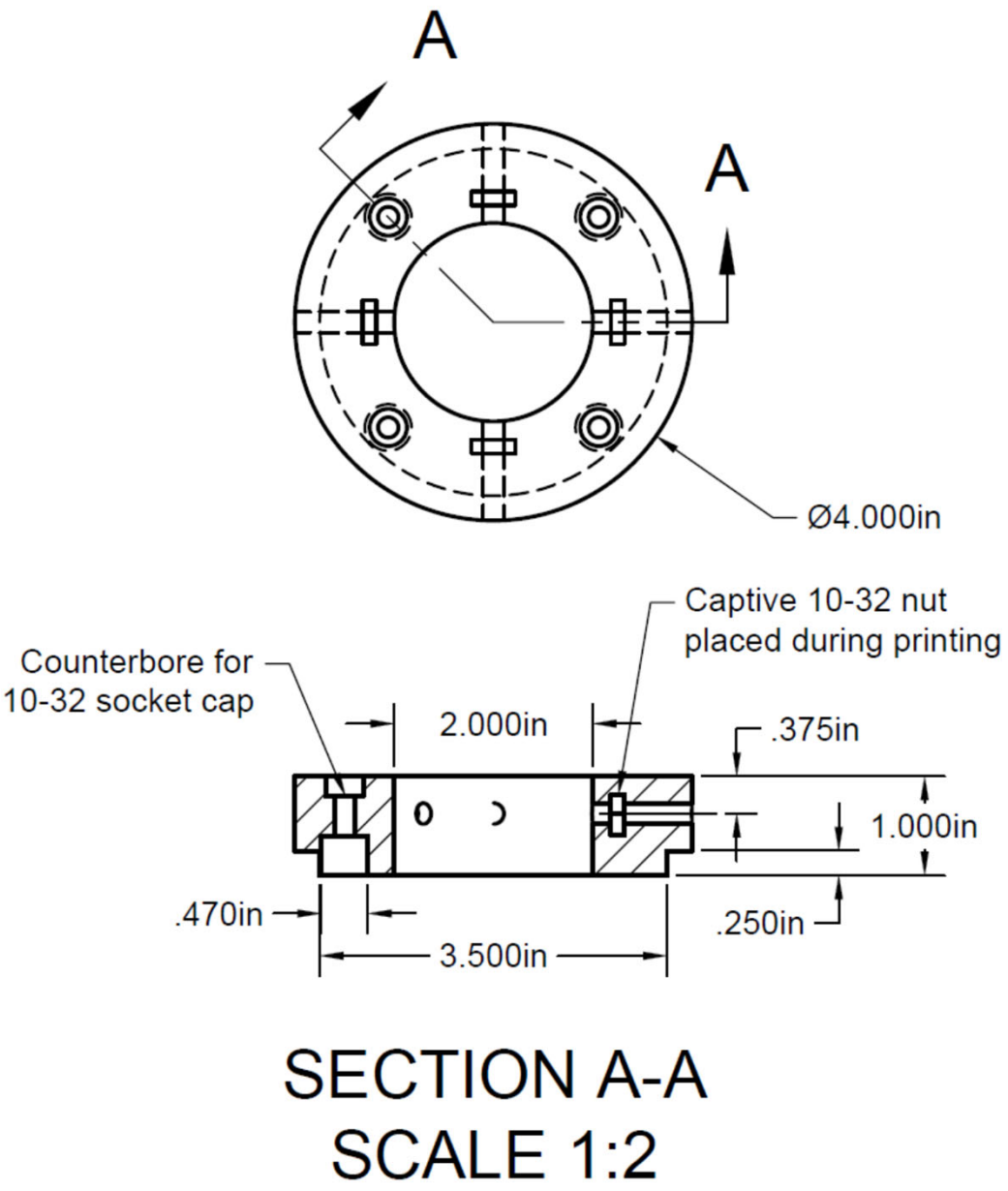
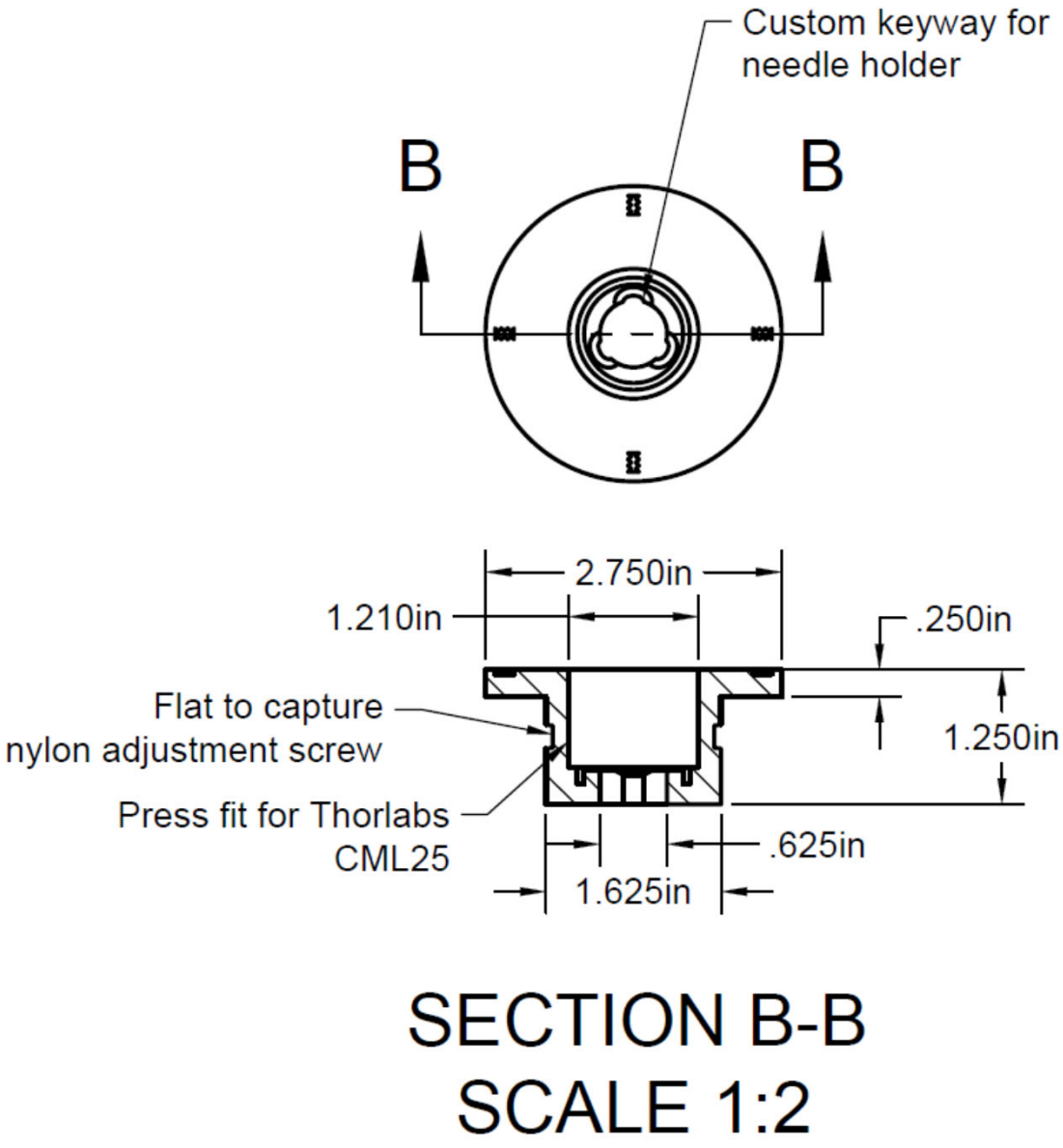


Figure A.10 Machine shop drawing of the 3D printed ESI source main body.

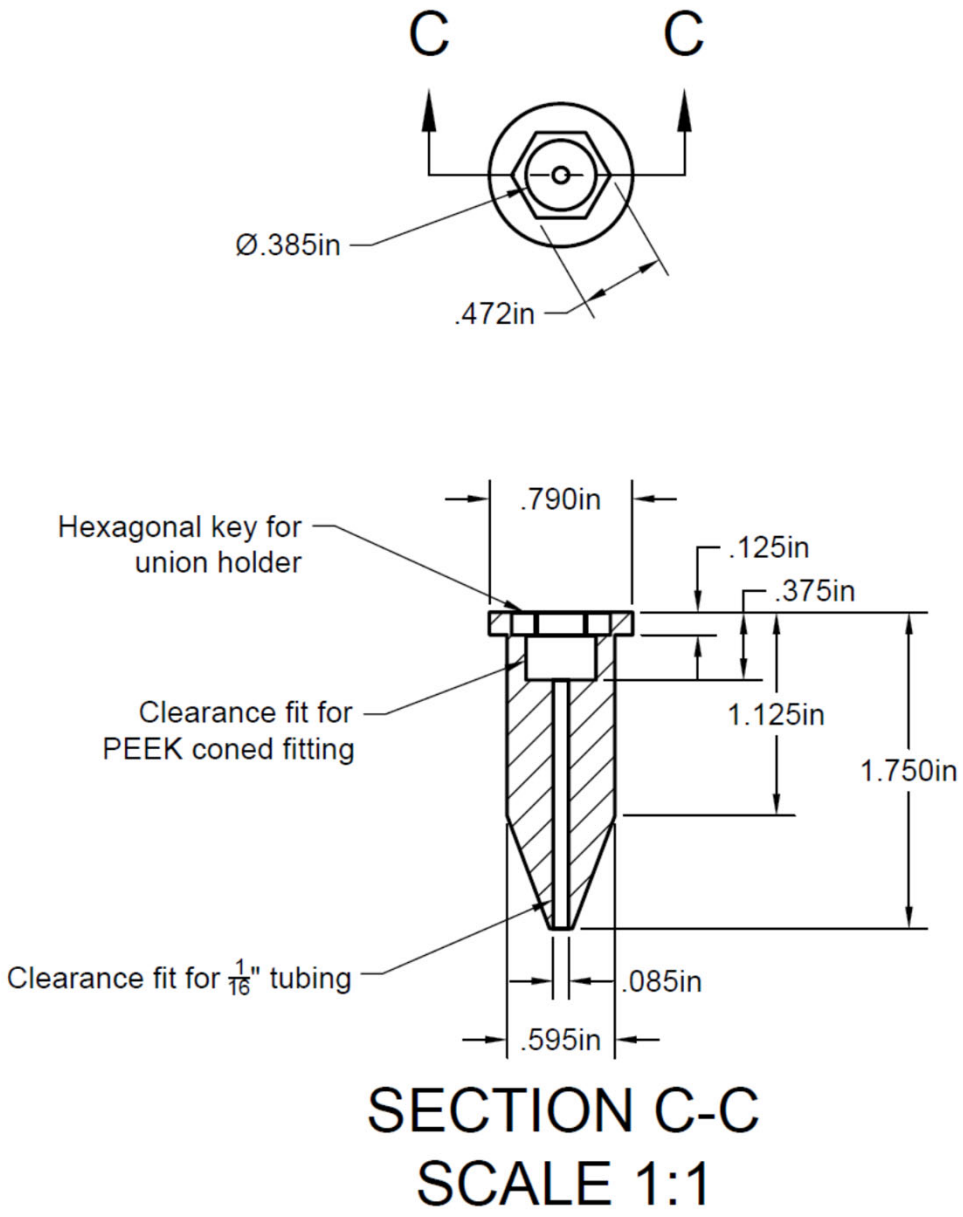
Figure A.11



Enclosed ESI Carriage
Material: 3D printed PETG
Scale 1:2

Figure A.11 Machine shop drawing of the 3D printed ESI source carriage.

Figure A.12



Enclosed ESI Needle Guide
Material: 3D printed PETG
Scale 1:1

Figure A.12 Machine shop drawing of the 3D printed ESI source needle guide.

Figure A.13

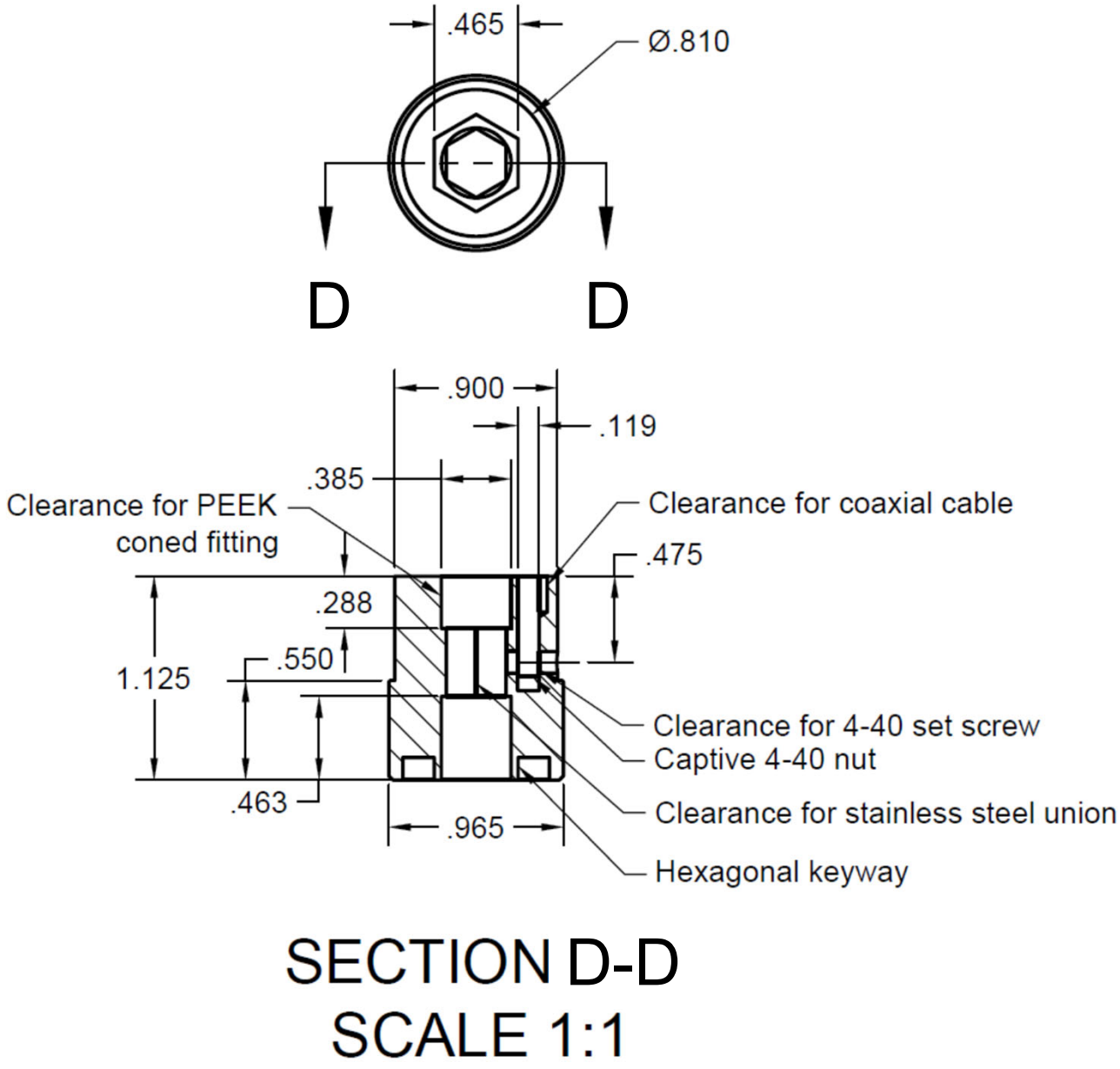


Figure A.13 Machine shop drawing of the 3D printed ESI source union holder.

Figure A.14

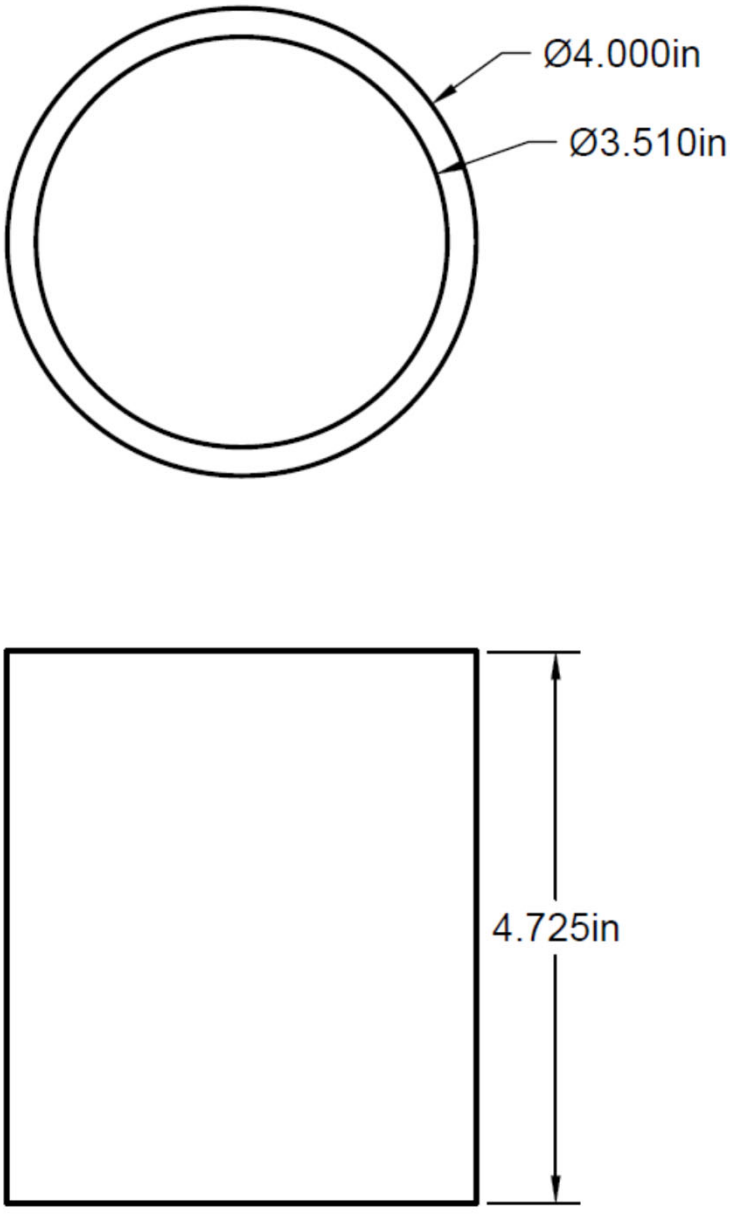
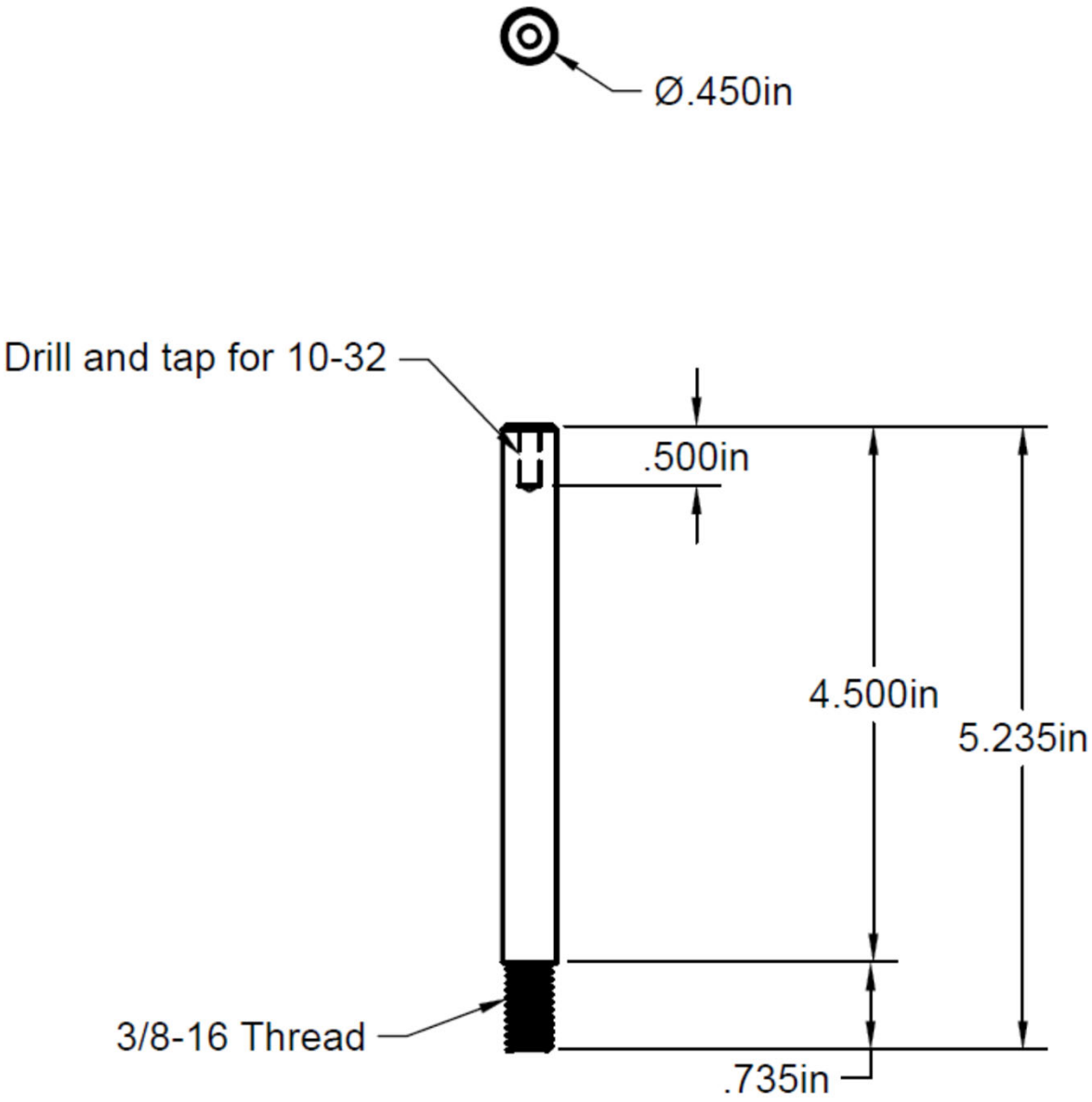


Figure A.14 Machine shop drawing of the 3D printed ESI source enclosure tube.

Figure A.15



Enclosed ESI Standoff
Material: 0.5” Aluminum Rod
Scale 1:2

Figure A.15 Machine shop drawing of the 3D printed ESI source standoffs.

Figure A.16

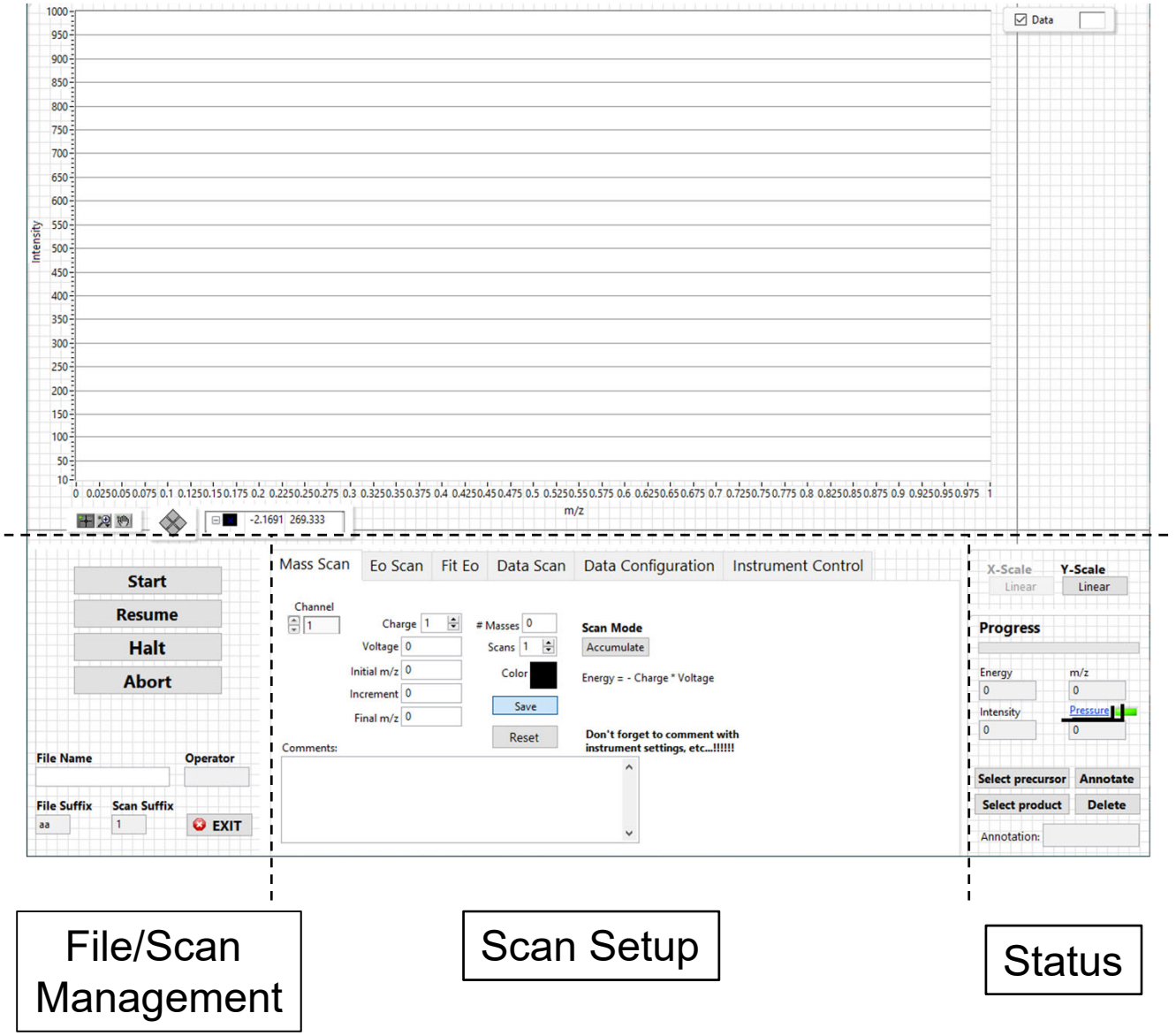


Figure A.16 Overview of the LabVIEW MS-EMP user interface. Data is plotted on an XY graph control and formatted appropriately to its scan mode. File and scan management is largely universal to the scan mode, as are the status indicators and graph scale controls. The central tab control contains all the unique parameters used by each of the individual scanning modes.

Figure A.17

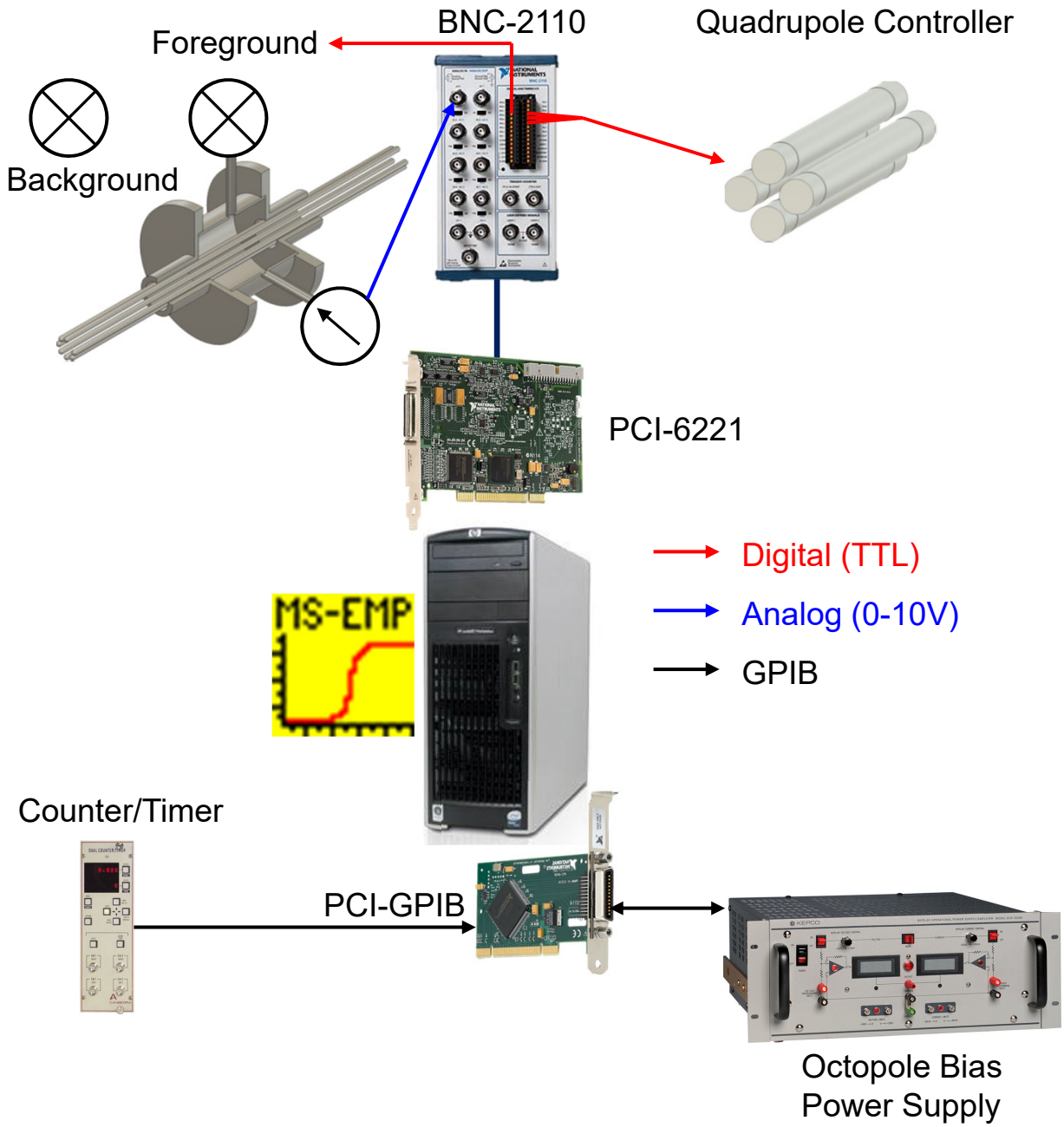


Figure A.17 Diagram of the communication interfaces that drive communication between MS-EMP and the GIBMS. Arrows indicate the direction of communication and signal type.

Figure A.18

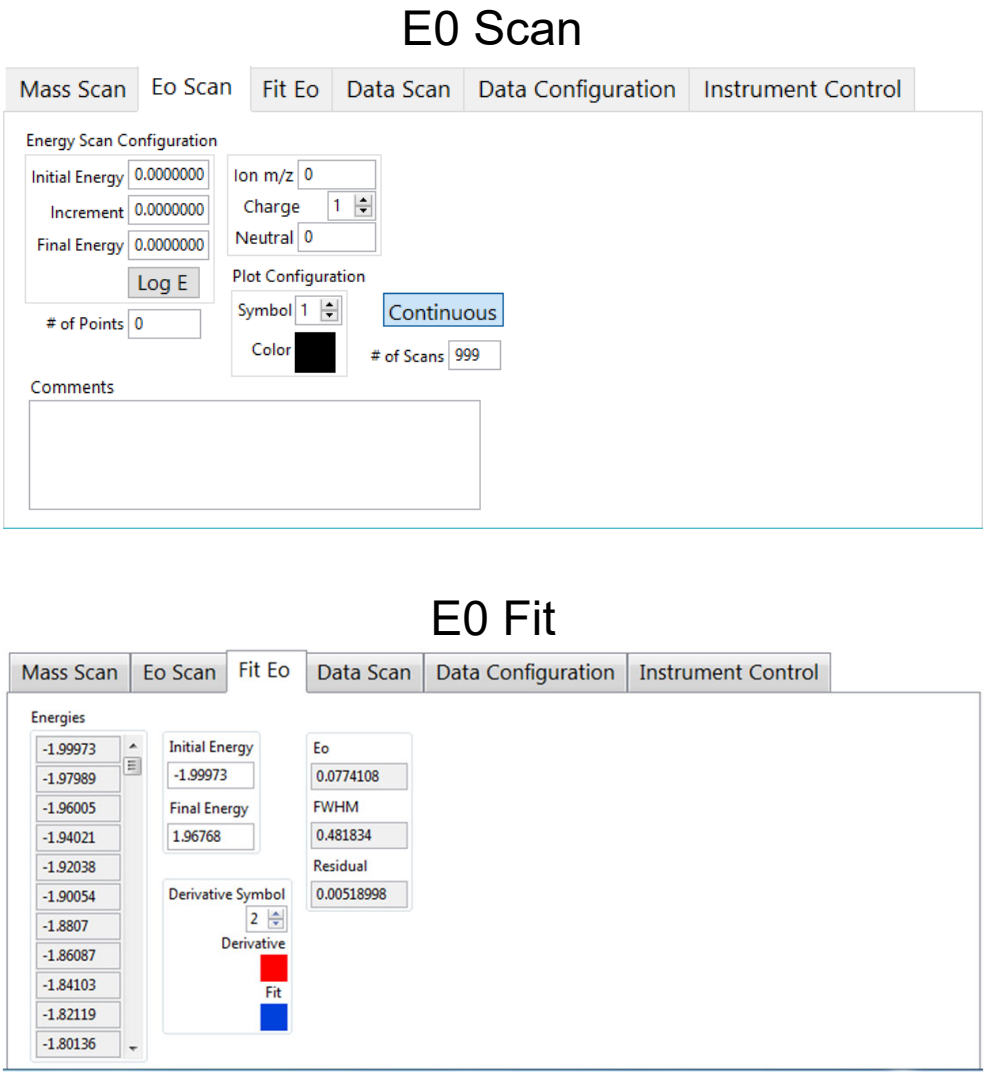


Figure A.18 The user controls for configuring an *E0 scan* and those for fitting the resulting data to determine the FWHM and E0.

Figure A.19

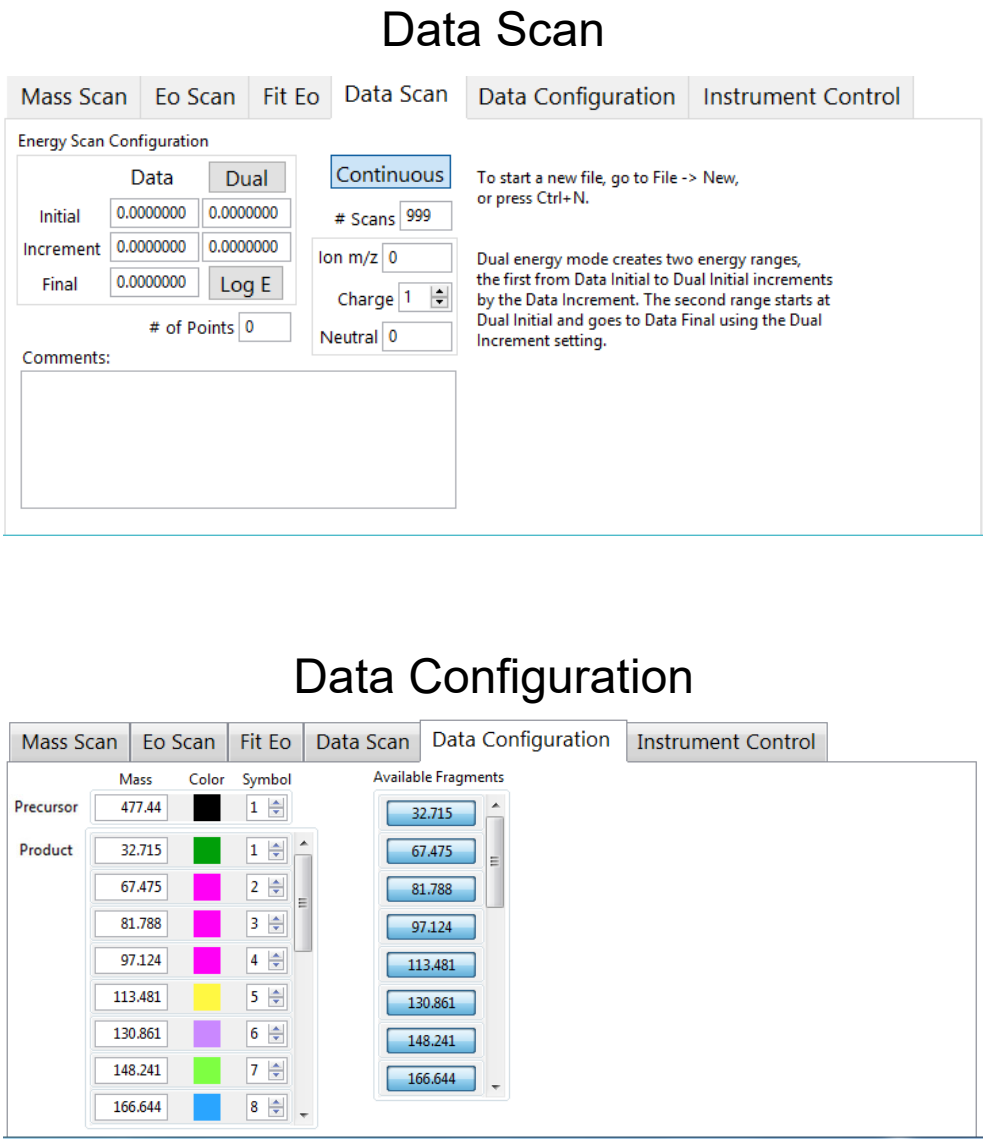


Figure A.19 The user controls for configuring the collision energy and mass scans of a *data scan*.

Appendix B Custom IRMPD Beam Path Components

Figure B.1

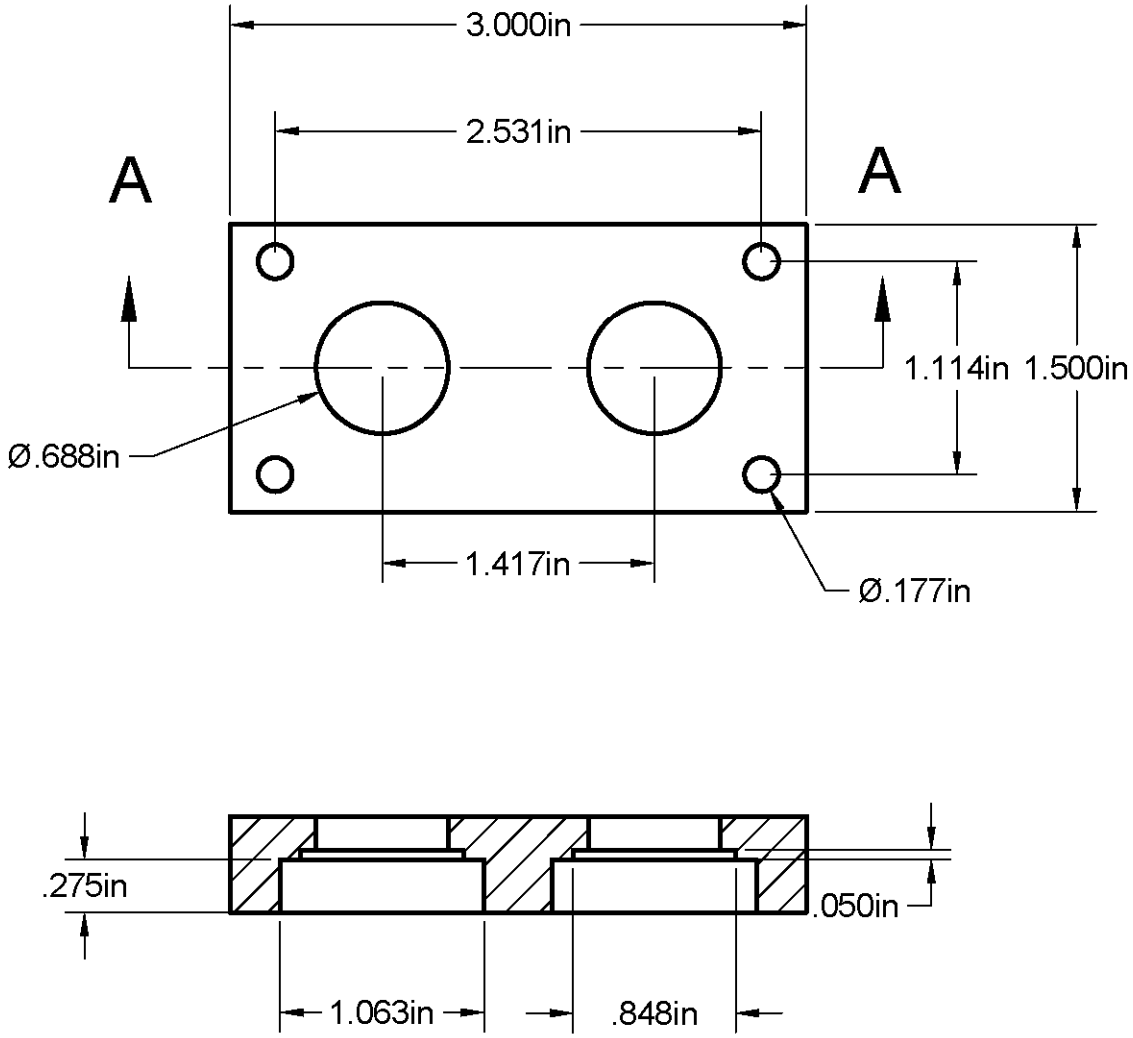
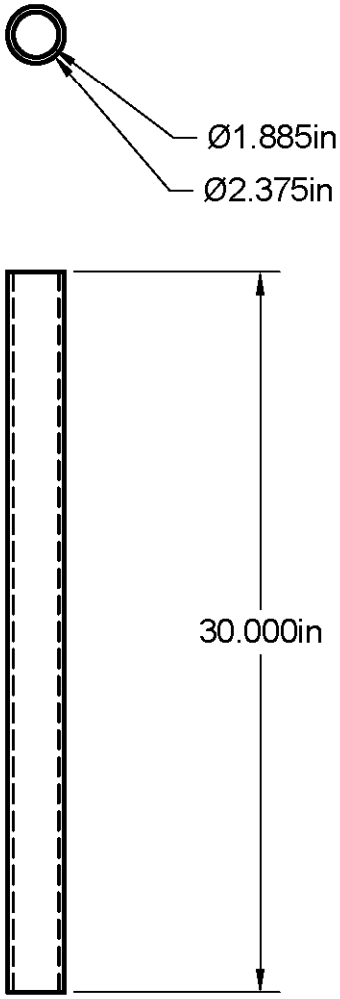


Figure B.1 Machine shop drawing of the window holder for the QIT-MS.

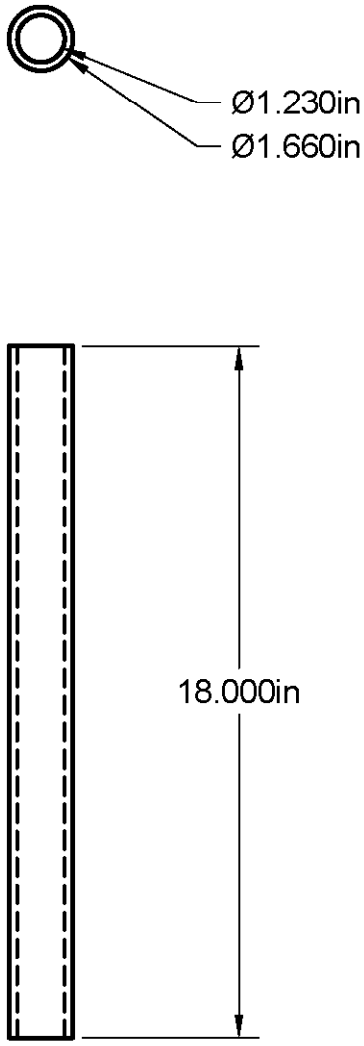
Figure B.2



Beam Enclosure Extension Large Tube
Material: Schedule 40 PVC Pipe
Scale 1:8

Figure B.2 Machine shop drawing of the large tube of the beam enclosure extension for the FT-ICR MS.

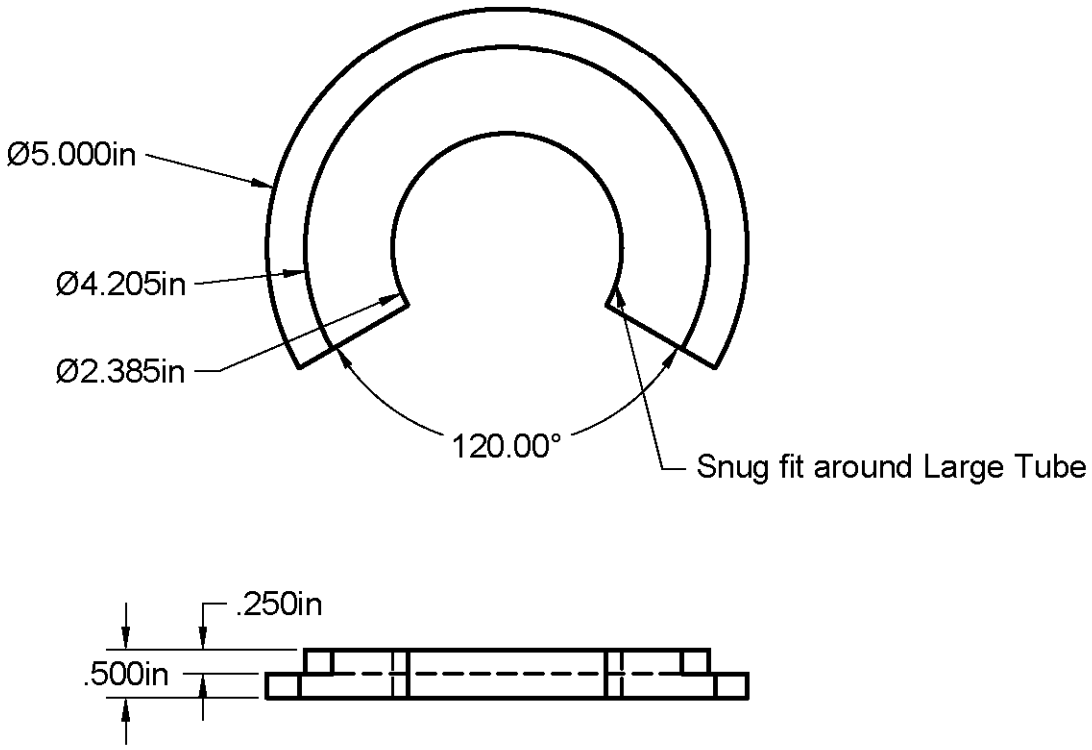
Figure B.3



Beam Enclosure Extension Small Tube
Material: Schedule 40 PVC Pipe
Scale 1:5

Figure B.3 Machine shop drawing of the small tube of the beam enclosure extension for the FT-ICR MS.

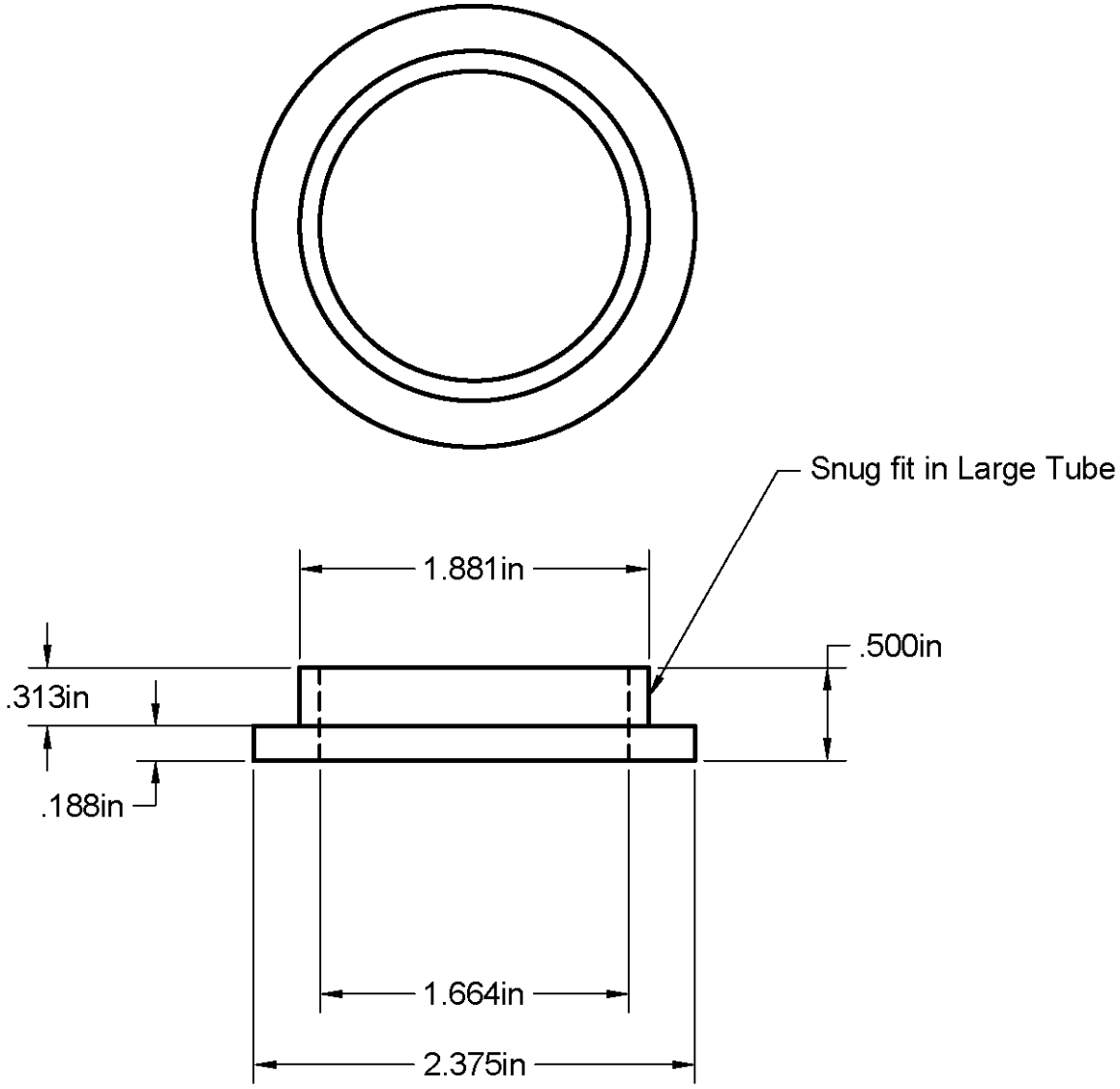
Figure B.4



FT-ICR Bore Cover
Material: Black Polypropylene Sheet
Scale 1:2

Figure B.4 Machine shop drawing of FT-ICR MS bore cover for the beam enclosure extension.

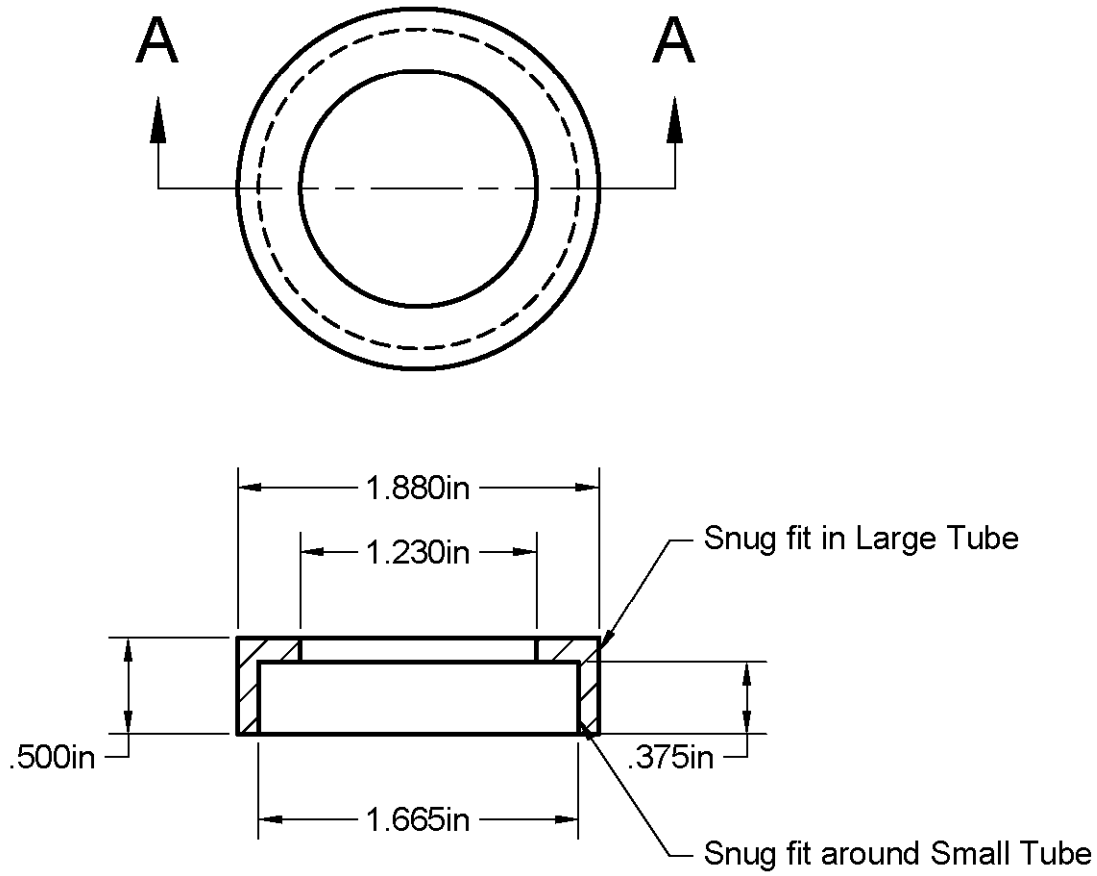
Figure B.5



Bearing A for Large Tube
Material: Black Polypropylene Sheet
Scale 1:1

Figure B.5 Machine shop drawing of the large bearing for the beam enclosure extension for the FT-ICR MS.

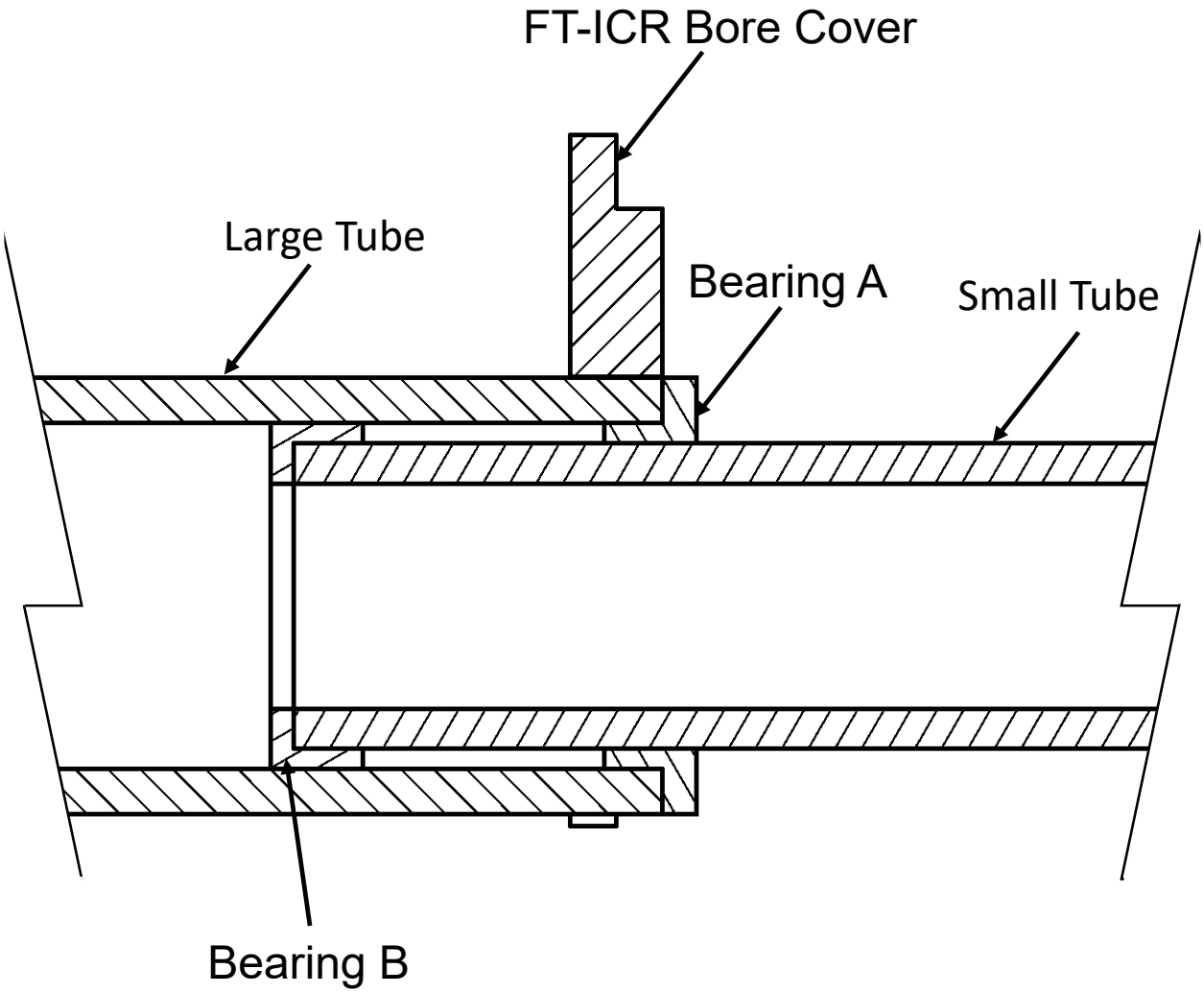
Figure B.6



Bearing B for Small Tube
Material: Black Polypropylene Sheet
Scale 1:1

Figure B.6 Machine shop drawing of the small bearing for the beam enclosure extension for the FT-ICR MS.

Figure B.7



Beam Enclosure Extension Assembly
Scale 1:1

Figure B.7 Machine shop drawing of the beam enclosure extension assembly for the FT-ICR MS.

APPENDIX C ENERGY-RESOLVED COLLISION-INDUCED DISSOCIATION OVERVIEW AND DATA ANALYSIS SOFTWARE

Energy-resolved collision-induced dissociation (ER-CID) experiments are performed in the Rodgers laboratory using the Bruker 3D quadrupole ion trap mass spectrometer described briefly in **Chapter 2**. Instead of irradiation by the IR OPO laser to produce photodissociation, they are kinetically excited by an auxiliary rf amplitude applied to the 3D quadrupole ion trap endcap electrodes to cause CID. The rf excitation amplitude (rf_{EA}) is scanned from 0.00 V until dissociation is complete, or no further dissociation is observed, by the smallest step size the instrument allows, 0.01 V. Although very similar in design to the ER-CID experiment described in **Appendix A**, the lack of stringent experimental control over ion internal energy, kinetic energy, and collision conditions prevents effective use of the TCID modeling to extract high quality energetics. Therefore, instead of calculating an absolute collisional cross section from the ion intensities, a survival yield is calculated via **Equation C.1** from the intensity of the precursor ion (I_p) and fragment ion(s) (I_f).

$$\text{Survival yield} = I_p \Big/ \left(I_p + \sum_i I_{f_i} \right) \qquad (\mathbf{C.1})$$

The survival yields are plotted as a function of rf_{EA} and a four parameter logistic dynamic curve fit, based upon **Equation C.2**, is used to determine the $CID_{50\%}$, i.e. the rf_{EA} required to produce 50% dissociation of the precursor ion.

$$\text{Survival yield} = \min + \frac{\max - \min}{1 + (rf_{EA}/CID_{50\%})^{CID_{slope}}} \qquad (\mathbf{C.2})$$

Min and max are the minimum (0) and maximum (1) values of the survival yield, rf_{EA} is the rf excitation amplitude, and CID_{slope} is the slope of the spine of the curve. Relative

stabilities for parallel systems are resolved by comparison of the determined $CID_{50\%}$ values.

The original workflow of energy-resolved collision-induced dissociation experiments (ER-CID) involved manual data extraction from individual data files collected at each rf excitation amplitude (rf_{EA}) for each of three replicate experiments. This process was particularly time consuming. Changes in the experimental method, from individual data scans, to continuous data scans leveraging the segment functionality, built into the Bruker QIT-MS data acquisition software, to control the scanning of the rf_{EA} led to even more rapid data acquisition and a greater need to handle the initial stage of data analysis automatically. The relatively straightforward nature of the ER-CID experiment led to it being an easily accessible tool in high demand, further increasing the demand for such a tool. In conjunction with previous student Yanlong Zhu, who was responsible for overseeing the ER-CID experiments at the time, the Survival Yield Analysis program was developed in LabVIEW™. The front panel of Survival Yield Analysis is shown in **Figure C.1**.

Data is extracted from tab-delimited text files, exported in ‘profile’ mode from Bruker’s Data Analysis program, both for easy verification by inexperienced users, and for easy data import. A downside of this method is the large file sizes and a somewhat prolonged import process, however a robust, easy-to-implement method for reading data directly from the native Bruker file format has not yet been found. Upon data import, all the profile spectra contained within a given data file are averaged and displayed in an XY graph. The averaged profile spectra of each individual data file are used for a default LabVIEW™ peak picking subVI to try and identify the peaks of important ions, which are

then displayed in the ‘Peaks’ listbox. The parameters of this peak picking subVI are available for editing through a drop-down menu option or the corresponding configuration file. Upon selecting an identified peak, the XY graph will zoom into that region of the averaged mass spectrum and the user is prompted to select the upper and lower bounds of the peak, used to determine individual peak widths and better determine peak intensities. Values for these bounds are automatically populated both on the graph as draggable cursors and by the corresponding array entry below the graph. It is important to note that these bounds must be set individually for each data file imported in order to force students to evaluate each data set individually. Ions included in the analysis are populated in a custom array where each element is comprised of a checkbox identifying it as either a precursor (P) or fragment (F) ion, and numeric controls for entry of the ion m/z, upper bound and lower bound. Selecting an ion from the ‘Peaks’ listbox not only automatically zooms the graph and sets default values for the upper and lower bounds but populates these values in this custom array. Likewise, dragging the upper and lower bounds cursors updates these array values, and vice versa. Once all the precursor and fragment ions have been selected and their bounds established for each data file, an output file is specified and the experimental segment length, initial rf_{EA} and rf_{EA} step are set. If necessary, a checkbox allows for the survival yield to be automatically normalized to one, which also uses a built-in LabVIEW™ subVI to detect the threshold in the sigmoid upon onset into the spine and set the survival yields for those rf_{EA} values less than the threshold to 1. Parameters for this threshold determination are accessible via a drop-down menu and configuration file.

After calculating the survival yield curve, it is displayed alongside normalized ion intensity curves to ensure the correct parameters were chosen and good quality data was collected. An output file can be saved via drop-down menu or the ‘Ctrl+S’ keyboard shortcut. The output file consists of the rf_{EA}, the survival yield calculated for each data file, an averaged survival yield and corresponding standard deviation, and the averaged intensities at each rf_{EA} for the selected precursor and fragment ions and corresponding standard deviations. This data file is formatted as a tab delimited .asc file for easy import into a separate spreadsheet or data analysis program in order to fit the survival yield with a four parameter logistic dynamic curve. The Survival Yield program requires no external libraries beyond the LabVIEW™ runtime and is therefore simple to install directly via the installer.

Figure C.1

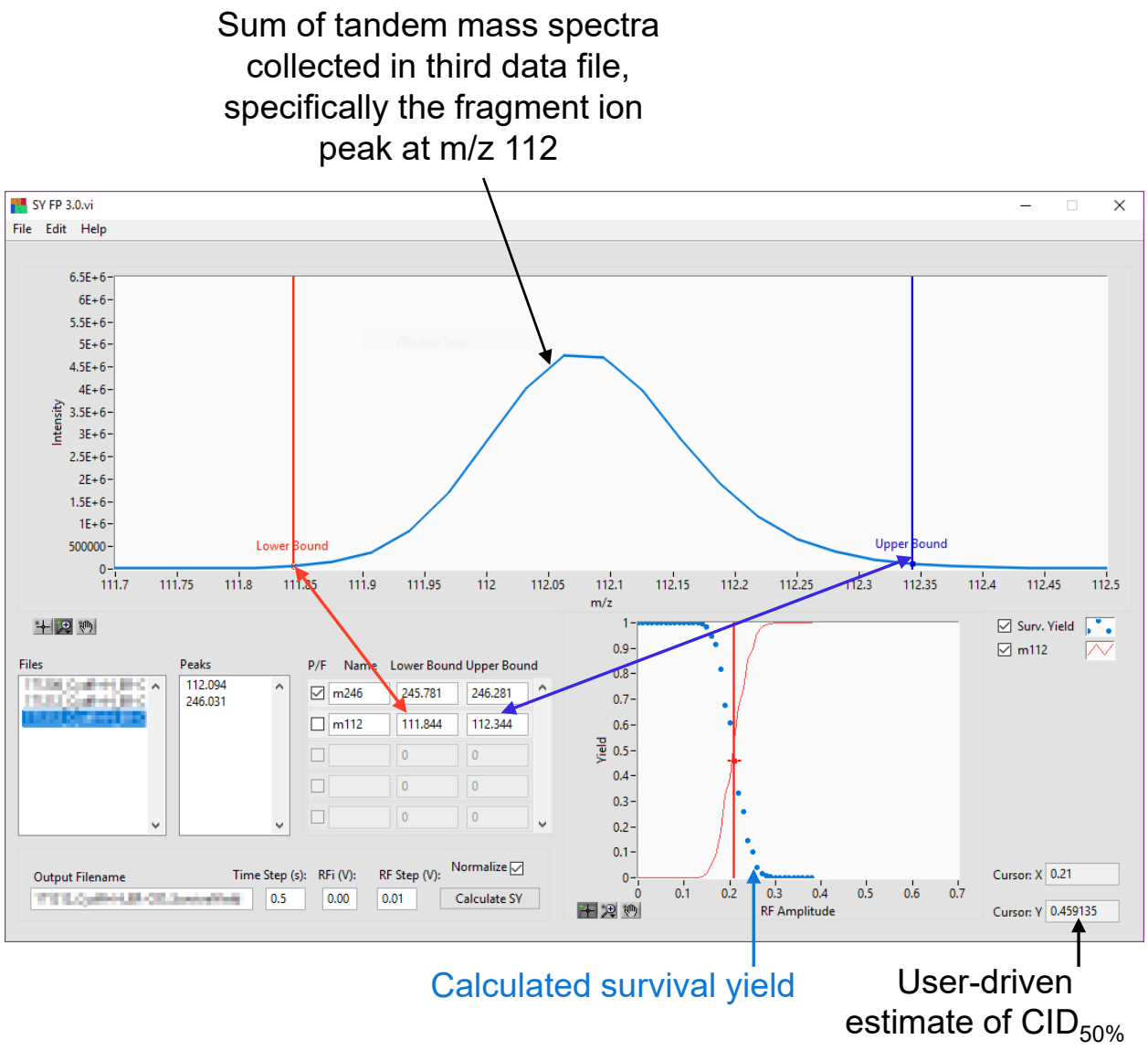


Figure C.1 Front panel interface of Survival Yield Analysis.

REFERENCES

1. Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J., *Molecular Cell Biology, 4th Edition*. W. H. Freeman: New York, 2000.

2. Tatum, E. L., Molecular Biology, Nucleic Acids, and the Future of Medicine. *Perspect. Biol. Med.* **1966,** *10*, 19-32.

3. Saenger, W., *Principles of Nucleic-Acid Structure*. Springer-Verlag: New York, 1984.

4. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** *171*, 737-738.

5. Neidle, S., *Oxford Gandbook of Nucleic Acid Structure*. Oxford University Press: Oxford ; New York, 1999; p xviii, 662 p.

6. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D* **2010,** *66*, 12-21.

7. Bewley, C. A.; Gronenborn, A. M.; Clore, G. M., Minor Groove-Binding Architectual Proteins: Structure, Function, and DNA Recognition. *Annu. Rev. Bioph. Biom.* **1998,** *27*, 105-131.

8. Chaires, J. B., Drug—DNA interactions. *Curr. Opin. Struc. Biol.* **1998,** *8*, 314-320.

9. Garvie, C. W.; Wolberger, C., Recognition of Specific DNA Sequences. *Mol. Cell* **2001,** *8*, 937-946.

10. Hamilton, P. L.; Arya, D. P., Natural Product DNA Major Groove Binders. *Nat. Prod. Rep.* **2012,** *29*, 134-143.

11. Pabo, C. O.; Sauer, R. T., Protein-DNA Recognition. *Annu. Rev. Biochem.* **1984,** *53*, 293-321.

12. Rana, T. M., Illuminating the Silence: Understanding the Structure and Function of Small RNAs. *Nat. Rev. Mol. Cell. Bio.* **2007,** *8*, 23-36.

13. Cullen, B. R., Is RNA Interference Involved in Intrinsic Antiviral Immunity in Mammals? *Nat. Immunol.* **2006,** *7*, 563-567.

14. Batey, R. T.; Rambo, R. P.; Doudna, J. A., Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Edit.* **1999,** *38*, 2326-2343.

15. Bevilacqua, P. C.; Ritchey, L. E.; Su, Z.; Assmann, S. M., Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* **2016,** *50*, 235-266.

16. Sigler, P. B., An Analysis of the Structure of tRNA. *Annu. Rev. Biophy. Bioeng.* **1975,** *4*, 477-527.

17. Decatur, W. A.; Fournier, M. J., RNA-Guided Nucleotide Modification of Ribosomal and Other RNAs. *J. Biol. Chem.* **2003,** *278*, 695-8.

18. Helm, M., Post-Transcriptional Nucleotide Modification and Alternative Folding of RNA. *Nucleic Acids Res.* **2006,** *34*, 721-33.

19. Vare, V. Y.; Eruysal, E. R.; Narendran, A.; Sarachan, K. L.; Agris, P. F., Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. *Biomolecules* **2017,** *7*, 29.

20. De Clercq, E., Dancing with chemical formulae of antivirals: A personal account. *Biochemical Pharmacology* **2013,** *86*, 711-725.

21. De Clercq, E.; Li, G. D., Approved Antiviral Drugs over the Past 50 Years. *Clin. Microbiol. Rev.* **2016,** *29*, 695-747.

22. Sommadossi, J.-P., Nucleoside Analogs: Similarities and Differences. *Clinical Infectious Diseases* **1993,** *16*, S7-S15.

23. Mathews, C. K., DNA Synthesis as a Therapeutic Target: The First 65 Years. *Faseb J.* **2012,** *26*, 2231-2237.

24. Mackey, J. R.; Baldwin, S. A.; Young, J. D.; Cass, C. E., Nucleoside transport and its significance for anticancer drug resistance. *Drug Resistance Updates* **1998,** *1*, 310-324.

25. Kaufman, H. E.; Dohlman, C.; Martola, E. L., Use of 5-Iodo-2'-Deoxyuridine (IDU) in Treatment of Herpes Simplex Keratitis. *Arch. Ophthalmol.-Chic.* **1962,** *68*, 235-239.

26. Cihlar, T.; Ray, A. S., Nucleoside and Nucleotide HIV Reverse Transcriptase Inhibitors: 25 years after Zidovudine. *Antivir. Res.* **2010,** *85*, 39-58.

27. Chowers, M.; Gottesman, B. S.; Leibovici, L.; Schapiro, J. M.; Paul, M., Nucleoside Reverse Transcriptase Inhibitors in Combination Therapy for HIV Patients: Systematic Review and Meta-Analysis. *Eur. J. Clin. Mirociol.* **2010,** *29*, 779-86.

28. Papatheodoridis, G. V.; Dimou, E.; Papadimitropoulos, V., Nucleoside Analogues for Chronic Hepatitis B: Antiviral Efficacy and Viral Resistance. *Am. J. Gastroenterol.* **2002,** *97*, 1618-28.

29. Galmarini, C. M.; Mackey, J. R.; Dumontet, C., Nucleoside Analogues: Mechanisms of Srug Resistance and Reversal Strategies. *Leukemia* **2001,** *15*, 875-90.

30. Dalton, J. G.; George, A. L.; Hruska, F. E.; McGaig, T. N.; Ogilvie, K. K.; Peeling, J.; Wood, D. J., Comparison of Arabinose and Ribose Nucleoside Conformation in Aqueous and Dimethylsulfoxide Solution 2',5'-Hydrogen Bonding in Arabinonucleosides. *Biochim. BioPhys. Acta.* **1977,** *478*, 261-73.

31. Hruska, F. E.; Wood, D. J.; McCaig, T. N.; Smith, A. A.; Holy, A., A Nuclear Magnetic Resonance Study of Nucleoside Conformation in Solution. The Effect of Structure and Conformation on the Magnetic Nonequivalence of the 5´-Methylene Hydrogens. *Can. J. Chem.* **1974,** *52*, 497-508.

32. Tollin, P.; Wilson, H. R.; Young, D. W., The Molecular Conformation of Uracil-β-D-Arabinofuranoside. *Nature-New Biol.* **1973,** *242*, 49-50.

33. Chwang, A. K.; Sundaralingam, M.; Hanessian, S., Crystal and Molecular Structure of a Purine Arabinonucleoside, 9-β-D-Arabinofuranosyladenine Hydrochloride. *Acta Crystallogr. B* **1974,** *30*, 2273-2277.

34. Saran, A.; Pullman, B.; Perahia, D., Molecular Orbital Calculations on the Conformation of Nucleic Acids and Their Constituents. X. Conformation of Beta-D-Arabinosyl Nucleosides. *Biochim. Biophys. Acta.* **1974,** *349*, 189-203.

35. Everaert, D. H.; Peeters, O. M.; Deranter, C. J.; Blaton, N. M.; Vanaerschot, A.; Herdewijn, P., Conformational-Analysis of Substituent Effects on the Sugar Puckering Mode and the Anti-Hiv Activity of 2',3'-Dideoxypyrimidine Nucleosides. *Antivir. Chem. Chemoth.* **1993,** *4*, 289-299.

36. Nasr, M.; Litterst, C.; McGowan, J., Computer-Assisted Structure-Activity Correlations of Dideoxynucleoside Analogs as Potential Anti-Hiv Drugs. *Antivir. Res.* **1990,** *14*, 125-148.

37. Nasr, M.; Litterst, C.; McGowan, J., Structure-Activity Correlations of Pyrimidine and Purine Dideoxynucleosides as Potential Anti-Hiv Drugs. *Ann. Ny. Acad. Sci.* **1990,** *616*, 579-587.

38. Plavec, J.; Koole, L. H.; Chattopadhyaya, J., Structural-Analysis of 2',3'-Dideoxyinosine, 2',3'-Dideoxyadenosine, 2',3'-Dideoxyguanosine and 2',3'-Dideoxycytidine by 500-Mhz H-1-Nmr Spectroscopy and Abinitio Molecular-Orbital Calculations. *J. Biochem. Bioph. Meth.* **1992,** *25*, 253-272.

39. Taylor, E. W.; Van Roey, P.; Schinazi, R. F.; Chu, C. K., A Stereochemical Rationale for the Activity of Anti-HIV Nucleosides. *Antivir. Chem. Chemoth.* **1990,** *1*, 163-173.

40. Van Roey, P.; Salerno, J. M.; Chu, C. K.; Schinazi, R. F., Correlation between Preferred Sugar Ring Conformation and Activity of Nucleoside Analogs against Human Immunodeficiency Virus. *Proc. Natl. Acad. Sci. U.S.A.* **1989,** *86*, 3929-3933.

41. Sundaralingam, M., Structure and Conformation of Nucleosides and Nucleotides and Their Analogs as Determined by X-Ray Diffraction. *Ann. Ny. Acad. Sci.* **1975,** *255*, 3-42.

42. Altona, C.; Sundaralingam, M., Conformational-Analysis of Sugar Ring in Nucleosides and Nucleotides - New Description Using Concept of Pseudorotation. *J. Am. Chem. Soc.* **1972,** *94*, 8205-8212.

43. Sundaralingam, M., Stereochemistry of Nucleic Acids and their Constituents. IV. Allowed and Preferred Conformations of Nucleosides, Nucleoside Mono-, Di-, Tri-, Tetraphosphates, Nucleic Acids and Polynucleotides. *Biopolymers* **1969,** *7*, 821-860.

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

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

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

47. Wu, R.; Rodgers, M., Tautomerization Lowers the Activation Barriers for N-Glycosidic Bond Cleavage of Protonated Uridine and 2′-Deoxyuridine. *Phys. Chem. Chem. Phys.* **2016,** *18*, 24451−24459.

48. Wu, R. R.; He, C. C.; Hamlow, L. A.; Nei, Y.-w.; Berden, G.; Oomens, J.; Rodgers, M. T., N3 Protonation Induces Base Rotation of 2'-Deoxyadenosine-5'-monophosphate and Adenosine-5'-monophosphate. *J. Phys. Chem. B* **2016,** *120*, 4616-4624.

49. Filippi, A.; Fraschetti, C.; Rondino, F.; Piccirillo, S.; Steinmetz, V.; Guidoni, L.; Speranza, M., Protonated Pyrimidine Nucleosides Probed by IRMPD Spectroscopy. *Int. J. Mass Spectrom.* **2013,** *354*, 54-61.

50. Ung, H. U.; Huynh, K. T.; Poutsma, J. C.; Oomens, J.; Berden, G.; Morton, T. H., Investigation of proton affinities and gas phase vibrational spectra of protonated nucleosides, deoxynucleosides, and their analogs. *Int. J. Mass Spectrom.* **2015,** *378*, 294–302.

51. Wu, R. R.; Yang, B.; Berden, G.; Oomens, J.; Rodgers, M. T., Gas-Phase Conformations and Energetics of Protonated 2'-Deoxyguanosine and Guanosine: IRMPD Action Spectroscopy and Theoretical Studies. *J. Phys. Chem. B* **2014,** *118*, 14774-14784.

52. Wu, R. R.; Yang, B.; Berden, G.; Oomens, J.; Rodgers, M. T., Gas-Phase Conformations and Energetics of Protonated 2'-Deoxyadenosine and Adenosine: IRMPD Action Spectroscopy and Theoretical Studies. *J. Phys. Chem. B* **2015,** *119*, 2795-2805.

53. Wu, R. R.; Yang, B.; Frieler, C. E.; Berden, G.; Oomens, J.; Rodgers, M. T., Diverse Mixtures of 2,4-Dihydroxy Tautomers and O4 Protonated Conformers of Uridine and 2'-Deoxyuridine Coexist in the Gas Phase. *Phys. Chem. Chem. Phys.* **2015,** *17*, 25978-25988.

54. Wu, R. R.; Yang, B.; Frieler, C. E.; Berden, G.; Oomens, J.; Rodgers, M. T., N3 and O2 Protonated Tautomeric Conformations of 2'-Deoxycytidine and Cytidine Coexist in the Gas Phase. *J. Phys. Chem. B* **2015,** *119*, 5773-5784.

55. Wu, R. R.; Yang, B.; Frieler, C. E.; Berden, G.; Oomens, J.; Rodgers, M. T., 2,4-Dihydroxy and O2 Protonated Tautomers of dThd and Thd Coexist in the Gas Phase: Methylation Alters Protonation Preferences versus dUrd and Urd. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 410-421.

56. Parneix, P.; Basire, M.; Calvo, F., Accurate modeling of infrared multiple photon dissociation spectra: The dynamical role of anharmonicities. *J. Phys. Chem. A* **2013,** *117*, 3954-3959.

57. Nei, Y.-w.; Hallowita, N.; Steill, J. D.; Oomens, J.; Rodgers, M. T., Infrared Multiple Photon Dissociation Action Spectroscopy of Deprotonated DNA

Mononucleotides: Gas-Phase Conformations and Energetics. *J. Phys. Chem. A* **2013,** *117*, 1319-1335.

58. Nei, Y.-w.; Akinyemi, T.; Kaczan, C.; Steill, J.; Berden, G.; Oomens, J.; Rodgers, M., Infrared Multiple Photon Dissociation Action Spectroscopy of Sodiated Uracil and Thiouracils: Effects of Thioketo-Substitution on Gas-Phase Conformation. *Int. J. Mass Spectrom.* **2011,** *308*, 191-202.

59. Nei, Y.-w.; Akinyemi, T. E.; Steill, J. D.; Oomens, J.; Rodgers, M. T., Infrared Multiple Photon Dissociation Action Spectroscopy of Protonated Uracil and Thiouracils: Effects of Thioketo-Substitution on Gas-Phase Conformation. *Int. J. Mass Spectrom.* **2010,** *297*, 139-151.

60. Wu, R.; He, C.; Hamlow, L.; Nei, Y.-w.; Berden, G.; Oomens, J.; Rodgers, M., Protonation Induces Base Rotation of Purine Nucleotides pdGuo and pGuo. *Phys. Chem. Chem. Phys.* **2016,** *18*, 15081–15090.

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

62. Yang, B.; Wu, R. R.; Berden, G.; Oomens, J.; Rodgers, M. T., Infrared Multiple Photon Dissociation Action Spectroscopy of Proton-Bound Dimers of Cytosine and Modified Cytosines: Effects of Modifications on Gas-Phase Conformations. *J. Phys. Chem. B* **2013,** *117*, 14191–14201.

63. Zhu, Y.; Hamlow, L. A.; He, C. C.; Lee, J. K.; Gao, J.; Berden, G.; Oomens, J.; Rodgers, M. T., Gas-Phase Conformations and N-Glycosidic Bond Stabilities of Sodium

Cationized 2'-Deoxyguanosine and Guanosine: Sodium Cations Preferentially Bind to the Guanine Residue. *J. Phys. Chem. B* **2017,** *121*, 4048-4060.

64. Zhu, Y.; Hamlow, L. A.; He, C. C.; Roy, H. A.; Cunningham, N. A.; Munshi, M. U.; Berden, G.; Oomens, J.; Rodgers, M. T., Conformations and N-Glycosidic Bond Stabilities of Sodium Cationized 2′-Deoxycytidine and Cytidine: Solution Conformation of $[Cyd+Na]^+$ is Preserved upon ESI. *Int. J. Mass Spectrom.* **2018,** *429*, 18-27.

65. He, C. C.; Hamlow, L. A.; Devereaux, Z. J.; Zhu, Y.; Nei, Y. W.; Fan, L.; McNary, C. P.; Maitre, P.; Steinmetz, V.; Schindler, B.; Compagnon, I.; Armentrout, P. B.; Rodgers, M. T., Structural and Energetic Effects of O2'-Ribose Methylation of Protonated Purine Nucleosides. *J Phys Chem B* **2018,** *122*, 9147-9160.

66. 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., Influence of Transition Metal Cationization versus Sodium Cationization and Protonation on the Gas-Phase Tautomeric Conformations and Stability of Uracil: Application to $[Ura+Cu]^+$ and $[Ura+Ag]^+$. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 2438-2453.

67. Crampton, K. T.; Rathur, A. I.; Nei, Y. W.; Berden, G.; Oomens, J.; Rodgers, M. T., Protonation Preferentially Stabilizes Minor Tautomers of the Halouracils: IRMPD Action Spectroscopy and Theoretical Studies. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1469-1478.

68. Gao, J.; Berden, G.; Rodgers, M. T.; Oomens, J., Interaction of Cu^+ with cytosine and formation of i-motif-like C-M^+-C complexes: alkali versus coinage metals. *Phys. Chem. Chem. Phys.* **2016,** *18*, 7269-77.

69. Sadeghian, K.; Flaig, D.; Blank, I. D.; Schneider, S.; Strasser, R.; Stathis, D.; Winnacker, M.; Carell, T.; Ochsenfeld, C., Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study. *Angew. Chem. Int. Edit.* **2014,** *53*, 10044-10048.

70. Wang, C.; Gao, H.; Gaffney, B. L.; Jones, R. A., Nitrogen-15-Labeled Oligodeoxynucleotides. 3. Protonation of the Adenine N1 in the A·C and A·G mispairs of the duplexes {d[CG(15N1)AGAATTCCCG]}$_2$ and {d[CGGGAATTC(15N1)ACG]}$_2$. *J. Am. Chem. Soc.* **1991,** *113*, 5486-5488.

71. Leroy, J. L.; Gehring, K.; Kettani, A.; Gueron, M., Acid Multimers of Oligodeoxycytidine Strands: Stoichiometry, Base-Pair Characterization, and Proton Exchange Properties. *Biochemistry* **1993,** *32*, 6019-31.

72. Gehring, K.; Leroy, J.-L.; Guéron, M., A Tetrameric DNA Structure with Protonated Cytosine-Cytosine Base Pairs. *Nature* **1993,** *363*, 561-565.

73. Russo, N.; Toscano, M.; Grand, A.; Jolibois, F., Protonation of Thymine, Cytosine, Adenine, and Guanine DNA Nucleic Acid Bases: Theoretical Investigation into the Framework of Density Functional Theory. *J. Comput. Chem.* **1998,** *19*, 989-1000.

74. Tureček, F.; Chen, X., Protonated adenine: Tautomers, solvated clusters, and dissociation mechanisms. *J. Am. Soc. Mass Spectrom.* **2005,** *16*, 1713-1726.

75. Touboul, D.; Bouchoux, G.; Zenobi, R., Gas-phase protonation thermochemistry of adenosine. *J. Phys. Chem. B* **2008,** *112*, 11716-11725.

76. Salpin, J.-Y.; Scuderi, D., Structure of Protonated Thymidine Characterized by Infrared Multiple Photon Dissociation and Quantum Calculations. *Rapid. Commun. Mass Sp.* **2015,** *29*, 1898-1904.

77. Thomas, G.; Favre, A., 4-Thiouridine Triggers Both Growth Delay Induced by near-Ultraviolet Light and Photoprotection. *Eur. J. Biochem.* **1980,** *113*, 67-74.

78. Zhang, R. B.; Eriksson, L. A., Theoretical Study on Conformational Preferences of Ribose in 2-Thiouridine - The Role of the 2' OH Group. *Phys. Chem. Chem. Phys.* **2010,** *12*, 3690-3697.

79. Agris, P. F., Wobble Position Modified Nucleosides Evolved to Select Transfer RNA Codon Recognition: a Modified-Wobble Hypothesis. *Biochimie* **1991,** *73*, 1345-9.

80. Agris, P. F., In *Progress in Nucleic Acid Research and Molecular Biology*, Cohn, W. E.; Moldave, K., Eds. Academic Press: 1996; Vol. 53.pp 79-129

81. Helm, M., Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* **2006,** *34*, 721-733.

82. Crick, F. H. C., Codon-Anticodon Pairing - Wobble Hypothesis. *J. Mol. Biol.* **1966,** *19*, 548-555.

83. Grosjean, H. J.; Dehenau, S.; Crothers, D. M., Physical Basis for Ambiguity in Genetic Coding Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1978,** *75*, 610-614.

84. McCloskey, J. A.; Nishimura, S., Modified Nucleosides in Transfer-RNA. *Accounts Chem. Res.* **1977,** *10*, 403-410.

85. Fourrey, J. L.; Gasche, J.; Fontaine, C.; Guittet, E.; Favre, A., Sequence Dependent Photochemistry of Di(Deoxynucleoside) Phosphates Containing 4-Thiouracil. *J. Chem. Soc. Chem. Comm.* **1989**, 1334-1336.

86. Mueller, E. G.; Buck, C. J.; Palenchar, P. M.; Barnhart, L. E.; Paulson, J. L., Identification of a Gene Involved in the Generation of 4-Thiouridine in tRNA. *Nucleic Acids Res.* **1998,** *26*, 2606-2610.

87. Khvorostov, A.; Lapinski, L.; Rostkowska, H.; Nowak, M. J., Unimolecular Photochemistry of 4-Thiouracils. *Photochem. Photobiol.* **2005,** *81*, 1205-1211.

88. Rabani, M.; Levin, J. Z.; Fan, L.; Adiconis, X.; Raychowdhury, R.; Garber, M.; Gnirke, A.; Nusbaum, C.; Hacohen, N.; Friedman, N.; Amit, I.; Regev, A., Metabolic Labeling of RNA Uncovers Principles of RNA Production and Degradation Dynamics in Mammalian Cells. *Nat. Biotechnol.* **2011,** *29*, 436-42.

89. Spitzer, J.; Hafner, M.; Landthaler, M.; Ascano, M.; Farazi, T.; Wardle, G.; Nusbaum, J.; Khorshid, M.; Burger, L.; Zavolan, M.; Tuschl, T., PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods Enzymol.* **2014,** *539*, 113-61.

90. Bergmann, W.; Burke, D. C., Contributions to the Study of Marine Products .39. The Nucleosides of Sponges .3. Spongothymidine and Spongouridine. *J. Org. Chem.* **1955,** *20*, 1501-1507.

91. Bergmann, W.; Feeney, R. J., Contributions to the Study of Marine Products .32. The Nucleosides of Sponges .1. *J. Org. Chem.* **1951,** *16*, 981-987.

92. Cullis, P. A.; Cushing, R., Vidarabine Encephalopathy. *J. Neurol. Neurosur. Ps.* **1984,** *47*, 1351-1354.

93. Abraham, A.; Devasia, A. J.; Varatharajan, S.; Karathedath, S.; Balasubramanian, P.; Mathews, V., Effect of Cytosine Arabinoside Metabolizing Enzyme Expression on Drug Toxicity in Acute Myeloid Leukemia. *Ann. Hematol.* **2015,** *94*, 883-885.

94. Absalon, M. J.; Smith, F. O., Treatment Strategies for Pediatric Acute Myeloid Leukemia. *Expert. Opin. Pharmaco.* **2009,** *10*, 57-79.

95. Saran, A.; Pullman, B.; Perahia, D., Molecular orbital calculations on the conformation of nucleic acids and their constituents. X. Conformation of beta-D-arabinosyl nucleosides. *Biochim Biophys Acta* **1974,** *349*, 189-203.

96. Chwang, A. K.; Sundaralingam, M.; Hanessian, S., Crystal and Molecular-Structure of a Purine Arabinonucleoside, 9-Beta-D-Arabinofuranosyladenine Hydrochloride. *Acta Crystallogr. B* **1974,** *30*, 2273-2277.

97. Tollin, P.; Wilson, H. R.; Young, D. W., The molecular conformation of uracil- -D-arabinofuranoside. *Nature-New Biol.* **1973,** *242*, 49-50.

98. Hruska, F. E.; Wood, D. J.; Mccaig, T. N.; Smith, A. A.; Holy, A., Nuclear Magnetic-Resonance Study of Nucleoside Conformation in Solution - Effect of Structure and Conformation on Magnetic Nonequivalence of 5'-Methylene Hydrogens. *Can. J. Chem.* **1974,** *52*, 497-508.

99. Perry, C. M.; Noble, S., Didanosine - An Updated Review of its use in HIV Infection. *Drugs* **1999,** *58*, 1099-1135.

100. Veal, G. J.; Barry, M. G.; Back, D. J., Zalcitabine (ddC) Phosphorylation and Drug-Interactions. *Antivir. Chem. Chemoth.* **1995,** *6*, 379-384.

101. Sanger, F.; Nicklen, S.; Coulson, A. R., DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **1977,** *74*, 5463-5467.

102. Ju, J.; Edwards, J. R.; Itagaki, Y., DNA Sequencing Using Biotinylated Dideoxynucleotides and Mass Spectrometry. *Nucleic Acids Res.* **2001,** *29*, e104-e104.

103. Prober, J.; Trainor, G.; Dam, R.; Hobbs, F.; Robertson, C.; Zagursky, R.; Cocuzza, A.; Jensen, M.; Baumeister, K., A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides. *Science* **1987,** *238*, 336-341.

104. Plavec, J.; Koole, L. H.; Chattopadhyaya, J., Structural-Analysis of 2',3'-Dideoxyinosine, 2',3'-Dideoxyadenosine, 2',3'-Dideoxyguanosine and 2',3'-Dideoxycytidine by 500-MHz H-1-NMR Spectroscopy and Ab-initio Molecular-Orbital Calculations. *J. Biochem. Bioph. Meth.* **1992,** *25*, 253-272.

105. Oomens, J.; Sartakov, B. G.; Meijer, G.; Von Helden, G., Gas-Phase Infrared Multiple Photon Dissociation Spectroscopy of Mass-Selected Molecular ions. *Int. J. Mass Spectrom.* **2006,** *254*, 1-19.

106. Ervin, K. M.; Armentrout, P. B., Translational Energy Dependence of $Ar^+ + XY \rightarrow ArX^+ + Y$ ($XY=H_2,D_2,HD$) from Thermal to 30 eV c.m. *J. Chem. Phys.* **1985,** *83*, 166-189.

107. Lapthorn, C.; Pullen, F.; Chowdhry, B. Z., Ion Mobility Spectrometry-Mass Spectrometry (IMS-MS) of Small Molecules: Separating and Assigning Structures to Ions. *Mass Spectrom. Rev.* **2013,** *32*, 43-71.

108. Rizzo, T. R.; Stearns, J. A.; Boyarkin, O. V., Spectroscopic Studies of Cold, Gas-Phase Biomolecular Ions. *Int. Rev. Phys. Chem.* **2009,** *28*, 481-515.

109. Agris, P. F.; Sierzputowskagracz, H.; Smith, W.; Malkiewicz, A.; Sochacka, E.; Nawrot, B., Thiolation of Uridine Carbon-2 Restricts the Motional Dynamics of the Transfer-RNA Wobble Position Nucleoside. *J. Am. Chem. Soc.* **1992,** *114*, 2652-2656.

110. Ly, T.; Julian, R. R., Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Edit.* **2009,** *48*, 7130-7137.

111. Watson, C. H.; Baykut, G.; Eyler, J. R., Laser Photodissociation of Gaseous Ions Formed by Laser Desorption. *Anal. Chem.* **1987,** *59*, 1133-1138.

112. Fridgen, T. D., Infrared Consequence Spectroscopy of Gaseous Protonated and Metal Ion Cationized Complexes. *Mass Spectrom. Rev.* **2009,** *28*, 586-607.

113. Bush, M. F.; O'Brien, J. T.; Prell, J. S.; Saykally, R. J.; Williams, E. R., Infrared Spectroscopy of Cationized Arginine in the Gas Phase: Direct Evidence for the Transition from Nonzwitterionic to Zwitterionic Structure. *J. Am. Chem. Soc.* **2007,** *129*, 1612-1622.

114. MacAleese, L.; Maitre, P., Infrared Spectroscopy of Organometallic Ions in the Gas Phase: From Model to Real World Complexes. *Mass Spectrom. Rev.* **2007,** *26*, 583-605.

115. Stedwell, C. N.; Galindo, J. F.; Roitberg, A. E.; Polfer, N. C., Structures of Biomolecular ions in the Gas Phase Probed by Infrared Light Sources. *Annu. Rev. Anal. Chem.* **2013,** *6*, 267-285.

116. Miyamoto, Y.; Majima, T.; Arai, S.; Katsumata, K.; Akagi, H.; Maeda, A.; Hata, H.; Kuramochi, K.; Kato, Y.; Tsukiyama, K., ^{13}C-Selective Infrared Multiple Photon Dissociation of Beta-Propiolactone by a Free Electron Laser. *Nucl. Instrum. Meth. B* **2011,** *269*, 180-184.

117. Valle, J. J.; Eyler, J. R.; Oomens, J.; Moore, D. T.; van der Meer, A. F. G.; von Helden, G.; Meijer, G.; Hendrickson, C. L.; Marshall, A. G.; Blakney, G. T., Free electron laser-Fourier transform ion cyclotron resonance mass spectrometry facility for obtaining infrared multiphoton dissociation spectra of gaseous ions. *Rev. Sci. Instrum.* **2005,** *76*, 023103-7.

118. Lemaire, J.; Boissel, P.; Heninger, M.; Mauclaire, G.; Bellec, G.; Mestdagh, H.; Simon, A.; Caer, S. L.; Ortega, J. M.; Glotin, F.; Maitre, P., Gas Phase Infrared Spectroscopy of Selectively Prepared Ions. *Phys. Rev. Lett.* **2002,** *89*.

119. Polfer, N. C.; Oomens, J.; Suhai, S.; Paizs, B., Infrared Spectroscopy and Theoretical Studies on Gas-Phase Protonated Leu-enkephalin and its Fragments: Direct Experimental Evidence for the Mobile Proton. *J. Am. Chem. Soc.* **2007,** *129*, 5887-5897.

120. Hoffmann, W.; Marianski, M.; Warnke, S.; Seo, J.; Baldauf, C.; von Helden, G.; Pagel, K., Assessing the Stability of Alanine-Based Helices by Conformer-Selective IR Spectroscopy. *Phys. Chem. Chem. Phys.* **2016,** *18*, 19950-4.

121. Boyall, D.; Reid, K. L., Modern Studies of Intramolecular Vibrational Energy Redistribution. *Chem. Soc. Rev.* **1997,** *26*, 223-232.

122. Nesbitt, D. J.; Field, R. W., Vibrational Energy Flow in Highly Excited Molecules: Role of Intramolecular Vibrational Redistribution. *The Journal of Physical Chemistry* **1996,** *100*, 12735-12756.

123. von Helden, G.; Holleman, I.; Meijer, G.; Sartakov, B., Excitation of C_{60} Using a Chirped Free Electron Laser. *Opt. Express* **1999,** *4*, 46-52.

124. Asmis, K. R.; Brummer, M.; Kaposta, C.; Santambrogio, G.; von Helden, G.; Meijer, G.; Rademann, K.; Woste, L., Mass-Selected Infrared Photodissociation Spectroscopy of $V_4O_{10}^+$. *Phys. Chem. Chem. Phys.* **2002,** *4*, 1101-1104.

125. Bakker, J. M.; Besson, T.; Lemaire, J.; Scuderi, D.; Maitre, P., Gas-Phase Structure of a pi-Allyl-Palladium Complex: Efficient Infrared Spectroscopy in a 7 T Fourier Transform Mass Spectrometer. *J. Phys. Chem. A* **2007,** *111*, 13415-13424.

126. Eyler, J. R., Infrared Multiple Photon Dissociation Spectroscopy of Ions in Penning Traps. *Mass Spectrom. Rev.* **2009,** *28*, 448-467.

127. Mac Aleese, L.; Simon, A.; McMahon, T. B.; Ortega, J. M.; Scuderi, D.; Lemaire, J.; Maitre, P., Mid-IR Spectroscopy of Protonated Leucine Methyl Ester Performed with an FTICR or a Paul Type Ion-Trap. *Int. J. Mass Spectrom.* **2006,** *249*, 14-20.

128. Martens, J.; Berden, G.; Gebhardt, C. R.; Oomens, J., Infrared Ion Spectroscopy in a Modified Quadrupole Ion Trap Mass Spectrometer at the FELIX Free Electron Laser Laboratory. *Rev. Sci. Instrum.* **2016,** *87*, 103108-8.

129. Martens, J.; Grzetic, J.; Berden, G.; Oomens, J., Structural Identification of Electron Transfer Dissociation Products in Mass Spectrometry Using Infrared Ion Spectroscopy. *Nat. Commun.* **2016,** *7*, 11754.

130. Papadopoulos, G.; Svendsen, A.; Boyarkin, O. V.; Rizzo, T. R., Spectroscopy of Mobility-Selected Biomolecular Ions. *Faraday Discuss.* **2011,** *150*, 243-255.

131. Polfer, N. C., Infrared Multiple Photon Dissociation Spectroscopy of Trapped Ions. *Chem. Soc. Rev.* **2011,** *40*, 2211-21.

132. Mino, W. K.; Gulyuz, K.; Wang, D.; Stedwell, C. N.; Polfer, N. C., Gas-Phase Structure and Dissociation Chemistry of Protonated Tryptophan Elucidated by Infrared Multiple-Photon Dissociation Spectroscopy. *J. Phys. Chem. Lett.* **2011,** *2*, 299-304.

133. He, C. C. Characterization of Protonated Modified Nucleosides and Amino Acid-Linked Cisplatin Analogues Using Tandem Mass Spectrometry Complemented By Theoretical Approaches. Wayne State University, 2018.

134. Wolinski, K.; Hinton, J. F.; Wishart, D. S.; Sykes, B. D.; Richards, F. M.; Pastone, A.; Saudek, V.; Ellis, P. D.; Maciel, G. E.; McIver Jr., J. W.; Blizzard, A. C.; Santry, D.

P.; Pople, J. A.; Ostlund, N. S.; Ducasse, L.; Hoarau, J.; Pesquer, M.; Kondo, M.; Ando, I.; Chujo, R.; Nishioka, A.; Vauthier, E. C.; Odiot, S.; Tonnard, F.; Baker, J. D.; Zerner, M. C.; Beveridge, D. V.; Anderson, W. P.; Cundari, T. R.; Bingham, R. C.; Dewar, M. J. S.; Lo, D. H.; Li, J.; Mello, P. C.; Jug, K.; Thiel, W.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P.; Fraser, M.; Hayes, D. M. HyperChem Computational Chemistry Software Package, version 8.0, 5.0; Hypercube Inc.: Gainesville, FL, 2004.

135. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179-5197.

136. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006,** *25*, 247-60.

137. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-713.

138. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004,** *25*, 1157-74.

139. Wu, R.; Yang, B.; Frieler, C.; Berden, G.; Oomens, J.; Rodgers, M., N3 and O2 protonated tautomeric conformations of 2′-deoxycytidine and cytidine coexist in the gas phase. *J. Phys. Chem. B* **2015,** *119*, 5773–5784.

140. Wu, R. R.; Yang, B.; Berden, G.; Oomens, J.; Rodgers, M. T., Gas-phase conformations and energetics of protonated 2′-deoxyguanosine and guanosine: IRMPD action spectroscopy and theoretical studies. *J. Phys. Chem. B* **2014,** *118*, 14774-14784.

141. Wu, R. R.; Yang, B.; Berden, G.; Oomens, J.; Rodgers, M. T., Gas-phase conformations and energetics of protonated 2′-deoxyadenosine and adenosine: IRMPD action spectroscopy and theoretical studies. *J. Phys. Chem. B* **2015,** *119*, 2795-2805.

142. Wu, R. R.; Yang, B.; Frieler, C. E.; Berden, G.; Oomens, J.; Rodgers, M. T., Diverse mixtures of 2,4-dihydroxy tautomers and O4 protonated conformers of uridine and 2′-deoxyuridine coexist in the gas phase. *Phys. Chem. Chem. Phys.* **2015,** *17*, 25978-25988.

143. Zhu, Y.; Hamlow, L. A.; He, C. C.; Roy, H. A.; Cunningham, N. A.; Munshi, M.; Berden, G.; Oomens, J.; Rodgers, M. T., Conformations and N-glycosidic bond stabilities of sodium cationized 2-deoxycytidine and cytidine: Solution conformation of $[Cyd+Na]^+$ is preserved upon ESI. *Int. J. Mass Spectrom.* **2018,** *429*, 18-27.

144. Nei, Y.-w.; Crampton, K. T.; Berden, G.; Oomens, J.; Rodgers, M. T., Infrared Multiple Photon Dissociation Action Spectroscopy of Deprotonated RNA Mononucleotides: Gas-Phase Conformations and Energetics. *J. Phys. Chem. A* **2013,** *117*, 10634-10649.

145. Cohen, A. J.; Mori-Sánchez, P.; Yang, W., Challenges for Density Functional Theory. *Chem. Rev.* **2012,** *112*, 289-320.

146. Oomens, J.; Tielens, A. G. G. M.; Sartakov, B. G.; von Helden, G.; Meijer, G., Laboratory Infrared Spectroscopy of Cationic Polycyclic Aromatic Hydrocarbon Molecules. *Astrophys. J.* **2003,** *591*, 968-985.

147. Oomens, J.; van Roij, A. J. A.; Meijer, G.; von Helden, G., Gas-Phase Infrared Photodissociation Spectroscopy of Cationic Polyaromatic Hydrocarbons. *Astrophys. J.* **2000,** *542*, 404-410.

148. Pople, J. A.; Scott, A. P.; Wong, M. W.; Radom, L., Scaling Factors for Obtaining Fundamental Vibrational Frequencies and Zero-Point Energies from HF/6–31G* and MP2/6–31G* Harmonic Frequencies. *Isr. J. Chem.* **1993,** *33*, 345-350.

149. Wong, M. W., Vibrational Frequency Prediction Using Density Functional Theory. *Chem. Phys. Lett.* **1996,** *256*, 391-399.

150. Asynchronously Calling VIs. http://zone.ni.com/reference/en-XX/help/371361P-01/lvconcepts/asynchronous_vi_calls/ (accessed 4/25/2019).

151. The Open Babel Package, version 2.3.1. https://openbabel.org (accessed 5/1/2019).

152. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011,** *3*, 33.

153. Zhu, Y.; Hamlow, L. A.; He, C. C.; Strobehn, S. F.; Lee, J. K.; Gao, J.; Berden, G.; Oomens, J.; Rodgers, M. T., Influence of Sodium Cationization versus Protonation on the Gas-Phase Conformations and Glycosidic Bond Stabilities of 2′-Deoxyadenosine and Adenosine. *J. Phys. Chem. B* **2016,** *120*, 8892–8904.

154. Zhu, Y.; Roy, H. A.; Cunningham, N. A.; Strobehn, S. F.; Gao, J.; Munshi, M. U.; Berden, G.; Oomens, J.; Rodgers, M. T., Effects of Sodium Cationization *versus* Protonation on the Conformations and *N*-Glycosidic Bond Stabilities of Sodium

Cationized Uridine and 2′-Deoxyuridine: Solution Conformation of $[Urd+Na]^+$ is Preserved Upon ESI. *Phys. Chem. Chem. Phys.* **2017,** *19*, 17637–17652.

155. Zhu, Y.; Roy, H. A.; Cunningham, N. A.; Strobehn, S. F.; Gao, J.; Munshi, M. U.; Berden, G.; Oomens, J.; Rodgers, M. T., IRMPD Action Spectroscopy, ER-CID Experiments, and Theoretical Studies of Sodium Cationized Thymidine and 5-Methyluridine: Kinetic Trapping During the ESI Desolvation Process Preserves the Solution Structure of $[Thd+Na]^+$. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 2423-2437.

156. Hamlow, L. A.; Zhu, Y.; Devereaux, Z. J.; Cunningham, N. A.; Berden, G.; Oomens, J.; Rodgers, M. T., Modified Quadrupole Ion Trap Mass Spectrometer for Infrared Ion Spectroscopy: Application to Protonated Thiated Uridines. *J. Am. Soc. Mass Spectrom.* **2018,** *29*, 2125-2137.

157. van Rossum, G.; Drake, F. L. Python Reference Manual. https://www.python.org/.

158. Walt, S. v. d.; Colbert, S. C.; Varoquaux, G., The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011,** *13*, 22-30.

159. Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open Source Scientific Tools for Python. https:///www.scipy.org/ (accessed 4/24/2019).

160. Kabsch, W., A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallogr. A* **1976,** *32*, 922-923.

161. Wagner, A.; Himmel, H.-J., aRMSD: A Comprehensive Tool for Structural Analysis. *J. Chem. Inf. Model.* **2017,** *57*, 428-438.

162. Popov, A. V.; Vorobjev, Y. N.; Zharkov, D. O., MDTRA: A Molecular Dynamics Trajectory Analyzer with a Graphical User Interface. *J. Comput. Chem.* **2013,** *34*, 319-325.

163. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996,** *14*, 33-38.

164. Calculate Root-Mean-Square Deviation (RMSD) of Two Molecules Using Rotation. http://github.com/charnley/rmsd (accessed 4/24/2019).

165. Yokoyama, S.; Watanabe, T.; Murao, K.; Ishikura, H.; Yamaizumi, Z.; Nishimura, S.; Miyazawa, T., Molecular Mechanism of Codon Recognition by Transfer-RNA Species with Modified Uridine in the 1st Position of the Anticodon. *Proc. Natl. Acad. Sci. U.S.A.* **1985,** *82*, 4905-4909.

166. Smith, W. S.; Sierzputowskagracz, H.; Sochacka, E.; Malkiewicz, A.; Agris, P. F., Chemistry and Structure of Modified Uridine Dinucleosides Are Determined by Thiolation. *J. Am. Chem. Soc.* **1992,** *114*, 7989-7997.

167. Hawkinson, S. W., Crystal and Molecular-Structure of 2-Thiouridine. *Acta Crystallogr. B* **1977,** *33*, 80-85.

168. Shukla, M. K.; Leszczynski, J., Electronic Transitions of Thiouracils in the Gas Phase and in Solutions: Time-Dependent Density Functional Theory (TD-DFT) Study. *J. Phys. Chem. A* **2004,** *108*, 10367-10375.

169. Hamlow, L. A.; He, C. C.; Devereaux, Z. J.; Roy, H. A.; Cunningham, N. A.; Soley, E. O.; Berden, G.; Oomens, J.; Rodgers, M. T., Gas-Phase Structures of Protonated Arabino Nucleosides. *Int. J. Mass Spectrom.* **2019,** *438*, 124-134.

170. Hamlow, L. A.; Devereaux, Z. J.; Roy, H.; Cunningham, N.; Berden, G.; Oomens, J.; Rodgers, M., Impact of the 2´-and 3´-Sugar Hydroxyl Moieties on Gas-Phase Nucleoside Structure. *J. Am. Soc. Mass Spectrom.* **2019,** *30*, 832-845.

171. Rodgers, M. T., Substituent Effects in the Binding of Alkali Metal Ions to Pyridines Studied by Threshold Collision-Induced Dissociation and Ab Initio Theory: The Methylpyridines. *J. Phys. Chem. A* **2001,** *105*, 2374-2383.

172. Rodgers, M. T.; Stanley, J. R.; Amunugama, R., Periodic Trends in the Binding of Metal Ions to Pyridine Studied by Threshold Collision-Induced Dissociation and Density Functional Theory. *J. Am. Chem. Soc.* **2000,** *122*, 10969-10978.

173. Amunugama, R. Noncovalent Interactions Between Metal Ions and Model Systems for Biologically Relevant Molecules: Threshold Collision-Induced Dissociation and Theoretical Studies. Wayne State University, 2002.

174. Moision, R. M.; Armentrout, P. B., An Electrospray Ionization Source for Thermochemical Investigation with the Guided Ion Beam Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2007,** *18*, 1124-1134.

175. Carpenter, J. E.; McNary, C. P.; Furin, A.; Sweeney, A. F.; Armentrout, P. B., How Hot are Your Ions Really? A Threshold Collision-Induced Dissociation Study of Substituted Benzylpyridinium “Thermometer” Ions. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1876-1888.

176. Shaffer, S. A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D., An Ion Funnel Interface for Improved Ion Focusing and Sensitivity Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1998,** *70*, 4111-4119.

177. Shaffer, S. A.; Tolmachev, A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D., Characterization of an Improved Electrodynamic Ion Funnel Interface for Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1999,** *71*, 2957-2964.

178. Aristov, N.; Armentrout, P. B., Collision Induced Dissociation of Vanadium Monoxide Ion. *J. Phys. Chem.-US* **1986,** *90*, 5135-5140.

179. Hales, D. A.; Armentrout, P. B., Effect of Internal Excitation on the Collision-Induced Dissociation and Reactivity of Co_2^+. *J. Clust. Sci.* **1990,** *1*, 127-142.

180. Dalleska, N. F.; Honma, K.; Sunderlin, L. S.; Armentrout, P. B., Solvation of Transition Metal Ions by Water. Sequential Binding Energies of $M^+(H_2O)_x$ (x = 1 - 4) for M = Ti - Cu Determined by Collision-Induced Dissociation. *J. Am. Chem. Soc.* **1994,** *116*, 3519-3528.

181. Teloy, E.; Gerlich, D., Integral Cross Sections for Ion-Molecule Reactions. 1. The Guided Beam Technique. *Chem. Phys.* **1974,** *4*, 417-427.

182. Daly, N. R., Scintillation Type Mass Spectrometer Ion Detector. *Rev. Sci. Instrum.* **1960,** *31*, 264-267.

183. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization - Principles and practice. *Mass Spectrom. Rev.* **1990,** *9*, 37-70.

184. Yamashita, M.; Fenn, J. B., Electrospray ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.-US* **1984,** *88*, 4451-4459.

185. Chen, F.; Gülbakan, B.; Weidmann, S.; Fagerer, S. R.; Ibáñez, A. J.; Zenobi, R., Applying Mass Spectrometry to Study Non-Covalent Biomolecule Complexes. *Mass Spectrom. Rev.* **2016,** *35*, 48-70.

186. Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C., Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** *246*, 64-71.

187. Finehout, E. J.; Lee, K. H., An Introduction to Mass Spectrometry Applications in Biological Research. *Biochem. Mol. Biol. Edu.* **2004,** *32*, 93-100.

188. John R. Yates, I., Mass Spectral Analysis in Proteomics. *Annu. Rev. Bioph. Biom.* **2004,** *33*, 297-316.

189. Chen, Y.; Rodgers, M. T., Structural and Energetic Effects in the Molecular Recognition of Protonated Peptidomimetic Bases by 18-Crown-6. *J. Am. Chem. Soc.* **2012,** *134*, 2313-2324.

190. Amunugama, R.; Rodgers, M. T., Absolute Alkali Metal Ion Binding Affinities of Several Azines Determined by Threshold Collision-Induced Dissociation and Ab Initio Theory. *Int. J. Mass Spectrom.* **2000,** *195/196*, 439-457.

191. Chen, Y.; Rodgers, M. T., Structural and Energetic Effects in the Molecular Recognition of Amino Acids by 18-Crown-6. *J. Am. Chem. Soc.* **2012,** *134*, 5863-5875.

192. Chu, Y.; Yang, Z.; Rodgers, M. T., Solvation of Copper Ions by Acetone. Structures and Sequential Binding Energies of $Cu^+(acetone)_x$, x = 1-4 from Collision-Induced Dissociation and Theoretical Studies. *J. Am. Soc. Mass Spectrom.* **2002,** *13*, 453-468.

193. Wu, R. R.; Rodgers, M. T., Tautomerization Lowers the Activation Barriers for N-Glycosidic Bond Cleavage of Protonated Uridine and 2'-Deoxyuridine. *Phys. Chem. Chem. Phys.* **2016,** *18*, 24451-24459.

194. Yang, B.; Rodgers, M. T., Base-Pairing Energies of Proton-Bound Heterodimers of Cytosine and Modified Cytosines: Implications for the Stability of DNA i-Motif Conformations. *J. Am. Chem. Soc.* **2014,** *136*, 282-290.

195. Yang, Z.; Rodgers, M. T., Influence of Methylation on the Properties of Uracil and its Noncovalent Interactions with Alkali Metal Ions. Threshold Collision-Induced Dissociation and Theoretical Studies. *Int. J. Mass Spectrom.* **2005,** *241*, 225-242.

196. McNary, C. P. Gas-phase Thermodynamics of Hydrazine Containing Compounds and Iron Cluster Cations: Guided Ion Beam and Theoretical Investigations. University of Utah, 2017.

197. Tutorial: State Machines. http://www.ni.com/tutorial/7595/en/ (accessed 4/25/2019).

198. Queued Message Handler Template documentation. http://www.ni.com/tutorial/53391/en/ (accessed 4/25/2019).

199. Kaufman, H. E., Epithelial Erosion Syndrome - Metaherpetic Keratitis. *Am. J. Ophthalmol.* **1964,** *57*, 983-987.

ABSTRACT

MODIFICATION OF BRUKER AMAZON ETD AND SOLARIX MASS SPECTROMETERS TOWARDS INFRARED MULTIPLE PHOTON DISSOCIATION: STRUCTURAL CHARACTERIZATION OF MODIFIED NUCLEOSIDES

by

LUCAS A. HAMLOW

December 2019

Advisor: Dr. Mary T. Rodgers

Major: Analytical Chemistry

Degree: Doctor of Philosophy

Infrared ion spectroscopy has become an increasingly powerful tool for examining the intrinsic structures of gas-phase ions. Infrared multiple photon dissociation (IRMPD) action spectroscopy has been particularly successful. Several free electron laser (FEL) facilities across the world have helped facilitate the growth of the IRMPD technique and increasing interest has driven the development of more accessible IRMPD instrumentation. The development of one such IRMPD instrumentation system is described in this work, based around commercially available 3D quadrupole ion trap mass spectrometers and Fourier-transform ion cyclotron resonance mass spectrometers.

The intrinsic gas-phase structures of nucleic acid monomers have been extensively studied by IRMPD and complimentary theoretical approaches. These studies have examined the common DNA and RNA nucleobases in several states of ionization. The common DNA and RNA nucleosides have also been extensively examined in several ionization states by several research groups. Studies of nucleotides have examined an even more diverse set of ionization states. Several studies have also examined the impact

of specific modifications on the intrinsic structures of specific nucleic acid monomers. However, the incredible wealth of available synthetic and naturally occurring modifications to these monomers still presents a substantial challenge in understanding the relationship between structure and function for modified nucleic acid monomers.

Thiation of uridine at the 2- or 4- position are important modifications for tRNA. 4-thiouridine is a naturally occurring modification in tRNA that is thought to offer some protection against near-UV exposure by cross-linking with a nearby cytidine. Whereas 2-thiouridine and 2-thiouridines further modified at the 5-position can be found at the wobble position of the tRNA anticodon. The altered base-pairing of 2-thiouridine and the modified 2-thiouridines is important to recognition of the codon on an mRNA. Previous study of the protonated 2-thiouracil $[s^2Ura+H]^+$ and 4-thiouracil $[s^4Ura+H]^+$ indicate a change in the protonation preference vs protonated uracil $[Ura+H]^+$. IRMPD action spectroscopy experiments in both the IR fingerprint region and hydrogen-stretching region are performed on protonated 2-thiouridine $[s^2Urd+H]^+$ and 4-thiouridine $[s^4Urd+H]^+$. Complimentary molecular dynamics simulations are used to explore the conformational space available to the protonated thiouridines and generate candidate structures. Density functional theory calculations at the B3LYP/6-311+G(2d,2p)//B3LYP/6-311+G(d,p) level of theory further optimize the candidate structures, predict their IR spectra, and calculate reasonably accurate energetics. Comparison of the measured IRMPD action spectra and the predicted IR spectra reveal the preferred conformations of $[s^2Urd+H]^+$ and $[s^4Urd+H]^+$ and those populated in the experiments. Protonation of 2-thiouridine prefers formation of the 2-sylfhydryl-4-hydroxyl, with some O4 protonated conformers present. Whereas protonation of 4-thiourudine prefers protonation at S4, with a minor contribution from 2-

hydroxyl-4-sulfhydryl tautomers. A mixture of C2′-*endo* and C3′-*endo* sugar puckering is observed, with *anti* oriented nucleobases preferred.

Common targets for modification of pharmaceutically active nucleoside analogues are the 2′- and 3′-hydroxy moieties on the sugar. Nucleosides with an arabinose sugar moiety are some of the oldest nucleoside analogue drugs. The arabinose analogues of cytidine, araCyd, and adenosine, araAdo, have both found use pharmaceutically. The arabinose sugar moiety inverts the stereochemistry at the 2′-position vs ribose, and previous studies by NMR, crystallography, and theoretical calculations are not consistent on the impact of this modification on nucleoside structure. Nucleosides based upon the 2′,3′-dideoxyribose sugar moiety are also common pharmaceutically. Understanding the impact of these two modifications on intrinsic nucleoside structure is important to understanding the basis for their pharmaceutical activity. IRMPD action spectroscopy experiments were performed for the protonated arabinose analogues of the common RNA nucleosides adenosine, guanosine, cytidine, and uridine. IRMPD experiments were performed for the protonated 2′,3′-dideoxyribose analogues of these RNA nucleosides as well as the analogue of thymidine. Complimentary theoretical calculations explore the conformational space of these ions to generate candidate structures and predict their IR spectra and energetics. Comparison of these predicted IR spectra and the experimental IRMPD spectra reveals the conformations that contribute to the experiments. An intramolecular O2′H···O5′ hydrogen-bonding interaction unique to the arabinose analogues is observed for each of the arabinose analogues alongside conformations parallel to those observed previously for the DNA nucleosides. This unique intramolecular hydrogen-bonding interaction strongly prefers C2′-*endo* sugar puckering, which does

have some impact on the overall sugar puckering preferences, but C3′-*endo* sugar puckering is also observed experimentally for the protonated arabinose analogues. Comparison of the experimental IRMPD spectra of the 2′,3′-dideoxyribose analogues reveals a stronger preference for C3′-*endo* sugar puckering. Otherwise, largely parallel conformations are observed between the protonated 2′,3′-dideoxyribose nucleosides and the previously studied protonated 2′-deoxyribose nucleosides.

AUTOBIOGRAPHICAL STATEMENT

Lucas A. Hamlow

EDUCATION

Ph.D. Analytical Chemistry **September 2012 – September 2019**

Wayne State University

Ph.D. Mentor: Professor Mary T. Rodgers

B.A. Chemistry **September 2007 – July 2011**

Wabash College

AWARDS and HONORS–Wayne State University

(1) Best Paper Award AnaChem/SAS Symposium, 2013

(2) Departmental Citation for Excellence in Teacher Service, 2014

(3) Schaap-Rumble Graduate Research Fellowship, 2015

(4) Dan Trivich Memorial Award for Research in Physical Chemistry, 2017

(5) Departmental Graduate Student Professional Travel Award, 2018

PUBLICATIONS–Wayne State University

(1) **L.A. Hamlow**, Y. Zhu, Zachary J. Devereaux, N.A. Cunningham, G. Berden, J. Oomens, and M.T. Rodgers, *J. Am. Soc. Mass Spectrom.* 2018, 29, 2125-2137

(2) **L.A. Hamlow**, C.C. He, Zachary J. Devereaux, H.A. Roy, N.A. Cunningham, E.O. Soley, G. Berden, J. Oomens, and M.T. Rodgers, *Int. J. Mass Spectrom.*, 2019, 438, 124-134

(3) **L. A. Hamlow**, Zachary J. Devereaux, H. A. Roy, N. A. Cunningham, G. Berden, J. Oomens, and M. T. Rodgers *J. Am. Soc. Mass Spectrom.* 2019, 30, 832-845